Biochemistry assignment

**initiation**

Formation of pre-replication complex.

For a [cell to divide](https://en.wikipedia.org/wiki/Cell_division), it must first replicate its DNA. 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](https://en.wikipedia.org/wiki/Pre-replication_complex).

**Pre-replication complex**

In late [mitosis](https://en.wikipedia.org/wiki/Mitosis) and early [G1 phase](https://en.wikipedia.org/wiki/G1_phase), a large complex of initiator proteins assembles into the pre-replication complex at particular points in the DNA, known as "[origins](https://en.wikipedia.org/wiki/Origin_of_replication)". In [*E. coli*](https://en.wikipedia.org/wiki/Escherichia_coli) the primary initiator protein is [DnaA](https://en.wikipedia.org/wiki/DnaA); in [yeast](https://en.wikipedia.org/wiki/Yeast), this is the [origin recognition complex](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/CDC6) and [Cdt1](https://en.wikipedia.org/wiki/DNA_replication_factor_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](https://en.wikipedia.org/wiki/Minichromosome_maintenance) onto the DNA. The Mcm complex is the helicase that will unravel the DNA helix at the replication origins and [replication forks](https://en.wikipedia.org/wiki/Replication_fork) 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](https://en.wikipedia.org/wiki/Cyclin)-[Cdk](https://en.wikipedia.org/wiki/Cyclin-dependent_kinase) 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](https://en.wikipedia.org/wiki/Budding_yeast), where the S cyclins [Clb5](https://en.wikipedia.org/wiki/Clb_5%2C6_%28Cdk1%29) and [Clb6](https://en.wikipedia.org/wiki/Clb_5%2C6_%28Cdk1%29) 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](https://en.wikipedia.org/wiki/Cell_division_cycle_7-related_protein_kinase) is also required through [S phase](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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.

**Preinitiation 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](https://en.wikipedia.org/wiki/DNA_polymerase_alpha) 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.

### Elongation

DNA polymerase has 5′–3′ activity. All known DNA replication systems require a free 3′ [hydroxyl](https://en.wikipedia.org/wiki/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:

1. All cellular life forms and many DNA [viruses](https://en.wikipedia.org/wiki/Virus), [phages](https://en.wikipedia.org/wiki/Phage) and [plasmids](https://en.wikipedia.org/wiki/Plasmid) use a [primase](https://en.wikipedia.org/wiki/Primase) to synthesize a short RNA primer with a free 3′ OH group which is subsequently elongated by a DNA polymerase.
2. The retroelements (including [retroviruses](https://en.wikipedia.org/wiki/Retrovirus)) employ a transfer RNA that primes DNA replication by providing a free 3′ OH that is used for elongation by the [reverse transcriptase](https://en.wikipedia.org/wiki/Reverse_transcriptase).
3. In the [adenoviruses](https://en.wikipedia.org/wiki/Adenovirus) and the φ29 family of [bacteriophages](https://en.wikipedia.org/wiki/Bacteriophage), 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.
4. In the single stranded DNA viruses—a group that includes the [circoviruses](https://en.wikipedia.org/wiki/Circovirus), the [geminiviruses](https://en.wikipedia.org/wiki/Geminivirus), the [parvoviruses](https://en.wikipedia.org/wiki/Parvovirus) and others—and also the many phages and [plasmids](https://en.wikipedia.org/wiki/Plasmid) 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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Processivity), while the lagging strand is extended discontinuously from each primer forming [Okazaki fragments](https://en.wikipedia.org/wiki/Okazaki_fragments). [RNase](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Ligase) works to fill these nicks in, thus completing the newly replicated DNA molecule.

The primase used in this process differs significantly between [bacteria](https://en.wikipedia.org/wiki/Bacteria) and [archaea](https://en.wikipedia.org/wiki/Archaea)/[eukaryotes](https://en.wikipedia.org/wiki/Eukaryote). Bacteria use a primase belonging to the [DnaG](https://en.wikipedia.org/wiki/DnaG) protein superfamily which contains a catalytic domain of the TOPRIM fold type. The TOPRIM fold contains an α/β core with four conserved strands in a [Rossmann-like](https://en.wikipedia.org/wiki/Rossmann_fold) topology. This structure is also found in the catalytic domains of [topoisomerase](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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*](https://en.wikipedia.org/wiki/Escherichia_coli), [DNA Pol III](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Pol_I) is the enzyme responsible for replacing RNA primers with DNA. DNA Pol I has a 5′ to 3′ [exonuclease](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/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](https://en.wikipedia.org/wiki/Eukaryote), 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 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](https://en.wikipedia.org/wiki/Replication_fork) with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process creates a "[theta structure](https://en.wikipedia.org/wiki/Theta_structure)" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

### Replication fork

Scheme of the replication fork.
a: template, b: leading strand, c: lagging strand, d: replication fork, e: primer, f: [Okazaki fragments](https://en.wikipedia.org/wiki/Okazaki_fragments)

Many enzymes are involved in the DNA 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 nascent 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 nascent (new) lagging strand DNA, whose direction of synthesis is opposite to the direction of the growing replication fork.

#### Leading strand

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

#### Lagging strand

The lagging strand is the strand of nascent 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](https://en.wikipedia.org/wiki/Primase) "reads" the template DNA and initiates synthesis of a short complementary [RNA](https://en.wikipedia.org/wiki/RNA) primer. A DNA polymerase extends the primed segments, forming [Okazaki fragments](https://en.wikipedia.org/wiki/Okazaki_fragment). The RNA primers are then removed and replaced with DNA, and the fragments of DNA are joined together by [DNA ligase](https://en.wikipedia.org/wiki/DNA_ligase).

#### Dynamics at the replication fork

The assembled human DNA clamp, a [trimer](https://en.wikipedia.org/wiki/Trimer_%28biochemistry%29) of the protein [PCNA](https://en.wikipedia.org/wiki/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.[[22]](https://en.wikipedia.org/wiki/DNA_replication#cite_note-replisome-in-Science-23)

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.[[23]](https://en.wikipedia.org/wiki/DNA_replication#cite_note-24) 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](https://en.wikipedia.org/wiki/DNA_gyrase)) achieve this by adding negative [supercoils](https://en.wikipedia.org/wiki/DNA_supercoil) to the DNA helix.

Bare single-stranded DNA tends to fold back on itself forming [secondary structures](https://en.wikipedia.org/wiki/Biomolecular_structure#Secondary_structure); these structures can interfere with the movement of DNA polymerase. To prevent this, [single-strand binding proteins](https://en.wikipedia.org/wiki/Single-strand_binding_protein) bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

[Clamp proteins](https://en.wikipedia.org/wiki/DNA_clamp) 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.

| **Important Enzymes in DNA Replication** |
| --- |
| **Enzyme** | **Function** |
| Topoisomerase | Relaxes the super-coiled DNA |
| DNA helicase | Unwinds the double helix at the replication fork |
| Primase | Provides the starting point for DNA polymerase to begin synthesis of the new strand |
| DNA polymerase | Synthesizes the new DNA strand; also proofreads and corrects some errors |
| DNA ligase | Re-joins the two DNA strands into a double helix and joins Okazaki fragments of the lagging strand |