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

Lecture: 55
Probing Protein Dynamics by NMR Spectroscopy V

So, students welcome to today's lecture. We are discussing protein dynamics probed by NMR spectroscopy and here we looked at how we can use the R_1, R_2, NOE to understand the protein dynamics. So, today I am going to explain more about how we can probably use these techniques for the interpretation of relaxation data. So, as you know relaxation equations are dependent on the calculation of a spectral density that we had discussed in the first class which is $J(\omega)$.

So, T_1, T_2, NOE are encoded in this spectral density function. So, like for T_1 or R_1, it depends upon frequency of $J(\omega_H)$ spectral density function of the proton frequency, $J(\omega_N)$ nitrogen frequency and so on and so forth. R_2 can be written in terms of the spectral density function like spectral density at frequency 0, spectral density at hydrogen frequency $J(\omega_H)$, nitrogen frequency $J(\omega_N)$ and so on and so forth.

R_2 also has a component R_{ex} here. The NOE is similarly can be written in terms of the spectral density function. So, essentially what I mean to say that all the relaxation equations are dependent on the calculation of spectral density function and or vice versa using these also we can calculate the spectral density function. Now, how to interpret this data? So, suppose we are dealing with a folded protein. So, typically for a folded protein these relaxation data are interpreted in terms of model free approach of Lipari and Szabo. Now, this model free approach talks about the separatability of internal and global motion, this is typically for a folded protein.

So, folded protein, consider it as a kind of a sphere. So, what is motion here this is a local motion and a global motion. Internal motion and global motion, they can be separated based on assumption and this dynamics is described in terms of the overall rotational correlation time like τ_m overall time, the internal time is τ_e like internal correlation time. Then another term called order parameter, which describes the amplitude of this internal motion, what is the amplitude of this internal motion. So, Lipari-Szabo basically gives you

three parameter to describe the dynamics, the overall tumbling time, the internal correlation time and the order parameter which describes the amplitude of internal motion.

Now, overall tumbling time of protein can be described by single correlation time and internal motion takes place much faster. So, it is overall tumbling happening slowly and internal motion happening very faster. So, the correlation for these two motions, the global motion and local motion, can be separated and the total correlation function can be given by this function.

$$C(t) = C_0(t) * C_I(t)$$

So, the correlation function for overall motion assuming that isotropic diffusion is given and these are the consideration for overall tumbling time and if you do Fourier transform of this we can get this spectral density function J(ω).

So now τ_m being a rotational correlation time for overall tumbling of a protein. The only consideration that internal motion is happening too fast here. Now, the internal motion can be written as

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

So, S is called generalized order parameter and τ_e is effective correlation time. For a completely restricted motion, the generalized order parameter $S^2 = 1$ and if it is completely disorder or flexible $S^2 = 0$.

So, typically for any realistic consideration the S^2 varies between 0 to 1. Basically, in Lipari-Szabo we are considering the Brownian rotational diffusion of any molecule in solution and this diffusion can happen in terms of ellipsoid. So, a protein is like ellipsoid or a spheroid or a sphere. How it is diffusing that basically comes from this Brownian rotation diffusion. So, these consideration are there in Lipari-Szabo interpretation of the dynamics data and this is called model free assumptions. So, in modal free assumptions essentially we are assuming that overall rotation of this protein molecule is isotropic, there is no anisotropy here, there is no orientation dependence here.

So, it is isotropic motion and this method for characterization the overall rotation is a priori implies that internal motion for most of the protein is too fast compared to the global

motion. Also overall rotation is an isotropic. So, usually we are measuring the parameters that govern the relaxation of N15 nuclei in this protein. They are kept fixed at some predetermined value in this model free analysis. And it assumes that intramolecular motion are independent or overall rotational motion is too fast and it is independent.

So, conventional model free approach assumes that protein does not aggregate. So, protein stays in a folded state without self-association happening at the concentration where we are measuring relaxation So, with this assumption this model free analysis can be done. So, what are assumptions? Isotropic motion, internal motion is too fast than the globular motion, and the actually intermolecular motion are independent of overall molecular rotation. So just to remind you again, spectral density function J(ω) can be written in terms of τ_c. Just to put things in perspective, if you plot J(ω) versus frequency and τ_c are denoted by this, so that if τ_c is 100 nanosecond it is tumbling slowly it is a spectral density function rapidly decay, if it is 1 ns that is fast, it starts with some value and slowly decay. So, what it implies a spectral density function J(ω) which is a Fourier transform of correlation function just says that rapidly relaxing domain signal give a broad line.

So, if something is rapidly relaxing that means a bigger protein it gives a broad line if slowly relaxing it gives a sharp line. So, this makes sense, right? A molecule that tumbles very rapidly can sample a large range of frequency ranging from say few megahertz to few like 100 megahertz and the molecule that some tumbles very slowly like a bigger molecule that can sample only a narrow range of frequency. So, very long correlation time like 100 nanosecond or so and samples fewer frequency that is spectral density function J(ω) says. So, taking that consideration the original Lipari-Szabo,

$$J(\omega) = \frac{S^2\tau_R}{1 + (\omega\tau_R)^2} + \frac{(1 - S^2)\tau'_e}{1 + (\omega\tau'_e)^2}$$

If there are two motion on two different time scale, you can just separate those two and you can add some anisotropic consideration as well. So, essentially you have to solve these functions get the generalized order parameter, overall rotational time and internal rotational time. So, essentially in the model free analysis data, the parameter that we get it from R_2/R_1 ratio is τ_e which is less than 100 picosecond and τ_R overall rotational time is more than 1 nanosecond. So, that is a wide range and these are adjustable parameter that can be obtained by fixing the relaxation data. You can calculate the hydrodynamic radii in an iterative manner.

So, this is kind of a work flow for selecting a suitable correlation function in a iterative manner. So, start with S^2 or S^2, R_{ex}; S^2, τ_e internal motion or S^2, τ_e, R_{ex}. So, all those motions with slow or fast motions can be fitted. So, if no then reject this model take another model and keep iterating to find the generalized order parameter. Another consideration is that we have the isotropic motion in a globular protein. But the proteins are not always globular.

There is a fair bit of anisotropy in a protein. So if there is anisotropy, which is commonly observed in disordered protein, this Lipari-Szabo model-free approach gets complicated. So in those cases where there is anisotropy, Peng and Wagner suggested the calculation of power spectral density function using six proton and nitrogen relaxation rate. So what are those six? We are going to come in a moment. Now this is called a spectral density mapping or spectral density function.

It does not depend upon any model. So it's just free of model and, no dependent on any form of time dependence or autocorrelation function, nor does it require any form of rotational diffuser tensor of the molecule, it is just does not require any of these. So, only thing you need a spectral density mapping at different frequencies. So, for N-H bond that are directly sampled at several relevant frequency or say 0 frequency or ω_N or $\omega_H - \omega_N$ like if you add the two frequency and subtract two frequency.

So, these are six J(0), J(ω_N) the frequency of nitrogen, J(ω_H) frequency of proton, J($\omega_H - \omega_N$) the difference of proton and nitrogen frequency and J($\omega_H + \omega_N$) sum of proton and nitrogen frequency. So, these are the spectral density functions that are there which needs to be considered. So, in this spectral density function the main goal is to understand the spectral density and several different methods has been derived to model the protein as different spectral density function with different parameters are fit to experimental data that is what proposed by Peng and Wagner. So, this is called spectral density mapping that we are discussing.

So, three mathematical relation between the relaxation parameter T_1 and T_2 and J(ω) that we had seen in previous slide and can be used for calculating this spectral density function. So, they come up with a much simpler method for interpreting this relaxation data and that was provided in terms of reduced spectral density function. So, what is reduced, that we

are going to just see it. So, if you consider this $J(\omega_H) \approx J(\omega_H + \omega_N) \approx J(\omega_H - \omega_N)$. Because spectral density function say we are doing at 600 megahertz.

So, proton spectral frequency is at 600 and nitrogen is at 60. So, 600 ± 60 is in the same range. So, one can roughly say that $J(\omega_H) \approx J(\omega_H + \omega_N) \approx J(\omega_H - \omega_N)$. So, this one can club into one. So three frequencies will be J(0), $J(\omega_H)$, and $J(\omega_N)$. So, instead of five we can just do with three spectral density function and as we have seen in the previous slide, they can be expressed in terms of three parameters that are R_1, R_2 and NOE.

So, these three measurable parameter R_1, R_2 and NOE, we can measure in a residue specific manner and using these three we can calculate the spectral density of J(0), $J(\omega_H)$, and $J(\omega_N)$ using this reduced spectral density function. So, these rates we can measure for proton and nitrogen at any frequency. Like say we are doing on 500 megahertz, which means 11.76 Tesla, as our allowed measurement happening at 0 megahertz, 50 megahertz or 500 megahertz. And these two like $\omega_H - \omega_N$ means 450 or $\omega_H + \omega_N$ means 550.

At 17.6 Tesla, 750 megahertz, it means 0 frequency, 750 frequency, 750 megahertz, 675 and 825, but these three are in the same range. So, we can do with only three of the spectral density functions. So, now we have come up to spectral density function. Using these three parameters that we have measured R_1, R_2 and NOE, our J(0) can be calculated or similarly $J(\omega_H)$ or $J(\omega_N)$ can be calculated.

So, as we know the R_1, R_2 and NOE. So, what is R_{NOE}? It can be calculated from our NOE experiment that we did.

$$R_{NOE} = [(\{1H\} - 15N)NOE - 1]R_1 \frac{\gamma_N}{\gamma_H}$$

Now, for calculating this we need two other constants called c^2 and d^2 and the value of these are have been calculated. So, the typical value of c^2 and d^2 at 600 MHz is $1.25*10^9$ $(rad/sec)^2$ and $1.35*10^9$ $(rad/sec)^2$. Relatively it can change little bit at 750 or 500 megahertz. Now, using these all parameter, if you have measured the R_1, R_2, and NOE, you can put all these value and can calculate the J(0), $J(\omega_H)$ and $J(\omega_N)$.

So, NOE which is fast amplitude motion is reporting more towards $J(\omega_H)$. Similarly, the J0 is contributed by R_1, R_2 and NOE. And here $J(\omega_N)$ which is at intermediate frequency is contributed by R_1 and NOE. So, what does this implies? So, $J(\omega_H)$ high amplitude motion like spectral density function at higher frequency.

Let us go here. So we are saying $J(\omega_H)$ at higher frequency. It captures fast motion. The slower motion is captured by J(0). So, $J(\omega_H)$ is largely determined by the heteronuclear NOE, that is a fast amplitude motion and most sensitive to high frequency motion of the protein backbone like the motion happening at picosecond time scale that will be given by this parameter $J(\omega_H)$.

Similarly, $J(\omega_N)$ contributes to R_1 and NOE. So, $J(\omega_N)$ is determined by R_1, whereas J(0) is contributed from R_1, R_2 as well as NOE. So, J(0) is sensitive to both nanosecond time scale motion as well as exchange phenomena that is happening at slower time scale microsecond to millisecond. So, if we are measuring spectral density function at zero frequency, it is sensitive to slower time scale motion, nanosecond as well as microsecond to millisecond.

The $J(\omega_N)$ is given by mostly nanosecond time scale motion, because this is contributed by R_1 and some also contributes from R_2 and NOE. So, nanosecond to picosecond time scale motion you can see here. This is nanosecond and microsecond to millisecond time scale motion this J(0) is giving and this $J(\omega_H)$ giving picosecond time scale motion. So, if you calculate this spectral density. From the R_1, R_2, and NOE data, we know exactly how the spectral density is mapped across the protein sequence.

And that gives the amplitude of the motion happening across the protein sequence. So you see this, the protein that we were dealing, SUMO1, which has an N-terminus tail, and at folded region and some loops again. So, here we had calculated in previous slide that if you remember the R_1 was like this here for flexible tail and this was for kind of a globular domain α-helix β-sheet over there and same little bit of the tails were here. That is R_1, the R_2 was low. NOE was negative here.

So, if we calculate $J(\omega_H)$ using this NOE value you can see here for the flexible portion the $J(\omega_H)$ is very high and for the all ordered, it is relatively low. So, now picosecond time scale motion in this protein, for the flexible tail is very high picosecond time scale motion

is higher here and lower here. Now, if we go to $J(\omega_N)$ which is dominated by R_1, you can see there is a one to one correlation between these two like a very large correlation. Higher $J(\omega_N)$ because this is capturing at nanosecond time scale motion and all these are lower, the loops are again higher. Now, J(0) which is contributed by R_1, R_2, and NOE, everything has an impact here.

So, you can see this is quite mirroring with our R_2 because it is capturing lot more from the slower time scale motion. So, spectral density mapping telling it how the protein is sampling various amplitude motion across the backbone which are calculated from three experimentally determined parameter R_1, R_2 and NOE. Fantastic. So we understood how the spectral density function can be mapped.

So what next? We find it out from the R_1, R_2, NOE. Now the R_2 you know that it has also R_{ex}. So if you map this it says that along this line there seems to be exchange phenomena happening in this region and that is why you see lots of variation. So the spectral density function is also capturing this because the slower time scale motion, microsecond to millisecond time scale motion. The exchange is happening and that essentially is nicely captured if you measure the spectral density function at a 0 frequency, at J(0). So, now using this, there is a linear relation that exists between J(0) and $J(\omega_H)$ and $J(\omega_N)$.

You can write this equation,

$$J_i(\omega_{N,H}) = \alpha_{N,H} J_i(0) + \beta_{N,H}$$

So, this is linear relation. Many time, this linearity is affected by contribution of this exchange happening, which affects the J(0) value. But if you rearrange this equation, one can find it out the τ. You can find τ and ω is a Larmor frequency for N and H. So, if you solve this equation linearly, if you plot (ω_H) or $J(\omega_N)$ versus J(0), you can find the α and β. Now, you put this α and β here and solve this equation, you can find the τ and that is correlation time.

So, now we are finding correlation time in a residue specific manner. So protein is tumbling and you can find the correlation time at different time scales. So, same data we are going to take it here, $J(\omega_H)$ or $J(\omega_N)$ and J(0) and plot as linear function. So, here is our $J(\omega_H)$, $J(\omega_N)$, J(0), this is my protein, this is N-terminus, this is C-terminus. Now, can we use this

equation to fit a linear equation and find out what is happening? So, we calculated for each of these amino acid $J(\omega_N)$ in ns/rad versus J(0) ns/radand we plotted it.

If you plot it, it is not very linear. There are some residue, which deviates from linearity. So, you need to fit two curve, one curve for these black residue, one curve for the open residue. If you plot these two, now they are looking more linear. Similarly, if you plot $J(\omega_N)$ versus J(0), again very nicely one portion here black and one portion for open circle residue.

Now, this telling us very interesting phenomena, because we started spectral density function assuming that anisotropy exist. So if you go in this protein structure, you can see that the amino acid that are there will have a different kind of motion than a globular protein. This you can consider as a sphere, spheroid, or all those. But these guys are tumbling in solution and because of that free tumbling, there is anisotropy exists and that anisotropy is being captured in the spectral density function.

So, if you plot this linear equation and deduce the parameter that is coming. So, if you fit this equation J(0) versus $J(\omega_N)$, fitting this section separately and fitting this section separately. We can calculate from the previous equation that we had α, β and τ. We can calculate this three parameter.

Now, just look at this interestingly important thing. For a low amplitude motion, what we are finding, two motions are in nanosecond, but there is a motion in microsecond. This is a slower motion happening for a globular protein. Some motions are in nanosecond, and some motion is in microsecond. However, for this high amplitude, you are getting a nanosecond time scale motion of 5 nanoseconds, typically for a protein. Okay, if we plot this J(0) versus $J(\omega_H)$, here again we are getting some τ_c of nanosecond, but for a high amplitude motion for the flexible portion, we are getting some of the motion as fast as picosecond.

Now that is a high amplitude motion. We can see for the flexible portion some picosecond motion. Now you see here is interesting phenomena happening. Here, this is a flexible tail, and it has a correlation time of few nanosecond? Yes, of course. Typical globular protein have 5 to 6 nanosecond, but one of the parameter that came here is 21 nanosecond, quite

high. Because it's a long open rope is tumbling in solution, that of course will have little longer correlation time. So now doing this spectral density mapping, solving these equations, finding it out what are the α, β and τ, we are deducing now all sorts of correlation time that can be seen in the protein.

This gives a wide range of the tumbling time and that is the power of NMR dynamics. If you use any other technique like a dynamic light scattering, it is a complementary technique for measuring the correlation time of a molecule. You mostly see one or two correlation time depending upon how it decays and gives you one correlation time of a protein whether it is a globular or elongated. But, NMR precisely gives you all the correlation time that is possible measuring the dynamics in a residue specific manner. So, measure this basic 3 experiment R_1, R_2, and NOE, from there we deduce the spectral density function J(0) and $J(\omega_H)$ and $J(\omega_N)$, fitted those equation and solve this quadratic equation to find it out what kind of correlation time we are getting.

Now we can reduce it, that why some of the most some of the correlation time are longer, because it is open chain and that are expected to show higher correlation. So, that is all about time and motion in protein. That is what we in this week wanted to give an impression to you that NMR is a beautiful tool to understand the protein dynamics. Because dynamics dictate the function, and we can measure the dynamics in a residue specific manner in more elegant way than possible from any other complementary technique. So, that is a strong point of NMR.

With this I am going to close the lecture for this week. Next week I would like to see you in the discussion where we are going to discuss more about protein-protein interactions and how we can use NMR to understand protein-protein interactions. So, looking forward to have you in the next class. Thank you very much.