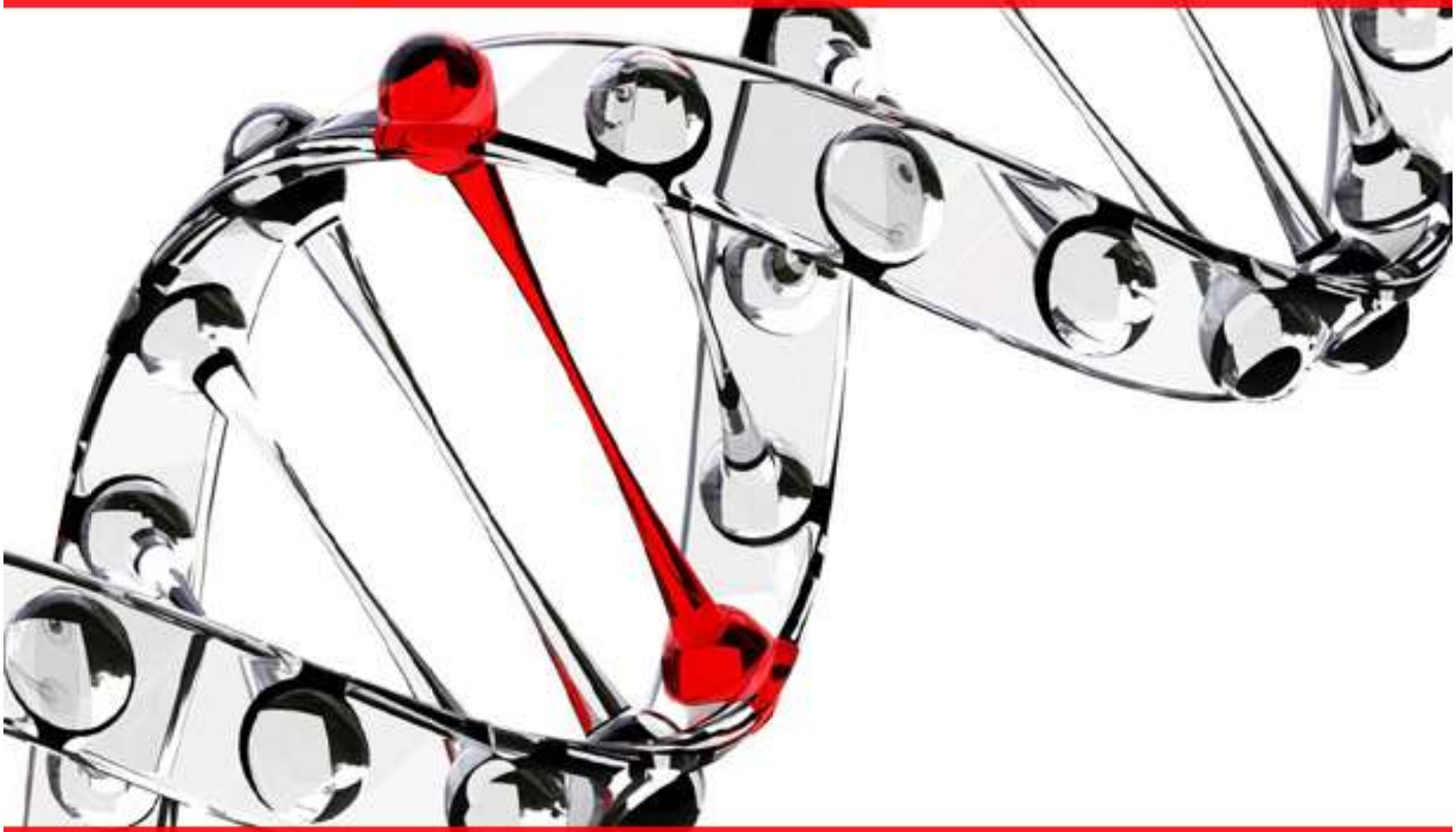
Handbook of DNA Repair and Replication



Jaydon Baez

Cleo Kahn

First Edition, 2012

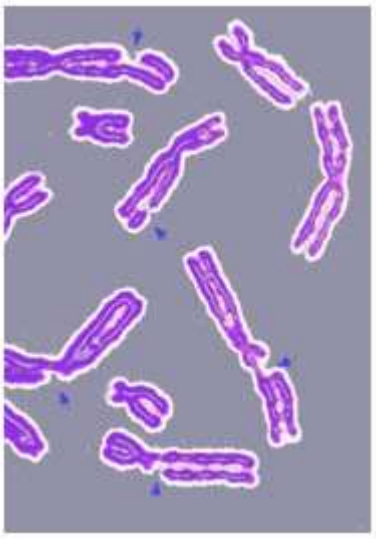
ISBN 978-81-323-0713-6

Published by:
Academic Studio
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter- 1

DNA Repair



DNA damage resulting in multiple broken chromosomes

DNA repair refers to a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome. In human cells, both normal metabolic activities and environmental factors such as UV light and radiation can cause DNA damage, resulting in as many as 1 million individual molecular lesions per cell per day. Many of these lesions cause structural damage to the DNA molecule and can alter or eliminate the cell's ability to transcribe the gene that the affected DNA encodes. Other lesions induce potentially harmful mutations in the cell's genome, which affect the survival of its daughter cells after it undergoes mitosis. As a consequence, the DNA repair process is constantly active as it responds to damage in the DNA structure. When normal repair processes fail, and when cellular apoptosis does not occur, irreparable DNA damage may occur, including double-strand breaks and DNA crosslinkages.

The rate of DNA repair is dependent on many factors, including the cell type, the age of the cell, and the extracellular environment. A cell that has accumulated a large amount of DNA damage, or one that no longer effectively repairs damage incurred to its DNA, can enter one of three possible states:

1. an irreversible state of dormancy, known as senescence

2. cell suicide, also known as apoptosis or programmed cell death
3. unregulated cell division, which can lead to the formation of a tumor that is cancerous

The DNA repair ability of a cell is vital to the integrity of its genome and thus to its normal functioning and that of the organism. Many genes that were initially shown to influence life span have turned out to be involved in DNA damage repair and protection. Failure to correct molecular lesions in cells that form gametes can introduce mutations into the genomes of the offspring and thus influence the rate of evolution.

DNA damage

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 1,000 to 1,000,000 molecular lesions per cell per day. While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs), unrepaired lesions in critical genes (such as tumor suppressor genes) can impede a cell's ability to carry out its function and appreciably increase the likelihood of tumor formation.

The vast majority of DNA damage affects the primary structure of the double helix; that is, the bases themselves are chemically modified. These modifications can in turn disrupt the molecules' regular helical structure by introducing non-native chemical bonds or bulky adducts that do not fit in the standard double helix. Unlike proteins and RNA, DNA usually lacks tertiary structure and therefore damage or disturbance does not occur at that level. DNA is, however, supercoiled and wound around "packaging" proteins called histones (in eukaryotes), and both superstructures are vulnerable to the effects of DNA damage.

Sources of damage

DNA damage can be subdivided into two main types:

1. endogenous damage such as attack by reactive oxygen species produced from normal metabolic byproducts (spontaneous mutation), especially the process of oxidative deamination
    1. also includes replication errors
2. exogenous damage caused by external agents such as
    1. ultraviolet [UV 200-300nm] radiation from the sun
    2. other radiation frequencies, including x-rays and gamma rays
    3. hydrolysis or thermal disruption
    4. certain plant toxins
    5. human-made mutagenic chemicals, especially aromatic compounds that act as DNA intercalating agents
    6. cancer chemotherapy and radiotherapy
    7. viruses

The replication of damaged DNA before cell division can lead to the incorporation of wrong bases opposite damaged ones. Daughter cells that inherit these wrong bases carry mutations from which the original DNA sequence is unrecoverable (except in the rare case of a back mutation, for example, through gene conversion).

Types of damage

There are five main types of damage to DNA due to endogenous cellular processes:

1. *oxidation* of bases [e.g. 8-oxo-7,8-dihydroguanine (8-oxoG)] and generation of DNA strand interruptions from reactive oxygen species,
2. *alkylation* of bases (usually methylation), such as formation of 7-methylguanine, 1-methyladenine, 6-O-Methylguanine
3. *hydrolysis* of bases, such as deamination, depurination, and depyrimidination.
4. "bulky adduct formation" (i.e., benzo[a]pyrene diol epoxide-dG adduct)
5. *mismatch* of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.

Damage caused by exogenous agents comes in many forms. Some examples are:

1. *UV-B light* causes crosslinking between adjacent cytosine and thymine bases creating *pyrimidine dimers*. This is called direct DNA damage.
2. *UV-A light* creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.
3. *Ionizing radiation* such as that created by radioactive decay or in *cosmic rays* causes breaks in DNA strands. Low-level ionizing radiation may induce irreparable DNA damage (leading to replicational and transcriptional errors needed for neoplasia or may trigger viral interactions) leading to pre-mature aging and cancer.
4. *Thermal disruption* at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single-strand breaks. For example, hydrolytic depurination is seen in the thermophilic bacteria, which grow in hot springs at 40-80 °C. The rate of depurination (300 purine residues per genome per generation) is too high in these species to be repaired by normal repair machinery, hence a possibility of an adaptive response cannot be ruled out.
5. *Industrial chemicals* such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic hydrocarbons found in smoke, soot and tar create a huge diversity of DNA adducts- ethenobases, oxidized bases, alkylated phosphotriesters and Crosslinking of DNA just to name a few.

UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage. Spontaneous damage can include the loss of a base, deamination, sugar ring puckering and tautomeric shift.

Nuclear versus mitochondrial DNA damage

In human cells, and eukaryotic cells in general, DNA is found in two cellular locations - inside the nucleus and inside the mitochondria. Nuclear DNA (nDNA) exists as chromatin during non-replicative stages of the cell cycle and is condensed into aggregate structures known as chromosomes during cell division. In either state the DNA is highly compacted and wound up around bead-like proteins called histones. Whenever a cell needs to express the genetic information encoded in its nDNA the required chromosomal region is unravelled, genes located therein are expressed, and then the region is condensed back to its resting conformation. Mitochondrial DNA (mtDNA) is located inside mitochondria organelles, exists in multiple copies, and is also tightly associated with a number of proteins to form a complex known as the nucleoid. Inside mitochondria, reactive oxygen species (ROS), or free radicals, byproducts of the constant production of adenosine triphosphate (ATP) via oxidative phosphorylation, create a highly oxidative environment that is known to damage mtDNA. A critical enzyme in counteracting the toxicity of these species is superoxide dismutase, which is present in both the mitochondria and cytoplasm of eukaryotic cells.

Senescence and apoptosis

Senescence, an irreversible state in which the cell no longer divides, is a protective response to the shortening of the chromosome ends. The telomeres are long regions of repetitive noncoding DNA that cap chromosomes and undergo partial degradation each time a cell undergoes division. In contrast, quiescence is a reversible state of cellular dormancy that is unrelated to genome damage. Senescence in cells may serve as a functional alternative to apoptosis in cases where the physical presence of a cell for spatial reasons is required by the organism, which serves as a "last resort" mechanism to prevent a cell with damaged DNA from replicating inappropriately in the absence of pro-growth cellular signaling. Unregulated cell division can lead to the formation of a tumor, which is potentially lethal to an organism. Therefore, the induction of senescence and apoptosis is considered to be part of a strategy of protection against cancer.

DNA damage and mutation

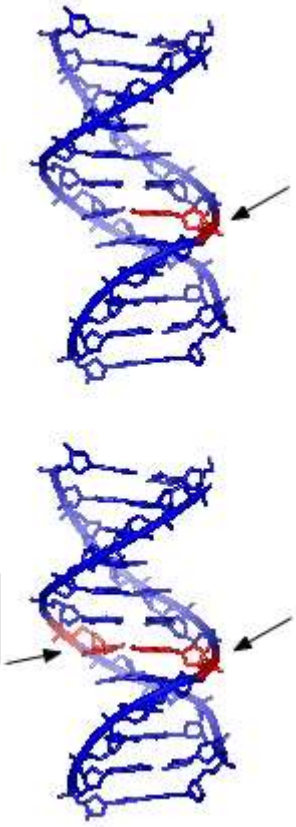
It is important to distinguish between DNA damage and mutation, the two major types of error in DNA. DNA damages and mutation are fundamentally different. Damages are physical abnormalities in the DNA, such as single- and double-strand breaks, 8-hydroxydeoxyguanosine residues, and polycyclic aromatic hydrocarbon adducts. DNA damages can be recognized by enzymes, and, thus, they can be correctly repaired if redundant information, such as the undamaged sequence in the complementary DNA strand or in a homologous chromosome, is available for copying. If a cell retains DNA damage, transcription of a gene can be prevented, and, thus, translation into a protein will also be blocked. Replication may also be blocked and/or the cell may die.

In contrast to DNA damage, a mutation is a change in the base sequence of the DNA. A mutation cannot be recognized by enzymes once the base change is present in both DNA

strands, and, thus, a mutation cannot be repaired. At the cellular level, mutations can cause alterations in protein function and regulation. Mutations are replicated when the cell replicates. In a population of cells, mutant cells will increase or decrease in frequency according to the effects of the mutation on the ability of the cell to survive and reproduce. Although distinctly different from each other, DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair; these errors are a major source of mutation.

Given these properties of DNA damage and mutation, it can be seen that DNA damages are a special problem in non-dividing or slowly dividing cells, where unrepaired damages will tend to accumulate over time. On the other hand, in rapidly dividing cells, unrepaired DNA damages that do not kill the cell by blocking replication will tend to cause replication errors and thus mutation. The great majority of mutations that are not neutral in their effect are deleterious to a cell's survival. Thus, in a population of cells comprising a tissue with replicating cells, mutant cells will tend to be lost. However, infrequent mutations that provide a survival advantage will tend to clonally expand at the expense of neighboring cells in the tissue. This advantage to the cell is disadvantageous to the whole organism, because such mutant cells can give rise to cancer. Thus, DNA damages in frequently dividing cells, because they give rise to mutations, are a prominent cause of cancer. In contrast, DNA damages in infrequently dividing cells are likely a prominent cause of aging.

DNA repair mechanisms



Single-strand and double-strand DNA damage

Cells cannot function if DNA damage corrupts the integrity and accessibility of essential information in the genome (but cells remain superficially functional when so-called "non-essential" genes are missing or damaged). Depending on the type of damage inflicted on the DNA's double helical structure, a variety of repair strategies have evolved to restore lost information. If possible, cells use the unmodified complementary strand of the DNA or the sister chromatid as a template to recover the original information. Without access to a template, cells use an error-prone recovery mechanism known as translesion synthesis as a last resort.

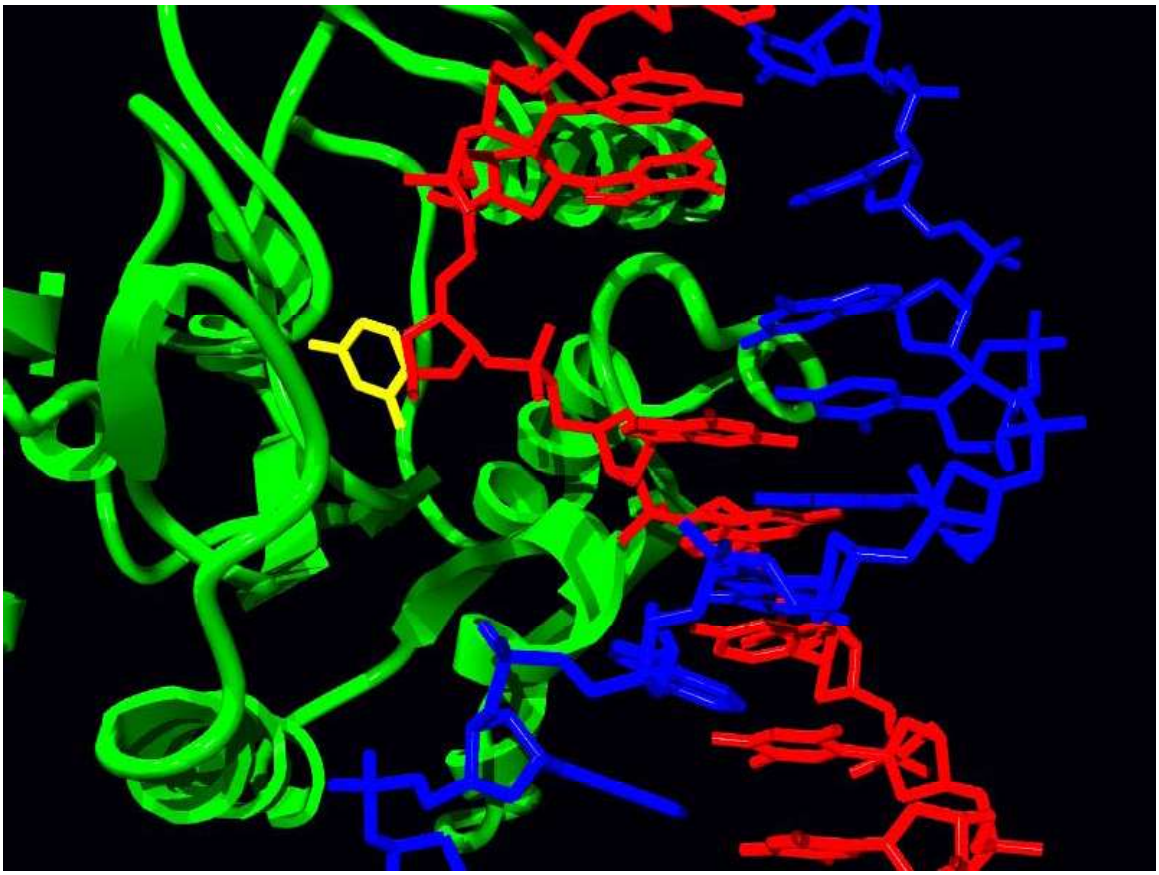
Damage to DNA alters the spatial configuration of the helix, and such alterations can be detected by the cell. Once damage is localized, specific DNA repair molecules bind at or near the site of damage, inducing other molecules to bind and form a complex that enables the actual repair to take place.

Direct reversal

Cells are known to eliminate three types of damage to their DNA by chemically reversing it. These mechanisms do not require a template, since the types of damage they

counteract can occur in only one of the four bases. Such direct reversal mechanisms are specific to the type of damage incurred and do not involve breakage of the phosphodiester backbone. The formation of pyrimidine dimers upon irradiation with UV light results in an abnormal covalent bond between adjacent pyrimidine bases. The photoreactivation process directly reverses this damage by the action of the enzyme photolyase, whose activation is obligately dependent on energy absorbed from blue/UV light (300–500 nm wavelength) to promote catalysis. Another type of damage, methylation of guanine bases, is directly reversed by the protein methyl guanine methyl transferase (MGMT), the bacterial equivalent of which is called ogt. This is an expensive process because each MGMT molecule can be used only once; that is, the reaction is stoichiometric rather than catalytic. A generalized response to methylating agents in bacteria is known as the adaptive response and confers a level of resistance to alkylating agents upon sustained exposure by upregulation of alkylation repair enzymes. The third type of DNA damage reversed by cells is certain methylation of the bases cytosine and adenine.

Single-strand damage



Structure of the base-excision repair enzyme uracil-DNA glycosylase. The uracil residue is shown in yellow.

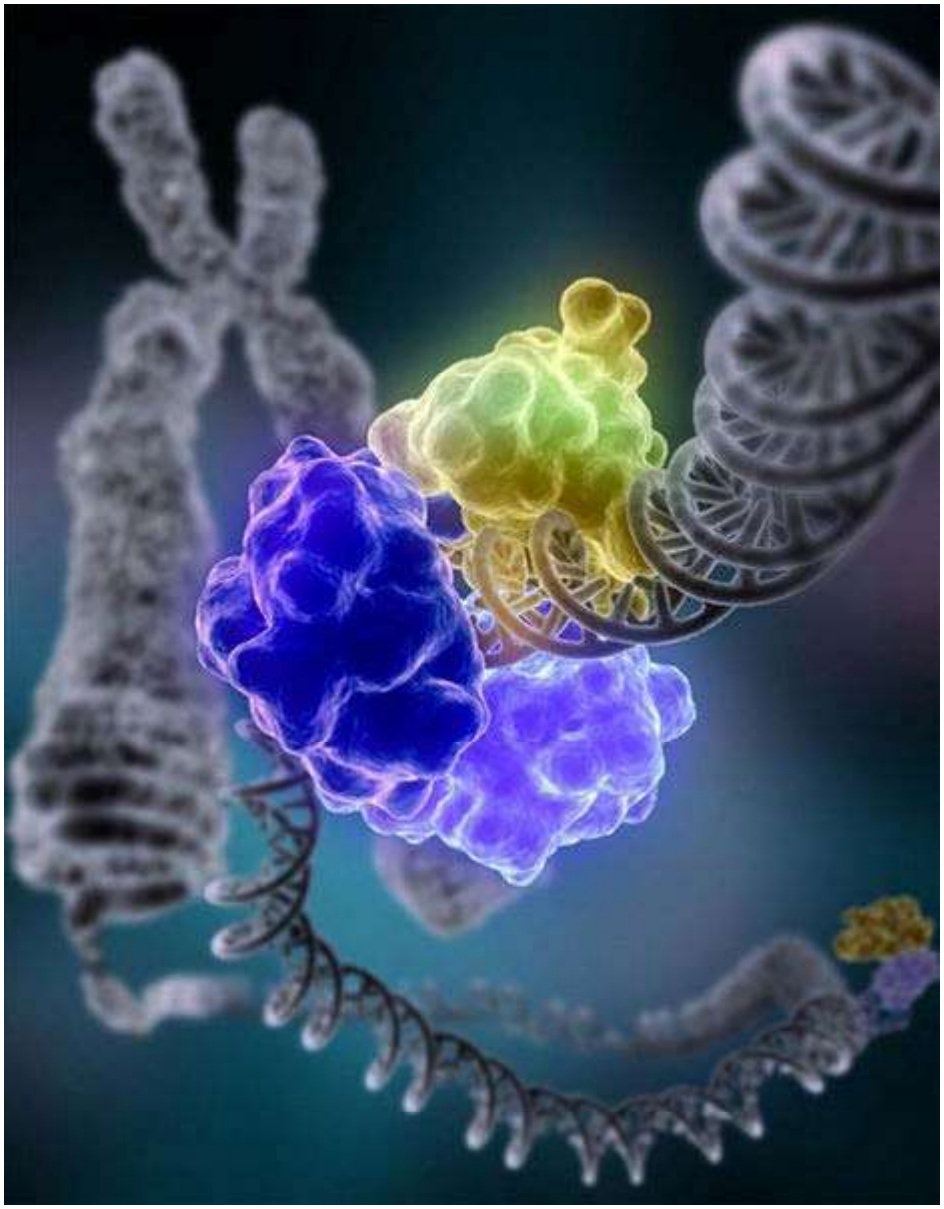
When only one of the two strands of a double helix has a defect, the other strand can be used as a template to guide the correction of the damaged strand. In order to repair

damage to one of the two paired molecules of DNA, there exist a number of excision repair mechanisms that remove the damaged nucleotide and replace it with an undamaged nucleotide complementary to that found in the undamaged DNA strand.

1. Base excision repair (BER), which repairs damage to a single base caused by oxidation, alkylation, hydrolysis, or deamination. The damaged base is removed by a DNA glycosylase. The "missing tooth" is then recognised by an enzyme called AP endonuclease, which cuts the Phosphodiester bond. The missing part is then resynthesized by a DNA polymerase, and a DNA ligase performs the final nick-sealing step.
2. Nucleotide excision repair (NER), which recognizes bulky, helix-distorting lesions such as pyrimidine dimers and 6,4 photoproducts. A specialized form of NER known as transcription-coupled repair deploys NER enzymes to genes that are being actively transcribed.
3. Mismatch repair (MMR), which corrects errors of DNA replication and recombination that result in mispaired (but undamaged) nucleotides.

Double-strand breaks

Double-strand breaks, in which both strands in the double helix are severed, are particularly hazardous to the cell because they can lead to genome rearrangements. Three mechanisms exist to repair double-strand breaks (DSBs): non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and homologous recombination. PVN Acharya noted that double-strand breaks and a "cross-linkage joining both strands at the same point is irreparable because neither strand can then serve as a template for repair. The cell will die in the next mitosis or in some rare instances, mutate."



DNA ligase, shown above repairing chromosomal damage, is an enzyme that joins broken nucleotides together by catalyzing the formation of an internucleotide ester bond between the phosphate backbone and the deoxyribose nucleotides.

In NHEJ, DNA Ligase IV, a specialized DNA ligase that forms a complex with the cofactor XRCC4, directly joins the two ends. To guide accurate repair, NHEJ relies on short homologous sequences called microhomologies present on the single-stranded tails of the DNA ends to be joined. If these overhangs are compatible, repair is usually accurate. NHEJ can also introduce mutations during repair. Loss of damaged nucleotides at the break site can lead to deletions, and joining of nonmatching termini forms translocations. NHEJ is especially important before the cell has replicated its DNA, since there is no template available for repair by homologous recombination. There are "backup" NHEJ pathways in higher eukaryotes. Besides its role as a genome caretaker,

NHEJ is required for joining hairpin-capped double-strand breaks induced during V(D)J recombination, the process that generates diversity in B-cell and T-cell receptors in the vertebrate immune system.

Homologous recombination requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is nearly identical to the machinery responsible for chromosomal crossover during meiosis. This pathway allows a damaged chromosome to be repaired using a sister chromatid (available in G2 after DNA replication) or a homologous chromosome as a template. DSBs caused by the replication machinery attempting to synthesize across a single-strand break or unrepaired lesion cause collapse of the replication fork and are typically repaired by recombination.

Topoisomerases introduce both single- and double-strand breaks in the course of changing the DNA's state of supercoiling, which is especially common in regions near an open replication fork. Such breaks are not considered DNA damage because they are a natural intermediate in the topoisomerase biochemical mechanism and are immediately repaired by the enzymes that created them.

A team of French researchers bombarded *Deinococcus radiodurans* to study the mechanism of double-strand break DNA repair in that organism. At least two copies of the genome, with random DNA breaks, can form DNA fragments through annealing. Partially overlapping fragments are then used for synthesis of homologous regions through a moving D-loop that can continue extension until they find complementary partner strands. In the final step there is crossover by means of RecA-dependent homologous recombination.

Translesion synthesis

Translesion synthesis is a DNA damage tolerance process that allows the DNA replication machinery to replicate past DNA lesions such as thymine dimers or AP sites. It involves switching out regular DNA polymerases for specialized translesion polymerases (e.g. DNA polymerase V), often with larger active sites that can facilitate the insertion of bases opposite damaged nucleotides. The polymerase switching is thought to be mediated by, among other factors, the post-translational modification of the replication processivity factor PCNA. Translesion synthesis polymerases often have low fidelity (high propensity to insert wrong bases) relative to regular polymerases. However, many are extremely efficient at inserting correct bases opposite specific types of damage. For example, Pol η mediates error-free bypass of lesions induced by UV irradiation, whereas Pol ζ introduces mutations at these sites. From a cellular perspective, risking the introduction of point mutations during translesion synthesis may be preferable to resorting to more drastic mechanisms of DNA repair, which may cause gross chromosomal aberrations or cell death. In short, the process involves specialized polymerases either bypassing or repairing lesions at locations of stalled DNA replication. A bypass platform is provided to these polymerases by Proliferating cell nuclear antigen (PCNA). Under normal circumstances, PCNA bound to polymerases replicates the DNA.

At a site of lesion, PCNA is ubiquitinated, or modified, by the RAD6/RAD18 proteins to provide a platform for the specialized polymerases to bypass the lesion and resume DNA replication.

Global response to DNA damage

Cells exposed to ionizing radiation, ultraviolet light or chemicals are prone to acquire multiple sites of bulky DNA lesions and double-strand breaks. Moreover, DNA damaging agents can damage other biomolecules such as proteins, carbohydrates, lipids, and RNA. The accumulation of damage, to be specific, double-strand breaks or adducts stalling the replication forks, are among known stimulation signals for a global response to DNA damage. The global response to damage is an act directed toward the cells' own preservation and triggers multiple pathways of macromolecular repair, lesion bypass, tolerance, or apoptosis. The common features of global response are induction of multiple genes, cell cycle arrest, and inhibition of cell division.

DNA damage checkpoints

After DNA damage, cell cycle checkpoints are activated. Checkpoint activation pauses the cell cycle and gives the cell time to repair the damage before continuing to divide. DNA damage checkpoints occur at the G1/S and G2/M boundaries. An intra-S checkpoint also exists. Checkpoint activation is controlled by two master kinases, ATM and ATR. ATM responds to DNA double-strand breaks and disruptions in chromatin structure, whereas ATR primarily responds to stalled replication forks. These kinases phosphorylate downstream targets in a signal transduction cascade, eventually leading to cell cycle arrest. A class of checkpoint mediator proteins including BRCA1, MDC1, and 53BP1 has also been identified. These proteins seem to be required for transmitting the checkpoint activation signal to downstream proteins.

P53 is an important downstream target of ATM and ATR, as it is required for inducing apoptosis following DNA damage. At the G1/S checkpoint, p53 functions by deactivating the CDK2/cyclin E complex. Similarly, p21 mediates the G2/M checkpoint by deactivating the CDK1/cyclin B complex.

The prokaryotic SOS response

The SOS response is the term used to describe changes in gene expression in *Escherichia coli* and other bacteria in response to extensive DNA damage. The prokaryotic SOS system is regulated by two key proteins: LexA and RecA. The LexA homodimer is a transcriptional repressor that binds to operator sequences commonly referred to as SOS boxes. In *Escherichia coli* it is known that LexA regulates transcription of approximately 48 genes including the lexA and recA genes. The SOS response is known to be widespread in the Bacteria domain, but it is mostly absent in some bacterial phyla, like the Spirochetes. The most common cellular signals activating the SOS response are regions of single-stranded DNA (ssDNA), arising from stalled replication forks or double-strand breaks, which are processed by DNA helicase to separate the two DNA

strands. In the initiation step, RecA protein binds to ssDNA in an ATP hydrolysis driven reaction creating RecA–ssDNA filaments. RecA–ssDNA filaments activate LexA autoprotease activity, which ultimately leads to cleavage of LexA dimer and subsequent LexA degradation. The loss of LexA repressor induces transcription of the SOS genes and allows for further signal induction, inhibition of cell division and an increase in levels of proteins responsible for damage processing.

In *Escherichia coli*, SOS boxes are 20-nucleotide long sequences near promoters with palindromic structure and a high degree of sequence conservation. In other classes and phyla, the sequence of SOS boxes varies considerably, with different length and composition, but it is always highly conserved and one of the strongest short signals in the genome. The high information content of SOS boxes permits differential binding of LexA to different promoters and allows for timing of the SOS response. The lesion repair genes are induced at the beginning of SOS response. The error-prone translesion polymerases, for example, UmuCD'2 (also called DNA polymerase V), are induced later on as a last resort. Once the DNA damage is repaired or bypassed using polymerases or through recombination, the amount of single-stranded DNA in cells is decreased, lowering the amounts of RecA filaments decreases cleavage activity of LexA homodimer, which then binds to the SOS boxes near promoters and restores normal gene expression.

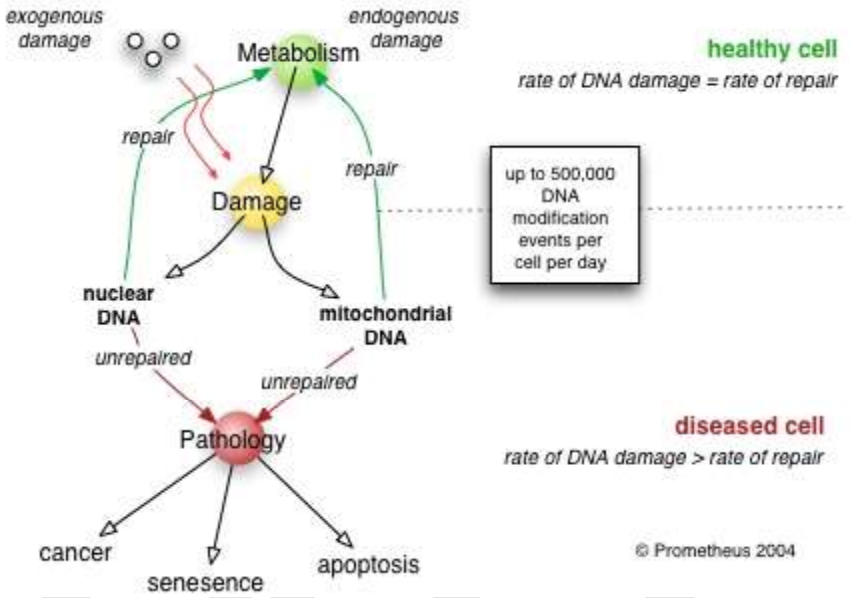
Eukaryotic transcriptional responses to DNA damage

Eukaryotic cells exposed to DNA damaging agents also activate important defensive pathways by inducing multiple proteins involved in DNA repair, cell cycle checkpoint control, protein trafficking and degradation. Such genome wide transcriptional response is very complex and tightly regulated, thus allowing coordinated global response to damage. Exposure of yeast *Saccharomyces cerevisiae* to DNA damaging agents results in overlapping but distinct transcriptional profiles. Similarities to environmental shock response indicates that a general global stress response pathway exist at the level of transcriptional activation. In contrast, different human cell types respond to damage differently indicating an absence of a common global response. The probable explanation for this difference between yeast and human cells may be in the heterogeneity of mammalian cells. In an animal different types of cells are distributed among different organs that have evolved different sensitivities to DNA damage.

In general global response to DNA damage involves expression of multiple genes responsible for postreplication repair, homologous recombination, nucleotide excision repair, DNA damage checkpoint, global transcriptional activation, genes controlling mRNA decay, and many others. A large amount of damage to a cell leaves it with an important decision: undergo apoptosis and die, or survive at the cost of living with a modified genome. An increase in tolerance to damage can lead to an increased rate of survival that will allow a greater accumulation of mutations. Yeast Rev1 and human polymerase η are members of Y family translesion DNA polymerases present during global response to DNA damage and are responsible for enhanced mutagenesis during a global response to DNA damage in eukaryotes.

DNA repair and aging

Pathological effects of poor DNA repair

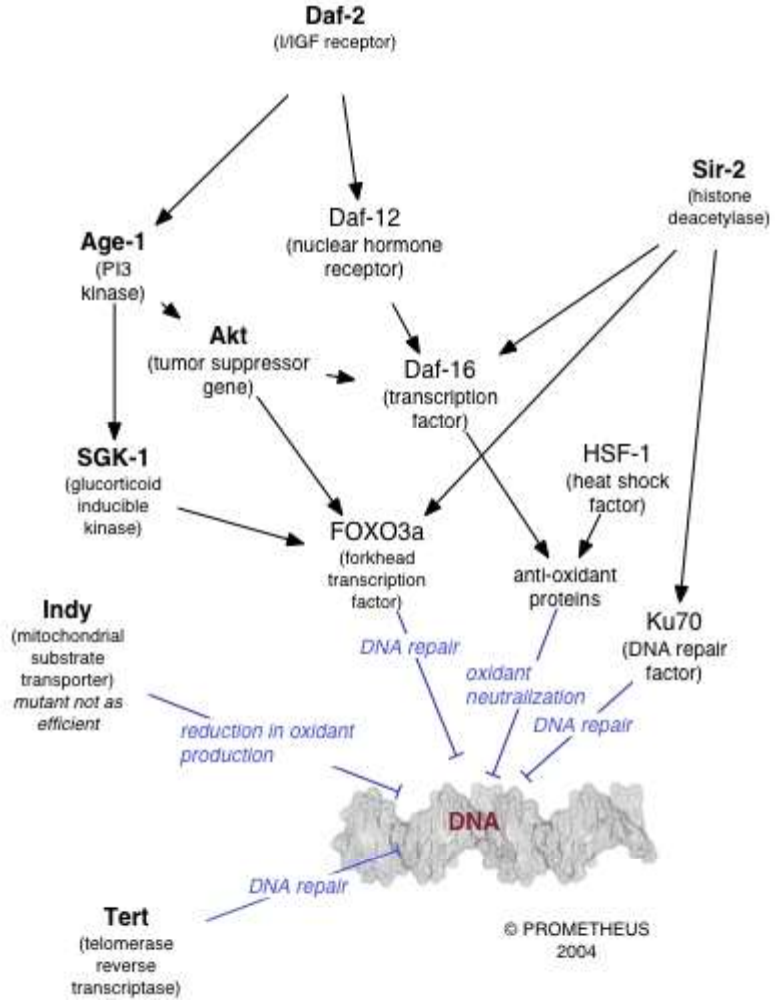


DNA repair rate is an important determinant of cell pathology

Experimental animals with genetic deficiencies in DNA repair often show decreased life span and increased cancer incidence. For example, mice deficient in the dominant NHEJ pathway and in telomere maintenance mechanisms get lymphoma and infections more often, and, as a consequence, have shorter lifespans than wild-type mice. In similar manner, mice deficient in a key repair and transcription protein that unwinds DNA helices have premature onset of aging-related diseases and consequent shortening of lifespan. However, not every DNA repair deficiency creates exactly the predicted effects; mice deficient in the NER pathway exhibited shortened life span without correspondingly higher rates of mutation.

If the rate of DNA damage exceeds the capacity of the cell to repair it, the accumulation of errors can overwhelm the cell and result in early senescence, apoptosis, or cancer. Inherited diseases associated with faulty DNA repair functioning result in premature aging, increased sensitivity to carcinogens, and correspondingly increased cancer risk (see below). On the other hand, organisms with enhanced DNA repair systems, such as *Deinococcus radiodurans*, the most radiation-resistant known organism, exhibit remarkable resistance to the double-strand break-inducing effects of radioactivity, likely due to enhanced efficiency of DNA repair and especially NHEJ.

Longevity and caloric restriction



Most life span influencing genes affect the rate of DNA damage

A number of individual genes have been identified as influencing variations in life span within a population of organisms. The effects of these genes is strongly dependent on the environment, in particular, on the organism's diet. Caloric restriction reproducibly results in extended lifespan in a variety of organisms, likely via nutrient sensing pathways and decreased metabolic rate. The molecular mechanisms by which such restriction results in lengthened lifespan are as yet unclear; however, the behavior of many genes known to be involved in DNA repair is altered under conditions of caloric restriction.

For example, increasing the gene dosage of the gene SIR-2, which regulates DNA packaging in the nematode worm *Caenorhabditis elegans*, can significantly extend lifespan. The mammalian homolog of SIR-2 is known to induce downstream DNA repair factors involved in NHEJ, an activity that is especially promoted under conditions of caloric restriction. Caloric restriction has been closely linked to the rate of base excision repair in the nuclear DNA of rodents, although similar effects have not been observed in mitochondrial DNA.

It is interesting to note that the *C. elegans* gene AGE-1, an upstream effector of DNA repair pathways, confers dramatically extended life span under free-feeding conditions but leads to a decrease in reproductive fitness under conditions of caloric restriction. This observation supports the pleiotropy theory of the biological origins of aging, which suggests that genes conferring a large survival advantage early in life will be selected for even if they carry a corresponding disadvantage late in life.

Medicine and DNA repair modulation

Hereditary DNA repair disorders

Defects in the NER mechanism are responsible for several genetic disorders, including:

- xeroderma pigmentosum: hypersensitivity to sunlight/UV, resulting in increased skin cancer incidence and premature aging
- Cockayne syndrome: hypersensitivity to UV and chemical agents
- trichothiodystrophy: sensitive skin, brittle hair and nails

Mental retardation often accompanies the latter two disorders, suggesting increased vulnerability of developmental neurons.

Other DNA repair disorders include:

- Werner's syndrome: premature aging and retarded growth
- Bloom's syndrome: sunlight hypersensitivity, high incidence of malignancies (especially leukemias).
- ataxia telangiectasia: sensitivity to ionizing radiation and some chemical agents

All of the above diseases are often called "segmental progerias" ("accelerated aging diseases") because their victims appear elderly and suffer from aging-related diseases at an abnormally young age, while not manifesting all the symptoms of old age.

Other diseases associated with reduced DNA repair function include Fanconi's anemia, hereditary breast cancer and hereditary colon cancer.

DNA repair and cancer

Inherited mutations that affect DNA repair genes are strongly associated with high cancer risks in humans. Hereditary nonpolyposis colorectal cancer (HNPCC) is strongly associated with specific mutations in the DNA mismatch repair pathway. BRCA1 and BRCA2, two famous mutations conferring a hugely increased risk of breast cancer on carriers, are both associated with a large number of DNA repair pathways, especially NHEJ and homologous recombination.

Cancer therapy procedures such as chemotherapy and radiotherapy work by overwhelming the capacity of the cell to repair DNA damage, resulting in cell death.

Cells that are most rapidly dividing - most typically cancer cells - are preferentially affected. The side-effect is that other non-cancerous but rapidly dividing cells such as stem cells in the bone marrow are also affected. Modern cancer treatments attempt to localize the DNA damage to cells and tissues only associated with cancer, either by physical means (concentrating the therapeutic agent in the region of the tumor) or by biochemical means (exploiting a feature unique to cancer cells in the body).

DNA repair and evolution

The basic processes of DNA repair are highly conserved among both prokaryotes and eukaryotes and even among bacteriophage (viruses that infect bacteria); however, more complex organisms with more complex genomes have correspondingly more complex repair mechanisms. The ability of a large number of protein structural motifs to catalyze relevant chemical reactions has played a significant role in the elaboration of repair mechanisms during evolution. For an extremely detailed review of hypotheses relating to the evolution of DNA repair, see.

The fossil record indicates that single-cell life began to proliferate on the planet at some point during the Precambrian period, although exactly when recognizably modern life first emerged is unclear. Nucleic acids became the sole and universal means of encoding genetic information, requiring DNA repair mechanisms that in their basic form have been inherited by all extant life forms from their common ancestor. The emergence of Earth's oxygen-rich atmosphere (known as the "oxygen catastrophe") due to photosynthetic organisms, as well as the presence of potentially damaging free radicals in the cell due to oxidative phosphorylation, necessitated the evolution of DNA repair mechanisms that act specifically to counter the types of damage induced by oxidative stress.

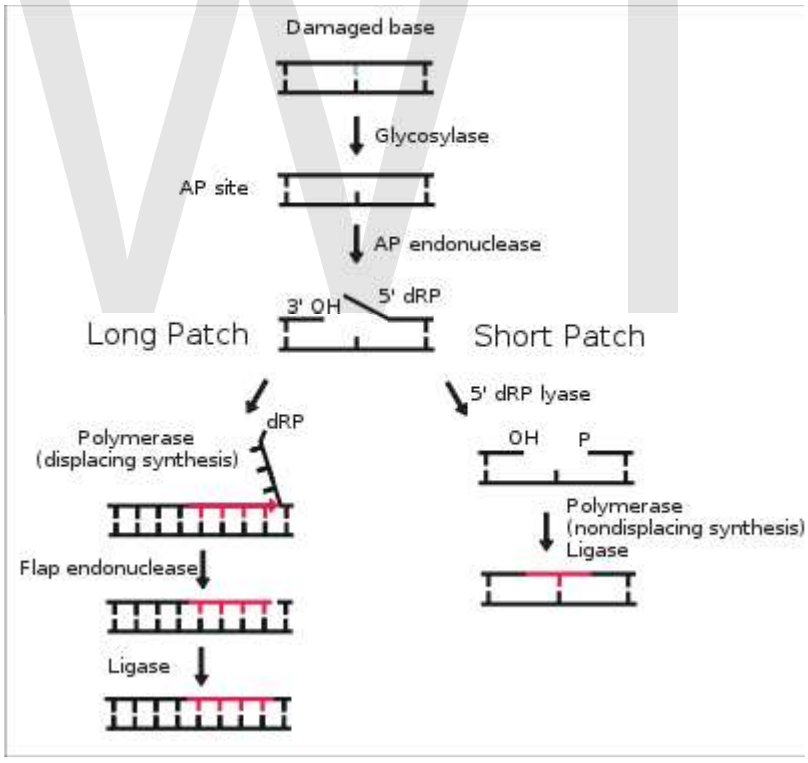
Rate of evolutionary change

On some occasions, DNA damage is not repaired, or is repaired by an error-prone mechanism that results in a change from the original sequence. When this occurs, mutations may propagate into the genomes of the cell's progeny. Should such an event occur in a germ line cell that will eventually produce a gamete, the mutation has the potential to be passed on to the organism's offspring. The rate of evolution in a particular species (or, in a particular gene) is a function of the rate of mutation. As a consequence, the rate and accuracy of DNA repair mechanisms have an influence over the process of evolutionary change.

Chapter- 2

Base Excision Repair and Nucleotide Excision Repair

Base excision repair

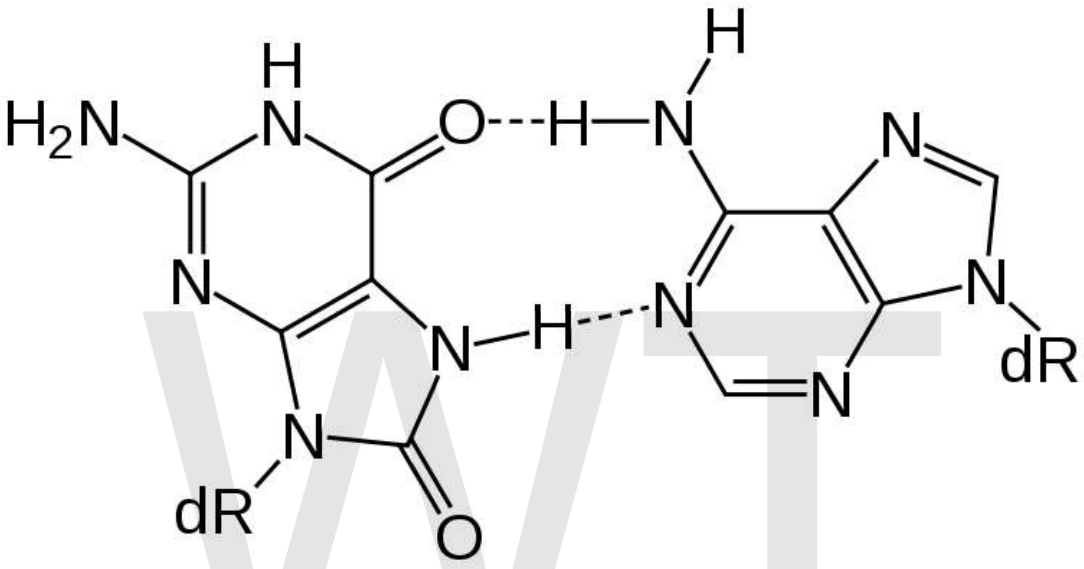


Basic steps of base excision repair

Base excision repair (BER) is a cellular mechanism that repairs damaged DNA throughout the cell cycle. It is responsible primarily for removing small, non-helix-distorting base lesions from the genome. The related nucleotide excision repair pathway repairs bulky helix-distorting lesions. BER is important for removing damaged bases that

could otherwise cause mutations by mispairing or lead to breaks in DNA during replication. BER is initiated by DNA glycosylases, which recognize and remove specific damaged or inappropriate bases, forming AP sites. These are then cleaved by an AP endonuclease. The resulting single-strand break can then be processed by either short-patch (where a single nucleotide is replaced) or long-patch BER (where 2-10 new nucleotides are synthesized).

Lesions processed by BER



8-oxoguanine forms a Hoogsteen base pair with adenine

Single bases in DNA can be chemically damaged by a variety of mechanisms, the most common ones being deamination, oxidation, and alkylation. These modifications can affect the ability of the base to hydrogen-bond, resulting in incorrect base-pairing, and, as a consequence, mutations in the DNA. For example, incorporation of adenine across from 8-oxoguanine (right) during DNA replication causes a G:C base pair to be mutated to T:A. Other examples of base lesions repaired by BER include:

- **Oxidized bases**: 8-oxoguanine, 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG, FapyA)
- **Alkylated bases**: 3-methyladenine, 7-methylguanine
- **Deaminated bases**: hypoxanthine formed from deamination of adenine. Xanthine formed from deamination of guanine. (Thymidine products following deamination of 5-methylcytosine are not repaired)
- **Uracil** inappropriately incorporated in DNA or formed by deamination of cytosine

In addition to base lesions, the downstream steps of BER are also utilized to repair single-strand breaks.

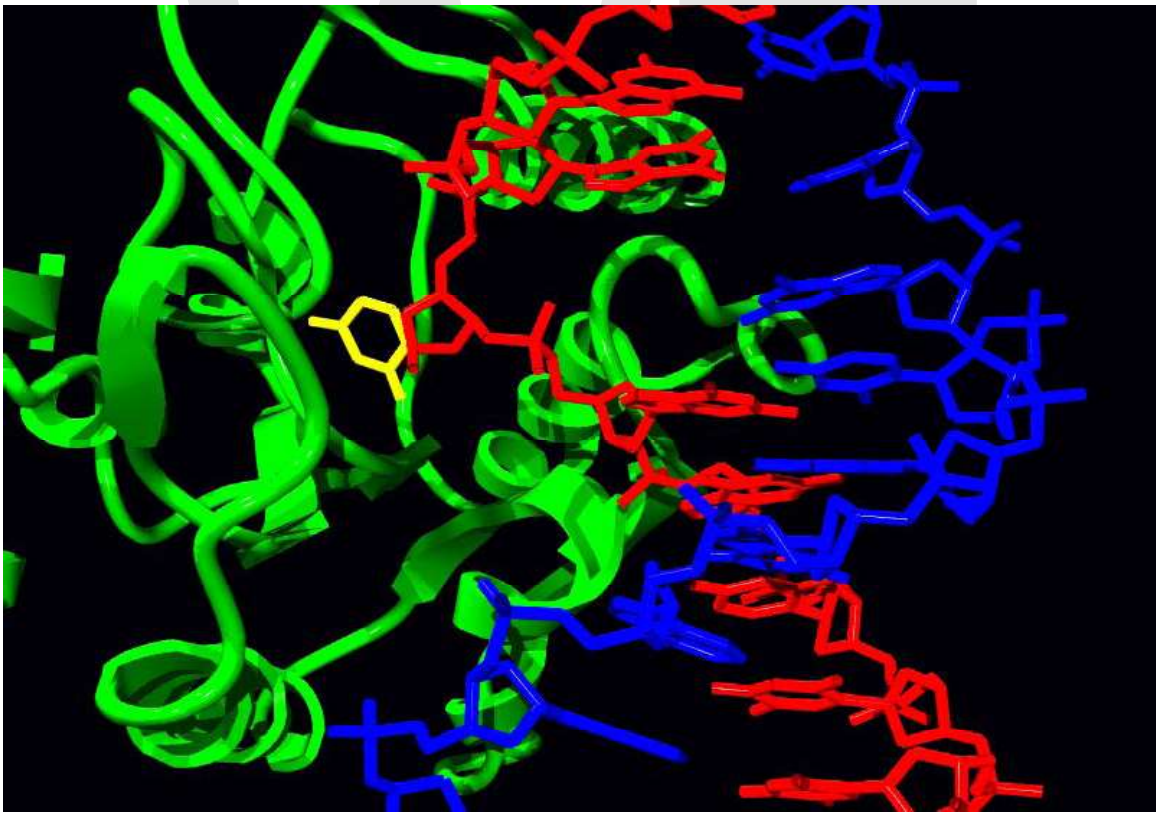
The choice between long-patch and short-patch repair

The choice between short- and long-patch repair is currently under investigation. Various factors are thought to influence this decision, including the type of lesion, the cell cycle stage, and whether the cell is terminally differentiated or actively dividing. Some lesions, such as oxidized or reduced AP sites, are resistant to pol β lyase activity and, therefore, must be processed by long-patch BER.

Pathway preference may differ between organisms, as well. While human cells utilize both short- and long-patch BER, the yeast *Saccharomyces cerevisiae* was long thought to lack a short-patch pathway because it does not have homologs of several mammalian short-patch proteins, including pol β, DNA ligase III, XRCC1, and the kinase domain of PNKP. The recent discovery that the poly-A polymerase Trf4 possesses 5' dRP lyase activity has challenged this view.

Proteins involved in base excision repair

DNA glycosylases



Uracil DNA glycosylase flips a uracil residue out of the duplex, shown in yellow.

DNA glycosylases are responsible for initial recognition of the lesion. They flip the damaged base out of the double helix, as pictured, and cleave the N-glycosidic bond of the damaged base, leaving an AP site. There are two categories of glycosylases:

monofunctional and bifunctional. Monofunctional glycosylases have only glycosylase activity, whereas bifunctional glycosylases also possess AP lyase activity. Therefore, bifunctional glycosylases can convert a base lesion into a single-strand break without the need for an AP endonuclease. β-Elimination of an AP site by a glycosylase-lyase yields a 3' α,β-unsaturated aldehyde adjacent to a 5' phosphate, which differs from the AP endonuclease cleavage product. Some glycosylase-lyases can further perform δ-elimination, which converts the 3' aldehyde to a 3' phosphate. A wide variety of glycosylases have evolved to recognize different damaged bases. Examples of DNA glycosylases include Ogg1, which recognizes 8-oxoguanine, Mag1, which recognizes 3-methyladenine, and UNG, which removes uracil from DNA.

AP endonucleases

The AP endonucleases cleave an AP site to yield a 3' hydroxyl adjacent to a 5' deoxyribosephosphate (dRP). AP endonucleases are divided into two families based on their homology to the ancestral bacterial AP endonucleaes endonuclease IV and exonuclease III. Many eukaryotes have members of both families, including the yeast Saccharomyces cerevisiae, in which Apn1 is the EndoIV homolog and Apn2 is related to ExoIII. In humans, only a single AP endonuclease, APE1, has been identified. It is a member of the ExoIII family.

End processing enzymes

In order for ligation to occur, a DNA strand break must have a hydroxyl on its 3' end and a phosphate on its 5' end. In humans, polynucleotide kinase-phosphatase (PNKP) promotes formation of these ends during BER. This protein has a kinase domain, which phosphorylates 5' hydroxyl ends, and a phosphatase domain, which removes phosphates from 3' ends. Together, these activities ready single-strand breaks with damaged termini for ligation. The AP endonucleases also participate in 3' end processing. Besides opening AP sites, they possess 3' phosphodiesterase activity and can remove a variety of 3' lesions including phosphates, phosphoglycolates, and aldehydes. 3'-Processing must occur before DNA synthesis can initiate because DNA polymerases require a 3' hydroxyl to extend from.

DNA polymerases

Pol β is the main human polymerase that catalyzes short-patch BER, with pol λ able to compensate in its absence. These polymerases are members of the Pol X family and typically insert only a single nucleotide. In addition to polymerase activity, these enzymes have a lyase domain that removes the 5' dRP left behind by AP endonuclease cleavage. During long-patch BER, DNA synthesis is thought to be mediated by pol δ and pol ε along with the processivity factor PCNA, the same polymerases that carry out DNA replication. These polymerases perform displacing synthesis, meaning that the downstream 5' DNA end is "displaced" to form a flap. Pol β can also perform long-patch displacing synthesis and can, therefore, participate in either BER pathway. Long-patch synthesis typically inserts 2-10 new nucleotides.

Flap endonuclease

FEN1 removes the 5' flap generated during long patch BER. This endonuclease shows a strong preference for a long 5' flap adjacent to a 1-nt 3' flap. The yeast homolog of FEN1 is *RAD27*. In addition to its role in long-patch BER, FEN1 cleaves flaps with a similar structure during Okazaki fragment processing, an important step in lagging strand DNA replication.

DNA ligase

DNA ligase III along with its cofactor XRCC1 catalyzes the nick-sealing step in short-patch BER in humans. DNA ligase I ligates the break in long-patch BER.

Links between BER and cancer

Defects in a variety of DNA repair pathways lead to cancer predisposition, and BER appears to follow this pattern. Deletion of BER genes increases the mutation rate in a variety of organisms, predicting that loss of BER could contribute to the development of cancer. Indeed, somatic mutations in Pol β have been found in 30% of human cancers, and some of these mutations lead to transformation when expressed in mouse cells. Mutations in the DNA glycosylase MYH are also known to increase susceptibility to colon cancer.

Nucleotide excision repair

Nucleotide excision repair is a DNA repair mechanism. DNA constantly requires repair due to damage that can occur to bases from a vast variety of sources including chemicals, radiation and other. Nucleotide excision repair (NER) is a particularly important mechanism by which the cell can prevent unwanted mutations by removing the vast majority of UV-induced DNA damage (mostly in the form of thymine dimers and 6-4-photoproducts). The importance of this repair mechanism is evidenced by the severe human diseases that result from in-born genetic mutations of NER proteins including Xeroderma pigmentosum and Cockayne's syndrome. While the base excision repair machinery can recognize specific lesions in the DNA it can correct only damaged bases that can be removed by a specific glycosylase, the nucleotide excision repair enzymes recognize bulky distortions in the shape of the DNA double helix. Recognition of these distortions leads to the removal of a short single-stranded DNA segment that includes the lesion, creating a single-strand gap in the DNA, which is subsequently filled in by DNA polymerase, which uses the undamaged strand as a template. NER can be divided into two subpathways (Global genomic NER and Transcription coupled NER) that differ only in their recognition of helix-distorting DNA damage.

Uvr Proteins

The process of nucleotide excision repair is controlled in *E. coli* by the UvrABC endonuclease enzyme complex, which consists of four Uvr proteins: UvrA, UvrB, UvrC, and DNA helicase II (sometimes also known as UvrD in this complex). First, a UvrA-UvrB complex scans the DNA, with the UvrA subunit recognizing distortions in the helix, caused for example by pyrimidine dimers. When the complex recognizes such a distortion, the UvrA subunit leaves and an UvrC protein comes in and binds to the UvrB monomer and, hence, forms a new UvrBC dimer. UvrB cleaves a phosphodiester bond 4 nucleotides downstream of the DNA damage, and the UvrC cleaves a phosphodiester bond 8 nucleotides upstream of the DNA damage and created 12 nucleotide excised segment. DNA helicase II (sometimes called UvrD) then comes in and removes the excised segment by actively breaking the hydrogen bonds between the complementary bases. The resultant gap is then filled in using DNA polymerase I and DNA ligase. The basic excision process is very similar in higher cells, but these cells usually involve many more proteins – *E.coli* is a simple example.

Nucleotide Excision Repair in Eukaryotes

Nucleotide excision repair has more complexity in eukaryotes. But the general principles upon which it operates are similar. There are 9 major proteins involved in NER in mammalian cells and their names come from the diseases associated with the deficiencies in those proteins. XPA, XPB, XPC, XPD, XPE, XPF, and XPG all derive from Xeroderma pigmentosum and CSA and CSB represent proteins linked to Cockayne syndrome. Additionally, the proteins ERCC1, RPA, RAD23A, RAD23B, and others also participate in nucleotide excision repair.

As described below, nucleotide excision repair can be categorized into two classes, global genome NER (GG-NER) and Transcription Coupled NER (TC-NER). Two different sets of proteins are involved in the distortion and recognition of the DNA damage in the two types of NER. In GG-NER, the XPC-Rad23B complex is responsible for distortion recognition, and DDB1 and DDB2 (XPE) can also recognize some types of damage caused by UV light. Additionally, XPA performs a function in damage recognition that is as yet poorly defined. In TC-NER, CS proteins CSA and CSB bind some types of DNA damage instead of XPC-Rad23B.

The subsequent steps in GG-NER and TC-NER are similar to each other and to those in NER in prokaryotes. XPB and XPD, which are subunits of transcription factor TFIIH have helicase activity and unwind the DNA at the sites of damages. XPG protein has a structure-specific endonuclease activity, which makes an incision 3’ to the damaged DNA. Subsequently XPF protein, which is associated with ERCC1 makes the 5' incision during the NER. The dual-incision leads to the removal of a ssDNA with a single strand gap of 25~30 nucleotides.

The resulting gap in DNA is filled by DNA polymerase δ or ε by copying the undamaged strand. Proliferating Cell Nuclear Antigen (PCNA) assists the DNA polymerase in the

reaction, and Replication protein A (RPA) protects the other DNA strand from degradation during NER. Finally, DNA ligase seals the nicks to finish NER.

Global genomic NER

Global genomic NER repairs damage in both transcribed and untranscribed DNA strands in active and inactive genes throughout the genome. This pathway employs several 'damage sensing' proteins including the DNA-damage binding (DDB) and XPC-Rad23B complexes that constantly scan the genome and recognize helix distortions. Upon identification of a damaged site, subsequent repair proteins are then recruited to the damaged DNA to verify presence of DNA damage, excise the damaged DNA surrounding the lesion then fill in the repair patch.

A 2002 paper published by Frit et al. showed that Global Genome Repair can be enhanced near transcribed regions. Their main conclusion is that the binding of transcription factors led to an increase in nucleotide excision repair on nearby regions of DNA. They showed that this enhancement was distinct from Transcription Coupled Repair in two ways: Repair was occurring on both the template and non-template strands. Secondly, when Pol-II activity was blocked, either by α-amanitin or modification of the TATA box, the repair still occurred. Through digestions with micrococcal nuclease they determined that the binding of transcription factors led to local chromatin remodeling, likely giving NER proteins greater access to the DNA.

Transcription coupled repair in Eukaryotes

There is a difference in NER efficiency between transcriptionally silent and transcriptionally active regions of the genome. This arises largely because- for many types of lesions- NER repairs the transcribed strands of transcriptionally active genes faster than it repairs the nontranscribed strands and faster than it repairs transcriptionally silent DNA- this particular mode of NER is called TC-NER. At any given time- most of the genome in an organism is not undergoing transcription. The genome wide process which repairs damage in both transcribed and untranscribed DNA strands in active and inactive genes is called GGR – this process is not dependent on transcription. TC-NER and GGR are two subpathways of NER- differing only in the initial steps of DNA damage recognition. The principle difference is that TC-NER does not require XPC or DDB proteins for distortion recognition in mammalian cells. TC-NER is instead thought to be initiated when RNA polymerase stalls at a lesion in DNA. In TC-NER the blocked RNA polymerase serves as a damage recognition signal, replacing the need for the distortion recognition properties of the XPC-RAD23B and DDB complexes. Subsequent steps in TC-NER utilise the NER factors XPA, TFIIH, and RPA as well as the nucleases ECC1-XPF and XPG for dual incision at a lesion. The repair patch size for mammalian cells in vivo is around 30 nucleotides, indistinguishable from the repair patch size for GGR.

Chapter- 3

DNA Glycosylase and Homology Directed Repair

DNA glycosylase

DNA glycosylases are a family of enzymes involved in base excision repair, classified under EC number EC 3.2.2. Base excision repair is the mechanism by which damaged bases in DNA are removed and replaced. DNA glycosylases catalyze the first step of this process. They remove the damaged nitrogenous base while leaving the sugar-phosphate backbone intact, creating an apurinic/apyrimidinic site, commonly referred to as an AP site. This is accomplished by flipping the damaged base out of the double helix followed by cleavage of the N-glycosidic bond. Glycosylases were first discovered in bacteria, and have since been found in all kingdoms of life.

Monofunctional vs. bifunctional glycosylases

There are two main classes of glycosylases: monofunctional and bifunctional. Monofunctional glycosylases have only glycosylase activity, whereas bifunctional glycosylases also possess AP lyase activity that permits them to cut the phosphodiester bond of DNA, creating a single-strand break without the need for an AP endonuclease. β-Elimination of an AP site by a glycosylase-lyase yields a 3' α,β-unsaturated aldehyde adjacent to a 5' phosphate, which differs from the AP endonuclease cleavage product. Some glycosylase-lyases can further perform δ-elimination, which converts the 3' aldehyde to a 3' phosphate.

Biochemical mechanism

The first crystal structure of a DNA glycosylase was obtained for E. coli Nth. This structure revealed that the enzyme flips the damaged base out of the double helix into an active site pocket in order to excise it. Other glycosylases have since been found to follow the same general paradigm, including human UNG pictured below. To cleave the N-glycosidic bond, monofunctional glycosylases use an activated water molecule to

attack carbon 1 of the substrate. Bifunctional glycosylases, instead, use an amine residue as a nucleophile to attack the same carbon, going through a Schiff base intermediate.

Types of glycosylases

Crystal structures of many glycosylases have been solved. Based on structural similarity, glycosylases are grouped into four superfamilies. The **UDG** and **AAG** families contain small, compact glycosylases, whereas the **MutM/Fpg** and **HhH-GPD** families comprise larger enzymes with multiple domains.

A wide variety of glycosylases have evolved to recognize different damaged bases. The table below summarizes the properties of known glycosylases in commonly studied model organisms.

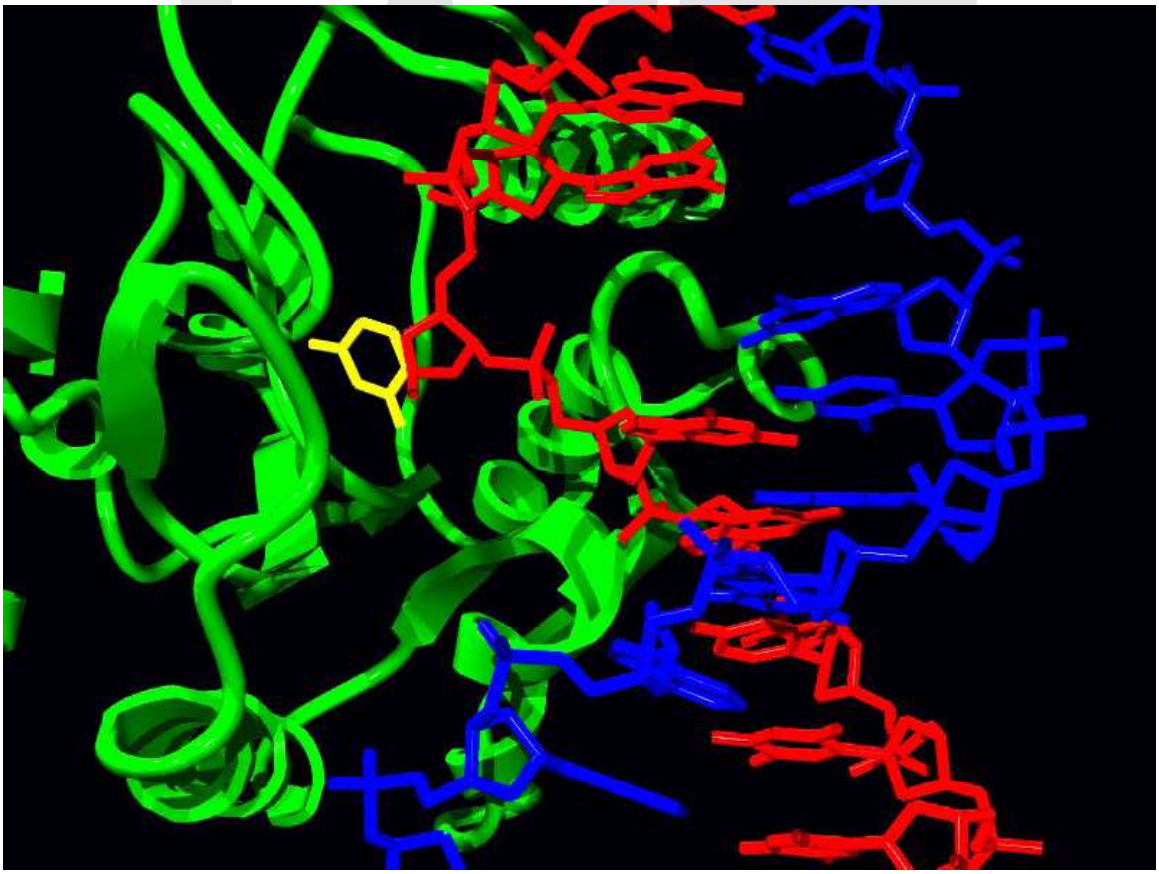
Glycosylases in bacteria, yeast and humans

E. coli	Yeast (S. cerevisiae)	Human	Type	Substrates
AlkA	Mag1		monofunctional	3-meA, hypoxanthine
UDG	Ung1	UNG	monofunctional	uracil
Fpg	Ogg1	hOGG1	bifunctional	8-oxoG, FapyG
Nth	Ntg1	hNTH1	bifunctional	Tg, hoU, hoC, urea, FapyG
	Ntg2			
Nei	Not present	hNEIL1	bifunctional	Tg, hoU, hoC, urea, FapyG, FapyA
		hNEIL2		AP site, hoU
		hNEIL3		unknown

MutY	Not present	hMYH	monofunctional	A:8-oxoG
Not present	Not present	hSMUG1	monofunctional	U, hoU, hmU, fU
Not present	Not present	TDG	monofunctional	T:G mispair
Not present	Not present	MBD4	monofunctional	T:G mispair

DNA glycosylases can be grouped into the following categories based on their substrate(s):

Uracil DNA glycosylases

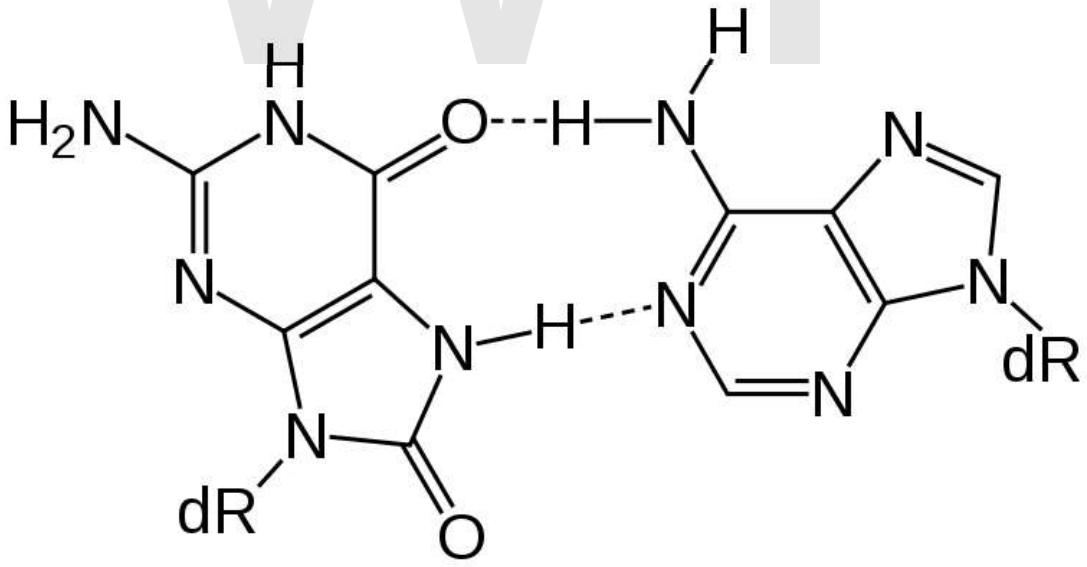


Structure of the base-excision repair enzyme uracil-DNA glycosylase. The uracil residue is shown in yellow.

Uracil DNA glycosylases remove uracil from DNA, which can arise either by spontaneous deamination of cytosine or by the misincorporation of dU opposite dA during DNA replication. The prototypical member of this family is E. coli UDG, which was among the first glycosylases discovered. Four different uracil-DNA glycosylase activities have been identified in mammalian cells, including UNG, SMUG1, TDG, and MBD4. They vary in substrate specificity and subcellular localization. SMUG1 prefers single-stranded DNA as substrate, but also removes U from double-stranded DNA. In addition to unmodified uracil, SMUG1 can excise 5-hydroxyuracil, 5-hydroxymethyluracil and 5-formyluracil bearing an oxidized group at ring C5. TDG and MBD4 are strictly specific for double-stranded DNA. TDG can remove thymine glycol when present opposite guanine, as well as derivatives of U with modifications at carbon 5. Current evidence suggests that, in human cells, TDG and SMUG1 are the major enzymes responsible for the repair of the U:G mispairs caused by spontaneous cytosine deamination, whereas uracil arising in DNA through dU misincorporation is mainly dealt with by UNG. MBD4 is thought to correct T:G mismatches that arise from deamination of 5-methylcytosine to thymine in CpG sites. MBD4 mutant mice develop normally and do not show increased cancer susceptibility or reduced survival. But they acquire more C T mutations at CpG sequences in epithelial cells of the small intestine.

The structure of human UNG in complex with DNA revealed that, like other glycosylases, it flips the target nucleotide out of the double helix and into the active site pocket. UDG undergoes a conformational change from an ‘‘open’’ unbound state to a ‘‘closed’’ DNA-bound state.

Glycosylases of oxidized bases



8-oxoG (*syn*) in a Hoogsteen base pair with dA (*anti*)

A variety of glycosylases have evolved to recognize oxidized bases, which are commonly formed by reactive oxygen species generated during cellular metabolism. The most

abundant lesions formed at guanine residues are 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 8-oxoguanine. Due to mispairing with adenine during replication, 8-oxoG is highly mutagenic, resulting in G to T transversions. Repair of this lesion is initiated by the bifunctional DNA glycosylase OGG1, which recognizes 8-oxoG paired with C. hOGG1 is a bifunctional glycosylase that belongs to the helix-hairpin-helix (HhH) family. MYH recognizes adenine mispaired with 8-oxoG but excises the A, leaving the 8-oxoG intact. OGG1 knockout mice do not show an increased tumor incidence, but accumulate 8-oxoG in the liver as they age. A similar phenotype is observed with the inactivation of MYH, but simultaneous inactivation of both MYH and OGG1 causes 8-oxoG accumulation in multiple tissues including lung and small intestine. In humans, mutations in MYH are associated with increased risk of developing colon polyps and colon cancer. In addition to OGG1 and MYH, human cells contain three additional DNA glycosylases, NEIL1, NEIL2, and NEIL3. These are homologous to bacterial Nei, and their presence likely explains the mild phenotypes of the OGG1 and MYH knockout mice.

Glycosylases of alkylated bases

This group includes E. coli AlkA and related proteins in higher eukaryotes. These glycosylases are monofunctional and recognize methylated bases, such as 3-methyladenine.

Homology directed repair

Homology directed repair (HDR) is a mechanism in cells to repair double strand DNA lesions. This repair mechanism can only be used by the cell when there is a homologue piece of DNA present in the nucleus, mostly in G2 and S phase of the cell cycle. When the homologue DNA piece is absent, another process called non-homologous end joining (NHEJ) can take place instead.

Cancer Suppression

HDR is important for suppressing the formation of cancer. HDR maintains the genomic stability by repairing the broken DNA strand, assumed error free because of the use of a template. When a double strand DNA lesion is repaired by NHEJ there is no validating DNA template present which may result in a non-original DNA strand formation with loss of information. A different nucleotide sequence in the DNA strand results in a different protein expressed in the cell. This protein may malfunction by which processes in the cell may fail. When, for example, a receptor of the cell that can receive a signal to stop dividing malfunctions, the cell ignores the signal and keeps dividing and can form a cancer. An example for the essence of HDR is the fact that the mechanism is conserved

throughout evolution. The HDR mechanism has also been found in more simple organisms, like in yeast.

Biological Pathway

The pathway of HDR is not totally enlightened yet (*March 2008*). Though there are a lot of experimental results which point to the validity of certain models. Generally accepted is the phosphorylation of histon H2AX (noted as γH2AX) within seconds after occurrence of the damage. H2AX is phosphorylated widely in the surrounding area of the damage and not only at the precise location. Therefore γH2AX is suggested to function as an adhesive component to attract proteins to the damaged location. A variety of research groups suggest that the phosphorylation of H2AX is done by ATM and ATR in cooperation with MDC1. Before or meanwhile H2AX is involved in the repair pathway the MRN complex (which exists of Mre11, Rad50 and NBS1) is suggested to stick at the broken DNA ends and other MRN complexes to keep the broken ends together. The act of the MRN complex might prevent chromosomal breaks. Somewhat further in the process the DNA ends are processed so unnecessary left overs of chemical groups etc. are removed and single strand overhands a formed. In the meanwhile, from the beginning, every piece of single stranded DNA is covered by the protein RPA (Replication Protein A). The function of RPA is likely to keep the single stranded DNA piece stable until the complementary piece is resynthesized by a polymerase. After this, Rad51 replaces RPA and forms filaments on the DNA. Together with BRCA2 (Breast Cancer Associated) it couples a complementary DNA piece which invades the broken DNA strand to form a template for the polymerase. The polymerase is hold on the DNA strand by PCNA (Proliferating Cell Nuclear Antigen). PCNA forms typical patterns in the nucleus of the cell by which the current cell cycle can be determined. The polymerase synthesizes the missing part of the broken strand. When the broken strand is rebuild both strands need to uncouple again. Multiple ways of "uncoupling" are suggested though there is no evidence yet to discard or accept a model (*March 2008*). After the strands are separated the process is done.

The co localization of Rad51 with the damage is accepted to be the definite point where HDR is initiated instead of NHEJ. For NHEJ the Ku complex (Ku70 and Ku80) is the point to accept where NHEJ is initiated instead of HDR.

HDR and NHEJ repair double strnd breaks. Other mechanisms such as NER (Nucleotide Excision Repair), BER (Base Excision Repair) and MMR recognise lesions and replace them via single strand perturbation.

Chapter- 4

DNA Mismatch Repair

DNA mismatch repair is a system for recognizing and repairing erroneous insertion, deletion and mis-incorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage.

Mismatch repair is strand-specific. During DNA synthesis the newly synthesised (daughter) strand will include errors commonly. In order to do this the mismatch repair machinery distinguishes the newly synthesised strand from the template (parental). In gram-negative bacteria transient hemimethylation distinguishes the strands (the parental is methylated and daughter is not). In other prokaryotes and eukaryotes the exact mechanism is not clear. It is suspected that in eukaryotes, newly synthesized lagging-strand DNA transiently contains nicks (before being sealed by DNA ligase) and provides a signal that directs mismatch proofreading systems to the appropriate strand. This implies that these nicks must be present in the leading strand, but it is unclear how.

Any mutational event that disrupts the superhelical structure of DNA carries with it the potential to compromise the genetic stability of a cell. The fact that the damage detection and repair systems are as complex as the replication machinery itself highlights the importance evolution has attached to DNA fidelity.

Examples of mismatched bases include a G/T or A/C pairing. Mismatches are commonly due to tautomerization of bases during synthesis. The damage is repaired by recognition of the deformity caused by the mismatch, determining the template and non-template strand, and excising the wrongly incorporated base and replacing it with the correct nucleotide. The removal process involves more than just the mismatched nucleotide itself. A few or up to thousands of base pairs of the newly synthesized DNA strand can be removed.

Mismatch repair proteins

Mismatch repair is a highly conserved process from prokaryotes to eukaryotes. The first evidence for mismatch repair was obtained from *S. pneumoniae* (the hexA and hexB genes). Subsequent work on *E. coli* has identified a number of genes that, when

mutationally inactivated, cause hypermutable strains. The gene products are therefore called the "Mut" proteins, and are the major active components of the mismatch repair system. Three of these proteins are essential in detecting the mismatch and directing repair machinery to it; MutS, MutH and MutL (MutS is a homologue of HexA and MutL of HexB).

MutS forms a dimer ($MutS_2$) that recognises the mismatched base on the daughter strand and binds the mutated DNA. MutH binds at hemimethylated sites along the daughter DNA, but its action is latent, being activated only upon contact by a MutL dimer ($MutL_2$) which binds the MutS-DNA complex and acts as a mediator between $MutS_2$ and MutH, activating the latter. The DNA is looped out to search for the nearest d(GATC) methylation site to the mismatch, which could be up to 1kb away. Upon activation by the MutS-DNA complex, MutH nicks the daughter strand near the hemimethylated site and recruits a UvrD helicase (DNA Helicase II) to separate the two strands with a specific 3' to 5' polarity. The entire MutSHL complex then slides along the DNA in the direction of the mismatch, liberating the strand to be excised as it goes. An exonuclease trails the complex and digests the ss-DNA tail. The exonuclease recruited is dependent on which side of the mismatch MutH incises the strand – 5’ or 3’. If the nick made by MutH is on the 5’ end of the mismatch, either RecJ or ExoVIII (both 5’ to 3’ exonucleases) is used. If however the nick is on the 3’ end of the mismatch, ExoI (a 3' to 5' enzyme) is used.

The entire process ends past the mismatch site - i.e. both the site itself and its surrounding nucleotides are fully excised. The single-stranded gap created by the exonuclease can then be repaired by DNA Polymerase III (assisted by single-strand binding protein), which uses the other strand as a template, and finally sealed by DNA ligase. Dam methylase then rapidly methylates the daughter strand.

MutS homologs

When bound, the $MutS_2$ dimer bends the DNA helix and shields approximately 20 base pairs. It has weak ATPase activity, and binding of ATP leads to the formation of tertiary structures on the surface of the molecule. The crystal structure of MutS reveals that it is exceptionally asymmetric, and while its active conformation is a dimer, only one of the two halves interact with the mismatch site.

In eukaryotes, MutS homologs form two major heterodimers: Msh2/Msh6 (MutSα) and Msh2/Msh3 (MutSβ). The MutSα pathway is involved primarily in base substitution and small loop mismatch repair. The MutSβ pathway is also involved in small loop repair, in addition to large loop (~10 nucleotide loops) repair. However, MutSβ does not repair base substitutions.

MutL homologs

MutL also has weak ATPase activity (it uses ATP for purposes of movement). It forms a complex with MutS and MutH, increasing the MutS footprint on the DNA.

However, the processivity (the distance the enzyme can move along the DNA before dissociating) of UvrD is only ~40–50bp. Because the distance between the nick created by MutH and the mismatch can average ~600 bp, if there isn't another UvrD loaded the unwound section is then free to re-anneal to its complementary strand, forcing the process to start over. However, when assisted by MutL, the *rate* of UvrD loading is greatly increased. While the processivity (and ATP utilisation) of the individual UvrD molecules remains the same, the total effect on the DNA is boosted considerably; the DNA has no chance to re-anneal, as each UvrD unwinds 40-50 bp of DNA, dissociates, and then is immediately replaced by another UvrD, repeating the process. This exposes large sections of DNA to exonuclease digestion, allowing for quick excision(and later replacement) of the incorrect DNA.

Eukaryotes have MutL homologs designated Mlh1 and Pms1. They form a heterodimer which mimics MutL in *E. coli*. The human homologue of prokaryotic MutL has three forms designated as MutLα, MutLβ and MutLγ. The MutLα complex is made of two subunits MLH1 and PMS2, the MutLβ heterodimer is made of MLH1 and PMS1, while MutLγ is made of MLH1 and MLH3. MutLα acts as the matchmaker or facilitator, coordinating events in mismatch repair. It has recently been shown to be a DNA endonuclease that introduces strand breaks in DNA upon activation by mismatch and other required proteins, MutSa and PCNA. These strand interruptions serve as entry points for an exonuclease activity that removes mismatched DNA. Roles played by MutLβ and MutLγ in mismatch repair are less well understood.

MutH: an endonuclease present in E. coli and Salmonella

MutH is a very weak endonuclease that is activated once bound to MutL (which itself is bound to MutS). It nicks unmethylated DNA and the unmethylated strand of hemimethylated DNA but does not nick fully methylated DNA. It has been experimentally shown that mismatch repair is random if neither strand is methylated. These behaviours led to the proposal that MutH determines which strand contains the mismatch. MutH has no eukaryotic homolog. Its endonuclease function is taken up by MutL homologs, which have some specialized 5'-3' exonuclease activity. The strand bias for removing mismatches from the newly synthesized daughter strand in eukaryotes may be provided by the free 3’ ends of Okazaki fragments in the new strand created during replication.

β-sliding clamp/PCNA

PCNA and the β-sliding clamp associate with MutSα/β and MutS, respectively. Although initial reports suggested that the PCNA-MutSα complex may enhance mismatch recognition, it has been recently demonstrated that there is no apparent change in affinity of MutSα for a mismatch in the presence or absence of PCNA. Furthermore, mutants of MutSα that are unable to interact with PCNA *in vitro* exhibit the capacity to carry out mismatch recognition and mismatch excision to near wild type levels. Curiously, such mutants are defective in the repair reaction directed by a 5' strand break, suggesting for the first time MutSα function in a post-excision step of the reaction.

Defects in mismatch repair

Mutations in the human homologues of the Mut proteins affect genomic stability, which can result in microsatellite instability (MI). MI is implicated in most human cancers. Specifically the overwhelming majority of hereditary nonpolyposis colorectal cancers (HNPCC) are attributed to mutations in the genes encoding the MutS and MutL homologues MSH2 and MLH1 respectively, which allows them to be classified as tumour suppressor genes. A subtype of HNPCC is known as Muir-Torre Syndrome (MTS) which is associated with skin tumors.

Chapter- 5

DNA Damage Theory of Aging

The **DNA damage theory of aging** proposes that aging is a consequence of unrepaired DNA damage accumulation. Damage in this context includes chemical reactions that mutate DNA and/or interfere with DNA replication. Although both mitochondrial and nuclear DNA damage can contribute to aging, nuclear DNA is the main subject of this analysis. Nuclear DNA damage can contribute to aging either indirectly (by increasing apoptosis or cellular senescence) or directly (by increasing cell dysfunction).

In humans, DNA damage occurs frequently and DNA repair processes have evolved to compensate. On average, approximately 800 DNA lesions occur per hour in each cell, or about 19,200 per cell per day (Vilenchik & Knudson 2000). In any cell some DNA damage may remain despite the action of repair processes. The accumulation of unrepaired DNA damage is more prevalent in certain types of cells, particularly in non-replicating or slowly replicating cells, which cannot rely on DNA repair mechanisms associated with DNA replication such as those in the brain, skeletal and cardiac muscle.

DNA damage and mutation

To understand the DNA damage theory of aging it is important to distinguish between DNA damage and mutation, the two major types of errors that occur in DNA. DNA damages and mutation are fundamentally different. Damages are physical abnormalities in the DNA, such as single and double strand breaks, 8-hydroxydeoxyguanosine residues and polycyclic aromatic hydrocarbon adducts. DNA damages can be recognized by enzymes, and thus they can be correctly repaired if redundant information, such as the undamaged sequence in the complementary DNA strand or in a homologous chromosome, is available for copying. If a cell retains DNA damage, transcription of a gene can be prevented and thus translation into a protein will also be blocked. Replication may also be blocked and/or the cell may die.

In contrast to DNA damage, a mutation is a change in the base sequence of the DNA. A mutation cannot be recognized by enzymes once the base change is present in both DNA strands, and thus a mutation cannot be repaired. At the cellular level, mutations can cause alterations in protein function and regulation. Mutations are replicated when the cell

replicates. In a population of cells, mutant cells will increase or decrease in frequency according to the effects of the mutation on the ability of the cell to survive and reproduce. Although distinctly different from each other, DNA damages and mutations are related because DNA damages often cause errors of DNA synthesis during replication or repair and these errors are a major source of mutation.

Given these properties of DNA damage and mutation, it can be seen that DNA damages are a special problem in non-dividing or slowly dividing cells, where unrepaired damages will tend to accumulate over time. On the other hand, in rapidly dividing cells, unrepaired DNA damages that do not kill the cell by blocking replication will tend to cause replication errors and thus mutation. The great majority of mutations that are not neutral in their effect are deleterious to a cell's survival. Thus, in a population of cells comprising a tissue with replicating cells, mutant cells will tend to be lost. However infrequent mutations that provide a survival advantage will tend to clonally expand at the expense of neighboring cells in the tissue. This advantage to the cell is disadvantageous to the whole organism, because such mutant cells can give rise to cancer. Thus DNA damages in frequently dividing cells, because they give rise to mutations, are a prominent cause of cancer. In contrast, DNA damages in infrequently dividing cells are likely a prominent cause of aging.

The first person to suggest that DNA damage, as distinct from mutation, is the primary cause of aging was Alexander (1967). By the early 1980s there was significant experimental support for this idea in the literature (Gensler & Bernstein, 1981). By the early 1990s experimental support for this idea was substantial, and furthermore it had become increasingly evident that oxidative DNA damage, in particular, is a major cause of aging (Bernstein & Bernstein, 1991; Ames & Gold, 1991; Holmes et al., 1992; Rao & Loeb, 1992; Ames et al., 1993).

In a series of articles from 1970 to 1977, PV Narasimh Acharya, Phd. (1924–1993) theorized and scientifically proved that cells undergo "irreparable DNA damage," whereby DNA crosslinks occur when both normal cellular repair processes fail and cellular apoptosis does not occur. Specifically, PVN Acharya noted that double-strand breaks and a "cross-linkage joining both strands at the same point is irreparable because neither strand can then serve as a template for repair. The cell will die in the next mitosis or in some rare instances, mutate." (PVN Acharya; PVN Acharya & Bjorksten et al.) Acharya's research also showed how irreparable DNA damage is caused by environmental pollutants, low dose ionizing radiation and food additives, particularly nitrites and nitrates and such damage to the DNA is a causal factor for pre-mature aging and cancer.

Age-associated accumulation of DNA damage and decline in gene expression

In tissues composed of non- or infrequently replicating cells, DNA damages can accumulate with age and lead either to loss of cells, or, in surviving cells, loss of gene expression. Accumulated DNA damages are usually measured directly. Numerous studies

of this type have indicated that oxidative damage to DNA is particularly important. The loss of expression of specific genes can be detected at both the mRNA level and protein level.

Brain

The adult brain is composed in large part of terminally differentiated non-dividing neurons. Many of the conspicuous features of aging reflect a decline in neuronal function. Accumulation of DNA damage with age in the mammalian brain has been reported during the period 1971 to the present in at least 29 studies, too many to describe here. A review of the role of DNA damage in aging, including a comprehensive summary of the studies showing DNA damage accumulation with age in brain, muscle, liver and kidney, is in press (Bernstein et al., 2008). Here, we mention only some recent studies involving rodents plus one human study. Rutten et al. (2007) showed that single-strand breaks accumulate in the mouse brain with age. Sen et al. (2007) showed that DNA damages which block the polymerase chain reaction in rat brain accumulate with age. Wolf et al. (2005) showed that the oxidative DNA damage 8-hydroxydeoxyguanosine (8-OHdG) accumulates in rat brain with age. As humans age from 48–97 years, 8-OHdG accumulates in the brain (Mecocci et al., 1993).

Decrements in function were noted in aging human brain, where transcription of a set of evaluated genes declines with age from 40 to 106 years (Lu et al., 2004). These genes play central roles in synaptic plasticity, vesicular transport and mitochondrial function. In the brain, promoters of genes with reduced expression have markedly increased DNA damage (Lu et al., 2004). In cultured human neurons, these gene promoters are selectively damaged by oxidative stress. Thus Lu et al. (2004) concluded that DNA damage may reduce the expression of selectively vulnerable genes involved in learning, memory and neuronal survival, initiating a program of brain aging that starts early in adult life.

Muscle

Muscle strength, and stamina for sustained physical effort, have decrements in function with age in humans and other species. Skeletal muscle is a tissue composed largely of multinucleated myofibers, elements that arise from the fusion of mononucleated myoblasts. Accumulation of DNA damage with age in mammalian muscle has been reported in at least 18 studies (Bernstein et al., in press 2008) since 1971. We will mention here only two of the more recent studies in rodents plus one in humans. Hamilton et al. (2001) reported that the oxidative DNA damage 8-OHdG accumulates in heart and skeletal muscle (as well as in brain, kidney and liver) of both mouse and rat with age. In humans, increases in 8-OHdG with age were reported for skeletal muscle (Mecocci et al., 1999). Catalase is an enzyme that removes hydrogen peroxide, a reactive oxygen species, and thus limits oxidative DNA damage. In mice, when catalase expression is increased specifically in mitochondria, oxidative DNA damage (8-OHdG) in skeletal muscle is decreased and lifespan is increased by about 20% (Schriner et al.,

2005; Linford et al., 2006). These findings suggest that mitochondria are a significant source of the oxidative damages contributing to aging.

Protein synthesis and protein degradation decline with age in skeletal and heart muscle, as would be expected, since DNA damages block gene transcription. In a recent study (Piec et al., 2005) found numerous changes in protein expression in rat skeletal muscle with age, including lower levels of several proteins related to myosin and actin. Force is generated in striated muscle by the interactions between myosin thick filaments and actin thin filaments.

Liver

Liver hepatocytes do not ordinarily divide and appear to be terminally differentiated, but they retain the ability to proliferate when injured. With age, the mass of the liver decreases, blood flow is reduced, metabolism is impaired, and alterations in microcirculation occur. At least 21 studies (Bernstein et al., in press 2008) have reported an increase in DNA damage with age in liver. For instance, Helbock et al. (1998) estimated that the steady state level of oxidative DNA base alterations increased from 24,000 per cell in the liver of young rats to 66,000 per cell in the liver of old rats.

Kidney

In kidney, changes with age include reduction in both renal blood flow and glomerular filtration rate, and impairment in the ability to concentrate urine and to conserve sodium and water. DNA damages, particularly oxidative DNA damages, increase with age (at least 8 studies)(Bernstein et al., in press 2008). For instance Hashimoto et al. (2007) showed that 8-OHdG accumulates in rat kidney DNA with age.

Long-lived stem cells

Tissue-specific stem cells produce differentiated cells through a series of increasingly more committed progenitor intermediates. In hematopoiesis (blood cell formation), the process begins with long-term hematopoietic stem cells that self-renew and also produce progeny cells that upon further replication go through a series of stages leading to differentiated cells without self-renewal capacity. In mice, deficiencies in DNA repair appear to limit the capacity of hematopoietic stem cells to proliferate and self-renew with age (Rossi et al., 2007). Sharpless and Depinho (2007) reviewed evidence that hematopoietic stem cells, as well as stem cells in other tissues, undergo intrinsic aging. They speculated that stem cells grow old, in part, as a result of DNA damage.

Mutation theories of aging

A popular idea, that has failed to gain significant experimental support, is the idea that mutation, as distinct from DNA damage, is the primary cause of aging. As discussed above, mutations tend to arise in frequently replicating cells as a result of errors of DNA synthesis when template DNA is damaged, and can give rise to cancer. However, in the

mouse there is no increase in mutation in the brain with aging (Dolle et al., 1997; Stuart et al., 2000; Hille et al., 2005). Mice defective in a gene (Pms2) that ordinarily corrects base mispairs in DNA have about a 100-fold elevated mutation frequency in all tissues, but do not appear to age more rapidly (Narayanan et al., 1997). On the other hand, mice defective in one particular DNA repair pathway show clear premature aging, but do not have elevated mutation (Dolle et al., 2006).

One variation of the idea that mutation is the basis of aging, that has received much attention, is that mutations specifically in mitochondrial DNA are the cause of aging. Several studies have shown that mutations accumulate in mitochondrial DNA in infrequently replicating cells with age. DNA polymerase gamma is the enzyme that replicates mitochondrial DNA. A mouse mutant with a defect in this DNA polymerase is only able to replicate its mitochondrial DNA inaccurately, so that the mutation rate is 500-fold higher than in normal mice. Yet these mice showed no obvious features of rapidly accelerated aging (Vermulst et al., 2007). The probable explanation for the apparent lack of effect of the additional mutations in mitochondrial DNA is that, within a typical cell, there are large numbers of mitochondria and each mitochondrion can have multiple copies of mitochondrial DNA. Since most mutations are recessive, any particular deleterious mutation would not be expected to have a pronounced effect when many copies of the correct DNA sequence are present in the same and in other mitochondria in the cell.

Dietary Restriction

In rodents, caloric restriction slows aging and extends lifespan. At least 4 studies have shown that caloric restriction reduces 8-OHdG damages in various organs of rodents. One of these studies (Hamilton et al., 2001) showed that caloric restriction reduced accumulation of 8-OHdG with age in rat brain, heart and skeletal muscle, and in mouse brain, heart, kidney and liver. More recently, Wolf et al. (2005) showed that dietary restriction reduced accumulation of 8-OHdG with age in rat brain, heart, skeletal muscle, and liver. Thus reduction of oxidative DNA damage is associated with a slower rate of aging and increased lifespan.

Inherited defects that cause premature aging

If DNA damage is the underlying cause of aging, it would be expected that humans with inherited defects in the ability to repair DNA damages should age at a faster pace than persons without such a defect. Numerous examples of rare inherited conditions with DNA repair defects are known. Several of these show multiple striking features of premature aging, and others have fewer such features. Perhaps the most striking premature aging conditions are Werner syndrome (mean lifespan 47 years), Huchinson-Gilford Progeria (mean lifespan 13 years), and Cockayne syndrome (mean lifespan 13 years). Werner syndrome is due to an inherited defect in an enzyme (a helicase and exonuclease) that acts in base excision repair of DNA (e.g. Harrigan et al., 2006). Hutchinson-Guilford Progeria is due to a defect in Lamin A protein which forms a scaffolding within the cell nucleus to organize chromatin and is needed for repair of

double-strand breaks in DNA (Liu et al., 2007). Cockayne Syndrome is due to a defect in a protein necessary for the repair process, transcription coupled nucleotide excision repair, which can remove damages, particularly oxidative DNA damages, that block transcription (D'Errico et al., 2007). In addition to these three conditions, several other human syndromes, that also have defective DNA repair, show several features of premature aging. These include ataxia telangiectasia, Nijmegan breakage syndrome, some subgroups of xeroderma pigmentosum, trichothiodystrophy, Fanconi anemia, Bloom syndrome and Rothmund-Thomson syndrome.

In addition to human inherited syndromes, experimental mouse models with genetic defects in DNA repair show features of premature aging and reduced lifespan (e.g. Vogel et al., 1999; Niedernhoffer et al., 2006; Mostoslavsky et al., 2006).

Lifespan in different mammalian species

Studies comparing DNA repair capacity in different mammalian species have shown that repair capacity correlates with lifespan. The initial study of this type, by Hart and Setlow (1974), showed that the ability of skin fibroblasts of seven mammalian species to perform DNA repair after exposure to a DNA damaging agent correlated with lifespan of the species. The species studied were shrew, mouse, rat, hamster, cow, elephant and human. This initial study stimulated many additional studies involving a wide variety of mammalian species, and the correlation between repair capacity and lifespan generally held up. In one of the more recent studies, Burkle et al. (2005) studied the level of a particular enzyme, poly(ADP-ribose) polymerase, which is involved in repair of single-strand breaks in DNA. They found that the lifespan of 13 mammalian species correlated with the activity of this enzyme. In addition, they found that humans who lived past 100 years had a significantly higher activity of this enzyme than younger individuals.

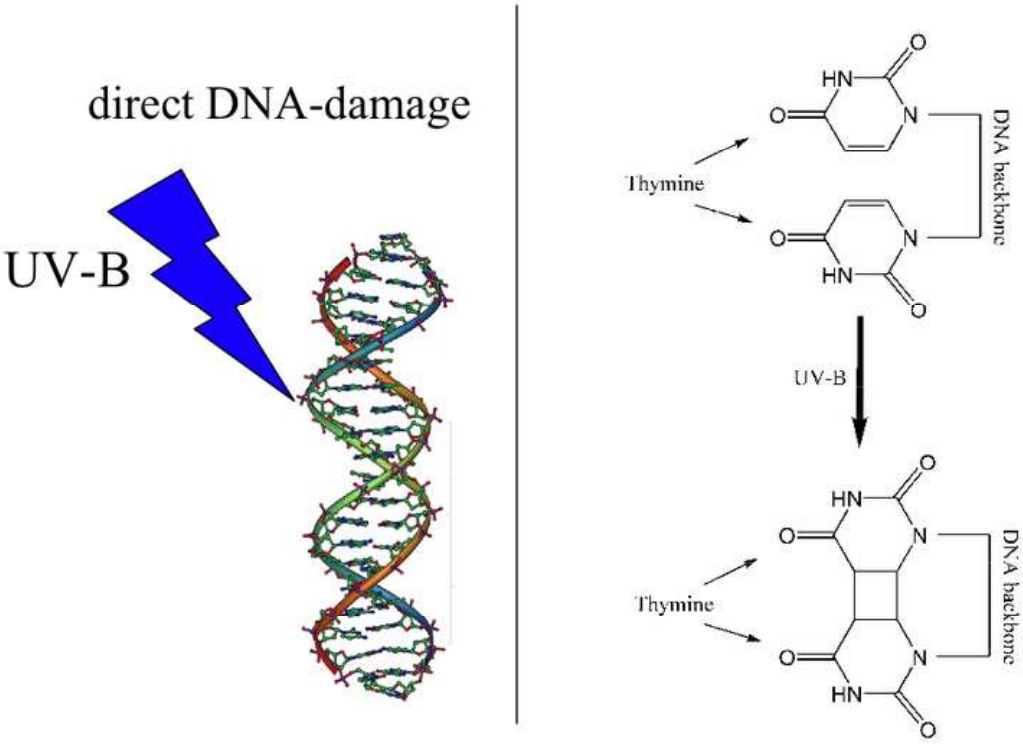
Conclusions

Numerous studies have shown that DNA damage accumulates in brain, muscle, liver, kidney, and in long-lived stem cell. These accumulated DNA damages are the likely cause of the decline in gene expression and loss of functional capacity observed with increasing age. On the other hand, accumulation of mutations, as distinct from DNA damages, is not a plausible candidate as the primary cause of aging. A calorie-restricted diet in mammals improves lifespan, and this improvement is associated with a decrease in oxidative DNA damage. Several inherited genetic defects in ability to repair DNA damage give rise to premature aging suggesting a causal relationship between DNA damage and aging. In comparisons of different mammalian species that differ in lifespan, DNA repair capacity is found to correlate with lifespan. The principal source of the DNA damages leading to normal aging appears to be reactive oxygen species, produced as byproducts of normal cellular metabolism.

Chapter- 6

Direct DNA Damage and Indirect DNA Damage

Direct DNA damage



Direct DNA damage: The UV-photon is directly absorbed by the DNA (left). One of the possible reactions from the excited state is the formation of a thymine-thymine cyclobutane dimer (right). The direct DNA damage leads to sunburn, causing an increase in melanin production, thereby leading to a long-lasting tan. However, it is responsible for only 8% of all melanoma.

Direct DNA damage can occur when DNA directly absorbs the UV-B-photon. UVB light causes thymine base pairs next to each other in genetic sequences to bond together into thymine dimers, a disruption in the strand, which reproductive enzymes cannot copy. It causes sunburn and it triggers the production of melanin.

Other names for the "direct DNA damage" are:

- thymine dimers
- pyrimidine dimers
- Cyclobutane Pyrimidine Dimers (CPD's).
- UV-endonuclease-sensitive-sites (ESS)

Due to the excellent photochemical properties of DNA, this nature-made molecule is damaged only by a tiny fraction of the absorbed photons. DNA transforms more than 99.9% of the photons into harmless heat (But the damage from the remaining < 0.1% of the photons is still enough to cause sunburn). The transformation of excitation energy into harmless heat occurs via a photochemical process called internal conversion. In DNA, this internal conversion is extremely fast - and therefore efficient. This ultrafast (subpicosecond) internal conversion is an extremely powerful photoprotection provided by single nucleotides. However, the Ground-State Recovery is much slower (picoseconds) in G·C−DNA duplexes and hairpins. Probably it is even slower for double stranded DNA in conditions of the nucleus. The absorption spectrum of DNA shows a strong absorption for UVB-radiation and a much lower absorption for UVA-radiation. Since the action spectrum of sunburn is identical to the absorption spectrum of DNA, it is generally accepted that the direct DNA damages are the cause of sunburn. While the human body reacts to direct DNA damages with a painful warning signal,no such warning signal is generated from indirect DNA damage.

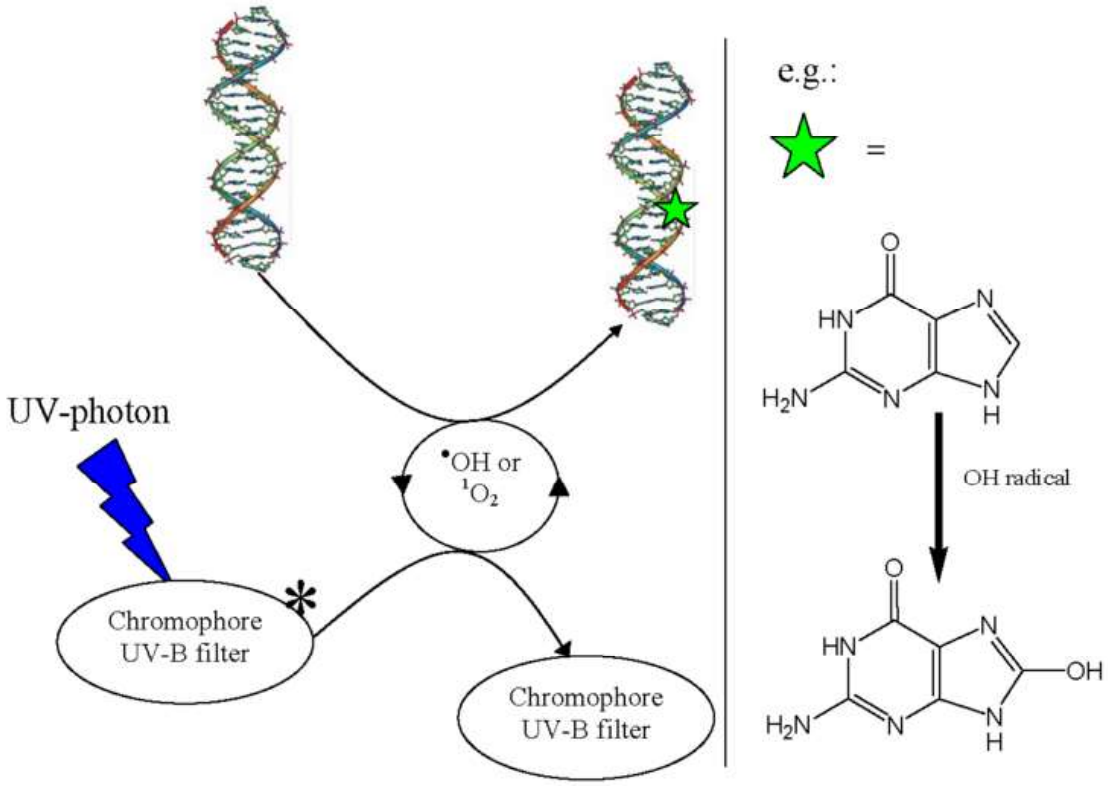
Sunscreen and melanoma

A study by Hanson suggests sunscreen that penetrates into the skin and thereby amplifies the amount of free radicals and oxidative stress contributes to the formation of melanoma, but this idea has not been validated by other researchers.

Effect of topical sunscreen and effect of absorbed sunscreen

The direct DNA damage is reduced by sunscreen. This prevents sunburn. When the sunscreen is at the surface of the skin, it filters the UV-rays, which attenuates the intensity. Even when the sunscreen molecules have penetrated into the skin, they protect against the direct DNA damage, because the UV-light is absorbed by the sunscreen and not by the DNA.

Indirect DNA damage



Indirect DNA damage: The chromophore absorbs UV-light (* denotes an excited state), and the energy of the excited state is creating singlet oxygen ($^1O_2$) or a hydroxyl radical (•OH), which then damages DNA through oxidation.

Indirect DNA damage occurs when a UV-photon is absorbed in the human skin by a chromophore that does not have the ability to convert the energy into harmless heat very quickly. Molecules that do not have this ability have a long lived excited state. This long lifetime leads to a high probability for reactions with other molecules - so called bimolecular reactions. Melanin and DNA have extremely short excited state lifetimes in the range of a few femtoseconds (10^{-15}s). The excited state lifetime of these substances is 1,000 to 1,000,000 times longer than the lifetime of melanin and therefore they may cause damage to living cells that come in contact with them.

The molecule that originally absorbs the UV-photon is called a "chromophore". Bimolecular reactions can occur either between the excited chromophore and DNA, or between the excited chromophore and another species, to produce free radicals and Reactive Oxygen Species. These reactive chemical species can reach DNA by diffusion

and the bimolecular reaction damages the DNA (oxidative stress). Importantly, indirect DNA damage does not result in any warning signal or pain in the human body.

The bimolecular reactions that cause the indirect DNA damage are illustrated in the figure:

$$(\text{Chromophore})^* + {}^3O_2 \rightarrow \text{Chromophore} + {}^1O_2$$

${}^1O_2$ is reactive harmful singlet oxygen:

$${}^1O_2 + \text{intact DNA} \rightarrow {}^3O_2 + \text{damaged DNA}$$

Location of the damage

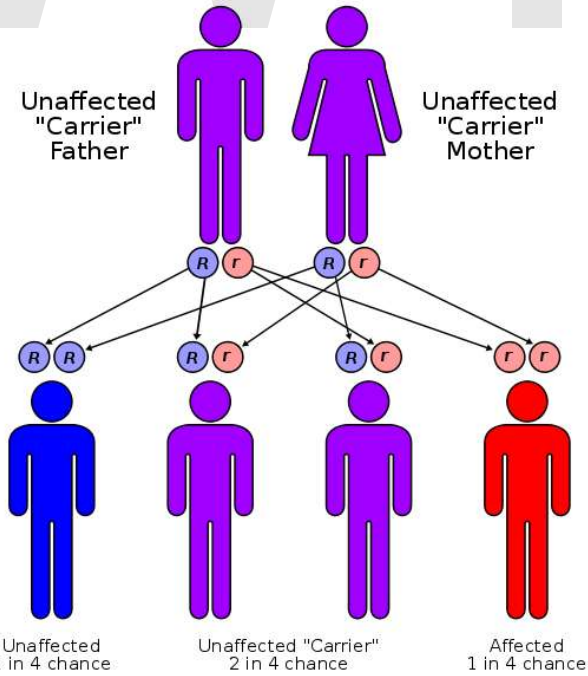
Direct DNA damage is confined to areas that can be reached by UV-B light. In contrast free radicals can travel through the body and affect other areas - possibly even inner organs. The traveling nature of the indirect DNA damage can be seen in the fact that the malignant melanoma can occur in places that are not directly illuminated by the sun—in contrast to basal-cell carcinoma and squamous cell carcinoma, which only appear on directly illuminated locations on the body.

Chapter- 7

Xeroderma Pigmentosum and Bloom Syndrome

Xeroderma pigmentosum

Xeroderma pigmentosum	
ICD-10	Q82.1
ICD-9	757.33
DiseasesDB	14198
eMedicine	derm/462 neuro/399
MeSH	D014983



Xeroderma pigmentosum has an autosomal recessive pattern of inheritance.

Xeroderma pigmentosum, or XP, is an autosomal recessive genetic disorder of DNA repair in which the ability to repair damage caused by ultraviolet (UV) light is deficient. In extreme cases all exposure to sunlight must be forbidden, no matter how small. Multiple basal cell carcinomas (basaliomas) and other skin malignancies frequently occur at a young age in those with XP. In fact, metastatic malignant melanoma and squamous cell carcinoma are the two most common causes of death in XP victims. thedoctorsdoctor.com reports XP involves both sexes and all races with an incidence of 1:250,000 and a gene frequency of 1:200. XP is roughly six times more common in Japanese people than in other groups.

The most common defect in xeroderma pigmentosum is an autosomal recessive genetic defect in which nucleotide excision repair (NER) enzymes are mutated, leading to a reduction in or elimination of Nucleotide Excision Repair. If left unchecked, damage caused by UV light can cause mutations in individual cell's DNA. If tumor suppressor genes (e.g. p53) or proto oncogenes are affected, the result may be cancer. Patients with XP are at a high risk for developing skin cancers, such as basal cell carcinoma, for this reason.

Normally, damage to DNA in epidermal cells occurs during exposure to UV light. The absorption of the high energy light leads to the formation of pyrimidine dimers, namely CPDs (cyclobutane-pyrimidine-dimers) and 6-4PPs (pyrimidine-6-4-pyrimidone photoproducts). In a healthy, normal human being, the damage is first excised, "cut out" by endonucleases. DNA polymerase then repairs the missing sequence, and ligase "seals" the transaction. This process is known as nucleotide excision repair.

Types

There are 7 complementation groups, plus one variant form:

Type	Diseases Database	OMIM	Gene	Locus	Also known as/Description
Type A, I, XPA	29877	278700	XPA	9q22.3	**Xeroderma pigmentosum group A**. Classical form of XP.
Type B, II, XPB	29878	133510	XPB	2q21	**Xeroderma pigmentosum group B**.
Type C, III, XPC	29879	278720	XPC	3p25	**Xeroderma pigmentosum group C**.
Type D, IV, XPD	29880	278730 278800	XPD ERCC6	19q13.2-q13.3, 10q11	**Xeroderma pigmentosum group D** or **De Sanctis-Cacchione syndrome**. De Sanctis-Cacchione syndrome can be considered a subtype of XPD.
Type	29881	278740	DDB2	11p12-	**Xeroderma pigmentosum group E**.

E, V, XPE				p11	
Type F, VI, XPF	29882	278760	ERCC4	16p13.3-p13.13	**Xeroderma pigmentosum group F**.
Type G, VII, XPG	29883	278780 133530	RAD2 ERCC5	13q33	**Xeroderma pigmentosum group G** and COFS syndrome type 3.
Type V, XPV		278750	POLH	6p21.1-p12	**Xeroderma pigmentosum variant**. XPV patients suffer from mutation in a gene that codes for a specialized DNA polymerase called polymerase-η (eta). Polymerase-η can replicate over the damage and is needed when cells enter S-phase in the presence of a DNA-damage.

Symptoms

Some of the most common symptoms of XP include:

- A severe sunburn when exposed to only small amounts of sunlight. Often occurring during a child's first exposure to sunlight.
- Development of many freckles at an early age
- Rough-surfaced growths (solar keratoses), and skin cancers
- Eyes that are painfully sensitive to the sun and may easily become irritated, bloodshot, and clouded
- Blistering or freckling on minimum sun exposure
- Spidery blood vessels
- Oozing raw skin surface
- Limited growth of hair on chest and legs
- Scaly skin
- Irregular dark spots on the skin

Treatment

The most obvious, and often important part of treatment is avoiding exposure to sunlight. Keratoses can also be treated using cryotherapy or fluorouracil.

Prognosis

Fewer than 40% of individuals with the disease survive beyond age 20 years. Some XP victims with less severe cases of the disease do manage to live well into their 40s.

Bloom syndrome

Bloom syndrome	
ICD-9	757.39
OMIM	210900
DiseasesDB	1505
eMedicine	derm/54
MeSH	D001816

Bloom syndrome (BLM), also known as **Bloom–Torre–Machacek syndrome**, is a rare autosomal recessive chromosomal disorder characterized by a high frequency of breaks and rearrangements in an affected person's chromosomes. The condition was discovered and first described by dermatologist Dr. David Bloom in 1954.

Presentation

Bloom syndrome is characterized by short stature and a facial rash that develops shortly after first exposure to sun. This rash can make a butterfly-shaped patch of reddened skin on the cheeks. The rash can develop on other sun-exposed areas such as the backs of the hands. Other clinical features include a high-pitched voice; distinct facial features, such as a long, narrow face, micrognathism of the mandible, and prominent nose and ears; pigmentation changes of the skin including hypo- and hyper-pigmented areas and cafe-au-lait spots; telangiectasias (dilated blood vessels) which can appear on the skin but also in the eyes; moderate immune deficiency, characterized by deficiency in certain immunoglobulin classes, that apparently leads to recurrent pneumonia and ear infections; hypo-gonadism characterized by a failure to produce sperm, hence infertility in males, and premature cessation of menses (premature menopause), hence sub-fertility in females. However, several women with Bloom syndrome have had children.

Complications of the disorder may include chronic lung problems, diabetes, and learning disabilities. In a small number of persons, there is mental retardation. The most striking complication of the disorder is susceptibility to cancer, as described in more detail in the next section.

Relationship to cancer

A greatly elevated *rate* of mutation in Bloom syndrome results in a high risk of cancer in affected individuals. The cancer predisposition is characterized by 1) broad spectrum, including leukemias, lymphomas, and carcinomas, 2) early age of onset relative to the same cancer in the general population, and 3) multiplicity. Persons with Bloom syndrome

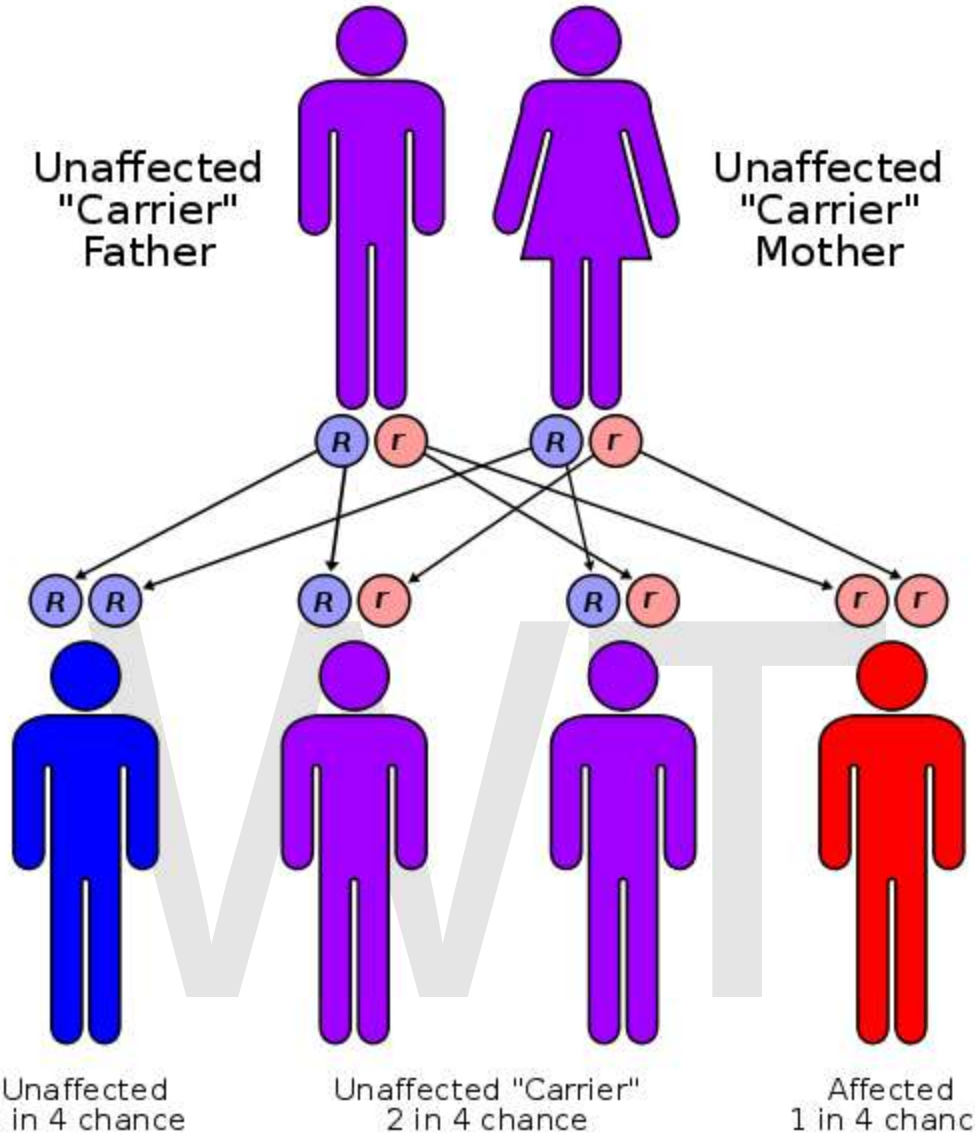
may develop cancer at any age. The average age of cancer diagnoses in the cohort is approximately 25 years old.

Pathophysiology

Mutations in the *BLM* gene, which is a member of the DNA helicase family, are associated with Bloom syndrome. DNA helicases are enzymes that unwind the two strands of a duplex DNA molecule. DNA unwinding is required for most processes that involve the DNA, including synthesis of DNA copies, RNA transcription, DNA repair, etc.

When a cell prepares to divide to form two cells, the chromosomes are duplicated so that each new cell will get a complete set of chromosomes. The duplication process is called DNA replication. Errors made during DNA replication can lead to mutations. The BLM protein is important in maintaining the stability of the DNA during the replication process. The mutations in the *BLM* gene associated with Bloom syndrome inactivate the BLM protein's DNA helicase activity or nullify protein expression (the protein is not made). Lack of BLM protein or protein activity leads to an increase mutations; however, the molecular mechanism(s) by which BLM maintains stability of the chromosomes is still a very active area of research.

Persons with Bloom syndrome have an enormous increase in exchange events between homologous chromosomes or sister chromatids (the two DNA molecules that are produced by the DNA replication process); and there are increases in chromosome breakage and rearrangements compared to persons who do not have Bloom syndrome. Direct connections between the molecular processes in which BLM operates and the chromosomes themselves are under investigation. The relationships between molecular defects in Bloom syndrome cells, the chromosome mutations that accumulate in somatic cells (the cells of the body), and the many clinical features seen in Bloom syndrome are also areas of intense research.



Bloom syndrome has an autosomal recessive pattern of inheritance.

Bloom syndrome is inherited in an autosomal recessive fashion. Both parents must be carriers in order for a child to be affected. The carrier frequency in individuals of Eastern European Jewish (Ashkenazi Jewish) ancestry is about 1/100. If both parents are carriers, there is a one in four, or 25%, chance with each pregnancy for an affected child. Genetic counseling and genetic testing is recommended for families who may be carriers of Bloom syndrome. For families in which carrier status is known, prenatal testing is available using cytogenetic or molecular methods. Molecular DNA testing for the mutation that is common in the Ashkenazi Jewish population is also available.

Other mutations in Bloom Syndrome

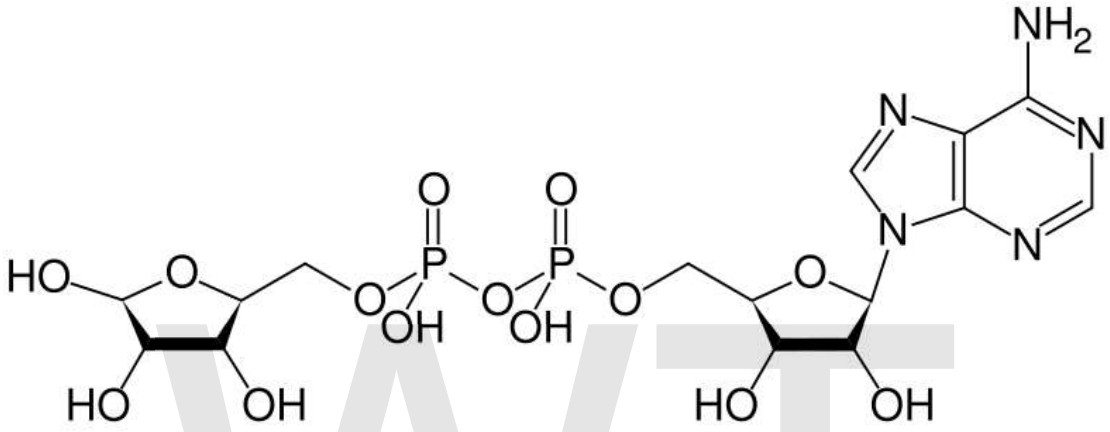
For the first time pyruvate kinase M2 enzyme was reported with two missense mutations, H391Y and K422R, found in cells from Bloom syndrome patients, prone to develop cancer. Results show that despite the presence of mutations in the inter-subunit contact domain, the K422R and H391Y mutant proteins maintained their homotetrameric structure, similar to the wild-type protein, but showed a loss of activity of 75 and 20%, respectively. Interestingly, H391Y showed a 6-fold increase in affinity for its substrate phosphoenolpyruvate and behaved like a non-allosteric protein with compromised cooperative binding. However, the affinity for phosphoenolpyruvate was lost significantly in K422R. Unlike K422R, H391Y showed enhanced thermal stability, stability over a range of pH values, a lesser effect of the allosteric inhibitor Phe, and resistance toward structural alteration upon binding of the activator (fructose 1,6-bisphosphate) and inhibitor (Phe). Both mutants showed a slight shift in the pH optimum from 7.4 to 7.0. The co-expression of homotetrameric wild type and mutant PKM2 in the cellular milieu resulting in the interaction between the two at the monomer level was substantiated further by in vitro experiments. The cross-monomer interaction significantly altered the oligomeric state of PKM2 by favoring dimerisation and heterotetramerization. In silico study provided an added support in showing that hetero-oligomerization was energetically favorable. The hetero-oligomeric populations of PKM2 showed altered activity and affinity, and their expression resulted in an increased growth rate of Escherichia coli as well as mammalian cells, along with an increased rate of polyploidy. These features are known to be essential to tumor progression.

PKM2(M2-PK); A potential multi-functional protein

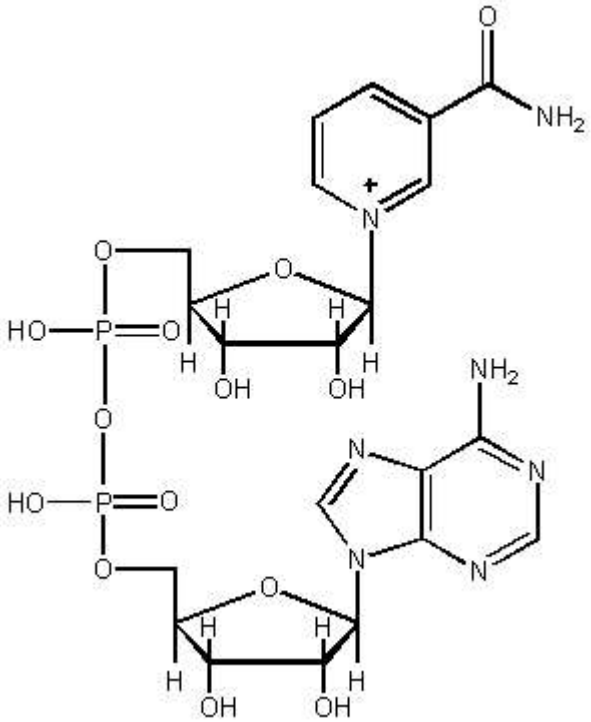
It is hypothesized that an aberrant PKM2 protein can justify other features of Bloom syndrome.

Chapter- 8

Poly ADP Ribose Polymerase



ADP ribose



Nicotinamide adenine dinucleotide

Poly (ADP-ribose) polymerase (**PARP**) is a protein involved in a number of cellular processes involving mainly DNA repair and programmed cell death.

Members of PARP family

The PARP family comprises 17 members (10 putative). They have all very different structures and functions in the cell.

- *PARP1*, *PARP2*, VPARP (*PARP4*), Tankyrase-1 and -2 (PARP-5a or *TNKS*, and PARP-5b or *TNKS2*) have a confirmed PARP activity.

- Others include *PARP3*, *PARP6*, *TIPARP* (or "PARP7"), *PARP8*, *PARP9*, *PARP10*, *PARP11*, *PARP12*, *PARP14*, *PARP15*, and *PARP16*.

PARP Structure

PARP is composed of four domains of interest: a DNA-binding domain, a caspase-cleaved domain domain(see below), an auto-modification domain, and a catalytic domain. The DNA-binding domain is composed of two zinc finger motifs. In the presence of damaged DNA (base pair-excised), the DNA-binding domain will bind the DNA and induce a conformational shift. It has been shown that this binding occurs independent of the other domains. This is integral in a programmed cell death model based on Caspase cleavage inhibition of PARP. The auto-modification domain is responsible for releasing the protein from the DNA after catalysis. Also, it plays an integral role in cleavage-induced inactivation.

Functions

PARP is found in the cell's nucleus, the main role is to detect and signal single strand DNA breaks (SSB) to the enzymatic machinery involved in the SSB repair. PARP activation is an immediate cellular response to metabolic, chemical, or radiation-induced DNA SSB damage. Once PARP detects a SSB it binds to the DNA, and, after a structural change, begin the synthesis of a poly(ADP-ribose)chain (PAR)as a signal for the other DNA repairing enzymes such as DNA ligase III (LigIII), DNA polymerase beta (polβ) and scaffolding proteins such as x-ray repair complementing gene 1 (XRCC1). After repairing, the PAR chains are degraded via PAR glycohydrolase (PARG). Interestingly, NAD+ is required as substrate for generating ADP-ribose monomers. The overactivation of PARP may deplete the stores of cellular NAD+ and induce a progressive ATP depletion, since glucose oxidation is inhibited, and necrotic cell death. In this regard, PARP is inactivated by caspase-3 cleavage (in a specific domain of the enzyme) during programmed cell death. PARP enzymes are essential in a number of cellular functions, including expression of inflammatory genes: PARP1 is required for the induction of ICAM-1 gene expression by smooth muscle cells, in response to TNF.

Activity

The catalytic domain is responsible for Poly (ADP-ribose) polymerization. This domain has a highly conserved motif that is common to all members of the PARP family. PAR polymer can reach lengths up to 200 bp before inducing apoptotic processes. The formation of PAR polymer is similar to the formation of DNA polymer from nucleoside triphosphates. Normal DNA synthesis requires that a pyrophosphate act as the leaving group, leaving a single phosphate group linking ribose sugars. PAR is synthesized using nicotinamide (NAM) as the leaving group. This leaves a pyrophosphate as the linking group between ribose sugars rather than single phosphate groups. This creates some special bulk to a PAR bridge, which may have an additional role in cell signaling.

Role in repairing DNA nicks

One important function of PARP is assisting in the repair of single-strand DNA nicks. It binds sites with single-strand breaks through its N-terminal zinc fingers and will recruit XRCC1, DNA ligase III, DNA polymerase beta, and a kinase to the nick. This is called *base excision repair* (BER). PARP-2 has been shown to oligomerize with PARP-1 and, therefore, is also implicated in BER. The oligomerization has also been shown to stimulate PARP catalytic activity. PARP-1 is also known for its role in transcription through remodeling of chromatin by PARylating histones and relaxing chromatin structure, thus allowing transcription complex to access genes.

Role of tankyrases

The tankyrases are PARPs that comprise ankyrin repeats, oligomerization domain (SAM), and a *PARP catalytic domain* (PCD). Tankyrases are also known as PARP-5a and PARP-5b. They were named for their interaction with the telomere-associated TRF1 proteins and ankyrin repeats. They may allow the removal of telomerase-inhibiting complexes from chromosome ends to allow for telomere maintenance. Through their SAM domain and ANKs they can oligomerize and interact with many other proteins, such as TRF1, TAB182 (*TNKS1BP1*), *GRB14*, IRAP, NuMa, EBNA-1, and Mcl-1. They have multiple roles in the cell, vesicular trafficking through its interaction in *GLUT4 vesicle* (GSVs) with *insulin-responsive amino peptidase* (IRAP). It also plays a role in spindle assembly through its interaction with *nuclear mitotic apparatus* (NuMa), therefore allowing bipolarity. In the absence of TNKs mitosis arrest is observed in pre-anaphase through Mad2 kinetochore checkpoint. TNKs can also PARsylate Mcl-1L and Mcl-1S and inhibit both their pro- and anti-apoptotic function. Relevance of this is not yet known.

Role in cell death

Upon DNA cleavage by enzymes involved in cell death (such as caspases), PARP can deplete the ATP of a cell in an attempt to repair the damaged DNA. ATP depletion in a cell leads to lysis and cell death. PARP also has the ability to directly induce apoptosis,

via the production of PAR, which stimulates mitochondria to release AIF. This mechanism appears to be caspase-independent.

Role in Epigenetic DNA modification

PARP-mediated post-translational modification of proteins such as CTCF can affect the amount of DNA methylation at CpG dinucleotides. This regulates the insulator features of CTCF can differentially mark the copy of DNA inherited from either the maternal or the paternal DNA through the process known as genomic imprinting. PARP has also been proposed to affect the amount of DNA methylation by directly binding to the DNA methyltransferase DNMT-1 after attaching poly ADP-ribose chains to itself after interaction with CTCF and affecting DNMT1's enzymatic activity.

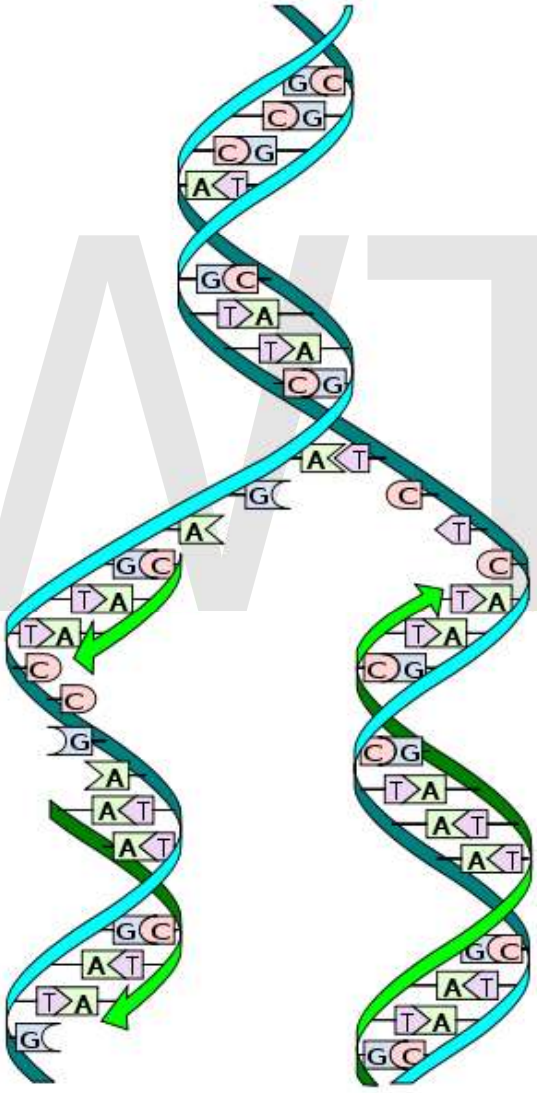
PARP Inactivation

PARP is inactivated by Caspase cleavage. It is believed that normal inactivation occurs in systems where DNA damage is extensive. In these cases, more energy would be invested in repairing damage than is feasible, so that energy is instead retrieved for other cells in the tissue through programmed cell death.

While in vitro cleavage by caspase occurs throughout the Caspase family, preliminary data suggest that Caspase-3 and Caspase-7 are responsible for in vivo cleavage. Cleavage occurs at Aspartic Acid 214 and Glycine 215, separating PARP into a 24kDA and 89kDA segment. The smaller moiety includes the zinc finger motif requisite in DNA binding. The 89 kDa fragment includes the auto-modification domain and catalytic domain. The putative mechanism of PCD activation via PARP inactivation relies on the separation of the DNA-binding region and the auto-modification domain. The DNA-binding region is capable of doing so independent of the rest of the protein, cleaved or not. It is unable, however, to dissociate without the auto-modification domain. In this way, the DNA-binding domain will attach to a damaged site and be unable to effect repair, as it no longer has the catalytic domain. The DNA-binding domain prevents other, non-cleaved PARP from accessing the damaged site and initiating repairs. This model suggests that this “sugar plug” can also begin the signal for apoptosis.

Chapter- 9

DNA Replication



DNA replication. The double helix is unwound and each strand acts as a template. Bases are matched to synthesize the new partner strands.

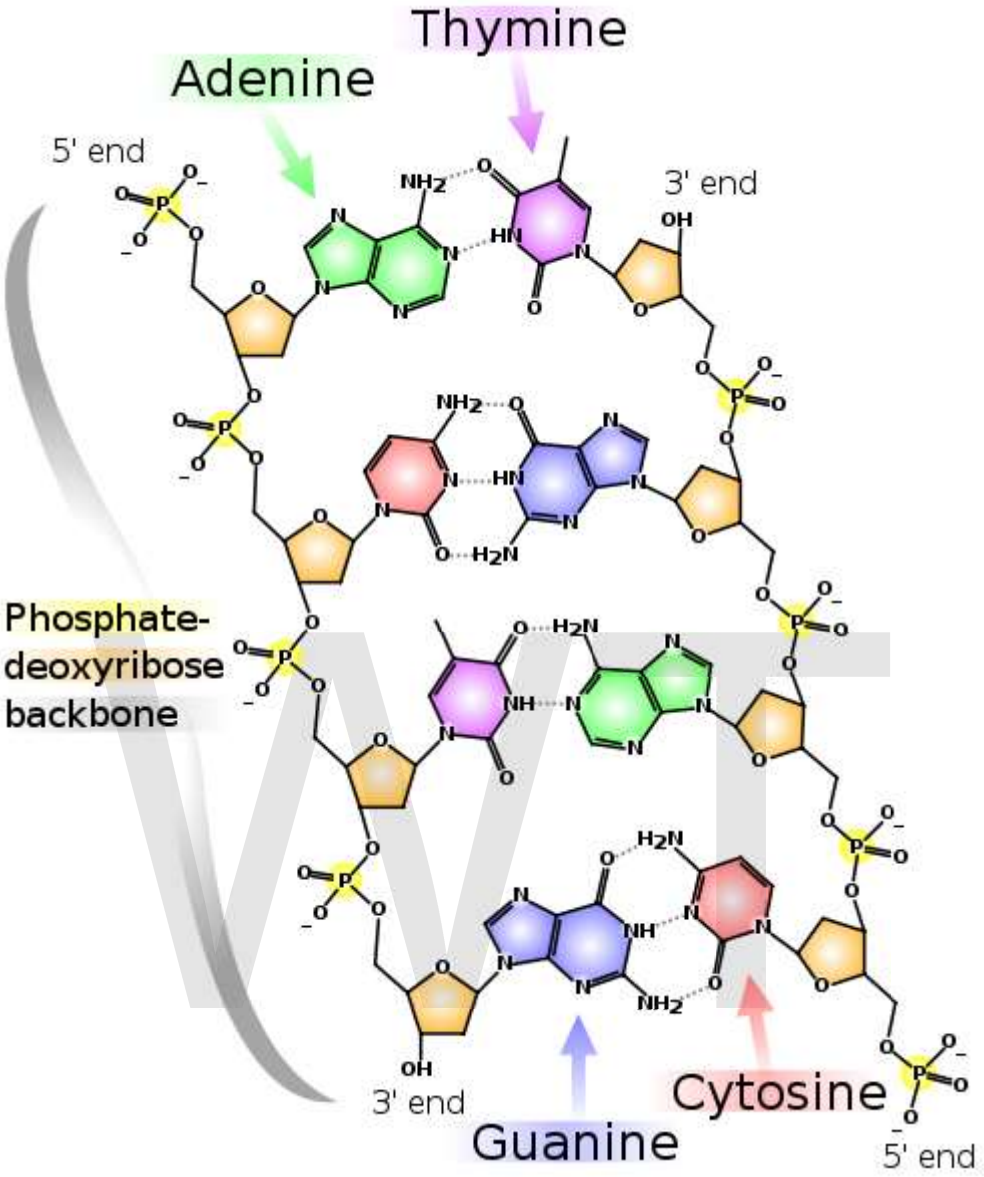
DNA replication, the basis for biological inheritance, is a fundamental process - that occurs in all living organisms - that copies their DNA. This process is *replication*, in that

each strand of the original double-stranded DNA molecule serves as template for the reproduction of the complementary strand. Therefore, following DNA replication, two identical DNA molecules have been produced from a single double-stranded DNA molecule. Cellular proofreading and error toe-checking mechanisms ensure near perfect fidelity for DNA replication.

In a cell, DNA replication begins at specific locations in the genome, called "origins". Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork. In addition to DNA polymerase, the enzyme that synthesizes the new DNA by adding nucleotides matched to the template strand, a number of other proteins are associated with the fork and assist in the initiation and continuation of DNA synthesis.

DNA replication can also be performed *in vitro* (outside a cell). DNA polymerases, isolated from cells, and artificial DNA primers are used to initiate DNA synthesis at known sequences in a template molecule. The polymerase chain reaction (PCR), a common laboratory technique, employs such artificial synthesis in a cyclic manner to amplify a specific target DNA fragment from a pool of DNA.

DNA structure



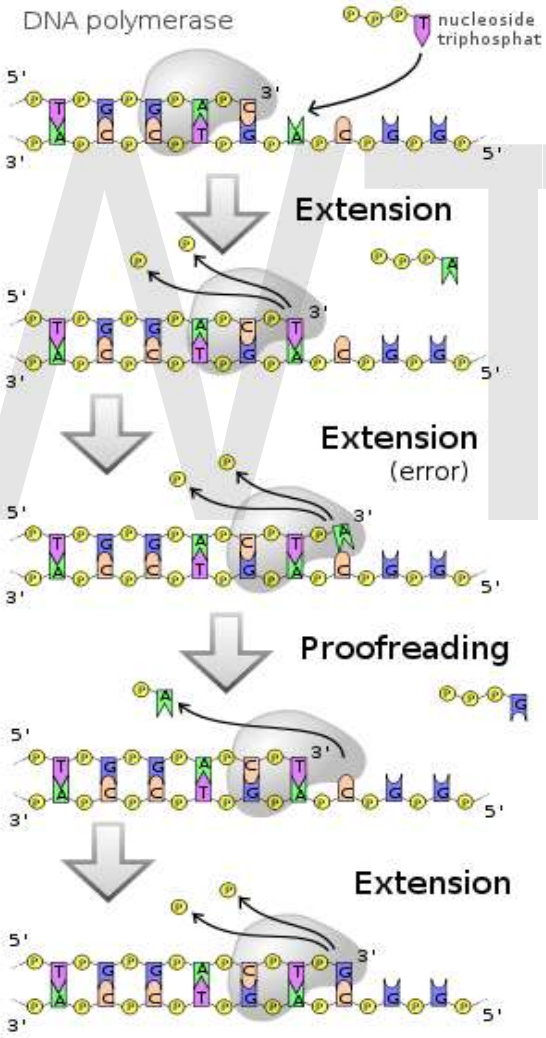
The chemical structure of DNA

DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types of nucleotides having the bases: adenine, cytosine, guanine, and thymine. A nucleotide is a mono-, di-, or triphosphate deoxyribonucleoside; that is, a deoxyribose sugar is attached to one, two, or three phosphates. Chemical interaction of these nucleotides forms phosphodiester linkages, creating the phosphate-deoxyribose backbone of the DNA double helix with the bases pointing inward. Nucleotides (bases) are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine and cytosine pairs with guanine.

DNA strands have a directionality, and the different ends of a single strand are called the "3' (three-prime) end" and the "5' (five-prime) end." These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. In addition to being complementary, the two strands of DNA are antiparallel: They are orientated in opposite directions. This directionality has consequences in DNA synthesis, because DNA polymerase can synthesize DNA in only one direction by adding nucleotides to the 3' end of a DNA strand.

The pairing of bases in DNA through hydrogen bonding means that the information contained within each strand is redundant. The nucleotides on a single strand can be used to reconstruct nucleotides on a newly synthesized partner strand.

DNA polymerase



DNA polymerases adds nucleotides to the 5' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues.

DNA polymerases are a family of enzymes that carry out all forms of DNA replication. A DNA polymerase can only extend an existing DNA strand paired with a template strand; it cannot begin the synthesis of a new strand. To begin synthesis of a new strand, a short fragment of DNA or RNA, called a primer, must be created and paired with the template strand before DNA polymerase can synthesize new DNA.

Once a primer pairs with DNA to be replicated, DNA polymerase synthesizes a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from two of the three total phosphates attached to each unincorporated base. (Free bases with their attached phosphate groups are called nucleoside triphosphates.) When a nucleotide is being added to a growing DNA strand, two of the phosphates are removed and the energy produced creates a phosphodiester (chemical) bond that attaches the remaining phosphate to the growing chain. The energetics of this process also help explain the directionality of synthesis - if DNA were synthesized in the 3' to 5' direction, the energy for the process would come from the 5' end of the growing strand rather than from free nucleotides.

In general, DNA polymerases are extremely accurate, making less than one error for every 10^7 nucleotides added. Even so, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a strand in order to correct mismatched bases. If the 5' nucleotide needs to be removed during proofreading, the triphosphate end is lost. Hence, the energy source that usually provides energy to add a new nucleotide is also lost.

DNA replication within the cell

Origins of replication

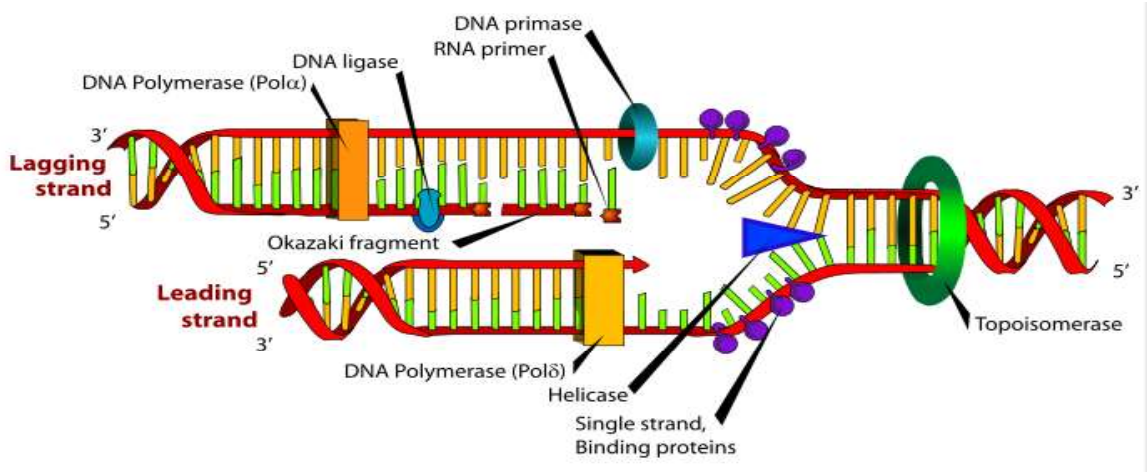
For a cell to divide, it must first replicate its DNA. This process is initiated at particular points within the DNA, known as "origins", which are targeted by proteins that separate the two strands and initiate DNA synthesis. Origins contain DNA sequences recognized by replication initiator proteins (e.g., dnaA in *E coli'* and the Origin Recognition Complex in yeast). These initiator proteins recruit other proteins to separate the two strands and initiate replication forks.

Initiator proteins recruit other proteins to separate the DNA strands at the origin, forming a bubble. Origins tend to be "AT-rich" (rich in adenine and thymine bases) to assist this process, because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair)—in general, strands rich in these nucleotides are easier to separate due the positive relationship between the number of hydrogen bonds and the difficulty of breaking these bonds. Once strands are separated, RNA primers are created on the template strands. To be more specific, the leading strand receives one RNA primer per active origin of replication while the lagging strand receives several; these several fragments of RNA primers found on the lagging strand of DNA are called Okazaki

fragments, named after their discoverer. DNA Polymerase extends the leading strand in one continuous motion and the lagging strand in a discontinuous motion (due to the Okazaki fragments). RNase removes the RNA fragments used to initiate replication by DNA Polymerase, and another DNA Polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming 2 replication forks. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

The replication fork



Many enzymes are involved in the DNA replication fork.

The replication fork is a structure that forms within the nucleus during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together. The resulting structure has two branching "prongs", each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; The templates may be properly referred to as the leading strand template and the lagging strand template.

Leading strand

The leading strand is the template strand of the DNA double helix so that the replication fork moves along it in the 3' to 5' direction. This allows the new strand synthesized complementary to it to be synthesized 5' to 3' in the same direction as the movement of the replication fork.

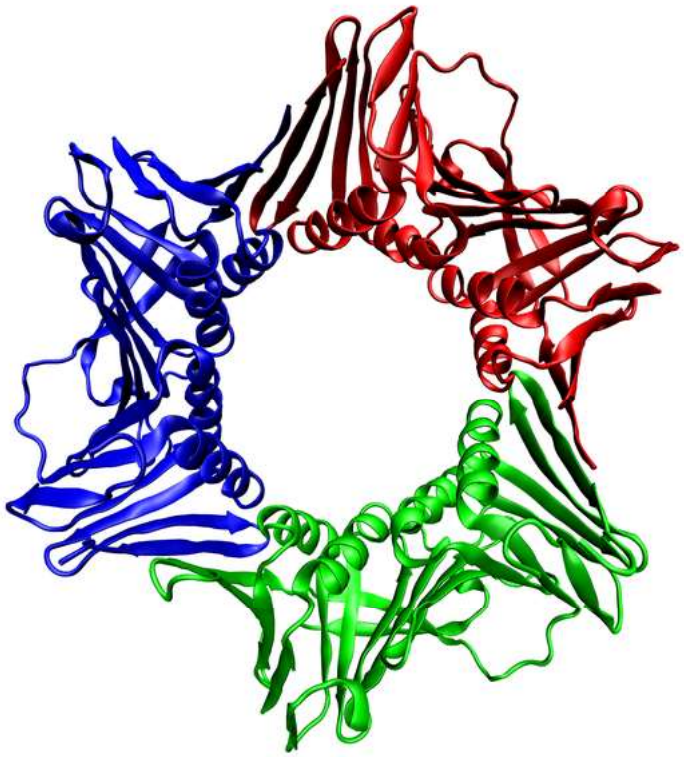
On the leading strand, a polymerase "reads" the DNA and adds nucleotides to it continuously. This polymerase is DNA polymerase III (DNA Pol III) in prokaryotes and presumably Pol ε in eukaryotes.

Lagging strand

The lagging strand is the strand of the template DNA double helix that is oriented so that the replication fork moves along it in a 5' to 3' manner. Because of its orientation, opposite to the working orientation of DNA polymerase III, which moves on a template in a 3' to 5' manner, replication of the lagging strand is more complicated than that of the leading strand.

On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to Pol α. DNA polymerase III or Pol δ lengthens the primed segments, forming Okazaki fragments. Primer removal in eukaryotes is also performed by Pol δ. In prokaryotes, DNA polymerase I "reads" the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5'-3' exonuclease activity of polymerase I [weaver, 2005], and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different kinds of nucleotides). DNA ligase joins the fragments together.

Dynamics at the replication fork



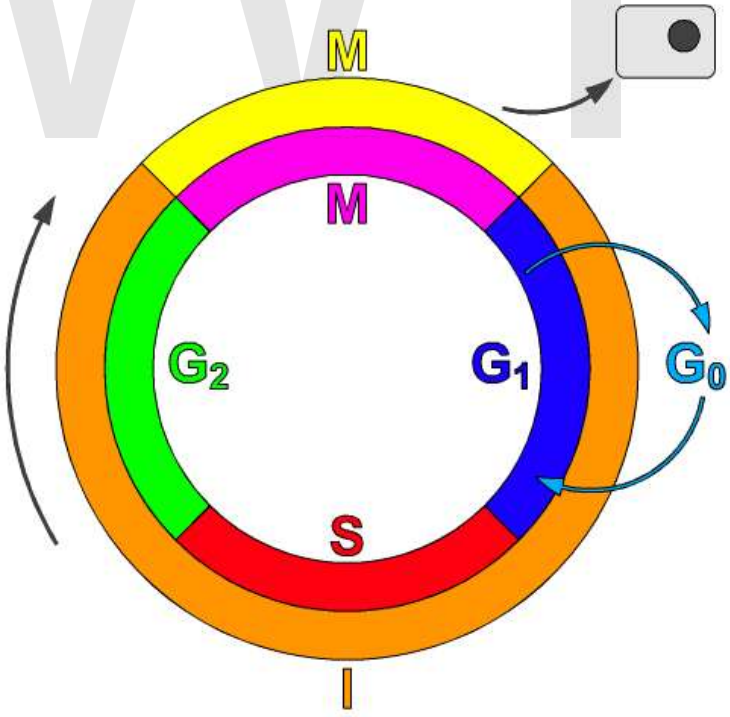
The assembled human DNA clamp, a trimer of the protein PCNA

As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead. This build-up would form a resistance that would eventually halt the progress of the replication fork. DNA topoisomerases are enzymes that solve these physical problems in the coiling of DNA. Topoisomerase I cuts a single backbone on the DNA, enabling the strands to swivel around each other to remove the build-up of twists. Topoisomerase II cuts both backbones, enabling one double-stranded DNA to pass through another, thereby removing knots and entanglements that can form within and between DNA molecules.

Bare single-stranded DNA has a tendency to fold back upon itself and form secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

Regulation of replication



The cell cycle of eukaryotic cells

Eukaryotes

Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication occurs during the S phase (**S**ynthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases.

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells that do not proceed through this checkpoint are quiescent in the "G0" stage and do not replicate their DNA.

Replication of chloroplast and mitochondrial genomes occurs independent of the cell cycle, through the process of D-loop replication.

Bacteria

Most bacteria do not go through a well-defined cell cycle and, instead, continuously copy their DNA; during rapid growth, this can result in the concurrent occurrences of multiple rounds of replication. Within *E coli*, the best-characterized bacteria, regulation of DNA replication can be achieved through several mechanisms, including: the hemimethylation and sequestering of the origin sequence, the ratio of ATP to ADP, and the levels of protein DnaA. These all control the process of initiator proteins binding to the origin sequences.

Because *E coli* methylates GATC DNA sequences, DNA synthesis results in hemimethylated sequences. This hemimethylated DNA is recognized by a protein (SeqA), which binds and sequesters the origin sequence; in addition, dnaA (required for initiation of replication) binds less well to hemimethylated DNA. As a result, newly replicated origins are prevented from immediately initiating another round of DNA replication.

ATP builds up when the cell is in a rich medium, triggering DNA replication once the cell has reached a specific size. ATP competes with ADP to bind to DnaA, and the DnaA-ATP complex is able to initiate replication. A certain number of DnaA proteins are also required for DNA replication — each time the origin is copied the number of binding sites for DnaA doubles, requiring the synthesis of more DnaA to enable another initiation of replication.

Termination of replication

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. *E coli* regulate this process through the use of termination sequences that, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result,

the replication forks are constrained to always meet within the termination region of the chromosome.

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular manner. Because eukaryotes have linear chromosomes, DNA replication is unable to synthesize to the very end of the chromosomes (telomeres), resulting in telomere shortening. This is a normal process in somatic cells — cells are able to divide only a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation. Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation.

Polymerase chain reaction

Researchers commonly replicate DNA *in vitro* using the polymerase chain reaction (PCR). PCR uses a pair of primers to span a target region in template DNA, and then polymerizes partner strands in each direction from these primers using a thermostable DNA polymerase. Repeating this process through multiple cycles produces amplification of the targeted DNA region. At the start of each cycle, the mixture of template and primers is heated, separating the newly synthesized molecule and template. Then, as the mixture cools, both of these become templates for annealing of new primers, and the polymerase extends from these. As a result, the number of copies of the target region doubles each round, increasing exponentially.

Chapter- 10

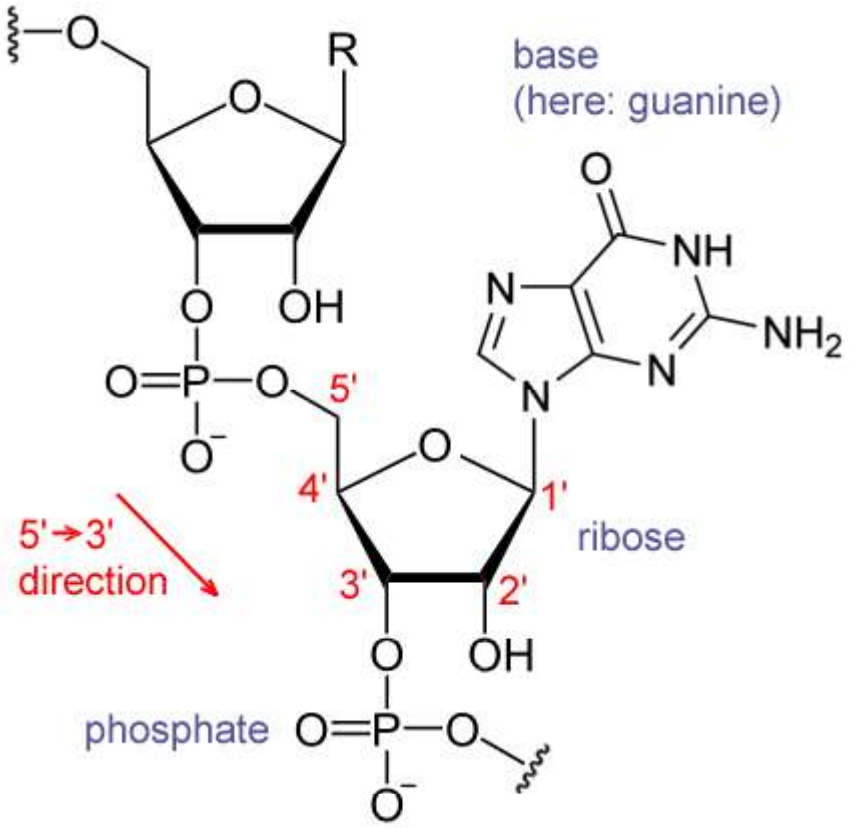
Nucleic Acid Sequence

The **sequence** or **primary structure of a nucleic acid** is the exact specification of its atomic composition and the chemical bonds connecting those atoms. As nucleic acids, e.g. DNA and RNA, are unbranched polymers, this is equivalent to specifying exact sequence of nucleotides that comprise the whole molecule. This sequence is written as a succession of letters representing a real or hypothetical DNA molecule or strand. By convention, the primary structure of a DNA or RNA molecule is reported from the 5' end to the 3' end.

The sequence has capacity to carry information. When used in reference to biological DNA, which carries the information which directs the functions of living beings, the term **genetic sequence** is often used. Sequences can be read from the biological raw material through DNA sequencing methods.

Primary structure is sometimes mistakenly termed *primary sequence*, but there is no such term, as well as no parallel concept of secondary or tertiary sequence.

Nucleotides



Chemical structure of RNA

Nucleic acids consist of a chain of linked units called nucleotides. Each nucleotide consists of three subunits: a phosphate group and a sugar (ribose in the case of RNA, deoxyribose in DNA) make up the backbone of the nucleic acid strand, and attached to the sugar is one of a set of nucleobases. The nucleobases are important in base pairing of strands to form higher-level secondary and tertiary structure such as the famed double helix.

The possible letters are *A*, *C*, *G*, and *T*, representing the four nucleotide bases of a DNA strand — adenine, cytosine, guanine, thymine — covalently linked to a phosphodiester backbone. In the typical case, the sequences are printed abutting one another without gaps, as in the sequence AAAGTCTGAC, read left to right in the 5' to 3' direction. With regards to transcription, a sequence is on the coding strand if it has the same order as the transcribed RNA.

One sequence can be complementary to another sequence, meaning that they have the base on each position is the complementary (i.e. A to T, C to G) and in the reverse order. For example, the complementary sequence to TTAC is GTAA. If one strand of the double-stranded DNA is considered the sense strand, then the other strand, considered the antisense strand, will have the complementary sequence to the sense strand.

Notation

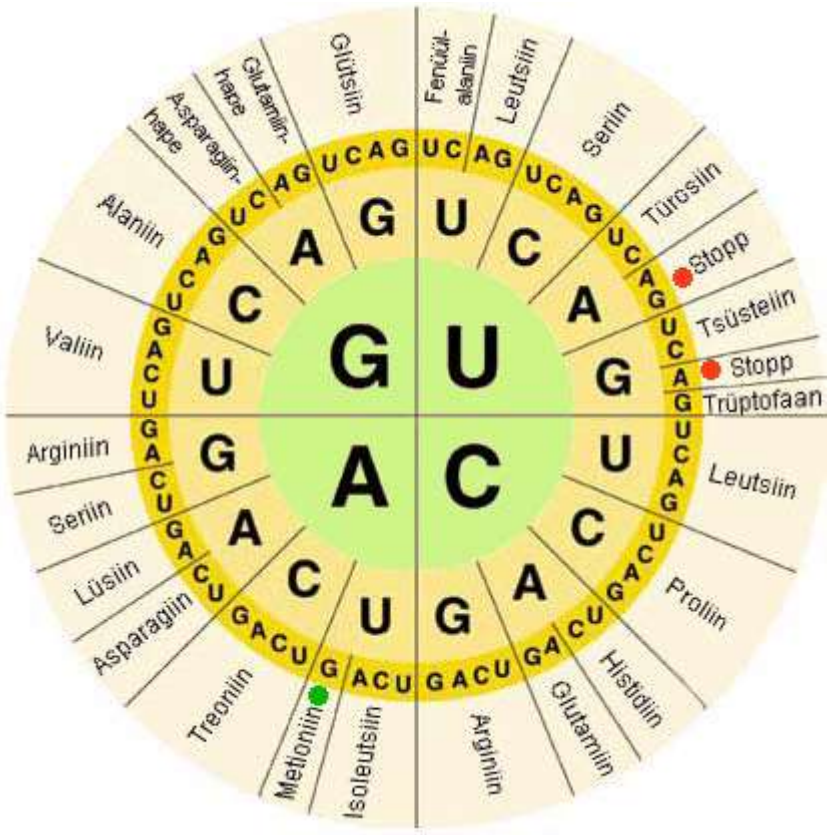
While A, T, C, and G represent a particular nucleotide at a position, there are also letters that represent ambiguity. Of all the molecules sampled, there is more than one kind of nucleotide at that position. The rules of the International Union of Pure and Applied Chemistry (IUPAC) are as follows:

- **A** = adenine
- **C** = cytosine
- **G** = guanine
- **T** = thymine
- **R** = G A (purine)
- **Y** = T C (pyrimidine)
- **K** = G T (keto)
- **M** = A C (amino)
- **S** = G C (strong bonds)
- **W** = A T (weak bonds)
- **B** = G T C (all but A)
- **D** = G A T (all but C)
- **H** = A C T (all but G)
- **V** = G C A (all but T)
- **N** = A G C T (any)

These symbols are also valid for RNA, except with U (uracil) replacing T (thymine).

Apart from adenine (A), cytosine (C), guanine (G), thymine (T) and uracil (U), DNA and RNA also contain bases that have been modified after the nucleic acid chain has been formed. In DNA, the most common modified base is 5-methylcytidine (m5C). In RNA, there are many modified bases, including pseudouridine (Ψ), dihydrouridine (D), inosine (I), ribothymidine (rT) and 7-methylguanosine (m7G). Hypoxanthine and xanthine are two of the many bases created through mutagen presence, both of them through deamination (replacement of the amine-group with a carbonyl-group). Hypoxanthine is produced from adenine, xanthine from guanine. Similarly, deamination of cytosine results in uracil.

Biological significance

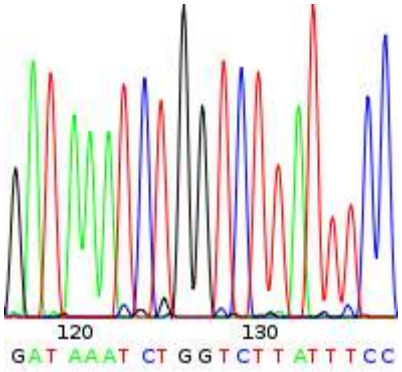


A depiction of the genetic code, by which the information contained in nucleic acids are translated into amino acid sequences in proteins.

In biological systems, nucleic acids contain information which is used by a living cell to construct specific proteins. The sequence of nucleobases on a nucleic acid strand is translated by cell machinery into a sequence of amino acids making up a protein strand. Each group of three bases, called a codon, corresponds to a single amino acid, and there is a specific genetic code by which each possible combination of three bases corresponds to a specific amino acid.

The central dogma of molecular biology outlines the mechanism by which proteins are constructed using information contained in nucleic acids. DNA is transcribed into mRNA molecules, which travels to the ribosome where the mRNA is used as a template for the construction of the protein strand. Since nucleic acids can bind to molecules with complementary sequences, there is a distinction between "sense" sequences which code for proteins, and the complementary "antisense" sequence which is by itself nonfunctional, but can bind to the sense strand.

Sequence determination



Electropherogram printout from automated sequencer for determining part of a DNA sequence

DNA sequencing is the process of determining the nucleotide sequence of a given DNA fragment. The sequence of DNA encodes the necessary information for living things to survive and reproduce. Determining the sequence is therefore useful in fundamental research into why and how organisms live, as well as in applied subjects. Because of the key nature of DNA to living things, knowledge of DNA sequence may come in useful in practically any biological research. For example, in medicine it can be used to identify, diagnose and potentially develop treatments for genetic diseases. Similarly, research into pathogens may lead to treatments for contagious diseases. Biotechnology is a burgeoning discipline, with the potential for many useful products and services.

RNA is not sequenced directly. Instead, it is copied to a DNA by reverse transcriptase, and this DNA is then sequenced.

Current sequencing methods rely on the discriminatory ability of DNA polymerases, and can therefore only distinguish four bases. An inosine (created from adenosine during RNA editing) will be read as a G, and 5-methyl-cytosine (created from cytosine by DNA methylation) will be read as a C. It is also currently difficult to sequence small amounts of DNA, as the signal will be too weak to measure. This is overcome by PCR amplification.

Digital format

Genetic sequence in digital format

Once a nucleic acid sequence has been obtained from an organism, it is stored *in silico* in digital format. Digital genetic sequences may be stored in sequence databases, be analyzed, be digitally altered and/or be used as templates for creating new actual DNA using artificial gene synthesis.

Sequence analysis

Digital genetic sequences may be analyzed using the tools of bioinformatics to attempt to determine its function.

Genetic testing

The DNA in an organism's genome can be analyzed to diagnose vulnerabilities to inherited diseases, and can also be used to determine a child's paternity (genetic father) or a person's ancestry. Normally, every person carries two copies of every gene, one inherited from their mother, one inherited from their father. The human genome is believed to contain around 20,000 - 25,000 genes. In addition to studying chromosomes to the level of individual genes, genetic testing in a broader sense includes biochemical tests for the possible presence of genetic diseases, or mutant forms of genes associated with increased risk of developing genetic disorders.

Genetic testing identifies changes in chromosomes, genes, or proteins. Most of the time, testing is used to find changes that are associated with inherited disorders. The results of a genetic test can confirm or rule out a suspected genetic condition or help determine a person's chance of developing or passing on a genetic disorder. Several hundred genetic tests are currently in use, and more are being developed.

Sequence alignment

In bioinformatics, a **sequence alignment** is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. If two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps as indels (that is, insertion or deletion mutations) introduced in one or both lineages in the time since they diverged from one another. In sequence alignments of proteins, the degree of similarity between amino acids occupying a particular position in the sequence can be interpreted as a rough measure of how conserved a particular region or sequence motif is among lineages. The absence of substitutions, or the presence of only very conservative substitutions (that is, the substitution of amino acids whose side chains have similar biochemical properties) in a particular region of the sequence, suggest that this region has structural or functional importance. Although DNA and RNA nucleotide bases are more similar to each other than are amino acids, the conservation of base pairs can indicate a similar functional or structural role.

Computational phylogenetics makes extensive use of sequence alignments in the construction and interpretation of phylogenetic trees, which are used to classify the evolutionary relationships between homologous genes represented in the genomes of divergent species. The degree to which sequences in a query set differ is qualitatively related to the sequences' evolutionary distance from one another. Roughly speaking, high sequence identity suggests that the sequences in question have a comparatively young most recent common ancestor, while low identity suggests that the divergence is more ancient. This approximation, which reflects the "molecular clock" hypothesis that a roughly constant rate of evolutionary change can be used to extrapolate the elapsed time since two genes first diverged (that is, the coalescence time), assumes that the effects of mutation and selection are constant across sequence lineages. Therefore it does not account for possible difference among organisms or species in the rates of DNA repair or the possible functional conservation of specific regions in a sequence. (In the case of nucleotide sequences, the molecular clock hypothesis in its most basic form also discounts the difference in acceptance rates between silent mutations that do not alter the meaning of a given codon and other mutations that result in a different amino acid being incorporated into the protein.) More statistically accurate methods allow the evolutionary rate on each branch of the phylogenetic tree to vary, thus producing better estimates of coalescence times for genes.

Sequence motifs

Frequently the primary structure encodes motifs that are of functional importance. Some examples of sequence motifs are: the C/D and H/ACA boxes of snoRNAs, Sm binding site found in spliceosomal RNAs such as U1, U2, U4, U5, U6, U12 and U3, the Shine-Dalgarno sequence, the Kozak consensus sequence and the RNA polymerase III terminator.

Chapter- 11

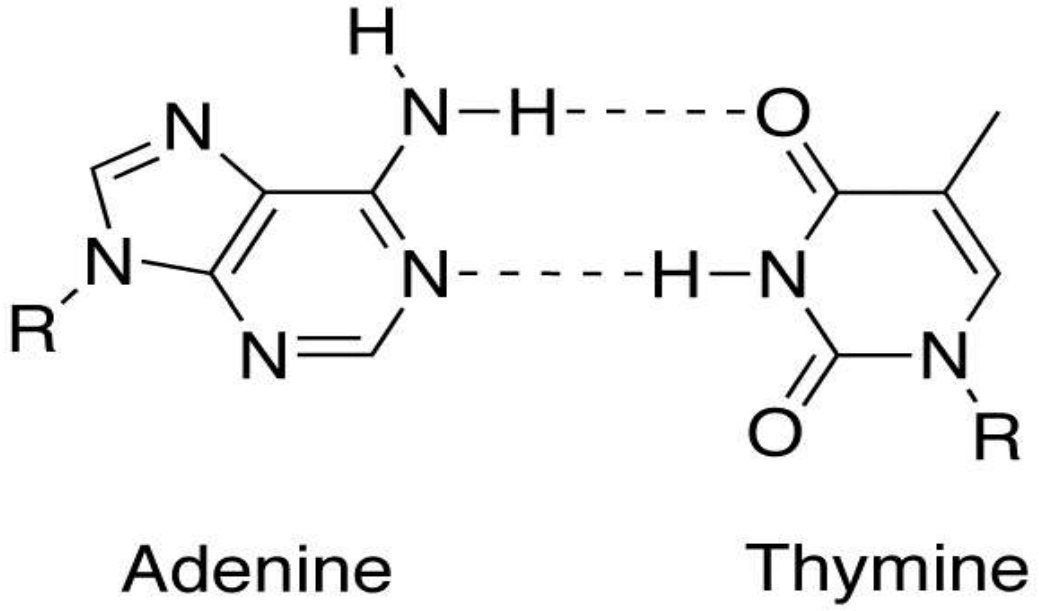
Nucleic Acid Secondary Structure

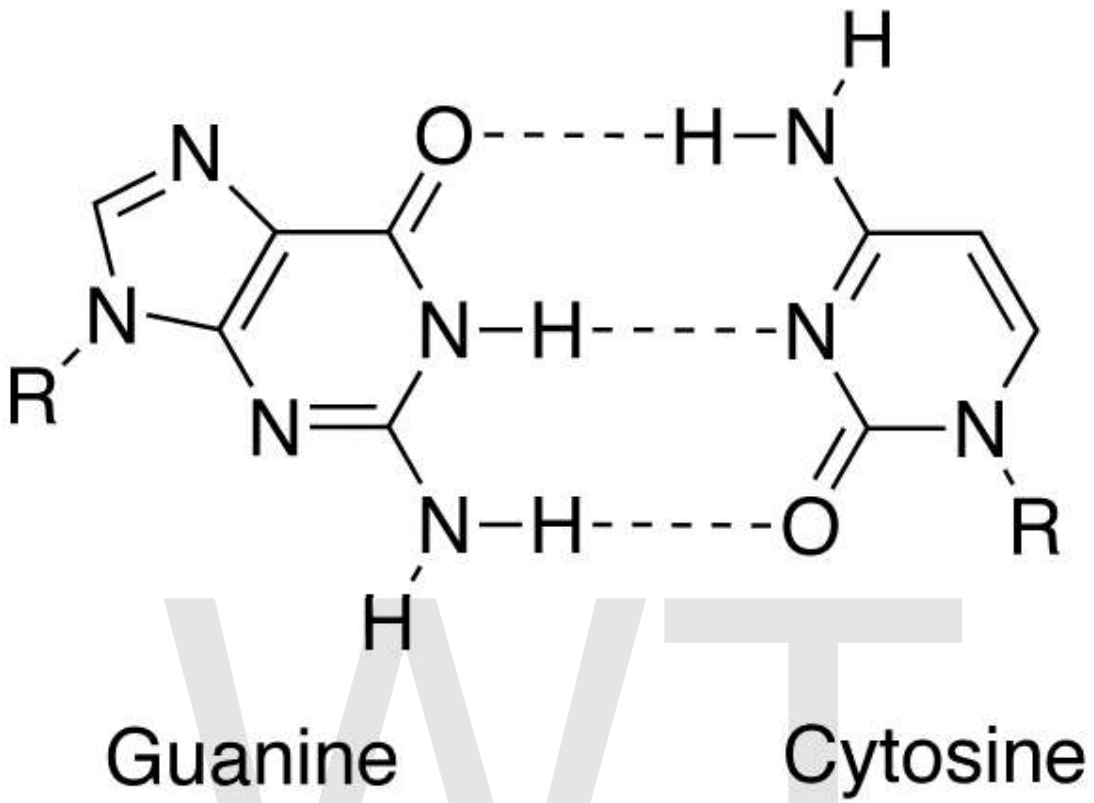
The **secondary structure of a nucleic acid molecule** refers to the basepairing interactions within a single molecule or set of interacting molecules, and can be represented as a list of bases which are paired in a nucleic acid molecule. The secondary structures of biological DNA's and RNA's tend to be different: biological DNA mostly exists as fully base paired double helices, while biological RNA is single stranded and often forms complicated base-pairing interactions due to its increased ability to form hydrogen bonds stemming from the extra hydroxyl group in the ribose sugar.

In a non-biological context, secondary structure is a vital consideration in the rational design of nucleic acid structures for DNA nanotechnology and DNA computing, since the pattern of basepairing ultimately determines the overall structure of the molecules.

Fundamental concepts

Base pairing





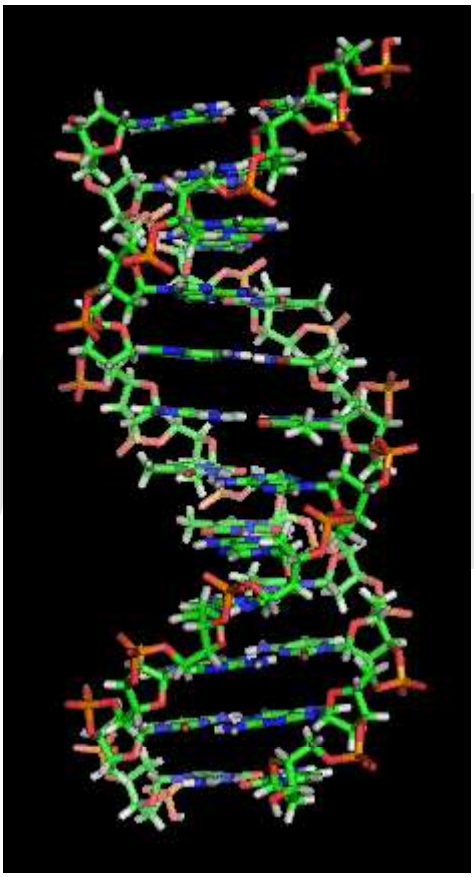
Top, an AT base pair demonstrating two intermolecular hydrogen bonds; *bottom*, a GC base pair demonstrating three intermolecular hydrogen bonds.

In molecular biology, two nucleotides on opposite complementary DNA or RNA strands that are connected via hydrogen bonds are called a base pair (often abbreviated bp). In the canonical Watson-Crick base pairing, adenine (A) forms a base pair with thymine (T), as and guanine (G) with cytosine (C) in DNA. In RNA, thymine is replaced by uracil (U). Alternate hydrogen bonding patterns, such as the wobble base pair and Hoogsteen base pair, also occur—particularly in RNA—giving rise to complex and functional tertiary structures. Importantly, pairing is the mechanism by which codons on messenger RNA molecules are recognized by anticodons on transfer RNA during protein translation. Some DNA- or RNA-binding enzymes can recognize specific base pairing patterns that identify particular regulatory regions of genes.

Hydrogen bonding is the chemical mechanism that underlies the base-pairing rules described above. Appropriate geometrical correspondence of hydrogen bond donors and acceptors allows only the "right" pairs to form stably. DNA with high GC-content is more stable than DNA with low GC-content, but contrary to popular belief, the hydrogen bonds do not stabilize the DNA significantly and stabilization is mainly due to stacking interactions.

The larger nucleobases, adenine and guanine, are members of a class of doubly-ringed chemical structures called purines; the smaller nucleobases, cytosine and thymine (and uracil), are members of a class of singly-ringed chemical structures called pyrimidines. Purines are only complementary with pyrimidines: pyrimidine-pyrimidine pairings are energetically unfavorable because the molecules are too far apart for hydrogen bonding to be established; purine-purine pairings are energetically unfavorable because the molecules are too close, leading to overlap repulsion. The only other possible pairings are GT and AC; these pairings are mismatches because the pattern of hydrogen donors and acceptors do not correspond. The GU wobble base pair, with two hydrogen bonds, does occur fairly often in RNA.

Nucleic acid hybridization



Two complementary regions of nucleic acid molcules will bind and from a double helical structure held together by base pairs.

Melting stability of base steps (B DNA)

Step	Melting ΔG /Kcal mol^{-1}
T A	-0.12
T G or **C A**	-0.78
C G	-1.44
A G or **C T**	-1.29
A A or **T T**	-1.04
A T	-1.27
G A or **T C**	-1.66
C C or **G G**	-1.97
A C or **G T**	-2.04
G C	-2.70

Hybridization is the process of complementary base pairs binding to form a double helix. Melting is the process by which the interactions between the strands of the double helix are broken, separating the two nucleic acid strands. These bonds are weak, easily separated by gentle heating, enzymes, or physical force. Melting occurs preferentially at certain points in the nucleic acid. **T** and **A** rich sequences are more easily melted than **C** and **G** rich regions. Particular base steps are also susceptible to DNA melting, particularly **T A** and **T G** base steps. These mechanical features are reflected by the use of sequences such as **TATAA** at the start of many genes to assist RNA polymerase in melting the DNA for transcription.

Strand separation by gentle heating, as used in PCR, is simple providing the molecules have fewer than about 10,000 base pairs (10 kilobase pairs, or 10 kbp). The intertwining of the DNA strands makes long segments difficult to separate. The cell avoids this problem by allowing its DNA-melting enzymes (helicases) to work concurrently with topoisomerases, which can chemically cleave the phosphate backbone of one of the strands so that it can swivel around the other. Helicases unwind the strands to facilitate the advance of sequence-reading enzymes such as DNA polymerase.

Secondary structure motifs

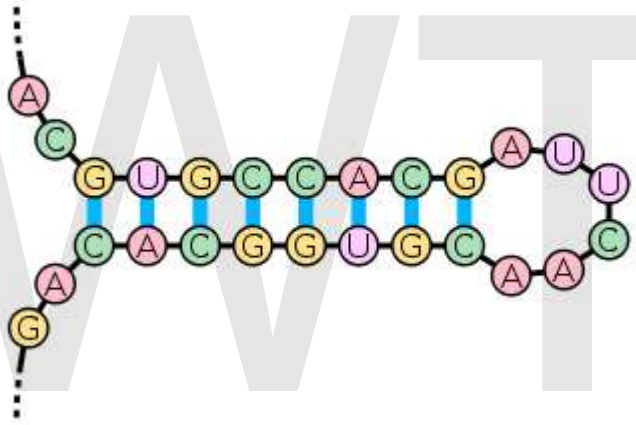
Nucleic acid secondary structure is generally divided into helices (contiguous base pairs), and various kinds of loops (unpaired nucleotides surrounded by helices). Frequently these elements, or combinations of them, can be further classified, for example, tetraloops, pseudoknots, and stem-loops.

Double helix

The double helix is an important tertiary structure in nuleic acid molecules which is intimately connected with the molecule's secondary structure. A double helix is formed by regions of many consecutive base pairs.

The DNA double helix is a right-handed spiral polymer of nucleic acids, held together by nucleotides which base pair together. A single turn of the helix constitutes about ten nucleotides, and contains a major groove and minor groove, the major groove being wider than the minor groove. Given the difference in widths of the major groove and minor groove, many proteins which bind to DNA do so through the wider major groove. Many double-helical forms are possible; for DNA the three biologically relevant forms are A-DNA, B-DNA, and Z-DNA, while RNA double helices have structures similar to the A form of DNA.

Stem-loop structures

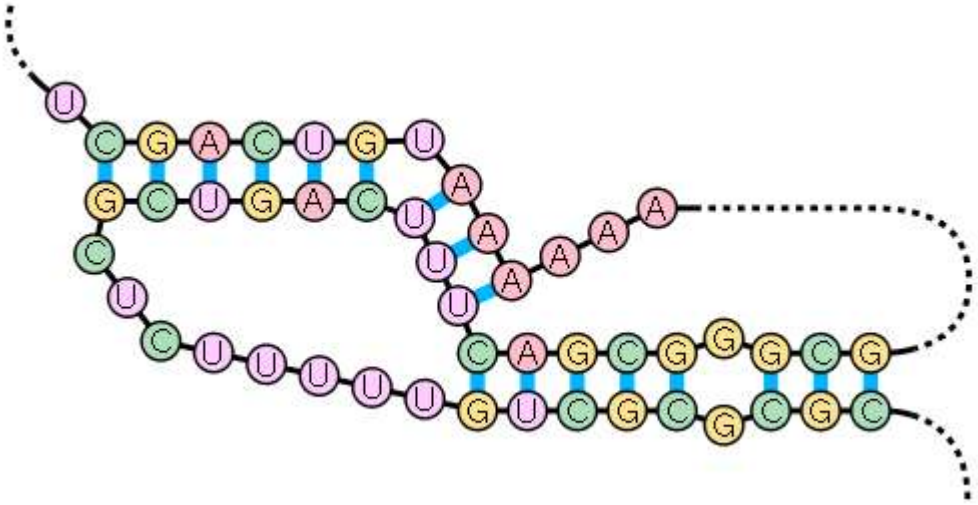


An example of an RNA stem-loop secondary structure

The secondary structure of nucleic acid molecules can often be uniquely decomposed into stems and loops. The stem-loop structure in which a base-paired helix ends in a short unpaired loop is extremely common and is a building block for larger structural motifs such as cloverleaf structures, which are four-helix junctions such as those found in transfer RNA. Internal loops (a short series of unpaired bases in a longer paired helix) and bulges (regions in which one strand of a helix has "extra" inserted bases with no counterparts in the opposite strand) are also frequent.

There are many secondary structure elements of functional importance to biological RNA's; some famous examples are the Rho-independent terminator stem-loops and the tRNA cloverleaf. There is a minor industry of researchers attempting to determine the secondary structure of RNA molecules. Approaches include both experimental and computational methods.

Pseudoknots



This example of a naturally occurring pseudoknot is found in the RNA component of human telomerase. Sequence from.

A pseudoknot is an RNA secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem. Pseudoknots fold into knot-shaped three-dimensional conformations but are not true topological knots. The base pairing in pseudoknots is not well nested; that is, base pairs occur that "overlap" one another in sequence position. This makes the presence of general pseudoknots in RNA sequences impossible to predict by the standard method of dynamic programming, which uses a recursive scoring system to identify paired stems and consequently cannot detect non-nested base pairs with the most common algorithms. Limited subclasses of pseudoknots can be predicted using dynamic programs described in. Newer structure prediction techniques such as stochastic context-free grammars also do not take pseudoknots into account.

Several important biological processes rely on RNA molecules that form pseudoknots. For example, the RNA component of human telomerase contains a pseudoknot that is critical for activity. Though DNA can also form pseudoknots, they are generally not present in biological DNA.

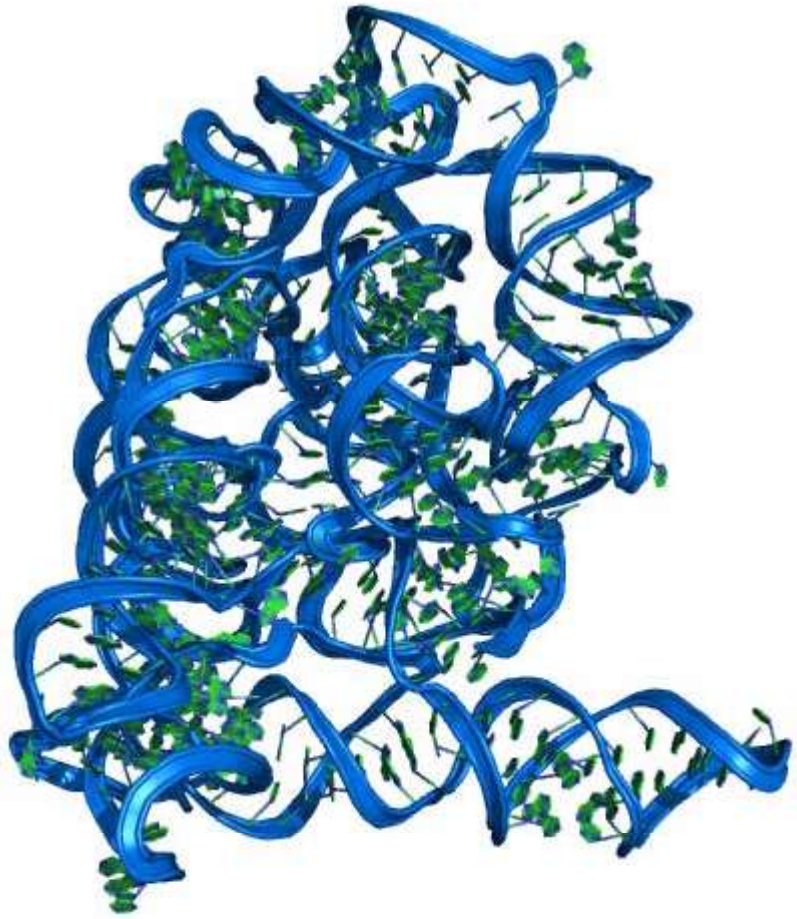
Secondary structure prediction

One application of bioinformatics uses predicted RNA secondary structures in searching a genome for noncoding but functional forms of RNA. For example, microRNAs have canonical long stem-loop structures interrupted by small internal loops. A general method of calculating probable RNA secondary structure is dynamic programming, although this has the disadvantage that it cannot detect pseudoknots or other cases in which base pairs are not fully nested. More general methods are based on stochastic context-free grammars. A web server that implements a type of dynamic programming is Mfold.

For many RNA molecules, the secondary structure is highly important to the correct function of the RNA — often more so than the actual sequence. This fact aids in the analysis of non-coding RNA sometimes termed "RNA genes". RNA secondary structure can be predicted with some accuracy by computer and many bioinformatics applications use some notion of secondary structure in analysis of RNA.

Chapter- 12

Nucleic Acid Tertiary Structure

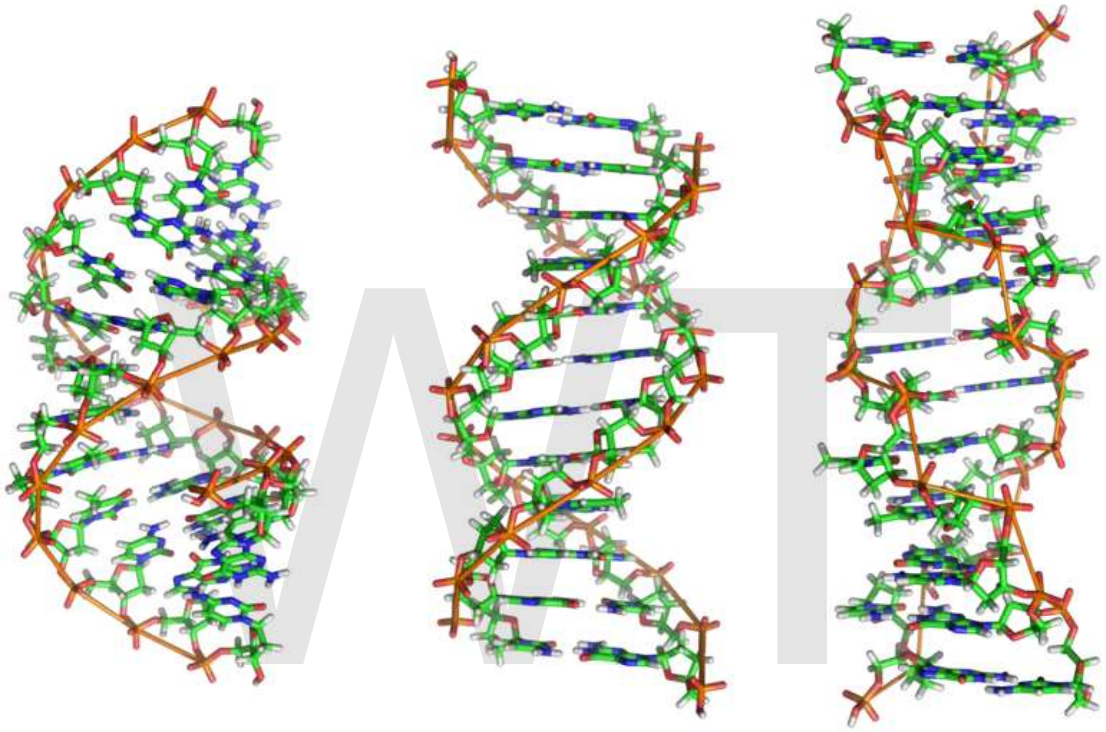


Example of a large catalytic RNA. The self-splicing group II intron from *Oceanobacillus iheyensis*.

The **tertiary structure of a nucleic acid** is its precise three-dimensional structure, as defined by the atomic coordinates. RNA and DNA molecules are capable of diverse functions ranging from molecular recognition to catalysis. Such functions require a

precise three-dimensional tertiary structure. While such structures are diverse and seemingly complex, they are composed of recurring, easily recognizable tertiary structure motifs that serve as molecular building blocks. Some of the most common motifs for RNA and DNA tertiary structure are described below, but it is important to remember that this information is based on a limited number of solved structures. Many more tertiary structural motifs will be revealed as new RNA and DNA molecules are structurally characterized.

Helical structures



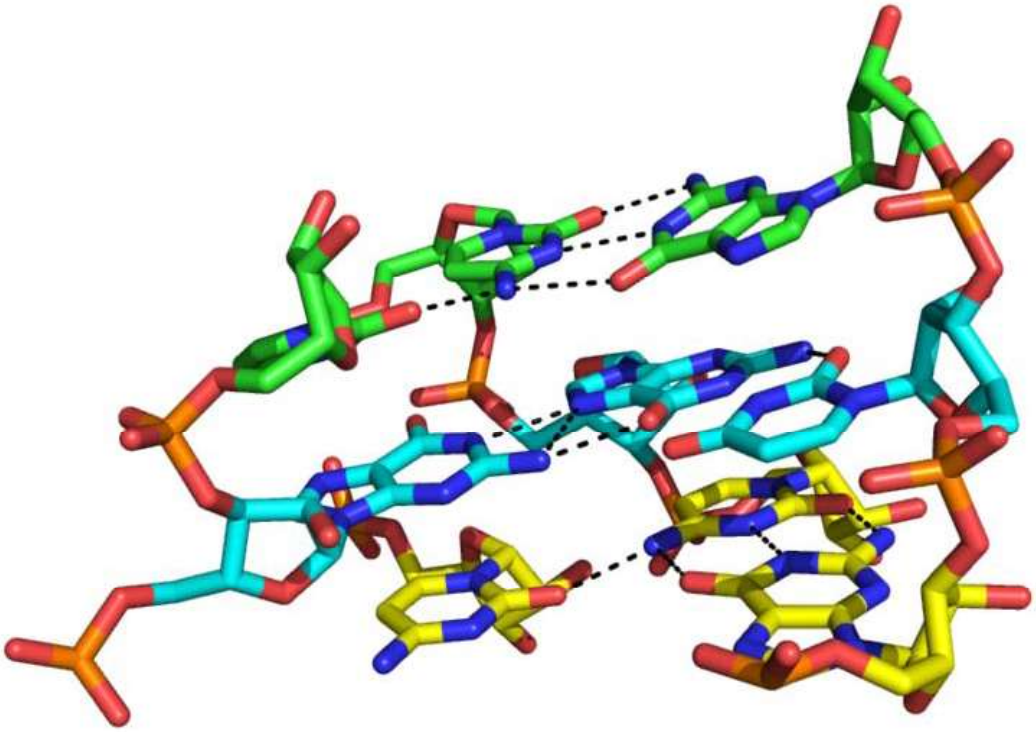
The structures of the A-, B-, and Z-DNA double helix structures

Double helix

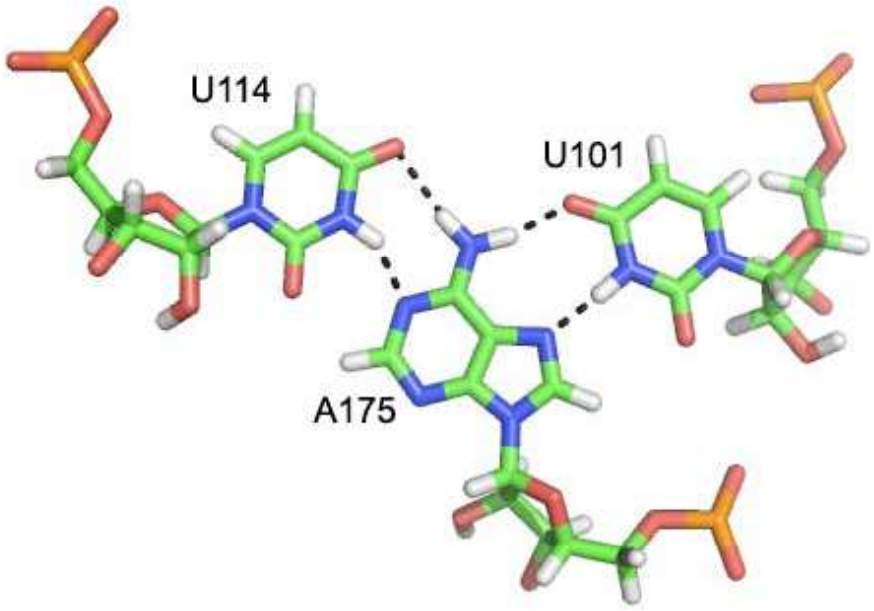
The double helix is the dominant tertiary structure for biological DNA, and is also a possible structure for RNA. Three DNA conformations are believed to be found in nature, A-DNA, B-DNA, and Z-DNA. The "B" form described by James D. Watson and Francis Crick is believed to predominate in cells. James D. Watson and Francis Crick described this structure as a double helix with a radius of 10 Å and pitch of 34 Å, making one complete turn about its axis every 10 bp of sequence. The double helix makes one complete turn about its axis every 10.4-10.5 base pairs in solution. This frequency of twist (known as the helical *pitch*) depends largely on stacking forces that each base exerts on its neighbours in the chain. Double-helical RNA adopts a conformation similar to the A-form structure.

Other conformations are possible; in fact, only the letters F, Q, U, V, and Y are now available to describe any new DNA structure that may appear in the future. However, most of these forms have been created synthetically and have not been observed in naturally occurring biological systems.

RNA triplexes



Major groove triples in the group II intron in *Oceanobacillus Iheyensis*. Each stacked layer is formed by one triplex with a different color scheme. Hydrogen bonds between triplexes are shown in black dashed lines. "N" atoms are colored in blue and "O" atoms in red. From top to bottom, the residues on the left side are G288, C289, and C377.



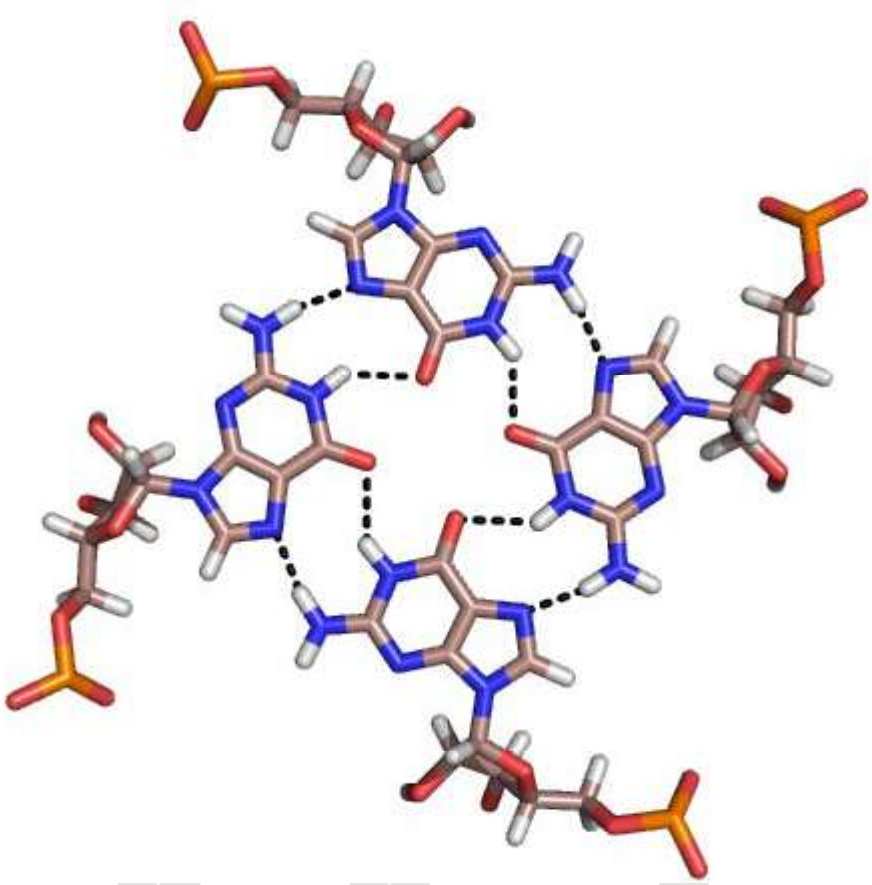
Close-up rendering of the U114:A175-U101 major groove (Hoogsteen base) triplex formed within the wild type pseudoknot of Human Telomerase RNA. Hydrogen bonds are shown in black dashed lines. "N" atoms are colored in blue and "o" atoms in red.

Major and minor groove triplexes

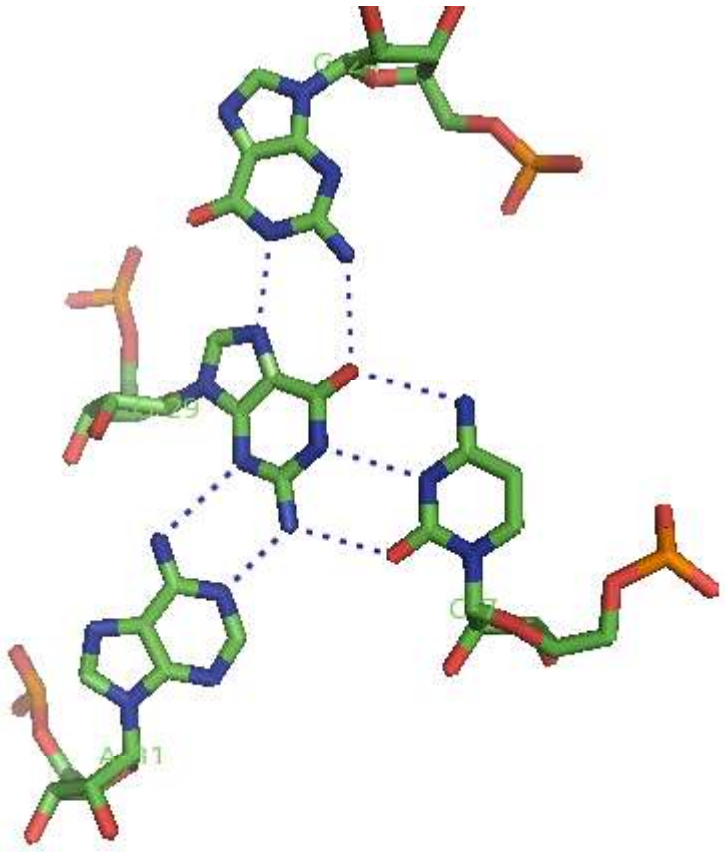
The minor groove triple is a ubiquitous RNA structural motif. Because interactions with the minor groove are often mediated by the 2'-OH of the ribose sugar, this RNA motif looks very different from its DNA equivalent. The most common example of a minor loop triple is the A-minor motif, or the insertion of adenosine bases into the minor groove (see above). However, this motif is not restricted to adenosines, as other nucleobases have also been observed to interact with the RNA minor groove.

The minor groove presents a near-perfect complement for an inserted base. This allows for optimal van der Waals contacts, extensive hydrogen bonding and hydrophobic surface burial, and creates a highly energetically favorable interaction. Because minor groove triples are capable of stably packing a free loop and helix, they are key elements in the structure of large ribonucleotides, including the group I intron, the group II intron, and the ribosome.

Quadruplexes



Above: Typical Ring Structure of a Hoogsteen paired G-quartet



Above: **Quadruplex** seen in crystal structure of Malachite Green RNA aptamer. G29 involved in major groove, minor groove, and Watson-Crick hydrogen-bonding with three other bases.

Although the major groove of standard A-form RNA is fairly narrow and therefore less available for triplex interaction than the minor groove, major groove triplex interactions can be observed in several RNA structures. These structures consist several combinations of base pair and Hoogsteen interactions. For example, the GGC triplex (GGC amino(N-2)-N-7, imino-carbonyl, carbonyl-amino(N-4); Watson-Crick) observed in the 50S ribosome, composed of a Watson-Crick type G-C pair and an incoming G which forms a pseudo-Hoogsteen network of hydrogen bonding interactions between both bases involved in the canonical pairing. Other notable examples of major groove triplexes include (i) the catalytic core of the group II intron shown in the figure at left (ii) a catalytically essential triple helix observed in human telomerase RNA and (iii) the SAM-II riboswitch.

Triple-stranded DNA is also possible from Hoogsteen or reversed Hoogsteen hydrogen bonds in the major groove of B-form DNA.

Quadruplexes

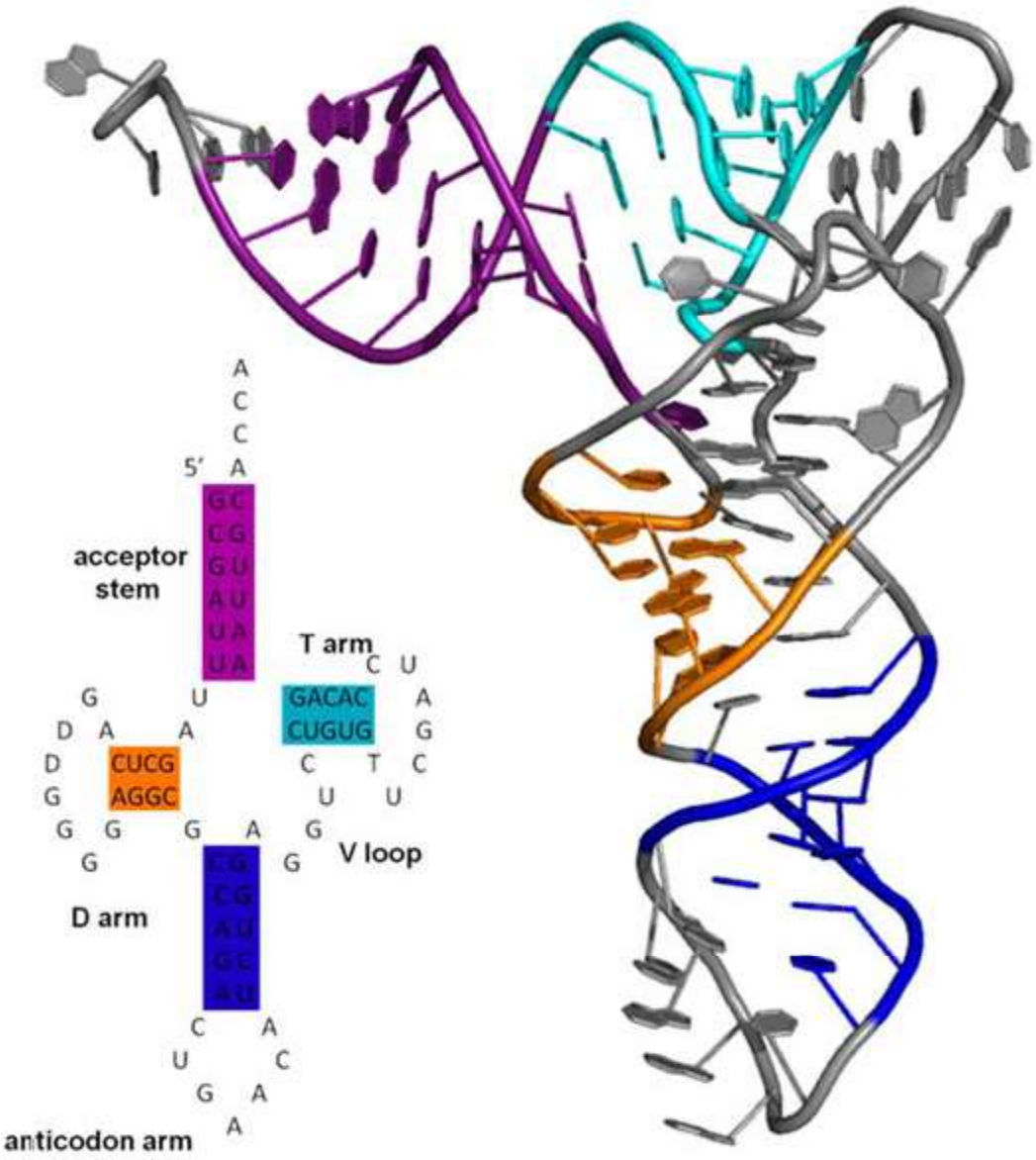
Besides double helices and the above-mentioned triplexes, RNA and DNA can both also form quadruple helices. There are diverse structures of RNA base quadruplexes. Four consecutive guanine residues can form a quadruplex in RNA by Hoogsteen hydrogen bonds to form a "Hoogsteen ring" (See Figure). G-C and A-U pairs can also form base quadruplex with a combination of Watson-Crick pairing and noncanonical pairing in the minor groove.

The core of malachite green aptamer is also a kind of base quadruplex with a different hydrogen bonding pattern (See Figure). The quadruplex can repeat several times consecutively, producing an immensely stable structure.

The unique structure of quadruplex regions in RNA may serve different functions in a biological system. Two important functions are the binding potential with ligands or proteins, and its ability to stabilize the whole tertiary structure of DNA or RNA. The strong structure can inhibit or modulate transcription and replication, such as in the telomeres of chromosomes and the UTR of mRNA. The base identity is important towards ligand binding. The G-quartet typically binds monovalent cations such as potassium, while other bases can bind numerous other ligands such as hypoxanthine in a U-U-C-U quadruplex.

Along with these functions, the G-quadruplex in the mRNA around the ribosome binding regions could serve as a regulator of gene expression in bacteria. There may be more interesting structures and functions yet to be discovered *in vivo*.

Coaxial stacking



Secondary (inset) and tertiary structure of tRNA demonstrating coaxial stacking

Coaxial stacking, otherwise known as helical stacking, is a major determinant of higher order RNA tertiary structure. Coaxial stacking occurs when two RNA duplexes form a contiguous helix, which is stabilized by base stacking at the interface of the two helices. Coaxial stacking was noted in the crystal structure of tRNAPhe. More recently, coaxial stacking has been observed in higher order structures of many ribozymes, including many forms of the self-splicing group I and group II introns. Common coaxial stacking motifs include the kissing loop interaction and the pseudoknot. The stability of these interactions can be predicted by an adaptation of "Turner's rules".

In 1994, Walter and Turner determined the free energy contributions of nearest neighbor stacking interactions within a helix-helix interface by using a model system that created a helix-helix interface between a short oligomer and a four-nucleotide overhang at the end of a hairpin stem. Their experiments confirmed that the thermodynamic contribution of base-stacking between two helical secondary structures closely mimics the thermodynamics of standard duplex formation (nearest neighbor interactions predict the thermodynamic stability of the resulting helix). The relative stability of nearest neighbor interactions can be used to predict favorable coaxial stacking based on known secondary structure. Walter and Turner found that, on average, prediction of RNA structure improved from 67% to 74% accuracy when coaxial stacking contributions were included. Theories of coaxial stacking can be tested using the technique of helical fusion. This approach was used by Murphy and Cech to confirm a coaxial stacking interaction between the P4 and P6 helices within the catalytic center of the Tetrahymena group I intron.

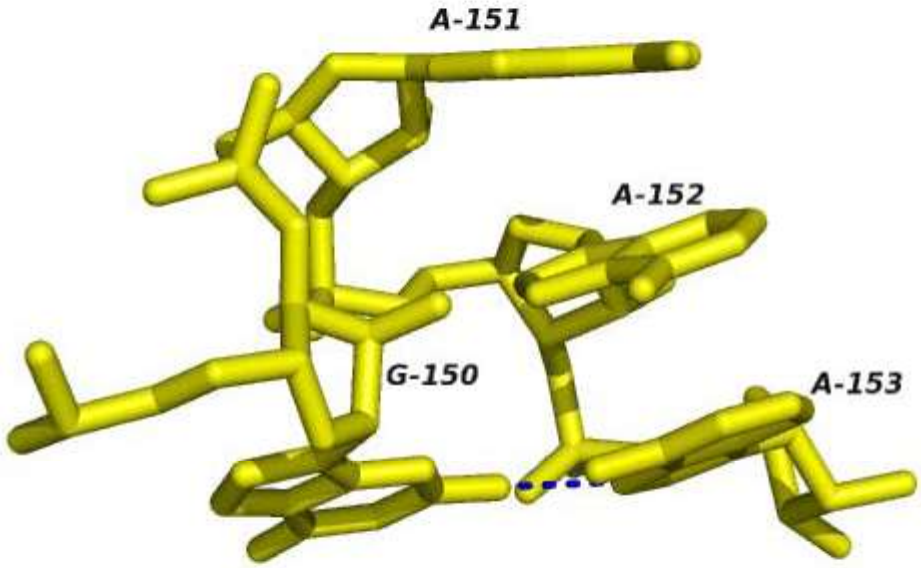
Most well-studied RNA tertiary structures contain examples of coaxial stacking. Some prominent examples are tRNA-Phe, group I introns, group II introns, and ribosomal RNAs. Crystal structures of tRNA revealed the presence of two extended helices that result from coaxial stacking of the amino-acid acceptor stem with the T-arm, and stacking of the D- and anticodon-arms. These interactions within tRNA orient the anticodon stem perpendicularly to the amino-acid stem, leading to the functional L-shaped tertiary structure. In group I introns, the P4 and P6 helices were shown to coaxially stack using a combination of biochemical and crystallographic methods. The P456 crystal structure provided a detailed view of how coaxial stacking stabilizes the packing of RNA helices into tertiary structures. In the self-splicing group II intron from Oceanobacillus iheyensis, the IA and IB stems coaxially stack and contribute to the relative orientation of the constituent helices of a five-way junction. This orientation facilitates proper folding of the active site of the functional ribozyme. The ribosome contains numerous examples of coaxial stacking, including stacked segments as long as 70 bp.

Two common motifs involving coaxial stacking are kissing loops and pseudoknots. In kissing loop interactions, the single-stranded loop regions of two hairpins interact through base pairing, forming a composite, coaxially stacked helix. Notably, this structure allows all of the nucleotides in each loop to participate in base-pairing and stacking interactions. This motif was visualized and studied using NMR analysis by Lee and Crothers. The pseudoknot motif occurs when a single stranded region of a hairpin loop basepairs with an upstream or downstream sequence within the same RNA strand. The two resulting duplex regions often stack upon one another, forming a stable coaxially stacked composite helix. One example of a pseudoknot motif is the highly stable Hepatitis Delta virus ribozyme, in which the backbone shows an overall double pseudoknot topology.

An effect similar to coaxial stacking has been observed in rationally designed DNA structures. DNA origami structures contain a large number of double helixes with exposed blunt ends. These structures were observed to stick together along the edges that contained these exposed blunt ends, due to the hydrophobic stacking interactions.

Other motifs

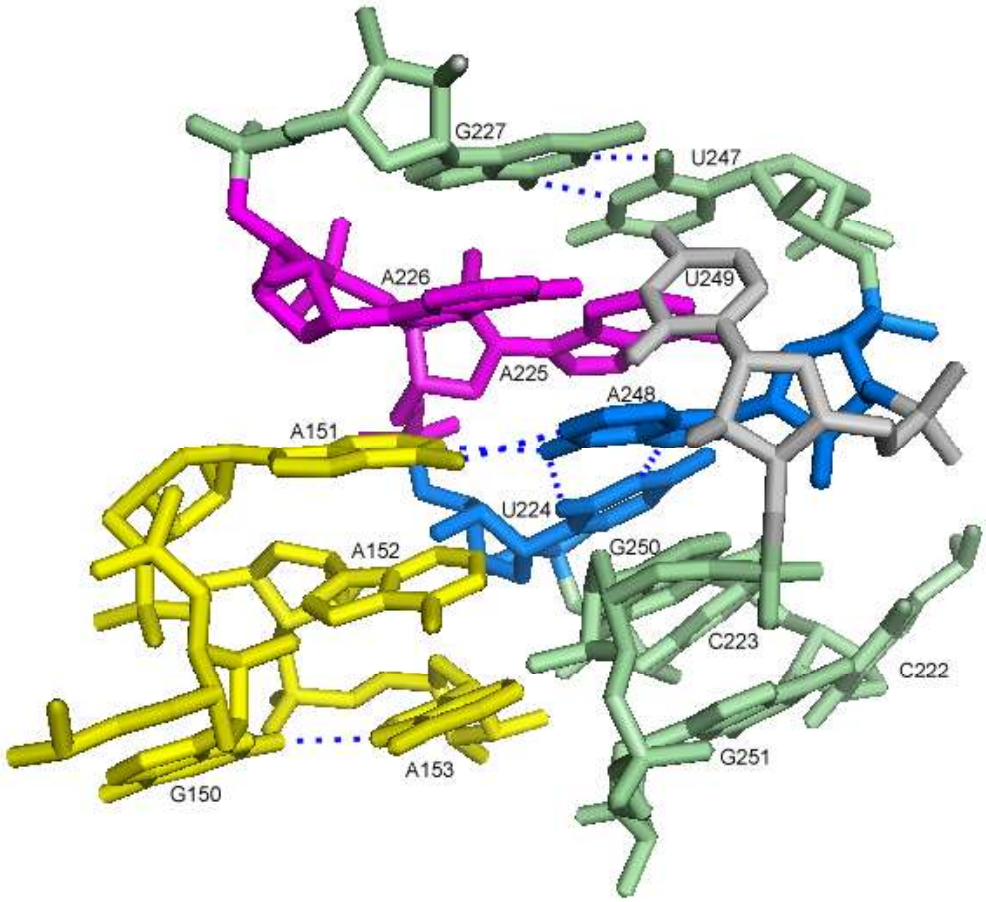
Tetraloop-receptor interactions



Stick representation of a GAAA tetraloop - an example from the GNRA tetraloop family.

Tetraloop-receptor interactions combine base-pairing and stacking interactions between the loop nucleotides of a tetraloop motif and a receptor motif located within an RNA duplex, creating a tertiary contact that stabilizes the global tertiary fold of an RNA molecule. Tetraloops are also possible structures in DNA duplexes.

Stem-loops can vary greatly in size and sequence, but tetraloops of four nucleotides are very common and they usually belong to one of three categories, based on sequence. These three families are the CUYG, UNCG and GNRA tetraloops. In each of these tetraloop families, the second and third nucleotides form a turn in the RNA strand and a base-pair between the first and fourth nucleotides stabilizes the stemloop structure. It has been determined, in general, that the stability of the tetraloop depends on the composition of bases within the loop and on the composition of this "closing base pair". The GNRA family of tetraloops is the most commonly observed within Tetraloop-receptor interactions.



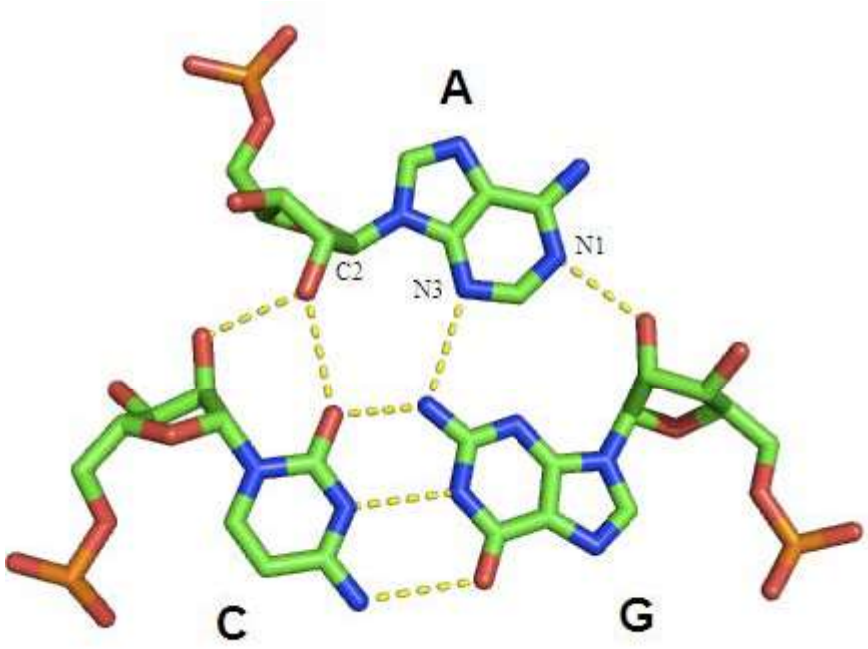
GAAA Tetraloop and Receptor: Stick representation of tetraloop (yellow) and its receptor, showing both Watson-Crick and Hoogsteen base-pairing.

"Tetraloop receptor motifs" are long-range tertiary interactions consisting of hydrogen bonding between the bases in the tetraloop to stemloop sequences in distal sections of the secondary RNA structure. In addition to hydrogen bonding, stacking interactions are an important component of these tertiary interactions. For example, in GNRA-tetraloop interactions, the second nucleotide of the tetraloop stacks directly on an A-platform motif (see above) within the receptor. The sequence of the tetraloop and its receptor often covary so that the same type of tertiary contact can be made with different isoforms of the tetraloop and its cognate receptor.

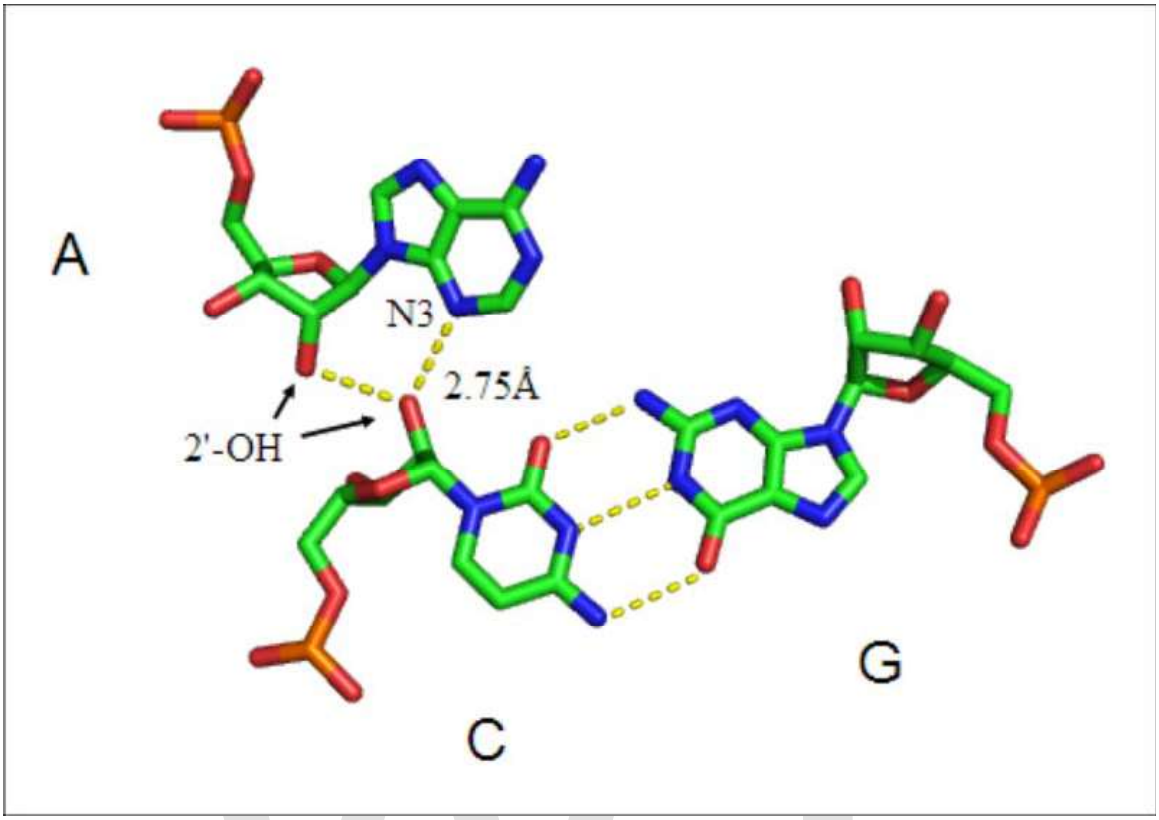
For example, the self-splicing group I intron relies on tetraloop receptor motifs for its structure and function. Specifically, the three adenine residues of the canonical GAAA motif stack on top of the receptor helix and form multiple stabilizing hydrogen bonds with the receptor. The first adenine of the GAAA sequence forms a triple base-pair with the receptor AU bases. The second adenine is stabilized by hydrogen bonds with the same uridine, as well as via its 2'-OH with the recptor and via interactions with the guanine of the GAAA tetraloop. The third adenine forms a triple base pair.

A-minor motif

A-minor Interactions



Type I A-minor interaction: Type I interactions are the most common, strongest A-minor interactions, as they involve numerous hydrogen bonds, and bury the incoming A base in the minor groove.



Type II A-minor interaction: Type II interactions involve the 2'-OH group and N3 of the adenosine. The adenosine interacts with the cytosine's 2'-OH group in the minor groove. The strength of this interaction is on the order of the Type I interaction.

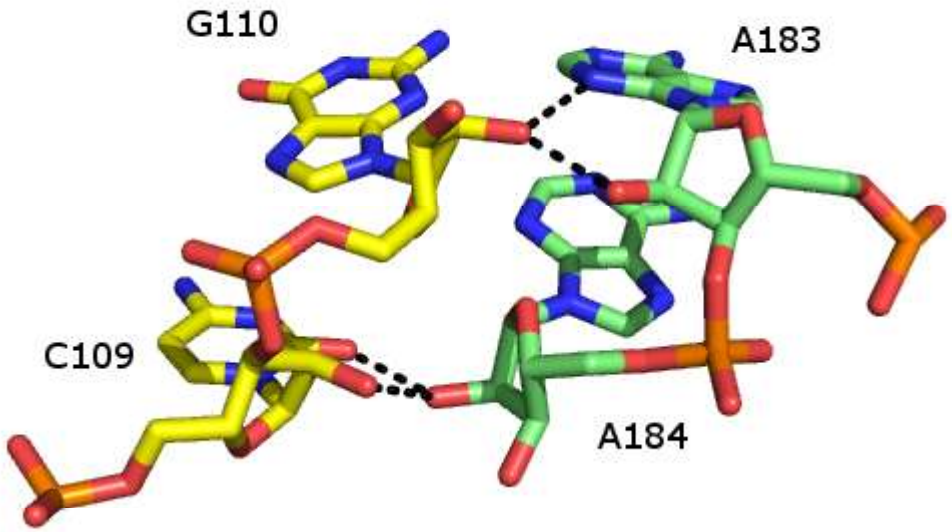
The A-minor motif is a ubiquitous RNA tertiary structural motif. It is formed by the insertion of an unpaired nucleoside into the minor groove of an RNA duplex. As such it is an example of a minor groove triple. Although guanosine, cytosine and uridine can also form minor groove triple interactions, minor groove interactions by adenine are very common. In the case of adenine, the N1-C2-N3 edge of the inserting base forms hydrogen bonds with one or both of the 2'-OH's of the duplex, as well as the bases of the duplex. The host duplex is often a G-C basepair.

A-minor motifs have been separated into four classes, types 0 to III, based upon the position of the inserting base relative to the two 2'-OH's of the Watson-Crick base pair. In type I and II A-minor motifs, N3 of adenine is inserted deeply within the minor groove of the duplex, and there is good shape complementarity with the base pair. Unlike types 0 and III, type I and II interactions are specific for adenine due to hydrogen bonding interactions. In the type III interaction, both the O2' and N3 of the inserting base are associated less closely with the minor groove of the duplex. Type 0 and III motifs are weaker and non-specific because they are mediated by interactions with a single 2'-OH.

The A-minor motif is among the most common RNA structural motifs in the ribosome, where it contributes to the binding of tRNA to the 23S subunit. They most often stabilize RNA duplex interactions in loops and helices, such as in the core of group II introns.

An interesting example of A-minor is its role in anticodon recognition. The ribosome must discriminate between correct and incorrect codon-anticodon pairs. It does so, in part, through the insertion of adenine bases into the minor groove. Incorrect codon-anticodon pairs will present distorted helical geometry, which will prevent the A-minor interaction from stabilizing the binding, and increase the dissociation rate of the incorrect tRNA.

Ribose zipper



Ribose Zippers: View of a canonical ribose zipper between two RNA backbones.

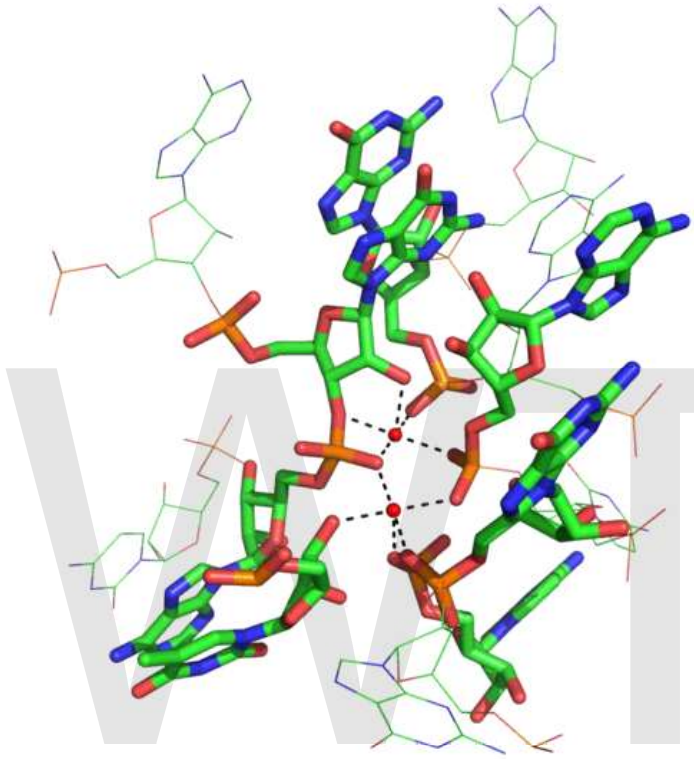
The ribose zipper is an RNA tertiary structural element in which two RNA chains are held together by hydrogen bonding interactions involving the 2'OH of ribose sugars on different strands. The 2'OH can behave as both hydrogen bond donor and acceptor, which allows formation of bifurcated hydrogen bonds with another 2' OH.

Numerous forms of ribose zipper have been reported, but a common type involves four hydrogen bonds between 2'-OH groups of two adjacent sugars. Ribose zippers commonly occur in arrays that stabilize interactions between separate RNA strands. Ribose zippers are often observed as Stem-loop interactions with very low sequence specificity. However, in the small and large ribosomal subunits, there exists a propensity for ribose

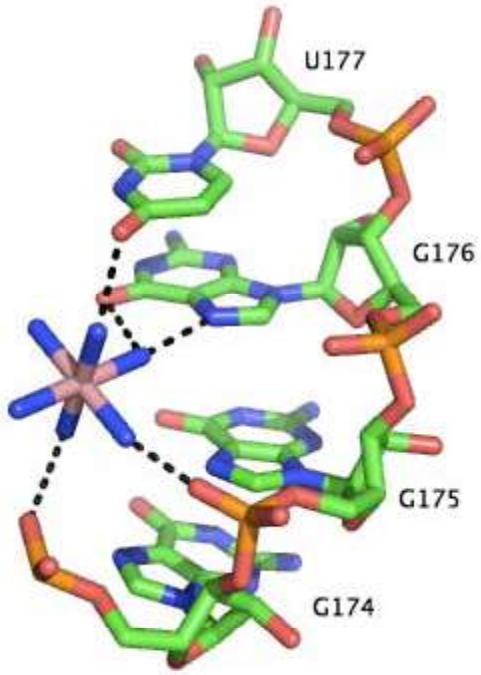
zippers of the CC/AA sequence- two cytosines on the first chain paired to two adenines on the second chain.

Role of metal ions

Metal Ion Binding in the Group I Intron



PDB rendering of Group I intron inner sphere magnesium coordination. The two red balls indicate magnesium ions and dashed lines coming from the ions indicate coordination with the respective groups on nucleotides. The color-coding scheme is as follows: green=carbon, orange=phosphate, pink=oxygen, blue=nitrogen.



PDB rendering of Group 1 intron P5c binding pocket demonstrating outer sphere coordination. Here, the six amines of osmium hexamine(III) fulfill the role generally served by water molecules and mediate the ion's interaction with the major groove. Coordination via hydrogen bonds is indicated by dashed lines and osmium is rendered in pink, all other colors are as above.

Functional RNAs are often folded, stable molecules with three-dimensional shapes rather than floppy, linear strands. Cations are essential for thermodynamic stabilization of RNA tertiary structures. Metal cations that bind RNA can be monovalent, divalent or trivalent. Potassium (K^+) is a common monovalent ion that binds RNA. A common divalent ion that binds RNA is magnesium (Mg^{2+}). Other ions including sodium (Na^+), calcium (Ca^{2+}) and manganese (Mn^{2+}) have been found to bind RNA *in vivo* and *in vitro*. Multivalent organic cations such as spermidine or spermine are also found in cells and these make important contributions to RNA folding. Trivalent ions such as cobalt hexamine or lanthanide ions such as terbium (Tb^{3+}) are useful experimental tools for studying metal binding to RNA.

A metal ion can interact with RNA in multiple ways. An ion can associate diffusely with the RNA backbone, shielding otherwise unfavorable electrostatic interactions. This charge screening is often fulfilled by monovalent ions. Site-bound ions stabilize specific elements of RNA tertiary structure. Site-bound interactions can be further subdivided into two categories depending on whether water mediates the metal binding. "Outer sphere" interactions are mediated by water molecules that surround the metal ion. For example, magnesium hexahydrate interacts with and stabilizes specific RNA tertiary structure motifs via interactions with guanosine in the major groove. Conversely, "inner sphere"

interactions are directly mediated by the metal ion. RNA often folds in multiple stages and these steps can be stabilized by different types of cations. In the early stages, RNA forms secondary structures stabilized through the binding of monovalent cations, divalent cations and polyanionic amines in order to neutralize the polyanionic backbone. The later stages of this process involve the formation of RNA tertiary structure, which is stabilized almost largely through the binding of divalent ions such as magnesium with possible contributions from potassium binding.

Metal-binding sites are often localized in the deep and narrow major groove of the RNA duplex, coordinating to the Hoogsteen edges of purines. In particular, metal cations stabilize sites of backbone twisting where tight packing of phosphates results in a region of dense negative charge. There are several metal ion-binding motifs in RNA duplexes that have been identified in crystal structures. For instance, in the P4-P6 domain of the *Tetrahymena thermophila* group I intron, several ion-binding sites consist of tandem G-U wobble pairs and tandem G-A mismatches, in which divalent cations interact with the Hoogsteen edge of guanosine via O6 and N7. Another ion-binding motif in the *Tetrahymena* group I intron is the A-A platform motif, in which consecutive adenosines in the same strand of RNA form a non-canonical pseudobase pair. Unlike the tandem G-U motif, the A-A platform motif binds preferentially to monovalent cations. In many of these motifs, absence of the monovalent or divalent cations results in either greater flexibility or loss of tertiary structure.

Divalent metal ions, especially magnesium, have been found to be important for the structure of DNA junctions such as the Holliday junction intermediate in genetic recombination. The magnesium ion shields the negatively-charged phosphate groups in the junction and allows them to be positioned closer together, allowing a stacked conformation rather than an unstacked conformation. Magnesium is vital in stabilizing these kinds of junctions in artificially designed structures used in DNA nanotechnology, such as the double crossover motif.

History

The earliest work in RNA structural biology coincided, more or less, with the work being done on DNA in the early 1950s. In their seminal 1953 paper, Watson and Crick suggested that van der Waals crowding by the 2`OH group of ribose would preclude RNA from adopting a double helical structure identical to the model they proposed - what we now know as B-form DNA. This provoked questions about the three dimensional structure of RNA: could this molecule form some type of helical structure, and if so, how?

In the mid 1960's, the role of tRNA in protein synthesis was being intensively studied. In 1965, Holley *et al.* purified and sequenced the first tRNA molecule, initially proposing that it adopted a cloverleaf structure, based largely on the ability of certain regions of the molecule to form stem loop structures. The isolation of tRNA proved to be the first major windfall in RNA structural biology. In 1971, Kim *et al.* achieved another breakthrough,

producing crystals of yeast $tRNA^{PHE}$ that diffracted to 2-3 Ångström resolutions by using spermine, a naturally occurring polyamine, which bound to and stabilized the tRNA.

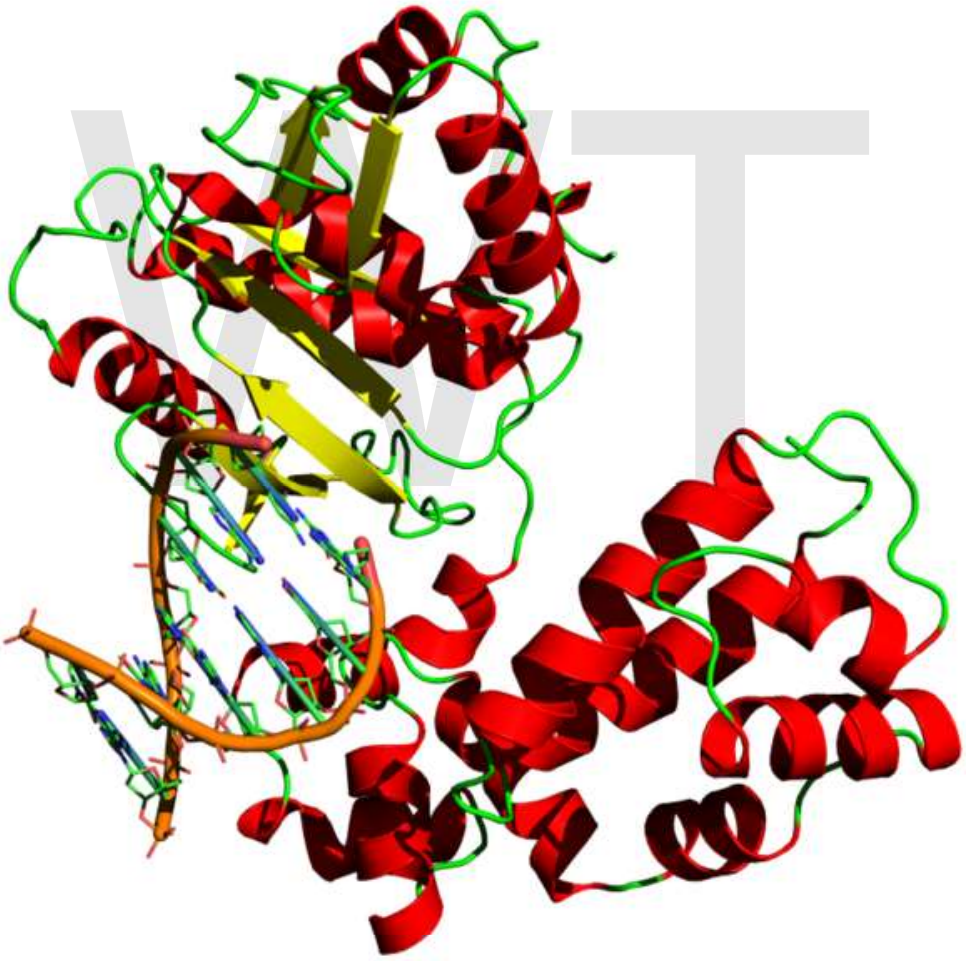
For a considerable time following the first tRNA structures, the field of RNA structure did not dramatically advance. The ability to study an RNA structure depended upon the potential to isolate the RNA target. This proved limiting to the field for many years, in part owing to the fact that other known targets - i.e. the ribosome - were significantly more difficult to isolate and crystallize. As such, for some twenty years following the original publication of the $tRNA^{PHE}$ structure, the structures of only a handful of other RNA targets were solved, with almost all of these belonging to the transfer RNA family.

This unfortunate lack of scope would eventually be overcome largely because of two major advancements in nucleic acid research: the identification of ribozymes, and the ability to produce them via *in vitro* transcription. Subsequent to Tom Cech's publication implicating the *Tetrahymena* group I intron as an autocatalytic ribozyme, and Sidney Altman's report of catalysis by ribonuclease P RNA, several other catalytic RNAs were identified in the late 1980s, including the hammerhead ribozyme. In 1994, McKay *et al.* published the structure of a 'hammerhead RNA-DNA ribozyme-inhibitor complex' at 2.6 Ångström resolution, in which the autocatalytic activity of the ribozyme was disrupted via binding to a DNA substrate. In addition to the advances being made in global structure determination via crystallography, the early 1990s also saw the implementation of NMR as a powerful technique in RNA structural biology. Investigations such as this enabled a more precise characterization of the base pairing and base stacking interactions which stabilized the global folds of large RNA molecules.

The resurgence of RNA structural biology in the mid-1990s has caused a veritable explosion in the field of nucleic acid structural research. Since the publication of the hammerhead and P_{4-6} structures, numerous major contributions to the field have been made. Some of the most noteworthy examples include the structures of the Group I and Group II introns, and the Ribosome. It should be noted that the first three structures were produced using *in vitro* transcription, and that NMR has played a role in investigating partial components of all four structures - testaments to the indispensability of both techniques for RNA research. Most recently, the 2009 Nobel Prize in Chemistry was awarded to Ada Yonath, Venkatraman Ramakrishnan, and Thomas Steitz for their structural work on the ribosome, demonstrating the prominent role RNA structural biology has taken in modern molecular biology.

Chapter- 13

DNA Polymerase

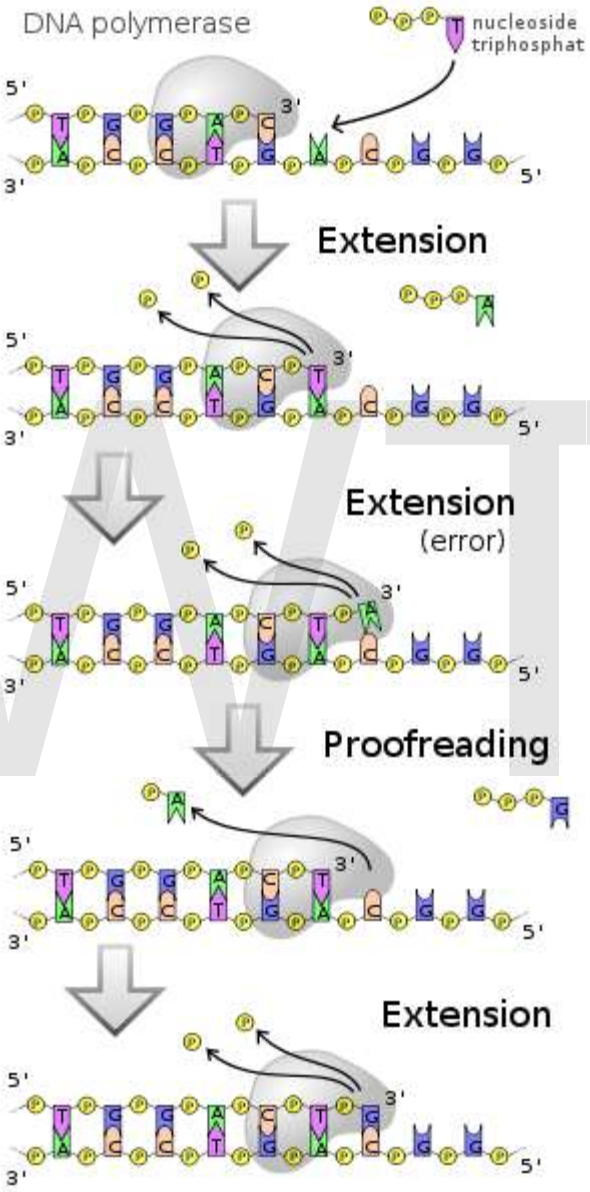


3D structure of the DNA-binding helix-turn-helix motifs in human DNA polymerase beta

A **DNA polymerase** is an enzyme that catalyzes 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.

Function



DNA polymerase with proofreading ability

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. An enzyme known as a helicase is required to unwind DNA from a double-strand structure to a single-strand structure to facilitate replication of each strand consistent with the semiconservative model of DNA replication.

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.

Various DNA polymerases are extensively used in molecular biology experiments.

Variation across species

DNA polymerases have highly-conserved structure, which means that their overall catalytic subunits vary, on a whole, very little from species to species. Conserved structures usually indicate important, irreplaceable functions of the cell, the maintenance of which provides evolutionary advantages.

Some viruses also encode special DNA polymerases, such as Hepatitis B virus DNA polymerase. These may selectively replicate viral DNA through a variety of mechanisms. Retroviruses encode an unusual DNA polymerase called reverse transcriptase, which is an RNA-dependent DNA polymerase (RdDp). It polymerizes DNA from a template of RNA.

DNA polymerase families

Based on sequence homology, DNA polymerases can be further subdivided into seven different families: A, B, C, D, X, Y, and RT.

Family A

Polymerases contain both replicative and repair polymerases. Replicative members from this family include the extensively-studied T7 DNA polymerase, as well as the eukaryotic mitochondrial DNA Polymerase γ. Among the repair polymerases are *Escherichia coli* DNA pol I, *Thermus aquaticus* pol I, and *Bacillus stearothermophilus* pol I. These repair polymerases are involved in excision repair and processing of Okazaki fragments generated during lagging strand synthesis.

Family C

Polymerases are the primary bacterial chromosomal replicative enzymes. DNA Polymerase III alpha subunit from *E. coli* is the catalytic subunit and possesses no known

nuclease activity. A separate subunit, the epsilon subunit, possesses the 3'-5' exonuclease activity used for editing during chromosomal replication. Recent research has classified Family C polymerases as a subcategory of Family X.

Family D

Polymerases are still not very well characterized. All known examples are found in the Euryarchaeota subdomain of Archaea and are thought to be replicative polymerases.

Family X

Contains the well-known eukaryotic polymerase pol β, as well as other eukaryotic polymerases such as pol σ, pol λ, pol μ, and terminal deoxynucleotidyl transferase (TdT). Pol β is required for short-patch base excision repair, a DNA repair pathway that is essential for repairing abasic sites. Pol λ and Pol μ are involved in non-homologous end-joining, a mechanism for rejoining DNA double-strand breaks. TdT is expressed only in lymphoid tissue, and adds "n nucleotides" to double-strand breaks formed during V(D)J recombination to promote immunological diversity. The yeast *Saccharomyces cerevisiae* has only one Pol X polymerase, Pol IV, which is involved in non-homologous end-joining.

Family Y

Y Polymerases differ from others in having a low fidelity on undamaged templates and in their ability to replicate through damaged DNA. Members of this family are hence called translesion synthesis (TLS) polymerases. Depending on the lesion, TLS polymerases can bypass the damage in an error-free or error-prone fashion, the latter resulting in elevated mutagenesis. Xeroderma pigmentosum variant (XPV) patients for instance have mutations in the gene encoding Pol η (eta), which is error-free for UV-lesions. In XPV patients, alternative error-prone polymerases, e.g., Pol ζ (zeta) (polymerase ζ is a B Family polymerase a complex of the catalytic subunit REV3L with Rev7, which associates with Rev1), are thought to be involved in mistakes that result in the cancer predisposition of these patients. Other members in humans are Pol ι (iota), Pol κ (kappa), and Rev1 (terminal deoxycytidyl transferase). In *E. coli*, two TLS polymerases, Pol IV (DINB) and Pol V ($UmuD'_2C$), are known.

Family RT

The reverse transcriptase family contains examples from both retroviruses and eukaryotic polymerases. The eukaryotic polymerases are usually restricted to telomerases. These polymerases use an RNA template to synthesize the DNA strand.

Chapter- 14

Prokaryotic DNA Replication and Eukaryotic DNA Replication

Prokaryotic DNA replication

DNA replication in prokaryotes is exemplified in *E. coli*. It is bi-directional and originates at a single origin of replication (OriC).

Initiation

The initiation of DNA replication is mediated by DnaA, a protein that binds to a region of the origin known as the DnaA box. In *E. coli*, there are 5 DnaA boxes, each of which contains a highly conserved 9 bp consensus sequence 5' - TTATCCACA - 3'. Binding of DnaA to this region causes it to become negatively supercoiled. Following this, a region of OriC upstream of the DnaA boxes (known as DnaB boxes) become melted. There are three of these regions, and each is 13 bp long, and AT-rich (which facilitates melting because less energy is required to break the two hydrogen bonds that form between A and T nucleotides). This region has the consensus sequence 5' - GATCTNTTNTTTT - 3. Melting of the DnaB boxes requires ATP (which is hydrolyzed by DnaA). Following melting, DnaA recruits a hexameric helicase (six DnaB proteins) to opposite ends of the melted DNA. This is where the replication fork will form. Recruitment of helicase requires six DnaC proteins, each of which is attached to one subunit of helicase. Once this complex is formed, an additional five DnaA proteins bind to the original five DnaA proteins to form five DnaA dimers. DnaC is then released, and the prepriming complex is complete. In order for DNA replication to continue, SSB protein is needed to prevent the single strands of DNA from forming any secondary structures and to prevent them from reannealing, and DNA gyrase is needed to relieve the stress (by creating negative supercoils) created by the action of DnaB helicase. The unwinding of DNA by DnaB helicase allows for primase (DnaG) an RNA polymerase to prime each DNA template so that DNA synthesis can begin.

Elongation

Once priming is complete, DNA polymerase III holoenzyme is loaded into the DNA and replication begins. The catalytic mechanism of DNA polymerase III involves the use of two metal ions in the active site, and a region in the active site that can discriminate between deoxyribonucleotides and ribonucleotides. The metal ions are general divalent cations that help the 3' OH initiate a nucleophilic attack onto the alpha phosphate of the deoxyribonucleotide and orient and stabilize the negatively charged triphosphate on the deoxyribonucleotide. Nucleophilic attack by the 3' OH on the alpha phosphate releases pyrophosphate, which is then subsequently hydrolyzed (by inorganic phosphatase) into two phosphates. This hydrolysis drives DNA synthesis to completion.

Furthermore, DNA polymerase III must be able to distinguish between correctly paired bases and incorrectly paired bases. This is accomplished by distinguishing Watson-Crick base pairs through the use of an active site pocket that is complementary in shape to the structure of correctly paired nucleotides. This pocket has a tyrosine residue that is able to form van der Waals interactions with the correctly paired nucleotide. In addition, dsDNA (double stranded DNA) in the active site has a wider and shallower minor groove that permits the formation of hydrogen bonds with the third nitrogen of purine bases and the second oxygen of pyrimidine bases. Finally, the active site makes extensive hydrogen bonds with the DNA backbone. These interactions result in the DNA polymerase III closing around a correctly paired base. If a base is inserted and incorrectly paired, these interactions could not occur due to disruptions in hydrogen bonding and van der Waals interactions.

DNA is read in the 3' → 5' direction, therefore, nucleotides are synthesized (or attached to the template strand) in the 5' → 3' direction. However, one of the parent strands of DNA is 3' → 5' while the other is 5' → 3'. To solve this, replication occurs in opposite directions. Heading towards the replication fork, the leading strand is synthesized in a continuous fashion, only requiring one primer. On the other hand, the lagging strand, heading away from the replication fork, is synthesized in a series of short fragments known as Okazaki fragments, consequently requiring many primers. The RNA primers of Okazaki fragments are subsequently degraded by RNAse H and DNA Polymerase I (exonuclease), and the gap (or nicks) are filled with deoxyribonucleotides and sealed by the enzyme ligase.

Termination

Termination of DNA replication in *E. coli* is completed through the use of termination sequences and the Tus protein. These sequences allow the two replication forks to pass through in only one direction, but not the other.

DNA replication initially produces two catenated or linked circular DNA duplexes, each comprising one parental strand and one newly synthesised strand (by nature of semiconservative replication). This catenation can be visualised as two interlinked rings

which cannot be separated. Topoisomerase IV in *E. coli* unlinks or decatenates the two circular DNA duplexes.

Regulation

Regulation of DNA replication is achieved through several mechanisms. Mechanisms involve the ratio of ATP to ADP, of DnaA to the number of DnaA boxes and the hemimethylation and sequestering of OriC.

The ratio of ATP to ADP indicates that the cell has reached a specific size and is ready to divide. This "signal" occurs because in a rich medium, the cell will grow quickly and will have a lot of excess DNA.

Furthermore, DnaA binds equally well to ATP or ADP, and only the DnaA-ATP complex is able to initiate replication. Thus, in a fast growing cell, there will be more DnaA-ATP than DnaA-ADP. Because the levels of DnaA are strictly regulated, and 5 DnaA-DnaA dimers are needed to initiate replication, the ratio of DnaA to the number of DnaA boxes in the cell is important. After DNA replication is complete, this number is halved, thus DNA replication cannot occur until the levels of DnaA protein increases.

Finally, DNA is sequestered to a membrane-binding protein called SeqA. This protein binds to hemi-methylated GATC DNA sequences. This four bp sequences occurs 11 times in OriC, and newly synthesized DNA only has its parent strand methylated. DAM methyltransferase methylates the newly synthesized strand of DNA only if it is not bound to SeqA. The importance of hemi-methylation is twofold. Firstly, OriC becomes inaccessible to DnaA, and secondly, DnaA binds better to fully methylated DNA than hemi-methylated DNA.

These mechanisms serve to downregulate DNA replication so that it only occurs once per cell cycle, preventing over-replication of DNA.

Eukaryotic DNA replication

DNA replication in eukaryotes is much more complicated than in prokaryotes, although there are many similar aspects. Eukaryotic cells can only initiate DNA replication at a specific point in the cell cycle, the beginning of S phase.

Mechanism

Location in cell cycle

DNA replication in eukaryotes occurs only in the S phase of the cell cycle. However, pre-initiation occurs in the G1 phase. Thus, the separation of pre-initiation and activation ensures that the origin can only fire once per cell cycle.

Due to the sheer size of chromosomes in eukaryotes, eukaryotic chromosomes contain multiple origins of replication. Some origins are well characterized, such as the autonomously replicating sequences (ARS) of yeast while other eukaryotic origins, particularly those in metazoa, can be found in spans of thousands of basepairs. However, the assembly and initiation of replicaton is similar in both the protozoa and metazoa.

Preparation in G1 phase

The first step in DNA replication is the formation of the pre-initiation replication complex (the pre-RC). The formation of this complex occurs in two stages. The first stage requires that there is no CDK activity. This can only occur in early G1. The formation of the pre-RC is known as licensing, but a licensed pre-RC cannot initiate replication in the G1 phase

Current models hold that it begins with the binding of the origin recognition complex (ORC) to the origin. This complex is a hexamer of related proteins and remains bound to the origin, even after DNA replication occurs. Furthermore, ORC is the functional analogue of prokaryotic DnaA. Following the binding of ORC to the origin, Cdc6/Cdc18 and Cdt1 coordinate the loading of the MCM (Mini Chromosome Maintenance) complex to the origin by first binding to ORC and then binding to the MCM complex. The MCM complex is thought to be the major DNA helicase in eukaryotic organisms. Once binding of MCM occurs, a fully licensed pre-RC exists.

Synthesis in S phase

Activation of the complex occurs in S-phase and requires Cdk2-Cyclin E and Ddk. The activation process begins with the addition of Mcm10 to the pre-RC, which displaces Cdt1. Following this, Ddk phosphorylates Mcm3-7, which activates the helicase. It is believed that ORC and Cdc6/18 are phosphorylated by Cdk2-Cyclin E. Ddk and the Cdk complex then recruits another protein called Cdc45, which then recruits all of the DNA

replication proteins to the replication fork. At this stage the origin fires and DNA synthesis begins.

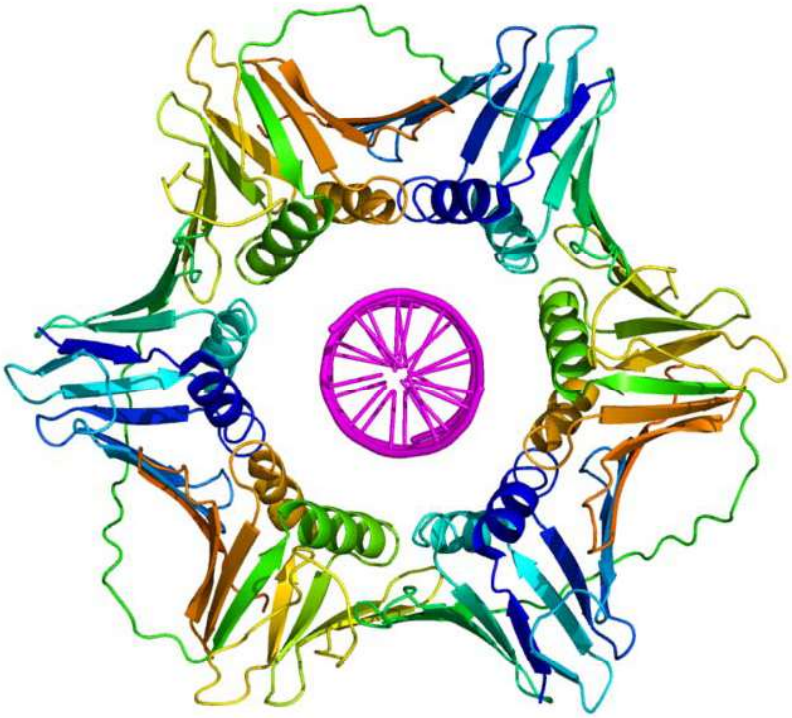
Activation of a new round of replication is prevented through the actions of the cyclin dependent kinases and a protein known as geminin. Geminin binds to Cdt1 and sequesters it. It is a periodic protein that first appears in S-phase and is degraded in late M-phase, possibly through the action of the anaphase promoting complex (APC). In addition, phosphorylation of Cdc6/18 prevent it from binding to the ORC (thus inhibiting loading of the MCM complex) while the phosphorylation of ORC remains unclear. Cells in the G0 stage of the cell cycle are prevented from initiating a round of replication because the Mcm proteins are not expressed.

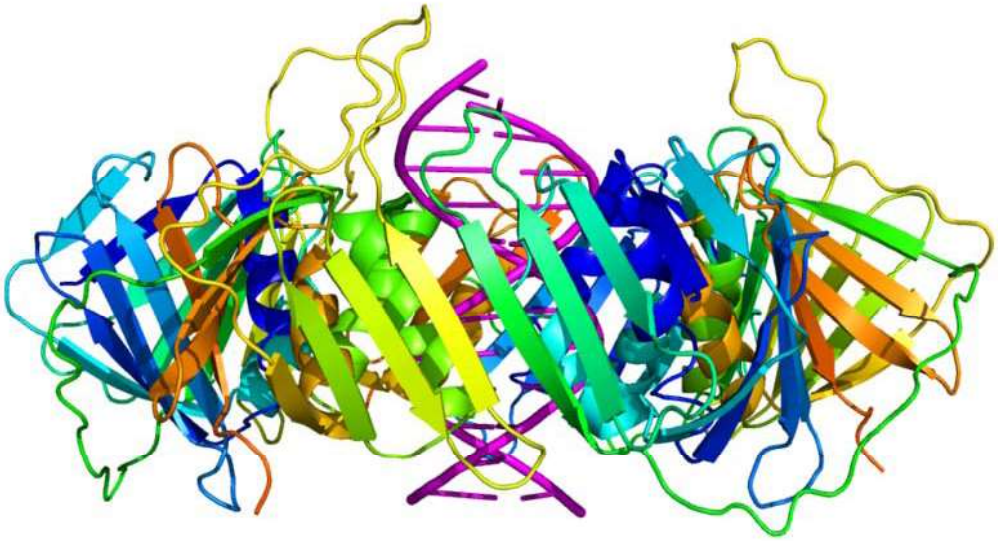
At least three different types of eukaryotic DNA polymerases are involved in the replication of DNA in animal cells (POL α, Pol δ and POL ε).

- **Pol α** forms a complex with a small catalytic (PriS) and a large noncatalytic (PriL) subunit, with the Pri subunits acting as a primase (synthesizing an RNA primer), and then with DNA Pol α elongating that primer with DNA nucleotides. After around 20 nucleotides elongation is taken over by Pol ε (on the leading strand) and δ (on the lagging strand).
- **Pol δ**: Highly processive and has proofreading 3'->5' exonuclease activity. Thought to be the main polymerase involved in lagging strand synthesis, though there is still debate about its role.
- **Pol ε**: Also highly processive and has proofreading 3'->5' exonuclease activity. Highly related to pol δ, and thought to be the main polymerase involved in leading strand synthesis, though there is again still debate about its role.

Chapter- 15

DNA Clamp





Top and side views of a homotrimer of the human PCNA sliding clamp (rainbow colored, N-terminus = blue, C-terminus = red) complexed with double stranded DNA (magenta).

A **DNA clamp**, also known as a **sliding clamp**, is a protein fold that serves as a processivity-promoting factor in DNA replication. As a critical component of the DNA polymerase III holoenzyme, the clamp protein binds DNA polymerase and prevents this enzyme from dissociating from the template DNA strand. The clamp-polymerase protein-protein interactions are stronger and more specific than the direct interactions between the polymerase and the template DNA strand; because the rate-limiting step in the DNA synthesis reaction is the association of the polymerase with the DNA template, the presence of the sliding clamp dramatically increases the number of nucleotides that the polymerase can add to the growing strand per association event. The presence of the DNA clamp can increase the rate of DNA synthesis up to 1,000-fold compared with a nonprocessive polymerase.

Structure

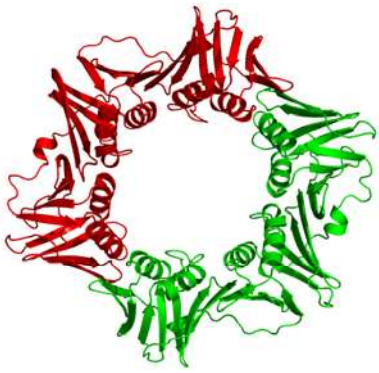
The DNA clamp fold is an α+β protein that assembles into a multimeric structure that completely encircles the DNA double helix as the polymerase adds nucleotides to the growing strand. The DNA clamp assembles on the DNA at the replication fork and "slides" along the DNA with the advancing polymerase, aided by a layer of water molecules in the central pore of the clamp between the DNA and the protein surface. Because of the toroidal shape of the assembled multimer, the clamp cannot dissociate from the template strand without also dissociating into monomers.

The DNA clamp fold is found in bacteria, archaea, eukaryotes and some viruses. In bacteria, the sliding clamp is a homodimer composed of two identical beta subunits of DNA polymerase III and hence is referred to as the beta clamp. In archaea and eukaryotes, it is a trimer composed of three molecules of PCNA. The T4 bacteriophage also uses a sliding clamp, called gp45 that is a trimer similar in structure to PCNA but lacks sequence homology to either PCNA or the bacterial beta clamp.

Kingdom	Sliding clamp protein	Aggregation state	Associated polymerase
Bacteria	beta subunit of pol III	dimer	DNA polymerase III
Archaea	archaeal PCNA	trimer	pol ε
Eukaryote	PCNA	trimer	DNA polymerase delta
Virus	gp43 / gp45	timer	RB69 Pol / T4 Pol

Bacterial

DNA polymerase III subunit beta



Crystallographic structure of the dimeric DNA polymerase beta subunit from *E. coli*.

Identifiers	
Organism	*Escherichia coli*
Symbol	dnaN
Entrez	948218
PDB	1MMI
RefSeq (Prot)	NP_418156
UniProt	P0A988
Other data	
EC number	2.7.7.7
Chromosome	*MG1655: 3.88 - 3.88 Mb*

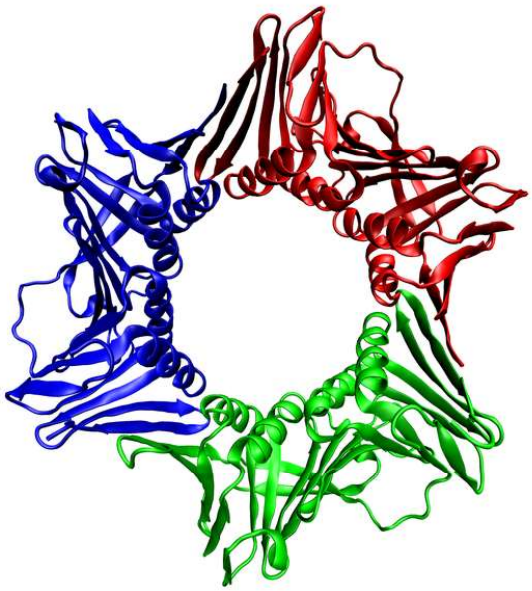
The **beta clamp** is a specific DNA clamp and a subunit of the DNA polymerase III holoenzyme found in bacteria. Two beta subunits are assembled around the DNA by the gamma subunit and ATP hydrolysis; this assembly is called the pre-initiation complex. After assembly around the DNA, the beta subunits' affinity for the gamma subunit is replaced by an affinity for the alpha and epsilon subunits, which together create the complete holoenzyme. DNA polymerase III is the primary enzyme complex involved in prokaryotic DNA replication.

The gamma complex of DNA polymerase III, composed of γδδ'χψ subunits, catalyzes ATP to chaperone two beta subunits to bind to DNA. Once bound to DNA, the beta subunits can freely slide along double stranded DNA. The beta subunits in turn bind the αε polymerase complex. The α subunit possesses DNA polymerase activity and the ε subunit is a 3'-5' exonuclease.

The beta chain of bacterial DNA polymerase III is composed of three topologically non-equivalent domains (N-terminal, central, and C-terminal). Two beta chain molecules are tightly associated to form a closed ring encircling duplex DNA.

Eukaryote

proliferating cell nuclear antigen



The assembled human DNA clamp, a trimer of the human protein PCNA.

Identifiers	
Symbol	PCNA

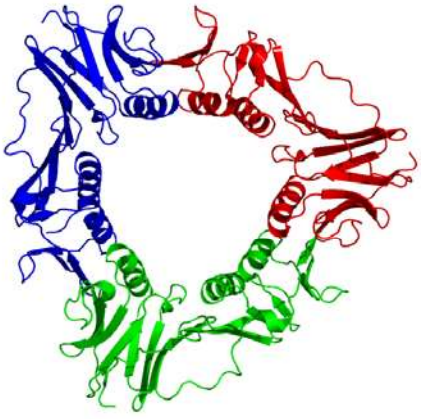
Entrez	5111
HUGO	8729
OMIM	176740
PDB	1axc
RefSeq	NM_002592
UniProt	P12004
Other data	
EC number	2.7.7.7
Locus	Chr. 20 *pter-p12*

The sliding clamp in eukaryotes is assembled from a specific subunit of DNA polymerase delta called the proliferating cell nuclear antigen (PCNA). The N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA.

The sequence of PCNA is well conserved between plants and animals, indicating a strong selective pressure for structure conservation, and suggesting that this type of DNA replication mechanism is conserved throughout eukaryotes. Homologues of PCNA have also been identified in the archaea (*Euryarchaeota* and *Crenarchaeota*) and in Paramecium bursaria Chlorella virus 1 (PBCV-1) and in nuclear polyhedrosis viruses.

Viral

DNA polymerase accessory protein 45



Crystallographic structure of the trimeric gp45 sliding clamp from *bacteriophage T4*.

Identifiers

Organism	*Enterobacteria phage T4*
Symbol	gp45
Entrez	1258821
PDB	1CZD
RefSeq (Prot)	NP_049666
UniProt	P04525
Other data	
EC number	2.7.7.7
Chromosome	*1: 0.03 - 0.03 Mb*

The viral gp45 sliding clamp subunit protein contains two domains. Each domain consists of two alpha helices and two beta sheets - the fold is duplicated and has internal pseudo two-fold symmetry. Three gp45 molecules are tightly associated to form a closed ring encircling duplex DNA.

Assembly

Sliding clamps are loaded onto their associated DNA template strands by specialized proteins known as "sliding clamp loaders", which also disassemble the clamps after replication has completed. The binding sites for these initiator proteins overlap with the binding sites for the DNA polymerase, so the clamp cannot simultaneously associate with a clamp loader and with a polymerase. Thus the clamp will not be actively disassembled while the polymerase remains bound. Although DNA clamps play a less significant role in associating with other DNA-interacting proteins, such as nucleosome assembly factors, Okazaki fragment ligases, and DNA repair proteins, all of these proteins also share a binding site on the DNA clamp that overlaps with the clamp loader site, ensuring that the clamp will not be removed while any enzyme is still working on the DNA. The activity of the clamp loader requires ATP hydrolysis to "close" the clamp around the DNA.

Chapter- 16

Polymerase Chain Reaction

A strip of eight PCR tubes, each containing a 100 µl reaction mixture

The **polymerase chain reaction** (**PCR**) is a scientific technique in molecular biology to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.

Developed in 1983 by Kary Mullis, PCR is now a common and often indispensable technique used in medical and biological research labs for a variety of applications. These include DNA cloning for sequencing, DNA-based phylogeny, or functional analysis of genes; the diagnosis of hereditary diseases; the identification of genetic fingerprints (used in forensic sciences and paternity testing); and the detection and diagnosis of infectious diseases. In 1993, Mullis was awarded the Nobel Prize in Chemistry for his work on PCR.

The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase (after which the method is named) are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase, an enzyme originally isolated from the bacterium *Thermus aquaticus*. This DNA polymerase enzymatically assembles a new DNA strand from DNA building blocks, the nucleotides, by using single-stranded DNA as a template and DNA oligonucleotides (also called DNA primers), which are required for initiation of DNA synthesis. The vast majority of PCR methods use thermal cycling, i.e., alternately heating and cooling the PCR sample to a defined series of temperature steps. These thermal cycling steps are necessary first to physically separate the two strands in a DNA double helix at a high temperature in a process called DNA melting. At a lower temperature, each strand is then used as the template in DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The selectivity of PCR results from the use of primers that are complementary to the DNA region targeted for amplification under specific thermal cycling conditions.

PCR principles and procedure

Figure 1a: A thermal cycler for PCR

Figure 1b: An older model three-temperature thermal cycler for PCR

PCR is used to amplify a specific region of a DNA strand (the DNA target). Most PCR methods typically amplify DNA fragments of up to ~10 kilo base pairs (kb), although some techniques allow for amplification of fragments up to 40 kb in size.

A basic PCR set up requires several components and reagents. These components include:

- *DNA template* that contains the DNA region (target) to be amplified.
- Two *primers* that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strand of the DNA target.

- *Taq polymerase* or another DNA polymerase with a temperature optimum at around 70 °C.
- *Deoxynucleotide triphosphates* (dNTPs), the building blocks from which the DNA polymerases synthesizes a new DNA strand.
- *Buffer solution,* providing a suitable chemical environment for optimum activity and stability of the DNA polymerase.
- *Divalent cations,* magnesium or manganese ions; generally Mg^{2+} is used, but Mn^{2+} can be utilized for PCR-mediated DNA mutagenesis, as higher Mn^{2+} concentration increases the error rate during DNA synthesis
- *Monovalent cation* potassium ions.

The PCR is commonly carried out in a reaction volume of 10–200 µl in small reaction tubes (0.2–0.5 ml volumes) in a thermal cycler. The thermal cycler heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction (see below). Many modern thermal cyclers make use of the Peltier effect which permits both heating and cooling of the block holding the PCR tubes simply by reversing the electric current. Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration. Most thermal cyclers have heated lids to prevent condensation at the top of the reaction tube. Older thermocyclers lacking a heated lid require a layer of oil on top of the reaction mixture or a ball of wax inside the tube.

Procedure

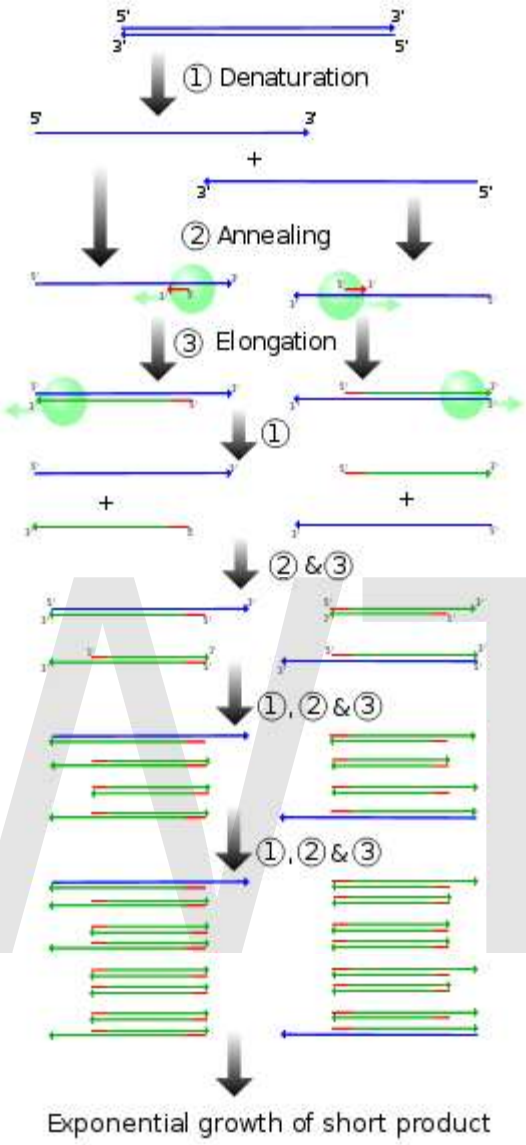


Figure 2: Schematic drawing of the PCR cycle. **(1) Denaturing at 94–96 °C. (2) Annealing at ~65 °C (3) Elongation at 72 °C**. Four cycles are shown here. The blue lines represent the DNA template to which primers (red arrows) anneal that are extended by the DNA polymerase (light green circles), to give shorter DNA products (green lines), which themselves are used as templates as PCR progresses.

Typically, PCR consists of a series of 20-40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2-3 discrete temperature steps, usually three (Fig. 2). The cycling is often preceded by a single temperature step (called *hold*) at a high temperature (>90°C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis,

the concentration of divalent ions and dNTPs in the reaction, and the melting temperature (Tm) of the primers.

- *Initialization step*: This step consists of heating the reaction to a temperature of 94–96 °C (or 98 °C if extremely thermostable polymerases are used), which is held for 1–9 minutes. It is only required for DNA polymerases that require heat activation by hot-start PCR.

- *Denaturation step*: This step is the first regular cycling event and consists of heating the reaction to 94–98 °C for 20–30 seconds. It causes DNA melting of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single-stranded DNA molecules.

- *Annealing step*: The reaction temperature is lowered to 50–65 °C for 20–40 seconds allowing annealing of the primers to the single-stranded DNA template. Typically the annealing temperature is about 3-5 degrees Celsius below the Tm of the primers used. Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA synthesis.

- *Extension/elongation step*: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75–80 °C, and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to be amplified. As a rule-of-thumb, at its optimum temperature, the DNA polymerase will polymerize a thousand bases per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

- *Final elongation*: This single step is occasionally performed at a temperature of 70–74 °C for 5–15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.

- *Final hold*: This step at 4–15 °C for an indefinite time may be employed for short-term storage of the reaction.

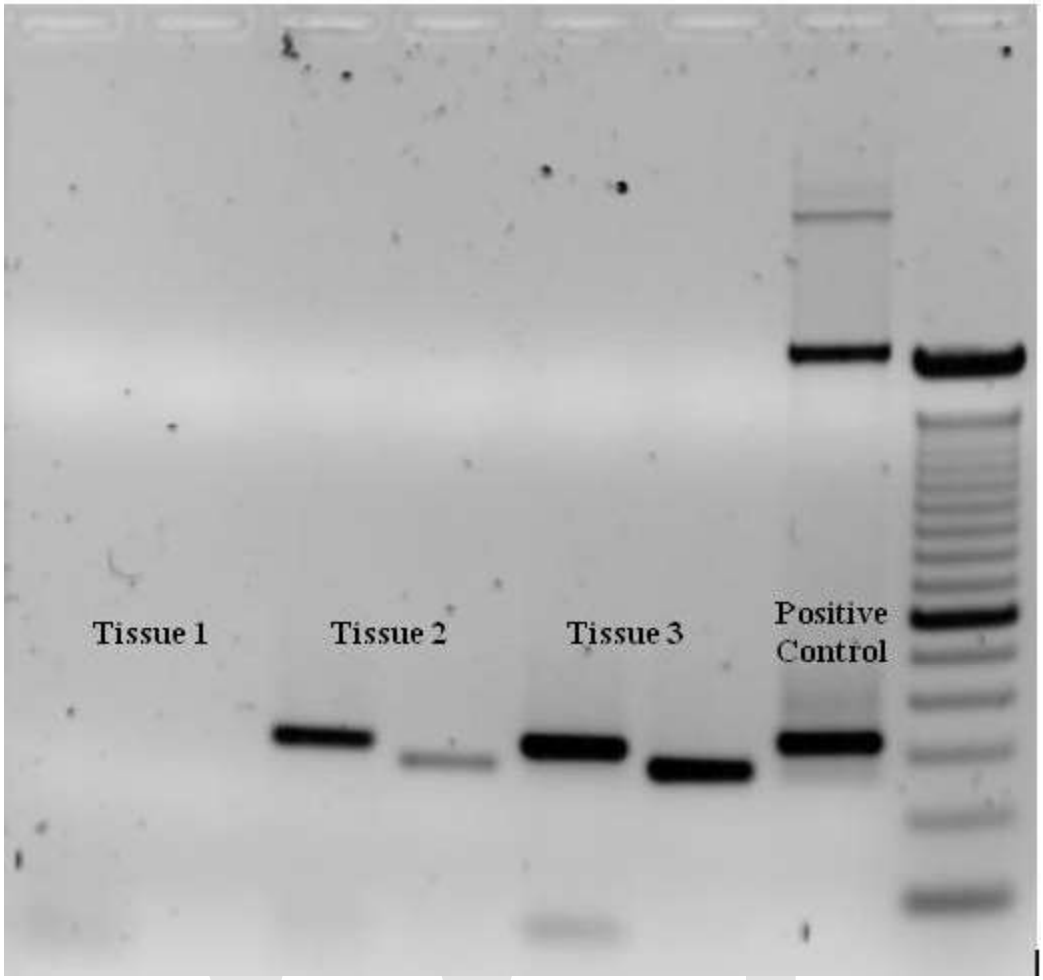


Figure 3: Ethidium bromide-stained PCR products after gel electrophoresis. Two sets of primers were used to amplify a target sequence from three different tissue samples. No amplification is present in sample #1; DNA bands in sample #2 and #3 indicate successful amplification of the target sequence. The gel also shows a positive control, and a DNA ladder containing DNA fragments of defined length for sizing the bands in the experimental PCRs.

To check whether the PCR generated the anticipated DNA fragment (also sometimes referred to as the amplimer or amplicon), agarose gel electrophoresis is employed for size separation of the PCR products. The size(s) of PCR products is determined by comparison with a DNA ladder (a molecular weight marker), which contains DNA fragments of known size, run on the gel alongside the PCR products (see Fig. 3).

PCR stages

The PCR process can be divided into three stages:

Exponential amplification: At every cycle, the amount of product is doubled (assuming 100% reaction efficiency). The reaction is very sensitive: only minute quantities of DNA need to be present.

Levelling off stage: The reaction slows as the DNA polymerase loses activity and as consumption of reagents such as dNTPs and primers causes them to become limiting.

Plateau: No more product accumulates due to exhaustion of reagents and enzyme.

PCR optimization

In practice, PCR can fail for various reasons, in part due to its sensitivity to contamination causing amplification of spurious DNA products. Because of this, a number of techniques and procedures have been developed for optimizing PCR conditions. Contamination with extraneous DNA is addressed with lab protocols and procedures that separate pre-PCR mixtures from potential DNA contaminants. This usually involves spatial separation of PCR-setup areas from areas for analysis or purification of PCR products, use of disposable plasticware, and thoroughly cleaning the work surface between reaction setups. Primer-design techniques are important in improving PCR product yield and in avoiding the formation of spurious products, and the usage of alternate buffer components or polymerase enzymes can help with amplification of long or otherwise problematic regions of DNA. Addition of reagents, such as formamide, in buffer systems may increase the specificity and yield of PCR.

Application of PCR

Selective DNA isolation

PCR allows isolation of DNA fragments from genomic DNA by selective amplification of a specific region of DNA. This use of PCR augments many methods, such as generating hybridization probes for Southern or northern hybridization and DNA cloning, which require larger amounts of DNA, representing a specific DNA region. PCR supplies these techniques with high amounts of pure DNA, enabling analysis of DNA samples even from very small amounts of starting material.

Other applications of PCR include DNA sequencing to determine unknown PCR-amplified sequences in which one of the amplification primers may be used in Sanger sequencing, isolation of a DNA sequence to expedite recombinant DNA technologies involving the insertion of a DNA sequence into a plasmid or the genetic material of another organism. Bacterial colonies (E. coli) can be rapidly screened by PCR for correct DNA vector constructs. PCR may also be used for genetic fingerprinting; a forensic technique used to identify a person or organism by comparing experimental DNAs through different PCR-based methods.

Some PCR 'fingerprints' methods have high discriminative power and can be used to identify genetic relationships between individuals, such as parent-child or between

siblings, and are used in paternity testing (Fig. 4). This technique may also be used to determine evolutionary relationships among organisms.

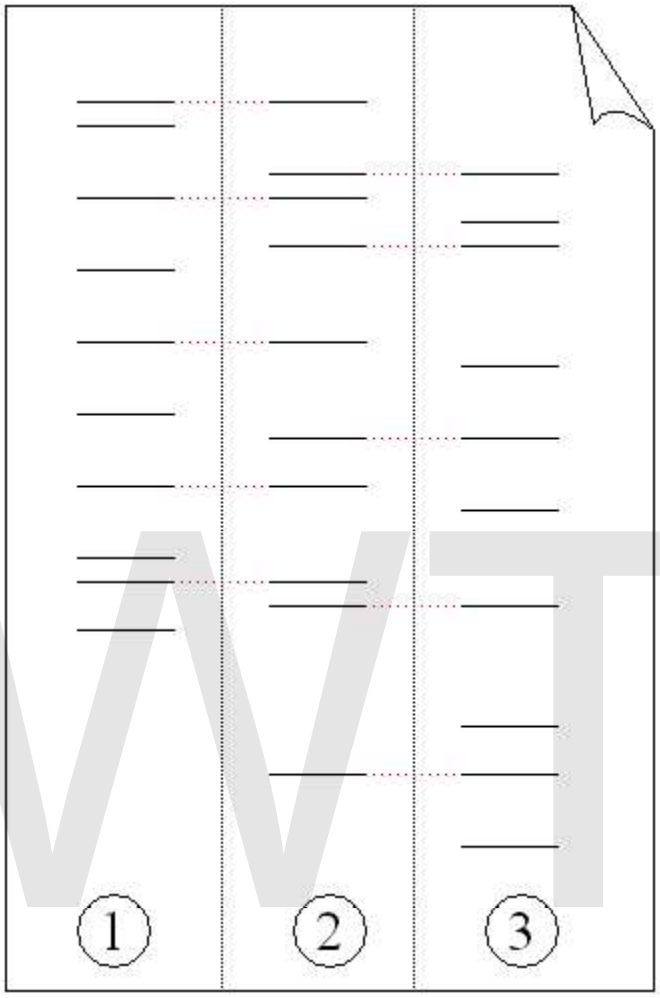


Figure 4: Electrophoresis of PCR-amplified DNA fragments. (1) Father. (2) Child. (3) Mother. The child has inherited some, but not all of the fingerprint of each of its parents, giving it a new, unique fingerprint.

Amplification and quantification of DNA

Because PCR amplifies the regions of DNA that it targets, PCR can be used to analyze extremely small amounts of sample. This is often critical for forensic analysis, when only a trace amount of DNA is available as evidence. PCR may also be used in the analysis of ancient DNA that is tens of thousands of years old. These PCR-based techniques have been successfully used on animals, such as a forty-thousand-year-old mammoth, and also on human DNA, in applications ranging from the analysis of Egyptian mummies to the identification of a Russian tsar.

Quantitative PCR methods allow the estimation of the amount of a given sequence present in a sample—a technique often applied to quantitatively determine levels of gene expression. Real-time PCR is an established tool for DNA quantification that measures the accumulation of DNA product after each round of PCR amplification.

PCR in diagnosis of diseases

PCR permits early diagnosis of malignant diseases such as leukemia and lymphomas, which is currently the highest developed in cancer research and is already being used routinely. PCR assays can be performed directly on genomic DNA samples to detect translocation-specific malignant cells at a sensitivity which is at least 10,000 fold higher than other methods.

PCR also permits identification of non-cultivatable or slow-growing microorganisms such as mycobacteria, anaerobic bacteria, or viruses from tissue culture assays and animal models. The basis for PCR diagnostic applications in microbiology is the detection of infectious agents and the discrimination of non-pathogenic from pathogenic strains by virtue of specific genes.

Viral DNA can likewise be detected by PCR. The primers used need to be specific to the targeted sequences in the DNA of a virus, and the PCR can be used for diagnostic analyses or DNA sequencing of the viral genome. The high sensitivity of PCR permits virus detection soon after infection and even before the onset of disease. Such early detection may give physicians a significant lead in treatment. The amount of virus ("viral load") in a patient can also be quantified by PCR-based DNA quantitation techniques (see below).

Variations on the basic PCR technique

- *Allele-specific PCR*: a diagnostic or cloning technique which is based on single-nucleotide polymorphisms (SNPs) (single-base differences in DNA). It requires prior knowledge of a DNA sequence, including differences between alleles, and uses primers whose 3' ends encompass the SNP. PCR amplification under stringent conditions is much less efficient in the presence of a mismatch between template and primer, so successful amplification with an SNP-specific primer signals presence of the specific SNP in a sequence.

- *Assembly PCR* or *Polymerase Cycling Assembly (PCA)*: artificial synthesis of long DNA sequences by performing PCR on a pool of long oligonucleotides with short overlapping segments. The oligonucleotides alternate between sense and antisense directions, and the overlapping segments determine the order of the PCR fragments, thereby selectively producing the final long DNA product.

- *Asymmetric PCR*: preferentially amplifies one DNA strand in a double-stranded DNA template. It is used in sequencing and hybridization probing where amplification of only one of the two complementary strands is required. PCR is

carried out as usual, but with a great excess of the primer for the strand targeted for amplification. Because of the slow (arithmetic) amplification later in the reaction after the limiting primer has been used up, extra cycles of PCR are required. A recent modification on this process, known as *L*inear-*A*fter-*T*he-*E*xponential-PCR (LATE-PCR), uses a limiting primer with a higher melting temperature (Tm) than the excess primer to maintain reaction efficiency as the limiting primer concentration decreases mid-reaction.

- *Helicase-dependent amplification*: similar to traditional PCR, but uses a constant temperature rather than cycling through denaturation and annealing/extension cycles. DNA helicase, an enzyme that unwinds DNA, is used in place of thermal denaturation.

- *Hot-start PCR*: a technique that reduces non-specific amplification during the initial set up stages of the PCR. It may be performed manually by heating the reaction components to the melting temperature (e.g., 95°C) before adding the polymerase. Specialized enzyme systems have been developed that inhibit the polymerase's activity at ambient temperature, either by the binding of an antibody or by the presence of covalently bound inhibitors that only dissociate after a high-temperature activation step. Hot-start/cold-finish PCR is achieved with new hybrid polymerases that are inactive at ambient temperature and are instantly activated at elongation temperature.

- *Intersequence-specific PCR* (ISSR): a PCR method for DNA fingerprinting that amplifies regions between simple sequence repeats to produce a unique fingerprint of amplified fragment lengths.

- *Inverse PCR*: is commonly used to identify the flanking sequences around genomic inserts. It involves a series of DNA digestions and self ligation, resulting in known sequences at either end of the unknown sequence.

- *Ligation-mediated PCR*: uses small DNA linkers ligated to the DNA of interest and multiple primers annealing to the DNA linkers; it has been used for DNA sequencing, genome walking, and DNA footprinting.

- *Methylation-specific PCR* (MSP): developed by Stephen Baylin and Jim Herman at the Johns Hopkins School of Medicine, and is used to detect methylation of CpG islands in genomic DNA. DNA is first treated with sodium bisulfite, which converts unmethylated cytosine bases to uracil, which is recognized by PCR primers as thymine. Two PCRs are then carried out on the modified DNA, using primer sets identical except at any CpG islands within the primer sequences. At these points, one primer set recognizes DNA with cytosines to amplify methylated DNA, and one set recognizes DNA with uracil or thymine to amplify unmethylated DNA. MSP using qPCR can also be performed to obtain quantitative rather than qualitative information about methylation.

- *Miniprimer PCR*: uses a thermostable polymerase (S-Tbr) that can extend from short primers ("smalligos") as short as 9 or 10 nucleotides. This method permits PCR targeting to smaller primer binding regions, and is used to amplify conserved DNA sequences, such as the 16S (or eukaryotic 18S) rRNA gene.

- *Multiplex Ligation-dependent Probe Amplification* (*MLPA*): permits multiple targets to be amplified with only a single primer pair, thus avoiding the resolution limitations of multiplex PCR (see below).

- *Multiplex-PCR*: consists of multiple primer sets within a single PCR mixture to produce amplicons of varying sizes that are specific to different DNA sequences. By targeting multiple genes at once, additional information may be gained from a single test run that otherwise would require several times the reagents and more time to perform. Annealing temperatures for each of the primer sets must be optimized to work correctly within a single reaction, and amplicon sizes, i.e., their base pair length, should be different enough to form distinct bands when visualized by gel electrophoresis.

- *Nested PCR*: increases the specificity of DNA amplification, by reducing background due to non-specific amplification of DNA. Two sets of primers are used in two successive PCRs. In the first reaction, one pair of primers is used to generate DNA products, which besides the intended target, may still consist of non-specifically amplified DNA fragments. The product(s) are then used in a second PCR with a set of primers whose binding sites are completely or partially different from and located 3' of each of the primers used in the first reaction. Nested PCR is often more successful in specifically amplifying long DNA fragments than conventional PCR, but it requires more detailed knowledge of the target sequences.

- *Overlap-extension PCR*: a genetic engineering technique allowing the construction of a DNA sequence with an alteration inserted beyond the limit of the longest practical primer length.

- *Quantitative PCR (Q-PCR)*: used to measure the quantity of a PCR product (commonly in real-time). It quantitatively measures starting amounts of DNA, cDNA or RNA. Q-PCR is commonly used to determine whether a DNA sequence is present in a sample and the number of its copies in the sample. *Quantitative real-time PCR* has a very high degree of precision. QRT-PCR methods use fluorescent dyes, such as Sybr Green, EvaGreen or fluorophore-containing DNA probes, such as TaqMan, to measure the amount of amplified product in real time. It is also sometimes abbreviated to RT-PCR (*R*eal *T*ime PCR) or RQ-PCR. QRT-PCR or RTQ-PCR are more appropriate contractions, since RT-PCR commonly refers to reverse transcription PCR, often used in conjunction with Q-PCR.

- *R*everse *T*ranscription PCR (RT-PCR): for amplifying DNA from RNA. Reverse transcriptase reverse transcribes RNA into cDNA, which is then amplified by

PCR. RT-PCR is widely used in expression profiling, to determine the expression of a gene or to identify the sequence of an RNA transcript, including transcription start and termination sites. If the genomic DNA sequence of a gene is known, RT-PCR can be used to map the location of exons and introns in the gene. The 5' end of a gene (corresponding to the transcription start site) is typically identified by RACE-PCR (*Rapid Amplification of cDNA Ends*).

- *Solid Phase PCR*: encompasses multiple meanings, including Polony Amplification (where PCR colonies are derived in a gel matrix, for example), Bridge PCR (primers are covalently linked to a solid-support surface), conventional Solid Phase PCR (where Asymmetric PCR is applied in the presence of solid support bearing primer with sequence matching one of the aqueous primers) and Enhanced Solid Phase PCR (where conventional Solid Phase PCR can be improved by employing high Tm and nested solid support primer with optional application of a thermal 'step' to favour solid support priming).

- *Thermal asymmetric interlaced PCR (TAIL-PCR)*: for isolation of an unknown sequence flanking a known sequence. Within the known sequence, TAIL-PCR uses a nested pair of primers with differing annealing temperatures; a degenerate primer is used to amplify in the other direction from the unknown sequence.

- *Touchdown PCR* (*Step-down PCR*): a variant of PCR that aims to reduce nonspecific background by gradually lowering the annealing temperature as PCR cycling progresses. The annealing temperature at the initial cycles is usually a few degrees (3-5°C) above the T_m of the primers used, while at the later cycles, it is a few degrees (3-5°C) below the primer T_m. The higher temperatures give greater specificity for primer binding, and the lower temperatures permit more efficient amplification from the specific products formed during the initial cycles.

- *PAN-AC*: uses isothermal conditions for amplification, and may be used in living cells.

- *Universal Fast Walking*: for genome walking and genetic fingerprinting using a more specific 'two-sided' PCR than conventional 'one-sided' approaches (using only one gene-specific primer and one general primer - which can lead to artefactual 'noise') by virtue of a mechanism involving lariat structure formation. Streamlined derivatives of UFW are LaNe RAGE (lariat-dependent nested PCR for rapid amplification of genomic DNA ends), 5'RACE LaNe and 3'RACE LaNe.

History

A 1971 paper in the Journal of Molecular Biology by Kleppe and co-workers first described a method using an enzymatic assay to replicate a short DNA template with primers *in vitro*. However, this early manifestation of the basic PCR principle did not receive much attention, and the invention of the polymerase chain reaction in 1983 is generally credited to Kary Mullis.

At the core of the PCR method is the use of a suitable DNA polymerase able to withstand the high temperatures of >90 °C (194 °F) required for separation of the two DNA strands in the DNA double helix after each replication cycle. The DNA polymerases initially employed for in vitro experiments presaging PCR were unable to withstand these high temperatures. So the early procedures for DNA replication were very inefficient, time consuming, and required large amounts of DNA polymerase and continual handling throughout the process.

The discovery in 1976 of Taq polymerase — a DNA polymerase purified from the thermophilic bacterium, *Thermus aquaticus*, which naturally lives in hot (50 to 80 °C (122 to 176 °F)) environments such as hot springs — paved the way for dramatic improvements of the PCR method. The DNA polymerase isolated from *T. aquaticus* is stable at high temperatures remaining active even after DNA denaturation, thus obviating the need to add new DNA polymerase after each cycle. This allowed an automated thermocycler-based process for DNA amplification.

When Mullis developed the PCR in 1983, he was working in Emeryville, California for Cetus Corporation, one of the first biotechnology companies. There, he was responsible for synthesizing short chains of DNA. Mullis has written that he conceived of PCR while cruising along the Pacific Coast Highway one night in his car. He was playing in his mind with a new way of analyzing changes (mutations) in DNA when he realized that he had instead invented a method of amplifying any DNA region through repeated cycles of duplication driven by DNA polymerase. In *Scientific American*, Mullis summarized the procedure: "Beginning with a single molecule of the genetic material DNA, the PCR can generate 100 billion similar molecules in an afternoon. The reaction is easy to execute. It requires no more than a test tube, a few simple reagents, and a source of heat." He was awarded the Nobel Prize in Chemistry in 1993 for his invention, seven years after he and his colleagues at Cetus first put his proposal to practice. However, some controversies have remained about the intellectual and practical contributions of other scientists to Mullis' work, and whether he had been the sole inventor of the PCR principle (see below).

Patent wars

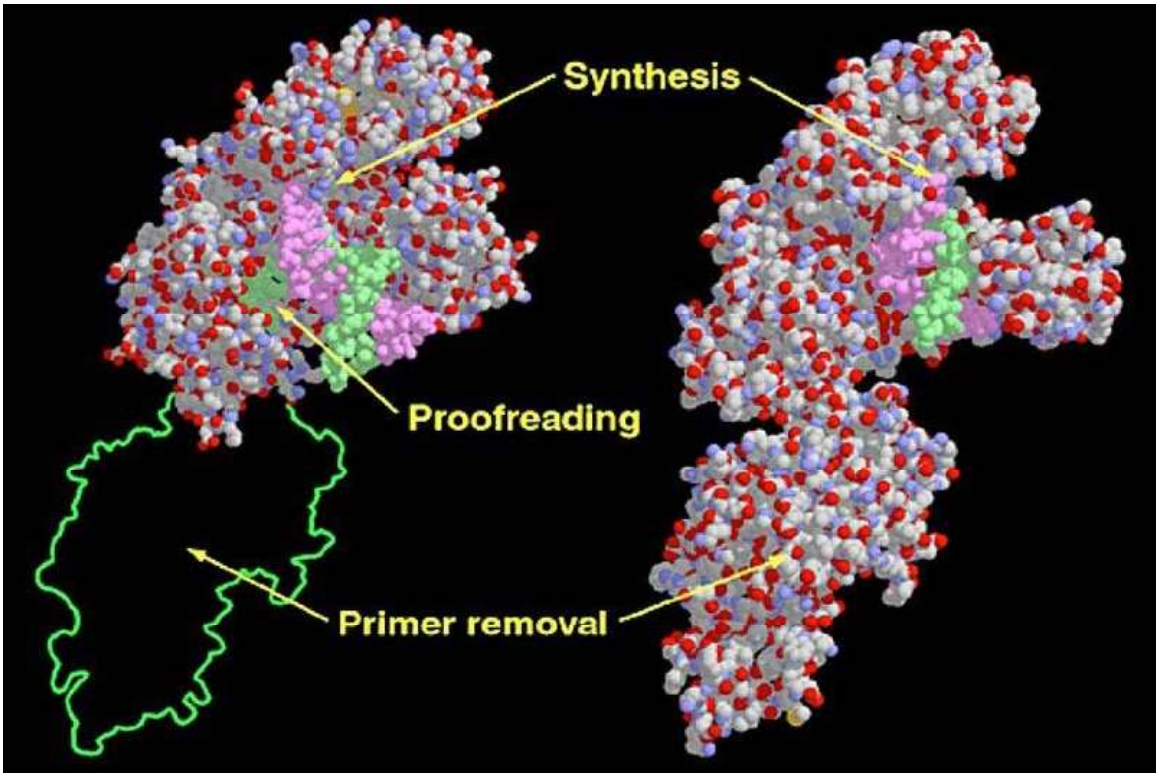
The PCR technique was patented by Kary Mullis and assigned to Cetus Corporation, where Mullis worked when he invented the technique in 1983. The *Taq* polymerase enzyme was also covered by patents. There have been several high-profile lawsuits related to the technique, including an unsuccessful lawsuit brought by DuPont. The pharmaceutical company Hoffmann-La Roche purchased the rights to the patents in 1992 and currently holds those that are still protected.

A related patent battle over the Taq polymerase enzyme is still ongoing in several jurisdictions around the world between Roche and Promega. The legal arguments have extended beyond the lives of the original PCR and Taq polymerase patents, which expired on March 28, 2005.

Chapter- 17

Klenow Fragment and Okazaki Fragment

Klenow fragment



Functional domains in the Klenow Fragment (left) and DNA Polymerase I (PDB).

The **Klenow fragment** is a large protein fragment produced when DNA polymerase I from *E. coli* is enzymatically cleaved by the protease subtilisin. First reported in 1970, it retains the 5'-3' polymerase activity and the 3’ → 5’ exonuclease activity for removal of precoding nucleotides and proofreading, but loses its 5' → 3' exonuclease activity.

The other smaller fragment formed when DNA polymerase I from *E. coli* is cleaved by subtilisin retains the 5'-3' exonuclease activity but does not have the other two activities exhibited by the Klenow fragment (i.e. 5'-> 3' polymerase activity, and 3'->5' exonuclease activity).

Research

Because the 5' → 3' exonuclease activity of DNA polymerase I from *E. coli* makes it unsuitable for many applications, the Klenow fragment, which lacks this activity, can be very useful in research. The Klenow fragment is extremely useful for research-based tasks such as:

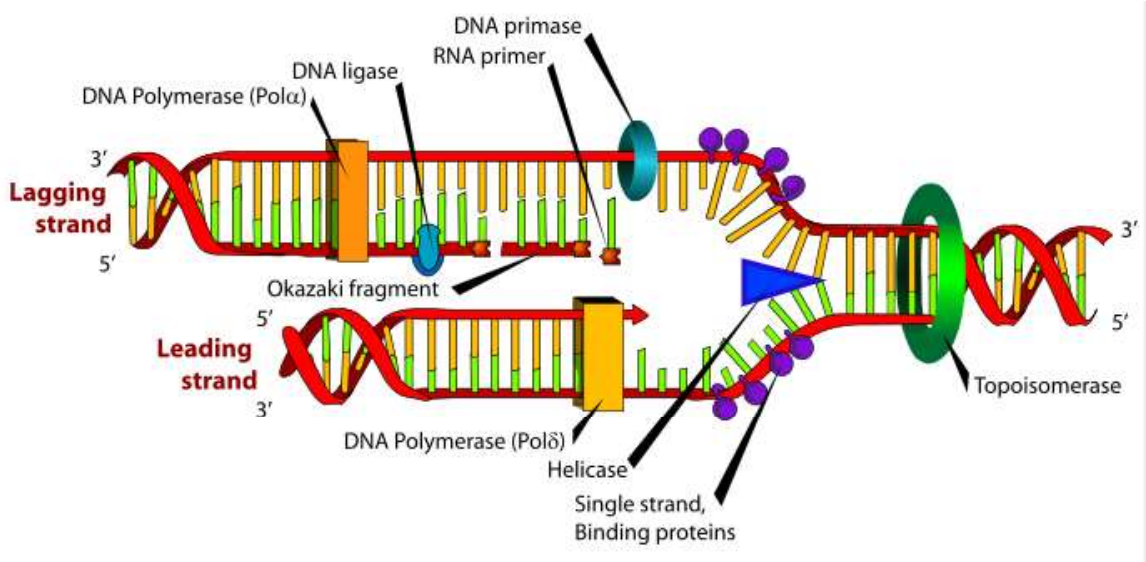
- Synthesis of double-stranded DNA from single-stranded templates
- Filling in (meaning removal of overhangs to create blunt ends) recessed 3' ends of DNA fragments
- Digesting away protruding 3' overhangs
- Preparation of radioactive DNA probes

The Klenow fragment was also the original enzyme used for greatly amplifying segments of DNA in the polymerase chain reaction (PCR) process, before being replaced by thermostable enzymes such as Taq polymerase.

The exo- Klenow fragment

Just as the 5' → 3' exonuclease activity of DNA polymerase I from *E.coli* can be undesirable, the 3' → 5' exonuclease activity of Klenow fragment can also be undesirable for certain applications. This problem can be overcome by introducing mutations in the gene that encodes Klenow. This results in forms of the enzyme being expressed that retain 5' → 3' polymerase activity, but lack any exonuclease activity (5' → 3' or 3' → 5'). This form of the enzyme is called the **exo- Klenow fragment**.

Okazaki fragment



DNA replication

An **Okazaki fragment** is a relatively short fragment of DNA (with no RNA primer at the 5' terminus) created on the lagging strand during DNA replication. The lengths of Okazaki fragments are between 1,000 to 2,000 nucleotides long in *E. coli* and are generally between 100 to 200 nucleotides long in eukaryotes. It was originally discovered in 1968 by Reiji Okazaki, Tsuneko Okazaki, and their colleagues while studying replication of bacteriophage DNA in *Escherichia coli*.

When the lagging strand is being replicated on the original strand, the 5'-3' pattern must be used; thus a small discontinuity occurs and an Okazaki fragment forms. These fragments are processed by the replication machinery to produce a continuous strand of DNA and hence a complete daughter DNA helix.

In dealing with the synthesis of complementary DNA strands the nascent leading strand always reads 3' to 5'. Its antiparallel complement strand, the nascent lagging strand reads from 5' to 3'. Because the original strands of DNA are antiparallel, and only one continuous new strand can be synthesised at the 3' end of the leading strand due to the intrinsic 5'-3' polarity of DNA polymerases, the other strand must grow discontinuously in the opposite direction. Regarding the lagging strand, the result of this strand's discontinuous replication is the production of a series of short sections of DNA called Okazaki fragments.

Each Okazaki fragment is initiated near the replication fork at an RNA primer created by primase, and extended by DNA polymerase III. In eukaryotes, lagging strand synthesis is carried out by the DNA polymerase δ. The primer is later removed by enzymes that have endonucleolytic activity such as Ribonuclease H (RNAse H), flap endonucleases (FENs) and Dna2 helicase/nucleases. In prokaryotes the FEN nuclease is a domain of DNA

polymerase I while in eukaryotes FENs are separate enzymes. The excised RNA bases are replaced with DNA by DNA polymerase I in prokaryotes or DNA polymerase δ in eukaryotes. Adjoining fragments are then linked together by DNA ligase, using phosphodiester bonds, to create a continuous strand of DNA.

Okazaki used a pulse chase type experiment to confirm discontinuous strand replication. He took actively replicating DNA, then added "hot" tritiated nucleotides for a short pulse of about 5 seconds. During the 5 seconds the radioactive nucleotides were incorporated into the growing DNA strands. After the pulse Okazaki chased with "cold" un-labeled nucleotides for varying amounts of time and quickly isolated the DNA. Then the DNA was centrifuged and analyzed for radioactivity. What Okazaki found was that with short chases of about 7 to 15 seconds most of the radioactivity was found in the small fragments higher in the tube after centrifuge. However with longer chases more radioactivity was found in the lower, larger strands. This confirmed that during synthesis first small fragments are formed on the lagging strand, then later these fragments are combined and incorporated into much larger strands. The small fragments found on the lagging strand are called Okazaki fragments.

Chapter- 18

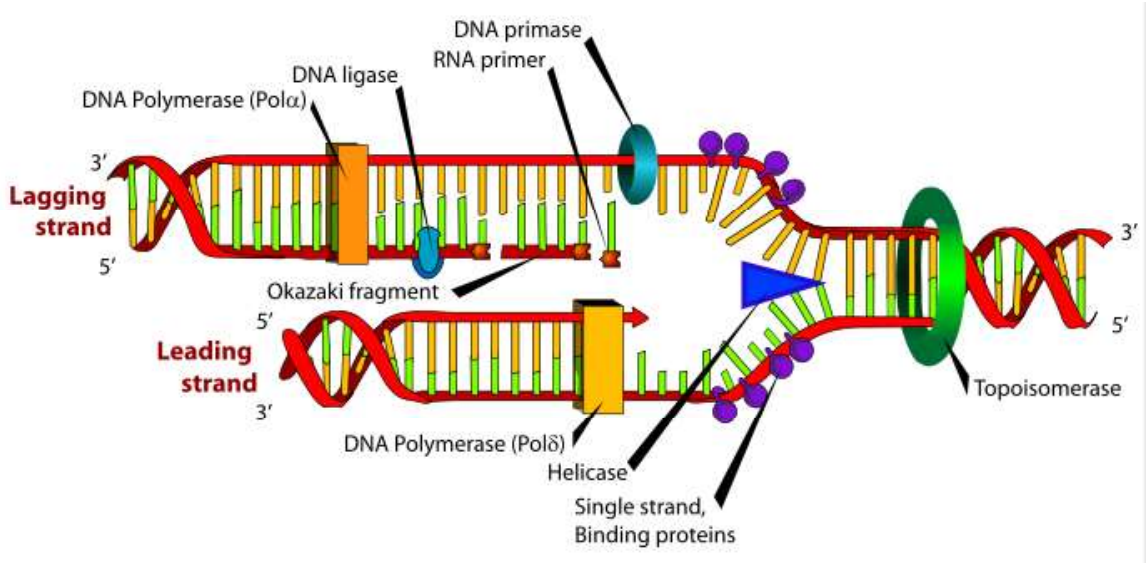
Primer (Molecular Biology)

Primer is a strand of nucleic acid that serves as a starting point for DNA synthesis. They are required because the enzymes that catalyze replication, DNA polymerases, can only add new nucleotides to an existing strand of DNA. The polymerase starts replication at the 3'-end of the primer, and copies the opposite strand.

In most cases of natural DNA replication, the primer for DNA synthesis and replication is a short strand of RNA (which can be made *de novo*).

Many of the laboratory techniques of biochemistry and molecular biology that involve DNA polymerase, such as DNA sequencing and the polymerase chain reaction (PCR), require DNA primers. These primers are usually short, chemically synthesized oligonucleotides, with a length of about twenty bases. They are hybridized to a target DNA, which is then copied by the polymerase.

Mechanism in vivo



The DNA replication fork. RNA primer labeled at top.

The lagging strand is that strand of the DNA double helix that is orientated in a 5' to 3' manner. Therefore, its complement must be synthesized in a 3'→5' manner. Because DNA polymerase III cannot synthesize in the 3'→5' direction, the lagging strand is synthesized in short segments known as Okazaki fragments. Along the lagging strand's template, primase builds RNA primers in short bursts. DNA polymerases are then able to use the free 3'-OH groups on the RNA primers to synthesize DNA in the 5'→3' direction.

The RNA fragments are then removed by DNA polymerase I for prokaryotes or DNA polymerase δ for eukaryotes (different mechanisms are used in eukaryotes and prokaryotes) and new deoxyribonucleotides are added to fill the gaps where the RNA was present. DNA ligase then joins the deoxyribonucleotides together, completing the synthesis of the lagging strand.

Primer removal

In eukaryotic primer removal, DNA polymerase δ extends the Okazaki fragment in 5' to 3' direction, and when it encounters the RNA primer from the previous Okazaki fragment, displacing the 5′ end of the primer into a single-stranded RNA flap, which is removed by nuclease cleavage. Cleavage of the RNA flaps involves either endonuclease 1 (FEN1) cleavage of short flaps, or coating of long flaps by the single-stranded DNA binding protein replication protein A (RPA) and sequential cleavage by Dna2 nuclease and FEN1.

This mechanism is a potential explanation to how HIV virus can transform its genome into double stranded DNA from the RNA-DNA formed after reverse transcription of its RNA. However, the HIV-encoded reverse transcriptase has own ribonuclease activity that degrades the viral RNA during the synthesis of cDNA, as well as DNA-dependent DNA polymerase activity that copies the sense cDNA strand into a *antisense* DNA to form a double-stranded DNA intermediate.

Uses of synthetic primers

DNA sequencing is used to determine the nucleotides in a DNA strand; the chain termination method (dideoxy sequencing or Sanger method) uses a primer as a start marker for the chain reaction.

In PCR, primers are used to determine the DNA fragment to be amplified by the PCR process. The length of primers is usually not more than 30 (usually 18–24) nucleotides, and they need to match the beginning and the end of the DNA fragment to be amplified. They direct replication towards each other – the extension of one primer by polymerase then becomes the template for the other, leading to an exponential increase in the target segment.

It is worth noting that primers are not essentially always necessary for DNA synthesis and can in fact be used by viral polymerases, e.g. influenza, for RNA synthesis.

PCR primer design

Pairs of primers should have similar melting temperatures since annealing in a PCR occurs for both simultaneously. A primer with a T_m significantly higher than the reaction's annealing temperature may mishybridize and extend at an incorrect location along the DNA sequence, while T_m significantly lower than the annealing temperature may fail to anneal and extend at all.

Primer sequences need to be chosen to uniquely select for a region of DNA, avoiding the possibility of mishybridization to a similar sequence nearby. A commonly used method is BLAST search whereby all the possible regions to which a primer may bind can be seen. Both the nucleotide sequence as well as the primer itself can be BLAST searched. The free NCBI tool Primer-BLAST integrates primer design tool and BLAST search into one application, so does commercial software product such as Beacon Designer. Mononucleotide repeats should be avoided, as loop formation can occur and contribute to mishybridization. Primers should not easily anneal with other primers in the mixture (either other copies of same or the reverse direction primer); this phenomenon can lead to the production of 'primer dimer' products contaminating the mixture. Primers should also not anneal strongly to themselves, as internal hairpins and loops could hinder the annealing with the template DNA.

When designing a primer for use in TA cloning, efficiency can be increased by adding AG tails to the 5' and the 3' end.

Degenerate primers

Sometimes *degenerate primers* are used. These are actually mixtures of similar, but not identical primers. They may be convenient if the same gene is to be amplified from different organisms, as the genes themselves are probably similar but not identical. The other use for degenerate primers is when primer design is based on protein sequence. As several different codons can code for one amino acid, it is often difficult to deduce which codon is used in a particular case. Therefore primer sequence corresponding to the amino acid isoleucine might be "ATH", where A stands for adenine, T for thymine, and H for adenine, thymine, or cytosine, according to the genetic code for each codon, using the IUPAC symbols for degenerate bases. Use of degenerate primers can greatly reduce the specificity of the PCR amplification. The problem can be partly solved by using touchdown PCR.

Degenerate primers are widely used and extremely useful in the field of microbial ecology. They allow for the amplification of genes from thus far uncultivated microorganisms or allow the recovery of genes from organisms where genomic information is not available. Usually, degenerate primers are designed by aligning gene sequencing found in GenBank. Differences among sequences are accounted for by using IUPAC degeneracies for individual bases. PCR primers are then synthesized as a mixture of primers corresponding to all permutations.