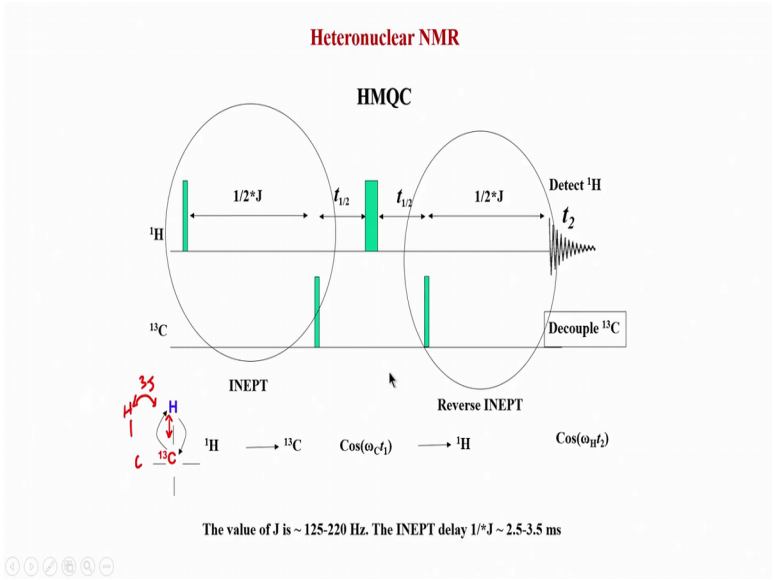
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules
Prof. Hanudatta S Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 14
Heteronuclear multiple quantum coherence (HMQC) and single quantum coherence (HSQC) – Part II

Welcome back. We are looking at HMQC in the last lecture, how the 2D HMQC works; what is the pulse sequence; how does it look like; we will continue with that today before we move on to the next experiment. So, this was the pulse sequence of HMQC which we looked at.

(Refer Slide Time: 00:45)



So, this sees as I said the information wise it is similar to a 2D HSQS experiment which we have seen, similar to that it has 3 different blocks. So, the first block is the inept block which is slightly different from what we saw in HSQS, it does not contain this pulses 180 degrees here, but the idea is similar to HSQS that you transfer the polarization from proton to carbon, using this delay period which is 1 by 2 J and this J is set according to the coupling constant, which we saw that it varies from methyl to aromatics. So, based on this coupling values, we choose a particular delay period and that ensures that the transfer takes place only for 1 bond.

So, we have to keep in mind we are transferring the magnetization from proton to carbon only through a 1 single bond. You are not going for multiple bond transfer which is a different experiment. So, is the second part of the experiment consist of this evolution of the carbon chemical shift, which is excited because of this 90 degrees pulse here and this carbon chemical shift now evolves during this period, and it is decoupled because of the 180 here from the proton.

So, remember we have to decouple proton from carbon and carbon from proton both. So, this is that during the detection period we decouple proton from carbon, but during the indirect period detection period which is indirect dimension, we decouple the proton from carbon and then the reverse in a block comes where you transfer back the polarization to the proton and the proton now starts evolving or getting detected during the t 2 period. So, that is a chemical shift of proton, which evolve some happens during the t 2 period and that is given by cosine omega t 2 and cosine omega t 1 is for carbon.

So, this correlates 2 chemical shifts carbon along t 1 with the directly attached proton along t 2. So, this if you recall is a same kind of an information, which we obtain from HSQC. So, HSQS and HMQC wise there is the information wise there is no difference, but what I would say is that the difference mainly relies on the sensitivity.

So, the HMQC as I said the 180 degree pulses which are not here, make it more sensitive compared to HSQS, because HSQS the 180 degree pulse degrade the intensity and that is because the 180 pulses are not broadband, and they effect the inversion that is 180 degrees are supposed to invert a magnetization similarly 180 degree pulses are suppose invert to for carbon, they are not really uniform over the entire chemical shift range.

So, because of that the all the chemical shifts of carbon are not equally effected by this 180 and at the edges of the spectrum the sensitivity goes down. The HMQC the problem with that is it has is resolution problem because during this period t 1 although I have not shown here, there is also a coupling term which will come along with chemical shift and that if the t 1 is very long that coupling term also becomes dominant.

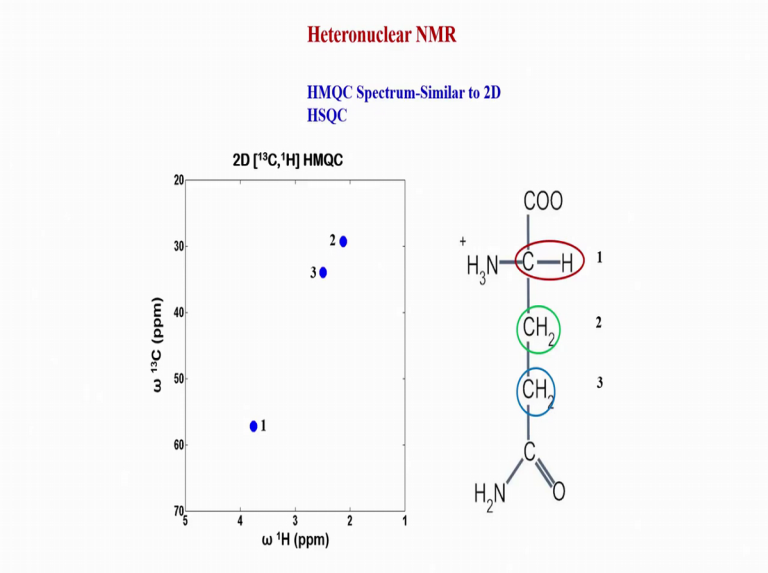
And therefore, that is proton proton coupling now, because that is not carbon proton we have decoupled from carbon, but we have not decoupled 2 protons from each other. So, if there is a proton for example, let me show you here, if there is a proton which is attached to another carbon here this coupling which is 3 bond J coupling is surely active,

it is not decoupled. What we have decoupled is proton to carbon; we have decoupled these 2 interactions, but we have not decoupled this interaction.

So, this coupling and not only to this proton there could be another proton this side. So, there could be several proton proton couplings, which are active which are actually present or revolving during this t 1 period. And that comes into the expression here which all though we are not showing for a simple the simplification and that causes a decrease in the resolution of the spectrum, because the peaks starts now splitting because of the J coupling and that degrades the resolution.

So, this is one disadvantage of HMQC over HSQS. So, therefore, depending on the applications people use either HMQC or HSQS. So, we will see that soon and we will come to biomolecules what type of experiments are carried out.

(Refer Slide Time: 05:17)

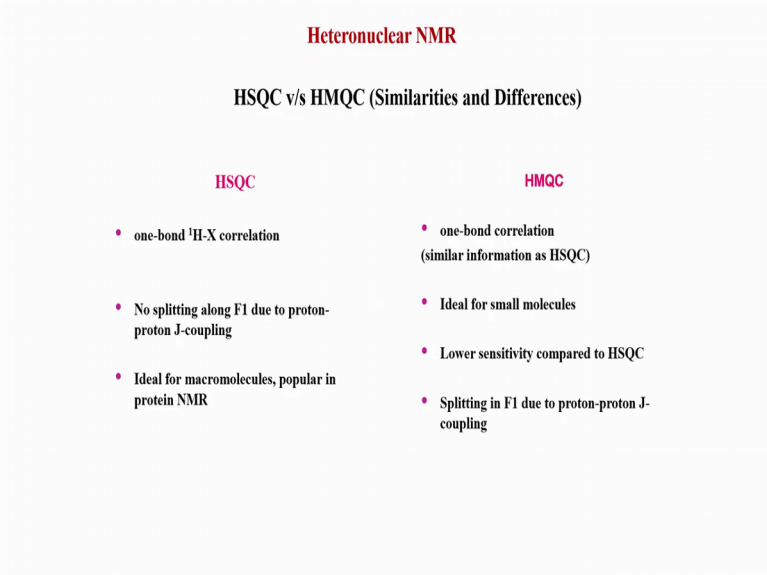


So, this is an example spectrum again schematic drawing of a molecule, this is again glutamine which we saw in the HSQS case, and you can see here again similar to HSQS we have 3 carbon proton pairs labeled as 1 2 and 3. So, 1 pair is 1 CH. So, this CH is coming up here and that is downfield because of this d shielding effect from this groups. It comes typically around 50 to 60 PPM between 50 and 60. Now the second group here pair is CH 2. Now remember CH 2 they this H 2 is not equivalent.

This is something you have to keep in mind because this is a chiral carbon. So, there is an asymmetry in the molecules. So, these are not equivalent, but here as I just to show for simplification, I have shown it as a single peak, but there actually there 2 peaks and the chemical shift of these 2 protons will be either very close or it can be quite different depending on the particular structure in the protein. So, that is rex number 2 and the third pair that is CH 2 gamma this is alpha, this is beta and this is gamma.

In the gamma case you can see now CH 2 again here its coming slightly downfield compared to 2, because of this d shielding effect. But this exactly similar to what we get in HSQS, the chemical shift wise there is no difference a spectral pattern wise there is no difference. The finite difference that is a coupling which I was mentioning is not shown here, but it will also effect in the spectrum.

(Refer Slide Time: 06:57)



So, these are the advantages and disadvantages where I would say similarities and differences between HSQS and HMQC. So, let us go through this once more although we have seen this in the previous slides. Number 1 is HSQS gives you 1 bond, direct 1 bond proton to carbon or proton to nitrogen a proton to phosphorus whichever is X nucleus correlation; similarly here we get the same thing in HSQS in HMQC we get the 1 bond correlation. So, there are these are the similarities.

Here there is no splitting along the t 1 or F 1 dimension that is because of proton to proton coupling and this is something which happens here in case of HMQC. There is

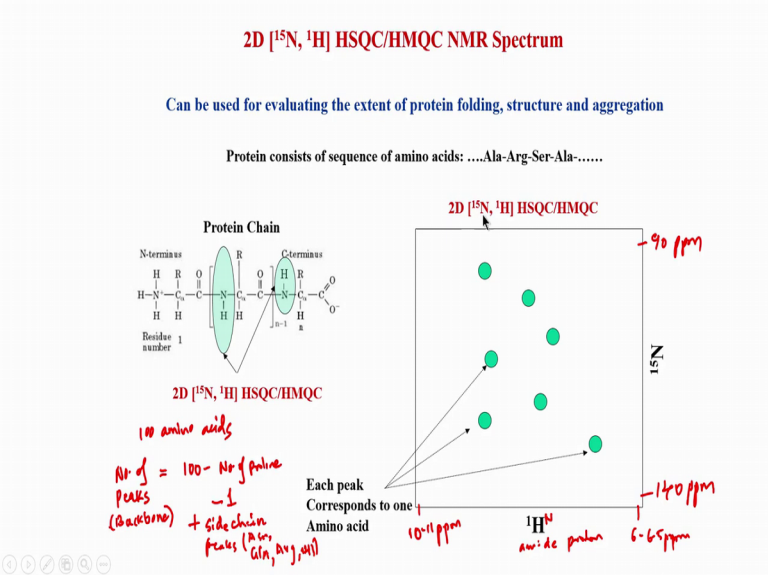
some which is what I told you that if there is a 3 bond coupling between 2 neighboring protons, you will end up seeing a coupling and therefore, that will split the peaks into multi plates similar to what we see in 1 D, but that will happen along t 1. So, that is not we did not see that, but that can happen if you have a very long t 1 value.

Now the third one the advantages as I said is very ideal from micro molecules and is very popularly used in protein NMR. It has a lower sensitivity compared to HSQS, reason being 1 is of course, the degradation because of this coupling and a second thing which I mentioned in the last class was that there is a t 2 relaxation. When you have multiple quantum coherence, HMQC stands for heteronuclear multiple quantum coherence we have sensitivity loss because of t 2 contribution relaxation from the protons.

So, it is ideal for small molecules it is also for proteins its not that it cannot be used. In fact, now a days we have fast NMR experiments and recent years which have been relying direct quing on HMQC. So, HMQC is also very good for biomolecules and if you want to do rapid NMR approaches, then HMQC is way to go.

But when you go for 3 D NMR experiments normally we incorporate HSQS into that and therefore, HSQS becomes more important in such scenarios. So, let me now give you a few examples actually practical examples of how HSQS is used in protein NMR.

(Refer Slide Time: 09:19)



Let us look at this protein structure for a very brief period, this is not a detail structure analysis just a quick overview. So, HSQS or HMQC can be used for evaluating the extent of protein folding structure and aggregation.

So, this is one of the very important application of HSQS, is suppose I you have a protein with you and you want to know whether this protein is folded in 3 dimensional space properly or not, it could be an unfolded protein, it could be intrinsically disorder protein, which does not have any structure or it could be a protein which has got aggregated. So, instead of a monomer you might be saying a dimer even be getting a trimer in your samples.

So, how do we identify those things with NMR? So, this is the application where HSQS become very useful. So, this is something which all of us know that proteins consists of sequence, a linear sequence of amino acids. So, this is just an example here. You have a several; let us say you have several amino acids in a protein you have alanine liquid arginine and this sequence are arbitrary. So, there is a sequence of amino acids and what also we know is that there is a peptide bond between neighboring amino acid.

So, this is a simple schematic drawing of a protein structure, its linear sequence structure is not a 3 dimensional structure this is just showing the sequence of amino acids. So, you can see here there is a residue number 1 is belongs to the N-terminus. Now within the N-terminus of a protein you have NH 3 it is not NH. NH comes only in the peptide bond ok.

So, therefore, the there is a terminal nuclear terminal amino acid is having NH 3 and that NH exchanges rapidly with the solvent, these exchanges very rapidly with the solvent. So, normally you will not be able to capture this NH correlation, but here you have a NH in peptide bond and here as written here there are suppose you have N amino acids, I can write it as 1 and first amino acid and n minus 1 and the last is the nth amino acid now this is the terminal COO carboxylic group.

So, this is again we know this is the last amino acid in the chain. So, it also has a NH, but here it has COO minus. So, now, if I do an experiment called N 15 proton HSQC. So, you can see here this is very specific experiment, where I am correlating the chemical shift of the nitrogen it is this nitrogen with this hydrogen, but what is important here is to note is at this nitrogen is now N 15, which means I am not looking at the natural

abundance and protein I am enriching this molecule I would have enrich using some methods which you will see later down the line.

But assuming that I have isotopically enriched this whole protein with N 15, I will have all the nitrogens in the protein has N 15 and I can do a experiment N 15 proton HSQS or HMQC, where I will can correlate the chemical shift of each nitrogen here with its corresponding directly attached proton ok. So, this is shown schematically in this spectrum here and this is a you can see here, this is HSQS spectrum and you can see this correlations what we are seeing here each of this parts here or peaks, and each of this peek actually correspond to 1 amino acid. So, 1 peek here has 1 nitrogen chemical shift in this axis and 1 hydrogen chemical shift on this axis. So, this is a nitrogen proton pair and that pairs are basically these we can see in this picture.

So, each of the amino acid because as I showed here these are n amino acids. So, if you have N amino acids you have N such combinations, N such pairs in your protein and each of this pair now we will show up in this case ok. So, that is what we will see, but n terminus here this one with is not actually an amide it is an amine and at because of this 3 hydrogen NH 3 exchanges rapidly with the solvent; whereas these hydrogens are not so, exchangeable although they can be exchanged.

But normally enfolded proteins they are involved in hydrogen bonds. So, therefore, they are protected from solvent exchange. So, and therefore, there are the peaks are not so labile are the protons are not so, labile as this protons these are amines. So, an amide proton we are looking at CO NH comes between somewhere between. So, let me write down the chemical shift range, typically it comes from 90 PPM to about 140 PPM.

So, this is what the typical range is, usually it comes somewhere from 100 PPM to 135. So, there are this specific range will come visit again, when we come to the proteins and here typical its about 6 to 6.5 PPM in the proton scale to about 10 to 11 PPM ok. So, this is typically the range of amide protons. So, we are looking at amide protons here. So, I would also specifically write this, this is amide protons. So, this is the typical range of chemical shift what we would get in a protein ok. So, this so, another thing one has to notice is that this peaks are spread out, they are not together. So, this spreading out of peaks in this region is very crucial for our understanding.

So, as I will show you the next few slides how this spreading out helps us to get some information from the protein. So, this is how HSQS spectrum of a protein or HMQC N 15 proton HSQS or N 15 proton HSQS is analyzed; that means, every peak here corresponds to 1 amino acid because each amino acid is NH, but the terminal the N terminus not the C, the N terminus residue will not given ep because of this NH 3. And there is one more amino acid which does not give any peak in the spectrum even if it is not in terminal even if it is in the centre or middle or anywhere and that is proline; because proline does not have NH it has only N and these are cyclic ring. So, NH is a absent in a proline. So, NH is absent I cannot see any correlation of N with H in that amino acid. So, that is why prolines are also absent in this spectrum.

So, we have to keep in mind that suppose we have 100 amino acid, you have to subtract 100 minus the number of prolines and minus 1; 1 because of this first one. So, that many number of peaks you should expect to get. So, let me write it down more clearly here suppose number of amino acid if you have 100 amino acids in your protein, the number of peaks in HSQS equal to 100 minus number of proline minus number of the first residue one. Now this is for backbone, this is backbone; what is a backbone? Back bone meaning the this chain here this line which you see here this is the backbone of the protein we are not looking at the side chain, side chain comes in this R.

So, if you recollect this is a side chain of amino acid, I we will see this as we go along in the protein structure part for those of you who are not aware of a how a protein structure looks like, but this R is called the side chain and this the backbone is C alpha C O NH these are the side backbones. Now in HSQS you are seeing the backbone amides. So, what we saw, but you will also see side chain amide. So, let me go back a few slides here you see there is an side chain amide here also, this is in case of glutamine. Similarly if you have asparagine there also you have CO NH 2. So, such kind of amino acid also this is an amide group. So, it will also appear in the HSQS spectrum. So, the number of peaks are not just simply given by this formula here which I wrote, it is also an additional peaks will come which will be additional peaks are from the side chain. So, let me add this side chain peaks.

So, which are the amino acid which can give side chain? We have asparagine, glutamine we can also have arginine histidine and so on. So, we will not go into details of how to right now, we will come back to this later, but these are the extra peaks apart from this

numbers which you would expect to see. But typically what happens is we know how to identify these. So, that is a way to do that which I will come back to it in the later part. And once you know that you identify the side chain and if you ignore that in your calculation then what you are left with this only this part. Now of course, this is ideal situation you would expect that for a 100 amino acids, suppose I have 5 prolines. So, 100 minus 6 94 peaks, but that is not always the case its very very rare to see 100 percent of a peaks in HSQC.

So, typically for a very good system I would expect somewhere between 90 to 100, 90 to 95 there will be always a few percentage of peaks, which are not observed in at all and why is that so? That is because they exchange rapidly with the solvent. So, remember this is an labile proton, as I said although we say its involved in hydrogen bond etcetera, but many of the hydrogens which are in the loop regions might be exposed solvent and therefore, they might get exchange with the solvent very rapidly and because of the exchange they do not appear in the spectrum because of the fast exchange with solvent.

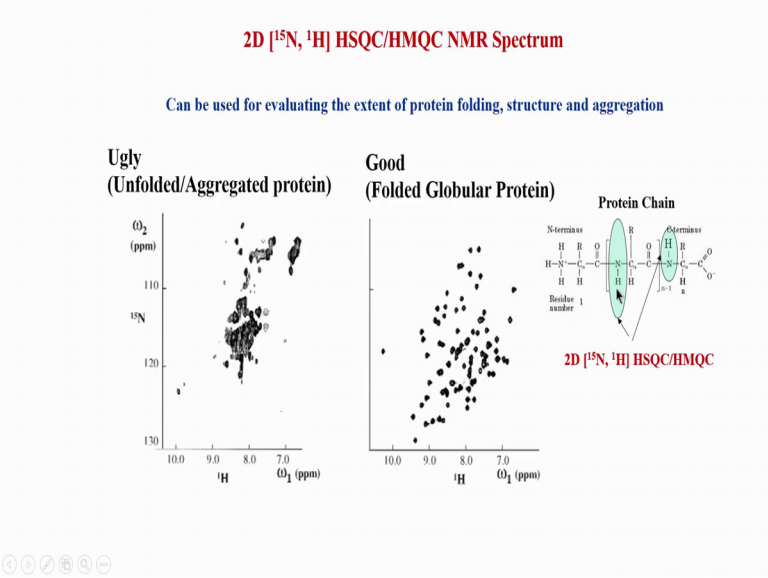
So, in such cases instead of like ideally expected 94 example for example, let us say I have 5 prolines, I expect 94 peaks I may not get 94 I may get somewhere around 90 or may be even 185. So, the missing peaks are really the challenging part in an HSQS or in a protein. In proteins generally we miss quite a number of peaks compared to what we expect from based on these kind of calculations.

But then we have to deal with them using some methods, we will see that if possible if time permits for this is the challenging part in a protein NMR experiment compared to organic molecules in a small organic molecules if you expect 7 or 8 peaks based on a structure, we normally always see those 7 or 8 peaks. Their intensity may vary we will definitely not loose them, but in protein the intensity can vary a lot. So, there for this peak is 100 in intensity the missing peak will be like less than 1.

So, it will be like a large difference. So, therefore, it is very easy to miss them or not observe them in the spectrum, even if you record the spectrum with a very high signal to noise. See even though I would say the signal to noise is very high, even then they can have many peaks which are missing. So, that is the problem with challenge with protein NMR and that affects structure calculation that affects the assignment of the protein and so on, but whatever we have with that itself we have to sometimes manage.

And we will see how to improve the spectrum quality in case we want to get more and more peaks. So, this is how HSQS N 15 becomes a very useful spectrum to start for any new protein. If you have any new protein you have to first label the protein with N 15 how we label them will see that in a second part of this course, but once you label a protein with N 15 you can record this spectrum and let me show you now how the information what kind information we can get from this.

(Refer Slide Time: 21:29)



So, as I said it can be used for estimating or evaluating the extent of protein folding and structure or aggregation.

So, this is a again same figure which I showed in the last class. So, now, look at these 2 spectra here. So, you can see this is for some protein which got aggregated or it became multimeric or it basically started degrading and you can see here I can hardly see then peaks separation here ok. So, these are the side change peaks which I showed you there are ways to figure out as we see later, but you can see here I can hardly count the number of peaks in this spectrum. So, this was a another protein where it gives a beautiful spectrum here.

So, you can see here it is a folded globular protein whereas this protein is either unfolded we do not know or it could be aggregated most likely this is an aggregated system. So, you can see here each of the spots here peaks are nicely separated and each one as I said in the previous slide it corresponds to one amino acid. So, the goal of protein NMR peak

community is to try to get this kind of a spectrum, this is if you count the number of peaks roughly it somewhere between a 90 to 100. So, this is a small protein this is not a very large protein.

But even then if you get a spectrum like this it becomes very easy to work such systems and one can get the structure of this protein relatively easier. In fact, this type of system here on the left is very difficult to work at. So, normally this type of proteins we normally we do not work with such system what is done is to try to improve the sample condition so, that we can reach this level ok. So, what are the things one can try one can vary the pH maybe this at a given pH it has got aggregated. So, one can try a slightly lower or higher pH, normally lower pH is preferred in NMR for bimolecular for proteins because of the exchange.

The exchange of this NH with the solvent depends on the pH and this is mentioned in varieties of books we will not go into detail, typically at lower pH the exchanges lower or lower. So, therefore, if you go to lower pH the peaks may become better less exchange. So, more sensitive, but at the same time very low pH can also result in precipitation of the protein and so on, one has to optimize the ph salt is another condition maybe this is aggregating because of low salt.

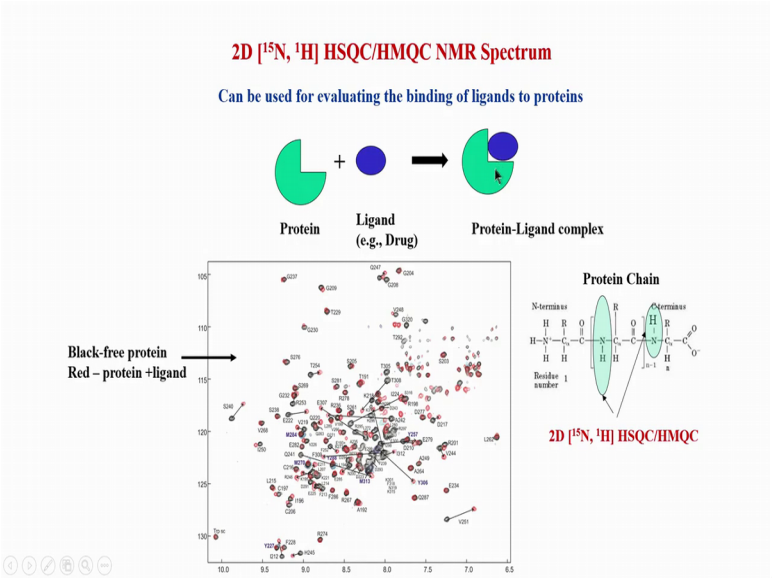
So, if you add little bit of salt it can prevent the aggregation and it can result in this condition. Sometime the temperature is a parameter which has to be optimized because temperature higher the temperature protein may start aggregating and therefore, we may have to go lower in temperature so on. So, this has to be optimized so, that we get as clean spectrum like this as possible, and if it is not possible to get to this level what typically is done is that you go back to the laboratory and reclone, means you remove some portion of the protein which may be causing this problem, you can chop the protein into smaller pieces and so on so forth or mutate some amino acids in the sequence to make it more less hydrophobic or more hydrophobic depending on what you are looking at.

And therefore, that kind of change in the sequence or protein that has to be carried out to get a good looking spectrum like this. So, that therefore, this kind of process involves time and effort and that is really a challenge again in proteins which are not amenable to NMR, because this kind of spectrum is very difficult to work.

Because I can hardly see separate peaks and therefore, one has to resort to some more technique. It can be also possible that this is not an aggregated protein this could be a monomer, but it could be high molecular weight let us say we are working with very high molecular weight like 50 kilodaltons or 70 kilodaltons. There we have to then use more sophisticated NMR experiments, a simple HSQS which is shown here is may not be sufficient and there we may have to record what is called at trosy 2D trosy experiment which will see down the line in the second half.

There are more advanced experiments which will probably help in improving the resolution when we go from here to here. So, these are the different challenges which 1 has to face, but you can see that that experiment is already helping us to give lot of information with it kind guides us that look this is not a really great system, this is a great system, can we do something to go from this particular state to this particular state by changing some conditions etcetera. So, HSQS is a there for N 15 proton HSQS is an important experiment, in any new protein before we start looking at the structure of the protein.

(Refer Slide Time: 26:21)



Second use of this experiment is in ligand binding. So, this is again we will cover in the second half of the course, but this is right now I am just give an example trying to make see the applications of the experiments. So, you can see for example, let us see I have a protein and have some drug molecule or a ligand to which the protein binds now this is a

binding. So, how do I detect that there is a binding? So, by non NMR techniques there are varieties of method we can use itc isothermal titration calorimetry or SPR and so on from NMR how can we you detect?

So, you can see this is an example of some protein and here the at color peaks, peaks which are colored in black are coming from the free protein, free meaning this protein here which is not having the ligand. But the moment the ligand was added to this protein in the NMR tube itself. So, we are directly adding the ligand in a specified concentration to the NMR in the NMR tube to the protein, we start seeing changes you can see now the red colour is a new set of peaks. So, the red color peaks are coming from a mixture or the addition of protein and ligand means it is coming from this here you can see this diagram. So, it is coming from the protein ligand complex.

So, the red color the brown black colors are the free protein. We can see there are so, much so, many peaks which are shifted, you can see this peak has moved here from red to blue black to red from here black to red and so on. So, each number here which is shown here indicates the amino acid number. So, this is G 204 means it is a glycine number 204 that sequence number the sequence is a long sequence and the 204th amino acid in that protein in the sequence was a glycine. Similarly here if you look at here serine 276 means the amino acid was number 276 in the chain and the serine was that amino acid name or type in that particular position.

So, in this case you can see it has not moved very much, it is in the same position. But look at this here serine 240 has moved quite a bit. Similarly here this valine 251 means, valine at position 251 has more substantial. So, by seeing at this moment of peaks we can actually find out what are the residues which are located at the interface this line here, the interface of the protein. Because those are the one which might be affected because of this binding we do not expect this binding to effect residues which are far away here. The residue which are at the interface are the one which are expected to be most affected and that are the one those are the one which are moved substantially.

So, therefore, this HSQS experiment by simply titrating or adding the ligand to the protein, we are able to find out which are the amino acids in the protein which I have got shifted, which are binding, which are located at the binding interface; this is a binding interface with the ligand. So, therefore, this is a very popular experiment we use the

word chemical shift perturbation, which is used very properly in drug industry in drug discovery and drug design, where we try to find out which are suppose I had a drug molecule and which are the part of the protein which it interacts with. So, of course, 1 thing if you notice here I should know every peak here the amino acid identity.

And that is the part of the whole process called sequence specific assignment, which you will go into the second when we do in the second half. So, you have to identify every amino acid here, identify the label with every amino acid. If you look back here I do not know the identity I just know there are some let us say 100 peaks in this spectrum, but I do not know each peak corresponds to which amino acid in the protein chain. That information is not available to us and that comes only from assignment and assignment itself is a very challenging task it takes sometimes months to assign a protein and figure out which amino acid corresponds to which peak here.

But once you have the assignment then you have lot of things which can be done such as this kind of experiments where you can titrate with any ligand and see what are the changes happening. So, these are the applications briefly I have gone through for HSQS and HMQC. In the next part lecture we will move on to a last experiment in heteronuclear 2D NMR spectroscopy which is known as HSQS TOCSY and that experiment actually it I will show us how to proceed from there to 3D NMR. So, we will see that in the next class how a 2D HSQS, TOCSY is made.