Eukaryotic Gene Expression

Prof. P. N. Rangarajan

Department of Biochemistry

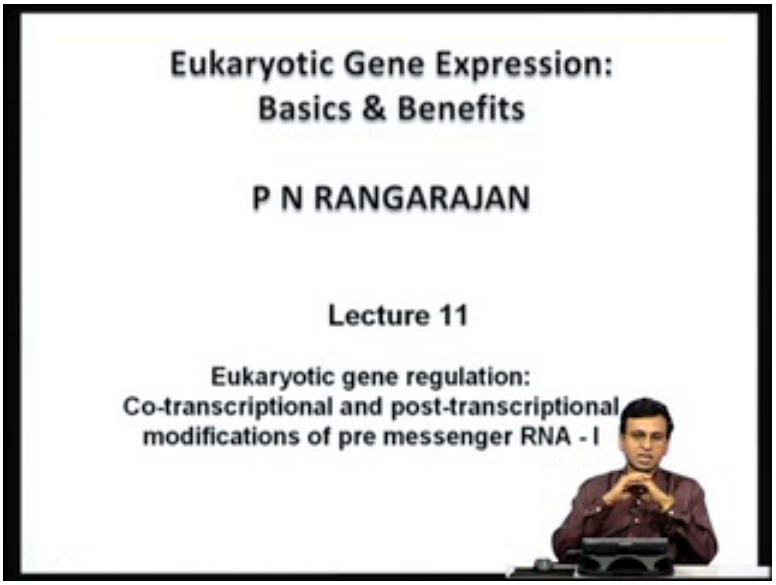
Indian Institute of Science, Bangalore

Model No. # 04

Lecture No. # 11

Eukaryotic Gene Regulation: Co-transcriptional and Post-transcriptional Modifications of Pre messenger RNA - I

(Refer Slide Time: 0016)



Welcome to this series of lectures on eukaryotic gene expression and basics and benefits. Today, we are going to be discussing about RNA processing; that is, co-transcriptional and post-transcriptional modifications of pre messenger RNA.

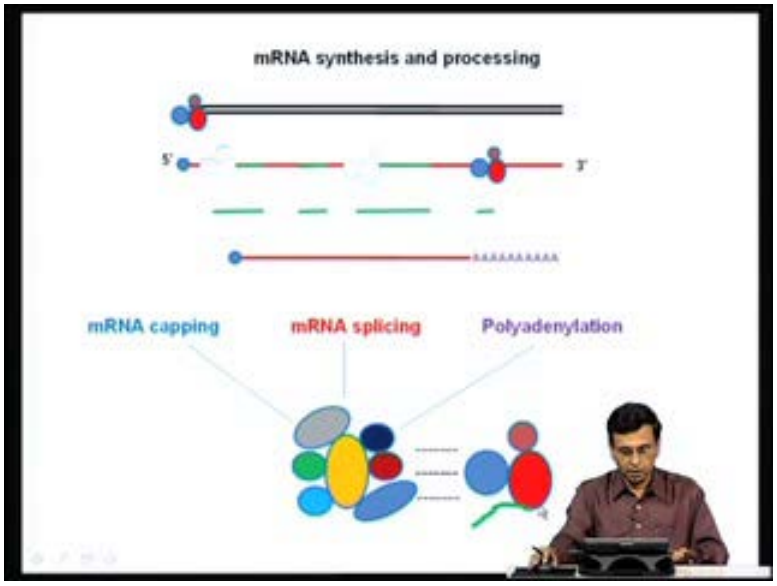
So far, the last ten lectures, the focus has been primarily on initiation of transcription. What we now discuss in the next two classes is, once transcription is initiated, how does the RNA polymerase transcribe the entire gene, how does the transcription elongation takes place and as the transcription elongation takes place, what kind of post-transcriptional or more appropriately co-transcriptional modifications that take place for the pre messenger RNA.

We are all aware that in the eukaryotes, when the pre messenger RNA is processed, after transcription or along with transcription, the transcript undergoes a number of post-transcriptional or co-transcriptional changes such as a cap is added at the five prime end, then the introns are spliced out and then a poly-A tail is added at the three prime end.

So, what we will discuss in the next two lectures is that how all these processes take place, and more importantly, the focus is primarily going to be to discuss the role of RNA polymerase. See, there is a misconception that the function of RNA polymerase II is only to be primarily at the level of transcription initiation.

But at the end of these two lectures, the point I would like to emphasize is that the RNA polymerase II plays a very important role, not only in initiation of transcription, but in the elongation as well as in other modifications of pre messenger RNA namely, splicing, polyadenylation as well as mRNA capping.

(Refer Slide Time: 02:19)



So, for all these processes, RNA polymerase II plays a very important role. So, the crux is going to be, discuss how RNA polymerase II is essential not only for transcription initiation and transcription elongation, but also for the other pre mRNA processing mechanisms that operate in the eukaryotic cells.

Now, let us discuss now, so far, the focus has been only on the initiation of transcription. I have depicted in this cartoon, the pre-initiation complex as a small, a multi subunit

complex here. So, far what we have discussed is that the RNA polymerase binds to specific promoter regions with the help of general transcription factors and when the RNA polymerase binds, it moves along, which is called as transcription elongation.

And as the RNA polymerase II moves along, you get transcription elongation and once the transcription elongation takes place, the pre messenger RNA, that is synthesized undergoes a number of modifications and there are three important modifications, which I have listed here; there are also other modifications, we will not discuss about. These are the major mechanisms that operate in living cells.

The first major co-transcriptional or post-transcriptional modification that the pre messenger RNA undergoes is what is called as five prime capping or mRNA capping. Then as we know, all the eukaryotic messenger RNAs contain introns as well as exons. I have depicted the introns in the green colour here and exons in the red colour.

So, the introns are spliced out, in is which is what is say, mRNA splicing and the exons are joined together to get a messenger RNA and finally, at the three prime end of this messenger RNA, a poly-A tail is added, which is what is called as a polyadenylation.

So, in the next two classes today, and the the this class and the next class, we will discuss how does mRNA capping, mRNA splicing and polyadenylation takes place and how exactly RNA polymerase II actively participates in all these three processes as well. So, the role of RNA polymerase II is confined not only to transcription initiation, but also for all these other post-transcriptional, co-transcriptional processing mechanisms.

Now, each one of these processes namely capping, splicing, polyadenylation again is carried out by a multi protein complexes. So, you the point I am trying to mention is that you look at by the time the RNA comes out of a nucleus, you can see how many multi protein complexes are involved in transcription initiation, transcription elongation and these three processes. So, a number of protein functions have to come together and active play their role in a very coordinative manner before a mature messenger RNA comes out of the nucleus and gets transcribed in the cytoplasm.

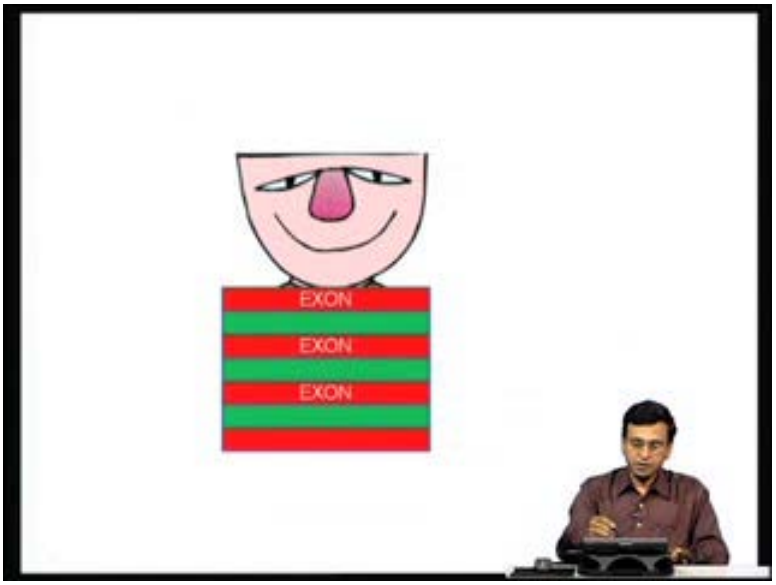
What is very important, what we are going to discuss today is that all these multi protein complexes, which are involved in this capping, splicing and polyadenylation, they

extensively interact with components of the preinitiation complex, especially the RNA polymerase II.

So, the focus today is going to be on how RNA polymerase II interacts with the multi protein complexes or machineries involved in mRNA capping, splicing and polyadenylation and how these interactions are very important for a regulation of eukaryotic gene expression.

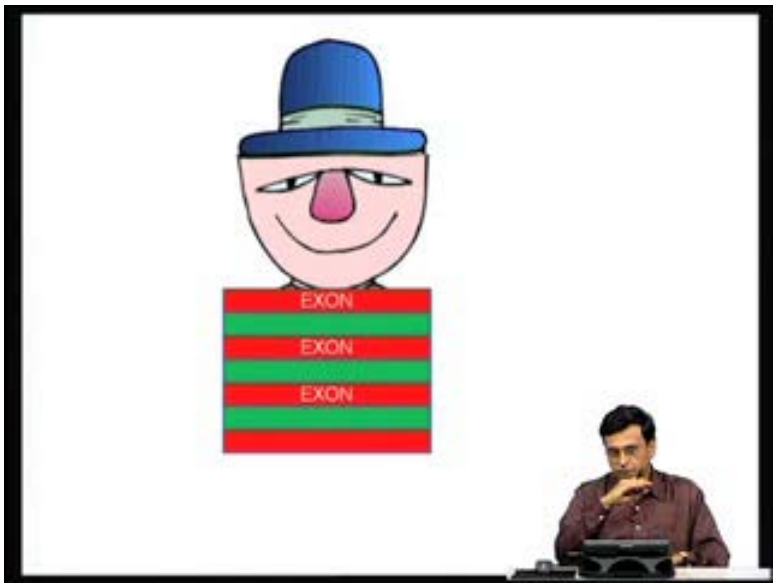
The other important point I would like to stress is that among the components of the preinitiation complex or even in the RNA polymerase II, the RNA polymerase C-terminal domain, what we call as RNA polymerase CTD, plays a very, very important role in all these three processes. So, we are going to discuss extensively what exactly is the role of RNA polymerase CTD in mRNA capping, mRNA splicing and polyadenylation.

(Refer Slide Time: 05:35)

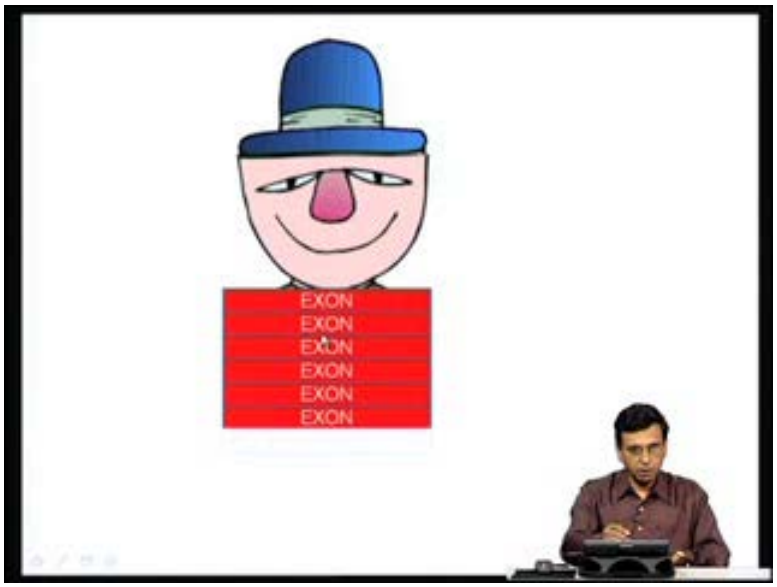


So, I just depict a cartoon here, for you to get an overall picture of what are you going to discuss today. I depicted this cartoon as the pre messenger RNA, which consists of exons, which I have shown in red and introns, which I have shown in green.

(Refer Slide Time: 05:48)

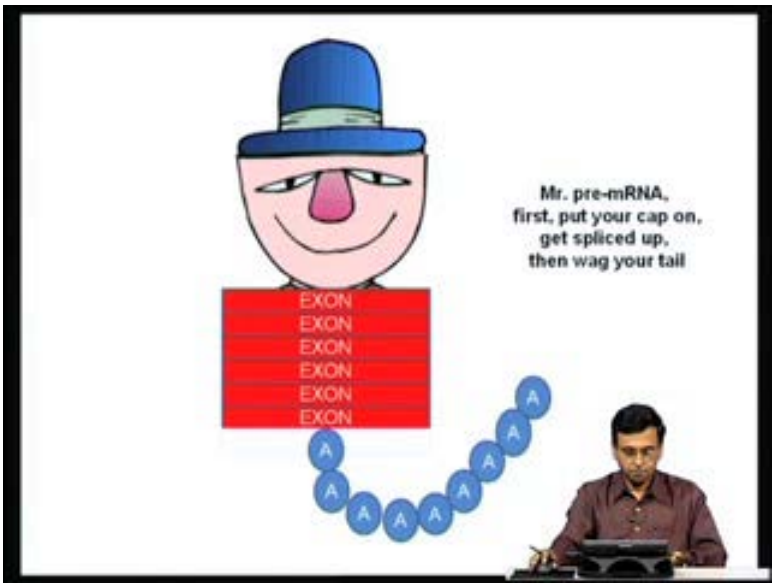


(Refer Slide Time: 06:08)



The first mechanism that I am going to discuss is that as the transcript is generated, as the RNA polymerase II leaves the promoters and starts elongation of the transcription, the first thing that happens is that a cap is added. So, I have just depicted in the form of a cap. So, we will discuss what exactly mRNA cap is and once an mRNA cap is added, the green colour - the introns are spliced out and the exons are joined together; that is what I have depicted here.

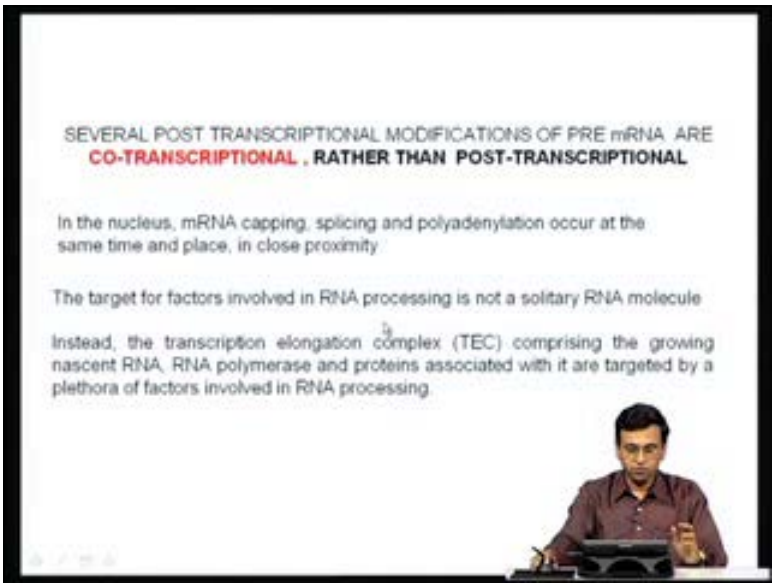
(Refer Slide Time: 06:15)



And once the capping and splicing takes place, at the end, when the RNA polymerase releases three prime end of the thing, a series of adenine residues are added and that is what I have shown here; this is what is called as a polyadenylation. And these number of A residues that are added ranges anywhere from 100 to 200 nucleotides are added at the end of the three prime end, which is the polyadenylation processes.

So, let us now discuss mister mRNA, how will you get your cap, how are you get capped and how you get spliced and how will you get polyadenylated, how a tail is added at the end. These are the three major mechanisms, we are going to discuss.

(Refer Slide Time: 06:48)



Now, all these three major modifications that a pre messenger RNA undergoes in a in about previously used to be described as post-transcriptional modifications. So, many of this splicing, capping and polyadenylation are often described as post-transcriptional modifications of pre messenger RNA, but later studies clearly reveal that many of these changes that take place in the messenger RNA are rather, co-transcriptional rather than post-transcriptional.

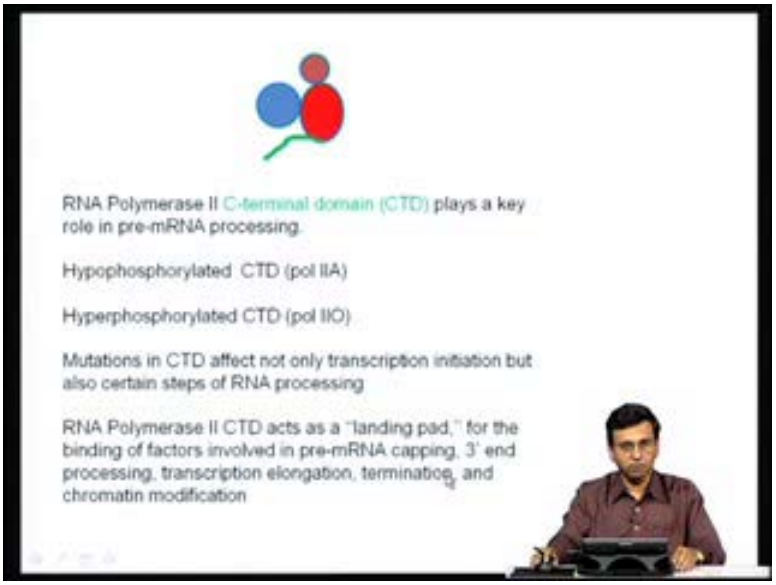
You know, what do you mean by this? What you mean by this is that as the transcription takes place, these modifications also take place simultaneously and the capping or splicing or polyadenylation does not wait till the entire messenger RNA is transcribed in the vein, but they simultaneously happen as the transcription proceeds.

So, all these events are primarily co-transcriptional rather than post-transcriptional, which is a very, very important. So, although we have discussed or in textbooks we discussed each one of them as a separate event, first we study transcription initiation, then we study transcription elongation, then we study splicing, capping and polyadenylation, inside the nucleus all these three process occur simultaneously in very close proximity, at the same time and at the same place.

So, in the nucleus, mRNA capping, splicing and polyadenylation occur at the same time, same place and in very close proximity. The target for factors involved in RNA's

processing is not a solitary RNA molecule. Remember all these processes, target not just RNA, but a multi protein complex or a multi ribonucleoprotein complex and which is what is called as the transcription elongation complex. This transcription elongation complex abbreviated TEC is the target for all these machineries and these comprises of the growing nascent RNA molecule. The RNA polymerase and the proteins associated with it, all these are targeted by the splicing machinery, capping machinery and polyadenylation machinery to finally bring out a mature messenger RNA molecule.

(Refer Slide Time: 08:48)



Now, the point I am going to emphasize and the point I am going to discuss very intensively in today's class, is the role of RNA polymerase C-terminal domain in pre mRNA processing. We all have discussed in the previous classes that this RNA polymerase CTD, which I have depicted in the green colour here, exists in two forms.

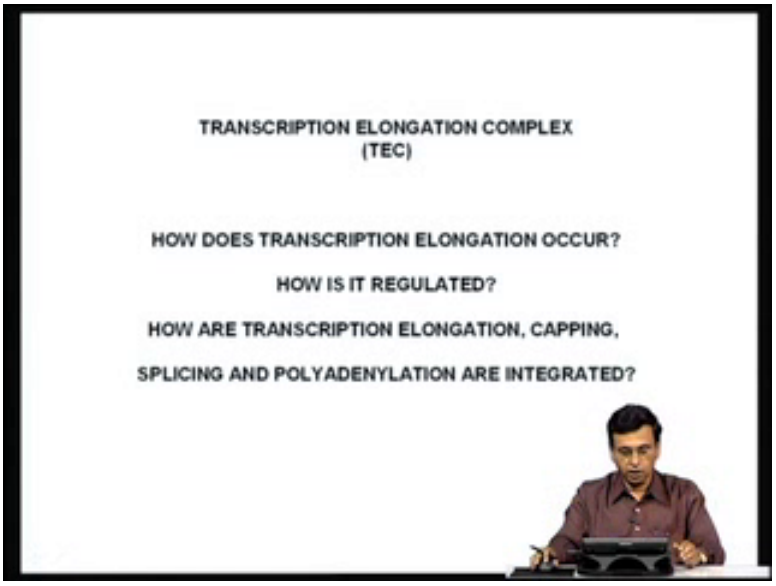
When all the factors are assembled, the general transcription factor TFIIH actually, phosphorylates this C-terminal domain of the RNA polymerase and this is the signal for RNA polymerase to leave the promoter and transcription elongation starts. So, basically the RNA polymerase II exists in two forms: a non-phosphorylated or a hypophosphorylated form, which is what is called as pol IIA and a hyperphosphorylated form, which is the pol IIO.

You know it turns out in a very, very important observation that was made is that when you make mutations in this C-terminal domain of RNA polymerase, especially those residues, which undergo phosphorylation, these mutations affect not only transcription initiation, but they also affect certain steps of RNA processing.

This was a very, very important observation that was made clearly indicated that the C-terminal domain of RNA polymerase has a role not only in transcription initiation, but may also play a very important role in other post-transcriptional, co-transcriptional events as well. So, the RNA polymerase II C-terminal domain acts as a landing pad for the binding of factors involved in pre-mRNA capping, three prime processing, transcription, elongation termination and chromatin modification.

And this is what we are going to discuss in this class and the subsequent class today. How the RNA polymerase II actually acts like a launching pad so that the capping factors, the splicing factors and polyadenylation factors can come there and do their job so that the final mature RNA comes out of the nucleus.

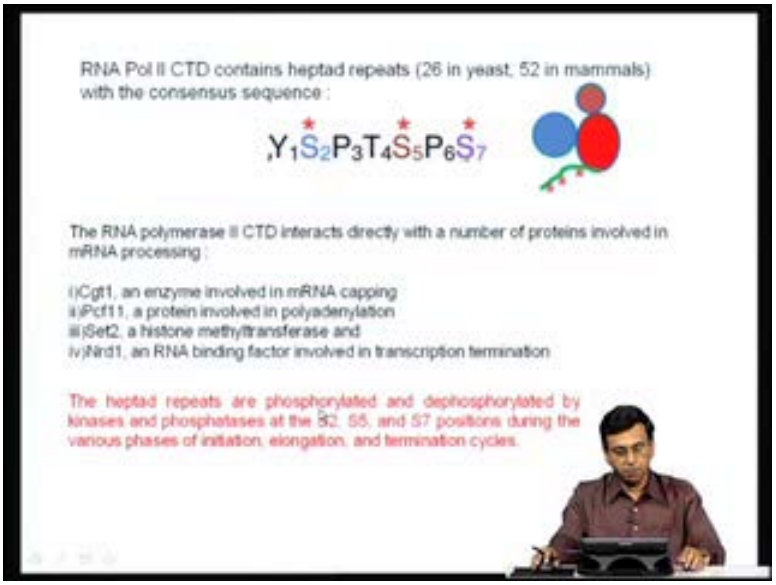
(Refer Slide Time: 10:40)



So, the focus of today's class is going to be on discussing extensively the transcription elongation complex. How does actually the transcription elongation occur and how it is regulated; how are transcription elongation, capping, splicing and polyadenylations are

integrated and how they are coordinately regulated so that all these modifications can simultaneously take place along with the transcription elongation.

(Refer Slide Time: 11:02)



Now, remember the RNA polymerase II CTD consists of heptad repeats, which I have repeatedly shown here with serine center space between the other residues and there are about twenty six such repeats in the yeast cells, whereas there are about fifty two repeats, such repeats in mammalian cells and this is the consensus sequence for these heptad repeats and the CTD, I have depicted as a green tail for the RNA polymerase of the pre-initiation complex. What has been observed is that the RNA polymerase II CTD interacts directly with a number of proteins involved in the mRNA processing.

So, this was the first evidence, which clearly indicated that in addition to a role in transcription initiation and transcription elongation, the RNA polymerase II CTD may play a very important role in other subsequent steps of RNA processing as well. For example, a protein called CGT one which is actually involved in mRNA capping has been shown to interact with RNA polymerase CTD. Similarly, a protein called PCF 11 involved in polyadenylation of mRNA also interacts with the CTD of RNA polymerase II and similarly, I have given couple of more examples as well.

So, these kinds of examples, where you take the RNA polymerase CTD and then do what is called as a coimmunoprecipitation experiments to see what kind of proteins can be

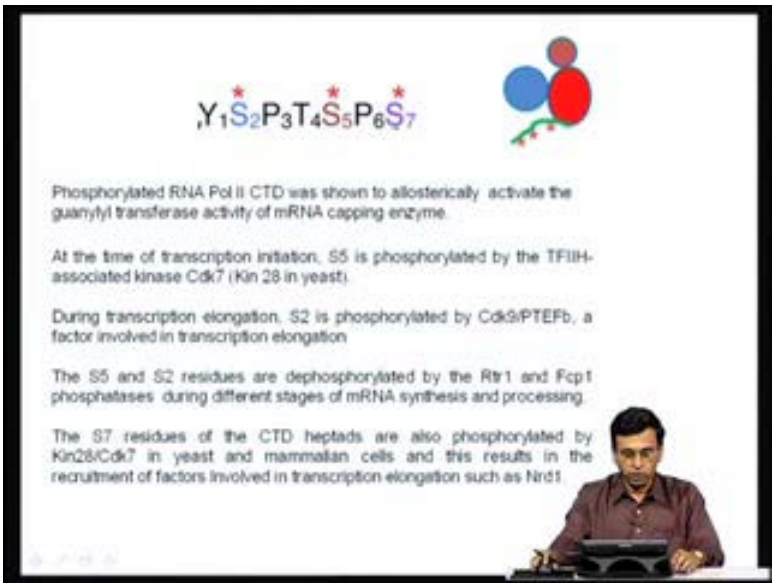
coimmunoprecipitated along with the RNA polymerase II CTD or what kind of proteins can be cross linked to the RNA polymerase CTD and then mutation analysis of CTD and see, if you mutate specific serine residues of CTD, what are all the processes that get affected.

All these studies clearly indicated that the RNA polymerase II CTD not only has a role in transcription initiation, but may play important role in all the subsequent processes as well. So, the serine residues which undergo phosphorylation from the RNA polymerase II CTD are the serine 2, serine 5 and serine 7; I have shown it in the asterisks here and heptad repeats are phosphorylated and dephosphorylated by kinases and phosphatases at these residues during the various phases of initiation, elongation and termination cycles.

So, remember in the previous classes, I just told you that once the pre-initiation complex is assembled, the final event before the train leaves the station or before the RNA polymerase start moving is the addition of phosphates or phosphorylation of C-terminal domain by TFIIH and that is the signal for the RNA polymerase to leave the initiation promoter and then start elongating.

But what we will now discuss is that this phosphorylation of CTD is a very dynamic process. Throughout the process of transcription elongation, these serine residues undergo phosphorylation and dephosphorylation and this facilitates interaction of the variety of proteins depending upon the phosphorylation status of the RNA polymerase II CTD.

(Refer Slide Time: 13:44)



Let us now discuss some of the evidences in favour of this particular hypothesis that RNA polymerase II CTD plays a very important role in RNA processing. For example, the RNA phosphorylated RNA polymerase II CTD was shown to allosterically activate guanylyl transferase activity of mRNA capping enzyme.

Now, just as we have a cell-free transcription system for studying transcription initiation, people have now established cell-free transcription system even to study mRNA capping, mRNA splicing and mRNA polyadenylation. mRNA polyadenylation So, using these kind of cell-free systems, people have been able to identify various components which are involved for or which essential for each one of these steps.

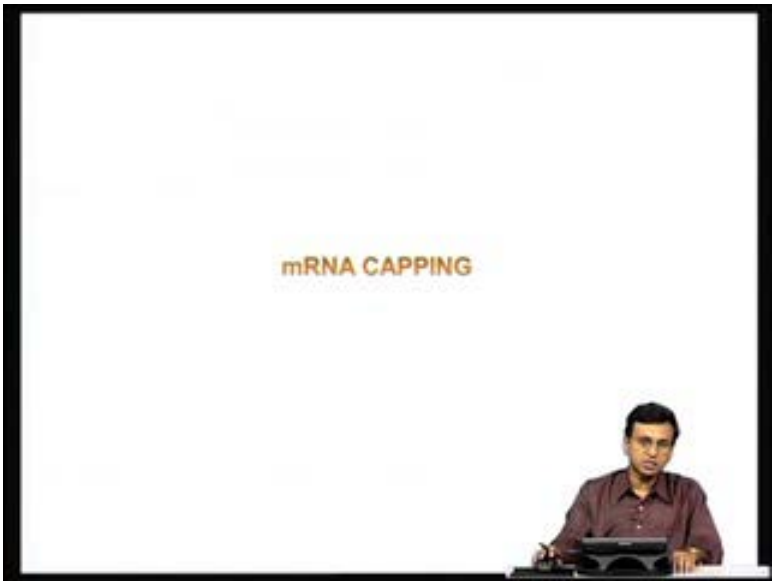
So, what has been shown is that if you now add phosphorylated RNA polymerase CTD in a invitro system, that is favourable for mRNA capping, the phosphorylated CTD actually was shown to stimulate the guanylation guanylation of the mRNA, which is basically the process involved in mRNA capping, clearly indicating that CTD phosphorylation plays a very important role in mRNA capping. Similarly, at the time of transcription initiation, when the serine five residue is phosphorylated by TFIIH, which is called kinase 28 in the case of yeast, whereas during transcription elongation serine 2 is phosphorylated by a different kinase and this factor is involved in transcription elongation.

So, transcription initiation involves phosphorylation of serine 5 by TFIIH, whereas transcription elongation involves phosphorylation of serine 2 of TFIID by another kinase. So, this is what I said different serine residues are phosphorylated during the process of transcription initiation, elongation, termination and so on and so forth. And turns out the serine 5 and serine 2 residues are dephosphorylated by specific phosphatases, Rtr 1 Fcp 1 during different stages of mRNA synthesis and processing. So, there are kinases which specifically phosphorylate specific residues of RNA polymerase CTD.

And at same time, there are actually phosphatases which remove these serines. So, it is a very dynamic equilibrium between phosphorylation, dephosphorylation of CTD, which actually plays a very important role in the assembly of either capping machinery or splicing machinery or polyadenylation machinery in the in the multi protein complexes.

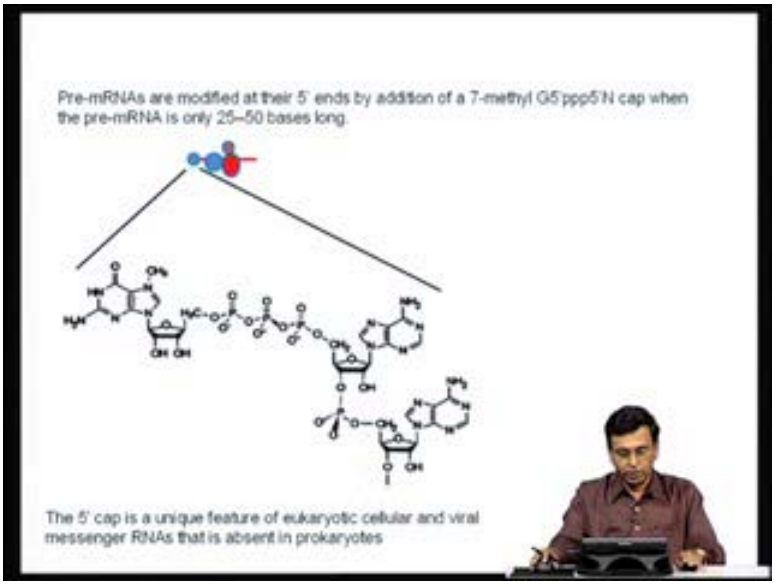
The serine 7 residues of CTD heptads are also phosphorylated by another kinase 28 in yeast and mammalian cells and this results in the recruitment of factors involved in transcription elongation such as Nrd 1. So, the point I am trying to make through the discussion here is that differential phosphorylation of the RNA polymerase CTD is essential either for transcription elongation or for mRNA capping or for mRNA splicing or polyadenylation. These differential phosphorylations may basically bring in these different factors and these processes are carried out accordingly.

(Refer Slide Time: 16:38)



So, let us now spend some time to understand what this mRNA is capping, what exactly happens during mRNA capping.

(Refer Slide Time: 16:45)



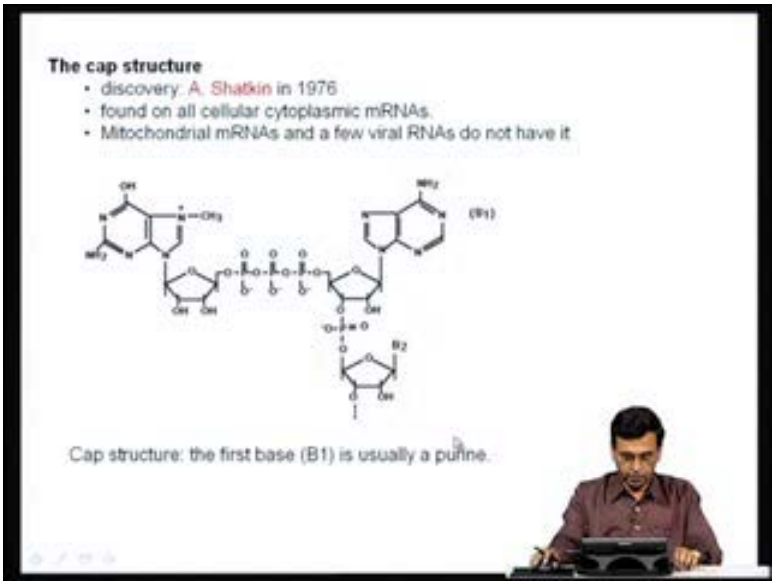
Now, as we know in eukaryotes, the pre-messenger RNAs are modified at their five prime end by the addition of what is called as a seven methyl guanosine cap, when the pre-mRNA is just twenty five to fifty bases long. So, as the RNA polymerase is phosphorylated by TFIIH and as it starts moving and about just about twenty five to fifty bases long RNA comes out of the RNA polymerase, immediately the capping is initiated. So, mRNA capping does not take place till the entire RNA transcription is over and RNA, the gene is completely transcribed, but even just when twenty to fifty bases RNA comes out, immediately the capping process is initiated.

So, what is this mRNA cap, which I have depicted in the blue dot here? This is what is the actual structure of an mRNA cap and this is the first base of the messenger RNA, which is usually a purine either adenine or guanine and this is the methyl guanosine, which is added to the five prime end of the messenger RNA and this is what is called as a mRNA capping.

You see the methyl group here and this is the guanine. So, the addition of the methyl guanine to the five prime end of mRNA is what is called as the methylation or mRNA capping. Remember, this five prime cap is a very, very unique feature of eukaryotic

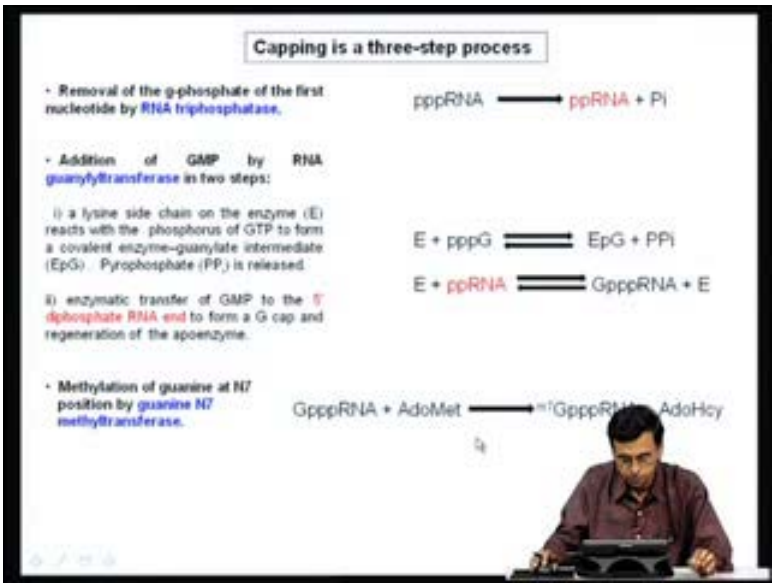
cellular RNAs and also, viruses which infect eukaryotic cells; not all viruses, but certain viruses and this kind of mRNA capping mechanism is absent in prokaryotes. So, it is a major difference between prokaryotic mRNA processing and eukaryotic mRNA processing.

(Refer Slide Time: 18:15)



This mRNA capping was first discovered by Aaron Shatkin in the year 1976 and soon, he showed that this mRNA cap is a common feature of almost all the cellular messenger RNAs and sometimes a few mitochondrial RNAs and certain few viral RNAs also are mR capped. So, this is the cap structure, which I have already shown here. This is the methyl guanosine which is linked by a phosphodiester bond to the first base of the messenger RNA and this first base is usually either an adenine or a guanine.

(Refer Slide Time: 18:48)



Let us now understand how exactly this capping takes place. What are the enzymes which are involved in mRNA capping? The mRNA capping takes place in at least three steps. So, there are three enzyme complexes which are involved; these are called an RNA triphosphatase, a guanylyltransferase and a guanine N7 methyltransferase.

So, let us just briefly see how exactly this capping mechanism takes place. The first step in the mRNA capping is the removal of the guanosine phosphate of the first nucleotide by RNA phosphatase. Remember, the first base is always a trinucleotide; the first nucleotide of the five prime end of N1 is a trinucleotide and the first step in the capping process is the removal of one of the phosphate so that it becomes a dinucleotide. So, pppRNA, becomes a triphosphate becomes a diphosphate with the removal of an inorganic phosphate.

Now, this is a very I have depicted this ppRNA in red colour because this has a very important role, which we will discuss in the next slide. What is the next step? The next step is the addition of GMP by RNA guanylyltransferase.

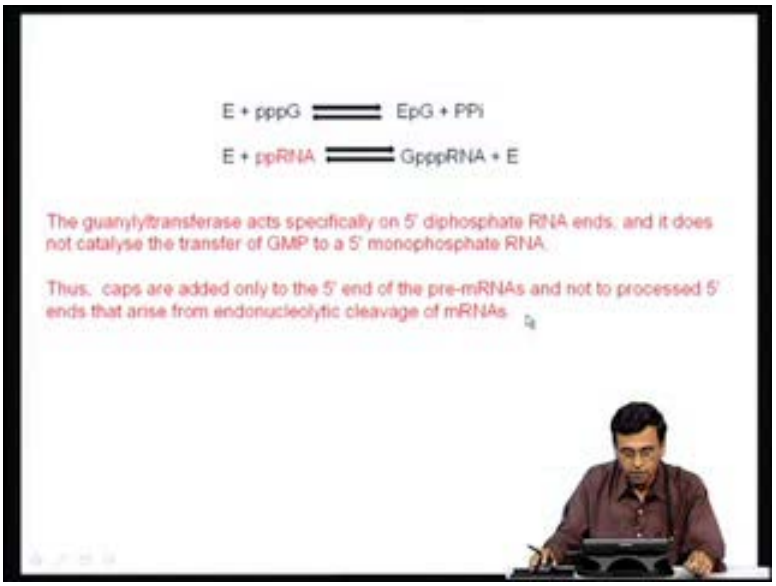
So, what happens, the lysine side chain on the enzyme, the guanylyltransferase reacts with the phosphorous of the GTP to first form a covalent enzyme guanylate intermediate; this is what I have shown here. this is a This is a two step reaction; first, the enzyme reacts with the GTP to form an enzyme guanylate complex and the pyrophosphate is

released in the process and in the next step, this enzyme transfers the GMP to this diphosphate RNA to form a trimethylguanosine cap and in the process, the apoenzyme gets regenerated.

So, guanylation takes place in two steps and the enzyme that is involved here is the guanylyltransferase. In the final step, the guanosine which is added to the five prime end then gets methylated by an enzyme called guanosine N methyltransferase and here adenosine S adenosylmethionine acts as a methyl donor and the methyl group is transferred from S adenosylmethionine to this guanylated RNA so that you get methylated guanosine at the five prime end.

In the process, adenosylmethionine get converted into S adenosylhomocysteine. Remember, in almost all the methylation reactions in biological system, S adenosylmethionine acts as a methyl donor and it plays a very, very important role in a number of methylation reactions in biological systems.

(Refer Slide Time: 21:19)



So, this is how a mRNA cap is formed in the eukaryotic messengers. Now, I told you that I have depicted this diphosphate RNA in the red and it has a very important reason for that. The guanylyltransferase acts specifically on the five primes diphosphate RNA and it does not catalyze the transfer of GMP to a monophosphate RNA. It is very, very important to remember that guanylation takes place only for a diphosphate RNA, but not

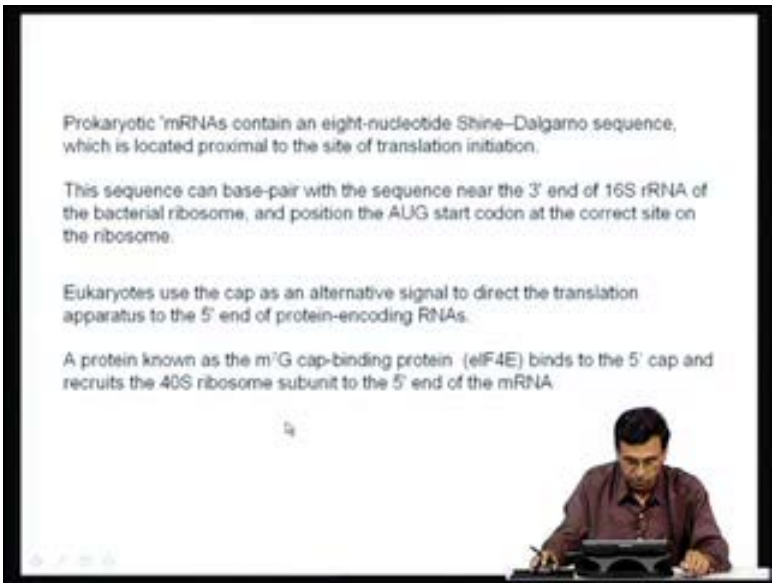
for a monophophate RNA. Why is this important? It is important because the caps are added only to the five prime end of pre-messenger RNAs, but not to processed five prime ends that arise from the endonucleolytic cleavage of RNA.

You know there are a number of cellular messenger RNAs which not only are synthesized, they are also being degraded. So, when these mRNAs are degraded by endonucleases, you generate five prime ends, but these five prime ends contain a monophosphate.

So, to distinguish between these RNAs, which are generated by endonucleolytic cleavage of pre-synthesized RNAs from a freshly made RNA, this kind of a feature has been introduced. So, only the pre-messenger RNA contains a diphosphate at the five prime end and only this diphosphate is substrate for guanylation, whereas all the five prime ends which are generated by a RNA degradation contain only a monophosphate; therefore, they are not capped.

So, you see how interestingly, the biological system has evolved a mechanism to distinguish between capping a pre-messenger RNA molecule at the five prime end versus capping five prime ends of RNAs which have been generated by RNA degradation. So, only the pre-messenger RNAs, which are freshly synthesized and which contain a diphosphate in the five prime end are capped, whereas those messenger RNAs, which have been, which contain five prime ends which are generated by endonucleolytic cleavage of RNA ribonucleases, they contain monophasphate and they will not undergo capping.

(Refer Slide Time: 23:02)

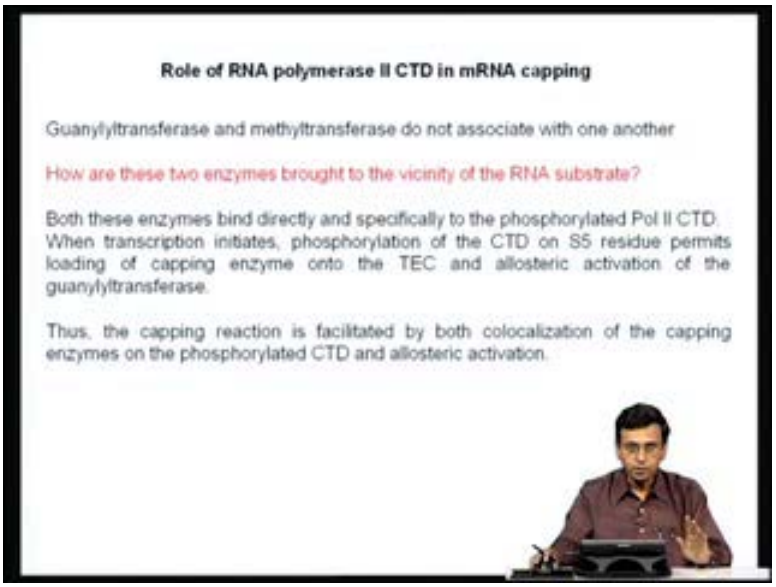


Now, I told you that this mRNA capping is a very, very unique feature of eukaryotes, and the whereas in the case of prokaryotes, they contain instead, an eight-nucleotide sequence called a Shine-Dalgarno sequence, which is located just about the five prime to the translation start site and this is what is recognized by the prokaryotic ribosomes and then translation is initiated from the first A2G codon, in the case of prokaryotes and this Shine-Dalgarno sequence actually can base-pair the sequence near the three prime end of the 16S ribosomal RNA of the bacterial ribosome and then translation initiates from the first AUG codon, it encounters. So, the Shine-Dalgarno sequence is involved in recognition by ribosome, in case of prokaryotes and accurate initiation of translation, whereas in the case of eukaryotes, this function is taken over by the mRNA cap structure.

So, the mRNA cap is actually recognized by the protein translating machinery of the eukaryotic cells and a protein known as 7 methyl guanosine cap binding protein, usually known as eukaryotic initiation factor 4E actually binds the five prime cap and results in the recruitment of 40S ribosome subunit to the five prime end of the RNA and therefore, and that is how translation initiated.

So, the five prime cap plays a very important role in initiation of translation because this five prime cap structure is recognized by one of the components of the translation machinery and this results in bringing the 40S ribosome to the five prime end and then translation is initiated.

(Refer Slide Time: 24:26)



So, what we will do now is, so far, we have very briefly discussed what exactly an mRNA cap is. It is nothing, but addition of five methylguanosine to the five prime end of the pre-messenger RNA by a three step enzymatic reaction; that is what is mRNA cap. Now, let us now discuss what exactly the role of RNA polymerase II is in mRNA capping.

It turns out I told you there are about three enzymes involved in mRNA capping of which the second and third enzyme, the guanylyltransferase and methyltransferase, they do not actually interact with each other; there is no association. Then the question comes how are these two enzymes brought in close vicinity because once the guanylyltransferase adds the guanosine to the five prime end, the next step is the methyltransferase should methylate the guanosine.

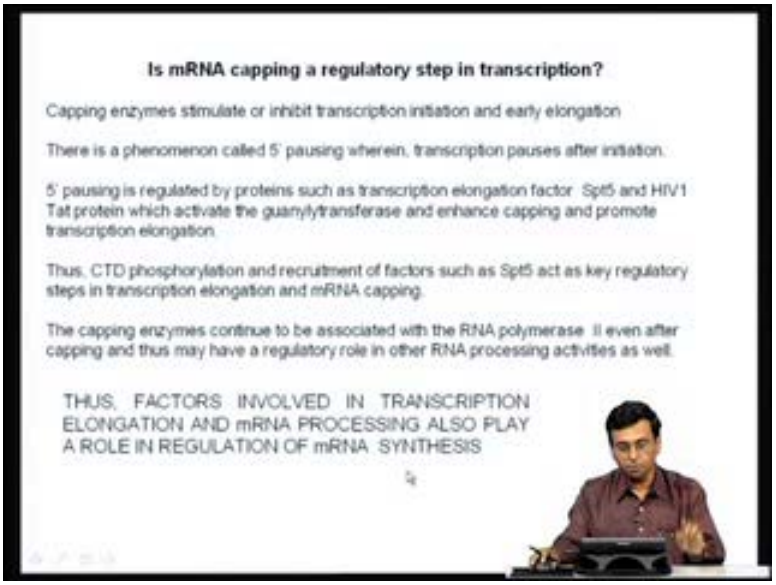
So, how do these two reactions take place together, when these two proteins do not interact? It turns out these enzymes have to be brought in the vicinity of the RNA substrate so that these two reactions can take one after another. It turns out both these enzymes namely the guanylyltransferase and methyltransferase bind directly and specifically to the phosphorylated polymerase II of CTD.

So, when transcription initiates, phosphorylation of the CTD on the serine five residue permits loading of the capping enzyme onto the transcription elongation complex as well as allosteric activation of the guanylyltransferase.

So, you can see the phosphorylation of a serine five residue not only serves as a signal for transcription elongation, the serine five residue also serves as a signal for the recruitment of the capping enzyme and also, allosteric activation of the guanylyltransferase activity of the capping enzyme.

So, the capping reaction is facilitated both by colocalization of the capping enzymes on the phosphorylation of CTD as well as allosteric activation. So, this is what is important, the CTD of the RNA polymerase not only serves as a launching pad for bringing the capping machinery to the vicinity of the RNA, but it also allosterically modifies this enzyme so that the active enzyme activity is enhanced.

(Refer Slide Time: 26:26)



So, there is a very active role of RNA polymerase CTD in the mRNA capping process. So, the next question one would ask is that is mRNA capping really a regulatory step. Is this process really regulated in eukaryotic cells?

It turns out capping enzymes also can stimulate or inhibit transcription initiation and early elongation and what is the evidence. It turns out there is a phenomenon called five prime pausing, wherein transcription actually pauses after initiation.

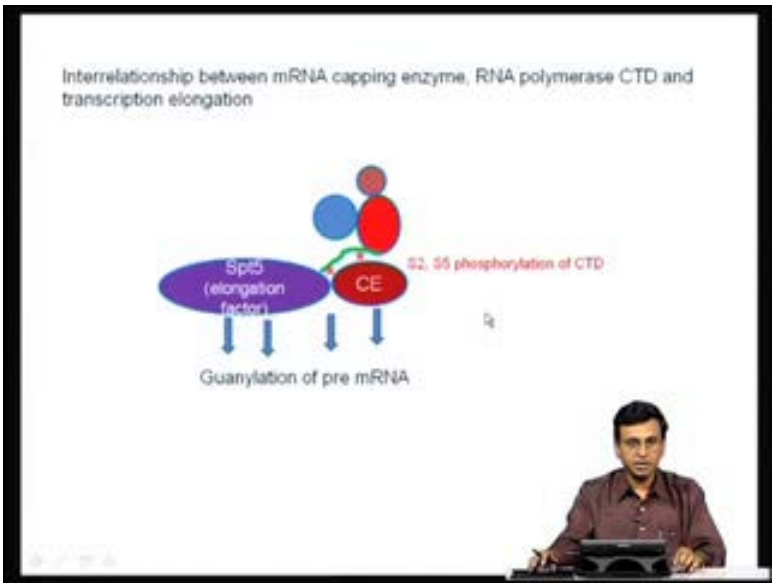
In fact, the rate of elongation of transcription varies for different genes and there are many instances, where the RNA polymerase after initiation actually pauses; it does not elongate further and this and this five prime pausing is regulated by proteins such as transcription elongation factor Spt5 as well as certain viral transcription factors like HIV1 tat protein, which actually activate the guanylyltransferase activity of the capping enzyme and therefore, enhance capping and for promote transcription elongation

So, many transcription elongation factors, by enhancing the guanylyltransferase activity of the capping enzymes, actually, enhance the rate of transcription elongation. So, the CTD phosphorylation recruitment of factors such as Spt5 act as the key regulatory steps in transcription elongation in mRNA capping

So, not only there is a regulation of regulation at the level of transcription initiation by enhancing the rate of transcription elongation by factors like the capping enzyme and transcription elongation factors, the mRNA synthesis is regulated even at the level of transcription elongation as well as mRNA capping.

So, the capping enzymes continue to be associated with the RNA polymerase II, even after capping and it is suggested that these capping enzymes may play a role not only in capping, but also in subsequent mRNA processing events. So, the take home message, I want to give you from this is that the factors involved in transcription elongation and mRNA processing play a very important role in mRNA synthesis in addition to factors involved in transcription initiation; this is the message.

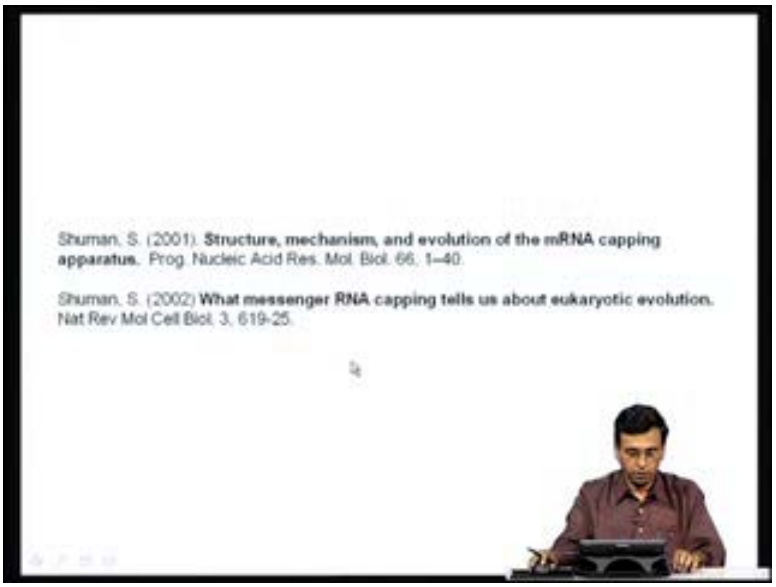
(Refer Slide Time: 28:13)



So, the summary of what I have told so far is, I have depicted in the form of a cartoon. There is a very strong interrelationship between the mRNA capping enzyme, RNA polymerase CTD and factors involved in transcription elongation. The RNA polymerase II, the final step in transcription initiation is the phosphorylation of the serine two and serine five residues of the CTD of the RNA polymerase II and these actually, serve as a signal for the recruitment of the capping enzyme.

And when capping enzyme binds this, not only this is involved in recruitment, it also allosterically modifies the guanylyltransferase activated with capping enzyme so that it can efficiently guanylate the pre messenger RNA, promote mRNA capping. And there are also other factors like those involved in transcription elongation such as a Spt5 factors which also interact with the capping enzyme and they also enhance the gualynation of pre-messenger RNA. As a result of all these things, the rate of transcription elongation is initiated.

(Refer Slide Time: 29:20)

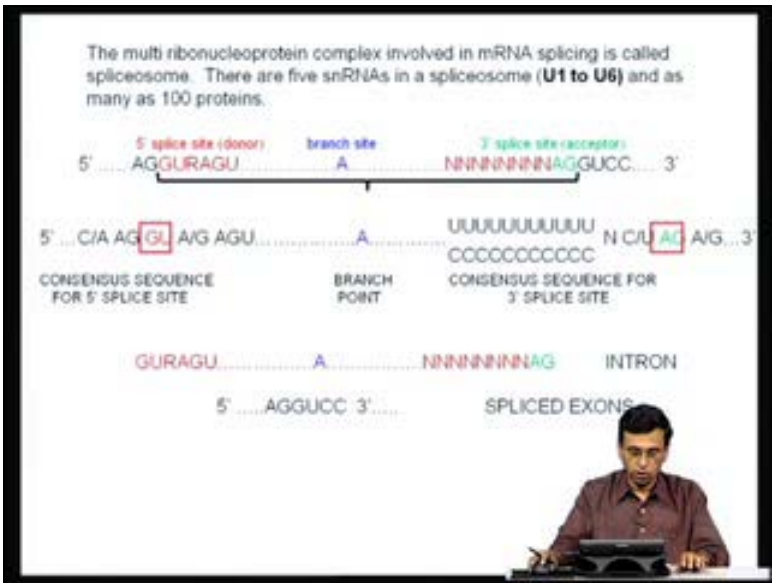


So, regulation can take place not only at the initiation level, but the rate of elongation can also be influenced by factors involved in RNA splicing and transcription elongation. There are very nice review articles which very well discuss, especially in the early 2000s. These papers discuss very nicely about the process of mRNA capping as well as the evolution significance. So, some of you, who are interested can go through some of these papers.

(Refer Slide Time: 29:36)

Now, let us now switch over and then come to the next RNA processing mechanism namely, mRNA splicing. So, as the RNA as the RNA polymerase moves from transcription initiation site and proceeds immediately after the addition of 20 to 50 bases, the cap structure is added and the next immediate event that has to take place is that splicing of the introns and joining of the exons.

(Refer Slide Time: 30:03)



Now, how does the interaction between splicing machinery and a RNA polymerase influence this particular RNA processing mechanism? I will not go and discuss in detail, the RNA splicing per say because that is a that that is a huge phenomenon by itself. So, I am going to just very briefly mention the basic aspects of mRNA splicing in this particular cartoon, just sufficient for you to understand and appreciate how RNA polymerase II regulates mRNA splicing mechanism.

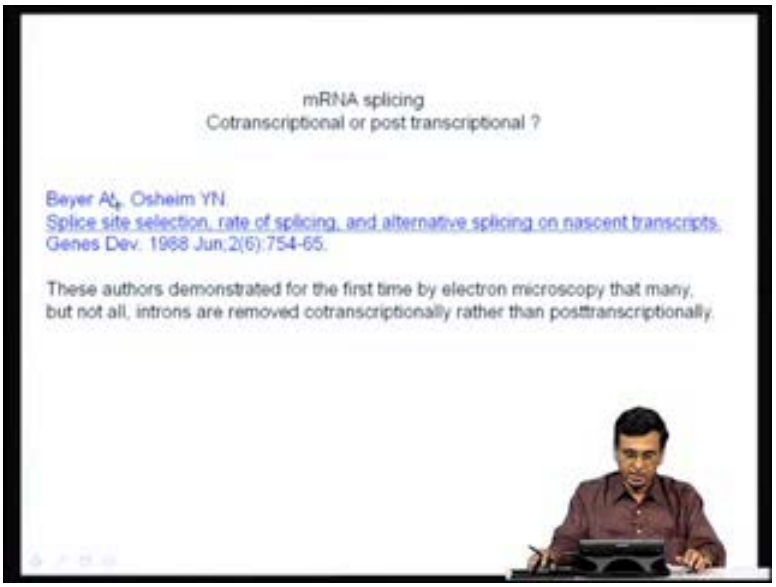
So, as you all know or you should study a little bit more of splicing before you proceed further. Now, mRNA splicing in eukaryotes is carried out by a multi ribonucleoprotein complex called spliceosome. And this spliceosome contains of five snRNAs; these are called U1 to U6 and also as many as hundred proteins. So, the splicing machinery again is a multi subunit complex or a multi protein complex containing not only RNA molecules, but also protein molecules.

So, the spliceosome is a big, is a huge ribonucleoprotein complex which takes plays a very important role in the mRNA splicing. How is splicing carried out? If this is the pre messenger RNA that is coming out and usually, there is what is called as a five prime splice site and a three splice site and usually, the most common five prime splice site is what is called as a GT or GU plus RNA and the three prime splice site, it is a AG. So, this is what serves as a five prime or the donor splice site and three prime splice or the acceptors splice site and in between, there is also what is called as a very important adenosine residue which is called as a branch site. These five prime splice site, branch site and three prime splice site are very essential for mRNA splicing.

So, in addition to this core GU sequence and AG sequence, by analyzing a number of eukaryotic messenger RNAs, people have identified that sequences flanking this GU residue as well as sequences flanking the AG residues also play a very important role in the recognition of the five prime and three prime splice of the RNA polymerase machinery, and a consensus sequence is accordingly drawn here and this is what is shown here.

So, the RNAs splicing machinery which consists of U1 to U6 snRNA and a number of proteins recognize this five prime splice site, branch site as well as three prime splice site and through a series of reactions bring about the splicing, which involves the removal of the intron between these five prime splice site and three prime splice site, and joining of the exons. So, that the two exons are spliced and you get the mature transcript. This is in a nutshell how messenger RNA splicing takes place in eukaryotes. We will not go into the exact mechanism by which the spliceosome carries out this reaction; that is not the crux of today’s talk.

(Refer Slide Time: 32:38)

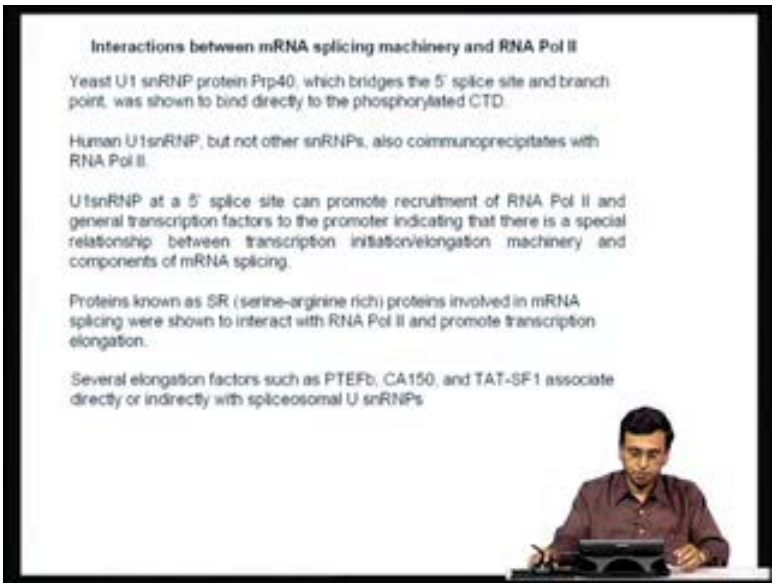


Now, the question is how does this what does this spliceosome do? Does the machinery, splicing machinery wait till the entire transcription is over and then, once the transcription is over, the RNAs undergo splicing or as the RNAs coming out of the d n a template, the splicing takes place simultaneously. So, is mRNA splicing co-transcriptional or post-transcriptional?

A very nice research work was done and is described in this paper, where they actually demonstrated using electron microscopy that many introns are actually removed co-transcriptionally rather than post-transcriptionally.

For reasons of copyright, I am not able to show of these electron micrograph pictures. So, I suggest you go and refer this paper in genes and development, where the authors have clearly shown electron micrographic pictures, which clearly shows that splicing, if not all, but at least majority of the splicing takes place as transcription proceeds. Therefore, splicing predominantly is co-transcriptional rather than post-transcriptional.

(Refer Slide Time: 33:40)



Now, what is the evidence to actually demonstrate that there are interactions between RNA polymerase II and the splicing machinery? If it is co-transcriptional and if both the transcriptional elongation and splicing have to take place simultaneously there has to be some kind of interactions between proteins involved in splicing as well as proteins associated with the RNA polymerase.

What are these associations? It turns out the U1snRNP, both in the case of yeast and human, actually, which is involved in a very important splicing reaction was shown to directly bind to the phosphorylated carboxy terminal domain of RNA polymerase II. So, one of the proteins, a ribonucleoprotein involved in splicing reaction directly interacts with the phosphorylated C-terminal domain of RNA polymerase II suggesting that there is an interaction between components of splicing machinery and the RNA polymerase II.

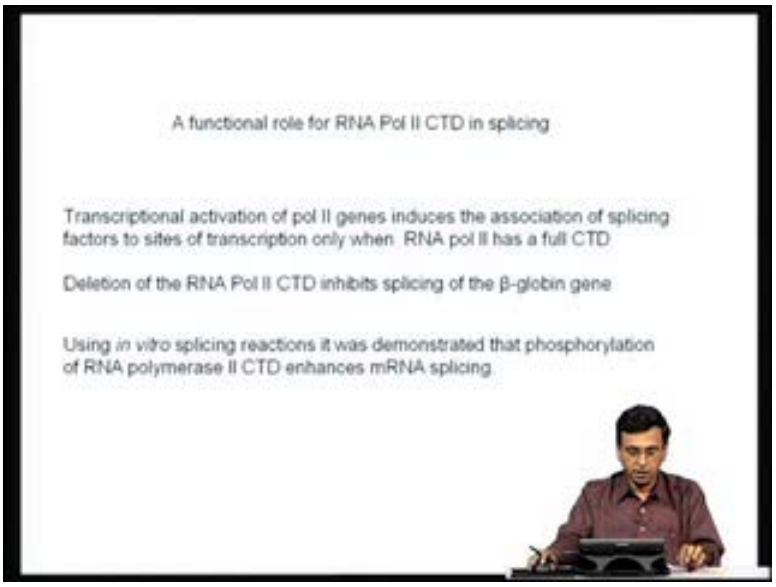
Similarly, the U1snRNP at a splice site cannot can promote recruitment of RNA polymerase II and general transcription factors to the promoter indicating that there is a special relationship between transcription initiation/elongation machinery as well as mRNA splicing.

Now, all our previous lectures, our focus had been on how activators and co-activators recruit the pre-initiation complex to the promoter and that is how they enhance the rate of transcription initiation. What has been found out, even a protein involved in mRNA

splicing can also enhance the rate of transcription initiation by enhancing the recruitment of RNA polymerase II clearly indicating that all these events are highly coordinated or highly integrated.

So, even a factor involved in splicing can enhance the recruitment of RNA polymerase II to the vicinity of the promoter and many proteins known as serine-arginine rich proteins that are usually called as SR proteins, which are again involved in mRNA splicing, again were shown to be involved in RNA polymerase II and promote transcription elongation. And several elongation factors such as PTEfb, CA150 and TAT-SF1 associate directly or indirectly with components of the spliceosomes. So, what I am trying to tell you with all these things is that protein components involved in splicing and protein complex involved in transcription initiation, transcription elongation, they all interact with each other and therefore, they promote one another; that is the crux of this entire discussion.

(Refer Slide Time: 35:56)



Now, just because I describe to you that there are proteins which interact with each other does not mean that it has a functional significance. So, we actually, I have to demonstrate this splicing factors actually play a very important role in the or RNA polymerase II really has an active functional role in mRNA splicing. How do you demonstrate this?

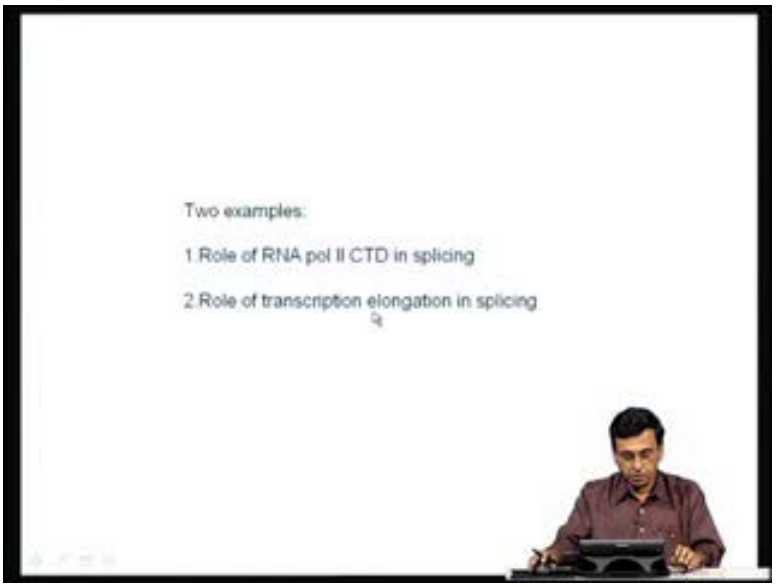
It turns out transcriptional activation of polymerase II genes induces the association of splicing factors to the sites of transcription only when RNA polymerase II has a full

carboxy terminal domain. So, if we take an RNA polymerase containing normal C-terminal domain and an RNA polymerase II having a truncated or a mutated C-terminal domain, only the RNA polymerase II, which contains a complete CTD can actually recruit actively splicing factor to the vicinity of transcription, whereas if you remove the CTD, splicing factor cannot be recruited to the vicinity of transcription clearly indicating that the CTD plays a very, very important role in the recruiting of the splicing machinery. And it turns out if you delete the RNA polymerase II CTD in the case of genes like beta-globin, splicing is inhibited clearly saying that the C-terminal domain of RNA polymerase II has a role, not only in transcriptional initiation and elongation as well as mRNA capping, it also has an important role in mRNA splicing.

And in fact, using in vitro splicing reactions, it was demonstrated that phosphorylation of RNA polymerase CTD actually enhances mRNA splicing. So, what you do? You have a cell-free transcription system, which in which you have all the factors required for an mRNA splicing.

Now, if you do the splicing reaction in the presence of RNA polymerase II, which is not phosphorylated or RNA polymerase II which which is phosphorylated in the CTD, the phosphorylated RNA polymerase II enhances, gives a better splicing, promotes better splicing reaction than the hypophosphorylated or a unphosphorylated CTD, clearly saying that phosphorylation of CTD actually enhances RNA splicing. So, these kinds of in vitro reactions clearly indicated that the pol II CTD has a very, very active role in mRNA splicing.

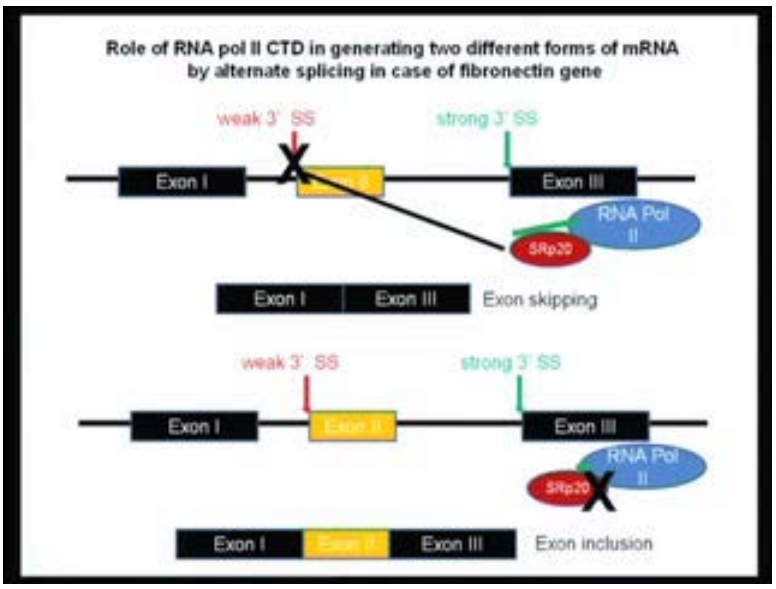
(Refer Slide Time: 37:56)



I want to give you just two examples to illustrate the point that how RNA polymerase II play plays a very important role in mRNA splicing.

One example, I would like to discuss with you is that how RNA polymerase II CTD plays a very important role in mRNA splicing. Then I will also tell you how transcription elongation itself plays a very important role in mRNA splicing.

(Refer Slide Time: 38:18)



So, what I will discuss I will give an example. I mean if you go into the literature and if you start looking research paper, there are a number of examples because people who are studying various gene systems have studied all these things and you can find a plethora of examples which illustrate that there is an active participation of RNA polymerase II in splicing machinery in a number of instances or a number of examples of individual genes.

I will only discuss one example here. Let us for example, see whether RNA polymerase II C-terminal domain plays a very important role in differential regulation in the case of fibronectin gene. What happens if the RNA polymerase II is normal and what happens the RNA polymerase II is mutated? This is a very, very interesting example. That is why I am quoting this here.

Now, let us assume that here is a gene which contains three exons - exon1, exon2 and exon3. Now, it has been shown, especially in the case of fibronectin gene, like when I discussed about the consensus sequences for splice sequence, I told you there is a GT and AG, but I have also told you that the sequences flanking the GT and flanking the AG, also play a very important role in the recognition of the donor splice site, in the acceptor splice site and based on the these things, a consensus sequence have been derived.

So, there are certain cases where they contain what is called as a weak three prime splice signals; that means they do not exactly conform to the consensus sequence. There are what are called as strong three prime splices, which means they exactly conform to the consensus sequence, which is very well recognized by the splicing machinery.

Now, here is an instance where, in the same gene, you have a weak splice three prime splice site as well as a strong three prime splice site. Now, let us see how does splicing take place when you have these two kinds of with two kinds of splice signals with differential affinities.

Now, it turns out when the RNA polymerase start transcribing this gene and when it pass the exon1, exon2, exon3 and if it has a normal CTD, there is a protein factor called SRp20, which actually interacts with the carboxyl terminal domain of the RNA

polymerase II and as a result of this interaction, the SRp20 prevents the splicing machinery from recognizing a weak three prime splice site.

Remember, I am repeating again. In this particular case, as the RNA polymerase transcribes the gene and when it passes the exon3, a protein factor called SRp20 interacts with the C-terminal domain of RNA polymerase and as a result of this interaction, the SRp20 prevents the splicing machinery from recognizing the weak three prime splice site and as a result, instead of, splicing this intron and joining exon1and exon2, the exon2 is skipped and exon1and exon3 are joined. So, this is what you get as a final product.

So, normally these two introns should have been removed so that exon1 should be joined to exon2, exon2 should be joined to exon3, but because of this interaction of SRp20 with the RNA polymerase II CTD and the masking of the weak three prime signal, instead of, joining exon2 with the exon1, exon1and exon3 are joined and this is what is called as a exon skipping.

So, what is the implication of this? That means the codons which are present in exon2 will not be represented in the final messenger RNA. Therefore, the protein sequence of this transcript is quite different than when all the 3 exons are together; that means one can generate two different kinds of transcript from the same gene by this particular mechanism.

It turns out, to prove that the RNA polymerase II CTD really has an important role in this process, what has been done is that if you now make cells express an RNA polymerase which instead of a normal C-terminal domain, which is pretty long contains a truncated C-terminal domain, then the SRp20 protein cannot interact with this truncated C-terminal domain and as a result, it cannot mask the weak three prime signal and therefore, the splicing machinery now recognizes three prime with the weak three prime signal as well and removes this intron and this intron joins exon1 to exon2 and then exon2 to exon3. Therefore, we get the exon is the exon2 is now included. So, you now get a different messenger RNA molecule.

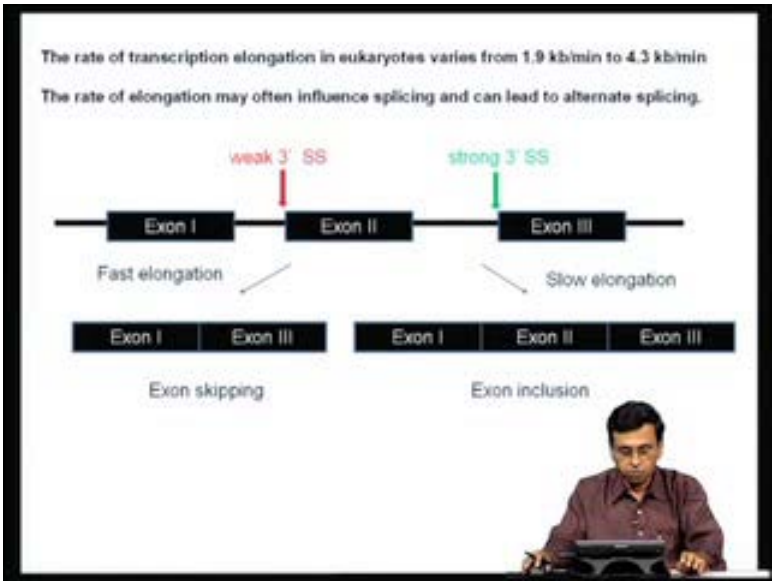
So, you can see how interaction of a protein factor with the RNA polymerase II is regulating splicing. So, by differential splicing through the involvement of RNA

polymerase II C-terminal domain, you can generate two different types of messenger RNAs: one which does not contain exon, one which contains a specific exon.

So, differential gene regulation or differential gene expression by generating two different messenger RNAs from a same gene is possible by interaction of a protein factor with the RNA polymerase II and by regulating the RNA polymerase by regulating the RNA splicing machinery.

So, you can see differential gene regulation can take place not only in the case of transcription initiation and elongation, it can also take place at the level of splicing and by these kinds of mechanism, you can generate more than one RNA from the same gene. So, this is one example to clearly indicate that the RNA polymerase II C-terminal domain plays a very important role in some genes and it regulates differential splicing, generating two different messenger RNAs from the same gene.

(Refer Slide Time: 43:36)



The other interesting example I want to give is to give an example how the rate of transcription elongation also can influence splicing. So, again I have taken the same gene, where you have weak three prime splice site and strong three prime splice site.

Now, it turns out the rate of elongation of transcription varies from gene to gene and it can be anywhere from 1.9 kb per minute to 4.3 kb per minute. So, the rate of depending

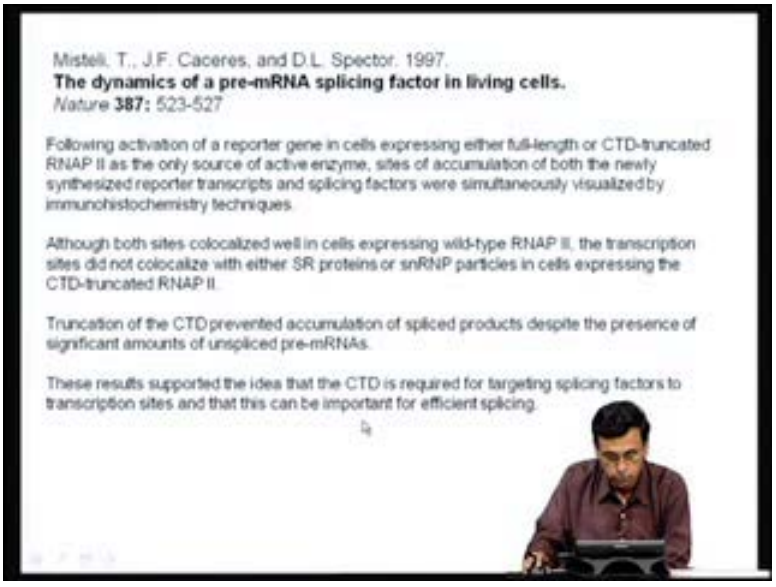
upon what kind of elongation factors are present, there are certain genes can be elongated in a slow rate and certain genes can be elongated at a very fast rate.

Turns out the rate of elongation may also influence splicing and can lead to alternating splicing and generate two different messenger RNAs. For example, in this particular case, when the transcription elongation is processing at a very fast space, then the splicing machinery does not have time to recognize this weak splicing site and as a result, this exon is skipped and this exon1is joined this exon3 and therefore, you get this kind of a messenger RNA transcript and this again is called exon skipping as we have seen in the previous slide.

Whereas if this gene is transcribed at a slow phase, then the splicing machinery has sufficient time to recognize this weak three prime splice site and as a result, this exon2 is included and these introns are spliced and you get a normal splicing reaction including all the three exons and is called as exon inclusion.

So, by differential rate of elongation, you can generate two different messenger RNAs by differential splicing. So, you can see, not only the RNA polymerase CTD can influence the rate of transcription the mRNA splicing reaction, but even the rate of transcription elongation can have a very important role in generating two different messenger RNAs from a single gene by influencing splicing reactions.

(Refer Slide Time: 45:26)



Now, I have just mentioned some papers here, just to give you an idea how people See, all these all these experiments on understanding transcription initiation, mRNA processing, all these groups were actually working in isolation for a long time, but then once people realized that all these factors actually interact with each other and they have to work together before messenger RNA comes out [in two case], microscopic techniques became very, very well developed and using these microscopic techniques, people started asking questions - do machinery involved in transcription initiation and elongation as well as machinery involved in splicing, will they also be present in the same place and same time inside the cell, inside the nuclei or are they present in different or they are present in different places inside the nucleus.

So, this particular paper for example, discusses wherein a reporter gene in the cells, where they are expressed as a full-length or a CTD, truncated CTD. That is, there are cells in which a reporter gene is getting expressed, there are cells where the source of transcription, the source of RNA polymerase is an RNA polymerase having a normal C-terminal domain and in another case, there is a the source of RNA polymerase is an RNA polymerase containing a truncated CTD and you have these two different scenario. It turns out accumulation of both newly synthesized reporter transcripts and splicing factor were simultaneously visualized by immunohistochemistry techniques under two different experimental conditions.

You have cells in which you are looking at the expression of a reporter gene as well as the presence of splicing factors, where the RNA polymerase has a normal C-terminal domain; you have other type of cells, where you are looking at the same phenomenon, but here the RNA polymerase is a truncated C-terminal domain.

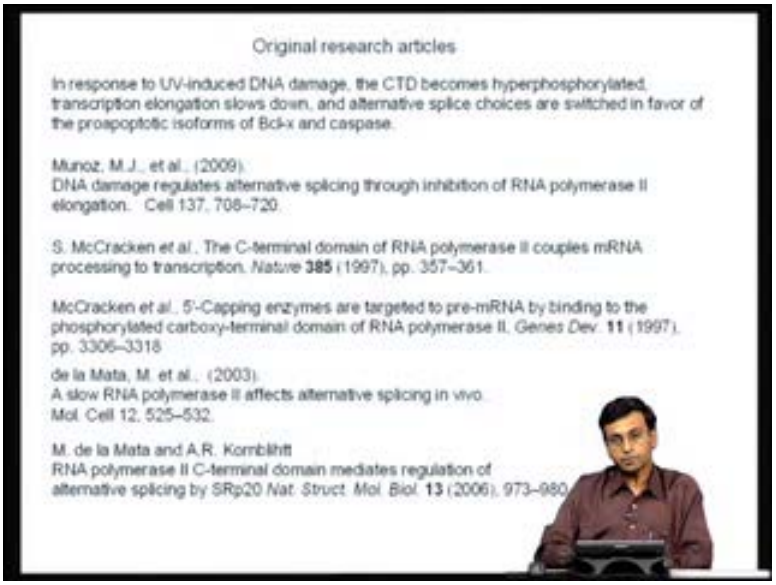
It turns out both the transcripts arising out of the reporter gene as well as splicing factors, they colocalized to the same place, when you have the normal RNA polymerase containing the normal C-terminal domain, whereas the transcripts or the transcription site did not colocalize with protein components involved in splicing, when you have a truncated C-terminal domain, clearly indicating that when you have a normal C-terminal domain of RNA polymerase, it actively recruits the splicing machinery to the vicinity of transcription and therefore, both RNA as well as splicing factor colocalized to the same place inside a nucleus.

Whereas if you have a truncated C-terminal domain, then this splicing factor cannot interact this with this truncated C-terminal domain and as a result, active splicing does not take place and therefore, the splicing factors do not colocalize with the RNA that is coming out of the transcription clearly indicated that the C-terminal domain of RNA polymerase plays a very, very important role in actively recruiting splicing factors to the vicinity of transcription and inside the nucleus both splicing and transcription takes place simultaneously and they colocalize to the same region.

So, these kinds of experiments, where you try to colocalize both transcripts they are coming out of the particular gene as well as the splicing factors, these kind of results clearly supported that the C-terminal domain is required here required actually for targeting splicing factors to transcription sites start site transcription sites and this can be very, very important for efficient splicing.

So, what I am trying to tell you from this particular paper is that the RNA polymerase CTD plays a very active role in recruiting not only the capping enzymes and as well as elongation factors, but also recruiting enzymes involved in splicing and that is why the splicing can be the rate of splicing can be enhanced by RNA polymerase II and if you have mutant RNA polymerase II, the efficiency of splicing is drastically reduced.

(Refer Slide Time: 49:10)



So, I have just mentioned some of the research articles, which I feel is very, very important for you to understand whatever I have told you so far. These are all the original research article that you have to go through you can go through. From these From these articles, I have picked up some of the important information I have discussed so far.

The take-home message I want to give you from today's discussion is two things. Number one - once transcription initiation is over, the next step in the regulation of gene expression is transcription elongation, RNA five prime capping, messenger RNA splicing and polyadenylation.

Now, the important message that I would like to convey from this lecture is that the RNA polymerase enzyme plays not only an important role in transcription initiation and subsequent transcription elongation, but it also plays a very, very active role in other RNA processing events such as mRNA capping, mRNA splicing and in the next class, I will tell you that RNA polymerase II plays a important role in polyadenylation as well.

I have clearly mentioned that phosphorylation of specific serine residues in the C-terminal domain of RNA polymerase results in the active recruitment of mRNA capping enzymes and not only it recruits the capping enzymes, but it also enhances the guanylyltransferase activity of the capping enzyme so that mRNA capping can be efficiently carried out.

So, the RNA polymerase II CTD allosterically modifies the capping enzyme so that the efficiency of mRNA capping is enhanced. So, it has a very, very active role in mRNA capping. Then, I briefly discussed how the splicing also plays a is also many of the splicing factors also, are associated with RNA polymerase II C-terminal domain and how when you have cells expressing normal C-terminal domain versus RNA polymerase within a truncated C-terminal domain, splicing can be drastically affected.

And this can be demonstrated even with the in vitro reactions, wherein if you take a test tube which contains all the components of RNA splicing, but if you now add an RNA polymerase containing a normal CTD versus an RNA polymerase containing a truncated CTD and the test tube which contains an RNA polymerase in normal CTD, the splicing is more efficient compared to the test tube containing a truncated CTD.

And I also gave you an example where the presence of a normal CTD versus truncated CTD can actually generate two different transcripts from the same gene clearly indicating that factors involved in splicing such as the Sp20 actually interacts with RNA C-terminal domain of RNA polymerase and this interaction influence the splicing machinery in recognition of certain weak versus strong three prime splice signals and as a result, two different transcription can be generated from the same gene.

I also gave an example where transcription elongation also plays a very important role and when the RNA is being translated at very fast rate, certain weak splice signals may not be recognized by the splicing machinery and as a result, this leads to what is called as an exon skipping and results in a different kind of messenger RNA, whereas when the same gene is transcribed at a slower rate, then even weak three prime splice signals can be recognized by the splicing machinery and you get what is called as exon inclusion and you can get a different kind of messenger RNA.

So, differential expression or differential regulation is possible even at the level of transcription elongation and factors which promote transcription elongation, factors which enhance transcription elongation may also affect things like RNA splicing and mRNA capping activities.

So, I think I will stop here clearly mentioning to you, the two important RNA processing activities namely mRNA capping and mRNA splicing are very coordinately or very intricately connected to the RNA polymerase C-terminal domain and then in the subsequent class, we will now discuss, even the final RNA processing mechanism namely, the polyadenylation or the addition of polyethyl at the three prime end RNA also requires very, very active role of RNA polymerase C-terminal domain.

Thank you.