Cell and Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Week 10
Central Dogma of Molecular Biology (Part 2)
Lecture - 35
Replication (Part 1)

Hello everyone, this is Dr. Vishal Trivedi from the Department of Biosciences and Bioengineering, IIT Guwahati. So in today's lecture, we are going to start discussing the first event, which is the synthesis of new DNA from the pre-existing DNA. If you recall when we were discussing cell division, we discussed that during the S phase, the DNA is synthesized and then you have two copies of the same DNA. And then these two copies of the same DNA are actually going to be shared between the cells, and that is how the cell is actually going to replicate and the cell is going to divide. So, the synthesis of the new DNA molecule from the pre-existing DNA molecule is known as replication.

Now replication is an important event. Because it allows for the synthesis of DNA, replication means the synthesis of DNA from pre-existing DNA. You can imagine that this could be a DNA of the parents and this could be a DNA of the offspring. And you can understand that this whole process has to be done with a lot of precision and a lot of accuracy so that you are not going to carry the bad information.

You want to carry the same information that is present in this particular DNA, and on the other hand, it has to be completed in a given time period, so replication has two tasks: one is the synthesis of the new DNA considering old DNA as template. And the second is that it should be accurate, so it should be almost an exact 100% replica of the older DNA. There should be no mutations or there will be no substitutions. There should be no alterations and all of that. So, to achieve this, the machinery has to be very robust.

Machinery should have the components so that it can do this job, the synthesis job. But at the other end, it should also have the components so that it can perform quality checking. And the replication is also related to the genome. the prokaryotic genome as well as the eukaryotic genome. If you recall from the previous lecture, we discussed the prokaryotic genome and the eukaryotic genome, and there is a significant difference between the two, right? And as a result, the machinery is also different in the prokaryotic system and the eukaryotic system.

So, in the prokaryotic system, we have different types of components that are required for the synthesis of the genome. Whereas in eukaryotic systems, the eukaryotic genome is

more complex, requiring different types of machinery. So, because the machinery is different, I have split this particular thing into two components. So, we are going to first discuss prokaryotic replication and then we are actually going to discuss eukaryotic replication because when we discuss this separately or when we discuss prokaryotic versus eukaryotic. You will understand how the machinery is being adopted from the prokaryotic system to the eukaryotic system so that it is more efficient and actually brings more accuracy to the processes.

So, the first question is, what is replication? We already discussed, right? So, DNA replication is a biological process that helps to transmit the DNA or genetic information from generation to generation by producing two identical DNA strands for the daughter cells from the double-stranded parental DNA. This is anyway we have discussed. What is the importance of replication? Replication is a way of duplicating all the genetic information from all living parent organisms to the daughter organisms. thereby helping maintain the genetic material's intactness and the organism's survival. Now, if this is such an important event and it mostly depends on the type of DNA you would like to do the replications and so on, it is important to understand the genetic material of the prokaryotic system because we are going to start the replication of the prokaryotic system first.

So, prokaryotic genetic material remember that when we have discussed in detail about the prokaryotic genetic material. So, prokaryotic genetic material is a single chromosome. And it has mostly been found in the cytoplasm, right? So, prokaryotic genomes can be found in the cytoplasm. They are negatively supercoiled and circular. Generally, they can be found in a singular number.

However, exceptions exist. For example, Vibrio cholerae has two copies of the genome, right? So, it has two genomes. So, prokaryotes such as E. coli contain self-replicating extracellular genetic material, which is also called plasmid. So, you have the circular genome or the circular chromosome.

So, it is called a chromosome. So, see you have a single copy of the chromosome. which is negatively supercoiled and is a circular DNA, and mostly there is only one copy of that circular genome present in the prokaryotic system, but there are exceptions, as in Vibrio cholera, where there are two genomes. Apart from this chromosome, you are also going to have the extra chromosomal self-replicating genome, which is called plasmids. When we talk about genetic material, we are actually going to talk about the composition of chromosomes as well as plasmids.

Now, as far as the plasmids are concerned, plasmids are small-sized chromosomes,

small-sized DNA. It is only between 1500 and 2000 KB. Therefore, it is very small compared to the actual prokaryotic DNA. They contain the origin of applications and replicate independently. So, a plasmid does not require any kind of help from the chromosome.

They are independent, self-replicating, extrachromosomal genetic material. Now let us first ask the question of how you can actually be able to have the replication. So there are three different modes that are being proposed as far as replication is concerned. Replication can be done in three different modes. One is called the conservative mode, the second is called the semi-conservative mode, and the third is called the dispersive mode.

So, in a conservative mode, if it is in conservative mode, then it produces two DNA helices from one single original DNA helix. One helix contains the parent's DNA while the other contains entirely new DNA. Accepted postulate but does not have much significance. So what is meant by the conservative mode is that from the parental DNA you are going to have after replication. So if you are going to do the replications, you are going to generate, for example, from the two strands, you are going to have the four strands, right? So four strands are actually going to be segregated, and as a result, you are going to have, for example, here you have one and two.

So, 1 and 2 will go into the parents, whereas the newly synthesized 3 and 4 will go into the daughter. So, this is called the conservative mode. In a conservative mode, there will be no mixing of the content from the parent's DNA. So, that is why in the second round of applications, you are going to have the parent DNA, and you are also going to have the daughter DNA separately. This means there will be no mixing of the previous copy, and it is actually going to have the pure DNA present in the parent DNA.

Then the second mode is called semi-conservative. So one of the two helices forms, each containing one new strand and one parental strand. According to Watson and Crick, one strand serves as a template during the replication process. It is mostly accepted that DNA polymerase needs a strand to form a new complementary strand. So what is meant by the semi-conservative mode is that from the parent DNA, for example, you have the parent DNA as 1 and 2; it is actually going to do the replications.

And once you do the replications, it is actually going to form the four copies: one, two, three, and four. So, 1 and 2 are the original copies, while 3 and 4 are actually going to be the newly synthesized copies because from 1, you are going to have copy number 3, and from 2, you are going to have copy number 4. Now, in the semi-conservative mode, the 1 is actually going to make a pair with 4, and the 2 is actually going to make a pair with 3.

So, this is actually going to be 1, 4, and 2, 3. There will be a dilution of the genetic material of the parents, and that's how the parents are actually going to share the 50% copy with the offspring.

That means the two offspring that are going to be produced after the first replication, or the two DNA molecules that are going to be formed after the first replication, are actually going to have the mixture. It means going to have 50% from the parents and 50% of the newly synthesized DNA. And this continues because even if I have a second replication discontinued, you are going to have that kind of dilution. So in the first applications, you are going to have the mixing of the DNA from the parent. So you are going to have the two strands of the parents that are going to be diluted.

Now the third method is called the dispersive. So, this mode generates the DNA helices with an alternating pattern of old and new DNA segments, and it does not have any kind of biological significance. So, what this means is that it is in the dispersive state. What it says is that you are going to have the parental DNA, and after the replication, 50% of the portion is going to be replicated with the original DNA, and the other 50% of the portion is going to be replicated with the new DNA. This means you are going to have a mixture of both strands and even both components.

This is not acceptable because it is not possible; when you have the copy numbers 1 and 2, you are going to generate the 3 and 4. You are not going to have 1 followed by 3, followed by 1, followed by 3 like that. So that kind of scheme is not possible because once the DNA polymerase sits and starts synthesizing the DNA, it will not come off, and if it comes off, then it is actually not going to have this particular type of pattern. So this is the most unacceptable method. Now, the question is, do we have the three different modes: the conservative mode, the semi-conservative mode, and the dispersive mode? And these are the three different types of models that have been proposed, and how the replication is going to happen, right? But the question is how scientifically we can prove this.

So, to prove this, there is a simple experiment and a classical experiment that is being conducted by some scientists. So, the Meselson-Stahl experiment, or Matthew Meselson and Franklin Stahl, actually performed the experiments in which they asked what the mode of replication would be in the bacterial system. So, what was the aim? The aim was to establish which method is applied in the prokinetic replication: is it the conservative mode, the semi-conservative mode, or the dispersive mode? So, to ask these questions, they have prepared and designed an experiment, right? And these are the requirements. So, you require media; you require media to grow the E. coli for several generations, and the media has two different types.

You are going to have the media with the standard form of nitrogen, which is N-14. So I will say this is the normal nitrogen, which is the N14 nitrogen. And then you also require a medium with the rare and heavy form of nitrogen, which is called N15. So these are the two different types of isotopes that you can actually use. One is normal nitrogen and the other is heavy nitrogen.

And then you can actually be able to, you require the analytical reagents, right? So you can actually require the techniques so that you can be able to separate the DNA that is being formed by the N14 versus N15. So you require a cesium chloride gradient to separate DNA based on its density. Remember that this is heavy and that this is light. So, it is actually going to have different types of densities, and that is why N14 can be separated from N15 if it is present in the same mixture. So, if you have the pure DNA of N14 or the pure DNA of N15, it is going to form different bands.

Into the cesium chloride gradient mixture. And if you are not sure about the gradient centrifugation, remember that we have already discussed that technique while we were discussing cellular fractionation and we have discussed density gradient centrifugation. So, I think we have taken an example of sucrose, but here they are using cesium chloride. So that is the only difference. And you know that the gradient can be in the upward direction or the downward direction.

And we have also discussed in detail the principles and how things are separated when you run them on the gradient. So the gradient is actually going to separate the molecules based on their density. So this is the procedure that you are going to follow. You are going to grow the culture of E. coli and 15 isotopic media for several generations so that the heavy isotope is incorporated into the purine and pyrimidine bases.

Remember that most of the nitrogenous bases are made up of N-15. So when you grow them in N15 media, all your purines and pyrimidines are actually going to be labeled. Means all the nitrogen that is present as N14 is actually going to be replaced by N15. So all the DNA is going to be heavy.

And then some of the E. coli from the heavy isotope media are taken and transferred into the normal media containing the N14 as the nitrogen source. So, now what will happen is that when it is going to synthesize the new DNA molecule, it has to utilize this N14, right? So, based on the density gradient, the DNA band will be generated in the centrifuge tubes. For equilibrium density gradient analysis, the DNA is collected from the media and put into 6 molar cesium chloride. DNA samples in a heavy salt gradient are taken into the centrifuge for 50 to 60 hours at 100,000 g rotations. So, this is what exactly

going to happen when you are actually going to have the E. coli into the N15 media. When you run them onto the cesium chloride density gradient, what you will see is that it actually forms a separate band, and this band is for the N15. Whereas when you transfer that into the N14, you will actually have the bands which are of intermediate density. So this is the region where you are going to get the band for the N14, and this is the region for the N15. So when are you going to grow the bacteria in the normal media after N15 media? It is actually going to have the N15 band in the intermediate spin, intermediate position. It is not going to be related to the N14 or the N15.

Now, if you grow further, if you go for several more generations, then what will happen is that you are actually going to start getting the band that corresponds to the N14. which means that there will be some DNA that is going to be served as the template for the N14 and that is how you are going to have the intermediate DNA, but you are also going to have the N14. Now, this will continue, and that is how this will be the final product. So, if you are only growing the bacteria in the heavy media, you are going to have a single band at the N15, but if you are running it in the N14 media, the DNA is initially N15. So N15 DNA is being transferred into the N15-labeled DNA-containing bacteria; when you put it into the normal media, it is going to have 50% N14 and 50% N15, and as a result, after the first generation, you are going to see a band that is of intermediate density.

So this is actually N14 by N15 media, right? and this is going to be the intermediate DNA. Now, if you continue for several more generations, then it is actually going to form the N14 dimers as well as the N14 and N15 intermediates. So it is actually going to form the intermediate DNA band as well as the N14 live bands. and this will continue because this is the actual thing. So remember that when we were talking about the semi-conservative mode, what we said is that if you have this DNA and you have the two strands at one and two, after the replication what will happen is that it is actually going to form the four strands.

It's going to form two strands, and it's going to form them. So, 1 is going to make a pair with 4, and 2 is going to make a pair with 3. This means, imagine that this is N15. So, if this is N15, then this is going to be N15. Now, the new DNA that is going to be formed is not going to be N15 because there is no N15 available in the media.

So, what will happen is that the N1 strand, which is going to be N15, the other strand is going to be N14, and the same is true here: N15 and N14; and that is why it is actually going to give you an intermediate density. Now, if this goes for another generation, then this N14 is going to replicate, this N14 is going to replicate, and N15 will again have

another copy. So, this is actually what happens: when you are going to have another replication, you are going to have 4 DNA molecules, right? So, you are going to have the N15 as an N14, right? So, 2 copies of this, and you are going to have the 2 copies of this, right? So, this is actually going to give you a band that corresponds to the N14, and this is going to give you a band for the intermediate. So, that is how you are going to get this actually here in subsequent generations. So, what you can conclude from this is that whenever there are DNA applications, they are actually going to be impure; they are going to give you DNA that is impure.

So, it is going to have both the original copy and the new copy. So, out of these three proposed replication modes, the semi-conservative mode is the selected mode that can be observed in both the prokaryotic and eukaryotic systems. Now let us talk about the replication machinery. So, first, what are replicons? So, replicon is the region of the DNA that is going to participate in the replication reactions. Replicon, or the small stretch of DNA that is going to be involved in the replication, is a DNA segment of the spore period that undergoes replication.

Replicants, what is there in the replicant? So, the replicant is going to have the origin of application, and it is also going to have the termination regions. For example, E. coli has one replicon in its genome. So, it is going to be monoreplenonic, which means it is going to start from one side and it is going to end on the other side.

So, it is going to have the monoreplicated. Whereas eukaryotes have multiple origins of replication, they are therefore multireplicated. So this means if it is a eukaryotic genome, then you may have multiple origins of replication. And why is it so? Because eukaryotic genomes are large compared to prokaryotic genomes. So they can actually afford to go with the single origin of replication so that by the time the replication is over, they are also going to have the synthesis and other kinds of preparations.

So, let us first talk about the origins of applications. So, the origin of applications. As the name suggests, one of the origins of replication represents the starting point of replication in prokaryotes. It is approximately 245 base pair 80 reach region cis-acting sequences. What is meant by the 6 acting sequences? Six acting sequences are that they can affect only the molecule of the DNA in which they reside. So, cis acting means they are actually going to affect the molecule.

Now, the second question is why it is AT-rich? So, the AT-rich sequence is preferred because AT-rich sequences are easy to melt. You know that A is making 2 base pairs with T, whereas G is making 3 base pairs with C. So, if you have G in GC regions, then you are supposed to break the 3 hydrogen bonds, and it is difficult to break the 3

hydrogen bonds compared to the 2 hydrogen bonds. So, for the melting or unwinding of duplex DNA, less energy is required to break the hydrogen bonds than for the GC sequences. The melting of duplex DNA is ATP hydrolysis-dependent, as the energy released helps break the hydrogen bonds between A and T.

The above-mentioned ATD sequence is recognized by the enzyme known as DNA helicase to initiate the unwinding process. So this is the typical original application where you are going to have the 200 base pair, 245 base pair long stretch, and it is going to have the 80 reach region in the region, and it is going to have the DUE and DOR. DUE means the Duplex Unbinding Element, whereas DOR is called Deno Oligomerization.

So, the origin of the application in E. coli is known as the ODC. It contains two short repeat motifs, such as five copies of the NINEMER. Sequences spread throughout the origin of replication. DOR site serves as the DNA binding site of the DNA A, which is a replication initiation protein, and the three copies of the 13 base pair or 13 mer 80 reach repeat, which is called the DO site, where the DNA starts unbinding. Upon the binding of DNA A at the 9-mer region, the 13-mer region starts melting. So, origins OREOC contain the 11 copies of the 5-prime GATC 3-prime repeat methylated on both strands, and only complete methylation can lead to the initiation of replication.

Hemimethylated origins cannot initiate replication until they are fully stored or methylated. So, it is important that the methylation is also going to control the replications because remember that the methylation is the defense mechanism, right. We are going to discuss some of these enzymes, right? So, which I think are going to be used in molecular cloning. So, the restriction enzyme, restriction methylated system is a kind of defense system, right? And that is how the machinery will not initiate the replication until the DNA is fully methylated. If the DNA is not methylated, hemimethylated, or unmethylated, then it is not going to be replicated because it is considered to be a host; it is going to be considered to be DNA of external origins.

Then we have the replication fork. So, a replication fork is going to be formed, and that is how it is actually going to initiate. So, once the DNA has been melted at the original replication, it is going to form the replication fork. So, the Y-shaped structure is generally found when the DNA starts melting and opening up. As the DNA opens up bidirectionally at the origin of C, two replication forks are generated. The extension of the two oppositely directed replication forks leads to a replication bubble.

So, once the replication fork moves in this direction as well as in that direction, it opens the DNA in both directions, and as a result, it is actually going to form a bubble-like situation, right? The replication fork is going to move in both directions, and this is going

to be called a replication bubble. And this replication bubble will move in both directions where it is actually going to start synthesizing the new DNA. So this is going to serve as a template for this machinery, and this is going to serve as the machinery for this replication. So you are going to have one machine that will run in this direction, and they are going to have another machine that will run in this direction.

So, that is how it is actually going to replicate both sites. So, you are going to have the replication in this direction, and you are going to have the replication in this direction. To maintain the single-stranded situation, the single-stranded DNA binding proteins are going to coat the single-stranded DNA to prevent the rewinding of the double-stranded DNA. So, these are some of the components that are going to be present within the DNA structure. So, you are going to have the original replication applications, and then it is going to form the replication forks. Apart from this, you also require a battery of the machinery that is being formed or that is being assembled onto the replication fork to start the replication.

So, these are some of the important enzymes that have been found in prokaryotic replication. So, the first enzyme is the DNA helicase, or helicase, right? So, it is called as DNAB. It melts or opens up the DNA at the replication fork. So, it is going to be the first enzyme that is going to sit at the original replications and then it is actually going to start the opening of the DNA. Then the second is the single-stranded DNA binding proteins, and these will prevent the unbinding of the single-stranded DNA from the double-stranded DNA.

So you can imagine that as soon as the bubble is formed, the single standard DNA binding protein will go and sit on the nucleotides so that they do not have any kind of interaction. Remember that these two molecules are complementary to each other, so as soon as they open, they are supposed to remain like that so that the other molecules will come and sit and do their jobs. But if you do not do that, then they will come and stick to each other because every "a" is going to have a "you know." Complementarity to T and every G is going to have the complementarity to C. So this complementarity can only be broken if you have a molecule that is sitting on this right.

So that is how you are going to have no interaction between these two, and that is how they are going to remain separated, and then the other machinery will come and sit and start the replications. Then you also have the topoisomerase or DNA G. Topoisomerase works in the region ahead of the replication fork to prevent supercoiling. And then you also have the DNAPol1, or it is also called the Kornberg enzyme because Kornberg is the first scientist who discovered DNAPol1.

So, the first DNA polymerase to be discovered in E. coli was by the Nobel Prize winner Arthur Kornberg. And the gene that is actually coding for DNA pol 1 is called pol A. And it is a monomeric protein of 928 amino acids or 109 kilodaltons.

And it has three different types of activities. It has 5' to 3' polymerase activity. It has a 3' to 5' exonuclease activity. And it also has a 5' to 3' exonuclease activity. So don't worry about these activities because we are going to elaborate on or discuss them when we talk about the other events. So, these three activities are really important. This activity is important for DNA synthesis, and the other two activities are required for proofreading as well as other kinds of activities.

DNA polymerase 1, or the Kornberg enzyme, has 5 prime to 3 prime exonuclease activity that is independent of the other activities. Then we have the proteases. Cleaves polymerase 1 into two fragments. So, if you treat DNA pol 1 with the protease like trypsin, then it is actually going to generate two distinct fragments. And then you are also going to have the smaller fragment which has the proofreading activity that is the 5 prime to 3 prime exonuclease activity.

Clino fragment is also very important and popular in terms of in vitro replication, such as PCR. So, sometimes people use the clino fragments rather than the complete enzyme because with the complete enzyme there is always a danger of having the 5 prime to 3 prime exonuclease activity, which may actually interfere with DNA replication under in vitro conditions. So, it has low processivity, which means only 200 nucleotides can be processed. Then it also has a low polymerization rate.

So, around 20 nucleotides can be added per second. Apart from these, you are also going to have other kinds of activities associated with DNA Pol I, which include RNA primer removal, gap filling, and DNA repair. These three will actually be discussed in detail when we talk about the elongations and terminations. And it is actually a metalloenzyme, so it is actually a zinc-dependent enzyme. Then we also have DNA Pol II.

So, the monomeric protein with Pol B as a structural gene is a size of 90 kilodaltons. DNA Pol II has two activities: 5' to 3' polymerase activity and 3' to 5' exonuclease activity. It has a low polymerization rate of about 40 nucleotides per second, and it also has a low processivity rate of about 1500 nucleotides. It mainly serves as the alternate DNA repair polymerase. Therefore, it can replicate DNA if the template is damaged.

It does not require ATP for any type of activity. Then, you also have DNA pol III. So, DNA pol III is a primary DNA replicase with the structural gene pol C and has a molecular weight of 900 kilodaltons. It is a multimeric protein complex of 10 different

polypeptides, such as alpha, epsilon, theta, z, tau, y, sigma, and others. It has a high polymerization rate of about 1000 nucleotides per second, and it has high processivity, which is 50,000 nucleotides. Serves as a holoenzyme during its applications, and holoenzyme refers to the multi-protein complex whose catalytic activity is associated with the extra components. Then we have DNA pol III, so this is the structure of DNA pol III, and you can see that all these components are assembled to form this particular enzyme, and there are four essential components.

You have the two copies of the catalytic core, which are the alpha subunit and the epsilon subunit. 5' prime to 3' prime polymerase activity, whereas the epsilon subunit has the 3' prime to 5' prime exonuclease activity, and the theta subunit is for increasing the efficiency of the epsilon subunit. Two copies of the dimerization component, which is the tau, and the two copies of the homodimer of the beta subunit ring for the processivity component, as well as one copy of the clamp loaders, constitute the enzymes; these are the subunits that are present in the clamp loaders. And how it is actually going to perform the DNA polymerization. So, the clamp loader links the two catalytic cores and the two beta clamps increase the possessivity of the DNA pol III holoenzyme, and the loading of clamps is done by the clamp loaders, while the dimer component helps the catalytic cores to function at the same time.

Then we have DNA polymerase 4 and DNA polymerase 5. Both are the Y family polymerases that do not have 3' to 5' exonuclease activity. It has low catalytic efficiency, low processivity, and low fidelity. It is involved in the transition of synthesis and replication damage to DNA by bypassing the nucleotides that can block the progression of a replication fork. And it has been synthesized by the structural gene like the DIN B and the UMU D2C or the pole 5.

Then we also require DNA primase. So, DNA primase is required for the synthesis of RNA primers that are complementary to the strands. Then you also require the DNA ligase to seal the DNA fragment gap into the strands. There are three major events that are going to happen when you have the DNA applications. So, you have the first stage that is called initiation, the second stage which is called elongation, and the third stage that is called termination. So, the first stage is going to have the recognition and initiation of the replication; then, during elongation, the replication fork leads the dNTP synthesis and proofreading, and the third stage is termination, where you will have the cessation of replication.

So, E. coli chromosome DNA is circular with no free end, and it is replicated bidirectionally. Therefore, it resembles the Greek letter Theta. Hence, this replication mode is also known as theta replication. It can be seen in the Gram-negative bacteria such

as Proteobacteria, some commonly used plasmids like Coli E1, RK2, F, and P1 bacteriophage as well. So, this mode of replication, which is called theta replication, is found in the prokaryotic system as well as in plasmids like Col1, RK2F, and P1 bacteriophage.

So, these are some of the steps in the initiation. So, initially, the 2 to 4 DNA-A protein binds to the Nynmer DOR region in the ODC using ATP to perform the initiation complex. Once this is done, it will enter the second stage where the DNA coils around the multiple copies of DNA-A, which leads to topological stress, and once the topological stress is generated. It is actually during the presence of ATP that the DNA A influences the 13 mer ATH DUE region to start melting. And once the DUE region starts melting, the further melting is carried out by the recruiting hexamer protein, which is called helicase or DNA B. DNA-Bay helicase clamps around each of the two single-stranded strands of the DUE site of the original C, and the clamping of DNA-U is supported by the clamp loader DNA-C.

They make the DNA-B and DNA-C complex, and that is how it is going to have the initial melting of the DNA at the original replication; that is how it is actually going to form the replication fork and the replication bubbles. Then, we have the DNA-C, which is going to open up the DNA-B ring and help in placing the ring around the single-stranded DNA at the origin. While DNA-B moves forward with the help of ATP hydrolysis, the single-stranded DNA binding protein covers the single-stranded DNA to prevent unwinding. Single standard DNA binds cooperatively in a sequence-dependent manner and a sequence-independent manner. So single SSBs are actually going to bind the nucleotides, and that is how they are going to destroy the affinity of the complementary strands, and that is how it is going to keep the strands in the single-stranded DNA.

Then we have the next step: DNA B recruits DNA G, which is called RNA primase, to synthesize the RNA primer on both strands, which are called the leading strand and the lagging strand. And then we are going to have the RNA primer, which influences the DNA C to release the DNA B from the site. To initiate the elongation phase, the primersome formation occurs, and as a result, it is actually going to synthesize the primers. And the primersome is actually a functional complex that is going to have the DNAG, DNAB, SSB, and some of the accessory proteins. So, in the initiation, exactly what will happen is that you are actually going to have the binding of some of the components at the DO site and some of the components at the DOR site.

Once these two sites are actually occupied by the initiation factors DNAB and other kinds of proteins. You are going to have the recruitment of the helicase, and helicase is a

hexagonal protein. So, it is actually going to run in both directions. So, it is going to run in this direction, and it is going to run in that direction, and that is how it is actually going to form the replication bubble. And once it is formed, it will allow the binding of the single-stranded DNA binding protein, and that is how it will actually allow the binding of the primase. And that is how there will be a synthesis of the primers, and one fork will run in both directions, and that is how it is actually going to have one strand which is going to be called the leading strand; the other strand is going to be called the lagging strand.

And once this initiation stage is over, it will enter the elongation stage. So, what are the basics of the elongations? So, mainly the DNAPol 3 does not do the polymerizations. Chain elongation happens when the free 3 prime OH group of the primers attacks the alpha phosphoryl group of the incoming TNTPs; as a byproduct of the reaction, the pyrophosphate is generated. The bond formed between known is at the phosphodiester bonds, and that is how it is actually going to start adding the nucleotides based on the base pairing information. So, once you have the Watson-Crick base pairing information, which is going to be available through DNA pol III, it is going to keep adding the nucleotides. And these nucleotides are going to be coupled with each other by the phosphodiester linkage, and the pyrophosphate is going to be released.

This pyrophosphate is going to be hydrolyzed by the pyrophosphatase, which is also going to generate the energy that is going to be utilized in the process. In this case, two replication forks are generated for the prokaryotic cell that move in opposite directions from each other. Replication forks proceed bidirectionally at a speed of 1000 base pairs per second per fork. Both the leading and the lagging strands are replicated simultaneously. What is the leading strand? The DNA polymerase III synthesizes the strand from the 5' to 3' direction continuously towards the replication fork.

So, one fork is running in the direction of the 5 prime to 3 prime, while the other is running in the reverse direction, which is called the lagging strand. So, in the lagging strands, the synthesis happens in the 5 prime to 3 prime direction but discontinuously away from the replication fork. So if you have a bubble, right, and if this is the bubble, right, and if the replication is going in this direction, the machinery is going in this direction. Then this is going to be the leading strand, and this is going to be the lagging strand because it will wait for this region to be available to open, and then only it is actually going to do the synthesis in the reverse direction.

And that's why this is going to be called the lagging strand and this is going to be called the leading strand. So, there are three different stages of elongation, right? So, DNAB helicase separates the two DNA molecules binding to the lagging strand template at the

replication forks and moving along according to the polarity of the 5' to 3' direction. Now, DNAG, the primase associated with DNAB, synthesizes the RNA primers complementary to the associated single-stranded strands. Interaction between DNA B and DNA G regulates the Okazaki fragment length. Tighter association results in more frequent short fragments, whereas the loose interaction will produce a longer lagging strand.

Length of the Okazaki fragments could vary between 1,000 and 2,000 nucleotides. Formation of the replisome. So, it contains the DNA Pol III holoenzyme and the associated proteins like DNA B and DNA J. DNAG. This replisome starts the joining of the dNTPs by forming the phosphodiester bonds. DNApol-1 removes the RNA primers from both strands through its 5' to 3' exonuclease activity, which generates a gap after the primer is removed from both strands.

DNApol-1 also fills in the gap between the lagging strand fragments after the primer is removed. Lastly, the leftover NICs are sealed with the help of DNA ligase and NAD plus as an energy source. So, this is exactly what is going to happen in the replication's elongation stage. So, in the elongation stage, what will happen is that you are going to have the replication fork moving in this direction, right? So, this is going to be considered the leading strand, and this is going to be considered lagging at its end. So, once this portion opens, the synthesis will start from this direction, right? So, it is going to happen in this direction. Now, what is the role of gyrase in elongation? So, due to the unbinding with the help of the DNA helicase, the double-stranded DNA in front of the fork becomes positively supercoiled.

If the supercoiling increases, the fork will halt. Therefore, to overcome the halting from the supercoiling, topoisomerases are needed. In E. coli, the DNA gyrase, a type 2 gyrase, was discovered by a scientist named Martin Gallet. And Gyrase contains two different subunits: GyrA, which cuts and rejoins the DNA, and GyrB, which is responsible for providing the energy through ATP hydrolysis. Then we also have the proofreading activity because when you are doing a DNA application, it is possible that the DNA polymerase could add some nucleotides that may not match, which may not be as per the information available on the template.

And in that case, there is a proofreading activity required because when you are synthesizing a product, it has to go into a testing stage or it is going to get into a stage where you should test whether the product you are synthesizing is of good quality or not. So, the proofreading activity is going to ensure that what is being produced is exactly identical to the template. So, DNA replication is amazingly accurate, with one error in one billion nucleotides incorporated. The above-mentioned tolerated mutant level is

especially desired for the large genome size.

DNA polymerase carries out the process by its 3 prime to 5 prime exonuclease activity. When an inaccurate nucleotide is incorporated, the synthesis rates are reduced due to the wrong positioning of the 3 prime OH. This works like a delete key, removing only the most recent error. So, what happens is that when you are actually synthesizing, it keeps going and keeps checking whether the attachment I have made is actually pairing with the template or not.

So, if it is not, then it is actually going to go back and do the corrections. Now the third process is called as the terminations. As there is bidirectional replication, the fork will melt at a position diametrically opposite to the ODC on the genome. The termination region contains multiple copies of 23-base-pair-long sequences, or TER sequences. Every ter sequence acts as a recognition site for a protein that is called TUS. So, terminate the utilization of substrate proteins. TUS-TER complex allows one replication fork to pass if it is moving in one direction, but blocks the progression if it is formed in the opposite direction.

The directionality depends on the TUS protein localization on the DNA helix. In E. coli, the orientation of the ter-ster complex is such that it ensures both the fork's movement is stopped at or near the same point. After the complete replication process, the two new circles are physically interlocked or catenated. This decatenation is carried out by topoisomerase IV. to generate the two separate double-stranded DNA molecules. So, termination is done by the places where you are going to have the TUR sequences, and these TUR sequences are going to be recognized by a protein called TUS.

All the termination utilization substrate proteins, the TUS, and the TUR are going to form a complex allowing only one replication fork to pass. But if there are two replication forks that are going to reach that particular point and they try to bypass it, then it is actually going to halt the replication. Now let us talk about the summary of the things we have discussed so far, right? So what will be the difference between the replication in prokaryotic and eukaryotic systems? So in the initiation, you are going to have the DNA A, which is actually going to participate.

So you are going to have the initiation reactions where you will have the DNA A. Whereas in prokaryotes, you have the protein called ORC. Then you are also going to have the activity of helicase, which is going to unwind the DNA, and it is going to be DNA B. Then that is done by the MCM complex in the eukaryotic system. Then you are going to have the helicase loader, which is going to be DNA C. Whereas the same function is being done by CDC6 and CDT1.

You do not have to worry about all these proteins because when we talk about eukaryotic applications, we will discuss all of them. Then, the single standard DNA binding proteins are prokaryotic SSBs, whereas in the case of eukaryotes, they are RPA. Then we have the primase, so it is going to be DNA G in the case of the prokaryotic system, whereas in the case of the eukaryotic system, you are going to have the Pol alpha primase. And then we have the polymerase, so you are going to have DNA Pol 3, which is the main polymerase required for replication in the prokaryotic system, whereas in the case of eukaryotes, you are going to have DNA Pol theta, delta, and epsilon.

clamp, you are going to have the beta clamp. Whereas, in the case of the prokaryote, you are going to have the PCNA ring. Then we have the clamp loader, so you are going to have the gamma complex, whereas in the case of the eukaryotes, you are going to have the RFC. Then we have the ligase; ligase is required for, you know, joining the lagging strands, right? So it is going to be a DNA ligase, whereas in the case of eukaryotes, it is going to be DNA ligase 1. Then we require the primase removal. So primase removal is done by DNAPol1 or ribonuclease H in prokaryotes, whereas in eukaryotes it is done by RNase H or Phen1.

So, these are some of the components that are different between the prokaryotic system and the eukaryotic system. And what we have discussed so far is the DNA replication machinery in the prokaryotic system, the origin of replication, features of the origin of replication, and so on. So, in a subsequent lecture, we are going to discuss some more aspects related to prokaryotic replication. Thank you.