Ministry of Higher Education

& Scientific Research

Al Muthanna University

College of Veterinary Medicine

Division of Public Health

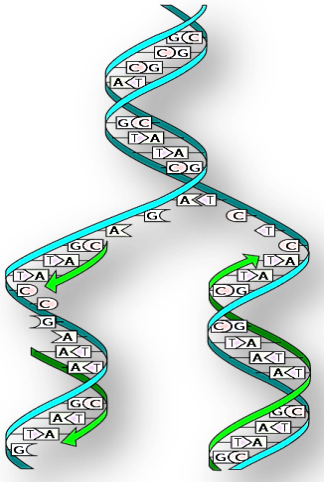
Subject: Genetics

Grade: 2

Lecture : 5

❖ DNA replication

In molecular biology, DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule.[1] DNA replication occurs in all living organisms acting as the most essential part for biological inheritance. This is essential for cell division during growth and repair of damaged tissues, while it also ensures that each of the new cells receives its own copy of the DNA. The cell possesses the distinctive property of division, which makes replication of DNA essential.



DNA structure

DNA exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types of nucleotides. Nucleotides in DNA contain a deoxyribose sugar, a phosphate, and a nucleobase. The four types of nucleotide correspond to the four nucleobases adenine, cytosine, guanine, and thymine, commonly abbreviated as A, C, G and T. Adenine and guanine are purine bases, while cytosine and thymine are pyrimidines. These nucleotides form phosphodiester bonds, creating the phosphate-deoxyribose

backbone of the DNA double helix with the nucleobases pointing inward (i.e., toward the opposing strand). Nucleobases are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine (two hydrogen bonds), and guanine pairs with cytosine (three hydrogen bonds (DNA strands have a directionality, and the different ends of a single strand are called the "3′ (three-prime) end" and the "5′ (five-prime) end". By convention, if the base sequence of a single strand of DNA is given, the left end of the sequence is the 5′ end, while the right end of the sequence is the 3′ end. The strands of the double helix are anti-parallel with one being 5′ to 3′, and the opposite strand 3′ to 5′. These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. Directionality has consequences in DNA synthesis, because DNA polymerase can synthesize DNA in only one direction by adding nucleotides to the 3′ end of a DNA strand.

The pairing of complementary bases in DNA (through hydrogen bonding) means that the information contained within each strand is redundant. Phosphodiester (intra-strand) bonds are stronger than hydrogen (inter-strand) bonds. The actual job of the Phosphodiester bonds is where in DNA polymers connect the 5' carbon of one nucleotide to the 3' carbon of another nucleotide, while the hydrogen bonds stabilize DNA double helices across the helix axis but not in the direction of the axis1. This allows the strands to be separated from one another. The nucleotides on a single strand can therefore be used to reconstruct nucleotides on a newly synthesized partner strand.

DNA polymerases

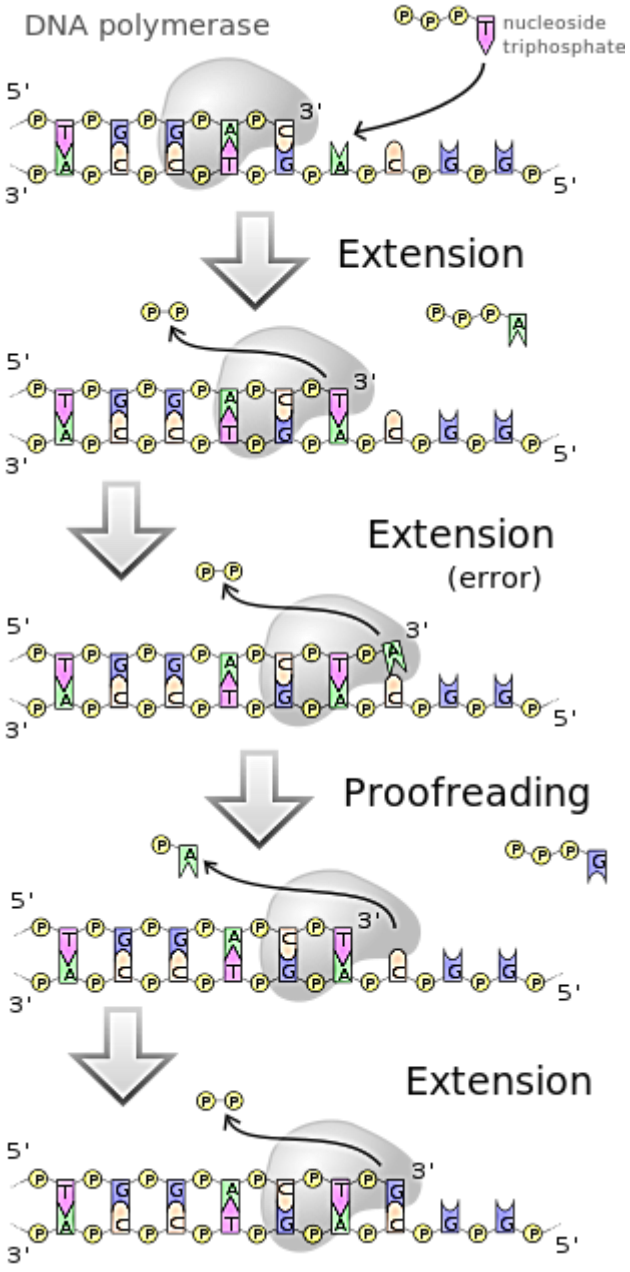
DNA polymerases are a family of enzymes that carry out all forms of DNA replication.[14] DNA polymerases in general cannot initiate synthesis of new strands, but can only extend an existing DNA or RNA strand paired with a template strand. To begin synthesis, a short fragment of RNA, called a primer, must be created and paired with the template DNA strand.

DNA polymerase adds a new strand of DNA by extending the 3’ end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from hydrolysis of the high-energy phosphate (phosphoanhydride) bonds between the three phosphates attached to each unincorporated base. Free bases with their attached phosphate groups are called nucleotides; in particular, bases with three attached phosphate groups are called nucleoside triphosphates. When a nucleotide is being added to a growing DNA strand, the formation of a phosphodiester bond between the proximal phosphate of the nucleotide to the growing chain is accompanied by hydrolysis of a high-energy phosphate bond with release of the two distal phosphates as a pyrophosphate. Enzymatic hydrolysis of the resulting pyrophosphate into inorganic phosphate consumes a second high-energy phosphate bond and renders the reaction effectively irreversible.

In general, DNA polymerases are highly accurate, with an intrinsic error rate of less than one mistake for every 107 nucleotides added. In addition, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a growing strand in order to correct mismatched bases. Finally, post-replication mismatch repair mechanisms monitor the DNA for errors, being capable of distinguishing mismatches in the newly synthesized DNA strand from the original

strand sequence. Together, these three discrimination steps enable replication fidelity of less than one mistake for every 109 nucleotides added.

The rate of DNA replication in a living cell was first measured as the rate of phage T4 DNA elongation in phage-infected E. coli. During the period of exponential DNA increase at 37 °C, the rate was 749 nucleotides per second. The mutation rate per base pair per replication during phage T4 DNA synthesis is 1.7 per 108.



Replication Process

DNA replication, like all biological polymerization processes, proceeds in three enzymatically catalyzed and coordinated steps: initiation, elongation and termination.

1- Initiation

Role of initiators for initiation of DNA replication. Formation of pre-replication complex. For a cell to divide, it must first replicate its DNA.[18] DNA replication is an all-or-none process; once replication begins, it proceeds to completion. Once replication is complete, it does not occur again in the same cell cycle. This is made possible by the division of initiation of the pre-replication complex.

2- Pre-replication complex

In late mitosis and early G1 phase, a large complex of initiator proteins assembles into the pre-replication complex at particular points in the DNA, known as "origins". In E. coli the primary initiator protein is DnaA; in yeast, this is the origin recognition complex. Sequences used by initiator proteins tend to be "AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) and thus are easier to strand-separate. In eukaryotes, the origin recognition complex catalyzes the assembly of initiator proteins into the pre-replication complex. Cdc6 and Cdt1 then associate with the bound origin recognition complex at the origin in order to form a larger complex necessary to load the Mcm complex onto the DNA. The Mcm complex is the helicase that will unravel

the DNA helix at the replication origins and replication forks in eukaryotes. The Mcm complex is recruited at late G1 phase and loaded by the ORC-Cdc6-Cdt1 complex onto the DNA via ATP-dependent protein remodeling. The loading of the Mcm complex onto the origin DNA marks the completion of pre-replication complex formation.

If environmental conditions are right in late G1 phase, the G1 and G1/S cyclin-Cdk complexes are activated, which stimulate expression of genes that encode components of the DNA synthetic machinery. G1/S-Cdk activation also promotes the expression and activation of S-Cdk complexes, which may play a role in activating replication origins depending on species and cell type. Control of these Cdks vary depending cell type and stage of development. This regulation is best understood in budding yeast, where the S cyclins Clb5 and Clb6 are primarily responsible for DNA replication. Clb5,6-Cdk1 complexes directly trigger the activation of replication origins and are therefore required throughout S phase to directly activate each origin.

In a similar manner, Cdc7 is also required through S phase to activate replication origins. Cdc7 is not active throughout the cell cycle, and its activation is strictly timed to avoid premature initiation of DNA replication. In late G1, Cdc7 activity rises abruptly as a result of association with the regulatory subunit Dbf4, which binds Cdc7 directly and promotes its protein kinase activity. Cdc7 has been found to be a rate-limiting regulator of origin activity. Together, the G1/S-Cdks and/or S-Cdks and Cdc7 collaborate to directly activate the replication origins, leading to initiation of DNA synthesis.

3- Pre-initiation complex

In early S phase, S-Cdk and Cdc7 activation lead to the assembly of the preinitiation complex, a massive protein complex formed at the origin. Formation of the preinitiation complex displaces Cdc6 and Cdt1 from the origin replication complex, inactivating and disassembling the pre-replication complex. Loading the preinitiation

complex onto the origin activates the Mcm helicase, causing unwinding of the DNA helix. The preinitiation complex also loads α-primase and other DNA polymerases onto the DNA. After α-primase synthesizes the first primers, the primer-template junctions interact with the clamp loader, which loads the sliding clamp onto the DNA to begin DNA synthesis. The components of the preinitiation complex remain associated with replication forks as they move out from the origin.

4- Elongation

DNA polymerase has 5′–3′ activity. All known DNA replication systems require a free 3′ hydroxyl group before synthesis can be initiated (note: the DNA template is read in 3′ to 5′ direction whereas a new strand is synthesized in the 5′ to 3′ direction—this is often confused). Four distinct mechanisms for DNA synthesis are recognized:

All cellular life forms and many DNA viruses, phages and plasmids use a primase to synthesize a short RNA primer with a free 3′ OH group which is subsequently elongated by a DNA polymerase. The retroelements (including retroviruses) employ a transfer RNA that primes DNA replication by providing a free 3′ OH that is used for elongation by the reverse transcriptase.

In the adenoviruses and the ϕ29 family of bacteriophages, the 3′ OH group is provided by the side chain of an amino acid of the genome attached protein (the terminal protein) to which nucleotides are added by the DNA polymerase to form a new strand.

In the single stranded DNA viruses—a group that includes the circoviruses, the geminiviruses, the parvoviruses and others—and also the many phages and plasmids that use the rolling circle replication (RCR) mechanism, the RCR endonuclease creates a nick in the genome strand (single stranded viruses) or one of the DNA strands (plasmids). The 5′ end of the nicked strand is transferred to a tyrosine residue on the nuclease and the free 3′ OH group is then used by the DNA polymerase to synthesize the new strand.

The first is the best known of these mechanisms and is used by the cellular organisms. In this mechanism, once the two strands are separated, primase adds

RNA primers to the template strands. The leading strand receives one RNA primer while the lagging strand receives several. The leading strand is continuously extended from the primer by a DNA polymerase with high processivity, while the lagging strand is extended discontinuously from each primer forming Okazaki fragments. RNase removes the primer RNA fragments, and a low processivity DNA polymerase distinct from the replicative polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

The primase used in this process differs significantly between bacteria and archaea/eukaryotes. Bacteria use a primase belonging to the DnaG protein superfamily which contains a catalytic domain of the TOPRIM fold type.[23] The TOPRIM fold contains an α/β core with four conserved strands in a Rossmann-like topology. This structure is also found in the catalytic domains of topoisomerase Ia, topoisomerase II, the OLD-family nucleases and DNA repair proteins related to the RecR protein.

The primase used by archaea and eukaryotes, in contrast, contains a highly derived version of the RNA recognition motif (RRM). This primase is structurally similar to many viral RNA-dependent RNA polymerases, reverse transcriptases, cyclic nucleotide generating cyclases and DNA polymerases of the A/B/Y families that are involved in DNA replication and repair. In eukaryotic replication, the primase forms a complex with Pol α.

Multiple DNA polymerases take on different roles in the DNA replication process. In E. coli, DNA Pol III is the polymerase enzyme primarily responsible for DNA replication. It assembles into a replication complex at the replication fork that exhibits extremely high processivity, remaining intact for the entire replication cycle. In contrast, DNA Pol I is the enzyme responsible for replacing RNA primers with DNA. DNA Pol I has a 5’ to 3’ exonuclease activity in addition to its polymerase activity, and uses its exonuclease activity to degrade the RNA primers ahead of it as it extends the DNA strand behind it, in a process called nick translation. Pol I is much less

processive than Pol III because its primary function in DNA replication is to create many short DNA regions rather than a few very long regions.

In eukaryotes, the low-processivity enzyme, Pol α, helps to initiate replication because it forms a complex with primase. In eukaryotes, leading strand synthesis is thought to be conducted by Pol ε; however, this view has recently been challenged, suggesting a role for Pol δ. Primer removal is completed Pol δ[27] while repair of DNA during replication is completed by Pol ε.

As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming a replication fork with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

5- The replication fork

The replication fork is a structure that forms within the long helical DNA during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together in the helix. The resulting structure has two branching "prongs", each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; the templates may be properly referred to as the leading strand template and the lagging strand template.

DNA is read by DNA polymerase in the 3’ to 5’ direction, meaning the new strand is synthesized in the 5' to 3' direction. Since the leading and lagging strand templates are oriented in opposite directions at the replication fork, a major issue is how to achieve synthesis of new lagging strand DNA, whose direction of synthesis is opposite to the direction of the growing replication fork.

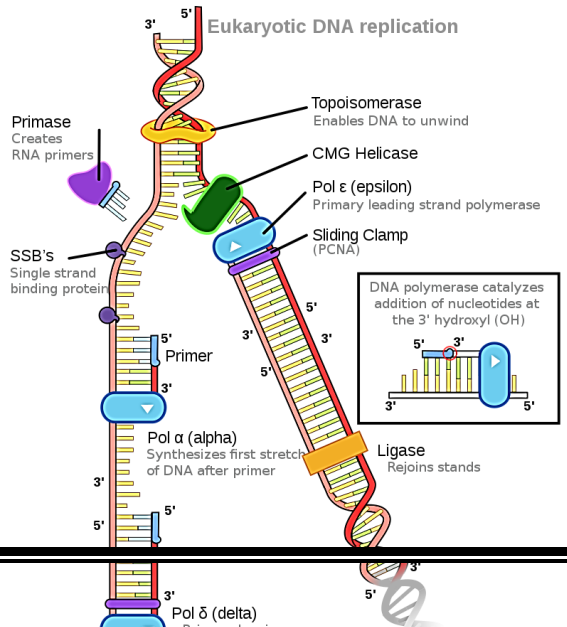
A- Leading strand

The leading strand is the strand of new DNA which is synthesized in the same direction as the growing replication fork. This sort of DNA replication is continuous.

B- Lagging strand

The lagging strand is the strand of new DNA whose direction of synthesis is opposite to the direction of the growing replication fork. Because of its orientation, replication of the lagging strand is more complicated as compared to that of the leading strand. As a consequence, the DNA polymerase on this strand is seen to "lag behind" the other strand.

The lagging strand is synthesized in short, separated segments. On the lagging strand template, a primase "reads" the template DNA and initiates synthesis of a short complementary RNA primer. A DNA polymerase extends the primed segments, forming Okazaki fragments. The RNA primers are then removed and replaced with DNA, and the fragments of DNA are joined together by DNA ligase.



Dynamics at the replication fork

The assembled human DNA clamp, a trimer of the protein PCNA.

In all cases the helicase is composed of six polypeptides that wrap around only one strand of the DNA being replicated. The two polymerases are bound to the helicase heximer. In eukaryotes the helicase wraps around the leading strand, and in prokaryotes it wraps around the lagging strand.

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead.[30] This build-up forms a torsional resistance that would eventually halt the progress of the replication fork. Topoisomerases are enzymes that temporarily break the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; topoisomerases (including DNA gyrase) achieve this by adding negative supercoils to the DNA helix.

Bare single-stranded DNA tends to fold back on itself forming secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

Double-stranded DNA is coiled around histones that play an important role in regulating gene expression so the replicated DNA must be coiled around histones at the same places as the original DNA. To ensure this, histone chaperones disassemble the chromatin before it is replicated and replace the histones in the correct place. Some steps in this reassembly are somewhat speculative.

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner

face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

DNA replication proteins

At the replication fork, many replication enzymes assemble on the DNA into a complex molecular machine called the replisome. The following is a list of major DNA replication enzymes that participate in the replisome

Enzyme	Function in DNA replication
DNA helicase	Also known as helix destabilizing enzyme. Helicase separates the two strands of DNA at the Replication Fork behind the topoisomerase.
DNA polymerase	The enzyme responsible for catalyzing the addition of nucleotide substrates to DNA in the 5’ to 3’ direction during DNA replication. Also performs proof-reading and error correction. There exist many different types of DNA Polymerase, each of which perform different functions in different types of cells.
DNA clamp	A protein which prevents elongating DNA polymerases from dissociating from the DNA parent strand.
Single-strand DNA-binding protein	Bind to ssDNA and prevent the DNA double helix from re-annealing after DNA helicase unwinds it, thus maintaining the strand separation, and facilitating the synthesis of the new strand.
Topoisomerase	Relaxes the DNA from its super-coiled nature.
DNA gyrase	Relieves strain of unwinding by DNA helicase; this is a specific type of topoisomerase

DNA ligase	Re-anneals the semi-conservative strands and joins Okazaki Fragments of the lagging strand.
Primase	Provides a starting point of RNA (or DNA) for DNA polymerase to begin synthesis of the new DNA strand.
Telomerase	Lengthens telomeric DNA by adding repetitive nucleotide sequences to the ends of **eukaryotic chromosomes**. This allows germ cells and stem cells to avoid the Hayflick limit on cell division.

Replication machinery

Replication machineries consist of factors involved in DNA replication and appearing on template ssDNAs. Replication machineries include primosotors are replication enzymes; DNA polymerase, DNA helicases, DNA clamps and DNA topoisomerases, and replication proteins; e.g. single-stranded DNA binding proteins (SSB). In the replication machineries these components coordinate. In most of the bacteria, all of the factors involved in DNA replication are located on replication forks and the complexes stay on the forks during DNA replication. These replication machineries are called replisomes or DNA replicase systems. These terms are generic terms for proteins located on replication forks. In eukaryotic and some bacterial cells the replisomes are not formed.

Since replication machineries do not move relatively to template DNAs such as factories, they are called a replication factory.[36] In an alternative figure, DNA factories are similar to projectors and DNAs are like as cinematic films passing constantly into the projectors. In the replication factory model, after both DNA helicases for leading strands and lagging strands are loaded on the template DNAs, the helicases run along the DNAs into each other. The helicases remain associated for the remainder of replication process. Peter Meister et al. observed directly replication sites in budding yeast by monitoring green fluorescent protein (GFP)-tagged DNA polymerases α. They detected DNA replication of pairs of the tagged loci

spaced apart symmetrically from a replication origin and found that the distance between the pairs decreased markedly by time. This finding suggests that the mechanism of DNA replication goes with DNA factories. That is, couples of replication factories are loaded on replication origins and the factories associated with each other. Also, template DNAs move into the factories, which bring extrusion of the template ssDNAs and new DNAs. Meister's finding is the first direct evidence of replication factory model. Subsequent research has shown that DNA helicases form dimers in many eukaryotic cells and bacterial replication machineries stay in single intranuclear location during DNA synthesis.

The replication factories perform disentanglement of sister chromatids. The disentanglement is essential for distributing the chromatids into daughter cells after DNA replication. Because sister chromatids after DNA replication hold each other by Cohesin rings, there is the only chance for the disentanglement in DNA replication. Fixing of replication machineries as replication factories can improve the success rate of DNA replication. If replication forks move freely in chromosomes, catenation of nuclei is aggravated and impedes mitotic segregation.

6- Termination

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome. Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes. Due to this problem, DNA is lost in each replication cycle from the end of the chromosome. Telomeres are regions of repetitive DNA close to the ends and help prevent loss of genes due to this shortening. Shortening of the telomeres is a normal process in somatic cells. This shortens the telomeres of the daughter DNA chromosome. As a result, cells can only divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation,

telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation. Increased telomerase activity is one of the hallmarks of cancer.

Termination requires that the progress of the DNA replication fork must stop or be blocked. Termination at a specific locus, when it occurs, involves the interaction between two components: (1) a termination site sequence in the DNA, and (2) a protein which binds to this sequence to physically stop DNA replication. In various bacterial species, this is named the DNA replication terminus site-binding protein, or Ter protein.

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. E. coli regulates this process through the use of termination sequences that, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome.

Dr. Naer Alkaabi