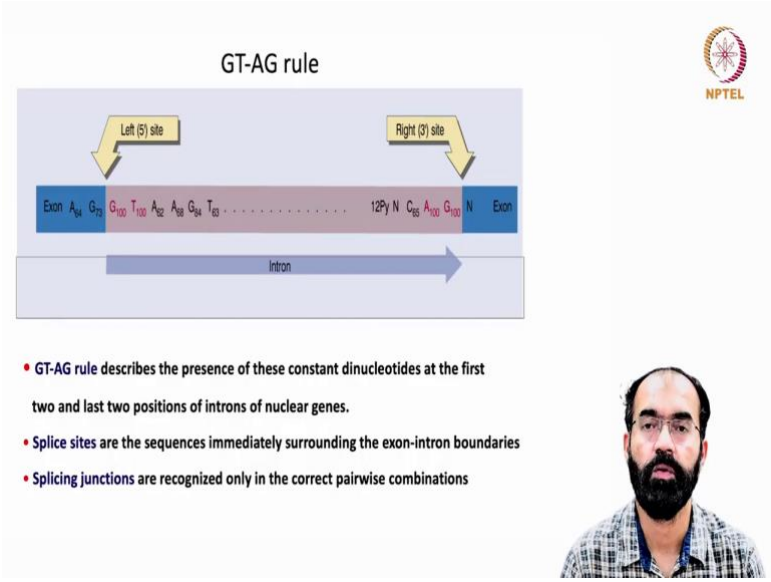
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 20
Alternative RNA Processing and Editing: Splicing and Pathology

(Refer Slide Time: 00:22)



Hello everyone, welcome back to another session of RNA Biology. In the previous class we were here that we are looking into the boundaries of exon and intron rather we were trying to find out is there any variations tolerated at certain basis because, we know that GU and AG are kind of unquestioned rule that intron starts with GT or GU in the case of RNA and ends with AG.

So, this base specificity is always followed in the case of introns. And of course, the polypyrimidine tract and the branch points etcetera also very important. But remember, as I told you in the previous class the branch point A is the one which participates in the first transesterification reaction on to the 3-prime end of the first exon or 3 prime end of the 5 prime exon.

So, 5 prime is a relative term which we are using 5 prime basically means towards the when you are looking into the screen towards the left that is. So, extreme end of the exon the first transesterification reaction initiated by the A. However, the free 3 prime OH can decide where it should participate in the second transesterification reaction, but it will

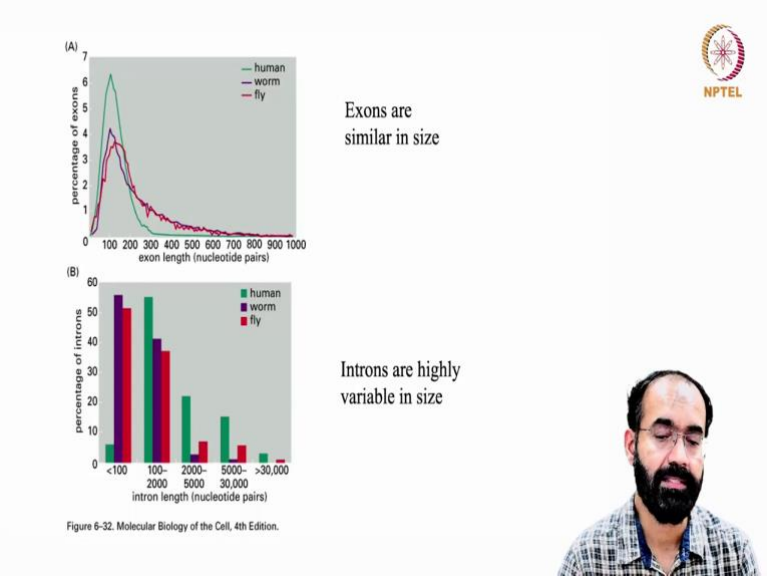
not go ever that you know around 10,000 basis or missed or something that does not happen, but few bases can miss in certain cell types which can kind of cause the alternative splicing.

But by and large as a rule it usually fuses soon after the AG part of the intron. What is important is same polypyrimidine tract same AG can come somewhere even inside the exon also. So, that time what happens some part of the exon is now included in the intron. Sometimes a intron an exon and the next introns like exon can be flanked by introns right. So, we saw in the previous class some examples that is we call it as exon skipping an entire intron exon and intron is considered as a huge single intron.

So, that time the available RNA or the mature RNA that comes out of such a splicing will completely miss out one exon part. So, this idea should be clear to you that while retaining the 100 percent situation 100 percent of the frequency for the beginning of intron with GT in the case of DNA gene what you are talking about otherwise it should be GU and then ending with AG is always followed.

So, GT-AG rule or GU AG rule is always followed when it comes to the boundary of an intron.

(Refer Slide Time: 03:35)



So, let us compare now what is the take of exons and intron when you compare them like exons and intron let us compare are the exons sizes are conserved or intron sizes are

conserved etcetera. So, looking little bit more detail into their conservative nature. So, in the y axis of this top graph A you can see the percentage of exon. And in the x axis you can see the length of the individual exon in nucleotide.

So, as you see in this graph itself no matter which animal you are talking about like you have human, you have worm basically means the C. elegans, Caenorhabditis elegans which is a model organism fly means drosophila which is fruit fly these are model organisms.

But when you compare these three animals, you can see no matter whether it is human no matter whether it is a worm no matter whether it is fly you can see the exon size more or less match when it comes to the size the length. So, 100, 200 or maybe a maximum 300 base length is a major representation of exons in all these individuals.

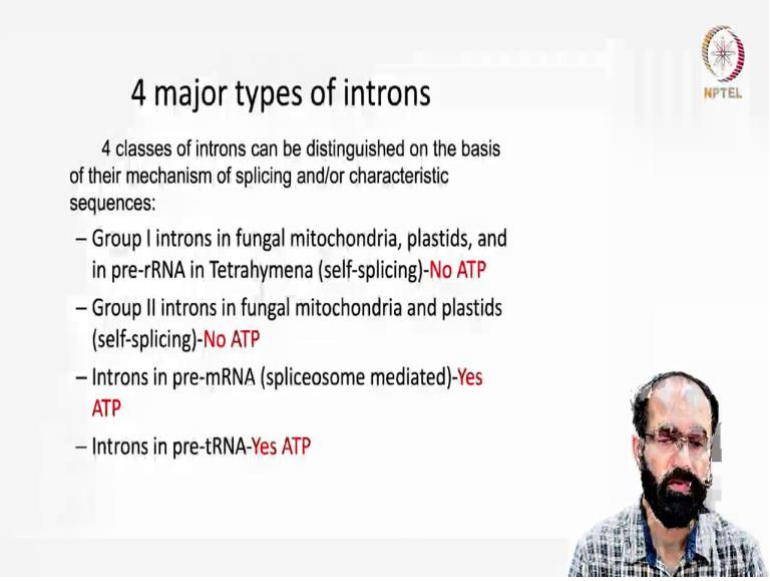
But some exons are there they can go as long or as big as 1000 base, but their frequency percentage representation is very low. So, what it indicates exons are reasonably smaller they do not have huge size whereas, introns when you are comparing same human, worm and fly when you are comparing remember exons were nucleotides length; whereas, introns also nucleotides, but they are now huge number because introns we know that by and large they are larger much much larger than the exons.

So, if you look the y axis is having percentage of introns and the x axis is having a bracketed number like less than 100 base pair 100 to 2000 and 2000 to 5000 and 5000 to 30,000 and more than 30,000 bases in length basically you are referring to intron size in Kb. Again, the intron size in human, worm and fly the intron size are much larger or much diverse compared to exon size, but we can always say a typical intron is less than 5000 on an average.

In human there are some introns which is ranging from 5000 to 30,000, but by and large the domination is on to the shorter side like less than 5000 or less than even 2000 is the predominance. And interestingly in humans you cannot find that much frequency of introns which is less than 100 bases with a reason because we know for a stretch of nucleated sequence to act as intron it need to have this branch point, polypyrimidine track all these features need to be there.

If it is too short it is hard to combine that such a small stretch of sequence less than 100 base pair having these features are a rarity. If it is larger introns are larger it is much much easy to accommodate these features.

(Refer Slide Time: 07:18)



And there are 4 major types of introns are there. 4 classes of introns can be distinguished on the basis of their mechanism of splicing and their characteristic sequences. Some we have already seen it in the previous class group I introns in fungal mitochondria, plastids and in pre ribosomal RNA of tetrahymena these are all self splicing. Group I and group II are self splicing introns and that do not require any energy source no ATP requirement.

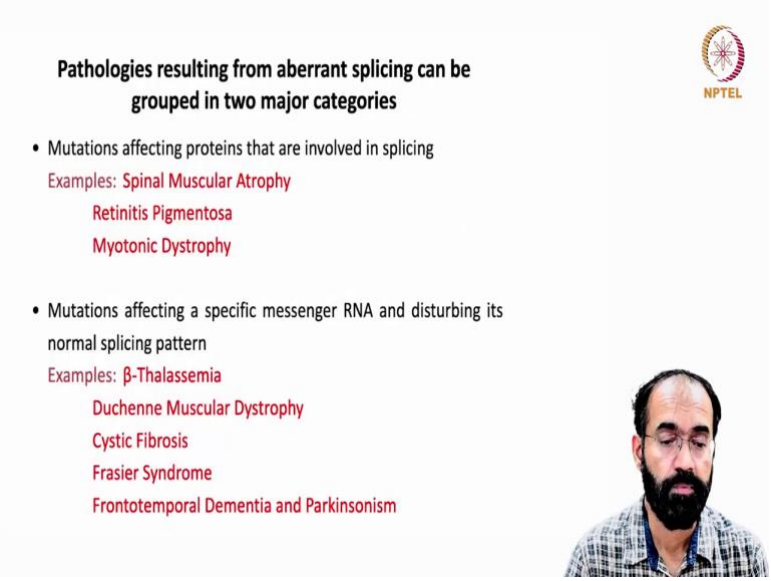
Here ATP what we are referring is whether consumption of energy takes place or not because they are ribozyme action, they are spontaneous. Whereas group II introns in the fungal mitochondria they are similar to group I except that in group I you have got a external guanosine recruitment whereas, in group II the endogenous A is acting as the branch point whereas, in group I a exogenously recruited G is acting as the branch point.

That is the only difference otherwise both are ribozyme and group II also do not require ATP. And introns in pre-mRNA that is spliceosome mediated they require ATP, they do spend make use of energy because they need plenty of spliceosomal machinery and this machinery can function if they have adequate amount of ATP available and introns of

pre-tRNA and they also contain the ATP they also have this ATP requirement, they also demand ATP.

So, all these 4 classes they can be easily identified or broadly classified whether they require ATP or they do not require and then we can split it whether they require the any assistance of spliceosome etcetera.

(Refer Slide Time: 09:23)



So, there are plenty of disorders or diseases come from the splicing defects and some of them can cause havoc. it is a dangerous situation. Pathologies resulting from aberrant splicing can be grouped into two major categories and we should understand some of the splicing defects are debilitating and even like threatening.

So, let us see some example. Mutations affecting proteins that are involved in splicing. So, there is a saying right if a person gone mad you can put them in chain what if chain goes mad, they say it's a trivial way of saying, but here we should understand mutations occur on those proteins that are helpful in splicing then plenty of the splicing events get affected because of the guardian or the guard of a building or guard of an ATM is turning out to be thief then lot of problems come right.

So, some examples of such disease are spinal muscular atrophy that is called SMA very serious problem affects children and some promising treatments are available which involves a gene therapy, but treatment should start before the age of two and its very

expensive as of now one trial medicine is costing around 18 crores of rupees, but some promising results those who are about to spend can get a life otherwise spinal muscular atrophy is a destined death there is no way you can recover from that naturally.

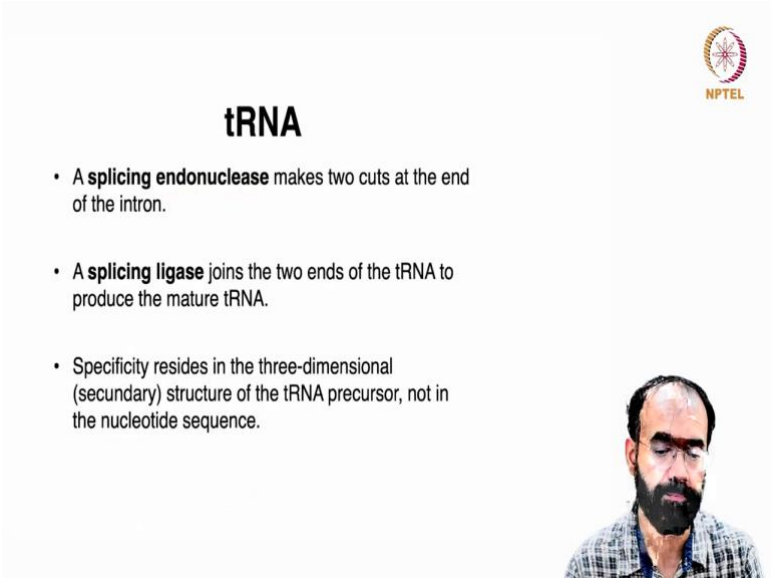
And then retinitis pigmentosa which affects very serious cancer condition a myotonic dystrophy. So, these are all some of the major or very prevalent disease that occurs in humans because of mutations affecting the splicing itself splicing event is affected. And then comes mutations affecting specific mRNA specific messenger RNA and disturbing its normal splicing pattern it is a candidate case like one it just like if one child or one man is or one student is not doing well that is that student's problem what if that sir the teaching person is not doing well that whole batch will go affected right.

So, this logic is very important. So, second example is where individual genes are affected one example is beta thalassemia and Duchenne muscular dystrophy and cystic fibrosis and Frasier syndrome, frontotemporal dementia and Parkinsonism these are all some of the diseases that is case by case.

A gene A is important or a protein A is important and the protein A is splicing is affected and only protein A's functions are affected, but if this protein is important in splicing of 100 mRNA, then all these 100 mRNAs are affected. So, the top one is very serious situation which can have a long lasting and widespread effect whether the lower one is not any less important, but it is a candidate gene problem.

So, understand splicing defects can cause havoc in the life of human beings. So, many people gets affected and of course, family history or genetics everything contributes to it. And; however, even randomly some genes if they are affected splicing defect is there, they can have more of a local effect rather than a widespread lineage specific effect.

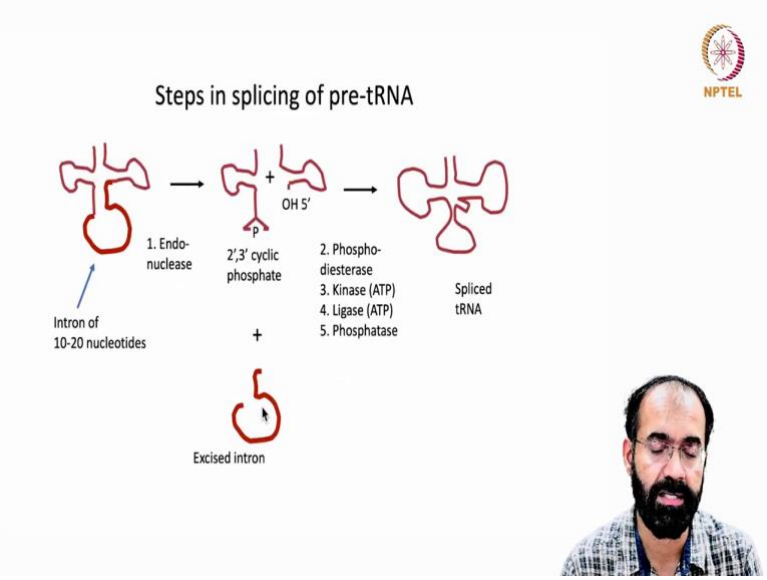
(Refer Slide Time: 13:17)



Let us look into the tRNA which is an important molecule which we know that it is an essential molecule for the protein translation. A splicing endonuclease makes two cuts at the end of the intron as we know tRNA also comes with pre-tRNA form that contains the intron. Now, a endonuclease cuts on both ends of the unwanted portion or the. So, called intronic portion then a splicing ligase joins the two ends of that tRNA to produce mature tRNA. This is how normally the maturation of the tRNA works.

The specificity resides in the three-dimensional secondary structure of the tRNA precursor this I have told you before also, RNAs identity is not on its sequence, RNAs identity is on its secondary structure. What shape the RNA assumes? Accordingly, you will end up having a recognition by an enzyme. Not based the identity of an RNA is not based on its nucleotide sequence.

(Refer Slide Time: 14:28)

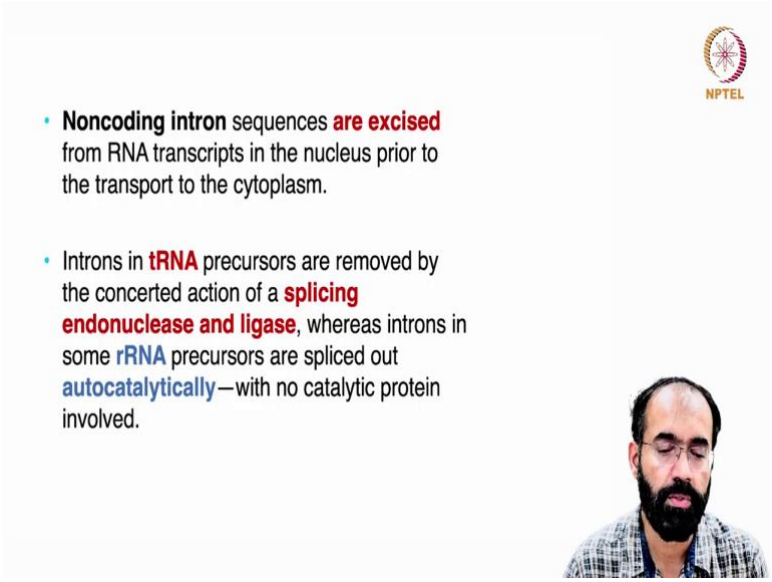


Let us see how the pre-tRNA maturation takes place. You have a pre-tRNA which is waiting for the maturation. The endonuclease cuts it, it cuts on two places and it makes two fragments and it gives a 2-prime, 3-prime cyclic phosphate stage and the excised intron is removed. Now, you have phosphodiesterase, kinase, ligase and phosphatase.

These enzymes comes together and fuse these fragments back into a functional or mature tRNA which is ready for act acted upon by aminoacyl tRNa synthetase various other modifying enzyme it has to have a CCA tail and also some pseudo-uridine etcetera, some base modification etcetera has to happen onto this tRNA.

But the splicing is very important. So, that this secondary structure now, earlier it had a particular secondary structure. Now, it has got a new secondary structure which will be recognized by a different set of enzyme.

(Refer Slide Time: 15:37)



So, non-coding intron sequences are excised from the RNA transcripts in the nucleus prior to its transport into the cytoplasm. So, introns present in the tRNA precursors are removed by a concerted action of both splicing endonuclease and the ligase. So, all these enzymes should act together as you saw in the previous slide you have to have a phosphodiesterase, kinase, kinase is an enzyme that can make use of the ATP and add a phosphate group to a target protein.

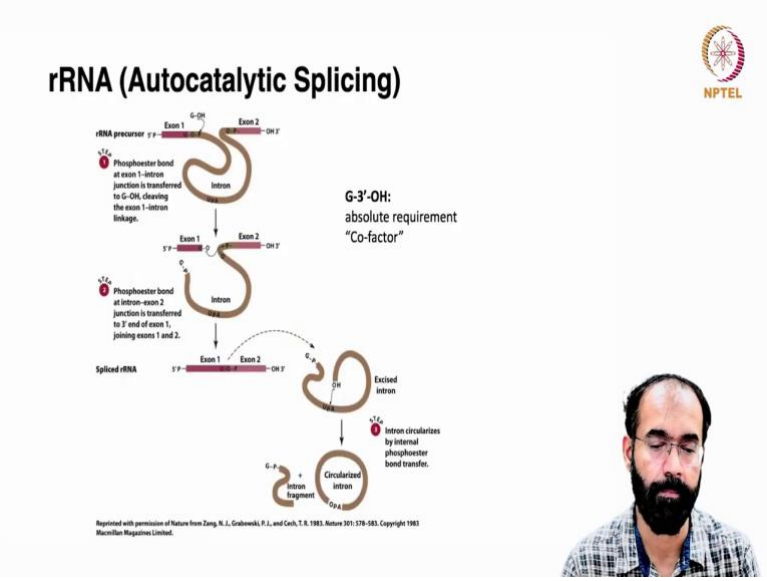
And then you have to have a ligase which also utilize the ATP and then you have a phosphatase. Sometimes an extra phosphate hanging around you need to remove it. So, that the enzyme will make the end of the nucleic acid blunt by getting removing the phosphate group. These are all requirement for the further action by the enzyme.

Sometimes removal of the phosphate group is important for preventing the concatemer formation. Nucleic acids will form join head to tail, head to tail it can join together if the phosphate group is there. So, phosphatase make sure that there is no such concatemerization happens.

So, introns in the tRNA precursors are move by the concerted action of many of these enzymes and the introns of some ribosomal RNA precursors are spliced out automatically and we call it as autocatalytic action, autocatalytically they are moved with no specific protein catalyst gets involved.

So, in some case it is not a rule, in some case it is capable of doing well similar to that what we saw in group I and group II introns. So, this is we are talking about certain ribosomal RNA.

(Refer Slide Time: 17:27)

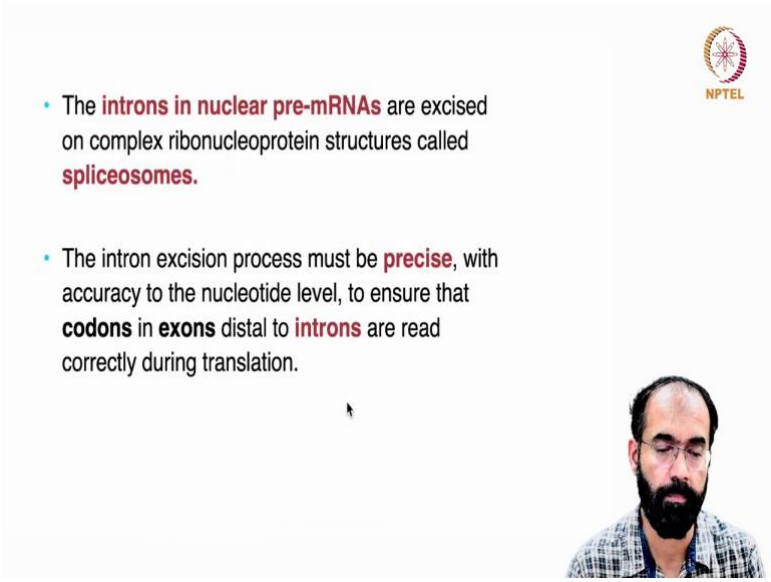


Let us see how does it do it. So, you have a ribosomal RNA precursor and it has got an exon and it has got a huge intron and then exon 2. So, the phosphodiester bond at the exon 1 and intron 1 junction that is the first exon intron boundary transferred to the GOH cleaving the exon 1 intron 1 linkage.

So, you can see here an extra G comes into picture and it transferred the shift and this G3 prime OH is an absolute requirement for this autocatalytic nature. And then as you saw in the case of group I intron the end of the exon 1 or the 3-prime end of the exon 1 participate in the second transesterification reaction and you end up getting a fused exon 1, exon 2 and the intron is released out.

So, intron later on circularizes by the formation of a phosphodiester bond and this will be cleaved off. So, this is the post processing of the released intron.

(Refer Slide Time: 18:36)



So, introns in the nuclear pre-mRNAs are excised on complex ribonucleoprotein structures and we call them as spliceosome. So, spliceosomes are very extensive and, in some animals, I mean some model organisms such as yeast there are bunch of other proteins which is not traditional spliceosomal proteins they also participate in this splicing reaction.

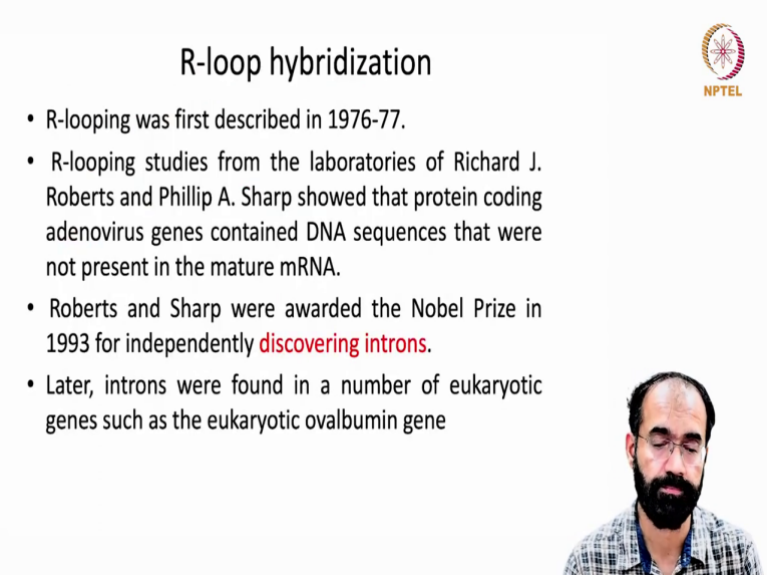
But remember the maturation of pre-mRNAs are very complex and stringently processed because if there is any error that happens in mRNA maturation that can have long lasting effect sometimes even deleterious effect; that means, it can cause the life of the organism because these mRNAs are important for the protein translation. So, the intron excision process must be precise with accuracy to single nucleotide level. You cannot have more accuracy than a single nucleotide.

You cannot have go to a half nucleotide level because the minimum or the maximum precision you can have is at single nucleotide level. To ensure that the codons in the exons distal to introns that is if you have an intron, it is upstream you have got exon down stream you have got exon. Many a times one codon may span between these two exon. Two base from exon 1 is now fused to single base from the exon 2.

So, if there is no such precision happens this codons fate is sealed and there can have frame shift mutation that can cause lot of problems. So, that is why the precision has to come up to single base and distal to introns are to be read correctly during translation. If

any problem happens with this fusion or joining then you will have very serious problem with this protein production. So, that is why mRNAs splicing is extremely important and stringently controlled.

(Refer Slide Time: 20:47)



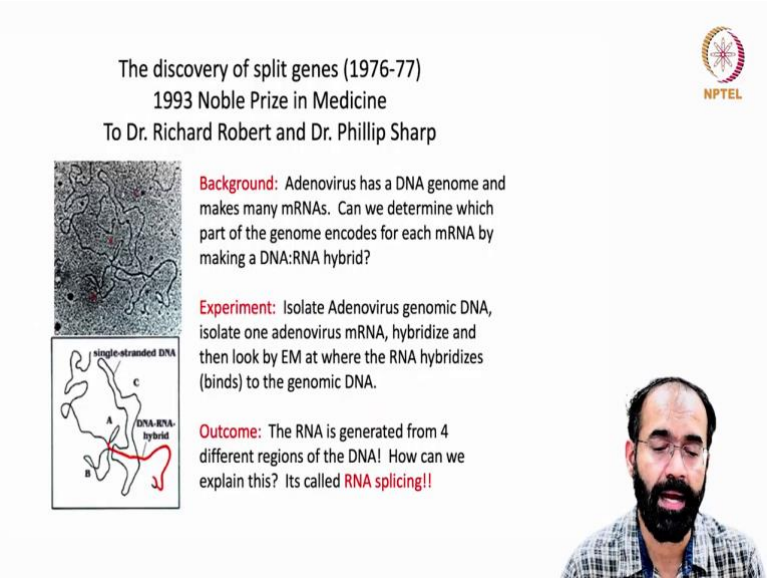
Now, let us think how do we know in the first-place introns are there. Now, we know that exons are there, introns are there we heard like a story, but who came and told us that you know genes have got intron bacteria do not have intron whereas, eukaryotes have introns. So, nobody came and told it is a observation done in the lab. Let us see how it is done. So, the technique used or the technique that time it was helpful in identifying the introns or the presence of introns is through R-loop hybridization.

So, name of the technique is R-loop hybridization. So, R-looping was first described in 1976, 77 in the year. So, R-looping studies from the laboratories of Richard Roberts and Philip Sharp these were the scientists who reported it showed that protein coding adenovirus genes contain DNA sequences that were not present in the mature mRNA. Very simple observation.

So, certain coding sequences present in the adenovirus genes which they were been working on they contain the DNA sequences that were not present where did they go mRNA things were missing. So, how did it disappear? This was an observation, but that is not an evidence it is a simple an observation. Roberts and Sharp was awarded Nobel Prize in 1993 for their discovery in 1976 and they are the discoveries of introns.

And we will see the mechanism of R-loop hybridization. Later introns were found in a number of eukaryotic genes such as eukaryotic ovalbumin gene etcetera etcetera. Plenty almost all of eukaryotic genes contain introns. Like I told you earlier also there are some genes without any introns. There are intron less genes also single exon genes also there, but their number is very low.

(Refer Slide Time: 22:44)



So, let us see how and these genes those discoverers like Richard Robert and Phillip sharp they named them as split genes because they know in the genome say 1 to 1000 basis there, but when it comes to RNA 1 to 1000 is not there say 1 to 100 is present and then 100 to 500 is missing 500 to 700 is present and 700 to 900 is missing 900 to 1000 is present.

So, they called it as split genes, but evidence was needed and you can see here their experiment with a proper designing using radioactive isotopes and electron microscopy they could prove it that presence of introns are nothing, but they are somehow removed post transcriptionally from the RNA. But evidence was needed.

What they did adenovirus has got a DNA genome and makes many mRNAs and can we determine which part of the genome encodes for each mRNA by making RNA-DNA hybrid or DNA-RNA hybrid. Like I told you DNA have a tendency to stay in double helical form and RNA also have a tendency to pair, but RNA and DNA also can pair.

We have seen during transcription the DNA literally pairs with the newly formed RNA and there also we used the term DNA-RNA hybrid. So, DNA-RNA hybrid can be used in experiments also in the laboratory condition. All we need to know is can I denature this double stranded DNA. Denaturing means separating the strand. You can denature them by heat and you can also denature them by chemical methods.

So, if you make double stranded DNA single stranded and then each of this strand is vulnerable for hybridization with another DNA strand or another RNA strand. So, this is the approach. So, DNA RNA hybrid can be created in the laboratory condition. So, what experiment they did?

Isolate adenovirus genomic DNA and isolate one adenovirus mRNA hybridize and then look by electron microscope electron microscopy and where the RNA hybridizes or binds. Hybridization is a term used where DNA and RNA are paired to the genomic DNA.

Now, what is the outcome? The RNA is generated from four different regions of the DNA that was amazing and surprising. So, actually they have the adenovirus DNA and they took one RNA and hybridized it. So, they would have expected this RNA will go and pair to a DNA strand because its denature, but they found this binding in A, B, C and D location only. Other parts it is not binding. The DNA is somehow standing out.

So, this is what they found. So, RNA generated from four different regions of the DNA and this is possible only by RNA splicing. So, you can see here in this picture this is the actual top panel is the actual electron micrograph and the lower panel is the cartoon of it. What you can see here? Single standard DNA is seen in black in color whereas, DNA RNA hybrid is seen in this red in color.

Many portions loop out and some portions are standing back. About this we will learn more in detail about this R-loop hybridization in the next class. Until then keep this information in your mind that the presence of introns were discovered through a technique called R-loop hybridization.

Thank you.