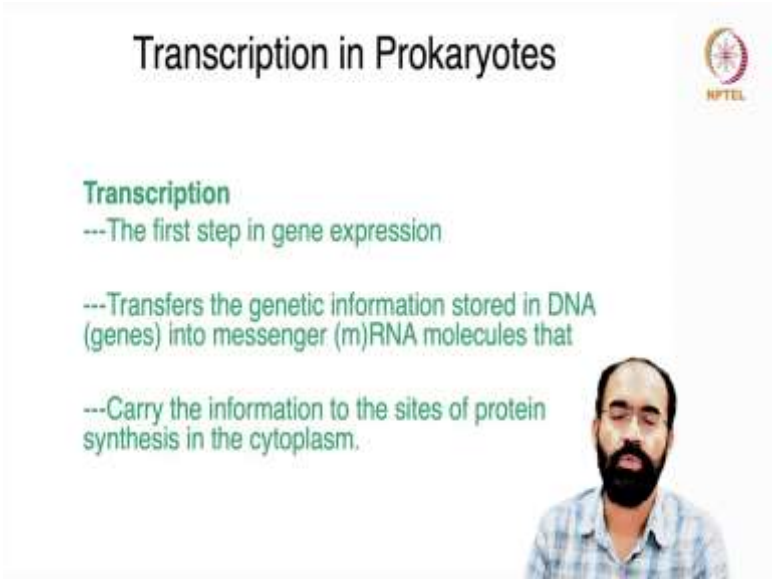
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 12
RNA Transcription: Different Stages

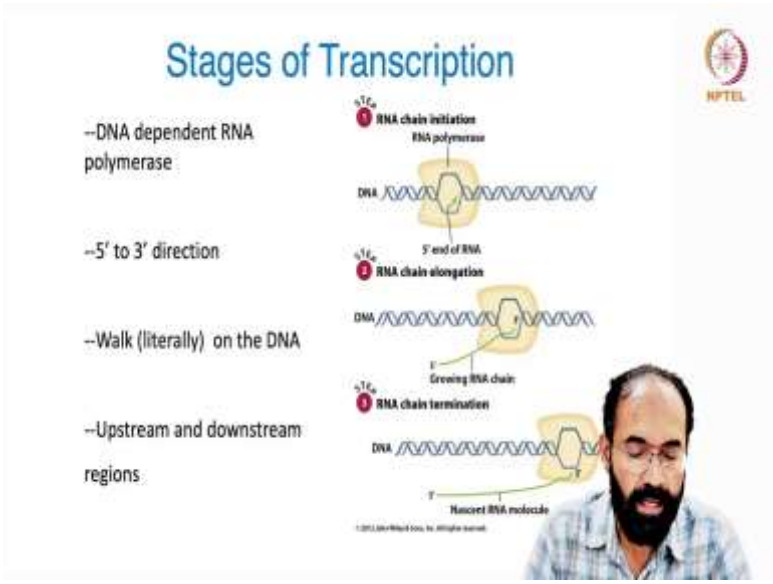
(Refer Slide Time: 00:21)



Welcome back to another session of RNA Biology and we left in this topic that is the Transcription in Prokaryotes. And, transcription in any animal is the first step in the gene expression and the gene expression as I already told you is a very loaded terminology, in the sense it can be influenced in multiple ways. It is not as simple as we portray or we assume.

So, as and when we interact with each sections or we will deal with each sections we will come to that topic and revisit the so called gene expression. And, remember in the simplistic term gene expression is the conversion of a existing genetic information into a meaningful or useful or functional stuff, let it be anything.

(Refer Slide Time: 01:31)

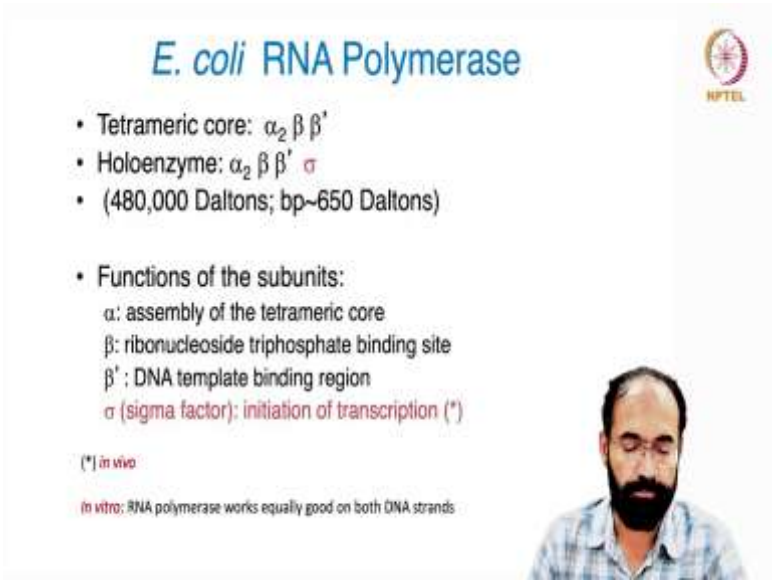


And, if you look into the stages of transcription, it is basically started with DNA dependent RNA polymerase and it happens unidirectionally unidirectionally that is 5 prime to 3 prime direction only. And, it the RNA polymerase walks literally on the DNA by unwinding it locally as and when required. And, it also have specific upstream and downstream sequence which enables the proper defining of a gene region or proper defining of the genes boundary.

So, the RNA polymerase initiates a RNA chain and that is called RNA chain initiation and then comes the RNA chain elongation that is the second step in the RNA production or the mRNA transcription. And, the third step is the chain termination. This production of RNA is varying or the rate of production of an RNA from a gene vary significantly from organism to organism and even from within one organism gene to gene.

Not all genes are produced at the same rate, not all organisms produce a given RNA at the same rate. So, it varies significantly.

(Refer Slide Time: 03:01)



And, if you look into E. coli RNA polymerase; E. coli is a bacterium which is living in our intestine, our body which is actual name is Escherichia coli and it is a well studied model organism for various aspects of you know bacterial physiology. Or even it has helped us in understanding how bacterial pathogens function in you know pathogens cannot be used for pathogen is those bacteria which can cause infection in you.

E. coli is rather harmless bacteria which is extensively used for studying several biological process related to prokaryotes. And it is a tetrameric core. So, the E. coli RNA polymerase contain tetrameric core. What are they? Two alpha units – alpha and alpha 2 alpha units and one beta and a beta dash.

So, this four protein subunits constitute the tetrameric core of the RNA polymerase in bacteria and then if you want to have a hollow enzyme or rather a complete enzyme, this tetrameric core is the crucial part, but it is not the complete enzyme. Say for example, if you someone ask you how do you eat food? So, you will say through the mouth. So, the core part is the mouth, but mouth cannot exist unless you have a face, unless you have a body. Mouth cannot exist in the atmosphere, right.

So, the total human or total organism have a mouth attached on to their face. So, a holoenzyme is a necessity for this tetrameric core to perform the actual task. So, tetrameric core plus a new factor that is sigma factor forms the holoenzyme or rather complete enzyme.

So, their molecular weight is roughly around 480,000 Dalton and can range up to 650 Dalton individual subunit and it each base pair is going up to 650 Daltons whereas the total protein is around 480,000 Dalton. This is roughly a rough estimate for us to understand how big the protein is. And, the molecular weight of proteins are usually referred in Daltons after the famous scientist Dalton.

And, the functions of individual subunits what is the function of alpha subunit? It is the assembly of the tetrameric core. So, it is an organizer of the enzyme. Beta, what it will do? It ribonucleoside triphosphate binding site; that means, beta subunit of the RNA polymerase enables the ribonucleoside triphosphate binding. And, beta dash which is related to beta but not exactly beta. Remember two alpha units we have which are identical.

So, the beta dash allows the DNA template binding region. The sigma factor allows the recognition of the region on the DNA where the transcription has to take place. Let us revisit once again alpha allow alpha subunit allow assembly of the tetrameric core. That means, all the 2 alpha beta and beta dash should come together whereas, the beta subunit allow the ribonucleoside triphosphate binding site.

And, beta dash what it will do? The DNA template binding region that means beta dash recognize, go and attach on to the templates DNA strand. And, the sigma factor tells where is it. So, we can call sigma factor as a supervisor. So, you have a cleaning staff or a construction worker is there, they are the one who is actually doing the job. But a cleaning staff or construction worker cannot know which part of the building we should be working today. So, that is governed by or directed by a supervisor.

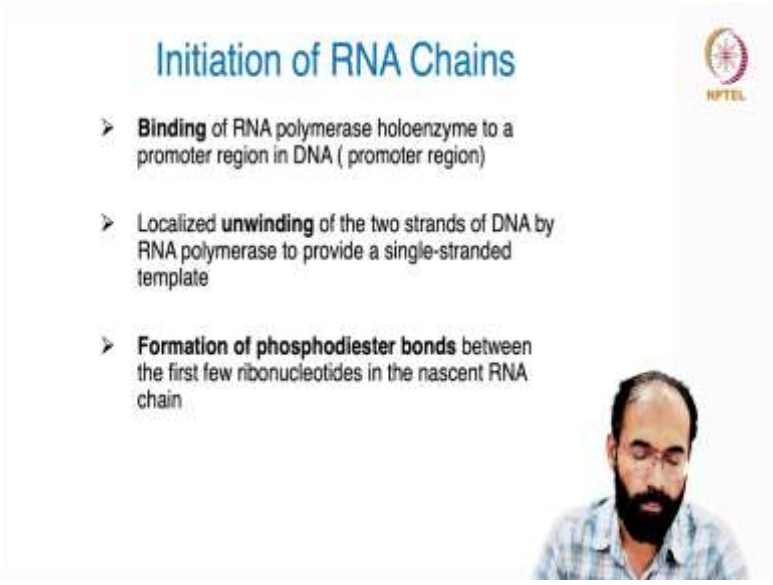
So, sigma factor is somewhat like that of a supervisor, but without supervisor function cannot happen effectively. Function will happen randomly somewhere in some way. So, sigma factor is very important in deciding where or how a transcription should initiate. So, sigma factor is a requirement for in vivo situation; in vivo means in real situation.

If a bacterial genome is present inside a bacteria, in real time situation sigma factor comes into picture and it decides how effectively a given gene is transcribed. But in vitro the RNA polymerase works equally good on both the DNA strands and sigma factor is not a necessity.

Because entire whichever region the RNA polymerase finds in the system you can always have a transcription taking place in vitro. In vitro means in laboratory condition in test tube condition where you do not need to have the organism. So, basically for a RNA polymerase to function all you need is a DNA strand and the enzyme RNA polymerase enzyme and the monomer that is ribonucleoside triphosphate.

They need that is enough and the polymerization will take place and the RNA will be formed from a given template DNA strand.

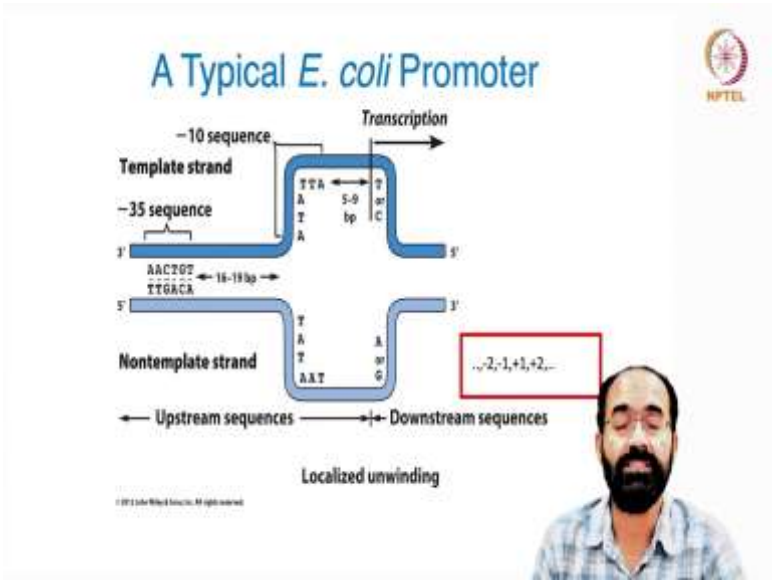
(Refer Slide Time: 08:49)



So, the initiation of RNA chains that is the binding of the RNA polymerase holoenzyme remember tetrameric core plus sigma factor is the holoenzyme to a promoter region onto the DNA. And, then you need to have localized unwinding of the two strands by RNA polymerase to provide a single stranded template to be accessed to make a copy of the template into a complementary strand which is in the RNA form.

And, then the formation of the phosphodiester bonds between the first few ribonucleotides in the nascent RNA chain that is growing up. So, the initiation of RNA chain contains these three steps. So, we saw transcription has three phase – initiation, elongation and termination. So, the initiation has got binding, unwinding and phosphodiester backbone until few bases.

(Refer Slide Time: 09:59)



A typical E. coli promoter if you are looking this slide may look little complicated, but do not worry it can be dissected out quite easily. Let us see there is an arrow given here that is the transcription and the directionality shows left to right for our convenience. So, in a cell there is no left or no right or no top or no bottom, but for our convenience on a paper we always give left to right as the transcription just in the similar to the convention of you write on a paper.

So, this is the place where the transcription is starting and this has to be a specific region associated with a regulatory region or a promoter region. And, upstream of this region around we refer to as minus 10 sequence. So, the transcription initiation, the first base that is transcribed we give a number plus 1 and the previous base immediate left side base we call it as minus 1, there is no 0.

So, unlike the graph you always say plus 1, 0 and then minus 1, no. There is no 0 here because you are counting on to a definitive base; so, plus 1 or minus 1. So, the start point is plus 1, now you go backward minus 1, minus 2, minus 3, minus 4, minus 5 like that you go up to minus 10. You have a specific sequence that reads AT AT TA. So, this is basically a minus 10 sequence which is very important in the recognition of the DNA by the holoenzyme.

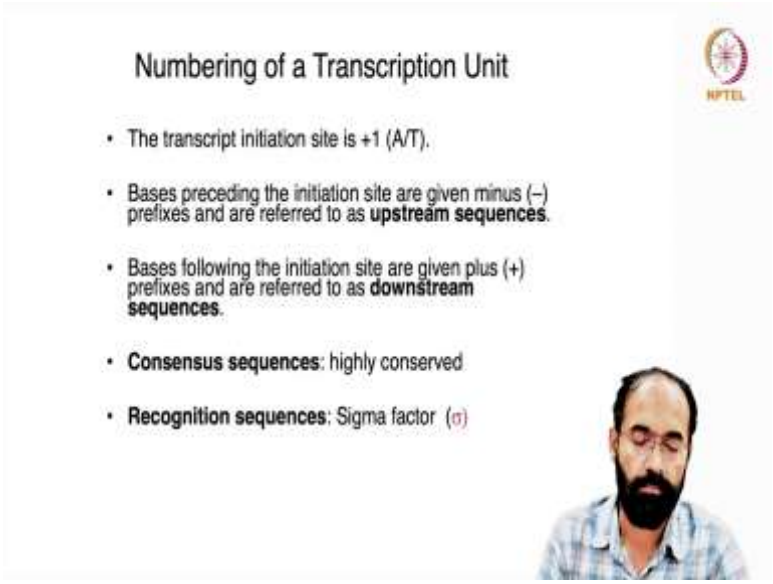
And, much upstream to that minus 10 then you have minus 11 minus 12 minus 13 minus 14 like that you are going 5 prime upstream or backwards more and more you have a

sequence AA C T GT and the opposite strand is TT GA CA. And, in between the minus 10 and minus 35 you have around 16 to 19 bases. So, this can vary from promoter to promoter or gene to gene, it can slightly vary.

And, we refer the top strand as the template strand because which is used to make a copy of the RNA and the non template strand is actually the sense strand which is what is made in the form of an RNA. So, in the right-hand side box the idea is made very clear plus 1 and minus 1; plus 2 and then minus 2. So, like that the sequences are moving downstream and upstream of the transcription initiation.

And, the localized unwinding is shown here in the form of a you know a widened double strand and this double strand is showing now like a railway track parallel for convenience. Actually, it is remaining like a double helical form for convenience we have opened it up and it is like a straightforward strand.

(Refer Slide Time: 13:06)



And, the numbering details are given for easier understanding in text. The transcript initiation site is plus 1 and invariably it has to be an A or T usually, the first base and the basis preceding the initiation site were given a minus notation that it prefixes and are referred to as upstream sequences. When you say upstream like in a river or in a journey you have an upstream journey or downstream journey.

So, same way for a river you are standing in a particular location where from the river is coming until the place where we are standing, we call it as an upstream sequence or upstream stream, upstream region and the place where you are standing where the river is going further, we call it as a downstream region. So, same way transcription initial point and before that we refer to as upstream and downward is called downstream sequence.

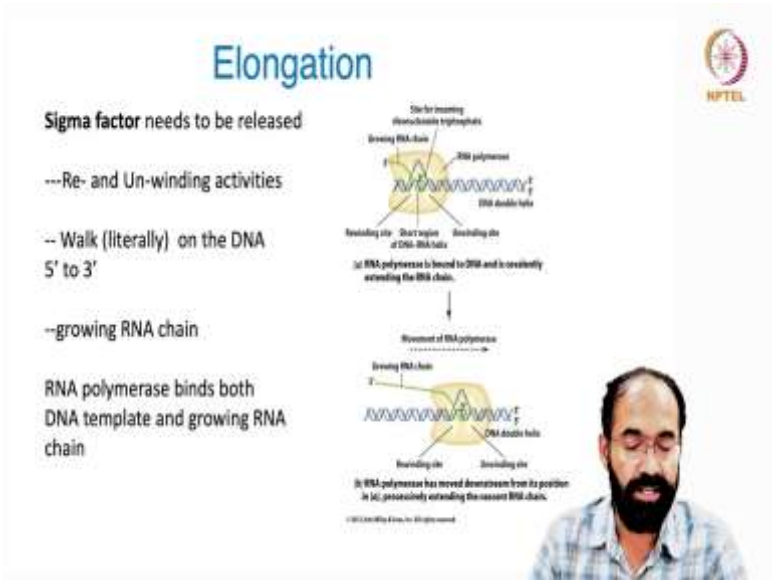
Bases following the initiation sites were always given a plus notation or plus symbol and they prefix a plus symbol and they are referred to as downstream sequence for our conventional understanding. So, keep in mind whenever someone says an upstream sequence of a gene that is more 5 prime towards the sequence and downstream means more towards the coding region of this gene.

And, then consensus sequence are those regions that are highly conserved especially the minus 10 and minus 35 sequences are highly conserved. They do not allow variation. Whereas, in between minus 35 and minus 10 variations in sequences are allowed. So, minus 10 and minus 35 regions are conserved.

Minus 10 when you say do not think that exactly the minus tenth base, when you say minus 35 do not think exactly the minus thirty fifth base that vicinity around 5, 6 base they usually remain conserved with a purpose because those regions are helpful in recognizing the DNA template or the region where the gene is present by the holoenzyme.

And, the recognition sequence predominantly are by the sigma factor. Sigma factor is the one which decides where the transcription must initiate.

(Refer Slide Time: 15:32)



Now, the step 2 of the transcription that is elongation; RNA initiation, transcription initiation and now elongation. Importantly, the sigma factor needs to be released; that means, sigma factors function was to identify the location of transcription and now it has to be released and then re and unwinding activities to continue. Initially unwinding of the DNA happened, the transcription bubble was created and now, the open region has to be closed.

So, there has to be a rewinding and then downstream region you have to continue to unwind because you want to access the template strand. So, this rewinding and unwinding will continue over and over throughout the transcription bubble when it is walking. So, this transcription bubble walk literally walk on the DNA strand in 5 prime to 3 prime direction, and as a result you end up getting a growing RNA chain.

And, the RNA polymerase binds both DNA template and growing RNA chain. So, RNA polymerase has to make sure that the growing RNA should not fall off from this complex. If it fall off it is a truncated RNA. So, it has to catch hold of both the DNA and the RNA. So, growing RNA should be held together and the DNA template also should be held together by the polymerase.

That is what you are seeing here. The transcription bubble and the RNA polymerase is in yellow colour. It continues, continues to grow and then there will be unwinding of the DNA double helix and see initially it was here and now it is here. And, it continues to

move on top of the DNA. As a result this green colour short RNA now became longer and longer and it will continue.

Some genes are pretty long. They can go in thousands of bases and it can take you know sometimes several minutes or maybe even hours to complete the transcription of one gene. Human muscular dystrophy is an example. It is a huge gene and it takes long time to complete the transcription.

(Refer Slide Time: 18:02)

And, now quickly look at the this table which talks about the recognition sequence of sigma factor. Like we saw tetrameric core remains constant the same, but the sigma factor is not the same. It can vary from gene to gene; that means, each sigma factor has the ability to allow the recognition of certain promoter regions, not all promoter regions.

Say if E. coli have got 100 genes. One sigma factor can govern the transcription of say some 10 genes randomly 10 genes and the other sigma factor can govern maybe another 15 or 20 genes, another sigma factor can govern maybe around similar number of genes. All these thing together can constitute the recognition of entire genome of an E. coli.

What is the difference? Say sigma 28, basically it is a 28 Dalton. 28 Dalton is it is molecular weight the sigma factor and sigma 32 means 32 Daltons, sigma 54 – 54 Daltons, sigma 70 – 70 Daltons. So, sigma factors are different and their molecular weight also differs.

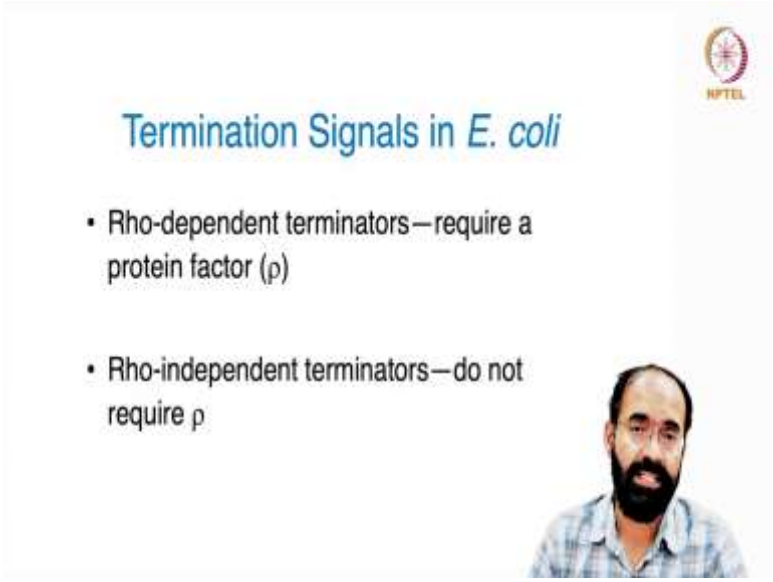
Accordingly, they have their own unique consensus sequence, minus 35 and minus 10. So, sigma 28 will prefer a minus 35 consensus of T triple A and I will not read into the details GCCGATAA. So, this is the minus 10 sequence and this sigma 28 governs the expression of genes important in flagella synthesis and chemotaxis. It do not bother about other genes at all sigma factor; whereas, tetrameric core is a must.

But sigma factor is the I told you that is the caretaker or that is the supervisor that allows the worker to go and bind where. Sigma 32 similarly have got its own unique. If sigma 32 also had the same as that of sigma 28, then sigma 28 will take care of the sigma 32 factor. That does not happen it has got a totally different sequence CTTGAA. Similarly, it has got a different minus 10 sequence.

And, they are important in heat shock genes; that means, if bacteria undergoes a high temperature region, some proteins has to be protected some proteins has to be produced in order to protect the bacterial cell. So, they are called heat shock genes. And, another sigma 54 meant for nitrogen metabolism genes. Sigma 70 meant for housekeeping genes.

Like that each sigma unit decides what sets of genes are produced whereas; once it is initiated the tetrameric core remains the same in all the RNA polymerases.

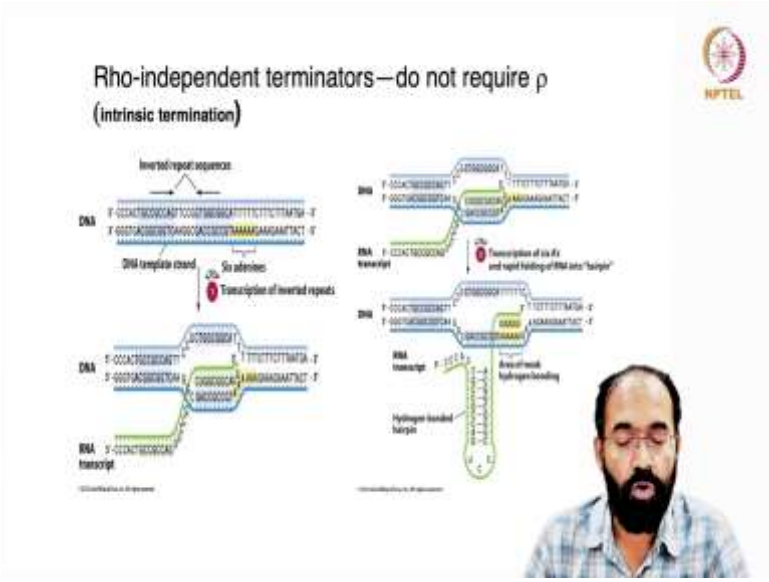
(Refer Slide Time: 21:04)



Now, we are talking about the termination signals in E. coli. Termination is the last phase in the transcription finalization or the termination is the stoppage of the

transcription and they are of two types – rho-independent and rho-dependent. So, first one is rho-dependent termination: as the name indicates it requires a new protein factor called rho factor and it is given in this notation of the letter that is rho. And, rho-independent terminators as the name indicates it do not require this protein rho for the termination to occur.

(Refer Slide Time: 21:51)



Let us see what is rho-independent terminators. And, rho-independent terminators are those which do not require the rho factor. So, let us see how the transcription is taking place this is the DNA strand. Let us note it has a interesting sequence called inverted repeat sequence.

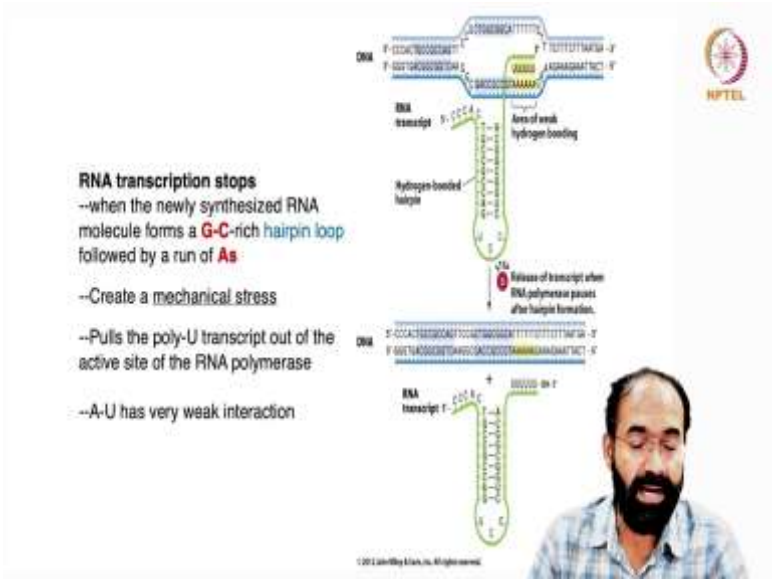
What is inverted repeat sequence? See the sequence TGCCGCCAG and same thing repeated here TGCCGCCAG from right to left, same way this is repeated here. And, this inverted repeat basically means they have the tendency to pair because they are complementary to each other.

So, when the transcription take place when this inverted repeat regions are transcribed because this happens towards the end of transcription. See this blue highlighted regions are transcribed; this green color is the RNA and now it has transcribed. What happens? When this completely transcribed it can form a loop because you can see a stem and a small loop region and it allows some steric pull or a some strong force that is acting on to the RNA polymerase and the transcription bubble.

And, subsequent to that immediate downstream you have got a A rich region. If it is an A rich region, the RNA will have a U rich region. So, there is no other base just AAAAA and complementary to that in the RNA you will have UUUUU sequence, and this sequence which is a weak area of hydrogen binding because bonding that is A double bond U, A double bond U, A double bond rich region are rather weak.

And AU pairing immediate downstream to this inverted repeat sequence allow a strong pulling action by the long RNA that is finished its transcription and this secondary structure that is formed in this RNA creates a drag a pull on to the growing RNA because of which the transcription bubble is no more able to hold on to this RNA and it falls apart. And, this is the completion of the RNA in a rho-independent manner.

(Refer Slide Time: 24:19)

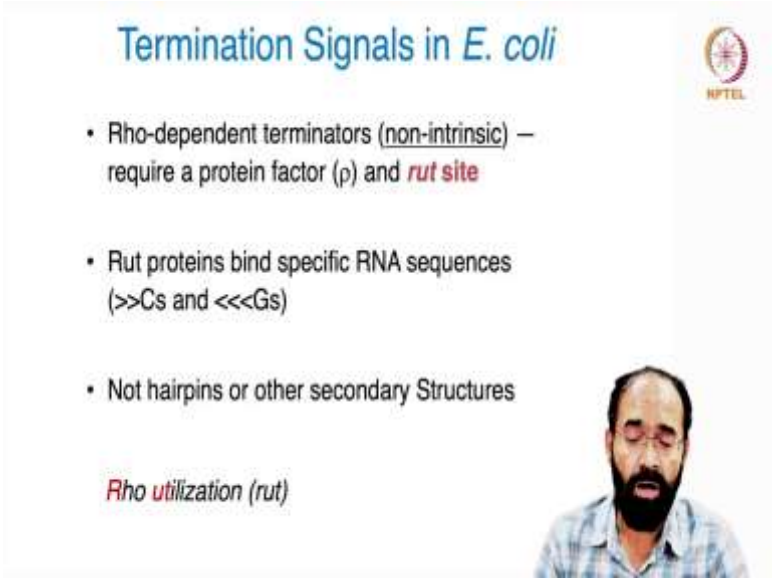


So, RNA transcription stops with the newly synthesized RNA molecule forms a G-C rich hair pin loop as you can see here and followed by a stretch of run of As. And, this create a mechanical stress and pulls the poly-U transcript because poly-U copied from the poly-A in the DNA. Poly-U transcript out of the active site of the RNA polymerase and A-U has got a weak double bond interaction with.

So, the DNA binding with the RNA is rather weak because of which you end up getting a release of the RNA in this weak binding region. So, the release of the transcript when the RNA polymerase posses after the hair pin. After the hair pin the RNA polymerase

posses and this mechanical stress creates the release of the RNA transcript. So, this terminates the RNA transcription.

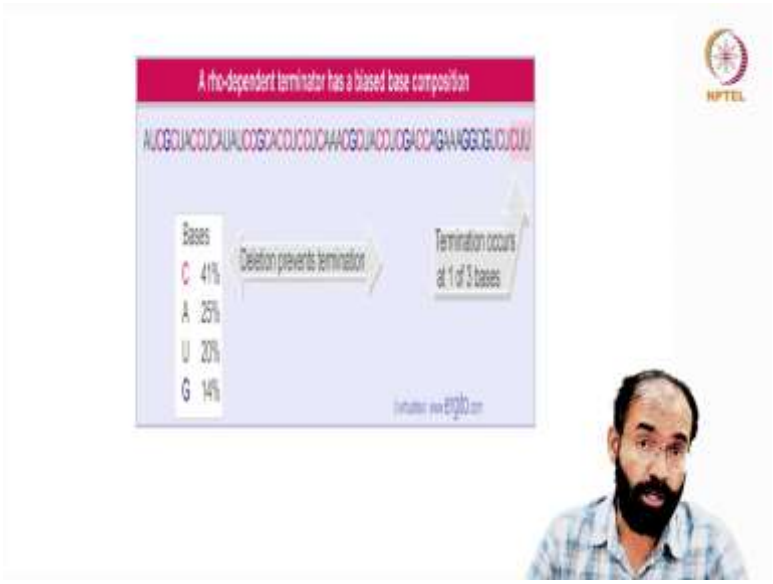
(Refer Slide Time: 25:17)



Let us see how the termination signal is made use in a rho-dependent manner. So, in rho-dependent termination that is non-intrinsic and it requires a rho utilization site or rut site and non-intrinsic means it depends on a external factor, intrinsic means it is self sufficient. Like rho-independent is a self sufficient or an intrinsic system.

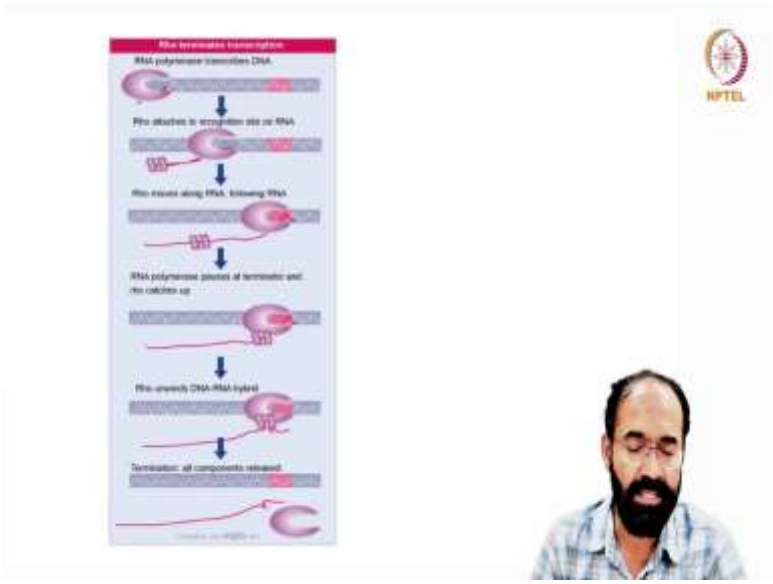
And, this rut region contains specific sequence that allow the binding of this rho protein and it is not meant for hairpins or other secondary structure, it is sequence dependent manner. So, the rho recognition sequence when present in the RNA, the rho protein will come and bind on to the are growing RNA molecule.

(Refer Slide Time: 26:11)



So, you can see here in the sequence there are long stretch of rho binding region, where the C's are 41 percent C's and 25 percent A's and 20 percent U's and 14 percent G's. So, predominantly this sequence is rich in C's and then A U and least by G's in this RNA sequence. And, this allows the rho recognition and this termination occurs roughly around 1 to 3 bases downstream of this rho utilization site or rut site.

(Refer Slide Time: 26:51)



So, you can see here this rho dependent termination whenever you have this RNA molecule that has come to the rho utilization region, the rho protein binds on to the RNA.

This again causes a mechanical stress because an RNA if a protein is bound onto it and the rho moves along the RNA and following this newly growing RNA chain. And, the RNA polymerase poses at the terminator sequence and the rho can easily catch it up on to that RNA and it binds stably.

So, the rho can start following the RNA and whenever the rut site comes it will be able to bind strongly and rho raw unwinds the DNA – RNA hybrid. And this rho protein at its extreme end it will cause a unwinding of the DNA – RNA hybrid and which will cause the release. And, that is called the termination of all components that takes place and the rho and the RNA gets separated.

So, we will continue the remaining part of the RNA biology in another lecture. I am stopping here.

Thank you.