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

Lecture: 60
Basics of solid state NMR spectroscopy - V

So welcome to today's lecture. This is going to be last lecture of this course. In this week, we are discussing basics of solid state NMR and its application in structural biology. So, till now we have discussed what kind of samples we can investigate in solid state NMR, what are the problem associated with solid state NMR and how we can resolve those. So, two of the main component CP and MAS, cross polarization and MAS help us to resolving the line broadness and low sensitivity. To some extent we can get rid of those and I also discussed how to do resonance assignment in solid state NMR, what kind of pulse sequence you use, how we reintroduce the lost interaction which were lost because of magic angle spinning. So today we are going to discuss more in detail two of the major application of solid state NMR in two kind of biological system.

One is amyloid fiber and another is membrane protein. Amyloid fiber, I will discuss little in more detail than the membrane protein. So let us start with amyloid fibers. What are actually amyloid fibers? So it was proposed by Christ Dobson that aggregation is a generic property of any protein.

Under certain physiochemical condition all protein aggregates, whether it is intrinsically disordered protein or β-sheet protein or α-helix protein. So, what happens when aggregation starts, they actually partially unfold and they make some kind of intermediates, which can form various kinds of aggregates. It can be amorphous aggregate or it can be ordered aggregate like a pore like aggregate. Now these pore like aggregates if it interacts with a biological membrane, they forms pore in the membrane and because of that actually it causes toxicity.

So, you see the membrane is very tightly regulated entry point for anything and this pore actually punctures the membrane and that is how it becomes toxic. These oligomers can associate themselves into fiber like a structure, which is called amyloid fibers. If you look at the tissue where aggregation of protein has been found, one can look at the TEM picture, the rod like a structure and this is amyloid fibers.

So still lots of people are debating because this is found in the aged tissue sample, the aggregates. So they are discussing whether the precursor of the amyloids like oligomeric states is toxic or the amyloid fiber itself is toxic or even does it has any protective function. So this debate is still going on. I will not go in this debate in this course.

This is matter of another discussion but just to emphasize upon that all sorts of protein has probability to form amyloid fibers and these amyloid fibers are generally not soluble molecule. So there are some way we can get their structural information. One is by staining. You can stain with some dye, certain dye and look at in the cross polarized light. So if you look at the cross polarized light, this is amyloid specific dye.

You can get an apple green birefrignence by staining those and amyloid is the word derived because it is very starch light. In cross-polarized light you get a staining of these amyloid fibers with a certain fluorescent dye. So that is the one of the characteristic for staining. If you look at the TEM and AFM you see fiber like structure.

So this is just morphology determination of these amyloid fibers. Typically with low resolution structural entity you can see with CD, that there is a minima at around 218 or 215, this confirms that these aggregated structure are rich in β-sheet. Similar thing is corroborated with FTIR, this is a migone stretching, you can see the wavenumber of 1635 which corresponds to the β-strand or β-sheet structure that is found in FTIR spectrum. So all those confirms that they look like rope, rope like a structure, fiber like a structure. They can be stained with a specific dye to detect it, they have beta like a structure.

You cannot do crystallography but still you can do fiber diffraction and you get two characteristic reflection one at 4.8 another at 10 and this particular pattern of reflection of the amyloid fibers in fiber diffraction pattern suggest that they have a cross β-structure where the β-sheet is in this axis and β-strands are arranged here. So inter sheet spacing is about 10 angstrom and inter strand spacing is about 4.8 angstrom. So these are the characteristic features but this is of course not the atomic resolution or high resolution structure. Now another important features of these amyloid fibers that they have various polymorphic structure. The strands that we saw in the previous slides they can be arranged in a various fashion like this is secondary structure where β-strands are there.

It can be arranged in a particular fashion to tertiary structures. It can have a parallel orientation; it can have anti-parallel orientation and those leads to the polymorphic structure of these proteins, the amyloid proteins. So actually, it is very important to understand looking at the just morphology is not enough, it is important to understand their quaternary level structural arrangement to investigate their property. Now today I am going to focus one of the protein, which is α-synuclein. This aggregates and cause the second common neurodegenerative disease, which is called Parkinson's disease.

It was first investigated by Sir William Richards Grover PD in 1986 and it has a clinical feature like movement disorder, muscle rigidity, tremor, slowing of the physical movement in extreme cases and loss of physical movement. So what happens in this disease? Some protein gets aggregated in the neurons of the brain and this portion where they get aggregated is called substantia niagra. So in normal because of aggregation what happens that cellular communication gets impaired. So in normal condition here is a neuron and that is communicating with another neurons using neurotransmitters.

However, in case of Parkinson's disease the protein aggregates makes Lewy body and because of that the communication between neurons is impaired and that leads to movement disorder. So this Lewy body is found in the brain. So these Lewy bodies are made up of various kind of β-sheet structural arrangement which can be seen in the amyloid fibers and one protein which is responsible for this amyloid formation is known as α-synuclein. However, over the years of 100 years of research proposed that although the clinical features look more or less same in Parkinson's disease but causative agent α-synuclein has various mutations and these are familial mutations. So there are mutation at 53 position A converts to T, E converts to K like E46K, A30P.

So these are various familial mutants across the globe people had investigated. The question to ask that in case of different Parkinson's disease in different people are the protein that is there having some mutation. Are they going to form same kind of fibers? So this we cannot answer unless until we have a detailed structural investigation. Just looking at the gross morphology of these amyloid fibers is not good enough to investigate what is a real causative species whether it is oligomer and oligomers are different or same or what is happening. So one need to have a detailed investigation of these.

Now Ronald Mielke found it out that there are various strain behavior that has been shown in α-synuclein propagation. So here is a neuron, this is the place where the protein gets aggregated, this is the protein α-synuclein and it forms two kind of polymer. The polymer one, which was looking more straight, and this was looking elongated and sheet like structure. So strain one spaghetti type, strain two was more straight kind of like more straight but twisted. These kind of structures they found and they saw that they are showing different strain behavior.

So various causative disease are caused by different strain. So here again it is a protein strain that is different in two different patients. So that kind of polymorphic nature also this protein exhibit and they have a different functional property as well. So the toxicity with different strains is different. The binding propensities to the cell membrane is different.

Also the seeding capacity, how fast they spread is also different. So this proves the case for looking at the detailed investigation of these amyloid fibers using one of the technique that we can do it. Now over the years solid state NMR has been proved to be a very aptly suited technique for detailed investigation of these, primarily reason that you do not require any long range order for doing solid state NMR. Neither you require solubility as we discussed. So these proteins can be isolated or grown in bacteria and obtain the particular kind of morphology, just go for solid state NMR to get the hydrodynamic and structural aspect.

So let us see what we can do with solid state NMR. So one thing is clear that all protein like all amino acid in protein will not forms the aggregated core, there will be probability that some remains still flexible. So first let us edit this amyloid fiber based on the motion. So now suppose a particular portion of protein is aggregating and rest remains unstructured, here is the aggregated protein structure. So we can employ some kind of a pulse sequence where we use solution like concept of transfer of the magnetization and here we are using solid state like concept to transfer the magnetization.

So if you remember this is J-based transfer we are doing, for J-based transfer what has to happen? Dipolar coupling should be average, right. So then only the dominant transfer mechanism will be J-based. So if you do that we can investigate the flexible portion of the protein and in this pulse sequence what we will do? We just using this T_2 filter out signal from the flexible portion and have only signal transferred through dipolar coupling for the rigid portion. So if you do this complementary experiment of the flexible portion as well

as for the rigid portion, we can find it out the portion which is involved in the aggregation and portion which just hangs around. So for this kind of flexible portion, mobile portion, we can use INEPT based transfer or INEPT-TOCSY and here we can use PDSD dipolar based transfer.

Using this two complementary solid state experiment, we already dissected the amyloid fibers in two portion, one which was rigid, one which is flexible. So that is already a good start, right. So we can identify the amino acids, that are not actually forming the core of the fiber and which forms the core of the fiber during this dipolar base transfer. Good, so after identification, let us now focus, this is easy, this we know how to identify using our liquid state concept.

Let us focus more into dipolar based transfer. So I am just starting with a PDSD that we discussed in our previous class. So if we take the amyloid fibers, this was 13C-N15 labelled, grown in bacteria, put at right condition which form the fibers and you can see beautiful fiber like structure using electron microscopy. So this is our materials that we are investigating. This resembles the brain fiber structure but for NMR purpose we have labeled with 13C-N15.

Now we record the carbon-carbon correlation spectrum using proton driven spin diffusion condition and here are the parameters recorded on 850 MHz, spinning speed was 11 kilohertz and mixing time of PDSD was 20 ms. So what we expect to see mostly intra-residue correlation and that is what we see here. So let us see from the previous class now you can appreciate, what this could be? It is a serine Cα-C β. Serine Cα Cβ, so if you draw it here, it is the same peak on two side of diagonal, here is the Cα backbone and this is the Cβ. So serine Cα Cβ, what is here? This is threonine Cα Cβ.

So what we see that in this we see at least 3 serine and about 10 threonine that can be investigated. Now looking at the other region, this region, if you remember this region, we had discussed this, this is actually alanine region. So we can identify asparagine region. This region is little crowded and here we have lots of amino acids like E, K and all those. This is again threonine Cγ, and here is our valine Cβ, and Cα-Cβ, Cα-Cγ correlation.

So in this region we can find it out alanine, serine, threonine, it can be D or N and this can be E, K, this is threonine again, this is K and this one is V. So these are various regions that we can identify. Now, here if you look at this region is isoleucine, so we can start assignment from all over the anywhere whatever I identified. You can use the PDSD or DARR of different mixing time to introduce the interresidue contacts and that will help us in assignment. So these are starting point of the assignments. Now if you look at some of the peaks looks really crowded.

So it is not so easy to assign this. But, then we have to combine this with some higher order structure or we have to do simple various sample preparation to reduce the complexity. So what kind of sample preparation we will do? Just briefly, I will discuss with you. So first scheme can be non-diluted sample where all the molecules of the amyloid fibers are going to be 13C-N15 label, uniformly 13C-N15 labeled. So we are growing our bacteria in 13C-N15 label media and all α-synuclein whatever it has made was 13C-N15 label. We aggregated those and this is our non-diluted sample.

The next sample we can do to reduce the inter-molecular contacts by diluting it. So what we can do, we can take one label molecule and mix with various un-label molecule. So this will be a statistically distributed and one can have diluted sample. Or we can even have a mixed sample where one molecule will be exclusively N15 labelled, another molecule can be exclusively 13C labelled, something like that.

So this will be mixed label. So for uniformly labelled sample you have to start with 13C-labelled glucose or glycerol and we can have even diluted samples depending upon how we want to spin dilute it, you can use 1,3-labelled glycerol or 2-labelled glycerol or various glucose as well. So these are various sample preparation schemes or labeling scheme that is employed for various sample preparation. But, let us start with simple uniformly labeled sample where we are combining now, here is our NCA spectrum, like nitrogen 15 and Cα correlation. Here is our NCO experiment and this is the section taken from PDSD.

This is also section taken from PDSD. This was carbonyl region and this is aliphatic region. So let us start with any amino acid that we like it. So as I said serine was very nice, coming at distinct position, so here is our serine. Now we are choosing one of the amino acid from serine, here is our S87, this is Cα-Cβ correlation. We can go up by aligning in NCA correlation, we can find this is serine 87 N, so here is Serine Cα-H-Cβ-OH.

So, here is CO. So, this serine Cα we are getting it here in NC correlation just above this, this is serine residue and this is Cα of serine residue, this is Cβ of serine residue. Now we can go one step backward and in NCO correlation we can find that at the same N15 position the CO of 86 is coming. So we identified i-1 CO which was this, so this correlation we identified it. So this is NCO correlation, so we identified this. Now we are using this CO we can go here and we can find it out the 86 Cα-CO correlation wonderfully depicted in the CC correlation spectrum of PDSD.

Now from here 87Cβ we can go to 87Cβ-CO correlation. So if you look at we identify Cα, Cβ, CO, N15 of 87. We can walk above and we can see a correlation of 87CO plus 88N and that helps us in identifying 88Cα. So now, we can come here and we can find it out 88Cα and then go to 88Cβ, Cγ, Cδ and so. So by doing this exercise for few days or few weeks, we can identify each of these amino acids from these, even not so high-resolution spectrum.

But then when we go to a region like whatever I showed you here, it becomes really difficult where to put it. So we have to get a better spectrum and that better spectrum can come with a smart sample preparation. So what we need to do, so if we just start with one of the carbon label in glucose like two carbon label glucose or one carbon label in glucose you see we enhance the resolution many fold. So this is the spectrum coming from uniformly label glucose this is coming only from the second carbon label glucose and you can see that spectrum becomes beautiful although this is very expensive compared to this but you enhance the resolution quite a bit. This is the same carbon-carbon spectrum that you see, of course some of these correlation will be missing but that biochemistry one has to work what will appear and what will not appear.

But the same spectrum you can resolve drastically. Now such a spectrum is really boon for the assignments. So this is called a sparse labelling where we can start with either first carbon label glucose or we can start with either second carbon label glucose and we get such beautiful spectrum where all peaks are separated. Then that can be used for resonance assignment that I discussed here is the carbon-carbon correlation spectrum, the carbon NCA spectrum, NCO spectrum. So you can see here 44T N-Cα. going to 44, 43 CO, coming to here Cβ, 43 Cα and 44 N.

Similarly we can walk across the 42Cα, Cβ and all those. So even not only the same residue, intraresidue correlation, we can get it. Now the long range, short-range correlation depending upon how long we have mixed it and that helps us in precise resonance assignment. Just to repeat once again, so we know that this region is serine Cα, we can go to N15 of serine Cα-Cβ. Now from here we can see that this peak was connected to another peak, which was Cα.

This is like a near neighbor residue correlation. So Cα of 43 is obvious choice which will be connected because this region comes as a Cα of 43. So this peak is 42 Cα with 43 Cα. Now we find it out 43 Cα and we can go this is your 43 ε, β and all those. So therefore by doing this exercise for months and months we can find it out all the resonance assignment and once we find it out what is the next? So this is one, we find all the resonances by doing month of exercise. But now the question comes that how these are arranged in a supramolecular structure.

This we find it out about a sheet, one white, one monomer. That here is my same residue assignment, here is my next residue assignment. How they are arranged? in different sheets. So for doing that we can use one of the pulse sequence, which is NHHC. It starts with a proton, transfer to nitrogen, then transfer to carbon, back to proton and detect on carbon. So that is a NHHC, we can have a smart sample preparation scheme that I discussed in the first slide here.

So now here we are preparing a mixed label sample where one molecule is N15, and another molecule is exclusively 13C label and if you apply that NHHC experiment we can find it out the orientation of these strands and that is what we are going to do now to find it out supramolecular arrangement of the fibers. So one, we have mixed it N15 label sample with 13C label sample here and then we have employed this NHHC. So this correlation can be only established if we have the like distance of these two strands about 5 angstrom. So start with a proton connected to N, transfer here and then using CP condition we establish this correlation, transfer to proton and then we mix the proton and detect on carbon.

Now if we do that what we found on mixed label samples many of these residue are appearing in the red and this black was typical NCA. If they are appearing what they are saying that these strands are arranged in a cross position, it is a very close position and all

those can be assigned because this was assigned, the black spectrum is assigned. All those wherever they are appearing one can find it out all these correlation. Now you can see, if this peak is appearing that clearly says distance between these two strands were 4.7 or 5 angstrom.

Now that is possible only if these strands will be arranged or the sheet will be arranged in a parallel in registered manner and that is what one can conclude here that different monomer in α-synuclein are arranged in a parallel in registered manner. Now we had all the assignments, so this was supramolecular arrangement but how these different strands are arranged in sheet that we can find it out typically using carbon chemical shift and secondary structure calculation. We can use all those carbon chemical shift or N15 chemical shift that we have assigned, use in predicting prediction algorithm that is TALOS, find it out that phi and psi torsion angle, and then using that we can just plot phi-psi torsion angle and find it out from where to where this strands exist along the sequence of the protein. Then we can find it out the different strains that we have talked about say strain number 1, strain number 2 and strain number 3, even the core is same but you can look at the secondary structural arrangement seems to be different in different core.

So that actually initiates the strain behavior in different fibers. Because the strands positions and location and the length of the strands are different in α-synuclein fibers. But, that is not enough for solving the structure. We found the secondary structure, the core of the fibers, even we got some idea about supramolecular arrangement but how they are arranged in three-dimensional space. For doing that you have to find it out long-range resistance constraints. So this can be done by experiment called PANE or PAR or long mixing time PDSD.

I will not go in detail but essentially what we have to do record longer and longer so that we see inter residue distance constraints, short range and long range distance constraints and if you can find all those assignments you are set to go for a structure calculation. Another thing you have to find mass per unit length, so for that you have to combine another experiment called STEM, give you mass per unit length, how many molecules will be arranged in particular mass. So here we get a tertiary arrangement, using this we can get a quaternary arrangement and using experiment like NHHC which gives us idea about the supramolecular arrangement, we club all these data and then using any of the structure calculation program like CYANA or SOV or CNS, we can solve the structure of the amyloid fibers.

So if you look at here, this was solved by Chad-Jinsta group of α-synuclein. He found that monomeric α-synuclein, the rigid core is from 44 amino acid to 97 amino acid and this expands up to 4.5 nanometer. Then using STEM and long-range distance constraint, he solved the atomic resolution structure and this monomer were stacked in a parallel in registered manner where some portion was flexible and rest portion was rigid.

He proposed a Greek key like arrangement for α-synuclein fibers. So this was the synuclein high resolution structure that was solved by Chad Jinstra group in 2016. This was fibers but microcrystalline protein was, this is the first structure I especially I want to mention this was solved by Oskinert group of this microcrystalline protein and they actually first showed how you can use solid state NMR to solve a high resolution structure of a spectrine SS3 domain and that was published in Nature in 2002. Since then lots of structure has been solved, I just showed you one of the amyloid fiber structure which was solved. So next 5 minutes or so I am going to give you a glimpse how you can use solid state NMR for investigating membrane proteins. So you know membrane proteins are very important proteins because lots of cellular functions are done by these proteins.

Like signal transductions for an example GPCR are important class of protein. This is a huge drug target G protein coupled receptors, molecular pumps like a calcium ATPase or ion channel like a potassium channel. All those are membrane protein and they perform extremely important function. But getting their high resolution structure is difficult task because their native setting is membrane, therefore they have a low solubility, they have difficulty in crystallization.

So we need to think to capture their structure and dynamics using native setup and membrane is their native setup. So we have to put them in a right condition. You see one-third of the proteins are membrane protein, genome codes one-third of membrane protein but structure solved is only 5 % or even less than that because of difficulty. So it is still wide open field for using various biophysical technique to getting the structural arrangement.

Membrane protein can be of different shape and size. It can be integral membrane protein, a single helix pass, multiple helix pass, like a helical bundle. This is enzyme and transporter receptor. This can be recognition receptor. β barrel protein like transporters, the channel

protein, it can be peripheral protein, only portion of this protein will be membrane, in membrane, rest hangs around.

So you can see here is enzymes or transporters or carrier are of peripheral protein. One thing to be important, the sample preparation is really crucial because their native setting is membrane. So you have to put in right condition, it can be like earlier people are doing in micelles, but micelle you can see it is a circular structure, so there is a possibility of putting lots of pressure and this is a single layer, right, a round shape. So detergent micelles is not so nice. Then people moved into bicellar conformation. So bicelles, here you have a linear arrangement, then you have circular arrangement, this is used for liquid state.

Then people evolved and they proposed with a nandisk where lipid is forming a nanodisk which was bound by a membrane scaffolding protein and membrane protein is embedded here. So nowadays this is one of the dominant system that is being used for studying the membrane protein. But the natural environment is liposome. Can you constitute your membrane protein liposome? If you constitute it, now molecular weight increases many fold and therefore here comes the solid state NMR to investigate this protein. So in next few minutes, I am just going to give you only one example of a membrane protein that is VDAC voltage dependent anion channel.

This is a mitochondrial outer membrane protein. It is involved in transport of NADH and ATP that you know that is energy currency in the cell. So this is in mitochondria. It has a profound role in energy transduction. So this has β barrel kind of a structure like here is a β-barrel and there is a short helix which works like a gate closing and opening mechanism. So here is a helix and it can flip here and out.

So to understand the precise mechanism you need to have a structural information of this. Over the years, people worked hard and solved this structure. So Sebastian Hiller, he solved the structure of VDAC using liquid state NMR that was published in Science. Then Christian Griesinger group using NMR and X-ray he combined these two techniques and solve it this structure. Then subsequently another X-ray structure came which we showed this structure. If you look at the important portion that is missing like if you look at the β barrel more or less remain same, but this helix which is the actual controlling unit of the VDAC function is different.

Here it is more elongated and random coil here, we see more formed helix. So what is actual conformation? Can we look at the details? Can we understand the mechanism of this using the better technique like a solid state NMR? So the published 3D structures define in functionality and important N-terminus region especially and native conformation of the lipid because this was in solution, lipid environment was different, here it is a kind of mixed NMR and X-ray and this was kind of frozen. So let us look at in native condition. The one-way approach again put a right kind of lipid and solved in the LDAO micelles.

So this was healer solved in micelles. You can see here the flexible portion is still there. This was a spectrum but in actual this should be an helix. That is what X-ray structure said. So Adam Lange group tried putting in different membrane environment and you can see depending upon what is your choice of membrane they can put a lateral pressure or various sorts of pressure. So choosing the membrane environment is very important that can change your spectrum.

So this optimization is required. Here I showed you a spectrum if you put in DMPC or DOPC. So if you see DMPC is of right length of your membrane protein, DOPC is slightly longer, DOPE is putting a pressure on the membrane and that changes the spectrum minutely. Because this minute change is very critical. So you have to choose a right kind of membrane environment and that depends upon your protein, what kind of membrane setup you have to have for bilayer composition. So you have to do little bit of optimization. But once they optimize it, they solve this structure and what they found that this membrane, this helix exists as a helix.

It has a nice connection with one of the β-strands here and during its functionalities, they showed it with various sample preparation. This helix can open it. So there is an interaction between seed β strand and helix one of the residue, this can open it. So there are lots of dynamics associated with these membrane proteins which can be nicely captured if we design right kind of solid-state experiment and that Robert Schneider published in 2010 in Angewadnte Cheme. So that's what a brief introduction or exposition of solid-state NMR to investigate two difficult biological macromolecular system, amyloid protein and membrane protein.

This gives you lots of excitement where else we can utilize and how we can use this wonderful technique for answering the subtle questions that exist in biological system. So

with this I would like to summarize what we discussed in this week. Solid state NMR has become a method of choice for complex biological macromolecules, which are not amenable to liquid state, which are not amenable to crystallography and which are not very easily tractable by the cryo electron microscopy. All those can be investigated with solid state NMR. The one important part, the dynamics associated with these molecules can be nicely investigated with solid state NMR.

Just now, I showed you for VDAC and that is an important advantage to have solid state NMR. Now in solid state NMR, we have to do a judicious combination of MAS and different RF like we saw it, that made it possible to look at the structural details and we need to also introduce recoupling sequence that does the selective manipulation for spin interaction that establish the long range interactions and that helps us in getting the structural and dynamics parameter from these macromolecules. So over the year a remarkable progress has already been made in various biomolecular system and that is how today solid state NMR has become so popular technique and lots of developments are happening to increase the sensitivity, to increase the resolution, to ask and answer difficult questions in biology. So this is the age that has come that Now you can combine solid state enamel as one of the major biophysical technique with other techniques to answer the questions in the native setting of the protein whether inside the cell or inside the membrane, it is your question and your answer. But, solid state NMR has a significant role to play in coming years. With this, I would like to close here and thank you very much for being part of this course. Thank you.