Supporting Information for

6-Thioguanine and S^6-Methylthioguanine Are Mutagenic in Human Cells

Bifeng Yuan[§], Timothy R. O'Connor[+] and Yinsheng Wang[§*]

[§]Department of Chemistry, University of California, Riverside, CA 92521; [+]Biology Division, Beckman Research Institute, City of Hope National Medical Center, Duarte, CA 91010.

The authors declare no conflict of interest.

[*]To whom correspondence should be addressed: Telephone: (951)827-2700. Fax: (951)827-4713. E-mail: yinsheng.wang@ucr.edu

Detailed Experimental Procedures

1. Construction of pTGFP-Hha10 genomes harboring a site-specifically inserted guanine, ^{S}G or S^{6}mG

The 17-mer ODN substrates containing a ^{S}G or S^{6}mG were prepared following previously published procedures (*1*), and the identities and purities of the modified ODNs were confirmed by ESI-MS and MS/MS analyses (Figures S1 and S2). To insert a structurely defined lesion into a unique site in the pTGFP-Hha10 vector, we first nicked the original vector with N.BstNB I to produce a gapped vector and a 33mer single-stranded DNA (Figure 2A in the main text). In this respect, the pTGFP-Hha10 vector contains two N.BstNB I recognition sites upstream of the coding region of the Turbo GFP (TGFP) gene, and N.BstNB I nicks only one strand of duplex DNA at the 4th nucleotide 3' to the GAGTC restriction recognition site (Figure 2B). The 33mer single-stranded ODN was subsequently removed from the nicked plasmid by annealing the cleavage mixture with the complementary 33mer ODN in 50-fold molar excess. The gapped plasmid was isolated from the mixture by using 100 kDa-cutoff ultracentrifugal filter units (Centricon 100 from Millipore). The purified gapped construct (see Lane 2 in Figure 2C) was subsequently annealed with a 5'-phosphorylated 17-mer lesion-carrying insert and a 16-mer unmodified ODN, and ligated with T4 DNA ligase (Figure 2B). In this regard, we chose to employ a 17-mer rather than 33-mer lesion-containing insert because of the relative ease in the synthesis and purification of the former substrate. The ligation mixture was incubated with ethidium bromide, and the resulting supercoiled, lesion-carrying double-stranded plasmid was isolated from the mixture by agarose gel electrophoresis (Lanes 5 & 6 in Figure 2C).

2. Cell culture, transfection and *in-vivo* replication

The 293T and human fibroblast cells were cultured at 37°C in 5% CO_2 atmosphere and in Dulbecco's Modified Eagle medium supplemented with 10% fetal bovine serum

(Invitrogen, Carlsbad, CA), 100 U/mL penicillin, and 100 µg/mL streptomycin (ATCC). Cells (3×10^5) were seeded in 6-well plates and cultured overnight, after which they were transfected with 500 ng of the lesion-containing or lesion-free pTGFP-Hha10 vector using Lipofectamine 2000 (Invitrogen). After *in-vivo* replication for 24 h, DNase I was added to digest the residual DNA adsorbed to the exterior of cells (*2*). Cells were subsequently detached by treating with trypsin-EDTA (ATCC, Manassas, VA), the progenies of the plasmid were isolated by using an alkali lysis method (*3*), and the residual unreplicated plasmid was removed by DpnI digestion.

3. Determination of the bypass efficiency and mutation frequency

The bypass efficiencies and mutation frequencies were determined by a method adapted from the REAP assay (*4*), which was initially developed by Essigmann and coworkers for assessing the cytotoxic and mutagenic properties of DNA lesions in *E. coli* cells using single-stranded M13 genome (*5, 6*). The progeny genomes were subsequently amplified by PCR using Phusion high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA). The primers flanking the site where the lesion was initially inserted were 5'-YGCAGAGCTGGTTTAGTGAACCGTCAG-3' and 5'-YCTCTGCTGAAGCCAGTTACCTTCGG-3' (Y is an amino group), and the amplification cycle was 35, each consisting of 10 s at 98°C, 30 s at 62°C, 2 min at 72°C, with a final extension at 72°C for 5 min. The resulting 3950 bp PCR products were purified by using QIAquick PCR purification kit (Qiagen, Valencia, CA). For determining the bypass efficiency and mutation frequency by using the modified REAP assay, a portion of the above PCR fragments was treated with 10 U Sac I (recognition site highlighted in bold in Figure 3A) and 1 U shrimp alkaline phosphatase in 10-µl NEB buffer 4 at 37°C for 1 h, followed by heating at 65°C for 20 min to deactivate the phosphatase. The above mixture was then treated with a 15-µl NEB buffer 4 containing 5 mM DTT, ATP (50 pmol cold, premixed with 1.66 pmol [γ-^{32}P]ATP) and 10 U

polynucleotide kinase. The reaction was continued at 37°C for 1 h, followed by heating at 65°C for 20 min to deactivate the polynucleotide kinase. To the reaction mixture was subsequently added 10 U Fsp I (recognition sites highlighted in bold in Figure 3A), and the solution was incubated at 37°C for 1 h, followed by quenching with 15 μl formamide gel loading buffer containing xylene cyanol FF and bromophenol blue dyes. The mixture was loaded onto a 30% denaturing polyacrylamide gel with 5 M urea (acrylamide:bis-acrylamide=19:1) and products were quantified by phosphorimager analysis (*4*). After these cleavages, the original lesion site was housed in a 14mer/10mer duplex, d(pGCAAAMCTAGAGCT)/d(p*CTAGNTTTGC) (bottom strand product), where 'M' represents the nucleobase incorporated at the initial damage site during *in-vivo* DNA replication, 'N' is the paired nucleobase of 'M' in the complementary strand and "p*" designates the 5'-radiolabeled phosphate (Figure 3A). The mutation frequencies were determined from the relative amounts of different 10-mer products from the gel band intensities. On the other hand, the restriction cleavage of the product arising from the replication of the lesion-free top-strand gave d(pGCAAAGCTTGAGCT)/d(p*CAAGCTTTGC). The 10-mer products were monitored instead of the 14-mer products because the former products could be resolved by denaturing polyacrylamide gel. The bypass efficiency was calculated using the following formula, %bypass = (lesion vector bottom-strand signal/lesion vector top-strand signal)/(control vector bottom-strand signal/control vector top-strand signal).

4. Identification of replication products by using LC-MS/MS

In order to further identify the replication products using LC-MS, PCR products were treated with 50 U Sac I, 50 U Fsp I and 20 U shrimp alkaline phosphatase in 200-μl NEB buffer 4 at 37°C for 2 h, followed by heating at 65°C for 20 min. The resulting solution was then extracted once with phenol/chloroform/isoamyl alcohol (25:24:1, v/v), and the aqueous portion was dried with Speed-vac and dissolved in 12 μl water. The ODN

mixture was subjected to LC-MS/MS analysis. A 0.5×150 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies) was used for the separation, and the flow rate was 8.0 μl/min, which was delivered by using an Agilent 1100 capillary HPLC pump. A gradient composed of 0-20% methanol in 5 min followed by 20-50% methanol in 35 min, both in 400 mM HFIP with pH being adjusted to 7.0 through the addition of triethylamine, was employed for the separation (*7*). The effluent from the LC column was coupled directly to an LTQ linear ion trap mass spectrometer (Thermo Electron, San Jose, CA), which was set up for monitoring the fragmentation of the $[M-4H]^{4-}$ ions of the 14-mer [d(GCAAAMCTAGAGCT), where "M" designates A, T, C, or G].

References:

(1) Xu, Y. (1996) Post-synthetic introduction of labile functionalities onto purine residues via 6-methylthiopurines in oligodeoxyribonucleotides. *Tetrahedron 52*, 10737-10750.

(2) Baker, D. J., Wuenschell, G., Xia, L., Termini, J., Bates, S. E., Riggs, A. D., and O'Connor, T. R. (2007) Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J. Biol. Chem. 282*, 22592-22604.

(3) Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol. 26*, 365-369.

(4) 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. USA 105*, 8679-8684.

(5) 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.

(6) Delaney, J. C., and Essigmann, J. M. (1999) Context-dependent mutagenesis by DNA lesions. *Chem. Biol. 6*, 743-753.

(7) Yuan, B., and Wang, Y. (2008) Mutagenic and cytotoxic properties of 6-thioguanine, S^6-methylthioguanine, and guanine-S^6-sulfonic acid. *J. Biol. Chem. 283*, 23665-23670.

Figure S1. The product-ion spectrum of the ESI-produced $[M-4H]^{4-}$ ion (*m/z* 1304.1) of d(GCGCAAA^{S}GCTAGAGCTC). Shown in the inset is the negative-ion ESI-MS for the ODN and a scheme summarizing the observed fragment ions.

Figure S2. The product-ion spectrum of the ESI-produced $[M-4H]^{4-}$ ion (*m/z* 1307.4) of d(GCGCAAAS^{S}mGCTAGAGCTC). Shown in the inset is the negative-ion ESI-MS for the ODN and a scheme summarizing the observed fragment ions.

Figure S3. LC-MS/MS for monitoring the restriction fragments of interest without mutation or with a G→A mutation at the original ^{S}G or *S*^{6}mG site [i.e., d(GCAAAGCTAGAGCT) and d(GCAAAACTAGAGCT)]. Shown in (A) and (B) are the MS/MS of the $[M–4H]^{4-}$ ions (*m/z* 1072.7 and 1068.7) of these two ODNs and depicted in the insets are schemes summarizing the observed fragment ions.

Figure S1

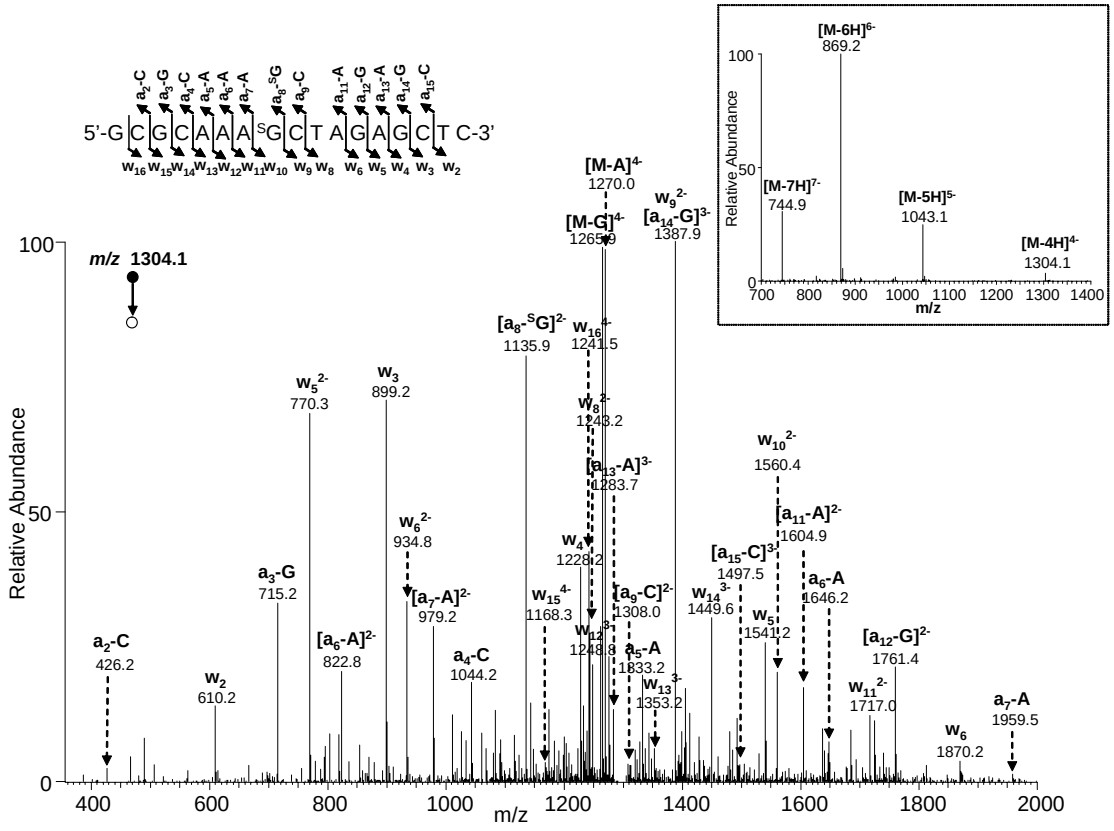

5'-G C G C A A A SG C T A G A G C T C-3'
m/z 1304.1
Relative Abundance
m/z
[M-6H]6- 869.2
[M-7H]7- 744.9
[M-5H]5- 1043.1
[M-4H]4- 1304.1
[M-A]4- 1270.0
[M-G]4- 1265.9
w9 2- [a14-G]3- 1387.9
[a8-SG]2- 1135.9
w16 4- 1241.5
w5 2- 770.3
w3 899.2
w8 2- 1243.2
w10 2- 1560.4
[a13-A]3- 1283.7
[a11-A]2- 1604.9
w6 2- 934.8
w4 1228.2
[a15-C]3- 1497.5
a6-A 1646.2
a3-G 715.2
[a7-A]2- 979.2
w15 4- 1168.3
[a9-C]2- 1308.0
w14 3- 1449.6
w5 1541.2
[a12-G]2- 1761.4
a2-C 426.2
[a6-A]2- 822.8
a4-C 1044.2
w12 3- 1248.8
a5-A 1833.2
w13 3- 1353.2
w11 2- 1717.0
a7-A 1959.5
w2 610.2
w6 1870.2

Figure S2

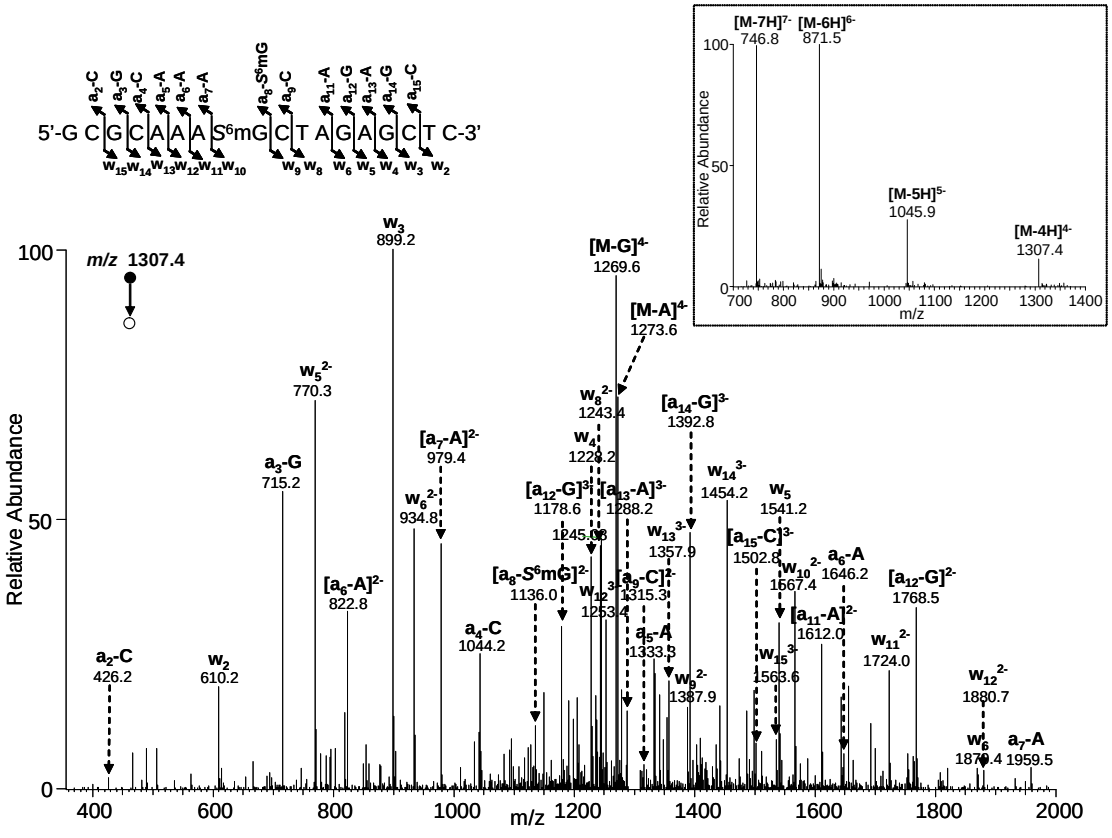
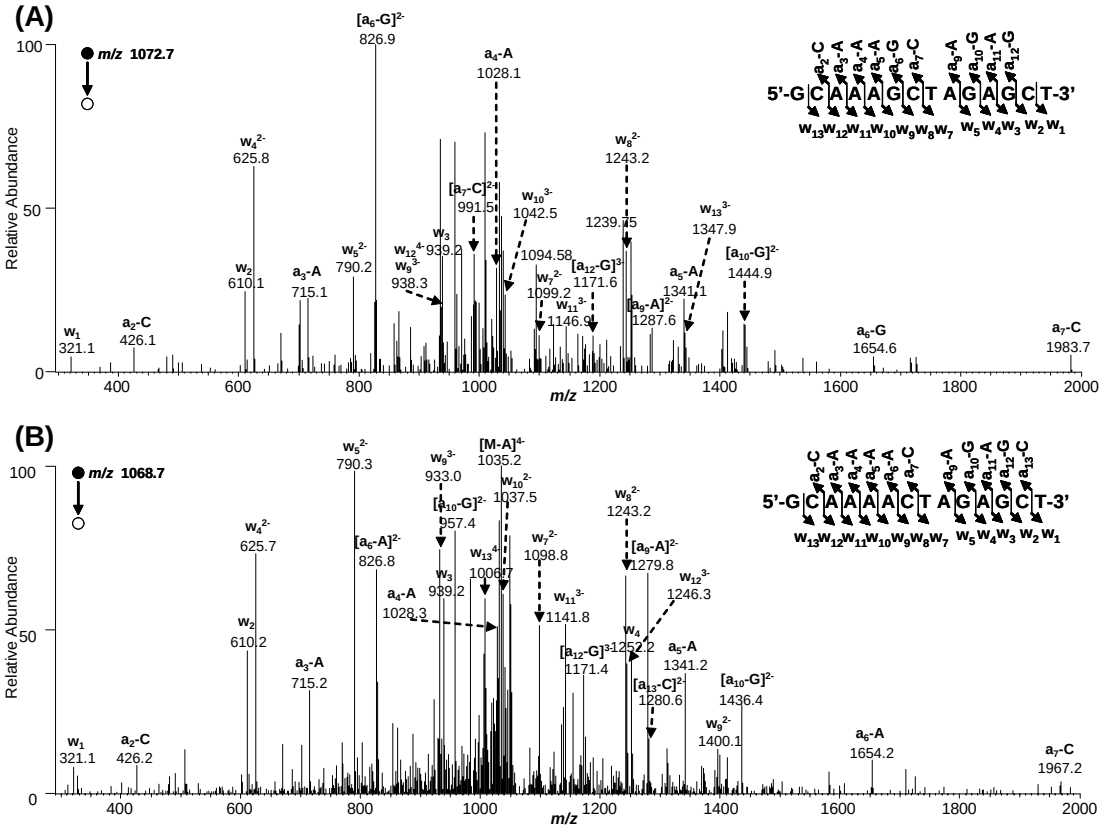

5'-G C G C A A A S6mG C T A G A G C T C-3'
m/z 1307.4
Relative Abundance
m/z
[M-7H]7- 746.8
[M-6H]6- 871.5
[M-5H]5- 1045.9
[M-4H]4- 1307.4
[M-G]4- 1269.6
[M-A]4- 1273.6
w3 899.2
w5 2- 770.3
a3-G 715.2
[a7-A]2- 979.4
w6 2- 934.8
[a6-A]2- 822.8
a2-C 426.2
w2 610.2
a4-C 1044.2
[a8-S6mG]2- 1136.0
w8 2- 1243.4
w4 1228.2
[a12-G]3- 1178.6
[a13-A]3- 1288.2
1245.08
w13 3- 1357.9
[a14-G]3- 1392.8
w14 3- 1454.2
w5 1541.2
[a15-C]3- 1502.8
w10 2- 1667.4
a6-A 1646.2
[a12-G]2- 1768.5
[a11-A]2- 1612.0
w11 2- 1724.0
w12 2- 1880.7
w6 1879.4
a7-A 1959.5
w12 3- 1253.4
a5-A 1333.3
a9-C]2- 1315.3
w15 3- 1563.6
w9 2- 1387.9

Figure S3


(A)
m/z 1072.7
Relative Abundance
m/z
5'-GCAAAAGCTAGAAGCT-3'
(B)
m/z 1068.7
Relative Abundance
m/z
5'-GCAAAACTAGAAGCT-3'