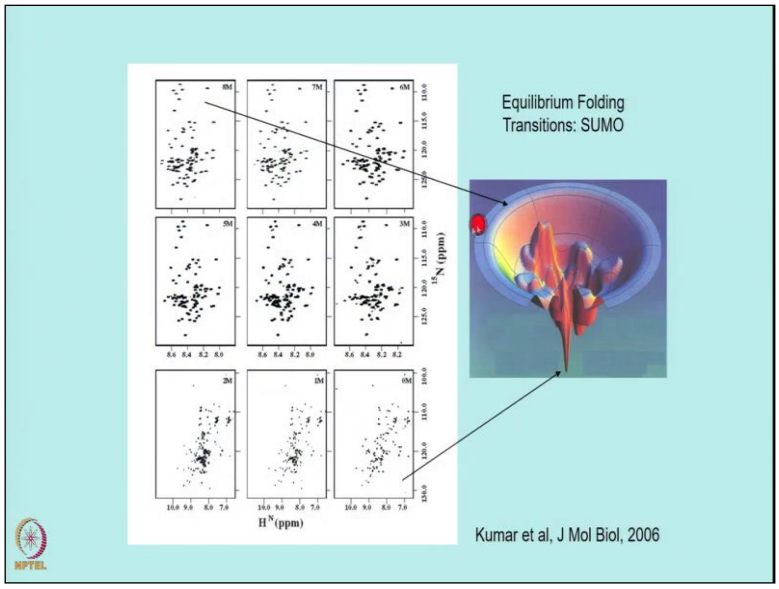
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 39
NMR Analysis of Protein Dynamics - III

(Refer Slide Time: 00:34)



So, we have been discussing about the protein folding phenomena. To just to have a quick recap, we said this sort of a diagram describes protein folding; this is called as a folding funnel. So, we can look at the folding transition this is the native state here and then the top you have the unfolded state which consists of trillions of millions of many different conformations and each one of the molecule in your ensemble may have different conformation.

And it will eventually start folding down the funnel crossing through various energy barriers and eventually lead to a single confirmation or maybe a related small number of confirmations here. This is called as the native state and symbol here. So, it may be few conformations which are very similar to each other by-and-large. So, to monitor these equilibrium folding transitions, they will also belong to the entire set of transformations that are possible for any protein.

We can start with the denatured state here. We create a denatured state by using urea or guanidine hydrochloride or whatever, some of the denaturing agents or DMSO; and then you slowly start changing the conditions to induce the protein to fold that means induce the protein

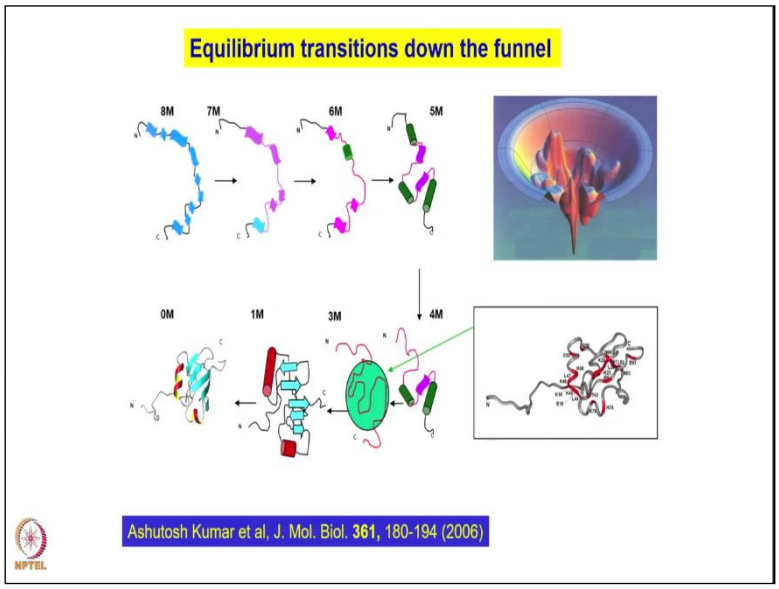
to go down the funnel here and then you will start seeing changes in the HSQC spectra. These are ^{1}H-^{15}N HSQC spectra, amide proton N15 HSQC spectra.

And you will see that there will be small shifts in the peaks and the peaks will start becoming narrower and things like that. See this spectral dispersion will increase as you go down and this is at 0 molar this is the folded state. In the native state, the spectral dispersion is very good, it goes all the way from 7 ppm to 10 ppm, and you see here in the unfolded state it is only 0.3 ppm dispersion there, therefore the peaks are looking broader in this in this situation.

And you start going down, there will be changes in the chemical shifts, there will also be changes in other parameters like the relaxation properties which we will use. Because as the protein starts to fold some new contacts will be made, when the contacts are made between different residues they will put restrictions on the free movements of the individual residues, therefore the dynamics in the protein gets reduced.

So, this will reflect in the relaxation times, the relaxation rates, particularly the R_2s, the transverse relaxation rates, they will get reduced. Therefore changes in the relaxation rates will be good monitors for seeing the folding of the particular protein. The motions will start slowing down and that will be reflected in the individual residue wise changes. So, that is how we can identify which residues are folding and what is the pathway, how the protein is trying to fold.

(Refer Slide Time: 03:16)



So, this is what we showed here in this particular slide, you start from an unfolded state which is completely random and very high frequency motions are present and as you start going down

the motions will change. So, this is indicated by the change in the colour here and the motions will start slowing down, that will reflect in the changes in the R_2 values, R_2 values will start going up.

And then as the structures are getting formed you will start seeing secondary structural preferences, in this area there are no structural preferences. But here you will get some structural preferences, they are not permanent structures still they are structural preferences. So, you get a helix made here and the motions will further increase, they get restricted; and as you go further down 5 molar, you will see more helices are formed transiently.

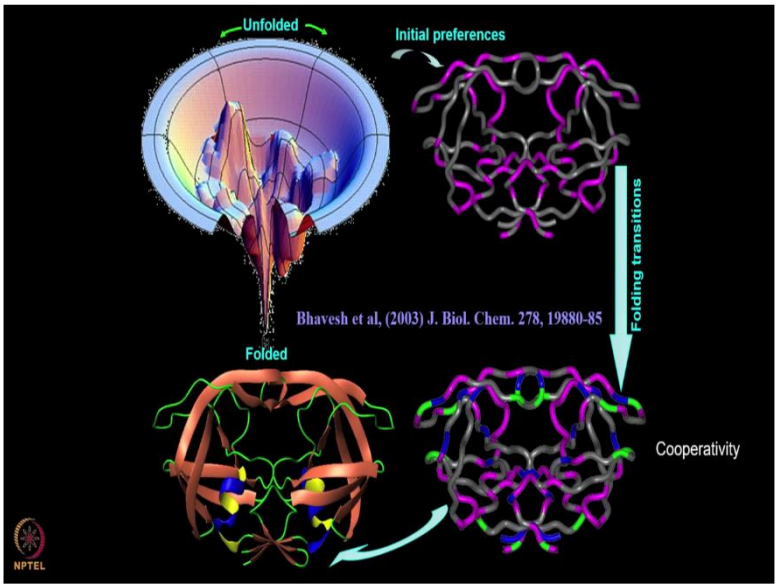
So, they make and break, make and break, that sort of a situation happens. Go to 4 molar some of these, one of this helix actually disappears and these two helices are there. When you go to 3 molar, so all the helices have disappeared and then you will see it has no structural propensities. But you see here the R_2 values have increased at various places and this is the final structure of the protein and what is coloured in red here, these are the places, where the R_2 values have gone up.

That means these residues are coming closer and going away, coming in contact and going away, things like forming and breaking, coming from close and going away, and things like that. So, there is exchange here, this is the kind of a folding-unfolding transitions happening here. These residues are coming close and then going away. So, therefore these are experiencing an exchange process.

The exchange process results in increase in the R_2 values and therefore you see, that is an interesting thing, that at this point the protein is getting ready to form the native kind of structures. These structures were non-native, therefore those structures were removed and then once it is formed here then this kind of a structural transitions are happening inside the ensemble; and then slowly further reduce the urea concentration, you start seeing the β-sheet formed and then the two helices are the native helices.

And then it forms a very stable structure at 0 molar urea concentrations, that means in absence of the urea. So, this is how you can monitor the residue wise changes that are happening in the in any given protein.

(Refer Slide Time: 05:34)



Now here I show you an example of the folding done in the similar manner using guanidine hydrochloride, but this is with HIV protease. HIV protease, it has an intrinsic preferences here, initial preferences that these are called the photons. So, these are the areas which the folding can get initiated. this is a dimer here. This is called the flap here. So, individual two monomers are held together by kind of a β-sheet at the N-terminal here and then the C-terminals come close.

And so, therefore this is the initial folding preferences in the in the protein. So, and then you initiate the folding here, then you start getting more folding preferences, that is again indicated by the colour code. So, you get the blue ones, and then the green ones, and then the red ones, you start getting these. On what basis? This is on the basis of the slowing down of the relaxation rates,

the R_2 values. Now you see, here the ones are adjacent to each other or the changes that are happening they are adjacent to each other, which means when a particular portion of the protein folds, it induces the neighbouring fellows to fold and this is called as a cooperativity. So, there is a cooperativity in the folding, this is clearly demonstrated in this example of HIV protease and this is the final structure of the protein.

So, it is a β-sheet these are all β-sheets here. What is shown here is of course I have shown the same structure not that these ones are in the initial phases, these structures are present, but these are the preferences which are present, and where the protein gets kind of propensities to fold in that manner. So, that is why these ones are indicated on the same structure here.

So, there is a cooperativity of the structure formation. The initially it will be preferences and then eventually they will form the structures and this is the flap barrier, this is the so-called active site of the protein. The active site of the protein is at the interface of the two monomers and this flap is also very dynamic, and this dynamicity of the flap is responsible for the protein function, the substrate will have to come in here and then go out, the flap opens for the substrate to come in, then flap closes.

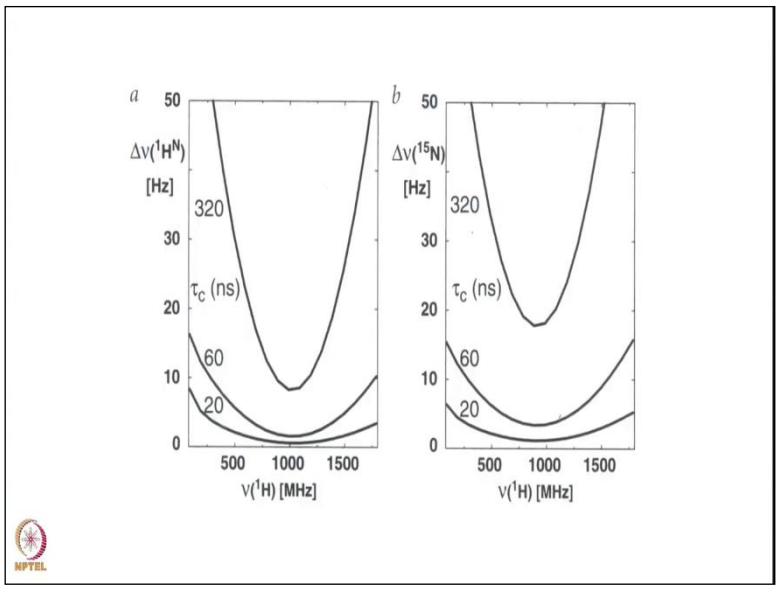
And then the cleavage happens, this is the protease right. So, it is supposed to cleave the proteins. So, the substrate comes here and once the reaction is over the flaps open it goes away. So, therefore dynamics of the protein also plays a very crucial role in the biological function. I will also illustrate this to you later with another example and this is how the protein folding happens.

So you have initial preferences and then you provide the folding conditions, there will be cooperativity in the folding transitions and they will eventually find way to the final folded structure. And the final folded structure here is it has one helix and it is a whole series of β-sheets. There are two β-sheets here, overlaid on the two monomers and it forms a flap.

(Refer Slide Time: 08:38)

So, now so far we are looking at small proteins. Now we see how NMR can be used to study large protein assemblies. When I say large not some 100 amino acid residues, 150 amino acids, things like that. But these are of several mega Daltons, 100s of kilo Daltons of molecular size. How do we study such large protein assemblies?

(Refer Slide Time: 09:05)



So, what is the problem in NMR to study such large protein assemblies? So, here is the slide which shows the line widths in the protein signals the according to the molecular size. What is plotted here is line width on this axis. This is for the amide protons and this is for the N15 and on this axis the spectrometer frequency in megahertz. And these various numbers indicated on the curves these actually indicate the correlation times.

The correlation time reflects on the size of the molecule, 20 nanosecond correlation times it is a relatively small molecule, 60 nanosecond is larger, and then this is 320 nanosecond, this is an extremely large molecule. They can be assemblies, these are multiple copies of the same protein can assemble together in an order form to form large assemblies. In such a situation, you can see how the line widths change in both the proton axis as well as the N15 axis.

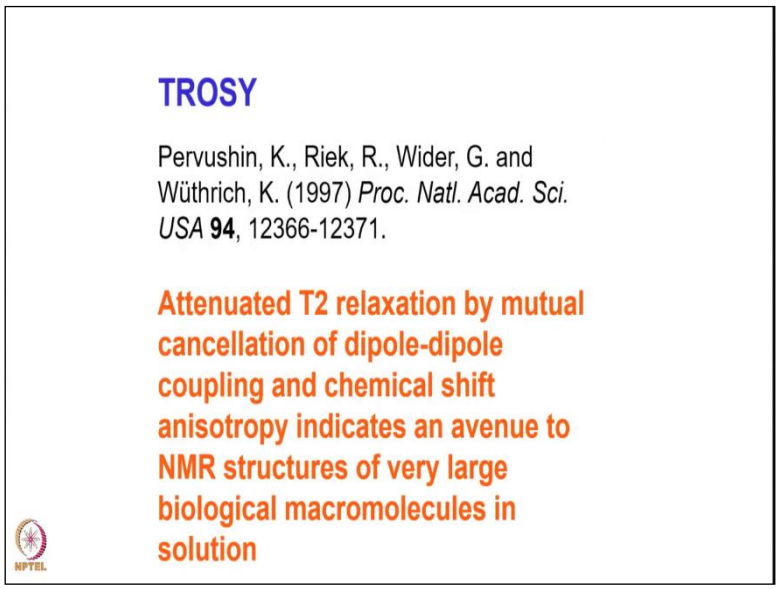
The line widths are so large, you see the 50 hertz and it is 30 hertz and things like that. So, these are very large line widths. So, when that is the case, it is very difficult to see the signals in the HSQC spectrum, and of course this is also field dependence, it is also indicated here. So, if you see the line widths are smaller at this particular frequency, this is almost one gigahertz. Actually, this makes the case for going to larger and larger spectrometer frequencies especially if you are going to larger molecules.

Of course, when you are dealing with small molecules, it does not matter. So, small molecules that 20 nanoseconds, you can anyway from here to here if you go, it does not matter the correlation time is 20 nanoseconds and the line widths have not changed much. So, from here

to here, it is about the same line widths. So, therefore it does not matter. When it goes to larger molecules, it becomes very crucial to go to higher magnetic fields where you have a smaller line width.

So, you see it is 10 hertz here at 1 gigahertz, if you take at 500 megahertz it is about 30 hertz, and you go further down of course it goes to 50, and so on so forth. So, therefore as you go higher in molecular size, you also need to go to higher in a spectrometer frequency to get sharper lines. It will have important implications, as we will see this is true for both the proton line widths as well as N15 line widths. What is the consequence of this, we will see here.

(Refer Slide Time: 11:29)

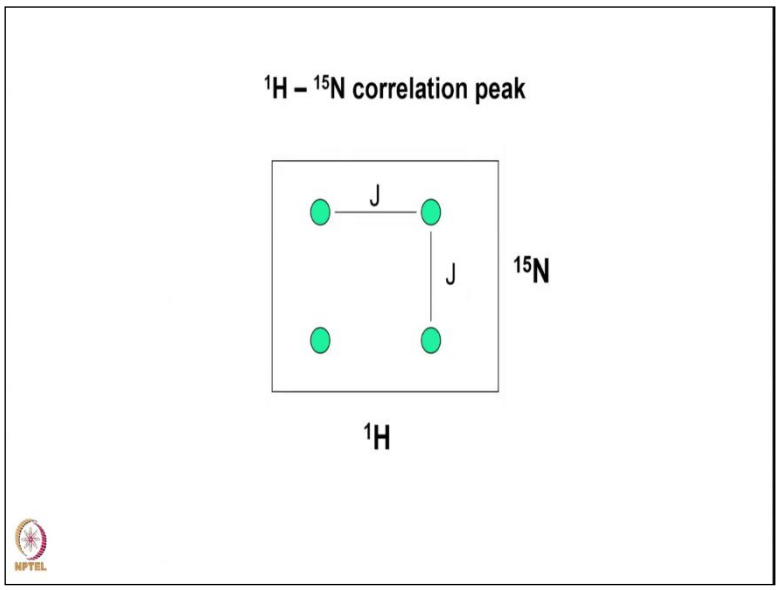


So, therefore to make use of this different line widths, depending upon the spectrometer frequency, to generate a better spectrum with the better line widths, a technique was developed which is called as the TROSY. This is Transverse Relaxation Optimized Spectroscopy. This was developed by Pervushin in the laboratory of Kurt Wuthrich.

And this actually became extremely useful, this is called as attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. We will not go into the theoretical details of this. Basically it is make use of a kind of interference between chemical shift anisotropy mechanism of relaxation and the dipole-dipole coupling mechanism of relaxation.

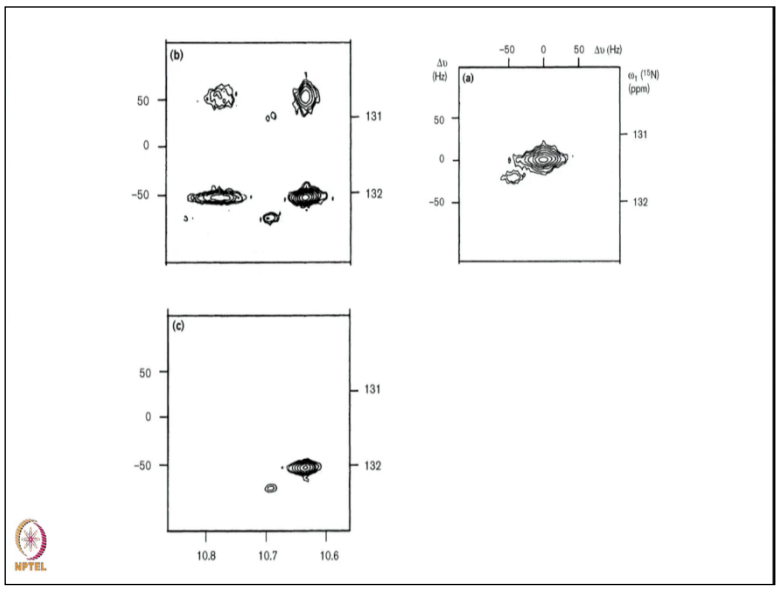
So, the two things go in opposite directions at some stage they will sort of cancel to produce very small very small line width

(Refer Slide Time: 12:34)



and that is indicated here. So, suppose you are taking in a HSQC spectrum; in an HSQC spectrum, it will have on the proton axis, the amide proton is a doublet with an proton N15 coupling, N15 is also doubled with the same coupling provided we did not decouple this amide-proton-nitrogen coupling. If we did not decouple, it will have 4 components here. But all these 4 peaks are different characteristics with regard to the relaxation.

(Refer Slide Time: 13:13)



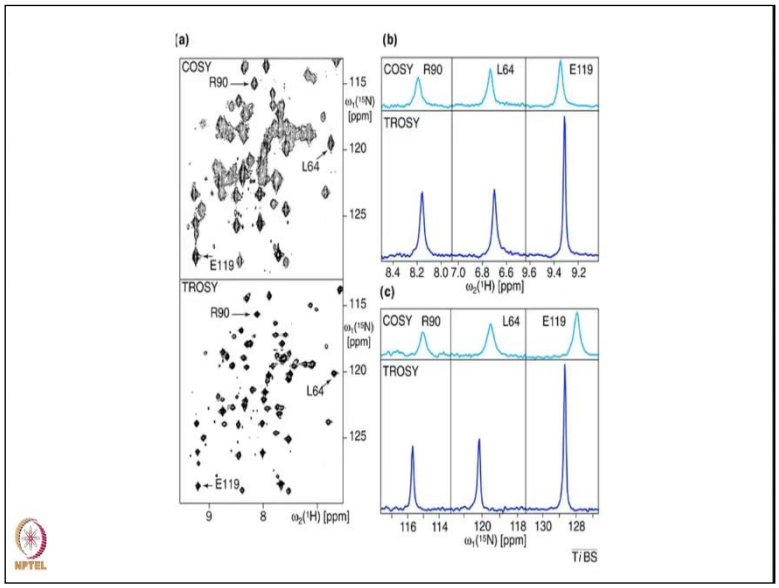
And that is indicated here. So, if you see here. So, typically in the normal HSQC spectrum this is what we will see, because we decouple this we remove this coupling, we remove this coupling, therefore it will appear a peak at the center here. This is one peak is what we will get

in the normal HSQC spectrum, which is what I showed you earlier. Now if we did not decouple, then the couplings are present in both cases there are 4 components here.

Now all these 4 components of the peaks have very different line widths and therefore you can see their intensities are also very different. So, among these, this is the sharpest and these ones are relatively less sharp and then of course the broadest. So, therefore different ones are different components and this is the point one makes use of in the TROSY to say let me pick up this only forget about these ones.

I will do a technique, whereby I can pick up only this component of the correlation peak. Then you see that is what you will get in this spectrum here. This is the so called TROSY spectrum and you will have only one peak here, this is much sharper in intensity and much larger. You will see that example immediately in the next slide.

(Refer Slide Time: 14:24)



See here. This is the normal HSQC spectrum, it is also called as COSY, because it is a N15-proton correlation spectrum, it is called COSY. But you see this, the normal COSY spectrum which is looking like this; now in the TROSY the same spectrum looks like this, because you have picked up the sharpest component of the multiplet. Notice therefore that you cannot afford to decouple the proton nitrogen 15 either in the ω_1 dimension or in the ω_2 dimension.

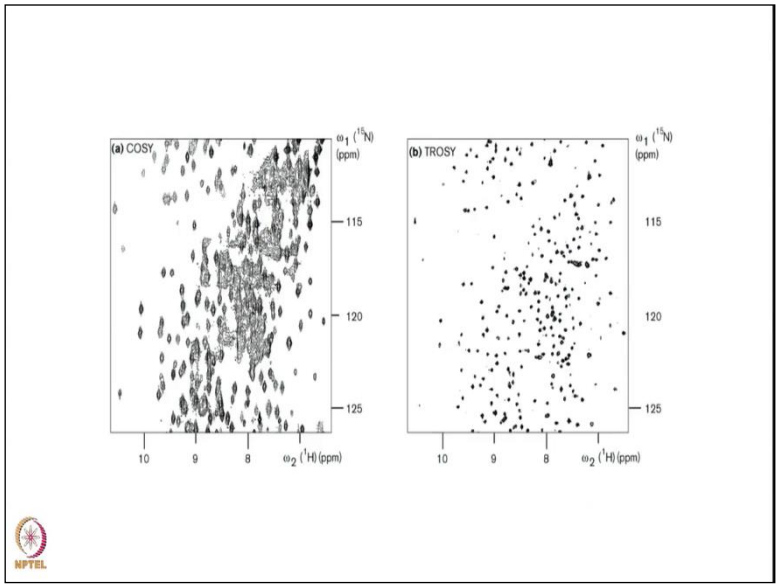
And then you to device the pulse sequence in such a way that you pick up only the one component of that 4 and that will be the sharpest signal and that is what is represented here. It will reflect your amino acid residue anyway. So, therefore you see how sharp these ones are.

Now you take this cross section, there are 3 cross sections are taken this peak, this peak, and this peak. These 3 cross sections are shown here, this is in the COSY.

So, you take a cross section like this, horizontal cross section in all these positions. So, it shows the line width along the F2 dimension. So, here it is case of the proton. So, cosy all these 3 look at the TROSY peak intensity, this is so much more larger. The noise level is the same, this is so sharp. Similarly in the N15 axis also, this is the COSY and this is the TROSY. So, such a sharp line here for each one of these.

And therefore naturally if you do this, your resolution in the spectrum is going to be higher and you will be able to see larger number of peaks in the spectrum.

(Refer Slide Time: 16:00)



And this is the very very exquisite illustration of this gain in the resolution. This is a large protein of 130 kilo Dalton or something like that. So, this is the COSY spectrum or the HSQC spectrum and you see here is the TROSY and the peaks are so well resolved and sharp, you can count these peaks, measure their intensities, and all kinds of things you can do. Therefore you can actually work with large proteins in this situation.

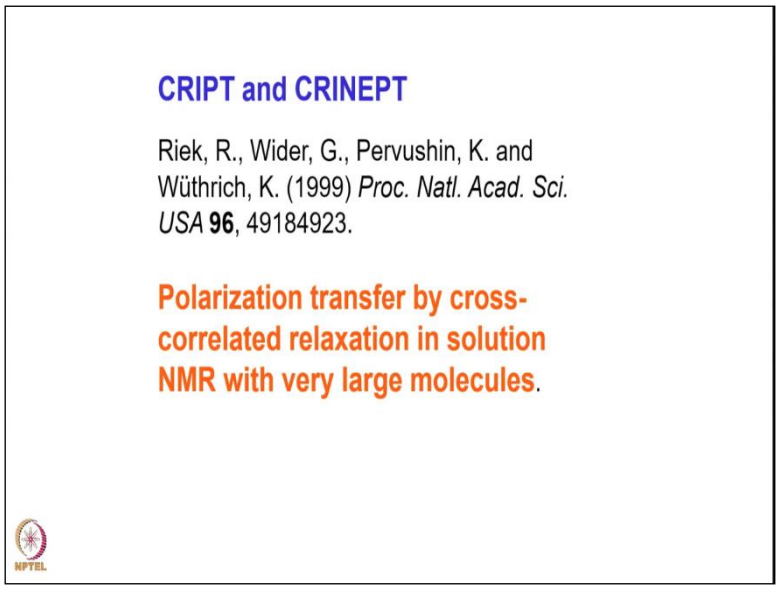
It may be a single protein or it may be and it may be an assembly of smaller molecules, but eventually lead into a larger molecular size. So then either case one can use this. Notice however, that if you use a very large spectrometer frequency for small molecules, you actually

stand to lose. Why you stand to lose in this case, because I will show you that here. So, you saw here that we are picking out one component of the 4.

In the normal case, we are getting contributions from all the peaks here. If you are working at a small molecule then all these components do not have very different intensities. Therefore, when you decouple this will be very sharp peak with high intensity. But if you did a TROSY experiment on a small molecule of 10, 20 kilo Dalton, then you are throwing away the intensities of all of these and picking out only one them, of course you actually tend to lose the intensity.

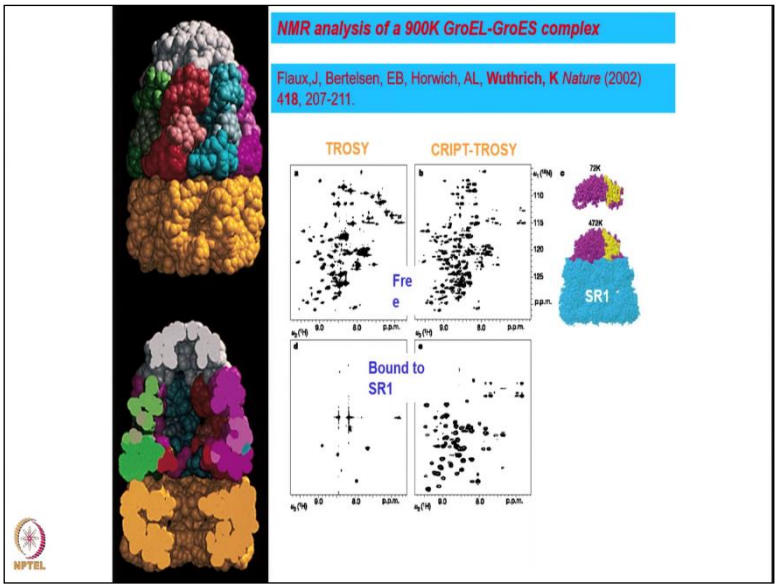
So, therefore we should not do that. You should use TROSY only when you are working with the larger systems and then go to higher spectrometer frequencies.

(Refer Slide Time: 17:39)



So, then of course one went even further to develop two more pulse sequences, this is done by the same group once more, this is called the CRIPT and the CRINEPT. This depends upon polarization transfer by cross-correlated relaxation in solution NMR with vary large molecules. So, we will not go into the theoretical details of this, but I will simply show you the result that was obtained in this.

(Refer Slide Time: 18:00)



Now this was an application to a 900 kilo Dalton GroEL-GroES complex and all of you know that this is actually a chaperone. This is the structure of this chaperone, this is determined by x-ray crystallography, this structure is known. So, it is not that you are determining the structure, what is the intention here is only to demonstrate the application of the technique to observe signals even in such large complexes.

Now you see here, this is the structure and the same thing is spread out in this particular form, these are 3 different ones here, you have SR1 and then you have this domain here, and then the domain here. So, all of these sit on top of each other there are multiple copies of a particular small protein here. In this case there are seven copies of a small protein which is 10 kilo dalton in size.

So, therefore there are seven copies of this, one is trying to look at the signals from this particular protein. What you do is? You can actually reconstitute this protein with some portions labelled and some portions unlabelled; and then you can see what you are going to get. So this is the spectra in the free this one assembly without these ones there and then of course you can see reasonably good signal here.

Now the moment this one is bound to SR1, that is one this all these are present there together then you will see the signals have completely disappeared, vanished in the TROSY. So, this is what is trying to demonstrate here, that where the TROSY also becomes insufficient, then you can use the CRIPT-TROSY and you see this is the spectrum which is recovered when it is bound to SR1.

And this is free when it is free of course you are seeing this in the CRIPT-TROSY this actually has much larger number of peaks compared to this and this already has shown a certain enhancement. But if even when it is bound to SR1 it is bound here then you will still have a large section of the peaks which are present here. So, this is how this technique can be used and this was published as you can see in nature in 2002.

And this was an important development. So, the point to remember here is that you must use this techniques only when you need to work with very large very large molecules, large protein assemblies it is possible to study the association process. Therefore, you can stretch your application to very very large systems from small molecules to large systems.

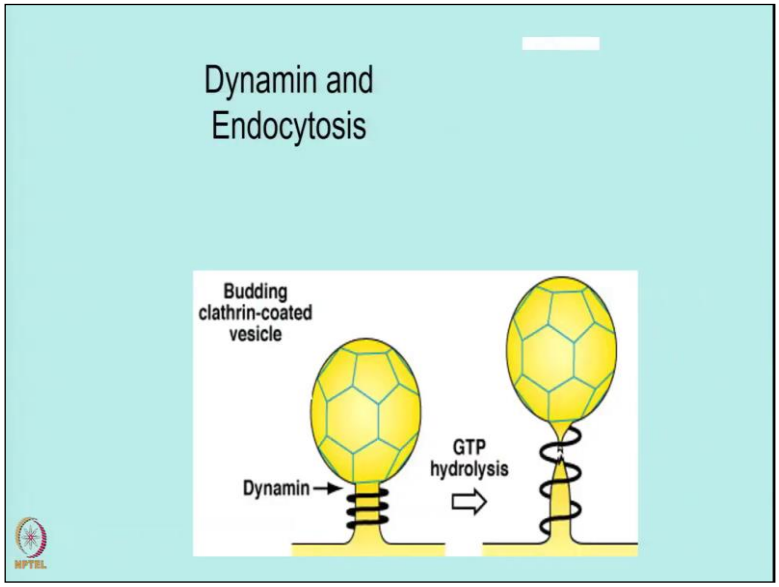
(Refer Slide Time: 20:52)

Now going further, from this we would like to see how one can see self-association. In the previous case what we saw was that the associated state is particularly known, the structure is known. How the association is happening is already known in these cases. So, we know the structure, therefore we know the pattern of association of the individual molecule, the structure is not changing from individual molecules to in the associated state.

Because the spectral peaks have the same pattern there and therefore, on application, this demonstrates that you can actually see the signals even in large ensembles. Now, when you are working with a small molecule which actually associates so, this is very common phenomena in protein NMR that or whenever expressed many proteins tend to self-associate.

Now how do we understand the self-association process? Does it lead to an association which is constructive or destructive? Is it functionally relevant or non-relevant? So, therefore self-association is an important phenomena to be studied in biology. Therefore we try to see how one can study this. I will illustrate this to you with particular example here.

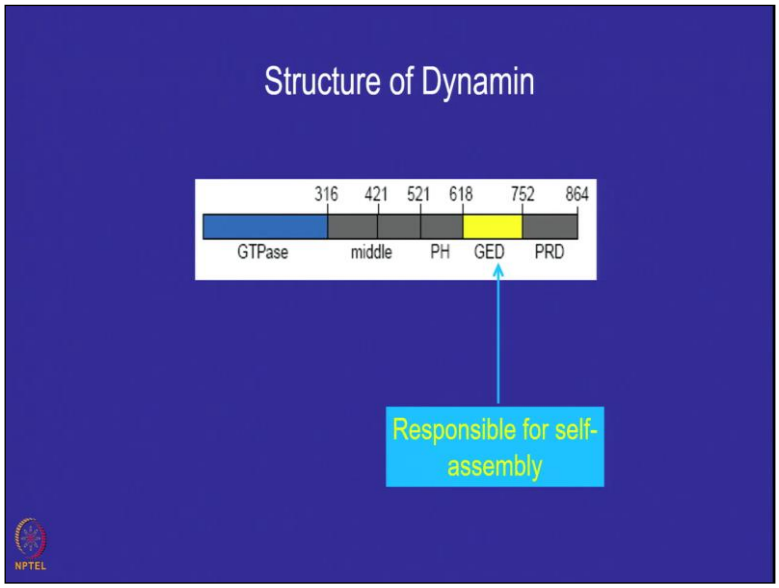
(Refer Slide Time: 22:04)



Now this is a process called as endocytosis. What does that mean? So, if a small molecule has to be internalized the small molecule comes here and then this will be engulfed by a membrane and then it actually forms a kind of an encapsulated system. And then there is a protein which actually binds at the neck of this formation here and this proteins name is dynamin.

And this is the budding clathrin coated vesicle, the dynamin actually forms a big rope kind of a thing and then when you have the GTP hydrolysis this actually gets separated out and then you have this molecule internalized. So, this is the process known as endocytosis. So, the interesting point here is this dynamin molecule. The dynamin molecule how does it work? What is dynamin? What is its role? So, this is therefore a functionally relevant process.

(Refer Slide Time: 23:09)

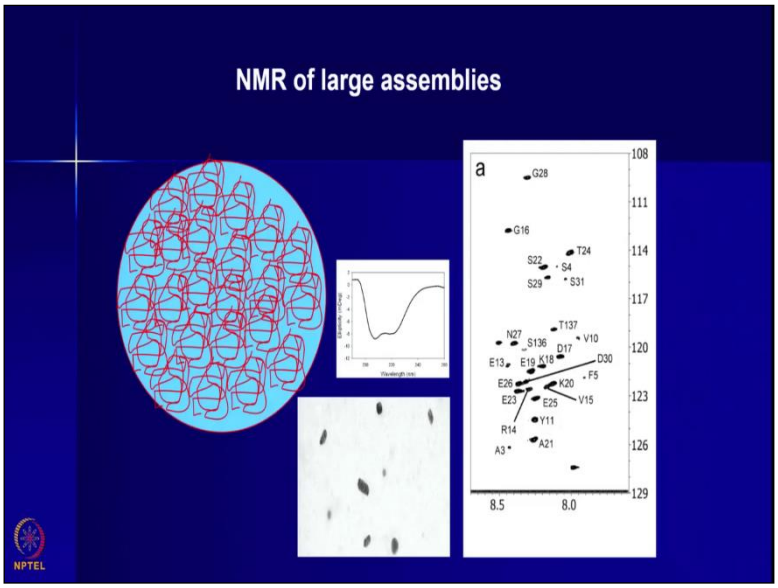


Now what is dynamin? Let us look at what is represent here, the structure of dynamin. So, the dynamin is a protein of 864 amino acid residues, it has this many domains the GTPase domain, then the middle domain, the pH domain, GED domain, and the PRD domain. These are 5 domains which are there. This is 864 amino acid residues and these protein actually self-associates to form huge assemblies, several mega Dalton size assemblies.

That is how it is able to form a big rope. When it forms a big rope, then it can actually bind at the neck, wrap around the neck of the budding vesicle and then of course it squeezes with GDP hydrolysis, as it puts pressure and then the whole thing gets separated out. So, among these domains, it is the GED domain which is responsible for the self-association of the dynamin.

So, therefore can we study this domain? So, let us see this, you can individually express these proteins, produce these proteins in solution and this is what you get.

(Refer Slide Time: 24:18)



So, you take a GED, when you take the GED molecule, express it and you see it has a nice structure. This is the circular dichroism spectrum and you see that it has a very beautiful structure, this is the helical signature. Of course, this will not tell you that what is the size of this molecule, whatever is the size. So, this is expressed protein has this sort of a thing.

Now and this is the electron microscopy image, the transmission electron microscopy, you can see that this is forming particles of this size. So, what that means this means, these are extremely large extremely large assemblies associated state. So, typically so 20 to 30 nanometers of size you will get here. So, which means the molecular weight if you translate that into it will be several mega Daltons 10-15 mega Daltons in size.

With such a large molecular weight, it will be impossible to get any signals. Now we look at the HSQC spectrum of that one here. So, you see how many peaks, you see only about 25 peaks here roughly. Typically, the GED protein itself has about 618 to 752. So, this is about 134 residues. So, 134 residues discount some prolines and things like that. So, you should expect about 125 to 130 peaks.

So, if you see look at that, in contrast to that, you are seeing only 25 peaks. So, what is happening? Of course where are these peaks coming from? Obviously, these peaks are coming from a certain section of the protein which is flexible. And therefore it has a different correlation time, the flexible domain has a certain correlation time and you will see that. So, I will be able to see the signals for these ones, and where are these? If you look at these numbers these ones one can analyze.

So, these are coming from a particular domain of this particular protein and these are in the N-terminal. See look at this residues number here 3 14 21 11 25 15. So, these are mostly in the N-terminal of the protein. So, therefore N-terminal is free, whereas the rest of the protein gets in a kind of an associated state to form a large assembly. So, when it forms a large assembly you will get electron microscopy pictures like this and all those portions which are in the interior of this assembly you will not see the peaks from there, because that forms a huge molecular mass.

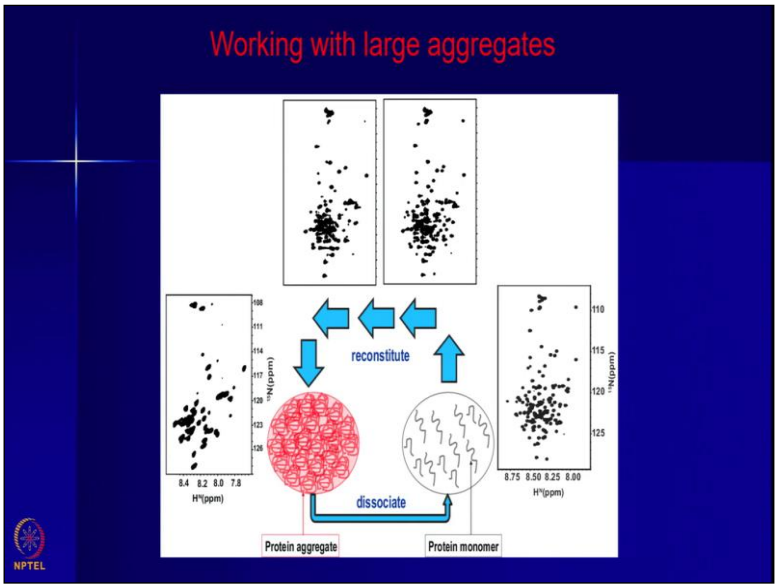
And that will be tumbling very sluggishly and you will have very large line widths and you will not be able to see those signals here. In these ones you can see particular types of residues here, these are mostly you can see few of the glycines, you can see the serines here and the threonines. And then you have largely charged residues either hydroxyl groups here or the charge residues, glutamates, aspartates, arginines and tyrosine here and one or two here valine.

And things like that. But by and large these are charged residues and it is reasonable to expect these charged residues are on the surface of the assembly. And from the different components, because their flexibility they do not show us different peaks, suppose there are some ten thousand molecules in the assembly, all these ten thousand molecules N-terminal segments are looking similar, they are producing one peak only, they are not producing different, different peaks.

So, therefore this ensemble average is what you are getting and these are the residues which are on the surface, these are charged residues therefore they are expected to be on the surface. And the ensemble average of all of those ones gives you one peak each for each of these residues and you are able to see those peaks and you can analyze this, you can assign these individual residues by using the standard procedures which you have discussed earlier for small peptides or small proteins.

Now how do we understand this process? Our objective is to understand this process here, this assembly process.

(Refer Slide Time: 28:26)



So, we have to use a use a particular strategy to identify this. So, I think I will take up in the next class we will stop here you.