Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Aritri Bir

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Week 01

Lecture 03 : Transcription (RNA Synthesis)

Namaskar. So, welcome to the next lecture on Continuing the NPTEL lecture series, Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. So, we are in the very first week where we are discussing the essential fundamentals of molecular biology and molecular diagnostics. In this class we are going to discuss the transcription process which is the RNA synthesis. In this class we are going to discuss broadly the gross mechanism of transcription how the different the RNA is synthesized in different types of cells of prokaryotic cell, eukaryotic cell. The main enzyme which is relevant with this transcription process that is RNA polymerase its function the structure of RNA polymerase different subunits or the subunit based functions in details then a bit of post transcriptional modification of the mRNA.

So, this figure you are quite acquainted with that is the flow of genetic information which is occurring in the under the heading of central dogma where the synthesis of DNA from DNA is the replication and from DNA the synthesis of RNA is known as transcription. So, transcription is basically nothing, but the synthesis the process by which RNA is synthesized from a specific region of DNA. Now, why I said specific region? In case of replication if you remember the whole DNA is copied or replicated, but in case of transcription when the RNA is synthesized the whole DNA is not generating the RNA completely at one time. So, when required the when the protein is required protein is coming from the mRNA.

So, the relevant mRNA is synthesized or rather transcribed from a specific portion of the genome in a specific point of time. So, the genes or the regions which are generating or transcribed into RNA those regions are called structural genes. Now, coming to the transcription process there are two types of strands related to it one is template strand another is coding strand. Now, because the RNA is synthesized taking the template of the DNA. So, here you can see this is the template of the DNA.

So, this is the template strand consider this is the two strand of the DNA. So, if from

these strand RNA is transcribed you can see this is just the complementary base pairing rule. Like this . So, this is our template strand, but if you see the sequence it is similar to the other strand or the non template strand. And the code in the RNA are just similar with the coding strand except instead of the uracil which is present in RNA there is thymine.

Now why I say it is the coding strand because the codes are similar with the RNA. So, sometimes this coding strand is also called the sense strand whereas, the template strand is also known as antisense or non coding strand. So, this concept is clear from the template strand directly RNA is transcribed, but the sequence is matched with the coding or non template strand. Now, like replication there are three phases for transcription that is initiation, elongation and termination and that is same for these three steps present both in eukaryote as well as prokaryote. Now the initiation is that where RNA polymer is the main transcription enzyme recognizes the region the promoter region from where the transcriptions transcription process begins.

And then it it is followed by elongation where definitely the elongation of the RNA strands keeps on growing after that termination of the transcription where RNA polymerase stops synthesizing the RNA and the nascent RNA strand which is just synthesized is separated from the template DNA strand. So, these are the three steps we are going to discuss it in details. But before that we are coming to the main enzyme of transcription which is RNA polymerase. Now RNA polymerase here you can see it is also known as DNA dependent RNA polymerase why because for the synthesis of RNA or by transcription or template is required and that template is provided by the DNA. So, the whole process or the function of RNA polymerase is basically dependent on the DNA.

So, that is DNA dependent RNA polymerase. Along with the DNA template what other things are required in transcription all the four nucleoside five triphosphate or the nucleotides along with that magnesium is required as a cofactor. Now if you remember the mechanism catalytic mechanism of the DNA polymerase RNA polymerase the polymerization process is same that is you can see the 3 prime hydroxyl group here acts as a nucleophile it attacks the phosphates of the incoming nucleoside triphosphates releases a pyrophosphate and forms the bond. Now RNA is synthesized from 5 prime to 3 prime end. So, the template DNA strand is red just the opposite direction that is 3 prime to 5 prime end.

In case of prokaryotic RNA polymerase it has multiple subunits. So, one single RNA polymerase has multiple subunits whereas, eukaryotic RNA polymerase are of different types where the prokaryote remember the prokaryotic RNA polymerase its multiple subunits show different types of function and those different types of functions are basically delivered by different types of RNA polymerase in case of eukaryotes. So,

RNA polymerase like replication unlike replication rather it does not need any primer and the process of transcription begins or initiates in the region of I mean when the RNA polymerase binds the region that is known as promoter. Now coming to the promoter sequence definitely promoter sequence are the consensus sequence present in the template DNA strand which is recognized by the RNA polymerase and then it begins the transcription. So, what is the promoter site? Now this is the this figure is the is resembling the prokaryotic transcriptions process.

So, the promoter site comprises different types of consensus sequence the very known or well characterized sequences are basically those which are present in E coli long studied long described well characterized. So, one of the promoter consensus sequence is present in minus 35 region. Now what is minus 35 region? Remember if that is the suppose this is the template strand this is our template strand and this is the first nucleotide which is transcribed. So, basically in the RNA strand this is the first nucleotide attached and RNA synthesized from 5 prime to 3 prime direction. So, in this way the nucleotides are getting attached to the nascent RNA strand.

So, this is our first nucleotide in the RNA that is the start point and that is denoted as plus 1. Whereas anything before that region or the transcription starts right are represented in minus like minus 1 remember there is no 0 that is plus 1 the next nucleotide is minus 1 minus 2 minus 3 like that. So, in this way in the minus 35 region there is a consensus sequence which resembles TTGACA and another consensus sequence which is present in minus 10 region which resembles TATAAT. These are the very common promoter sequences which are present in prokaryotic templates or prokaryotic transcription. Now in the minus 35 region is the recognition and binding site of RNA polymerase.

Whereas the minus 10 region in this region there is a stable or strong bonding or stable complex formation between the DNA and RNA polymerase. Now the prokaryotic RNA polymerase as I told you it has multiple subunit very commonly it has 6 subunit amongst them there is 2 alpha alpha 2 means there are 2 alpha subunits 1 beta subunit 1 beta dash subunit 1 omega subunit and 1 sigma subunit. Now this sigma subunit excluding that sigma subunit these 5 subunits form the core enzyme core subunit and when with this core enzyme the core subunit sigma is attached that is known as the hollow enzyme. Now all these subunits are performing different types of function the alpha subunits basically determines the which part of the DNA to be transcribed beta subunit performs the main polymerization function beta dash subunit it binds and open the DNA template basically it is showing the helicase function and the sigma subunit that is very important sigma subunit it is responsible for identifying the promoter region. So, sigma identifies the promoter region alpha beforehand decides which region to be transcribed after that beta causes polymerization with the help of beta dash.

Now there are certain drugs which can inhibit this RNA polymerase one such is rifampicin is a very well known drug used as anti tubercular therapy. Now this rifampicin can bind the beta subunit of RNA polymerase and in this way it can inhibit the RNA synthesis definitely for this the protein synthesis is inhibited. Now unlike DNA polymerase which is used in replication RNA polymerase does not have a separate proofreading 3 prime 5 prime exonuclease activity then how the proofreading is delivered by RNA polymerase in case of transcription there are several undefined process one such is the direct reversal of polymerase reaction. So, whenever there is a wrong nucleotide is included there is direct reversal of polymerase reaction and that mistakenly attached nucleotide is removed. Now basically by this it is very much evident that the error is more in case of transcription in comparison to the replication remember replication has a lot of proofreading function and the process is process is conducted with high fidelity replication whereas, transcription the proofreading capability is low.

So, what happens the in comparison the RNA synthesis is more error prone than the DNA synthesis and it is acceptable why because replication is a permanent process remember all the RNA which are synthesized in our body at the end of the procedure those RNA's in the cytosols are degraded by different types of nucleases. So, the function of the RNA is not permanent whereas, replication gives rise to that DNA template which is persistent persistently it is synthesizing different types of RNA at different phases of the cycle. So, the permanent process needs to be more accurate and should be done with more sincerity. So, coming to the prokaryotic transcription process as we discussed initiation is or initiation began when the sigma factor of the RNA polymerase identifies the promoter region and causes the binding. Now, RNA polymerase binds non specifically to one site of the double stranded DNA with low affinity after that this sigma factor starts scanning the whole strands of the DNA till it finds this promoter region and that is where the RNA polymerase is bound with the sigma subunit and forms the close initiation complex.

Now, what is close initiation complex where the two strands of the DNA are yet not separated. When this sigma factor reaches the promoter site the DNA and RNA polymerase interaction the association becomes strong. Now, as I discussed in case of E coli the region promoter region spans from minus 70 to minus 30 and that whole region is scanned by the sigma factor. Now, as I told there is one upstream promoter AT region where the alpha subunit of the alpha subunit of the RNA polymerase binds and decides which segment to be transcribed. Now, all these sequences promoter region sequences minus 35 minus 10 the AT regions all these sequences basically decides the efficiency of the RNA polymerase.

Remember the efficiency is how strongly the RNA polymerase can bind with the

template DNA strand the association also. Remember the function of RNA polymerase is such when it starts polymerization or adding the nucleotides in the nascent RNA after adding some nucleotides it gets dissociated. So, in between the one cycle of association and dissociation how much nucleotide is attached to the to the nascent RNA strand is basically the efficiency of the RNA polymerase. So, we were discussing the prokaryotic transcriptional initiation and we told that there is a closed initiation complex. So, similarly there should be a an open initiation complex as well.

So, open initiation open complex is basically where the 2 strands of DNA are opened up or means the hydrogen bonding bonds are dissociated and forms the open complex with the RNA polymerase. Now, there is formation of another structure which is known as transcription bubble. Transcription bubble is basically formed at that minus 10 region which is AT reach. Now, in comparison to the GC region which has 3 hydrogen bond AT has 2. So, definitely those region which are reach in AT those are the weakest site where the hydrogen bonds can be opened up.

So, at that region unwinding of the DNA occurs and in that region only transcription bubble is formed. So, you can see this is the transcription bubble here the 2 strands of DNA are opened up and this is our template strand where the new RNA strand is getting synthesized this is the new RNA strand and this is our RNA polymerase. The this orange thing this orange thing is the whole RNA polymerase and you can see that a bubble like structure is formed here that is the transcription bubble. Now, after RNA polymerase is binding to the promoter region the first nucleotide is attached the that is the initiation or the initiation plus one site generally that first nucleotide is a purine nucleotide. Now, as I told you this RNA polymerase after adding a few nucleotides gets dissociated from the transcription bubble or the complex.

Now, these are the types of abortive transcription what happens when a very short chain is formed like 5 to 6 nucleotides long chain is formed and the RNA polymerase is dissociated those that nascent chain along with that RNA polymerase dissociates and gets dissolved. Whereas, after this such type of multiple abortive transcription a stable mRNA or nascent mRNA is formed which is the 12 around 12 base pair long. Now, this is the longer one nascent mRNA. So, here it can bind this mRNA nascent mRNA can bind strongly with the template strand after that the sigma subunit gets detached. So, basically when that sigma subunit is detached from the RNA polymerase what remains are the 5 subunits which forms the core polymerase enzyme and this is the signal for the end of initiation.

So, remember when the when the RNA polymerase binds with the RNA template strand from then only the initiation begins begins and then the sigma factor starts scanning finds the promoter region after a few abortive transcription when a stable nascent RNA

mRNA is formed sigma subunit then as a whole gets dissociated from the RNA polymerase leaving the core enzyme and that is the signal for the end of the initiation then what begins is the elongation phase. Now, what happens when the transcription goes on there is some conformational changes which converts the complex to the elong I mean it it helps the complex converted to the elongation form and in that time what happens as a whole the RNA polymerase moves after because it has already transcribed the initiation initiating part of the template DNA it moves along the promoter and cleared the promoter region it was initially has the RNA polymerase initially has occupied the promoter region now it has been cleared and that is known as promoter clearance. So, basically what happens RNA polymerase if that is your promoter region and here the first initiation has begun. So, the RNA polymerase move or slides along the DNA strand clearing this promoter region. So, the next RNA polymerase will be like this after sliding towards this.

So, the promoter region is cleared and that is promoter clearance. Now, sigma cycle is basically associated with one cycle of transcription over and beginning of another cycle. So, what happens at the end of the whole process of transcription the termination of transcription sigma once again is attached to the RNA polymerase and as a whole the whole RNA polymerase now is detached from the template strand. So, this is known as sigma cycle.

Next coming to the elongation. So, here you can see the transcription bubble where this is the template strand this is the non template or coding strand here you can see this green color is our nascent RNA this orange is as a whole the RNA polymerase and you can see this part this 2 strands of DNA have opened up in the template strand the nascent DNA is attached forming RNA DNA hybrid and the transcription is going on to this direction and that is our transcription bubble. Now, transcription bubble contains RNA polymerase anion DNA strand both the strands newly synthesized mRNA strand which is 18 to 20 base paired 18 to 20 base paired spanning 18 to 20 base pair anion region newly synthesized RNA is hanging from the DNA. So, those part which are attached here others are hanging and approximately 8 to 9 base pair long nascent segment is basically paired with the template strand. Now, unwinding and rewinding. So, this part when it is detached and this is our unwinding part.

So, the when this unwinding moves on this way this segment is rewind rewound rather with the enzymes and this is occurring simultaneously. Now, what is formed here is the binding of here once again some super coils which are released by DNA topoisomer. Now, prokaryotic transcription is terminated following 2 types of signals one is row factor dependent another is row factor independent. Now, coming to the row factor independent type of transmission termination here is an RNA sequence which is a palindromic sequence. So, it can form self complementary sequences thus form the

hairpin structure how you can see here the sequence is like that if you take this sequence these are our self complementary sequence.

So, like that. So, this segment if it is folded upon each other the segment other can form a hairpin structure. So, row factor in case of row factor independent termination this type of region is present that is GC ridge palindromic region followed by multiple A in the template strand. So, the template strand is having GC ridge region followed by multiple A. So, what happens when transcribed RNA forms this type of hairpin structure. So, these are folded upon each other see these are folded upon each other.

So, this is the hairpin loop formation. Now, again the UA bond is the weakest bond because this is having only 2 double bonds. So, all these UA bonds are basically the poor one and it is considered as the termination signal. Now, coming to the row dependent termination where there is one specific row factor. Now, row is one hexameric protein which binds to a nucleotide region which is around 72 nucleotide long. Now, there is a region which is CA ridge region where row utilization element is present also the row factor it shows ATPase or ATP dependent helicase function which causes the dissolution of the hydrogen bonds created between the RNA and DNA in RNA-DNA hybrid.

Now, what happens row protein here you can see this row factor moves on with the RNA polymerase till it finds the a signal which is here you can see that region is a CA ridge region and the row protein is found. Now, row protein while migrating towards 5 prime to 3 prime direction when it finds this signal gives a pause and releases the RNA transcript from the RNA-DNA hybrid via utilizing the helicase property of its own. So, this is the row dependent termination. Now, if we consider the eukaryotic transcription process the process as a whole is seen in similar with the prokaryotic one, but definitely there are differences like here the RNA polymerase is of different types instead of having different multiple subunits here it is of different types. The polymerase 1 is synthesizing the ribosomal RNA or pre rRNA or the precursors of 18 S 5 point S and 20 S rRNA polymerase 3 is synthesizing the very important mRNA polymerase sorry polymerase 2 and polymerase 3 is synthesizing the remaining ribosomal rRNA that is 5 S rRNA and also the tRNA.

So, based on the their specific type of RNA specific synthesis of different types of RNA there are different types of RNA polymerase. Now, polymerase 2 definitely is the main enzyme which causes the synthesis of mRNA and it is also a more complex counterpart of the bacterial RNA polymerase. It has 12 different subunit in case of the yeast RNA polymerase which is well studied, but the subunits of the prokaryotic RNA polymerase their functions are more or less similar with the RNA polymerase 2 subunit like the 2 large subunit resembles the beta and beta dash subunits of the prokaryotic RNA polymerase. Apart from that there is one C terminal or carboxy terminal domain in RNA

polymerase 2 that helps in the activation of the enzyme by phosphorylation and it is this CTD is basically formed by repeats of consensus heptad amino acid sequence. Then in case of initiation in eukaryotic transcription there are different basal transcription factors or general transcription factors which are responsible for identification of the promoter site.

After the promoter site identification by these general transcription factors there is formation of pre initiation complex which was not present in prokaryotes. Then there are core promoter element of the polymerase 2 helps in the transcription process also there is enhancer and silencer sequence. So, enhancer sequence which gives signal for the transcription more transcription whereas, silences causes a stop in the transcription. So, elongation more or less same, but there are different types of elongation factors present in eukaryotes and the process as a whole takes long time so slow. Then termination is associated with one polyadenylation in the 3 prime end.

So, these are the difference in case of eukaryotic transcription with comparison to the prokaryotic one. Now if we come to the initiation where I told that there are different general transcription factor which causes identification of the promoter site. You can see there are different types of transcription factors from tata box binding protein which binds to the promoter sequence tata box. Then it is identified by the transcription factor 2 B followed by the transcription factor 2 F and polymerase 2 basically this transcription factor 2 F helps the binding of polymerase 2 to the complex followed by addition of transcription factor 2 E and 2 H here the transcription factor 2 H helps or rather acts as helicase. So, all these factors are named as 2 because it is associated with polymerase 2 and here the different types of initiation signal in case of eukaryotes are tata box where the tata box binding protein is attached there is one initiator sequence one downstream promoter element.

Now how frequently the transcription will happen in a region is denoted by some frequency signals like g c box or cat box also there are enhancer and silencers. So, all these basically decide how frequently the transcription will happen where it will happen and how it will happen. So, you can see there is formation of a close initiation complex which when is conduct or acted upon by transcription factor 2 H which causes opening up of the 2 DNase transforms the open complex and on dissociation of this transcription factor 2 H and 2 E elongation begins. Now elongation has different other elongation factors like ELL or elongin and finally, termination is associated with addition of a poly H and E tail in the 3 prime end. Now polymerase 2 as a whole is continuously attached with the template RNA strand during activation remember CTD or the carboxyl terminal domain was has been phosphorylated for the activation of polymerase 2.

So, similarly on dephosphorylation factor it gets dissociated and recycled once again for

synthesizing another mRNA transcript. So, these are the differences in case of eukaryotic transcription which does not occur in prokaryotes. Next we are coming to one very important thing that is post transcriptional modification. Now after the mRNA has been synthesized there are multiple modifications over this mRNA. So, here you can see on transcription the mRNA which is synthesized as the primary transcript which is known as pre mRNA which contains all the exons and introns inside it.

So, remember exons and introns which we have decides in the in our very first class that is the coding part and the non coding part. So, the RNA is processed in such a way that the non coding part or the introns are basically cut off from it along with that there is addition of one cap at the 5 prime end and one poly A tail at the 3 prime end. So, all these are combined under the heading of post transcriptional modification. So, let us check one by one.

So, there is a cap added at the 5 prime end. So, that cap is formed by 7 methyl GTP the enzyme which helps in the addition is a guanilyltransferase and guanilyltransferase cause the methylation of guanine by S adenosyl methionine. Now, why it is helpful because the 5 prime end is open before capping it is vulnerable to digestion by different 5 prime exonucleases. Remember the transcription of RNA after the transcription of mRNA it is it is transferred from the nucleus to the cytosol and the cytosol contains a number of exonucleases. So, for the protection of from the for the protection from these exonucleases the 5 prime end is capped. Also it helps in this transfer of mRNA from nucleus to cytosol also different transcription transcription factor bindings are also helped.

Then coming to poly A tailing or addition of around 80 to 250 adenine nucleotides at the 3 prime end with the help of the enzyme poly adenylate polymerase. This also stabilize the mRNA by giving protection from the 3 prime exonuclease. Next coming to the splicing. Splicing is basically cutting of the introns or the non coding regions from the pre mRNA and joining all the coding region or exons here. So, this splicing mechanism can be self splicing where the enzymes are located within the mRNA or the pre mRNA itself or can be done by some other factors which together forms a spliceosome.

So, let us come to the gist of this session where we have discussed that there are 3 steps of transcription that is initiation elongation and termination. The main enzyme which is responsible for the transcription is a DNA dependent RNA polymerase. In case of prokaryote that RNA polymerase has 6 subunit whereas, eukaryotes have different types of RNA polymerase. Anyway the transcription begins after identification of the promoter site and finally, mRNA undergoes multiple post transcriptional modifications like 5 prime capping, 3 prime poly A tailing or splicing by cutting of the non coding regions of

the pre mRNA.

These are my references. Thank you all. See you in the next class.