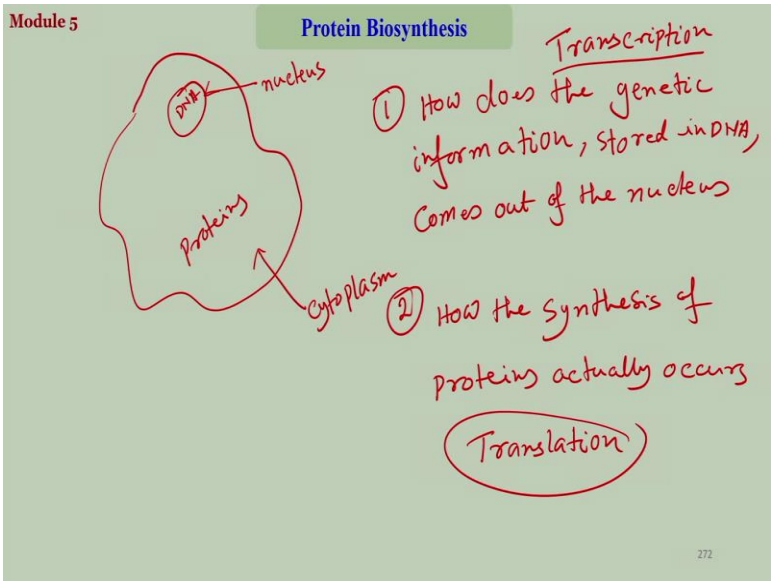
Essentials of Biomolecules: Nucleic Acids, Peptides and Carbohydrates

Prof. Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology-Guwahati

Lecture No. 20
Transcription - The Transfer of Genetic Information from DNA to mRNA

Hello everybody and welcome back so today we will talk about another very interesting and very important biological process that is biosynthesis of proteins. How proteins are synthesized in our in that living cells, so, this is going to be our module 5. So, previously we have seen how the DNA are synthesized in the biological cells through DNA amplification process. So, today we are going to see how the genetic information stored in the DNA is transferred to synthesize a predetermined specific sequence of a protein. So, the title of the module is this DNA to proteins transcription translation and genetic code.

(Refer Slide Time: 01:44)



So, let us say this is our cell and this is the nucleus, this is the nucleus of our cell where is our DNA the other part is our cytoplasm. So, now we have seen that DNA is the genetic storage unit it carries all the genetic information and it passes on the information when the cell division happens to the newly generated cells. So, DNA we call it as headquarter of the cell which controls all the cellular functions. Most of the cellular functions are actually controlled or dictated by DNA.

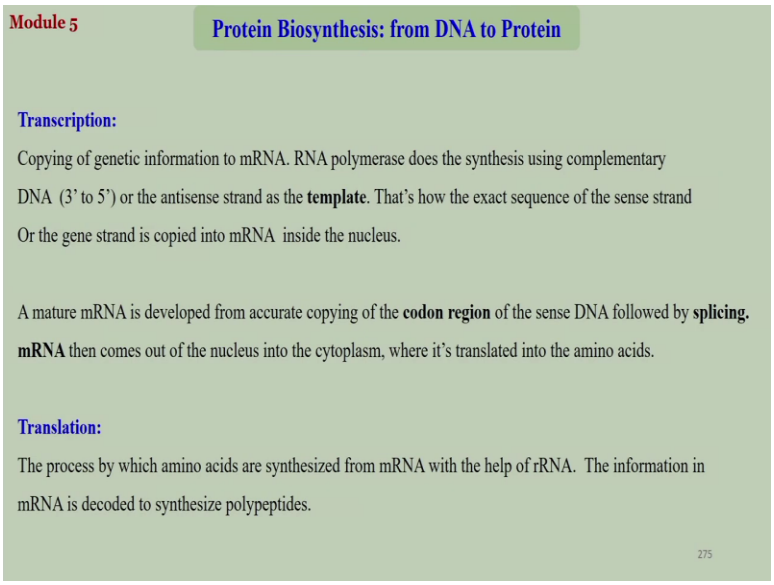
On the other hand protein is the biological macromolecule which is very, very important biological molecule or biomolecule which carries out all the reactions or that catalyzes all the reactions that happens in the biological cells. So, proteins or the enzymes are almost inevitable in any biological reaction that goes on and protein stays in the cytoplasm of the cell here are proteins here is DNA. Now the question is there must be a relation if DNA controls all the cellular programs and if protein actually does the work catalyzes the cellular processes then there must be a relation between the genetic code that is stored in the DNA and the sequence of the protein that is being synthesized.

However there is one concern the concern is DNA stays in the nucleus of the cell it never comes out and protein stays in the cytoplasm. So, how does the information that is stored in the DNA or their code comes out of the nucleus of the cell and dictates the protein synthesis that is one. Second is of course DNA has one kind of form proteins we have seen have completely different form different chemical structures. So, how does DNA Nick again or transfers its genetic code or changes its form from one to the other from the DNA to the protein code, how does it change.

So, these are the two important points that or the conscience that have to be resolved. So, first point is how does the genetic information that is stored in DNA, comes out of the nucleus that is question number one and two is how the synthesis of proteins actually happens mechanism actually occurs. So, these two are the key points that needs to be resolved. And each of these processes are actually involved with this chain of events that we will see one by one.

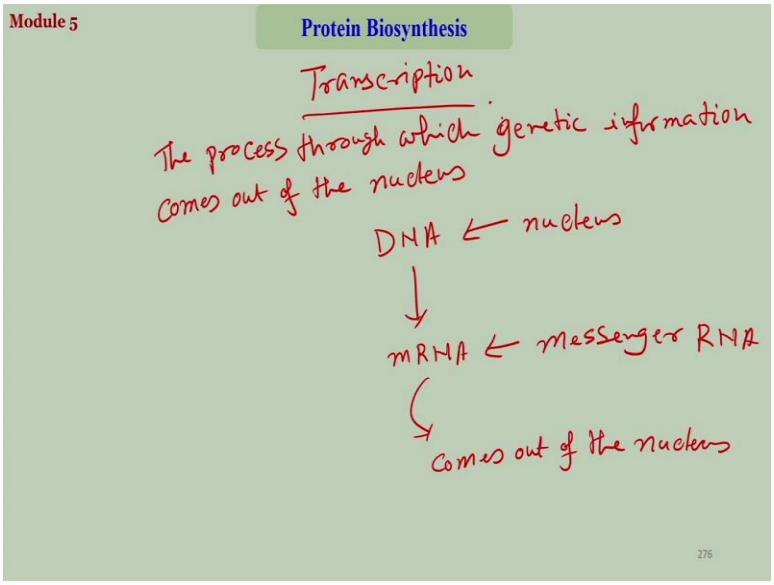
So the first process the way through which the information present in DNA comes out of the nucleus that is one process and this is known as transcription. And the second the way or the mechanism through which a protein is synthesized taking the information from DNA is known as translation. So, these are the two main processes that are associated with the biosynthesis of protein and we will discuss them one by one. And as I say again each of these processes are actually involved with a chain of events.

(Refer Slide Time: 07:56)



So, here the first step is called the transcription which means the copying of the genetic information and second stage is called the translation the process by which the amino acids are synthesized from the taking the genetic code, so, why the name transcription why the name translation that will understand as we move on.

(Refer Slide Time: 08:32)



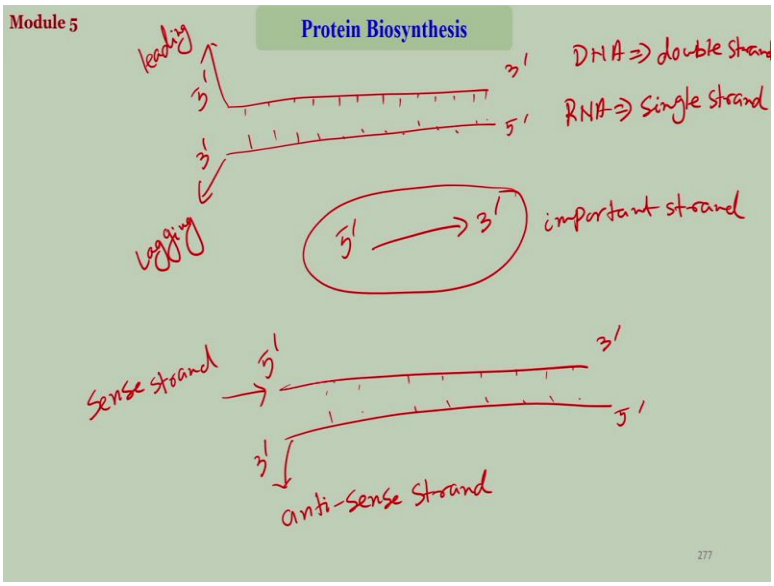
So, let us take the first process that is transcription in the process through feet in genetic information comes out of the cell nucleus. This is actually very interesting and you can see that how clever the cells are when they are doing this transcription process. So, we have the DNA in nucleus and we know that DNA is acting as a headquarter or kind of the control room of the cell

it is the brain of the cell. And therefore it cannot come out of the nucleus. So, if you think of a police headquarter or a police control room you cannot keep a control room empty.

Because in that case all communications will be destroyed, so DNA cannot come out of the nucleus protein leaves stays in the cytoplasm. So, somehow the information in DNA or the code that is present in DNA has to be transferred outside the nucleus of the cell and that is done very cleverly by synthesis of RNA. A specific kind of RNA known as mRNA m means messenger. So, DNA synthesizes one specific kind of RNA that contains all the information that the original DNA has and this RNA comes out of the nucleus.

So, in this way DNA does not have to come out by itself it sends another molecule with the information in it outside the cell and this information will further be used for the synthesis of the protein. Now how the mRNA is synthesized and how does it contain the exact information that is stored in the DNA.

(Refer Slide Time: 11:50)



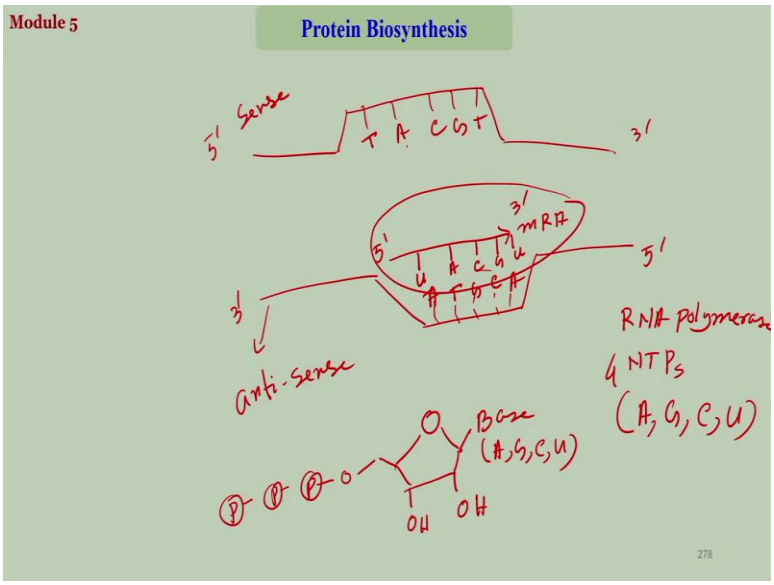
So, if you have a double-stranded DNA this is the double-stranded DNA that is in the nucleus present in the nucleus of the cell. Now you have to synthesize one RNA that will contain the genetic information and RNA is single-stranded DNA is double strand on the other hand RNA are usually single strand. So, if a single strand RNA is to be synthesized then which template it is going to use out of this double helix which strand is going to be used as the template.

And you have to remember all the time that the sequence of the RNA has to contain all the genetic code. So, remember we used to call the 5 prime to 3 prime strands as the leading strand and this we used to call as the lagging strand. And I have also maintained all the time throughout this course that whenever you want to present a sequence of a DNA then you just write only the 5 prime to 3 prime direction sequence and that is enough you do not have to write down the sequence of the other strand sequence of the lagging strand.

Because once you write this sequence the other strand is already meant but we usually write this; the prime strand 5 prime to 3 prime direction and that is a usual practice for biochemist. Apparently it is not only for our writing purpose cells also think similarly. In biological cells they also regard the 5 prime to 3 prime sequence as the most important strand. And it regards this sequence that this particular one has or will contain the genetic information. So, all the genetic information that we talk about are actually because of the sequence of this strand 5 prime to 3 prime strands.

And now onwards we will call these people have give it a different name instead of lagging strand and leading strand this is now called sense strand and this is now called anti sense strand. Now we have to synthesize an mRNA which contains all the genetic information that is present in the sense strand because biologically this is taken as the unit which stores the information. So, therefore we need to synthesize the exact sequence of this DNA into the RNA form.

(Refer Slide Time: 16:20)



How can it be done the cells are very clever. So, 5 prime to 3 prime so this is your anti-sense friend this is your sense strand, now when it opens off the double helix when it opens up to synthesize a new RNA then an RNA single strand so you have to synthesize only one strand. Then it takes the anti same strand as the template and it synthesizes our RNA here which is 5 prime to 3 prime direction taking the anti sense strand as the template not the sense strand and as instant as the template.

If it takes this as the template let us say this is A T G C A so this has to be in originally T A C G and T. Now it takes the enters in strand as the template and the new mRNA strand is started to be synthesized with the help of course the enzymes. This kind of replication process but it is not DNA replication in this case you are using synthesizing RNA so you need RNA polymerase and you need the nucleobases not the GNTP's but the normal entities with the ribose all four in NTP's which are basically ribose-adenine, ribose-guanine, ribose-cytosine and not thymine but eurosine right.

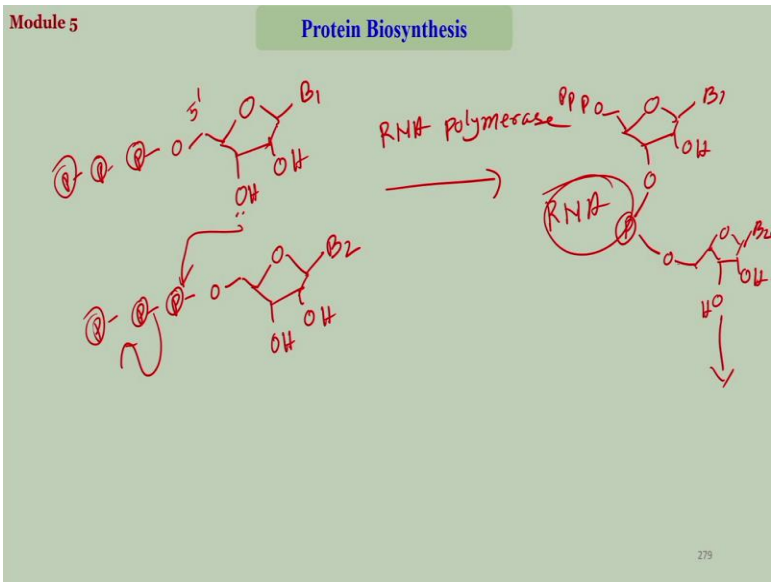
The general structure being here is the base A G C or U here you have a hydroxyl group. Now here you also have a hydroxyl group and as usual you have phosphate, phosphate and phosphate. So, this is the NTP and RNA polymerase would be taken instead of the DNA polymerase. It will use this as the template and will synthesize the complementary strand RNA strand. So, what would be the sequence of the RNA strand here the sequence has to be opposite to A is to T where

RNA case it will be U, this has to be against T it should be A against G it should be C against C it should be G and here against A it should be T.

But for RNA it is U, U A C G U 5 prime to 3 prime sequence. Now let us see whether this mRNA contains the genetic code that is present in the prime spread sense strand U C A A C C G G U S D so by using of course you can understand by using the anti sense strand as the template you can synthesize an exact copy of the same strand in the RNA format. And that is why this is this process is called transcription. So, when you talk about transcription when you get your grid cuts or your MUC sheets from colleges or universities or other or other institutes like IITs NIT's wherever you are you call them transcripts which means these are basically the copies authorized copies.

Of course they are basically the copies of the master one the master copy or the original copy escaped with the headquarter of the university or the Institute so they are called transcripts. But they are the exact replica of the original one. The same thing here the code or the sequence of the newly synthesized mRNA is exactly the same as that of the same strand. So, it is basically an transcript of the sense strand and that is why this whole process is known as transcription.

(Refer Slide Time: 22:11)

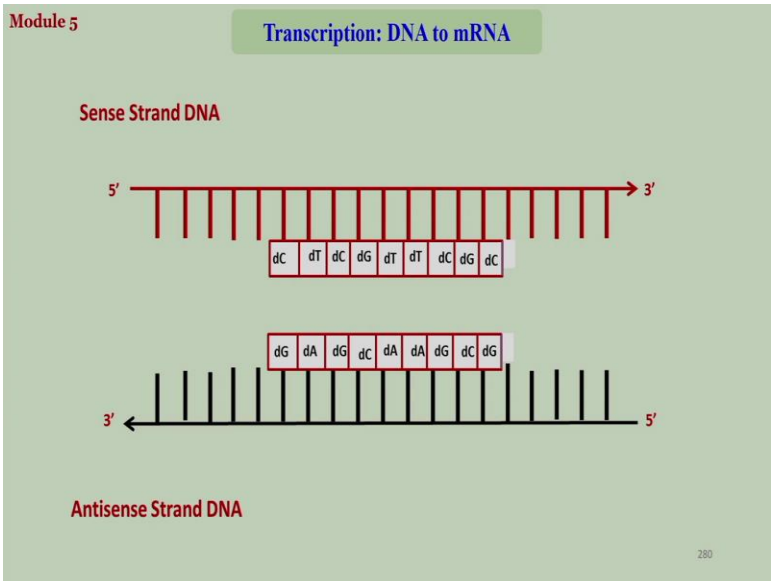


So, now let us look at the reaction when you are synthesizing the mRNA what would be the reaction that would be going on quickly let us say this is B 1 or in a 3 prime hydroxyl free O

phosphate-phosphate-phosphate this is the 5 prime end. And then you have B 2 OH OH 5 prime phosphate-phosphate-phosphate. So, RNA polymerase and these are the dint apiece the rival sugars RNA polymerase will catalyze this reaction the reaction is as usual a diphosphate would be out.

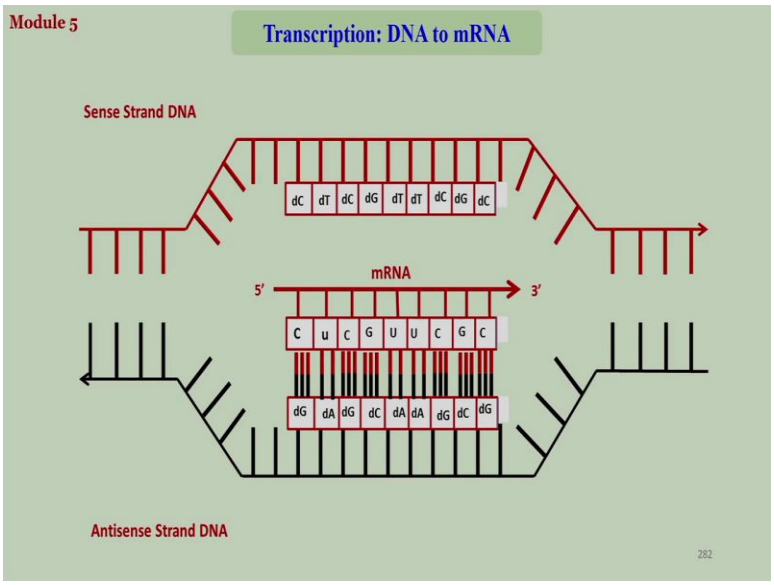
O P P P here O phosphate O CH2 basically here be to here there is OH here there is OH and here would be the OH and this will go on like this so you will have your RNA synthesized there.

(Refer Slide Time: 24:11)



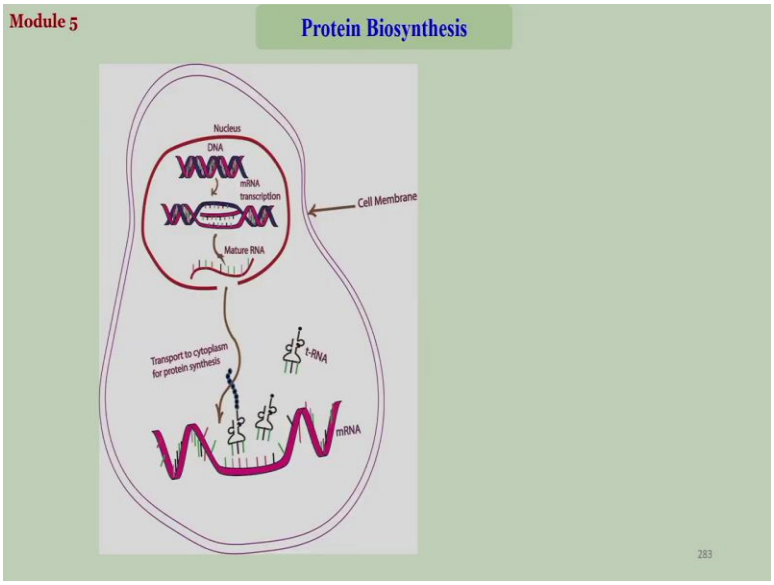
So, it is given here sense strand 5 prime to 3 prime anti sense strand 3 prime to 5 prime and let us take a sequence here I have only written a portion of the sequence not the whole one so CT CG TT CG C however this is of course the opposite sequence this is the double helix DNA your gene that contains the information.

(Refer Slide Time: 24:43)



Now you need to synthesize the mRNA and that is done where this anti sense strand is used as the template and the new our mRNA is synthesized using this as the template. Then instead of D G it comes the C ribose cytosine against that adenine takes a euracil GC CG AU AU GC CG GC now if you compare this sequence is of course the exact of the sense strand. So, you that is how you get the transcripted version of the sense strand in the RNA format.

(Refer Slide Time: 25:31)

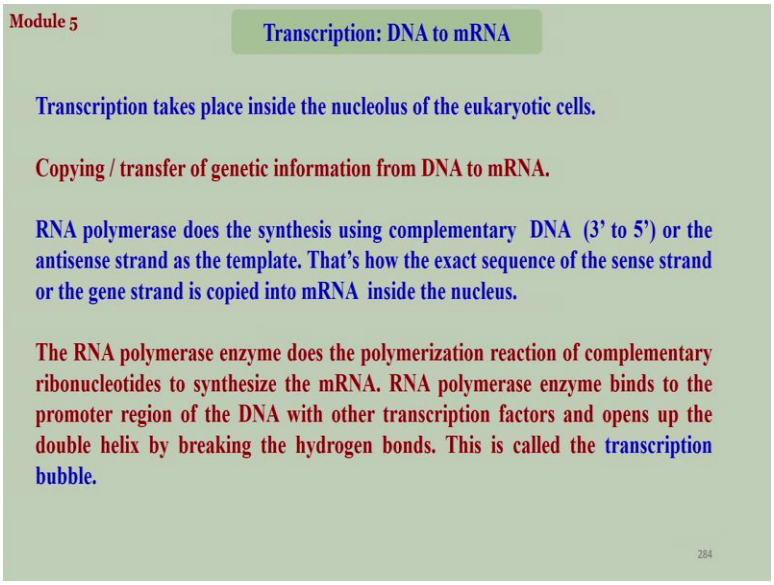


Here is the process how it comes out of the cell this is the nucleus you have the DNA here and then DNA this double helix DNA is used to synthesize a mRNA which is this. And then when the mRNA is synthesized it gets matured I will show you what the process is it includes a couple of more events but this is the simplified version that once this the mRNA is synthesized that

contains the exact copy of the prime strand or the sense strand and it gets matured this matured RNA that comes out of the nucleus of the cell.

And this process until it comes out of the cell nucleus is called the transcription. Now once it comes out of the cell this mRNA is used or the sequence of the mRNA is used to synthesize the protein that is called the translation.

(Refer Slide Time: 26:42)

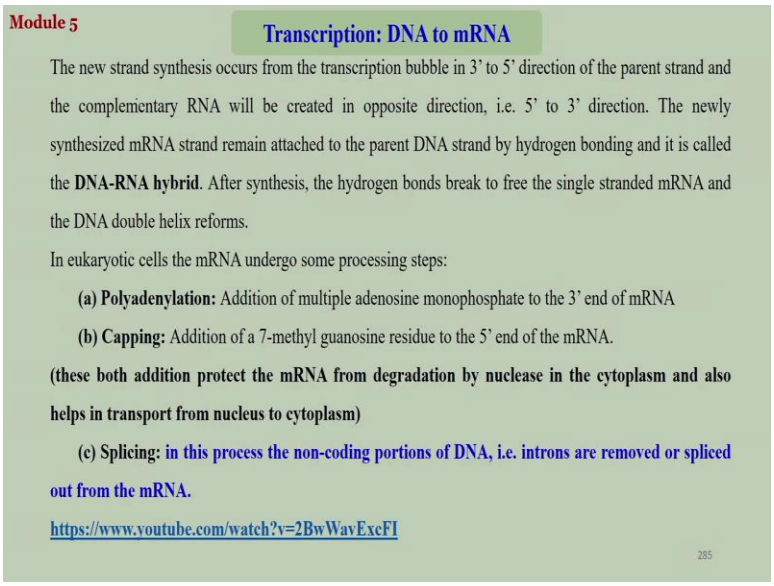


So, that process of transcription takes place inside the nucleolus of the eukaryotic cells remember I have been or whenever I am talking about the protein biosynthesis I am primarily talking about the eukaryotic cells the higher organisms copying or transfer of genetic information from DNA to mRNA. RNA polymerase does the synthesis as we have seen using the complimentary DNA that is the 3 prime to 5 prime sequence or the anti sense or the antigen strand as the template.

That is how the exact sequence of the sense strand or the genes strand is copied into mRNA inside the nucleus. Now how the RNA polymerase works RNA polymerase binds to the promoter region of the DNA so it out of the long DNA it has certain sequences of the binding sites for the RNA polymerase that is called the promoter region of the DNA with the other transcription factors are also included that I am not going to talk about and open up the double helix by breaking the hydrogen bonds.

So that is the driving force this is called the transcription bubble. So, this one where you have this part out of the whole DNA you have opened a particular part of the double helix of double helix structures so and this is called the transcription bubble.

(Refer Slide Time: 28:22)



The new strand synthesis occurs from the transcription bubble in the 3 prime to 5 prime direction of the parent strand and the complementary RNA will be created in opposite direction of course that is from that 5 prime to 3 prime directions that is what we have seen. The newly synthesized mRNA strand remain attached to the parent DNA strand by hydrogen bonding and it is called the DNA-RNA hybrid. So, initially this mRNA with the DNA when you synthesize this, this is basically a double strand with one RNA and one DNA.

So this is called the RNA-DNA hybrid this is this. After the synthesis the hydrogen bonds break to free the single-stranded mRNA and the DNA double helix reforms. So, once RNA get out of the helix you get back your double helix and the gene retains its property or which means the ability to replicate again in eukaryotic cells the mRNA undergo some processing steps. So, once the RNA comes out of the nucleus that is our ORNA that actually that contains all the genetic code that or that is the copy of the full-length DNA.

But full-length DNA is not required to synthesize a protein a whole DNA a major part of our gene or the major part in of our genome is actually not transferred into synthesis of the protein

only a short portion are actually important that is called the coding region or the code on that coding region of the DNA is actually important to synthesize the specific unit of protein. When the mRNA is synthesized initially then the whole of the apparent DNA is copied to the mRNA and then mRNA needs to be edited to remove all the unwanted part that is the un-coded part so this is how it is done.

In eukaryotic cells the mRNA undergo some processing steps after the synthesis. First one is very important polyadenylation at the 3 prime end of the mRNA a number of adenosine moieties would be added there this is to cap it actually. We will later on see that they are meant for initial at the one end it will be initiation and the other end it will be stop code-on. So, polyadenylation means the addition of multiple number of adenosines at the 3 prime end of the mRNA.

And then capping so that the mRNA remains intact a does not father undergo replication or does not father undergo extension capping his addition of a 7 methyl guanosine, here we see again a modified version of the guanidine 7 methyl guanidine we have already seen when you are when we have talked about mismatch repair they are the DNA methylation sites were the recognition sites and DNA methylation sites means the methyl guanosine methyl adenine and methyl cytosine.

So 7 methyl guanine that is the modified version of the guanine that is added to the 5 prime end of the mRNA once that is there then the RNA is capped it can go no further because this is not regarded as a base pair unit. So, this both additions protect the mRNA from degradation by nucleus in the cytoplasm and also helps the transport from nucleus to the cytoplasm. So, that this is about the maturation of the mRNA.

So once is the mRNA get matured then comes the splicing, splicing here in this process the non-coding portions of DNA that is introns are removed are spliced out from the mRNA. So, non-coding part is actually removed the non-coding part is known as the introns. The introns are removed or spliced out from the mRNA and then the mRNA comes out of the nucleus and that is how the process of transcription occurs. So, this is the process of transcription or the copying of the information in DNA to the mRNA.

One more thing is why this is called the messenger RNA. So, in olden days when there was no postal systems forget about Mobile's now. The kings or the or the very important people they used to send secret information's or news to the other part of the world through specific persons they are known as messengers. So, those were those people or those persons used to be very important or they was to very close to the king and they were confident with the news or the with the information from the king.

And they take those information's to the other part. So, they were usually regarded as kind of the loyal servants of the not the servant sometimes friends even very loyal to the king and they used to speak on behalf of the king. So, here also the RNA that we are synthesizing contains the exact information it carries the information that is present in the headquarter that is present in the nucleus that is present in the DNA.

And then that this mRNA which is synthesized new is bringing that information outside the nucleus to the cytoplasm. So, that is where they are called the messenger RNA. So, in the next phase in the process of translation we will see how this coding information that is present now in mRNA is used to synthesize proteins, thank you.