Unit – 1
Properties of DNA and Gene Expression

IDENTIFICATION OF DNA

Deoxyribonucleic acid (DNA) was first described by Friedrich Miescher in 1869, only four years after Mendel's work was published. But it took over 80 years for its role as the genetic material of most organisms to become firmly established. DNA was first characterized as acid-precipitable material from the cell nuclei of pus and fish sperm. He called it nuclein, because it was associated exclusively with the nucleus. Further work demonstrated that nuclein is a complex of protein and DNA.

DNA AND HEREDITY

Clear experiments linking DNA to heredity were not performed until the mid 1940s. DNA was found almost exclusively in the chromosomes. Chromosomes also contain histone proteins and RNA, which did not seem as genetic material because sperms do not contain RNA, and the histones are replaced in sperm by a different protein, protamine. DNA (deoxyribose nucleic acid) has been found to be a genetic material in all the living beings except few plant viruses where RNA is the genetic material because DNA is not found in such viruses.

DNA as genetic material Experimental proof

The concept that DNA is the genetic material has been supported by the following evidences:

1. Bacterial Transformation or Transforming Principle (Griffith's Experiment) and identification of Transforming Principle by Avery, MacLeod and McCarty
2. Bacteriophage Infection (Hershey and Chase Experiment)

1. Bacterial Transformation or Transforming Principle (Griffith's Experiment) and identification of Transforming Principle by Avery, MacLeod and McCarty

The first unambiguous evidence that DNA was the hereditary material came from Frederick Griffith's studies in 1928. Griffith used the bacterium that causes pneumonia, *Diplococcus pneumoniae*. Virulence required the presence of a polysaccharide capsule around the bacterium.

This bacterium is found in two forms.

i) Virulent (disease producing), Smooth colony producing (S form)
ii) Non-virulent (not disease producing), Rough colony producing (R form)

i) Virulent (disease producing) Smooth colony producing (S form)

Whose cells produce capsule of polysaccharides (mucous), causing the colonies on agar Smooth (S). This strain is virulent (pathogenic) and causes pneumonia.

i) Non-virulent (not disease producing), Rough colony producing (R form)

Whose cells do not produce capsule of polysaccharides (mucous), causing the colonies on agar Rough (R). This strain is non-virulent (non pathogenic) and does not cause pneumonia.

Griffith performed his experiment by injecting the above bacteria into mice and found the following results:

a. S (virulent) bacteria were injected into mice; the mice developed pneumonia and finally died.
b. R (non-virulent) bacteria were injected into mice; the mice suffered no illness because R strain was non-pathogenic.
c. When Griffith injected heat killed S bacteria into mice, they did not suffer from pneumonia and thus survived.
d. A mixture of R (non-virulent) and heat killed S bacteria were injected into mice; the mice developed pneumonia and died.

Thus some genetic factor from dead S cells converted the live R cells into live S cells and the latter produced the disease. In short, living R cells were somehow transformed. So the Griffith effect gradually became known as transformation and turned out to be the first step in the identification of genetic material.

Identification of Transforming Principle by Avery, MacLeod and McCarty

In 1944, sixteen years after Griffith's experiment, Avery, MacLeod and McCarty reported successfully repetition of bacterial transformation, but in vitro. They were able to identify the transforming genetic material. They tested fractions of heat killed cells for transforming ability. Their findings were as under.

Mixture injected into healthy mice	Result obtained
1. R type living cells + Capsule of heat killed S type.	Mice did not develop pneumonia.
2. R type living cells + Cell wall of heat killed S type.	As above.
3. R type living cells + Cytoplasm of heat killed S type (without DNA)	As above.
4. R type living cells + DNA of heat killed S type.	Mice developed pneumonia and died.
5. R type living cells + DNA of heat killed S type + DNAase enzyme	Mice did not develop pneumonia.
6. R type living cells + DNA of heat killed S type + protease	Mice developed pneumonia and died.
7. R type living cells + DNA of heat killed S type + RNAse	Mice developed pneumonia and died.

Their findings were:

a. DNA alone from S bacteria caused R bacteria to become transformed.
b. They found that proteases (protein digesting enzymes) and RNAse (RNA digesting enzymes) did not affect transformation.
c. Digestion with DNAase inhibit transformation.

Thus they finally concluded that DNA is the hereditary material. Therefore, it is now beyond any reasonable doubt that DNA is the hereditary material.

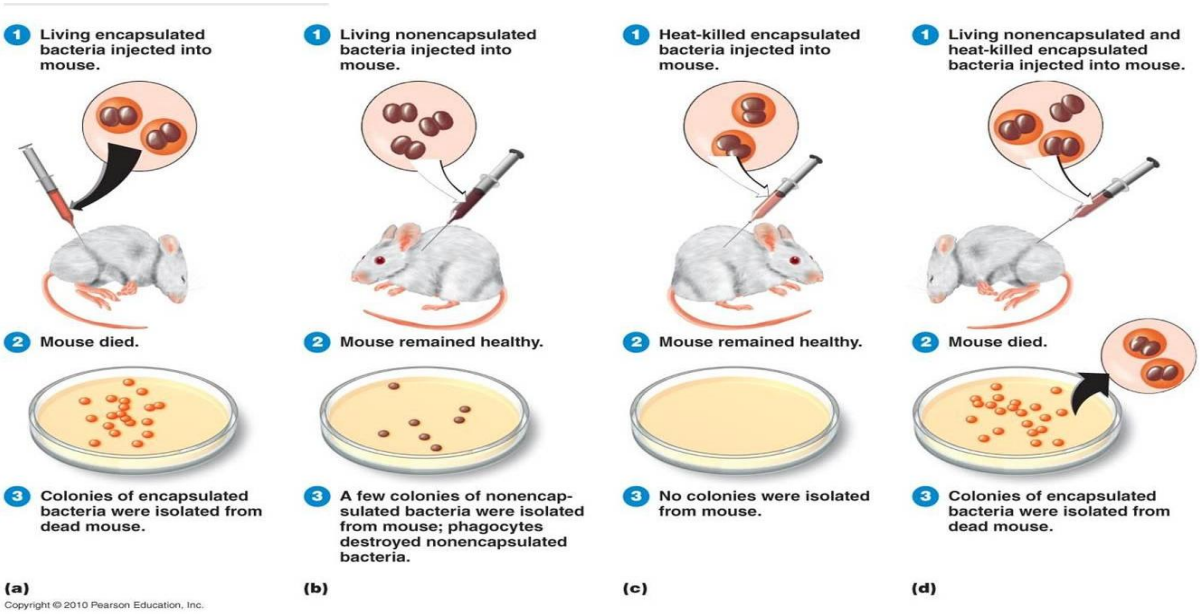


Fig. Griffith's discovery of the "transforming principle."

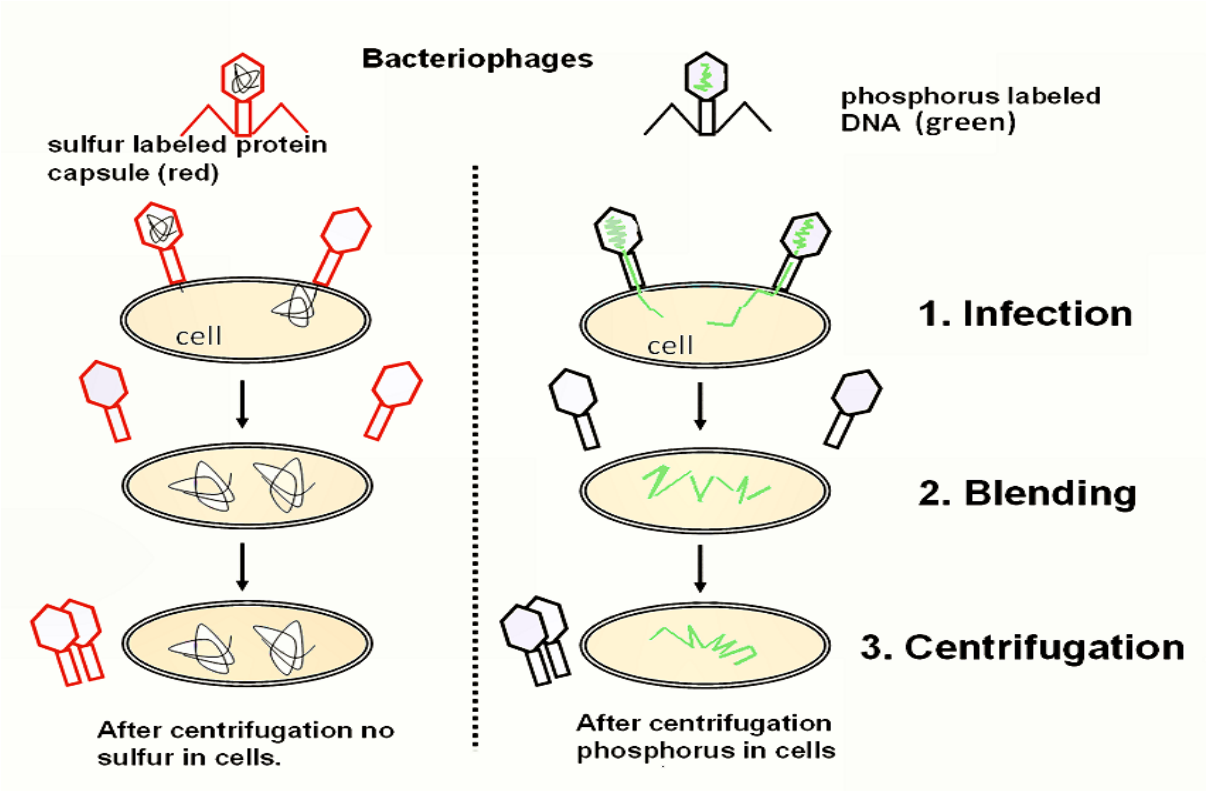
2. Bacteriophage Infection (Hershey and Chase Experiment)

In 1952, **Hershey and Chase performed experiment** by using bacteriohage T_2 and the bacterium *Escherichia coli* and proved DNA is the genetic material in some viruses. Viruses are small, very simple aggregates of nucleic acid and protein. T_2 contains only protein and DNA; the DNA forms the central core of the virus, while the protein surrounds the core like a coat. Phages infect bacteria by adsorbing to the cell walls and injecting the genetic material into the bacteria. This material causes the production of many new viruses within the cell. Eventually the cell is ruptured (lysed), and the new viruses are released.

The chemical make-up of protein and of DNA is quite different. Hershey and Chase used these differences to distinguish between them. DNA contains phosphorus and proteins do not contain phosphorus. Proteins contain sulfur, and DNA does not contain sulfur. By specifically labeling the phosphorus ^{32}P and sulfur ^{35}S atoms with radioisotopes, Hershey and Chase could distinguish between the protein and the DNA of the phage and determined whether either or both were injected into the bacterial cell during the course of infection.

i. Few bacteriophages were grown in bacteria having ^{32}P. This radioactive ^{32}P was incorporated to DNA of phage particles.
ii. Some other bacteriophages were grown in bacteria containing ^{35}S. This radioactive ^{35}S was incorporated into the cysteine and methionine amino acids of proteins that form outer coat of T_2 phage.
ii. These two radioactive phage preparations (one with radioactive DNA and another with radioactive proteins) were allowed to infect the culture of *E. coli.*
iii. It was observed that bacteriophage labeled with ^{32}P DNA were allowed to infect a cell, almost all the label entered the cell. If such infected cells were allowed to lyse, the label was found among the progeny viruses.
iv. The opposite occurred when bacteriophage labeled with ^{35}S protein infected a bacterial culture. Almost all label remains on the outside of the bacterium, bound to fragments of the cell wall.
v. If the purified ^{32}P was first treated with DNAase, no progeny phages were produced.

So, it can be safely concluded that during infection by bacteriophage T_2, it was DNA which entered the bacteria contained all the information necessary to produce new virus particles.



The Hershey-Chase experiment with bacterial viruses.

- **Molecular properties of DNA**

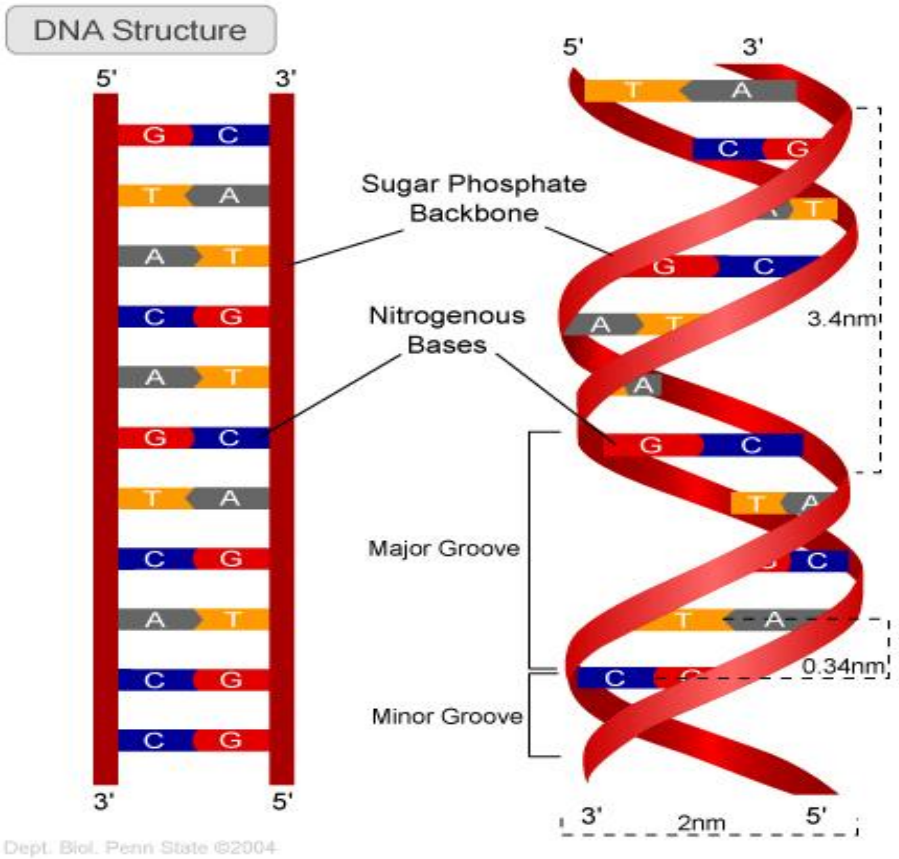
i. DNA was first identified and isolated by Miescher in 1869 at the University of Tübingen, a substance he called *nuclein*, and the double helix structure of DNA was first discovered in 1953 by Watson and Crick at the University of Cambridge, using experimental data collected by Franklin and Wilkins.

ii. **Deoxyribonucleic acid (DNA)** is a molecule that carries most of the genetic instructions used in the development, functioning and reproduction of all living organisms and many viruses.

iii. Within cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Within the chromosomes, chromatin proteins such as histones compact and organize DNA.

iv. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm.

v. DNA is a nucleic acid. DNA is a long polymer made from repeating units called nucleotides. DNA molecules consist of two biopolymer strands coiled around each other to form a double helix (the double helix), (duplex DNA). The structure of DNA of all species comprises two helical chains each coiled round the same axis, and each with a pitch of 34 ångströms (3.4 nanometres) and a diameter of 20 ångströms (2.0 nanometre). The general appearance of the polymer shows a periodicity of 3.4 A°, corresponding to the distance between 2 bases, and another one of 34 A°, corresponding to one helix turn (and also to 10 bases pairs).The two strands are twisted together like a twisted ladder -- the **double helix** (like a screw; right-handed). The two strands of DNA run in opposite directions to each other and are therefore anti-parallel.



vi. The two DNA strands are known as polynucleotides since they are composed of simpler units called nucleotides. DNA polymers can be very large molecules containing millions of nucleotides.

vii. Each nucleotide is composed of 3 components; a nitrogen-containing nucleobase—either Adenine (A), Guanine (G), Cytosine (C) or Thymine (T), a monosaccharide sugar called deoxyribose and a phosphate group.

viii. The nucleotides are joined to one another in a chain by **covalent bonds** between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone. Thus sides of the ladder comprise the sugar-phosphate portions.

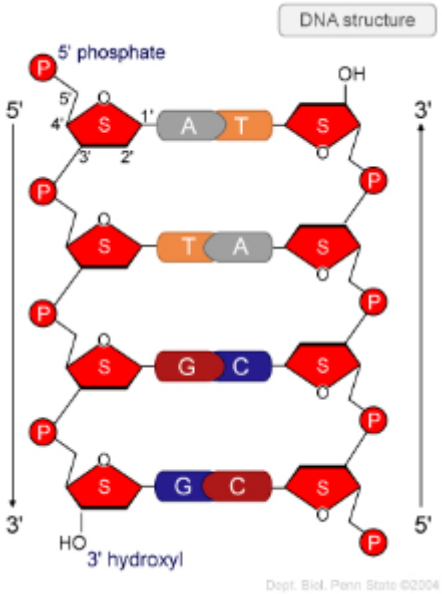
ix. The hydrogen bonds between phosphates cause the DNA strand to twist. The nitrogenous bases point inward on the ladder and form pairs with bases on the other side, like rungs.

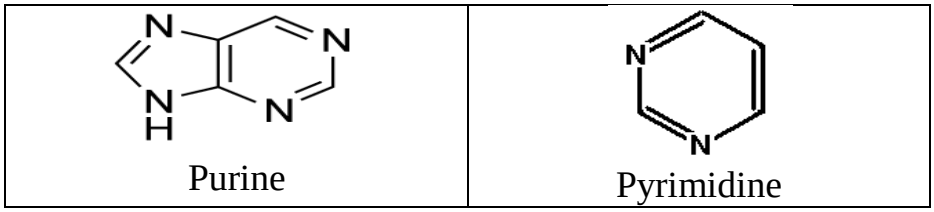
x. Each base pair is formed from two complementary nucleotides (purine with pyrimidine) bound together by hydrogen bonds. According to base pairing rules **adenine pairs to thymine with 2 hydrogen bonds** and **cytosine pairs to guanine with 3 hydrogen bonds** (A with T, and G with C). Hydrogen bonds bind the nitrogenous bases of the two separate polynucleotide strands to make double-stranded DNA.

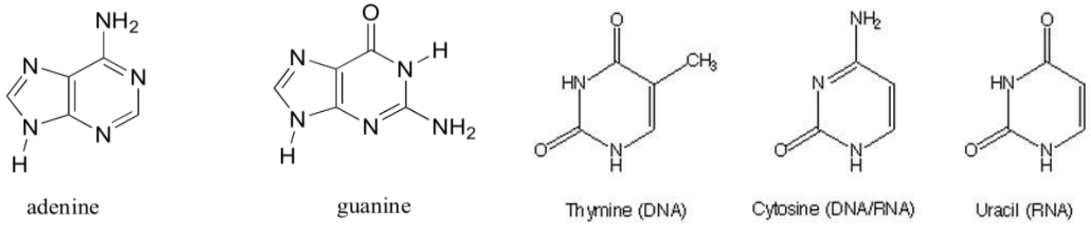
xi. **Nucleoside** - A nucleobase linked to a sugar is called a nucleoside.

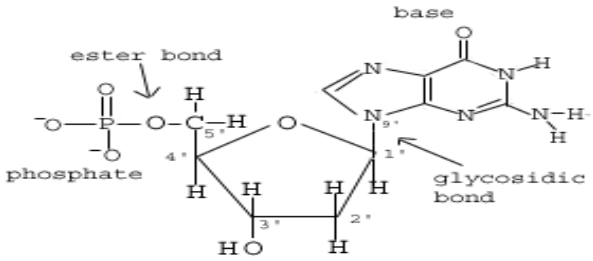
xii. **Nucleotide**- A base linked to a sugar and one or more phosphate groups is called a nucleotide. A polymer comprising multiple linked nucleotides is called a polynucleotide.

xiii. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5′ end having a terminal phosphate group and the 3′ end a terminal hydroxyl group.





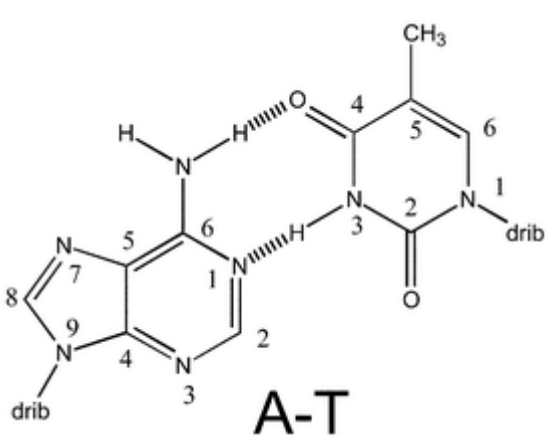


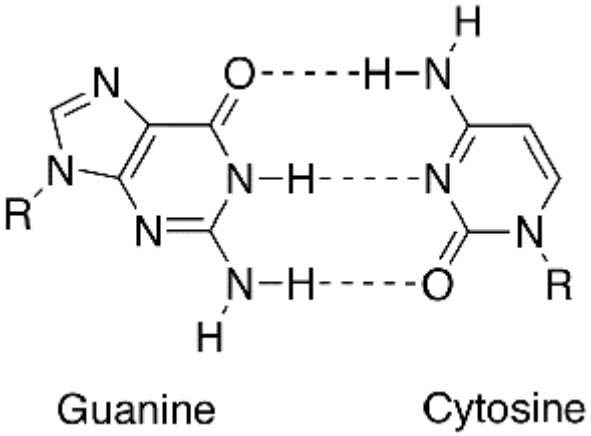


xiv. The DNA double helix is stabilized primarily by two forces: hydrogen bonds between nucleotides and base-stacking interactions among aromatic nucleobases.

xv. In a DNA double helix, each type of nucleobase on one strand bonds with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with adenine bonding only to thymine in two hydrogen bonds, and cytosine bonding only to guanine in three hydrogen bonds. This arrangement of two nucleotides binding together across the double helix is called a base pair.

xvi. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature. As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication.





==========

- **Molecular properties of DNA**

- **Denaturation / Melting of DNA**

DNA denaturation, also called DNA melting, is the process by which double-stranded deoxyribonucleic acid unwinds and separates into single-stranded strands through the breaking of hydrogen bonding between the bases. Both terms are used to refer to the process as it occurs when a mixture is heated.

When a DNA solution is heated enough, the double-stranded DNA unwinds, and the Hydrogen bonds that hold the two strands together weaken and finally break. The process of breaking a double-stranded DNA into single strands is known as DNA denaturation, or DNA melting.

Melting temperature(Tm): The temperature at which the DNA strands are half denatured, meaning half double-stranded, half single-stranded, is called the **melting temperature(Tm)**. The T_m is defined as the temperature in degrees Celsius, at which 50% of all molecules of a given DNA sequence are hybridized into a double strand, and 50% are present as single strands.

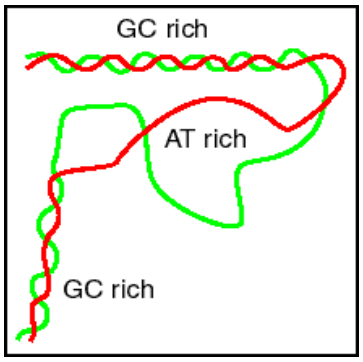
The amount of strand separation, or melting, is measured by the absorbance of the DNA solution at 260nm. Nucleic acids absorb light at this wavelength because of the electronic structure in their bases, but when two strands of DNA come together, the close proximity of the bases in the two strands reduce some of this absorbance. When the two strands separate, the absorbance rises 30%-40%.This is called Hyperchromicity. The Hypochromic effect is the effect of stacked bases in a double helix absorbing less ultra-violet light.

Note that 'melting' in this sense is not a change of aggregate state, but simply the dissociation of the two molecules of the DNA double helix.

The Tm is affected by a number of factors:

- Concentration of DNA
- Concentration of ions in the solution, most notably Mg^+ and K^+
- DNA sequence
- Length of DNA

While the ratio of G to C and A to T in an organism's DNA is fixed, the GC content (percentage of G +C) can vary considerably from one DNA to another. The percentage of GC content of DNA has a significant effect on its T_m. Because G-C pairs form three hydrogen bonds, while A-T pairs form only two, the higher the percentage of GC content, the higher its T_m. Thus, double-stranded DNA rich in G and C needs more energy to be broken than one that is rich in A and T, meaning higher melting temperature(T_m). Above the T_m, DNA denaturizes, below it, DNA anneals. Annealing is the reverse of denaturation.



❖ **Uses of DNA melting**

➢ **GC content analysis**

The process of DNA denaturation can be used to analyze some aspects of DNA. Because cytosine / guanine base-pairing is generally stronger than adenosine / thymine base-pairing, the amount of cytosine and guanine in a genome (called the "GC content") can be estimated by measuring the temperature at which the genomic DNA melts. Higher temperatures are associated with high GC content.

➢ **DNA-DNA hybridization**

DNA denaturation can also be used to detect sequence differences between two different DNA sequences. DNA is heated and denatured into single-stranded state, and the mixture is cooled to allow strands to rehybridize. Hybrid molecules are formed between similar sequences and any differences between

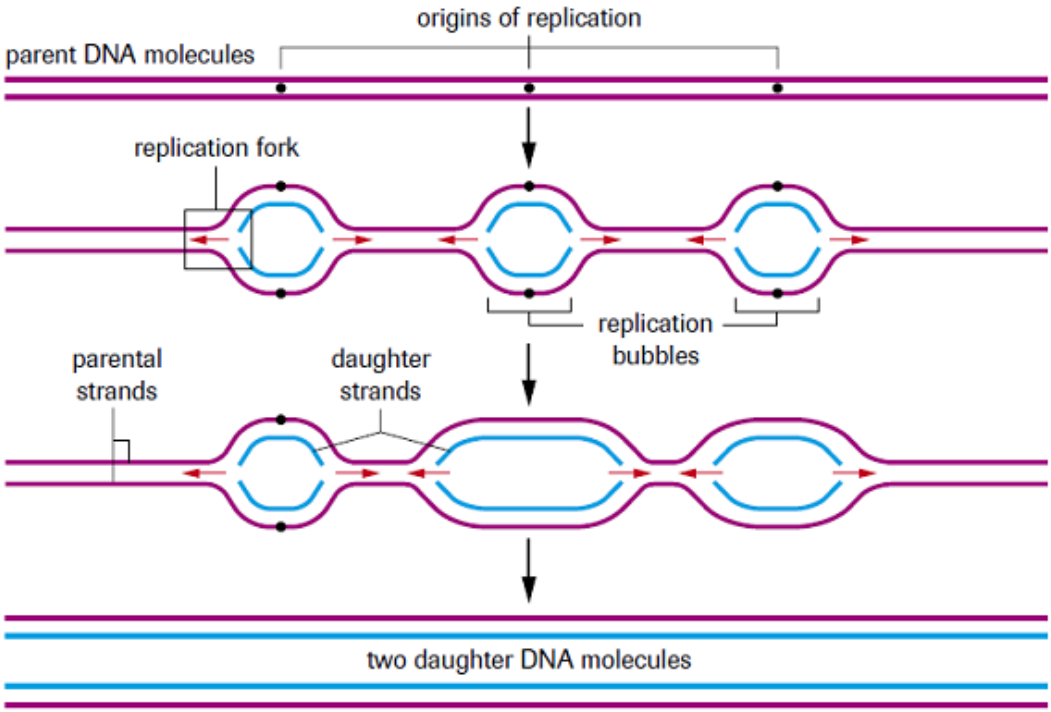
those sequences will result in a disruption of the base-pairing. On a genomic scale, the method has been used by researchers to estimate the genetic distance between two species, a process known as DNA-DNA hybridization.

- **DNA microarrays**

The process of DNA melting is also used in molecular biology techniques, notably in the polymerase chain reaction (PCR). Although the temperature of DNA melting is not diagnostic in the technique, methods for estimating T_m are important for determining the appropriate temperatures to use in a protocol. DNA melting temperatures can also be used as a proxy for equalizing the hybridization strengths of a set of molecules, e.g. the oligonucleotide probes of DNA microarrays.

- **DNA breathing**

Under physiological conditions, local DNA-breathing occurs spontaneously due to thermal fluctuations. This opens up transient bubbles of a few tens of base pairs. These breathing fluctuations may be supported by single-strand binding proteins, thereby lowering the DNA base pair stability. DNA breathing and the lability of local stretches of the DNA double-helix is essential for numerous physiological processes such as the association of single-strand binding proteins, and the initiation of replication and transcription.



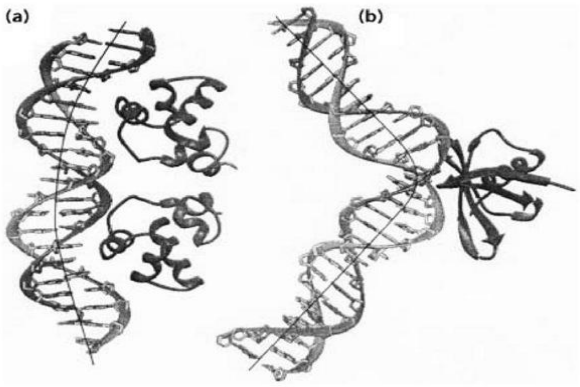
- **DNA Bending**

DNA is a relatively rigid polymer, typically modeled as a worm-like chain. It has three significant degrees of freedom; bending, twisting and compression, each of which causes particular limitations on what is possible with DNA within a cell. Twisting/torsional stiffness is important for the circularization of DNA and the orientation of DNA bound proteins relative to each other and bending/axial stiffness is important for DNA wrapping and circularization and protein interactions. Compression/extension is relatively unimportant in the absence of high tension.

DNA in solution does not take a rigid structure but is continually changing conformation due to thermal vibration and collisions with water molecules, which makes classical measures of rigidity impossible. Hence, the bending stiffness of DNA is measured by the persistence length, defined as:

> "The length of DNA over which the time-averaged orientation of the polymer becomes uncorrelated by a factor of e".

This value may be directly measured using an atomic force microscope to directly image DNA molecules of various lengths. In an aqueous solution, the average persistence length is 46-50 nm or 140-150 base pairs (the diameter of DNA is 2 nm), although can vary significantly. This makes DNA a moderately stiff molecule.



Models for DNA bending

The entropic flexibility of DNA is remarkably consistent with standard polymer physics models such as worm-like chain model. Consistent with the worm-like chain model is the observation that bending DNA is also described by Hooke's law at very small forces. However for DNA segments less than the persistence length, the bending force is approximately constant and behavior deviates from the worm-like chain predictions.

This effect results in unusual ease in circularizing small DNA molecules and a higher probability of finding highly bent sections of DNA.

Bending preference

DNA molecules often have a preferred direction to bend, ie. anisotropic bending. This is, again, due to the properties of the bases which make up the DNA sequence - a random sequence will have no preferred bend direction, i.e. isotropic bending.

Preferred DNA bend direction is determined by the stability of stacking each base on top of the next. If unstable base stacking steps are always found on one side of the DNA helix then the DNA will preferentially bend away from that direction. **A** and **T** residues will be preferentially be found in the minor grooves on the inside of bends. This effect is particularly seen in DNA-protein binding where tight DNA bending is induced, such as in nucleosome particles.

- **DNA flexibility**

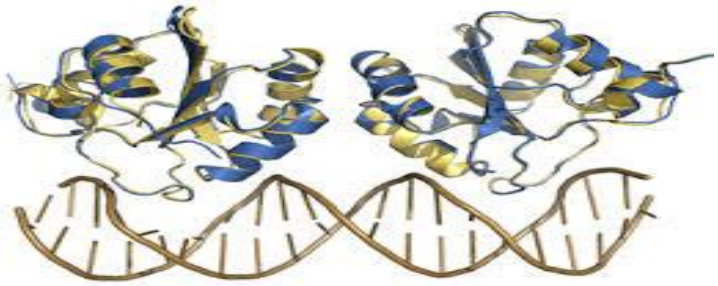
The early representations of DNA presented the molecule as a static straight rod. Yet in reality the DNA double helix is a highly dynamic structure that can both bend and twist. This flexibility can be described by two parameters: torsional flexibility, characterized by alterations in the twist angle between adjacent base pairs, and bending flexibility, in which the long axis of the double helix deviates, both locally and globally, from a straight trajectory.

These properties depend primarily on the physico-chemical properties of individual base steps and to a lesser extent on the DNA sequence context in which they are embedded. Since the physical characteristics of the 10 possible base pairs vary substantially, the flexibility of DNA is a sequence-dependent property and varies both locally, depending on the presence of a particular sequence, and globally, depending on the overall base composition of the molecule in question.

The major determinants of flexibility are the stacking energy of a given base step, the number of hydrogen bonds in a base pair, and the occupancy of both the major and minor grooves by exocyclic groups. Only the last parameter imparts directionality or anisotropy to bending flexibility and together with the dipole on the contiguous base pairs defines the conformational space available to a base step.

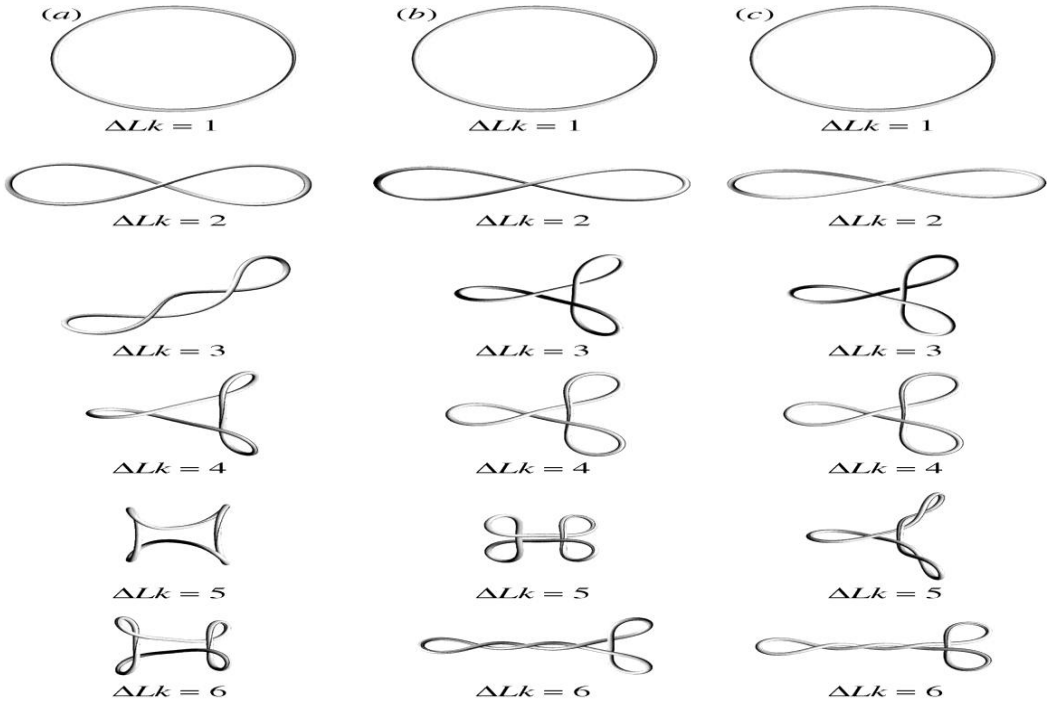
Both the torsional and bending flexibility of DNA have important biological consequences. The processes of transcription and DNA replication both require that the DNA be untwisted prior to the initiation of the copying of the DNA sequence. This untwisting is often initiated at particularly thermodynamically labile sequences.

Bending flexibility is also an essential aspect of biological function. In all living cells the DNA genome is packaged in a small volume and the necessary compaction is achieved by the tight bending of DNA around a protein scaffold. Similarly, the manipulation of DNA during transcription or DNA recombination often requires the tight bending of DNA around the appropriate protein assembly.



- **Linking Number of DNA**

The linking number, Lk, is the defining property of DNA elementary topological domains. Generally speaking, Lk is a measure of the total number of complete revolutions that either strand makes about the other. As long as the strands remain intact, Lk is a fixed quantity. Two otherwise identical closed duplex DNAs (cdDNA) that differ only in Lk are termed topoisomers. Even though they contain exactly the same nucleotide sequence and covalent connections, the properties of the members of a family of topoisomers depend strongly on Lk. The linking number of DNA can be described formally in several different ways, all of which may be readily generalized to any two closed curves in space.



Properties of the Linking Number

1. Lk is an integer for superhelical DNA (but not necessarily for protein-sealed DNA loops)—this follows from the requirement that both strands be closed curves.
2. In general, Lk is constant as long as the topological domain remains intact—a topological domain may be broken by a single- or double-stranded DNA chain scission or, if appropriate, by the disruption of links to the protein sealing the domain.
3. The linking number is a topological quantity, and its value is independent of DNA geometry—that is, Lk does not vary with deformation of the trajectory of either strand or with changes in the characteristic duplex geometric quantities (pitch, roll, twist, tilt, propeller twist, etc).
4. The linking number is independent of the ordering of the two curves; thus, for two DNA strands this independence of ordering clearly distinguishes the linking number from the twist
5. Finally, the double-helix structure of DNA defines a duplex axis (A) that is considered to be a continuous line along the center of the helix. Since either of the two backbone chains can be continuously deformed into the axis, the linking number of either strand about the axis is the same as the linking number of the strands about one another.

- **Major groove and minor groove**

The sugar-phosphate backbones spiral around the outer surface of DNA. Between these backbones are two different sized grooves, designated the major and minor grooves. The major groove is approximately 50% wider than the minor.

As the strands are not symmetrically located with respect to each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide. The width of the major groove means that the edges of the bases are more accessible in the major groove than in the minor groove. As a result, proteins such as transcription factors that can bind to specific sequences in double-stranded DNA usually make contact with the sides of the bases exposed in the major groove.

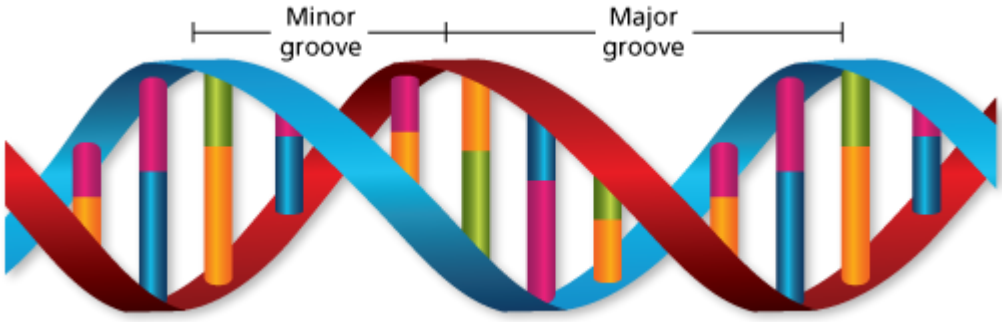
The double helix is a quite rigid and viscous molecule of an immense length and a small diameter. It presents a major groove and a minor groove. The major groove is deep and wide, the minor groove is narrow and shallow.

DNA-protein interactions are major/essential processes in the cell life (transcription activation or repression, DNA replication and repair).

Proteins bind at the floor of the DNA grooves, using specific binding: hydrogen bounds, and non specific binding: van der Waals interactions, generalized electrostatic interactions.

Proteins recognize H-bond donors, H-bond acceptors, methyl groups (hydrophobic), the later being exclusively in the major groove; there are 4 possible patterns of recognition with the major groove, and only 2 with the minor groove (see iconography).

Some proteins bind DNA in its major groove, some other in the minor groove, and some need to bind to both.



==========

- **Novel DNA Structures**

DNA can exist as single, double-stranded, or mixed forms. dsDNA has a linear sequence (primary structure), secondary structure (right handed double helix), and tertiary/quaternary structure (it is folded and packed in the cell).The **primary structure** is the sequence of nucleoside monophosphates (usually written as the sequence of bases they contain).The **secondary structure** refers to the shape a nucleic acid assumes as a result of the **primary structure**. A-DNA, B-DNA and Z-DNA are forms of **secondary structure. Tertiary structure** refers to large-scale folding in a linear polymer that is at a higher order than **secondary structure**. The **tertiary structure** is the specific three-dimensional shape into which an entire chain is folded.

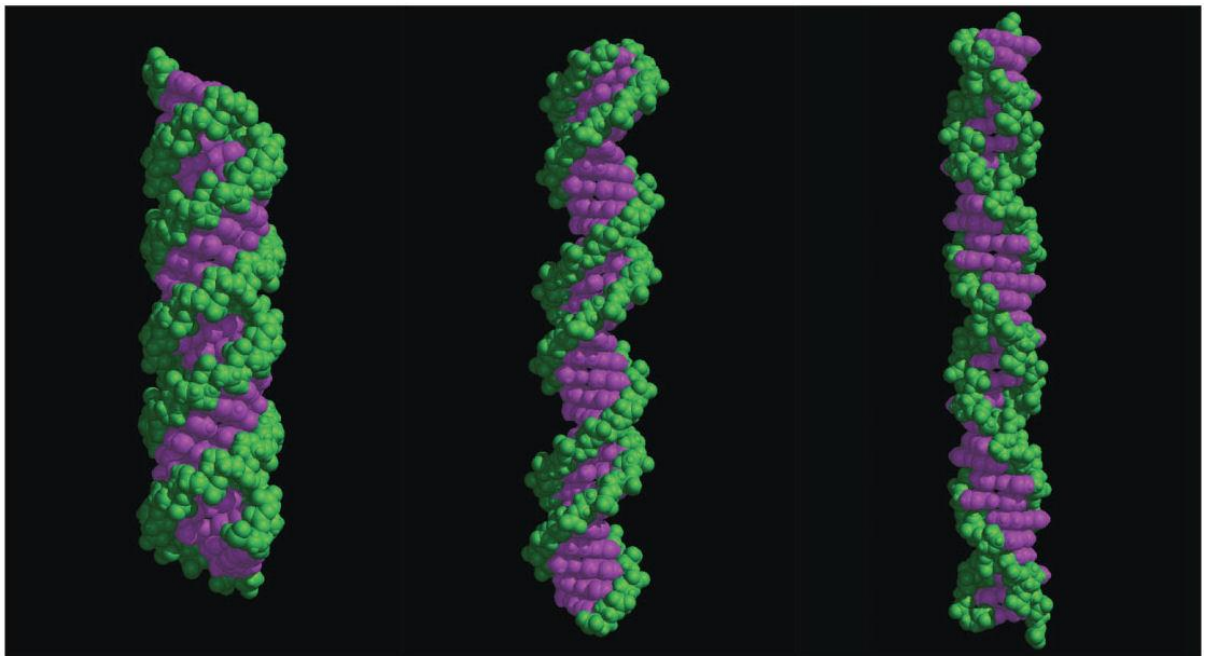
A DNA: - The **right-handed 'A' form** occurs in crystalline structures where the water concentration is reduced. The structure is distorted and the bases are no longer co-planar. A-DNA occurs when DNA is dehydrated, but also in DNA/RNA hybrids. It is a right-handed double helix fairly similar to the more common and well-known B-DNA form, but with a shorter more compact helical structure whose base pairs are not perpendicular to the helix-axis as in B-DNA. The same helical conformation is the most commonly seen one in double-stranded RNA's. There is a slight increase in the number of base pairs (bp) per rotation (resulting in a tighter rotation angle), and smaller rise per turn. This

results in a deepening of the major groove and a shallowing of the minor. A thicker right-handed duplex with a shorter distance between the base pairs has been described for RNA-DNA duplexes and RNA-RNA duplexes.

B DNA: - The **right-handed 'B' form** is the standard form in biological systems. The double stranded DNA molecule is a right-handed helix as determined by Watson and Crick using Franklin's x-ray diffraction images. This B-form of DNA has approximately 10 nucleotides per turn of the helix and is the most common form of DNA found in nature. The most common form, present in most DNA at neutral pH and physiological salt concentrations, is B-form. That is the classic, right-handed double helical structure. The major difference between A-form and B-form nucleic acid is in the conformation of the deoxyribose sugar ring. It is in the C2' endoconformation for B-form, whereas it is in the C3' endoconformation in A-form. In the C2' endoconformation, the C2' atom is above the plane, whereas the C3' atom is above the plane in the C3' endoconformation. The latter conformation brings the 5' and 3' hydroxyls (both esterified to the phosphates linking to the next nucleotides) closer together than is seen in the C2' endoconfromation. Thus the distance between adjacent nucleotides is reduced by about 1 Angstrom in A-form relative to B-form nucleic acid. In B-form, the base-pairs are almost centered over the helical axis, but in A-form, they are displaced away from the central axis and closer to the major groove. The result is a ribbon-like helix with a more open cylindrical core in A-form.

Z DNA: - The **left-handed 'Z' form** occurs occasionally in the middle of a **B**-form molecule. A small amount of the DNA in a cell exists in the Z form. Z-DNA is a radically different duplex structure, with the two strands coiling in left-handed helices and a pronounced zig-zag (hence the name) pattern in the phosphodiester backbone. Bases are co-planar in both the **B** and **Z** forms. The high salt and GC base-pairs, used to form the DNA crystals cause the helix to twist in a left-handed way, creating a structure called Z-DNA. A third form of duplex DNA has a strikingly different, left-handed helical structure. This Z DNA is formed by stretches of alternating purines and pyrimidines, e.g. GCGCGC, especially in negatively supercoiled DNA. The big difference is at the G nucleotide. It has the sugar in the C3' endoconformation (like A-form nucleic acid, and in contrast to B-form DNA) and the guanine base is in the synconformation. This places the guanine back over the sugar ring, in contrast to the usual anticonformation seen in A- and B-form nucleic acid. Note that having the base in the anticonformation places it in the position where it can readily form H-bonds with the complementary base on the opposite strand. The duplex in Z-DNA has to accommodate the distortion of this G nucleotide in the synconformation. The cytosine in the adjacent nucleotide of Z-DNA is in the "normal" C2' endo, anticonformation.

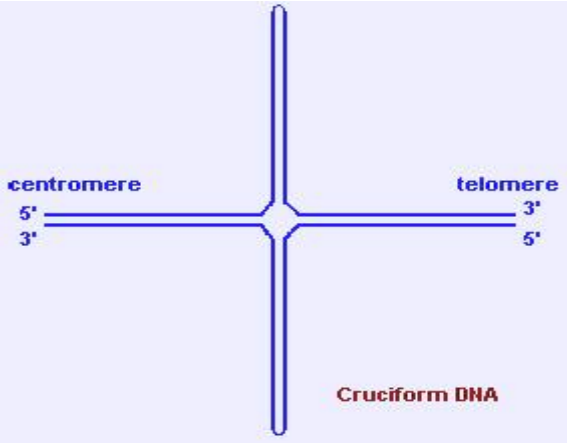
a) A-DNA b) B-DNA c) Z-DNA

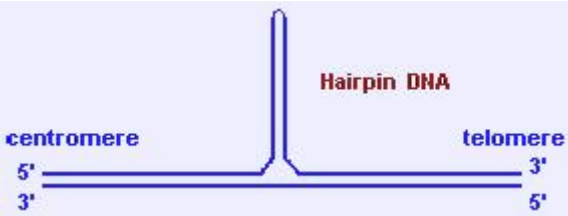


Feature	A-DNA	B-DNA	Z-DNA
Type of helix	Right-handed	Right-handed	Left-handed
Helical diameter (nm)	2.55	2.37	1.84
Rise per base pair (nm)	0.29	0.34	0.37
Distance per complete turn (pitch) (nm)	3.2	3.4	4.5
Number of base pairs per complete turn	11	10	12
Topology of major groove	Narrow, deep	Wide, deep	Flat
Topology of minor groove	Broad, shallow	Narrow, shallow	Narrow, deep

Cruciform DNA and hairpin DNA

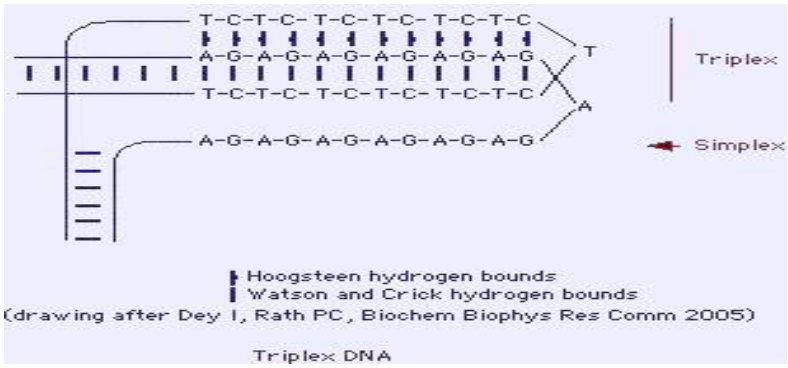
Junctions formed during recombination are cruciform structures. Inverted (or mirror) repeats (palindromes) of polypurine / polypyrimidine DNA stretches can also form cruciform or hairpin structures through intra-strand pairing. Palindromic AT-rich repeats are found at the breakpoints. Nucleases bind and cleave junctions after recombination. Other well known proteins such as HMG proteins and MLL can also bind cruciform DNA.





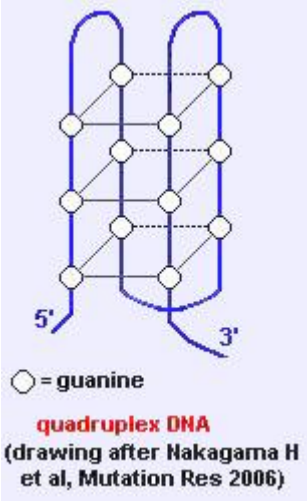
H-DNA or triplex DNA

Inverted repeats (palindromes) of polypurine/polypyrimidine DNA stretches can form triplex structures (triple helix). A triple-stranded plus a single stranded DNA are formed. H-DNA may have a role in functional regulation of gene expression as well as on RNAs (e.g. repression of transcription).



G4-DNA

G4 DNA or quadruplex DNA: folding of double stranded GC-rich sequence onto itself forming Hoogsteen base pairing between 4 guanines ("G4"), a highly stable structure. Often found near promotors of genes and at the telomeres. Role in meiosis and recombination; may be regulatory elements.



Quaternary structure of the molecule - Chromatin

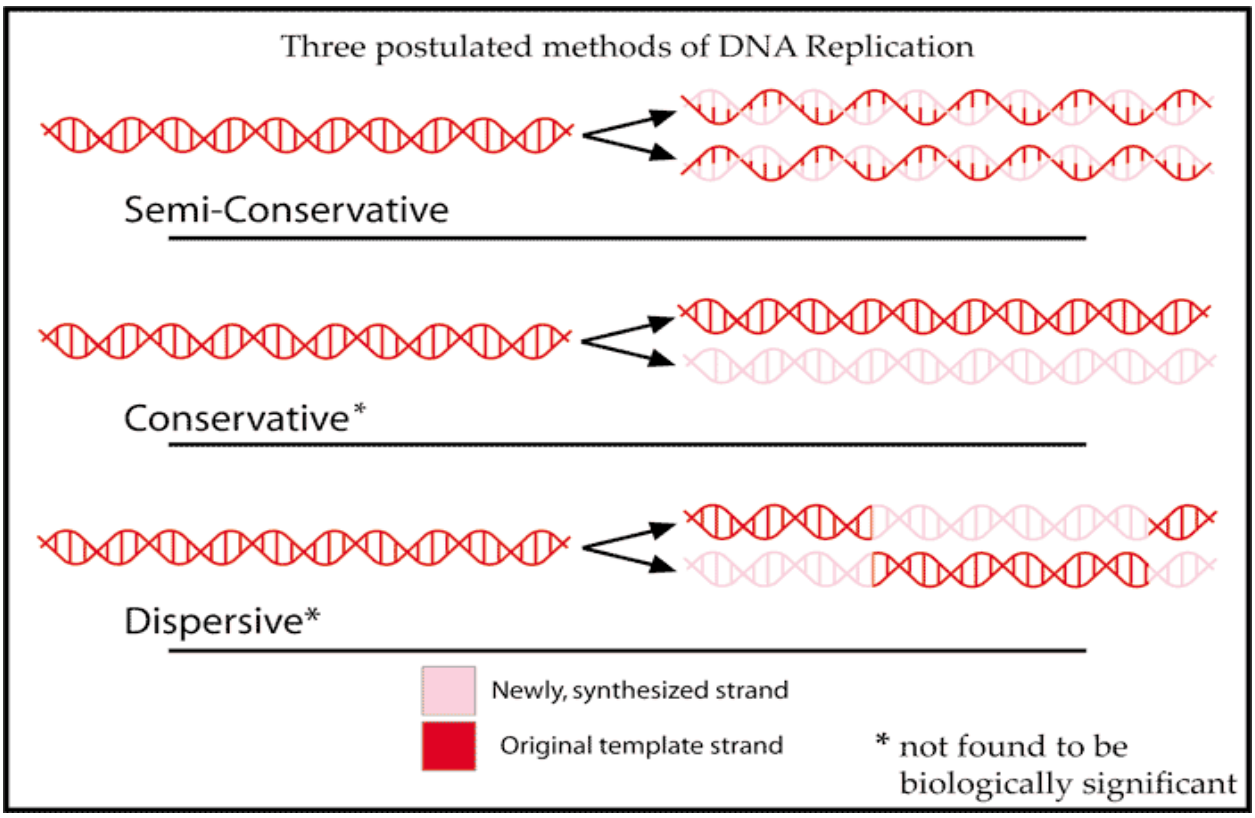
DNA is associated with proteins: histones and non histone proteins, to form the chromatin. DNA as a whole is acidic (negatively charged) and binds to basic (positively charged) proteins called histones. There is 3 x 10^{9} nucleotide pairs in the human haploid genome representing about 30 000 genes dispersed over 23 chromosomes for a haploid set.

- **DNA Replication**

- **DNA Replication models**

The mechanism of replication was one of three models originally proposed for DNA replication:

- **Semiconservative replication** would produce two copies that each contains one of the original strands and one new strand.
- **Conservative replication** would leave the two original template DNA strands together in a double helix and would produce a copy composed of two new strands containing all of the new DNA base pairs.
- **Dispersive replication** would produce two copies of the DNA, both containing distinct regions of DNA composed of either both original strands and both new strands.



The semiconservative model was confirmed by the Meselson-Stahl experiment. The **Meselson–Stahl experiment** is an experiment by Matthew Meselson and Franklin Stahl in 1958 which supports the hypothesis that DNA replication is semiconservative. Semiconservative replication means that when the double stranded DNA helix replicates, each of the two double stranded DNA helices consists of one strand coming from the original helix and one newly synthesized. It has been called the most beautiful experiment in biology.

Experimental procedure and results

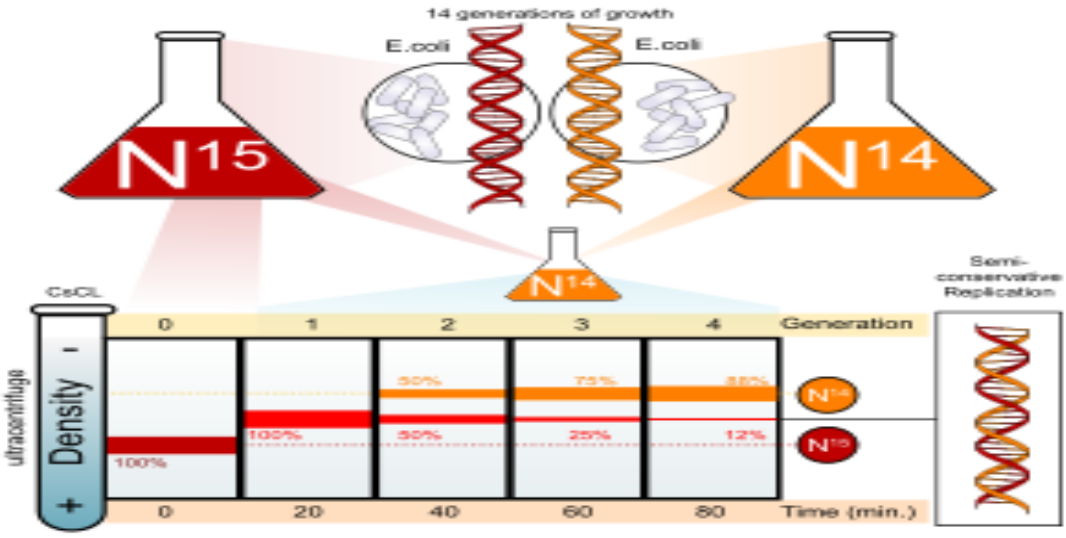


Fig. Meselson and Stahl's experiment

Nitrogen is a major constituent of DNA. ^{14}N is by far the most abundant isotope of nitrogen, but DNA with the heavier (but non-radioactive) ^{15}N isotope is also functional.

E. coli were grown for several generations in a medium with ^{15}N. When DNA is extracted from these cells and centrifuged on a salt density gradient, the DNA separates out at the point at which its density equals that of the salt solution. The DNA of the cells grown in ^{15}N medium had a higher density than cells grown in normal ^{14}N medium. After that, *E. coli* cells with only ^{15}N in their DNA were transferred to a ^{14}N medium and were allowed to divide; the progress of cell division was monitored by measuring the optical density of the cell suspension.

DNA was extracted periodically and was compared to pure ^{14}N DNA and ^{15}N DNA. After one replication, the DNA was found to have close to the intermediate density. Since conservative replication would result in equal amounts of DNA of the higher and lower densities (but no DNA of an intermediate density), conservative replication was excluded. However, this result was consistent with both semiconservative and dispersive replication. Semiconservative replication would result in double-stranded DNA with one strand of ^{15}N DNA, and one of ^{14}N DNA, while dispersive replication would result in double-stranded DNA with both strands having mixtures of ^{15}N and ^{14}N DNA, either of which would have appeared as DNA of an intermediate density.

The authors continued to sample cells as replication continued. DNA from cells after two replications had been completed was found to consist of equal amounts of DNA with two different densities, one corresponding to the intermediate density of DNA of cells grown for only one division in ^{14}N medium, the other corresponding to DNA from cells grown exclusively in ^{14}N medium. This was inconsistent with dispersive replication, which would have resulted in a single density, lower than the intermediate density of the one-generation cells, but still higher than cells grown only in ^{14}N DNA medium, as the original ^{15}N DNA would have been split evenly among all DNA strands. The result was consistent with the semiconservative replication hypothesis

- **Mechanism of DNA replication**

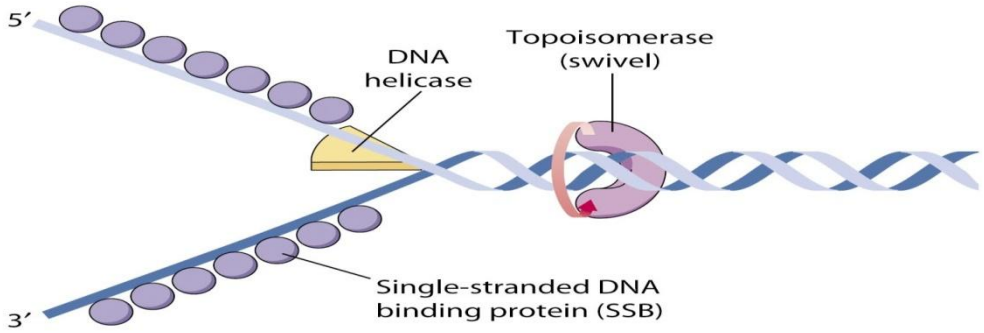
DNA replication is a biological process that occurs in all living organisms and copies their DNA; it is the basis for biological inheritance. The process starts with one double-stranded DNA molecule and produces two identical copies of the molecule. Each strand of the original double-stranded DNA molecule serves as template for the production of the complementary strand. Cellular proofreading and error toe-checking mechanisms ensure near perfect fidelity for DNA replication. DNA replicates by the **semiconservative**

model of replication. This model states that during replication of double stranded DNA, the two strands separate and each separate strand is then used as a **template** to make a new strand. In this manner, the cell ends up with two identical molecules of the original DNA, with each molecule containing one strand from the original piece of DNA and one newly synthesized strand.
DNA replication consists of three steps--

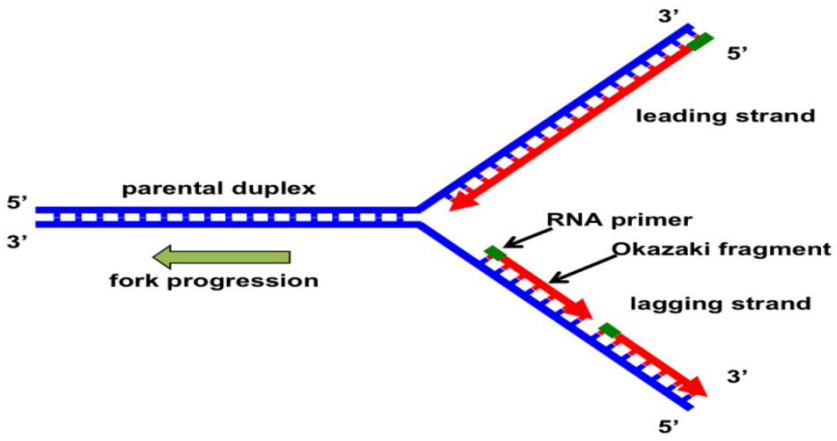
1. Initiation
2. Elongation
3. Termination

1. **Initiation:** - The initiation point where the splitting starts is called "origin of replication". The structure that is created is known as "**Replication Fork**". The first major step for the **DNA Replication** to take place is the breaking of hydrogen bonds between bases of the two antiparallel strands. During replication the DNA double helix must be unwound to generate separate single strands. The unwinding of the two strands is the starting point the two strands. The DnaA protein binds to *origin* while hydrolyzing ATP. This leads to the initial unwinding of double-stranded DNA at the initiation site. Further unwinding occurs through the activity of the DnaB protein, a helicase. The enzyme that separates the strands is known as **helicase**. This enzyme first binds to a specific site on the DNA known as the origin of replication and separates the strands - this allows the DNA polymerase to bind. As the replication process continues, the helicase moves down the DNA and to continue separating the DNA strands.

Rapid unwinding can lead to tension and formation of supercoils or supertwists in the helix, just as rapid separation of two strands of a rope can lead to knotting or coiling of the rope. The tension generated by unwinding is relieved, and the unwinding process is promoted by enzyme known as **topoisomerase (**also known as DNA **gyrase** in bacteria**).** This enzyme changes the structure of DNA by transiently breaking one or two strands in such a way that it remains unaltered as its shape is changed. Once the strands have separated, they are kept single through specific binding with **single-stranded DNA binding proteins (SSBs).**

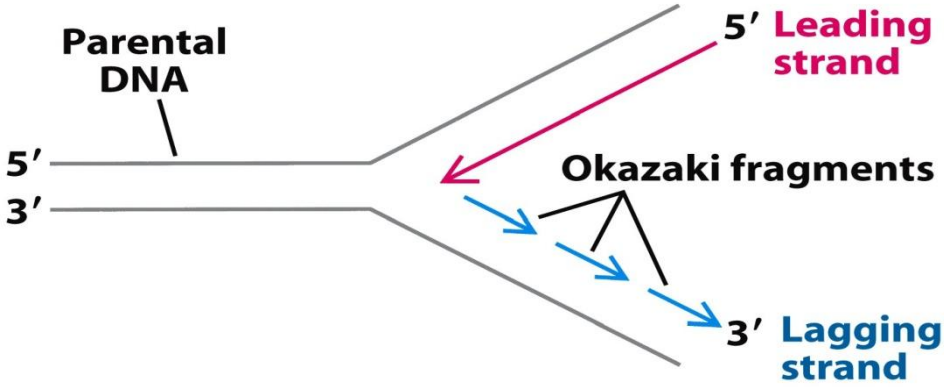


a) One of the most important **steps of DNA Replication** is the binding of **RNA Primase (RNA polymerase)** in the initiation point of the 3'-5' parent chain. Special RNA polymerase called a **primase** synthesizes a short RNA primer. DNA polymerase requires a primer to operate. That is, the polymerase can only add nucleotides to an existing strand (the primer). This existing strand is laid down by the enzyme **primase**. Primase lays down a strand of about 8-10 nucleotides that is complementary to the replicating strand. DNA polymerase III holoenzyme can then add the remaining nucleotides to the 3' end of the primer. RNA nucleotides are the primers (starters) for the binding of DNA nucleotides.



2) Elongation: - This process is different for the 5'-3' and 3'-5' template.

a)**5'-3' Template**: The proceeding daughter strand -that uses a **5'-3' template**- is called **leading strand** because **DNA Polymerase III** can "read" the template and continuously adds nucleotides (complementary to the nucleotides of the template, for example Adenine opposite to Thymine etc).



b) **3'-5'Template**: The **3'-5' template** cannot be "read" by DNA Polymerase III. The replication of this template is complicated and the new strand is called **lagging strand**. In the lagging strand the RNA Primase adds more RNA Primers. **DNA polymerase III** reads the template and lengthens the bursts. The

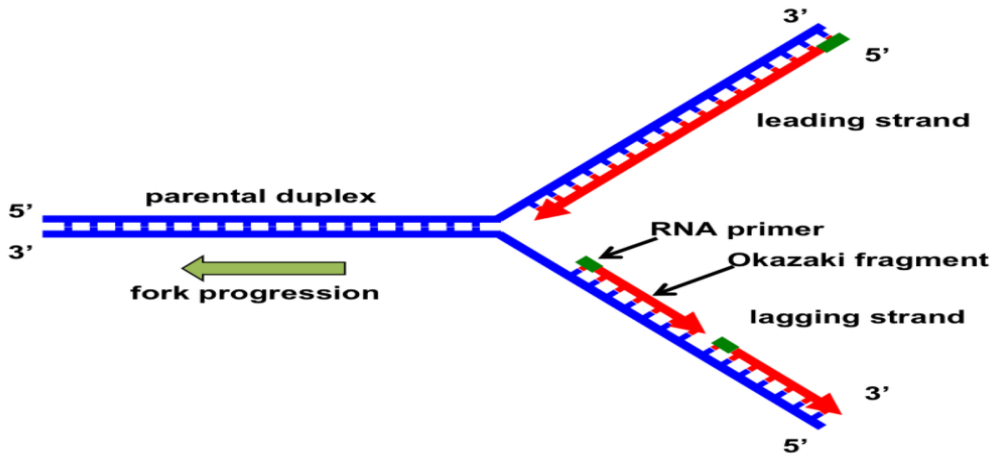
gap between two RNA primers consisting of single-stranded DNA is called "**Okazaki Fragments**". The RNA Primers are necessary for DNA Polymerase to bind Nucleotides to the 3' end of them. The daughter strand is elongated with the binding of more DNA nucleotides.

Okazaki fragments

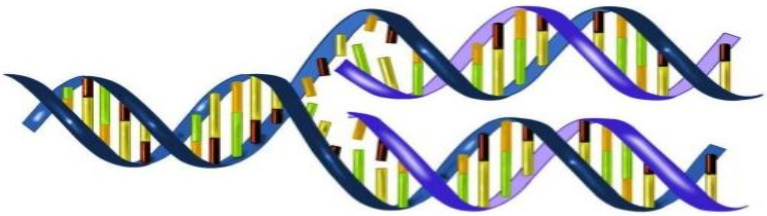
Okazaki fragments are short molecules of single-stranded DNA that are formed on the lagging strand during DNA replication. They are in between 1,000 to 2,000 nucleotides long in Escherichia coli and are in between 100 to 200 nucleotides long in eukaryotes.

On the leading strand DNA replication proceeds continuously along the DNA molecule as the parent double-stranded DNA is unwound. But on the lagging strand the new DNA is made in installments, which are later joined together by a DNA ligase enzyme. This is because the enzymes that synthesise the new DNA can only work in one direction along the parent DNA molecule. On the leading strand this route is continuous, but on the lagging strand it is discontinuous.

They were originally discovered in 1966 by Kiwako Sakabe and Reiji Okazaki in their research on DNA replication of *Escherichia coli*.



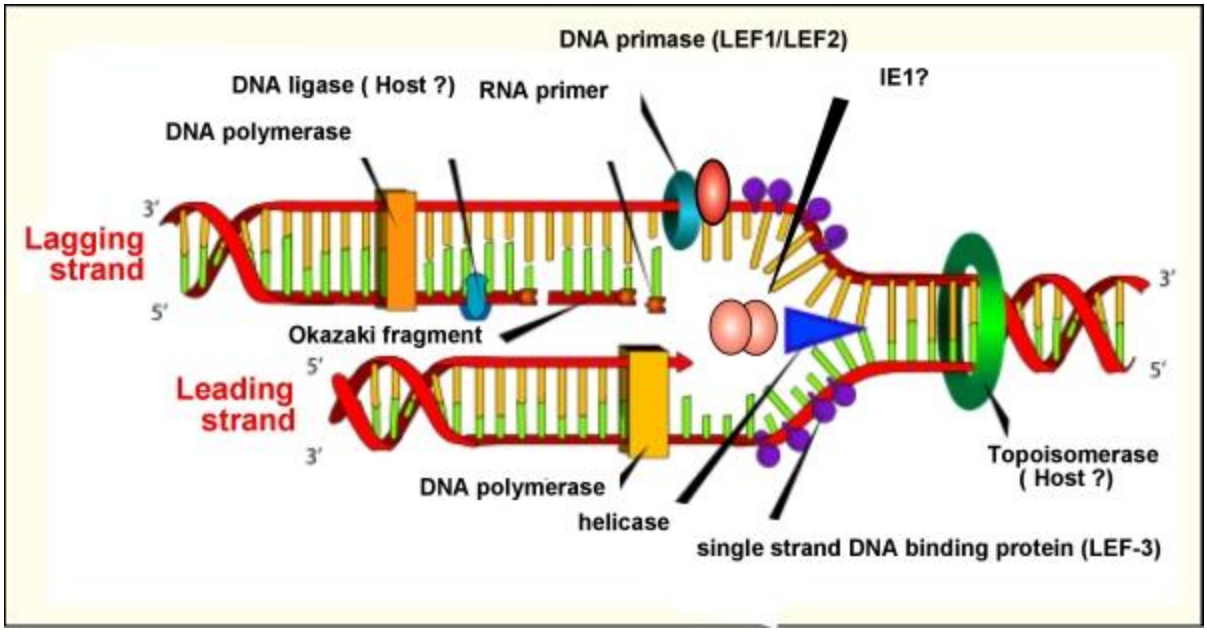
c) In the lagging strand the **DNA Pol I -exonuclease-** reads the fragments and removes the RNA Primers. The gaps are closed with the action of DNA Polymerase III (adds complementary nucleotides to the gaps) and DNA Ligase (adds phosphate in the remaining gaps of the phosphate - sugar backbone). Each new double helix is consisted of one old and one new chain. This is what we call **semiconservative replication**.



3) **Termination:-** The last **step of DNA Replication** is the **Termination**. This process happens when the DNA Polymerase reaches to an end of the strands. We can easily understand that in the last section of the lagging strand, when the RNA primer is removed, it is not possible for the DNA Polymerase to seal the gap (because there is no primer). So, the end of the parental strand where the last primer binds isn't replicated. These ends of linear (chromosomal) DNA consist of noncoding DNA that contains repeat sequences and are called **telomeres**. As a result, a part of the telomere is removed in every cycle of DNA Replication.

The DNA Replication is not completed before a **mechanism of repair** fixes possible errors caused during the replication. Enzymes like **nucleases** remove the wrong nucleotides and the DNA Polymerase fills the gaps.

Similar processes also happen during the **steps of DNA Replication** of prokaryotes though there are some differences.



==========

- **Structural models of DNA replication**

a) Theta (θ) Model of DNA Replication
b) Rolling circle model or Sigma (σ) model of DNA replication

a) Theta (θ) Model of DNA Replication

The *E. Coli* DNA is a circular. DNA is replicated while maintaining the integrity of the circle i.e., the circle does not appear to be broken in the process of DNA replication; an intermediate theta structure is formed which is due to the formation of replication eye. Replication of the DNA seems to be occurring at one or two moving Y-junctions in the circle Replication forks.

DNA replication is initiated at a specific nucleotide sequence, the **origin of replication**, and proceeds in two directions until the forks meet at the opposite side of the circular chromosome. At completion, the two circular daughter molecules are interlinked and are separated by topoisomerase II.

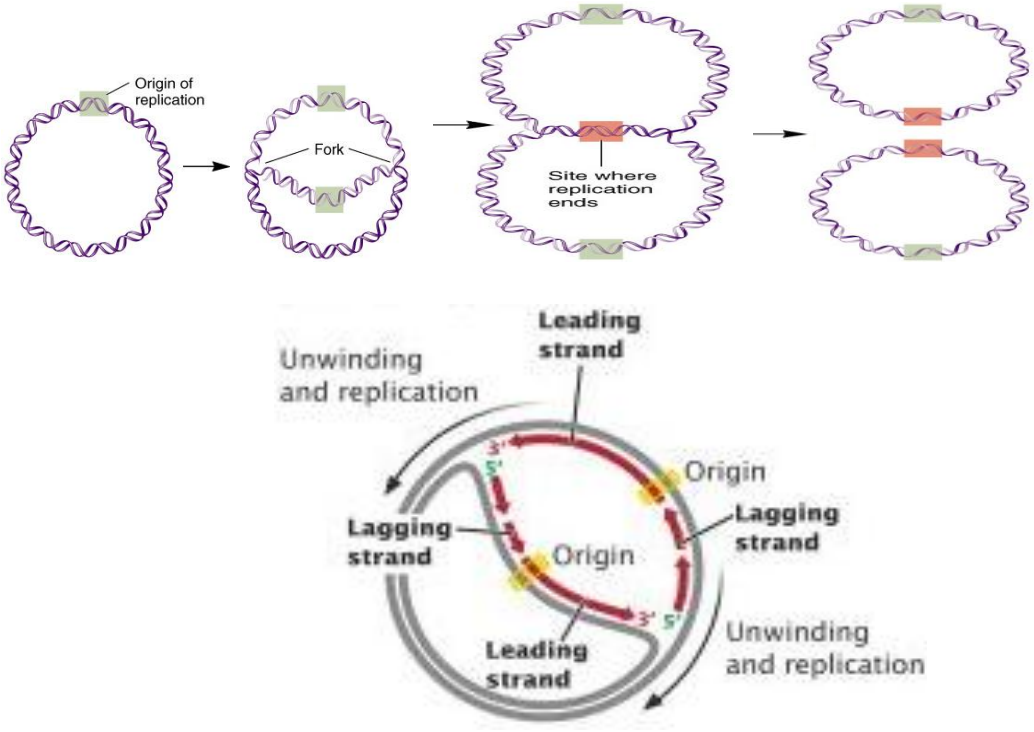


Fig. Theta (θ) Model of DNA Replication

b) Rolling circle model or Sigma (σ) model of DNA replication

Rolling circle replication is the unidirectional mode of DNA replication employed by circular DNA molecules, such as plasmids and the genomes of bacteriophages and some eukaryotic viruses. In viruses with linear genomes, the ability to circularize once inside a cell is a crucial prerequisite for rolling circle replication. By replicating in this fashion, the virus can ensure that no genetic

material is lost from its genome as a consequence of successive rounds of replication. Circularization of linear phage genomes occurs by the interaction between *cos* sites (cohesive sites) in the viral genome.

The process of replication begins with a plasmid or phage-encoded enzyme called *relaxase*, which creates a nick in the circular DNA at a site called the *double-strand origin (DSO)*; the relaxase remains bound to the 5' phosphate at the site of this nick, so the 3' OH group is available as a primer for DNA synthesis by DNA polymerase III. The polymerase moves along the nicked strand, using the un-nicked strand as a template for replication, and a helicase displaces the nicked strand behind polymerase as a single-stranded DNA molecule. This procedure can be repeated multiple times to create numerous linear copies in a continuous head-to-tail series called a *concatemers*.

To make these linear strands double-stranded and circular again, an initiator protein makes another nick to terminate DNA synthesis. DNA polymerase III and RNA polymerase then work in conjunction to replicate the *single-strand origin (SSO)* of a linear strand to make it double-stranded. Finally, DNA polymerase I removes the primer, replacing it with DNA, and DNA ligase covalently binds the strands end-to-end to make the final circular structure.

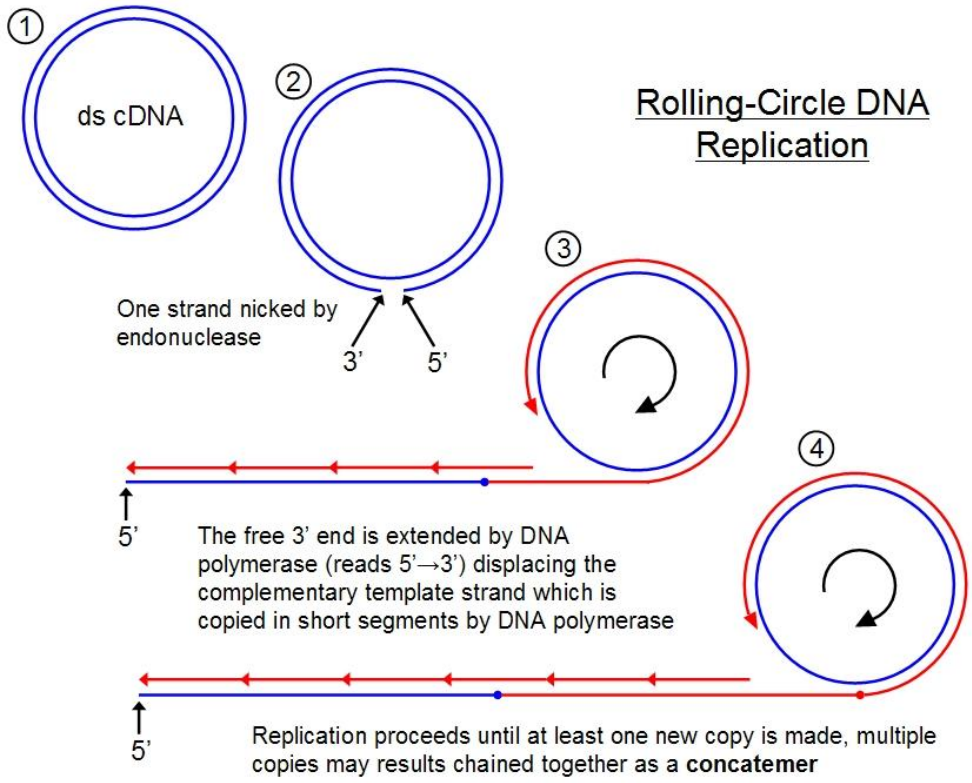


Fig. Rolling circle model OR Sigma (σ) model of DNA replication

==========

- **Enzymology of DNA replication**

Enzymes Involved in DNA Replication

There are several enzymes involved simultaneously to make DNA replication a successful process. Each enzyme has a distinct function and acts accordingly. The enzymes have an active site which recognizes a specific substrate. No enzyme interferes the action of another enzyme. Below is the list of Enzymes involved in DNA replication with their functions.

Enzyme	**Function for leading and lagging strands**	
Helicase	Unwinds parental double helix at replication forks	
Single stranded binding protein	Binds to and stabilizes single stranded DNA until it can't be used as a template	
Topoisomerase	Corrects 'overwinding' ahead of replication forks by breaking, swivelling and rejoining DNA strands	
	Function for leading strand	**Function for lagging strand**
Primase	Synthesizes a single RNA primer at the 5' end of the leading strand	Synthesizes an RNA primer at the 5'end of each Okazaki fragment
DNA pol III	Continuously synthesizes the leading strand adding on to the primer	Elongates each Okazaki fragment adding on to its primer
DNA pol I	Removes primer from the 5'end of leading strand and replaces it with DNA, adding on to the adjacent 3' end	Removes the primer from the 5'end of each fragment and replaces it with DNA, adding on to the 3'end to the adjacent fragment
DNA ligase	Join the 3'end of the DNA that replaces the primer to the rest of the lagging strand	Joins the Okazaki fragment

SINGLE STRAND BINDING (SSB) PROTEINS:

Single strand binding protein named because its nature to bind to single strand of DNA which will be formed during open complex formation during initiation of replication. It contains four subunits. The major function of SSB is to prevent recoiling of DNA strands after it's unwinding by helicases. Thus, SSB plays vital role in replication.

Helicases

These are proteins, which are involved in the unwinding of DNA molecule. There are four kinds of helicases namely Dna A, Dna B, Rep Proteins and DNA Helicase - II.

Dna-A protein (Mw 48,000) - It binds to 4 of 9mer sequence and unwinds a 3 of 13mer sequence at Ori C site and forms an open complex during initiation of replication. It is the first protein which binds to DNA to initiate DNA replication.

Dna-B proteins (Mw- 3, 00,000) - It is a primosome constituent and consists of six subunits. It unwinds DNA during replication. It is responsible for the extension of open complex during replication.

Rep Proteins (Mw 65,000) - It is a helicase consisting of one subunit. It binds to the 5' to 3' template and moves in 3' to 5' direction. It actively participates in leading strand synthesis in replication.

DNA Helicase - II (Mw 75,000) - It is helicase consisting of only one subunit. It binds with 3' to 5' template stand and moves along in 5' to 3' direction. It is involved in lagging strand synthesis in replication.

Helicases are often utilized to separate strands of a DNA double helix or a self-annealed RNA molecule using the energy from ATP hydrolysis, a process characterized by the breaking of hydrogen bonds between annealed nucleotide bases.

Topoisomerases

The normal biological function of DNA occurs only if it is in proper topological state, i.e. to stay in having proper supercoiling or super helical tension. Topoismerases are a group of enzymes which control supercoiling of DNA thereby maintaining it in the proper topological state or superhelical tension. There are two classes of topoisomerases. They are type – I topoisomerase and type – II topoisomerase.

Type - I Topoisomerase:

Type-I topoisomerase (Nicking-Closing enzymes) are monomeric 100kd proteins that are widespread in both prokaryotes and eukaryotes. They can remove negative supercoils without leaving nicks in the DNA molecule.

Mechanism:

After the enzyme binds to a DNA molecule and cuts on strand, the free 5' phosphate on the DNA is covalently attached to a tyrosine residue in the enzyme in the case of prokaryotes (the free 3'-phosphate on the DNA is covalently attached in the case of eukaryotes). The DNA strand that has not been cleaved is then passed through the single stranded break. The cleaved strand is then resealed. By this mechanism, the enzyme removes one negative supercoil at a time there by increasing the LK by one in the case of prokaryotes. By forming free 3' or free 5' end DNA- protein covalent intermediates, the free energy of the cleaved phosphodiester bond is preserved so that no energy input is required to reseal the nick.

Type-II Topoisomerases:

Prokaryotic type-II topoisomerases, which are also known as DNA gyrases, are 375kd proteins that consists of two pairs of subunits designated A and B. DNA gyrase converts a right-handed toroidal supercoil to a left-handed toroidal supercoil. This mechanism is called as sign inversion mechanism. DNA gyrase requires ATP for its function. Hydrolysis of ATP is not required to induce supercoils but is required for the enzyme to turnover and introduces additional supercoils.

Primase and Primosome complex

DNA primase is an RNA Polymerase enzyme involved in the replication of DNA. Primase catalyzes the synthesis of a short RNA segment (called a primer) complementary to a ssDNA template. Primase is of key importance in DNA replication because no known DNA polymerases can initiate the synthesis of a DNA strand without an initial RNA or DNA primer (for temporary DNA elongation).

In DNA replication RNA primer is synthesized by two different enzymes namely RNA polymerase and Primosome complex. RNA polymerase synthesize RNA primer for synthesize of leading strand where as Primosome synthesize RNA primer for lagging strand synthesis. In addition to this, RNA polymerase synthesizes only one primer whereas Primosome synthesizes many primers.

DNA polymerases

A **DNA polymerase** is an enzyme that helps catalyze in the polymerization of deoxyribonucleotides into a DNA strand. DNA polymerases

are best known for their feedback role in DNA replication, in which the polymerase "reads" an intact DNA strand as a template and uses it to synthesize the new strand. This process copies a piece of DNA. The newly polymerized molecule is complementary to the template strand and identical to the template's original partner strand. DNA polymerases use magnesium ions as cofactors. Human DNA polymerases are 900-1000 amino acids long.

DNA polymerase can add free nucleotides to only the 3' end of the newly forming strand. This results in elongation of the new strand in a 5'-3' direction. No known DNA polymerase is able to begin a new chain (de novo). DNA polymerase can add a nucleotide onto only a preexisting 3'-OH group, and, therefore, needs a primer at which it can add the first nucleotide. Primers consist of RNA and/or DNA bases. In DNA replication, the first two bases are always RNA, and are synthesized by another enzyme called primase.

Error correction is a property of some, but not all, DNA polymerases. This process corrects mistakes in newly synthesized DNA. When an incorrect base pair is recognized, DNA polymerase reverses its direction by one base pair of DNA. The 3'-5' exonuclease activity of the enzyme allows the incorrect base pair to be excised (this activity is known as proofreading). Following base excision, the polymerase can re-insert the correct base and replication can continue.

There are 3 DNA polymerases

DNA Polymerase I (or **Pol I**) is an enzyme that participates in the process of DNA replication in prokaryotes. It is composed of 928 amino acids, and is an example of a processive enzyme - it can sequentially catalyze multiple polymerizations

It was initially characterized in *E. coli*, although it is ubiquitous in prokaryotes. In E. coli and many other bacteria, the gene which encodes Pol I is known as **polA**.

DNA Polymerase I possesses three enzymatic activities:

1. A 5' → 3' (forward) DNA polymerase activity, requiring a 3' primer site and a template strand.
2. A 3' → 5' (reverse) exonuclease activity that mediates proofreading. This is the proofreading or editing activity of DNA polymerase, which gets activated when a wrong base pair is incorporated into the chain. DNA pol-I removes deoxy nucleotide at 3'-end of the progeny DNA. Because of this, the activity is referred as 3'-->5' exonuclease activity. Due to this activity, it repairs DNA during replication.

3. A 5'→3' (forward) exonuclease activity mediating nick translation during DNA repair. With the help of this activity, enzyme removes nucleotides from 5'-end. Two functions can be carried out by this activity. They are removal of RNA primer and DNA repair.

DNA polymerase II (also known as **DNA Pol II** or **Pol II**) is a prokaryotic DNA polymerase most likely involved in DNA repair. The enzyme is 90 kDa in size and is coded by the polB gene. DNA Pol II can synthesize DNA new base pairs at an average rate of between 40 and 50 nucleotides/second. Structure and function are not completely elucidated. It has polymerase activity along with proof reading function, but it does not have 5'-->3' exonuclease activity.

DNA polymerase III holoenzyme is the primary enzyme complex involved in prokaryotic DNA replication. The complex has high processivity (i.e. the number of nucleotides added per binding event). Being the primary holoenzyme involved in replication activity, the DNA Pol III holoenzyme also has proofreading capabilities that correct replication mistakes by means of exonuclease activity working 3' to 5'. DNA Pol III is a component of the replisome, which is located at the replication fork.

Holoenzyme of DNA pol-III consists of ten subunits. They are aeqbgdd'cjt . Core enzyme of DNA poly-III contains three subunit aeq. Of three subunits alpha subunit posses polymerizing activity and epsilon posses proof reading activity. These are the two function of the enzyme. It does not have 5'-->3' exonuclease activity.

gdd'c and j subunits called as a gamma complex, which mediates transfer of the beta subunit to the duplex DNA primer strand. DNA pol-III possesses high processivity than other DNA polymerases. This is because of beta subunit. It forms a donut shaped dimers around duplex DNA and then associate with and hold the catalytic core aeq polymerase at the primer template terminus.

Once tightly associated with DNA, the beta subunit dimers functions like a "clamp" which can move freely along the DNA, like a ring on a string carrying the associated core polymerase with it. In this way, the active sites remain near the growing fork and the processivity of the core polymerase maximized.

Processivity: Number of nucleotides added by the enzyme before it gets released from the template strand after its attachment.

DNA ligase

DNA ligase has applications in both DNA repair and DNA replication. DNA Ligase is a protein which functions during the replication process of DNA to ensure that the complementary, double-stranded structure of DNA results. DNA Ligase functions in piecing (joining) together the Okazaki fragments of the lagging strand to form a continuous complementary strand. The reaction it

catalyzes requires the use of ATP. AMP is added to the active site releasing pyrophosphate. DNA then enters the active site and the AMP is used to attach together the 3' and 5' ends of the lagging strand.

The DNA ligases are responsible for connecting DNA segments during replication, repair and recombination. They are class of enzymes that catalyze the formation of alpha-phosphodiester bond between two DNA chains. This enzyme requires the free OH group at the 3' end of other DNA strand and phosphate group at 5' end of the other. The formation of a phosphodiester bond between these groups is an endergonic reaction. Hence energy source required for ligation. In *E. coli* and other bacteria NAD^+ supplies the energy whereas in animals i.e. eukaryotes ATP play the role.

This enzyme seals the nick and joints the Okazaki fragments in the lagging strand. It is also involved in the repair of damaged DNA. DNA ligases are extensively used as a joining tool in genetic engineering.

==========

- **POST-REPLICATIVE MODIFICATION OF DNA**

Postreplication repair is the repair of damage to the DNA that takes place after replication.

Restriction-Modification system occurs in many bacterial species and constitutes a defense mechanism against the introduction of foreign DNA into the cell. Luria Bertani et al observed that the phage that grew well in one bacterial strain often grew poorly in a second time (forming only a few plaques). Phages isolated from these plaques were able to re-infect the second strain and grow well, but lost the ability to grow on the original strain.

These initial observations led to the eventual discovery of restriction enzymes in the early 1950s. They postulated that certain bacterial strains contain an **endonuclease** that is able to cleave DNA, and that some strains contain a strain specific modification system that is responsible for protecting host DNA from the **action of its own endonuclease**.

The unmodified foreign DNA, such as of an infecting phage, is degraded by the endonuclease, restricting phage infection (hence the term restriction endonuclease). However, a small proportion of the phage DNA is modified prior to degradation by the endonuclease. This modified DNA is able to successfully replicate and infect the second host but since that host does not contain the same modification system as the first, the modified phage lost their ability to replicate on the original host.

Restriction-Modification system consists of two components:

1. Restriction endonuclease: This recognizes a short, symmetrical DNA sequence and hydrolyses DNA backbone in each strand at specific sites within that sequence - so that the foreign DNA will be degraded to relatively short fragments. There are different classes of restriction endonucleases namely **type I, type II, and type III restriction endonucleases**.

2. Methylase or Methylation dependant restriction enzymes: -

In addition to the above three types of restriction endonucleases, *E. coli* possess one more type of restriction enzymes system that restricts to cut the DNA, which has been specifically methylated. These are called methylation dependant restriction enzymes as they cut the DNA at their recognization site only if it is methylated.

Modification in some of the bases of DNA is carried out in bacteria with the objective of protection from autorestriction. Modifications involve methylation of certain bases by enzymes called as methylases. Methylase adds a methyl group to Adenine (A) or Cytosine (C) base within the same recognition sequences in the cellular DNA. This modification renders the host DNA resistant to degradation by the endonuclease.

E. coli DNA contains small amounts of 6-methyladenine and 5-methyl cytosine. These bases are generated by the action of three different types of methylases enzyme system on DNA, namely hsd system, dam system and dcm system.

1) Hsd system: -

This enzyme system is coded by hsd genes (host specific DNA and it possesses endonuclease and methylase activity (type I & type III restriction modification enzyme systems). The enzyme acts as a methylase when the recognization site is methylated. When the reorganization site is completely unmodified, the enzyme acts as endonucleases and hydrolyzes DNA. It is highly specific enzyme and it methylates adenine in AAC (N) 6 GTGC and GCAC (N) 6 GTT system. Its methylated DNA is recognized as a self and thus protected from auto restriction. Examples of such enzyme system in *E. coli* are EcoK1 & EcoB1.

2) Dam system (DNA adenine methylase system):-

It is the enzyme system that methylates only N-6 position of the adenine and produce N-6-methyladine in the sequence GATC present at the ori C region the DNA. It is coded by a set of gene called as dam gene.

Dam methylase enzyme system has to important functions-

1. It is involved in the control of DNA replication by regulating the initiation step, and
2. It makes DNA strands for post replication mismatch repair.

a) Role of dam in controlling DNA replication: - Initiation of DNA replication is associated with some changes. Replication initiates only after full methylation of GATC sequence present in the Ori C region. Before replication, palindromic sequence (GATC) is present in fully methylated form. (Adenine on both strands is methylated). During replication, a normal (non-modified) base gets inserted into the daughter strand. This result in the generation of hemimethylated daughter DNAs in which one strand is methylated and other is non-methylated. Before using such daughter hemimethylated DNA for the initiation of replication, it is converted to fully methylated state (both strands are methylated) by Dam system. DNA replicating gets initiated only after full methylation. Thus dam system controls the DNA replication by regulating the initiation of replication in Ori C region of the DNA.

b) Roll of dam in marking DNA strands for post replication mismatch repair: - Dam enzyme system also plays an important role in the post replication mismatch repair system in *E. coli* by marking DNA strands for repair. When the DNA is grossly damaged due to the insertion or creation of wrong base that it is not naturally found in DNA, it is recognized and removed by repair system restoring the wild type sequence. However if the target for repair is a mismatched in partnership of normal base, it is very difficult for the repair system to recognize which is the wild type base and which is the mutant. It recognizes both the bases in improper. Each of the bases of this pair can provide target for excision repair.

If the mutated base is excised, the wild type sequence is restored. But if wild type base is excised, the new (mutant) sequence becomes fixed. When mismatch error occurs during replication in *E. coli*, it is possible to distinguish, the original strand of DNA. Immediate after DNA replication, the parental strand is methylated while the newly synthesized strand is waiting for methylation, hence two strands can be distinguished. Thus, successful post replication repair depends upon the ability of enzyme to distinguish between old and new synthesized DNA strands. The same target sites are used by the replication related repair system. The system attaches to the GATC sequence and excise the base from strand that lack the methylation. The gap is then filled by DNA-polymerase I. Dam methylation is done after sealing the gap.

c) Dcm system (DNA cytosine methylase): -

It is coded by dcm gene and it methylates internal cytosine at the C_5 position on strands within the sequence CCAGG or CCTGG.

==========

- **History of molecular biology and genetic engineering**

The **history of molecular biology** begins in the 1930s with the convergence of various, previously distinct biological disciplines: biochemistry, genetics, microbiology, and virology. With the hope of understanding life at its most fundamental level, numerous physicists and chemists also took an interest in what would become molecular biology.

In its modern sense, molecular biology attempts to explain the phenomena of life starting from the macromolecular properties that generate them. Two categories of macromolecules in particular are the focus of the molecular biologist: 1) nucleic acids, among which the most famous is deoxyribonucleic acid (or DNA), the constituent of genes, and 2) proteins, which are the active agents of living organisms

1) 1910: - Mendelian-chromosome theory of heredity.
2) 1938: - Molecular biology—the name was coined by Warren Weaver.
3) 1940: - George Beadle and Edward Tatum demonstrated the existence of a precise relationship between genes and proteins.
4) 1944: - Avery-MacLeod-McCarty experiment demonstrated that genes are made up of DNA.
5) 1952:- Hershey-Chase confirmed that the genetic material of the bacteriophage is made up of DNA.
6) 1953: - James Watson and Francis Crick discovered the double helical structure of the DNA molecule.
7) 1960: - Monod and Jacob also demonstrated how certain specific proteins, called regulative proteins, control the transcription of these genes into messenger RNA.
8) 1961: - Francois Jacob and Jacques Monod hypothesized the existence of an intermediary between DNA and its protein products, which they called messenger RNA.
9) 1961 to 1965: - The genetic code, discovered.
10) 1965 to 1990: - The chief discoveries of molecular biology took place.
11) 1990 to 2005: - New and more sophisticated technologies, united today under the name of genetic engineering, would permit the isolation and characterization of genes.