DNA DEMAGE AND REPAIR

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 10,000 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 and contribute to tumour heterogeneity.

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 200-400 nm] 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. 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 several 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-methylguanosine, 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, aristolactam I-dA 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.

6. Monoadduct damage cause by change in single nitrogenous base of DNA
7. Diadduct damage

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. Intermediate-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 aromatic 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 chromatinduring 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 mitochondriaorganelles, 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 process 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 (see Hayflick limit). In contrast, quiescence is a reversible state of cellular dormancy that is unrelated to

genome damage (see cell cycle). 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 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 composing 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

DNA repair is 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 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 (interstrand crosslinks or ICLs). This can eventually lead to malignant tumors, or cancer as per the two hit hypothesis.

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 the normal functionality of that organism. Many genes that were initially shown to influence life span have turned out to be involved in DNA damage repair and protection.

DNA REPAIR MECHANISMS

Cells cannot function if DNA damage corrupts the integrity and accessibility of essential information in the genome (but cells remain superficially functional when 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.

1. **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 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.

2. **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. Photolyase, an old enzyme present in bacteria, fungi, and most animals no longer functions in humans, who instead use nucleotide excision repair to repair damage from UV irradiation. 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.

3. **Single-strand damage**

   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)** repairs damage to a single nitrogenous base by deploying enzymes called glycosylases. These enzymes remove a single nitrogenous base to create an apurinic or apyrimidinic site (AP site). Enzymes called AP endonucleasesnick the damaged DNA backbone at the

AP site. DNA polymerase then removes the damaged region using its 5’ to 3’ exonuclease activity and correctly synthesizes the new strand using the complementary strand as a template.

2. **Nucleotide excision repair (NER)** repairs damaged DNA which commonly consists of bulky, helix-distorting damage, such as pyrimidine dimerization caused by UV light. Damaged regions are removed in 12-24 nucleotide-long strands in a three-step process which consists of recognition of damage, excision of damaged DNA both upstream and downstream of damage by endonucleases, and resynthesis of removed DNA region. NER is a highly evolutionarily conserved repair mechanism and is used in nearly all eukaryotic and prokaryotic cells. In prokaryotes, NER is mediated by Uvr proteins. In eukaryotes, many more proteins are involved, although the general strategy is the same.
3. **Mismatch repair systems** are present in essentially all cells to correct errors that are not corrected by proofreading. These systems consist of at least two proteins. One detects the mismatch, and the other recruits an endonuclease that cleaves the newly synthesized DNA strand close to the region of damage. In *E. coli* , the proteins involved are the Mut class proteins. This is followed by removal of damaged region by an exonuclease, resynthesis by DNA polymerase, and nick sealing by DNA ligase.

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.

Translesion synthesis

Translesion synthesis (TLS) 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 (i.e. DNA polymerase IV or V, from the Y Polymerase family), 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) on undamaged templates 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 ιintroduces mutations at these sites. Pol η is known to add the first adenine across the T^T photodimer using Watson-Crick base pairing and the second adenine will be added in its syn conformation using Hoogsteen base pairing. 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. For example, Human DNA polymerase eta can bypass complex DNA lesions like guanine-thymine intra-strand crosslink, G[8,5-Me]T, although can cause targeted and semi-targeted mutations. Paromita Raychaudhury and Ashis Basu[studied the toxicity and mutagenesis of the same lesion in *Escherichia coli* by replicating a G[8,5-Me]T-modified plasmid in *E. coli* with specific DNA polymerase knockouts. Viability was very low in a strain lacking pol II, pol IV, and pol V, the three SOS-inducible DNA polymerases, indicating that translesion synthesis is conducted primarily by these specialized DNA polymerases. 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. After translesion synthesis, extension is required. This extension can be carried out by a replicative polymerase if the TLS is error-free, as in the case of Pol η, yet if TLS results in a mismatch, a specialized polymerase is needed to extend it; Pol ζ. Pol ζ is unique in that it can extend terminal mismatches, whereas more processive polymerases cannot. So when a lesion is encountered, the replication fork will stall, PCNA will switch from a processive polymerase to a TLS polymerase such as Pol ι to fix the lesion, then PCNA may switch to Pol ζ to extend the mismatch, and last PCNA will switch to the processive polymerase to continue 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.

DNA damage checkpoint is a signal transduction pathway that blocks cell cycle progression in G1, G2 and metaphase and slows down the rate of S phase progression when DNA is damaged. It leads to a pause in cell cycle allowing the cell time to repair the damage before continuing to divide.

Checkpoint Proteins can be separated into four groups: phosphatidylinositol 3-kinase (PI3K)-like protein kinase, proliferating cell nuclear

antigen (PCNA)-like group, two serine/threonine(S/T) kinases and their adaptors. Central to all DNA damage induced checkpoints responses is a pair of large protein kinases belonging to the first group of PI3K-like protein kinases-the ATM (Ataxia telangiectasia mutated) and ATR (Ataxia- and Rad-related) kinases, whose sequence and functions have been well conserved in evolution. All DNA damage response requires either ATM or ATR because they have the ability to bind to the chromosomes at the site of DNA damage, together with accessory proteins that are platforms on which DNA damage response components and DNA repair complexes can be assembled.

An important downstream target of ATM and ATR is p53, as it is required for inducing apoptosis following DNA damage. The cyclin-dependent kinase inhibitor p21 is induced by both p53-dependent and p53-independent mechanisms and can arrest the cell cycle at the G1/S and G2/M checkpoints by deactivating cyclin/cyclin-dependent kinase complexes.

The prokaryotic SOS response

The SOS response is the 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.