

## **Eukaryotic Gene Expression: Basics and Benefits**

**Prof. P N RANGARAJAN**

**Department of Biochemistry**

**Indian Institute of Science Bangalore**

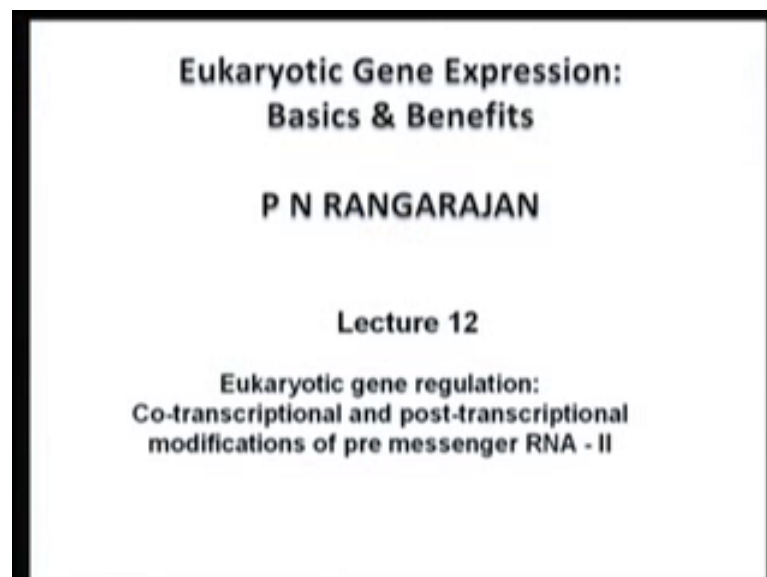
**Module No #04**

**Lecture No # 12**

### **Eukaryotic gene Regulation: Co-transcriptional and post-transcriptional modifications of pre messenger RNA- 2**

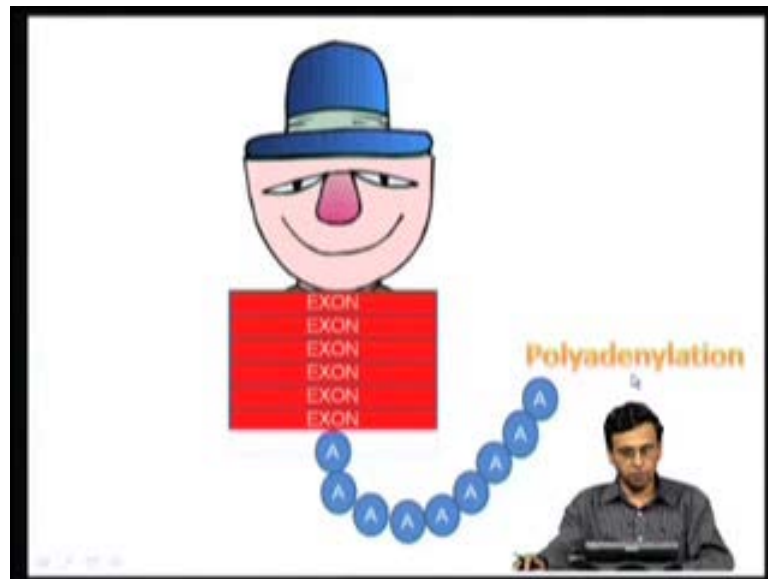
In the previous class, while we discussing about the regulation of the Eukaryotic gene expression, we primarily discussed about once the transcription is over, how does the messenger RNA, what kind of post-transcriptional modifications the messenger RNA undergo and what I basically told you in the last class is that there is what is called as a both co-transcriptional as well as post-transcriptional modifications that the pre messenger RNA undergoes during transcription or after transcription. And we discussed two such co-transcriptional processing mechanisms namely mRNA capping and mRNA splicing. And the third most important co or post-transcriptional modification that a pre messenger RNA undergoes is polyadenylation. So, today we will discuss about the mechanism by which a polyadenylate is added to messenger RNA and how again RNA polymerase 2 plays a very important role in this event as well.

(Refer Slide Time: 01:12)



So, today's topic we are going to talk about is co-transcriptional and post-transcriptional modifications of pre messenger RNA part 2. The part one we discussed about mRNA capping and mRNA splicing and in part two we are going to discuss about polyadenylation. How does messenger RNA get polyadenylated, what are the enzymes are involved and how RNA polymerase 2 plays a very important role.

(Refer Slide Time: 01:33)

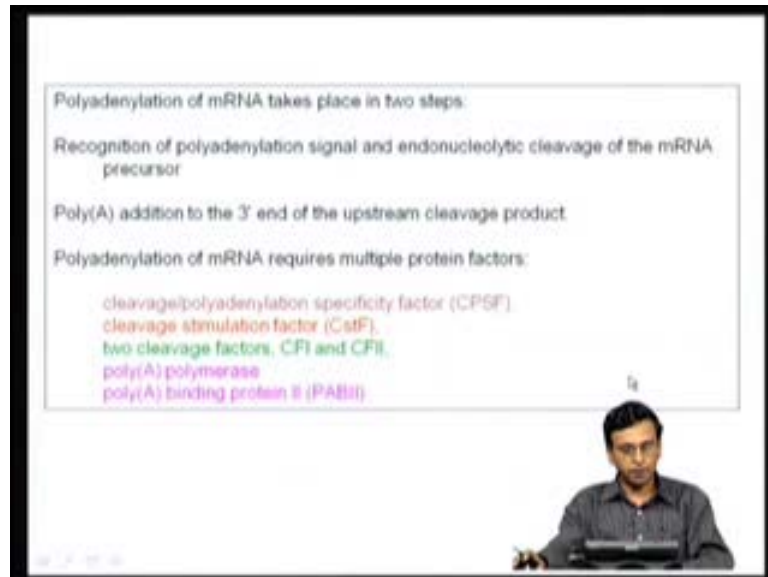


It is a cartoon I have drawn just to stay in your memory what we are going to just talk about. Once the pre messenger RNA is made which contains both introns and exons, a cap is added. That is one of the first co or post-transcription modifications a pre messenger RNA undergoes and we discussed how clearly the RNA polymerase 2 carboxy terminal domain plays a very important role in the addition of this cap. And then we also discussed about how the splicing machinery removes the introns and joins the exon together. And even the splicing is highly interconnected the RNA polymerase 2 and again RNA polymerase 2 plays a very important role, especially in the carboxy terminal domain. Moreover acts like a launching pad for the assembly of both capping factors as well as a splicing factors and as a result both splicing and capping are carried out simultaneously as RNA transcription takes place.

So, what will discuss today is about the third important co or post-transcriptional event namely the polyadenylation signal. That is what I depicted here. How a series of A residues are added to the messenger RNA, so that you get a polyadenylated and this is

what I have shown. So, what we will discuss today is that what is the mechanism by which polyadenylation is takes place and a polyadenylated is added to the pre messenger RNA during, as well as once the transcription is over.

(Refer Slide Time: 02:55)



Now, the polyadenylation is a very unique feature of a Eukaryotic messenger RNAs. It does not take place in prokaryotic species. And not all messenger RNAs are polyadenylated. All the majority of them undergo polyadenylation, but there are a small population of RNAs for example, histone messenger RNAs, they do not undergo polyadenylation and I also discussed in the previous class that this addition of polyadenylated has actually given a very important unique research tool to separate out messenger RNA molecules from ribosomal RNA and transfer RNA molecule. So, what you basically do? If you want to isolate messenger RNA from a cell from eukaryotic tissue or from a eukaryotic cells. You isolate the total RNA and then simply take this total RNA and pass it through what is called as a ((C)) cellulose column where cellulose is couple to a series of T residues and because the messenger RNA alone contains A residues. The A will base pair with T and only the messenger RNA will be retained on the Oligo D T affinity column whereas, the ribosomal RNA transform a pass through and this is how you can specifically enrich or you can specifically isolate the polyadenylated messenger RNA from total RNA of eukaryotic cell extracts.

So, the presence of polyadenylated as actually a given a unique tool for purifying polyadenylate messenger RNA from total RNA in eukaryotic cells. Now let us see how exactly the polyadenylation takes place in the case of eukaryotic cells. The polyadenylation of messenger RNA basically takes place in two steps. Recognition of polyadenylation signal and endonucleolytic cleavage of messenger RNA precursor is the first step, because somewhere there has to be a signal to tell that this is the place where a polyadenylated has to be added. So, there has to be what is called as a polyadenylation signal. So, will now analyze and see what is a polyadenylation signal? Is it conserved among all messenger RNAs or different in between different messenger RNAs and once the machinery involved in polyadenylation recognizes that this is the polyadenylation signal then the RNA is cleaved, downstream of this polyadenylation signal and then another enzyme takes over which actually adds poly A at the three prime end of the cleaved RNA.

So, there are two important events of polyadenylation. One is recognition of the polyadenylation signal and cleavage of the pre messenger RNA and once that is over a series of A residues are added by two enzymes called poly A polymerase and poly A binding protein 2. Now there are three protein factors which play very important role in the recognition of the polyadenylation signal and endonucleolytic cleavage of the pre messenger RNA. And they are called as CPSF, which is cleavage polyadenylation specificity factor, cleavage stimulation factor are CstF, and there are two cleavage factors called CF 1 and CF 2. These 3 are involved in the first step of the polyadenylation and once these three enzymes do their job and cut the pre messenger RNA 2 enzymes called poly A polymerase and poly A binding protein actually add a polyadenylate, so that you get a polyadenylation messenger RNA spaces.

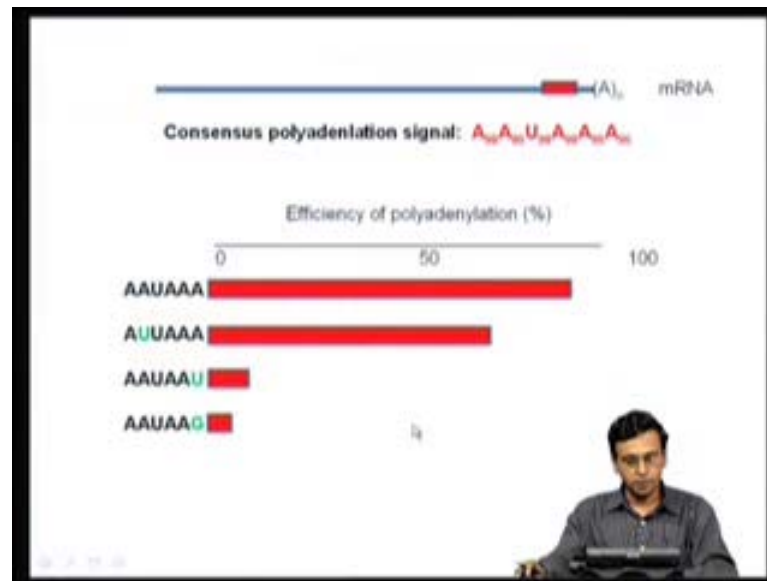
(Refer Slide Time: 05:49)



Let us now take a very brief look at what are these enzymes? What exactly they do? As I mentioned earlier CPCF is nothing but cleavage and polyadenylation specificity factor. It actually is the one which recognizes the poly A signal and in majority of the eukaryotic messenger RNAs, the pol A relation signal is highly conserved and it is a sequence called AAUAAA, repeat AAU triple A. This is the sequence called polyadenylation signal and this is the signal that is recognized by the CPCF and it is required for both cleavage as well as polyadenylation. So, once the transcription is over, once the RNA polymerase transcribed this poly A signal and this poly A signal present in the RNA is immediately recognized by the CPCSF and then it recognizes that the transcription has to stop here. And it cleaves the messenger RNA and adds a polyadenylate.

And there are two proteins involved in the cleavage of the once the CPCF recognizes the AAUAAA sequence, there are two proteins called CF 1 and CF 2 which will actually involved in the cleavage process and then is what is called as CstF cleavage stimulating factor which is actually required for cleavage and its binds to a region called GU slash U region, which is immediately downstream of the AAUAAA region. And then once all these events are over, two enzymes called poly A polymerase and poly A binding protein 2 play important role in the addition of the A residues to the end of the RNA. Initially the poly A polymerase in that is involved in the polyadenylation process it adds about 10 to 20 residues of A and then, the poly A binding protein takes over and its adds up to 100 residues of A to get a completely polyadenylate messenger RNA.

(Refer Slide Time: 07:29)

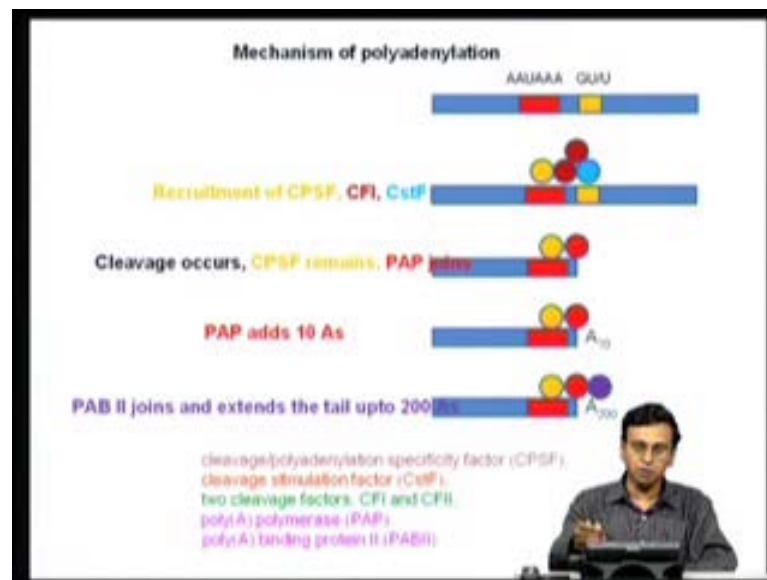


And now let us now look at the sequence, which plays a very important role in the polyadenylation namely, the polyadenylation signal. The question we ask is that since there are number of protein coding genes, there must be number of messenger RNA synthesized not only in a given cell type, but in various tissues and also different eukaryotic species. The question people ask is that is the machinery that is involved in polyadenylation, does it recognize the same polyadenylation signal? Is it conserved? Just like we have Tata box which is conserved in most of the eukaryotic promoters whereas, the peoples are asking the question, are there conserved polyadenylation signal? So, when they analyze a number of eukaryotic messenger RNA sequences, they found out that at the three prime end of the RNA, in majority of the cases this is the sequence that is highly conserved. You can see 98 means in 98 out of 100, the 98 percent of the genes the first residue is A, similarly 96 percent of the genes the first residue is A, the next residue is A and so on and so forth.

So, majority of the genes this is the signal that is highly conserved at the three prime end of the messenger RNA. And to actually find out whether this sequence really has any role in polyadenylation signal, people did what is called as a **(C)** experiment. So, you can see if you look at the efficiency of poly relation of a normal messenger RNA containing AAUAAA sequences 100 percent. You see if we now mutate the second A residues to U the efficiency drop down to more than 60 percent. Similarly, if you now mutate the last A residue to U, there is a dramatic drop in the polyadenylation efficiency.

Similarly, mutation of A to G also result decrease in the efficiency of polyadenylation. clearly indicating that these sequence AAUAAA is very essential for efficient polyadenylation of eukaryotic messenger RNA. And as we go along we realize that it is the one of the poly relation proteins for binding factors actually recognizing the sequence and this is what the signaling first signal for polyadenylation of messenger RNA. So, remember a signal called AAU triple A, we call as the polyadenylation signal is highly conserved in majority of the eukaryotic messenger RNAs and these signal is very essential for efficient polyadenylation of eukaryotic messenger RNAs.

(Refer Slide Time: 09:45)



Let us now briefly discuss about the mechanism by which polyadenylation takes place. Like I told you the cis acting elements which are involved in polyadenylation of this signal called AAUAAA sequence and also a adjacent sequence called GU slash U region and these are the cis acting sequences which are recognized by specific transacting factors. The first step in this polyadenylation is recruitment of three proteins the cleavage and polyadenylation specificity factors CPCSF, the cleavage stimulation factor and the cleavage factor 1. These proteins come and they recognize the AAU sequence as the GU sequence and they bind and they are recruited to the cis acting elements of the messenger RNA with these elements are present in three prime end. And once these assemble and recognize the poly A sequence the cleavage immediately about 20 to 30 base pairs downstream of the polyadenylation signal, cleavage takes place and after the cleavage takes place of these three proteins, the CPSF remains bound and a new protein called the

poly A polymerase joins and this poly A polymerase which are shown as a red ball here, now adds about 10 residues of A residues. Then another enzyme called the poly A binding protein takes over and elongates further and adds up to 100 or 200 A residues further.

So, this is in a nutshell the mechanism by which polyadenylation of messenger RNAs takes place in the case of eukaryotic cells. So, you can see here. First is the recognition of the polyadenylation signal which consists of the AAU triple A by three polyadenylation factors the CPSF, CF 1 and CstF and after this, about 20 to 30 base pairs downstream of the signal the RNA is cleaved and once the cleavage is over the poly A polymerase comes and joins the RNA it adds about 10 A residues and then the poly A binding protein joins and then expands or extends the A further up to 200 residues.

(Refer Slide Time: 11:52)

**Role of RNA polymerase II in polyadenylation and transcription termination**

RNAs transcribed by CTD-truncated RNAP II were not efficiently polyadenylated in transiently transfected cells.

It was also shown that CPSF and CstF present in unfractionated nuclear extracts could bind GST-CTD, and both were present in an RNAP II holoenzyme preparation.

Unlike capping and splicing, the phosphorylation status of the CTD was found not to affect binding.

These results suggested that the CTD functions to help recruit polyadenylation factors to sites of RNAP II transcription, increasing their local concentration and thereby facilitating efficient processing.

Proudfoot NJ (1989) *TIBS* 14: 105-110  
McCracken et al. (1997) *Nature* 385: 357-361

The slide also features a small video inset in the bottom right corner showing a man in a dark shirt speaking at a podium.

So, this is mechanism by which polyadenylation takes place. Now the important point that I would like to discuss, like the way we discussed in the previous lectures is that RNA polymerase 2 not only plays an important role in initiation of transcription, but it also plays a very important role in mRNA capping, in splicing and today will tell also plays a very important role in polyadenylation of messenger RNA. This is the message I want to convey. So, most of the people think the only job of RNA polymerase is to initiate transcription and make RNA. But what is very important that take home message I want to give end of these previous lecture and this lecture is that RNA polymerase 2



plays a very important role not only in the RNA synthesis, but also in RNA processing. Pre mRNA processing and especially the carboxy terminal domain of the largest sub unit of RNA polymerase two is very essential for all these three post-transcriptional or more appropriately co-transcriptional process to takes place, namely capping, splicing and polyadenylation.

And what is the evidence? Why do you think that RNA polymerase 2 plays a very important role in this? The evidence is that RNAs transcribed by a CTD truncated RNA polymerase 2 were found to be not efficiently polyadenylated in transiently transcribed cells. So, we have cells which are expressing normal RNA polymerase 2 with its normal carboxy terminal domain and if we have cells in which the RNA polymerase carboxy terminal domain is truncated and then, if you isolate RNA from these two different cell types and look at the efficiency of polyadenylation in the cells expressing the normal RNA polymerase the mRNAs are very efficiently poly related whereas, the cells which are expressing a truncated or a deleted form of carboxy terminal domain you do not see very efficiently efficient poly relation.

Clearly indicating that the carboxy terminal domain of RNA polymerase is playing a very important role for polyadenylation of messenger RNAs. In fact, it was shown that the two factors involved in polyadenylation which would just discussed in the previous slide namely the CPSF and CstF is actually present in the unfractionated nuclear extracts they could actually bind to GST-CTD. Now what is the GST-CTD? Now, if you want to study protein-protein interactions what you do is suppose I would like to know or any of the proteins involved in polyadenylation are the actually interacting with the RNA polymerase 2 carboxy terminal domain. So, what do I do, take the carboxy terminal domain of RNA polymerase 2 and fuse it to glutathione S-transferase. So, I have a glutathione S-transferase C-terminal domain fusion protein and this can be now made in large amounts in equalized cells and since the glutathione S-transferase has very high affinity for glutathione. If you now takes this whole cell extract containing the glutathione S-transferase CTD fusion protein and you can bind this protein specifically to a glutathione affinity column.

And now you take nuclear extracts which contains all kinds of proteins and if you pass this nuclear extracts to glutathione affinity column, all those proteins which interact with this carboxy terminal domain of RNA polymerase 2 will now be retained on the column

and then you can elute these proteins and identify this protein either by N-terminal sequence of proteins or by mass spectrometry, so on and so forth. In fact, this is one of the techniques by which people have actually identified what kind of proteins are actually interacting in the RNA polymerase carboxy terminal domain and when they actually did these kinds of experiments, found two of the proteins involved in polyadenylation namely CPSF and CstF can actually be retained on GST affinity columns which contain the GST-CTD.

So, the control is only the GST whereas, the actual protein will be GST-CTD and proteins which are retained only on GST-CTD column, but not on a GST column indicates that, it is the CTD of RNA polymerase 2 which is interacting with these proteins. So, when these kind of experiments were done they actually found both two of the proteins involved in polyadenylation namely CPSF and CstF can be retained on GST-CTD columns indicating that these proteins actually interact with the carboxy terminal domain of RNA polymerase 2.

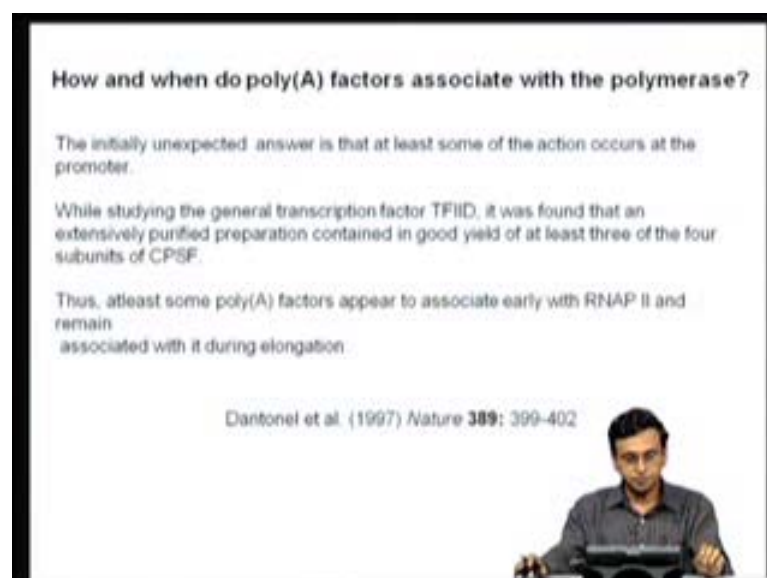
These kinds of biochemical experiments (( )) to the in vivo experiments which I told you earlier that, if you truncate CTD and in such cells the efficiency of polyadenylation is very low. These kinds of experiments clearly indicated that the RNA polymerase 2 C-terminal domain is very essential for the polyadenylation to take place. And see these kind of experiments were also done in the previous cases where people were actually shown enzymes involved in mRNA capping, enzymes are RNA splicing, they also interact with RNA polymerase two C-terminal domain and in fact, phosphorylated forms of RNA polymerase CTD are more efficient in promoting mRNA capping as well as mRNA splicing. We discussed this aspect in detail in the previous class.

But the important distinction between the role of or interaction of RNA polymerase 2 CTD between mRNA capping enzymes and splicing enzymes and the interaction of RNA polymerase 2 CTD polyadenylation machinery is that in the previous two cases the phosphorylated status of RNA polymerase 2 plays a very important role both in interactions of the splicing machinery as well as capping machinery with the RNA polymerase 2 CTD phosphorylation status. But in the case of polyadenylation phosphorylation of CTD does not seem to play a very important role in the interaction with the polyadenylation machinery or proteins involved in polyadenylation.

So, unlike capping and splicing phosphorylation status of CTD was not found to affect interaction of the CPCF and CstF with the RNA polymerase 2 CTD. So, all these results clearly indicated that the polymerase 2 carboxy terminal domain plays a very important role in poly relation. It does help to recruit poly relation factors to the site of RNA polymerase 2 transcription increasing their local concentration and thereby pro facilitating efficient processing or efficient polyadenylation. I got two references where one in TIBS and one in nature. This is the time during the late 90s, late 80s and early 90s that is the time, a lot of effort went on to identify polyadenylation mechanisms, understand the factors involved in this and people are actually developed invitro systems just like a have a selfy transcription system, people developed selfy systems for demonstrating mRNA capping, selfy systems for demonstrating splicing and also selfy systems for studying polyadenylation. So, where we can actually purify factors and put them in a test tube and actually demonstrate polyadenylation, can actually takes place.

And in fact, using such kind of experiments people are actually shown, if you have extracts in which the RNA polymerase 2 CTD or RNA polymerase 2 is depleted, you find the polyadenylation efficiency is the very low and if you now put back RNA polymerase 2 you see the polyadenylation how takes place very well. So, both in vivo experiments as well as biochemical experiments involving protein proteins studies as well as selfy polyadenylation systems demonstrated that the carboxy terminal domain of RNA polymerase 2 plays a very important role in efficient poly relation process.

(Refer Slide Time: 18:45)



**How and when do poly(A) factors associate with the polymerase?**

The initially unexpected answer is that at least some of the action occurs at the promoter.

While studying the general transcription factor TFIID, it was found that an extensively purified preparation contained in good yield of at least three of the four subunits of CPSF.

Thus, atleast some poly(A) factors appear to associate early with RNAP II and remain associated with it during elongation.

Dantoni et al. (1997) Nature 389: 399-402

Now, how and when do poly A factors associate with the RNA polymerase? Now people thought as we can see polyadenylation is almost the last step in messenger RNA synthesis because once the RNA is made then it gets capped, then it gets spliced and finally, the RNA gets cleaved and polyadenylate is added. So, people thought proteins which are involved in poly relation probably has nothing do with RNA polymerase and even if they interact the RNA polymerase it must be happening at the towards the fag end of the transcription process. But very surprisingly people found that even at the stage of transcription initiation itself, many of the proteins involved in the polyadenylation actually, interact with the RNA polymerase 2 and they remain bound to the part of the pre initiation complex. So, this was the very unexpected answer where, they found that the enzymes involved in polyadenylation actually, associated with pre initiation complex and there actually present in the promoter itself.

So, although polyadenylation is the last step, almost the last step in the mRNA processing the enzymes involved in polyadenylation, seem to be associated with the promoter itself. And some of the experimental evidences are shown here for example, while studying the general transcription factor TF 2 D, it was found that an extensively purified preparation contained in good yield of at least 3 of the 4 subunits of CPCF. TF 2 D which so far, we thought is involved only in transcription initiation and its job is to make a assemble of pre initiation complex and to recruit the RNA polymerase 2 promoter, the TF 2 D is also seem to be interacting with CPSF which is a polyadenylation factor.

So, at least some of the poly A factors appear to associate very early during RNA synthesis and during the transcription initiation a step itself and they remain associated with RNA polymerase throughout the process of transcription elongation. So, the cuts of what this slide actually tolds you is that although, polyadenylation is the last step in pre mRNA processing. Some of the factors involved in polyadenylation seem to be associated with the RNA polymerase and the pre initiation complex at the stage of transcription initiation itself and right from the beginning they remain associated with RNA polymerase throughout the transcription process.

(Refer Slide Time: 21:04)

**Biochemical evidence demonstrating a more direct role of RNAP II in polyadenylation**

Purified RNAP IIA and IIO were both found to activate the first step of polyadenylation, 3' cleavage, in a reconstituted system containing all the other polyadenylation factors.

In addition, both unphosphorylated and hyperphosphorylated GST-CTD proteins activated cleavage just as efficiently as did RNAP II, although in this case the hyperphosphorylated CTD was more active than the nonphosphorylated form.

Further, 3' cleavage in nuclear extracts could be inhibited by immunodepletion of RNAP II and rescued by add-back of the purified enzyme.

These results suggested that the CTD of RNAP II participates directly in the formation and/or function of a stable, catalytically active processing complex (through direct interaction with polyadenylation factors).

Hirose & Manley (1998) Nature 395: 93-96

Now, what is the biochemical evidence demonstrating a more direct role for RNA polymerase 2 in poly relation? This is where I discussed just little while ago. If we now take the purified RNA polymerase 2 and RNA polymerase 2, both were found to activate the first step in poly relation. Remember the difference between polyadenylation 2 A and O is that one is a non-phosphorylated form and another is a phosphorylated form. As I mentioned previously, for the interaction of the poly A factors with RNA polymerase 2 phosphorylation does not need to play important role. So, whether the CTD is phosphorylated or not the RNA polymerase 2 efficiently interacts with poly A factors and therefore, both of them are capable of promoting polyadenylation. In a reconsidered system containing polyadenylation factors and RNA polymerase 2. So, we have a selfy transcription system which contains all the factors which I just mentioned previously. So, that we have a selfy transcription where, if we now add a pre messenger RNA it is capable of getting poly related.

Then such a system, if you now add an RNA polymerase 2 with its carboxy terminal domain the efficiency of poly relation is dramatically enhanced, clearly indicating that the polymerase 2 especially the carboxy terminal domain plays a very important role. And both unphosphorylated and hyper phosphorylated GST-CTD proteins activated the cleavage just as efficiently as RNA polymerase 2 although in this case the hyper-phosphorylated was little bit more active than the non-phosphorylated form. Some

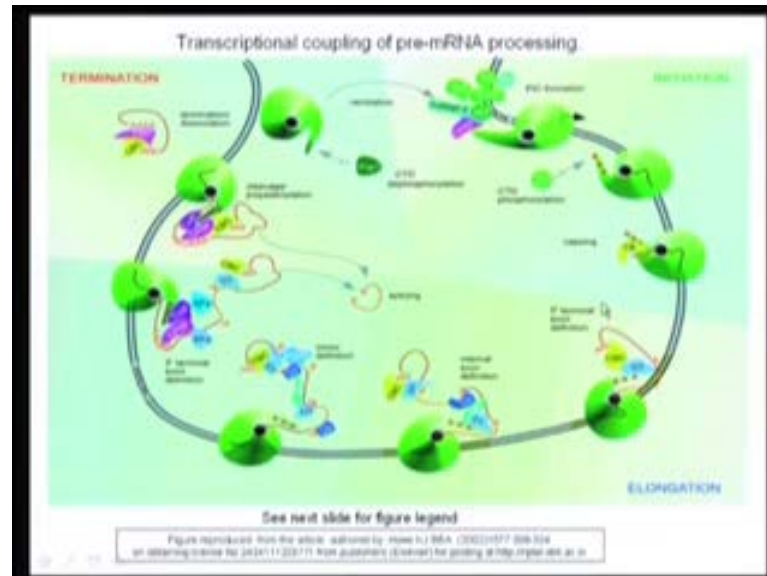
experiment did suggest that there is some difference in the whether the CTD is phosphorylated or non-phosphorylated.

And the three prime cleavage in nuclear extracts could be inhibited by immunodepletion of RNA polymerase two and can be rescued back by adding purified enzymes. So, all these studies, I am telling you because during the late 80s and in early 90s a lot of effort went in to really understand that the RNA polymerase 2, in addition to having a role in transcription also plays a very important role in other sub sequence steps of RNA processing namely capping, splicing and poly relation and since this is against the dog mark, nobody thought enzymes which are involved in polyadenylation could be assembled at the time of the pre initiation complex itself in the promoter and RNA polymerase to actually be interact with some of these factors.

These for all not unimaginable those days. So, to convince people that yes, all these things are happening together and RNA polymerase 2 CTD playing a very important role a number of experiments have to be done. For example, in the last step what I said you actually take an nuclear extract which is capable of polyadenylation and in such an extract, if you know immunodeplet. That means, you pass this extract through a column which contains antibodies against the carboxy terminal domain of RNA polymerase 2, so that the RNA polymerase 2 along with the CTD can be depleted and a such extracts, they found the poly relation does not takes place efficiently. To actually show that you have only depleted RNA polymerase 2 and its CTD, now if you add that purified RNA polymerase along with CTD, now there is a dramatic increase in the poly relation efficiency. Clearly indicating that the RNA polymerase 2 and CTD is playing a very important role in enhancing the efficiency of poly relation and they are very essential for poly relation to take place.

So, all these kinds of experiments are clearly suggested that the carboxy terminal domain of RNA polymerase 2 participates directly in the formation of a stable catalytically active processing complex through direct interaction with the poly relation factors. A person called James Manley played a very important role. He has published a number of papers to study polyadenylation factors, mechanism of poly relation, developed selfy transcription and so on. I will just show one of the key papers published in late 1990s to demonstrate the link between the RNA polymerase 2 and polyadenylation.

(Refer Slide Time: 24:44)



I now shown a cartoon here which actually demonstrates all that we discussed in the last two classes. Basically, this cartoon which was actually taken from a paper published in BBA in 2002, it is a nice review. I have given the reference also here. And this also actually tells you how transcription is couple to RNA processing. So, you can now look at this. What basically is cartoon tells you is that we know how the pre initiation complex for many form. This is the RNA polymerase 2 and we have all the general transcription factors like TF 2 ABD, so on and so forth. But what was surprising is that the purple color, what I have shown here is the poly relation factor.

Now, those days especially in early 90s nobody thought an enzymes CPCF which is involved in poly relation is actually interact is part of a pre initiation complex at the time of transcription initiation. This was a very surprising finding. So, in addition to the general transcription factors in RNA polymerase 2 a factor which is involved in the last step of RNA processing also is associated and the beginning of the transcription itself. So, throughout the transcription the C P S F remains bound to the RNA polymerase 2 and it is only when the RNA polymerase passes through the AAATAA signal that this poly A factors now jump into the RNA and then cleave the RNA till then, they remain associated with the RNA polymerase 2 CTD.

Now, let us see what happens. Once the pre initiation complex is formed, the general transcription factor TF 2 H. Now phosphorylates the RNA polymerase 2 that is shown in

the red asterisk here and once the RNA polymerase 2 is phosphorylated that is the signal elongation to start. So, the PSE leaves the promoter and starts transcribing and when the RNA is about 20 nucleotides old the capping enzyme, all the enzymes are mRNA capping, now hop on to the phosphorylated RNA polymerases CTD and stop doing the mRNA capping. They are the 5 meters Cytosine cap to the five prime end of the messenger RNA. And once the capping is over then, enzymes involved in splicing. Now hop on to the RNA polymerase 2 CTD and they carry out the splicing process and these three steps actually show various splicing taking place and introns getting spliced out. Basically, what I am telling you is that both the capping machinery as well as splicing machinery are happening as the transcription is proceeding all these RNA processing does not.

So, the earlier concept that mRNA capping and splicing are post-transcriptional has now paved way to being defined as co-transcriptional. This is the message I want to convey. mRNA capping and at least some of the splicing reactions are co-transcriptional rather than post-transcriptional processes. And once the capping is over, once the splicing is taking place and once the RNA polymerases traverses, the poly A signal in the gene and the poly A RNA synthesise, the capping enzyme which was actually interact with the RNA polymerase CTD, now hops on to the AAUAAA sequence of the RNA and then immediately cleaves and then the poly A binding protein now, adds poly A residues and now, I have got the polyadenylated messenger RNA coming out into the nucleus.

Now, there are also two important steps which I have not discussed during the RNA processing steps because it is little bit advanced, I do not want to confuse the readers and there are also a regulations taking place at the level of transcription elongation. So far we have discussed regulation at lower transcription initiation and also during the RNA processing how alternate splicing can lead to more than one messenger RNA, so long and so forth. But transcription regulation can also takes place **the below** transcription elongation. But maybe I will just show one or two slides to actually, show the role of elongation factors in regulating transcription and there are also regulation taking place at the lower transcription termination. For example, a protein called FCP 1 plays a very important role and this is the one that actually, now dephosphorylates the carboxy terminal domain and therefore, the RNA polymerase can be now recycle back and can



again go back to the promoter, again interact with the general transcription factor and starts one more round of transcription.

So, this is what is called a transcription cycle. We have initiation, elongation and termination and the message, I want to give is that the pre mRNA processing which involves capping, splicing and polyadenylation. All these three are coupled to transcription and there are highly integrated and all these three take place in a very harmonious manner and RNA polymerase 2 especially, C-terminal domain plays a very important role not only in transcription initiation, but also elongation capping, splicing and polyadenylation

(Refer Slide Time: 29:28)

**Transcriptional coupling of pre-mRNA processing.**  
Figure legend

The transcription cycle is depicted starting at the top right and proceeding in a clock-wise fashion.

Processing factors may be recruited to the PIC at the promoter (TATA) by GTFs (TFIID, 3D; TFIIB, IIF) and transcriptional activators (TAs).

Hypophosphorylated RNAAP II (green "tailpole" depicted with a CTD "tail").


TFIIB phosphorylates the CTD (red stars) soon after initiation, which recruits CE.

CBC mediates 5' terminal exon definition. IVSs are demoted by the gold DNA, SR proteins and the CTD facilitate internal exon definition. Intron definition occurs when splice sites are paired across introns (Pip40 is shown assisting this process).

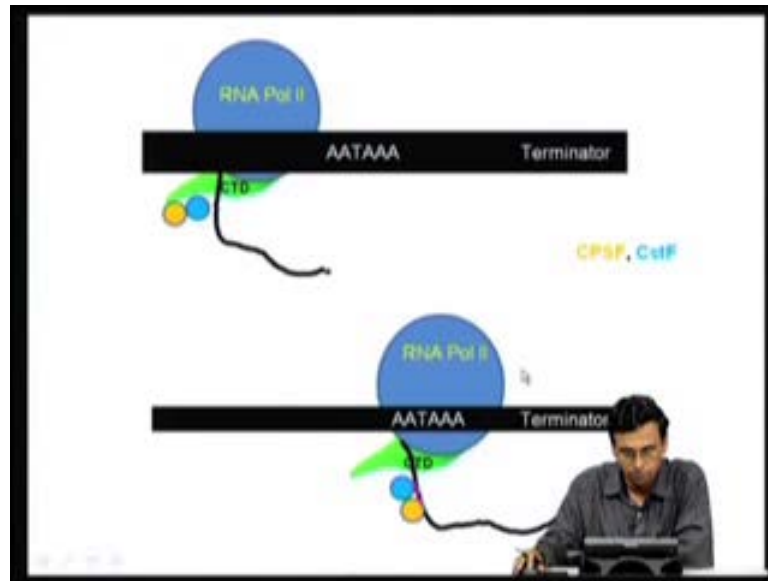
3'-End processing is mediated by cleavage/polyadenylation factors assembled at the poly(A) site (AATAAA) and is assisted by the CBC and splicing factors (SFs). These interactions mediate 3' terminal exon definition.

Transcription termination is accompanied by CTD dephosphorylation (proposed to occur 5

Figure legend reproduced from the article: authored by Jesse H. BBA (2002) 917 808-816  
on obtaining license to 2023/11/20/211 from publishers ( Elsevier) for posting on the repository at ii

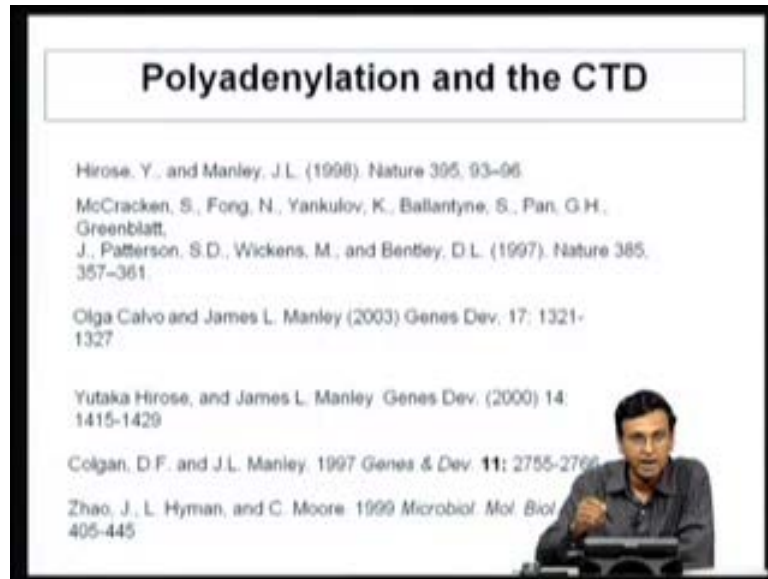


(Refer Slide Time: 29:33)



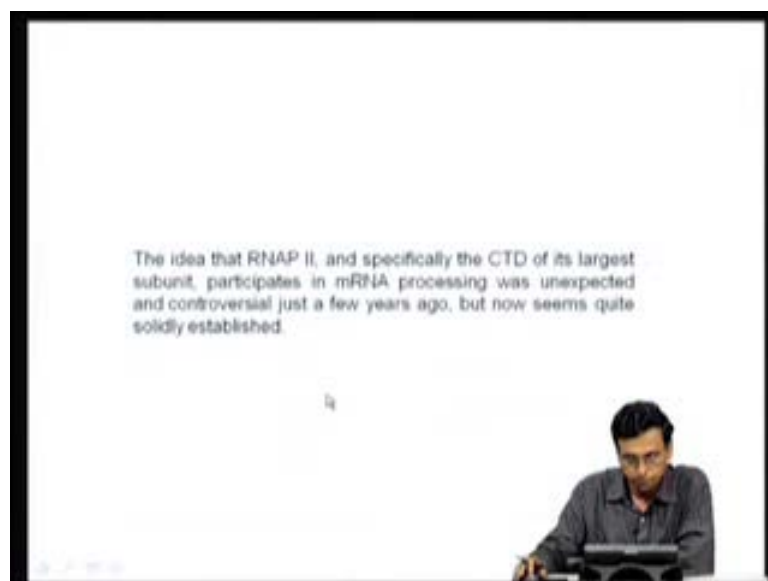
So, this is just the legend for the previous figure, we can just go through this. This is just the cuts of what I told you that the two of the enzymes proteins involved in polyadenylation namely CPSF and CstF, they hop on or hop on to the carboxy terminal domain of the RNA polymerase 2 and remain associated with the carboxy terminal domain throughout the transcription process, but once the RNA polymerase 2 whole enzyme crosses the poly A sequence and once the RNA gets transcribed and once you have the poly A the polyadenylated messenger RNA coming out of the RNA polymerase, immediately these enzymes leave the RNA polymerase and hop on to the RNA and immediately they cleave the messenger RNA and then the polyadenylated messenger RNA is added and polyadenylated messenger RNA is released. So, this is the mechanism by which poly A addition takes place and the interactions between machinery of polyadenylation and RNA polymerase in C-terminal domain.

(Refer Slide Time: 30:24)



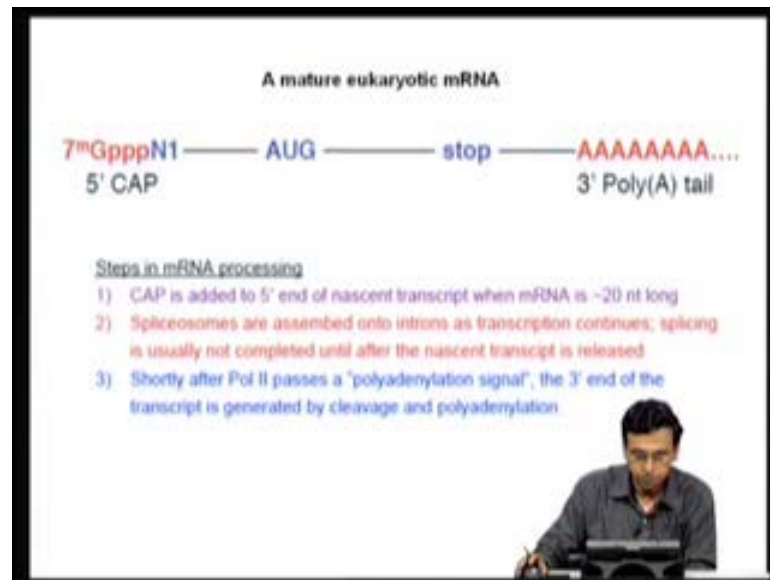
I just given a number of papers which describe the major advances that took place in the discovery of decipher in the mechanism of poly relation, especially the coupling of or the role of RNA polymerase C-terminal domain poly relation. So, you can just go through the some of these papers and then read little bit more about ((C)) and as I told you James Manley has published a number of papers during the late 90s and early 2000 to decipher or understand the mechanism of polyadenylation.

(Refer Slide Time: 30:56)



So, what is surprising is that the RNA polymerase 2 especially the CTD the C-terminal domain participates in mRNA processing was a kind of unexpected and there was highly controversial. In the initial stages, nobody believed that the RNA polymerase 2 which is actually involved in transcription initiation may be having a role in the subsequent steps of RNA processing. So, it took a lot of time and lot of experimental evidence to convince biochemical experiments in vivo experiments, selfy studies to actually convince that all these process are integrated. So, if you do not have RNA polymerase CTD, not only transcription initiation affected. Capping is affected, splicing is affected, polyadenylation is affected and to convince people that is real a lot of experiment is to be done during the late 90s and early 2000.

(Refer Slide Time: 31:47)



So, the bottom line what came out of this lecture and previous lecture is that a matured eukaryotic messenger RNA contains a five prime cap, a five prime (( )) at the five prime end, so, it gets capped and then all the introns are removed and then it contains the poly A (( )) at the three prime end. And there are three major RNA processing mechanism that we have discussed so far. A cap is added to the five prime end of nascent transcript when the mRNA is about 20 nucleotide long. So, capping does not wait till the entire transcription is completed. Even as the RNA is coming out of the RNA polymerase elongation complex, even RNA is just 20 nucleotide old, the capping takes place. And as the RNA synthesise continues, the spliceosomes are assembled onto introns as transcription continues, splicing usually it is not completed until the nascent transcript is

related. So, unlike capping, so, I can actually call capping as a co-transcriptional process, but splicing is probably both co-transcriptional as well as post-transcriptional process.

So, while capping is completed even before the RNA synthesis is completed, splicing is initiated co-transcriptionally and completed post-transcriptionally. This is the message I want to give. So, the two distinction between capping and splicing is, capping is done in a co-transcriptional manner is co-transcriptional whereas, splicing is initiated co-transcriptionally and completed post-transcriptionally. Now, shortly after the RNA polymerase 2 pauses the polyadenylation signal the three prime end of the transcriptional generated by cleavage and then, poly relation takes place. So, three RNA processing mechanisms capping splicing and poly relation or either co-transcriptional or co-transcriptional as well as post-transcriptional and RNA polymerase 2 C-terminal domain plays a very important role in all the three processes.

(Refer Slide Time: 33:39)

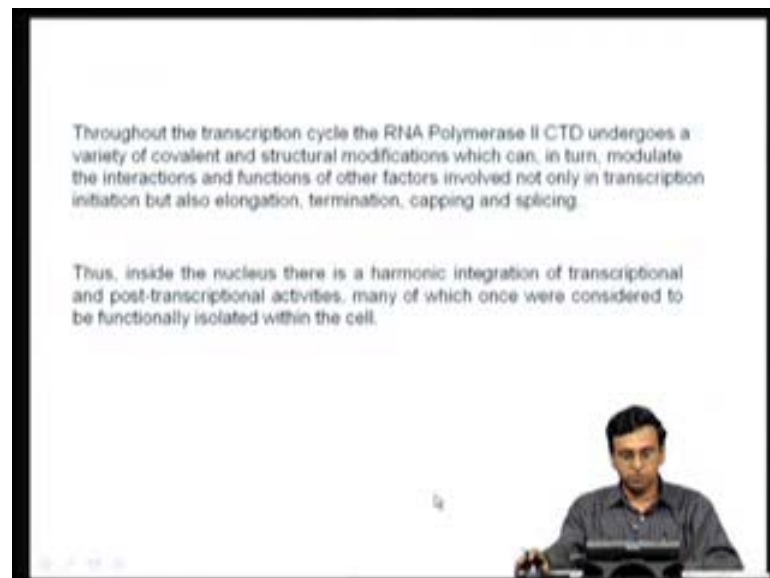


This is a cartoon for it easily remember, you have RNA polymerase and a RNA polymerase CTD and the RNA polymerase 2 CTD is involved not only initiation of transcription, elongation, capping, polyadenylation, splicing. So, this is a guy who is conduct RNA polymerase 2, CTD is the one who is conduct taken orchestra and who are the players, these are all the people playing various instruments and all these instruments are being controlled by one person who conducts the orchestra and then same way RNA

polymerase 2 CTD not only controls initiation, it also controls elongation, capping, polyadenylation as well as splicing.

And all these factors which are involved, many of these factors which all these processes they interact with either unphosphorylated or phosphorylated form of C-terminal domain of RNA polymerase 2. So, this is the take home message of this lecture and the previous lecture, that mRNA processing is a co-transcriptional and post-transcriptional mechanism and it happens as the transcriptional proceeds and RNA polymerase 2 plays a very important role in not only initiation, but also in the subsequent RNA processing events.

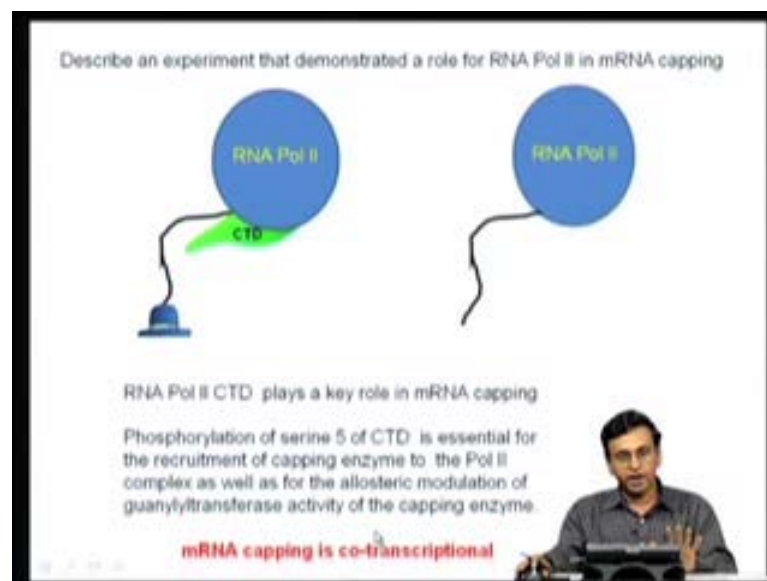
(Refer Slide Time: 34:46)



So, throughout the transcriptional cycle the RNA polymerase 2 carboxy terminal domain undergoes a variety of covalent and structural modifications, which can in turn modulate the interactions and functions of the other factors involved not only in transcription initiation, but also elongation termination capping and splicing. So, this whole thing is a very dynamic process. So, many residues of CTD are getting phosphorylated, dephosphorylated depending upon which residues getting phosphorylated and which residues getting dephosphorylated. It interacts either with capping enzymes or splicing enzymes or polyadenylation enzymes and the whole thing is a very dynamic process. So, throughout the process of transcription elongation and termination, the CTD is undergoing dynamic changes of phosphorylation, dephosphorylation and differential protein-protein interactions.

So, inside the nucleus there is a harmonic integration of transcriptional post-transcriptional activities many of which once were considered functionally isolated in the cell. So, the take home message I want to give from this lecture is that, the early 80s or late 80s people thought initiation is a different event and once the transcription initiation is over then the mRNA processing takes place, capping is added separately, splicing is added separately and polyadenylation adds separately and these all happens in different places in the nucleus. But what we now, know today is that these are highly integrate. They are all happen both in space as well as time, they happen together. They are all highly linked and all the proteins or enzymes involved. All these process they interact with each other and there are highly linked. This is the one of the major findings of research coming out of studying transcription initiation as well as post-transcriptional process of messenger RNA.

(Refer Slide Time: 36:25)



So, a number of experiments as I said, it took hell lot of time or lot of effort to actually prove this point, that the RNA polymerase 2 especially, C-terminal domain plays a very very important role in the co-transcriptional processes and I am just listing one or two experiments. So, that it registers in your mind for example, here is a cartoon how do you demonstrate that a RNA polymerase 2 especially, C-terminal domain really has a role in mRNA capping. If somebody ask the question, what you do? Take cell type, there are two cell types, one expressing the normal RNA polymerase with C-terminal domain, another expressing the truncated RNA polymerase 2 without C-terminal domain and if

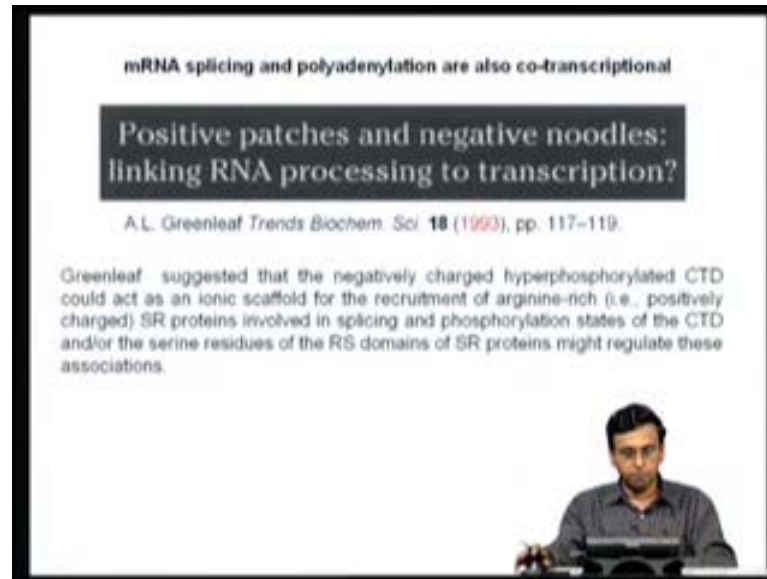
we now look at the messenger RNA which have been synthesized and you find that only in the cells which contain the normal RNA polymerase (( )) complete C-terminal domain messenger RNAs are capped whereas, cells expressing RNA polymerase two in which the CTD is not there RNA is made, but the RNA is not efficiently capped.

And what people like to find in these kinds of experiments? Not only capping is getting affected in these kinds of cells where CTD is truncated or mutated. Other events like splicing, polyadenylation also getting affected. So, these kinds of in vivo experiments clearly demonstrate that the C-terminal domain of RNA polymerase 2 plays a very important role in pre mRNA processing. See for example, the phosphorylation of serine 5 of C-terminal domain is essential for the recruitment of capping enzyme to polymerase 2 complex as well as for the allosteric modification of glutathione transferase translation of the capping enzyme. Now, the other important point I want to mention is that not only RNA polymerase CTD acts like a launching pad, for proteins to come do their job and go back. Like for example, the capping enzymes come sit on the RNA polymerase 2 cap the messenger RNA and go back.

The splicing machinery again come back, sit on the RNA polymerase 2, it is more like acts like a launching pad for all these enzymes, but is not just that. It also has an active catalytic role, for example, in the case of capping the C-terminal domain not only just serving as a platform for the assembly of these factors, but it actively enhances the rate of mRNA capping. So, it actually modulates the activity of enzymes which are involved in mRNA capping. So, it has a very active role in all these post-transcriptional processes and mRNA capping is co-transcriptional.

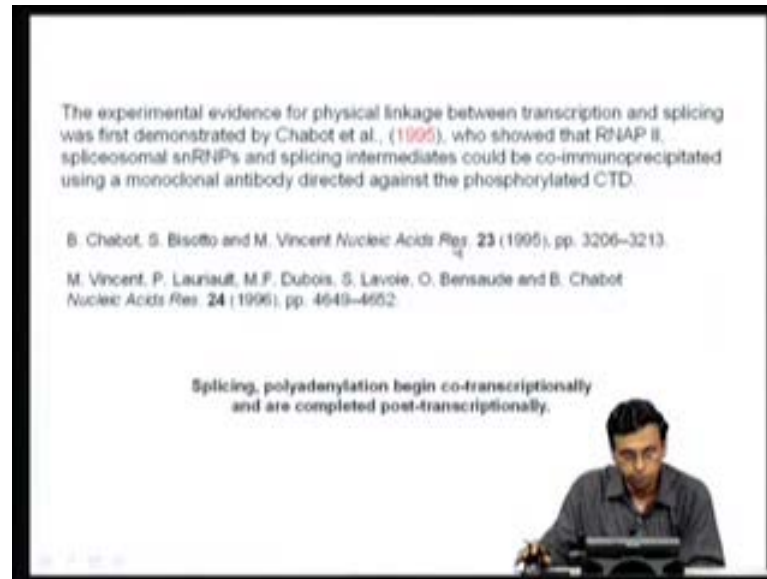


(Refer Slide Time: 38:42)



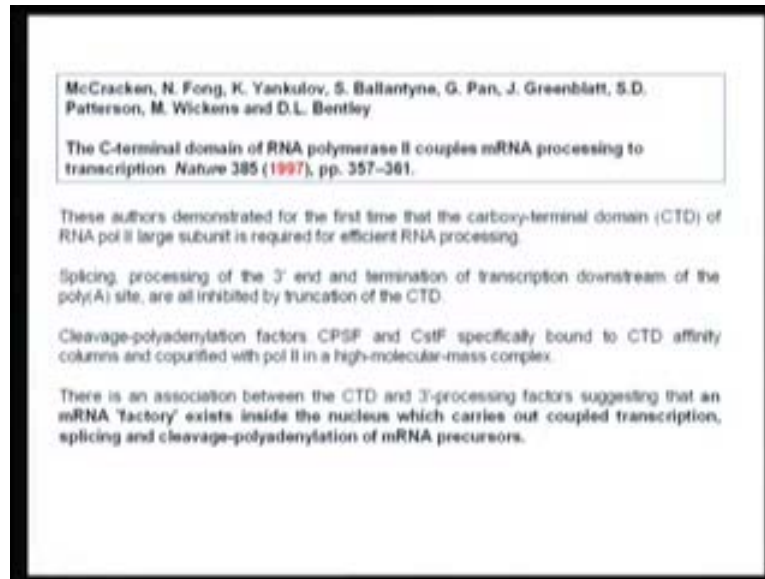
And mRNA splicing and polyadenylation also co-transcriptional. Again very elegant experiments were done for example, here is a paper in 1993 where a person called Greenleaf actually suggested that the negatively charged hyperphosphorylated CTD could act as a ionic scaffold for the recruitment of arginine and rich proteins called as SR proteins which are in actually involved in splicing and phosphorylation states of CTD and the serine residues of the RS domains of SR proteins might regulate these interactions. So, this paper in 1992 through review in tapes actually, talks about how RNA polymerase 2 can play very important role in splicing. And depending on the phosphorylation status of the RNA polymerase CTD as far the phosphorylation status of the serine rich proteins or SR proteins which are involved in splicing, the splicing may be actually regulated by RNA polymerase 2.

(Refer Slide Time: 39:32)



Again the experimental evidence for physical linkage between transcription splicing was actually demonstrated first by Chabot et al in 1995, who actually show that RNA polymerase 2 spliceosomal snRNPs and splicing intermediates could be co-immunoprecipitated using a monoclonal antibodies direct against the phosphorylation form of CTD. So, if we now take cell extracts which contain RNA polymerase 2 with its CTD and other factors and if we now have an anti-body which is specially raised again the C-terminal domain of RNA polymerase 2 and if we do immunoprecipitaion not only the RNA polymerase 2 CTD comes along with that, you can also pull down proteins involved in splicing, indicating that many of the proteins involved in splicing actually interact with the phosphorylated form of RNA polymerase 2. So, these kinds of protein-protein interaction studies clearly demonstrated there is linking between splicing and RNA polymerase 2. So, as I said the take home message is that splicing polyadenylation begin co-transcriptionally and are completed post-transcriptionally whereas, mRNA capping is primarily co-transcriptional.

(Refer Slide Time: 40:36)



These are the two papers which actually demonstrate for the first time that the carboxy terminal domain of RNA polymerase 2 subunit is actually required for efficient RNA processing. One published a nature 1997 the C-terminal domain of RNA polymerase 2 couples mRNA processing to transcription. a very important paper where, they actually shown splicing processing of three prime end and termination of transcription downstream of the poly A state, all are inhibited when you have cells expressing truncated form of C-terminal domain. So, if you read this paper it becomes very clear what kind of experiments they actually done to demonstrate that if you truncate the C-terminal domain of CTD of RNA polymerase 2, splicing as well as three prime end formation of poly A relation, all are affected indicating the CTD plays a very important role and they have also actually done experiments to actually demonstrate that cleavage poly relation factors like CPSF and CstF actually binds to CTD using affinity chromatography studies and they actually co-purify with RNA polymerase 2 as a high molecular protein complex.

So, these authors based on their experiments actually demonstrated proposed in 1997 that there is an association with the carboxy terminal domain of RNA polymerase 2 and three prime passing factors suggesting that inside the nucleus, an mRNA factory exists. And this mRNA factory carries out not only transcription, but also splicing, cleavage, polyadenylation altogether. So, these are all co-transcriptional and they do not happen in an isolated marines like nucleus, but they are all highly integrated.

(Refer Slide Time: 42:18)

**REGULATION OF TRANSCRIPTION ELONGATION**

Eukaryotic transcription factors involved in regulation of transcription elongation:

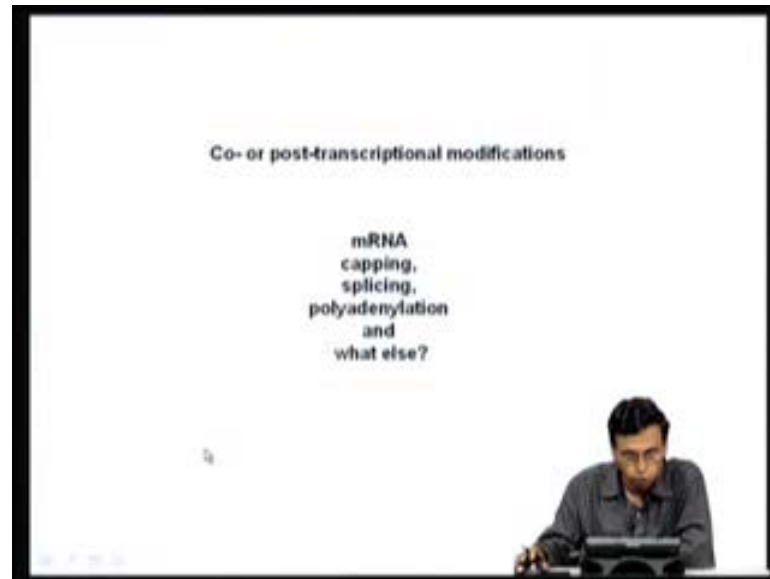
- I: Those that suppress transient pausing and stimulate the rate of transcript elongation by the polymerase: TF2F, ELL, and Elongin
- II: Factors that reactivate a RNA polymerase molecule that has arrested during transcription: the eukaryotic TF2S proteins which stimulate a cleavage activity that is intrinsic to the RNA polymerase in order to rescue it from arrest.
- III: Factors that stimulate elongation on chromatin templates: HMG14, HMG17, FACT and nucleosome remodeling complexes like SWI/SNF.

Other factors: P-TEFb, DSIF, NELF, RIF1, Spt4/5, and Elongator

Rachel N. Fish and Caroline M. Kane (2002)  
Promoting elongation with transcript cleavage stimulatory factors  
BBA 577:287-307

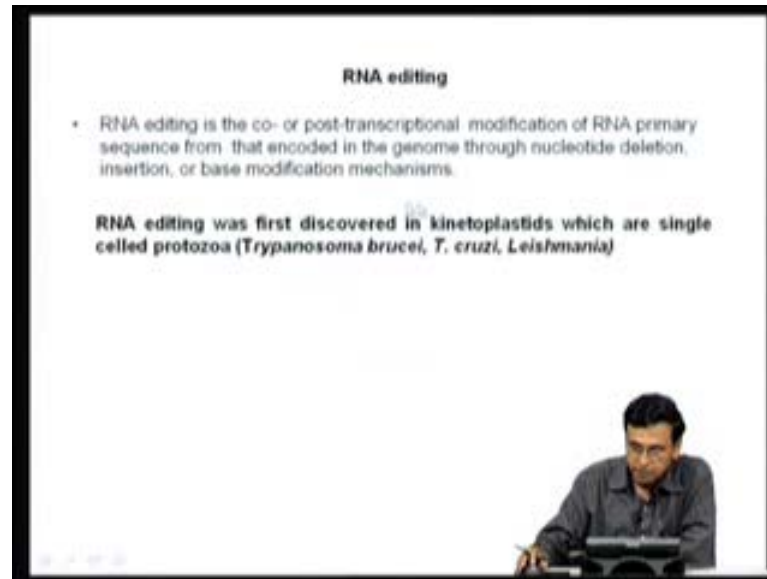
So, these are all very important factors that actually are responsible for some of the major advances that took place in the area of RNA processing. I just shown one or two slides to actually demonstrate that regulation of transcription also takes place at the level of elongation, but I will not elaborate too much on that. Just like we have factors which regulate transcription initiation, splicing, capping and polyadenylation and so on, there are also factors which enhance or inhibit the rate of transcription elongation. All the genes are not elongated at the same time. The RNA polymerase does not make all the RNAs at the same rate and there are factors in the presence of which RNA polymerase 2 either the process to RNA polymerase is either high or slow. For example, there are proteins which actually suppress transient pausing and stimulate the rate of transcription elongation of RNA polymerase. These proteins are called TF 2 F, E L L, Elongin and so on, so forth. Similarly there are factors that reactivate the RNA polymerase molecule which was arrested during transcription these are called TF 2 S proteins.

(Refer Slide Time: 43:34)



And there are actually factors that simulate elongation and chromatin templates called HMG14, HMG17 and FACT so on and so forth. So that, on chromatin templates the rate of transcription elongation can be regulated. There are number of other factors, I will not go to the details. Those who are keen actually can study a nice review which are paid in BBA sometime in the year 2002. Now, I want to spend another ten minutes to just explain another very important RNA processing mechanism. So far, we discussed mRNA capping, splicing and polyadenylation and we made it very clear, these are either co-transcriptional or post-transcriptional modifications, but are there any other RNA processing mechanisms? Are there any other co or post-translation modifications that pre messenger RNA undergoes?

(Refer Slide Time: 44:03)



There are many others but one important that I would like to mention in the next may be five, ten minutes is a mechanism called RNA editing. Now RNA, So, we discussed mRNA splicing, mRNA capping, mRNA polyadenylation. There is also a mechanism called RNA editing, now what is RNA editing? RNA editing is again a co or post-transcriptional modifications of RNA primary sequence from that encode in the genome through nucleotide deletion insertion or base modification mechanisms. That means once the pre messenger RNA synthesized either certain bases can be added or certain bases can be deleted or some bases can be modified. That means, the pre messenger RNA sequence need not be exactly same as the mature messenger RNA because of this modifications. We already know one of the major differences between pre messenger RNA sequence and mature RNA is the mature RNA does not contain an introns.

In addition to that there also certain nucleotides which may be different between the two because RNA editing. So, let us now see how this takes place. Now, this process of RNA editing was first discovered in kinetoplastids which are single cell protozoa, namely the Trypanosoma brucei and Trypanosoma cruzi, Leishmania and etc. So, this process of RNA editing first was discovered in some of the protosomal parasites.

(Refer Slide Time: 45:16)

**RNA editing**

Two types

Base modification (substitution editing) (Adenosine deaminases)

- A to I seen in viruses, human genes
- C to U, U to C seen in chloroplasts, plant mitochondria, human genes

Insertion/deletion editing

- U insertion/deletion, seen in kinetoplastid protozoa
- mono/di nucleotide insertion, seen in *Physarum*
- nucleotide replacement, seen in *Acanthamoeba* tRNAs

Pre-mRNA 5'-GGGGGAGGAGAGAGAAGAAGGGAAAGUGAUUUUGGAGUUAUAGA-3'

Edited mRNA 5'-GGGGGAGGAGAGAGAAGAAGGGAAAGUGUUUAUUUGGAGUUAUAGA-3'

Now, what is this RNA editing? There are basically two types of RNA editing that takes place in eukaryotic cells. One is called as the base modification or substitution addition. These are primarily carried out by enzymes called adenosine deaminases. There are two types of base modifications which can happen. The adenosine residue in the primary messenger RNA can be converted to inosine by deamination reaction. This is primarily seen in viruses, certain human genes and another important RNA editing that takes place is conversion of cytosine to uracil and uracil to cytosine. This is seen in chloroplasts, plant mitochondria and again certain human genes.

So, this kind of modification is called where existing base is either converted into another base. Either A is converted into inosine or cytosine is converted into uracil or uracil converted to cytosine. This is called as the base modification mechanism and these kinds of changes are primarily carried out by a family of enzymes called as adenosine deaminases. The other important method of RNA editing involves the insertion or deletion. Here you are just modifying a base, but here it is actually either inserting or deleting a base. The most common RNA editing that happens is the insertion of U residues or deletion of U residues, primarily seen in the again kinetoplastid protozoan's and certain other cases or organisms, you can also insert results in the insertion of dinucleotides or they can actually result in the replacement of one nucleotide by another. So, all these involve RNA editing. Base modification, insertion or deletion of nucleotides and I just shown one example here.

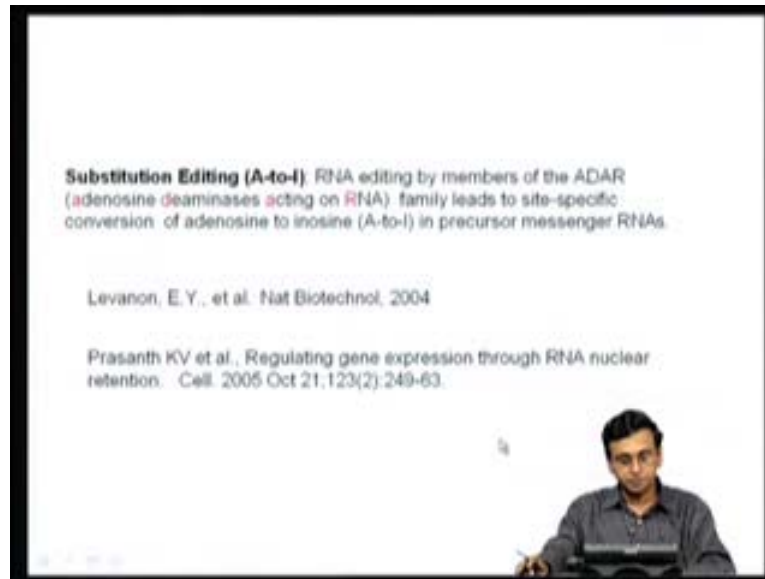
For example, if this is the pre messenger RNA sequence here and if it undergoes an insertion editing what happens, if U residues are inserted for example, you can see here after the triple AG, 2 U's are inserted here and similarly after U G again 2 U residues are inserted here and as a result the open reading frame of pre messenger RNA will be completely different from the open reading frame of the mature messenger RNA.

So, if you look at the genomic sequence you will see this, but if you actually sequence the mature RNA and if you deduce the amino acid nucleotide sequence based on the protein, you will find the deduce them nucleotides sequence based on the sequence does not match with the genomic DNA sequence because of this RNA editing phenomenon. So, you conclude that, how do you conclude that, the gene sequence is different from the mature RNA sequence when you deduce the amino acid sequence from the gene and if it does not match with that of the actual messenger RNA, if sequence the cDNA and the cDNA sequence does not match with the actual genomic sequence you conclude that some kind of a co or post-transcriptional modification has taken place. I repeat again, how do you demonstrate that some kind of a post-translational or co-transcriptional modifications are taken place for a particular gene.

If you know sequence for cDNA because now cDNA primarily made from messenger RNA and if you sequence this cDNA and if you also sequence the gene from the genomic DNA and the genomic DNA sequence and the cDNA sequence do not exactly match, one conclusion is that naturally because introns may be spliced, but another could be some kind of RNA editing might have taken place and as a result where for example, if a base modification as taking place there may be inosine in the cDNA whereas, it may be adenosine in the genomic DNA. Similarly, you find that certain extra U's are there when you sequence cDNA, if that kind of U's or T's are not there in genomic DNA it again means that a post-transcriptional modification has taken place and an RNA editing is probably the cause why the cDNA sequence is different from that of the genomic DNA sequence.



(Refer Slide Time: 48:58)

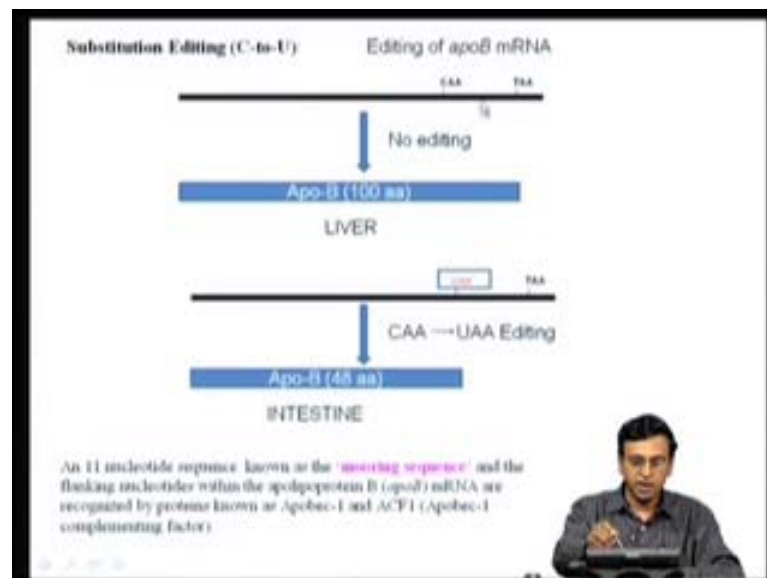


Substitution editing example I told here, the most common is what is called as adenosine and inosine modification. RNA editing by members of ADAR which means adenosine deaminases acting on RNA. This family of enzymes convert adenosine into inosine in some of the precursor messenger RNAs. So, these are actually known as inosine containing messenger RNAs and what is the significance of this? There is a very nice review nature by technology and it turns out the especially in the as early as late of 2005 and the early 2000s, people have now realize that many of the transcribes, see usually after transcription and after RNA processing the messenger RNA comes out of the nucleus, goes in the cytoplasm, gets translated and the protein is made.

This is the common mechanism. But people are found out there are many messenger RNAs which are actually transcribed RNA polymerase 2, but they do not actually leave the nucleus, they stay inside the nucleus. They do not come out. And when they actually found out what they are, they come out only after some kinds of modifications. So, they have actually analyzed what are these messenger RNAs, which are transcribed RNA polymerase 2, but they stay inside a nucleus, they found that many of these RNAs actually contain inosines. So, the adenosine to inosine modification is one of the mechanisms by which transcripts are retained in the nucleus. So, the A to I RNA editing is actually involved in the retention of messenger RNA inside the nucleus. Such RNAs contains inosine residues are usually not seen in the cytoplasm, they are primarily present in the nucleus. So, this is one mechanism by which certain RNA molecules are

selectively retained within the nucleus and they do not come out into the cytoplasm at all.

(Refer Slide Time: 50:53)



So, this is one of the major significance of the major RNA editing mechanism. I will give you one other very important and very classical example of a substitution editing. There in the previous slide, I showed you adenosine to inosine conversion and that kind of RNA editing seems to play very important role in nuclear retention of messenger RNAs. Here is an example where a cytosine is getting converted into U and as a result it results in the formations of two different proteins.

For example there is a gene that codes for a protein called Apolipoprotein B, known as, abbreviated as apoB. Now, if you look at the gene of the Apolipoprotein B gene, there are two sequences which are highlighted here. One is CAA another is TAA, now when there is no RNA editing that is taking place especially this happens in liver the RNA is transcribed and the RNA is translated and since, TAA is a stop code on it, stops here and as a result we generate an Apolipoprotein B protein which is about 100 amino acids long. So, if you purify Apolipoprotein B from liver cells and you find out that this protein contains 100 amino acids and it starts from ATG and ends at the start code on here. Interestingly, in intestine the same gene the RNA that is transcribed from the same gene in liver undergoes a C to U editing, the substitution editing and as a result this particular C is converted into U. So, CAA now has become UAA and that UAA is a stop codon. So,

as a result when this RNA comes from cytoplasm translation does not proceed beyond this point and therefore, you get a truncated protein which is only 48 amino acids. So, you can see from the same gene the gene is the same, whether it is a liver or intestine, but the messenger RNA if does not undergo RNA editing, it generates Apolipoprotein protein of 100 amino acids that is an another tissue because the RNA undergoes editing, it generates a protein of just half the size. So, you can see how from a single gene you can generate two different proteins by RNA editing mechanism.

And I have just mentioned here why RNA editing takes place only this particular CAA is recognized. there is something called as a (( )) nucleotide sequence which is called as mooring sequence, which is in and around this CAA and this flanking nucleotides in and around CAA is within the A polipoprotein B are actually recognized by these proteins involved in RNA editing in the case of Apolipoprotein B these proteins are called Apobec-1 and ACF1, that is Apobec-1 complementing factor. So, these protein factors recognize the 11 base per sequence in and around CAA and as a result they bring about RNA editing especially, in this particular region and therefore, a CAA which coach for an amino acid is converted into a stop codon, as a result translation stops here and you get a truncated protein from a single gene by RNA editing, you generates two proteins in two different tissues.

(Refer Slide Time: 53:52)

MULTIPLE mRNAs CAN BE GENERATED FROM A SINGLE GENE BY PROCESSES SUCH AS RNA EDITING, RNA SPLICING ETC.

**Systematic identification of abundant A-to-I editing sites in the human transcriptome**  
Levanon, E. F., et al. Nat Biotechnol, 2014

They predict 12723 editing sites in 1637 genes and experimentally validate 26 genes.  
They offer many new editing sites and experimental evidence.

Based on EST and other expression data, up to 80% of human genes are believed to produce alternatively spliced mRNAs

With now, genome sequence being out we now know, how many protein coding genes are present and so on, so forth. What now people are now examining is that how many genes actually undergo RNA editing? How many genes undergo splicing? Because by these kinds of post-transcriptional mechanisms although the number of genes may be only 30 or 40000, you can now realize because of alternate splicing, differential splicing, different RNA editing the number of transcription number of proteins, you can generate or many times more than the actual number of genes. Because you can generate more than one messenger RNA, consider more than one protein by many of these post-transcriptional mechanisms. And for example, systematic identification of abundant adenosine to inosine editing sites in human transcriptome. This is published in 2004 and these people predict that there are something like 12723 editing sites in 1637 genes and they are actually experimentally validate 26 such genes. Indicating that there are these many genes that undergo RNA editing and there are these many sites which are potentially, can undergo RNA editing.

And they offer many new editing sites and they actually experimentally evidences how RNA editing may be a much more common mechanism than people previously thought. In fact, based on the EST and other expression data, it is now predicted that up to 80 percent of the human genes are believed to undergo or produce alternatively spliced messenger RNA. See when, human genome sequence or (( )) genome sequence are published and when people realize it that contains only 30000 or 50000 genes people were surprised. But you can see by these kinds of mechanisms, you can generate lot more proteins and lot more messenger RNAs by all these post-transcriptional mechanisms and therefore, from one gene you can always generate more than multiple messenger RNAs and multiple proteins.

(Refer Slide Time: 55:43)

**Sub nuclear structures having a role in transcriptional regulation**

- Cajal Bodies:** involved in small nuclear and nucleolar RNA metabolism, snRNP biogenesis
- Gems:** assembly of snRNP and snRNP maturation
- Nucleolus:** ribosome synthesis and assembly
- Heterochromatin:** inactive chromatin
- PcG bodies:** contain polycomb group proteins (silencing proteins) such as RING1, BMI1 and hPc2
- Nuclear Speckles:** contains groups of pre-mRNA splicing factors
- Paraspeckles:** storage of processed mRNA (Ex. CTN RNA), contains proteins such as PSP1 and noncoding RNAs such as VINCENEA1/Men epsilon RNA
- IGC:** interchromatin granule clusters assembly, modification of pre-mRNA splicing factors
- OPT Domains:** (Otl1/PTF/transcription) appear in G1 but disappear in S phase; contain transcription factors but not RNA processing factors
- Cleavage Bodies:** cleavage and polyadenylation of pre-mRNA processing



I will just take one minute to explain that. So, far if we have been studying in text books you will realize that nucleus contains only what is called as nucleolus, but what studies have now shown is that once people realize splicing, polyadenylation all these things are coupled to transcription, people are actually identified a number of sub nuclear structures. I will not go into the details of this because we will discuss some of these little later time point in the course. I am just mentioned some of these things here, a number of sub nucleus structures have been recognized in the recent studies and we will discuss many of these things later during the course.

(Refer Slide Time: 56:17)

**Transcription and processing are indeed linked physically and functionally in the nuclei of mammalian cells.**

Mitel and Spector 1999 *Nature* 387: 523-527  
Spector 1993 *Annu. Rev. Cell. Biol.* 9: 265-315

Nuclear speckles represent storage sites, or perhaps sites of assembly or recycling of splicing complexes, rather than sites of active splicing.

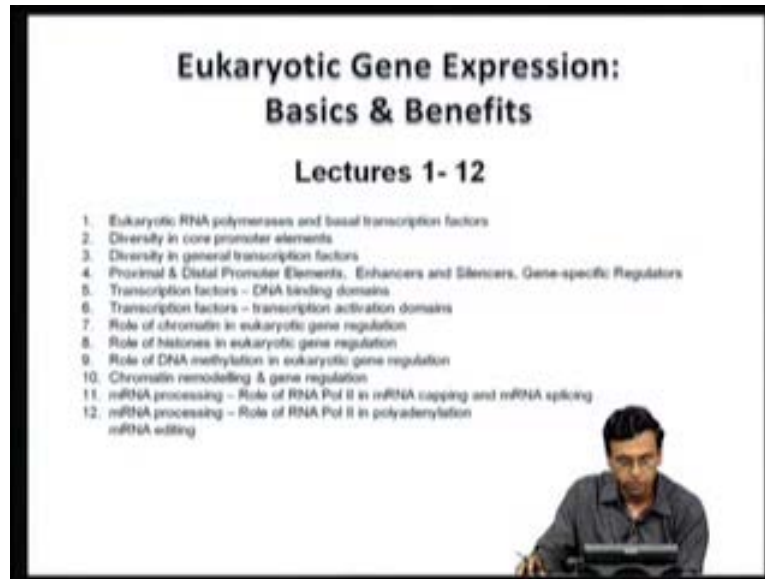
Both hyperphosphorylated RNAP II and polyadenylation factors are associated with periphery of speckles. Transcriptional activation seems to result in a redistribution of factors from the speckles, and indeed RNAP II and the SR protein ASF/SP2 appear to migrate from speckles to sites of transcription.

In cells expressing an RNAP II with a truncated CTD, relocalization of splicing factors to transcription sites does not occur, further confirming a role for CTD in RNA splicing.

Thus the localization and dynamics of transcription and mRNA processing factors within the nucleus is consistent with the functional interactions observed in vitro.

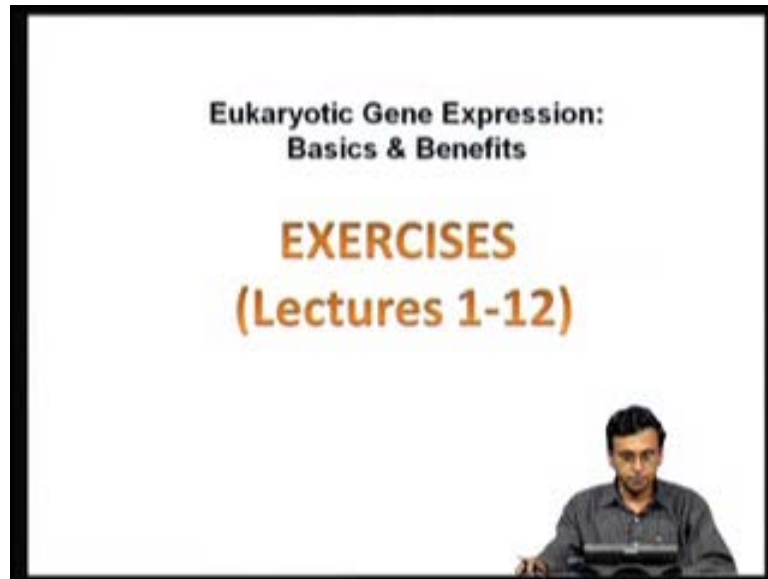


(Refer Slide Time: 56:29)

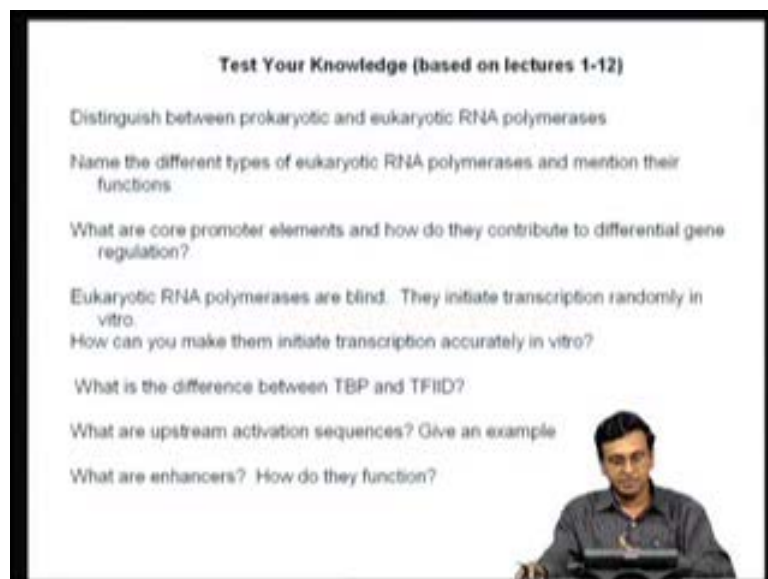


There are something called as nuclear speckles for example, which play very important role in storage of certain RNA molecules, we will discuss this again later during the course. So, we now come to the conclusion of this particular lecture now this is the twelfth in this particular course. So, this lectures from lecture one to lecture twelve primarily we have focused our attention on the synthesis of messenger RNA. How messenger RNA synthesis regulated inside the cells we started from basal transcription factors, promoter elements, we discussed transcriptional factors, then we discussed the role of chromatin, then we studied epigenetic regulation, then we discussed very interestingly role of chromatin and then we discussed in the last two classes, what kind of processings that mRNA undergoes and with this, one module of this course has been completed and these twelve lectures, if you study you will get a very good idea about how regulation of RNA polymerase 2 transcription takes place inside the mammalian cells.

(Refer Slide Time: 57:21)



(Refer Slide Time: 57:35)



(Refer Slide Time: 57:37)

**Test Your Knowledge (based on the lectures 1-12)**


Give atleast three examples of transcriptional activators that bind to specific enhancer elements and activate transcription

Discuss:  
Thyroid hormone receptor – co-repressor interactions  
Activation of transcription by cyclicAMP

What are the major functional domains of eukaryotic transcriptional activators?

Define the function of the following domains:

- a) zinc finger
- b) helix-turn-helix
- c) leucine zipper
- d) acidic activation domain



(Refer Slide Time: 57:39)

**Test Your Knowledge (based on lectures 1-12)**

Functional significance of RNA Polymerase CTD

Name the general transcription factor which phosphorylates RNA polymerase II CTD

What are RNA polymerase II A and RNA Polymerase II O

What are the components of a nucleosome


Histones have only a structural role in cells – state true or false, give reasons

What are core histones and linker histone

Mention various post translational modifications (PTMs) of histones

What is the significance of histone acetylation?

Which PTMs of histones are involved in negative regulation of gene expression?





(Refer Slide Time: 57:46)

**Test Your Knowledge (based on lectures 1-12)**

What are HATs and HDACs? How are they recruited near promoters?


What is histone code?

How are cytosines methylated in a chromatin template?

How are methylated cytosines recognized in a chromatin template?

What is the interrelationship between histone deacetylation and DNA methylation?

What is epigenetic code?



(Refer Slide Time: 57:49)

**Test Your Knowledge (based on lectures 1-12)**

Distinguish between Transcription initiation complex and transcription elongation complex

What is mRNA capping and discuss its functional significance

What is the role of RNA polymerase II in mRNA capping?

How do you prove that RNA Pol II CTD is involved in mRNA splicing?

Transcription initiation and polyadenylation are coupled – Discuss


What is DNA foot printing?

What is an electrophoretic mobility shift assay?

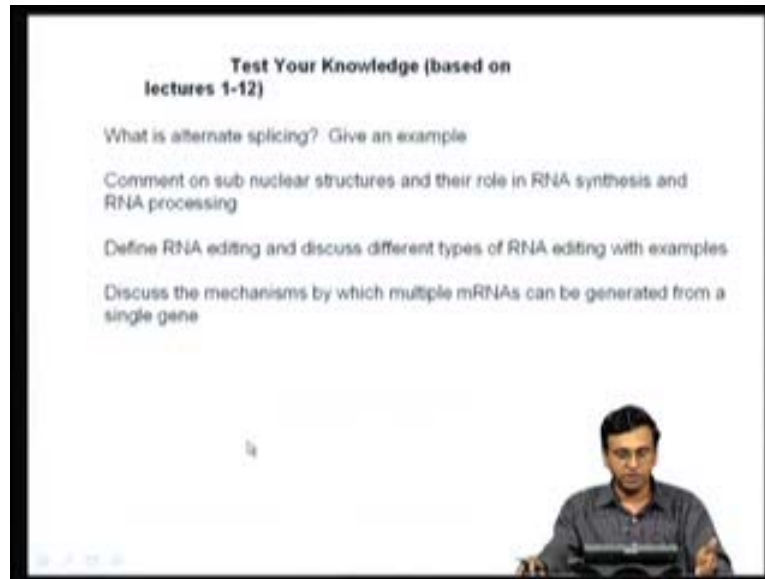
How do you assay the activity of a promoter in vivo and in vitro?

What is a G-free cassette?

A reconstituted cell-free transcription system containing TBP instead of TFIID does not support activator-dependent transcription - Discuss



(Refer Slide Time: 57:54)



So, based on all these two lectures I have listed a number of exercises here. So, if you have really understood all these twelve lectures and have really got a good knowledge about messenger RNA synthesis and regulation, you should be able to answer many of these exercises. I have listed a number of questions here. I suggest you go through many of these questions and then see where you can answer many of these questions. If you cannot answer, go back to our lectures and read some of these things and see whether you can answer some of these questions and this is the way by which you can get a very comprehensive knowledge of how messenger RNA regulated inside the cells. I will not go through many of these questions, but you can read up and see whether you are able to answer some of these questions.