MOLECULAR BIOLOGY

Dr.J.P.Patel

DNA REPLICATION

KEY POINTS:

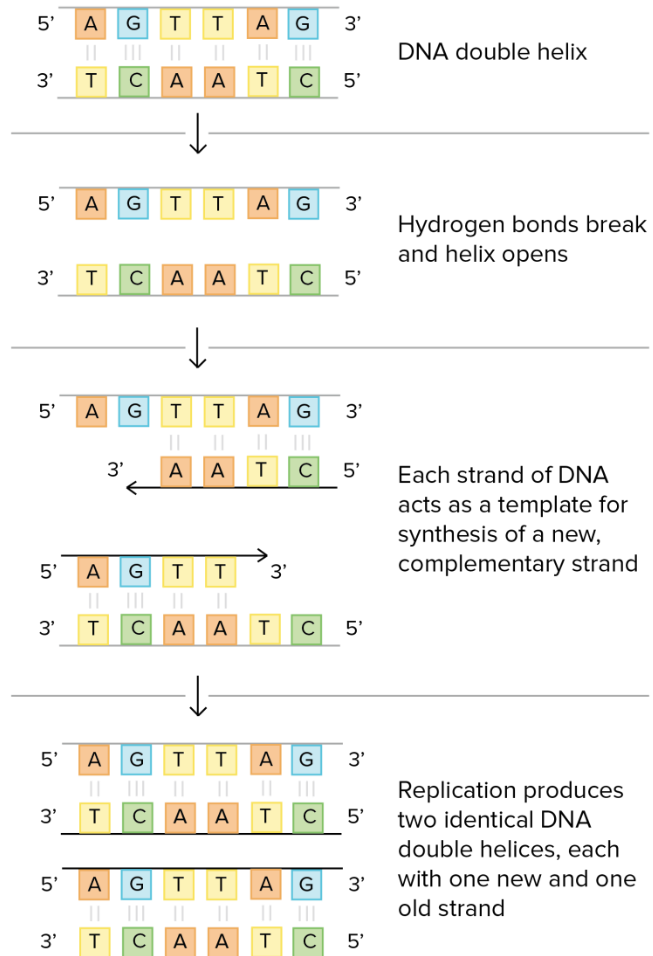
- DNA replication is **semiconservative.** Each strand in the double helix acts as a template for synthesis of a new, complementary strand.
- New DNA is made by enzymes called **DNA polymerases**, which require a template and a **primer** (starter) and synthesize DNA in the 5' to 3' direction.
- During DNA replication, one new strand (the **leading strand**) is made as a continuous piece. The other (the **lagging strand**) is made in small pieces.
- DNA replication requires other enzymes in addition to DNA polymerase, including **DNA primase**, **DNA helicase**, **DNA ligase**, and **topoisomerase**.

INTRODUCTION

- **DNA replication**, or the copying of a cell's DNA, is no simple task! There are about 333 billion base pairs of DNA in your genome, all of which must be accurately copied when any one of your trillions of cells divides.
- The basic mechanisms of DNA replication are similar across organisms. In this article, we'll focus on DNA replication as it takes place in the bacterium *E. coli*, but the mechanisms of replication are similar in humans and other eukaryotes.
- Let's take a look at the proteins and enzymes that carry out replication, seeing how they work together to ensure accurate and complete replication of DNA.

THE BASIC IDEA

- DNA replication is **semiconservative**, meaning that each strand in the DNA double helix acts as a template for the synthesis of a new, complementary strand.
- This process takes us from one starting molecule to two "daughter" molecules, with each newly formed double helix containing one new and one old strand

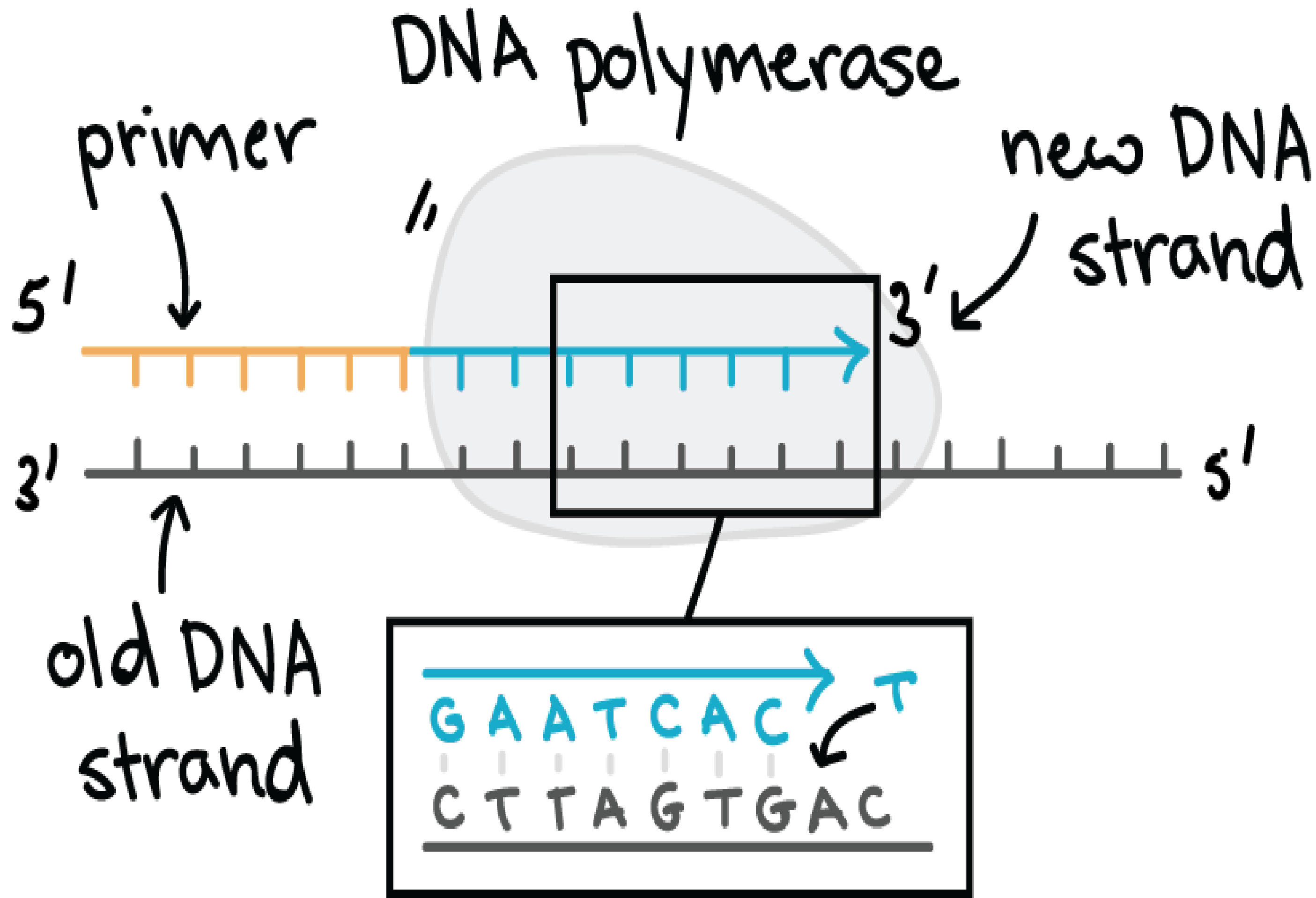


5’
A
G
T
T
A
G
3’
3’
T
C
A
A
T
C
5’
DNA double helix
Hydrogen bonds break and helix opens
Each strand of DNA acts as a template for synthesis of a new, complementary strand
Replication produces two identical DNA double helices, each with one new and one old strand

- In a sense, that's all there is to DNA replication! But what's actually most interesting about this process is how it's carried out in a cell.
- Cells need to copy their DNA very quickly, and with very few errors (or risk problem such as cancer). To do so, they use a variety of enzymes and proteins, which work together to make sure DNA replication is performed smoothly and accurately.

DNA POLYMERASE

- One of the key molecules in DNA replication is the enzyme **DNA polymerase**. DNA polymerases are responsible for synthesizing DNA: they add nucleotides one by one to the growing DNA chain, incorporating only those that are complementary to the template.


DNA polymerase
primer
new DNA strand
5'
3'
3'
5'
old DNA strand
GAATCAC
T
CTTAGTGAC

- Here are some key features of DNA polymerases:
- They always need a template
- They can only add nucleotides to the 3' end of a DNA strand
- They can't start making a DNA chain from scratch, but require a pre-existing chain or short stretch of nucleotides called a **primer**
- They **proofread**, or check their work, removing the vast majority of "wrong" nucleotides that are accidentally added to the chain

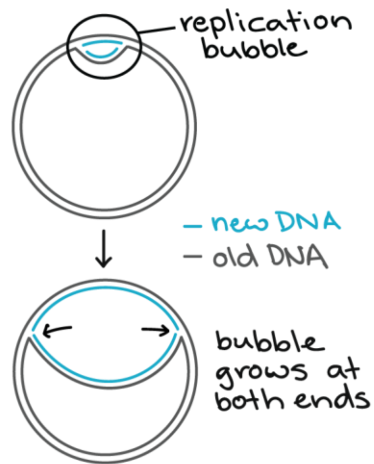
- The addition of nucleotides requires energy. This energy comes from the nucleotides themselves, which have three phosphates attached to them (much like the energy-carrying molecule ATP). When the bond between phosphates is broken, the energy released is used to form a bond between the incoming nucleotide and the growing chain.

- In prokaryotes such as *E. coli*, there are two main DNA polymerases involved in DNA replication: DNA pol III (the major DNA-maker), and DNA pol I, which plays a crucial supporting role we'll examine later.

STARTING DNA REPLICATION

- How do DNA polymerases and other replication factors know where to begin? Replication always starts at specific locations on the DNA, which are called **origins of replication** and are recognized by their sequence.
- *E. coli*, like most bacteria, has a single origin of replication on its chromosome. The origin is about 245245245 base pairs long and has mostly A/T base pairs (which are held together by fewer hydrogen bonds than G/C base pairs), making the DNA strands easier to separate.

- Specialized proteins recognize the origin, bind to this site, and open up the DNA. As the DNA opens, two Y-shaped structures called **replication forks** are formed, together making up what's called a **replication bubble**. The replication forks will move in opposite directions as replication proceeds.

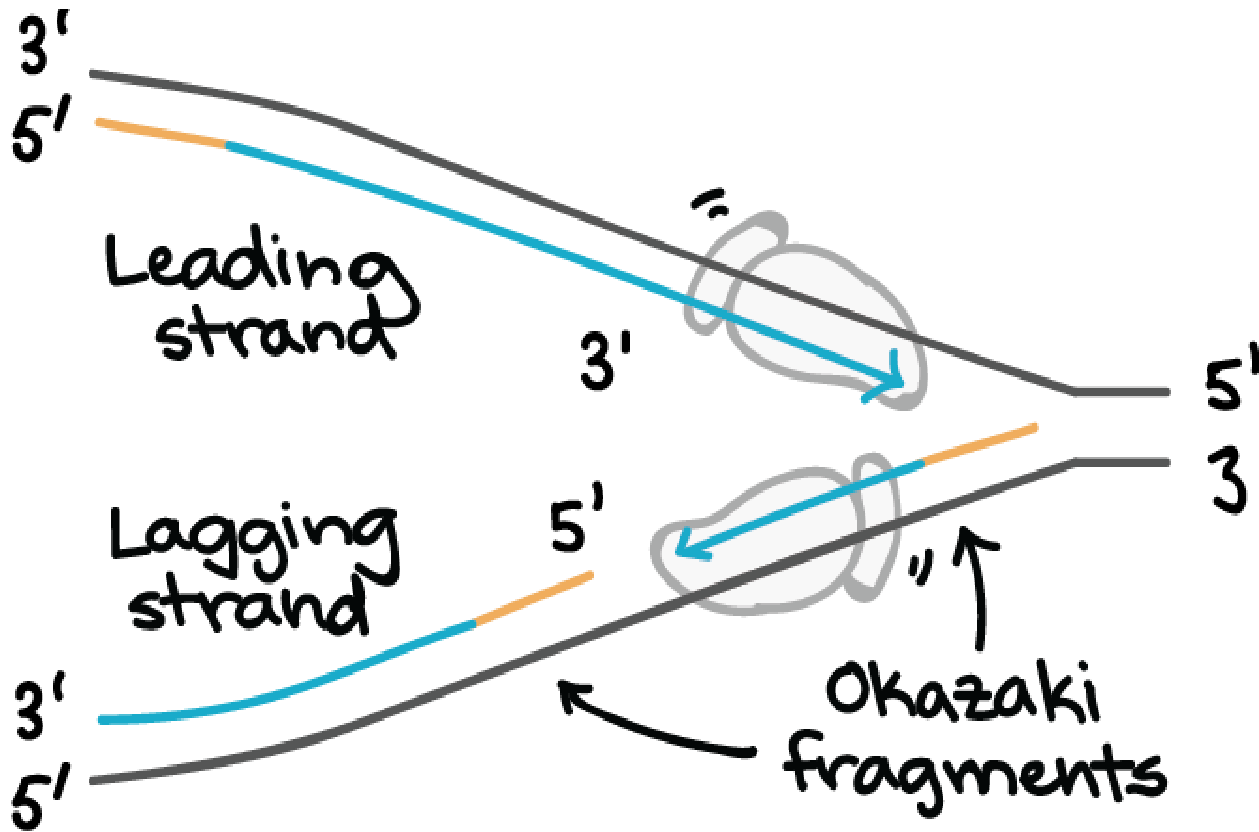

replication bubble
new DNA
old DNA
bubble grows at both ends

- How does replication actually get going at the forks? **Helicase** is the first replication enzyme to load on at the origin of replication^33cubed. Helicase's job is to move the replication forks forward by "unwinding" the DNA (breaking the hydrogen bonds between the nitrogenous base pairs).
- Proteins called **single-strand binding proteins** coat the separated strands of DNA near the replication fork, keeping them from coming back together into a double helix.

- **Primers and primase**
- DNA polymerases can only add nucleotides to the 3' end of an existing DNA strand. (They use the free -OH group found at the 3' end as a "hook," adding a nucleotide to this group in the polymerization reaction.) How, then, does DNA polymerase add the first nucleotide at a new replication fork?
- Alone, it can't! The problem is solved with the help of an enzyme called **primase**. Primase makes an RNA **primer**, or short stretch of nucleic acid complementary to the template, that provides a 3' end for DNA polymerase to work on. A typical primer is about five to ten nucleotides long. The primer *primes* DNA synthesis, i.e., gets it started.
- Once the RNA primer is in place, DNA polymerase "extends" it, adding nucleotides one by one to make a new DNA strand that's complementary to the template strand.

LEADING AND LAGGING STRANDS

- In *E. coli*, the DNA polymerase that handles most of the synthesis is DNA polymerase III. There are two molecules of DNA polymerase III at a replication fork, each of them hard at work on one of the two new DNA strands.
- DNA polymerases can only make DNA in the 5' to 3' direction, and this poses a problem during replication. A DNA double helix is always anti-parallel; in other words, one strand runs in the 5' to 3' direction, while the other runs in the 3' to 5' direction. This makes it necessary for the two new strands, which are also antiparallel to their templates, to be made in slightly different ways.



3'
5'
Leading strand
3'
5'
3
Lagging strand
5'
3'
5'
Okazaki fragments

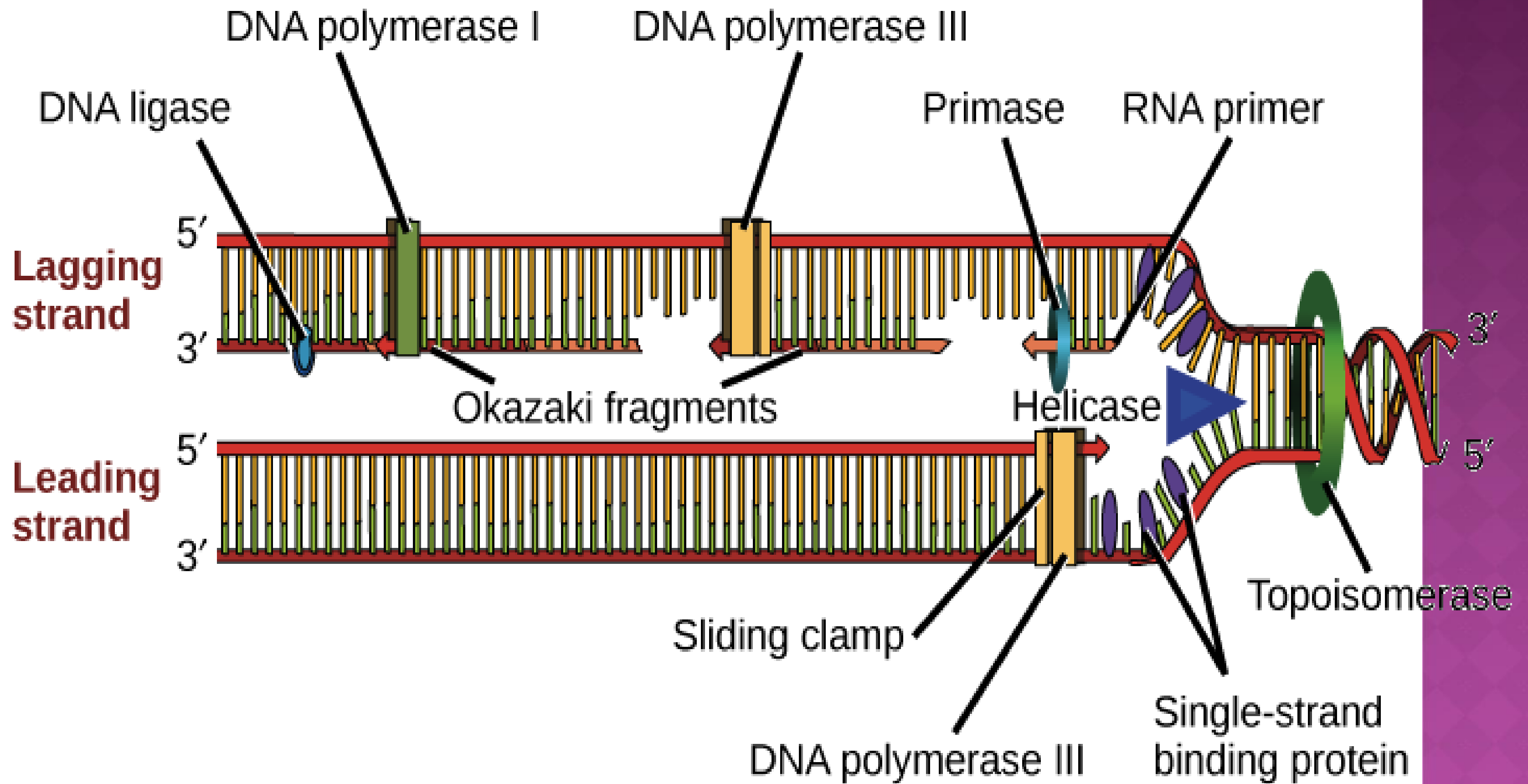
- One new strand, which runs 5' to 3' towards the replication fork, is the easy one. This strand is made continuously, because the DNA polymerase is moving in the same direction as the replication fork. This continuously synthesized strand is called the **leading strand**.

- The other new strand, which runs 5' to 3' away from the fork, is trickier. This strand is made in fragments because, as the fork moves forward, the DNA polymerase (which is moving away from the fork) must come off and reattach on the newly exposed DNA. This tricky strand, which is made in fragments, is called the **lagging strand**.
- The small fragments are called **Okazaki fragments**, named for the Japanese scientist who discovered them. The leading strand can be extended from one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments.

THE MAINTENANCE AND CLEANUP CREW

- Some other proteins and enzymes, in addition the main ones above, are needed to keep DNA replication running smoothly. One is a protein called the **sliding clamp**, which holds DNA polymerase III molecules in place as they synthesize DNA. The sliding clamp is a ring-shaped protein and keeps the DNA polymerase of the lagging strand from floating off when it re-starts at a new Okazaki fragment^44start superscript, 4, end superscript.
- **Topoisomerase** also plays an important maintenance role during DNA replication. This enzyme prevents the DNA double helix ahead of the replication fork from getting too tightly wound as the DNA is opened up. It acts by making temporary nicks in the helix to release the tension, then sealing the nicks to avoid permanent damage.
- Finally, there is a little cleanup work to do if we want DNA that doesn't contain any RNA or gaps. The RNA primers are removed and replaced by DNA through the activity of **DNA polymerase I**, the other polymerase involved in replication. The nicks that remain after the primers are replaced get sealed by the enzyme **DNA ligase**.

SUMMARY OF DNA REPLICATION IN *E. COLI*



- **Helicase** opens up the DNA at the replication fork.
- **Single-strand binding proteins** coat the DNA around the replication fork to prevent rewinding of the DNA.
- **Topoisomerase** works at the region ahead of the replication fork to prevent supercoiling.
- **Primase** synthesizes RNA primers complementary to the DNA strand.
- **DNA polymerase III** extends the primers, adding on to the 3' end, to make the bulk of the new DNA.
- RNA primers are removed and replaced with DNA by **DNA polymerase I**.
- The gaps between DNA fragments are sealed by **DNA ligase**.

DNA REPLICATION IN EUKARYOTES

- The basics of DNA replication are similar between bacteria and eukaryotes such as humans, but there are also some differences:
- Eukaryotes usually have multiple linear chromosomes, each with multiple origins of replication. Humans can have up to 100,100,100, comma000000000 origins of replication^55start superscript, 5, end superscript!
- Most of the *E. coli* enzymes have counterparts in eukaryotic DNA replication, but a single enzyme in *E. coli* may be represented by multiple enzymes in eukaryotes. For instance, there are five human DNA polymerases with important roles in replication^55start superscript, 5, end superscript.
- Most eukaryotic chromosomes are linear. Because of the way the lagging strand is made, some DNA is lost from the ends of linear chromosomes (the telomeres) in each round of replication.

MODE OF DNA REPLICATION: MESELSON-STAHL EXPERIMENT

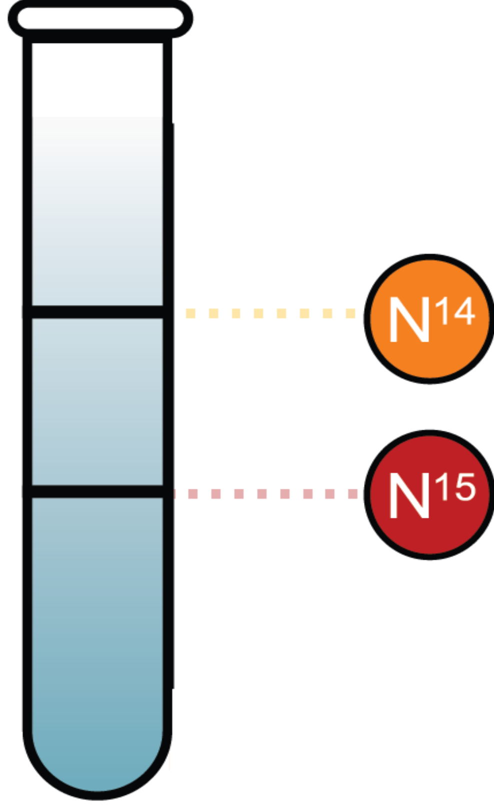
- **Key points:**
- There were three models for how organisms might replicate their DNA: semi-conservative, conservative, and dispersive.
- The **semi-conservative** model, in which each strand of DNA serves as a template to make a new, complementary strand, seemed most likely based on DNA's structure.
- The models were tested by Meselson and Stahl, who labeled the DNA of bacteria across generations using isotopes of nitrogen.
- From the patterns of DNA labeling they saw, Meselson and Stahl confirmed that DNA is replicated semi-conservatively.

THE THREE MODELS FOR DNA REPLICATION

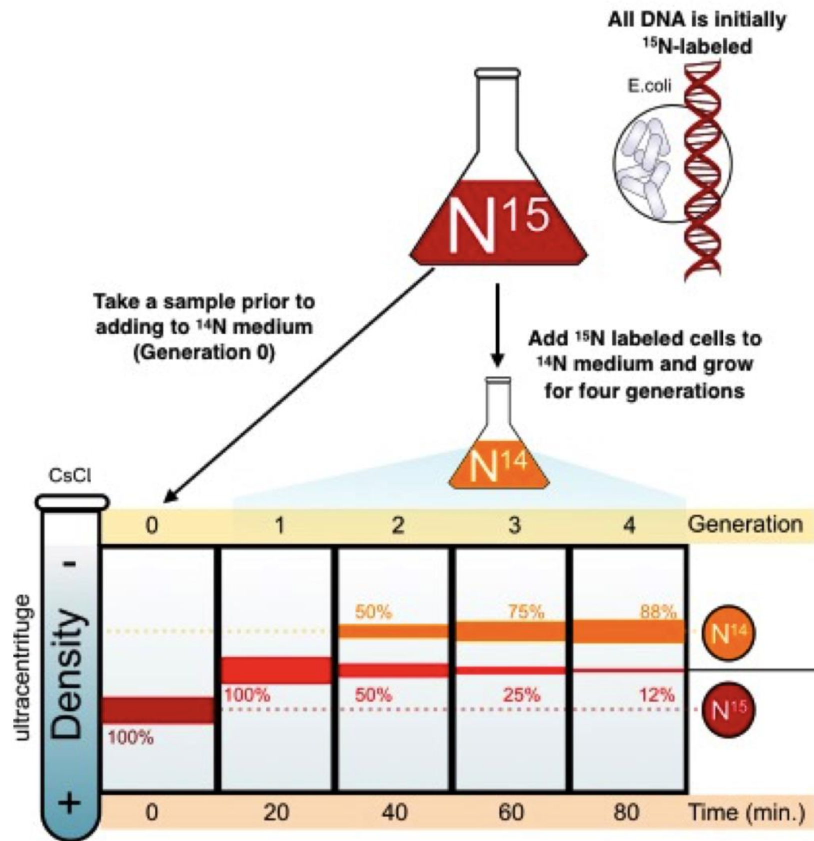
- **Semi-conservative replication.** In this model, the two strands of DNA unwind from each other, and each acts as a template for synthesis of a new, complementary strand. This results in two DNA molecules with one original strand and one new strand.
- **Conservative replication.** In this model, DNA replication results in one molecule that consists of both original DNA strands (identical to the original DNA molecule) and another molecule that consists of two new strands (with exactly the same sequences as the original molecule).
- **Dispersive replication.** In the dispersive model, DNA replication results in two DNA molecules that are mixtures, or “hybrids,” of parental and daughter DNA. In this model, each individual strand is a patchwork of original and new DNA.

- **The Meselson-Stahl experiment**
- Meselson and Stahl conducted their famous experiments on DNA replication using *E. coli* bacteria as a model system.
- They began by growing *E. coli* in medium, or nutrient broth, containing a "heavy" isotope of nitrogen, ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text. (An **isotope** is just a version of an element that differs from other versions by the number of neutrons in its nucleus.) When grown on medium containing heavy ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text, the bacteria took up the nitrogen and used it to synthesize new biological molecules, including DNA.
- After many generations growing in the ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text medium, the nitrogenous bases of the bacteria's DNA were all labeled with heavy ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text. Then, the bacteria were switched to medium containing a "light" ^{14}\text N14Nstart superscript, 14, end superscript, start text, N, end text isotope and allowed to grow for several generations. DNA made after the switch would have to be made up of ^{14}\text N14Nstart superscript, 14, end superscript, start text, N, end text, as this would have been the only nitrogen available for DNA synthesis.

- Meselson and Stahl knew how often *E. coli* cells divided, so they were able to collect small samples in each generation and extract and purify the DNA. They then measured the density of the DNA using **density gradient centrifugation**.
- This method separates molecules such as DNA into bands by spinning them at high speeds in the presence of another molecule, such as cesium chloride, that forms a density gradient from the top to the bottom of the spinning tube. Density gradient centrifugation allows very small differences—like those between N14 andN15labeled DNA—to be detected.



Density
-
+
N^{14}
N^{15}



All DNA is initially
15N-labeled
E.coli
N15
Take a sample prior to
adding to 14N medium
(Generation 0)
Add 15N labeled cells to
14N medium and grow
for four generations
N14
CsCl
0
1
2
3
4
Generation
ultracentrifuge
Density
-
+
50%
75%
88%
100%
50%
25%
12%
100%
N14
N15
0
20
40
60
80
Time (min.)

GENERATION 0

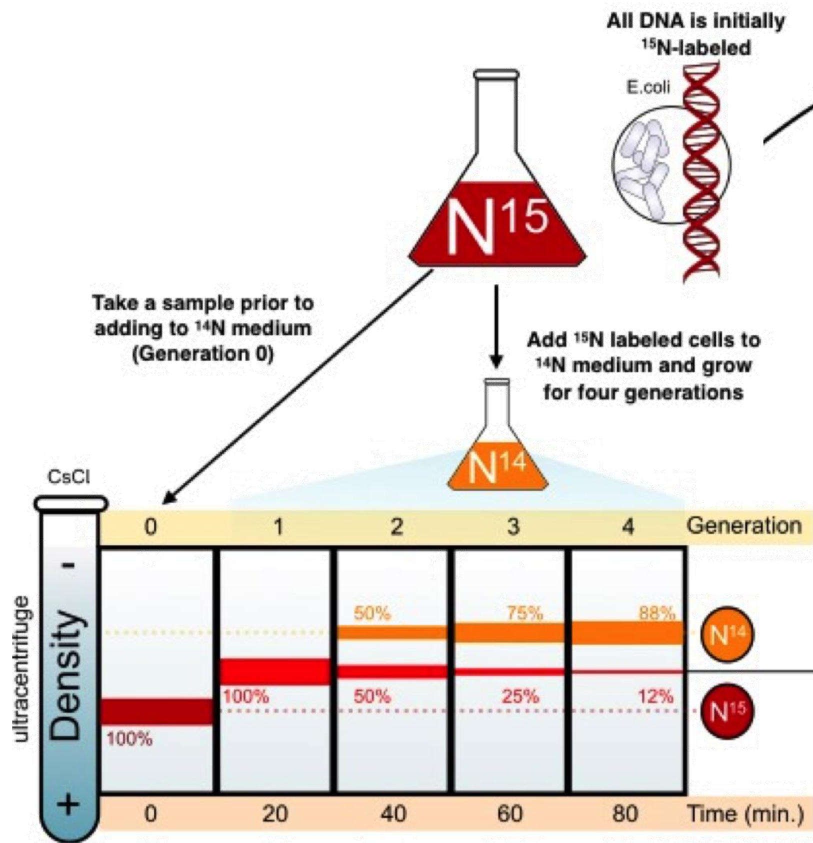
- DNA isolated from cells at the start of the experiment ("generation 0," just before the switch to ^{14}\text N14Nstart superscript, 14, end superscript, start text, N, end text medium) produced a single band after centrifugation. This result made sense because the DNA should have contained only heavy ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text at that time.

GENERATION 1

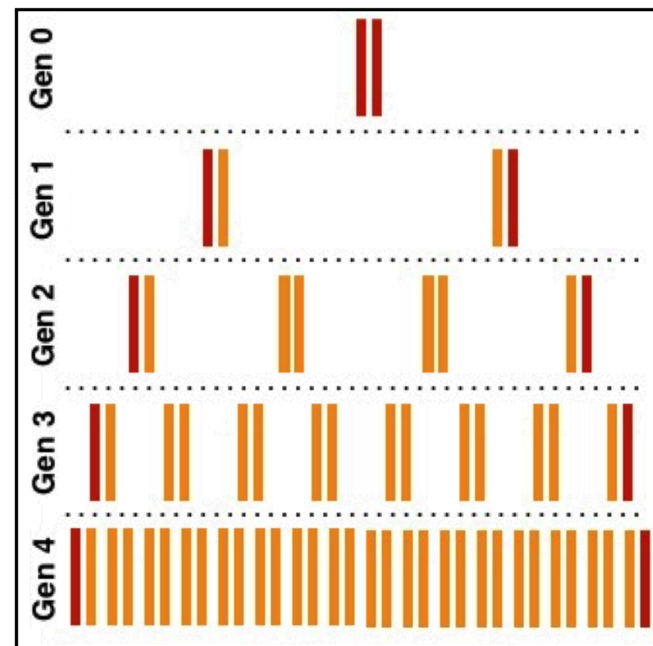
- DNA isolated after one generation (one round of DNA replication) also produced a single band when centrifuged. However, this band was higher, intermediate in density between the heavy ^{15}\text N15Nstart superscript, 15, end superscript, start text, N, end text DNA and the light ^{14}\text N14Nstart superscript, 14, end superscript, start text, N, end text DNA.
- The intermediate band told Meselson and Stahl that the DNA molecules made in the first round of replication was a hybrid of light and heavy DNA. This result fit with the dispersive and semi-conservative models, but not with the conservative model.
- The conservative model would have predicted two distinct bands in this generation (a band for the heavy original molecule and a band for the light, newly made molecule).

GENERATION 2

- Information from the second generation let Meselson and Stahl determine which of the remaining models (semi-conservative or dispersive) was actually correct.
- When second-generation DNA was centrifuged, it produced two bands. One was in the same position as the intermediate band from the first generation, while the second was higher (appeared to be labeled only with ^{14}\text N14Nstart superscript, 14, end superscript, start text, N, end text).
- This result told Meselson and Stahl that the DNA was being replicated semi-conservatively. The pattern of two distinct bands—one at the position of a hybrid molecule and one at the position of a light molecule—is just what we'd expect for semi-conservative replication (as illustrated in the diagram below). In contrast, in dispersive replication, all the molecules should have bits of old and new DNA, making it impossible to get a "purely light" molecule.



All DNA is initially
15N-labeled
E.coli
N15
Take a sample prior to
adding to 14N medium
(Generation 0)
Add 15N labeled cells to
14N medium and grow
for four generations
N14
CsCl
ultracentrifuge
Density
0
1
2
3
4
Generation
50%
75%
88%
100%
50%
25%
12%
100%
0
20
40
60
80
Time (min.)



Gen 0
Gen 1
Gen 2
Gen 3
Gen 4

GENERATIONS 3 AND 4

- In the semi-conservative model, each hybrid DNA molecule from the second generation would be expected to give rise to a hybrid molecule and a light molecule in the third generation, while each light DNA molecule would only yield more light molecules.
- Thus, over the third and fourth generations, we'd expect the hybrid band to become progressively fainter (because it would represent a smaller fraction of the total DNA) and the light band to become progressively stronger (because it would represent a larger fraction).
- As we can see in the figure, Meselson and Stahl saw just this pattern in their results, confirming a semi-conservative replication model.

CONCLUSION

- The experiment done by Meselson and Stahl demonstrated that DNA replicated semi-conservatively, meaning that each strand in a DNA molecule serves as a template for synthesis of a new, complementary strand.
- Although Meselson and Stahl did their experiments in the bacterium *E. coli*, we know today that semi-conservative DNA replication is a universal mechanism shared by all organisms on planet Earth. Some of your cells are replicating their DNA semi-conservatively right now!

DNA PROOFREADING AND REPAIR

- **Key points:**
- Cells have a variety of mechanisms to prevent **mutations**, or permanent changes in DNA sequence.
- During DNA synthesis, most DNA polymerases "check their work," fixing the majority of mispaired bases in a process called **proofreading**.
- Immediately after DNA synthesis, any remaining mispaired bases can be detected and replaced in a process called **mismatch repair**.
- If DNA gets damaged, it can be repaired by various mechanisms, including **chemical reversal**, **excision repair**, and **double-stranded break repair**.

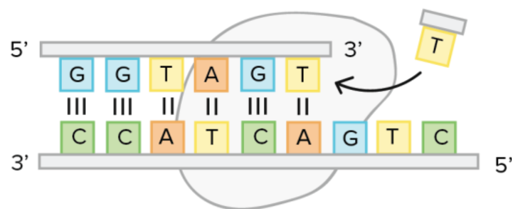
INTRODUCTION

- What does DNA have to do with cancer? Cancer occurs when cells divide in an uncontrolled way, ignoring normal "stop" signals and producing a tumor. This bad behavior is caused by accumulated **mutations**, or permanent sequence changes in the cells' DNA.
- Replication errors and DNA damage are actually happening in the cells of our bodies all the time. In most cases, however, they don't cause cancer, or even mutations. That's because they are usually detected and fixed by DNA proofreading and repair mechanisms. Or, if the damage cannot be fixed, the cell will undergo programmed cell death (apoptosis) to avoid passing on the faulty DNA.

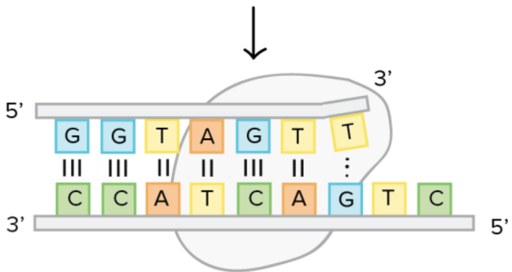
- Mutations happen, and get passed on to daughter cells, only when these mechanisms fail. Cancer, in turn, develops only when multiple mutations in division-related genes accumulate in the same cell.
- In this article, we'll take a closer look at the mechanisms used by cells to correct replication errors and fix DNA damage, including:
- Proofreading, which corrects errors during DNA replication
- Mismatch repair, which fixes mispaired bases right after DNA replication
- DNA damage repair pathways, which detect and correct damage throughout the cell cycle

PROOFREADING

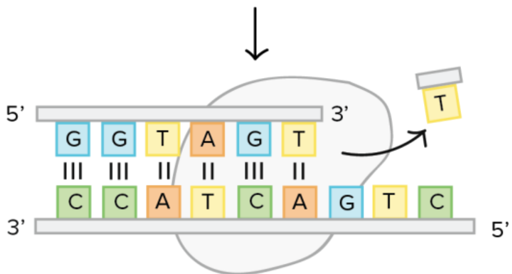
- **DNA polymerases** are the enzymes that build DNA in cells. During DNA replication (copying), most DNA polymerases can “check their work” with each base that they add. This process is called **proofreading**. If the polymerase detects that a wrong (incorrectly paired) nucleotide has been added, it will remove and replace the nucleotide right away, before continuing with DNA synthesis^11start superscript, 1, end superscript.



Polymerase adds an incorrect nucleotide to the new strand of DNA.



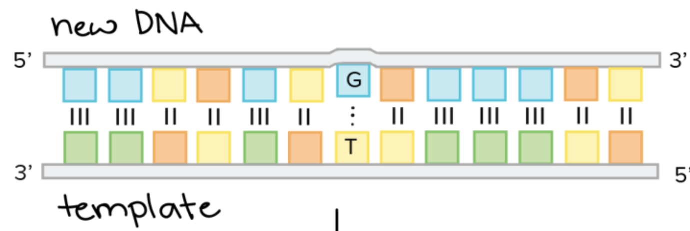
Polymerase detects that bases are mispaired.



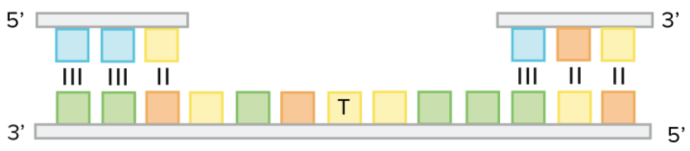
Polymerase uses 3’→ 5’ exonuclease activity to remove incorrect nucleotide.

MISMATCH REPAIR

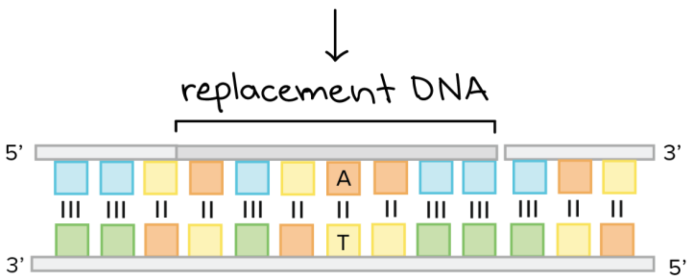
- Many errors are corrected by proofreading, but a few slip through. **Mismatch repair** happens right after new DNA has been made, and its job is to remove and replace mis-paired bases (ones that were not fixed during proofreading). Mismatch repair can also detect and correct small insertions and deletions that happen when the polymerases "slips," losing its footing on the template^22squared.
- How does mismatch repair work? First, a protein complex (group of proteins) recognizes and binds to the mispaired base. A second complex cuts the DNA near the mismatch, and more enzymes chop out the incorrect nucleotide and a surrounding patch of DNA. A DNA polymerase then replaces the missing section with correct nucleotides, and an enzyme called a DNA ligase seals the gap^22squared.



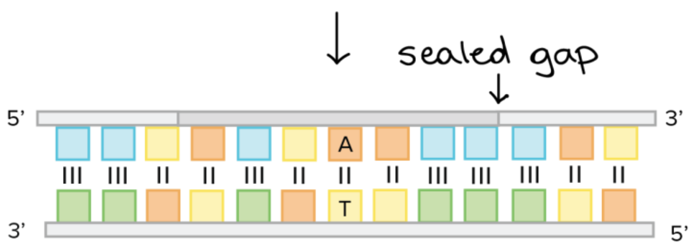
A mismatch is detected in newly synthesized DNA.



The new DNA strand is cut, and the mispaired nucleotide and its neighbors are removed.



The missing patch is replaced with correct nucleotides by a DNA polymerase.



A DNA ligase seals the gap in the DNA backbone.

- One thing you may wonder is how the proteins involved in DNA repair can tell "who's right" during mismatch repair. That is, when two bases are mispaired (like the G and T in the drawing above), which of the two should be removed and replaced?
- In bacteria, original and newly made strands of DNA can be told apart by a feature called *methylation state*. An old DNA strand will have methyl (-\text{CH}_3−CH3minus, start text, C, H, end text, start subscript, 3, end subscript) groups attached to some of its bases, while a newly made DNA strand will not yet have gotten its methyl group^33cubed.
- In eukaryotes, the processes that allow the original strand to be identified in mismatch repair involve recognition of nicks (single-stranded breaks) that are found only in the newly synthesized DNA^33cubed.

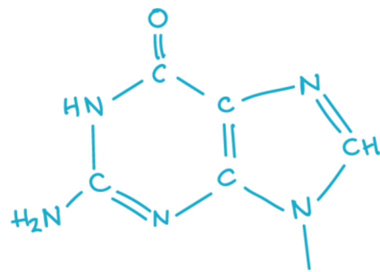
DNA DAMAGE REPAIR MECHANISMS

- Bad things can happen to DNA at almost any point in a cell's lifetime, not just during replication. In fact, your DNA is getting damaged all the time by outside factors like UV light, chemicals, and X-rays—not to mention spontaneous chemical reactions that happen even without environmental insults!^44start superscript, 4, end superscript
- Fortunately, your cells have repair mechanisms to detect and correct many types of DNA damage. Repair processes that help fix damaged DNA include:

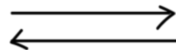
- **Direct reversal:** Some DNA-damaging chemical reactions can be directly "undone" by enzymes in the cell.
- **Excision repair:** Damage to one or a few bases of DNA is often fixed by removal (excision) and replacement of the damaged region. In **base excision repair,** just the damaged base is removed. In **nucleotide excision repair**, as in the mismatch repair we saw above, a patch of nucleotides is removed.
- **Double-stranded break repair:** Two major pathways, non-homologous end joining and homologous recombination, are used to repair double-stranded breaks in DNA (that is, when an entire chromosome splits into two pieces).

REVERSAL OF DAMAGE

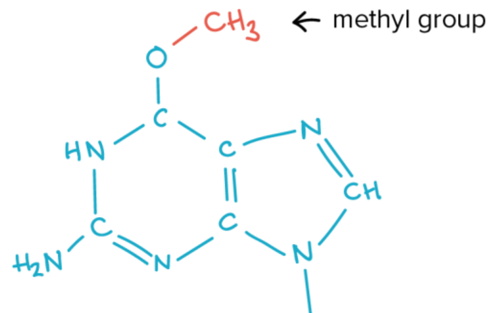
- In some cases, a cell can fix DNA damage simply by reversing the chemical reaction that caused it. To understand this, we need to realize that "DNA damage" often just involves an extra group of atoms getting attached to DNA through a chemical reaction.
- For example, guanine (G) can undergo a reaction that attaches a methyl (-\text{CH}_3−CH3minus, start text, C, H, end text, start subscript, 3, end subscript) group to an oxygen atom in the base. The methyl-bearing guanine, if not fixed, will pair with thymine (T) rather than cytosine (C) during DNA replication. Luckily, humans and many other organisms have an enzyme that can remove the methyl group, reversing the reaction and returning the base to normal^55start superscript, 5, end superscript.



reaction with
harmful chemical
adds a methyl group



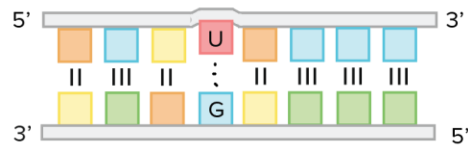
enzyme removes
methyl group



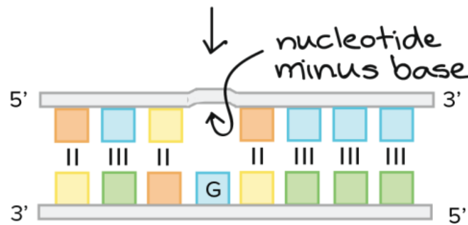
← methyl group

BASE EXCISION REPAIR

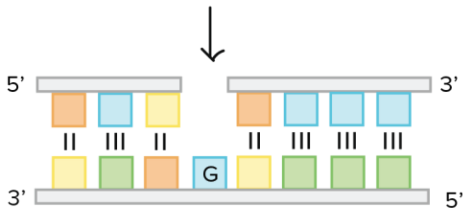
- **Base excision repair** is a mechanism used to detect and remove certain types of damaged bases. A group of enzymes called glycosylases play a key role in base excision repair. Each glycosylase detects and removes a specific kind of damaged base.
- For example, a chemical reaction called deamination can convert a cytosine base into uracil, a base typically found only in RNA. During DNA replication, uracil will pair with adenine rather than guanine (as it would if the base was still cytosine), so an uncorrected cytosine-to-uracil change can lead to a mutation^55start superscript, 5, end superscript.
- To prevent such mutations, a glycosylase from the base excision repair pathway detects and removes deaminated cytosines. Once the base has been removed, the "empty" piece of DNA backbone is also removed, and the gap is filled and sealed by other enzymes^66start superscript, 6, end superscript.



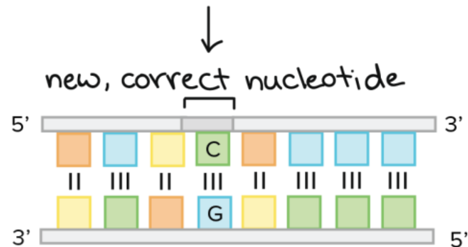
Deamination converts a cytosine base into a uracil.



The uracil is detected and removed, leaving a base-less nucleotide.



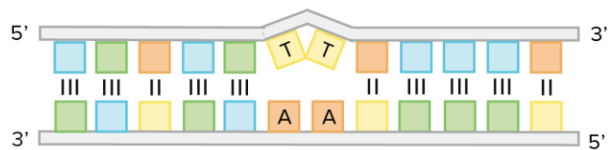
The base-less nucleotide is removed, leaving a small hole in the DNA backbone.



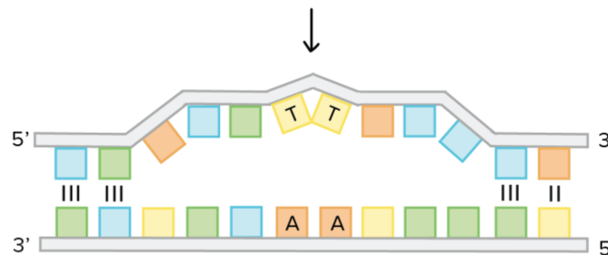
The hole is filled with the right base by a DNA polymerase, and the gap is sealed by a ligase.

NUCLEOTIDE EXCISION REPAIR

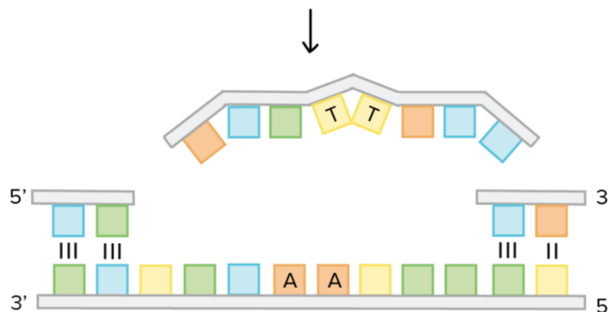
- **Nucleotide excision repair** is another pathway used to remove and replace damaged bases. Nucleotide excision repair detects and corrects types of damage that distort the DNA double helix. For instance, this pathway detects bases that have been modified with bulky chemical groups, like the ones that get attached to your DNA when it's exposed to chemicals in cigarette smoke^77start superscript, 7, end superscript.
- Nucleotide excision repair is also used to fix some types of damage caused by UV radiation, for instance, when you get a sunburn. UV radiation can make cytosine and thymine bases react with neighboring bases that are also Cs or Ts, forming bonds that distort the double helix and cause errors in DNA replication. The most common type of linkage, a **thymine dimer**, consists of two thymine bases that react with each other and become chemically linked^88start superscript, 8, end superscript.
-



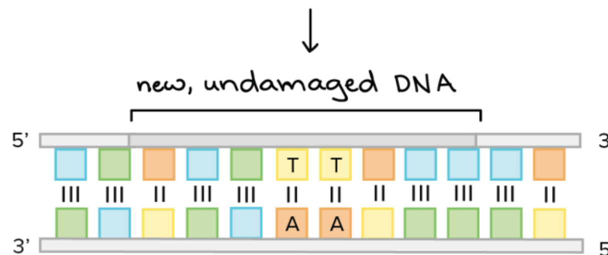
UV radiation produces a thymine dimer.



Once the dimer has been detected, the surrounding DNA is opened to form a bubble.



Enzymes cut the damaged region out of the bubble.

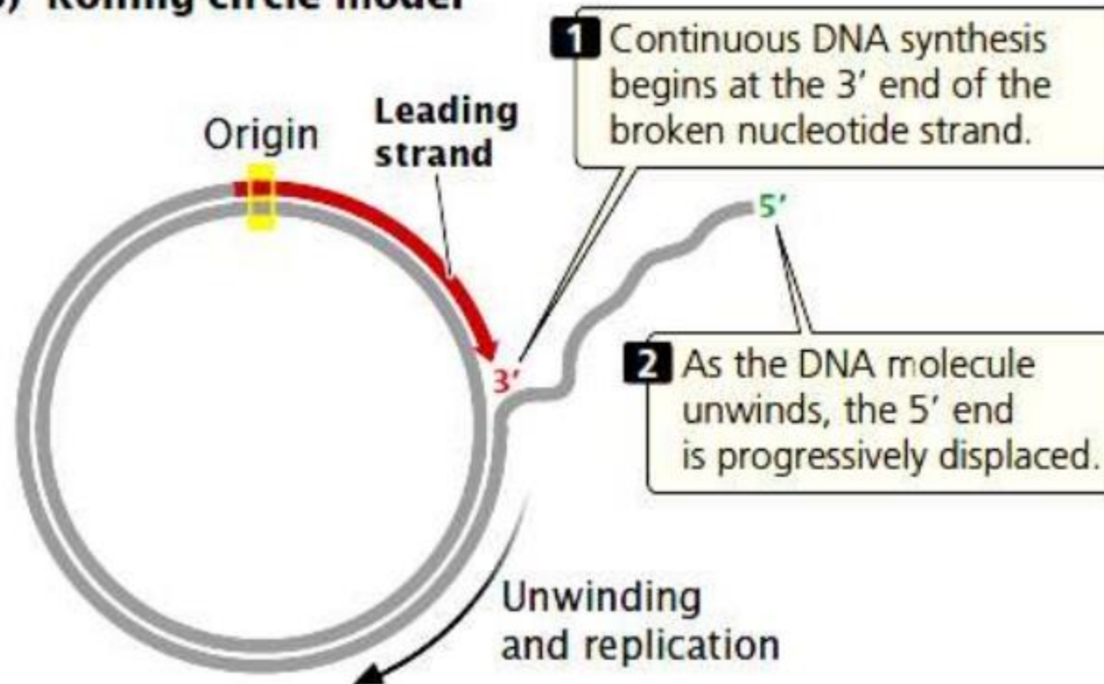


A DNA polymerase replaces the excised (cut-out) DNA, and a ligase seals the backbone.

- In nucleotide excision repair, the damaged nucleotide(s) are removed along with a surrounding patch of DNA. In this process, a helicase (DNA-opening enzyme) cranks open the DNA to form a bubble, and DNA-cutting enzymes chop out the damaged part of the bubble. A DNA polymerase replaces the missing DNA, and a DNA ligase seals the gap in the backbone of the strand^9

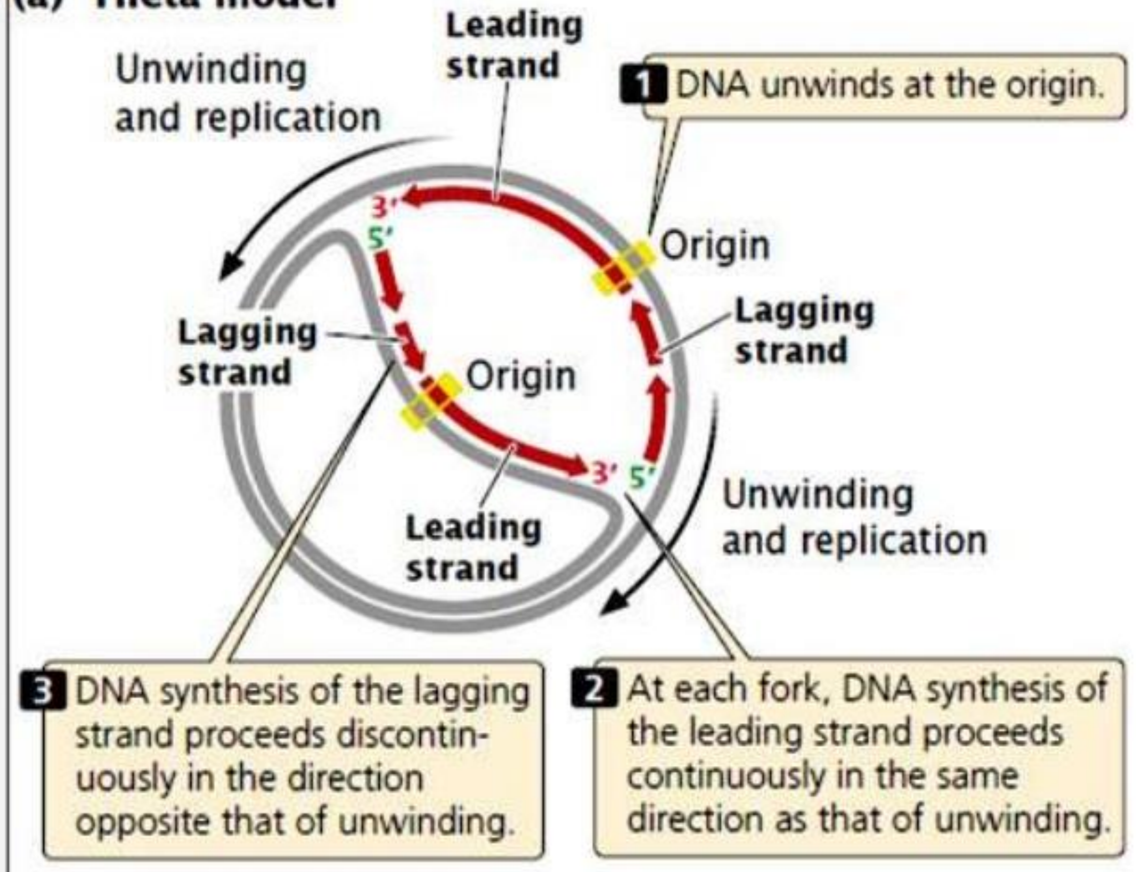
ROLLING CIRCLE REPLICATION (RCA

- **Rolling circle replication** (RCA) is a process of unidirectional nucleic acid **replication** that can rapidly synthesize multiple copies of circular molecules of DNA or RNA, such as plasmids, the genomes of bacteriophages, and the circular RNA genome of viroids


(b) Rolling-circle model
Origin
Leading strand
1 Continuous DNA synthesis begins at the 3′ end of the broken nucleotide strand.
5′
3′
2 As the DNA molecule unwinds, the 5′ end is progressively displaced.
Unwinding and replication

THETA DNA REPLICATION

- A **theta** structure is an intermediate structure formed during the **replication** of a circular DNA molecule. Two **replication** forks can proceed independently around the DNA ring and when viewed from above the structure resembles the Greek letter "**theta**" (θ).


(a) Theta model
Leading strand
Unwinding and replication
1 DNA unwinds at the origin.
3'
5'
Origin
Lagging strand
Lagging strand
Origin
3'
5'
Unwinding and replication
Leading strand
3 DNA synthesis of the lagging strand proceeds discontinuously in the direction opposite that of unwinding.
2 At each fork, DNA synthesis of the leading strand proceeds continuously in the same direction as that of unwinding.