Eukaryotic Gene Expression

Prof. P. N. Rangarajan
Department of Biochemistry
Indian Institute of Science, Bangalore

Lecture No. # 31

Eukaryotic Protein Expression Systems - II

Welcome to this lecture on Eukaryotic Protein Expression Systems - II. In the previous lecture, we had discussed about the expression of proteins in yeast and insect cells, and before that we have discussed about introducing genes into E coli cells, and how we can make proteins in E coli cells. And when your protein actually requires specific post translation modifications such as glycosylation, phosphorylation or specific folding involving disulfide bonds, and in such cases, many a times proteins of eukaryotic carrageen, when they are expressed in E coli cells you will not happen, and therefore, the protein may not have proper biological activity.

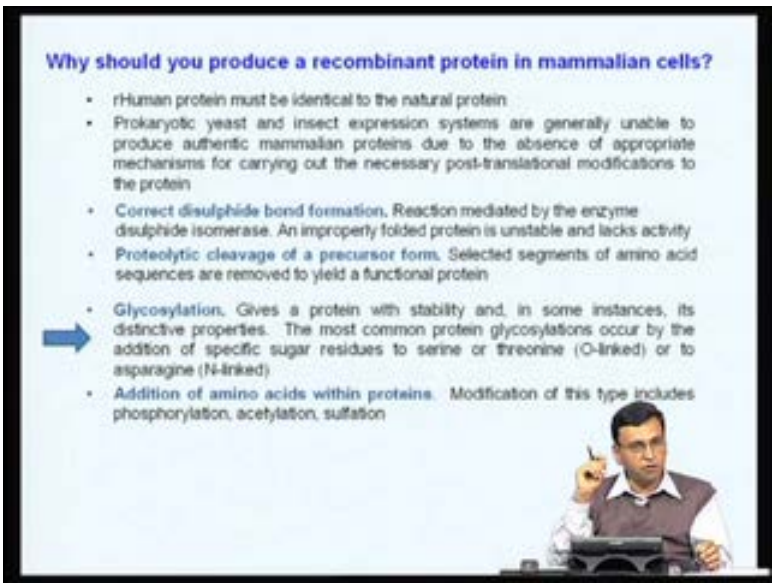
So, in such cases, you have no other options, but to take your gene and express then in eukaryotic expression systems under the control of a eukaryotic promoter in a eukaryotic cell line. And the first choice to express a protein of eukaryotic origin and a eukaryotic expression system is again microbial systems and yeast is an ideal choice to express some of these mammalian proteins in each cells. In fact, there are number of mammalian proteins have been expressed successfully in yeast cells and are now available as either diagnostic as therapeutics in the market.

But many a times, when you express the proteins in either insect or yeast cells the protein may not be biological active, because the folding or the glycosylation pattern or other post translation modifications is not exactly same as it happens in a mammalian cell types.

So, in such a scenario, we have no other choice, but to go and then clone your gene in a mammalian expression vector and express this in a mammalian expression system. So, let us now discuss today or as to see what kind of mammalian expression systems are available and how does one go about expressing a gene in a mammalian expression system, which has a number of advantages as well as disadvantages, over expressing genes in cells and cell lines. And in the next class, we are going to talk about the

expression of proteins in plants and farm animals and then will talk about human gene therapy where instead of taking your gene and expressing in cells outside you now directly going a express your gene in the body itself and see, whether you can correct some of the genetic disorders by using this kind of a human gene therapy approaches.

(Refer Slide Time: 02:51)



So, let us today focus on the expression of genes in mammalian cells, how it can make r eukaryotic proteins using mammalian cells as factories. As I said, let us now spend time to understand this logic why should you express protein or why should we produce a recombinant protein in mammalian cells.

As I said, when you want your recombinant human protein or a recombinant mammalian protein to be exactly the same as this occurs in the native form. If the human, if you want express a human protein in recombinant form and if you want this recombinant human protein to be exactly the same as it is present in the native human cells, then you have no other option, but to go for a mammalian expression system.

The reason is many a times the prokaryotic as well as yeast and insect expression systems may not be able to produce mammalian proteins in the same way as it happens in the mammalian cells. And when they post translate the modifications or the folding patterns in these recombinant systems are not exactly the same as that happens in the mammalian cells; then the mammalian protein that is expressed in either yeast insect or prokaryotic system will not be the same as that happens in the vivo.

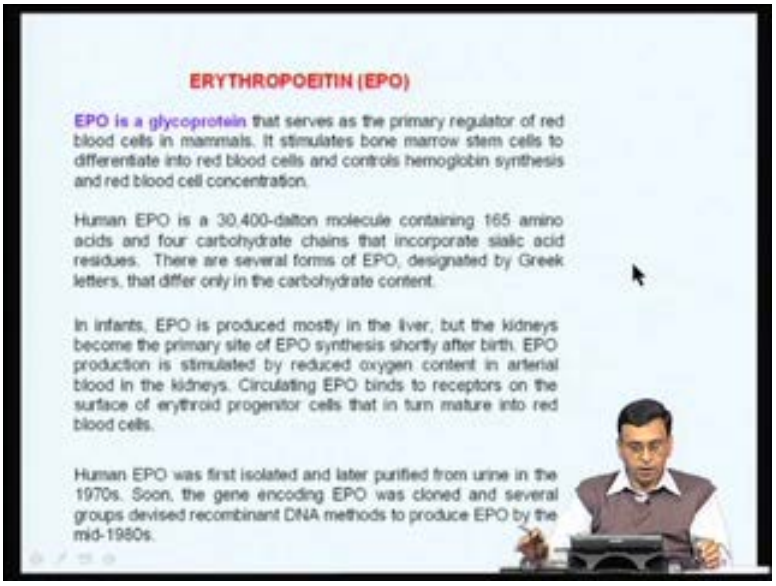
And therefore, you have to choose a mammalian expression system for your recombinant protein production. One of the most important things that often happens is the proper disulphide bond formation as you know, many proteins contain the number of disulphide bonds present between the two cystine residues and this disulphide bond formation is catalyzed by the enzyme disulphide isomers. And many times, the proper formation of the disulphide bonds is very very important for the correct folding of the molecule.

And when these disulphide bonds are not properly formed, then the protein that is produced in this recombinant form will not be biological active and therefore, it will be useless, so proper disulphide bond formation often happens only when you express these human proteins in a mammalian expression system rather than expressing in yeast or insect systems, so you have no other option better to go for a mammalian expression system for making your recombinant protein.

The other thing that normally one encounters when you want to make a recombinant proteins is photolytic cleavage of precursor forms; many times in human and other mammalian cells, the proteins are often proteolytically processed to get a final matured form, and I mean in this kind of proteolytic processing is different in either in insect or yeast cells then you cannot make this protein in those systems, so we have to come back to the mammalian expression systems for making your protein.

And the most important thing that often happens for you to choose a mammalian expression system rather than a yeast or insect expression system is the glycosylation pattern. Now, although the yeast and insect cells to glycosylation proteins, many time the exact glycosylation that happens in mammalian cells may not be reproduced in the yeast or the insect cells. And if this glycosylation differences are going to affect the biological activity of protein, then you have no other choice, but you express your protein in a mammalian expression system.

(Refer Slide Time: 06:34)



And as you know glycosylation primary takes place on serine and heroine residues, when it is actually called as the o linked glycosylation, and when this glycosylation takes place in the asparagines residues it is called as n linked glycosylation. This o linked and n linked glycosylation patterns are often different in different expression systems and I am going to actually give you one example as to how glycosylation plays a very important role in defect method in the biological activity of a protein.

So, in addition there could also be differences in other modifications such as phosphorylation, acetylation, and sulphation and so on and so forth, and when you want all these changes that take place exactly as it happens in mammalian cells; then you have to choose for a mammalian expression system to make your recombinant protein.

Now, I am going to give you one example just to tell you, how glycosylation affects the biological activity of a protein. I am going to give you one example, a protein called erythropoietin, which is a very very important protein, it is a glycoprotein that serves as the primary regulator of red blood production in mammalian cells.

So, whenever the red blood cell counts in our blood goes down, it is sensed by the kidney and the kidney now starts making a hormone called erythropoietin and this erythropoietin now goes and stimulates red cell production in the bone marrow. So, the circulating levels of erythropoietin play a very very important role in determining the hematocrit or the erythrocyte counts in your blood. So, the erythropoietin stimulates

bone marrow stem cells to differentiate into red blood cells and controls hemoglobin synthesis and red blood cell concentration in the blood.

Now, what is erythropoietin, it is about a 30 to 31 kilo dalton protein, because it is about 30,400-dalton molecule contains about 165 amino acids and four carbohydrate chains that incorporates silica acid residues. There are several forms of erythropoietin designated by Greek letters and these different forms of erythropoietin differ only in their glycosylation pattern or the carbohydrate content.

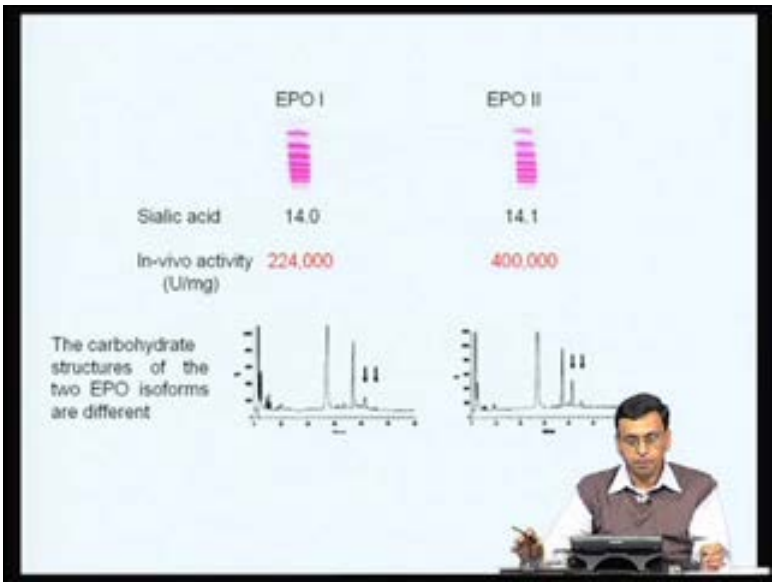
In the infants erythropoietin is produced mostly in liver, but the kidneys become the primary site of erythropoietin synthesis in the neonates and in the adult life, erythropoietin production is stimulated by reduced oxygen content in the arterial blood in the kidneys. So, when you do not have sufficient erythrocytes, then your oxygenation in the blood becomes less and the reduced oxygen content in the arterial blood is sensed by the kidney, and now the kidney start the erythropoietin gene in the kidney start getting expressed and you get erythropoietin. Now, erythropoietin which is synthesized by the kidneys, now is released in the blood strain it now goes and binds to receptors present on the elytroid progenitor cells in the bone marrow stem cells and stimulates the converts them into red blood cells.

The human erythropoietin was first isolated and later purified from the urine in the 1970’s, but once the recombinant DNA technology became available the gene for erythropoietin was cloned and several groups especially a company called Amgen in the United States actually devised the method to actually produce the recombinant erythropoietin by using various expression systems.

(Refer Slide Time: 08:48)

So, the erythropoietin today is available in a in a wide different generic names, a number of companies actually produce the erythropoietin under various trade names; and it became a very very important recombinant protein, that as tremendous use for treating anemia as well as cancer.

(Refer Slide Time: 09:06)



Now, people actually found out when you produce when the erythropoietin was produced by two different expression systems, they actually found out their glycosylation pattern; these multiple bands actually show the different glycosylations of

the erythropoietin that is actually produced in the mammalian expression systems. And the silica acid content of these 2 erythropoietin, which are produced in two different cell lines is more or less the same, but however when they look at the biological active of these two erythropoietin one was at least two fold more active than the other.

To find out what could be the cause because the amino acids sequence is the same the, folding is the same, the glycosylation pattern seems to be the same, but still the biological activity was very different. So, when they actually look at the carbohydrate structure which is present in two forms they actually found out that the form, which as lower activity and the form, which as higher activity, the glycosylation patterns are very different.

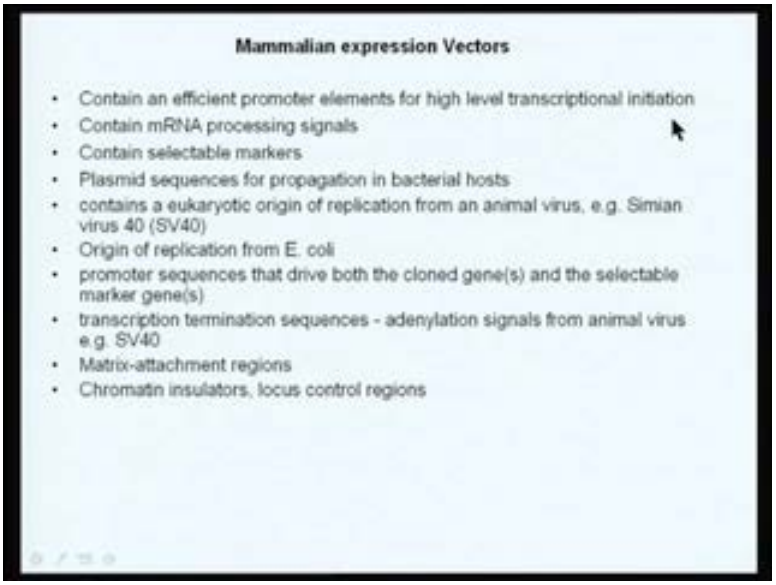
And you can see these peaks which dipped the branched carbohydrate chains are quite different in the two forms. So, the kind of carbohydrates structures in the erythropoietin can play very very important role in term of the biological activity, and if these kinds of differences it is very very important, you actually produce the erythropoietin in the same expression system as the happen in the native and close as as comes to is a mammalian expression system. So, producing this erythropoietin will in yeast or insect expression system will have a different glycosylation pattern and therefore, their their biological activity will be pretty different. So, let us now spend some time to understand how you can produce proteins like erythropoietin in a mammalian expression system, as this true for any other expression system the first step that is required if you want to express a gene in a expression system is you have to develop a expression vector.

So, let us now try to understand, how do you design and develop a mammalian expression vector. Are in this lecture today, we are going to focus primarily about non viral vectors you can introduce, you can make expression vectors for expressing genes in mammalian cells using either non viral vectors or viral vectors. Now, as you know a number of viruses infect us all the time and many of times you can actually take these viruses, manipulate these genomes in such a way, that you can actually introduce gene of interest from the viral genome; and then take this viral genome and make a recombinant virus. And these recombinant viruses can be made to infect mammalian cells and your gene can be expressed in large amounts, and a protein of your interest can be made in huge amounts.

We are going to discuss about this kind of a viral vectors or recombinant viruses for expressing recombinant proteins, at a later stage in this lecture series especially when you are going to talk about gene therapy. Today, we are going to talk about, how you can express genes in mammalian cell lines using non viral mammalian expression vectors.

Now, as is to for the yeast expression vectors or e coli expression vector or insect expression vector in the case of the mammalian expression vector also they need to have a minimal characteristic features. For example, they should contain a efficient promoter elements for high level transcriptional initiation, they should contain messenger RNA processing signals, they should contain selectable markers, because we have distinguish between cells which are taken up the vector and cells, which are not taken up the vector they should have plasmids sequences for propagation of this vectors in bacterial hosts, because first we have to clone your gene in a mammalian vector, put them in e coli cells, make this vector in large amounts then only you can go back and put them in a mammalian vector for expression.

(Refer Slide Time: 12:12)



So, you need to have plasmid sequences for propagation in the bacterial hosts, the mammalian expression vector also should contain a eukaryotic origin of expression. Eukaryotic origin of replication from usually from animal virus such as S V 40 virus, so that you can also replicate in mammalian cells the origin of replication from a, e coli is a must for all the mammalian expression vectors because the first once you cloned your

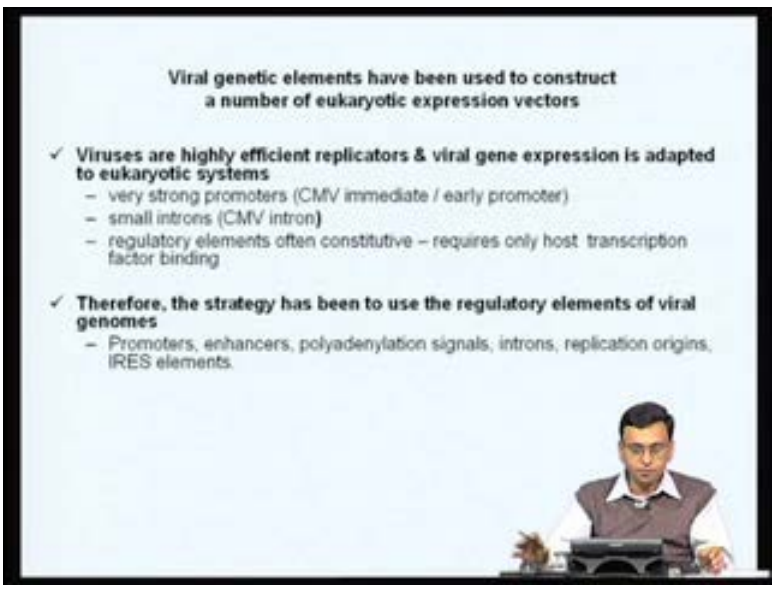
gene into the vector, you have to make this plasmid in large amounts in bacterial cells. So, these plasmids have to multiply in e coli cells therefore, you should have an origin of replication that functions from e coli. You should also have promoter sequences that drive not only your cloned gene, but also the selectable marker genes.

So, we will discuss about what are selectable markers are actually used for expression of genes in mammalian cells. Off course, you should also have multiple cloning sites for cloning your genes as well as transcription termination sequences as well as polyadenylation signals from viral or mammalian origin.

So that, from the promoter the gene is expressed and the transcription properly terminated and then your messenger RNA is polyadenylated; and these polyadenylated messengers can then be translated by the eukaryotic transcription missionary or translation missionary.

In addition many of the modern mammalian expression vectors contain, what are called as matrix attachment regions as well as chromatin insulators and locus control regions. We will discuss these new features of mammalian expression vectors at the late Stage in the lecture series.

(Refer Slide Time: 14:04)



Now, viral genetic elements have been extensively used to construct a number of eukaryotic expression vectors. Now, viruses are highly efficient replicators and viral

gene expression is adapted to eukaryotic systems as I told the animals as well as humans are often bombarded by a number of viruses, we are infected by number of viruses they have natural tropism to infect select cell types in our body. So, these viruses when they enter the body, they actually use the host transcription missionary to to replicate there, to transcribe their RNA and therefore, they contain very powerful promoters to which the host transcription factors go and bind and activate the transcription.

So, if they can choose some of these promoters and these viral genomes which control the expression of the viral genes and you can take these promoters and design eukaryotic expression vectors, so that you can express your gene of interest in very high amounts.

So, some of the very strong viral promoters which are very efficient, very routinely used in expressing genes in eukaryotic cells are the cytomegalovirus immediately early promoter, this is one of the most widely used mammalian expression vectors contains the cytomegalovirus promoter, and usually these mammalian expression vectors also contains a intron that is derived from the cytomegalovirus promoter, cytomegalovirus genome.

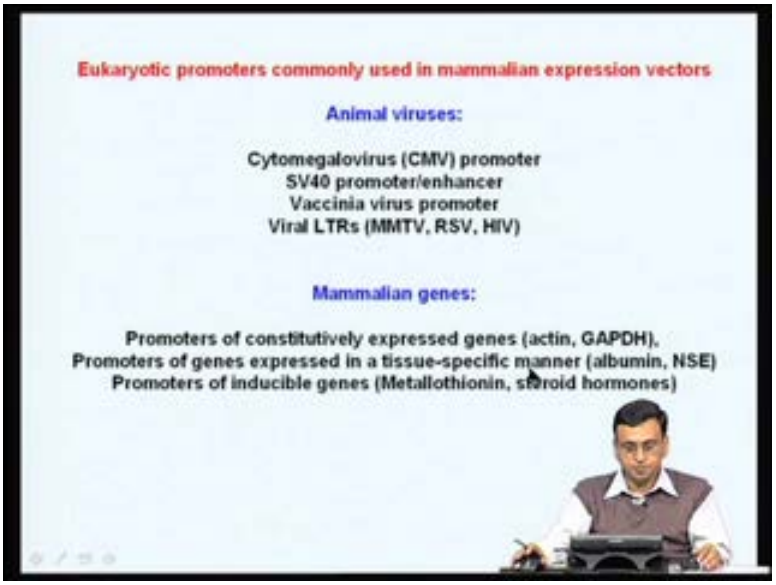
If you remember some of the early lectures that I gave in this lecture series, the splicing when the messenger RNA synthesis in eukaryotic cells as the transcription proceeds some of the post transcriptional modifications such as, splicing, polyadelation they all occurs simultaneously and all these process are coupled, and there have been many instances where, when you have an intone these RNA's are spliced efficiently and they are also polyadenylated efficiently, and if you do not the RNA'S, which are not spliced are not efficiently polyadenylated and therefore, if you want a efficiently transcribed gene. And if you want a good amounts of polyadenylated are to be made it is better to include an intron in the vector, so that they RNA that is synthesized from these vectors are spliced and therefore, they are also efficiently polyadenylated, because splicing and polyadenylated is linked.

So, most of the eukaryotic expression vectors also contain an intron, so that immediately after transcription the RNA is spliced and such RNA'S are likely to be Polyadenylated more efficiently. And the promoter region should contain regulatory elements, and if you want to a gene to be constitutive expressed all the time you include constitutive

promoters and if you want to be regulated, then you have to include regulatable or a inducible promoter depending upon the choice.

So, if you want to express your gene constitutively mostly use viral promoters and not only we use viral promoters you also use enhancers, polyadenylated signals, intones, replication origins and what are called as internal ribosome, intercedes or I RES elements.

(Refer Slide Time: 17:09)



Most of these actually incorporate from the animal viruses or human viruses, because these are the been very well characterized. So, the eukaryotic promoters which are commonly used in mammalian expression vectors are as I said the cytomegalovirus promoters is one of the most extensively used, and very highly expressed very powerful promoters that are routinely used in mammalian expression vectors; you also use the seeing virus 40 promoter and enhancer or sometimes you can use vaccinia virus promoters you can also use, what are called as viral long terminal repeats or LTRS, which contain very powerful promoters for example, the mouse mummeries tumor virus contains a long terminal repeat, which actually contains glucocorticoids response elements and therefore, when you use mammalian M M T V L T R S for in your expression vectors such vectors they expression of your transgenic can be regulated by glucocorticoids.

So, you can turn on the expression by adding glucocorticoids culture medium and you can turn on the gene by removing the glucocorticoids in the medium, but if you want to constitutive expression you can use glucocorticoids virus long terminal repeat or the H I V, L T R.

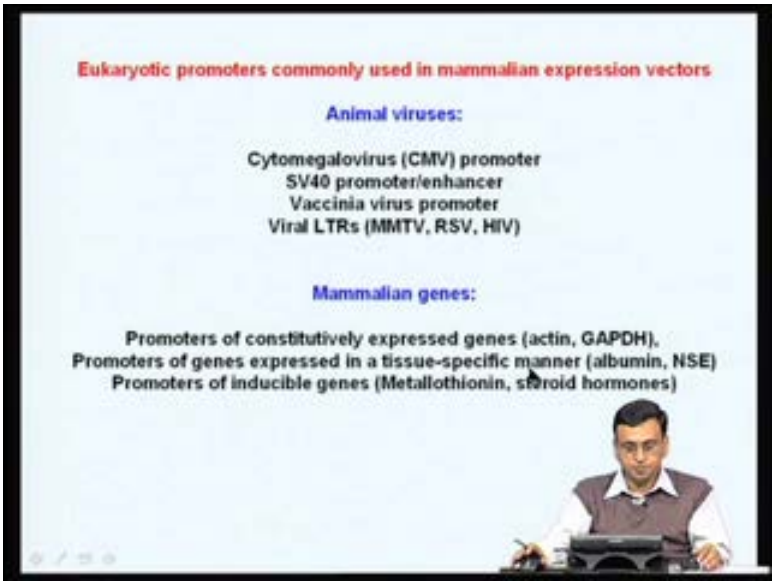
So, depending upon whether you want to express the gene constitutively or in an inducible manner you can use different viral regulatory elements; in addition to these promoters and enhancer sequences from the animal viruses, you can also use the promoters and enhancer sequences from a number of mammalian genes, which have been very well characterized for example, you can use actin or glyceraldehydes 3-phosphate dehydrogenase promoters; as you know acting is expressed in all most all cell types and is one of the most abundant proteins in the cell, so it is a very powerful promoter. Similarly, gaped again because of the house keeping functions is abundantly expressed in a number of cell types, so if you want to express your gene constitutively in a number of cell types you can use either acting or gaped as your promoter of your choice.

You can also use promoters of genes, which are expressed in tissue-specific manner for example, if it is an albumin promoter you can express your trans gene only in liver cells or a liver cell cell lines. You can use for example, promoters like neuron specific enlace, which is a very powerful promoter and when you use this promoter you can express your genes only in neuronal cells or neuronal cell lines.

Similarly, you can use mammalian promoters, which are inducible for example, metallothionein, you can use metallothionein promoter in such cases, you can induce the expression of your trans gene by adding metals into the medium into the culture medium or you can also clone your genes subschema of promoter containing steroid hormone response elements, one example I gave you is the mouse memory tumor virus which contains the glucocorticoids response elements and in such cases, the expression of your gene can be induced by adding glucocorticoids.

So, a wide variety of promoters are available either of animal virus origin or of various mammalian origins and you can use any one of these promoters and design a expression vector of your choice.

(Refer Slide Time: 20:07)



So, a typical mammalian expression vector contains a strong promoter and as I said this strong promoter can be either of a animal viral origin or it can be of a mammalian promoter and usually you include an intone immediately downstream of the promoter sequence, and then you contain a multiple cloning site into which you introduce your trans gene and a trans gene should contain a ATG cordon, so that it is with actually kozak consensus sequence.

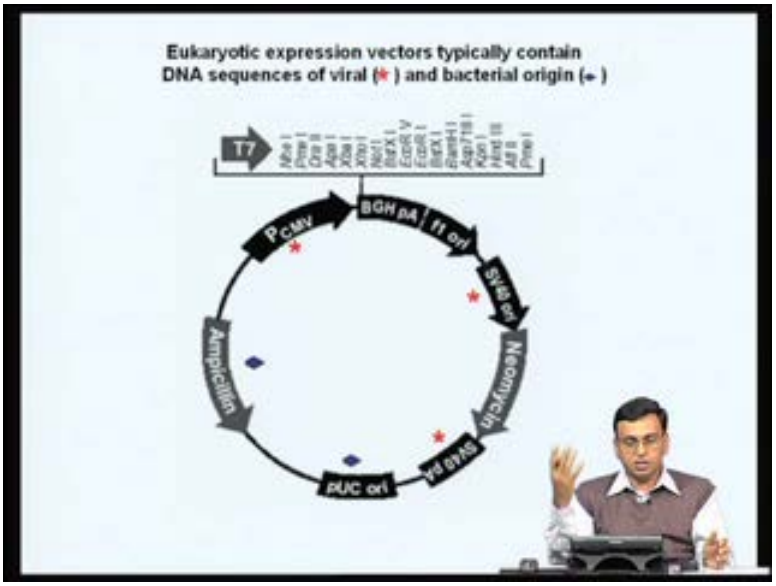
So, that it is properly translated efficiently should also have a proper stop cordon at the end and it should also have three prime untranslated sequences, so that such as the AAUAA or the polyadenylated sequences, so that the RNA is transcription stops and RNA is properly polyrelated. And many a times when you want to purify your protein that is made from your construct a very simple using very simple methods you can actually include certain tags for example, you can use a purification tags such as it includes six hesterines in the amino terminal region of the all the carboxidal region of your gene then you can purify the such recombinant proteins usually using a nickel of any tag, because hesterine binds to nickel and using the nickel column you can easily purify a recombinant proteins by virtue containing a histerine type. And many a times, you can also add tags like myc tag or what is called as flag epitopes these are all very very short amino acid sequences, which are usually incorporate in front of a recombinant proteins and in such cases, detection of your protein becomes very very easy, because antibodies against these epitopes, flag epitopes or myc epitopes are commercially

available and therefore, you do not have to actually have an antibody against your recombinant protein by by incorporating this epitopes into the amino terminals or carboy terminal region of your gene, you can express these proteins containing myc epitopes or flag epitopes and expression of your.

Recombinant protein can be easily detected using antibodies raised against these particular epitopes. So, by using histamine tags you can purify your recombinant protein very easily by nickel affinity column or by introducing epitopes like myc epitopes and flag epitopes you can easily identify the expression of your proteins using antibodies available against these epitopes.

So, there are also many modern expression vectors, which also contain as I said certain barrier elements and insulated and so on forth.

(Refer Slide Time: 22:38)



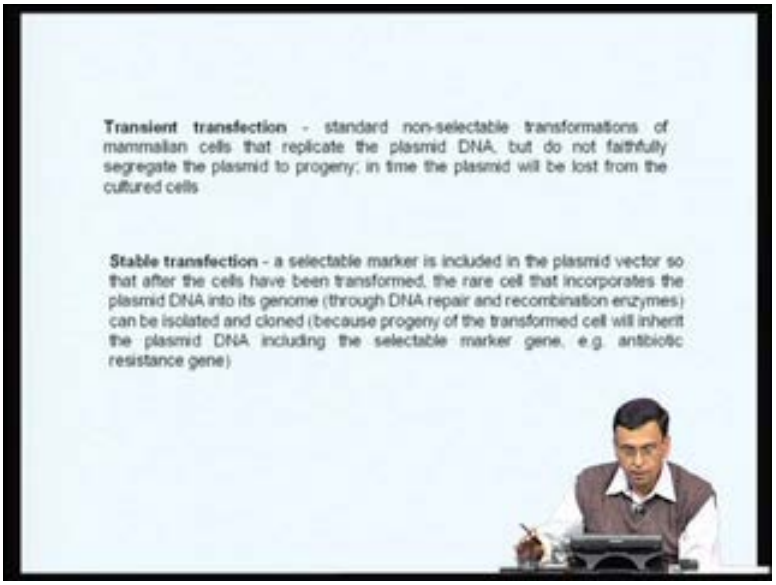
We will discuss some of these very unique developments a little later in the lecture series, this is a typical eukaryotic expression vector I have shown here, which contains a both viral as well as bacterial sequences; the viral sequence of viral origin are shown by asterisk in red color for example, you can you can cytomegalovirus promoter it contains a SV 40 origin of replication. So, that if now I now, introduce this vector in mammalian cell lines which have been transformed with S V 40 large t antigen then every time the cell divides your vector also divides. So, when you transfer to cells, and cells start dividing the daughter cells also will contain your plasmid, your vector; it also contains

the SV 40 poly relation signals. So, that your RNA gets polyadenylated the the blue diamond's actually denotes the sequences derived from the bacterial origin the ampicillin resistance gene.

So, that you clone your gene first in the multiple cloning site and then make this pro vector in large amounts or select select this bacterial cells containing the harboring with gene by ampicillin resistance marker, it also contains a PUC origin of replication, so that this plasmid can be efficiently replicated in the e coli cells.

So, a typically mammalian expression vectors contains a number of sequences from animal viruses such as, cytomegalovirus promoter SV 40 polyadenylated signal and so on and so forth. And it also contains, several bacterial sequences for them to be introduced to bacteria, and select the bacterial cells as well as for replication of bacterial cells.

(Refer Slide Time: 24:06)



Usually, when you want to introduce your expression vector, mammalian expression vector into eukaryotic cells you can do two things, you can express your trans gene or you can express your expression vector, you can induce the expression of the gene in the expression vector in a transient mode or you can permanently integrate your gene of your interest into the genome of the mammalian cell.

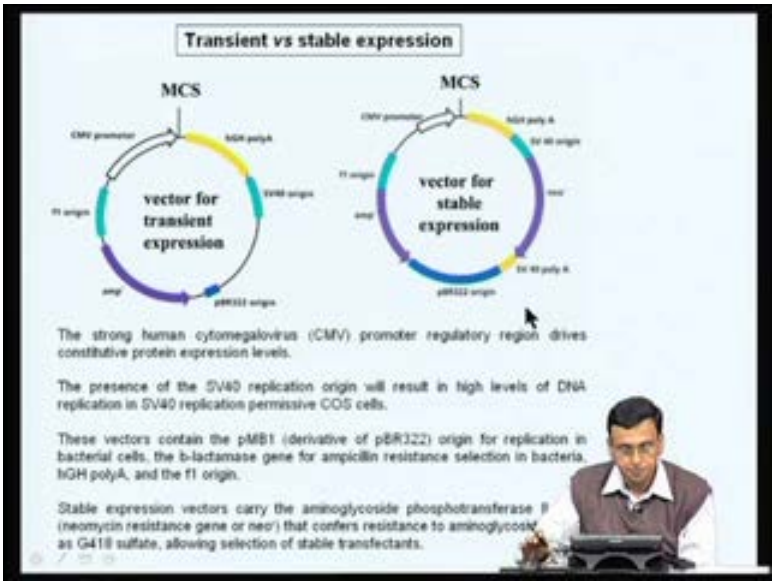
So, when you want expression only for a short period of time and when you would not do not really want to transition to be integrated into the host chromosome such transfections known as the transient transfection. So, in the case of a transient transfection, standard non selectable transformations of mammalian cells that replicate the plasmid DNA, but do not faithfully segregate the plasmid DNA to; progeny in the time of the plasmid will be lost from the cultured cells.

So, if you just want for example, to study certain promoter sequences or study or express a gene only temporarily and you do not want all the cells to be transfected then you adopt for a transient transfection. Usually, for many basic research studies for example, when you want some functional studies or when you want to make some deletion mutants of certain genes and look at their functional domains you use transient transfection studies, but usually for recombinant protein production in large scale you go for, what is called as a stable transfection.

So that, not only the cell once the gene gets introduced into the mammalian cell it gets stably incorporate to the host genome and therefore, every time the cell divides your trans gene along with the promoter also divides and therefore, the daughter cells also will contain your gene of your interest. And so in the case of your stable transfection you have to actually select those cells which have taken up your vector and therefore, you need to have appropriate markers, so that you can select those cells using their marker. So, the main difference between a transient transfection and a stable transfection is, in the transient transfection the gene is not incorporate in to the host genome.

And you do not actually select the cells for which have got stable transfection gene of interest whereas; in the case of stable transfection you actually induce a selectable marker in the vector, so that it can actually select those cells which have actually taken up the vector. So, a selectable marker is included in the plasmid vector, so that after the cells has been transfected the rare cells that incorporate the plasmid DNA into the genome through either DNA repair or homologous recombination can be isolated and cloned. So, the progeny of the transformed cell will inherit the plasmid including the selectable marker gene such as an antibiotic resistance gene. So, these are the two different methods you can either you can use a transient transfection method or you can use a stable transfection method for expressing your genes in a mammalian cells.

(Refer Slide Time: 26:45)



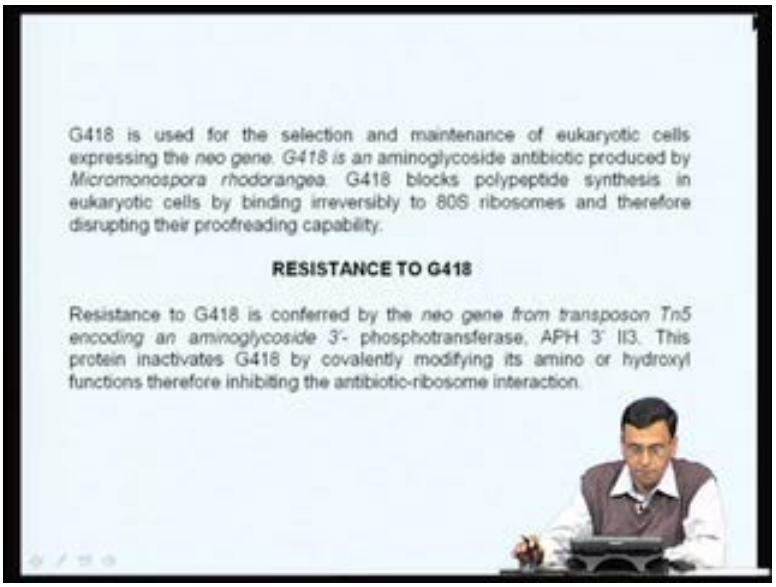
Accordingly if you want to do a transient transfection, you have to choose a different kind of a vector if you want to do a stable transfection you choose a different kind of a vector. So, in both the cases the vector should contain a strong promoter in this case for example, the vector is contain a cytomegalovirus promoter, so that the protein can be expressed constitutively in the cell line it also contains for a SV 40 origin of replication, so that when you introduce this vector in a mammalian cell line which contains or which has been transformed with SV 40 large t antigen every time the cell divides the vector also divides.

They also contain an origin of replication from p B R 322; so that these can be replicated efficiently in bacteria they also have an ampicillin resistance marker. So, that they can be e coli cells are taken up the plasmid can be selected, but the only difference between a vector for a transient expression vector for a stable expression is that, stable expression vector actually contain say a selectable marker gene in this case the aminoglycoside phosphotransferase two gene often referred to as the neo origin or neomycin resistance genes.

And you can see the neomycin resistance gene which is under the control of a eukaryotic promoter is present in a vector meant for a stable expression whereas, in a neomycin selection marker is absent in a vector which is meant for a stable expression. Now, let us

spend some time to understand how you actually select the eukaryotic mammalian cells which are taken up the plasmid DNA or the vector DNA.

(Refer Slide Time: 28:16)



The very commonly used selection marked for making stable cell line expressing your gene of interest is the neomycin resistance gene and you actually select these cells by adding a compound called G418.

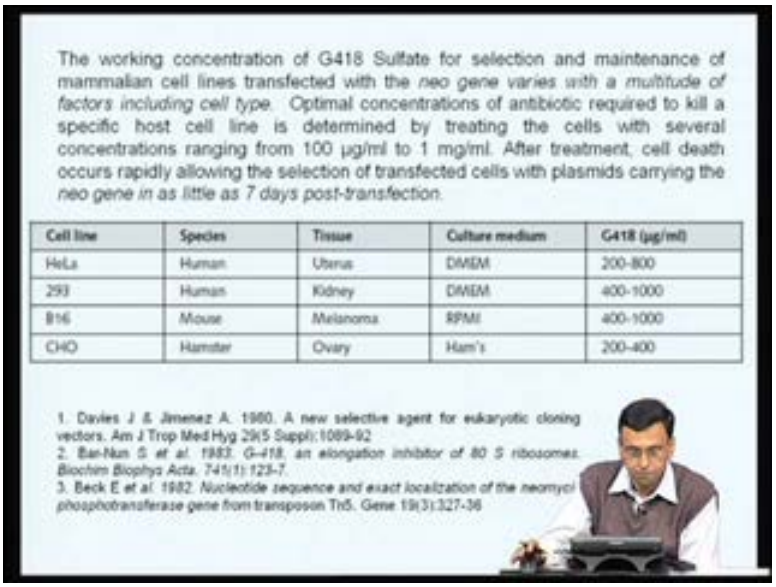
Now, G418 is actually used for the selection and maintenance of eukaryotic cells which express the neomycin resistance gene. Now, what is G418, G418 is an aminoglycocide antibiotic produced by Micromonospora rhodorangea. The G418 blocks polypeptide synthesis in eukaryotic cells by binding irreversibly to the 80 S ribosomes and therefore, disrupting their proofreading capability, so this is the mechanism by which the neomycin resistance genes half.

So, how do you develop resistance to G418, the resistance to G418 is conferred by the neo resistance neo or gene from the transposon Tn 5 encoding an aminoglycoside 3-phosphotransferase. So, this protein inactivates G418 by covalently modifying its amino or hydroxyl functions therefore inhabiting the antibiotic-ribosome interaction.

So, those cells which express the neo resistance gene modify the antibiotic, so that the translation is not blocked whereas, those cells which are not taken up the neo resistance gene, when you add this antibiotic it goes and blocks the translation and therefore, these

cells die. So, cells which have taken up your plasmid become resistant to G418 cells which are not taken up your vector become sensitive and therefore, they die.

(Refer Slide Time: 29:41)

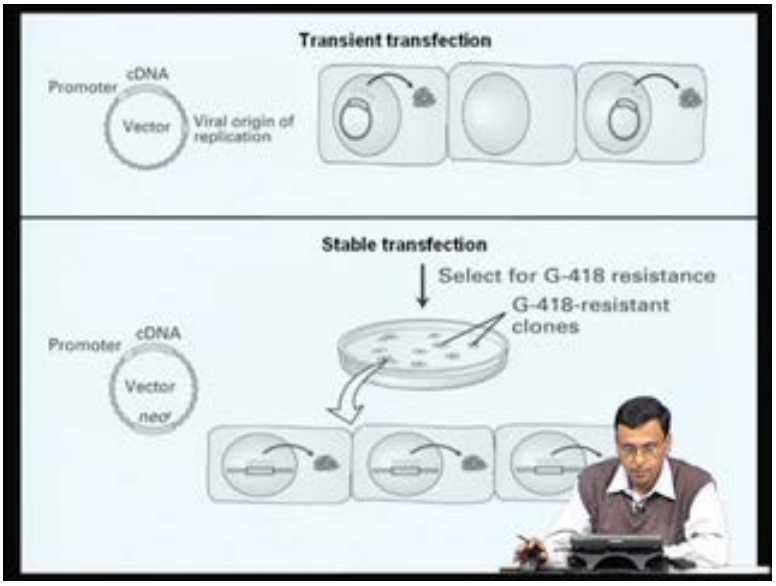


Many a times the exact concentration of G418 that you have used to select mammalian cells, which have taken up this vectors often varies from cell line to cell line. The optimal concentrations of G418 required to kill a specific host cell line, has to be determined by treating the cells with several concentrations ranging from 100 microgram per ml to 1 milligram per ml, and after the treatment the cell that occurs rapidly allowing the selection of the transfected cells with factors carrying numerous verses genes within about weak style.

So, if you once you do the transfection with your stable sector you add G418 and allow the cells maintain observer the cells for a week time all the cells which are not taken up the G formulate they all die whereas, cells which have taken up the vector become resist G foraminate and you get what are called as distant colonies. And these colonies be cloned and then be used for further analyses.

So, different cell lines have required different amounts of G418 from cells require less G418 to be killed and some cells required more G418 to be killed, these as to be slandered depending upon what kind of cell line you are using using for expressing your gene.

(Refer Slide Time: 30:47)



So, the two major distinction between a tense and transient transfection is, stable transfection is the vectors which are use for transient transfections, the vector which are used for transient tranfections is do not contain the selection mark such as resistance gene. So, when you introduce this trans and transaction, the gene may not be introduce in all the cells, only some genes take up the genes and these genes remain extra chromosomal and the DNA does not you are not actually selecting for those genes, with the in which the DNA is gone stably incorporated.

So, they remaining extra chromosomal and therefore they replicate and some cells may not take up the plasmid, but you are not selecting for those cells which are not taken up the plasmid. In the case of table transfection, following transfection of your vector you actually add antibiotic such as, G-418 and all those cells are taken up your gene and have got stably integrator these gene every time these cells divide, your gene of interest also divide and then therefore, those cells get selected they become resistance for G-418 and there four ultimately end up with clones in which your trans gene has got stably integrate in the whole chromosome, because not only there there progeny also become drag resistance.

(Refer Slide Time: 31:53)

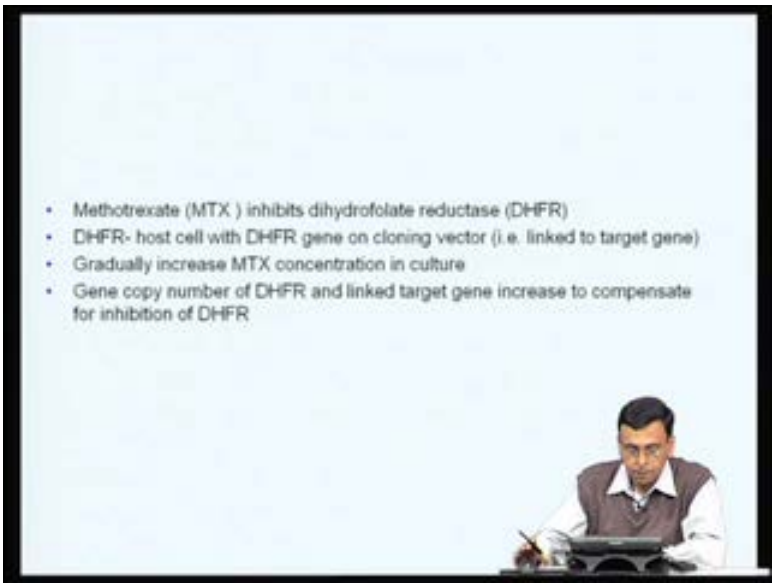
So, in addition to G-418 other very popular selectable markers which are used for selecting vectors following transfect to mammalian cells are, what is called is methotrexate is another drug, that is actually use and the selection marker which is used is called as the dihydrofolate reductase gene and by those cells which express the hydrofolate reductase become resistance methotrexate therefore, you can select these cells.

There also you can use Hygromycin B phosphotransferase gene as a selectable marker and we can actually select those cells using hygromycin B these again inhibitor of protein synthesis you can ask Thymidine kinase, again you can select those cells which express thymidine kinas using actually aminopterin which actually inhibits de novo Purine and thymidine synthesis you can use XGPRT are the genuine phosphor ribosomal transferase again the compound you can use to select cells are mycophenolic acid which inhibits de novo gene synthesis, you can also use adenosine delaminate is selection marker in this case you can use a modified nucleotide for selecting those cells, which are resistance to this which contain the express this particular gene.

So, a number of selectable markers can be used are selecting for developing or for isolation for stably transfected mammalians cells which are express your gene of your interest. I am going to spend some time to express about the how you select for methotrexate resistance (()) for express, because this is one of the markers which are

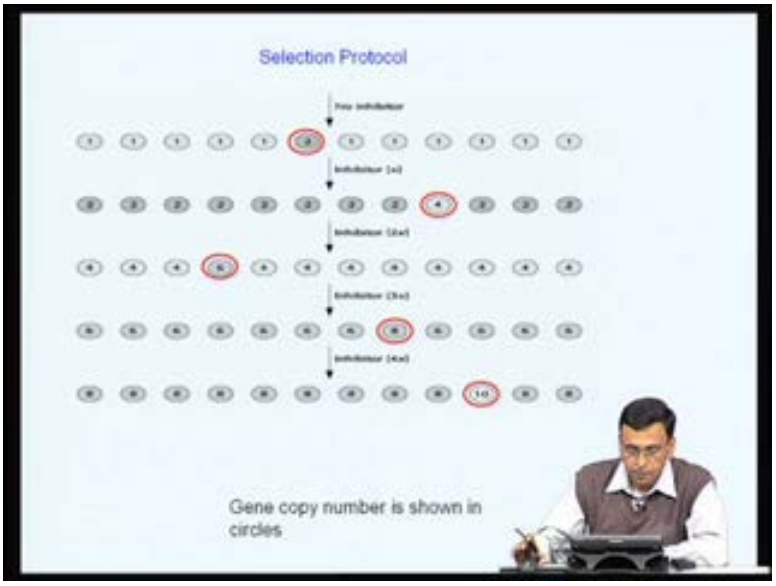
very efficiently use for recombinant protein production in a cells called Chinese hamster ovary cells.

(Refer Slide Time: 33:24)



So, methotrexate say it actually inhibits the enzyme dihydrofolate reductase, the enzyme dihydrofolate reductase host cell with DHFR gene is present on the cloning vector, and is therefore, linked to the target gene. And what do you do is, once you do the transfection with your stable expression vector contain gene of interest you slowly start increasing, expose the cell to increasing concentration methotrexate and when you do that, the cells some of the cells which contain this DHFR gene, the DHFR gene gets amplified to cope up the increase concentrated methotrexate. And therefore, by increasing with different generations by slowly increasing methotrexate concentration, you can increase the copy of number of your gene and therefore, more the copy number they can call at higher concentration methotrexate. Therefore, by gradually increasing methotrexate concentration; you can finally select the cell line which contains very high copy number of your trans gene and therefore, expressing very high levels of your protein.

(Refer Slide Time: 34:19)



This is the brief how do you the selection protocol, so you do the transfection in the hormones of the inhibiter, and usually most of the cells will contain one copy of your vector, very rarely do you have two copies, and then you start selecting those cells in the presence of the inhibiter.

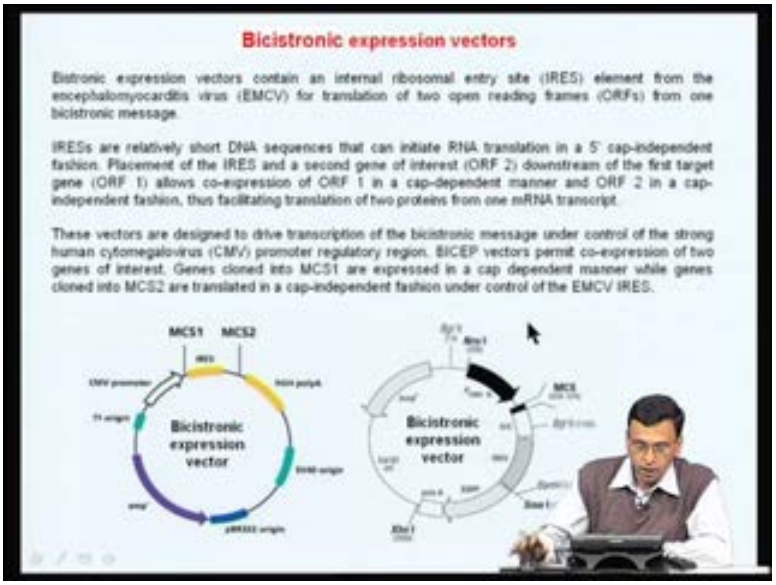
So, those cells which are not taken up the vector they all die, they become sensitive methotrexate, but there be some cells has it start increasing methotrexate concentration of every generation slowly in some of the cells the DHFR gene gets amplified (()). For example, in this case many cells contain only two copies for some cells may contain amplify the DHFR gene, so they contain four copies or if an express now higher concentration, they may now increase to six copies, eight copies, and ten copies so and so forth.

So, by gradually exposing the cells to higher concentration of methotrexate, you can actually score for cells, which are tolerant to higher concentrations, because the gene got amplified and therefore, they have more and more of DHFR, protein present and accordingly your trans gene copy number also goes up therefore, the expression level of your protein of interest also goes up therefore, the expression level of the protein of interest also goes up.

So, this is in brief how, one you how you express a, once you how design and expression by mammalian expression vector, and using a either viral enhancer or viral regulatory

elements or mammalian promoter elements and once it design you can either design a transient expression vector or a stable expression vector and then you select when become a stable transfaction using compounds like G418 or methotrexate, you can actually select for stable transformate exposing high high levels of your Tran protein of your interest. I am going to a spend some time to express another very important feature of new kind of expression is called as Bicistronic expression vectors.

(Refer Slide Time: 36:07)



Now, Bicistronic expression, so for we talked about expression vectors, in which you introducing one gene and therefore, only one protein can be made from the vector. There is also what is called as Bicistronic expression vector and they incorporate a very unique feature what is called as an internal ribosome entry site.

Now, let us now discus what these are, bicistronic expression vectors actually contain what is called as an internal ribosome entry site, abbreviated as IRES element and this is usually derived from encephalomyocarditis virus are the EMCV for translation of open reading frames.

As you all know for the normal translation of the eukaryotic messenger RNA's you required a cap the 5 prime cap, because the ribosome translation machinery actually recognize ribosome cap and then only the two subunits of symbol.

So, if for efficient translation of eukaryotic messenger RNA's is you require a cap transcript or a 5 prime cap is very very important, but there are certain viruses, which whose massager RNA's can be translated, even in the absence of the cap structure. And it turned out these viruses contain, what is called as a, internal ribosome entry site, this is actually DNA sequence, and when he had a rays sequence ribosome still to this IRES elements and do not actually require a 5 prime cap. So, many of this viruses have developed a mechanism by which, thereby they actually inhibit the capping mechanism and therefore, the host messenger RNA will not be translate efficiently, but their own messenger RNA's by virtue of which of containing these IRES, internal ribosomal intercedes can be efficient to translate.

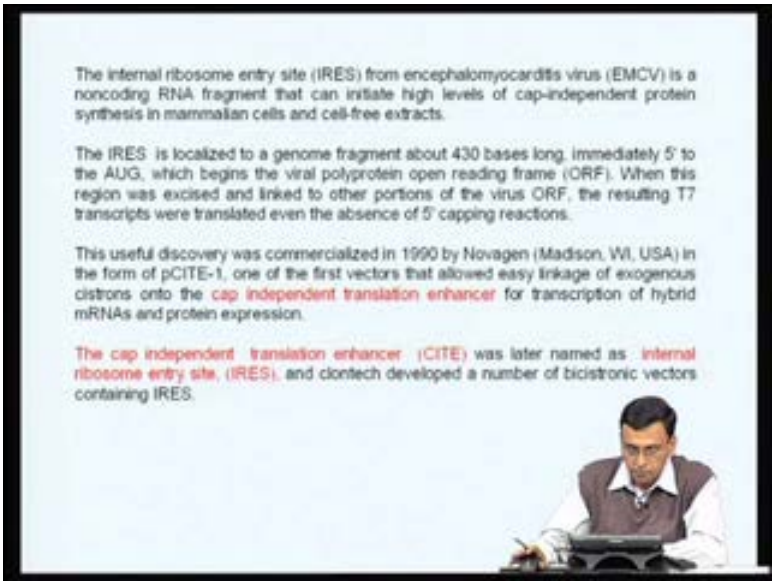
So, they make use of the host translation machinery for the translation of their messenger RNA organizes rather than translation of the host messenger RNA. So, people have actually take an advantage of this internal ribosome entry elements and incorporate this central ribosome entry elements, in an expression vector for example, here you have a expression vector, where you have a two multiple cloning sites and the first multiple cloning sites is downstream of the virus promoter.

So, when you clone you a gene of interest here the RNA polymerase, host RNA polymerase binds to promote transcribes your RNA, but you also have internal ribosome entry site and you can clone your second gene in to the multiple cloning, which is present here therefore, two genes can be clone, one before the IRES and one after the IRES. So, a single messenger RNA which is transcribe from this expression vector now, gets translated in a cap-independent fashion the first gene is translated in a cap in dependent fashion, where the second gene is translate in a cap-independent faction, because ribosome's can bind to both the cap structure they also can bind to the internal ribosome entry site. So, simultaneously two proteins can be synthesis from a single massagers RNA that is why they called as Bicistronic vectors.

So, the Bicistronic vectors are designed to drive the transcription of a Bicistronic message and to the control of a strong viral promoter, such as a same promoters and these vector permit co-expression of two genes of interest. Genes clone in to MCS1 are express in a cap dependent manner and genes cloned in the sub sequent MCS2 are translated in a cap-independent fashion under the control of the internal ribosome entry side.

So, you can make two protein from a single vector for example, in this vector you can clone your gene of your interest downs number cytomegalovirus promoter in this MCS and you also have another gene cloning for the green florescent protein; it is a very good reporter gene and which is under the control IRE sector. So, when you your clone gene of enter the MCS and you need to transfect this vector in to mammalian cells line, both the protein will be synthesis and you can usually score genes, which got transected by scoring for the expression of the UFB by just looking under the UV light you can see which kind of cells have taken up your plasmid, what are the cells which are expressing.

(Refer Slide Time: 39:49)

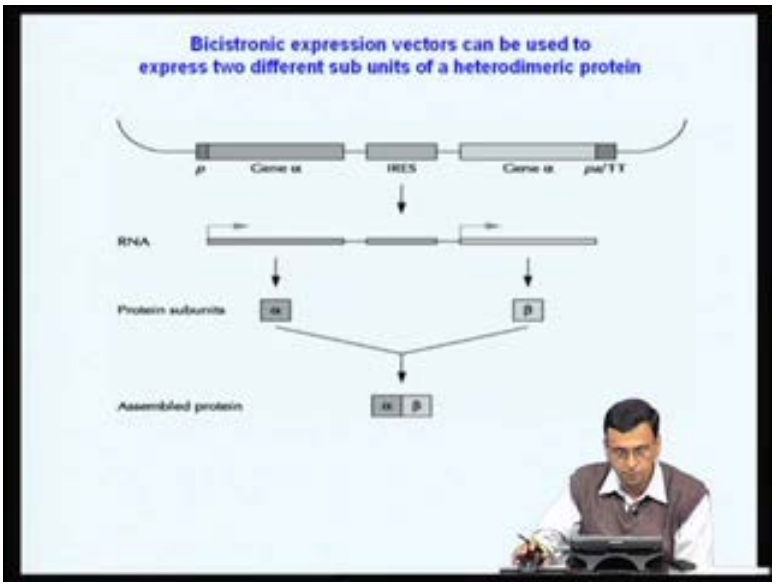


So, the internal ribosome entry side from the encephalomyocarditis virus is a non-coding RNA fragment that can initiate high levels of cap-independent protein synthesis in mammalian cells as well as cell-tree extract. The IRES is localized to genome fragment of about 430 bases immediately 5 prime to the translation cordon the AUG, which begins the viral poly protein open reading frame.

When this region was excised and linked to other protons of the virus ORF, the resulting transcription were translated even in the absence of 5 prime capping. This discovery, that these virus, animal virus such as the encephalomyocarditis required as well as polio viruses, they contain these kind of internal ribosome entry sites, has this actually commercializing 1990 by company called Novagen to form the first bioistronic vector called as pCITE-1. So, sites caps work transfer, cap independent translation enhancer.

So, you do not require capping anymore and therefore, you can use this vector to generate RNA'S in a cap, translate RNA in a cap independent fraction. Following this cap independent translation enhancer was later named as internal ribosome entry site, and a company called Clontech developed a number of bicistronic contains vectors which contains this internal ribosome entry site.

(Refer Slide Time: 41:04)

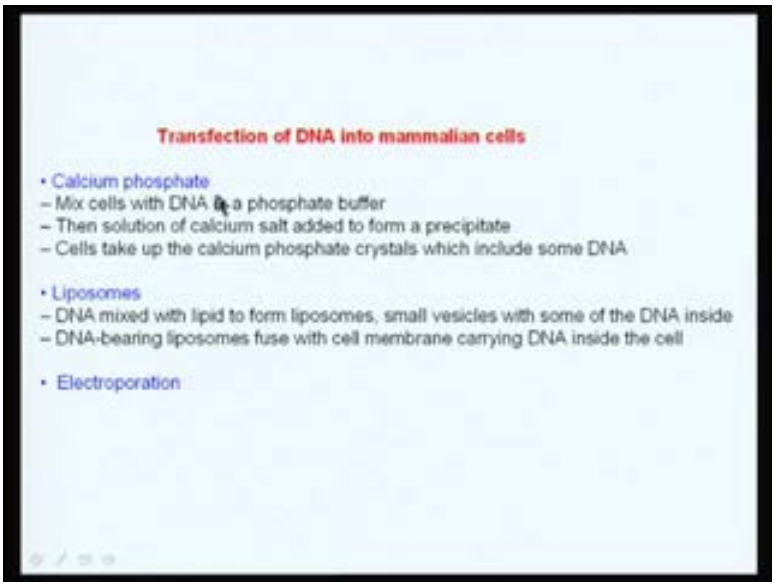


What are the advantages of using these kinds of Bicistronic factors is that, suppose you want to express a protein, which is heterodimeric it contains two different sub units. So, what you can do is, you can clone this two genes coding for each sub unit, one under the control of the promoter, another downstream of IRES and by just transecting one vector both this sub units can be made in the single cell and heterodimeric and you can get a functional protein.

So, when you want to express many proteins or heterodimeric you can express both sub units using this kinds of biosystematics vectors, and you can get a functional protein by using a single vector; rather than cloning each one of this sub unit genes in to two different vectors and co-tranfecting this plasmids. So, so for we have discuss about the what kind of vectors of people use for expressing genes in mammalians cells, let us now try to understand, what kind of a transfection methods, one has to employ if they have to introduce this gene in to vector in to mammalians cells.

As now, where many of this vectors that were discussing the are all negatively charged. The mammalian cells have a bye layer, which is also negatively charge therefore; mammalian cells cannot take up the DNA just like that. So, we have to introduce your vector in to mammalians cells, you need to some specific mechanisms or specific treatments.

(Refer Slide Time: 42:33)



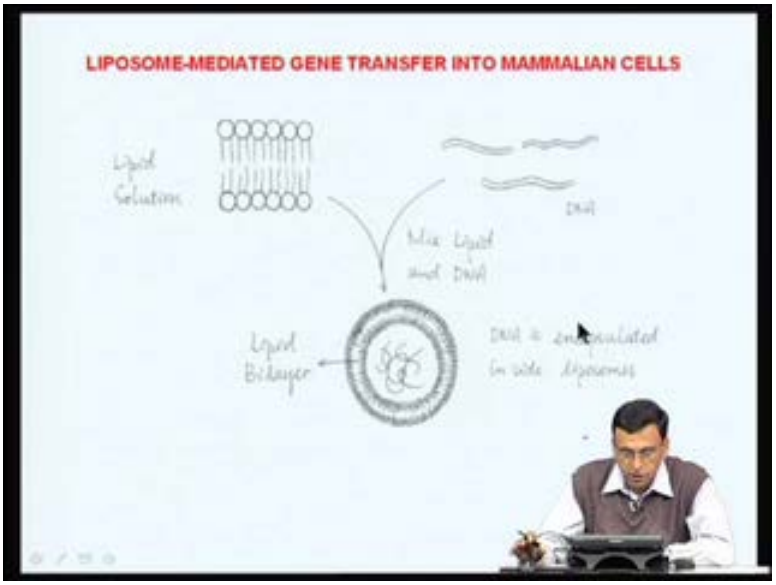
So, some of the methods that are routinely used for introducing these vectors, expression vectors in mammalians cells are by do what is called as calcium phosphate precipitation. What you do is, you take your DNA which is negatively charged, and mix the DNA with calcium salts usually calcium chloride or calcium chloride.

So, calcium which is negatively charges, this is positively charged binds to the negative pass fate group of the DNA and therefore, makes it kind of neutral and therefore, this calcium phosphate crystals are now, efficiently taken up by the mammalian cells.

So, you are a temporally neutralizing negative charge of the DNA and therefore, the DNA can taken up by the plasma membrane inside. This is a very old method of transfection, which is been very successfully used for number of basic applications for understanding eukaryotic gene expression in the 90's and early 2000, but many modern methods of transfectional taken over very rarely now people use, calcium transaction phosphate media transfaction these days.

The most prefer method of introducing vectors, expression vectors in to mammalian cells is now using, what are called as liposomes or cationic lipids as well as by electro poration. So, in the case of liposomes the vector of your choice is mixed with lipids to form what are called as liposomes, which are nothing but, small vesicles with a DNA inside and these DNA bearing liposomes fuse the cell membrane and therefore, the DNA inside the cell.

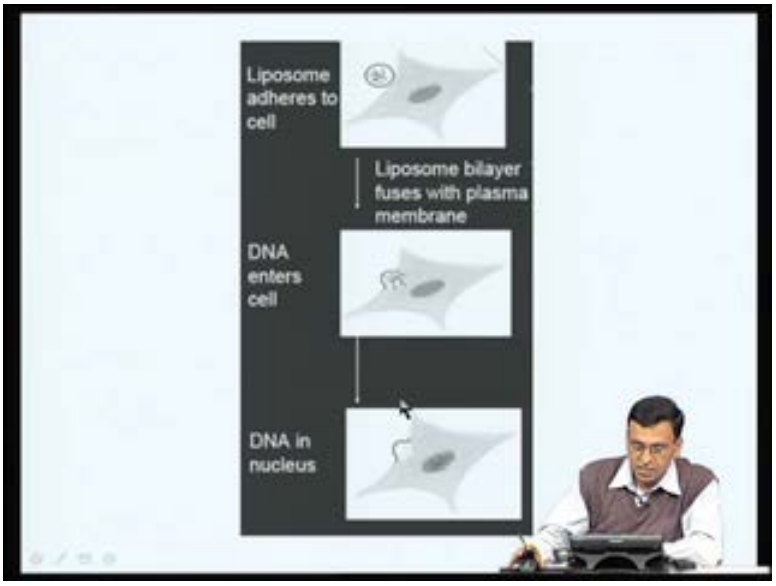
(Refer Slide Time: 43:47)



This is now shown victory here, so you actually vertex your DNA of interest with, what are called as the lipid solutions as lipid has both the hydrophobic, and hydrophilic positions. So, they form actually lipid vesicles and your DNA gets trapped inside these leaped vesicles and if you now, add this in to the cells in culture of the lipid actually fuses with the lipid by layer and the DNA gets deliver inside.

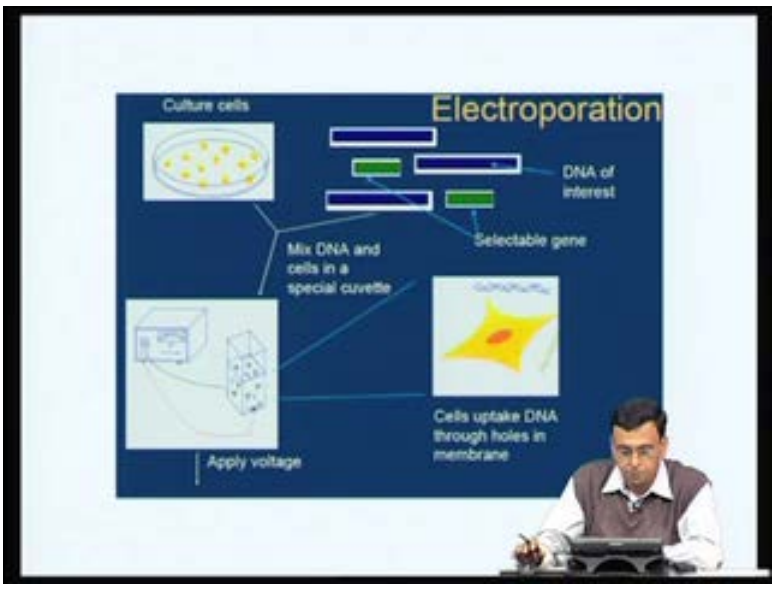
So, is the very popular method of tranfecting or some mammalians cells by using expression vectors. These cells also have variation of this liposomes you also have what are called as cationic lipids which are nothing but, positively charged lipid and therefore, you can use this positively charge lipids when you mix your DNA with this positive charge lipids; the positively charged head of the cationic lipid, now binds with the DNA and they form neutral lipid particle, which is now efficiently taken up by the mammalians cells.

(Refer Slide Time: 44:36)



In the case of, so this what happens the liposome adhere to the cell membrane, and the lipids fuse able cell membrane and the DNA gets inside and therefore, they go inside nucleus and the DNA enter cells transcribed.

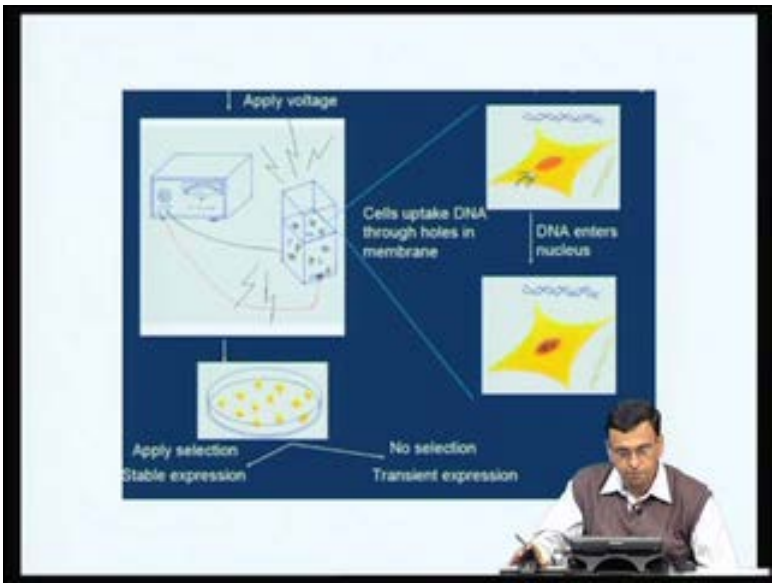
(Refer Slide Time: 44:51)



These how, cationic lipids are liposomes can be use for a introducing your genes, in the case of electro proration you actually take your DNA of interest mixed with the cell of your choice and then subject them to a short electric current. And when you give this

apply a specific voltage he creates pores inside the membrane cells, cell membrane and DNA gets inside the mammalians cells.

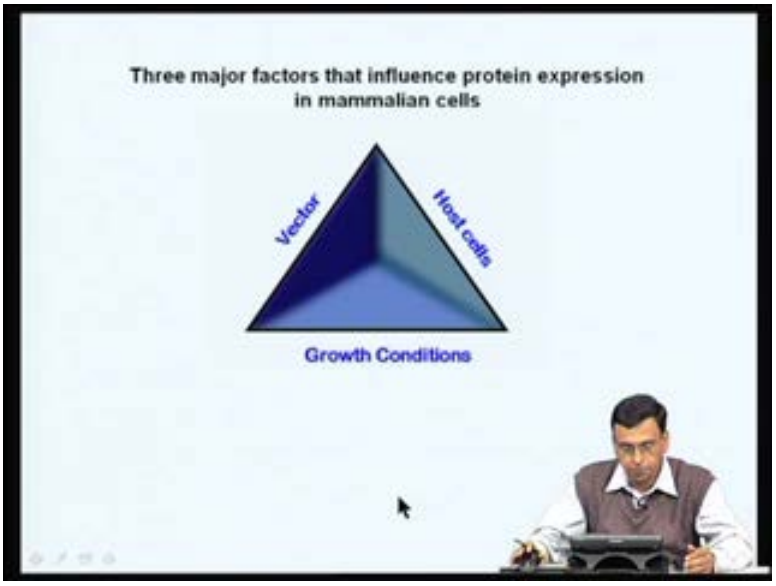
(Refer Slide Time: 45:10)



So, when you apply in electric pulse, when the cells contain DNA the pores are made inside the cell membrane the DNA gets inside and therefore, and once the DNA gets inside they can, then either if the vector contains the selection marker, you can apply the selection pressure and select for stably transected cells or you can just go ahead with the trans in transfection.

So, three very population methods of introducing mammalian vectors in to mammalian cells, is calcium phosphate precipitation, electro portion, as well as lipid mediation gene transfer using a liposomes or cationic lipid gene transfer.

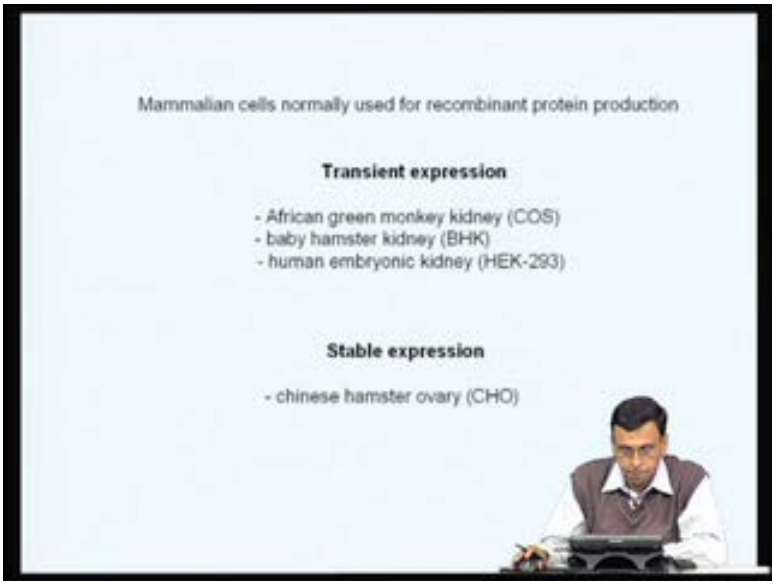
(Refer Slide Time: 45:45)



So, some of the very important factors which effect the expression of your genes in mammalian cells are, what kind of vector you design as I said depending upon the promoter as well as other sequence like inclusion of intone proper poly relation signal some so and so forth, efficient mammalian character has been designed. So, designing an appropriate highly efficient mammalian expression vector is one of the major requisites for efficient expression of your trans gene or producing recombinant proteins in large amounts.

The next important step is, what kind of host cells you are choosing? There are many host cells, which cannot be efficiently tranfected and there are also many host cells which are do not actually carry out appropriate post translation modifications. Therefore, choosing the appropriate host cells is the very very important once you design appropriate vector, then the most important job, another important criterion is the growth conditions. So, what kind of growth conditions you can actually grow this host cells in to high cell densities, so that you can produce protein in large amounts.

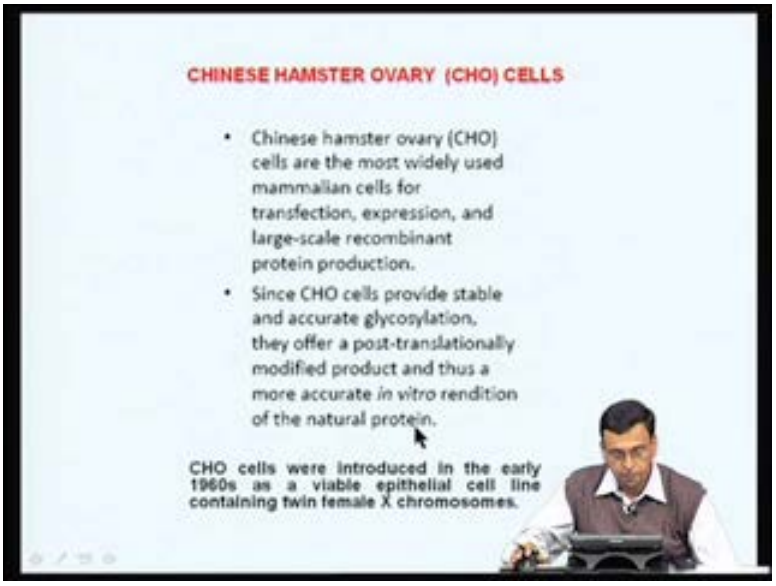
(Refer Slide Time: 46:49)



So, number of parameters I can determine by the produce recombinant protein in high levels or not. So, as I said, if you want to transient transfection usually people choose something like the cause cells are the baby hamster kidney cells or human embryonic kidney 293 cells, these are the usual mammalian cells people choose for, doing transient transfection studies are you want to express your gene in a transiently.

Whereas, if you want a express your gene in stable form usually the Chinese hamster ovary cells are mostly prefer for stably transforming genes, and especially for producing recombinant proteins in large amounts.

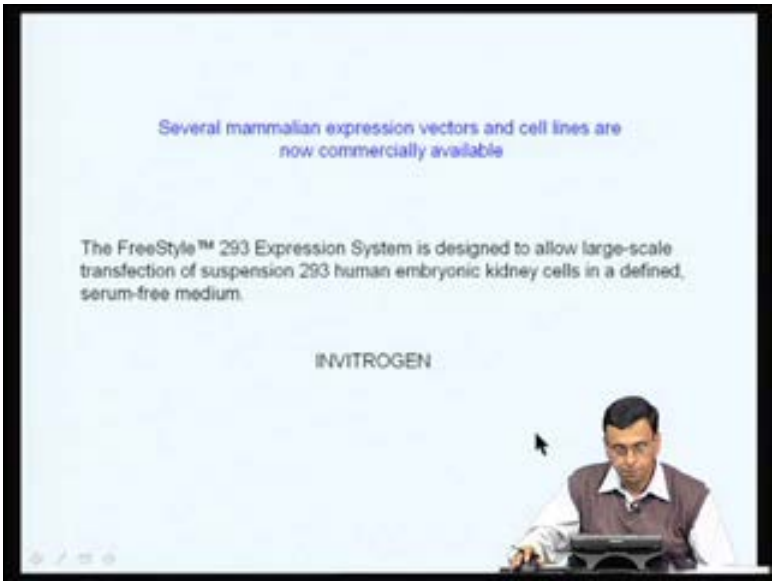
(Refer Slide Time: 47:24)



For example erythropoietin is primarily produce using the Chinese hamster ovary cells. So, the Chinese hamster ovary cells are the most widely used cells for mammalian expression of mammalian genes especially for large scale recombinant protein production and the CHO cells provides very stable and accurate Glycosylation they half of post translation modified product that is moral that is that is almost same as the natural protein. In fact, the CHO cells are introduced in the early 1960s as a viable epithelial cell and containing two female x chromosomes.

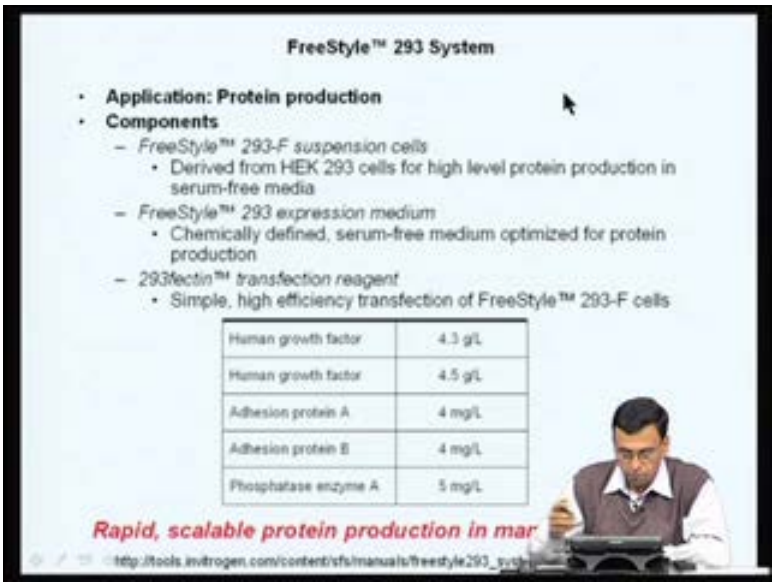
So, the CHO cells are as it is called is very widely used for recombinant protein production, and erythropoietin which we discussed in the beginning of this lecture, is actually produced using the CHO cells.

(Refer Slide Time: 48:02)



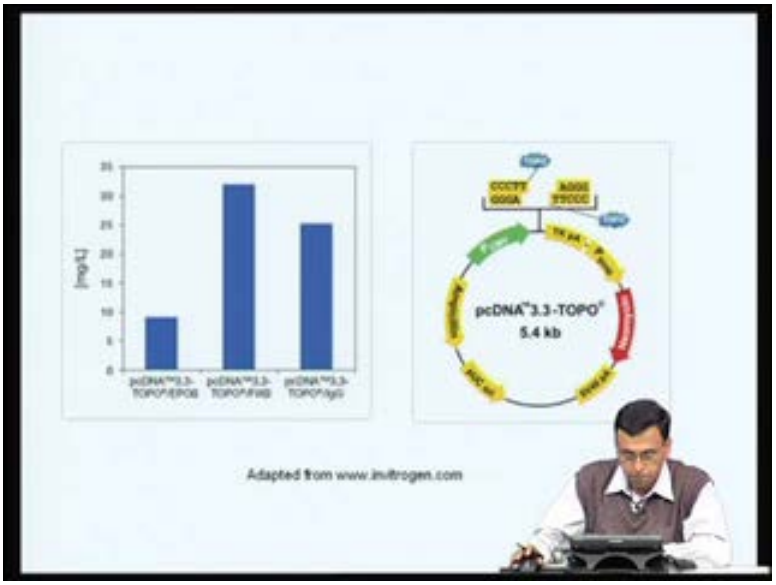
Several mammalian expression vectors and cell lines are now commercially available, for example, nitrogen cells actually what is called as the free style 293 expression system the if you just by a kit they supply the vector, they can supply some all the transaction reagents, as well as for methods for stably select in this stable cell line. And you can also the growth conditions and all this can be, you do not have to slag you can simply by this states and you can express your gene of your interest.

(Refer Slide Time: 48:31)



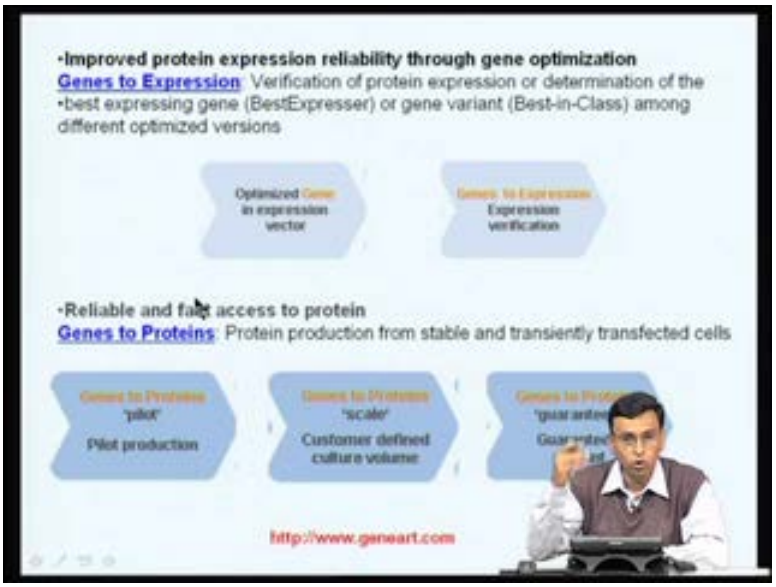
The for example, using this kind of a free style 293 system it has been shown that a number of proteins of, can be expressed as very high levels from gram per liters to even mile grams per liter, a very high level of expression can be obtained using these kinds of a commercially available vectors and commercially available cell lines.

(Refer Slide Time: 48:50)



Here for example, using this particular vector and using the cell lines nitrogen actually tells you that, number of proteins like erythropoietin factor nine eminoglobin has been expressed at mile grams level using mammalian cell lines.

(Refer Slide Time: 49:06)

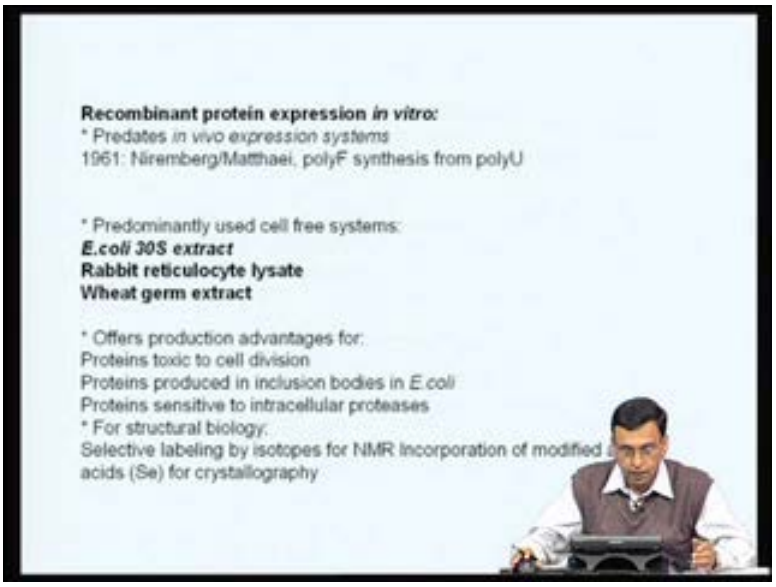


There are actually companies, such as G not where you do not have to learn, how to express proteins if you simply tell them, what kind of mammalian gene you want to be expressed and pay the money, they will design the appropriate expression system, they will make the expression vectors, they will introduce in to the cell lines and they will express the protein and give you the recombinant protein of your choice.

So, these are called the custom protein expression, now systems are now available you can go to this companies and just pay the money and they will express the protein of your interest and give you the protein.

So, let us now spend about some time to understand cell-free protein synthesis. So, for all our focus has been on expressing genes inside the cells, where it is e coli cells, where it is yeast cells or insect cells or mammalian cells, but there also now cell-free systems which are available. So, you can make your protein in a test tube without any cells any cell free manner.

(Refer Slide Time: 49:59)



Now, the first recombinant protein expression in a cell-free system dates back to 1961 wherein Nirenberg and Matthaei actually demonstrated, if you have an RNA which contains (()), you can actually make a polypeptide containing (( )) this is one of the first successful demonstration of a cell freed protein synthesis. In fact, it made a very important contribution for the understanding of the genetic code and got the Nobel Prize for it.

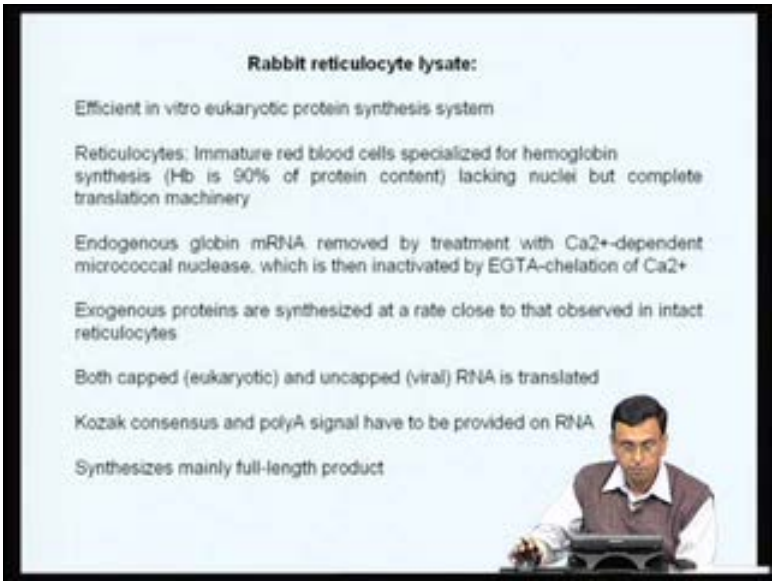
Following it people started making this kind of cell free extracts for example, cell-free extracts have been made. So, a cell-free extracts actually should contain, what it should contain all the amino acids it should contain the all the amino acids TRNA's should contain the riposomes and therefore, if you have and also proper energy source therefore, you simply add an RNA the riposome it can translate your RNA.

So, cell-free extracts have been made cell-free extracts capable of protein synthesis have been made from E coil's, they have been made from Rabbit reticulocyte tysate and have been made from wheat germ. What is the advantage of a cell-free protein expression system them.

You can actually use these cell-free protein systems to make proteins, which are actually toxic to cells you can use proteins which are, which can be made soluble proteins can be made using cell-free systems, which are normally going to inclusion bodies in E coli. Proteins which are sensitive to intracellular proteases, can be now made using cell-free proteases systems and if you want to label proteins for with certain heavy isotopes for a NMR studies cell free systems can be made used of.

So, the cell-free protein expression system is a viable alternative for making proteins inside the cells. As I said, the E coli cell-free protein system have been expressed, have been or now commercially available; it as a very simple operate translation apparatus and a comparative simple control of initiation. Following this the Rabbit reticulocyte lysate systems, have been developed for efficient in vitro protein synthesis system.

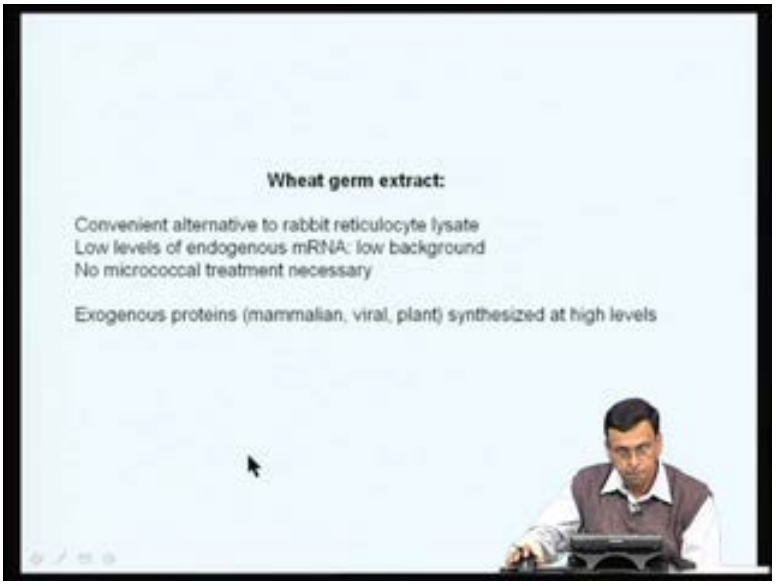
(Refer Slide Time: 51:43)



Reticulocytes are nothing but, immature red blood cells specialized for hemoglobin synthesis lacking nuclei, but they have complete translation machinery. So, what you do is you take you take this reticulocytes and degrade the endogenous globin messenger RNA present in the reticulocytes by treating them with micrococcal nuclease, which choose upon the RNA. And then you inactivate the nucleus by adding, by chelating with EGTA; and then exogenous proteins are synthesized at a rate close to that observed intact reticulocytes by adding messenger RNA synthesized at rate reticulocytes.

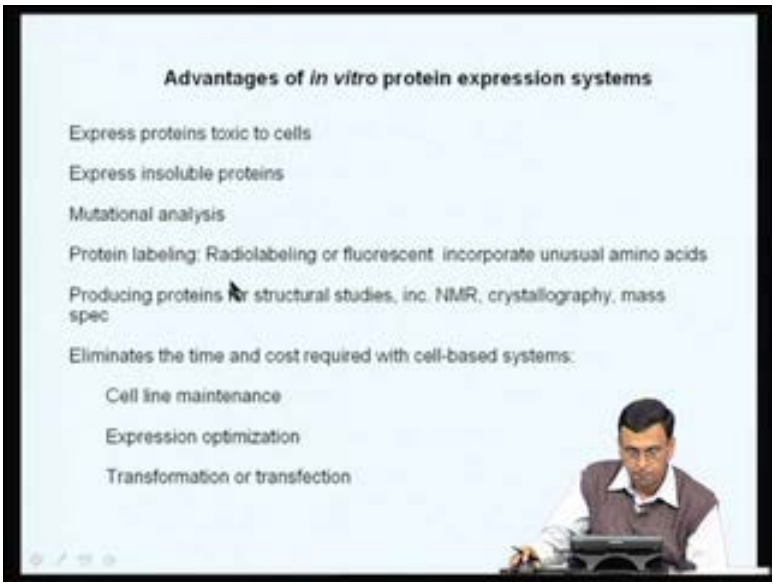
So, both caped as well as uncapped RNA is can be translated, using this kind of a rabbit reticulocytes systems, and you can get nearly full-length protein system synthesis in the rapid reticulocytes.

(Refer Slide Time: 52:30)



The same way using a plant source you can actually make a wheat gremials it, again the using the similar procedure you destroy the endogenous RNA by micrococcal nucleus; and then you add your messenger RNA your messenger RNA will get translated and your protein can be made.

(Refer Slide Time: 52:45)



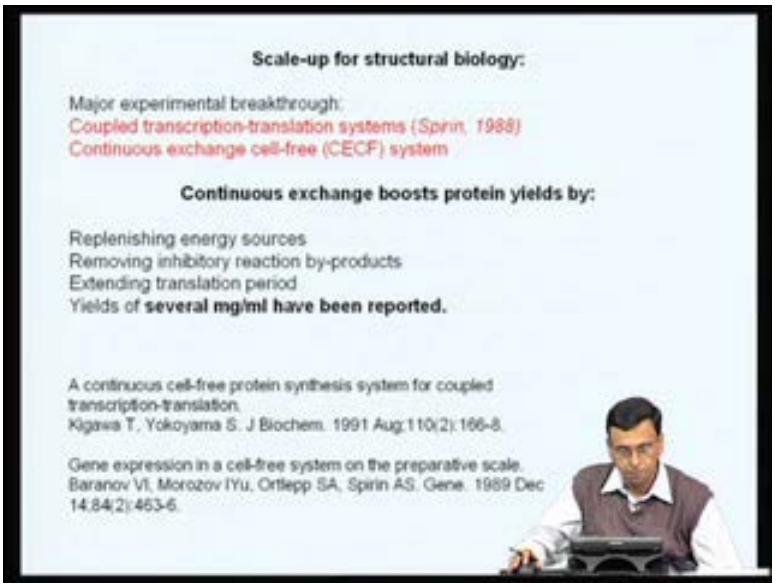
As I said, it as an advantage many protein system toxic cells can be expressed using in vitro, protein expression systems, they can be used for number of structural studies for like crystallography mass spec NMR and so on and so forth. It eliminates the cost and

time required for cell based systems, you do not have to maintain the cell lines, you do have to optimize the expression, you do not have to transformation or tranfaction, you have life cell-free system for making your proteins.

(Refer Slide Time: 53:09)

A number of commercial extracts are available for example, promega cells all these E coli cell-free system, rabbit reticulocytes as well as wheat germ extracts, simply by the cell-free systems and simply put your RNA of interest there, and you can get protein of your interest.

(Refer Slide Time: 53:26)

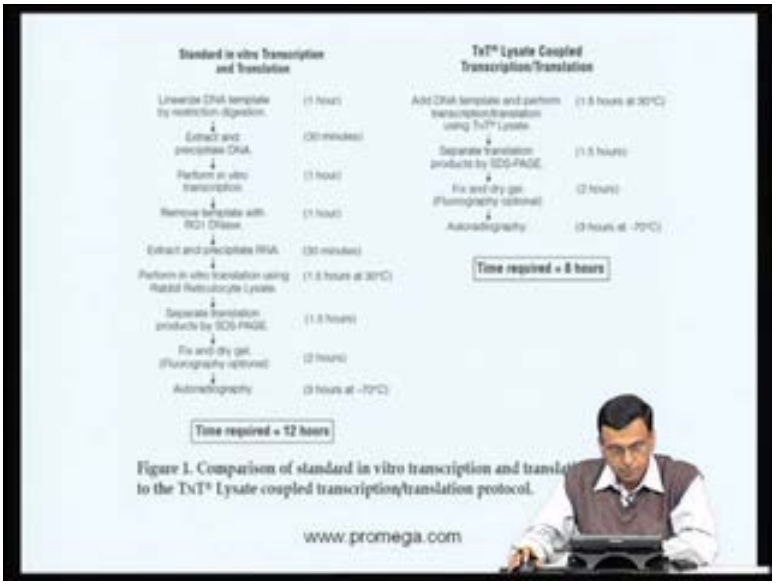


Two important breakthroughs occurred, in this cell-free expression systems, for mammalian systems. One is called as a coupled transcription translation systems or what are called as a continuous exchange cell-free system or the CECF system.

Now, in the early in cell-free systems you have to first transcribe your gene using certain prokaryotic RNA polymerases, like three seven RNA polymerases. And then, take this RNA purify this RNA, and then add this RNA to a cell-free translation system make your protein. Now, in the case of a coupled transcription translation systems you simply add your plasmid, which contains your gene, the in the same system the gene will get transcribed by the phase RNA polymerase and then the RNA will get translated. So, in one shot both transcription and translation takes place the same system and therefore, it is called as a coupled transcription translation system. Whereas, in the case of the continuous exchange cell-free system, is a very important it happen because, in the normal cell-free one of the important problems is you cannot make large number of large amounts of your protein because the energy gets depleted.

Whereas in the case of continuous exchange cell-free systems energy is continuous replicate therefore, a large number of protein can be made even mile grams amounts of proteins, will be made using this kind of a cell-free translation systems now.

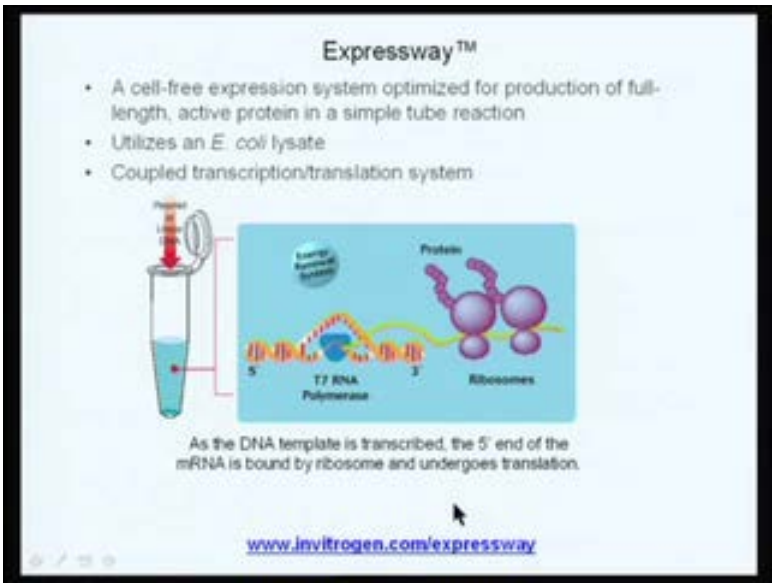
(Refer Slide Time: 54:44)



So, I just taken some of the protocols, which are given when the company websites like promega actually gives you, how actually you can do both transcription and translation

in the using this coupled transcription translation systems, and you just have to buy this sites and you can make your protein of your interest.

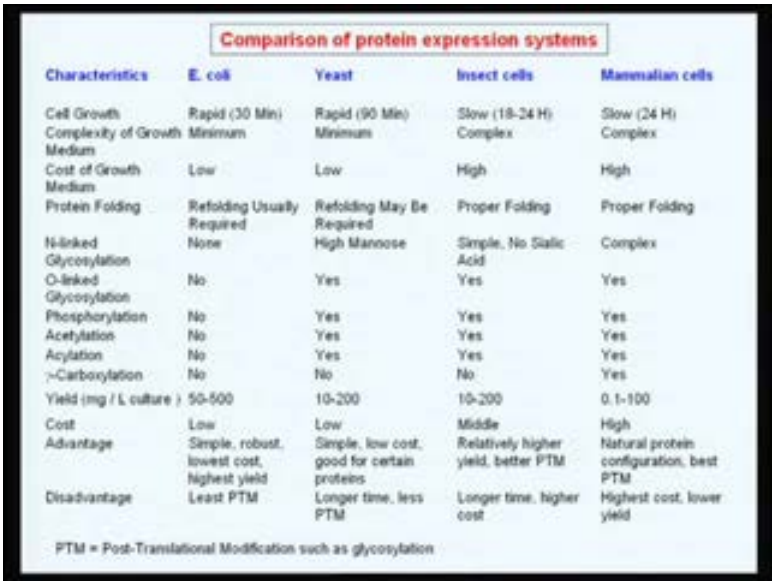
(Refer Slide Time: 55:00)



Similarly, using the commercial kits like expressway, you can actually synthesis large amounts of this kinds of a proteins using this kind of a coupled transcription transition systems and again the company gives the exact protocols, how to go about and do this cell-free transcription machineries.

Rosh again cells very efficient cell-free transcription translation systems for making many gram amounts of proteins or one can go to the Rosh website, and get the mechanism by which you can actually make these proteins.

(Refer Slide Time: 55:33)



Comparison of protein expression systems

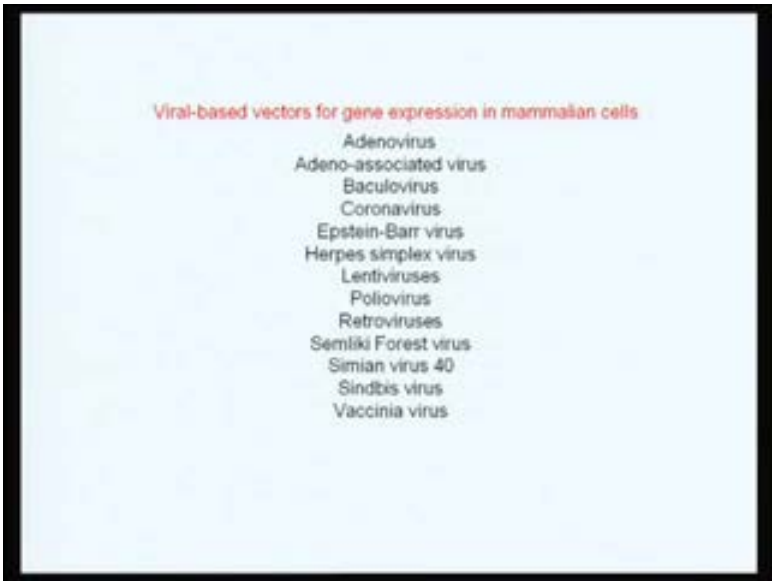
Characteristics	E. coli	Yeast	Insect cells	Mammalian cells
Cell Growth	Rapid (30 Min)	Rapid (90 Min)	Slow (18-24 H)	Slow (24 H)
Complexity of Growth Medium	Minimum	Minimum	Complex	Complex
Cost of Growth Medium	Low	Low	High	High
Protein Folding	Refolding Usually Required	Refolding May Be Required	Proper Folding	Proper Folding
N-linked Glycosylation	None	High Mannose	Simple, No Sialic Acid	Complex
O-linked Glycosylation	No	Yes	Yes	Yes
Phosphorylation	No	Yes	Yes	Yes
Acetylation	No	Yes	Yes	Yes
Acylation	No	Yes	Yes	Yes
γ-Carboxylation	No	No	No	Yes
Yield (mg / L culture)	50-500	10-200	10-200	0.1-100
Cost	Low	Low	Middle	High
Advantage	Simple, robust, lowest cost, highest yield	Simple, low cost, good for certain proteins	Relatively higher yield, better PTM	Natural protein configuration, best PTM
Disadvantage	Least PTM	Longer time, less PTM	Longer time, higher cost	Highest cost, lower yield

PTM = Post-Translational Modification such as glycosylation

So, in this chart I have actually listed all the expression system discussed so far, we have discussed about E coli expression systems, we discussed yeast expression systems, insect expression systems and mammalian expression systems.

And again you have to choose your basic you , and your decision to see what kind of expression system, which you have to choose each one has its own advantages and disadvantages ultimately the choice depends on the particular gene of your interest you are interested in what kind of protein you are interested in and if you have to make a protein which is you have to make a protein, which does not require post translation modification you go from E coli system and then gradually try out other systems and ultimately if none of the systems were ultimately chosen an expression, because that is the most expensive.

(Refer Slide Time: 56:23)



So, with this we have come to the end of this lecture, so we have discussed various aspects of non viral methods of introducing genes and making proteins and mammalian expression systems, what will do in the sub sequent class is to look at some of the viral-based vectors for gene expression mammalian cells.

What kind of viruses are people using to introduce genes in to mammalian cells, and how various viruses such as adenovirus, retrovirus, poliovirus, lent viruses and vaccine viruses these viruses can be modified. So, that genes can be introduced in to the genomes and recombinant viruses can be generated, and this recombinant virus can now be infected various mammalian cells and proteins can be expressed in very efficiently non genomes, I think I will stop here.