

7,8-Dihydro-8-oxo-1,*N*⁶-ethenoadenine: an exclusively Hoogsteen-paired thymine mimic in DNA that induces A→T transversions in *Escherichia coli*

Andrey V. Aralov^{1,*}, Nina Gubina^{2,3,†}, Cristina Cabrero^{4,†}, Vladimir B. Tsvetkov^{5,6,†}, Anton V. Turaev⁵, Bogdan I. Fedeles², Robert G. Croy², Ekaterina A. Isaakova⁵, Denis Melnik⁷, Svetlana Dukova⁷, Dmitriy Y. Ryazantsev¹, Alexei A. Khrulev¹, Anna M. Varizhuk^{5,8}, Carlos González⁴, Timofei S. Zatsepin^{7,9,*} and John M. Essigmann²

¹Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow 117997, Russia, ²Department of Biological Engineering, Department of Chemistry and Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, ³Institute of Theoretical and Experimental Biophysics RAS, Pushchino 142290, Russia, ⁴Instituto de Química-Física Rocasolano (IQFR-CSIC), Madrid 28006, Spain, ⁵Federal Research and Clinical Center of Physical Chemical Medicine of Federal Medical Biological Agency, Moscow 119435, Russia, ⁶World-Class Research Center “Digital biodesign and personalized healthcare”, Sechenov First Moscow State Medical University, Moscow 119146, Russia, ⁷Center for Life Sciences, Skolkovo Institute of Science and Technology, Moscow 143026, Russia, ⁸Moscow Institute of Physics and Technology, Dolgoprudny 141701, Russia and ⁹Chemistry Department, Lomonosov Moscow State University, Moscow 119992, Russia

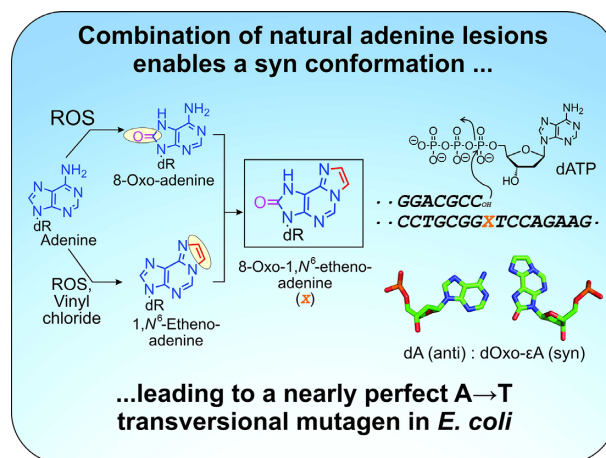
Received August 11, 2021; Revised February 09, 2022; Editorial Decision February 14, 2022; Accepted February 25, 2022

ABSTRACT

This work investigated the structural and biological properties of DNA containing 7,8-dihydro-8-oxo-1,*N*⁶-ethenoadenine (oxo-εA), a non-natural synthetic base that combines structural features of two naturally occurring DNA lesions (7,8-dihydro-8-oxoadenine and 1,*N*⁶-ethenoadenine). UV-, CD-, NMR spectroscopies and molecular modeling of DNA duplexes revealed that oxo-εA adopts the non-canonical syn conformation ($\chi = 65^\circ$) and fits very well among surrounding residues without inducing major distortions in local helical architecture. The adduct remarkably mimics the natural base thymine. When considered as an adenine-derived DNA lesion, oxo-εA was >99% mutagenic in living cells, causing predominantly A→T transversion mutations in *Escherichia coli*. The adduct in a single-stranded vector was not repaired by base excision repair enzymes (MutM and MutY glycosylases) or the AlkB dioxygenase and did not detectably affect the efficacy of DNA replication *in vivo*. When the biological and structural data are viewed together, it is likely that the nearly exclusive syn conformation and thymine mimicry of

oxo-εA defines the selectivity of base pairing *in vitro* and *in vivo*, resulting in lesion pairing with A during replication. The base pairing properties of oxo-εA, its strong fluorescence and its invisibility to enzymatic repair systems *in vivo* are features that are sought in novel DNA-based probes and modulators of gene expression.

GRAPHICAL ABSTRACT



*To whom correspondence should be addressed. Tel: +7 9266062910; Email: Baruh238@mail.ru
Correspondence may also be addressed to Timofei S. Zatsepin. Tel: +7 9265248570; Email: t.zatsepin@skoltech.ru
†The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

INTRODUCTION

Modified nucleosides, nucleotides and nucleic acids are finding an expanding range of applications in basic research, biotechnology and drug development. Point chemical modifications of nucleobases or the sugar-phosphate backbone facilitate study of DNA–protein interactions (1), enable development of specific inhibitors (2), permit the tuning of duplex stability (3), improve selectivity of siRNA (4) or aptamers (5), and improve pharmaceutical properties of antisense oligonucleotides (6), siRNA (7) and mRNA vaccines (8). Modified bases can improve stability to nucleases, enhance immune responses and promote favourable pharmacokinetics and pharmacodynamics by blocking or enhancing interactions with specific proteins. The opportunity to program these interactions in a prescribed manner helps promote success in the development of oligonucleotide drugs (more than ten of which are already approved by the FDA) and mRNA vaccines for infectious diseases such as COVID-19. The design of such drugs has been informed by results of numerous studies on nuclease-resistant base modifications that create unnatural base pairing yet do not interfere with the machinery of DNA replication (9–11). As one example, isocytidine (isoC) : isoguanosine (isoG) was the first unnatural base pair used in technology (Branched DNA (bDNA) testing kit (Roche)) (12) to increase the thermodynamic stability of a DNA duplex (13). Later it was shown that incorporation of the unnatural base pairs 5-Me-isoC:isoG and A:2-thioT into double helices enhances their thermal stability and resistance to T7 exonuclease digestion (14).

In the present work, we studied a novel DNA lesion, 7,8-dihydro-8-oxo-1, N^6 -ethenoadenine (oxo- ϵ A) by the tools of structural biology and mutagenesis to define its properties and potential for use in basic research and biotechnology. Oxo- ϵ A combines the chemistry and structure of several well-known DNA modifications formed by reactive oxygen species (ROS), such as 7,8-dihydro-8-oxoguanine (oxoG) (15,16), Fapy-dG, uracil, 7,8-dihydro-8-oxoadenine (oxoA) (17) and others (18). In addition to direct genome damage, ROS can oxidize membrane and other lipids leading to the formation of alkyl bridge-adducts (e.g. etheno adducts) with DNA adenines, cytosines and guanines. Most of these lesions can be efficiently removed by base or nucleotide excision repair systems (BER and NER) or by direct oxidative reversal by AlkB family enzymes.

Among the common oxidative DNA lesions, oxoG is one of the most prominent and its mutagenic properties have been well studied both *in vitro* and *in vivo* (15,16). Mutagenesis stemming from oxoG can be explained by the ability of the nucleobase to rotate between its syn and anti conformations about the glycosidic bond in the DNA helix (19,20). OxoG in the non-mutagenic anti conformation can form three Watson–Crick hydrogen bonds with cytosine, whereas when it rotates to the syn conformation, two Hoogsteen hydrogen bonds are possible for pairing with adenine (Figure 1A) (20–24). The syn-oxoG:A base pair resembles the cognate pyrimidine (T):purine (A) pair, which explains why oxoG formation in DNA leads to G→T transversion mutations (25).

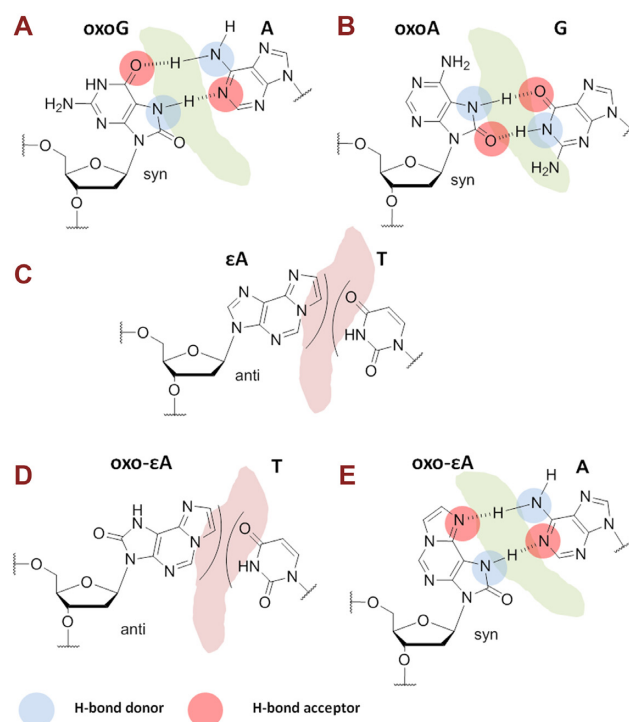


Figure 1. Proposed base pairing patterns of modified nucleobases in different conformations. *Syn*-8-oxoG:*anti*-A (A) and *syn*-8-oxoA:*anti*-G (B) can promote G→T and A→C transversion mutations, respectively. *Anti*- ϵ A cannot pair with T (C). *Anti*-oxo- ϵ A also cannot pair with T (D) and suggested base pairing of *syn*-oxo- ϵ A:*anti*-A (E). Pairings (E) can promote the observed A→T transversion mutations.

A structurally similar oxidative DNA lesion is oxoA, which can form a stable Hoogsteen base pair in the *syn* conformation with an opposing guanine (Figure 1B) (26). The presence of the 8-oxo group in oxoA increases the population of the *syn* conformer in comparison to native adenine (27,28) and leads to A→C transversion mutations (26). The third natural oxidative stress-induced DNA lesion of relevance to the present work is 1, N^6 -ethenoadenine (ϵ A), which is also mutagenic (29). This DNA adduct is formed following exposure to the carcinogen vinyl chloride through reaction with adenine of its 2-chloroacetaldehyde metabolite (30) or by the reaction of adenine with endogenous aldehyde by-products of lipid peroxidation (31). In the *anti* conformation ϵ A is unable to pair through hydrogen bonds with any canonical base in DNA duplexes (Figure 1C). However, the presence of the 1, N^6 -etheno group can partially shift the *syn*-*anti* equilibrium towards the *syn* conformation of the glycosidic bond (32), resulting in the incorporation of A, C and G opposite ϵ A (33–35). Taken together, the observations on oxoG, oxoA and ϵ A suggest a common theme for their mechanism of mutagenesis: mutagenic pairing is achieved when the damaged base adopts a *syn* conformation, which enables the formation of Hoogsteen H-bonds with non-cognate native bases in the opposite DNA strand. H-Bonding is not the only driving force for base pairing (36,37), and hydrophobic and stacking interactions becomes particularly important in the design of unnatural

base pairs (UBPs) (38). However, the structural and geometric similarity of all the aforementioned purine adducts and relevant pairing data make it possible to revise the importance of H-bonding in this context. Indeed, the constructed 2'-deoxy-7,8-dihydro-8-oxo-1,*N*⁶-ethenoadenosine modification should block canonical Watson–Crick base pairing (Figure 1D) and provide both an optimal Hoogsteen base pair geometry and a perfect hydrogen bonding pattern (one H-bond donor and one H-bond acceptor) (Figure 1E).

An adenine derivative featuring both an 1,*N*⁶-etheno bridge and an 8-oxo group of the above-described lesions is an unlikely occurrence *in vivo*, but it represents an excellent opportunity to evaluate the structural and biological consequences of rotamerism (the ability of a base to rotate about its glycosidic bond) coupled with a novel predefined H-bonding pattern.

For our applications, we sought a DNA modification that would be both mutagenic and insensitive to the pathways of adduct repair. Here we report the synthesis of oxo-εA phosphoramidite and its use to prepare modified oligodeoxynucleotides bearing the designed modification with varied flanking nucleotides. CD-spectroscopy and UV-melting analyses revealed that the oxo-εA modification placed opposite adenine in a DNA duplex did not significantly disturb helix geometry and stability in comparison to the cognate T:A base pair. The structure of the DNA duplex bearing oxo-εA was determined by 2D NMR techniques, indicating that oxo-εA opposite adenine displayed a *syn* conformation without significant local duplex deformations. The stability of the oxo-εA:A base pair was also evaluated by molecular modelling, which lends support to the empirical data.

To evaluate the base pairing properties of oxo-εA in living cells, the modification was introduced into a single-stranded M13mp7(L2) phage genome *via* ligation of oligonucleotides that site-specifically contain the modification. Subsequently, the modified phage genomes were replicated in *E. coli* strains proficient or deficient in DNA repair systems capable of acting on similar lesions (e.g. cells lacking MutM, MutY or AlkB), and phage progeny were recovered and their DNA analyzed by next-generation sequencing (NGS). We observed that oxo-εA is remarkably mutagenic, inducing almost exclusively A→T transversions in any nucleotide context in all genetic backgrounds evaluated. Thus, our *in vivo* data provide corroborating evidence that oxo-εA strongly prefers to pair with adenine during replication. Additionally, we found no indication that oxo-εA, under the conditions in which it was studied, was recognized and processed by any DNA repair system in *E. coli*. In sum, our study provides an in depth look at the impact on base-pairing specificity of a novel purine base containing a combination of common chemical DNA alterations that adopts a predominantly *syn*-pairing conformation in duplex DNA. As DNA–protein interactions play a pivotal role in transcriptional regulation of gene expression, oxo-εA can become a valuable tool for both *in vitro* and *in vivo* studies. Being a fluorescent thymine mimic, oxo-εA can help monitor changes of secondary structure in DNA–protein complexes and aptamers without risk of repair in the cell. This fluorophore, which is well tolerated in cells, is a prospective

component of FRET systems, such as aptamer switches or FRET relay logic gates, in next-generation molecular sensors (39). Also, oxo-εA nucleobase can be combined with modified sugar-phosphate backbones that could be used in a wide range of RNA-based applications.

MATERIALS AND METHODS

For a detailed description of the phosphoramidite and oligonucleotide synthesis, circular dichroism, UV-melting experiments, as well as molecular modelling and NMR spectroscopy, see Supporting Information.

Primer-extension past oxo-εA

The specificity of dATP insertion opposite oxo-εA and subsequent primer elongation were studied on native (ODN11; see Supplementary Table S1 for oligodeoxynucleotide (ODN) structures) and modified (ODN12) templates by primer extension experiments using primer (ODN13) bearing Cy5 residue at 5'-end and four native dNTPs (Supplementary Table S1). The reactions were performed by mixing a preliminary annealed primer-template duplex (5 μM) in 1× corresponding reaction buffer, 750 μM each of dNTPs (with or without dATP), and 2 units of either *E. coli* DNA Polymerase I, Large fragment (Klenow fragment) or T4 DNA polymerase (SibEnzyme, Russia). Primer extension was carried out at 37°C for 4 h, with the analysis of a reaction mixture every hour. Formamide-containing dye was used to terminate the reaction and products were evaluated by 19% denaturing PAGE. The gels were scanned using a Typhoon RGB scanner (Amersham, UK). Bands with the products of the primer extension reaction were cut out, and extension products were eluted and analyzed by mass spectrometry.

Building control and lesion-containing M13mp7(L2) genomes

The M13mp7(L2) genome was prepared as described in Delaney and Essigmann (40). Barcoded lesion-containing genomes were constructed based on a previously reported method (41) with some modifications. Briefly, the 16-mer with adenine or oxo-εA in the TXG context (X = test site), or a pool containing an equimolar mixture of sixteen 16-mers with adenine or oxo-εA in 16 contexts, were ligated to 18-mer containing a trinucleotide barcode unique for a single genome or for a pool of 16-mers. The 34-mer ligation product was then ligated to linearized M13mp7(L2) single-stranded bacteriophage genomic DNA. The ligated M13 genome was extracted with phenol/chloroform/isoamyl alcohol (25:24:1) (Invitrogen) and purified with QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol.

Replication of the phage genome construct in *E. coli*

Electrocompetent cells were prepared as described in Delaney and Essigmann (40). One hundred fmol of a single genome or 480 fmol of a pool (NAN or NXN) were electroporated into 100 μl of competent *E. coli* cells. Each electroporation was done in triplicate. All electroporations resulted in at least 5.3×10^4 initial transformation events

(infective centers), with most of the samples producing more than 10^5 events. The progeny phage were amplified in SCS110 cells to dilute out the residual genomic DNA used for electroporation. Single-stranded DNA from the progeny phage preparation was isolated with the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. A ~1 kb sequence covering the region of interest was amplified in triplicate by PCR; the PCR product was purified with the QIAquick PCR Purification Kit (Qiagen).

Library preparation and next-generation sequencing

Triplicates of purified 1kb PCR amplicons were pooled and submitted for next generation sequencing as three technical replicates for each biological replicate. Libraries were prepared using Nextera DNA Flex Library Prep (now called Illumina DNA Prep, Illumina), pooled together and submitted for sequencing on a MiSeq machine (Illumina).

Next-generation sequencing data analysis

After evaluating the quality of reads and removing the adapters, read pairs were concatenated, and only fragments having a 34-mer insert (18-mer barcode and 16-mer lesion oligonucleotide) were selected for the further analysis. Then the reads were split into groups associated with lesions, and within each lesion-related group into subgroups associated with 16 trinucleotide contexts, where applicable, and mapped to the M13mp7(L2) genome containing TAG 18-mer and TAG 16-mer insertions.

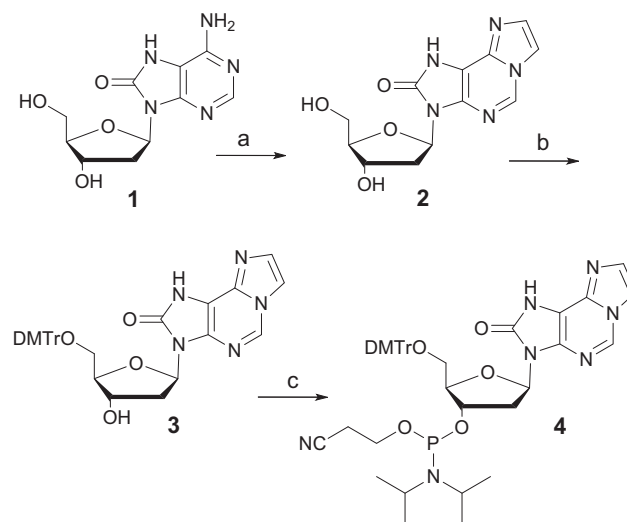
The mutation rate was calculated from the piled-up reads for each base inside the region of interest (6244...6280 bases). The miscoding percentage was obtained by calculating the relative proportion of each native base (A, C, G or T) at the site previously occupied by the lesion. Statistical comparison among genotypes and among sequence contexts was performed using ANOVA with the Tukey post-hoc test; $P < 0.05$ was taken as the significance threshold.

RESULTS

Evaluation of context-dependent oxo- ϵ A base pairing within DNA duplexes by UV melting and CD spectroscopy

Oligonucleotides were synthesized by using phosphoramidite chemistry, purified by HPLC and characterized by LC-MS. The phosphoramidite **4** was prepared in three steps (Scheme 1). Treatment of 2'-deoxy-7,8-dihydro-8-oxoadenosine **1** (42) with chloroacetaldehyde gave nucleoside derivative **2**, which was successively 5'-*O*-dimethoxytritylated and 3'-*O*-phosphitylated. Detailed synthetic procedures and NMR/MS data are presented in Supporting Information (Materials and Methods).

To assess the effects of oxo- ϵ A on DNA duplex stability, we synthesized eight modified homologous decamers (ODN1-ODN8) with different central triplets (the triplets are schematically shown in Figure 2; for full sequences, see Supplementary Table S1). By using UV melting and CD spectroscopy, we evaluated the effect of oxo- ϵ A on duplex stability in oligonucleotides in which the modification was



Scheme 1. Synthesis of oxo- ϵ A phosphoramidite. Reagents and conditions: (a) chloroacetaldehyde, aq. sodium acetate buffer, pH 4.6, 84%; (b) DMTr-Cl, pyridine, 79%; (c) NCCH₂CH₂OP(Cl)N-*i*Pr₂, DIPEA, CH₂Cl₂, 81%.

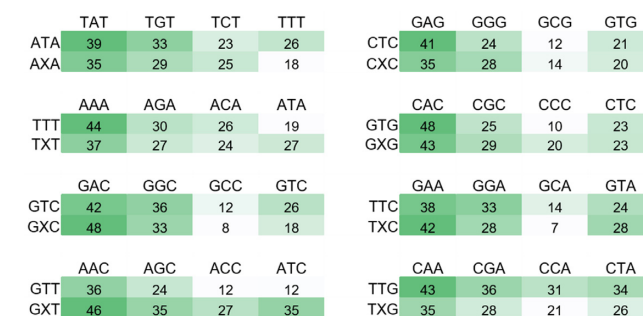


Figure 2. Heat maps summarizing relative thermal stabilities ($T_m \pm 2^\circ\text{C}$) of the duplexes formed by a native (5'-GCANTNTACG-3') or modified (5'-GCANXNTACG-3', X = oxo- ϵ A) strand and a complement (5'-CGTANNNTGC-3') (for detailed information, see Supplementary Table S1).

situated in eight flanking base contexts, and where the modification was matched with all possible pairing partners.

Based on the T_m values derived from melting curves (Supplementary Figure S1), we obtained a 'heat map' of relative duplex stabilities (Figure 2) and calculated the modification-induced stabilization/destabilization of the duplex (ΔT_m) (Supplementary Table S2, for thermodynamic parameters of duplex formation from single strands see Supplementary Table S3). According to these data, oxo- ϵ A behaves similarly to thymine, the canonical base pairing partner of adenine. At the same time, oxo- ϵ A modulates duplex stability in a flanking base-dependent manner and ΔT_m varies from $-7 \pm 2^\circ\text{C}$ (TXT) to $+10 \pm 2^\circ\text{C}$ (GXT) (Figure 2). The hybridization specificity is also context-dependent, with the lowest value observed for the GXT triplet. CD spectra of the duplexes with varied flanks (5'-NXN-3'; where X = oxo- ϵ A and N = any base) are close to the 5'-NTN-3' set and consistent with the typical CD signature of mixed-sequence dsDNA with negative and posi-

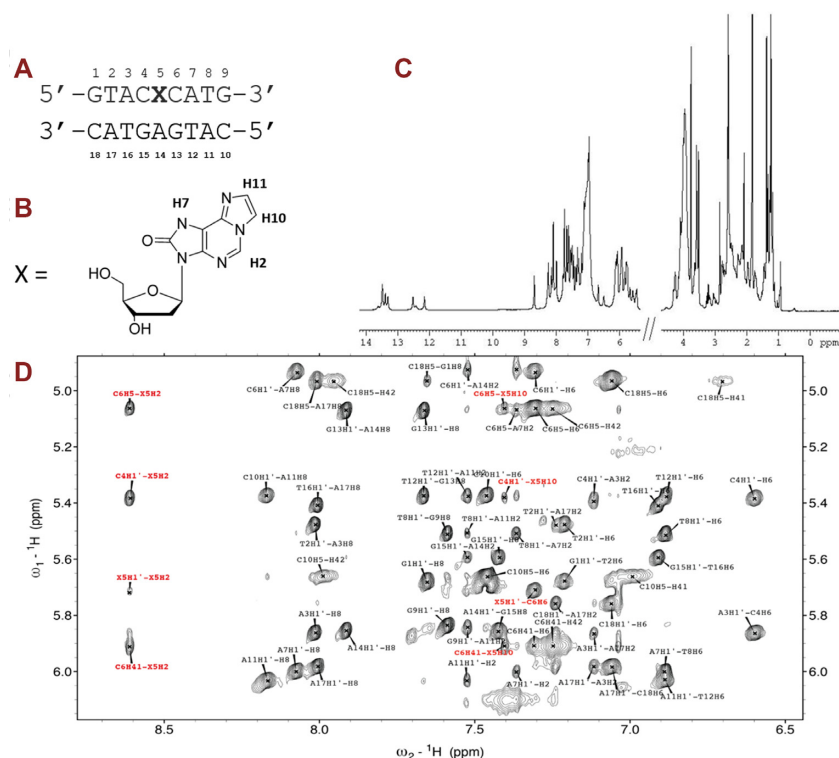


Figure 3. NMR analysis of oxo- ϵ A in duplex DNA. (A) DNA duplex used in the NMR and MD studies; (B) H-atoms numbering for oxo- ϵ A nucleoside; (C) 1D NMR spectrum; (D) H1'-aromatic region of the NOESY spectra (150 ms mixing time, $T = 3^\circ\text{C}$ (X5 proton resonances are marked in red).

tive bands around 250 nm and 280 nm, respectively (Supplementary Figure S2). Taken together, the data lead to the conclusion that oxo- ϵ A does not substantially alter duplex stability and geometry and likely forms an almost perfect base pair with adenine.

NMR and molecular modelling of the DNA duplex containing oxo- ϵ A support adoption of the syn conformation and formation of a stable base pair with the opposite adenine

To uncover the structural basis underlying the unusual selective pairing of oxo- ϵ A with an opposing strand adenine, we performed NMR studies that were further refined using molecular simulations. We synthesized the 9-mer DNA duplex used previously to determine the ϵ A conformation opposite T or G (32,43), but the central base pair was replaced with oxo- ϵ A:adenine (X:A) (Figure 3A, B). The 1D NMR spectrum is shown in Figure 3C, and the detailed assignment of protons in Supplementary Table S4. Non-exchangeable protons were assigned using established techniques for right-handed, double-stranded nucleic acids using DQF-COSY, TOCSY and 2D NOESY spectra. The base-H1' region is illustrated in Figure 3D. Three signals in the guanine imino protons region are observed in the NMR spectra (Figure 4A). Two of the signals could be assigned to G13 and G15 by following their connection to H5 of cytosines (H5C \rightarrow HN4C \rightarrow H1G) involved in G:C base pairs (Figure 4B). Most probably, the broad signal at 12.53 ppm (Figure 4A) corresponds to the terminal guanines (G1 and/or G9) which is further confirmed by the ^1H -NMR melting experiment (Supplementary Figure S5). Thymine

imino H3 protons were identified by their strong inter-strand cross-peaks with H2 of their base-paired adenines. NOE connectivities between imino protons confirmed the sequential assignment of all imino and most amino protons of non-terminal base pairs. The cross-peak patterns observed indicate that all bases are forming Watson-Crick base pairs. The intensities of intra-residual H1'-H6/H8 NOEs indicate that the glycosidic angles of all the non-modified bases, including A14, are in anti conformation (see Figure 3D). Many NOEs involving protons of oxo- ϵ A residue could be detected (Supplementary Figures S3 and S4). Of particular importance are the NOEs H7X5-H2A14 (Figure 4B) and H11X5-H6A14 (Supplementary Figure S4), supporting the oxo- ϵ A:A base pair shown in Figure 4C. The chemical shift of one of the A14 amino protons (9.62 ppm) is also indicative of being involved in a hydrogen bond.

The solution structure of the modified duplex was determined on the basis of 196 experimental distance constraints derived from NOESY experiments. The structure is well-defined, with an RMSD for all heavy atoms below 0.7 Å. Statistical data are summarized in Supplementary Table S5 and the structure ensemble is shown in Figure 5A. Overall, the DNA structure is almost a standard B-form double helix. The modified oxo- ϵ A base adopts a syn conformation ($\chi = 65^\circ$) and fits very well within the surrounding residues without inducing distortions in local DNA architecture (see Figure 5D). Short distances and appropriate geometries indicate the formation of two hydrogen bonds between oxo- ϵ A (X5) and A14 (N4X5-H61A14 and N1A14-H2X5), as shown in Figure 5C. This structure is supported by nu-

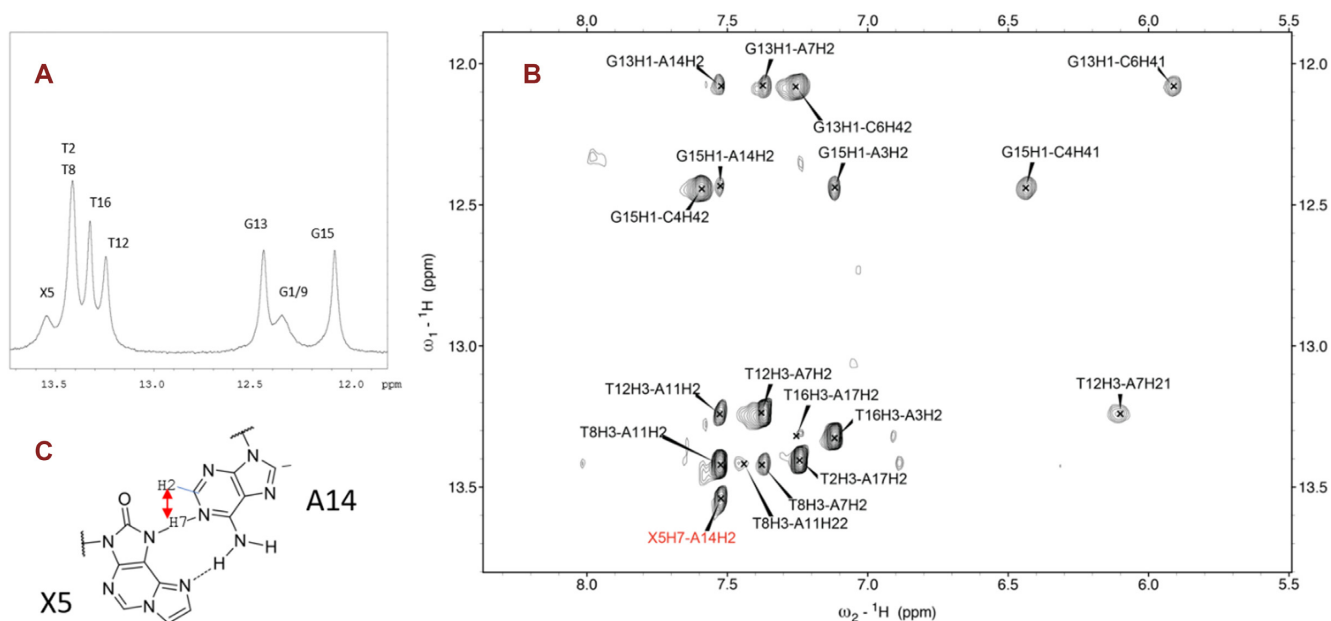


Figure 4. Exchangeable protons region of the NMR spectra (A); NOESY spectra (mixing time 50 ms, $T = 3^{\circ}\text{C}$) (B). All non-terminal base pairs could be identified. H7 proton of X5 exhibits a clear NOE with H2A14 (labelled in red), confirming the proposed base pairing X5A14 shown in (C).

merous experimental NOEs involving oxo- ϵ A (X5) protons, shown in Supplementary Table S6. Final structures are deposited in the PDB (7NBP) and their geometrical parameters are shown in Supplementary Table S7. The resulting structure is very similar to a standard B-form duplex with only very minor distortions. A comparison with a canonical B-duplex with an A:T base pair is shown in Supplementary Figure S6. Oxo- ϵ A fits very well within the surrounding nucleobases. The minor changes in the backbone observed are sufficient to mitigate potential steric clashes due to the syn conformation of the oxo- ϵ A residue.

Next, computational modeling was performed to evaluate the stability of the oxo- ϵ A:adenine base pair. We compared syn (X^S) and anti (X^A) conformations of oxo- ϵ A in the DNA duplex used for NMR studies. Distances between NH donors and N acceptors and initial and final conformations for both variants are shown in Figure 5E-F and Supplementary Figure S7. For the syn conformation of oxo- ϵ A, two NH-N hydrogen bonds can exist with adenine, while only one H-bond is available in case of the anti orientation (Supplementary Figure S7, Supplementary Table S8). At the same time, the contribution of electrostatic interactions was higher for the syn variant (Supplementary Figure S8B). In the case of the anti conformation, the aromatic part of the X^A :A pair does not perturb stacking with neighboring bases (Supplementary Figures S7 and S8C), but causes significant positive stress and deformations of the sugar-phosphate backbone (Supplementary Figure S8E). Thus, the calculated free energy is much lower for the syn conformation, mainly due to the Coulombic contributions and stress energies (Supplementary Figure S8A). As a result, we conclude that the anti conformation of oxo- ϵ A is not likely to exist in a B-form DNA duplex. Taken together, the NMR and molecular dynamics simulation data are consistent with the UV melting and CD spectroscopy data and confirm that oxo- ϵ A functions as a close mimic of thymine.

***In vitro* bypass of oxo- ϵ A by DNA polymerase I, Klenow fragment or T4 DNA polymerase**

First, we carried out primer-extension experiments, using native DNA (ODN11 with native T at the 7-position) or modified DNA template (ODN12 with oxo- ϵ A residue at the 7-position), a primer bearing 5'-terminal Cy5 residue (ODN13) and the large fragment of *E. coli* DNA polymerase I (Klenow fragment) or T4 DNA polymerase (44). For both polymerases, after a 4 h primer extension reaction under optimal conditions, an extension product was found only in the presence of all four nucleoside triphosphates (dNTPs), whereas in the absence of deoxyadenosine triphosphate (dATP), the reaction did not yield any measurable extension product (lanes 2 and 3, 4 and 5, respectively in Supplementary Figure S9, upper panel). Both polymerases were equally efficient in primer elongation on the unmodified template ODN11, and T4 DNA polymerase also effectively extended the primer on modified template ODN12 within 1 h. In turn, we did not observe complete primer elongation with the Klenow fragment on modified template even after 4 h, suggesting that oxo- ϵ A can be a partial replication blocker for some polymerases *in vitro*. MS analysis showed the presence of a fully elongated product for the both polymerases with the incorporation of a deoxyadenosine residue opposite oxo- ϵ A, as well as certain intermediate products of the extension reaction (see Supplementary Figure S9, lower panel).

Evaluation of oxo- ϵ A base pairing properties in living cells

Oxo- ϵ A is not known to be a naturally formed pre-mutagenic lesion, but studies of how it is processed *in vivo* are relevant for two reasons. First, oxo- ϵ A applications in biotechnology would be limited if it were to be repaired inside cells. And second, as a fundamental issue, it was of interest to see if observed mutational outcomes would be in

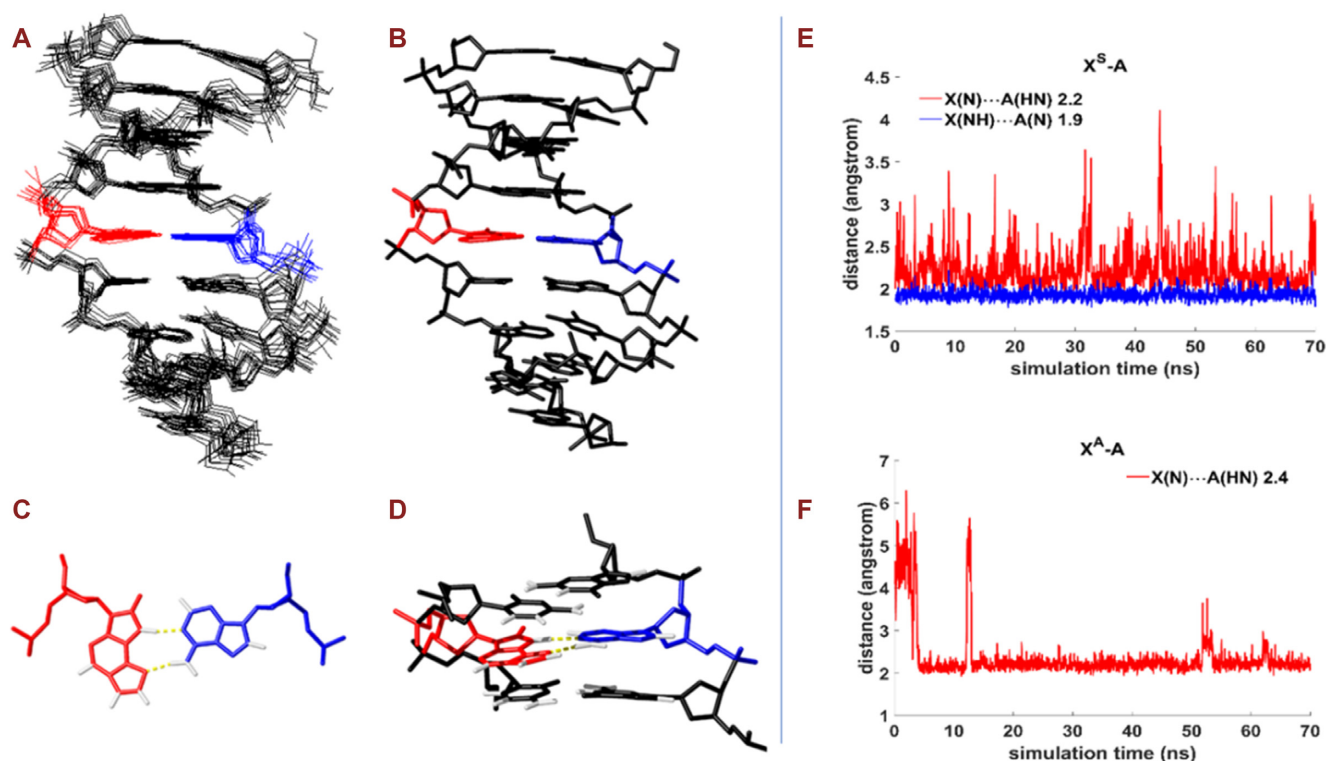


Figure 5. Ensemble of 10 resulting structures determined from NMR data (A) (7NBP). Averaged solution structure (B). Detail of the X5:A14 base pair (C), and their neighboring base pairs (D). X5 is shown in red, and A14 in cyan. The distance between hydrogen atom of H-bond donor and nitrogen of H-bond acceptor in X^S -A (E) and X^A -A (F) base pairs.

alignment with expectations from the spectroscopic, modelling and *in vitro* data that indicate a syn-oxo- ϵ A:adenine base pair.

The spectroscopic data show the ability of oxo- ϵ A to form a stable base pair with adenine. Here, we consider oxo- ϵ A as a modification of the cognate adenine base, and thus, any base-pairing behavior deviating from that of normal adenine is considered mutagenic. In this context, we anticipated, based on the spectroscopy data, that oxo- ϵ A was very likely to induce mutations. Because oxo- ϵ A is a hybrid combining the structural features of oxoA and ϵ A, we expected that the mutagenic properties of oxo- ϵ A might reflect the mutagenic features of either one or both lesions. In *E. coli*, the repair of ϵ A in single-stranded DNA is performed primarily by AlkB, a direct DNA damage reversal enzyme that removes the etheno bridge by oxidation and restores the canonical base, adenine, within the DNA sequence (45). Less is known about repair of oxoA *in vitro*, but the bacterial MUG enzyme (mismatch uracil glycosylase) is able to remove oxoA from duplexes, primarily when mispaired with C, but less able to repair the lesion in an oxoA:A pair (46). In eukaryotes, oxoA is repaired by the OGG1 glycosylase, a component of the base excision repair pathway (BER) (47); the TDG glycosylase also plays a role, but, as with the bacterial MUG, repair is dependent on the type of mismatch (46). With regard to oxoG, MutM, the bacterial homolog of human OGG1, removes oxoG paired with C, while the MutY glycosylase prevents the mutagenic outcome of oxoG by removing A from the oxoG:A pair (48,49). Given that both

MutM and MutY have a broad substrate specificity (50), we decided to probe both repair pathways - direct reversal by AlkB, and MutM/MutY-mediated BER to examine the mutagenic properties of oxo- ϵ A.

Oxo- ϵ A is not successfully repaired by the *E. coli* base excision repair machinery

To assess mutagenic potential of oxo- ϵ A in MutM/MutY-proficient and -deficient strains of *E. coli*, we took an advantage of the high-throughput sequencing technology developed in our lab whereby the mutagenic properties of a site-specific DNA lesion can be probed in parallel in multiple bacterial strains and/or sequence contexts using multiplexing and NGS (Figure 6) (41). In brief, M13 single-stranded vectors with either A or oxo- ϵ A at a defined position in the 5'-TXG-3' sequence context and a specific barcode sequence, are mixed in a 1:1 ratio and introduced by electroporation into wild type (WT) *E. coli* or MutM⁻, MutY⁻, MutM⁻/MutY⁻ strains. After *in vivo* replication, progeny DNA from each repair background was isolated, the 1 kb fragment around the lesion site was PCR amplified and fragmented to generate sequencing libraries, where each individual sample was given a second set of specific Illumina barcodes. This step allowed us to pool all samples and sequence them in a single run on an Illumina MiSeq instrument. NGS provides data on the mutagenicity of oxo- ϵ A as a function of genetic background (the detailed procedure is described in Supporting Information). The percentage of

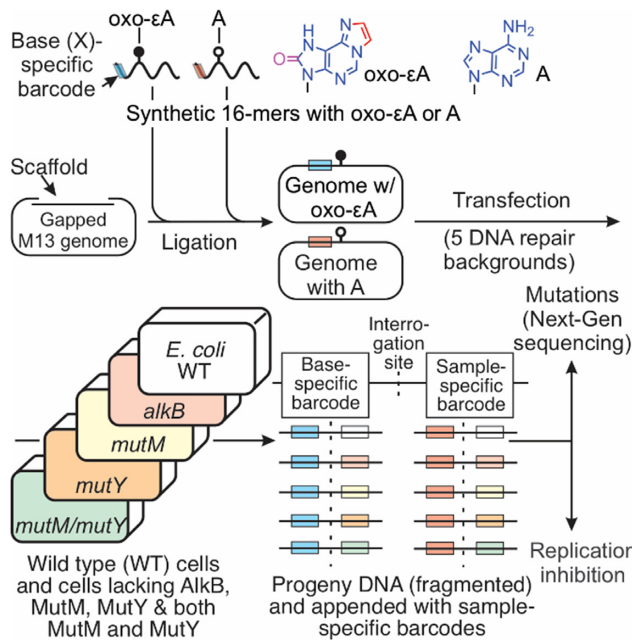


Figure 6. Schematic representation of the *in vivo* mutagenesis assay with Next-Gen sequencing. X is either A (adenine) or oxo-εA at the lesion site. The colored box to the left of the interrogated site symbolizes the lesion-specific trinucleotide barcode; the box to the right of the interrogated site represents a second barcode introduced during Illumina library preparation (indexing barcode), corresponding to each sample.

mutations is given in Supplementary Table S9, and statistical comparison between genotypes using ANOVA with the Tukey post-hoc test is given in Supplementary Table S10. The results show that oxo-εA is ~97% mutagenic, independent of the MutM/MutY status of the host cell in which the modification is replicated (Figure 7A), compared to the adenine control (Figure 7B). Notably, 95% of all mutations induced by oxo-εA are A→T transversions, in alignment with our predictions based on structural and physicochemical characteristics of oxo-εA in DNA. The results suggest that an oxo-εA:A mispairing in the TXG context is unlikely to be a substrate for the MutM glycosylase (Figure 7A). Unexpectedly, in the experiment with oxo-εA in 16 contexts, there are nearly 2% of oxo-εA→C mutations in the GXC, GXG and GXT contexts (Figures 8A and C, Supplementary Figures S11A and S11C), implying that in these cases oxo-εA could be paired with G and could become a substrate for MutM, as MutM can excise oxoG paired with G (51). However, even if MutM removes oxo-εA from oxo-εA:G pairs, in further repair reactions C will be incorporated opposite to G leading to the A→C mutation observed in the GXC, GXG and GXT contexts, and therefore, the potential activity of MutM against oxo-εA:G mispairing will be camouflaged.

The same situation seems to apply to MutY. Even if it recognizes oxo-εA:A pairings and successfully removes A from that pair, in the next rounds of repair, this A will be replaced by another A, and the mutagenic potential of oxo-εA will not be decreased. Therefore, we cannot completely exclude the opportunity of oxo-εA recognition by MutM or MutY, but their activity will be unnoticed due to the futile

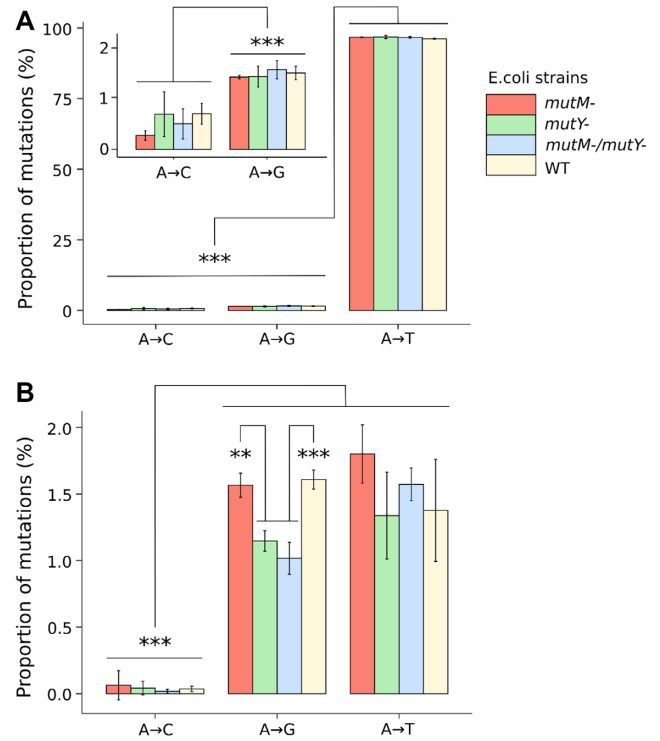


Figure 7. Nucleotide composition at the lesion site in M13 phage DNA after replication in *E. coli* cells either proficient or deficient in MutM, MutY or both (A) oxo-εA lesion (B) control adenine. X-axis denotes a type of mutation; y-axis indicates the relative quantity of a given base at the lesion site.

cycle of further incorporation of the complementary base, which will finally lead to nearly 100% mutagenesis of oxo-εA even in WT cells.

Oxo-εA is not recognized by an *E. coli* direct reversal repair system, independent of the nucleotide context

The experiments above were conducted in a single sequence context around the lesion site, namely 5'-TXG-3'. Since mutagenesis and DNA repair can be context-dependent (52,53). It is possible that the choice of sequence context may partially account for the high mutation rate and low extent of repair observed. To address this issue, we set out to investigate the mutagenic properties of oxo-εA in all 16 possible 3-base sequence contexts, as well as to determine the contribution of the direct reversal repair enzyme AlkB, which is very effective at repairing εA. We constructed 32 single-stranded M13 vectors, each containing either adenine or oxo-εA in one of 16 trinucleotide contexts, and a trinucleotide barcode corresponding to adenine or oxo-εA (detailed description is given in Supporting Information). Sixteen adenine-containing genomes were mixed in an equimolar proportion into a control pool, and 16 oxo-εA-containing genomes were prepared as the experimental pool, and electroporated into AlkB-proficient and AlkB-deficient strains (HK81 and HK82, respectively). The subsequent steps, including phage replication, DNA isolation, PCR amplification, sequencing and data analysis, were performed as described above. The percentage of mutations is

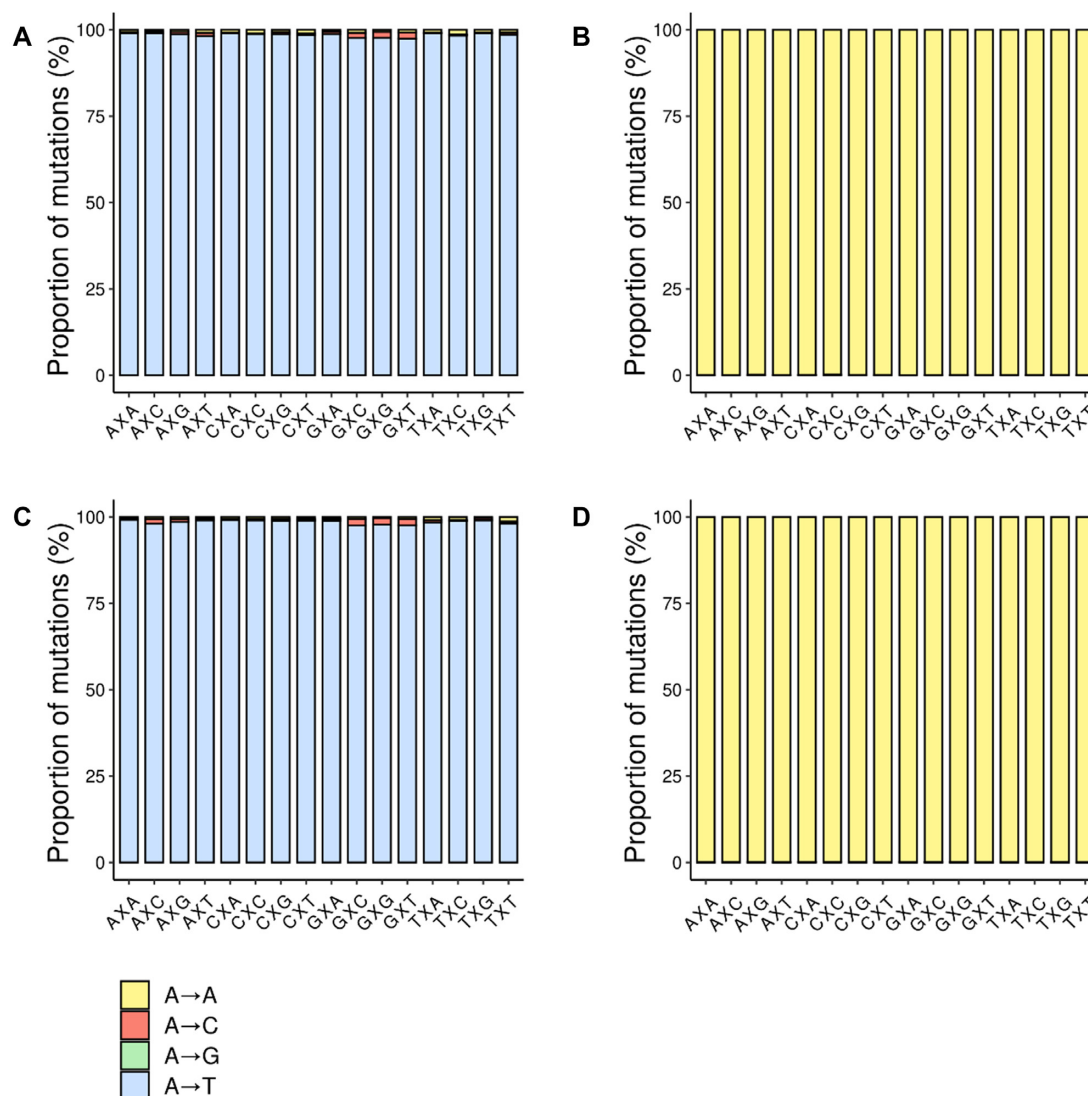


Figure 8. Sequence-context dependent mutagenic potential of oxo- ϵ A in *E. coli* cells possessing and lacking the AlkB protein. (A) Mutagenesis of oxo- ϵ A in the WT strain (*alkB*⁺). (B) Mutagenesis of the unmodified A control genome in the WT strain (*alkB*⁺). (C) Mutagenesis of oxo- ϵ A in *alkB* (AlkB-deficient) cells. (D) Mutagenesis of the A control in the *alkB* strain. The x-axis shows the trinucleotide contexts in which the modification or control bases were evaluated. The y-axis denotes relative percentage composition of a base at the lesion site (with respect to the total coverage).

given in Supplementary Table S11 for AlkB deficient strain, and the results of statistical comparison between sequence contexts are provided in Supplementary Table S12. Consistent with the previous experiment, we found that oxo- ϵ A is nearly 100% mutagenic in both the WT strain (Figure 8A, Supplementary Figure S11A) and the AlkB-deficient strain (Figure 8C, Supplementary Figure S11C). Moreover, high levels of miscoding are seen uniformly across all sequence contexts. The dominant mutation is the A→T transversion (98.5%), which is consistent with the hypothesis that oxo- ϵ A in the syn conformation strictly pairs with dATP during replication. The control genome, featuring a normal A in all sequence contexts shows only a background level of mutations, well in line with the sensitivity and level of noise of the method (Figures 8B and D; Supplementary Figures S11B and D).

Evaluation of oxo- ϵ A genotoxicity in *E. coli*

Lesions such as ϵ A are very toxic to replication, so it was of interest to determine the level of genotoxicity associated with replication of DNA containing oxo- ϵ A. One established method of estimating the ability of the DNA lesion to inhibit replication is the competitive replication of adduct bypass (CRAB) assay based on mixing the lesion-bearing genome with a non-lesion competitor genome prior to transfection and calculating the replication blocking ability from the decrease in the lesion-to-competitor output signal ratio (40). More recently, we have shown that the same result can be inferred from comparing the number of sequenced fragments (i.e. sequencing reads) corresponding to the lesion-containing genomes and lesion-free control genomes; when starting with an equimolar mix of genomes, the ratio between the number of sequencing reads provides a

direct estimate of the bypass efficiency, similarly to the traditional CRAB assay (41). We electroporated an equimolar mixture of adenine- and oxo- ϵ A-bearing genomes into MutM/MutY-proficient and -deficient strains of *E. coli*, which allowed us to use the adenine-bearing genome as a competitor for oxo- ϵ A. In these settings, a post-replicative ratio of lesion genome to control genome equal to 1 reflects 100% bypass. Distribution of the coverage at the lesion site (Supplementary Figure S10) showed no statistically significant difference between the number of reads corresponding to the adenine control genome and to the oxo- ϵ A genome, indicating that oxo- ϵ A does not impact the efficacy of replication in any of the *E. coli* strains analysed.

DISCUSSION

We describe here a novel DNA modification, 7,8-dihydro-8-oxo-1, N^6 -ethenoadenine, which possesses a precise arrangement of H-bond donors and acceptors that enables pairing between its Hoogsteen face and an opposite adenine in the double helix. Oxo- ϵ A, being a combination of two naturally occurring forms of DNA damage, oxoA and ϵ A, was anticipated to be mutagenic due to the predicted ability to interact with adenine. Thorough evaluation of the influence of oxo- ϵ A on DNA duplex thermal stability confirmed the hypothesis that predicted a thymine-like behaviour of oxo- ϵ A (Figure 2). The melting temperature trend for oxo- ϵ A:A within eight nucleotide neighboring contexts (AXA, TXT, CXC, GXG, GXC, GXT, TXC and TXG) was consistent with data on the T:A pair. Moreover, the pK_a of the N7 atom of the model oxo- ϵ A derivative (9-ethyl-8-oxo- ϵ A), which acts as an H-bond acceptor, is equal to 9.49 (54). This value is close to the pK_a of the thymine N3 atom and, therefore, the modification mimics well enough thymine from the point of view of H-bond proton donating ability (55). In agreement with CD spectroscopy results, the NMR structure clearly shows that the oxo- ϵ A:A base pair does not provoke major distortions in the local duplex structure. Additional evidence of the oxo- ϵ A syn conformation was provided by molecular modelling showing that oxo- ϵ A in the anti conformation should lead to significant deformations of the sugar-phosphate backbone. Thus, oxo- ϵ A in the anti conformation is an unlikely occurrence in DNA duplex (Supplementary Figure S8A). Given its base-pairing preference, oxo- ϵ A constitutes a purine that functions as a structural and functional thymine mimic in DNA duplexes.

Many DNA lesions are mutagenic because they undergo rotation about glycosidic bond and can interact during replication with non-classically complementary opposite-strand nucleobases. Furthermore, transient Hoogsteen base pairs exist even in canonical duplex DNA (56,57) and can contribute to specific DNA-protein interactions and formation of non-canonical DNA structures (e.g. G-quadruplexes). So, oxo- ϵ A can be thought of as a limit case where the equilibrium is completely shifted towards a Hoogsteen base pair (Figure 1E). To test the hypothesis that a combination of chemical modifications of adenine can fix the syn conformation of the nucleobase, we studied its mutagenic properties. We found that oxo- ϵ A in DNA overwhelmingly pairs with adenine during elongation *in vitro* and replication *in vivo*, which, when considering A as the

reference base, leads to a predominant A \rightarrow T mutational outcome. Our *in vivo* studies indicate that oxo- ϵ A is extraordinarily mutagenic in *E. coli*, independent of the trinucleotide sequence context. While the chemical parents of oxo- ϵ A (oxoA and ϵ A) are individually vulnerable to repair by BER and AlkB, respectively, their combination in a single molecule is a powerful mutagen that cannot be repaired by these systems. Because some DNA lesions are repaired in a sequence context-dependent manner, we thought that the potent mutagenic properties and resistance to DNA repair might be attributable to the 3-base context in which the lesion was present in the genome. As one example, context-dependent DNA repair *in vivo* is observed with the methylated base, O^6 -methylguanine (52). We found that oxo- ϵ A causes A \rightarrow T transversions in all 16 possible trinucleotide contexts with almost equal efficacy in repair-proficient WT cells and in cells lacking AlkB (Figure 8). Thus, oxo- ϵ A is a potent mutagen in all possible contexts with an inherent mutation frequency above 97%. Such strong mutagenicity together with low genotoxicity suggests that oxo- ϵ A pairs with adenine as efficiently as thymine—the canonical adenine partner—and cannot be detected as DNA damage by *E. coli*. The lack of genotoxicity is remarkable and probably reflects the ability of the modification to masquerade as thymine, allowing facile polymerase bypass, as well as being invisible to repair systems.

It is puzzling that oxo- ϵ A is not an obvious substrate for common repair pathways in *E. coli*. However, it should be borne in mind that DNA with oxo- ϵ A in our study was single-stranded when entering the cell, and therefore, some repair factors might not be fully operative against single-stranded DNA. It is known that both MutM and MutY will ignore oxo- ϵ A in single stranded DNA but, once replication occurs, oxo- ϵ A predominantly exists in an oxo- ϵ A:A base pair (Figures 4 and 5). If this pair is a substrate for MutM, this enzyme will excise oxo- ϵ A and leave behind the A on the opposite strand, which would direct other BER enzymes to place a T opposite this A, leading to the observed A \rightarrow T mutations. Similarly, if the oxo- ϵ A:A pair is recognized by MutY, which normally excises A from oxoG:A mismatches, the A will be removed leaving an abasic site opposite oxo- ϵ A. It is possible that the BER polymerase (pol I) would repeat the mistake and place A once again across from oxo- ϵ A, leading to the same mutagenic outcome, or worse, futile cycling and delayed phage replication (Figure 9). Taken together, the *E. coli* MutM and MutY glycosylases are at best inactive against oxo- ϵ A, and at worst, they may even facilitate mutation formation rather than participate in lesion repair. These considerations are consistent with the data of primer elongation *in vitro* and the proposed model that oxo- ϵ A is smoothly bypassed without attendant toxicity and mostly pairs with adenine, yielding the almost quantitative A \rightarrow T mutation observed.

Another DNA repair pathway that could operate on oxo- ϵ A is AlkB-mediated direct damage reversal (45). The Fe(II)- and alpha-ketoglutarate-dependent dioxygenase AlkB recognizes ϵ A and directly removes its etheno bridge to restore adenine in both single-stranded and double-stranded DNA. The innate activity of AlkB renders ϵ A non-mutagenic in WT *E. coli* (34,58). However, in the present study, the mutagenic ability of oxo- ϵ A was unaf-

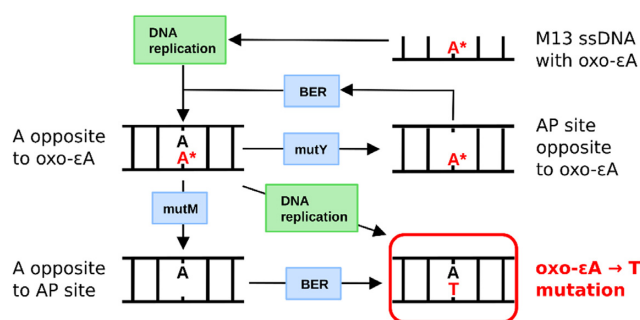


Figure 9. Proposed scheme of oxo-εA-driven mutagenesis in *E. coli* and the involvement of the BER pathway. AP site is an apurinic site.

ected by the AlkB status of the cells (Figure 8). This observation indicates that the addition of the C8 oxygen to the εA lesion eliminated the affinity of AlkB to oxo-εA. We note that AlkB-mediated repair requires coordination of Fe²⁺ atom between two oxygens of alpha-ketoglutarate and two oxygens of Asp133 of the enzyme (45). For oxo-εA the presence of the extra oxygen on C8 may disrupt this coordination environment, as the protonated N7 would fail to build a bridge with Q132, crucial for positioning of the damaged base in the enzyme active site (59). As AlkB is a fast enzyme that repairs εA in both single-stranded and double-stranded DNA, the syn conformation of the lesion should not influence the repair process (60). Yet another possibility is that even if the AlkB catalyzed oxidation occurs, the epoxide and diol intermediates may be longer-lived compared to the intermediates of the εA reaction with AlkB. Regardless of these possible scenarios, our data show that AlkB activity is unable to prevent the oxo-εA:A mispairing and its subsequent mutagenic outcome.

In addition to direct damage reversal and base excision, one could speculate that oxo-εA could be repaired by other DNA repair systems, such as nucleotide excision repair (NER) and mismatch repair (MMR). NER recognizes the distortions in the DNA double helix caused by the DNA lesions, rather than the lesions themselves (61). However, oxo-εA placed opposite adenine in the DNA duplex did not significantly alter the geometry and stability of helix, and therefore, it is unlikely that NER could contribute to the repair of oxo-εA in our settings.

The *E. coli* MMR system recognizes a mismatched base and nicks the strand with a wrong base to initiate the excision reaction. This system is highly efficient in reducing G:C→T:A transversions and moderately efficient in preventing A:T→T:A transversions (62), and its proximal enzyme, MutS, was shown to recognize 8-oxoG incorporated from the nucleotide pool into the nascent DNA chain (63). MMR determines the newly synthesized strand by the absence of adenine methylation at a GATC site. [reviewed here (64)]. The single-stranded M13 phage DNA used in our experiments to replicate lesion-bearing and control nucleotides in *E. coli* is not methylated, and theoretically any strand—either the original or complementary one—could be considered as a ‘nascent’ strand by MMR, and therefore could be repaired via MMR. The opportunity of either DNA strand repair was demonstrated on the *E. coli* cells

with inactivated *dam* gene responsible for DNA methylation (65). However, as oxo-εA pairs with A, whichever strand is excised, the result will be the same: an A:T→T:A mutation at the lesion position. Therefore, even if MMR would act upon the oxo-εA:A pairing, its effect would be undistinguishable.

One additional important and interesting feature of oxo-εA is its high efficiency of replication. Unlike other bridged adenine lesions such as εA, 1,N⁶-ethano-adenine and 1,N⁶-butadiene-adenine, which are strong replication blockers and not very mutagenic (29,66,67), oxo-εA is highly mutagenic and does not affect replication rates in *E. coli* (Supplementary Figure S10). Once again, the best mechanistic explanation is that, unlike other alkyl adenine lesions, oxo-εA is seamlessly accommodated within the DNA duplex without provoking significant distortions that could disrupt DNA polymerase activity (Figure 5) or attract the attention of the repair enzymes. The 8-oxo group of oxo-εA and/or the protonated N7 offer the possibility of a new hydrogen bond (with an incoming A), a bond that does not exist for any of the aforementioned adenine lesions. Hence, we cannot exclude the possibility that oxo-εA can be recognized as a non-instructional base/lesion because, as a default, bacterial DNA polymerases tend to insert A opposite to non-instructional lesions and AP sites (the A-rule) (68). This phenomenon would also account for the dominant A→T transversion that we observe here. However, in most cases the A-rule is invoked only when bypass polymerases are induced (the SOS response). While various stresses can induce the SOS response, common electroporation techniques that we used in this study, are known to be safe and pose no risk of genotoxicity (69). Thus, our results do not support the potential contribution of SOS-inducible DNA polymerases in the observed A→T transversions.

The absence of a repair system for oxo-εA in *E. coli* together with the efficient replicative bypass of the lesion can offer unique opportunities for using oxo-εA as a synthetic model base for applications in biotechnology and medicine. For example, oxo-εA could be used for studying cellular repair processes (70) or developing MutY/MutT specific inhibitors that may find utility as antitumor drugs (71–73). In addition, it could find utility in constructing thermodynamically stable and high affinity aptamers and other complex supramolecular DNA structures to target living cells and organisms. Aptamers are single stranded DNA and RNA oligonucleotides that can fold into defined architectures and, like antibodies, can recognize other types of molecules, including metal ions, small molecules, toxins, proteins and DNA *in vitro* and *in vivo* (74–76). Since oxo-εA has an extended aromatic system which stacking properties differ from thymine according to our UV melting studies, it can provide desirable hydrophobic and stacking contacts with target nucleic acids, proteins, peptides or small molecules, while maintaining the thymine-like H-bonding pattern. In this regard, oxo-εA can be useful in heavily-modified aptamers like SOMAmers with functional groups absent in natural nucleic acids that can improve SELEX success rates, especially for difficult protein targets (77). As oxo-εA can participate in metal-mediated pairing (54), one additional level of complexity can be added to form desired folded non-canonical 3D structures based on nucleic acids.

Moreover, the fluorescent properties of oxo- ϵ A (54) provide an additional axis for the development of oligonucleotide therapeutics that can be used to visualize unbiased intracellular and intra-organismal distribution of nucleic acids (78), analysis of DNA structures and DNA–protein interactions.

CONCLUSIONS

Oxo- ϵ A is a DNA modification that combines two naturally-occurring lesions, 7,8-dihydro-8-oxoadenine and 1, N^6 -ethenoadenine. Using UV-, CD- and NMR spectroscopies and molecular modeling, this modification in DNA duplexes was shown to adopt primarily a syn conformation and mimic the base-pairing behavior of the natural base thymine. Our molecular design strategy conceptualized a modified nucleobase for which predefined hybridization properties are achieved by blocking ability of one of the rotamers to interact with the opposite nucleobase. Importantly, the modification caused mostly A→T transversion mutations *in vitro* and in *E. coli* as evidence of its thymine mimicry, and was not repaired by either BER enzymes (MutM and MutY) or the direct-reversal AlkB enzyme, which removes the etheno bridge from 1, N^6 -ethenoadenine. In addition, oxo- ϵ A was not a significant DNA replication inhibitor in *E. coli* and its mutagenic properties did not depend on sequence context. Moreover, the ability of oxo- ϵ A to mimic thymine by adopting a syn conformation during replication emphasizes the role of conformational equilibria in nucleic acids in living cells and can be beneficial for multiple practical applications. Indeed, its minimal local distortions of nucleic acid architecture and ability to evade DNA repair systems are assets for those applications. Thus, oxo- ϵ A, being a near perfect and fluorescent thymine mimic, will find applications in DNA-based molecular tools to monitor rearrangements and intracellular distribution of DNA and DNA–protein complexes without distortion of DNA architecture and to develop chemically diverse modified aptamers. We expect that the opportunity to stabilize the syn conformation of nucleobase and to fine-tune Hoogsteen base-pairing ability opens up prospects for the creation of similar modified DNA-based molecular tools.

DATA AVAILABILITY

Atomic coordinates and structural factors have been deposited with the Protein Data bank under accession number 7NBP.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Lina Kim (ETH, Zurich) and Dr Maxim Pyatkov (Boston University) for technical advice regarding NGS data analysis pipeline. We acknowledge the “Manuel Rico” NMR laboratory (LMR), a node of the Spanish ICTS R-LRB.

FUNDING

MIT Skoltech Next Generation Program Pilot Grant (to J.M.E.); National Institutes of Health (NIH) [R01-CA080024 to J.M.E.]; NIEHS Center Grant [P30-ES002109 (to Center for Environmental Health Sciences, which provided access to NGS facilities)]; Skoltech (to T.S.Z.); MICINN [PID2020-116620GB-I00 to C.G.]; Ministry of Science and Higher Education Russian Federation [07515-2021-1049 to A.V.A. – synthesis and UV/CD studies]. Funding for open access charge: Skoltech. *Conflict of interest statement.* None declared.

REFERENCES

- Anisenko, A.N., Knyazhanskaya, E.S., Zalevsky, A.O., Agapkina, J.Y., Sizov, A.I., Zatselin, T.S. and Gottikh, M.B. (2017) Characterization of HIV-1 integrase interaction with human Ku70 protein and initial implications for drug targeting. *Sci. Rep.*, **7**, 5649.
- Clark, J.L., Hollecker, L., Mason, J.C., Stuyver, L.J., Tharnish, P.M., Lostia, S., McBrayer, T.R., Schinazi, R.F., Watanabe, K.A., Otto, M.J. *et al.* (2005) Design, synthesis, and antiviral activity of 2'-deoxy-2'-fluoro-2'-C-methylecytidine, a potent inhibitor of hepatitis C virus replication. *J. Med. Chem.*, **48**, 5504–5508.
- Freier, S.M. and Altmann, K.H. (1997) The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.*, **25**, 4429–4443.
- Schlegel, M.K., Foster, D.J., Kef'ina, A.V., Zlatev, I., Bisbe, A., Jayaraman, M., Lackey, J.G., Rajeev, K.G., Charisse, K., Harp, J. *et al.* (2017) Chirality dependent potency enhancement and structural impact of glycol nucleic acid modification on siRNA. *J. Am. Chem. Soc.*, **139**, 8537–8546.
- Mayer, G. (2009) The chemical biology of aptamers. *Angew. Chem. Int. Edit.*, **48**, 2672–2689.
- Crooke, S.T., Seth, P.P., Vickers, T.A. and Liang, X.H. (2020) The interaction of phosphorothioate-containing RNA targeted drugs with proteins is a critical determinant of the therapeutic effects of these agents. *J. Am. Chem. Soc.*, **142**, 14754–14771.
- Liczner, C., Duke, K., Juneau, G., Egli, M. and Wilds, C.J. (2021) Beyond ribose and phosphate: selected nucleic acid modifications for structure/function investigations and therapeutic applications. *Beilstein J. Org. Chem.*, **17**, 908–931.
- Beck, J.D., Reidenbach, D., Salomon, N., Sahin, U., Tureci, O., Vormehr, M. and Kranz, L.M. (2021) mRNA therapeutics in cancer immunotherapy. *Mol. Cancer*, **20**, 69.
- Chen, T., Hongdilokkul, N., Liu, Z.X., Thirunavukarasu, D. and Romesberg, F.E. (2016) The expanding world of DNA and RNA. *Curr. Opin. Chem. Biol.*, **34**, 80–87.
- Malyshev, D.A., Dhama, K., Lavergne, T., Chen, T.J., Dai, N., Foster, J.M., Correa, I.R. and Romesberg, F.E. (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature*, **509**, 385–388.
- Hirao, I. and Kimoto, M. (2012) Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma. *P. Jpn. Acad. B-Phys.*, **88**, 345–367.
- Benner, S.A., Yang, Z.Y. and Chen, F. (2011) Synthetic biology, tinkering biology, and artificial biology. What are we learning? *Cr. Chim.*, **14**, 372–387.
- Horn, T., Chang, C.A. and Collins, M.L. (1995) Hybridization properties of the 5-methyl-isocytidine/isoguanosine base pair in synthetic oligodeoxynucleotides. *Tetrahedron Lett.*, **36**, 2033–2036.
- Liu, Q., Liu, G.C., Wang, T., Fu, J., Li, R.J., Song, L.L., Wang, Z.G., Ding, B.Q. and Chen, F. (2017) Enhanced stability of DNA nanostructures by incorporation of unnatural base pairs. *Chem. Phys. Chem.*, **18**, 2977–2980.
- Grollman, A.P. and Moriya, M. (1993) Mutagenesis by 8-oxoguanine – an enemy within. *Trends Genet.*, **9**, 246–249.
- Kamiya, H. (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides. *Nucleic Acids Res.*, **31**, 517–531.

17. Bonicel, A., Mariaggi, N., Hughes, E. and Teoule, R. (1980) *In vitro* gamma-irradiation of DNA - identification of radio-induced chemical modifications of the adenine moiety. *Radiat Res.*, **83**, 19–26.
18. Cooke, M.S., Evans, M.D., Dizdaroglu, M. and Lunec, J. (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.*, **17**, 1195–1214.
19. Wood, M.L., Esteve, A., Morningstar, M.L., Kuziemko, G.M. and Essigmann, J.M. (1992) Genetic effects of oxidative DNA damage - comparative mutagenesis of 7,8-dihydro-8-oxoguanine and 7,8-dihydro-8-oxoadenine in *escherichiacoli*. *Nucleic Acids Res.*, **20**, 6023–6032.
20. Batra, V.K., Shock, D.D., Beard, W.A., McKenna, C.E. and Wilson, S.H. (2012) Binary complex crystal structure of DNA polymerase β reveals multiple conformations of the templating 8-oxoguanine lesion. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 113–118.
21. Hsu, G.W., Ober, M., Carell, T. and Beese, L.S. (2004) Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature*, **431**, 217–221.
22. Brieba, L.G., Eichman, B.F., Kokoska, R.J., Double, S., Kunkel, T.A. and Ellenberger, T. (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J.*, **23**, 3452–3461.
23. Rechkoblit, O., Malinina, L., Cheng, Y., Geacintov, N.E., Broyde, S. and Patel, D.J. (2009) Impact of conformational heterogeneity of oxoG lesions and their pairing partners on bypass fidelity by γ family polymerases. *Structure*, **17**, 725–736.
24. Beard, W.A., Batra, V.K. and Wilson, S.H. (2010) DNA polymerase structure-based insight on the mutagenic properties of 8-oxoguanine. *Mutat. Res.-Gen. Environ.*, **703**, 18–23.
25. Wood, M.L., Dizdaroglu, M., Gajewski, E. and Essigmann, J.M. (1990) Mechanistic studies of ionizing-radiation and oxidative mutagenesis - genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. *Biochemistry*, **29**, 7024–7032.
26. Koag, M.C., Jung, H. and Lee, S. (2019) Mutagenic replication of the major oxidative adenine lesion 7,8-dihydro-8-oxoadenine by human DNA polymerases. *J. Am. Chem. Soc.*, **141**, 4584–4596.
27. Leonard, G.A., Guy, A., Brown, T., Teoule, R. and Hunter, W.N. (1992) Conformation of guanine:8-oxoadenine base pairs in the crystal structure of d(Cgcaatt(O8a)Gcg). *Biochemistry*, **31**, 8415–8420.
28. Koag, M.C., Jung, H. and Lee, S. (2020) Mutagenesis mechanism of the major oxidative adenine lesion 7,8-dihydro-8-oxoadenine. *Nucleic Acids Res.*, **48**, 5119–5134.
29. Delaney, J.C., Smeester, L., Wong, C.Y., Frick, L.E., Taghizadeh, K., Wishnok, J.S., Drennan, C.L., Samson, L.D. and Essigmann, J.M. (2005) AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nat. Struct. Mol. Biol.*, **12**, 855–860.
30. Guengerich, F.P., Crawford, W.M. and Watanabe, P.G. (1979) Activation of vinyl-chloride to covalently bound metabolites - roles of 2-chloroethylene oxide and 2-chloroacetaldehyde. *Biochemistry*, **18**, 5177–5182.
31. Elghissassi, F., Barbin, A., Nair, J. and Bartsch, H. (1995) Formation of 1,N-6-ethenoadenine and 3,N-4-ethenocytosine by lipid-peroxidation products and nucleic-acid bases. *Chem. Res. Toxicol.*, **8**, 278–283.
32. Santos, C.D., Kouchakdjian, M., Yarema, K., Basu, A., Essigmann, J. and Patel, D.J. (1991) NMR studies of the exocyclic 1,N6-ethenodeoxyadenosine adduct (epsilon dA) opposite deoxyguanosine in a DNA duplex. edA(syn)-dG(anti) pairing at the lesion site. *Biochemistry*, **30**, 1828–1835.
33. Pandya, G.A. and Moriya, M. (1996) 1,N-6-ethenodeoxyadenosine, a DNA adduct highly mutagenic in mammalian cells. *Biochemistry*, **35**, 11487–11492.
34. Levine, R.L., Yang, I.Y., Hossain, M., Pandya, G.A., Grollman, A.P. and Moriya, M. (2000) Mutagenesis induced by a single 1,N-6-ethenodeoxyadenosine adduct in human cells. *Cancer Res.*, **60**, 4098–4104.
35. Tolentino, J.H., Burke, T.J., Mukhopadhyay, S., McGregor, W.G. and Basu, A.K. (2008) Inhibition of DNA replication fork progression and mutagenic potential of 1,N-6-ethenoadenine and 8-oxoguanine in human cell extracts. *Nucleic Acids Res.*, **36**, 1300–1308.
36. Moran, S., Ren, R.X.F., Rumney, S. and Kool, E.T. (1997) Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.*, **119**, 2056–2057.
37. Moran, S., Ren, R.X.F. and Kool, E.T. (1997) A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10506–10511.
38. Malyshev, D.A. and Romesberg, F.E. (2015) The expanded genetic alphabet. *Angew. Chem. Int. Ed.*, **54**, 11930–11944.
39. Steinmetzger, C., Bauerlein, C. and Hobartner, C. (2020) Supramolecular fluorescence resonance energy transfer in nucleobase-modified fluorogenic RNA aptamers. *Angew. Chem. Int. Edit.*, **59**, 6760–6764.
40. Delaney, J.C. and Essigmann, J.M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions in vivo. *Method Enzymol.*, **408**, 1–15.
41. Chang, S.C., Fedeles, B.I., Wu, J., Delaney, J.C., Li, D.Y., Zhao, L.L., Christov, P.P., Yau, E., Singh, V., Jost, M. *et al.* (2015) Next-generation sequencing reveals the biological significance of the N-2,3-ethenoguanine lesion in vivo. *Nucleic Acids Res.*, **43**, 5489–5500.
42. Chatgililoglu, C., Navacchia, M.L. and Postigo, A. (2006) A facile one-pot synthesis of 8-oxo-7,8-dihydro-(2'-deoxy)adenosine in water. *Tetrahedron Lett.*, **47**, 711–714.
43. Kouchakdjian, M., Eisenberg, M., Yarema, K., Basu, A., Essigmann, J. and Patel, D.J. (1991) NMR studies of the exocyclic 1,N(6)-ethenodeoxyadenosine adduct (ϵ -dA) opposite thymidine in a DNA duplex - nonplanar alignment of ϵ -dA(anti) and dT(anti) at the lesion site. *Biochemistry*, **30**, 1820–1828.
44. Chen, J.J., Tsai, C.H., Cai, X., Horhota, A.T., McLaughlin, L.W. and Szostak, J.W. (2009) Enzymatic primer-extension with glycerol-nucleoside triphosphates on DNA templates. *PLoS One*, **4**, e4949.
45. Fedeles, B.I., Singh, V., Delaney, J.C., Li, D.Y. and Essigmann, J.M. (2015) The AlkB family of Fe(II)/alpha-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.*, **290**, 20734–20742.
46. Talhaoui, I., Couve, S., Ishchenko, A.A., Kunz, C., Schar, P. and Saparbaev, M. (2013) 7,8-dihydro-8-oxoadenine, a highly mutagenic adduct, is repaired by *escherichiacoli* and human mismatch-specific uracil/thymine-DNA glycosylases. *Nucleic Acids Res.*, **41**, 912–923.
47. Girard, P.M., D'Ham, C., Cadet, J. and Boiteux, S. (1998) Opposite base-dependent excision of 7,8-dihydro-8-oxoadenine by the ogg1 protein of *Saccharomyces cerevisiae*. *Carcinogenesis*, **19**, 1299–1305.
48. Tchou, J., Kasai, H., Shibutani, S., Chung, M.H., Laval, J., Grollman, A.P. and Nishimura, S. (1991) 8-Oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate-specificity. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4690–4694.
49. Michaels, M.L., Cruz, C., Grollman, A.P. and Miller, J.H. (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 7022–7025.
50. David, S.S. and Williams, S.D. (1998) Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.*, **98**, 1221–1261.
51. Hazra, T.K., Hill, J.W., Izumi, T. and Mitra, S. (2001) Multiple DNA glycosylases for repair of 8-oxoguanine and their potential *in vivo* functions. *Prog. Nucleic Acid Res.*, **68**, 193–205.
52. Delaney, J.C. and Essigmann, J.M. (2001) Effect of sequence context on O-6-methylguanine repair and replication in vivo. *Biochemistry*, **40**, 14968–14975.
53. Fedeles, B.I. and Essigmann, J.M. (2018) Impact of DNA lesion repair, replication and formation on the mutational spectra of environmental carcinogens: aflatoxin B-1 as a case study. *DNA Repair*, **71**, 12–22.
54. Schonrath, I., Tsvetkov, V.B., Barcelo-Oliver, M., Hebenbrock, M., Zatsepin, T.S., Aralov, A.V. and Muller, J. (2021) Silver(I)-mediated base pairing in DNA involving the artificial nucleobase 7,8-dihydro-8-oxo-1,N-6-ethenoadenine. *J. Inorg. Biochem.*, **219**, 111369.
55. Krishnamurthy, R. (2012) Role of pK(a) of nucleobases in the origins of chemical evolution. *Acc. Chem. Res.*, **45**, 2035–2044.
56. Nikolova, E.N., Kim, E., Wise, A.A., O'Brien, P.J., Andricioaei, I. and Al-Hashimi, H.M. (2011) Transient hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–502.
57. Alvey, H.S., Gottardo, F.L., Nikolova, E.N. and Al-Hashimi, H.M. (2014) Widespread transient hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat. Commun.*, **5**, 4786.

58. Basu, A.K., Wood, M.L., Niedernhofer, L.J., Ramos, L.A. and Essigmann, J.M. (1993) Mutagenic and genotoxic effects of vinyl chloride-induced DNA lesions - 1,N(6)-ethanoadenine, 3,N(4)-ethenocytosine, and 4-amino-5-(Imidazol-2-yl)imidazole. *Biochemistry*, **32**, 12793–12801.
59. Yi, C.Q., Jia, G.F., Hou, G.H., Dai, Q., Zhang, W., Zheng, G.Q., Jian, X., Yang, C.G., Cui, Q.A. and He, C.A. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330–333.
60. Baldwin, M.R., Admiraal, S.J. and O'Brien, P.J. (2020) Transient kinetic analysis of oxidative dealkylation by the direct reversal DNA repair enzyme alkB. *J. Biol. Chem.*, **295**, 7317–7326.
61. Scharer, O.D. (2007) Achieving broad substrate specificity in damage recognition by binding accessible nondamaged DNA. *Mol. Cell.*, **28**, 184–186.
62. Long, H.A., Miller, S.F., Williams, E. and Lynch, M. (2018) Specificity of the DNA mismatch repair system (MMR) and mutagenesis bias in bacteria. *Mol. Biol. Evol.*, **35**, 2414–2421.
63. Colussi, C., Parlanti, E., Degan, P., Aquilina, G., Barnes, D., Macpherson, P., Karran, P., Crescenzi, M., Dogliotti, E. and Bignami, M. (2002) The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool. *Curr. Biol.*, **12**, 912–918.
64. Fukui, K. (2010) DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids*, **2010**, 260512.
65. Marinus, M.G. and Morris, N.R. (1974) Biological function for 6-methyladenine residues in the DNA of *Escherichia coli* K12. *J. Mol. Biol.*, **85**, 309–322.
66. Frick, L.E., Delaney, J.C., Wong, C., Drennan, C.L. and Essigmann, J.M. (2007) Alleviation of 1,N-6-ethanoadenine genotoxicity by the *escherichiacoli* adaptive response protein alkB. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 755–760.
67. Chang, S.C., Seneviratne, U.I., Wu, J., Tretyakova, N. and Essigmann, J.M. (2017) 1,3-Butadiene-induced adenine DNA adducts are genotoxic but only weakly mutagenic when replicated in *escherichiacoli* of various repair and replication backgrounds. *Chem. Res. Toxicol.*, **30**, 1230–1239.
68. Strauss, B., Rabkin, S., Sagher, D. and Moore, P. (1982) The role of DNA-polymerase in base substitution mutagenesis on non-instructional templates. *Biochimie*, **64**, 829–838.
69. Gusbeth, C., Frey, W., Volkmann, H., Schwartz, T. and Bluhm, H. (2009) Pulsed electric field treatment for bacteria reduction and its impact on hospital wastewater. *Chemosphere*, **75**, 228–233.
70. Majumdar, C., McKibbin, P.L., Krajewski, A.E., Manlove, A.H., Lee, J.K. and David, S.S. (2020) Unique hydrogen bonding of adenine with the oxidatively damaged base 8-oxoguanine enables specific recognition and repair by DNA glycosylase mutY. *J. Am. Chem. Soc.*, **142**, 20340–20350.
71. Gad, H., Koolmeister, T., Jemth, A.S., Eshtad, S., Jacques, S.A., Ström, C.E., Svensson, L.M., Schultz, N., Lundbäck, T., Einarsson, B.O. *et al.* (2014) MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature*, **508**, 215–221.
72. Berglund, U.W., Sanjiv, K., Gad, H., Kalderén, C., Koolmeister, T., Pham, T., Gokturk, C., Jafari, R., Maddalo, G., Seashore-Ludlow, B. *et al.* (2016) Validation and development of MTH1 inhibitors for treatment of cancer. *Ann. Oncol.*, **27**, 2275–2283.
73. Sharbeen, G., Youkhana, J., Mawson, A., McCarroll, J., Nunez, A., Biankin, A., Johns, A., Goldstein, D. and Phillips, P. (2017) MutY-homolog (MYH) inhibition reduces pancreatic cancer cell growth and increases chemosensitivity. *Oncotarget*, **8**, 9216–9229.
74. Bouvier-Muller, A. and Duconge, F. (2018) Application of aptamers for *in vivo* molecular imaging and theranostics. *Adv. Drug. Deliver. Rev.*, **134**, 94–106.
75. Ismail, S.I. and Alshaer, W. (2018) Therapeutic aptamers in discovery, preclinical and clinical stages. *Adv. Drug. Deliver. Rev.*, **134**, 51–64.
76. Hermann, T. and Patel, D.J. (2000) Biochemistry - adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
77. Davies, D.R., Gelinas, A.D., Zhang, C., Rohloff, J.C., Carter, J.D., O'Connell, D., Waugh, S.M., Wolk, S.K., Mayfield, W.S., Burgin, A.B. *et al.* (2012) Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 19971–19976.
78. Nilsson, J.R., Baladi, T., Gallud, A., Baždarević, D., Lemurell, M., Esbjörner, E.K., Wilhelmsson, L.M. and Dahlén, A. (2021) Fluorescent base analogues in gapmers enable stealth labeling of antisense oligonucleotide therapeutics. *Sci. Rep.*, **11**, 11365.