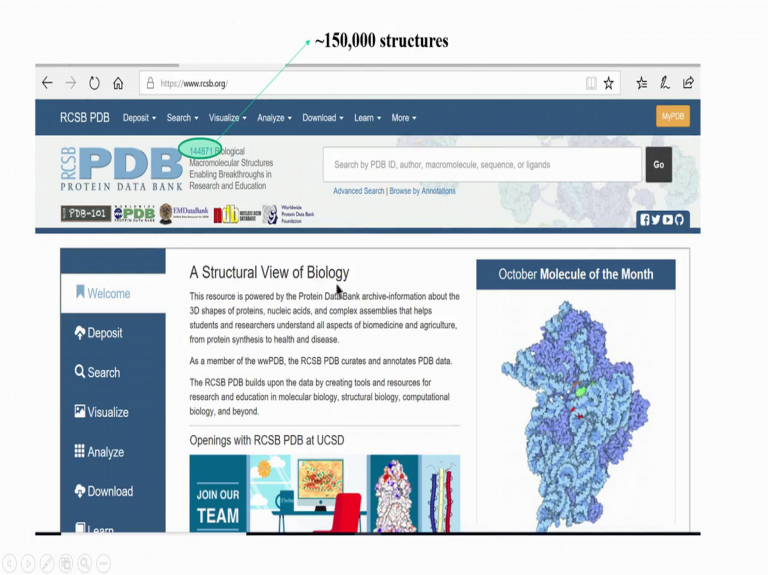
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture - 22
Introduction to Structure Determination of Bio Molecules by NMR

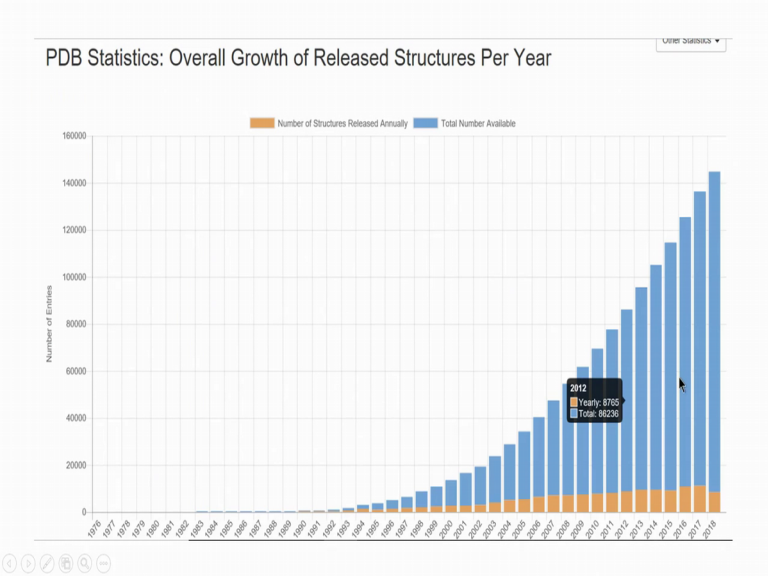
In the last class we were looking at protein databank. This is the PDB structural database where all protein structures determined by different techniques are deposited ok. So, this is shown here, this is what we looked at so, nearly about 150,000 structures are known today, by different techniques this includes both theory theoretical approaches and experimental approaches.

(Refer Slide Time: 00:55)



Now, one thing if you look at this database we can actually analyse. How many structures are available which is shown here or what are the different types of experimental structures and so on. So, very interestingly you can actually look at this here and analyse statistics which is shown here.

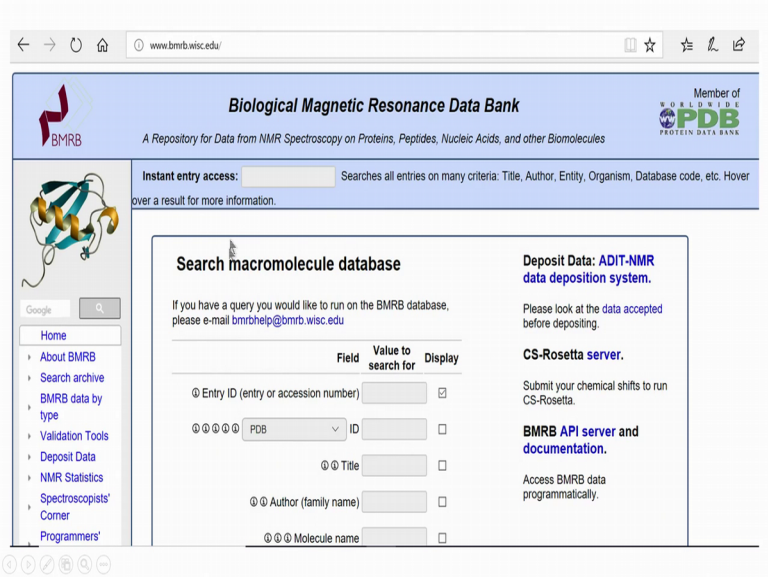
(Refer Slide Time: 01:11)



So, one of the statistics which is shown here is you can see in this graph. So, actually the PDB statistics is the overall growth of release structures per year.

So, we can see how exponentially it is actually increasing. So, every year the number of protein structures which are getting deposited or available is getting higher and higher. So, it is kind if exponential growth in the protein structure. So, this of course, is complete total which includes all the techniques, which I showed you in the previous slide.

(Refer Slide Time: 01:43)



Now, there is another database which one has to be aware of and this concerns only

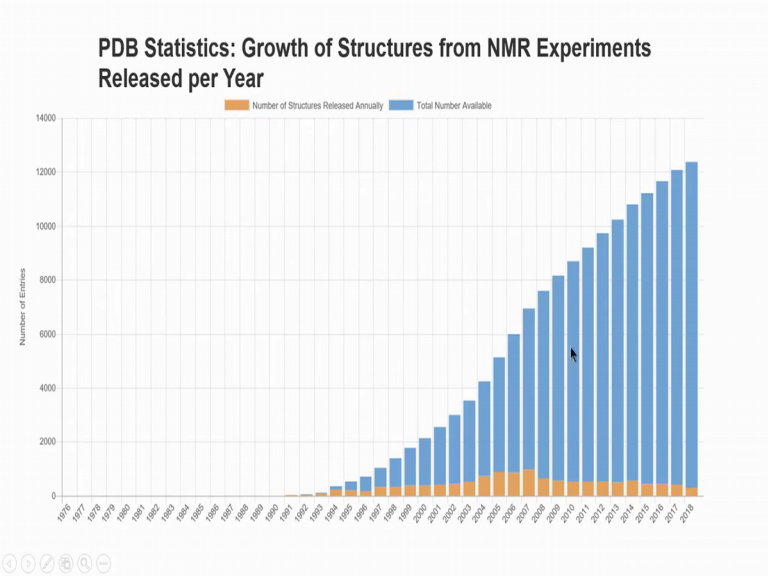
NMR database so, NMR data. So, this is called Bio Mag Res Bank BMRB for short. So, BMRB can be visited with using this link here this url. And, this is a very very important database for the NMR community. And this of course, is linked to PDB as well as shown here so; that means; if you solve any NMR structure or any structure of a protein by NMR you have to deposit the NMR data here.

And, the structure goes to PDB. And, those 2 data that is NMR data of that structure are linked to each other so, if somebody goes to PDB and he clicks on the protein structure and if it is solved by NMR, it will now bring you to this website where the NMR data of that protein will be available. So, this is good because you want to know how, what kind of data what kind of experiments in NMR were used for solving the structure of that particular protein, what techniques were used and what are the chemical shifts which were obtained for that particular protein.

Therefore, this is Bio Mag Res Bank or BMRB is a very popular and database for another NMR people. And, here again you can use different links here and look at statistics look at how many structures have deposited and you can deposit your own data and so on.

So, again I would recommend strongly to visit this website just have a look at this try all these different links here and you can see what you get from that. So, that is a good exercise for a new person who is new to NMR, especially new to biomolecular NMR to have a look at this database.

(Refer Slide Time: 03:37)

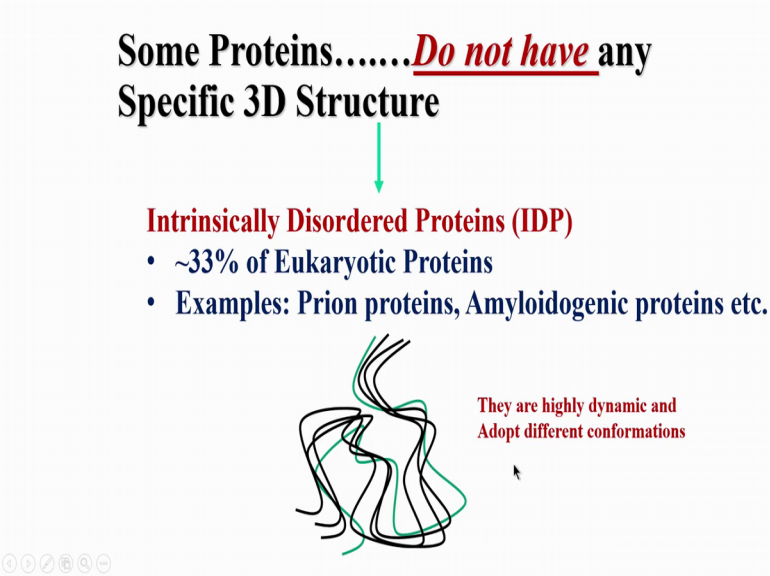


So, moving further on the PDB statistics this is the statistics for NMR structures. So, you can see here also the growth is there, but is although it is not exponential, but there is a growth steadily over the years. The NMR structures which are deposited now roughly about 12,000.

So, this is about less than 10 percent about 9 percent, 8 percent of the total structures which are available, support yeah suppose less than 10 percent. So, this is what basically shows because NMR remember X-ray crystallography is one of the very popular techniques and older much older technique than NMR. And, therefore, the technique has now been refined and now it is pretty automated compared to NMR. NMR still has lot of manual work to be done and required whenever you solve a protein structure which we will see now shortly and it requires sample preparation and so on.

But, in a crystallography what you need is a crystal. Once, you have the crystal of a protein then getting the structure is not. So, time consuming of course, it requires skills and people who are very skilled in this can do it faster, but compared to NMR techniques X-ray can be fast. So, NMR is now growing slowly and the higher and larger protein structures are being solved because of improvements in the technology.

(Refer Slide Time: 05:01)



So, one protein some protein which I have mentioned earlier, some proteins have specific 3 dimensional structure, but there are some proteins which do not have any structure at all and this is the class of proteins which I will look at it now we look at it this now. So, these are called intrinsically disordered proteins, in short they called IDPs. And, this have really come up in recent years, because it have been now found that nearly 33 percent of Eukaryotic Proteins. Eukaryotic means which includes human beings mammals higher organisms, then this kind of organisms of almost one-third of the protein are actually intrinsically disordered, means they do not have any specific 3 dimensional structures.

The examples of this protein which can create problems these are prion proteins, amyloidogenic proteins. So, on, but one should keep in mind is not that all IDPs are problem case. I only listed the disease causing proteins which are IDPs, their IDPs itself are very useful in our body, because once if you do not have a structure they looks something like this, which does not have any alpha helix beta sheets, which means they can actually bind multiple partners.

So, they are available multiple sites where they can bind compared to a rigid structured protein which can only bind to one particular ligand. So, IDPs are very now well established well known that they can have multifunctional roles and they play very important role in our body.

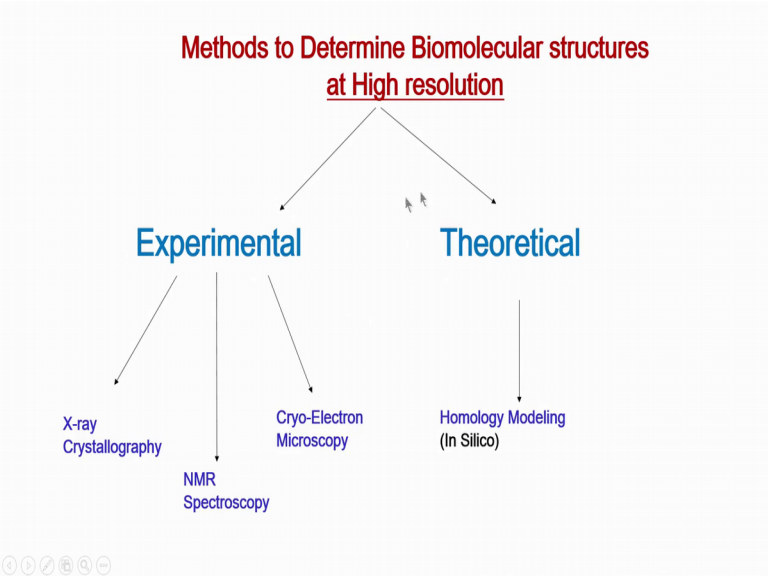
A second thing is the IDPs are actually can take adopt a structure if they bind to the

ligand. So, for example, some IDPs are known to actually adopt alpha helix or a beta sheets form once they bind to a ligand. And, what could be the ligand? Ligand can be a DNA. For example, DNA binding protein, and DNA binding protein can be an IDP, but it becomes structured or adopt some structure local structure of course, not the entire protein, but the part which interacts with DNA starts becoming more rigid and that could happen also in many cases.

But, in many other cases it just binds to the ligand and there is no change in the structure, it remains flexible. So, that is basically the point here in IDP is that IDP are very dynamic. They are highly dynamic in nature and they can adopt different conformations. And, that is what is the beauty of IDPs they can adopt different conformation again remember confirmations is not same as structure, conformation is the local geometry, or orientations of the different group and chains, structure is something when we refer to we talk about alpha helix beta sheet and so on. So, it can adopt different conformation and therefore, IDP is a very dynamic.

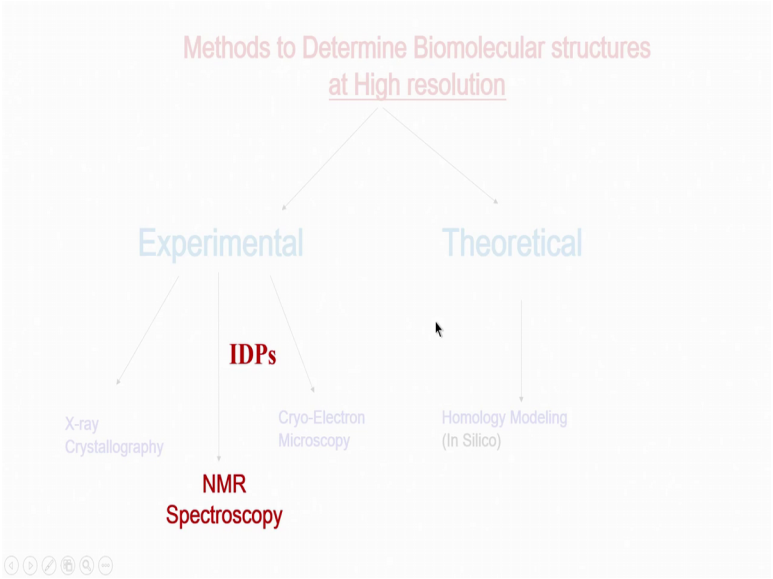
Now, given that there is no structure possible for an IDP means a well-defined, well-defined 3 dimensional specific structure which is written here. Then you cannot use X-ray crystallography, or theoretical methods. Because how do you use X-ray crystallography when there is no structure. So, when we talk about IDPs we do not talk about the structure, because they do not have any particular structure.

(Refer Slide Time: 08:21)



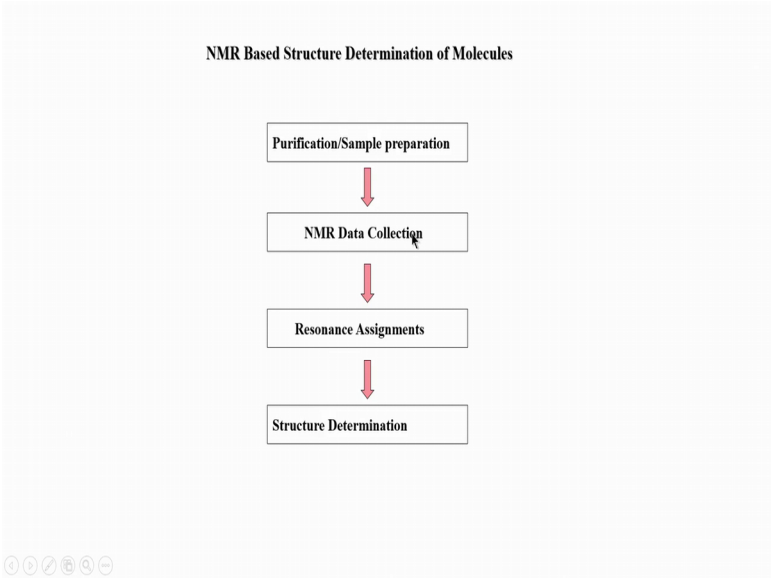
So, therefore, if you look at the techniques which we looked at for different for solving biomolecular structures, the only technique which we can use for IDPs is NMR.

(Refer Slide Time: 08:25)



So, this is where NMR stands out as a unique technique for solving the structures of IDPs. Because, there is no structure sorry first looking at the IDPs how they interact with their partners, how they interact with their ligands and so on. In fact, many of the drugs nowadays are targeting IDPs. So, therefore, it is important to know from NMR how do we understand, how do we look at, the interaction of an IDP with ligand. And, that is something can be only done with NMR, because IDP is inherently lack 3 dimensional, specific 3 dimensional structure.

(Refer Slide Time: 09:05)



So, now coming having come looked at how nm different protein structures can be solved by NMR. Now let us go in to detail of how we solve the structure of biomolecules by NMR. So, this one flow chart is something which was shown earlier in the previous lectures is that, if you want to solve the structure of a protein by NMR you need to follow these steps.

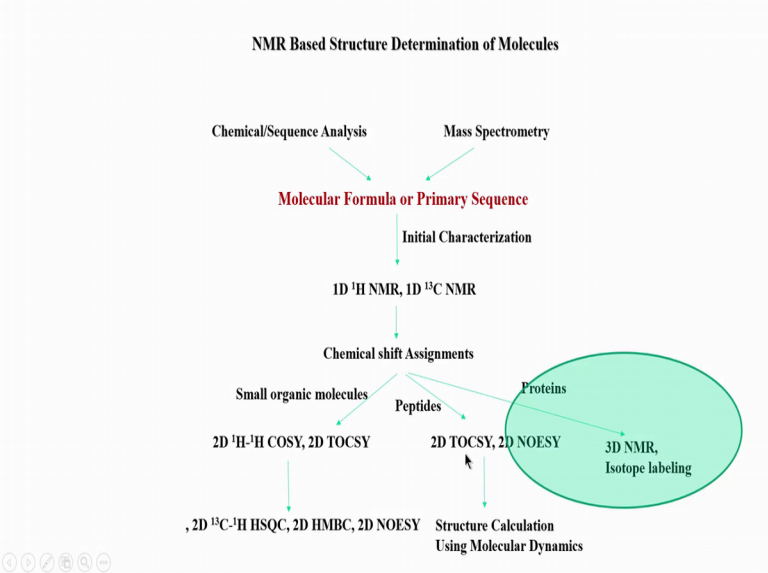
So, typically we start from sample preparation, in the case of proteins this involves isotope labelling which we will see in detail in this part. Then once you have the protein sample prepared, we go to the NMR spectrometer and record data. This includes 1 dimensional, 2 dimensional, and 3 dimensional data, for proteins. Again remember if you want to do 3 D NMR or 4 D NMR you need to have the protein isotopically labelled. At natural abundance is not possible to do 3 D NMR or 4 D. Then once you have the data that is 3 D NMR or 2 D NMR.

We need to now assign, we need to carry out this important step here. And, this is something which we will spend some time down the line, because that is a major part before without resonance assignment meaning, without assigning the chemical shifts of a protein we cannot go to structure. So, structure determination requires this as a first prior step. So, that is, what is the next step.

And, then once you have the assignments we can then start solving the structure of the 3 D structure of the protein and we need a very important experiment here called as

NOESY. That is something we looked at, in the previous half of this course in the very beginning and that is again we will look at it in a 3 D manner if required for 3 D proteins structure.

(Refer Slide Time: 10:59)



So, now let us see overall broad picture of how a protein structures are solved or a molecular structures in general are solved by NMR. So, this is the general schematic for any molecule not just by a molecule. So, typically for organic molecules we do chemical analysis. And, we do mass spectrometry and we get the idea of the molecular formula. We can do the same thing with protein of course, in protein as I said we need to know the sequence of the protein ok.

So, the primary sequence of the protein should be known, without that we cannot do any NMR structural studies. So, this is something which many do not grasp the gravity of this is very important point. And this how do you get the sequence? Sequence can be done by varieties of methods. Typically once you know the DNA of the where it is generated the protein in the cell, then from the nucleic acid that is from the DNA you can actually know the sequence of the protein.

So, protein sequence once it is known and if you know the mass of the protein, example from mass you may know whether it is a multimer monomer so on, you can now generate a molecular possible molecular formula for the protein or a molecule. So, this is what is written here that you need to know the primary sequence for the protein case, or

you need to know the molecular formula for a small protein for a small molecule organic molecule.

Now, once you have that information we can then do the initial characterisation and that is carried out with the help of 1 D NMR. And, this is something which all of us are aware of how it is carried out you simply record a 1 D spectrum of protein and take case of organic molecules we also record 1 D carbon 13 NMR, but in protein we normally do not record this experiment for a protein we normally record only this.

Now, based on 1 D proton NMR one can get some information on the protein. For example, you can look at the stability of the protein, you can see if the protein remain stable for long time. Because remember 3 D NMR 2 D NMR takes long time to record some 3 D NMR takes few days of NMR instrument time. So, you cannot put a sample for few days. So, you have to therefore, keep it for a short duration.

So, you have to know whether this protein is stable or not for that duration if let us say you have to record for 4 days 3 D NMR you should be able to know, whether the protein is stable. So, typically initial characterisation we use 1 D NMR to see whether protein is stable whether it is degrading with time and so on. You can also use 2 D for that 2 D NMR.

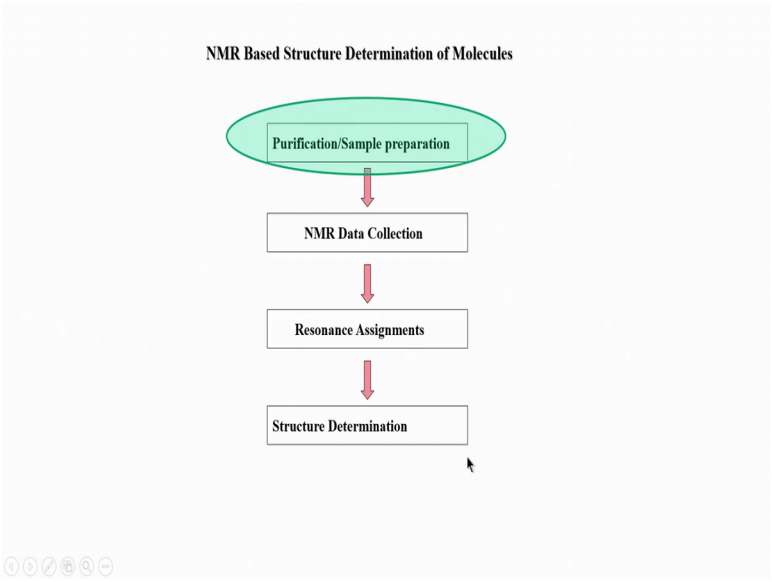
Then once you have this experiment of course, here is written chemical shift assignment, but we need to also record multidimensional which is not written here. It is followed after 1 D NMR we do 2 D 3 D NMR and do this resonance assignments which I already mentioned in the previous slide. And, that is done with 2 D COZY and TOCSY for small organic molecules. And, you can also use HMBC HSQC and NOESY and this part was what was covered in the previous course in NMR.

But, when it comes to biomolecules for example, if you have small peptides, then the chemical shift assignments can be done with just these 2 experiments TOCSY and NOESY 2 D. So, here there is no need for any labelling. So, we are not doing any labelling of a peptide, it is a natural peptide, we would have extracted, or synthesise in unlabelled form and that be used. And, this can be solved then you can look at the structure of those peptides by using computers modelling that is molecular dynamics.

So, when it comes to proteins there as I mentioned we need to do 3 D NMR, and isotope

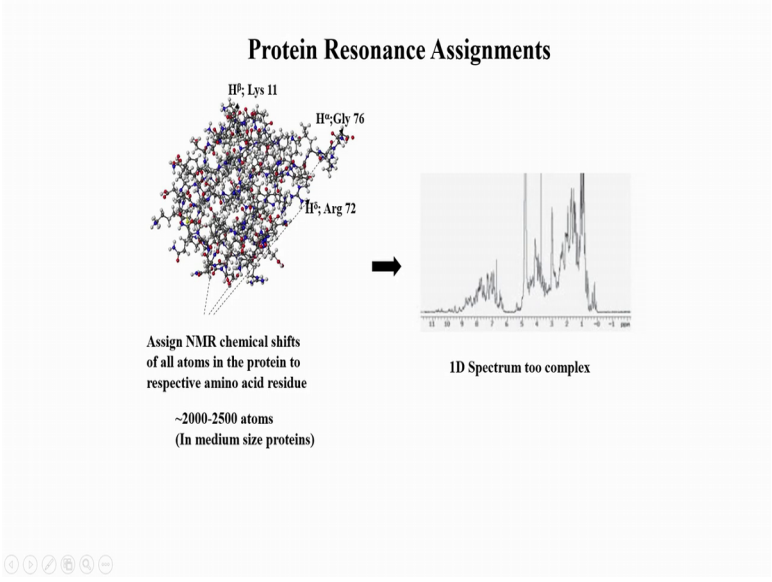
labelling along with that of course, as I said 3 NMR need isotope labelling. So, we need to somehow label the protein with carbon 13 and N 15 and then carry out 3 NMR. One of the questions is why do we need 3 D NMR here and here we do not need is because of this problem of spectral overlap ok.

(Refer Slide Time: 15:13)



So, let us now start how this part from purification of the sample, how do we actually make a sample and this is where we begin our isotope labelling part.

(Refer Slide Time: 15:27)

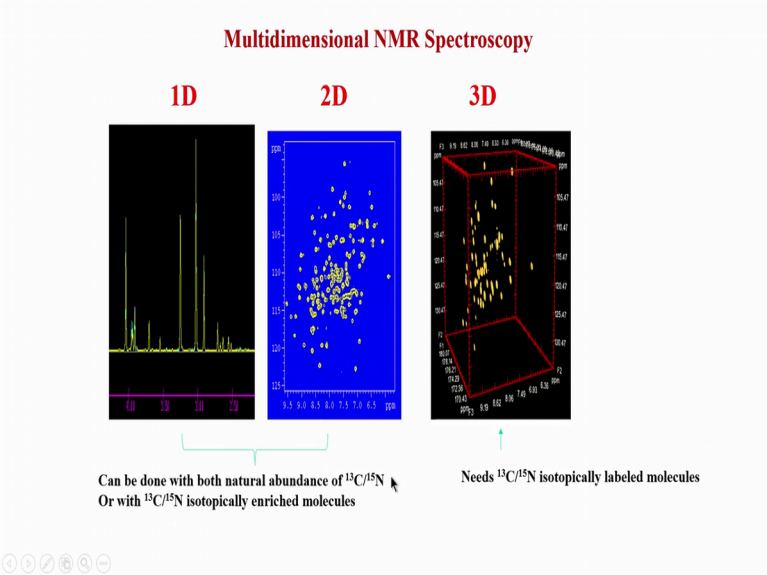


Now, the question is why do you need to do isotope labelling and that is what is trying to

show shown here. So, as I said the assignment when we say protein resonance assignment, what does it mean? It means for every atom in this protein as much as possible of course, not possible for every atom or for every hydrogen here the example H delta arginine 72, or H alpha glycine 76 this is the ubiquitin protein structure or H beta of some lysine 11, you need to actually know the chemical shifts. That means, we need to assign NMR chemical shifts of all atoms in the protein to the respective amino acid residue. And, that is the key here, you want to get the chemical shift of every atom and this could become very complicated because you may have over 2000 atoms in even a smaller or medium sized protein.

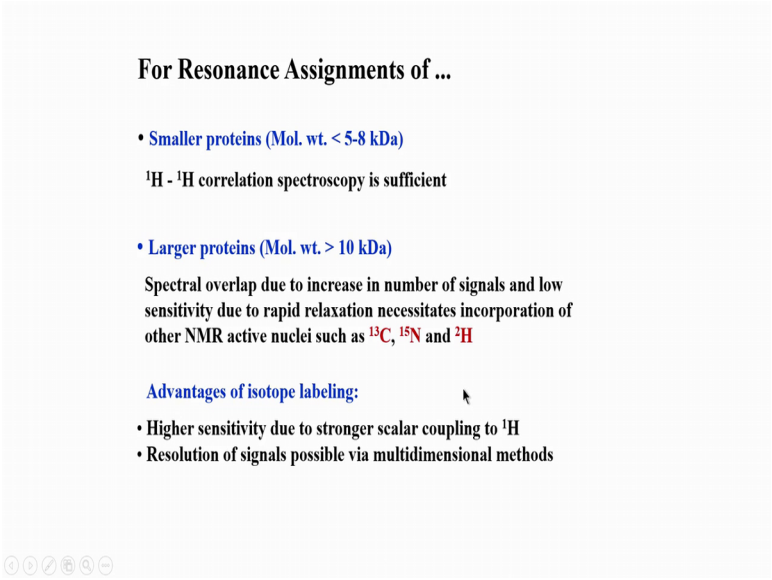
So, in such a case you will get a spectrum which is 1 D look very complex so; obviously, 1 D NMR cannot be used for getting any information on the assignment of each and every atom here, because they are all overlapped.

(Refer Slide Time: 16:39)



So, that is why we need to go for multidimensional NMR spectroscopy and this is something we have covered in the first part, we looked at different 2 D NMR experiments that can be used, we have looked at different 3 D NMR experiments that can be used, and as I said they need labelling in this case 3 NMR, but for 1 D and 2 D you can still do that without any labelling and we can do with simple natural abundance of nitrogen and carbon.

(Refer Slide Time: 17:07)



So, now let us see why when do we need labelling and when we do not need labelling? Isotope labelling in terms of the size of course, for every 3 D NMR we definitely need labelling, but if you see look at smaller proteins that is which have roughly 5 to 8 kilodaltons. We can see here what is the number of amino acid, if it is a 5 kilo Dalton protein is roughly about 40 45 amino acids. So, if you have a small protein which is almost like a peptide, you do not need labelling, you do not need carbon 13 or and 15 3 NMR experiments. You can simply do away or get the structure solved by only proton proton correlation that is 2 D NOESY 2 D TOCSY etcetera.

But, as the protein becomes larger in size for example, look at this here it could be 100 amino acid, with 90 amino acids, these are slightly getting bigger and bigger then you have spectral overlap. Due to increase in number of signal, because is a signal comes in the same ppm values. So, I have now more signals, because I have more number of amino acids.

So, that will; obviously, result in a signal overlap. In such cases I have to resolve the signals in separate them and that can only be done with 3 D and 2 D NMR. So, this is why we need to now start labelling the protein with different nuclei. So, for a medium size suppose I have up to 20 kilo Daltons, we can do away or we can be sufficient with C 13 and 15. And, when you go to 25 kilo Dalton and bigger size, then you need to also label with another nucleus called deuterium. And, this is very special specially done this

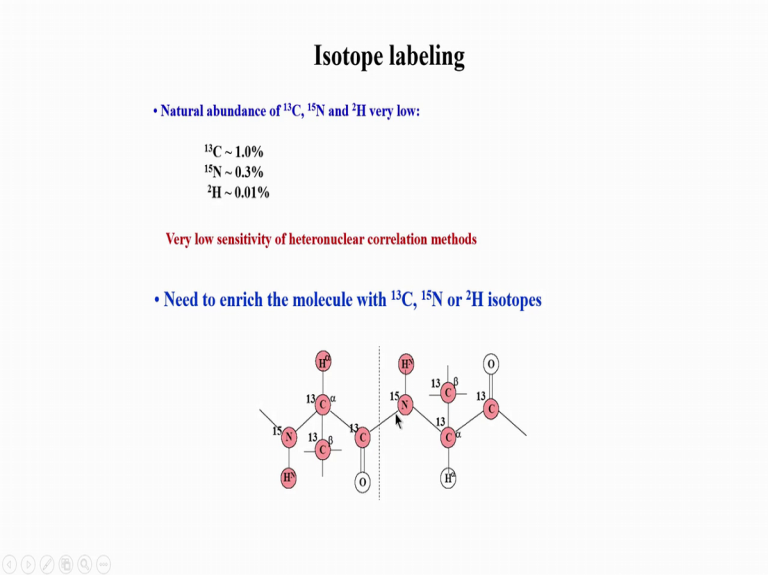
is not simpler or similar to this and that requires different type of procedure which we look at.

And, what are the advantages once we label this isotopes within the protein. The advantages are these 2 written here; one is the higher sensitivity. From where does higher sensitivity come, it comes because we have very strong coupling to proton. So, remember carbon proton coupling means how much is typically about 140 hertz, which we saw in the 3 NMR part.

Similarly, N 15 proton coupling is 90 hertz; you can compare this with proton coupling. A proton to proton coupling is about 6 hertz, 10 hertz, or maybe 15 hertz, but now we are looking at 140 hertz here and 90 hertz here. So, because of a stronger scalar coupling or j coupling to proton it is possible to design NMR experiment which are very sensitive if you label it. For example, the experiments we saw is HNCO, HNCOCA and so on which we saw in the previous class is a very useful because the sensitivity of that is very high.

A second advantages now with the 3 D experiments which we again in the multidimensional triple resonance experiments, we can start resolving the signal. So, when you start resolving the signal, the signal becomes better and better we separated and that helps us to look at the chemicals more easily than in the 1 D case or 2 D case.

(Refer Slide Time: 20:13)



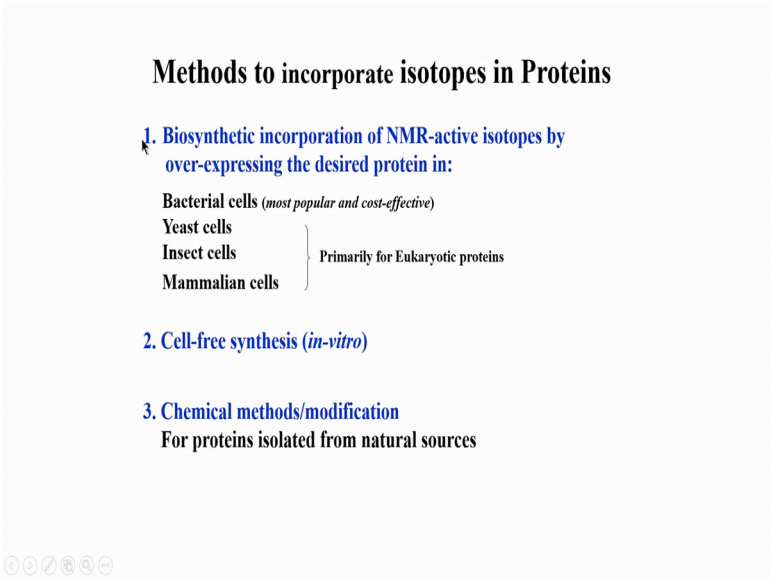
So, now again on the reason why we need isotope labelling is that look at the natural

abundance of these nuclei is shown here. So, you can see carbon 13 we know this is about 1 percent, natural abundance and nitrogen 15 is even less, and deuterium is even much less than both of these. So, you can see that the natural abundance we cannot work at natural abundance with these nuclei, if you want to 3 D NMR because of the low sensitivity they are very less in abundance. So, therefore, we need to actually enrich meaning make the protein rich in C 13 and 15.

So, basically what it means is that we need to convert. So, you can see here every carbon and every nitrogen now is shown as N 15 and C 13. So, this is what we need to generate. So, this is a dipeptide, this is the protein part of a protein. So, the protein runs from this side n terminal to C terminal. And therefore, it has now become I mean they have shown for a dipeptide. So, we need to label everywhere C 13; that means, I have to convert this molecule normally, they can naturally in all molecule will exist in C 12 N 14 as the original natural abundance, but now I have to convert this molecule in to C 13 and 15.

So, how do I achieve this conversion, how do I convert a protein molecule, or how do I not convert, but I do I generate a protein molecule, which will have labelled in these positions? So, this is the whole topic of isotope labelling which we will now look at.

(Refer Slide Time: 21:43)



So, what are the different methods to incorporate isotopes in proteins? So, there are different methods with verities of methods. So, number must most important category the first method approach is what is called as biosynthetic incorporation of NMR active

isotopes by over expressing the desired protein in bacteria, over expressing means expressing the protein.

So, our protein which is enough interest should now be produced by bacteria, or east, or insect, or mammal. So, you can see there are different types of cells which can be used. So, most of the time, most popular is bacteria. And, it is because it is very cost effective is very easy to produce bacterial to grow bacterial cells in the laboratory in a large scale, and bacterial cells also multiply very fast. And therefore, it is very convenient to make protein in bacterial cells; that means, our protein which we want to study, we have to somehow express in bacterial cells we will see how we do that, but many proteins are not it is not possible to study with bacterial cells.

For example; let us say you are protein which you are interested as what is called post translational modification. Example it undergoes it functions only if it glycolytic group or a glucose group is attached to the protein. How does how bacterial cells do not have that machinery in the cells. Bacterial cells do not cannot do post translational modification to your protein.

So, therefore, you cannot use bacterial cells to make your protein. You have to when go to higher cells, higher eukaryotic cells. And, that is yeast or insect or mammalian cells, that is why I written here this kind of cells are mainly used for eukaryotic protein, but you have to keep in mind that as you go to these cells the cost becomes exorbitantly high, or prohibitively high is very high very costly to make proteins in this mammalian cell for example, if you want to go C 13 and 15. So, very few laboratory lab labs in the world actually do this. In the most popular as I said is bacteria or yeast.

So, we will look at bacterial cells in more detail, we will not look these in more detail, you can refer to this book on isotope labelling which I refer showed you in the very beginning where these are covered in that book. So, for this course we will focus mainly on different isotope labelling methods, which we can do with bacterial cells, now this is one approach. So, what are you doing we are using biologically we are expressing the protein which we want in a cell; that means, it is a live cell you are making the protein in the cell.

The second approach is we can get rid of the cell, we can do what is called cell free synthesis of a protein again this is for isotope labelling. So, cell free isotope labelling of

proteins. And, this is the in vitro approach why in vitro because here it is inside the cell; we are doing the making the protein here is outside the cell; cell free. Now, how do you do that outside the cell? The main thing is to keep in mind is remember we need what do we need for a cell to be made. So, we saw the central dogma of biology that a DNA is transcribed into RNA and RNA is translated into a protein.

So, now suppose I take the DNA of the protein, the desired protein I take the DNA part. And, I transcribed into a into RNA and into ribosome, I take a ribosome. And, then put all these enzymes needed to make the protein. I do not really need a living cell, a real cell to do that. If, I can take that basic machinery which is available which is needed for the protein synthesis, which is basically the DNA. Then followed by transcription to an RNA followed by translation to protein these are the 3 steps; I can do that without any cell, I can take those enzymes required for those conversions, and I can do it in a test tube that is why we call it in vitro outside the cell.

So, therefore, if I take those machinery from the cell we did not really need a cell at all we can do it outside the cell. Now, what is the advantage of this the advantage are many for example, many proteins are not possible to express in bacteria cells, because they may be toxic to the cell and so on. In such cases I can simply go outside the cell and take the machinery of course, I need those enzymes which are needed to convert DNA to RNA. I need those enzymes which cannot RNA to protein and a ribosomes all those can machineries those those basic things, but that can be obtained and mixed together and without actually a living cell a real cell we can do that. So, that is the concept of cell free synthesis again very popular, but those type of proteins which cannot be expressed in bacteria.

And, the final of course, method is you have to chemical method, which is basically mainly the synthesis part. Chemical synthesis of peptides and proteins can be done using solid phase peptide synthesis, but that typically we do not go beyond, we cannot go today beyond 50 amino acids. Typically, it is done for smaller peptides like 20 25 amino acids, beyond that going for 40 or 50 amino acid chemical methods become very costly, and time consuming. Because you have to now purify a desired protein from the mixture you will generate a mixture of amino proteins with different chemical modifications. So, we have to purify our sample that becomes complicated if your protein size is large.

So, this is for essentially this is something which we cannot do labelling for example, if it is coming from a natural source, there is no way I can express that protein in bacterial cell, I am extracting directly from nature. So, for such kind of proteins we have to basically use chemical methods to purify and then like using HPLC and then directly go for NMR.

So, labelling may not be possible then in that case. So, isotope labelling essentially is restricted to these 2 methods; it is biosynthetic approaches, which is outside the cell and cell free synthesis sorry this is inside the cell and cell free synthesis which is outside. So, again in this particular course we will be focusing mainly on this part, bacterial cells. So, in the next class we will see how we can use bacterial cells to now express the protein of our interest and carry out the desired isotope labelling.