Eukaryotic Gene Expression

Prof. P. N. Rangarajan

Department of Biochemistry

Indian Institute of Science, Bangalore

Module No. # 01

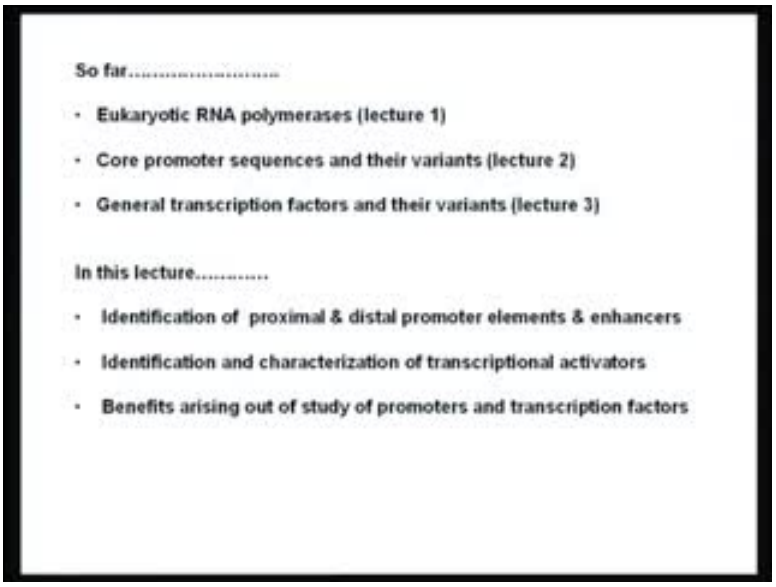
Lecture No. # 04

Gene Regulation in Eukaryotes:

Proximal and Distal Promoter Elements Enhancers and Silencers Gene-Specific Regulators

This is Rangarajan here. We are going to continue the fourth lecture in the course on eukaryotic gene expression.

(Refer Slide Time: 00:29)



What we did in the previous three lectures is to briefly describe about eukaryotic RNA polymerases. We discussed about the various components of eukaryotic RNA polymerases - polymerase 1, 2 and 3, how they function and in the activation of gene expression in eukaryotes. Then we stuck confined ourselves to understanding the core promoter elements and how variations within this core promoter elements can actually lead to differential gene regulation.

Then, I also very briefly mentioned about the various general transcription factor, which interact with RNA polymerase and constitute what is called as a preinitiation complex, and how variations within the general transcription factors itself, like their tissue specific expression and variations in recognition of core promoter elements, that itself contribute to differential gene regulation.

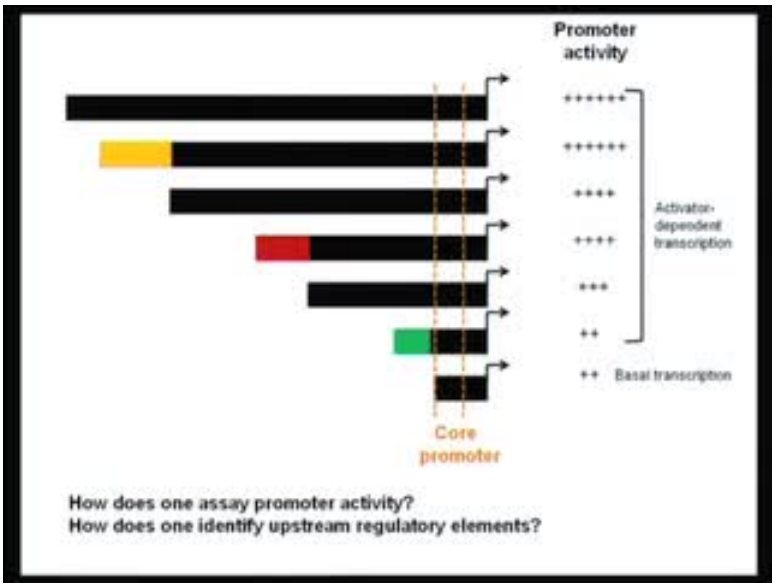
Now, what we are going to do today is to move a little bit away from the core promoter, and then, focus more about move towards proximal and distal promoter elements, and by end of this class, you will realize that the promoter actually consists of several kilo bases of sequences upstream of the transcription start site and a number of cis-acting elements, you can call them as proximal promoter elements or distal promoter elements or enhancers, all these things together actually play a very important role in the regulation of gene expression eukaryotes.

So, we will now study little about the identification and characterization of various transcriptional activators, which actually go and bind to this kind of a proximal and distal promoter elements and binding of this transcriptional activators actually play a very important role in the modulation of levels of gene activity and also, the title of this course is eukaryotic gene expression basics and benefits.

So far, I have not talked anything about the benefits arising out of eukaryotic gene regulation. Today, maybe, we will just spend about maybe, 5 minutes to tell you how understanding the various promoter elements and transcription factors actually paved way for development of what are called as recombinant protein expression systems by which you can actually express a number of recombinant proteins and this has now become a big billion dollar biotechnology industry.

So, the way they have structured the course is that we will talk mostly about basics, but at the end of maybe about 4 or 5 lectures, we will review the whole thing and really see, having studied the knowledge in the basic resource, has it really led to any applications or any uses, or any benefits have come out of some of this basics. So, that is the way we are going to proceed further.

(Refer Slide Time: 03:03)



So, let us now focus our attention on promoter elements, which are kind of further away from the core promoter. So far, we had we have been confining ourselves only to the core promoter region of a eukaryotic gene. Now, we all know the transcription starts from the transcription start site and in all my lectures, I indicate this as with an arrow. So, this is what is called as the plasma.

Now, what has been shown, as people started studying about the various cis-acting elements and various promoters sequences, it became very clear that in addition to the core promoter region, there is also a huge region, some sometimes this region which is upstream region from the transcription start site can run to several kilobases upstream and if you actually make some deletions within this promoter region also, the expression of the downstream gene is modulated.

You can see this is what I have written as a promoter activity and for example, let us now take this promoter region, which in addition to core promoter consists of further upstream sequences and let us designate the activity of this promoter; that is this promoter activity has let us say 6 pluses, means very highly expressed.

Now, people have started asking the question, in addition to the core promoter region are there other promoter elements which are essential for the activity of the promoter. The way they went about is that we actually start doing deletion mapping, wherein you start

deleting different regions of the upstream region and ask the question, how do these deletions in the further upstream region of the promoter affect the promoter activity.

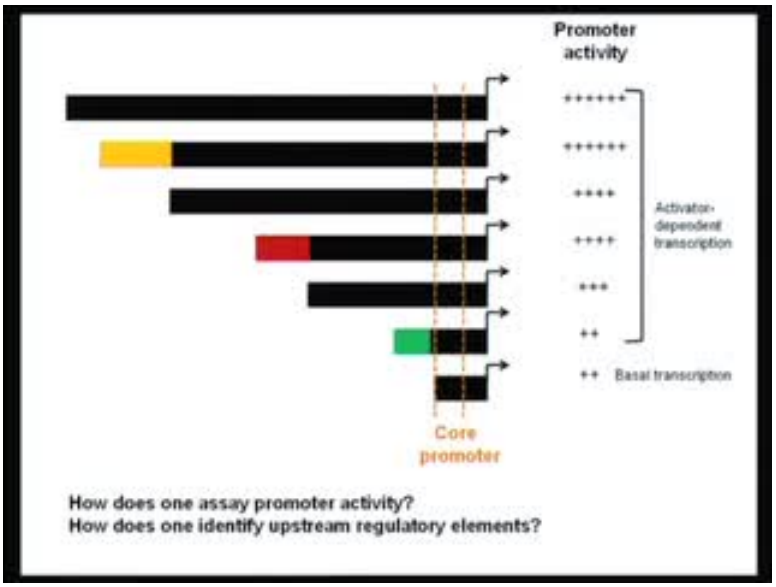
And this particular example, you can see when you make a deletion from here, when you delete this particular region, there is not much change in the promoter activity suggesting that there may not be very important regulatory elements within this region of the promoter. But you can see when you delete further, when you delete from here to here, there is a drop in the promoter activity and similarly, when you make a deletion from here to here, there is a drop in the promoter activity and so on and so forth.

So, this is what I have illustrated here. So, what this means is that when you make a deletion in this upstream region, you there is a decrease in the promoter activity suggesting that this region which I (( )) an orange box, may actually harbour some very important cis-acting elements to which certain protein factors may be binding and they inturn may be interacting with the general transcription factors and RNA polymerase (( )) and enhance the rate of initiation of transcription.

Similarly, you can see, there may be in this particular case, there may be two other totally, there are about three regions and if you delete these promoter regions, there can be a decrease in the promoter activity suggesting that this region of these promoters may harbour some important regulatory elements to which certain protein factor may be binding and they inturn may be interacting with the core promoter region and contributing to transcription initiation.

So, it became very clear; when you start studying about eukaryotic gene regulation, so far, we have been talking only about basal transcription where RNA polymerase II and general transcription factors are binding to a core promoter region, but it is not becoming very clear in addition to this region, there are also other upstream regions, which are essential for transcription activation of eukaryotic genes and therefore, these regions which are further upstream of the core promoters to which certain transcription activators may be binding and further activating transcription, refers to what is called as an activator dependent transcription.

(Refer Slide Time: 03:03)



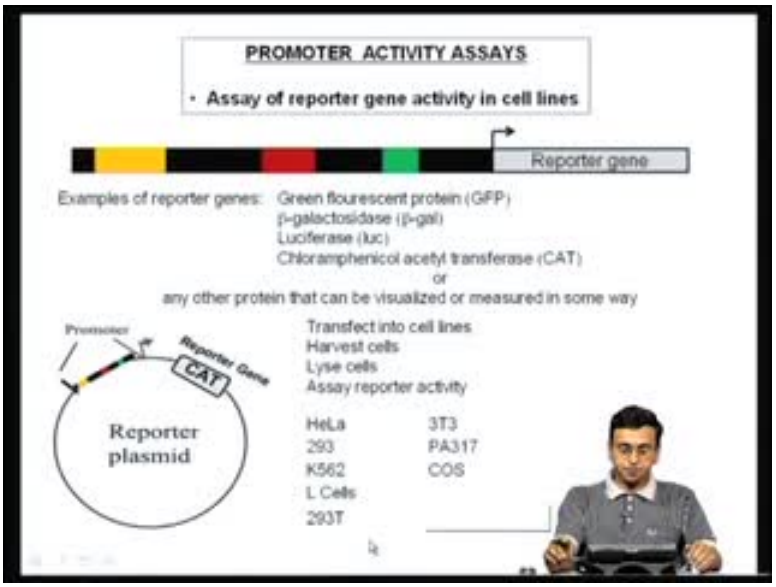
So, you have core transcription or basal transcription contributed general transcription factor and RNA polymerase II and you have further upstream elements to which the other transcription activators may be binding and this is often referred to as the activator dependent transcription.

Now, we are now going to spend may be about 20 or 30 minutes because so far, I have been talking mostly about theoretical aspects. We also It is very important for us also, to understand how do people study the promoter activity, how to assay, what kind of experiments people normally do to assay promoters, what kind of experiments we actually do.

So, what you are going to study in the next few minutes is to understand how does one assay the promoter activity. You know, how do you show that when you have this proper region, the promoter is highly active and when you start making deletions, the promoter activity goes down?

What kind of promoter activity assays people use to study actually the regulation of gene expression and how does actually one identify the upstream regulatory elements. These are the two important techniques which have made a very important difference in eukaryotic gene regulation. Therefore, we will spend some time to understand these experimental techniques.

(Refer Slide Time: 07:16)



Now, the promoter activity - how do you assay the promoter activity? One of the very important experimental systems that people have developed to assay promoter activity is by what are called as reporter gene assay in cell lines. Now, as you know, there are number of eukaryotic cell lines, which have become available like hela cells, which are derived actually from a mammary tumor cell line and you have liver cell lines, you have fiber blog cell line and so on, so forth.

So, once you have this cell lines, these cell lines can be indefinitely cultured in the laboratory in vitro, in a carbon dioxide incubator. So, you can actually use this cell lines and you want to study a particular eukaryotic promoter, what you do is for example, in the previous promoter, we know that there are probably three promoter regions which have been identified and how does how do people identify these three promoter regions?

What you actually do is you take this entire promoter region and downstream of the transcription (()) , you link them to a reporter gene. This is using recombinant DNA technologies. You can actually have restriction sites; by using restriction, you can actually put different reporter genes. Now, what is a reporter gene? A reporter gene is one which codes for a protein or an enzyme, whose activity can be easily assayed or whose presence can be easily assayed, if it is a protein or if it is an enzyme whose activity can be easily measured.

So, it becomes very easy to assay the promoter activity. The most common reporter genes which are used in eukaryotic gene expression studies is the one that codes for a what is called as a green fluorescent protein. You all know the people who have actually discovered this green fluorescent protein, they actually won noble prize couple of years ago and then there is something called beta galactosidase, which when you add a substrate, it actually when the enzyme cleaves and you get a blue color; then you have what is called as luciferase, which actually can convert luciferin in the presence of ATP and (( )) and they emit light.

So, the light can be very easily measured or you have reporter genes like chloramphenicol acetyl transferase, which converts chloramphenicol into acetyl chloramphenicol and these non-acetyl, acetylated forms of chloramphenicol can be easily separated on a thin layer chromatography and the number acetyl the percent of acetylated form that is formed is actually a measure of the promoter activity. So, these are the most commonly used reporter genes in eukaryotic gene expression studies.

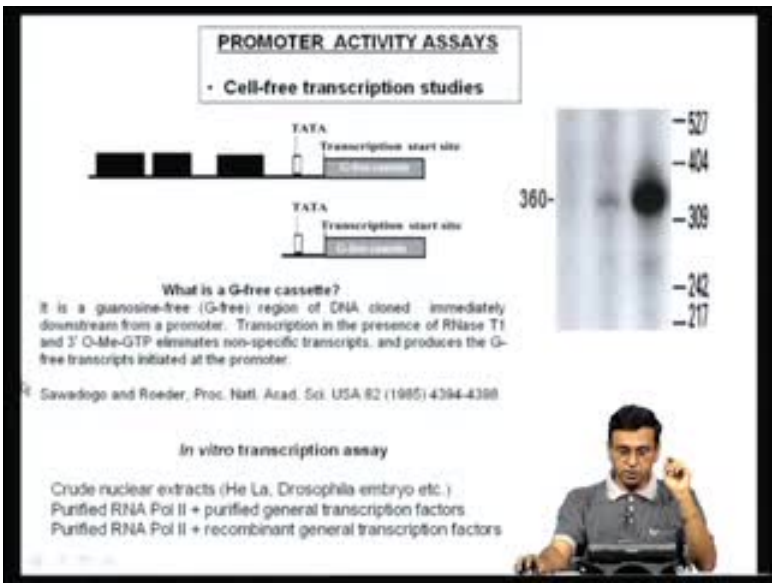
So, when you want to study a particular promoter, for example, in this particular case, let us say, you are interested in here, you just clone the reporter gene downstream of this promoter and this entire thing is actually put in a plasmid and this is called as a reporter plasmid, where you have a promoter of your interest and you have a downstream reporter gene and you then transfer this into various cell lines. And there are various methods for transfecting gene plasmids or genes into cell lines.

You can do what is called as a calcium phosphate precipitation; you can do what is called as a lipofection and so on; there are a number of techniques, but it is sufficient to know that you can actually introduce a exogenous gene into eukaryotic mammalian cell lines and culture and this plasmid goes inside and the various protein factors present inside a nuclei of these cells, actually will now go and bind to this promoter, activate reporter gene expression and then you actually harvest this cell, lyse the cell and assay for reporter activity.

So, this is how in the previous slide, I showed you that when you start cloning different deletion regions of the promoter as a downstream of reporter gene, the reporter gene activity started varying and this kinds of studies actually told in that case that there are actually three promoter regions, which when you make deletion, the reporter gene

expression actually goes down. So, the assaying the promoters, using reporter genes in various cell lines made a very important contribution for understanding the promoters of a number of eukaryotic genes.

(Refer Slide Time: 10:45)



The other important promoter activity study that actually made a very important contribution for understanding eukaryotic gene regulation is using, what is called as cell free transcription studies. So, here what you do is suppose, you want to study a gene which is present in a liver cell line or a say hela cells.

So, you actually make whole cell extracts or nuclear extracts from this cells and this preparation of nuclear extracts which are transcriptionally competent, requires a lot of require a lot of effort and lot of studies have gone into it, just like (( )), there is a very standard procedure for actually making nuclear extracts, which are capable of transcribing eukaryotic promoters in vitro. So, making transcription competent nuclear extracts or whole cell extract from mammalian cell lines has made a very important contribution for understanding eukaryotic gene expression.

So, once you have what is called as a cell free extract or a soluble extract, which contains all the factors including RNA polymerase II, general transcription factors and your other upstream activators in that particular cell line, then what you do, you put what is called as a DNA template for in vitro transcription and this DNA template, actually, is a lot of

people have been using lot kind all kinds of templates, but one very interesting and very innovative template which was designed by Sawadogo and Roeder is what is called as a G-free cassette.

Now, what you actually do for example, if you are interested in studying this particular promoter, which contains a TATA box and transcription starts site here, a cell-free transcription system or an in vitro transcription system, what you do is, just like we cloned a reporter gene downstream of the transcription site for the transcription assays, in this particular case, you actually clone what is called as a G-free cassette downstream of this promoter region. Now, what is a G-free cassette? A G-free cassette is a guanosine free region or a g-free region cloned immediately downstream of the promoter. That means, this template actually contains only C, A and T residues; it has no G residue.

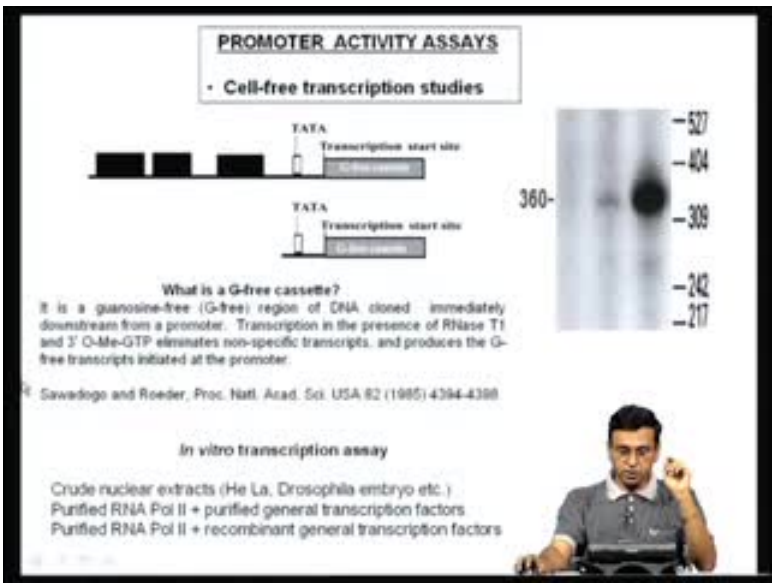
So, this template can actually transcribe in the absence of GTP because GTP will not be incorporated in the RNA; therefore, G-free template. So, when you actually carry out the transcription in presence of RNase T1 and O-methyl GTP, the transcription which have initiated from places other than transcription start site, will be aborted and only those transcripts, which are originating from the transcription start site will be visualized and by actually cloning G-free cassettes of different lengths, you can actually get different sizes of RNA because immediately after the g-free region, you again have a normal DNA, which contains G and therefore, the O-methyl GTP which is there in a reaction mixture will get incorporated and transcription will not proceed further.

So, only your G-free template will be transcribed and depending upon the size of your G-free cassette, you will get that particular size RNA because you are actually, in all this assays, use Alpha P32 UTP, radio label UTP as one of the nucleotides and therefore, the RNA gets radio labeled and you can actually run this labeled RNA, after the reaction is over and then analyze them on a acrylamide gel and depending upon the size of the G-free cassette, you can get a very specific size product.

For example in this case, the G-free cassettes of 360 base percent lengths and as you can see, if you can only actually use a promoter which contains only the core promoter, you get very low levels of transcript, but the moment you use a template, which also contains upstream region, you can see there is a many fold increase in the level of RNA. That is actually made in the cell-free system clearly indicating that this region of the promoter

actually harbours very important cis-acting elements to which certain transcription factors are going and binding and as a result of it, the transcription is enhanced several fold.

(Refer Slide Time: 10:45)



So, the difference between this and this, this and this is actually contributed by the transcriptional activators. So, this is basal transcription; this is the activator depended transcription that you are measuring. So, by using transfection assays in vivo in cell lines or by using cell-free transcription studies, one can actually assay the activity of various promoters.

A number of such cell-free transcription systems have been developed over the years. Some of the most popular examples are for example, you can actually make nuclear extract from hela cells, as I told you, such cells are derived from a mammary tumor; you can actually make such kind of a transcription competent cell free extracts from Drosophila embryos, you can actually make from liver, rat liver and so on and so forth.

So, in all these cases, you actually can actually either make a crude nuclear extract which contains everything, and but as research proceeded, especially in the late eighties and nineties and once people started purifying various general transcription factor, instead of using crude nuclear extract, people actually started using what is called as a in vitro reconstituted system, wherein you purified you have purified RNA polymerase and its

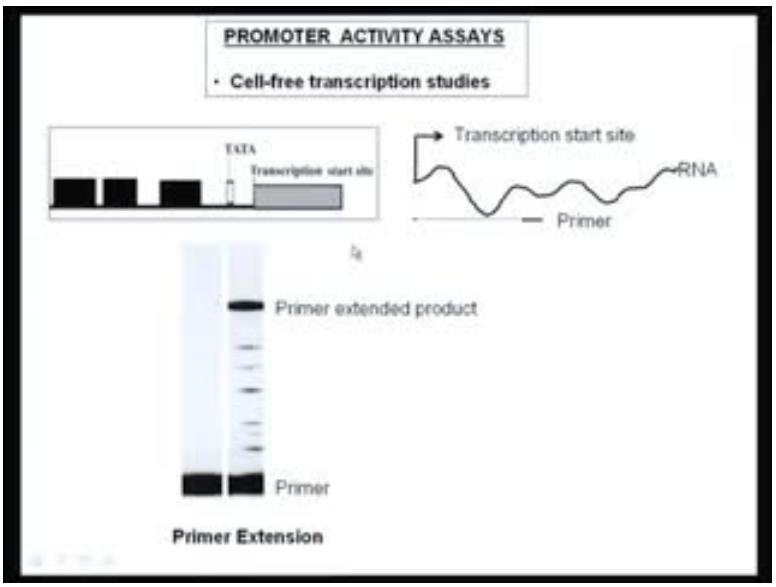
sub unit and you actually have purified general transcription factors. You add this very specific components and this is actually called as a recombinant or reconstituted in vitro transcription system, where instead of taking this crude nuclear extracts, you have purified components, which contains RNA polymerase II and general transcription factor. You can put them together and then you add your DNA template, nucleotides and you can actually measure the rate of transcription.

And as people started cloning the various genes for the general transcription factor, they have also been expressed they have also been able to express this general transcription factors in E.coli systems and people started using recombinant general transcription factor in this in these cell-free transcription systems.

So, you have RNA polymerase II purified from a particular source and then you add recombinant TF2D, recombinant TF2B and so on and so forth.

So, now we have a reconstituted system where you know exactly what you are adding. So, from crude nuclear extracts, purified components and recombinant general transcription factors, you know how they evolve using highly specified in vitro transcription systems. Now, in addition to this G-free cassette which is one of the very popular methods of measuring in vitro transcription assays, you can also do what is called as a primary extension.

(Refer Slide Time: 16:46)



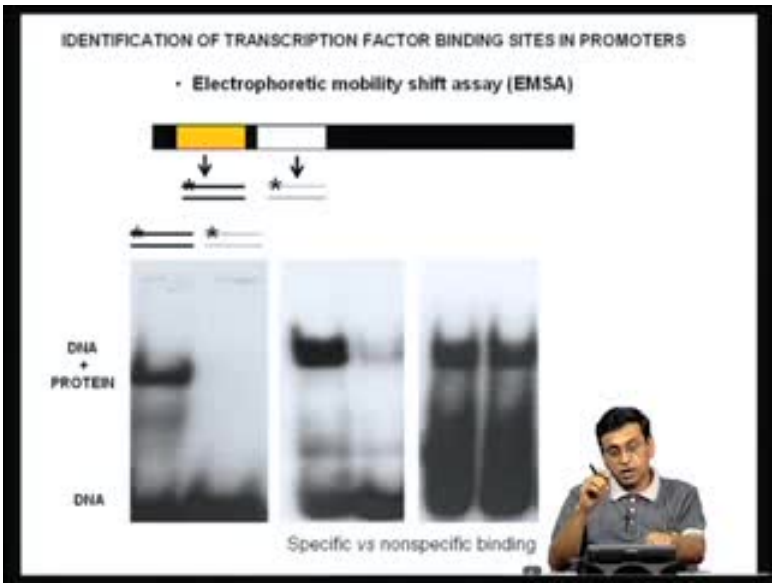
Where So, the template here is slightly different from what I showed in the previous slide. Here, you do not have a G-free cassette, but you have your normal gene sequence which is present and which is the coding region of the normal gene and you do the transcription and then once you have done the transcription, you do what is a called as a primer extension.

So, here the RNA is not labeled. You actually carry out a regular non-label RNA. You take this RNA and then once the RNA is synthesized in the cell-free transcription system, you hybridize to a small primer and then extend the primer using a reverse transcription.

So, suppose, the transcription starts here, the primer will be extend up to the transcription start site and then you can this will be radio labeled and this cannot be analyzed on a gel and you can see, when you have no template, only the radio label nucleotide only the other components of the thing, you do not get a primer extended product. But when you put the template and when you have all the components for transcription, this template, the transcription start and RNA will be made and you can get a specific extended product, which is called as the primary extended product.

So, either using a G-free transcription template or using what are called as primer extension, you can actually quantify the RNA that is synthesized in a cell-free transcription system and by using promoters [whether they] contain mutations or deletions, you can assay all these promoters in these in vitro transcription systems and you can actually demonstrate what kind of regions in the promoter are actually essential for activator depended transcription. So, this is how people have actually identified various promoter regions, which are essential for transcription activation of genes in eukaryotic cells.

(Refer Slide Time: 18:06)



Now, once you know that this particular promoter region is essential or plays a very important role in the activation of transcription, the next question comes how do you identify what kind of transcription factors are actually going and binding to these promoter regions.

As you know, if a particular promoter region, when you delete and thereby there is a decrease in promoter activity, it actually means that there are certain protein factors which are transcription factors, which are actually going and binding to those regions and that is how this promoter region became very important. So, if you delete that promoter region, these transcription factors can no longer bind to those regions and therefore, there is a decrease in promoter activity.

So, once you identify these kinds of promoter region which are important for the promoter activity, the next question is how do you actually identify, what kind of protein factors are actually going and binding to this promoter region.

No there are many, many assays many, many experimental techniques by which you can actually identify these transcription factors, binding sites in the promoter as well as the transcription factors themselves.

So, let us now study one or two techniques, which people routinely use to identify transcription factor binding sites in the promoter regions. One of the very, very popular

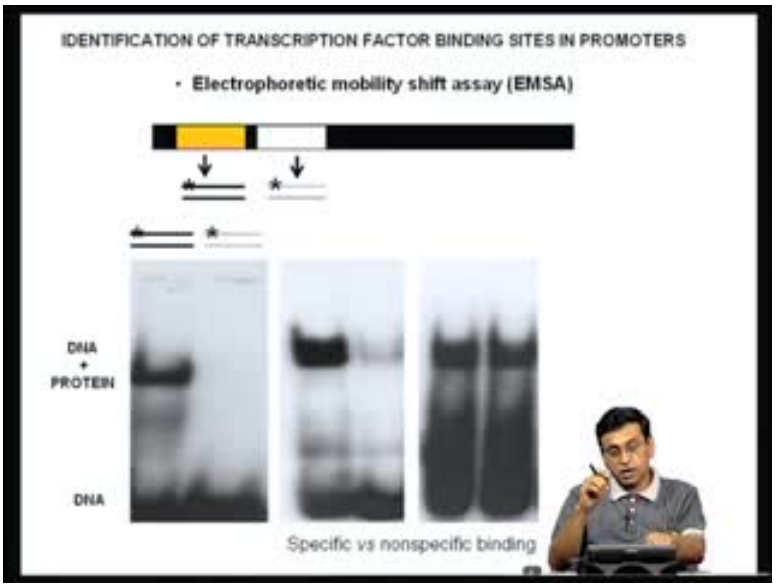
assays that people use for identifying the various promoter regions, which are essential for binding protein factor is called as electrophoretic mobility shift assay, is very popular known as EMSA. E M S A - electrophoretic mobility shift assay.

Now, let us understand what this EMSA actually is. Now, for example, I told you in my previous study, I had a promoter and when I deleted this particular region of the promoter, there was a decrease in the promoter activity. Now, I want to really see does this promoter region really bind to any transcription factor and it is because of this transcription factor binding that this promoter region actually contributed to promoter activity. How do I go about? What you do is that let us say this is about say 50 or 60 base per region which when I delete it the promoter activity drop down.

So, what I do is I make a double stranded oligonucleotide; these days we can actually make oligonucleotides of varying length. You can in fact synthesize genes these days without any problem. People have been making synthetic genomes these days. So making small DNA fragments is not a big deal. Therefore, you actually make a double stranded DNA oligonucleotide and then you radioactively label this oligonucleotide.

So you have a region, which is actually essential for promoter activity in vivo and you take this DNA fragment and as a control, you also take an adjacent promoter DNA fragment which when I deleted, it had no effect on the promoter activity. So, I have an oligonucleotide, which in the promoter context has an important role in the promoter function. I have an adjacent region which when deleted nothing happened to promoter activity. So, this serves as a control.

(Refer Slide Time: 18:06)



Now, what I do I take a cell extract or I take a tissue extract in which this gene is normally active and ask the question, what kind of protein factors within this tissue extract or cell extract is actually binding to this region and as a result, there is an increase in the promoter activity.

So, what I do is that I take this radio label DNA fragment and then incubate it with a nuclear extract or a whole cell extract of my interest. From whatever tissue I am interested, I make this whole cell extract and once I incubate this radio label DNA with this nuclear extract, in presence of what are called as certain amounts of non-specific DNA, therefore, the non-specific interactions are brought about or knocked off and then you actually do what is called as an electrophoresis; in a polyacrylamide gel, you do an electrophoresis.

When you do an electrophoresis what happens, when we have when the DNA is not bound to any particular protein, it will move faster and therefore, it will have faster mobility. But if this DNA fragment is interacting with a specific protein inside this nuclear extracts and is forming a complex and because of the formation of the DNA protein complex, the mobility of this DNA gets shifted; that is what is called as an electrophoretic mobility shift assay.

So, the mobility of the DNA gets shifted and therefore, you have a retarded complex and therefore, this is even called as gel retardation assay, where the mobility of the DNA gets retarded because of the formation of DNA protein complex. So, in this particular case, this particular DNA fragment was able to form a DNA protein complex and therefore, the mobility of this DNA got retarded here or shifted here, whereas this particular promoter region, which did not have any role in the promoter function, did not bind to any particular protein in the particular cell extract and therefore, did not form any complex and it was moving much faster.

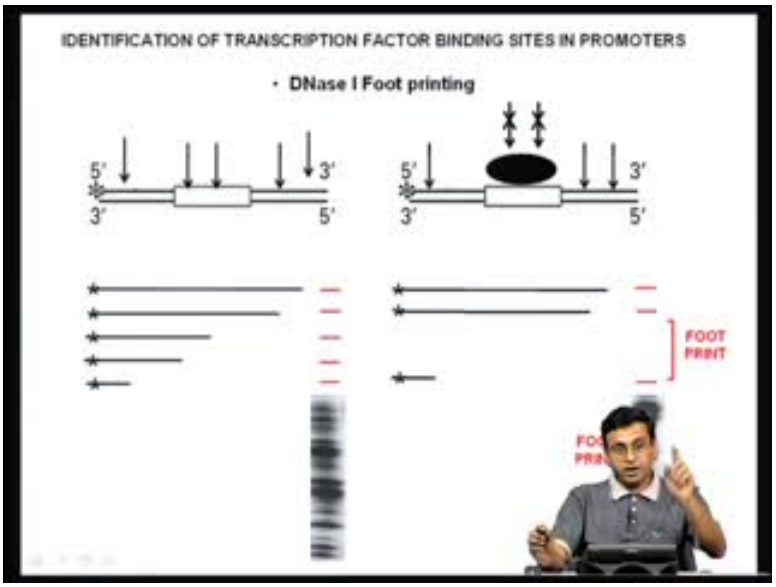
So, you can see by using these kind of electrophoretic mobility shift assays, you can actually identify what kind of protein factors are actually present in this region, in this particular cell extracts and are they really binding to this promoter region, and this assay is very popular and very easy to use because you can also distinguish between specific versus non-specific binding.

In this case for example, I have radio labeled both the specific DNA as well as non-specific DNA and I am showing that only the specific DNA is binding. There are other variations by which you can actually do this. You can also do for example, you can once you get this radio labeled DNA protein complex, you now add excess amounts 10 fold 20 fold 50 fold or 100 fold excess of the unlabeled DNA, of the of the same unlabeled DNA and you can see now the complex has completely chased, whereas when you have the same unlabeled excess competitive of the other non-specific DNA or DNA which does not bind, it does not chase.

So, clearly this is also called as competition experiments, which again tells you that you are actually looking at a very specific DNA protein interaction. Only a homologous DNA fragment is able to chase the binding of the protein to radio label DNA, whereas a heterologous DNA fragment cannot chase or cannot compete for the binding of the protein in the nuclear extract to the specific DNA.

So, by using these kinds of competition experiments, you can actually convince yourself that what you are actually measuring is a very specific DNA protein interaction. The protein that is actually binding to this DNA fragment is very specific, but it does not bind anywhere else.

(Refer Slide Time: 24:15)



Now, there is another very, very popular experimental technique that people have been using and this is actually called as a DNase I foot printing. So, in addition to the EMSA, this DNase I foot printing also is a very popular and very important technique that people have been using to identify transcription factor binding sites within the eukaryotic promoters.

Now, let us see how exactly it is done. Now, let us say for example, I have a promoter region and I suspect there is a protein actually that is binding to this particular region within this promoter. Now, when I take this promoter DNA and then radio label it at one end of this DNA. Remember this is very important; when you do DNase one foot printing you have to radio label the DNA either at the 5 prime end or at this end; you cannot label both ends; you can only label one end of the DNA is radioactively labeled.

Then you start treating (( )) with very limited amounts of DNase I. As you know DNase I introduces double stranded DNA nicks; when you have manganese and magnesium as cations in the buffer, it will introduce double stranded DNase and it randomly cleave the DNA. In this particular case, for example, these are all DNA for example, the first instance of DNase I will make a nick here and as a result, you will get a radio labeled DNA fragment of this size; when it makes a nick here, you will get a radio labeled fragment of this size and when it makes a nick here, you will get a fragment of this size and so on and so forth.

So, depending upon where the DNase I is making a nick, you will get different sizes of the radio labeled fragment, but here the DNase I does not distinguish; this is a very random process. So, it has equal property of cutting the DNA at any particular region within the DNA. So, almost all the regions of the DNA are equally susceptible for DNase I attack.

Now, let us see what happens? But suppose now, you take this radio labeled DNA and then incubate it with the protein of your interest like what we did in the previous experiment. Suppose, now you incubated the nuclear extract so that this particular specific protein present in the DNA now goes and binds to this specific region within the promoter region. So, you know how a DNA protein complex is formed here.

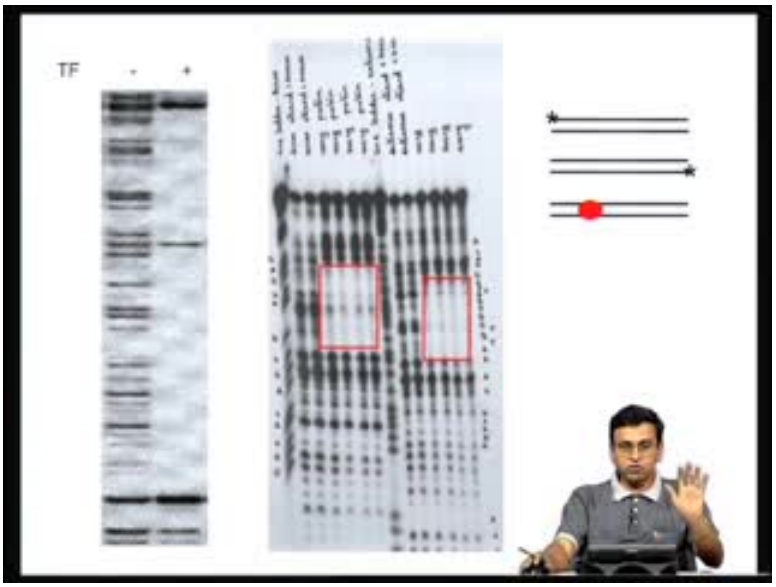
Now, when you take this DNA template and then treat it with DNase, as you can see here, the DNase cannot, while it has cut here, it cannot cut this region because there is a transcription factor binding here and because of steric hindrance, it will not allow DNase one to act there.

So, therefore, the DNase only can cleave here and here, but not in this region. Therefore, it cannot generate these two radio labeled fragments here. So, when you take this mixture of fragments and this mixture of fragments and analyze them on a polyacrylamide gel, on a denaturing polyacrylamide gel, while you get uniformly radio labeled DNA fragments of all size, when there is no protein binding in this particular case, you will not see DNA fragments corresponding to this region.

So, this region actually now forms what is called as a foot print. So, the region, where exactly the transcription factor bound actually leaves a foot print and this is what is called as a DNase one foot printing studies.

This is how an actual gel looks. From our own lab, we have done some of this foot printing studies and when you do this label this promoter and do this kind of DNase I foot printing in the absence of the binding protein, you see you have radio label fragments generated all through, but when you now incubated the same DNase with a specific protein, which, you know binding a specific region and you can see this particular region is not susceptible DNase attack and therefore, I conclude this is the region, where my transcription factor is actually binding.

(Refer Slide Time: 27:36)



These are the actual gels; this is from literature that people have actually done. You can see in the absence of protein, you have radio label fragments of all sizes when you have transcription factor and when it is binding in this region, DNase I cannot choose this region.

Therefore, you do not see band in this region. In our own lab, actually, we have done some of these DNase I foot printing studies. I will discuss these experiments in much more detail in the next class, when I talk about transcription activators and so on and so forth.

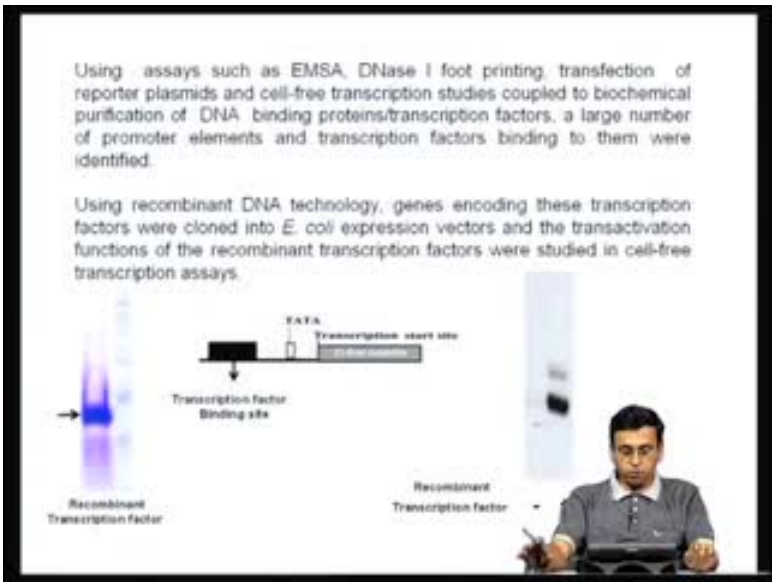
But it is suffice for you know, here for example, in this case, we have taken a promoter region of a gene called alcohol oxidase in a yeast called pichia pastoris and we are interested in finding out what kind of transcription factors are actually binding to the promoter region of this particular gene.

So, we radioactive label this promoter region and did this DNase I foot printing study and you can see here. Now, I have got a foot print here. So, the first 2 the these 2 lines are without protein and these 4 lines are with protein; where all other regions are getting cleaved, you can see this region is not susceptible to DNase I. So, I conclude, this is where the protein is binding. Now, I can also label the other end of the DNA and repeat

the same thing. So, I will get the protein binding region both in the top strand as well as the bottom strand.

So, this is the foot print in the top strand of the DNA; this is the foot print in the bottom strand of DNA; by putting this together, I can actually tell you this is exactly the region where the protein is actually binding. The protein is making contact both in the top strand as well as the bottom strand and therefore, it is binding like that. So, this is how one can precisely map transcription factor binding sites in the promoter region using DNase I foot printing studies.

(Refer Slide Time: 29:07)



So, using assay such as EMSA, DNase one foot printing as well as transfection of reporter plasmids into various cell lines, making promoter deletion studies as well as cell-free transcription studies, either using crude nuclear extracts or using purified components of general transcription factor, people have been able to and using these assays, people have been able to purify a number of transcription factors from various eukaryotic cells and they have also been able to identify the various promoter elements to which these transcription factors are going to bind.

So, once you have purified these proteins by biochemical means, then you can actually use these purified proteins and then you can, by raising antibodies against it and by using routine recombinant DNA and genetic engineering techniques, you can actually clone the

genes. So, over about between actually 1980s and 2000, almost 2 decades, a number of transcription factor genes were actually cloned. So, genes encoding various transcription factors were actually cloned

So, we have actually the genes, which are actually coding for this transcription factors and we have also gained knowledge where exactly in a promoter region of a particular gene, this protein factors are actually binding. So, once you have the gene coding for this eukaryotic transcription factors, what you now do is using recombinant DNA technology, you can actually express these recombinant proteins in a E coli system.

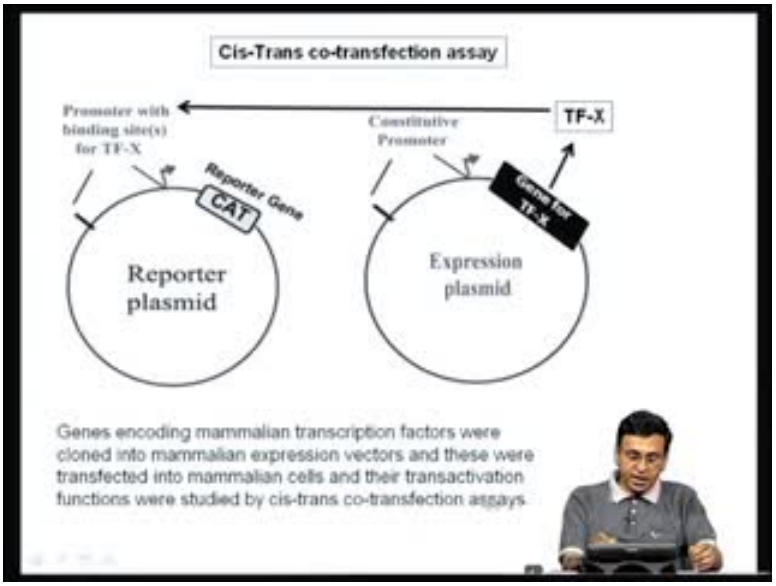
So, you have a eukaryotic transcription factor, which has been made produced in large amounts in E coli cells. Here, I have shown for example, we have actually over expressed one of the transcription factors which we are interested. We actually clone the gene coding for this transcription factor; we have over expressed this protein E coli cells; we have actually purified this recombinant protein from E cells and this is what is called as a coomassie blue stain it has s d s polyacrylamide gel electrophoresis; this is with the purified transcription factor

And now actually, I can take this transcription factor and put them in my in vitro transcription system or a cell-free transcription system and actually show that when I have a transcription when in the absence of the transcription factor, I do not see very high levels of transcription, but the moment I put the transcription factor that goes and binds to this transcription binding site in the promoter region and therefore, I get very high levels of the transcription. clearly indicating that so this is

So, I first took the promoter region, made some deletion on mutations and identified what are the regions essential for promoter activity, then by using EMSA, DNase I foot printing as an assay system, I went ahead and purified transcription factor, which actually bind into this region. Then I have actually cloned the actual gene that codes for the transcription factor and then I have taken the gene, put the gene in E coli systems, make the recombinant transcription factor, now, I put this recombinant transcription factor back into my cell-free system and convincingly, show yes, this recombinant transcription factor can in fact activate transcription from that particular promoter region.

So, the circle is complete. So, this is how people have functionally demonstrated that the genes that they have cloned or they have identified actually encode transcription factors and functionally, they can demonstrate that these genes can actually function as the recombinant proteins can function as transcription factors in in vitro systems.

(Refer Slide Time: 32:19)



You can also assay the transcription factor function of these cloned genes in a in vivo system and a very popular assay which has again made a very important contribution to understanding gene regulation is called as cis-trans co-transfection assay. What did you see in the previous experiment, I took the recombinant protein made in E coli cells, put it in a cell-free system and showed that this transcription factor can activate transcription from their particular promoter region. But in this case But I can also argue that you have only shown in a vitro system, you have actually not showed in a in vivo these gene actually codes for a transcription factor.

Now, further what do I do is I do I make two plasmids: one is called as a reporter plasmid; another is called as an expression plasmid. So, I take the gene which is actually coding for a particular transcription factor and clone it in front of what is called as a constitutive promoter; that is, this is a promoter which is expressed all the time. Remember in our eukaryotic cells, we have promoters which are constitutively expressed and there are promoters which are inducible because we do not have to make all the proteins all the time.

We only make proteins all the time, probably those involved in glycolysis, pepcycle, fatty acid metabolism, protein metabolism, these are all every cell in the body requires these enzymes and proteins and therefore, many of these gene coding for these enzymes and proteins will be constitutively expressed. So these are called as housekeeping genes. despite Irrespective of what cell type you are dealing with, almost all the cells actually express these set of enzymes. So, these are called housekeeping functions.

So, you need these proteins and so, all the cells will make these proteins, but there are many proteins which are actually expressed in specific cell types under specific periods; that is, these genes are expressed only when there is a need for them. So, these are called as inducible genes. So, by studying a number of eukaryotic promoters, eukaryotic genes over the years, people have identified a number of inducible promoters, number of constitutive promoters.

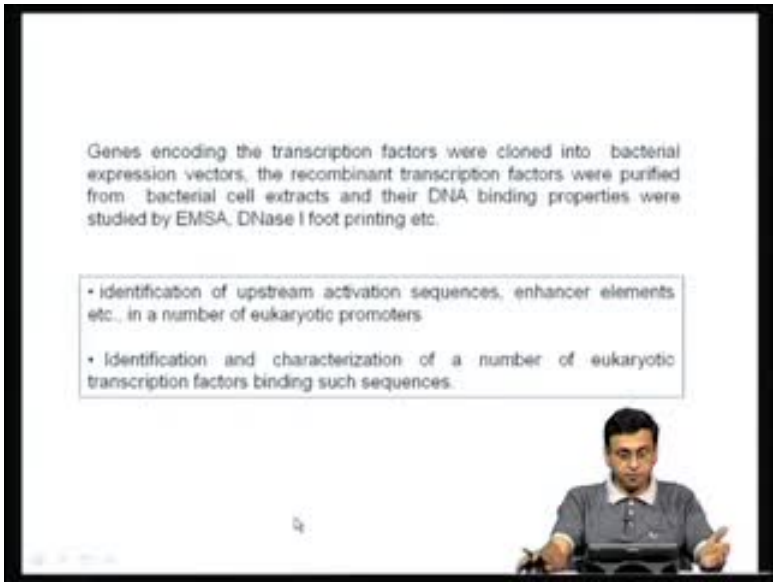
So, you take the gene coding for your particular transcription factor; you clone it in front of a constitutive promoter and you also take the promoter region, which you suspect is the binding site for this particular transcription factor and clone that in front of a reporter gene, which we have discussed earlier. Now, you mix both this plasmid and transfer them into eukaryotic cells of your choice. So, that is why it is called this is called as cis-plasmid; this is called as trans-plasmid. That is why this assay is called cis-trans co-transfection assay.

So, two plasmids are transfected simultaneously. Now, what happens, this transcription factor gene will now be expressed because it has a constitutive promoter and the transcription factor, which is made from this particular plasmid will now go and bind to this particular binding sight present in the promoter region and then as a result, it will activate this promoter, the reporter gene will be expressed and now you can easily assay the reporter gene activity.

So, by making recombinant transcription factors and examining their transcription activation potential in cell-free system as soon as and by doing these kinds of study in vivo, you can actually convincingly demonstrate, yes, what you really have is a gene that codes for a transcription factor, which can activate transcription either in a cell-free system or in a in vivo system using cell lines.

So, genes encoding mammalian transcription factors were cloned into mammalian expression factors and these were transfected into mammalian cells and their trans activation functions were studied by using these kinds of a cis-trans co transfection assays.

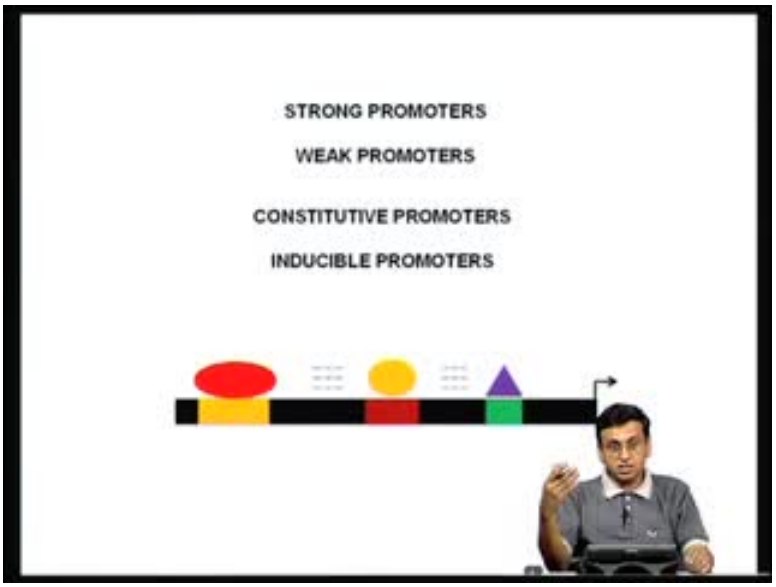
(Refer Slide Time: 35:30)



So, genes encoding transcription factors - that is what I mentioned here, all list of assays were done and as a result of all these experiments, a number of upstream activation sequence, enhancer elements have been identified in a number of eukaryotic promoters.

So, all that you have to do suppose, I want to study the particular promoter, I can make this deletions, put it in front of reporter gene, test them in cell lines or I can also make these, put it in front of the g-free cassette and then assay them in a cell-free system and then I can also use techniques like EMSA, DNase foot printing and then really see what kind of protein factors are actually binding into these promoter elements and by doing this combination of studies, you can actually identify exactly what kind of promoter regions are actually acting as enhancer sequencers and what kind of transcription factors are actually going and binding to these sequences.

(Refer Slide Time: 36:17)



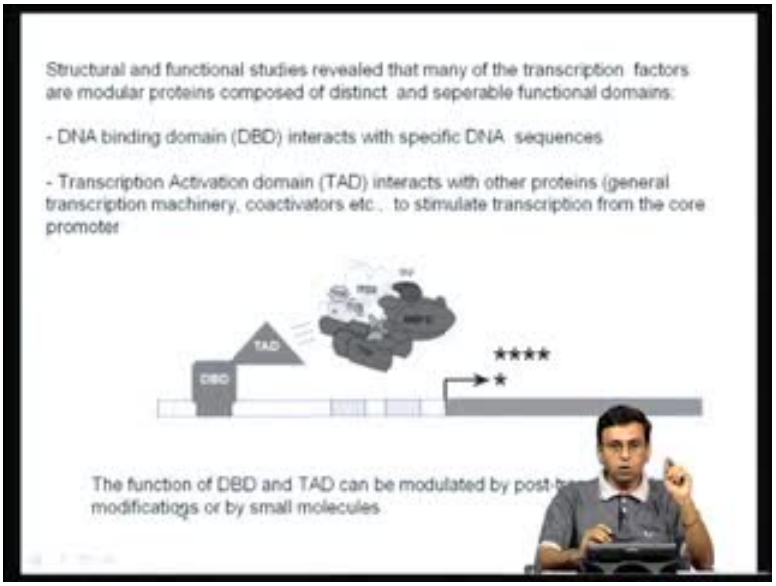
So, as a result today, we are actually flooded with hundreds and hundreds of eukaryotic promoters and eukaryotic gene sequences and you can actually classify many of these promoters as strong promoters, weak promoters, constitutive promoters, inducible promoters and so on, which I explain here. When you say a strong promoter, what it actually means is that probably it has more than, it has multiple enhancer elements or multiple upstream elements to which a number of transcription factors are binding and as a result, they can synergistically interact with the general transcription machinery and therefore, can recruit RNA polymerase II much more efficiently promoter and therefore, the rate of initiation is very high. It can initiate multiple rounds of transcription activation initiation in a given time. So, it is a strong promoter.

In a weak promoter, probably a weak transcription factor is binding to it or probably not many transcription factor binding sites are present at the promoter and as a result, they cannot efficiently recruit the general transcription machinery to the transcription initiation site and therefore, these promoters are transcribed rather weakly.

And then I mentioned in the just previously that there are also what is called constitutive promoters and inducible promoters. In the case of inducible promoters, the transcription factors which go and bind to this enhancer elements, they bind only under certain conditions which you have to induce.

We will discuss little bit more in detail in the next few minutes. So, it became very clear, transcription activation in mammalian cells is a very, very complicated process. So, it not only involves a core promoter region to which the general transcription factors and RNA polymerase two binds and then constitute what is called as the basal transcription, but there are also many important factors which are binding to many upstream elements in the promoter region and it is the interactions between these transcription factors amongst the general transcription machinery, these synergistic interaction, ultimately contribute to the increase in levels of transcription of eukaryotic protein coding genes.

(Refer Slide Time: 38:18)



So, this structural and functional Now, we will go now one step further and see, So, you had purify transcription factor from various sources. They have actually cloned the genes by over expressive proteins and once they demonstrated that they have indeed a cloned the gene coding for a transcription factor or they have lots large amounts of these eukaryotic transcription factors in their hands, people started looking at how do this transcription factor go and bind DNA and how do they activate transcription, either in a cell-free system or in a cell line transfection system.

So, that people by these kinds of studies, it became very clear that the transcription factors are actually modular proteins and they actually are composed of distinct and separable functional domains. Now, the two important domains actually emerge. Now, remember, when you say, you have a transcription factor, it is a it has to go and bind to a

specific sequence of the promoter and once it binds, it had to interact with the general transcription machinery and active transcription.

So, most of the transcription factors which people studied, they at least contain two major domains. One is called as DNA binding domain for interaction specific DNA sequences and it also contains what is the transcription activation domain, which will now, once it binds DNA through this transcription activation domain, it interacts with general transcription factor and enhance the rate of transcription initiation.

So, almost all the eukaryotic transcription activators contain two major function domains: a specific DNA binding domain and a specific transcription activation domain. So, these are the two minimal domains or modules that eukaryotic transcription factor must have and we will study this little bit later, but suffice for you to know that the function of both the DNA binding domain and the transcription activation domain of eukaryotic transcription factors can also be modulated by post translational modifications and the next few classes, I will elaborate on many of these examples.

Post translational modification of transcription factors is one of the major mechanism by which the transcription factor function is modulated in eukaryotic cells. For example, there may be a transcription factor present in the cell, but not only it may be present, it may be even binding with DNA, but it will not be able to activate transcription.

But if there is kinase and this kinase goes and phosphorylates the transcription activation domain, now the transcription factor becomes transcriptionally active and then it can interact with general transcription machinery and activate transcription. So, in this case, phosphorylation of a specific amino acid residue in the transcription activation domain of the particular transcription factor is the mechanism by which the activity of the particular gene is modulated. So, phosphorylation, which is a post translational modification is a major regulatory mechanism this particular example.

Then we have also examples, where the transcription factor actually may be present in the nucleus, in the cytoplasm, in the conditions when the gene is not active and when the gene is to be activated, a specific stimulus takes this transcription factor and translates the protein to nucleus so that it can actually bind to DNA and then activate transcription of certain genes.

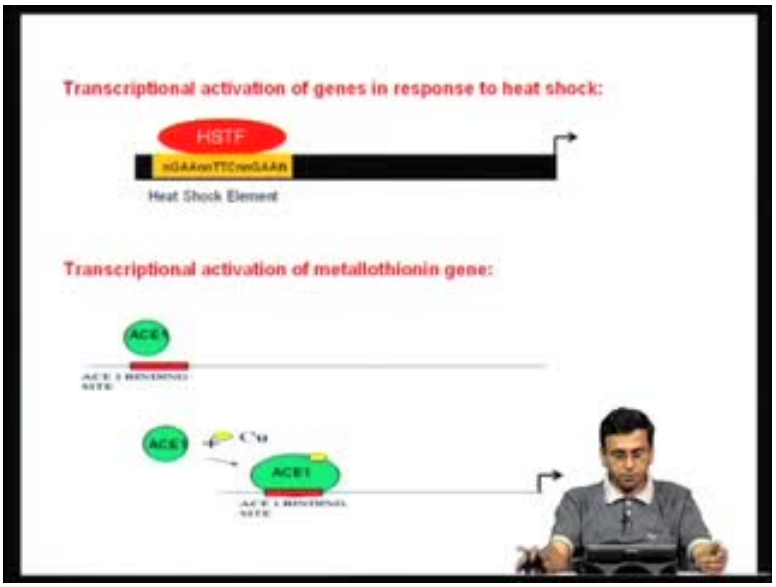
So, there are a number of mechanisms and in many case, post translation modification such as phosphorylation, acetylation, ubiquitylation and so many other post translation modification of specific amino acids of the transcription factor play a very, very important role in regulating the activity of these various transcription factors.

(Refer Slide Time: 41:46)

If you now today want to know what kind of eukaryotic transcription factor are actually present, how many transcription factors have been identified, what kind of sequences they go and bind, you simply go and click on this website and you have a whole bunch of information about the various eukaryotic of transcription factors which have been identified so far and what kind of sequences they actually go and bind.

So, a number of transcription factors have been identified, the genes have been cloned and people have even identified what kind of exact promoter sequences these transcription factors will go and bind. So, this kind of information is now available. All that you have to do is to go to this particular website and there are many such websites; this is just one example, where you can get this kind of information.

(Refer Slide Time: 42:26)



So, what I will do in the next few minutes is to actually give you some key examples just for you to get a flavour. But remember, in the next few classes, we are going to take specific examples of certain genes on certain transcription factors and study in detail. But for all that I have told you so far, for you to get an idea how actually transcription factors were activated, I am going to give you two or three examples. For example, there are genes called as heat shock genes and these genes are active only under conditions of heat shock.

Now, as long as you grow these mammalian cells at 37 degree centigrade, which is the normal temperature - body temperature, these genes are not active; these genes are not transcribed, but when you increase the temperature from 37 degrees to 42 degrees, this is what is called as heat shock, then certain genes called heat shock genes are activated.

Now, people have gone ahead and asked the question how does this heat shock help the activation of transcription of this particular heat shock genes. What people have actually found out in such cases using the various techniques that I have described, people have actually shown that there is an enhancer element in these promoters of this heat shock genes called as heat shock element and all these genes, which are activated by heat shock and if you go to the promoters, they contain this heat shock elements.

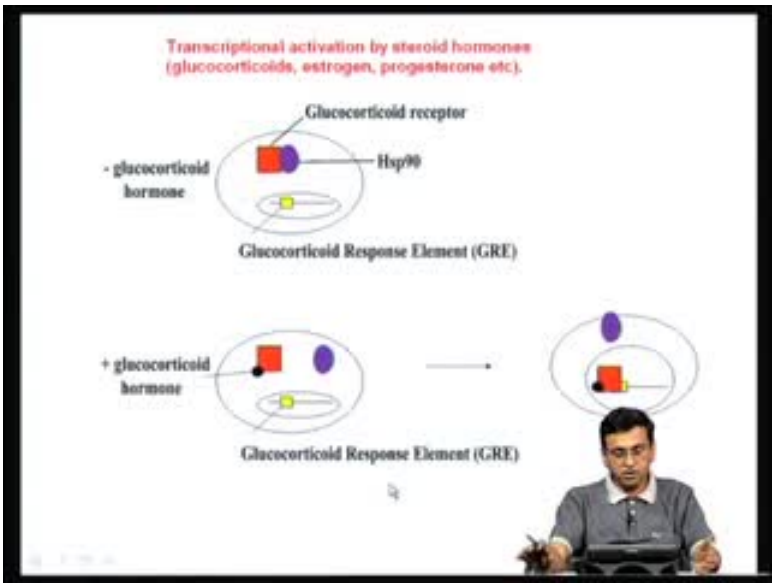
And it turns out, there is a transcription factor called heat shock transcription factor which actually can recognize this particular sequence. So, when you give a heat shock, this transcription factor goes and binds to this particular heat shock element and then enhances the rate of transcription initiation of these genes.

So, all these genes which are induced in response to heat shock, you can see now, there is common mechanism. Whenever there is a heat shock, whenever there is an increase in temperature from 37 to 42, this heat shock transcription factor goes and binds the heat shock elements of these various promoters, of these various heat shock genes and all these genes are getting transcribed.

Similarly, there are what are called as metallothionin genes. Now, this gene need not be expressed in normal times. This gene will be expressed only, when there is a metals metal concentration goes up inside the cells. So it turns out when you now look at the promoter regions of this metallothionin genes, there is what is called as binding site for a transcription factor called ACE 1 in E cells. In this case, we are talking about E cells and it turns out when you have when there is an increase in metal concentration, therefore for example, copper sulphate, when you put copper sulphate or copper chloride in the culture medium of E cells, now this copper now goes and binds this particular transcription factor and as a result, in presence of copper, this ACE1 can bind to this particular sequence in metallothionin promoter and enhance the rate of transcription initiation.

So, metallothionin genes are transcribed only when you have metals like copper in large amounts because now this copper binds to this transcription factor and makes it transcriptionally active. Therefore, it can go and then activate the transcription of these genes.

(Refer Slide Time: 45:16)



The most versatile or the most well studied example of inducible gene expression in higher eukaryotes is what is called as the mechanism by which genes are actually activated by various steroid hormones. You know there are many steroid hormones. Steroid hormones are nothing, but very small hydrophobic molecules. For example, you have glucocorticoids, you have progesterone and you have estrogen, androgens, mineralocorticoids; these are all called as steroid hormones.

Now, all these steroid hormones function, you know the estrogen has a specific function during reproduction, progesterone has a different function, glucocorticoid has a totally different function, but it turns out all these molecules, although they are structurally different and have different physiological function, the mechanism by which they act appears to be similar. It turns out all these molecules, which are all called as collectively called as steroid hormones, when you expose cells to these, steroid hormones being hydrophobic, these steroid hormones traverse the cell membrane and then they enter the cytoplasm and inside the cytoplasm, they had actually receptors which are present. For example, if you take glucocorticoids, there is a specific receptor called glucocorticoid receptor, which is present in the cytoplasm of this eukaryotic cell.

Now, the glucocorticoid hormone, when you add, it goes and binds to this glucocorticoid receptor and when there is no hormone, this glucocorticoid receptor is actually present in the cytoplasm. This is because in the absence of the hormone, this glucocorticoid

receptor interacts with the protein called heat shocks protein 90 - Hsp 90 and as a result of this interaction, this receptor cannot go inside the nucleus. Remember, if you want to activate the transcription of genes, the transcription factor has to go and bind to promoter elements of the genes, which are actually inside the nucleus.

So, when there is no hormone, the glucocorticoid receptor stays in the cytoplasm because it is complex to the protein called Hsp 90 and therefore, it cannot activate the transcription of genes, which are responsible glucocorticoid hormone. Now, the moment, you add glucocorticoid hormone, being hydrophobic, the glucocorticoid hormones simply traverse the cell membrane. Normally, it goes and binds to the glucocorticoid receptor.

Now, when the hormone binds to the receptor, it induces a conformational change in the glucocorticoid receptor and as a result, the heat shock protein can no longer bind to the receptors. So, the heat shock protein dissociates. Now, you have a glucocorticoid receptor which is not in complex with heat shock 90 and it turns out with this kind of conformational change, the hormone binding also exposes what is called as a nuclear localization signal.

Now, for many of the proteins to go inside a nucleus, they have to contain what is called as nuclear localization signals and so, only those proteins which have these nuclear localized signal NLS can actually go into the nucleus. So, as long as the glucocorticoid receptor in the absence of hormones is interacting with the Hsp 90, this nuclear localized signal is buried inside the protein and is therefore, not exposed and that is the reason the protein cannot go inside the nucleus.

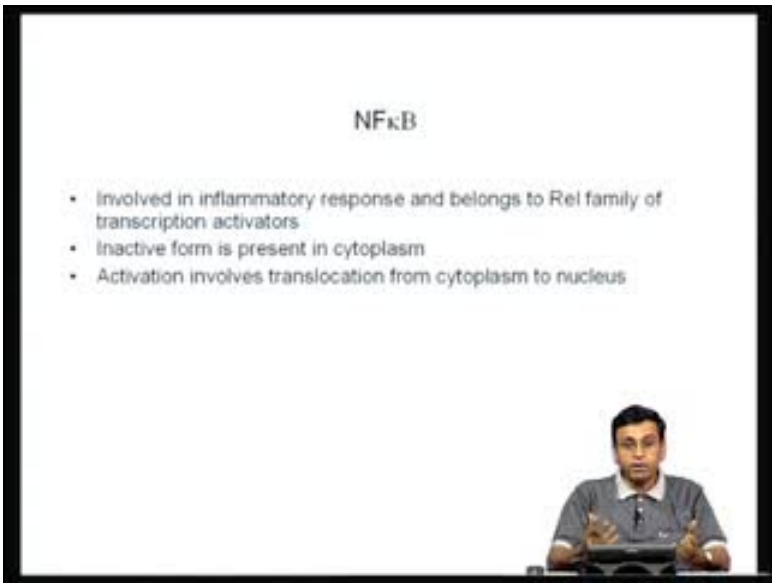
But once the Hsp 90 dissociate, this nuclear localized signal now gets exposed and as a result, the receptor can go now inside the nucleus and bind to what are called as glucocorticoid response elements which are specific binding sites for this particular receptor, which are present in all the genes and which are responsible to glucocorticoids. Therefore, now the receptor binds to these specific elements and then activates the transcription of all these genes.

So, you can see in one case, the heat shock induces the transcription factor and that went and bound to the promoter activated genes. I give another example, where in the

presence of a metal, activity of the transcription factor is modulated, in the presence of metal, the protein is able to bind to DNA and therefore, activate transcription.

Here, I have another small molecule, which actually the regulation is at the level of nucleocytoplasmic transport of the transcription factor. When this small molecule is there, the transcription factor translocates to the cytoplasm of the nucleus, then binds to specific response elements or specific enhancer elements in the promoter regions and activates the transcription of the downstream genes.

(Refer Slide Time: 49:12)



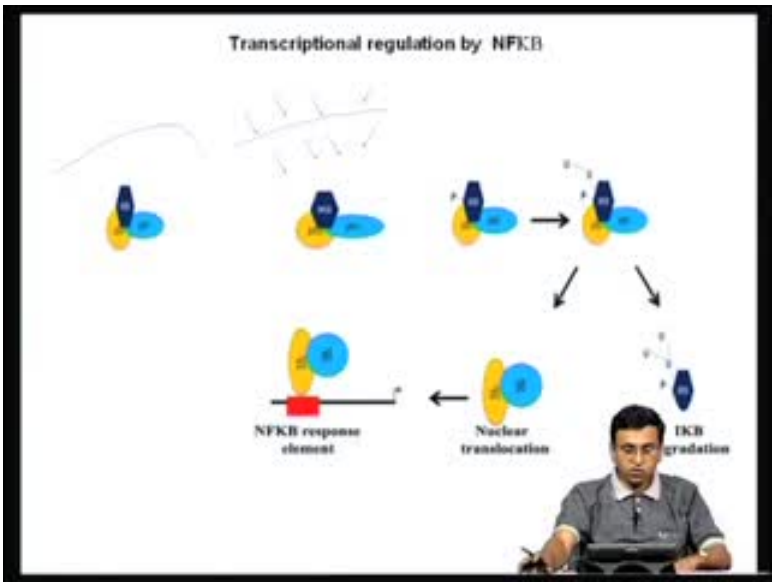
There is another very interesting example and a very, very important transcription factor called nuclear factor kappa B, which is a very important transcription factor involved in inflammatory response and it belongs to a super family of transcription factor, which is called as the Rel family of transcription factors.

Again, I will not go into details at this point. We will discuss each one of this example in greater detail at a later stage during this particular course. Now, again just as we have seen in the case of the glucocorticoid receptor, this NF kappa B is actually present in an inactive form in this cytoplasm.

The NF kappa B is normally present in the inactive form in the cytoplasm of cells, but in response to certain stimulus, for example, there are many situations where there is an

inflammatory response, there is inflammation; in such cases, there are a number of genes which need to be activated in response to this inflammation

(Refer Slide Time: 50:08)



So, how these genes are activated in response to inflammation? So, what happens during this inflammation In response to inflammatory signals, this is what happens. Normally, the protein, the NF kappa B is actually is a heterodimer. Remember, this is another important point to note down. Many of the eukaryotic transcription factors bind to DNA as dimers not as monomers; they actually bind as dimers.

We will discuss in more detail in the next class. In the case of previous case for example, glucocorticoid receptor is called as a homodimer because two monomers of glucocorticoid receptor actually go and bind to the DNA; so, it is called as a homodimer. In the case of NF kappa B, it is an example of a heterodimeric transcription factor, where it has two different subunits: one is called as p 65 and another is called as t 50. So, while glucocorticoid receptor is a homodimer, NF kappa B is a heterodimeric transcription factor.

But I want to give this example because you can see the mechanism of nuclear translational of glucocorticoid receptor is different. There the interaction between the Hsp 90 and glucocorticoid receptor is abolished, when the hormone actually binds and that is how the protein goes to a nucleus.

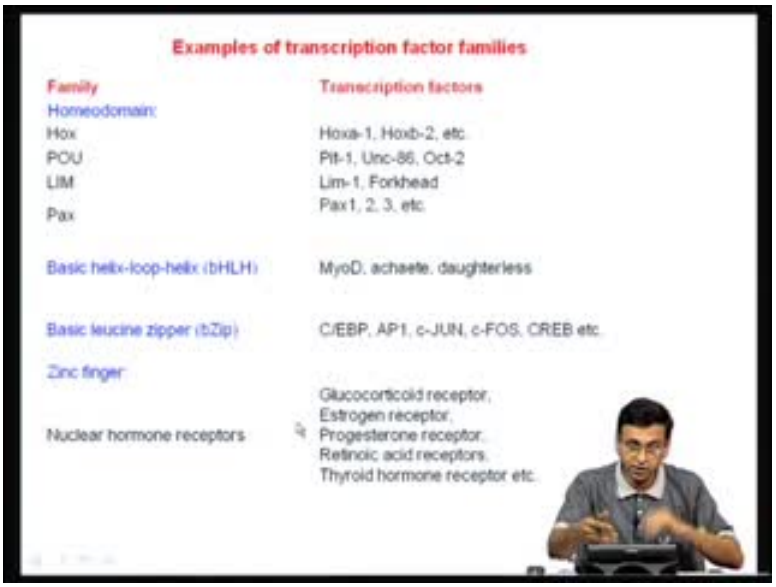
In this case, the protein is prevented from going into the nucleus because it is interacting with another protein called I kappa B. It is called I kappa B because it is called inhibitor of kappa B, nuclear NF kappa B. Now, as long as this heterodimeric NF kappa b is interacting in the I kappa B, the protein stays in the cytoplasm. Now, when there are specific situations like an inflammatory response, when you have to activate the genes in response to this kind of inflammatory signals, now, this NF kappa B has to be translocated inside a nuclei. What happens? Through a cascade of events which again we will discuss later, this I kappa B protein gets phosphorylated.

Now, when I kappa B gets phosphorylated, the phosphorylated form of I kappa B is actually a substrate for what is called as a ubiquitylation of I kappa B. Again, we will discuss these things later. Ubiquitylation is another important post translational modification and usually proteins, which are ubiquitylated, they are degraded. So, ubiquitylation of protein is actually signal for it to be taken to a degradation path way.

Once the I kappa B is ubiquitylated, the protein degradation machinery goes and degrades the I kappa B and therefore, in the absence of I kappa B, the nuclear factor actually the NF kappa B now, can go inside the nucleus, bind to specific response elements called as NF kappa B response elements of various promoter of genes, which are actually need to be expressed in response to inflammation and all these genes are getting activated.

So, I have given all these various examples just to tell you that over the years, using the various experimental approaches I have described, people have started understanding how various eukaryotic genes are getting activated. I gave you some 4 examples, but you can go and then find out a number of such examples.

(Refer Slide Time: 52:56)



So, as I showed here, there are now families of transcription factors. There are a number of eukaryotic transcription factor families and many times these transcription factor families have been named based on the kind of DNA binding domain they have.

I told you the transcription factor contains a DNA binding domain and a transcription activator domain and depending upon what kind of DNA binding domain they have, these various transcription factors have been classified into for example, homeodomain family, basic helix loop helix family, basic leucine zipper family, zinc finger family and so on and so forth. We will discuss these things more detail in the next class.

(Refer Slide Time: 53:43)

So, what is the crux of the matter, which I told you so far is that transcription activation in mammals depends on more than 2000 site specific transcription factors, which interact with general transcription machinery, either directly or through intermediate proteins known as coactivators and as a result of this, transcription activation actually takes place.

Now, again, I have not touched so far about the interaction of the transcription factor with general transcription machinery. How exactly the interaction takes place, we will discuss couple of classes later. So far, I have been telling about only binding to a specific DNA sequence, but once the transcription factor binds to a specific DNA sequence, it has to interact with the general transcription machinery and RNA polymerase. That itself is a huge problem which has been solved and how exactly they interact? Do they directly interact with general transcription factors or do they interact through certain proteins called as coactivators or corepressors, we will discuss much later.

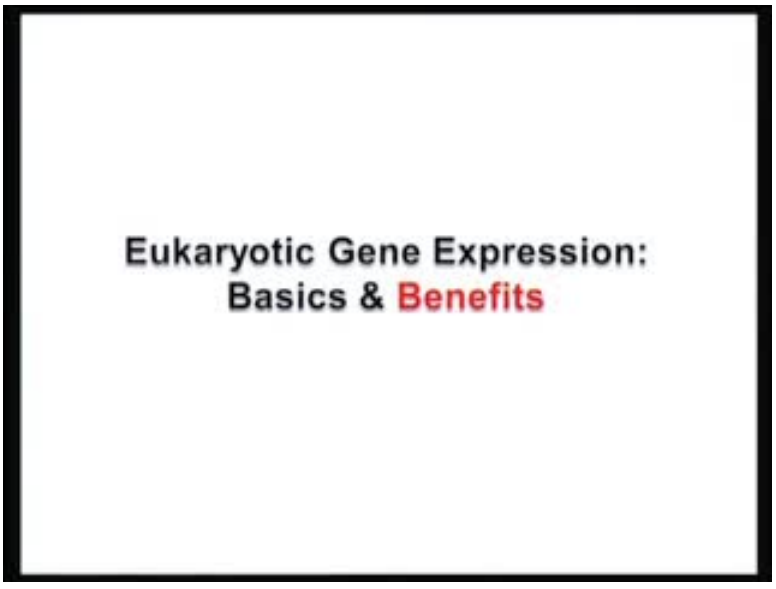
So, coming back to this point, as development and differentiation proceeds selection among these more than 2000 transcription factors ensures regulation of cell specific gene expression. So, there are more than 2000 transcription factors, each of which are selected at different time points during development or during the adult (( )) and these combination of this transcription factors then go and activate the transcription of specific genes and this kind of a differential gene regulation is very important during embryonic development. Again, we will discuss this later, much later.

So, during embryonic development, promoters of specific genes contain transcription factor binding sites that are often organized into clusters called cis-regulatory module or CRMs. Now, so far, I have been talking about only one, one particular element. I told you glucocorticoid goes and binds with this element and activates transcription, heat shock transcription factor goes and binds to heat shock element and activates transcription.

But during development, if you look at certain promoters of certain genes, it is not one or two elements. They are a combination of multiple cis acting regulatory elements to which a number of different sequence specific DNA binding proteins or sequence specific transcription factors bind and it is the synergy synergistic effect of all these transcription factors that is ultimately used in the activation of that particular gene.

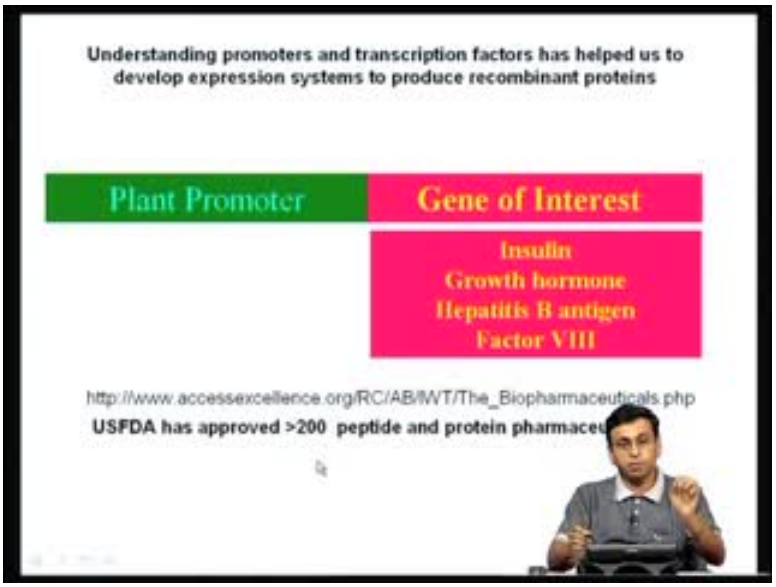
So, development and regulation of gene expression is much more complicated, which involves very complex cis acting regulatory elements called as cis-regulatory modules or CRMs, which contain multiple transcription factor binding sites and synergistic action of all these things ultimately lead to transcription activation.

(Refer Slide Time: 56:03)



Now, I am going to just take 2 or 3 minutes to explain to you that So, people can ask you have studied all these promoters, you have identified so many transcription factors, has there been any benefits arising out of this. What kind of benefits have we got?

(Refer Slide Time: 56:08)



I will just give you one example. For example, from all that I have told you now, it is becoming very clear that you have what are called as promoters. When transcription factors go and bind to these promoters, the downstream gene gets activated. But this principle itself has now led to what is called as production of recombinant proteins in a expression system of your choice. So, understanding promoters and transcription factors has helped us to develop expression systems to produce a number of recombinant proteins.

For example, you want to make insulin, you want to make growth hormone, you want to make recombinant hepatitis b vaccine by expressing hepatitis b antigen or you want to make factor VIII, which is a very important clotting factor. All that what you have to do, you have to take the gene coding for these proteins and then put in front of promoter of your choice.

For example, you want to make a protein in bacteria, you put a bacterial promoter and put this plasmid bacterial cells. Now, bacteria will start making protein of your interest. On the contrary, if you want to make this protein yeast cells, you simply put an yeast promoter and now, yeast cells will start making their protein of your interest.

Similarly, if you put an insect promoter, your protein will be made in insect cells or if you put a mammalian promoter, your protein can be made in mammalian cells or if you

put a plant promoter, you can make you protein of interest in plant cells. And you can see, so, this particular ability to make recombinant proteins and expression systems of your choice has now became a billion dollar industry, you know.

Recombinant protein production growth hormone, hepatitis b, antigen, insulin various (( )), there are so many useful proteins can now be made in expression system, in mammalian cells, plant cells, insect cells, bacterial cells and can be used for diagnostic and therapeutic purposes and if you actually go to this particular website, they actually tells you that the US food and drug administration has actually approved more than 200 such recombinant proteins or peptides, which have been produced using this recombinant DNA technology.

(Refer Slide Time: 57:58)

So, in addition, there are many transcription factors and promoter elements themselves are actually targets for a number of diseases. By turning on a specific promoter or turning off a specific promoter, you can actually prevent or cure a particular disease. Again, we will discuss this in greater detail, as we understand little bit better about how transcription factors function. So, remember, there have been a number of benefits arising out of all these studies on eukaryotic gene expression.

So, I think I will end my talk here. So, what I actually told you so far is that in addition to the core promoter elements and general transcription factors binding in the near

upstream region, there are multiple upstream sequence called enhancers or proximal or distal promoter elements to which specific sequence, specific transcription factors bind and as a result of all these things, ultimately, the rate of initiation of transcription of a particular eukaryotic gene is either up regulated or down regulated. Thank you.