



Evolutionary origins of DNA repair pathways: Role of oxygen catastrophe in the emergence of DNA glycosylases

Paulina Prorok, Inga R Grin, Bakhyt T Matkarimov, Alexander A Ishchenko, Jacques Laval, Dmitry O Zharkov, Murat Saparbaev

► To cite this version:

Paulina Prorok, Inga R Grin, Bakhyt T Matkarimov, Alexander A Ishchenko, Jacques Laval, et al.. Evolutionary origins of DNA repair pathways: Role of oxygen catastrophe in the emergence of DNA glycosylases. *Cells*, 2021, 10 (7), pp.1591. 10.3390/cells10071591 . hal-04289943v2

HAL Id: hal-04289943

<https://hal.science/hal-04289943v2>

Submitted on 16 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolutionary origins of DNA repair pathways: Role of oxygen catastrophe in the emergence of DNA glycosylases

Paulina Prorok¹, Inga R. Grin^{2,3}, Bakhyt T. Matkarimov⁴, Alexander A. Ishchenko⁵, Jacques Laval⁵, Dmitry O. Zharkov^{2,3,*} and Murat Saparbaev^{5,*}

- ¹ Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany; prorok@bio.tu-darmstadt.de
- ² SB RAS Institute of Chemical Biology and Fundamental Medicine, 8 Lavrentieva Ave., Novosibirsk 630090, Russia; grin.inga@gmail.com
- ³ Center for Advanced Biomedical Research, Department of Natural Sciences, Novosibirsk State University, 2 Pirogova St., Novosibirsk 630090, Russia; dzharkov@niboch.nsc.ru
- ⁴ National Laboratory Astana, Nazarbayev University, Nur-Sultan 010000, Kazakhstan; bmakarimov@nu.edu.kz
- ⁵ Groupe «Mechanisms of DNA Repair and Carcinogenesis», Equipe Labellisée LIGUE 2016, CNRS UMR9019, Université Paris-Saclay, Gustave Roussy Cancer Campus, F-94805 Villejuif Cedex, France; laual-j@orange.fr; Alexander.ISHCENKO@gustaveroussy.fr
- * Correspondence may be addressed. M.S.: tel. +33(1)42115404, email murat.saparbaev@gustaveroussy.fr D.O.Z.: tel. +7(383)3635187, email dzharkov@niboch.nsc.ru

Citation: Prorok, P.; Grin, I.; Matkarimov, B. T.; Ishchenko, A.A.; Laval, J.; Zharkov, D. O. and Saparbaev, M. Evolutionary origins of DNA repair pathways: Role of oxygen catastrophe in the emergence of DNA glycosylases. *Cells* **2021**, *10*, x. <https://doi.org/10.3390/xxxxx>

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: It was proposed that the last universal common ancestor (LUCA) evolved under high temperatures in an oxygen-free environment, similar to those found in deep-sea vents and on volcanic slopes. Therefore, spontaneous DNA decay such as base loss and cytosine deamination were the major factors affecting LUCA's genome integrity. Cosmic radiation due to weak Earth's magnetic field and alkylating metabolic radicals added to these threats. Here, we propose that ancient forms of life had only two distinct repair mechanisms: versatile apurinic/apyrimidinic (AP) endonucleases to cope with both AP sites and deaminated residues, and enzymes catalysing direct reversal of UV and alkylation damage. The absence of uracil–DNA *N*-glycosylases in some Archaea, together with the presence of an AP endonuclease that can cleave uracil-containing DNA, suggest that the AP endonuclease-initiated nucleotide incision repair (NIR) pathway evolved independently from glycosylase-mediated base excision repair. NIR may be a relic that appeared in an early thermophilic ancestor to counteract spontaneous DNA damage. We hypothesize that a rise in the oxygen level in the Earth's atmosphere ~2 Ga triggered the narrow specialization of AP endonucleases and DNA glycosylases to cope efficiently with a widened array of oxidative base damage and complex DNA lesions.

Keywords: DNA repair; DNA glycosylases; AP endonucleases; protein folds; structural homology.

1. Introduction

This theoretical essay was inspired by Miroslav Radman's works on the mechanisms of SOS response and mismatch repair in bacteria. In their seminal paper, Watson and Crick did not point out the key role of the double-stranded DNA structure for maintaining genome stability thanks to DNA repair. Indeed, to carry out its hereditary function DNA requires not only replication for self-reproduction and transcription for converting genetic information into RNA and then to proteins, but also repair, to fix eventual DNA damage that otherwise can induce mutations and block template reading. During the course of the evolution, cellular organisms conceived several distinct strategies to counteract mutagenesis and DNA damage to their double-stranded genomes, such as direct reversal (DR) (Figure 1A), DNA excision repair (Figure 1B–J), lesion tolerance/bypass, nucleotide pool sanitization, and recombination repair. It should

be stressed that the mechanistic principles used in each repair strategy are fundamentally different. In DR and nucleotide pool sanitization pathways, both of which do not require double-stranded structure, DNA damage is either removed or prevented without the need for excision of modified nucleotides and *de novo* DNA synthesis (Figure 1A). DNA damage tolerance strategy employs both template switching and lesion bypass and, similar to DR, does not require excision of damaged base. In contrast, DNA excision repair that includes pathways such as the base excision repair (BER) (G-H), nucleotide incision repair (NIR) (F), alternative excision repair (AER) (I), nucleotide excision repair (NER) (J), and mismatch repair (MMR), proceeds through phosphodiester bond cleavage, removal of DNA damage *via* excision, *de novo* DNA synthesis with the non-damaged strand as a template, and ligation. Recombination repair also relies on *de novo* synthesis but the template is provided by another copy of double-stranded DNA.

The very multiplicity of DNA repair pathways begs for a question how they, particularly excision repair now in charge of the elimination of deleterious DNA base lesions, could emerge very early in the evolution? Could the copious details known about the molecular mechanisms of repair pathways yield insights into hypothetical DNA repair mechanisms in the last universal common ancestor (LUCA) of all living organisms? Obviously, the genotoxic factors faced by the most ancient life forms depend both on the environment of the early Earth and on the metabolic reactions used by life then. The earliest commonly accepted geochemical evidence of life are dated at least ~3.77 Ga, whereas molecular evidence suggest a date >3.9 Ga, before the Late Heavy Bombardment events [1-3]. For a significant part of the early biotic evolution, life was likely exposed to hostile conditions such as high temperatures, extreme pH, anoxic UV-transparent atmosphere and weak Earth's magnetic field. The genetic evidence also suggest that the metabolism in LUCA was replete in carbon radical-based reactions [4,5]. Thus, several candidates for primeval genotoxic factors appear feasible: spontaneous DNA hydrolysis, short-wavelength UV (UVC) light, ionizing radiation and alkylating compounds. On the other hand, there is a general agreement that oxygen-rich atmosphere and widespread oxidative metabolism appeared much later in the history of Earth, thus making oxidative DNA damage of substantially less concern than it is today.

The main products of spontaneous DNA decay in water are apurinic/apyrimidinic (AP) sites, generated by the hydrolysis of *N*-glycosidic bonds, mainly in purine deoxynucleotides, and uracil (U) residues, arising from hydrolytic deamination of cytosines; hypoxanthine (Hx) and xanthine (Xan) residues are also produced by deamination of adenine and guanine, respectively, but with lower yields (Figure 2A). The main products of UV radiation are cyclobutane pyrimidine dimers (CPD) and pyrimidine-pyrimidone (6-4) photoproducts (6-4PP) (Figure 2B) [6]. Both photoproducts are cytotoxic (block DNA replication and transcription) and mutagenic. Notably, fossilized intertidal microbial mats with evidence of periodic atmosphere exposure suggest that the pathways to cope with UV damage were already in place as early as ~3.4 Ga [7]. DNA strand breaks with 3'-blocking groups, abasic sites and modified bases are major products of ionizing radiation [8,9]. Pyrimidine hydrates in DNA such as 6-hydroxy-5,6-dihydrothymine occur spontaneously and can be also generated by exposure to UV light and ionizing radiation under anoxic conditions (Figure 2A) [10-12]. Electrophilic agents, including carbon radicals, avidly react with electron-rich positions of nucleobases, forming several classes of their *N*- and *O*-alkylated derivatives [13,14]. Last but not least, the replication machinery is not a perfect copier and can make mistakes, misincorporating ribonucleotides or non-complementary dNMPs [15,16].

At least four major DNA repair pathways can efficiently deal with these abnormal chemical structures in DNA: photolyase-, alkyltransferase- and dioxygenase-catalyzed DR, DNA glycosylase-initiated BER, AP endonuclease-mediated NIR and multiprotein complex-dependent NER (Figure 1). DR of a damaged base to its native state without excision and *de novo* DNA synthesis provides the simplest, most efficient and accurate biochemical mechanism of DNA repair. However, DR mechanisms are known only for a handful of UV-induced and alkylated base lesions. NER is the main pathway used by present-day cells to remove a variety of structurally unrelated bulky DNA lesions such as those generated by UV radiation and xenobiotics including environmental toxins and anti-cancer drugs. Here DNA lesions are eliminated by a multiprotein

complex through dual incision bracketing the adduct in the damaged strand to form a short oligonucleotide containing the lesion and a single-stranded gap serving as a template to restore DNA integrity through the action of DNA polymerases and ligases.

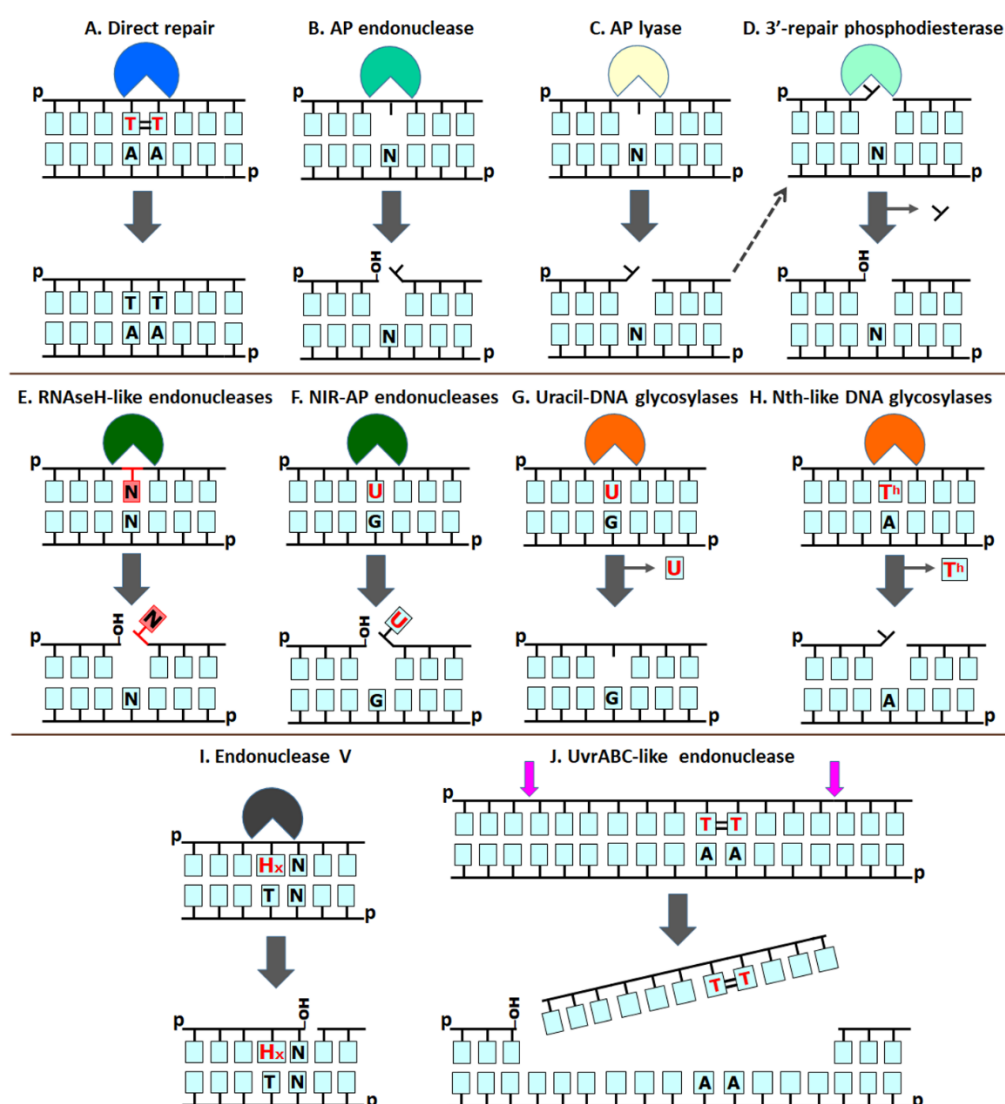


Figure 1. Principal DNA repair mechanisms employed by cellular organisms. N, deoxyribonucleotide; N with the red backbone, ribonucleotide; T^h, thymine hydrate; U, uracil; Hx, hypoxanthine; T=T, UV induced pyrimidine dimers.

The classic BER pathway is initiated by a DNA glycosylase excising a modified or mismatched base by hydrolysis of its *N*-glycosidic bond, generating a free base and an AP site, which is a cytotoxic and mutagenic repair intermediate (Figure 1G). Then either a hydrolytic AP endonuclease or an AP lyase activity present in certain DNA glycosylases nicks DNA 5' or 3' to the AP site, respectively (Figure 1B,C). The incised DNA is further processed by either 2'-deoxyribose-5'-phosphate lyase (dRpase) or 3'-cleaning repair activity (3'-phosphodiesterase or 3'-phosphatase), and then by a DNA polymerase, DNA ligase, and other accessory proteins to restore the integrity of DNA. The flow of the DNA glycosylase-initiated BER, which requires the sequential action of two enzymes for proper DNA incision, must be tightly coordinated because of the genotoxic intermediates generated. In the alternative DNA glycosylase-independent NIR pathway, an AP endonuclease directly cleaves damaged DNA 5' to various modified nucleotides, generating 3'-OH and 5'-phosphate termini thus avoiding potentially genotoxic AP site formation (Figure 1F) [17]. The NIR pathway, which can be genetically separated from BER, is evolutionary

conserved in all three domains of life [18,19]. BER and NIR are the most versatile pathways that recognize a wide range of DNA modifications from small, non-bulky ones to bulky lesions that distort the helical DNA structure.

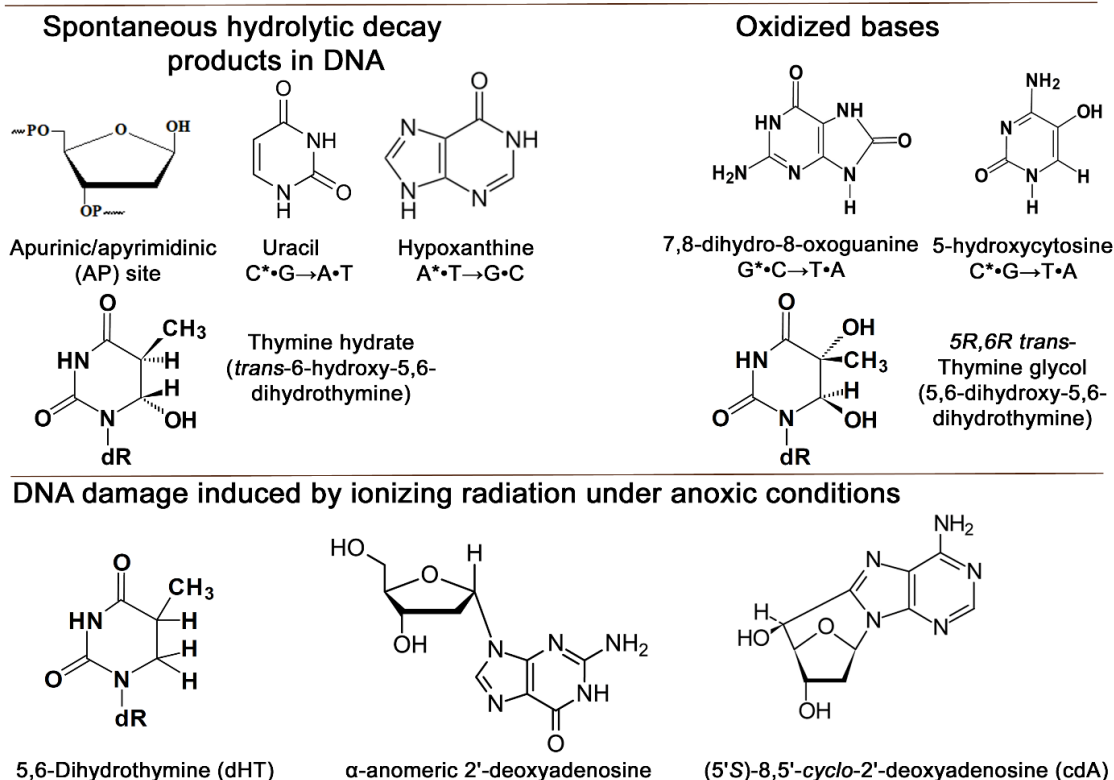
From the mechanistic point of view, DR and NIR pathways are the simplest, requiring few steps and energy, as compared to the BER and NER machineries, which employ multiple sequential steps and more than one protein to cleave damaged DNA and generate proper ends. Recent findings shed light on the evolutionary origins of DNA repair pathways and mechanisms of DNA damage recognition. In this scenario, following the Oxygen Catastrophe there was an increase in the spectrum and complexity of DNA damage, which overwhelmed the capacity of ancient DR and NIR system to maintain genome stability and resulted in the appearance of new pathways such as BER and NER. Here, we propose that the DNA glycosylase-initiated BER and multi-protein complex mediated NER systems appeared later in evolution, offering more versatile and efficient catalytic mechanisms that involve sequential actions of several repair proteins to remove DNA damage and generate a single-stranded gap of varying length with a proper 3'-end that can be used as a primer for DNA repair synthesis and ligation (Figure 1G,H,J). Here, we advance the idea that under the conditions on early Earth the spectrum of DNA damage was limited and only few singular repair pathways such as DR and NIR were sufficient to protect the genome of LUCA. We propose that LUCA employed a photolyase, and O⁶-alkylguanine alkyl transferase to repair UV and alkylation DNA damage, respectively, whereas spontaneous DNA decay and ionizing radiation induced DNA damage were counteracted by the NIR-specific AP endonucleases, AP lyases and 3'-repair phosphodiesterases (Figure 1B,C,F,F).

2. Types of DNA damage

Cellular macromolecules such as proteins, lipids and nucleic acids are prone to spontaneous decomposition because of their intrinsic chemical instability. In addition, endogenous oxidative stress and number of exogenous environmental factors, such as alkylating and oxidizing agents, photosensitizers, antibiotics, ionizing and UV radiation also contribute to structural damage to cellular components. Cellular DNA constantly undergoes structural alterations, resulting in DNA lesions that are cytotoxic, miscoding, or both and are believed to be at the origin of cell death and mutations [20,21]. DNA damage can be classified by its nature: spontaneous *vs* induced and by structure of the produced lesions: singular *vs* complex.

2.1. Hydrolytic DNA damage. As today, the ancestral life would encounter spontaneous DNA decay as a primary challenge to the genetic inheritance, further promoted by elevated temperatures. In aqueous solutions, sufficiently long DNA molecules are prone to spontaneous N-glycosidic bond hydrolysis and deamination at significant rates [22,23] (Figure 2A). Under normal physiological conditions, ~1 per 10⁵ purines are lost from DNA per day, generating AP sites [22]. AP sites have miscoding properties, since the replication machinery follows the "A rule", preferentially incorporating dAMP opposite them [24,25]. Spontaneous loss of the exocyclic amino groups in C, A and G bases generates highly mutagenic U, Hx and Xan residues, which induce C→T, A→G, and G→A transitions, respectively (Figure 2A) [26-28]. Under typical intracellular conditions, deamination in DNA occurs in about 1 per 10⁷ C residues per day, whereas purines are deaminated at ~1/10 of this rate [23]. Importantly, C deamination in single-stranded nucleic acids is an order of magnitude faster than in duplex DNA, and the need to distinguish an instructive genomic base from the damage product is believed to play a critical role in T substituting for U upon transition from the RNA world to the modern DNA world [29].

A. Singular DNA base damage



B. Complex DNA damage

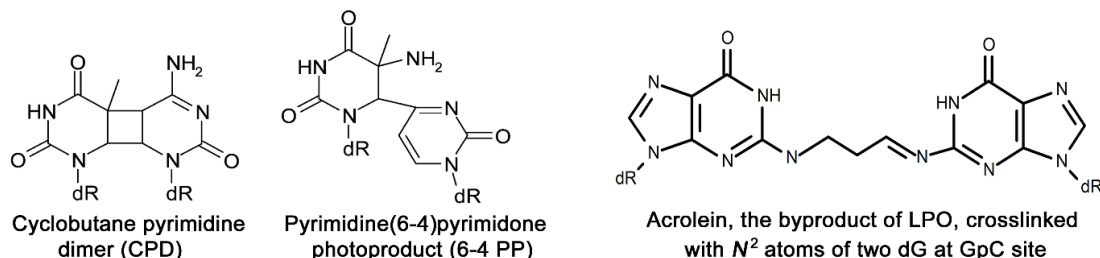


Figure 2. Chemical structures of spontaneous and induced DNA lesions. (A) Singular DNA base damage, including products of spontaneous DNA decay, oxidative DNA base damage and DNA base modifications generated by exposure to ionizing radiation under anoxic conditions. (B) Complex DNA damage, including bulky DNA adducts induced by UV light and inter-strand DNA crosslink generated by reactive aldehydes.

2.2. UV damage. Solar light is, and was, the primary source of UV radiation from the very beginning of life on Earth. The UV radiation spectrum can be divided into three bands: the least energetic long-wavelength UVA (315–400 nm), medium-wavelength UVB (280–315 nm), and the most energetic short-wave UVC (100–280 nm). The UV radiation, especially UVC, is very efficient in killing bacteria, and germicidal UVC-emitting lamps are frequently used for sterilization purposes. The most frequent and detrimental effect in DNA of UVC and UVB is the formation of pyrimidine dimers (Figure 2B), first discovered over sixty years ago by Beukers and Berends [30]. Two most common UV products are CPDs formed by cyclization of C5=C6 bonds of two adjacent pyrimidines, and 6–4PPs in which two adjacent pyrimidines are connected by a covalent bond between C6 and C4 (Figure 2B). 6–4PPs can be converted into Dewar valence isomers upon UVB irradiation [31]. CPDs are more frequent than 6–4PPs and constitute ~80% of the total UV-induced lesions [6]. In addition to bulky DNA adducts, UV radiation induces genotoxic small base modifications – pyrimidine hydrates such as 6-hydroxy-5,6-dihydrothymine, 6-hydroxy-5,6-dihydrocytosine and 6-hydroxy-5,6-dihydrouracil (Figure 2A) [32,33]. Appearance of oxygen

in the atmosphere ~2.4 Ga lead to the formation of the ozone layer, which absorbs radiation below ~310 nm, thus blocking all UVC and 95% of UVB from the surface of the Earth and offering an efficient protection for living organisms from excessive UV radiation [34,35].

2.3. Ionizing radiation damage. Ionizing radiation consists of subatomic particles or electromagnetic waves that are sufficiently energetic to tear electrons off atoms or molecules. Cosmic rays and the decay of radioactive isotopes are the primary sources of natural ionizing radiation on Earth contributing to background radiation. Since the Earth's core solidified only ~1.5 Ga, the early Earth's magnetic field was much weaker than it is today, resulting in ~100-fold higher solar high-energy flux [36,37]. Ionizing radiation induces high-energy radiolysis of water molecules producing highly reactive species such as hydrated electrons, $\cdot\text{H}$ and $\cdot\text{OH}$ radicals. The major products generated by ionizing radiation are DNA strand breaks with 3'-phosphoglycolate and 3'-phosphate termini, which arise by the abstraction of a hydrogen from deoxyribose at C1' or C4' by $\cdot\text{OH}$ radicals following base loss and sugar fragmentation [38]. A salient feature of ionizing radiation is the generation of closely spaced DNA lesions forming clusters, *i.e.* two or more damaged bases or strand breaks spaced by less than one helical turn of duplex DNA, either on the same strand or on the opposite DNA strands [8]. It should be stressed that the exposure to IR under low oxygen concentration favours the formation of 5,6-dihydrothymine (DHT), thymine hydrates, α -anomeric 2'-deoxynucleotides (αdN) and 8,5'-cyclopurine deoxynucleosides, but not the classic oxidized bases such as 7,8-dihydro-8-oxoguanine (8oxoG), 5-hydroxycytosine (OHC) or thymine glycol (Tg) (Figure 2A) [39-45]. This might suggest that such "anoxic" lesions were more physiologically relevant as damage generated by ionizing radiation in the cellular DNA on the early Earth.

2.4. Alkylation damage. Nitrosamines are formed under acidic pH by a reaction between nitrates or nitrites and secondary amines generated during protein degradation. Most of nitrosamines are potent mutagens and carcinogens. Metabolic activation of nitrosamines convert them to alkylating electrophilic compounds that react with DNA bases at all exocyclic oxygens and most ring nitrogens [46,47]. These agents can also alkylate non-bridging oxygen atoms of internucleoside phosphates, generating alkylphosphotriesters. $\text{S}_{\text{N}}1$ alkylating agents, such as *N*-methyl-*N*-nitrosourea (MNU) and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) use a monomolecular mechanism to alkylate at N and O atoms, while the $\text{S}_{\text{N}}2$ agents, which include methyl methanesulfonate (MMS) and dimethylsulfate (DMS), act through a bimolecular mechanism to alkylate at N atoms only. Ethyl methanesulfonate (EMS) uses both $\text{S}_{\text{N}}1$ and $\text{S}_{\text{N}}2$ mechanisms. 7-methylguanine (7meG), 3-methyladenine (3meA) and O^6 -methylguanine (O^6meG) are the most abundant alkylated lesions [48,49]. 1-Methyladenine (1meA) and 3-methylcytosine (3meC) are also generated by $\text{S}_{\text{N}}2$ agents, such as MMS, preferentially in single-stranded DNA and RNA [50]. The presence of 3meA, 1meA and 3meC residues in DNA template can block DNA replication, making them highly cytotoxic lesions, whereas 7meG is non-toxic and non-mutagenic [51,52]. $\text{S}_{\text{N}}1$ type agents such as MNNG and MNU are most potent chemical carcinogens because they react more readily with exocyclic oxygen of DNA bases to generate the highly miscoding O^6meG and O^4 -methylthymine (O^4meT) adducts [53]. The major adduct O^6meG mispairs with T during DNA replication, resulting in G→A transitions. Interestingly, mutagenic $\text{S}_{\text{N}}1$ nitroso-compounds can arise endogenously in bacteria through the nitrosation of amino acids, peptides or polyamines [54]. Endogenous methyl donor *S*-adenosylmethionine (SAM) can also act as a weak $\text{S}_{\text{N}}2$ type of alkylating agent *in vivo* and contribute to endogenous DNA alkylation [55]. Moreover, SAM is cleaved by a variety of enzymes (radical SAM enzymes) producing a 5'-deoxyadenosyl radical that is further used to generate other free radical metabolic species [56]. Radical SAM reactions likely constituted a significant part of LUCA metabolism [5]. Thus, exogenously and endogenously generated alkylating species would present a danger for the genetic material of ancestral life forms.

2.5. Oxidative damage. The appearance of oxygen-rich atmosphere resulted in the wide adoption of oxidative phosphorylation as the major pathway to produce adenosine triphosphate (ATP). In a canonical electron transport chain, free radicals are generated when an oxygen molecule promiscuously reacts with one of the transported electrons before it reaches the final com-

plex IV (cytochrome c oxidase). Superoxide anion radical ($\cdot\text{O}_2^-$) is generated, which is then converted to hydrogen peroxide (H_2O_2). Reduced transition metals such as Fe^{2+} and Cu^+ react with H_2O_2 to produce a highly reactive hydroxyl radical $\cdot\text{OH}$ via Fenton reactions [57], which in turn damages cellular DNA, proteins and membrane lipids [58]. About 80 different types of base and sugar lesions induced by these reactive oxygen species (ROS) have been identified [59]. ROS can damage both nucleobases and sugars in DNA either directly or indirectly. Hydroxyl radicals, which are the most active species, predominantly react at C8 of purines forming 8oxoG [60] and imidazole ring-opened products such as 2,6-diamino-5-formamidopyrimidine (FapyG) or 4,6-diamino-5-formamidopyrimidine (FapyA) [61], at the C5=C6 double bond of pyrimidines to form glycols [39,62], and at C8 and C5' of purine deoxynucleosides forming their 8,5'-cyclopurine derivatives [63]. Major oxidized bases due to endogenous ROS, such as 8oxoG, 5OHC, and 5-hydroxyuracil (5OHU) are miscoding and, if not repaired, lead to mutation upon replication [16,64,65]. Others, such as oxidized AP sites, thymine glycols and fragmentation products of DHT, cause replication block and therefore have a strong cytotoxic effect [66-69]. Indirectly, ROS can damage DNA through membrane lipid peroxidation (LPO) generating highly reactive α,β -unsaturated aldehydes (enals), e. g., acrolein, crotonaldehyde, and 4-hydroxynonenal (4-HNE), which can react with nucleobases forming exocyclic adducts such as 1, N^2 - γ -hydroxypropano-dG [70-72]. When present in DNA, these exocyclic adducts can further yield inter-strand DNA crosslinks (ICLs), often between guanines in CpG context, as a result of exocyclic ring opening and interaction with undamaged guanine in opposite DNA strand (Figure 2B) [73]. ICLs are highly lethal DNA lesions that block DNA replication, transcription and recombination by preventing strand separation. In addition to ICLs, chemical agents such as formaldehyde and physical factors, such as the ionizing and UV radiations, can crosslink proteins to an undisrupted DNA strand generating DNA-protein crosslinks (DPC) [74].

3. DNA repair in reconstructed and minimal genomes

The explosive growth of the number of fully or partially sequenced genomes and the appearance of tools to establish their evolutionary relationships led to several attempts to reconstruct both minimal genomes allowing modern cells to survive [75-81] and putative ancestral genomes [5,82-84]. Last Universal Common Ancestor is a hypothetical organism or, more precisely, a common gene pool from which all presently living organisms descend [85-87]. A recent reconstruction of the LUCA genome outlined 355 protein families potentially present in this pool [5]. Table 1 lists representative proteins from *E. coli* and humans that are related to DNA repair in today's species and have their putative ancestry among the LUCA genes.

Table 1. DNA repair protein-coding genes from reconstructed and minimal genomes.

Structural superfamily	Process	<i>E. coli</i>	human	Function ^a
LUCA genome reconstruction: Mirkin <i>et al.</i> [82] ^b				
O ⁶ -methylguanine-DNA-protein-cysteine methyltransferase	DR	Ada, Ogt, Atl1	MGMT	Exocyclic O-alkyltransferases
NUDIX hydrolases	nucleotide sanitization pool	MutT	MTH1	oxidized NTP/dNTP hydrolysis

Uracil–DNA glycosylase	BER	Ung, Mug	UNG, TDG, SMUG1	uracil–DNA glycosylase
Helix–hairpin–helix	BER	Nth, MutY, AlkA	NTHL1, OGG1, MUTYH, MBD4	DNA glycosylases for various lesions
TIM barrel	BER	Nfo	–	AP endonuclease
5'→3' exonucleases	various	PolA, Exo	FEN1	normal and damaged DNA degradation; flap endonucleases
Family A DNA polymerases; DnaQ-like 3'→5' exonucleases	repair DNA synthesis, proofreading	PolA	POLN, POLQ	DNA polymerases; proofreading activity may or may not be present
DEAD-box helicases	NER, recombination repair	UvrD, RecB	PARPBP1	monomeric DNA helicases involved in recombination, replication, and repair
ATP-binding cassette transporters	MMR	MutS	MSH2–MSH6	mismatch detection
S5-like fold; histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	MMR	MutL	MLH1, MLH3, PMS1, PMS2	mismatch signal transduction
P-loop NTPases	recombination repair; non-homologous end joining	SbcC	SMC3, RAD50	DSB recognition and/or processing
Calcineurin-like phosphoesterases	recombination repair; non-homologous end joining	SbcD	MRE11	DSB recognition and/or processing
Haloacid dehalogenases	recombination repair	RecJ	–	5'→3' exonucleases
P-loop NTPases	recombination repair	RecA	RAD51, DMC1	strand exchange
LUCA genome reconstruction: Weiss <i>et al.</i> [5]				
PDDEXK nucleases	specialized mismatch repair	Vsr	–	T:G mismatch cleavage

PDDEXK nucleases	specialized mismatch repair	–	–	archaeal mismatch endonucleases (NucS/EndoMS)
RNase H-like	AER	Nfi	ENDOV	deoxyinosine 3'-endonuclease
P-loop NTPases	recombination repair	RuvB	WRNIP1	hexameric DNA helicases driving branch migration during recombination
Family Y DNA polymerases	translesion synthesis	DinB, UmuC	POLH, POLI, POLK, REV1	DNA polymerases capable of translesion synthesis
DEAD-box helicases	various	RecQ	BLM, WRN	monomeric DNA helicases involved in recombination, replication, and repair
DEAD-box helicases	unknown, DNA damage-induced	Lhr	ASCC3	DNA damage-inducible DNA:RNA helicase
DEAD-box helicases	NER	RadD (YejH)	ERCC3	DNA helicase that functions in transcription and NER
DNA topoisomerases, subtype IA	possibly break repair and chromatin remodelling/maintenance	–	–	archaeal reverse gyrase; introduces positive supercoils into DNA
Radical_SAM	?	–	–	putative DNA modification/repair FeS-containing radical SAM protein
Minimal genome: Glass <i>et al.</i> [80] ^c				
Uracil–DNA glycosylase	BER	Ung	UNG	uracil–DNA glycosylase
TIM barrel	BER	Nfo	–	AP endonuclease
ATP-binding cassette transporters	NER	UvrA		NER lesion search
ATP-binding cassette transporters	NER	UvrB		NER lesion search
RNase H-like	NER	UvrC	–	NER nuclease

Minimal genome: Venetz <i>et al.</i> [81] ^d				
dUTPases	nucleotide pool sanitization	Dut	DUT	trimeric dUTP diphosphatase
Uracil–DNA glycosylase	BER	Ung	UNG	uracil–DNA glycosylase
Haloacid dehalogenases	recombination repair	RecJ	–	5'→3' exonucleases

^aGenes with an essential function in replication are not included except DNA polymerase I (PolA).

^bLUCA1.0 from [82].

^cMinimal genome reconstruction based on combined data from [75–78,80]. The experimental gene essentiality data were from *Mycoplasma genitalium*, *M. pneumoniae* and *Bacillus subtilis* genomes.

^dMinimized *Caulobacter crescentus* genome.

Several notes of caution are imperative when trying to use these reconstructions to understand the evolution of DNA repair systems. Minimal genomes are based on experimental data from a very limited number of bacterial species and thus necessarily reflect the kinds of genotoxic challenges faced by them under the laboratory growth conditions. Even if the gene inactivation experiments would be performed in another environment, or with other species, the gene sets identified as essential might be different. Reconstructed ancestral genomes, on the other hand, strongly depend on assumed evolutionary trees of the major phylogenetic groups included into the analysis, which are still far from being firmly established. The early evolutionary split between Bacteria and Archaea is especially obscure, hindering the analysis of gene ancestry and lateral transfer in the eukaryotic lineage [88–90]. For example, attempts to uncover eukaryotic genes of bacterial origin produced widely divergent results, at least concerning DNA repair genes [91,92]. It is also important to realize that the reconstructed LUCA genome reflects the metabolism and environment characteristic of the last genetic bottleneck, which very likely far post-dated both the appearance of life and its transition to the DNA world.

This said, and despite little congruence between essential DNA repair genes in minimal and reconstructed ancestral genomes evident from Table 1, these exercises may be useful to understand the biology of LUCA and how it coped with the genome damage. A recent reconstruction focused on LUCA's metabolism revealed an abundance of iron–sulfur (FeS) clusters, radical SAM reactions, and a variety of nucleobase modifications [5]. FeS clusters are important modules in many DNA repair enzymes from BER and NER pathways and in some DNA polymerases, often serving as at-a-distance sensors of oxidative DNA damage [93,94]. Radical SAM enzymes catalyse a diverse set of reactions and are based on the ancient TIM barrel fold, which also includes one major group of AP endonucleases [56,95,96]. Among other cofactors, LUCA appears to rely on flavin and pterin moieties, critical elements in UV-protecting DR. Finally, LUCA's genetic makeup is consistent with the hot environment, making hydrolytic reactions – base loss and deamination – possibly the strongest threat for its genome.

4. Spontaneous DNA damage: To BER or not to BER?

4.1. DNA glycosylases: primordial players or a later adaptation? Today, deaminated bases (mostly U) are removed by one of multiple uracil–DNA glycosylases, whereas AP sites, either spontaneous or glycosylase-formed, are nicked by an AP endonuclease (Figure 1G and B). However, the appearance of toxic and mutagenic AP sites in the course of canonical BER raises a question why such a potentially hazardous pathway evolved at all. A nuclease activity recognizing a lesion and cleaving DNA next to it (Figure 1E,F and see next sections) would not be entailed with this drawback. However, uracil–DNA glycosylase is one of few enzymes that

both apparently existed in LUCA and are essential in minimal genomes (Table 1). Did ancestral DNA repair pathways require DNA glycosylases?

Notably, at least seven protein folds can catalyze base excision from DNA: the uracil–DNA glycosylase-like α/β -fold, helix–hairpin–helix motif (HhH)-containing enzymes, helix–two–turn–helix motif (H2tH)-containing enzymes, the formyltransferase C-terminal domain fold, the T4 pyrimidine dimer DNA glycosylase fold, the tetratricopeptide-like repeats fold, and the HALFPIPE fold [97,98]. Even more strikingly, none of the known DNA glycosylases are based on structural folds of ubiquitous metabolic enzymes capable of hydrolyzing *N*-glycosidic bonds between sugars and aromatic heterocycles (including nucleobases), such as nucleoside hydrolases, purine/uridine nucleoside phosphorylases, or *N*-(deoxy)ribosyltransferases, or more specialized SMF/DprA-LOG superfamily that includes enzymes producing heterocyclic signaling molecules, or ribosome-inactivating RNA *N*-glycohydrolases. One is tempted to speculate that the weaker nature of the *N*-glycosidic bond in DNA compared with that in RNA makes its hydrolysis so inherently easy that many DNA-binding proteins could evolve an ability to catalyze this reaction with biologically acceptable rates. It is interesting to note that almost all DNA glycosylases and several other structurally unrelated classes of proteins (e. g., AP endonucleases, nucleobase dioxygenases, DNA methyltransferases, and epigenetic readers) kink DNA and flip out the target nucleotide [99,100], and it has been suggested based on molecular dynamics modeling that the configuration of DNA phosphates in this strained intermediate mainly provides the catalytic power for the *N*-glycosidic bond breakage [101]. An intriguing corollary of the differential *N*-glycosidic bond stability was noted by Rios and Tor [102] who hypothesized that the RNA world used an extended repertoire of nucleobases due to the need to combine information storage and catalysis in one type of nucleic acid, and the primeval glycosylases might have evolved with the appearance of DNA genomes to eliminate non-standard bases no longer needed, including U, from the genome carrier. This suggestion aligns both with the large repertoire of nucleobase modifications in LUCA [5] and with the ubiquitousness of UNG-like enzymes, which seem to represent an ancient uracil-binding fold [103,104]. Interestingly, the widest phylogenetic diversity today can be found not among Family 1 uracil–DNA glycosylases where extensively characterized human and *E. coli* enzymes belong but among Family 4 and Family 5 enzymes that were initially discovered in extremophilic archaea and bacteria and contain a FeS cluster [105–108].

In total, eight families of UNG-like enzymes have been defined so far, of which six are biochemically characterized. Some of them, such as Family 1 where extensively studied human and *E. coli* enzymes belong, and Family 4 initially discovered in extremophilic archaea and bacteria and containing a FeS cluster, seem to be bona fide uracil–DNA glycosylases with little activity on other substrates. Others, like SMUG1 (Family 3), FeS-containing Family 5, and the recently discovered SMUG2 and *Bradyrhizobium diazoefficiens* uracil–DNA glycosylase (BdiUng)-like enzymes have wider substrate specificity that may additionally include other U derivatives (5OHU, 5-hydroxymethyluracil, 5-formyluracil), Hx, and Xan [109–112]. Thymine–DNA glycosylase (TDG) present in animals and fungi is involved in active epigenetic demethylation removing oxidized and/or deaminated derivatives of 5-methylcytosine [113,114], and together with its bacterial homolog Mug may be the primary glycosylase for exocyclic pyrimidine adducts such as 3,*N*⁴-methylcytosine, or for 7,8-dihydro-8-oxoadenine (8oxoA) [115,116]. Finally, Family 6 lacks uracil-removing activity completely, excising hypoxanthine instead [107]. The most versatile group of DNA glycosylases is the HhH superfamily, the members of which are present in all domains of life, remove all major kinds of damaged bases and some mismatched bases, and can be monofunctional or have an associated β -elimination activity (Figure 1H). The HhH motif itself is a small non-specific DNA-binding element also found in DNA and RNA polymerases, transposases, Holliday junction branch migration protein RuvA, RecD helicase and several other DNA-binding proteins [117–119]. In DNA glycosylases with an associated β -elimination activity (Nth, OGG1, OGG2), the second helix bears a catalytic Lys residue

that acts as a nucleophile to displace the target base. The homologous Lys in DNA polymerases β and λ serves as a nucleophile in the dRpase reaction catalyzed by these enzymes [120,121]. In monofunctional enzymes (MutY, AlkA, MBD4) this Lys is replaced with different non-basic residues, indicating that the ability to perform β -elimination was likely ancestral in the glycosylases belonging to this superfamily. In DNA glycosylases, the HhH motif is followed by a Gly/Pro-rich stretch bearing a second catalytic residue, an Asp side chain (GPD motif) [122,123], absent from the non-glycosylase members of the superfamily. Some of HhH DNA glycosylases also contain a FeS cluster. Thus, HhH DNA glycosylases likely arose as a result of evolution of an ancient universal DNA-binding HhH motif.

The last major group of DNA glycosylases is built on another small DNA-binding module, the H2tH motif. Unlike the HhH motif, it is found only in a limited group of polypeptides, including ribosomal protein S13 (S18 in eukaryotes) and several ribosome quality control proteins. These glycosylases also carry a DNA-binding β -ribbon zinc finger or a structurally equivalent β_2 -module called a “zincless finger”, and a unique N-terminal β -sandwich domain with the catalytic N-terminal amino group and a catalytic acidic side chain [124,125]. Members of this group are missing from Archaea, and, as they are exclusively involved in oxidative damage repair, the H2tH glycosylases represent good candidates for acquisition from symbiotic mitochondria functionally associated with the evolution of oxidative metabolism.

4.2. AP endonucleases: the archetypal repair endonucleases. The occurrence of an enzyme recognising AP sites in *E. coli* was first described in the seminal work of Verly [126]. AP endonucleases cleave DNA 5' to an AP site by a hydrolytic mechanism [127,128]. Majority of known hydrolytic AP endonucleases are divided into two distinct families based on amino acid sequence identity to either *E. coli* exonuclease III (Xth) or endonuclease IV (Nfo) [129]. Mg^{2+} -dependent Xth-like AP endonucleases belong to a large endonuclease/exonuclease/phosphatase (EEP) superfamily that also includes DNase I, tyrosyl-DNA phosphodiesterase 2, retrotransposon endonucleases and various phosphoester hydrolysis enzymes [130–134]. Nfo-like AP endonucleases are totally different in their structure, belonging to the multifunctional TIM barrel superfamily, which probably originally contained sugar metabolism enzymes [95,135], and utilize Zn^{2+} ions for catalysis. Despite these differences, the geometry of the active sites in Xth- and Nfo-like AP endonucleases is very similar, underlying a conserved catalytic mechanism [136]. Genetic data indicate that AP endonuclease-deficient bacterial, yeast and mammalian mutant cells are highly sensitive to alkylating agents, oxidative stress and ionizing radiation, indicating critical role of these enzymes in protecting cells from AP sites and DNA strand breaks with 3'-blocking groups [137,138]. Furthermore, complete lack of AP site repair enzymes in *S. cerevisiae* is lethal, indicating that the background level of AP sites is sufficient to cause cell death, at least in yeast [139,140]. Interestingly, an Nfo-like AP endonuclease is present in a minimal *Mycoplasma* genome and in one of LUCA genome reconstructions (Table 1). Taken together, these observations suggest that LUCA contained an AP endonuclease (possibly Nfo-like) to counteract the effects of spontaneous base loss in its genome. However, as we argue below, AP sites might not be its only, and even not the primary substrate.

4.3. Archaeal AP endonuclease Mth212 in the repair of deaminated DNA bases. C→T transitions, likely arising through cytosine deamination, are the most frequently occurring base substitutions observed in living organisms [141–143]. In addition, uracil in DNA can be generated through incorporation of dUMP instead of dTMP from the nucleotide pool; this process results in U:A pairs that are not mutagenic but can be lethal when accumulated at a high level in genomic DNA [144,145]. Indeed, as dU(M/D/T)P pool is a metabolic precursor of thymine nucleotides, the threat of dUMP incorporation is unavoidable, and the need to counteract the buildup of genomic U may be the main reason for the universal persistence of UNG across all major domains of life including some viruses.

Surprisingly, no homologues of the major UNG-like families could be detected in the complete genome sequences coming from several orders of Euryarchaeota such as Methanopyrales [146], Methanococcales (*Methanocaldococcus jannaschii* [147], *Methanococcus maripaludis* [148]), two species of Methanobacteriales (*Methanothermobacter thermautotrophicus* [149] and *Methano-*

sphaera stadtmanae [150]), suggesting the existence of alternative mechanisms for removal of deaminated DNA bases. Analysis of protein extracts from *M. thermautotrophicus* revealed a nicking activity towards U in ssDNA and dsDNA [18]. The purified protein cleaved the 5'-phosphodiester bond next to dU in a Mg^{2+} -dependent manner, and was identified as Mth212, a distinct member of the EEP family. Mth212 cleaves duplex DNA containing AP site or dU paired with any base but not DNA with mismatched natural bases. Additionally, the U-endonuclease activity was followed by 3'→5' exonuclease in the lesion vicinity which extended the nick to a single-stranded gap upstream of the lesion [151]. The repair of dU in dsDNA was successfully reconstituted *in vitro* with purified enzymes. In the first step Mth212 cleaves duplex next to the lesion and generates the free 3'-OH end extended by DNA polymerase B; the displaced DNA strand is cleaved by 5'-flap endonuclease producing a nick, which is then sealed by DNA ligase to accomplish the repair [152].

The crystal structure of Mth212 exhibits no apparent homology to UNG-like DNA glycosylases but is very similar to Xth-like AP endonucleases. Attempt to obtain co-crystal structure of Mth212 and dU-containing DNA yielded several abortive enzyme/DNA complexes in which either one duplex oligonucleotide was bound at both of its ends by two wild-type Mth212 molecules, or two duplex oligonucleotides were bound to one molecule of catalytically inactive Mth212 D151N mutant forming a pseudo-continuous DNA duplex [153]. In the latter structure, one oligonucleotide was bound in the same way as with wild-type Mth212, and the second oligonucleotide extended the first one at the enzyme's active site resulting in a joint DNA molecule (Figure 3A). Nevertheless, these aberrant Mth212/DNA complexes revealed the mechanism of dU recognition in duplex DNA surprisingly similar to DNA glycosylases. Mth212 inserts Arg209 amino acid into the DNA base stack and stabilizes extra-helical flipped-out conformation of dU in specific active site pocket. The specific recognition of the U base is attained by interactions with Lys125, Ser171 and Gln153. Modelling of dU and dC in active site of Mth212 showed that these amino acids can discriminate between damaged and regular DNA base. Thus, nucleotide flipping is a common mechanism used by NIR-specific Xth family AP endonucleases and uracil-DNA glycosylases.

Further characterization of the Mth212 substrate specificity demonstrated that the archaeal repair enzyme, similar to NIR-proficient AP endonucleases such as *E. coli* Nfo and human APE1, is able to cleave at others lesions such as α dA and 5OHC in duplex DNA. Furthermore, under low Mg^{2+} concentrations (≤ 1 mM) APE1 also can cleave duplex DNA 5' to U opposite G and then extend the resulting nick to a single-stranded gap by its non-specific 3'→5'-exonuclease activity [154]. Thus, the direct AP endonuclease-catalyzed DNA incision 5' to a damaged nucleotide (the NIR pathway) is evolutionary conserved in Xth-like AP endonucleases from Archaea to humans, and can serve as a back-up – or sometimes the only – mechanism to repair genomic dU residues.

4.4. Archaeal endonuclease Q in the repair of deaminated DNA bases. Another study by Ishino and colleagues showed that cell extracts of hyperthermophilic archaeon *Pyrococcus furiosus* contain a new DNA damage-specific endonuclease, endonuclease Q (EndoQ), which cleaves phosphodiester bond in DNA at the 5' side of Hx, U, Xan and AP sites [155]. The repertoire of *P. furiosus* EndoQ (PfuEndoQ) substrates was later extended to include oxidized pyrimidines such as 5,6-dihydrouracil, 5-hydroxyuracil and 5-hydroxycytosine residues, reminiscent of *E. coli* Nfo and human APE1 [156]. The activity of EndoQ is affected by the nature of the base opposite to the lesion [156].

Enzymes belonging to the EndoQ family, not related to other endonucleases, have been identified mainly in the species belonging to Euryarchaeota, such as the extremely hyperthermophilic *Thermococcus kodakarensis*, growing optimally at temperatures over 100°C, and a mesophilic methanogene *Methanosarcina acetivorans*, which possesses three enzymes for the removal of U from DNA: EndoQ, Xth and Family 4 UNG [157].

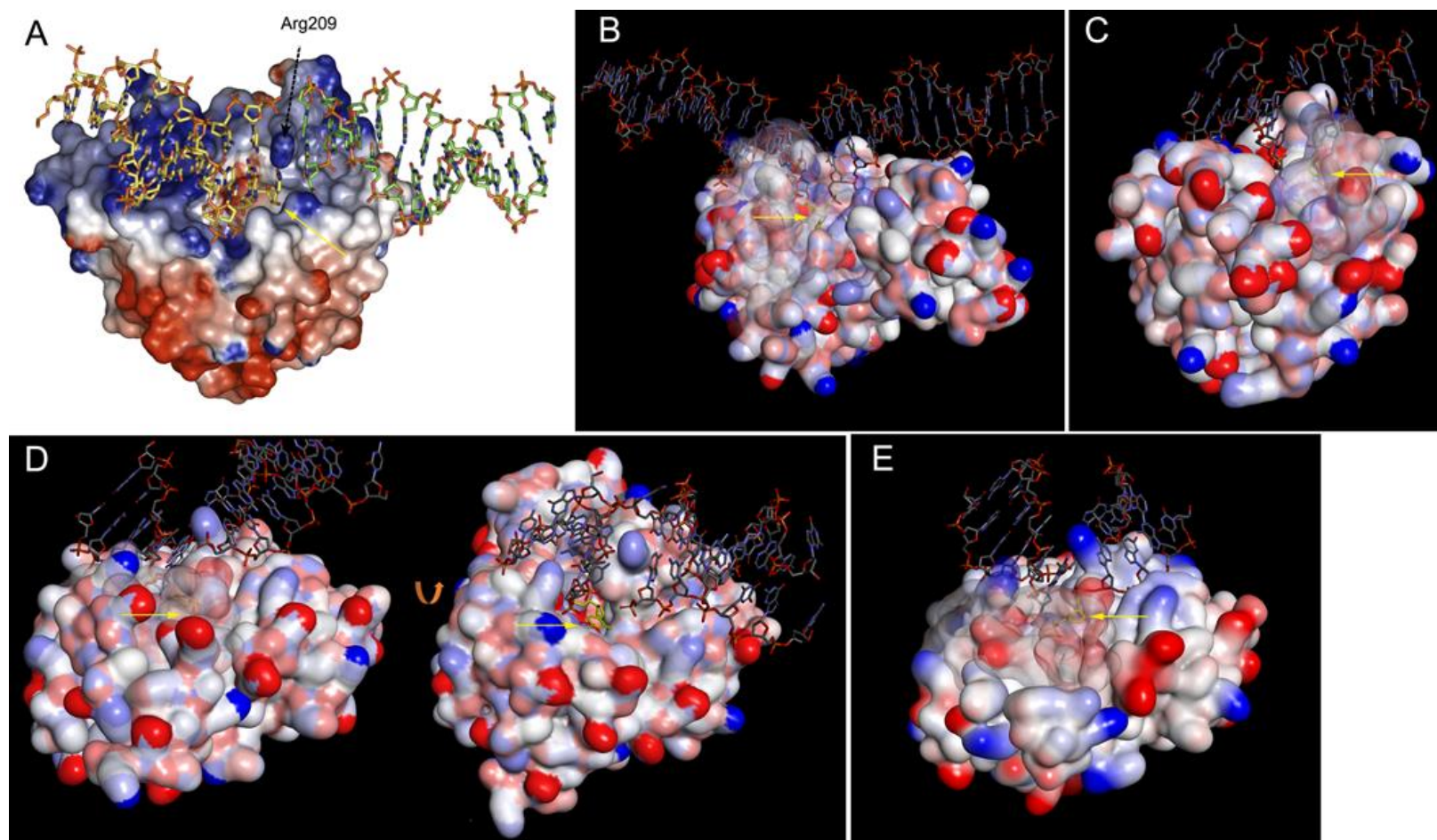


Figure 3. Nucleotide flipping out of DNA helix is the conserved DNA base damage recognition mechanism among NIR AP endonucleases. (A) Enzyme–substrate complex of the Mth212 D151N mutant obtained with dsDNA (exonuclease reaction substrate with a terminal dC in the active site) (PDB: 3GA6) [153]. (B) PfuEndoQ complex with a 27 mer duplex substrate with dU at the active site (PDB: 7K30) [160]. (C) Human APE1 bound to non-cleaved THF-DNA (PDB: 1DE8) [132]. (D) *E. coli* Nfo H69A mutant bound to a cleaved DNA duplex containing α DA (PDB: 4K1G) [165]. (E) *E. coli* Nfo Y72A mutant bound to THF-DNA (PDB: 2NQ9) [166]. Yellow arrows point to the nucleotides flipped into the active sites.

475
476
477
478
479
480
481

Interestingly, it has been shown that EndoQ interacts with PCNA through its PIP-motif, which stimulates the endonucleolytic activity of the repair enzyme [158]. Direct interaction with PCNA, as well as activity towards duplex and ssDNA suggest that EndoQ can be a part of both DNA repair machinery and DNA replication complex. More extensive phylogenetic studies permitted to identify members of EndoQ family in some bacteria (*Bacillus pumilus* and *B. subtilis*), where the enzymes may participate in antiviral defense mechanism beside their role in DNA repair [159]. The activity of the *B. subtilis* EndoQ is not inhibited by Ugi, the specific UNG inhibitor from PBS1 phage [156].

Recent X-ray crystallography studies provided insights into the molecular mechanism of EndoQ catalyzed endonucleolytic cleavage, showing that the damaged base is flipped out into an extrahelical position and placed in the catalytic pocket of the enzyme adjacent to a Zn^{2+} -binding site [160]. The enzyme binds three Zn^{2+} ions; two of them reside in the PHP domain and are directly involved in the phosphodiester bond hydrolysis, while the third Zn^{2+} is found in a separate Zn-binding domain [160]. More recently, the crystal structures of PfuEndoQ bound to dU-, dI-, and AP site-containing DNA duplexes were solved (Figure 3B) [161]. The highest-resolution structure of PfuEndoQ with the U:G substrate revealed an active site with a single Zn^{2+} ion coordinated by Glu76, His84, His139, and the oxygen atom of the scissile phosphodiester bond [161]. The E76A, H84A, and H139A single amino acid substitutions completely abolish the endonucleolytic activity of EndoQ, suggesting that this Zn^{2+} is essential for the phosphodiester bond cleavage [160]. EndoQ binding induces a highly distorted conformation of the damaged DNA strand up- and downstream of the lesion. Three structural domains of EndoQ form a positively charged cleft that accommodates the sharply bent DNA duplex. The damaged nucleotide is at the apex of the kink, rotated out of the duplex and inserted into a deep active site pocket formed by Zn-binding and catalytic domains of the protein (Figure 3B). Interestingly, U and Hx bases are specifically recognized by different set of amino acid residues, and this mode of recognition is distinct from that of DNA glycosylases. Noteworthy, the His139, Gly169 and Ser171 residues in the catalytic pocket of EndoQ present a steric obstacle for the binding of regular T and A bases because of the presence of a C⁵-methyl group and an N⁶-amino group absent from U and Hx, respectively.

Noteworthy, the nucleotide flipping accompanied with specific interactions in the active site pocket are universally employed by EEP superfamily AP endonucleases, TIM superfamily AP endonucleases, and archaeal EndoQ. Unlike DNA glycosylases and sequence-specific DNA methyltransferases, which have to flip out the target nucleotide in order to get access to the reacting atoms, endonucleases, in principle, could reach the scissile bond without nucleotide eversion. Indeed, structures of substrate and product complexes of DNase I, a non-specific endonuclease also belonging to the EEP superfamily, show no signs of nucleotide eversion [162,163]. Furthermore, closer inspection of the available Nfo/DNA and APE1/DNA structures revealed that these NIR AP endonucleases also utilize the nucleotide flipping mechanism to recognize both AP sites and 3'-terminal nucleotide residues at the DNA strand break (Figure 3C-E). Structural and mutational analysis of EEP superfamily NIR-competent AP endonucleases revealed that amino acid residues in their active site pocket do not interact with damaged DNA bases and instead recognize DNA damage via changes in the sugar puckering and in the distance between two neighboring phosphates in the DNA backbone [164].

Moreover, some residues located outside of the active site (e. g., Thr268 of APE1) can be involved in additional DNA contacts resulting in a dramatic effect on the NIR activity and Mg^{2+} dependence [165]. Notably, NIR-deficient *E. coli* Xth from the EEP superfamily is a very efficient AP endonuclease and has a more compact active site in comparison to NIR-proficient homologs (APE1, Mth212 and *B. subtilis* ExoA) [164]. Taken together, these observations suggest that nucleotide flipping is a common mechanism of lesion recognition employed by both

DNA glycosylases and AP endonucleases. This may be a perfect case when we can trace convergent evolution of various protein folds to exploit a common physicochemical feature – the inherent thermodynamic instability of damaged DNA.

5. Nucleotide incision repair: an ancient and versatile mechanism to counteract spontaneous DNA decay and ionizing radiation induced DNA damage

5.1. NIR as a remnant of the pre-BER world. Generally, it is agreed that the environment of early Earth around the time of LUCA existence was characterized by high temperatures and reducing anoxic atmosphere. Under this harsh conditions the spectra of DNA damage consisted mainly the products of base loss and deamination, UV light-, ionizing radiation- and alkylation agents-induced DNA lesions with no or very little oxidative DNA damage. The efficiency of archaeal Mth212 and EndoQ-like endonucleases in the repair of deaminated nucleobases without the involvement of DNA glycosylases may suggest that in early evolution, to counteract the deleterious effects of spontaneous DNA decay, the primordial organisms were endowed with universal endonucleases that dealt with both AP sites and U residues in a single step, as in the present-day NIR pathway (Figure 4). Nevertheless, today's NIR-competent AP endonucleases remove AP sites more efficiently than modified bases. We propose that latter in course of evolution, due to a dramatic change in the spectra of DNA damage and for the sake of efficiency, uracil–DNA glycosylases substituted for endonucleases as the primary dU-recognizing enzymes, while the endonucleases evolved towards more stringent specificity for AP sites. Metal cofactors could had become the key limiting resource for endonucleases as life moved from the first poorly compartmentalized, easily perfused forms to tight self-contained cells; interestingly, EEP enzymes show excellent performance with many divalent metal ions (Mn^{2+} , Co^{2+} , Cd^{2+} , Fe^{2+}) but use catalytically less preferable Mg^{2+} due to easier availability [167]. With the separation of functions between DNA glycosylases and AP endonucleases, BER as we know it finally appeared.

Another testimony for the likely existence of universal endonucleases in LUCA is the remarkable specificity of today's NIR for radiation-induced anoxic DNA base damage. The Earth's magnetic field protects life from ionizing radiation coming in form of solar wind and cosmic rays. Measurements from 3.3–4.2 Ga old minerals point to the weak magnetic field about $1/10^{th}$ of that at present [36,37]. Consequently, life on the ancient Earth was exposed to increased levels of ionizing radiation. As mentioned above, the spectra of radiation DNA damage drastically depend on the amount of oxygen dissolved in water. Under anoxic conditions, ionizing radiation generated preferentially DHT, DHU and α dN in DNA [42,44], which are preferred substrates for the NIR AP endonucleases such as archaeal Mth212, *E. coli* Nfo and human APE1 [17,154,168,169]. It should be noted that α dN in DNA cannot be recognized by DNA glycosylases, which require the presence of deoxynucleotides in β -anomeric configuration in order to flip them out from the DNA helix. Based on these observations, it is tempting to speculate that the AP endonuclease-initiated NIR pathway appeared very early in evolution as an adaptation to handle both spontaneous and induced DNA damage generated under anoxic conditions.

5.2. Specialized NIR: UVDE-initiated repair of UV photoproducts and RNaseH2-catalyzed ribonucleotide excision repair. Oxygen-free atmosphere of ancient Earth offered much less protection from short-wavelength UVC radiation. Therefore, ancient forms of life that lived in surficial environments such as shallow ponds of condensed geothermal vapor were constantly exposed to UV light, which induces highly genotoxic CPD and 6-4PP adducts in cellular DNA. Although they are usually repaired by DR and NER pathways, strikingly, these lesions are also substrates for endonucleases. Genetic and biochemical studies identified the UV damage endonuclease (UVDE) that recognizes and incises duplex DNA 5' to CPD and 6-4PP [170,171]. Similar to NIR AP endonucleases, UVDE can cleave DNA duplex containing AP sites, U, DHU and

cleans single strand break termini from 3'-blocking groups to generate 3'-OH end for DNA repair synthesis [172,173]. This UV damage endonuclease is not universally conserved and is found only in fungi (e. g., *Schizosaccharomyces pombe* and *Neurospora crassa*), Bacteria (though not in *E. coli*) and in a few archaeal lineages (Sulfolobaceae, Methanomicrobia and Halobacteria). The crystal structure of UVDE protein from the thermophilic bacterium *Thermus thermophilus* showed essential structural features of the TIM barrel fold very similar to that of Nfo [135,174], suggesting that UVDE probably evolved from the Nfo family AP endonucleases. Later, the crystal structure of archaeal UVDE from *Sulfolobus acidocaldarius* (SacUVDE) in a pre-catalytic complex with DNA duplex containing a 6-4PP has been resolved [175]. It revealed a novel dual dinucleotide flip mechanism for recognition of bulky crosslinked dipyrimidines. SacUVDE flips the two purines opposite to the damaged pyrimidine bases into a dipurine-specific pocket, whereas the damaged pyrimidines are also flipped into another cleft [175]. Based on these observations, we propose that earlier in evolution the ancient life forms did not possess sophisticated repair systems such as NER machinery but rather employed either photolyases to remove UV-DNA adducts or NIR-like endonucleases to counteract spontaneous, radiation- and UV-induced damage.

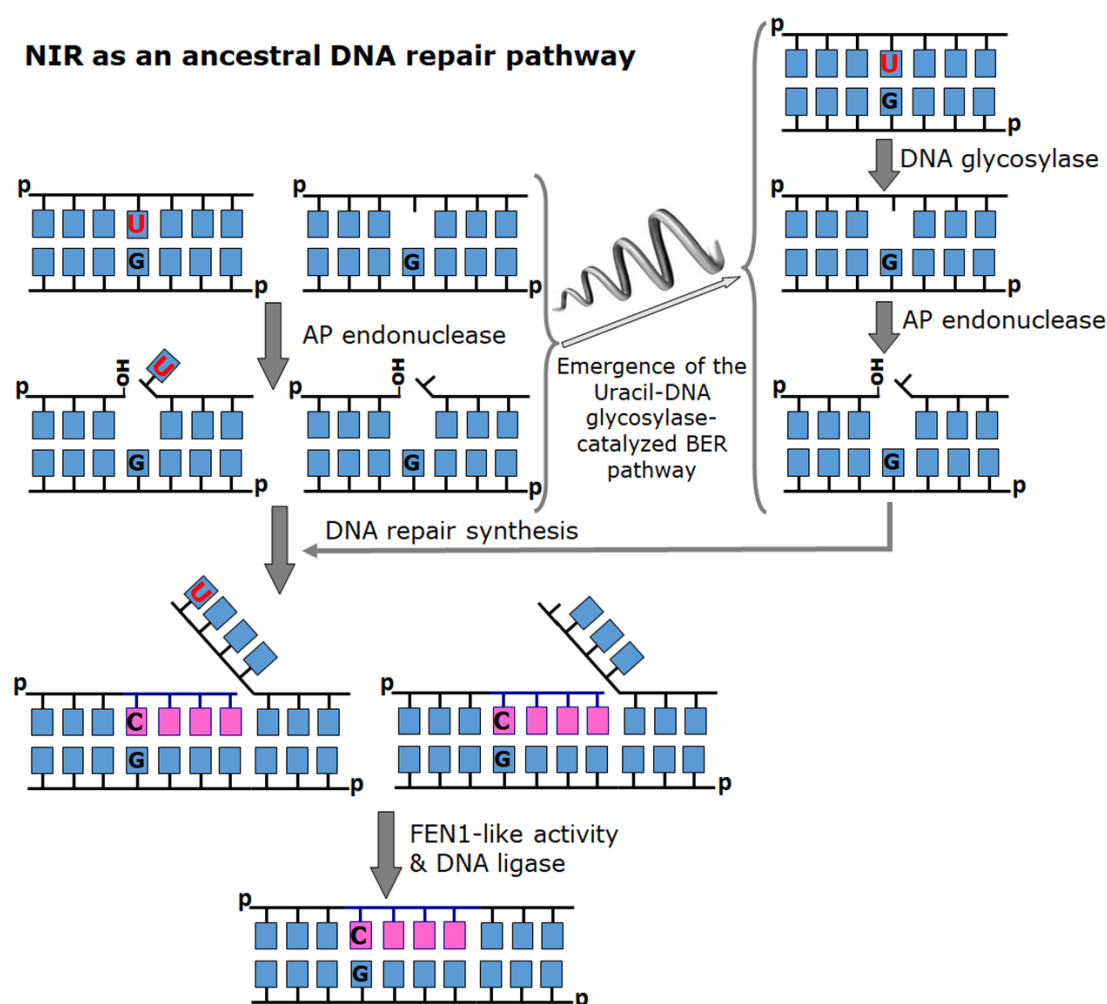


Figure 4. Putative evolutionary origin of the NIR and BER pathways for spontaneous DNA damage. In LUCA the products of spontaneous DNA decay were repaired by AP endonucleases in the NIR pathway in DNA glycosylase-independent manner. Mono-functional DNA glycosylases specific for uracil and hypoxanthine appeared later in evolution. Inclusion of additional base excision step into NIR pathway resulted in the appearance of modern BER pathway in three domains of life.

Ribonucleoside monophosphates (rNMPs) can be erroneously incorporated in cellular DNA during DNA replication [176,177], and this problem was likely even more acute shortly after the transition from the RNA to DNA world. These misincorporated ribonucleotides are eliminated via RNase H2-catalyzed cleavage on the 5' side of a ribonucleotide in DNA [178,179] to initiate a NIR-like pathway referred as ribonucleotide excision repair (RER) [180]. Indeed, the mechanism of action of RNase H2 towards ribonucleotides in DNA is very similar to that of NIR AP endonucleases towards their substrates (Figure 1E and F). RNase H fold is shared by many endo- and exonucleases, including those participating in DNA repair, such as *E. coli* endonuclease V (see below), RuvC resolvase, UvrC NER endonuclease, and proofreading domains of A and B family DNA polymerases [181,182]. Based on this mechanistic resemblance we propose that RNase H2-mediated DNA incision is an ancient NIR pathway evolved to repair distinct set of endogenous DNA damage occurred in LUCA. Importantly, RNases H2 are present in all three domains of life and frequently found together with type 1 RNases H, which have the same fold and mechanism but cannot cleave single rNMPs embedded in DNA [182]. Noteworthy, many Archaea contain only type 2 RNases H [183], furthermore, catalytic subunits of Archaeal and eukaryotic RNaseH2s have higher similarity to each other, both structurally and biochemically, compared with their bacterial counterparts [182].

6. Alternative excision repair pathway: an unusual variation on the endonuclease theme

Endonuclease V (Nfi/EndoV) is another enzyme performing the cleavage of lesion-containing DNA. Nfi was originally isolated from *E. coli* in 1977 [184] and characterized as a Hx-specific enzyme in 1994 [185]. Nfi homologs are present in all domains of life and show activity towards DNA and RNA substrates. In contrast to Xth, Nfo and EndoQ endonucleases, the Nfi protein cleaves the second phosphodiester bond 3' to a deaminated base. Further steps of the Nfi-mediated repair remain unclear, since extension of the 3'-OH terminus will not lead to damage removal. The purified *E. coli* Nfi protein exhibits activities towards U, Hx, Xan, urea [185] and towards regular DNA containing mismatched bases, flap and pseudo-Y structures [186,187]. The Nfi-catalyzed activity towards Hx is ~20-fold higher than with any other DNA substrate [187]. The purified Nfi protein from hyperthermophilic bacteria *Thermotoga maritima* exhibits similar DNA substrate specificity with *E. coli* Nfi [188]. Interestingly, Nfi homologues are present in most Archaea except methanogenic species. EndoV from *P. furiosus* (PfuEndoV) is also similar to *E. coli* Nfi in its substrate preferences, but additionally recognizes and cleaves Hx-containing ssRNA, dsRNA and DNA/RNA hybrid with the same efficiency as the DNA substrates, suggesting a role in RNA editing [189]. PfuEndoV demonstrates much higher affinity and cleavage activity for Hx than other deaminated or mismatched bases in both ds- and ssDNA. *P. furiosus* also contains EndoQ (but no Hx-specific Family 6 UNG), which exhibits higher Hx cleavage catalytic efficiency and turnover rate and is apparently present in the cells at higher levels as compared to Nfi, suggesting that EndoQ-initiated NIR plays a major role in the removal of deaminated DNA bases in Archaea [190].

Despite the apparently higher specificity of Nfi towards Hx lesions, its function in *E. coli* remains controversial. An analysis of an *E. coli nfi* mutant showed an increased frequency of nitrous acids-induced mutations at A:T base pairs, suggesting a major role for Nfi in the *in vivo* Hx repair [191,192]. However, a later study found no increase in the level of Hx in the genomic DNA of *E. coli nfi* cells [193]. Indeed, the monofunctional alkylpurine DNA glycosylases AlkA in *E. coli* and MPG in human cells can efficiently remove Hx [194,195].

At present, the role of eukaryotic Nfi homologs in DNA repair are unclear, since these enzymes have no or very weak activity on DNA bearing Hx residues [196-200]. Instead, further studies revealed ribonuclease activity of human endonuclease V (ENDO V) on ssRNA and, with lower efficiency, dsRNA containing a single Hx residue [199,200]. In these substrates, ENDO V cleaves the second phosphodiester on the 3' side of Hx, similarly to prokaryotic Nfi acting on

DNA [199,200]. Importantly, *E. coli* Nfi is also able to cleave RNA substrates containing a Hx residue [200].

Although the Nfi-initiated “alternative excision pathway” remains on the list of minor repair schemes for specific lesions, Nfi after all may be not a true repair enzyme. To summarize the arguments against its role in DNA repair, (i) it is unusual among BER and NIR enzymes in the mechanism of DNA cleavage, which is not coupled to known repair cleansing activities and DNA strand displacement synthesis; (ii) it has non-specific DNA cleavage activity; (iii) it shows no DNA strand discrimination when incising duplex DNA with mismatches; (iv) both prokaryotic and eukaryotic homologs have high affinity for RNA. The specificity of Nfi-like proteins for RNA loops and Hx-containing RNA may rather suggest their role in RNA processing and editing [200,201].

7. Oxygen Catastrophe: back to BER

The known NIR-competent AP endonucleases can repair a wide variety of structurally distinct DNA lesions including oxidized pyrimidines, but neither highly mutagenic 8oxoG, the major oxidized base occurring in DNA under normoxic conditions, nor 8oxoA [116,154]. A huge increase in the oxygen concentration ~2.4–2.0 Ga is believed to have occurred due to the emergence of photosynthesis leading to the so-called Oxygen Catastrophe, or Great Oxidation Event (GOE) [34,35]. The abundance of O₂ in the atmosphere led to dramatic changes in the cellular metabolism and the appearance of oxidative phosphorylation that supplied sufficient amounts of energy for complex forms of life. However, leakage from the respiratory chain is an abundant ROS source, far exceeding the amounts produced by ionizing radiation. Moreover, oxygenated solutions produce other kinds of lesions than irradiation under anoxic conditions. Eventually, the evolution of life from oxygen-resistant to oxygen-dependent had to deal with new kinds of oxidized DNA bases, and also with complex DNA lesions such as bulky DNA adducts and inter-strand DNA crosslinks generated by the products of lipid peroxidation. GOE definitely post-dated the separation between Bacteria and Archaea. Regarding eukaryotes, it had long been held that their appearance broadly coincided with GOE; however, a recent integral analysis of genomic and fossil evidence put the emergence of eukaryotes after ~1.84 Ga [2]. In fact, due to multiple sinks, accumulation of O₂ in the Earth’s atmosphere cannot be described by a single oxidation event; at least two periods of O₂ buildup are distinguished: a rise from nearly zero to ~0.02–0.04 atm pO₂ ~2.4–2.0 Ga and a much more recent rise to the present-day ~0.2 atm pO₂ starting ~0.7–0.8 Ga [34,35]. This complex profile might leave its footprint in late diversification of DNA repair pathways such as the existence of multiple H2tH DNA glycosylases for oxidative lesions in multicellular eukaryotes.

Since LUCA existed long before GOE, solutions to combat the oxidative stress burden should have evolved in Bacteria and Archea/Eukaryota independently, yet they had to be rooted in already existing pathways. An instructive census of archaeal BER/NIR glycosylases and endonucleases by Ishino and colleagues [157] shows that the most conserved and diverse DNA repair enzyme in Archaea is endonuclease III (Nth), an HhH, FeS-containing bifunctional DNA glycosylase that excises a wide variety of oxidized pyrimidines. Nth is also universally found in Bacteria, making it a good candidate for a possible ancestor of the HhH glycosylases (Figure 5). Nth can excise ionizing radiation-induced anoxic nucleobase lesions (DHT, DHU) and UV-induced pyrimidine hydrates [32,202,203], and might had been the primary enzymes for the repair of the latter damage type. Alternatively, or additionally, Nth-like HhH enzymes could be responsible for AP site repair using their AP lyase activity. AlkA, which is present in bacteria and many archaeal species, probably specialized in alkylation damage repair before the separation of these two domains. However, other oxidative damage HhH DNA glycosylases seem to evolve after this split. MutY, which is uniquely suited for post-replicative repair of

8oxoG, is found only in bacteria and eukaryotes. On the other hand, 8oxoG-specific DNA glycosylases OGG1 and OGG2 are restricted to Eukaryota and Archaea (with rare findings of OGG1 in Firmicutes, possibly due to lateral gene transfer), while bacteria employ Fpg, an H2tH 8oxoG DNA glycosylase.

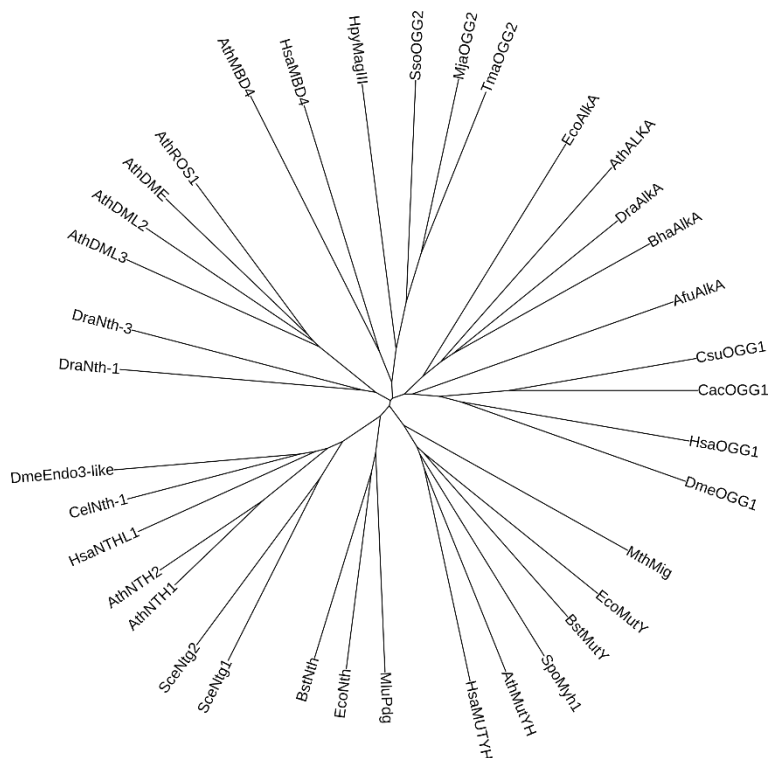


Figure 5. Unrooted tree of selected HhH DNA glycosylases. The species are: Afu, *Archaeoglobus fulgidus*; Ath, *Arabidopsis thaliana*; Bha, *Alkalihalobacillus* (formerly *Bacillus*) *halodurans*; Bst, *Geobacillus* (formerly *Bacillus*) *stearothermophilus*; Cac, *Clostridium acetobutylicum*; Cel, *Caenorhabditis elegans*; Csu, *Caldanaerobacter subterraneus* (formerly *Thermoanaerobacter tengcongensis*); Dme, *Drosophila melanogaster*; Dra, *Deinococcus radiodurans*; Eco, *Escherichia coli*; Hpy, *Helicobacter pylori*; Hsa, *Homo sapiens*; Mja, *Methanocaldococcus jannaschii*; Mlu, *Micrococcus luteus*; Mth, *Methanothermobacter thermautotrophicus* (formerly *Methanobacterium thermoformicum*); Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Sso, *Saccharolobus* (*Sulfolobus*) *solfataricus*; Tma, *Thermotoga maritima*. The selected proteins are HhH DNA glycosylases characterized either biochemically or structurally. The phylogenetic tree was produced by Clustal Omega [204] and visualized using iTOL [205].

So, if primordial DNA repair was probably performed by NIR-like endonucleases, why was it eventually supplanted by BER? One possible driving force could be the widened repertoire of lesions due to GOE. Since deaminated, alkylated and oxidative lesions are non-bulky and do not greatly distort DNA structure, there is no common mode of their recognition, as in NER. Ultimately, specialization or repair enzymes towards their cognate lesions had to occur. A critical consideration in favor of specialized DNA glycosylases rather than 5'-endonucleases is that all glycosylases produce a common product whatever the initial lesion was. In effect, today's BER has to deal with only three type of intermediates downstream of DNA glycosylases: AP sites (products of monofunctional glycosylases), 3'-terminal α,β -unsaturated aldehydes (products of β -elimination), and 3'-terminal phosphates (products of β,δ -elimination). On the contrary, specialized 5'-endonucleases would still leave the lesion in DNA, necessitating its removal by a 5'→3' exonuclease or displacement during DNA repair synthesis. Many today's

exonucleases and polymerases have problems encountering certain kinds of damage in the degraded or displaced strand, so in effect the specialized 5'-endonucleases would probably require specialized downstream repair enzymes to accommodate all kinds of DNA damage. Ultimately, the BER scheme with specialized DNA glycosylases followed by an AP endonuclease or a 3'-phosphodiesterase or a 3'-phosphatase turned out more streamlined and better suited for removal of the most frequently occurring genome lesions.

8. Direct repair as a primordial mechanism to counteract alkylation and UV damage

Direct repair is a highly efficient and energy-saving mechanism that can remove a limited range of DNA lesions induced by alkylating agents and UV light. As compared to other more sophisticated DNA repair systems such as BER and NER, the DR pathway employs a single protein and does not require multi-protein machinery, or excision of damaged nucleotides, or *de novo* DNA synthesis. Why this simple scheme does not apply to more types of lesions? We hypothesize that the reducing conditions on the primordial Earth limited induced DNA damage to only two types of bulky UV lesions and several DNA bases modifications induced by alkylation agents and ionizing radiation. Therefore, it is reasonable to presume that at the beginning LUCA evolved the easiest mechanism to counteract just a few DNA base modifications. Moreover, since DR does not depend on template synthesis, the pathway may even pre-date the DNA world.

Alkylation damage of DNA induced by SAM radicals or nitrosamines at acidic pH was probably a major challenge to genome maintenance of ancient life. Metabolically activated nitrosamines react predominantly by an S_N1 mechanism and generate preferentially 7meG, 3meA and O^6 meG residues. Due to their positive charge, 7meG and 3meA have weak glycosidic bonds and undergo accelerated depurination to AP sites, which can be efficiently repaired by AP endonucleases. On the contrary, O^6 meG, a highly mutagenic base, is quite stable even at elevated temperatures, and therefore presented a higher risk than other alkylated DNA bases. O^6 -alkylguanine DNA alkyltransferases (AGTs) that revert O^6 meG and O^4 meT back to guanine and thymine, respectively, belong to a family of highly conserved proteins that are widely represented in the three domains of life, but are apparently absent from plants. AGTs recognize the damaged nucleobase in DNA and irreversibly transfer the alkyl group to a Cys residue in their active site [206]. Since after the methyl transfer AGTs are inactivated, they are not true enzymes but rather suicidal proteins. Although removal of a single lesion by AGTs requires one protein molecule at a time, this repair mechanism is highly conserved in evolution and present in higher eukaryotes suggesting that AGTs are essential to preserve the genome stability. Moreover, today's FeS-dependent radical SAM enzymes include methyltransferases that use their Cys residues to transfer methyl groups [56], and they were also present in the LUCA genome [5], so it is quite possible that early AGTs were not suicidal.

$N1$ -methyladenine (m1A) and $N3$ -methylcytosine (m3C) are major toxic and mutagenic lesions induced by S_N2 type of alkylating agents in single-stranded DNA. In bacteria and eukaryotes, m1A and m3C are removed by oxidative demethylation catalyzed by α -ketoglutarate-dependent dioxygenases: AlkB and AlkB-like (ABH) proteins, respectively [207-209]. To repair these lesions, AlkB and ABH use Fe^{2+} to activate O_2 and oxidize the methyl group, producing an unstable hydroxymethylated intermediate that spontaneously decays to yield formaldehyde and the original DNA base [210,211]. Another atom of the O_2 molecule is consumed during the conversion of the α -ketoglutarate co-substrate to CO_2 and succinate. The requirement of oxygen in AlkB-mediated direct repair suggests that oxidative DNA demethylation appeared later in evolution after oxygen catastrophe. Indeed, no AlkB homologues have been identified in Archaea, instead, as shown for *Archaeoglobus fulgidus*, m1A and m3C are removed through the BER pathway initiated by an AlkA-like DNA glycosylase [212]. This observation may suggest

that the need for removal of some alkylated bases from DNA could be one of the factors driving the evolution of early BER. 783 784

Photoreactivation is the visible light-dependent recovery from cytotoxic DNA damage caused by UVC and UVB. It was proposed that DNA photolyase-mediated photoreactivation was one of the first DNA repair pathway to evolve [213] and the most effective one in the repair of UV induced CPD adducts [214]. DNA photolyases catalyze conversion of UV pyrimidine dimers into a normal pair of pyrimidine bases driven by absorption of photons with wavelengths 319–490 nm (reviewed in [215]). Photolyases belong to the large family of cryptochrome/photolyase proteins that are implicated in DNA repair, circadian photoreceptor, and transcriptional regulation. Members of the family are widely distributed in all three domains of life and can be divided in three classes according to their function: CPD-photolyases that repair CPD lesions, 6-4PP-photolyases that repair 6-4PP photoproducts, and cryptochromes. Cryptochromes do not participate in DNA repair, although they have probably evolved from photolyases, and instead employ their blue-light harvesting apparatus to mediate light-dependent growth in plants [216] and regulate the circadian clock in animals [217]. Photolyases are flavoproteins composed of two structural domains, an α/β HUP (HIGH-signature proteins, UspA, and PP-ATPase) domain and a helical domain [218,219]. They contain two light-absorbing chromophores: the catalytic two-electron-reduced deprotonated flavine adenine dinucleotide (FADH⁻) bound by the helical domain and a second chromophore bound by the HUP domain and acting as a light-harvesting antenna, which, in different species, can be either 8-hydroxy-7,8-didemethyl-5-deazariboflavin (8-HDF) [220,221] or methenyltetrahydrofolate (MTHF) [222,223]. The HUP domain has a very old ancestry pre-dating LUCA and rooted in a common nucleotide-binding precursor of photolyases, electron transfer flavoproteins, class I aminoacyl-tRNA synthetases, and nucleotidyl transferases [224]. Both flavins and MTHF were in the repertoire of cofactors likely used by LUCA [5], and it is possible that photolyases first appeared as RNA- or DNA-binding proteins that stabilized photoactive dimer-splitting molecules near DNA and later evolved into highly efficient two-chromophore enzymes. Within the cryptochrome/photolyase family, it has been suggested that its common ancestor might have been a flavoprotein involved in the both UV damage repair and light-dependent regulation of gene expression [225]. 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812

Unexpectedly, a recently discovered 6-4PP photolyase, PhrB from *Agrobacterium tumefaciens*, defines a new class of photolyases that lack 8-HDF and MTHF and seem to use a FeS cluster as a primary light-harvesting antenna [226]. This mechanistically connects photolyases with another group of UV damage DR proteins, so-called “spore product lyases”, which repair 5-thymine-5,6-dihydrothymine, a pyrimidine dimer that arises upon UV irradiation of dehydrated, A-form DNA [227]. These enzymes are typical FeS/radical SAM proteins, currently narrowly restricted to spore-forming bacteria, mostly Firmicutes [228]. While it is unclear whether spore product lyases are remnants of an early repair pathway or appeared later in evolution, their ancient protein fold is consistent with the idea that LUCA could already possess several ways to directly reverse UV damage. 813 814 815 816 817 818 819 820 821 822 823

9. Putative origins of NER 824

Photolyases very likely appeared in evolution before the occurrence of more sophisticated NER pathways [213]. Although the NER machinery has the advantage over photolyases because of its broader repertoire and the ability for dark repair of UV lesions, bacterial taxa nearly universally contain both UvrABC proteins and DNA photolyases. This suggests that the main biological role of the NER pathway is not to repair CPDs, but rather to deal with other bulky helix-distorting DNA adducts induced by other factors than UV. 825 826 827 828 829 830

NER by its nature is lesion-unspecific, repairing almost any lesion that strongly distorts the double helix. Strikingly, the NER machinery in eukaryotes has evolved completely independently from the bacterial and archaeal one. In eukaryotes, the central NER sensor, XPC/Rad4, descends from bacteriophage anti-restriction ArdC proteins fused with an archaeal-derived papain-like peptidase domain [229]. In bacteria, the same role is played by UvrA₂B complex, with both UvrA and UvrB belonging to a large and diverse family of P-loop NTPases that also includes MutS mismatch repair protein, RecA and many helicases [230]. In Archaea, only a few groups possess the UvrABC system, likely acquired from Bacteria via horizontal transfer, and some archaeal groups possess homologs of several eukaryotic NER proteins but no XPC or XPA [231]. Nucleases involved in 5'- and 3'-incisions in NER can be reliably traced to other nuclease families. Eukaryotic Rad2/XPG/ERCC5, the 3'-incision nuclease, is closely related to FEN1 flap endonuclease, exonuclease I and 5'→3' exonuclease domains of Family A DNA polymerases [232,233]. Both subunits of the 5'-nuclease, comprised of ERCC1 and ERCC4/XPF proteins, share a common fold with many proteins grouped together as PD-(D/E)XK family nucleases, including the MMR endonuclease MutH, specialized mismatch endonuclease Vsr, RecB and RecC components of RecBCD exonuclease, eukaryotic, archeal and phage Holliday junction resolvases and a variety of restriction endonucleases [234,235]. Bacterial UvrC proteins use two domains to nick DNA: the 5' incision domain is related to Nfi, RNases H, DNA polymerase A and B family 3'→5' exonucleases, RuvC and phage transposases [236], whereas the 3' incision GIY-YIG nuclease domain comes from a diverse group of proteins that comprises class I homing endonucleases, some II restriction enzymes and retrotransposon endonucleases [237]. It seems that eukaryotes and bacteria have independently adapted existing convenient nuclease modules to be guided by conformation-recognizing modules, also independently evolved in these two domains of life, and that in today's Archaea we can see traces of these pre-existing nucleases as they were before the appearance of NER.

What new lesions could drive NER appearance and evolution? Oxidative damage may not had been the primary cause since 8,5'-cyclopurine lesions – the only class of oxidative lesions repaired predominantly by NER – are generated under anoxic rather than normoxic conditions. An interesting possibility is that NER could have emerged as a response to ever more complex biogenic chemosphere. Today's NER is critical to cope with DNA damage due to metabolically activated xenobiotics, such as benz[a]pyrene, aminofluorenes, aflatoxins, etc. Characteristically, even today specialized varieties of NER exist to counteract metabolically related DNA damage. For example, the genomes of Streptomyces that produce doxorubicin and daunorubicin, two anthracycline intercalating antibiotics, encode the DrrC protein, a homolog of UvrA involved in the NER of lesions caused by these compounds [238]. In *E. coli*, Cho endonuclease is a homolog of UvrC 3'-incision domain with a specificity for a subset of bulky adducts (cholesterol, menthol) and psoralen-induced interstrand cross-links [239,240]. The ability of NER to fight a wide range of lesions could become its critical advantage in a world full of more and more diverse genotoxic compounds.

6. Conclusion: “Nothing in biology makes sense except in the light of evolution”

Modern molecular biology took its origin after Watson and Crick's discovery of the double helix structure of DNA and since then made great achievements in the understanding what life is. Molecular biology is inherently reductionist, trying to understand biological systems in terms of their components and ultimately reduce the complexity and diversity of living matter to universal physical and chemical principles. However, many questions in biology remain difficult to address solely in a reductionist way. For example, solving the atomic structure of a protein provides critical information on its functions, but offers little insight into its origins. We know that this protein fold can catalyze this reaction, but how was it selected from many other

880 folds that could buttress the same arrangement of the reacting groups? On the other hand, evo-
881 lutionary theory is inherently non-reductionist because of the randomness of variation and the
882 principle of natural selection, which acts on the whole organisms and populations. We can use
883 sequences of genes and whole genomes to interpret large-scale patterns of the evolution, per-
884 haps up to the earliest steps of life, but cannot see with this information how the parts interact
885 with each other in the cell and the body. Only combining Darwinian principles of evolution
886 with molecular biology we can arrive to a holistic approach to address the most interesting
887 questions.

888 DNA repair field had its start even before Watson and Crick, from the seminal works of
889 Albert Kelner and Renato Dulbecco who independently discovered photoreactivation in the
890 late 1940s. A wealth of molecular biology information accumulated since then allowed us to
891 decipher mechanisms of DNA repair in the three domains of life and even in viruses. Now we
892 have enough reductionist tiles to start the assembly of the evolutionary puzzle and understand
893 the origins of repair pathways starting from (and maybe even before) LUCA. The analysis of
894 the repair pathways in Archaea lead us to propose that AP endonucleases and DNA damage
895 reversal enzymes were the earliest DNA repair players that counteracted genotoxic effects of
896 spontaneous decay of DNA (deamination and base loss) and high-energy electromagnetic radi-
897 ation (UV pyrimidine dimers and α -anomeric 2'-deoxynucleotides). Thus, AP endonucleases
898 entered the repair field before monofunctional uracil-DNA glycosylases, and photolyases, be-
899 fore sophisticated NER machinery. Yet bifunctional DNA glycosylases exemplified by the Nth-
900 like proteins probably appeared very early, together with AP endonucleases, but their activities
901 were not directly coupled to NIR AP endonucleases and the modern BER pathway described in
902 bacteria and eukaryotes appeared much later.

903 Primordial DNA excision repair mechanisms most likely appeared with the separation of
904 functions between RNA and DNA, which in all probability pre-dated cellular forms of life. The
905 primeval AP endonucleases and DNA glycosylases might have evolved with the appearance of
906 DNA genomes not only to protect them from spurious damage but also to eliminate non-stand-
907 ard bases, including U, Hx, pseudouridine, N^6 -methyladenine, 5meC and other base modifica-
908 tions, from the genome carrier.

909 We propose that the Great Oxidation Event played a crucial role in the origin of modern
910 BER and NER pathways. Appearance of oxidative phosphorylation provided energy for devel-
911 opment of complex multicellular forms of life, but on the other side, extended the spectra of
912 genotoxic DNA base damage. Increased structural variety of DNA damage resulted in the spe-
913 cialization of DNA repair enzymes towards their cognate lesions. This specialization resulted in
914 the appearance of multiple DNA glycosylases which converted every distinct base modification
915 into only three type of DNA repair intermediates, which are then taken care of by more special-
916 ized AP endonucleases/3'-repair phosphodiesterases, rather than by a single universal NIR AP
917 endonuclease.

918
919 **Author Contributions:** Writing—original draft preparation J.L., D.O.Z and M.S.; figures and table
920 preparations, A.A.I., D.O.Z. and M.S.; writing—review and editing: P.P. and A.A.I.; funding acquisition,
921 I.G., A.A.I., B.T.M., D.O.Z. and M.S.; All authors discussed and contributed to analysis of published litera-
922 ture and to writing the manuscript. All authors have read and agreed to the published version of the man-
923 uscript.

924 **Funding:** This research was supported by grants to Murat Saparbaev from la Ligue National Contre le
925 Cancer “Equipe Labellisee LIGUE 2016” (<https://www.ligue-cancer.net>) and Electricité de France (RB
926 2021-05) (<http://www.edf.fr>); to Alexander Ishchenko from the French National Research Agency (ANR-
927 18-CE44-0008) and Fondation ARC (PJA-20181208015); to Bakhyt T. Matkarimov from Nazarbayev Uni-
928 versity Oak Ridge Associated Universities (ORAU) grant 091019CRP2111; to Inga Grin from Russian Sci-
929 ence Foundation (grant 17-14-01190; analysis of DNA glycosylases); to Dmitry Zharkov from the Russian

Ministry of Science and Higher Education (AAAA-A17-117020210023-1). Paulina Prorok was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [Project-ID 393547839 – SFB 1361, CA198/9-2, CA 198/12-1].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank other members of Drs Saparbaev and Zharkov's laboratories for their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Dodd, M.S.; Papineau, D.; Grenne, T.; Slack, J.F.; Rittner, M.; Pirajno, F.; O'Neil, J.; Little, C.T.S. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **2017**, *543*, 60–64. 10.1038/nature21377
- Betts, H.C.; Puttick, M.N.; Clark, J.W.; Williams, T.A.; Donoghue, P.C.J.; Pisani, D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2018**, *2*, 1556–1562. 10.1038/s41559-018-0644-x
- Pearce, B.K.D.; Tupper, A.S.; Pudritz, R.E.; Higgs, P.G. Constraining the time interval for the origin of life on Earth. *Astrobiology* **2018**, *18*, 343–364. 10.1089/ast.2017.1674
- Kozbial, P.Z.; Mushegian, A.R. Natural history of S-adenosylmethionine-binding proteins. *BMC Struct. Biol.* **2005**, *5*, 19. 10.1186/1472-6807-5-19
- Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. 10.1038/nmicrobiol.2016.116
- Douki, T.; Court, M.; Sauvaigo, S.; Odin, F.; Cadet, J. Formation of the main UV-induced thymine dimeric lesions within isolated and cellular DNA as measured by high performance liquid chromatography-tandem mass spectrometry. *J. Biol. Chem.* **2000**, *275*, 11678–11685. 10.1074/jbc.275.16.11678
- Westall, F.; de Ronde, C.E.J.; Southam, G.; Grassineau, N.; Colas, M.; Cockell, C.; Lammer, H. Implications of a 3.472–3.333 Gyr-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental conditions on the early Earth. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 1857–1875. 10.1098/rstb.2006.1896
- Ward, J.F. Complexity of damage produced by ionizing radiation. *Cold Spring Harb. Symp. Quant. Biol.* **2000**, *65*, 377–382. 10.1101/sqb.2000.65.377
- Dizdaroglu, M.; Jaruga, P. Mechanisms of free radical-induced damage to DNA. *Free Radic. Res.* **2012**, *46*, 382–419. 10.3109/10715762.2011.653969
- Breimer, L.H.; Lindahl, T. Thymine lesions produced by ionizing radiation in double-stranded DNA. *Biochemistry* **1985**, *24*, 4018–4022. 10.1021/bi00336a032
- Boorstein, R.J.; Hilbert, T.P.; Cunningham, R.P.; Teebor, G.W. Formation and stability of repairable pyrimidine photohydrates in DNA. *Biochemistry* **1990**, *29*, 10455–10460. 10.1021/bi00498a004
- Ganguly, T.; Duker, N.J. Stability of DNA thymine hydrates. *Nucleic Acids Res.* **1991**, *19*, 3319–3323. 10.1093/nar/19.12.3319
- Lawley, P.D.; Phillips, D.H. DNA adducts from chemotherapeutic agents. *Mutat. Res.* **1996**, *355*, 13–40. 10.1016/0027-5107(96)00020-6

14. Nakamura, J.; Mutlu, E.; Sharma, V.; Collins, L.; Bodnar, W.; Yu, R.; Lai, Y.; Moeller, B.; Lu, K.; Swenberg, J. The endogenous exposome. *DNA Repair* **2014**, *19*, 3–13. 10.1016/j.dnarep.2014.03.031
15. Williams, J.S.; Kunkel, T.A. Ribonucleotides in DNA: Origins, repair and consequences. *DNA Repair* **2014**, *19*, 27–37. 10.1016/j.dnarep.2014.03.029
16. Kunkel, T.A.; Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **2000**, *69*, 497–529. 10.1146/annurev.biochem.69.1.497
17. Ischenko, A.A.; Saparbaev, M.K. Alternative nucleotide incision repair pathway for oxidative DNA damage. *Nature* **2002**, *415*, 183–187. 10.1038/415183a
18. Georg, J.; Schomacher, L.; Chong, J.P.J.; Majerník, A.I.; Raabe, M.; Urlaub, H.; Müller, S.; Ciirdaeva, E.; Kramer, W.; Fritz, H.-J. The *Methanothermobacter thermautotrophicus* ExoIII homologue Mth212 is a DNA uridine endonuclease. *Nucleic Acids Res.* **2006**, *34*, 5325–5336. 10.1093/nar/gkl604
19. Ishchenko, A.A.; Deprez, E.; Maksimenko, A.; Brochon, J.-C.; Tauc, P.; Saparbaev, M.K. Uncoupling of the base excision and nucleotide incision repair pathways reveals their respective biological roles. *Proc. Natl Acad. Sci. U.S.A.* **2006**, *103*, 2564–2569. 10.1073/pnas.0508582103
20. Imlay, J.A.; Linn, S. DNA damage and oxygen radical toxicity. *Science* **1988**, *240*, 1302–1309. 10.1126/science.3287616
21. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709–715. 10.1038/362709a0
22. Lindahl, T.; Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **1972**, *11*, 3610–3618. 10.1021/bi00769a018
23. Lindahl, T.; Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **1974**, *13*, 3405–3410. 10.1021/bi00713a035
24. Boiteux, S.; Laval, J. Coding properties of poly(deoxycytidylic acid) templates containing uracil or apyrimidinic sites: In vitro modulation of mutagenesis by deoxyribonucleic acid repair enzymes. *Biochemistry* **1982**, *21*, 6746–6751. 10.1021/bi00269a020
25. Strauss, B.S. The ‘A rule’ of mutagen specificity: A consequence of DNA polymerase bypass of non-instructional lesions? *Bioessays* **1991**, *13*, 79–84. 10.1002/bies.950130206
26. Krokan, H.E.; Drabløs, F.; Slupphaug, G. Uracil in DNA – occurrence, consequences and repair. *Oncogene* **2002**, *21*, 8935–8948. 10.1038/sj.onc.1205996
27. Alseth, I.; Dalhus, B.; Bjørås, M. Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.* **2014**, *26*, 116–123. 10.1016/j.gde.2014.07.008
28. Wuenschell, G.E.; O’Connor, T.R.; Termini, J. Stability, miscoding potential, and repair of 2’-deoxyxanthosine in DNA: Implications for nitric oxide-induced mutagenesis. *Biochemistry* **2003**, *42*, 3608–3616. 10.1021/bi0205597
29. Poole, A.; Penny, D.; Sjöberg, B.-M. Confounded cytosine! Tinkering and the evolution of DNA. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 147–151. 10.1038/35052091
30. Beukers, R.; Ylstra, J.; Berends, W. The effect of ultraviolet light on some components of the nucleic acids: II: In rapidly frozen solutions. *Rec. Trav. Chim.* **1958**, *77*, 729–732. 10.1002/recl.19580770806
31. Douki, T.; Sage, E. Dewar valence isomers, the third type of environmentally relevant DNA photoproducts induced by solar radiation. *Photochem. Photobiol. Sci.* **2016**, *15*, 24–30. 10.1039/C5PP00382B
32. Boorstein, R.J.; Hilbert, T.P.; Cadet, J.; Cunningham, R.P.; Teebor, G.W. UV-induced pyrimidine hydrates in DNA are repaired by bacterial and mammalian DNA glycosylase activities. *Biochemistry* **1989**, *28*, 6164–6170. 10.1021/bi00441a007
33. Ganguly, T.; Weems, K.M.; Duker, N.J. Ultraviolet-induced thymine hydrates in DNA are excised by bacterial and human DNA glycosylase activities. *Biochemistry* **1990**, *29*, 7222–7228. 10.1021/bi00483a009

34. Holland, H.D. The oxygenation of the atmosphere and oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 903-915. 10.1098/rstb.2006.1838
35. Lyons, T.W.; Reinhard, C.T.; Planavsky, N.J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **2014**, *506*, 307-315. 10.1038/nature13068
36. Tarduno, J.A.; Cottrell, R.D.; Davis, W.J.; Nimmo, F.; Bono, R.K. A Hadean to Paleoproterozoic geodynamo recorded by single zircon crystals. *Science* **2015**, *349*, 521-524. 10.1126/science.aaa9114
37. Doglioni, C.; Pignatti, J.; Coleman, M. Why did life develop on the surface of the Earth in the Cambrian? *Geosci. Front.* **2016**, *7*, 865-873. 10.1016/j.gsf.2016.02.001
38. Pogozelski, W.K.; Tullius, T.D. Oxidative strand scission of nucleic acids: Routes initiated by hydrogen abstraction from the sugar moiety. *Chem. Rev.* **1998**, *98*, 1089-1107. 10.1021/cr960437i
39. Teoule, R.; Bert, C.; Bonicel, A. Thymine fragment damage retained in the DNA polynucleotide chain after gamma irradiation in aerated solutions. II. *Radiat. Res.* **1977**, *72*, 190-200. 10.2307/3574690
40. Teoule, R.; Bonicel, A.; Bert, C.; Fouque, B. Isolation and identification of thymine products from DNA γ irradiated in oxygen-free aqueous solutions. *J. Am. Chem. Soc.* **1978**, *100*, 6749-6750. 10.1021/ja00489a034
41. Bonicel, A.; Mariaggi, N.; Hughes, E.; Teoule, R. In vitro γ irradiation of DNA: Identification of radioinduced chemical modifications of the adenine moiety. *Radiat. Res.* **1980**, *83*, 19-26. 10.2307/3575255
42. Furlong, E.A.; Jorgensen, T.J.; Henner, W.D. Production of dihydrothymidine stereoisomers in DNA by γ -irradiation. *Biochemistry* **1986**, *25*, 4344-4349. 10.1021/bi00363a025
43. Akhlaq, M.S.; Schuchmann, H.-P.; von Sonntag, C. The reverse of the 'repair' reaction of thiols: H-abstraction at carbon by thiyl radicals. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **1987**, *51*, 91-102. 10.1080/09553008714550531
44. Lesiak, K.B.; Wheeler, K.T. Formation of α -deoxyadenosine in polydeoxynucleotides exposed to ionizing radiation under anoxic conditions. *Radiat. Res.* **1990**, *121*, 328-337.
45. von Sonntag, C. *Free-Radical-Induced DNA Damage and Its Repair: A Chemical Perspective*; Springer: Berlin - Heidelberg, 2006; pp. 523.
46. Lawley, P.D. Effects of some chemical mutagens and carcinogens on nucleic acids. *Prog. Nucleic Acid Res. Mol. Biol.* **1966**, *5*, 89-131. 10.1016/S0079-6603(08)60232-9
47. Singer, B.; Kuśmirek, J.T. Chemical mutagenesis. *Annu. Rev. Biochem.* **1982**, *51*, 655-691. 10.1146/annurev.bi.51.070182.003255
48. Singer, B. All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature* **1976**, *264*, 333-339. 10.1038/264333a0
49. Pegg, A.E. Methylation of the O⁶ position of guanine in DNA is the most likely initiating event in carcinogenesis by methylating agents. *Cancer Invest.* **1984**, *2*, 223-231. 10.3109/07357908409104376
50. Bodell, W.J.; Singer, B. Influence of hydrogen bonding in DNA and polynucleotides on reaction of nitrogens and oxygens toward ethylnitrosourea. *Biochemistry* **1979**, *18*, 2860-2863. 10.1021/bi00580a029
51. Boiteux, S.; Laval, J. Mutagenesis by alkylating agents: Coding properties for DNA polymerase of poly (dC) template containing 3-methylcytosine. *Biochimie* **1982**, *64*, 637-641. 10.1016/S0300-9084(82)80103-X
52. Larson, K.; Sahm, J.; Shenkar, R.; Strauss, B. Methylation-induced blocks to in vitro DNA replication. *Mutat. Res.* **1985**, *150*, 77-84. 10.1016/0027-5107(85)90103-4
53. Swann, P.F. Why do O⁶-alkylguanine and O⁴-alkylthymine miscode? The relationship between the structure of DNA containing O⁶-alkylguanine and O⁴-alkylthymine and the mutagenic properties of these bases. *Mutat. Res.* **1990**, *233*, 81-94. 10.1016/0027-5107(90)90153-U

54. Taverna, P.; Sedgwick, B. Generation of an endogenous DNA-methylating agent by nitrosation in *Escherichia coli*. *J. Bacteriol.* **1996**, *178*, 5105-5111. 1054 1055
55. Posnick, L.M.; Samson, L.D. Influence of S-adenosylmethionine pool size on spontaneous mutation, dam methylation, and cell growth of *Escherichia coli*. *J. Bacteriol.* **1999**, *181*, 6756-6762. 10.1128/JB.181.21.6756-6762.1999 1056 1057
56. Broderick, J.B.; Duffus, B.R.; Duschene, K.S.; Shepard, E.M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **2014**, *114*, 4229-4317. 10.1021/cr4004709 1058 1059
57. Imlay, J.A.; Chin, S.M.; Linn, S. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. *Science* **1988**, *240*, 640-642. 10.1126/science.2834821 1060 1061
58. Henle, E.S.; Linn, S. Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. *J. Biol. Chem.* **1997**, *272*, 19095-19098. 10.1074/jbc.272.31.19095 1062 1063
59. Bjelland, S.; Seeberg, E. Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutat. Res.* **2003**, *531*, 37-80. 10.1016/j.mrfmmm.2003.07.002 1064 1065
60. Kasai, H.; Tanooka, H.; Nishimura, S. Formation of 8-hydroxyguanine residues in DNA by X-irradiation. *Cann* **1984**, *75*, 1037-1039. 1066 1067
61. Dizdaroglu, M.; Bergtold, D.S. Characterization of free radical-induced base damage in DNA at biologically relevant levels. *Anal. Biochem.* **1986**, *156*, 182-188. 10.1016/0003-2697(86)90171-5 1068 1069
62. Schuchmann, M.N.; Steenzen, S.; Wroblewski, J.; von Sonntag, C. Site of OH radical attack on dihydrouracil and some of its methyl derivatives. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **1984**, *46*, 225-232. 10.1080/09553008414551341 1070 1071
63. Dizdaroglu, M. Free-radical-induced formation of an 8,5'-cyclo-2'-deoxyguanosine moiety in deoxyribonucleic acid. *Biochem. J.* **1986**, *238*, 247-254. 10.1042/bj2380247 1072 1073
64. Grollman, A.P.; Moriya, M. Mutagenesis by 8-oxoguanine: An enemy within. *Trends Genet.* **1993**, *9*, 246-249. 10.1016/0168-9525(93)90089-Z 1074 1075
65. Kreutzer, D.A.; Essigmann, J.M. Oxidized, deaminated cytosines are a source of C → T transitions *in vivo*. *Proc. Natl Acad. Sci. U.S.A.* **1998**, *95*, 3578-3582. 1076 1077
66. Laval, J.; Boiteux, S. Repair of cyclic nucleic acid adducts and adverse effects of apurinic sites. *IARC Sci. Publ.* **1986**, *70*, 381-385. 1078 1079
67. Basu, A.K.; Loechler, E.L.; Leadon, S.A.; Essigmann, J.M. Genetic effects of thymine glycol: Site-specific mutagenesis and molecular modeling studies. *Proc. Natl Acad. Sci. U.S.A.* **1989**, *86*, 7677-7681. 1080 1081
68. Ide, H.; Petrullo, L.A.; Hatahet, Z.; Wallace, S.S. Processing of DNA base damage by DNA polymerases: Dihydrothymine and β-ureidoisobutyric acid as models for instructive and noninstructive lesions. *J. Biol. Chem.* **1991**, *266*, 1469-1477. 1082 1083
69. Demple, B.; DeMott, M.S. Dynamics and diversions in base excision DNA repair of oxidized abasic lesions. *Oncogene* **2002**, *21*, 8926-8934. 10.1038/sj.onc.1206178 1084 1085
70. Marnett, L.J.; Burcham, P.C. Endogenous DNA adducts: Potential and paradox. *Chem. Res. Toxicol.* **1993**, *6*, 771-785. 10.1021/tx00036a005 1086 1087
71. Burcham, P.C. Genotoxic lipid peroxidation products: Their DNA damaging properties and role in formation of endogenous DNA adducts. *Mutagenesis* **1998**, *13*, 287-305. 10.1093/mutage/13.3.287 1088 1089
72. Chung, F.-L.; Nath, R.G.; Nagao, M.; Nishikawa, A.; Zhou, G.-D.; Randerath, K. Endogenous formation and significance of 1,N²-propanodeoxyguanosine adducts. *Mutat. Res.* **1999**, *424*, 71-81. 10.1016/S0027-5107(99)00009-3 1090 1091
73. Stone, M.P.; Cho, Y.-J.; Huang, H.; Kim, H.-Y.; Kozekov, I.D.; Kozekova, A.; Wang, H.; Minko, I.G.; Lloyd, R.S.; Harris, T.M., et al. Interstrand DNA cross-links induced by α,β-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc. Chem. Res.* **2008**, *41*, 793-804. 10.1021/ar700246x 1092 1093 1094

74. Barker, S.; Weinfeld, M.; Murray, D. DNA–protein crosslinks: Their induction, repair, and biological consequences. *Mutat. Res.* **2005**, *589*, 111–135. 10.1016/j.mrrev.2004.11.003 1095
1096
75. Mushegian, A.R.; Koonin, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. U.S.A.* **1996**, *93*, 10268–10273. 10.1073/pnas.93.19.10268 1097
1098
76. Hutchison, C.A., III; Peterson, S.N.; Gill, S.R.; Cline, R.T.; White, O.; Fraser, C.M.; Smith, H.O.; Venter, J.C. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **1999**, *286*, 2165–2169. 10.1126/science.286.5447.2165 1099
1100
77. Kobayashi, K.; Ehrlich, S.D.; Albertini, A.; Amati, G.; Andersen, K.K.; Arnaud, M.; Asai, K.; Ashikaga, S.; Aymerich, S.; Bessieres, P., et al. Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. U.S.A.* **2003**, *100*, 4678–4683. 1101
10.1073/pnas.0730515100 1102
1103
78. Gil, R.; Silva, F.J.; Peretó, J.; Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 518–537. 10.1128/MMBR.68.3.518–537.2004 1104
1105
79. Forster, A.C.; Church, G.M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2006**, *2*, 45. 10.1038/msb4100090 1106
80. Glass, J.I.; Assad-Garcia, N.; Alperovich, N.; Yooseph, S.; Lewis, M.R.; Maruf, M.; Hutchison, C.A., III; Smith, H.O.; Venter, J.C. Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. U.S.A.* **2006**, *103*, 425–430. 1107
10.1073/pnas.0510013103 1108
1109
81. Venetz, J.E.; Del Medico, L.; Wölfe, A.; Schächle, P.; Bucher, Y.; Appert, D.; Tschan, F.; Flores-Tinoco, C.E.; van Kooten, M.; Guennoun, R., et al. Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl Acad. Sci. U.S.A.* **2019**, *116*, 8070–8079. 10.1073/pnas.1818259116 1110
1111
1112
82. Mirkin, B.G.; Fenner, T.I.; Galperin, M.Y.; Koonin, E.V. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* **2003**, *3*, 2. 10.1186/1471-2148-3-2 1113
1114
1115
83. Tuller, T.; Birin, H.; Gophna, U.; Kupiec, M.; Ruppín, E. Reconstructing ancestral gene content by coevolution. *Genome Res.* **2010**, *20*, 122–132. 10.1101/gr.096115.109 1116
1117
84. Lake, J.A.; Larsen, J.; Tran, D.T.; Sinsheimer, J.S. Uncovering the genomic origins of life. *Genome Biol. Evol.* **2018**, *10*, 1705–1714. 10.1093/gbe/evy129 1118
1119
85. Woese, C. The universal ancestor. *Proc. Natl Acad. Sci. U.S.A.* **1998**, *95*, 6854–6859. 10.1073/pnas.95.12.6854 1120
86. Koonin, E.V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, *1*, 127–136. 10.1038/nrmicro751 1121
1122
87. Theobald, D.L. A formal test of the theory of universal common ancestry. *Nature* **2010**, *465*, 219–222. 10.1038/nature09014 1123
88. Cavalier-Smith, T. The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 7–76. 10.1099/00207713-52-1-7 1124
1125
89. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K., et al. A new view of the tree of life. *Nat. Microbiol.* **2016**, *1*, 16048. 10.1038/nmicrobiol.2016.48 1126
1127
90. Cavalier-Smith, T.; Chao, E.E.-Y. Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaeobacteria). *Protoplasma* **2020**, *257*, 621–753. 10.1007/s00709-019-01442-7 1128
1129
91. Ku, C.; Nelson-Sathi, S.; Roettger, M.; Sousa, F.L.; Lockhart, P.J.; Bryant, D.; Hazkani-Covo, E.; McInerney, J.O.; Landan, G.; Martin, W.F. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* **2015**, *524*, 427–432. 1130
10.1038/nature14963 1131
1132
92. Méheust, R.; Bhattacharya, D.; Pathmanathan, J.S.; McInerney, J.O.; Lopez, P.; Baptiste, E. Formation of chimeric genes with essential functions at the origin of eukaryotes. *BMC Biol.* **2018**, *16*, 30. 10.1186/s12915-018-0500-0 1133
1134
93. Fuss, J.O.; Tsai, C.-L.; Ishida, J.P.; Tainer, J.A. Emerging critical roles of Fe–S clusters in DNA replication and repair. *Biochim. Biophys. Acta* **2015**, *1853*, 1253–1271. 10.1016/j.bbamcr.2015.01.018 1135
1136

94. Khodour, Y.; Kaguni, L.S.; Stiban, J. Iron–sulfur clusters in nucleic acid metabolism: Varying roles of ancient cofactors. *Enzymes* **2019**, *45*, 225–256. 10.1016/bs.enz.2019.08.003 1137
1138
95. Copley, R.R.; Bork, P. Homology among ($\beta\alpha$)₈ barrels: Implications for the evolution of metabolic pathways. *J. Mol. Biol.* **2000**, *303*, 627–641. 10.1006/jmbi.2000.4152 1139
1140
96. Sofia, H.J.; Chen, G.; Hetzler, B.G.; Reyes-Spindola, J.F.; Miller, N.E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: Functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **2001**, *29*, 1097–1106. 10.1093/nar/29.5.1097 1141
1142
1143
97. Hitomi, K.; Iwai, S.; Tainer, J.A. The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair* **2007**, *6*, 410–428. 10.1016/j.dnarep.2006.10.004 1144
1145
98. Mullins, E.A.; Rodriguez, A.A.; Bradley, N.P.; Eichman, B.F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* **2019**, *44*, 765–781. 10.1016/j.tibs.2019.04.006 1146
1147
99. Roberts, R.J.; Cheng, X. Base flipping. *Annu. Rev. Biochem.* **1998**, *67*, 181–198. 10.1146/annurev.biochem.67.1.181 1148
100. Hong, S.; Cheng, X. DNA base flipping: A general mechanism for writing, reading, and erasing DNA modifications. *Adv. Exp. Med. Biol.* **2016**, *945*, 321–341. 10.1007/978-3-319-43624-1_14 1149
1150
101. Dinner, A.R.; Blackburn, G.M.; Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **2001**, *413*, 752–755. 10.1038/35099587 1151
1152
102. Rios, A.C.; Tor, Y. Refining the genetic alphabet: A late-period selection pressure? *Astrobiology* **2012**, *12*, 884–891. 10.1089/ast.2011.0789 1153
1154
103. Aravind, L.; Koonin, E.V. The α/β fold uracil DNA glycosylases: A common origin with diverse fates. *Genome Biol.* **2000**, *1*, research0007. 10.1186/gb-2000-1-4-research0007 1155
1156
104. Lucas-Lledó, J.I.; Maddamsetti, R.; Lynch, M. Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol. Biol. Evol.* **2011**, *28*, 1307–1317. 10.1093/molbev/msq318 1157
1158
105. Hoseki, J.; Okamoto, A.; Masui, R.; Shibata, T.; Inoue, Y.; Yokoyama, S.; Kuramitsu, S. Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus thermophilus* HB8. *J. Mol. Biol.* **2003**, *333*, 515–526. 10.1016/j.jmb.2003.08.030 1159
1160
106. Kosaka, H.; Hoseki, J.; Nakagawa, N.; Kuramitsu, S.; Masui, R. Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *J. Mol. Biol.* **2007**, *373*, 839–850. 10.1016/j.jmb.2007.08.022 1161
1162
107. Lee, H.-W.; Dominy, B.N.; Cao, W. New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J. Biol. Chem.* **2011**, *286*, 31282–31287. 10.1074/jbc.M111.249524 1163
1164
108. Engstrom, L.M.; Partington, O.A.; David, S.S. An iron–sulfur cluster loop motif in the *Archaeoglobus fulgidus* uracil–DNA glycosylase mediates efficient uracil recognition and removal. *Biochemistry* **2012**, *51*, 5187–5197. 10.1021/bi3000462 1165
1166
109. Wibley, J.E.A.; Waters, T.R.; Haushalter, K.; Verdine, G.L.; Pearl, L.H. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol. Cell* **2003**, *11*, 1647–1659. 10.1016/S1097-2765(03)00235-1 1167
1168
110. Xia, B.; Liu, Y.; Li, W.; Brice, A.R.; Dominy, B.N.; Cao, W. Specificity and catalytic mechanism in family 5 uracil DNA glycosylase. *J. Biol. Chem.* **2014**, *289*, 18413–18426. 10.1074/jbc.M114.567354 1169
1170
111. Pang, P.; Yang, Y.; Li, J.; Wang, Z.; Cao, W.; Xie, W. SMUG2 DNA glycosylase from *Pedobacter heparinus* as a new subfamily of the UDG superfamily. *Biochem. J.* **2017**, *474*, 923–938. 10.1042/BCJ20160934 1171
1172
112. Chembazhi, U.V.; Patil, V.V.; Sah, S.; Reeve, W.; Tiwari, R.P.; Woo, E.; Varshney, U. Uracil DNA glycosylase (UDG) activities in *Bradyrhizobium diazoefficiens*: Characterization of a new class of UDG with broad substrate specificity. *Nucleic Acids Res.* **2017**, *45*, 5863–5876. 10.1093/nar/gkx209 1173
1174
1175
113. Cortázar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A.L.; Siegrist, F., et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **2011**, *470*, 419–423. 10.1038/nature09672 1176
1177
1178

114. Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D., et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **2011**, *146*, 67–79. 10.1016/j.cell.2011.06.020
115. Saparbaev, M.; Laval, J. 3,*N*⁴-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *Proc. Natl Acad. Sci. U.S.A.* **1998**, *95*, 8508–8513.
116. Talhaoui, I.; Couvé, S.; Ishchenko, A.A.; Kunz, C.; Schär, P.; Saparbaev, M. 7,8-dihydro-8-oxoadenine, a highly mutagenic adduct, is repaired by *Escherichia coli* and human mismatch-specific uracil/thymine-DNA glycosylases. *Nucleic Acids Res.* **2013**, *41*, 912–923. 10.1093/nar/gks1149
117. Doherty, A.J.; Serpell, L.C.; Ponting, C.P. The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res.* **1996**, *24*, 2488–2497. 10.1093/nar/24.13.2488
118. Aravind, L.; Walker, D.R.; Koonin, E.V. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* **1999**, *27*, 1223–1242. 10.1093/nar/27.5.1223
119. Shao, X.; Grishin, N.V. Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res.* **2000**, *28*, 2643–2650. 10.1093/nar/28.14.2643
120. Deterding, L.J.; Prasad, R.; Mullen, G.P.; Wilson, S.H.; Tomer, K.B. Mapping of the 5′-2-deoxyribose-5-phosphate lyase active site in DNA polymerase β by mass spectrometry. *J. Biol. Chem.* **2000**, *275*, 10463–10471. 10.1074/jbc.275.14.10463
121. García-Díaz, M.; Bebenek, K.; Kunkel, T.A.; Blanco, L. Identification of an intrinsic 5′-deoxyribose-5-phosphate lyase activity in human DNA polymerase λ: A possible role in base excision repair. *J. Biol. Chem.* **2001**, *276*, 34659–34663. 10.1074/jbc.M106336200
122. Thayer, M.M.; Ahern, H.; Xing, D.; Cunningham, R.P.; Tainer, J.A. Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **1995**, *14*, 4108–4120.
123. Nash, H.M.; Bruner, S.D.; Schärer, O.D.; Kawate, T.; Addona, T.A.; Spooner, E.; Lane, W.S.; Verdine, G.L. Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr. Biol.* **1996**, *6*, 968–980. 10.1016/S0960-9822(02)00641-3
124. Zharkov, D.O.; Golan, G.; Gilboa, R.; Fernandes, A.S.; Gerchman, S.E.; Kycia, J.H.; Rieger, R.A.; Grollman, A.P.; Shoham, G. Structural analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J.* **2002**, *21*, 789–800. 10.1093/emboj/21.4.789
125. Prakash, A.; Doublé, S.; Wallace, S.S. The Fpg/Nei family of DNA glycosylases: Substrates, structures, and search for damage. *Prog. Mol. Biol. Transl. Sci.* **2012**, *110*, 71–91. 10.1016/B978-0-12-387665-2.00004-3
126. Verly, W.G.; Paquette, Y. An endonuclease for depurinated DNA in *Escherichia coli* B. *Can. J. Biochem.* **1972**, *50*, 217–224. 10.1139/o72-029
127. Barzilay, G.; Hickson, I.D. Structure and function of apurinic/apyrimidinic endonucleases. *Bioessays* **1995**, *17*, 713–719. 10.1002/bies.950170808
128. Mol, C.D.; Hosfield, D.J.; Tainer, J.A. Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: The 3′ ends justify the means. *Mutat. Res.* **2000**, *460*, 211–229. 10.1016/S0921-8777(00)00028-8
129. Wilson, D.M., III; Barsky, D. The major human abasic endonuclease: Formation, consequences and repair of abasic lesions in DNA. *Mutat. Res.* **2001**, *485*, 283–307. 10.1016/S0921-8777(01)00063-5
130. Suck, D.; Oefner, C. Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA. *Nature* **1986**, *321*, 620–625. 10.1038/321620a0
131. Mol, C.D.; Kuo, C.-F.; Thayer, M.M.; Cunningham, R.P.; Tainer, J.A. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* **1995**, *374*, 381–386. 10.1038/374381a0

132. Mol, C.D.; Izumi, T.; Mitra, S.; Tainer, J.A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature* **2000**, *403*, 451–456. 10.1038/35000249
133. Whisstock, J.C.; Romero, S.; Gurung, R.; Nandurkar, H.; Ooms, L.M.; Bottomley, S.P.; Mitchell, C.A. The inositol polyphosphate 5-phosphatases and the apurinic/apyrimidinic base excision repair endonucleases share a common mechanism for catalysis. *J. Biol. Chem.* **2000**, *275*, 37055–37061. 10.1074/jbc.M006244200
134. Shi, K.; Kurahashi, K.; Gao, R.; Tsutakawa, S.E.; Tainer, J.A.; Pommier, Y.; Aihara, H. Structural basis for recognition of 5'-phosphotyrosine adducts by Tdp2. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1372–1377. 10.1038/nsmb.2423
135. Hosfield, D.J.; Guan, Y.; Haas, B.J.; Cunningham, R.P.; Tainer, J.A. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **1999**, *98*, 397–408. 10.1016/S0092-8674(00)81968-6
136. Tsutakawa, S.E.; Shin, D.S.; Mol, C.D.; Izumi, T.; Arvai, A.S.; Mantha, A.K.; Szczesny, B.; Ivanov, I.N.; Hosfield, D.J.; Maiti, B., et al. Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013**, *288*, 8445–8455. 10.1074/jbc.M112.422774
137. Demple, B.; Halbrook, J.; Linn, S. *Escherichia coli xth* mutants are hypersensitive to hydrogen peroxide. *J. Bacteriol.* **1983**, *153*, 1079–1082.
138. Cunningham, R.P.; Saporito, S.M.; Spitzer, S.G.; Weiss, B. Endonuclease IV (*nfo*) mutant of *Escherichia coli*. *J. Bacteriol.* **1986**, *168*, 1120–1127.
139. Guillet, M.; Boiteux, S. Endogenous DNA abasic sites cause cell death in the absence of Apn1, Apn2 and Rad1/Rad10 in *Saccharomyces cerevisiae*. *EMBO J.* **2002**, *21*, 2833–2841. 10.1093/emboj/21.11.2833
140. Karumbati, A.S.; Deshpande, R.A.; Jilani, A.; Vance, J.R.; Ramotar, D.; Wilson, T.E. The role of yeast DNA 3'-phosphatase Tpp1 and Rad1/Rad10 endonuclease in processing spontaneous and induced base lesions. *J. Biol. Chem.* **2003**, *278*, 31434–31443. 10.1074/jbc.M304586200
141. Fryxell, K.J.; Zuckerkandl, E. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* **2000**, *17*, 1371–1383.
142. Hershberg, R.; Petrov, D.A. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.* **2010**, *6*, e1001115. 10.1371/journal.pgen.1001115
143. Ognjanovic, S.; Olivier, M.; Bergemann, T.L.; Hainaut, P. Sarcomas in *TP53* germline mutation carriers: A review of the IARC *TP53* database. *Cancer* **2012**, *118*, 1387–1396. 10.1002/cncr.26390
144. Warner, H.R.; Duncan, B.K.; Garrett, C.; Neuhaard, J. Synthesis and metabolism of uracil-containing deoxyribonucleic acid in *Escherichia coli*. *J. Bacteriol.* **1981**, *145*, 687–695.
145. Guillet, M.; Auffret Van Der Kemp, P.; Boiteux, S. dUTPase activity is critical to maintain genetic stability in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2006**, *34*, 2056–2066. 10.1093/nar/gkl139
146. Slesarev, A.I.; Mezhevaya, K.V.; Makarova, K.S.; Polushin, N.N.; Shcherbinina, O.V.; Shakhova, V.V.; Belova, G.I.; Aravind, L.; Natale, D.A.; Rogozin, I.B., et al. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 4644–4649. 10.1073/pnas.032671499
147. Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D., et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **1996**, *273*, 1058–1073. 10.1126/science.273.5278.1058
148. Hendrickson, E.L.; Kaul, R.; Zhou, Y.; Bovee, D.; Chapman, P.; Chung, J.; Conway de Macario, E.; Dodsworth, J.A.; Gillett, W.; Graham, D.E., et al. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J. Bacteriol.* **2004**, *186*, 6956–6969. 10.1128/JB.186.20.6956-6969.2004

149. Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K., et al. Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics. *J. Bacteriol.* **1997**, *179*, 7135-7155. 1263 1264 1265
150. Fricke, W.F.; Seedorf, H.; Henne, A.; Krüer, M.; Liesegang, H.; Hedderich, R.; Gottschalk, G.; Thauer, R.K. The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *J. Bacteriol.* **2006**, *188*, 642-658. 10.1128/JB.188.2.642-658.2006 1266 1267 1268
151. Schomacher, L.; Chong, J.P.J.; McDermott, P.; Kramer, W.; Fritz, H.-J. DNA uracil repair initiated by the archaeal ExoIII homologue Mth212 via direct strand incision. *Nucleic Acids Res.* **2009**, *37*, 2283-2293. 10.1093/nar/gkp102 1269 1270
152. Schomacher, L.; Schürer, K.A.; Ciirdaeva, E.; McDermott, P.; Chong, J.P.J.; Kramera, W.; Fritz, H.-J. Archaeal DNA uracil repair *via* direct strand incision: A minimal system reconstituted from purified components. *DNA Repair* **2010**, *9*, 438-447. 10.1016/j.dnarep.2010.01.004 1271 1272 1273
153. Lakomek, K.; Dickmanns, A.; Ciirdaeva, E.; Schomacher, L.; Ficner, R. Crystal structure analysis of DNA uridine endonuclease Mth212 bound to DNA. *J. Mol. Biol.* **2010**, *399*, 604-617. 10.1016/j.jmb.2010.04.044 1274 1275
154. Prorok, P.; Alili, D.; Saint-Pierre, C.; Gasparutto, D.; Zharkov, D.O.; Ishchenko, A.A.; Tudek, B.; Saparbaev, M. Uracil in duplex DNA is a substrate for the human nucleotide incision repair pathway. *Proc. Natl Acad. Sci. U.S.A.* **2013**, *110*, E3695-E3703. 10.1073/pnas.1305624110 1276 1277 1278
155. Shiraishi, M.; Ishino, S.; Yamagami, T.; Egashira, Y.; Kiyonari, S.; Ishino, Y. A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*. *Nucleic Acids Res.* **2015**, *43*, 2853-2863. 10.1093/nar/gkv121 1279 1280 1281
156. Shiraishi, M.; Iwai, S. Molecular basis of substrate recognition of endonuclease Q from the euryarchaeon *Pyrococcus furiosus*. *J. Bacteriol.* **2020**, *202*, e00542. 10.1128/JB.00542-19 1282 1283
157. Shiraishi, M.; Ishino, S.; Heffernan, M.; Cann, I.; Ishino, Y. The mesophilic archaeon *Methanosarcina acetivorans* counteracts uracil in DNA with multiple enzymes: EndoQ, ExoIII, and UDG. *Sci. Rep.* **2018**, *8*, 15791. 10.1038/s41598-018-34000-x 1284 1285 1286
158. Shiraishi, M.; Ishino, S.; Yoshida, K.; Yamagami, T.; Cann, I.; Ishino, Y. PCNA is involved in the EndoQ-mediated DNA repair process in *Thermococcales*. *Sci. Rep.* **2016**, *6*, 25532. 10.1038/srep25532 1287 1288
159. Shiraishi, M.; Ishino, S.; Cann, I.; Ishino, Y. A functional endonuclease Q exists in the bacterial domain: Identification and characterization of endonuclease Q from *Bacillus pumilus*. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 931-937. 10.1080/09168451.2016.1277946 1289 1290 1291
160. Miyazono, K.-i.; Ishino, S.; Makita, N.; Ito, T.; Ishino, Y.; Tanokura, M. Crystal structure of the novel lesion-specific endonuclease PfuEndoQ from *Pyrococcus furiosus*. *Nucleic Acids Res.* **2018**, *46*, 4807-4818. 10.1093/nar/gky261 1292 1293
161. Shi, K.; Moeller, N.H.; Banerjee, S.; McCann, J.L.; Carpenter, M.A.; Yin, L.; Moorthy, R.; Orellana, K.; Harki, D.A.; Harris, R.S., et al. Structural basis for recognition of distinct deaminated DNA lesions by endonuclease Q. *Proc. Natl Acad. Sci. U.S.A.* **2021**, *118*, e2021120118. 10.1073/pnas.2021120118 1294 1295 1296
162. Lahm, A.; Suck, D. DNase I-induced DNA conformation. 2 Å structure of a DNase I-octamer complex. *J. Mol. Biol.* **1991**, *222*, 645-667. 10.1016/0022-2836(91)90502-W 1297 1298
163. Weston, S.A.; Lahm, A.; Suck, D. X-ray structure of the DNase I-d(GGTATACC)₂ complex at 2.3 Å resolution. *J. Mol. Biol.* **1992**, *226*, 1237-1256. 10.1016/0022-2836(92)91064-V 1299 1300
164. Redrejo-Rodríguez, M.; Vigouroux, A.; Mursalimov, A.; Grin, I.; Alili, D.; Koshenov, Z.; Akishev, Z.; Maksimenko, A.; Bissenbaev, A.K.; Matkarimov, B.T., et al. Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. *Biochimie* **2016**, *128-129*, 20-33. 10.1016/j.biochi.2016.06.011 1301 1302 1303 1304

165. Mazouzi, A.; Vigouroux, A.; Aikeshchev, B.; Brooks, P.J.; Saparbaev, M.K.; Morera, S.; Ishchenko, A.A. Insight into mechanisms of 3'-5' exonuclease activity and removal of bulky 8,5'-cyclopurine adducts by apurinic/aprimidinic endonucleases. *Proc. Natl Acad. Sci. U.S.A.* **2013**, *110*, E3071-E3080. 10.1073/pnas.1305281110
166. Garcin, E.D.; Hosfield, D.J.; Desai, S.A.; Haas, B.J.; Björas, M.; Cunningham, R.P.; Tainer, J.A. DNA apurinic-aprimidinic site binding and excision by endonuclease IV. *Nat. Struct. Mol. Biol.* **2008**, *15*, 515-522. 10.1038/nsmb.1414
167. Gottesman, M.E.; Chudaev, M.; Mustaev, A. Key features of magnesium that underpin its role as the major ion for electrophilic biocatalysis. *FEBS J.* **2020**, *287*, 5439-5463. 10.1111/febs.15318
168. Ide, H.; Tedzuka, K.; Shimzu, H.; Kimura, Y.; Purmal, A.A.; Wallace, S.S.; Kow, Y.W. α -Deoxyadenosine, a major anoxic radiolysis product of adenine in DNA, is a substrate for *Escherichia coli* endonuclease IV. *Biochemistry* **1994**, *33*, 7842-7847. 10.1021/bi00191a011
169. Gros, L.; Ishchenko, A.A.; Ide, H.; Elder, R.H.; Saparbaev, M.K. The major human AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway. *Nucleic Acids Res.* **2004**, *32*, 73-81. 10.1093/nar/gkh165
170. Bowman, K.K.; Sidik, K.; Smith, C.A.; Taylor, J.-S.; Doetsch, P.W.; Freyer, G.A. A new ATP-independent DNA endonuclease from *Schizosaccharomyces pombe* that recognizes cyclobutane pyrimidine dimers and 6-4 photoproducts. *Nucleic Acids Res.* **1994**, *22*, 3026-3032. 10.1093/nar/22.15.3026
171. Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A. A eukaryotic gene encoding an endonuclease that specifically repairs DNA damaged by ultraviolet light. *EMBO J.* **1995**, *14*, 2393-2399.
172. Avery, A.M.; Kaur, B.; Taylor, J.-S.; Mello, J.A.; Essigmann, J.M.; Doetsch, P.W. Substrate specificity of ultraviolet DNA endonuclease (UVDE/Uve1p) from *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **1999**, *27*, 2256-2264. 10.1093/nar/27.11.2256
173. Kanno, S.-i.; Shigenori, I.; Takao, M.; Yasui, A. Repair of apurinic/aprimidinic sites by UV damage endonuclease; a repair protein for UV and oxidative damage. *Nucleic Acids Res.* **1999**, *27*, 3096-3103. 10.1093/nar/27.15.3096
174. Paspaleva, K.; Thomassen, E.; Pannu, N.S.; Iwai, S.; Moolenaar, G.F.; Goosen, N.; Abrahams, J.P. Crystal structure of the DNA repair enzyme ultraviolet damage endonuclease. *Structure* **2007**, *15*, 1316-1324. 10.1016/j.str.2007.05.010
175. Meulenbroek, E.M.; Peron Cane, C.; Jala, I.; Iwai, S.; Moolenaar, G.F.; Goosen, N.; Pannu, N.S. UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize different DNA lesions. *Nucleic Acids Res.* **2013**, *41*, 1363-1371. 10.1093/nar/gks1127
176. Nick McElhinny, S.A.; Kumar, D.; Clark, A.B.; Watt, D.L.; Watts, B.E.; Lundström, E.-B.; Johansson, E.; Chabes, A.; Kunkel, T.A. Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **2010**, *6*, 774-781. 10.1038/nchembio.424
177. Reijns, M.A.M.; Rabe, B.; Rigby, R.E.; Mill, P.; Astell, K.R.; Lettice, L.A.; Boyle, S.; Leitch, A.; Keighren, M.; Kilanowski, F., et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **2012**, *149*, 1008-1022. 10.1016/j.cell.2012.04.011
178. Eder, P.S.; Walder, R.Y.; Walder, J.A. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* **1993**, *75*, 123-126. 10.1016/0300-9084(93)90033-O
179. Rydberg, B.; Game, J. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 16654-16659. 10.1073/pnas.262591699
180. Sparks, J.L.; Chon, H.; Cerritelli, S.M.; Kunkel, T.A.; Johansson, E.; Crouch, R.J.; Burgers, P.M. RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **2012**, *47*, 980-986. 10.1016/j.molcel.2012.06.035
181. Rice, P.; Craigie, R.; Davies, D.R. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **1996**, *6*, 76-83. 10.1016/S0959-440X(96)80098-4

182. Hyjek, M.; Figiel, M.; Nowotny, M. RNases H: Structure and mechanism. *DNA Repair* **2019**, *84*, 102672. 10.1016/j.dnarep.2019.102672 1346
1347
183. Tadokoro, T.; Kanaya, S. Ribonuclease H: Molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.* **2009**, *276*, 1482-1493. 10.1111/j.1742-4658.2009.06907.x 1348
1349
184. Gates, F.T., III; Linn, S. Endonuclease V of *Escherichia coli*. *J. Biol. Chem.* **1977**, *252*, 1647-1653. 10.1016/S0021-9258(17)40598-9 1350
1351
185. Yao, M.; Hatahet, Z.; Melamed, R.J.; Kow, Y.W. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from *Escherichia coli*. *J. Biol. Chem.* **1994**, *269*, 16260-16268. 1352
1353
186. Yao, M.; Kow, Y.W. Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from *Escherichia coli*. *J. Biol. Chem.* **1996**, *271*, 30672-30676. 10.1074/jbc.271.48.30672 1354
1355
187. Yao, M.; Kow, Y.W. Further characterization of *Escherichia coli* endonuclease V: Mechanism of recognition for deoxyinosine, deoxyuridine, and base mismatches in DNA. *J. Biol. Chem.* **1997**, *272*, 30774-30779. 10.1074/jbc.272.49.30774 1356
1357
188. Huang, J.; Lu, J.; Barany, F.; Cao, W. Multiple cleavage activities of endonuclease V from *Thermotoga maritima*: Recognition and strand nicking mechanism. *Biochemistry* **2001**, *40*, 8738-8748. 10.1021/bi010183h 1358
1359
189. Kiyonari, S.; Egashira, Y.; Ishino, S.; Ishino, Y. Biochemical characterization of endonuclease V from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J. Biochem.* **2014**, *155*, 325-333. 10.1093/jb/mvu010 1360
1361
190. Ishino, S.; Makita, N.; Shiraishi, M.; Yamagami, T.; Ishino, Y. EndoQ and EndoV work individually for damaged DNA base repair in *Pyrococcus furiosus*. *Biochimie* **2015**, *118*, 264-269. 10.1016/j.biochi.2015.06.015 1362
1363
191. Schouten, K.A.; Weiss, B. Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. *Mutat. Res.* **1999**, *435*, 245-254. 10.1016/S0921-8777(99)00049-X 1364
1365
192. Weiss, B. Endonuclease V of *Escherichia coli* prevents mutations from nitrosative deamination during nitrate/nitrite respiration. *Mutat. Res.* **2001**, *461*, 301-309. 10.1016/S0921-8777(00)00062-8 1366
1367
193. Pang, B.; McFaline, J.L.; Burgis, N.E.; Dong, M.; Taghizadeh, K.; Sullivan, M.R.; Elmquist, C.E.; Cunningham, R.P.; Dedon, P.C. Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. *Proc. Natl Acad. Sci. U.S.A.* **2012**, *109*, 2319-2324. 10.1073/pnas.1118455109 1368
1369
1370
194. Saparbaev, M.; Laval, J. Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. U.S.A.* **1994**, *91*, 5873-5877. 1371
1372
195. O'Brien, P.J.; Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **2004**, *279*, 9750-9757. 10.1074/jbc.M312232200 1373
1374
196. Moe, A.; Ringvoll, J.; Nordstrand, L.M.; Eide, L.; Bjørås, M.; Seeberg, E.; Rognes, T.; Klungland, A. Incision at hypoxanthine residues in DNA by a mammalian homologue of the *Escherichia coli* antimutator enzyme endonuclease V. *Nucleic Acids Res.* **2003**, *31*, 3893-3900. 10.1093/nar/gkg472 1375
1376
1377
197. Fladeby, C.; Vik, E.S.; Laerdahl, J.K.; Neurauter, C.G.; Heggelund, J.E.; Thorgaard, E.; Strøm-Andersen, P.; Bjørås, M.; Dalhus, B.; Alseth, I. The human homolog of *Escherichia coli* endonuclease V is a nucleolar protein with affinity for branched DNA structures. *PLoS ONE* **2012**, *7*, e47466. 10.1371/journal.pone.0047466 1378
1379
1380
198. Mi, R.; Alford-Zappala, M.; Kow, Y.W.; Cunningham, R.P.; Cao, W. Human endonuclease V as a repair enzyme for DNA deamination. *Mutat. Res.* **2012**, *735*, 12-18. 10.1016/j.mrfmmm.2012.05.003 1381
1382
199. Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I. Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nat. Commun.* **2013**, *4*, 2273. 10.1038/ncomms3273 1383
1384
200. Vik, E.S.; Nawaz, M.S.; Strøm Andersen, P.; Fladeby, C.; Bjørås, M.; Dalhus, B.; Alseth, I. Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* **2013**, *4*, 2271. 10.1038/ncomms3271 1385
1386

201. Dalhus, B.; Alseth, I.; Bjørås, M. Structural basis for incision at deaminated adenines in DNA and RNA by endonuclease V. *Prog. Biophys. Mol. Biol.* **2015**, *117*, 134–142. 10.1016/j.pbiomolbio.2015.03.005 1387–1388
202. Dizdaroglu, M.; Laval, J.; Boiteux, S. Substrate specificity of the *Escherichia coli* endonuclease III: Excision of thymine- and cytosine-derived lesions in DNA produced by radiation-generated free radicals. *Biochemistry* **1993**, *32*, 12105–12111. 10.1021/bi00096a022 1389–1391
203. Jurado, J.; Saparbaev, M.; Matray, T.J.; Greenberg, M.M.; Laval, J. The ring fragmentation product of thymidine C5-hydrate when present in DNA is repaired by the *Escherichia coli* Fpg and Nth proteins. *Biochemistry* **1998**, *37*, 7757–7763. 10.1021/bi972982z 1392–1394
204. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J., et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. 10.1038/msb.2011.75 1395–1397
205. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. 10.1093/nar/gkw290 1398–1399
206. Daniels, D.S.; Woo, T.T.; Luu, K.X.; Noll, D.M.; Clarke, N.D.; Pegg, A.E.; Tainer, J.A. DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **2004**, *11*, 714–720. 10.1038/nsmb791 1400–1401
207. Aas, P.A.; Otterlei, M.; Falnes, P.Ø.; Vågbø, C.B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjørås, M.; Slupphaug, G.; Seeberg, E., et al. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **2003**, *421*, 859–863. 10.1038/nature01363 1402–1404
208. Falnes, P.Ø.; Rognes, T. DNA repair by bacterial AlkB proteins. *Res. Microbiol.* **2003**, *154*, 531–538. 10.1016/S0923-2508(03)00150-5 1405–1406
209. Sedgwick, B. Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 148–157. 10.1038/nrm1312 1407
210. Treweek, S.C.; Henshaw, T.F.; Hausinger, R.P.; Lindahl, T.; Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **2002**, *419*, 174–178. 10.1038/nature00908 1408–1409
211. Falnes, P.Ø.; Johansen, R.F.; Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **2002**, *419*, 178–182. 10.1038/nature01048 1410–1411
212. Leiros, I.; Nabong, M.P.; Grøsvik, K.; Ringvoll, J.; Haugland, G.T.; Uldal, L.; Reite, K.; Olsbu, I.K.; Knævelsrud, I.; Moe, E., et al. Structural basis for enzymatic excision of N¹-methyladenine and N³-methylcytosine from DNA. *EMBO J.* **2007**, *26*, 2206–2217. 10.1038/sj.emboj.7601662 1412–1414
213. Eisen, J.A.; Hanawalt, P.C. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* **1999**, *435*, 171–213. 10.1016/S0921-8777(99)00050-6 1415–1416
214. Schul, W.; Jans, J.; Rijksen, Y.M.A.; Klemann, K.H.M.; Eker, A.P.M.; de Wit, J.; Nikaido, O.; Nakajima, S.; Yasui, A.; Hoeijmakers, J.H.J., et al. Enhanced repair of cyclobutane pyrimidine dimers and improved UV resistance in photolyase transgenic mice. *EMBO J.* **2002**, *21*, 4719–4729. 10.1093/emboj/cdf456 1417–1419
215. Sancar, A. Structure and function of photolyase and *in vivo* enzymology: 50th anniversary. *J. Biol. Chem.* **2008**, *283*, 32153–32157. 10.1074/jbc.R800052200 1420–1421
216. Ahmad, M.; Jarillo, J.A.; Smirnova, O.; Cashmore, A.R. Cryptochrome blue-light photoreceptors of *Arabidopsis* implicated in phototropism. *Nature* **1998**, *392*, 720–723. 10.1038/33701 1422–1423
217. Panda, S.; Hogenesch, J.B.; Kay, S.A. Circadian rhythms from flies to human. *Nature* **2002**, *417*, 329–335. 10.1038/417329a 1424
218. Park, H.-W.; Kim, S.-T.; Sancar, A.; Deisenhofer, J. Crystal structure of DNA photolyase from *Escherichia coli*. *Science* **1995**, *268*, 1866–1872. 10.1126/science.7604260 1425–1426
219. Tamada, T.; Kitadokoro, K.; Higuchi, Y.; Inaka, K.; Yasui, A.; de Ruiter, P.E.; Eker, A.P.M.; Miki, K. Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat. Struct. Biol.* **1997**, *4*, 887–891. 10.1038/nsb1197-887 1427–1428

220. Eker, A.P.M.; Dekker, R.H.; Berends, W. Photoreactivating enzyme from *Streptomyces griseus*—IV. On the nature of the chromophoric cofactor in *Streptomyces griseus* photoreactivating enzyme. *Photochem. Photobiol.* **1981**, *33*, 65-72. 10.1111/j.1751-1097.1981.tb04298.x
221. Eker, A.P.M.; Kooiman, P.; Hessels, J.K.C.; Yasui, A. DNA photoreactivating enzyme from the cyanobacterium *Anacystis nidulans*. *J. Biol. Chem.* **1990**, *265*, 8009-8015. 10.1016/S0021-9258(19)39031-3
222. Johnson, J.L.; Hamm-Alvarez, S.; Payne, G.; Sancar, G.B.; Rajagopalan, K.V.; Sancar, A. Identification of the second chromophore of *Escherichia coli* and yeast DNA photolyases as 5,10-methenyltetrahydrofolate. *Proc. Natl Acad. Sci. U.S.A.* **1988**, *85*, 2046-2050. 10.1073/pnas.85.7.2046
223. Li, Y.F.; Sancar, A. Cloning, sequencing, expression and characterization of DNA photolyase from *Salmonella typhimurium*. *Nucleic Acids Res.* **1991**, *19*, 4885-4890. 10.1093/nar/19.18.4885
224. Aravind, L.; Anantharaman, V.; Koonin, E.V. Monophyly of class I aminoacyl tRNA synthetase, USPA, ETPP, photolyase, and PP-ATPase nucleotide-binding domains: Implications for protein evolution in the RNA world. *Proteins* **2002**, *48*, 1-14. 10.1002/prot.10064
225. Vechtomova, Y.L.; Telegina, T.A.; Kritsky, M.S. Evolution of proteins of the DNA photolyase/cryptochrome family. *Biochemistry (Mosc)* **2020**, *85*, S131-S153. 10.1134/S0006297920140072
226. Oberpichler, I.; Pierik, A.J.; Wesslowski, J.; Pokorny, R.; Rosen, R.; Vugman, M.; Zhang, F.; Neubauer, O.; Ron, E.Z.; Batschauer, A., et al. A photolyase-like protein from *Agrobacterium tumefaciens* with an iron-sulfur cluster. *PLoS ONE* **2011**, *6*, e26775. 10.1371/journal.pone.0026775
227. Setlow, P.; Li, L. Photochemistry and photobiology of the spore photoproduct: A 50-year journey. *Photochem. Photobiol.* **2015**, *91*, 1263-1290. 10.1111/php.12506
228. Berteau, O.; Benjdia, A. DNA repair by the radical SAM enzyme spore photoproduct lyase: From biochemistry to structural investigations. *Photochem. Photobiol.* **2017**, *93*, 67-77. 10.1111/php.12702
229. Krishnan, A.; Burroughs, A.M.; Iyer, L.M.; Aravind, L. Unexpected evolution of lesion-recognition modules in eukaryotic NER and kinetoplast DNA dynamics proteins from bacterial mobile elements. *iScience* **2018**, *9*, 192-208. 10.1016/j.isci.2018.10.017
230. Geourjon, C.; Orelle, C.; Steinfels, E.; Blanchet, C.; Deléage, G.; Di Pietro, A.; Jault, J.-M. A common mechanism for ATP hydrolysis in ABC transporter and helicase superfamilies. *Trends Biochem. Sci.* **2001**, *26*, 539-544. 10.1016/S0968-0004(01)01907-7
231. White, M.F.; Allers, T. DNA repair in the archaea—an emerging picture. *FEMS Microbiol. Rev.* **2018**, *42*, 514-526. 10.1093/femsre/fuy020
232. Matelska, D.; Steczkiewicz, K.; Ginalska, K. Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Res.* **2017**, *45*, 6995-7020. 10.1093/nar/gkx494
233. Senissar, M.; Manav, M.C.; Brodersen, D.E. Structural conservation of the PIN domain active site across all domains of life. *Protein Sci.* **2017**, *26*, 1474-1492. 10.1002/pro.3193
234. Bujnicki, J.M. Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of protein structures. *J. Mol. Evol.* **2000**, *50*, 39-44. 10.1007/s002399910005
235. Kinch, L.N.; Ginalska, K.; Rychlewski, L.; Grishin, N.V. Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Res.* **2005**, *33*, 3598-3605. 10.1093/nar/gki676
236. Karakas, E.; Truglio, J.J.; Croteau, D.; Rhau, B.; Wang, L.; Van Houten, B.; Kisker, C. Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *EMBO J.* **2007**, *26*, 613-622. 10.1038/sj.emboj.7601497

-
237. Dunin-Horkawicz, S.; Feder, M.; Bujnicki, J.M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* **2006**, *7*, 98. 10.1186/1471-2164-7-98 1470
1471
238. Furuya, K.; Hutchinson, C.R. The DrrC protein of *Streptomyces peucetius*, a UvrA-like protein, is a DNA-binding protein 1472
whose gene is induced by daunorubicin. *FEMS Microbiol. Lett.* **1998**, *168*, 243-249. 10.1111/j.1574-6968.1998.tb13280.x 1473
239. Moolenaar, G.F.; van Rossum-Fikkert, S.; van Kesteren, M.; Goosen, N. Cho, a second endonuclease involved in 1474
Escherichia coli nucleotide excision repair. *Proc. Natl Acad. Sci. U.S.A.* **2002**, *99*, 1467-1472. 10.1073/pnas.032584099 1475
240. Vidya Perera, A.; Mendenhall, J.B.; Tan Courcelle, C.; Courcelle, J. Cho endonuclease functions during DNA interstrand 1476
cross-link repair in *Escherichia coli*. *J. Bacteriol.* **2016**, *198*, 3099-3108. 10.1128/JB.00509-16 1477