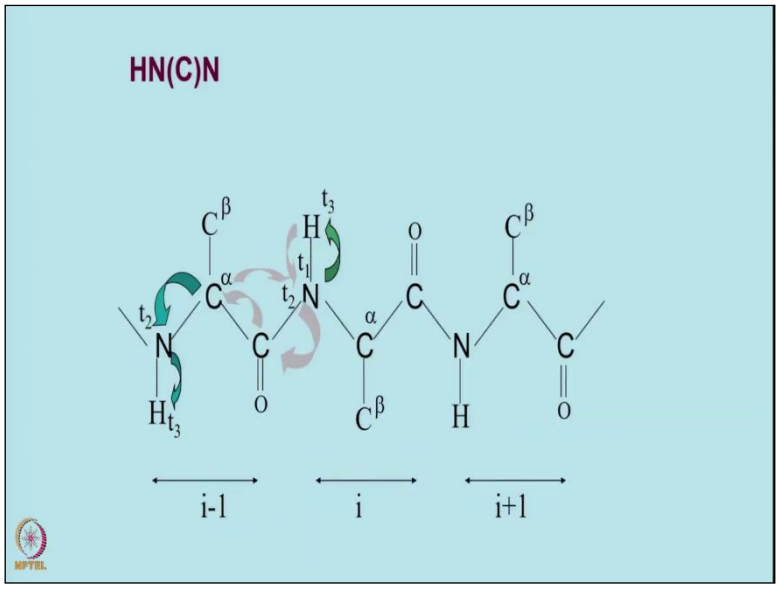
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 36
Determination of Structure and Dynamics of Proteins - 6

So, let us continue the discussion using different kinds of experiments. So we talked about the HNN experiment which is particularly useful for understanding or studying disorder proteins or unfolded proteins because the N15 dispersion was not very crucial there. We made use of that and we generated a kind of a triplet filter through the HSQC spectrum giving correlations between the amides and the N15 of residues i, (i – 1) and (i + 1).

(Refer Slide Time: 00:58)



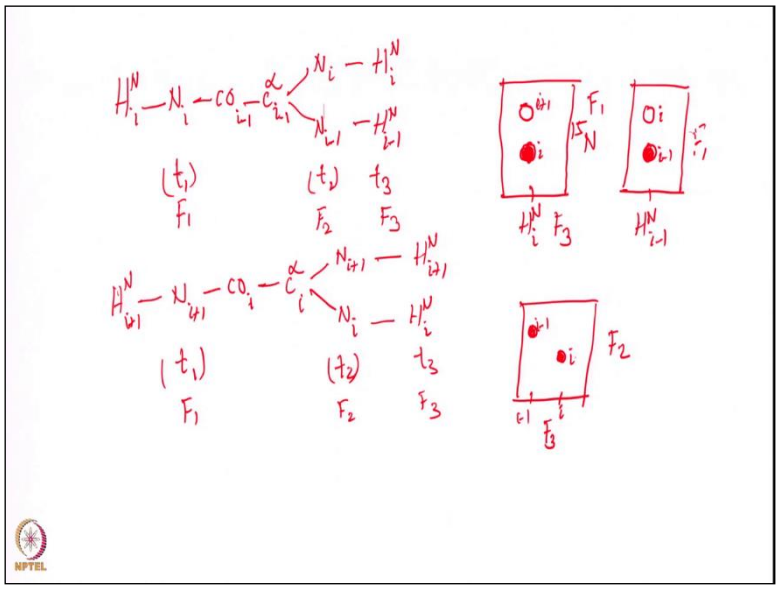
But often it is difficult to figure out which one is (i – 1) and which is (i + 1). So, a complementary experiment is what I am going to describe now. That is called as the HN(C)N experiment which is indicated here. So, the pathway of magnetization transfer is indicated here. Let us say you start from the amide proton of residue i, you go to the nitrogen-15 of residue i and then from here,

now we go to the carbonyl of residue (i – 1). So, this is the residue (i – 1). Now the i is going NH-Cα-CO and this is NH-Cα-CO of residue (i – 1). So, from the nitrogen of i you go to the carbonyl of (i – 1) and then from the carbonyl you will transfer to the Cα of residue (i – 1).

Now from the Cα we go back to the nitrogen of this residue i and also the nitrogen of this residue (i – 1).

So, from here onwards you go to the amide proton of residue (i – 1) and then you go to the amide proton of residue i. So, therefore you get 2 correlations here. You do not have to (i + 1). So, therefore you will have 2 peaks coming from there and this will help us resolve as to which is the residue i and which is residue (i – 1).

(Refer Slide Time: 02:28)



Let us draw this picture here, draw this pattern of connectivities which are coming here. So, I showed you here the residue HNi we go to the nitrogen i, and then from here we go to the carbonyl of i, then from here we go Cα of i. Now you branch. So, from the Cα of i we go to the nitrogen of i and then the nitrogen of (i – 1). And so, and then this one will give me HN of i, this will give me HN of (i – 1).

This is my t3, this is my t2, and this is my t1. So, I did not do anything in between, I did not label the carbonyl chemical shift, labelling was not done. I only label the nitrogen here. So, along the t1 dimension that means along the F1 dimension I will have nitrogen, F2 dimension I will also have nitrogen, and F3 dimension I will have amide protons. Let me do the same for and I will write here also from let us say I start from HN(i - 1).

So, what do I do, I go to nitrogen (i - 1), this is Cα of (i – 1). Then from nitrogen (i – 1) and this is Ni. So, now CO of i and then I go from here to Cα of i then where do I go? From here I

go to the nitrogen of (i + 1), and this will be nitrogen of i, then I go here and this is HN of (i + 1) and this will be HN of i.

Now this is my t2, this is my t1, and this is my t3. So, correspondingly this will be F1, this will be F2, and this will be F3. Now look where does this HNi appear? HNi comes here, if I draw this plane here, I draw here HN chemical shift of residue i and this side is the N15 chemical shift in the F1 dimension. So, this is the F1 dimension, and the F2-F3 dimensions also have the nitrogen.

Now which of these 2 will give me peaks in the F1 dimension? F1 dimension Ni is giving me HN of i, I am getting this peak here. Now N(i + 1) in the F1 dimension is also giving me HNi. Therefore I will have another peak, this is (i + 1) and this is i. So, therefore and this is at a particular amide proton chemical shift, this is along F3.

I am plotting the F1-F3 plane. These 2 starting points Ni and N(i + 1) they both resulted in HN of i and therefore if you were to drew of HN(i – 1), then what will I get? I will get 2 peaks again and this will be HN (i – 1) and this will be i. Now what if I were to look at the F2-F3 plane at a particular F1 position, what will be the peaks in the F2 plane?

Let us draw it here, this is F3 and this is my F2, and let us say I look at the HN of i and where does it come from? HN of i comes from, I will have in the F2 plane which are the things which well I will see a peak which is corresponding to the Ni. Because I am looking at the F2, I should be looking here, HNi is giving me Ni.

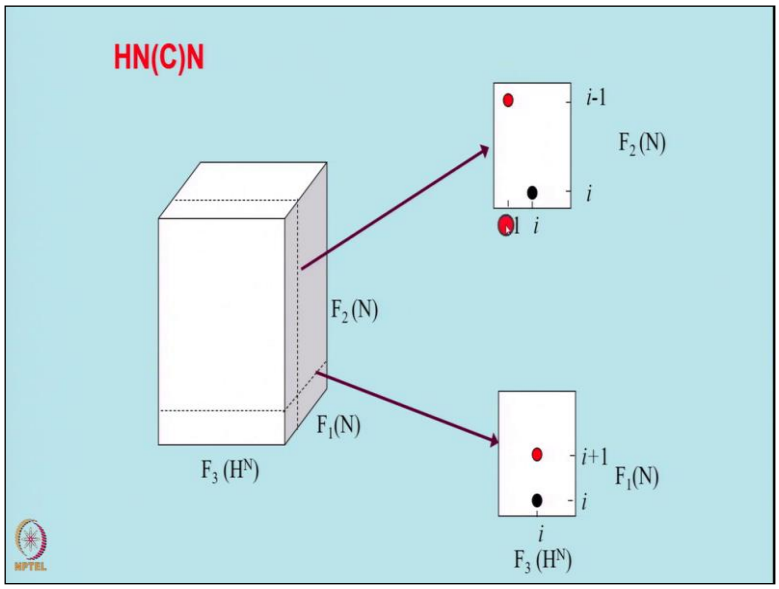
But I also have HN(i – 1). So, I will also have this is the peak and also I have in the same experiment I also have the HN of (i – 1). Originating from HN of i, I am getting HN(i – 1) and HNi both. Therefore I also have HN(i – 1) here and I will have a peak which corresponds to its own peak right and this will be the N15, this is of i and this is of (i – 1).

So, you see in this experiment I will have both the directions established. So, if I took the F1-F3 plane, I know the N15 of the same residue and also of the next residue the (i + 1). But if I took the F2-F3 plane then in a particular plane I will have the peaks at the individual HSQC positions that is i, its own peak and (i – 1) also its own peak, this is the F2-F3 plane. So,

therefore this produces an indication of which direction one is walking along the polypeptide chain.

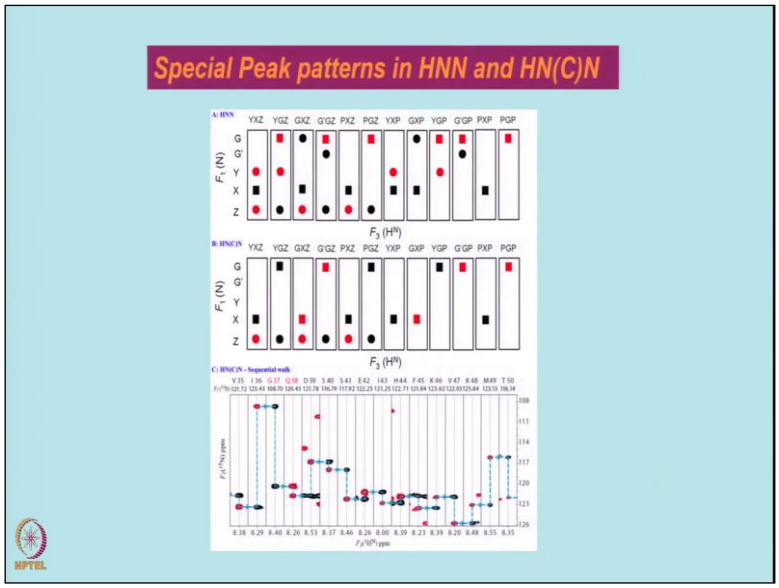
So, this is an extremely useful technique. Then you can straight away jump to which N15 plane you should go there is no need of scanning because immediately that information is already available to you. So, that saves a lot of trouble with regard to the degeneracies in various other chemical shifts. This is particularly useful for IDPs or Intrinsically Unfolded Proteins.

(Refer Slide Time: 09:58)



Now this is the same thing I indicated here, you can see in the F1-F3 plane I have i and (i + 1) at the HN chemical shift of residue i and in the orthogonal plane the F2-F3 plane, I will have the self-peaks here i and (i – 1) and these are at respective amide proton chemical shifts as well. That is what I showed you explicitly by the calculation.

(Refer Slide Time: 10:26)



Now there is one other important feature, if you noticed there is, in the previous slides, I had used different colours for the self-peak and the sequential peak. Similarly I had used different colours in the HNN experiment also for the self-peaks and the sequential piece. So, there is a particular significance to that. The significance is the following, these indicate positive and negative peaks.

If I take the black peaks as the positive peaks, the red peaks are the negative peaks. So, typically this is the HNN spectrum. In the spectrum we showed already that there is going to be three peaks, as the squares are the self peaks. For example, I took this particular sequence GG'YXZ. This is the particular sequence which I am looking at. This is the protein sequence going from GG'.

There are 2 G's here and Y here and X here and Z there. This is the sequence what you are looking at. So, in this case if I take the residue X, what are its neighbours the neighbours are Y and Z. What are Y and Z? They can be any residue other than the glycine, it can be any residue. So, then what I see, for the residue X, the self-peaks it is a positive peak and the 2 sequential peaks which are there are negative peaks.

Now the situation changes if there is a glycine somewhere. The glycine actually produce a whole different lot of patterns. For example, if I look here for the pattern, if there is a YGZ and then, for G as the central residue, I will have the self-peaks which is negative and the self-peak for Y, this will also be negative and GZ which is on the (i + 1) side that will be positive.

Therefore you see here, the self-peak is negative, the sequential peak on the (i – 1) side is negative, but the one on the (i + 1) side is positive. Now we look at GXZ, if there is a triplet sequence like this, then the X is the central residue for the triplet, the central residue X has a positive diagonal means the self-peak and the G is (i – 1) that is a positive peak, and the sequential (i + 1) peak is negative.

If I have this sort of a thing G'GZ, suppose this is the kind of a sequence I will have, then the central residue is G, (i – 1) residue is also G, and the (i + 1) residue is Z. In this case, the G in the center has a negative peak which is the square and its one which is on the (i – 1) side is positive and the (i + 1) side is also positive. Of course if there is a proline, then you do not have a peak because proline does not have the amide proton. Here you will have the X, the central residue has a positive peak and the one which is on the (i + 1) side that is negative.

If I have a PGZ then the central residue is G that is again negative and the Z is positive. So, one thing we generalize here is that the self-peak for all the glycines no matter where it is it will always be negative. The self-peak of G will be negative and the sequential peaks will depend what the residue on the (i – 1) is, on the (i + 1) is. So, you will get therefore different kinds of peak patterns here, combinations of positive and negative peaks in this kind of a sequence.

So, this is typically indicated here, how these ones are appearing. Now in the case of HNCN, this will also have similar kind of a features. Now as I said, I will have the i and the (i + 1) in the case of HNCN. Suppose I have this YXZ, gives me the self-peak as positive and the sequential peak is negative, which is the similar to this except that I do not have the (i – 1); but now you see here, this fellow the YGZ the G is in the middle.

I will see only i to (i + 1) right. So, i to (i + 1) this both the peaks are positive. In this sequence the G is in the middle, both the self-peak and (i + 1) peak, they are both positive. Now if it is GXZ, if G is on the (i – 1) side, X is in this middle, and Z is on the (i + 1) side, both these peaks are negative. X is also negative. I do not see to the (i – 1) you see to the (i + 1) therefore the (i + 1) is here and this both will be negative. This is very interesting here you see here these 2 are negative YGZ, GXZ.

So, these are negative and if you have the G'GZ. So, the middle one is G, G that is always negative, and the (i + 1) here this is now positive. So, like that you get different kinds of sequences here and that is what is illustrated with this particular experiment here. This is the experimental spectrum, this is the sequential walk through the experimental spectrum.

So, what is the sequence here V35-I36-G37-Q38-D39-S40-S41-I42-H43-F44-V45-K46. So, therefore there is a G here, these are called as check points. Wherever you see a G, the glycine it provides a peak pattern change with regard to the combinations of positive and negative peaks. Therefore, these become checkpoints along your polypeptide chain. They will allow you to identify the peaks in an unambiguous manner sequence.

So, let us look at that here, let us say we start from this. So, what we have here? This V35-I36-G37. Now this fellow at this V35, I see 2 peaks and this shows a self-peak and the sequential peak, the sequential peak is negative. So, what is the sequence, this is very similar to this YXZ. So, I have here a positive self-peak and a negative sequential peak here. This is like the X and Z here, because the next residue is also I, V and I.

Now we come here to this. So, this is now, i and (i + 1), (i + 1) is a glycine. How does this correspond to? That this situation is here. So, this is I36. So, this pattern is where we should take, that is this one here, G is in the middle. So, we have to take actually this one only YXZ. Now the next one is gly. So, we should check YXG.

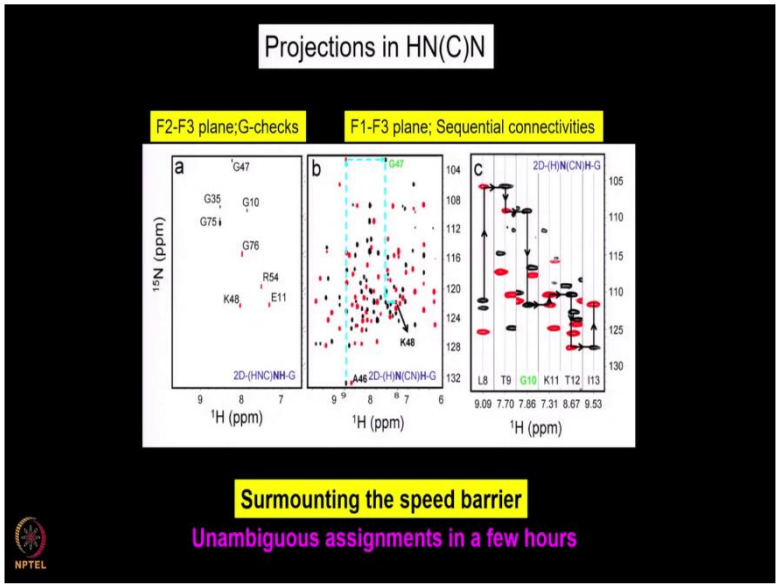
So, we should see YXG, it is there. Now that is this is like a YXG. So, this fellow has a positive peak for the self and a negative peak for the (i + 1) right. So, (i + 1) is G and therefore that is here. In this situation, this (i + 1) peak will be negative if it is that pattern is not shown here. We can see here this pattern will be negative, the self-peak is positive and the sequential peak is negative.

Now you go to the next one, the next one is G here and what will be its peak pattern. So, this will be similar to YGZ, the central residue is G and the other 2 residues are non G's, and therefore what I should get I should get this sort of a pattern YGZ and you see you have both the peaks positive, the self-peak is positive, sequential peak is also positive. And the immediate next ones will be negative, negative, that is the same as here GXZ.

Now this becomes a GXZ, therefore this will be negative, negative. Therefore you see you can walk along the polypeptide chain. You check here go to this pattern of positive-positive and negative-negative here. Then after that it will all follow, positive negative positive negative and so, on so forth. So, you can walk along this way and reach up to the end point. So, T50 here and once again we see for the T50, we see set of a positive peak because the next residue is similar to what is the pattern which is present here.

So, this is how you walk along the polypeptide chain by making use of this pattern of peaks. Therefore these various glycines which are present these provide extremely useful checkpoints in your sequential assignment procedure. And there can be no errors here, because if you made an error then you will not hit the proper peak pattern.

(Refer Slide Time: 20:18)



Now you can go to the step forward and suppose you took projections of this spectra, these are three-dimensional spectra and if you took projections this is the way they will look. What these projections are, I have this projection, the F2-F3 plane that means I go down the F1 plane and take the projection in the F2-F3 plane and plot.

If I do that, because of the various cancellations that will happen you will only see peaks coming from the G's and the residues next to them. For example here you see G47-K48, G10-E11 and then you also have this R54 and these are also coming G75 and G76. So, therefore you will see only this i and (i + 1) projections which are coming out here and they also have different signs.

So, therefore by looking at this alone you can figure out where your glycines are, this is F2-F3 plane projection. On the other hand, if I took F1-F3 plane projection, I have 2 kinds of things here one is the normal HSQC spectrum, then I will have the sequential peaks the sequential peaks to the (i + 1) from every residue and they have different sign patterns as in the case here.

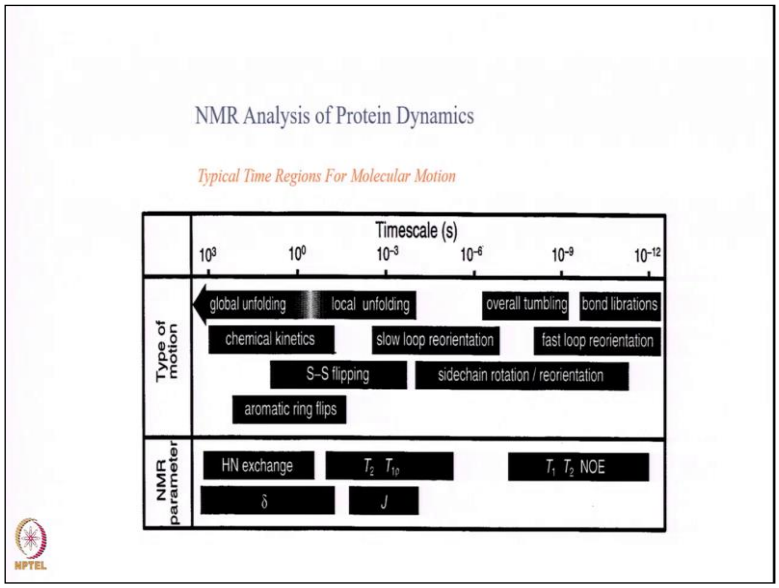
Therefore, all of them will be present in one plane. Therefore this will allow you to walk along this entire sequence of the polypeptide chain in one spectrum and this spectrum can be recorded very rapidly. This is because this will be 2 dimensional spectrum. What you need to do is if I want to record only the F1-F3 plane I do not increment the t2 time period, therefore the t2 time period is kept constant.

So, I will record directly the F1-F3 plane only. If I want to record F2-F3 plane directly, then I don't do the t1 increment I will get only the F2-F3 plane. Now this is an illustration of the sequential connectivities using this spectrum. This is just an expansion of that which is from the same here. So, you can see you can make a walk through the polypeptide chain here from L8 to T9 and you see those thing here is the line is drawn at the center of the of the column.

So, you go from here to here this positive negative, you go here positive to negative, you go here. Now you see positive-positive this is because of the glycine. You got the glycine here you got the positive-positive peak pattern, then you go immediately after that you have the negative-negative peak pattern. So, that is what it should be, right. So, you get the negative-negative and then you go to the next once again you have the positive-positive, this is the sequence it goes in the normal manner if there is no glycine.

So, wherever there is a glycine, the one the same number and the one which is after that these 2 have very characteristic peak patterns, you have the negative-negative here and positive-positive there. So, this will allow you to identify where you are along the polypeptide chain in a simple manner. So, therefore you can obtain such large number of assignments very rapidly for unfolded proteins and that we call it the surmounting the speed barrier you can get the assignments in a few hours.

(Refer Slide Time: 23:52)



So, now I think we are going into a slightly different topic that, once we have done this analysis of the individual protein chains then you have to study what are the kinds of motions present. We are talking about the intrinsically disordered proteins. Intrinsically disorder proteins have various kinds of dynamics, all other proteins also have the dynamics present there. And protein dynamics becomes an important part to study that. So, there are protein dynamics will have various time scales of operations and these are the typical time regions for molecular motions and that is indicated here.

So, what we have done here is classified the various time scales. So, we may not go into the entire details that I only indicate here. What sort of a time scales appear in the proteins and the disorder proteins will have all of these kinds of time scales that is the reason I am showing here. So, we have been talking about the intrinsically disorder proteins various kinds of motions are possible there.

The conformational fluctuations can happen, but this general slide, which will show you what kind of time scales are available for what kind of processes. So, typically the time scale can run from this 10^3 seconds to 10^{-12} seconds. So, these are 10^3 is extremely slow and this is extremely fast. So, what sort of time scales are they, this 10^{-12} of bond liberations, 10^{-9} are overall tumbling, this is the nanosecond time scale.

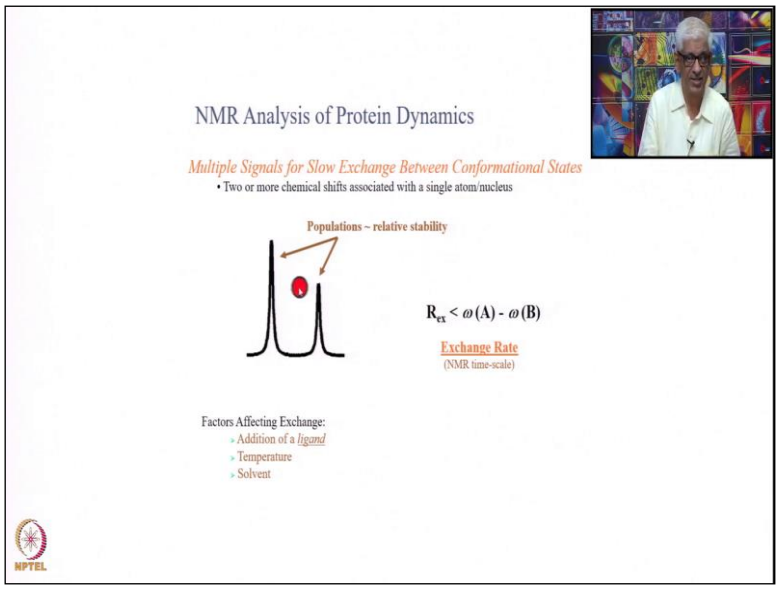
And then you have in the microsecond to millisecond time scale, you will have slow loop reorientations. This will happen in your unfolded proteins. Then you can have local unfolding, there is the local fluctuations there, this also this will also happen in the millisecond time scale

motions. And global unfolding, entire protein to unfold, this will happen in seconds to hours-time scales.

And typically of course chemical kinetic reactions also happen in this area and S-S flipping and side chain rotation etcetera these will happen in this time scale. Now how do we measure these ones, this goes into different kinds of experiments and those are various tools available and those are typically indicated here, which time scales can be studied by what NMR technique. You have the hydrogen exchange possible, then you have the relaxation time measurements there, and these are heteronuclear NOEs and relaxation time measurements.

And the coupling constant of the chemical shifts these are the various parameters which one uses to obtain this kind of information.

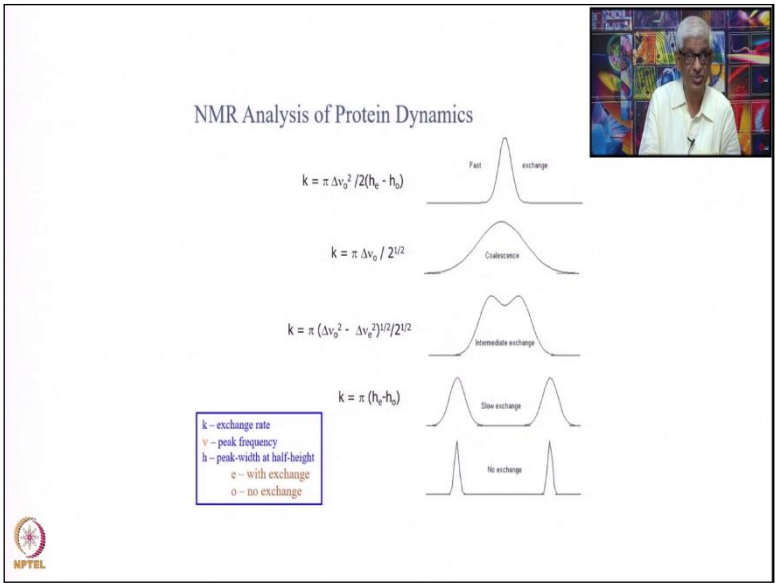
(Refer Slide Time: 26:33)



A simple example is indicated here. So, if you have the chemical exchange happening on a slow exchange between conformational states. You have one population here, another population there, and these ones are 2 different signals for the same proton or the same signal which is undergoing chemical exchange, you will have 2 different intensities. The populations here and here may not be the same.

So, that depends on the energy difference between the 2 conformational states. So, you have different chemical shifts for the 2, and if the exchange between these 2 is slow compared to the chemical shift difference between these 2, this we have studied earlier also. And then you will have this kind of a peak pattern which is present.

(Refer Slide Time: 27:23)



But if the exchange rate increases, it goes faster then you start seeing changes in the peak patterns and that is what is indicated here. So, depending upon what sort of exchange rate is present, you will have the peak patterns appearing like this. This is slow exchange, this is also slow exchange but the lines are broadened. You increase the exchange rate by increasing the temperature or different things then they slowly start merging here. Then the coalescence temperature, from the coalescence you can calculate the exchange rates here; and very much fast exchange everything will merge and you will find a single line here in the middle and that will be at the weighted average chemical shift of the 2 populations there. So, this is what will happen in the case of intrinsically disordered proteins. Even in the folded proteins, there can be certain domains where there are slow exchanges, there are domains, which can be fast exchanges and all of these can be studied by using NMR. I think we can we can stop here.