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

Lecture: 54
Probing Protein Dynamics by NMR Spectroscopy IV

So, students this week we are discussing protein dynamics as probed by NMR spectroscopy and we discussed how you can measure protein dynamics, what experiments are to be done. So, we already discussed the longitudinal relaxation and transverse relaxation. So, we will start with revising the previous class. So, experiment for measurement of relaxation parameters in protein, we are going for HSQC or HMQC based experiment. As we discussed proton relaxations are complicated therefore, we measure on heteronuclei like 13C or N15.

And, this scheme for all these 2D heteronuclear relaxation measurement is simple like any 2D experiment. We have to start with a preparation stage, creating a desirable coherence, then we introduce a time delay of variable length which can captures the auto or cross correlation rate of this selected coherence. Then, we introduce a T1 period and this T_1 period is for frequency encoding in the indirect dimension. Then finally, we transfer our magnetization to the proton, and we detect on proton while decoupling the heteronuclei. There has to be a delay between the each scan and this delay should be about 5 times of T_1.

So, typically delay in relaxation experiment is longer and we already discussed longitudinal relaxation rate and transverse relaxation rate R_1 and R_2. We will be focusing on the heteronuclear NOE, but before going to that we discussed this transverse relaxation rate is extremely important to understand the protein dynamics – the conformational exchange, because R_2 is ($R_{2,intrinsic} + R_{ex}$). So, it also gives the information of the exchange phenomena that happens in the time regime of millisecond to microsecond. Now, just one example I want to give you before I really move. So, this is a transverse relaxation rate measured for a protein at 0 M urea concentration and we add a mild denaturant and like a 0.9 M and we measure again the dynamics.

So, as expected the termini has a lower or slower rates than the folded protein region. A folded domain has a higher R_2 rates and again termini here have a lower R_2 rate. So, if you

take the difference 0.9 M versus 0 M, what we see here that lots of residue shows higher R_2. They are saying that they are already going into some kind of conformational exchange.

Just by doing this simple experiment of recording R_2 at two different urea concentration telling a lot about this contribution that is coming because of exchange. So, R_2 is a good experiment to measure the exchange contribution as well as how protein folding happens, how destabilization happens. So, many things can be simply investigated and studied using the transverse relaxation rate which is essentially R_2.

Coming to the heteronuclear NOE, this essentially captures the fast amplitude motion that is in picosecond time scale. So, this heteronuclear NOE is determined from the change in the intensity of the NMR signal of a heteronucleus say 13C or N15, when the equilibrium magnetization of a proton in vicinity is perturbed by saturation. So, in the first experiment we saturate the proton signal and then we can measure the how much this perturbation is affecting the heteronuclei X when the equilibrium magnetization of proton is perturbed. So, we measure it either as a transient NOE or a steady state NOE.

So, experiment is something like this. Here 3 channels that we are showing, proton, nitrogen-15 channel, and the gradient channel. You start perturbing the proton signal exclusively. So, here are those saturation pulse. So, we saturated the proton. Now, we are looking at the effect of that saturation on the relaxation rate of the nitrogen.

So, after saturating we are starting with a proton magnetization transferring to nitrogen and then we are encoding here nitrogen, you can see here $T_1/2$, frequency encoding nitrogen while decoupling the proton. Then we transfer back to proton and then do sensitivity enhancement, we record on the proton while decoupling the nitrogen. These are gradient for coherent selection as well as the spurious magnetization separation. So, this is typical pulse sequence. We start with a saturation on proton, transferring that magnetization on the nitrogen, and then detecting back on proton.

So, like a first 90° pulse that we are applying on proton, this is combined with a gradient. If you look at here, it is combined with a gradient and that ensures that N15 magnetization is the only magnetization in the experiment. So, actually we have saturated, started with a 90° on proton here and then we have a gradient here. So, that ensures that magnetization

only goes to nitrogen and after the first 90° pulse which is this, the chemical shift of the heteronuclei, N15 in this case, is evolving and that is evolving during the T_1 period.

And then as we discussed we transfer the magnetization back to the proton, which can go for detection. So, here we decouple so that it does not evolve under coupling during this period. So, there is a 180° pulse here and finally, this orthogonal magnetization component is generated during the T_1 period which is refocused in the PEP sequence and then we simultaneously detect by inverting the phase φ4 here and do the coherence selection. So, then what typically we do in these experiments we do two experiments.

So, this is the experiment. Quadrature detection we do in F1 dimension by shifting the phase of this pulse, first pulse and the receiver together so that we detect in a state TPPI manner. So, this is typical way of doing experiment saturating it, starting with a proton transferring to nitrogen encoding the T_1 frequency transferring back for PEP on the proton, detecting proton and then here is for TPPI detection. You phase shift this together with this. So, that we select the quadrature detection for this and the coherent selection is done by cycling this φ. So, to measure the NOE, we do two experiment a pair of experiment in one case where we are saturating that is called NOE and in one case where we are not saturating that is called NONOE and then we take the intensity of this NOE and NONOE, saturated and not saturated.

So, here if you look at the intensity is given by

$$I_{sat} = \langle X_z \rangle_{eq} + \frac{\sigma_{XH}}{R_1} \langle H_z \rangle_{eq} = I_{unsat} \left(1 + \frac{\sigma_{XH}}{R_1} \frac{\gamma_H}{\gamma_X} \right)$$

I_{sat} and I_{unsat} are the intensities. σ_{XH} is the rate constant of the cross correlation and as we discussed γ_H and γ_X are gyromagnetic ratio. The NOE is given by I_{sat}/I_{unsat}, where we saturate for say 3 second and then we do not saturate.

So, that is how you do. The error in the measurement of NOE is given something like this

$$\frac{\sigma_{NOE}}{NOE} = \sqrt{\frac{\sigma_{I_{sat}}^2}{I_{sat}^2} + \frac{\sigma_{I_{unsat}}^2}{I_{unsat}^2}}$$

So, this is the error in the measurement of NOE.

So, typically $\sigma_{I_{sat}}$ and $\sigma_{I_{unsat}}$ represents the standard deviation. So, that can be calculated separately and then we can find it out how much is the NOE value. Now, typically for N15 nuclei the NOE varies between 1 to –4 and this is because of the negative gyromagnetic ratio of nitrogen. As you know the gyromagnetic ratio of nitrogen is negative whereas, for carbon 13 it varies from 1 to 5. So, 1 means NOE, 1 means this is coming from the rigid portion of the protein, –4 means it is extremely flexible.

So, as we discussed NOE reports about the fast amplitude motion, NOE value closer to 1 means this is coming from the rigid portion and –4 is for flexible portion. So, how do we set up the experiment? Typically we have a recycle delay of about 5 second right. So, because we are starting with N15 allows for longitudinal relaxation. So, all the time what we want that before we start our next scan magnetization should go back to equilibrium. We started from somewhere, but it should go to equilibrium.

So, that is how we have a longer recycle delay and usually NOE and NONOE recorded in an interleaved manner to reduce the artifact, typically you record 128 complex t1 point. FID can be processed with 0 filling of 512 in direct dimension and about k in the direct dimension. Then you do processing by just 90 degree shifted square sign bell in both dimension and then you calculate basically I ratio of the peak I_{sat}/I_{unsat} for cross peak intensity. So, here we are getting 2 for like a we can record in an interleaved manner, then do a split NOE, it gives me two spectrum here is say I_{unsat}.

And then you have another spectrum where intensity are little weaker, this is I_{sat}. So, you measure the intensity of I_{sat} say x amount and divide that by corresponding intensity by y. So, this is our NOE value and then error in NOE can be calculated using the formula. So, you can have residue specific NOE value and if you do for this protein that I was discussing with you SUMO-1. Now, what we see this is these are the NOE value.

So, this is actual NOE value and these are errors basically you can see here these are error bars. So, in this protein let us look at the structure like our residue wise the NOE value. If you look at starting from 22, 23 and going up to about 90, we see that all NOE are positive with little variations. And this says that most of the protein is quite rigid in this domain right starting from here will be 22 and here will be about 93, 94 right. So, all these shows positive NOE values are quite close to 1.

And when we look at the termini here and here the value is negative, here all the way up to –4. As we discussed this is N15-NOE, so the maximum negative value can go up to –4 and here also the flexible protein shows the negative NOE. So, all the flexible portion in this protein is showing negative NOE, all the rigid portion is showing the positive NOE that is how typically the heteronuclear NOE value for N15 nuclei one can obtain it. So, experiment is very simple you record an NOE where relaxation delay is 5 second in one case you saturate about 3 second in another case you do not saturate, record 2 HSQC spectrum, measure the intensity of each peak. and then you calculate the heteronuclear NOE, dividing I_{sat}/I_{unsat} that is a residue wise your NOE value. Then you can calculate the error and then get information of the fast amplitude motion in a residue specific manner good.

So, now we summarize R_1, R_2 and NOE. Let us look at together what happens when we measure for the same protein. This is our N-terminus, this is our C-terminus, this protein is called SUMO1, small ubiquitin related modifier 1. So, here are the secondary structure element. If you look at here, now let us look at all these three R_1, R_2 and NOE.

For the flexible portion, we have a high R_1 value, low R_2 value and negative NOE value. For all structured portion you look at the beta sheet position or alpha helix position, we have a low R_1 value, high R_2 value and positive NOE. Again, for the loop you have flexibility, so high R_1 value, relatively like some of those are low R_2 value, but because of exchange contribution some of these are still high and positive NOE value. So, by comparing it you know where is the fast amplitude motion.

This also reports about nanosecond time scale motion, this also nanosecond to picosecond time scale motion, whereas R_2 reports picosecond to nanosecond and microsecond to millisecond time scale motion. So, that is what we can get information and just to for your ready reference if you denature this protein R_2 and R_1 in a sequence wise manner do not show much variation they are more or less flat. So, for a structured protein we know where is the flexibility and that are very well captured in the experiments R_1, R_2 and NOE, that is what is being shown here. One important thing, the R_2 has one important contribution that we had call it R_{ex}. Now, that is exchange contribution. So, here is a protein where you see R_1 value, R_2 value, and NOE.

In this case, you do not see much variation only you see that there are some loops here, but not like whatever SUMO-1 we had. Now, in this case you see that R_2 value for some region you have relatively high R_2 value. So, this high R_2 value and here if you look at NOE it is about 0.6 so that means, relatively rigid not so flexible, few of them NOE are less, but protein seems to be quite rigid, the R_1 value also do not show much variation, but in R_2 there are some portion which shows high R_2. Now, one can measure the $R_{ex,}$ which is exchange contribution and you can see the portion where high R_2 values are there you see high R_{ex}.

That means the value in R_2 are coming from that exchange contribution that we are talking. In a simplistic term, you do not need to really do the R_{ex} experiment separately, if you have measured the R_1 and R_2, you do R_2/R_1 and those portion will be quite exposed. So, you know that this portion is coming from this high value is coming because of R_{ex} measurement. You can even do the analysis called R_2*R_1 and that also reports about this R_{ex} contribution.

Another parameter, which we are going to talk about, is S^2, the order parameter. Now, order parameter also tells about the rigidity in the molecule that I am going to talk after a few slides the how it is calculated, but this tells about the rigidity of the molecule. If you calculate the order parameter, you can also see it is quite flat. So, order parameter about 1 is showing rigidity and coming towards 0 shows flexibility. So, here you can see on an average the order parameter is 0.8 which says that protein is quite rigid.

So, if you do all these experiments R_1, R_2, NOE, we learn a lot about the protein dynamics and possible alternate states that it might be capturing. So, we measure R_1, R_2, NOE, now can we go and do more analysis. So, you know that if you remember from the initial lectures we had talked about the how $1/T_1$ which is R_1 is captured in different spectral density function like $J_{\omega H}$ the spectral density at proton frequency, $J_{\omega N}$ nitrogen frequency, $J_{(\omega H+\omega N)}$ sum of this frequency, $J_{(\omega H-\omega N)}$ difference of all this frequency.

So, similarly T_2 can be explained in terms of $J_{\omega 0}$, $J_{(\omega H-\omega N)}$ the difference in the frequency and all those. NOE can again be measured or it can be represented in terms of the sum of the frequency and difference in the frequency. So, essentially if you measure R_1, R_2, NOE we can get the spectral density function of a protein that we are going to discuss in the coming slides. So, essentially you need three parameters this R_1 measurement, R_2

measurement and NOE. So, if we have these then how to interpret the relaxation data that we can now look at. So, relaxation data for a folded protein can be analyzed using something called modal-free approach of Lipari-Szabo.

Lipari-Szabo gave a model free approach where we do not consider a model. So, this is based on assumption that separate internal or globular motion. Like a globular motion here is a protein how it is tumbling and internal motion how the loops residue in the protein are tumbling. So, it is based on an assumption that these two can be separated and the dynamics can be described in term of the overall tumbling.

So, here is my molecule how it is overall tumbling that is t_m, how internal correlation time for say particular this portion is happening that is t_e and the generalized order parameter that describes the amplitude of the of the internal motion. So, using this experimental parameter and model one can describe the motion that are present in the protein. The only thing that there are some assumptions in doing the Lipari-Szabo calculation. So, what it says the when overall tumbling of a protein, so say protein is quite rigid molecule and you are considering that as a globular. So, if you are considering globular the overall tumbling in a protein that happens can be defined by single correlation time.

So, here is my protein and it is stumbling. So, you know that there is a only one τ_c, the correlation time and internal motion can take on a time scale which is happening much faster, what it means say the protein is tumbling at a nanosecond time scale, but internal motion is much faster than that. So, that is one of the assumptions that basically this analysis of model-free takes. So, if this is the case then the correlation of two motion internal and overall can be separated and total correlation time can be given as that. So, correlation time at time t will be $C_0(t)^* C_I(t)$. So, the correlation function of overall motion assuming it is a isotropic distribution.

So, everything is moving in a isotropic manner and that is how you give this correlation time. So one can have a C0t which will be given by

$$C_0(t) = \frac{1}{5} exp\left(\frac{-t}{\tau_m}\right)$$

and if you do Fourier transform of this you can get this spectral density function which is

$$J(\omega) = \frac{2}{5}\left(\frac{\tau_c}{1 + \omega^2\tau_c^2}\right)$$

So, τ_m being the rotational correlation time for overall tumbling of a protein that is how you essentially measure in this case considering the isotropic motion. Now, the correlation function for internal motion, which was this component. This is external motion like overall tumbling and for internal motion τ_e you can get as

$$C_I(t) = [S^2 + (1 - S^2)\exp\left(\frac{-t}{\tau_e}\right)]$$

So, for a completely restricted motion the S^2 value goes to 1, that shows the protein is quite rigid and for completely disorder region the S^2 is 0. So, now coming back to this whatever we discussed few slides back here if you look at the S^2 for this protein is more or less 0.8.

What it means? This protein seems to be quite rigid. If the protein has a flexibility, you see the order parameter will be 0.2, 0.3 or something.

Now, for doing this analysis Lipari-Szabo has considered that this molecule is quite globular and it is having isotropic motion and that is how you get these values of S^2 1 for rigid and S^2 equal to 0 for flexible. So model-free assumptions, it is assuming the molecular overall motion is an isotropic. That's a bigger assumption. And the method for characterization over all motion, overall rotation is primarily implying that motion of most of the protein are very fast. If you remember we are saying that here is my molecule, it is considering overall motion to be τ_m to be isotropic and internal motion is very fast like a less than 100 picosecond. Now, usually the parameter governing the relaxation for N15 nuclei are kept fixed in this predetermined value in modal free analysis. The modal free approach assume that intramolecular motion are independent of overall tumbling. Why? Because this is happening very fast compared to the overall tumbling and the conventional model free approach implicitly assume that protein does not aggregate.

So, at any concentration, protein is well behaved and typically NMR relaxation there is no aggregation during the experiment that we are doing. So all these assumptions are there in the modal free analysis and that's how one can do it. So, now let us see how model free analysis is done. Very briefly if you remember,

$$J(\omega) = \frac{2}{5}\left(\frac{\tau_c}{1 + \omega^2\tau_c^2}\right)$$

J(ω) means how much power it has.

So, it captures the power with the frequency. So, for a fast tumbling and a slow tumbling if the τ_c is of 1 nanosecond, it is a fast tumbling. A bigger protein has less J(ω) and slow tumbling, a smaller protein has a more and it rapidly decays it. So, this is the J(ω) that we had seen in the earlier slide. So, just to remind, spectral density function J(ω) is the Fourier transform of correlation function. And just as rapidly relaxing domain signal gives a broad signal like if it is rapidly relaxing it is a broad signal, if it is slowly relaxing it has a sharp signal and that basically is given by the spectral density function. So, this makes sense. Molecule that tumbles very rapidly samples a wide range of frequency and molecules that some tumble slowly have a very long correlation times and only samples free frequency.

So, with this assumptions we delve deeper and do the modal free analysis to understand what parameters can be extracted. Model free analysis is for isotropic globular protein. It considers that the isotropic motion is happening in the protein and the internal motion is very fast. With this I will stop here. We can take the analysis of model free in the next class and if your protein is not behaving isotropically there is lots of anisotropy if it is not a globular protein there are some portion which is very rigid and some portion which is flexible what needs to be done. These two questions we are going to take it in the next class and we will discuss model free analysis as well as the reduced spectral density analysis for protein to interpret the data of relaxation.

With this I will stop it here for today. Thank you very much.