Essentials of Biomolecules : Nucleic Acids, Peptides and Carbohydrates

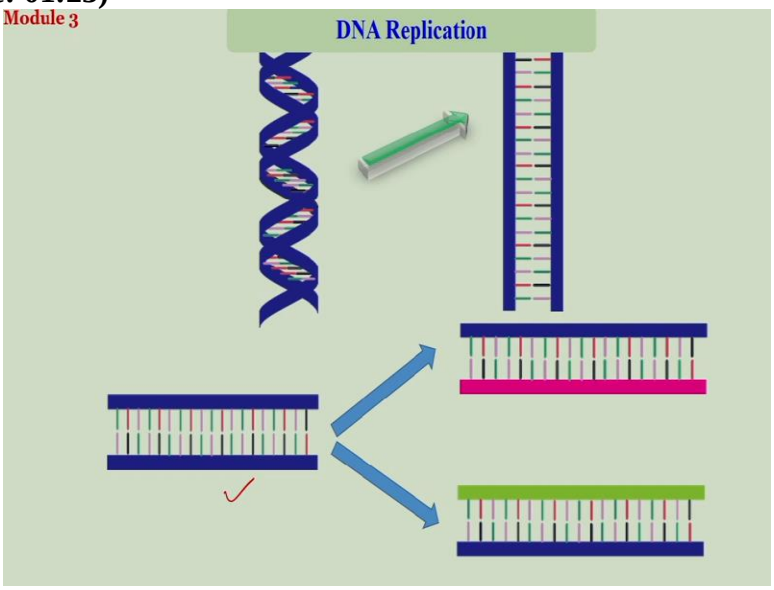
Dr.Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology Guwahati

Lecture-10
DNA Replication, Polymerases, DNA Sequencing and PCR

Hello, everybody, and welcome back. So today, I will start the module 3, which is about DNA replication and DNA followed by then we will cover DNA sequencing and the PCR, which is known as polymerase chain reactions. So after some organic synthesis we are coming back. Let us come back to some biology again. So I will start with DNA replication first, and then one by one we will move on to the other topics such as the how to sequence a DNA.

And then how can we amplify DNA sample which we have isolated from on an organism or a diseased cell? How can you amplify that DNA in the laboratory using a technique known as polymerase chain reaction?

(Refer Slide Time: 01:23)



So, first is the DNA replication. So, DNA replication is a process that happens in our cells. In eukaryotic cells, the process is about from 1 DNA. How is the DNA copied to the next generations, which means how the DNA is amplified to the newly synthesized cells? So, if this is a DNA, we usually denote a double stranded DNA as this. This is double stranded DNA with the

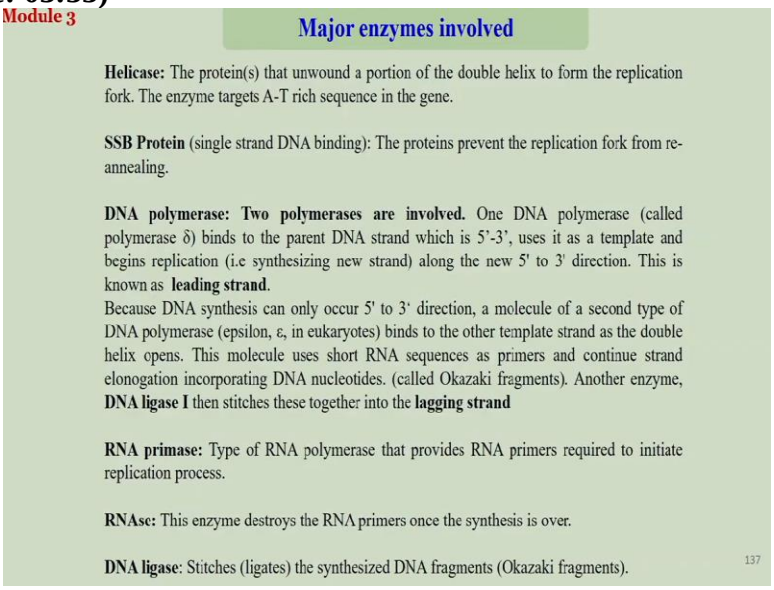
base pairings here, this is the notations we are using now, so, if you have that double stranded DNA.

How that is divided into the 2 newly formed DNA, 1 DNA is multiplied into 2 for the new generations of the cells. So, essentially what happens is that each of this strand in this in the parent DNA is used as a template and then our new strand is synthesized taking the template as the complimentary of it. So, this new strand would be synthesized, taking the parent strand as the template. Again, the other template, other strand would be taken as a template to synthesize and other strand.

So, ultimately you get to new DNA exactly looking like the same as the parent DNA. So, this is what the DNA replication is and this happens in our cells as cell division occurs eukaryotic cells. So I will restrict myself to show the different steps that are involved for DNA replications. In eukaryotic cells, this is very complicated process. Lots of enzymes are involved, but beautiful biology and very nice chemistry also.

So as we move on, we will see many Nobel Prizes have been given have been awarded for this to understand the whole of the application process. Nowadays, it is quite well understood. And as we see that I will show you.

(Refer Slide Time: 03:33)



At the beginning, the enzymes that are involved in whole of the replication process. So helicase, as I mentioned, that it is a combination of the functions of many different enzymes that acts for various purpose that serves various purposes, actually, and I will first show you the whole of the process, and then I will take each step and then we will discuss them. So this is one enzyme called helicase.

Helicase is a protein that actually unfolds our DNA. And then you will need the SSP protein. I will come back again later SSP protein is called the single stranded binding protein. Then comes the most important enzyme for this DNA replication is the DNA polymerase. DNA polymerase was the key invention to study the whole of the replication process. And as I am saying that a few number of Nobel Prizes have been given for these to study the whole process.

Because, of course, you can understand how complicated the and the biology might be here. The father figure in this field is Arthur Kornberg. Kornberg who has received Nobel Prize in I guess 1959 for his discovery, or for his isolation of the enzyme DNA polymerase, and this is the key enzyme that is involved for the all of the application process. All the other enzymes will be acting as a supporting enzyme.

So Arthur Kernberg, his son, Roger Kornberg also later on, we will see also He also received a Nobel Prize for his study of the synthesis biosynthesis of protein using the DNA as the genetic code that we will see when we talk about DNA transcription and translation. And then his another son Thomas he is also a famous scientist who has isolated a different kind of polymerase. That is also a DNA polymerase but DNA polymerase 2 we call it so pretty successful scientific family actually.

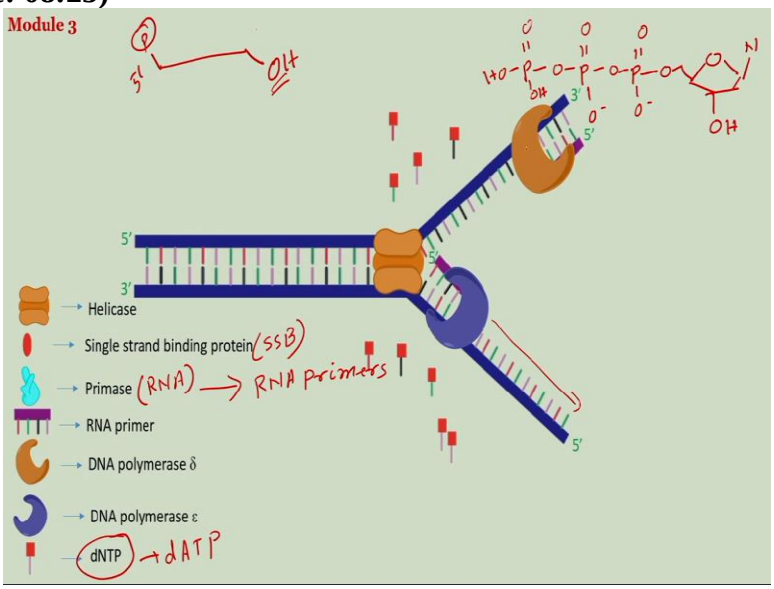
So, here DNA polymerase is the enzyme that actually helps synthesizing the complimentary strand and there will be 2 different DNA polymerases that will be needing will come back again later. And then you need another enzyme called RNA primase. RNA primase is the protein that synthesizes RNA primers, primers are short sequences of DNA. The last lecture I had shown when you have synthesized the DNA oligonucleotides I have mentioned that the DNA oligonucleotides are usually being used as primers.

Which means short length of the DNA oligonucleotides they can act as primers. The term primer will this will discuss in detail. So, in this case, this is a RNA primer which is involved in our cellular system, we do not have short sequences of DNA in our body. DNA means for our body or the eukaryotic of systems means large genes, the very long chain of DNA, or we call them gene or chromosome. So primers, which mean the shorter version of the DNA, they do not exist in our body.

But what we do have are the shorter sections of the RNA. Those are called RNA primers. And those RNA primers are synthesized by this enzyme RNA primase and then will lead to cleave those RNA primers later on. And that is why we need this enzyme RNAase. RNAase is the enzyme that destroys the RNA, or that destroys the RNA primers. And finally, we will see that we need to stitch the fragment together that or we need to ligate things together that is done by the enzyme DNA ligase.

So these are the key proteins that are involved in the whole of the replication process. So, as I am saying it is a very complicated biological process, but we will make it simple. We will discuss only the key steps and the key proteins that are involved the important enzymes that are involved in the process.

(Refer Slide Time: 08:25)



So, let us look at the animation first. So, this is your double stranded DNA, the parent or the double stranded gene actually. So the first thing that we need, if you want to synthesize a complimentary strand here or a complimentary strand there, what is the first thing that is needed, that you need to unwind the DNA you need to break the double helix into 2 single stranded DNA then only the new strand can be synthesized. And that is done by this enzyme helicase.

Helicase what it does, it opens up or it unwinds certain portion of the DNA or the gene, because the gene is so long, you cannot really open or you cannot really make the double strand fully into the single strands, they are too large. So, what it does, it opens that the DNA in a short sequence, at least as a fraction of the DNA double helix is broken. So, that enzyme helicase will come bind to the double stranded DNA.

And then it will help open up partially the double strand into the 2 single strands and it looks like fork this is known as a replication fork. This is the start of it, replication fork. And now, what may happen is that once they open up, they might again come closer to form the hybridized strand, because DNA hybridization is quite a strong binding property. So, in order that these 2 newly formed single strands do not come again to the hybridized strand, another set of enzyme comes and stabilize the single strands.

These are known as single stranded binding protein or In short, they are called SSB protein. SSB proteins are the proteins which bind to the single strands. And generally, enzyme DNA binding is quiet wrong. So it is a stabilizing factor. So once the proteins comes and bind to the single strands, they do not form the double strand again, because these bindings are stronger compared to the double helix formation. So that will ensure you that the partially opened single strand will remain as single strand.

So, that is how the single stranded binding protein will stabilize the replication fork. And now comes in order to start now you are going to start the synthesis or you are ready to start the synthesis of the new strands complimentary strand, but you need a base to start that and that is where the primase is coming. As I have said, this primase is actually the RNA primase what they

do they synthesize RNA primers. I will show later I will discuss later on also, the synthesize RNA primers and RNA primers will give you the base of it.

The word primer actually means it is almost the same meaning as you understand the primer of a when you coat your wall. So, before coloring your house or before coloring your wall, we usually make a simple coat which is known as primer that is to ensure that the next coat will add here to the surface of the wall strongly. So the primer this primer means the same thing. It will give you a base on which you can start your synthesis next phase easily.

So, RNA primers will develop RNA and those RNA, the short sequences of RNA, they will actually hybridize they are complimentary. So, these RNA primer is complimentary to the this strand and this RNA primer is complimentary to the other strand. So, they are short segments, they will be hybridized to the single strand temporarily and then you have the base. Now, you can go ahead with the synthesis, I will discuss why this these are needed.

So, RNA primers are generated and they are hybridized 2 different strands and then come the polymerases there are 2 strands. So, you need 2 different kinds of polymerases so, this is a polymerases the key enzyme which actually does the replication process, so, polymerase will come and bind in this region, where you are in a primer was bound. Similarly, this polymerase will come and bind with the double strand and then what do you need for the synthesis of the complimentary strand; you need the individual basis or individual nucleotides.

These are the individual nucleotides, they are denoted as dNTP. dNTP means, if it is AT D ATP, then you call it deoxyribose adenosine triphosphates. So, this is in means the nuclear base it can be any AT GC. This is the 3 prime hydroxyl. And this is your phosphate. OH or O minus. So it is a triphosphates deoxyribose nucleic acid triphosphates basically, and this is the unit these are the monomers that are already present in our body in the form of triphosphates.

So in our body adenine, thymine, guanine and cytosine all these 4 nucleobases are usually present in this form the sugar form and triphosphate sugar form and those are this as a general talk. We call them NTP nucleotide triphosphates D means deoxyribose because we are

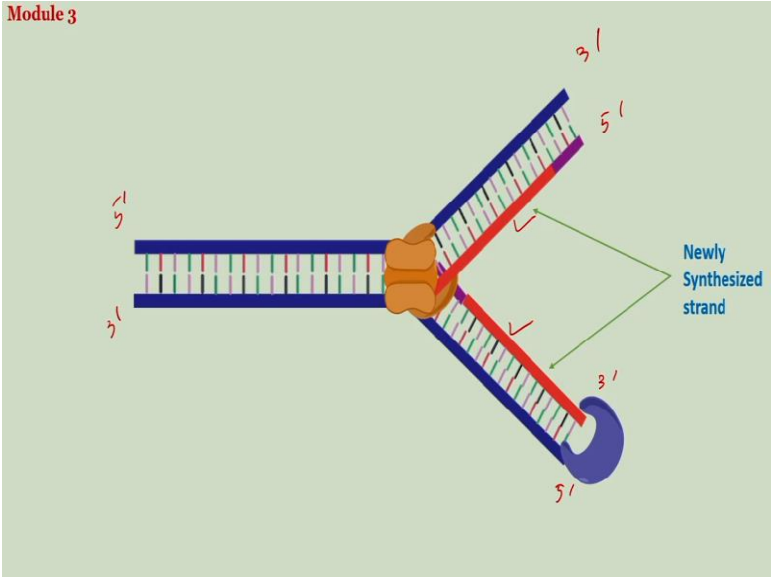
synthesizing the DNA. So, those will are floating around in your cell in the sweet table one will be picked up by the polymerase and they will be joined here one by one.

Now, here is the tree. So, this strand is 5 prime to 3 prime strands and the other strand is 3 prime to 5 prime. So, the RNA primer that was there in this strand would be the complimentary of it is means 5 prime to 3 prime directions. Similarly, so here why this is not coming here, because here would be 3 prime and your DNA or RNA whatever it is, they are always in the 5 prime to 3 prime directions and I have mentioned if you remember that whenever you have a DNA or whenever you have an RNA.

This is always if this is a strand, then your 5 prime ends has a phosphate always. This is 5 prime and 3 prime and has a free hydroxyl. Here you can see, if you make a DNA and move on this way, then the starting one will always return the triphosphate. And as you move on, there are very last one will have a hydroxyl free at the 3 prime and that is what is this and so, your functional growth that reacts is this. Therefore, here the 5 prime to 3 prime it has to be there.

So, if this is the complimentary 5 prime, that has to be the 3 prime and if you have the 3 prime and then the elongation can only occur in this direction, not to this direction, if you have the prime on here, then you can make the elongation in this direction, because here you have the 3 prime hydroxyl free and that is a reactive species. So, a suitable according to the template, the suitable nuclear basis will be picked up by the polymerase and will be stitched together.

(Refer Slide Time: 18:25)

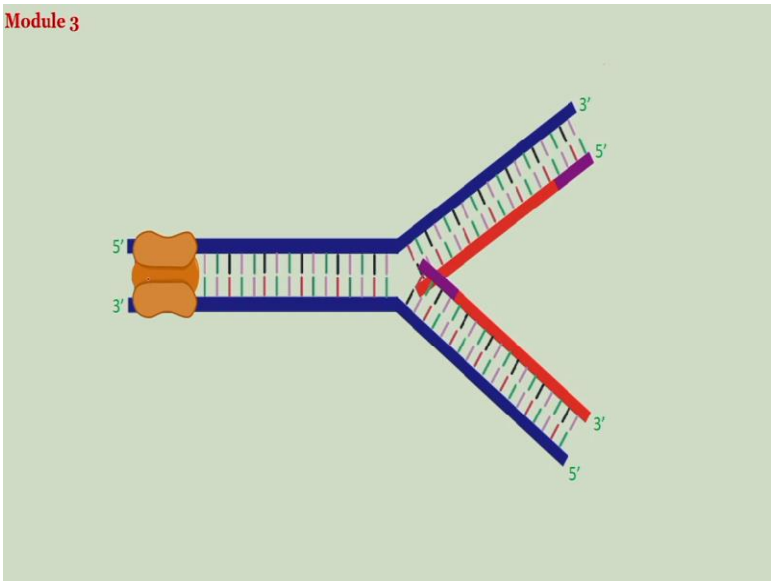


So, polymerase will help synthesize the new strands. This is 5 prime, 3 prime. So, this is 5 prime and this is 3 prime, 5 prime. So, this is your 3 prime end. So, here you see the different. For this strand the synthesis happened from the 5 prime to 3 prime directions and for the other strand, the synthesis happens in the opposite direction. And it will finish here; it cannot go any longer because your DNA is complete.

And here we will see how the 2 strands actually behave differently. So when you write our DNA strand, we always say that DNA is double stranded DNA, and they are complementary to each other. So it should be that the both strands will have that similar kind of functionalities, or seen similar behavior in biological systems. And this would be kind of behaving as twins. It is not really so for the process of replication, the 2 strands, 5 prime to 3 prime and the 3 prime to 5 prime will behave very differently.

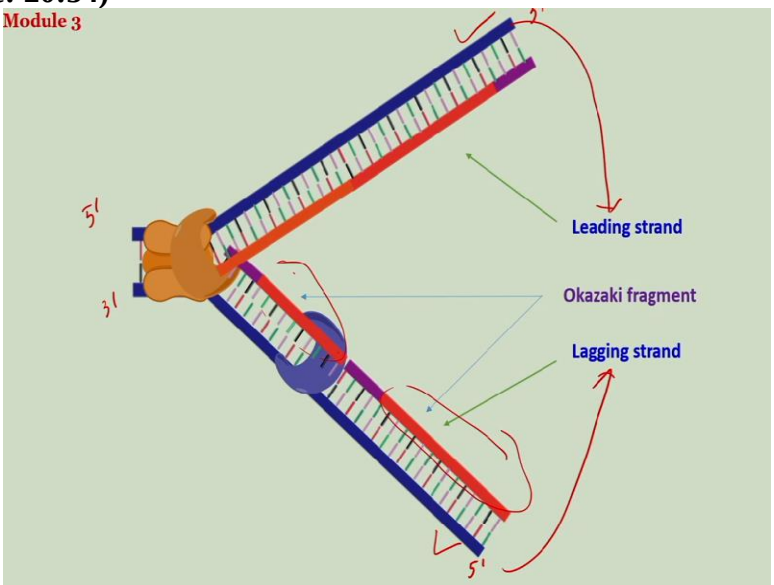
As we will see, you can already see here, you need a primer for this direction. On the other hand, for this strand is going on the other direction. So, this and these are now the newly synthesized strands, DNA strands.

(Refer Slide Time: 20:09)



3 prime, 5 prime 5 prime 3 prime. Now, this so far and new double helix is synthesized by the set of the enzymes. Now, what next? This is only a partial replication; you have to synthesize the complementary strands for the whole rest of the DNA or the rest of the gene. How that will happened, which means, the helicase has to unwind another portion of the double helix and it will do that so, helix moves there and it tries to open the other portion.

(Refer Slide Time: 20:54)



The same way as happened before. So, you now have the new replication fork with the new opening of the double helix. So you have now the single stranded DNA to some extent again. And now you can see the difference between the 2 strands so for this strand, how the synthesis will happen. This is the 5 prime, this is so this is 3 prime with the free hydroxyl here, this is the 3

prime with the free hydroxyl this is the 5 prime of the new strand. If the synthesis has to go continuously, then it can only move this direction and there is no point of doing that.

So, if the new synthesis has to happen in this strand, you have to start again from the beginning with the new set of primer, the new primer has to be there, then only you can have a 3 prime hydroxyl free here and that can be continued on the other hand, this strand, you have the hydroxyl free hydroxyl in this direction and therefore, it can moves easily in the same direction as was moving.

So, it can be for this strand the synthesis of the complimentary DNA would be a continuous process, it will not stop anywhere. With this starting with the single RNA primer, it will move on move on move on until the end of the gene but that will not be show for this strand. Because for this strand every time it unfolds, you have to start again from the beginning. So, that is the very difference between the 2 strands and we call one of them this trend we call as leading strand and this trend we call as lagging strand.

Lagging strand means it really lacks behind in synthesizing the complementary strand and that takes a lot of other factors, other some other enzymes has to get involved for this. And that is the point why you need 2 different kinds of polymerases because for these strands, every time a new polymerase has to come bind in a fraction in this case, the same polymerase can shift and can move along. So, 2 different types of polymerizes will be needing here.

So, again, once the replication fork is created, you need to stabilize them, and that is done again by the single stranded binding protein. For this strand, you need the RNA primer again. Before this trend you do not need so polymerase will come and bind here straight because it already has the base. In this case, you have to create the base again the primer again. So, RNA primer has come in and the second polymerase has to go and bound there.

Now, the individual dNTPs would be picked up from the cellular system, which are floating around and appropriately they will be taken up by the polymerase and the complimentary strand will be synthesized. Now, for this strand as I am saying it is a continuous way without a break.

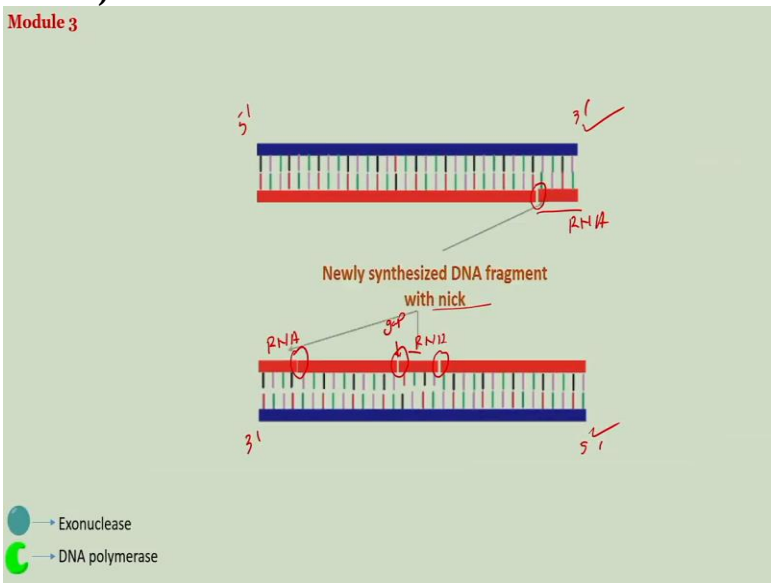
But for this trend, you have started here you have synthesize the newly formed strand here and there will be a gap because that was a different sections.

So, it can synthesize up to the base, which was present opposite direction. So, you see, this is called the leading strand actually, this strand is called the leading strand, and this the; parent strand is called the leading strand. This parent strand is known as the lacking strand so, this one was 5 prime to 3 prime. 5 prime to 3 prime strands is the leading strand 3 prime to 5 prime is the lagging strands.

So, that is why when that is also a reason why we usually write, when you write a DNA sequence, if you look at the previous lectures, I have always written the DNA in the 5 prime to 3 prime directions, because that is the main strand we usually use for the notations and you see, so, for the leading strand the continuous synthesis for the lagging strand your synthesis is partial. So, this is the new DNA that has been synthesized.

This is the new DNA that has been synthesized. So, discrete synthesis and they create their synthesized in fragments. These fragments are known as Okazaki fragments, all of the steps we will discuss in detail one by one.

(Refer Slide Time: 26:09)



So now that is how you have synthesized the 2 different DNA taking the template of the parent's strand these 2 are the templates. So, 5 prime to 3 prime and this is 3 prime to 5 prime, what is

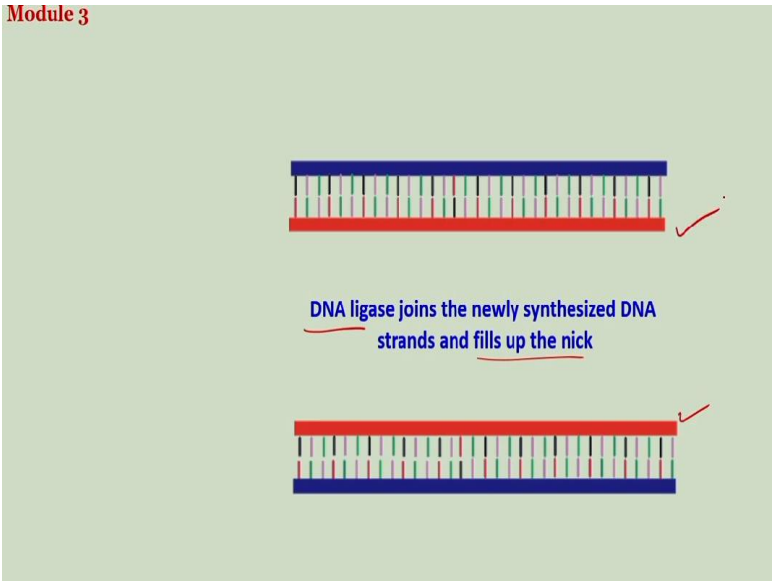
left? You have RNA left here, you have RNA here. You have RNA here and there was a gap. There was a gap here in this spot. This is a gap there was a gap here is so do not be touched. So now the RNA fragment should not exist in your newly synthesize DNA because you are creating a new gene and it is all about DNA.

So, RNA should not exist there. So, you have to replace that RNA by DNA. So, the first thing is that you have to destroy the RNA. So, all the RNA will be destroyed. So, these are your newly synthesize DNA molecules with RNA impurity. You to destroy the RNA and that is done by the enzyme. RNA is the kind of exonuclease this RNAase. RNAase is the enzyme which is in the family of exonuclease and that will destroy the RNA. RNA are destroyed.

That creates gap, creates gaps here, creates gap there were RNA was there. Now, already formed DNA strands, they will act as the template and they will fill up this gap with the DNA. Again polymerase has to come and synthesize the DNA to fill this gap falling what is will come and they will synthesize this new strand. Now, the new strands are synthesized according to the basis that you have in the opposite strand.

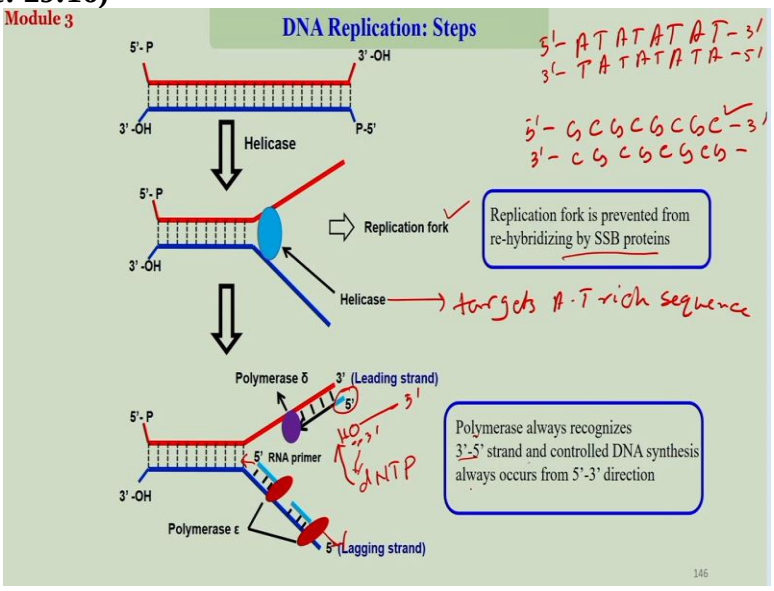
But these 2 will not be connected, this 2 will not be connected, these 2 will not be connected, this will not be connected. So you have to connect them together, then only you are going to get intact double helix. That is what you have to stitch the fragments together. That is where the ligation is coming. So, these are the nicks.

(Refer Slide Time: 28:43)



So this is the enzyme ligase. DNA ligase was the enzyme that actually joins the newly synthesized DNA strands and complete to fill up the nick, now you have the newly synthesized DNA. So, now you have 2 different complete sequences of genes in the newly generated cells.

(Refer Slide Time: 29:16)



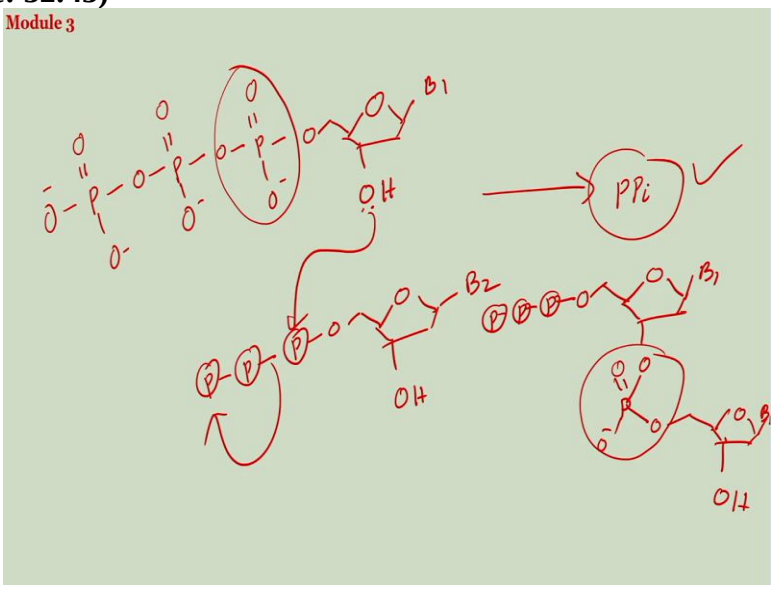
So, all these things I have talked about. So, one by one if we take first thing, so, this is 5 prime, you have the phosphate; you have the 3 prime free hydroxyl. The other way around, this is your parent strand, first enzyme is helicase comes, binds here that opens the double helix partially and forms the replication fork. Now, this helicase is the protein or it is actually can be a group of proteins also that targets a certain part or the certain sequence of the DNA. So, if you have a double strand DNA maybe all AT AT AT AT AT 5 prime to 3 prime.

So complimentary is TA TA TA TA.If this is one of your double stranded DNA, just consider short version and you have GC GC GC GC 3 prime would be CG, CG, CG, CG. Then which one is stronger which DNA stronger obviously this DNA is stronger because GC spears means 3 hydrogen bonds. Three hydrogen bonding said there in AT, there are 2 hydrogen bonds and therefore this is weaker. So helicase targets the AT rich sequence of the gene where there is multiple AT sequences.

That region the enzyme weaker targets because it is easier to open. Since this is weaker hydrogen bonding, it also easier for the enzyme to open that part. So helicase usually targets AT rich sequence of the gene and that what it opens up forms a replication fork. And once a little bit part is open, that actually forces the other but also to open and then it will be stabilized by the SSB protein.

So, that they do not come closer together. And then of course, once you have the replication fork, you have the RNA primase now, when you look at all the enzymes, so, many enzymes are involved in the whole of the replication process, when you look at the all the enzymes, you know, they actually all of these proteins, they actually do a single thing, if you look at the chemistry of it, chemically, what is the chemistry behind all this replication process, you know, it is a single chemical reaction.

(Refer Slide Time: 32:43)



And that reaction is if you have a nuclear base, B1 your 3 prime hydroxyl is free and you have the phosphate at the 5 prime it can be OH or O minus. In echo solution in right pH it will be O minus. So this is what exists in their cells. Second base B2 now I am not writing the whole of the phosphate, alright P circle, that means basically this part with the whole this part. The chemical reaction is only this.

Three prime hydroxyl of the first ways will react the phosphate will eliminate a diphosphate PPi will be eliminated and you will get B1 this is P B2 OH here you will have O 3 phosphate. Three prime hydroxyl is acting as the nucleophile attacking the phosphorus eliminating a pyrophosphate. And this is actually the driving force for the reaction because it produces a lot of entropy. So, degrees up from one molecule to large molecules you are creating one small molecule which has high degrees of freedom.

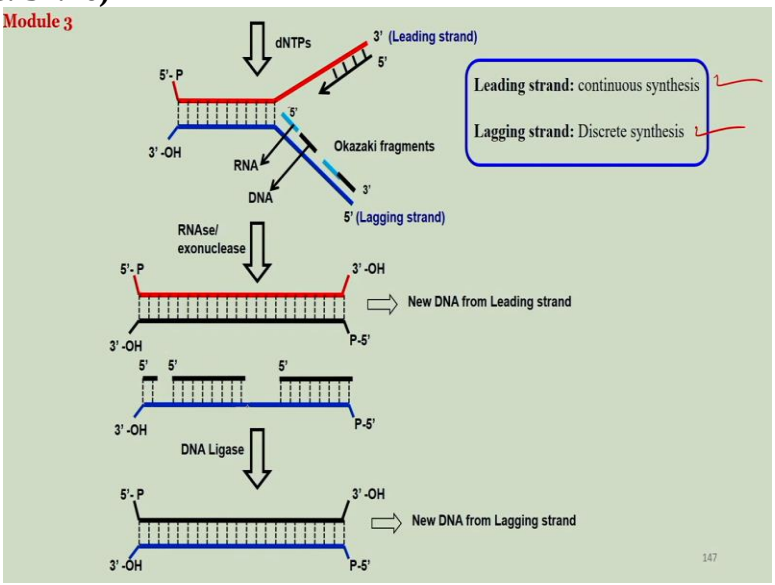
That is what is the is the reaction and that is the driving force for this reaction. Now, this is what happens in our cell in the biological systems to form the phosphate linkage. So, in the whole of the replication process, whatever enzymes whatever processes that you are steps that you are seeing, so, they are basically to facilitate this reaction only this reaction. Now, let us go back. So, here now, if you see, the DNA primer RNA primer came in, and as I was saying, it has the hydroxyl group free. So, this is 5 prime, 3 prime would be 3 hydroxyl group.

Now, this will react to the incoming dNTP at the 5 prime ends so, this will react with the 5 prime of it, eliminating the part of phosphate. That is the reaction. It cannot be the other way around. The 5 prime cannot react with the 3 prime, because you see you are actually in this case what you are doing, you are going opposite to the entropy. In order to form this bond, your respective dNTP has to come and stay here.

So it was freely available in the solution, it had high degrees of freedom at the beginning. Now, when it comes and sits there, then its degrees of freedom is restricted. So, reaction will not happen in opposite directions. Since your primer is available primaries already hybridized there, the direction from 3 prime hydroxyl groups reacting to the 5 prime end of the dNTPs more special and that is why it will go in this direction as shown.

On the other hand, if you look at the lagging strand, as I was saying, it cannot go this direction it can only happen in the forward process. That is why you need to have the discrete number of RNA primers. Polymerase always recognizes 3 prime to 5 prime strand and controlled DNA synthesis. Control DNA synthesis always occurs from 5 prime to 3 prime directions. That is where the newly strand new strand would be synthesized in the direction of 5 prime to 3 prime.

(Refer Slide Time: 37:26)



Now, once you do that, you can synthesize the leading strand continuously because it is always 5 prime to 3 prime. On the other hand, if you look at the lagging strand, your 5 prime to 3 prime is in this direction. So, once this is done, the next phase of replication forks when that is open, it cannot go in this direction. So it has to start again from the beginning. Another one so this is RNA.

This is DNA this is RNA this is DNA. That is how the Okazaki fragments should be forming, leading strand continuous synthesis lagging strand discrete synthesis. So once you have that RNAase. Which is a part of exonuclease will chop off all the RNA followed by the so in this strand here you will have the discrete DNAase after the elimination of the RNAse. And then the DNA ligase will come stitch everything together and you will get a newly formed the lagging strand. So, that is how you can get to individual DNA strands there.

So, now let us go back to the all the enzymes and look at them again. So now I can describe the exact functions of all the enzymes helicase. That was the first enzyme; the protein that unwound a portion of the double helix to form the replication fork, the engine targets A-T rich sequences in the gene. Next comes the SSB protein, single stranded binding proteins, the proteins prevent the replication fork from reanncaling so that they do not rehybridrates again.

DNA polymerase 2 polymerases are involved as we have seen, one DNA polymerase that is known as polymerase delta. That is for the leading strand binds to the parent DNA strand, which is 5 prime to 3 prime direction uses it as a template and begins replication along the new 5 prime to 3 prime direction this is known as leading strand. Now, as I have mentioned that because the DNA synthesis can only occur from the 5 prime to 3 prime directions, a molecule of a second type of DNA polymerase, polymerase epsilon that binds to the other template.

This molecule uses short RNA sequences as primers that is of course also needed at the beginning for the leading strand and continue strand elongation incorporating the DNA nucleotides. And that is how you get the Okazaki fragments. Then comes the another enzyme DNA ligase that stitches, the fragments together into the lagging strand that is also needed for the leading strand as well RNA primers as I mentioned, its type of RNA polymerase that provides of RNA primers.

So DNA polymerase synthesizes DNA, RNA polymerase will synthesize RNA. So RNA primase is a variety of RNA polymerases that will synthesize small RNA primase required to initiate the replication process. Then you need to destroy the RNAase that is the enzyme RNAase. This enzyme destroys the RNA primers once the synthesis is over. And finally, you need to stitch things together.

That is done by the DNA ligase that stitches, the synthesized DNA fragments, or that stitches the Okazaki fragments basically. So these are more or less the simplified version of the DNA replication process. Thank you.