



Reading targeted DNA damage in the active demethylation pathway: role of accessory domains of eukaryotic AP endonucleases and thymine-DNA glycosylases

Alexander V Popov, Inga R Grin, Antonina P Dvornikova, Bakhyt T Matkarimov, Regina Groisman, Murat A Saparbaev, Dmitry O Zharkov

► To cite this version:

Alexander V Popov, Inga R Grin, Antonina P Dvornikova, Bakhyt T Matkarimov, Regina Groisman, et al.. Reading targeted DNA damage in the active demethylation pathway: role of accessory domains of eukaryotic AP endonucleases and thymine-DNA glycosylases. *Journal of Molecular Biology*, 2020, 432 (6), pp.1747-1768. 10.1016/j.jmb.2019.12.020 . hal-03060641

HAL Id: hal-03060641

<https://hal.science/hal-03060641>

Submitted on 14 Dec 2020

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.

**Reading targeted DNA damage in the active demethylation
pathway: role of accessory domains of eukaryotic AP endonucleases
and thymine-DNA glycosylases**

**Alexander V. Popov¹, Inga R. Grin^{1,2}, Antonina P. Dvornikova^{1,2},
Bakhyt T. Matkarimov³, Regina Groisman^{4,5}, Murat Saparbaev^{4,5*},
Dmitry O. Zharkov^{1,2,*}**

¹SB RAS Institute of Chemical Biology and Fundamental Medicine, 8 Lavrentieva Ave.,
Novosibirsk 630090, Russia

²Department of Natural Sciences, Novosibirsk State University, 2 Pirogova St.,
Novosibirsk 630090, Russia

³National Laboratory Astana, Nazarbayev University, 53 Kabanbay Batyr Ave., Astana 010000,
Kazakhstan

⁴Groupe «Réparation de l'ADN», Equipe Labellisée par la Ligue Nationale contre le Cancer,
CNRS UMR8200, Université Paris-Sud, Université Paris-Saclay, F-94805 Villejuif Cedex,
France

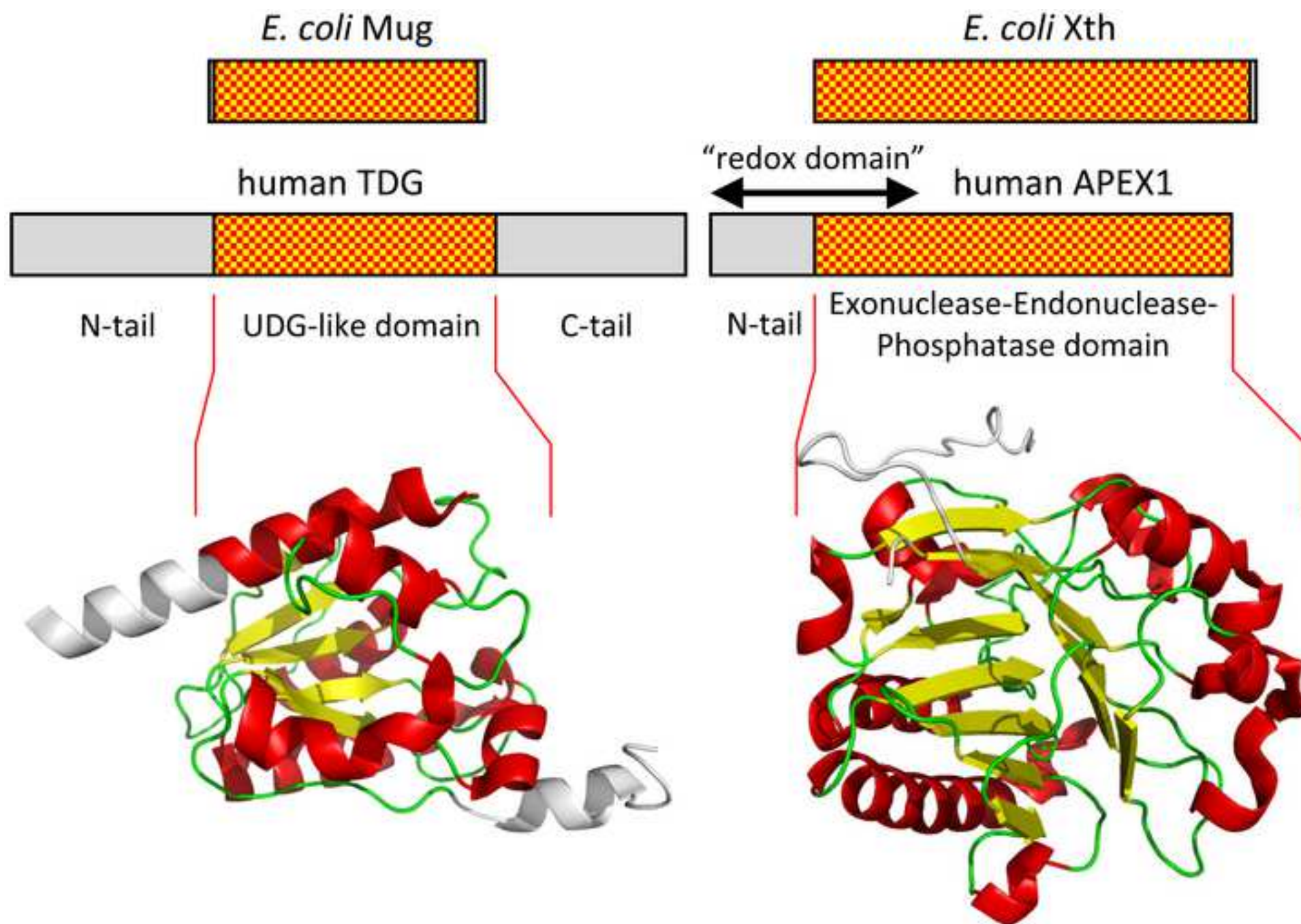
⁵Gustave Roussy Cancer Campus, F-94805 Villejuif Cedex, France

*To whom 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

Keywords: DNA repair; DNA glycosylases; AP endonucleases; protein domains; structural
disorder

ABBREVIATIONS

ROS, reactive oxygen species; AP, apurinic/apyrimidinic site; BER, base excision repair;
NIR, nucleotide incision repair; APE1 (a.k.a. APEX1, HAP-1, or Ref-1), human major AP
endonuclease 1; TDG, T:G mismatch-specific thymine-DNA glycosylase. mC, 5-
methylcytosine; hmC, 5-hydroxymethylcytosine; fC, 5-formylcytosine; caC, 5-carboxylcytosine;
fU, 5-formyluracil; hmU, 5-hydroxymethyluracil; TF, transcription factor; CGIs, CpG islands;
Rrp1, *Drosophila* AP endonuclease 1.



Highlights_Popov et al 2019

HIGHLIGHTS

- Eukaryotic BER proteins contain extended non-catalytic tails absent from bacterial and archaeal homologs;
- The tails of TDG and APE1 homologs are more diverse than their catalytic cores and contain diverged short repeats;
- The recombination function cannot be universally ascribed to the N-tail of insect APE1 homologs;
- The tails of TDG and APE1 participate in the formation of oligomeric aggregates on DNA and enable active epigenetic demethylation;
- The APE1 and TDG aggregates on DNA might be involved in liquid phase separation.

Abstract

Base excision DNA repair (BER) is an important process used by all living organisms to remove non-bulky lesions from DNA. BER is usually initiated by DNA glycosylases that excise a damaged base leaving an apurinic/apyrimidinic (AP) site, an AP endonuclease then cuts DNA at the AP site, and the repair is completed by correct nucleotide insertion, end processing, and nick ligation. It has emerged recently that the BER machinery, in addition to genome protection, is crucial for active epigenetic demethylation in vertebrates. This pathway is initiated by TET dioxygenases that oxidize the regulatory 5-methylcytosine, and the oxidation products are treated as substrates for BER. T:G mismatch-specific thymine-DNA glycosylase (TDG) and AP endonuclease 1 (APE1) catalyze the first two steps in BER-dependent active demethylation. In addition to the well-structured catalytic domains, these enzymes possess long tails that are structurally uncharacterized but involved in multiple interactions and regulatory functions. In this review, we describe known roles of the tails in TDG and APE1, discuss the importance of order and disorder in their structure and consider the evolutionary aspects of these accessory protein regions. We also propose that the tails may be important for the enzymes' oligomerization on DNA, an aspect of their function that only recently gained attention.

1. Introduction

DNA repair, a system of several partially overlapping enzymatic pathways, has evolved to protect genetic material from ongoing damage and is present in some form in all cellular organisms [1]. Of these pathways, base excision repair (BER) removes small non-bulky lesions from DNA; lesions of this kind are the most abundant source of mutations and are mostly unavoidable since they are generated through nucleotide hydrolysis and oxidation [1-4]. In the simplest form of BER, the damaged base is located and excised by one of the lesion-specific enzymes belonging to the group of DNA glycosylases, the nascent abasic (apurinic/apyrimidinic, AP) site is cut by an AP endonuclease, the correct dNMP is inserted by a DNA polymerase with simultaneous removal of the hanging deoxyribose remnant, and the resulting nick is sealed by a DNA ligase [1-4]. Several variations on this general mechanism operate in mammalian cells, differing in the requirements for DNA glycosylases and AP endonucleases and generating single-strand breaks with different termini [2-5]. Deficiencies in the key components of BER are embryonic lethal [6-7] while certain functional variants are associated with an increased risk of cancer [8-10].

In recent years, much attention has been drawn to non-canonical functions of BER, not related to genome integrity maintenance in a strict sense. Obviously, the versatile BER machinery can be easily adopted to processing regulated targeted DNA modification in the cell. BER has been shown to be involved in somatic hypermutation and class switch recombination during immunoglobulin gene maturation [11-12], virus life cycle and cell antiviral defense [13-14] and active epigenetic cytosine demethylation [15-17]. In higher eukaryotes, this last pathway is initiated when the regulatory DNA base 5-methylcytosine is oxidized by TET family dioxygenases and/or deaminated by AID and APOBEC family deaminases and then removed preferentially by mismatched thymine-DNA glycosylase, TDG. The resulting AP site is then

hydrolyzed by the major AP endonuclease APE1 and further processed *via* regular BER mechanisms.

Many protein molecules in living organisms consist of two or more structural domains, tens to hundreds amino acids long [18-19]. A domain is usually defined as a separate protein part that has a defined structure and function and may evolve largely independently of the rest of the protein. In the case of enzymes, their catalytic activity is often localized in one or several domains, which form a core fragment, while other domains of the same protein may participate in activity regulation, protein–protein interactions, localization in the cell, etc. In multidomain proteins, often the function of the core fragment is known whereas the roles of other domains are unclear. The development of methods for protein domain prediction based on homology and structural modeling [20-21] led to description of hundreds of domains of unknown functions. Moreover, there are many cases when a protein possesses known domains and long tails or linkers that are not identified as domains. Structurally they may be true domains (i. e., have a defined structure and be separated from other domains), but more often they are disordered and either not solved by X-ray crystallography or prevent crystallization and thus have to be removed to solve the core structure of the protein. The lack of structural information much complicates the analysis of their functions.

Available structural data show that many eukaryotic BER enzymes consist of a well-resolved core domain (or several domains) responsible for catalysis and/or DNA binding, and terminal tails that are, in most cases, disordered [22-23]. These tails generally can be removed without affecting the activity of the enzyme, and the bacterial counterparts of eukaryotic BER enzymes are most often compact proteins with little extra sequence outside the catalytic domain borders. The functions of these tails in eukaryotic BER, except in a very few occurrences, remain enigmatic but they are usually considered to be involved either in protein–protein interactions [24-25] or in non-specific DNA binding contributing to the affinity of the protein for DNA or facilitating its movement along DNA [26-27].

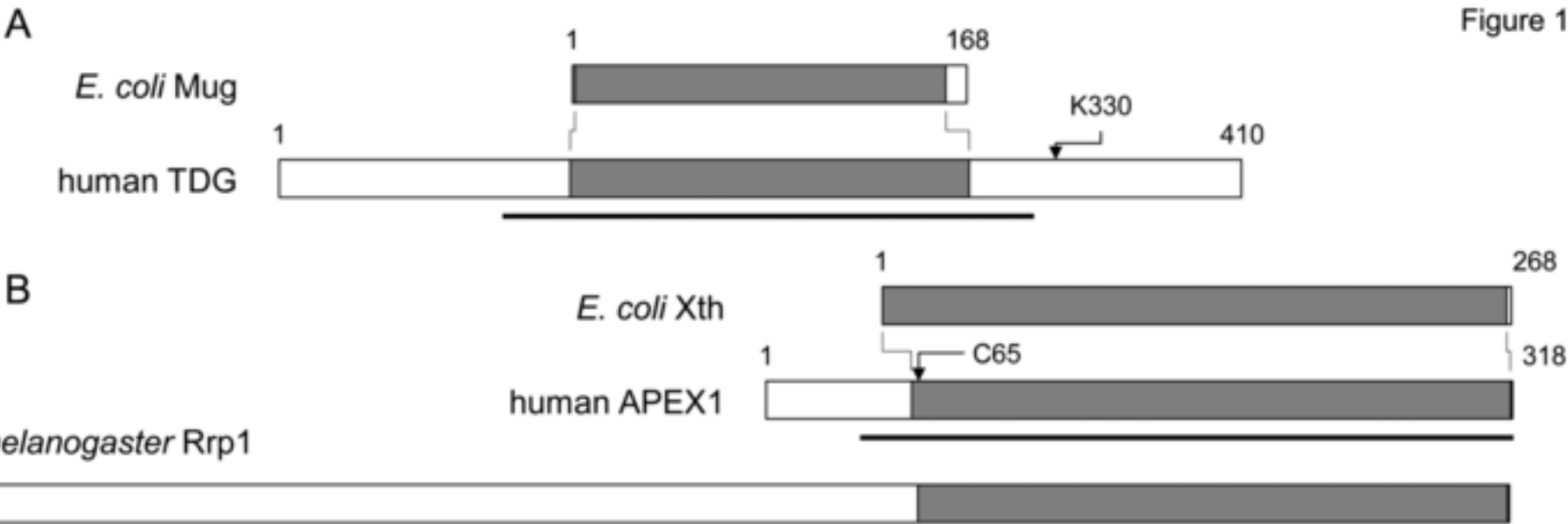
Based on the structure of the core domain, human DNA glycosylases can be assigned to four superfamilies: UDG-like (uracil-DNA glycosylase homologs), HhH (containing a helix–hairpin–helix DNA binding motif), H2tH (containing a helix–two-turn–helix DNA binding motif), and FMT_C (homologs of the C-terminal domain of methionyl-tRNA formyltransferase). AP endonucleases belong to two major superfamilies, EEP (exonuclease–endonuclease–phosphatase) and TIM (triosephosphate isomerase) barrel. Human TDG and APE1, which initiate the BER stage of active DNA demethylation, belong to the UDG-like and EEP superfamilies, respectively. Both TDG and APE1 fit the structural paradigm of the ordered core domain and disordered tails (Figure 1). In this review, we summarize what is known about the functions of the tails of TDG and APE1 and discuss their evolution and the possible role in the regulation of epigenetic demethylation¹.

2. TDG: a hub of epigenetic DNA demethylation

Methylation of cytosine at the C5 position with the formation of 5-methylcytosine (mC) in DNA provides the molecular basis of epigenetic regulation of gene expression in several major groups of eukaryotes [28-30]. DNA methylation is essential for the organism development, cell differentiation, genomic imprinting and suppression of mobile genetic elements. A drawback of this mode of regulation is that spontaneous deamination of mC generates thymine, resulting in a T:G mismatch which, if not repaired, leads to C→T transitions at CpG dinucleotides [31]. In mammalian cells, two DNA glycosylases, mismatch-specific thymine–DNA glycosylase (TDG) and methyl-binding domain protein 4 (MBD4/MED1) prevent the mutagenic impact of mC deamination by excising T from T:G mispairs in CpG context thus initiating BER that restores C [32-33].

¹ In the literature, the term “domain” is often used indiscriminately to describe true structural domains, or structurally non-discrete parts of the protein responsible for specific functions, or disordered regions. Here, we will consistently refer to structural domains as “domains” (or “cores” specifically when describing the main catalytic domains), parts responsible for specific functions as “parts”, and terminal disordered regions as “tails”.

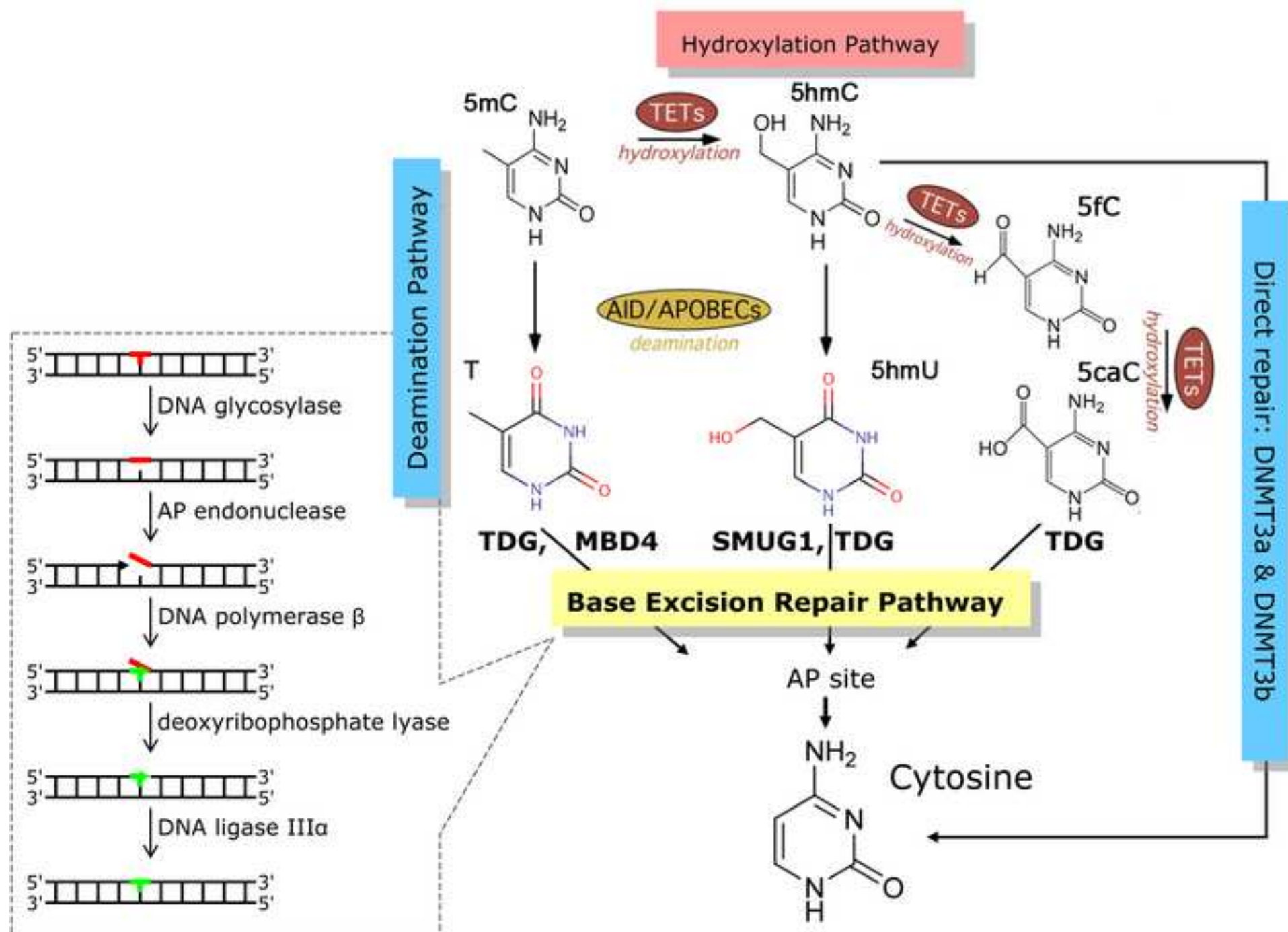
Figure 1
[Click here to download high resolution image](#)



1 T:G mismatch-specific thymine–DNA glycosylase (TDG) was initially discovered as an
2 enzymatic activity excising Thy bases from T:G mispairs [32,34-35], purified and cloned,
3 revealing significant homology with *E. coli* mismatched uracil–DNA glycosylase (Mug) [36].
4 Subsequent discovery of additional Mug/TDG homologs showed that they form a separate
5 family within the uracil–DNA glycosylase-like superfamily [37-39]. A more detailed enzyme
6 activity inquiries showed that TDG exhibits wide substrate specificity, excising 3,*N*⁴-
7 ethenocytosine [40-41], thymine glycol [42], 5-hydroxycytosine [43], 7,8-dihydro-8-oxoadenine
8 [44], mismatched U [35] and its C5-modified derivatives [43,45]. In contrast, MBD4 has narrow
9 DNA substrate specificity, excising T, U, 5-fluorouracil and 5-hydroxymethyluracil in a XpG
10 context from duplex DNA [33,46-47].

22 Until recently, TDG had been considered as functional “back-up” for the major vertebrate
23 uracil–DNA glycosylase, UNG, with somewhat broadened damaged base specificity. The
24 discovery of the central place of TDG in active epigenetic demethylation revealed what is likely
25 the true biological function of at least the vertebrate homologs; no similar role outside the basic
26 BER is presently known for bacterial Mug proteins. Active DNA demethylation in vertebrates is
27 dependent on the Ten-eleven translocation family of proteins (TETs; TET1, TET2, and TET3),
28 iron/ α -ketoglutarate-dependent dioxygenases that convert mC to 5-hydroxymethylcytosine
29 (hmC) and then further oxidize it to 5-formylcytosine (fC) and 5-carboxylcytosine (caC) [48-51]
30 (Figure 2). TDG excises fC and caC from CpG context with high efficiency, suggesting direct
31 involvement of the TDG-initiated BER in the active erasure of mC epigenetic marks [50,52-57].
32 In addition, mC and its C5-oxidized derivatives can be deaminated by enzymes belonging to the
33 activation-induced cytidine deaminase (AID) / Apolipoprotein B mRNA-editing enzyme
34 complex (APOBEC) family of cytidine deaminases [58-60], with at least 5-hydroxymethyluracil
35 (hmU) and 5-formyluracil (fU) also being substrates for TDG [45,53-54,61-62]. In agreement
36 with these observations, *Tdg* knockouts in mice are embryonic lethal due to aberrant *de novo*
37 DNA methylation of CpG islands in promoters of developmental genes and failure to establish

Figure 2
[Click here to download high resolution image](#)



and/or maintain cell type-specific gene expression programs during embryonic development [62-63].

In addition to TETs/TDG-dependent active erasure of 5mC residues, a second mechanism was shown, in which AID catalyses deamination of 5mC to thymine generating a G•T mismatch base pair that is repaired by MBD4 [58]. Moreover, it was shown that AID/APOBEC catalyze deamination of 5hmC to hmU residue, which is in turn excised by the MBD4, and single-strand specific monofunctional uracil-DNA glycosylase 1 (SMUG1) [59,62]. MBD4/MED1 is a bipartite protein that belongs to the family of methyl-CpG-binding domain (MBD) proteins, and consists of an N-terminal MBD domain that is linked to a C-terminal DNA glycosylase domain [33,46]. Mutations of MBD4 gene were identified in tumors with defective DNA mismatch repair, however, knockout of MBD4 in mouse causes only a small 2-3 fold increase in C→T mutations at CpG sites and did not increase mini-satellite instability implying that MBD4 may act as a modifier and not as driver of tumorigenesis [64-65]. MBD4 interacts with the mismatch repair protein MLH1 [66] and DNA methyltransferases DNMT1 and DNMT3b [67-68] suggesting a potential link between post-replication repair and DNA methylation. Noteworthy, the interaction between MBD4 and DNMT1 is involved in the repair of oxidative DNA damage and control of gene expression at methylated CpG islands [69]. These observations suggest that MBD4 is a back up DNA glycosylase involved in the epigenetic regulation *via* removal of oxidation and deamination products of 5mC.

Structurally, TDG consists of the Mug-like catalytic core 171 amino acid residues (aa) long and two tails, N-terminal (N-tail, 123 aa) and C-terminal (C-tail, 116 aa). The available crystal structures of TDG leave the tails mostly unresolved (Figure 1) [54-56,70-76]. NMR studies suggest that the N-tail is largely disordered, however, its part proximal to the catalytic core shows interactions both with the core and with bound DNA [75,77]. No experimental information regarding the C-tail is available.

From the moment of TDG discovery, the N-tail has been suspected to be somehow involved in the substrate specificity: progressive deletions into the N-tail eventually abolish the activity on T:G mismatches but hardly affect or even modestly increase the uracil excision from U:G mismatches [37]. Residues 82–110 appear to be critical for the efficient T:G binding by TDG [75]. Additionally, substrate processing by TDG can be regulated by the N-tail at the level of enzyme turnover: full-length TDG has very high affinity for the AP site left after base excision and normally requires stimulation by APE1 for product release [78-79], whereas deletion of the N-tail dramatically increases the turnover on the U:G substrate [80]. The importance of the N-tail for the processing of substrates other than T:G and U:G remains largely unknown.

Another factor facilitating TDG enzymatic turnover is its covalent modification by small ubiquitin-like modifier proteins SUMO-1 and SUMO-2/3 at Lys330 in the C-tail, a residue absolutely conserved within Deuterostomia [80-81]. The sumoylation drastically reduces TDG affinity for AP sites or normal and damaged DNA and modifies its substrate specificity, enhancing the cleavage of U:G and suppressing the cleavage of T:G mismatches [81]. It also improves TDG stimulation by APE1 [81]. Weakening a network of hydrogen bonds between SUMO and the C-tail by site-directed mutagenesis abolishes the effect on sumoylation on DNA binding [71]. A comparison of the structures of unmodified and sumoylated TDG [71,74-75] suggest that the rigidly attached SUMO molecule clashes with DNA near the lesion, which likely should lead to the protein–DNA complex falling apart. Noteworthy, the sumoylation effect on TDG turnover also requires an intact N-tail [80,82], and a site for non-covalent SUMO binding is found in the N-terminal part of the catalytic core [82]. As evidenced by NMR, covalent conjugation and non-covalent binding of SUMO to the respective sites induce distinct conformational changes in the TDG tails [83]. During active demethylation, SUMO1 is first attached to XRCC1, the main BER scaffold protein, which enhances the assembly of the BER

complex involving TDG, to which SUMO1 is later transferred to promote further repair steps [84].

Disordered tails in proteins are often involved in protein–protein interactions. TDG is implicated in the regulation of transcription by structurally diverse factors such as retinoic acid receptors (RAR α /RXR α) [85], estrogen receptor ER α [86], several other nuclear receptor family proteins [87], basic leucine zipper protein c-Jun [85,88], homeodomain-containing thyroid transcription factor 1 [89], p53 and p73 [90]. The interactions with the individual transcription factors have been mapped to the catalytic core. However, TDG complexes with transcription factors can be elaborated by including transcriptional co-activators – CBP/p300 lysine acetylases that initiate chromatin remodeling. Taken alone, TBP can bind CBP/p300 through both N- and C-tails [91]. The resulting complex is competent in both BER and histone acetylation. TDG enhances the CBP/p300 transcriptional activity and is in turn acetylated by CBP/p300, negatively regulating the recruitment of APE1. The TDG–CBP/p300 complexes can also include RAR α /RXR α [92] and LEF1/TCF4, a complex involved in the Wnt/ β -catenin signaling pathway [93]. Although the role of the glycosylase activity is not evident in all these cases, multiple interactions of TDG with transcription regulation machinery including CBP/p300 may couple DNA demethylation with chromatin unfolding for gene activation.

Finally, TDG expression is strictly cell-cycle regulated: it is present in cells throughout the G2-M and G1 phases, but rapidly disappears in the S phase [94]. This regulation involves TDG proteasomal degradation initiated by the association of TDG with PCNA through the “PIP degron”, a PCNA-interacting peptide that triggers recognition by CRL4^{Cdt2} E3 ubiquitin ligase upon binding [95-96].

CpG islands (CGIs) are short interspersed CG-rich genomic regions that are abundant near the transcription start sites of genes [97]. In mammalian genomes, CGIs are typically 500–3,000 base pairs in length, and have been found in or near half of the promoters [98]. As key elements in transcription regulation, CGIs are enriched in permissive histone modifications, poor

in DNA cytosine methylation and contain multiple sites for transcription factors [99]. Despite CpG are significantly underrepresented in mammalian genomes, their enrichment at transcription start sites appeared early in the evolution of vertebrates [100]. It was hypothesized that the emergence and stabilization of CGIs could be due to a hypodeamination regime associated with low level of DNA methylation or GC-biased gene conversion – a non-reciprocal copying of a DNA sequence from one homologous chromosome to the other during meiotic recombination [101-102]. At present, the molecular mechanisms underlying the CpG enrichment at regulatory regions and the emergence of CGIs in evolution remain unclear. Interestingly, the nuclear fraction of TDG associates tightly with euchromatin and binds to CpG-rich promoters of transcribed genes including the pluripotency and developmental genes [63,91]. *Tdg*^{-/-} mouse embryonic fibroblasts show imbalanced histone modifications and CpG methylation at the promoters of the affected genes, suggesting that the TDG-initiated BER is required to support correct methylation patterns and chromatin remodeling in specific DNA regions.

Recently, it has been demonstrated that TDG, and to a lesser extent MBD4, is able to excise T opposite to various adenine lesions with good efficiency from the specific 5'-TpG-3'/5'-CpX-3' context (where X is hypoxanthine (Hx), ethenoadenine, 8-oxoadenine, or AP site) [103]. This ability was classified as aberrant DNA repair, since it can lead to misincorporation opposite the damaged base [104]. The isolated catalytic domain of TDG (residues 111–308) lacking both N- and C-tails, while maintaining the uracil removal activity, excises T from these substrates very weakly, suggesting that the N- and C-terminal portions of TDG are required for the aberrant activity. In vitro reconstitution of TDG-initiated BER with DNA containing a T:Hx pair results in the misincorporation of C opposite to Hx. Furthermore, comparison of the spontaneous genome-wide mutation spectra with those at the CGIs revealed that CGIs exhibit a strong mutational bias for TpG, CpA→CpG mutations [103]. These findings demonstrate that TDG may catalyze sequence context-dependent aberrant removal of T that results in the appearance of

a CpG dinucleotide, thus providing a plausible mechanism for the evolutionary origin of CpG islands in mammalian genomes.

3. APE1: a dual function enzyme

AP endonuclease 1 (variously called APE1, APEX1, HAP-1 or Ref-1 in the literature) is the major AP endonuclease supporting BER in vertebrate cells both for genome defense and for active DNA demethylation, and is also involved in RNA biogenesis [105-106]. APE1 was independently discovered as an AP site-specific DNA endonuclease homologous to the *E. coli* Xth protein [107-108] and as a redox factor (Ref-1) that regulates the DNA-binding domains of the Fos and Jun subunits of AP-1 transcription factor [109]. Based on the biochemical characterization of N- and C-terminal deletion mutants, the APE1 polypeptide can be subdivided into two functionally distinct and partially overlapping regions, with the N-terminal part (residues 1–127) possessing the “redox activity”, and the C-terminal part (residues 61–318) responsible for the AP site cleavage activity [110-111]. Structurally, however, the redox part does not fold into a separate domain, and APE1 contains an Xth-like catalytic core (residues 62–316) and an N-tail unresolved in the available crystal structures over the first 2/3 of its length [112-123].

The redox function localized to the N-tail and the adjacent part of the catalytic domain has been a matter of extensive studies, which nevertheless still have not come to an agreeable model; even the redox nature of this activity is controversial. Initially, it was shown that spurious oxidation of Cys272 in Jun and Cys154 in Fos subunits of AP-1 transcription factor prevent its binding to DNA [124], and that this binding can be restored by Ref-1, a protein factor that turned out to be identical to APE1 [109,125]. Truncated APE1 containing only residues 1–127 is fully able to stimulate AP-1 DNA binding [110]. The N-tail is essential for the stimulation of AP-1 binding to DNA, and its deletion results in a > 10-fold loss in the stimulation activity but does not affect AP site cleavage [110-111,126]. In addition to AP-1, APE1 has been reported to

enhance DNA binding by NF- κ B [109,127-129], p53 [130-132], Egr-1 [133], Pax-5, Pax-8 [134-135], HIF-1 α and HLF [136-137], and YB-1 [138], although the localization of the stimulation activity to the Ref-1 part in many of these cases has not been confirmed.

Based on mutagenesis results, APE1 was initially proposed to activate AP-1 and NF- κ B via a redox mechanism by mediating the reduction of a Cys residue in their DNA-binding domains [115,126]. A disulfide bond would form in APE1 between Cys65 (residing in the catalytic domain) and presumably Cys93, which is then regenerated by thioredoxin as a reducing molecule [139-140]. However convenient, this model is plagued by a number of contradictory findings. First, crystallographic studies revealed that Cys65 and Cys93 are buried inside the APE1 globule and are unlikely to interact with other proteins [112-113]. Second, the ablation of either the homologous Cys64 residue in mouse Apex1, or of all seven cysteines in human APE1 does not affect the stimulation of DNA binding by AP-1 and NF- κ B [129,141]. Third, the *Arabidopsis* AP endonuclease ARP, which is only distantly related to APE1, lacks a Cys65 homolog and shares very little homology with the Ref-1 part of APE1, is able to stimulate the DNA-binding activity of human AP-1 [142]. Finally, at least in AP-1, the Cys oxidation does not produce a disulfide bond but rather involves other oxidized thiol derivatives [124]. On the other hand, in support of the thiol-mediated redox reaction model, it was shown that the zebrafish homolog of APE1, which also lacks a Cys65 homolog, does not activate human AP-1 but gains this ability if a Cys is reinstated [115]. For these reasons, and considering recent data on APE1–DNA interactions, an alternative model of transcription factor stimulation has been proposed, discussed in the last section.

Besides transcription factors, APE1 can also stimulate DNA glycosylases in a redox-independent manner. As mentioned above for TDG, many DNA glycosylases bind tightly to their abasic site product. Kinetic characterization of TDG, MBD4, single-strand-specific monofunctional uracil–DNA glycosylase (SMUG1), *N*-methylpurine–DNA glycosylase (MPG/ANPG/AAG), 8-oxoguanine–DNA glycosylase (OGG1), endonuclease III homolog

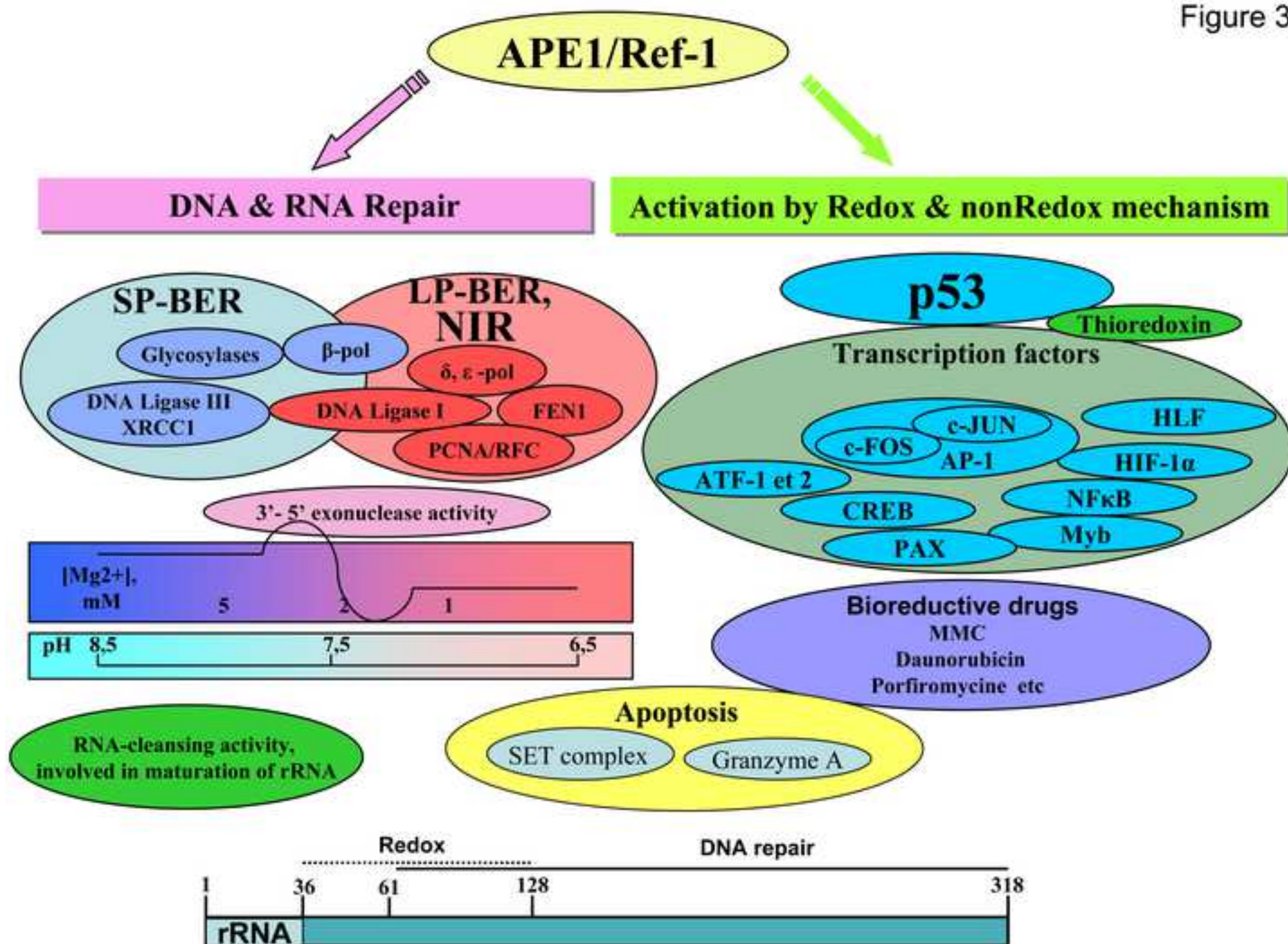
(NTHL1) and MutY homolog (MUTYH) reveals that the product release is rate limiting during the steady-state phase of the reaction [78,143-148]. Full-length APE1 promotes dissociation of the enzyme-product complex, increasing the glycosylase turnover rates [79,148-151]. In contrast, the truncated APE1-NΔ61 mutant lacking the N-tail cannot stimulate the multiple turnover of OGG1, MBD4, and MPG [152].

The N-tail of APE1 is highly enriched in positively charged Lys and Arg residues, which can engage in electrostatic interactions with DNA phosphates, and are involved into multiple posttranslational modifications with a regulatory role. The first 33 residues of the N-tail mediate the binding of APE1 to RNA and to negative Ca^{2+} responsive sequence elements (nCaRE) of certain gene promoters [153-154]. Deletion, mutation, or acetylation of Lys residues within this region increase the AP site cleavage activity, suggesting that the extra charge at the N-tail could retain APE1 on DNA [153,155]. APE1 is acetylated at Lys6 and Lys7 by CBP/p300 and deacetylated by SIRT1, apparently switching between transcription mode (acetylated) and DNA repair mode (deacetylated): the acetylation is required for efficient binding of nCaRE elements [156] and for efficient stimulation of YB-1 and YB1-dependent transcription [138], whereas deacetylated APE1 is bound to XRCC1 [157]. Lys residues 24, 25, and 27 are sites of ubiquitylation, which targets APE1 for degradation and retain it in cytoplasm in a MDM2/p53-dependent manner, possibly indicating shutdown of DNA repair in apoptosis-committed cells [158]. Interestingly, in a specific form of apoptosis initiated by granzyme A protease, APE1 is cleaved at Lys31 and then fully degraded [159]. The first 13 residues of APE1 contain the nuclear localization signal [160]; at the same time, the intact N-tail is suggested to mask the mitochondrial targeting sequence located in the C-terminal region of the polypeptide [161].

In addition to BER, APE1 plays a key role in the removal of damaged DNA bases *via* the nucleotide incision repair (NIR) pathway [162-164] (Figure 3). In BER, APE1 acts downstream of DNA glycosylases by incising a DNA duplex at AP sites and removing 3'-blocking sugar-phosphate moieties. In NIR, APE1 bypasses the glycosylase action and directly generates a

Figure 3
[Click here to download high resolution image](#)

Figure 3

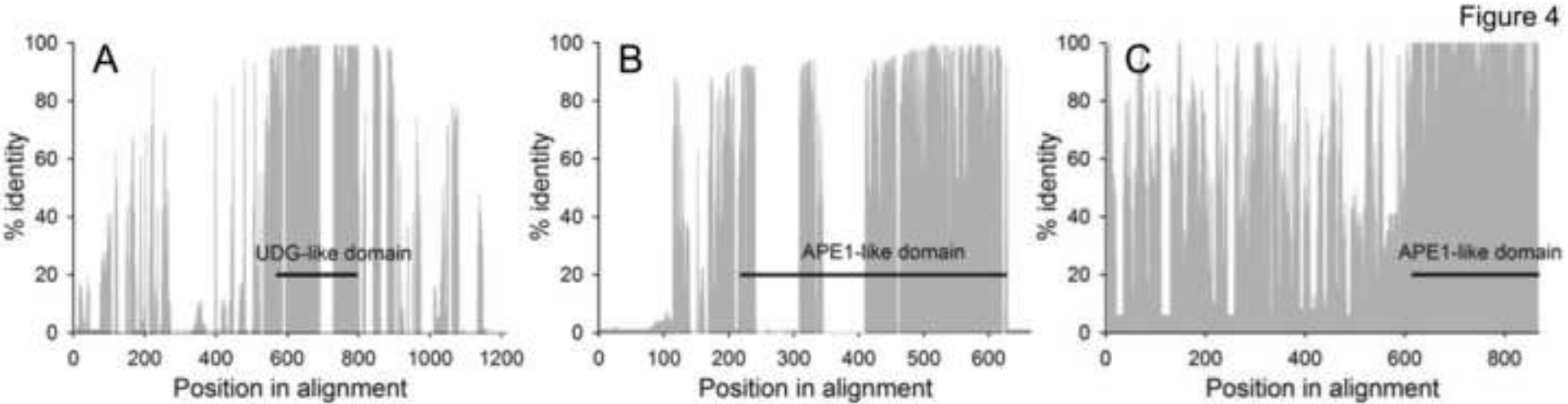


single-strand break with a 5'-dangling damaged nucleotide and a 3'-hydroxyl group [163,165]. In the NIR mode, APE1 can recognize diverse types of DNA base lesions including α -anomeric deoxynucleotides (α dN), oxidized pyrimidines [163,166], formamidopyrimidines [167], exocyclic DNA bases and uracil [168-169], bulky lesions such as benzene-derived DNA adducts [170] and a UV-induced 6-4 photoproduct [171]. The chemical structures of these DNA lesions have very little in common, implying that in, contrary to DNA glycosylases, APE1 tends to recognize damage-induced structural distortions of the DNA helix and not the modified base itself. The BER/NIR dichotomy in the APE1 action is modulated by concentrations of divalent cations, pH, and ionic strength in an apparently allosteric manner [163]. Changes in intracellular Mg^{2+} concentration induce conformational rearrangements in APE1, changing its DNA substrate specificity [163,172]. Notably, the truncated APE1-N Δ 61 protein exhibits a dramatic decrease in the NIR activity, suggesting a role of the N-tail in its regulation [163,166].

4. Evolution of tails

Given the apparent involvement of TDG and APE1 into transcription regulation and epigenetic demethylation in mammals, and possible involvement of the tails of the mammalian proteins into the interactions important for these functions, it is of great interest to follow the evolution of sequences extending the catalytic domains of these proteins. As the features of cytosine methylation and active demethylation are quite different between vertebrates and other eukaryotes, we have limited the initial analysis to Deuterostomia, a clade that includes vertebrates, echinoderms, and several minor taxa. A comparison of 731 sequences from genomes of 290 Deuterostomia species retrieved from the NCBI RefSeq database as having a TDG, Mug-like catalytic core showed that TDG is present across this clade, and, as expected, possesses a conserved catalytic core, with the similarity in the tails fading towards the polypeptide termini (Figure 4A).

Figure 4
[Click here to download high resolution image](#)

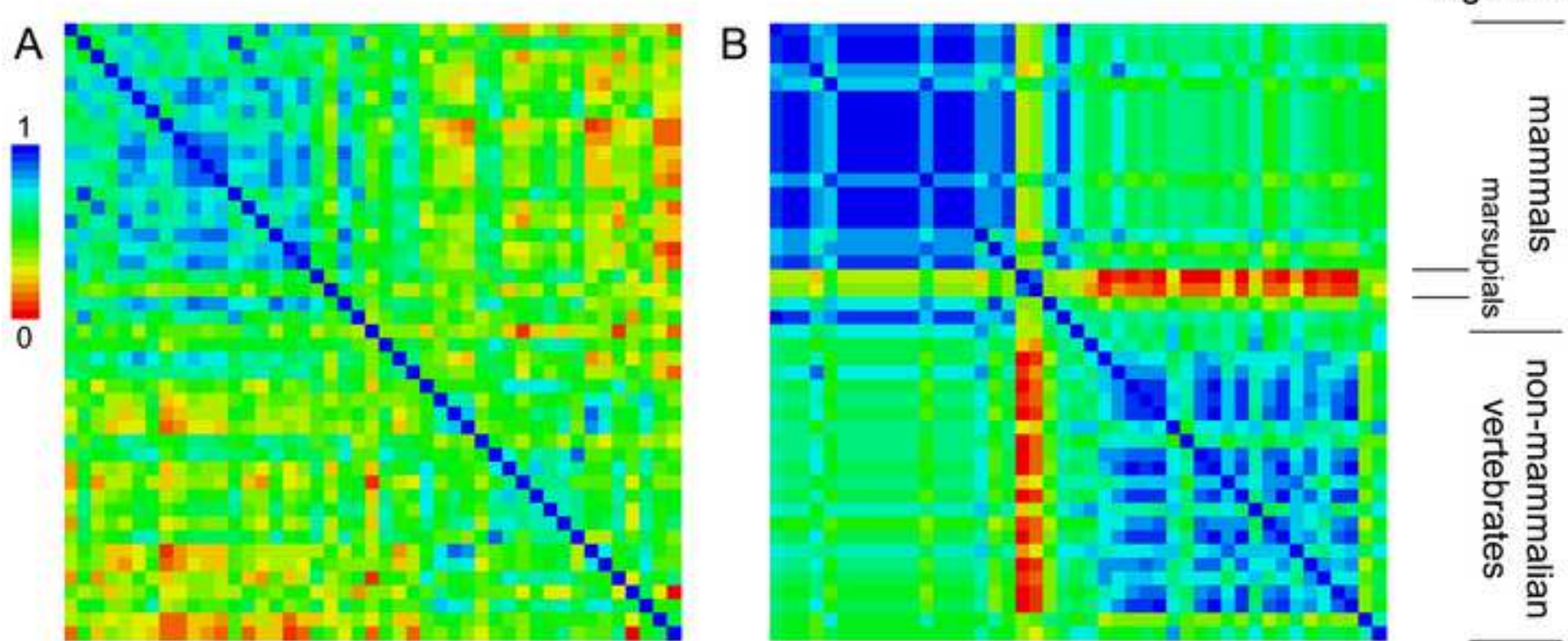


An early limited analysis of APE1 conservation in sequences from 28 vertebrate species suggested that the “redox function” might be featured only in mammals [115]. We have analyzed 290 unique sequences containing the APE1-like domain from Deuterostomia (184 species, including 97 mammals, 81 vertebrate non-mammals, and 6 non-vertebrate chordates and non-chordate Deuterostomia). Whereas the catalytic domain was well conserved, the N-tails were much more diverse (Figure 4B). Most of this diversity was probably produced by alternative translation start sites and exon use in different species: whereas all four known RNA isoforms of human APE1 encode the same polypeptide, at least 20 various types of N-tails could be identified in vertebrates. Cys in the position corresponding to human Cys65 was present only in mammals, with Ser/Thr occupying this position in all analyzed non-mammalian vertebrates.

In the absence of clear homology, the basic secondary structure elements can still be conserved between proteins from different species. To address whether this is the case for APE1 N-tail, we had run secondary structure predictions for 45 vertebrate APE1 sequences (23 mammalian and 22 non-mammalian species). The predicted structures (classified as helix, sheet, or random coil for each position) were compared using the Levenstein distance, a string metric for measuring the difference between non-numerical sequences. Moderate conservation of the N-tail secondary structure for mammals was revealed, while lacking in other vertebrates (Figure 5A). In comparison, the structures of C-terminal sequences of equal lengths are strongly conserved in the clusters of mammals (with an exception of marsupials) and non-mammalian vertebrates (Figure 5B). Overall, it seems that N-tails in mammalian APE1 proteins have indeed evolved some functions making them distinct from their non-mammalian counterparts.

Of known AP endonucleases, at least one more has a long functionally significant extension: the *Drosophila* protein, Rrp1, carries a BER function in its C-terminal part and a homologous recombination function (albeit of unknown mechanism) in its N-terminal part. Rrp1 was first isolated from fruit fly embryos as a protein capable of promoting strand annealing and strand exchange between single- and double-stranded DNA molecules in an energy-independent

Figure 5
[Click here to download high resolution image](#)



manner [173]. The compact C-terminal part (251 aa long; [Figure 1B](#)) was later shown to bear homology with EEP superfamily AP endonucleases [174]. It catalyzes cleavage of AP sites, resection of blunt or recessed 3'-ends in a DNA duplex, possesses 3'-phosphatase and 3'-phosphodiesterase activities on recessed damaged 3'-ends [175-178] and complements the oxidation- and alkylation-sensitive phenotype of AP endonuclease-null *E. coli* [179], showing all attributes of a functional AP endonuclease participating in BER. Although no structural information on Rrp1 is available, limited proteolysis combined with circular dichroism and sedimentation velocity analysis indicate that the N-tail (426 aa) is mostly unstructured [180]. The physiological role of Rrp1 in recombination is not clear: overexpression of full-length Rrp1 in *Drosophila* larvae reduces rather than increases the level of DNA damage-induced genetic marker conversion but this could reflect more efficient repair by BER and not suppression of recombination [181]. Rrp1 might contribute to recombination indirectly: its N-tail interacts with both Rev3 and Rev7 subunits of DNA polymerase ζ [182], an error-prone DNA polymerase that extends from terminally mismatched DNA primers and, from genetic evidence, is involved in homologous recombination in *Drosophila* [183].

A protein sequence database search reveals that functional AP endonucleases are linked with the recombination functional part over a very narrow phylogenetic range. Rrp1 sequences are available for 25 *Drosophila* species, with up to four protein isoforms predicted for some genomes. Their comparison clearly shows that the N-tail is much less conserved than the C-terminal APE1 homology domain ([Figure 4C](#)). Thus, even within the same genus, the N-terminal sequences, while still recognized as homologous, diverged much faster than AP endo-like domain sequences. Other Diptera also possess APE1-like AP endonucleases with long N-tails, their homology to the recombination part of *Drosophila* proteins are below the limit of positive identification. In other insects, the length of the N-tail of Rrp1/APE1 fluctuate greatly. In Lepidoptera and Coleoptera, the Rrp1 part was as extensive or even longer as in Diptera, while

in other insect orders it could be less than 70 aa. Clearly, the recombination function cannot be universally ascribed to the N-tail of insect APE1 proteins.

The failure to obtain crystal structures of TDG and APE1 tails suggests that they are likely disordered. We have used ESpritz bi-directional recursive neural network [184] trained on the training set of the DisProt database of partially or completely disordered proteins [185] to predict disorder in the tails of human APE1 and TDG and *D. melanogaster* Rrp1. Indeed, the tails showed higher disorder probability than the catalytic cores, and the boundaries of the predicted disordered regions well coincided with the borders between the core and the tails (Figure 6). It is known that terminal regions of protein-coding sequences in both mammals and insects are evolutionarily more labile, accumulating more non-synonymous mutations than the central part that presumably contains the functional core of the protein [186]. For an important class of proteins – those composed a mammalian cell mRNA interactome – high intrinsic disorder was shown to be associated with the enrichment in short repetitive amino acid motifs [187]. To assess the representation of short repeats in the tails and the core of APE1 and related AP endonucleases we have analyzed 2766 sequences containing the APE1-like domain using rapid automatic detection and alignment of repeats (RADAR) algorithm [188]. Of these, 370 contained an N-tail > 40 aa long (only 22 sequences contained a C-tail, which was not enough for a valid comparison). As an example of such analysis, Figure 7A and B illustrates the results for Rrp1, in which 74% of the N-tail turned out to be covered by short repeats. N-tails were significantly enriched in short repeats in comparison with the catalytic cores ($p < 10^{-6}$, Kolmogorov–Smirnov test; Figure 7C). TDG sequences produced a similar picture for both N- and C-tails. Thus, as in the case of mRNA-binding proteins, disorder in the tail of APE1-like endonucleases could be due to the presence of repetitive protein motifs. It should be emphasized that the disorder by itself may have functional significance and the evolution might favor accumulation of short diverged repeats in the appropriate protein regions.

Figure 6
[Click here to download high resolution image](#)

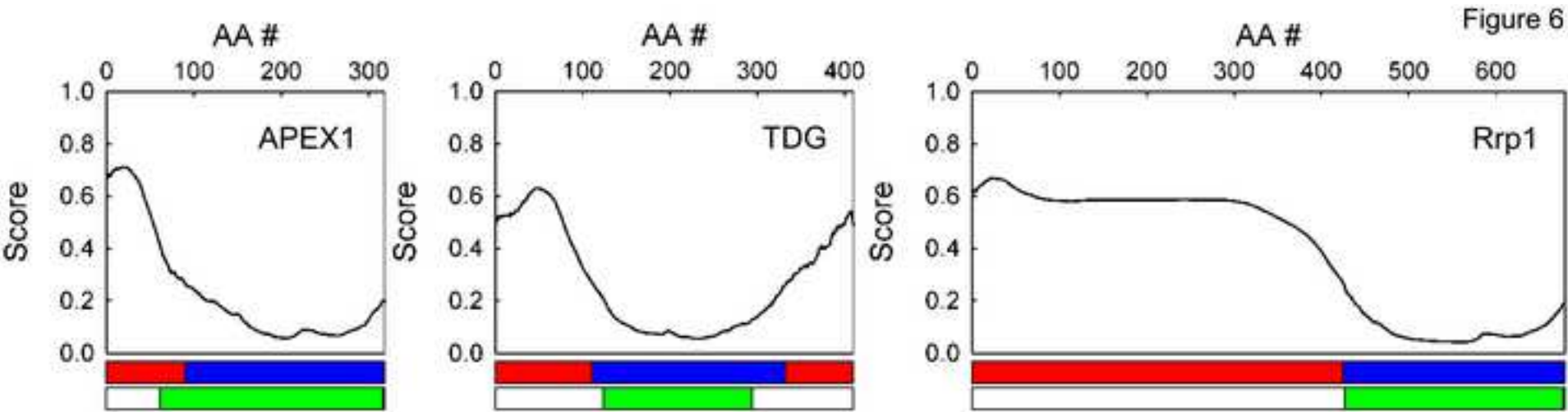
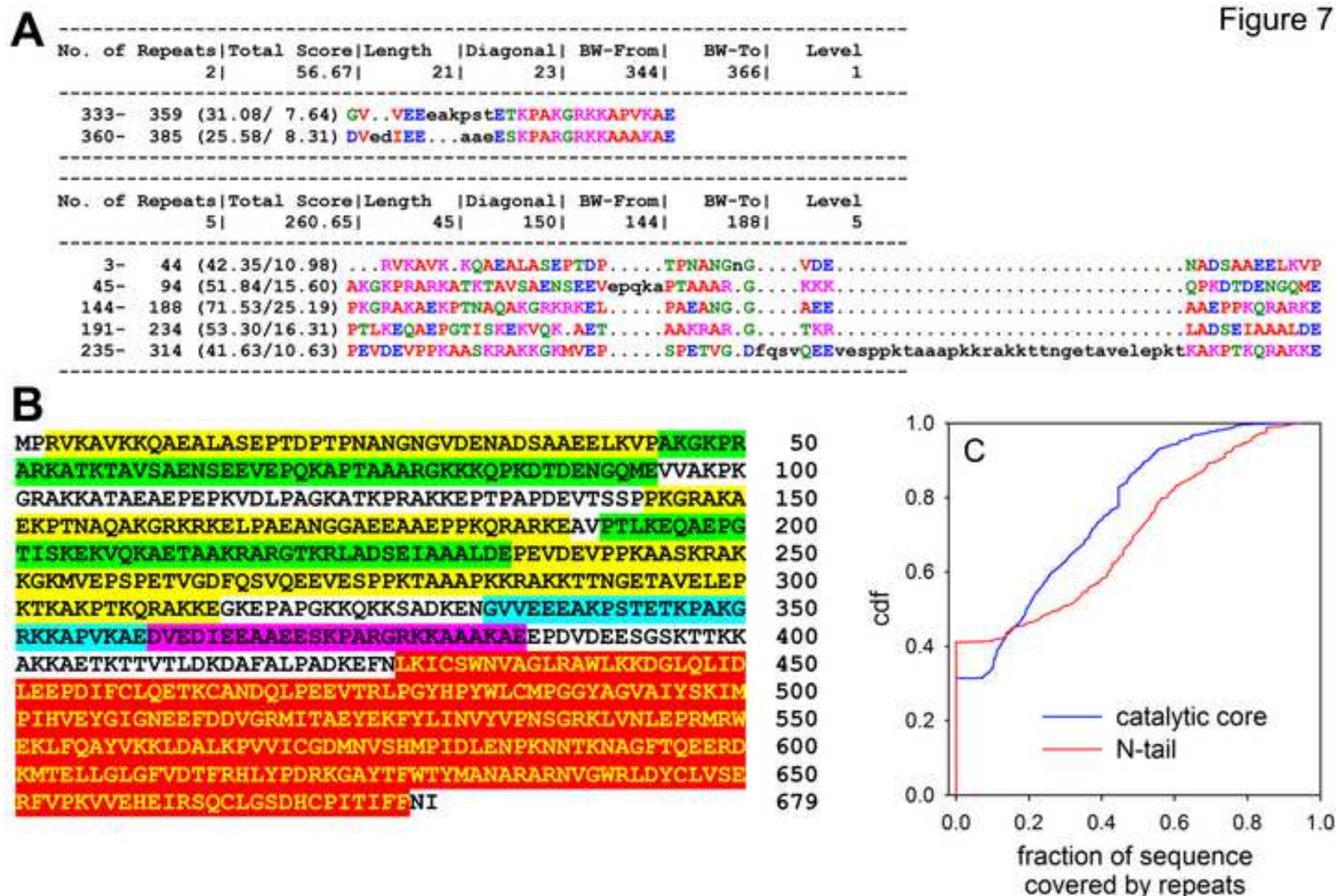


Figure 7
[Click here to download high resolution image](#)

Figure 7



5. APE1 and TDG: cooperative mediators?

As discussed above, the N-tail of APE1 is indispensable for the functions that appear very different mechanistically, namely the stimulation of DNA binding by transcription factors, turnover of DNA glycosylases, and the BER/NIR switch. Recent observations brought forward a somewhat unexpected hypothesis that unifies all these functions under a model of APE1 multimerization on DNA. Transmission electron microscopy and atomic force microscopy revealed that the full-length APE1 can form filaments on both undamaged and damaged DNA fragments and introduce bends and kinks that can be used to detect an AP site [149,189]. Several APE1 molecules was observed to bind both undamaged and damaged oligonucleotides in an apparently cooperative mode, with single-stranded DNA or nicks acting as preferential binding sites [189]. Combined with the quantitative proteomics estimates revealing that APE1 is an abundant protein in human cells, exceeding the amount of most DNA glycosylases and transcription factors by > 2 orders of magnitude [190-191], these observations suggest that APE1 monomers could dynamically assemble on DNA, starting from DNA nicks, gaps or breaks. Random binding and dissociation of terminal APE1 monomers effectively amounts to a stochastic walk of the multimer along the DNA contour, mechanistically equivalent to DNA scanning by individual molecules of many DNA repair proteins, restriction endonucleases and transcription factors [192-193]. In this mode, an APE1 multimer may “search” for some specific sites in DNA. Importantly, truncated APE1-NΔ61 multimerizes on DNA much less efficiently than the full-length protein [152], and this decreased capacity strongly correlates with the reduced ability of APE1-NΔ61 to perform NIR and to stimulate transcription factors and DNA glycosylases [110-111,126,152,163,189].

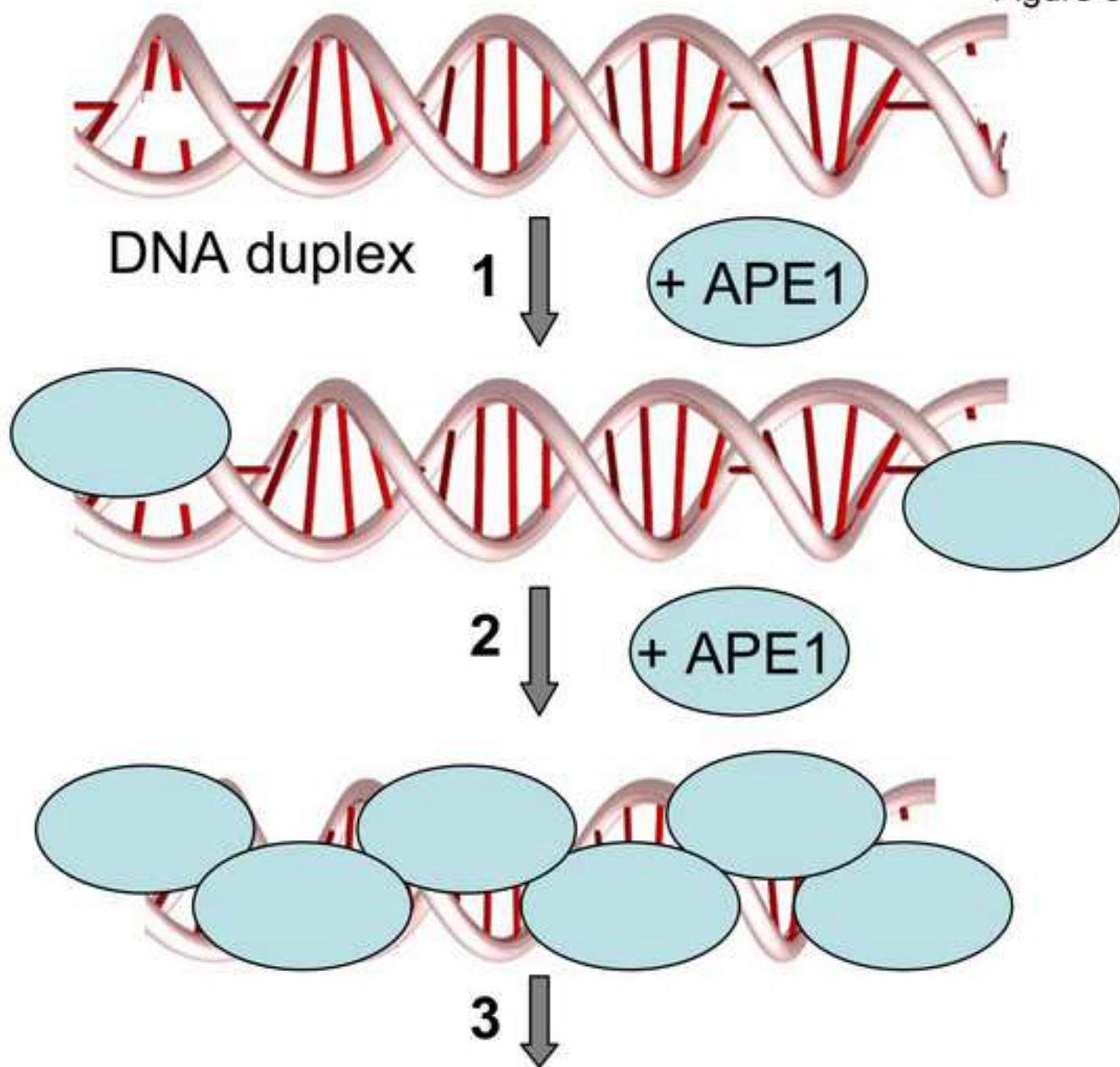
In addition, the modeled structure of the full-length APE1 could be docked into oligomers with translational symmetry, in which the DNA-binding site is exposed, consistent with the APE1 multimer assembly on DNA [189]. Binding of proteins in the DNA minor groove is known to induce groove dehydration and easy DNA bending [194]. Based on these

1 observations, it has been proposed that APE1 multimerization induces DNA helix distortion that
2 exacerbates pre-existing anisotropic flexibility at AP sites, oxidized nucleotides and intrinsically
3 bent DNA, which in turn enables NIR activity, stimulation of DNA binding by transcription
4 factors, and stimulation of DNA glycosylases turnover [152,189] (Figure 8).
5
6

7
8 What could be the mechanism responsible for the stimulation of transcription factors by
9 an APE1 multimer? DNA bending due to bound proteins also induces distortions at neighboring
10 sequences, which may promote the binding of transcription factors; a good example is p53
11 binding to its cognate sequence, which is favored by the DNA curvature induced by HMGB1
12 chromatin architectural protein [195]. Generally, the proteins that induce DNA bending exhibit
13 better affinity when their recognition sites are located in intrinsically bent or flexible regions, and
14 intrinsic DNA bending plays important role in transcription regulation [196-198]. Both NF- κ B
15 and AP-1 have been demonstrated to bend DNA [199-200]. Even though the DNA bending by
16 isolated AP-1 was debated [201], accessory factors such as NFAT1 assist bending and AP-1
17 binding to its recognition sites [202-203]. Essentially, bending induced by APE1 can play the
18 same role, sculpting DNA for easier binding of transcription factors. Although no comprehensive
19 analysis of sequence preferences was done, APE1 may itself favor intrinsically bent DNA: it
20 binds to nCaRE-A and nCaRE-B sites in the promoters of the parathyroid hormone (*PTH*) gene
21 and the own *APE1* gene [154,204-205]. nCaRE sequences contain intrinsically bent A-tracts
22 d(TTTTTG)/d(CAAAAA) flanking the recognition site for nCaRE-binding proteins. Overall,
23 APE1 might specifically bind to and multimerize on intrinsically bent cognate sites of
24 transcription factors and promote further DNA distortion, which in turn stimulates transcription
25 factor binding.
26
27

28
29 There are also some evidence for functionally relevant oligomerization of DNA
30 glycosylases, including TDG. Structural studies revealed that the catalytic domain of TDG binds
31 to 23–28-mer DNA duplexes in a 2:1 ratio, with a second TDG molecule bound in a non-
32 catalytic orientation, whereas the protein dimer complex cannot be formed on a short 15-mer
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 8



Stimulation of enzymatic turnover of DNA glycosylases and DNA binding of transcription factors.

DNA duplex [55,72]. Full-length TDG forms dimers on short duplexes with much higher efficiency, suggesting that N- and C-tails participate in the assembly of TDG on DNA. In relation to these observations, it has been shown that *E. coli* Mug, human O⁶-alkylguanine-DNA alkyltransferase (MGMT), and N-truncated *E. coli* MutY DNA glycosylase exhibit cooperative binding to DNA substrates and can form protein oligomers on DNA with 4–12 nucleotides periodicity [206-208]. It was hypothesized that the cooperative mode of binding could enable more efficient lesion search or protect repair intermediates before handing them over to downstream BER machinery. The stoichiometric difference between TDG complexes formed on on 28- and 15-mer DNA duplexes has no measurable effect on the U excision activity [72]. However, TDG does not excise T opposite ethenoadenine or hypoxanthine, or 8-oxoadenine opposite to T from 15-mer duplexes, suggesting that the cooperative binding of TDG to DNA may be required for efficient processing of a wider range of DNA substrates [103]. At present, the molecular mechanism and structural basis for this molecular recognition phenomenon remain unclear.

At least for APE1, the ability to assemble on DNA can have profound consequences for gene regulation. A testable prediction from this model is that regulatory sites exhibiting anisotropic flexibility or intrinsic DNA bending would be prone to APE1-mediated stimulation of DNA binding by transcription factors. Similarly, multimerization of APE1 at specific regulatory elements may enhance the removal of oxidized bases via stimulation of the DNA glycosylases. Noteworthy, it has been shown that activation of several promoters, including those responsive to ER α and c-Myc and also those containing potential quadruplex-forming sequences, requires guanine oxidation followed by BER in an OGG1, APE1-dependent manner [209-211]. Conversely, displacement of OGG1 by APE1 could compete with the activation of a number of immune- and stress-responsive genes, that relies on reversibly oxidized OGG1 that binds but does not excise 8-oxoguanine [212-213]. Of course, stimulation of TDG and MBD4 by APE1 may also contribute to gene regulation by TET-dependent active demethylation, both

through the erasure of mC marks and through recruitment of poly(ADP-ribose) polymerase 1 (PARP1) to the nicked repair intermediates, which poly(ADP-ribosyl)ates histones and recruits of chromatin remodeling factors. Finally, DNA curvature introduced by APE1 and enhanced by transcription factors binding, or PARP1 recruited to the APE1-produced nicks could promote DNA loop formation and interactions with distantly located enhancer or insulator elements.

6. Conclusion

The possibility of BER proteins combining their BER function with some other activity and containing structurally separated determinants for both, is attractive in the light of recent discoveries of the role of the BER pathway in processes not involved with genome integrity maintenance. In particular, in the last years it has been firmly established that mammalian cells rely on TDG-initiated BER to actively demethylate epigenetically regulated genes. Mammalian zygotes or reprogrammed cells undergo massive demethylation through this mechanism to erase epigenetic marks and re-establish pluripotency. Another well-documented regulatory role belongs to APE1, which, in addition to its role in BER, stimulates DNA binding by many transcription factors.

Both TDG and APE1 have homologs in bacteria but have a notable feature in the presence of extended tails in addition to the catalytic cores. Notably, their archaeal homologs also lack the tails, indicating that these structural elements emerged during the evolution of eukaryotes. However, even in eukaryotes, the tails of TDG and APE1 homologs are much more diverse than their catalytic cores, as evidenced, for example, by human APE1 and *Drosophila* Rrp1 that have totally different sequences, structures, and functions ascribed to their N-tails. The evolution of the tails might proceed through accumulation and fast divergence of short repeats, which would be initially close to neutral outside of the catalytic cores, and then switch to a directional selection mode once some useful function was acquired.

It is clear that tails of TDG and APE1 are involved in a multitude of protein–DNA and protein–protein interactions and serve as elements for regulatory protein modification. Currently, the mechanistic understanding of the functions of TDG and APE1 tails is limited by the lack of structural information. Given the intrinsically disordered nature of the tails, their structure in isolation is unlikely to be determined. However, it is possible that the conformations of the tails are better defined in some stable or metastable protein–protein complexes, including those involved in epigenetic DNA demethylation, which can be analyzed by conventional methods of structural biology. Moreover, new experimental methods to study mechanistically important features of intrinsically disordered protein regions are being developed using tools from NMR [214], single-molecule fluorescence microscopy [215], isotope exchange mass spectrometry [216], etc. The characterization of the tails' behavior is especially important in view of the explosive development of a new paradigm of biomolecular ensemble organization based on liquid phase separation, which is now believed to underlie many cell regulation processes including DNA repair and chromatin remodeling [217-220]. It cannot be excluded that tail-dependent assembly of APE1 and TDG aggregates on DNA might be involved in the liquid demixing in DNA repair and active epigenetic demethylation.

Acknowledgments

This research was supported by grants to Dmitry O. Zharkov from Russian Science Foundation (17-14-01190, evolutionary analysis); to Bakhyt T. Matkarimov from Nazarbayev University Oak Ridge Associated Universities (ORAU) grant, 091019CRP2111 and Ministry of Education and Science of the Republic of Kazakhstan grants AP05133910, AP05134683 program BR05236508; and to Murat Saparbaev from la Ligue National Contre le Cancer "Equipe Labellisee", Electricité de France (RB 2017) and French National Center for Scientific Research (PRC CNRS/RFBR n1074 REDOBER).

References

- [1] Friedberg E. C., Walker G. C., Siede W., Wood R. D., Schultz R. A., Ellenberger T. (2006). DNA repair and mutagenesis. Series. ASM Press.
- [2] Zharkov D. O. (2008). Base excision DNA repair. *Cell. Mol. Life Sci.* 65, 1544-1565.
- [3] Krokan H. E., Bjoras M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol.* 5, a012583. <https://doi.org/5/4/a012583> [pii] 10.1101/cshperspect.a012583.
- [4] Wallace S. S. (2014). Base excision repair: A critical player in many games. *DNA Repair (Amst)*. [https://doi.org/S1568-7864\(14\)00101-3](https://doi.org/S1568-7864(14)00101-3) [pii] 10.1016/j.dnarep.2014.03.030.
- [5] Fortini P., Dogliotti E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst)*. 6, 398-409.
- [6] Friedberg E. C., Meira L. B. (2006). Database of mouse strains carrying targeted mutations in genes affecting biological responses to DNA damage Version 7. *DNA Repair (Amst)*. 5, 189-209.
- [7] Kim D. V., Makarova A. V., Miftakhova R. R., Zharkov D. O. (2019). Base Excision DNA Repair Deficient Cells: From Disease Models to Genotoxicity Sensors. *Curr Pharm Des.* 25, 298-312. <https://doi.org/CPD-EPUB-97414> [pii] 10.2174/1381612825666190319112930.
- [8] Nohmi T., Kim S. R., Yamada M. (2005). Modulation of oxidative mutagenesis and carcinogenesis by polymorphic forms of human DNA repair enzymes. *Mutat Res.* 591, 60-73. [https://doi.org/S0027-5107\(05\)00316-7](https://doi.org/S0027-5107(05)00316-7) [pii] 10.1016/j.mrfmmm.2005.03.033.
- [9] Chan K. K., Zhang Q. M., Dianov G. L. (2006). Base excision repair fidelity in normal and cancer cells. *Mutagenesis*. 21, 173-178.
- [10] Karahalil B., Bohr V. A., Wilson D. M., 3rd. (2012). Impact of DNA polymorphisms in key DNA base excision repair proteins on cancer risk. *Hum Exp Toxicol.* 31, 981-1005. <https://doi.org/0960327112444476> [pii] 10.1177/0960327112444476.
- [11] Di Noia J. M., Neuberger M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem.* 76, 1-22.
- [12] Kavli B., Otterlei M., Slupphaug G., Krokan H. E. (2007). Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst)*. 6, 505-516.
- [13] Priet S., Sire J., Querat G. (2006). Uracils as a cellular weapon against viruses and mechanisms of viral escape. *Curr HIV Res.* 4, 31-42. <https://doi.org/10.2174/157016206775197673>.
- [14] Sousa M. M., Krokan H. E., Slupphaug G. (2007). DNA-uracil and human pathology. *Mol Aspects Med.* 28, 276-306.
- [15] Zhu J. K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* 43, 143-166.
- [16] Franchini D. M., Schmitz K. M., Petersen-Mahrt S. K. (2012). 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu Rev Genet.* 46, 419-441. <https://doi.org/10.1146/annurev-genet-110711-155451>.
- [17] Shen L., Song C. X., He C., Zhang Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu Rev Biochem.* 83, 585-614. <https://doi.org/10.1146/annurev-biochem-060713-035513>.
- [18] Apic G., Gough J., Teichmann S. A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol.* 310, 311-325. <https://doi.org/10.1006/jmbi.2001.4776> S0022-2836(01)94776-X [pii].

- [19] Kolodny R., Pereyaslavets L., Samson A. O., Levitt M. (2013). On the universe of protein folds. *Annu Rev Biophys.* 42, 559-582. <https://doi.org/10.1146/annurev-biophys-083012-130432>.
- [20] Letunic I., Doerks T., Bork P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257-260. <https://doi.org/10.1093/nar/gku949> [pii]
- [21] Marchler-Bauer A., Derbyshire M. K., Gonzales N. R., Lu S., Chitsaz F., Geer L. Y., Geer R. C., He J., Gwadz M., Hurwitz D. I., Lanczycki C. J., Lu F., Marchler G. H., Song J. S., Thanki N., Wang Z., Yamashita R. A., Zhang D., Zheng C., Bryant S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222-226. <https://doi.org/10.1093/nar/gku1221> [pii]
- [22] Huffman J. L., Sundheim O., Tainer J. A. (2005). DNA base damage recognition and removal: new twists and grooves. *Mutat. Res.* 577, 55-76.
- [23] Hitomi K., Iwai S., Tainer J. A. (2007). The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst).* 6, 410-428.
- [24] Vidal A. E., Boiteux S., Hickson I. D., Radicella J. P. (2001). XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *Embo J.* 20, 6530-6539.
- [25] Fan J., Wilson D. M., 3rd. (2005). Protein-protein interactions and posttranslational modifications in mammalian base excision repair. *Free Radic Biol Med.* 38, 1121-1138. <https://doi.org/10.1016/j.freeradbiomed.2005.01.012> [pii]
- [26] Crane-Robinson C., Dragan A. I., Privalov P. L. (2006). The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem Sci.* 31, 547-552. <https://doi.org/10.1016/j.tibs.2006.08.006> [pii]
- [27] Vuzman D., Azia A., Levy Y. (2010). Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J Mol Biol.* 396, 674-684. <https://doi.org/10.1016/j.jmb.2009.11.056> [pii]
- [28] Feng S., Cokus S. J., Zhang X., Chen P. Y., Bostick M., Goll M. G., Hetzel J., Jain J., Strauss S. H., Halpern M. E., Ukomadu C., Sadler K. C., Pradhan S., Pellegrini M., Jacobsen S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A.* 107, 8689-8694. <https://doi.org/10.1073/pnas.1002720107> [pii]
- [29] Zemach A., McDaniel I. E., Silva P., Zilberman D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 328, 916-919. <https://doi.org/10.1126/science.1186366> [pii]
- [30] Schmitz R. J., Lewis Z. A., Goll M. G. (2019). DNA Methylation: Shared and Divergent Features across Eukaryotes. *Trends Genet.* 35, 818-827. <https://doi.org/10.1016/j.tig.2019.07.007> [pii]
- [31] Bird A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8, 1499-1504.
- [32] Neddermann P., Jiricny J. (1993). The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J Biol Chem.* 268, 21218-21224.
- [33] Hendrich B., Hardeland U., Ng H. H., Jiricny J., Bird A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature.* 401, 301-304.
- [34] Wiebauer K., Jiricny J. (1989). In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. *Nature.* 339, 234-236.

- [35] Neddermann P., Jiricny J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A.* 91, 1642-1646.
- [36] Neddermann P., Gallinari P., Lettieri T., Schmid D., Truong O., Hsuan J. J., Wiebauer K., Jiricny J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* 271, 12767-12774.
- [37] Gallinari P., Jiricny J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature.* 383, 735-738.
- [38] Aravind L., Koonin E. V. (2000). The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.* 1, RESEARCH0007.
- [39] Lucas-Lledo J. I., Maddamsetti R., Lynch M. (2011). Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol Biol Evol.* 28, 1307-1317. <https://doi.org/msq318> [pii] 10.1093/molbev/msq318.
- [40] Hang B., Medina M., Fraenkel-Conrat H., Singer B. (1998). A 55-kDa protein isolated from human cells shows DNA glycosylase activity toward 3,N4-ethenocytosine and the G/T mismatch. *Proc Natl Acad Sci U S A.* 95, 13561-13566.
- [41] Saparbaev M., Laval J. (1998). 3,N4-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.* 95, 8508-8513.
- [42] Yoon J. H., Iwai S., O'Connor T. R., Pfeifer G. P. (2003). Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. *Nucleic Acids Res.* 31, 5399-5404.
- [43] Bennett M. T., Rodgers M. T., Hebert A. S., Ruslander L. E., Eisele L., Drohat A. C. (2006). Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* 128, 12510-12519.
- [44] Talhaoui I., Couve S., Ishchenko A. A., Kunz C., Schar P., Saparbaev M. (2013). 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.* 41, 912-923. <https://doi.org/gks1149> [pii] 10.1093/nar/gks1149.
- [45] Hardeland U., Bentele M., Jiricny J., Schar P. (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, *Drosophila* and fission yeast orthologs. *Nucleic Acids Res.* 31, 2261-2271.
- [46] Petronzelli F., Riccio A., Markham G. D., Seeholzer S. H., Genuardi M., Karbowski M., Yeung A. T., Matsumoto Y., Bellacosa A. (2000). Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J. Cell. Physiol.* 185, 473-480.
- [47] Hashimoto H., Liu Y., Upadhyay A. K., Chang Y., Howerton S. B., Vertino P. M., Zhang X., Cheng X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* <https://doi.org/gks155> [pii] 10.1093/nar/gks155.
- [48] Tahiliani M., Koh K. P., Shen Y., Pastor W. A., Bandukwala H., Brudno Y., Agarwal S., Iyer L. M., Liu D. R., Aravind L., Rao A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 324, 930-935. <https://doi.org/1170116> [pii] 10.1126/science.1170116.
- [49] Ito S., D'Alessio A. C., Taranova O. V., Hong K., Sowers L. C., Zhang Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature.* 466, 1129-1133. <https://doi.org/nature09303> [pii] 10.1038/nature09303.
- [50] He Y. F., Li B. Z., Li Z., Liu P., Wang Y., Tang Q., Ding J., Jia Y., Chen Z., Li L., Sun Y., Li X., Dai Q., Song C. X., Zhang K., He C., Xu G. L. (2011). Tet-mediated formation of 5-

carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 333, 1303-1307. <https://doi.org/science.1210944> [pii]
10.1126/science.1210944.

[51] Ito S., Shen L., Dai Q., Wu S. C., Collins L. B., Swenberg J. A., He C., Zhang Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 333, 1300-1303. <https://doi.org/science.1210597> [pii]
10.1126/science.1210597.

[52] Maiti A., Drohat A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334-35338. <https://doi.org/C111.284620> [pii]
10.1074/jbc.C111.284620.

[53] Morera S., Grin I., Vigouroux A., Couve S., Henriot V., Saparbaev M., Ishchenko A. A. (2012). Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res.* 40, 9917-9926. <https://doi.org/gks714> [pii]
10.1093/nar/gks714.

[54] Hashimoto H., Hong S., Bhagwat A. S., Zhang X., Cheng X. (2012). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* 40, 10203-10214. <https://doi.org/gks845> [pii]
10.1093/nar/gks845.

[55] Zhang L., Lu X., Lu J., Liang H., Dai Q., Xu G. L., Luo C., Jiang H., He C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* 8, 328-330. <https://doi.org/nchembio.914> [pii]
10.1038/nchembio.914.

[56] Hashimoto H., Zhang X., Cheng X. (2013). Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *J Mol Biol.* 425, 971-976. [https://doi.org/S0022-2836\(13\)00029-6](https://doi.org/S0022-2836(13)00029-6) [pii]
10.1016/j.jmb.2013.01.013.

[57] Hashimoto H., Zhang X., Cheng X. (2013). Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA Repair (Amst)*. 12, 535-540. [https://doi.org/S1568-7864\(13\)00084-0](https://doi.org/S1568-7864(13)00084-0) [pii]
10.1016/j.dnarep.2013.04.003.

[58] Rai K., Huggins I. J., James S. R., Karpf A. R., Jones D. A., Cairns B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*. 135, 1201-1212. [https://doi.org/S0092-8674\(08\)01517-1](https://doi.org/S0092-8674(08)01517-1) [pii]
10.1016/j.cell.2008.11.042.

[59] Guo J. U., Su Y., Zhong C., Ming G. L., Song H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. 145, 423-434. [https://doi.org/S0092-8674\(11\)00299-6](https://doi.org/S0092-8674(11)00299-6) [pii]
10.1016/j.cell.2011.03.022.

[60] Olinski R., Starczak M., Gackowski D. (2016). Enigmatic 5-hydroxymethyluracil: Oxidatively modified base, epigenetic mark or both? *Mutat Res Rev Mutat Res.* 767, 59-66. [https://doi.org/S1383-5742\(16\)30008-4](https://doi.org/S1383-5742(16)30008-4) [pii]
10.1016/j.mrrev.2016.02.001.

[61] Liu P., Burdzy A., Sowers L. C. (2003). Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair (Amst)*. 2, 199-210. <https://doi.org/S1568786402001982> [pii]
10.1016/s1568-7864(02)00198-2.

[62] Cortellino S., Xu J., Sannai M., Moore R., Caretti E., Cigliano A., Le Coz M., Devarajan K., Wessels A., Soprano D., Abramowitz L. K., Bartolomei M. S., Rambow F., Bassi M. R., Bruno T., Fanciulli M., Renner C., Klein-Szanto A. J., Matsumoto Y., Kobi D., Davidson I., Alberti C., Larue L., Bellacosa A. (2011). Thymine DNA glycosylase is essential for active DNA

demethylation by linked deamination-base excision repair. *Cell*. 146, 67-79. [https://doi.org/S0092-8674\(11\)00662-3](https://doi.org/S0092-8674(11)00662-3) [pii]
10.1016/j.cell.2011.06.020.

[63] Cortazar D., Kunz C., Selfridge J., Lettieri T., Saito Y., MacDougall E., Wirz A., Schuermann D., Jacobs A. L., Siegrist F., Steinacher R., Jiricny J., Bird A., Schar P. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*. 470, 419-423. <https://doi.org/nature09672> [pii]
10.1038/nature09672.

[64] Millar C. B., Guy J., Sansom O. J., Selfridge J., MacDougall E., Hendrich B., Keightley P. D., Bishop S. M., Clarke A. R., Bird A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*. 297, 403-405.

[65] Wong E., Yang K., Kuraguchi M., Werling U., Avdievich E., Fan K., Fazzari M., Jin B., Brown A. M., Lipkin M., Edelmann W. (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14937-14942.

[66] Bellacosa A., Cicchillitti L., Schepis F., Riccio A., Yeung A. T., Matsumoto Y., Golemis E. A., Genuardi M., Neri G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci U S A*. 96, 3969-3974.

[67] Ruzov A., Shorning B., Mortusewicz O., Dunican D. S., Leonhardt H., Meehan R. R. (2009). MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development*. 136, 2277-2286.

[68] Boland M. J., Christman J. K. (2008). Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J. Mol. Biol.* 379, 492-504. [https://doi.org/S0022-2836\(08\)00248-9](https://doi.org/S0022-2836(08)00248-9) [pii]
10.1016/j.jmb.2008.02.049.

[69] Laget S., Miotto B., Chin H. G., Esteve P. O., Roberts R. J., Pradhan S., Defossez P. A. (2014). MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics*. 9, 546-556. <https://doi.org/27695> [pii]
10.4161/epi.27695.

[70] Baba D., Maita N., Jee J. G., Uchimura Y., Saitoh H., Sugawara K., Hanaoka F., Tochio H., Hiroaki H., Shirakawa M. (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature*. 435, 979-982. <https://doi.org/nature03634> [pii]
10.1038/nature03634.

[71] Baba D., Maita N., Jee J. G., Uchimura Y., Saitoh H., Sugawara K., Hanaoka F., Tochio H., Hiroaki H., Shirakawa M. (2006). Crystal structure of SUMO-3-modified thymine-DNA glycosylase. *J Mol Biol*. 359, 137-147. [https://doi.org/S0022-2836\(06\)00372-X](https://doi.org/S0022-2836(06)00372-X) [pii]
10.1016/j.jmb.2006.03.036.

[72] Maiti A., Morgan M. T., Pozharski E., Drohat A. C. (2008). Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8890-8895.

[73] Maiti A., Noon M. S., MacKerell A. D., Jr., Pozharski E., Drohat A. C. (2012). Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8091-8096. <https://doi.org/1201010109> [pii]
10.1073/pnas.1201010109.

[74] Malik S. S., Coey C. T., Varney K. M., Pozharski E., Drohat A. C. (2015). Thymine DNA glycosylase exhibits negligible affinity for nucleobases that it removes from DNA. *Nucleic Acids Res*. 43, 9541-9552. <https://doi.org/gkv890> [pii]
10.1093/nar/gkv890.

- [75] Coey C. T., Malik S. S., Pidugu L. S., Varney K. M., Pozharski E., Drohat A. C. (2016). Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* 44, 10248-10258. <https://doi.org/gkw768> [pii] 10.1093/nar/gkw768.
- [76] Pidugu L. S., Flowers J. W., Coey C. T., Pozharski E., Greenberg M. M., Drohat A. C. (2016). Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry.* 55, 6205-6208. <https://doi.org/10.1021/acs.biochem.6b00982> 10.1021/acs.biochem.6b00982 [pii].
- [77] Smet-Nocca C., Wieruszeski J. M., Chaar V., Leroy A., Benecke A. (2008). The thymine-DNA glycosylase regulatory domain: residual structure and DNA binding. *Biochemistry.* 47, 6519-6530. <https://doi.org/10.1021/bi7022283>.
- [78] Waters T. R., Swann P. F. (1998). Kinetics of the action of thymine DNA glycosylase. *J. Biol. Chem.* 273, 20007-20014.
- [79] Waters T. R., Gallinari P., Jiricny J., Swann P. F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* 274, 67-74.
- [80] Steinacher R., Schar P. (2005). Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol.* 15, 616-623. [https://doi.org/S0960-9822\(05\)00227-7](https://doi.org/S0960-9822(05)00227-7) [pii] 10.1016/j.cub.2005.02.054.
- [81] Hardeland U., Steinacher R., Jiricny J., Schar P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *Embo J.* 21, 1456-1464.
- [82] Mohan R. D., Rao A., Gagliardi J., Tini M. (2007). SUMO-1-dependent allosteric regulation of thymine DNA glycosylase alters subnuclear localization and CBP/p300 recruitment. *Mol Cell Biol.* 27, 229-243.
- [83] Smet-Nocca C., Wieruszeski J. M., Leger H., Eilebrecht S., Benecke A. (2011). SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity. *BMC Biochem.* 12, 4. <https://doi.org/1471-2091-12-4> [pii] 10.1186/1471-2091-12-4.
- [84] Steinacher R., Barekati Z., Botev P., Kusnierczyk A., Slupphaug G., Schar P. (2019). SUMOylation coordinates BEROsome assembly in active DNA demethylation during cell differentiation. *Embo J.* 38, <https://doi.org/embj.201899242> [pii] 10.15252/embj.201899242.
- [85] Um S., Harbers M., Benecke A., Pierrat B., Losson R., Chambon P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* 273, 20728-20736.
- [86] Chen D., Lucey M. J., Phoenix F., Lopez-Garcia J., Hart S. M., Losson R., Buluwela L., Coombes R. C., Chambon P., Schar P., Ali S. (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J Biol Chem.* 278, 38586-38592. <https://doi.org/10.1074/jbc.M304286200> M304286200 [pii].
- [87] Cortazar D., Kunz C., Saito Y., Steinacher R., Schar P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst).* 6, 489-504. [https://doi.org/S1568-7864\(06\)00322-3](https://doi.org/S1568-7864(06)00322-3) [pii] 10.1016/j.dnarep.2006.10.013.
- [88] Chevray P. M., Nathans D. (1992). Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc Natl Acad Sci U S A.* 89, 5789-5793. <https://doi.org/10.1073/pnas.89.13.5789>.
- [89] Missero C., Pirro M. T., Simeone S., Pischetola M., Di Lauro R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription

factor-1-activated transcription. *J Biol Chem.* 276, 33569-33575.
<https://doi.org/10.1074/jbc.M104963200>
M104963200 [pii].

[90] Kim E. J., Um S. J. (2008). Thymine-DNA glycosylase interacts with and functions as a coactivator of p53 family proteins. *Biochem Biophys Res Commun.* 377, 838-842.
[https://doi.org/S0006-291X\(08\)02028-7](https://doi.org/S0006-291X(08)02028-7) [pii]
10.1016/j.bbrc.2008.10.058.

[91] Tini M., Benecke A., Um S. J., Torchia J., Evans R. M., Chambon P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell.* 9, 265-277.

[92] Leger H., Smet-Nocca C., Attmane-Elakeb A., Morley-Fletcher S., Benecke A. G., Eilebrecht S. (2014). A TDG/CBP/RARalpha ternary complex mediates the retinoic acid-dependent expression of DNA methylation-sensitive genes. *Genomics Proteomics Bioinformatics.* 12, 8-18. [https://doi.org/S1672-0229\(13\)00139-3](https://doi.org/S1672-0229(13)00139-3) [pii]
10.1016/j.gpb.2013.11.001.

[93] Xu X., Yu T., Shi J., Chen X., Zhang W., Lin T., Liu Z., Wang Y., Zeng Z., Wang C., Li M., Liu C. (2014). Thymine DNA glycosylase is a positive regulator of Wnt signaling in colorectal cancer. *J Biol Chem.* 289, 8881-8890. <https://doi.org/M113.538835> [pii]
10.1074/jbc.M113.538835.

[94] Hardeland U., Kunz C., Focke F., Szadkowski M., Schar P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res.* 35, 3859-3867.

[95] Shibata E., Dar A., Dutta A. (2014). CRL4Cdt2 E3 ubiquitin ligase and proliferating cell nuclear antigen (PCNA) cooperate to degrade thymine DNA glycosylase in S phase. *J Biol Chem.* 289, 23056-23064. <https://doi.org/M114.574210> [pii]
10.1074/jbc.M114.574210.

[96] Slenn T. J., Morris B., Havens C. G., Freeman R. M., Jr., Takahashi T. S., Walter J. C. (2014). Thymine DNA glycosylase is a CRL4Cdt2 substrate. *J Biol Chem.* 289, 23043-23055. <https://doi.org/M114.574194> [pii]
10.1074/jbc.M114.574194.

[97] Bird A., Taggart M., Frommer M., Miller O. J., Macleod D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell.* 40, 91-99. [https://doi.org/0092-8674\(85\)90312-5](https://doi.org/0092-8674(85)90312-5) [pii]
10.1016/0092-8674(85)90312-5.

[98] Akan P., Deloukas P. (2008). DNA sequence and structural properties as predictors of human and mouse promoters. *Gene.* 410, 165-176. [https://doi.org/S0378-1119\(07\)00613-0](https://doi.org/S0378-1119(07)00613-0) [pii]
10.1016/j.gene.2007.12.011.

[99] Deaton A. M., Bird A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010-1022. <https://doi.org/25/10/1010> [pii]
10.1101/gad.2037511.

[100] Sharif J., Endo T. A., Toyoda T., Koseki H. (2010). Divergence of CpG island promoters: a consequence or cause of evolution? *Dev Growth Differ.* 52, 545-554. <https://doi.org/DGD1193> [pii]
10.1111/j.1440-169X.2010.01193.x.

[101] Duret L., Galtier N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 10, 285-311. <https://doi.org/10.1146/annurev-genom-082908-150001>.

[102] Cohen N. M., Kenigsberg E., Tanay A. (2011). Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell.* 145, 773-786. [https://doi.org/S0092-8674\(11\)00482-X](https://doi.org/S0092-8674(11)00482-X) [pii]
10.1016/j.cell.2011.04.024.

- [103] Talhaoui I., Couve S., Gros L., Ishchenko A. A., Matkarimov B., Sapparbaev M. K. (2014). Aberrant repair initiated by mismatch-specific thymine-DNA glycosylases provides a mechanism for the mutational bias observed in CpG islands. *Nucleic Acids Res.* 42, 6300-6313. <https://doi.org/gku246> [pii] 10.1093/nar/gku246.
- [104] Talhaoui I., Matkarimov B. T., Tchenio T., Zharkov D. O., Sapparbaev M. K. (2017). Aberrant base excision repair pathway of oxidatively damaged DNA: Implications for degenerative diseases. *Free Radic Biol Med.* 107, 266-277. [https://doi.org/S0891-5849\(16\)31074-7](https://doi.org/S0891-5849(16)31074-7) [pii] 10.1016/j.freeradbiomed.2016.11.040.
- [105] Demple B., Sung J. S. (2005). Molecular and biological roles of Ape1 protein in mammalian base excision repair. *DNA Repair (Amst).* 4, 1442-1449.
- [106] Tell G., Wilson D. M., 3rd, Lee C. H. (2010). Intrusion of a DNA repair protein in the RNome world: is this the beginning of a new era? *Mol. Cell. Biol.* 30, 366-371.
- [107] Demple B., Herman T., Chen D. S. (1991). Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 88, 11450-11454.
- [108] Robson C. N., Milne A. M., Pappin D. J., Hickson I. D. (1991). Isolation of cDNA clones encoding an enzyme from bovine cells that repairs oxidative DNA damage in vitro: homology with bacterial repair enzymes. *Nucleic Acids Res.* 19, 1087-1092.
- [109] Xanthoudakis S., Miao G., Wang F., Pan Y. C., Curran T. (1992). Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. *Embo J.* 11, 3323-3335.
- [110] Xanthoudakis S., Miao G. G., Curran T. (1994). The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains. *Proc. Natl. Acad. Sci. U. S. A.* 91, 23-27.
- [111] Izumi T., Mitra S. (1998). Deletion analysis of human AP-endonuclease: minimum sequence required for the endonuclease activity. *Carcinogenesis.* 19, 525-527.
- [112] Gorman M. A., Morera S., Rothwell D. G., de La Fortelle E., Mol C. D., Tainer J. A., Hickson I. D., Freemont P. S. (1997). The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *Embo J.* 16, 6548-6558.
- [113] Mol C. D., Izumi T., Mitra S., Tainer J. A. (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature.* 403, 451-456.
- [114] Beernink P. T., Segelke B. W., Hadi M. Z., Erzberger J. P., Wilson D. M., 3rd, Rupp B. (2001). Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism. *J. Mol. Biol.* 307, 1023-1034.
- [115] Georgiadis M. M., Luo M., Gaur R. K., Delaplane S., Li X., Kelley M. R. (2008). Evolution of the redox function in mammalian apurinic/apyrimidinic endonuclease. *Mutat Res.* 643, 54-63. [https://doi.org/S0027-5107\(08\)00097-3](https://doi.org/S0027-5107(08)00097-3) [pii] 10.1016/j.mrfmmm.2008.04.008.
- [116] Manvilla B. A., Pozharski E., Toth E. A., Drohat A. C. (2013). Structure of human apurinic/apyrimidinic endonuclease 1 with the essential Mg²⁺ cofactor. *Acta Crystallogr D Biol Crystallogr.* 69, 2555-2562. <https://doi.org/S0907444913027042> [pii] 10.1107/S0907444913027042.
- [117] 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., Pique M. E., Frankel K. A., Hitomi K., Cunningham R. P., Mitra S., Tainer J. A. (2013). Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J Biol Chem.* 288, 8445-8455. <https://doi.org/M112.422774> [pii] 10.1074/jbc.M112.422774.

- [118] He H., Chen Q., Georgiadis M. M. (2014). High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease I. *Biochemistry*. 53, 6520-6529. <https://doi.org/10.1021/bi500676p>.
- [119] Freudenthal B. D., Beard W. A., Cuneo M. J., Dyrkheeva N. S., Wilson S. H. (2015). Capturing snapshots of APE1 processing DNA damage. *Nat Struct Mol Biol*. 22, 924-931. <https://doi.org/nsmb.3105> [pii] 10.1038/nsmb.3105.
- [120] Redrejo-Rodriguez M., Vigouroux A., Mursalimov A., Grin I., Alili D., Koshenov Z., Akishev Z., Maksimenko A., Bissenbaev A. K., Matkarimov B. T., Saparbaev M., Ishchenko A. A., Morera S. (2016). 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*. 128-129, 20-33. [https://doi.org/S0300-9084\(16\)30120-1](https://doi.org/S0300-9084(16)30120-1) [pii] 10.1016/j.biochi.2016.06.011.
- [121] Fairlamb M. S., Whitaker A. M., Freudenthal B. D. (2018). Apurinic/apyrimidinic (AP) endonuclease 1 processing of AP sites with 5' mismatches. *Acta Crystallogr D Struct Biol*. 74, 760-768. <https://doi.org/S2059798318003340> [pii] 10.1107/S2059798318003340.
- [122] Whitaker A. M., Flynn T. S., Freudenthal B. D. (2018). Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat Commun*. 9, 399. <https://doi.org/10.1038/s41467-017-02175-y> 10.1038/s41467-017-02175-y [pii].
- [123] Trilles R., Beglov D., Chen Q., He H., Wireman R., Reed A., Chennamadhavuni S., Panek J. S., Brown L. E., Vajda S., Porco J. A., Jr., Kelley M. R., Georgiadis M. M. (2019). Discovery of Macrocyclic Inhibitors of Apurinic/Apyrimidinic Endonuclease 1. *J Med Chem*. 62, 1971-1988. <https://doi.org/10.1021/acs.jmedchem.8b01529>.
- [124] Abate C., Patel L., Rauscher F. J., 3rd, Curran T. (1990). Redox regulation of fos and jun DNA-binding activity in vitro. *Science*. 249, 1157-1161.
- [125] Xanthoudakis S., Curran T. (1992). Identification and characterization of Ref-1, a nuclear protein that facilitates AP-1 DNA-binding activity. *Embo J*. 11, 653-665.
- [126] Walker L. J., Robson C. N., Black E., Gillespie D., Hickson I. D. (1993). Identification of residues in the human DNA repair enzyme HAP1 (Ref-1) that are essential for redox regulation of Jun DNA binding. *Mol. Cell. Biol*. 13, 5370-5376.
- [127] Mitomo K., Nakayama K., Fujimoto K., Sun X., Seki S., Yamamoto K. (1994). Two different cellular redox systems regulate the DNA-binding activity of the p50 subunit of NF-kappa B in vitro. *Gene*. 145, 197-203. [https://doi.org/0378-1119\(94\)90005-1](https://doi.org/0378-1119(94)90005-1) [pii].
- [128] Nishi T., Shimizu N., Hiramoto M., Sato I., Yamaguchi Y., Hasegawa M., Aizawa S., Tanaka H., Kataoka K., Watanabe H., Handa H. (2002). Spatial redox regulation of a critical cysteine residue of NF-kappa B in vivo. *J Biol Chem*. 277, 44548-44556. <https://doi.org/10.1074/jbc.M202970200> M202970200 [pii].
- [129] Ando K., Hirao S., Kabe Y., Ogura Y., Sato I., Yamaguchi Y., Wada T., Handa H. (2008). A new APE1/Ref-1-dependent pathway leading to reduction of NF-kappaB and AP-1, and activation of their DNA-binding activity. *Nucleic Acids Res*. 36, 4327-4336. <https://doi.org/gkn416> [pii] 10.1093/nar/gkn416.
- [130] Jayaraman L., Murthy K. G., Zhu C., Curran T., Xanthoudakis S., Prives C. (1997). Identification of redox/repair protein Ref-1 as a potent activator of p53. *Genes Dev*. 11, 558-570.
- [131] Hanson S., Kim E., Deppert W. (2005). Redox factor 1 (Ref-1) enhances specific DNA binding of p53 by promoting p53 tetramerization. *Oncogene*. 24, 1641-1647. <https://doi.org/1208351> [pii] 10.1038/sj.onc.1208351.

- [132] Seemann S., Hainaut P. (2005). Roles of thioredoxin reductase 1 and APE/Ref-1 in the control of basal p53 stability and activity. *Oncogene*. 24, 3853-3863. <https://doi.org/1208549> [pii]
10.1038/sj.onc.1208549.
- [133] Huang R. P., Adamson E. D. (1993). Characterization of the DNA-binding properties of the early growth response-1 (Egr-1) transcription factor: evidence for modulation by a redox mechanism. *DNA Cell Biol.* 12, 265-273. <https://doi.org/10.1089/dna.1993.12.265>.
- [134] Tell G., Pellizzari L., Cimarosti D., Pucillo C., Damante G. (1998). Ref-1 controls pax-8 DNA-binding activity. *Biochem Biophys Res Commun.* 252, 178-183. [https://doi.org/S0006-291X\(98\)99548-1](https://doi.org/S0006-291X(98)99548-1) [pii]
10.1006/bbrc.1998.9548.
- [135] Tell G., Zecca A., Pellizzari L., Spessotto P., Colombatti A., Kelley M. R., Damante G., Pucillo C. (2000). An 'environment to nucleus' signaling system operates in B lymphocytes: redox status modulates BSAP/Pax-5 activation through Ref-1 nuclear translocation. *Nucleic Acids Res.* 28, 1099-1105.
- [136] Carrero P., Okamoto K., Coumailleau P., O'Brien S., Tanaka H., Poellinger L. (2000). Redox-regulated recruitment of the transcriptional coactivators CREB-binding protein and SRC-1 to hypoxia-inducible factor 1alpha. *Mol Cell Biol.* 20, 402-415. <https://doi.org/10.1128/mcb.20.1.402-415.2000>.
- [137] Lando D., Pongratz I., Poellinger L., Whitelaw M. L. (2000). A redox mechanism controls differential DNA binding activities of hypoxia-inducible factor (HIF) 1alpha and the HIF-like factor. *J Biol Chem.* 275, 4618-4627. <https://doi.org/10.1074/jbc.275.7.4618>.
- [138] Chattopadhyay R., Das S., Maiti A. K., Boldogh I., Xie J., Hazra T. K., Kohno K., Mitra S., Bhakat K. K. (2008). Regulatory role of human AP-endonuclease (APE1/Ref-1) in YB-1-mediated activation of the multidrug resistance gene MDR1. *Mol. Cell. Biol.* 28, 7066-7080.
- [139] Qin J., Clore G. M., Kennedy W. P., Kuszewski J., Gronenborn A. M. (1996). The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal. *Structure*. 4, 613-620.
- [140] Hirota K., Matsui M., Iwata S., Nishiyama A., Mori K., Yodoi J. (1997). AP-1 transcriptional activity is regulated by a direct association between thioredoxin and Ref-1. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3633-3638.
- [141] Ordway J. M., Eberhart D., Curran T. (2003). Cysteine 64 of Ref-1 is not essential for redox regulation of AP-1 DNA binding. *Mol. Cell. Biol.* 23, 4257-4266.
- [142] Babiychuk E., Kushnir S., Van Montagu M., Inze D. (1994). The Arabidopsis thaliana apurinic endonuclease Arp reduces human transcription factors Fos and Jun. *Proc. Natl. Acad. Sci. U. S. A.* 91, 3299-3303.
- [143] Hill J. W., Hazra T. K., Izumi T., Mitra S. (2001). Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *Nucleic Acids Res.* 29, 430-438.
- [144] Petronzelli F., Riccio A., Markham G. D., Seeholzer S. H., Stoerker J., Genuardi M., Yeung A. T., Matsumoto Y., Bellacosa A. (2000). Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. *J. Biol. Chem.* 275, 32422-32429.
- [145] Nilsen H., Haushalter K. A., Robins P., Barnes D. E., Verdine G. L., Lindahl T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *Embo J.* 20, 4278-4286.
- [146] Yang H., Clendenin W. M., Wong D., Demple B., Slupska M. M., Chiang J. H., Miller J. H. (2001). Enhanced activity of adenine-DNA glycosylase (Myh) by apurinic/apyrimidinic endonuclease (Ape1) in mammalian base excision repair of an A/GO mismatch. *Nucleic Acids Res.* 29, 743-752.

- [147] Marenstein D. R., Chan M. K., Altamirano A., Basu A. K., Boorstein R. J., Cunningham R. P., Teebor G. W. (2003). Substrate specificity of human endonuclease III (hNTH1). Effect of human APE1 on hNTH1 activity. *J. Biol. Chem.* 278, 9005-9012.
- [148] Xia L., Zheng L., Lee H. W., Bates S. E., Federico L., Shen B., O'Connor T. R. (2005). Human 3-methyladenine-DNA glycosylase: effect of sequence context on excision, association with PCNA, and stimulation by AP endonuclease. *J Mol Biol.* 346, 1259-1274. [https://doi.org/S0022-2836\(05\)00036-7](https://doi.org/S0022-2836(05)00036-7) [pii] 10.1016/j.jmb.2005.01.014.
- [149] Sidorenko V. S., Nevinsky G. A., Zharkov D. O. (2007). Mechanism of interaction between human 8-oxoguanine-DNA glycosylase and AP endonuclease. *DNA Repair (Amst).* 6, 317-328.
- [150] Esadze A., Rodriguez G., Cravens S. L., Stivers J. T. (2017). AP-Endonuclease 1 Accelerates Turnover of Human 8-Oxoguanine DNA Glycosylase by Preventing Retrograde Binding to the Abasic-Site Product. *Biochemistry.* 56, 1974-1986. <https://doi.org/10.1021/acs.biochem.7b00017>.
- [151] Endutkin A. V., Yudkina A. V., Sidorenko V. S., Zharkov D. O. (2019). Transient protein-protein complexes in base excision repair. *J Biomol Struct Dyn.* 37, 4407-4418. <https://doi.org/10.1080/07391102.2018.1553741>.
- [152] Kladova O. A., Bazlekowa-Karaban M., Bacconnais S., Pietrement O., Ishchenko A. A., Matkarimov B. T., Iakovlev D. A., Vasenko A., Fedorova O. S., Le Cam E., Tudek B., Kuznetsov N. A., Saparbaev M. (2018). The role of the N-terminal domain of human apurinic/apyrimidinic endonuclease 1, APE1, in DNA glycosylase stimulation. *DNA Repair (Amst).* 64, 10-25. [https://doi.org/S1568-7864\(17\)30288-4](https://doi.org/S1568-7864(17)30288-4) [pii] 10.1016/j.dnarep.2018.02.001.
- [153] Fantini D., Vascotto C., Marasco D., D'Ambrosio C., Romanello M., Vitagliano L., Pedone C., Poletto M., Cesaratto L., Quadrifoglio F., Scaloni A., Radicella J. P., Tell G. (2010). Critical lysine residues within the overlooked N-terminal domain of human APE1 regulate its biological functions. *Nucleic Acids Res.* 38, 8239-8256. <https://doi.org/gkq691> [pii] 10.1093/nar/gkq691.
- [154] Antoniali G., Lirussi L., D'Ambrosio C., Dal Piaz F., Vascotto C., Casarano E., Marasco D., Scaloni A., Fogolari F., Tell G. (2014). SIRT1 gene expression upon genotoxic damage is regulated by APE1 through nCaRE-promoter elements. *Mol Biol Cell.* 25, 532-547. <https://doi.org/mbc.E13-05-0286> [pii] 10.1091/mbc.E13-05-0286.
- [155] Lirussi L., Antoniali G., Vascotto C., D'Ambrosio C., Poletto M., Romanello M., Marasco D., Leone M., Quadrifoglio F., Bhakat K. K., Scaloni A., Tell G. (2012). Nucleolar accumulation of APE1 depends on charged lysine residues that undergo acetylation upon genotoxic stress and modulate its BER activity in cells. *Mol Biol Cell.* 23, 4079-4096. <https://doi.org/mbc.E12-04-0299> [pii] 10.1091/mbc.E12-04-0299.
- [156] Bhakat K. K., Izumi T., Yang S. H., Hazra T. K., Mitra S. (2003). Role of acetylated human AP-endonuclease (APE1/Ref-1) in regulation of the parathyroid hormone gene. *Embo J.* 22, 6299-6309. <https://doi.org/10.1093/emboj/cdg595>.
- [157] Yamamori T., DeRicco J., Naqvi A., Hoffman T. A., Mattagajasingh I., Kasuno K., Jung S. B., Kim C. S., Irani K. (2010). SIRT1 deacetylates APE1 and regulates cellular base excision repair. *Nucleic Acids Res.* 38, 832-845. <https://doi.org/gkp1039> [pii] 10.1093/nar/gkp1039.
- [158] Busso C. S., Iwakuma T., Izumi T. (2009). Ubiquitination of mammalian AP endonuclease (APE1) regulated by the p53-MDM2 signaling pathway. *Oncogene.* 28, 1616-1625. <https://doi.org/onc20095> [pii] 10.1038/onc.2009.5.

- [159] Fan Z., Beresford P. J., Zhang D., Xu Z., Novina C. D., Yoshida A., Pommier Y., Lieberman J. (2003). Cleaving the oxidative repair protein Ape1 enhances cell death mediated by granzyme A. *Nat. Immunol.* 4, 145-153.
- [160] Jackson E. B., Theriot C. A., Chattopadhyay R., Mitra S., Izumi T. (2005). Analysis of nuclear transport signals in the human apurinic/apyrimidinic endonuclease (APE1/Ref1). *Nucleic Acids Res.* 33, 3303-3312. <https://doi.org/33/10/3303> [pii] 10.1093/nar/gki641.
- [161] Li M., Zhong Z., Zhu J., Xiang D., Dai N., Cao X., Qing Y., Yang Z., Xie J., Li Z., Baugh L., Wang G., Wang D. (2010). Identification and characterization of mitochondrial targeting sequence of human apurinic/apyrimidinic endonuclease 1. *J Biol Chem.* 285, 14871-14881. <https://doi.org/M109.069591> [pii] 10.1074/jbc.M109.069591.
- [162] Ischenko A. A., Saparbaev M. K. (2002). Alternative nucleotide incision repair pathway for oxidative DNA damage. *Nature.* 415, 183-187.
- [163] Gros L., Ishchenko A. A., Ide H., Elder R. H., Saparbaev M. K. (2004). The major human AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway. *Nucleic Acids Res.* 32, 73-81.
- [164] Yasui A. (2013). Alternative excision repair pathways. *Cold Spring Harb. Perspect. Biol.* 5, 1-8. <https://doi.org/cshperspect.a012617> [pii] 10.1101/cshperspect.a012617.
- [165] Redrejo-Rodriguez M., Saint-Pierre C., Couve S., Mazouzi A., Ishchenko A. A., Gasparutto D., Saparbaev M. (2011). New insights in the removal of the hydantoins, oxidation product of pyrimidines, via the base excision and nucleotide incision repair pathways. *PLoS One.* 6, e21039. <https://doi.org/10.1371/journal.pone.0021039> PONE-D-11-03778 [pii].
- [166] Daviet S., Couve-Privat S., Gros L., Shinozuka K., Ide H., Saparbaev M., Ishchenko A. A. (2007). Major oxidative products of cytosine are substrates for the nucleotide incision repair pathway. *DNA Repair (Amst).* 6, 8-18.
- [167] Christov P. P., Banerjee S., Stone M. P., Rizzo C. J. (2010). Selective Incision of the alpha-N-Methyl-Formamidopyrimidine Anomer by *Escherichia coli* Endonuclease IV. *J. Nucleic. Acids.* 2010, <https://doi.org/10.4061/2010/850234>.
- [168] Prorok P., Alili D., Saint-Pierre C., Gasparutto D., Zharkov D. O., Ishchenko A. A., Tudek B., Saparbaev M. K. (2013). Uracil in duplex DNA is a substrate for the nucleotide incision repair pathway in human cells. *Proc Natl Acad Sci U S A.* 110, E3695-3703. <https://doi.org/1305624110> [pii] 10.1073/pnas.1305624110.
- [169] Prorok P., Saint-Pierre C., Gasparutto D., Fedorova O. S., Ishchenko A. A., Leh H., Buckle M., Tudek B., Saparbaev M. (2012). Highly mutagenic exocyclic DNA adducts are substrates for the human nucleotide incision repair pathway. *PLoS One.* 7, e51776. <https://doi.org/10.1371/journal.pone.0051776> PONE-D-12-20360 [pii].
- [170] Guliaev A. B., Hang B., Singer B. (2004). Structural insights by molecular dynamics simulations into specificity of the major human AP endonuclease toward the benzene-derived DNA adduct, pBQ-C. *Nucleic Acids Res.* 32, 2844-2852.
- [171] Vrouwe M. G., Pines A., Overmeer R. M., Hanada K., Mullenders L. H. (2011). UV-induced photolesions elicit ATR-kinase-dependent signaling in non-cycling cells through nucleotide excision repair-dependent and -independent pathways. *J. Cell Sci.* 124, 435-446. <https://doi.org/jcs.075325> [pii] 10.1242/jcs.075325.
- [172] Yu E., Gaucher S. P., Hadi M. Z. (2010). Probing conformational changes in Ape1 during the progression of base excision repair. *Biochemistry.* 49, 3786-3796.

- [173] Lowenhaupt K., Sander M., Hauser C., Rich A. (1989). *Drosophila melanogaster* strand transferase. A protein that forms heteroduplex DNA in the absence of both ATP and single-strand DNA binding protein. *J Biol Chem.* 264, 20568-20575.
- [174] Sander M., Lowenhaupt K., Lane W. S., Rich A. (1991). Cloning and characterization of Rrp1, the gene encoding *Drosophila* strand transferase: carboxy-terminal homology to DNA repair endo/exonucleases. *Nucleic Acids Res.* 19, 4523-4529. <https://doi.org/10.1093/nar/19.16.4523>.
- [175] Sander M., Lowenhaupt K., Rich A. (1991). *Drosophila* Rrp1 protein: an apurinic endonuclease with homologous recombination activities. *Proc Natl Acad Sci U S A.* 88, 6780-6784. <https://doi.org/10.1073/pnas.88.15.6780>.
- [176] Nugent M., Huang S. M., Sander M. (1993). Characterization of the apurinic endonuclease activity of *Drosophila* Rrp1. *Biochemistry.* 32, 11445-11452.
- [177] Sander M., Huang S. M. (1995). Characterization of the nuclease activity of *Drosophila* Rrp1 on phosphoglycolate- and phosphate-modified DNA 3'-termini. *Biochemistry.* 34, 1267-1274.
- [178] Sander M., Benhaim D. (1996). *Drosophila* Rrp1 3'-exonuclease: demonstration of DNA sequence dependence and DNA strand specificity. *Nucleic Acids Res.* 24, 3926-3933. <https://doi.org/10.1093/nar/24.20.3926> [pii]
- [179] Gu L., Huang S. M., Sander M. (1993). *Drosophila* Rrp1 complements *E. coli* xth nfo mutants: protection against both oxidative and alkylation-induced DNA damage. *Nucleic Acids Res.* 21, 4788-4795. <https://doi.org/10.1093/nar/21.20.4788>.
- [180] Reardon B. J., Lombardo C. R., Sander M. (1998). *Drosophila* Rrp1 domain structure as defined by limited proteolysis and biophysical analyses. *J Biol Chem.* 273, 33991-33999. <https://doi.org/10.1074/jbc.273.51.33991>.
- [181] Szakmary A., Huang S. M., Chang D. T., Beachy P. A., Sander M. (1996). Overexpression of a Rrp1 transgene reduces the somatic mutation and recombination frequency induced by oxidative DNA damage in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 93, 1607-1612.
- [182] Takeuchi R., Ruike T., Nakamura R., Shimanouchi K., Kanai Y., Abe Y., Ihara A., Sakaguchi K. (2006). *Drosophila* DNA polymerase zeta interacts with recombination repair protein 1, the *Drosophila* homologue of human abasic endonuclease 1. *J Biol Chem.* 281, 11577-11585.
- [183] Kane D. P., Shusterman M., Rong Y., McVey M. (2012). Competition between replicative and translesion polymerases during homologous recombination repair in *Drosophila*. *PLoS Genet.* 8, e1002659. <https://doi.org/10.1371/journal.pgen.1002659> PGENETICS-D-11-01925 [pii].
- [184] Walsh I., Martin A. J., Di Domenico T., Tosatto S. C. (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics.* 28, 503-509. <https://doi.org/10.1093/bioinformatics/btr682> [pii]
- [185] Sickmeier M., Hamilton J. A., LeGall T., Vacic V., Cortese M. S., Tantos A., Szabo B., Tompa P., Chen J., Uversky V. N., Obradovic Z., Dunker A. K. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 35, D786-793. <https://doi.org/10.1093/nar/gkl893> [pii]
- [186] Yang H., Wu Y., Feng J., Yang S., Tian D. (2009). Evolutionary pattern of protein architecture in mammal and fruit fly genomes. *Genomics.* 93, 90-97. <https://doi.org/10.1016/j.ygeno.2008.09.009> [pii]
- [187] Castello A., Fischer B., Eichelbaum K., Horos R., Beckmann B. M., Strein C., Davey N. E., Humphreys D. T., Preiss T., Steinmetz L. M., Krijgsveld J., Hentze M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell.* 149, 1393-1406. <https://doi.org/10.1016/j.cell.2012.04.031> [pii]

- [188] Heger A., Holm L. (2000). Rapid automatic detection and alignment of repeats in protein sequences. *Proteins*. 41, 224-237. [https://doi.org/10.1002/1097-0134\(20001101\)41:2<224::AID-PROT70>3.0.CO;2-Z](https://doi.org/10.1002/1097-0134(20001101)41:2<224::AID-PROT70>3.0.CO;2-Z) [pii]
10.1002/1097-0134(20001101)41:2<224::aid-prot70>3.0.co;2-z.
- [189] Bazlekowa-Karaban M., Prorok P., Bacconnais S., Taipakova S., Akishev Z., Zembrzuska D., Popov A. V., Endutkin A. V., Groisman R., Ishchenko A. A., Matkarimov B. T., Bissenbaev A., Le Cam E., Zharkov D. O., Tudek B., Saparbaev M. (2019). Mechanism of stimulation of DNA binding of the transcription factors by human apurinic/aprimidinic endonuclease 1, APE1. *DNA Repair (Amst)*. 82, 102698. [https://doi.org/S1568-7864\(19\)30019-9](https://doi.org/S1568-7864(19)30019-9) [pii]
10.1016/j.dnarep.2019.102698.
- [190] Beck M., Schmidt A., Malmstroem J., Claassen M., Ori A., Szymborska A., Herzog F., Rinner O., Ellenberg J., Aebersold R. (2011). The quantitative proteome of a human cell line. *Mol Syst Biol*. 7, 549. <https://doi.org/msb201182> [pii]
10.1038/msb.2011.82.
- [191] Wang M., Herrmann C. J., Simonovic M., Szklarczyk D., von Mering C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics*. 15, 3163-3168. <https://doi.org/10.1002/pmic.201400441>.
- [192] Zharkov D. O., Grollman A. P. (2005). The DNA trackwalkers: principles of lesion search and recognition by DNA glycosylases. *Mutat Res*. 577, 24-54. [https://doi.org/S0027-5107\(05\)00170-3](https://doi.org/S0027-5107(05)00170-3) [pii]
10.1016/j.mrfmmm.2005.03.011.
- [193] Lee A. J., Wallace S. S. (2017). Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases? *Free Radic Biol Med*. 107, 170-178. [https://doi.org/S0891-5849\(16\)31046-2](https://doi.org/S0891-5849(16)31046-2) [pii]
10.1016/j.freeradbiomed.2016.11.024.
- [194] Privalov P. L., Crane-Robinson C. (2018). Forces maintaining the DNA double helix and its complexes with transcription factors. *Prog Biophys Mol Biol*. 135, 30-48. [https://doi.org/S0079-6107\(17\)30295-X](https://doi.org/S0079-6107(17)30295-X) [pii]
10.1016/j.pbiomolbio.2018.01.007.
- [195] Stros M., Kucirek M., Sani S. A., Polanska E. (2018). HMGB1-mediated DNA bending: Distinct roles in increasing p53 binding to DNA and the transactivation of p53-responsive gene promoters. *Biochim Biophys Acta Gene Regul Mech*. 1861, 200-210. [https://doi.org/S1874-9399\(17\)30396-6](https://doi.org/S1874-9399(17)30396-6) [pii]
10.1016/j.bbagr.2018.02.002.
- [196] Perez-Martin J., Rojo F., de Lorenzo V. (1994). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol Rev*. 58, 268-290.
- [197] Falvo J. V., Thanos D., Maniatis T. (1995). Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell*. 83, 1101-1111. [https://doi.org/0092-8674\(95\)90137-X](https://doi.org/0092-8674(95)90137-X) [pii].
- [198] Ohyama T. (2001). Intrinsic DNA bends: an organizer of local chromatin structure for transcription. *Bioessays*. 23, 708-715. <https://doi.org/10.1002/bies.1100> [pii]
10.1002/bies.1100.
- [199] Schreck R., Zorbas H., Winnacker E. L., Baeuerle P. A. (1990). The NF-kappa B transcription factor induces DNA bending which is modulated by its 65-kD subunit. *Nucleic Acids Res*. 18, 6497-6502.
- [200] Leonard D. A., Kerppola T. K. (1998). DNA bending determines Fos-Jun heterodimer orientation. *Nat. Struct. Biol*. 5, 877-881. <https://doi.org/10.1038/2316>.
- [201] McGill G., Fisher D. E. (1998). DNA bending and the curious case of Fos/Jun. *Chem Biol*. 5, R29-38. [https://doi.org/S1074-5521\(98\)90139-9](https://doi.org/S1074-5521(98)90139-9) [pii].
- [202] Chen L., Glover J. N., Hogan P. G., Rao A., Harrison S. C. (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature*. 392, 42-48. <https://doi.org/10.1038/32100>.

- [203] Diebold R. J., Rajaram N., Leonard D. A., Kerppola T. K. (1998). Molecular basis of cooperative DNA bending and oriented heterodimer binding in the NFAT1-Fos-Jun-ARRE2 complex. *Proc Natl Acad Sci U S A.* 95, 7915-7920.
- [204] Okazaki T., Chung U., Nishishita T., Ebisu S., Usuda S., Mishiro S., Xanthoudakis S., Igarashi T., Ogata E. (1994). A redox factor protein, ref1, is involved in negative gene regulation by extracellular calcium. *J Biol Chem.* 269, 27855-27862.
- [205] Izumi T., Henner W. D., Mitra S. (1996). Negative regulation of the major human AP-endonuclease, a multifunctional protein. *Biochemistry.* 35, 14679-14683. <https://doi.org/10.1021/bi961995u>
- [206] Lee C. Y., Bai H., Houle R., Wilson G. M., Lu A. L. (2004). An Escherichia coli MutY mutant without the six-helix barrel domain is a dimer in solution and assembles cooperatively into multisubunit complexes with DNA. *J Biol Chem.* 279, 52653-52663. <https://doi.org/10.1074/jbc.M405271200>
- [207] Rasimas J. J., Kar S. R., Pegg A. E., Fried M. G. (2007). Interactions of human O6-alkylguanine-DNA alkyltransferase (AGT) with short single-stranded DNAs. *J. Biol. Chem.* 282, 3357-3366. <https://doi.org/M608876200>
- [208] Gripon S., Zhao Q., Robinson T., Marshall J. J., O'Neill R. J., Manning H., Kennedy G., Dunsby C., Neil M., Halford S. E., French P. M., Baldwin G. S. (2011). Differential modes of DNA binding by mismatch uracil DNA glycosylase from Escherichia coli: implications for abasic lesion processing and enzyme communication in the base excision repair pathway. *Nucleic Acids Res.* 39, 2593-2603. <https://doi.org/gkq913>
- [209] Perillo B., Ombra M. N., Bertoni A., Cuozzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., Avvedimento E. V. (2008). DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science.* 319, 202-206. <https://doi.org/319/5860/202>
- [210] Amente S., Bertoni A., Morano A., Lania L., Avvedimento E. V., Majello B. (2010). LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene.* 29, 3691-3702. <https://doi.org/nc2010120>
- [211] Fleming A. M., Ding Y., Burrows C. J. (2017). Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc Natl Acad Sci U S A.* 114, 2604-2609. <https://doi.org/1619809114>
- [212] Pan L., Zhu B., Hao W., Zeng X., Vlahopoulos S. A., Hazra T. K., Hegde M. L., Radak Z., Bacsí A., Brasier A. R., Ba X., Boldogh I. (2016). Oxidized Guanine Base Lesions Function in 8-Oxoguanine DNA Glycosylase-1-mediated Epigenetic Regulation of Nuclear Factor kappaB-driven Gene Expression. *J Biol Chem.* 291, 25553-25566. <https://doi.org/M116.751453>
- [213] Ba X., Boldogh I. (2018). 8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions. *Redox Biol.* 14, 669-678. [https://doi.org/S2213-2317\(17\)30531-1](https://doi.org/S2213-2317(17)30531-1)
- [214] Gibbs E. B., Cook E. C., Showalter S. A. (2017). Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys.* 628, 57-70. [https://doi.org/S0003-9861\(17\)30297-7](https://doi.org/S0003-9861(17)30297-7)

- [215] LeBlanc S. J., Kulkarni P., Weninger K. R. (2018). Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules*. 8, <https://doi.org/biom8040140> [pii]
10.3390/biom8040140.
- [216] Oganessian I., Lento C., Wilson D. J. (2018). Contemporary hydrogen deuterium exchange mass spectrometry. *Methods*. 144, 27-42. [https://doi.org/S1046-2023\(17\)30452-8](https://doi.org/S1046-2023(17)30452-8) [pii]
10.1016/j.ymeth.2018.04.023.
- [217] Chong P. A., Forman-Kay J. D. (2016). Liquid-liquid phase separation in cellular signaling systems. *Curr Opin Struct Biol*. 41, 180-186. [https://doi.org/S0959-440X\(16\)30107-5](https://doi.org/S0959-440X(16)30107-5) [pii]
10.1016/j.sbi.2016.08.001.
- [218] Pancsa R., Schad E., Tantos A., Tompa P. (2019). Emergent functions of proteins in non-stoichiometric supramolecular assemblies. *Biochim Biophys Acta Proteins Proteom*. 1867, 970-979. [https://doi.org/S1570-9639\(19\)30043-3](https://doi.org/S1570-9639(19)30043-3) [pii]
10.1016/j.bbapap.2019.02.007.
- [219] Sawyer I. A., Bartek J., Dundr M. (2019). Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. *Semin Cell Dev Biol*. 90, 94-103. [https://doi.org/S1084-9521\(18\)30120-4](https://doi.org/S1084-9521(18)30120-4) [pii]
10.1016/j.semcdb.2018.07.001.
- [220] Schrank B., Gautier J. (2019). Assembling nuclear domains: Lessons from DNA repair. *J Cell Biol*. 218, 2444-2455. <https://doi.org/jcb.201904202> [pii]
10.1083/jcb.201904202.
- [221] Katoh K., Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30, 772-780. <https://doi.org/mst010> [pii]
10.1093/molbev/mst010.
- [222] Garnier J., Gibrat J. F., Robson B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*. 266, 540-553. [https://doi.org/S0076-6879\(96\)66034-0](https://doi.org/S0076-6879(96)66034-0) [pii]
10.1016/s0076-6879(96)66034-0.
- [223] Deza M. M., Deza E. (2009). *Encyclopedia of Distances*. Series. Springer Berlin Heidelberg.
- [224] Noivirt-Brik O., Prilusky J., Sussman J. L. (2009). Assessment of disorder predictions in CASP8. *Proteins*. 77 Suppl 9, 210-216. <https://doi.org/10.1002/prot.22586>.

Legends to figures

Figure 1. Schematic domain structure of *E. coli* Mug and human TDG (A) and *E. coli* Xth, human APE1, and *Drosophila* Rrp1 (B). The shaded blocks represent the enzymes' conserved catalytic domains. The lines below indicate structurally characterized part of the human proteins. Functionally important Lys330 in TDG and Cys65 in APE1 are marked by arrows.

Figure 2. Schematic representation of the active DNA demethylation pathways. mC can be hydroxylated by TET1, TET2, or TET3 proteins to form hmC or further oxidized fC and caC. In addition, mC and hmC can be deaminated by the AID/APOBEC family members to form T and hmU. Removal of these demethylation intermediates by DNA glycosylases in the BER pathway results in the erasure of mC and its replacement with C. Alternatively, de novo DNA methyltransferases DNMT3a/3b are able to directly convert hmC to C without base excision and repair synthesis. The inset shows the essential short-patch BER pathway in mammalian cells: the modified base (red) is removed by a DNA glycosylase, an AP endonuclease then nicks 5' to the nascent AP site, DNA polymerase β inserts the correct nucleotide (green) and excises the remaining deoxyribose phosphate, and DNA ligase III α seals the nick.

Figure 3. Mammalian AP endonuclease 1 (APE1), an essential multifunctional repair protein. (Left) APE1 is involved in the BER and NIR pathways to remove non-bulky DNA base damage and degrades damaged abasic RNA. BER/NIR switch and DNA substrate specificity of APE1 are modulated by concentrations of divalent cations, pH and ionic strength. (Right) APE1 stimulates DNA binding of various transcription factors *via* redox- and nonredox-dependent mechanisms and also participates in the activation of bio-reductive drugs. APE1 is a member of the SET complex and promotes the Granzyme A activated cell death response. (Bottom) Schematic structure of APE1 showing amino acids involved in nuclear localisation signal (NLS), redox function and DNA repair activities.

Figure 4. Conservation plot of the alignment of TDG (A, 731 sequences from genomes of 290 Deuterostomia species), APE1 (B, 290 sequences from 184 Deuterostomia species) and Rrp1 (C, 49 sequences from genomes of 25 *Drosophila* species). The alignment was done using MAFFT [221]. The fraction of the most frequent residue in every position is shown.

Figure 5. Conservation of the secondary structures at the N- and C-terminus of APE1 sequences. Secondary structure predictions were run for 45 vertebrate APE1 sequences (23 mammalian and 22 non-mammalian species, limited to one per taxonomic order to minimize similarity by descent) that have the N-tail > 40 aa long (44–67 in the set) using GOR IV [222]. The predicted strings of secondary structure assignments were compared by Levenstein distance between them [223]. Color-coded normalized distance matrices are shown for the complete N-tails (A); and for the last 70 C-terminal residues in the catalytic domain (B).

Figure 6. Predicted disordered regions in human APE1 and TDG and *D. melanogaster* Rrp1. ESpritz probability score is plotted over the length of the protein. Best weighted score method for disorder prediction was used [224]. Red bars show the predicted disordered regions, blue bars, the ordered regions. Green bars show the positions of the catalytic cores.

Figure 7. Detection of repeats in the APE1 family N-tails. (A), list of repeats in *Drosophila* Rrp1 identified by RADAR [188]. (B), repeats mapped onto the Rrp1 sequence (red, catalytic domain; cyan and magenta, alternating level 1 repeats; yellow and green, alternating level 5 repeats). (C), cumulative distribution function (cdf) for the fraction of the length of the catalytic domain and the N-tail covered by repeats in 2766 APE1 proteins (370 containing N-tails longer than 40 aa).

Figure 8. Proposed unified mechanism of the APE1-catalyzed stimulation of DNA glycosylases and transcription factors. (1) APE1 binds to transient single-stranded structures present at the termini of duplex DNA. (2) Multiple molecules of APE1 assemble on the DNA fragment and form multimer complexes. (3) APE1 multimers induce conformational changes in

the DNA helix, which in turn stimulates DNA glycosylase activities and sequence-specific binding of transcription factors to DNA.