Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

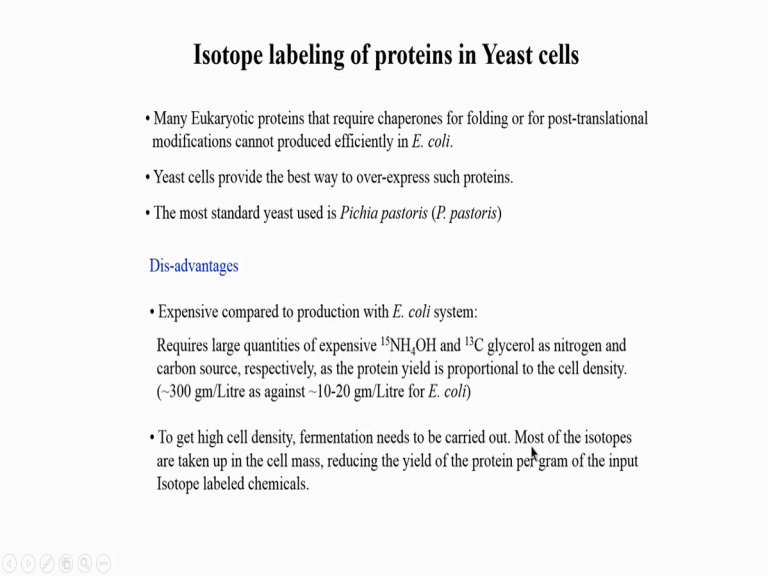
Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 27
Isotope labeling of proteins for NMR studies – Part IV

In the last class we started looking at isotope labeling of proteins in bacteria and we completed that. So, let us move on to the next option for isotope labeling this is done in yeast cells as shown here. So, what happens is many of the eukaryotic proteins for example, proteins which are made in human beings or higher mammals they may sometimes not be expressible not be possible to express in bacteria.

This is because many proteins in our body require what is called as post translational modification. Meaning after the protein is expressed in the cell after it is translated into a protein chain it has to be further modified by adding a some glucose group or some kind of a modification to the protein sequence. Not in the sequence level, but at the structural level for example, disulphide bonding and so on.

(Refer Slide Time: 01:25)



So, such kind of post translational modifications cannot be done in E. coli. So, this is what is the written here. So, therefore, you have to go to other than E. coli. So, what is the next possible higher organism? Yeast, yeast is a eukaryotic cell. So, we can use yeast now to express our protein. Because yeast has all the machinery, all that possible thing

which can be done with E. coli. So, yeast cells provide the best way to over express such proteins ok.

So, we have to choose a particular strain of yeast like in bacteria in E. coli. We chose BL 21 DE 3. In yeast the most popular strain which is used is Pichia pastoris. So, we will see now this will be a very brief introduction to yeast cells in the sense isotope labeling. We will not going to detail as we did for bacteria. So, these are the reasons why we go for E. coli, but this is for yeast, but there are several disadvantages compared to E. coli.

So, as I said E. coli is always the best possible system if we can do that. The following disadvantages are there for yeast. Number 1, it is very expensive compared to E. coli; that means, to make labeled isotope label protein in yeast it requires a lot of effort and time and money. For example, here is written so as written here requires large quantities of expensive ammonium chloride or hydroxide as a nitrogen source glycerol is expensive carbon 13 labeled and that has to be used and they can see there is a lot of amount is used.

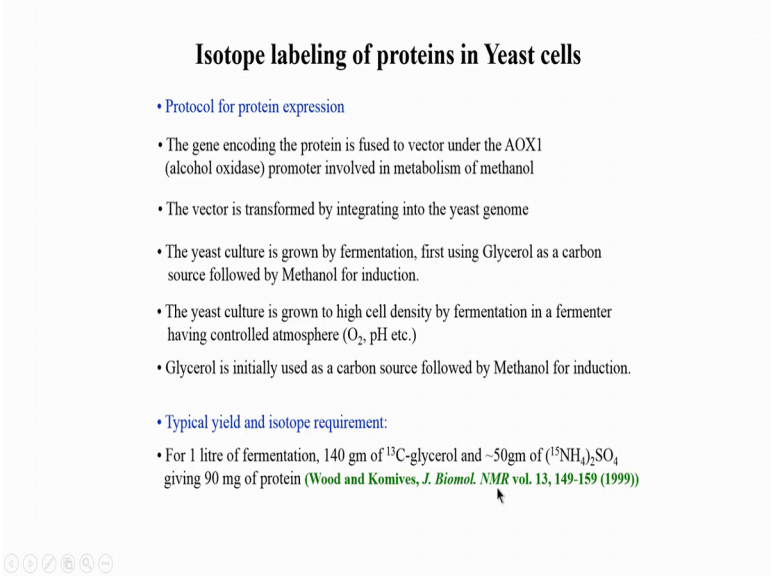
For example, we have to use 10 to 20 grams in E. coli, but we use about 30000 grams per litre for the yeast. And why is that we use this large amount for yeast? Is this because the cell density, the amount of cell which is grown per litre is also very high in yeast. So, when you have more number of cells you need more amount of nutrients. So, that is the reason why do we need high amount of this labeled materials.

Now another thing is if you want to get high cell density then you have to use what is called as fermenters or we need to ferment. So, yeast is then grown by fermentation. So, what happens is most of the isotopes that is N 15 and C 13 what we supply are taken up in the cell mass. That means, the yield of the protein per gram of the isotope input is reduced. For example in E. coli we use 4 gram per litre and that the entire 4 gram is efficiently use per gram.

But here the per gram yield that is per gram of how much you use and how much you get protein out of it efficiency is not very high in yeast. So, therefore, you are using more amount of labeled material and you are getting less of the protein proportionately. So, therefore, it is not considered very efficient, but of course, they are ways to optimize it various ways to improve it also, but overall compared to E. coli yeast labeling is

definitely expensive approach, but as I said we cannot avoid this in case we have proteins which cannot be expressed effectively in E. coli.

(Refer Slide Time: 04:51)



So, now let us go little bit further what is a protocol that is used for expressing protein in yeast cells. So, less it is similar to E. coli we need a gene encoding the protein of our interest. So, whichever protein you want to study that gene of that protein has to be prepared synthesized and it has to be fused means it has to be put in a vector a specific vector which is at particular vector is given here you can take some other vector depending on your a project and that is now put it into the cell ok.

So, this is similar to what is that vector is transformed by integrating it into yeast genome. So, it is unlike the plasmid here we are integrating into the yeast genome. So, in bacteria we did it in a plasmid cell a plasmid DNA ok. Now the yeast culture is then has to be grown. So, remember what is this is typical like the recombinant DNA approach what we do; we are basically taking the DNA of our interest and we are putting in into the cell like we did in E. coli.

And then we have to grow the cell because the cell has to may cover protein. So, for that it has to grow and it is grown in a particular medium where in this case. So, in the case of yeast glycerol is used as a carbon source ok. In the case of E. coli we use glucose as a carbon source here glycerol is used. And for induction that is remember induction is the set is a point where actual production of our protein starts in the cell. Till then the cell is

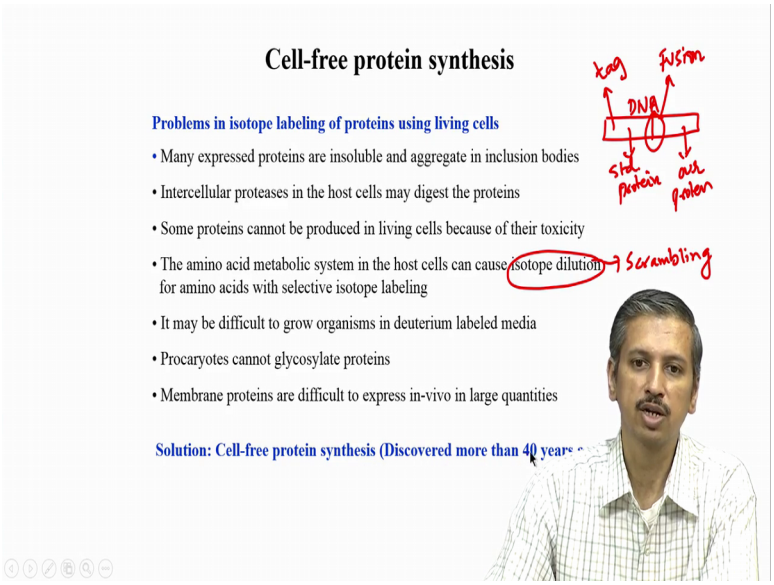
basically growing and then at some point we induce the production of our protein by adding some inducing agent.

In case of E. coli the BL21 DE3 cells we saw the standard inducer is IPTG whereas, here in yeast methanol is used for induction. So, now, the yeast culture is grown to verify cell density ok. So, where it becomes a very thick density and that is achieved by fermenter in a fermenter. So, we ferment the system fermentation is carried out in a particular controlled atmosphere of a particular pH, oxygen etcetera. And then that is now the whole cell density is further the cells are now collected and then harvested and used for further processing.

So, here as I mentioned here again is a glycerol is initially used as a carbon source. Then it is changed to methanol at the time of induction. So, after that induction methanol takes over. Basically you need some carbon source and that should be the only source of carbon there should not be 2 or 3 different sources of carbon. So, you use one source depending on a requirement and then change over to methanol for induction. So, what is the yield? Yield means finally, how much protein do you get the yield of protein is very high in E. coli in yeast complete E. coli. For 1 litre of fermentation we use this much amount.

So, it is a huge amount be of course, use of carbon labeled glucose, glycerol, and ammonium chloride or sulphate, but the yield we typically get 90 mg which is huge. for a for E. coli you will typically get 10 times less, but of course, in E. coli we also use less of the glucose and ammonium chloride. So, this paper is there is a details the typical protocol which is used for yeast. So, therefore, as I said we are not going into detail in this course on the details of how labeling takes place in yeast cells, but I would recommend you to refer to this paper to get more details.

(Refer Slide Time: 08:37)



So, let us move on to the next approach for isotope labeling in cells and that is known as cell free synthesis. Now again this are the reasons why we need cell free synthesis because sometimes your proteins may not be expressible in E. coli or it may not be even expressible in yeast as well. So, therefore, we have to basically try some other approach and that approach is called cell free synthesis. So, as the name suggest here to see that word cell free means that are there is no cell involve means we are not growing any bacterial or yeast culture.

So, there is no biological organism involved here, but; obviously, without biology we cannot make a protein. So, there is something which what is done is let us see in this slide. So, here you can see these are the reasons mentioned why do we need cell free synthesis and what are the problems if you want to grow our cells and you make protein inside the cell. So, the problem, one of the problem is that many of the proteins which are expressed are actually insoluble in the cell.

So, inside the cell when a protein is getting made it is not soluble in the cytoplasm and therefore, it start aggregating and precipitating and that is called as inclusion bodies. So, this word some of you may have heard some me of you may not have heard. But the point here is this proteins which are expressed in our E. coli sometimes do not solubilize and therefore, come out or come into the cell precipitate and they form lumps call as inclusion bodies.

So, how do we deal with that problem? That is a separate story here we are not going into details. For that there are different approaches which have been suggested to avoid the formation of inclusion bodies. For example, one could try different cell temperature growth temperature and so on. try may reduce the production sometime the protein is made in very large amount it contain to aggregate and so on. Sometimes the protein itself you have to change the sequence of the protein it may happen the protein has a particular sequence primary sequence which is causing the problem.

So, by changing the primary sequence mutations etcetera you can again get rid of inclusion bodies. Another option which is propose suggested is you can actually do what is called a fusion protein. This is something which we have not discussed in this course, but I will briefly mention what is the fusion protein. In a fusion protein you take a protein a standard protein. For example let us take maltose binding protein or GST and this kind of proteins are well expressed in protein in a in the cell in E. coli.

So, you take that protein and then the DNA the protein of your interest is fused to this protein; meaning the DNA which we generate now the DNA of the protein and the DNA of the standard protein is included together as a single chain in the cell. That means, when the cell makes the protein it makes a single chain which consists of the standard protein plus our protein. So, let me draw it here as a schematic to illustrate the point.

So, let us say this is the DNA of the protein. So, what is done is there will be a standard protein standard meaning one which can be expressed very easily in E. coli and our protein. So, this is called fusion. So, we are fusing two DNA's. So, and of the one of this protein which is very standard or which is getting express easily in the E. coli and one which is our protein of a interest. So, now, our protein because it is tagged so we can call it as a tag.

So, because our protein is now tagged with another protein which is expressible in the cell which may not have be a problem. So, our protein also now the total system this whole thing solubilizes it does not precipitate in the cell. So, of that happens then we can express our protein as a fusion protein with another protein. Then after purification this separating this from the cell I said we can cut it here we can cut it by some enzymes which can cut this protease.

Proteases are there which will cut at a particular sequence here and then our protein will get separated from the other protein. So, this is a very standard approach we call as fusion systems and it is done routinely in cases where the protein of the interest is not soluble in the cell or is not able to express well in the cell. But this which is express very nicely if it is a protein known like this then our protein can be joined or fused with that in the data DNA level.

And it comes out also in the protein level as joint, but can be separated by proteases by putting a protease which will chop the protein here and our protein becomes separated from the tagged protein. So, this is the approach which is used for cases as I said where it forms inclusion bodies or it is not expressing well. Sometimes it can happen that your proteases in the bacteria starts digesting means degrading the protein which is of our interest.

So, this can happen if the protein of our interest has something sequence which is recognized by proteases and they start digesting it. So, these kind of problems can arise several times in E. coli and therefore, one has to be decide what to do. So, one option is you can go to yeast, but in yeast also there is a problem then we have to go to cell free methods. Some proteins cannot be produced in living cells because of their toxicity.

So, this is another reason why we cannot expressing E. coli. For example, let us take we are expressing a protein which is which kills the bacteria so it is kind of an antibiotic. So, is that is a case then the protein; obviously, inside the cell we start killing the cells. So, for cells will not be able to survive because our protein is actually poisonous to the cell so therefore, we cannot make it in bacteria. Similarly yeast maybe a problem for some other reason.

So, in that case again cell free method helps because there is no cell here. So, there is no question of poison to the cell in this case. So, that is a reason why we need. So, we can see there are so many other reasons which we may need. for example, another very important reason we use cell free method is because of the scrambling; is a isotope dilution what is written here is nothing, but scrambling. So, I can write this here it is scrambling.

So, this is something we referred to in the selective labeling approach. and that is creates a problem because suppose I want to label a particular amino acid let us say alanine or

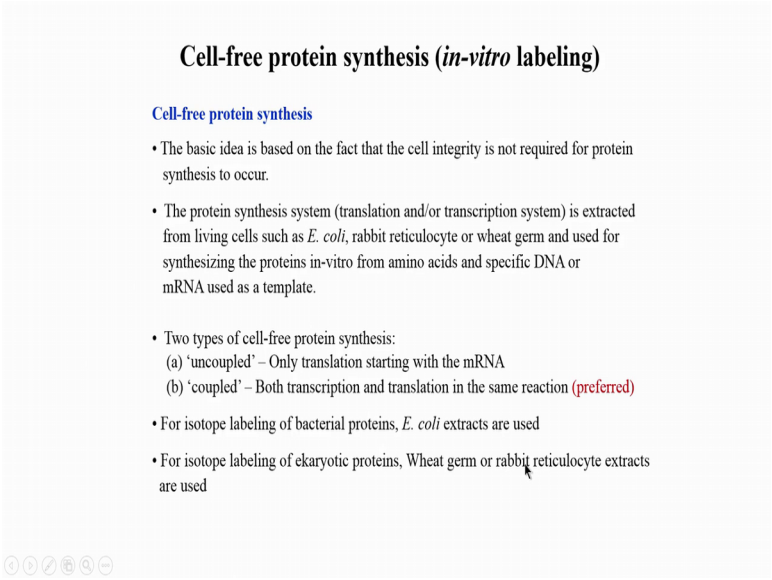
aspartic acid. And that because of this scrambling problem it may label some other amino acid also. So, aspartic acid may label alanine, espazine and so on. So, such kind of problems can be avoided if we go to cell free because in cell free we do not need to have a cell. So, therefore, all that enzymes which are causing the scrambling problem are not present in the cell free approach.

They are present in the cell, but not outside the cell. So, we can take make a get rid all that problem with cell free approach. Now it is another issue that can be that sometimes the bacteria does not grow well in deuterium labeled medium. So, this is face very important for deuteration if we recall in deuteration for deuteration we need the cell to be grown adapted to D2O. And they can be is possible that many times that is a cell is not growing in D2O well.

So, again for that reason we need to go for cell free again the E. coli this was something mentioned earlier. E. coli cannot glycosylate proteins; which means in any glycosylation that is called post translational modification cannot be done in E. coli. And therefore, again we need to resort to some other approach that is cell free. And membrane proteins are another set of proteins which we did not go into detail in this we are not going to go detail in this course. in membrane proteins is very difficult to express in E. coli or in vivo because of the solubility problem.

So, therefore, we again have to go to other approaches as cell free. So, these are the reasons why expressing a protein in E. coli is not always possible although equalize a most popular approach many of the proteins are simply not possible to study with or express in an isotope label. So, one of the very approach is cell free synthesis was discovered actually more than 40 years ago, but it is use started only recently.

(Refer Slide Time: 17:57)



So, let us see what is happens in cell free approach what are the things we need. So, what are the basic idea here? The basic idea is that the cell is not required for protein synthesis to occur. In the sense I do not need a full living cell remember a cell consists of organelles it consists in E. coli it consists of many other proteins many other things in the cell. But where is the protein actually being made? The protein is made by from DNA it is transcribed to RNA from RNA it translated to protein by on the ribosomes.

So, basically and other enzymes are needed tRNA is also needed. So, if I can collect all this machinery or the transcription and translation system all the materials which is proteins and other chemicals which are important for transcription and translation. Then I can take those things out of the cell means I can extract it from cells and then use it as such. Because now that powder that extract will contain all the enzymes and ribosomes and tRNA and all the things which are required for protein synthesis.

So, I really do not need a whole entire living cell to make my protein now. I have the all the required ingredients or material which can now be use for expressing or a translating a DNA into protein. So, for example, from DNA to protein you need RNA polymerase because RNA polymerase will make the mRNA. So, if I have RNA polymerase in my solution and I add DNA it can be transcribed into the RNA of my interest so like that.

So, like that similarly if I have tRNA for all the amino acids and I have this mRNA and ribosome and others required enzymes it can be translated into protein in the solution in a

in vitro. So, we do not really need a cell or in vivo approach for isotope labeling. So, this is basically the idea that you take the DNA of your interest again this is similar to the DNA which you put it in the plasmid in E. coli the same type of plasmid DNA can be taken.

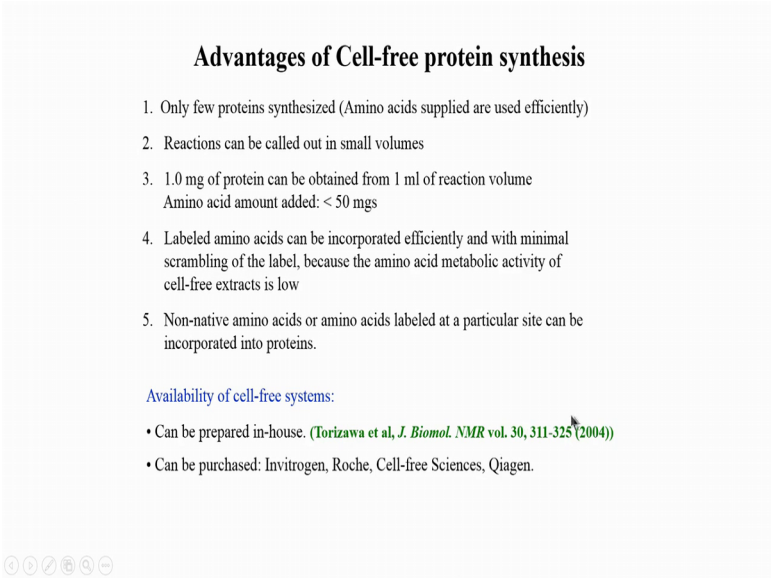
And you can use now mRNA also if required if you have mRNA. But DNA is sufficient DNA is good enough and once you have the DNA of your interest you can then transcribe into RNA with polymerase and other things and then convert that into protein. So, these are the two little bit detail of cell free synthesis the there are two types of approaches in cell free synthesis. One is called uncoupled and coupled.

So, in coupled both transcription and translation takes place in the same reaction. Meaning I take a test tube I add all the mixture I add the DNA of my interest and I add all the amino acids and then start the reaction. So, first the DNA will get converted to the RNA and then RNA will get translated into the protein in the same test tube so it is a one part reaction. But this is happening again remember there is no cell growth here. Of course, that means we need to add all the amino acids we need it can be now glucose or ammonium chloride.

Because ammonium chloride and glucose were used for the cell to grow and it was converted in the cell into amino acids, but here there is no such conversion no growth. So, we have to add directly the amino acids which are required. So, for isotope labeling E. coli extracts are used; that means, suppose I am take making a bacterial protein which can be suitable in bacteria. Then I can extract all the machinery the machinery here translation transcription machinery from the E. coli.

And use that powder that kind of a system that is the my in the test tube and make the protein. And for a which if we want eukaryotic protein to be expressed or made or isotope labeled then you extract these machinery translation and transcription from this wheat germ or rabbits reticulocytes ok. So, these are the two cell systems from where you extract. Remember again we are not going to grow the cells we are going to extract those machinery or system from those cells and use it use that extract in a test tube.

(Refer Slide Time: 22:27)



So, let us go little bit further what are the other ingredients which are needed for a cell free synthesis. So, you can see here we need basically in small volume. So, we do not need to really use very large volumes like 1 litre in as we did in E. coli is very small volume. And amino acids have to be supplied because as I said there is no growth of cells here so glucose and ammonium chloride do not exist I mean to be used you have to use amino ammonium acid supplied.

But the advantage of cell free is you do not have to supply a large amount like we did in the E. coli for selective labeling where you can use very less amount of the amino acids and we can save the cost. So, you can see here 1 mg of a protein can be obtained from just 1 ml of your reaction volume ok. So, 1 ml of a reaction volume in a test tube will give almost 1 mg of protein. Typically we need about 5 to 10 mg for a medium small to medium size protein.

So, you can see if I go 10 ml reaction volume 5 to 10 ml I may be able to get all my requires protein. But so in E. coli cell real cell if I am growing getting 5 to 10 mg protein may require 1 or 2 litres or more of the culture and that becomes very expensive. So, this is kind of a less expensive approach and how much amino acids we have added for getting this much? We would have added about 50 mega milligrams you can see compared to an E. coli we were adding 100 mg per litre of culture or more, but here we are adding very less amount.

So, you can that the why is it we need less here compared to E. coli, is a very simple answer. The answer is in E. coli when we are growing the cells all the amino acids which we added are not just for our protein it is used for the full sell the entire cell all the proteins in the cell are now using those amino acid which we added. Whereas here only the DNA of our interest is added there is no other protein which is getting made. So, therefore, entire set of amino acid which we use is needed only for making our protein.

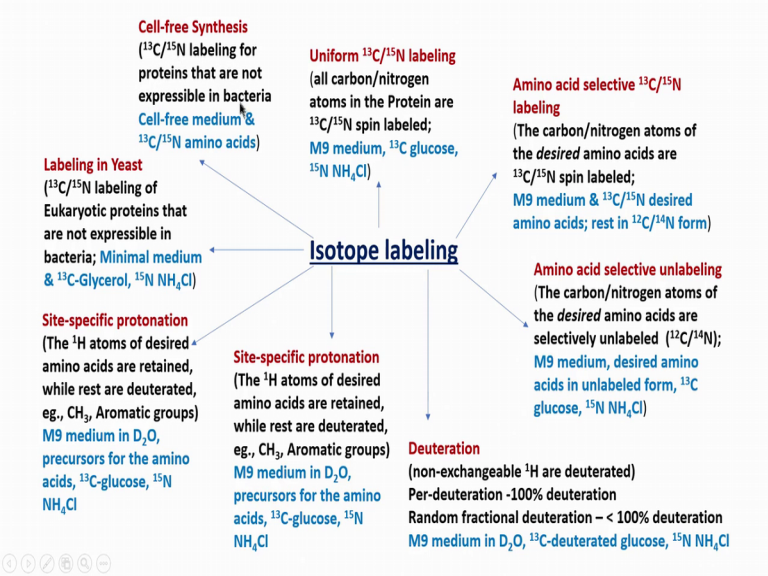
So, obviously, we do not need to have very large amount of amino acid because we are only channelizing the entire machinery to make our protein. So, therefore, the amount of amino acid requires is much less. Another advantage of having doing cell free synthesis is shown here. For example, many times in a protein you may want to introduce incorporate non native amino acid. So, typically there are 20 amino acids standard in biological systems in living systems, but there are many more non natural or non native amino acid.

Now, if you want to incorporate that into your protein you cannot expect that to be done by E. coli. Because our yeast because if they do not have that machinery they only work with the standard amino acids. So, if you want to make them incorporate or a strange or a normal or a non natural amino acid into your protein then you have to do many more things in cells. But in cell free synthesis it is very easy you have to find we have to basically tweak the DNA to change the code of the DNA.

And you code it for a non natural amino acid which can be added into the reaction volume. So, by doing that non native amino acids can also be incorporated easily in a protein compared to what we can do inside the cell in a cell based method. So, there are two methods cell based and cell free. So, the cell free systems can be prepared in house all you need is basically the reaction volume which you can extract from the as I said all the machinery can be extracted from E. coli.

Or you can buy this as a kit from different companies. So, again I recommend you to go through this paper which gives you much more detailed information on how this cell free extracts are prepared in the laboratory. So, that anybody can make them and use them.

(Refer Slide Time: 26:37)



So, now let us summarize all the isotope labeling schemes that we have learnt in this part of the course. So, we have basically we started from looking at uniform labeling that is a first thing. And you what is the approach in uniform C 13 N 15 labeling that all carbon nitrogen atoms in the protein are now see C 13 and N 15 spin labeled ok. So, you can see this slash basically means or so you do not have to always do both of them you can choose either [N] C 13 label only or 15 only or you can choose both double so this is called double labeling.

And double labeling means both are labeled in the protein single labeling means either C 13 or N 15 is labeled in the protein. So, we can choose whichever we want we want uniform C 13 labeling or uniform 15 labeling. So, what is the media which is written very briefly here of course, this is not a complete medium list you need some more items. But the idea basic idea is unit of specific minimal medium called M9 medium which is supplemented means in that you are adding C 13 glucose because you need to lose C 13 labeling and N 15 ammonium chloride.

Actually M9 medium already contains C 12 glucose and N 14 ammonium chloride. So, you can remove that and replace it with C 13 glucose and N 15 ammonium chloride. Other things in this medium consists of salts and other things you have to add a supplements or vitamins and so on, but that is not listed here, but this is the basic idea.

So, the second thing which we saw was selective labeling the selective labeling what is the aim of that?

The aim of the selective labeling is so the carbon nitrogen atoms of a desired amino acid or particular set of amino acids or label with C 13 or N 15 spins ok. So, we choose a set of amino acid not necessarily only 1 amino acid you can chose 2 or 3 amino acids for which you want to label them and for that you have to add the desired amino acid in the medium. So, the medium is M9 medium, but you add these amino acid whichever you want to label in the medium.

But that is let us say you want to label alanine then you add alanine in this labeled form, but the other amino acid can be in the unlabeled form unlabeled meaning C 12 or N 14. So, you have to add all the 20 amino acids, but 1 or 2 which are to be labeled will be added in this form and remaining 17, 18, or 19 whichever you do not want to label will be in this form. Another approach we saw is a amino acid selective unlabeling.

This is reverse of this approach so it is also called as reverse labeling ok. So, selective unlabeling the goal is that you want the carbon nitrogen atoms of some amino acids or again desired or particular amino acid in the unlabeled form. So, you do not want them to be C 13 and N 15, but you want them to remain unlabeled. And how do you achieve that you achieve that again using a M9 medium and you add your amino acids which you want to unlabel in the unlabeled form.

But you have to remember that there is the remaining amino acids you do not have to add like here the remaining amino acids or they will come from C 13 glucose ammonium chloride which you will be adding in the medium. So, when you add C 13 glucose N 15 ammonium chloride they become carbon 13 N 15 labeled. But our amino acid which we want to unlabel is added exogenously from outside in the unlabeled form.

So, the bacteria will take up our amino acid which we add and it will not label it, but all the other amino acid in the protein will be labeled so it is opposite of selective labeling. The next approach which we saw in this course was deuteration. This is very useful important for large proteins typically 25 kilo Daltons and higher there what we do is the all non exchangeable the CH hydrogens are deuterated so they become CD.

Now there are two approaches in deuteration per deuteration where you do 100 percent of deuteration means all the non exchangeable or replace entirely like uniform deuteration. So, in equivalent to this uniform labeling uniform deuteration or we can do random fractional deuteration means we choose a less than 100 percent deuteration you can chose 60 percent, 70 percent, and that many protons roughly approximately are replaced or not replace or deuterated and the remaining become protonated.

So, this is useful if you want to keep the hydrogens and per deuteration is useful if you do not care for the non exchangeable hydrogen. But remember that n h the amide which is exchangeable hydrogens they are normally protonated so they remain in the proton form. Now what is the medium which is required here this is now M9 medium again, but you have to make it in D2O the medium has to be made in D2O. And if you want 100 percent deuteration you should use C 13 deuterated glucose means the glucose also is having all hydrogens which are now deuterated.

And this is a very expensive chemical, but you have to use if you want very good deuteration level. And ammonium chloride remains the same we can use N 15 ammonium chloride. So, after this we look a site specific protonation where a specific site like for example, methyl or aromatic protons you want to keep them in a protonated for we do not want to deuterate, why? Because this is very useful for structure determination of proteins so we do not want to deuterate that and we use that.

So, this is basically again the medium is D2O and the precursor amino acid of that which we want to keep it protonated are used. And C 13 glucose and ammonium chloride is used for the remaining amino acid which we want to label. But the methyl groups are less the precursors are used which then make sure that the sites are kept protonated they are not deuterated all the others are deuterated because of D2O.

The next thing we saw is site specific protonation. Again this is similar to what we saw here sorry this is repetition of that then we are done labeling in yeast and yeast basically we are looked at how to grow proteins which are not expressible in bacteria. So, we saw that the medium we use is minimal medium with C 13 glycerol. And finally, we look that cell free synthesis where we need cell free we do not need the cells to grow any cells.

And but all the cell machinery is extracted from the cell and used in a test tube form. And we add the DNA of our interest in the test tube and we start the reaction which makes our

protein. But what is required here is we have to supplement where we add all these amino acids in the test tube. Because that is not there is no other there is no glucoses and ammonium chloride which will be used by the cell free extract it has to be added in the cell free in this form. So, these are the different isotope labeling schemes that we have gone through.

We will now finish this we will in the next class we will begin with the assignment of a protein. Once your protein is isotope labeled, how do we then do 3DMNR to assign it, we will see that in the subsequent classes.