Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Module - 06
Transcription
Lecture-27 Post Transcriptional Modifications

Hello everyone, this is Doctor Vishal Trivedi from department of biosciences and bioengineering IIT Guwahati and what we were discussing we were discussing about the different properties of the molecules within the course molecular biology. So far what we have discussed we have discussed about the basic properties of the cells within that we have discussed about the cellular structures. So we have discussed about the prokaryotic and as well as the eukaryotic structures and then we have also discussed about the organelles structures. So we have discussed about the structure of the different organelles. And in the previous module we have also discussed about the different types of biomolecules and what are the different structure and functional properties of these biomolecules. So we have discussed about the DNA, RNA, proteins, enzymes and how the activity of these molecules are being utilized by the cell so that they could not be able to perform the different types of functions.

And in the previous module we have discussed about the discovery of the genomes and how the genome is playing a crucial role in relaying the information from the one generation to the next generation and in addition to that we have also discussed about the central dogma of molecular biology. So when you talk about the central dogma of molecular biology it is a series of reactions which are being required by individual cell or even the organisms to produce the proteins and these events are being tightly controlled and regulated at multiple steps. So in this process we have the three different processes. In the process one you are actually going to have the synthesis of the new DNA from the pre-existing DNA through a process which is called as replications.

These reactions are being catalyzed by the enzyme which is called as DNA dependent DNA polymerase. Then subsequent to that the DNA is also been responsible for production of the or synthesis of the RNA and this process is called as transcription and this process is also been is catalyzed by the DNA dependent RNA polymerase and once the RNA is been formed it is actually going to be utilized by the protein synthesis machinery to produce the protein and then this process is also called as translation. And in the current module we are discussing about the transcriptions. So if you recall in the previous two lectures we have discussed about the transcription in the prokaryotes and in the transcription as well as the transcription in the eukaryotes. So when we were discussing about the transcription in prokaryotes we discuss about the different events

like the initiation, elongation and terminations.

Within the initiation we discuss about the how the RNA polymerase is been assembled onto the promoter side and what is the composition of the promoters and so on. So and subsequent to that we have also discussed about the eukaryotic promoter, eukaryotic transcription and within the eukaryotic transcription we discuss about how the pre initiation complex is been formed, how the RNA polymerase is different from the eukaryotic from the prokaryotic RNA polymerase and so on and what are the different events are happening within the eukaryotic transcriptions. So in today's lecture we are going to discuss about the another important topic that is related to the transcription and that is called as post transcriptional modifications. So transcription the generation of the RNA from the DNA is called as transcription and as a result of transcription you are going to have the three different types of RNA you ribosomal RNA you going to have the tRNA and you are going to have the messenger RNA. So formation of RNA from the DNA is known as the transcription and at the result of transcription you are going to have t3 different types of RNA species you going to have a ribosomal RNA you are going to have tRNA and you are going to have messenger RNA.

Now if you recall when we were discussing about the transcription in the prokaryotes the transcription in prokaryotes you are going to have the single RNA species a single RNA polymerase to do this job right. Whereas in the case of eukaryotic transcription you are going to have the three different types of RNA molecules you RNA polymerase molecules you are going to have the RNA pol 1, pol 2 and pol 3 and all these are actually going to have the separate set of the genes which are responsible for the production of ribosomal RNA, tRNA and messenger RNA. Now these messenger RNA and ribosomal RNA are actually going to be presented as a crude molecule ok. So once they are been synthesized they are actually going to be refined or I will say they are actually going to be modified in such a way that they will be more competent in terms of doing the job. So they are actually going to go through with the process of called as post-transcriptional modifications.

So once you have synthesized they are actually going to be present as a crude molecules and then you are going to have the post-transcriptional modification. The purpose of post-transcriptional modification could be very different. One of the major process is that it is actually going to be required for increasing the stability of the molecule. The second is it is actually going to provide the attachment site for many types of molecules. So it is going to be for attachment site and the third is it is actually going to be required for the making the molecules more versatile.

So that it is actually going to be interacting with more number of biomolecules and they

will be going to do many more functions. So first we will going to talk about the post-transcriptional modification within the messenger RNA and then we are going to take up the post-transcriptional modification in the ribosomal RNA and as well as the tRNA. Now messenger RNA so messenger RNA there will be three different types of post-transcriptional modifications what is going to happen in the messenger RNA you are going to have the addition of a cap to the 5 prime end then you are going to have a poly tail at the 3 prime end and then you are going to have the splicing of the introns from the gene. So this these two events are actually been required for increasing the stability of the molecules. So RNA polymerase is sorry so messenger RNA is been synthesized as a crude messenger RNA right and then it is actually going to be modified.

So initially you are going to put the 5 prime cap and you are going to have the coding sequence and then you are going to put the 3 prime poly tail and you are also going to remove the unwanted regions within the gene and that is how it is actually going to be a mature messenger RNA. So this is going to be mature messenger RNA where you are actually going to have so you are going to remove the introns right and you are also going to put the 5 prime cap and you are also going to put the 3 prime poly tail. Now let us first discuss about adding of the 5 prime end cap. So adding of cap to the 5 prime end so 5 prime and cap the as I said you know already that it is actually been required for providing the stability of the molecule. So capping in eukaryotic cell the messenger RNA is in integrably unstable at the end.

So it needs to be modified at the end to protect it against the ribonucleases. Messenger RNA is capped so that it is protected from the ribonucleases as well as it is important in the binding of messenger RNA to the ribosome for the translation. It uses the certain cap binding protein complexes. Capping reaction start soon after the transcription has started right. Remember that the as soon as the transcription start the messenger RNA 5 prime end is actually going to be out right and that is how the capping reaction will start so that the 5 prime end should be get protected from the ribonucleases and it is also going to serve as a docking site for the ribosome assembly and it is going to be assembled on that particular 5 prime end.

So that anyway we are going to discuss when we are going to discuss about the translation. So as soon as the 20 to 30 nucleotides are formed the capping occurs. At the 5 prime end the capping process occurs a slightly modified guanine the 7 methyl guanine is attached backward by a 5 prime to 5 prime linkage to the triphosphate of the first transcribed base. Capping reaction includes the condensation of the GTP with the triphosphate at the 5 prime end followed by the methylation of the guanine at 7 side. Further methylation occur at the 2 prime end hydrolysis of the second and third nucleotide adjacent to the cap.

So this is actually the 7-methanine-guanine-cane cap which is actually going to be placed onto the 5 prime end of the messenger RNA and how this is actually going to be synthesized that you are actually going to have the nucleotide. So you are going to have the guanine triphosphate or GTP and from guanine GTP there will be phosphohydrolase enzyme which is actually going to remove the PI and as a result of this it is actually going to remove the gamma phosphate. And then you are going to have this and then this is actually the messenger RNA and then you are going to have the addition of the GTP. So GTP also is going to have the 3 phosphate rings you are going to have the alpha, beta and gamma and then you are going to have the guanine transferase and guanine transferase is actually going to transfer the GTP onto this and as a result and there will be a removal of PPI. So these two are actually going to be removed and these are actually going to be attached onto this and then there will be a release of PPI.

This release of PPI is again going to be in form the 2 PI and that also is going to give you the energy. And then you are going to have this complex and then this complex is going to be get methylated by the guanine 7 methyl transferase enzyme and that is how the there will be addition of methane at the G site right on the guanine site. So there will be addition of methane, methionine methyl group at this site and then there will be another methylation at the 2 prime site and that reaction is going to be catalyzed by the 2 methyl transferase and that is how you are going to have the 7 methyl guanine scene which is actually going to be attached onto the messenger RNA. And this cap is very stable because this cap is not going to be recognized by the RNA polymerase because this cap is not neither because this cap had so many methylations and all those kind of modifications. So they are actually not going to be recognized by the RNA is what is present in the cytosol and then on the other hand this cap is actually going to be specifically be recognized by the ribosome and that is how it is actually going to be responsible for initiating the protein synthesis from this end.

Then the second modification is the adding addition of a poly A tail on the 3 prime end. So on this side you are going to have the poly A tail. So eukaryotic messenger RNA has a series of adenosine residue ranging from the 80 to 250 in the number forming a poly A tail at the 3 prime end of the primary transcripts. This poly A tail has several uses. It can be export mature messenger RNA out of the nucleus site.

It increases the stability of the messenger RNA and number 3 it serves as a recognition of the sequence for the binding of translational factor during the initiation of the translation. The process requires the template independent RNA polymerase called a poly A polymerase. So this is going to be the reactions site. So then the first step what you are going to have is you are going to have the binding of the CPSF. So that CPSF

will go and bind to the primary transcripts.

And then there will be a binding of the additional factors which is called as STSF and CSF and things are actually going to bind the polyadenylation site. And then there will be a binding of the poly A polymerase. Remember that the poly A polymerase is a template independent polymerase. Remember that the DNA polymerase requires a template. RNA polymerase also requires a template.

But in this case the poly A polymerase does not require a template and then the polymer A polymerase is actually going to add the series of A's on to the 3 prime end of the primary transcripts. The length of this poly A tail like for example it can range from the 80 to 250. So depending upon the number of A's it is actually going to say that what will be the age of the messenger RNA. Then we have the third step and the third step is that splicing or the intron removal of the intron. So remember that in a eukaryotic system what you have is you have the genomic region.

The genomic region is actually having the two regions. One is called as the axon and the other is called as the intron. And so you can have the multiple axons and introns and these introns are actually the non-coding region. So these introns are the non-coding region of a non-coding region what is present in the gene. So since these are non-coding region and they will be present within the coding region this non-coding region has to be removed and then only you can be able to attach the axon 1.

So this is the axon 1 and this is axon 2. So you can actually need to attach the axon 1 to the axon 2 by removing the intron through a process which is called as the splicing. So this is been achieved after the RNA is or messenger RNA is been synthesized. So introns are the non-coding nucleotide sequence within a gene that do not code for the protein and do not appear in the final messenger RNA molecule with that is removed by a process which is called as splicing. Protein coding sequence of a gene known as the axon which are interrupted by the intron.

The vast majority of the eukaryotic genes are interrupted by non-coding regions that is intron which needed to be spliced out. However histone protein coding gene is in the vertebrate in a one exception. So in a histone protein coding gene you are actually going to have the nose splicing. The occurrence of the intron varies in the eukaryotic species some each species slacks intron and many genes in the eukaryotes carry the a dozen of them. New bacterial and archal genes also have the introns.

So introns can vary in a length from 50 to 20,000 nucleotides. In higher animals such as human the introns are more than the axon which means in higher animals like mammals

you are going to have the more bigger region of the non-coding region rather than the coding regions. There are four classes of the introns. You can have the group one introns, you can have the group two introns both are the self-splicing intron and does not involve any of the protein enzymes. Then you have the spliceosomal introns they are the not self-splicing introns and then you also have the introns that require the ATP for the splicing.

So depending upon the structures and the other kinds of features the introns could be of four different types. Group one, group two, spliceosomal introns and the intron that do not require the ATP for the splicing. So splicing is a very very important feature, but it is mostly been associated with the eukaryotic genes rather than cochariotic genes. So as a thumb rule most of the eukaryotic genes are actually having the intron although there are exception that in some of the yeast gene there is no intron present and also the prokaryotic genes do not contain the introns, but there are exception that some of the bacteria and the archaebacterias are actually having the introns in their genes. So keeping the exception on a side introns are present in the eukaryotic system and introns are absent in the prokaryotic system.

So let us discuss first the splicing by the different mechanisms. So splicing mechanisms first is the splicing mechanism in the group one and group two introns. Splicing mechanism of both the group one and group two remember these are the self splicing introns involves the similar steps of the two trans-terrification reaction in which a ribose two prime or three prime hydroxyl group makes a nucleophilic attack onto the phosphorus and the new phosphodiester bond is formed at the expense of the old. So this is the mechanism in the group one and this is the reactions in the group two introns. So what happened is that this is the region what you are supposed to remove.

So this is the intron and intron is always been characterized by a specific pattern of the nucleotides what is present in that particular region. For example if you are having the G and all that. So what will happen is there will be a two event of the nucleophilic attacks by the two prime and three prime hydroxyl group and as a result so what will happen is the first nucleophile attack would be on this side. So this is going to be first nucleophile attack and as a result this bond between the pi prime this bond between the this is going to be broken and then this is actually going to have the nucleophilic attack on this right. This is going to have the nucleophilic attack on this and as a result this particular portion is actually going to be removed and this is actually going to form a bond.

So this is actually exon one and two and this is actually going to be the intron what is present. So in a group one intron the there will be a nucleophilic attack from the this OH right what is present on the interface on the boundary of the first exon and the intron and

then there will be a second nucleophilic attack from this OH on to this phosphate and as a result there will be a bond which is going to be formed and this the in exon intron is actually going to be removed. In the group two where you are actually going to have the lariat formation there will be exactly the same way that you are going to have the two rounds of the transesterification reactions. So and that actually is going to result into a addition of the one and two exon and there will be a removal of lariates.

Then we have the alternate splicing. So the alternate splicing mechanism is a method which is substantially used for many mammalian genes can result in the multiple product that vary structurally and functionally from the same primary transits. Some type alternate splicing is unregulated phenomena which is some in strictly regulated. One of the best example of regulated alternate splicing occurs in the sex determination in Drosophila. In Drosophila three genes are involved in sex determinations sex lethal cell that is the XL transformer gene that is the tra genes and double sex gene that is the ESX. So in the Drosophila what you have or in the alternate splicing what we have is you are actually going to have a gene for example, this gene has four exons followed by the intron.

So you are going to have the exon number one two three and four. Now what will happen is that you are going to have the primary transcript that is going to be formed as messenger RNA. And then you can actually have the multiple combinations you can have the one two three you can have the one three four. So this one is actually going can come along with two and three and that is how you are going to have this. This is the transcript one and if one comes with the three and four then it is going to form this and there are many more other combinations also like for example, one can come with one two and four this is going to be the third combination and you can also have another combination that is the like one followed by three followed by four or one followed by two followed by followed by so one followed by three one followed by four.

So all these combinations could be possible or two three four actually there will be another combination that is like two followed by three followed by four. So these are the some of the alternate splicing where one exon is making a combination with two and three or one exon is making a combination with three four or one exon or two exon is making a combination with three four. So these are the different combinations what could be possible and as a result of this only the mammalian genome has the potential to produce different types of proteins and different types of proteins even from the single gene. Due to the alternate splicing the functional genes are produced in females and non functional genes are produced in the male. So this is a just example that where the one example is in Drosophila where the sex determination is being done by the three genes that is called as sex lethal genes, transformer gene and double sex genes and as a result

of alternate splicing the functional genes are produced in the females whereas, non functional genes are produced in males.

Alternate splicing occurs using two mechanism one when we two one when two poly A or cleavage sites are available in the primary transcripts cleavage occurs at either site resulting in the two products. Such mechanism is followed by the variable domain of the immunoglobulin heavy chain and their diversity is due to the mechanism of alternate splicing. Similarly the alternate splicing with such mechanism results in the product of two different hormones calcium regulating hormone in the red thyroid and the calcitonin gene related peptide in the red brain. Other mechanism involves more than three prime site or five prime site hence splicing occurs by taking either of those three prime splice site resulting in the different products.

Then we have the another messenger RNA splicing. Such mechanism is followed by a variable domain of the immunoglobulin heavy chain and other variant is due to the mechanism of alternate splicing. So the alternate splicing is a very robust phenomena what is happening within the eukaryotic system and it actually allows the production of different types of variations within a biological system and as a result you are actually going to have the different types of protein what is going to be produced even from the single gene. Now let us move on to the next RNA that is the ribosomal RNA. So processing of ribosomal RNA eukaryotes have 87 ATS ribosomes whereas the prokaryotes have the 70S ribosomes. Ribosomal RNAs are transcribed as a long precursor sequence which is then modified at a specific basis and cleavage to give the mature product.

In both bacteria and eukaryotic ribosomal RNA processing involve two basic steps of cleavage and the base modifications. So ribosomal RNA processing in bacteria. So ribosomal RNA precursor in bacteria is a 30S RNA ribosomal RNA which is modified and cleaved to give 23S ribosomal RNA, 16S ribosomal RNA, 5S ribosomal RNA and some tRNA segment in between are also there sometime. 30S pre ribosomal transcript consists of the 16S ribosomal sequence followed by the spacer which may have tRNA sequence in some cases and there is a 23 ribosomal RNA sequence followed by the 5S ribosomal sequence near 3 prime end. At times there is one more tRNA sequence after the 5S ribosomal RNA sequence at the 3 prime end.

There are several different genes for ribosomal RNA in E. coli they are essentially similar in sequence of ribosomal segment, but differ with number and sequence of tRNA segments. Maturation process involves the methylation of the 3 prime 30S ribosomal precursor at a specific site occurring at 2 prime hydroxyl group of bases. Some bases such as uridine is modified to pseudo-uridine or dihydro-uridine. Further cleavage

process is carried out using the enzyme RNase 3, RNase P and RNase E at the site 1, 2 and 3 respectively.

Intermediate products are formed mainly as 70S ribosomal RNA, tRNA, 25S and 5S. These are acted on by certain nuclei to give the final product of 16S tRNA, 23S, 5S respectively. So this is actually going to be the primary transcript where you are going to have the 16S ribosomal RNA, tRNA gene, then you are going to have 23R ribosomal RNA and the 5S ribosomal. And after this you are going to have the mature RNA where you are going to have the 16S ribosomal RNA, you are going to have the tRNA, you are going to have 23S ribosomal RNA and you are going to have the 5S ribosomal RNA. And you know that all of these are actually going to come together to give you a 70S ribosomal RNA.

Then ribosomal RNA processing in the eukaryotes. So in the eukaryotes nucleolus is a site of processing the ribosomal RNA. A 45S ribosomal precursor is formed by the RNA pol I and processed in the 90S pre ribosomal nuclear complex to give the 18S, 28S and 5.8S ribosomal RNA. There is a tight coupling of RNA processing with the ribosomal assembly 5S ribosomal RNA is transcribed by the RNA pol III from a separate gene. Precursor RNA undergoes methylation at more than 100 bases from the 14,000 nucleotide at 2 prime and hardoxyl groups.

Furthermore, there is a modification of bases such as uridine to serodilidine etc. followed by a series of cleavage reactions. Cleavage and modifications are guided by the small nuclear roller RNA in yeast the entire processing involves a pre RNA, 170 non ribosomal protein, 70SNO RNA and 78 ribosomal protein. SNO RNA are supposed to be the remnant of the nucleosomes. So this is going to be the ribosomal RNA pre ribosomal RNA which actually contains the gene for 18S, 5.

8S and 28S and then this is going to be cleaved and from separately you are going to have the 5S ribosomal RNA. And these all going to be combined and it will actually going to give you a mature ribosomal RNA that is the 80S ribosomal RNA. Now let us talk about the processing of the tRNA. So we do you do not have to worry about the structure of the tRNA because that anyway we are going to discuss when we are going to discuss in the into the because the tRNA has a major role in the protein production. So that we are going to discuss when we are going to discuss about the translation.

So in both the eukaryotic and prokaryotic tRNA processing occurs it is transcribed as a long precursor sometime as a single primary transcript carry more than tRNA one tRNA segment which is separated by the cleavage. The processing of tRNA involved cutting off the extra sequences by the endonucleases such as RNase P at the 5 prime end and

RNase D at the 3 prime end. RNase P is a ribozyme which is a RNA exhibiting catalytic activity. After removal of the sequences from the 3 prime end the CCS sequence is added by the enzyme tRNA nucleotide transferase. This enzyme binds to the sequence at its active site and phosphodiester bond is formed with the 3 prime end.

Furthermore there is a base modifications occurring simultaneously such as methylation, deaminations or reduction. In case of the pseudo uridine the uracil is removed and reattached to the sugar through the 5 end. So when you when you will when we will going to show you the structure of the tRNA and you will see that there are different regions in that within the tRNA you are going to have the different types of CCA and you are going to have all those kind of things you are going to have anticodon chain and all that. So that time you will be able to understand why the tRNA is actually going to be modified for all these modifications and that is how you are actually going to have the mature product and these mature product are actually going to participate into the protein synthesis machinery. So what we have discussed so far we have discussed about the replications we have discussed about the transcriptions and in the current module we have discussed about the prokaryotic transcription we have discussed about the eukaryotic transcription and in the current lecture we have discussed about the post transcriptional modifications.

So within the post transcriptional modification we discuss about the how the messenger RNA is actually going to be modified so messenger RNA is going to be capped at both the ends it is actually going to have the 5 prime cap and it is going to be protected from the 3 prime end by the having a poly a tail apart from that the messenger RNA is also going to be modified by removing the introns and this process is known as the splicing. So there are going to be 4 different types of splicing introns you are going to have group 1 tron you are going to have group 2 introns you are going to have the spliceosomal introns and you also require the you also going to have the splicing where you does not require the ATP. So these are the 4 different types of splicing different types of introns what are present in the eukaryotic system. In general the introns are not present in the prokaryotic system and they are only present in the eukaryotic system but there are exceptions both in the prokaryotic system and as well as in the eukaryotic system. For example there are yeast genes where there is no introns present and there are prokaryotic bacterias or the RK bacterias where the introns are present.

So these are some exceptions but in general the introns are absent in the prokaryotic system and introns are present in the eukaryotic system. Now introns are actually playing a very crucial role because it is allowing the alternate splicing and alternate splicing is a robust mechanism and a tool through which the eukaryotic organisms are actually been able to produce the multiple types of different types of proteins by the single transcripts.

So within the transcripts the multiple types of genes can come together right for example if within the one transcript you have the 4 genes 1 2 3 4. So either the 1 2 3 can come together or 1 3 4 can come together and that is how you are going to have a protein which is made up of 1 2 3 or 1 3 4. So from the single transcript you can have the different types of proteins.

So and at the end we have also discussed about the both transcriptional modification in the ribosomal RNA and as well as tRNA. You will be able to understand more when we are going to discuss about the ribosomal RNA structure of the ribosomal RNA and the structure of the tRNA in the subsequent modules and why there will be a modification required in these structures. So with this I would like to conclude my lecture here in our subsequent lecture we are going to discuss some more aspects related to molecular biology. Thank you.