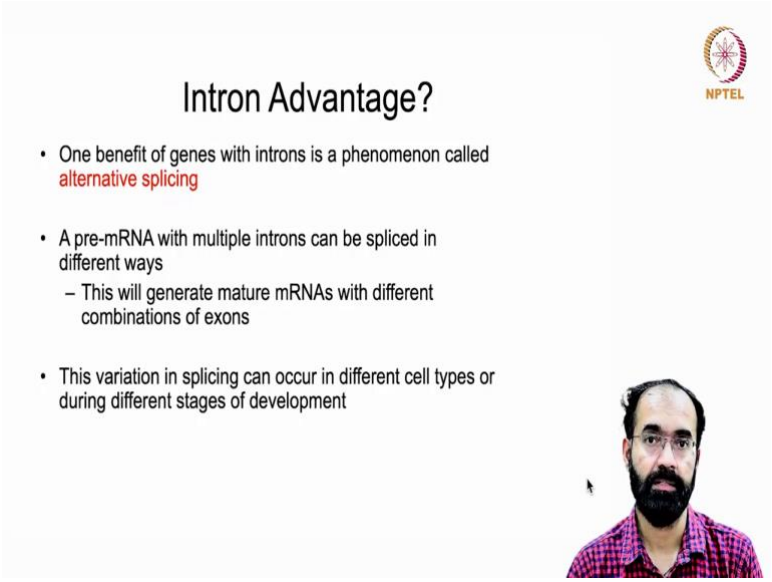
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

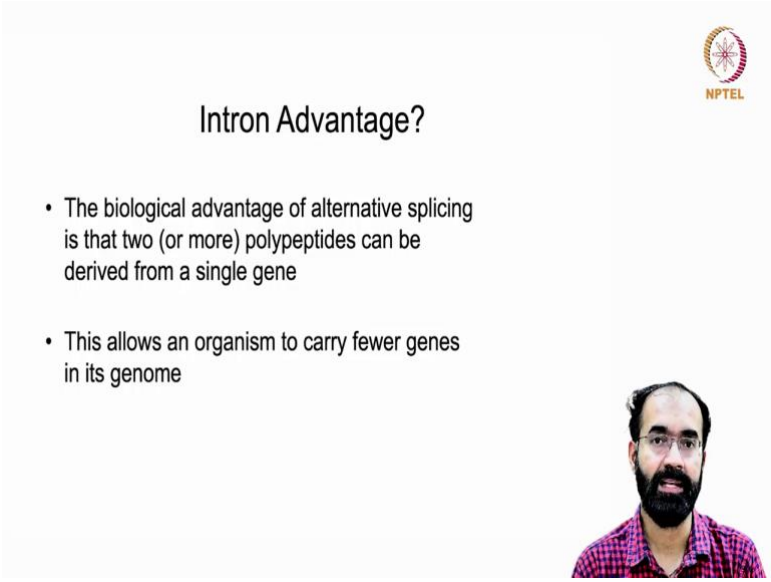
Lecture - 19
Alternative RNA Processing and Editing: Implications of Introns

(Refer Slide Time: 00:21)



Welcome back to another session of RNA Biology. And we were discussing about the advantages of having intron. And the answer is introns are indeed important in multiple ways. One is for expression of genes and also for creating diversity, plenty more. And we will see them, address them one by one. Because in nutshell, in a nutshell, we can say introns are extremely important and eukaryotes are eukaryotes credit goes to introns to a great extent compared to prokaryotes.

(Refer Slide Time: 01:04)



So, the biological advantage of alternative splicing is that two or more polypeptides can be derived from a single gene. And two or more is a very small or silly number when you are going to see a real time example, it will baffle you completely. But even two is definitely more than one, right? So, two is definitely handy.

So, alternative splicing minimum assures that there are two proteins coming from one gene. And hence with less number of genes you are able to do more. Coming back to the example which I told you with assumed 23000 genes you are able to perform the task of or you are able to make a complex organism that will require at least 1 lakh genes is, credit goes to alternative splicing and it is an enigmatic phenomenon.

Nobody completely understands the real essence of alternative splicing and I am sure in future someone may be able to come up with a clear-cut answer to alternative splicing and that will be one of the Nobel Prize winning discovery. Because alternative splicing is that important and we do not even now, completely understand how many expressed genes are there in humans.

We do not know; all we know is the possible coding sequences which can act as gene from the entire genome. So, total number of genes expressed in the lifetime because some genes are important only during developmental stage after that they are permanently shut off, they will never turn on back again.

So, it is very tricky and very difficult to identify that which are the genes, what are the possible alteration or alternative versions possible in from a gene in a simplistic form especially in human. This allows an organism to carry fewer genes in their genome. Because during evolution, evolution has happened in both direction.

In some direction the genome size started increasing. In another direction the genome size started decreasing. Genome means you may wonder what is this genome you are talking about. Human genome is around 3000 mega base means one base, once it becomes 1000 base it is one kilo base, 1000 kilo base is 1 mega base and we have 3000 mega base is our genome whereas, equally bacteria has got only 4 MB, four mega base.

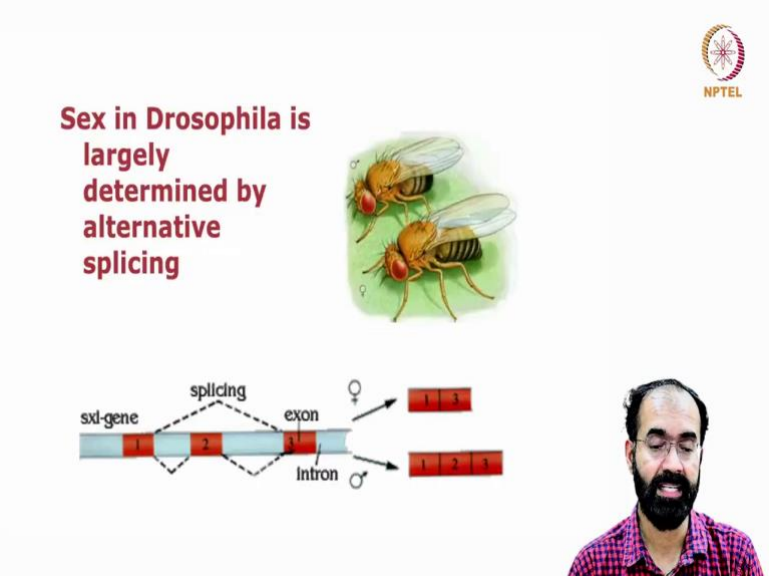
Drosophila and zebrafish all these animals have got differing size of their genome. Like some animals have got extremely large. Say for example, axolotl which is a model organism which has got 12 times bigger than human genome. That means 3000 into 12 that is the size of axolotl genome whereas, zebrafish have got only 1700 MB which is almost half of human genome.

But remember they all have got more or less the same number of genes. Do not think fish have got less number of genes than human because it is very small, not like that. Size or elephant have got larger number of genes or dinosaurs have much bigger number of genes than a worldly set. So, these are all completely different events which you should understand the size of an organism is not. Some worms have extremely large genome and they have large number of chromosomes also.

So, evolution has happened in both directions. So, we cannot simply say in one answer we cannot say more genome is more complex organism more genome means more genes more genes means, you can have complex or more ability to perform larger number of functions not like that. It always depends which organism you are talking about which gene you are talking about.

And also, you should not think every gene have got at least five alternatives splice to form. No, there are plenty of genes which have only one alternative splice form. And we should not think every human gene have got introns. There are many intron less genes also somewhat like bacteria just one exon that is the gene. So, this is also possible. So, there are less number of concrete rules in especially in genomic research.

(Refer Slide Time: 06:07)



Let us see one example how this alternative splicing is bringing in complexity. Sex in drosophila is largely determined by alternative splicing. Let us quickly see an example. This gene sxl-gene which has got exon 1, exon 2, exon 3. If the splicing happens between exon 1 and exon 3 then you end up getting a female. If splicing happens between exon 1, exon 2, and exon 3 you end up getting a male.

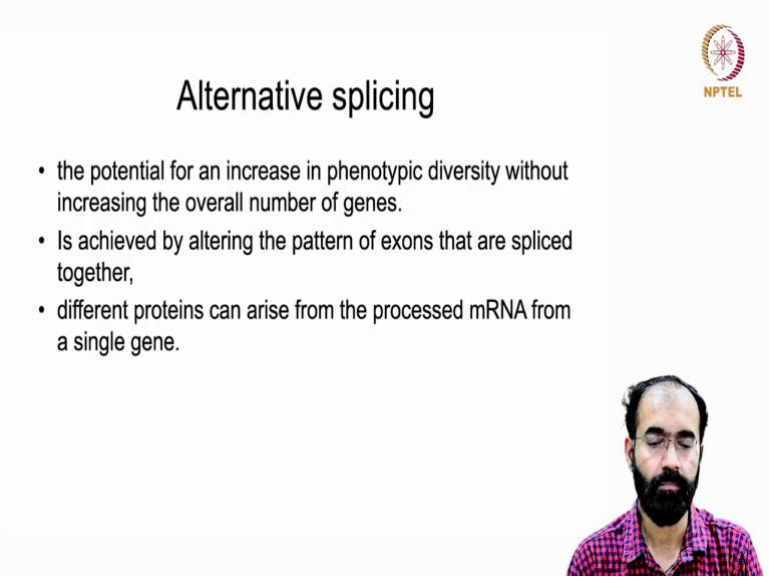
So, this is one easier example to understand because many organisms do not have a dedicated sex determination gene. Even if it do not have it will be able to perform a differential function during the development and subsequent animal physiology. It will be able to bring in, like human like human body like human body by default is female in structure. So, the role of the testosterone or the hormones is to resist or fighting 24X7 not to become female body.

So, if an organism human is born and that person do not have hormones testosterone or male hormones are not there or male hormone have got its corresponding receptor. If that receptor is mutated you cannot distinguish that person from a typical female. No way you can distinguish unless you do anatomical study or chromosomal preparation. So, what it indicates every organism have got a default body structure.

So, in the case of human the default body structure is feminine it is the pattern of a female. So, many organisms the hormones or the genes are resisting to become another

one. So, this is how the evolution has put forward individual organism we will see them when we study about the dosage compensation etcetera.

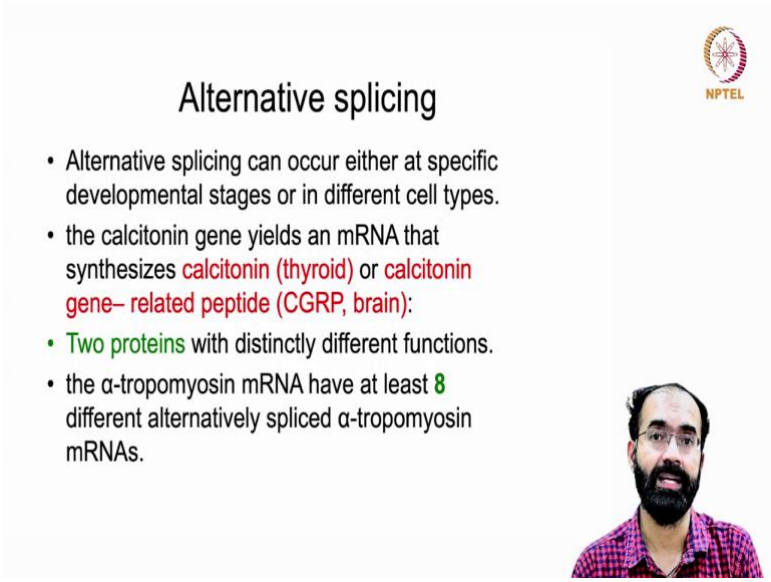
(Refer Slide Time: 08:07)



So, coming back to alternative splicing the potential for an increase in the phenotypic diversity without increasing the overall number of genes; this is the foremost advantage that with less you can do more because each protein is a functional molecule. It is achieved by altering the pattern of exons that are spliced together.

So, which are the exons that should be spliced together or joined together depends on which tissue or which organism you are referring to and which gene you are referring to. Different proteins can arise from the processed mRNA from a single gene because each RNA have got different coding sequence and each of them can perform different task.

(Refer Slide Time: 08:58)



So, alternative splicing can occur either at specific developmental stages or in different cell type. Sometimes a given protein is needed only for maybe for two days during the development. Then dedicating a full-fledged gene for that can be you know counterproductive. So, instead make an alternative spliced form of that particular protein make use of it. And rest of the time do not go for that alternative spliced form because that gene can perform a different task throughout the organism's life.

So, this is how you make use of alternative splicing during the development. Of course, we cannot make use of at our will it is how nature has evolved or an organism have evolved in during the evolutionary time scale. Let us see an example the calcitonin gene yields an mRNA that synthesizes calcitonin which is in the thyroid gland.

Thyroid gland is somewhere here near your neck. And we know thyroxine is an important hormone which is important for you know if you do not have thyroxine during your childhood you end up, I think that disease is called cretinism, you will be dwarfed and you will not have proper intellectual development.

But if you have dwarfism because of growth hormone deficiency then you will be short, but you will be normal, you are intellectually not challenged. But thyroid thyroxine is so important. And you may have heard many people will be taking thyroid tablet that T3, TSH you would have said that metabolism will be very low. We call them as

hypothyroid condition. Means their body has proper thyroxine production, but not adequate.

So, you have to supplement with tablet it is very important. Many times, for miscarriage during pregnancy etcetera will be a major problem if the thyroid hormone levels are under the normal level or below the normal level. So, it is very important throughout the lifespan for maintaining the body's metabolism.

So, calcitonin is a gene that is important in regulating the calcium balance in the body. So, it is normally expressed in the thyroid gland. Whereas, the calcitonin Gene related peptide it is called CGRP which is expressed in the brain both are coming from the one gene.

And two proteins with distinctly different function in different tissue so, one is in the brain another is in the thyroid so, coming from the same gene. And the alpha tropomyosin mRNA have at least 8 different alternatively spliced alpha tropomyosin mRNAs. This is something interesting and very important for us to keep in mind.

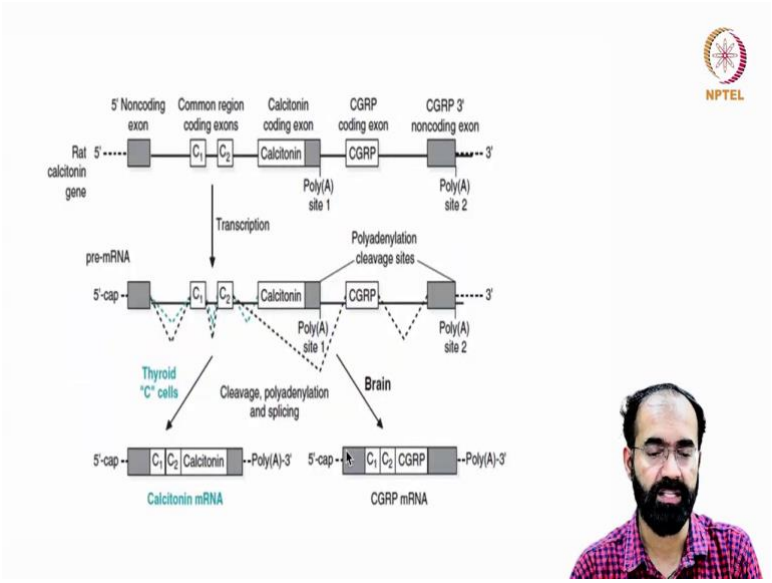
Because several structural protein because tropomyosin troponin these are all structural proteins in the muscle like you have muscle contraction, you are able to move your hand, finger, your facial muscles are moving, I am able to give a lecture all credit goes to muscle contraction happening at different parts of my body.

So, you need to have structural proteins to perform the task. So, heart have a different type of muscle, I my skeletal muscle have got a different type of muscle, my blood vessels have a different type of muscle. So, we will not go into those details. But understand structural proteins often need more diversity than an enzymatic protein or a hormonal protein because they have to perform differentially in different tissue.

So, one example is alpha tropomyosin which is a structural protein, structural protein means which provides structural feature or that provides function because of its structure. If an enzyme say salivary amylase it does not provide any structural advantage, it is provides a functional advantage. When you eat chapati and if chew it salivary amylase act down to the starch and convert it into glucose, it is a function where a structural protein it has to be there in order to perform a task.

So, structure related function so, we usually call them as a structural protein. So, alpha tropomyosin it is a structural protein, but it can give rise to mRNA can give rise to 8 different alternatively spliced alpha tropomyosin mRNAs and subsequently that much different proteins.

(Refer Slide Time: 13:36)



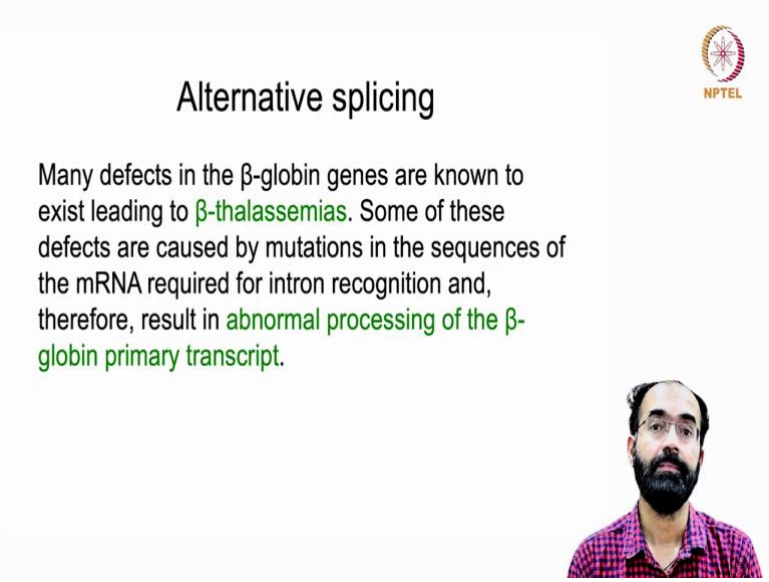
Let us see how this is done. We will see the calcitonin example also. So, this is an example of rat calcitonin it has got a 5 prime noncoding exon and it has got common region coding exon C_1 and C_2 and then you have another big exon and followed by a poly a tail and we call it calcitonin coding exon. And then you have downstream.

Remember you have a poly a site and then downstream you have a unique exon that is CGRP, CGRP coding exon and then you have a CGRP noncoding exon you have a poly a site. So, what do we understand? You have a common region 1st exon, 2nd, 3rd, 4th like that, but you have two different poly A site.

So, during transcription what it happens? The pre mRNA you have this 5 prime noncoding exon capping happen and you have proper splicing $C_1$to C_2 and calcitonin specific and the poly a site is made use of and it goes smooth and you end up getting calcitonin mRNA in the thyroid cells. You have 5 prime cap C_1 C_2 Calcitonin and poly A tail it goes smooth.

Whereas, in brain what happens? It completely skips this calcitonin, it includes as a big intron. So, it rather makes use of the CGRP. Remember it has got this noncoding exon C_1 and C_2 common noncoding exon C_1 C_2 common, but this calcitonin region is now replaced by CGRP because calcitonin region is skipped. And this poly A site was not used in the brain because it is masked and the second poly a site is made use of and you end up getting CGRP RNA.

(Refer Slide Time: 15:30)

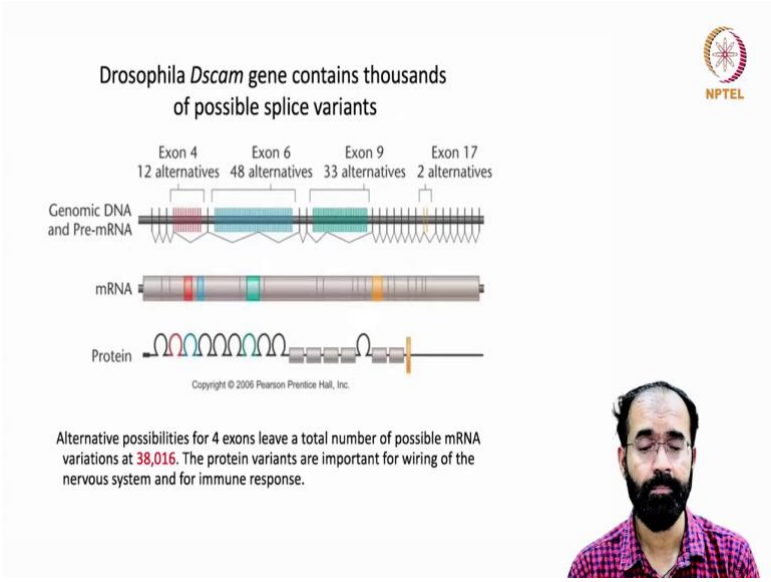


Then comes alternative splicing and its other features. Many defects in the beta globin genes are known to exist leading to beta-thalassemias. You may have heard about this this is condition because pregnant women are often screened for the embryonic baby whether it is vulnerable for beta-thalassemia.

So, parents usually their blood sample will be tested. So, to check it whether they are vulnerable for beta-thalassemia by gene sequencing etcetera. Some of these defects can be caused by mutations in the sequences of mRNA required for intron recognition. And therefore, it can result in abnormal processing of beta globin primary transcript.

When beta globin is abnormally processed you will have a defective less functional or non-functional hemoglobin. So, beta-thalassemia arise from this defective hemoglobin structure. Like I told gave an example of sickle cell anemia and like that beta-thalassemia also a structural defect in the globin. So, globin is an important component of hemoglobin and also in the formation of the RBCs.

(Refer Slide Time: 16:53)



And let us see some real classic example of drosophila and the gene name is Dscam. What is Dscam? Drosophila cell addition molecule that is was what Dscam gene. It contains thousands of possible splice variants. So, genomic DNA and the pre mRNA if you are seeing you have got exon 4, exon 6, exon 9 and exon 17. They participate in a diverse variety of alternative splicing.

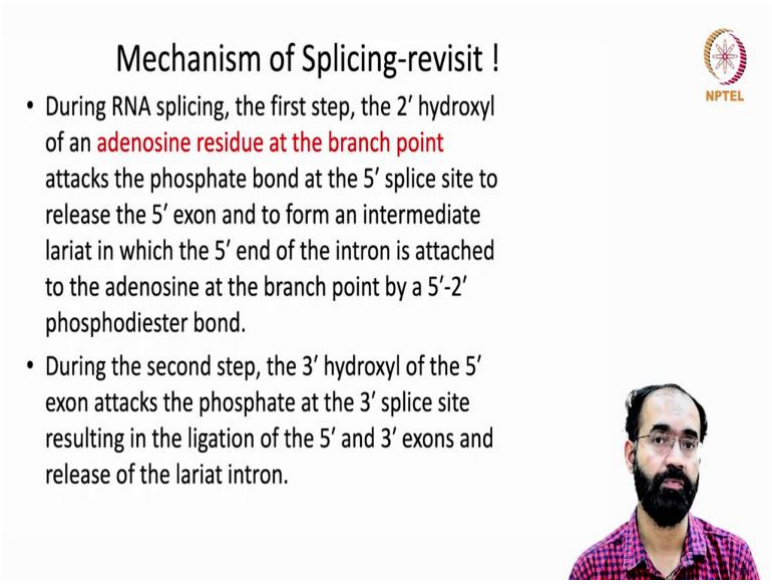
That means exon 4 can contribute 12 alternatives, exon 6 can cause 48 alternatives, exon 9 can bring in 33 alternatives and exon 17 can bring 2 alternatives. In the end of the day once the mRNA is formed it can bring up to 38,016 alternatives spliced product which is larger than human gene genome in itself. Human genome is supposed to have 23,000. So, 1 Dscam gene in drosophila is capable of encoding up to 38,016 variety.

So, remember it has got plenty of exons all are not participating. 4 exons leave a total number of possibilities. What are they? Exon 4, exon 6, exon 9, exon 17 with just 4 exons contributing to alternative splicing itself you can bring in around 38,000 plus varieties. The protein variants are important for the wiring the nervous system and immune response.

So, during the development of a complex organism like human you can imagine how much diverse, how much variety of proteins will be needed if a nervous system of drosophila require the cell addition molecule Dscam of up to around 38,000 different varieties.

So, it is just a question of being detected. It is not a question of if. No question exists if there is an alternative splice method of introducing complexity in human. It is just a question of how many and when these are alternatively spliced.

(Refer Slide Time: 19:18)



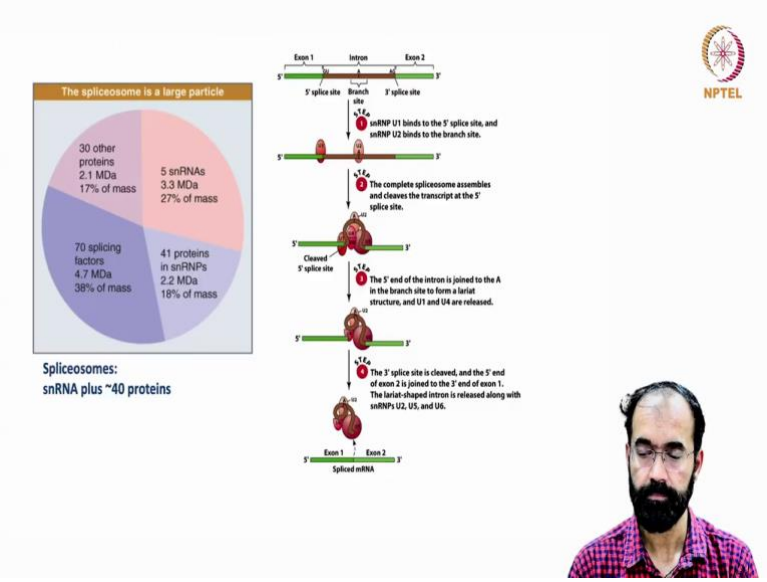
So, let us see the splicing once again for our easier understanding. The during the mRNA splicing the first step is the 2-prime hydroxyl of the adenosine residue at the branch point attacks the phosphate bond at the 5 prime splice site to release the 5 prime exon and to form an intermediate lariat of the intron.

Intron is still remaining attached to the second exon which at the 5-prime end of the intron is attached to the adenosine at the branch point that is in the intron. And this pairing is the 5 prime 2 prime phosphodiester bond that is the lariat. And during the second step the 3-prime hydroxyl of the 5-prime exon. Because 5 prime exon is now completely free, but it is not releasing credit goes to the snRNA or the spliceosomal machinery.

Exon attacks the phosphate at the 3 prime splice site resulting in the ligation of the 5 prime and 3 prime exon. Here comes the catch. Now, the 3 prime OH of the 5 prime exon where, is it going to start its hybridization or its fusion or its ligation or its splicing? Where is it going to start? And wherever it is starting it will include the portion as the intron.

So, first step is the lariat is formed because of the adenosine at the branch point and now the first exon or the 5 prime exon decides where it should fuse. So, cell can have different ways of regulating the second transesterification reaction or the second ligation reaction. Accordingly, you can decide an exon should be longer, shorter or where the second fusion should happen. That is why we are revisiting the splicing once again although we have already addressed the splicing.

(Refer Slide Time: 21:24)



So, now if you see closely this spliceosome is a large particle which we know, it is not a simpler thing and it is a highly highly what you call it is a highly organized structure and it is regulated in a stringent manner and it may sound like it is everything is perfect. And to some extent it is perfect everything is perfect. But splicing defects can also happen sometimes with a huge price and we will see some example.

Let us see more in detail that why we call spliceosome as complex. Some examples will make you understand better. Around 2.1 mega-Dalton mass is there and around 30 different proteins are present which is constituting 17 percent of the mass. And this is done by 5 snRNA which is around 3.3 mega-Dalton and 27 percent of the mass. Then again you have 41 proteins and which is constituting to the snRNP which will add to around 2.2 mega-Dalton and 18 percent of the mass.

And 70 splicing factors; splicing factors which is around 4.7 mega-Dalton and 38 percent of the mass. So, we should understand spliceosomes are forming only a minority

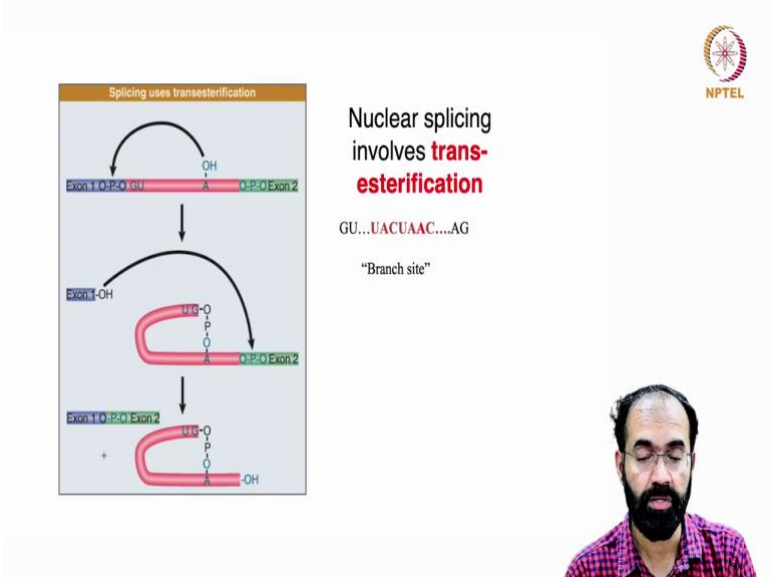
compared to a large number of proteins that can bring in variations in the during the splicing process.

So, we are seeing here quickly the splicing that is the exon 1, exon 2 and the intron 1 and you know U1 and U2 binds. U1 binds in exon 1 and intron 1 boundary, U2 binds in the branch point and then the complete spliceosome assembled once U1 and U2 are paid and cleaves the transcript at the 5 prime splice site.

Because the assembly of the spliceosome causes a bending. Because unless the intron is bent you will not be able to form the lariat. And the lariat is formed here the 5-prime end of the intron 1 is joined to the A in the branch point and to form the lariat structure and the U1 and U4 are now released. And now you have to have the second transesterification reaction.

The 3 prime splice site is cleaved at the 5-prime end of the exon 2 and is joined to the 3-prime end of the exon 1 and the lariat shaped intron is now released and it is free to go. And the exon 1 the 3 prime OH of the exon 1 has to be very crucial in deciding that where the fusion has to happen at what point onto the exon.

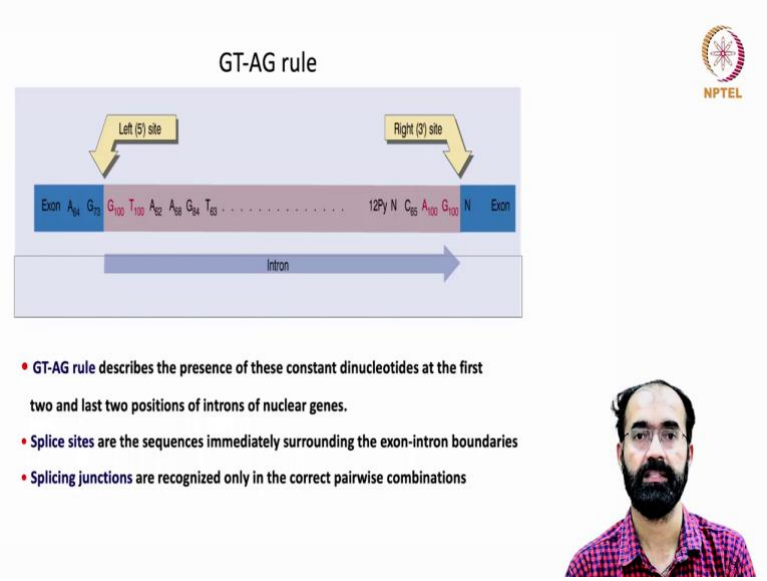
(Refer Slide Time: 24:27)



So, the nuclear splicing involves the so-called transesterification reaction. You are seeing this splicing in a less complex structure you have got this A; it has got this branch point OH and this OH is 2 prime OH, 2 prime OH participate in the first transesterification

reaction and then it goes into the second transesterification reaction here and you end up getting fusion of these 2 exons and the lariat is released. And the branch site structure is shown here, but the red color label to the A is the branch site A.

(Refer Slide Time: 25:06)



And the GT-AG rule is compulsory in RNA it is GU-AG rule. And let us see this picture shows how important they are this blue color one is exon and this little dull pink color portion is the intron and this exon is having ending A and G exon ok, and 64 percent is A and 73 percent is G. Now, let us see the G and T in the intron 100 percent is G 100 percent is T; that means, that rule is not flexible.

And same way you see the other end of the intron 100 percent is A in the towards the end of the intron and 100 percent is G that rule is not fixed. On the other hand, you see here this is an N first base of an exon is an N means it can be any nucleotide there is no compulsory; that means, this exon is free to fuse as and when it requires, wherever it required there is no rule that it has to it can have variation. So, this adds to the complexity.

So, GT-AG rule describes the presence of this constant dino nucleotide at the first and two and the last two positions GT and AG positions of the introns. And the splice sites are the sequences immediately surrounding the exon intron boundary, 5 prime splice site, 3 prime splice site. The splicing junctions are recognized only in the correct pair wise combinations. So, we will revisit more about the splicing in the subsequent classes.

Thank you.