Eukaryotic Gene Expression

Prof.P.N.Rangarajan

Department of Biochemistry

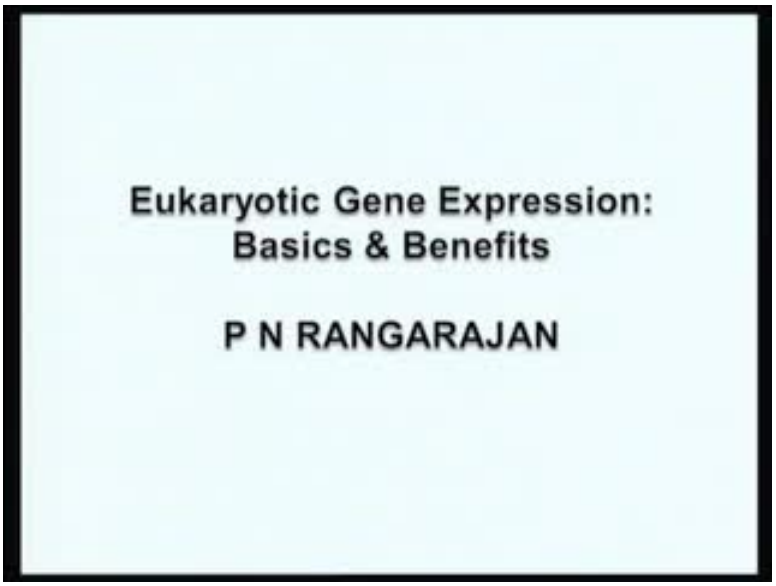
Indian Institute of Science, Bangalore

Module No.# 01

Lecture No.# 01

Introduction to Gene Regulation in Eukaryotes: Eukaryotic RNA Polymerases and Basal Transcription Factors

(Refer Slide Time: 00:25)



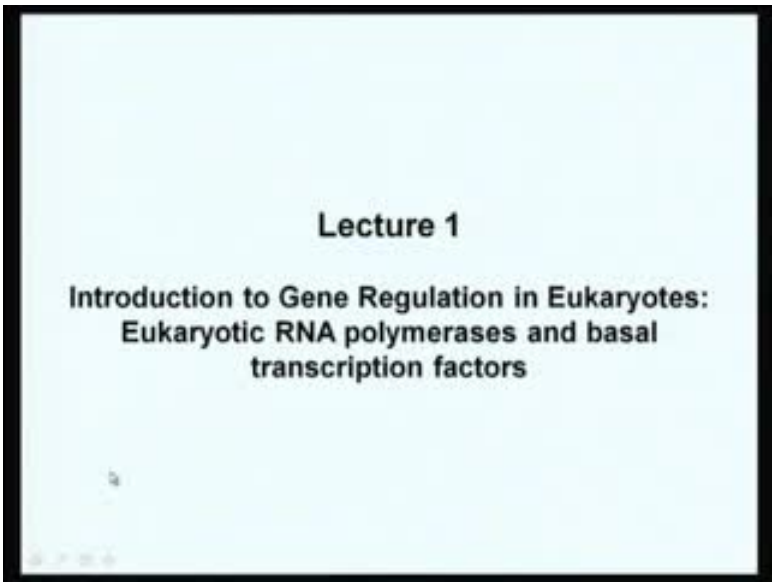
Friends, I am going to now start the first lecture in this course, which is entitled eukaryotic gene expression, on basics and benefits. Eukaryotic gene expression has become a very, very important topic, and this century actually is going to witness a number of very important key discoveries in this area.

So, I have structured this course in such a way that I am going to take you through a history of how many basic research has been initiated in this area, maybe over the last 8 or 9 decades, and how things have progressed and what stage we are; and during this entire course of research on basic aspects of gene regulation and gene expression, people

have also been in parallel being trying to understand how we can translate some of these basic research findings for the benefit of mankind.

So, the way I have structured this course is to first tell you little bit about the basics of gene regulation; and once as we start understanding some of the basic aspects of gene regulation, I am going to now mingle this basic research into some of the important advances, or important key discoveries, that has really led to some improvement of vaccine development, new therapeutics, new drugs, new preventive tools, etcetera and so on so forth, so that you start appreciating why it is important to understand gene expression, and how studying this gene expression is going to help us to understand and find cures, or therapies, for many important diseases.

(Refer Slide Time: 01:49)

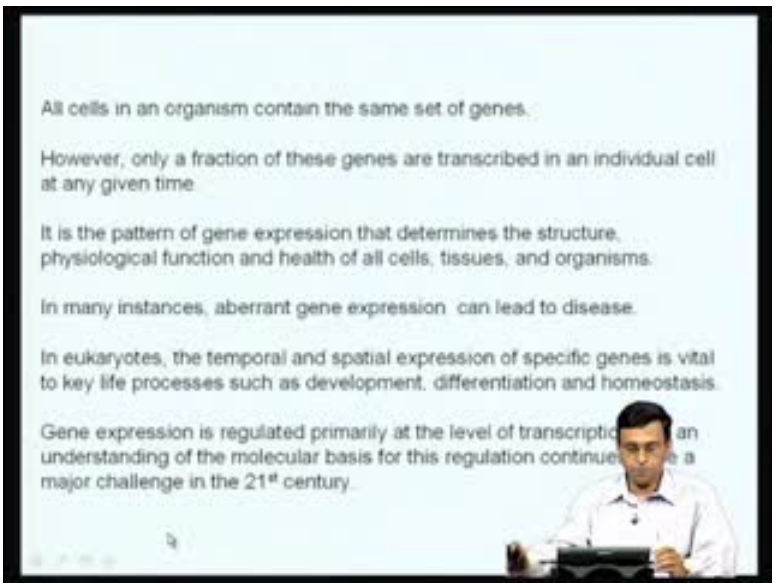


So, the first lecture in this series, I am going to give a very brief introduction about gene regulation in eukaryotes. We are basically going to talk about the workhorses for gene expression, namely the RNA polymerases. These are the enzymes, which are actually involved in synthesis of RNA from DNA.

We are also going to understand how this RNA polymerase is able to transcribe a gene, and how certain accessory proteins called basal transcription factors actually help the RNA polymerase to transcribe a gene. This is what the crux of the today's class.

How RNA polymerase transcribe a gene, and how certain protein factors called as basal transcription factors assist the RNA polymerase to go and recognize what we call as a promoter elements and then regulate the expression of various genes?

(Refer Slide Time: 02:35)



Now, let us now try to understand that what is the need for regulating gene expression. Now, all the cells in organism contain the same set of genes. Now, whether you take a muscle cell or whether you take a liver cell, whether you take a brain cell, the number of genes is more or less the same in all the cases.

In fact, this. For example, today, if I say the human genome encodes about 30000 to 40000 genes, and every cell of our body has the same number of genes. But, what is very important, as far as the gene regulation is concerned, not all the genes are actually transcribed in any individual cell at any given time.

The genes which are transcribed in a muscle cell are different from the genes which are being transcribed in a liver cell or a brain cell, and so on and so forth. So, the same complement of genes are not transcribed in every cell of our body. So, this is what is called as a differential regulation of gene expression.

This is what is responsible for a liver to look like a liver, or a muscle tissue to look like a muscle, or a brain tissue to look like a brain, and so on and so forth. So, only fraction of

the genes, or the total complement of genes, are transcribed in an individual cell at any given time.

So, it is this pattern of gene expression that determines the structure, physiological function, and health of all cells and tissues in the organisms. So, the differential gene expression is very important and, although all the cells of our body has a same complement of genes, not all genes are transcribed at same time. And only some genes are transcribed where other remains silent; and this differential regulation is what is responsible for the development and differentiation of an organism.

Now, what is very important is that if this gene expression does not take place in a normal manner, or aberrant gene expression can manifest in the form a disease. A number of genetic disorders, including cancer, can actually manifest if there are aberrations in this regulation of gene expression. So, what kind of genes have to be expressed at a given time in a given cell is very important, and this is very finely regulated in our body.

So, in fact, one of the biggest challenge in the twenty first century is to understand how this differential gene expression is brought about. Now, as far as eukaryotes is concerned, the temporal and spatial expression of specific genes is vital to key life processes, such as development, differentiation, and homeostasis.

Especially, for those of you have studied developmental biology, you will understand and appreciate, from the single cells I got through a series of divisions, it develops into an adult organism; and each and every step of this development and differentiation specific genes are expressed at specific time points, and this spatial and temporal expression of these genes is actually responsible for development of a fertilized egg into an adult organism.

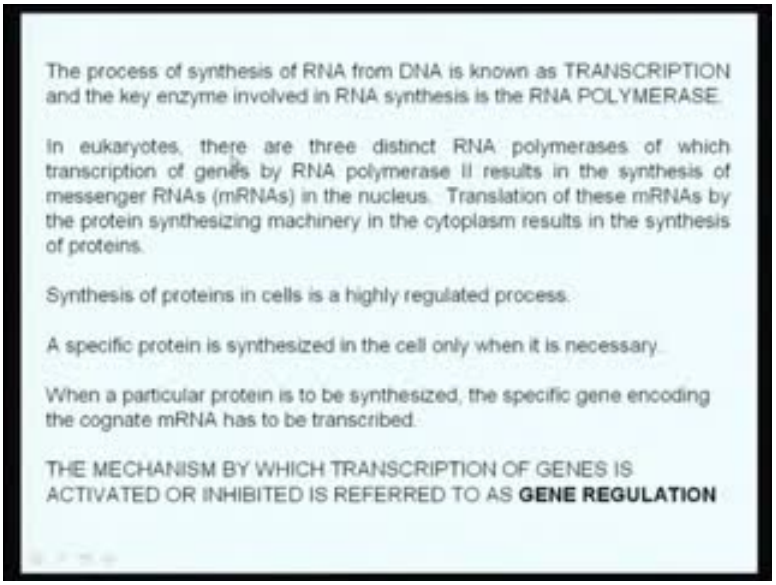
Even in the adult, once you become an adult, specific gene expression in specific tissues and their proper regulation is responsible for the correct functioning of the various tissues in our body.

So, temporal and spatial expression of specific genes is vital for all the cell life processes especially development, differentiation, and the homeostasis. Now, gene expression is

regulated primarily at the level of transcription, and an understanding of the molecular basis for this regulation continues to be a major challenge in twenty first century.

So, what are going to now try to understand is, what we know about gene regulation in the last few decades, and what are the challenge that we are going to face in the twenty first century, and what are the key aspect that we are going to understand in the next few decades.

(Refer Slide Time: 06:00)



Now, let us start with some very basics. Now, the process of synthesis of RNA from DNA is known as transcription, which I am sure all of you know; I do not have to elaborate on this; and the key enzyme that is involved in the RNA synthesis is RNA polymerase. So, the enzyme which is responsible making RNA from DNA is kind RNA polymerase.

So, in today's class, we are going to focus entirely about this RNA polymerase and try to understand how this enzyme is going to make RNA from DNA, and we have in the next 2 classes we are going elaborate little bit more about what kind of transcription factors are actually helping the RNA polymerase to transcribe by a gene.

Now, in eukaryotes, there are 3 distinct types of RNA polymerases, of which the RNA polymerase, which transcribes protein coding genes or which makes messenger RNA, is

called RNA polymerase 2. So, the transcription of genes by RNA polymerase 2 results in the synthesis of messenger RNAs in the nucleus.

Translation of these mRNAs in the protein by the protein synthesizing machinery in the cytoplasm results in the synthesis of proteins. So RNA polymerase 2 in eukaryotes is primarily responsible for making proteins.

So, the polymerase 2 goes and binds the promoters of protein coding genes. It makes messenger RNA, and this messenger RNA comes into cytoplasm gets translated into proteins. What there are 2 other RNA polymerases. We will talk about it little bit later.

Now, coming back to the other point, the synthesis of this protein in cells is a highly regulated process. I told you, in a sense, not all the proteins are made in the all the cells of our body. Specific proteins are made in specific cell types, and we need to understand how this regulation is brought about. Although RNA polymerase is there in all the cells, and all those genes are there in the cells, the RNA polymerase does not transcribe all the genes, all the time, and all the tissues. It is very selective.

Now, when a particular protein is to be synthesized, the specific gene coding for the cognate mRNA has to be transcribed. This is where the regulation is comes into the picture. Only those proteins needs to be synthesized, and only those genes need to be transcribed the RNA polymerase, whereas other genes whose proteins are not required, those genes should not be transcribed by RNA polymerase. This is the molecular basis for differential gene regulation.

So, as we have for a very basic definition of what we are going to study today, the mechanism by which transcription of genes is activated or inhibited is referred to as gene regulation. This is the primary objective of this entire course; trying to understand how this gene regulation is brought about, how this RNA polymerase is selectively able to activate the expression of certain genes and inhibit the expression of other genes, is what is going to be the crux of this particular course.

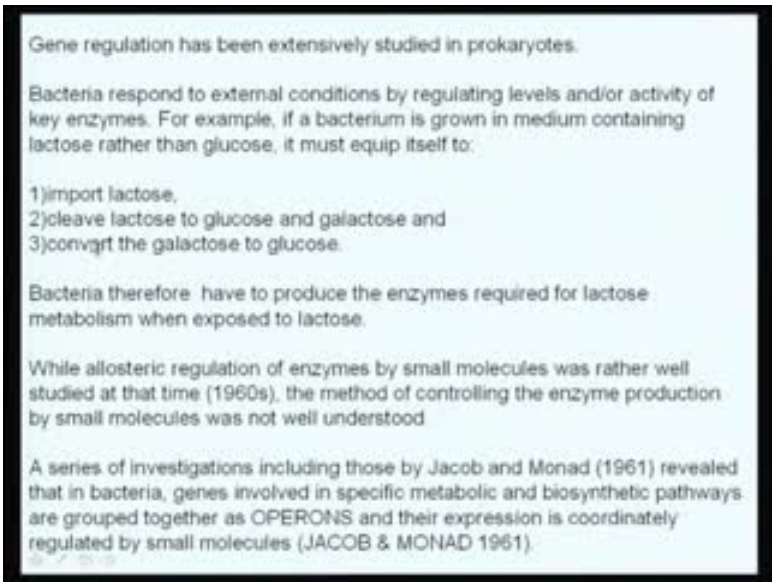
Now, before we go to the eukaryotes, let us understand, gene… the gene regulation has been extensively studied in prokaryotes because prokaryotes are very simple organisms.

They are unicellular organisms and they have much less complicated, much less complex than eukaryotes.

So, in the last century, early part of the last century, a lot of effort has gone in to understand how gene regulation is takes place in prokaryotes. So, let us take bacteria, which are simplest form of prokaryotes.

Bacteria respond to external condition by regulating levels or the activity of certain key enzymes. For example, if a bacterium is grown in a medium containing lactose, normally glucose is the most preferred carbon source because the simplest form of sugar and it can be readily metabolized, and very easily energy can be generated without expending energy.

(Refer Slide Time: 09:43)



But, suppose there is no glucose in the medium and there is only lactose, and lactose as you know, is a disaccharide comprising of glucose and galactose, and therefore, if the organism has to metabolize lactose, how does it react?

Now, it is what does the cell has to do, if it has to know metabolize lactose instead of glucose. First, it has to import the lactose from the medium, and then it has to cleave the lactose to glucose and galactose, because as I said, galactose is disaccharide; it has to be first broken down into glucose and galactose, and then the galactose has to be converted

into glucose, and then glucose then enter by… enters the metabolism like glycolysis, Kreb cycle, and so on so forth, and energy is generated.

Now, bacteria, therefore, have to produce the enzymes required for lactose metabolism only when exposed to lactose; it makes common sense. When there is no lactose in the medium, why should the organism make enzymes that is required for import of lactose, or to cleave lactose to glucose or galactose, or convert galactose to glucose? It is totally a waste of energy.

Therefore, bacteria need to express the genes that code for the enzymes involved in these processes only when lactose is present in the medium. Now, in the early part of the last century, the twentieth century, allosteric rare variation, allosteric regulation of enzymes by smaller molecules was very well studied, and enzymology always had an upper hand compared to molecular biology and understanding gene regulation. Therefore, researchers were studying, actually, how enzymes function and how small molecules are metabolized, and so on and so forth.

But, however, how actually enzyme production is actually regulated by small molecules were not very well understood. This is, I am talking sometimes early 1950s and 1960s. So, while you very well know how small molecules are actually catalytically activated by enzymes, and how enzymes metabolize many of the small molecules, is rather well understood. How exactly the small molecules activate expression of genes leading to synthesis of enzymes is not very well understood in the early part of the last century.

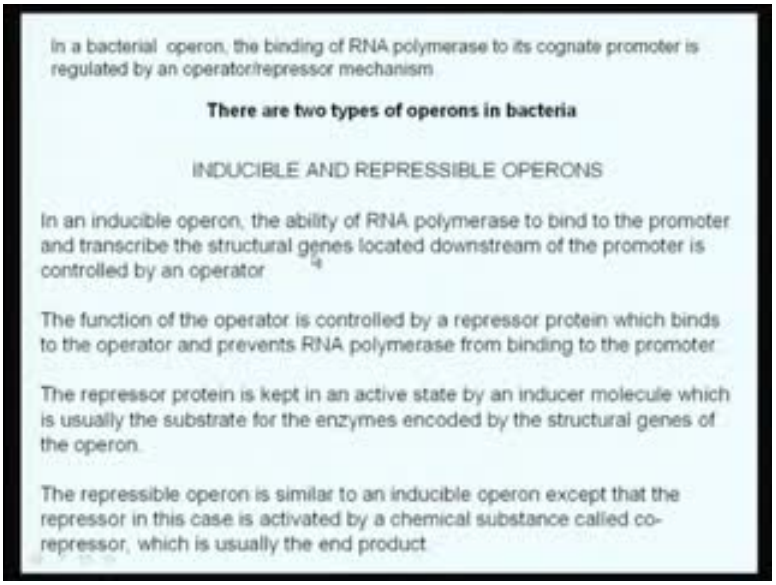
A series of investigations, especially those by the Jacob and Monod, who actually won the Nobel prize at, later revealed that in bacteria, genes involved in specific metabolic of biosynthetic pathways are grouped together as operons and their expression is coordinately regulated by small molecules.

Now, I am not going spend too much time trying to explain to you prokaryotic gene regulation. In fact, I have organized this course with the basic understanding that you people have already studied how gene regulation takes place in, in prokaryotes.

The concept of operons, and how e coli RNA polymerase goes and binds to promoter sequence and activate transcription. These, I am sure, you must be having some basic knowledge. So, I am not going to tell too much about the operons.

Just supposed to know that, in the case of bacteria, genes involved in specific metabolic pathways or biosynthetic pathways are organized in the form of operons, and therefore, these genes are coordinately regulated. This is what we need to know.

(Refer Slide Time: 12:23)



Now, in a bacterial operon, the binding of RNA polymerase to its cognate promoter is regulated by an operator repressor mechanism. I am sure this also you have very well understood, and you have studied some of the basic textbooks on how actually a something like a lac operon or trp operon is regulated. I am going to give a very brief introduction about it, just to recapitulate what already about bacterial gene regulation.

I am sure you are also aware there are actually 2 types of operons in bacteria. One is inducible operon, another is repressible operon. Now, the best example, I am sure most of you have studied it for inducible operons, is— In a inducible operon, the ability of an RNA polymerase to bind to the promoter and transcribe the structural genes located downstream of the promoter is controlled by an operator sequence.

So, you have an operator sequence linked to a downstream promoter sequence, and usually a repressor molecule goes on binds in the operator and prevents the RNA

polymerase from binding, and that is what I have written here. The function of an operator is controlled by a repressor protein, which binds to the operator and prevents RNA polymerase from binding the promoter. So, this is how the operon is not transcribed or it is kept in a silent mode.

So, as long as the repressor binds the operator sequence and prevents the RNA polymerase from binding the promoter, the operon is not functional, and therefore, these genes are not transcribed. And therefore, enzymes of that particular pathway is not synthesized.

Now, the repressor protein is kept in an active state by an inducer molecule, which usually is the substrate for the enzymes encoded by the structural genes of the operon. For the best example is, for example, lactose. I am sure all of you have studied.

When there is a lactose in the medium, the lactose goes and binds the lac repressor and prevents the lac repressor from binding to the operator, and therefore, RNA polymerases can now go and bind to the promoter sequences and transcribe the structural genes.

Now, all these genes now result in the synthesis of enzymes, which will then import lactose in large amounts, cleave the lactose into galactose and glucose, and then the galactose is further epimerized into glucose, and then glucose is further metabolized and derive energy.

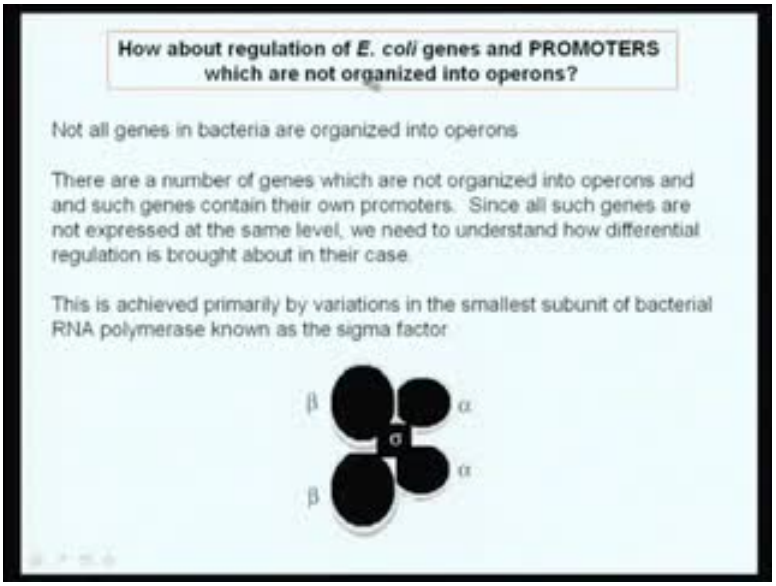
You also have examples of a repressible operon. For example, the tryptophan operon wherein, it is also similar to inducible operon, except that the repressor, in this case, is activated by a chemical substance called co-repressor, which is usually the end product.

Now remember, inducible operons usually are involved in metabolic pathways, in catabolic pathways like lactose degradation and so on and so forth, whereas repressible operons are usually involved in biosynthetic pathways like, for example, tryptophan biosynthesis. It is actually the trp repressor, which is the end product of the tryptophan biosynthesis that acts as a inhibitor of trp operon.

Now, I am not going to talk a little bit more about the operons, because I am sure all of you aware of it. But, one point I want to make, before I switch to eukaryotic gene regulation, is that not all the genes in in prokaryotes are always in the form of operons.

There are many genes which are have their own individual promoters, and then you also question— do all these genes which have their own individual promoters, are they all regulated in the same manner? They are all expressed at the… Do they make same levels of RNAs? The answer is no.

(Refer Slide Time: 15:33)

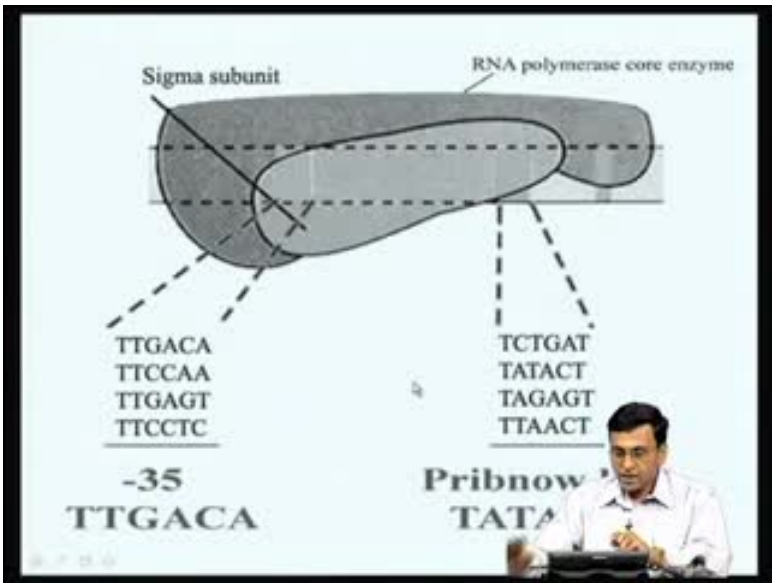


So, there is differential gene regulation of these individual genes. That means, there has to be regulatory mechanisms of these genes, which are not organized into the operon. So, what I want to now stress, may be couple of minutes, is, how about the regulation of e coli genes and promoters, which are not organized into operons?

So, there are a number of genes which are not organized into operons and such genes also contain their own promoters. And since all the such genes are not expressed the same level, we need to understand how differential regulation is brought in this case.

Now, this is primarily achieved by the variation of the smallest subunit of bacterial RNA polymerase, known as the sigma factor. I am sure all of you are aware that the e coli RNA polymerase or the bacterial RNA polymerase is a tetramer, which consists of 2 alpha subunits and 2 beta subunits, and it also has a very small subunit called sigma subunit.

(Refer Slide Time: 16:18)



Now, the sigma subunit is actually responsible for gene regulation in the case of eukaryotes, and these genes which have their own individual promoters, these genes are differentially expressed by, actually, by synthesizing different sigma factors.

Now, it turns out, if you look at the e coli promoters, the among the in the RNA polymerase holoenzyme when I say RNA polymerase holoenzyme, which means, the RNA polymerase core enzyme, which consists of the 2 alpha and the 2 beta subunits, and the sigma subunit.

It is the sigma subunit which actually recognizes specific promoter sequences in the upstream region of the genes, and these are all actually called as the minus 35 sequence and the minus 10 sequence of the Pribnow box.

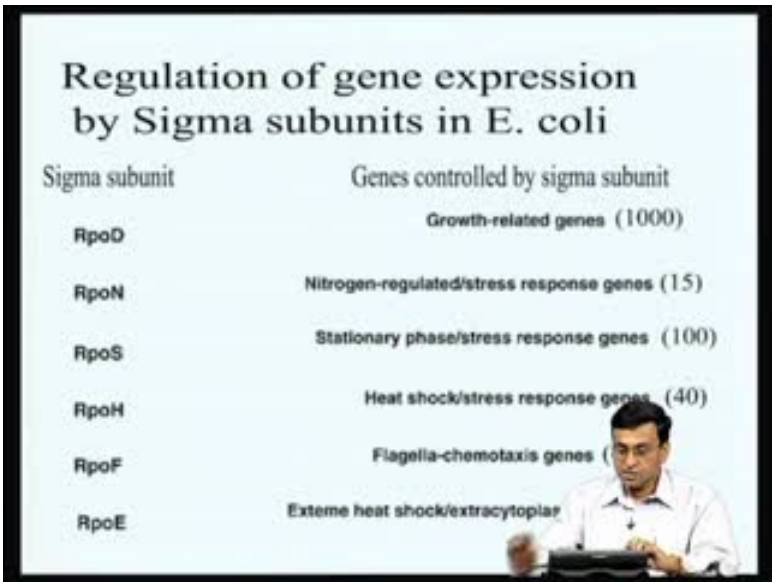
Now, the consensus sequence for the sigma subunit is the Pribnow box usually contains TATAAT sequence, and a minus 35 box usually contains TTGACA sequence, and these are the 2 sequences which are actually recognized by sigma subunit of RNA, e coli RNA polymerase.

But, it turns out, not all the e coli genes contain exactly the same sequence. There are minor variations within the sequence, and I just gave here four examples. There are four different genes where you can see, there are minor variations in this consensus sequence.

It turns out, these minor variations in this minus 35 and minus 10 sequence is actually responsible for differential regulation of gene expression in bacteria. Now, there are certain sigma factors which will only recognize this minus 35 and this minus 10, but not this minus 35 and minus 10.

So, only those genes, which contain this particular type of sequence, will be activated by that particular sigma factor. So, by minor variations in minus 35 and minus 10 sequence, different sigma factors can bind different promoter sequences and activate or repress different genes. I just give an example here.

(Refer Slide Time: 17:59)

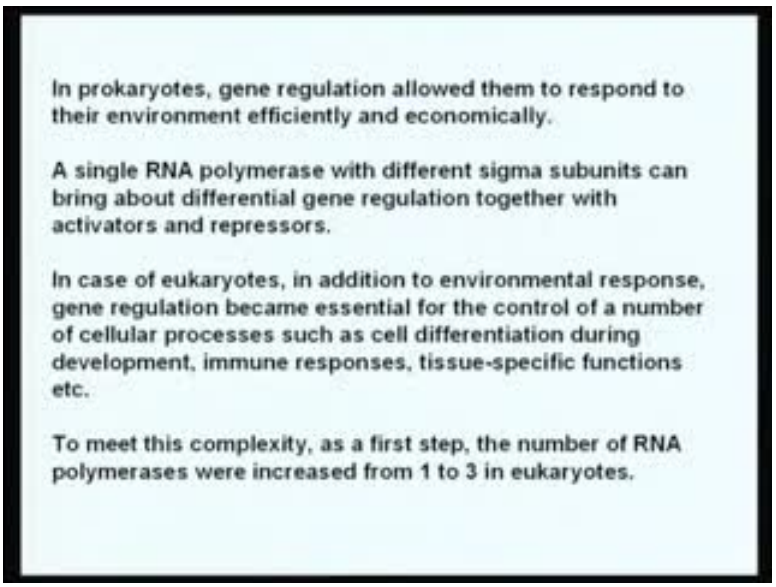


For example, if for example, let us say the e coli has to activate, has to now grow, and there are about 1000 genes which now have been activated for the e coli cells to grow. And what does it do? The e coli now makes a specific sigma factor called RpoD, and this RpoD goes and binds to specific minus 35 and minus 10 sequences of these 1000 growth related genes, and activates all these genes.

Whereas, if suppose, there is a starvation, there is a nitrogen regulation, there is a stress response, there is a nitrogen starvation, or there is a stress response; and let us say, there is above fifteen genes which now need to activated, a specific sigma factor called RpoN now goes and binds to these minus 35 and 10 sequences of these genes and activates the transcription of these genes.

So, by synthesizing different sigma subunits, which have the ability to recognize different minus 35 and minus 10 sequences, e coli can bring about differential regulation of gene expression.

(Refer Slide Time: 18:55)



Now, so, what I told so far is that in prokaryotes, gene regulation allowed them to respond to their environment efficiently and economically. Either the genes can be grouped in the form of operons as I told in the beginning, or there are variations in the minus 35 and minus 10 sequences recognized by the sigma factor, and using these variations, different genes can be transcribed by using different sigma factors. This is how differential gene regulation is brought about in bacteria. There are other variations, which we will not dwell about at this time.

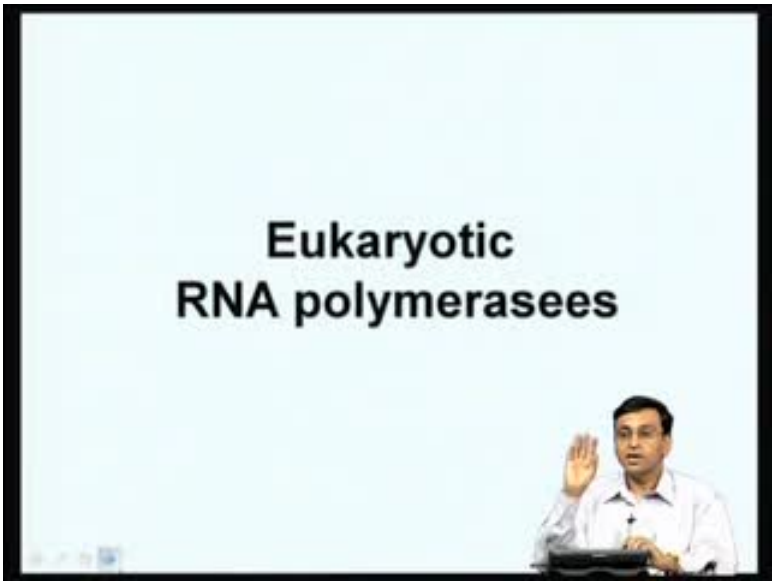
So, in the e coli, a single RNA polymerase with the help of different sigma subunits can bring about differential gene regulation, together with specific the activators or repressors, in the case of operons. Now, now let us come to eukaryotes.

Now, in the case of eukaryotes, in addition to these environmental responses or environmental stress signals, gene regulation became essential for the control of a number of cellular processes, such as cellular differentiation during development, immune responses, tissue-specific functions, and so many other processes.

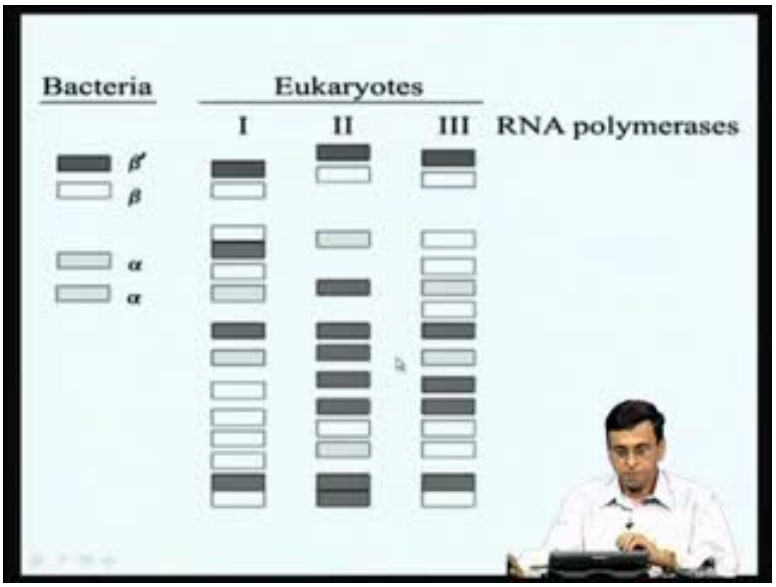
Because the eukaryotes become more and more complicated, you have various tissues and you have developmental band, which is very different. There are many proteins have to be made, and they have to be regulated properly, and the nervous system is much more complex. So, there are all kinds of complications. So, to take care of all these things, the gene expression has to be regulated in much more complicated manner. Then what does happen in the case of e coli?

So, e coli is able to rule just 1 e coli RNA polymerase and a bunch of sigma factors, but this alone is not sufficient. When it comes to eukaryotes, you require lot more complex regulatory machines.

(Refer Slide Time: 20:38)



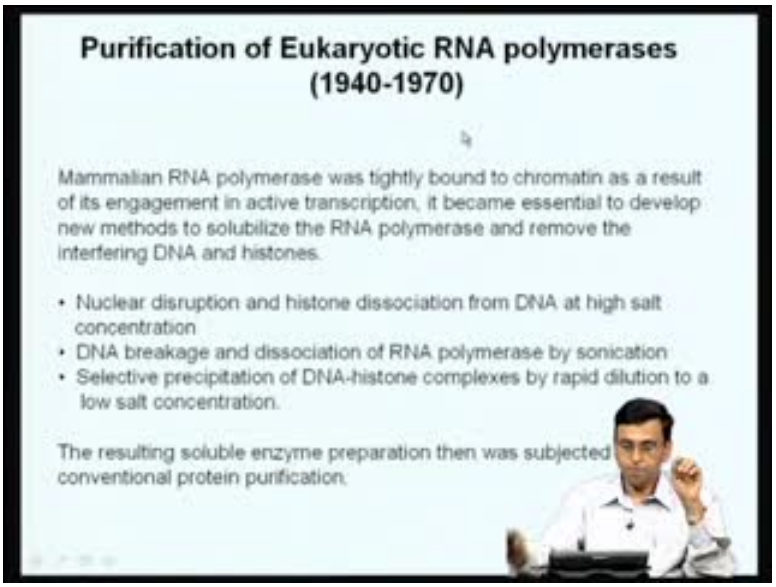
(Refer Slide Time: 20:55)



So, to meet this complexity of this eukaryote, as a first step, the number of RNA polymerases were first increased from 1 to 3 in eukaryotes. So, while in the case of e coli, just 1 RNA polymerase was able to do everything, in the case of eukaryotes, says it, the 1 RNA polymerase became 3 to take care of this complexity.

So, let us now spend some time to understand what are these eukaryotic RNA polymerases, and how they have evolved. So, as I told you, in the case of bacteria, the core RNA polymerase consists of 2 alpha subunits and 2 beta subunits and with the help of different sigma factors, it can bring about a whole bunch of gene regulation.

(Refer Slide Time: 21:18)



Now, when it comes to eukaryotes, instead of one RNA polymerase, now you have got 3 different RNA polymerases, which are usually designated as polymerase 1, polymerase 2, and polymerase 3.

Now, I am going now spend some time on the history. Now, it is always nice to remember some historical aspects and let us see what kind of effort actually went to understand the various eukaryotic RNA polymerases, and biochemistry plays a, played a very important role in a understanding the function of various eukaryotic RNA polymerases.

So, the first step, you want to understand the function of an enzyme you have to now purify. You have to, because when this when a mixture of protein, you cannot really study and then understand the function of a particular enzyme.

RNA polymerases are nothing but enzymes. These are enzymes which make RNA. So, as is true for any enzymes, if you want to know, study, and understand how these RNA polymerase function, the first thing you have to do is to purify the eukaryotic RNA polymerases. But, there was a problem.

Well, in the case of mammalian cells or in eukaryotes, the RNA polymerase was tightly bound to chromatin; know in the case of e coli, you do not have a very well-organized chromatin structure. But, in the case of eukaryotes, you have a very compact chromatin

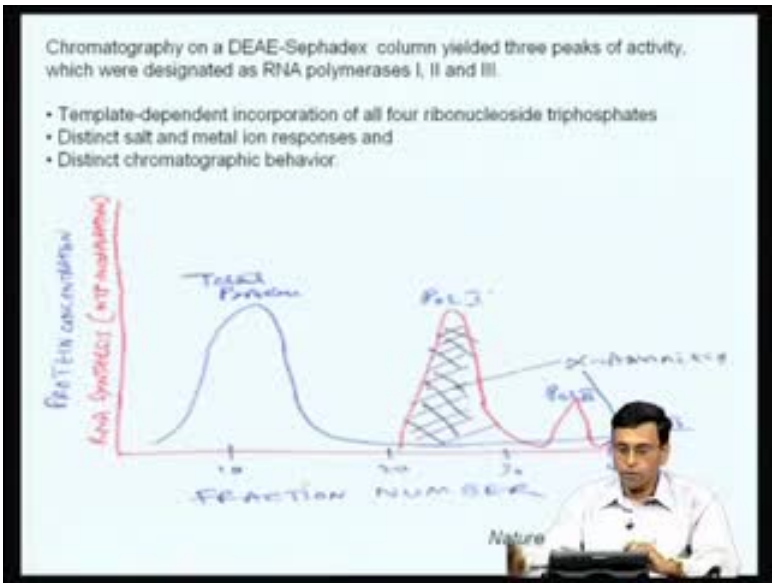
structure, which is organize in the form of a… I mean, inside the nucleus, and this RNA polymerase is very tightly bound to chromatin as a result of its engagement to active transcription; it first became essential to develop new methods to solubilize RNA polymerase and remove the interfering DNA and histones.

Because DNA is tightly bound with histones and is organized in the form of chromatin, and if you want to study, understand, how RNA polymerase regulate, and you want to purify RNA polymerase from this, first you have to dissociate histone and DNA and then remove the RNA polymerase; then only you can study function.

So, a lot of effort went into under to purify this RNA polymerases, or isolate this RNA polymerase from this DNA and histones. So, the key, or the 3 key steps which were actually discovered to find the purification of RNA polymerases— first, you have to disrupt the nuclei, and then dissociate the histones from DNA using very high salt.

Now, histones are positively charged and DNA is negatively charged. Therefore, their association is very tight and therefore, you have to first dissociate the histones and DNA using very high salt concentration; something like 2 molar NaCl, and once the histone dissociates from the DNA, then you break the DNA and dissociate RNA polymerase by sonication, and then you selectively precipitate the DNA protein complex. And now, you have the RNA polymerase with other proteins in the solution, and that is what your starting material for purifying the RNA polymerase.

(Refer Slide Time: 23:45)



So, this soluble enzyme preparation, which is devoid of DNA and histones, was then used to purify RNA polymerase. Now, I want to spend some time on a very important, very key experiment, which was actually described in nature by Robert Roeder's group in 1969, which is how exactly the 3 RNA polymerase were discovered in eukaryotes.

Now, what basically it is, once you have a soluble enzyme fraction which contained RNA polymerase and many other soluble proteins, the next step is to you have to purify. Now, protein purification as you know, is done conventionally by protein chromatography techniques.

You are aware that proteins can be separated based on their charge, or proteins can be separated based on their mass or molecular weight. Now, based on charge it is called ion exchange chromatography, and if you want to separate based on their mass or molecular weight, you can use what it is called as a gel filtration chromatography.

So, what these people did in the late 1960s? They took the soluble enzyme preparation which is devoid of DNA and histones, which are either prepared from sea urchin embryos, or yeast cells, or drosophila embryos, or hela cells, and so on and so forth; and then put it on a ion exchange column.

For example, in this case, DEAE Sephadex column. Now, what happens when you put this mixture of proteins on a DEAE column? Proteins bind to this column depending

upon their charge, and they can be… proteins may be separated based on their charge, and proteins which do not bind this column, they first elute in the void volume, and then you get different, various other proteins. And once you are sure the proteins can be… Once the proteins are bound to this ion exchange column, they can now be eluted by increasing the salt concentration.

So, depending upon the affinity of the proteins, that is, proteins which have, for example, very low affinity to the ion exchange resin will elute first with very low salt concentration, whereas proteins which have very high affinity for the ion exchange resin will get eluted later.

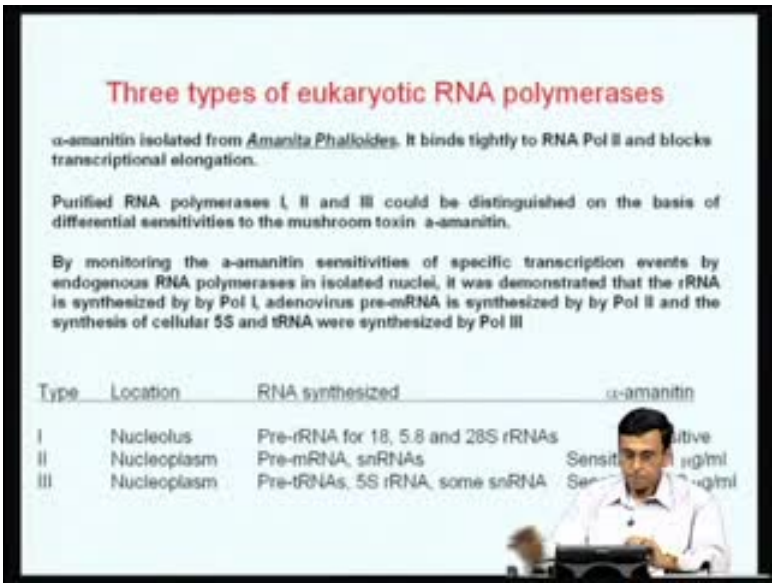
So, we will get… so if you now start collecting fractions with increasing salt concentration, you can collect various proteins in different fractions, and you can see, accordingly, you will get what is called as a elution profile of the protein.

Now, you take these different fractions of these various proteins, which are now separated, based on their charge, and then ask the question— which of these fractions actually contains RNA polymerase activity? When they did this experiment, they found there was no RNA polymerase activity in this total, in this major protein peak, but the RNA polymerase activity was actually present in three distinct peaks, which I have shown here, in the red.

So, the total protein peak is different, and the RNA polymerase activity peak is different, and surprisingly, they got 3 distinct peaks of RNA polymerase activity, which they designated as polymerase 1, polymerase 2, and polymerase 3.

Now, this they have got in a very reproducible fashion. Every time they took this soluble enzyme extract and put it on a ion exchange column, they always got 3 peaks of RNA polymerase activity, clearly saying that this is not an artifact. There are, in fact, some 3 different kinds of RNA polymerase activities in these cells, and they probably have some very important key differences.

(Refer Slide Time: 26:45)



Now, one important difference that I actually found out, is that there is a fungal toxin called alpha amanitin, which was isolated from a fungus called Amanita phalloides, and this alpha amanitin binds very tightly to RNA polymerase 2 and blocks transcription elongation. So, if you add this alpha amanitin to cells, it will go and bind to RNA polymerase 2 and prevent transcription by RNA polymerase 2.

Very interestingly, this alpha amanitin does not bind to RNA polymerase 1, and it requires very high concentration to bind to RNA polymerase 3. So, of the 3 RNA polymerase that they identified, the RNA polymerase 1 is insensitive to alpha amanitin, whereas RNA polymerase 2 is highly sensitive, whereas RNA polymerase 3 requires slightly, for example, ten microgram per ml of alpha amanitin is required to inhibit the activity of RNA polymerase 2.

So, by just, say, simply look at this sensitive of this RNA polymerase to the alpha amanitin, you demonstrate there were actually three distinct types of RNA polymerase—alpha amanitin insensitive RNA polymerase 1, highly sensitive RNA polymerase 2, and an intermediately sensitive RNA polymerase 3, which was highly present in the eukaryotic cells.

By using this particular mechanism, or using this particular observation, what they by simply monitoring the alpha amanitin sensitivities of specific transcription events by

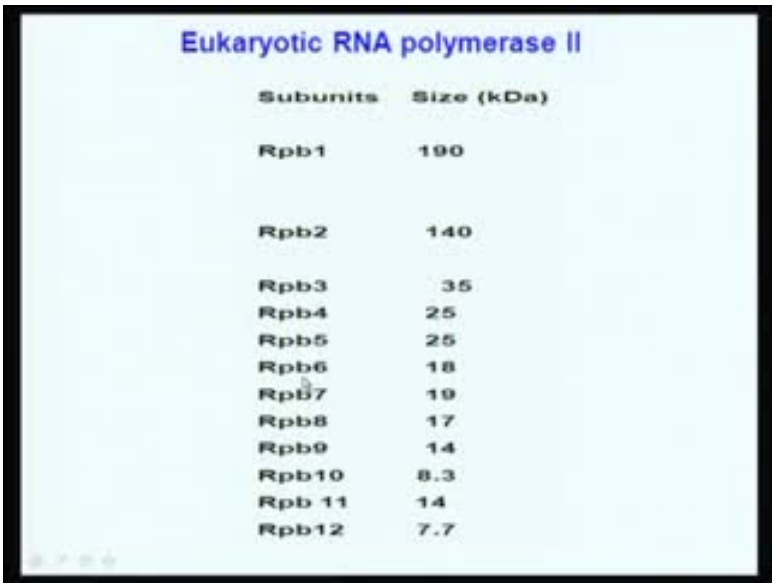
endogenous RNA polymerase in the isolated nuclei, it was actually demonstrated that rRNA is synthesized by RNA polymerase 1, whereas messenger RNA synthesized by polymerase 2, and the 5S and tRNA is actually synthesized by polymerase 3.

That means, what you do, you take the nuclei, and now if you add alpha amanitin, and you find that ribosomal RNA is still being synthesized, which means, that since we already know polymerase 1 is actually responsible for making ribosomal RNA and polymerase 1 is not (( )) to alpha amanitin, it meant that ribosomal RNA is actually being made by the alpha amanitin insensitive RNA polymerase 1.

Whereas in the same nuclei, if you now add a very small (( )) of alpha amanitin, no mRNA could be synthesized, and since in the… from in vitro experiments we already know the RNA polymerase 2 peak is highly sensitive to alpha amanitin, it concluded that, for synthesis of messenger RNA by this RNA polymerase 2 which is highly sensitive to alpha amanitin, RNA polymerase 2 is responsible for mRNA synthesis, and so on and so forth.

Now, today we actually know that, not only there are these distinct RNA polymerases, they actually have very very specific functions, and we know, that the RNA polymerase 1, which actually makes the ribosomal RNA, is actually present in nucleolus, whereas the other 2 RNA polymerase, both 2 and 3, are present a nucleoplasm; one of them actually responds for making messenger RNA and snRNAs, and the RNA polymerase 3 is actually responsible for synthesizing transfer RNA, 5S RNA, and several other small nuclear RNAs.

(Refer Slide Time: 29:31)

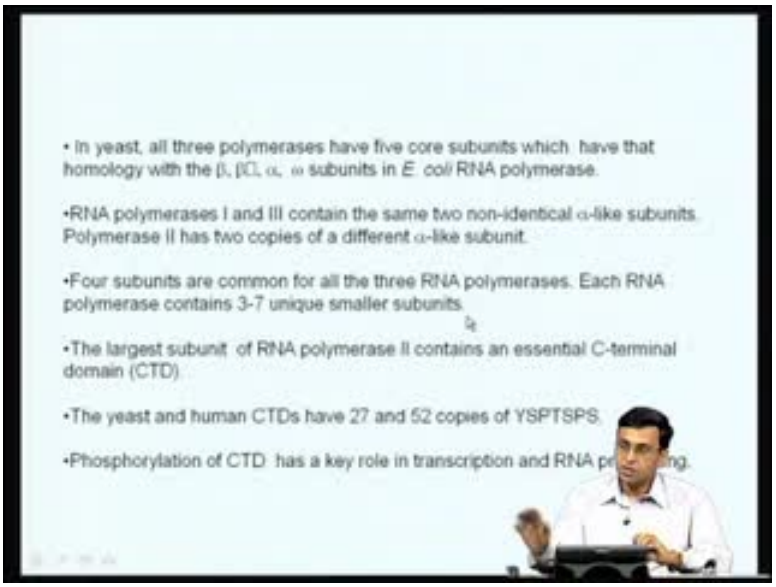


So, this is how the 3 RNA polymerase were actually discovered, now. So, once you started purifying the identify the RNA polymerase activity by using series of chromatographic steps, the RNA polymerases were kept on… were being purified, and once they have got a pure RNA polymerase 2 activity, and when they know, subject to what is called as SGS polyatomic gel electrophoresis, they found this pure RNA polymerase actually has a number of subunits.

So, you can see the equal RNA polymerase core enzyme had only 2 alpha and 2 beta subunits, whereas if you now take the purified RNA polymerase 2, it had at least 9 different subunits or 12 different subunits, as I showed here, and each of them has a different molecular weight.

So, this clearly told that the eukaryotic RNA polymerase is much more complex than the prokaryotic column RNA polymerase and it has many more subunits than the… its prokaryotic counterpart.

(Refer Slide Time: 30:22)



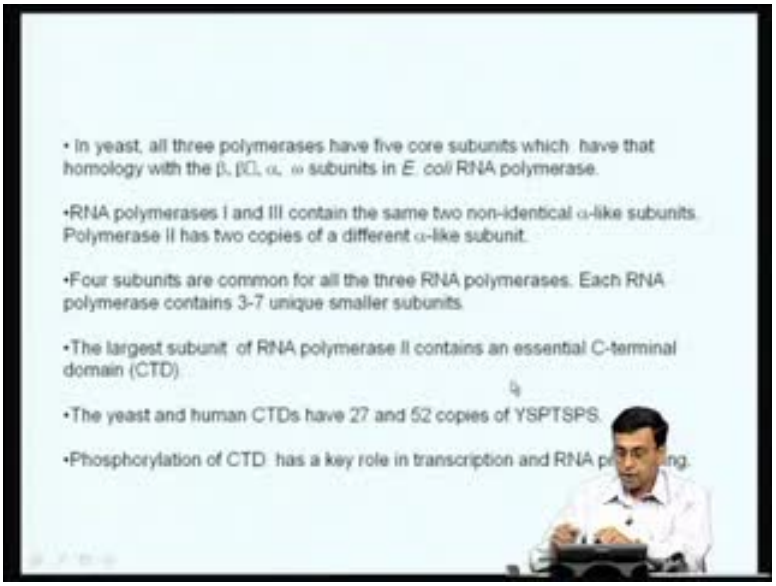
Now, let us now trying to see what are the commonalities and what are the differences in the RNA polymerases between the various RNA polymerases in eukaryotes. Now, in the case of yeast cells, which is also an eukaryote, all the 3 RNA polymerases have 5 core subunits, which have homology with the some of the subunits of e coli RNA polymerases; clearly telling that they are all originated from the e coli. E coli RNA polymerase is the ancestor for evolution of this eukaryotic RNA polymerase.

Now, the RNA polymerase 1 and 2 contain the same two non-identical alpha-like subunits. Polymerase 2 has two copies of a different alpha-like subunits. The four subunits which are present are common for all the three RNA polymerases, that is, of the various subunits of each of these RNA polymerases. Four are common for all the three RNA polymerases, and each of these RNA polymerase have at least 3 to 7 unique smaller subunits.

Now, this is very important, what I am telling you. Now, the largest subunits of RNA polymerase 2, the largest subunit of RNA polymerase 2 means, that is, the 100 and 90 kilo dalton Rpb1. This largest subunit of RNA polymerase 2 has, what is called as a C-terminal domain, where in the next couple of classes this is going to be very important, and I am going to tell you how important is the C-terminal domain of this large subunit of RNA polymerase 2.

Remember, the largest subunits of RNA polymerase 2 contain a very important part called as C-terminal domain or called as a CTD. Now, what is so unique about the CTD, both in yeast as well as humans, and all other mammalian cells, this CTD contains what is called as a YSPTSPS repeat motif.

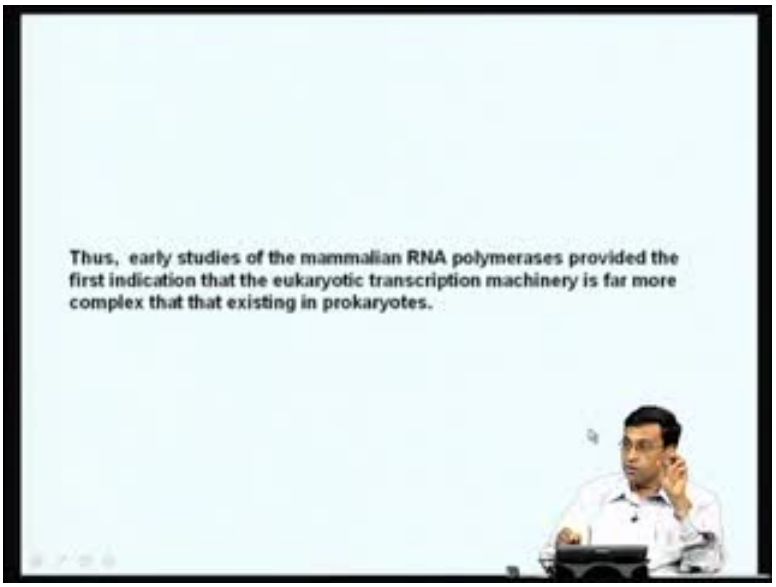
(Refer Slide Time: 30:22)



As you know, Y is a tyrosine; S is serine; proline, threonine. So, tyrosine, serine, proline, threonine, serine, proline, serine; and as you know, serine and threonine are the residues which can be phosphorylated. So, the C-terminal domain has a highly serine and threonine rich motif, and in the subsequent process I will tell you that phosphorylation of the serine and threonine residues in the CTD plays a very important role in regulating the activity of RNA polymerase 2.

So, just remember; the largest subunit of RNA polymerase 2 contains a C-terminal domain, which is has a number of repeats containing serine and threonine residues, and at a later stage I will show you that phosphorylation of this and this plays a very important role in regulating the activity of RNA polymerase. That is what I have written here.

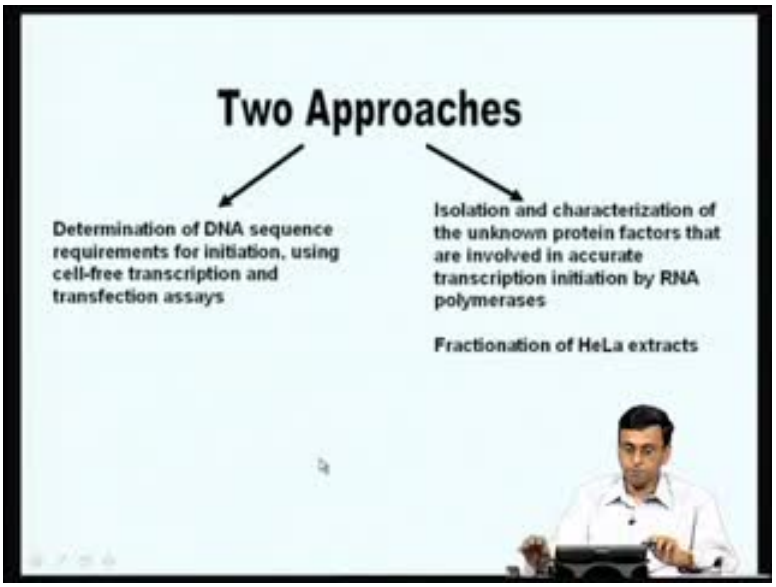
(Refer Slide Time: 32:10)



Now, so what I told you so far, is, early studies of the mammalian RNA polymerases provided the first indication that eukaryotic transcription machinery is far more complex than that existing in prokaryotes.

I think, you will now agree, because of e coli you just had 4 subunits of the core enzyme and you had may be a, a dozen or so of dozen or 2 dozens of sigma subunits, and with that, e coli was able to manage differential gene regulation. But, when we came to eukaryotes, the number of RNA polymerase itself was increased to 3, and as we proceed further, I will now tell you it is not just the increasing number of RNA polymerase 2 subunits, RNA polymerase subunits, but even a number of other accessory proteins are also essential in order to accurately initiate transcription by RNA polymerases in the case of eukaryotes.

(Refer Slide Time: 33:39)



Now, so what I told you is that, historically, around the late 1960s, people began to purify RNA polymerases from various eukaryotic cells, and they still did not understanding that these RNA polymerases are much more complex and they contain multi-subunit processes. And as one group started purifying this RNA polymerase and try to understand how these RNA polymerases are regulating, and isolate and characterize what are the subunits, and so on and so forth, there was also another group which started looking at what are the sequences to which these RNA polymerases go and bind and activate gene regulation.

So, 2 distinct approaches were being simultaneously being followed to understand gene regulation in eukaryotes. In one case, determination of DNA sequence requirements for initiation using cell-free transcription and transfection assays. I would elaborate little bit on this later.

That means, you need to identify how it is exactly that even the eukaryotic RNA polymerase goes and binds the promoter sequence. Like in the case of e coli, I told you, the sigma factor binds to this minus 35 and minus 10 sequence, and that is what is responsible for differential gene regulation.

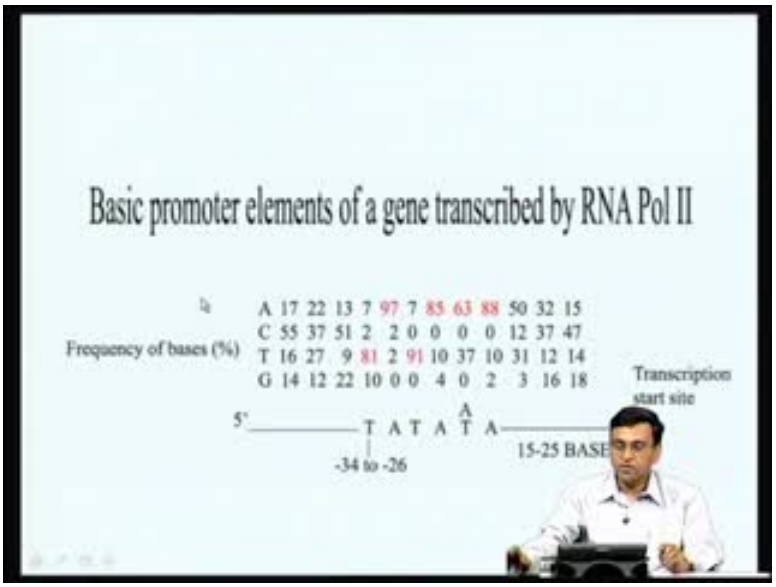
So, are there similar elements in the case of eukaryotic promoters, and what are these promoter sequences, and how these promoter sequences function in vivo by using what

are called transfection assays, or what are called as a cell-free transcription assays, where you actually make cell-free extracts from various cells, and in a test tube, you now add DNA templates, you add various ribonucleotides, and you add ATP and ask the question, whether now the purified RNA polymerase can initiate transcription accurately in a test tube; that this is what is the in vitro or cell-free transcription system.

So, 2 approaches were being followed. On one hand, people were trying to purify RNA polymerases from various cell extracts and what kind of subunit structure, and so on and so forth; on the other hand, people were asking the question, if we now take this purified RNA polymerases and add to a test tube, which contains the promoter region of a gene, and now if you add all the factors which are required for it, such as the template DNA containing promoter sequence, if we now include all the energy, and if you require all the NTPs, will this RNA polymerase be able to initiate transcription? That is, the unless RNA polymerase goes and binds to specific sequences in the promoter, it cannot initiate transcription, and what are these sequences? Are the sequence similar to what is there in the prokaryotic promoters in the eukaryotes?

So, isolation and characterization of unknown protein factors that are involved in the accurate initiation of transcription by RNA polymerases has been carried out on one hand, and when they add all these purified RNA polymerases, they are putting these RNA polymerases in a cell-free system on a transfection assay and ask the question—how these RNA polymerases are able to initiate transcription?

(Refer Slide Time: 36:08)



Now, such studies came into, gave a very important and interesting observation. It turns out, when people started looking at the sequence of this promoter region, especially, the region is about, within about 50 bases from the transcription start site of eukaryotic promoters, they came up with a very interesting observation.

If you now look at some this sequence, you can see, if you take, for example, the numbers I have given is for about is a percentage. If you take, for example, 100 genes and look at the promoter regions of this 100 genes, for example, in this particular position, in 97 out of 100 genes, always it was an A.

Whereas in this position, for example. Similarly, in this particular position, in 85 out of 100 genes, there was again a A, whereas in this particular position, 81 percent of the genes had a T. Here, 97 percent of the genes had a A, 91 percent here had a T residue, 85 percent had a A residue, 63 percent had A or T, 88 percent of genes had a A residue.

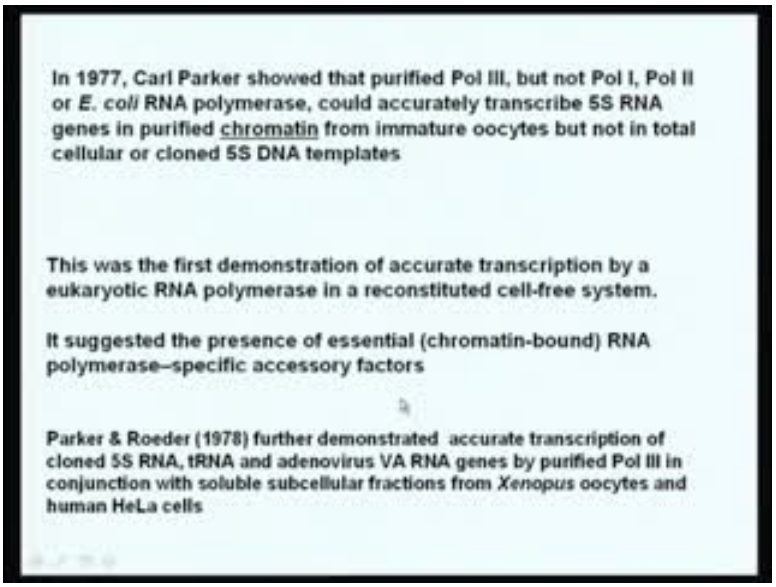
So, which clearly tell that there is some…, this TATATA motif seems to be highly conserved in the promoter region of a majority of protein coding genes in eukaryotic genes, eukaryotic promoters. So, this is now being now recognized as what is called, say, TATA box.

As we go along, you realize that this TATA box is the place where actually transcription RNA polymerase and other factors actually go and bind, and this assembly of RNA

polymerase in and around TATA is actually responsible for accurate initiation of transcription in the RNA eukaryotic promoters.

So, the first major finding about the organization of eukaryotic promoters is the presence of a highly conserved sequence called TATATA motif. This is known as the TATA box, and this TATA box seems to be conserved in a number of eukaryotic promoters. So, this is the first important finding from the analysis of various genes transcribed by RNA polymerase 2.

(Refer Slide Time: 38:02)



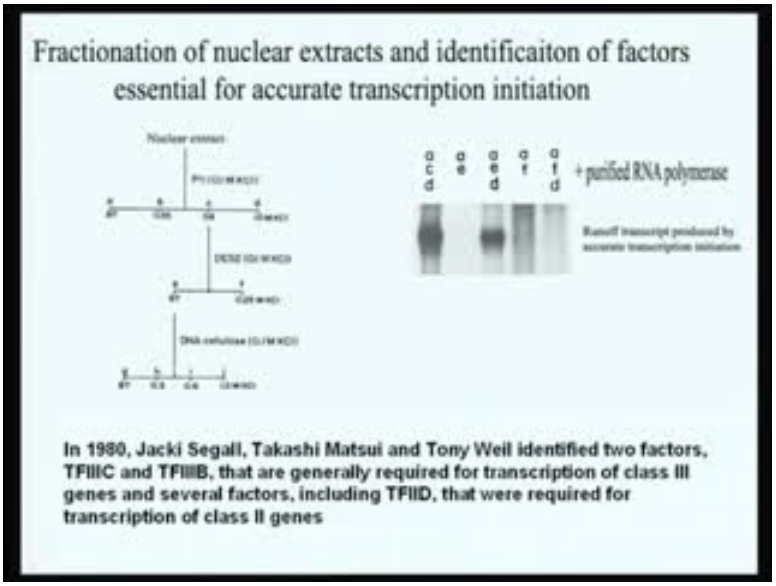
Now, so in the late seventies, people actually started doing experiments, not only in the cell-free systems containing DNA templates, but they also started using chromatin templates. And such studies, for example, by Carl Parker's group actually showed that in the case of RNA polymerase 3, it actually can transcribe the 5S RNA genes in purified chromatin from immature oocytes, but not in total cellular or cloned 5S DNA templates.

That means, the RNA polymerase 2 actually can transcribe the 5S or polymerase 3 can transcribe 5S RNA only in a chromatin-based template, but not in DNA template devoid of chromatin, just plain, naked, DNA template.

Which actually told that, for RNA polymerase to transcribe accurately genes, some factors in the chromatin are also required. Just naked DNA alone is not sufficient. Some chromatin templates containing other accessory proteins are actually extremely required.

So, the… So, this expert actually suggested that there are chromatin bond factors which are actually essential for RNA polymerase 2 actually transcribe some of these genes. There are many other experiments which are carried out. I will give a key references at the end of the presentation.

(Refer Slide Time: 39:24)



You can go through some of these experiments and read a little bit more on how people actually started fractionating extraction, start putting this extract in the cell-free system, and started identifying what kind of factors are actually required for RNA polymerase to accurate initiate transcription. I am just giving one example here. For example, fractionation of nuclear extracts and identification of factors essential for accurate initiation of transcription.

So, as I said, they took the soluble nuclear extract first, put it on a, for example, a phosphocellulose column, and eluted the protein with varying salt concentration; for example, here, the point 35 molar NaCl, point 6 molar NaCl, 1 molar KCl.

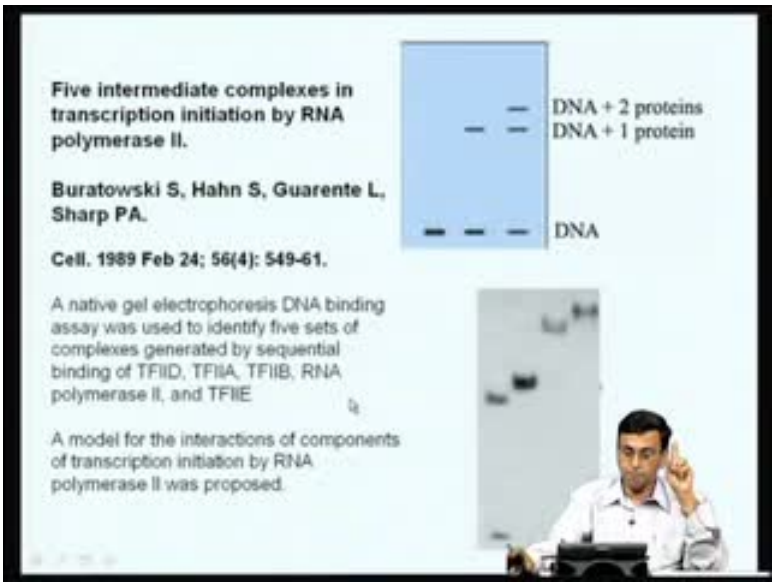
Now, you now take these proteins which got eluted from the point 6 molar KCl fraction, and again put it on a DEAE cellulose column, and then proteins bound to the DEAE cellulose column, again you start eluting at different salt concentration. And again, take this fraction, put on another DNA cellulose column, start eluting at different salt concentration, and you now start looking at each one of these fractions, and see which

one of this fractions contain protein factors which are essential for RNA polymerase to accurately initiate transcription.

Now, I want to describe here a very important assay, which is acts known as the runoff transcription assay. What actually you do in this assay is that you take a DNA template which contains the promoter sequence, and then ask the question— how RNA polymerase will go and bind to this promoter sequence? And it binds, you have what is called as a linear template. Let us say, for example, a 400 base pairs downstream of a promoter. So, when RNA polymerase binds and then starts transcribing the gene it will fall off at the end of this 500 base pairs, and you will get a 500 base pair RNA. So is called a runoff transcript. So, this what I have shown here.

If you now take some of this chromatographic fractions, and then add it to RNA polymerase, some of these chromatographic fraction, for example, a, c, and d. That is, a, c, and d was able to support accurate transcription by RNA polymerase 2, but for example, if you now add a, add only a and e, it does not support, give any runoff transcript.

(Refer Slide Time: 41:18)



So, these… By using these kind of runoff transcription assays, people started identifying which chromatographic fraction actually contains those protein components, which are actually essential for RNA polymerase to accurately initiate transcription.

At the same time, another important study; as people started identifying some of these protein factors which are actually helping the RNA polymerase to bind and accurately initiate transcription, people also start using what is called as the super shift assays.

Now, here, what you do, once you purify each one of these transcription factors, which are actually as protein factors which are helping the RNA polymerase to accurately initiate transcription, you would now start asking the question— are these protein factors interacting with each other or not?

So, what you do is a very important. We have a shift, what is called as a gel electrophoresis mobility shift assay, where you take the promoter DNA fragment, like the containing the TATA sequence, radio label it, and now you add, for example, a protein factor which is, which you have purified in the, as I have showed in the last slide.

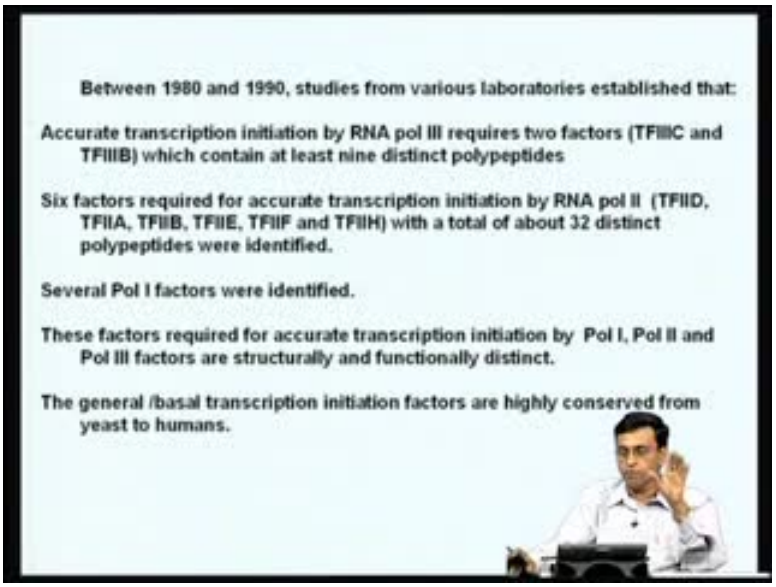
When this protein factor binds this DNA sequence, it causes a mobility shift, and you can see the probe will get shifted here. Now, you add another protein component, for example, when this is just the naked radio labeled promoter DNA alone, which moves here. Now, when this DNA binds to protein, it causes the mobility shift and you get a complex here.

Now, if one more protein is binding to this, now you will see a further shift in the mobility and this indicates that, what it tells you is that, by one by one, number of protein factors can actually assemble over the promoter sequence.

By doing this, generally, people have actually demonstrated there is a very sequential assembly of protein factors on the promoter DNA template. So, this is a key factor which actually demonstrated that certain fine intermediate complex in the transcription initiation were actually identified by RNA polymerase 2.

Based on this thing, they proposed a model by which sequential assembly of transcription factors on the promoter sequence is actually responsible for accurate initiation of transcription by RNA polymerase 2.

(Refer Slide Time: 42:59)



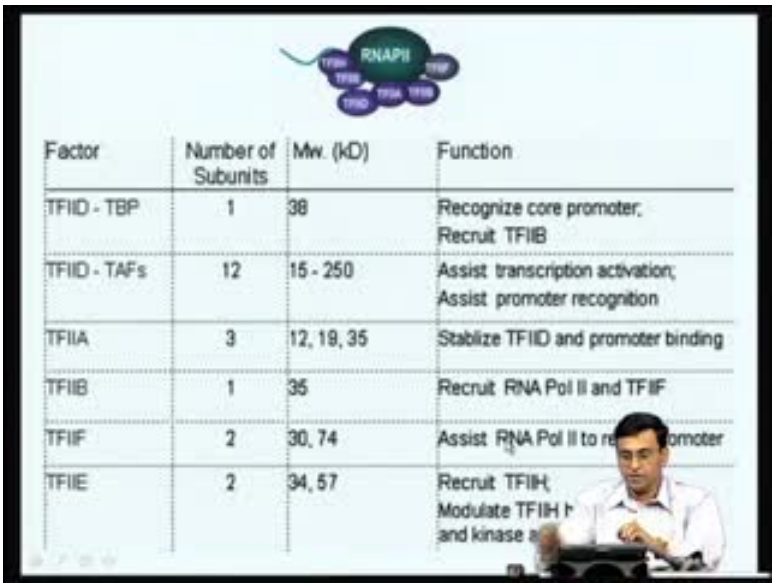
So, between 1980 and 1990s, studies from various laboratories have actually established that accurate transcription initiation by RNA polymerase 2 requires polymerase 3, requires 2 transcription factors called as TF3C and TF3B, which contain at least 9 distinct polypeptides.

So, using runoff transcription assays and also using these kinds super shift experiments, people have identified, at least, there are 2 protein factors which are actually required for RNA polymerase 3 to transcribe tRNA and 5S rRNA genes.

Similarly, in the case of RNA polymerase 2, you require a number of protein factors which are named as TF2D, TF2A, B, E, F, and H, and which contains about, in all put together, they are about 32 different polypeptides, which actually are essential for RNA polymerase to accurately initiate transcription in the case of protein coding genes.

In the same way, several polymerase 1 factors are also identified, and all these factors were found to be structurally and functionally very distinct. So, and they also found out, whether you purify these transcription factors from the yeast, or drosophila, or mammalian cells, or sea urchin cells, they all seem to be highly conserved.

(Refer Slide Time: 44:18)



So, which clearly told you that these protein factors which help the RNA polymerase to accurately initiate transcription, they are highly conserved from yeast to man. That was the outcome of the various research that took place between 1980s and 1990s.

So, I have just given you the summary, here, of the various research effort that went into to identification of all these key accessory proteins that help the RNA polymerase from for transcribing RNA protein coding genes.

For example, you have what is so called as a TF2D, which means transcription factor 2D. One of the important components of transcription of 2D, what is called as a TATA binding protein. This is the one that actually recognizes the TATA box and brings responsible for specific assembly of a RNA polymerase similar in the TATA box. Each contains about, the TBP contains about a 38 kilo dalton protein. Its main function is to recognize the core promoter sequence, that is the TATA box, and also now, then that would be next transcription factor called TF2B.

There is another; the TF2D, also, in addition to TBP, contains what are called as TBP associated factors. There are about 12 of them with a molecular weight ranging from 15 to 200 and 50 kilo daltons and their job is to actually help the RNA polymerase in transcription activation, and also in promoter reorganization. Similarly, you have protein

factors called TF2A, TF2B, TF2F, and TF2E, and each of them has a very specific function in the initiation of transcription, especially, the TF2H.

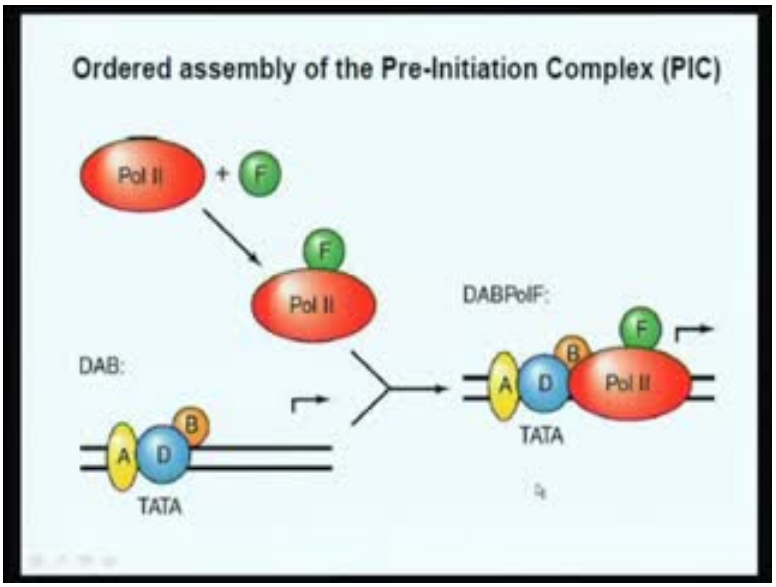
For example, TF2E is actually involved in the recruitment of what is called as TF2H, and the TF2H actually has what is called as a helicase activity, and this helicase activity is actually responsible for modulating the activity for RNA polymerase 2. And I told you in the in the previous 2 slides that RNA polymerase 2 actually contains what is called as a helicase domain, sorry, carboxy terminal domain, which is rich in serine and threonine routines, and it is the TF2H which actually phosphorylates it is carboxy terminal domain of the RNA polymerase, and it actually responsible for converting what is called as initiation the complex to a elongating complex.

So, it is almost like a train that, if the train is there in the station, and the train wants to move, the guard has to wave his green flag, or a green signal has to come and say. Only then, the train starts moving.

The same way, if the RNA polymerase has to leave the promoter region and start transcribe the gene, it has to receive some specific signals. And when that when does, when does that happen, the transcription initiation can be take place when the cell has to that all the components require for transcription initiation is actual present in the cell.

For example, it has all the required ribonucleotides, all the factors required for transcription is actually assembled, and the entire RNA polymerase holoenzyme has actually assembled in the promoter region, and then, when we have all the ribonucleotides, the TF2H now comes and phosphorylates this carboxy terminal domain of the RNA polymerase 2, and that is the signal for RNA polymerase to now leave the promoter, or leave the station and stop moving.

(Refer Slide Time: 47:18)

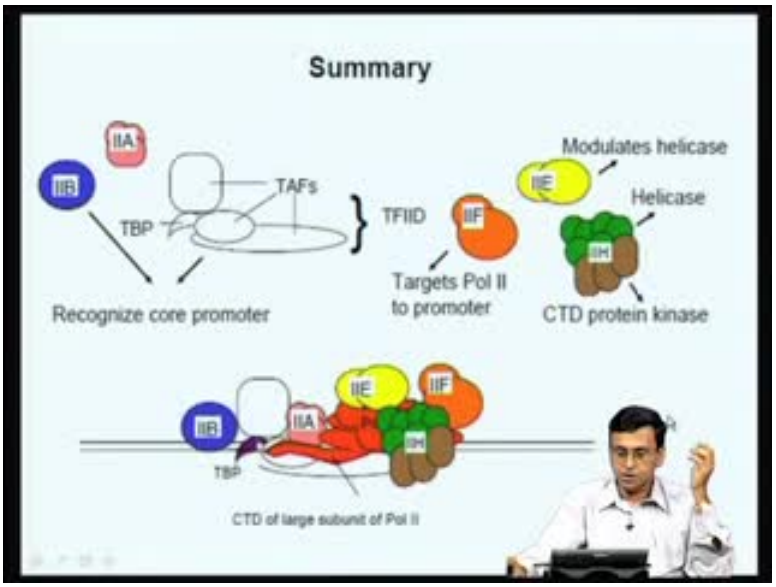


So, remember, the initiation RNA polymerase 2 in the initiation complex is non-phosphorylated, whereas in the elongating RNA polymerase 2 the carboxyl terminal domain is highly phosphorylated, and this phosphorylation is actually done by one of the accessory proteins required for transcription initiation, namely, the TF2H.

It also has other helicase and kinase activities, and so on so forth. So, what I told you so, far, is that there is a very ordered assembly, or what is called as a pre-initiation complex formation in an eukaryotic protein coding gene promoter.

You have, for example, the RNA polymerase 2, which itself contains up to 12 subunits in the case of yeast, and this polymerase 2 first has to associate with the protein factor called TF2F, and this polymerase 2 TF2F is then goes and binds to the promoter region. And in this promoter region, first the TF2D, which contains the TATA binding protein, comes and binds. Then, it recruits TF2B and TF2A, and this is actually called as a formation of a DAB complex.

(Refer Slide Time: 48:18)



So, once the DAB complex is formed, that is the signal for the pol 2 and F and come and join the promoter region, and once the DAB F and pol 2 complex is assembled, now, the TF2H comes and binds here, and it makes sure that everything is ready. Then, it phosphorylates the carboxy terminal domain of RNA polymerase 2, and then the polymerase 2 starts moving and starts the transcription of the various protein coding genes.

So, what I told you in the summary, so far, is, I gave a very brief overview of how actually the transcription initiation takes place in e coli, and how the e coli RNA polymerase regulates gene expression, either in using accessory proteins like activators and repressors in the case of operons, or by using specific sigma factors, it can differentially regulate various bacterial genes.

But, then I told, I had showed, I discussed with you, as we moved from prokaryotes to eukaryotes the RNA polymerase itself became very complex. So, instead of 1 RNA polymerase, you got 3 RNA polymerase in the case of eukaryotes, and in these 3 RNA polymerase, again, the each of them became a multi-subunit complex, whereas eukaryote RNA polymerase has only 4 subunits and the sigma factor in the case of eukaryotic animals, you have anywhere from 9 to 12 subunits.

Then, not only have different subunits of the RNA polymerase, you also need a wide variety of accessory proteins in order for RNA polymerase to go and bind to the promoter and recognize. And you can see, here, this is what is called the… what I have shown here is what is called as a pre-initiation complex, which, actually, is responsible for initiation of transcription in the eukaryotes.

As I explained in my previous slides, you have what is called as the TATA binding protein and you have what is called as the TF2B you have the TF2A and what is called as a TBP associated factors, and these TBP associated factors and TBP together constitutes what is called as a TF2B, TF2D.

These TF2D, in addition to, and along with TF2A and TF2B, are responsible for recognizing the core promoter sequence, namely, the TATA box, in the case of the protein coding genes of the eukaryotes. Now, the job of TF2F is actually to target the RNA polymerase to this DAB complex, which are actually formed, and once the RNA polymerase 2F complex comes and binds here, then the TF2H, which itself consists of a multi-subunit complex, comes and joins the party, and TF2H actually has 2 major activities, namely, it has a…, it is a helicase activity and the TF2H actually modulates the helicase activity.

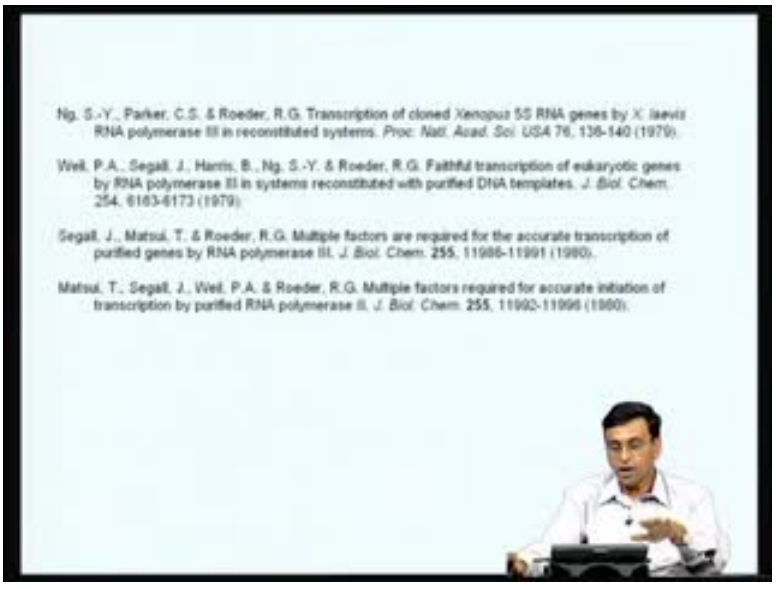
These factors then come and join the pre-initiation complex, and when once all these entire complex is assembled in and around the TATA box, and if ribonucleotides are present in the cell, and if ATP is present in the cell, then the TF2H actually phosphorylates this C-terminal domain of RNA polymerase 2; and this phosphorylation of CT domain is the actual signal that RNA polymerase can now leave the promoter, and then go and start transcription of the RNA polymerase 2.

So, you can see now how complex the whole situation is; just to transcribe one gene, you require close to some 40 to 50 different polypeptides, maybe 60 polypeptides. You require RNA polymerase 2, which itself is a multi-subunit complex. You require TF2A, B, D, E, and F. Each of them, again, is a multi-subunit protein complex and assembly of the all the subunits in and around the TATA box results in the formation of what is called, say, pre-initiation complex, and this pre-initiation complex is actually responsible for initiation of transcription in eukaryotes.

(Refer Slide Time: 51:26)

So, what I actually done in the next few slides is actually give you some of the original research articles, starting way back from 1960 all the way down, and these are all very some of the keys experiments which are actually done, to actually, to demonstrate how the eukaryotic RNA polymerase actually function, how the purification of RNA polymerase were studied, and how the general transcription factors or the key factors required for RNA polymerase 2 transcription were identified.

(Refer Slide Time: 51:53)



These key original articles if you read, you now, you get actually historical perspective of how difficult it was to study and understand the transcription initiation process in eukaryotes, and it has not been a breezy road to some of these studies, and lot of effort as actually gone into some of these things, and a number of groups have actually contributed to understand transcription initiation eukaryotes.

If you, some of these are all what I call as a very important research articles, original research articles. They have actually laid foundation for our basic understanding of eukaryotic transcription initiation. So, with this, I will now close this particular lecture, and in the next subsequent lectures we are going to study little bit more details about how actually is transcription initiation takes place; how each of the general transcription factors function, and then, we will move on to other regions, the distal promoter elements in the case of eukaryotes, and what happens there, and how other transcription factors, transcription activators, interact with RNA polymerase 2.

In the subsequent classes, I am going to tell you, so far, I had discussed as if DNA is a naked template, but I am going to tell you that DNA is not naked in eukaryote. It is actually present as a chromatin. So, all this RNA polymerase and general transcription factors that we have discussed so far, they have to actually activate transcription not from a naked DNA template, but they actually activate transcription from a chromatin template; and how actually these transcription factors was able to recognize chromatin, and then remove histones from the promoter regions, and then make RNA polymerase go and bind and activate transcription in the eukaryotic promoters.

This is what we are going to discuss the subsequent classes.

Thank you.