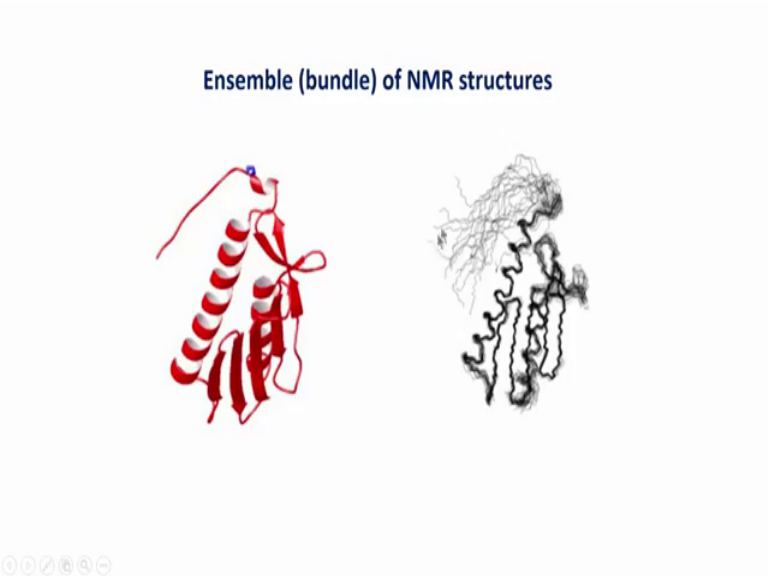
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

Prof. Hanudatta S Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 36
Understanding Protein ligand interaction by NMR: Chemical Shift perturbation

So, in the last class we looked at how protein structures are determined by NMR. And we also saw that we do not show one particular structure as a representative structure in NMR. We show what is known as the bundle or ensemble of NMR structure which is shown here.

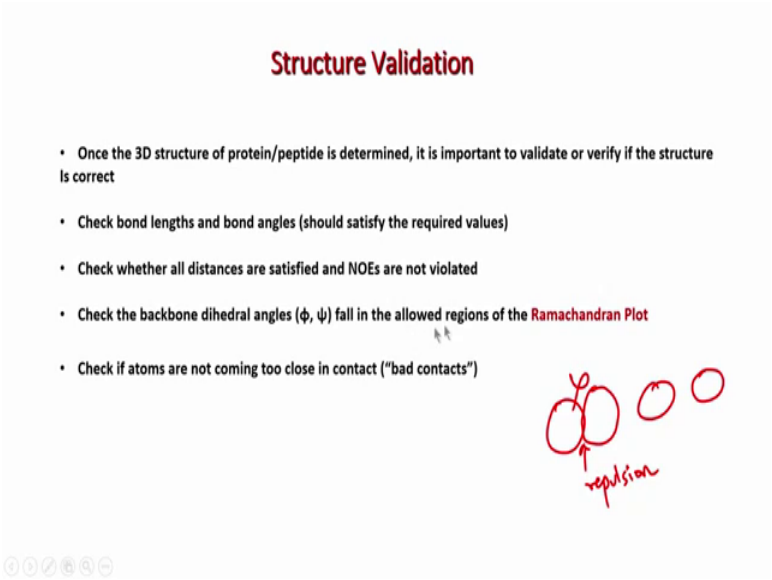
(Refer Slide Time: 00:42)



Ah Because all the distance is an angle constraints that we gave, would have been satisfied by all the 20, 25 structure shown here. So, therefore, all of them equally correct. And actually if you look at it they even have all of them superimposed on each other pretty well. Except this part of this protein, here looks very unsuper imposable, but that is because in this region there was not many constraints given to us. So, the software program could not determine a unique structure for this part. So, these are typically the do loop regions. And the end of the protein typically they behave like this in NMR structures because they are not well defined meaning there is no particular distance value we are able to find between atoms in this part because they are very flexible.

So, the NOESY correlations in a NOESY spectrum are not very strong they are weak. So, when the correlations are not present so, there is no distance information between neighbouring atoms or far away atoms. So, therefore, this part of the protein is not having a well defined confirmation which is not bad which is considered is fine because that shows it is a flexible part of the protein. Similarly if you look at this helical part here it seems to be well defined and all of them are almost super imposable because the distance values are satisfied equally in all the 20, 25 structure which are shown here. So, typically this is how NMR structures are published and shown and we normally take the average structure or we can take the first out of the 25. Remember the 25 structures here are given some ranking. Ranking from 1 to 25 and the first best structure can be shown as the representative model ok.

(Refer Slide Time: 02:32)



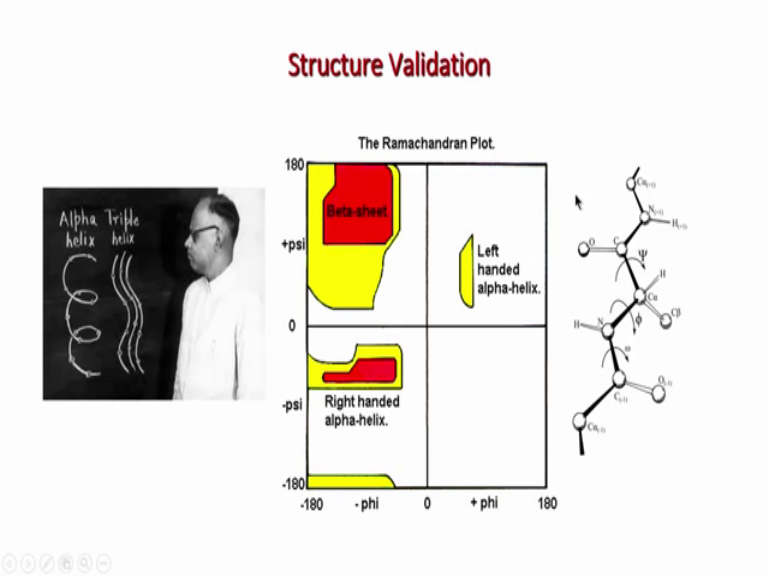
So, now the next part in NMR structure determination is to verify. So, remember we have solved the structure but how do we trust that the structure is correct or not? How do we believe? How do we validate? So, therefore, this is a very important step in protein structure validation where we have to see whether the structure is correct or not. So, typically this is done by many software, normally it is done by the database where we deposit the structures. So, database already checks thus lot of pre checking before it accepts your structure. If the structure is not satisfactory or violating some basic rules then it is not allowed to deposit.

So, what are the checks we do? Typically we check the bond lengths bond angles because CH or a CC bond angle a bond length, should not vary much even if your alpha, helixa or beta sheet. CC their angles and bond lengths are normally fixed by whether it is any structure you take. So, one should check whether you are structure which you have got at least is having the correct values for these angles and bond lengths. The next check is done is whether any NOE which you have given was violated or not. Violated means suppose you gave a distance value of 3.1 for two atoms, but the in the real struc[ture]-final structure, it was found to be 4.1 which is a huge violation. So, that is not acceptable.

So, that kind of distances checks are carried out, and then you reject those structures where the violation or deviation between the given NOE distance and the observed distance is very high. So, there is some sort of a cut off, which is used and without be anything beyond that cut off is rejected. Then you check the backbone angle phi and psi which falls in the allowed regions. So, this is a very important plot which is named after Indian scientist, who worked in Indian Institute of Science. And he discovered this in university he was in University of Madras and his very important plot called Ramachandran Plot that is something I will discuss next.

And the last thing is you have to check if the atoms are not coming to close to each other o, this is called bad contacts. So, remember two atoms let me show in a diagrammatic way two atoms cannot bump into each other this is not allowed. They have to have sufficient distance. Because this will cause a repulsion and therefore, it is not in nature you cannot expect two atoms to be closer than certain distance. So, that is called a bad contact and if this happens, then be a structure has to be rejected. It has to be. So, again for all this there is a cut off, means it is not that it is yes or no there is a cut off so, threshold. So, if it is beyond the threshold then it is rejected, but if is within a threshold it is acceptable.

(Refer Slide Time: 05:31)



So, let us look at now at this important plot called Ramachandran Plot. This is basically what was invented, the idea was invented by Professor, G N Ramachandran and idea is follows. That look at this, here on the right hand side is drawn the backbone structure of a amino acid of a protein chain. So, this is something we have already seen and you see there are what is called the torsion angle. So, this is a torsion angle phi between the H N, H alpha and this is a torsion angle psi. So, these phi and psi are very important indicators of secondary structure.

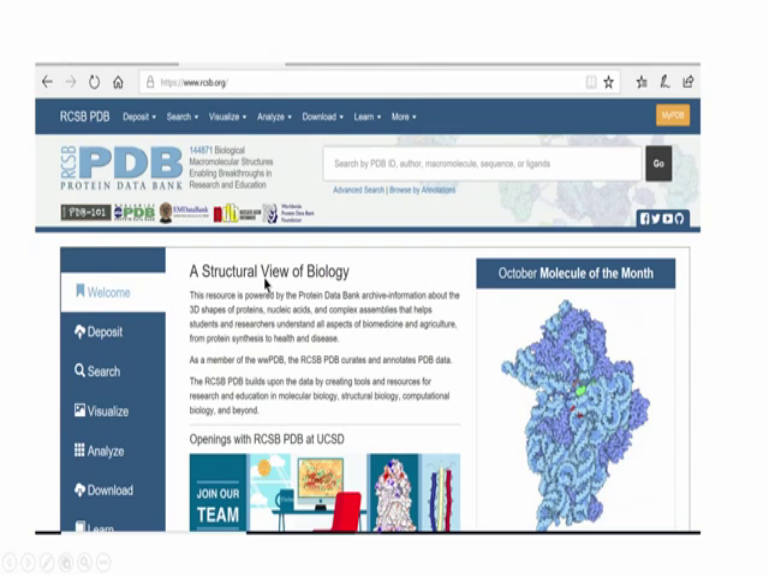
So, it turns out, he did lot of calculation on different structures, and he found out that if you are this chain. if is in alpha helical geometry confirmation. Then the phi and psi value should fall in this region only. So, for example, let us say this is some amino acid lysine. And if the lysine is an alpha helix then, that phi and psi of that lysine should be in this zone. Yellow and red colour zone, ok. And if it is in a beta sheet conformation then it has to be in this only. And if it is a left handed alpha helix, which is normally rare, but a still possible then it will be here. So, you see that this phi and psi are very highly restricted. They cannot take any arbitrary values they have to fall in this zones of secondary structures only. But the only amino acid which is exempted from this criteria or it can have any value is glycine.

So, glycine can come anywhere in this picture, but other than glycine all other amino acid should form. When, why is glycine exempted? Because glycine is a flexible amino

acid it has only two protons. Remember here there is no C beta. So, this restrictions what we are seeing here is actually coming because of the steric factor. Steric meaning there is the restriction because the C beta and N and all these atoms cannot come very close to each other. So, because of this geometrical restrictions there is a restriction on the phi and psi and that is reflected in this picture.

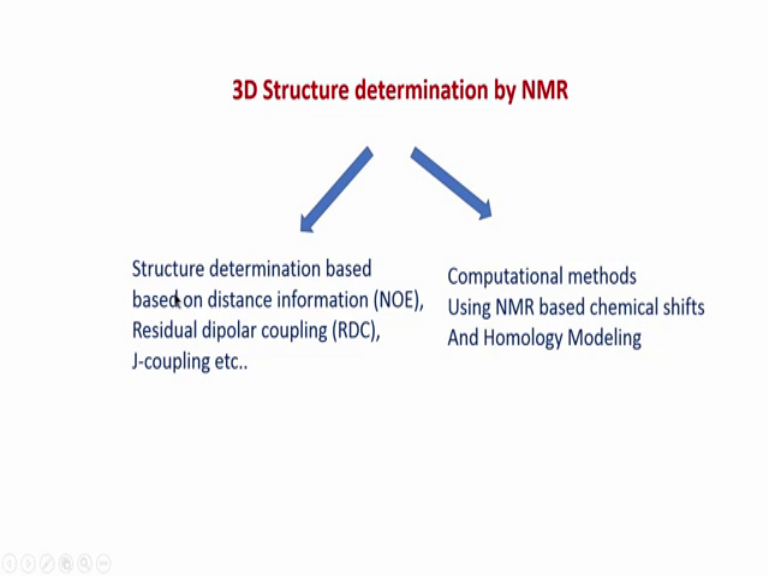
So, that it was very important plot and now in every paper, every protein structure which we publish, one has to show this graph and tell that look for my protein which I have solved all the phi psi value are coming in the right place. If I claim that my protein is, beta sheet protein I will show that everything is here. If I claim it is a mixture of alpha and beta, then alpha part should come in this and all the amino acids for the beta part should come in this side. So, this is called structure validation. You are validating with this plot whether your structure what you claim is correct or not. For example, if you claim it is a beta sheet, but then if the phi and psi of your beta sheet is coming somewhere outside this, then that is getting violated. So, that is what are the checks which are carried out. So, these are all carried out by a software which are available in the database.

(Refer Slide Time: 08:28)



So, this is the database which I mentioned earlier. In this database all the structures are deposited. And here itself the checks are carried out before you deposit and before it is released for the public.

(Refer Slide Time: 08:41)

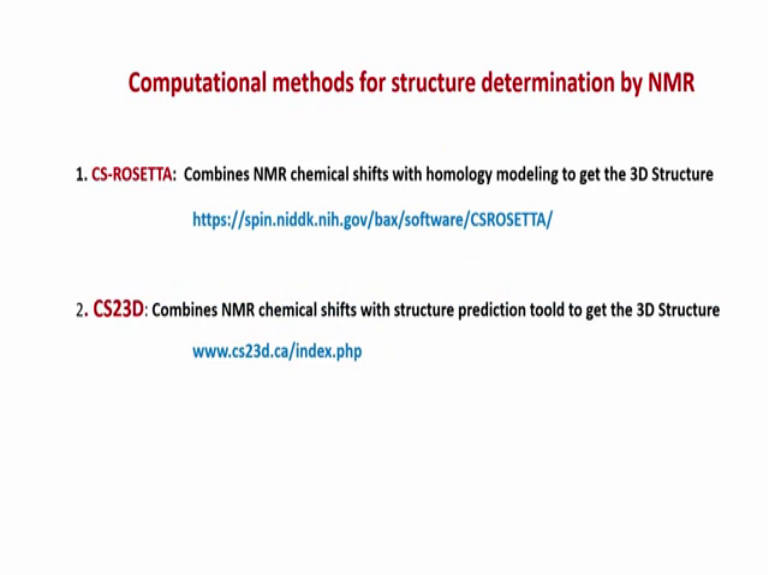


So, this is basically brings us to the end of the secondary structure part. I mean sorry the structure determination by NOE or by the distance this information wise. But there is another method which I told you in the beginning is that we do not have to go through this laborious, procedure of NOE. Remember NOE ah, distance determination structure determination is quite laborious. It takes lot of time, because the 3D NOESY experiment is not something very easy to interpret. Number one, number two as I said its an iterative procedure. You have to give distance information, that to a structure find out what are the violations, why the distances are violated, what could be going wrong and then go back start again and look at different set of NOE's. So, remove some bad NOE's and so on so forth.

So, that takes quite a long time, if you want a very good structure. But if you look at computational methods these are newly emerging methods which have come up in last ten years or so. Here what you do is you first you assign the proteins. Assignment is something which I said, is mandatory, you cannot avoid that step. So, once you assign a protein then you have the protein chemical shifts with you, take that chemical shifts and combine with homology modelling. Meaning the sequence which you are trying, the structure which are trying to solve a similar structure may already be available in the database. Which database? PDB the database.

So, if the similar structure, similar sequence. A structure of a similar sequence protein is available. Then you do not have to really solve the structure again. You can use homology modelling. But the information you are going to give extra is the NMR chemical shift of your protein, which you are trying to solve. So, this is a important point here, that one should look at. The homology based modelling if it the structure is already known for a similar protein.

(Refer Slide Time: 10:40)



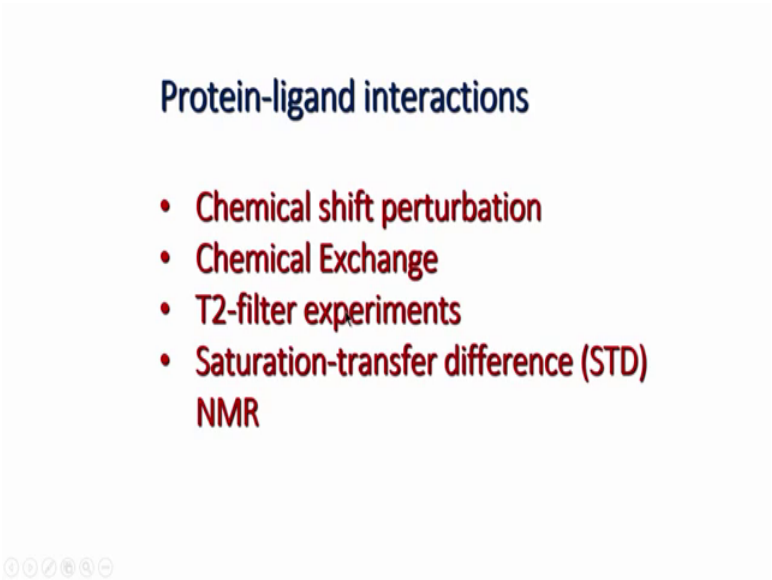
So, there are different software we will not go in this course, a particular course on all the details of this because this is quite involved in terms of the algorithms, methodology that is used. So, we will just I want to name two softwares which are popularly used one is called CS-ROSETTA, which combines NMR chemical shifts with homology modelling to get the 3D structure. So, you can look at this website. Where you can download and use it for your protein or upload your data and they will solve the structure. Similarly CS2 3D, CS2 means chemical shift to 3D structure combines NMR chemical shift structure prediction tools ok. So, there are many structure prediction tools in, in bioinformatics which are known on well established today those can be used and combined with NMR.

So, essentially what we are doing here and also here is we are using computational methods in this case it is homology modelling. But in this case it is abinitio or it could be a structure prediction method. But what we are doing is once we get a structure or when

we are predicting the structure we are constraining it means we are trying to constrain such that the chemical shift what we have observed should be satisfied ok. So, each time, when a structure is predicted or homology model structure is generated a chemical shift is predicted from the structure and when the chemical shift prediction and the observed chemical shift should match.

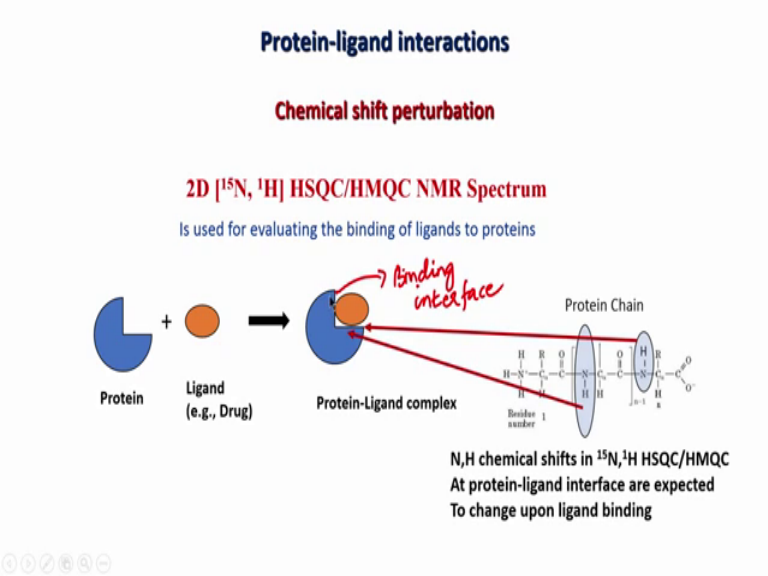
So, that is a kind of a constraint, which is given to the software and they generate the structure. So, this is another alternative to NOE, but again the resolution which we get from these may not match the one which we get from NOE. So, NOE is still the more preferred way, but that is a slow approach this is must faster approach ok.

(Refer Slide Time: 12:32)



So, we come to the next part of this course, where we look at now having assigned a protein, having known the structure of the protein can we do now study? Can we do something to study? Protein to ligand interaction. So, this is a very important topic because most of the drug discovery depends on this interaction. So, we are trying to suppose discover a new drug molecule, we need to know if it binds the target protein or not. So, these are the different techniques, which you will try to see in this course and so know what is the ways, different methods which we can use to characterize. So, our goal here is to characterize or figure out whether a particular ligand is binding to a protein or not ok. So, let us start from the first most basic experiment or most basic method called Chemical shift perturbation.

(Refer Slide Time: 13:34)



So, we have already seen this Chemical shift perturbation in the previous part of this course, where I showed you for HSQC; how HSQC is useful for perturbation for ligand interaction. So, I am repeating a slight repetition of that, so typically chemical shift perturbation means, we are trying to find out what chemical shifts are changing in an HSQC spectrum. Again specifically we use N 15 proton HSQC. So, N 15 proton HSQC or HMQC can be used and then we titrate, titrate means we add the ligand and then try to monitor the changes happening to the protein signals.

So, what this HSQC is now for the protein which means I should know, I should label the protein with N 15. Number 1, second I should know the chemical shift of every peak in that protein. So, the again this a makes bring are backs to the importance of assignment. So, suppose let us say this is a protein, and it is interacting with a drug molecule or any ligand it could be another protein a peptide, any other partner binding partner. So, let us say they interact with each other, and it forms a complex. So, what we do is, we study this complex formation by HSQC that is N 15 HSQC or HMQC.

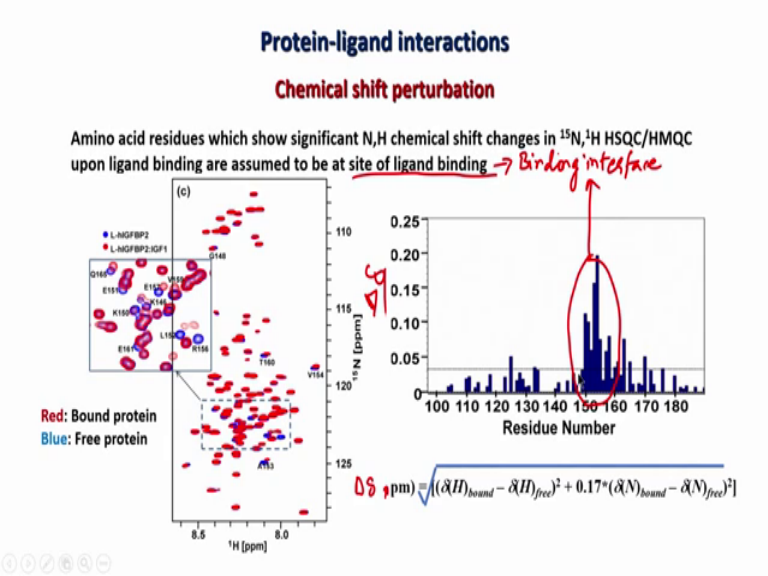
So, in HSQC or HMQC, if you recall by now you should be now familiar with this that we observe the N H correlations in the spectrum ok. So, this N H correlations are seen. Now if this for a given in the protein chain. Let us say this is a big protein chain and the amino acids which are on this interface means sitting on the binding interface with the

ligand. They are expected to change chemical, shift of those amino acids. These amino acids here and here, are expected to change when the ligand binds to the protein.

Why do they expect, Why do we expect it to change? The reason is that there is a small change in the structure will happened at the interface. Not only that the chemical environment before adding the ligand of these amino acids here, will be different from the chemical environment when the ligand is bound to the protein. So, therefore, there is a change in the chemical environment or change in a structure can lead to changes in this backbone n h chemical shift.

Of course, all chemical shift will change, but we are only looking in HSQC spectrum. Because it is a very convenient, very easy to do the experiment and in that we see each N H correlation they are expected to shift, shift meaning change. We will show this example soon; we will see some real case studies. But those shifts are what is indicating, whether the amino acid, if the ligand is binding or not ok. So, let us look at this more in more detail now.

(Refer Slide Time: 16:44)



So, there are different types of ligand binding. So, this is basically amino acid residues, so, let me go back to this here. So, as I mentioned with this arrow marks here they are pointing to the interface, so, let me write this word very clearly this is called a binding interface. So its a Binding interface and the amino acids which are located in this interface now will experience at difference in a chemical shift before and after, before

and after the ligand interaction. So, this change in this chemical shift is what we are looking for ok. Now what about the amino acids here inside the protein? There we do not expect much change because it will not be affected by a ligand which is sitting far away.

The only the one which are close to the interface, Binding interface these are the one which we expect to really undergo a big change in or small change in the chemical shift ok. That change will depend on what kind of ligand is this ok. It will depend on how much structural changes happening and so on. It is not that every time you will get the same type of changes for every protein. It depends on protein to protein it depends on ligand also ok. So, the amino acids which show a significant changes in your spectrum, will be the one which are assumed to be at the site of ligand binding.

What is this site of ligand binding? A site of ligand binding means, binding interface. So, binding in the interface, the residues which are located and the interface of the binding will be the one which will show significant changes ok. So, let us take an example this is a real case example of a protein. So, we can see these are all like this is HSQC spectrum N 15 proton. And this red colour is a bound protein, and the blue colour is coming from the free protein. So, free means the normal protein. And after adding the ligand the blue the peaks have moved you can see here if you zoom this doted box here you can see in this here, there are changes shifts happening. Shift means they are all moving.

So, the blue has moved here this blue has moved there and so on. So, these are the changes happening in the HSQC spectrum of this protein, when the ligand was added. Now you can see here this numbers E 151 and so on. These are the residue numbers. They are the ones which were undergo change. That means, I should know which peak here corresponds to which residue and that I would expected to be knowing at this stage because I would have assigned the protein, so I will be knowing all the assignments of each peak. So, now, this changes what you are saying here can be plotted in a graph, graphical manner.

So, what you see here in this graph on the right side is that there is a residue number on the x axis. And this vertical blue colour bars are telling me how much chemical shift change is happening so, I can write is as delta. Delta delta means how much change in the chemical shift is taking place so, this change is what you see here, but if you notice the change is here is happening in both N 15 direction and the amide horizontal

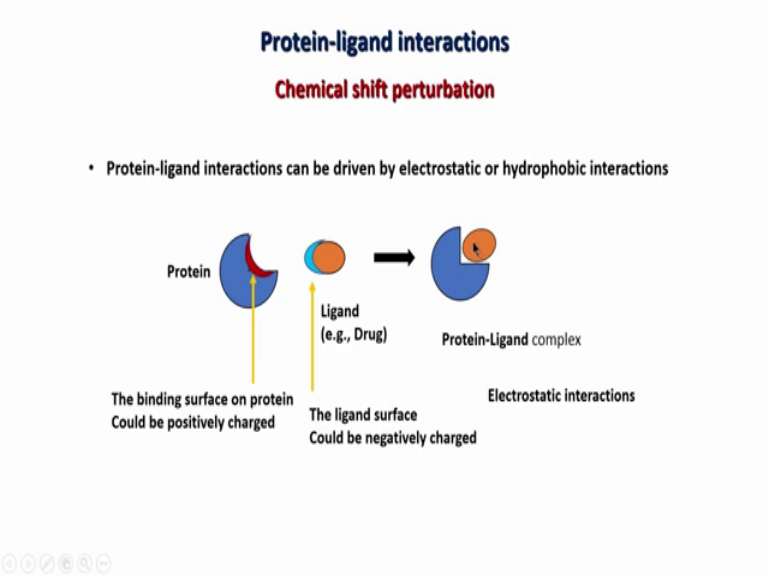
direction. So that means, a change is not just in the amide region. I mean amide chemical shift it is also change in the N 15 chemical shift.

So, how do we combine this change in the two dimensions? We combined with some formula. So, this is shown here so for example, this change delta delta. So, this is got cut here delta delta in ppm scale, which can be calculated like this. So, you have a chemical shift of proton in the bond form, minus chemical shift of the proton in the free form. For a given amino acid. Take for example, R 156 you calculate this difference for that amino acid between the red and blue colour. The proton dimension proton axis means the horizontal this horizontal axis.

Then you add the shift in the nitrogen axis. Nitrogen axis is N 15 the small shift here you see if there is a vertical shift also for between these R 56 red and blue and then that difference in N 15 dimension is also taken, but it is scaled down because of the gyromagnetic ratio of N is small. So, the scaling down is done and then added to this, and then use a square root you calculate the final delta delta and that is what is plotted here in this graph so, you can see for the residues between 150 to 160. This part that is, this portion seems to be changing a lot. Ok. It is beyond the threshold; threshold meaning the standard deviation.

So, therefore, I can say that this is my Binding interface. So, I can say this is my Binding interface because the Binding interface is what is expected to show significant shift when you add the protein sorry the ligand to the protein. So, this is basically what is this is how we do normally all the chemical shift perturbation analysis. You add the ligand to the protein. First you record the spectrum the HSQC spectrum, of the protein alone, and 15 labelled and once you have assigned the protein. Once you add the ligand to that sample, and then the sample the ligand will now cause a change in the chemical shift if it is binding, and wherever the changes are taking place. That can be plotted graphically and wherever there is a cluster, means a group of changes that group is involved in the ligand interface binding interface ok. So, but now let us see further, there are some little complications which can happen.

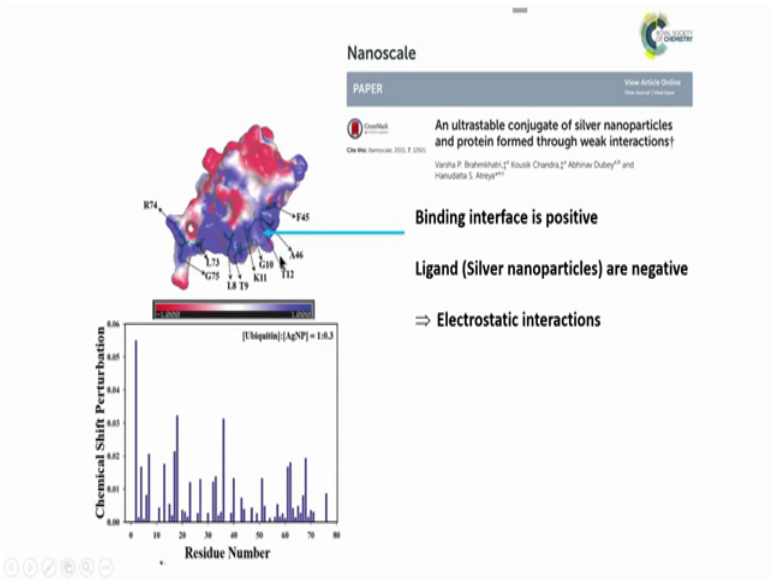
(Refer Slide Time: 22:54)



First of all one should know what are the different types of interaction? So, a protein to ligand or a drug molecule there are possible interactions are it could be electrostatic in nature or driven by electrostatic interaction or this ligand interaction could be driven or dominated by hydrophobic interactions, these are the two broad possibilities. I can so, how do we know the electrostatic or not? We can find out like this, suppose the binding interface on the protein could be a positively charged residues or surface could be positively charged. And this side of the ligand could be negatively charged.

So, when these two interact with each other the positive surface of the protein interacts with a negatively charged surface of the ligand and that means, it is an electrostatic interaction. So, that can be found out by based on the charge of the surface, where the protein the ligand is binding and based on the charge of the ligand. So, that is possible and that can be also the opposite. I can have this as negatively charged or this and this could be positively charged so it could be opposite, but still the interaction is driven means governed by the electrostatic interaction.

(Refer Slide Time: 24:12)

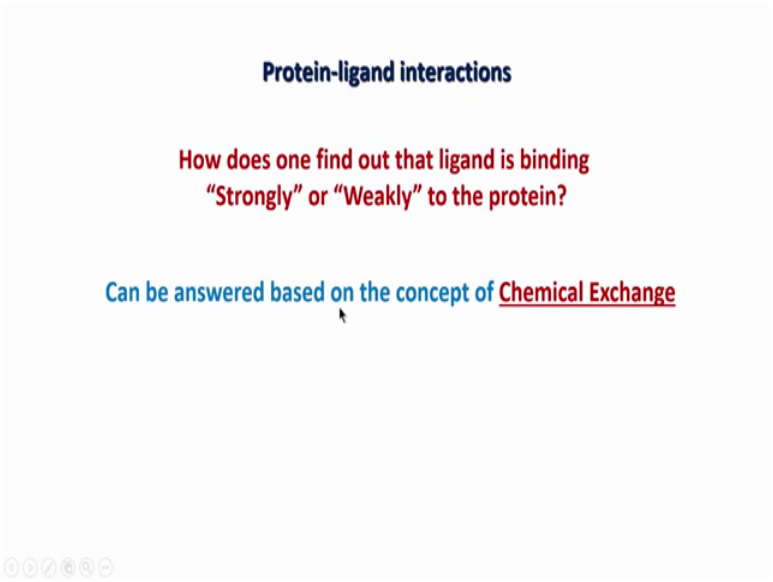


So, this is I give I will give a real example, from the publication we can go to this paper to get details but what was seen here is this is a protein called ubiquitin. So, when you add a silver nanoparticles. Silver nanoparticle is a negatively charged system. It goes and interacts with the positively charged surface. You see the blue colour here is a positively charged surface of the protein and these amino acids which are indicated are the ones which show very high chemical shift perturbation. Delta delta that is what we saw in the previous slide.

So, those which you undergo very high delta delta, they were plotted on the surface of the protein and it was found that this all of them lye, means either located on the positive charge surface. Blue colour is positive, of the protein which means this nanoparticle is interacting because is negative it is interacting with the positive surface of the protein, which means it is this interaction between the nanoparticle and the protein is driven by electrostatic interactions ok.

So, this is very important to know this is how NMR can give information, on what is the nature of interaction between two ligand and a protein. Based on the Chemical shift perturbation plot, one can figure out from the structure where it corresponds to whether it is on the positive side or negative side or it could be hydrophobic amino acids, and in such cases it could be hydrophobic interactions. So, that kind of information comes out from such analysis like this ok.

(Refer Slide Time: 25:56)



Now, the second next question you may want to ask is ok I know there is an interaction. But is it a strong interaction or weak interaction? Mean how do I find out the ligand is binding strongly or weakly, so this is shown in a quotes, because strong and weak are always relative to some measurement. It is not a absolutely strong or absolutely weak relative to something, but how do I still figure out or how do I characterized whether a ligand is binding weekly or strongly to a protein and that can be answered based on the concept of chemical exchange.

Chemical exchange is very important concept in protein NMR. We will take it up in the next class, I will see how this helps us to find out whether a protein and ligand having or having a strong interaction or weak interaction. So, we will see that in the next class.