Published online 02 April 2015 *Nucleic Acids Research, 2015, Vol. 43, No. 11* ***5489–5500***
doi: 10.1093/nar/gkv243

NAR Breakthrough Article

Next-generation sequencing reveals the biological significance of the N^2,3-ethenoguanine lesion *in vivo*

Shiou-chi Chang[1,2], Bogdan I. Fedeles[1,2,3], Jie Wu[4], James C. Delaney[1,2,3], Deyu Li[1,2,3], Linlin Zhao[5,6], Plamen P. Christov[5,6,7], Emily Yau[1,2], Vipender Singh[1,2,3], Marco Jost[3], Catherine L. Drennan[3,8,9], Lawrence J. Marnett[5,6,7,10], Carmelo J. Rizzo[5,6,7,10], Stuart S. Levine[4], F. Peter Guengerich[5,6,10] and John M. Essigmann[1,2,3,*]

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States, [2]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, United States, [3]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, United States, [4]BioMicro Center, Massachusetts Institute of Technology, Cambridge, MA 02139, United States, [5]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, United States, [6]Center in Molecular Toxicology, Vanderbilt University, Nashville, TN 37232, United States, [7]Department of Chemistry, Vanderbilt University, Nashville, TN 37232, United States, [8]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, United States, [9]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, United States and [10]Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN 37232, United States

Received January 25, 2015; Revised March 06, 2015; Accepted March 09, 2015

ABSTRACT

Etheno DNA adducts are a prevalent type of DNA damage caused by vinyl chloride (VC) exposure and oxidative stress. Etheno adducts are mutagenic and may contribute to the initiation of several pathologies; thus, elucidating the pathways by which they induce cellular transformation is critical. Although N^2,3-ethenoguanine (N^2,3-εG) is the most abundant etheno adduct, its biological consequences have not been well characterized in cells due to its labile glycosidic bond. Here, a stabilized 2′-fluoro-2′-deoxyribose analog of N^2,3-εG was used to quantify directly its genotoxicity and mutagenicity. A multiplex method involving next-generation sequencing enabled a large-scale *in vivo* analysis, in which both N^2,3-εG and its isomer 1,N^2-ethenoguanine (1,N^2-εG) were evaluated in various repair and replication backgrounds. We found that N^2,3-εG potently induces G to A transitions, the same mutation previously observed in VC-associated tumors. By contrast, 1,N^2-εG induces various substitutions and frameshifts. We also found that N^2,3-εG is the only etheno lesion that cannot be repaired by AlkB, which partially explains its persistence. Both εG lesions are strong replication blocks and DinB, a translesion polymerase, facilitates the mutagenic bypass of both lesions. Collectively, our results indicate that N^2,3-εG is a biologically important lesion and may have a functional role in VC-induced or inflammation-driven carcinogenesis.

INTRODUCTION

The integrity of genomic DNA is under constant pressure from the attacks by endogenous and exogenous DNA damaging agents. Damaged DNA bases, also known as DNA lesions or adducts, can lead to mutations or cell death if not properly repaired. Decades of research have shown that specific lifestyles and environments in which there is an increased exposure to DNA damaging agents are associated

*To whom correspondence should be addressed. Tel: +1 617 253 6227; Fax: +1 617 253 5445; Email: jessig@mit.edu
Present addresses:
James C. Delaney, Visterra, Inc., Cambridge, MA 02139, United States.
Deyu Li, Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI 02881, United States.
Linlin Zhao, Department of Chemistry and Biochemistry, Central Michigan University, Mount Pleasant, MI 48859, United States.
Plamen P. Christov, Chemical Synthesis Core, Vanderbilt Institute of Chemical Biology, Nashville, TN 37232, United States.

with an increased risk for specific cancers (1–4). Therefore, it is generally believed that the accumulation of mutations resulting from DNA lesions is one of the drivers of carcinogenesis. However, DNA damaging agents usually produce a spectrum of lesions, making the identification of the biologically relevant DNA lesions a challenging task.

Vinyl chloride (VC) carcinogenesis is a model for environmentally induced disease because it induces a rare type of tumor, angiosarcoma of the liver (5). Since the 1970s, hepatic angiosarcoma has been reported in workers industrially exposed to VC (5). Two independent analyses of the human hepatic angiosarcoma samples associated with VC exposure report a high number of G:C to A:T transitions in the K-*ras* gene, many of which are found at codon 13 (6,7). The resulting mutant $p21^{ras}$ protein is also detected in the sera of workers exposed to VC and its detection correlates with the VC exposure level (5). Interestingly, mutations at codon 13 of K-*ras* are less commonly observed in other types of cancer (8–13). Mutations in K-*ras* have also been detected in nonneoplastic liver tissues and a preangiosarcoma lesion from patients exposed to VC (6,7). Furthermore, VC and its metabolites have been shown to be mutagenic in *Salmonella typhimurium* (14) and induce G:C to A:T transitions in *Escherichia coli* (15). These findings collectively suggest that the G:C to A:T mutation in the K-*ras* gene may be a consequence of VC exposure.

The two-carbon exocyclic bridged nucleobase adducts, commonly known as the etheno adducts, have been proposed to be the mutagenic DNA lesions formed by VC metabolites reacting with DNA (5,16,17). In addition to exogenous sources, etheno adducts can also be formed when DNA is attacked by lipid peroxidation products, which are generated under inflammation and oxidative stress (18–22). Indeed, etheno lesions are detected in tissues of humans and rats without known exposure to exogenous carcinogens (23–26). Thus, understanding the mutagenic potential and the repair of etheno lesions is important for gaining more insights into the molecular mechanisms of VC-induced carcinogenesis as well as inflammation-driven cancers.

A total of four etheno lesions have been identified in DNA: 1,N^6-ethenoadenine (εA), 3,N^4-ethenocytosine (εC), 1,N^2-ethenoguanine (1,N^2-εG) and N^2,3-ethenoguanine (N^2,3-εG) (26). Their mutagenic potentials have been evaluated to various degrees *in vitro* and *in vivo* (reviewed in (16,17,27)). Although all etheno adducts have been found to be mutagenic, it remains unclear which etheno lesion is most associated with VC carcinogenesis. Interestingly, N^2,3-εG is the most abundant etheno lesion present endogenously (17,20) as well as after VC exposure (5), which highlights its potential biological importance. However, although relatively stable in DNA, the monomeric N^2,3-εG nucleoside or nucleotide rapidly depurinates due to its labile glycosidic bond (28), which has prevented detailed investigations on the mutagenicity and repair of N^2,3-εG. Using an indirect method, Cheng *et al.* measured a G to A mutation frequency of 0.5% for N^2,3-εG in *E. coli*, which was adjusted to 13% after applying a large correction factor (29). All these considerations motivated us to re-evaluate the biological consequences of N^2,3-εG *in vivo* using a direct method with high resolution.

Recently, we demonstrated that the glycosidic bond of N^2,3-εG can be stabilized by using fluorine as a non-classical isostere replacing a 2′-β-hydrogen on the deoxyribose (30,31). The stabilized 2′-fluoro-N^2,3-ε-2′-deoxyarabinoguanosine analog (2′-F-N^2,3-εG) can be site-specifically incorporated into oligonucleotides, thus allowing detailed mutational analyses to be performed. In this study, we analyzed the biological consequences of both εG lesions (2′-F-N^2,3-εG and 1,N^2-εG, Figure 1A) *in vivo* under various repair and replication states. Previously, we showed that AlkB, an iron(II)- and α-ketoglutarate-dependent dioxygenase, can repair εA and εC via a direct reversal mechanism (32). Thus, we asked whether AlkB could also repair the two εG lesions. The role of DinB (DNA polymerase IV) and the SOS response on the mutagenesis of the εG lesions were also investigated. DinB is a Y-family DNA polymerase specialized in translesion synthesis, a damage tolerance mechanism that allows cells to replicate DNA containing unrepaired, damaged bases (33,34). Although they increase cell survival, translesion polymerases have lower fidelities than regular polymerases and are believed to be responsible for the majority of lesion-induced mutagenesis (33,34). As DinB can efficiently bypass N^2-alkylguanine lesions (35–37), we hypothesized that it could play a role in the bypass of 1,N^2-εG and N^2,3-εG. Since the level of DinB expression increases 10-fold when the SOS response is induced in *E. coli* (34), the impact of DinB on lesion mutagenesis was studied in *dinB*$^+$ versus *dinB*$^-$ cells under SOS induction.

Traditionally, the *in vivo* evaluation of lesion genotoxicity and mutagenicity has been done by site-specifically inserting the lesion of interest into a vector, allowing the vector to replicate in host cells and then interrogating the progeny DNA biochemically or via mass spectrometry (MS) analysis (36,38). Although these techniques are effective in generating quantitative measurements, they suffer from low throughput as different lesion-containing vectors in different host cells have to be analyzed separately. Next-generation sequencing technology has offered an affordable and reliable way to perform massively parallel sequencing (39,40). In this work, we adopted and improved on a previously described next-generation sequencing approach (41,42), which enabled us to multiplex our site-specific mutagenesis assay (38) and quantify insertion and deletion mutations readily. Since multiple lesions were investigated in multiple repair and replication backgrounds in all possible combinations (Figure 1B), the resulting comprehensive dataset allowed us to gain deep insights into the repair and bypass mechanisms of these lesions. We found that N^2,3-εG potently induces G to A mutations and is not a substrate for the AlkB repair system. Our data, in conjunction with previous studies on N^2,3-εG, present a compelling argument that N^2,3-εG may have a higher biological significance than previously anticipated.

MATERIALS AND METHODS

Oligonucleotide synthesis

All lesion-free oligonucleotides were obtained from Integrated DNA Technologies. Oligonucleotide 16-mers (5′-GAAGACCTXGGCGTCC-3′, where X is the lesion) con-

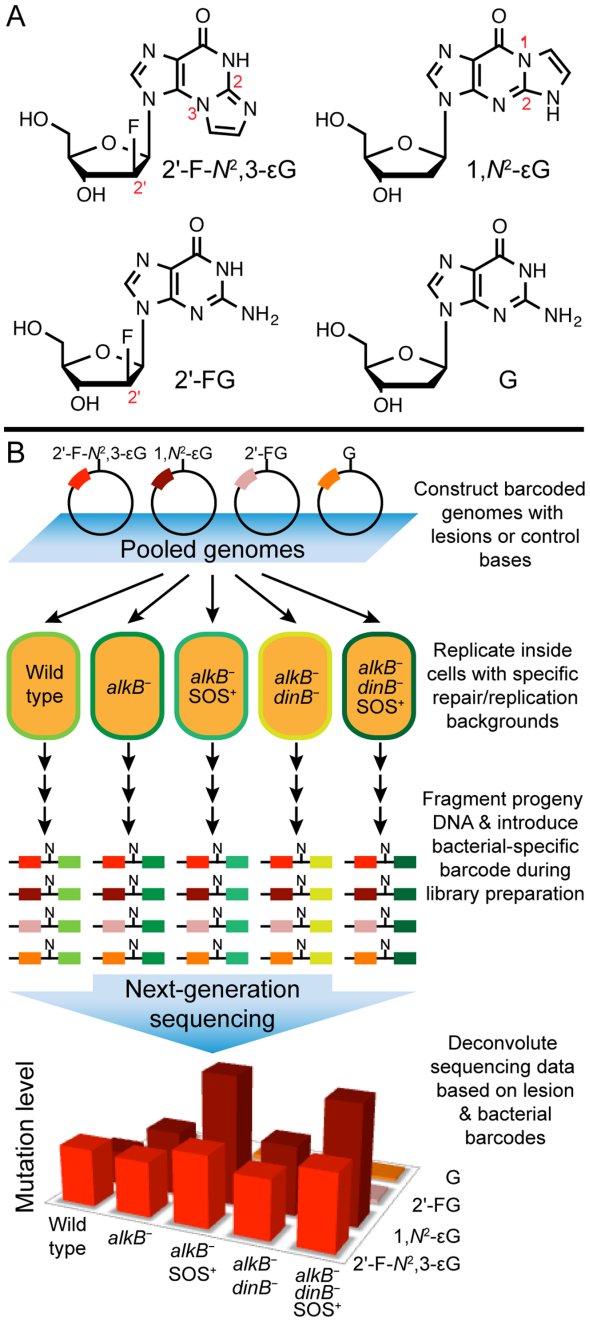


Figure 1. Experimental overview. (**A**) Structures of the modified DNA bases and controls (shown as deoxynucleosides) investigated for genotoxic and mutagenic properties. Numbers in red show the key atom positions on the nucleosides. (**B**) Schematic representation of the *in vivo* mutagenesis assay with next-generation sequencing. M13 single-stranded vectors, each containing a site-specific lesion and a lesion-specific barcode sequence, were mixed in a known ratio and introduced into cells with specific repair and replication backgrounds. After *in vivo* replication, progeny DNA from each repair/replication background was isolated, amplified and fragmented to generate sequencing libraries. N represents the site, in progeny, that had originally contained the lesion, and the colored box to the left of N symbolizes the lesion-specific barcode (Barcode 1). A second set of barcodes (Barcode 2, to the right of N), designating the repair/replication backgrounds and biological replicates were also introduced at the library preparation step. The resulting DNA was pooled and subjected to next-generation sequencing. The genotoxicity and mutagenicity of each lesion under each bacterial condition were determined from the sequencing data, which were sorted according to the two sets of barcodes.

taining a modified base, 2′-F-N^2,3-εG, 1,N^2-εG, 2′-F-1,N^2-εG or 2′-FG, were synthesized by using phosphoramidite solid-phase methods as described before (30,43). The 2′-F-1,N^2-εG phosphoramidite was synthesized similarly to the synthesis of the 1,N^2-εG phosphoramidite (44) but with 2′-FG as the starting material. The purity of oligonucleotides was evaluated by Electrospray Ionization Time Of Flight Mass Spectrometry (ESI-TOF MS).

Lesion-containing M13 genome construction

Barcoded lesion-containing genomes were constructed based on a previously reported method (38) with some modifications. First, 30 pmol of lesion-containing 16-mer (5′-GAAGACCTXGGCGTCC-3′, where X is the lesion) was ligated to 30 pmol of 18-mer containing a trinucleotide barcode unique to each lesion (5′-CACGGTB$_1$B$_2$B$_3$TGCTCTGAC-3′, where B$_1$B$_2$B$_3$ is the trinucleotide barcode) with T4 DNA ligase (New England Biolabs) and a scaffold sequence complementary to the 3′ end of the barcode-containing oligonucleotide and the 5′ end of the lesion-containing oligonucleotide (5′-GGTCTTCGTCAGAGCA-3′). The 34-mer ligation product was then ligated to 20 pmol of EcoRI (New England Biolabs) linearized M13mp7(L2) single-stranded bacteriophage genomic DNA with additional T4 DNA ligase and two genome construction scaffolds (5′-ACCGTGCACTGAATCATGGTCATAGC-3′ and 5′-AAAACGACGGCCAGTGAATTGGACGC-3′). After the removal of scaffolds by the 3′-exonuclease activity of T4 DNA polymerase (New England Biolabs), the ligated M13 genome was extracted with phenol/chloroform/isoamyl alcohol (25:24:1) (Invitrogen) and desalted with QIAquick PCR Purification Kit (Qiagen) following the manufacturer's protocol.

Genome concentration normalization

The following procedure determines the relative concentrations of different genome constructs, which is necessary for measuring lesion bypass efficiencies. The procedure is similar to that described previously (38) with one major difference in that a 69-mer oligonucleotide (5′- GGAGAA AATCAGATGGAAAGTATCAACACCGAGAGGAA CGCTCTGTCTACGGCTAACGACGCTCGTGAT-3′) was added as an internal standard to correct for sample loss during the procedure. Each desalted genome construct (5 μl of 25 nM) was mixed with 125 fmol of the internal standard, which was designed to be resistant to digestion by the restriction enzymes used in this procedure. The two genome construction scaffolds were annealed to the re-circularized genomes, which were then digested with HinFI (New England Biolabs) and dephosphorylated with shrimp alkaline phosphatase (Roche). The dephosphorylated 5′ ends and the internal standard were radiolabeled with [γ-^{32}P]ATP (Perkin Elmer) and OptiKinase (Affymetrix). Further digestion of the genome with HaeIII (New England Biolabs) yielded a 52-mer radiolabeled fragment. The digestion products were separated by denaturing polyacrylamide gel electrophoresis. Band intensities were quantified by using phosphorimagery (Typhoon 7000, GE

Healthcare Life Sciences). After accounting for sample loss based on the internal standard band intensities, the corrected intensities of the 52-mer bands reflected the relative concentrations of different genome constructs. To obtain accurate normalization ratios, this procedure was repeated three times and the averaged normalization ratios were used in the calculation of bypass efficiencies.

In vivo replication of DNA lesion-containing genomes

Electrocompetent cells of HK81 (as AB1157, but *nalA*; this is the wild type strain), HK82 (as HK81, but *alkB*$^-$), HK84 (as HK82, but *dinB*$^-$), HK82 SOS$^+$ (SOS-induced strain of HK82) and HK84 SOS$^+$ (SOS-induced strain of HK84) were prepared as described previously (38). SOS induction was achieved by irradiation of cells with 254 nm UV light at 45 J/m^2 (45). Genomes containing 2′-F-N^2,3-εG, 1,N^2-εG, 2′-F-1,N^2-εG, 2′-FG or G were mixed together at a 3:3:3:1:1 ratio. Higher amounts of lesion-containing genomes relative to the control genomes were used to ensure that a sufficient number of progeny was generated from the lesion-containing genomes. A total of 250 fmol of the genome mixture mixed with 150 μl of electrocompetent cells was exposed to ∼2.5 kV, typically delivered in >4.5 ms, in a 2-mm electroporation cuvette. After electroporation, cells were immediately transferred to 10 ml LB medium and grown for 6 h at 37°C. The number of successful transformation events for each electroporation was estimated as described previously (38). All electroporations resulted in at least 2.9 × 10^4 initial events (infective centers), with most of the samples producing >10^5 events. After the 6 h incubation, the progeny phage were amplified in SCS110 cells at 37°C for 7 h to dilute out the residual genomic DNA used for electroporation. Cells were pelleted and the supernatant containing progeny was isolated and stored at 4°C. Single-stranded DNA from the progeny phage was isolated with the QIAprep Spin M13 Kit (Qiagen). A ∼1 kb sequence covering the region of interest was PCR amplified using 1 fmol of the M13 template, 50 pmol of each PCR primers (forward: 5′-CGATTTCGGAACCACCATCAAACAGG-3′, reverse: 5′-TGAGAGTCTGGAGCAAACAAGAGAATCG-3′) and PfuTurbo polymerase (Agilent) for 25 cycles of 95°C for 30 s, 68°C for 30 s and 72°C for 1.25 min. The PCR product was purified with the QIAquick PCR Purification Kit (Qiagen) and stored at −20°C.

Sequencing library preparation and next-generation sequencing

For each of the PCR products originating from one independent electroporation, an Illumina sequencing library was prepared from 50 ng of the PCR product via the 'tagmentation' reaction by using the Nextera DNA Sample Preparation Kit (Illumina). Tagged DNA fragments were amplified and indexed with a second set of barcodes, designating samples from cells with a specific repair/replication background and from different biological replicates, by using the same kit following the manufacturer's instructions. Equal amounts of amplified DNA fragments from different samples were pooled and sequenced on two lanes of an Illumina MiSeq sequencer (200 + 200 paired-end, v3 chemistry).

Sequencing data analysis to determine lesion genotoxicity and mutagenicity

The sequencing reads were first separated into different groups based on the barcode designating different repair/replication backgrounds and biological replicates. The sequencing adaptors were trimmed with Trimmomatic 0.32 (46) (options used: 'ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:1:true') and the paired-end reads were concatenated using PEAR v0.9.3 (47) (with options '−n 10' and '−b 64'). Afterwards, an in-house script was used to extract the reads that originated from different lesions based on the trinucleotide lesion barcodes. In this step, we required perfect matches of the 13-mer around the lesion barcode (flanking ±5 nucleotides plus the trinucleotide barcode) and extracted the downstream sequences from the reads. To improve accuracy, we also replaced all low-quality bases that had 64-based Phred scores lower than 74 with 'N's (where N = an unidentified base). The extracted sequences were aligned to the M13 reference genome using Bowtie 2 version 2.1.0 (48). SAMtools 0.1.19 (49) was used to pileup the mapping results to find mutations (options used: 'mpileup −AB −d1000000'). Finally, the pileup files were parsed to compute the per-base mutation rate inside the region of interest.

Lesion bypass efficiency was calculated based on the principle that more genotoxic lesions would produce fewer progeny relative to the control due to their hindrance of DNA polymerase activity. The total number of reads containing each lesion barcode was counted. The bypass percentage of a particular lesion was calculated by dividing the number of sequencing reads from the lesion, after accounting for the mixing and normalization ratios, by that of the control.

In vitro repair of DNA lesions by AlkB

AlkB protein was purified based on a previously reported procedure (50) and all *in vitro* AlkB repair reactions utilized conditions similar to those described previously (32). For each incubation, 5 μM of the lesion-containing 16-mer oligonucleotide (5′-GAAGACCTXGGCGTCC-3′, where X is the lesion) was incubated with 10 μM of AlkB protein (or just the reaction buffer in case of no enzyme controls) in the reaction buffer containing 45 mM HEPES (pH 8.0), 100 μM $Fe(NH_4)_2(SO_4)_2$, 0.9 mM α-ketoglutarate and 1.8 mM ascorbate. After incubation at 37°C for 1 h, the reaction mixtures were analyzed by HPLC–ESI-TOF MS. HPLC separation was performed by using a Zorbax SB-Aq column (2.1 × 150 mm, 3.5 um; Agilent) with 10 mM ammonium acetate (A) and 100% acetonitrile (B) at a flow rate of 0.2 ml/min (gradient profile: 1% B isocratic hold from 0 to 5 min; 1–60% B gradient from 5 to 35 min; 60% B to 98% B from 35 to 36 min; isocratic hold at 98% B from 36 to 46 min). HPLC effluents were analyzed on an Agilent 6510 Q-TOF mass spectrometer. The most abundant 16-mer oligonucleotide ion observed was the −4 charged species.

Modeling etheno lesions in the AlkB active site

Coordinates for the N^2,3-εG, 1,N^2-εG and εC lesion bases were obtained from existed structural data (30,51,52). Models of AlkB with N^2,3-εG and 1,N^2-εG lesions were generated from the reported structure of AlkB with an εA lesion (PDB ID: 3O1P) (53) by replacing the εA base with the corresponding εG bases and subjecting the models to energy minimization in CNS (54,55). Similarly, the model of AlkB with an εC lesion was generated from the reported structure of AlkB with a 3-methylcytosine lesion (PDB ID: 3O1M) (53). The structure of the active iron(IV)-oxo intermediate was generated by mapping recently reported QM/MM coordinates of the intermediate (the 'reactant complex' in (56)) onto the structures of AlkB with lesion bases, with an iron to oxygen distance of 1.62 Å. This particular structure was chosen because the iron(IV)-oxo intermediate initiates catalysis in AlkB (56).

RESULTS

Investigating site-specific, lesion-induced mutagenesis *in vivo* using next-generation sequencing

We combined our traditional site-specific mutagenesis approach with next-generation sequencing to evaluate the biological consequences of multiple lesions under multiple replication and repair conditions (Figure 1B). Next-generation sequencing provided the power to analyze samples from more than 90 experiments in parallel. We constructed single-stranded M13 vectors, each containing a site-specific, structurally-defined lesion and a triplet oligonucleotide barcode sequence unique to each lesion. The barcoded control and lesion-containing vectors were mixed in a known ratio and electroporated into isogenic cell strains, each with a specific repair and replication status. To investigate the role of AlkB, DinB and the SOS response, we used a wild type *E. coli* strain (HK81), *alkB*$^-$ strain (HK82), *alkB*$^-$/*dinB*$^-$ strain (HK84) as well as SOS-induced HK82 and HK84 strains (HK82 SOS$^+$ and HK84 SOS$^+$). The analyses on DinB and the SOS response were performed in cells with an *alkB*$^-$ background to eliminate any potential contribution from AlkB repair. Experiments in each of these five bacterial conditions were done in triplicate to monitor biological variability.

After *in vivo* replication, M13 progeny DNA were isolated and a 1 kb region containing the lesion barcode and the lesion site was PCR amplified. One major difference between our approach and similar methods reported previously (41,42) was the generation of sequencing libraries. Illumina libraries were prepared by using modified transposases, which randomly fragmented the PCR products to generate high complexity libraries. This approach circumvents any potential complication with the Illumina sequencers when handling libraries with a high degree of sequence homology, such as unprocessed PCR amplicons. During the library generation, a second set of barcodes was introduced to designate the respective five bacterial conditions and the three biological replicates. Equal amounts of fragmented DNA from each bacterial condition and biological replicate were pooled and sequenced on an Illumina MiSeq sequencer.

The sequencing data were sorted based on lesion and bacterial barcodes. Mutations, insertions and deletions of each lesion were obtained by mapping sequencing reads against the expected sequence. The replicative bypass efficiency, a metric of lesion genotoxicity, was calculated based on the change in number of sequencing reads from the lesion-containing genome relative to that from the lesion-free control.

Achieving high coverage at the lesion site is crucial in generating reliable genotoxicity and mutagenicity measurements. We obtained on average ∼7300-fold coverage at the lesion site per experiment that analyzed one lesion in a single cellular environment, which provided a sufficiently large sample size to generate a statistically robust result. For lesions displaying potent genotoxicity, the coverage dropped due to fewer progeny being produced; however, we still achieved greater than 200-fold coverage. By looking at the mutation rate at the conserved bases surrounding the lesion site, we estimated the error rate of this method is ∼0.5% (Supplementary Table S1). A higher error rate (1–2%) for the lesion-free control at the lesion site was observed, which was likely due to the presence of a low level of impurities in the starting oligonucleotide or in the barcode oligonucleotide.

DinB plays a significant role in the bypass of both εG lesions

Bypass efficiency is a measurement of how well polymerases can replicate DNA past a lesion present in the template strand. In wild type cells, both εG lesions had low bypass efficiencies, indicating they significantly hindered polymerase activity (Figure 2A). The bypass efficiency of 2′-F-N^2,3-εG was 21% relative to the lesion-free control and the bypass efficiency of 1,N^2-εG was only 4.4%. To ensure that the presence of a 2′-fluoro group did not hinder replication, we also investigated the bypass efficiency of 2′-FG, which was essentially the same as the lesion-free control (Figure 2A).

Inducing the SOS response increases the level of translesion polymerases in cells. Thus, it was not surprising that lesion toxicities were partially alleviated when the SOS response was induced (Figure 2B). In *alkB*$^-$/*dinB*$^+$ cells (thus eliminating any potential contribution from AlkB repair), the bypass efficiency of 2′-F-N^2,3-εG increased three-fold from 26 to 76% and that of 1,N^2-εG increased seven-fold from 1.8 to 13% upon SOS induction (Figure 2B, left). When DinB was absent, however, a more modest alleviation of toxicity was observed when SOS response was induced (Figure 2B, right). For example, in *alkB*$^-$/*dinB*$^-$ cells, the bypass efficiency improved only 2-fold from 14 to 29% for 2′-F-N^2,3-εG and 2.5-fold from 1.8 to 4.7% for 1,N^2-εG after SOS induction. Collectively, the data suggest that DinB plays a significant role in the bypass of both εG lesions, while other translesion polymerases (Pol II and Pol V) may play a secondary role.

2′-F-N^2,3-εG produces mainly G to A transitions

Following replication of a lesion-containing genome in cells, examining the resulting sequence of M13 progeny at and near the site originally containing the lesion revealed the mutagenic properties of each lesion. In each repair and

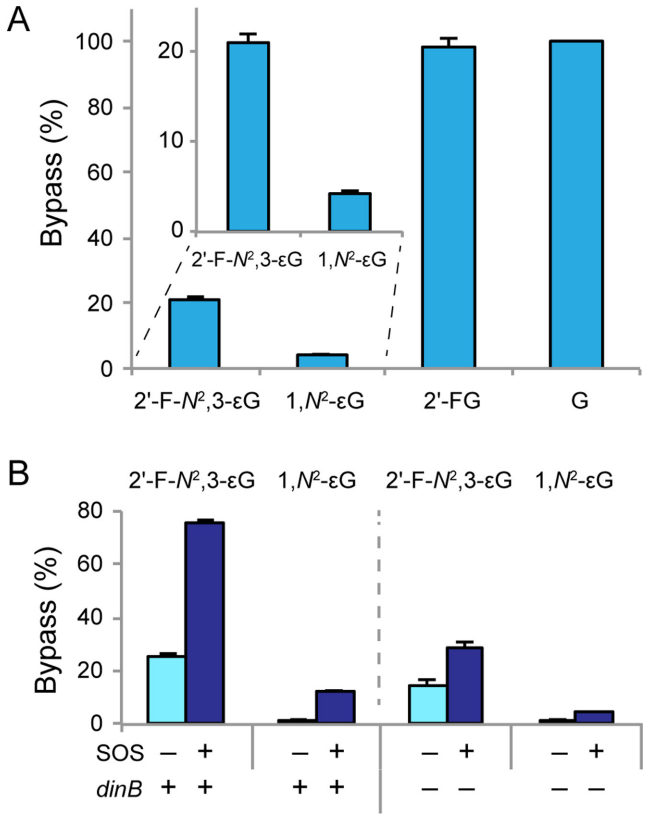


Figure 2. Replicative bypass efficiencies of the εG lesions. Lower bypass efficiency indicates higher lesion genotoxicity. (**A**) Bypass efficiencies of 2′-F-N^2,3-εG and 1,N^2-εG lesions, as well as 2′-FG and G controls in wild type (HK81) cells, with error bars representing one standard deviation (SD) ($N = 3$). Inset shows zoomed-in details of the 2′-F-N^2,3-εG and 1,N^2-εG results. (**B**) Bypass efficiencies of the εG lesions with or without SOS induction and in the presence or absence of DinB. The results shown in this figure (also tabulated in Supplementary Table S2) were obtained in an *alkB*$^-$ background. Error bars represent one SD ($N = 3$).

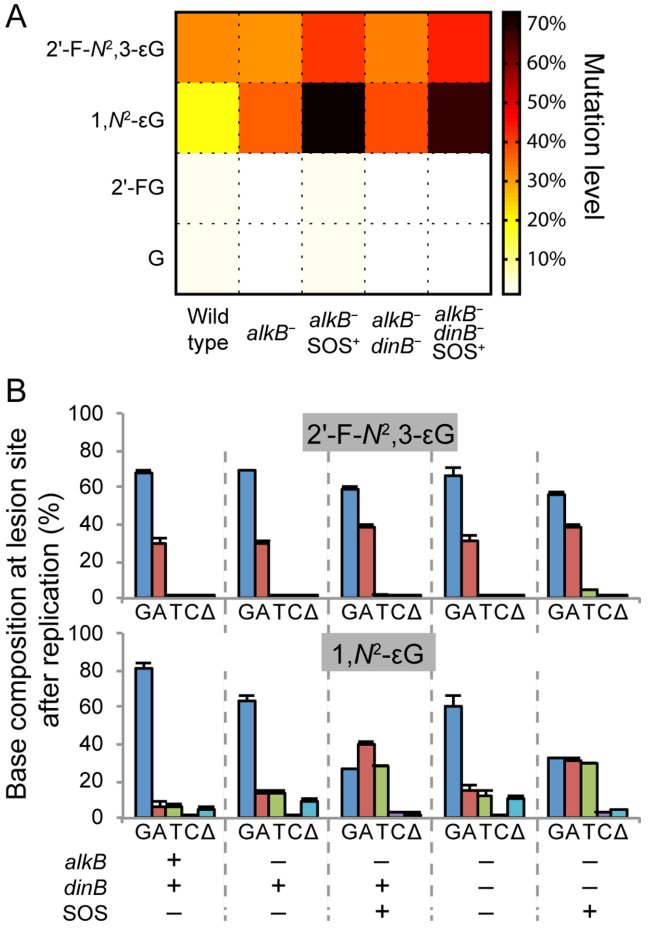


Figure 3. Lesion mutation frequencies. (**A**) A heat map representation of the mutation frequencies of 2′-F-N^2,3-εG, 1,N^2-εG, 2′-FG and G in all five repair/replication backgrounds investigated. (**B**) Mutation frequency and specificity of 2′-F-N^2,3-εG and 1,N^2-εG under the five bacterial conditions investigated, with error bars representing one SD ($N = 3$). G, A, T and C indicate the possible bases present at site N (Figure 1) in progeny genomes (i.e. the base present at the lesion site after *in vivo* replication), and Δ denotes the occurrence of deletions at the lesion site after replication. Results presented in this figure are also tabulated in Supplementary Table S4.

replication condition studied both εG lesions produced high levels of mutations after *in vivo* replication (Figure 3A). As expected, the controls 2′-FG and G were not mutagenic.

Looking at the specific types of mutations induced by the lesions, we found that 2′-F-N^2,3-εG produced almost exclusively G to A transitions in all five bacterial cellular environments studied (Figure 3B, top). In the non-SOS-induced cells, 2′-F-N^2,3-εG produced ∼30% G to A transitions with no significant level of other mutations. In both *alkB*$^-$ and *alkB*$^-$/*dinB*$^-$ cells, SOS induction increased the G to A mutation level to ∼38% and generated an additional low level (2–4%) of G to T transversions.

1,N^2-εG induces all possible base substitutions as well as frameshifts

While 2′-F-N^2,3-εG gave almost exclusively G to A transitions, 1,N^2-εG induced a wide range of mutations after *in vivo* replication (Figure 3B, bottom). In wild type cells, there were 6.4% G to A transitions, 6.0% G to T and 1.7% G to C transversions. Additionally, 4.9% of the isolated progeny had deletions at the lesion site, implicating the 1,N^2-εG lesion as a frameshift inducer. The mutation frequencies increased to 13% G to A, 13% G to T, 1.2% G to C and 8.7% deletion mutations in *alkB*$^-$ cells, suggesting that AlkB participates in the cellular defense against this lesion. In contrast to 2′-F-N^2,3-εG, inducing the SOS response greatly increased the frequency of substitution mutations for 1,N^2-εG; specifically, SOS induction in *alkB*$^-$/*dinB*$^+$ cells increased G to A mutations from 13 to 40%, G to T from 13 to 28% and G to C from 1.2 to 3.1%. Interestingly, deletion mutations decreased four-fold from 8.7 to 2.0% after SOS induction. In *alkB*$^-$/*dinB*$^-$ cells, SOS induction also substantially increased the frequency of substitutions but the increase was slightly smaller (from 15 to 31% of G to A, 12 to 30% of G to T and 1.6 to 2.9% of G to C) compared to that in the *dinB*$^+$ cells.

The level of deletions generated by 1,N^2-εG was the highest in *alkB*$^-$/SOS$^-$ cells, intermediate in wild type cells as well as *alkB*$^-$/*dinB*$^-$/SOS$^+$ cells, and the lowest in *dinB*$^+$/SOS$^+$ cells (Figure 4). The position of the deletions,

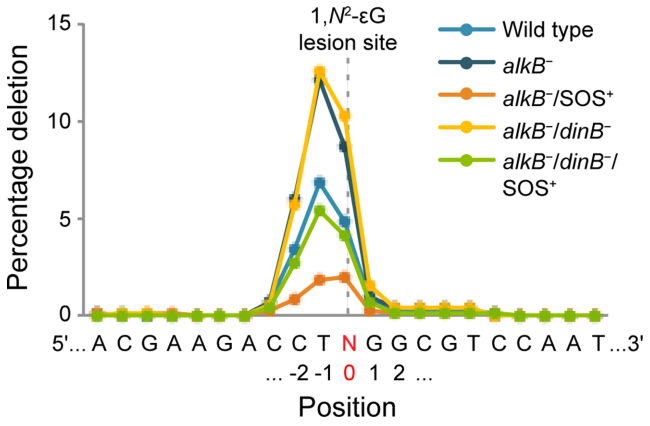


Figure 4. The frequency of deletions induced by 1,N^2-εG at the lesion site and adjacent positions. Each line represents the averaged results from the three independent biological replicates of a specific repair and replication background. N, highlighted in red, denotes the lesion site (position 0).

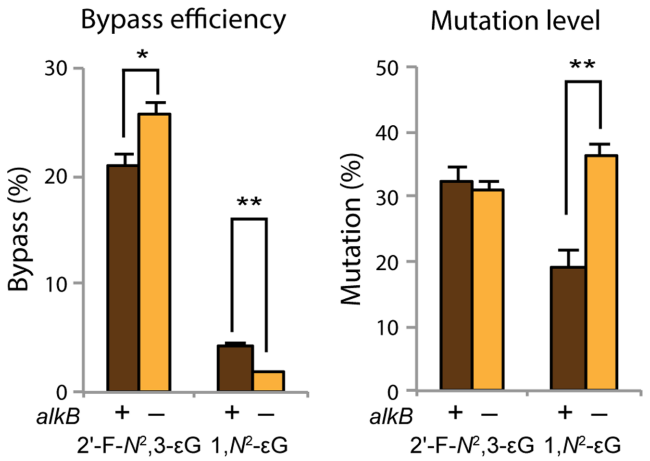


Figure 5. Replicative bypass efficiencies and mutation levels of 2′-F-N^2,3-εG and 1,N^2-εG in the presence and absence of AlkB. Error bars represent one SD ($N = 3$). Statistical significance was determined by a two-tailed, heteroscedastic, Student's t-test (*$P < 0.05$, **$P < 0.002$).

mostly 5′ to the lesion site, indicated that they occurred after the polymerase bypassed the 1,N^2-εG lesion. Most deletions were −2 frameshifts, deleting the two bases at either the −2 and −1 positions or at the −1 and 0 positions, with position 0 indicating the lesion site (Table 1). One-base deletions were also observed at a lower frequency (Table 1). The spectra of deletions were similar in all replication / repair environments examined.

2′-F-N^2,3-εG is not repaired by AlkB

To determine whether AlkB plays a role in the repair of the εG adducts, we compared lesion mutagenicity and toxicity in the presence and absence of AlkB. In AlkB-deficient cells, 1,N^2-εG was a stronger replication block and more mutagenic than in wild type cells; however, the same was not observed for 2′-F-N^2,3-εG (Figure 5), suggesting that AlkB repair provided some protection against the toxic and mutagenic effects of 1,N^2-εG, but not 2′-F-N^2,3-εG. This hypothesis was further evaluated using *in vitro* DNA repair reactions, followed by MS analyses. Repair intermediates and products (restored, unmodified G bases) were observed when oligonucleotides containing 1,N^2-εG or its 2′-fluoro analog were incubated with AlkB, Fe(II), ascorbate and α-ketoglutarate in a HEPES buffer (Figure 6A). The proposed repair mechanism as well as the theoretical m/z values for putative intermediates and final products are shown in Figure 6B. By contrast, no evidence of repair was observed for 2′-F-N^2,3-εG following a 1 h incubation with AlkB. The incubation with 2′-F-1,N^2-εG served as a control to ensure that the outcomes of the AlkB repair reactions were due to the lesion structures and not due to the presence of the 2′-fluoro group. The presence of repair intermediates and products for 2′-F-1,N^2-εG at comparable levels with those observed in the AlkB reaction of 1,N^2-εG suggests that the 2′-fluoro modification on the deoxyribose is tolerated by AlkB. These *in vitro* data, together with our *in vivo* observations detailed earlier establish that 1,N^2-εG is a substrate for AlkB repair, while 2′-F-N^2,3-εG and, by extension, the native lesion N^2,3-εG are not.

Comparison with the results from the traditional site-specific mutagenesis assays

In order to validate the sequencing approach of measuring *in vivo* lesion genotoxicity and mutagenicity, the two εG lesions were also analyzed with our traditional assays (38). The results obtained by the sequencing and the traditional methods were consistent with each other (Supplementary Tables S2, S4, S5 and S6 and Figures S4 and S5), engendering confidence in the new sequencing approach. Thus, our sequencing methodology is capable of generating measurements that are consistent with those from the traditional approach but at a much higher throughput. The measurements are also highly precise, as evidenced by the small standard deviations among the three independent biological replicates within each experiment.

DISCUSSION

In this study, we utilized a multiplex sequencing approach to measure quantitatively the *in vivo* mutagenicity and genotoxicity of the two prevalent εG DNA lesions. Next-generation sequencing not only provided the capability to multiplex, generating data at an unprecedented throughput, but also allowed us to identify and quantify easily the different types of lesion-induced insertion and deletion mutations. By requiring perfect matches only at the 13-mer sequence covering the trinucleotide lesion barcode, our algorithm allows sequence flexibility in the downstream sequences that contain the lesion site. Any insertion or deletion induced by the lesion can therefore be detected when mapping the reads against the expected sequence. This was demonstrated in the results of 1,N^2-εG, which is a known frameshift inducer. In addition to substitution mutations, we observed a significant level of −1 and −2 deletions at or 5′ to the lesion site (Table 1). Although a different sequence context and experimental design were used in a previous study (57), many of our observed substitution and frameshift products can be explained by the mechanisms proposed in that study. Hence, our observations on the

Table 1. The abundance of different deletion types induced by 1,N^2-εG in all five repair and replication backgrounds and the resulting sequence near the lesion site

Deletion type	Resulting sequence	Deletion abundance				
		Wild type	*alkB*$^-$	*alkB*$^-$/SOS$^+$	*alkB*$^-$/*dinB*$^-$	*alkB*$^-$/*dinB*$^-$/SOS$^+$
−1C @ −2	AC−TNGGCG	0.48 ± 0.16%	0.31 ± 0.40%	0.26 ± 0.07%	0.12 ± 0.11%	0.26 ± 0.23%
−2CT @ −2	AC−−NGGCG	3.00 ± 0.59%	5.49 ± 1.93%	0.73 ± 0.20%	5.09 ± 1.44%	2.46 ± 0.84%
−2TN @ −1	ACC−−GGCG	3.21 ± 0.90%	5.97 ± 1.53%	0.89 ± 0.22%	6.70 ± 2.49%	2.73 ± 0.37%
−1N @ 0	ACCT−GGCG	0.99 ± 0.71%	1.74 ± 0.51%	0.88 ± 0.23%	1.98 ± 0.96%	0.87 ± 0.24%
−2NG @ 0	ACCT−−GCG	0.32 ± 0.28%	0.61 ± 0.51%	0.15 ± 0.09%	0.97 ± 0.47%	0.37 ± 0.05%

The expected full-length sequence of the nine bases surrounding the site of interest is 5′-ACCTNGGCG-3′. Position 0 indicates the lesion site (site N in Figure 4). The "−2CT @ −2" deletion, for example, represents a two base deletion deleting bases C at position −2 and T at position −1. Results are presented as average ± one SD of the three biological replicates.

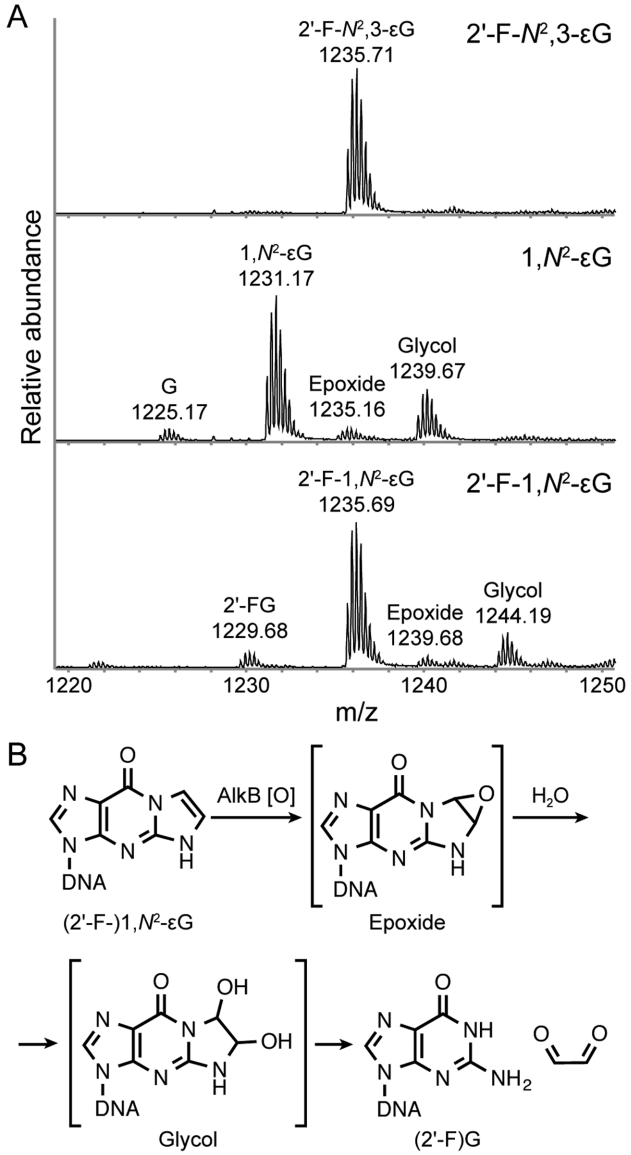
propensity of 1,N^2-εG to induce mainly G to A and G to T mutations, as well as −1 and −2 frameshift deletions, corroborate and expand previous investigations (57–59).

Investigating the mis-coding property of N^2,3-εG has been challenging since its instability has precluded its incorporation into oligonucleotides via the conventional phosphoramidite method. Cheng *et al.* described an alternative strategy that utilized DNA polymerases to incorporate the N^2,3-εG triphosphate into M13 vectors (29). However, in order to establish the mutagenicity of N^2,3-εG in *E. coli*, they had to apply a correction factor to account for incorporation inefficiency and phenotypic penetrance; thus, adjusting the initially measured mutation frequency of 0.5% to an estimated 13% (29). The accuracy of this estimation was further complicated by the potential loss of N^2,3-εG due to spontaneous depurination or exonuclease removal from the polymerase during oligonucleotide synthesis (29). In the present study, we utilized a chemical manipulation to stabilize the labile glycoside bond of N^2,3-εG. The 2′-fluoro substituted, stabilized analog, 2′-F-N^2,3-εG, can be incorporated site-specifically via the phosphoramidite method, and thus we were able to measure directly the lesion mutagenicity *in vivo*. Moreover, the current method allowed us to also measure the lesion genotoxicity *in vivo*, which has not been quantified previously.

We first established that 2′-F-N^2,3-εG is a suitable analog for investigating the biological consequences of N^2,3-εG by confirming that the 2′-fluoro modification on deoxyribose is neither genotoxic nor mutagenic *in vivo*, as evidenced by our data on 2′-FG. Our findings that N^2,3-εG potently and almost exclusively induces G to A transitions after *in vivo* replication are consistent with previous observations (29–31,60). However, the mutation frequency we observed in wild type cells was approximately 30%, significantly higher than the frequency previously estimated by others (29). Given that N^2,3-εG is produced in DNA after VC exposure, the ability of N^2,3-εG to induce G to A transitions suggests that this lesion may play a functional role in generating the G:C to A:T mutations found in the K-*ras* gene of VC-induced angiosarcoma of the liver as well as in cells treated with VC and its metabolites. The molecular basis of the miscoding property of N^2,3-εG can be attributed to its ability to pair with a T. In previous X-ray crystallography studies, we have shown that N^2,3-εG can pair with a T as a sheared base pair in *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) (30) and can adopt the *syn* conformation to pair with a T in human DNA polymerase ι (31).

Although N^2,3-εG is labile in monomeric form, it is stable in duplex DNA (28). In fact, previous studies have found that the half-life of N^2,3-εG in liver and lung of rodents is ∼150 days whereas that of εA is only about one day (61,62). The level of N^2,3-εG is also higher than that of other etheno lesions, derived either endogenously or after VC exposure (5,17,20). These observations suggest that N^2,3-εG may be poorly repaired. Although both εG lesions have been shown to be poorly repaired by the base excision repair pathway (whereas εA and εC are efficiently removed by glycosylases) (63), our AlkB repair results offer an additional insight into the persistence of N^2,3-εG. In concert with our previous study (32), our data collectively indicate that AlkB cannot repair N^2,3-εG whereas it can repair the other three etheno lesions. We ruled out the possibility that the 2′-fluoro group may inhibit AlkB repair by showing that AlkB was comparably proficient at repairing both 1,N^2-εG and 2′-F-1,N^2-εG *in vitro* (Figure 6). Under the same condition, 2′-F-N^2,3-εG was left unrepaired (Figure 6), suggesting that the inability of AlkB to repair N^2,3-εG is likely attributable to the structure of the lesion itself. As an aside, our *in vivo* results comparing 1,N^2-εG and 2′-F-1,N^2-εG in *alkB*$^+$ and *alkB*$^-$ cells suggest that the 2′-fluoro group may interfere with AlkB repair when the enzyme is present at low concentration, as might occur in non-adapted cells (Supplementary Figures S1 and S2). By contrast, the *in vitro* repair experiments were carried out using AlkB at a higher concentration, more akin to an adaptive response-induced state.

Modeling etheno lesions in the AlkB active site offers a structural rationale for explaining why AlkB can repair all etheno lesions except N^2,3-εG (Figure 7). Steric hindrance is unlikely to be the cause, as N^2,3-εG can easily fit into the active site based on the model. The two etheno carbon atoms of N^2,3-εG, however, are positioned 3.8 and 4.2 Å away from the iron(IV)-oxo intermediate, whereas all other etheno lesions (εA, εC and 1,N^2-εG) have at least one carbon atom within 3.3 Å of the oxo group (Figure 7). Similarly, modeling the iron(IV)-oxo intermediate into the crystal structure of AlkB with a 3-methylcytosine lesion, also a known substrate of AlkB, places the 3-methyl group 3.3 Å away from the oxo group (Supplementary Figure S6) (53). Therefore, we hypothesize that N^2,3-εG cannot be repaired because its exocyclic etheno carbons are too dis-



Theoretical m/z	1,N^2-εG	Epoxide	Glycol	G
With 2'-H	1231.21	1235.21	1239.71	1225.21
With 2'-F	1235.70	1239.71	1244.21	1229.71

Figure 6. Susceptibility of εG lesions to AlkB repair *in vitro*. (**A**) Mass spectra of 16-mer oligonucleotides containing either a 2′-F-N^2,3-εG, 1,N^2-εG or 2′-F-1,N^2-εG following 1 h incubation with AlkB. Data represent the −4 charge envelopes and the observed monoisotopic peak value is labeled above each peak envelope. (**B**) The proposed mechanism of AlkB repair on (2′-F-)1,N^2-εG. Theoretical *m/z* values of the −4 charged ions of the 16-mer oligonucleotides starting materials, putative reaction intermediates and repair products are listed for both the 2′-deoxy (2′-H) and 2′-fluoro (2′-F) versions of the bases. A recent theoretical study proposed that repair of εA by AlkB may be mediated by a zwitterionic species, and that some of the species observed by the MS analysis could be byproducts rather than repair intermediates (56). However, the ability of AlkB to repair the εG lesions is not affected by this hypothesis.

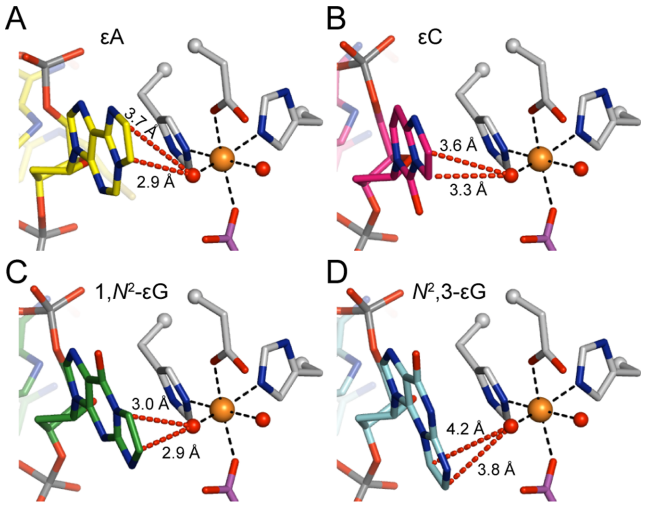


Figure 7. Models of the four etheno lesions in the active site of AlkB. (**A**) AlkB active site with an εA lesion (yellow carbons, PDB ID: 3O1P), a known good substrate for AlkB, and with the iron(IV)-oxo intermediate modeled (see below). (**B**–**D**) Models of AlkB with (**B**) εC (pink carbons), (**C**) 1,N^2-εG (green carbons), and (**D**) N^2,3-εG (cyan carbons) lesions in the active site. Models of 1,N^2-εG and N^2,3-εG are based on the crystal structure of εA in AlkB (PDB ID: 3O1P) and the model of εC is based on the crystal structure of 3-methylcytosine (3mC) in AlkB (PDB ID: 3O1M, Supplementary Figure S6). In all panels, selected AlkB amino acid residues are shown in grey, iron-bound succinate in purple, the iron ion as an orange sphere, and iron-bound oxygens (or water molecules) as red spheres. The iron(IV)-oxo intermediate is modeled based on a recent study, with an iron-oxygen distance of 1.62 Å (56). Distances from the iron(IV)-oxo oxygen atom to the exocyclic etheno carbon atoms are shown as red dashed lines.

tant from the iron(IV)-oxo center to be oxidized by the enzyme. In contrast, the other three etheno lesions can be repaired by AlkB because their etheno bridges are closer to the iron(IV)-oxo center.

Accumulation of mutations in the genome is believed to be a driver of carcinogenesis (64,65). Understanding the mutagenicity and repair of individual DNA lesions can help to bridge the gap between DNA damage and the carcinogenic process (66) and, on a practical level, identify biomarkers that can be used to assess risk (67). For a biomarker to be useful, one critical criterion is that it has to be quantifiable non-invasively. DNA adducts that are exclusively repaired by the direct reversal pathway, such as by AlkB, are destroyed during the repair process and thus, are not detectable without invasively extracting DNA from tissue samples. On the other hand, excision repair removes the intact lesions from the genome and the excised lesions are often excreted in urine. The lack of repair of N^2,3-εG by the AlkB pathway suggests that removal of this adduct from the genome, albeit slow removal as evidenced by its long half-life, is probably by excision repair or spontaneous hydrolysis. Indeed, one study has demonstrated the detection of the N^2,3-εG base in urine samples with high sensitivity and specificity (68). The excretion of this powerfully mutagenic lesion in urine affords an opportunity to measure non-invasively the level of this potential biomarker for risk assessment of diseases that are associated with inflammation or the exposure to certain industrial chemicals such as VC.

In addition to DNA repair pathways, translesion synthesis represents an important mechanism by which cells tolerate DNA damage, even though this process is often mutagenic. Here, we have shown that induction of the SOS response alleviated some of the toxicity by the εG lesions, and most of the lesion bypass could be attributed to the activity of the DinB translesion polymerase in *E. coli*. The bypass of 1,N^2-εG is a highly mutagenic event, which is presumably due to the two exocyclic carbons blocking the canonical Watson-Crick base-pairing region. In the *dinB*$^+$ cells, inducing the SOS response doubled the mutation level from 36 to 73%. In contrast, the bypass of N^2,3-εG is more accurate. SOS induction increased the bypass of N^2,3-εG from 26 to 76%, while the mutation frequency was only increased from 31 to 41%.

Despite the bypass of N^2,3-εG being more accurate compared with the 1,N^2-εG bypass, the end result of the increased level of bypass is an effective increase in the population of progeny harboring a G to A mutation at the lesion site. Therefore, translesion synthesis may allow for the survival and proliferation of the cells carrying N^2,3-εG by preventing replication arrest or fork collapse, outcomes that often lead to apoptosis, but at the cost of an increased G to A mutation load. Additionally, several studies have found overexpression of translesion polymerases in various cancers (69–72). This overexpression of lower-fidelity polymerases may not only increase the level of spontaneous mutations, but also enhance the survival of cancer cells exposed to additional DNA damaging agents or chemotherapeutics. Thus, the overexpression of translesion polymerases may provide for a hyper-mutagenic phenotype and give pre-cancerous and cancerous cells a selective advantage over normal cells, likely exacerbating disease progression.

Even though our study was conducted in *E. coli* cells, the results are relevant for human pathophysiology in part because both AlkB and DinB proteins have human counterparts. AlkB has nine human homologs; two of them, ALKBH2 and ALKBH3 have been shown to have similar substrate specificity as AlkB (73). DNA polymerase κ (POLK) is the human ortholog of DinB and shares the ability to bypass many N^2-G adducts (74). Thus, our study in *E. coli* cells could be an informative model system for understanding the biological consequences of the εG lesions in human cells.

Although to date there is no direct evidence that N^2,3-εG is the lesion directly responsible for the mutations seen in the K-*ras* gene of VC-induced tumors, our present work, together with other studies, presents three key pieces of evidence that strongly support a functional role of the N^2,3-εG lesion in carcinogenesis. First, this lesion is the most abundant etheno lesion, formed both endogenously and after VC exposure. Second, it is persistent inside cells, which can be partially explained by the fact that the lesion is not a good substrate for AlkB repair. Third, our *in vivo* results indicate that N^2,3-εG is highly mutagenic, producing the same type of G to A transition mutations as seen in VC-induced tumors. Additionally, as N^2,3-εG is also produced under conditions of inflammation and oxidative stress, our results have broader implications beyond VC-induced carcinogenesis. Since N^2,3-εG can be detected and quantified in urine samples, it could serve as a functional biomarker for evaluating the mutational burden in individuals, which could correlate with the risk of VC-induced or inflammation-driven cancers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Koli Taghizadeh for assistance on HPLC and MS analyses, the Massachusetts Institute of Technology (MIT) Center for Environmental Health Sciences for providing mass spectrometry and imaging facilities and the MIT BioMicro Center for conducting the next-generation sequencing experiments. C.L.D. is a Howard Hughes Medical Institute Investigator.

FUNDING

This work was supported by the National Institutes of Health [P30 ES002109, T32 ES007020, R37 CA080024 and P01 CA026731 to J.M.E., R01 GM069857 to C.L.D., R01 ES010546 to F.P.G., P01 ES05355 and P01 CA160032 to C.J.R., P30 ES000267 and P30 CA068485 to F.P.G. and C.J.R.]. Funding for open access charge: National Institutes of Health [R37 CA080024].
Conflict of interest statement. None declared.

REFERENCES

1. Breslow,L., Hoaglin,L., Rasmussen,G. and Abrams,H.K. (1954) Occupations and cigarette smoking as factors in lung cancer. *Am. J. Public Health Nations Health*, **44**, 171–181.
2. Abdulla,F.R., Feldman,S.R., Williford,P.M., Krowchuk,D. and Kaur,M. (2005) Tanning and skin cancer. *Pediatr. Dermatol.*, **22**, 501–512.
3. Wogan,G.N. (1975) Dietary factors and special epidemiological situations of liver cancer in Thailand and Africa. *Cancer Res.*, **35**, 3499–3502.
4. Triantafillidis,J.K., Nasioulas,G. and Kosmidis,P.A. (2009) Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. *Anticancer Res.*, **29**, 2727–2737.
5. International Agency for Research on Cancer. (2008) *IARC Monogr. Eval. Carcinog. Risks Hum*. International Agency for Research on Cancer, Lyon, Vol. **97**, pp. 311–443.
6. Marion,M.J., Froment,O. and Trepo,C. (1991) Activation of Ki-*ras* gene by point mutation in human liver angiosarcoma associated with vinyl chloride exposure. *Mol. Carcinog.*, **4**, 450–454.
7. Weihrauch,M., Bader,M., Lehnert,G., Koch,B., Wittekind,C., Wrbitzky,R. and Tannapfel,A. (2002) Mutation analysis of K-*ras*-2 in liver angiosarcoma and adjacent nonneoplastic liver tissue from patients occupationally exposed to vinyl chloride. *Environ. Mol. Mutagen.*, **40**, 36–40.
8. Barbacid,M. (1987) *ras* genes. *Annu. Rev. Biochem.*, **56**, 779–827.
9. Kozma,S.C., Bogaard,M.E., Buser,K., Saurer,S.M., Bos,J.L., Groner,B. and Hynes,N.E. (1987) The human c-Kirsten *ras* gene is activated by a novel mutation in codon 13 in the breast carcinoma cell line MDA-MB231. *Nucleic Acids Res.*, **15**, 5963–5971.
10. Liu,E., Hjelle,B., Morgan,R., Hecht,F. and Bishop,J.M. (1987) Mutations of the Kirsten-*ras* proto-oncogene in human preleukaemia. *Nature*, **330**, 186–188.
11. Vogelstein,B., Fearon,E.R., Hamilton,S.R., Kern,S.E., Preisinger,A.C., Leppert,M., Nakamura,Y., White,R., Smits,A.M. and Bos,J.L. (1988) Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.*, **319**, 525–532.

12. Nagata,Y., Abe,M., Kobayashi,K., Yoshida,K., Ishibashi,T., Naoe,T., Nakayama,E. and Shiku,H. (1990) Glycine to aspartic acid mutations at codon 13 of the c-Ki-*ras* gene in human gastrointestinal cancers. *Cancer Res.*, **50**, 480–482.
13. Kraus,M.C., Seelig,M.H., Linnemann,U. and Berger,M.R. (2006) The balanced induction of K-*ras* codon 12 and 13 mutations in mucosa differs from their ratio in neoplastic tissues. *Int. J. Oncol.*, **29**, 957–964.
14. McCann,J., Simmon,V., Streitwieser,D. and Ames,B.N. (1975) Mutagenicity of chloroacetaldehyde, a possible metabolic product of 1, 2-dichloroethane (ethylene dichloride), chloroethanol (ethylene chlorohydrin), vinyl chloride, and cyclophosphamide. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 3190–3193.
15. Barbin,A., Besson,F., Perrard,M.H., Bereziat,J.C., Kaldor,J., Michel,G. and Bartsch,H. (1985) Induction of specific base-pair substitutions in *E. coli trpA* mutants by chloroethylene oxide, a carcinogenic vinyl chloride metabolite. *Mutat. Res.*, **152**, 147–156.
16. Dogliotti,E. (2006) Molecular mechanisms of carcinogenesis by vinyl chloride. *Ann. Ist. Super. Sanita*, **42**, 163–169.
17. Swenberg,J.A., Lu,K., Moeller,B.C., Gao,L., Upton,P.B., Nakamura,J. and Starr,T.B. (2011) Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol. Sci.*, **120**(Suppl. 1), S130–S145.
18. el Ghissassi,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.
19. Ham,A.J., Ranasinghe,A., Koc,H. and Swenberg,J.A. (2000) 4-Hydroxy-2-nonenal and ethyl linoleate form N^2, 3-ethenoguanine under peroxidizing conditions. *Chem. Res. Toxicol.*, **13**, 1243–1250.
20. Chung,F.L., Chen,H.J. and Nath,R.G. (1996) Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts. *Carcinogenesis*, **17**, 2105–2111.
21. Marnett,L.J. (2000) Oxyradicals and DNA damage. *Carcinogenesis*, **21**, 361–370.
22. Lee,S.H., Arora,J.A., Oe,T. and Blair,I.A. (2005) 4-Hydroperoxy-2-nonenal-induced formation of 1, N^2-etheno-2′-deoxyguanosine adducts. *Chem. Res. Toxicol.*, **18**, 780–786.
23. Nair,J., Barbin,A., Guichard,Y. and Bartsch,H. (1995) 1,N^6-Ethenodeoxyadenosine and 3,N^4-ethenodeoxycytine in liver DNA from humans and untreated rodents detected by immunoaffinity/^{32}P-postlabeling. *Carcinogenesis*, **16**, 613–617.
24. Barbin,A., Ohgaki,H., Nakamura,J., Kurrer,M., Kleihues,P. and Swenberg,J.A. (2003) Endogenous deoxyribonucleic Acid (DNA) damage in human tissues: a comparison of ethenobases with aldehydic DNA lesions. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 1241–1247.
25. Morinello,E.J., Ham,A.J., Ranasinghe,A., Nakamura,J., Upton,P.B. and Swenberg,J.A. (2002) Molecular dosimetry and repair of N^2, 3-ethenoguanine in rats exposed to vinyl chloride. *Cancer Res.*, **62**, 5189–5195.
26. Nair,U., Bartsch,H. and Nair,J. (2007) Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. *Free Radic. Biol. Med.*, **43**, 1109–1120.
27. Shrivastav,N., Li,D. and Essigmann,J.M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, **31**, 59–70.
28. Kusmierek,J.T., Folkman,W. and Singer,B. (1989) Synthesis of N^2, 3-ethenodeoxyguanosine, N^2, 3-ethenodeoxyguanosine 5′-phosphate, and N^2, 3-ethenodeoxyguanosine 5′-triphosphate. Stability of the glycosyl bond in the monomer and in poly(dG, εdG-dC). *Chem. Res. Toxicol.*, **2**, 230–233.
29. Cheng,K.C., Preston,B.D., Cahill,D.S., Dosanjh,M.K., Singer,B. and Loeb,L.A. (1991) The vinyl chloride DNA derivative N^2,3-ethenoguanine produces G—-A transitions in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9974–9978.
30. Zhao,L., Christov,P.P., Kozekov,I.D., Pence,M.G., Pallan,P.S., Rizzo,C.J., Egli,M. and Guengerich,F.P. (2012) Replication of N^2, 3-ethenoguanine by DNA polymerases. *Angew. Chem. Int. Ed. Engl.*, **51**, 5466–5469.
31. Zhao,L., Pence,M.G., Christov,P.P., Wawrzak,Z., Choi,J.Y., Rizzo,C.J., Egli,M. and Guengerich,F.P. (2012) Basis of miscoding of the DNA adduct N^2,3-ethenoguanine by human Y-family DNA polymerases. *J. Biol. Chem.*, **287**, 35516–35526.
32. Delaney,J.C., Smeester,L., Wong,C., 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.
33. Jarosz,D.F., Beuning,P.J., Cohen,S.E. and Walker,G.C. (2007) Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol.*, **15**, 70–77.
34. Walsh,J.M., Hawver,L.A. and Beuning,P.J. (2011) *Escherichia coli* Y family DNA polymerases. *Front Biosci. (Landmark Ed.)*, **16**, 3164–3182.
35. Jarosz,D.F., Godoy,V.G., Delaney,J.C., Essigmann,J.M. and Walker,G.C. (2006) A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature*, **439**, 225–228.
36. Yuan,B., Cao,H., Jiang,Y., Hong,H. and Wang,Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2′-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 8679–8684.
37. Shrivastav,N., Fedeles,B.I., Li,D., Delaney,J.C., Frick,L.E., Foti,J.J., Walker,G.C. and Essigmann,J.M. (2014) A chemical genetics analysis of the roles of bypass polymerase DinB and DNA repair protein AlkB in processing N^2-alkylguanine lesions *in vivo*. *PLoS ONE*, **9**, e94716.
38. Delaney,J.C. and Essigmann,J.M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions *in vivo*. *Methods Enzymol.*, **408**, 1–15.
39. Mardis,E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, **9**, 387–402.
40. Metzker,M.L. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.*, **11**, 31–46.
41. Yuan,B., Wang,J., Cao,H., Sun,R. and Wang,Y. (2011) High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing. *Nucleic Acids Res.*, **39**, 5945–5954.
42. Xing,X.W., Liu,Y.L., Vargas,M., Wang,Y., Feng,Y.Q., Zhou,X. and Yuan,B.F. (2013) Mutagenic and cytotoxic properties of oxidation products of 5-methylcytosine revealed by next-generation sequencing. *PLoS ONE*, **8**, e72993.
43. Shanmugam,G., Goodenough,A.K., Kozekov,I.D., Guengerich,F.P., Rizzo,C.J. and Stone,M.P. (2007) Structure of the 1,N^2-etheno-2′-deoxyguanosine adduct in duplex DNA at pH 8.6. *Chem. Res. Toxicol.*, **20**, 1601–1611.
44. Goodenough,A.K., Kozekov,I.D., Zang,H., Choi,J.Y., Guengerich,F.P., Harris,T.M. and Rizzo,C.J. (2005) Site specific synthesis and polymerase bypass of oligonucleotides containing a 6-hydroxy-3, 5, 6, 7-tetrahydro-9*H*-imidazo[1, 2-*a*]purin-9-one base, an intermediate in the formation of 1,N^2-etheno-2′-deoxyguanosine. *Chem. Res. Toxicol.*, **18**, 1701–1714.
45. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 14051–14056.
46. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
47. Zhang,J., Kobert,K., Flouri,T. and Stamatakis,A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.
48. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357–359.
49. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
50. 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 *Escherichia coli* adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 755–760.
51. Shanmugam,G., Kozekov,I.D., Guengerich,F.P., Rizzo,C.J. and Stone,M.P. (2010) Structure of the 1, N^2-etheno-2′-deoxyguanosine lesion in the 3′-G(εdG)T-5′ sequence opposite a one-base deletion. *Biochemistry*, **49**, 2615–2626.
52. Cullinan,D., Johnson,F., Grollman,A.P., Eisenberg,M. and de los Santos,C. (1997) Solution structure of a DNA duplex containing the

exocyclic lesion 3,N^4-etheno-2′-deoxycytidine opposite 2′-deoxyguanosine. *Biochemistry*, **36**, 11933–11943.
53. Yi,C., Jia,G., Hou,G., Dai,Q., Zhang,W., Zheng,G., Jian,X., Yang,C.G., Cui,Q. and He,C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330–333.
54. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
55. Brunger,A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
56. Wang,B., Usharani,D., Li,C. and Shaik,S. (2014) Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. *J. Am. Chem. Soc.*, **136**, 13895–13901.
57. Zang,H., Goodenough,A.K., Choi,J.Y., Irimia,A., Loukachevitch,L.V., Kozekov,I.D., Angel,K.C., Rizzo,C.J., Egli,M. and Guengerich,F.P. (2005) DNA adduct bypass polymerization by Sulfolobus solfataricus DNA polymerase Dpo4: analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N^2-ethenoguanine. *J. Biol. Chem.*, **280**, 29750–29764.
58. Langouët,S., Muller,M. and Guengerich,F.P. (1997) Misincorporation of dNTPs opposite 1,N^2-ethenoguanine and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1, 2-*a*]purine in oligonucleotides by *Escherichia coli* polymerases I exo$^-$ and II exo$^-$, T7 polymerase exo$^-$, human immunodeficiency virus-1 reverse transcriptase, and rat polymerase β. *Biochemistry*, **36**, 6069–6079.
59. Langouët,S., Mican,A.N., Muller,M., Fink,S.P., Marnett,L.J., Muhle,S.A. and Guengerich,F.P. (1998) Misincorporation of nucleotides opposite five-membered exocyclic ring guanine derivatives by *Escherichia coli* polymerases *in vitro* and *in vivo*: 1,N^2-ethenoguanine, 5,6,7,9-tetrahydro-9-oxoimidazo[1, 2-*a*]purine, and 5,6,7,9-tetrahydro-7-hydroxy-9-oxoimidazo[1, 2-*a*]purine. *Biochemistry*, **37**, 5184–5193.
60. Singer,B., Spengler,S.J., Chavez,F. and Kusmierek,J.T. (1987) The vinyl chloride-derived nucleoside, N^2, 3-ethenoguanosine, is a highly efficient mutagen in transcription. *Carcinogenesis*, **8**, 745–747.
61. Mutlu,E., Collins,L.B., Stout,M.D., Upton,P.B., Daye,L.R., Winsett,D., Hatch,G., Evansky,P. and Swenberg,J.A. (2010) Development and application of an LC-MS/MS method for the detection of the vinyl chloride-induced DNA adduct N^2, 3-ethenoguanine in tissues of adult and weanling rats following exposure to [$^{13}C_2$]-VC. *Chem. Res. Toxicol.*, **23**, 1485–1491.
62. Ham,A.J., Engelward,B.P., Koc,H., Sangaiah,R., Meira,L.B., Samson,L.D. and Swenberg,J.A. (2004) New immunoaffinity-LC-MS/MS methodology reveals that *Aag* null mice are deficient in their ability to clear 1, N^6-etheno-deoxyadenosine DNA lesions from lung and liver *in vivo*. *DNA Repair (Amst.)*, **3**, 257–265.
63. Dosanjh,M.K., Chenna,A., Kim,E., Fraenkel-Conrat,H., Samson,L. and Singer,B. (1994) All four known cyclic adducts formed in DNA by the vinyl chloride metabolite chloroacetaldehyde are released by a human DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 1024–1028.
64. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
65. Vogelstein,B., Papadopoulos,N., Velculescu,V.E., Zhou,S., Diaz,L.A. Jr and Kinzler,K.W. (2013) Cancer genome landscapes. *Science*, **339**, 1546–1558.
66. Smela,M.E., Currier,S.S., Bailey,E.A. and Essigmann,J.M. (2001) The chemistry and biology of aflatoxin B_1: from mutational spectrometry to carcinogenesis. *Carcinogenesis*, **22**, 535–545.
67. Groopman,J.D., Johnson,D. and Kensler,T.W. (2005) Aflatoxin and hepatitis B virus biomarkers: a paradigm for complex environmental exposures and cancer risk. *Cancer Biomark.*, **1**, 5–14.
68. Gonzalez-Reche,L.M., Koch,H.M., Weiss,T., Muller,J., Drexler,H. and Angerer,J. (2002) Analysis of ethenoguanine adducts in human urine using high performance liquid chromatography-tandem mass spectrometry. *Toxicol. Lett.*, **134**, 71–77.
69. Albertella,M.R., Lau,A. and O'Connor,M.J. (2005) The overexpression of specialized DNA polymerases in cancer. *DNA Repair (Amst.)*, **4**, 583–593.
70. O-Wang,J., Kawamura,K., Tada,Y., Ohmori,H., Kimura,H., Sakiyama,S. and Tagawa,M. (2001) DNA polymerase κ, implicated in spontaneous and DNA damage-induced mutagenesis, is overexpressed in lung cancer. *Cancer Res.*, **61**, 5366–5369.
71. Yang,J., Chen,Z., Liu,Y., Hickey,R.J. and Malkas,L.H. (2004) Altered DNA polymerase ι expression in breast cancer cells leads to a reduction in DNA replication fidelity and a higher rate of mutagenesis. *Cancer Res.*, **64**, 5597–5607.
72. Wang,H., Wu,W., Wang,H.W., Wang,S., Chen,Y., Zhang,X., Yang,J., Zhao,S., Ding,H.F. and Lu,D. (2010) Analysis of specialized DNA polymerases expression in human gliomas: association with prognostic significance. *Neuro Oncol.*, **12**, 679–686.
73. Yi,C., Yang,C.G. and He,C. (2009) A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.*, **42**, 519–529.
74. Waters,L.S., Minesinger,B.K., Wiltrout,M.E., D'Souza,S., Woodruff,R.V. and Walker,G.C. (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.*, **73**, 134–154.