Structure, Function and Evolution of the Helix-hairpin-Helix DNA Glycosylase Superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms

C.H. Trasviña-Arenas*, Merve Demir, Wen-Jen Lin, and Sheila S. David*

Department of Chemistry, University of California, Davis, California 95616

**Corresponding authors: Department of Chemistry, University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA. E-mail: ssdavid@ucdavis.edu (S.S. David) and trasvina@ucdavis.edu (C.H. Trasviña-Arenas).*

Abstract

The Base Excision Repair (BER) pathway is a highly conserved DNA repair system targeting subtle chemical base modifications that arise from oxidation, deamination and alkylation reactions. BER features lesion-specific DNA glycosylases (DGs) which recognize and excise modified or inappropriate DNA bases to produce apurinic/apyrimidinic (AP) sites and coordinate AP-site hand-off to subsequent BER pathway enzymes. The DG superfamilies identified have evolved independently to cope with a wide variety of nucleobase chemical modifications. Most DG superfamilies recognize a distinct set of structurally related lesions. In contrast, the Helix-hairpin-Helix (HhH) DG superfamily has the remarkable ability to act upon structurally diverse sets of base modifications. The versatility in substrate recognition of the HhH-DG superfamily has been shaped by motif and domain acquisitions during evolution. In this paper, we review the structural features and catalytic mechanisms of the HhH-DG superfamily and draw a hypothetical reconstruction of the evolutionary path where these DGs developed diverse and unique enzymatic features.

1. Introduction to the HhH DNA glycosylase superfamily: A versatile platform for DNA repair

Base Excision Repair (BER) is a highly conserved DNA repair pathway that is tasked with the repair of a wide array of DNA nucleobase lesions and mismatches [1]. BER relies on a crew of lesion-specific DNA glycosylases (DGs) that recognize a particular type of DNA base lesion among a large excess of canonical DNA and hydrolyze the N-glycosidic linkage to create an apurinic/apyrimidinic (AP) site. The subsequent sequential activity of other BER enzymes including, but not limited to, an AP endonuclease, DNA repair polymerase and DNA ligase mediate full repair of the original damaged site. As DNA damage-specific enzymes, DGs are the "adapters" of BER, recognizing a wide variety of structurally diverse DNA base lesions and channeling its AP product through a conserved series of downstream BER reactions. To cope with the structural diversity of DNA base lesions, different DG superfamilies with specific substrate preferences have evolved. Several DG superfamilies have been identified: Uracil-DNA Glycosylase (UDG) [2], Heat-Like Repeat (HLR) [3], Fpg/Nei [4], Methyl-Purine DNA glycosylase (MPG) [5] and Helix-Hairpin-Helix (HhH) [6] superfamilies.

The HhH superfamily is one of the most versatile of the DG superfamilies in terms of the diversity of lesions and mismatches it targets and its biological functions [6] (**Figure 1**). The HhH-DG superfamily includes the MutY/MIG, EndoIII/Nth, OGG1/OGG2/AGOG, TAG (also referred to as 3-methyladenine DNA glycosylase I), AlkA (also known as MagII), MagI, MagIII, MBD4 and DEMETER DNA glycosylase families with broad presence in Bacteria, Archaea and Eukaryotes (**Figure 1B**). The sequence similarities and monophyletic clustering suggest that the HhH-DGs are products of a common ancestor that diversified to alter substrate specificity and biological function (**Figure 1A**). Indeed, HhH-DGs mediate repair of oxidized purine and pyrimidines (OGG1/EndoIII/MutY), alkylated purines (TAG/AlkA/MagI/MagIII), uracil-containing DNA (MBD4/MIG) and base mispairs (MutY/MBD4/MIG) (**Figure 1C**). Evolution has shaped the HhH-DG superfamily to provide for unique substrate preferences and has exploited the versatility of this superfamily to perform functions beyond BER in diverse processes such as gene regulation and cell cycle control [7-10].

The defining features of the HhH-DG superfamily are a highly conserved HhH motif followed by a Gly/Pro-rich loop and a conserved aspartic acid residue (HhH–GPD motif) [11-14]. These conserved motifs are part of the catalytic domain of the enzyme that is generally composed by two α-helical (αH) barrel subdomains bridged through the HhH motif (**Figure 2**). Despite the differences in substrate preference and amino acid sequence, the folding of the catalytic domain of HhH-DGs is highly conserved [13, 15-17]. In addition, despite structurally similar catalytic domains, evolution has fine-tuned the catalytic strategies used by the different superfamily enzymes to remove different DNA lesions or mismatched bases. HhH-DGs have also acquired domains and motifs to alter substrate recognition and biological activity (see Section 2). The gain of motifs or domains might have occurred by illegitimate DNA repair (homologous or nonhomologous recombination) or by means of mobile genetic elements such as DNA transposons or retrotransposons [18-20].

Herein, we review the structural features and catalytic mechanisms of the HhH-DG superfamily and draw a hypothetical reconstruction of the evolutionary path whereby these DGs developed the ability to target diverse DNA base lesions. The information available on plant 5-methylcytosine (5mC) DEMETER DGs is presently too limited for comparison to other DGs [21-24]. In addition, we discuss OGG1 briefly since the OGG1/OGG2/AGOG family has been recently reviewed in detail [25]. We focus on MutY/MUTYH glycosylases leveraging our expertise [1, 26, 27] to draw parallels in recognition and base excision mechanisms to other HhH-DGs.

2. Biological and structural insights of HhH-DGs.

2.1. MutY/MIG family

MutY and MIG are structurally and phylogenetically related monofunctional DGs sharing up to 30% amino acid identity (**Figure 1A** and **2C**) [28]. MutY is a ubiquitous adenine DNA glycosylase present in eukaryotes, bacteria, and archaea [6, 29]. MutY enzymes have been shown to remove adenine mispaired to 8-oxo-7,8-dihydroguanine (OG), G or C *in vitro* [30-33]. Notably, however, mismatch affinity and glycosylase activity of MutY enzymes is greatest for OG:A substrates [32, 34], strongly supporting the OG:A mismatch as the biologically relevant substrate. Moreover, cellular assays and mutation frequency data strongly support the role of MutY as a highly specialized glycosylase that initiates repair of OG:A mispairs [32]. The lack of the *mutY*

gene is associated with increased G:C to T:A transversion mutations due to "T"-like coding propensity of OG [35]. Moreover, mutations frequencies in bacteria are significantly higher in the absence of OG glycosylase MutM (aka Fpg), indicating that both glycosylases act on the same lesion [36, 37]. Indeed, G:C to T:A transversions form the genomic signature of MUTYH-associated polyposis (MAP), a human genetic disorder characterized by defects in the *MUTYH* gene. Impairment of MUTYH increases the frequency of somatic mutations in colorectal cells and subsequent inactivation of the tumor suppressor gene *APC*, leading to multiple colorectal adenomas and carcinoma [26, 38].

In contrast to MutY, MIG is a mismatch-specific DNA glycosylase with substrate preference for uracil and thymine mispaired with G that arise from hydrolytic deamination of cytosine and 5-mC, respectively [39]. Failure to repair deaminated cytosine and 5-mC leads to C:G to T:A mutations upon replication [40]. Interestingly, MIG has been identified in all domains of life, but most often in hyperthermophilic microorganisms (**Figure 1B**). MIG along with other uracil-DNA repair enzymes [41] expand archaeal uracil DNA repair activities. The high frequency of cytosine and 5mC deamination at high temperatures and the inability of hyperthermophilic archaeal DNA polymerases to bypass uracil may have provided the evolutionary pressure to evolve multiple uracil DNA repair enzymes in these organisms [42-44].

Sequence and structural alignment (1.91 Å RMSD) between MIG and the catalytic core of *Geobacillus stearothermophilus* (*Gs*) MutY along with their phylogenetic distribution (**Figure 1A** and **2C**) indicate a close evolutionary origin. Both enzymes have the same structural folding of the two helical subdomains plus an $[4Fe\text{-}4S]^{2+}$ cluster coordinated by four highly conserved cysteines within the helical subdomain 2 (**Figure 2**). Our laboratory has extensively evaluated the functional role of the $[4Fe\text{-}4S]^{2+}$ cluster in MutY [26]. We have established the importance of DNA phosphodiester backbone contacts through positively-charged residues harbored in the $[4Fe\text{-}4S]^{2+}$ cluster loop (FCL) in substrate recognition [45]. Site-directed mutagenesis of the $[4Fe\text{-}4S]^{2+}$ cluster cysteinyl ligands and surrounding residues typically compromises protein stability and activity [45-48]. However, the impact on activity and stability is highly dependent on the type and position of the substitution. For instance, mutation of one of the cysteinyl ligand to His (Cys 199 in *E. coli* MutY) results in biochemical and *in vivo* repair activity similar to WT [48]. Importantly,

the [4Fe-4S]$^{2+}$ cluster in MutY has been shown to be more readily oxidized in the presence of DNA [49-51]. The DNA dependent redox cycling has been proposed to facilitate DNA lesion localization via long-distance charge transport with other cluster containing enzymes [49, 50, 52]. Similar DNA dependent redox cycling has been observed in other DGs [49, 53], DNA repair [54, 55] and replicative enzymes [56, 57], suggesting that the [4Fe-4S]$^{2+}$ cluster is an important functional and structural element in DNA transactions.

Our laboratory also identified a novel "Zinc Linchpin motif" comprised within the interdomain connector (IDC) region and N-terminal domain of mammalian MUTYH enzymes and have shown the importance of the coordinated Zn(II) for lesion substrate engagement required for the adenine glycosylase activity and the ability to suppress DNA mutations [58, 59]. The IDC in bacterial and mammalian MutY/MUTYH DGs connects the catalytic domain and OG recognition domain (**Figure 2** and **3**). Our lab identified that the Zn(II) in the Zn linchpin motif in MUTYH is coordinated by three Cys residues in the IDC, and by a proposed fourth Cys ligand that resides near the [4Fe-4S]$^{2+}$cluster [58, 59]. However, mouse MUTYH structure recently solved shows that the Zn(II) coordination sphere involves three Cys residues and one His [60]. Notably, the region coordination the Zn(II) was highly disordered suggesting potential flexibility in this region, and perhaps the ability to adopt alternative ligand coordination. This flexibility of the IDC and Zinc linchpin motif may also play a role in its function as a protein-protein interaction scaffold. Indeed, this region has been shown to be the locus for interactions with APE1, SIRT6 and Hus1, that play roles in BER, DNA damage response and cell cycle control [61, 62].

MutY is a peculiar DG with its high specificity for OG:A mismatches and removal of the undamaged adenine within the mispair. This substrate duality requires lesion base pair recognition to be tightly coupled with base excision. Such challenges were solved evolutionarily through the acquisition of a specific OG-recognition or MutT-like domain to the HhH-DG scaffold. MutT is a d(OG)TPase in charge of removing the oxidized dGTP from the cellular dNTP pool [35]. Several structural studies have shown that the recognition of the OG moiety in MutT and MutT-like domain in MutY are carried out at different regions. MutT recognizes and hydrolyzes d(OG)TP in a catalytic pocket centered at the core of the MutT structure [63]. On the other hand, MutT-like domain of MutY recognizes OG using a solvent accessible "FSH" loop (**Figure 2 and 3**), that

has been shown to mediate specific recognition of OG over G [64]. Additionally, histidine (H309 in Gs MutY) appears to be the primary "sensor" residue responsible for interhelical detection of OG:A lesion bps based on cellular repair and single molecule assays [31].

The modularity associated with the gain of the MutT-like domain during MutY evolution provides several advantages. The two domains coordinate to interrogate the DNA helix for hidden OG:A mispairs; the MutT-like domain searches for OG and the catalytic core (HhH scaffold) for adenine. Additional contacts of the catalytic core with OG serve to confirm OG identity once located and facilitate proper positioning of A to support base excision catalysis [32, 33] (**Figure 3**). Unlike other HhH-DGs, MutY engages both DNA strands to enhance specificity and activity [15]. These features are reflected in the observed ~30-fold more efficient removal of A from OG:A over G:A mismatches by MutY *in vitro* [32, 65]. In addition, cellular repair relies heavily on the presence of OG, with minimal observed repair of G:A mismatches [66]. Structure-Activity Relationships (SAR) studies using A and OG analogs reveal the importance of the unique structure of OG_{syn}:A_{anti} base-pairs for efficient recognition and repair by MutY [32, 33] (**Figure 3A)**. Indeed, the unique position of the 2-amino group of OG in the major groove in OG:A mismatches provides for its detection by the His residue within the FSH loop of MutY.

2.2. EndoIII/Nth family.

In terms of structure and phylogenetic distribution, the EndoIII/Nth family is closely related to the MutY/MIG family. The EndoIII/Nth family is also ubiquitous across the entire tree of life. Furthermore, like MutY, EndoIII/Nth family also contains a $[4Fe\text{-}4S]^{2+}$ cluster and the associated FCL motif [13, 67]. Functional studies have demonstrated that the $[4Fe\text{-}4S]^{2+}$ cluster in EndoIII has similar functions as that in MutY [68-71]. In contrast to MutY/MIG family, the EndoIII/Nth family has a broad substrate scope and has been shown to remove a variety of oxidized pyrimidines (**Figure 1C**). For example, EndoIII/Nth DGs remove C oxidation product 5-hydroxycytosine, and its deamination products, 5-hydroxyuracil, and uracil glycol that ultimately result in C:G to T:A transition mutations. EndoIII/Nth DGs also remove the T oxidation product thymine glycol that are potent blockades in DNA replication [72, 73]. Recently, inherited mutations in the human EndoIII gene, *NTHL1*, have been correlated with a colorectal cancer

predisposition mechanism referred to as NTHL1-associated polyposis (NAP). Cancer cells expressing NTHL1 variants have been shown to exhibit a unique mutational signature (signature 30, characterized by C to T transitions) and increased DNA double-stranded breaks resulting in genomic instability [74, 75].

2.3. MBD4 family

MBD4, like MutY, is a modular HhH-DG, that harbors an N-terminal methyl-CpG binding (MBD) domain in addition to a C-terminal catalytic glycosylase domain (**Figure 2A**). Unlike MutY, MBD4 is found only in eukaryotes [76-78]. MBD4 removes a variety of base modifications such as halogenated pyrimidines [79, 80], demethylation intermediates such as 5-hydroxymethyluracil [81, 82], and thymine/uracil mispaired with G or O6-methylguanine (**Figure 1C**) [83-85]. There have been several individual structures of the MBD and DG domain of MBD4 in complex with DNA [82, 86, 87]. The MBD consists of one α-helix and three β-sheet components that recognizes 5mC through an Arginine finger motif [87]. The DG domain of MBD4 has the characteristic folding of HhH-DGs [87], sharing 17% amino acid identity with the catalytic HhH motif of MUTYH; the two structures superimpose with an RMSD of 1.02 Å for 55 Cα atoms (**Figure 2C**).

The presence of an MBD in MBD4, along with information from biochemical and cellular studies, suggests that MBD4 targets its glycosylase activity toward damaged DNA on genomic regions with high GC content (so-called CpG sites) [88, 89]. CpG sites are abundant at regulatory regions in eukaryotic genomes and are subject to methylation as part of epigenetic gene regulation. The high abundance of cytosine and its methylation to 5mC make CpG sites prone to hydrolytic deamination, causing the emergence of uracil and thymine mismatches, leading to C:G to T:A transition mutations [90, 91]. The sensitivity of these regulatory regions likely provided the evolutionary pressure to evolve MBD4 as CpG site-damage specific DGs in eukaryotes. Similar to MUTYH and NTHL1, impairment of MBD4 function has been associated with the etiology of cancer [92-95].

2.4 OGG1 subfamily

The 8-oxoguanine HhH-DG family present in archaea, bacteria and eukaryotes includes OGG1, OGG2 and AGOG subfamilies [25, 96]. OGG1 is the only subfamily with ubiquitous

presence, while OGG2 (archaea/bacteria) and AGOG (archaea) have a more restricted distribution. OGG1 has a restricted substrate scope, with high specificity for OG and FapyG lesions. These DNA lesions occur by ROS-mediated guanine oxidation and are miscoding lesions leading to G:C to T:A transversions [97]. Deficiency of mouse MUTYH and OGG1 leads to increased OG-associated mutations [98]. Reduced activity of human OGG1 has been linked to the development of several pathologies including cancer [98], metabolic dysfunction [99, 100], Alzheimer's disease [101, 102] and dysregulation of inflammatory responses [103, 104].

OGG1 and the alkyl-purine HhH-DG, AlkA (see **section 3.5**), share 23% amino acid identity and superimpose with an RMSD of 3.81 Å for 112 Cα atoms. Both enzymes contain an antiparallel β-sheet domain that resembles the TATA-binding protein (TBP) and therefore is referred to as TBP-like domain [15]. This subdomain appears as an N-terminal extension of the HhH-DG domain and does not directly contact the DNA (**Figure 2A** and **C**) making it unlikely to participate in the DG activity. More likely the TBP-like domain mediates protein-protein interactions with downstream BER enzymes [105], other DNA repair systems [106-109] and cell cycle control machinery [110]. This notion is supported by the fact that Poly [ADP-ribose] polymerase 1 (PARP1), a sensor and repair modulator of single-stranded DNA breaks, physically interacts with the TBP-like region of OGG1 [111].

2.5. Alkyl-purine HhH DNA glycosylases

Evolution has armed life with several repair mechanisms to protect cells from alkylation DNA damage [112]. Alkylated DNA bases are produced by exposure to alkylating agents from the environment [113] and also have endogenous origins such as *in situ* nitrosation [114]. The wide variety of alkylated bases that may be produced results in different consequences on genomic integrity. For example, alkylated bases such as 2-methylthymine, 2-methylcytosine and O6-methylguanine are mutagenic while 3-methyladenine blocks DNA replication [112, 115]. There are four alkyl-purine DNA glycosylases within the HhH-DG superfamily: TAG, AlkA, MagI and MagIII. While TAG, MagI and MagIII are specialized to remove 7-methylguanine and 3-methyladenine, AlkA is an extraordinary DNA glycosylase that evolved an expansion on its substrate preference, being able to remove alkylated pyrimidines such as 2-O-methylcytosine

and 2-O-methylthymine and hypoxanthine [5]. Moreover, AlkA has been reported to be able to remove normal bases, with its overexpression inducing mutagenesis in *E. coli* [116] (**Figure 1C**). These HhH-DGs have variations in their taxonomic distribution. For instance, AlkA is a DG exclusive to bacteria, TAG has presence in Bacteria and some eukaryotic organisms, while MagI and MagIII are present in all three domains of life.

Most of the HhH-DGs contain additional motifs, such as the [4Fe-4S]$^{2+}$ cluster and FCL motif in MutY/EndoIII, and additional domains, such as the MBD in MBD4. In this sense, TAG and MagI/MagIII are the most austere HhH DGs since they contains only one α-helical subdomain and two two α-helical subdomains, respectively, in addition to the HhH domain [117-119] (**Figure 2C)**. In addition, AlkA stands-out among alkyl-purine HhH-DGs in harboring a TBP-like domain at its N-terminal region, similar to OGG1 [11, 120]. The biological function of TBP-like domain in AlkA remains unknown but likely plays roles similar to those suggested for the TBP-domain in OGG1.

3. HhH-DG catalytic pocket architecture and its *modus operandi:* Variations of the same tricks.

3.1 A closer view into the catalytic pocket of HhH-DG subdomains.

HhH-DGs bear remarkable structural conservation of the catalytic pocket (**Figure 2B** and **4**). The catalytic pockets of HhH-DGs are composed of three α-helices (αH) followed by two loops (L) and an unstructured region (UR). The αH1 and Loop 1 (L1), are structural elements harbored in the α-helical subdomain 1. The αH2 with L2, and UR with αH3 are part of the HhH motif, and α-helical subdomain 2, respectively. The HhH motif is a widespread structural element among enzymes involved in DNA transactions, participating as a structural scaffold for non-sequence-specific protein-DNA interactions through electrostatic interactions between the DNA phosphodiester backbone and the HhH hairpin [121]. In the HhH-DG superfamily, the HhH motif participates in non-specific DNA-protein interaction through L2. Additionally, the motif takes on an active role in catalysis and substrate recognition (see **section 4.2** and **Figure 5** and **6**) through the αH2, which consist of residues Tyr/Lys/Trp in position R2 across the HhH-DGs and a Tyr in position R3 in AlkA (see below and **Figure 2**). L1 participates in interactions with the DNA phosphodiester backbone in OGG1 and MutY, suggesting a general role in stabilizing DNA-protein complexes. Given that αH1 of OGG1 does not interact with the substrate directly, it might be an

important structural scaffold of the active site. In contrast, Glu43 (R1) of αH1 in MutY has been proposed to play an active role as both a general acid/base in the glycosidic bond cleavage mechanism (See **section 3.2** and **Figure 6**). In addition, αH3 and the UR participate in multiple intimate interactions with the phosphodiester backbone and the base in MutY and OGG1 (**Figure 5**). Notably, the UR harbors the invariable Asp in R4 position (Asp144 and Asp268 in MutY and OGG1) that is a key catalytic residue in most HhH-DGs (**Figure 5**).

3.2 Topology of the catalytic pocket.

The catalytic pockets of HhH-DGs exhibit similar shapes with the exception of AlkA (**Figure 3A**). EndoIII, OGG1 and MutY contain the smallest catalytic pockets with a volume of 446, 527 and 453 $Å^3$, respectively. Despite the small catalytic pocket size, EndoIII displays notable substrate promiscuity, processing oxidized bases of different sizes ranging from 54 to 135 $Å^3$ (**Figure 1C**). This feature, along with molecular dynamics simulations, suggests that flexibility of the EndoIII active site allows for accommodation of the specific base lesion to be excised [122]. In contrast, MutY and OGG1 may have a low substrate promiscuity due a more structurally restricted active site with reduced flexibility.

The catalytic pocket of AlkA is much larger than the other HhH-DGs with a catalytic volume of 894 $Å^3$. The pocket has a lateral expansion that may provide the means to accommodate the diversity of potential alkylated bases that may be encountered by the enzyme [123-125]. Notably, the expanded active site may also have liabilities and be responsible for the observed pro-mutagenic removal of normal bases by AlkA [116]. A proportional relationship between active site size and substrate promiscuity has been reported with other enzymes [126-128]. Curiously, MagI displays are broader substrate scope and possesses a slightly larger catalytic pocket than both MutY and OGG1 that have a more restricted substrate scope (**Figure 4**). The larger active site cavity in MagI and AlkA may also be related to recognition of positively charged alkylated bases that relies on π-donor/acceptor interactions with electron-rich residues such as Trp, Phe and Tyr (R3 Y221 in AlkA and F158 in MagI) [5, 117, 120]. Indeed, these electron-rich residues appear with high frequency in the active site of TAG, MagI, MagIII and AlkA (**Figure 5A**).

3.3 Catalytic mechanisms of HhH superfamily through evolution: residue conservation and adaption

Despite the high degree of conservation of active site residues, the HhH-DG superfamily contains monofunctional glycosylases and bifunctional glycosylase/β-lyases [27]. Monofunctional DGs only catalyze the N-glycosidic bond hydrolysis between the damaged base and the deoxyribose while strand scission at the resulting AP site is catalyzed by other enzymes. On the other hand, bifunctional (glycosylase/lyase) DGs use a lysine side chain or the N-terminus to catalyze base cleavage which forms a covalent Schiff-base that results in a strand break at the 3' phosphate side of the sugar [27]. To provide insight into how evolution may have shaped the HhH-DG catalytic mechanisms, we will discuss details of the MutY mechanism and how specific residue changes (R1-4, see **Figure 2B**) may be related to differences in mechanism with other HhH-DGs. We will focus on discussing the roles of residues R1-4 whose functions have been supported by biochemical and structural studies, and evolutionary conservation.

The MutY glycosylase mechanism is the most fully defined of the HhH-DGs by kinetic isotope effect (KIE) studies, a plethora of structural studies and NMR determination of product stereochemistry [26, 129]. KIE measurements have revealed that MutY uses a stepwise S_N1 ($D_N{*}A_N^{\ddagger}$) mechanism that features the formation of a highly reactive oxacarbenium ion intermediate [130]. Of note, the KIE studies with MutY are consistent with a mechanism similar to acid-catalyzed depurination involving protonation of AN7 to enhance the base lability [131]. Indeed, in the crystal structures of *Geobacillus stearothermophilus (Gs)* MutY bound to a duplex containing the substrate analog FA (where FA = 2'-fluoroA) opposite OG, Glu43 (R1) is positioned near N7 of adenine, suggesting a role as the general acid [12, 16]. The significance of the corresponding Glu in *Escherichia coli* (Ec) MutY has also been demonstrated through site-directed mutagenesis, pH-dependent glycosylase assays and cellular assays [132].

Asp144 (R4) residue in Gs MutY corresponds to the universally covered Asp in HhH-DG. Site-directed mutagenesis and pH-dependent glycosylase assays are consistent with the Asp participating in catalysis as the carboxylate anion [12, 16, 132]. A structure of *Gs* MutY bound to a transition-state (TS) mimic (1-azaribose, 1N, referred to as transition state analog complex

(TSAC)), showed close approach of Asp144 to the position corresponding to C1' and a water nucleophile positioned close to Glu43 on the opposite side of the sugar in the space previously occupied by adenine [133]. The disposition of catalytic residues suggested retention of configuration that was confirmed by 2D-NMR determination of the stereochemistry of the methanolysis product. Based on the accumulated data, our laboratory proposed a revised mechanism for MutY similar to that for "retaining" O-glycosidases [133] (**Figure 6A**). In the proposed MutY mechanism, upon adenine departure, the oxacarbenium ion intermediate is stabilized by formation of a transient covalent intermediate with the catalytic Asp residue. The acetal intermediate is hydrolyzed by a water molecule activated by Glu43 to form the AP product [133]. The TSAC structure also highlighted the role of an active site Tyr residue (Tyr126 and R2) in potentially aiding in TS stabilization and hydrogen bonding with Glu43 (**Figure 6A**).

It is likely that all HhH-DGs use an S_N1 (or S_N1-like) mechanism that involves the universally conserved Asp (R4) residue stabilizing the developing charge in the TS and intermediate (**Figure 2B**). Moreover, similar to O-glycosidases, there may be both "retaining" and "inverting" HhH-DGs. In the structure of AlkA bound to the 1N TS mimic, there is similar close approach of the catalytic Asp and no room for a water molecule, suggesting approach of the water nucleophile from the opposite side [17]. In the case of MBD4, a structure with a G:T substrate suggested a covalent complex; however, the resolution was not sufficient to draw a definitive conclusion [82]. More recent structural studies with a suite of substrate, product and TS mimics suggests that MBD4 may be an inverting DG [134]. In fact, subtle differences in the active site may be sufficient to convert between retaining or inverting glycosylases, similar to what has been observed in O-glycosidases [135]. Further studies to define product stereochemistry may provide support for the existence of both retaining and inverting BER glycosylases.

In the HhH-DG superfamily, only MutY and MIG subfamilies contain the highly conserved residue R1 as Glu in their active site. As mentioned in **Section 3.1**, Glu in MutY has been proposed to act as a general acid due to the need to protonate adenine in purine N-glycosidic bond hydrolysis [130]. Non-enzymatic and enzymatic hydrolysis of pyrimidine nucleotides does not require an acid catalyst [136-141]. For this reason, the Glu in the thymine DG, MIG, has been suggested to participate in substrate stabilization through interactions to O4 and N3 of thymine

based on modeling of the nucleobase in the active site of a crystal structure of MIG [14]. Similarly, the substrates of AlkA are readily depurinated and therefore do not require a general acid catalyst. Consistent with this feature, R1 is an Ala in AlkA (**Figure 2B** and **5A**).

In MutY, residue R2 is a Tyr that H-bonds to the catalytic Glu. The identity of R2 in MIG and MBD4 is also a Tyr, though it is at a greater distance away from R1 compared to that observed in MutY. In a structure of MBD4 bound to substrate and the structure of MIG with thymine modeled in its active site, the contact of R2 Tyr to the thymine O2 suggests that the Tyr may play a role in substrate recognition and in catalysis by stabilizing the anionic nucleobase leaving group [14, 82]. For alkylpurine HhH-DGs like AlkA and MagI, R2 position is a Trp (W218 and W150, respectively) which forms a hydrogen bond contact with R4 (Asp) via the indole sidechain [17, 118]. Most conspicuously, the R2 residue corresponds to a Lys in bifunctional HhH-DGs (EndoIII, OGG1). The Lys side chain in bifunctional HhH-DGs participates in base excision and forms a transient covalent Schiff-base intermediate with the resulting AP site that facilitates an enzyme catalyzed β-elimination (β-lyase) strand scission reaction. Notably, the role of the "Lys" is analogous to that of piperidine treatment in Maxim-Gilbert sequencing reactions [27].

Many previously proposed mechanisms have implied that the Lys residue directly attacks C1' of the target nucleotide in an S_N2 mechanism to produce the imine; however, there are several lines of evidence that suggest that monofunctional and bifunctional HhH-DGs likely have mechanisms that are more similar. Outstandingly, MutY and EndoIII show mechanistic plasticity, where a single mutation turns monofunctional into a bifunctional DG and vice versa [142-144]. This suggests that the evolutionary path that separates the catalytic modes can be only one mutation away**.** In the case of hOGG1, identification of separation of function mutations have indicated that its glycosylase and lyase activities are uncoupled, and have further suggested that OGG1 may function as a monofunctional glycosylase in cells [145, 146]. All HhH-DGs evaluated exhibit extremely high affinity to duplex DNA containing positively charged azaribose nucleotides that mimic the positive charge build-up in S_N1 transition state(s) or intermediate(s) [133]. Based on these arguments, we suggest that the HhH-DGs use a general catalytic pathway with subtle differences in the extents of C-N glycosidic bond breakage in the TS due to strategic variations in the active site to cope with the biologically relevant lesion(s) (**Figure 6**).

As mentioned before, the different catalytic components of HhH-DG are positioned at both α-helical subdomains and HhH motif. Nonetheless, TAG is a HhH-DG that lacks α-helical subdomain 2 and its HhH motif appears to primarily participate in making non-specific DNA-protein interaction [147, 148]. This implies that the catalytic elements in TAG are organized within the α-helical subdomain 1. Interestingly, R1 residue is a Glu (E38) as is present in MIG and MutY [147, 148]. Although, the role of the E38 in TAG is not clear, it has been proposed to have a similar role to R1 Glu of MutY and MIG, promoting recognition and removal of 3-methyladenine by hydrogen bonding the 6-amino and N7 position [148]. The Asp at position R4 has been reported as a missing component in TAG [147]. However, in our sequence alignment, including all HhH-DNAG families, D174 in Salmonella typhi TAG aligns exactly at position R4 and is highly conserved among other TAGs (**Figure 2B**). Nonetheless, in spite of the conservation at the sequence level of Asp in TAG, it is not been established to play a role in catalysis. Lastly, similarly to MagI and AlkA, TAG utilizes electron-rich residues such as Trp, and Tyr to recognize alkylated bases through π-donor/acceptor interactions.

5. Final remarks; reconstruction of the evolutionary pathway of HhH-DNAG superfamily.

The multifaceted features of HhH-DGs reveals a robust structural/functional malleability that has been exploited during evolution. The monophyletic and taxonomic distributions of HhH-DGs suggest a shared evolutionary origin in the last ultimate common ancestor (LUCA) around 3.65 billion years ago [149]. It also suggests that all the HhH families arose by means of gene duplications followed by substrate specificity divergence and specialization as described for other hydrolases [150, 151] and DNA repair enzymes [152] (**Figure 1**). The existence of paralogues in EndoIII and MagI families with different substrate specificities [37, 119, 153] and the substrate [154] and mechanistic [142-144] exchangeability between HhH-DG families support the possibility of this evolutionary scenario. Based on the structural, biochemical and functional properties discussed herein, we reconstructed a hypothetical evolutionary path for HhH-DG superfamily (**Figure 7**).

The fact that TAG, MagI and MagIII are structurally the most austere HhH-DGs and that alkylpurine DGs are clustered at four different clades with different phylogenetic levels (**Figure 1A**) makes a Mag-like DG the most parsimonious candidate for the last common ancestor (LCA)

of the HhH-DG superfamily. It might have had a narrow substrate preference being able to process only 3mA and 7mG lesion, since they are the most common substrates in HhH alkylpurine DGs. Of note, crystal structures of TAG and MagIII with alkylated adenines in its active site show mainly π-stacking interactions with aromatic residues for substrate recognition almost devoid of polar interaction [117]. This recognition strategy may resemble the substrate recognition pattern of HhH-DG LCA. Subsequently, multiple gene duplications of the HhH-DG LCA rendered common intermediate ancestors that were further evolved to produce the distinct HhH-DG family clades. During this process, the acquisition of new domains and motifs was a strong evolutionary driver that led to substrate specialization and coordination with other cellular pathways. Acquisition of MutT-like and MBD domain acquisition in MutY and MBD4 provided a means to target their glycosylase activity toward mismatches with OG and CpG regulatory regions, respectively. The acquisition of [4Fe-4S]-cofactor was a milestone in EndoIII, MutY and MIG evolution since it provided a new functional tool to address structural and lesion recognition challenges. Similarly, acquisition of TBP-like domain in OGG1/AlkA ancestor aided in diversifying function in HhH-DGs in a manner that is still unclear. Given that MBD4 is an exclusive eukaryotic enzyme and structural similarity between its DG domain and MUTYH catalytic core, MBD4 might have close evolutionary origin to MutY family whose origin might have emerged at some point of the eukaryotic lineage evolution. Evolutionary pressure induced changes to the active site of HhH-DGs provided for fine-tuning of the substrate specificity resulting in the observed broad DNA lesion repair scope. The end result, is an extraordinary and versatile DG superfamily that is able to meet the challenges imposed by DNA base damage.

Acknowledgements

We gratefully acknowledge support for our research on DNA glycosylases for many years by NIH and NSF to S.S.D. The graduate student authors M.D. and W.-J. L. were supported by NSF CHE-1905304 and the Department of Chemistry, UC Davis. C.H.T.A was supported by a UCMEXUS-CONACYT Postdoctoral Fellowship. The authors also thank Nhu Tran for critical reading. We also thank our crystallographer collaborator Professor Martin Horvath for his keen insights into MutY mechanisms. We apologize to authors whose beautiful work we were unable to highlight in this review due to limited space!

References

[1] S.S. David, V.L. O'Shea, S. Kundu, Base-excision repair of oxidative DNA damage, Nature, 447 (2007) 941-950.
[2] J.I. Lucas-Lledó, R. Maddamsetti, M. Lynch, Phylogenomic analysis of the uracil-DNA glycosylase superfamily, Molecular biology and evolution, 28 (2011) 1307-1317.
[3] B. Dalhus, I.H. Helle, P.H. Backe, I. Alseth, T. Rognes, M. Bjørås, J.K. Laerdahl, Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats, Nucleic acids research, 35 (2007) 2451-2459.
[4] A. Prakash, S. Doublié, S.S. Wallace, The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage, Progress in molecular biology and translational science, 110 (2012) 71-91.
[5] M.D. Wyatt, J.M. Allan, A.Y. Lau, T.E. Ellenberger, L.D. Samson, 3-methyladenine DNA glycosylases: structure, function, and biological importance, Bioessays, 21 (1999) 668-676.
[6] D.R. Denver, S.L. Swenson, M. Lynch, An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases, Molecular biology and evolution, 20 (2003) 1603-1611.
[7] A.G. Raetz, D.M. Banda, X. Ma, G. Xu, A.N. Rajavel, P.L. McKibbin, C.B. Lebrilla, S.S. David, The DNA repair enzyme MUTYH potentiates cytotoxicity of the alkylating agent MNNG by interacting with abasic sites, Journal of Biological Chemistry, 295 (2020) 3692-3707.
[8] A.G. Raetz, S.S. David, When you're strange: unusual features of the MUTYH glycosylase and implications in cancer, DNA repair, 80 (2019) 16-25.
[9] A.C. Drohat, C.T. Coey, Role of base excision "repair" enzymes in erasing epigenetic marks from DNA, Chemical reviews, 116 (2016) 12711-12729.
[10] A.M. Fleming, Y. Ding, C.J. Burrows, Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair, Proceedings of the National Academy of Sciences, 114 (2017) 2604-2609.
[11] B.R. Bowman, S. Lee, S. Wang, G.L. Verdine, Structure of Escherichia coli AlkA in complex with undamaged DNA, Journal of Biological Chemistry, 285 (2010) 35783-35791.
[12] S. Lee, G.L. Verdine, Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase, Proceedings of the National Academy of Sciences, 106 (2009) 18497-18502.
[13] J.C. Fromme, G.L. Verdine, Structure of a trapped endonuclease III–DNA covalent intermediate, The EMBO journal, 22 (2003) 3461-3471.
[14] C.D. Mol, A.S. Arvai, T.J. Begley, R.P. Cunningham, J.A. Tainer, Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases, Journal of molecular biology, 315 (2002) 373-384.
[15] S.D. Bruner, D.P. Norman, G.L. Verdine, Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA, nature, 403 (2000) 859-866.
[16] Y. Guan, R.C. Manuel, A.S. Arvai, S.S. Parikh, C.D. Mol, J.H. Miller, R.S. Lloyd, J.A. Tainer, MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily, Nature structural biology, 5 (1998) 1058-1064.
[17] T. Hollis, Y. Ichikawa, T. Ellenberger, DNA bending and a flip-out mechanism for base excision by the helix–hairpin–helix DNA glycosylase, Escherichia coli AlkA, The EMBO journal, 19 (2000) 758-766.
[18] L. Patthy, Genome evolution and the evolution of exon-shuffling—a review, Gene, 238 (1999) 103-114.
[19] J.A. Marsh, S.A. Teichmann, How do proteins gain new domains?, Genome biology, 11 (2010) 1-4.
[20] M. Buljan, A. Frankish, A. Bateman, Quantifying the mechanisms of domain gain in animal proteins, Genome biology, 11 (2010) 1-15.

[21] Y.G. Mok, R. Uzawa, J. Lee, G.M. Weiner, B.F. Eichman, R.L. Fischer, J.H. Huh, Domain structure of the DEMETER 5-methylcytosine DNA glycosylase, Proceedings of the National Academy of Sciences, 107 (2010) 19225-19230.
[22] Y. Choi, M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S.E. Jacobsen, R.L. Fischer, DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis, Cell, 110 (2002) 33-42.
[23] M. Gehring, J.H. Huh, T.-F. Hsieh, J. Penterman, Y. Choi, J.J. Harada, R.B. Goldberg, R.L. Fischer, DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation, cell, 124 (2006) 495-506.
[24] V.K. Schoft, N. Chumak, Y. Choi, M. Hannon, M. Garcia-Aguilar, A. Machlicova, L. Slusarz, M. Mosiolek, J.-S. Park, G.T. Park, Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte, Proceedings of the National Academy of Sciences, 108 (2011) 8042-8047.
[25] F. Faucher, S. Doublié, Z. Jia, 8-oxoguanine DNA glycosylases: one lesion, three subfamilies, International journal of molecular sciences, 13 (2012) 6711-6729.
[26] D.M. Banda, N.N. Nuñez, M.A. Burnside, K.M. Bradshaw, S.S. David, Repair of 8-oxoG: A mismatches by the MUTYH glycosylase: Mechanism, metals and medicine, Free Radical Biology and Medicine, 107 (2017) 202-215.
[27] S.S. David, S.D. Williams, Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair, Chemical Reviews, 98 (1998) 1221-1262.
[28] H. Yang, S. Fitz-Gibbon, E.M. Marcotte, J.H. Tai, E.C. Hyman, J.H. Miller, Characterization of a Thermostable DNA Glycosylase Specific for U/G and T/G Mismatches from the Hyperthermophilic ArchaeonPyrobaculum aerophilum, Journal of bacteriology, 182 (2000) 1272-1279.
[29] C.H. Trasviña-Arenas, L.M. Lopez-Castillo, E. Sanchez-Sandoval, L.G. Brieba, Dispensability of the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY, The FEBS journal, 283 (2016) 521-540.
[30] M.A. Pope, S.S. David, DNA damage recognition and repair by the murine MutY homologue, DNA repair, 4 (2005) 91-102.
[31] A.J. Lee, C. Majumdar, S.D. Kathe, R.P. Van Ostrand, H.R. Vickery, A.M. Averill, S.R. Nelson, A.H. Manlove, M.A. McCord, S.S. David, Detection of OG: A lesion mispairs by MutY relies on a single His residue and the 2-amino group of 8-oxoguanine, Journal of the American Chemical Society, 142 (2020) 13283-13287.
[32] A.H. Manlove, P.L. McKibbin, E.L. Doyle, C. Majumdar, M.L. Hamm, S.S. David, Structure–activity relationships reveal key features of 8-oxoguanine: a mismatch detection by the MutY glycosylase, ACS chemical biology, 12 (2017) 2335-2344.
[33] C. Majumdar, P.L. McKibbin, A.E. Krajewski, A.H. Manlove, J.K. Lee, S.S. David, Unique Hydrogen Bonding of Adenine with the Oxidatively Damaged Base 8-Oxoguanine Enables Specific Recognition and Repair by DNA Glycosylase MutY, Journal of the American Chemical Society, 142 (2020) 20340-20350.
[34] S.L. Porello, A.E. Leyes, S.S. David, Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates, Biochemistry, 37 (1998) 14756-14764.
[35] M. Michaels, J.H. Miller, The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7, 8-dihydro-8-oxoguanine), Journal of bacteriology, 174 (1992) 6321.
[36] P.A. Lind, D.I. Andersson, Whole-genome mutational biases in bacteria, Proceedings of the National Academy of Sciences, 105 (2008) 17878-17883.
[37] C.H. Trasviña-Arenas, S.S. David, L. Delaye, E. Azuara-Liceaga, L.G. Brieba, Evolution of Base Excision Repair in Entamoeba histolytica is shaped by gene loss, gene duplication, and lateral gene transfer, DNA repair, 76 (2019) 76-88.

[38] N. Al-Tassan, N.H. Chmiel, J. Maynard, N. Fleming, A.L. Livingston, G.T. Williams, A.K. Hodges, D.R. Davies, S.S. David, J.R. Sampson, Inherited variants of MYH associated with somatic G: C→ T: A mutations in colorectal tumors, Nature genetics, 30 (2002) 227-232.
[39] J. Horst, H. Fritz, Counteracting the mutagenic effect of hydrolytic deamination of DNA 5-methylcytosine residues at high temperature: DNA mismatch N-glycosylase Mig. Mth of the thermophilic archaeon Methanobacterium thermoautotrophicum THF, The EMBO journal, 15 (1996) 5459-5469.
[40] B.K. Duncan, J.H. Miller, Mutagenic deamination of cytosine residues in DNA, Nature, 287 (1980) 560-561.
[41] M. Shiraishi, S. Ishino, M. Heffernan, I. Cann, Y. Ishino, The mesophilic archaeon Methanosarcina acetivorans counteracts uracil in DNA with multiple enzymes: EndoQ, ExoIII, and UDG, Scientific reports, 8 (2018) 1-14.
[42] K.J. Fryxell, E. Zuckerkandl, Cytosine deamination plays a primary role in the evolution of mammalian isochores, Molecular biology and evolution, 17 (2000) 1371-1383.
[43] R.S. Lasken, D.M. Schuster, A. Rashtchian, Archaebacterial DNA polymerases tightly bind uracil-containing DNA, Journal of Biological Chemistry, 271 (1996) 17692-17696.
[44] D.W. Grogan, Understanding DNA repair in hyperthermophilic Archaea: persistent gaps and other reasons to focus on the fork, Archaea, 2015 (2015).
[45] C.L. Chepanoske, M.-P. Golinelli, S.D. Williams, S.S. David, Positively Charged Residues within the Iron–Sulfur Cluster Loop of E. coli MutY Participate in Damage Recognition and Removal, Archives of biochemistry and biophysics, 380 (2000) 11-19.
[46] T.E. Messick, N.H. Chmiel, M.-P. Golinelli, M.R. Langer, L. Joshua-Tor, S.S. David, Noncysteinyl coordination to the [4Fe-4S] 2+ cluster of the DNA repair adenine glycosylase MutY introduced via site-directed mutagenesis. Structural characterization of an unusual histidinyl-coordinated cluster, Biochemistry, 41 (2002) 3931-3942.
[47] S.L. Porello, M.J. Cannon, S.S. David, A substrate recognition role for the [4Fe-4S] 2+ cluster of the DNA repair glycosylase MutY, Biochemistry, 37 (1998) 6465-6475.
[48] M.-P. Golinelli, N.H. Chmiel, S.S. David, Site-directed mutagenesis of the cysteine ligands to the [4Fe− 4S] cluster of Escherichia coli MutY, Biochemistry, 38 (1999) 6997-7007.
[49] A.K. Boal, E. Yavin, O.A. Lukianova, V.L. O'Shea, S.S. David, J.K. Barton, DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters, Biochemistry, 44 (2005) 8397-8407.
[50] K.J. McDonnell, J.A. Chemler, P.L. Bartels, E. O'Brien, M.L. Marvin, J. Ortega, R.H. Stern, L. Raskin, G.M. Li, D.H. Sherman, J.K. Barton, S.B. Gruber, A human MUTYH variant linking colonic polyposis to redox degradation of the [4Fe4S](2+) cluster, Nature chemistry, 10 (2018) 873-880.
[51] E.M. Boon, A.L. Livingston, N.H. Chmiel, S.S. David, J.K. Barton, DNA-mediated charge transport for DNA repair, Proceedings of the National Academy of Sciences, 100 (2003) 12543-12547.
[52] A.K. Boal, J.C. Genereux, P.A. Sontz, J.A. Gralnick, D.K. Newman, J.K. Barton, Redox signaling between DNA repair proteins for efficient lesion detection, Proceedings of the National Academy of Sciences, 106 (2009) 15237-15242.
[53] L.M. Engstrom, O.A. Partington, S.S. David, An Iron–Sulfur Cluster Loop Motif in the Archaeoglobus fulgidus Uracil–DNA Glycosylase Mediates Efficient Uracil Recognition and Removal, Biochemistry, 51 (2012) 5187-5197.
[54] R.M. Silva, M.A. Grodick, J.K. Barton, UvrC Coordinates an O2-Sensitive [4Fe4S] Cofactor, Journal of the American Chemical Society, 142 (2020) 10964-10977.
[55] T.P. Mui, J.O. Fuss, J.P. Ishida, J.A. Tainer, J.K. Barton, ATP-stimulated, DNA-mediated redox signaling by XPD, a DNA repair and transcription helicase, Journal of the American Chemical Society, 133 (2011) 16378-16381.

[56] E. O'Brien, M.E. Holt, L.E. Salay, W.J. Chazin, J.K. Barton, Substrate binding regulates redox signaling in human DNA primase, Journal of the American Chemical Society, 140 (2018) 17153-17162.
[57] P.L. Bartels, J.L. Stodola, P.M. Burgers, J.K. Barton, A redox role for the [4Fe4S] cluster of yeast DNA polymerase δ, Journal of the American Chemical Society, 139 (2017) 18339-18348.
[58] L.M. Engstrom, M.K. Brinkmeyer, Y. Ha, A.G. Raetz, B. Hedman, K.O. Hodgson, E.I. Solomon, S.S. David, A zinc linchpin motif in the MUTYH glycosylase interdomain connector is required for efficient repair of DNA damage, Journal of the American Chemical Society, 136 (2014) 7829-7832.
[59] N.N. Nuñez, C. Khuu, C.S. Babu, S.J. Bertolani, A.N. Rajavel, J.E. Spear, J.A. Armas, J.D. Wright, J.B. Siegel, C. Lim, The Zinc Linchpin Motif in the DNA Repair Glycosylase MUTYH: Identifying the Zn2+ Ligands and Roles in Damage Recognition and Repair, Journal of the American Chemical Society, 140 (2018) 13260-13271.
[60] T. Nakamura, K. Okabe, S. Hirayama, M. Chirifu, S. Ikemizu, H. Morioka, Y. Nakabeppu, Y. Yamagata, Structure of the mammalian adenine DNA glycosylase MUTYH: insights into the base excision repair pathway and cancer, Nucleic Acids Research, (2021).
[61] B.-J. Hwang, J. Jin, Y. Gao, G. Shi, A. Madabushi, A. Yan, X. Guan, M. Zalzman, S. Nakajima, L. Lan, SIRT6 protein deacetylase interacts with MYH DNA glycosylase, APE1 endonuclease, and Rad9–Rad1–Hus1 checkpoint clamp, BMC molecular biology, 16 (2015) 1-16.
[62] P.J. Luncsford, B.A. Manvilla, D.N. Patterson, S.S. Malik, J. Jin, B.-J. Hwang, R. Gunther, S. Kalvakolanu, L.J. Lipinski, W. Yuan, Coordination of MYH DNA glycosylase and APE1 endonuclease activities via physical interactions, DNA repair, 12 (2013) 1043-1052.
[63] T. Nakamura, S. Meshitsuka, S. Kitagawa, N. Abe, J. Yamada, T. Ishino, H. Nakano, T. Tsuzuki, T. Doi, Y. Kobayashi, Structural and dynamic features of the MutT protein in the recognition of nucleotides with the mutagenic 8-oxoguanine base, Journal of Biological Chemistry, 285 (2010) 444-452.
[64] L.P. Russelburg, V.L. O'Shea Murray, M. Demir, K.R. Knutsen, S.L. Sehgal, S. Cao, S.S. David, M.P. Horvath, Structural Basis for Finding OG Lesions and Avoiding Undamaged G by the DNA Glycosylase MutY, ACS chemical biology, 15 (2019) 93-102.
[65] D.M. Noll, A. Gogos, J.A. Granek, N.D. Clarke, The C-Terminal domain of the Adenine-DNA glycosylase MutY confers specificity for 8-Oxoguanine⊙ adenine mispairs and may have evolved from MutT, an 8-Oxo-dGTPase, Biochemistry, 38 (1999) 6374-6379.
[66] A.L. Livingston, V.L. O'Shea, T. Kim, E.T. Kool, S.S. David, Unnatural substrates reveal the importance of 8-oxoguanine for in vivo mismatch repair by MutY, Nature chemical biology, 4 (2008) 51-58.
[67] M.M. Thayer, H. Ahern, D. Xing, R.P. Cunningham, J.A. Tainer, Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure, The EMBO journal, 14 (1995) 4108-4120.
[68] C.A. Romano, P.A. Sontz, J.K. Barton, Mutants of the base excision repair glycosylase, endonuclease III: DNA charge transport as a first step in lesion detection, Biochemistry, 50 (2011) 6133-6145.
[69] W. Fu, S. O'Handley, R. Cunningham, M.K. Johnson, The role of the iron-sulfur cluster in Escherichia coli endonuclease III. A resonance Raman study, Journal of Biological Chemistry, 267 (1992) 16135-16137.
[70] A.A. Gorodetsky, A.K. Boal, J.K. Barton, Direct electrochemistry of endonuclease III in the presence and absence of DNA, Journal of the American Chemical Society, 128 (2006) 12082-12083.
[71] A. Hassan, L.J. Macedo, J.C. de Souza, F.C. Lima, F.N. Crespilho, A combined Far-FTIR, FTIR Spectromicroscopy, and DFT Study of the Effect of DNA Binding on the [4Fe4S] Cluster Site in EndoIII, Scientific reports, 10 (2020) 1-12.
[72] P. Aller, M.A. Rould, M. Hogg, S.S. Wallace, S. Doublié, A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol, Proceedings of the National Academy of Sciences, 104 (2007) 814-818.

[73] J.M. McNulty, B. Jerkovic, P.H. Bolton, A.K. Basu, Replication inhibition and miscoding properties of DNA templates containing a site-specific cis-thymine glycol or urea residue, Chemical research in toxicology, 11 (1998) 666-673.
[74] L. Das, V.G. Quintana, J.B. Sweasy, NTHL1 in genomic integrity, aging and cancer, DNA repair, 93 (2020) 102920.
[75] C.G. Marsden, P. Jaruga, E. Coskun, R.L. Maher, D.S. Pederson, M. Dizdaroglu, J.B. Sweasy, Expression of a germline variant in the N-terminal domain of the human DNA glycosylase NTHL1 induces cellular transformation without impairing enzymatic function or substrate specificity, Oncotarget, 11 (2020) 2262–2272.
[76] A. de Mendoza, D. Poppe, S. Buckberry, J. Pflueger, C.B. Albertin, T. Daish, S. Bertrand, E. de la Calle-Mustienes, J.L. Gómez-Skarmeta, J.R. Nery, The emergence of the brain non-CpG methylation system in vertebrates, Nature Ecology & Evolution, (2021) 1-10.
[77] F. Nota, A.C. DamiA-n, P. Ribone, E.A. MarA-a, Expression and function of AtMBD4L, the single gene encoding the nuclear DNA glycosylase MBD4L in Arabidopsis, Plant Science, 235 (2015) 122-129.
[78] L.M. Iyer, S. Abhiman, L. Aravind, Natural history of eukaryotic DNA methylation systems, Progress in molecular biology and translational science, 101 (2011) 25-104.
[79] D.P. Turner, S. Cortellino, J.E. Schupp, E. Caretti, T. Loh, T.J. Kinsella, A. Bellacosa, The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine, Cancer research, 66 (2006) 7686-7693.
[80] M.A. Aziz, J.E. Schupp, T.J. Kinsella, Modulation of the activity of methyl binding domain protein 4 (MBD4/MED1) while processing iododeoxyuridine generated DNA mispairs, Cancer biology & therapy, 8 (2009) 1156-1163.
[81] H. Hashimoto, Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, X. Zhang, X. Cheng, Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic acids research, 40 (2012) 4841-4849.
[82] H. Hashimoto, X. Zhang, X. Cheng, Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation, Nucleic acids research, 40 (2012) 8276-8284.
[83] F. Petronzelli, A. Riccio, G.D. Markham, S.H. Seeholzer, M. Genuardi, M. Karbowski, A.T. Yeung, Y. Matsumoto, A. Bellacosa, Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): Fundamental role of the catalytic domain, Journal of cellular physiology, 185 (2000) 473-480.
[84] F. Petronzelli, A. Riccio, G.D. Markham, S.H. Seeholzer, J. Stoerker, M. Genuardi, A.T. Yeung, Y. Matsumoto, A. Bellacosa, Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase, Journal of Biological Chemistry, 275 (2000) 32422-32429.
[85] S. Cortellino, D. Turner, V. Masciullo, F. Schepis, D. Albino, R. Daniel, A.M. Skalka, N.J. Meropol, C. Alberti, L. Larue, The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity, Proceedings of the National Academy of Sciences, 100 (2003) 15071-15076.
[86] S. Morera, I. Grin, A. Vigouroux, S. Couve, V. Henriot, M. Saparbaev, A.A. Ishchenko, Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA, Nucleic acids research, 40 (2012) 9917-9926.
[87] J. Otani, K. Arita, T. Kato, M. Kinoshita, H. Kimura, I. Suetake, S. Tajima, M. Ariyoshi, M. Shirakawa, Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4, Journal of Biological Chemistry, 288 (2013) 6351-6362.
[88] B. Hendrich, U. Hardeland, H.-H. Ng, J. Jiricny, A. Bird, The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites, Nature, 401 (1999) 301-304.

[89] B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl CpG-binding proteins, Genetics Research, 72 (1998) 59-72.
[90] K.J. Fryxell, W.-J. Moon, CpG mutation rates in the human genome are highly dependent on local GC content, Molecular Biology and Evolution, 22 (2005) 650-658.
[91] R.C. Poulos, J. Olivier, J.W. Wong, CpG methylation accounts for genome-wide C> T mutation variation and cancer driver formation across cancer types, bioRxiv, (2017) 106872.
[92] M. Rodrigues, L. Mobuchon, A. Houy, S. Alsafadi, S. Baulande, O. Mariani, B. Marande, K.A. Rais, M.K. Van der Kooij, E. Kapiteijn, Evolutionary routes in metastatic uveal melanomas depend on MBD4 alterations, Clinical Cancer Research, 25 (2019) 5513-5524.
[93] M.A. Sanders, E. Chew, C. Flensburg, A. Zeilemaker, S.E. Miller, A.S. Al Hinai, A. Bajel, B. Luiken, M. Rijken, T. Mclennan, MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML, Blood, The Journal of the American Society of Hematology, 132 (2018) 1526-1534.
[94] P. Repo, J.E. Jäntti, R.S. Järvinen, E.S. Rantala, M. Täll, V. Raivio, T.T. Kivelä, J.A. Turunen, Germline loss-of-function variants in MBD4 are rare in Finnish patients with uveal melanoma, Pigment Cell & Melanoma Research, 33 (2020) 756-762.
[95] R. Tricarico, S. Cortellino, A. Riccio, S. Jagmohan-Changur, H. Van der Klift, J. Wijnen, D. Turner, A. Ventura, V. Rovella, A. Percesepe, Involvement of MBD4 inactivation in mismatch repair-deficient tumorigenesis, Oncotarget, 6 (2015) 42892–42904.
[96] S.M. Robey-Bond, R. Barrantes-Reynolds, J.P. Bond, S.S. Wallace, V. Bandaru, Clostridium acetobutylicum 8-oxoguanine DNA glycosylase (Ogg) differs from eukaryotic Oggs with respect to opposite base discrimination, Biochemistry, 47 (2008) 7626-7636.
[97] N. Jena, P. Mishra, Formation of ring-opened and rearranged products of guanine: mechanisms and biological significance, Free Radical Biology and Medicine, 53 (2012) 81-94.
[98] Y. Xie, H. Yang, C. Cunanan, K. Okamoto, D. Shibata, J. Pan, D.E. Barnes, T. Lindahl, M. McIlhatton, R. Fishel, Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors, Cancer research, 64 (2004) 3096-3102.
[99] H. Sampath, V. Vartanian, M.R. Rollins, K. Sakumi, Y. Nakabeppu, R.S. Lloyd, 8-Oxoguanine DNA glycosylase (OGG1) deficiency increases susceptibility to obesity and metabolic dysfunction, PloS one, 7 (2012) e51697.
[100] V. Vartanian, J. Tumova, P. Dobrzyn, A. Dobrzyn, Y. Nakabeppu, R.S. Lloyd, H. Sampath, 8-oxoguanine DNA glycosylase (OGG1) deficiency elicits coordinated changes in lipid and mitochondrial metabolism in muscle, PLoS One, 12 (2017) e0181687.
[101] G. Mao, X. Pan, B.-B. Zhu, Y. Zhang, F. Yuan, J. Huang, M.A. Lovell, M.P. Lee, W.R. Markesbery, G.-M. Li, Identification and characterization of OGG1 mutations in patients with Alzheimer's disease, Nucleic acids research, 35 (2007) 2759-2766.
[102] P.-C. Pao, D. Patnaik, L.A. Watson, F. Gao, L. Pan, J. Wang, C. Adaikkan, J. Penney, H.P. Cam, W.-C. Huang, HDAC1 modulates OGG1-initiated oxidative DNA damage repair in the aging brain and Alzheimer's disease, Nature communications, 11 (2020) 1-17.
[103] T. Visnes, A. Cázares-Körner, W. Hao, O. Wallner, G. Masuyer, O. Loseva, O. Mortusewicz, E. Wiita, A. Sarno, A. Manoilov, Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation, Science, 362 (2018) 834-839.
[104] J.G. Mabley, P. Pacher, A. Deb, R. Wallace, R.H. Elder, C. Szabó, Potential role for 8-oxoguanine DNA glycosylase in regulating inflammation, The FASEB journal, 19 (2005) 1-18.
[105] S. Marsin, A.E. Vidal, M. Sossou, J. Ménissier-de Murcia, F. Le Page, S. Boiteux, G. de Murcia, J.P. Radicella, Role of XRCC1 in the coordination and stimulation of oxidative DNA damage repair initiated by the DNA glycosylase hOGG1, Journal of Biological Chemistry, 278 (2003) 44068-44074.

[106] B. Hang, B. Singer, Protein− Protein Interactions Involving DNA Glycosylases, Chemical research in toxicology, 16 (2003) 1181-1195.
[107] S. Jang, N. Kumar, E.C. Beckwitt, M. Kong, E. Fouquerel, V. Rapić-Otrin, R. Prasad, S.C. Watkins, C. Khuu, C. Majumdar, Damage sensor role of UV-DDB during base excision repair, Nature structural & molecular biology, 26 (2019) 695-703.
[108] M. d'Errico, E. Parlanti, M. Teson, B.M.B. De Jesus, P. Degan, A. Calcagnile, P. Jaruga, M. Bjørås, M. Crescenzi, A.M. Pedrini, New functions of XPC in the protection of human skin cells from oxidative damage, The EMBO journal, 25 (2006) 4305-4315.
[109] N.C. de Souza-Pinto, S. Maynard, K. Hashiguchi, J. Hu, M. Muftuoglu, V.A. Bohr, The recombination protein RAD52 cooperates with the excision repair protein OGG1 for the repair of oxidative lesions in mammalian cells, Molecular and cellular biology, 29 (2009) 4441-4454.
[110] M.J. Park, J.-H. Park, S.-H. Hahm, S.I. Ko, Y.R. Lee, J.H. Chung, S.Y. Sohn, Y. Cho, L.-W. Kang, Y.S. Han, Repair activities of human 8-oxoguanine DNA glycosylase are stimulated by the interaction with human checkpoint sensor Rad9–Rad1–Hus1 complex, DNA repair, 8 (2009) 1190-1200.
[111] N.N. Hooten, K. Kompaniez, J. Barnes, A. Lohani, M.K. Evans, Poly (ADP-ribose) polymerase 1 (PARP-1) binds to 8-oxoguanine-DNA glycosylase (OGG1), Journal of Biological Chemistry, 286 (2011) 44679-44690.
[112] B. Sedgwick, P.A. Bates, J. Paik, S.C. Jacobs, T. Lindahl, Repair of alkylated DNA: recent advances, DNA repair, 6 (2007) 429-442.
[113] E.C. Friedberg, G.C. Walker, W. Siede, R.D. Wood, DNA repair and mutagenesis, American Society for Microbiology Press, 2005.
[114] P. Taverna, B. Sedgwick, Generation of an endogenous DNA-methylating agent by nitrosation in Escherichia coli, Journal of bacteriology, 178 (1996) 5105-5111.
[115] J.-H. Yoon, J.R. Choudhury, J. Park, S. Prakash, L. Prakash, Translesion synthesis DNA polymerases promote error-free replication through the minor-groove DNA adduct 3-deaza-3-methyladenine, Journal of Biological Chemistry, 292 (2017) 18682-18688.
[116] K.G. Berdal, R.F. Johansen, E. Seeberg, Release of normal bases from intact DNA by a native DNA repair enzyme, The EMBO journal, 17 (1998) 363-367.
[117] B.F. Eichman, E.J. O'Rourke, J.P. Radicella, T. Ellenberger, Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases, The EMBO journal, 22 (2003) 4898-4909.
[118] S. Adhikary, B.F. Eichman, Analysis of substrate specificity of Schizosaccharomyces pombe Mag1 alkylpurine DNA glycosylase, EMBO reports, 12 (2011) 1286-1292.
[119] S. Adhikary, M.C. Cato, K.L. McGary, A. Rokas, B.F. Eichman, Non-productive DNA damage binding by DNA glycosylase-like protein Mag2 from Schizosaccharomyces pombe, DNA repair, 12 (2013) 196-204.
[120] J. Labahn, O.D. Schärer, A. Long, K. Ezaz-Nikpay, G.L. Verdine, T.E. Ellenberger, Structural basis for the excision repair of alkylation-damaged DNA, Cell, 86 (1996) 321-329.
[121] A.J. Doherty, L.C. Serpell, C.P. Ponting, The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA, Nucleic acids research, 24 (1996) 2488-2497.
[122] A. Sarre, M. Stelter, F. Rollo, S. De Bonis, A. Seck, C. Hognon, J.-L. Ravanat, A. Monari, F. Dehez, E. Moe, The three Endonuclease III variants of Deinococcus radiodurans possess distinct and complementary DNA repair activities, DNA repair, 78 (2019) 45-59.
[123] P.J. O'Brien, T. Ellenberger, The Escherichia coli 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site, Journal of Biological Chemistry, 279 (2004) 26876-26884.
[124] P.J. O'Brien, T. Ellenberger, Human alkyladenine DNA glycosylase uses acid− base catalysis for selective excision of damaged purines, Biochemistry, 42 (2003) 12418-12429.
[125] W.B. Mattes, C.S. Lee, J. Laval, T.R. O'Connor, Excision of DNA adducts of nitrogen mustards by bacterial and mammalian 3-methyladenine-DNA glycosylases, Carcinogenesis, 17 (1996) 643-648.

[126] O. Khersonsky, D.S. Tawfik, Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective, Annu Rev Biochem, 79 (2010) 471-505.
[127] Y. Zhang, J. An, G.-Y. Yang, A. Bai, B. Zheng, Z. Lou, G. Wu, W. Ye, H.-F. Chen, Y. Feng, Active site loop conformation regulates promiscuous activity in a lactonase from Geobacillus kaustophilus HTA426, PLoS One, 10 (2015) e0115130.
[128] B.J. Jones, R.L. Evans III, N.J. Mylrea, D. Chaudhury, C. Luo, B. Guan, C.T. Pierce, W.R. Gordon, C.M. Wilmot, R.J. Kazlauskas, Larger active site in an ancestral hydroxynitrile lyase increases catalytically promiscuous esterase activity, PloS one, 15 (2020) e0235341.
[129] A.H. Manlove, N.N. Nuñez, S.S. David, The GO repair pathway: OGG1 and MUTYH, in: The Base Excision Repair Pathway: Molecular Mechanisms and Role in Disease Development and Therapeutic Design, World Scientific, 2017, pp. 63-115.
[130] J.A. McCann, P.J. Berti, Transition-state analysis of the DNA repair enzyme MutY, Journal of the American Chemical Society, 130 (2008) 5789-5797.
[131] X.-Y. Chen, P.J. Berti, V.L. Schramm, Transition-state analysis for depurination of DNA by ricin A-chain, Journal of the American Chemical Society, 122 (2000) 6527-6534.
[132] M.K. Brinkmeyer, M.A. Pope, S.S. David, Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays, Chemistry & biology, 19 (2012) 276-286.
[133] R.D. Woods, V.L. O'Shea, A. Chu, S. Cao, J.L. Richards, M.P. Horvath, S.S. David, Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases, Nucleic acids research, 44 (2016) 801-810.
[134] L.S. Pidugu, H. Bright, W.-J. Lin, C. Majumdar, R.P. Van Ostrand, S.S. David, E. Pozharski, A.C. Drohat, Structural Insights into the Mechanism of Base Excision by MBD4, Journal of Molecular Biology, 433 (2021) 167097.
[135] Q. Wang, R. Graham, D. Trimbur, R.A.J. Warren, S. Withers, Changing enzymic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme, Journal of the American Chemical Society, 116 (1994) 11594-11595.
[136] T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid, Biochemistry, 11 (1972) 3610-3618.
[137] T. Lindahl, O. Karlstrom, Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution, Biochemistry, 12 (1973) 5151-5154.
[138] R. Shapiro, S. Kang, Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine, Biochemistry, 8 (1969) 1806-1810.
[139] R. Shapiro, M. Danzig, Acidic hydrolysis of pyrimidine deoxyribonucleotides, Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis, 319 (1973) 5-10.
[140] J. Dong, A.C. Drohat, J.T. Stivers, K.W. Pankiewicz, P.R. Carey, Raman Spectroscopy of Uracil DNA Glycosylase− DNA Complexes: Insights into DNA Damage Recognition and Catalysis, Biochemistry, 39 (2000) 13241-13250.
[141] A.C. Drohat, J. Jagadeesh, E. Ferguson, J.T. Stivers, Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase, Biochemistry, 38 (1999) 11866-11875.
[142] X. Liu, R. Roy, Mutation at active site lysine 212 to arginine uncouples the glycosylase activity from the lyase activity of human endonuclease III, Biochemistry, 40 (2001) 13617-13622.
[143] S.D. Williams, S.S. David, A single engineered point mutation in the adenine glycosylase MutY confers bifunctional glycosylase/AP lyase activity, Biochemistry, 39 (2000) 10098-10109.
[144] T.J. Begley, R.P. Cunningham, Methanobacterium thermoformicicum thymine DNA mismatch glycosylase: conversion of an N-glycosylase to an AP lyase, Protein engineering, 12 (1999) 333-340.

[145] B. Dalhus, M. Forsbring, I.H. Helle, E.S. Vik, R.J. Forstrøm, P.H. Backe, I. Alseth, M. Bjørås, Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase, Structure, 19 (2011) 117-127.
[146] J.C. Fromme, S.D. Bruner, W. Yang, M. Karplus, G.L. Verdine, Product-assisted catalysis in base-excision DNA repair, Nature Structural & Molecular Biology, 10 (2003) 204-211.
[147] A.H. Metz, T. Hollis, B.F. Eichman, DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG), The EMBO journal, 26 (2007) 2411-2420.
[148] C. Cao, K. Kwon, Y.L. Jiang, A.C. Drohat, J.T. Stivers, Solution structure and base perturbation studies reveal a novel mode of alkylated base recognition by 3-methyladenine DNA glycosylase I, Journal of Biological Chemistry, 278 (2003) 48012-48020.
[149] M.C. Weiss, F.L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W.F. Martin, The physiology and habitat of the last universal common ancestor, Nature microbiology, 1 (2016) 1-8.
[150] B. van Loo, S. Jonas, A.C. Babtie, A. Benjdia, O. Berteau, M. Hyvönen, F. Hollfelder, An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily, Proceedings of the National Academy of Sciences, 107 (2010) 2740-2745.
[151] J. Intra, G. Pavesi, D.S. Horner, Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family, BMC evolutionary biology, 8 (2008) 214.
[152] P.J. O'Brien, Catalytic promiscuity and the divergent evolution of DNA repair enzymes, Chemical reviews, 106 (2006) 720-752.
[153] S. Sentürker, M. Dizdaroglu, P.A. van der Kemp, H.J. You, P.W. Doetsch, S. Boiteux, Substrate specificities of the Ntg1 and Ntg2 proteins of Saccharomyces cerevisiae for oxidized DNA bases are not identical, Nucleic acids research, 26 (1998) 5270-5276.
[154] Y.N. Fondufe-Mittendorf, C. Härer, W. Kramer, H.-J. Fritz, Two amino acid replacements change the substrate preference of DNA mismatch glycosylase Mig. Mth I from T/G to A/G, Nucleic acids research, 30 (2002) 614-621.
[155] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Molecular biology and evolution, 35 (2018) 1547-1549.
[156] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface, Nucleic acids research, 36 (2008) W5-W9.
[157] O. Adebali, I.B. Zhulin, Aquerium: A web application for comparative exploration of domain-based protein occurrences on the taxonomically clustered genome tree, Proteins: Structure, Function, and Bioinformatics, 85 (2017) 72-77.
[158] L. Sumbalova, J. Stourac, T. Martinek, D. Bednar, J. Damborsky, HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information, Nucleic acids research, 46 (2018) W356-W362.

Figure captions.

Figure 1. Helix-hairpin-Helix DNA glycosylase evolution and substrate specificity. The phylogenetic tree (A) was constructed with ML algorithm from an amino acid sequence alignment constructed with MUSCLE algorithm in MEGA software [155]. For the taxonomic distribution (B) the amino acid sequences analyses were obtained from NCBI database with iterated profile searches with PSI-BLAST algorithm [156] and Aquerium server [157]. The substrate preference

and bases removed by each HhH-DG (C) is indicated with colored circles; EndoIII/red, MagI/gray, MBD4/green, MagIII/golden, AlkA/pink, OGG1/blue, MIG/cyan, MutY/orange TAG/light gray. DEMETER DNA glycosylase activity for 5-methylcytosine is indicated.

Figure 2. Sequence and structural similarity of HhH-DGs (A) Schematic representation of amino acid sequence alignment and motifs of HhH-DGs. (B) Sequence and structural analyses of the active site of HhH-DGs illustrate locations of key residues R1-R4. (C) Structures of HhH-DGs highlighting particular domains and motifs. Structure/PDB ID; MagI/3S6I [118], MagIII/1PU8 [117], TAG/2OFI [147], MIG/1KEA [14], EndoIII/1ORN [13], MutY/3G0Q [12], OGG1/3KTU [To be published], AlkA/1DIZ [17], MBD5/4E9F [86].

Figure 3. MutY has an HhH-DG and MutT-like domain (A) *G. stearothermophilus* MutY X-ray crystal structure in complex with DNA containing transition state analog 1N across OG (PDB ID; 5DPK) [133]. The hydrogen bond network implied in the recognition of OG (gray sticks) by the catalytic domain (pink) and Ser308 within the FSH loop (red) is displayed. The DNA phosphodiester backbone is shown in gray. (B) Structural and sequence alignment of MutT-like domain of MutY and MutT protein. The d(OG)TP (orange sticks) recognition conformation within MutT enzyme (PDB ID; 3A6U [63]) is highlighted.

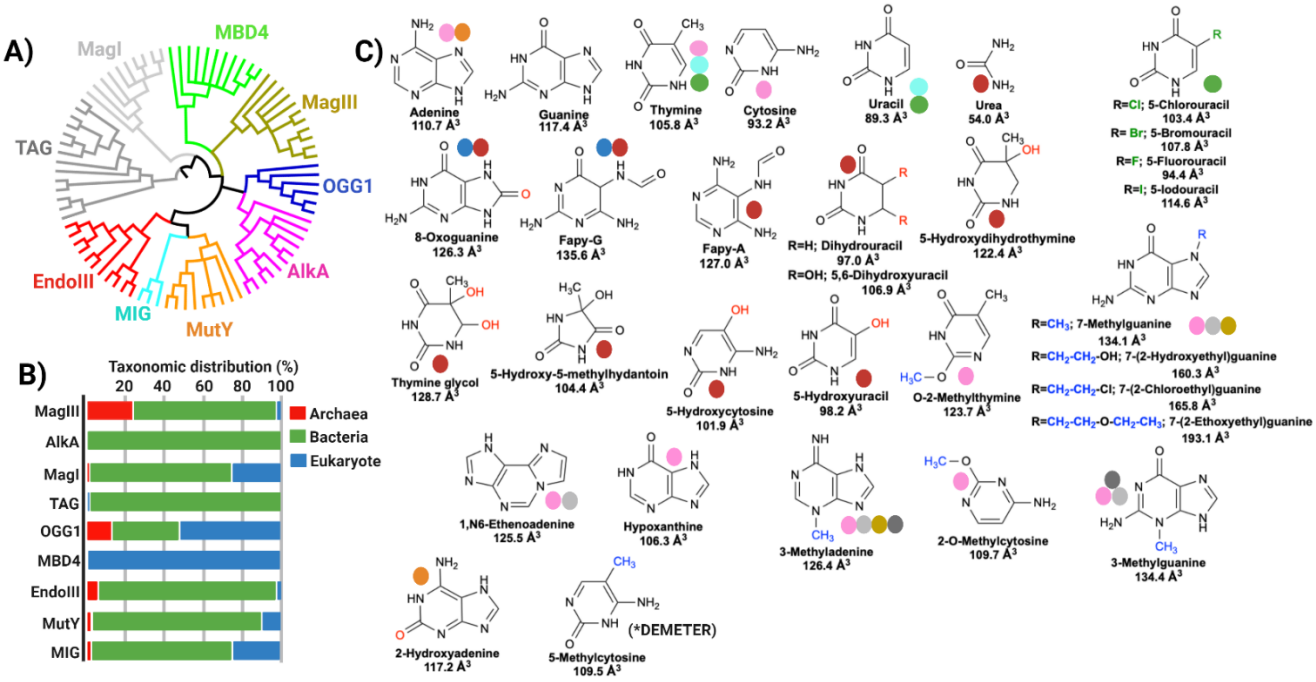
Figure 4. Shapes and Sizes of Catalytic Pockets in HhH-DGs (A) Catalytic pocket architecture of HhH-DGs. The active site pocket topologies and volumes were calculated with Hotspot Wizard 3.0 server using only DNA-protein complexes to aid with pocket boundary delimitation [158]. Important residues for catalysis and lesion recognition are shown on red sticks. The catalytic pocket of MutY and OGG1 are shown with fluorinated adenine (fA) and OG (fOG) as in the lesion recognition complexes from 3G0Q [12] and 3KTU PDB entries, respectively. (B) Catalytic pocket/substrate volumes relationship plot based on the substrates displayed in Figure 1.

Figure 5. Key Residues involved in Lesion Recognition. (A) Logo representation of sequence alignments of active site components of HhH-DG superfamily. Lower panel shows the schematic representation of MutY (B) and OGG1 (C) lesion recognition complexes from 3G0Q [12] and 3KTU PDB entries. MutY and OGG1 are in complex with fluorinated adenine (fA) and OG (fOG), respectively.

Figure 6. Unified mechanisms for HhH-DGs (A) Proposed mechanism for MutY: The catalytic residue E43 acts as a general acid to protonate the base which is followed by the glycosidic bond cleavage and formation of an oxacarbenium ion intermediate. D144 is proposed to attack the oxacarbenium ion at C1' to stabilize it by a covalent intermediate. Finally, E43 activates a water molecule for nucleophilic attack at C1 to form AP site product [133]. There may be similar "retaining" mechanisms for other HhH-DGs (B) A proposed general mechanism for "inverting" monofunctional HhH-DGs would follow an S_N1-like mechanism where an oxacarbenium ion intermediate is formed with the N-glycosidic bond cleavage. A water molecule attacks C1 to form the AP site product in alpha position. R1* corresponds to a Glu only in MutY and MIG. (C) A general mechanism for bifunctional HhH enzymes is shown with the base excision step similar to monofunctional glycosylases with TS/intermediate intercepted by a Lys residue, rather than water. The resulting Schiff-base formed with the Lys undergoes enzyme-catalyzed β-elimination leading to a strand break at 3' position. B* in the mechanism has been proposed to be OG in OGG1 mechanism, while a water molecule activated by Asp45 in EndoIII [13, 146].

Figure 7. Hypothetical evolutionary pathway of HhH-DG superfamily. Red double-headed arrows indicate gene duplication events, single arrows; domain or motif acquisitions, and colored squares; substrate promiscuity degree. Substrate preferences are in colored boxes. The degree of substrate promiscuity is indicated with colored bar. LCA: Last common ancestor. LUCA: Last universal common ancestor. The types of chemical base modifications that are removed by HhH-DGs are shown in boxes with different shading: green/alkylated lesions, green/oxidized lesions, blue/canonical bases within mispairs, yellow/halogenated bases and pink/canonical bases.

Figure 1


A)
MBD4
MagI
MagIII
TAG
OGG1
EndoIII
AlkA
MIG
MutY
B)
Taxonomic distribution (%)
20
40
60
80
100
MagIII
AlkA
MagI
TAG
OGG1
MBD4
EndoIII
MutY
MIG
Archaea
Bacteria
Eukaryote
C)
Adenine 110.7 Å3
Guanine 117.4 Å3
Thymine 105.8 Å3
Cytosine 93.2 Å3
Uracil 89.3 Å3
Urea 54.0 Å3
R=Cl; 5-Chlorouracil 103.4 Å3
R= Br; 5-Bromouracil 107.8 Å3
R=F; 5-Fluorouracil 94.4 Å3
R=I; 5-Iodouracil 114.6 Å3
8-Oxoguanine 126.3 Å3
Fapy-G 135.6 Å3
Fapy-A 127.0 Å3
R=H; Dihydrouracil 97.0 Å3
R=OH; 5,6-Dihydroxyuracil 106.9 Å3
5-Hydroxydihydrothymine 122.4 Å3
R=CH3; 7-Methylguanine 134.1 Å3
R=CH2-CH2-OH; 7-(2-Hydroxyethyl)guanine 160.3 Å3
R=CH2-CH2-Cl; 7-(2-Chloroethyl)guanine 165.8 Å3
R=CH2-CH2-O-CH2-CH3; 7-(2-Ethoxyethyl)guanine 193.1 Å3
Thymine glycol 128.7 Å3
5-Hydroxy-5-methylhydantoin 104.4 Å3
5-Hydroxycytosine 101.9 Å3
5-Hydroxyuracil 98.2 Å3
O-2-Methylthymine 123.7 Å3
1,N6-Ethenoadenine 125.5 Å3
Hypoxanthine 106.3 Å3
3-Methyladenine 126.4 Å3
2-O-Methylcytosine 109.7 Å3
3-Methylguanine 134.4 Å3
2-Hydroxyadenine 117.2 Å3
5-Methylcytosine 109.5 Å3
(*DEMETER)

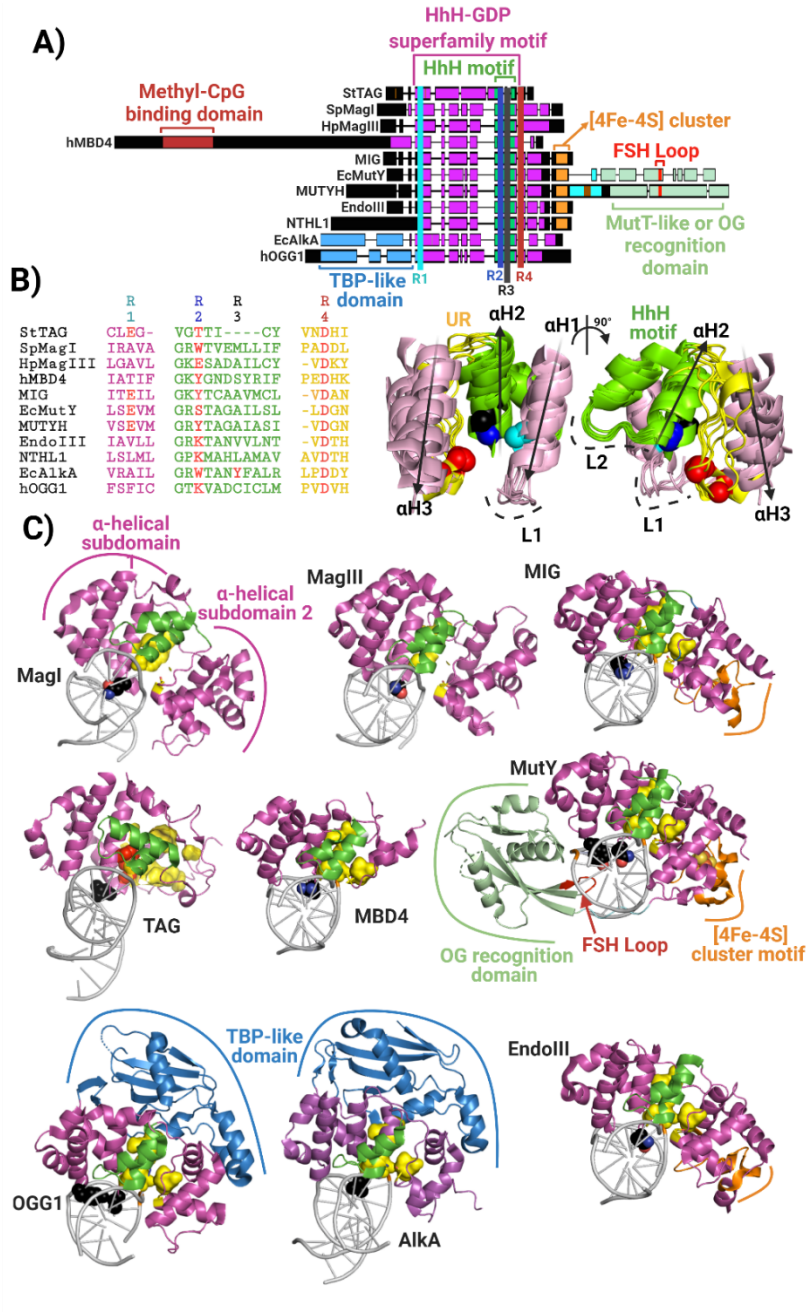

A)
HhH-GDP
superfamily motif
HhH motif
Methyl-CpG
binding domain
hMBD4
StTAG
SpMagI
HpMagIII
MIG
EcMutY
MUTYH
EndoIII
NTHL1
EcAlkA
hOGG1
[4Fe-4S] cluster
FSH Loop
MutT-like or OG
recognition
domain
TBP-like
domain
R1
R2
R3
R4
B)
R 1
R 2
R 3
R 4
StTAG CLEG- VGTTI----CY VNDHI
SpMagI IRAVA GRWTVEMLLIF PADDL
HpMagIII LGAVL GKESADAILCY -VDKY
hMBD4 IATIF GKYGNDSYRIF PEDHK
MIG ITEIL GKYTCAAVMCL -VDAN
EcMutY LSEVM GRSTAGAILSL -LDGN
MUTYH VSEVM GRYTAGAIASI -VDGN
EndoIII IAVLL GRKTANVVLNT -VDTH
NTHL1 LSLML GPKMAHLAMAV AVDTH
EcAlkA VRAIL GRWTANYFALR LPDDY
hOGG1 FSFIC GTKVADCICLM PVDVH
UR
αH2
αH1
90°
HhH
motif
αH2
L2
αH3
L1
L1
αH3
C)
α-helical
subdomain
α-helical
subdomain 2
MagI
MagIII
MIG
MutY
TAG
MBD4
OG recognition
domain
FSH Loop
[4Fe-4S]
cluster motif
TBP-like
domain
EndoIII
OGG1
AlkA

Figure 2

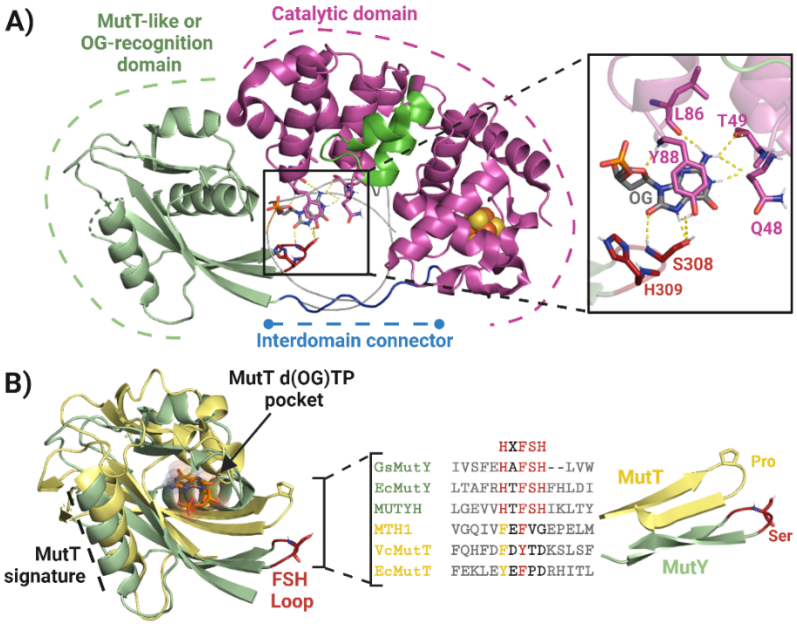

A)
MutT-like or
OG-recognition
domain
Catalytic domain
L86
T49
Y88
OG
Q48
S308
H309
Interdomain connector
B)
MutT d(OG)TP
pocket
MutT
signature
FSH
Loop
HXFSH
GsMutY IVSFEHAFSH--LVW
EcMutY LTAFRHTFSHFHLDI
MUTYH LGEVVHTFSHIKLTY
MTH1 VGQIVFEFVGEPELM
VcMutT FQHFDFDYTDKSLSF
EcMutT FEKLEYEFPDRHITL
MutT
Pro
Ser
MutY

Figure 3

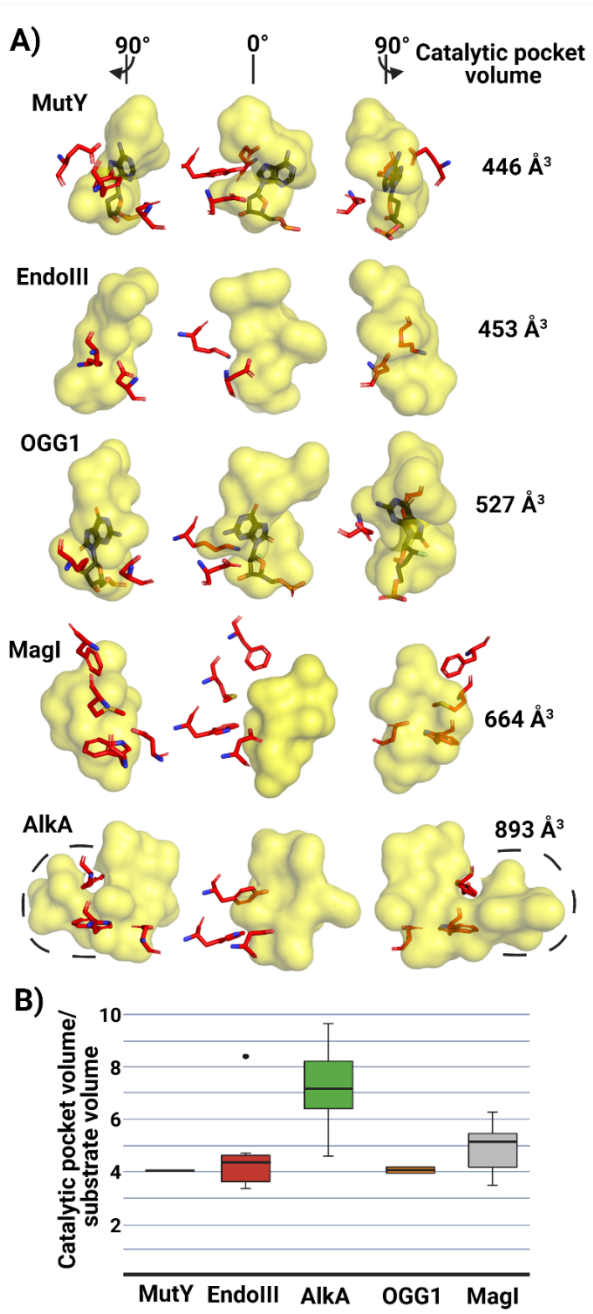

A)
90°
0°
90°
Catalytic pocket volume
MutY
446 Å³
EndoIII
453 Å³
OGG1
527 Å³
MagI
664 Å³
AlkA
893 Å³
B)
Catalytic pocket volume/ substrate volume
10
8
6
4
2
MutY
EndoIII
AlkA
OGG1
MagI

Figure 4

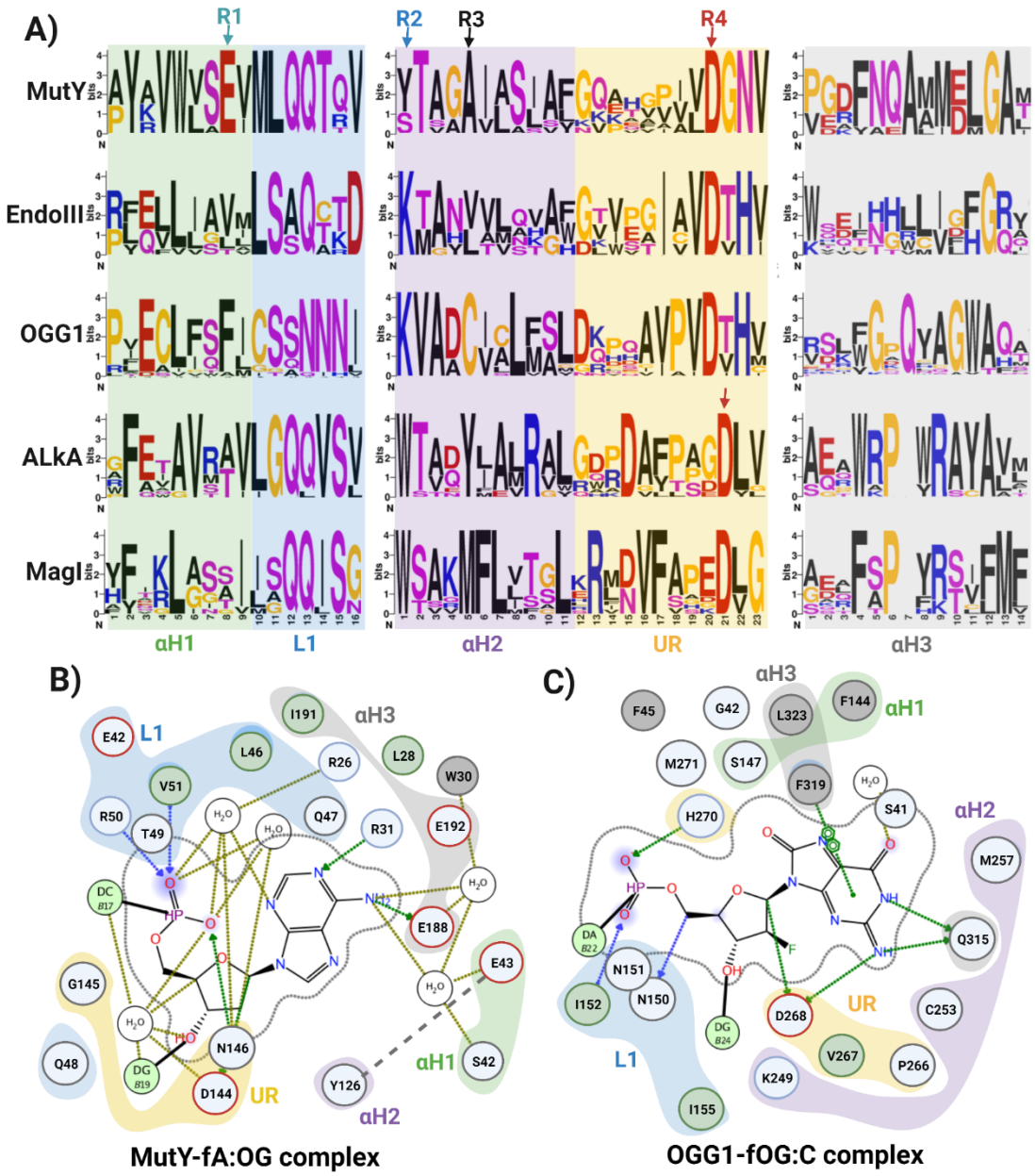

A)
R1
R2
R3
R4
MutY
EndoIII
OGG1
ALkA
MagI
αH1
L1
αH2
UR
αH3
B)
E42
L1
I191
αH3
L46
R26
L28
W30
V51
R50
T49
Q47
R31
E192
DC
817
E188
E43
G145
Q48
DG
819
N146
D144
UR
Y126
αH2
αH1
S42
MutY-fA:OG complex
C)
αH3
F45
G42
L323
F144
αH1
M271
S147
F319
S41
αH2
H270
M257
DA
822
Q315
N151
N150
I152
DG
824
D268
UR
C253
L1
V267
P266
K249
I155
OGG1-fOG:C complex

Figure 5

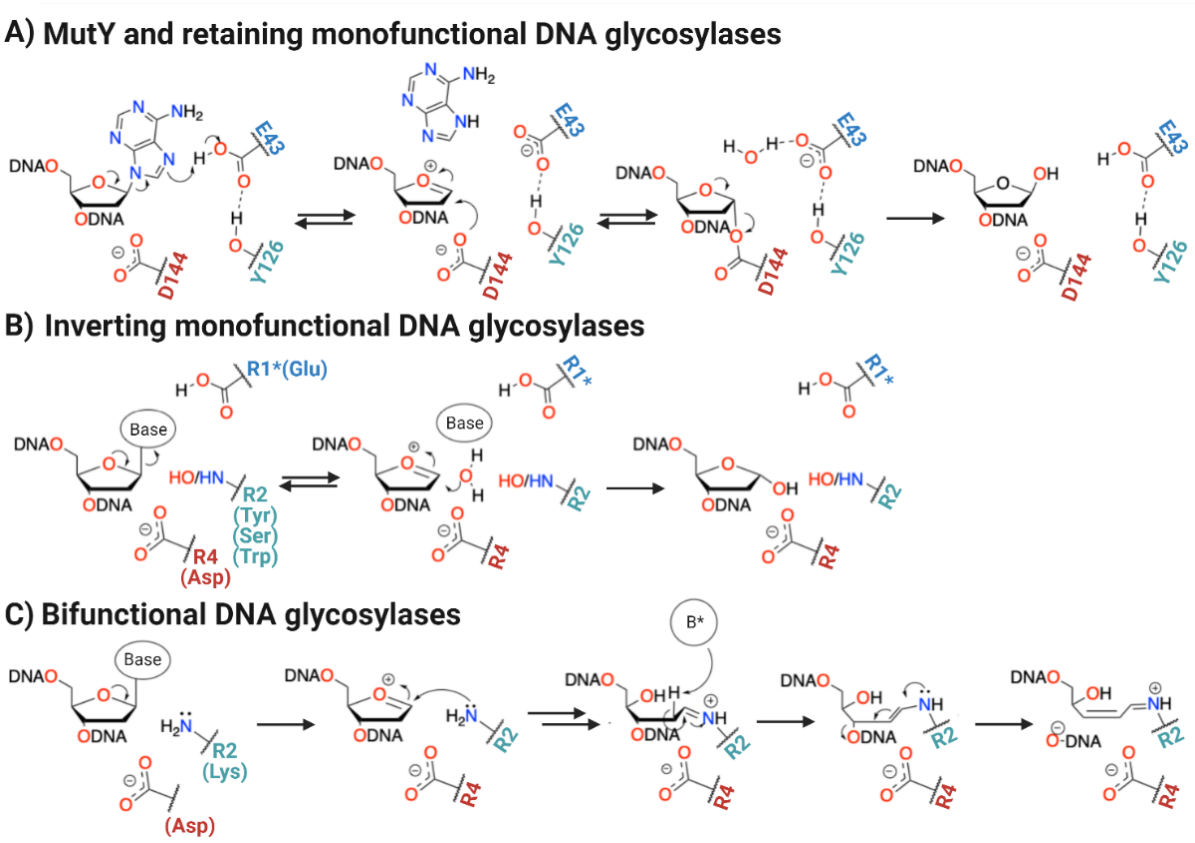

A) MutY and retaining monofunctional DNA glycosylases
B) Inverting monofunctional DNA glycosylases
C) Bifunctional DNA glycosylases

Figure 6

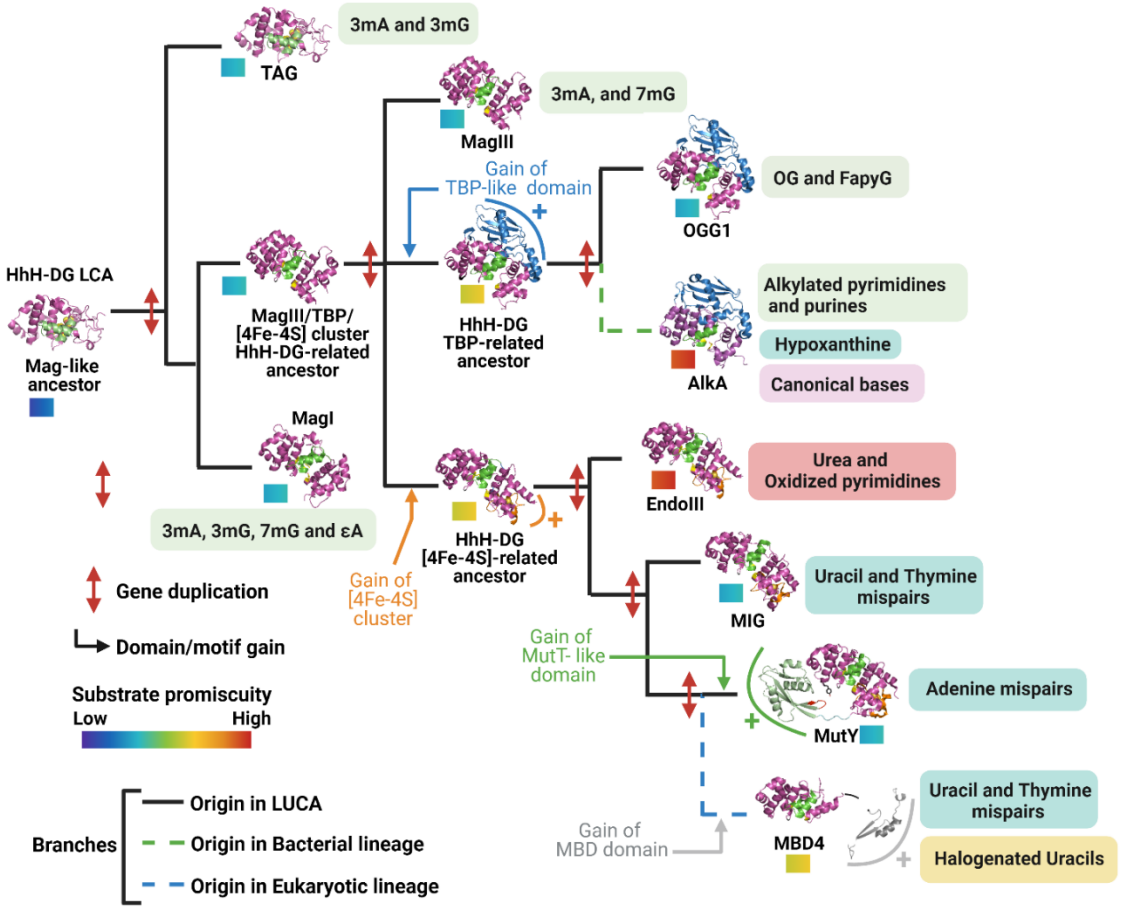

3mA and 3mG
TAG
3mA, and 7mG
MagIII
Gain of
TBP-like domain
OG and FapyG
OGG1
HhH-DG LCA
MagIII/TBP/
[4Fe-4S] cluster
HhH-DG-related
ancestor
HhH-DG
TBP-related
ancestor
Alkylated pyrimidines
and purines
Hypoxanthine
Canonical bases
AlkA
Mag-like
ancestor
MagI
3mA, 3mG, 7mG and εA
HhH-DG
[4Fe-4S]-related
ancestor
Gain of
[4Fe-4S]
cluster
EndoIII
Urea and
Oxidized pyrimidines
Gene duplication
Domain/motif gain
Substrate promiscuity
Low
High
MIG
Uracil and Thymine
mispairs
Gain of
MutT- like
domain
MutY
Adenine mispairs
Gain of
MBD domain
MBD4
Uracil and Thymine
mispairs
Halogenated Uracils
Branches
Origin in LUCA
Origin in Bacterial lineage
Origin in Eukaryotic lineage

Figure 7