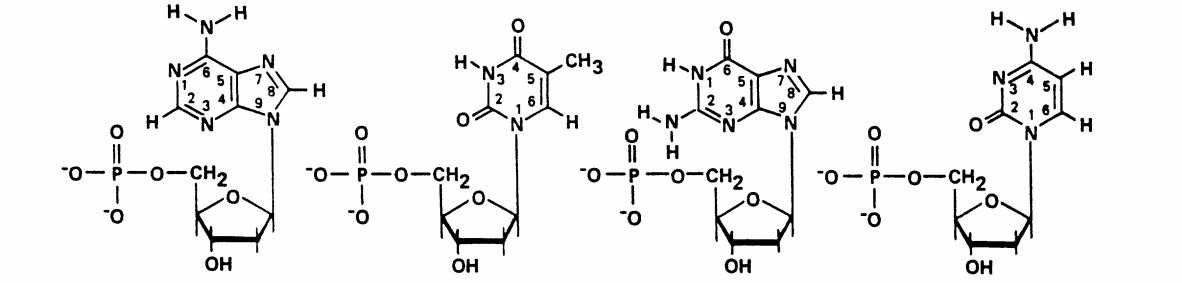
DNA AND CHROMOSOME STRUCTURE

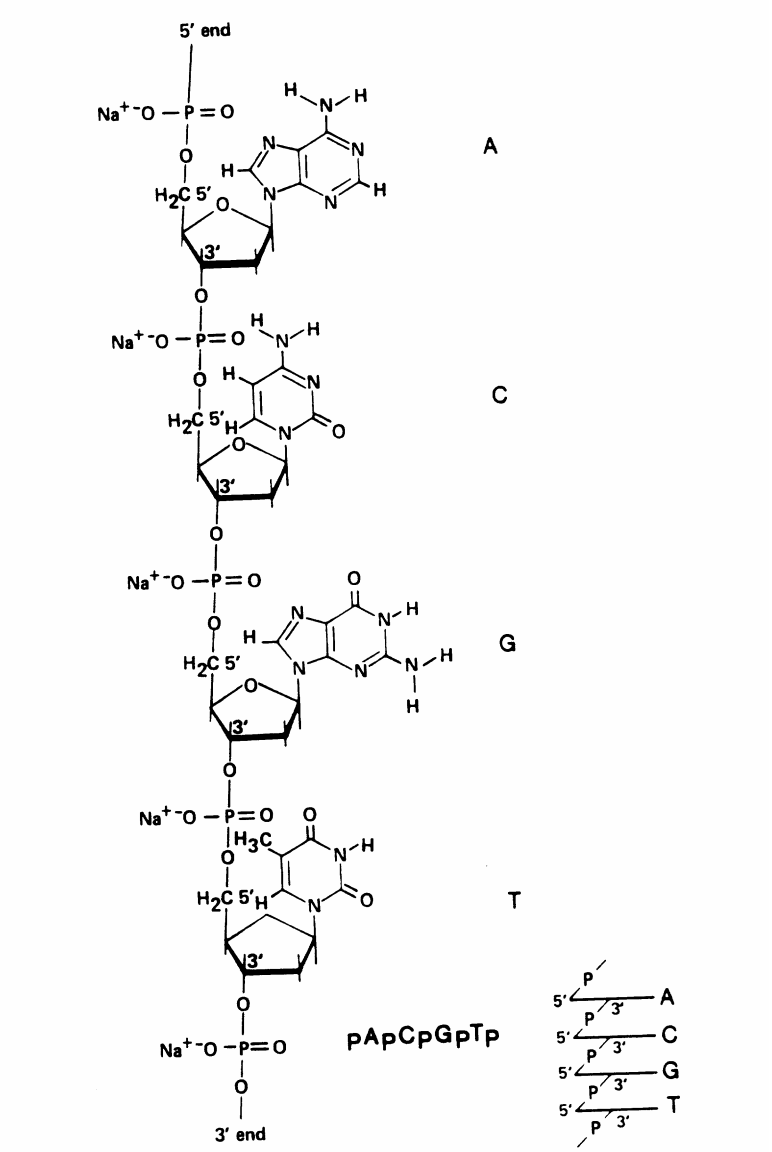
The seminal work of Avery and colleagues in 1944 established DNA as the genetic material. Specifically, they demonstrated that an extract from one type of bacteria, that normally formed a capsular structure and was pathogenic in animals, could transfer this pathogenecity and capsular structure to a bacterium that normally did not exhibit these properties. They went on to show that the active material in this extract, that was capable of "transforming" the non-virulent strain to the virulent state, was DNA. Further work in other organisms, and in particular the bacteriophage T2, verified the fact that DNA was the genetic material which then culminated in the landmark work of Watson and Crick in 1953 that established the double helical structure of the DNA molecule and provided the basis for understanding the replication and flow of genetic information. Obviously, an understanding of genetic mechanisms, including the molecular basis for human disease, is critically dependent on a detailed understanding of the structure and properties of the DNA molecule.

DNA Primary Structure

DNA is comprised of the four deoxyribonucleoside monophosphates (also called deoxyribonucleotides or simply dNMPs):



A DNA strand is an unbranched polymer of dNMPs linked together in a linear fashion by phosphodiester bonds:



Polarity is conferred on the DNA chain by virtue of the asymmetry about the plane of the deoxyribose ring. This polarity is described using the terms 3' and 5' reflecting the disposition of the 5' carbon and the 3' OH on opposite sides of the sugar ring.

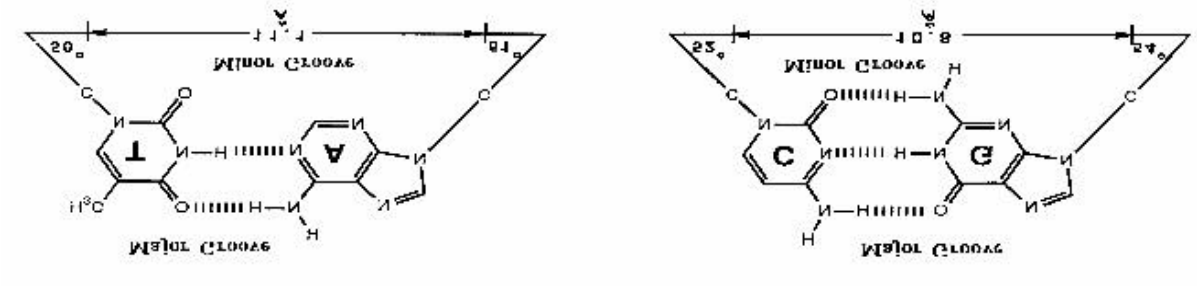
Note the shorthand abbreviations used to designate a DNA chain.

DNA Secondary Structure

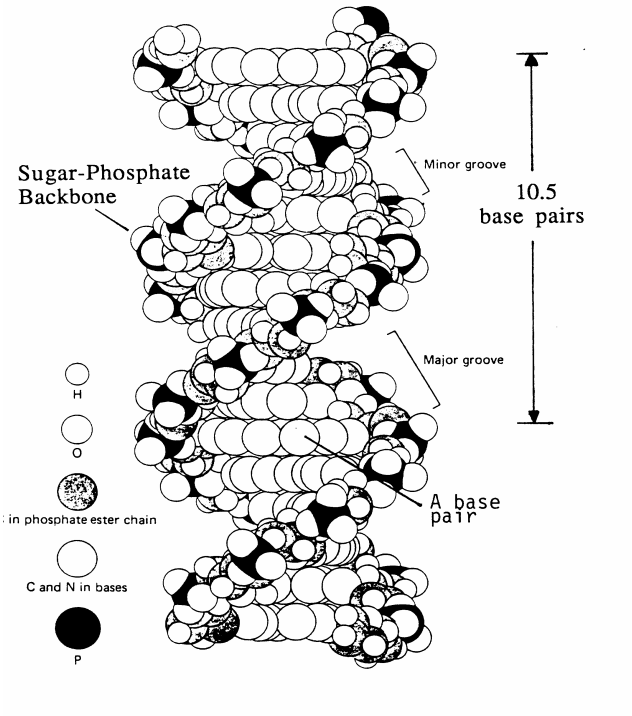
A. The Watson-Crick base pairs T-A and C-G.

i) The planes of the purine and pyrimidine rings in a base pair are tilted slightly relative to each other (a "propeller" twist) in B-form DNA.

ii) T-A and C-G base pairs are very similar in size and shape. This allows them to be accommodated within the helix with only small effects on local helix structure.



B. The DNA double helix (B-form DNA) -- The B-conformation, a right-handed double helix, is the average conformation adopted by the majority of biologically active DNA sequences.

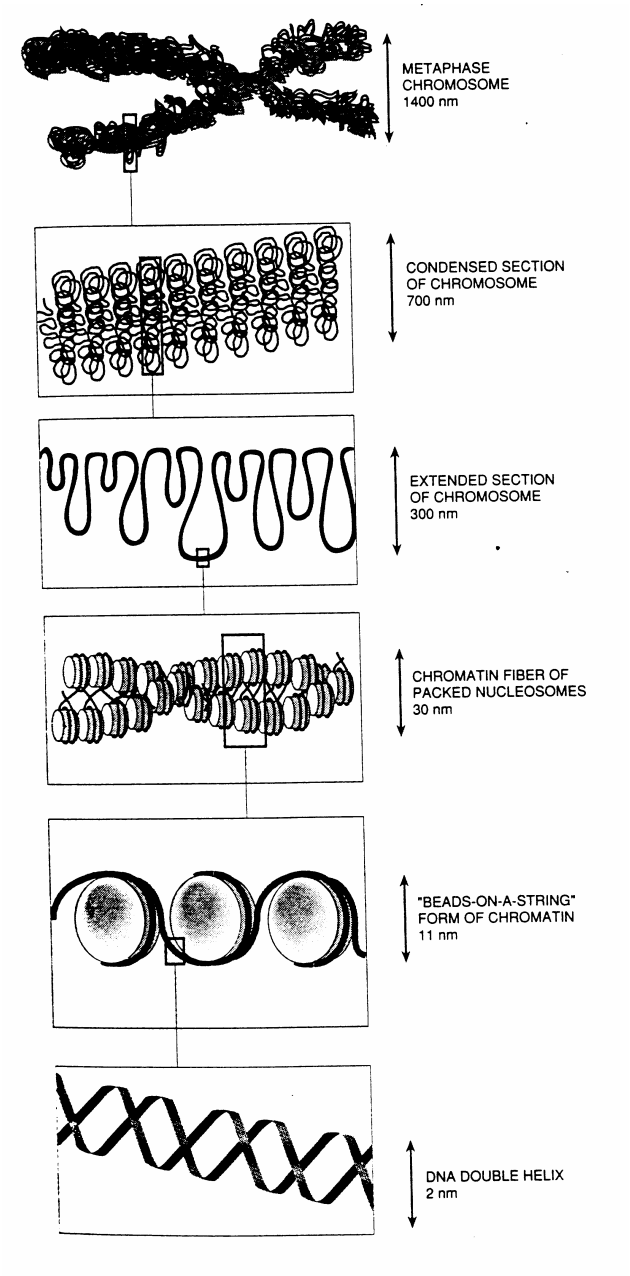


The properties of the DNA double helix can be summarized as:

- ? The negatively charged sugar-phosphate backbone is located on the outside of the helix.
- ? The two strands of the helix are oriented in an anti-parallel fashion and are held together by non-covalent bonds.
- ? Base pairs are stacked on top of each other inside the helix. For B-DNA in solution, one turn of the helix corresponds to 10.5 base pairs (about 34Å).
- ? Base pair functional groups are exposed in the major and minor grooves of the helix (see 2A above). Genetic regulatory proteins, which act by recognizing certain DNA sites, typically recognize a particular sequence by "reading" this groove information.

Sizes of DNA Molecules and Chromosome Structure

The smallest biologically active DNA's (chromosomes of small viruses) are about 5000 base pairs (bp) in size corresponding to a helix contour length of about 0.0017 mm. The haploid content of the human genome is 3×10^9 bp (990 mm). Thus, each human somatic (diploid) cell contains about 2 meters of DNA distributed among 22 pairs of autosomes and one pair of sex chromosomes (XX - female, XY - male). The average size of a human chromosome is therefore 130 million bp (4.3 cm). Moreover, there is good evidence that the DNA within each chromosome is present as a single, continuous double helix. Although the manner in which the cell deals with such large molecules is not fully understood, the outline below illustrates the current view of DNA condensation within the nucleus.



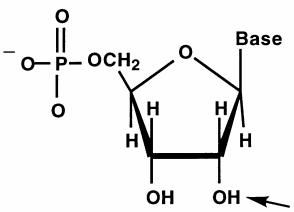
The manner in which the 30 nm chromatin filament is condensed to yield the 300 nm and 70 nm fibers that have been observed by microscopy is not understood

Histone H1 mediates condensation of nucleosome "beads", resulting in a 30 nm fiber comprised of a helical arrangement of nucleosome particles stabilized by H1.

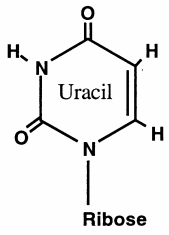
The basic unit of DNA packing in eukaryotic chromosomes is the nucleosome "bead". This structure is comprised of a protein core - an octamer of four basic histone proteins: $(H2A)_2(H2B)_2(H3)_2(H4)_2$ with about 200 bp of DNA wrapped around its surface.

RNA Primary Structure

Cells contain a second kind of nucleic acid called RNA, which has a primary structure similar to that of DNA. An RNA chain, like a DNA strand, is comprised of mononucleotide units linked by phosphodiester bonds. However, in RNA the nucleotide units are derivatives of the sugar ribose, with the four major constituents of RNA being AMP, GMP, CMP, and UMP:

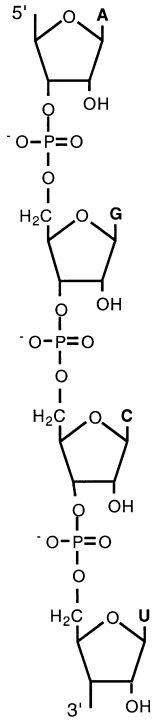


Bases: The bases adenine, guanine, and cytosine found in DNA are also present in RNA. However, RNA has the base uracil (U) instead of thymine (thymine is 5-methyluracil).



A ribonucleoside-5'-monophosphate
(rNMP, ribonucleotide, mononucleotide)

A portion of an RNA strand:



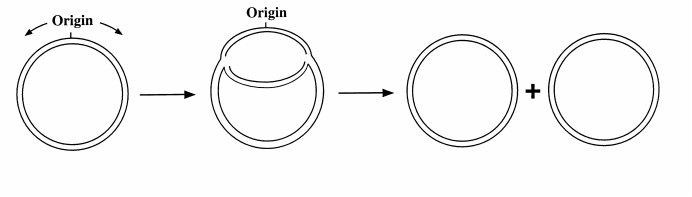
DNA Replication

The double helical duplex structure of the DNA molecule, as determined in the work of Watson and Crick, immediately predicted a simple mechanism for replication of the genetic material whereby one DNA strand would serve as a template for the linear synthesis of the complementary strand. Although the replication of DNA indeed does involve a copying mechanism, it is anything but a simple process.

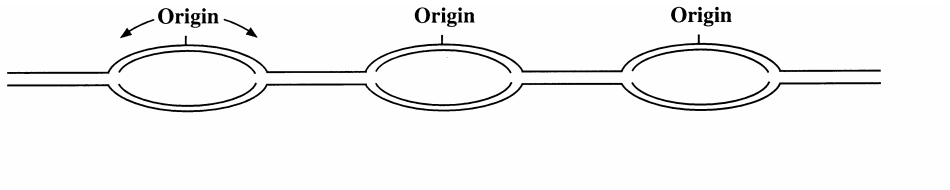
Replication Units

The chromosomes of viruses, bacteria, and eukaryotic cells replicate
in discrete units. Replication of each unit is initiated at a defined origin, with replication usually proceeding in a bidirectional fashion from this site. Origin initiation appears to be a major site of replication control.

A. **Bacteria and many viruses** -- the unit of replication is the chromosome:



B. **Man and other eukaryotes** -- multiple replication units per chromosome:



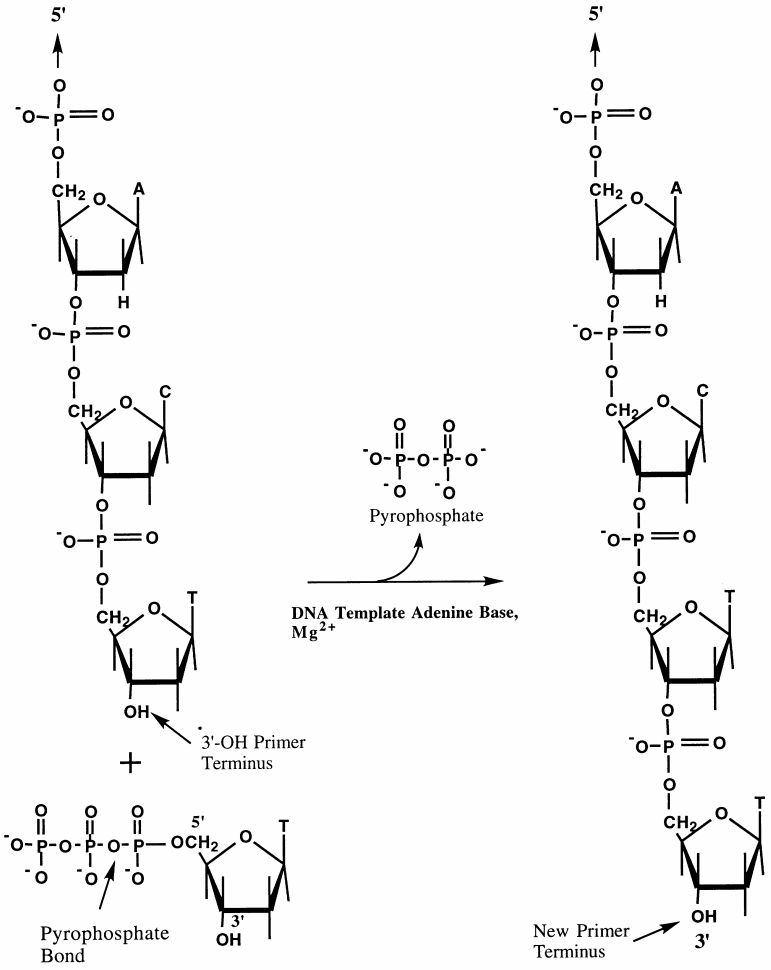
Accuracy of Replication

The mutation rate in human cells is about 10^{-10} per bp per cell division. Since the mutation rate includes genetic alterations from all sources, this value represents an upper limit for the error rate of chromosome replication. Replication precision reflects the intrinsic accuracy of the DNA

polymerases responsible for DNA biosynthesis, as well as correction of DNA biosynthetic errors by the mismatch repair system which will be described below. It is also pertinent to note that given the large size of the human genome (3 x 10^9 bp haploid), a mutation rate of 10^{-10} per bp per division means that human cells accrue about one mutation for each cell division.

The DNA Polymerase Reaction

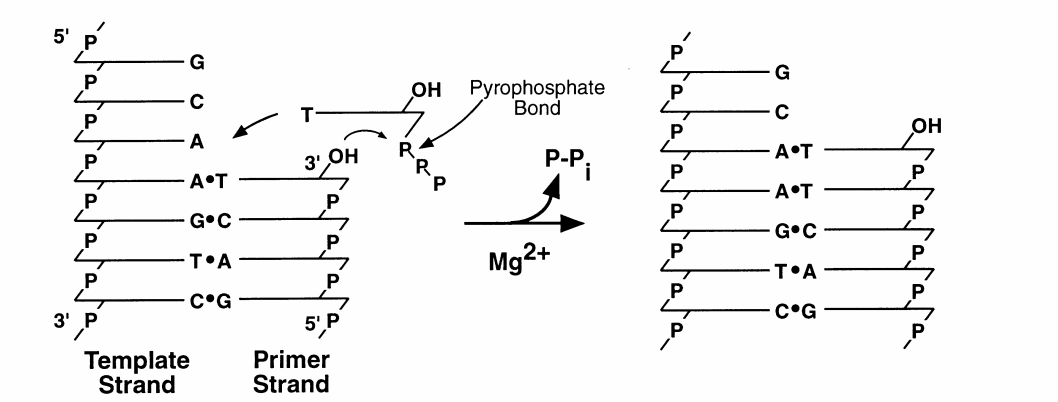
DNA polymerases from all organisms catalyze the same reaction: the template-directed addition of deoxyribonucleotides to the 3' end of a growing chain. DNA polymerases cannot start new chains: their action is restricted to extension of pre-existing strands. The chain that undergoes extension is called the primer strand to emphasize this fact.



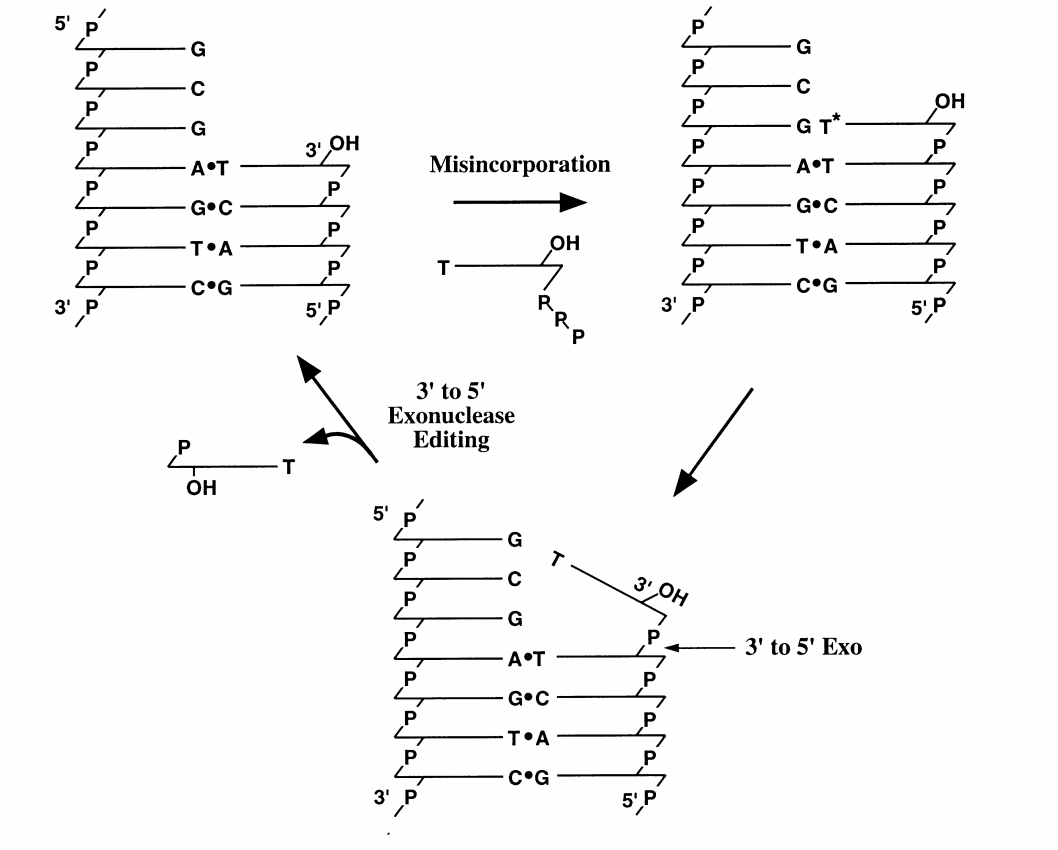
The precursors for DNA synthesis are the four deoxyribonucleoside-5'-triphosphates (dATP, dGTP, dCTP, dTTP). The chemistry of polymerase-catalyzed phosphodiester bond formation and the

template-directed nature of this reaction are illustrated below. As can be seen, the incoming nucleotide is added to the 3'-terminus of the growing DNA chain, with formation of the phosphodiester linkage driven by hydrolysis of a pyrophosphate bond. Elongation of the 3'-end of the primer strand thus corresponds to 5' to 3' chain growth.

The template-directed nature of this reaction can be shown schematically as:



4. **The Editing Exonuclease**

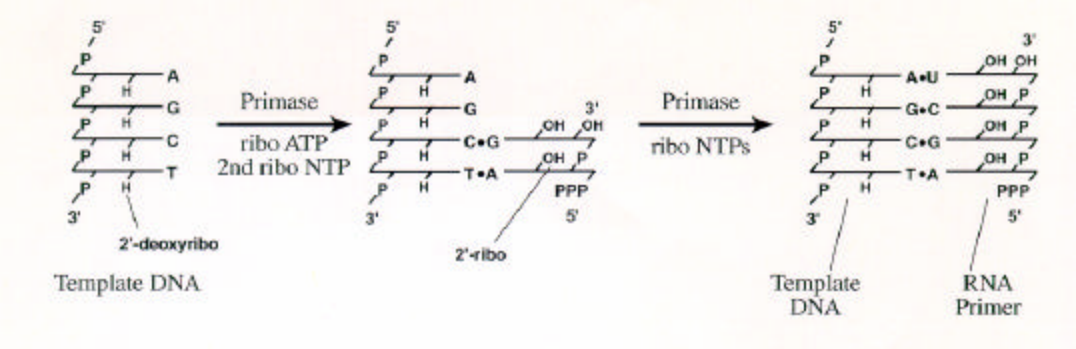


Many, but not all, DNA polymerases have an associated 3' to 5' exonuclease activity that hydrolyzes single strand DNA. When given a DNA single strand, these activities initiate hydrolysis at the 3'-end and sequentially remove 5'-dNMPs. Enzymes possessing this activity include the human d and e DNA polymerases. Proofreading by the editing exonuclease removes 95-99% of the errors that result from misincorporation of incorrect nucleotides by the polymerase function. Polymerase selectivity coupled with the action of the editing exonuclease confers an overall error rate of about 10^{-6} to 10^{-7} per nucleotide copied. This high degree of accuracy is not sufficient for biological systems (it would correspond to about 600 to 6000 mutations per human cell per division). Residual errors are removed by mismatch repair systems (discussed below), which scan newly synthesized DNA for errors, selectively eliminating incorrect nucleotides from the daughter DNA strand.

Other Proteins Required for DNA Replication

The participation of DNA polymerases in chromosome replication is proven. This may seem puzzling in view of the restricted directionality of the enzymes (5' to 3' chain growth only) and their inability to initiate new DNA chains. These issues were resolved with the identification of additional enzymes and proteins required for replication, and the discovery that one strand of a replicating helix is copied in a discontinuous fashion. The figure showing the model for a human replication fork presents the current view of the replication. Note that the **leading strand** grows continuously while the **lagging strand** is synthesized discontinuously in the form of short DNA chains. These short chains are initiated *de novo* via synthesis of an RNA primer. After removal of the RNA primer by RNAse H and repair of the ensuing gap, the short chains are joined together by DNA ligase. The functions of the proteins shown are:

Primase: Synthesizes very short RNA chains (about 10 ribonucleotides long) that serve as primers for initiation of synthesis by DNA polymerase a. Primase forms a tight complex with DNA polymerase a and catalyzes the reaction shown below:⁇



In contrast to DNA polymerases, RNA polymerases can initiate a new chain, with the new chain containing a 5'-triphosphate terminus from the initiating ATP. RNA primers synthesized by primase are usually about 10 residues long.

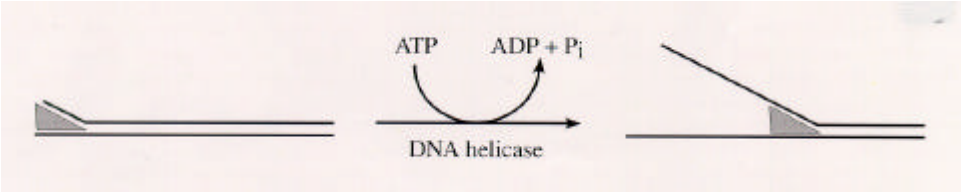
DNA Polymerase a: Extends RNA primers synthesized by primase to initiate DNA synthesis on the lagging strand.

DNA Polymerase d and ?: DNA polymerase ? extends RNA-primed DNA chains synthesized by DNA polymease ?. DNA polymerase a adds about 30 dNMPs to an RNA primer (pol a-synthesized material is called iDNA for "initiating DNA") before d takes over to extend the chain to about 500 nucleotides. DNA polymerase ? is probably responsible for leading strand synthesis.

RFC (Replication Factor C) Involved in loading DNA polymerase?d (and possibly ? as well) onto a primer terminus. Also functions to displace the DNA polymerase a/primase complex from iDNA so d can take over lagging strand synthesis.

PCNA (Proliferative Cell Nuclear Antigen): Stabilizes binding of polymerase d and ? to a primer-template.

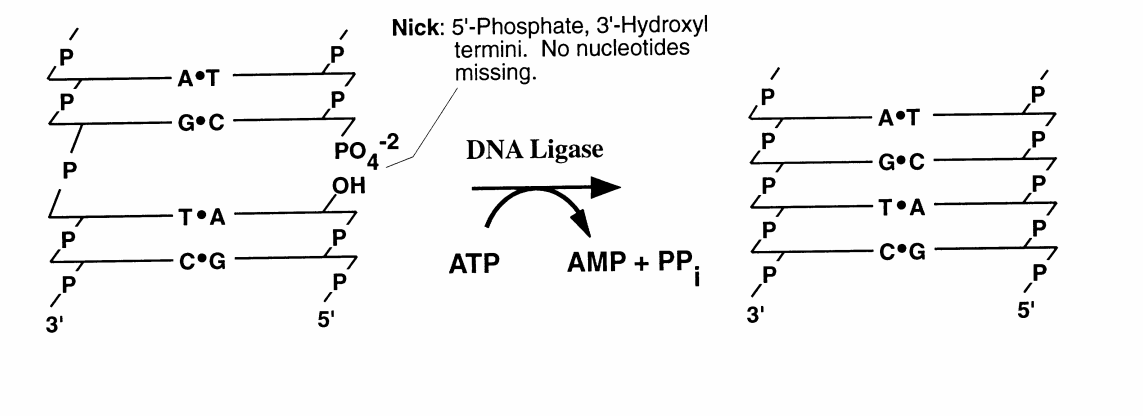
DNA helicase: Unwinds the strands of the parental helix in an ATP-driven reaction.



SSB (Single-stranded DNA binding protein; the human SSB is also called RPA for replication protein A): Binds to single-stranded DNA and eliminates intra-strand base pairing, rendering the strand a better template for DNA polymerase. SSB is displaced from the template strand by the DNA polymerase during the course of synthesis.

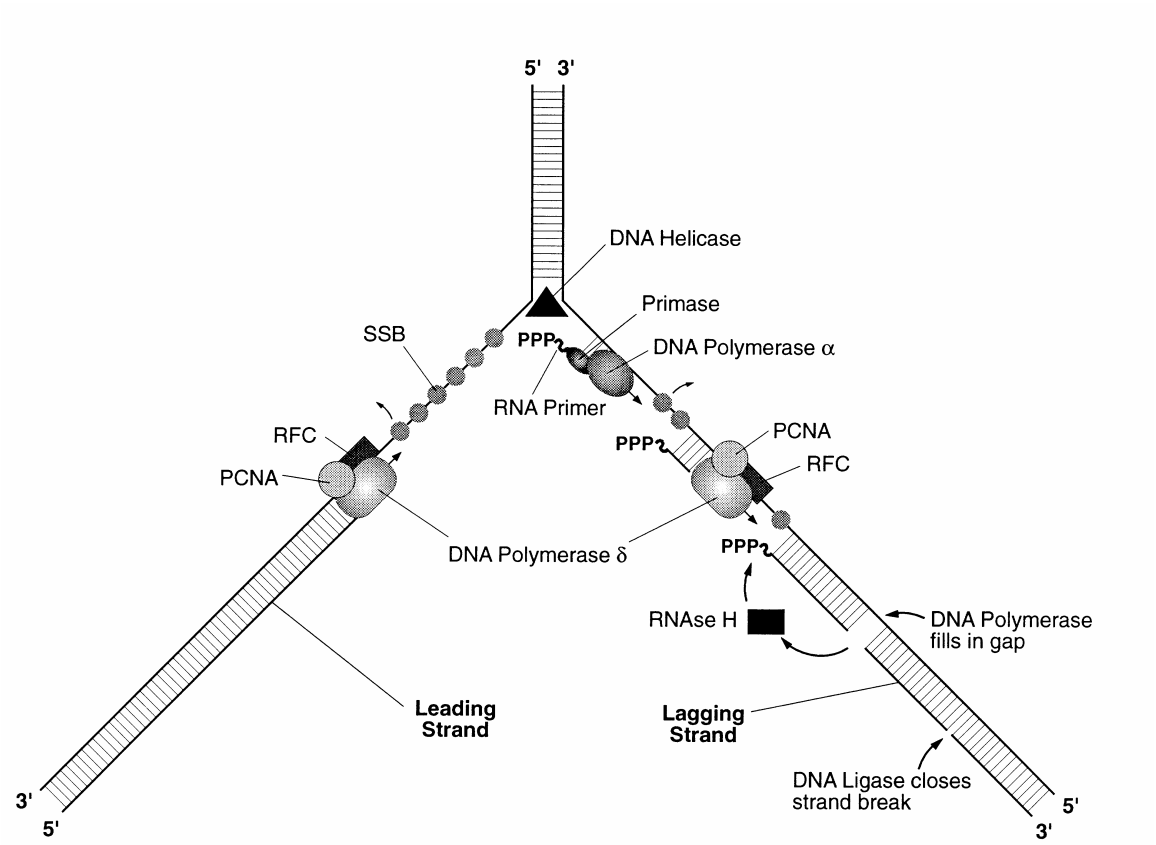
RNAse H: Hydrolyzes RNA in a RNA-DNA hybrid helix and thus serves to eliminate RNA primers. The gap resulting from RNAse H action is filled in by a DNA polymerase (probably DNA polymerase d) to generate the substrate for ligase.

DNA ligase: Catalyzes phosphodiester bond formation at a single strand break (nick):



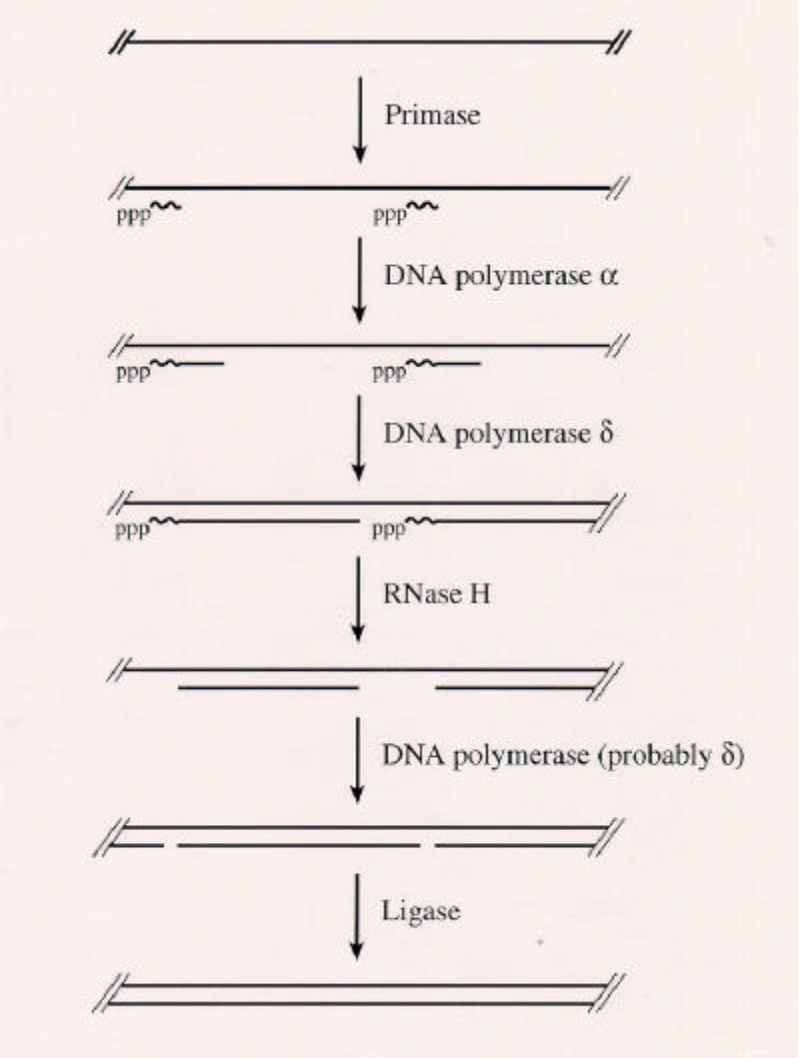
Model for the human replication fork

The molecular activities of the set of proteins described above can account for the events at the human replication fork as depicted below:



Model for Human Replication Fork

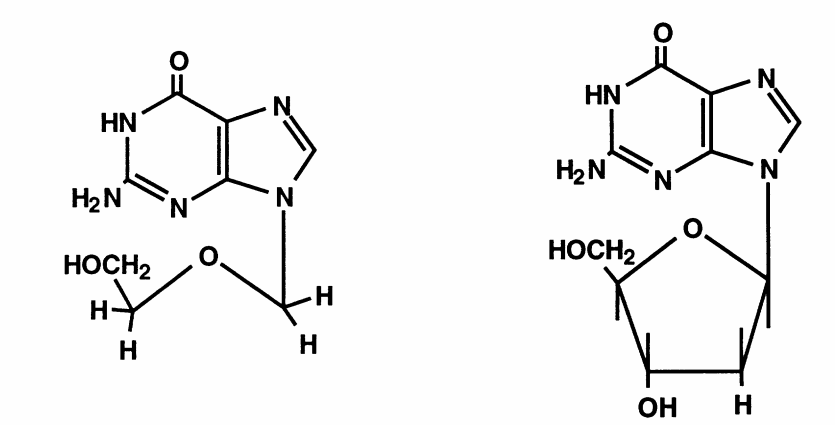
The steps involved in DNA synthesis on the lagging strand are outlined in more detail below:



Chain Terminating Nucleotides

Analogues of the dNTPs that lack a 3'-OH group can lead to chain termination by DNA polymerases, and several of the corresponding nucleosides have proven useful as antiviral agents.

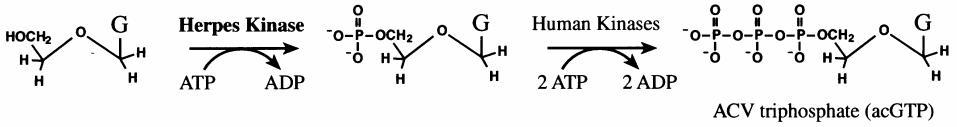
A. **Acycloguanosine** – an anti-herpes compound. Compare with deoxyguanosine that occurs naturally in DNA:



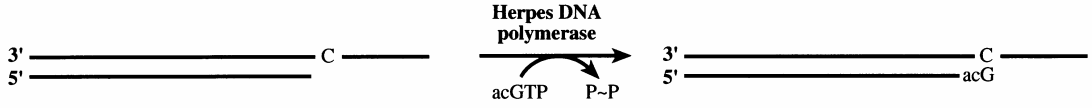
acycloguanosine (acyclovir)

deoxyguanosine

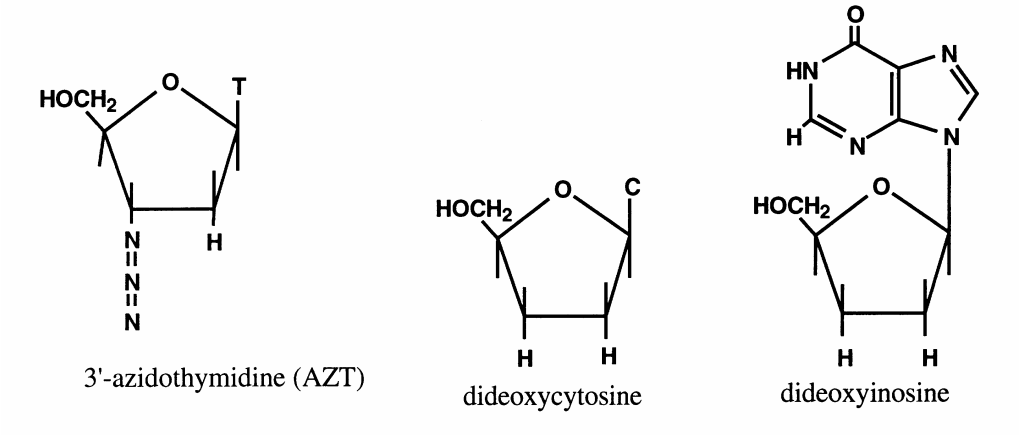
As depicted below, Acyclovir (ACV) is efficiently converted to the corresponding triphosphate in Herpes-infected cells, but not in uninfected human cells. Furthermore, although acG triphosphate is an extremely poor substrate for the human DNA polymerases, it is efficiently utilized by the DNA polymerase encoded by the virus. Addition of the analogue to growing Herpes DNA strands results in chain termination events and loss of the viral DNA's biological activity. In addition, the Herpes DNA polymerase binds tightly to the acycloguanosine-terminated DNA strand. This effectively inactivates the enzyme.



ACV triphosphate (acGTP)



B. **Chain Terminators as Anti-HIV-1 Drugs** – Dideoxycytosine (ddC) and dideoxyinosine (ddI) contain a 3'-H substituent rather than the 3'-OH present in deoxynucleosides. Like deoxycytosine, ddC pairs with G. ddI, an analogue of G, pairs with C forming two hydrogen bonds (compare base with G in deoxyguanosine above). AZT is a derivative of thymidine containing an azido group in place of the 3'-OH.



Each of these compounds is converted to the triphosphate form upon entry into human cells. While the corresponding nucleotides are poorly utilized by the human replicative polymerases ?, ?, and ?, they are efficiently added to growing retroviral DNA chains by the HIV-1 reverse transcriptase, the DNA polymerase responsible for replication of the viral genome. This mode of action is illustrated for AZT below.