I. DNA Replication
   A. Requirement for high-fidelity DNA replication
   B. Enzymology and mechanism of DNA synthesis
      1. General description of replication process
      2. Properties of DNA polymerases
      3. Therapeutic nucleotides
      4. Okazaki fragment synthesis
         Priming and processing
      5. Other replication factors
         Sliding clamps, DNA helicases, etc.
   C. Initiation of DNA replication
   D. Telomeres

II. DNA Repair
   A. Types of DNA damage
      Spontaneous and environmental
   B. Base excision repair
      DNA glycosylases, AP endonuclease
   C. Nucleotide excision repair
      \textit{E. coli} UvrABC endonuclease, helicase II (uvrD)
      Human xeroderma pigmentosum (XP) complementation groups
   D. Mismatch repair
      \textit{E. coli} mutS, mutH, mutL and determination of incorrect strand
      human hMSH2, hMLH1 and hereditary nonpolyposis colon cancer
   E. Defects in cell cycle checkpoints
      Ataxia Telangiectasia
      p53 mutation
Overview of DNA replication

Specialized elements termed 'origins of DNA replication' occur many times on a chromosome.

Initiation of DNA replication from origin of replication generates structures termed 'DNA replication bubbles'. (One of many replication bubbles is shown)

Each replication bubble contains two 'DNA replication forks' moving in opposite directions. (An enlarged view of one fork is shown here)

Action of DNA replication complex (acting on fork) duplicates both parental strands.
Replication employs two different mechanisms to replicate the leading and lagging strands. DNA synthesis on the leading strand occurs in the same direction as fork movement, is continuous and generally highly processive. Lagging strand DNA synthesis occurs in the direction opposite to fork movement, is discontinuous and uses DNA polymerases with moderate processivity.
Nascent strand synthesis by DNA polymerases occurs only in the 5’ to 3’ direction. Nucleotide incorporation utilizes two high-energy phosphate bonds, driving the reaction in the forward direction.
Anomalous base pairing can occur that involves rare tautomeric forms of the bases. In the example shown, when cytosine is in the rare imino form (C*), it can form a base pair with the common amino form of adenine (A). The rare form of adenine (A*) can also pair with the common form of cytosine (C). G*:T and G:T* base pairs can also form. Defective base insertion can be corrected either by DNA polymerase proofreading or mismatch repair.

An important process required for high fidelity DNA replication is proofreading. DNA molecules with a mismatched nucleotide at the 3' end of the primer strand (i.e., because of DNA polymerase misincorporation) are poor templates for continued DNA synthesis and usually cause stalling of the DNA polymerase. A 3'-to-5' proofreading exonuclease associated with the polymerase enzyme removes unpaired nucleotides from the 3' end of the nascent strand. The DNA polymerase can therefore act as a self-correcting enzyme to remove its own polymerization mistakes.
The activities of key replication enzymes are conserved in bacteria (E. coli) and eukaryotes (human). For example, the overall structure of the E. coli β protein and the human PCNA are similar, even though the gene sequences are divergent.

<table>
<thead>
<tr>
<th>Functions</th>
<th>E. coli</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA polymerase; 3' to 5' exonuclease</td>
<td>Pol III core</td>
<td>Pol δ</td>
</tr>
<tr>
<td></td>
<td>(αδε)</td>
<td></td>
</tr>
<tr>
<td>Primase</td>
<td>DnaG</td>
<td>Primase</td>
</tr>
<tr>
<td></td>
<td>Ssb</td>
<td>Pol α*</td>
</tr>
<tr>
<td>Sliding clamp</td>
<td>β</td>
<td>PCNA</td>
</tr>
<tr>
<td>Clamp loader</td>
<td>γ complex</td>
<td>RFC</td>
</tr>
<tr>
<td>DNA helicase</td>
<td>DnaB</td>
<td>MCM family?</td>
</tr>
<tr>
<td>Single-stranded DNA-binding protein</td>
<td>SSB</td>
<td>RPA</td>
</tr>
<tr>
<td>Ligation of DNA</td>
<td>Ligase</td>
<td>Ligase</td>
</tr>
<tr>
<td>Nuclease for removal of RNA primers</td>
<td>RNase H</td>
<td>RNase H</td>
</tr>
</tbody>
</table>

*The human DNA pol α and primase activities function in a multiprotein complex to synthesize RNA-DNA primers.*
<table>
<thead>
<tr>
<th>Polymerase name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA polymerase α (pol α)</td>
<td>Primer function in DNA replication</td>
</tr>
<tr>
<td>pol β</td>
<td>Base excision repair</td>
</tr>
<tr>
<td>pol γ</td>
<td>Mitochondrial DNA replication and repair</td>
</tr>
<tr>
<td>pol δ</td>
<td>DNA replication; nucleotide and base excision repair</td>
</tr>
<tr>
<td>pol ε</td>
<td>DNA replication; nucleotide and base excision repair</td>
</tr>
<tr>
<td>pol ζ</td>
<td>Translesion synthesis?</td>
</tr>
<tr>
<td>pol η</td>
<td>Accurate translesion synthesis</td>
</tr>
<tr>
<td>pol θ</td>
<td>DNA repair of crosslinks?</td>
</tr>
<tr>
<td>pol ι</td>
<td>Somatic hypermutation</td>
</tr>
<tr>
<td>pol κ</td>
<td>Error-prone repair</td>
</tr>
<tr>
<td>pol λ</td>
<td>Meiosis-associated DNA repair?</td>
</tr>
<tr>
<td>pol μ</td>
<td>Somatic hypermutation?</td>
</tr>
</tbody>
</table>

Human DNA-directed DNA polymerases. Human cells, similar to prokaryotic and other eukaryotic cells, contain multiple DNA polymerases. The activities of these DNA polymerases differ to a certain extent because of the specific roles they play in DNA replication, repair, and recombination. The most important human DNA polymerases for general DNA replication are DNA polymerase α (which has an associated DNA primase activity) and DNA polymerase δ. DNA polymerase δ is the major DNA polymerase required for processive leading strand synthesis, but also functions to complete Okazaki strand synthesis on the lagging strand template. DNA polymerases α to ε have been known for a number of years and are fairly well characterized. Most of the other DNA polymerases have been identified only very recently, and the characterization of their functional role is ongoing. Note that the list is not comprehensive – additional DNA polymerases have been identified which are not included in this list.


Therapeutic nucleoside analogs

1. Anticancer

<table>
<thead>
<tr>
<th>Cytosine</th>
<th>Converted to the nucleoside triphosphate. Inhibits elongation of DNA synthesis by competing with dCTP for DNA polymerase. Can act as a chain terminator - selectively kills growing cells. Antileukemic.</th>
</tr>
</thead>
</table>

![](cytosine.png)
cytosine arabinoside (araC)

2. Antiviral

A. Anti-HIV

<table>
<thead>
<tr>
<th>Thymine</th>
<th>Converted to the nucleoside triphosphate. Selectively inhibits HIV reverse transcriptase activity (an essential HIV DNA polymerase that synthesizes the HIV DNA strands from the RNA template). Acts as a chain terminator.</th>
</tr>
</thead>
</table>

![](thymine.png)
3'-deoxy-3'-azidothymidine (AZT)

<table>
<thead>
<tr>
<th>Inosine</th>
<th>Inosine is a precursor to purines guanine and adenine. Converted to the nucleoside triphosphate. Selectively inhibits HIV reverse transcriptase activity. Acts as a chain terminator.</th>
</tr>
</thead>
</table>

![](inosine.png)
dideoxyinosine (ddI)

B. Other Anti-Viral

| --- | --- |

![](guanine.png)
Acyclovir
Okazaki fragment (OF) synthesis in human cells

- Normally occurs on lagging strand
- Two DNA polymerases are used

**Step 1. Priming**
DNA polymerases can not start nascent strand synthesis de novo. Need RNA priming by DNA primase associated with DNA pol α. Primer is ~10 nt in length.

**Step 2. Synthesis of initiator DNA (iDNA).**
DNA pol α extends the RNA primer. The iDNA is ~35 nt in length, and acts as a DNA primer.

**Step 3. Extension of iDNA.**
DNA pol δ (which has no priming activity) extends the iDNA. Extension continues ~200 nt to 5' end of previous OF.

**Step 4. Removal of RNA primer.**
The RNA primer is removed by RNase H (which cleaves the RNA portion of a DNA-RNA hybrid). Most of the iDNA is also removed by a nuclease. Occurs while next OF is being synthesized.

**Step 5. Ligation.**
In the final reaction, the OFs are joined.
Eukaryotic and prokaryotic replication utilize a generally similar mechanism of lagging strand synthesis. In the case of the eukaryotic replication, the DNA polymerase alpha/DNA primase (pol α/primase) first makes a short RNA/DNA hybrid, which DNA polymerase delta (pol δ) then extends. Processing enzymes including RNase H and DNA ligase join the Okazaki fragments. For prokaryotic DNA synthesis, primase synthesizes the RNA primer, which then is used by the pol III holoenzyme complex. A second polymerase (pol I), in combination with RNase H, processes the 5’ end of the previous Okazaki fragment. After the pol III holoenzyme fills the gap, DNA ligase joins the two strands. In each case, therefore, two distinct DNA polymerases are used for lagging strand synthesis.
Three-dimensional view of a replication fork. Current work indicates that, both for prokaryotic and eukaryotic DNA replication, the primary leading strand and lagging strand DNA polymerase are coupled. For human DNA replication (shown), two molecules of DNA polymerase delta (pol δ) are joined. Upon completion of the Okazaki fragment, the lagging strand pol δ would transfer to the most recently formed RNA/DNA primer (synthesized by the DNA polymerase α/DNA primase complex; pol α/primase). For the \textit{E. coli} system, two molecules of the pol III holoenzyme would be joined. Because the lagging strand loop would vary in size during the synthesis of each Okazaki fragment, this mode of DNA synthesis has been termed the ‘trombone model’ of DNA synthesis.

The eukaryotic cell cycle generally has four phases termed M (mitosis and cytokinesis), G1 (first gap phase), S (DNA synthesis), and G2 (the second gap phase). Chromosomal DNA replication occurs during S phase. During S phase, all genes are replicated only a single time (the ‘once and only once’ mechanism of eukaryotic DNA replication). Because of this type of regulation, no genes are lost (because they were not replicated) or re-duplicated (because they underwent a second round of replication during a single S phase).
Mechanisms of topoisomerase I and topoisomerase II activity. In each case, the topoisomerase makes reversible covalent bonds with the DNA. Each topoisomerase activity has somewhat different roles. Topoisomerase I serves to relieve the torsional strain ahead of the replication fork generated as a result of DNA unwinding. Topoisomerase II generally separates the two daughter molecules at the final stages of DNA replication (and just prior to mitosis). Because of their critical roles during DNA replication, various therapeutic agents are directed against both these enzymes.
Mechanism for replication initiation at oriC, the replication origin for *E. coli*. The dnaA protein first binds specific sequences in oriC (the 9-mers). In the presence of ATP, dnaA causes a localized melting of DNA elements that have a relatively unstable duplex (the 13-mers) to form the open complex. The replicative DNA helicase (dnaB) can then join the complex and extends the melted bubble. The remainder of the DNA polymerase machinery can then bind to the exposed single-strand and catalyze nascent strand synthesis.

Mechanism for replication initiation in eukaryotic cells. The reaction mechanism is significantly more complex than used for prokaryotic initiation. The origin recognition complex (ORC) first recruits cdc6 and the MCM complex to the origin during the G1 phase. At the G1/S phase transition, cell cycle kinases (e.g., cdk2/cyclin E and cdc7/dbf4) become activated and phosphorylate the origin-bound proteins. Other replication factors are subsequently recruited leading to melting of the origin DNA and loading of the DNA polymerase machinery. Many of the details of the process are still not understood.
Replication fork initiation. The replication initiation complex binds origin sequences and induces the local opening of the DNA duplex. In the case of eukaryotic DNA replication, the DNA pol α/primase complex then binds and synthesizes an RNA/DNA primer to initiate leading strand synthesis. The formation of this replication fork generates longer ssDNA regions that serve to bind additional replication components. In this way, two replication forks are formed, moving in opposite directions, that synthesize both nascent leading and lagging strands.

Telomeres and telomerase. The ends of linear chromosomes in eukaryotes are ‘capped’ by repeating G-rich sequences, which comprise telomeres. For certain ciliates (shown), the repeats have a GGGTTG sequence, but in humans they are GGGTTA. The shorter bottom strand was initially synthesized as a nascent lagging strand, but is shortened upon processing (i.e., by removal of the RNA/iDNA primer). Without telomerase, therefore, each chromosome loses a few telomere repeats and becomes shorter every cell cycle. Telomerase contains an RNA template which allows the synthesis of more telomere repeats at the 3’ end of telomeric DNA. Thus, shortening (which could eventually lead to loss of genetic information) is prevented.
Telomere shortening in the absence of telomerase. Normal human somatic cells express very low levels of telomerase. As a person ages, the telomeres shorten because post-replication processing (particularly the removal of the RNA primer from the lagging strand product) results in the loss of DNA. Over many generations, the telomere length becomes sufficiently short (the Hayflick limit) that it causes cells to become senescent, and cells will no longer divide. This may be an evolutionary device to inhibit cell immortalization and tumorigenesis. If cells lack p53 (p53 is often mutated in cancer cells), they do not undergo senescence but instead continue to divide. In these cells, telomere length becomes so short that it causes cells to undergo apoptosis (crisis phase). Occasionally, a rare cell survives which expresses telomerase activity and can therefore become immortal. In contrast to somatic cells, germline cells and almost all tumor cells express telomerase and maintain telomere length. Because telomerase activation appears to be a critical step in cell immortalization, anti-telomerase agents are being developed that could potentially lead to novel treatments against cancer cells.
Overview of DNA Repair

Base damage. A very common type of DNA damage is spontaneous deamination. In the above example, the cytidine to uridine conversion occurs at a rate of ~100 bases per genome per day. Other common spontaneous types of DNA damage are depurination and depyrimidation (creating AP sites), and oxidative damage to bases. Base damage also occurs from environmental sources such as ionizing or ultraviolet radiation, and chemical agents. If the damage occurs on a single base and does not result in a bulky adduct, it is repaired by base excision repair (BER). In the first step of BER, a DNA glycosylase (many different DNA glycosylases exist) hydrolyzes the N-glycosidic bond between the damaged base and the sugar-phosphate backbone, giving an AP site and releasing the free base.
Schematic representation of base excision repair. For simplicity, only the relevant DNA strand is shown for most of the figure. Certain forms of base damage are recognized by DNA glycosylases (step 1), that catalyze excision of the free base by hydrolysis of the N-glyosyl bond linking the base to the sugar-phosphate backbone. This reaction leaves an apurinic/apyrimidinic (AP) site in the DNA. Attack at such sites by an AP endonuclease (step 2) results in a strand break with a 5’ phosphate and a 3’ hydroxyl. The abasic sugar-phosphate residue is excised by a phosphodiesterase. The resulting single nucleotide gap is filled by repair synthesis, and DNA repair is completed by DNA ligase.

The exposure of cells to UV light (such as contained in sunlight) can result in the formation of thymine dimers in chromosomal DNA. The formation of these dimers is excised in the process of nucleotide excision repair, involving the UvrABC endonuclease in *E. coli* (below), and XP proteins in humans.
Nucleotide excision repair in bacteria. This repair process excises thymine dimers (shown) from DNA although other bulky adducts are also substrates for this reaction. In the first step (A), the thymine dimer (T = T) is recognized by the UvrA and UvrB proteins. UvrC replaces UvrA (step B) and the UvrB/UvrC complex nicks the DNA strand on either side of the damage (step C). In the next step D, Helicase II (also called uvrD) removes the excised strand (12 to 13 nts in length). The gap is then filled by DNA polymerase I (pol I) and the strands joined by the action of DNA ligase. The recognition/excision proteins are usually termed the UvrABC endonuclease.

Model for the strand-specific transcription-coupled DNA repair in *E. coli*. (A) RNA polymerase is shown transcribing a template strand that contains base damage ahead of the transcription complex. (B to D) Stalling of RNA polymerase at the site of base damage in the transcribed strand (B) results in the binding of the transcription-repair coupling factor (TRCF; C) and displacement of the polymerase and the truncated transcript, leaving TRCF bound at the site of damage (D). TRCF then binds the UvrABC endonuclease (E) and nucleotide excision repair takes place (F).
Nucleotide Excision Repair | Role in Repair
---|---
XPA | Damage recognition
RPA | Damage recognition
TFIIH (XPB, XPD) | Formation of preincision complex
XPC | Stabilization of preincision complex
XPF | 5’ excision
XPG | 3’ excision

Proteins required in human nucleotide excision repair (NER). Most repair factors were named after xeroderma pigmentosum (XP), an inherited disease in which patients show an extreme sensitivity to sunlight. The process of NER in humans is generally similar to NER in bacteria although the excised fragment is larger (the predominant species 27 to 29 nts is in length).

Microsatellite Instability. Normal human cells show a low level of polymorphism of repeat or ‘microsatellite’ sequences such as (CA)_n. Occasional slippage is corrected by mismatch repair. Cells from hereditary nonpolyposis colon cancer (HNPCC) patients show a greatly increased level of sequence instability, often resulting from mutation of mismatch repair genes (usually hMSH2 and hMLH1). In the figure, a specific DNA region containing a polymorphic site was amplified by PCR and the products were separated by gel electrophoresis. The PCR products of colon cancer tissue DNA (T) from seven different HNPCC patients or from adjacent normal tissue DNA (N) are shown. For most of the patients, the tumor tissue shows an increased level of polymorphism at these (and other) sites, compared to normal tissue.
Mismatch repair in bacteria. Three E. coli proteins (mutS, mutL, and mutH) are primarily responsible for mismatch recognition. The mutH protein nicks the DNA strand containing the incorrect base at a distal site, and that strand is then degraded back to the mismatch. Damage to the genes coding for their human counterparts appears to trigger hereditary colon cancer and other forms of cancer.