Molecular Biology

Prof. Vishal Trivedi

Department of Biosciences and Bioengineering

Indian Institute of Technology, Guwahati

Module - 10

Molecular Cloning

Lecture-40 Cloning (Part 1)

Hello everyone, this is Dr. Vishal Trivedi from Department of Biotransition and Bioengineering IIT Guwahati. So, in today's lecture, we are going to discuss about the enzyme production. So, enzyme production as the name suggests is where you are actually going to use the where you are going to prepare the large amount of enzyme and we are there are two approaches one can actually be able to use. First approach is that one can actually be able to purify the enzyme purify the enzyme from the rich sources. For example, if I have to purify the enzyme of the glycolysis for example, if I have to purify the enzyme of hexokinase, then what I can do is I can just isolate the cytosol of a cell and then I can use the different types of conventional chromatography techniques and I can be able to get the pure enzyme.

Same is true for many other enzymes like I can use for lysosomes, I can use the lipase, I can use any other method. So, in this particular approach, you have to go and first identify the rich source. For example, if I have to purify the lysosome, lysosome is very heavily available and in many of the biological fluids. So, I have to identify those biological fluids and then only I can be able to use the different chromatography techniques to get the pure enzyme.

The second approach, which is more straightforward and which actually gives you the, you know, the enzyme with lot of ease is that you can actually be able to identify. So, enzymes are made up of the proteins and the proteins are being synthesized from the RNA, which is actually going to be coded from the DNA. And this DNA is nothing which is also called as G. So, for example, if I have to work or I have to use an enzyme for my experimental applications or if I want an enzyme for structural elucidations or if I want an enzyme for the industrial applications, then I can just first do the enzyme, I should identify the corresponding gene. And once I identify the corresponding gene, then what I can do is I can just clone this gene into a suitable vector.

And once I have the gene into the suitable vector, I can do the over expression and I can be able to purify this protein with the help of the conventional chromatography and as well as the specialized chromatography. So, once I purify the protein, I am going to get the pure enzyme and that pure enzyme I can be able to use for many applications. For example, if I

need an enzyme for studying the DNA recombination or if I need an enzyme for running the transcription, translations, replications, all that, so that all I can be able to prepare in large quantities. And so that you can be able to do because there is no, once you have done that into a, once you have cloned the gene into a suitable vector, then you can actually be able to, you know, transform that into a suitable host and that is how you can be able to make the proteins in large quantities. So, this is what we are going to discuss in this particular module and in the subsequent module as well, that how you can be able to use this kind of approach to produce a large quantity of enzyme because this enzyme actually be able to use for many applications such as we can use it for the structure elucidations, we can use this for catalyzing the reactions and we can also use this for the industrial applications.

Now, if I have to go with the approach number 2 or if I have to go with the approach number 1, we have to have a very clear understanding because the, see with what is the drawback of the approach 1 is that first of all you have to go with and identify a rich source, the second is you have to purify this protein with utilization of the different additional chromatography technique and even then the production is going to be very, very limited because it depends on the amount of enzyme which is present in this particular rich source. Whereas in the approach number 2, once you have identified the gene, you can actually be able to overexpress that into a very large quantity and also since you have the gene, you can be able to do the fine tuning. For example, what is mean by the fine tuning is that suppose I am working with an enzyme and I want to identify its structure with a larger objective is that I want to study how this enzyme is interacting with the substrate and how it is actually forming the complexes with the substrate and product and so on so that I can be able to get the inner insight into the enzyme catalyzed reactions. In that case, I might actually require the enzyme alone and I might also require the mutants, mutant of those enzymes so that I can actually be able to study the enzyme, how the enzyme is interacting with the substrate and then I also can test the mutants. The mutant approach is not going to work when you are going to work with the approach number 1, but the mutant approach will work in the approach number 2 because you have the gene and you know at what point you are actually trying to make the mutations and that you can actually be able to insert into this particular gene and that is how you can be able to change the overproduced enzyme.

So this is one of the major advantage of going with the approach number 2. So in the approach number 2, what you are going to do is you are going to first identify the gene. So the first step is that you are actually going to identify the gene and that you are going to do by screening the different types of clones that we are anyway going to discuss in detail. So that you require a screen, right? Because you have to screen the gene in a pool of genes so that you can be able to identify the gene which is responsible for the production of the enzyme X for example, okay? So once you identify the gene, then you are actually going to clone that gene into a suitable vector. This cloning also has the multiple steps.

In the step number 1, you are going to amplify this gene with the side-specific, with the help of the side-specific primers and these side-specific primer are actually going to have

the restriction enzyme. So once you have the restriction enzyme, they will be able to, you can be able to cleave that amplify gene and that is how it is actually going to produce the sticky ends. And these sticky ends are then going to be, you know, be facilitate the entry of this particular gene into a suitable vector and that is how you are actually going to produce the clone. So in the step number 2, you are going to do the restriction digestion so that the sticky ends are going to be generated and then you are actually going to clone that into a suitable vector. Once you clone it into a vector, then you are also going to do a screening of these clones so that you can be able to select the suitable clone.

So you are going to use the suitable clone. Now once you got the suitable clone, then you are going to use that suitable clone for overexpression. So in the step number 3, you are going to do the overexpression. In the overexpression, the first step is that you are going to choose the host cells. Host cell means the cell where you are actually going to make the proteins.

It could be very simple such as the bacteria or E. coli. It could be as complicated as the mammalian cells. So depending on the host cells, it can be yeast, it could be insect cell lines and all that. So these are the multiple options what you have at this stage when you want to use the host cells.

And choosing the host cell is very crucial because that is going to decide the final production. Then once you have selected the host, you are going to do the transformation of this particular clone what you have produced in the step number 2. So this is the step number 2. This is the step number 3. And once you do the transformation, after that again you have to do a screening.

The third step, you are going to do a screening to select the suitable transformed clone. And then you are going to do overexpression. For overexpression also you are going to have the different strategies depending on what host you are going to use. And then ultimately you are going to do the analysis of the overexpression which means you are going to see whether the protein is been, protein or the enzyme is being produced or not.

Okay. So once you are confirmed that the enzyme is getting produced, then you are actually going to make the large cultures, you are going to produce large amount of host cells and that is how you are going to have the large amount of proteins. But these proteins are actually also going to have the host proteins. So when you overexpress you are going to have two proteins, the enzyme X, which is going to be in an overstressed state and then you also going to have the host proteins. This means you are going to get a mixture of the enzyme X and the host protein after the overexpression. So then you are going to enter into the step number four where you are going to do the purification so that you can be able to separate out the enzyme X from the host protein and that is how you are going to get the pure enzyme X.

Now when you want to do the purification, you also again going to do the multiple step. In the step number one, you are going to do the cell disruption. In the step number two, you are going to do or you are going to perform the different chromatography techniques such as ion exchange chromatography, hydrophobic interaction chromatography, gel filtration chromatography or sometime you might also do the affinity chromatography. Once you are done with the chromatography, then you are going to do the estimation of the or the yield of your purification or the yield of what amount of enzyme X you have produced and then you are also going to do the further analysis for seeing how much is the purification. So, all these you are going to do.

So, all these four steps are required if you want to produce the enzyme under the approach number two. So, let us start discussing about these approaches and these steps one after the other. So, today we are going to discuss about how you can be able to identify the gene of corresponding to the protein X or enzyme X. Now there are different approaches for isolation of the gene of interest like for example, in this case, we want to identify the gene which is responsible for the enzyme X. So, what are the things you can actually be able to use as the starting material? So, if I have the enzyme X and if I so, it depends on what kind of phenomena we what kind of you know information we have.

So, for example, if the gene sequence is not known, for example, this is one condition where the gene sequence is not known. If the gene sequence of this enzyme X is not known, then you are going to use these two approaches and if the gene sequence is known, then you are going to use this approach. So, in that if the gene sequence is not known, you are actually going to know either two information one, you are going to have a gene fragment, which you are sure that it is actually expressing a part of enzyme X or you are actually going to have the messenger RNA or you are actually going to have the active enzyme, which is going to be active. So, if you have the gene fragment, you can be able to go and look at into the genome of their that particular organisms. So whether you want to use the genome or whether you want to use the expression profiling, in both of these cases, you might have to prepare a genomic library.

So, in this case, when you want to look for the genome, you have to prepare a genomic library. So, genomic library is nothing but a collection of the clones, collection of clones corresponding to different genes. Similarly, so all the genes, whether these genes are expressing and giving you the protein or not, they are actually going to give you all genes. Whereas when you are going to go for the expression analysis, you say that you have the active enzyme but you do not know which gene actually, then you can actually be go with the transcriptome or you are actually going to go with the cDNA library. So, when you are going to go with the cDNA library, cDNA library is actually a collection of the X gene expressing in a cell.

So, all the genes, so this is actually a expression library, this is the genomic library. So, it is actually going to take care of the all the genes, which means genomic library will contain

the cDNA library, but cDNA library will not going to contain the genomic library. This means the seed whatever the genes you will see under the cDNA library will also be present in the genomic library, but whatever the genes you will see under the genomic library may or may not present in the in the cDNA library. Once you got the library site, just like as you have the library in your college or you know in your institution, and you want to see that, I want to today I want to study a book which is about the protein biochemistry or protein structure determinations, then what you are going to do, you are going to take these keywords and you are going to put it into the software. And what the software is going to say, software is actually going to go through with the your institutional library and it will actually going to tell you, these are the books which are available which is related to the protein structure determinations.

So, then you are actually going to screen out the suitable book actually. Similarly, these are the collection of the clones, these are the collections of the genes, one is for the whole genome, the other one is only for the expressions of expression status of the protein, expression status of that particular organism. So, then the second step is, so in the first step, you are actually going to prepare the genomic library or the cDNA library. In the step two, you are going to screen the library with the help of the gene fragment or the active enzyme what you have with you and then you are actually going to use the screening. In the step three, you are going to take out or you are going to isolate these gene fragments and then you are actually going to clone it into a suitable vector.

And that is how you are actually going to get the final clone. So, this is all about the approach when you do not have, when you do not know the genomic sequence, which means the pre genomic era, before the pre genomic era, people were actually going with this approach where they are actually only having the, they have isolated an enzyme with the help of the traditional approach, but they do not know gene sequence. So then they actually go with these kind of approaches. After the post genomic era, when the people actually know the genomic sequences, then you are very, because you know the genomic sequences, you can actually be able to use a technique which is called as polymerase chain reactions. And that will, it is actually going to give you the amplified gene product and that you can be able to clone it into a suitable vector.

And that actually can you can be used further into the over expression and purification steps. So, first we will discuss about the approach number one where the genomic sequence is not known. And then we are going to discuss about the gene, the approach where the gene sequence is known. Now the first approach is when how to prepare the genomic library. So preparation of genomic library is a multi step process.

In the step one, you are going to isolate the genomic DNA of the host or of the, you know the, of the organism from which you are interested to isolate the enzyme. Then you are going to do the restriction digestion of the genomic DNA or in, or in general what you are going to do is you are going to make the fragments. Then the step three, you are going to

purify these fragments so that you can be able to use that. And then you are going to ligate these fragments into a suitable vector. So in this case, we have taken an example of VAC, YAC vector.

And then in the step five, all these are going to transform into the yeast so that you are going to get a library of clones. And once you got the library of clones, that is actually going to give you a genomic library which is actually going to represent the, all the genes which is present in this particular organisms. So how you are going to do this? You have, what you are going to do is you are going to, first what you are going to do is you are going to isolate the genomic DNA or the human genome, because we are talking about the preparation of the genomic library from the human. And then you are going to digest that human genome. So all the chromosomes are going to be digested with the help of the excision enzymes.

And that is actually going to give you the different fragments, all these fragments, which actually contains one or more genes in each fragment. Ideally, you are going to produce the fragment so that it represent at least one gene, it may have the additional fragments or the 50% fragments more from the other genes. And the same way you are actually going to take the YAC vectors. So YAC is the yeast artificial cloning vectors. And that actually you are also going to have the EcoR1 site.

So you are actually going to digest that with the EcoR1 and BamH1 and that is actually going to give you the linear fragments. And then what you are going to do is you are going to do the ligation. So once you go to ligation, all these fragments randomly and are going to be get ligated into the YAC vectors. And then you are going to do the yeast transformations and that is how you are going to get a clone. So this is one clone, this is second clone, this is third clone and so on.

So these are the different clones and all these clones are actually going to constitute the genomic library. So let us discuss about each and every step how you are going to first prepare or how you are going to first isolate the genomic DNA. So in the step one, you are going to do the isolation of the genomic DNA. So isolation of the genomic DNA is a multi-step process. In the step one, you are going to do the lysis of the cell with a detergent called lysis buffer.

With a detergent containing lysis buffer. So if the cells are cultured cells or the tissue, you are actually going to do the homogenization and that is actually going to give you the single cell. So once you do the, if it is like tissue, for example, if it is a liver or spleen and you want to prepare the genomic library, then you are going to use, first you are going to homogenize these tissues so that they will actually going to give you the single cell suspension and then these homonized projects or the single cell is actually going to incubate with the lysis buffer so that it is actually going to lysis the cells. So you are going to get the lysate and this lysate is actually going to contain three components. One, it is actually going to contain the genomic DNA or gDNA. It is actually going to contain the protein and it is also going to

contain the lipid and so and it may also contain the RNA which is also present in the cell.

So these are the four material what you have and you are only looking for the genomic DNA. So what we are going to do is we are going to do the purification steps. So incubation of cell with the digestion buffer containing the proteinase K, SDS will release the genomic DNA from the DNA protein complex and this genomic DNA which you are going to get just after the lysis is actually going to be present as a complex with the protein and you are going to incubate that within a protease and that protease is actually going to chew up all the proteins and that is how your genomic DNA is going to be released. Then the step three, you are going to isolate the genomic DNA by the absolute alcohol precipitation. So we are going to precipitate the DNA but this DNA is also going to contain the contaminating agents like the proteins, lipids and RNA.

So it is actually going to not only going to contain the purified genomic DNA, it is going to contain the other molecules. So in that case what you are going to do is then you are going to do a purification of the genomic DNA with the phenol chloroform mixture and chloroform mixture has two phases. One is aqueous phase and other one is a organic phase. This is what you see right, it has an aqueous phase and organic phase. In this step the phenol denature the remaining protein and keep the protein in the organic phase.

So protein is actually going to be separated out and it will go into the organic phase. So the genomic DNA, the blushed colored phase what you see is actually going to be isolated again and it is actually going to be precipitated with the help of the absolute alcohol. Now, in the last step because it is a genomic DNA, so you also have to test that the genomic DNA is perfect. So you are going to analyze the genomic DNA on 0.8% agarose and a good preparation of genomic DNA give a intact band with a no visible strain.

So this is what you see, this is genomic DNA, this is the control where no genomic DNA and this is the genomic DNA. So what you see here is that we have the intact one band and that migrates very slowly because genomic DNA is very big. So it does not migrate very fast. So this is all about the genomic isolation of the genomic DNA and then once you have generated genomic DNA then the step two is the generation of the suitable size fragments.

Suitable size fragment can be done by the two methods. The first one is a restriction digestion method where genomic DNA can be digested with a frequent DNA cutting enzyme such as the EcoR1 or the BamH1 or SAW3A to generate the random sizes of genomic DNA. The criteria to choose the restriction enzyme or the pair of enzyme in such way such that the reasonable size DNA fragment will be generated. What is reasonable size is that it should contain at least one gene. It may have a flanking sequences that is not an issue but it should at least contain one gene in one particular fragment. So that it should not be the case that it you fragments are so big that it actually contains multiple genes because in that case it is actually going to you know fail the purpose of making a genomic DNA.

Ideally the genomic library means that you every clone which you are going to make should represent only the single gene or it should may have some flanking genes but only very fragments. It may have the fragments of other genes but at least it should have one gene. It should not have like four or five or six or more genes. So that is why you have to generate the fragment which are reliable and relatively big enough. It is likely that each and every genomic sequence is represented in the pool.

As size of the DNA fragment is large the complete genome will be presented in very few number of clones which means the clone number should be very high so that it should represent the all the genes which represent in that particular genome. So that is the step number the method number one where you can do the restriction digestion. In the step number two you can also do a mechanical sharing. So what you can do is just vertex the genomic DNA and when you vertex the genomic DNA it is actually going to break the genomic DNA into the multiple fragments. So genomic DNA can be fragmented using a mechanical sharing but restriction digestion method is more popular and more precise because here you are using an enzyme so enzymes are going enzyme mediated events are going to be more and more regular and more controlled.

Now for example how much would be the fragment size. So for example if an organism has a genome size of 2 to power 10 to power 2 into 10 to power 7 kB and an average size of fragment is 20 kB right. Then number of fragment would be n is equal to 10 to power 6. So you divide this number by this number you will get this number. In reality this is the minimum number of clones to represent the fragment in the library where the actual number is much larger it almost be a 10 times the number what you have to put or what you have to get. So the probability of finding a particular genomic sequence in a random library of n independent clone is n is equal to lin 1 minus p divided by lin n minus 1 1 by n where n is the number of clones what you have in the genomic library p is the probability of finding a fragment into the genomic library and n is the size of the average fragment size.

So you can use this particular formula to calculate how many clones I should have to generate and that actually is going to increase your probability of finding the clone or finding your gene into a genomic library. So once you generated the suitable size fragment then it has to be ligated into the suitable vector. So in the step 3 you are going to use the cloning into the suitable vector. So depending on the size of the you know the size of the different organisms for example the mycoplasma bacteria or like the flowering plants you have to choose the suitable vector. So cloning into the suitable vector the suitable vector to prepare the genomic library can be selected based on the size of the fragment of the genomic DNA and carrying capacity of the vector.

Size of the average factor can be calculated and accordingly a suitable vector can be choose. In the case of the fragment generated by the restriction enzyme the vector can be digested with the same enzyme and put for ligation to get the clone. In the case of mechanical sharing mediated fragment generation putting these fragment needs additional

efforts. In one of the approach adapter molecule can be used to generate the sticky ends. Alternatively, endonubrious can be used to generate the sticky ends.

So these are the some of the you know the vector carrying capacity of the some of the vectors or the fragment carrying capacity of some of the vectors. For example the plasmids they can actually be able to carry the 15 MB phage lambda, cosmids, bacteriophage, back and the yeast. So depending on the size of the fragment what you are going to get from a particular genome you can be able to use the different vectors. Then you are in the step 4 you are going to do the transformation to get the colonies. So the post ligation the clones are been transformed into a suitable host to get the colonies.

A suitable host can be bacterial stain or the yeast. Different methods of delivering a clone into the host is discussed in a future lecture. So when we are going to discuss about the over expression that time we will discuss about the transformation how the different methods can be used to deliver the DNA into the different host cells like the bacterial cells, yeast cells, mammalian cells and so on. So this is all about the genomic library approach where you have now we have just prepared the genomic library. And now we will move on to the next approach where we are going to use the cDNA library. So that we are going to discuss in our subsequent lecture. Thank you.