Supplementary Materials for

"DNA replication studies of *N*-nitroso compound-induced O^6-alkyl-2′-deoxyguanosine lesions in *Escherichia coli*"

Pengcheng Wang[1,2], Jiapeng Leng[1] and Yinsheng Wang[1*]

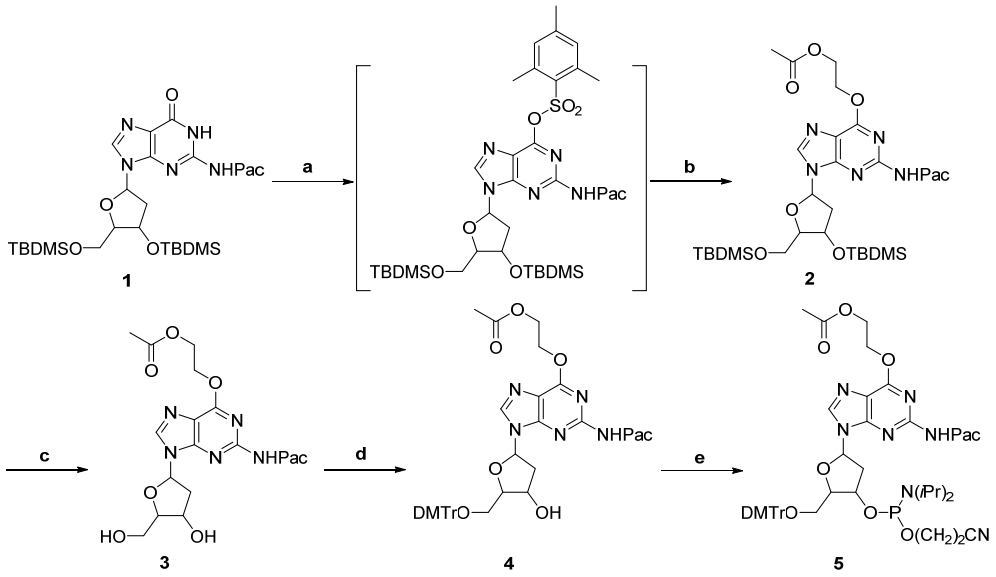
[1]Department of Chemistry, University of California, Riverside, California 92521-0403, United States

[2]Institute of Surface Analysis and Chemical Biology, University of Jinan, Jinan, Shandong 250022, China

*To whom correspondence should be addressed: Tel. (951)827-2700; E-mail: Yinsheng.Wang@ucr.edu

Supplementary Materials and Methods:

Scheme S1. Synthetic procedures for the phosphoramidite building block of O^6-HOEt-dG*



*Reagents and conditions: a) 2-mesitylenesulfonyl chloride, DMAP, 40 min; b) *N*-methylpyrrolidine, 30 min; ethylene glycol monoacetate, DBU, 6 h; c) TBAF, THF, r.t., 1 h; d) DMTr-Cl, DMAP, pyridine, r.t., 10 h; e) 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphoramidite, DIEA, CH_2Cl_2, 1 h.

Reaction yields, and NMR and mass spectrometric characterizations of the synthetic products:

3′,5′-*O*-Bis(*tert*-butyldimethylsilyl)-N^2-(phenylacetyl)-O^6-acetyloxyethoxyl-2′-deoxyguanosine (2). Obtained as brownish solid (42% yield); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.21 (s, 1H), 7.92 (s, 1H), 7.43 – 7.28 (m, 5H), 6.42 (t, *J* = 6.4 Hz, 1H), 4.76 – 4.72 (m, 2H), 4.60 (dt, *J* = 5.8, 3.6 Hz, 1H), 4.50 – 4.46 (m, 2H), 4.17 (s, 2H), 4.02 (dd, *J* = 6.7, 3.3 Hz, 1H), 3.83 (ddd, *J* = 14.3, 11.3, 3.4 Hz, 2H), 2.55 (dt, *J* = 12.7, 6.3 Hz, 1H), 2.47 – 2.39 (m, 1H), 2.07 (s, 3H), 0.93 (s, 9H), 0.92 (s, 9H), 0.11 (s, 12H). HRMS (ESI) calcd for $C_{34}H_{54}N_5O_7Si_2$ $[M+H]^+$ 700.3556, found 700.3562.

N^2-(phenylacetyl)-O^6-acetyloxyethoxyl-2′-deoxyguanosine (3). Obtained as yellowish film (56% yield); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.20 (s, 1H), 8.16 (s, 1H), 7.43 – 7.28 (m, 5H), 6.34 (t, *J* = 6.7 Hz, 1H), 5.32 (s, 1H), 4.93 – 4.88 (m, 1H), 4.75 – 4.69 (m, 2H), 4.48 – 4.42 (m, 2H), 4.17 – 4.12 (m, 1H), 4.01 – 3.81 (m, 4H), 3.36 – 3.28 (m, 1H), 2.92 (dt, *J* = 13.3, 6.6 Hz, 1H),

2.47 – 2.39 (m, 1H), 2.05 (s, 3H). HRMS (ESI) calcd for $C_{22}H_{26}N_5O_7$ $[M+H]^+$ 472.1827, found 472.1836.

5′-*O*-(4,4′-dimethoxyltrityl)-N^2-(phenylacetyl)-O^6-acetyloxyethoxyl-2′-deoxyguanosine (4). Obtained as yellowish foam (51% yield); ^{1}H NMR (400 MHz, CD_3OD) δ 8.26 (s, 1H), 7.35 – 7.13 (m, 14H), 6.70 (dd, *J* = 17.3, 8.9 Hz, 4H), 6.46 (t, *J* = 6.5 Hz, 1H), 4.84 – 4.81 (m, 1H), 4.78 (dd, *J* = 5.8, 3.2 Hz, 2H), 4.52 – 4.48 (m, 2H), 4.15 – 4.10 (m, 1H), 3.86 (s, 2H), 3.75 (s, 3H), 3.74 (s, 3H), 3.45 (dd, *J* = 10.4, 6.4 Hz, 1H), 3.25 (dd, *J* = 10.4, 2.8 Hz, 1H), 3.06 – 2.97 (m, 1H), 2.54 – 2.45 (m, 1H), 2.03 (s, 3H). HRMS (ESI) calcd for $C_{43}H_{44}N_5O_9$ $[M+H]^+$ 774.3134, found 774.3143.

5: ^{31}P NMR (120 MHz. $CDCl_3$): δ 149.85, 149.79.

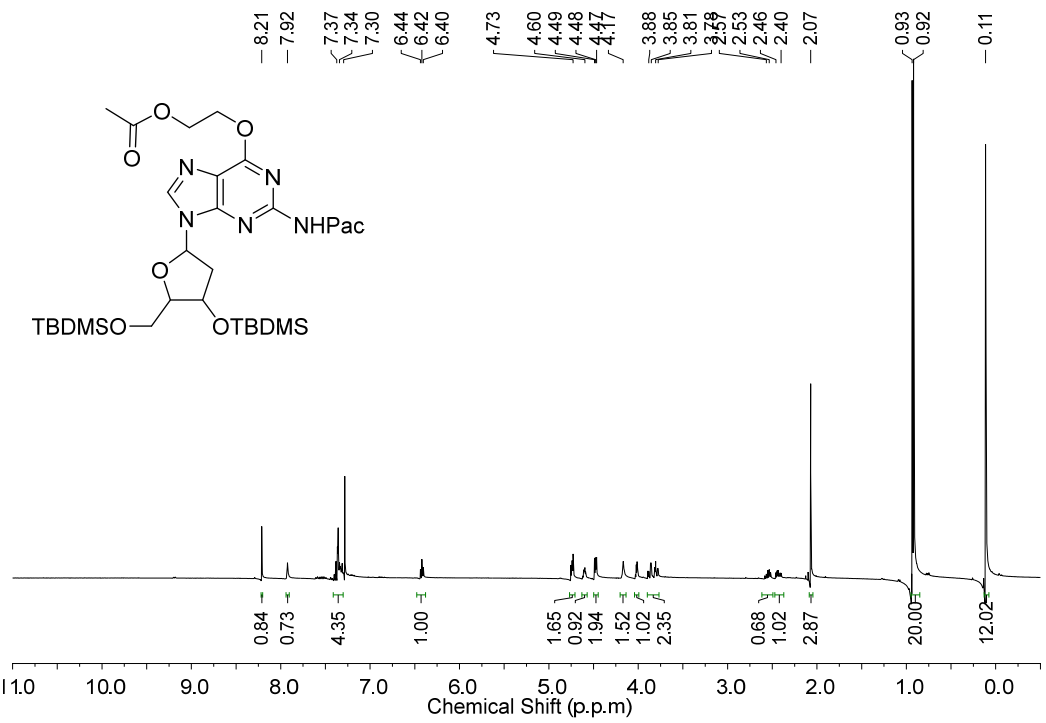


Figure S1. ^{1}H NMR spectrum of **2** (400 MHz, $CDCl_3$, 25°C).

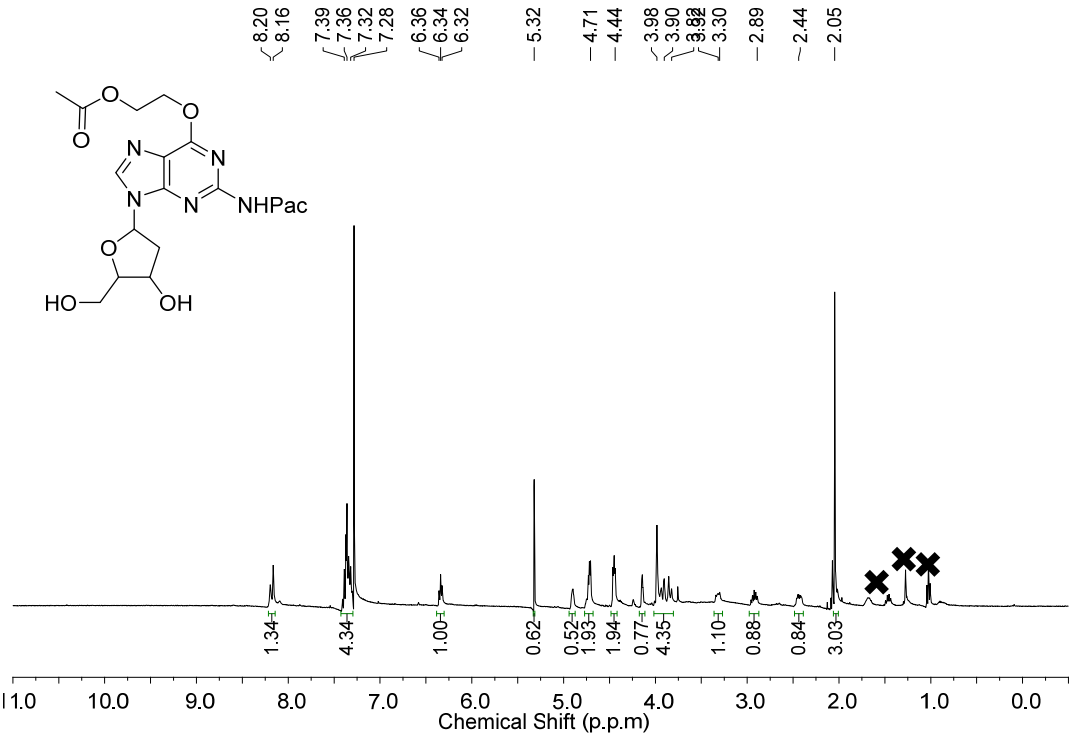


Figure S2. ^{1}H NMR spectrum of **3** (400 MHz, $CDCl_3$, 25°C). 'X' indicates peaks from residual TBAF.

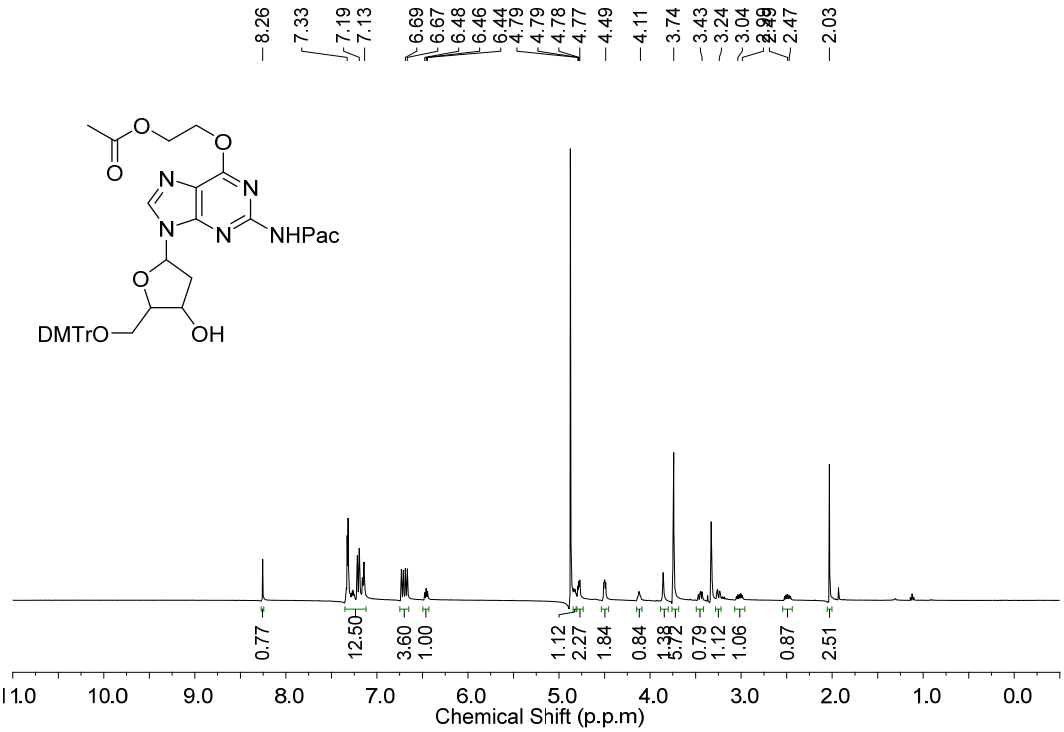


Figure S3. ^{1}H NMR spectrum of **4** (400 MHz, CD_3OD, 25°C).

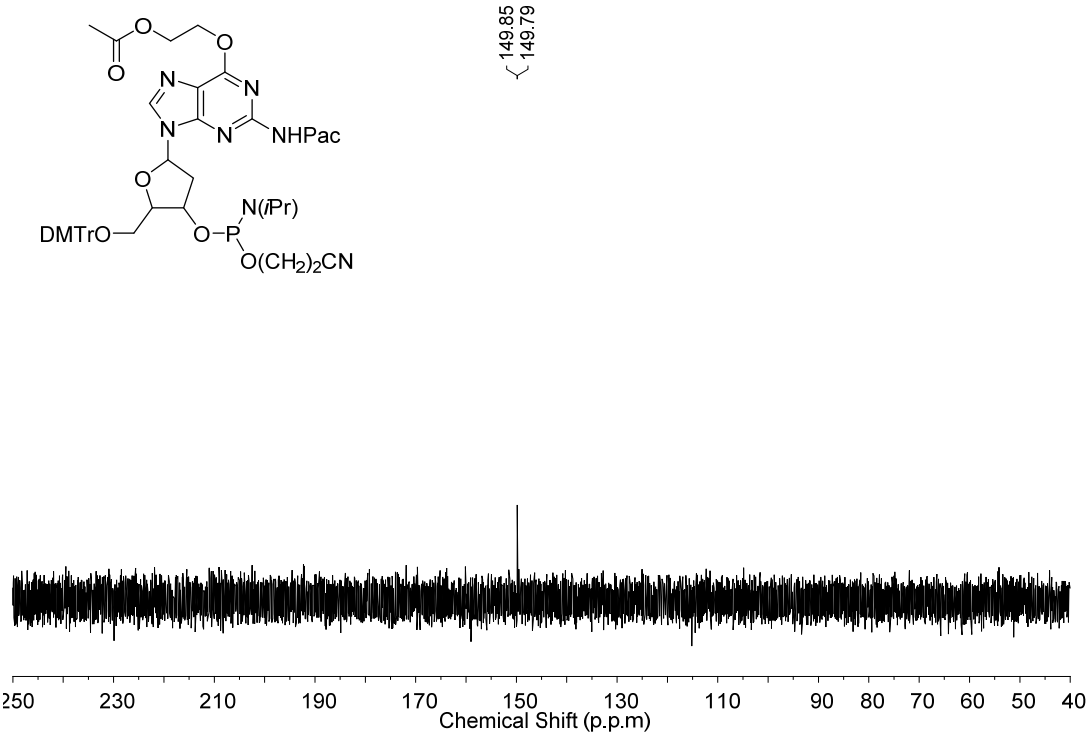


Figure S4. ^{1}H NMR spectrum of **5** (80 MHz, $CDCl_3$, 25°C).

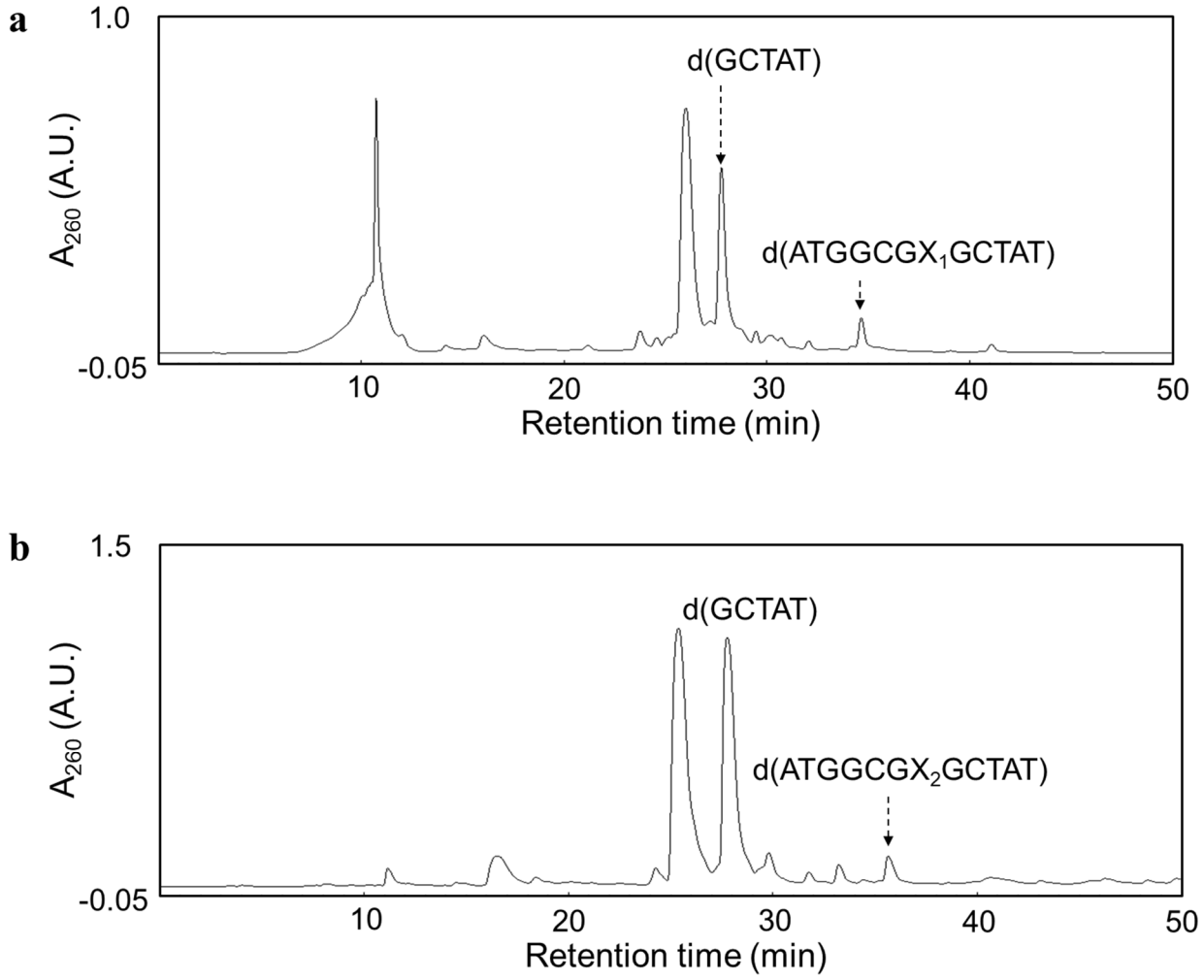


Figure S5. HPLC traces for the separation of the synthesized 12mer ODNs: (a) X_1 = O^6-ACM-dG and (b) X_2 = O^6-HOEt-dG.

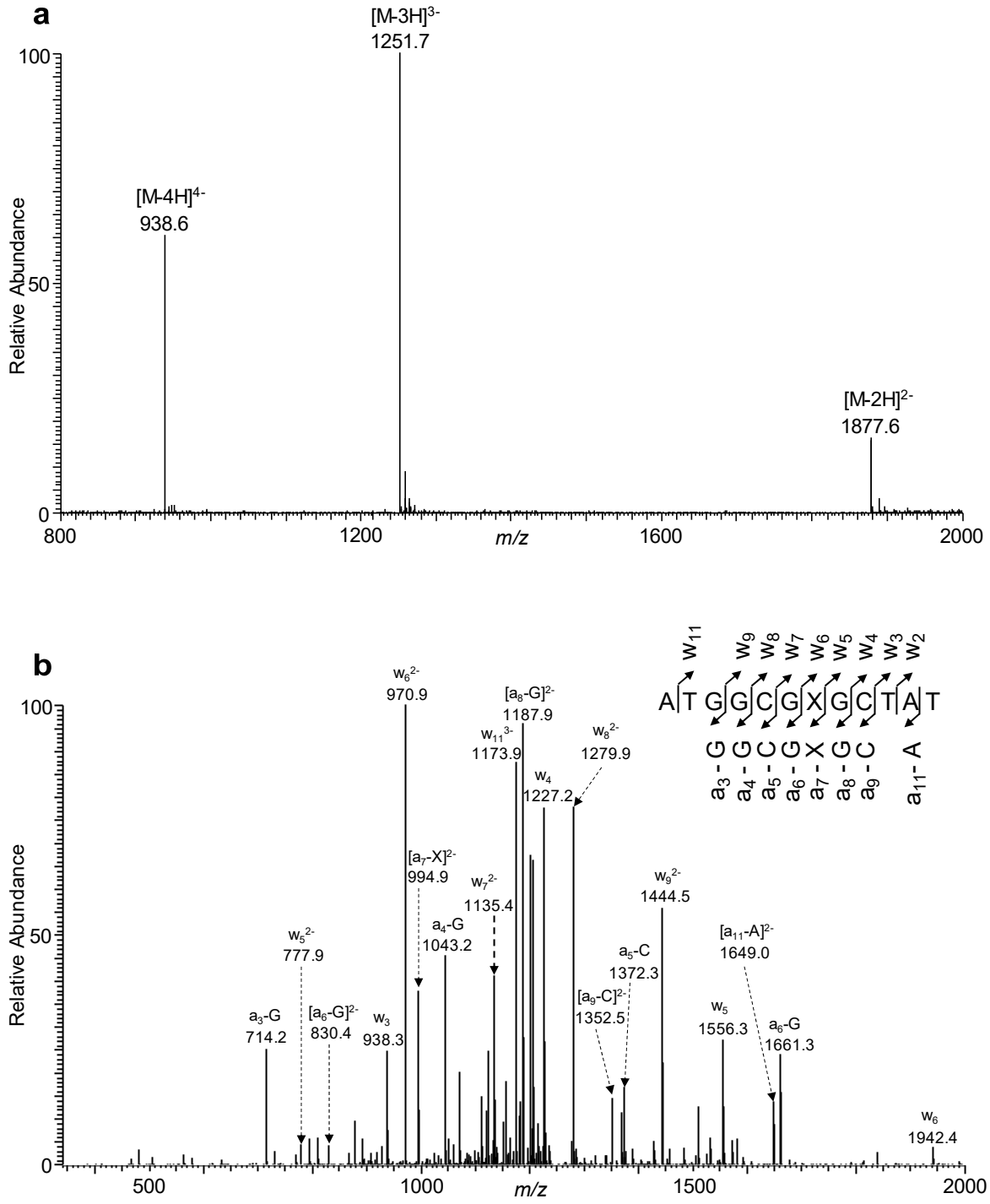


Figure S6. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^6-ACM-dG. (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1251.7).

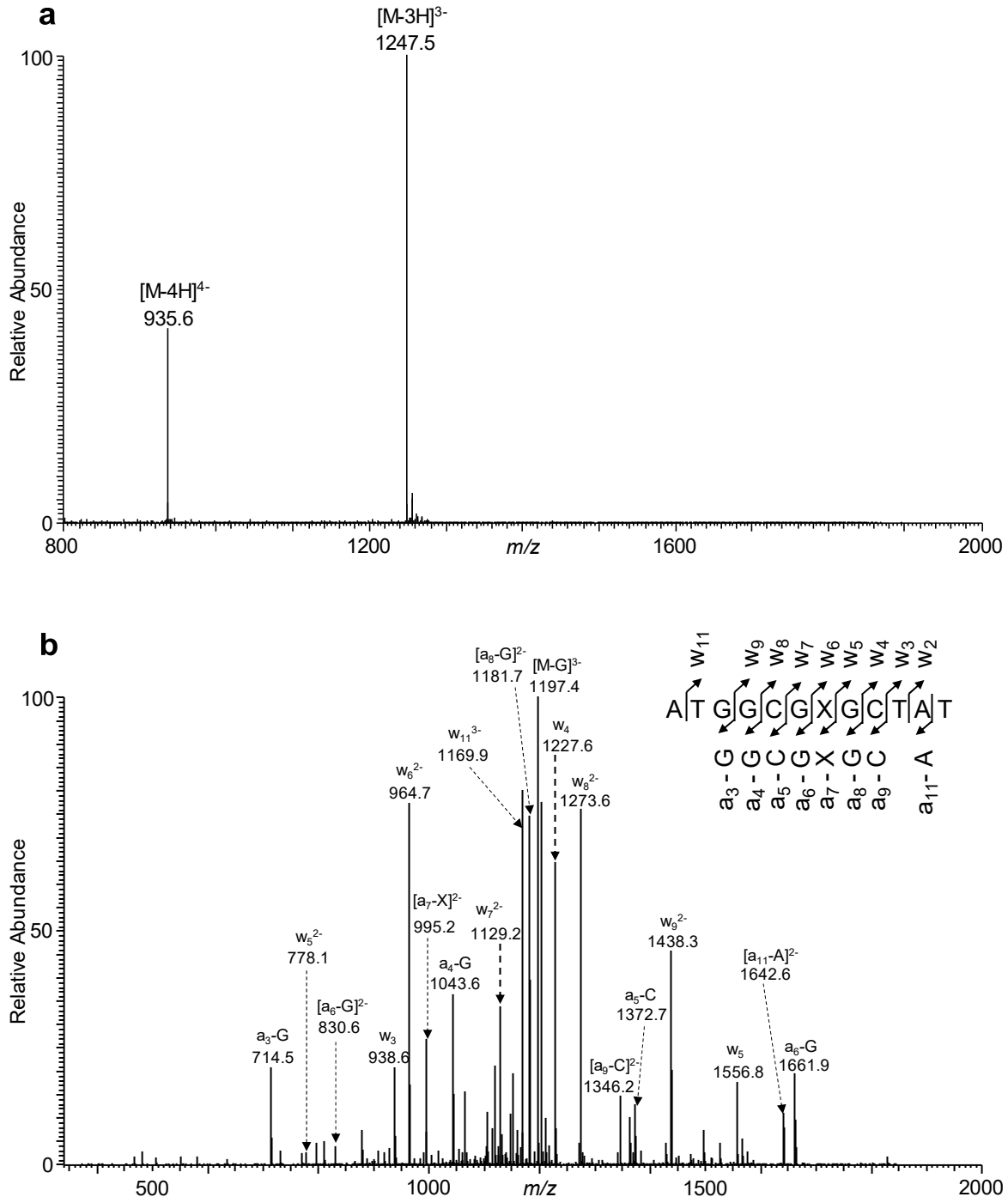


Figure S7. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^6-HOEt-dG. (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1247.5).

a. Construction of the Lesion-containing genome:

5′-GAAAGCTATGACCATGATTCAGTGAGTGGAAGACATGGCGXGCTATAATTCACTGGCCGTCGTTTT-3′
3′-CTTTCGATACTGGTACTAAGTCACTCACCTTC-3′ 3′-CGATATTAAGTGACCGGCAGCAAAA-5′

b. Construction of the competitor genome:

5′-AGCTATGACCATGATTCAGTGAGTGGAAGACATGGCGATAAGCTATAATTCACTGGCCGTCGTT-3′
3′-TCGATACTGGTACTAAGTCACTCACCTTC-5′ 3′-CGATATTAAGTGACCGGCAGCAA-5′

Figure S8. Schematic diagrams showing the constructions of the lesion-containing (a) and competitor (b) genomes. Displayed are the partial sequence of the linearized M13, the 22-mer lesion-containing or the 25-mer lesion-free competitor insert (shown in red), and the two scaffolds employed for the ligation reactions. The lesion site is underlined (X).

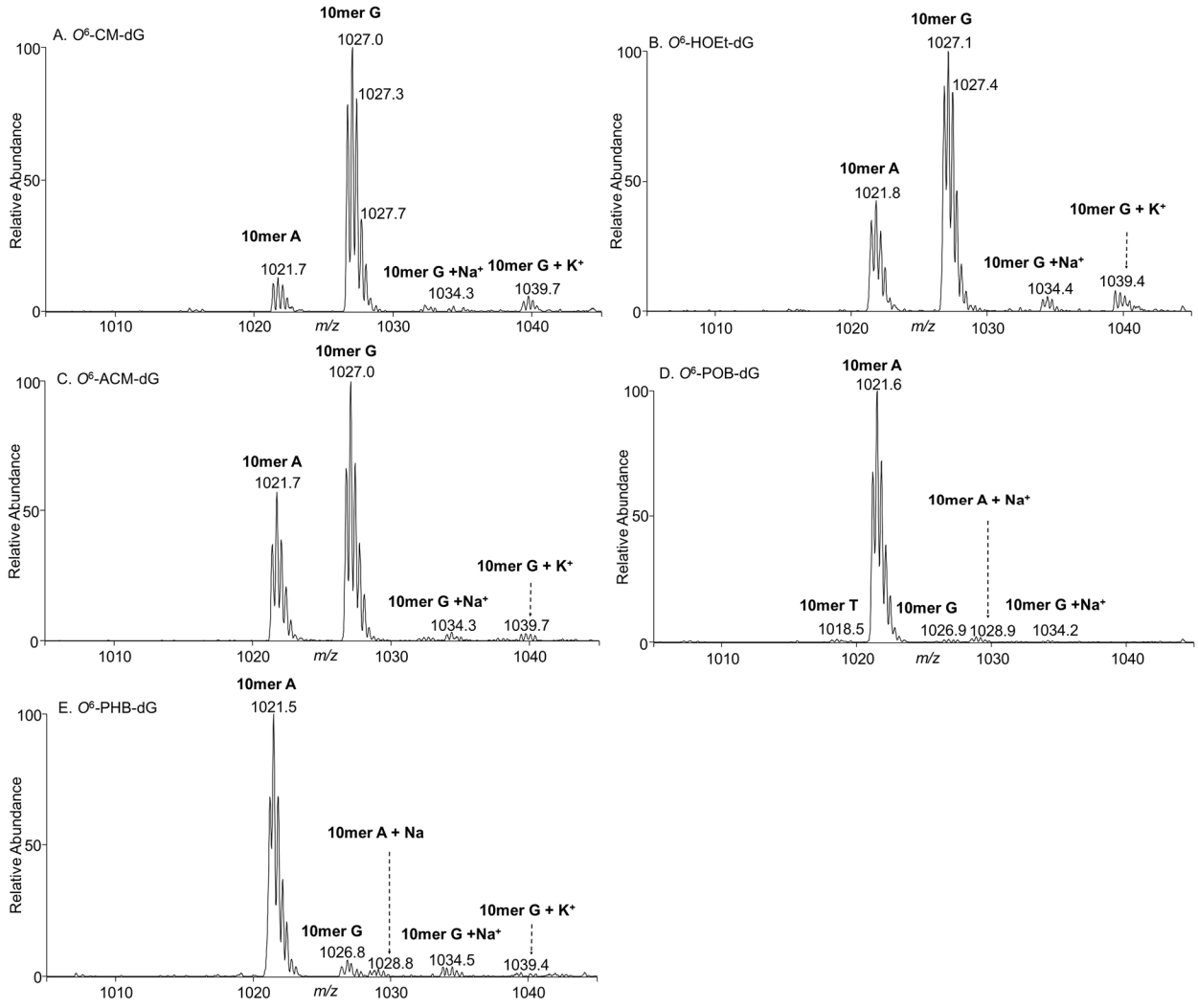


Figure S9. Higher-resolution "zoom" scan ESI-MS for monitoring the $[M - 3H]^{3-}$ ions of the lesion-containing strand of the restriction fragments of the PCR products from the replication of: (A) O^6-CM-dG-, (B) O^6-HOEt-dG-, (C) O^6-ACM-dG-, (D) O^6-POB-dG-, and (E) O^6-PHB-dG-bearing single-stranded M13 genomes in SOS-induced wild-type AB1157 cells. The G→A and G→T mutation products were further confirmed by MS/MS analyses, and representative MS/MS results for O^6-POB-dG are shown in Figure S10.

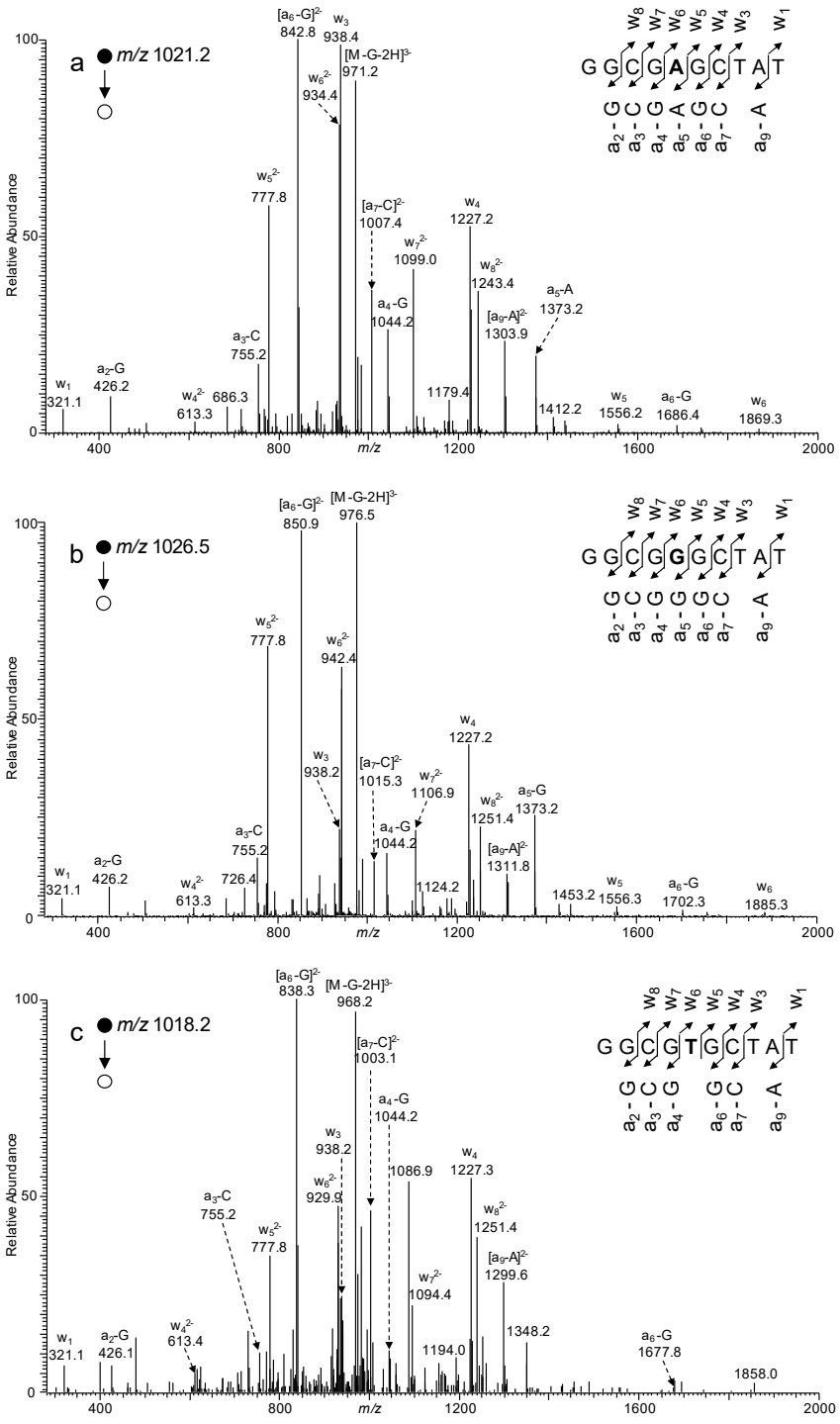


Figure S10. Representative LC-MS/MS data for the identification of (a) G→A mutation and (b) non-mutagenic product observed for the O^6-CM-dG, and (c) G→T mutation observed for O^6-POB-dG in SOS-induced wild-type (WT) AB1157 *E. coli* cells. Shown are the MS/MS for the $[M - 3H]^{3-}$ ions of (a) 10 mer A (G→A mutation), (b) 10 mer G (non-mutagenic product), and (c) 10 mer T (G→T mutation).

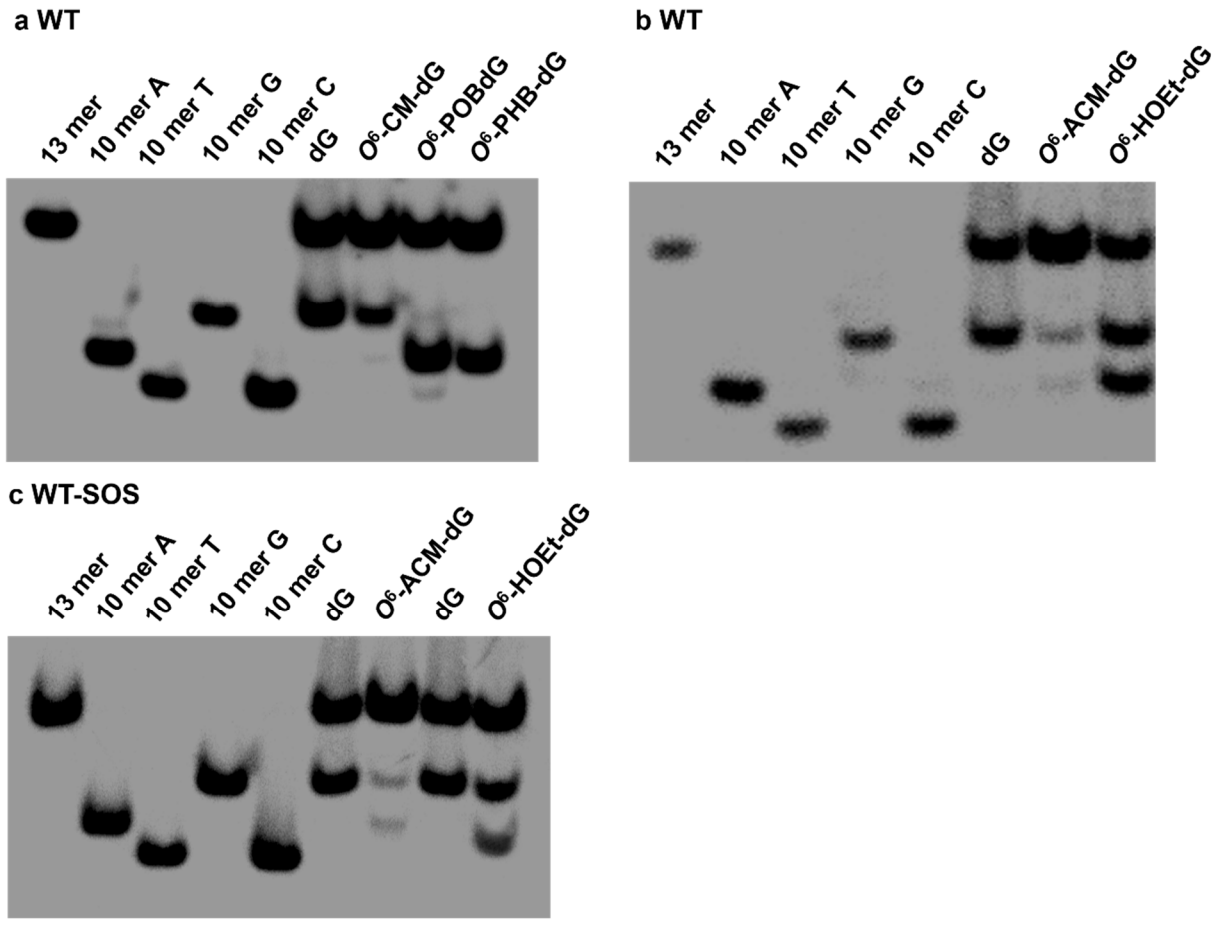


Figure S11. Native PAGE (30%) for monitoring the bypass efficiencies and mutation frequencies of (a) O^6-CM-dG, O^6-POB-dG, and O^6-PHB-dG in wild-type (WT) AB1157 *E. coli* cells; and O^6-ACM-dG, and O^6-HOEt-dG in (b) WT and (c) SOS-induced WT AB1157 *E. coli* cells. Gel image showing the 13-mer and 10-mer products released from the original lesion-situated strand of the PCR products of the progeny of the competitor genome (13-mer, i.e. 5′-GGCGATAAGCTAT-3′) and the control or lesion-carrying genome (10-mer), where 10mer A, 10mer C, 10mer G, and 10mer T represent the [5′-^{32}P]-labeled standard ODNs 5′-GGCGMGCTAT-3′, with 'M' being A, C, G, and T, respectively.

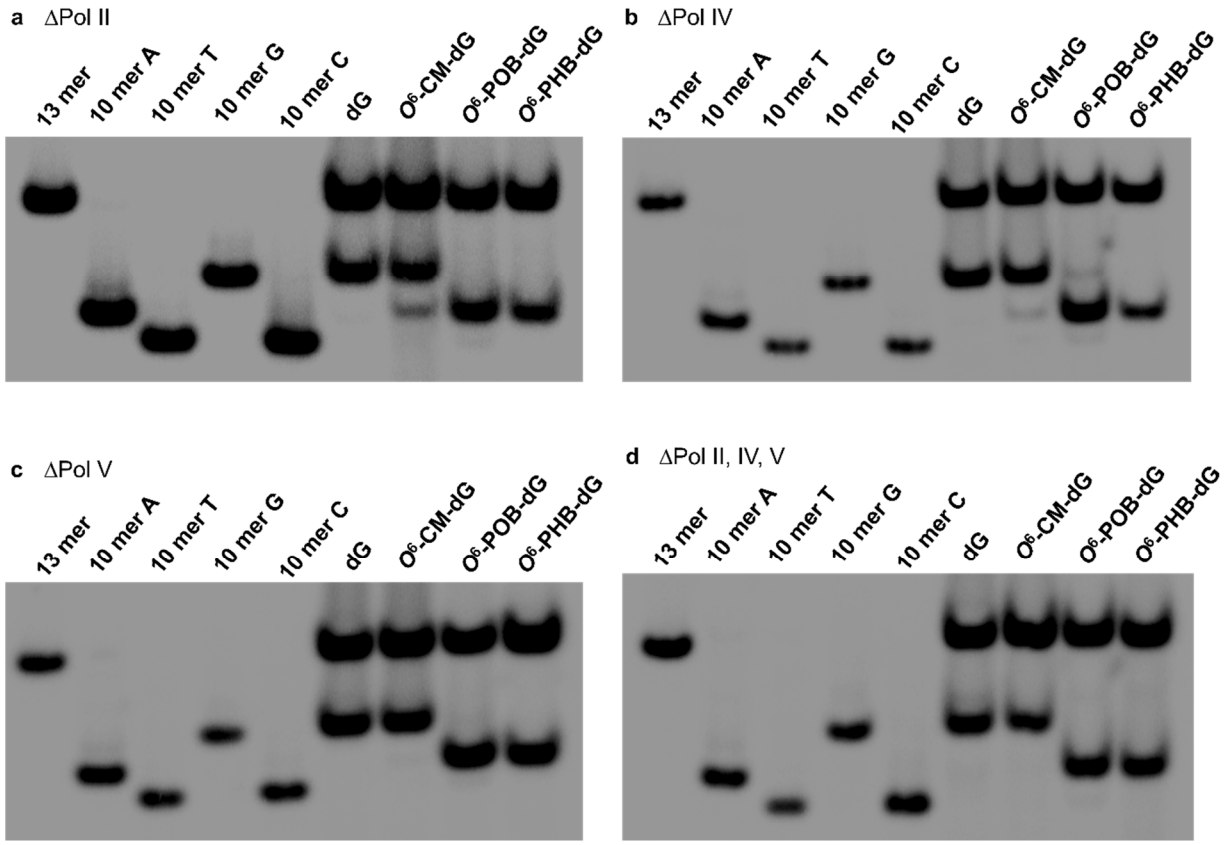


Figure S12. Native PAGE (30%) for monitoring the bypass efficiencies and mutation frequencies of O^6-CM-dG, O^6-POB-dG, and O^6-PHB-dG in SOS-induced AB1157 *E. coli* cells that are deficient in Pol II (a), Pol IV (b), Pol V (c), or all three translesion synthesis polymerases (d). Gel image showing the 13-mer and 10-mer products released from the original lesion-situated strand of the PCR products of the progeny of the competitor genome (13-mer, i.e. 5′-GGCGATAAGCTAT-3′) and the control or lesion-carrying genome (10-mer), where 10mer A, 10mer C, 10mer G, and 10mer T represent the [5′-^{32}P]-labeled standard ODNs 5′-GGCGMGCTAT-3′, with 'M' being A, C, G, and T, respectively.

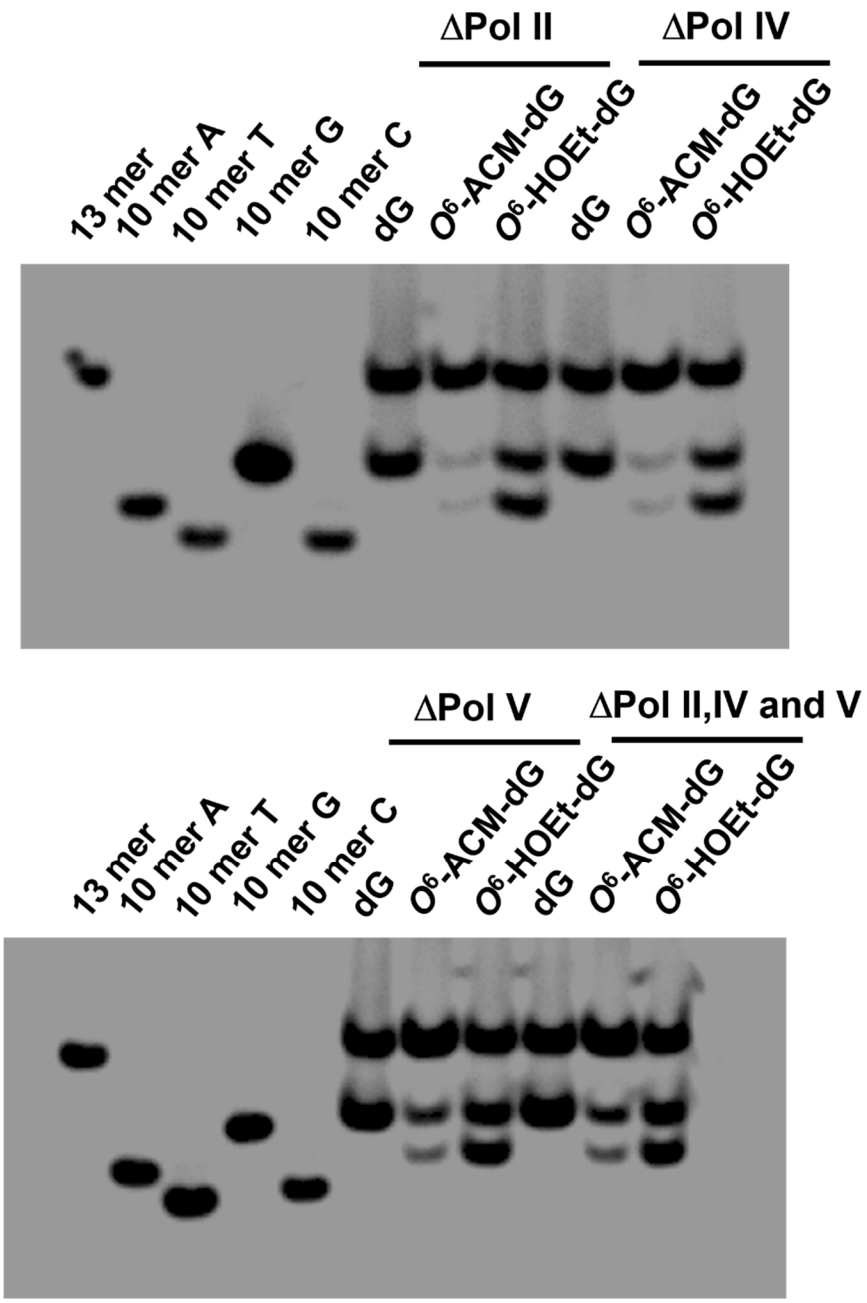


Figure S13. Native PAGE (30%) for monitoring the bypass efficiencies and mutation frequencies of O^6-ACM-dG, and O^6-HOEt-dG in SOS-induced AB1157 *E. coli* cells that are deficient in Pol II, Pol IV, Pol V, or all three translesion synthesis polymerases. Gel image showing the 13-mer and 10-mer products released from the original lesion-situated strand of the PCR products of the progeny of the competitor genome (13-mer, i.e. 5′-GGCGATAAGCTAT-3′) and the control or lesion-carrying genome (10-mer), where 10mer A, 10mer C, 10mer G, and 10mer T represent the [5′-^{32}P]-labeled standard ODNs 5′-GGCGMGCTAT-3′, with 'M' being A, C, G, and T, respectively.

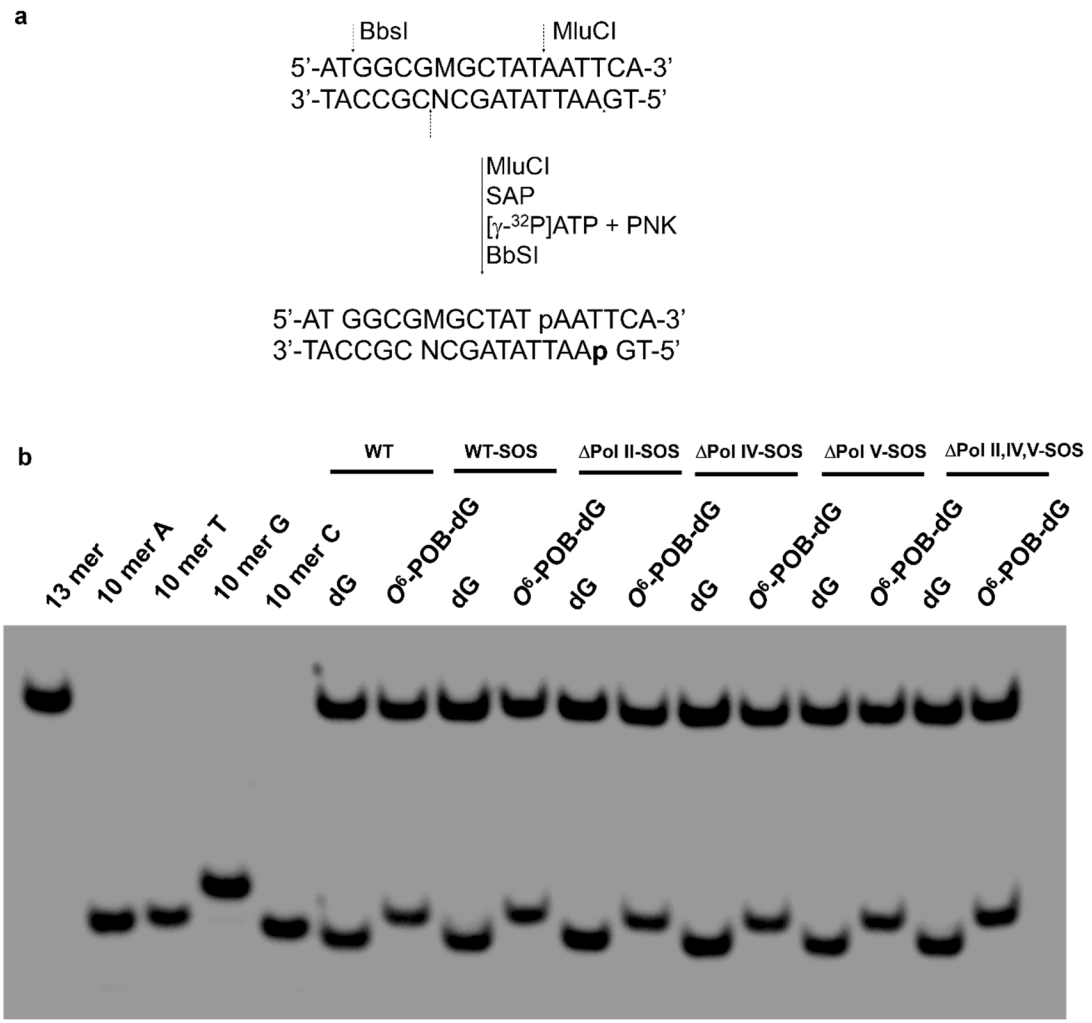


Figure S14. Native PAGE (30%) for monitoring the bypass efficiencies and mutation frequencies of O^6-POB-dG lesions in AB1157 *E. coli* cells that are proficient or deficient in translesion synthesis polymerases. (A) Sequential restriction enzyme digestion and selective radio-labeling of the strand complementary to the strand initially containing the lesion. 'SAP' and 'PNK' designate shrimp alkaline phosphatase and T4 polynucleotide kinase, respectively. (B) Gel image showing the 13-mer and 10-mer radio-labeled restriction fragments released from the original lesion-containing strand of the PCR products of the progeny of the competitor genome and the control or lesion-carrying genome, where 10mer A, 10mer C, 10mer G, and 10mer T represent the [5′-^{32}P]-labeled standard ODNs 5′-AATTATAGCN-3′, with 'N' being A, C, G, and T, respectively.