Mechanisms of DNA damage, repair, and mutagenesis

<table>
<thead>
<tr>
<th>Citation</th>
<th>Chatterjee, Nimrat, and Graham C. Walker. “Mechanisms of DNA Damage, Repair, and Mutagenesis.” Environmental and Molecular Mutagenesis 58, 5 (May 2017): 235–263 © 2017 Wiley Periodicals, Inc</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1002/EM.22087">http://dx.doi.org/10.1002/EM.22087</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Wiley Blackwell</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Jan 03 15:36:55 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/116957">http://hdl.handle.net/1721.1/116957</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Mechanisms of DNA damage, repair and mutagenesis

Nimrat Chatterjee* and Graham C. Walker
Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02138

Abstract

Living organisms are continuously exposed to a myriad of DNA damaging agents that can impact health and modulate disease-states. However, robust DNA repair and damage-bypass mechanisms faithfully protect the DNA by either removing or tolerating the damage to ensure an overall survival. Deviations in this fine-tuning are known to destabilize cellular metabolic homeostasis, as exemplified in diverse cancers where disruption or deregulation of DNA repair pathways results in genome instability. Because routinely used biological, physical and chemical agents impact human health, testing their genotoxicity and regulating their use have become important. In this introductory review, we will delineate mechanisms of DNA damage and the counteracting repair/tolerance pathways to provide insights into the molecular basis of genotoxicity in cells that lays the foundation for subsequent articles in this issue.

Keywords

Base excision repair; mismatch repair; nucleotide excision repair; single and double strand break repair; translesion synthesis; telomeres

Introduction

Preserving genomic sequence information in living organisms is important for the perpetuation of life. At the same time, mutagenesis plays an indispensible part in its maintenance and evolution, while also contributing to cancer, certain human diseases and aging. It is known that DNA, the basic unit of inheritance, is an intrinsically reactive molecule and is highly susceptible to chemical modifications by endogenous and exogenous agents. Furthermore, the DNA polymerases engaged in DNA replication and repair make mistakes, thereby burdening cells with potentially disadvantageous mutations. However, cells are equipped with intricate and sophisticated systems—DNA repair, damage tolerance, cell cycle checkpoints and cell death pathways—that collectively function to reduce the deleterious consequences of DNA damage.

Cells respond to DNA damage by instigating robust DNA damage response (DDR) pathways, which allow sufficient time for specified DNA repair pathways to physically remove the damage in a substrate-dependent manner. At least five major DNA repair pathways—base excision repair (BER), nucleotide excision repair (NER), mismatch repair

*Correspondence author: nimratc@mit.edu, Phone: 832-702-0106.
There are no conflicts of interests.
(MMR), homologous recombination (HR) and non-homologous end joining (NHEJ)—are active throughout different stages of the cell cycle, allowing the cells to repair the DNA damage. A few specific lesions can also be removed by direct chemical reversal and interstrand crosslink (ICL) repair. These repair processes are key to maintaining genetic stability in cells. In addition, certain types of DNA damage are substrates for the DNA damage tolerance pathways. In higher eukaryotes, for example, a well-orchestrated group of five main translesion synthesis (TLS) polymerases—REV1, POL ζ, POL η, POL κ and POL ι—bypass the damage to enable the continuation of replication, but with the possibility of a concurrent introduction of an incorrect base that can be fixed into a mutation in the subsequent round of replication. Under the circumstances, when the damaged DNA persists, programmed cell death or apoptosis, a regulatory response to DNA damage, is activated to get rid of cells with extensive genome instability.

Not surprisingly, in many cancers, DNA repair, DNA damage tolerance and DDR pathways are disrupted or deregulated, which increases mutagenesis and genomic instability, thereby promoting cancer progression [Bouwman and Jonkers, 2012; Ghosal and Chen, 2013; Wolters and Schumacher, 2013]. likewise, aging is attributed to attrition of chromosomal ends and failing capacities of a combination of these pathways. Other diseases, such as neurodegenerative disorders, result from a combinatorial failure of more than one of these processes. The 2015 Nobel Prize in Chemistry to Drs. Lindahl, Modrich and Sancar highlights the importance of mechanisms of DNA damage and repair and their implications for human health. In this review we will discuss the details of various types and mechanisms of DNA damage and the compensatory repair and tolerance pathways.

Types of DNA damage

DNA damage can be categorized into two main classes based on its origin: endogenous and exogenous. The majority of the endogenous DNA damage arises from the chemically active DNA engaging in hydrolytic and oxidative reactions with water and reactive oxygen species (ROS), respectively, that are naturally present within cells. Such inherently predisposed reactions of DNA with molecules from its immediate surroundings fuel the development of hereditary diseases and sporadic cancers [Visconti and Grieco, 2009; Reuter et al., 2010; Perrone et al., 2016]. Exogenous DNA damage, on the other hand, occurs when environmental, physical and chemical agents damage the DNA. Examples include UV and ionizing radiation, alkylating agents, and crosslinking agents. We offer here a brief summary of the main endogenous and environmental agents that produce the different classes of DNA damage that then become substrates for the specific DNA repair pathways discussed in the subsequent section.

Endogenous DNA damage

Replication errors, DNA base mismatches and topoisomerase-DNA complexes

Every time a human cell replicates, approximately $3 \times 10^9$ bases are copied over by high fidelity replicative polymerases (δ and ε). However, a battery of other DNA polymerases (α, β, σ, γ, λ, REV1, ζ, η, ι, κ, θ, ν, μ, Tdt and PrimPol) can carry out lower fidelity DNA synthesis during DNA replication or repair (Table 1) [Loeb and Monnat (2008)]. High
fidelity DNA synthesis is a consequence of structural and biochemical attributes of replicative DNA polymerases, which ensure the insertion of a correct complementary deoxynucleotide opposite the template base. This is accomplished, for instance, by: 1) the thermodynamic stability and base-pair energetics of the incoming dNTP and template base, 2) the geometric selection of a correctly shaped and sized dNTP in the polymerase’s active site, and 3) removing an incorrectly inserted deoxynucleotide by a 3′-5′ deoxynucleotide exonuclease. In addition, the mismatch repair (MMR) pathway contributes to replication fidelity by more than 100-fold by correcting the rare errors that have escaped proofreading by replicative polymerases [Kunkel, 2004; Kunkel, 2009; Kunkel, 2011].

Nevertheless, base substitutions and single base insertion and deletion errors still accumulate at a frequency of 10^{-6} to 10^{-8} per cell per generation [Kunkel, 2004; Kunkel, 2009]. Additional replication errors accumulate from strand slippage events at repetitive sequences causing insertions and deletions of nucleotides that can potentially change the reading frame [Viguera et al., 2001; Chatterjee N., 2013]. Other times, the replicative polymerases incorrectly incorporate uracil in the DNA or end up with a compromised fidelity because of the alterations of the relative and absolute concentrations of dNTPs and rNTPs within the cell’s environment [Andersen et al., 2005; Vertessy and Toth, 2009; Kumar et al., 2011; Clausen et al., 2013; Buckland et al., 2014; Potenski and Klein, 2014]. These incorrectly paired/incorporated nucleotides that escape proofreading and MMR become mutations in the next round of replication and are a major source of spontaneous mutagenesis.

Another source of endogenous DNA damage results from the action of topoisomerase enzymes (for example: TOP I, TOP II, TOP III; 7 TOP genes are found in the human genome), which primarily remove superhelical tension on DNA during replication and transcription [Wang, 2002; Pommier et al., 2006]. TOP1, for example, transiently nicks the supercoiled DNA and facilitates rotation of the broken strand around the TOP1-bound DNA strand to relax the DNA. Thereafter, TOP1 religates the breaks by aligning the 5′-OH group of the DNA with the tyrosine-DNA phosphodiester bond to resolve the complex [Stewart et al., 1998; Carey et al., 2003]. Misalignment of the 5′-OH DNA end stabilizes the cleavage complex to form a DNA lesion [Pommier and Cherfils, 2005; Pommier and Marchand, 2005]. Interestingly, anticancer drugs such as camptothecin and many naturopathic compounds are known to stabilize the TOP1-DNA cleavage complexes [Staker et al., 2002; Han et al., 2008]. Additionally, DNA adducts (from UV and benzene derivatives) and aberrant DNA structures (nicks, mismatches, abasic sites) can also irreversibly trap the TOP1-DNA cleavage complex into DNA lesions called suicidal complexes [Burgin et al., 1995; Pourquier and Pommier, 2001; Meng et al., 2003]. TOP1-associated DNA damage is usually repaired by reversal of these complexes or is excised by TDP1 (tyrosyl DNA phosphodiesterase) and endonucleases [Pommier et al., 2006].

**Spontaneous base deamination**

Base deamination is a major source of spontaneous mutagenesis in human cells, where cytosine (C), adenine (A), guanine (G), and 5-methyl cytosine (5mC) in DNA lose their exocyclic amine to become uracil (U), hypoxanthine, xanthine and thymine (T), respectively (Figure 1B). Interestingly, these base deamination events occur at a much higher frequency
in single-stranded versus double-stranded DNA and are often exacerbated by transient single strandedness during active replication, transcription and recombination [Lindahl, 1993; Yonekura et al., 2009]. In the case of deamination of cytosine, for instance, the native C:G base pairing alters to a U:A base pair in the first round of replication, which in the next round of replication results in a CG→TA mutation. Cytosine and 5-methyl cytosine are the most frequently deaminated, but 5-methyl cytosine is deaminated three to four times more frequently than cytosine [Lindahl, 1979]. While deaminated cytosine is rapidly removed from DNA by uracil-DNA glycosylase, the G:T base pair resulting from deamination of 5-methylcytosine is instead a substrate for the thymine DNA glycosylase (TDG) and the relatively slow MMR process [Lindahl, 1979; Wiebauer and Jiricny, 1990; Waters and Swann, 1998]. Consequently, the GC→AT transition at the CpG sequences accounts for one-third of the single site mutations responsible for hereditary diseases in humans [Cooper and Youssoufian, 1988; De Bont and van Larebeke, 2004].

Paradoxically, cytosine deamination is also a normal route for somatic hypermutagenesis during antibody development due to the action of the deaminase enzymes AID (activation-induced deaminase) and APOBEC1 (Apolipoprotein B mRNA editing enzyme catalytic polypeptide 1), which mediate host defense against retroviruses [Goff, 2003; Blanc and Davidson, 2010; Chandra et al., 2015]. In addition to the endogenous deamination sources, environmental exposure to UV radiation, intercalating agents, nitrous acid and sodium bisulfite can in general enhance base deamination rates in the DNA [Chen and Shaw, 1993; Moyer et al., 1993; Pfeifer et al., 2005; d’Ischia et al., 2011]. From an evolutionary standpoint, cytosine deamination from endogenous and exogenous sources may serves as a source for genetic diversity [Fryxell and Zuckerandl, 2000; Nabel et al., 2012].

Abasic sites

Abasic or AP (apurinic/apyrimidic) sites are continuously created in the DNA when the N-glycosyl bond, which links the nitrogenous base and the sugar phosphate backbone, either hydrolyzes spontaneously or gets cleaved by a DNA glycosylase to generate an intermediate in the BER pathway. For example, AP sites are formed when uracil is removed from the DNA by uracil-DNA glycosylase [Lindahl and Barnes, 2000]. In a human cell, about 10,000 abasic sites are created per day; both extreme pH conditions and high temperatures positively impact their generation [Lindahl, 1993; Tropp, 2011]. Abasic sites are inherently unstable and readily convert into single strand breaks (SSBs) from a β-elimination reaction that targets the 3′ phosphodiester bond of the leftover deoxyribose [Bailly and Verly, 1988; Waters and Walker, 2006; Tropp, 2011; Chan et al., 2013]. Most AP sites are effectively removed by AP endonucleases that cleave at their 5′ end and allow the BER pathway to repair them. Alternatively, AP sites can be bypassed by TLS polymerases [Chan et al., 2013]. It is not known whether other exogenous stresses could also directly propel the formation of AP sites in the genome.

Oxidative DNA damage

Reactive oxygen species (ROS) are the typical byproducts of the electron transport chain (ETC) during cellular respiration in aerobic organisms, and are additionally derived from catabolic oxidases, anabolic processes and peroxisomal metabolism [Henle and Linn, 1997].
At low levels, ROS species perform important cellular functions such as serving as cellular messengers in redox signaling reactions and effecting important defense responses to invading pathogens by the immune system [Errol C. Friedberg, 2005; Segal, 2005; Malle et al., 2007]. However, in excess, ROS species can cause a total of approximately 100 different oxidative base lesions and 2-deoxyribose modifications [Bjelland and Seeberg, 2003; Cadet et al., 2010; Cadet et al., 2011; Cadet et al., 2012; Ravanat et al., 2012; Cadet and Wagner, 2014]. Ordinarily, the deleterious consequences of ROS are regulated in cells by 1) restricting respiration in the mitochondrial compartment, thereby protecting other cellular components, 2) protecting DNA by complexing it with histones, and 3) quenching of surplus ROS species by the anti-oxidant enzymes superoxide dismutase, catalase, and peroxiredoxin [Riley, 1994; Mates et al., 1999; Mates and Sanchez-Jimenez, 1999]. Despite this, an overabundance of ROS species is notably associated with the development of human diseases, such as cancer, Alzheimer’s disease, Parkinson’s disease, diabetes, and heart failure [Giacco and Brownlee, 2010; Liou and Storz, 2010; Mohsenzadegan and Mirshafiey, 2012; Dias et al., 2013; Hafstad et al., 2013].

The most conspicuous of the ROS species are the superoxide radicals (•O$_2^-$), hydrogen peroxide (H$_2$O$_2$), and the hydroxyl radical (•OH) [Tropp, 2011]. Amongst these ROS species, the •OH radical, produced as a byproduct of a Fenton’s reaction of H$_2$O$_2$ with Fe$^{2+}$, is by far the most reactive, and is capable of damaging DNA, proteins and lipids [Imlay et al., 1988; Dizdaroglu et al., 1991]. These electrophilic •OH radicals react with DNA bases by 1) adding to their double bonds, 2) abstracting hydrogen atoms from their methyl groups, and 3) attacking the sugar residue in their immediate vicinity [Breen and Murphy, 1995; Winterbourn, 2008]. For example, thymine glycol residues are generated from a •OH attack on the C5/C6 double bonds of thymine (Figure 1C). Similarly, the •OH radical produced as a byproduct of the Fenton reaction of H$_2$O$_2$ and Fe$^{2+}$ induces an imidazole ring opening in guanine and adenine to form the fragmented purine structure formamidopyrimidine (Figure 1C) [Chetsanga et al., 1981; Errol C. Friedberg, 2005; C., 2006]. Another biologically significant and major oxidative base lesion formed from hydroxylation of the C-8 residue of guanine is the saturated imidazole ring 7,8 dihydro-8-oxoguanine (8-oxo-G) (Figure 1C). 8-oxo-guanine pairs incorrectly with adenine instead of cytosine, thereby adding to the overall mutational load, and is further oxidized to other deleterious secondary DNA lesions because of its low oxidation potential [Kasai and Nishimura, 1984; Cheng et al., 1992; Cadet et al., 1999; Cadet et al., 2010].

Other than attacking DNA bases, ROS radicals can also compromise the DNA backbone causing an estimated 2300 single strand breaks per cell per hour in mammalian cells [Gilioni et al., 1981; R, 1981; Henner et al., 1983a; Henner et al., 1983b]. While the BER pathway repairs the oxidized bases, the breaks in the DNA backbone are repaired by the single strand break repair (SSBR) pathways or the double strand break repair (DSBR) pathways [Henner et al., 1983a; Demple and Harrison, 1994]. Finally, lipid peroxidation, the oxidation of lipid molecules by hydroxyl radicals, generates aldehyde products such as malondialdehyde and 4-hydroxynonenal, which can react with adenine, guanine and cytosine to form mutagenic adducts [Marnett, 2000; Plastaras et al., 2000; VanderVeen et al., 2001]. About 1 adduct per $10^6$ – $10^7$ parent DNA bases results from lipid peroxidation events and the number of mutagenic adducts are expected to be even higher for metal storage diseases such as...
Wilson’s disease and hemochromatosis [Carmichael et al., 1995; Luczaj and Skrzydlewska, 2003; Broedbaek et al., 2009].

DNA methylation

S-adenosylmethionine (SAM), which is used as a methyl donor by methyl transferases during normal methylation reactions, can also spontaneously generate up to 4000 N7-methylguanine, 600 N3-methyladenine and 10–30 O6-methylguanine residues per cell per day in mammals (Figure 1d) [Rydberg and Lindahl, 1982; Holliday and Ho, 1998; De Bont and van Larebeke, 2004]. Other methylating agents include endogenous nitrosated bile salts, betaine, choline, and environmental agents such as tobacco smoke, diet, pollution or derivatives of N-nitroso compounds [O’Driscoll et al., 1999; Zhao et al., 1999]. O6-methylguanine and the related residues O6-methylthymine and O6-ethylthymine are highly mutagenic, producing G:C→A:T and T:A→C:G transition mutations, respectively. In contrast, N3-methyladenine is only partly cytotoxic due to inhibition of DNA synthesis, while the N7-methylguanine residue is essentially harmless unless it undergoes a spontaneous cleavage to generate an AP site or opens the imidazole ring to form formamidopyrimidine [Loveless, 1969; Loechler et al., 1984; Larson et al., 1985; Preston et al., 1986; Preston et al., 1987; O’Connor et al., 1988; Singer et al., 1989; Tudek et al., 1992]. Other minor methyl lesions produced by SAM are the mutagenic N3-methylthymine and N3-methylcytosine (Figure 1D) [Boiteux and Laval, 1982; Saffhill, 1985; Huff and Topal, 1987].

Methylated bases are removed from DNA by two main pathways: 1) direct reversal of the DNA damage by O6-methylguanine DNA methyltransferase or by oxidation by an α-ketoglutarate-dependent dioxygenase AlkB homolog, and 2) BER, which is initiated by DNA glycosylases to remove the methylated bases by catalyzing the cleavage of their glycosidic bonds [Sakumi and Sekiguchi, 1990; Tudek et al., 1992; Huang et al., 1994; Zak et al., 1994; Ye et al., 1998]. In addition, the O6-methylguanine DNA lesion interestingly triggers a cytotoxic and futile cycle of MMR, by means of its abnormal base pairing with other residues [Branch et al., 1993; Kat et al., 1993]. If left unrepaired, methylated DNA bases are a major source of spontaneous DNA damage. Alkylated DNA damage from exogenous compounds will be discussed later in the article.

Exogenous DNA damage

Ionizing Radiation (IR)

Ionizing radiation, composed of alpha, beta, gamma, neutrons and X-rays, is abundant in our environment, being produced from diverse sources ranging from rocks, soil, and radon, to cosmic radiation and medical devices. Each type of radiation can be classified to describe its effect (direct or indirect) and ionization density (linear energy transfer (LET)). Depending on the amount of energy transferred to matter, radiations are classified as either high LET (alpha rays) or low LET (beta and gamma). Cumulatively, IR can damage the DNA either directly, or by indirect means, such as by radiolysis of the surrounding water to generate a cluster of highly reactive hydroxyl radicals (•OH) [Errol C. Friedberg, 2005; Omar Desoukya, 2015]. The presence of oxygen and other reactive species in the surrounding also
potentiates the formation of other DNA-reactive free radicals by IR [Wardman, 2009]. In fact, indirect DNA damage from •OH radicals accounts to about 65% of the radiation-induced DNA damage [Vignard et al., 2013]. Because of this, IR produces a spectrum of base lesions that is similar to that generated by ROS species (see previous section). Major lesions include 8-oxo-guanaine, thymine glycol and formamidopyrimidines (Figure 1C).

Apart from causing base lesions, ionizing radiation also causes single strand breaks with a unique signature, where the DNA breaks have 3′ phosphate or 3′-phosphoglycolate ends rather than 3′-OH ends. In addition, fragmented sugar derivatives and loss of terminal base residues culminate into clustered damage or single stranded gaps [Henner et al., 1982; Henner et al., 1983b; Obe et al., 1992]. AP endonucleases, polynucleotide kinase/phosphatase (PNKP) and tyrosyl DNA phosphodiesterase 1 (TDP1) can efficiently process the modified ends and enable repair of the IR-induced single strand breaks [Price, 1993; Jilani et al., 1999; Zhou et al., 2005; El-Khamisy et al., 2007]. A particularly important radiation-induced lesion is the double strand break, formed from multiple damaged sites closely positioned on both DNA strands [Hutchinson, 1985; Iliaakis, 1991]. Although toxic, IR-induced double strand breaks can be repaired by the HR pathway [Lomax et al., 2013].

Ultraviolet (UV) radiation

UV radiation emanating from the sun is the leading cause of skin cancers in humans [Davies, 1995; KIEFER, 2007]. Typically, UV radiation is categorized into three classes based on the range of wavelength: UV-C (190–290 nm), UV-B (290–320 nm) and UV-A (320–400nm). DNA absorbs maximal UV radiation at 260 nm, beyond which the photoabsorption drops dramatically. Sunlight is composed of 5.1% UV-A, 0.3% UV-B, 62.7% visible light and 31.9% infrared, as the hazardous UV-C is mostly filtered out by the ozone layer [Davies, 1995]. The effects of UV on matter are disseminated in two ways. First, if the UV is absorbable, molecules in matter are excited leading to their photochemical alteration. Second, if UV cannot be directly absorbed, energy transfer from nearby molecules called photosensitizers indirectly affects matter. UV damage DNA by both pathways.

Laboratory studies have shown that UV-C damages DNA primarily by causing covalent linkages between two adjacent pyrimidines. Here the two major photoproducts are the cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6 – 4) pyrimidone photoproducts ((6 – 4) PPs) (Figure 2). Their relative formation frequency depends on wavelength and dose of light [Varghese, 1972; Mitchell and Nairn, 1989; Davies, 1995], although, the yield of (6 – 4) PP is slightly less than CPDs [Mitchell and Nairn, 1989]. Other minor photoproducts are also generated, such as pyrimidine hydrate, thymine glycols, and dipurine adducts [Demple and Linn, 1982; Bose et al., 1983; Kumar et al., 1991; Mitchell et al., 1991]. In CPDs, a cyclobutane ring covalently links the two adjacent pyrimidines, whereas in (6 – 4) PP, the C6 position of one pyrimidine is covalently linked to C4 position of the adjacent pyrimidine. These bulky dimers distort the helix, requiring TLS polymerases for replication past them, thereby contributing to mutagenicity. For example, C:G→T:A, T:A→C:G and characteristic tandem CC→TT transition mutations result from pyrimidine dimers [Chan et al., 1985; Dumaz et al., 1993; Gentil et al., 1996; Naegeli, 1997]. An interesting attribute of the (6 – 4) PP is that it undergoes photoisomerization to a Dewar valence isomer in the presence of UV-
B, while reverting to the conventional (6 – 4) PP structure when exposed to UV-C light [Mitchell and Nairn, 1989; Davies, 1995]. If these lesions are left unrepaired or are not bypassed, they result in cytotoxicity.

UV-C is widely used in laboratory investigations because of its maximal absorption by DNA, producing more photoproducts than the UV-A and UV-B radiations, which are also physiologically relevant UV wavelengths that also cause DNA damage [Kiefer, 2007]. UV-B for instance causes the formation of pyrimidine dimers, but does so less efficiently than UV-C [You et al., 2000; Errol C. Friedberg, 2005; Rastogi et al., 2010]. UV-A damages DNA by inducing DNA adduct formation by photooxidation reactions and by the excitation of endogenous (porphyrins and flavins) and exogenous (psoralens, tetracycline, promazine and methylene blue) photosensitizers [Epe, 1991; Kvam and Tyrrell, 1997; Douki et al., 1999]. In addition, UV-A-mediated photosensitization can induce 8-oxoG formation or an excess accumulation of cyclobutane dimers [Epe, 1991; Rochette et al., 2003]. In mammalian cells, near and far UV radiations are known to cause DNA protein crosslinks, while UV-A radiation results in DNA strand breakages [Peak and Peak, 1986; Errol C. Friedberg, 2005]. UV lesions are repaired by direct reversal of UV-damaged bases, NER, interstrand crosslink (ICL) repair, translesion synthesis, or homologous recombinations (HR), all of which either repair the lesions or enable cells to tolerate their presence [Sancar, 1996; Errol C. Friedberg, 2005; Waters and Walker, 2006; Eppink et al., 2011].

**Exogenous chemical agents**

**Alkylating agents**

Exogenous alkylating agents are primarily produced from dietary components, tobacco smoke, biomass burning, industrial processing and chemotherapeutic agents [Lawley, 1966; AE, 1990; Crutzen and Andreae, 1990]. The electrophilic alkylating agents react with increased affinity to the highly nucleophilic base ring nitrogens, especially the N7 of guanine and N3 of adenine, and slightly less so with the oxygens. Examples of added DNA bases include modified adenine (at N1, N3, N6 and N7), guanine (N1, N2, N3, N7 and O6), cytosine (N3, N4 and O3), thymine (N3, O2 and O4), and alkyl phosphates in the DNA backbone (exocyclic positions on DNA bases are in italicized superscripts) [Singer and Kusmirek, 1982; Singer, 1986; Errol C. Friedberg, 2005]. Mechanistically, alkylating agents add the alkyl group by either 1) an SN1 substitution reaction that progresses via the first order kinetics and involves a carbonium ion intermediate, or, 2) an SN2 substitution reaction that follows the second order kinetics, and in general produces adducts that are less mutagenic and carcinogenic than those of the SN1 pathway [Naegeli, 1997], although evidence has been presented that some SN1 alkylating agents may not proceed via the carbonium ion intermediate [Loechler, 1994].

Most common alkylating agents that are regularly used in labs, including methyl methanesulfonate (MMS), ethyl methanesulfonate (EMS), N-methyl-N′-nitro-N-nitrosoguanidine (MNNG) and methylnitrosourea (MNU) (Figure 3A), react with DNA to generate mutagenic and carcinogenic lesions. For example, MMS produces the mutagenic N7-methylguanine and N3-methyladenine, both of which are susceptible to cleavage of the N-glycosidic bond, thereby generating AP sites, while MNNG and MNU produce O6.
methylguanine, which mispairs with T and induces G:C→A:T mutations [Loechler et al., 1984; Beranek, 1990; Wyatt and Pittman, 2006].

Other classical examples of alkylating agents are sulfur and the nitrogen mustards, first used in World War I, and in many other conflicts since including the present day Syria. The mustards drive SN1 reactions, and are bifunctional in that they carry two reactive groups, instead of one as in monofunctional alkylating agents, and thus have the potential to react with two different sites on the DNA. Such bifunctional reactions result in intra- and interstrand crosslinks, along with the DNA-protein crosslinks, which block DNA metabolic activity [Lawley, 1966; AE, 1990]. These properties of the mustards have been exploited in their use as chemotherapeutic alkylating agents [DeVita and Chu, 2008]. One clinically relevant alkylating agent for chemotherapy is cyclophosphamide (Figure 3B) used in the treatment of lymphomas, leukemias and solid tumors [Emadi et al., 2009]. Another class of crosslinking agents that are used in chemotherapy includes cisplatin (Figure 3B), the first FDA approved platinum compound that is used to treat a wide variety of cancers [Kelland, 2007; Dasari and Tchounwou, 2014]. A crosslinking agent that is not an alkylating agent, psoralen (a furocoumarin) (Figure 3B), intercalates into DNA and cause both interstrand crosslinks and pyrimidine adducts upon photoactivation by UV-A [Yurkow and Laskin, 1991]. The combined psoralen+UV-A or PUVA has been effectively used for treating skin conditions such as psoriasis, eczema and cutaneous T-cell lymphomas. Direct damage reversal, BER and ICL repair are the putative repair pathways that respond to alkylated base damage [Wyatt and Pittman, 2006].

Aromatic amines

Aromatic amines are principally produced from cigarette smoke, fuel, coal, industrial dyes, pesticides and everyday high temperature cooking [Sugimura, 1986; Skipper et al., 2010]. Upon activation by the P450 monoxygenase system, aromatic amines are converted into the carcinogenic (ester and sulfate) alkylating agents that attack the C8 position of guanine [Hammons et al., 1997; Naegeli, 1997]. The most intensively studied examples of aromatic amines are 2-aminofluorene (AF) and its acetylated derivative N-acetyl-2-aminofluorene (AAF) (Figure 3C), which were originally used as insecticides until they were recalled due to their carcinogenic properties [Kriek, 1992]. C8-guanine lesions formed from aminofluorenes are known to form persistent lesions that ultimately give rise to base substitutions and frameshift mutations [Mah et al., 1989; Heflich and Neft, 1994; Shibutani et al., 2001]. The mutagenic properties of the C8-guanine lesion come from its characteristic ability to adopt two conformations while on the DNA [Eckel and Krugh, 1994a]. In the external conformation where the fluorene moiety protrudes out, there is minimal disturbance to Watson-Crick base pairing, which allows these isomers to be effectively bypassed by TLS polymerases [Vooradi and Romano, 2009]. While in internal conformation, the C8-guanine lesion and its partner cytosine are displaced into the minor groove, completely altering the geometry and acting as a very mutagenic substrate on the DNA [Kriek, 1992; Eckel and Krugh, 1994a; Eckel and Krugh, 1994b]. The NER pathway is known to repair C8-guanine adducts in human cells [Mu et al., 2012].
Polycyclic aromatic hydrocarbon (PAH)

Polycyclic aromatic hydrocarbons are carbon compounds with two or more aromatic rings and are generally known to be inert, nonpolar and widely distributed carcinogens in the environment [Harvey, 1991]. Common sources include tobacco smoke, automobile exhaust, charred food and incomplete combustion of organic matter and fossil fuels [Schoket, 1999; Yu, 2002]. The carcinogenicity of these compounds was first documented in 1775, followed by their isolation from coal tar and the later elucidation of their mechanism of action [Butlin, 1892; Phillips, 1983; Fujiki, 2014]. PAHs depend on the P-450 system of the liver to generate reactive intermediates that react with DNA [Phillips, 1983]. Photo-oxidation, one electron oxidation, multiple ring-oxidation and nitrogen-reduction pathways are also known to activate the PAHs [Strniste et al., 1980; Fu, 1990; RamaKrishna et al., 1992; Rogan et al., 1993; Flowers et al., 1997; Penning et al., 1999; Yu, 2002].

Prominent examples of PAHs are naphthalene, anthracene, pyrene, 1-hydroxypyrene, 1-nitropyrene, benzo(a)pyrene and dibenzo[a,l]pyrene. Of these, the most well studied is benzo(a)pyrene (Figure 3D). Upon P-450 activation, benzo(a)pyrene generates the ultimate carcinogen (+)-anti-BPDE [(+)-7,8-hydroxy-9α, 10α-epoxy-7,8,9,10-tetrahydrobenzo(a)pyrene], along with the (+)-anti-BPDE and the (−)-anti-BPDE intermediates. These intermediates first intercalate into DNA, then the C10 position of the BPDE binds to the N2 exocyclic position of guanine to form DNA adducts [Geacintov, 1986; Graslund and Jernstrom, 1989; Cosman et al., 1992]. In terms of carcinogenicity, dibenzo[a,l]pyrene (Figure 3D) is the most potent PAH and poses a major cancer risk to humans [Luch, 2009]. Normally, the excision repair pathways such as NER and BER repair the PAH DNA lesions if they are not bypassed by TLS polymerases [Braithwaite et al., 1998; Jha et al., 2016].

Other reactive electrophiles

Given the space constraint and scope of this manuscript, we will only briefly touch upon a few other relevant reactive electrophiles that damage the DNA. N-nitrosamines, which are potent carcinogens, are byproducts of tobacco smoke and are also encountered by humans in preserved meats. N-nitrosamines have been implicated in the development of esophagus, stomach and nasopharynx cancers [Bartsch and Montesano, 1984; Tricker and Preussmann, 1991; Hecht, 1999; Herrmann et al., 2015]. Another reactive electrophile, 4-nitroquinoline 1-oxide, has both carcinogenic and mutagenic properties (Figure 3E). Upon its metabolic activation to 4-acetoxyaminoquinoline 1-oxide (Ac-4HAQO), 4NQO1 forms covalent adducts with C8 or N2 of guanine and N6 of adenine, as well as causing oxidative stress that results in 8-hydroxyguanine lesion, all of which significantly adds to the strand breakage events and oral carcinogenesis [Galiegue-Zouitina et al., 1985; Galiegue-Zouitina et al., 1986; Kohda et al., 1986; Hawkins et al., 1994; Kanojia and Vaidya, 2006].

Our final notable compound is the hormone estrogen, frequently used in hormonal replacement therapy, which poses a cumulative increased cancer risk after its prolonged use [Cavalieri et al., 2000; Yager and Davidson, 2006]. Epidemiological and clinical trial studies indicate an increased breast cancer risk and other health issues from a combinatorial use of estrogen and progesterone compared to estrogen alone [Yager and Davidson, 2006].
P-450 1BI enzyme complex, constitutively expressed in breast and other tissues, hydroxylates estrogen at position 4 to produce reactive catechol estrogens, which are either oxidized to semiquinones and quinones that react with N3 and N7 position of purines, or generate ROS species [Nutter et al., 1991; Nutter et al., 1994; Errol C. Friedberg, 2005]. Both of these unstable bulky adducts and oxidants produce AP sites and strand breakages [Errol C. Friedberg, 2005]. Estrogen is also implicated in the development of prostate cancer, where strand breaks and lipid peroxidation were the phenotypic readouts in a prostate rat model [Ho and Roy, 1994; Nelles et al., 2011].

**Toxins**

Natural toxins constitute a class of genotoxic and carcinogenic compounds, which are normally used by microorganisms or fungi in defense responses [Ames et al., 1990]. Human and animal exposures result from contaminated cereals, oilseeds, spices, tree nuts, milk and milk products [Lopez et al., 2002]. Aflatoxins are naturally occurring toxins from *Aspergillus flavus* and *Aspergillus parasiticus*, of which aflatoxin B1 is the most potent liver carcinogen [Bennett and Klich, 2003]. After passively diffusing into cells, aflatoxin B1 (Figure 3F) is metabolized by the P-450 complex into the active form, aflatoxin B1-8,9-epoxide. This reactive electrophile then adducts with N7 of guanine to form a positively charged product, 8,9-dihydro-8-(N7-guanyl)-9-hydroaflatoxin B1, which weakens the glycosidic bond resulting in depurination [Essigmann et al., 1977; Smela et al., 2001].

**Environmental stresses**

Environmental sources of stress such as extreme heat or cold, hypoxia, and oxidative stress have been shown to cause DNA damage in human cells [Gregory and Milner, 1994; Gafter-Gvili et al., 2013; Luoto et al., 2013; Neutelings et al., 2013; Kantidze et al., 2016]. These stresses have also been shown to cause mutagenesis at trinucleotide repeats, which are implicated in the development of neurodegenerative disorders via the alt-NHEJ DNA repair pathway [Chatterjee et al., 2015; Chatterjee et al., 2016b]. Equally compelling is the observation that the pathway of environmental stress-induced mutagenesis is akin to the physiological genome instability program operational in many cancer cells [Chatterjee et al., 2015; Chatterjee et al., 2016b]. It is of interest to know whether similar environmental stress-induced phenotypes can be recapitulated in mouse studies.

Other everyday use biological products have increasingly been associated with DNA damage. For example, butyl paraben (BP) and bisphenol A (BPA), found in cosmetics, pharmaceuticals, food-products and beverage processing, are linked to DNA damage in sperm cells [Oishi, 2002; Meeker et al., 2010a; Meeker et al., 2010b; Meeker et al., 2011]. Food preservatives [(sodium benzoate (SB), potassium benzoate (PB) and potassium sorbate (PS)] and food additives [(citric acid (CA), phosphoric acid (PA), brilliant blue (BB) and sunset yellow (SY)] are all known to cause DNA damage [Mamur et al., 2010; Zengin et al., 2011; Yilmaz et al., 2014; Pandir, 2016]. Additionally, plant protection products (PPPs) regularly used by orchard workers have also been associated with DNA damage [Kasiotis et al., 2012]. Such instances stress the importance of global regulatory requirements on the use of chemicals that risk human health, as there may be yet unknown chemicals that have health risks.
DNA damage response (DDR)

After the DNA is damaged, lesion-specific sensor proteins initiate a DNA damage response. The DDR is a collection of mechanisms that sense DNA damage, signal its presence and promote subsequent repair [Harper and Elledge, 2007]. Recruitment of DDR factors is a spatiotemporally regulated process, in which the DDR factors are assembled at the site of damage in a sequential and coordinated manner, as verified by time-lapse microscopy of discrete foci [Harper and Elledge, 2007; Ciccia and Elledge, 2010; Polo and Jackson, 2011]. In addition, chromatin remodeling is an important modulator of DDR response, whereby key post-translational modifications allow assembly of specific DDR and repair factors [Bekker-Jensen et al., 2006; Harper and Elledge, 2007; Misteli and Soutoglou, 2009; Polo and Jackson, 2011; Altmeyer and Lukas, 2013a; Altmeyer and Lukas, 2013b; House et al., 2014]. Mutations affecting DDR network components are the cause of several cancer predisposition syndromes, reflecting their overall importance in avoiding DNA damage-induced human diseases [Ciccia and Elledge, 2010]. However, DNA repair pathways (Figure 4) effectively remove most DNA lesions, which could otherwise result in the formation of mutations or block metabolic processes such as replication and transcription thereby causing senescence and cell death as we discuss below. Readers are directed to excellent reviews on the role of histone modifications during the DNA damage response [van Attikum and Gasser, 2005; Altaf et al., 2007; Zhu and Wani, 2010].

Repair of base DNA damage

Reversal of DNA damage

Small subsets of DNA lesions—UV photolesions and alkylated bases—are simply reversed in an error-free manner. Readers are directed to excellent literature sources on the photolyase-mediated photoreactivation of UV lesions found in lower organisms and marsupials [Kato et al., 1994; Errol C. Friedberg, 2005; Yi and He, 2013]. Here we will briefly discuss the reversal of alkylated DNA damage.

Two different classes of enzymes reverse alkylated bases in humans and mammals. First, the $O^6$-alkylguanine-DNA alkyltransferase (AGT/MGMT) enzyme reverses $O$-alkylated DNA lesions, such as the $O^6$-methyl, ethyl, 2-chloroethyl, benzyl and aliphatic groups, the pyridyloxobutyl adducts of guanine, and even repair the $O^6$-G-alkyl-$O^6$-G interstrand cross-links [Tubbs et al., 2007; Fang et al., 2008; Pegg, 2011]. A single AGT molecule removes the alkylation adduct in a one step reaction by transferring the alkyl group from the oxygen of the DNA base to the cysteine residue in its catalytic pocket [Kaina et al., 2007]. AGT has a special and complex significance in the cancer field. On one hand, AGT’s potential to target a diverse set of substrates is exploited to synthesize pseudosubstrates that can be used in combination with therapeutic alkylating agents to circumvent resistance to cancer chemotherapy [Tubbs et al., 2007]. On the other hand, lack of AGT expression is associated with certain group of cancers [Lee et al., 2011; Mokarram et al., 2013]. In addition, alkyltransferase-like proteins (ATLs), a family of AGT homologs, inhibit the AGT enzyme by directing the repair of bulky alkyl damage to the NER pathway [Margison et al., 2007; Tubbs et al., 2009].

*Environ Mol Mutagen.* Author manuscript; available in PMC 2018 June 01.
The second class of direct reversal enzymes, the AlkB-related α-ketoglutarate-dependent dioxygenases (AlkB), reverse N-alkylated base adducts. There are 9 human homologs of *E. coli* AlkB, which are designated as ALKBH1-8 (Alkylation Repair Homologs) and FTO (Fat Mass and Obesity associated) in human cells [Kurowski et al., 2003; Gerken et al., 2007; Sanchez-Pulido and Andrade-Navarro, 2007; Yi and He, 2013]. For demethylation, the AlkB family proteins hydroxylate the alkyl group in a α-ketoglutarate and iron(II) dependent manner. The oxidized alkyl group is released as formaldehyde, leaving behind the original base [Drablos et al., 2004; Falnes et al., 2007].

**Base excision repair (BER)**

BER corrects those forms of oxidative, deamination, alkylation and abasic single base damage that are not perceived as significant distortions to the DNA helix. In the nucleus, this repair process is mainly active in the G1 phase of the cell cycle [Dianov and Hubscher, 2013]. For BER transactions, chromatin remodeling at the DNA damage site is followed by lesion recognition by a DNA glycosylase [Odell et al., 2013]. At least 11 different DNA glycosylases can recognize and excise a damaged base from undistorted helices, as well as ones flipped out from the major groove [Huffman et al., 2005; Krokan and Bjoras, 2013]. In terms of function, DNA glycosylases are either monofunctional, with only a glycosylase activity, such as the uracil glycosylases, N-methylpurine DNA Glycosylase (MPG), and MutY Homolog (MUTYH), or are bifunctional with a glycosylase and an additional β-lyase activity. Examples of the latter include the Nth-like DNA glycosylase 1 (NTHL1), Nei-like DNA glycosylase 1 (NEIL1) and Nei-like DNA glycosylase 2 (NEIL2) [Jacobs and Schar, 2012]. It should be noted that 8-oxoguanine DNA glycosylase (OGG1) and NEIL3 function as both mono- and bifunctional glycosylases [Svilar et al., 2011]. An abasic site created from the monofunctional glycosylases gets committed to the short-patch-repair pathway, while the bifunctional glycosylases initiate the long-patch repair pathway of BER [Dianov and Hubscher, 2013].

In short patch repair, the abasic site is the substrate for the AP endonuclease (APE1 in human cells), which cleaves the phosphodiester bond 5′ to the abasic site and generates a hydroxyl residue at the 3′-end while leaving a deoxyribose phosphate (dRP) at the 5′-end. This repair gap is tailored by the 5′-dRP lyase activity of POL β (gap tailoring), followed by filling the single nucleotide gap by POL β and ligation by either LIG1 (DNA ligase 1) or a complex of LIG3 (DNA ligase 3) and XRCC1 (X-ray repair cross-complementing protein 1) [Almeida and Sobol, 2007]. In long patch repair, the repair gap left behind from the bifunctional glycosylase is tailored by the 3′ phosphodiesterase activity of APE1. Thereafter, POL β (in non-proliferating cells) or POL δ/ε (in proliferating cells) synthesize in a strand-displacement manner, which is then followed by flap removal by the flap endonuclease and a LIG1-mediated ligation [Akbari et al., 2009; Svilar et al., 2011].

While BER of 8-oxo-G lesions at CAG repeats is implicated in triplet repeat instability, downregulation of OGG1 is associated with aging, neurodegenerative disorders and cancer [Kovtun et al., 2007; Tian et al., 2009; Curtin, 2012; Mollersen et al., 2012; Chatterjee N., 2013; Krokan and Bjoras, 2013; Chatterjee et al., 2015]. Specifically, mutations in POL β are found in solid cancers and POL β variants can act as dominant negative and sequence...
specific mutators [Wang et al., 1992; Starcevic et al., 2004; Lang et al., 2007; Murphy et al., 2012]. In addition, PARP1 (Poly [ADP-ribose] polymerase 1) has also been shown to be required for the repair of single strand breaks and damaged purine bases by a sub-pathway of BER [Krokan and Bjoras, 2013; Reynolds et al., 2015]. Finally, mitochondria are also known to carry out both short and long patch BER, where the synthesis step is carried out by POL γ; all of which adds to the significance of this repair pathway in the maintenance of global genome stability [Akbari et al., 2008; Liu and Demple, 2010]. Readers are directed to these reviews for an overview of mitochondrial BER [Bauer et al., 2015; Prakash and Doublie, 2015].

Repair of multiple and bulky base damage

Nucleotide excision repair (NER)

Nucleotide excision repair is the pathway of choice to remove bulky lesions such as CPDs and (6 – 4)PP from UV radiation, benzo[a]pyrene adducts, or damage from chemotherapeutic agents. NER deficiency results in a number of different human syndromes: Xeroderma Pigmentosum (XP), which is associated with a predisposition to skin cancers; Cockayne Syndrome (CS); rare UV-Sensitive Syndrome (UV³S³); and Cerebro-Oculo-Facio-Skeletal syndrome (COFS) [Errol C. Friedberg, 2005; Vermeulen and Fousteri, 2013]. However, like the BER pathway, NER contributes to the instability mechanisms in triplet repeat disorders [Lin et al., 2006; Hubert et al., 2011; Dion, 2014]. To begin NER, chromatin remodeling mediated both by chromatin and NER components makes way for the NER machinery on the specified DNA lesions [Scharer, 2013]. There are two major branches of NER: global genome NER (GG−NER) and transcription−coupled NER (TC−NER).

In GG-NER, the main DNA damage sensor is the XPC (Xeroderma Pigmentosum, complementation group C) protein, complexed with RAD23B (UV excision repair protein Radiation sensitive 23B) protein and CETN2 (Centrin 2). This complex scans for the presence of transient single−stranded DNA (ssDNA) caused by disrupted base pairing due to the lesion [Masutani et al., 1994; Nishi et al., 2005]. For repair of UV-induced CPDs, the ultraviolet-damaged DNA damage−binding protein (UV–DDB) complex, consisting of DDB1 (XPE–binding factor) and the GG−NER−specific protein DDB2, directly binds to UV−radiation−induced lesions and then stimulates the binding of XPC [Chu and Chang, 1988; Wakasugi et al., 2002; Scrima et al., 2008]. XPC bound to the lesion becomes the substrate for the transcription initiation factor II H (TFIIH) complex, a transcription initiation and repair factor composed of ten protein subunits that can switch functions in both transcription initiation and in NER [Yokoi et al., 2000; Volker et al., 2001; Compe and Egly, 2012]. The final step of dual excision and gap filling is coordinated to prevent the ssDNA gap formation that can potentially trigger DDR signaling [Marini et al., 2006; Marti et al., 2006; Mocquet et al., 2008].

The incision step of GG-NER commits the assemblage of all the proteins to NER. It involves the use of structure specific endonucleases XPF–ERCC1 and XPG (also known as ERCC5), which cuts the damaged strand short distances away from the 5’ and 3’ end of the lesion respectively [Fagbemi et al., 2011]. The replication proteins PCNA (proliferating cell
nuclear antigen), RFC (replication factor C), POL δ, POL ε or POL κ, and LIG1 or XRCC1–LIG3 carry out the final step of gap–filling synthesis and ligation. Proliferative status of the cell determines the choice of polymerase used. For example, POL ε–dependent repair predominates in non–replicating cells, while POL δ and POL κ are the main NER polymerases in replicating cells. LIG1-dependent ligation occurs in replicating cells. However, XRCC1–LIG3 complex seals the gap in non-proliferating cells because of low expression of dNTPs and LIG1 in these cells, [Moser et al., 2007; Ogi et al., 2010].

The second NER pathway, TC–NER, is initiated by a lesion-stalled RNA polymerase II, which begins with the recruitment of TC–NER–specific proteins CSA (Cockayne syndrome WD repeat protein A; also known as ERCC8) and CSB (Cockayne syndrome protein B; also known as ERCC6), which are essential for additional assembly of other TC–NER components [Fousteri et al., 2006]. These include the core NER factors (except for the GG–NER– specific UV–DDB and XPC complexes) and TC–NER–specific proteins, such as UVSSA (UV–stimulated scaffold protein A), USP7 (ubiquitin–specific–processing protease 7; also known as ubiquitin C–terminal hydrolase 7), XAB2 (XPA–binding protein 2; also known as pre–mRNA–splicing factor SYF1) and HMGN1 (high mobility group nucleosome–binding domain–containing protein 1; also known as non–histone chromosomal protein HMG14) [Fousteri et al., 2006; Schwertman et al., 2012]. Once localized at the lesion site, the CSA-CSB complex backtracks (or reverse translocates) RNA polymerase II, exposing the lesion site. TFIIH is recruited to the lesion. The subsequent sequence of events is predicted to be the same as in GG–NER as the lesion is removed from the transcribed strand [Marteijn et al., 2014].

Mismatch repair (MMR)

MMR is an evolutionarily conserved, post replicative repair pathway that contributes to replication fidelity by at least 100-fold [Kunkel, 2009; Arana and Kunkel, 2010]. Typical substrates for the MMR pathway are base mismatches that have arisen during replication and the insertion-deletion loops (IDLs) within repetitive DNA sequences that have resulted from strand slippage events [Errol C. Friedberg, 2005; Jiricny, 2006]. MMR is also implicated in a variety of cellular processes including microsatellite stability, meiotic and mitotic recombination, DNA-damage signaling, apoptosis, class-switch recombination, somatic hypermutation and triplet-repeat expansion [Jiricny, 2006; Jiricny, 2013; Chatterjee et al., 2016a]. Germline mutations in the MMR genes result in Lynch syndrome (also known as hereditary nonpolyposis colorectal cancer or HNPCC), which presents as a familial susceptibility to colon and ovarian cancers as well to a number of other cancers [Peltomaki, 2001]. Chromatin modifications have recently been shown to pave the way for the MMR proteins to gain access to the DNA lesion and initiate repair [Li et al., 2013; Li, 2014].

Of the eight known MSH (MutS homolog) polypeptides in eukaryotes, humans employ the MutSa heterodimer (MSH2/MSH6) to recognize base mismatches and one-to-two nucleotide IDLs, and the MutSβ heterodimer (MSH2/MSH3) to recognize large IDLs [Kunkel and Erie, 2005; Sachadyn, 2010]. The previously accepted model was that after the lesion recognition step, the MutS complex translocates along the DNA in an ATP-dependent manner to make way for the downstream MMR components [Jiricny, 2013]. Recently, the
Modrich lab has shown that MutL can trap MutS at the mismatch before it forms a sliding clamp [Qiu et al., 2015]. Next, the MutL complexes are recruited on to DNA and among the 4 known human MutL homologs; the MutLα heterodimer (MLH1/PMS2 heterodimer) plays a major role in MMR [Nicolaides et al., 1994; Papadopoulos et al., 1994; Li and Modrich, 1995; Lipkin et al., 2000]. MutLα regulates termination of mismatch-provoked excision, and its endonuclease activity plays a role in the 3′ nick-directed digestion by EXO1 (Exonuclease 1) in a PCNA/RFC dependent manner [Zhang et al., 2005; Kadyrov et al., 2006]. EXO1 also carries out the 5′ directed mismatch excision creating a gap that is stabilized by RPA [Genschel and Modrich, 2003; Zhang et al., 2005]. POL δ, RFC, HMGB1 (high mobility group box 1 protein) and LIG1 orchestrate the final steps of new DNA synthesis and ligation [Genschel and Modrich, 2003; Yuan et al., 2004; Guo et al., 2006]. PCNA plays an important role in both the initiation step of MMR and in the subsequent DNA synthesis by interacting and localizing MutSa/β and MutLα complexes at the lesion site [Umar et al., 1996; Lau and Kolodner, 2003; Jiricny, 2006]. In addition to mismatch repair and other cellular functions, the mismatch repair genes have recently been shown to be repressed in response to environmental stresses, such as hypoxia, benzo[a]pyrene, inflammation and even tumor microenvironment [Mihaylova et al., 2003; Bindra and Glazer, 2007; Nakamura et al., 2008; Edwards et al., 2009; Chen et al., 2013]. It remains to be seen whether other exogenous stresses can also suppress the expression of MMR genes.

**Interstrand cross-link repair (ICL)**

Interstrand crosslinks are lesions in which two bases from complementary strands are covalently linked due to damage to the DNA from crosslinking agents such as platinum compounds, nitrogen mustards, MMC, psoralens and alkylating agents [Clauson et al., 2013]. Additional modifications from these crosslinking agents include bases monoadducts, intrastrand crosslinks, and DNA-protein crosslinks. These lesions are recognized and repaired by the Fanconi anemia (FA) proteins. Mutations in the FA genes are the cause of the autosomal recessive FA disorder. FA disorder is a heterogeneous and rare genetic disorder characterized with a high frequency of hematological abnormalities, congenital anomalies and a general predisposition to cancers [Kee and D’Andrea, 2012]. Classically, FA is diagnosed by assessing cellular hypersensitivity—chromosomal breaks and chromosomal radial formations—to DNA ICL agents such as diepoxybutane (DEB) and MMC [D’Andrea, 2010]. In addition, DEB-induced chromosome breakage assay is widely used for the primary diagnosis of FA [Auerbach, 1993].

Interstrand crosslink repair is initiated by chromatin loading of the FA proteins in a cell cycle-dependent manner [Mi and Kupfer, 2005; Kim et al., 2008]. The FA family consists of 21 different functional complementation groups (A, B, C, D1, D2, E, F, G, I, J, L, M, N, O, P, Q, R, S, T, U, V), which are known to suppress ICL sensitivity [Bluteau et al., 2016; Michl et al., 2016]. Upon ICL damage, FANCM is recruited to the damaged site along with FAAP24 (Fanconi Anemia associated protein of 24 kDa) and MFH (histone fold protein complex) [Ciccia et al., 2007; Niedernhofer, 2007; Yan et al., 2010]. Replication fork remodeling stimulated by MFH and FANCM promotes Holliday junction migration and the
creation of ssDNA gaps [Gari et al., 2008a; Gari et al., 2008b; Huang et al., 2010]. RPA bound ssDNA signals ATR activation [Zou and Elledge, 2003; Ben-Yehoyada et al., 2009]. ATR phosphorylates downstream target CHK1, which in turn phosphorylates FANCE, FANCD2, FANC1 and MRN [Andreassen et al., 2004; Smogorzewska et al., 2007; Wang et al., 2007a; Cui et al., 2009; Duquette et al., 2012]. In a yet unknown way, other FA core complex components (FANCA/B/C/E/F/G/L/T) assemble at the damaged site and activate the phosphorylated FANCI–FANCD2 heterodimer through FANCL-mediated monoubiquitination [Smogorzewska et al., 2007]. This activation of FANCI-D2 marks as the major activation switch for the FA pathway [Wang, 2008; Tomida et al., 2013].

Subsequently, excision of the DNA strand (5′ and 3′) of the lesion is coordinated by structure specific endonucleases—XPF-ERCC1, MUS8-EME1, SLX4-SLX1, FAN1, SNM1A/SNM1B—in an as-yet unclear fashion [Clauson et al., 2013]. Next, depending upon the proliferation state of the cells, ICL repair bifurcates into either of the two pathways below.

In replicating cells, the presence of ICL stalls the ongoing replication on the leading strand, as well as on the 5′ end of the lagging strand at some distance from the lesion [Raschle et al., 2008; Knipscheer et al., 2009]. Next, XPF-ERCC1 and SNM1A induce incisions on either side of the lesion that unhook the ICL from the lagging strand, thereby producing a gap [Wang et al., 2011]. The leading strand with the ICL becomes the template for new DNA synthesis—by TLS polymerases POL ι, POL κ, POL ν and REV1—that proceeds up to the lesion, bypasses it, and extends beyond the lesion until it reaches the first downstream Okazaki fragment [Minko et al., 2008; Raschle et al., 2008; Yamanaka et al., 2010; Ho et al., 2011; Klug et al., 2012]. After this step, the 3′ overhang of the leftover lagging strand invades the newly synthesized strand in a RAD51-dependent manner in a tightly coordinated manner [Long et al., 2011]. Interestingly, resolution of this HR intermediate depends on the FANCD2-FANCI complex. NER pathway eventually removes the ICL hook that was still hanging on to the leading strand.

In non-replicating cells, ICL repair of psoralen, MMC, cisplatin, and alkyl ICLs depend on NER and TLS polymerases such as REV1 and POL ζ [Clauson et al., 2013]. Helix-distorting lesions are recognized by both the GG-NER and the TC-NER pathways to initiate repair, although some lesions such as cisplatin may escape recognition [Enoiu et al., 2012]. After lesion recognition, the components of the NER pathway are known to cut only the 5′ side of the lesion, with further incisions possibly aided by the MutSβ complex create a ssDNA-gap [Bessho et al., 1997; Mu et al., 2000; Smeaton et al., 2008; Zhao et al., 2009]. Next, the error-prone TLS polymerase synthesize across the gap and finally a second round of NER removes the ICL hook on the other strand [Clauson et al., 2013].

Recent studies provide a striking evidence of crosstalk between the FA and other repair pathways. For example, the FA pathway suppresses non-homologous end joining (NHEJ) by interacting with CtIP, and recruiting the NHEJ-inhibiting molecules—PARP1 and RAD18 to the DNA [Saberi et al., 2007; Ceccaldi et al., 2016a]. Similarly, inhibition of NHEJ components alleviates the sensitivity of FA deficient cells to crosslinking agents, while the same FA deficient cells show enrichment of 53BP1, RIF1 and RAP80 components at damaged chromatin [Adamo et al., 2010; Pace et al., 2010; Ceccaldi et al., 2016b; Renaud et
al., 2016]. A second intriguing example of the FA pathway’s crosstalk is its undefined role in promoting alternate end joining events, as seen in patients with FANCA mutations who lack immunoglobulin class switch recombination. In addition, loss of FANCD2 confers a synthetic lethal phenotype in POL θ null mice [Nguyen et al., 2014; Ceccaldi et al., 2015; Howard et al., 2015]. Finally, the FA pathway has now been implicated in trinucleotide repeat instability [Chatterjee N., 2016].

**Translesion Synthesis (TLS)**

Translesion synthesis is carried out by highly conserved TLS polymerases. TLS polymerases are specialized DNA polymerases that can replicate opposite and past aberrant DNA lesions in a relatively lower fidelity manner than replicative DNA polymerases [Sale, 2013]. If incorrect nucleotides were incorporated by TLS polymerases, they would become mutations in the next round of replication, which propel tumorigenesis and disease, but can also contribute to the overall fitness and evolution of organisms. A total of eleven TLS polymerases are known (REV1, POL η, POL ι, POL ξ, POL ζ, POL μ, POL β, POL ν, POL θ), which are distributed in four families (Y, B, X and A) and PrimPol (Table 1). Although all TLS polymerases are less accurate than replicative polymerases, certain TLS polymerases are able to copy relatively accurately over certain cognate lesions. For example, cyclobutane thymine-thymine CPDs are cognate lesions for POL η (Table 1). The frequency of DNA synthesis errors during translesion synthesis depends on several factors, such as whether the lesion is a cognate for the particular TLS DNA polymerase, the biochemical characteristics of the particular TLS polymerase and the DNA sequence context [Pages and Fuchs, 2002; McCulloch et al., 2004; Waters and Walker, 2006]. XPV patients, who exhibit a photosensitive phenotype with a high incidence of skin cancer, highlight the physiological significance of certain TLS DNA polymerases bypassing particular lesions. These patients lack the POL η enzyme and are highly susceptible to UV radiation because alternate TLS polymerases (POL ι and POL κ) instead bypass the UV-induced cyclobutane dimers (CPD) in an error-prone fashion [Yamada et al., 2000; Sweasy et al., 2006; Wang et al., 2007b; Ziv et al., 2009].

The TLS polymerases’ fascinating ability to help cells tolerate DNA damage arises from their structural and biochemical features [Rothwell and Waksman, 2005; Pavlov et al., 2006; Waters and Walker, 2006]. Unique functional attributes of TLS polymerases that distinguishes them from the classical replicative polymerases, stems from their discrete physical features. Notable features include the very limited sequence homology to replicative DNA polymerases, the absence of a 3′-5′ exonuclease domain to proofread incoming nucleotides, and their smaller thumb and finger domains, which make fewer contacts with the DNA than those found in replicative DNA polymerases [Rothwell and Waksman, 2005; Waters and Walker, 2006; Sale, 2013]. These structural differences orient the thumb, fingers, palm, and little finger catalytic domains into a relatively larger open active site, while being aided by other physical features such as the polymerase-associated domain (PAD), wrist, and N-clasp region found in Y Family DNA polymerases that facilitate additional DNA binding. Together, these structural features provide the TLS polymerases with a unique architecture that enable them to bypass DNA damage or fill ssDNA gaps.
Two models have been proposed to explain the DNA lesion bypass process via translesion synthesis. In the polymerase switch model, TLS polymerases come together sequentially in a two-step process to replicate past the DNA lesion at a stalled replication fork. First, an ‘inserter’ TLS enzyme, usually a POL $\eta$, POL $\iota$, or POL $\kappa$, and less often REV1 or POL $\zeta$, incorporates a nucleotide opposite the DNA lesion [Korzhnev and Hadden, 2016]. In the second step, an extender TLS enzyme, a role usually fulfilled by POL $\zeta$ exclusively but in some cases by POL $\kappa$, replaces the inserter polymerase and extends the primer-template termini [Washington et al., 2002; Yuji Masuda, 2016]. This two-step model is proposed to direct both the error-free and error-prone translesion DNA synthesis across the damage [Shachar et al., 2009]. The central molecule that orchestrates both the insertion and extension step is REV1, and by way of its unique scaffolding function facilitates an assemblage of the TLS polymerases by binding to an RIR-containing polymerase—POL $\eta$, POL $\iota$, or POL $\kappa$—via one interface, and also to POL $\zeta_4$ (REV3-REV7-POLD2-POLD3), via a second interface, a central step in this model’s execution [Wojtaszek et al., 2012a; Wojtaszek et al., 2012b]. Moreover, POLD3, which is part of the POL $\zeta_4$ complex, interacts with REV1 via its RIR, thereby assisting the switch from RIR directed insertion to POL $\zeta_4$-mediated extension during damage bypass [Pustovalova et al., 2016].

In the gap-filling model, single strand gaps left behind by replicative polymerases during replication or via an incomplete DNA repair process, such as during immunoglobulin gene hypermutation, are the targets of TLS synthesis [Sale et al., 2009]. Usually, these type of TLS events are expected to fall outside of the S phase, but based on the type of DNA lesion, a cell-cycle independence is sometimes conferred [Quinet et al., 2016]. Using a gapped plasmid assay, it has been shown that TLS is as high or higher in G2 compared to S phase in human cells, with slightly higher amounts of POL $\eta$ in G2 compared to S phase of the cell cycle [Diamant et al., 2012]. An exact order of events for a gap-TLS is still unknown, with only a few isolated studies implying the role of TLS polymerases in gap filling. For example, REV1 is very important in mouse cells for synthesis across post-replicative gaps where REV1 gets recruited to the gaps by the 5$'$-end, unlike the gap-filling step of NER, where POL $\kappa$ is the polymerase of choice [Ogi and Lehmann, 2006; de Groote et al., 2011; Sale, 2013]. Likewise, REV3 engages in TLS across gaps opposite 6-4 photoproducts [Quinet et al., 2016], all suggestive of a role of TLS at replicating across ssDNA gaps.

In addition to their traditional DNA damage bypass functions, TLS polymerases are now known to play a role in other cellular pathways. For instance, as previously discussed, TLS polymerases are required for ICL repair and can also play role in the BER and NER pathways to synthesize new DNA after the excision step. Exogenous stressors—for example UV-C, MNNG, and BPDE—regulate the transcriptional expression of POL $\eta$, POL $\iota$, POL $\kappa$ and POL $\zeta$ [Zhu et al., 2003; Yu et al., 2004; Liu and Chen, 2006; Zhu et al., 2010; Zhu et al., 2012]. Likewise, an HSP90 inhibitor reduces expression of REV1 and POL $\eta$ in human cells, indicating an evolutionary regulation of these polymerases [Sekimoto et al., 2010; Pozo et al., 2011]. Interestingly, the TLS polymerases, REV1 and REV3, were also implicated in the development of chemoresistance in human cells and mice models, opening
the possibility for a whole new class of promising chemotherapeutic drugs [Doles et al., 2010; Xie et al., 2010; Xu et al., 2013].

Repair of DNA breaks

Single stranded break repair (SSBR)

Single strand breaks (SSBs) are often generated from oxidative damage to the DNA, from abasic sites, or from erroneous activity of the DNA topoisomerase 1 (TOP1) enzyme [Wang, 2002; Hegde et al., 2008]. Unresolved SSBs often collapse DNA replication, stall ongoing transcription, and effect PARP1 activation, which releases cellular NAD\(^{+}\), ATP and apoptosis inducing factor (AIF) in cells [Zhou and Doetsch, 1993; Heeres and Hergenrother, 2007]. At least two human genetic disorders, spinocerebellar ataxia with axonal neuropathy 1 (SCAN1) and ataxia-oculomotor apraxia 1 (AOA1), are associated with an abortive SSBR. These patients often manifest genetic instability and high incidence of cancers [El-Khamisy et al., 2005; Reynolds et al., 2009]. SSBR is predicted to occur through three different pathways depending on the source of SSB.

In the long patch SSBR pathway, SSBs are transiently detected by PARP1, which undergoes a rapid cycle of poly(ADP) ribosylation and dissociates to detect the next SSB [D’Amours et al., 1999; Davidovic et al., 2001]. After this, the ends undergo end processing by the apurinic-apyrimidic endonuclease 1 APE1, PNKP (polynucleotide kinase 3′-phosphate) and aprataxin (APTX) [McKinnon and Caldecott, 2007]. Next, FEN1 removes the damaged 5′ termini aided by PARP1 and PCNA, leaving behind a ssDNA gap, which is filled by POL \(\beta\), in combination with POL \(\delta/\epsilon\). The final step of ligation is carried out by the LIG1, which is dependent on the presence of PCNA and XRCC1 [Lan et al., 2004; Mortusewicz et al., 2006; McKinnon and Caldecott, 2007]. In the short patch SSBR pathway, SSBs generated during the BER are recognized by APE1 followed by a similar end-processing pathway as the long patch repair. The gap-filling step, however, is only carried out by POL \(\beta\) enzyme, followed by LIG3-catalyzed ligation [McKinnon and Caldecott, 2007]. Finally, the TOP1-SSB pathway is a variant of the PARP1-dependent long patch repair in which the end-processing is carried out by the TDP1 (tyrosyl-DNA phosphodiesterase 1) enzyme that removes the TOP1 from the 3′-end [Caldecott, 2008].

Double strand break repair (DSBR)

Highly toxic DSBs are induced by various chemical and physical DNA damaging agents [Pfeiffer et al., 2000]. Unresolved DSBs are implicated in various human disorders and cancers [Jackson and Bartek, 2009]. We will briefly discuss the two major pathways—homologous recombination (HR) and non-homologous end joining (NHEJ)—that organisms have evolved to resolve the DSBs. Chromatin modification is the first event that registers the presence of a DSB and triggers a cascade of events including ATM activation, targeted phosphorylation of H2AX, chromatin PARylation, MDC1 recruitment and finally recruitment of 53BP1 and BRCA1 [Rogakou et al., 1998; Rothkamm et al., 2003; Gottschalk et al., 2009; Chou et al., 2010; Lukas et al., 2011; Price and D’Andrea, 2013; Liu et al., 2014]. Interestingly, both 53BP1 and BRCA1 exhibit an antagonistic influence on
each other and 53BP1 depletion rescues embryonic lethality of BRCA1 null mice [Xie et al., 2007; Cao et al., 2009; Bunting et al., 2010].

In the NHEJ pathway of DSBR, 53BP1 plays an important regulatory role by recruiting the NHEJ components to the break site, activating checkpoint signaling and facilitating synapsis of the two ends [Panier and Boulton, 2014]. The Ku (Ku70 and Ku80) heterodimer is the first to recognize and bind the DSBs within seconds to prevent end resection and serves as a scaffold to recruit other NHEJ components [Pang et al., 1997; Mari et al., 2006; Soutoglou et al., 2007; Mimitou and Symington, 2010]. Other recruited components include DNA-PKcs, XRCC4, LIG4 and XLF (XRCC4-like factor), APLF (Aprataxin-and-PNK-like factor) [and also TdT (terminal deoxynucleotidyl transferase) in lymphocytes] [Gottlieb and Jackson, 1993; Nick McElhinny et al., 2000; Costantini et al., 2007; Yano et al., 2008; Grundy et al., 2013]. Recent studies indicate that the order of recruitment of these components may depend on the complexity of DNA damage; for instance, DNA-PKcs recruitment depends on the nature of the break [Mari et al., 2006; Yano and Chen, 2008]. However, once DNA-PKcs gets recruited, it is activated in a DNA dependent manner; it pushes Ku inwardly on the DNA and then phosphorylates other nearby components, including autoautophosphorylating itself [Gottlieb and Jackson, 1993; Yoo and Dynan, 1999; Weterings and Chen, 2008]. At the same time, XRCC4 is believed to help stabilize the NHEJ complex by tethering the ends and acting as an additional scaffold with Ku to recruit other components [Malivert et al., 2010; Hammel et al., 2011; Andres et al., 2012]. Once the ends are bridged and stabilized, Artemis, PNKP, APLF, WRN, Aprataxin and Ku initiate DNA end processing, which involves removing groups that are blocking the ends and resecting the resultant naked strands [Ma et al., 2002; Bernstein et al., 2005; Ahel et al., 2006; Perry et al., 2006; Roberts et al., 2010; Li et al., 2011]. The gaps left behind after resection are filled by family X polymerases in a template-dependent (POL μ) or template-independent (POL λ) manner [Ramadan et al., 2004; Roberts et al., 2010]. LIG4 joins the ends and completes the NHEJ process [Grawunder et al., 1997].

The HR pathway consist of a set of related sub-pathways that utilize DNA strand invasion and template-directed DNA repair synthesis to effect a high-fidelity repair [Li and Heyer, 2008]. In addition to the traditional DSBR-induced HR pathway, synthesis-dependent strand annealing (SDSA) and break-induced repair (BIR) are two other variations following the HR premise [Li and Heyer, 2008]. Here, we will very briefly summarize the HR pathway of DSBR.

The MRN (MRE11-RAD50-NBS1) complex initiates HR at a DSB, where it recognizes and binds the DSB and then recruits ATM and TIP60 to the DNA [Sun et al., 2005; Stracker and Petriti, 2011]. Activated ATM (from TIP60) phosphorylates H2AX, which then serves as an anchor for MDC1 [Bhatti et al., 2011]. Next, MDC1 is phosphorylated by ATM, and the phosphorylated MDC1 functions as a scaffold to bring in the ubiquitin E3 ligases RNF8 and RNF168 [Altmeyer and Lukas, 2013b]. Both of these E3 ligases ubiquitinate H2AX, which then serves as a docking site for 53BP1 and BRCA1. In the S/G2 phase where HR is predominant, BRCA1 (recruited by ubiquitinated chromatin) successfully counteracts 53BP1 and initiates ubiquitination of the downstream component, CtIP [Yu et al., 2006;
Chapman et al., 2012]. At this time, the other HR components, RPA, and RAD51 proteins make their way on to the DNA.

The next step of end resection involves a 5′-to-3′ nucleolytic degradation to generate 3′ overhangs, committing cells to the HR pathway. Initial resection occurs by the endonuclease activity of MRN, with the help of CtIP, followed by long-range resection by EXO1 or BLM together with DNA2 [Chen et al., 2008; Nimonkar et al., 2011]. Next, RPA coats the 3′ overhang, which is then displaced by RAD51, generating a nucleoprotein filament. BRCA2 and PALB2 aid in the formation of the nucleoprotein filament formation that invades a nearby duplex DNA forming a D-loop [Zhang et al., 2009; Holloman, 2011]. Several other proteins function together at this step. For the strand to invade the template DNA, RAD54 and RAD54B remove RAD51 and allow the 3′-OH group to prime synthesis by Polymerases δ, κ and ν [Mazin et al., 2010; Sebesta et al., 2013]. If the new DNA synthesis stops after a limited distance, as is the case in SDSA, the RTEL1 enzyme dissolves the D-loop [Barber et al., 2008]. Otherwise, the Holliday junction is collectively processed by the BLM-TOPOIII-RMI1-RMI2 complex, GEN1 endonuclease, the MUS81-EME1 complex and the SLX1-SLX4 complex [Chang et al., 2008; Ciccia et al., 2008; Xu et al., 2008; Fekairi et al., 2009; Rass et al., 2010].

DNA damage and telomeres

Telomeres are well-conserved nucleoprotein structures found at the end of linear chromosomes that help differentiate normal chromosomal ends from DSBs [Longhese, 2008; Shammas, 2011]. Telomeric DNA consists of tandem repetitive DNA (TTAGGG in humans), where the G-rich strand (also called the G-tail), bound by shelterin protein POT1 (protection of telomeres 1), extends beyond the complementary C-rich strand and invades into the double-stranded telomeric DNA. The t-loop thus generated complexes with other shelterin proteins such as TRF1 (telomeric-repeat binding factor 1), TRF2, TIN2 (TRF-interacting protein 2), the transcriptional repressor/activator protein RAP1, and the TPP1 (POT1- and TIN2- organizing protein), which together prevent the chromosomal ends from being recognized as DNA damage [Takai et al., 2003; d’Adda di Fagagna et al., 2004; Liu et al., 2004; de Lange, 2005]. In addition, telomeric DNA is replicated and maintained by a specialized ribonucleoprotein complex called telomerase [composed of a telomere RNA component (TERC) and a telomere reverse transcriptase (TERT)], which is the only positive regulator of telomere length [Bachand et al., 2001; Blasco, 2003]. A decline in telomerase activity contributes to telomere attrition, which is associated with aging, cancer and several inherited bone marrow failure (IBMF) disorders [Chang and Harley, 1995; Shay et al., 2001; Alter et al., 2015].

Deprotected telomeres elicit a DNA damage response, recruiting DSBR components that attempt to repair the exposed ends, causing deleterious nucleolytic degradation, recombination, and chromosomal fusions [Longhese, 2008]. For example, short telomeres often assemble DDR factors such as 53BP1, ATM, γH2AX, and MRE11 as foci, called Telomere Dysfunction-Induced Foci (TIF), which are highly prone to NHEJ-mediated end-to-end fusion [Takai et al., 2003; Hewitt et al., 2012; Marcand, 2014]. Surprisingly, telomere maintenance requires the presence of some of the same components of DSBR/NHEJ
components—for example, Ku and the MRN complex—complicating our understanding of the exact mechanism of telomere stability [Maser and DePinho, 2004; Marcand, 2014]. Establishing the dynamics of this telomere biology is an active area of research.

Recently, several environmental toxins have been implicated in telomere shortening. For example, tobacco smoke causes telomere shortening via oxidative stress [Valdes et al., 2005; McGrath et al., 2007; Song et al., 2010; Babizhayev and Yegorov, 2011]. Obesity has also been associated with accelerated shortening of telomeres in adipose tissue of mice, where levels of ROS were high [Song et al., 2010]. Likewise, white blood cells of obese women harbor shorter telomeres than lean women [Valdes et al., 2005]. Genotoxic stressors such as certain pollutants (e.g. toluene and benzene, and PAHs) are also associated with telomere shortening [Hoxha et al., 2009; Pavanello et al., 2010; Trusina, 2014]. Finally, everyday stresses, including psychological stress, is known to shorten telomeres, whereas meditation and mindfulness may bring about an opposite effect [Epel et al., 2004; Cherkas et al., 2006; Simon et al., 2006; Epel et al., 2009; Mathur et al., 2016]. Critically shortened telomeres are associated with aging and cancer [Shammas, 2011].

**Conclusion**

In conclusion, DNA is continually being exposed to both endogenous and exogenous DNA damaging agents that chemically modify the DNA constituents. Unresolved DNA damages are implicated in human diseases and cancers. However, robust DNA repair and damage tolerance pathways help remove or tolerate the lesions to allow survival (Figure 4). An understanding of these pathways helps evaluate possible toxic exposures and design strategies to control deleterious consequences on human health.

**Acknowledgments**

The writing of this review was supported by National Institute of Environmental Health Sciences grants ES-015818 to G.C.W. and P30 ES-002109 to the MIT Center for Environmental Health Sciences. G.C.W. is an American Cancer Society Professor. Nimrat Chatterjee and Graham Walker wrote the manuscript.

**References**


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Potenski CJ, Klein HL. How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. Nucleic Acids Res. 2014; 42(16):10226–10234. [PubMed: 25159610]


Price BD. Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Environ Mol Mutagen. Author manuscript; available in PMC 2018 June 01.


Figure 1.
Common DNA base lesions. A) Normal structures of DNA bases: adenine (A), guanine (G), cytosine (C) and thymine (T). B) Deaminated bases: hypoxanthine, xanthine, uracil and thymine arising from deamination of exocyclic bases of adenine, guanine, cytosine and 5-methylcytosine (5-mC) respectively. C) Oxidized DNA bases: formamidopyrimidine derivative of adenine (Fapy-A), 7,8 dihydro-8-oxoguanine (8-oxo-G) and thymine glycol. D) Methylated DNA bases: N3-methyladenine, N7-methylguanine, O\(^6\)-methylguanine, N3-methylcytosine, O\(^4\)-methylthymine, O\(^4\)-ethylthymine and N3-methylthymine.
Figure 2.
Main UV radiation-induced DNA base lesions. A) Representative cyclobutane pyrimidine dimers (CPD). Shown here are cyclobutane thymine dimers. B) Representative pyrimidine (6 – 4) pyrimidone photoproduct [(6 – 4)PP]. Shown here are derivatives of two thymine bases linked via C6 of one thymine base and C4 of the other thymine base.
Figure 3.
Figure 4.
Schematic of various DNA damage-induced DNA repair pathways. A variety of DNA damaging agents can induce DNA damage, which becomes substrate for specific DNA repair pathways. Upper panel shows representative DNA damaging agents: errors from replication, spontaneous base deamination, alkylating agents, toxins, oxidative agents, ionizing radiation, UV radiation, crosslinking agents, aromatic compounds and environmental agents such as heat, cold and hypoxia. Middle panel represents different kinds of damaged DNA: base mismatches (C:T), uracil from deamination of cytosine, an abasic site from the loss of a base from one DNA strand, methylated guanine, methylated adenine, 8-oxo-G lesion, thymine glycols, single strand breaks, double strand breaks, intrastrand cyclobutane thymine dimers and interstrand guanine crosslinks. The lower panel lists the specific DNA repair pathways that are instigated to repair DNA damages: mismatch repair corrects replication errors and other base mismatches; base excision repair removes base adducts, uracil, abasic sites and oxidative lesions; single strand break repair pathways repairs single stranded breaks in the DNA backbone; double strand break repair pathway repair double strand breaks; nucleotide excision repair removes bulky lesions and intrastrand
crosslinks; interstrand crosslink repair removes interstrand linkages and translesion synthesis bypasses intrastrand crosslinks and bulky lesions.
Table 1

Shows representative human DNA polymerases. BER (base excision repair), MMR (mismatch repair), NER (nucleotide excision repair), DSBR (double strand break repair), dCTP (deoxycytidine phosphate), FA (fanconi anemia), TLS (translesion synthesis), SHM (somatic hypermutation), ICL (interstrand cross link), dRP (deoxyribosephosphate), TdT (Terminal deoxynucleotidyl transferase), RT (reverse transcriptase), AEP (archaeo-eukaryotic primases).

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>Family</th>
<th>Error Rate</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>α (POLA)</td>
<td>B</td>
<td>$10^{-4} - 10^{-5}$</td>
<td>An RNA primase during replication; role in S-phase checkpoint</td>
</tr>
<tr>
<td>β (POLB)</td>
<td>X</td>
<td>5 X $10^{-4}$</td>
<td>A dRP and AP lyase; role in BER</td>
</tr>
<tr>
<td>δ (POLD)</td>
<td>B</td>
<td>$10^{-5} - 10^{-6}$</td>
<td>Has a 3′-5′ exonuclease activity; role in replication; additional roles in BER, MMR, DSBR, NER</td>
</tr>
<tr>
<td>ε (POLE)</td>
<td>B</td>
<td>$10^{-6} - 10^{-7}$</td>
<td>Has a 3′-5′ exonuclease activity; role in replication; additional roles BER, MMR, DSBR, NER and S-phase checkpoint</td>
</tr>
<tr>
<td>REV1 (REV1)</td>
<td>Y</td>
<td>Incorporates only dCTPs</td>
<td>Incorporates only dCTPs and mediate protein-protein interactions during TLS; role in FA, HR</td>
</tr>
<tr>
<td>ζ (REV3)</td>
<td>B</td>
<td>$10^{-1}$</td>
<td>Roles in TLS, DSBR, FA and SHM</td>
</tr>
<tr>
<td>η (POLH)</td>
<td>Y</td>
<td>3.5 X $10^{-2}$</td>
<td>Roles in TLS, SHM, BER</td>
</tr>
<tr>
<td>ι (POLI)</td>
<td>Y</td>
<td>$10^{-1} - 10^{-2}$</td>
<td>Roles in TLS, SHM, BER</td>
</tr>
<tr>
<td>κ (POLK)</td>
<td>Y</td>
<td>$10^{-2} - 10^{-3}$</td>
<td>Roles in TLS and NER</td>
</tr>
<tr>
<td>θ (POLQ)</td>
<td>A</td>
<td>2.4 X $10^{-3}$</td>
<td>Has a helicase motif; ICL repair</td>
</tr>
<tr>
<td>γ (POLG)</td>
<td>A</td>
<td>$10^{-5}$</td>
<td>Has a 3′-5′ exonuclease activity; role in mitochondrial replication; BER</td>
</tr>
<tr>
<td>λ (POLL)</td>
<td>X</td>
<td>1.5 X $10^{-4}$</td>
<td>A dRP lyase; roles in V(D)J recombination, NHEJ and BER</td>
</tr>
<tr>
<td>μ (POLM)</td>
<td>X</td>
<td>$10^{-3} - 10^{-5}$</td>
<td>A terminal transferase; roles in V(D)J recombination and NHEJ</td>
</tr>
<tr>
<td>ν (POLN)</td>
<td>A</td>
<td>3.5 X $10^{-3}$</td>
<td>Possibly TLS</td>
</tr>
<tr>
<td>σ (POLS)</td>
<td>X</td>
<td>unknown</td>
<td>Has a 3′-5′ exonuclease activity; role in sister chromatid exchange</td>
</tr>
<tr>
<td>Tdt</td>
<td>X</td>
<td>unknown</td>
<td>Has a 3′-5′ exonuclease activity; role in sister chromatid exchange</td>
</tr>
<tr>
<td>Telomerase</td>
<td>RT</td>
<td>2 X $10^{-5}$</td>
<td>Replicates the ends of chromosomes</td>
</tr>
<tr>
<td>PrimPol</td>
<td>AEP</td>
<td>unknown</td>
<td>Translesion polymerase with high efficiency</td>
</tr>
</tbody>
</table>