Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

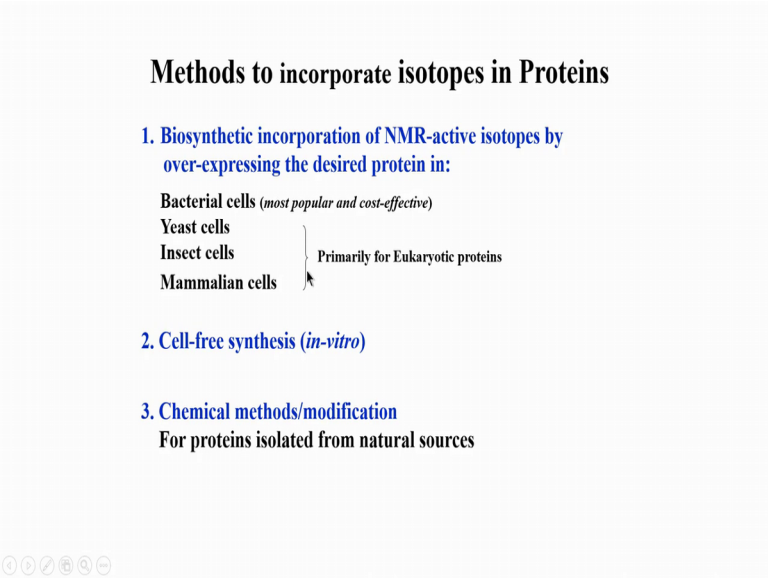
Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture - 23
Over Expression of Proteins in Bacteria

In the last class we started looking at isotope labelling. And, we saw that there are different methods different ways to do it. We can do the biosynthetic incorporation of isotopes which is done I shown here in bacterial cells yeast and mammalian cells. You can also extract the protein from insect cells; remember these are all recombinant DNA approaches. Meaning you take the DNA of your protein of interest and that is actually cloned into the organism. So, you are creating a copy of the DNA of the protein into the organism.

So, we can do this either in bacterial cells E coli which is the most popular approach or we can do it in yeast insect or mammalian cells. There is another approach where we do not actually need the cells at all, because when you grow cells you have to supplied with nutrients, you have to monitor the growth of the cells, you have to harvest the cells, and then you will also have to extract the protein from the cell and purify.

(Refer Slide Time: 01:29)



So, that is the approach which we did do in the cell based approach, that is in this the first approach then you use bacteria or yeast or mammalian cells.

But, if I can use another approach which is called as cell free synthesis in which we do not need any organism like here, which we saw yesterday briefly the idea here is that whatever is needed to make a protein inside a cell, those materials can actually be taken outside the cell and you can mix with chemicals and actually get the protein synthesised in isotope labeled form.

So, here you do not have the hassle or the problem of growing the cells and monitoring their growth, extracting from the cells and purifying. The protein comes out in a pure form of course, slight purification is needed, but not as extensive as here and also here you can make a protein of only your interest.

So, here what happens is if I grow a my protein or if I express my protein in a bacterial cell, what am I doing, I am actually over expressing or making my protein in large quantities inside the cell, which we will see shortly how it is done. But in addition to my protein the bacteria or the yeast or insect or any of the cells are actually making their own protein as well, because they have to grow. Said every step of their growth they are actually making not only our protein of interest, which we have cloned into it is DNA, but it is also making it is own protein also.

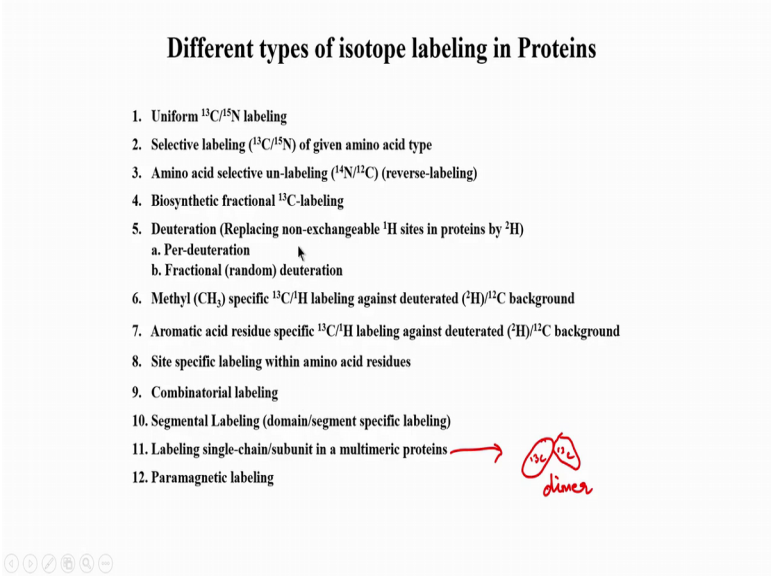
So, therefore, it is actually not very cost effective or not very efficient, because all we need is our protein of interest, but instead we are making the cell make all the proteins of it is own along with our protein. So, now, we will see shortly that this is not actually very efficient approach and this is also now in recent times is very useful approach. In fact, this is the only way to get proteins if you cannot express in any of these.

For example, there could be many proteins which are toxic to the cell they may precipitate inside the cell and so on. So, those kind of proteins you have to use this approach alone only. So, this is how basically cell free synthesis has become very popular in recent times. And with this another approach which of course, as I said not very rarely which is rarely used it is chemical methods of modification. So, you can actually synthesise a peptide or a protein, if you know the sequence. And if you know the sequence you can actually take choose your amino acids and accordingly also you can use labeling you can label it.

So, we will not go in detail in to this approach, we will basically look at these two in this course. And in this particular we will look at bacterial cells in more detail because that is

the most popular and cost effective approach whereas, yeast insect and mammalian become very expensive if you have to label the protein.

(Refer Slide Time: 04:31)



So, let us see how we can do this in bacteria. So, there are varieties of labeling schemes different types of isotope labeling which is possible. So, we can see there is this is just a dozen of them listed here, there are more in literature which are known. For example, you can do uniform C 13 labeling this is the most standard thing. Here what we do is in this approach, we basically the entire protein chain is labelled with C 13 and N 15 uniformly, which means every amino acid of the protein amino acid residue every atom carbon and nitrogen atom in every amino acid of the protein gets converted to C 13 N 15. So, there is no distinction between amino acid type.

For example, alanine versus, glutamine versus, glutamine acid etcetera everything is labelled uniformly across whole protein VC 13 N 15. So, this is the most first basic approach which people do. The second way of labeling is you can do selective labeling here. So, selective labeling basically means you choose an amino acid type which you want to isotopically labelled.

For example let us say I want to label only alanine in my sample why would you need to do that we will see that shortly, it helps to assign simplify the spectrum make this spectrum less complex or you can choose to label only let us say asparagine or arginine or lysine in a particular protein. So, in such cases we need to now do another approach

which I will we will see how it is done, but in such cases only that amino acid type means all arginines in a given protein will get labelled with C 13 N 15 if I want to label arginine.

If I want to label alanine, I will do in such a manner such that all the alanines in that protein and only alanines, no other amino acid in that protein will be C 13 N 15. All the other amino acids will be C 12 at natural abundance and C 9 protein natural abundance, but only that particular amino acid which we chose can be labelled selectively with C 13 and N 15.

So, we will see as we go along how we can achieve that? Third way of an selective labeling is actually the opposite of the first one, here opposite of this one, so, here we are doing the reverse. So, we use the word unlabelling; unlabelling meaning it is also actually referred to as reverse labeling in the literature.

What basically it means is here we are choosing the other way round, we choose a particular amino acid which we do not want to be labelled. So, which are let us say alanine again as an example. Let us say in all the alanine suppose I have 7 alanines in my protein in my protein sequence and suppose I desire to unlabelled only alanine means, I want to alanine to remain C 12 N 14 a natural abundance, but I want all the other amino acid in that protein to be labelled with C 13 N 15.

So, we can see this is exactly the reverse opposite here. We chose a particular amino acid to be labeled all the others to be unlabeled. Here we are choosing a particular amino acid to be unlabeled and all the others to be labeled. And again why would you need to do this we will see as we go along, it very much helps to simplify the spectrum not only that it is a very cost effective approach compared to this approach of selective labeling.

So, another option is fractional labeling and here what we do is here we want only about let us say 10 to 15 percent of the carbon in the entire protein to be labeled. So, it is not 100 percent labeling it is about fractional labeling fractional meaning about 10 to 15 percent, again there are various reason and when we need to do that and will see that soon.

Deuteration is the most next important isotope labeling approach in which we replace all non-exchangeable hydrogen in the protein by 2 H. Now, what are non-exchangeable

those which are attached to CH carbon a CH proton typically. So, these are nonexchangeable whereas, NH OH in the protein they get exchange they get exchange with water whereas, CH CH 2 CH 3 they can be deuterated, so, therefore, that is one approach.

So, there are again different types of deuteration in a given protein we will not look at it now will see it when we come to the topic. And there is another labeling scheme where only the Methyl's are specifically protonated, but all the other protons are replaced with deuteron and C 12.

So, you see here what we are going to do is only the Methyl of every an amino acids, which are the amino acids which are Methyl's is alanine, methionine, (Refer Time: 09:26) isoleucine, threonine and so on. So, these are the amino acid which have methyl groups for some of the this methyl groups we will keep it as CH 3; that means, we will keep C 13 here and proton here. But all the other hydrogens in the protein will be changed to 2 H.

And, why do we need to again do that the reason is Methyl's are very interesting moiety's in proteins they are very strong in NMR. So, their signals are very highly intense. So, very useful to have Methyl's as CH 3 that is carbon 13 CH 3, but all other protons are not. So, useful and we can get rid of them by deuteration. So, we can retain only the methyl's as protonated and all the protons.

Again remember when I say all the remaining proton I am meaning only the nonexchangeable. The exchangeable that is amide OH NH they remain always protonated I am assuming they remain protonated. So, that and Methyl will now get remain protonated in this scheme number 6 whereas, all the nonexchangeable CH carbon protons are now replaced with these 2 that is this is the natural abundance and deuteration for that particular proton.

So, we can do this as I said for several other types of labeling. Similarly, we can do the same thing for aromatic amino acid. In aromatic amino acid I can have specific C 13 proton labeling whereas, the remaining background; that means, all the aromatic amino acids will have carbon 13 proton in their side chain, but all the other remaining protons will get deuterated.

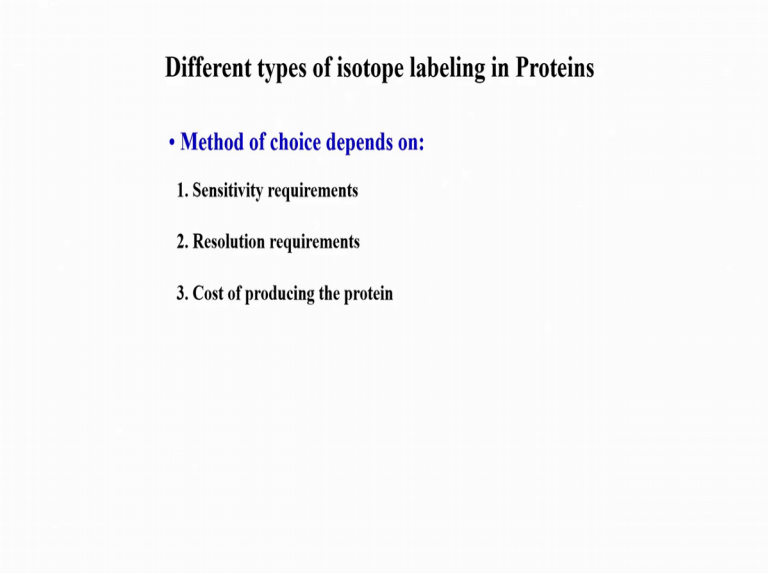
So, this is basically another approach where we can use this, there are other combinations other labeling approach which we will not go into detail in this course. Because these are more advance level of isotope labeling schemes we will basically try to cover this first 6 or 7 of them as we go along.

There is another very interesting thing is here number 11 here what we do is suppose you have a multimeric protein. For example, you have a monomer dimer or a trimmer of a protein. So, what we can do is let us let me show you this with the picture here. So, I am referring to this labeling here. Suppose, I have a protein which is existing as a dimer so, this is a dimer.

What I can do is I can specifically label this part is C 13 and keep this as C 12 that is natural abundance; so, I can do that trick. So, once I have this one portion I see 13 other one as C 12, then there are lot of experiments I can do to look at this molecule. So, this is another way to label a single chain; that means, this is one chain this is another chain, they are joined by noncovalent interaction because it is a dimer. And one of the chain only I am specifically labeling with C 13 and the remaining I am leaving it as e 12. So, this is possible to do with different approaches we will see that and so on.

So, we let us start from the very basic first one that is uniform C 13 N 15 labeling how do we do that in again remember we are looking at e coli. So, all these labeling schemes will be done in e coli. So, we will see how we can do that in bacteria.

(Refer Slide Time: 12:47)



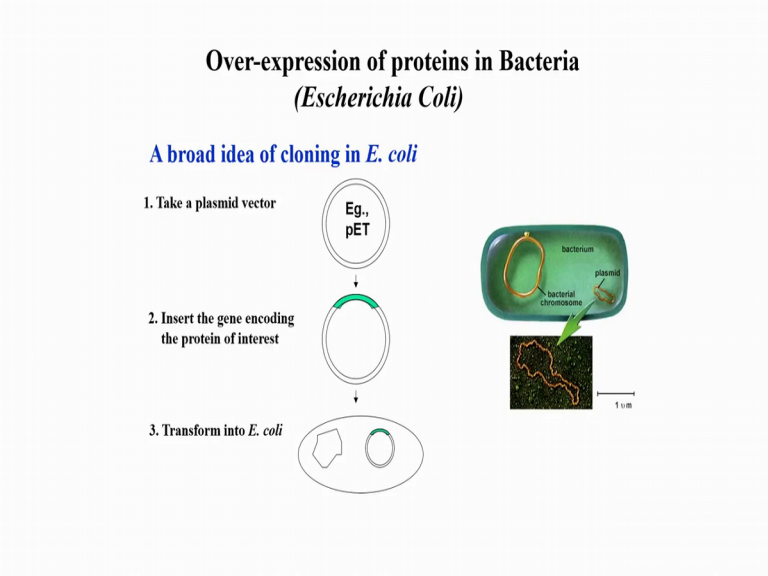
So, the question is out of this all these labeling schemes which I showed earlier how do you decide which one is good for you. So, all depends on these 3 factors number 1, what is that sensitivity are looking for?.

So, if you want a very good sensitivity, then basically you have to choose a particular type for example, you are looking at very high sensitivity you have to deuteration is the best approach, you have to deuterate all the nonexchangeable. So, that you get rid off any relaxation contribution and that is helps to do that or you can look at resolution suppose you want to have only a certain signals, you want to suppress signals from some amino acids, and you want keep signals only from some other amino acids selectively. So; that means, you want to resolve the signals from different amino acids, then you need the selective labeling or selective unlabelling scheme which I mentioned. So, that helps you in getting good resolution.

And 3rd thing is a cost it is actually remember this cost of labeling is very high. It is not very easy to it is not very inexpensive it is expensive to label proteins with carbon 13 and 15 and deuteration. In fact, if you want to deuterate in Indian currency today, it can run into lakhs of rupees for making one sample. So, you can see it is a very costly approach, but sometimes there is no choice for us we have to spend that money to make our protein otherwise we would not be able to study that protein.

So, cost also matters when it comes to what kind of labeling scheme you want to choose from the list which I showed in the previous slide.

(Refer Slide Time: 14:25)



So, now let us begin with the very simple idea of how we can express protein in bacteria? So, this is a general scheme this is nothing to do with what kind of labeling you are going to do isotope labelling, it is a broad idea of this approach. So, this is called cloning or we can also use the approach recombinant DNA approach.

So, what we do is the following. So, this is a picture of E coli cell. And E coli if you recall it has 2 types of DNA in it is cell; one is called the chromosomal DNA and other is called the plasmid DNA. This plasmid DNA is a small circular DNA, which you can see looks although not exactly circular, but it is a close loop. So, this is the electron micrograph of a real plasmid.

So, now that plasmid is shown as a circle here ok. So, what we do is typically we take this kind of a plasmid we use the word plasmid vector is the word used. So, let us say this is an example what is called a pet vector. These are different vectors available in literature in commercial market. So, one can choose a particular vector again depends on many factors.

So, we let us say you have chosen one vector of your interest or a plasmid of your interest what we do next is we make a DNA a gene which encodes our protein. So, now, remember a protein comes we know the central dogma of biology, DNA is transcribed to RNA, RNA is translated to protein. So, basically proteins come from finally, from the DNA.

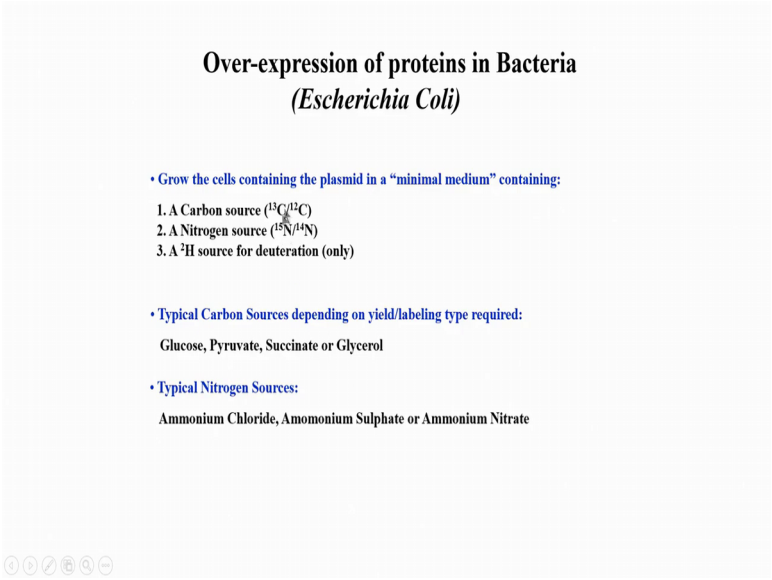
So, DNA means you have the code genetic code for a DNA. So, sorry for a protein say every amino acid in a protein is coded by 3 nucleotides in the DNA that is the triplet codon principle.

So, we what we do is we take the amino acid sequence of our interest. So, remember we should know the sequence this is something I have been telling in the previous class as well, the most important requirement for any NMR studies of a protein is that you need to know the sequence amino acid primary sequence of that protein, without that we cannot do anything much. So, once you know the protein amino acid sequence you can generate a gene or a DNA corresponding to that protein; that means, it will be the complimentary; it will create a DNA which will code for that protein using the triplet codon idea.

So, you take the gene the DNA what you do is you insert into this vector what so, what we do is we break this here ok. So, we can break it at some place particular place it is not arbitrarily, it is depending on where the operator and so on are located, but you break the DNA you insert the gene of our interest our protein into the cell sorry into the plasmid. And, then you can join it like this this is called ligation; ligation means you have joined the two ends.

Now, once you have joined the plasmid now looks like this and our DNA of interest is sitting in the particular location. So, what we do is take this plasmid and we put it inside the cell like this here. This is called transfection or transformation. So, we transform this plasmid into the cell it basically the cell, now contains it is own circular I mean it is own chromosomal DNA plus it contains the DN plasmid which we have inserted into the cell. Now, what you do next with this? Now, we have the cell containing our gene of interest ok.

(Refer Slide Time: 18:03)



So, now next step is you grow the cells in a particular medium ok. So, this particular medium we call it as a minimal medium, this is required for isotope labeling some the minimal medium basically contains a carbon source. So, remember a carbon is energy source we need the bacteria to grow.

So, if you want something to grow it needs energy and energy basically comes from carbon proton glucose essentially from glucose. So, bacteria can grow on glucose. So, you need a medium; that means, in which it has to grow and that medium will should contain a carbon source, typically we use carbon glucose that is glucose.

Now, that source can now we can buy from company's either labelled carbon or unlabelled. So, unlabelled basically is standard glucose which is available very routinely we use in the laboratory or you can buy a special glucose which will have C 13 labelled so, that is carbon 13 labelled glucose which also is available commercially. Of course, this is the expensive part this is not very this is a lot expensive compared to standard glucose that is unlabelled, but if you want to labelled with C 13 we have to purchase this carbon 13 labelled glucose.

So, that will be the source of carbon means any carbon in the whole of the cell will come only from that glucose. There is no other source of carbon which will be given to the bacteria.

So, every carbon in that bacteria whenever it multiplies has to come from this carbon; that means, every carbon will be C 13 because the glucose is itself C 13. So, when glucose itself is C 13 labelled every carbon in the cell of the bacteria will start getting labelled. You also need a nitrogen source because remember we the said the ammonia that NH 2 the amide of every protein in the cell has to come from nitrogen. So, how do we do that we have to supply; that means, our medium should contain a nitrogen source.

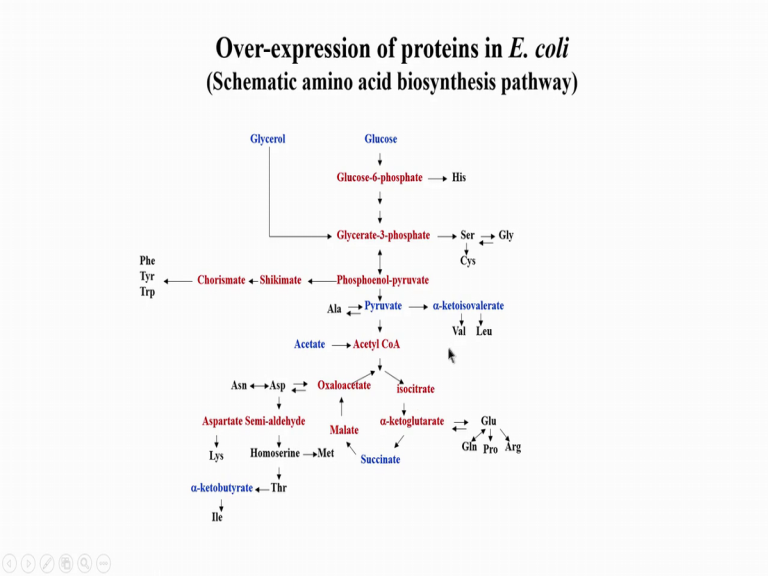
So, what are the typical nitrogen sources are listed here is ammonium chloride ammonium sulphate or ammonium nitrate. Again remember these have to be in the labelled form means I should take N 15 ammonium chloride, N 15 ammonium sulphate, or N 15 ammonium nitrate, if I want N 15 labelling. But if I do not care for N 15 I want a normal protein with N 14 abundance, then I can choose a regular ammonium chloride or ammonium sulphate or ammonium nitrate ok.

And, if you want the deuteration to be done additionally then you need a source of 2 H. And, what could be that source the source there are 2 sources; number one the glucose also should have the protons in the glucose should be deuterated. So, CD glucose carbon 13 labelled deuterated glucose, and second is remember the bacteria the minimal medium is made in water that is H 2 O. So, if I want to deuterate I need to now make use D 2 O not H 2 O, because D 2 O is the source of deuterium. So, this is how we can do for deuteration we will see that as we go along.

So, this is a list of different carbon source of carbon which are used in labeling some in bacteria the most popular is glucose or pyruvate is also use certain times, but succinate and glycerol are not so, popular. So, we will stick to glucose in our case when we look at our examples. So, again I repeat the glucose which we use has to be either C 13 labelled, if you want carbon 13 labelling, or it can be regular D glucose which has C 12 as the natural abundance ok.

So, these how we have to basically go about we have to take this I go back to the previous slide, we take this cell ok. The of course, we would not have a single cell, we will have a lot of cells, colony of cells and that cells are now grown in a medium. So, we put the cells in a medium and start growing them.

(Refer Slide Time: 22:11)



So, let us go through the steps one by one. So, what is the point here; the point here is how are we able to make those labelled protein is because look at this chart here. This is a very standard schematic biosynthesis pathway taken from biochemistry textbook you can find this in lehninger, Biochemistry by Lehninger or Biochemistry by Stryer, these are the 2 most popular biochemistry books. The idea here is that you have you start from a labelled glucose source that is the or unlabelled that is the energy source and it gets now in to different pathways throughout the cell.

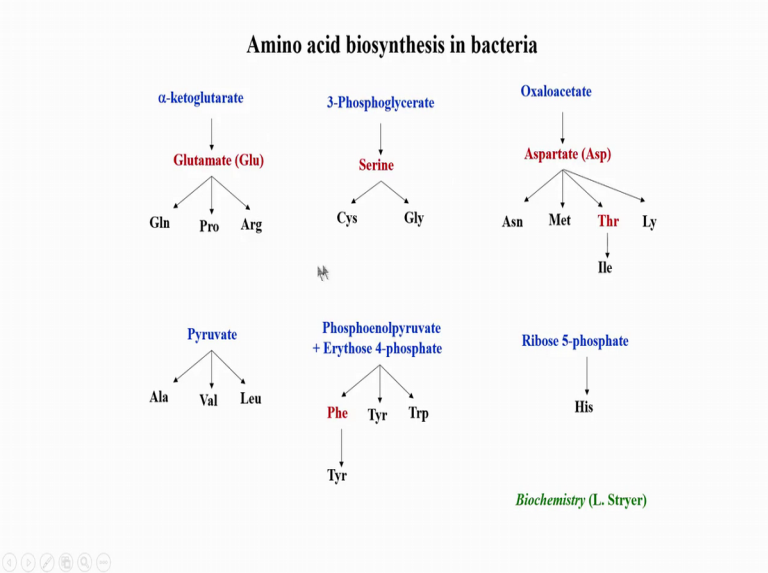
And you can see the different amino acids gets synthesise based on the different intermediate metabolites. So, these are all metabolites and these are all metabolic pathways. So, you can see that alanine is coming from pyruvate and so on, serine Glycine's cysteine are inter convertible here and aromatic amino acid comes from shikimate pathway, which is a very specific pathway and so on.

So, one has to understand one has to know this pathways. In case suppose let us say I want do not want the protein to make aromatic amino acid. I do not want the bacteria to make aromatic amino acid I want to supply it from outside. In such a case then I have to stop this pathway here right. So, this kind of approaches are needed to be known so, that we can see which pathways result in which amino acid.

So, that we can control that pathway and we can control the synthesis of those amino acids ok. So, this is very useful when you look at isotope labeling of selective amino

acids that is selective labelling.

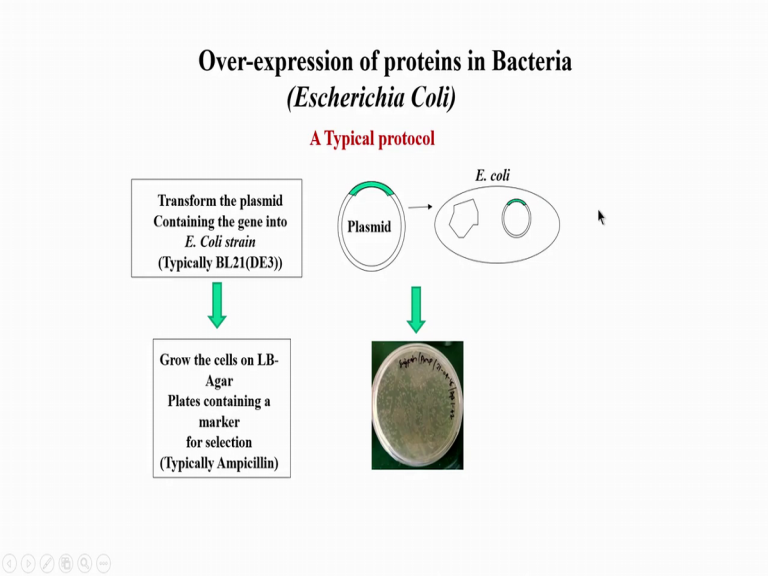
(Refer Slide Time: 23:51)



So, here is another example a few more examples of what are the different metabolites intermediates. So, these are intermediate metabolites shown in blue colour, which are generated in the cell and then they give rise to amino acids. And each of these amino acids further they get converted into different amino acids. So, you can see this here phenylalanine tyrosine, phenylalanine can give rise to tyrosine or it can directly come from here. Similarly, isoleucine and threonine is a precursor for isoleucine especially the delta carbon and so on. So, this has been taken from this book biochemistry by one stryer.

So, the idea here is basically to understand that how amino acids are synthesise that is biosynthetic pathway in the bacteria. So, that we can control how we want to label, suppose I want to label only glycine, can I do that then in that case I have to label serine because serine will give rise to glycine. But, now if I label glycine serine also gets labelled. Similarly, cysteine also will get labelled if I try to label serine. So, you can see that selective labeling very much depends on how selective are these pathways ok. So, we will see that as we go along and we go in deeper into this.

(Refer Slide Time: 25:05)



So, this is let us start from a now a typical protocol. Typical a protocol basically for over expressing the proteins in bacteria and I can our ideas is for labeling here. So, we will see from the labeling point of view what, what is the procedure that we follow for isotope labeling for NMR studies.

So, you can essentially start from as transformation transformed plasmid. So, this is what I mention, that you take the plasmid vector of your choice, you put the DNA of our interest in to insert that into the plasmid, and after insertion you take this entire plasmid put it into the cell inside the cell that is the transformation. So, we chose a typical bacterial strain which is written here BL 21 DE 3 you can refer to all the books and they will show this particular strain why and so on there are many reasons, you can refer to the textbooks for those details, but we that strain is a typically the popular strain which is used.

So, once you have this bacterial cell containing this plasmid and next step used to basically plate it on a an agar plate. So, we can see here if you look carefully, you can see this white spots here. So, what are this white spots these are the colonies; colonies means collection of bacterial cells these are not individual cells individual cells of course, we cannot see with our naked eye, but this is a collections. So, we can see what has done is this is the plate, you can see this is a petri dish. On a petri dish we have an agar media and the same medium that is our minimal medium on the agar plate ok. And, that and the

bacteria is grown on that plate and what does not grow like a liquid state it grows like colonies it develops into small colonies.

And, now one thing is here very important point is that you have to you need a marker for selection, what do you mean by a marker? Marker means, suppose I am growing this bacteria cell in a medium. How do I prevent the other competing bacteria in the same medium to grow, because remember bacteria in our in our world is very ubiquitous. Means everywhere bacteria is there it is there in our body it is there on our hands on our everywhere in the in the in the in the laboratory. So, whenever I put a medium I take a medium. Even though I will sterilize the medium, means I will autoclave the medium, I will make a sterile still bacteria can come from anywhere it can come from the air and so on.

So; that means, if I am growing the protein sorry the cell of my interest, which has my plasmid I can still have to prevent any other bacteria coming from contamination to grow, I should not allow the contamination to grow. So, how do I achieve that can be achieved by using a specific marker meaning, let us say I take an ampicillin and antibiotic ampicillin is a well-known antibiotic. Antibiotic means it kills microbes or bacteria.

So, if I take an if I in this DNA suppose I have a another marker; that means, ampicillin resistant gene is there for the bacteria so; that means, my bacteria becomes resistance to ampicillin, but the one which is the other contaminant may not be resistant to ampicillin. Because, that is coming from natural sources it is not engineered like this this is an engineered bacteria.

So, you see actually in a way we have engineered it to become resistant to ampicillin which we can do very easily, but the other cells which are not part of this are contaminations are not actually having resistance to ampicillin. So, when I grow in a medium which contains ampicillin like this agar plates the ampicillin will kill all the contaminants. It will take get rid off all the contaminants and only the bacteria of my interest, that is the cells which are having this plasmid which is contains the DNA of my interest.

And it of course, it is resistant to ampicillin only that those cells will develop and may become colonies, but all the other contaminations are gone. So, I can be very sure that all

these tiny spots which I am seeing here are actually coming from the bacteria of my interest which contains the plasmid of my interest.

So, therefore, that is one that is the call concept of markers for selection. So, there are different antibiotics which can be made resist use, but we will assume that in the standard approach this is ampicillin resistance. So, we will now continue in the next class on the further steps which we need to do to purify and extract the protein and how to label it.