Organic Chemistry In Biology And Drug Development

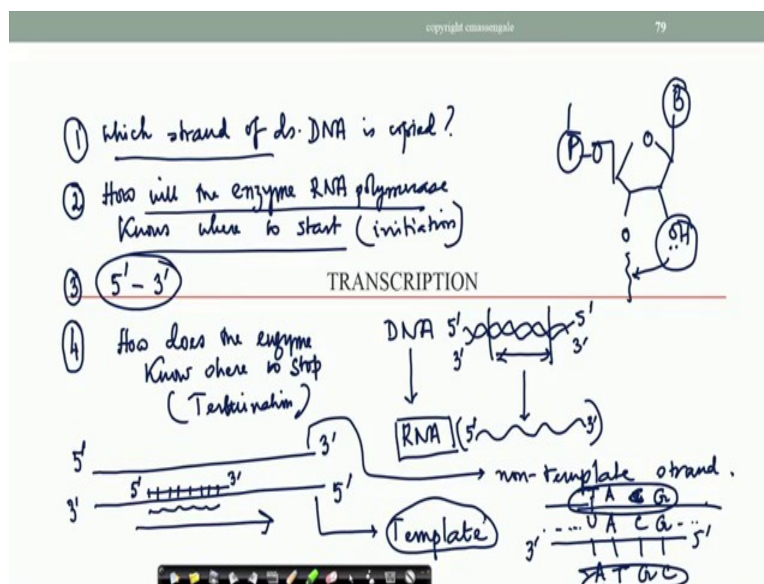
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 29
Central Dogma: DNA Replication, Transcription and Translation (Contd.)

Welcome back to this course on Organic Chemistry in Biology and Drug Development. In the last session basically we have discussed the first stage in the flow of information in the biological system. We have read about this central dogma in biology.

So, for cell division process several steps have to be taken and the first step is that you have to copy the existing DNA into daughter DNAs. We know that was called replication and the replication process is semi conservative.

(Refer Slide Time: 00:23)



Now, that what is the next step? The next step is now what is called transcription. Now the information has to flow from the DNA and according to the information stored in the DNA you will make ribonucleic acids . Ribonucleic acids are little different from the deoxyribonucleic acid. It is clear from the name that in ribonucleic acid you have is the ribose unit; that means, that 2 prime and the 3 prime hydroxyl groups are present in ribonucleic acid. So, it is a full ribose. So, there is that 2 prime OH, the base and the phosphodiester linkage.

Here another difference is that thymine is replaced by This is connected via phosphodiester bond and these two prime OH can hydrolyze the phosphodiester bond by a process what is known as neighboring group participation that is very common in organic reactions.

We are going to talk about transcription. Transcription is the process where a DNA is copied into RNA. Now, the DNA is double stranded. But RNAs that is going to be made from the DNA is single stranded. Now, the question is that how this process takes place because there are two strands of DNA. The first question is there are two strands of DNA and one strand runs from 5 prime to 3 prime and the other in the anti-parallel direction 3 prime to 5 prime.

These two strands are complementary and their complementarity comes from the Watson-Crick base pairing. When it is transcribed to make the mRNA, mRNA will also have a direction. Suppose this is the 5 prime to 3 prime. Now, the question is which of these strands is actually copied because information is stored in both the strands. Although these informations are complimentary, one of the strands information has to be copied to make the RNA when it is transcribed into RNA.

Which strand of the double stranded DNA is copied? When this transcription takes place you have a well-defined size of the RNA. This is different from the replication process because replication process needs to copy the entire piece of DNA. But when RNA is made from some segment of the DNA, full DNA has need not to be copied. So, part of the DNA has to be copied and transcribed into the RNA.

Which is making the RNA molecules? It is a polymerization reaction. Now, instead of DNA polymerase you have is RNA polymerase. How the enzyme RNA polymerase knows where to start? Suppose this is the information that needs to be transcribed here. How does the enzyme knows that we have to start from this side. So, that is the second question.

What is the direction of the synthesis of the mRNA that is again 5 prime to 3 prime direction? So, there is no change in the oligonucleotides and the DNA chain extends in a 5 prime to 3 prime direction. The 3 prime hydroxyl attacks the 5 prime triphosphate. The same thing happens here. How does the enzyme know where to stop? That means, there should be a termination mechanism.

So, this can be called a termination mechanism and this can be called an initiation mechanism or an initiation step.

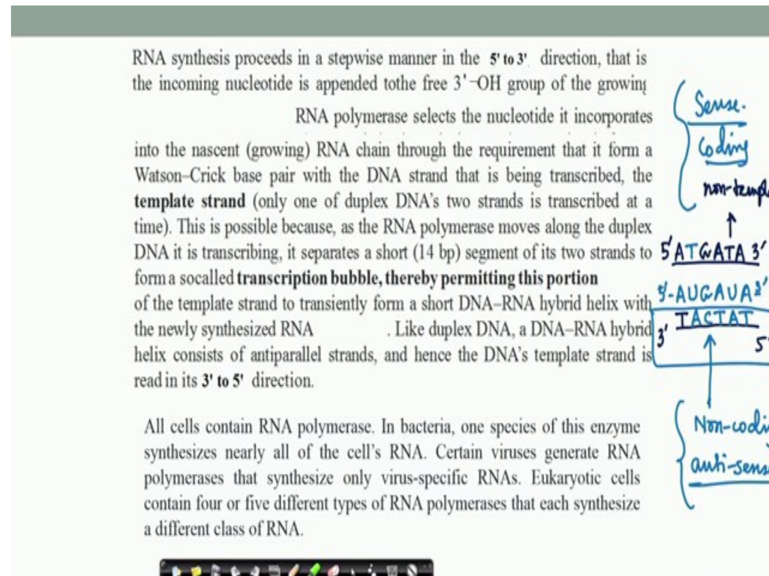
Which strand of the DNA is copied? Now, remember there are two strands of DNA. Suppose I write it in this fashion in a linear fashion. Suppose, this is your 5 prime to 3 prime end and this is the 3 prime to 5 prime end.

Now, if one of them actually acts as the template on which the RNA polymerase works to make the RNA. Suppose the RNA polymerase is bound to these strands runs from 3 prime to 5 prime and then it starts to make the RNA. If this is the strand then there will be no direction problem. The direction of synthesis is following this 5 prime to 3 prime.

This is now acting as a template strand. It acts as a template and this is DNA dependent RNA polymerase. This template is actually a DNA molecule or a deoxy oligonucleotide. So, this is a template strand and this will be the non-template strand. We are now distinguishing between the two strands. The strand of DNA which is copied to make the RNA is called the template strand and the other strand is called non template.

When the RNA is copied it will take the complementary bases of the template strand. Just as an example, base sequence of template DNA strand is ATGC. This runs from 3 prime to 5 prime. Now, the RNA will be made here. The sequence of RNA will be UACG. Now in RNA, you have U instead of T. Here non-template strand of DNA will have a sequence of T, A, C, G. The sequences are virtually same except where T is replaced by U in the RNA.

(Refer Slide Time: 12:31)



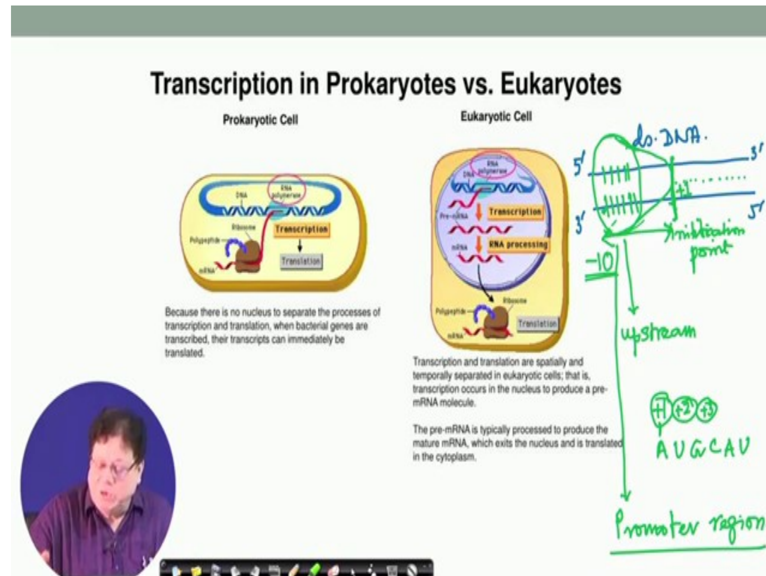
Suppose I have a non-template strand which is in the 5 prime to 3 prime directions and suppose the sequence is TACTAT.

Now what is the sequence of the mRNA that is synthesized? So sequence of non-template strand will be same as sequence of m RNA. Only T will be replaced by U.

The base sequence of template strand and base sequence mRNA will be complementary to each other. Other names have been given to this template strand and the non-template strand. Template strand is also called the non-coding. Non template strand is known as coding strand because the codes match with each other; except that T and U.

Template strand is also known as antisense strand and non-template strand is also known as sense strand.

(Refer Slide Time: 16:11)



How does the polymerase know that where to bind and where to start? Suppose you have this duplex DNA. So, this is 3 prime, this is 5 prime. This is your DNA which needs to be copied and transcribed into the RNA.

Now, there is a big RNA polymerase which will traverse to it and it is searching for a particular sequence. Once finds that sequence in a region called a promoter region then it binds here. It is usually 6 base pair sequence. RNA polymerase is a very big molecule. It can cover the graph up to the initiation point. The RNA was actually traversing on the surface.

When it is traversing on the surface it was looking for some sequence. The whole molecule stretches far beyond this sequence and this is the initiation point from where it starts. Then the oligonucleotides are added one after another.

Now, this region is called the upstream region because your synthesis is going in the reverse direction. So, actually the synthesis starts in the upstream direction. We are talking about the sequence which the RNA polymerase recognizes. So, this is in the upstream. How many bases are there in this gap?

During transcription of RNA, the first nucleotide is added at the terminal 5 prime end. The first nucleotide is called plus 1. If the first nucleotide has a sequence of A U G C A

U then this is called a plus 1 base, this is the plus 2. So, in that way you are increasing the number of bases that to be added.

At the initiation point, the first one is called plus 1 and then this is upstream. I said how much it is far from the plus 1 site. That is usually minus 10. There are about 10 nucleotide bases. Now, this is very important region because if this is not present then RNA cannot tell where to start. So, the first the RNA has to find this sequence at the promoter region.

6 bases are present in the promoter region. Here the RNA polymerase binds and accordingly 10 bases downstream the RNA synthesis starts.

(Refer Slide Time: 21:36)

Scientists have found that promoter region is present in DNA. Once it binds to the promoter the DNA double helix has to be separate into the template and the non-template strand. The template strand is then copied and RNA is made.

So, the first step is binding of the RNA polymerase to the double stranded DNA. Here the slide finds the promoter region. Once it finds the promoter region it knows that 10 bases downstream. It finds the promoter region, unwind the double helix and then synthesis of the RNA strand starts at the site. Position of initiation site is called the plus 1. So, I think that is more or less clear.

(Refer Slide Time: 23:06)

It is schematically shown here - this is the double stranded DNA and DNA promoter site is somewhere here. Then the DNA gets melted but not the completely melted. The position where DNA is melted is called a bubble. This is like a transient bubble. Now, it starts the synthesis of the RNA and as this RNA polymerase moves and this transient bubble will now also proceed along with it.

So, the RNA polymerase moves from 5 prime to 3 prime direction resulting in the movement of the transient bubble. This bubble is basically a region where it is inflated. It is inflated because the two strands are further apart.

(Refer Slide Time: 24:43)

Now, there are two kinds major division in prokaryotes. They are simple cells having no further inner compartmentalization. Now, the steps of replication, transcription and translation are very well studied in prokaryotes. In eukaryotes, these things become more complicated. For this course, we will try to stick to the mechanism that happens in a prokaryotic system.

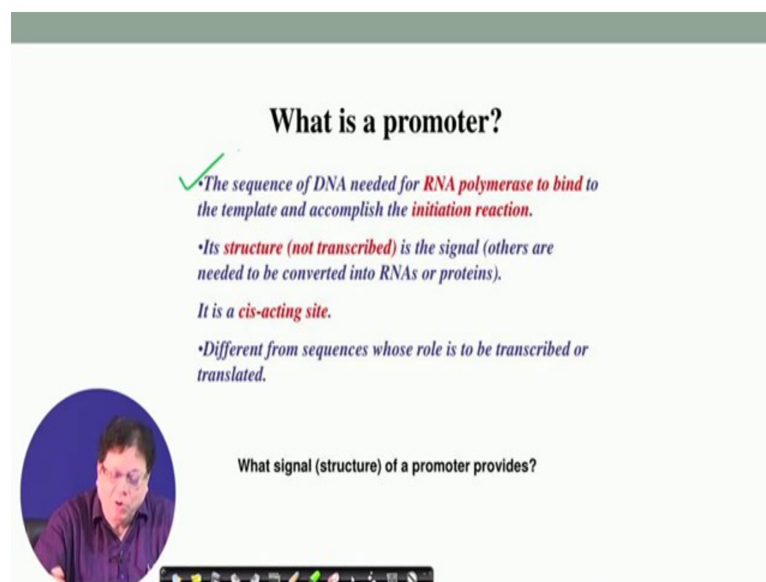
Now, the prokaryotes first of all use a single RNA polymerase enzyme. There are major three classes of RNA. One is the messenger RNA, another is ribosomal RNA and the third one is tRNAs. But there are other types of RNA like siRNA, miRNA. But for this transcription and translation, we will just stick to these three broad classifications of RNAs. Because for the transcription and translation, we need all these three types of RNAs. Only a single RNA polymerase enzyme synthesizes all types of RNA.

Now, I will briefly tell you what the function of these RNAs is. mRNA is the template for protein synthesis; that means, according to the best sequence in mRNA the protein will be synthesized. The amino acid sequence in the protein will depend upon the base sequence in the mRNA. Ribosomal RNA actually helps the mRNA to bind to the ribosome. Ribosome is where the protein synthesis takes place. The translation is taking place at mRNA. The binding process between the mRNA and the ribosome is dependent on the rRNA. Besides that rRNA also helps to synthesize the proteins. It acts as a catalyst to make the peptide bond. We will go to that little later.

tRNA i.e. transfer RNA transfers the required amino acids to a going to a peptide chain. So, all the addition of amino acids is basically through the tRNA. tRNA takes of the amino acid and then brings it to the ribosome and then the amino acid is delivered to make the peptide.

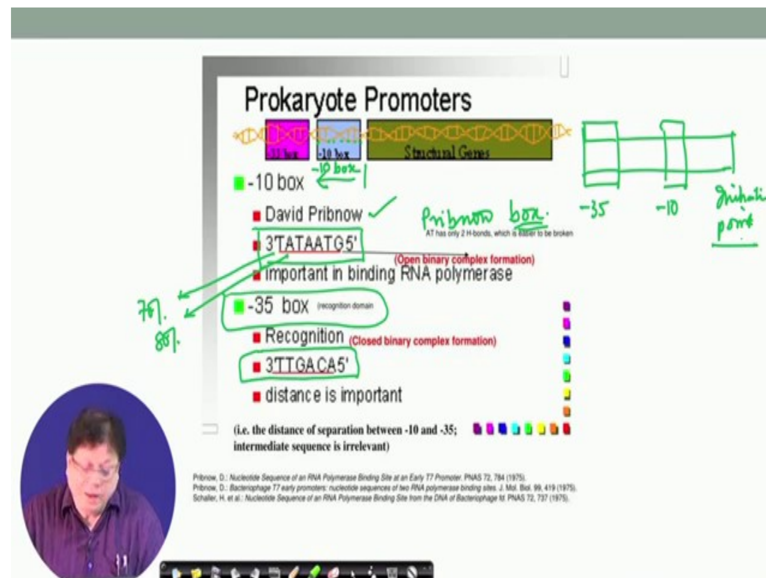
On the other hand, eukaryotes have three RNA polymerase. RNA polymerase 1 that synthesizes rRNA, polymerase 2 synthesizes mRNA, polymerase 3 synthesizes tRNA and part of the ribosomal RNA.

(Refer Slide Time: 29:15)



I have already told the base sequence of DNA in promoter region that needed for RNA polymerase to bind to the template and accomplish the initiation reaction.

(Refer Slide Time: 39:37)



What is the sequence of the promoter region? This is the upstream region. So, there is something called a minus 10 box. There are some base sequence that is called minus 10 box and that is also called Pribnow box. David Pribnow is the person who first discovered this binding of RNA polymerase during transcription. At the minus 10 upstream, this sequence is present and that is known as Pribnow box minus 10.

Now, what is the sequence? The usual sequence for the Pribnow box is TATAATG. But, this sequence is not consensus. All prokaryotes may not have the same sequence whatever is written here. There is a Pribnow box which has got a particular sequence that may vary little bit from one bacteria to another bacteria. But there is some consensus sequence at different locations.

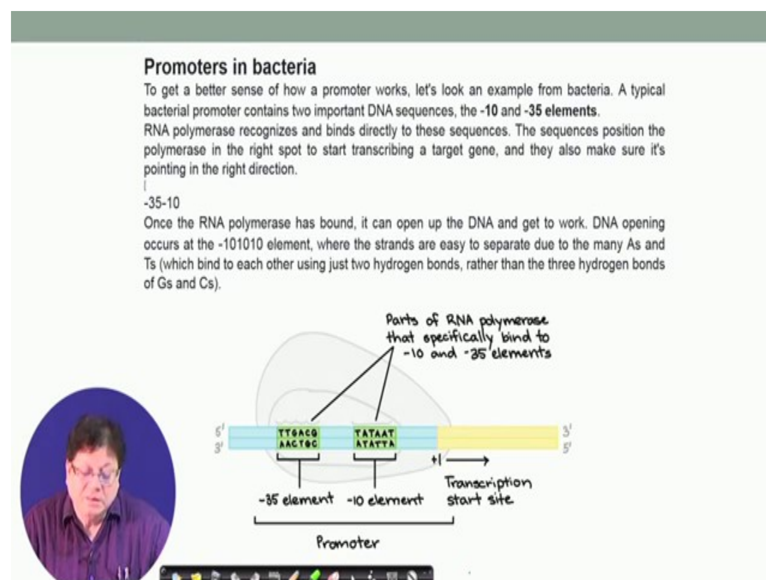
Suppose in one bacteria you need a T at this point. But if you screen several bacteria you will find that maybe 70 percent of the bacteria have There. Similarly, 80 percent of the bacteria have A here. So, this is little bit variable.

This variation actually is translated into the capacity of the RNA polymerase to bind to the DNA. Stronger is the binding to the DNA transcription process will be expressed in a better way. For weaker binding, the transcription process may not be efficient. This transcription process can be modulated by changing the sequence of the Pribnow box.

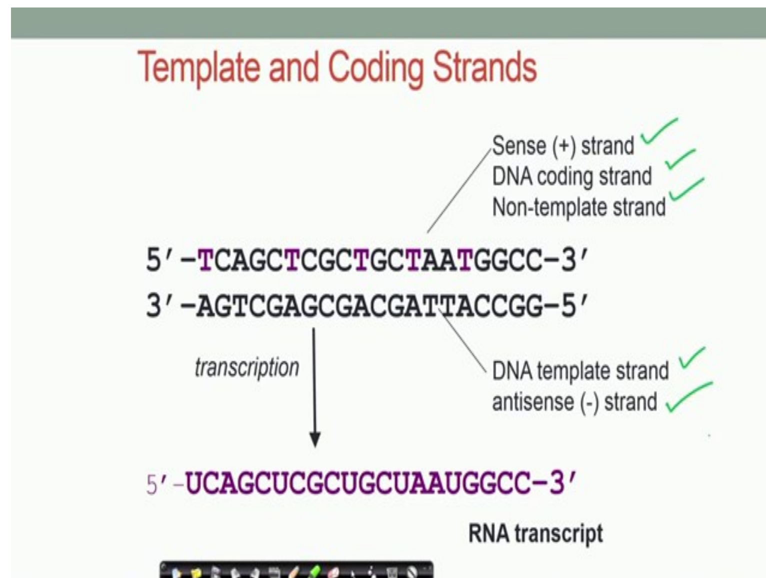
This promoter region can be manipulated and you can have different expression levels. This is called expression level of RNAs. If that is different expression level of RNA there will be different expression level of the protein. Ultimately it is translated into the protein. Apart from this minus 10 box scientists have also identified another recognition domain that is called minus 35 box. The sequence is more or less consensus. This sequence is TTGACA.

In prokaryotes, there are two types of sequences- one is at the minus 35 which is called the Pribnow box and other is minus 10 which is called initiation point.

(Refer Slide Time: 33:48)

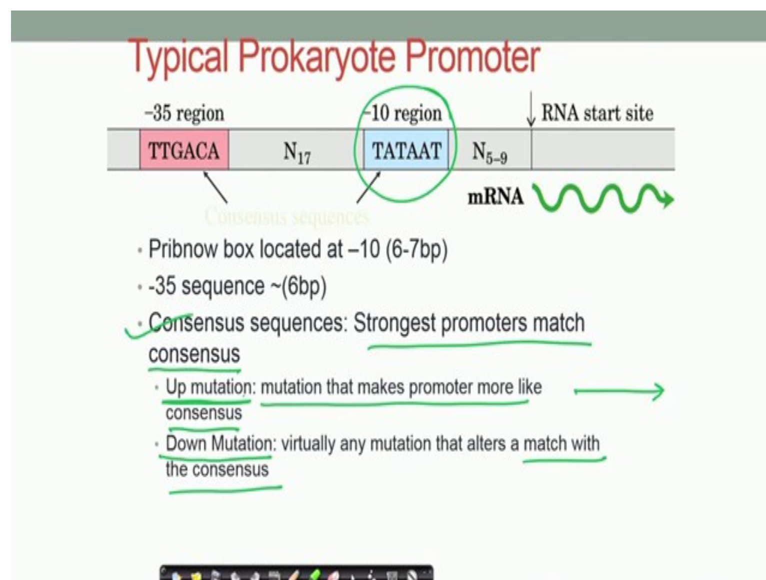


(Refer Slide Time: 34:01)



Let us come to the template. Non-template strand is also called coding strand and template strand is called anti-sense strand. This is the antisense strand.

(Refer Slide Time: 34:28)

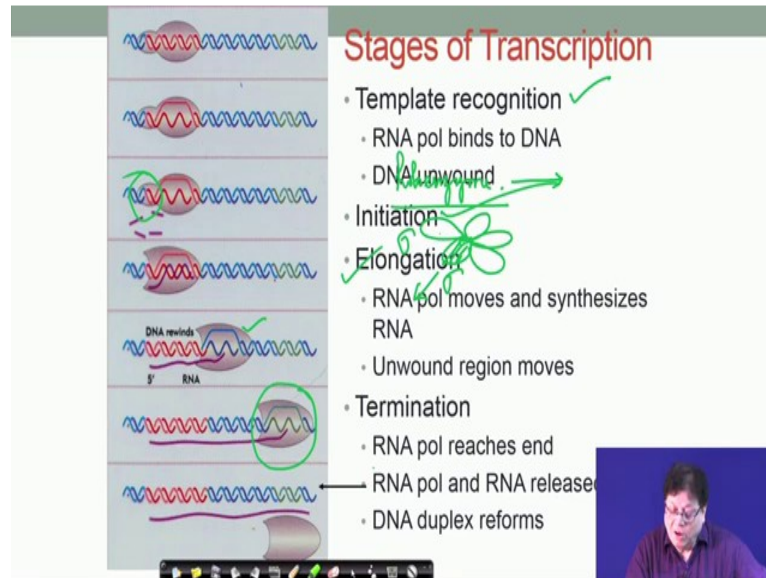


In case of eukaryotes, there is also a promoter region but it may vary from species to species ok. Promoter region has consensus sequence.

Whatever RNA will be synthesized that will be called up mutation. Your ultimate expression level will increase and if there is less consensus promoter region does not bind that well. Then that is called down mutation. Virtually any mutation that alters a

match with the consensus is called down mutation up mutation. The expression level of the final transcribed RNA and that will have a consequence of the finally transcribed protein.

(Refer Slide Time: 35:52)



We know the stages of a transcription that is shown here. First it reconiges the template. RNA polymerase has got a quaternary structure. There are five subunits- two alpha subunit, one beta subunit, one beta prime subunit and one sigma factor. Now, when the sigma factor is present it is called the holoenzyme.

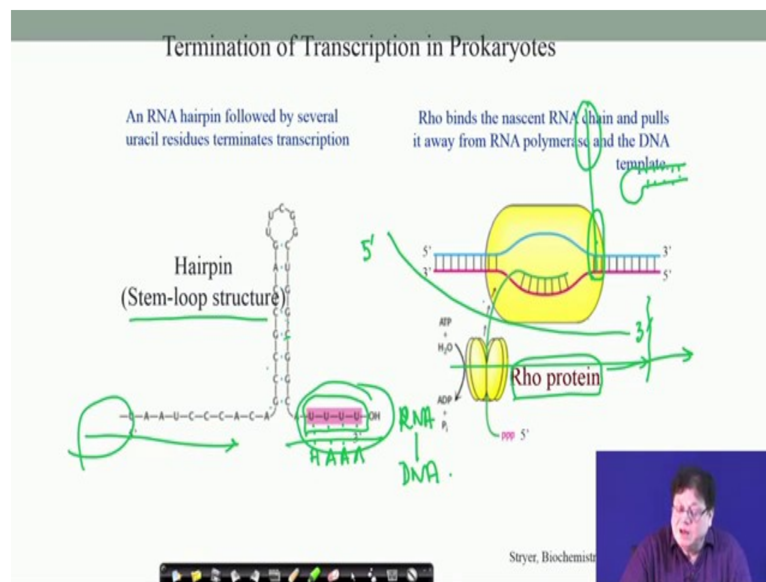
The holoenzyme is basically applicable to the enzyme chemistry. The small molecule is combined with the enzyme is called holoenzye. When the small molecule acting as cofactor or coenzyme goes off from the enzyme then it is called apoenzyme.

When sigma is present in RNA polymerase the polymerase recognizes the promoter chain and then this sigma falls off. When the sigma falls off no longer binding is in the promoter region. This is your promoter region. So, now the polymerase is free. It can now melt the DNA i.e. forms the bubble. Then slowly transcribe - one by one bring that RNA bases and then form the growing RNA chain.

Then it enters the terminal stage. My next question is how it terminates this RNA chain synthesis. So, there must be something which the RNA polymerase is recognizing. Now, I will show these steps again. First is RNA template recognition, second is initiation,

third is elongation i.e. bring more bases synthesized. As the RNA polymerase moves forward the DNA gets melted. That means double helix becomes separated from each other forming a transcription bubble and then finally undergoes termination. So, the polymerase reaches the end.

(Refer Slide Time: 29:04)



Now, how the polymerase does knows that it has reached the end. So, when these polymerase is making the RNA, that is slowly moving towards this direction. If the polymerase moves in this direction, the RNA is synthesized like this. This is your 5 prime end and this is your 3 prime end and this is the movement of the RNA polymerase. When it is towards the end RNA can form a hairpin like structure.

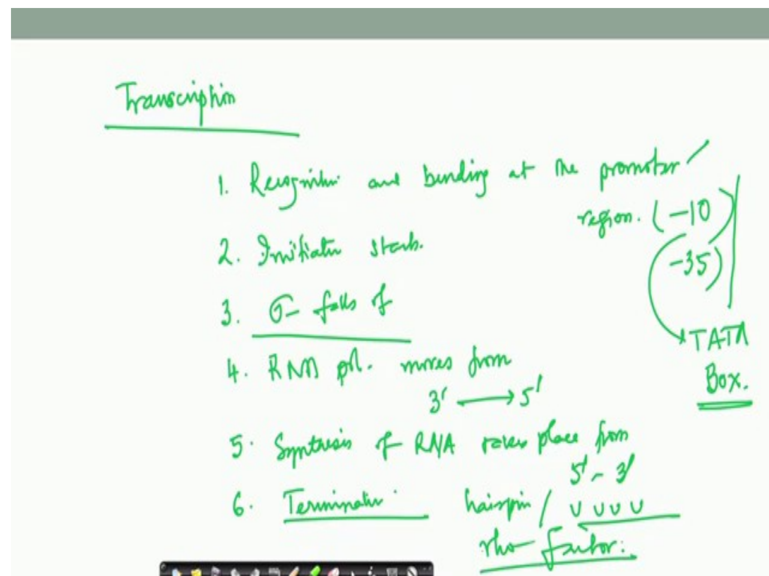
Some portion of the RNA is complementary to the sequence that is present in this. The RNA can take a shape like this because this is complementary to that part. It can form a hairpin. Hairpin means the pin or clip that is used to tie up the hair. So, that is called hairpin. Then after the hairpin, there is lot of Uracil.

If this happens then the RNA polymerase is moving towards this direction. Now, it suddenly sees a loop. So, there is a barrier here. Once it sees the barrier the RNA polymerase stops bringing more oligonucleotides and it falls off. Why there is lot of U? Because we know that U is actually complimentary to A. You know that A and U are actually having only two hydrogen bonds.

So, to release or to melt this part of the RNA-DNA duplex will be easier. The stronger hydrogen bonding pair is the GC. If there is lot of GCs then it will require much more energy to separate this RNA from the DNA. So, nature has done these arrangements for termination step. Hairpin is very strong because there are lots of GCs complementary to each other.

The mechanism that is also operative. This is called a rho dependent release. There is an enzyme which is called rho protein that binds here and then using ATP as the energy source it releases the RNA from the RNA-DNA duplex.

(Refer Slide Time: 43:11)



Now in summary of transcription, the first step is recognition and binding at the promoter region. Promoter region has particular sequence. The sequence in the promoter region is usually minus 10 base pair upstream. Then there is also minus 35 in prokaryotes. By the way, there are a lot of Ts. This is called a TATA box because there are lots of Ts. In case of eukaryotes, this is much more complicated. There are different types of promoter regions, but they are wide apart in eukaryotes. One could be at minus 40 another could be at minus 70, 80 or 100. So, there is a large gap between the two.

In the second step, initiation starts. Until sigma falls the RNA polymerase cannot move. RNA polymerase moves from 3 prime to 5 prime directions and synthesis of RNA takes place from 5 prime to 3 prime directions. Then the next step is termination. Termination can be of two mechanisms- one is a hairpin followed by lot of Uracil to melt the DNA or

to release the DNA from the RNA DNA hybrid or by a mechanism which is dependent on a rho factor. So, that is all about transcription.

Thank you.