MOLECULAR BIOLOGY: DNA STRUCTURE AND DNA REPLICATION

Lecture 1
Lecture slides

Nucleotides and Nucleic Acids

- Nucleotides are building blocks of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).
- Nucleotides have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose sugar, and (3) a phosphate.
- The molecule without the phosphate group is called a **nucleoside**.
- The nitrogenous bases are derivatives of two parent compounds, pyrimidine and purine.
- In the pentoses of nucleotides and nucleosides the carbon numbers are given a prime (') designation to distinguish them from the numbered atoms of the nitrogenous base.

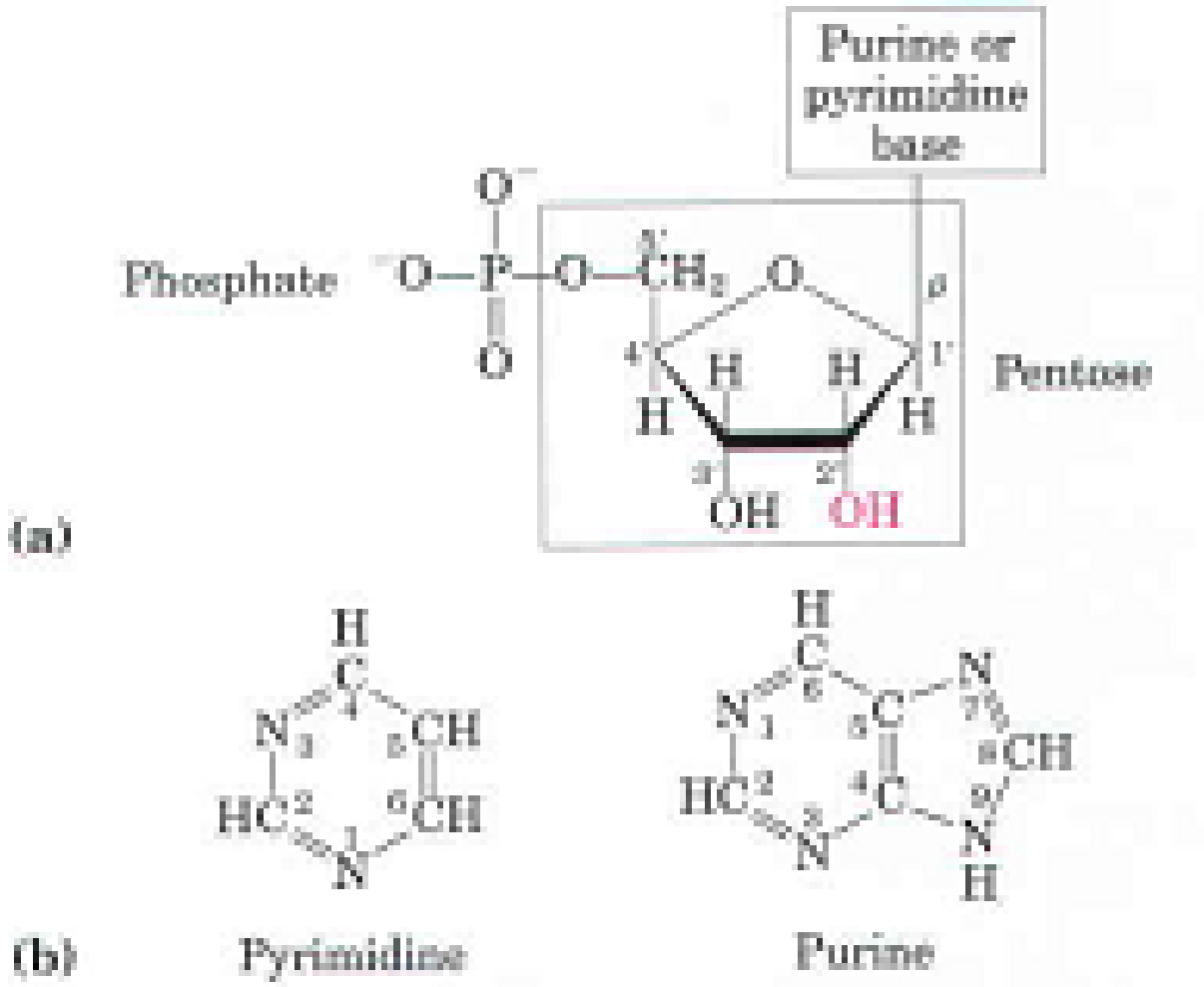


Figure 1.1 Structure of nucleotides. (a) General structure showing the numbering convention for the pentose (ribose in this case) ring. (b) The parent compounds of the pyrimidine and purine bases of nucleotides and nucleic acids, showing the numbering conventions.

- The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an N-β-glycosyl bond to the 1' carbon of the pentose, and the phosphate is esterified to the 5' carbon.
- The N-β-glycosyl bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base).
- Both DNA and RNA contain two major purine bases, adenine (A) and guanine (G), and two major pyrimidines. In both DNA and RNA one of the pyrimidines is cytosine (C), but the second major pyrimidine is not the same in both: it is **thymine** (T) in DNA and **uracil** (U) in RNA.
- The structures of the five major bases are shown in Figure 1.2 and the nomenclature of their corresponding nucleosides and nucleotides is shown in Table 1.1

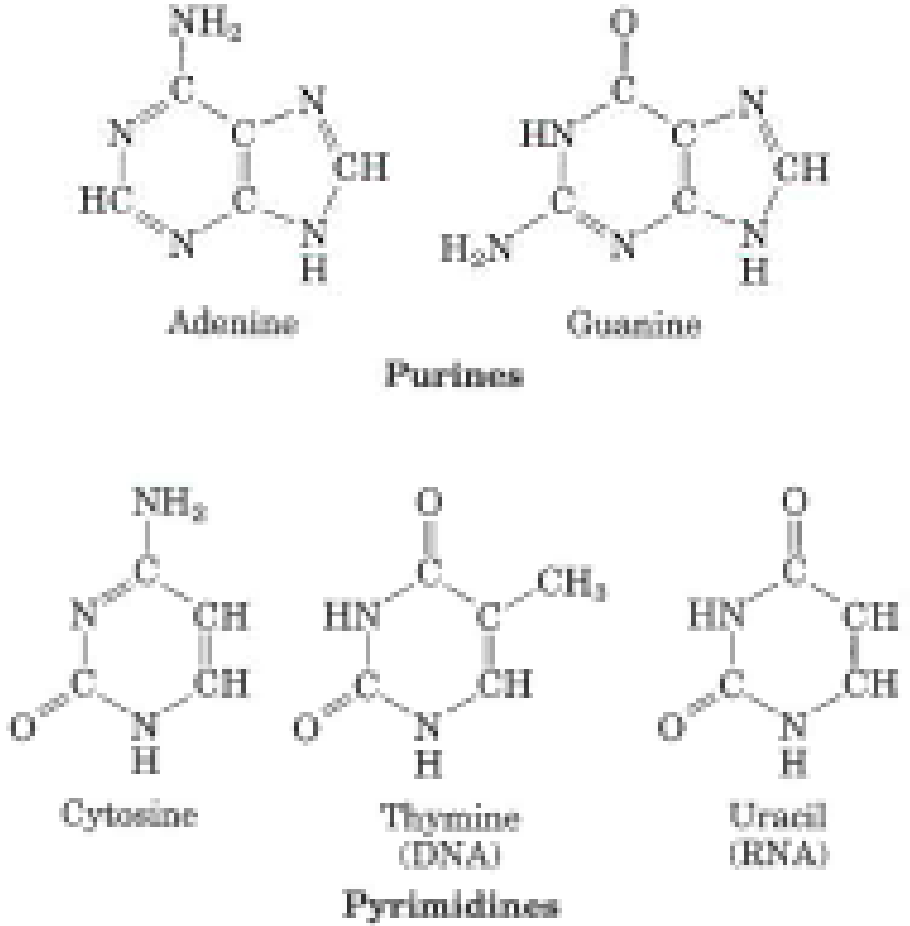


Figure 1.2 Major purine and pyrimidine bases of nucleic acid

Table 1.1 Nucleotide and nucleic acid nomenclature

Base	Nucleoside	Nucleotide	Nucleic acid
Purines			
Adenine	Adenosine	Adenylate	RNA
	Deoxyadenosine	Deoxyadenylate	DNA
Guanine	Guanosine	Guanylate	RNA
	Deoxyguanosine	Deoxyguanylate	DNA
Pyrimidines			
Cytosine	Cytidine	Cytidylate	RNA
	Deoxycytidine	Deoxycytidylate	DNA
Thymine	Thymidine or deoxythymidine	Thymidylate or deoxythymidylate	DNA
Uracil	Uridine	Uridylate	RNA

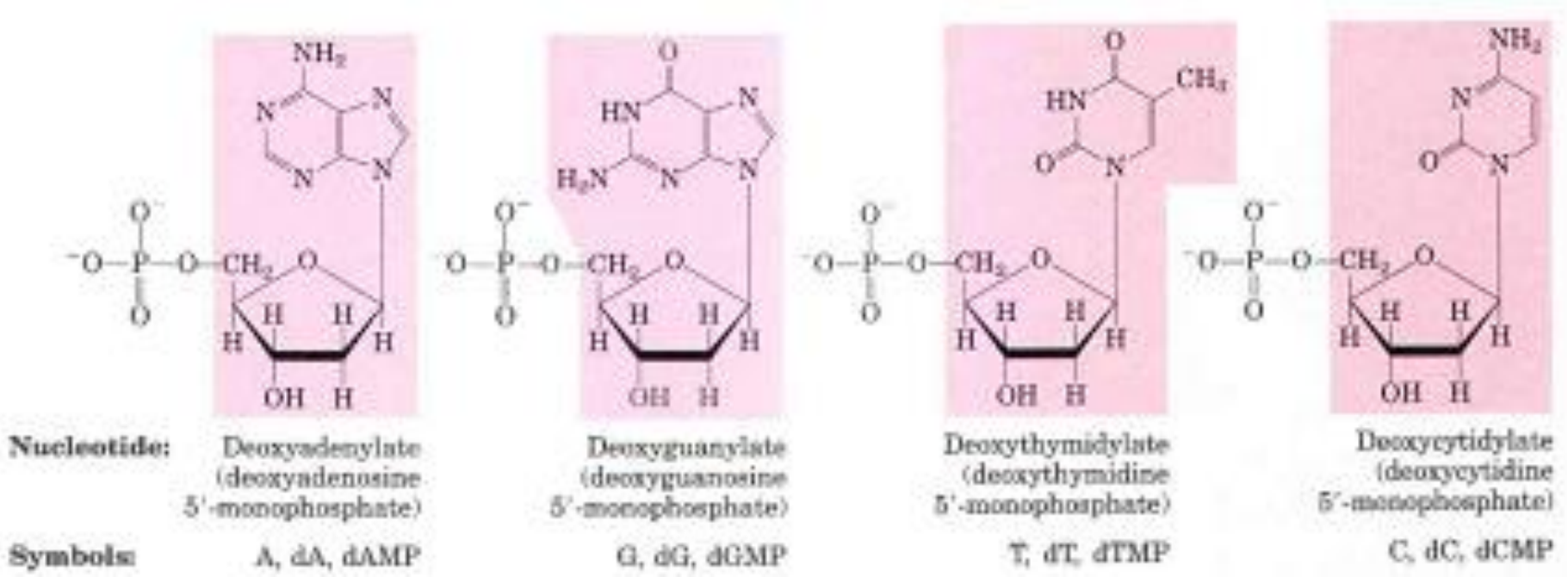


Figure 1.3 Deoxyribonucleotides of nucleic acid (i.e. DNA)

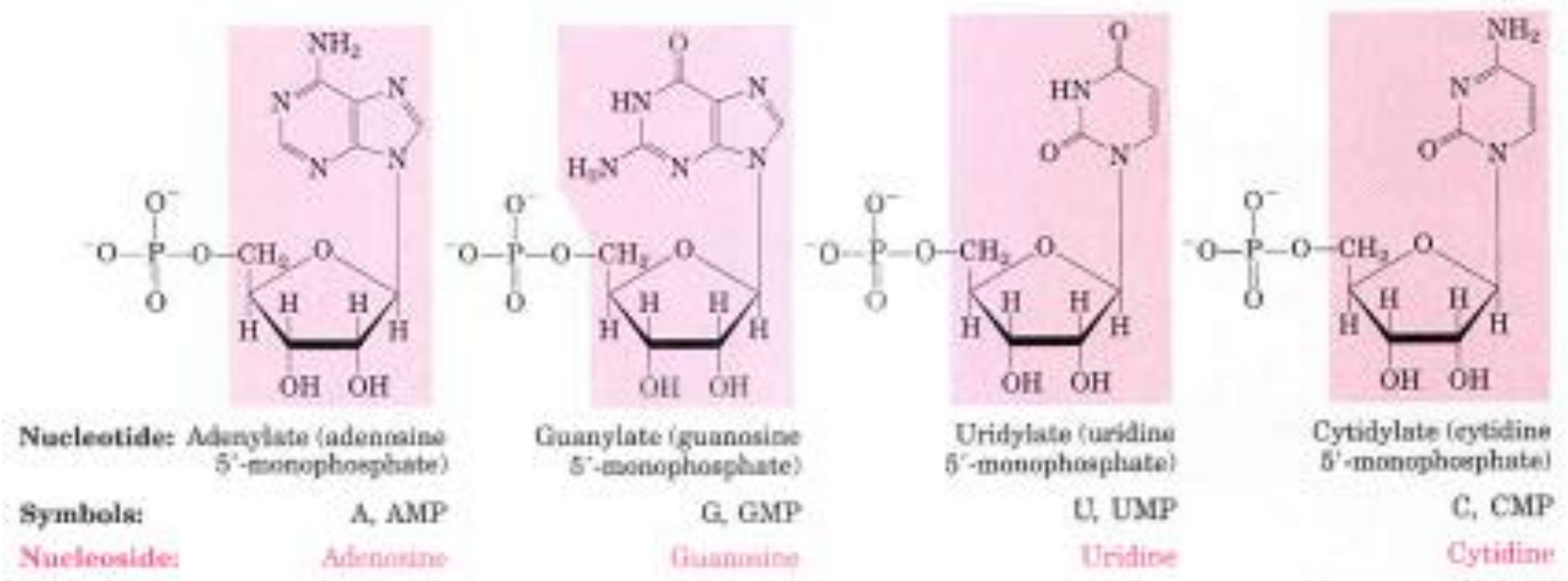


Figure 1.4 Ribonucleotides of nucleic acid (i.e. RNA)

- The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group "bridges," in which the 5'-phosphate group of one nucleotide unit is joined to the 3'-hydroxyl group of the next nucleotide, creating a phosphodiester linkage (Figure 1.5).
- All the phosphodiester linkages in DNA and RNA have the same orientation along the chain (Figure 1.5).
- Two nucleotides connected to each other are called a dinucleotide and three are called a trinucleotide.
- A short nucleic acid is referred to as an oligonucleotide. The definition of "short" is somewhat arbitrary but polymers containing 50 or fewer nucleotides are generally called oligonucleotides. Numerous nucleotides connected in a long chain is termed a polynucleotide.
- The sequence of a single strand of nucleic acid is always written with the 5' end at the left and the 3' end at the right-that is, in the 5' → 3’ direction.

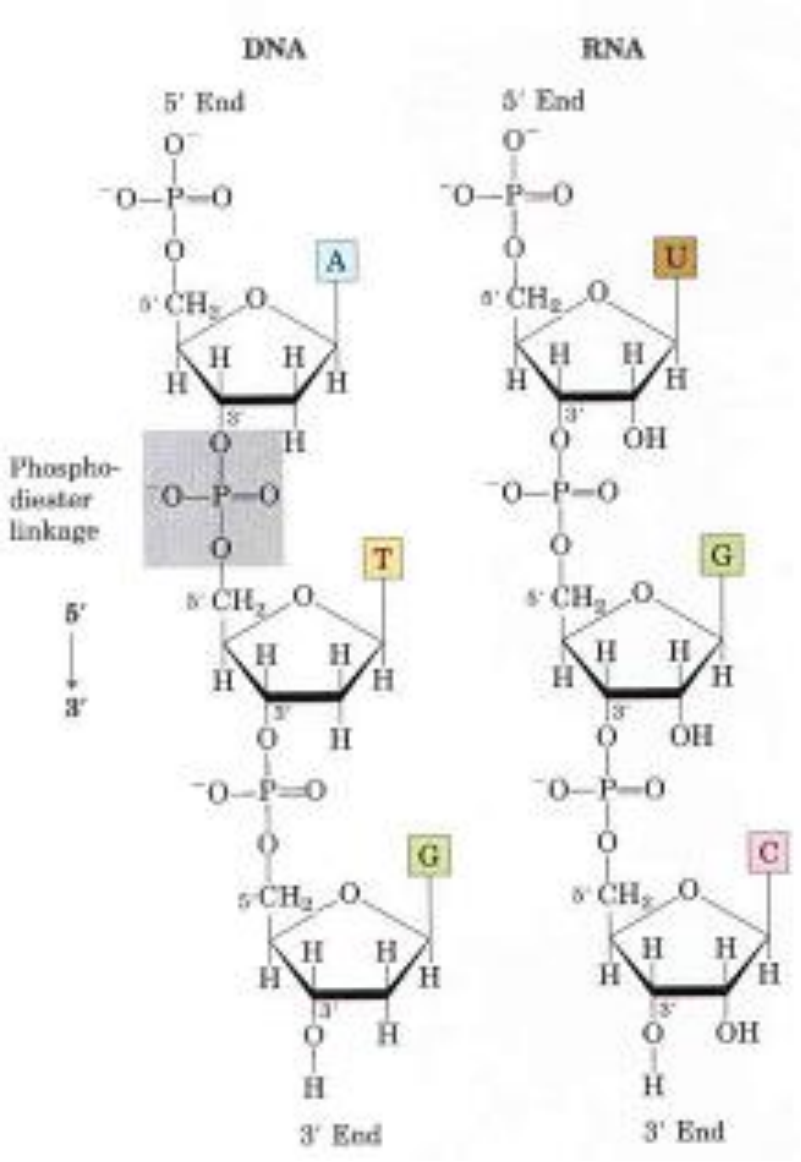


Figure 1.5 Phosphodiester linkages in the covalent backbone of DNA and RNA

- In the early 1950s, the biochemist Erwin Chargaff performed experiments and showed that the relative ratios of the four bases in DNA were not equal, but were also not random.
- The number of adenine (A) residues in all DNA samples was equal to the number of thymine (T) residues, while the number of guanine (G) residues equaled the number of cytosine (C) residues.
- Chargaff's rules state that for any given species:

1. A = T and G = C
2. Sum of the purines = sum of the pyrimidines
3. The percentage of (C + G) does not necessarily equal the percentage of (A + T).

The antiparallel double helix of DNA

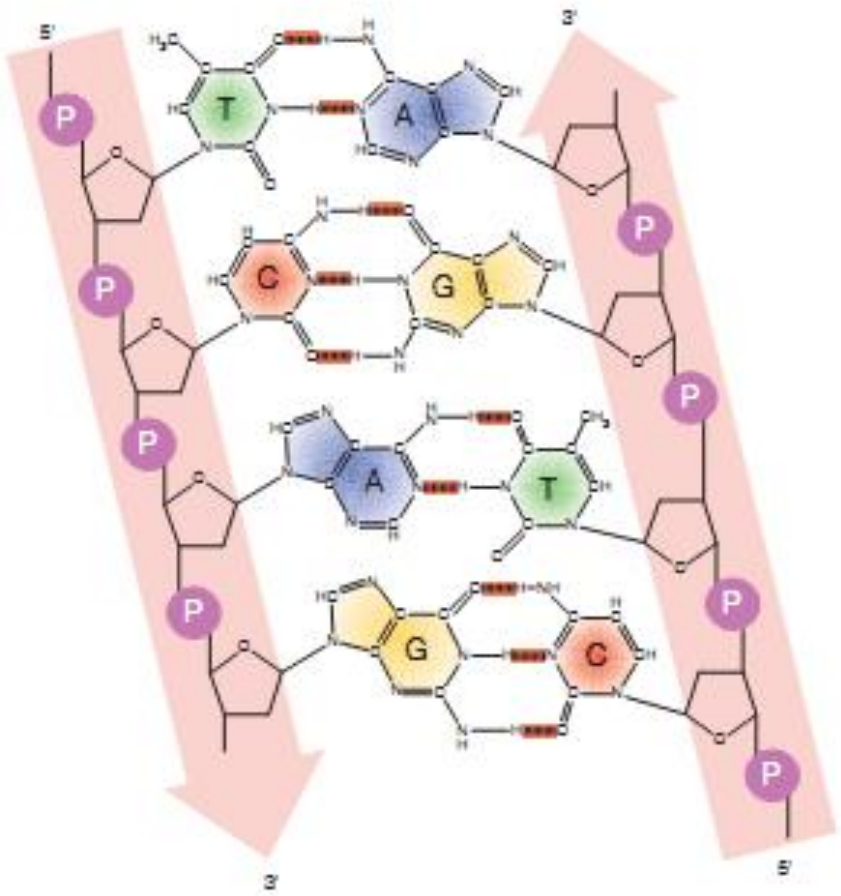


Figure 1.6 DNA base pairing and complementation. The two chains of the helix, arrowed in the 5’ to 3’ direction, are antiparallel. The bases on one strand of the helix are complementary to those on the opposite strand, A always base pairs with T and G always base pairs with C

DNA Replication

- One strand of DNA is the complement of the other. The strict base-pairing rules mean that each strand provides the template for a new strand with a predictable and complementary sequence.
- The fundamental properties of the DNA replication process and the mechanisms used by the enzymes that catalyze it are essentially identical in all species.
- DNA Replication follows a set of fundamental rules:

1. **DNA Replication is semiconservative**

   Each DNA strand serves as a template for the synthesis of a new strand, producing two new DNA molecules, each with one new strand and one old strand. This is semiconservative replication.

The Meselson-Stahl experiment

- Meselson and Stahl grew *E. coli*, cells for many generations in a medium in which the sole nitrogen source (NH_4Cl) contained ^{15}N, the "heavy" isotope of nitrogen, instead of the normal, more abundant "light" isotope, ^{14}N.
- The *E. coli*, cells grown in the ^{15}N medium were transferred to a fresh medium containing only the ^{14}N isotope where they were allowed to grow until the cell population doubled.
- The DNA isolated from these first-generation cells formed a single band in the CsCl gradient at a position indicating that the double-helical DNA molecules of the daughter cells were hybrids containing one new ^{14}N strand and one parent ^{15}N strand.
- Cells were again allowed to double in number in the ^{14}N medium. The isolated DNA product of this second cycle of replication exhibited 2 bands in the density gradient, one with a density equal to that of light DNA and the other with the density of the hybrid DNA.

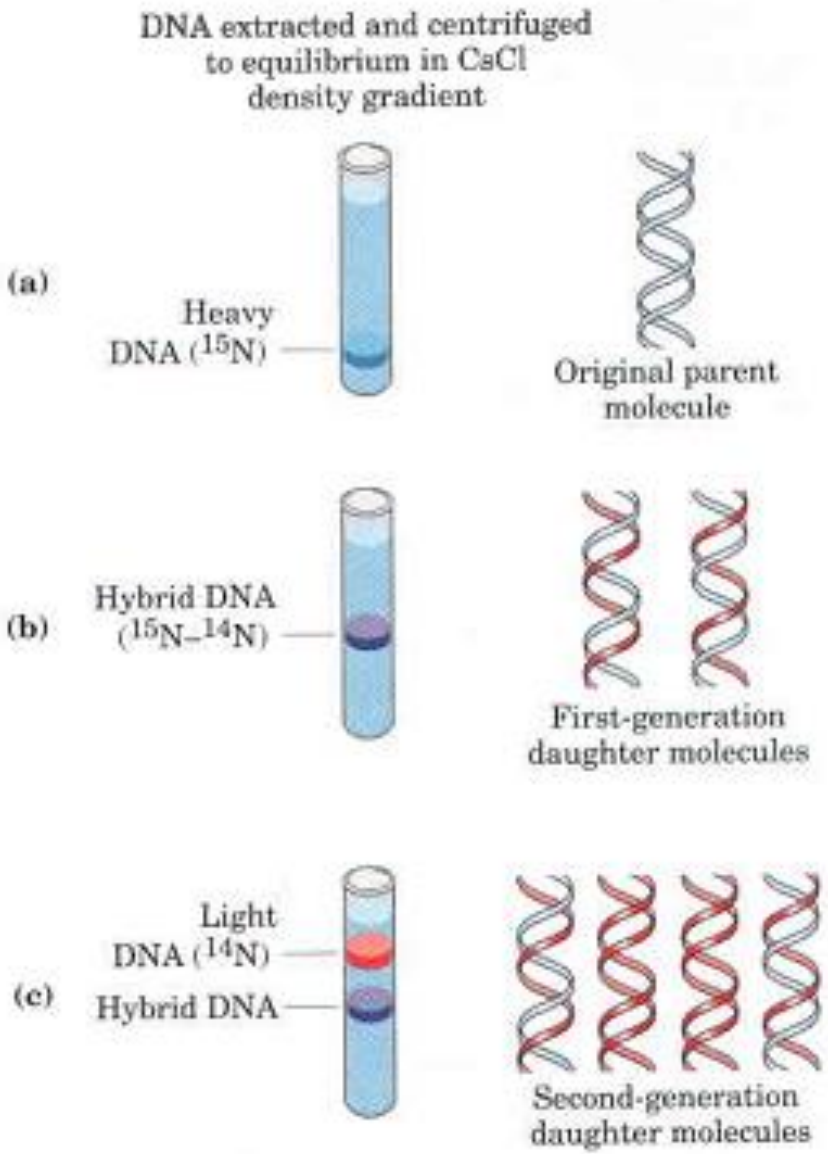
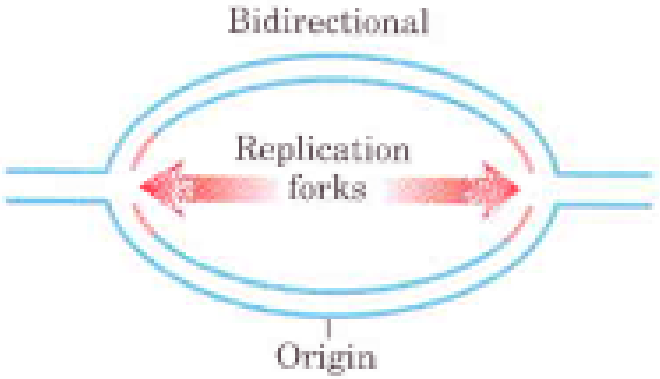


Figure 1.7 (a) Cells were grown for many generations in a medium containing only heavy nitrogen, so that all the nitrogen in their DNA was ^{15}N, as shown by a single band (blue) when centrifuged in a CsCl density gradient.
(b) Once the cells had been transferred to a medium containing only light nitrogen, ^{14}N, cellular DNA isolated after one generation equilibrated at a higher position in the density gradient (purple band).
(c) A second cycle of replication yielded a hybrid DNA band (purple) and another band (red), containing only ^{14}N DNA, confirming semiconservative replication

2 **Replication begins at an origin and usually proceeds bidirectionally**



3. **DNA Synthesis proceeds in a 5’ → 3’ direction and is semidiscontinuous**

A new strand of DNA is always synthesized in the 5’→3' direction, with the free 3’-OH as the point at which the DNA is elongated.

Because the two DNA strands are antiparallel, the strand serving as the template is read from its 3' end toward its 5’ end.

- one strand is synthesized continuously and the other discontinuously. The continuous strand, or leading strand, is the one in which 5' → 3' synthesis proceeds in the same direction as replication fork movement.
- The discontinuous strand, or lagging strand, is the one in which 5' → 3' synthesis proceeds in the direction opposite to the direction of fork movement. These DNA strands synthesized in short pieces are called **Okazaki fragments**.

4. **DNA is degraded by nucleases**
5. **DNA is synthesized by DNA Polymerases**
6. **Replication is very accurate**. DNA polymerase has a proof reading activity

The Replication Process

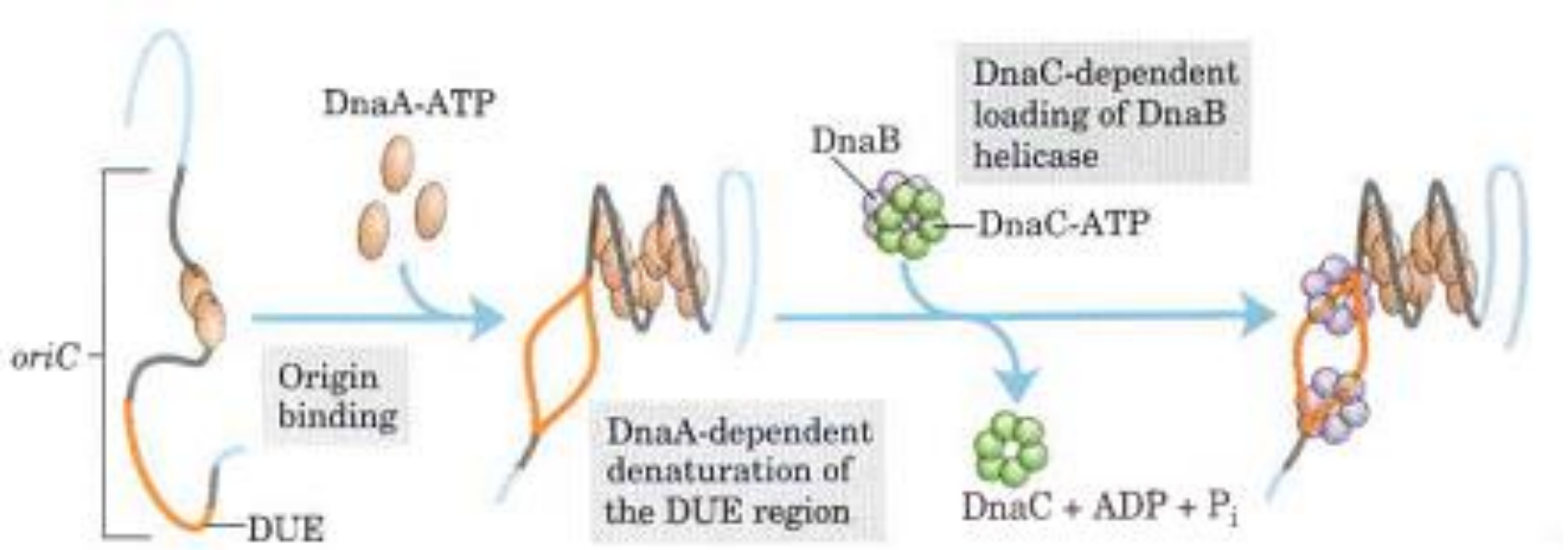
- The synthesis of a DNA molecule can be divided into three stages: **initiation**, **elongation**, and **termination**, distinguished both by the reactions taking place and by the enzymes required.

Initiation of replication

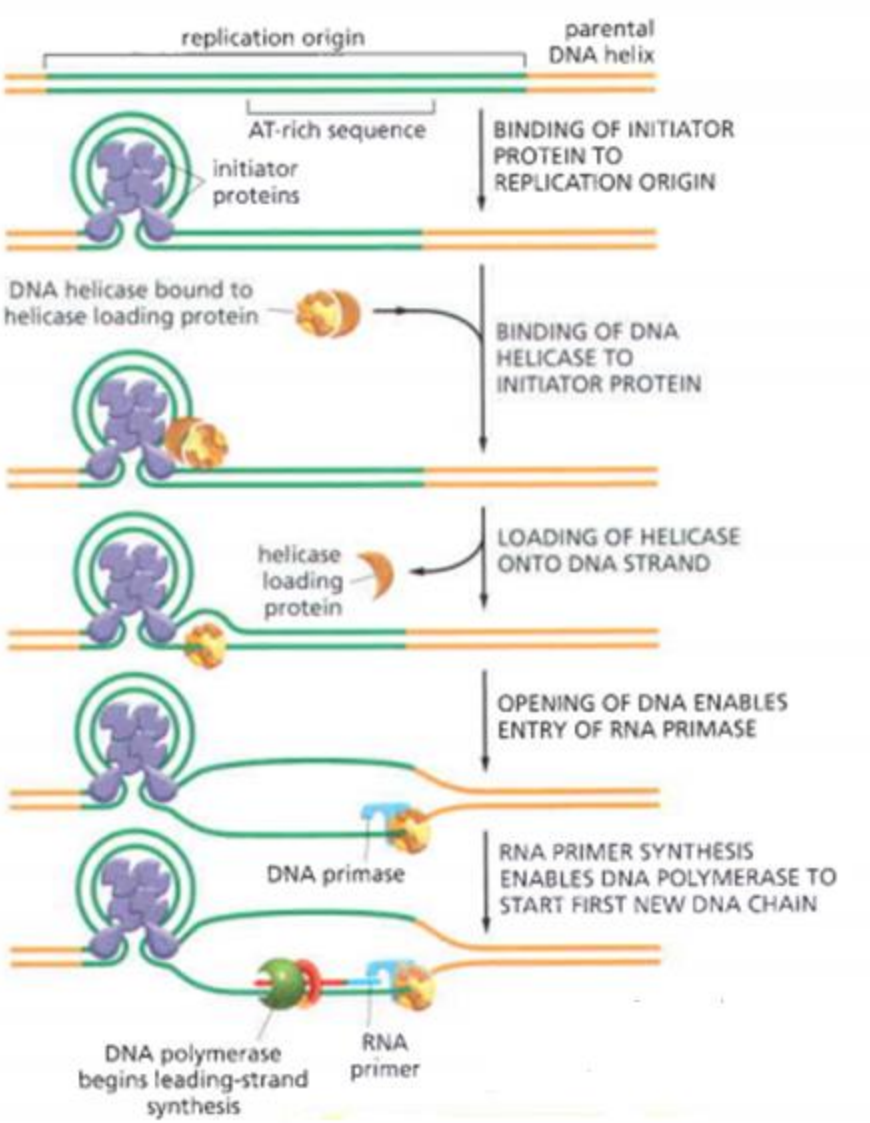
- The initiation of DNA replication occurs at specific points called origins of replication (e.g. OriC in *E. coli*).
- Once DNA synthesis has been initiated, two replication forks, extending in either direction from the origin of replication, proceed to allow the full replication of the genome.
- OriC is the binding site of proteins DnaA, B and C that promote the melting (opening) of the DNA helix, a process that is essential so that DNA replicating enzymes can read the base sequence.
- The polymerase can only function if a free 3’-OH group is present.

- This hydroxyl group is provided by an RNA primer (which is complementary to the DNA) that is 5–15 nucleotides long.
- The synthesis of the primer is directed by a form of RNA polymerase (called **primase**).
- DNA is unwound into the polymerase complex with the help of **DNA helicases**.
- **Topoisomerase** enzymes (e.g. DNA gyrase) are required to relieve tension in the helix that results as a consequence of the unwinding process.
- Single-stranded DNA produced during replication are stabilized through the binding of single-stranded binding proteins (SSBs)

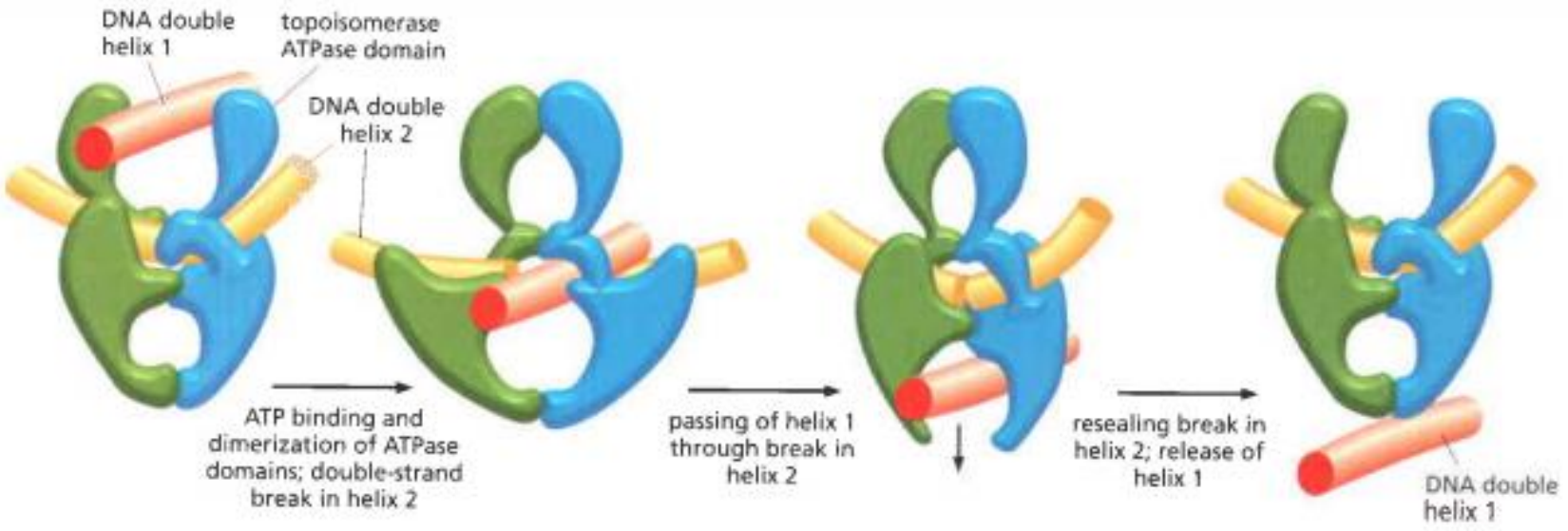
Model for initiation of replication at the *E coli* origin, oriC



DnaA = Initiator protein
DnaB = Helicase
DnaC =is helicase loading protein


replication origin
parental DNA helix
AT-rich sequence
initiator proteins
BINDING OF INITIATOR PROTEIN TO REPLICATION ORIGIN
DNA helicase bound to helicase loading protein
BINDING OF DNA HELICASE TO INITIATOR PROTEIN
helicase loading protein
LOADING OF HELICASE ONTO DNA STRAND
OPENING OF DNA ENABLES ENTRY OF RNA PRIMASE
DNA primase
RNA PRIMER SYNTHESIS ENABLES DNA POLYMERASE TO START FIRST NEW DNA CHAIN
DNA polymerase begins leading-strand synthesis
RNA primer

A model for topoisomerase II action



- Type II topoisomerases are largely confined to proliferating cells in eukaryotes; partly for that reason they have been popular targets for anticancer drugs

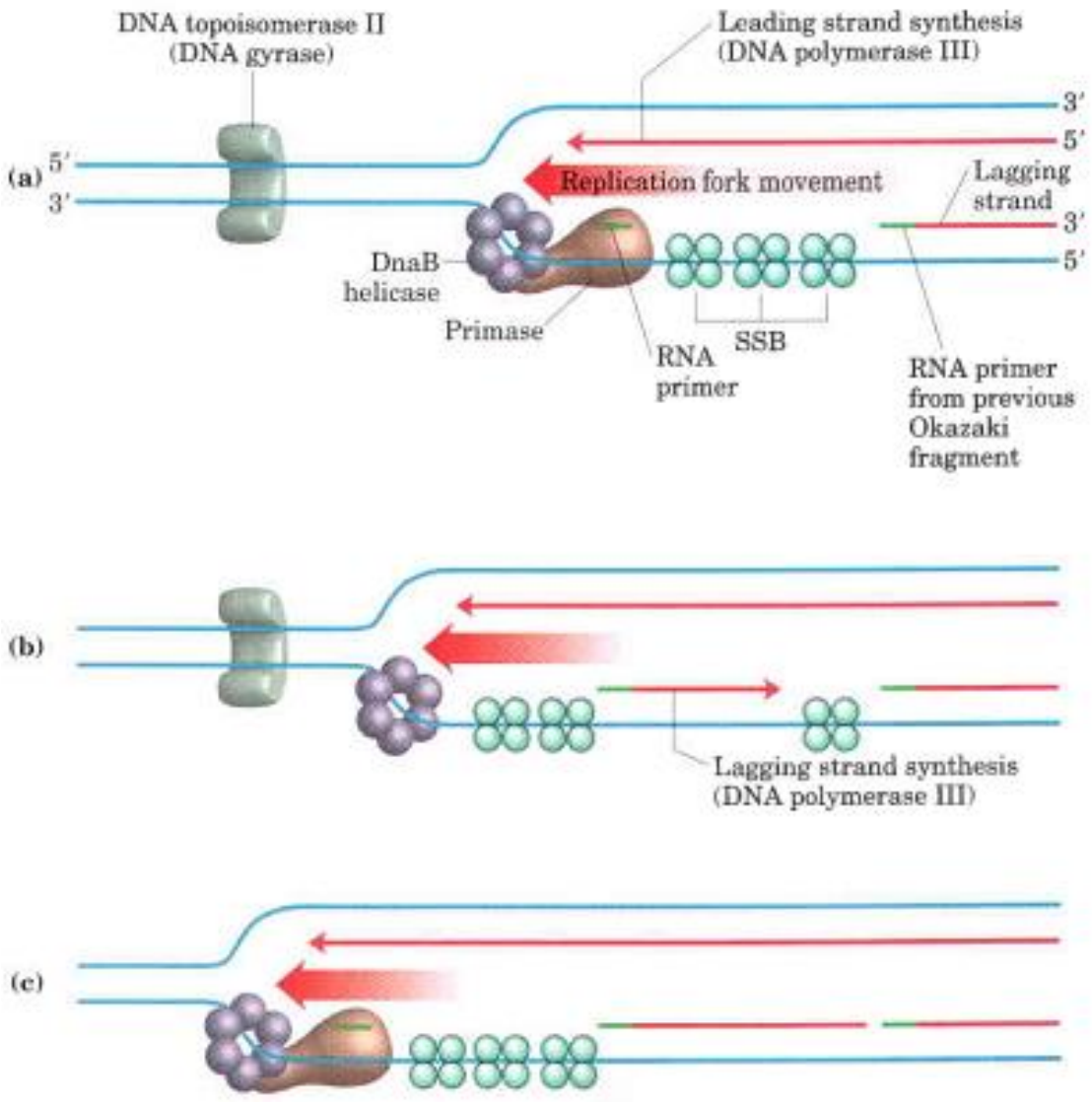
Table 1.2 Proteins required to initiate replication at the *E. coli* origin

Protein	M_r	Number of subunits	Function
DnaA protein	52,000	1	Recognizes ori sequence; opens duplex at specific sites in origin
DnaB protein (helicase)	300,000	6*	Unwinds DNA
DnaC protein	174,000	6*	Required for DnaB binding at origin
HU	19,000	2	Histonelike protein; DNA-binding protein; stimulates initiation
FIS	22,500	2*	DNA-binding protein; stimulates initiation
IHF	22,000	2	DNA-binding protein; stimulates initiation
Primase (DnaG protein)	60,000	1	Synthesizes RNA primers
Single-stranded DNA-binding protein (SSB)	75,600	4*	Binds single-stranded DNA
DNA gyrase (DNA topoisomerase II)	400,000	4	Relieves torsional strain generated by DNA unwinding

Elongation

- The elongation phase of replication includes two distinct but related operations: leading strand synthesis and lagging strand synthesis.
- Leading strand synthesis begins with the synthesis by primase of a short (10 to 60 nucleotide) RNA primer at the replication origin.
- Deoxyribonucleotides are added to this primer by a DNA polymerase III.
- Leading strand synthesis then proceeds continuously, keeping pace with the unwinding of DNA at the replication fork.
- The lagging strand is formed so that nucleotide polymerization can occur on both template strands in a 5'to 3'direction.
- DNA ligase is then required to join the phosphodiester backbone of the Okazaki fragments to form a complete strand.

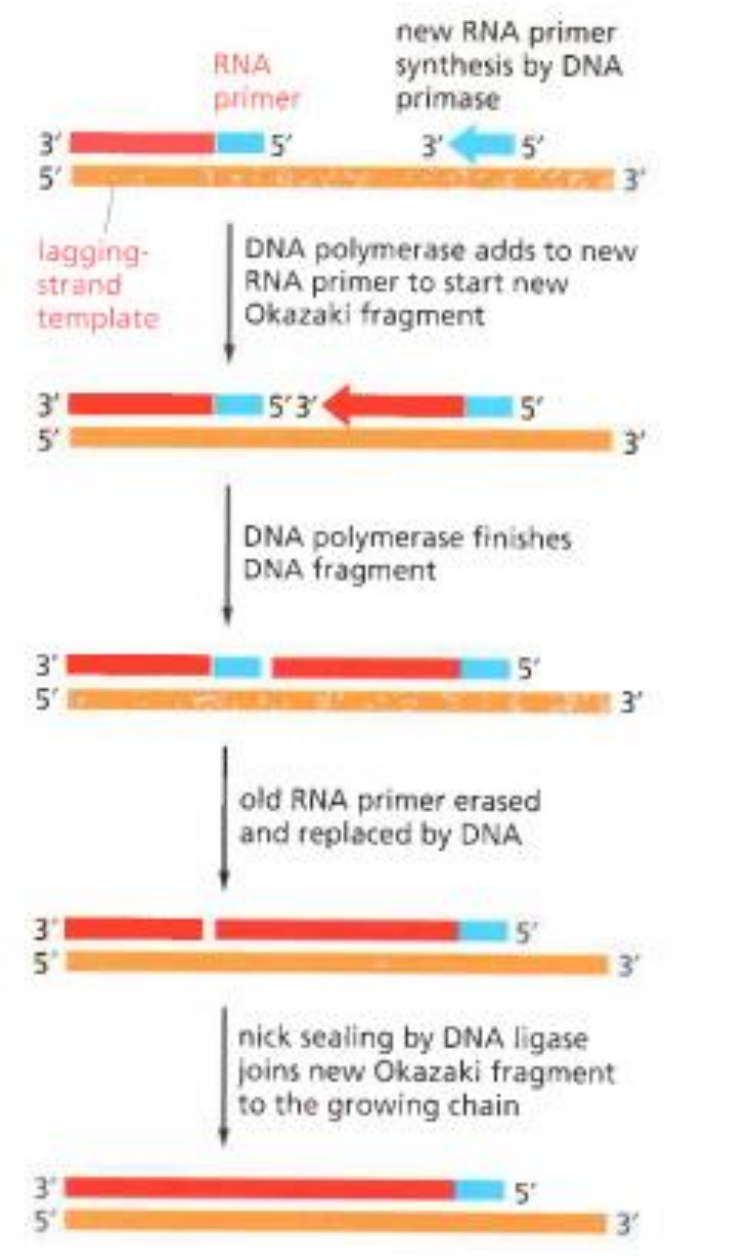
DNA synthesis on the leading and lagging strand



(a) At intervals, primase synthesizes an RNA primer for a new Okazaki fragment.

(b) Each primer is extended by DNA polymerase lll.

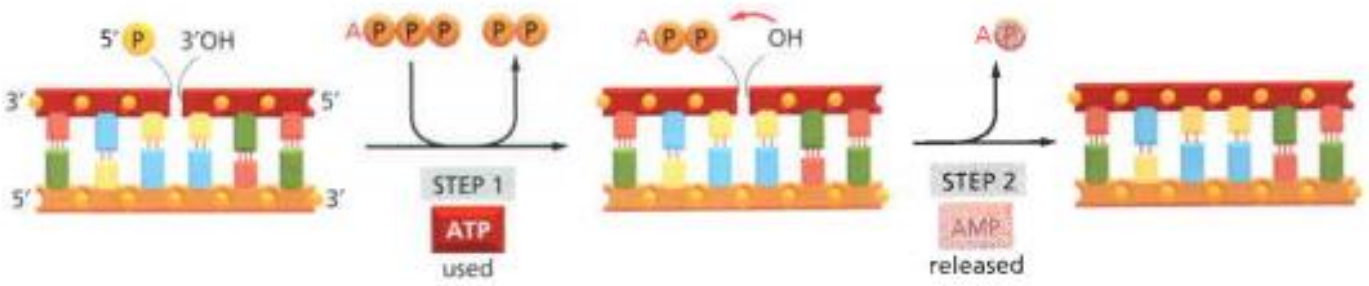
(c) DNA synthesis continues until the fragment extends as far as the primer of the previously added Okazaki fragment A new primer is synthesized near the replication fork to begin the process again.

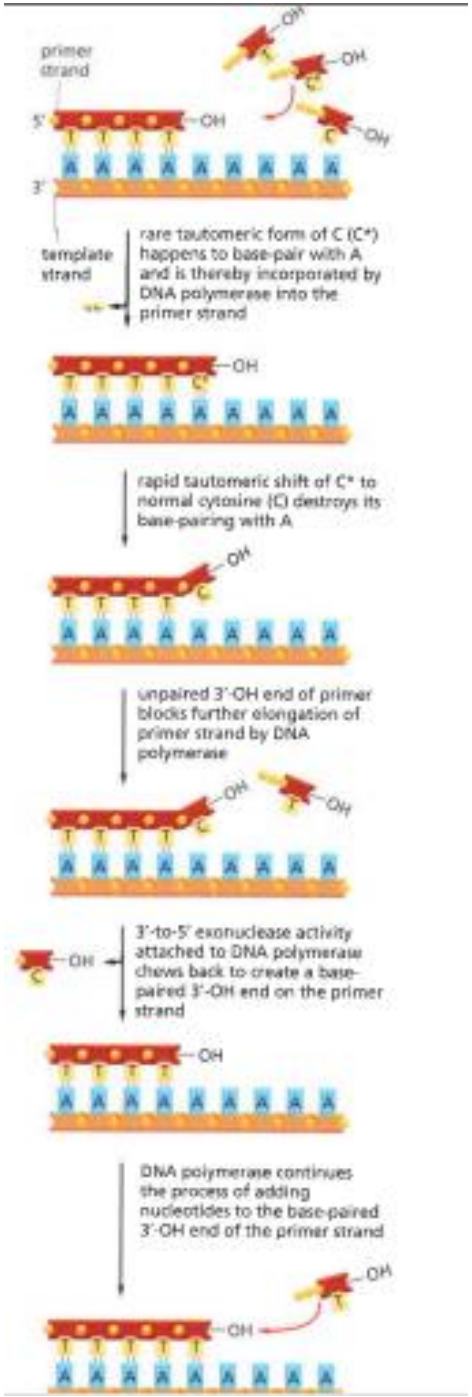


The synthesis of DNA fragment on lagging strand.

In eukaryotes RNA primers are made at intervals spaced about 200 nucleotides on the lagging Strand.
Each RNA primer is ~ 10 nucleotides long.
This primer is removed by a special DNA repair enzyme , **RNase H** that Recognises an RNA strand in RNA/DNA Hybrid and cleaves it.
The gaps are filled in by **DNA Polymerase** and **DNA ligase**

Reaction catalyzed by DNA ligase





DNA polymerase has a 3’-5’ Proofreading exonuclease Activity

Termination

- Termination occurs at defined DNA sequences (called terminator sequences) that act as binding sites for a protein called Tus (terminus utilization substance).
- The Tus-Ter complex can arrest a replication fork from only one direction.