Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 5
Mechanistic Overview of DNA Replication

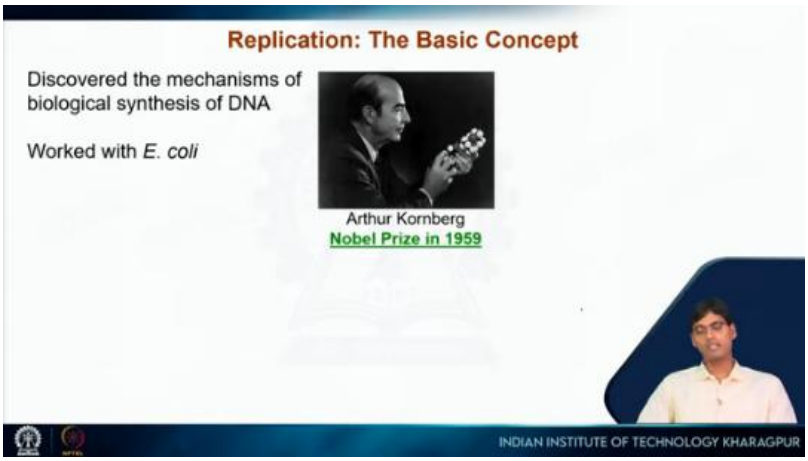
Hello everyone, I am Dibyendu Samanta from IIT Kharagpur and today I am going to discuss about the last lecture of this module, the module 1. Here I am going to discuss about mechanistic overview of DNA replication. During my last lecture, mostly I concentrated on the basics of DNA replication, how Meselson and Stahl proved that DNA replication is semi conservative, but here mostly I will focus on the mechanism of DNA replication and how it actually works. So, I must say here that the mechanism of DNA replication is very complex, there are many problems, there are many challenges and accordingly because of the long time of evolution, different life forms, I would say in the living system they also evolve with this problem.

They can take care of this because DNA replication is a very fundamental step in life. So, now in the first place I am going to discuss about the template and raw material for DNA replication and then the more complex part the mechanism of DNA replication and associated complexities. There are many enzymes that also involved, many different complex nomenclatures, but I will try to keep it very basic fundamental idea and then if you are getting more interested about this idea and this field then you can refer to any standard textbook. So here, Arthur Kornberg, he received the Nobel Prize in 1959 because of his phenomenal work on DNA replication. He particularly discovered the mechanism of biological synthesis of DNA that how DNA is getting synthesized by DNA polymerase.

It is an enzyme DNA polymerase. This enzyme name is very much logical, most of those names are very much logical in the biological world. Most of the enzymes end with ASE as polymerase. So, polymerase means some enzyme which will make the polymer. What kind of polymer in this case? This is DNA polymerase.

So, they make the DNA polymerase. As a result of that, the function of DNA polymerase is that they will add deoxyribonucleotides one after another and they will make the DNA polymer. So, now again Arthur Kornberg works with E. coli because E. coli is a model organism that is a simple bacteria and it has just one circular DNA molecule. This is the

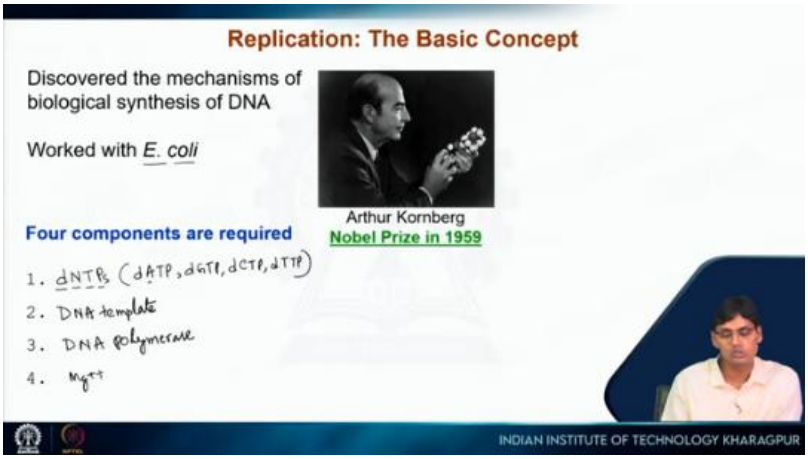
difference between our DNA and bacterial DNA. So bacterial DNA, they are circular, but in case of human or any eukaryotic DNA, they are linear and also in case of E. coli they have only one DNA molecule or one chromosome.



Now, what are the ingredients required in order to carry out DNA replication? So, to understand that if I say briefly that there are four major components those are required for DNA. So what are those four components? So, if we properly understand the DNA structure, what is present inside DNA then we will be able to provide the answer to this question. So, I would say the first component should be the raw material, which constitutes the DNA.

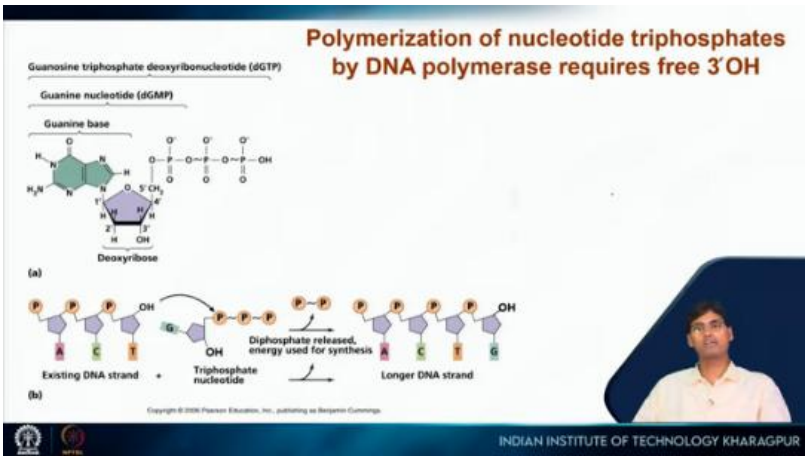
So, those are the components. D-oxyribonucleotide triphosphate. So, I am just writing here dNTP. So, that means here it can be dATP, A stands for adenosine here, dGTP, dCTP and dTTP because A, T, G, C those are present in DNA. So, those are the four major ingredients, the raw materials that should be required for the DNA replication dNTPs. Then as we know from the Meslson and Stahl's experiment that it is semi-conservative that means, we need a template and from that template we will get the new strand. As a result of that here also we need DNA templates that are the basic requirement for DNA replication so that we can copy the information from that template to make the newer copy. Now, some enzyme is required which will actually synthesize this DNA or synthesize this polymer.

The name of this enzyme is DNA polymerase. So, the first DNA polymerase was isolated I would say purified and characterized by Arthur Kornberg and that is why the name of that enzyme at that time was commonly called that Kornberg enzyme and then the last, but very important thing is magnesium ion. So, magnesium ion is also very crucial for this DNA replication step, because DNA polymerase cannot function without magnesium, it requires magnesium for its function. So, these four are absolutely required for DNA replication.



As I already told this is just for fun, you know this car actually belongs to Arthur Kornberg. He was so obsessed with protein, particularly about the purification and characterization of DNA polymerase so that he can understand the replication mechanism. If you see the number plate of his car here is purified. So, he is so obsessed with the purification of enzymes, particularly the DNA polymerase.

So, this is in California and then followed by the number here, the special number purify, but anyway that is not that important. Now, the chemistry of this reaction is how DNA polymerase actually polymerizes. So, the polymerization of nucleotide triphosphate by DNA polymerase require free 3 prime OH group, this is very important and that is why I told several times that any nucleic acid whether it is DNA or RNA it always extended from 5 prime to 3 prime direction, again I will discuss here so as you can see so this is an existing DNA strand. So, here in this existing DNA strand you can see that 5 prime position, why because in this strand the 5 prime carbon of this sugar is capped with the phosphate group. And here if you see that this is the 3 prime because the 3 prime carbon of this sugar is with the OH group here. So, this is free to react.



So, only this DNA can grow in this direction that is why I always say that 5 prime to 3 prime extensions but why in this extension? Soon you will know. Now what will happen? This OH group available at the 3 prime end of this existing DNA strand, this OH can acts as nucleophile and it will create some nucleophilic attack on this incoming nucleotide. In this case here as you can see this is GTP. So, here I already mentioned that you can see the guanine base here and then this is the sugar and this is the triphosphate.

So, this is GTP. Now because of this nucleophilic attack you can see that 2 phosphate here, diphosphate it is getting released and now you will see a new phosphodiester bond here. So, as a result of that now in this existing DNA we have 3 nucleotides A, C and T, but here this is longer DNA. Whatever we had before now we have one extra nucleotide here, here this is G residue or the G nucleotide added here.

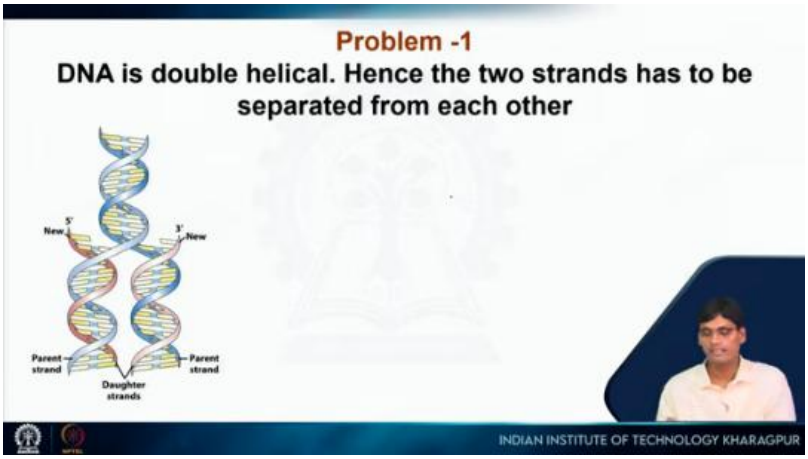
But if you see at the end of this G residue again you have the 3 prime free OH group. So, again one nucleotide can be attached here and that is why DNA or any nucleotide, it does not matter whether it is DNA or RNA the basic mechanism is the same and it always grows from 5 prime to 3 prime direction. But, now the question is why is this G coming here? Why not A? Why not P and C? So, here for the explanation, I am just showing here that G is getting added here and you have G residue already added, but this is template dependent synthesis. So, as a result of that I would say that here this is 5 prime and this is 3 prime.

So, you have here TG and A and if this is so then in the growing chain here as you can see this 5 prime A C and T and then the next one here is G that means, the guanosine is added here. As a result of that since it is template dependent that means, in the template you must have here C and that is why this G is added here. Similarly, the next nucleotide will come depending on what is present in your template. If you have G here then the next one will be C because inside the cell you have all these 4 nucleotides available in the surrounding, but it is the template sequence which is present here.

So depending on the template sequence the enzyme DNA polymerase will put the correct nucleotide, this is a very precise thing. So, DNA polymerase has a very precise role here. So, it will always put the correct nucleotide, but if you study in detail, you will find that rarely but sometimes some wrong nucleotide can be added there also. Fortunately DNA polymerase also has the proofreading activity.

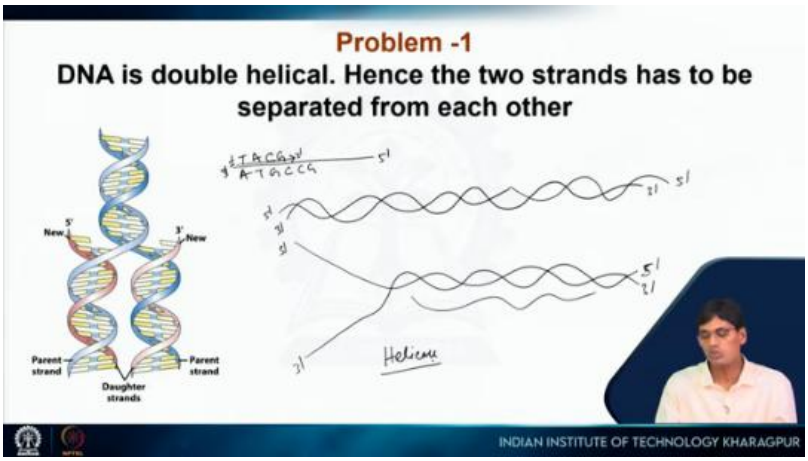
So, they can also understand that the wrong nucleotide has been added. So, they will rectify that mistake. So, there are many other mechanisms to correct the information. So, if something is wrong during DNA replication that should be corrected because DNA contains the instruction of life.

So, every information is actually present in DNA and finally, from DNA we will get the RNA and then protein. So, as a result of that the main information should be intact and should not be spoiled with some wrong basis or some kind of mutation. So, this is the chemistry of DNA replication, how the next nucleotide is attached to the existing DNA strand and which is catalyzed by DNA polymer is a specific enzyme which requires magnesium ions. Now we are going to discuss about the first problem here because whenever I am telling you that this is template dependent, I am just showing that here you have one template that this is 3 prime to 5 prime, you have sequence here A T G C C G and the new stand as a result of that this should be T A C G it is going like this. So, this is 5 prime to 3 prime, but this is possible. I am explaining in a very simple manner that you have the single strand available as a template, but inside our body, inside E. coli cells everywhere the DNA is double stranded. We discussed that if you increase the temperature of DNA, high temperature, I would say 70 degree 80 degree centigrade then those two strands will be separated, but our body temperature is much lower than that so as a result of that DNA is always double stranded and double helical in our body.



So, then the problem is how we will get the single strand template in order to replicate. I would say if we draw here like this, this is a DNA molecule. So this is 5 prime, this is 3 prime, 3 prime, 5 prime, DNA those two strands are antiparallel and that is why we are writing in this way. Now, the thing is the temperature is the ambient temperature in our body. So the hydrogen bond between these bases will not be broken and you will not get the single stranded DNA in order to replicate. As a result of that we have some more complicated mechanism and that will solve this problem. So, as a result of that what will happen? So, we should get this kind of form.

So, this is 5 prime and 3 prime. So, if you are getting this kind of single standard from then we can synthesize the DNA. Then how can we do that? If we pull apart these 2 strands of DNAs but that it is not possible to pull apart these 2 strands in this way then the front part here, the DNA will be tangled together. So, it is a real problem that the DNA will be tangled there and also how we will pull that. So, for example, if we believe that we have some enzyme so that the enzyme will break this hydrogen bond between the bases of double stranded DNA in ambient temperature. So, the name of this enzyme is helicase. If helicase works here and it actually destabilizes this hydrogen bond, those two strands will be separated. But still there will be some problem because it is not just double stranded DNA, it is also helical so they have kind of the coil to each other because of their helical nature.



So, as a result of that we need to solve this problem also. So, that is why we have another enzyme that is called topoisomerase. It seems like a very complicated name, but it is not actually very complex. See the function of the topoisomerase is isomer. So, I am saying that

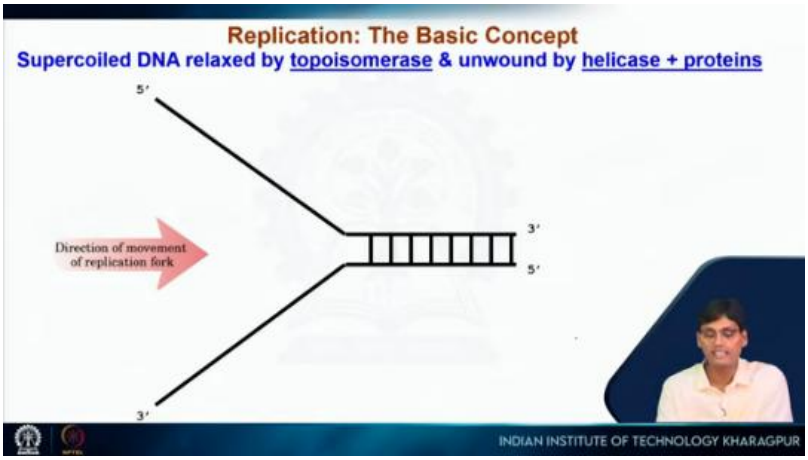
when we are trying to separate these two strands by this helicase then the front part of the DNA will be tangled. Since we are drawing in this linear form of DNA that is why I am telling that, but if you now imagine that the bacterial DNA, E. coli DNA is circular it will be more problem. We cannot separate in that way. So, as a result of that topoisomerase is one enzyme.

So, this is called the replication fork where you can see these are two strands which are getting separated. So, this is called replication fork that means, here the replication will be happening. So, just ahead of the replication fork, this enzyme will work, it will cut DNA strand and then they release the turn.

So if that is happening just ahead of this fork then helicase can work. This is the function of topoisomerase. So the name is topoisomerase because the DNA double stranded double helical structure, if their sequence is same, but one is getting relaxed, they are this turn coiling that is getting reduced. but if you see their sequence their molecular weight are exactly the same so they are isomers but what kind of isomer? Those two are like differing because of the topoisomerase action their topological problem has been solved so they are getting like they are differing by their topology so that is why this enzyme is named as topoisomerase. So, the topological problem is getting solved by this enzyme just ahead of the replication fork.

So, this is called topoisomerase and together like helicase and topoisomerase, they work together in order to get a single standard template at the replication fork. So, that DNA polymerase can synthesize DNA smoothly.

So, I am going to the next slide here you can see whatever I explain that supercoiled DNA is getting relaxed by topoisomerase and unwound by helicase and some associated proteins are there also but many other small things are also involved here. So now if you see this is the direction of movement of the replication fork. So, that you know right now the replication fork is here after some time when replication will be happening.

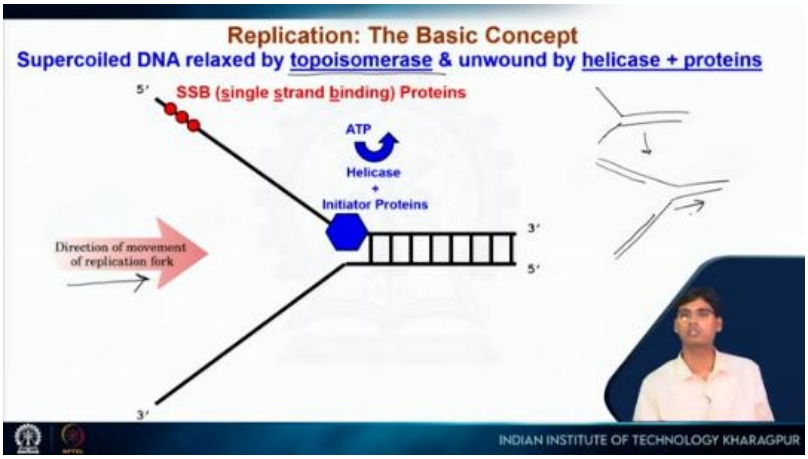


Now, after some time I would say the replication fork moved here. So, here already, maybe DNA has been synthesized in this place. So, over the time that way from 1 DNA molecule you will get 2 DNA molecules.

Here this is helicase which requires ATP and some other proteins are also involved and that actually break hydrogen bonds in between these two strands and topoisomerase is also

functioning just ahead of it to relax DNA molecules and that way it solves the problem. So, in normal body temperature we still have a single standard template in order to carry out DNA replication and some other protein which is called single stranded binding protein, those are also attached to the single stand. So in this temperature again those two complementary DNA strands will not be coming together just to help the process of replication.

Now I will discuss a few problems and I will discuss their solutions also.



First one I told about the topological problem and now the second problem is DNA polymerase needs a free 3 prime OH of a pre-existing nucleotide to extend the chain. As I already explained during the discussion of the mechanism how the next nucleotide is getting attached. During that time I mentioned that it requires a free 3 prime OH group as you can see so as a result of that we are trying to make the DNA, but from where we will get the initial DNA, the pre existing DNA so that the reaction can be carried out because DNA polymers cannot add independently. As a result of that we need to solve this problem also in the next slide I am going to discuss.

So, here as you can see DNA should be like I mentioned always that any nucleases should grow from 5 prime to 3 prime end so I would expect that this is the 3 prime end. So we can see that 5 prime to 3 prime but the thing is that DNA polymerase requires a pre-existing 3 prime OH group. That is why DNA polymerase doesn't start the replication. Instead of that we have some other specific enzyme, which is called primase. So primase is the enzyme people actually named it primase so that they can start the process, they can make the primer, but this primer is made up of ribonucleotides. So, initially we are actually discussing about the synthesis of DNA, now we see that at the beginning we are making the RNA primer and that is why we are telling it primer because this is not the final thing, this is primer. So, for example, when we paint some house or apartment, before putting the actual colour we put something which is called primer so after that we put the actual colon, similarly this is primer. So, it is just starting the process, but finally, we have to remove this thing because it is made up of ribonucleotides and this is called RNA primer. So, finally, we will remove this by some specific mechanism and it will be replaced by ah DNA itself.

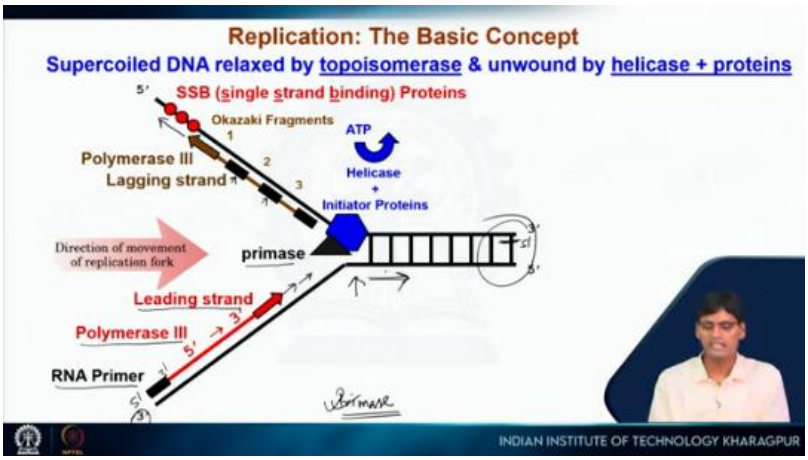
But in order to start this process we first need to put the RNA primer here. So, this is 5 prime end and this is the 3 prime end of RNA primer which is polymerized by the enzyme which is called primase. So now in this short segment of RNA primer you have a free OH group at the 3 prime end. So now DNA polymerase can extend this chain.

So, as you can see here now the DNA polymerase in this case specifically this is called DNA polymerase 3. So, you will see DNA polymerase 1, 2, 3 and many other things. Most of those things are named based on their sequential discovery, but it is ok if we just mention DNA polymerase. So, DNA polymerase will extend this from 5 prime to 3 prime direction and it is synthesizing this strand and this strand is going steadily in this direction and this stand is called leading strand as you can see here.

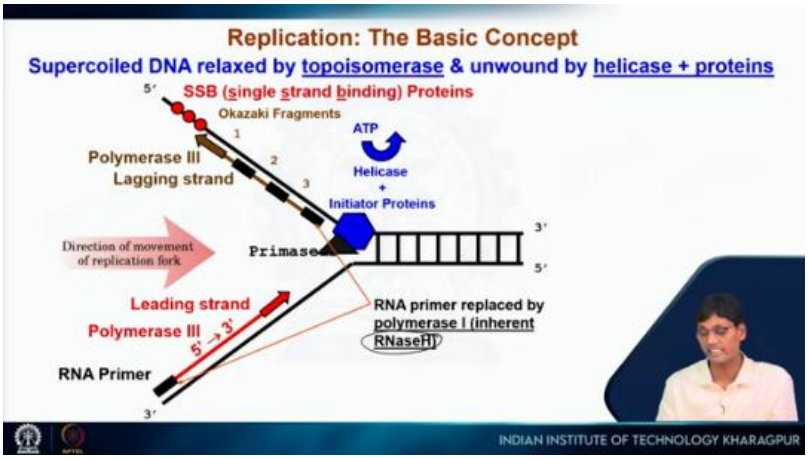
Why leading strand? Because it is going very steadily and then if you see here in this portion the fork will move towards right, this side after some time then this strand will be extended like this, it is kind of continuously getting synthesized. So as a result of that it will go in this direction, but now the other strand has the problem because like always DNA is getting synthesized from 5 prime to 3 prime. So there is no option of getting synthesized from this to this because 5 prime positions are present here and 3 prime should be here because DNA is antiparallel. So, then it has to go from 3 prime to 5 prime, but that is not possible because there is some problem in this place. So as a result of that what happened is in this strand the DNA synthesis happened in a different orientation. So, here in this strand you will see that one short primer is synthesized by primers and then the polymerase will be extending this primer towards this direction. So, this is a short end.

Why sort? Because this end is not yet available, if this end is available then one primer can sit here and it can quickly go in this direction steadily just like leading stem, but this end is not yet available. So, it is still a double stranded and helical structure. So, as a result of that this strand will be synthesized in a sequential manner. So, as a result of that you can see another primer came here and this short segment of DNA will be synthesized then another primer will be there. So, then when the fork will move in this direction then another primer will be attached.

So, it needs multiple small primers here and then small segments of DNA will be synthesized and those small segments of DNA are called Okazaki fragments based on the scientist name who discovered this, Japanese scientist Okazaki. He actually discovered this mechanism here in this lagging strand synthesis and that is why those are called Okazaki fragments. So, this way you can see the leading strand and the lagging strand, lagging strand because this is lagging behind because it is getting synthesized a little slower than the leading strand and it requires multiple primers and in a sequential manner it is getting synthesized.



So, this way both the strands will be synthesized. But now the third problem is that the RNA primers need to be removed because our goal was to synthesize DNA, this is DNA replication. So there is again another specialized enzyme, which is called polymerase, 1 DNA polymerase 1 actually. It has inherent RNAs, I would say H like structure activity so this enzyme whenever will see some kind of DNA RNA hybrid. So, you can see this is RNA.



So, now this is DNA. So, DNA RNA they have this despairing here also. So, whenever it sees a DNA RNA hybrid, it will destroy the RNA part. So, as a result of that, the RNA segments, the RNA fragments will be removed from this lagging strand and also from the leading strand. As a result of that no more RNA here, but now if you see the RNA is not there, but DNA is present here. So, this is one Okazaki fragment here. So, here you have the 3 prime OH available so again DNA polymerase can synthesize this part. As you can see that DNA polymerase will synthesize this part and the RNA primer will be removed and DNA will be there and finally, the gap, the phosphodiester bond will be formed by one enzyme this is called ligase. This is nothing but DNA ligase. Again ligase in common term ligation means to join two things together. So, similarly the name of the enzyme is DNA ligase so this DNA ligase will fill the gap.

So, as a result of that now you have continuous DNA molecules and this is not made up of RNA anymore, everything is DNA here. Now from one DNA molecule you are getting two DNA molecules. This is the basic mechanism of DNA replication but there is some problem. If you see how this part of this DNA will be synthesized when we remove the RNA from this end because again the 5 prime 3 prime the directional issue is there. So, this is particularly a big problem for eukaryotic replication because the eukaryotic DNA is linear. In the case of circular DNA it is not a problem, it will be synthesized if you try to draw in pen and pencil you will be seeing that it will get the 3 prime OH group and it will be synthesized.

But in case of linear DNA it is a real problem that how this end will be synthesized, the linear in case of eukaryote the end is called telomeres. The end of this DNA or the chromosome is called telomeres. There are some specific problems and some specific solutions are there in order to replicate the telomeres, the end of the DNA. If you want to go into more details, you can follow any standard textbook.

In this class, I am not going to discuss that. This is much more complicated stuff, but whatever I have discussed, this is the fundamental understanding, the basic understanding about the DNA replication, which is happening inside every living organism and that is all.

Thank you very much.