# Genetics & molecular biology

◉ Sheet ○ Slide

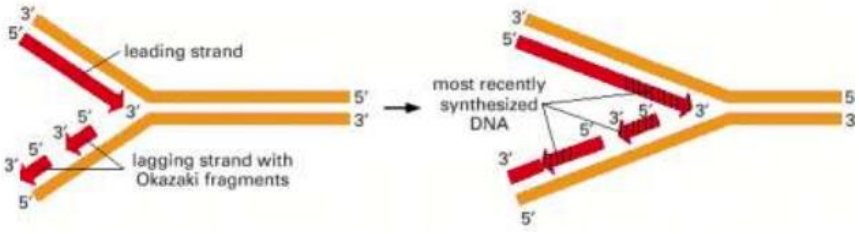
| | |
|---|---|
| Number: | 4 |
| Done by: | Joud Alkarmi |
| Corrected by: | Faisal Nimri |
| Doctor: | Mamoun Ahram |

# DNA replication

Since the two strands of the double helical DNA run in opposite **(antiparallel)** directions, continuous synthesis of two new strands at the replication fork would require that one strand be synthesized in the 5’ to 3’ direction while the other is synthesized in the opposite (3’ to 5’) direction. But DNA polymerase catalyses the synthesis **in the 5’ to 3’ direction only**, how then can the other daughter strand be synthesized?

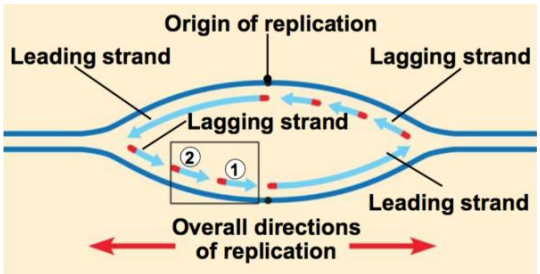
It has been showed that only one strand is synthesized **in a continuous manner** in the 5’ to 3’ direction, and the other is formed of **short, discontinuous** pieces of DNA. These small pieces of newly synthesized DNA are called **Okazaki fragments** which are then joined together by **DNA ligase** forming **the lagging strand**. And the continuously synthesized strand is called **the leading strand**. Moreover, DNA polymerase has to wait for the leading strand to open up the fork further creating enough space for synthesizing another Okazaki fragment on the other strand and that’s why these strands are given their names:

Lagging means slow and dragging; and the synthesis of this strand takes more time to be completed.
Leading means driving and directing; and the elongation of this strand exposes the template used for the synthesis of Okazaki fragments.



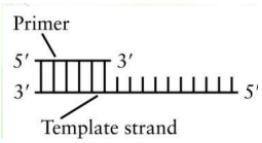

Since DNA replication starts **at the middle of the replication bubble**; this applies for one half of it, what do you think will occur in the other half?

Notice that replication is **bidirectional** and that the template that forms the leading strand in one half forms the lagging strand in the other half of it.




remember that each half of the bubble is called **a replication fork.**

## Molecular components of DNA replication:

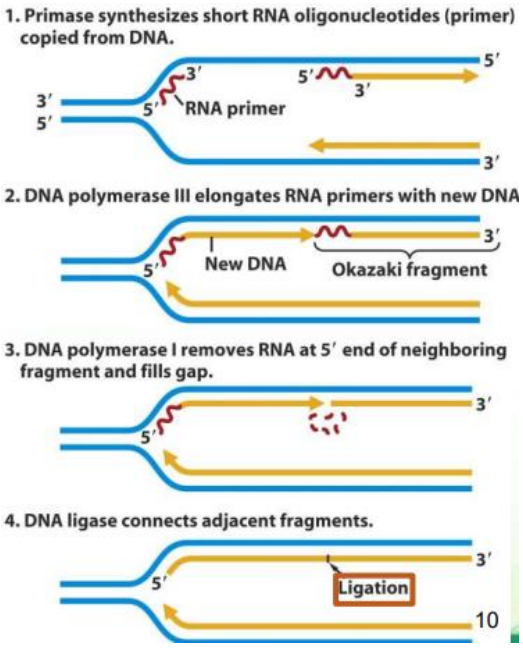



DNA polymerase can't start DNA synthesis de novo (de novo means from scratch); it requires a starting point which is known as **"a primer"** (primer means initiator). So, **DNA primase** synthesizes the primer which is a short RNA strand.

So, after the primase synthesizes a short RNA sequence, **DNA polymerase** comes in and starts the DNA synthesis.

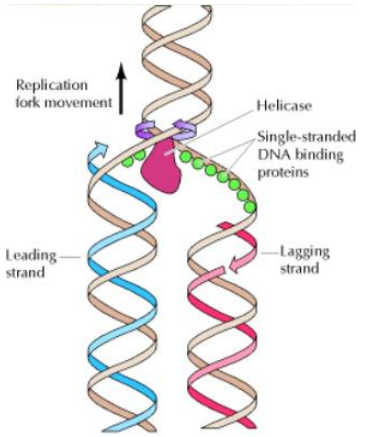
For the leading strand, only one primer is needed. However, each Okazaki fragment requires a primer of its own.

So, DNA polymerase synthesizes an Okazaki fragment until it hits the primer of the following fragment. Then the primer is removed, DNA polymerase completes the Okazaki fragment and then **DNA ligase** comes in and connects the fragments together.




The DNA must be single stranded so that replication can occur, the enzyme required for separating the two strands from each other is known as **DNA helicase**.

In bacteria, the helicases form a complex with the primase called **primosome** (-some means body); so the helicase separates the two strands then the primase synthesizes the primer.




*From slide:* DNA helicases use ATP to open up the double helical DNA as they move along the strands, in other words, they are **ATP dependent**.

However, it's abnormal for cells to have single stranded DNA, as the cells think that it's a viral DNA thus degrading it! So, this ssDNA should be protected. Moreover, when the DNA strands are separated from each other, they can easily come back together.

And that's why single **stranded DNA binding protein "SSB protein"** (specifically called **replication protein A or RPA**) is needed.

**Functions of SSB protein:**

1. Protects the ssDNA from being degraded.
2. Prevents the reformation of dsDNA. In other words; it aids helicases by stabilizing the unwound single stranded conformation.
3. Prevents the formation of the short **hairpin structures**, which occur when there are, in a single stranded DNA, two sequences that are complementary to each other; so they form double stranded regions on the same strand. And when the polymerase hits these double stranded structures it can't move any further so, DNA replication stops.


From slide: SSB proteins bind tightly to exposed ssDNA without covering the bases which remain available for templating.

## DNA polymerases in prokaryotes:

There are different types of polymerases (in E. coli for example):

1. **DNA polymerase I**, which is responsible for:
   - DNA synthesis; as it has a polymerase activity for filling the gaps between the lagging-strand fragments
   - 5' to 3' exonuclease activity (removal of RNA primer) of each Okazaki fragment.
   - DNA repair

   So, we can say that it has a **dual activity**; DNA polymerization and nuclease activity.
2. **DNA polymerase II**
3. **DNA polymerase III,** which is responsible for DNA polymerization at the growing fork.
   Note: The complex of primosome and polymerase is known as **replisome.**
4. **DNA polymerase IV**
5. **DNA polymerase V**

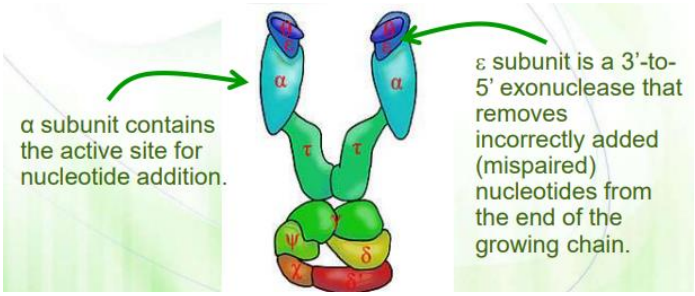
DNA polymerases II, IV, V function in DNA repair.

As we said before, DNA polymerase III synthesizes the Okazaki fragments, then it falls off when it hits the following primer. So, DNA polymerase I uses its exonuclease activity to remove the RNA primer, one nucleotide at a time, replacing it with a deoxynucleotide through its polymerase activity.

10:00

DNA polymerase III is a very large protein composed of **10 subunits** (named α, β, γ, δ…) with different functions. We will focus on two subunits:

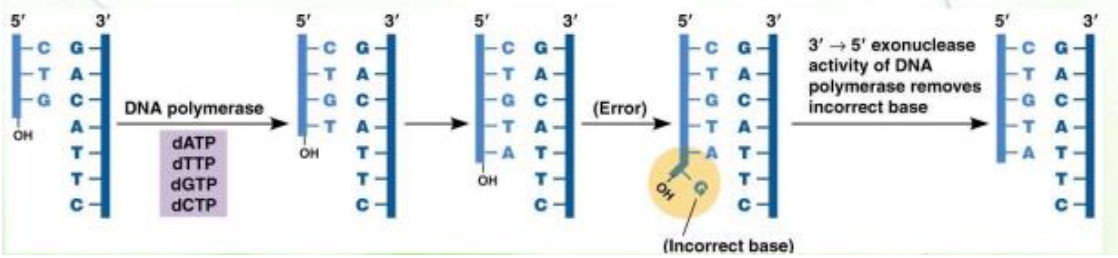
1. **Alpha α subunit**: which is responsible for DNA polymerization (synthesis) as it contains the active site for nucleotide addition from 5' to 3'.
2. **Epsilon ε subunit**: which is a 3' to 5' exonuclease that removes incorrectly added (mispaired) nucleotides from the end of the growing chain.
   So, ε subunit is important for the accuracy of the DNA. DNA synthesis must be accurate, otherwise, mutations occur causing cell death.



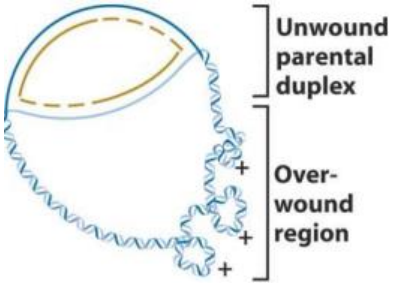

**Accuracy of DNA synthesis is achieved by two mechanisms:**

1. **The enzymatic activity is accurate itself** since the DNA polymerase can catalyse the formation of phosphodiester bonds when the right hydrogen bonding takes place between the bases. In other words; when there is an A it adds a T and when there is a C it adds a G and so on because its active site can only fit the right complementary base when it's bound to a specific base.
   Here, the accuracy is **1/1000** →for each 1000 nucleotide added, one is wrong. However, this is a large number when taking the whole genome, so there must be another mechanism.
2. **Proofreading mechanism:** (a 3'→5' exonuclease activity)- Remember ε subunit of DNA polymerase III.
   By this mechanism, when the enzyme adds a wrong nucleotide, it senses the error, so it moves a step backward (3' to 5' direction), removes the incorrect base and places the right one, then it continues moving forward.

That's why the frequency of errors during replication is only **one incorrect base per $10^8$ nucleotides incorporated.**

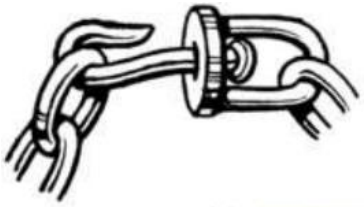


When the DNA replicates separating the two strands from each other, knots will be formed (overwound regions of DNA). So, the DNA will be tangled up and thus DNA synthesis can't move forward. Luckily, this is not what really happens due to the presence of another enzyme called **topoisomerase**.

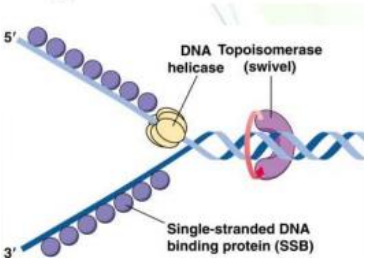



Remember that an isomerase is an enzyme that catalyses the conversion of a specified compound from one structure to another. So, topoisomerase is an enzyme that changes the topology of the DNA; altering the supercoiled form of the DNA. And there are several types of topoisomerases.

**DNA topoisomerase I:**



- Doesn't require energy **(ATP independent)**
- It creates **a swivel**, so that if the DNA in one side of the swivel is rotated the DNA in the other side won't be affected.
- So, it produces **a transient single strand break (nick)** in one of the strands as it breaks then reforms phosphodiester bonds in a DNA strand.



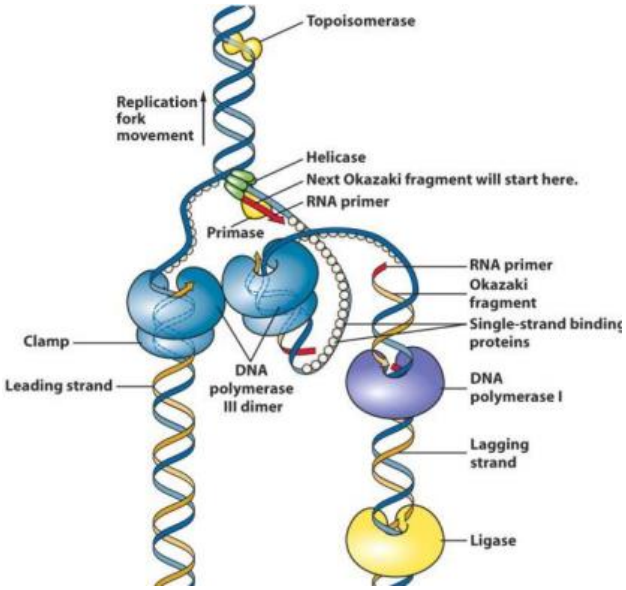

## To summarize the DNA replication machinery:

The primase is associated with helicase in E. coli, separating the two strands from each other and forming a primer.

At the same time, replication protein A prevents the reformation of dsDNA, prevents the reformation of dsDNA and formation of hairpin structures and protects the ssDNA from degradation.

When the primer is formed, DNA pol III starts DNA synthesis until it hits the primer of the Okazaki fragment in front of it, so it falls off. Then, DNA pol I uses its exonuclease activity to remove the primer and it adds DNA instead of it until it hits the Okazaki fragment in front of it, it stops and then the DNA ligase connects the Okazaki fragments with each other, forming the lagging strand.

Quite ahead of all of these actions, topoisomerase I creates a single stranded break called nick preventing the formation of overwound DNA right ahead of the complex.




## Where does DNA synthesis start?

It starts at a certain point, not randomly, this point is known as **the origin of replication (OriC)**. The same concept exists in plasmids, since we took that plasmids should be able to replicate in order to be used as vectors, so these plasmids also have their own OriC.

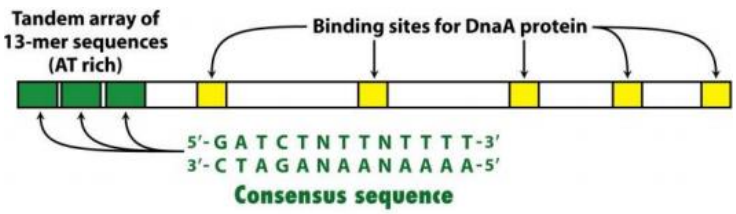
Both plasmids and bacterial DNA contain **only one** OriC for each.

So, what is special for OriC?

It contains **consensus sequences** (consensus means general agreement) which are sequences of DNA that are found in different organisms (different bacterial species and also the same sequence might be found in human's genome for example). These consensus sequences must be preserved.

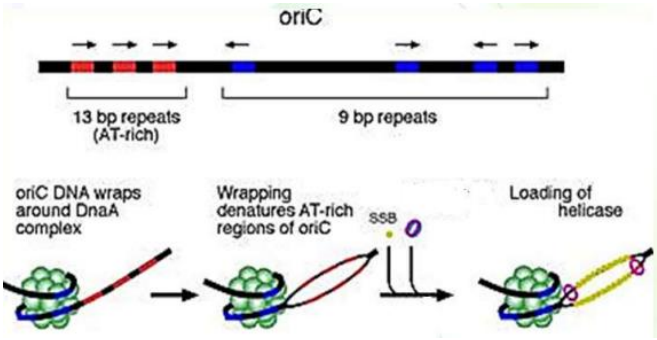
Note: the amino acid sequence of histones that are found in cows differs from the sequence of histones in humans by only two amino acids. This indicates how important histones are; and therefore they mustn't be subjected to evolutionary changes.

20:00



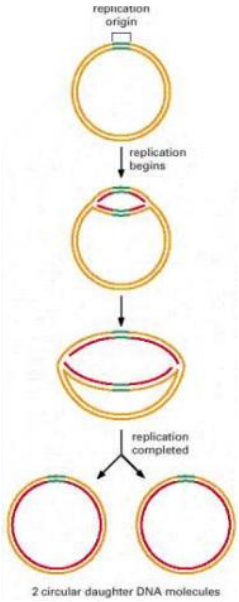
**There are two types of consensus sequences:**

1. **13-mer:**
   - composed of 13 base pairs
   - they are **tandem organized** which means that a 13-mer is followed by another 13-mer and so on …
   - **AT rich**; so, the binding is somehow weak (there are only two hydrogen bonds between each A and T), and this facilitates separation of dsDNA.
2. **9-mer:**
   - Composed of 9 base pairs
   - They are **dispersed**; as they exist in the OriC but they are separated by different sequences. So, if the sequence of any 9-mer is changed, DNA replication can't start. On the other hand, there won't be a problem if any of the sequences between two 9-mers is changed.
   - **Binding sites for DnaA protein:**
     When this protein binds to 9-mers in bacteria, the OriC wraps around the DnaA complex, which squeezes on the DNA applying stress on the AT rich region, therefore resulting in its denaturation and so it becomes a single stranded DNA. After that, other DNA replication machinery proteins (helicase, SSB proteins, polymerase…) can jump on the DNA and start DNA synthesis.

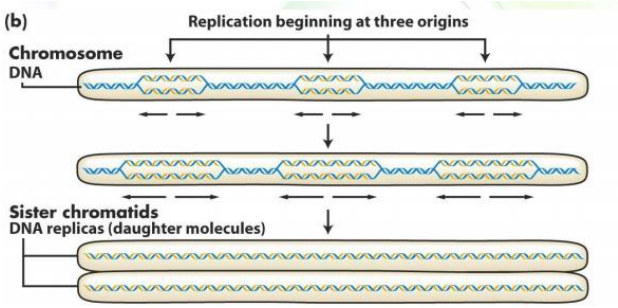



For better understanding: the mechanism by which DnaA protein works is similar to how a long balloon gets larger at one side when squeezing it at the other side.

DNA replication starts at the OriC of bacteria and proceeds bidirectionally (in opposite directions) until they meet up in a point in the middle (halfway around the chromosome) and the two chromosomes get separated from each other.

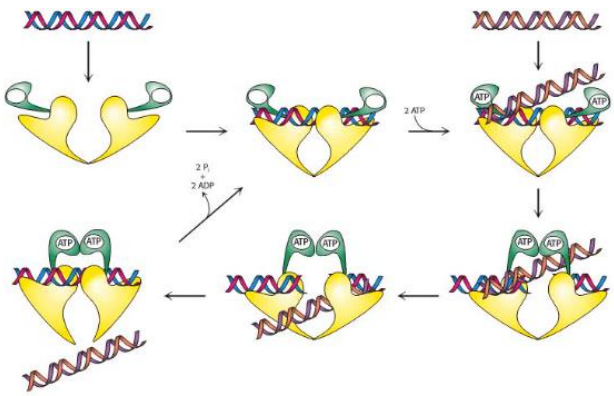



An average human chromosome has several hundred replicators (origins of replications) since it would take forever to synthesize and replicate the DNA if only one OriC was found. These replicators start **all at the same time** and eventually they meet halfway and the chromosomes get separated from each other.




However, the process is not that simple, since another protein called **topoisomerase II** is needed to separate the two chromosomes from each other.

Since during DNA replication the chromosomes are tangled up (in both human cells and bacterial cells), topoisomerase II is responsible for untangling chromosomes by creating a nick on one strand and another nick on the other strand **(transient double strand break)**, separating the chromosomes from each other.

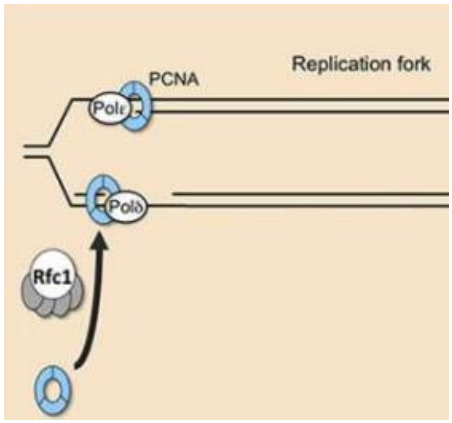



This protein is ATP dependent and it's known as **DNA gyrase** in bacteria.
Moreover, DNA topoisomerase II is responsible for **chromosome condensation** during cell cycle, since normally, before cell division, the DNA is found as **chromatin** which is the complex of DNA and histones by which chromosomes are packaged together, so they must be separated from each other and the chromosome must be condensed (to look like X-shape).

We can conclude that, if these DNA topoisomerases are inhibited, cell division will stop. So, **topoisomerase II inhibitors** are commonly used as drugs to target cells that are proliferating, specifically cancer cells. In other words, they are used in cancer therapy.

**PCNA protein "proliferating cell nuclear antigen protein":**

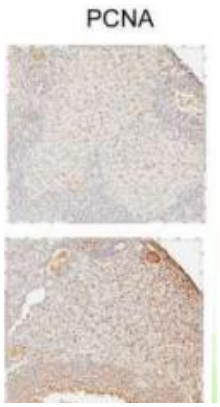
A protein that guides DNA polymerase to the primers.




Since during cellular division and DNA replication each Okazaki fragment has a primer, then high level of PCNA proteins is needed; that's why it's used as a marker for proliferating cells.

For example, when deciding whether a tumour is aggressive or not, we look for proliferating markers, if their level is high, then the cells are actively dividing and so they give an indication of how bad the cancer is.

Here, the PCNA protein is stained by a brown stain:




In the first picture, the brown colour is almost not seen so the cells are not proliferating.
However, in the second picture, the stain is dense, so the cells are actively synthesizing DNA and dividing. And if we were talking about a tumour then it would be an aggressive one.
Notice also that staining is more intense at the periphery because cell division is more active there.

## DNA polymerase in eukaryotic cells:

Eukaryotic cells contain **9 DNA polymerases;** most of them for DNA repair. And they are called (α, β, γ, δ…).

Remember that in bacteria DNA pol III is made up of several subunits called (α, β, γ, δ…) but in eukaryotic cells the polymerases themselves are called (α, β, γ, δ…) despite that they are also made up of multiple subunits which are given other names.

**We will focus only on four of the polymerases:**

1. **Gamma γ polymerase:**
   A mitochondrial protein that is responsible for the synthesis of mitochondrial DNA.
   Remember that mitochondrial DNA is circular, same as bacterial DNA and this supports the endosymbiotic theory.
2. **Alpha α polymerase**

3. **Delta δ polymerase**
4. **Epsilon ε polymerase**

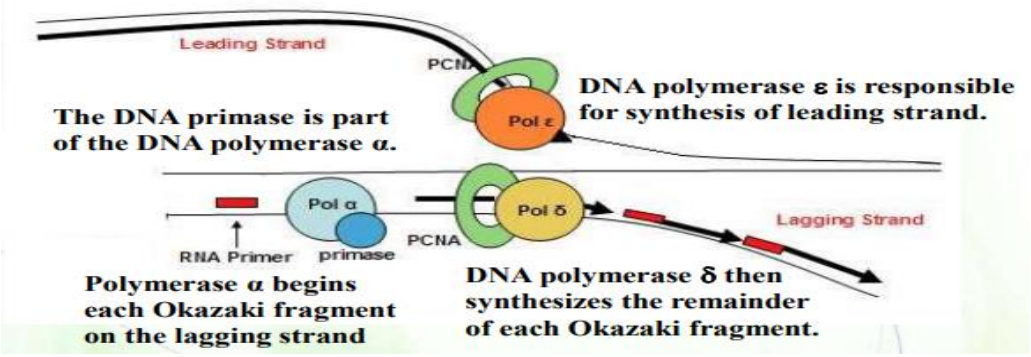
Important notes on these four DNA polymerases:

- **All α, δ, ε polymerases have DNA polymerase activity**; so all are responsible for DNA synthesis
- **The primase is associated with α polymerase only**; so it is responsible for initiating the DNA synthesis.
- **α polymerase doesn't have 3' to 5' exonuclease activity;** so it doesn't have a proofreading mechanism.
- **δ, γ and ε polymerases have 3' to 5' exonuclease activity**
- **For the processivity:** α polymerase has low processivity while δ and ε polymerases have high processivity (processivity refers to velocity).
- **Fidelity (related to the accuracy) is high for all of them.**

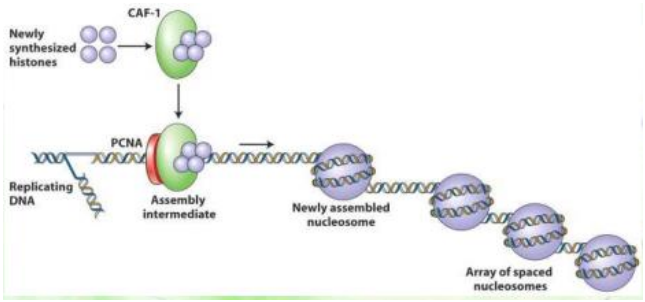
**The Biochemical Properties of Eukaryotic DNA Polymerases**

| | α | δ | ε | β | γ |
|---|---|---|---|---|---|
| Mass (kDa) | | | | | |
| Native | >250 | 170 | 256 | 36-38 | 160-300 |
| Catalytic core | 165-180 | 125 | 215 | 36-38 | 125 |
| Other subunits | 70, 50, 60 | 48 | 55 | None | 35, 47 |
| Location | Nucleus | Nucleus | Nucleus | Nucleus | Mitochondria |
| Associated functions | | | | | |
| 3' → 5' exonuclease | No | Yes | Yes | No | Yes |
| Primase | Yes | No | No | No | No |
| Properties | | | | | |
| Processivity | Low | High | High | Low | High |
| Fidelity | High | High | High | Low | High |
| Replication | Yes | Yes | Yes | No | Yes |
| Repair | No | ? | Yes | Yes | No |

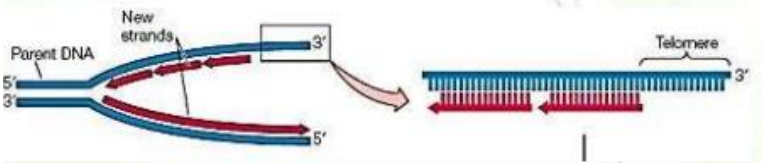
So, the α polymerase adds the primer and then synthesizes a short part of the DNA before falling off due to its low processivity. Then, δ and ε polymerases jumps up to the DNA, where δ polymerase synthesizes the Okazaki fragments and ε polymerase synthesizes the leading strand. They continue until they hit the primers in front of them and then special enzymes (their names are not important) remove the primers, and δ polymerase fills the gaps (almost similar to DNA polymerase I activity in bacteria).



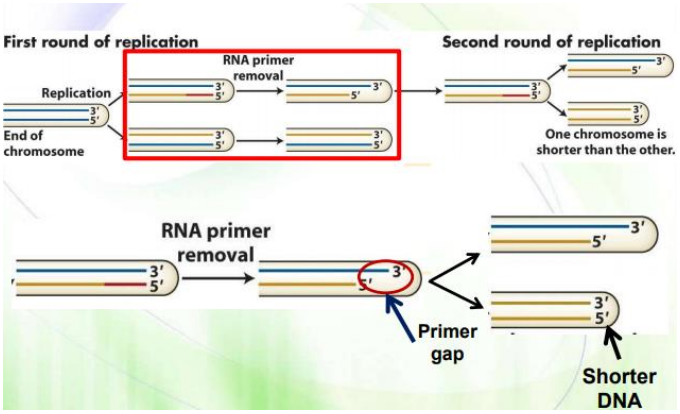
Chromatin is a combination of DNA plus proteins, however, DNA polymerase can't move along the strand when there are histones on the eukaryotic DNA (not found in prokaryotic). So, histones must be removed in order for the replication process to begin and this is the function of **chromatin remodelling proteins CRP.** At the same time, the newly synthesized DNA has to be packaged again so histones should be reassembled and this is done by the **chromatin assembly factor CAF**; which wraps the DNA behind each replication fork around new histones.
So, DNA synthesis and packaging occur **simultaneously.**




Now, unlike bacterial chromosomes, the chromosomes of eukaryotes are linear, meaning that they have ends. These ends pose a problem for DNA replication. The DNA at the end of the chromosome cannot be fully copied in each round of replication, resulting in a slow, gradual shortening of the chromosome.




Why is this case? When DNA is being copied, the **leading strand** is made continuously. However, the **lagging strand** is produced from many Okazaki fragments, each of which begins with its own RNA primer. And in most cases, the primers of the Okazaki fragments can be easily replaced with DNA and the fragments are connected to form an unbroken strand. However, when the replication fork reaches the end of the chromosome, a short stretch of DNA does not get covered by an Okazaki fragment—essentially, since there's no way to get the



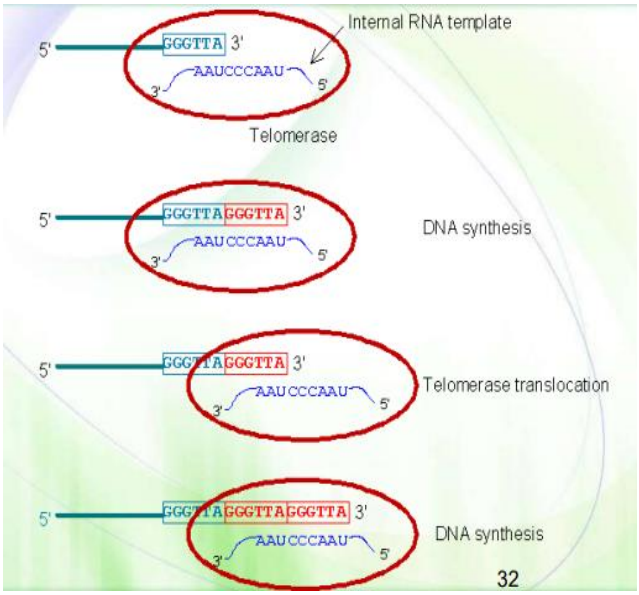
fragment started because the primer would fall beyond the chromosome end. Also, the primer of the last Okazaki fragment made can't be replaced with DNA like other primers.

This problem is solved by an enzyme called **telomerase** which is a **ribonucleoprotein** that is composed of a protein component plus an RNA component.

40:00

So, its RNA molecule act as a template that is used to elongate the DNA, in other words; it synthesizes DNA out of RNA, functioning **as a reverse transcriptase**.

So, the telomerase sits at the end of the DNA and uses its template to synthesize DNA in **the 5' to 3' direction** and it keeps synthesizing the same sequence over and over again since its template remains the same.




The sequence is **"GGGTTA"** and it is repeated thousands of times in the telomeres at the end of the chromosomes.

That's how the ends of the chromosomes are preserved in order to have enough space for the primase to add a primer for DNA synthesis, thus the length of the DNA is kept long enough; stabilizing the chromosome.

Please refer to the animation in the video lecture for better understanding. But notice that there are two mistakes in it:

1. In the upper lagging strand, delta polymerase is the one responsible of synthesizing Okazaki fragments, alpha just adds the primer and a short DNA sequence!
2. In the lower leading strand, epsilon polymerase is the one responsible of synthesizing this strand, not delta!

**So, why do we age?**

When we are young, the telomerase activity is high and its level is also high. But as we grow older, the activity and level of telomerase are reduced. So, the gradual shortening of the chromosomes decreases its stability, DNA gets broken up which leads to cell death.

It has even been suggested that life span can be determined by the length of telomeres and that's why telomerase is called **"elixir of youth"**.

Also remember that the level of telomerase is high in cancer cells so they keep dividing and thus they are immortal!

**Last thing I promise; do you remember Dolly the sheep?**

Dolly the sheep, is the first mammal to be cloned from an adult cell. A sheep like Dolly has a life expectancy of 11-12 years, but Dolly lived for 6.5 years and signs of aging like organ failure and slow activity appeared early on it, why?

The DNA they have used for cloning was 5 years old so, the DNA was already old with telomeres shorter than those found in other animals of similar age to Dolly, telomeres got shorter and shorter leading to DNA instability and death!

---

*Because Neveen wanted her name to be written somewhere; she wishes you a good luck ☺*

---