Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Replication of N^2,3-Ethenoguanine by DNA Polymerases**

*Linlin Zhao, Plamen P. Christov, Ivan D. Kozekov, Matthew G. Pence, Pradeep S. Pallan, Carmelo J. Rizzo, Martin Egli, and F. Peter Guengerich**

anie_201109004_sm_miscellaneous_information.pdf

SUPPORTING INFORMATION

TABLE OF CONTENTS

Scheme S1. Procedure for site-specific synthesis of oligonucleotides containing the stereochemically defined adduct 2'-F-N^2,3-ε-2'-deoxyarabinoguanosine (2'-F-N^2,3-εdG).

Figure S1. Kinetics of depurination of N^2,3-εG

Table S1. Steady-state kinetic parameters of DNA polymerase-catalyzed single-base insertion opposite X in a template sequence of 3'−CCCCCGAGCATTCCTAAGXTACT−5', where X= 2'-F-N^2,3-εdG, 2'-F-dG, or dG.

Table S2. Steady-state kinetic parameters of DNA polymerase-catalyzed single-base insertion opposite X in a template sequence of 3'−CCCCCGAGCATTCCTAAGXCACT−5', where X= 2'-F-N^2,3-εdG, 2'-F-dG, or dG.

Table S3. Products of extension of template-primer complexes containing 2'-F-N^2,3-εdG, 2'-F-dG, and unmodified dG by Dpo4.

Figure S2. LC-MS/MS analysis of full-length extension products by Dpo4 formed with the sequence c (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S4. Observed and theoretical CID fragmentation of *m/z* 934.7 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Table S5. Observed and theoretical CID fragmentation of *m/z* 727.3 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Figure S3. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S6. Observed and theoretical CID fragmentation of *m/z* 942.2 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Table S7. Observed and theoretical CID fragmentation of *m/z* 732.1 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Figure S4. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S8. Observed and theoretical CID fragmentation of *m/z* 946.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Table S9. Observed and theoretical CID fragmentation of *m/z* 954.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Table S10. Observed and theoretical CID fragmentation of *m/z* 790.1 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG).

Figure S5. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S11. Observed and theoretical CID fragmentation of *m/z* 926.7 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG).

Table S12. Observed and theoretical CID fragmentation of *m/z* 721.7 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG).

Figure S6. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S13. Observed and theoretical CID fragmentation of *m/z* 934.2 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG).

Figure S7. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of four dNTPs.

Table S15. Observed and theoretical CID fragmentation of *m/z* 938.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X=2'-F-N^2,3-εdG).

Table S16. Observed and theoretical CID fragmentation of *m/z* 782.1 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X=2'-F-N^2,3-εdG).

Table S17. Crystal data collection and refinement statistics.

Figure S8. Crystal structures of Dpo4•N^2,3-εdG-DNA complex (Z=T in the template) and possible base pairing mechanisms.

Table S18. Comparisons of kinetic parameters of KF^--, Dpo4-, and human pol κ-catalyzed single base insertion opposite two different etheno DNA adducts, X=2'-F-N^2,3-εdG and W=1,N^2-εdG.

Figure S9. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-dG)GAATCCTTACGAGCCCCC-3'.

Figure S10. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-dG)GAATCCTTACGAGCCCCC-3'.

Figure S11. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3'.

Figure S12. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3'.

Figure S13. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'.

Figure S14. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'.

Figure S15. HPLC analysis of enzymatic digest of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'.

Figure S16. HPLC analysis of enzymatic digest of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'.

REFERENCES FOR SUPPORTING INFORMATION SECTION

EXPERIMENTAL PROCEDURES

Materials. All commercial chemicals were of the highest quality available and used without further purification. Klenow fragment exonuclease$^-$ (KF$^-$),[1] bacteriophage T7 DNA polymerase exonuclease$^-$ (T7$^-$),[1] and *Escherichia coli* thioredoxin[2] were expressed and purified previously in this laboratory. Dpo4 was expressed and purified as described previously.[3] The recombinant catalytic fragment of human DNA pol κ (amino acids 19–526)[4] was expressed as an maltose binding protein-DNA pol κ (gift from Prof. F. W. Perrino, Wake Forest University, Winston-Salem, NC) fusion protein as previously reported. Catalytic core (amino acids 1-513) of *Saccharomyces cerevisiae* pol η K140A/S144W was expressed and purified according to the literature.[5] Unlabeled deoxynucleotide triphosphate (dNTP), T4 polynucleotide kinase, uracil DNA glycosylase (UDG), and restriction endonucleases were purchased from New England Biolabs (Ipswich, MA). [γ-^{32}P]ATP (specific activity 3×10^3 Ci/mmol) was purchased from PerkinElmer Life Sciences (Boston, MA). Biospin columns were purchased from Bio-Rad (Hercules, CA). Dideoxynucleotide triphosphate (ddNTP) was from GE healthcare (Piscataway, NJ). Unmodified oligonucleotides were purchased from Midland Certified Reagent Co. (Midland, TX) or Integrated DNA Technologies (Coralville, IA), and were purified using high-performance liquid chromatography by the manufacturers. 2-Amino-9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine was purchased from Metkinen (Finland). The reactions were monitored by thin-layer chromatography (TLC) on Merck silica gel 60 F_{254} plates, with visualization with 254 nm UV light followed by a *p*-anisaldehyde stain.

Oligonucleotide Synthesis, Purification and Characterization

NMR Spectroscopy. ^{1}H, ^{13}C, ^{19}F, and ^{31}P NMR spectra were recorded on Bruker NMR spectrometers operating at 600, 150, 380, and 160 MHz, respectively, in the indicated solvents.

Mass Spectrometry. FAB mass spectra (low and high resolution) were obtained at the Mass Spectrometry Facility at the University of Notre Dame, Notre Dame, IN. MALDI-TOF mass spectra (negative ion) of modified oligonucleotides were obtained on a Voyager Elite DE instrument (Perseptive Biosystems) at the Vanderbilt Mass Spectrometry Facility using a 3-hydroxypicolinic acid (HPA) matrix containing ammonium hydrogen citrate.

2-Amino-9-*(*2-deoxy-2-fluoro-β-D-3,5-*O*-diacetyl-arabinofuranosyl)-guanine (2). 2-Amino-*9*-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-guanine (**1**, 430 mg, 1.5 mmol) was co-evaporated with anhydrous pyridine and dried overnight under high vacuum. The residue was dissolved in dry pyridine (15 mL), and acetic anhydride (1.22 g, 12 mmol, 1.14 mL) was added. The solution was stirred at room temperature for 4 h. The solvent was removed in vacuo (rotary evaporator) and the crude product was purified by flash chromatography on silica gel (95:5 CH_2Cl_2:CH_3OH, v/v) to afford **2** (500 mg, 92%). ^{1}H NMR (DMSO-d_6) ^{1}H δ 10.76 (br s, 1H, NH), 7.71 (d, 1H, J = 3.0, H-8), 6.62 (s, 2H, NH_2), 6.14 (dd, 1H, J = 3.7 Hz, $^3J_{H\text{-}F}$ = 23.3 Hz, H-1'), 5.44 (dd, 1H, J = 3.5 Hz, $^2J_{H\text{-}F}$ = 32.0 Hz, H-2'), 5.36 (d, 1H, J = 3.0 Hz, H-3'), 4.37 (m, 1H, H-4'), 4.27 (m, 2H, H-5'), 2.10 (s, 3H, Ac), 2.03 (s, 3H, Ac); ^{13}C NMR (DMSO-d_6) δ 20.9 (5'-Ac), 21.0 (3'-Ac), 63.2 (C-5'), 75.7 (d, $^2J_{C\text{-}F}$ = 43.0, C-3'), 78.9 (C-4'), 81.9 (d, $^2J_{C\text{-}F}$ = 26, C-1'), 92.8 (d, $^1J_{C\text{-}F}$ = 285, C-2'), 116.2 (C-5), 136.1 (C-8), 151.4 (C-4), 154.6 (C-2), 156.5 (C-6), 169.8 (5'-Ac-C=O), 170.5 (3'-Ac-C=O); HRMS (FAB$^+$) *m/z* calcd for $C_{14}H_{17}FN_5O_6$ $[M + H]^+$ 370.1171, found 370.1157.

2-Amino-*9*-*(*2-deoxy-2-fluoro-β-D-3,5-diacetyl-arabinofuranosyl)-*O*^{6}-(*p*-nitrophenethyl)-guanine (3). *p*-Nitrophenethyl alcohol (443 mg, 2.65 mmol), triphenylphosphine (694 mg, 2.65 mmol) and diisopropyl azodicarboxylate (DIAD, 536 mg, 2.65 mmol, 522 μL) were sequentially added to a stirred solution of **2** (490 mg, 1.32 mmol) in anhydrous 1,4-dioxane (80 mL). The reaction mixture was stirred at 90° C for 2 h. The dioxane was removed in vacuo (rotary evaporator) and the residue was purified by flash chromatography on silica gel (50:50 to 20:80 hexane:ethyl acetate, v/v) to afford the desired product (584 mg, 85%). ^{1}H NMR (DMSO-d_6) δ 8.16 (d, 2H, J = 9.0 Hz, Ph-H), 7.89 (s, 1H, H-8), 7.62 (d, 2H, J = 9.0 Hz, Ph-H), 6.59 (s, 2H, NH_2), 6.27 (dd, 1H, J = 6.3 Hz, $^3J_{H\text{-}F}$ = 19.2 Hz, H-1'), 5.47 (dd, 1H, J = 3.6 Hz, $^2J_{H\text{-}F}$ = 34.8 Hz, H-2'), 5.43 (d, 1H, J = 3.0 Hz, H-3'), 4.67 (t, 2H, J = 6.6 Hz, CH_2-$C\underline{H_2}$-Ph-NO_2), 4.38 (m, 1H, H-4'), 4.28 (m, 2H, H-5'), 3.24 (t, 2H, J = 6.6 Hz, $C\underline{H_2}$-CH_2-Ph-NO_2), 2.11 (s, 3H, Ac), 2.03 (s, 3H, Ac); ^{13}C NMR (DMSO-d_6) δ 20.5 (5'-OAc), 20.6 (3'-OAc), 34.4 (CH_2), 62.9 (CH_2), 65.6 (CH_2), 75.3 (d, C-3', $^2J_{C\text{-}F}$ = 28.0 Hz), 78.34 (C-4'), 81.9 (d, $^2J_{C\text{-}F}$ =16.0 Hz, C-1'), 92.5 (d, $^1J_{C\text{-}F}$ = 187, C-2'), 112.8 (C-5), 123.4 (C-Ar), 130.3 (C-Ar), 138.1 (C-8), 146.3 (C-Ar), 146.6 (C-Ar), 154.0 (C-6), 160.0 (C-2), 160.2 (C-4), 169.5 (5'-Ac-C=O), 170.12 (3'-Ac-C=O). HRMS (FAB$^+$) *m/z* calcd for $C_{22}H_{24}FN_6O_8$ $[M + H]^+$ 519.1648, found 519.1634.

2-Amino-9-*(*2-deoxy-2-fluoro-β-D-arabinofuranosyl)-*O*^{6}-(*p*-nitrophenethyl)-guanine (4). **3** (570 mg, 1.1 mmol) was dissolved in ammonia in CH_3OH (0.2 M, 50 mL) and the reaction mixture was stirred for 8 h at room temperature. The CH_3OH

---

[1] Abbreviations used: BSA, bovine serum alumin; KF$^-$, Klenow fragment (exonuclease$^-$); T7$^-$, bacteriophage T7 DNA polymerase (exonuclease$^-$); Dpo4, *S. Solfataricus* S2 DNA polymerase IV; dNTP, deoxynucleoside triphosphate; UDG, uracil DNA glycosylase; ddNTP, dideoxynucleotide triphosphate; 2'-F-N^2,3-εdG, 2'-fluoro-N^2,3-ethenoarabinoguanosine; 2'-F-dG, 2'-fluoro-2'-deoxyarabinoguanosine; dG, 2'-deoxyguanosine; ddC, 2',3'-dideoxycytidine; dU, 2'-deoxyuracil; UPLC, ultraperformance liquid chromatography.

was removed in vacuo (rotary evaporator) and the residue was purified by flash chromatography on silica gel (98:2 to 96:4 CH_2Cl_2:CH_3OH) to afford **4** (410 mg, 86%). ^{1}H NMR (DMSO-d_6) δ 8.16 (d, 2H, J = 9.0 Hz, Ph-H), 7.95 (s, 1H, H-8), 7.61 (d, 2H, J = 9.0 Hz, Ph-H), 6.53 (s, 2H, NH_2), 6.23 (dd, 1H, J= 4.2 Hz, $^3J_{H\text{-}F}$ = 15.6 Hz, H-1'), 5.93 (d, 1H, J = 4.8 Hz, 3'-OH), 5.12 (dd, 1H, J= 3.8 Hz, $^2J_{H\text{-}F}$ = 32 Hz, H-2'), 5.06 (d, 1H, J = 3.0, 5'-OH), 4.66 (t, 2H, J = 6.6 Hz, CH_2-C$\underline{H_2}$-Ph-NO_2), 4.38 (m, 1H, H-4'), 3.61 (m, 2H, H-5'), 3.24 (t, 2H, J = 6.6 Hz, C$\underline{H_2}$-CH_2-Ph-NO_2). ^{13}C NMR (DMSO-d_6) δ 34.4 (CH_2), 60.4 (CH_2), 65.6 (C-5'), 73.7 (d, $^2J_{C\text{-}F}$ = 24.0 Hz, C-3'), 81.2 (d, $^2J_{C\text{-}F}$ = 16.5 Hz, C-1'), 83.6 (d, $^3J_{C\text{-}F}$ = 4.2 Hz, C-4'), 95.2 (d, $^1J_{C\text{-}F}$ = 190 Hz, C-2'), 113.0 (C-5), 123.4 (C-Ar), 130.3 (C-Ar), 138.1 (C-8), 146.9 (C-Ar), 146.7 (C-Ar), 153.9 (C-6), 160.0 (C-2), 160.1 (C-4). HRMS (FAB$^+$) *m/z* calcd for $C_{18}H_{20}FN_6O_6$ [M + H]$^+$ 435.1435, found 435.1423.

***9-(*2-Deoxy-2-fluoro-β-D-arabinofuranosyl]-*O*^{6}-(*p*-nitrophenethyl)-*N*^{2},3-ethenoguanine (5)**. An aqueous bromoacetaldehyde solution (10 mL) was prepared by stirring bromoacetaldehyde dimethylacetal (5.72 g, 0.033 mol, 4 mL), ethanol (4 mL), hydrochloric acid (6 M, 4 mL) and water (20 mL) for 3 days at room temperature.[6] The product was added to a solution of **4** (550 mg**,** 1.15 mmol) dissolved in a mixture of CH_3OH (30 mL) and acetate buffer (30 mL, pH 4.5; prepared by mixing 7.5 mL of 0.2 M sodium acetate, 7.5 mL of 0.2 M CH_3CO_2H, and 15 mL of H_2O). The pH of the reaction mixture was adjusted to 4.5 by addition of 3 M NaOH using a pH meter. The reaction mixture was stirred for 48 h at 40° C, then neutralized. The solvents were removed in vacuo (rotary evaporator) and the residue was purified by flash chromatography on silica gel (98:2 to 94:6 CH_2Cl_2:CH_3OH, v/v) to afford **5** (432 mg, 82%). ^{1}H NMR (DMSO-d_6) δ 8.47 (d, 1H, J = 1.8 Hz, (etheno) 5-H), 8.15 (d, 2H, J = 10.0 Hz, Ph-H), 8.09 (s, 1H, H-8), 7.63 (d, 2H, J = 10.0 Hz, Ph-H), 7.45 (d, 1H, J = 1.8 Hz, (etheno) 6-H), 6.75 (dd, 1H, J= 4.8 Hz, $J_{H\text{-}F}$ = 8.4 Hz, H-1'), 5.48 (ddd, 1H, J= 6.0 Hz, J = 12.0 Hz, $J_{H\text{-}F}$= 53 Hz, H-2'), 4.77 (t, 2H, J = 6.6 Hz, CH_2-C$\underline{H_2}$-Ph-NO_2), 4.43 (ddd, 1H, J = 6.0 Hz, J = 12.0 Hz, $J_{H\text{-}F}$ = 18.0 Hz, H-3'), 3.85 (m, 1H, H-4'), 3.69 (m, 2H,, H-5'), 3.30 (t, 2H, J = 6.6 Hz, C$\underline{H_2}$-CH_2-Ph-NO_2). ^{13}C NMR (DMSO-d_6) δ 34.2 (CH_2), 58.8 (CH_2), 65.9 (C-5'), 70.6 (d, C-3', $^2J_{C\text{-}F}$ = 22.0 Hz), 82.6 (C-4'), 83.4 (d, C-1', $^2J_{C\text{-}F}$ = 18.0 Hz), 96.1 (d, $^1J_{C\text{-}F}$ = 192 Hz, C-2'), 109.3 (C-5), 115.8 (etheno-C-2"), 123.4 (C-Ar), 130.3 (C-Ar), 130.3 (etheno-C-1"), 134.9 (C-4), 136.5 (C-8), 146.3 (C-Ar), 146.4 (C-Ar), 146.8 (C-2), 155.0 (C-6). HRMS (FAB$^+$) *m/z* calcd for $C_{20}H_{20}FN_6O_6$ [M+H]$^+$ 459.1417, found 459.1423.

***9-(*2-Deoxy-2-fluoro-β-D-arabinofuranosyl]-*N*^{2},3-ethenoguanine**. 1,8-Diazabicycloundec-7-ene (DBU, 245 mg, 1.74 mmol, 241 μL) was added to a stirred solution of **5** (40 mg, 0.087 mmol) in dry CH_3CN (10 mL), and the reaction mixture was stirred at room temperature for 3 h. The solvent was removed under high vacuum, and the residue was purified by flash chromatography on silica gel (95:4.5:0.5 to 90: 9.5:0.5 CH_3CN:H_2O:NH_4OH) to afford the desired product (22 mg, 81%). ^{1}H NMR (DMSO-d_6) δ 8.33 (s, 1H, (etheno) 5-H), 7.79 (s, 1H, H-8), 7.10 (s, 1H, (etheno) 6-H), 6.61 (dd, 1H, J= 5.4 Hz, H-1'), 5.47 (dd, 1H, J= 5.3 Hz, $J_{H\text{-}F}$ = 53.0 Hz, H-2'), 4.39 (m, 1H, H-3'), 3.86-3.67 (m, 3H, H-4', H-5'). ^{13}C NMR (DMSO-d_6) δ 58.8 (C-5'), 70.5 (d, $^2J_{C\text{-}F}$ = 22 Hz, C-3'), 82.5 (d, $^3J_{C\text{-}F}$ = 7.5 Hz, C-4'), 83.16 (d, $^2J_{C\text{-}F}$ = 18.0 Hz, C-1'), 96.0 (d, $^1J_{C\text{-}F}$ = 192 Hz, C-2'), 109.9 (C-5), 119. (etheno-C-2"), 128.2 (etheno C-1"), 133.3 (C-4), 135.6 (C-8), 142.9 (C-2), 156.0 (C-6). ^{19}F NMR (DMSO-d_6) δ -200.16. HRMS (FAB$^+$) *m/z* calcd for $C_{12}H_{13}FN_5O_4$ [M+H]$^+$ 310.0920, found 310.0946.

5′-*O*-(4,4′-Dimethoxytrityl)-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-*O*^{6}-(*p*-nitrophenethyl)-*N*^{2},3-ethenoguanine (6). ***5*** (80 mg, 0.17 mmol) was co-evaporated with anhydrous pyridine and dried overnight under high vacuum. The residue was dissolved in dry pyridine (20 mL), and DMTrCl (230 mg, 0.68 mmol) was added. The solution was stirred at room temperature for 8 h. The solvent was removed in vacuo (rotary evaporator) and the crude product was purified by flash chromatography on silica gel (98:1:1 to 95:4:1 CH_2Cl_2:CH_3OH:Et_3N) to afford **6** (90 mg, 68%, 15 mg of starting material was recovered). ^{1}H NMR (DMSO-d_6) δ 8.17 (d, 1H, J = 6.0 Hz, (etheno) 5-H), 8.16 (d, 2H, J = 6.0 Hz, Ph-H), 8.1 (s, 1H, H-8), 7.63 (d, 2H, J = 6.0 Hz, Ph-H), 7.36 (d, 1H, J = 6.0 Hz, (etheno) 6-H), 7.29 (m, 9H, Ar), 6.85 (m, 4H, ArH), 6.27 (m, 1H, H-1'), 5.48 (ddd, 1H, J= 6.0 Hz, J = 13.0 Hz, $J_{H\text{-}F}$ = 50 Hz, H-2'), 4.77 (t, 2H, J = 6.8 Hz, C$\underline{H_2}$-CH_2-Ph-NO_2), 4.56 (ddd, 1H, J = 6.0 Hz, J = 12.0 Hz, $J_{H\text{-}F}$= 18.0 Hz, H-3'), 4.06 (m, 1H, H-4'), 3.71 (s, 6H, OCH_3), 3.36 (m, 2H, H-5'), 3.30 (t, 2H, J = 6.8 Hz, C$\underline{H_2}$-CH_2-Ph-NO_2). ^{13}C NMR (DMSO-d_6) δ 34.1 (CH_2), 55.0 (OCH_3), 61.9 (CH_2), 59.9 (C-5'), 72.2 (d, $^2J_{C\text{-}F}$ = 23.0 Hz, C-3'), 81.23 (C-4'), 83.5 (d, $^2J_{C\text{-}F}$ = 17.0 Hz, C-1'), 85.9 ($\underline{C}Ph_3$), 96.2 (d, $^1J_{C\text{-}F}$ = 190 Hz, C-2'), 109.6 (C-5), 113.2 (Ph), 113.2 (etheno-C-2"), 123.5 (C-Ar, C-Ph), 126.8 (C-Ph), 127.9 (C-Ph), 127.9 (C-Ph), 129.7 (C-Ar), 130.3 (C-Ph), 131.3 (etheno C-1"), 135.0 (C-Ph), 135.6 (C-4), 136.6 (C-8), 144.6 (C-Ph), 146.3 (C-Ar), 146.4 (C-Ar), 146.8 (C-2), 156.0 (C-Ph), 158.1 (C-6). HRMS (FAB$^+$) *m/z* calcd for $C_{41}H_{38}FN_6O_8$ [M+H]$^+$ 761.2768, found 761.2730.

3′-*O*-[*N,N*-Diisopropylamino)-2-cyanoethoxyphosphinyl]-5′-*O*-(4,4′-dimethoxytrityl) -*9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl]-*O*^{6}-(*p*-nitrophenethyl)-*N*^{2},3-ethenoguanine (7). **6** (82 mg, 0.11 mmol) was evaporated from anhydrous pyridine (3 × 5 mL) and dried overnight under vacuum. The oily residue was dissolved in freshly distilled CH_2Cl_2 (25 mL) and a 0.45 M solution of 1*H*-tetrazole in CH_3CN (11 mg, 0.15 mmol, 330 mL) and 2-cyanoethyl-*N,N,N′,N′*-tetraisopropyl-phosphoramidite (52 mg, 0.174 mmol, 55 mL) were added. The solvent was removed in vacuo and the residue purified by flash column chromatography on silica gel (98:1:1 to 95:4:1 CH_2Cl_2:CH_3OH:Et_3N, v/v/v) to afford the desired product (60 mg, 57 %). ^{1}H NMR (CD_2Cl_2) δ 8.18 (d, 1H, (etheno) 5-H, J = 8 Hz) 8.00 (s, 1H, H-8), 7.96 (m, 2H, Ph-H), 7.58 (d, 1H, J = 8.0 Hz, (etheno) 6-H), 7.48 (m, 2H, Ph-H), 7.29 (m, 9H, Ar), 6.85 (m, 4H, ArH), 6.4 (m, 1H, H-1'), 5.5 (dt, 1H, J= 6.0 Hz, $J_{H\text{-}F}$= 50.0 Hz, H-2'), 4.90 (t, 2H, J= 9.0 Hz, CH_2-C$\underline{H_2}$-Ph-NO_2), 4.24 (m, 1H, H-3'), 4.22 (m, 1H, H-4'), 3.36 (m, 2H, H-5'), 3.89-3.85 (m, 2H, $POCH_2$), 3.80 (s, 6H, OCH_3), 3.63-3.66 (m, 2H, isopropyl CH), 3.37-3.36 (m, 2H, H-5') 2.63 (t, 1H, J = 6.0 Hz,

CH_2CN), 2.59 (t, 1H, J = 6.0 Hz, CH_2CN), 1.21-1.12 (m, 12H, isopropyl CH_3). ^{13}C NMR (CD_2Cl_2) δ 19.9, 19.9, 20.0, 23.9, 23.9, 24.0, 24.1, 34.5, 43.0, 43.1, 43.2, 57.9, 58.0, 58.1, 58.2, 61.2, 61.5, 82.8, 83.2, 84.3 (d, $^2J_{C\text{-}F}$ = 17.0 Hz, C-3'), 84.7 (d, $^2J_{C\text{-}F}$ = 15.0 Hz, C-2'), 86.3, 94.6 ($^1J_{C\text{-}F}$ = 194 Hz, C-2'), 107.7, 112.8, 116.8, 123.2, 126.6, 127.3, 127.5, 127.7, 127.7, 129.6, 129.7, 129.7, 129.7, 131.9, 134.7, 134.9, 135.0, 135.8, 144.1, 144.1, 145.9, 146.4, 147.1, 155.3, 158.4. ^{31}P-NMR (CD_2Cl_2) 152.4, 152.2; HRMS (FAB^+) *m/z* calcd for $C_{50}H_{55}FN_8O_9P$ $[M + H]^+$ 961.3801, found 961.3808.

***N*^{2}-[(Dimethylamino)methylene]-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine.** 2-Amino-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine (100 mg, 0.35 mmol) was co-evaporated with anhydrous pyridine and dried overnight under high vacuum. Dry methanol (20 mL) and *N*,*N*-dimethylformamide dimethyl acetal (334 mg, 2.8 mmol, 372 µL) were added. The reaction was stirred at 70° C for 6 h, after which time the solvent was removed in vacuo (rotary evaporator). The crude product was purified by flash chromatography on silica gel (95:5 CH_2Cl_2:CH_3OH) to afford the desired product (107 mg, 90%). ^1H NMR (DMSO-d_6) δ 12 (br s, 1H, NH), 8.55 (s, 1H, N=CH), 7.87 (s, 1H, H-8), 6.28 (dd, 1H, J = 6.0 Hz, $J_{H\text{-}F}$ = 16.0 Hz, H-1'), 5.70 (d, 1H, J = 3.0 Hz, 3'-OH), 5.21 (dd, 1H, J = 6.0 Hz, $^2J_{H\text{-}F}$ = 54.0 Hz, H-2'), 4.42 (ddd, 1H, J = 6.0 Hz, ^2J = 12.0 Hz, $^3J_{H\text{-}F}$ = 18.0 Hz, H-3'), 3.78 (m, 1H, J = 6.0 Hz, H-4'), 3.62 (m, 2H, H-5'), 3.24 (s, 3H, CH_3), 3.00 (s, 3H, CH_3). ^{13}C NMR (DMSO-d_6) δ 35.8, 40.6, 63.4, 72.8 (d, $^2J_{C\text{-}F}$ = 22.0, C-3'), 81.1 (d, $^2J_{C\text{-}F}$ = 16.0 Hz, C-1'), 83.5, 95.4 (d, $^1J_{C\text{-}F}$ = 190, C-2'), 119.0, 137.0, 149.9, 158.0, 162.3. HRMS (FAB^+) *m/z* calcd for $C_{13}H_{18}FN_6O_4$ $[M + H]^+$ 341.1352, found 341.1368.

***N*^{2}-[(Dimethylamino)methylene]-5′-*O*-(4,4′-dimethoxytrityl)-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine.** *N*^{2}-[(Dimethylamino)methylene]-*9*-(2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine (80 mg, 0.23 mmol) was co-evaporated with anhydrous pyridine and dried overnight under high vacuum. The residue was dissolved in dry pyridine (20 mL) and DMTrCl (477 mg, 1.41 mmol) was added. The solution was stirred at room temperature for 6 h. The solvent was removed in vacuo (rotary evaporator) and the crude product was purified by flash chromatography on silica gel (98:1:1 to 95:4:1 CH_2Cl_2:CH_3OH:Et_3N, v/v/v) to afford the desired product (121 mg, 82%). ^1H NMR (DMSO-d_6) δ 11.39 (br s, 1H, NH), 8.55 (s, 1H, N=CH), 7.72 (d, 1H, J = 2.4, H-8), 7.38 (m, 2H), 7.25 (m, 7H), 6.85 (m, 4H), 6.35 (dd, 1H, J = 4.2 Hz, $^3J_{H\text{-}F}$ = 16 Hz, H-1'), 6.10 (s, 1H, 3'-OH), 5.21 (dd, 1H, J = 3.0 Hz, $^2J_{H\text{-}F}$ = 52.0 Hz, H-2'), 4.42 (ddd, 1H, J = 6.0 Hz, J = 12.0 Hz, $^3J_{H\text{-}F}$ = 24.0 Hz, H-3'), 4.00 (m, 1H, H-4'), 3.72 (s, 3H, OCH_3), 3.23 (m, 2H, H-5'), 3.12 (s, 3H, CH_3), 3.02 (s, 3H, CH_3). ^{13}C NMR (DMSO-d_6) δ 34.7, 55.1, 63.5, 73.8 (d, $^2J_{C\text{-}F}$ = 24.0 Hz, C- 2'), 81.0 (d, $^2J_{C\text{-}F}$ = 16.0 Hz, C-3'), 92.3 (d, $^1J_{C\text{-}F}$ = 191 Hz, C-2'), 131.2, 118.9, 126.7, 127.7, 127.8, 129.7, 129.7, 135.4, 135.5, 137.0, 144.8, 149.8, 157.6, 158.1. HRMS (FAB^+) *m/z* calcd for $C_{34}H_{36}FN_6O_6$ $[M+H]^+$ 643.2667, found 643.2675.

***N*^{2}-[(Dimethylamino)-methylene]-3′-*O*-[*N*,*N*-diisopropylamino)-2-cyanoethoxyphosphinyl]-5′-*O*-(4,4′-dimethoxytrityl)-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine**. *N*^{2}-[(Dimethylamino)methylene]-5′-*O*-(4,4′-dimethoxytrityl)-*9-(*2-deoxy-2-fluoro-β-D-arabinofuranosyl]-guanine (100 mg, 0.155 mmol) was co-evaporated with anhydrous pyridine and dried overnight under high vacuum. The gummy residue was dissolved in dry CH_2Cl_2 (20 mL) and a solution of anhydrous 1*H*-tetrazole (21.70 mg, 0.31 mmol) and 2-cyanoethyl-*N*,*N*,*N*',*N*'-tetraisopropylphosphorodiamidite (137.71 mg, 0.31 mmol) were added simultaneously. The solution was stirred at room temperature for 4 h. The solvent was removed in vacuo (rotary evaporator) and the crude product was purified by flash chromatography on silica gel (98:1:1 to 95:4:1 CH_2Cl_2:CH_3OH:Et_3N, v/v/v) to afford the desired product (83 mg, 65%). ^1H NMR (CD_2Cl_2) δ 9.40 (br s,1H, NH), 8.63 (s, 1H, N=CH), 7.76 (s, 1H, H-8), 7.48 (m, 2H), 7.31 (m, 7H), 6.84 (m, 4H), 6.35 (dd, 1H, J = 4.0 Hz, $^3J_{H\text{-}F}$ = 22.0 Hz, H-1'), 5.10 (dd, 1H, $^2J_{H\text{-}F}$ = 51.0 Hz, H-2'), 4.62 (m, 1H, H-3'), 4.22 (m, 1H, H-4'), 3.78-3.66 (m, 2H, $POCH_2$), 3.85 (s, 6H, OCH_3), 3.63-3.61 (m, 2H, isopropyl CH), 3.37-3.36 (m, 2H, H-5'), 3.15 (s, 3H, CH_3-N-CH=), 3.10 (s, 3H, CH_3-N-CH=), 2.61 (t, 1H, J = 6.0 Hz, CH_2-CN), 2.47 (t, 1H, J = 6.0 Hz, CH_2-CN), 1.21-1.12 (m, 12H, isopropyl CH_3). ^{13}C NMR (CD_2Cl_2) δ 19.8, 19.9, 19.9, 23.9, 34.6, 40.8, 42.9, 42.9, 42.9, 43.0, 54.8, 58.0, 58.2, 58.2, 58.3, 82.0 ($^2J_{C\text{-}F}$ = 21.0 Hz, C-3'), 82.6 ($^2J_{C\text{-}F}$ = 17.0 Hz, C-1'), 85.8, 85.9, 86.0, 94.1 ($^1J_{C\text{-}F}$ = 193 Hz, C-2'). 112.7, 117.1, 117.2, 119.0, 123.3, 126.4, 126.4, 127.5, 127.6, 127.7, 129.6, 135.2, 135.3, 137.0, 144.4, 149.3, 149.9, 149.8, 157.5, 157.8. ^{31}P-NMR (CD_2Cl_2) 152.2, 152.0; HRMS (FAB^+) *m/z* calcd for $C_{43}H_{53}FN_8O_7P$ $[M+H]^+$ 843.3768, found 843.3753.

Oligonucleotide Synthesis. The oligodeoxyribonucleotides were synthesized on a Perseptive Biosystems Model 8909 DNA synthesizer on a 1-µmol scale using Expedite reagents (Glen Research, Sterling, Virginia) with the standard synthetic protocol for the coupling of the unmodified bases. The coupling of the 2'-F-*N*^{2},3-ε-deoxyguanosine phosphoroamidite was performed off–line for 30 min as previously described.[7] The remainder of the synthesis was performed on–line using standard protocols. The DMTr group of the last base was retained. The beads were treated with DBU (0.5 M, dry acetonitrile) at 45° C for overnight to remove the *O*^{6}-*p*-nitrophenethyl group. The liquid was decanted and the beads were washed with CH_3OH (3 × 1 mL). The oligonucleotides were further deprotected and removed from the solid support with concentrated ammonia at 45° C for 1 h. The ammonia was removed on a centripetal evaporator and the oligonucleotides were purified by reversed-phase HPLC. The lyophilized oligonucleotides were dissolved in H_2O (0.3 mL) and 2% (v/v) CF_3CO_2H (1.6 mL) was added. After 15 min, the mixture was neutralized, filtered through filter (Millex-GV®, 0.22 µm), and desalted. The coupling of the 2'-F-deoxyguanosine phosphoroamidite was performed off-line for 30 min. The remainder of the synthesis was performed on–line using standard protocols. The modified oligodeoxynucleotides were cleaved from the solid support and the exocyclic amino groups were deprotected in a single step using 0.1 M NaOH at room temperature overnight.

5'-TCAC(2'-F-dG)GAATCCTTACGAGCCCCC-3' Purified by reversed-phase HPLC using gradient *2*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^–$, 6946.6; found 6946.0. (Fig S9)

5'-TCAT(2'-F-dG)GAATCCTTACGAGCCCCC-3' Purified by reversed-phase HPLC using gradient *2*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^-$, 6962.5; found 6963.5. (Fig S10)

5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3' Purified by reversed-phase HPLC using gradient *3*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^-$, 6971.5; found 6970.4. (Fig S11)

5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3' Purified by reversed-phase HPLC using gradient *3*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^-$, 6986.5; found 6985.6. (Fig S12)

5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3' Purified by reversed-phase HPLC using gradient *3*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^-$, 5412.5; found 5412.1. (Fig. S13)

5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3' Purified by reversed-phase HPLC using gradient *3*. MALDI-TOF MS (HPA) *m/z* calcd for $[M–H]^-$, 5397.5; found 5397.5.(Fig. S14)

HPLC Purification. A YMC ODS-AQ column (250 mm × 4.6 mm, flow rate 1.5 mL/min, 250 mm × 4.6 mm, flow rate 10 mL/min) or Phenomenex Gemini-C18 column (250 mm × 4.6 mm, flow rate 1.5 mL/min, 250 mm × 4.6 mm, flow rate 5 mL/min) was used to monitor reactions and for oligonucleotide purifications. Oligonucleotides were detected by their UV absorbance at 254 nm. The mobile phase consisted of acetonitrile and 100 mM ammonium formate buffer, pH 3.5.

Gradient 1: initial conditions were 1% CH_3CN (v/v); a linear gradient to 10% CH_3CN (v/v) over 15 min; a linear gradient to 20% CH_3CN (v/v) over 5 min; isocratic at 20% CH_3CN (v/v) for 5 min; a linear gradient to 80% CH_3CN (v/v) over 3 min; isocratic at 80% CH_3CN (v/v) for 2 min; then a linear gradient to the initial conditions over 3 min.

Gradient 2: initial conditions were 1% CH_3CN (v/v); a linear gradient to 7% CH_3CN (v/v) over 20 min; isocratic at 7% CH_3CN (v/v) for 5 min; a linear gradient to 80% CH_3CN (v/v) over 3 min; isocratic at 80% CH_3CN (v/v) for 3 min; then a linear gradient to the initial conditions over 3 min.

Gradient 3: initial conditions were 5% CH_3CN (v/v); a linear gradient to 50% CH_3CN (v/v) over 20 min; a linear gradient to 80% CH_3CN (v/v) over 3 min; isocratic at 80% CH_3CN (v/v) for 2 min; then a linear gradient to the initial conditions over 3 min.

Enzymatic Digestion and Analysis of Oligonucleotides. The enzymatic digestion of oligonucleotides was carried out in a single step as follows: the oligonucleotide (0.5 A_{260} units) was dissolved in 70 μL of buffer (10 mM Tris-HCl pH 7.0, containing 10 mM $MgCl_2$). DNase I (5 units), alkaline phosphatase (1.7 units), and snake venom phosphodiesterase I, type II (0.02 units) were added and the solution was incubated at 37 °C overnight. HPLC analysis was performed using solvent gradient 1. The results are shown in Figs. S15, S16.

Steady-state Kinetics. Unless indicated otherwise, all polymerase reactions were performed at 37 °C in 50 mM Tris-HCl buffer, pH 8.0, containing 5% glycerol (v/v), 5 mM DTT, 50 mM NaCl, 5 mM $MgCl_2$ and 50 mg of bovine serum albumin (BSA) mL^{-1}. In the case of pol $T7^-$, *E. coli* thioredoxin was added in a 20-fold molar excess of pol T7. A primer-template DNA complex containing ^{32}P-labeled primer and an unmodified or adducted template had a final concentration of 60 nM in the reaction and different polymerases were added, ranging from 2 to 30 nM to achieve the steady-state kinetics of dNTP insertion. Reactions were initiated by adding 1 μL dNTP stock solution at each of eight different concentrations to an aliquot of 4 μL of DNA complex-polymerase mixture. Aliquots (1.5 μL) of reaction mixtures, taken at two different time points, were quenched with 9 μL of 20 mM EDTA (pH 9.0) in 95% formamide (v/v). Products were separated using 20% acrylamide (w/v) electrophoresis gels, and results were visualized using a phosphorImaging system (Bio-Rad, Molecular Imager® FX) and analyzed by Quantity One software as previously described.[3] The 23-mer template used in the steady-state kinetic analysis is 5'-TCAZXGAATCCTTACGAGCCCCC-3', Z=T or C, X=2'-F-N^2,3-εdG, 2'-F-dG, or dG; the primer sequence is 5'-GGGGGCTCGTAAGGATTC-3'.

LC-MS Analysis of Full-length Extension Product by Dpo4. An 18-mer primer 5'-GGGGGCTCGTAAGGAT(dU)C-3' was annealed to the same 23-mer oligomer as described above at a 1:1 molar ratio. Reaction conditions were similar to those used in steady-state kinetic assays except the final concentrations are as follows, Dpo4, 1 μM; primer-template complex, 12.5 μM; glycerol, 2% (v/v) in a total volume of 80 μL. Reactions were carried out in the presence of four dNTPs (1 mM each) for 3.5 h at 37 °C. The reactions were terminated by spin-column separations to extract dNTP and Mg^{2+}, and the resulting product was treated with 50 units of UDG and 0.25 M piperidine following a previous protocol.[3]

LC-MS/MS analyses were performed on an Acquity ultraperformance liquid chromatography (UPLC) system (Waters Corp.) connected to a Finnigan LTQ mass spectrometer (Thermo Scientific Corp., San Jose, CA), operating in the ESI negative ion mode and using an Acquity UPLC system BEH octadecylsilane (C18) column (1.7 μm, 1.0 mm × 100 mm). UPLC conditions were modified from previous procedures:[8] buffer A contained 10 mM $NH_4CH_3CO_2$ plus 2% CH_3CN (v/v), and buffer B contained 10 mM $NH_4CH_3CO_2$ plus 95% CH_3CN (v/v). The following gradient program was used with a flow rate of 150 μL min^{-1}: 0–3 min, linear gradient from 100% A to 97% A/3% B (v/v); 3–5 min, linear gradient to 80% A/20% B (v/v); 5–6 min, linear gradient to 100% B; 6–8 min, hold at 100% B; 8–10 min, linear gradient to 100% A; 11–13 min, hold at 100% A. A 20 μL aliquot was injected with an autosampler system, and the column temperature was maintained at 50 °C. ESI conditions were as follows: source

voltage 4 kV, source current 100 μA, auxiliary gas flow rate setting 20, sweep gas flow rate setting 5, sheath gas flow setting 34, capillary voltage –49 V, capillary temperature 350 °C, tube lens voltage –90 V. MS/MS conditions were as follows: normalized collision energy 35%, activation Q 0.250, and activation time 30 ms. Product ion spectra were acquired over the range *m/z* 300–2000. The most abundant species (−2 or −3 charged) was used for CID analysis; calculation of the CID fragmentation of oligonucleotide sequence was done using a program linked to the Mass Spectrometry Group of Medicinal Chemistry at the University of Utah.[9] The relative yields of various products were calculated based on the peak areas of extracted ion chromatograms from LC-MS analyses. The sum of the peak areas was used for multi-charged species.

Crystallization of Dpo4-DNA Complexes. The 18-mer oligomers used for the crystal structures were 5'-TCAZXGAATCCTTCCCCC-3', X=2'-F-*N*^{2},3-εdG, Z=T or C. The 13-mer primer used was 5'-GGGGGAAGGATTY-3', Y= dC or ddC. Crystallization conditions were similar to conditions reported previously in this laboratory[3] and others.[10] Ternary complexes Dpo4-1 (X=2'-F-*N*^{2},3-εdG, Z=C, Y=ddC) and Dpo4-2 (X=2'-F-*N*^{2},3-εdG, Z=T, Y=ddC) were obtained by incubating 200 μM Dpo4, 240 μM primer-template DNA complex, 5 mM $MgCl_2$, 1 mM dCTP, 20 mM Tris-HCl (pH 7.4), 60 mM NaCl, 2% glycerol (v/v), and 5 mM β-mercaptoethanol at 37 °C for 5 min prior to setting up crystal trays. Binary complex Dpo4-3 (X=2'-F-*N*^{2},3-εdG, Z=C, Y=C) was obtained under similar conditions, except that the concentrations of Dpo4 was 220 μM and the DNA complex 260 μM, and 1 mM ddTTP was used as the incoming nucleotide. Crystals were grown using the hanging-drop vapor diffusion method at 20 °C. Droplets (a 1:1 mixture of Dpo4-DNA complex mixture and the reservoir solution, v/v) were equilibrated against a reservoir solution containing 0.1M Tris-HCl (pH 7.4), 15% polyethylene glycol 3350 (w/v), 0.1 M $Ca(CH_3COO)_2$, and 2% glycerol (v/v).

X-ray Diffraction Data Collection and Processing. Crystals generally appeared within a week using the described conditions. Crystals were swiped through mother liquor containing 25% polyethylene glycol 3350 (w/v), 15% ethylene glycol (v/v) and 1 mM dCTP (Dpo4-1 and Dpo4-2) or ddTTP (Dpo4-3) as cryoprotectant and flash frozen in a stream of liquid N_2. X-ray diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) on the 21-ID-D and 21-ID-G (LS-CAT) beam lines. All data sets were recorded from cryoprotected crystals using a wavelength of 0.98 Å at 110 K, with individual diffraction resolution summarized in Table S17. Both Dpo4-1 and Dpo4-2 diffracted to 2.3 Å resolution; repeated attempts to grow and diffract Dpo4-3 crystals only yielded 3.5 Å resolution. Individual frames were indexed and scaled with the program HKL2000.[11] All three complex crystals belong to space group $P2_1$. X-ray diffraction data collection and processing statistics are listed in Table S17.

Structure Determination and Refinement. A previously refined Dpo4-DNA complex (PDB accession code 2bq0)[3] minus solvent molecules and metal ions was used as the starting model for Dpo4-1. The refined Dpo4-1 structure was used as a model for Dpo4-2 and Dpo4-3. Molecular Molecular replacement was performed using MOLREP as a part of the CCP4 program suite,[12] followed by rigid body refinement in Refmac 6.0.[13] Repeated manual model rebuilding and refinement were carried out using the program COOT[14] and restrained and TLS refinement were done in Refmac 6.0.[13] In the Dpo4-1 and Dpo4-2 structures, the 3'-G of the 2'-F-*N*^{2},3-εdG was flipped to a *syn* conformation, based on the positive density in the (*2Fo−Fc*) Fourier difference electron density map, and the close distance between the 3'-G and the pairing C in the primer when the 3'-G was modeled in an *anti* conformation. All three structures contained two Dpo4-DNA complexes in one asymmetric unit, and generally a more complete electron density map was observed in one copy than the other. The figures were prepared using PyMOL.[15]

Scheme S1. Procedure for site-specific synthesis of oligonucleotides containing the stereochemically defined adduct 2'-F-N^2,3-ε-2'-deoxyarabinoguanosine (2'-F-N^2,3-εdG).

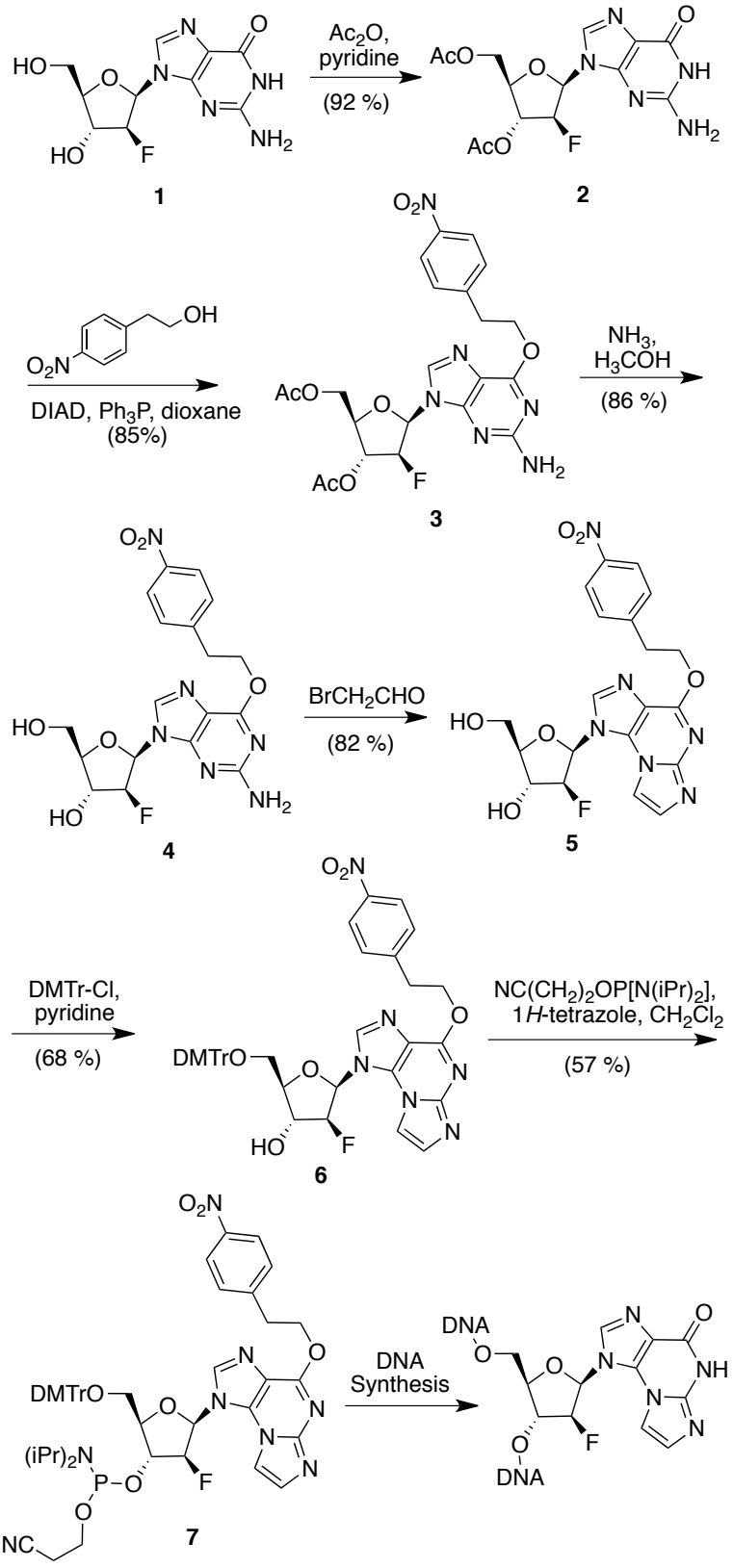


Figure S1. Kinetics of depurination of N^2,3-εG at pH 7.0, 37 °C from the 23-mer oligomer 3'−CCCCCGAGCATTCCTAAGXCACT−P^{32}, where X is 2'-F-N^2,3-εdG. a) The oligomer is single-stranded and b) the oligomer is annealed with a 23-mer complementary DNA strand with C opposite 2'-F-N^2,3-εdG. Plots where generated using the relationship $\ln[A_t/A_0]=-kt$, where A_t and A_0 are the amounts of intact oligomer quantified by Quantity One™ software in phosphorimages. Half-lives ($t_{1/2} = 0.693/k$) calculated from three experiments were 23 ± 4 days for single-stranded and 33 ± 6 days for double-stranded oligonucleotides.

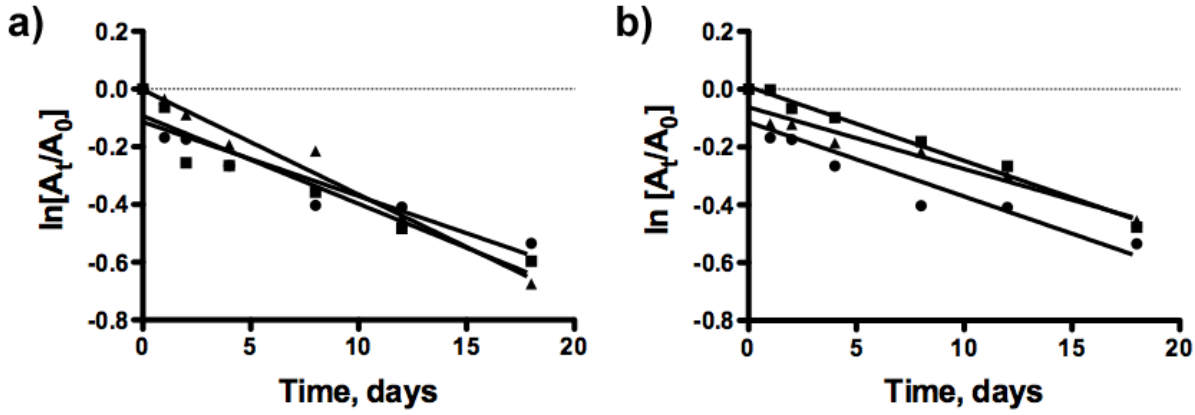


Table S1. Steady-state kinetic parameters of DNA polymerase-catalyzed single-base insertion opposite X in a template sequence of 3'−CCCCCGAGCATTCCTAAGXTACT−5', where X is 2'-F-N^2,3-εdG, 2'-fluoro-2'-deoxyarabinoguanosine (2'-F-dG), or 2'-deoxyguanosine (dG).

Template	dNTP	k_{cat} (min^{-1})	$K_{M,dNTP}$ (μM)	$k_{cat}/K_{M,dNTP}$ (min^{-1} μM^{-1})	f[a]
Dpo4					
2'-F-N^2,3-εdG	A	0.060 ± 0.001	52 ± 5	0.0012	0.048
	G	0.048 ± 0.001	55 ± 6	0.00087	0.035
	T	0.52 ± 0.03	96 ± 16	0.0054	0.22
	C	0.37 ± 0.04	15 ± 5	0.025	
2'-F-dG	A	0.10 ± 0.001	110 ± 10	0.00091	0.0014
	G	0.14 ± 0.006	160 ± 21	0.00088	0.0014
	T	0.42 ± 0.04	100 ± 29	0.0042	0.0067
	C	0.63 ± 0.08	1.0 ± 0.02	0.63	
dG	A	0.13 ± 0.003	640 ± 50	0.00020	0.0011
	G	0.16 ± 0.01	160 ± 19	0.0010	0.0056
	T	0.27 ± 0.01	130 ± 16	0.0021	0.012
	C	1.40 ± 0.03	7.7 ± 1.0	0.18	
Human pol κ					
2'-F-N^2,3-εdG	A	0.063 ± 0.002	55 ± 7	0.0011	0.05
	G	0.22 ± 0.01	210 ± 22	0.0010	0.045
	T	0.90 ± 0.04	110 ± 14	0.0082	0.37
	C	1.6 ± 0.1	73 ± 13	0.022	
2'-F-dG	A	0.42 ± 0.02	150 ± 23	0.0028	0.0041
	G	0.032 ± 0.005	350 ± 110	0.00091	0.0013
	T	0.75 ± 0.09	640 ± 140	0.0012	0.0018
	C	1.9 ± 0.1	2.8 ± 0.3	0.68	
dG	A	0.12 ± 0.02	150 ± 70	0.0008	0.0089
	G	0.21 ± 0.04	940 ± 340	0.00022	0.0024
	T	0.24 ± 0.06	1140 ± 470	0.00021	0.0023
	C	1.8 ± 0.1	20 ± 1	0.090	
Yeast pol η					
2'-F-N^2,3-εdG	A	0.023 ± 0.002	530 ± 210	0.000043	0.010
	G	0.0028 ± 0.0004	860 ± 200	0.0000032	0.0078
	T	0.38 ± 0.015	3300 ± 1400	0.00012	0.29
	C	0.38 ± 0.05	930 ± 210	0.00041	
2'-F-dG	A	0.021 ± 0.001	500 ± 4	0.000042	0.0021
	G	0.0092 ± 0.0007	490 ± 70	0.000019	0.00095
	T	0.31 ± 0.07	2110 ± 640	0.00015	0.0075
	C	0.53 ± 0.02	26 ± 6	0.020	
dG	A	0.0038 ± 0.0004	520 ± 110	0.0000073	0.0013
	G	0.0036 ± 0.07	760 ± 150	0.0000047	0.00081
	T	0.090 ± 0.033	3780 ± 1670	0.000024	0.0041
	C	0.26 ± 0.02	45 ± 8	0.0058	
Pol T7$^-$					
2'-F-N^2,3-εdG	A	0.74 ± 0.06	1000 ± 130	0.00074	0.17

	G	0.079 ± 0.009	1330 ± 220	0..000059	0.013
	T	0.29 ± 0.03	120 ± 20	0.0024	0.55
	C	0.27 ± 0.02	62 ± 9	0.0044	
2'-F-dG	A	0.65 ± 0.16	860 ± 350	0.00076	0.021
	G	0.28 ± 0.08	4370 ± 1120	0.00023	0.0062
	T	0.16 ± 0.05	5040 ± 870	0.000032	0.00086
	C	0.44 ± 0.03	12 ± 2	0.037	
dG	A	0.55 ± 0.04	710 ± 96	0.00077	0.00077
	G	0.51 ± 0.04	800 ± 96	0.00064	0.00064
	T	0.38 ± 0.04	500 ± 110	0.00076	0.00076
	C	1.1 ± 0.04	1.1 ± 0.2	1.0	
KF^{-}					
2'-F-N^2,3-εdG	A	0.88 ± 0.04	51 ± 8	0.017	0.074
	G	0.49 ± 0.02	70 ± 12	0.007	0.030
	T	3.4 ± 0.2	14 ± 3	0.24	1
	C	5.4 ± 0.4	24 ± 5	0.23	
2'-F-dG	A	0.67 ± 0.01	22 ± 1	0.030	0.010
	G	0.23 ± 0.02	260 ± 54	0.00088	0.00030
	T	0.18 ± 0.01	360 ± 57	0.0005	0.00017
	C	4.3 ± 0.4	1.5 ± 0.6	2.9	
dG	A	0.51 ± 0.01	38 ± 3	0.013	0.0054
	G	0.27 ± 0.02	62 ± 15	0.0044	0.0018
	T	0.53 ± 0.01	76 ± 7	0.0070	0.0029
	C	3.6 ± 0.4	1.5 ± 0.6	2.4	

[a] f (misinsertion frequency) = $(k_{cat}/K_{M,dNTP})_{incorrect}/(k_{cat}/K_{M,dNTP})_{correct}$

Table S2. Steady-state kinetic parameters of DNA polymerase-catalyzed single-base insertion opposite X in a template sequence of 3'−CCCCCGAGCATTCCTAAGXCACT−5', where X= 2'-F-N^2,3-ε3-a 2'-F-dG, or dG.

Template	dNTP	k_{cat} (min^{-1})	$K_{M,dNTP}$ (μM)	$k_{cat}/K_{M,dNTP}$ (min^{-1} μM^{-1})	f[a]
Dpo4					
2'-F-N^2,3-εdG	A	0.12 ± 0.01	220 ± 26	0.00054	0.016
	G	0.048 ± 0.001	53 ± 5	0.00091	0.027
	T	0.72 ± 0.03	89 ± 12	0.0081	0.24
	C	0.54 ± 0.05	16 ± 5	0.034	
2'-F-dG	A	0.16 ± 0.01	390 ± 70	0.00041	0.0011
	G	0.073 ± 0.003	53 ± 8	0.0014	0.0036
	T	0.37 ± 0.02	130 ± 15	0.0028	0.0072
	C	0.97 ± 0.02	2.50 ± 0.02	0.39	
dG	A	0.24 ± 0.02	460 ± 70	0.00052	0.0015
	G	0.20 ± 0.01	180 ± 18	0.0011	0.0031
	T	0.53 ± 0.04	230 ± 40	0.0023	0.0065
	C	1.1 ± 0.1	3.1 ± 0.3	0.35	
Human pol κ					
2'-F-N^2,3-εdG	A	0.044 ± 0.002	360 ± 33	0.00012	0.004
	G	0.13 ± 0.006	410 ± 38	0.00032	0.011
	T	0.78 ± 0.04	43 ± 10	0.018	0.60
	C	1.7 ± 0.1	57 ± 8	0.03	
2'-F-dG	A	0.025 ± 0.002	110 ± 28	0.00023	0.001
	G	0.29 ± 0.03	240 ± 65	0.0012	0.0052
	T	0.27 ± 0.03	460 ± 87	0.00059	0.0026
	C	2.7 ± 0.1	12 ± 1	0.23	
dG	A	0.095 ±0.005	71 ± 16	0.0013	0.017
	G	0.36 ± 0.01	330 ± 25	0.0011	0.014
	T	0.11 ± 0.02	200 ± 90	0.00055	0.0071
	C	2.3 ± 0.3	30 ± 8	0.077	
Yeast pol η					
2'-F-N^2,3-εdG	A	0.011 ± 0.002	720 ± 210	0.000015	0.026
	G	0.0019 ± 0.0004	510 ± 200	0.0000037	0.0065
	T	1.5 ± 0.02	17400 ± 500	0.000086	0.15
	C	0.16 ± 0.01	280 ± 57	0.00057	
2'-F-dG	A	0.090 ± 0.011	1530 ± 280	0.000059	0.0045
	G	0.011± 0.003	150 ± 90	0.000073	0.0056
	T	0.31 ± 0.10	4300 ± 1500	0.000072	0.0055
	C	0.21 ± 0.01	16 ± 5	0.013	
dG	A	0.0043 ± 0.0011	530 ± 260	0.0000081	0.0017
	G	0.0060 ± 0.0014	700 ± 280	0.0000085	0.0018
	T	0.050 ± 0.019	2400 ± 800	0.000021	0.0045
	C	0.20 ± 0.02	43 ± 11	0.0047	
Pol T7$^-$					
2'-F-N^2,3-εdG	A	1.4 ± 0.2	2490 ± 450	0.00056	0.56
	G	0.084 ± 0.006	3060 ± 1940	0.000027	0.027

	T	0.20 ± 0.05	210 ± 74	0.00095	0.95
	C	0.26 ± 0.11	260 ± 160	0.001	
2'-F-dG	A	1.7 ± 0.4	1420 ± 530	0.0012	0.036
	G	0.24 ± 0.04	1440 ± 350	0.00017	0.0052
	T	0.30 ± 0.10	4850 ± 1340	0.000062	0.0019
	C	1.1 ± 0.1	33 ± 7	0.033	
dG	A	1.1 ± 0.1	540 ± 50	0.0020	0.0011
	G	0.55 ± 0.07	610 ± 150	0.00090	0.00047
	T	0.82 ± 0.21	1500 ± 550	0.00055	0.00029
	C	1.2 ± 0.04	0.62 ± 0.16	1.9	
KF$^-$					
2'-F-N^2,3-εdG	A	0.92 ± 0.04	61 ± 10	0.015	0.079
	G	1.1 ± 0.05	35 ± 7	0.031	0.16
	T	4.6 ± 0.3	13 ± 3	0.35	1.8
	C	8.9 ± 0.8	47 ± 10	0.19	
2'-F-dG	A	3.3 ± 0.2	70 ± 12	0.047	0.021
	G	0.66 ± 0.06	78 ± 26	0.0085	0.0039
	T	0.57 ± 0.06	280 ± 68	0.0020	0.00091
	C	5.1 ± 0.3	2.3 ± 0.5	2.2	
dG	A	2.4 ± 0.1	63 ± 11	0.038	0.0066
	G	1.0 ± 0.1	56 ± 18	0.018	0.0031
	T	2.8 ± 0.2	140 ± 20	0.02	0.0034
	C	3.5 ± 0.3	0.60 ± 0.23	5.8	

[a] f(misinsertion frequency) = $(k_{cat}/K_{M,dNTP})_{incorrect}/(k_{cat}/K_{M,dNTP})_{correct}$

Table S3. Products of extension of template-primer complexes containing 2'-F-N^2,3-εdG, 2'-F-dG, and unmodified dG by Dpo4. Mass spectrometry data used to derive these data are given in Figs. S2−S7 and Tables S4−S16.

3'-CCCCCGAGCATTCCTAAGXCACT 5'-GGGGGCTCGTAAGGATUC		%	Inserted base
X=2'-F-N^2,3-εdG	CCATGA	45	C
	CCATGAA	3	
	CTATGA	44	T
	CTATGAA	4	
	CAATGA	1	A
	CGATGA	1	G
	CATGA	1	Deletion
2'-F-dG	CCATGA	100	C
dG	CCATGA	100	C

Figure S2. LC-MS/MS analysis of full-length extension products by Dpo4 formed with the sequence c (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 934.7. b) CID spectrum of *m/z* 934.7. See Table S4 for a full list of assignment of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 727.3. d) CID spectrum for *m/z* 727.3. See Table S5 for a full list of assignments of CID ions.

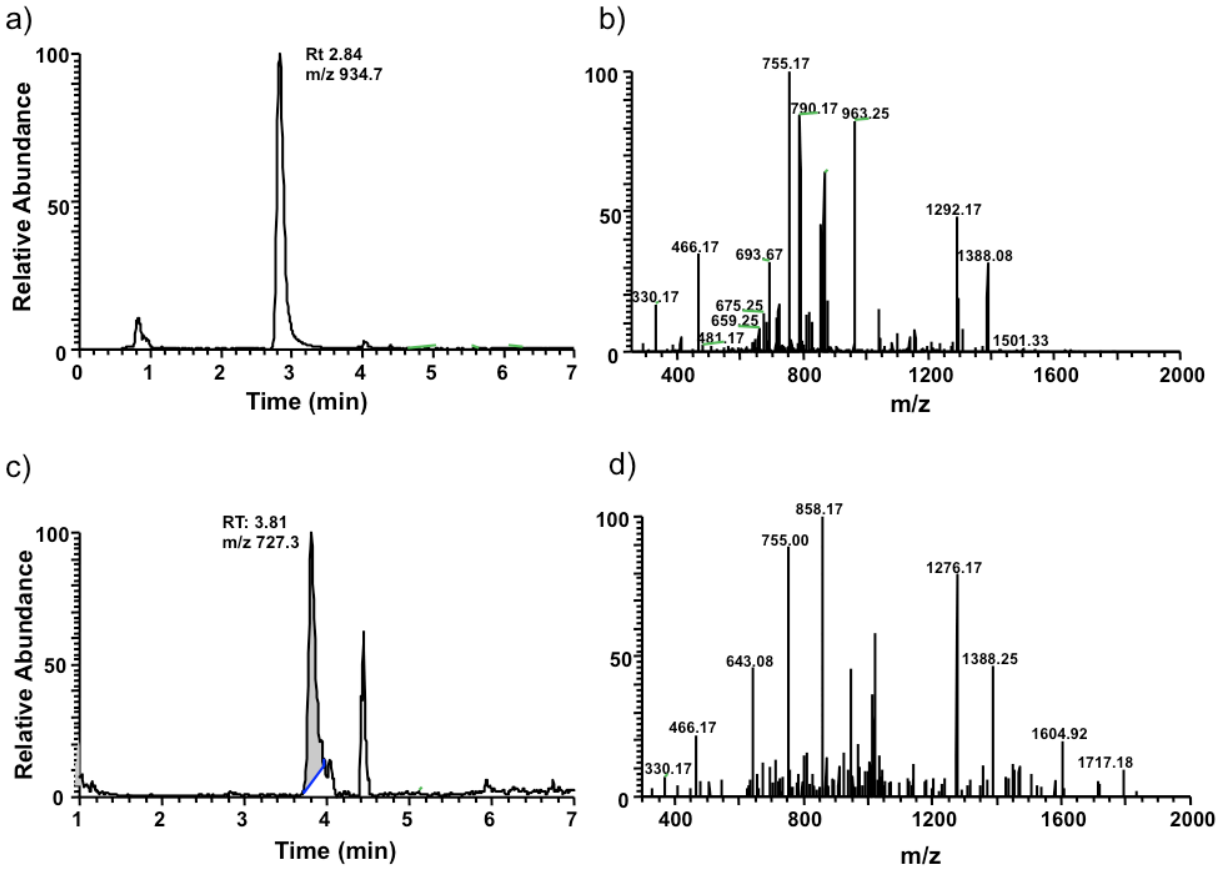


Table S4. Observed and theoretical CID fragmentation of *m/z* 934.7 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CCGTGAA-3' (Figs. S2a, S2b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCC (a_3−B, -1)	755.17	755.09
5'-pCCG (a_4−B, -1)	1084.1	1084.14
5'-pCCGT (a_5−B, -1)	1388.08	1388.18
5'-pCCGT (a_5−B, -2)	693.67	693.59
CCGTA-3' (y_5, -1)	1501.33	1501.29
pCCGTA-3' (w_5, -2)	790.17	790.12
pCGTA-3' (w_4, -1)	1292.17	1292.21
pGTA-3' (w_3, -1)	963.25	963.16
pGTA-3' (w_3, -2)	481.17	481.07
pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.17	330.06

Note: a−B ions carry a 2'-deoxyribose remnant on the 3'-end which is not indicated in the sequence assigment. This applies to all the following tables for sequence assigment.

Table S5. Observed and theoretical CID fragmentation of *m/z* 727.3 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CCGTGAA-3' (Figs. S2c, S2d).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCC (a_3−B, -1)	755.00	755.09
5'-pCCGT (a_5−B, -1)	1388.25	1388.18
5'-pCCGTA (a_6−B, -1)	1717.18	1717.24
5'-pCCGTA (a_6−B, -2)	858.17	858.11
pCCGTAA-3' (w_6, -2)	946.58	946.65
pCGTAA-3' (w_5, -1)	1604.92	1605.27
pGTAA-3' (w_4, -1)	1276.17	1276.21
pA-3' (w_1, -1)	330.17	330.06

Figure S3. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 942.2 (-2). b) CID spectrum of *m/z* 942.2. See Table S6 for a full list of assignments of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 732.1 (-3). d) CID spectrum for *m/z* 732.1. See Table S7 for a full list of assignments of CID ions.

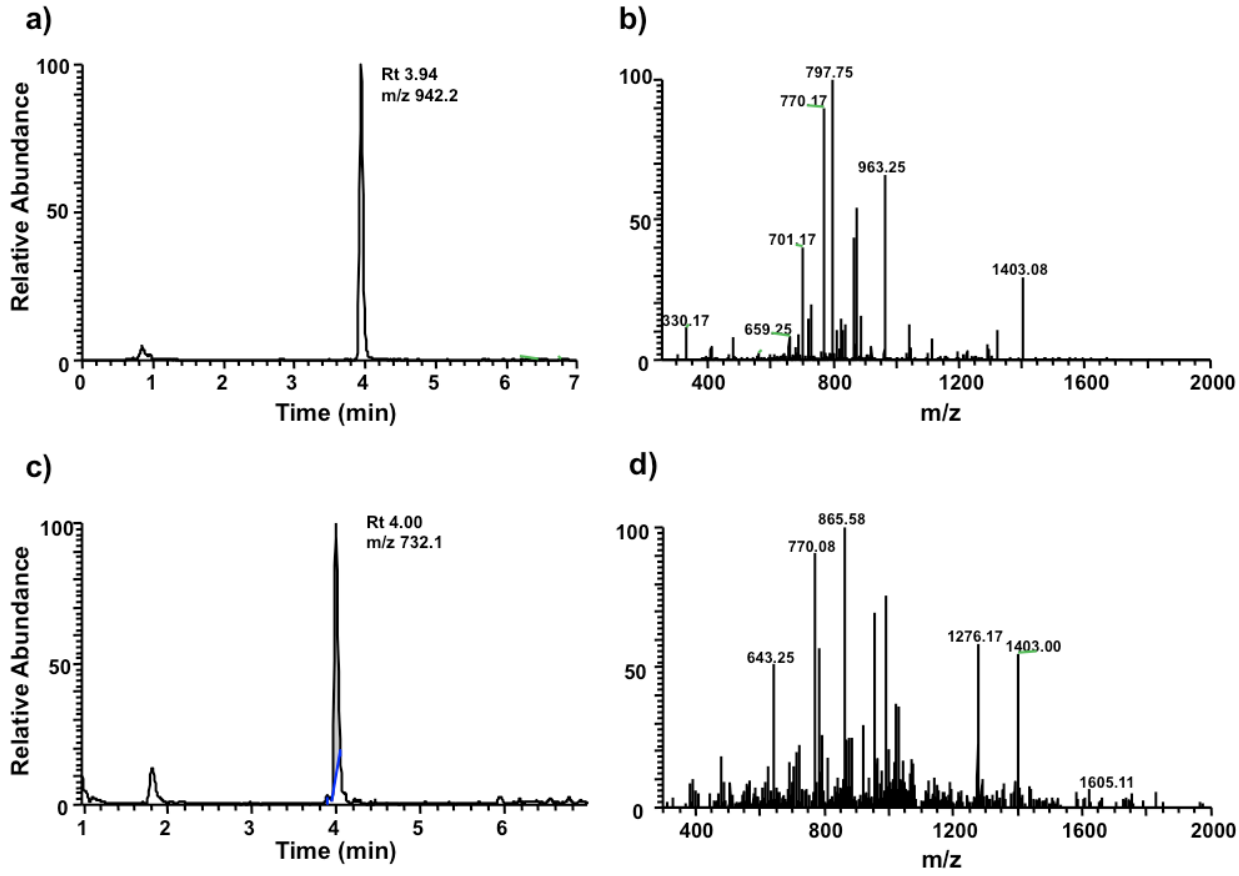


Table S6. Observed and theoretical CID fragmentation of *m/z* 942.2 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CTGTGA-3' (Figs. S3a, S3b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pCT (a_3−B, -1)	770.17	770.09
5'-pCTGT (a_5−B, -1)	1403.08	1403.18
5'-pCTGT (a_5−B, -2)	701.17	701.89
pTCGTA-3' (w_5, -2)	797.75	797.62
pGTA-3' (w_3, -1)	963.25	963.16
pGTA-3' (w_3, -2)	481.17	481.07
pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S7. Observed and theoretical CID fragmentation of *m/z* 732.1 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CTGTGAA-3' (Figs. S2c, S2d).

Fragment assignment	***m/z* observed**	***m/z* theoretical**
5'-pCT (a_3−B, -1)	770.17	770.09
5'-pCTGT (a_5−B, -1)	1403.08	1403.18
5'-pCTGTA (a_6−B, -2)	865.58	865.61
pTCGTAA-3' (w_6, -2)	954.08	954.15
pCCGTA-3' (w_5, -2)	1605.11	1605.27
pCGTA-3' (w_4, -1)	1276.17	1276.21
pTA-3' (w_2, -1)	643.25	643.11

Figure S4. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 946.6 (-2). b) CID spectrum of *m/z* 946.6. See Table S8 for a full list of assignment of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 954.6 (-2). d) CID spectrum for *m/z* 954.6. See Table S9 for a full list of assignments of CID ions. e) Reconstructed extracted ion chromatogram for *m/z* 790.1 (-2). d), CID spectrum for *m/z* 709.1. See Table S10 for a full list of assignments of CID ions.

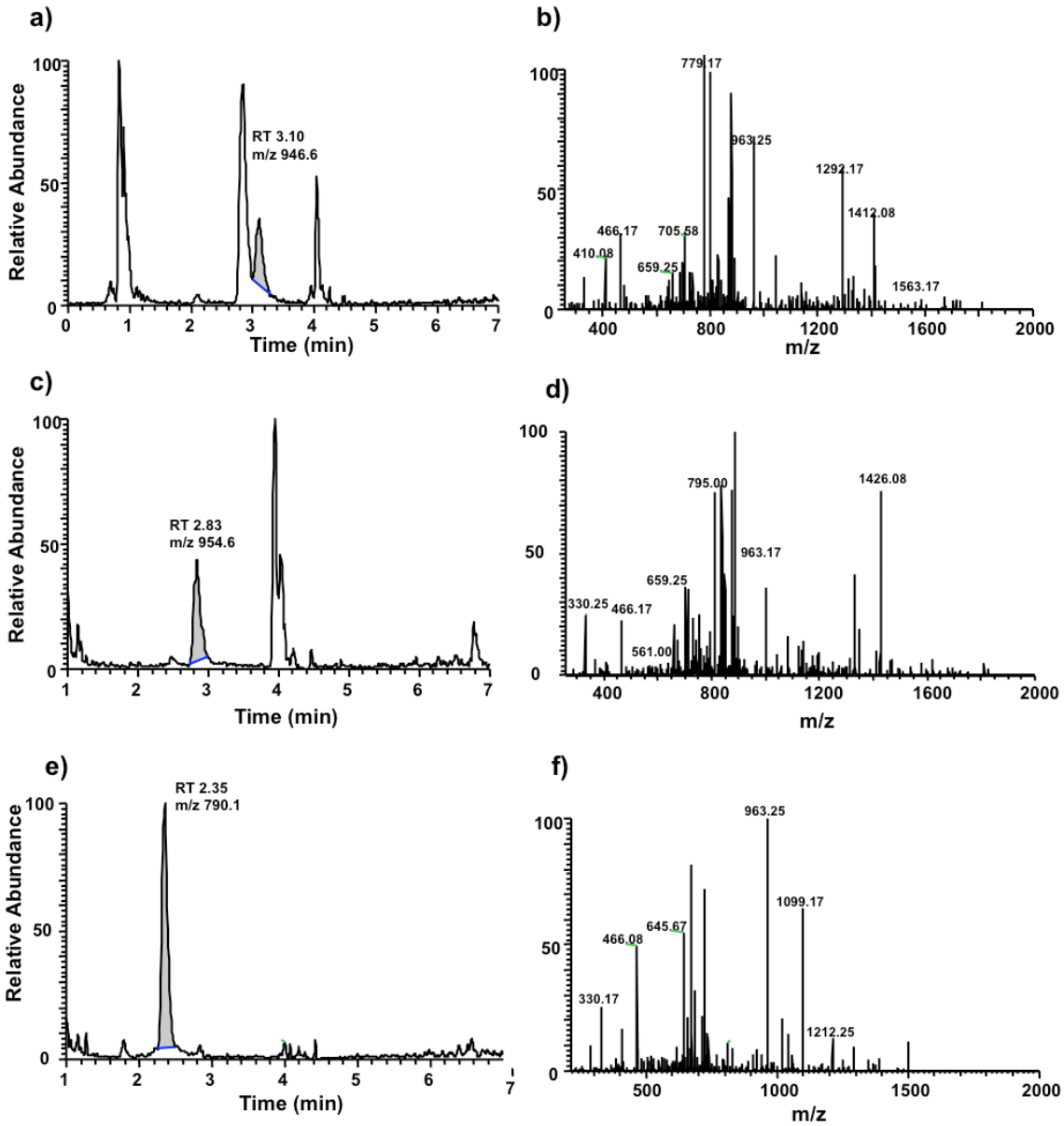


Table S8. Observed and theoretical CID fragmentation of *m/z* 946.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-*N*^{2},3-εdG). The assigned sequence is 5'-CAGTGA-3' (Figs. S4a, S4b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCA (a_3−B, -1)	779.17	779.10
5'-pCAGT (a_5−B, -1)	1412.08	1412.19
5'-pCAGT (a_5−B, -2)	705.59	705.58
pCAGTA-3' (w_5, -2)	790.17	790.12
pCGTA-3' (w_4, -1)	1292.17	1292.21
pGTA-3' (w_3, -1)	963.25	963.16

pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S9. Observed and theoretical CID fragmentation of *m/z* 954.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CGGTGA-3' (Figs. S4c, S4d).

Fragment assignment	***m/z* observed**	***m/z* theoretical**
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCG (a_3−B, -1)	795.00	795.09
5'-pCGG (a_4−B, -2)	561.00	561.57
5'-pCGGT (a_5−B, -2)	1428.17	1428.19
pCGGTA-3' (w_5, -1)	1621.58	1621.26
pGTA-3' (w_3, -1)	963.25	963.16
pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.25	330.06

Table S10. Observed and theoretical CID fragmentation of *m/z* 790.1 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CGGTGA-3' (Figs. S4e, S4f).

Fragment assignment	***m/z* observed**	***m/z* theoretical**
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCGT (a_4−B, -1)	1099.17	1099.14
GTGA-3' (y_4, -1)	1212.25	1212.24
pGTGA-3' (w_4, -1)	1292.15	1292.21
pGTGA-3' (w_4, -2)	645.67	645.60
pTGA-3' (w_3, -1)	963.25	963.16
pA-3' (w_1, -1)	330.25	330.06

Figure S5. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 926.7 (-2). b) CID spectrum of *m/z* 926.7. See Table S11 for the full list of assignments of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 721.7 (-3). d), CID spectrum for *m/z* 721.7. See Table S12 for the full list of assignments of CID ions.

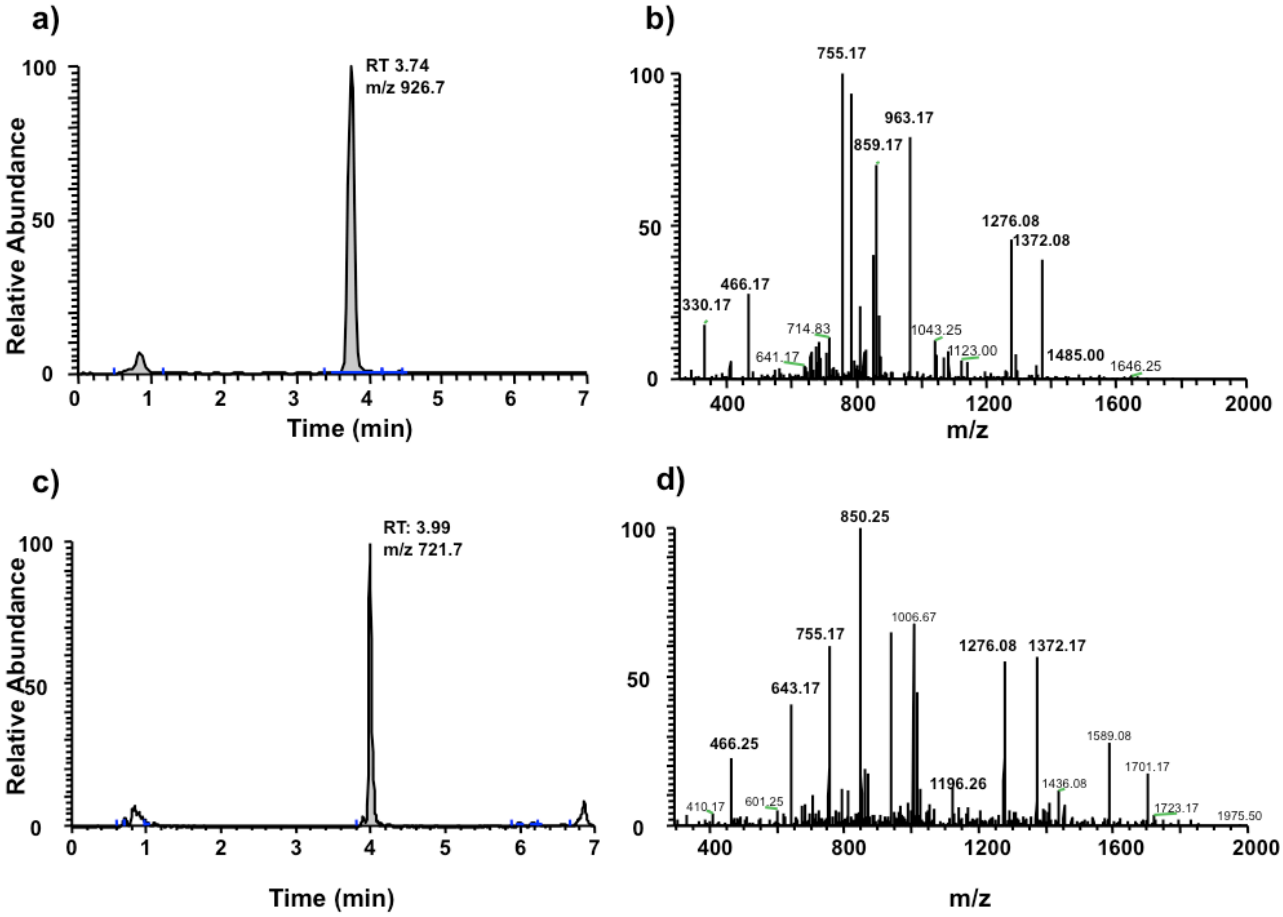


Table S11. Observed and theoretical CID fragmentation of *m/z* 926.7 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CCATGA-3' (Figs. S5a, S5b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCC (a_3−B, -1)	755.17	755.09
5'-pCCAT (a_5−B, -1)	1372.08	1372.19
CATGA-3' (y_5, -1)	1485.00	1485.29
pATGA-3' (w_4, -1)	1276.08	1276.21
pGTA-3' (w_3, -1)	963.17	963.16
pGTA-3' (w_3, -2)	481.02	481.07
pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S12. Observed and theoretical CID fragmentation of *m/z* 721.7 (-3) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CCATGAA-3' (Figs. S5c, S5d).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.25	466.04
5'-pCC (a_3−B, -1)	755.17	755.09
5'-pCCAT (a_5−B, -1)	1372.08	1372.19
5'-pCCATG (a_6−B, -2)	850.12	850.25
pCATGAA-3' (w_6, -2)	938.53	938.66
pATGAA-3' (w_5, -1)	1589.06	1589.27
pTGTA-3' (w_4, -1)	1276.08	1276.21
TGTA-3' (y_4, -1)	1196.26	1196.25
pTA-3' (w_2, -1)	643.17	643.12

Figure S6. LC-MS/MS analysis of the full-length extension product by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 934.2 (-2). b) CID spectrum of *m/z* 934.2. See Table S13 for the full list of assignments of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 1090.7 (-2). d) CID spectrum for *m/z* 1090.7. See Table S14 for the full list of assignments of CID ions.

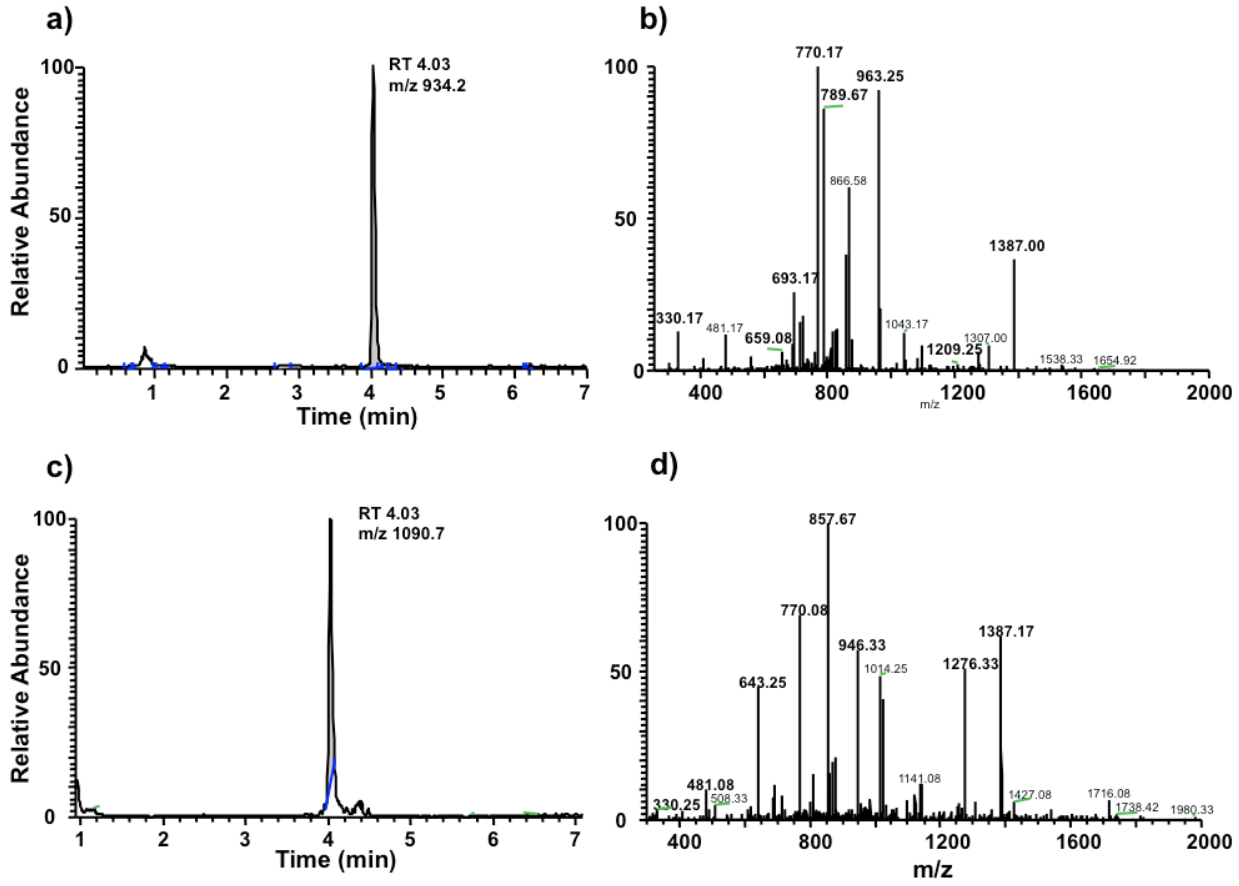


Table S13. Observed and theoretical CID fragmentation of *m/z* 934.2 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CTATGA-3' (Figs. S6a, S6b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCA (a_3−B, -1)	770.17	770.09
5'-pCAGT (a_4−B, -1)	1083.17	1083.14
5'-pCAGT (d_4−H_2O, -1)	1029.25	1029.19
5'-pCAGTG (a_5−B, -1)	1387.00	1387.19
5'-pCAGTG (a_5−B, -2)	963.25	963.09
pAGTGA-3' (w_5, -1)	1580.13	1580.26
pAGTGA-3' (w_5, -2)	789.62	789.67
pTGA-3' (w_3, -1)	963.25	963.16
pGA-3' (w_2, -1)	659.08	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S14. Observed and theoretical CID fragmentation of *m/z* 1090.7 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG). The assigned sequence is 5'-CTATGAA-3' (Figs. S6c, S6d).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pCT (a_3−B, -1)	770.08	770.09
5'-pCTAT (a_5−B, -1)	1387.17	1387.19
5'-pCTATG (a_6−B, -1)	1716.08	1716.24
5'-pCTATG (a_6−B, -2)	857.67	857.62
pTATGAA-3' (w_6, -2)	946.33	946.15
pTGAA-3' (w_4, -1)	1276.33	1276.21
pAA-3' (w_2, -1)	643.25	643.12
pA-3' (w_1, -1)	330.17	330.06

Figure S7. LC-MS/MS analysis of the full-length extension by Dpo4 formed with the sequence *c* (Scheme 1, Z = T, X= 2'-F-N^2,3-εdG) in the presence of all four dNTPs. a) Reconstructed extracted ion chromatogram for *m/z* 938.6(-2). b) CID spectrum of *m/z* 938.6. See Table S15 for the full list of assignments of CID ions. c) Reconstructed extracted ion chromatogram for *m/z* 782.1(-2). d) CID spectrum for *m/z* 782.1. See Table S16 for the full list of assignments of CID ions.

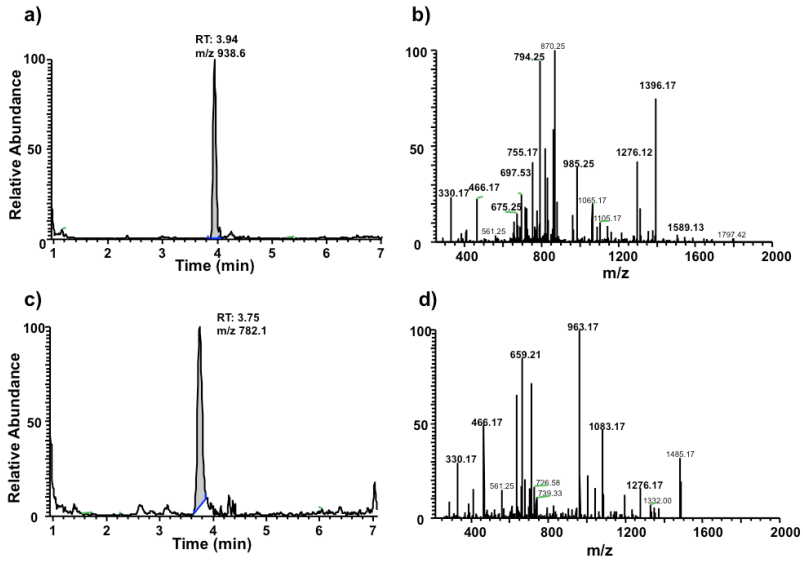


Table S15. Observed and theoretical CID fragmentation of *m/z* 938.6 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X=2'-F-N^2,3-εdG). The assigned sequence is 5'-CAATGA-3' (Figs. S7a, S7b).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCA (a_3−B, -1)	779.25	779.10
5'-pCAATG (a_5−B, -2)	679.53	679.60
pAATGA-3' (w_5, -2)	794.25	794.13
pACGTA-3' (w_5, -1)	1589.13	1589.27
pCGTA-3' (w_4, -1)	1276.12	1276.21
pGTA-3' (w_3, -1)	963.17	963.16
pTA-3' (w_2, -1)	659.25	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S16. Observed and theoretical CID fragmentation of *m/z* 782.1 (-2) for Dpo4-catalyzed extension in sequence *c* (Scheme 1, Z = C, X=2'-F-N^2,3-εdG). The assigned sequence is 5'-CATGA-3' (Figs. S7c, S7d).

Fragment assignment	*m/z* observed	*m/z* theoretical
5'-pC (a_2−B, -1)	466.17	466.04
5'-pCA (a_3−B, -1)	779.25	779.10
5'-pCAA (a_4−B, -1)	1083.14	1083.17
pCGTA-3' (w_4, -1)	1276.17	1276.21
pGTA-3' (w_3, -1)	963.17	963.16
pTA-3' (w_2, -1)	659.21	659.11
pA-3' (w_1, -1)	330.17	330.06

Table S17. Crystal data collection and refinement statistics for the ternary complexes (Dpo4•N^2,3-εG •dCTP) Dpo4-1 and Dpo4-2 and the binary (Dpo4• N^2,3-εG-DNA) complex Dpo4-3 (RCSB ID codes rcsb069655, rcsb069657, and rcsb069658; PDB ID codes 3V6H, 3V6J, and 3V6K, respectively).

Data collection	Dpo4-1 N^2,3-εG •dCTP	Dpo4-2 N^2,3-εG •dCTP	Dpo4-3 N^2,3-εG •ddT
Beamline	21-ID-D	21-ID-D	21-ID-G
Space group	$P2_1$	$P2_1$	$P2_1$
Unit cell (a, b, c) (Å)	52.53, 111.27, 98.94	52.70, 111.42, 98.81	51.86, 110.95, 100.24
Unit cell (α, β, γ) (°)	90.00, 102.68, 90.00	90.00, 102.57, 90.00	90.00, 102.70, 90.00
Resolution (Å) [a]	2.30 (2.38−2.30)	2.30 (2.38−2.30)	3.50 (3.63−3.50)
No. of measured reflections	246801	221187	54603
No. of unique reflections	48539	48579	14236
Completeness (%)	98.2 (86.5)	97.8 (85.7)	100.0 (100.0)
Redundancy	5.1	4.6	3.8
R_{merge} (%) [b]	8.6 (60.5)	8.7 (53.4)	13.5 (66.5)
Signal to noise ($I/\sigma I$)	17.2 (2.4)	15.3 (2.3)	11.7 (2.2)
Model composition (asymmetric unit)			
No. of amino acid residues	685	684	676
No. of water molecules	203	202	30
No. of Mg^{2+} ions	6	5	4
No. of template nucleotides	30	30	27
No. of primer nucleotides	23	23	20
No. of dCTP	2	2	0
No. of ddT	0	0	1
Refinement			
Resolution range (Å)	30.00−2.30	30.00−2.30	50.00−3.60
Reflections	46041	48579	11927
R_{work}(%) [c]	21.5	21.0	22.9
R_{free} (%) [d]	26.7	27.3	30.3
Root mean square deviation bond length (Å)	0.006	0.016	0.011
Root mean square deviation bond angles (°)	1.25	2.14	1.60
Mean B-factor (Å^2)	71.1	62.4	111.1

[a] Values for outermost shells are given in parentheses.
[b] $R_{merge} = \Sigma| I - \langle I \rangle |\Sigma I$, where I is the integrated intensity of a given reflection.
[c] $R_f = \Sigma||F_{observed} - F_{calculated}||/\Sigma|F_{observed}|$.
[d] R_{free} was calculated using 5% (for Dpo4-1 and Dpo4-2) or 8% (for Dpo4-3), with random data omitted from the refinement.

Figure S8. Crystal structures of Dpo4•N^2,3-εG-DNA complexes (Z=T in the template). a) Ternary complex of Dpo4•N^2,3-εG•dCTP, the quality of the data is demonstrated using non-biased omit electron density maps, displayed as red mesh at 1.0 σ; (b) the orientation of the bases with two possible hydrogen bonding mechanisms (distances in Å). (c) Superimposed DNA structures of three complexes in the current study, i.e. Dpo4-1 (red), Dpo4-2 (blue), and Dpo4-3 (green).

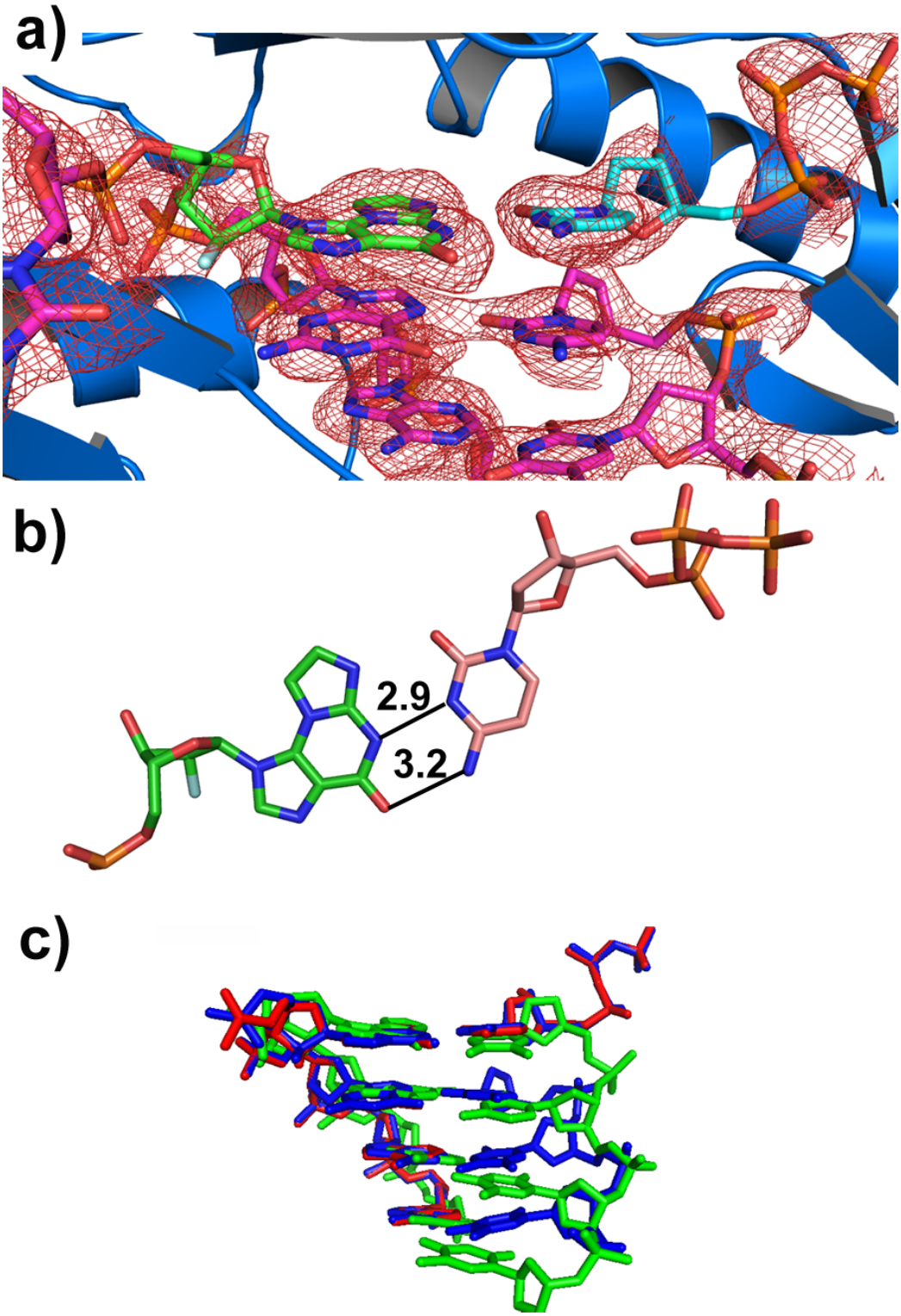


Table S18. Comparisons of kinetic parameters of KF^{-}, Dpo4^{-}, and human pol κ-catalyzed single base insertion opposite two different etheno DNA adducts: X is 2'-F-N^2,3-εdG, W is 1,N^2-εdG. Data for 1,N^2-εdG are from previous work.[1b, 16]

	Template	dNTP	$k_{cat}/K_{M,dNTP}$	f[a]	Template	dNTP	$k_{cat}/K_{M,dNTP}$	f[a]
KF^{-}	3'-XTA-	C	0.23	1	3'-WTG-	C	0.0081	1
		T	0.24	1		G	0.0087	1.1
						A	0.0016	1.2
Dpo4	3'-XTA-	C	0.025	1	3'-WTA-	C	0.00006	1
		T	0.0054	0.22		A	0.0008	14
Human pol κ	3'-XTA-	C	0.022	1	3'-WTA-	C	0.0012	1
		T	0.0081	0.37		T	0.0012	1.0

[a] f (misinsertion frequency) = $(k_{cat}/K_{M,dNTP})_{incorrect}/(k_{cat}/K_{M,dNTP})_{correct}$

Figure S9. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-dG)GAATCCTTACGAGCCCCC-3'. *m/z* calculated for [M–H]$^{-}$, 6946.6; found 6943.0.

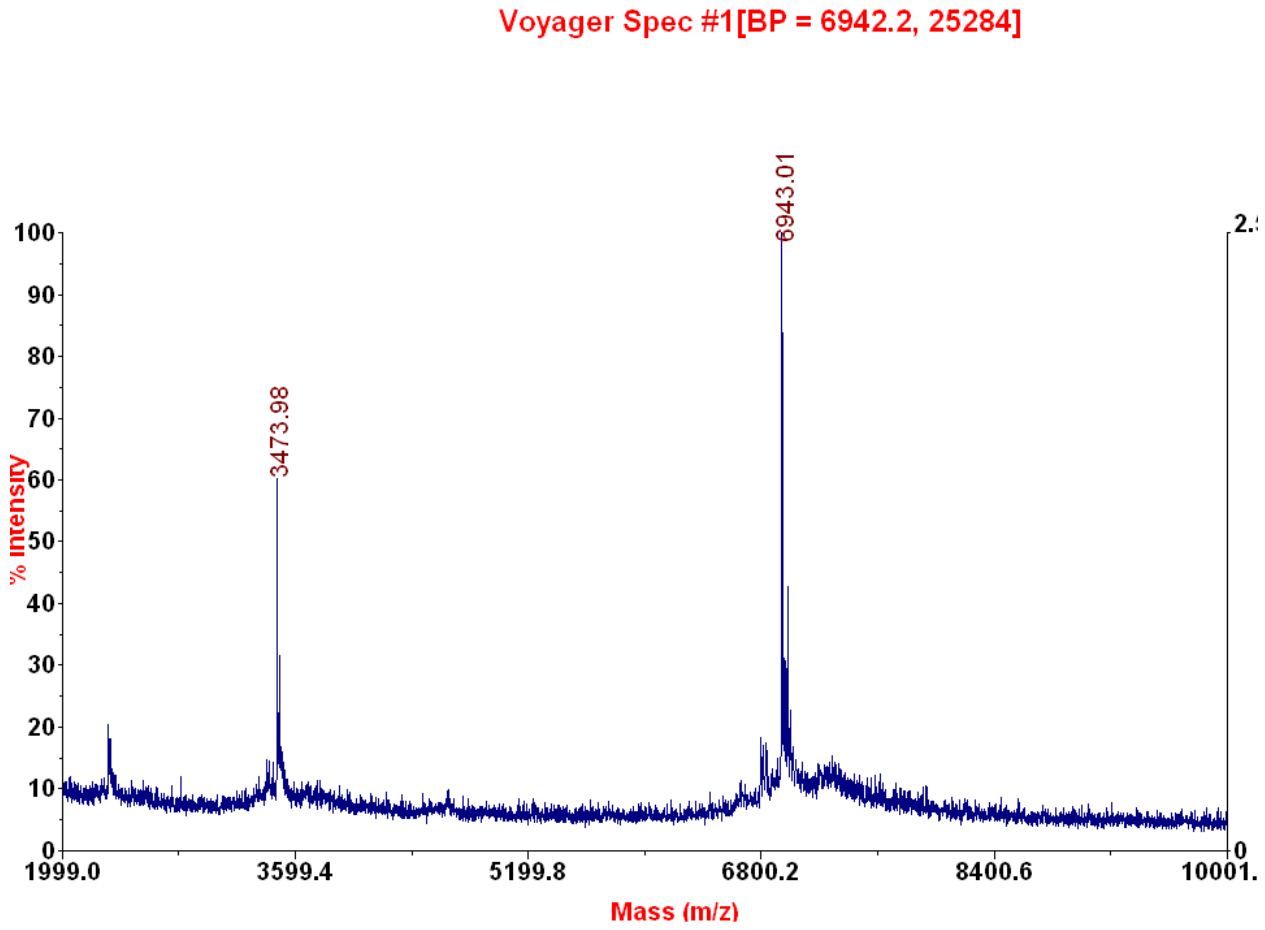


Figure S10. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-dG)GAATCCTTACGAGCCCCC-3'. *m/z* calculated for $[M–H]^-$, 6962.5; found 6965.4.

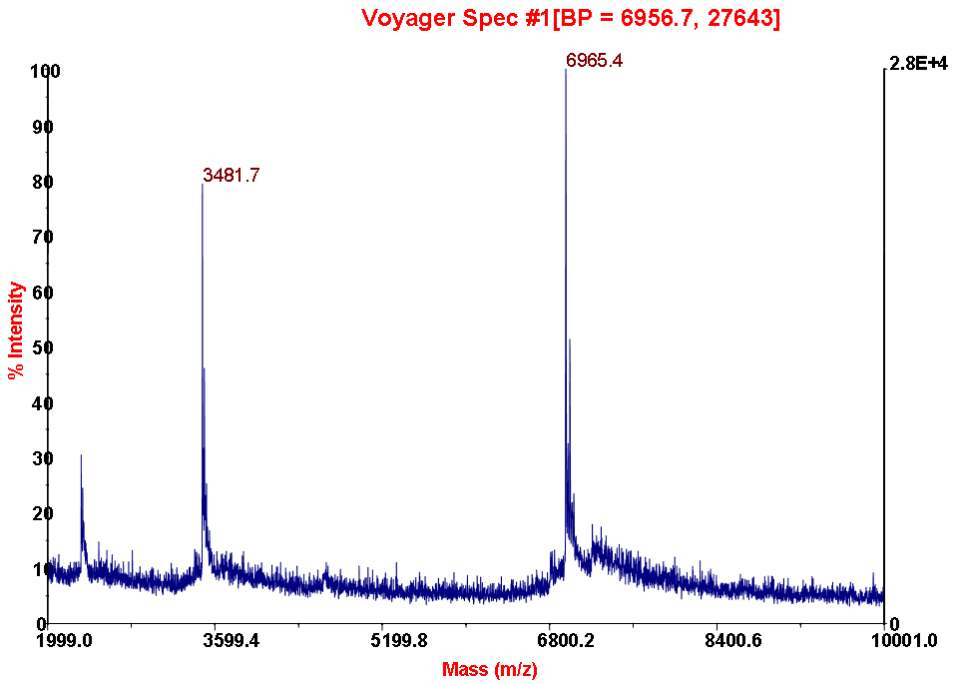


Figure S11. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3'. *m/z* calculated for $[M–H]^-$, 6971.5; found 6970.4.

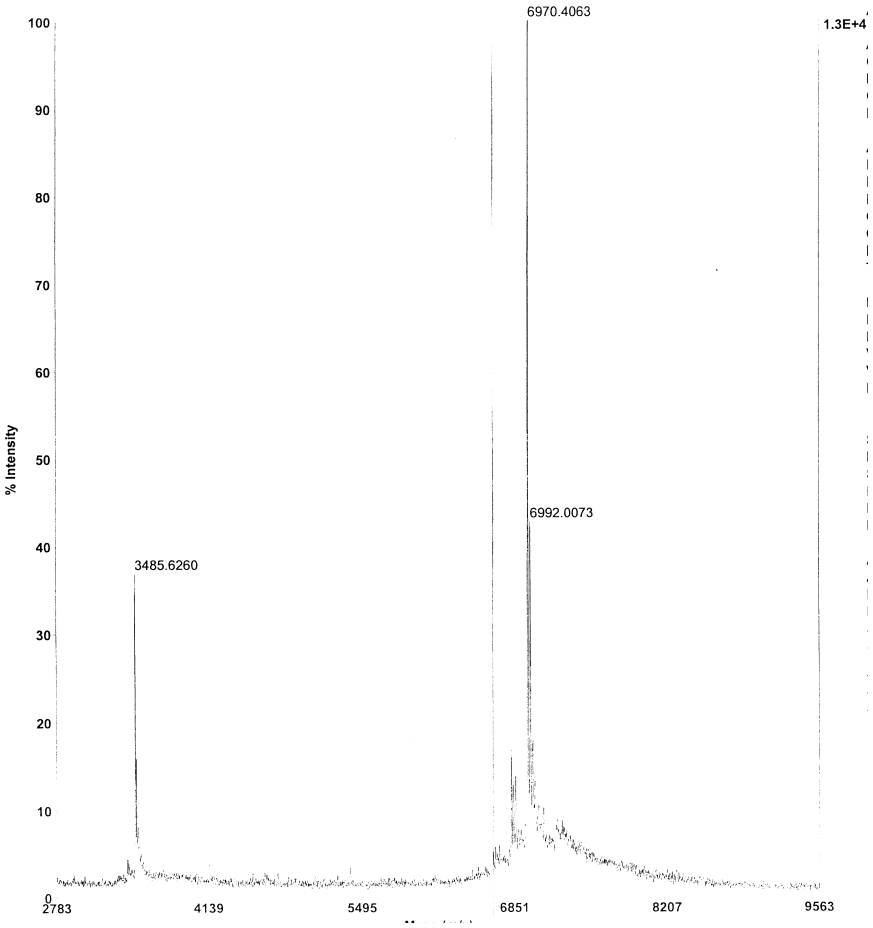


Figure S12. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTACGAGCCCCC-3'. *m/z* calculated for $[M–H]^-$, 6986.5; found 6985.6.

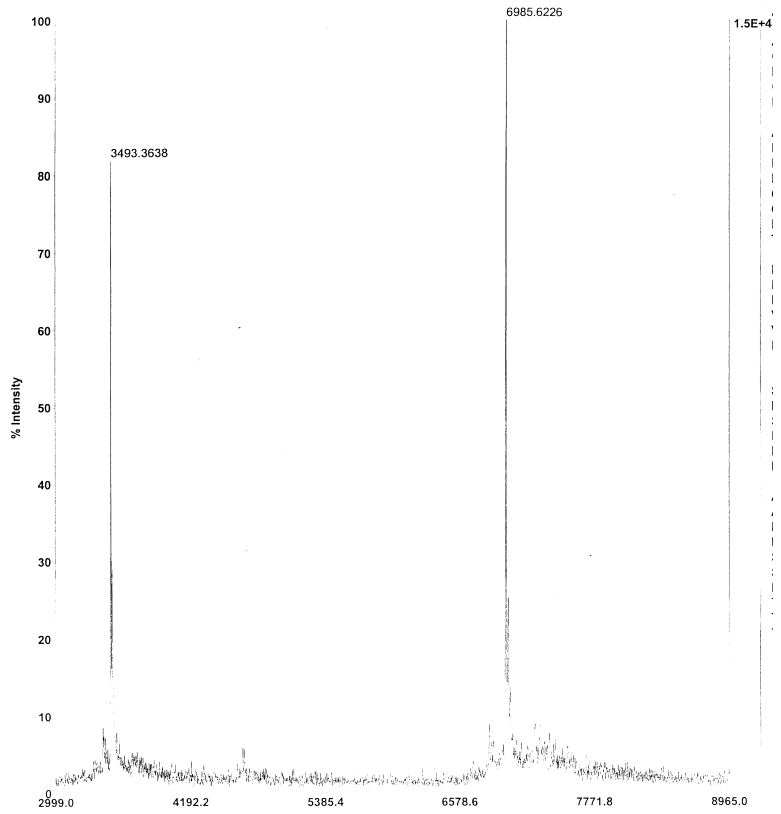


Figure S13. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'. *m/z* calculated for $[M–H]^-$, 5412.5; found 5412.1.

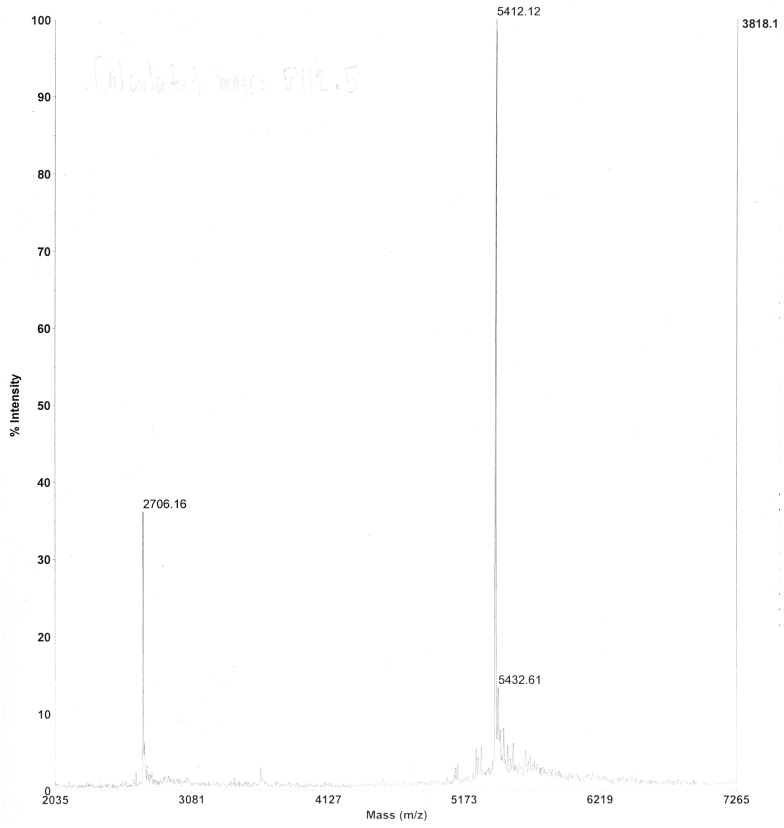


Figure S14. MALDI-TOF mass spectrometry characterization of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'. *m/z* calcd for $[M–H]^-$, 5397.5; found 5397.5.

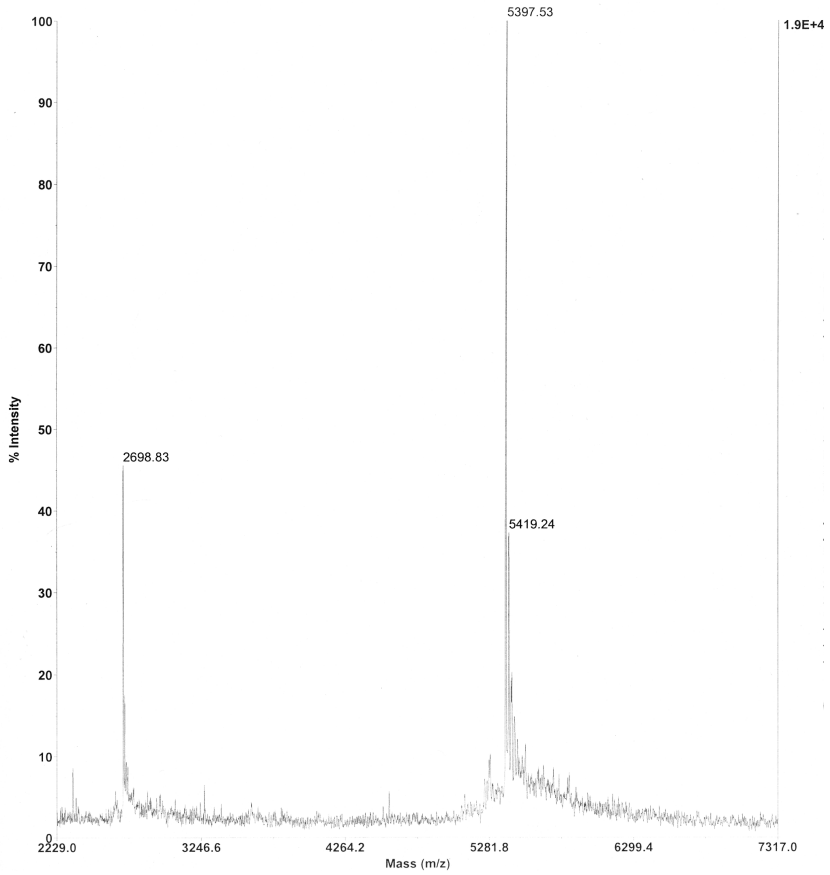
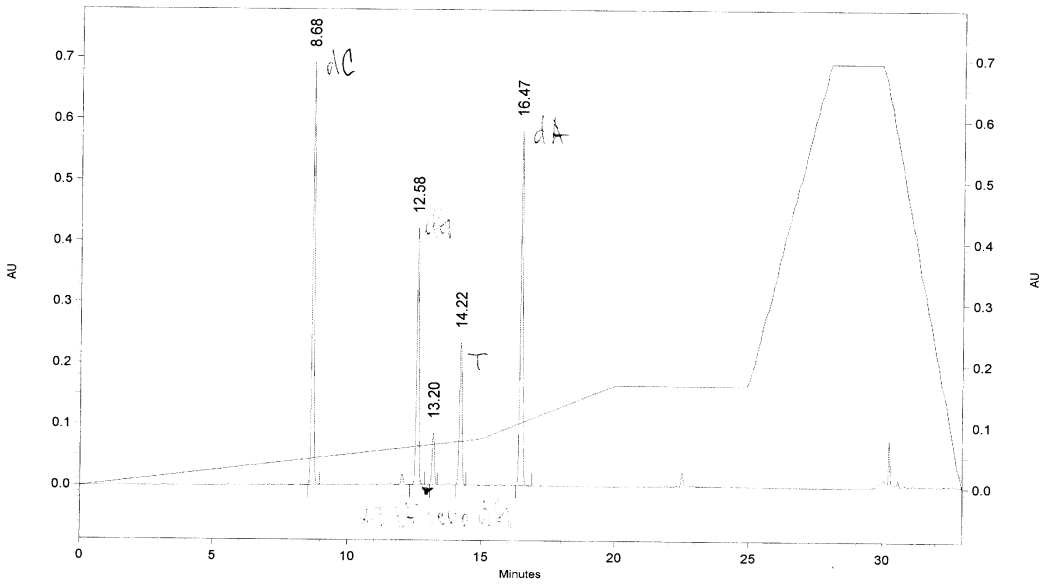


Figure S15. HPLC analysis of enzymatic digest of the oligonucleotide 5'-TCAC(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'. The species with their peak area percent and retention times are as follows: dC (32%), 8.68 min; dG (20%), 12.58 min; 2'-F-N^2,3-εdG (4.3%), 13.20 min; dT (13%,), 14.22 min; dA (31%), 16.47 min.

Enzyme Digest Angela 23mer C 2,3 Etheno dG

HPLC: HPLC4 (Offline)
Run Time: 11/19/2010 10:30:21 AM
Analysis Time: 11/19/2010 11:11:11 AM
Method Name: D:\32Karat\HPLC Files\Methods\Ivan\IKHM2533-1.0ml.met
File Name: D:\32Karat\HPLC Files\Data\Albena\Enzyme Digest Angela 23mer C 2,3 Etheno dG-1.dat



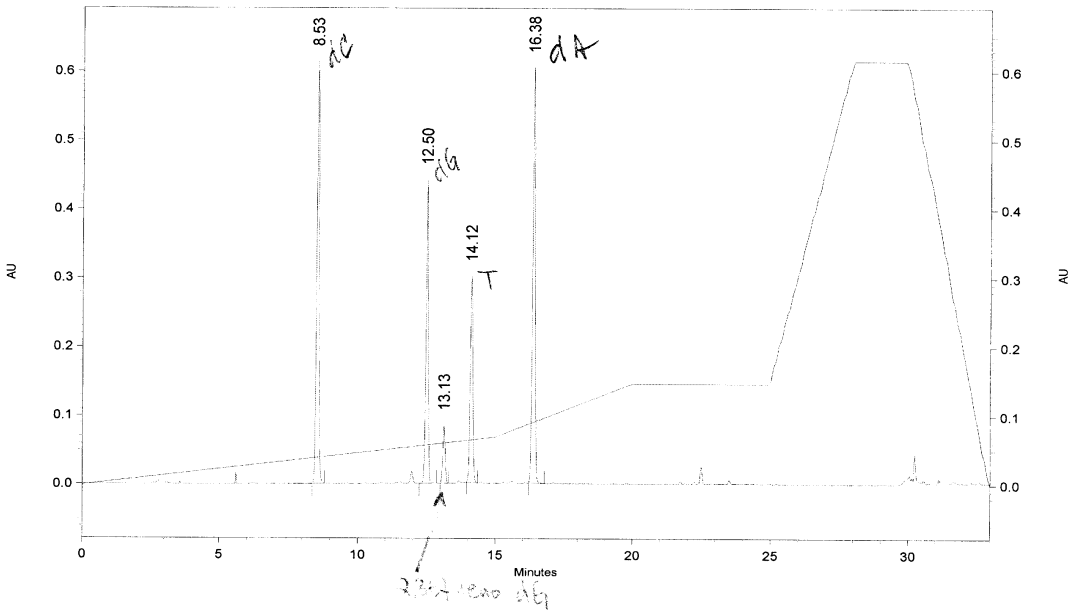
Det 168-254nm Results

Pk #	Retention Time	Area	Area Percent
1	8.683	3701767	31.489
2	12.583	2383056	20.271
3	13.200	499703	4.251
4	14.217	1493840	12.707
5	16.467	3677450	31.282
Totals		11755816	100.000

Figure S16. HPLC analysis of enzymatic digest of the oligonucleotide 5'-TCAT(2'-F-N^2,3-εdG)GAATCCTTCCCCC-3'. The species with their peak area percent and retention times are as follows: dC (29%), 8.53 min; dG (20%), 12.50 min; 2'-F-N^2,3-εdG (4.1%), 13.13 min; dT (16%,), 14.12 min; dA (31%), 16.38 min.

Enzyme Digest Angela 23mer T 2,3 Etheno dG

HPLC: HPLC4
Run Time: 11/19/2010 11:09:49 AM
Analysis Time: 11/19/2010 11:52:57 AM
Method Name: D:\32Karat\HPLC Files\Methods\Ivan\IKHM2533-1.0ml.met
File Name: D:\32Karat\HPLC Files\Data\Albena\Enzyme Digest Angela 23mer T 2,3 Etheno dG-1.dat



Det 168-254nm Results

Pk #	Retention Time	Area	Area Percent
1	8.533	3606759	28.513
2	12.500	2584714	20.433
3	13.133	517434	4.091
4	14.117	1982202	15.670
5	16.383	3958330	31.293
Totals		12649439	100.000

REFERENCES FOR SUPPORTING INFORMATION SECTION

[1] a) S. S. Patel, I. Wong, K. A. Johnson, *Biochemistry* **1991**, *30*, 511-525; b) H. Zang, T. M. Harris, F. P. Guengerich, *J. Biol. Chem.* **2005**, *280*, 1165-1178.

[2] L. L. Furge, F. P. Guengerich, *Biochemistry* **1997**, *36*, 6475-6487.

[3] H. Zang, A. K. Goodenough, J. Y. Choi, A. Irimia, L. V. Loukachevitch, I. D. Kozekov, K. C. Angel, C. J. Rizzo, M. Egli, F. P. Guengerich, *J. Biol. Chem.* **2005**, *280*, 29750-29764.

[4] M. G. Pence, P. Blans, C. N. Zink, J. C. Fishbein, F. W. Perrino, *DNA Repair* **2011**, *10*, 56-64.

[5] T. D. Silverstein, R. E. Johnson, R. Jain, L. Prakash, S. Prakash, A. K. Aggarwal, *Nature* **2010**, *465*, 1039-1043.

[6] R. Khazanchi, P. L. Yu, F. Johnson, *J. Org. Chem.* **1993**, *58*, 2552-2556.

[7] C. E. Elmquist, J. S. Stover, Z. Wang, C. J. Rizzo, *J. Am. Chem. Soc.* **2004**, *126*, 11189-11201.

[8] J.-Y. Choi, G. Chowdhury, H. Zang, K. C. Angel, C. C. Vu, L. A. Peterson, F. P. Guengerich, *J. Biol. Chem.* **2006**, *281*, 38244-38256.

[9] J. Rozenski, J. A. McCloskey, *J. Am. Soc. Mass. Spectrom.* **2002**, *13*, 200-203.

[10] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, *Cell* **2001**, *107*, 91-102.

[11] Z. Otwinowski, W. Minor, *Macromolec. Crystallogr., Pt A* **1997**, *276*, 307-326.

[12] a) A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022-1025; b) S. Bailey, *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **1994**, *50*, 760-763.

[13] a) M. D. Winn, G. N. Murshudov, M. Z. Papiz, *Methods Enzymol.* **2003**, *374*, 300-321; b) M. D. Winn, M. N. Isupov, G. N. Murshudov, *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **2001**, *57*, 122-133.

[14] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D., Biol. Crystallogr.* **2004**, *60*, 2126-2132.

[15] W. L. DeLano, *The PyMOL Molecular Graphics System* 2002, DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

[16] a) J.-Y. Choi, H. Zang, K. C. Angel, I. D. Kozekov, A. K. Goodenough, C. J. Rizzo, F. P. Guengerich, *Chem. Res. Toxicol.* **2006**, *19*, 879-886; b) S. Langouët, M. Müller, F. P. Guengerich, *Biochemistry* **1997**, *36*, 6069-6079.