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

Lecture: 46
NMR in Drug Design

Thank you. Okay students, so good morning. Last week we were discussing how to understand protein-ligand and protein-protein interaction using NMR spectroscopy. So we actually looked at that we can adopt two approach, in one where we can detect on ligand, in another where we can detect on protein. In ligand-detected NMR experiment, you do not need to isotopically label anything. There were some experiments like STD-NMR, saturation transfer difference, water-LOGSY, looking at the linewidth or so and so forth can be used to understand protein ligand interaction.

On the other hand, when we are using protein as a probe, we have to isotopically label this, so that we can investigate this interaction in a residue specific manner. So now onwards this week we are going to look at how this information like a protein-ligand interaction, protein-protein interaction can be used to design and develop a drug. So which will be more specific, more precise and even it can venture into like unprecedented territory where the drug has not been desired. Like looking at the disorder protein, looking at the interfaces, which is quite

wide and broad. Now in those cases also NMR plays very very important role in drug discovery, design and development. So, here onwards we will be looking and more concentrating on those aspects of drug designing and development. So let us start this week. So how drug development actually happens in a pharma industry? For targeting a new drug is like a fishing right, you are fishing it, this I have taken from internet. So, you are fishing a molecule out of a pond where there are many molecules, many receptors and you want to target one, like here say this is some food for the fishes.

Now you as a fisherman, attach food and there are so many fishes that can eat it. One of them by luck by chance will catch your molecule and then you can fish it. Drug discovery and design is essentially like that. Most often the development of a new medicine starts when a basic scientist like all of us learn about a biological target, which we want to target by a molecule; the target can be a receptor or an enzyme or a protein and gene and just by

curiosity, we want to target it, we want to block this, we want to like have a binding pocket where we want to put a molecule, that basic questions leads to a drug discovery. So, do not forget in all this translational research, the foundation is led by the basic research. The intense involved basic research is behind any of the translational research.

So basic scientist with a question and query come to target any of the receptor, enzyme, protein, gene and that leads to a profound drug discovery. So it is like a many years of work, many failures, some success, much uncertainty, that all involves in a drug discovery process. So how do we start? So even like in a programmed manner that most pharma company likes to do is they want to target a particular disease like you have to choose a disease, it can be obesity or diabetes or a cancer or autoimmune disorder. So first thing, you have to choose a defined disease against which you want to develop a drug. Once you identify then you are going to do lots of basic research to identify what are the critical parameters involved in the disease, what are the genes involved there, are they getting up-regulated, are they getting down-regulated, what is the cascading effect of that up-regulation or down-regulation, what are the proteins that are involved, how signaling cascading changes in those proteins.

So eventually, you are targeting a gene or a protein by a drug target that you have to find it out. So once you identify a disease and you identify a drug target by doing the essentially fishing, you have to identify an bioassay, a readout which determines the biological activity. So suppose your protein, name a protein, is involved or is responsible for particular disease, which has been up-regulated or down-regulated. You also identify a drug target that possibly can bind it.

But how to read out that whether my drug target is binding to that protein, doing something? So for that, you need to develop a bioassay where you can look at the binding, you can look at the catalysis, you can look at the gene expression. So some readout has to be defined for a particular target, the drug target and protein target, that is called bioassay. So you develop a bioassay like ELISA based assay, can be fluorescence based assay, can be the surface plasma resonance, where two protein or protein ligand binding has to be established.

So all of these one need to establish this. And once you have a lead compound which binds to your target molecule, you identify something called pharmacophore. It is a structural

features that directly responsible for the activity of the drug. So that is a pharmacophore you bind. Like these are the epitopes, that is what we studied in the last week that we can identify the binding epitopes which binds essentially to the target molecule. Identify a pharmacophore, the entity that is responsible for the activity.

So once you obtain this pharmacophore, you need to optimize the structure of that pharmacophore that improves the interaction with the target like you can do the chemical modifications. So you can talk to your medicinal chemistry friend and see look these guys seems to be playing an important role. Can you change some chemistry here? Can you put methyl group, ethyl group or make an ether, make an ester, whatever like you can ask them to do. They will do a like a knowledge based modification of those molecule. Again you perform an bioassay to find it out whether the readout has improved or it has like a readout has become worse.

So looking at this in an iterative fashion, you modify your drugs, tune your drugs for a better efficacy. So that is typically what we are doing? We are finding a lead compound, which binds to a target. So lead compound is a structure that has some activity against a chosen target.

Chosen target is a protein or a gene and lead compound typically here we can think of a small molecule. So we identify a small molecule based lead compound that binds to a target, shows some activity but yet not good enough to call it drug, right. So one need to later on modify based on this pharmacophore mapping. So if not known, determine the structure of this lead compound that will be next step. You identify suppose your lead compound from a library of a compound or a plant source, right.

You have, say I am just giving an example. You are like explorer, you went to the seashore and collected some cone shell that basically has some potential to be used as an anti-venom, right, drug, antidote. So now you identify and you are doing some bioassay, you find it out probably this is a good compound. But what is the structure of that? So that you have to purify that compound, find it out the structure and that is your lead compound. Then you can do again more chemical modification that we were discussing.

So essentially the process which takes many years, many failures and much uncertainty, for a better medicine, basically it is an iterative improvement on the current medicine. You find a lead compound, then you can go for like improving it, may be your improvement whatever you thought is not working, so it is a failure, again you come back, start modifying somewhere else and then probably this is also not working, so you do third time. So many times you have to do it and that takes time, it takes money and there is an uncertainty, but you do it because you want to make an improved quality drug on the current medicine. These improved medicines offers a benefit over existing medicine in terms of like they are more potent, you require a minimal dose, they are more safe, they are tolerable and easy to like administer to the patient, convenience is there. It has less side effect right, it produce less toxic output, you know body system it can be easily cleared. So all those consideration constantly goes in the mind of researcher when they want to develop a drug.

So you start with a hypothesis and you design a molecule, you deduce a molecule using say chemical intuition, various combinatorial approach, molecular modeling or de novo methods. You synthesize these molecules, you test these molecules. After testing, you develop an bioassay and assay which basically identify whether it is happening or not and then again, once you identify this, establish some bioassay. If it is giving some bioassay, again you do lots of test by chemical intuition, by modification, by intelligence, by machine learning all of these required. And then again you learn many things and then you again go for a hypothesis, modify this drug and finally when it is coming good, you end the process. So therefore if you look at, this is not an individual work, this is totally a teamwork where a basic chemist, the medicinal chemist or a biological chemist, the biophysicist, the machine learning guy, computational guy, optimization guy, chemical engineer, lots of people are involved. In the drug development, then bioassay person has to be there, then efficacy person, toxicology person has to be there.

So, this is for development, after that finance and marketing and selling and doctors are required. So essentially drug development and design is a teamwork, a totally totally teamwork, right. So it is a complex work. We are like we as a basic scientist, a chemist or biologist play a small role in developing a drug. So, this is small, this is very very important because you are working towards betterment of humanity, healthcare and all those.

So historically how drugs were developed, right? So it is all by luck by chance you can say. It is primarily empirical method where you investigate a substance or some preparation from material such as a plant part, you took it, extract something. And then you given like

by some knowledge from your ancestor, you find it out, okay this molecule is good. Like if you have grown up in village, we all know some medicinal property of some plants, right. As a kid we were knowing, we were playing, breaking some like or we create some wound, we will take some grass, rub it, put it on the wound and that helps, that was helping.

How this was coming? This was coming because we were knowing this knowledge in Indian village from our ancestor. If you do this, you will be cured. Now that is the typical empirical methods, right. So people have investigated the medicinal property of some of those plants or plant extract or even the animal extract found in the local environment and that were typically used for some rudimentary medicine, okay. But when you are doing a systematic drug design, you have to isolate those compounds, you have to purify the active material in this.

So suppose I am taking a plant like leaf and rubbing it, we get some decontamination or whatever the extract of that, use that is used for the healing your wound. But that is not an active compound because if you want to scale it up, make many of the tablets or whatever, you have to find it out which molecule in that state was active principles, active pharmaceutical ingredient, API. So that once you identify that then it can be optimized and scaled up. So, like see morphines were typically empirically it was identified as a pain reliever.

Like a Hoyoscyamine, stomach and intestinal cure. Quinine for years it was used in malaria; then digitoxin for heart failure. Many of such compounds were typically used for curing some medicine and this came from the empirical methods. So chemical structures of these naturally compounds were determined later on and then find it out that these are actually active compound.

And now, taking this knowledge one can modify, start modifying it. So say let us take an example, this quinine right. Even in childhood I was learning that the, when Britishers were ruling India, they were giving tonics to whatever the soldiers and tonic has this quinine. Why? Because you saw India has lots of forest, grassland and all those and there are lots of mosquito. So these soldiers were getting actually sick.

So they developed a technique where they were mixing this quinine and giving to the patients and that were curing, right. So and even in the Indian village people were giving quinine. So they basically learned from the village Vaidya that this has some medicinal effect and that is how it started. So once we know this compound quinine, one can modify it. So biologically active natural product that is structurally modified in order to optimize their pharmacology and drug property.

One of the modification you can say is hydroxychloroquine. So some chemical modification came, lots of test came and then this found to be very good for fever and all those. So this opens a new dimension and opportunity for deeper understanding of the cell biology, genetics and bio-structure and one can make a better analog starting from the natural compound which was like a quinine or so. So they started with a basic curiosity, can these compounds treat something, that was the original idea of any of the person who started. Once they were saying that okay, this is doing something, it was extracted, their structure was determined, the bioassay was established, then medicinal chemistry was done, made it better, so new dimension and opportunity came for deeper understanding, the better understanding how this is happening, mechanistic insights, right.

Why it is important to know the structure and the binding property? One of the classic example that I can give you is this compound thalidomide, right. And why exact configuration is need to be known, this is the classic example of thalidomide. Actually many of you know that during pregnancy women gets morning sickness and all those of the problem. So the doctors were prescribing a sleeping agent which treats the morning sickness in pregnant women and this guy, this compound is basically was used for a sedative. Now later on what they found that the child that were born after this medicine was given to the pregnant women, they were having serious fatal abnormality.

Now, we want to treat the morning sickness, but what is the result? It is a disaster, right? Disaster because the child which are born, which are getting born or they have a defect. Now later on they find it out, this is because this compound exists in two configuration R or S. The R is actually sedative, but if you do not purify it, if you are giving S version of this thalidomide, actually this causes serious fatal abnormality. So not only identifying compound is important but identifying the correct configuration of this compound is very important if you want to take it as a drug.

Therefore intense research is very much important, right. You cannot give anything and everything because it has some effect and that is the necessity to do a serious precise research when you are developing a drug. And in those case, our technique NMR helps a lot. NMR based strategy can identify even weak binding compound and that can be used to develop into potent drug like inhibitor which can be used as a lead compound in drug discovery. Even after modification at each step, NMR plays a profound role.

So already we have looked at but just I will summarize what NMR can offer so that we are in the same platform. NMR can provide a structure of a macromolecule in a solution. So in cell most of the proteins are soluble. Their precise structure with a dynamic information needs to be investigated. And NMR has unprecedented power to do it, to give the structure at a high resolution as well as capturing the dynamic feature of that.

So, molecular or even supramolecular dynamics can be investigated by NMR. Not only this we have already seen, it can characterize even the weak interaction between the molecules like in the last week we see protein-ligand or protein-protein interactions. So what is intercalating like a water-LOGSY kind of experiment or STD in a micromolar range, we can characterize basically the weak interactions. So that is what we have protein-protein, protein-ligand interaction very well can be characterized. Not only for a soluble protein because many of the target of the drugs are receptors which are in the membrane system, membrane-embedded.

NMR can also provide structural insights into membrane embedded system, actually extracellular domain of the membrane embedded system. Basically, here is my membrane-embedded system, here is the extracellular domain. Most of the drug binds to this extracellular domain. Now NMR can investigate this extracellular domain in a profound manner and many people use this for finding it out. So taken together it can offer a lot in drug discovery and design. So, one can investigate the physical and chemical principles, property of the protein-ligand, protein-nucleic acid, protein-protein interactions and those are of primary interest in structural biology as well as drug discovery.

So if you want to know the structure you have to probe these interactions. Now as we are mentioning NMR can characterize the high affinity or low affinity macromolecular complexes. In the last week we looked at how we can do that. The only question if you are looking at the protein we need to label with N15 or C13, right. So that is what we need to

do it. And we can find it out binding of a ligand or macromolecule that basically alters the chemical environment around the binding site.

So when we are doing the binding of a molecule, we are looking in a residue specific manner for a labeled protein. If the molecule is interacting, it should alter something, what it alters? The chemical environment, where this ligand is binding. So here say my protein is like this and drug is binding somewhere here, so upon binding the chemical environment of the residue in the protein essentially changes. And that is what we can probe by monitoring the chemical shift of these magnetic nuclei at that site and in typically HSQC based experiment we can detect it. So, for that we need labeled protein N15 or C13 or both right and then we can do HSQC based correlation experiment to find it out exactly where it binds.

Before going to HSQC, I will just tell you if you are not labeling your protein is very good, well dispersed. You can even use un-labeled protein, record a high resolution experiment and you can find it out whether the ligand is binding or not. So here is a protein 1D spectrum, you can say this protein is very well resolved. So red line in absence of any drug, blue line in presence of a drug. Now you can look at resonance by resonance, this is the aliphatic region, these are the backbones, the Hα, here are this side chain of like aromatics and also the side chains like asparagine, glutamine, lysine and all those, and here are the backbone, the last one is tryptophan side chain.

So, this is the typical 1D spectrum that you have studied. Now we are titrating with a drug and looking what the shift is happening. So let us look at closely. Even in the aromatic you can see from red to blue, there is some shift. This one peak has become two peaks, right, in presence of drug. So what is happening? Basically, upon binding it is in slow exchange, right, last week we looked at.

Same peak is now populating to probably one with a bound, another with an unbound. Now again some of the peaks you see shifting, here tryptophan completely shift to a new position, here you can see it. Now shift is happening, even the side chains, backbone everything is shifting. So looking at the shift in the resonance frequency or chemical shift, perturbation in the chemical shift, you can identify the binding event between drug and protein. So this was 1D, you can do even more fun way by doing 2D. You can start with a N15 proton labeled protein, you titrate it and you see now few peaks are shifting here, few

of them are not shifting like here red and blue are same, but few are shifting here, you can see it in a zoomed version. So now, the shifting again tells that protein is interacting with a ligand.

More precisely, you can know where actually it is interacting. So that starts showing some kind of activity or some kind of SAR like a binding or so. So this is the 2D and you can identify even using 13C, using N15 both you can do. So here you can identify using 13C, if you are not labeling with N15, you can label with a 13C and identify the resonances that are shifting here. So this is N15 again, you can find it out. You can similarly do 13C like experiment. Here is a 13C axis, here is a proton axis, you can probe the 13C-HSQC, find it out the peaks that are shifting from here to here.

So N15 labeled or 13C labeled both of these can be used for identifying the interactions. Now these are typically small proteins of say 100 or 200 amino acids. What happens if the protein size increases significantly? If it becomes like 200 kiloDalton. Now what you know that line becomes broader and in those cases, it becomes extremely difficult to have a decent resolution.

In those cases, we have to do some biochemical treatment. So like here say it is a 200 kiloDalton protein. But in 200 kiloDalton protein there are some probes that are very nice to be like nicely to be exploited. Those are methyl groups. So you know 6 amino acid in out of 20 have a methyl, alanine, leucine, valine, isoleucine, methionine and threonine, right.

So those 6 can essentially be labeled by tuning the metabolic pathway of like in bacteria. So you feed selectively labeled compound to bacteria and when they produce protein, they will produce where methyl groups of these will be labeled only. Now methyl have a nice phenomena because of interesting relaxation properties, they give even sharper line. So you can use a TROSY-based experiment to exploit the methyl groups in the protein. Now if you look at the distribution of methyl, they are distributed all over the protein and they are at the end of the side chain, right.

So essentially they capture the compact or the tertiary structure of a protein. Now if you probe these methyl groups, we can essentially probe the structural features, the higher order

structural features of a protein. Now by selectively labeling we can illuminate this methyl group, so like here in cyan color we are illuminating one kind of methyl group, here in another kind of methyl group, in third blue is another kind of methyl group. So with this resolution now if we titrate with a drug we can find it out like which methyl groups are precisely binding to a drug molecule. So these are the some of the biochemical trick that we can do it and then using these essentially we can find it out the SAR. So suppose a lead compound is binding to a protein molecule at some position and now you titrate with this probe, using NMR you find it out precisely where actually it binds.

So it binds to a helix that can be illuminated. Now looking at which residue appear or disappear. So, here I am showing you example of some MCL1 protein that binds to a compound and you can see binding happening here. So, you find it out actual position of the binding by SAR in NMR. Now, using this information you can like a, you can do the NMR guided fragment based drug design. So, you can find it out where actually it binds, you looking at the nearby residue in the protein can start growing the molecule and that is called fragment based drug design.

So using all these information, you can basically grow your molecule and lead compound can be modified, adapting more groups, making more complex molecule and having more potent inhibitors. So this information is provided by NMR which helps you in a precise manner, in a profound manner to develop your drug. We will be continuing discussing all these aspects in the coming lectures. So here I end with the note that NMR gives you the residue specific information if we are probing on protein which are the residue that are binding and on the other side you know from STD-NMR we can find it out epitopes on the small molecule that are engaged with a protein interaction. Combining these two information doing the structure based drug design you can grow your molecule and then find a better potent molecule in the drug design.

So let us continue this discussion in the next class. Hope to see you how we can take this knowledge forward for a drug discovery and design. Thank you very much.