Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

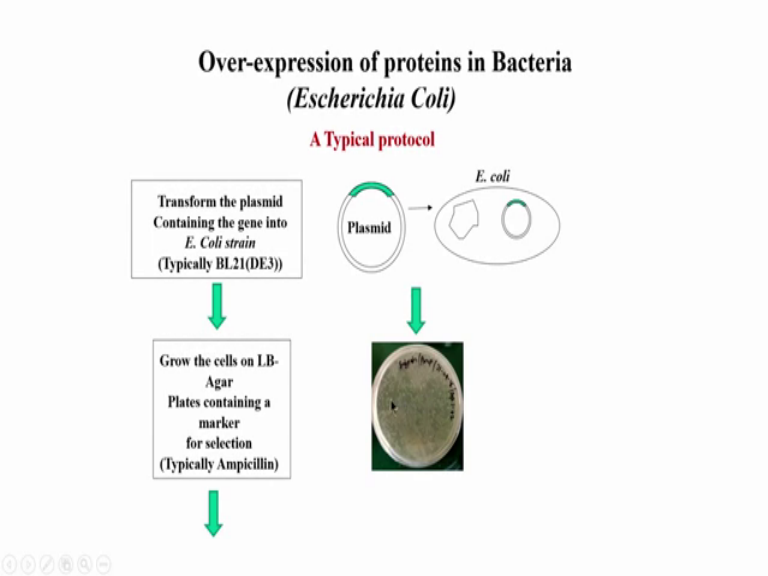
Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture - 24
Isotope labeling of proteins for NMR studies - Part I

In the last class we were we started looking at how the bacterial cells are grown for Isotope Labelling. So, the first step we saw is transformation where we take the DNA of our interest, put it in the vector or plasmid and that is inserted into the cell and then we grow it on a agar solid plate which is nothing but a petri dish and in that dish we start seeing the colonies. This typically takes about 16-20 hours, let us say a day, and after a day you can come and see this colonies would have grown on your plate.

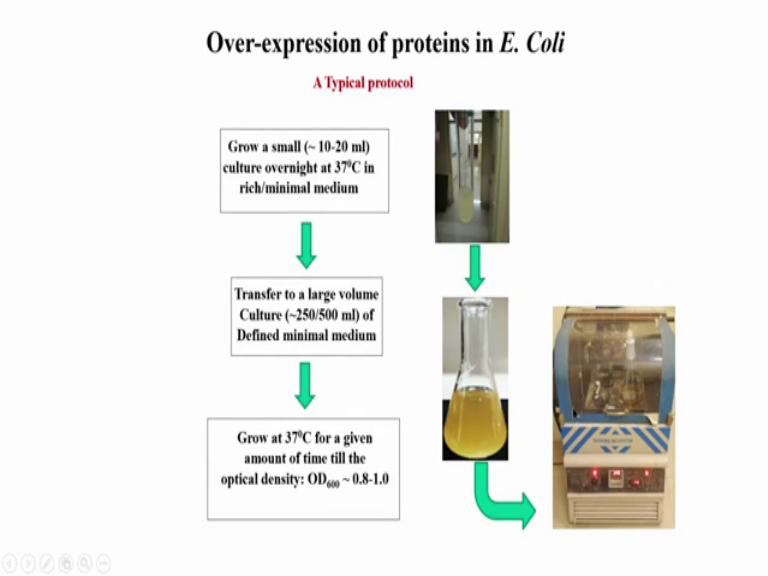
Provided, you have selected for the correct marker because remember ampicillin if it is not resistant to the contaminants also can grow. So, we need to have ampicillin resistance built into the cells. So, this is important point. We need the ampicillin resistance built into the cell which I said we can do with simple recombinant DNA approaches.

(Refer Slide Time: 01:13)



So, now once this colonies have come, we can choose any of this colonies. You can see there are millions of them we can choose any of this.

(Refer Slide Time: 01:28)



And then what we do is we put it in a small medium, again this is 10 ml to 20 ml you can see this is like a test tube here. So, in a test tube we have we can put that colonies and then grow it overnight at 37 degrees in the minimal medium. So, this is how it will become turbid after it has grown for overnight, grown at overnight, ok. So, this is the turbid you can see this is a milky type of a solution it has become because this is full of bacterial cells, ok.

So, again remember in this medium what are the sources of carbon and nitrogen? This is going to be a minimal medium for labelling isotope labelling, in a minimal medium we have carbon 13 or carbon 12 glucose as a source of carbon, and ammonium chloride as a source of nitrogen. And other than these two we have to add some more components such as a salt or a buffer, it has to be buffered and also we need magnesium chloride, magnesium sulphide, zinc and vitamins, and those are in very small quantities. The major component here we are talking about is carbon 13 or carbon 12 glucose and nitrogen ammonium chloride.

So, overnight this 10 to 20 ml culture is grown like this and once it becomes turbid we transform it or we transfer it to a larger volume. So, we can say this is a medium which consist of the again the same minimal medium we take a larger volume around 250 or 500 ml and this whole solution is just transported or transferred or you can take the cells you can centrifuge it, take the cells and put it in this medium, ok. This medium is again

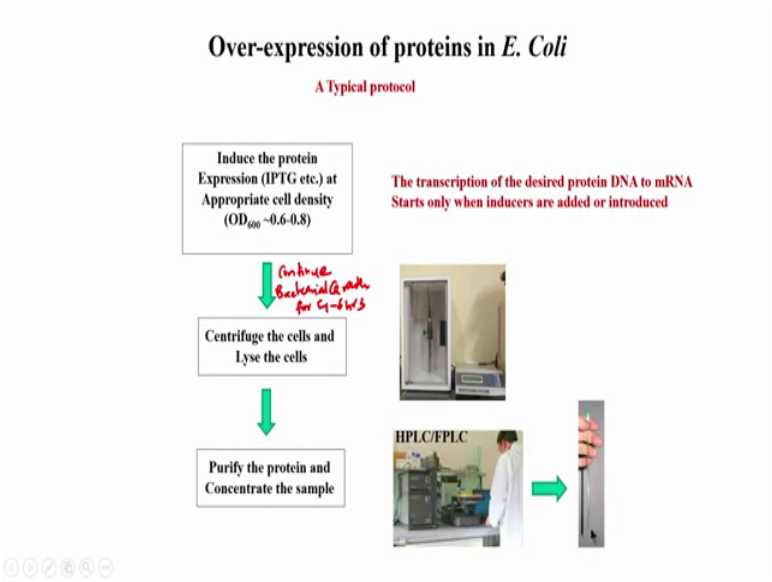
the defined minimal medium which contains the carbon 13 as a source of glucose as a source of carbon, and ammonium chloride hence other materials ammonium chloride mainly for nitrogen. So, this is very important and this is all throughout we are using minimal medium everywhere, ok. Sometimes people use some other medium here at this stage and then transfer it to minimal that is also, ok.

Now, after this transfer after this transferring this liquid from here to here we need to grow it again this particular culture. So, this is grown in incubator shakers which look like this, this is a small incubator shaker you can see the flask is here and this is actually it will shake, ok. So, this is rotated or shaken at a particular speed that is called rotations per minute rpm, ok. So, there and the temperature inside is kept at 37 degrees because that is optimal temperature at which E Coli cells can grow.

Now, the question is how long do you grow these cells? You grow the cells till the turbidity. So, this is the turbidity which looks like this is measured with a spectrophotometer. So, you take 1 ml of this take it to a spectrophotometer. So, you need a UV visible spectrophotometer in your lab. So, you take that to that spectrometer and at 600 nanometers, wavelength is set to 600 nanometer you measure what is the absorbance or we use the word optical density. You that absorbance is around this value then that is enough for the growth of the cells, we do not have to grow it more, ok. So, this is important point. The cell density this is called the log space means cells are very high number they are dividing very rapidly and we have to now do the next procedure at this stage. So, you should not overgrow the cells. If you allow it to grow to 1.2, 1.3 then the cells will starts dying.

Similarly, if you grow very less 0.2 or 0.3 then the cells totally basically would have not grown very enough. So, your protein will not come out in the good quantity. So, this is the stage two which the cells are grown and as I said they are monitored using UV visible spectrophotometer by setting the wavelength to 600 nanometers, and in this wavelength you are measuring basically the scattering, ok. So, this is not that is observing the slide it is the scattering by this the turbid the cells because the cells are also turbid. So, we are measuring the turbidity of the solution.

(Refer Slide Time: 05:49)



So, once it reaches 0.8 to 1 what do we do next. The next step is called induction. So, this is shown here. So, now, this is what we do now. So, what is induction? Induction is the transcription of the desired DNA protein, DNA that is remember we took the gene of our interest and we have put it in the plasmid. Now, the plasmid it is now it is a DNA form. So, it has to be transcribed or the transcription of the DNA to MRNA has to take place because without that our protein will not be made, and that transcription begins or starts only when the inducers are added. What are these inducers? These are chemical inducers or physical. Like for example, I can add a molecule known as IPTG.

See if I add IPTG to my solution when the density as reach this point then that protein that under the RNA of that protein of our interest starts getting out a large numbers, ok. So, the theory is that the till we add this induce inducers our protein is not so much expressed means our protein is not express at all, it is blocked. But a moment we add this inducers which we do not add in the beginning we add only after some time when the cell density has reach appropriate value, and that point we add this inducers these are chemicals or it can be heat shock, I can do something else for induction but typically in a standard very common approach is induction by IPTG a molecule called IPTG. I do not have the molecular formula here but you can look up in internet or any book.

So, once I add IPTG to the cells, the cells starts now making the RNA of my protein of interest, because the DNA which was there in the plasmid is convertor now transcribed to

MRMA. And this is a step very important because now this at this stage is when our protein of interest is being made in the cell, ok. So, that is done only at this appropriate stage.

So, now what happens is you know have C 13 labelled glucose let us say in your medium and my protein of interest need C 13 because it needs amino acids, it needs all the other things to make the protein. Now, all the amino acids have come from glucose because glucose is the only source of carbon. So, if glucose is only source of carbon amino acids are C 13 labelled and all the protein the entire protein which is coming now from this DNA of my interest is also carbon 13 labelled.

So, you see what we have done essentially, we have we have basically smuggled the DNA of our interest into the cell and by smuggling we are basically fooling the bacteria, we are acts in the bacteria to grow and make our protein in large quantity although that is not its native protein. The protein which we are making a something foreign means something alien to the cell the cell naturally does not contain that protein, but we have smuggled the DNA of our interest into the cell and we are making the bacteria make that protein for us in large quantities. And since the bacteria is growing on a C 13 labelled medium the entire protein is also carbon 13 labelled, ok.

So, that is how this labelling part comes into E Coli through this kind of approach in which the cells are grown in C 13 labelled glucose and the protein which comes out is also. Now, C 13 labelled because we have made that about DNA to express the protein in large quantity. So, this is the part known as induction.

Now, once I IPTG I need to grow further. Grow means, I need to keep the bacteria growing for some time. So, that is that step is not shown here, but the step after this that after induction you have to still grow the bacteria for some time. So, I will write it here you continue bacterial growth for let us say 4 to 6 hours. This is something which has to be optimised of course, there is no golden rule whether it should be 4 hours, 2 hours, 6 hours, 8 hours it depends on the temperature in which you are growing.

For example, let us say your induction, induction happens at 30 degrees, then you need to grow little longer with induction happens at 37 then typically 6-8 hour should be good enough. So, this has to be optimised for your protein, if you want to get a very good yield means the output this number of amount of protein.

So, once that 4 to 6 hours is over we have to now our job is done we have to centrifuge the cells and then break the cell, lyse means we have to. Why do we need to break open? Because remember our protein which is coming from DNA to RNA is sitting in the cell inside the cell. So, the cell has to be opened to remove our protein. But again if I open the cell remember there are many more proteins in that cell it is like a soup. So, the entire soup of the cytoplasm contains our protein as well as other protein. So, we have to do something more later further which we will see. But worst thing is we have to open the cell break open the cell and take out all the cell proteins in the cytoplasm, that is provided our protein is actually in the cytoplasm and not in what is called as inclusion bodies we will not worry about that. We will assume that our protein is soluble in the cytoplasm of the bacteria and what we are basically doing is taking out the entire soup out of the cell.

Now, this breaking, break opening is done typically with this instrument called as a sonicator or there is another instrument called French press which we will not see we look at this sonicator. So, sonicator is basically something like this which has a probe which is goes into the cell or solution of cells we take a make a solution or what is called suspension. We take the centrifuge cells and resuspend it in some solution and then we break open the cell wall. So, essentially, we are breaking the cell wall so that the entire cytoplasmic proteins come out. Once they come out, we have to purify our protein of interest because remember as I said it is a soup it contains many other proteins in that in the solution. So, we need to separate those proteins from our protein and that typically is done by HPLC or SPLC. So, you need to use HPLC as for some cases, sometimes HPLC sufficient and we can do that and get it in this form.

Now, once we do that protein may be dilute it may be not be very concentrated. So, if you recollect you need a very high concentration of protein sample in NMR tube, when we do NMR experiment. So, this is important, remember we typically we say that about 1 milli molar for is required for a protein of course, for some proteins 1 milli molar may be too high a concentration it may start precipitating so on, but as high as possible we have to do because NMR sensitivity very much depends on how much is the concentration.

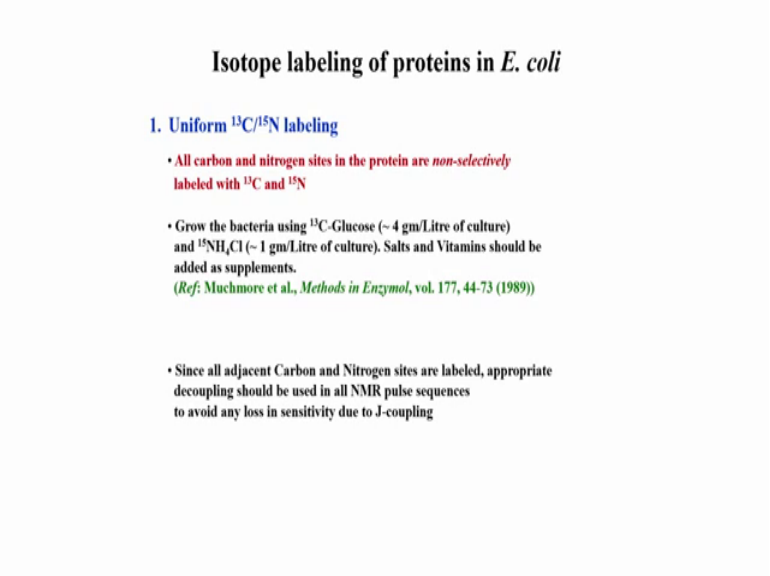
So, this is the final step where we take the purified protein coming from which we will see and we concentrate with some techniques like dialysis or you can use centric on and

so on, that different small instruments and from based on that we can concentrate and put it in a NMR tube, after this you record NMR data.

So, now this is a labelled sample. Now and why do I say it is labelled? Because remember right from beginning we have been growing the bacteria in a carbon 13 labelled medium or N 15 labelled ammonium chloride. So, therefore, every carbon every nitrogen of that protein of the cell is now, carbon 13. And for every protein as now carbon 13 labelled not only our protein, also the other protein in the bacterial cell. But remember we have separated those other proteins from our protein by using HPLC or SPLC or purifying our protein.

So, what this NMR tube contains is purely our proteins. The purity is it should be very high typically the purity we can obtain is about 90 to 95 percent. So, 95 percent purity is good enough for NMR, I figured 99 or 100 that is even better but typically that is the level of concern purity. And now but every carbon in that sample should be C 13 and N 15 because we have labelled it. So, this is how an uniform labelling of proteins is carried out, uniform C 13 N 15 because remember we are not selectively choosing an particular amino acid type for labelling, you have chosen all the amino acid, everything is labelled with C 13 and all N 15 in case, we want N 15 also.

(Refer Slide Time: 14:37)



So, now let us say this is what basically is the is the idea of uniform labelling, that is all the carbon and nitrogen in the protein are non-selectively labelled with C 13, they are all

uniformly labelled. So, this is what we said you grow the bacteria. So, this is the question is how much the glucose do we need? We typically for 1 liter of the bacterial culture we use 2 to 4 grams, 4 grams is on the higher side you can also use 2 grams per liter means culture volume of liters.

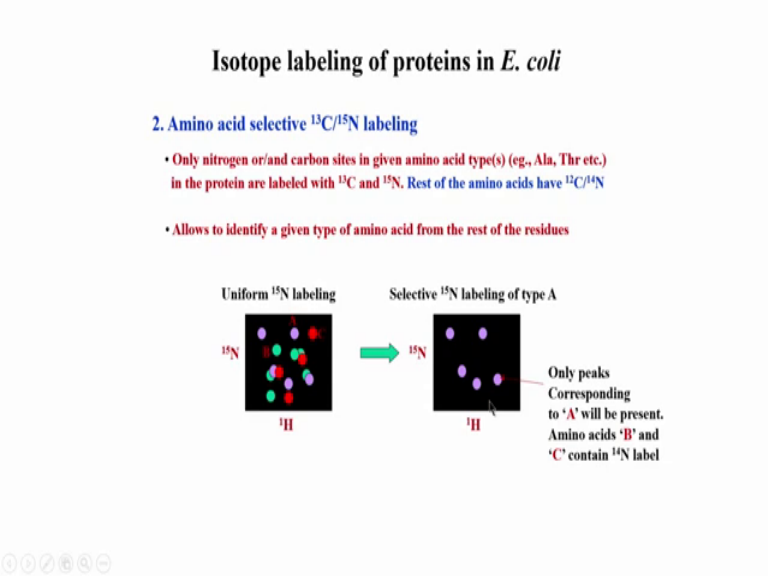
So, I am let me go back and show you what is it volume I am referring to I am referring to this volume here. So, this is actually the flask may be large, flask may be 500 ml but the culture is not 500 ml, this will be about 200 ml. So, if you have a 1 liter of this volume in a big flask for each liter of culture volume you need to use 4 grams of glucose carbon 13 or 2 3, 2 to 4 grams. For ammonium chloride the requirement is less it is only 1 gram is enough for 1 liter of culture volume. So, these are the two things which we have to add. And there of course, as I said you have to also add salts and vitamins as a supplement, ok.

So, I would refer request you recommend you to go to this paper and see more details of how we get this labelling done. Now, one more important thing if you do uniform labelling. So, we are require talking about uniform C 13 N 15 labelling which means every carbon in the protein you see 13 labelled so that means, my neighbouring carbon is also carbon coupling is present now in the molecule. Remember for natural in the case of natural abundance we ignored the carbon 13, carbon 13 coupling because the because of one person abandons the chance the two neighbouring carbons are C 13 was very low.

But in this case we have now every carbon as C 13. So, therefore, we need to worry about C 13 C 13 decoupling, that is what is mentioned here that once you have uniform labelling you have to now use appropriate decoupling in all NMR experiments or pulse sequences, because if you do not do that you will lose sensitivity by decoupling. The peaks will start splitting so you need to decouple carbon from carbon, carbon from nitrogen, and nitrogen from hydrogen, and carbon from hydrogen. So, all possible decoupling have to be implemented or used to avoid any loss of sensitivity.

This is something which is already build into the NMR experiments. Like for example, when we discuss HNCA, HNCOCA and so on. All those are already taken care in those experiments, so in a standard NMR experiment if you are using normally, we do not practically we do not have to worry about this because this is built into it is default present in those sequences.

(Refer Slide Time: 17:30)



So, let us move on. Now, for a selective labelling concept. So, now, what is selective labelling? So, selective labelling basically as I mentioned is selective labelling of a particular amino acid type. For example, as I said alanine or threonine if I want to label only that amino acid then only carbon nitrogen in that amino acid is labelled with C 13 N 15. And the entire remaining rest of the amino acids in the protein are now having carbon 12 or nitrogen 14, they are not labelled isotopically labelled with C 13 or N 15, ok.

So, how do we achieve that we will see shortly. For first let us appreciate why do we actually need that, why do we need to do the selective labelling because of the following reason. So, look at this schematic drawing here this is a uniform N 15 labelling protein, labelled protein which means all the peaks each peak corresponds to 1 amino acid here. So, this is something which we have seen in the first part of the course of an HSQC gives rise to different peaks. So, each peak in HSQC corresponds to 1 amino acid.

But let us say I have 3 types of amino acid in my protein, 3 types means let us say 1 type alanine, B something else, C could be something else. So, let us say only 3 types of amino acids are present in my protein. So, all the Bs will come like this, all the as will also show up and all the C will come. But remember this colours are only just to show you but actually in a spectrum we cannot figure out which is A, which is B, which is C. It is just to illustrate this point I have shown in different colours, but in a real spectrum NMR spectrum in HSQC you will not be able to distinguish which amino acid type is

each peak. They all come this, they look all the same.

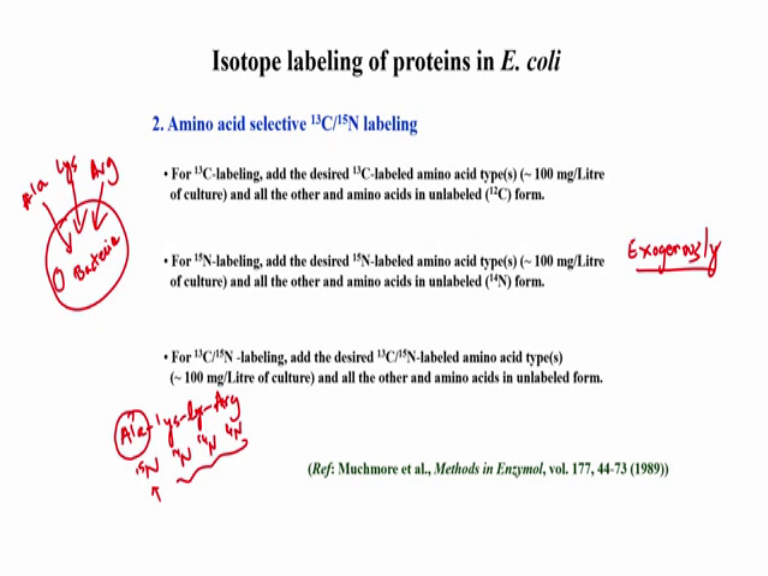
But let us say now this looks little crowded here, and I want to get rid or I want to simplify the complexity of this spectrum. What I can do, I can only label A, I can choose to label A. If I do that then only A will come in the spectrum because only A this will be labelled with N 15 and all the B and C will not be there will be unlabelled that is what was mentioned here B and C will be N 14, but A will be N 15.

So, in that case only peaks corresponding to A will be present and amino acids B and C will contain N 14 and therefore, they will not appear in this spectrum. So, by doing this I am actually you can see here have reduce the complexity, have reduce the number of peaks, I have made my life simpler. I can do the same with B; now, I can do selective labelling of B and record get another sample. Remember each one is one separate sample. So, this sample only can now, have a has labelled I need to prepare another sample go through the procedure which we saw in the previous slides and grow bacterial cells again but now, that is cells will have only B labelled. So, that is another sample. I can make another sample where only C is labelled.

So, like this for each amino acid I can prepare sampler separate samples. But do we actually do that? Practically no, practically we do not do that way. We actually what we do is typically we choose a few type of amino acids which we are interested which are for example, very important for us, alanine could be important, or lysine may be very important for the function of that protein. So, we choose a few amino acids which are in abundance that is another issue.

Suppose I have a lot of lysine's in my protein and I want to selectively look at them then I need to do selective lysine labelling. So, we do not as I said we do not really each take all the 20 amino acid and do one by one, that is not practically a good idea because selective labelling is little expensive compared to uniform labelling. So, in uniform labelling like this it is not so costly, but if I want to do selectively look at A or B or C then I need to do something which requires money or cost, extra cost. And how do we achieve this? We will see shortly in the next slide. But remember we have to choose a particular amino acid or few amino acid and label them selectively.

(Refer Slide Time: 21:43)



So, let us see how the selective labelling is carried out. Again, this is very nicely explained in this paper. So, what we do is we take this suppose we want to selectively label only alanine, but again remember I want alanine carbon 13 label. So, we take that amino acid we purchase that remember commercially, so it is available for example, in many companies. So, I can buy a carbon 13 labelled alanine for example. Let us say I am looking at alanine. Then how much alanine should I add? What I do is in the medium I will add this much per liter of culture typically 100 to 200 milligrams, ok.

So, what we are doing is we are taking the medium, that that is culture cell culture remember the cell culture I showed in the previous slides we take the cell culture medium into that we add 100 mg of C 13 labelled alanine, ok. And we add all other amino acids which are unlabelled.

So, now what is happening? We are not adding glucose. We remove the glucose out of the medium that means, I do not use glucose in the medium but rather I am using each amino acid. So, I take 20 amino acids but out of the 20, the 19 are unlabelled. So, they are added in unlabelled form which is very not very costly, one particular amino acid which I want to label is added in a labelled form, ok. So, all of these are added in the same quantity 100 milligram per litre of culture.

So, now what does the bacteria do? The bacteria actually now if you want N 15 labelling, then in that case you take the N 15 amino acid from the company. You buy this, labelled

amino acid from a company and that is added in the same quantity and other amino acids which you do not want to be labelled has to be in N 14. For example, in our previous slide we wanted A to be labelled. So, if I go to the previous slide this is what we want to be selectively N 15 labelled, but B and C do not have to be labelled. So, in such cases in this example we will take A in N 15 form, and B and C will be unlabelled form. So, we add those 3 to the culture, and the culture does not contain glucose but rather it contains these amino acids.

Now, what is the reason for that? The reason for that is if I add glucose my carbon from glucose will start getting into the amino acid. So, that is something I do not want. I want the bacteria to only take those particular amino acids which are labelled. So, what the bacteria will do is that it will stop making its own amino acid. It will take the amino acid which you have given outside them in the medium, this is known as exogenous. So, this is known as exogenous, exogenous amino acid. So, I am adding exogenously my amino acid, means I am adding from outside these amino acids. And what the bacteria does? It takes those exogenously added amino acid, and takes it into its N. It does not make its own amino acid. Why? Because you have given it for free in the medium.

So, by doing that it takes the amino acid which we have supplied for example, alanine and all other amino acids, and start using it for its protein preparation. So, whatever protein is made in the cell is now, taken from these amino acids. So, this is basically how the protein will make the N 15 label because now the N 15 label amino acid is actually incorporated into the cell. So, that means, if I have am I amino acid sequence like alanine, lysine, lysine, arginine and so on, my only one amino acid N 15 will be this and all the other amino acids will be N 14, because I am supplying this in the unlabelled form and this is supplied in the labelled form.

So, this is how the selective labelling of amino acids are done. And you can see that if I want to have both N 15 C 13 labelling that means, I want this entire alanine to have both carbon 13 labelled atoms, and N 15 then I need to do by this type of amino acid, means where C 13 and N 15 both are labelled in that amino acids. These are specially made amino acids which we have to buy from commercial agents. So, from companies. So, once you buy those amino acids you buy these also but these are all unlabelled, but these are this is labelled. So, you mix all of these in this each is 100 mg per liter of culture and put it in the medium. But remember the medium does not contain glucose now, it is

replaced by these amino acids.

And what the bacteria does is simply one by one it takes all these amino acid into its cell, so all these amino acids which are present in the medium is they all get into the cell, and the bacteria now starts making all this protein using this amino acids, ok. So, these are, this is how the proteins are made because now the all the amino acids are available to it inside the cell. So, it takes this amino acids and makes our protein of interest which is actually also made in the plasmid which is sitting inside the cell also gets now this labelled alanine, also gets this unlabelled glycine, also gets this unlabelled arginine and all those things. So, that is what is very important to know.

And there is another thing which we will see in the next class that this is not so simple as it looks here. Sometimes when I take alanine into my cell it gets converted into another amino acid. For example, we saw in the previous few slides in the last class from a biochemistry perspective that amino acids interconvert. They convert from one amino acid to another. So, there is a danger that if I use want to label lysine, it may also label some other amino acid because inside the cell the bacteria converts 1 amino acid into another acid.

So, these are the little difficulties which we have to deal with when we do selective labelling because of this inter conversion. For example, let us say I am using glutamic acid, glutamic acid or glutamine is basically known to its transfers N 15 atom to all other amino acids. The all amino acids actually get their N 15 or N nitrogen from glutamine that is called transaminase. We will see that in the next class, and see how we can avoid such scrambling. We use the word scrambling of amino acids and that we will see how we can deal with and avoid for improved efficiency.