Enzyme Science and Technology

Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati

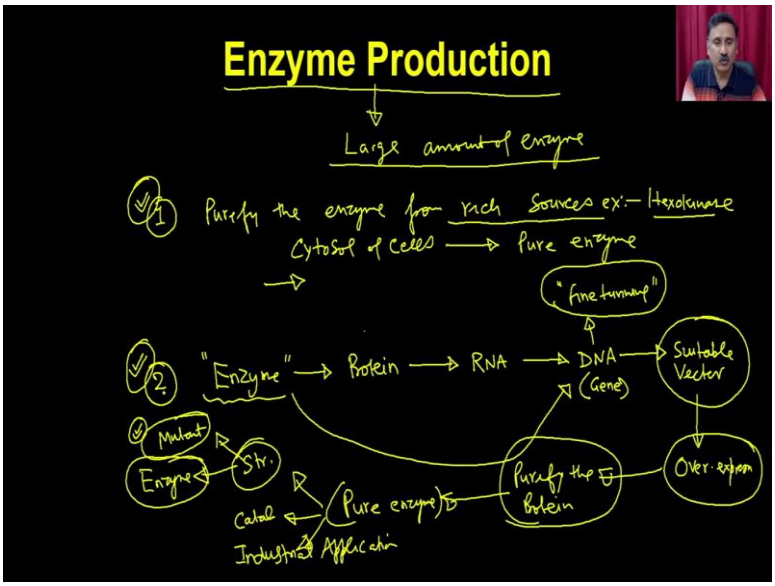
Module - III
Enzyme Production (Part 1: Cloning)
Lecture - 12
Identification of Enzyme Gene (Part-I)

Hello everyone. This is Dr. Vishal Trivedi from Department of Biosciences and Bioengineering, IIT, Guwahati. And what we were discussing in the course Enzyme Science and Technology. So far what we have discussed? We have discussed about the history of the development of the enzymology, then we have discussed about the general properties of the enzyme.

And in the previous module, we have discussed about the protein structures, and we have discussed about the primary structure, secondary structure, tertiary structure and quaternary structures. And we have discussed various methods through which you can be able to determine the primary structure, secondary structure, tertiary structure and as well as the quaternary structures.

Now, since you require the enzyme for all these type of actions or all these type of reactions, it is important to understand and also to discuss how you can be able to prepare the enzyme in your laboratory. So, in today's lecture, we are going to discuss about the Enzyme Production.

(Refer Slide Time: 01:53)



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 where there are two approaches one can actually be able to use, 1st 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 hetero kinase, then what I can do is I can just isolate the cytosol of a cell, right. 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 lysozymes, 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 lysozyme. Lysozyme is very heavily available in many of the biological fluids. So, I have to identify those biological fluid and then only I can be able to use the different chromatography techniques to get the pure enzyme.

The 2nd 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, right. And the proteins are being synthesized from the RNA, which is actually going to be coded from the DNA, ok. And this DNA is nothing which is also called as gene.

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 illustrations or if I want an enzyme for the industrial applications, then I can just first do the enzyme. I should identify the corresponding gene, ok.

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 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 re-combinations 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, ok. And so, that you can be able to do, because 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 why 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 what is the drawback of the approach 1 is that first of all you have to go with and identify the rich source.

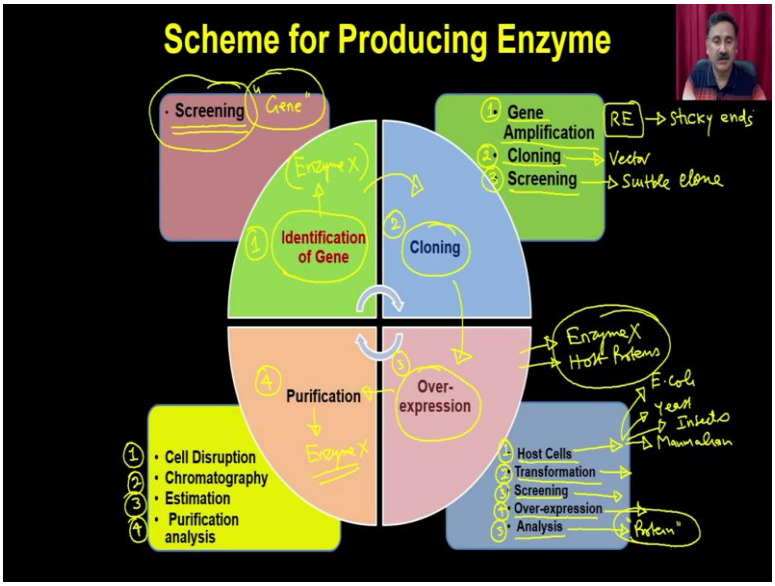
The second is you have to purify this protein with utilization of the different additional chromatography technique and even then it the production is going to be very very limited. Because it depends on the amount of enzyme which is present in this particular rich source.

Whereas, in the approach number 2, once you have identified the gene you can actually be able to over express 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, within larger objective is that I want to study how this enzyme is interacting with the substrate and how it is actually forming the you know 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, right, 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 have also can test the mutants, ok.

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 over produced enzyme. So, this is one of the major advantage of going with the approach number 2

(Refer Slide Time: 08:29)



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, ok. So, first you once you identify the gene, then you are actually going to clone that gene into a suitable vector.

This cloning also has a multiple steps. In the step number 1, you are going to amplify this gene with the help of the site specific primers and these site specific primer are actually going to have the restriction enzyme. So, once you have the restriction enzyme, you can be able to cleave amplify a 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 we facilitate the entry of this particular gene into a suitable vector and that is how you are actually going to produce the clone. So, 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, ok. 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 over expression.

So, in the step number 3, you are going to do the over expression. In the over expression, the step 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 in section lines and all that, ok. 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 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 over expression.

For the over expression 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 over expression, which means you are going to see whether the protein or the enzyme is being produced or not, ok.

And once you are confirmed that the enzyme is getting produced, then you are actually going to make the large cultures, you are going to produced 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, right. So, when you over express, you are going to have two proteins, the enzyme X which is going to be in a 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 over expression. So, then you are going to enter into the step number 4, 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 steps. In the step number 1, you are going to do the cell disruption.

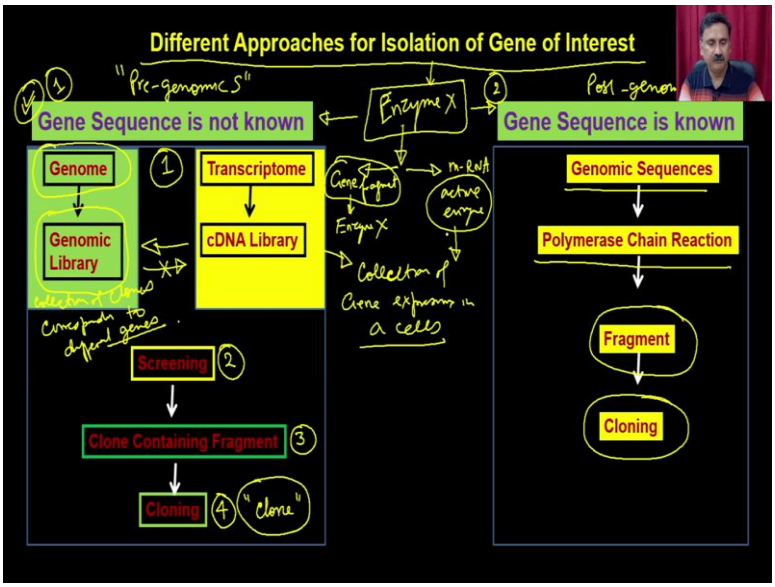
In the step number 2, you are going to do or you are going to perform the different chromatography techniques, such as ion exchange chromatography, hydrophobic interaction chromatography, gel filtration chromatography or sometime you might also do the affinity chromatography.

Once you are done with the chromatography, then you are going to do that 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 4 steps are required, if you want to produce the enzyme under the approach number 2. So, let us start discussing about these

approaches and the 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.

(Refer Slide Time: 13:41)



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, ok. 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 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, ok. And if the gene sequence is known then you are going to use this approach.

So, in the 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, ok.

So, if you have the gene fragment, ok you can be able to go and look at into the genome of the that particular organisms. So, whether you want to use the genome or whether you

want to use the expression profiling, in both of that 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 clones corresponding to different genes. Similarly, so all the genes, ok 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, ok 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 libraries, cDNA library is actually a collection of the X gene expressing in a cell, ok.

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, ok.

This means 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 cDNA library, ok. Once you got the library, right just like as you have the library in your college or you know in your institution, right, and today I want to study a book which is about the protein biochemistry or protein structure determinations, right. Then, what you are going to do? You are going to take these keywords and you are going to put it into the software, right. 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, ok these are the books which are available which as 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 expression status of that (Refer Time: 17:38)

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 2, 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.

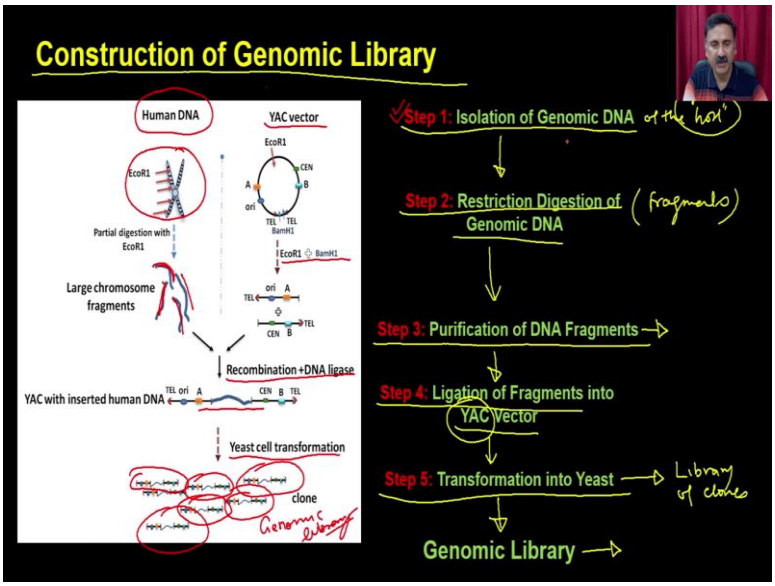
In the step 3, you are going to take out or you are going to isolate these gene fragments and then you are actually going to clone it into a suitable vector, ok. And that is how you are actually going to get the final clone. So, this is all about the approach when you do not have when you do not know the genomic sequence.

Which means the pre-genomic era, before the pre-genomic era people were actually going with the 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 easy 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 you can be used further into the over expression and purification steps. So, first, we will discuss about the approach number 1, where the genomic sequence is not known and then we are going to discuss about the approach where the gene sequence is known.

(Refer Slide Time: 19:19)



Now, the first approach is when how to prepare the genomic library, ok. So, preparation of genomic library is a multi-step process, right. In the step 1, 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 general what you are going to do is you are going to make the fragments. Then, the step 3, 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 5, all these are going to transform into the yeast, so that you are going to get a library of clones, ok. And once you got the library of clones that is actually going to give you a genomic library; all that genes which is present in this particular organisms.

So, how you are going to do this? The 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, right because we are talking about the preparation of the genomic library from the human, right.

And then, you are going to digest that human genome. So, see all the chromosomes are going to be digested with the help of the thin enzymes and that is actually going to give

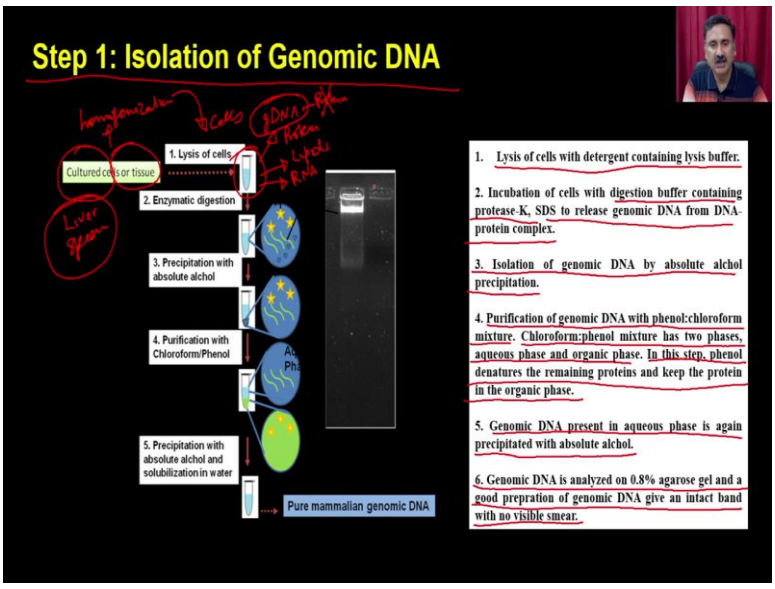
you the different fragments, right all these fragments which actually contains one or more genes in each fragment.

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 percent 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 EcoR-1and 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 do the 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 why you are going to get the clones.

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.

(Refer Slide Time: 22:22)



So, in the step one you are going to do the isolation of the genomic DNA. So, the isolation of the genomic DNA is a multi-step process. In the step 1, you are going to do

the lysis of the cell with the detergent called lysis buffer, ok, with 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 cells.

So, once you do the if it is a 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 tissue. So, that they will actually going to give you the single cell suspension. And then these homogenized project or the single cell is actually going to incubate with the lysis buffer, so that it is actually going to lyse the cells. So, you are going to get the lysate.

And this lysate is actually going to contain three component, one it is actually going to contain the genomic DNA or gDNA, it is actually going to contain the protein, and it will also going to contain the lipid. And so, and it may also contain the RNA which is also present in the cell, right. So, these are the 4 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 protease-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 why your genomic DNA is going to be released.

Then the step 3, 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.

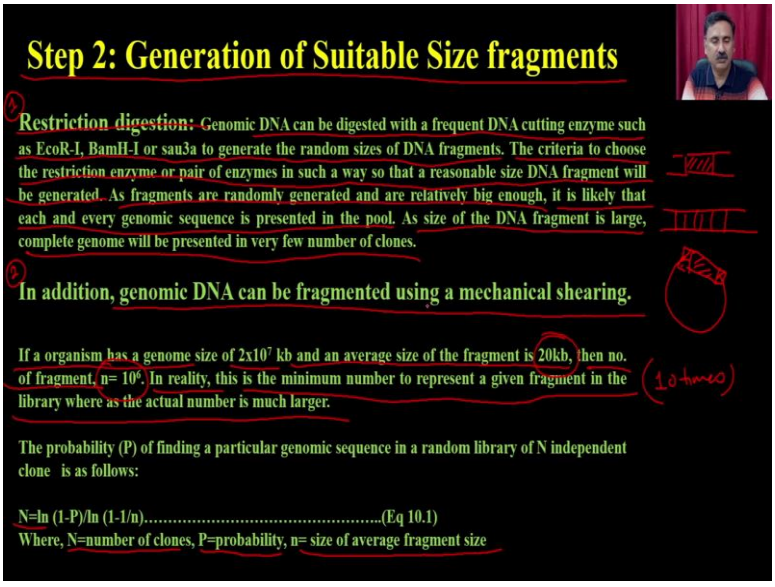
So, in that case what you are going to do is, then you are going to do a purification of the genomic DNA with the phenol chloroform mixture. And chloroform mixture has two phases, one is aqueous phase, then other one is a organic phase. This is what you see, right. It has an aqueous phase and an organic phase.

In this step, the phenol denatured 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 blue coloured 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 percent agarose and a good preparation of genomic give a intact band with a no visible strain. So, this is what you see. This is the genomic DNA, right.

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 the genomic DNA is very big. So, it does not migrate very fast. So, this is all about the genomic, isolation of the genomic DNA.

(Refer Slide Time: 26:00)



And then once you have generated genomic DNA, then the step 2 is the generation of the suitable size fragments. Suitable size fragment can be done by the two method, one is a restriction digestion method where you genomic DNA can be digested with a frequent DNA cutting enzyme. Such as, the EcoR-1 or the BamH1 or sau3a to generate the random sizes of genomic DNA.

The criteria to choose the restriction enzyme or the pair of enzyme in such a way, so 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 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 content, genomic DNA, 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.

So, it may have the fragments of other genes, but at least it should have one gene. It should not have like 4 or 5 or 6 or more genes. So, that is why you have to generate the fragments 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 is present in that particular genome.

So, that is the method number 1, where you can do the restriction digestion. In the step number 2, you can also do a mechanical sharing. So, what you can do is just vertex the genomic DNA and the 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.

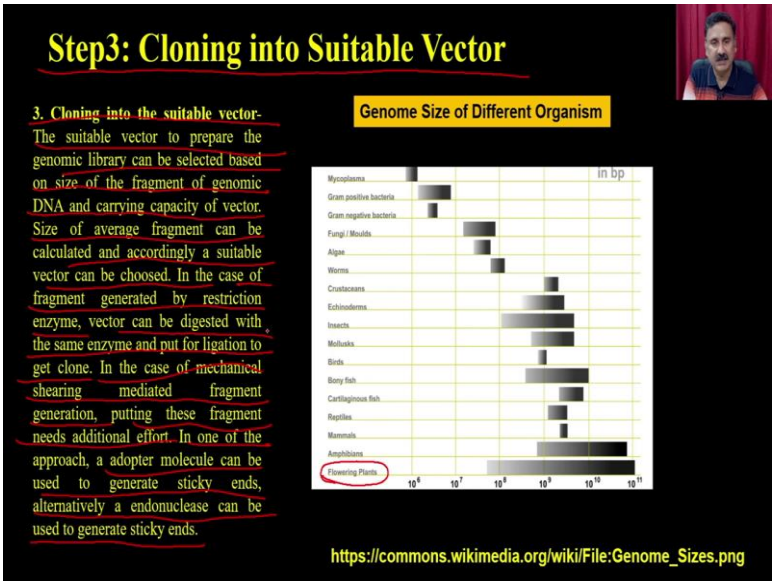
Now, for example, how much would be the fragment size. So, for example, if an organism has a genome size of 2 into 10 to power 7 kb and an average size of fragment is 20 kb, right. Then, number of fragment would be n is equal to 10 to power 6. So, you divide this number by this number, you will get this number.

In reality this is the minimum number of clones to represent a fragment in a 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, ok. So, the probability of finding a particular genomic sequence in a random library of n independent clone is N is equal to ln, 1 minus P divided by ln 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 it 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.

(Refer Slide Time: 29:35)



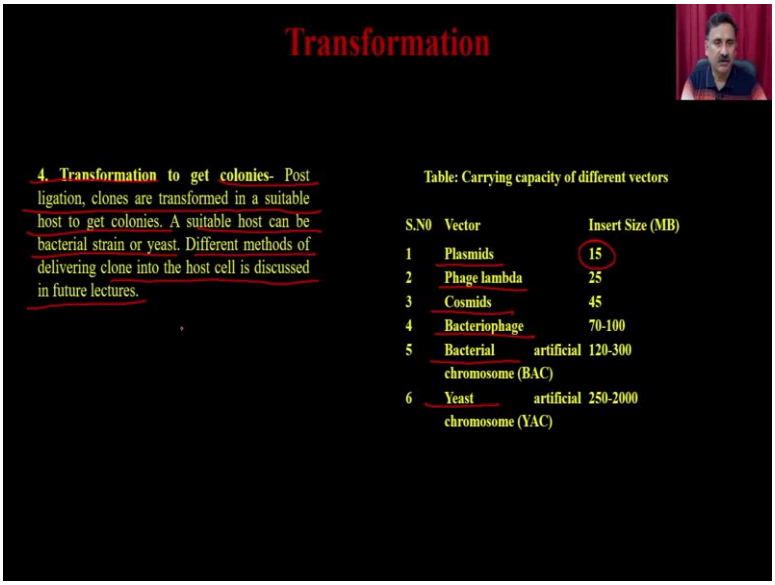
So, in the step 3, you are going to use the cloning into the suitable vector. So, depending on the size of the you know the size of the different organisms, for example, the mycoplasma, bacteria or like the flowering plants, you have to choose the suitable vector.

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, an accordingly a suitable vector can be choosed. 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, a endonuclease can be used to generate the sticky ends.

(Refer Slide Time: 30:29)



So, these are some of the you know 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, BAC and the yeast, ok.

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 in the step 4, you are going to do the transformation to get the colonies. So, the post ligation, the clones are being 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 method 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, ok. 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.