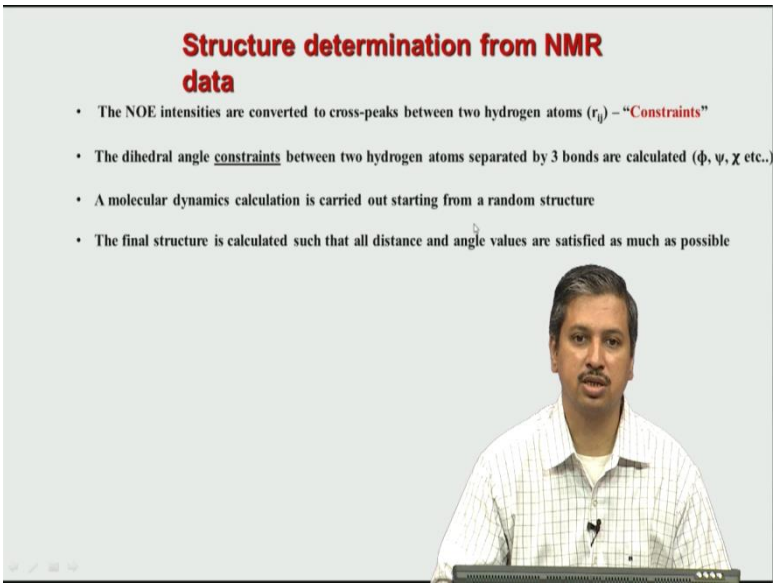
Principles and Applications of NMR Spectroscopy
Professor Hanudatta S. Atreya
NMR Research Centre
Indian Institute of Science Bangalore
Module 7
Lecture No 36

In the last class, we started looking at how peptide structures can be determined by NMR, we looked at how NOESY and ROESY peaks are integrated. And once the ROESY and NOESY peaks are integrated they are converted into the distance using the relation that the intensity is proportional to 1 over R to the power 6, but usually we do not do that practically. Practically the distances are never calculated exactly but they are estimated in a rough manner. So, typically what is done is that you consider a strong intensity peak as a short distance. Let us say between 2.5 to 3.5 a medium intensity you can considered as distance having 3.5 to 4.5 and a weak peak will be beyond 4.5 to 6, so this is what we will see now in today's class.

(Refer Slide Time: 1:13)



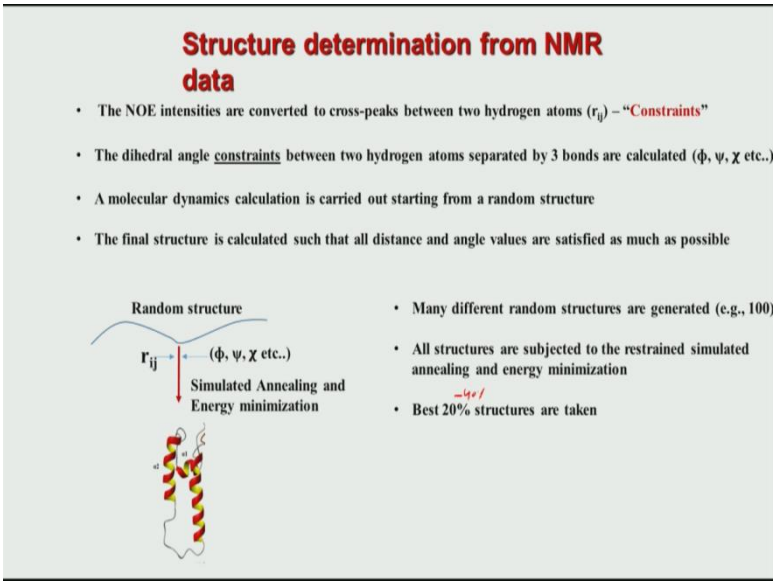
So, you can see the typical protocol is shown here. So, as we discussed first the NOESY intensities are converted to distance and we use the word constraints and why is that called a constraint? The constraint is called because we are now going to determine a 3 dimensional structure using distance as the information between the atoms. So, therefore the structure will be constraint to satisfy these distances. Another type of constraints or information which we give is the j3 bond j-coupling between hydrogen atoms. So typically, this is the backbone Phi angle, Psi angle and it can be for the side chain Chi angles and so on.

So, this is another set of values which we input to the computer program which does the structure determination, so this is what is being going shown here. So, we basically do what is called as molecular dynamics calculation carried out. So, what is done is you would take a random structure first, in case random structure is because initially we do not have any information of the structure of the protein because that is our goal, we want to determine the protein structure or peptide structure, so we do not have any information on how the structure looks like? So, we can start in a very random coil structure and then you start giving these constraints. And based on these constraints that is distance and angle the program will now try to let to a structure arrive at 3 dimensional structures which will satisfy these constraints.

So, the final structure which is calculated is such that all distance and angle value which we have given as input are satisfied as much as possible because it is never possible to completely satisfy all distances, the reason being many a times the intensities what we measure from ROESY may not be the correct intensity it may be distorted because of overlap of peaks on other residues, it could because of noise problem or it may be weak signal to noise and so on. So, therefore the distances which we get from the intensities are never the correct distance values and therefore, it can be it not possible that all distances are satisfied for a given structure.

So, what we do is we try to satisfy as much as possible and those which are not satisfied we use the ward violation. So we say that these distances are violated some of them so we look at those violations, why are those violating? What is wrong with those distances? Why did they get not satisfied and what are the angle problem values was values problem of angle values? Why did they get satisfy and so on. So, that detail analysis has to be looked again that means one has to go back to the spectrum and look at those intensities and recalculate distances.

(Refer Slide Time: 4:19)



So, this is a iterative procedure which is depicted in this picture here. So, as shown here we start from the random structure random coil structure and then we input this information of distances where r ij is the distance between atoms i and j 2 hydrogen atoms. Remember we are working with hydrogen distances hydrogen atom, so 2 hydrogen atoms i and j their distance between them and how do you get the distance? The cross peak in the ROESY or NOESY spectrum between these 2 atoms ij will that intensity gives you this distance.

Uhh similarly, we get the phi and psi values based on certain experiments we discussed like j HN H alpha gives you the angle phi and so on and we input all this and we carry out what is called as simulated aniline. Simulated aniline is basically a protocol for a molecular dynamic simulation, where you carry out first you straight generate a random structure at high temperature and then slowly cool it to a down and while cooling it is constraint to these structures that is the distances constraints and it finally gives you some structure which will satisfy or which is trying to satisfy all the distances and angles which you have given and the energy there is also energy minimization step so that the inner structure which you get in the end is energy minimized.

So, this is the typical protocol which is followed in any structure determination program which uses NMR data and what happens is because we are trying to satisfy a distance and a constraint. We do not know the structure as such so that is why we do not call this as a structure. In many literatures, many papers or many groups would tend to refer to this as a model because it is not a structure, we never really know the structure of the molecule, and we only know the model representative which looks like that structure. So, why it is called a

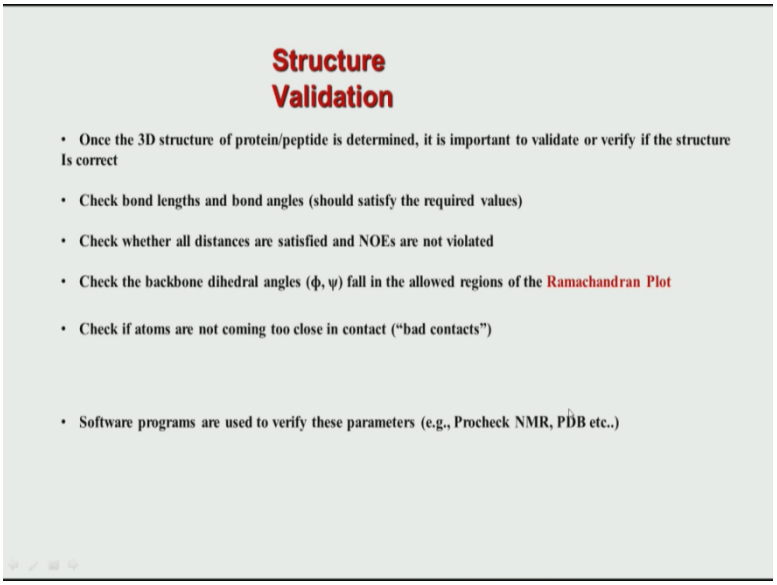
model because we are building a model based on the distance values and the angle values which we get from the NMR data NMR spectrum. So, based on those values we are building a model which can serve satisfies and may look like a structure. So, remember we never get the exact structure because NMR can only be it is a derivation but deriving the structure based on some distance information.

So, what is done typically we do not start from single structure, we generate 100s of random structures because remember random structures can be any number, it can any randomized structures, so we can typically generate about 100 randomized structures in each of them is subjected to this molecular dynamics analysis simulation aniline and slowly the distance values are given for each of these 100 structures and you a finally get 100 final structures, but all 100 may not be correct because many of the 100 or many out of the 100 there will be violations of distance values.

Violations means a distance values which we give may not be satisfied, so therefore we take best out of those 100 structures. So, this is what is shown here that all structures are basically subjected so we will call the word "restrained simulated aniline" and best 20 structures are taken as shown here. So, the best 20 structures typically mean 20% structure means 20 out of 40 out of 100, sorry this is actually it should be 40%, so let me correct this, it is 20 to 40 percent. So, that means if I generate 100 random structures I will take the best 40 structures taken and those best 40 structures out of 100 all of them can be considered to represent our structure. So, therefore when you look at NMR data which is presented in papers and on the websites we will see that they do not show you one structure they will show what is called a bundle, a bundle of 20 to 40 structures

And bundle means an assemble or collection of 20 to 40 structure because those collection they have been arrived by generating 100 random structures each one subjected to this conditions and then energy minimized and out of those close which are having the energy or least number of violations are considered as best structures and they are based on that those 20 to 40 structures collected and all of them are now equally good to represent the main structure. So, therefore it is NMR compared to other techniques like crystallography you do not generate one structure but you generate a set of structures some ensemble or bundle of structures all of which have been derived starting from random structures and all of which have been subjected to the same constraints of angle and distances and out of all the random structures the best ones are taken.

(Refer Slide Time: 9:38)



So, this is the typical protocol of NMR structure determination by NMR for peptides, proteins and even for nucleic acids and so on. Whenever, you use NMR data based or NOESY, this is usually what we do and but one thing is once you have the structures determined, the next question which comes to mind is how do we know it is correct okay so remember this is model building process. We have built a model of the protein structure based on the information we have given is angle and distances. So, the model can be wrong if the distance and angles are wrong, angle meaning torsional angle. So, how do you find out if the structures are correct what we have generated? So, this step is known as structure validation, so we validate or verify whether are the structures we have derived is correct or not.

So, number one thing to verify the structure is to find out whether it satisfies a general structural if in general whatever the structure the structure of peptides and proteins which exist in our literature in database, whether our new protein structure satisfies a general constraint of this proteins structure. What does it mean? This means that it should at least have the same bond lengths and bond angles within the limit is which are given in the literature for the standard bond angle and bond lengths, so that should be satisfied. So there is a required values are given in the literature or in database. So, all the bond angles and bond lengths in our peptides which we have stalled the structure should satisfy those values.

They cannot be arbitrarily different from these bond lengths and bond angles which are known in literature because in database typically these are derived from high resolution X-ray crystal structures for a large number of proteins, so it has been now found out that the bond lengths will typically for in a particular value for carbon-carbon single bond, carbon-nitrogen

double bond so on, the partial double bond and the peptide and so on. So these lengths should not vary beyond in certain limit. So within that limit if it is satisfied that is the first check which is done then the second check which we do to verify the structures is good is that all the distances which we have given we use our distance constraints.

So, all the distance constraints that we have given should be satisfied and there should be not too many violations. Violations meaning that it is not following the value which we have given. So, that means let us say between two atoms A and B I have given distance of 5.5 but in the real structure let us say it is coming to as 7.5. So, what basically it means that the limit should be 5.5 but it is not being satisfied so that case we called it as a violation and that is a very big violation for example, 5.5 to 7.5 is like a 2 Armstrong violations which is huge and that should not have happened that means something is wrong with some other distances or we just particular distances. So, one has to go back and look at his data and then recalculate distances or relook at the assignments.

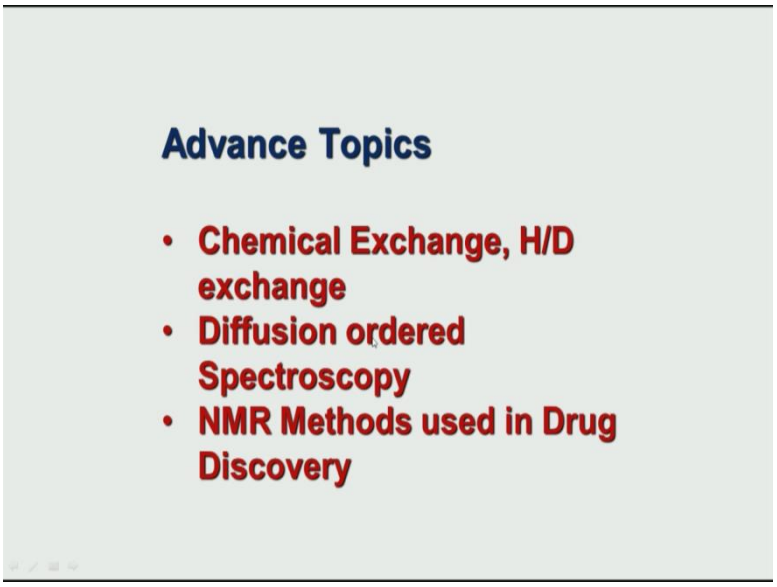
Many a times typically what happens is the assignments means given rij in a distance between 2 atoms i and j is calculated based on the cross peak between the 2 atoms i and j If the cross peak may not be correctly assign. So, in that case you result in wrong distance value and that can also create problems. So other thing which is checked is that the backbone dihedral angles which is phi and psi that is for the backbone peptide bonds surrounding the peptide bond they should fall in what is known as the allowed regions of the famous Ramchandran plot. So Ramchandran plot is a plot of phi verses psi, this is a very standard plot you would have seen in textbooks and in papers, is a standard plot of phi verses psi plot and it is said that for a given alpha helix or beta sheet one should have a particular value of phi and psi for the particular for all residues.

So, these regions where they should fall is known as the allowed regions, so if the phi and psi value for given amino acid in your peptide falls outside that region that is called disallowed region and why it is coming in the wrong region it is because there is something wrong with the structure and therefore, it is adopting a wrong conformation and therefore, that is not possible if the conformation was right and therefore, that region is known as disallowed region. And that also has to be checked that majority of the residues amino acids in a peptide should fall in the allowed region and then we also look at what is called bad contacts. Bad contacts are means two atoms are coming very close to each other within the vander waals distance and that cannot happen because that will lead to repulsion.

But that aspect since we have given around distance value that aspect does not come into play in the calculation, so therefore they may end up coming very close which should not be the case. So this all these checks have to be done to make sure that our structures are not too much away from the required constraint or required values for different distances angle and so on. So, all this these checks are done very automatically nowadays automated way by software programs and typically the names of software programs are given here, there is very famous program called procheck NMR.

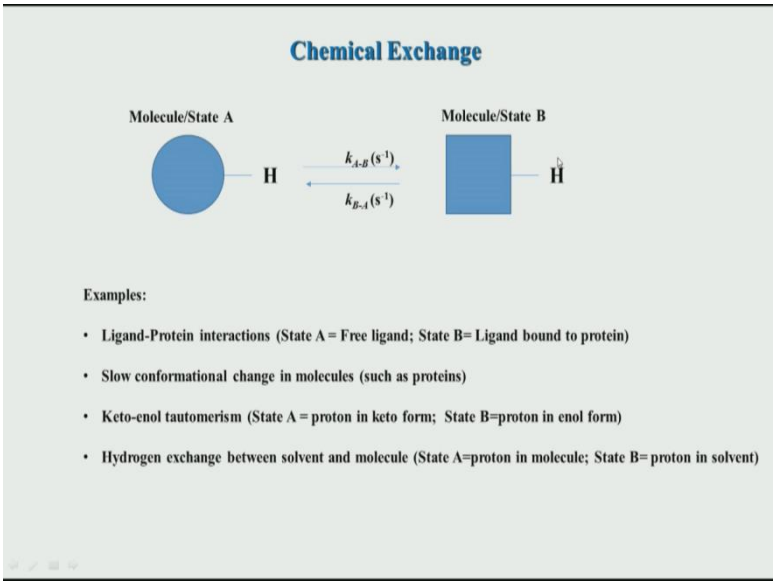
So, procheck is used in fact when we it deposit our structures in PDB in the data bank and PDB also has a inbuilt validation in a software. So that is also with checks before you finally deposit the structure to the data bank. So this is the structure validation we have looked at very briefly there are many more parameters as we checked, but these are the major ones.

(Refer Slide Time: 16:05)



So, this brings end to the this part of the course and we now move on to the next part the last part of the course where we looked at some advance topics in this course and this is where we will see these are the following topics. We will go one by one through this and these are basically either using 1D NMR or it can be using 2D NMR and we will see how these are basically applied in all of them in fact are used typically when we are looking at protein drug interactions in drug discovery, in protein structure analysis and (())(16:25) to protein binding and so on.

(Refer Slide Time: 16:37)



So, we will start from first one today so that is chemical exchange, so let us see what is a chemical exchange? Chemical exchange is nothing but a simple exchange of hydrogen atom between 2 different states of a system. So, if you look at this for example here, we label it as a molecule, so consider this whole thing as a molecule which we can call it as state A and this has some hydrogen atom schematically shown here. Now, this state or this molecule suppose undergoes a change to another molecule which may be the same molecule or different structure or different molecule. There the same hydrogen now is represented like this and this hydrogen now we can label it as state B and this is a dynamic exchange that means this whole state molecule this hydrogen atom goes from state A in this molecule to state B in this molecule and comes back to the original state and goes forward and back, so this is like an equilibrium steady state system.

So, we can call it as forward rate of exchange and a reverse rate of exchange. So, this exchange rate is measured in the unit is of second inverse, this is typically the hertz scale also which we use for frequency. So now what is the whole game of chemical exchange all about? The chemical exchange basically (())(18:08) to study in NMR, what is the rate of this exchange? How fast is this exchanging between these two or is it a very slow exchange? And so on.

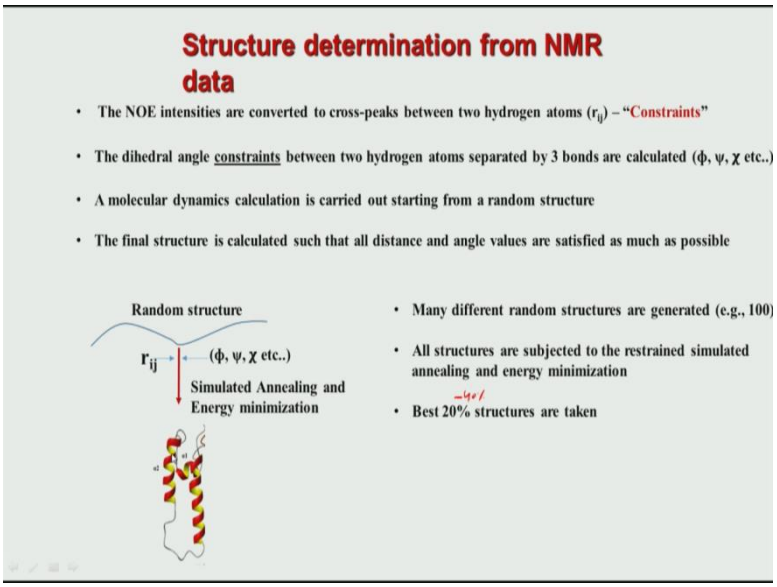
So, characterizing the rate of change of a hydrogen atom from one state to another state that process comes under the purview of chemical exchange and this has lot of applications which is shown here the few examples, for example, when you are studying ligand to protein interaction. So, suppose your ligand in state A this hydrogen atom is attached to ligand, now

this ligand goes and binds to a protein and this whole box here could be the protein + the ligand that is the bound form, in the bound form the same hydrogen now can be called as in a another state B.

So, ligand environment it is an environment around the free ligand, here it is still attached to the ligand but it is now the ligand is bound to the protein. So this hydrogen now experiences the different chemical environment and that is called state B. So, this is one of the first cases which we will come across ligand-protein interactions is maximum and wherever the chemical exchange is useful and typically when you look at drug to protein interaction or drug to receptor interaction or enzyme substrate interactions, all of this can be studied by this process of chemical exchange.

Second possibility, where it can be used is conformational change and for example, let us say there is a protein which undergoes a conformational change from one state to another state. So, how do we talk about that? So for example, let me go back to the previous slide here. We saw this particular protein. Now, let us say this helix which you are seeing here which is known as alpha 2.

(Refer Slide Time: 20:08)



Suppose this helix is moving little bit left and right, left and right like this. It is basically having a dynamic motion okay. So if that happens that can be called as chemical or conformational exchange okay. So hydrogen atoms sitting on this helix1 helix here and the hydrogen atom on the same helix, but when the helix moves out can be called as second state.

So, this is called conformational exchange and that is also very useful to study using this phenomenon of chemical exchange.

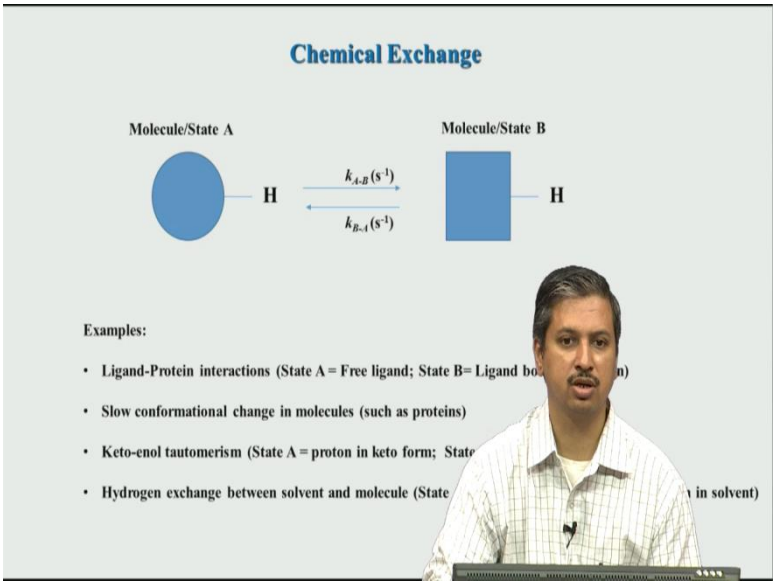
When in chemistry we often come across Keto-enol tautomerism where we can say that the Keto form the proton which in the keto form you can say state A one state and the same proton same hydrogen atom in the enol form can be called as state B. So, remember we are always looking at the same hydrogen atom between one position to another position. So, we are not looking at 2 different hydrogen atoms, so we are going to monitor the same hydrogen atom, how it behaves from one state to another state?

So, the same hydrogen atom which is keto form will also now become an enol form and the same hydrogen now will have a different chemical environment and therefore, there will be a exchange of the environments between the two based on this kind of a chemical equilibrium process and that now can be studied by NMR using the ideas of chemical exchange. And one of the other very important example hydrogen exchange, this is very commonly used in peptides proteins or even small organic molecules where we study how the hydrogen atom moves from the solute to the solvent.

For example, hydrogen atom let us consider this case let us suppose this blue color circle is a molecule, so the hydrogen is attached to the molecule but that hydrogen now goes and exchanges with the solvents. So suppose this square is a solvent here, so this hydrogen atom from the solute goes and attaches to the solvent so that is a different molecule now, it is sitting on the solvent, here it was setting on the solute. So this hydrogen atom therefore, experiences an exchange or a interchange between the molecule and the solvent.

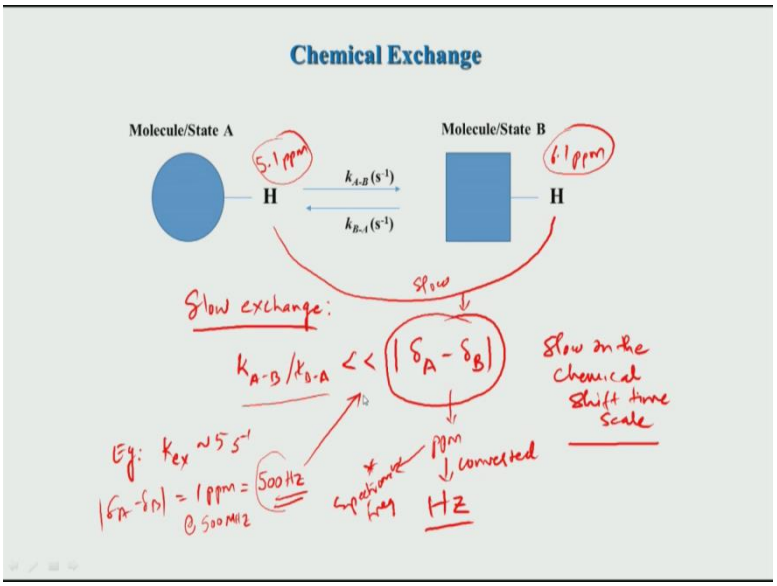
So, this process if it is fast or slow can be characterized by NMR and that gives you a lot of information about the environment of this hydrogen atom. Let us say suppose it is involved in hydrogen bonding with some other hydrogen in the molecule. If that is so, then then this hydrogen will not like to move to the solvent, so its exchange rate from here to here then will be slow because it is hydrogen bonded to some other hydrogen so it is not available for exchange. So, based on this exchange rates you can now figure out whether this hydrogen is hydrogen bonded or not or is it buried inside the molecule or not and that information becomes very useful to have if you want to look at structures. So, hydrogen exchange is a very important experiment which is carried out in NMR to find out the state of the hydrogen bonds.

(Refer Slide Time: 23:40)



So, when we look at chemical exchange between again between two states we talk about three different type of exchange. Number one it is called slow exchange, slow exchange means this hydrogen is going slow between these two, but remember, slow or fast is only relative. So what is it related to? Means that cannot there is no absolutely we can say there is slow or fast. It is always related to something that the exchange can be called as slow or fast. So that is what is shown here that we talk about slow exchange when the exchange rate between the two hydrogen. So, let me write this now here so we will talk about slow exchange what is slow exchange?

(Refer Slide Time: 25:10)



The slow exchange is when the exchange A to B or B to A is much slower than the difference in the chemical shift between A and B. Okay. So this is a very important parameter in terms of we say slow exchange slow on the chemical shift time scale okay. So we can say that these two hydrogens these two states this and this are undergoing slow if the exchange rate between them the value of exchange rate is much smaller than this.

So remember this is in ppm, so this has to be converted in hertz, so how do you convert this into Hertz, you have to multiply with the spectrometer frequency, okay. So for example, let us say the exchange rate if there is second example suppose exchange rate I will call it as K exchange which we basically sometime it is sum of these two. Suppose exchange rate is let us say 5 per second that means the molecules jumps between A to B 5 times in the second and let us say the chemical shift between the 2 peaks so let us say this is 5.1ppm and suppose this is 6.1ppm okay.

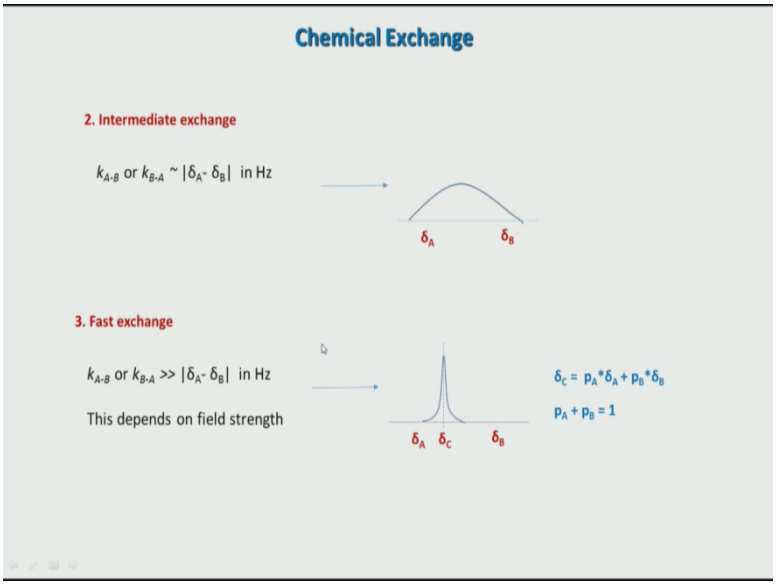
So, now what is the difference? The delta A - delta B if you take the modulus is 1ppm, which is supposed let us say we are doing at 500 megahertz spectrometer, and then it will be 500 hertz okay. So now you can see satisfies this condition because 5 per second is definitely much smaller than 500 Hertz. So, then we can say that these two hydrogens are undergoing a slow exchange means they are going a slow. So slow is always relative to the value and the difference in the chemical shift between the two atoms A and B, the 2 states A and B.

So in state A, it will have one particular value, in state B it will have another value that means there should definitely have 2 different values. They cannot have the same value or nearby, if it is nearby then this condition is not satisfied. So if it is far away far in the sense the difference in Hertz value should be small or should be large compare to the chemical shift exchange value rate value. So this is how we say whether an exchange can be considered as fast or slow. Suppose the distance between these 2 atoms saw in the chemical shift suppose that it is very small, suppose this is 5.9 and or 6ppm, suppose this is 6ppm and this is 6.1 then your difference is only 0.1ppm.

Now, 0.1ppm on 500 Mega Hertz is 50 Hertz, so even 50 Hertz is still much bigger than 5 Hertz, but suppose the Kexchange is also 50 okay. So, let us imagine that K exchange means that hydrogen atom jumps from here to this molecule to this molecule 50 times in a second, so it is 50 Hertz. Then the chemical shift exchange suppose this is 6ppm and 6.1, then the difference is 0.1 and 0.1ppm on 500 Mega Hertz is 50 Hertz. So 50 and 50 both become

almost equal, then this condition is not satisfied, so therefore we have to consider that as a different type of exchange and that is called as intermediate exchange.

(Refer Slide Time: 29:00)



So, intermediate exchange is a scenario where the exchange rate comes very close to the difference in the chemical shift, which is calculated in Hertz and if you go further suppose down the exchange is opposite it is very fast compared to the difference in Hertz then that is called as a fast exchange. So, in a intermediate exchange so we will look at in the next class having look at this 3 different regions, we will see how the peak shapes which are shown here can be explained in a fast exchange and then slow exchange region. So, this we will continue in the next class.