CONSTANCY OF DNA – *DNA REPLICATION*

Ho Huynh Thuy Duong

University of Science

GENETICS - STUDYING THE LAWS OF HEREDITY

- **Gregor Mendel set up the basic principles of Classical Genetics. His first and second laws about independent segregation and assortment of "genetic factors" implied that "genetic factors" are separate entities (1865).**

- **Walter Sutton (1903) observed from cytological samples that diploid chromosomes consist of two morphologically similar chromosomes which are equally distributed to each gamete during meiosis → "genetic factors" could be parts of the chromosomes.**

- **Thomas Morgan *et al* (1908), from their work on *Drosophila* mutants, found out that : (1) Groups of linked genes correspond to haploid chromosome numbers, (2) Recombination (crossing over) frequency between genes reflects the distance between them . This observation set up the chromosomal basis of heredity of which application concerned the construction of the first genetic map.**

- **Archibald Garrod (1909) suggested that Alcaptonuria, an inherited disorder, is due to the absence of a metabolic enzyme for phenylalanine → This was the first attempt to find gene-protein relationship.**

- **Frederick Griffith (1928) observed the transfer of virulent characteristics between two *Streptococci* strains. Oswald Avery, Colin McLeod, Maclyn McCarty (1944) proved that the transforming factor in Griffith's experiment was DNA.**

- **Alfred Hershey & Martha Chase (1952) definitely established DNA as genetic material by their experiments on bacteriophage (virus of bacteria)**

FROM MENDELIAN LAWS TO MOLECULAR GENETICS

ᘓ Maurice Wilkins & Rosalind Franklin (1950s) reported X-ray diffraction photographs of DNA molecule, suggesting that its helical structure is composed of more than one strand.

ᘓ Erwin Chargaff (1949) set up Chargaff's rules : Cellular ratios of A =T and G=C

→ Based on these findings, Francis Crick & James Watson (1953) announced the double-helical structure of DNA.

ᘓ Francis Crick (1956) proposed the Central Dogma of Molecular Biology :

TRANSCRIPTION ***TRANSLATION***

REPLICATION ↳**DNA** ➔ **RNA** ➔ **Protein**

ᘓ Matthew Meselson and Franklin Stahl (1958) proved that DNA replication is semi-conservative

ᘓ Discovery of DNA polymerases, RNA polymerases, mRNA, tRNA, rNA, …

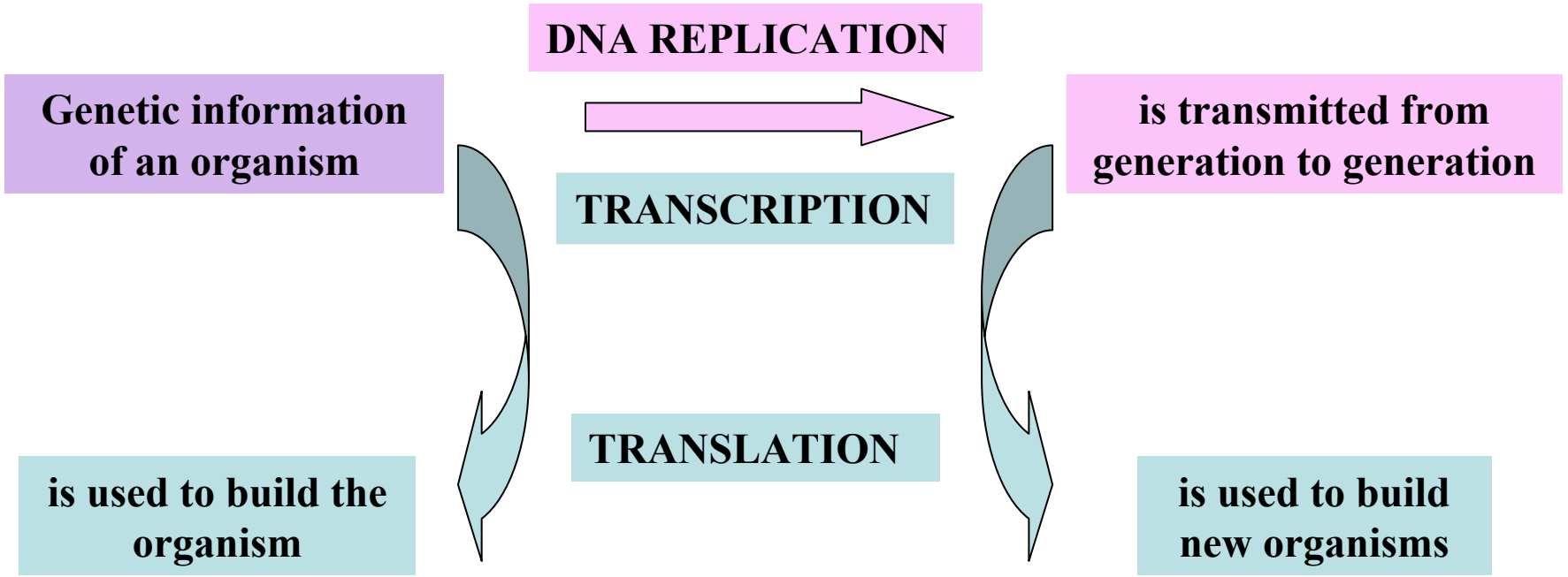
ᘓ Marshall Nirenberg & Heinrich Matthaei (1961), Har Gobind Khorana (1968) established the Genetic Code

• • •

ERA OF GENOMICS

The maintenance of DNA constancy is assured by : (1) DNA replication mechanism and (2) Repair systems of the cell

The ultimate goal of DNA replication is to maintain the constancy of genetic material throughout generations.



Nevertheless, genetic material of an organism is continuously altered throughout its life. Mutation, Recombination and Transposition are important factors causing these alterations. On the scale of the living world, variations of DNA provide new material for natural selection.

DNA REPLICATION

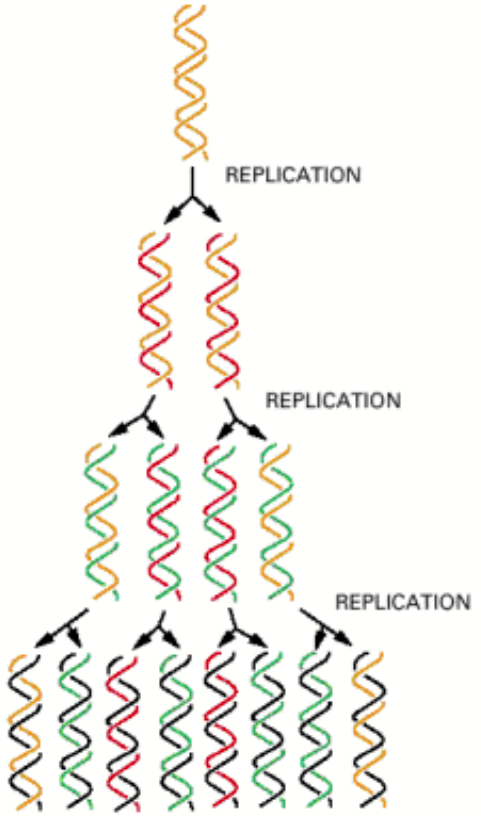
WHAT IS THE BIOLOGICAL ROLE OF DNA REPLICATION ?

Living organisms tend to perpetuate themselves by transmitting to their descendants their own characteristics. When a cell divides to give rise to two daughter cells, it tries to transmit to them as precisely as possible its genetic material. During cell division, the DNA double helix replicates to form two copies, each of which is distributed to each daughter cells.

HOW CAN DNA REPLICATION GIVE RISE TO TWO IDENTICAL COPIES ?

The double-helix structure of DNA and the sequence complementarity between the two strands suggested a semi-conservative model for DNA replication.

Meselson & Stahl (1958), by their experiments on bacteria grown in media containing either light or heavy nitrogen isotopes (^{14}N or ^{15}N), proved that DNA replication is semi-conservative. In the model they proposed, during DNA replication, the two parental strands separate, each strand serves as template to synthesize a complementary daughter strand *(Figure)*.



DNA REPLICATION PROCESS

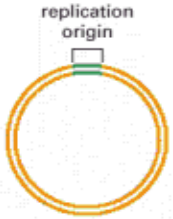
- **Mechanisms of DNA replication in prokaryotes and eukaryotes are essentially similar with some differences due to differences in cell structure and functions between the two groups**

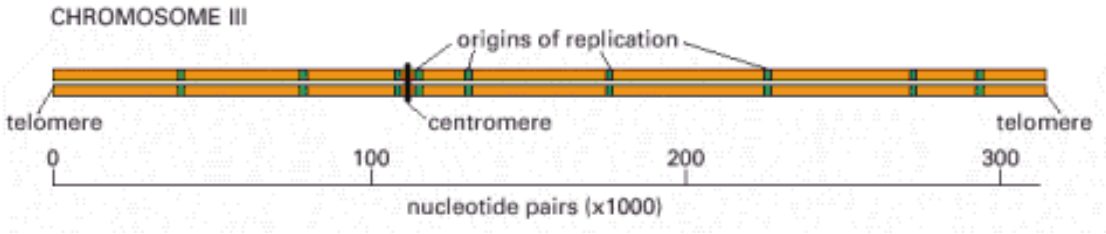
- **DNA replication process includes Initiation, DNA synthesis and Termination steps**

INITIATION OF DNA REPLICATION

- **The semi-conservative synthesis of DNA requires the separation of parental strands to serve as single-stranded template. The site where the two strands initially separate from each other is called the origin of replication.**

- **The segments of DNA replicated from an origin of replication form a replicon. The whole prokaryotic chromosome is a replicon *(Figure - left)* whereas eukaryotic chromosome is composed of many replicons (*Figure - right)***





The origin of DNA replication on E. coli chromosome

The origins of DNA replication on chromosome III of the yeast S. cerevisiae.

REPLICATOR AND INITIATOR

A replicon includes two components, the replicator and the initiator :

- **Replicators comprise all DNA sequences necessary to the initiation of DNA replication**

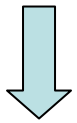
Replicators include two types of sequences : initiator-binding sites and AT-rich sites where the two strands can be easily unwound.

In *E. coli*, the replicator is called *oriC*

- **Initiators are proteins which interact with replicator to initiate DNA replication**

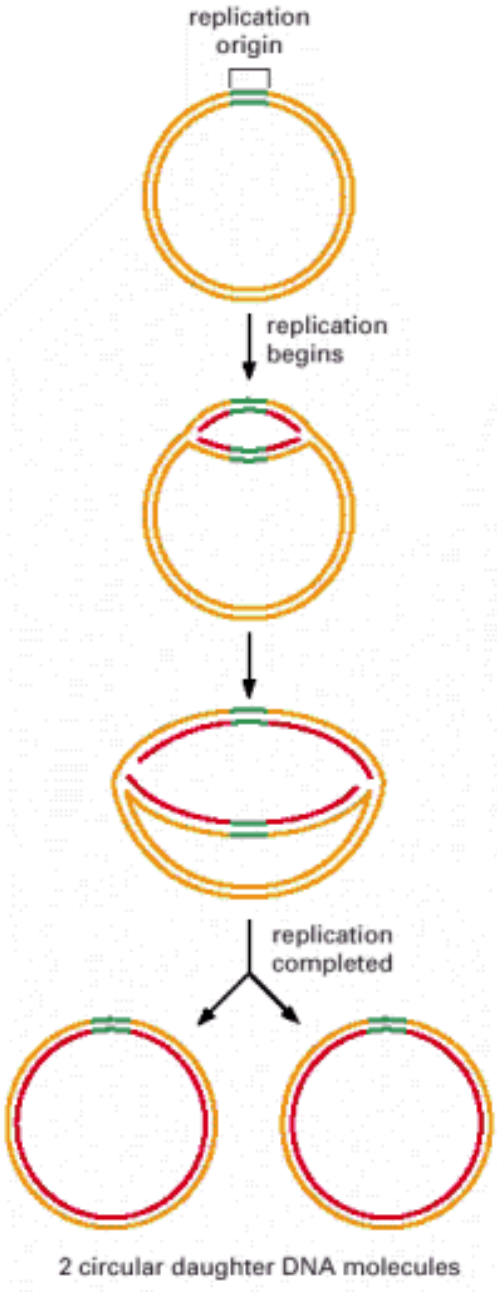
Initiators have three functions : (1) recognize and bind to a sequence within the replicator, (2) unwind the AT-rich region of the replicator, (3) recruit other proteins to efficiently initiate DNA replication.

Initiator protein in *E. coli* is called DnaA. Its homolog in eukaryotes is the Origin Recognition Complex (ORC)



INTERACTION BETWEEN REPLICATOR AND INITIATOR LEAD TO THE FORMATION OF TWO REPLICATION FORKS

INITIATION OF REPLICATION IN *E. COLI*



- **The replicator - *OriC* - includes five 9-mer repeats (TTATNCANA) and three 13-mer repeats (GATCTNTTNTTTT).**

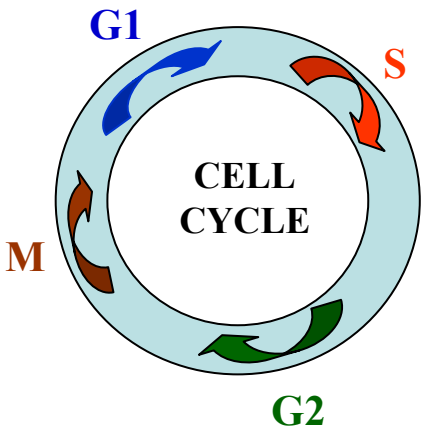
- **The initiator proteins DnaA of *E. coli* bind to the 9-mer repeats, unwind a sequence between the 13-mer repeats and recruit the helicase loader DnaC and helicase DnaB to the origin of replication.**

- **DnaC assembles DnaB on the single strand fragment at the origin of replication. The movement of DnaB along the single strand fragment dissociates the two parental strands and initiates the formation of two replication forks.**

Replication fork is the junction between the still intact double helix and the two newly synthesized strands. The movement of the replication fork progressively separates and copies each strand of the double helix until the whole DNA molecule is replicated.

Due to antiparallel orientation of the two DNA single strands, DNA synthesis using these template strands proceeds in opposite direction.

INITIATION OF REPLICATION IN EUKARYOTES



In eukaryote, the replicator sequence is less well defined than in prokaryotes. Initiators have to scan a region containing the replicator

Initiation of replication proceeds in two steps :

(1) Replicator selection

- **Occurs in G1 phase**
- **Replicator selection is mediated by protein complexes called pre-replicative complexes (pre-RCs)**
- **Pre-RCs are formed firstly by ORC (initiator) recognition of the replicator. Binding of ORC to the replicator recruits the helicase loaders Cdc6 and Cdt1. Helicase loaders will then recruit the helicase Mcm 2-7 complex to the origin of replication**

(2) Origin activation

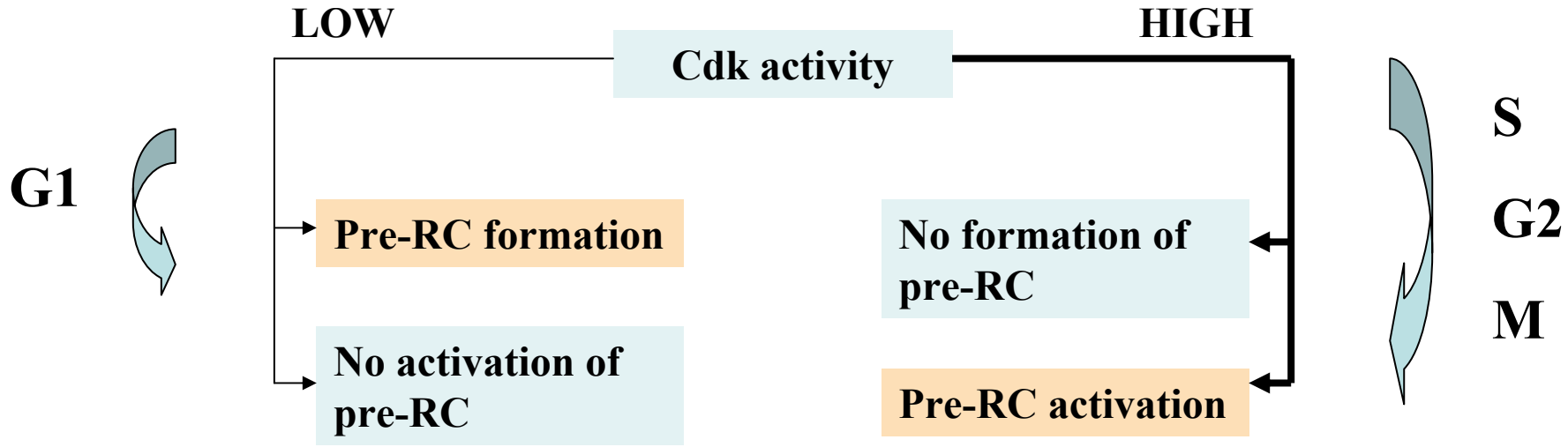
- **Begins in S phase of the cell cycle**
- **Pre-RCs, activated by two protein kinases Cdk and Ddk, unwind DNA and recruit DNA polymerases and other replicative proteins to the origin of replication**

DNA polα/primase are recruited first, followed by DNA pol δ and ε

PROBLEMS OF DNA REPLICATION IN EUKARYOTES

The crucial problem for eukaryotes is to replicate DNA once and only once per cell cycle.

The solution lies in the relationship between Cdk activity, pre-RC formation and cell cycle phase

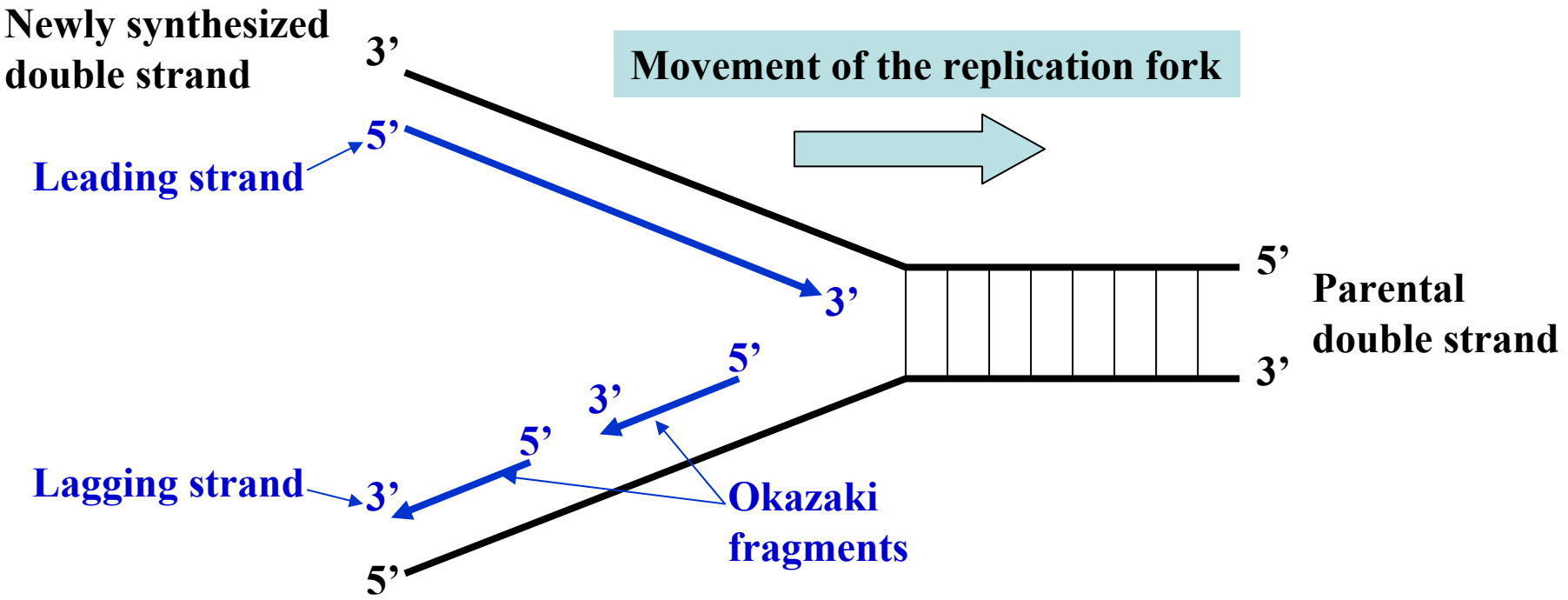


- **Cdk activity is low during G1 phase, allowing pre-RC formation. During the remaining phases (S, G2, M), high activity of Cdk activates the pre-RC formed and inhibits new pre-RC formation → for each cell cycle, only one pre-RC is formed at an origin of replication.**

- **On the whole eukaryotic chromosome having many origins of replication, the activation of one replicator will : (1) Inactivate the adjacent replicators and also (2) inactivate the corresponding replicators on daughter molecules**

This ensures the complete replication of DNA once per cell cycle.

WHAT HAPPENS AT ONE REPLICATION FORK ?



- On one template strand, the newly synthesized strand grows in the same direction of replication fork movement and is called "leading strand". The leading strand is synthesized continuously from the beginning to the end.

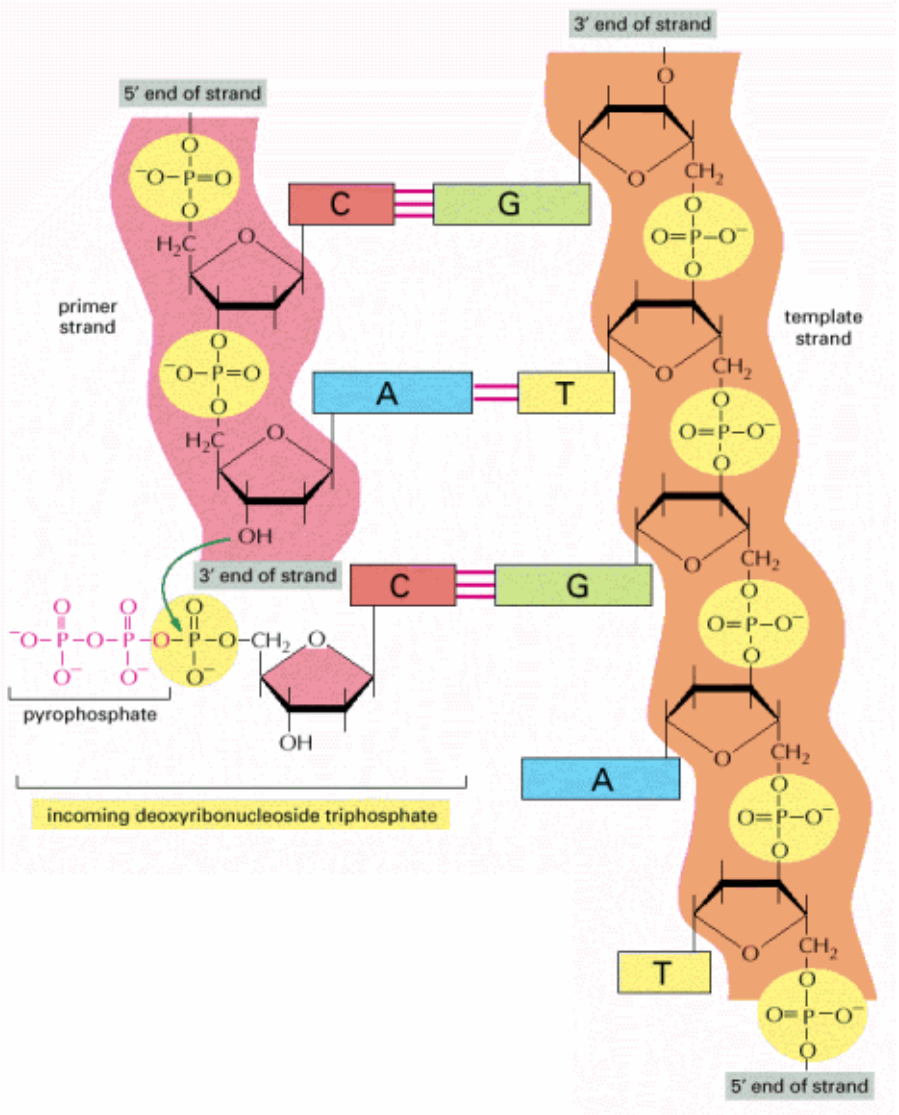
- On the other template strand, the daughter strand is synthesized in the opposite direction and is called "lagging strand". Synthesis of the lagging strand is discontinuous and must be re-initiated each time the replication fork moves and exposes a segment of the template. The lagging strand is composed of Okazaki fragments of about 1,000-2,000 nucleotides in prokaryotes and 100-200 nucleotides in eukaryotes. These Okazaki fragments are then joined to form an intact strand.

THE CHEMISTRY OF DNA SYNTHESIS

During DNA synthesis, one of the four deoxynucleoside triphosphates dATP, dTTP, dCTP and dGTP is added to the 3'end of the growing chain.

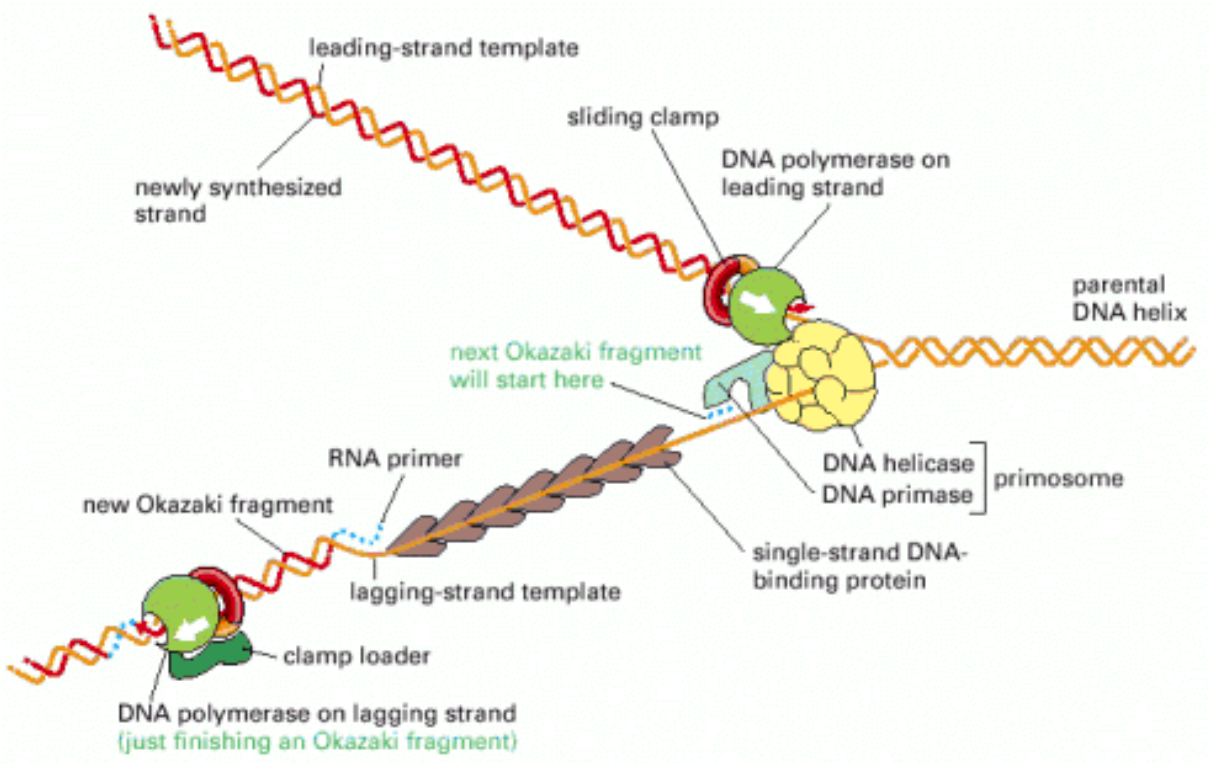
Each nucleotide added is linked to the terminal nucleotide of the growing chain by a phosphodiester bond. The formation of phosphodiester bonds uses energy caused by the release of a pyrophosphate from dNTP and the subsequent hydrolysis of this pyrophosphate into two inorganic phosphates.

Each added nucleotide base-pairs with the complementary nucleotide in the other chain in a Watson-Crick pairing.



HOW DOES THE REPLICATION FORK FUNCTION ?

FACTORS ACTING AT THE REPLICATION FORK



Factors acting at the replication fork include :

- **Two DNA polymerases, each catalyzes the synthesis of one daughter strand.**
- **DNA primases synthesize primers**
- **Single-Stranded Binding Proteins (SSB) bind to the separated parental strand**
- **DNA Topoisomerase unwinds the double helix ahead the moving replication fork**

- **The double helix must be separated into two single strands by DNA helicases.**

DNA helicases are composed of 6 protein molecules forming a ring which encircles one of the template strands at the moving point of the replication fork. The sliding of helicase on one strand displaces the other strand

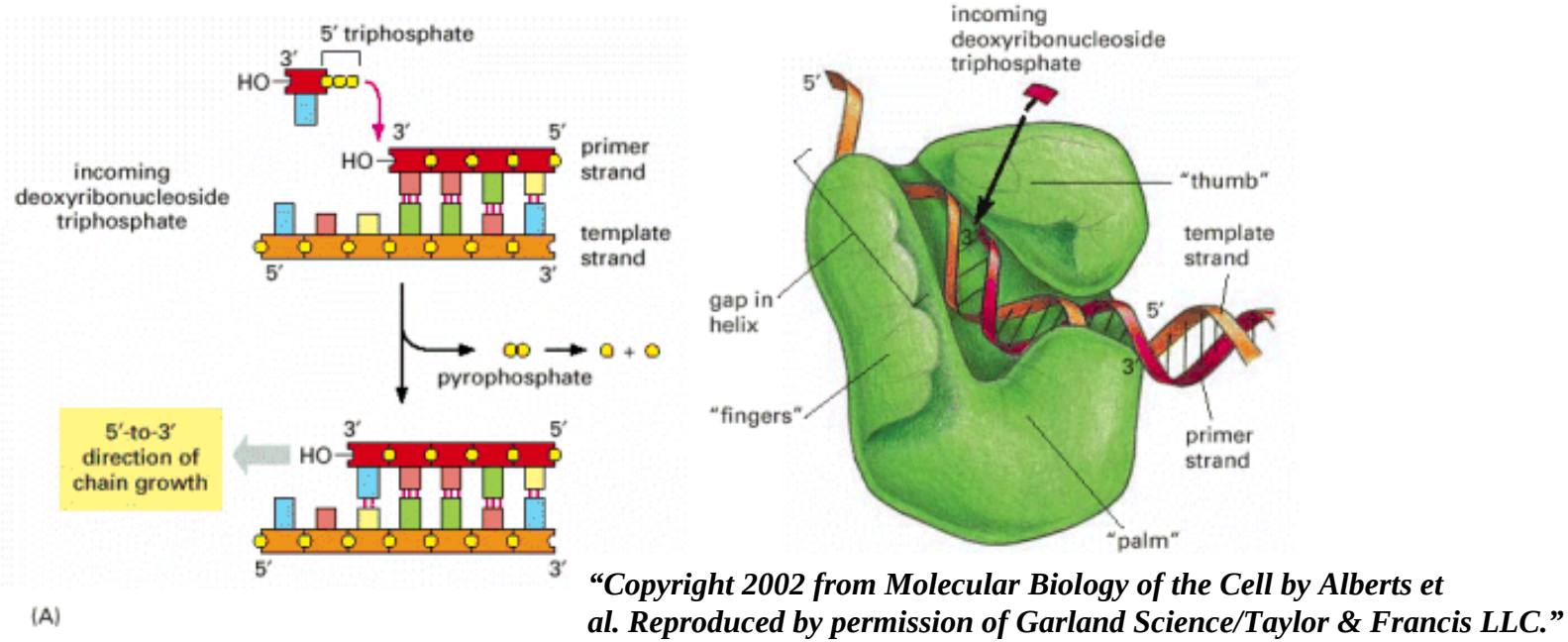
DNA SYNTHESIS IS CATALYZED BY DNA POLYMERASES

These enzymes have some characteristics :

(1) They synthesize DNA strand only in a 5'→3' direction

(2) They synthesize DNA strand by extending the 3'end of a short nucleotide chain (oligonucleotide), called "primer", annealed to the 3'terminus of the template strand.

(3) Many DNA polymerases possess 3' → 5' or 5'→ 3' exonuclease activities.



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

DNA polymerases catalyze the incorporation of nucleotides into the growing chain preferentially when the incoming nucleotide correctly base pairs with the template

Structure of an E.coli DNA polymerase which looks like a right hand with the "palm", "fingers" and "thumb" holding the DNA double helix

FACTORS ACTING AT THE REPLICATION FORK : DNA POLYMERASES

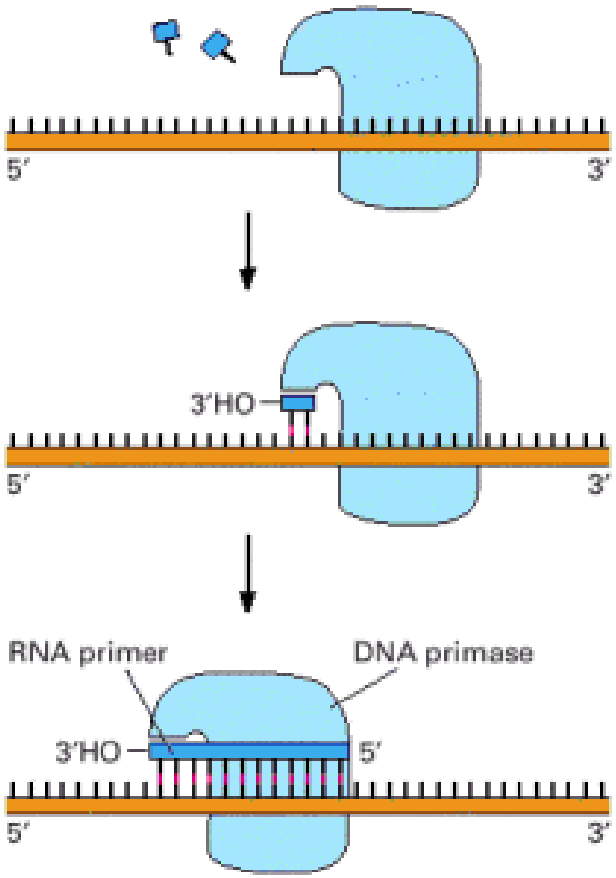
In prokaryotes (*E. coli*), there are five DNA polymerases :

- **DNA polymerase III holoenzyme : the main factor for chromosome replication, has high processivity and proofreading activity.**
- **DNA polymerase I : removes RNA primers and replaces them by DNA fragments, has proofreading activity.**
- **DNA polymerases II, IV, V : have essential roles in DNA repair systems**

In eukaryotes (*E. coli*), there are many DNA polymerases :

- **DNA polymerase α : combined with primase for RNA primers synthesis to initiate DNA replication**
- **DNA polymerases δ and ε : highly processive enzymes having the main role in replicating DNA, replace DNA polymerase α/primase after DNA synthesis has been initiated**
- **DNA polymerases β, γ, θ, ζ, λ, μ, κ, η, ι and Rev1 : have essential roles in DNA repair systems**

FACTORS ACTING AT THE REPLICATION FORK :
DNA PRIMASE



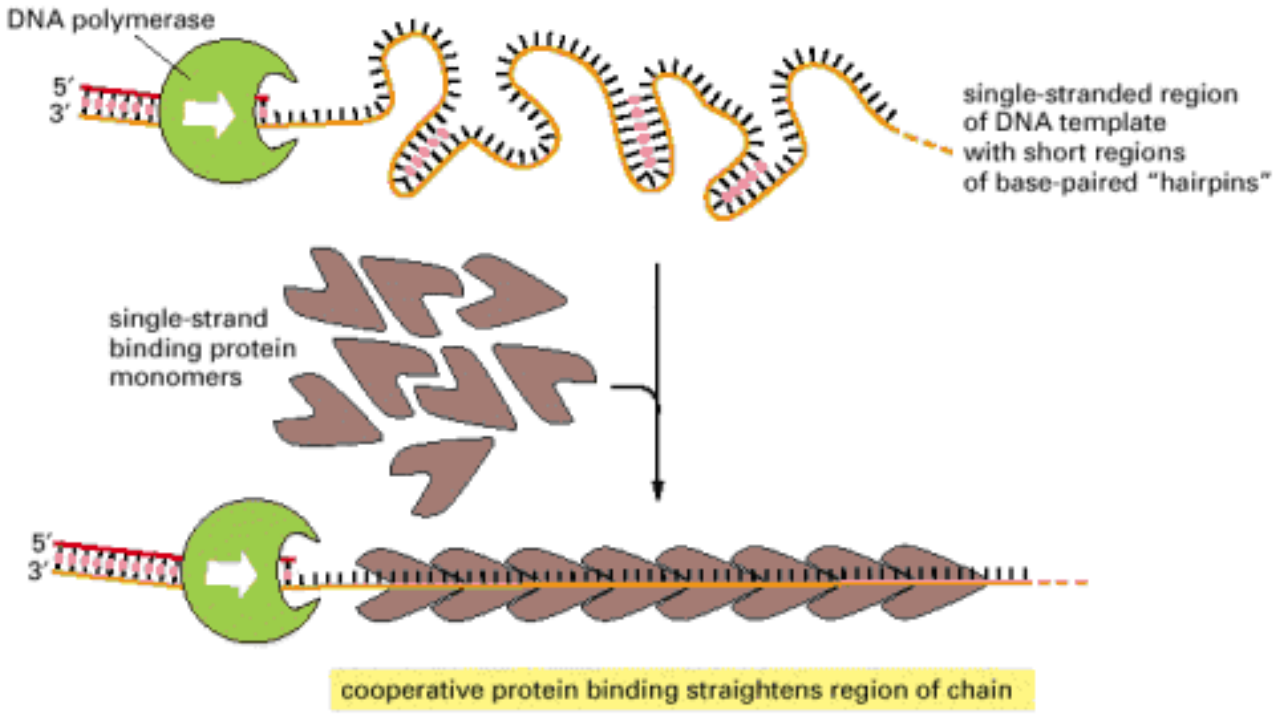
DNA polymerases require a free 3'OH group of an annealed primer to initiate the synthesis of the new strand. In the cell, these primers are short RNA fragments, synthesized by a type of RNA polymerase called primase.

Primases synthesize one primer for each leading strand. Whereas a lagging strand requires hundreds to thousands primers to prime its Okazaki fragments

At the replication fork, primase is associated with DNA helicase to form the primosome

Primer synthesis catalyzed by DNA primase

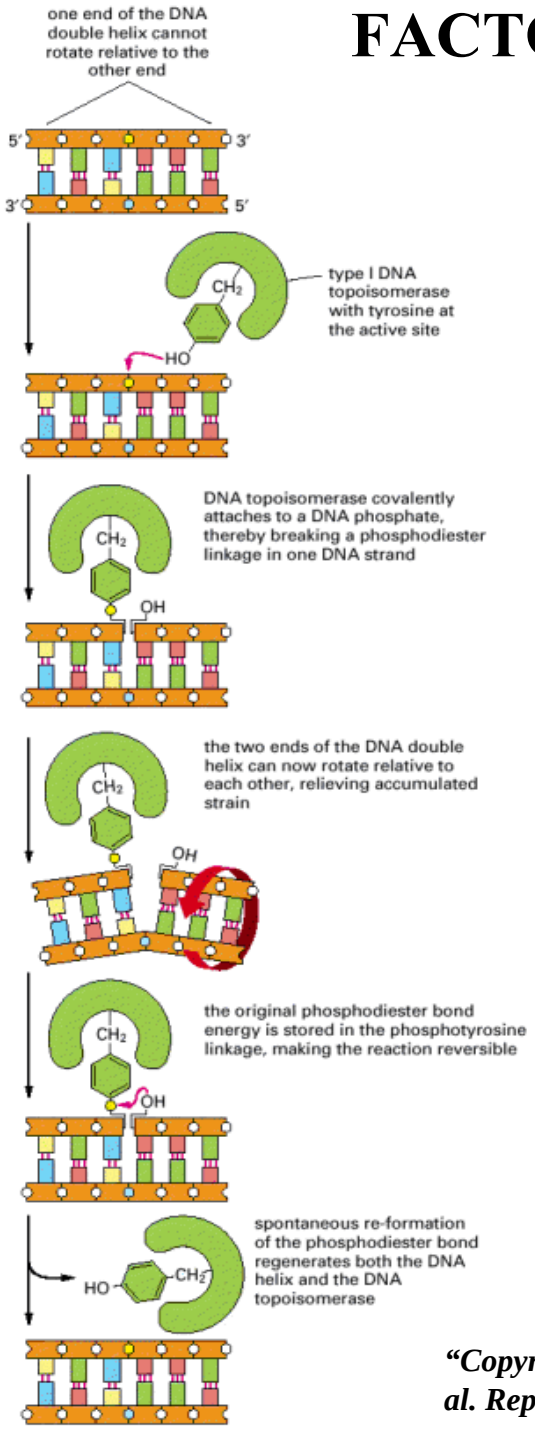
FACTORS ACTING AT THE REPLICATION FORK :
SSB PROTEIN



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

As soon as the two parental strands are separated, single-stranded DNA binding proteins (SSB protein) bind and stabilize these single strands, preventing their reassociation as well as inter-molecular secondary structures (hairpins) formation. Template strands are thus maintained in an elongated structure which facilitates the synthesis of daughter strands.

FACTORS ACTING AT THE REPLICATION FORK - DNA TOPOISOMERASE



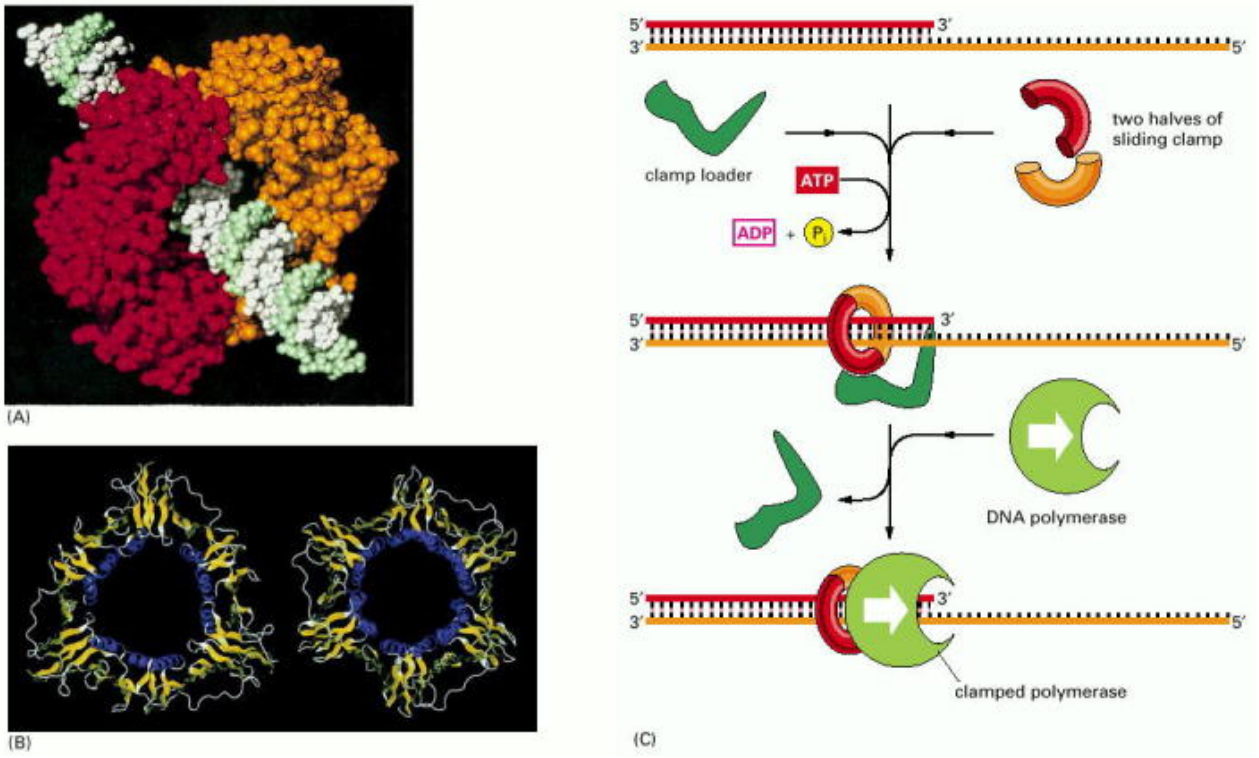
As DNA helicase unwinds double-stranded DNA behind the moving fork, the double helix in front of it will be increasingly supercoiled. Increasing supercoiling exerts a great pressure which will stop the replication machinery.

DNA topoisomerases induce a single or double strand break ahead of the replication fork, thus allowing a rotation of the two strands relative to each other. This rotation will remove the accumulated supercoiling. Then DNA topoisomerases reseal the break by reforming phosphodiester bonds.

DNA topoisomerases not only have a crucial role in bacterial circular choromosome but also in eukaryotic chromosomes which have very large size impeding the free rotation of the two strands during replication.

The reversible nicking reaction catalyzed by a eucaryotic DNA topoisomerase I enzyme

FACTORS ACTING AT THE REPLICATION FORK : SLIDING DNA CLAMP



"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

A sliding clamp (E. coli) with two halves encircling DNA molecule (left, above) ; Similar structures of a sliding clamp in E. coli (left, below) and its homolog in eukaryotes (right, below).

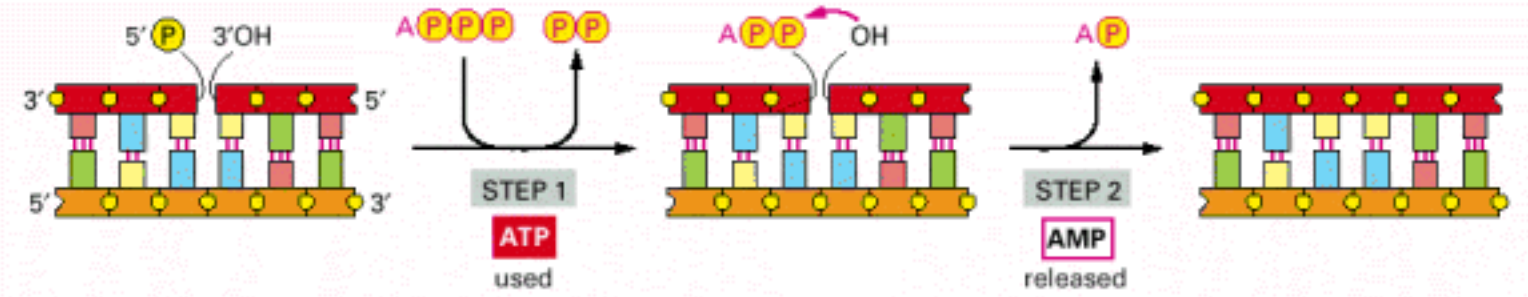
Assembly of sliding clamp by clamp loader and its association with DNA plymerase (right)

Sliding DNA clamps are proteins composed of many subunits forming a ring encircling the DNA molecule. When associated with DNA polymerase at the replication fork, sliding clamps make these enzymes capable of synthesizing continuously thousands to millions base pairs before being released from the template → DNA polymerase become highly processive.

Sliding clamps are assembled on DNA by a protein complex called clamp loader, using ATP-stored energy.

Sliding clamps are removed only when the ssDNA strand is completely copied.

COMPLETION OF DNA SYNTHESIS - RNASE H AND DNA LIGASE

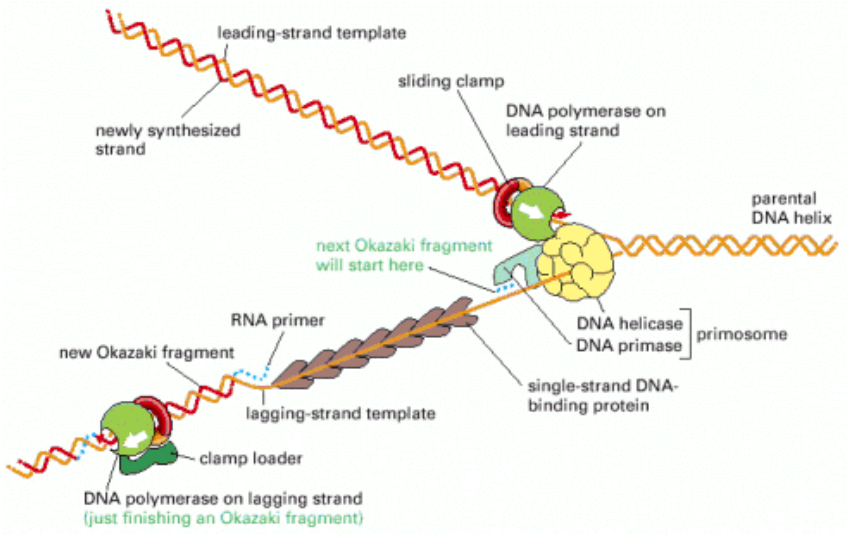


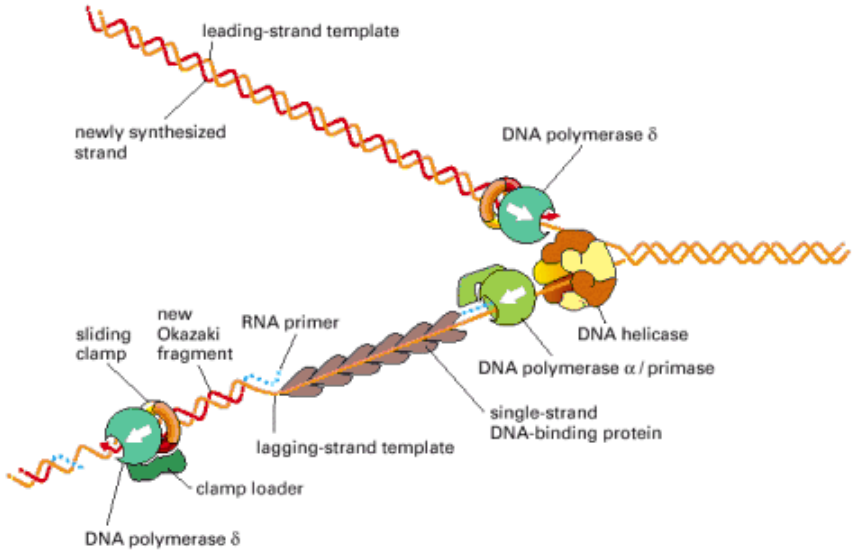
"Copyright 2002 from Molecular Biology of the Cell by Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis LLC."

Ligation reaction catalyzed by DNA ligase

- **RNA primers used to initiate DNA replication are removed by RNase H, except for the only ribonucleotide directly linked to DNA 5'phosphate end. This ribonucleotide is removed by DNA polymerase I which also fills in the gap left by RNA primer removal.**
- **The "nick" remaining between existing DNA fragments, such as Okazaki fragments on the lagging strand, and the DNA fragments replacing RNA primers are then sealed by DNA ligase. This enzyme uses energy from ATP to catalyze the ligation reaction.**

REPLICATION FORK IN PROKARYOTES AND EUKARYOTES





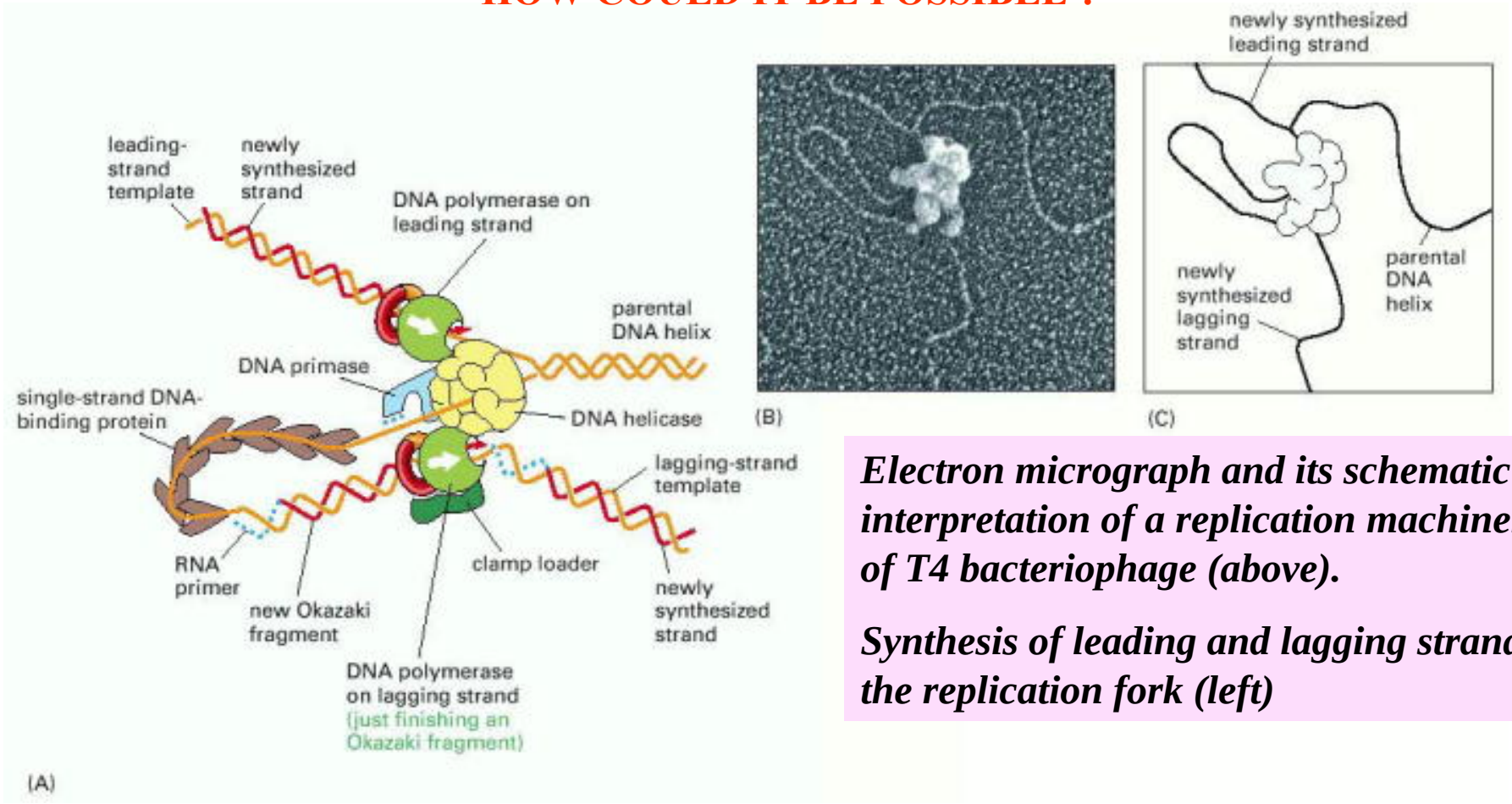
ENZYMES	PROKARYOTE (*E. COLI*) (left)	EUKARYOTE (HUMAN) (right)
Primase	DnaG	Primase
Helicase	DnaB	Mcm complex
SSB protein	SSB protein	RPA
Topoisomerase	Gyrase	Topo I, II
DNA polymerase	DNA pol III, I, II, IV, V	DNA pol δ, ε, α, β, γ, ζ, θ, η, ι,κ, λ
Ligase	Ligase	Ligase

These proteins interact to form the replisome at the replication fork

PROBLEMS OF LAGGING STRAND SYNTHESIS

Due to antiparallel orientation of the two template strands, the lagging strand and the leading strand are elongated in opposite directions whereas the two DNA polymerases synthesizing these strands are linked in a DNA replication complex !

HOW COULD IT BE POSSIBLE ?



Electron micrograph and its schematic interpretation of a replication machinery of T4 bacteriophage (above).

Synthesis of leading and lagging strands at the replication fork (left)

PROBLEMS OF LAGGING STRAND SYNTHESIS

A widely admitted model proposes that :

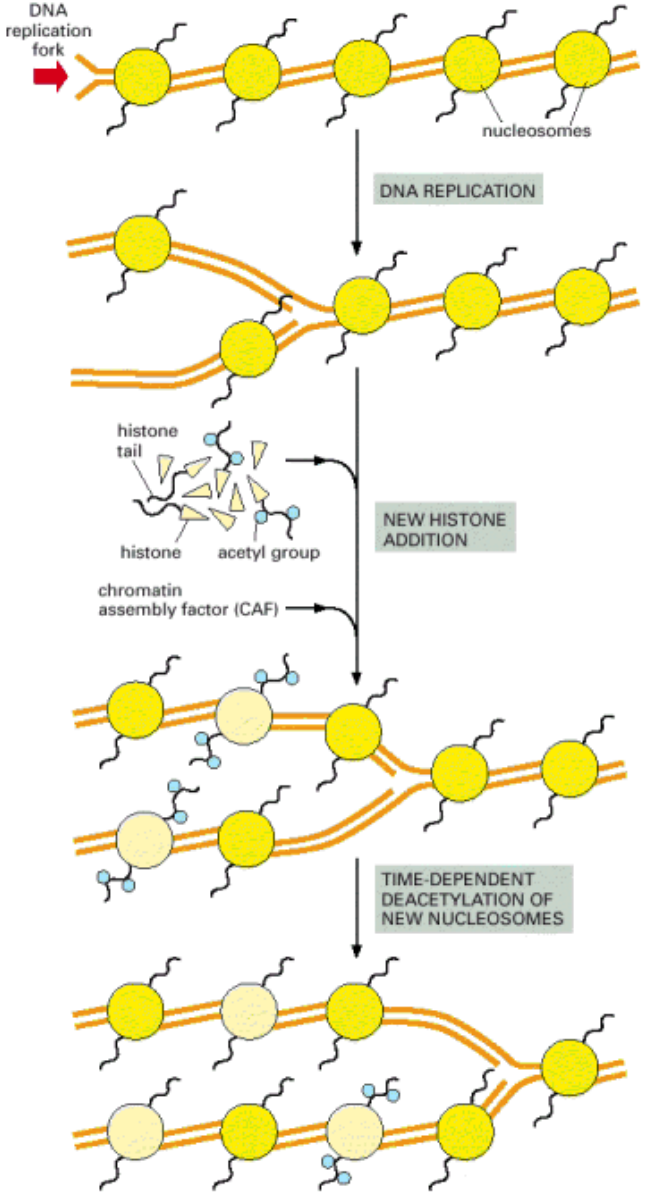
As the replication fork moves forward, the leading strand is elongated while the lagging strand is folded to form a loop so that at the moving point of the fork the two DNA polymerases move in the same direction.

The lagging strand-DNA polymerase is released from the template after finishing one Okazaki fragment but still binds to the leading strand-DNA polymerase at the moving point → this keeps the lagging strand-DNA polymerase in the proximity of the moving point.

During this time, primase synthesizes a new primer which anneals to a site near the moving point. Lagging strand-DNA polymerase binds to this primer, forms a loop and initiates the synthesis of the following Okazaki fragment.

This model is also used to explain DNA synthesis at the replication fork in eukaryotes

NUCLEOSOME DISASSEMBLY AND ASSEMBLY IN EUKARYOTES



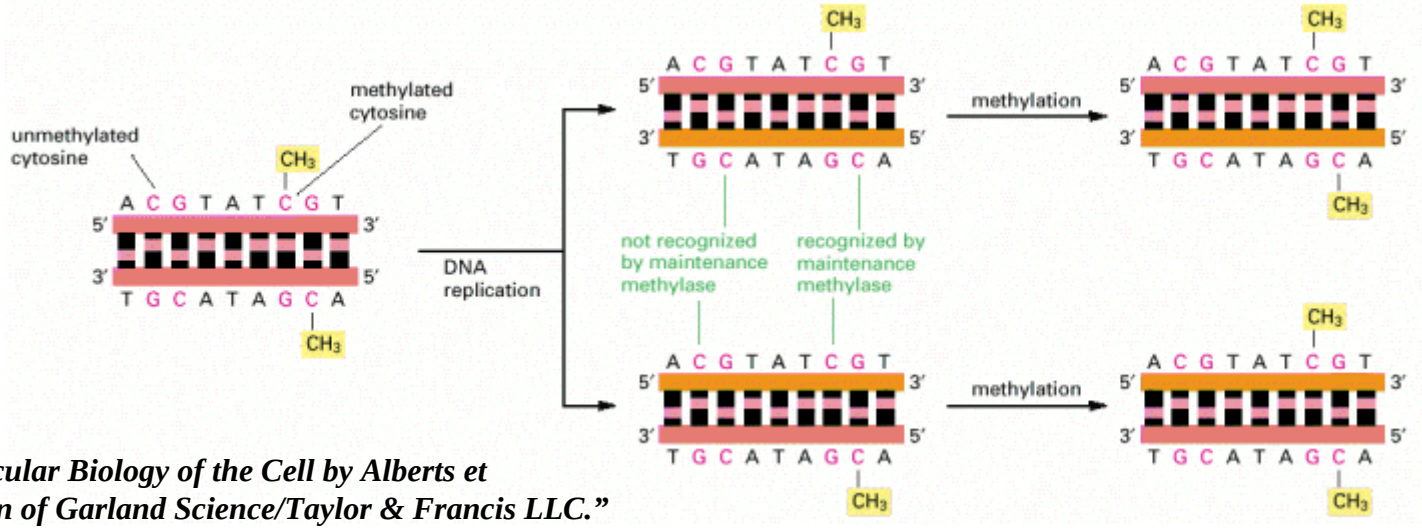
- **During DNA replication, nucleosomes are disassembled as the replication fork passes. Nevertheless, only H2A-H2B dimers are released from DNA, tetramers H3-H4 are randomly distributed to one of the two daughter molecules and are continuously bound to them.**
- **The daughter DNA molecules are rapidly packaged by the assembly of nucleosome components, beginning with H3-H4 tetramers binding, followed by the association of two H2A-H2B dimers and the final histone H1.**

Besides the remaining old histones, new histones are synthesized.

Nucleosomes on the daughter molecule are thus composed of old and new histones. The old histones restore the parental histone modification profile for each daughter strands.

- **Histone chaperones, such as CAF-1, bound to the PCNA – the sliding DNA clamp in eukaryotes, retain the H3-H4 tetramers in proximity of the replication fork.**

DNA METHYLATION INHERITANCE

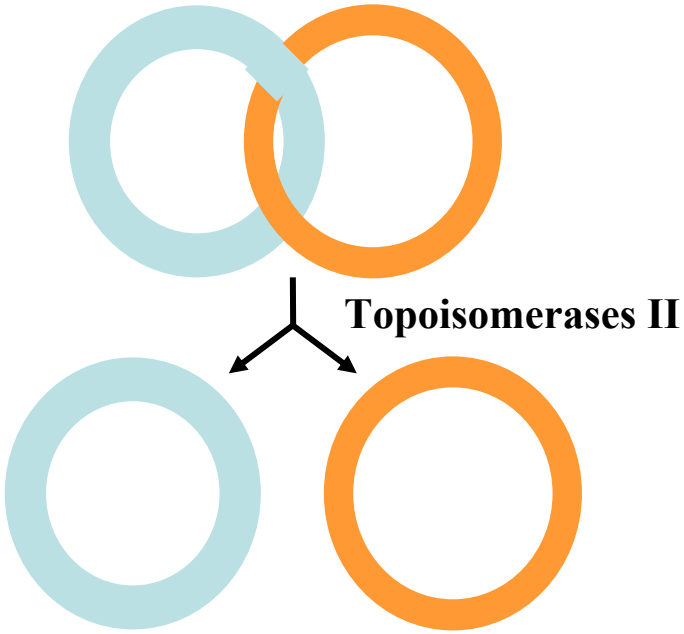


In mammals genome, unmethylated clusters of CpG (CpG "islands") are found in the promoter region of "housekeeping" and some tissue-specific genes which are actively transcribed. Methylated CpGs, in contrast, correspond mostly to inactive genes.

DNA methylation patterns are faithfully maintained throughout cell division. The example above shows DNA methylation inheritance in vertebrate genomic DNA. There are enzymes called Maintenance Methylases which recognize the hemi-methylated site, site where only the C in one strand is methylated, and add a methyl group to the non-methylated C in the opposite strand.

The inheritance pattern of DNA methylation state throughout generations belongs to a transcriptional control mechanism called epigenetic inheritance.

TERMINATION OF REPLICATION



"Adapted from Watson J.D. et al. 2004. Molecular Biology of the Gene. 5th edition, p.228, fig 8.33. Benjamin Cummings., CSHL Press"

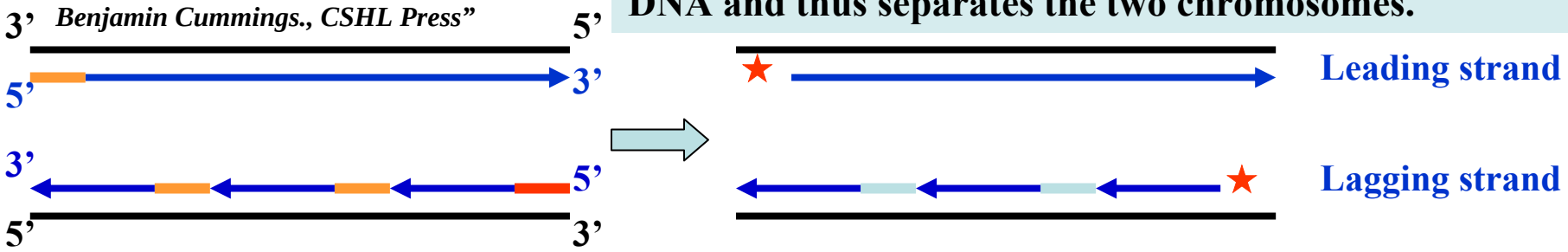
- In prokaryotes, there are some **termination sites** at about 180° from the start site, each group of termination sites corresponds to one replication fork

The sequence of a termination site :

AATTAGTATGTTGTAACTAAAGT
TTAATCATACAACATTGATTTCA

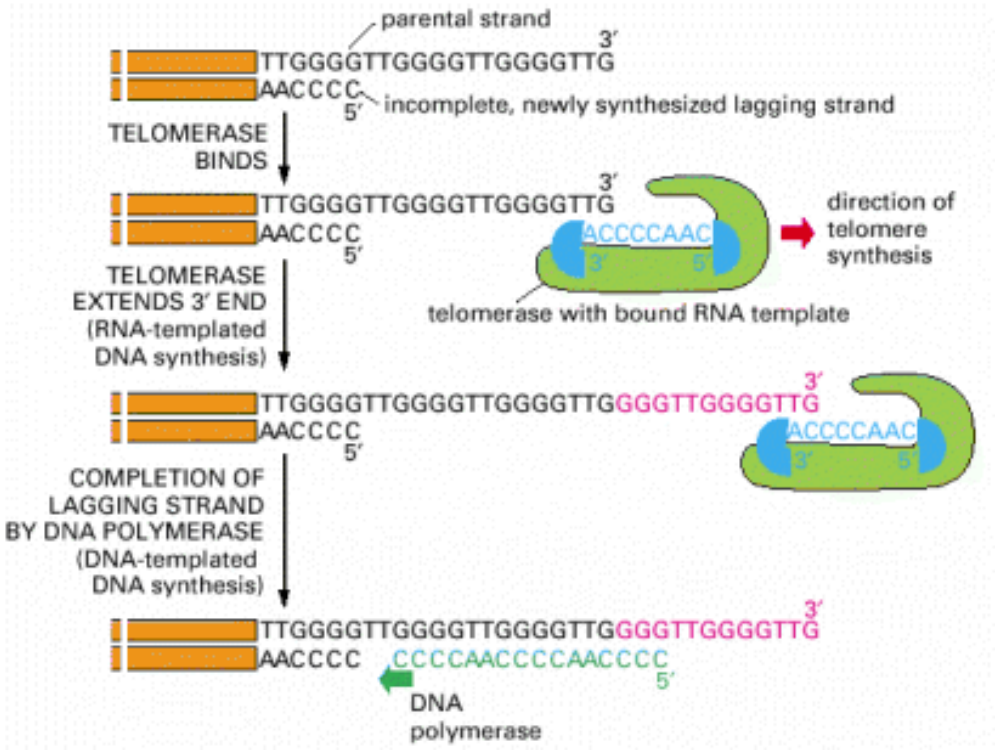
- Replicating circular chromosome (in prokaryotes) results in two daughter molecules linked together as catenanes (left, above).

The decatenation of this structure is catalyzed by **Topoisomerases II** which can break a double-stranded DNA and thus separates the two chromosomes.



Replicating linear chromosomes of eukaryotes encounters some problems, in particular for the lagging strand : (1) There could be no enough space to synthesize the last RNA primer so that the lagging strand template could not be completely copied, (2) When RNA primers are removed, the last RNA primer (★) can not be replaced.

FINISHING REPLICATION IN EUKARYOTES



The end of eukaryotic chromosomes are called telomeres and are replicated by enzymes called telomerase.

Telomerase is composed of a catalytic protein and an RNA. This RNA has sequence complementary to telomere repeats.

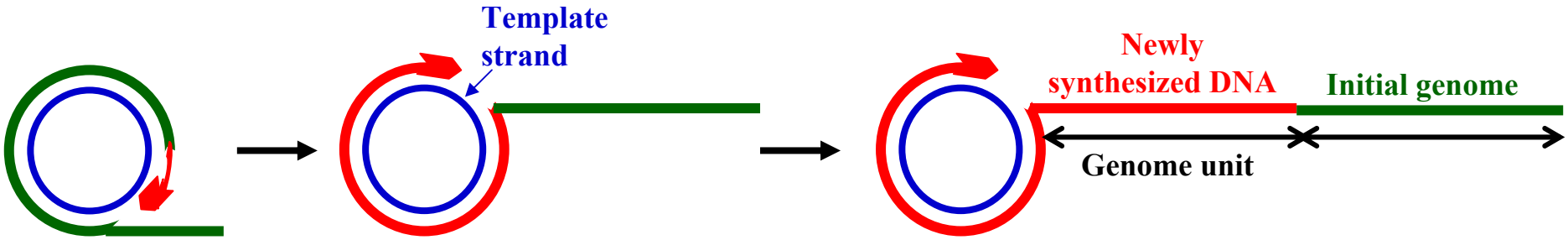
Telomerase binds to the 3' end of the overhang DNA and uses its RNA component (*in blue*) as template to elongate this DNA strand. After elongating the end of overhang strand, telomerase rebinds at the new formed 3'end and repeats the process. After many rounds of 3'end elongation (*in pink*), there will be enough place for new RNA primers and thus for new Okazaki fragments (*in green*) synthesis to avoid the shortening of telomeres after each DNA replication round.

DNA REPLICATION IN ROLLING CIRCLE

ꕤ Rolling circle is a mode of DNA replication in which a circular template is copied many times by a unidirectional replication fork giving rise to multiple replication units.

At each replication round, the newly synthesized DNA displaces the previously synthesized strand resulting in a long 5' end including many genome units.

Rolling circle can be used to replicate one or both strands to produce single-stranded or double-stranded DNA.



ꕤ Rolling circle is mainly used by bacteriophages to replicate their DNA ;

e.g φX174 phage has a single strand circular genome called plus (+) strand. To initiate genome replication, a minus (-) strand is synthesized and serves as template. The long newly synthesized strand resulting from many replication rounds is cleaved at the origins of replication to produce many genome units.

SUMMARY

ᘓ The transmission of the genetic material from generation to generation is assumed by DNA replication which is a semi-conservative process using both parental strands as template to synthesize two daughter strands.

ᘓ DNA replication process includes three steps : initiation, elongation and termination.

1. The replication initiation requires the recognition and binding of initiators to replicators, also called origin of replication. The parental double strand is melted leading to the formation of a replication fork. In eukaryotes, the initiation step is strictly controlled during the cell cycle

2. In the elongation step, the replication fork moves away from the origin of replication synthesizing a continuous leading strand and a discontinuous lagging strand. The components of the replication fork include : helicase, topoisomerase, primase, SSB proteins, DNA polymerases, RNase H and ligase.

3. The termination of replication in prokaryotes require Topoisomerase II activities to decatenate the final products. In eukaryotes, the end-of-replication problem is due to the shortening of linear chormosomes after each replicative round and can be resolved in some cases by the action of telomerase, an RNA-protein enzyme complex, which elongates the 3' end

ᘓ In most cases, the DNA replication is a bidirectional DNA synthesis produced by the movement in opposite direction of two replication forks

In some other cases, such as in many bacteriophages, DNA replication is done by rolling circle consisting of only one replication fork.