Molecular Biology

Prof. Vishal Trivedi

Department of Biosciences and Bioengineering

Indian Institute of Technology, Guwahati

Module - 10

Molecular Cloning

Lecture-46 Protein Over-expression

Hello everyone, this is Dr. Vishal Trivedi from department of biosciences and bioengineering IIT Guwahati. So, for what we have discussed, we have discussed about the cloning isolating the gene into from the genome either by utilizing the either by utilizing the PCR or with the help of the isolating the clone from the genomic library or the cDNA library. Subsequent to that we have also discussed how you can be able to clone this fragment into a suitable vector and once you got the clone into the vector, you can be able to deliver this DNA into a suitable host and once you got the transformed colonies, you can be able to utilize them for the screening and as well as subsequent to that for the protein production. So, what we have discussed so far is that you are going to get the transformed bacteria and or the transform host that you are going to put it for the screening and once you got the screen, you got the clone containing host cells.

Now this clone can be used for the protein production or the enzyme production. Now when we talk about the enzyme production, enzyme production will depend on the type of host what you are going to use for the production. So as far as the cloning is concerned, the cloning can be done into two different types of vectors. It can be done into the cloning vector or it can be done into the expression vector.

In some cases when you know that the enzyme or the protein what you are expressing is toxic in nature, so in those cases you do not do the cloning related performance or cloning related operations into a expression vector because then you are going to produce a protein and then eventually it is going to kill the cell. So, that is how you are going to use in a cloning vector. Apart from this cloning vector can also be able to use for studying the mechanism of the transcription, replications and the preparation of the genomic and as well as the cDNA library. Whereas the expression vector is exclusively been used for studying the mechanism of the translation or as well as the enzyme production and ultimately it is going to give you the enzyme of your interest. Now when you talk about the enzyme production, the enzyme production is complicated process and it requires the discrete steps.

So before getting into the different steps what you require in the recombinant with the recombinant DNA we should first understand how the protein production occurs in a

particular cell. So, protein production is a multi step process and these steps you have to follow following steps. In the step 1, the binding of the RNA polymerase to the promoter element to start the transcription to form the messenger RNA. So in the step 1 you are actually going to produce the messenger RNA and then this messenger RNA is going to be utilized in the step 2. So as soon as the messenger RNA is synthesized a translational machinery starts the synthesis of the protein or the enzyme.

So protein synthesis starts usually at a start codon which is called as AUG and it is at the stop codon which are called as UAE, UGA or UAG. In bacteria, transcription and translation occurs simultaneously because there is no nucleus right. So, transcription and translation occurs simultaneously whereas in the eukaryotic system, you are going to have the transcription inside the nucleus and then, The translation is going to be in the outside the cytosol. So, That is why in a eukaryotic-system transcription and translation are not going to be together whereas in the bacterial system, you are going to have the transcription and translation together. So, in the step 1 from the gene, you are going to produce the messenger RNA.

So, RNA polymerase is going to sit on to the promoter region and then, it is actually going to synthesize the messenger RNA which is responsible for this. This messenger RNA is going to be modified you know post translational modification and all that. And then ultimately, the first codon which is the AUG is going to be the place where the ribosome is going to sit and that is how it is actually going to start forming the synthesis of the proteins. So, and then it is going to synthesize the polypeptide. This polypeptide is going to be get folded and that is how you are going to get the folded functional proteins.

You can actually be able to read this so that you can be able to understand the subsequent process what we are going to do inside the host so that and it will actually be able to helpful. So, I have given you a reference which you can actually be below to understand all of these processes. So, you actually required to understand the transcription and translation events, then only you can be able to modulate and you can be able to understand how we are actually overexpressing the protein in the host system. Now, when we talk about the under the in vitro system, under the in vitro system you have the two different species. One is you have the host cells where you are going which you are going to use as the protein production machinery and then you also going to have the transforming agents which will actually going to use for providing the instructions.

This means you are going to first use the instructions and then you put it that into the production machinery and that is how you are going to use that production machinery for the protein production. So, as far as the host cell is concerned you have the multiple choices. You can use the prokaryotic system either the E. coli or other bacterial cells, you can use the eukaryotic system such as you can use the yeast, animals and plants and as far as the transforming agent is concerned you can use the different types of plasmids, you can use the mammalian vectors, you can use the yeast vectors and that all we have discussed

when we were discussing about how you can be able to clone a particular gene fragment into a vector of your choice. So, what we are going to discuss is what we are just going to discuss about the prokaryotic and thresholding of the eukaryotic expression system we are going to discuss about yeast expression system.

Then we are going to discuss about the animal expression system. So, these are the things we are going to discuss. E. coli as a expression system which is going to be for the prokaryotic system, yeast as the expression system which is be part of eukaryotic system, then we also going to insect cell line as the expression system and the mammalian expression system. Now, before getting into of these choices, so you for a particular gene you have the following choices, you have four choices or even more than that.

The first question comes how you can be able to select and select the particular expression system because every expression system has its positive and negatives. So, the number of factor need to be considered to choose the host expression system suitable for the over expression of a protein. The first factor is that the quantity of the desired protein. If the quantity protein required in a small quantity any host expression system can be suitable for the purpose. If the large quantity of protein is required such a E.

coli or yeast or baculoe expression system might be more suitable than the mammalian expression system because mammalian expression system is going to give you a very small amount of proteins. Then it also depends on the size of the protein, the E. coli expression system is not preferred for a large protein size of the protein, but an E. coli expression system is more suitable for the large size proteins. Then we have the compatibility, compatibility between the source organism and the expression system.

So, in general a close distance between the source organism and the expression system is preferred at may increase the chances of getting the expression of the clone gene and the presence of protein in the soluble fraction. Then we also require we have to see the downstream application. So, this is the most important criteria to choose a host factor system. If the protein production is for generating the antibody any expression system may suit for this purpose, but if the protein is required for activity or for ELISA then a compatible expression system is preferred. So, downstream application which means where you are going to use this particular protein is very, very important criteria to select any of these hosts.

For example, you cannot use the bacterial expression system in case you are going to use the downstream product for developing the vaccine or utilizing them for using for production generation of the antibodies or something. Because then there is a chance that you might actually be able to get some bacterial products and these bacterial products are many time causes the allergic reactions into the patients. So, if you downstream application is actually going to decide what expression system is going to use. So, in the E. coli expression system, so in a typical component of an E.

coli expression system additional structural features are essential for an expression vector. What we have discussed when we were discussing about the cloning vector we said that it should have the origin of replication, it should have a multiple cloning site and so on. But if you want to talk about the expression vector what you require is you require the promoter. So, for a cloning vector what you require? You require the origin of replications, number 1, number 2 you require the multiple cloning site and number 3 you also require a promoter because the promoter is going to decide the protein production. So, promoter this is the upstream sequence to the gene and provide the docking site for the RNA polymerase.

Then you also require the ribosome binding sites. So, ribosome binding site includes the Shinder-Gano sequences and it is a docking site for the assembly of ribosomes. So, you also require the RBS and RBS is a ribosome binding site and ribosome binding site is very important for the binding of the ribosome so that it is actually going to initiate the translation. Then you also require the termination sites. So, it terminate the synthesis of the messenger RNA and then some cases you also require the affinity tag.

So, affinity tag is not essential. These are the component which are essential. The origin of replication, multiple cloning site, promoter. The presence of affinity tag either before or after the gene sequence provide a mean to purify the protein using the affinity chromatography. So, these are we are going to discuss then only you will understand what is mean by the affinity tag and how it is actually makes the life easy for the researchers to purify the enzyme in a bulk quantities.

Now, as far as the promoter is concerned in a prokaryotic system, the promoter is containing the some of the classical features such as minus 35 regions and minus 10 region. So, TATA box is there. So, you have the TATA box and you also have the minus 35 regions. So, sequence at the minus 10 and minus 35 are crucial to facilitate the RNA polymerase and the subsequent determination of the strength of the promoter. So, as good these sequences are like minus 10 region and minus 35 region, it actually going to decide how efficiently the RNA polymerase will go and sit to these region and that is how they are actually going to give you the better transcription.

The nucleotide substitution in this region is severely affecting the turnover number of RNA polymerase binding and the transcription initiation side. Subsequently, a number of promoters are designed for the over expression of the protein in E. coli using a strong or weak promoter to see it is the over expression strategies. So, we have the IPTG inducible promoter. IPTG is stands for the indolel pyros is a synthetic analog of lactose and it has been widely been used for the to the construct different expression vector to express the protein in E.

coli, the different vector contain the lac promoter or its derivatives. You have the three different types of promoters. You have the lac promoters, examples of the plasmid is Puck

series and PGM. Then you also have the Taq promoters. So, Taq is a hybrid promoter where you have the some region of the tryptophan promoter and the lac promoters.

So, it is a hybrid promoter where minus 10 region is a from lac UV pipe promoter and it is used with the minus 35 region of the tryptophan promoter. Example is PKK223-3. Then you also have the Tric promoter. So, Tric promoter it is similar to the Taq promoter except that the distance separating the minus 10 and minus 35 region of the promoter is different from the Taq promoter.

The example is Ptric99A. So, either of these plasmids can be used for generating the recombinant DNA and then you can be able to transform that into the suitable host and then you can be able to use this for protein production. Then we have the bacteriophage lambda promoters. So, this promoter keeps a tight control over the protein production. It is regulated by the presence of repressor CLTPs857 to either the repressor repress the transcription or not. CLTP5857 is a temperature sensitive and degraded at high temperature and consequently in a temperature dependent fashion it represses the transcription at low temperature, but not at a high temperature.

This promoter is useful in cases where the protein is toxic in nature. So, then we also have the bacteriophage T7 promoters. So, similar to the bacteriophage PL promoter the T7 promoters is used to design the plasmid with tight control on the protein production. These vectors contain most of the structural blocks from the PVRT22 and the MCS in front of the T7 promoter to drive the transcription of the insert. Hence vector containing the foreign gene in front of the T7 promoter for the expression.

So, T7 host E. coli also needs the modification to shoots the T7 promoter and host E. coli is been transformed with the plasmid which carry the T7 RNA polymerase gene or the T7 RNA polymerase gene is integrated into the bacterial chromosome. In few host strain T7 RNA polymerase is placed under the tight control of IPTG inducible LAC UV5 promoter to tightly control the production of the T7 polymerase. So, either of these promoters first step is that you are going to transform the recombinant DNA into the host and then you are going to do the protein production. How you are going to do the protein production? In the step 1 you are going to do the transformations.

So, what you are going to do is you are going to take the recombinant plasmid and you are going to do the transformation into the suitable bacterial species or bacterial strain and that is how you are going to get the transformed bacteria. And you can use the multiple method of transformation, you can use a calcium chloride method or you can use the electroporation. Then the step 2 you are going to inoculate the single colony into suitable bacterial media such as LB media and you can allow them to grow into a 37 incubator. So, a single colony of the transformed colony is inoculated into a suitable media and it can grow up to a log phase such as the OD is 0.

6 to 0.7. And then what you are going to do is you are going to induce, so you are going to in the third step you are going to induce the bacterial species with the inducer such as IPTG for 3 to 6 hours to produce the proteins. So, you can do like that and then you are going to do a centrifugation or the collection of the bacterial cells. So, in the step 4 you are going to recover the bacterial bacteria and analyze the protein expression. So, bacteria can be recovered from the culture with a brief centrifugation at 8000 to 9000 RPM and analyze on to the SDS page. The detail of the SDS page will be discussed in a future experiment.

The SDS page analysis of a particular expression study in given and it indicates a prominent expression of the target protein in the induced cell as compared to the uninduced cell. So, once the induction is over you can actually be able to do the centrifugation and that will actually going to give you the bacterial pellet. This bacterial pellet can be analyzed for the protein production. So, what you can see here is this is uninduced cells and these are the IPTG induced cells and what you see here is a very prominent band of the protein of your interest. Now there are many factors which are actually going to decide the protein production into the E.

coli expression system. So, factor affecting the protein synthesis in E. coli. So, first factor is the translational efficiency. So, translational efficiency is governed by the composition of the promoter especially the sequence of the Shine-Dargano sequences which enables the binding of the ribosome protein production machinery. In addition, the distance between the Shine-Dargano sequence and the start codon is also important for the efficient translation.

Moreover, secondary structure of the promoter elements also affect the efficiency of the gene expression. Then the step 2 is the growth conditions. Growth media has a drastic effect on the protein production. Either the media component provide the raw material for the synthesis of the amino acid or provide the amino acid for the synthesis of a protein. In addition, the growth media rich with carbon source may provide high cell mass and as a result it will give you the more amount of proteins.

Then the third is the codon usage. So, third is codon usage. Genetic codes are degenerate and there are 61 codes which are available for the 20 amino acids. So, this is the genetic code what you see. And there are 61 codes which codes for the proteins whereas 3 codons which are for the stop codon. So, these are the stop codons what you see and which does not code for any amino acid.

Except these you are going to have the codon which is coding for one or other amino acids. As a result the organism has a preference towards a set of genetic code. Expressing these sequence requires the tRNA to recognize the genetic code. But if the host expression system has no tRNA or low level of particular tRNA then it will either delay the synthesis or stop the synthesis of a particular amino acid. Consequently, either it will produce less protein or the truncated protein.

So, what it mean is that every organism has a preference over using some of the codons. For example, in the case of the phenylalanine you have the two codons UUU, UUC right. It is possible that in E. coli probably the UUU is more preferred codon. So, in that case it will actually going to have the tRNAs only for the UUU.

But if you are taking a protein which is you know which does not have this codon which you have the other codon then either it will take the time for synthesis of these tRNA molecules or it will actually going to truncate the protein synthesis at that stage. So, that is a very very important criteria to select the host as per the codon what is present in your gene. Then we have the expression of the fusion protein in E.

coli. The protein in the E. coli expression system can be expressed as a hybrid protein where the reading frame of the two gene one for the fusing tag and the other one is for the foreign gene RNA frame. So, fusion tag can be placed either at the N-terminus or the C-terminus. So, these are the some of the fusion tag beta galactosidase, MBP, thirodoxin, poly histidine, GST and alkaline phosphatase and you can actually be able to use these are the vectors what you can use for providing the for tag. And what is the advantage for example, if you have beta galactosidase you can use the you can use that for blue white screening and as well as for affinity purification. Mostly these fusion tags are being used for affinity purification so that you can be you know avoid the contaminating protein because the fusion tag will not be present in other protein but only present in your protein.

So, if you pass through this to affinity column only this protein is actually going to bind and the rest protein will not bind and that is how you will get the purification in single band or single step. The advantage of the fusion proteins it is going to make the easy purification you might have seen that the with the help of the affinity purification it can be a single step purification. Sometime the tags are being put so that you can be able to target a protein into a particular compartment. So, fusion protein can be targeted to the different cellular compartment for various regions such as the periplasm targeting sequences will allow the protein to accumulate into the periplasm and hence can help to the easy isolation. It can also modulate the half life of the protein in many cases a fusion tag hides the potential protease site which are present on the foreign protein and enhances its half life.

Then it also increases the solubility keeping the tag at n-terminal direct the protein synthesis and helps in increasing the solubility of the foreign protein. Then we have the how we can be able to remove the fusion tag. So for many biotechnology application a protein is expressed as a fusion tag with n-terminal or c-terminal tag to easily purify the protein. But after the purification the tag need to be removed for the downstream applications such as vaccine or the protein crystallographic studies.

A list of reagent is given. So, these are the reagent what you can use you can use the cyanogen bromide you can use hydroxyl amine, enterokinase, factor 10a, alpha-thrombin,

trypsin and subtlin and they are all mostly the proteases except the cyanogen bromide which is a chemical and that is going to cleave just after the methionine. So if you have a tag and if you have a methionine here, so this is the tag actually and this is your gene of interest. So it is having a methionine in between. So what will happen is if you treat that with the cyanogen bromide it is actually going to cleave and that is how you are going to get two fragments your tag and the gene of your interest. So in many cases it is very very essential especially the places where you are going to use this protein as a drug for example like for example if you are going to use that like insulin for example.

So if you are going to put a affinity tag on the insulin it may actually cause the allergic reactions to the patient that is why this affinity tag has to be removed. How you are going to remove the fusion tag? So what you are going to do is in general the fusion tag junction point has either the protease cutting site or the site is sensitive for the chemical treatment. Treating the fusion protein with the protease or the chemical agent cut the fusion tag to release the target protein. So passing the cleavage mixture allows the binding of the tag into the affinity column whereas the target protein does not bind and comes out in the flow through. Target protein free of fusion tag can be collected and used for the downstream application.

So this is what you have in this particular we are taking an example of the his tag. So you have made a chimeric protein so that is having the his tag on one side and the protein of your interest on the other side and then what you can do is you can have you can actually be able to treat this with the thrombin because it has a thrombin cleavage site in between. So what thrombin is going to do is it is going to remove the his tag and the protein and now what you do is you load this onto a affinity column and as a result what will happen is that his tag will go and bind to the beads whereas the protein of your protein is going to come out into the flow through and that is how you can be able to separate the tag from the protein of your interest. So this is all about the bacteria as the host system for protein production and what we have discussed we have discussed about the different steps what you have to follow which means in the step one you are going to transform the bacterial species, bacteria with your recombinant DNA and we have discussed many methods what you can use for transformations. Number two you are going to do the screening or the selection of the transformed clones.

Number three you are going to grow a single colony you are going to inoculate into the media and that is how you are going to induce and the step four you are going to induce that bacterial culture with the help of the inducer. So in this particular example we have taken an inducer as the IPTG and then once the induction is over for 3 to 4 hours then you are going to collect these cells by centrifugations and then you are going to utilize these cells for analyzing the protein production in the SDS page and once you are sure that the protein is being produced then you can be able to lyse the cells and you can purify the protein for downstream applications. So this is all about the different aspects of the protein

production in E. coli as an expression system. So with this I would like to conclude my lecture here. Thank you.