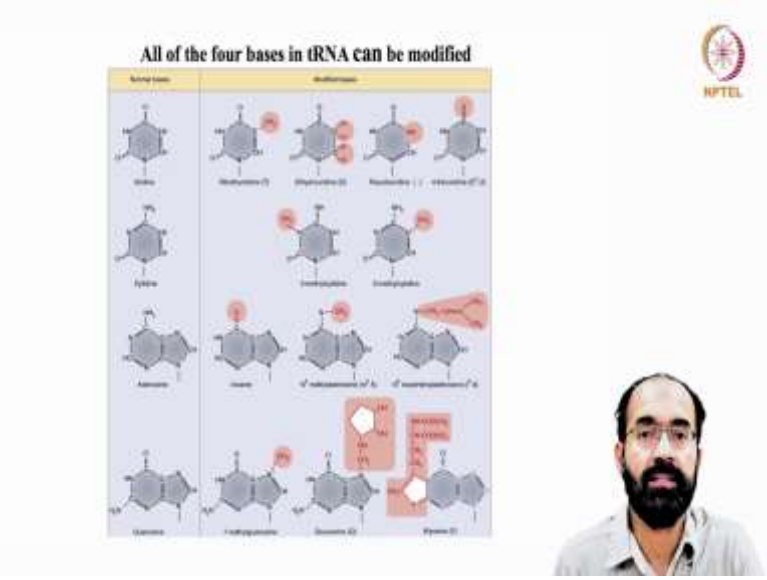
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 16
RNA Processing and Life Cycle: RNA Splicing

(Refer Slide Time: 00:22)



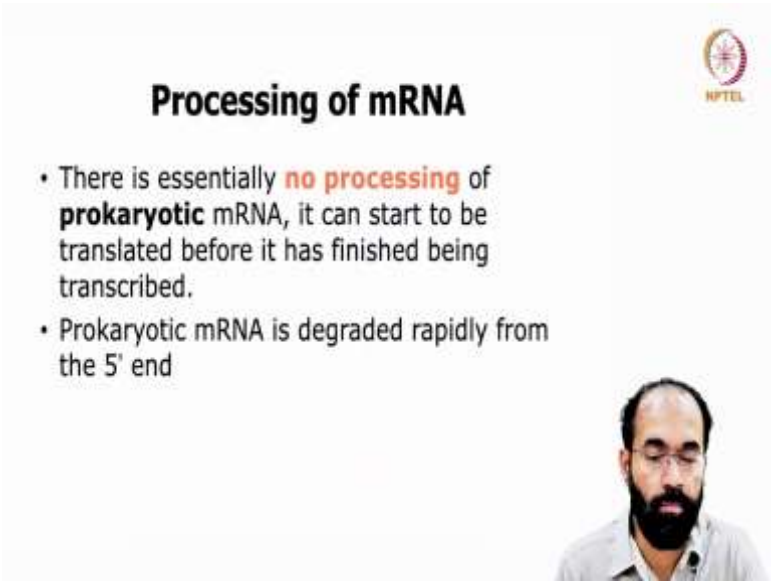
Hello everyone, welcome back to another session of RNA Biology. So, in the previous class we were discussing about the possible bases that can be modified in the tRNA because we have seen that tRNA can undergo post-transcriptional modification especially for enzyme recognition, protection from the nucleases etcetera.

So, the uridine base can be changed into ribothymidine, dihydrouridine and pseudouridine, thiouridine etcetera. So, like that, cytidine can be modified into 3 methyl cytidine, 5 methyl cytidine, etcetera. Adenosine can be changed to inosine and N N 6 methyl adenosine and isopentenyl adenosine etcetera.

So, names are not important. What is important is certain bases in certain arms or certain regions of the nucleic acid is vulnerable to change and such changes provide some protection to the molecule from degradation and also it will become a substrate for enzymes which are essential for their further journey of the molecule.

Say if a tRNA is not aminoacylated, it is basically useless. So, this is what we should keep in mind that it can always undergo modification. Now, let us come back into the routine processing of the mRNA and we will understand them more in detail.

(Refer Slide Time: 02:05)

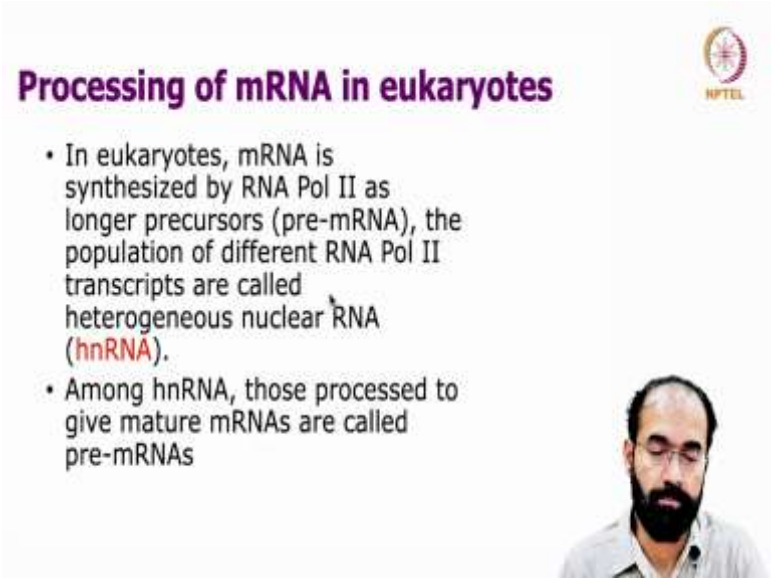


So, in bacteria, there is essentially no processing occurs, there is no processing of mRNA is necessary. Of course, it need to have some you know some modifications that occurs in the 3 prime tail etcetera. We will see them one by one, but there is no extensive processing as it occurs in the case of eukaryotes. And especially the translation, bacterial translation can start even when the RNA is not completed its transcription.

So, this we have seen it that coupled transcription and translation. This provides protection to the RNA and also enhances the rate of production of a protein. However, once that RNA is completed its synthesis then when it is detached from the DNA then this RNA is vulnerable for degradation and of course, some protection mechanism especially polyadenylation etcetera comes into picture.

So, prokaryotic mRNA is degraded rapidly from the 5 prime end when not in use or when not in occupied by ribosomal machinery.

(Refer Slide Time: 03:23)



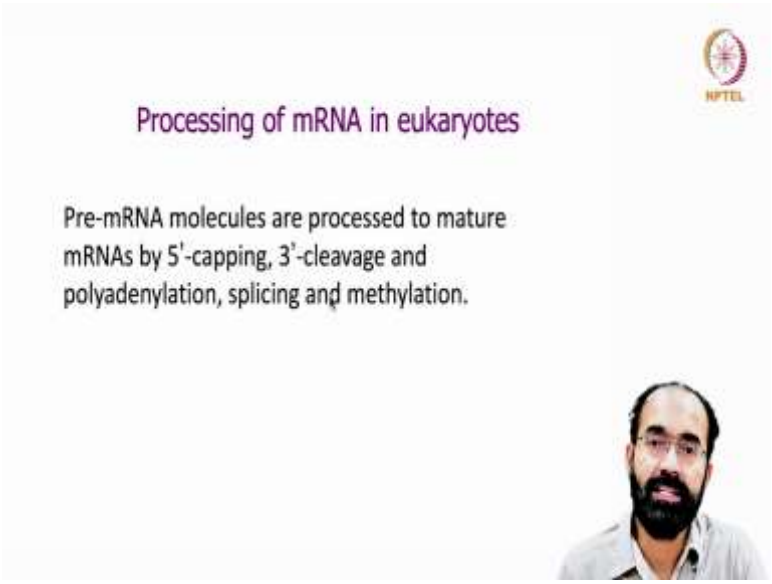
So, let us see how the processing of mRNA takes place in the eukaryotes. Eukaryotes the mRNA is synthesized mainly by the RNA polymerase II. We know that total 5 RNA polymerases are there. Polymerase II is the one which governs the production of mRNA in eukaryotes. And this give rise to extremely long pre-mRNAs. Why they are long? Because they have introns in between.

So, eukaryotic mRNA majority of them will have introns and these introns are pretty large compared to exons and this makes the pre-mRNA pretty large. And the population of different RNA polymerase II transcripts are called heterogeneous nuclear RNA, hnRNA because they can associate also with specific protein making them hnRNPs.

And this provides stability and this hnRNAs are always found to be interacting with some proteins to prevent their degradation and also for recognition by downstream events. If an RNA is a linear strand which do not have its proper folding; folding here what I mean to say is secondary structure then it will not be recognized as a bona fide molecule for downstream processing.

So, RNA secondary structure is very very important. So, it is applicable even to hnRNAs also. And among the hnRNA those process to give mature RNAs are now called pre-mRNAs. Why they are called pre-MRNA? They have introns still present that have to be removed and then become proper mRNA that will be migrating to the cytoplasm.

(Refer Slide Time: 05:17)

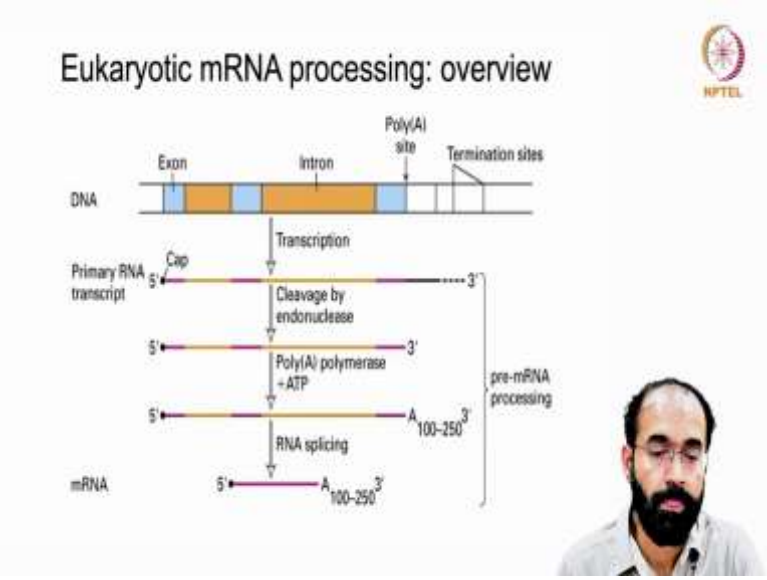


So, let us see how the processing of mRNA takes place in eukaryotes. The pre-mRNA molecules are processed to mature mRNA usually starts with a 5-prime capping. We have seen 7 methyl guanosine cap that is added by three enzyme or 3-step enzymatic process the 7 methyl guanosine cap is added because of which the 5-prime now resembles a 3-prime end because the 7-methyl guanosine is adding to phosphate group with a phosphate group of the guanosine.

So, phosphate group phosphate group pairing is taking place and this leads to a 3-prime OH exposed to the 5-prime end to deceive the degradation machinery. And at the 3-prime end once the transcription is completed at the 3-prime end it has to have a polyadenylation event that is a stretch of A-tails. It can be up to 200 A's that can be added after which the RNA will undergo splicing exons have to be spliced together and introns have to be spliced out.

And some of the RNA have to undergo process called methylation that allows proper localization and stability of the RNA and we will see them case by case as we progress in the class.

(Refer Slide Time: 06:44)



So, let us see in a cartoon manner how eukaryotic mRNA processing takes place. What you say in the topmost panel is the DNA which is shown in a double stranded structure like a railway track. You have exon which is given blue in color and introns are given yellow in color and you can see exons are intervening introns present exons are interrupted by intronic sequences.

And then after the poly A site, poly A site is a site where the transcription will come to an end which will be recognized by the poly A polymerase enzyme and it will add a stretch of A's. It will not add G, it will not add C, it will not add T usually add only A type that is how system have evolved and downstream of that you have got some termination sites.

During the transcription this will become handy because remember the poly A is not present in the DNA. You have a poly A site and then you have a termination signal which allow the transcription termination and the poly A site will be utilized by the poly A polymerase to recognize where to start adding the poly A tail.

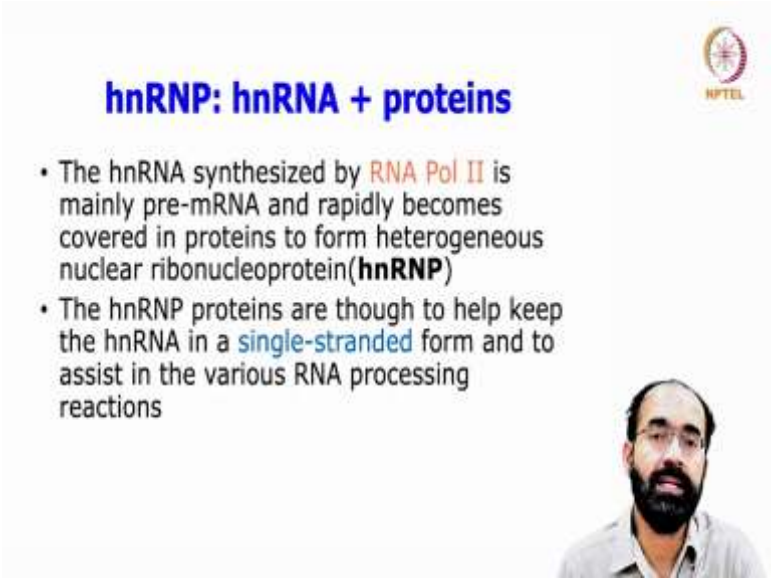
So, as soon as the transcription is completed the primary RNA transcript will have a 5 prime cap, 7 methyl guanosine cap and it also will have a 3 prime. This will be further cleaved by because downstream of the poly A site it will have a cleavage site by the endonuclease. Usually it is around 16 to 18 base pair downstream of the poly A site

usually. Poly A site we have seen AAUAAA or AUUAAA. This we have seen in the previous classes and there will be a cut created by the endonuclease.

And now this RNA will become an attractive candidate for the poly A polymerase enzyme which make use of ATP to add because ATP is the monomer which will be used for adding a stretch of A's. And this can the A can length can have a length of 100 to 200 or 250 long and it varies from RNA to RNA.

And now a cap and polyadenylated RNA pre-mRNA can undergo process called RNA splicing because of which this yellow colored introns will be removed and the pink color exons will remain. So, this is a processed RNA in a simplistic form.

(Refer Slide Time: 09:34)



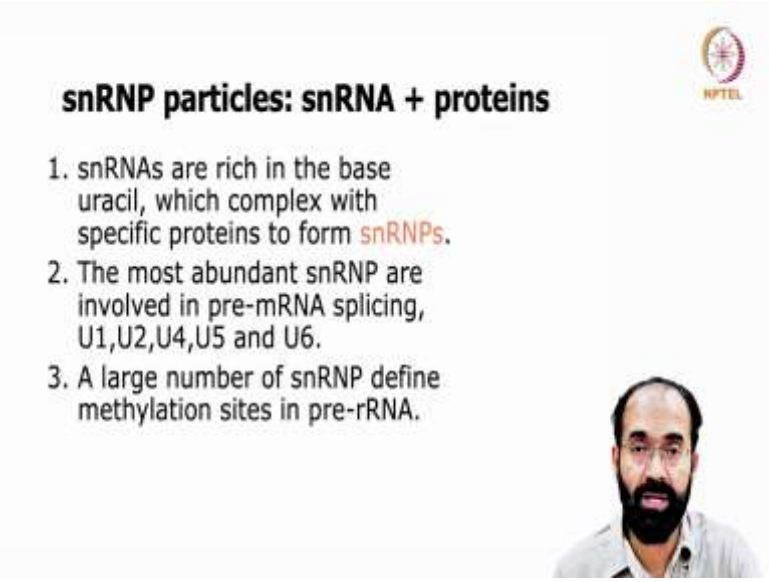
Now, let us quickly see what is hnRNP. hnRNA we have already seen the heterogeneous nuclear RNA is the pre-mRNA which has got all the introns and other you know it is quite crude and this bound with protein we call it as hnRNP. The hnRNA synthesized by RNA polymerase II is mainly pre-mRNA and rapidly becomes covered by in proteins to form hnRNP.

And the hnRNP proteins are although to help to keep the hnRNA in a single stranded form. Remember hnRNP or the proteins associated with hnRNA have a role to prevent a unwanted secondary structure being formed because sometimes some secondary structure make the RNA quite stable and resistant to any enzymatic action. This hnRNP

or the association of protein onto this hnRNA and that will help to assist in the various RNA processing reaction.

So, these proteins associated with hnRNA they are not junk they assist in the downstream processing of hnRNA.

(Refer Slide Time: 10:57)

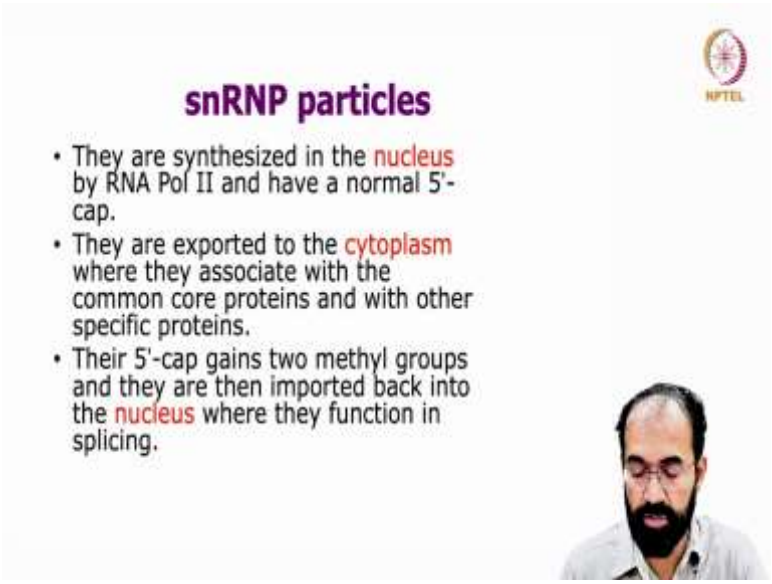


Like we know many processing are there capping is there, splicing is there, polyadenylation is there etcetera. And this snRNA also another example of this RNPs that is these are spliceosomal RNA we call it as snRNA and they are they are not mRNA their role is to help in the splicing.

So, they are encoded as separate genes elsewhere in the genome. So, snase snRNAs are rich in the base uracil, which complex with specific proteins and to form snRNPs and the most abundant snRNPs are involved in the pre-mRNA splicing. What are they U1, U2, U4, U5, U6 snRNA because since they are abundant in uracil, they are being given a name U1, U2, U4, U5 and U6. So, these are the major snRNAs that participate in the pre mRNA splicing.

A large number of snRNPs define the methylation sites in the pre ribosomal RNA because RNAs can undergo modification at various bases and they can also contribute to the stability as well as enzymatic identity to the RNA molecule. So, a variety of snRNPs are helpful in providing or identifying the methylation site on pre ribosomal RNA.

(Refer Slide Time: 12:38)



And let us look more into the snRNP particle. They are synthesized in the nucleus by RNA polymerase II and they usually have as soon as they are produced a few bases say 10 base, 15 base they will have a 5 prime cap that is made of 7 methyl guanosine. They are exported to the cytoplasm where they associate with the common core proteins and with other specific proteins.

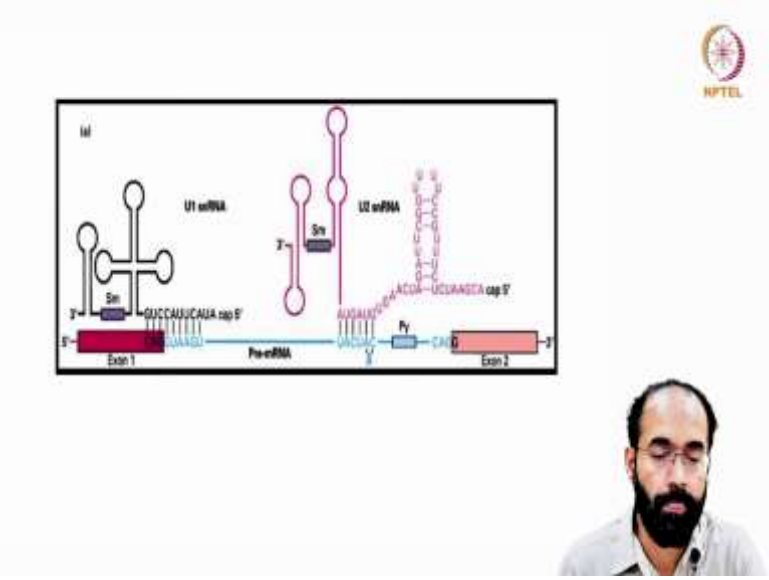
This is not for mRNA it is for some snRNP because they have to get exported to the cytoplasm and where they mature and again come back to the nucleus because their site of function is in the nucleus and their 5 prime cap gains 2 more methyl group. Initially they have got one methyl group and they are then imported back to the nucleus where they function in this splicing.

So, as soon as a snRNAs being produced it will have some protein association, it will have a monomethylation in the 5 prime cap and then it moves to the cytoplasm where it will acquire more proteins, more proteins that have to be present in this snRNP, but they are not available in the nucleus.

Because these proteins also have to be produced in the cytoplasm and some of these proteins have to undergo some modifications and then only it can bind on to this sn-RNP monomethylated at the 5 prime cap that is coming from the nucleus it will interact and then it will go back and going back at the time of going back it will have instead of one

methyl group it will have two methyl group. So, these are all some of the changes that take place during the maturation process of snRNP.

(Refer Slide Time: 14:25)



So, you can see here how snRNAs function you know this pink color box is exon 1 and this orange or saffron color box is exon 2 and in between this blue line is the intron. And important point to understand is that exon 1 and intron 1 boundary towards the 3 prime end of the exon and the 5 prime end of the intron you have a common stretch of sequence that is CAGGUAAGU.

This sequence is bound by U1 snRNA and U1 snRNA what it will do U1 snRNA can happily pair onto this exon intron boundary and this is very important for the splicing event. Then at downstream you have few features. What are they? You have a intron exon boundary you have it and then just upstream of that you have a polypyrimidine track, this polypyrimidine track much upstream of this polypyrimidine track you have got a branch point A site.

So, what is important for this is you may remember in group 1 and group 2 self splicing intron you studied about that in the ribozyme part we remember that one of them it recruited a exogenous guanosine and in another group 2 it had a branch point; it had a branch point that associated with the splicing event ok.

So, what we should understand; what we should understand is the polypyrimidine track and the intron exon boundary are quite important, it cannot be too far it has to be within 100 base without which it will not be able to perform the task of splicing. And in this branch point what we are seeing?

It can interact with specific base specific base of it can interact with the specific bases of U2 snRNA and because of this it can effectively pair and this pairing brings in some uniqueness. What is that uniqueness? This will have a specific secondary structure. So, what you are seeing here?

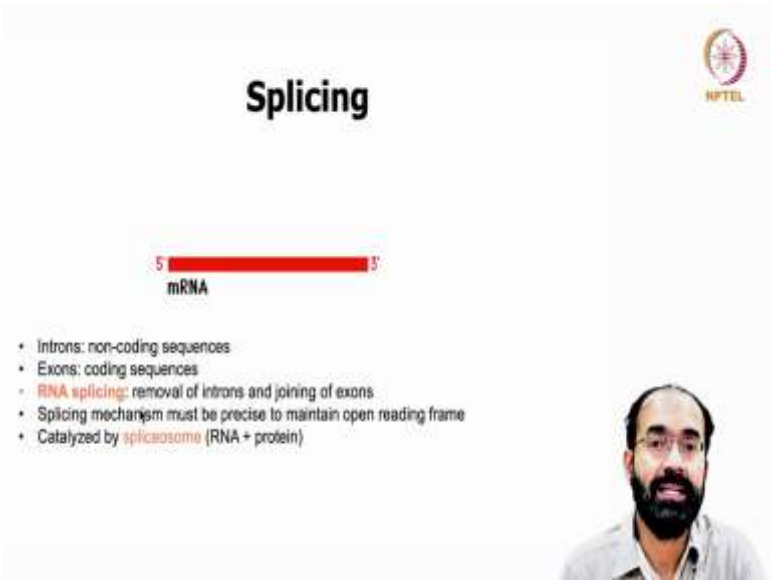
U1 snRNA bound onto the exon intron boundary exon 1 and intron 1 boundary whereas, U2 snRNA paired at the branch point we call it branch point why because this A and you also should not one more interesting thing that this pairing actually this sequence blue color sequence is UACUAAC.

But this U2 snRNA came and ignored this A, this A is pushed out of this pairing and because of this this A and of course, its an RNA it has got this 2 prime hydroxyl group reactive hydroxyl group is there which will be now available freely to interact for the splicing event to happen.

So, understand that this way of interaction is very important for the first transesterification reaction or that is important for the splicing event that is basically a nucleophilic attack. We will see that we have seen that in group 1 and group 2 introns, but we will see about it more in detail in the subsequent classes, but this slide is basically for you to understand that U1 snRNA and U2 snRNA physically pairs at the specific loci in the exon intron boundary and also at the branch point.

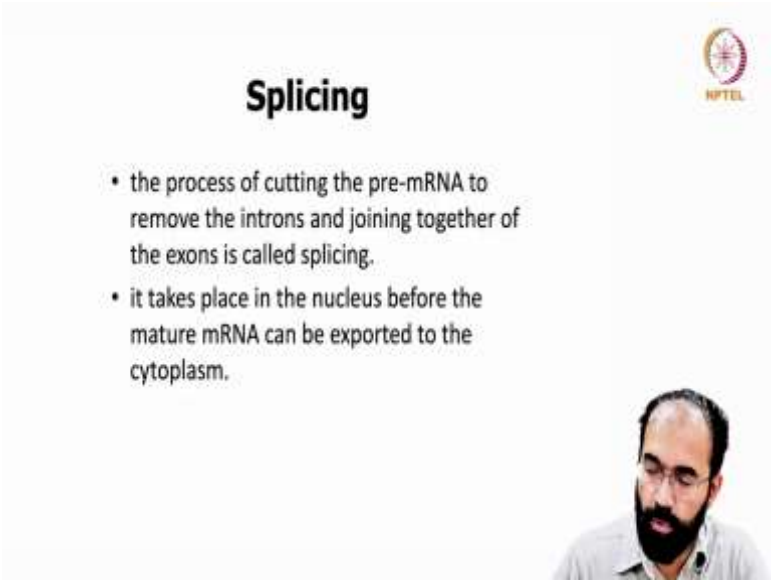
So, that the facilitation of splicing can happen such a thing was not necessary in the group 1 and group 2 introns because the secondary structure of the intron itself was enough to form the proper structures and also for facilitating the splicing, but here all introns do not have that proper secondary structure. So, hence the U1 and U2 snRNA become handy for those events.

(Refer Slide Time: 19:10)



Now, let us see how the splicing is happening. Introns are usually the non-coding sequence and the mRNA when they are produced, they have to undergo some maturation event.

(Refer Slide Time: 19:23)

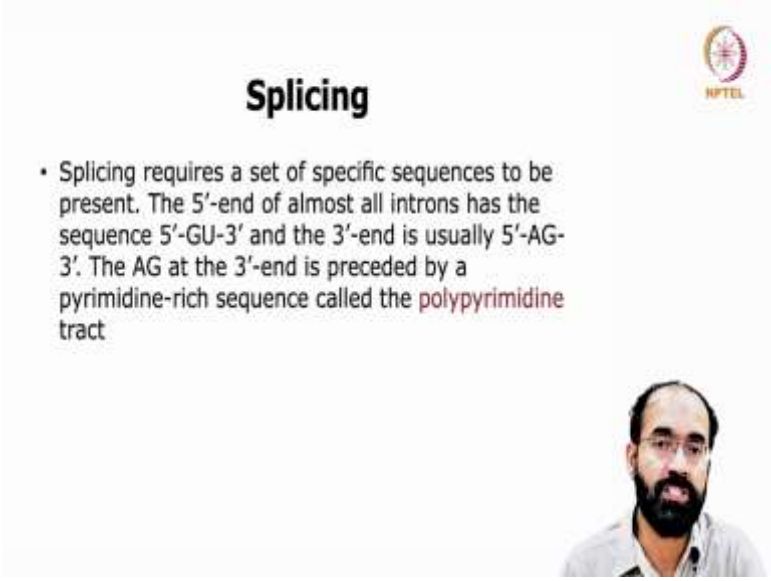


And the exons are the coding sequence they must be present they must be retained and the RNA splicing is very precise even one base error is not possible and the reason you know because the coding sequence are in triplet codon, if a triplet become quadruplet or one triplet became a duplex then you have a frame shift mutation.

So, the precision has to be at the level of single base. So, you cannot have one base extra, you cannot have one base less it has to be so precise. And the splicing action is facilitated by the spliceosomal machinery. Remember U1 and U2 are not the two spliceosomal machinery required there are other spliceosomal machinery also required we will see them in detail, but time being you will get some clue that U1 and U2 snRNA U1 binds on to the exon intron boundary and U2 binds at the branch site.

So, keep that information in the mind and when we study more about the splicing, we can have that in the mind quite handy. And coming back to splicing little bit more in detail that the process of cutting the pre mRNA to remove the introns and joining together of the exon is called simplistic term as the splicing. It takes place in the nucleus before the mature mRNA can be exported to the cytoplasm.

(Refer Slide Time: 21:03)



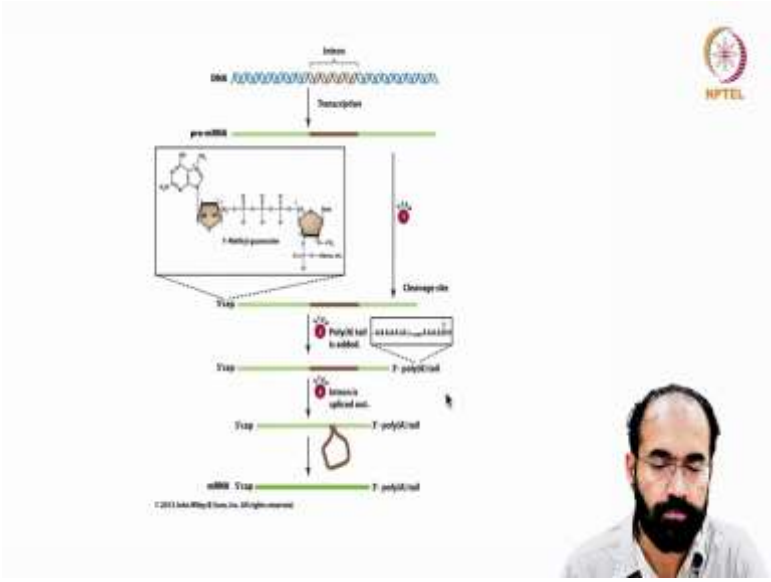
And the splicing require a set of specific sequences to be present splicing cannot happen randomly. Remember one important thing enzymes do not know or snRNA's do not know which is exon which is intron it is we who for convenience classify them they are blind they look only for specific sequence regions and it is the fusion post fusion event that makes an RNA useful for protein translation or not.

So, the splicing require a set of specific sequences to be present on the pre mRNA the 5 prime end of almost all introns has the sequence GU. So, the first base of an intron is G and the second base is always U and last, but one base of an intron has to be A and the

extreme last base of an intron is G this does not mean that every intron or every sequence starting from GU and ending with AG is an intron because we also need to have the polypyrimidine track and also, we need to have the branch point.

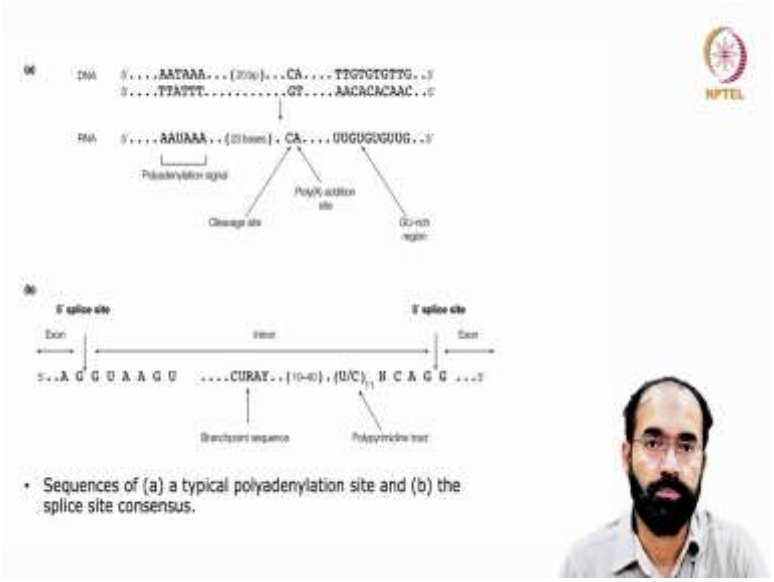
So, this has to be met, but no matter what the introns start with GU end with AG the AG at the 3 prime end is preceded by a pyrimidine rich sequence called polypyrimidine track and upstream of that you have got a branch point that must be an A.

(Refer Slide Time: 22:37)



So, coming back to the RNAs processing once again you have pre mRNA that is produced from the gene along with an intron and it undergoes the process called capping even before it is completed. And you have a cleavage site for the nucleus and then a poly A tail is added and this poly A tail added 5 prime capped RNA undergoes the splicing and this spliced RNA which has got the cap and also the poly A tail is now ready to be exported to the cytoplasm.

(Refer Slide Time: 23:15)



And this picture shows that what are the unique features that is required that has to be present in the DNA and also in the subsequent RNA. You can see here you have a sequence that is AAT triple A and then you have 20 base pair any random base pair then you have got a CA and followed by T T G T G T G T T G and of the sequence what is important to understand that the polyadenylation signal we are talking about the extreme end of a transcription unit and AAU AAA is a polyadenylation signal and you have any random 20 bases ok.

And this random 20 bases is followed by a CA sequence that is the cleavage site where the nucleus is acting and much downstream you have a GU-rich region. For every RNA before termination, it has got a GU-rich region, but that will not be part of the mature RNA because there will be a cleavage that is happening here in the CA site and once it has come out after cleavage this is ready for polyadenylation.

So, you can see here in the 5-prime splice site of an intron you have this GU that is starting with a intron GU and then you have got some AAGU sequence for the convenient pairing with the U1 snRNA then you have got a branch point sequence that is CURAY. So, these are all unique codes for the bases and C stands for cytosine, U stands for uridine, R stands for purine and A stands for A also its a purine, but G is not allowed it has to be an A.

And then it has to have a pyrimidine and then what happens you have a polypyrimidine track and followed by the polypyrimidine track you have got N and then CAG and this intron is ending with AG. So, these sequences are very important for the proper functioning of proper functioning of the splicing.

So, we will end today's class in here and in next class we will learn more in detail about the spliceosome because splicing is a complex process and we must understand step by step in a smooth manner. So, we will take one at a time and so that we will get a complete understanding of RNA splicing.

Thank you.