DNA Repair

Jaydon Baez

First Edition, 2012

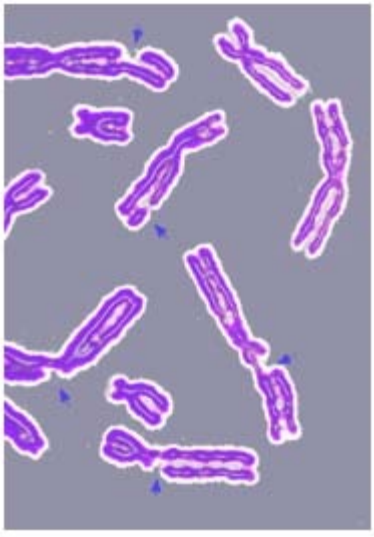
ISBN 978-81-323-3270-1

Published by:
Research World
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

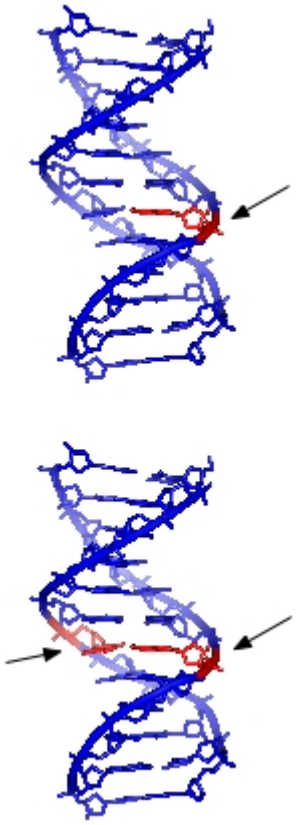
It is important to distinguish between DNA damage and mutation, the two major types of error in DNA. DNA damages and mutation are fundamentally different. Damages are physical abnormalities in the DNA, such as single- and double-strand breaks, 8-hydroxydeoxyguanosine residues, and polycyclic aromatic 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.

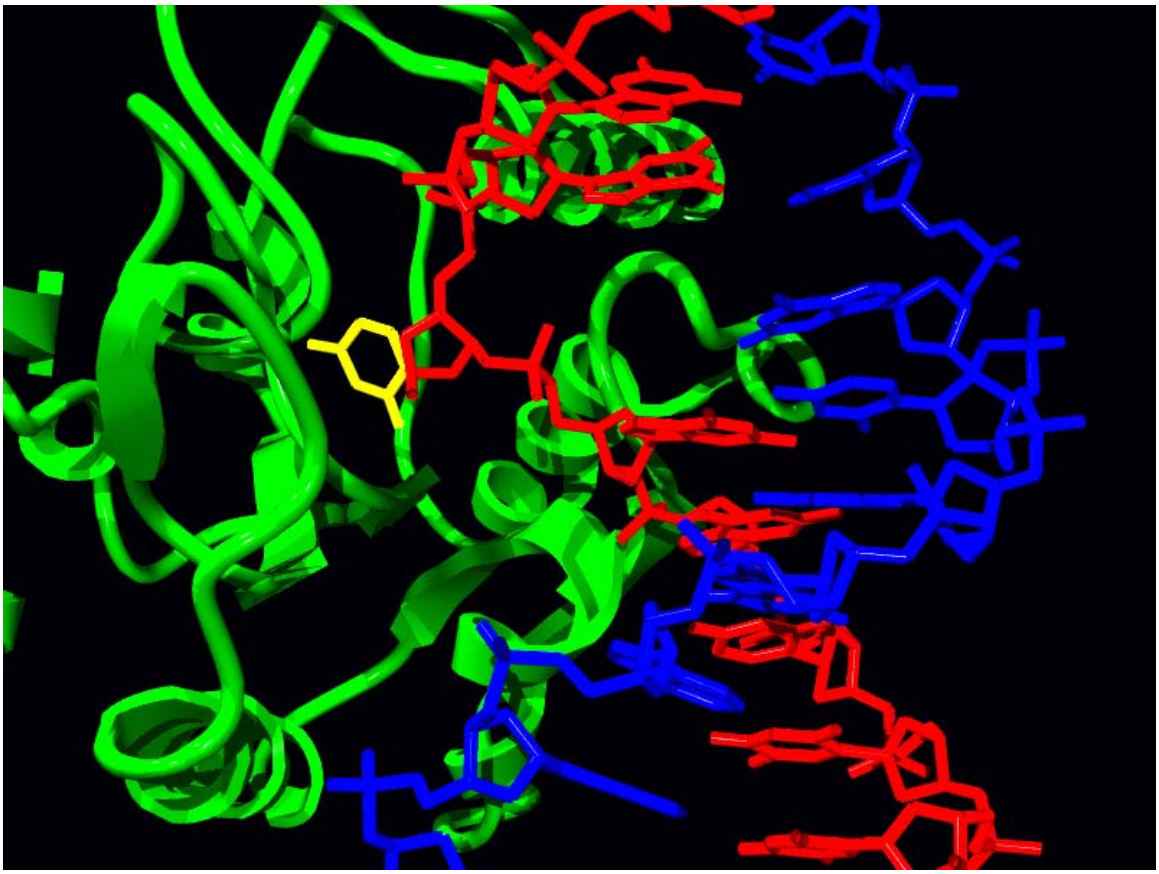
Damage to DNA alters the spatial configuration of the helix, and such alterations can be detected by the cell. Once damage is localized, specific DNA repair molecules bind at or near the site of damage, inducing other molecules to bind and form a complex that enables the actual repair to take place.

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.

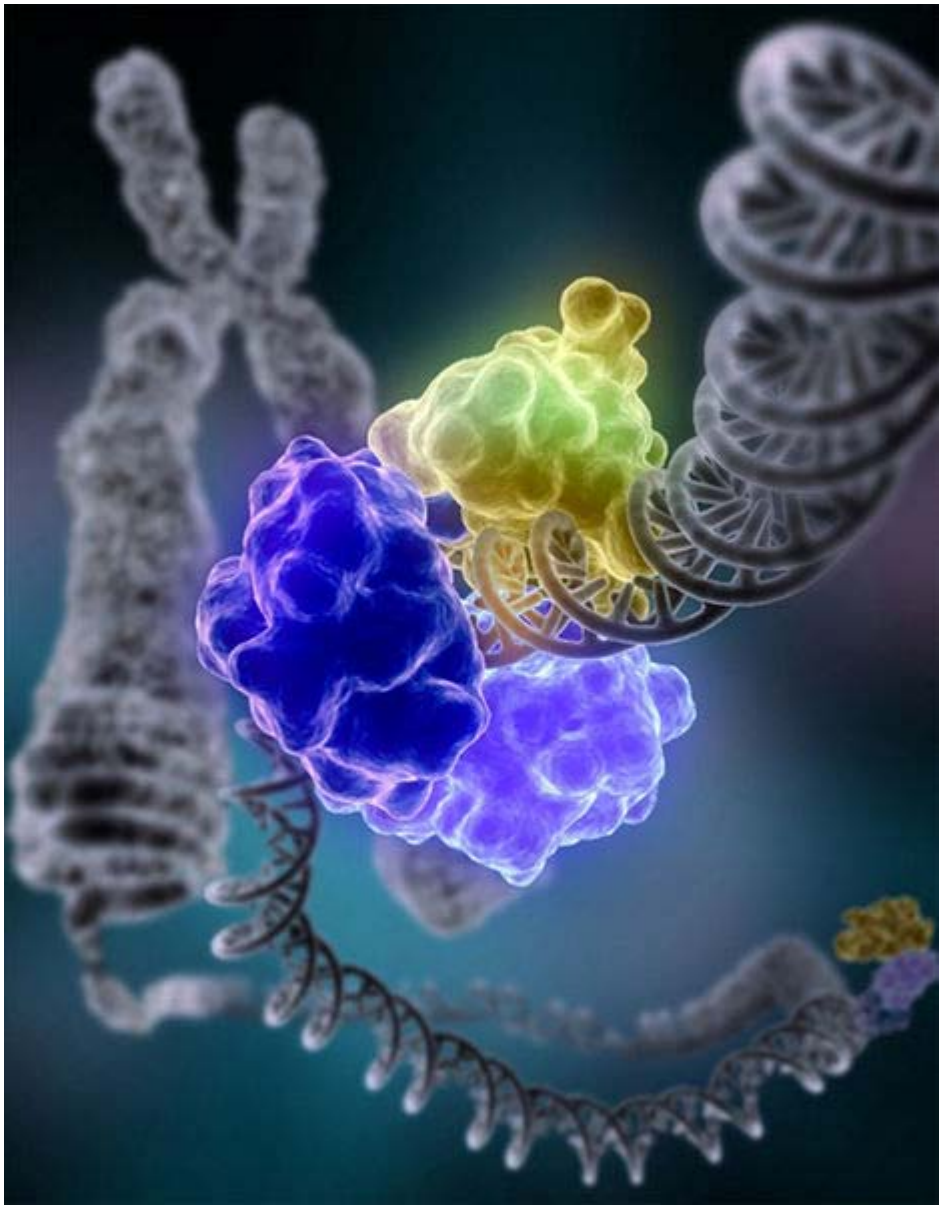
When only one of the two strands of a double helix has a defect, the other strand can be used as a template to guide the correction of the damaged strand. In order to repair

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.

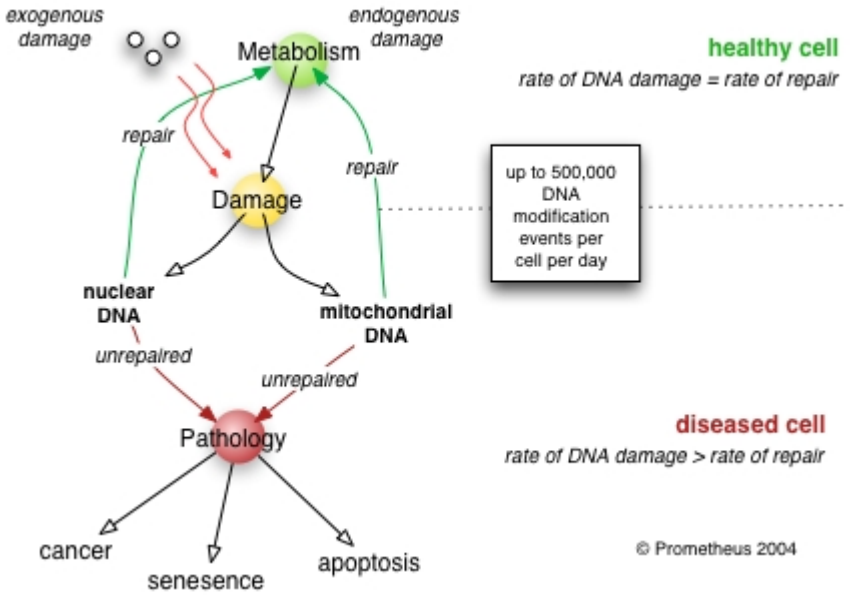
Eukaryotic transcriptional responses to DNA damage

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

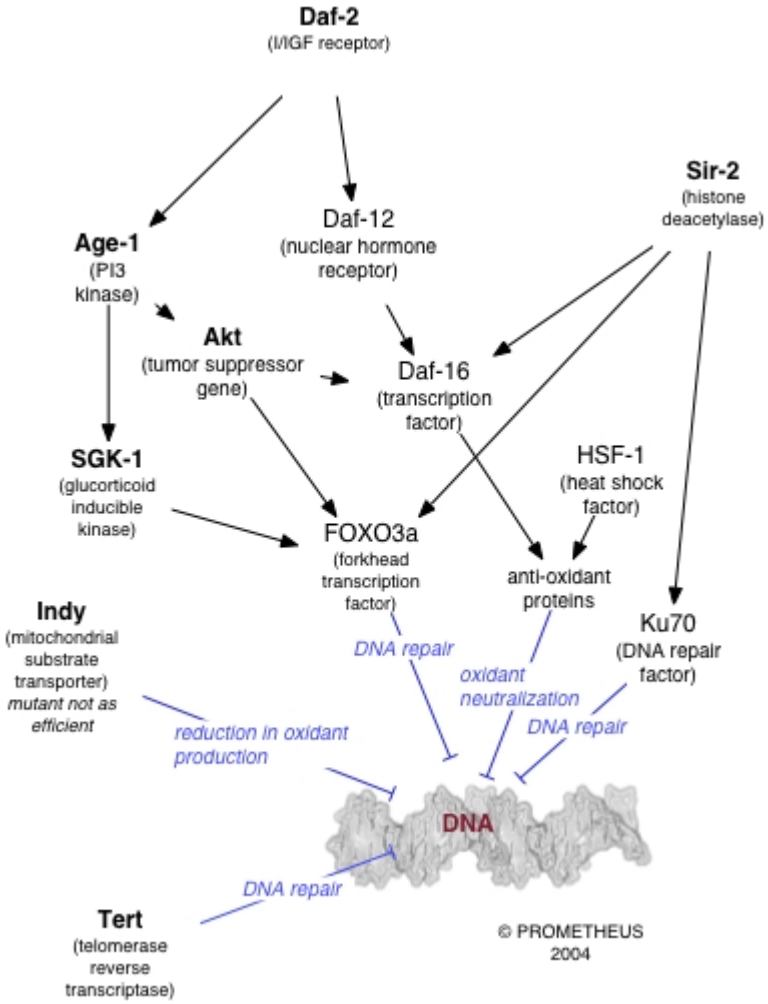


DNA repair rate is an important determinant of cell pathology

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.

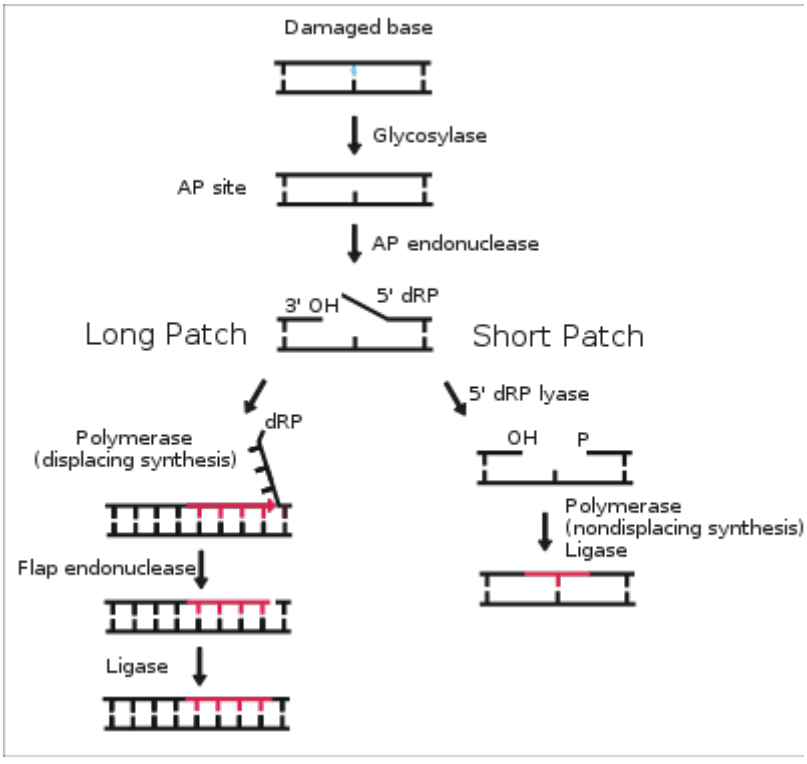
Rate of evolutionary change

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

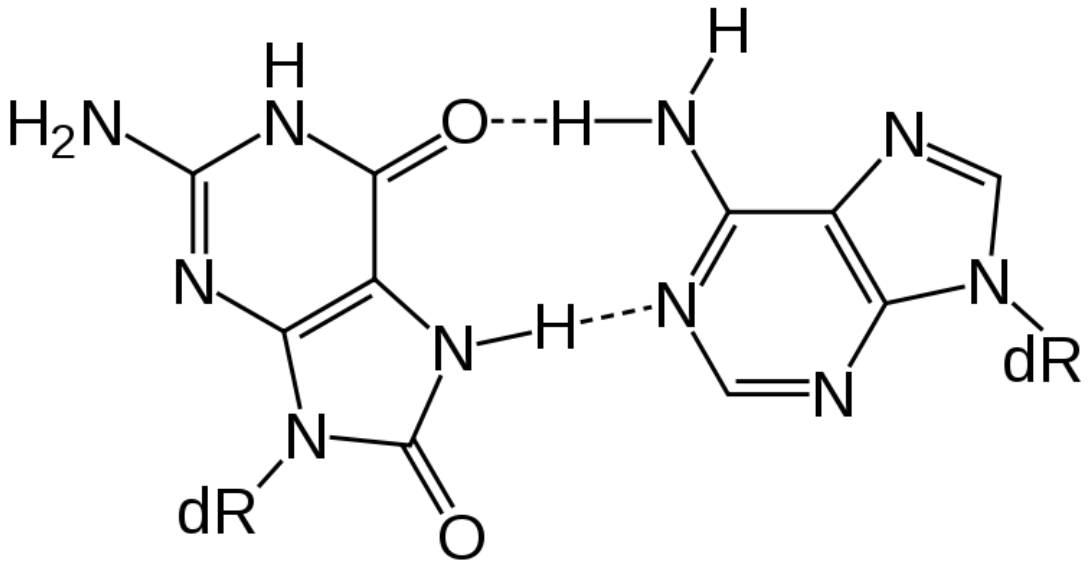


Basic steps of base excision repair

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.

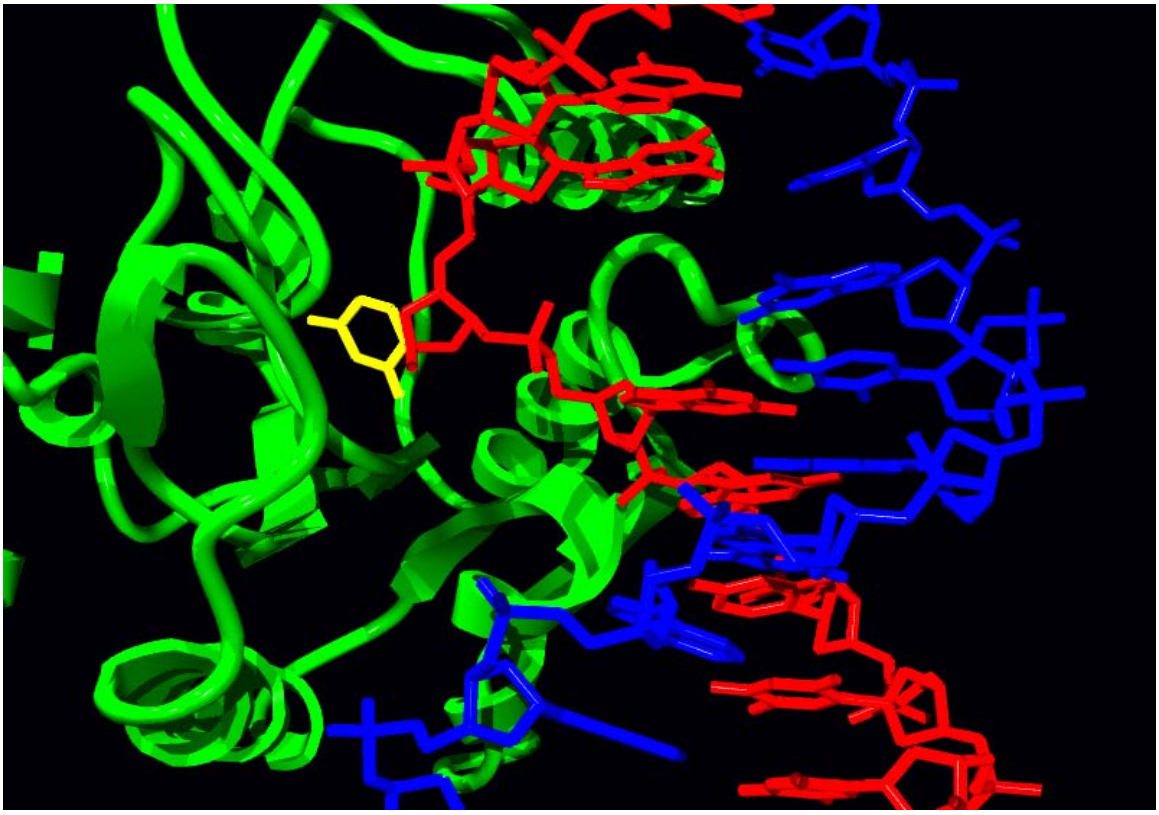
The choice between long-patch and short-patch repair

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.

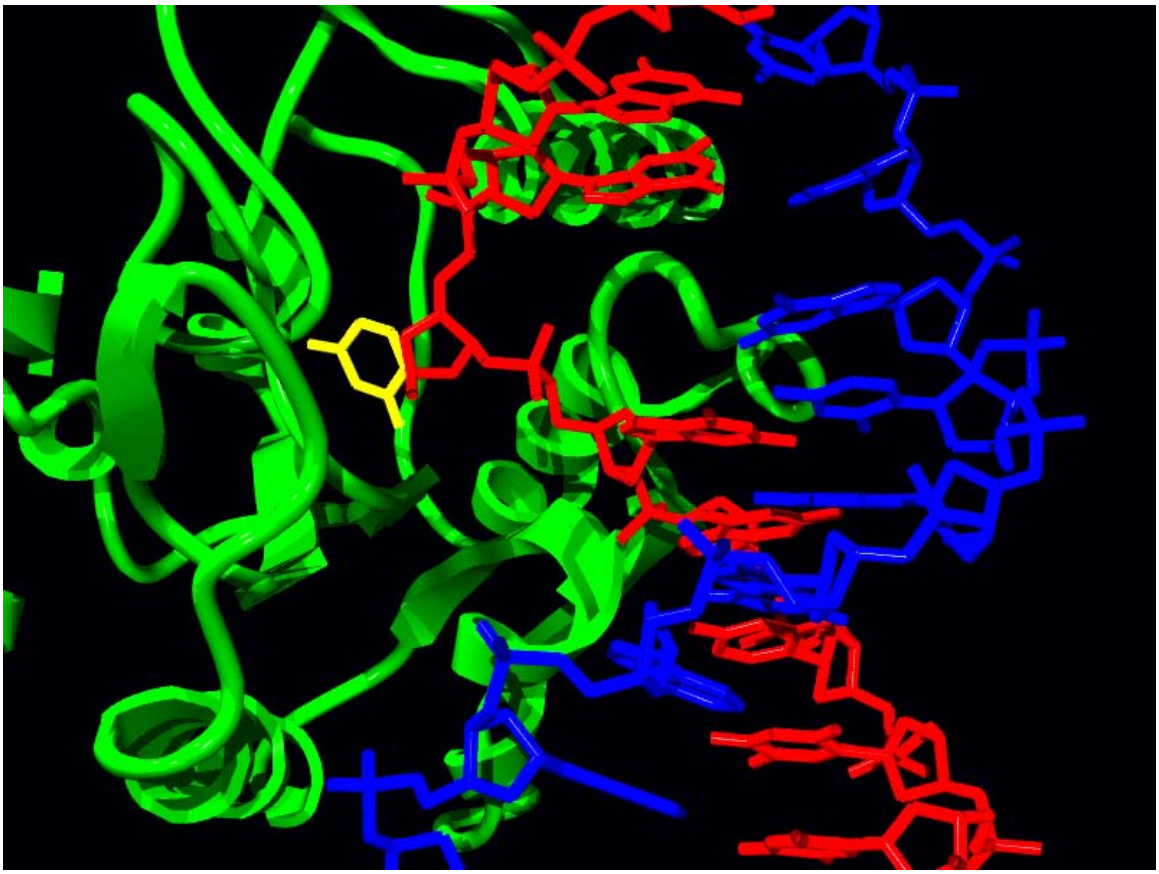
Glycosylases in bacteria, yeast and humans

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

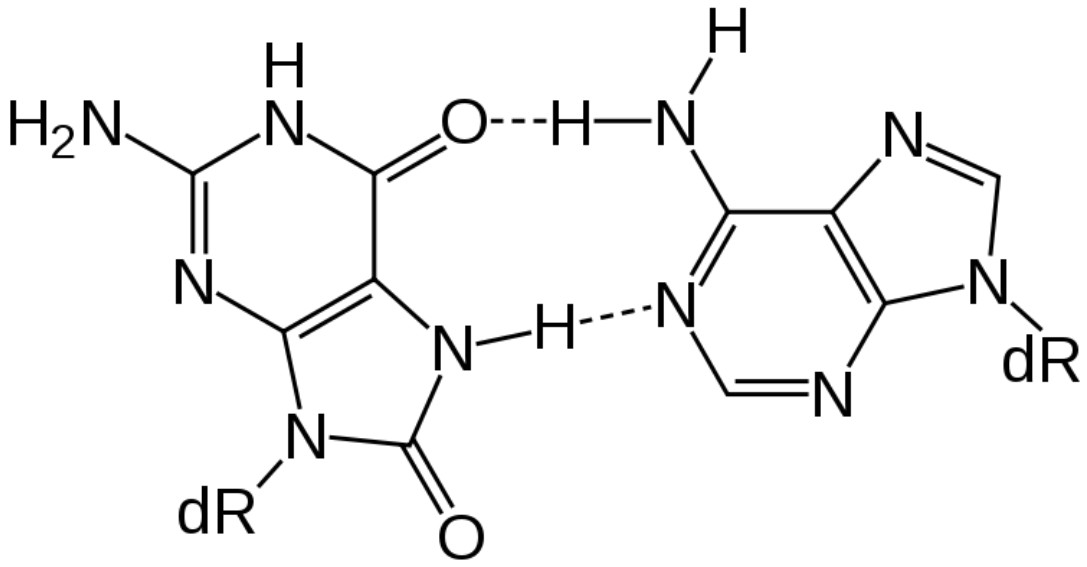


Structure of the base-excision repair enzyme uracil-DNA glycosylase. The uracil residue is shown in yellow.

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.

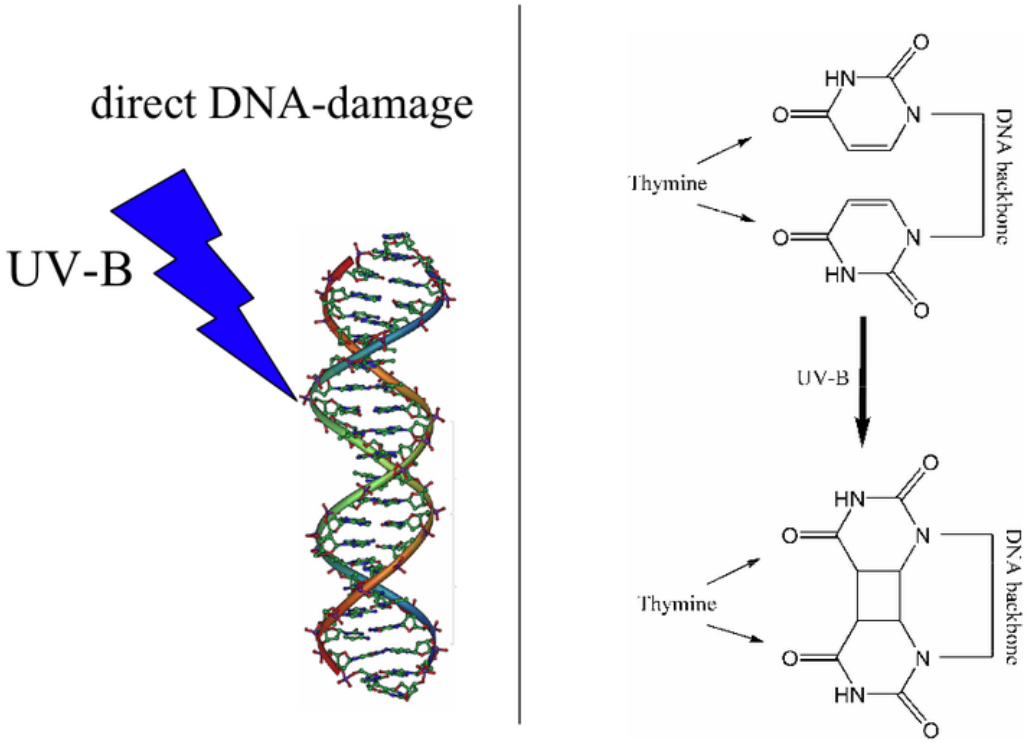
Conclusions

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.

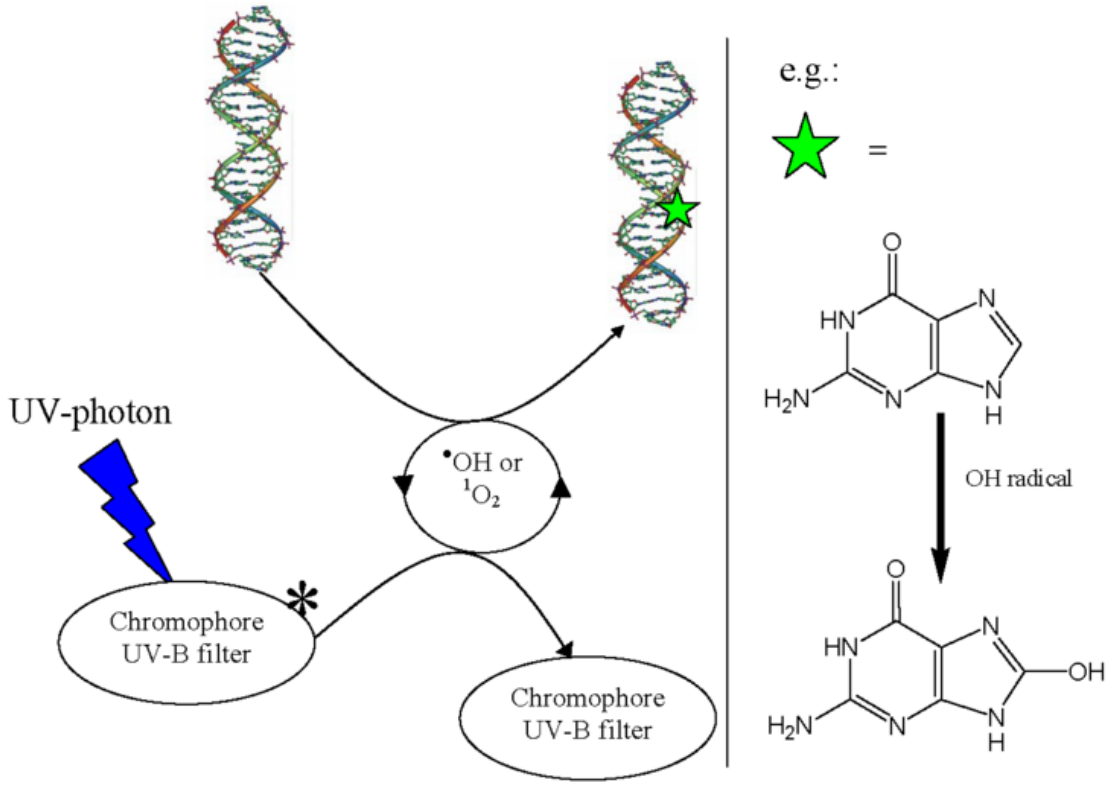
Sunscreen and melanoma

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

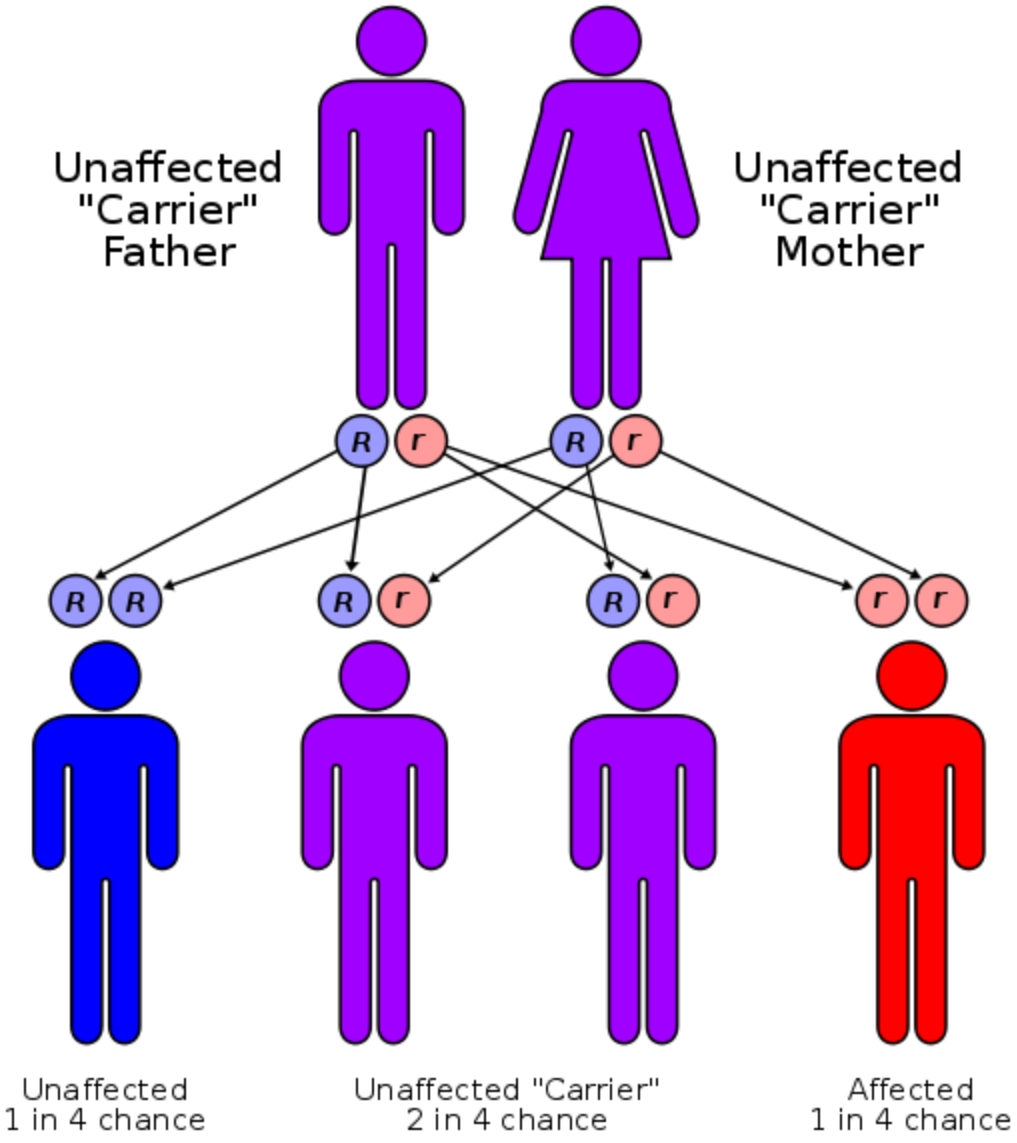
Direct DNA damage is confined to areas that can be reached by UV-B light. In contrast free radicals can travel through the body and affect other areas - possibly even inner organs. The traveling nature of the indirect DNA damage can be seen in the fact that the malignant melanoma can occur in places 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 E, V, XPE	29881	278740	DDB2	11p12-p11	**Xeroderma pigmentosum group E**.
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

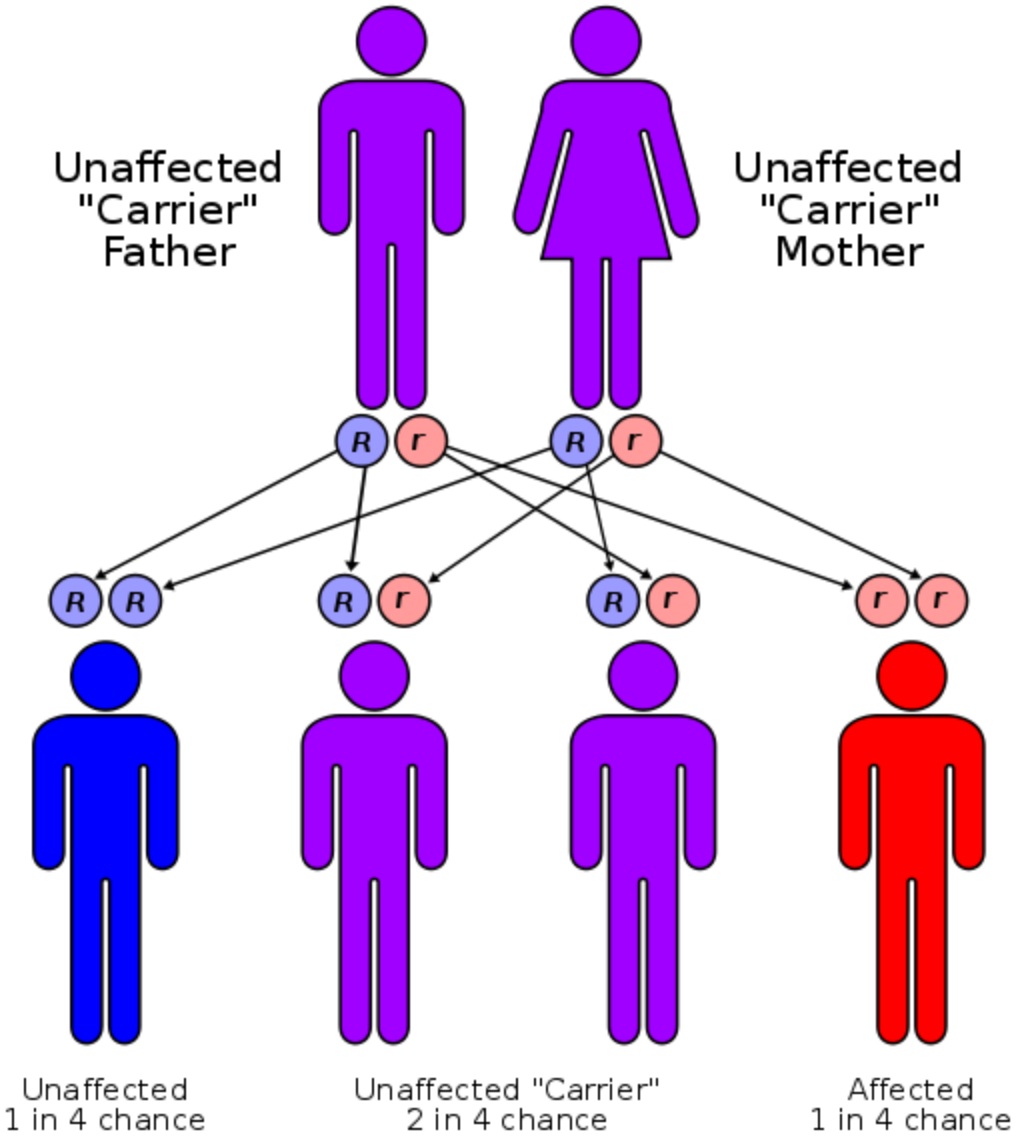
may develop cancer at any age. The average age of cancer diagnoses in the cohort is approximately 25 years old.

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

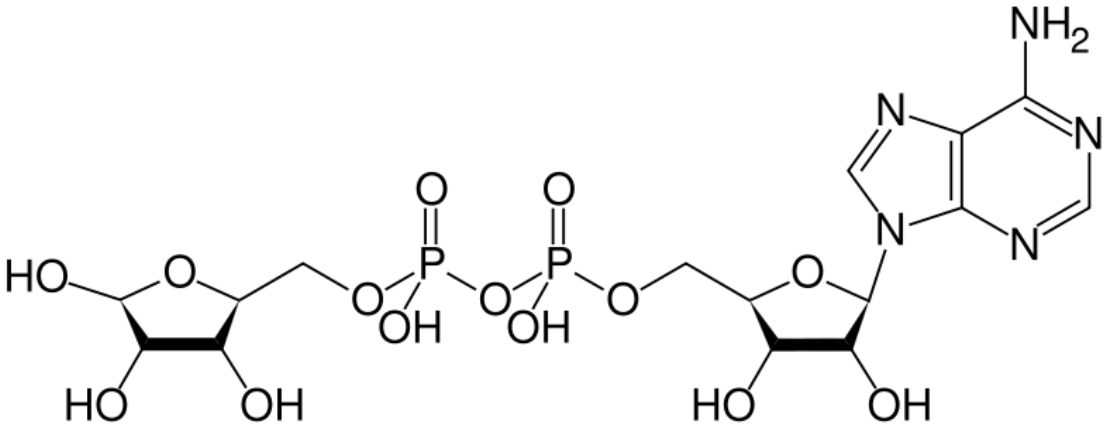
For the first time pyruvate kinase M2 enzyme was reported with two missense mutations, H391Y and K422R, found in cells from Bloom syndrome patients, prone to develop cancer. Results show that despite the presence of mutations in the inter-subunit contact domain, the K422R and H391Y mutant proteins maintained 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

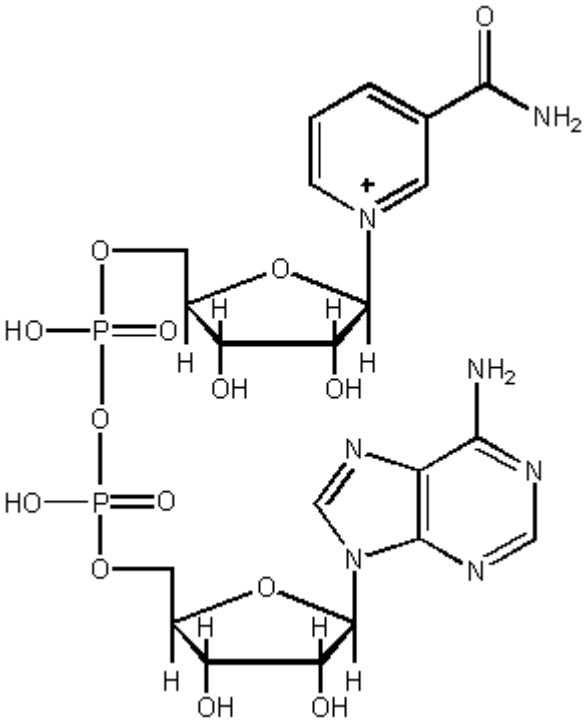
It is hypothesized that an aberrant PKM2 protein can justify other features of Bloom syndrome.

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

Non-Homologous End Joining and Proliferating Cell Nuclear Antigen

Non-homologous end joining

Non-homologous end joining (**NHEJ**) is a pathway that repairs double-strand breaks in DNA. NHEJ is referred to as "non-homologous" because the break ends are directly ligated without the need for a homologous template, in contrast to homologous recombination, which requires a homologous sequence to guide repair. The term "non-homologous end joining" was coined in 1996 by Moore and Haber.

NHEJ typically utilizes short homologous DNA sequences called microhomologies to guide repair. These microhomologies are often present in single-stranded overhangs on the ends of double-strand breaks. When the overhangs are perfectly compatible, NHEJ usually repairs the break accurately. Imprecise repair leading to loss of nucleotides can also occur, but is much more common when the overhangs are not compatible. Inappropriate NHEJ can lead to translocations and telomere fusion, hallmarks of tumor cells.

NHEJ is evolutionarily conserved throughout all kingdoms of life and is the predominant double-strand break repair pathway in mammalian cells. In budding yeast (*Saccharomyces cerevisiae*), however, homologous recombination dominates when the organism is grown under common laboratory conditions.

When the NHEJ pathway is inactivated, double-strand breaks can be repaired by a more error-prone pathway called microhomology-mediated end joining (MMEJ). In this pathway, end resection reveals short microhomologies on either side of the break, which are then aligned to guide repair. This contrasts with classical NHEJ, which typically uses microhomologies already exposed in single-stranded overhangs on the DSB ends. Repair by MMEJ therefore leads to deletion of the DNA sequence between the microhomologies.

In bacteria

Many species of bacteria, including *Escherichia coli*, lack an end joining pathway and thus rely completely on homologous recombination to repair double-strand breaks. NHEJ proteins have been identified in a number of bacteria, however, including *Bacillus subtilis*, *Mycobacterium tuberculosis*, and *Mycobacterium smegmatis*. Bacteria utilize a remarkably compact version of NHEJ in which all of the required activities are contained in only two proteins: a Ku homodimer and the multifunctional ligase/polymerase/nuclease LigD. In mycobacteria, NHEJ is much more error prone than in yeast, with bases often added to and deleted from the ends of double-strand breaks during repair. Many of the bacteria that possess NHEJ proteins spend a significant portion of their life cycle in a stationary haploid phase, in which a template for recombination is not available. NHEJ may have evolved to help these organisms survive DSBs induced during dessication. Corndog and Omega, two related mycobacteriophages of *Mycobacterium smegmatis*, also encode Ku homologs and exploit the NHEJ pathway to recircularize their genomes during infection. Unlike homologous recombination, which has been studied extensively in bacteria, NHEJ was originally discovered in eukaryotes and was only identified in prokaryotes in the past decade.

In eukaryotes

In contrast to bacteria, NHEJ in eukaryotes utilizes a number of proteins, which participate in the following steps:

End binding and tethering

In yeast, the Mre11-Rad50-Xrs2 (MRX) complex is recruited to DSBs early and is thought to promote bridging of the DNA ends. The corresponding mammalian complex of Mre11-Rad50-Nbs1 (MRN) is also involved in NHEJ, but it may function at multiple steps in the pathway beyond simply holding the ends in proximity. DNA-PKcs is also thought to participate in end bridging during mammalian NHEJ.

Eukaryotic Ku is a heterodimer consisting of Ku70 and Ku80, and forms a complex with DNA-PKcs, which is present in mammals but absent in yeast. Ku is a basket-shaped molecule that slides onto the DNA end and translocates inward. Ku may function as a docking site for other NHEJ proteins, and is known to interact with the DNA ligase IV complex and XLF.

End processing

End processing involves removal of damaged or mismatched nucleotides by nucleases and resynthesis by DNA polymerases. This step is not necessary if the ends are already compatible and have 3' hydroxyl and 5' phosphate termini.

Little is known about the function of nucleases in NHEJ. Artemis is required for opening the hairpins that are formed on DNA ends during V(D)J recombination, a specific type of

NHEJ, and may also participate in end trimming during general NHEJ. Mre11 has nuclease activity, but it seems to be involved in homologous recombination, not NHEJ.

The X family DNA polymerases Pol λ and Pol μ (Pol4 in yeast) fill gaps during NHEJ. Yeast lacking Pol4 are unable to join 3' overhangs that require gap filling, but remain proficient for gap filling at 5' overhangs. This is because the primer terminus used to initiate DNA synthesis is less stable at 3' overhangs, necessitating a specialized NHEJ polymerase.

Ligation

The DNA ligase IV complex, consisting of the catalytic subunit DNA ligase IV and its cofactor XRCC4 (Dnl4 and Lif1 in yeast), performs the ligation step of repair. XLF, also known as Cernunnos, is homologous to yeast Nej1 and is also required for NHEJ. While the precise role of XLF is unknown, it interacts with the XRCC4/DNA ligase IV complex and likely participates in the ligation step. Recent evidence suggests that XLF re-adenylates DNA ligase IV after ligation, recharging the ligase and allowing it to catalyze a second ligation.

Other

In yeast, Sir2 was originally identified as an NHEJ protein, but is now known to be required for NHEJ only because it is required for the transcription of Nej1.

Regulation

The choice between NHEJ and homologous recombination for repair of a double-strand break is regulated at the initial step in recombination, 5' end resection. In this step, the 5' strand of the break is degraded by nucleases to create long 3' single-stranded tails. DSBs that have not been resected can be rejoined by NHEJ, but resection of even a few nucleotides strongly inhibits NHEJ and effectively commits the break to repair by recombination. NHEJ is active throughout the cell cycle, but is most important during G1 when no homologous template for recombination is available. This regulation is accomplished by the cyclin-dependent kinase Cdk1 (Cdc28 in yeast), which is turned off in G1 and expressed in S and G2. Cdk1 phosphorylates the nuclease Sae2, allowing resection to initiate.

V(D)J recombination

NHEJ plays a critical role in V(D)J recombination, the process by which B-cell and T-cell receptor diversity is generated in the vertebrate immune system. In V(D)J recombination, hairpin-capped double-strand breaks are created by the RAG1/RAG2 nuclease, which cleaves the DNA at recombination signal sequences. These hairpins are then opened by the Artemis nuclease and joined by NHEJ. A specialized DNA polymerase called terminal deoxynucleotidyl transferase (TdT), which is only expressed in lymph tissue, adds nontemplated nucleotides to the ends before the break is joined.

This process couples "variable" (V), "diversity" (D), and "joining" (J) regions, which when assembled together create the variable region of a B-cell or T-cell receptor gene. Unlike typical cellular NHEJ, in which accurate repair is the most favorable outcome, error-prone repair in V(D)J recombination is beneficial in that it maximizes diversity in the coding sequence of these genes. Patients with mutations in NHEJ genes are unable to produce functional B cells and T cells and suffer from severe combined immunodeficiency (SCID).

At telomeres

Telomeres are normally protected by a "cap" that prevents them from being recognized as double-strand breaks. Loss of capping proteins causes telomere shortening and inappropriate joining by NHEJ, producing dicentric chromosomes which are then pulled apart during mitosis. Paradoxically, some NHEJ proteins are involved in telomere capping. For example, Ku localizes to telomeres and its deletion leads to shortened telomeres. Ku is also required for subtelomeric silencing, the process by which genes located near telomeres are turned off.

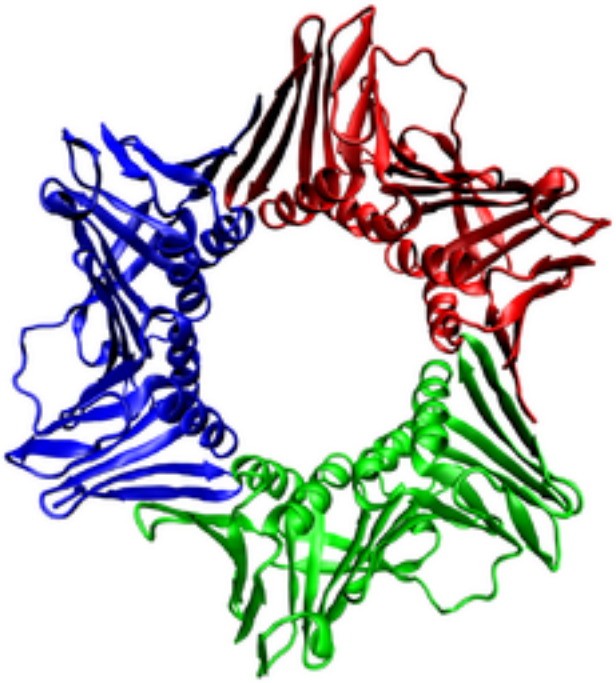
Consequences of dysfunction

Several human syndromes are associated with dysfunctional NHEJ. Hypomorphic mutations in LIG4 and XLF cause LIG4 syndrome and XLF-SCID, respectively. These syndromes share many features including cellular radiosensitivity, microcephaly and severe combined immunodeficiency (SCID) due to defective V(D)J recombination. Loss-of-function mutations in Artemis also cause SCID, but these patients do not show the neurological defects associated with LIG4 or XLF mutations. The difference in severity may be explained by the roles of the mutated proteins. Artemis is a nuclease and is thought to be required only for repair of DSBs with damaged ends, whereas DNA Ligase IV and XLF are required for all NHEJ events.

Many NHEJ genes have been knocked out in mice. Deletion of XRCC4 or LIG4 causes embryonic lethality in mice, indicating that NHEJ is essential for viability in mammals. In contrast, mice lacking Ku or DNA-PKcs are viable, probably because low levels of end joining can still occur in the absence of these components. All NHEJ mutant mice show a SCID phenotype, sensitivity to ionizing radiation, and neuronal apoptosis.

Proliferating cell nuclear antigen

Proliferating cell nuclear antigen



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

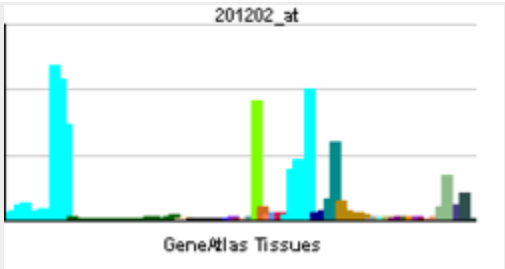
Available structures	
PDB	1axc, 1u76, 1u7b, 1ul1, 1vyj, 1vym, 1w60

Identifiers	
Symbols	PCNA; MGC8367
External IDs	OMIM: 176740 MGI: 97503 HomoloGene: 1945 GeneCards: PCNA Gene

Gene Ontology	
Molecular function	• DNA binding • protein binding • DNA polymerase processivity factor activity • MutLalpha complex binding
Cellular component	• cyclin-dependent protein kinase holoenzyme complex • nucleus

	• nuclear lamina • DNA replication factor C complex • PCNA complex
Biological process	• regulation of progression through cell cycle • DNA replication • regulation of DNA replication • DNA repair • base-excision repair, gap-filling • intracellular protein transport • cell proliferation • phosphoinositide-mediated signaling

RNA expression pattern



Orthologs

Species	**Human**	**Mouse**
Entrez	5111	18538
Ensembl	ENSG00000132646	ENSMUSG00000027342
UniProt	P12004	Q542J9
RefSeq (mRNA)	NM_002592	NM_011045
RefSeq (protein)	NP_002583	NP_035175
Location (UCSC)	Chr 20: 5.04 - 5.06 Mb	Chr 2: 131.94 - 131.94 Mb

Proliferating Cell Nuclear Antigen, commonly known as **PCNA**, is a protein that acts as a processivity factor for DNA polymerase δ in eukaryotic cells. It achieves this

processivity by encircling the DNA, thus creating a topological link to the genome. It is an example of a DNA clamp.

The protein encoded by this gene is found in the nucleus and is a cofactor of DNA polymerase delta. The encoded protein acts as a homotrimer and helps increase the processivity of leading strand synthesis during DNA replication. In response to DNA damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway. Two transcript variants encoding the same protein have been found for this gene. Pseudogenes of this gene have been described on chromosome 4 and on the X chromosome.

Expression in the nucleus during DNA synthesis

PCNA was originally identified as an antigen that is expressed in the nuclei of cells during the DNA synthesis phase of the cell cycle. Part of the protein was sequenced and that sequence was used to allow isolation of a cDNA clone. PCNA helps hold DNA polymerase delta (**Pol δ**) to DNA. PCNA is clamped to DNA through the action of replication factor C (RFC), which is a heteropentameric member of the AAA+ class of ATPases. Expression of PCNA is under the control of E2F transcription factor-containing complexes.

Role in DNA repair

Since DNA polymerase delta is involved in resynthesis of excised damaged DNA strands during DNA repair, PCNA is important for both DNA synthesis and DNA repair.

PCNA is also involved in the DNA damage tolerance pathway known as post-replication repair (PRR). In PRR, there are two sub-pathways: (1) a translesion pathway, which is carried out by specialised DNA polymerases that are able to incorporate damaged DNA bases into their active sites (unlike the normal replicative polymerase, which stall), and hence bypass the damage, and (2) a proposed "template switch" pathway that is thought to involve damage bypass by recruitment of the homologous recombination machinery. PCNA is pivotal to the activation of these pathways and the choice as to which pathway is utilised by the cell. PCNA becomes post-translationally modified by ubiquitin. Mono-ubiquitin of lysine number 164 on PCNA activates the translesion synthesis pathway. Extension of this mono-ubiquitin by a non-canonical lysine-63-linked poly-ubiquitin chain on PCNA is thought to activate the template switch pathway. Furthermore, sumoylation (by small ubiquitin-like modifier, SUMO) of PCNA lysine-164 (and to a lesser extent, lysine-127) **inhibits** the template switch pathway. This antagonistic effect occurs because sumoylated PCNA recruits a DNA helicase called Srs2, which has a role in disrupting Rad51 nucleoprotein filaments fundamental for initiation of homologous recombination.

PCNA-binding proteins

DNA polymerases • Clamp loader • Flap endonuclease • DNA ligase • Topoisomerase • Replication licensing factor • E3 ubiquitin ligases • E2 SUMO-conjugating enzyme • Helicases, ATPases • Mismatch repair enzymes • Base excision repair enzymes • Nucleotide excision repair enzyme • Poly ADP ribose polymerase • Histone chaperone • Chromatin remodeling factor • Histone acetyltransferase • Histone deacetyltransferase • DNA methyltransferase • Sister-chromatid cohesion factors • Protein kinases • Cell-cycle regulators • Apoptotic factors

Interactions

PCNA has been shown to interact with Ku70, MSH3, Werner syndrome ATP-dependent helicase, RFC2, RFC3, RFC1, RFC4, RFC5, GADD45G, CDC25C, MUTYH, Flap structure-specific endonuclease 1, Cyclin O, CHTF18, Y box binding protein 1, Cyclin D1, Annexin A2, MSH6, DNMT1, HDAC1, KCTD13, XRCC1, Cyclin-dependent kinase 4, Ku80, HUS1, GADD45A, POLD2, ING1, POLH, KIAA0101, POLDIP2, EP300, MCL1, POLD3, Cyclin-dependent kinase inhibitor 1C, POLL, Ubiquitin C and P21.

Uses

Antibodies against proliferating cell nuclear antigen (PCNA) or monoclonal antibody termed Ki-67 can be used for grading of different neoplasms, *e.g.* astrocytoma. They can be of diagnostic and prognostic value.

Chapter- 10

Homologous Recombination

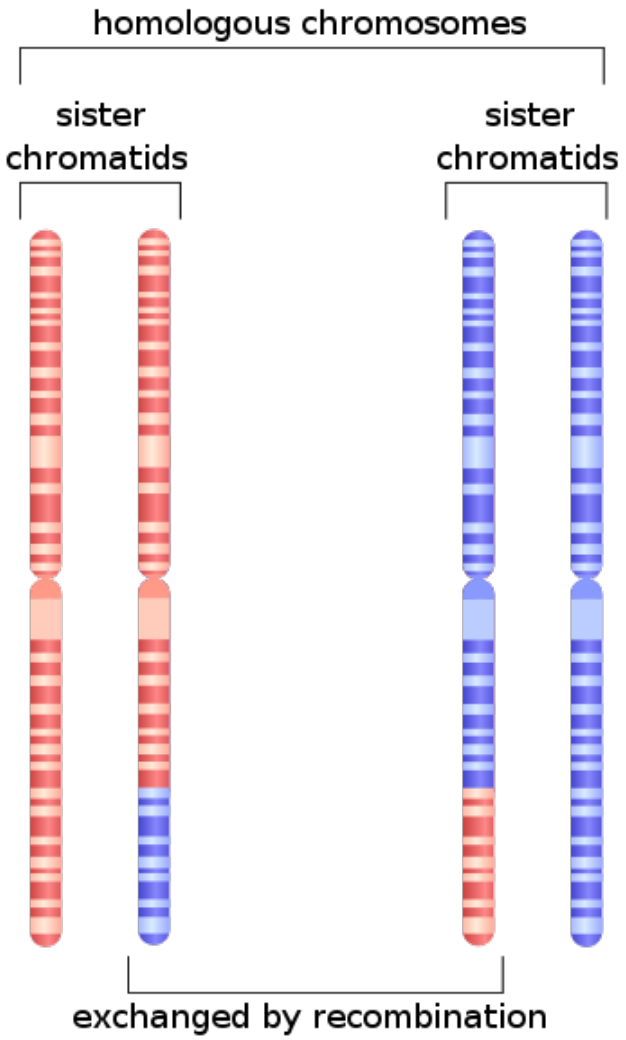


Figure 1. During meiosis, homologous recombination can produce new combinations of genes as shown here between similar but not identical copies of human chromosome 1.

Homologous recombination is a type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. It is most widely used by cells to accurately repair harmful breaks that occur on both strands of DNA, known as double-strand breaks. Homologous recombination also produces new combinations of DNA sequences during meiosis, the process by which eukaryotes make gamete cells, like sperm and egg cells in animals. These new combinations of DNA represent genetic variation in offspring, which in turn enables populations to adapt during the course of evolution. Homologous recombination is also used in horizontal gene transfer to exchange genetic material between different strains and species of bacteria and viruses.

Although homologous recombination varies widely among different organisms and cell types, most forms of it involve the same basic steps. After a double-strand break occurs, sections of DNA around the 5' ends of the break are cut away in a process called *resection*. In the *strand invasion* step that follows, an overhanging 3' end of the broken DNA molecule then "invades" a similar or identical DNA molecule that is not broken. After strand invasion, one or two cross-shaped structures called Holliday junctions connect the two DNA molecules. Depending on how the two junctions are cut by enzymes, the type of homologous recombination that occurs in meiosis results in either chromosomal crossover or non-crossover. Homologous recombination that occurs during DNA repair tends to result in non-crossover products, in effect restoring the damaged DNA molecule as it existed before the double-strand break.

Homologous recombination is conserved across all three domains of life as well as viruses, suggesting that it is a nearly universal biological mechanism. The discovery of genes for homologous recombination in protists—a diverse group of eukaryotic microorganisms—has been interpreted as evidence that meiosis emerged early in the evolution of eukaryotes. Since their dysfunction has been strongly associated with increased susceptibility to several types of cancer, the proteins that facilitate homologous recombination are topics of active research. Homologous recombination is also used as a technique in molecular biology for introducing genetic changes into target organisms. The development of gene targeting techniques that rely on homologous recombination was the subject of the 2007 Nobel Prize for Physiology or Medicine.

History and discovery

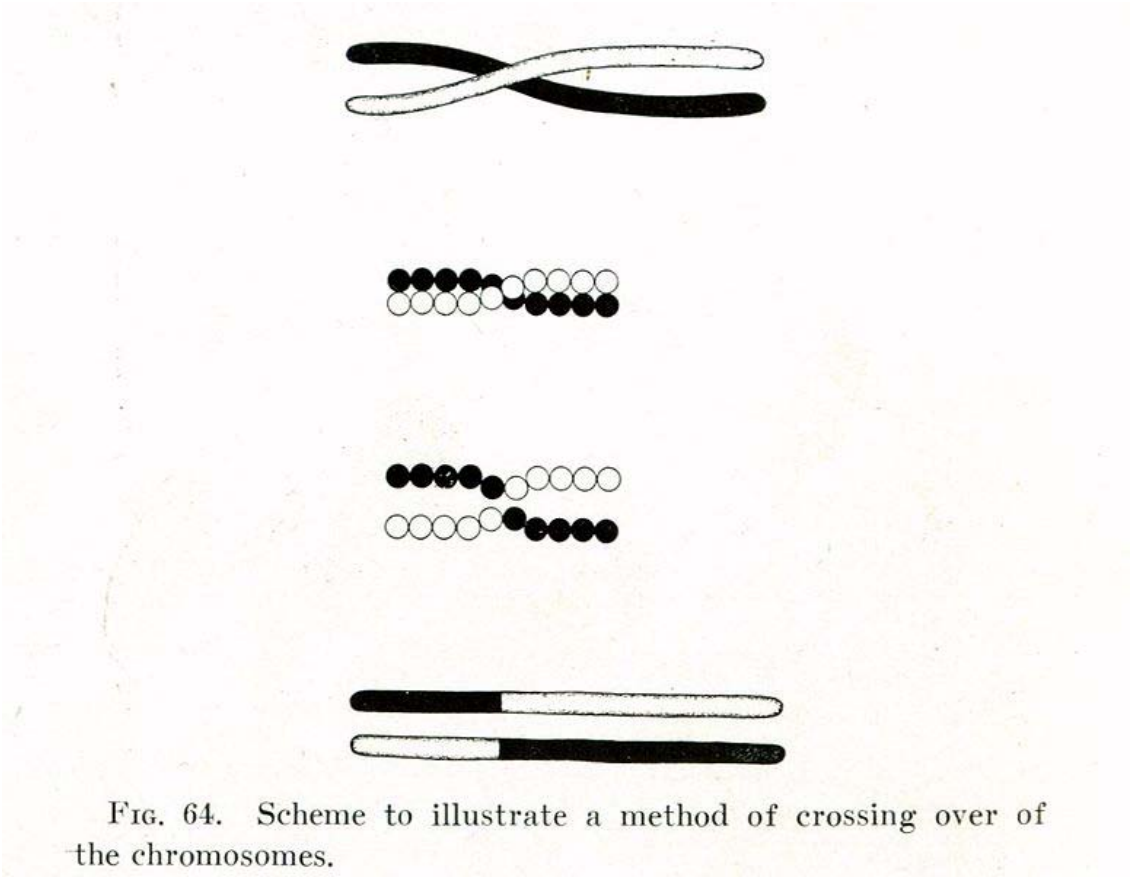


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.

Figure 2. An early illustration of crossing over from Thomas Hunt Morgan

In the early 1900s, William Bateson and Reginald Punnett found an exception to one of the principles of inheritance originally described by Gregor Mendel in the 1860s. In contrast to Mendel's notion that traits are independently assorted when passed from parent to child—for example that a cat's hair color and its tail length are inherited independent of each other—Bateson and Punnett showed that certain genes associated with physical traits can be inherited together, or genetically linked. In 1911, after observing that linked traits could on occasion be inherited separately, Thomas Hunt Morgan suggested that "crossovers" can occur between linked genes, where one of the linked genes physically crosses over to a different chromosome. Two decades later, Barbara McClintock and Harriet Creighton demonstrated that chromosomal crossover occurs during meiosis, the process of cell division by which sperm and egg cells are made. Within the same year as McClintock's discovery, Curt Stern showed that crossing over—later called "recombination"—could also occur in somatic cells like white blood cells and skin cells that divide through mitosis.

In 1947, the microbiologist Joshua Lederberg showed that bacteria—which had been assumed to reproduce only asexually through binary fission—are capable of genetic recombination, which is more similar to sexual reproduction. This work established *E.*

coli as a model organism in genetics, and helped Lederberg win the 1958 Nobel Prize in Physiology or Medicine. Building on studies in fungi, in 1964 Robin Holliday proposed a model for recombination in meiosis which introduced key details of how the process can work, including the exchange of material between chromosomes through Holliday junctions. In 1983, Jack Szostak and colleagues presented a model now known as the DSBR pathway, which accounted for observations not explained by the Holliday model. During the next decade, experiments in *Drosophila*, budding yeast and mammalian cells led to the emergence of other models of homologous recombination, called SDSA pathways, which do not always rely on Holliday junctions.

In eukaryotes

Homologous recombination is essential to cell division in eukaryotes like plants, animals, fungi and protists. In cells that divide through mitosis, homologous recombination repairs double-strand breaks in DNA caused by ionizing radiation or DNA-damaging chemicals. Left unrepaired, these double-strand breaks can cause large-scale rearrangement of chromosomes in somatic cells, which can in turn lead to cancer.

In addition to repairing DNA, homologous recombination also helps produce genetic diversity when cells divide in meiosis to become specialized gamete cells—sperm or egg cells in animals, pollen or ovules in plants, and spores in fungi. It does so by facilitating chromosomal crossover, in which regions of similar but not identical DNA are exchanged between homologous chromosomes. This creates new, possibly beneficial combinations of genes, which can give offspring an evolutionary advantage. Chromosomal crossover begins when a protein called Spo11 makes a targeted double-strand break in DNA. The sites of these double-strand breaks often occur at recombination hotspots, regions in chromosomes that are about 1,000–2,000 base pairs in length and have high rates of recombination. The absence of a recombination hotspot between two genes on the same chromosome often means that those genes will be inherited by future generations in equal proportion. This represents linkage between the two genes greater than would be expected from genes that independently assort during meiosis.

Timing within the cell cycle

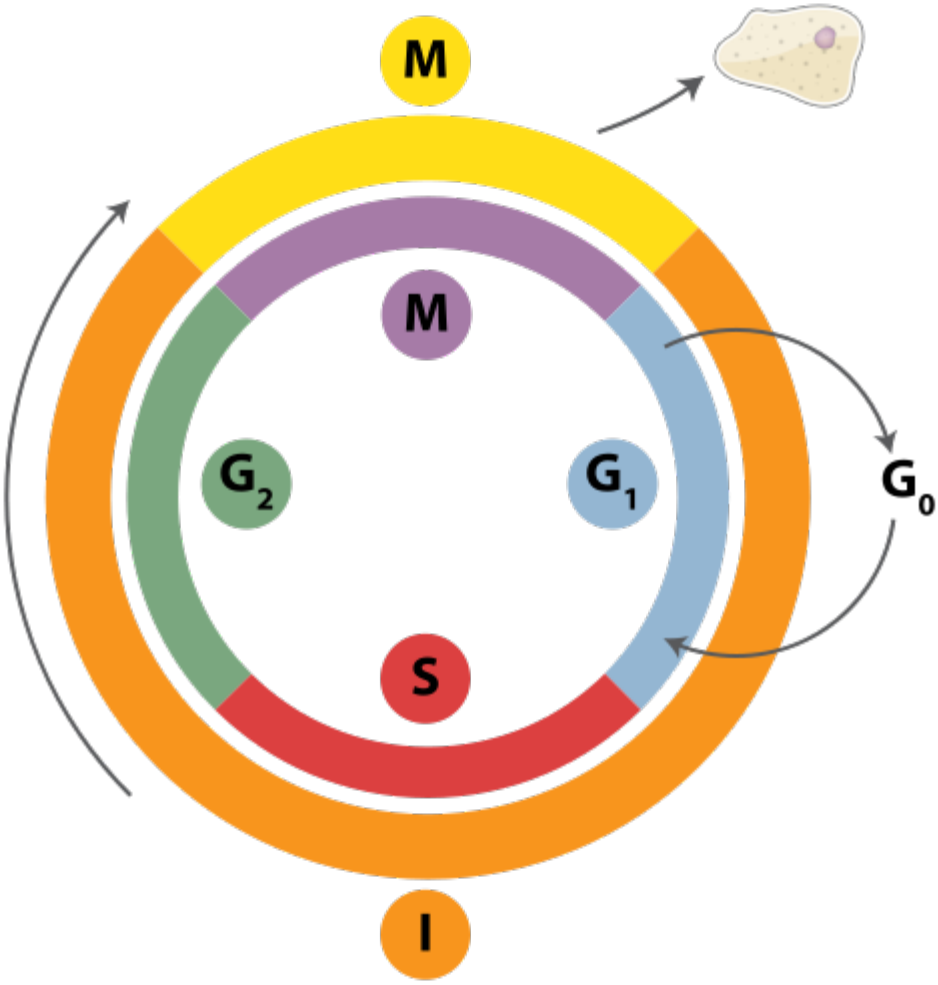


Figure 3. Homologous recombination repairs DNA before the cell enters mitosis (M phase). It occurs only during and shortly after DNA replication, during the S and G_2 phases of the cell cycle.

Double-strand breaks can be repaired through homologous recombination or through non-homologous end joining (NHEJ). NHEJ is a DNA repair mechanism which, unlike homologous recombination, does not require a long homologous sequence to guide repair. Whether homologous recombination or NHEJ is used to repair double-strand breaks is largely determined by the phase of cell cycle. Homologous recombination repairs DNA before the cell enters mitosis (M phase). It occurs during and shortly after DNA replication, in the S and G_2 phases of the cell cycle, when sister chromatids are more easily available. Compared to homologous chromosomes, which are similar to another chromosome but often have different alleles, sister chromatids are an ideal template for homologous recombination because they are an identical copy of a given chromosome. In contrast to homologous recombination, NHEJ is predominant in the G_1 phase of the cell cycle, when the cell is growing but not yet ready to divide. It occurs less frequently after the G_1 phase, but maintains at least some activity throughout the cell

cycle. The mechanisms that regulate homologous recombination and NHEJ throughout the cell cycle vary widely between species.

Cyclin-dependent kinases (CDKs), which modify the activity of other proteins by adding phosphate groups to (that is, phosphorylating) them, are important regulators of homologous recombination in eukaryotes. When DNA replication begins in budding yeast, the cyclin-dependent kinase Cdc28 begins homologous recombination by phosphorylating the Sae2 protein. After being so activated by the addition of a phosphate, Sae2 uses its endonuclease activity to make a clean cut near a double-strand break in DNA. This allows a three-part protein known as the MRX complex to bind to DNA, and begins a series of protein-driven reactions that exchange material between two DNA molecules.

Models

Two primary models for how homologous recombination repairs double-strand breaks in DNA are the DSBR pathway (sometimes called the *double Holliday junction model*) and the synthesis-dependent strand annealing (SDSA) pathway. The two pathways are similar in their first several steps. After a double-strand break occurs, the MRX complex (MRN complex in humans) binds to DNA on either side of the break. Next a resection, in which DNA around the 5' ends of the break is cut back, is carried out in two distinct steps. In the first step of resection, the MRX complex recruits the Sae2 protein. The two proteins then trim back the 5' ends on either side of the break to create short 3' overhangs of single-strand DNA. In the second step, 5'→3' resection is continued by the Sgs1 helicase and the Exo1 and Dna2 nucleases. As a helicase, Sgs1 "unzips" the double-strand DNA, while Exo1 and Dna2's nuclease activity allows them to cut the single-stranded DNA produced by Sgs1.

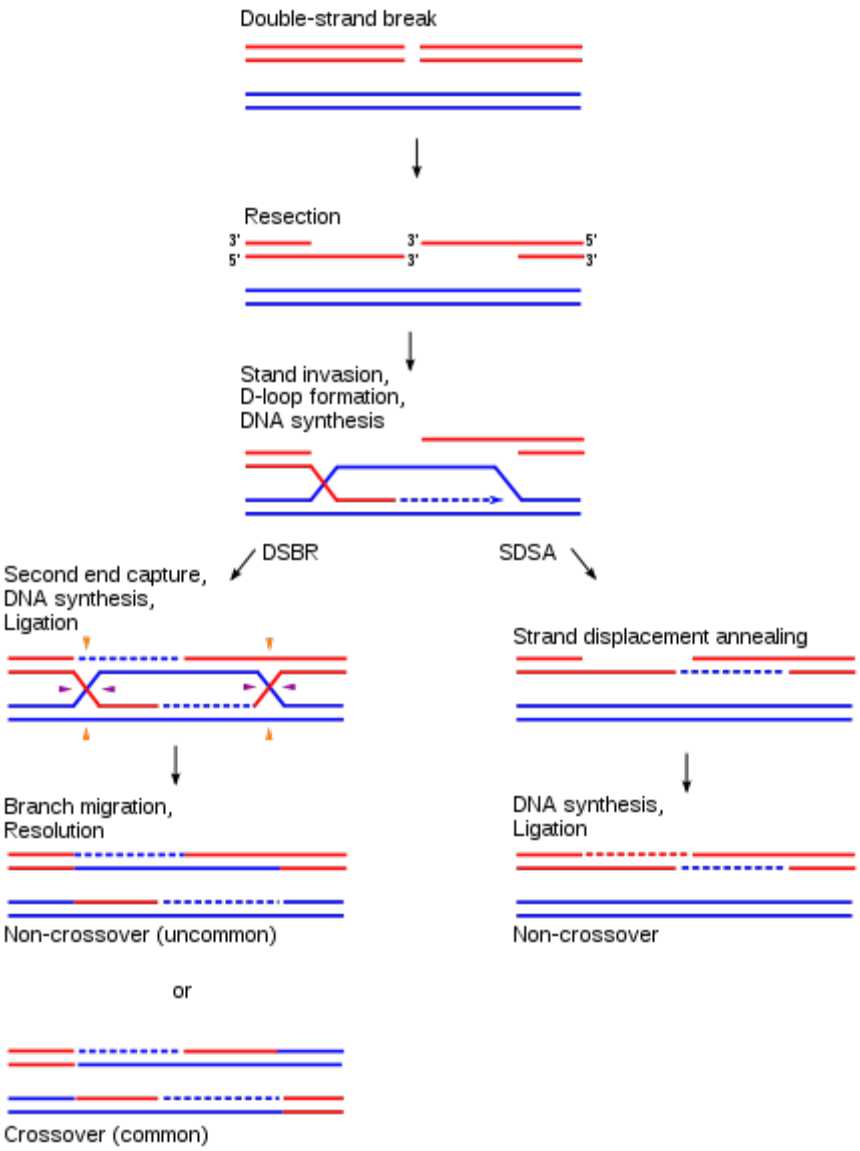


Figure 4. The DSBR and SDSA pathways follow the same initial steps, but diverge thereafter. The DSBR pathway most often results in chromosomal crossover (bottom left), while SDSA always ends with non-crossover products (bottom right).

The RPA protein, which has high affinity for single-stranded DNA, then binds the 3' overhangs. With the help of several other proteins that mediate the process, the Rad51 protein (and Dmc1, in meiosis) then forms a filament of nucleic acid and protein on the single strand of DNA coated with RPA. This nucleoprotein filament then begins searching for DNA sequences similar to that of the 3' overhang. After finding such a sequence, the single-stranded nucleoprotein filament moves into (invades) the similar or identical recipient DNA duplex in a process called *strand invasion*. In cells that divide through mitosis, the recipient DNA duplex is generally a sister chromatid, which is identical to the damaged DNA molecule and provides a template for repair. In meiosis,

however, the recipient DNA tends to be from a similar but not necessarily identical homologous chromosome. A displacement loop (D-loop) is formed during strand invasion between the invading 3' overhang strand and the homologous chromosome. After strand invasion, a DNA polymerase extends the end of the invading 3' strand by synthesizing new DNA. This changes the D-loop to a cross-shaped structure known as a Holliday junction. Following this, more DNA synthesis occurs on the invading strand (i.e., one of the original 3' overhangs), effectively restoring the strand on the homologous chromosome that was displaced during strand invasion.

DSBR pathway

After the stages of resection, strand invasion and DNA synthesis, the DSBR and SDSA pathways become distinct. The DSBR pathway is unique in that the second 3' overhang (which was not involved in strand invasion) also forms a Holliday junction with the homologous chromosome. The double Holliday junctions are then converted into recombination products by nicking endonucleases, a type of restriction endonuclease which cuts only one DNA strand. The DSBR pathway commonly results in crossover, though it can sometimes result in non-crossover products. Because of this tendency for chromosomal crossover, the DSBR pathway is a likely model of how homologous recombination occurs during meiosis.

Whether recombination in the DSBR pathway results in chromosomal crossover is determined by how the double Holliday junction is cut, or "resolved". Chromosomal crossover will occur if one Holliday junction is cut on the crossing strand and the other Holliday junction is cut on the non-crossing strand (in Figure 4, along the horizontal purple arrowheads at one Holliday junction and along the vertical orange arrowheads at the other). Alternatively, if the two Holliday junctions are cut on the crossing strands (along the horizontal purple arrowheads at both Holliday junctions in Figure 4), then chromosomes without crossover will be produced.

SDSA pathway

Homologous recombination via the SDSA pathway occurs in cells that divide through mitosis and results in non-crossover products. In this model, the invading 3' strand is extended along the recipient DNA duplex by a DNA polymerase, and is released as the Holliday junction between the donor and recipient DNA molecules slides in a process called *branch migration*. The newly synthesized 3' end of the invading strand is then able to anneal to the other 3' overhang in the damaged chromosome through complementary base pairing. After the strands anneal, a small flap of DNA can sometimes remain. Any such flaps are removed, and the SDSA pathway finishes with the resealing, also known as *ligation*, of any remaining single-stranded gaps.

SSA pathway

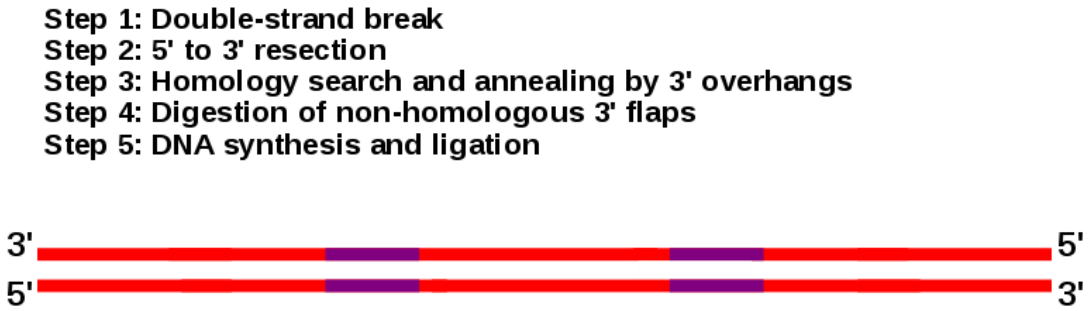


Figure 5. Recombination via the SSA pathway occurs between two repeat elements (purple) on the same DNA duplex, and results in deletions of genetic material.

The single-strand annealing (SSA) pathway of homologous recombination repairs double-strand breaks between two repeat sequences. The SSA pathway is unique in that it does not require a separate similar or identical molecule of DNA, like the DSBR or SDSA pathways of homologous recombination. Instead, the SSA pathway only requires a single DNA duplex, and uses the repeat sequences as the identical sequences that homologous recombination needs for repair. The pathway is relatively simple in concept: after two strands of the same DNA duplex are cut back around the site of the double-strand break, the two resulting 3' overhangs then align and anneal to each other, restoring the DNA as a continuous duplex.

As DNA around the double-strand break is cut back, the single-stranded 3' overhangs being produced are coated with the RPA protein, which prevents the 3' overhangs from sticking to themselves. A protein called Rad52 then binds each of the repeat sequences on either side of the break, and aligns them to enable the two complementary repeat sequences to anneal. After annealing is complete, leftover non-homologous flaps of the 3' overhangs are cut away by a set of nucleases, known as Rad1/Rad10, which are brought to the flaps by the Saw1 and Slx4 proteins. New DNA synthesis fills in any gaps, and ligation restores the DNA duplex as two continuous strands. The DNA sequence between the repeats is always lost, as is one of the two repeats. The SSA pathway is considered mutagenic since it results in such deletions of genetic material.

BIR pathway

During DNA replication, double-strand breaks can sometimes be encountered at replication forks as DNA helicase unzips the template strand. These defects are repaired in the *break-induced replication* (BIR) pathway of homologous recombination. The precise molecular mechanisms of the BIR pathway remain unclear. Three proposed mechanisms have strand invasion as an initial step, but differ in how they model the migration of the D-loop and later phases of recombination.

The BIR pathway can also help to maintain the length of telomeres, regions of DNA at the end of eukaryotic chromosomes, in the absence of (or in cooperation with) telomerase. Without working copies of the telomerase enzyme, telomeres typically shorten with each cycle of mitosis, which eventually blocks cell division and leads to senescence. In budding yeast cells where telomerase have been inactivated through mutations, two types of "survivor" cells have been observed to avoid senescence longer than expected by elongating their telomeres through BIR pathways.

Maintaining telomere length is critical for cell immortalization, a key feature of cancer. Most cancers maintain telomeres by upregulating telomerase. However, in several types of human cancer, a BIR-like pathway helps to sustain some tumors by acting as an alternative mechanism of telomere maintenance. This fact has lead scientists to investigate whether such recombination-based mechanisms of telomere maintenance could thwart anti-cancer drugs like telomerase inhibitors.

In bacteria

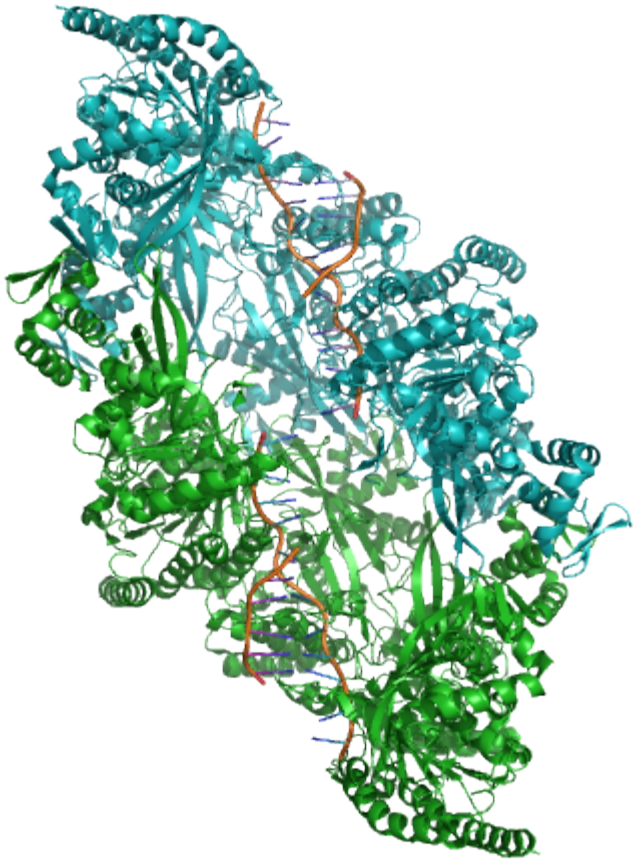


Figure 6. Crystal structure of two chains of the RecA protein bound to DNA. A double-strand break and two adjacent 3' overhangs are visible.

Homologous recombination is a major DNA repair process in bacteria. It is also important for producing genetic diversity in bacterial populations, although the process differs substantially from meiotic recombination, which brings about diversity in eukaryotic genomes. Homologous recombination has been most studied and is best understood for *Escherichia coli*. Double-strand DNA breaks in bacteria are repaired by the RecBCD pathway of homologous recombination. Breaks that occur on only one of the two DNA strands, known as single-strand gaps, are thought to be repaired by the RecF pathway. Both the RecBCD and RecF pathways include a series of reactions known as *branch migration*, in which single DNA strands are exchanged between two intercrossed molecules of duplex DNA, and *resolution*, in which those two intercrossed molecules of DNA are cut apart and restored to their normal double-stranded state.

RecBCD pathway

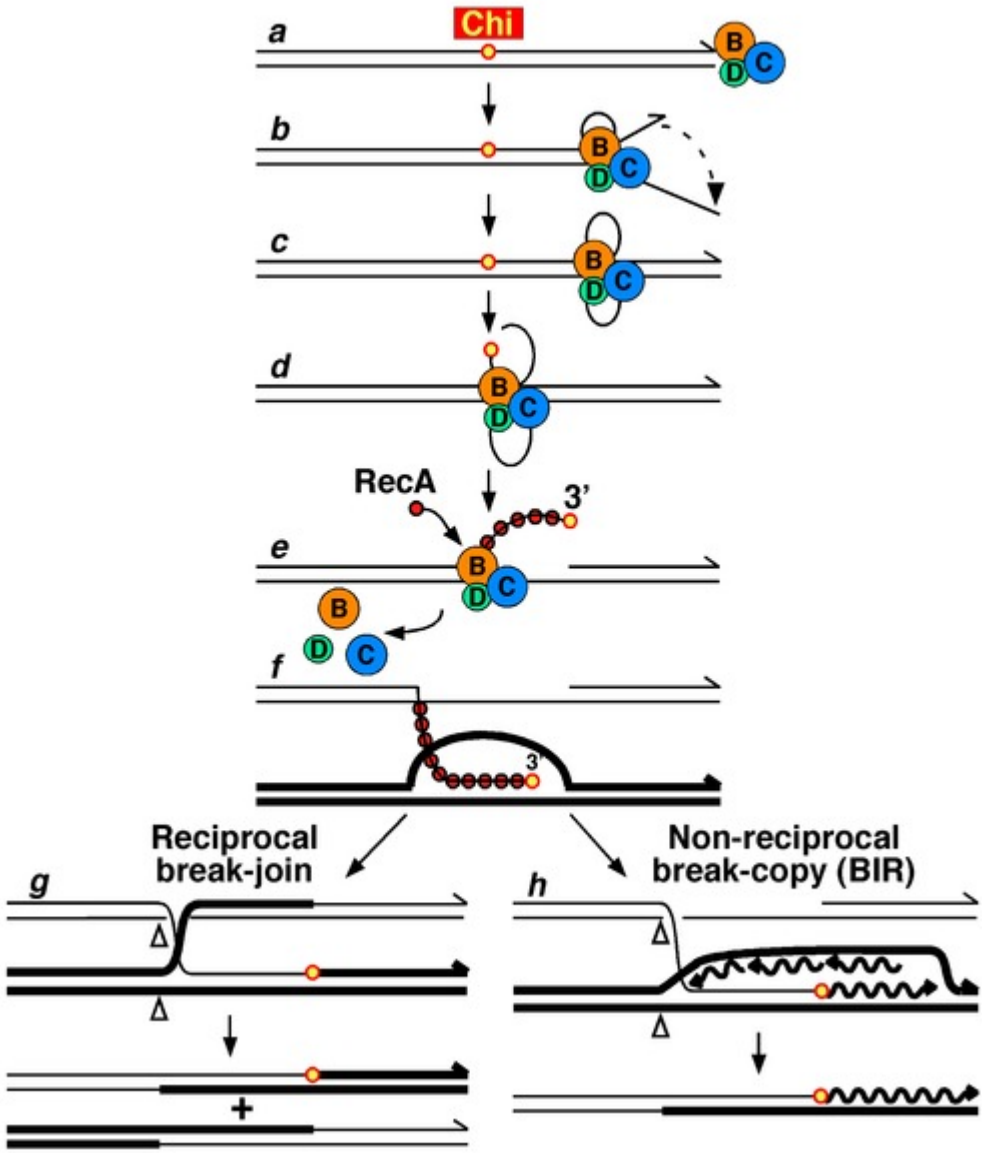


Figure 7A. Molecular model for the RecBCD pathway of recombination. This model is based on reactions of DNA and RecBCD with ATP in excess over Mg^{2+} ions. Step a: RecBCD binds to a double-stranded DNA end. Step b: RecBCD unwinds DNA. RecD is a fast helicase on the 5'-ended strand, and RecB is a slower helicase on the 3'-ended strand (that with an arrowhead). This produces two single-stranded (ss) DNA tails and one ss loop. The loop and tails enlarge as RecBCD moves along the DNA. Step c: The two tails anneal to produce a second ss DNA loop, and both loops move and grow. Step d: Upon reaching the Chi hotspot sequence RecBCD nicks the 3'-ended strand. Further unwinding produces a long 3'-ended ss tail with Chi near its end. Step e: RecBCD loads RecA protein onto the Chi tail. At some undetermined point, the RecBCD subunits disassemble. Step f: The RecA-ssDNA complex invades an intact homologous duplex DNA to produce a D-loop, which can be resolved into intact, recombinant DNA in two

ways. Step g: The D-loop is cut and anneals with the gap in the first DNA to produce a Holliday junction. Resolution of the Holliday junction (cutting, swapping of strands, and ligation) at the open arrowheads by some combination of RuvABC and RecG produces two recombinants of reciprocal type. Step h: The 3' end of the Chi tail primes DNA synthesis, from which a replication fork can be generated. Resolution of the fork at the open arrowheads produces one recombinant (non-reciprocal) DNA, one parental-type DNA, and one DNA fragment.

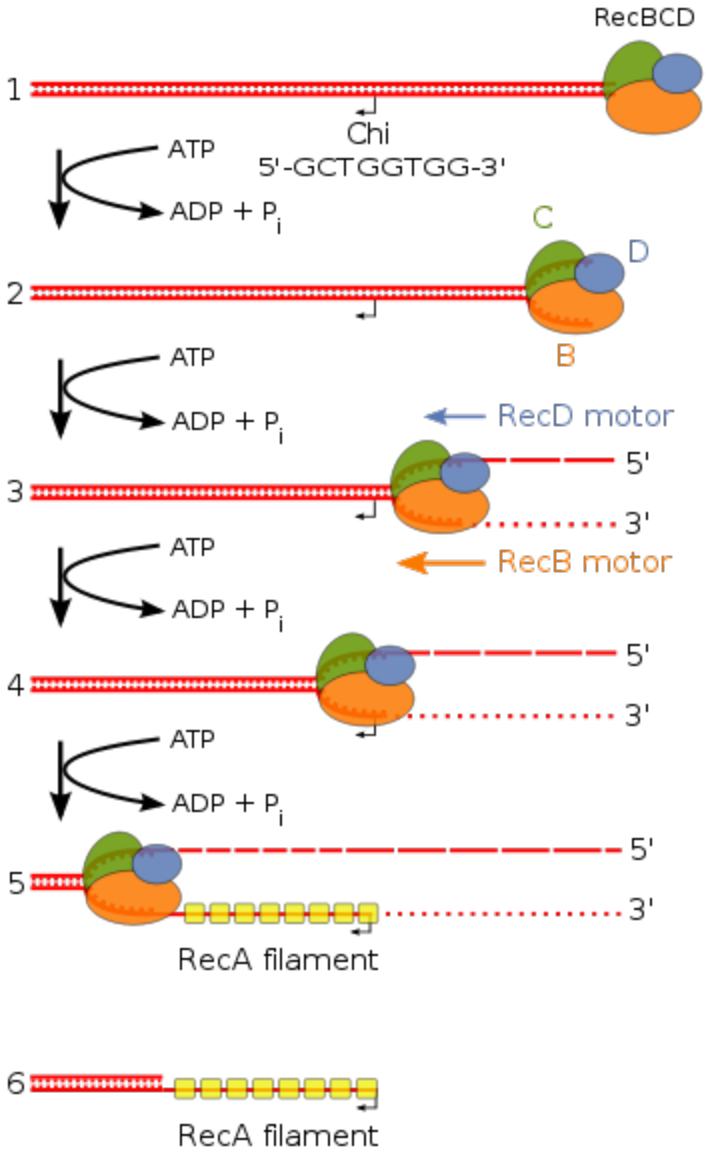


Figure 7B. Beginning of the RecBCD pathway. This model is based on reactions of DNA and RecBCD with Mg^{2+} ions in excess over ATP. Step 1: RecBCD binds to a DNA double strand break. Step 2: RecBCD initiates unwinding of the DNA duplex through ATP-dependent helicase activity. Step 3: RecBCD continues its unwinding and moves down the DNA duplex, cleaving the 3' strand much more frequently than the 5' strand.

Step 4: RecBCD encounters a Chi sequence and stops digesting the 3' strand; cleavage of the 5' strand is significantly increased. Step 5: RecBCD loads RecA onto the 3' strand. Step 6: RecBCD unbinds from the DNA duplex, leaving a RecA nucleoprotein filament on the 3' tail.

The RecBCD pathway is the main recombination pathway used in bacteria to repair double-strand breaks in DNA. These double-strand breaks can be caused by UV light and other radiation, as well as chemical mutagens. Double-strand breaks may also arise by DNA replication through a single-strand nick or gap. Such a situation causes what is known as a collapsed replication fork and is fixed by several pathways of homologous recombination including the RecBCD pathway.

In this pathway, a three-subunit enzyme complex called RecBCD initiates recombination by binding to a blunt or nearly blunt end of a break in double-strand DNA. After RecBCD binds the DNA end, the RecB and RecD subunits begin unzipping the DNA duplex through helicase activity. The RecB subunit also has a nuclease domain, which cuts the single strand of DNA that emerges from the unzipping process. This unzipping continues until RecBCD encounters a specific nucleotide sequence (5'-GCTGGTGG-3') known as a Chi site.

Upon encountering a Chi site, the activity of the RecBCD enzyme changes drastically. DNA unwinding pauses for a few seconds and then resumes at roughly half the initial speed. This is likely because the slower RecB helicase unwinds the DNA after Chi, rather than the faster RecD helicase, which unwinds the DNA before Chi. Recognition of the Chi site also changes the RecBCD enzyme so that it cuts the DNA strand with Chi and begins loading multiple RecA proteins onto the single-stranded DNA with the newly generated 3' end. The resulting RecA-coated nucleoprotein filament then searches out similar sequences of DNA on a homologous chromosome. Upon finding such a sequence, the single-stranded nucleoprotein filament moves into the homologous recipient DNA duplex in a process called *strand invasion*. The invading 3' overhang causes one of the strands of the recipient DNA duplex to be displaced, to form a D-loop. If the D-loop is cut, another swapping of strands forms a cross-shaped structure called a Holliday junction. Resolution of the Holliday junction by some combination of RuvABC or RecG can produce two recombinant DNA molecules with reciprocal genetic types, if the two interacting DNA molecules differ genetically. Alternatively, the invading 3’ end near Chi can prime DNA synthesis and form a replication fork. This type of resolution produces only one type of recombinant (non-reciprocal).

RecF pathway

Bacteria appear to use the RecF pathway of homologous recombination to repair single-strand gaps in DNA. When the RecBCD pathway is inactivated by mutations and additional mutations inactivate the SbcCD and ExoI nucleases, the RecF pathway can also repair DNA double-strand breaks. In the RecF pathway the RecQ helicase unwinds the DNA and the RecJ nuclease degrades the strand with a 5’ end, leaving the strand with the 3’ end intact. RecA protein binds to this strand and is either aided by the RecF, RecO,

and RecR proteins or stabilized by them. The RecA nucleoprotein filament then searches for a homologous DNA and exchanges places with the identical or nearly identical strand in the homologous DNA.

Although the proteins and specific mechanisms involved in their initial phases differ, the two pathways are similar in that they both require single-stranded DNA with a 3’ end and the RecA protein for strand invasion. The pathways are also similar in their phases of *branch migration*, in which the Holliday junction slides in one direction, and *resolution*, in which the Holliday junctions are cleaved apart by enzymes. The alternative, non-reciprocal type of resolution may also occur by either pathway.

Branch migration

Immediately after strand invasion, the Holliday junction moves along the linked DNA during the branch migration process. It is in this movement of the Holliday junction that base pairs between the two homologous DNA duplexes are exchanged. To catalyze branch migration, the RuvA protein first recognizes and binds to the Holliday junction and recruits the RuvB protein to form the RuvAB complex. Two sets of the RuvB protein, which each form a ring-shaped ATPase, are loaded onto opposite sides of the Holliday junction, where they act as twin pumps that provide the force for branch migration. Between those two rings of RuvB, two sets of the RuvA protein assemble in the center of the Holliday junction such that the DNA at the junction is sandwiched between each set of RuvA. The strands of both DNA duplexes—the "donor" and the "recipient" duplexes—are unwound on the surface of RuvA as they are guided by the protein from one duplex to the other.

Resolution

In the resolution phase of recombination, any Holliday junctions formed by the strand invasion process are cut, thereby restoring two separate DNA molecules. This cleavage is done by RuvAB complex interacting with RuvC, which together form the RuvABC complex. RuvC is an endonuclease that cuts the degenerate sequence 5'-(A/T)TT(G/C)-3'. The sequence is found frequently in DNA, about once every 64 nucleotides. Before cutting, RuvC likely gains access to the Holliday junction by displacing one of the two RuvA tetramers covering the DNA there. Recombination results in either "splice" or "patch" products, depending on how RuvC cleaves the Holliday junction. Splice products are crossover products, in which there is a rearrangement of genetic material around the site of recombination. Patch products, on the other hand, are non-crossover products in which there is no such rearrangement and there is only a "patch" of hybrid DNA in the recombination product.

Facilitating genetic transfer

Homologous recombination is an important method of integrating donor DNA into a recipient organism's genome in horizontal gene transfer, the process by which an organism incorporates foreign DNA from another organism without being the offspring

of that organism. Homologous recombination requires incoming DNA to be highly similar to the recipient genome, and so horizontal gene transfer is usually limited to similar bacteria. Studies in several species of bacteria have established that there is a log-linear decrease in recombination frequency with increasing difference in sequence between host and recipient DNA.

In bacterial conjugation, where DNA is transferred between bacteria through direct cell-to-cell contact, homologous recombination helps integrate foreign DNA into the host genome via the RecBCD pathway. The RecBCD enzyme promotes recombination after DNA is converted from single-strand DNA–in which form it originally enters the bacterium–to double-strand DNA during replication. The RecBCD pathway is also essential for the final phase of transduction, a type of horizontal gene transfer in which DNA is transferred from one bacterium to another by a virus. Foreign, bacterial DNA is sometimes misincorporated in the capsid head of bacteriophage virus particles as DNA is packaged into new bacteriophages during viral replication. When these new bacteriophages infect other bacteria, DNA from the previous host bacterium is injected into the new bacterial host as double-strand DNA. The RecBCD enzyme then incorporates this double-strand DNA into the genome of the new bacterial host.

In viruses

Homologous recombination occurs in several groups of viruses. In DNA viruses such as herpesvirus, recombination occurs through a break-and-rejoin mechanism like in bacteria and eukaryotes. There is also evidence for recombination in some RNA viruses, specifically positive-sense ssRNA viruses like retroviruses, picornaviruses, and coronaviruses. There is controversy over whether homologous recombination occurs in negative-sense ssRNA viruses like influenza.

In RNA viruses, homologous recombination can be either precise or imprecise. In the precise type of RNA-RNA recombination, there is no difference between the two parental RNA sequences and the resulting crossover RNA region. Because of this, it is often difficult to determine the location of crossover events between two recombining RNA sequences. In imprecise RNA homologous recombination, the crossover region has some difference with the parental RNA sequences – caused by either addition, deletion, or other modification of nucleotides. The level of precision in crossover is controlled by the sequence context of the two recombining strands of RNA: sequences rich in adenine and uracil decrease crossover precision.

Homologous recombination is important in facilitating viral evolution. For example, if the genomes of two viruses with different disadvantageous mutations undergo recombination, then they may be able to regenerate a fully functional genome. Alternatively, if two similar viruses have infected the same host cell, homologous recombination can allow those two viruses to swap genes and thereby evolve more potent variations of themselves.

Effects of dysfunction

Without proper homologous recombination, chromosomes often incorrectly align for the first phase of cell division in meiosis. This causes chromosomes to fail to properly segregate in a process called nondisjunction. In turn, nondisjunction can cause sperm and ova to have too few or too many chromosomes. Down's syndrome, which is caused by an extra copy of chromosome 21, is one of many abnormalities that result from such a failure of homologous recombination in meiosis.

Deficiencies in homologous recombination have been strongly linked to cancer formation in humans. For example, each of the cancer-related diseases Bloom's syndrome, Werner's syndrome and Rothmund-Thomson syndrome are caused by malfunctioning copies of RecQ helicase genes involved in the regulation of homologous recombination: BLM, WRN and RECQ4, respectively. In the cells of Bloom's syndrome patients, who lack a working copy of the BLM protein, there is an elevated rate of homologous recombination. Experiments in mice deficient in BLM have suggested that the mutation gives rise to cancer through a loss of heterozygosity caused by increased homologous recombination. A loss in heterozygosity refers to the loss of one of two versions—or alleles—of a gene. If one of the lost alleles helps to suppress tumors, like the gene for the retinoblastoma protein for example, then the loss of heterozygosity can lead to cancer.

Decreased rates of homologous recombination cause inefficient DNA repair, which can also lead to cancer. This is the case with BRCA1 and BRCA2, two similar tumor suppressor genes whose malfunctioning has been linked with considerably increased risk for breast and ovarian cancer. Cells missing BRCA1 and BRCA2 have a decreased rate of homologous recombination and increased sensitivity to ionizing radiation, suggesting that decreased homologous recombination leads to increased susceptibility to cancer. Because the only known function of BRCA2 is to help initiate homologous recombination, researchers have speculated that more detailed knowledge of BRCA2's role in homologous recombination may be the key to understanding the causes of breast and ovarian cancer.

Evolutionary origins

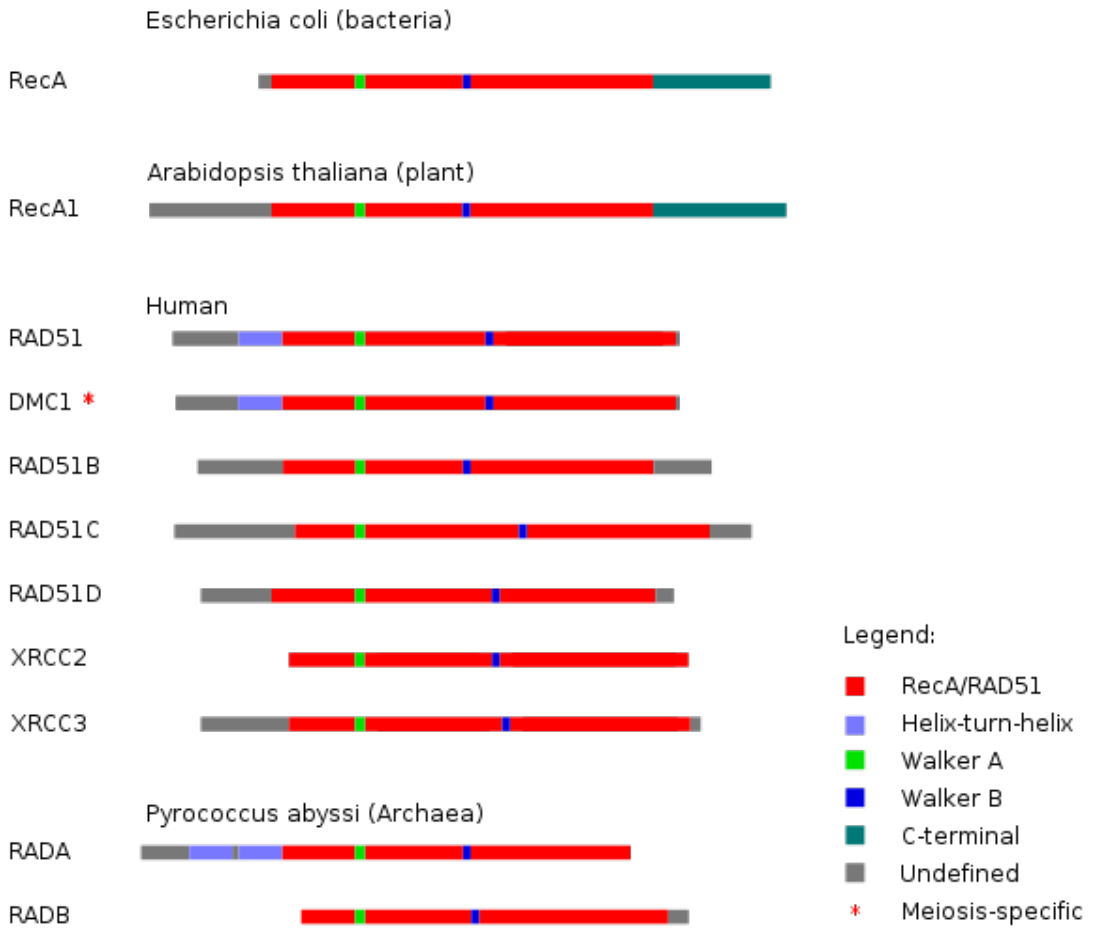


Figure 8. Protein domains in homologous recombination-related proteins are conserved across the three main groups of life: archaea, bacteria and eukaryotes.

Based on the similarity of their amino acid sequences, sets of proteins involved in homologous recombination are thought to share common evolutionary origins. One such set of proteins is the RecA/Rad51 protein family, which includes the RecA protein from bacteria, the Rad51 and Dmc1 proteins from eukaryotes and the RadA and RadB proteins from archaea. These proteins play key roles in the beginning stages of homologous recombination in the organisms that express them. The proteins in the RecA/Rad51 protein family share a long conserved region known as the RecA/Rad51 domain. Within this protein domain are two sequence motifs, Walker A and Walker B. The Walker A and B motifs allow members of the RecA/Rad51 protein family to engage in ATP hydrolysis, which provides energy for the proteins to drive reactions in homologous recombination.

Studies modeling the evolutionary relationships between the Rad51, Dmc1 and RadA proteins indicate that they are monophyletic, or that they share a common molecular

ancestor. Within this protein family, Rad51 and Dmc1 are grouped together in a separate clade from RadA. One of the reasons for grouping these three proteins together is that they all possess a modified helix-turn-helix motif, which helps the proteins bind to DNA, toward their N-terminal ends. An ancient gene duplication event of a eukaryotic RecA gene and subsequent mutation has been proposed as a likely origin of the modern RAD51 and DMC1 genes.

The discovery of Dmc1 in several species of *Giardia*, one of the earliest protists to diverge as a eukaryote, suggests that meiotic homologous recombination—and thus meiosis itself—emerged very early in eukaryotic evolution. In addition to research on Dmc1, studies on the Spo11 protein have provided information on the origins of meiotic recombination. Spo11 is a type II topoisomerase that initiates homologous recombination in meiosis by making targeted double-strand breaks in DNA. Phylogenetic trees based on the sequence of genes similar to SPO11 in animals, fungi, plants, protists and archaea have led scientists to believe that the version Spo11 currently in eukaryotes emerged in the last common ancestor of eukaryotes and archaea.

Technological applications

Gene targeting

Figure 9. As a developing embryo, this chimeric mouse had the agouti coat color gene introduced into its DNA via gene targeting. Its offspring are homozygous for the agouti gene.

Many methods for introducing DNA sequences into organisms to create recombinant DNA and genetically modified organisms use the process of homologous recombination. Also called gene targeting, the method is especially common in yeast and mouse genetics. The gene targeting method in knockout mice uses mouse embryonic stem cells to deliver artificial genetic material (mostly of therapeutic interest), which represses the target gene of the mouse by the principle of homologous recombination. The mouse thereby acts as a working model to understand the effects of a specific mammalian gene. In recognition of their discovery of how homologous recombination can be used to introduce genetic modifications in mice through embryonic stem cells, Mario Capecchi,

Martin Evans and Oliver Smithies were awarded the 2007 Nobel Prize for Physiology or Medicine.

Advances in gene targeting technologies which hijack the homologous recombination mechanics of cells are now leading to the development of a new wave of more accurate, isogenic human disease models. These engineered human cell models are thought to more accurately reflect the genetics of human diseases than their mouse model predecessors. This is largely because mutations of interest are introduced into endogenous genes, just as they occur in the real patients, and because they are based on human genomes rather than rat genomes. Furthermore, certain technologies enable the knock-in of a particular mutation rather than just knock-outs associated with older gene targeting technologies.

Protein engineering

Protein engineering with homologous recombination develops chimeric proteins by swapping fragments between two parental proteins. These techniques exploit the fact that recombination can introduce a high degree of sequence diversity while preserving a protein's ability to fold into its tertiary structure, or three-dimensional shape. This stands in contrast to other protein engineering techniques, like random point mutagenesis, in which the probability of maintaining protein function declines exponentially with increasing amino acid substitutions. The chimeras produced by recombination techniques are able to maintain their ability to fold because their swapped parental fragments are structurally and evolutionarily conserved. These recombinable "building blocks" preserve structurally important interactions like points of physical contact between different amino acids in the protein's structure. Computational methods like SCHEMA and statistical coupling analysis can be used to identify structural subunits suitable for recombination.

Techniques that rely on homologous recombination have been used to engineer new proteins. In a study published in 2007, researchers were able to create chimeras of two enzymes involved in the biosynthesis of isoprenoids, a diverse class of compounds including hormones, visual pigments and certain pheromones. The chimeric proteins acquired an ability to catalyze an essential reaction in isoprenoid biosynthesis—one of the most diverse pathways of biosynthesis found in nature—that was absent in the parent proteins. Protein engineering through recombination has also produced chimeric enzymes with new function in members of a group of proteins known as the cytochrome P450 family, which in humans is involved in detoxifying foreign compounds like drugs, food additives and preservatives.

Cancer therapy

In 2009, researchers reported trial results of a cancer therapy that exploits deficiencies in homologous recombination in certain types of cancer. The drug, named olaparib, a PARP1 inhibitor, was shown to shrink or stop the growth of tumors from breast, ovarian and prostate cancers caused by mutations in the BRCA1 or BRCA2 genes. BRCA1 and BRCA2 are necessary for DNA repair by homologous recombination. When BRCA1 or

BRCA2 are absent, another type of DNA repair mechanism called base-excision repair usually compensates for the lack of DNA repair by homologous recombination. However, when the PARP1 protein – which is necessary for base-excision repair – is inhibited, DNA repair is drastically reduced, and the cell dies.

By stopping DNA repair in such a fashion, olaparib applies the concept of synthetic lethality to specifically target cancer cells. An article in the *New England Journal of Medicine* noted that exploiting synthetic lethality was a new direction in anti-cancer drug development. While PARP1 inhibitors represent a novel approach to cancer therapy, researchers have cautioned that they may prove insufficient for treating late-stage, metastatic cancers. Cancer cells can become resistant to a PARP1 inhibitor if they experience deletions of mutations in BRCA2. This undermines the drug's synthetic lethality by restoring cancer cells' ability to repair DNA by homologous recombination.