Supplementary data

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

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

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

* To whom correspondence should be addressed. Tel: +1-617-253-6227; Fax: +1-617-253-5445; Email: jessig@mit.edu

Present Address:
[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

Table of Contents

Table S1. Error rate of the conserved sequence surrounding the site of interest in the control genome in wild type cells.

Table S2. Lesion bypass efficiency of 2′-F-N^2,3-εG, 1,N^2-εG, 2′-F-1,N^2-εG, 2′-FG and G in all five repair and replication backgrounds investigated.

Table S3. Number of reads obtained for each lesion from next-generation sequencing.

Table S4. Mutation frequency and specificity of all lesions and controls in all five repair and replication backgrounds investigated.

Table S5. Lesion bypass efficiency of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional CRAB assay.

Table S6. Mutation frequency and specificity of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional REAP assay.

Figure S1. Lesion bypass efficiency of 2′-F-1,N^2-εG in all five repair and replication backgrounds investigated.

Figure S2. Mutation frequency and specificity of 2′-F-1,N^2-εG in all five repair and replication backgrounds investigated.

Figure S3. Mass spectra demonstrating the activity of AlkB *in vitro* on εG-containing 16-mers.

Figure S4. Lesion bypass efficiency of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional CRAB assay.

Figure S5. Mutation frequency and specificity of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional REAP assay.

Figure S6. AlkB active site with a 3-methylcytosine lesion and with the iron(IV)-oxo intermediate modeled.

Table S1. Error rate of the conserved sequence surrounding the site of interest (site N in Figure 1 and this table) in the control genome in wild type cells. The expected sequence of the 21 bases encompassing the site of interest is 5′-ACGAAGACCTNGGCGTCCAAT-3′ where N is the site of interest at position 6260. The average error rate from the three biological replicates at each of the 21 sites is presented along with the standard deviation.

Position	Expected base	Average error rate	SD
6250	A	0.14%	0.04%
6251	C	0.26%	0.03%
6252	G	0.15%	0.05%
6253	A	0.16%	0.06%
6254	A	0.39%	0.02%
6255	G	0.26%	0.04%
6256	A	0.56%	0.10%
6257	C	0.64%	0.10%
6258	C	0.33%	0.04%
6259	T	0.48%	0.05%
6260	N	2.43%	0.25%
6261	G	0.60%	0.10%
6262	G	0.42%	0.05%
6263	C	0.29%	0.08%
6264	G	0.42%	0.01%
6265	T	0.38%	0.04%
6266	C	0.42%	0.14%
6267	C	0.45%	0.16%
6268	A	0.11%	0.01%
6269	A	0.09%	0.01%
6270	T	0.06%	0.03%

Table S2. Lesion bypass efficiency of 2′-F-N^2,3-εG, 1,N^2-εG, 2′-F-1,N^2-εG, 2′-FG and G in all five repair and replication backgrounds investigated. Results are presented as average ± one standard deviation (SD) of the three biological replicates.

Repair/replication background	Lesions				
	2′-F-N^2,3-εG	1,N^2-εG	2′-F-1,N^2-εG	2′-FG	G
HK81 (Wild type, *alkB*$^+$ / *dinB*$^+$ / SOS$^-$)	21.0 ± 1.0%	4.4 ± 0.1%	4.6 ± 0.4%	98.0 ± 3.3%	100%
HK82 (*alkB*$^-$ / *dinB*$^+$ / SOS$^-$)	25.7 ± 1.2%	1.8 ± 0.2%	4.4 ± 0.1%	107.0 ± 2.1%	100%
HK82 SOS$^+$ (*alkB*$^-$ / *dinB*$^+$/ SOS$^+$)	75.9 ± 0.5%	12.7 ± 0.2%	13.3 ± 0.5%	95.7 ± 0.9%	100%
HK84 (*alkB*$^-$ / *dinB*$^-$ / SOS$^-$)	14.3 ± 2.6%	1.8 ± 0.1%	4.1 ± 0.1%	107.5 ± 3.0%	100%
HK84 SOS$^+$ (*alkB*$^-$ / *dinB*$^-$ / SOS$^+$)	28.9 ± 2.2%	4.7 ± 0.2%	6.6 ± 0.5%	100.5 ± 2.6%	100%

Table S3. Number of reads obtained for each lesion from next-generation sequencing. These data are the basis for determining the bypass efficiencies (see Materials and Methods). Mixing ratio is the ratio in which different lesion-containing genomes were mixed together before electroporation. Normalization ratio is the concentration correction factor obtained from the genome concentration normalization step in the assay (see Materials and Methods).

Lesions		2′-F-N^2,3-εG	1,N^2-εG	2′-F-1,N^2-εG	2′-FG	G
Mixing ratio		3	3	3	1	1
Normalization ratio		0.78	1.02	0.92	0.99	1
Cell	Replicate					
HK81	1	9611	2733	2305	19989	19949
	2	6694	1715	1803	12545	12901
	3	3769	1031	1016	7460	7965
HK82	1	3259	274	647	5638	5205
	2	9196	923	1796	16028	15242
	3	4692	430	1024	8568	8204
HK82 SOS^+	1	8384	1795	1775	4389	4685
	2	21662	4758	4297	11659	12267
	3	14999	3332	3112	8060	8447
HK84	1	5690	1024	2202	21618	19727
	2	4458	789	1580	14654	14121
	3	6603	952	1944	17345	16392
HK84 SOS^+	1	11859	2627	3159	19162	19209
	2	17103	3651	4739	24981	24493
	3	21745	4565	5617	29454	30407

Table S4. Mutation frequency and specificity of all lesions and controls in all five repair and replication backgrounds investigated. Data in each row are presented from the three biological replicates. %G, %A, %T and %C indicate the percentage of each base present at the lesion site (site N in Figure 1) of progeny DNA and %Δ denotes the frequency of deletion.

2′-F-N^2,3-εG										
	Average					Standard deviation				
Cells	%G	%A	%T	%C	%Δ	%G	%A	%T	%C	%Δ
HK81	67.4	30.2	0.4	1.2	0.7	2.0	1.8	0.1	0.1	0.1
HK82	68.7	29.7	0.6	0.1	0.7	1.2	1.1	0.3	0.1	0.4
HK82 SOS^+	59.1	38.1	2.1	0.2	0.3	1.3	1.4	0.2	<0.1	<0.1
HK84	66.1	31.2	1.2	0.2	1.1	3.9	3.3	0.5	0.1	0.1
HK84 SOS^+	56.6	38.6	4.0	0.2	0.4	0.7	0.7	0.3	<0.1	<0.1
1,N^2-εG										
	Average					Standard deviation				
Cells	%G	%A	%T	%C	%Δ	%G	%A	%T	%C	%Δ
HK81	81.0	6.4	6.0	1.6	4.9	2.8	2.0	1.3	0.4	1.2
HK82	63.8	13.3	13.0	1.2	8.7	2.0	1.6	1.9	0.7	2.4
HK82 SOS^+	26.6	39.7	28.4	3.1	2.0	0.8	1.3	0.2	0.2	0.5
HK84	60.4	15.0	12.4	1.6	10.2	5.4	3.5	3.2	0.5	2.5
HK84 SOS^+	32.2	30.8	29.8	2.9	4.1	0.3	1.3	0.3	0.4	0.4
2′-F-1,N^2-εG										
	Average					Standard deviation				
Cells	%G	%A	%T	%C	%Δ	%G	%A	%T	%C	%Δ
HK81	41.4	6.3	2.5	2.0	47.7	1.6	1.2	0.2	0.3	1.0
HK82	42.5	6.5	2.8	0.9	47.2	2.7	2.0	0.2	0.1	1.4
HK82 SOS^+	17.9	30.0	32.3	3.4	16.3	1.0	1.7	0.8	0.2	0.7
HK84	41.4	6.5	3.8	1.8	46.3	2.5	1.0	0.8	0.9	1.1
HK84 SOS^+	28.8	18.4	20.7	3.4	28.7	2.1	1.6	2.4	0.4	2.2

(Continued onto the next page)

2′-FG										
	Average					Standard deviation				
Cells	%G	%A	%T	%C	%Δ	%G	%A	%T	%C	%Δ
HK81	97.4	0.8	0.2	1.0	0.5	0.4	0.4	0.1	0.3	0.1
HK82	98.0	1.1	0.2	0.1	0.5	0.8	0.7	0.1	<0.1	0.1
HK82 SOS^+	96.5	2.0	1.0	0.1	0.2	0.9	0.7	0.1	0.1	<0.1
HK84	98.4	0.5	0.3	0.1	0.5	0.3	0.1	<0.1	<0.1	0.1
HK84 SOS^+	97.7	1.0	0.6	0.1	0.4	0.2	0.1	<0.1	<0.1	<0.1
G										
	Average					Standard deviation				
Cells	%G	%A	%T	%C	%Δ	%G	%A	%T	%C	%Δ
HK81	97.6	0.8	0.1	1.1	0.3	0.3	0.2	<0.1	0.1	<0.1
HK82	98.2	0.9	0.2	0.1	0.4	0.4	0.3	<0.1	0.1	0.1
HK82 SOS^+	97.0	1.7	0.7	0.1	0.2	0.5	0.4	0.1	0.1	0.1
HK84	98.6	0.5	0.2	<0.1	0.3	0.2	0.1	0.1	<0.1	<0.1
HK84 SOS^+	98.1	0.8	0.5	0.1	0.2	0.1	0.1	<0.1	<0.1	<0.1

Table S5. Lesion bypass efficiency of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional Competitive Replication of Adduct Bypass (CRAB) assay (1). Only a subset of the strains shown in Table S2 was analyzed by this method. Results are presented as average ± one SD of the three biological replicates. A graphical representation of these data is shown in Figure S4.

Repair/replication background	Lesions	
	2′-F-N^2,3-εG	1,N^2-εG
HK81	34.8 ± 6.1%	4.5 ± 0.6%
HK82	36.7 ± 11.8%	4.4 ± 1.4%
HK82 SOS^+	112 ± 4%	9.8 ± 1.7%

Table S6. Mutation frequency and specificity of 2′-F-N^2,3-εG and 1,N^2-εG at the lesion site as measured by the traditional Restriction Endonuclease and Postlabeling (REAP) assay (1). Only a subset of the strains shown in Table S4 was analyzed by this method. Data in each row represent the three biological replicates. %G, %A, %T and %C indicate the percentage of each base present at lesion site of progeny DNA. A graphical representation of these data is shown in Figure S5.

2′-F-N^2,3-εG								
	Average				Standard deviation			
Cells	%G	%A	%T	%C	%G	%A	%T	%C
HK81	56.8	42.0	1.0	0.2	1.5	1.4	0.1	<0.1
HK82	62.7	36.3	0.9	0.1	4.8	4.9	0.2	0.1
HK82 SOS$^+$	34.6	60.9	2.9	1.6	4.8	3.8	1.4	0.7
1,N^2-εG								
	Average				Standard deviation			
Cells	%G	%A	%T	%C	%G	%A	%T	%C
HK81	80.0	7.5	7.4	5.1	2.8	1.5	0.8	1.4
HK82	50.5	26.5	16.5	6.6	10.4	5.1	5.6	0.5
HK82 SOS$^+$	15.4	47.2	33.4	4.0	2.3	1.0	1.7	1.8

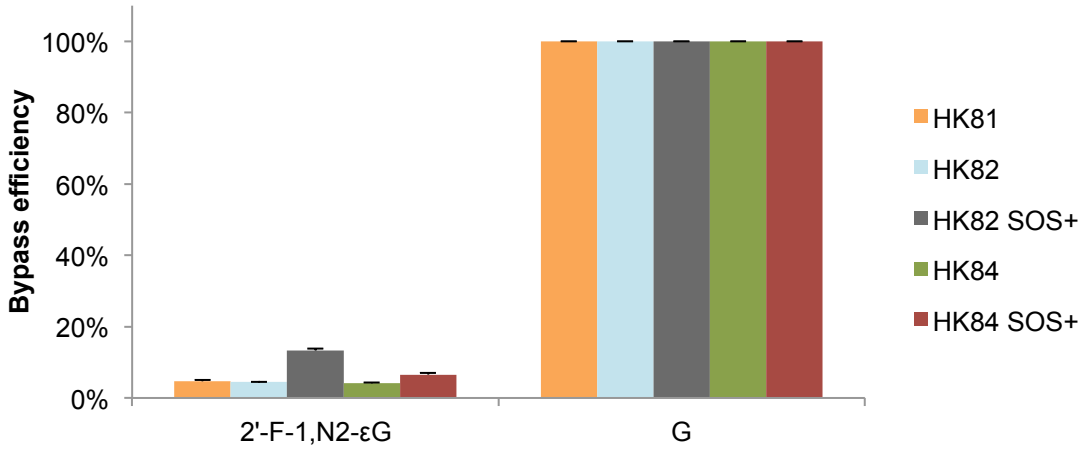


Figure S1. Lesion bypass efficiency of 2′-F-1,N^2-εG relative to the control (G) in all five repair and replication backgrounds investigated. Error bars represent one SD (N = 3).

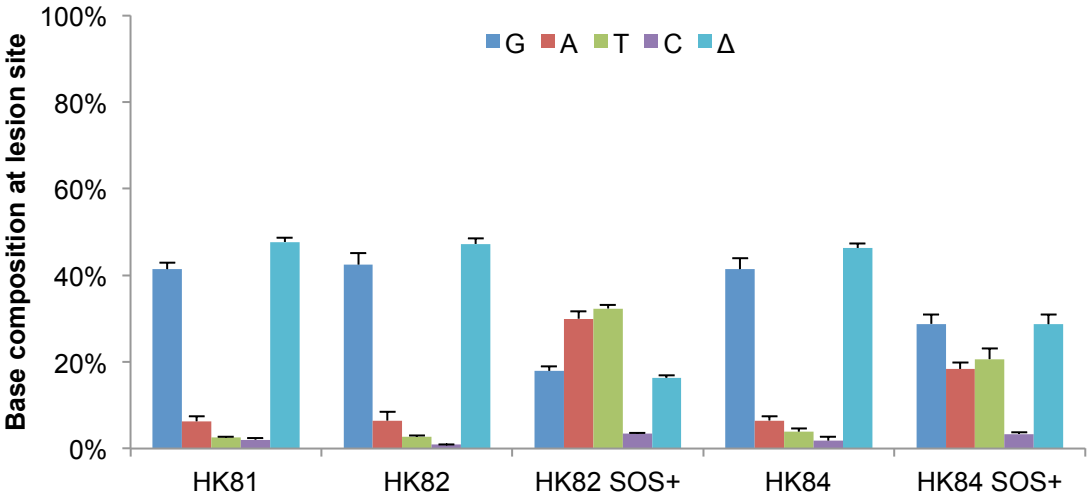


Figure S2. Mutation frequency and specificity of 2′-F-1,N^2-εG in all five repair and replication backgrounds investigated. G, A, T and C indicate the possible bases present at the lesion site (site N of Figure 1) after *in vivo* replication, and Δ denotes the occurrence of deletions at the lesion site. The apparent high occurrence of deletions was determined to be due to the presence of a small sub-population of oligonucleotides with deletions in the starting material (i.e., oligonucleotides lacking the lesion site), which was disproportionally amplified due to the lacking of the highly toxic 2′-F-1,N^2-εG lesion. Although the deletion mutation may be overly represented, nevertheless, the relative amounts of G, A, T and C bases at site N showing the point mutations induced by 2′-F-1,N^2-εG are accurate. Error bars represent one SD (N = 3).

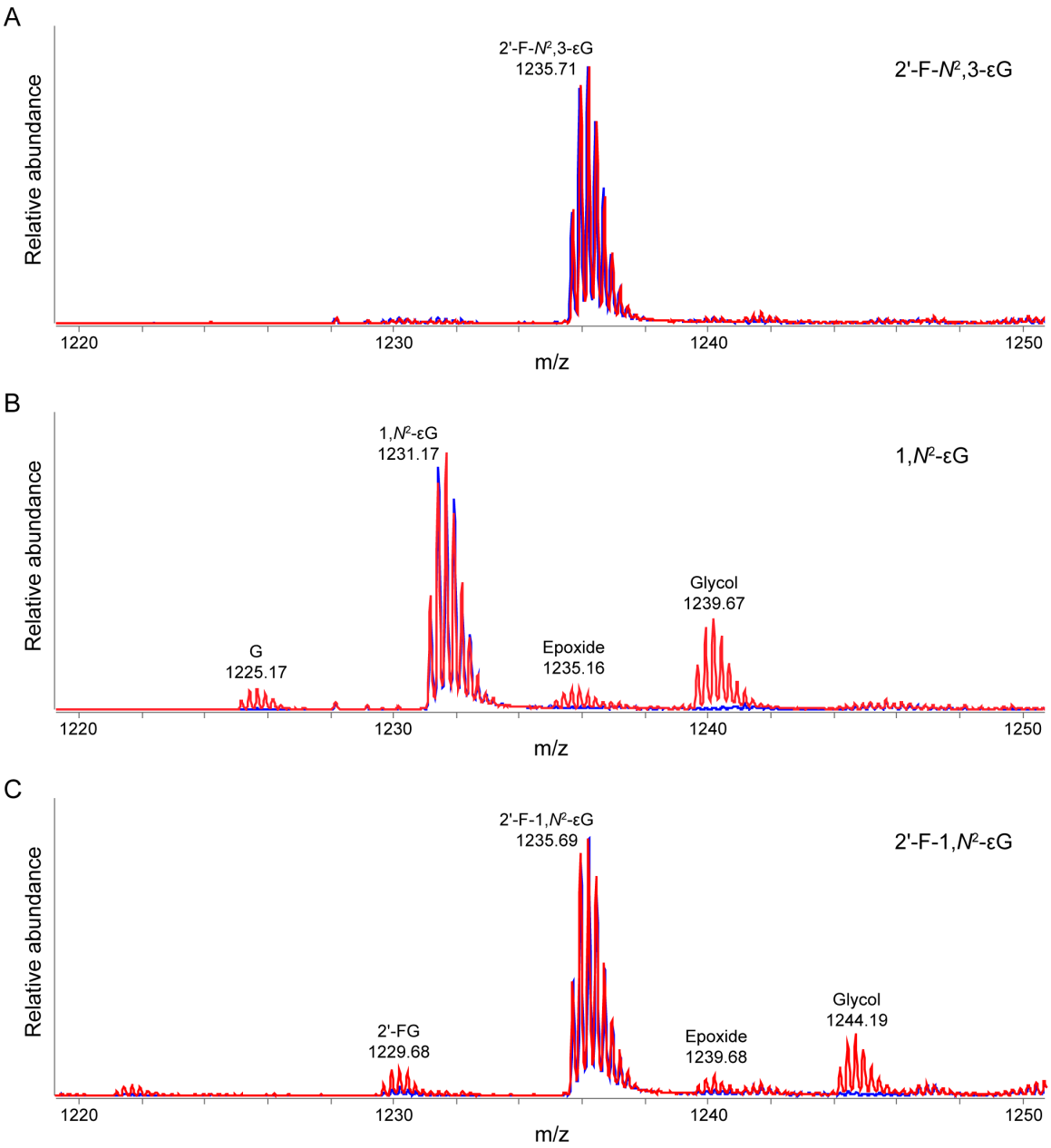


Figure S3. Mass spectra demonstrating the activity of AlkB *in vitro* on εG-containing 16-mers. Oligonucleotides were incubated in the presence (red) or absence (blue) of AlkB for one hour. Results shown here are the same as in Figure 6 with the addition of the negative controls (blue lines; no AlkB). Data represent the −4 charge envelopes and the observed monoisotopic peak value is labeled above each peak envelope. Incubation with 16-mer containing a 2′-F-N^2,3-εG (**A**), 1,N^2-εG (**B**) or 2′-F-1,N^2-εG (**C**).

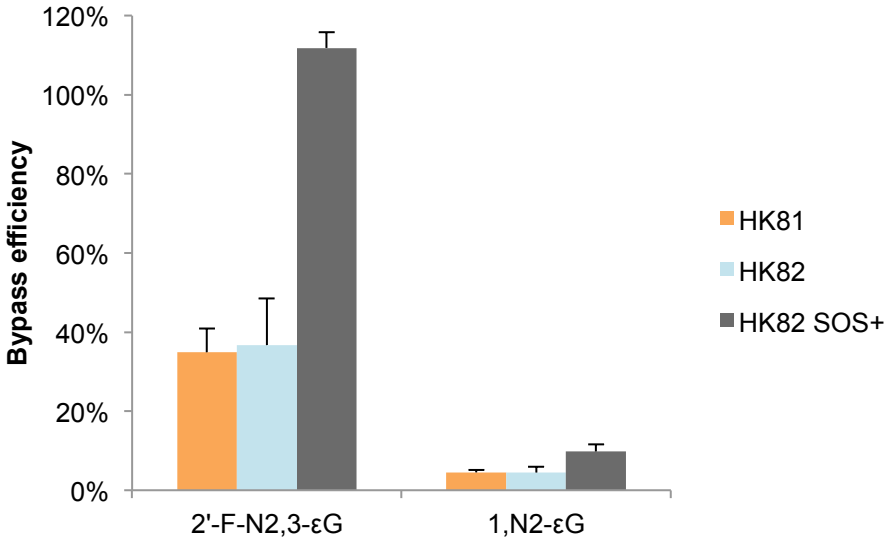
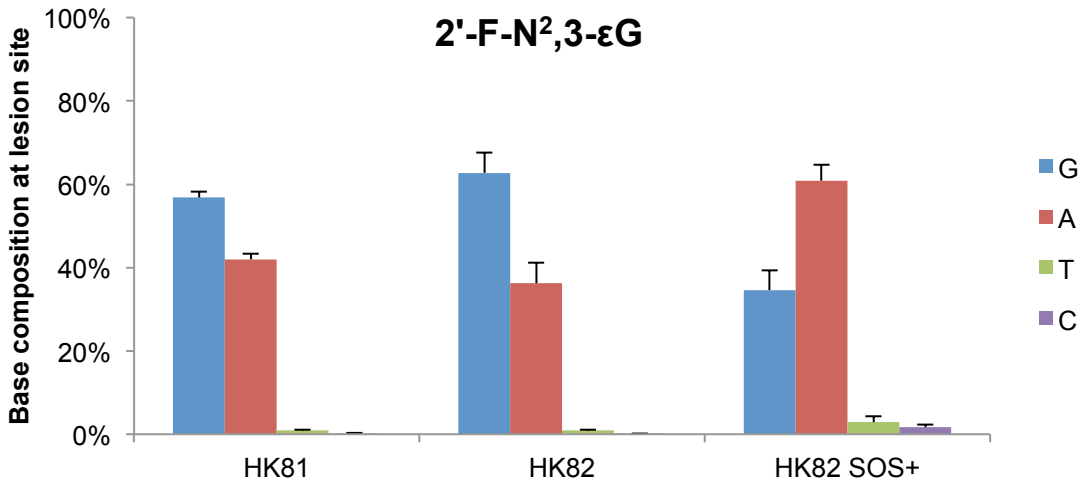


Figure S4. Lesion bypass efficiency of 2′-F-N^2,3-εG and 1,N^2-εG as measured by the traditional CRAB assay (1). Data are tabulated in Table S5. Error bars represent one SD (N = 3).

A



B

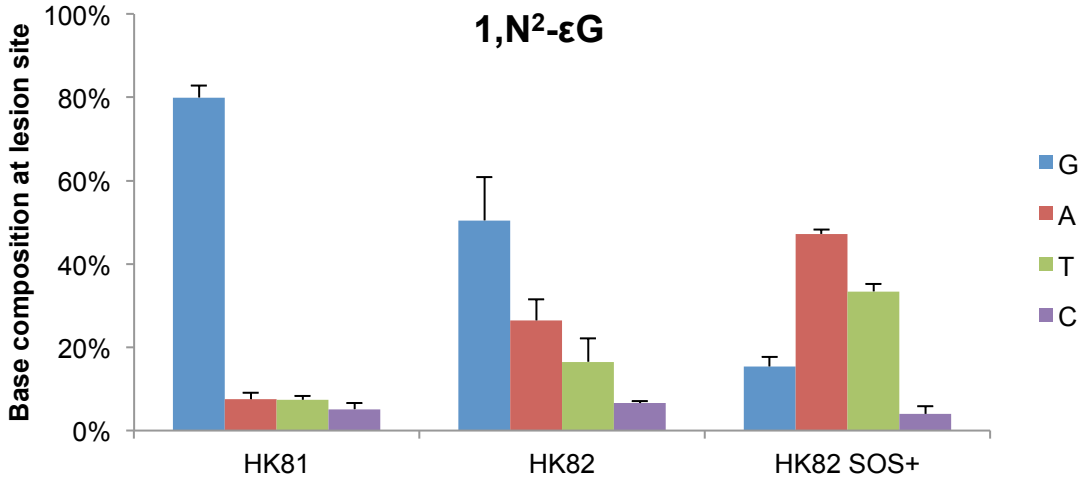


Figure S5. Mutation frequency and specificity of (**A**) 2′-F-N^2,3-εG and (**B**) 1,N^2-εG as measured by the traditional REAP assay (1). Data are tabulated in Table S6. G, A, T and C indicate the possible bases present at lesion sites in progeny DNA after *in vivo* replication. Error bars represent one SD (N = 3).

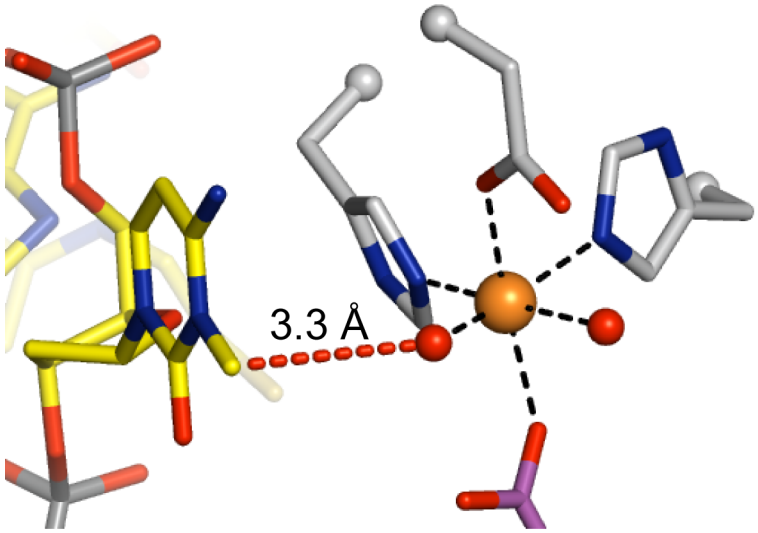


Figure S6. AlkB active site with a 3-methylcytosine lesion (yellow carbons, PDB ID 3O1M), a known good substrate for AlkB, and with the iron(IV)-oxo intermediate modeled. Selected AlkB amino acid residues are shown in grey, iron-bound succinate in purple, the iron ion as an orange sphere, and iron-bound 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. The red dashed line shows the distance from the iron(IV)-oxo center to the carbon of the methyl group on the base.

SUPPLEMENTARY REFERENCES

1. 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.