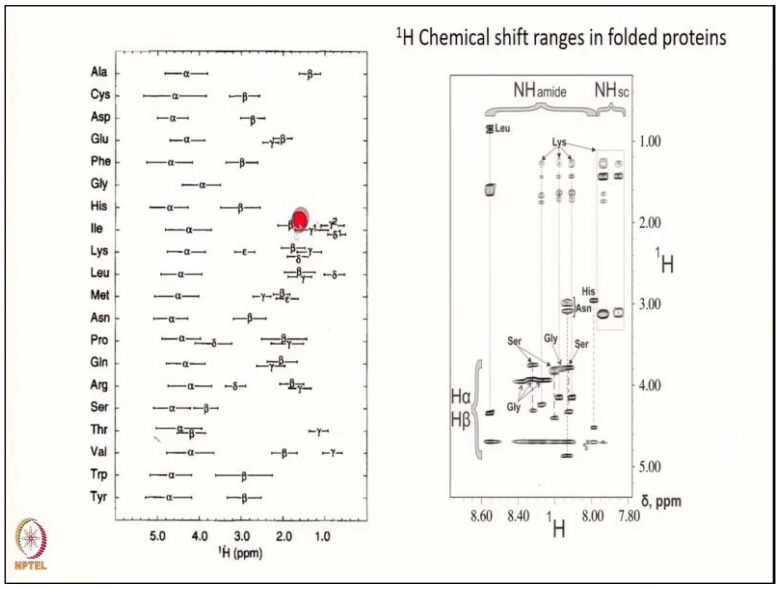
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 33
Determination of Structure and Dynamics of Proteins - 3

So, just a brief recap of the previous class we talked about the 20 different amino acids and what are the chemical shift ranges of the different protons in the different amino acids.

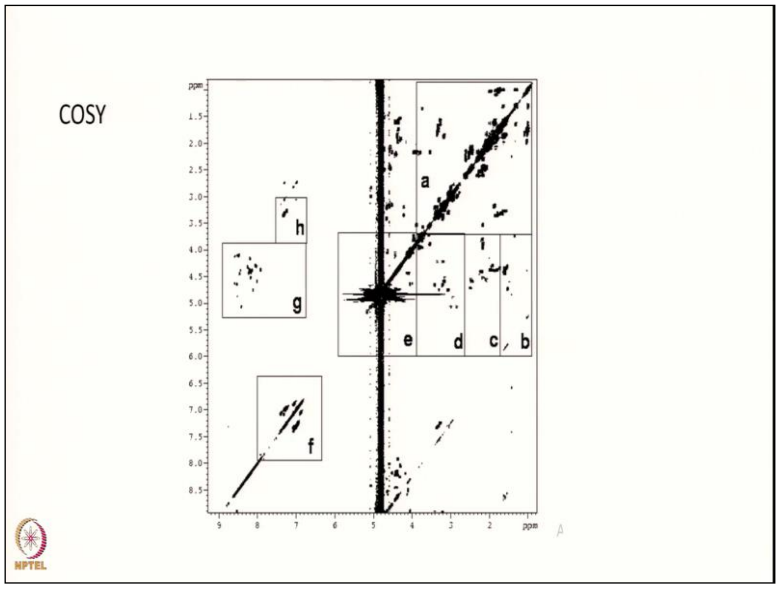
(Refer Slide Time: 00:33)



These are indicated in this plot here. For the different amino acids, you have different ranges of chemical shifts for the different types of protons. So, we also talked about the different proton types and the nomenclature as was discussed in the previous class. And these show up in different kinds of 2-dimensional spectra a typical one is shown here. This is the chemical shifts of the TOCSY spectrum of the folded protein.

So, here only one particular region is shown. So, the amide protons to the various alpha protons and the beta protons and things like that and also the side chains here.

(Refer Slide Time: 01:13)



So, a whole COSY spectrum is shown like this here. So, this is the area of the amide protons and the aromatic protons, and you have the correlations in the COSY spectrum to the alpha protons here and then you also have from the aromatic protons to its beta protons and so on. Then of course when you go here then you will have the correlation from the alpha protons to the beta protons and then from the beta to the gamma, gamma to the delta and things like that.

So, that is how the connectivities of the various protons are established in the COSY spectrum. The same thing will be available in the double quantum filtered COSY. TOCSY goes a little bit further and it establishes a relay of information in the spin system. So, here is a comparison of a particular segment of the COSY spectrum which is showing you the five peaks here, these five peaks are also present here in the TOCSY spectrum but it shows a relay to other protons and that is indicated by the different peaks which are present at the same chemical shift here.

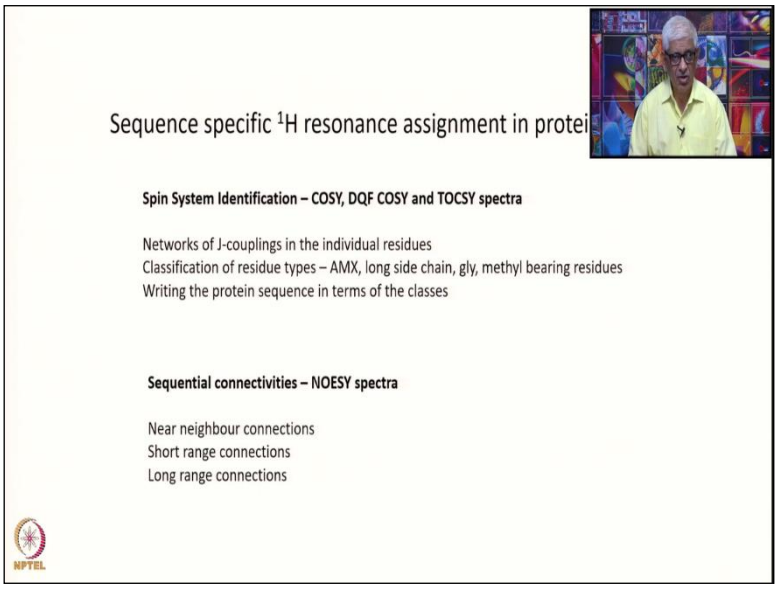
For example, for this particular peak here, there are others which are present here which shows the correlation to the other protons which are connected to this proton through a network of coupled spins. Similarly, for this one, you also have a correlation here. So, like that for every proton, there is a relay that happens in the TOCSY spectrum that establishes the spin system network.

And this is another region of a particular spectrum. You have the COSY here and this is the TOCSY spectrum here and you can see a full network of correlations in the TOCSY spectrum. Only near-neighbour interactions, the three bond couples, or the 2 bond coupled ones are shown

in the COSY spectrum. In the TOCSY spectrum, you have a whole range of connectivities from the alpha to the beta and then alpha to the gamma alpha to the delta, and so on.

That will appear because of the relay that happens in the TOCSY spectra this is what we have seen earlier when we discussed the methods.

(Refer Slide Time: 01:13)



So, now we come to the next step which is the information in the J correlated spectra. So, what do we use ultimately our objective is to obtain the so-called sequence-specific proton resonance assignment in proteins. You will have to be able to assign the individual protons to individual residues. There are different types of residues, which we discussed there. So, we have to establish the networks of J couplings in the individual residues.

This comes from the COSY, double quantum filtered COSY or the TOCSY spectra this is known as the spin system identification. What is the meaning of the spin system, so, let me also just briefly give a recap of that one. So, if you have a three-spin system you may have a system like this. You may say this is the AMX or if you have a linear system AMXQ and so on.

These are called the spin systems, there are couplings at each stage, there is a coupling here, coupling here, and a coupling there, and here, there will be coupling there. There is coupling there. So, all of them are in that way connected. So, therefore this is known as what type of spin system is present in your spectra? So, if you have a protein chain which consists of various amino acid residues, for example, if I want to write an amino acid chain as ATGCLQ, and so

on. Each one of them has a particular type of spin system connectivities and that is called as the spin system identification and this is known from the COSY and double quantum filtered COSY. So, many of these amino acid residues may belong to a similar type of spin system. That is why we need to classify these amino acid residues as AMX systems or long-side chains.

So, you have the classification of the residue types AMX type or that means it will be something like this here, or the long side chain it may be something like this here, or the glycine which does not go too far, or there are some residues which contain the methyls these are typically the leucine, isoleucine, alanine, and threonine, valine. These are the methyl-bearing residues. They produce characteristic patterns in the COSY and the TOCSY spectra.

So, therefore, from the COSY and the TOCSY spectra, what we do is, with a particular chemical shift what type of a spin system is present, whether it is an AMX system or an AMXQ type of system or AMXQR kind of a spin system, how many protons are connected in the network. So, that is what you establish in this. The long side-chain means there are many protons, which are connected in the network. Glycine has only 2 protons.

And the methyl-bearing residues also may have long side chains. All of these can come inside this network of spin systems and we classify the amino acid residues in that. Then after we have done that we do not read it like a polypeptide, we will read it as a spin system sequence. There is the three-spin system, long side chain sequence, glycine, and then of course again this long side chain, again a small spin system, and so on.

So, you will categorize them based on the spin systems. Therefore we write the protein sequence in terms of the classes. There will be several classes as we indicated here. These are the types of amino acid residues. Immediately we do not know which residue is where. So, you may have several glycines in your polypeptide chain, several phenylalanines, there will be several tyrosines, all of which belong to a particular type of residue or the spin system.

Therefore all of them belong to the same, whether you have glycine, whether you have tyrosine or phenylalanine, they belong to the same type of spin system, or you have threonine there is a different spin system, alanine is a different spin system. And likewise the long side chain ones. Suppose you have arginine or lysine or glutamate and glutamine, these have long side chains.

So, they will classify them as long-side chains. You will write the protein sequence in terms of the classes, that is the first step of information obtained from the COSY, DQF-COSY and TOCSY spectra. Because these display the J coupling correlations. Notice here of course that we only have these things from there within the same amino acid residues.

These do not connect one residue to the next residue on either side. Therefore, within the same amino acid residue, what sort of spins systems are present and what are the correlations established? So, from the COSY spectrum, you identify that, this may be an AMX spin system, this is a glycine spin system, or this may be again an AMX spin system and like that.

So, at the particular chemical shift, I will put this as an AMX spin system or a long side chain spin system or the methyl bearing residue spin system etc. So, this is how you obtain this information from the TOCSY spectra. So, this is the COSY and you compare that with the TOCSY here you get the connectivities and on the basis of that you say now here it is a system which has a long side chain.

Many cross peaks are originating from this particular diagonal peak. Here you have a peak, a peak there, and a peak there, and a peak there. So, therefore many correlations are present which means that the particular spin system is a long-side chain one. So, it has the alphas the betas and the gammas and the deltas and so on. So, you first classify your polypeptide chain along these lines. Here it is a long side chain, short side chain, methyl-bearing, AMX spin system and so forth.

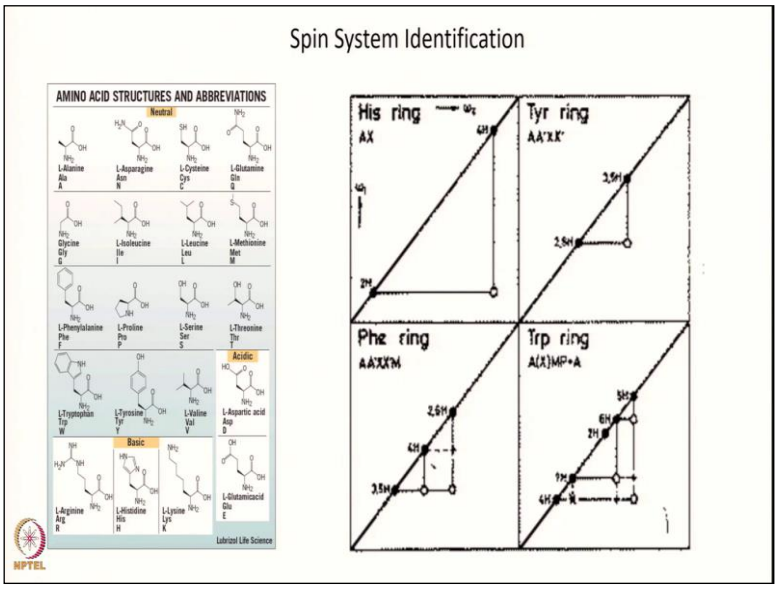
So, that is what is called writing the protein sequence in terms of the classes. Next, we have to connect these spin systems sequentially. So, we have these neighbouring residues, near-neighbour connections. These near-neighbour connections do not come from the COSY spectra or the TOCSY spectra, there are no J couplings there and there are also no relays. So, therefore this information will not be obtainable from the J-correlated spectra.

So, we have to resort here to the NOESY spectra. The NOESY spectrum displays distance correlations and there are short distances between the near-neighbours. There can be short-range connections over a longer distance, in the sense that three residues 4 residues apart or there can also be long-range connections which are because of the folding of the protein. There may be things which are coming which are 10 residues, 15 residues, and 20 residues apart.

And you may have that sort of connection also, because the protons come close by in space and that is what the NOESY spectrum displays. the NOESY spectrum displays those proton-proton correlations which have short distances, less than five angstroms. So, therefore we have to distinguish all three different types of connections here. First thing is to use the near neighbour connections to identify the sequential connectivities.

So, individually you walk from one residue to the next residue along the polypeptide chain. So, these are the 2 steps which are there and then once you do that you obtain the sequence-specific resonance assignments in proteins.

(Refer Slide Time: 10:36)



Now, how do the spectra look like for the different amino acid residues? This depends upon the chemical shift ranges. Now here again I have put in the various amino acid structures and the corresponding type of correlation you will see in the COSY or the TOCSY spectrum. For example, for the histidine ring. This has 2 protons, there is a proton here, and there is a proton there. There can be a correlation between the amide proton to this proton, you will have a correlation from this proton to this proton, and you will see that sort of a system. So, you will have the 2H and the 4H, which are the 2H and 4H? This is the 2H, this is the 4H.

So, when you are recording the spectra in D2O, the amide protons do not show up or the imino protons or the NH protons do not show up because they all exchange with water, exchange with the D2O and they will not be present. So, these spectra are recorded in D2O, so when we do that, then of course you will see the correlation from this proton, both these protons are

attached to the carbon C2H and C4H. Those are these 2 positions and you will see a correlation between them.

So, this will become like an AX spin system. Similarly, tyrosine is among the aromatic residues, and we have this is phenylalanine and threonine. Let us look at where is the tyrosine. Now here is tyrosine, this is Y. Now how many protons are there? So, this OH is of course not visible. There are 4 protons.

There typically these 2 may be equivalent and these 2 may be equivalent, but they may not be also. So, if they are not equivalent then you write it as AA'XX'. But if they are equivalent. If they are equivalent then they will say this is in A2 and if these are equivalent they are going to say X2. So, you do it can be an A2X2 spin system or if they are non-equivalent then they may be AA'XX' spin system.

And the non-equivalence can arise in the protein structure where there is a free rotation is restricted and therefore they may see different environments and their chemical shifts can be different. So, what is drawn here is when the 2 are equivalent, their chemical shifts are the same but their coupling constant may be different. So, though you may write in general as AA'XX' to be more general.

So, the chemical shifts of these 2 are the same therefore they are typically represented as one chemical shift here and another chemical shift there because you use the same symbol A. So, the chemical shifts are the same but the coupling constants may not be the same that is why you say to make them magnetically non-equivalent but chemically equivalent. When you have that of course there is only one coupling. So, the coupling is from this proton to this proton and that is this cross peak. You will have from the aromatic ring, you will only have this; but of course in this in the backbone, you will have the beta protons also.

The alpha protons will show to beta 1 and beta 2, what is shown here is only the ring protons. How are they connected in the ring? These are special. These are aromatic protons that are special they have a ring and therefore what is shown here is the connectivity in the ring. Now if you go to the phenylalanine that is phenylalanine is here. Now, this has five protons there.

Once again there can be equivalence hear, and equivalence there, but the third one will not be equivalent. So, therefore you have here 2-6, 3-5 and 4. So, the 4 is the central fellow here and the other ones are 2-6 and 3-5 and you will see correlations from 3-5 to 2-6. So, you will see this cross peak 3-5 to 2-6. This cross peak is the prominent one. These are the 3-5 and those show to the 4. Therefore you see that cross peak here. You do not see 2-6 to 4. 2-6 are these, and you may not see the cross peak here to the 4. So, therefore you do not see the 2-6 to 4 but you will see from 3-5 to 4 and 3-5 to 6. What is shown here is on only one side of the diagonal, but these peaks will also be present on the other side. See this 2-6 to 4 it can be a very weak one this may not be present.

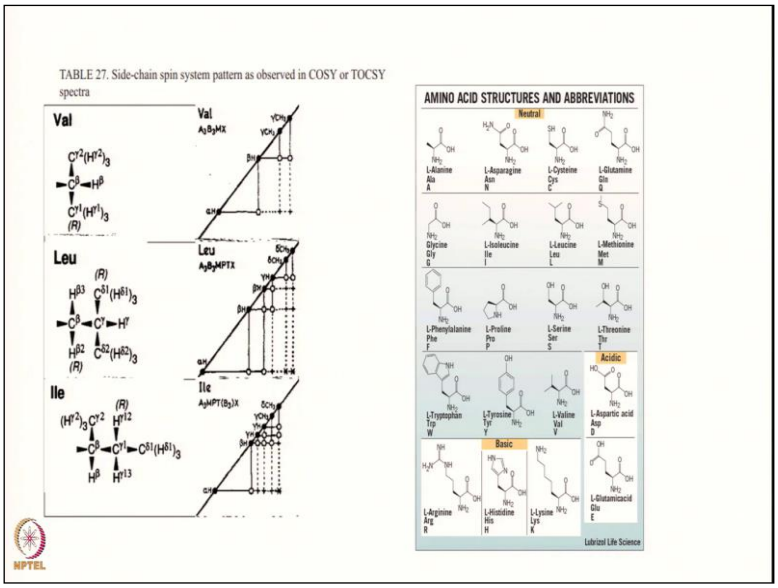
But it may be seen in the TOCSY spectra, these may not be seen in the COSY spectra. Therefore this spin system becomes AA'XX' and M, and that is the one, the single fellow which is the 5 and that will be this and the 4 this will be this M and these 2 are the AA'XX'. These are the 3-5 and 2-6 and this one is your M.

And similarly, if you go to the tryptophan ring, the tryptophan ring is this one here and you have the correlations of these protons there, big empty circles are the ones which are seen in the COSY. So, you will see from 5 to 6 and then you will see from 5 to 4 then you will see from 1 to 6 and that is the nomenclature which if you remember that is represented them as the various alpha-beta gamma etc but within the ring, the nomenclature goes by the numbers in this one. So, you will see this sort of pattern for tryptophan.

You will see these three peaks in the COSY spectrum but in the TOCSY spectrum, you may see these intermediate peaks as well. You may see this peak and those will be present in the case of tryptophan. These 2-protons, this one stands alone and does not have any coupling to anybody. So, that is a single peak which is which will not be seen.

Therefore you will not see any correlation from this proton to any other proton, and all others are in the six-membered ring and you will see correlations between the protons in the six-membered ring.

(Refer Slide Time: 17:13)



Now if we go further to the other side chains. So, you have here valines the side chains are already indicated here. So, you have this say γ1 γ2. So, this is the valine. Valine has this there is a methyl here. So, you will see this sort of a pattern, beta to the 2 gammas. So, it is coming from the alpha, the alpha proton is here on the side.

Only the side chain beyond the alpha proton is shown. You have one beta proton, that is attached to 2 methyls, gamma 1 and gamma 2. These are the 2 methyls, this is NH-C alpha - COOH and this is C alpha this is the C beta and that C beta there is one proton there and it is connected to the 2 methyls there. So, that is shown here. So, you have the beta proton there and these are the gamma 1 and gamma 2 and this is the alpha proton.

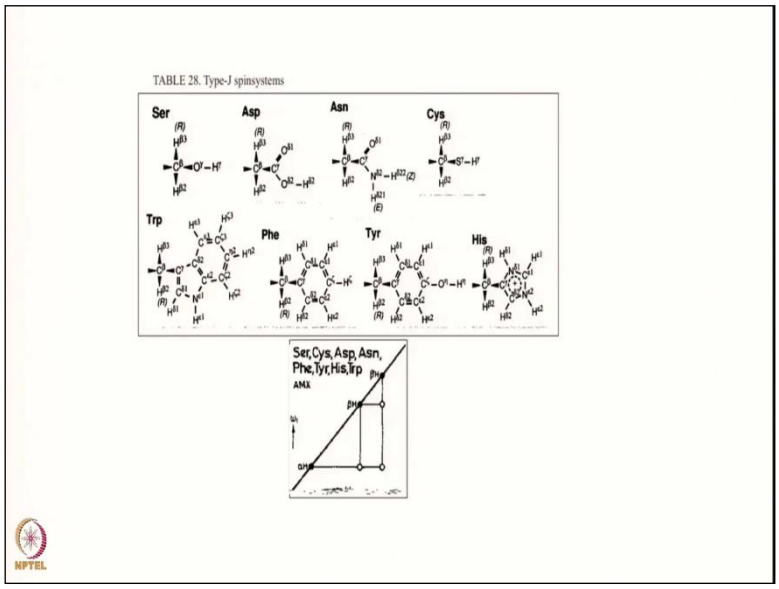
Now from the alpha proton, you are seeing to the beta proton, that is this. The alpha proton is here. Onto the C alpha which is present there to the beta proton, you see this peak here and from the beta proton you see to the 2 methyls. These correlations you will see from the beta proton to the 2 methyls here. And you do not see between the 2 gamma methyls because there is no coupling there therefore you will not see that cross peak.

Leucine if you see, again it has the alpha to the beta and then beta there are 2 beta protons you see to both the beta protons: the beta 1 and beta 2, and then from the beta as you see to the gamma, there is one gamma proton there. So, in the gamma, you see to 2 methyls. There are 2 delta methyls So gamma proton you see to the two methyls you see to these peaks there.

And of course, if you are looking at the TOCSY, the dashed lines and dotted lines indicate that you may see these additional peaks when you go to the TOCSY spectra. In the COSY you will see only these or in the double quantum filtered COSY you will see only these, empty circles which are indicated there. So, you will see these cross-peaks and if you go to the TOCSY you might see these individual correlations as well.

So, similarly, isoleucine has this network of couplings there, you have the from the alpha proton to the beta proton from the beta to the gamma one and gamma 2, and these gamma protons are connected to the methyl to the delta and therefore you will see from both the gamma protons you see to the methyls. So, that is how this spin system goes for these amino acid residues.

(Refer Slide Time: 20:05)



And likewise, the serine and others, are called the ABX spin systems or the AMX spin systems, because this side chain has only alpha to beta. That is alpha to beta, there is nothing beyond the beta proton. For example, you can see here, serine has alpha to beta and there is nothing and these ones will not be seen. The aspartate, is similar, to the beta and you do not have any other coupling there.

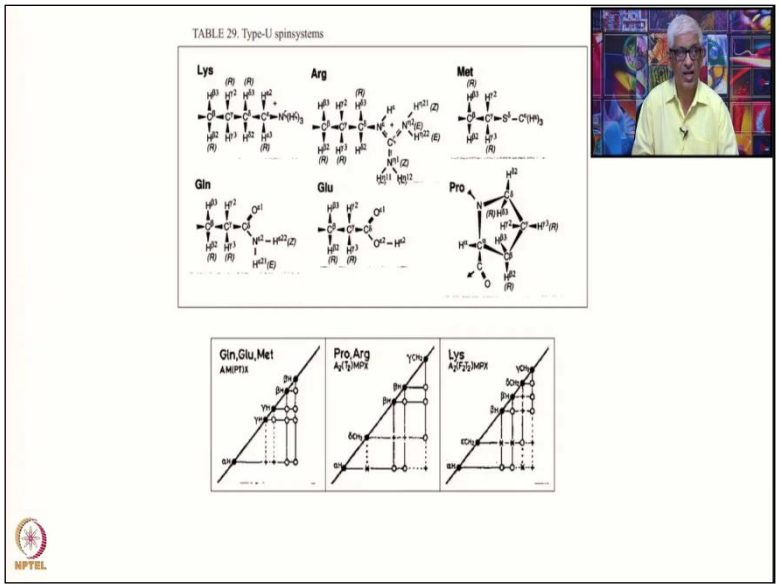
These ones are gone asparagine, this alpha to the 2 beta protons. So, an alpha proton is there on the side when the C alpha is there and so similarly the cysteine alpha to the 2 betas. Here the tryptophan, this alpha-beta proton region is also shown. The aromatic ones also have this alpha to the beta, we talked earlier about the rings here, we only talk about what is present within the ring. Now you see including the alpha you will see alpha to the betas, the 2 beta protons and so on.

And there is no coupling from the beta protons to the aromatic protons, therefore, you will not see a cross peak from the aromatic protons from the beta pronouns to the aromatic proton these appear separately and they can have to be identified differently. Now all these many 8 amino acid residues we classify as AMX spin systems or sometimes ABX spin systems.

So, because there are three spin systems so, wherever this serine, cysteine, or asparagine is present, you may call this kind of particular spin system called the J-spin system. It is represented as J J J J wherever this is present you write it as J and the long side chain is presented differently. So, you have this sort of spin system. And what you have here is only alpha to the beta, beta 1 and beta 2 and then you also have a cross peak between the beta 1 and the beta 2 protons.

So, this is only one side of this diagonal shown and you have this cross peak. The TOCSY and the COSY do not make a difference at all because there is a beta 1 and beta 2 and there is no other relay. So, there is a direct coupling here and all of these peaks can be seen in the COSY, DQF-COSY and TOCSY will be identical.

(Refer Slide Time: 22:15)



Now, this is another type of spin system as indicated these are the long side chains. So, from the alpha, you go to the beta, beta to the gamma, gamma to the delta, and then delta to the epsilon and things like that. So, there are long-side chain ones. So, how many are there? Lysine, arginine, methionine, glutamine, glutamic acid, and proline are called the long-side chain ones.

Because the network goes quite long you go from the alpha you see to the 2 betas, 2 betas to the 2 gammas, 2 gammas to this one, there of course you may not go beyond here something. These are the amino protons and these will exchange out and you will not see. So, it will be seen only up to this point. If you see lysine, from the alpha to the 2 betas, 2 betas to the 2 gammas, then to the 2 deltas, and then from there you go to the 2 epsilons and after that, you do not see.

Because this one goes to the NH group and this will be exchanged out in the D2O spectra. Arginine is similar, alpha to the 2 betas, betas to the 2 gammas, gammas to the 2 deltas, and after that you do not see. So, and therefore lysine is one more non-exchangeable proton compared to arginine. Whereas arginine has more exchangeable protons these will be seen more in the H2O spectra. There also you will see one of them that is typically here but these ones also typically exchange out in the water because of the exposure.

And the methionine has alpha to the 2 betas, 2 betas to the 2 gammas, and then, of course, you do not have any other thing there, there is a CH_3 group there, but it has no coupling to any of this one this is the long 4-bond and you will not see a correlation between these here and the glutamine, glutamine has once again alphas to the 2 betas, beta to the 2 gammas, and then after that, you do not see anything there and because these are exchanging out with the water. The glutamate and glutamate are similar. And now the proline is interesting. If you see alpha to the beta, there are 2 betas, 2 betas to the 2 gammas, and the gammas to the deltas here and that is what you have. So, therefore this is a closed ring here. So, and you will therefore because of the ring closure of course its chemical shifts are also different.

And how does it look in the spectrum? The spectrum I am showing you here, the alpha proton region, alpha proton region is here, this is typically alpha to the 2 betas. Now you see that the betas are more up-field compared to the gammas, in the residues, Glutamine, glutamate, and methionine. These betas are up-field compared to the gammas, why is it so, because these gammas are attached to either sulphur or oxygen or nitrogen and therefore they get downfield shifted in these three residues. Therefore the betas are above and the gammas are below.

So, you see from the alpha to the 2 betas, and from both the betas you see to the 2 gammas. Therefore you have 4 peaks there, and you also have the beta-beta cross peak, and now in proline and arginine, you have the alpha to the 2 betas.

And now this one is in the middle, look at the chemical shift range. It is important to notice here how the chemical shifts vary depending on the structure of the side chains there. So, the 2 betas are here, and the gamma is above, 2 betas are connected to the gamma by these 2 cross peaks, and the gamma is connected to the delta which is down here, much closer to the alpha proton here gamma is down here.

So, these characteristic patterns are very interesting same thing happens with the arginine. In the case, why the delta is below? Because they are attached to the nitrogen seen both in the arginine as well as in proline, the delta proton is attached to the nitrogen and therefore this comes much downfield and is closer to the alpha proton. Therefore this pattern is very distinct from the other one there. Although they are all long-side chains,

but the patterns of the peaks are different there. Lysine if you see, the alpha to the betas, which are down here and beta to the 2 gammas, these peaks there, and of course you have the gamma to the delta, which is very close to this one. Often you may not be able to distinguish between these because these are too close and then from the delta to the epsilon.

Now the epsilon is distinct, epsilon comes very far away because it is attached to the nitrogen, and this comes quite downfield there. So, this is how you get a very distinct pattern for this spin system. So, these are called U-pin systems. Because these are long-side chain ones, they are called U-spin systems. So, this is the nomenclature which was adopted in the book by Professor Wüthrich.

And these pictures are taken from the Wüthrich book, which is NMR of Proteins and Nucleic Acids, and that is published quite a long time ago. He was a pioneer in doing all of these assignments and the protein structure determination for which he of course won the Nobel prize in 2002. The next step is to do connections from one residue to another residue and this will come from the NOESY spectrum. I think we will we will stop here when we take it to the next class.