

DNA REPLICATION

Mechanism, Enzymology, and a Detailed Comparison of Prokaryotic and Eukaryotic Replication

1. Introduction and General Principles

DNA replication is the biological process by which a cell copies its DNA before division, so that each daughter cell receives a complete and identical set of genetic instructions. It is one of the most accurate biochemical processes known, with an overall error rate of roughly one mistake per 10^9 – 10^{10} base pairs once proofreading and repair are taken into account.

(a) Semiconservative Replication

- Proposed by **Watson and Crick (1953)** as a natural consequence of the double-helix structure, and proven experimentally by **Matthew Meselson and Franklin Stahl (1958)** using $^{15}\text{N}/^{14}\text{N}$ density-labelled *E. coli* DNA and caesium chloride density-gradient centrifugation.
- Each of the two strands of the parental double helix acts as a **template**; the new double helix that results is therefore "semiconservative" — one old (parental/conserved) strand plus one newly synthesised strand.
- The Meselson–Stahl experiment ruled out the two rival models: **conservative replication** (the original duplex stays fully intact and an entirely new duplex is made) and **dispersive replication** (both strands end up as a patchwork of old and new segments).

(b) Origins, Bidirectionality, and Replication Bubbles

- Replication begins at a specific site called the **origin of replication**, where the double helix is locally unwound to form a **replication bubble**.
- From each origin, replication is generally **bidirectional** — two **replication forks** move outward in opposite directions from the origin, each fork being the Y-shaped point where the parental duplex is actively being unwound and copied.
- Bacteria typically have a **single origin** per chromosome, while eukaryotic chromosomes — being far longer — use **multiple origins** so that the entire genome can be copied within a limited window of the cell cycle (S phase).

(c) Base Pairing and Polarity Rules

- New nucleotides are added following strict **Watson–Crick base pairing**: A pairs with T, and G pairs with C.
- All known DNA polymerases synthesise new DNA only in the **5'→3' direction**, reading the template strand 3'→5'. This single rule is the root cause of the leading/lagging-strand asymmetry described below.
- DNA polymerases cannot start a new chain from scratch (*de novo*) — they can only **extend an existing 3'-OH end**. This is why every new stretch of DNA must first be primed by a short RNA primer laid down by a primase.
- The two strands of the parental duplex are **antiparallel** (one runs 5'→3', the other 3'→5'), so at a single replication fork the two template strands are copied by fundamentally different mechanisms — continuous on one side, discontinuous on the other.

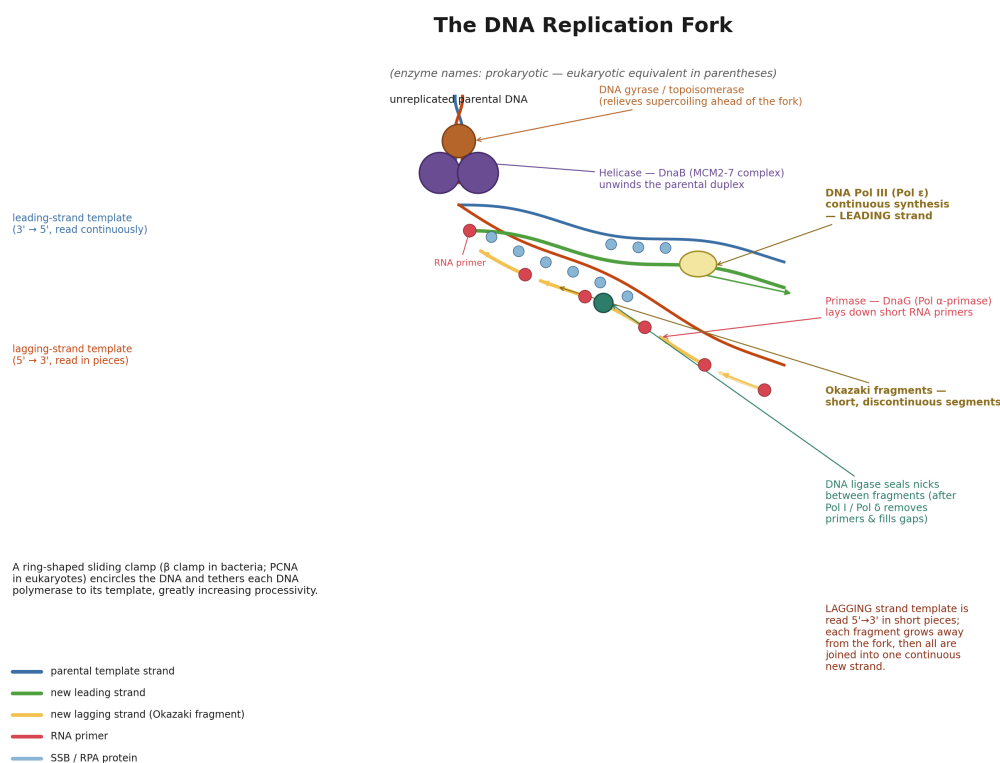


Fig. 1 — Structure of the replication fork: helicase unwinding, topoisomerase relief of supercoiling, continuous leading-strand synthesis, and discontinuous lagging-strand synthesis via Okazaki fragments.

(d) Leading and Lagging Strands

- **Leading strand** — synthesised continuously in the same direction as fork movement, needing only a single RNA primer at the origin.
- **Lagging strand** — synthesised discontinuously, in short stretches called **Okazaki fragments**, each primed separately and each synthesised in the direction away from the fork (i.e. back toward the previous fragment); the fragments are subsequently joined into one continuous strand.
- Okazaki fragments are named after **Reiji and Tsuneko Okazaki**, who first detected them in the late 1960s.

Why a lagging strand is needed at all

Because the two template strands run antiparallel but the fork opens up in only one direction, one template is exposed 3' → 5' (allowing continuous 5' → 3' synthesis — the leading strand) while the other is exposed 5' → 3' (which would require synthesis in the 3' → 5' direction if it were continuous). Since no polymerase can synthesise 3' → 5', this second strand is instead made in short, backward-priming bursts as the fork opens further — the lagging strand.

2. Prokaryotic DNA Replication (e.g. *Escherichia coli*)

Bacterial replication has been studied most thoroughly in *E. coli*, whose circular chromosome (~4.6 million base pairs) is replicated from a single origin at a rate of roughly **500–1000 nucleotides per second per fork**, completing replication in about 40 minutes.

(a) Initiation at *oriC*

- Replication starts at a unique ~245 bp origin sequence called ***oriC***, which contains AT-rich repeats that are relatively easy to melt (unwind) because A–T base pairs have only two hydrogen bonds.
- The initiator protein **DnaA** binds specific sequences within *oriC* and, with the help of ATP, causes local unwinding of the AT-rich region, forming an open complex.
- **DnaC** then loads the replicative helicase **DnaB** onto each of the exposed single strands, one helicase for each of the two forks that will move off in opposite directions.

(b) Unwinding and Stabilisation

- **DnaB helicase** uses ATP hydrolysis to translocate along the DNA and unwind the parental duplex ahead of each fork, breaking the hydrogen bonds between base pairs.
- Unwinding generates **positive supercoiling** ahead of the fork; this torsional strain is relieved by **DNA gyrase** (a type II topoisomerase), which introduces transient double-strand breaks to relax the supercoils, and by topoisomerase I, which relaxes negative supercoils behind the fork.
- **Single-strand binding proteins (SSB)** coat the exposed single-stranded template, preventing the strands from re-annealing or being degraded, and keeping them extended for the polymerase to read.

(c) Priming

- DNA polymerases cannot initiate a new strand unaided, so the primase **DnaG** synthesises short **RNA primers** (about 10–12 nucleotides) that provide the free 3'-OH group needed to start DNA synthesis.
- DnaB helicase and DnaG primase, together with accessory proteins, form a mobile complex sometimes called the **primosome**.

(d) Elongation — the Replisome

- The main replicative enzyme is **DNA polymerase III (Pol III) holoenzyme**, a large multi-subunit complex that synthesises both the leading and lagging strands with high speed and high processivity.
- Processivity is achieved through the β **sliding clamp**, a ring-shaped protein (loaded onto DNA by the clamp-loader complex) that encircles the DNA and tethers the polymerase to its template so it does not fall off after adding only a few nucleotides.
- On the **leading strand**, Pol III extends a single primer continuously in the same direction the fork is moving.
- On the **lagging strand**, Pol III synthesises one Okazaki fragment at a time (in bacteria, roughly 1000–2000 nucleotides long), starting each time from a new RNA primer laid down by DnaG; the template is looped so that the two Pol III core enzymes of the holoenzyme (one per strand) can travel together as a single replisome even though the lagging strand is being copied "backwards" relative to fork movement (the **trombone model**).

(e) Primer Removal, Gap Filling, and Ligation

- **DNA polymerase I (Pol I)** removes each RNA primer using its 5'→3' exonuclease activity and simultaneously fills the resulting gap with DNA using its 5'→3' polymerase activity (a process called nick translation). Pol I also possesses 3'→5' proofreading exonuclease activity.
- **DNA ligase** then seals the remaining nick between the 3'-OH end of the newly filled gap and the 5'-phosphate end of the adjacent Okazaki fragment, forming a continuous phosphodiester backbone.

(f) Termination

- Because the *E. coli* chromosome is circular, the two forks that started at *oriC* eventually meet on the opposite side of the circle, in a termination region containing several **ter sites**.
- The protein **Tus** binds to *ter* sites and acts as a one-way "molecular fork trap," allowing a fork to enter the terminus region but not to leave it, so that the two forks are made to stop close to one another.
- The two completed daughter circles, which are often interlinked (catenated), are separated by **topoisomerase IV**.

Enzyme / factor	Role in prokaryotic replication
DnaA	Recognises <i>oriC</i> and initiates unwinding of the origin
DnaB (helicase)	Unwinds the parental duplex ahead of the fork
DnaC	Loads DnaB helicase onto the origin
DNA gyrase (topoisomerase II)	Relieves positive supercoiling generated ahead of the fork
SSB protein	Coats and stabilises single-stranded template DNA
DnaG (primase)	Synthesises short RNA primers
DNA polymerase III	Main replicative enzyme; extends primers on both strands
β sliding clamp	Ring-shaped processivity factor that holds Pol III on the DNA
DNA polymerase I	Removes RNA primers and fills the resulting gaps with DNA
DNA ligase	Seals nicks between Okazaki fragments
Tus protein	Binds <i>ter</i> sites and halts the forks during termination
Topoisomerase IV	Decatenates (separates) the two interlinked daughter chromosomes

3. Eukaryotic DNA Replication

Eukaryotic replication follows the same fundamental chemistry as in bacteria, but is considerably more complex: the genome is far larger, it is packaged into chromatin, replication is confined to the **S phase** of the cell cycle, and it must be tightly regulated so that every stretch of DNA is copied **once, and only once**, per cell cycle.

(a) Multiple Origins and Licensing

- Each eukaryotic chromosome contains **hundreds to thousands of replication origins**, allowing the large genome to be copied within a limited S phase.
- During **G1 phase**, the **origin recognition complex (ORC)** binds each origin and, together with the loading factors **Cdc6** and **Cdt1**, recruits the **MCM2–7 complex** (the eukaryotic replicative helicase, loaded as an inactive double hexamer) to form the **pre-replication complex (pre-RC)** — a step called **origin licensing**.
- At the **G1/S transition**, cyclin-dependent kinases (CDKs) and the Dbf4-dependent kinase (DDK) activate a subset of licensed origins, converting MCM2-7 into an active helicase and recruiting further factors (Cdc45, the GINS complex — together forming the **CMG helicase**) to fire the origin.
- Once an origin has fired, re-licensing is blocked until the next G1, which is the key safeguard ensuring the genome is replicated exactly once per cycle.

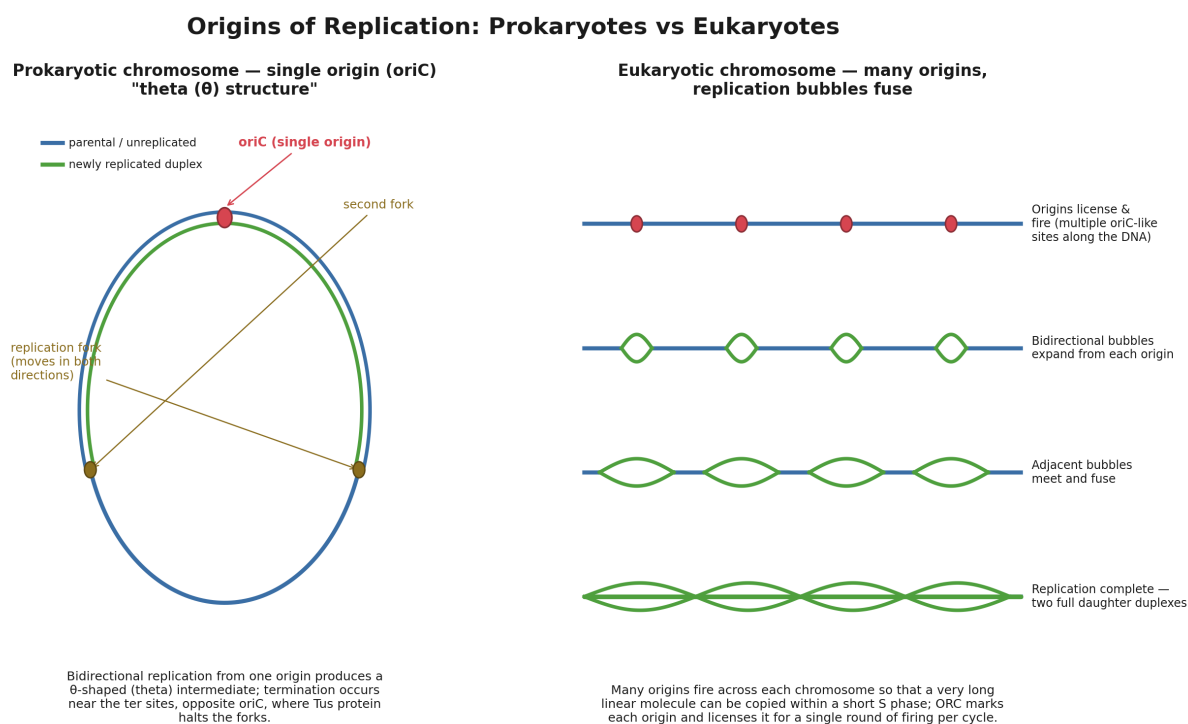


Fig. 2 — A single origin producing a theta-shaped bacterial chromosome versus many origins producing and fusing replication bubbles along a eukaryotic chromosome.

(b) Unwinding and Strand Stabilisation

- The **CMG (Cdc45–MCM2-7–GINS) helicase complex** unwinds the parental duplex at each active origin, moving outward as two forks (or a single fork near a chromosome end).
- Topoisomerases I and II** relieve the supercoiling generated ahead of the forks, analogous to bacterial gyrase.

- **Replication Protein A (RPA)** is the eukaryotic single-strand binding protein, coating and protecting the exposed template strands.

(c) Priming and Elongation — Three Main Polymerases

- **DNA polymerase α -primase (Pol α)** initiates synthesis: its primase subunit lays down a short RNA primer, and Pol α itself then extends this with a short stretch of DNA, producing a combined RNA–DNA primer.
- **Replication Factor C (RFC)** then loads the ring-shaped clamp **PCNA (Proliferating Cell Nuclear Antigen)** — the eukaryotic equivalent of the bacterial β clamp — onto the primer/template junction.
- PCNA hands off the primer to the main elongating polymerases: **DNA polymerase ϵ (Pol ϵ)**, which is thought to carry out the bulk of continuous **leading-strand** synthesis, and **DNA polymerase δ (Pol δ)**, which extends each Okazaki fragment on the **lagging strand**.
- Eukaryotic Okazaki fragments are much **shorter (about 100–200 nucleotides)** than bacterial ones (1000–2000 nt), roughly matching the ~200 bp spacing of nucleosomes along the DNA.

(d) Primer Removal, Gap Filling, and Ligation

- As Pol δ extends one Okazaki fragment and reaches the RNA primer of the previous one, it displaces the primer into a short single-stranded flap.
- The flap endonuclease **FEN1** cleaves off this flap, removing the RNA primer (with help from RNase H, which first degrades most of the RNA portion).
- **DNA ligase I** seals the remaining nick, joining adjacent fragments into a continuous lagging strand.

(e) Chromatin Reassembly

- Because eukaryotic DNA is packaged as chromatin, replication must be immediately followed by reassembly of nucleosomes on both daughter duplexes.
- The histone chaperone **CAF-1 (Chromatin Assembly Factor 1)**, recruited via PCNA, deposits new H3–H4 histone tetramers onto newly replicated DNA just behind the fork, while parental histones are recycled and redistributed to both daughters — helping to preserve epigenetic marks.

(f) Termination

- Forks moving from adjacent origins meet and fuse; the last few Okazaki fragments and primers between them are processed and ligated exactly as elsewhere along the lagging strand.
- Because eukaryotic chromosomes are **linear**, replication forks eventually run off the physical ends of the DNA, which creates the special problem addressed in Section 5 below — the end-replication problem.

Enzyme / factor	Role in eukaryotic replication
ORC, Cdc6, Cdt1	Recognise origins and license them by loading MCM2-7 (pre-RC formation)
MCM2-7 / CMG complex	The replicative helicase; unwinds the parental duplex
Topoisomerase I & II	Relieve supercoiling generated ahead of the forks
RPA	Single-strand binding protein; stabilises unwound template
Pol α-primase	Synthesises the initial RNA primer plus a short DNA extension
RFC	Clamp-loader that places PCNA at primer/template junctions

Enzyme / factor	Role in eukaryotic replication
PCNA	Ring-shaped sliding clamp; confers processivity on Pol δ and Pol ϵ
Pol ϵ	Main leading-strand polymerase
Pol δ	Main lagging-strand (Okazaki fragment) polymerase
FEN1 / RNase H	Remove RNA primers from Okazaki fragments
DNA ligase I	Seals nicks between adjacent Okazaki fragments
CAF-1	Assembles new nucleosomes on replicated DNA
Telomerase	Extends the 3' ends of linear chromosomes (see Section 5)

4. Prokaryotic vs Eukaryotic Replication — Side by Side

Feature	Prokaryotic (e.g. <i>E. coli</i>)	Eukaryotic
Genome / chromosome shape	Single circular chromosome	Multiple linear chromosomes
Number of origins	One (<i>oriC</i>)	Hundreds to thousands per genome
Initiator	DnaA	ORC + Cdc6 + Cdt1 (licensing)
Helicase	DnaB	MCM2-7 (as the CMG complex)
Primase	DnaG	Pol α -primase (primase + Pol α)
Single-strand binding protein	SSB	RPA
Main replicative polymerase(s)	DNA polymerase III	DNA polymerase ϵ (leading), Pol δ (lagging)
Sliding clamp	β clamp	PCNA
Clamp loader	γ complex	RFC
Primer removal	DNA polymerase I	FEN1 (with RNase H)
Ligase	DNA ligase	DNA ligase I
Okazaki fragment size	~1000–2000 nucleotides	~100–200 nucleotides
Replication rate per fork	~500–1000 nucleotides/second	~50 nucleotides/second
Timing in cell cycle	Continuous / coupled to cell growth	Restricted to S phase
Chromatin assembly	Not applicable (no histones)	CAF-1 deposits new nucleosomes behind the fork
End-replication problem	Absent (circular chromosome)	Present — solved by telomerase
Termination	Tus protein at <i>ter</i> sites; topoisomerase IV decatenates	Fusion of forks from adjacent origins

5. The End-Replication Problem and Telomerase

Linear eukaryotic chromosomes face a problem that circular bacterial chromosomes do not: the very mechanism that makes the lagging strand possible causes the extreme 5' end of each new strand to go unreplacated, so chromosomes would shorten with every round of division if nothing counteracted it.

The End-Replication Problem and Telomerase

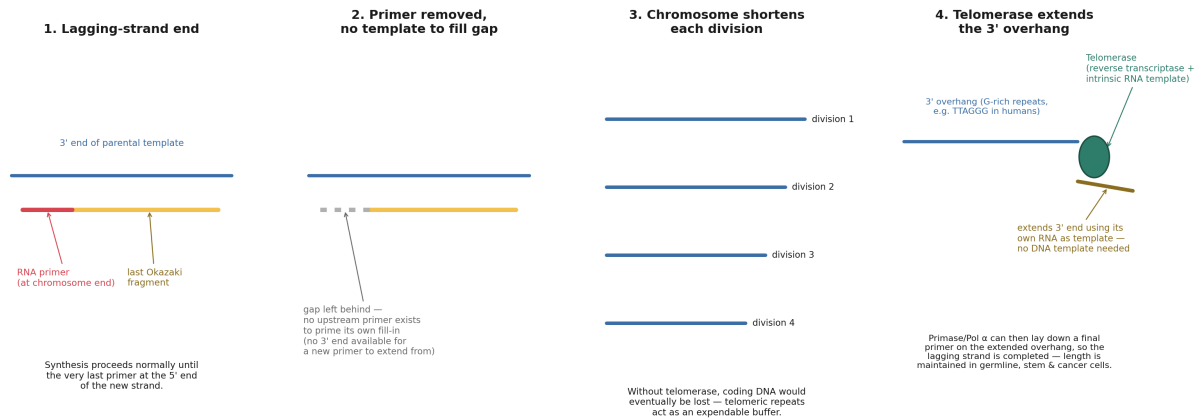


Fig. 3 — The end-replication problem: the terminal RNA primer on the lagging strand cannot be replaced by normal DNA synthesis, leading to progressive shortening unless telomerase intervenes.

(a) Why the Problem Arises

- At the very end of a linear template, the lagging strand's final Okazaki fragment is still primed by a short RNA primer sitting right at the chromosome tip.
- When that primer is later removed (by FEN1/RNase H), there is no further upstream 3'-OH from which a polymerase could fill the resulting gap — unlike every other gap along the lagging strand, which is filled by extending the previous fragment.
- The result is that each round of replication leaves the new lagging strand slightly shorter at that end, so the daughter chromosome is a little shorter than its parent.

(b) Telomeres as a Buffer

- The ends of eukaryotic chromosomes are capped by **telomeres** — long stretches of short, non-coding, tandemly repeated sequences (in humans, the repeat is **5'-TTAGGG-3'**), ending in a single-stranded 3' overhang.
- Because telomeric DNA does not code for protein, its gradual erosion over many cell divisions acts as an expendable buffer that protects adjacent essential genes from being lost.
- Telomeres also fold back and are capped by protective proteins (in humans, the **shelterin** complex), which distinguish natural chromosome ends from broken DNA and prevent the ends from triggering DNA-damage responses or end-to-end fusions.

(c) Telomerase — Restoring the Ends

- Telomerase** is a specialised ribonucleoprotein reverse transcriptase, discovered by **Elizabeth Blackburn and Carol Greider (1984)**, that can extend the 3' end of the DNA using an **intrinsic RNA template** carried within the enzyme itself, rather than copying an existing DNA template.

- It repeatedly adds telomeric repeat units to the 3' overhang; once the overhang is long enough, conventional primase/Pol α -primed synthesis can then fill in a complementary strand, restoring (or even lengthening) the chromosome end.
- Telomerase activity is high in **germline cells, stem cells, and most cancer cells** (allowing indefinite division), but is low or absent in most normal somatic cells, which is why human somatic chromosomes shorten gradually with age — a process linked to cellular senescence.

Clinical / biological relevance

Telomere shortening is considered a contributor to the ageing of cells (the "Hayflick limit" on the number of divisions a normal cell can undergo), and reactivation of telomerase is one of the hallmark changes seen in the majority of human cancers, making it an active target for anti-cancer drug development.

6. Quick Revision Summary

Aspect	Key point
Mode of replication	Semiconservative (Meselson & Stahl, 1958); each daughter duplex has one old + one new strand
Direction of synthesis	Always 5'→3'; polymerases can only extend an existing 3'-OH, hence the need for primers
Leading strand	Continuous synthesis from a single primer, in the direction of fork movement
Lagging strand	Discontinuous synthesis as Okazaki fragments, each separately primed
Prokaryotic origin	Single oriC; initiator DnaA; helicase DnaB; main polymerase Pol III
Eukaryotic origins	Many, licensed in G1 by ORC/Cdc6/Cdt1/MCM2-7, fired in S phase
Prokaryotic vs eukaryotic Okazaki fragments	~1000–2000 nt vs ~100–200 nt respectively
Primer removal	Pol I (prokaryotes) vs FEN1/RNase H (eukaryotes); ligase seals the final nick
Sliding clamp	β clamp (prokaryotes) vs PCNA (eukaryotes) — confers processivity
Linear-chromosome problem	End-replication problem, solved by telomerase adding TTAGGG repeats to 3' ends

Compiled study note — molecular biology reference. Figures are original schematic diagrams prepared for this note and are not to scale.