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

Lecture: 53
Probing Protein Dynamics by NMR Spectroscopy III

So, this week we are discussing about protein dynamics, which can be probed by NMR spectroscopy. So, in last lectures I discussed why dynamics is important, what all methods are there to measure dynamics and why NMR based dynamics. So, we will continue our discussion for the NMR based dynamics, how we can do the experiments. So, typically we discuss that we measure the spin dynamics, spin relaxation rate by carrying out two dimensional heteronuclear experiments such as HSQC (heteronuclear single quantum coherence) or HMQC (heteronuclear multiple quantum coherence) experiment, and then we can measure the 13C or N15 relaxation. Proton relaxation are rather complicated.

So, they are not directly measured in protein NMR. So, typical scheme for 2D heteronuclear relaxation measurement is something like this. We have to start with a preparation phase, where we start preparing a desirable coherence like we start with proton transfer to N15 and there we introduce the variable delay of various length which actually captures the auto or cross relaxation rate. After this, we introduce a T_1 period, which indirectly encodes the frequency such as N15 or 13C.

So, after T_1 period, we transfer the magnetization to proton because we are going to detect on proton. So finally, transferring the magnetization back to proton and acquiring on proton while decoupling the heteronuclei like a 13 C or N15 and then after acquisition. There is a delay period (D1) and this D1 period ensures that our magnetization are again back to the z-direction before we start the next scan. So, then we go back and start the next scan. So, typically this is the experiment that is done: preparation, T_1 delay, here we are encoding that like a in case of say T_1 experiment we will be encoding here longitudinal relaxation rate. So, then followed by a T_1 period, indirectly frequency encoding, transferring the magnetization on proton and finally acquiring it.

That is a typical acquisition scheme that we have in the 2D. So, if you look at the pulse sequence of a longitudinal relaxation or T_1 relaxation or longitudinal relaxation. This is the

kind of pulse sequence that we use, This is water flip back sensitivity enhanced T_1-HSQC experiment. So, water flip back is required for suppressing water.

As you know in protein, the concentration of protein is typical of few hundred micromolar or maximum one or two millimolar, but the water concentration is very high. So, we need to kill the water. Water is generally 55 M. So, we have to kill the water and that is how all the protein experiments include this water flip back or water suppression schemes. Here is also water flip experiment.

So if you look at typically we are starting here with an INEPT transfer, we are coming on the nitrogen. Then here we are encoding the T_1 that is a T_1 variable and here is frequency encoding and after that we are transferring back to proton and then we are doing the sensitivity enhancement. Finally, we are acquiring on proton while decoupling nitrogen. These are gradient pulse for coherence selection as well as killing the undesired magnetization. So, this is the T_1 period.

This is being introduced to measure the longitudinal relaxation rate in T_1 pulse sequence. So, what we have typically 180 degree inversion pulse, that inverse the magnetization from Sz direction to –Sz direction. T period is given to allow magnetization to come back to the z direction. So, we first we invert it and then it does the precessional motion to come back to z direction.

So, this T-period is basically varied and that is what we measures. So, you start with a – z direction and then with a different T period we are measuring how the intensity is changing and coming back to equilibrium. So, that is what you measure

$$M_z(t) = M_{z,eq}(1 - 2e^{-t/T_1})$$

So, this is the rate and this t which we are giving and $1/T_1$ is the rate.

So, a 90° pulse creates the observable transverse magnetization for detection that is why we apply a 90°, and then for T_1 measurement, this relaxation block is inserted in the sensitivity enhanced HSQC experiment. So that is how we do the experiment. So let me summarize again. We start with proton magnetization, transfer to nitrogen,

then we invert it, and then allow a T_1 period to spin to relax and then we apply a 90° pulse which creates the observable transverse magnetization and transfer back to proton for detection. So, in terms of product operator if you look at we start with a z direction H_z which with a 90° pulse converted to $-H_y$. then we are evolving under coupling and finally, transferring back to the nitrogen Ny. Then we introduce the T_1 time period, here and finally it goes to some ξN_z.

Now, this factor ξ is a time dependent magnetization signal amplitude. So, you can write

$$\xi = 1 - 2e^{-T/T_1} = 1 - 2e^{TR_1}$$

So, this gives us how much the magnetization is changing and that is how we measure it. So, this is the conceptual framework for doing the T_1 experiment.

Let us go to how we set up the experiment and how we process the data in this case. So, we start with a usual stuff. We take a protein sample whatever concentration you can afford, take it about 500 µl at 10% of D_2O, this is for locking, then you do the sample height adjustment. So, in a typical NMR tube you have to have something around 2.1 cm or 20 or 21 mm height adjustment.

Then you tune the magnet, shim the magnet, set the O1 frequency, O1 setting. That is the usual stuff in any protein NMR experiment. Then we determine the 90° pulse, there we calibrated for transmitter and decoupler and we choose the spectral window, how much we want to keep. For protein, generally you want to 0 to 12 ppm and here say 100 to 130 ppm should be fine for recording an N15 HSQC experiment. Now important point because we are doing relaxation experiment. So, typically we have to set a D1 which will be 5 times of T_1.

So, typically setup of recycle delay should be 1.5 to 2 times T_1 and generally we have to signal average good, right. So, number of scans should be 4 or 8. In the experiment different T point, we are recording right, where the spin is relaxing. So, we need typically of 8 to 12 T delay which you can range from 5 millisecond to 1.5 second for N15.

2D experiment we do with 5 millisecond, then 50 millisecond, 100 millisecond, 500 millisecond, 800 millisecond something like that at least 8 to 12 we need to do. Since these are INEPT based sequence, so we have to also look at what is the typical INEPT sequence. So, you can say you set the time delay which is here τ as 2.7 millisecond. So, that is what we set essentially and then we set the carrier frequency.

So, in proton dimension it should be set at 4.7 ppm or something at water. In the N15 dimension, you have to set it somewhere in between. So, if we are doing at 100 to 130 we typically set at 115 or something like that 118, 115. Now, you acquire the data, so like minimum should be 128 complex point in T_1 and for proton it can be 1k (1024) T2 points.

You can also record with this as 256 and 2k. So, that is a typical parameters that we are using. So, as we said we are recording 8 to 12 2D points. So, here is a representation of 2D we are recording and with a different time point, different T_1 delay, the signal is going to change. So, essentially we start with an inverted signal then little bit less, little bit less and that is how it goes something like this.

But in 2D, the signal is going to decay as we increase the T_1 point. So, we record this data, the 2D data and then do the Fourier transform for each of these 2D data. Before doing Fourier transform, we just do the usual stuff like a multiplying with a 90° shift square sine bell function or cosine bell function or Gaussian whatever you fit it, then you zero fill at least twice of the digital resolution. So, you can zero fill up to 2k and 1k or 512 in direct dimension. So, that digital resolution is typically of 2 hertz per point and then you can apply a 90° pulse on proton dimension or nitrogen dimension, zero-filling is required before Fourier transform. If required you can do linear prediction that improves the digital resolution.

But all the time it may not be required. So you can take a call whether you want to have a linear prediction or not. So these are cosmetics or data processing steps. Once you have processed data, all 2D, we are recording they should be identically processed. So for measuring the intensity, so once we process identically using these steps. Typically if you open Bruker Topspin that is what I have taken from. This is the zero-filling, you can do up to 512, 2048. This is the frequency like spectrometer frequency, the SR values, then window function what we want cosine or sine, line broadening what is the SSB. You can choose all of these parameter that you need, phase correction value. BC mode, all of these

you can choose, whatever is needed for data processing, getting a nicer well separated peaks because you cannot measure confidently intensity of the peaks, which are merged something like this. So, you need to process so that all peaks are very well separated and one can only take well separated peak for data analysis.

So, we create this series of 8 or 10 or 12 2D, then for each peak, we are going to measure the intensity. So, series of these spectra is phase corrected and then proton dimension is adjusted according to the first FID. N15 phase is corrected here. So, phase you have to correct for the both dimension. Now, essentially we are measuring the amplitude of cross peaks or we can measure the volume integral of each peaks and signal should not be overlapped.

So, you need to have all separated signal in the 2D for the data analysis. We measure the volume integral or intensity say *I*j at any point any t point, say 2 millisecond, 5 millisecond, 500 millisecond. So, all those intensities are measured and that is fitted in a equation to measure the longitudinal relaxation time constant. So, what is the equation that we are fitting? Essentially, this is the equation, that we had discussed earlier to get the T_1 parameter.

$$\xi = 1 - 2e^{-T/T_1} = 1 - 2e^{TR_1}$$

So if you measure it for say various amino acid that I am showing from one of the protein work. So here for leucine 24, phenylalanine 36, lysine 46 you see with the T_1 time which is varying from 0, 100 to say 1.2 millisecond intensity is plotted and you can see with a time this intensity decreases and that is what we are going to fit into this. So, we fit this equation and find the T_1 time or the R_1 rate that is coming.

So, this fitting we can done in residue specific manner for all non-overlapping residues. Once we fit it we can get the T_1 relaxation rate for each of the amino acid and now one can plot it in a residue specific manner. So, here I am showing you a protein which is called human SUMO that I had worked on. Now, this protein has a globular domain, which is here. It has a similar fold like a ubiquitin, ββαββα fold.

And it has a long N-terminus tail of about 20 amino acid and short C-terminus tail of about 4 to 5 amino acid. There are loops here. So now we are going to discuss the T_1 data that

we have recorded for this protein. So, here is the residue specific manner plotting of this data. What we see here, so initially the T_1 of these 20 amino acids shows higher value.

Now for all the well-folded region, α-helix, β-sheet or so, we are seeing the R_1 value the relaxation rate for the flexible domain is higher. For well-folded domain is lower. That means T_1 value for this is going to be shorter and for this is going to be longer.

So T_1 and R_1 has inverse relation. If you look at all the flexible portion has a high longitudinal relaxation rate. Again loops here, which connects the two β from here and some of the C-termini again have high, here loops are high. So all the loops you can see shows high longitudinal relaxation rate and all well-structured domain in this protein whether it is α-helix or β-sheet shows relatively less relaxation like longitudinal relaxation rate. Now suppose I put this protein in urea, where all major secondary structure is removed. So suppose I am doing the same experiment in 8 M urea, which denatures the protein.

So, now the typically α-helix β-sheets all of those are gone and now you see what are the rates. So, if you look at here residue specific manner 0 to 100. But we see here that more or less it becomes flat. So whatever you see sequence-wise variation here, when the protein is folded, if you denature, the relaxation rate is gone. So if we measure the T_1 relaxation rate of a protein, it tells the motion in a residue specific manner, the longitudinal relaxation rate in a residue specific manner.

And if we remove the structural elements from the protein, this T_1 relaxation rate becomes majorly flat. That means all the relaxation, sequence-wise variation that we were getting is becoming absolutely similar. So that was about longitudinal relaxation rate, which is called spin-lattice relaxation rate. Now coming back to another relaxation rate, which is called transverse relaxation rate, the T_2 relaxation rate. T_2 or $T_1\rho$, T1 in a rotating frame are more or less similar, I will just discuss the T_2, how we measure the transverse relaxation rate.

So, the sequence of T_2 was initially developed by Farrow et al. that I am going to show you in the next slide, it is similar to T_1. It is a similar to T_1, instead of creating a –z direction by applying of 180° pulse here we create a coherence. So, here the inversion scheme is replaced by a CPMG or spin lock sequence, CPMG in case of R_2 and spin lock in case of

$T_1\rho$. So, like here in T_1 what we saw our magnetization started with –z direction and it went back to the +z direction. Now, in this case we are starting with a transverse relaxation rate.

So, rather than decaying along the longitudinal relaxation, the heteronuclear magnetization relax in the transverse plane during this time T in the T_2 pulse sequence. So, here we are not going like this, here just by applying a 90° pulse or CPMG pulse we are coming in the transverse plane and from here, now our spin is going back to the equilibrium state. So, what contributes? So, in addition to spin-spin interaction, because this is transverse relaxation rate is spin-spin relaxation rate. So, other than the spin-spin interaction, the magnetic field inhomogeneity also contributes to transverse relaxation rate. So, to remove this contribution coming from field homogeneity, the CPMG spin echo sequence which was developed by Carr, Purcell, Meiboom and Gill was introduced in the T_2 scheme.

So, CPMG scheme of heteronuclear magnetization, what happens here, we started with a Sz and by applying a 90° pulse we came to Sx. So, Sx evolved during this period say ε, under the interaction of chemical shift and field inhomogeneity. Then we apply a 90° pulse, which reverse the direction of precession. So, we started here, we went back here and now it is slowly dephasing and then we apply a 180° pulse. So, it goes like this and then slowly it dephases and then finally, it refocus during the second period of ξ.

So, first it dephases, then you apply it and then it come back. So, the provided spin being refocused remain in the identical magnetic field during this both period of ε. So, then we can measure the rate. So, resulting transverse magnetization at the end of the even echo period of the CPMG pulse strain has an amplitude something like this, I is an intensity at any time point and I_0 is the initial intensity. Now, T is the time period with which we have like waited for and R_2 is the spin relaxation rate.

$$I = I_0 e^{TR_2}$$

So, that is what we measure basically for the T.

So, now $T = 2n(2\varepsilon + pw_{180^\circ})$

So, that is a CPMG pulse strain that we are measuring and essentially we are measuring again intensity during the T2 period. So, this is our pulse sequence that we are using. We started with INEPT block, we are coming back to here nitrogen, then we are introducing the CPMG pulse and during this the T_2 relaxation is happening, then we are frequency encoding going back to proton, detecting on proton while decoupling the nitrogen. So, this

is the T period during which transverse relaxation is happening and that is T_2 happening on nitrogen nuclei.

So again we record some 8 to 12 of 2D with different t to time, the Fourier transform them, process them and then we measure the intensity in a residue specific manner as we did for T_1 and you can get the R_2 rate. So here typically the duration that we keep is 0 to about 200 millisecond. So here you can see for different residue like Q53, K25, and isoleucine 88. Here the signal intensity decay are measured. So, now each of this fitting gives us a rate of a particular N15 nuclei.

So, data processing is done in a similar manner, typically 8 to 12 relaxation delay ranging from 5 millisecond to 150 millisecond is recorded and data processed similar like T_1. So if you do that and plot in a residue specific manner, now what we get is this one. So here, this is again rate. So here, for the flexible portion, we are getting the less R_2 value, whereas all the structured portion, we are getting more R_2 value.

Remember, this is spin-spin relaxation. Now, this was slightly different in case of T_1. Here, if you look at this reverse here for the flexible portion, we are getting the higher R_1 rate, lower R_2 value. For the structured portion, we are getting typically lower R_1 value, but higher R_2 value. So, this is the typical signature we get. Again at the end you see the lower R_2 values are there. Now, some of these peaks are also showing the high value, this may consists of ($R_{2,intrinsic} + R_{ex}$).

Now this is exchange phenomena. So R_2 also encodes the chemical exchange that is happening. For doing that you require relaxation dispersion experiment. But looking at this, you can say that this protein is geared to show some chemical exchange which needs to be further probed. So, I am not going in detail of the Rex at the moment, but that information is hidden there.

Now, similarly if you look at the denatured protein again. If you put this SUMO protein in 8 M urea, you can see R_2 becomes flat. So all exchange are suppressed, all the sequence wise variation is suppressed. So here is the folded protein in 0 M urea R_2, and for ready reference I am also showing in denatured state. So everything is suppressed, now protein

behaves quite uniformly because it was quite folded and when you put in 8 molar urea it become like a disordered, right.

So, just to show that the dynamics that we are measuring also tells a lot about structural compaction, where is the flexibility, where is the rigidity. So, not only structure, but also NMR can probe dynamic in a residue specific manner. Now, you can even use this concept for protein folding study. So, here I show you how urea concentration wise R_2 is changing. So, in 8 M urea you have more or less flat profile of relaxation rates, but when you reduce the urea you see some central portion is coming up, like it has a profile if you go to 5 M it has this profile.

So, all the central portion where a structure was supposed to be there basically building up and that is how measuring the R_2 dynamics also tells about what is the folding hotspot, where the structure is forming and that is what you can measure it in a residue specific manner to understand the protein folding. So, with this I just want to summarize today, R_1 and R_2 are two basic experiments that are used for the protein NMR dynamics. Next class I am also going to discuss about heteronuclear NOE. R_1 measures the longitudinal relaxation rate, R_2 measures the spin-spin relaxation rate, and for the folded region, there will be low R_1, high R_2. For a disordered region or the flexible portion, it will have high R_1, low R_2.

R_2 can be used to monitor the structure that could be formed during the protein folding that is what I am showing you here. So, with this I am going to close it today and in next class we can take heteronuclear NOE and how we can combine all these basic experiment of protein NMR T_1, T_2 and NOE, or R_1, R_2 and NOE to deduce spectral density function or even modal free.

So, with this I am closing it here for today. See you in the next class. Thank you very much.