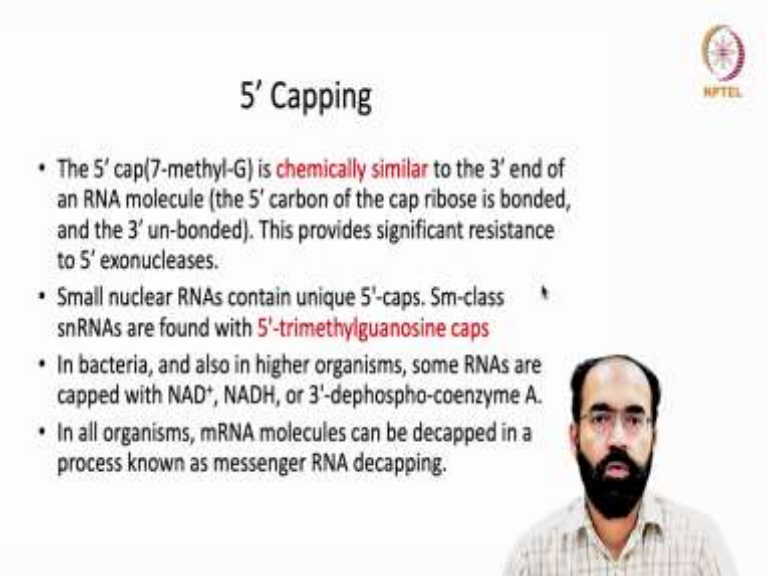
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

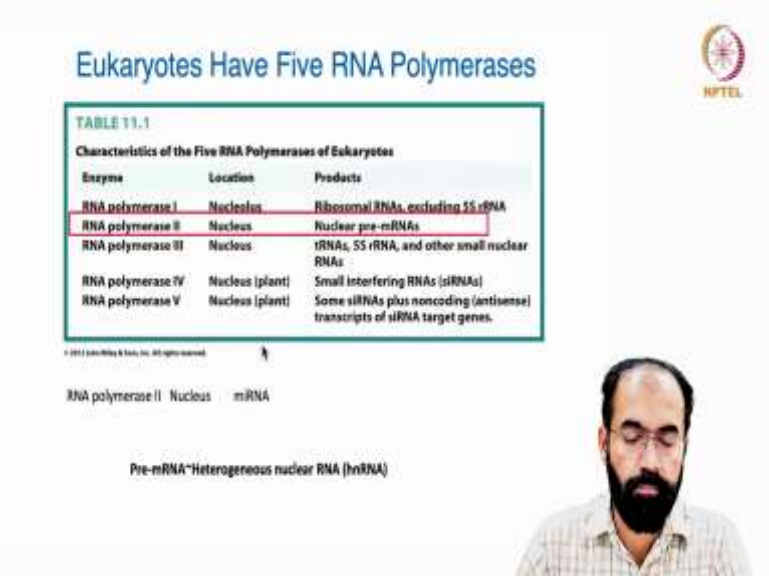
Lecture - 14
RNA Transcription: Different Polymerases

(Refer Slide Time: 00:21)



Welcome back to another session of RNA Biology. So, we were here in the previous class we are talking about the 5’ Cap. So, 5’ Cap is a structure that stabilizes the RNA from degradation. So, it can be any molecule any molecule means not any random molecule there are different possibilities of 5’ Capping exist, but in general it is a 7 methyl Guanosine cap that exists as the 5’ cap in every RNA molecule.

(Refer Slide Time: 00:58)



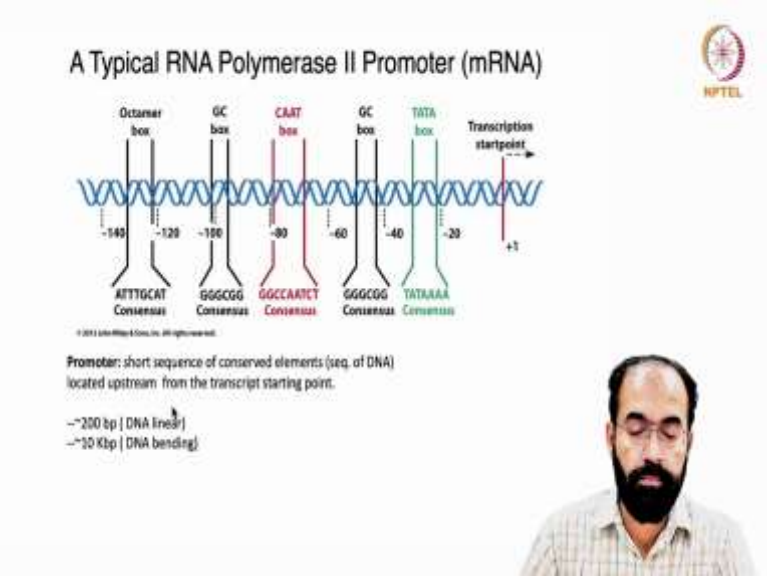
So, eukaryotes have around 5 different types of RNA polymerase I polymerase II polymerase III polymerase IV and polymerase V. So, RNA polymerase I is predominantly in the Nucleolus, nucleolus is a structure that can be seen densely stained in the cross section of a cell etcetera.

You can never see a nucleolus lighter in stain, it is because the a ribosomal RNA production is quite high in it is quite high in the nucleolus, because of which there will always be a intense staining. Now RNA polymerase II which is mainly in the nucleus throughout the nucleus and the Nuclear pre-mRNAs are transcribed with the help of RNA polymerase II.

RNA polymerase III is mainly transcribing transfer RNA and the 5 S ribosomal RNA, RNA polymerase II is mRNA RNA polymerase one ribosomal RNA with an exception of 5 S ribosomal RNA and RNA polymerase III is handling tRNAs, 5 S ribosomal RNA and other small nuclear RNAs.

Whereas, polymerase IV and polymerase V are a feature in plants and what they do they produce siRNA they are mainly meant for silencing some of the mRNA transcripts in plants the gene regulation through RNA mediated gene silencing is quite high, that allows protection from invading pathogenic bacteria and viruses. So, polymerase V is also undertaking siRNAs and some non coding RNAs which can help in silencing of the target genes.

(Refer Slide Time: 03:15)



And if you look into a typical RNA polymerase, RNA polymerase II is studied well studied as a candidate gene simply because the gene regulations are more applied to RNA polymerase II than polymerase I or polymerase III.

It is because a cell have got it is identity mainly because of the total number of protein, total diversity of protein present in that cell. So, RNA polymerase II being the polymerase considering or doing the transcription of gene encoding mRNAs it is transcription or it is study or it is regulation becomes more important than polymerase I and polymerase III.

So, in a cell there are group of proteins called specific transcription factor or tissue specific transcription factors that allow where a polymerase II should bind and start it is transcription. If you look into a typical RNA polymerase II promoter you can see a plus 1 sequence which is the place where the transcription initiation happens that is the first base of the RNA or this is a picture of the DNA. So, the first base of the RNA encoded on the template strand.

So, that is what plus 1 and downstream of that you have plus 2 plus 3 plus 4 like that, upstream of that you have got minus 1 minus 2 minus 3 like that up around minus 20 region you have the TATA box and TATA box have a consensus sequence TATAAAA.

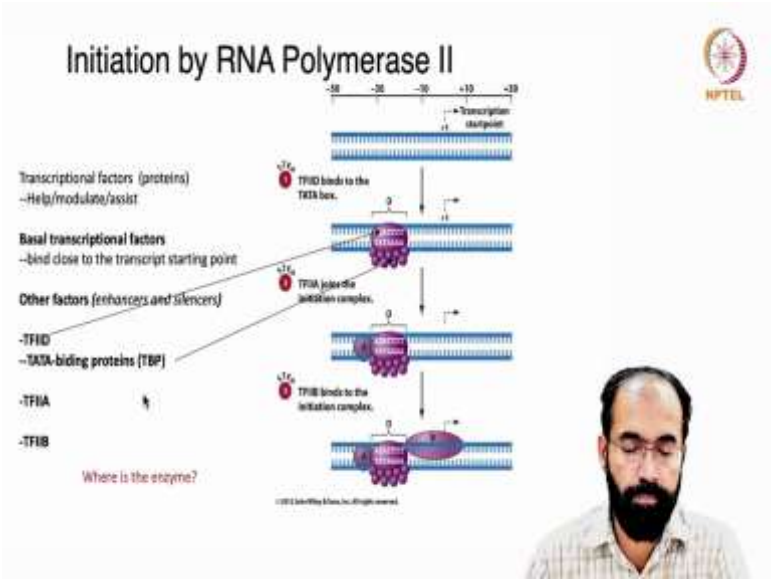
Consensus basically mean something predominantly seen or commonly seen that is the meaning of consensus, this does not mean that every gene must have only this sequence it will allow maybe one base alteration etcetera. But usually unlike the neighbouring parts this consensus sequence are more or less going to be the same and much upstream up around minus 40 region you have a GC box which has a sequence of GGGCGG that is the consensus.

And much upstream of that you have CAAT box and it has a consensus of GGCCAATCT and much upstream of that you have another GC box that is having a same sequence as the GC box you saw earlier, that is GGGCGG that is the consensus sequence of a GC box and then much upstream of that you have an Octamer box and Octamer box have a sequence of ATTTGCAT that is the consensus sequence.

So, this comprise the so called promoter region or promoter sequence. Why this is important? Because around 200 base pair long region can contain the necessary elements required for acting as a promoter and it can stretch up to around 10 kilo bases in some rare cases the promoter region can stretch up to mega bases.

That means, 1000 kb 1000 base 1000 nucleotide is 1 kb 1000 kb is 1 mb. So, it can be stretching much upstream also. So, we cannot define a specific boundary common to all the genes it vary from gene to gene.

(Refer Slide Time: 07:07)



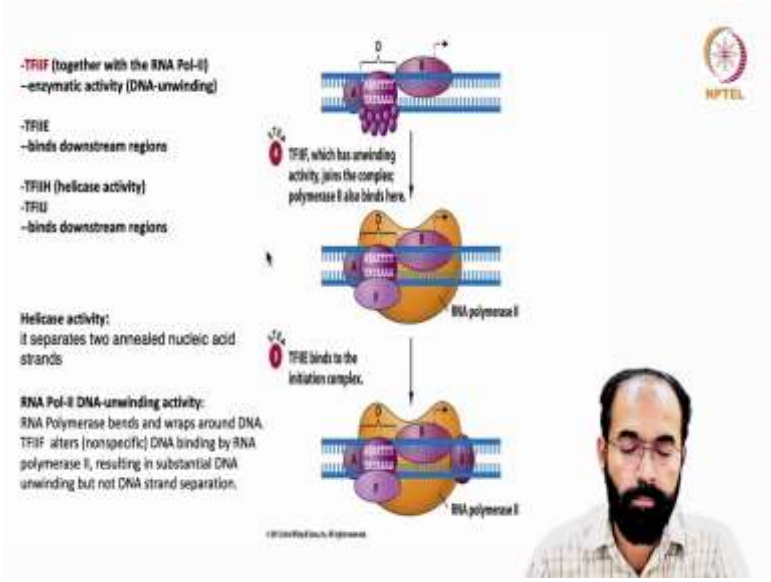
So, if you see initiation of transcription by RNA polymerase II it starts with one component of the RNA polymerase, there are t12 different subunits are there in the RNA polymerase complex. So, TFIID is a initiator molecule of the RNA polymerase II transcription it binds on to the TATA box region of the promoter and transcription factors are nothing, but proteins that help or modulate the transcription of mRNA in this case and they are called Basal transcription factors.

Basal transcription factors are essential they are inevitable for the transcription to take place and they usually bind pretty close to the transcription initiation point. And there are many other factors that contribute to the regulation of transcription they are called enhancers and silencer enhancers accelerate the transcribability of a gene and silencers reduce the transcribability of a gene.

TFIID is the first factor that binds and that is further assisted by other factors and TFIID is also called as TBP or TATA box binding protein this is joined by TFIIA and this is further joined by TFIIB. So, each of these factors are joining sequentially all these factors cannot join together because that will cause lot of steric hindrance and disturbance and also each of this factor need to have a specific stoichiometry in the DNA, which is possible only when one factor is bound.

If TFIIA is coming to a promoter at first then it may not find this spot attractive or find this spot worth to start a transcription initiation. So, the presence of TFIID is very important for the TFIIA to find this place attractive.

(Refer Slide Time: 09:22)



And then TFIIF together with the RNA polymerase II causes the enzymatic activity that is the DNA unwinding activity Then TFIIE joints that binds to downstream regions of the transcription initiation region, then to TFIIH which has got helicase activity and then TF2J that has got a binding capability to the downstream regions.

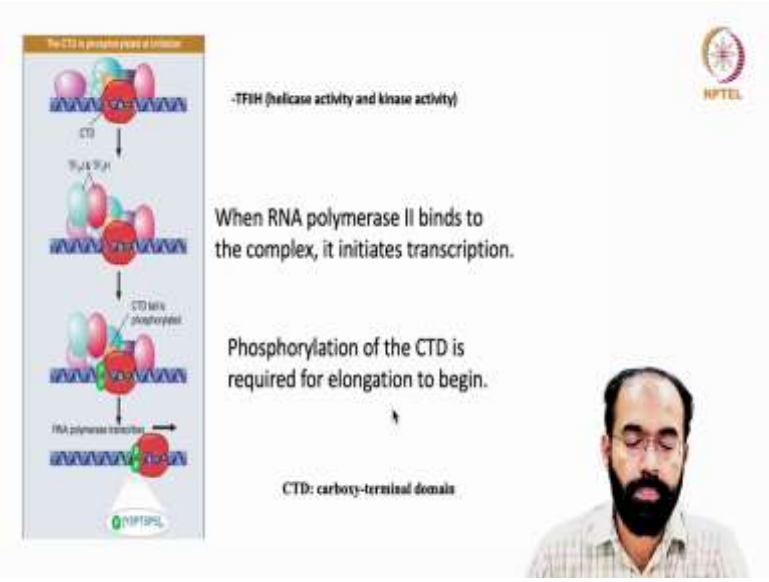
Then comes Helicase activity, so helicase activity is an important feature that is required for the strand separation, separation of 2 DNA strands without which you cannot access one template strand which makes copy of which makes a copy of the DNA RNA copy from a DNA template.

So, RNA polymerase II has to have the unwinding activity and then upstream of that it has to continue with the rewinding of the opened-up DNA double helix. So, RNA polymerase bends and wraps the DNA and TF2F alters the non-specific DNA binding by RNA polymerase II. So, it also imparts some specificity to the region and this results in substantial DNA unwinding but not DNA strand separation. Strand separation it can be facilitated if there is an unwinding happens.

So, like we already know DNA is not like a railway track not although it is double stranded it is not like a railway track, it is bend around like a spiral staircase. So, you have to unspiral it, so that it will look like a railway track and now this strands has to be separated.

So, that is what happens as a step 4TF2 F joints and that unwinding activity is very important for the polymerase II to initiate and then TFIIE binds to the pre initiation complex. And the RNA polymerase II you can see as a big yellow color structure here you can see and the other factors are small factors like A F E B etcetera and d they are all present in this complex.

(Refer Slide Time: 11:56)



And then comes the most important step that is the CTD Phosphorylated initiation. How does this happen we will see them in detail. TF2H has got the helicase activity we already know and it also have got a kinase activity. What is kinase activity?

Kinase activity allows the protein or an enzyme what you are talking about can catch hold of a ATP molecule and add the phosphate group to a target. And in this way what happens the molecule will get energized and it can change it is conformation and it can behave in slightly differently than what it is already doing.

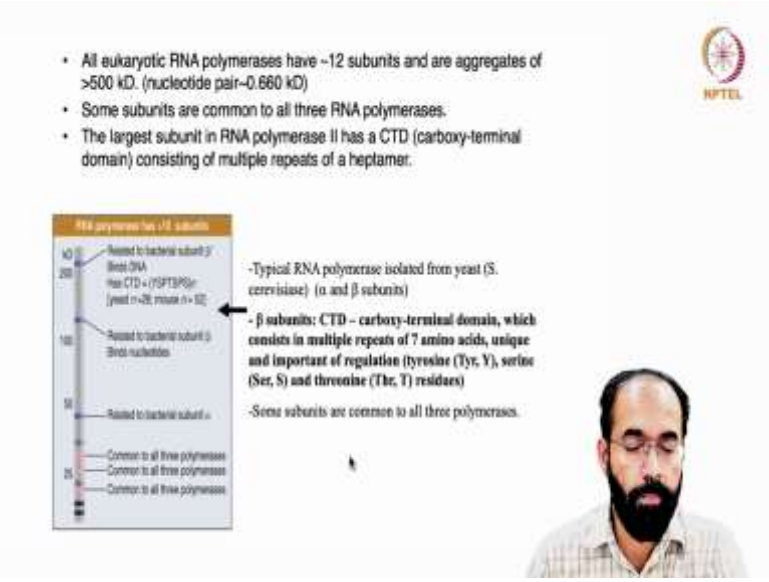
So, the kinase activity is very important. So, when the RNA polymerase II binds to the complex it initiates a transcription which is the complex made of TFIID A H like that. So, the phosphorylation of the C terminal domain that is called CTD is required for the elongation to begin.

So, what is C terminal domain? C terminal domain is the carboxyl terminal domain every protein have got 2 domains one is an N terminus and another is a C terminus because

every amino acid we have got an amino group and every amino acid you have got a carboxyl group, just like the 5'end 3' end.

So every protein starts with ANH 2 end that is where corresponding to the 5'end of an RNA you have got ANH 2 that is located in the N terminus and towards the 3 prime end or towards the end of the protein you have got a carboxyl group. So, that is called C terminal of a protein. So, we call it as CTD carboxyl terminal domain.

(Refer Slide Time: 13:57)



So, all eukaryotic RNA polymerase have about 12 subunits and they are aggregates of the order of 500 kD and this is a pretty large molecule and some subunits are common to all 3 RNA polymerase like is like polymerase I polymerase II and polymerase III can have common subunit.

And the largest subunit in RNA polymerase II has a C terminal domain and we were talking about the kinase action onto the C terminal domain for the initiation or the movement of the transcription. So, the carboxyl terminal domain it consists of several repeat elements of a heptamer hepta means 7.

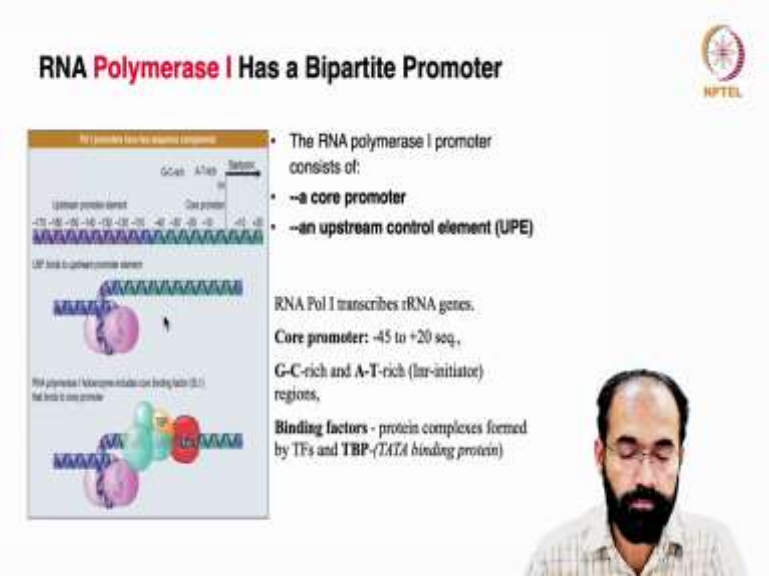
So, you may have heard about this mono di tri tetra penta hexa and hepta. So, this heptamer is consisting of a unique set of amino acid that is present in the C terminal domain. So, what you can see that the c terminal domain which consists of multiple repeats of seven amino acid they are unique and what are they tyrosine, serine, threonine

these are repeated up to 7 times and these ones are which is going to be important for the phosphorylation event.

So, this is an RNA polymerase that is multiple subunits are seen and the CTD is specifically seen here in the largest molecule in this is a western blotting YSPTSPS. So, these are single letter code for different amino acid like tyrosine Y, S for serine and P for proline and T for threonine S for serine like that it continues and that repeats multiple time in yeast it is around 26 times in mouse it is around 52 times like that.

So, this domain are very important for the phosphorylation event without which the RNA polymerase will not go downwards or downstream.

(Refer Slide Time: 16:12)



So, now if you look into the RNA polymerase I we have seen RNA polymerase II we can look is there any unique feature in the RNA polymerase I, this is called a bipartite promoter. That means, it has got sequence elements upstream as well as downstream and this promoter has a core promoter region and an upstream control element UPE.

So, upstream control element and the core promoter both are important for the functionality of the polymerase I, but it is less complex compared to polymerase II and this transcribes RNA polymerase I transcribes ribosomal RNA genes except 5 S 5 S ribosomal RNA and this core promoter consists of minus 45 till plus 20. So, it is

interesting to note that even after plus 1 the promoter region is expanding in this case of polymerase 1.

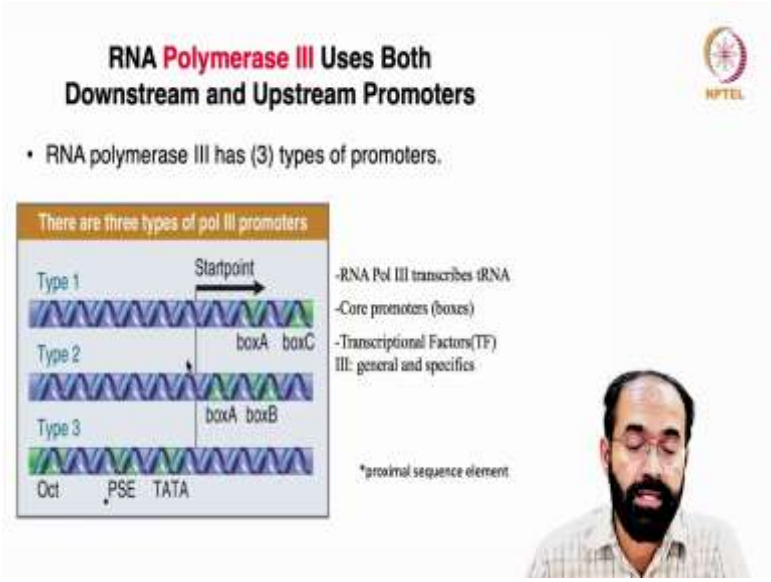
And it has got G-C rich and A-T rich regions and they also have got specific inter initiator regions and that allow the specificity to these promoters. And the binding factors are typical as you saw in case of RNA polymerase II and this complex are formed by transcription factors especially the TBP.

That means, data binding proteins becomes handy even in polymerase I and it is cartoon is shown here in the left-hand side you have this DNA and there is a start point here and it has got an at rich immediate upstream region, because the reason you know because that place is much easy for opening up just like you saw in the case of transcription termination.

A-T rich region in the promoter allows the opening of the DNA strand much easily or the separation of DNA strand much easily and immediate upstream you have G-C rich region and reason is quite obvious because G-C rich region do not allow further opening of the bubble you do not want to fall apart. So, you need you want to have some what you call some strength. So, at rich region upstream of that you have a G-C rich region, so that it do not get opened up further.

And then you have got the upstream promoter element which can spread much upstream which is called basically an upstream control element and once it is bound this protein factors are bound then it can form a loop it can form like a knot like structure, which in turn attracts more and more transcription factors and then it can act like a holoenzyme and this core binding factor facilitate the transcription initiation, in the case of polymerase I.

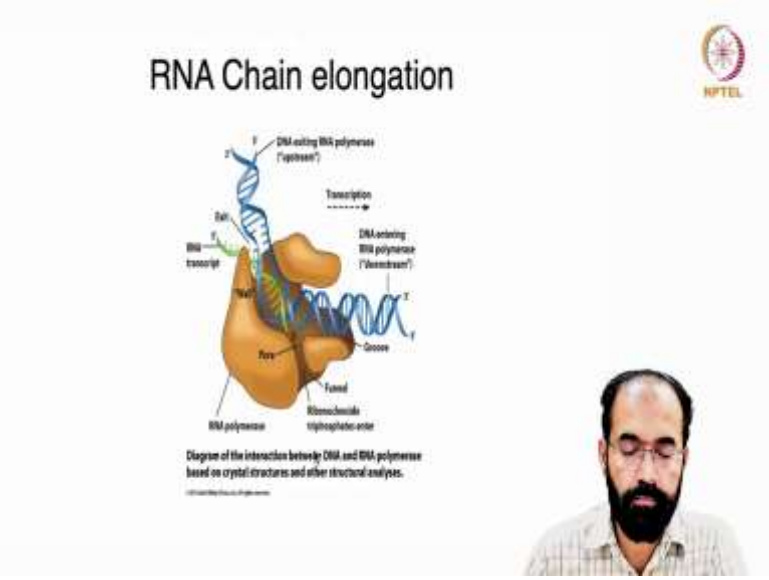
(Refer Slide Time: 19:07)



And now if you look into polymerase III it uses both downstream and upstream promoter sequences and polymerase III has 3 different types of promoters are there, like you can see type 1 type 2 and type 3. It is rather simple compared to polymerase II promoter and RNA polymers 3 transcribes a predominantly transfer RNA and also 5 S ribosomal RNA.

And this core promoter are seen in different names called boxes box A and box C is seen in type 1 box A and box B is seen in type 2 and there is specific domains are seen in the case of type 3 that is an Oct domain and PSE domain and the typical TATA box they are seen in type 3. So, transcription factors are quite unique in the case of pol III, because they can identify different boxes based on which the subcategory of pol III promoters are present.

(Refer Slide Time: 20:20)

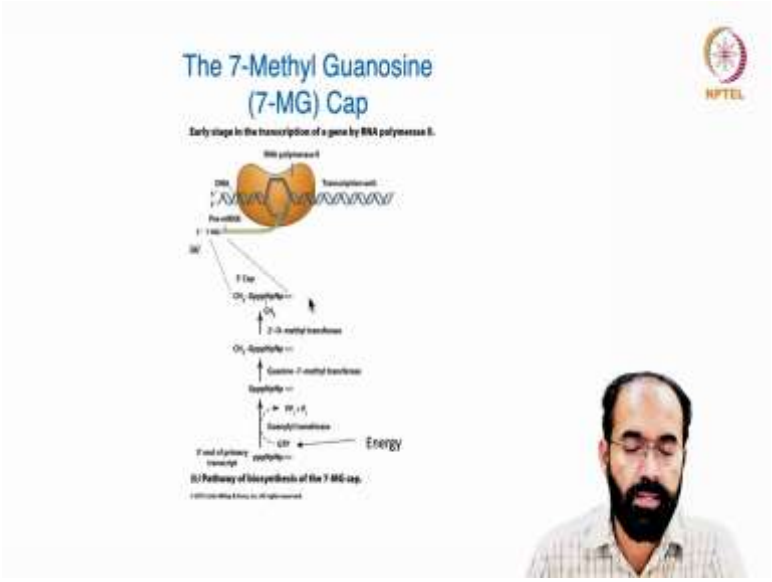


So, if you look into RNA chain elongation, RNA chain so far we have seen chain initiation that is the assembly of transcription factors are important for the chain initiation and the elongation when it comes about the elongation this require the phosphorylation of the CTD without which it will not kick start.

So, the existing DNA polymerase upstream has to retain it is double helical structure and then downstream it has to again retain it is double helical structure, in the middle it opens up the transcription bubble happens and you can see the RNA polymerase have got some lumen a loop region through which the RNA growing RNA transcript comes out that you can see in this green color.

The DNA strand is shown in blue color and there is a specific opening through which the ribonucleoside triphosphates enter, that is the raw material for the polymerization event and there is also a groove for the entry and exit of the DNA. So, it has got grooves for opening of the DNA entry and exit and also opening for the RNA exit and also entry for the polymerase sub polymerase raw material, that is the ribonucleoside triphosphates that is the raw material for the polymerization event.

(Refer Slide Time: 21:50)



And the important thing we need to know that is the soon after the production of the RNA, RNA has been completed. Soon after the RNA polymerase is released from the complex it need to have the 7 methyl cap. So, early stage in the transcription of a gene in this case it is RNA polymerase II it has to undergo capping, without which the newly formed RNA will be degraded.

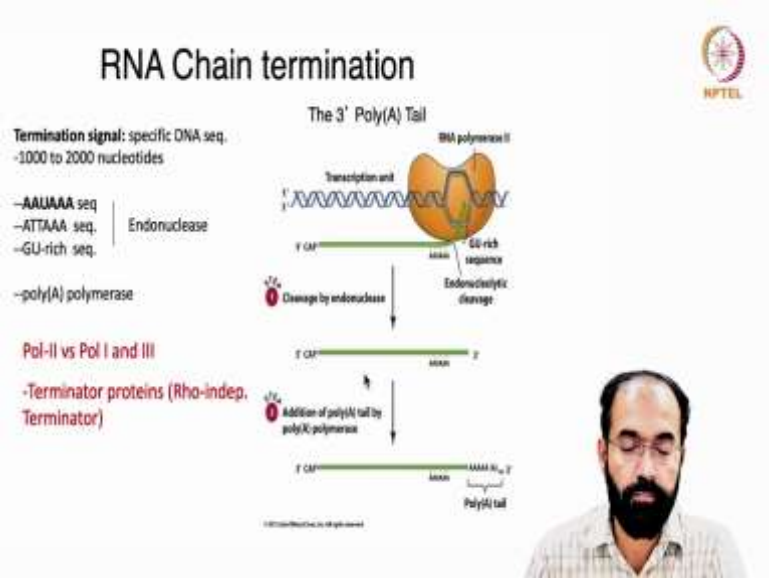
So, you can see in this picture this yellow color bubble is the transcription bubble and the RNA polymerase II, you can see the DNA opened up and then the nascent RNA that has come out and this RNA has to undergo the capping event. And how does it happen?

So, at the final step is the cap. So, to start with it starts from step number one and you have a utilization of GTP and you have to use Guanylyl transferase that will convert GpppNpNp a intermediate which is acted upon by guanylyl 7 methyl transferase and you end up getting a methylated GpppNpNp. What is Gpp indicates?

That means, it has a guanosine with all the 3 phosphate intact and downstream you have a nucleotide phosphodiester backbone again nucleotide, phosphodiester backbone that is why NpNp stands for. Then a third enzyme acts that is 2 prime o methyl transferase which allows another methylation onto the first base that is written as N, N stands for nucleotide first base.

So, there is a methyl group and this already had a methyl group and this is now added as a cap which we call it as 7 methyl guanosine cap on to the RNA So, it is a multi step process 3 enzymes are involved for the addition of a 7 methyl cap onto the RNA.

(Refer Slide Time: 24:02)



So, then comes RNA Chain termination. So, the termination signal has to be very specific because the specific DNA sequence can span around 1000 to 2000 nucleotides and there are specific sequence that can act as a termination signal, that is what are the AAUAAA that is one termination signal or ATTAAA.

So, TT can be UU in the case of RNA or when it is in DNA it is T when it is in RNA it is U. So, these sequences have to be present which will be recognized by the endonuclease, so that it can chop at the specific location. So, that the growing RNA can fall apart from the transcription bubble and poly A polymerase is next enzyme that comes handy for the 3‘ elongation 3 ‘ end elongation.

So, pol II pol I and pol III all of them have to be having a proper termination event without which the transcription will continue, let us see in detail. So, 3 prime poly A tail has to be formed soon after the transcription termination. So, there is a 5’ 7 methyl gaunosine cap is there then there is a AAUAAA signal is there and soon after that around 16 to 18 base downstream of this signal you have an endonucleolytic cleavage happen.

And much downstream you have a GU rich sequence GU rich sequence will prevent further movement of this transcription bubble and as soon as the RNA is cleaved then it has to be recognized by the poly A polymerase. So, that downstream of this Poly A signal AAUAAA that is Poly A signal, that is recognized by the Poly A polymerase and it will add a Poly A tail. Remember to distinguish between poly a signal and Poly A tail.

Poly A signal is a sequence that is unique as you saw here AAUAAA or AUUAAA or in DNA it is AATAAA or ATTAAA. So, only 2 sequences are predominant Poly A signal which is recognized by the endonucleases to cleave and which in turn is recognized by Poly A polymerase for the addition of Poly A tail.

Once poly A polymerase have acted then the RNA is now having a Poly A tail functional Poly A tail. So, with this I end this topic of transcription termination and we will resume in another class.

Thank you.