Supplementary Materials for

"Syntheses and Characterizations of the *in-vivo* Replicative Bypass and Mutagenic Properties of the Minor-groove O^2-Alkylthymidine Lesions"

by Qianqian Zhai, Pengcheng Wang, Qian Cai and Yinsheng Wang

***Nucleic Acids Res.*, 2014**

Supplementary Materials and Methods:

Mass spectrometry (MS) and NMR

Electrospray ionization-MS (ESI-MS) and tandem MS (MS/MS) experiments were carried out on an LCQ Deca XP ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). A mixture of acetonitrile and water (50:50, v/v) was used as solvent for electrospray. The spray voltage was 3.0 kV, and the temperature for the ion transport tube was maintained at 275°C. High-resolution mass spectra (HRMS) were acquired on an Agilent 6210 ESI-TOF instrument (Agilent Technologies, Palo Alto, CA). ^{1}H and ^{31}P NMR spectra were recorded at 300 MHz and 80 MHz, respectively, on a Varian Inova 300 instrument (Varian Inc., Palo Alto, CA).

2D NMR spectra were recorded at 25°C in CD_3OD or DMSO-d_6 using a Bruker Unity spectrometer operated at 600 MHz (Bruker, Co., Fremont, CA). Resonance assignments for O^2-*i*PrdT, O^2-*s*BudT and O^2-*i*BudT were made based on ^{1}H-^{13}C heteronuclear multi-bond correlation (HMBC) experiments, where the HMBC spectra were acquired using sweep widths of 6613.8 Hz and 30183.6 Hz for ^{1}H and ^{13}C, respectively. The first and second delays were set to match a 145-Hz coupling constant and a long-range coupling constant of 10 Hz, respectively.

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

***O*^{2}-*n*-propylthymidine (*O*^{2}-*n*PrdT, 2c)**. Obtained as a white solid (41% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.80 (s, 1H), 6.19 (t, J = 5.9 Hz, 1H), 4.57 (s, 1H), 4.38 (t, J = 6.6 Hz, 2H), 4.08 – 3.93 (m, 3H), 3.04 (d, J = 4.1 Hz, 1H), 2.92 (t, J = 4.9 Hz, 1H), 2.47 – 2.37 (m, 1H), 2.28 – 2.20 (m, 1H), 1.96 (s, 3H), 1.78 (dq, J = 14.5, 7.1 Hz, 2H), 0.99 (t, J = 7.4 Hz, 3H). HRMS (ESI) calcd for $C_{13}H_{21}N_2O_5$ $[M+H]^+$ *m/z* 285.1455, found *m/z* 285.1471.

***O*^{2}-*iso*-propylthymidine (*O*^{2}-*i*PrdT, 2d)**. Obtained as a white solid (47% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.81 (s, 1H), 6.16 (t, J = 5.8 Hz, 1H), 5.46 (dt, J = 11.6, 5.8 Hz, 1H), 4.57 (s, 1H), 3.98 (s, 2H), 3.51 (s, 1H), 3.21 (s, 1H), 3.09 (s, 1H), 2.47 – 2.16 (m, 2H), 1.95 (s, 3H), 1.36 (s, 3H), 1.34 (s, 3H). HRMS (ESI) calcd for $C_{13}H_{21}N_2O_5$ $[M+H]^+$ *m/z* 285.1455, found *m/z* 285.1469.

***O*^{2}-*n*-butylthymidine (*O*^{2}-*n*BudT, 2e)**. Obtained as a colorless film (36% yield). ^{1}H NMR (300

MHz, $CDCl_3$): δ 8.01 (s, 1H), 6.14 (dd, *J* = 6.3, 4.8 Hz, 1H), 4.58 (dd, *J* = 11.1, 6.1 Hz, 1H), 4.40 (t, *J* = 6.6 Hz, 2H), 3.98 (s, 3H), 2.90 (br, 2H), 2.50 – 2.37 (m, 1H), 2.28 – 2.16 (m, 1H), 1.94 (s, 3H), 1.72 (dd, *J* = 14.1, 7.2 Hz, 2H), 1.42 (dd, *J* = 15.2, 7.4 Hz, 2H), 0.95 (t, *J* = 7.4 Hz, 3H). HRMS (ESI) calcd for $C_{14}H_{23}N_2O_5$ $[M+H]^+$ *m/z* 299.1601, found *m/z* 299.1622.

O^2-*sec*-butylthymidine (O^2-*s*BudT, 2f). Obtained as a colorless film (53% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.00 (s, 1H), 6.14 (dd, *J* = 11.1, 5.6 Hz, 1H), 5.36 – 5.21 (m, 1H), 4.58 (dd, *J* = 11.0, 6.4 Hz, 1H), 4.39 – 4.22 (m, 1H), 3.98 (s, 2H), 2.96 (br, 2H), 2.49 – 2.38 (m, 1H), 2.28 – 2.17 (m, 1H), 1.96 (s, 3H), 1.77 – 1.60 (m, 2H), 1.31 (dd, *J* = 6.2, 2.3 Hz, 3H), 0.93 (td, *J* = 7.4, 2.4 Hz, 3H). HRMS (ESI) calcd for $C_{14}H_{23}N_2O_5$ $[M+H]^+$ *m/z* 299.1601, found *m/z* 299.1624.

O^2-isobutylthymidine (O^2-*i*BudT, 2g). Obtained as a colorless film (58% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.01 (s, 1H), 6.15 (dd, *J* = 6.2, 4.7 Hz, 1H), 4.57 (s, 1H), 4.17 (d, *J* = 6.6 Hz, 2H), 4.01 – 3.96 (m, 3H), 3.50 (s, 1H), 3.48 (s, 1H), 2.48 – 2.40 (m, 1H), 2.29 – 2.17 (m, 1H), 2.13 – 1.99 (m, 1H), 1.93 (s, 3H), 0.99 (s, 3H), 0.97 (s, 3H). HRMS (ESI) calcd for $C_{14}H_{23}N_2O_5$ $[M+H]^+$ *m/z* 299.1601, found *m/z* 299.1615.

5′-*O*-(4,4′-dimethoxytrityl)-O^2-*n*-propylthymidine (DMTr-O^2-*n*PrdT, 3c). Obtained as a white foam (59% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.64 (s, 1H), 7.45 – 7.22 (m, 9H), 6.83 (d, *J* = 8.8 Hz, 4H), 6.27 (t, *J* = 6.6 Hz, 1H), 4.58 (dt, *J* = 6.2, 3.3 Hz, 1H), 4.39 (t, *J* = 6.7 Hz, 2H), 4.13 – 4.07 (m, 1H), 3.79 (s, 6H), 3.44 (ddd, *J* = 29.2, 10.6, 3.3 Hz, 2H), 2.46 – 2.32 (m, 2H), 1.78 (dt, *J* = 14.2, 7.1 Hz, 2H), 1.60 (s, 3H), 0.98 (t, *J* = 7.4 Hz, 3H). HRMS (ESI) calcd for $C_{34}H_{39}N_2O_7$ $[M+H]^+$ *m/z* 587.2752, found *m/z* 587.2776.

5′-*O*-(4,4′-dimethoxytrityl)-O^2-*iso*-propylthymidine (DMTr-O^2-*i*PrdT, 3d). Obtained as a white foam (41% yield). ^{1}H NMR (300 MHz, acetone-d_6): δ 7.69 (s, 1H), 7.54 – 7.24 (m, 9H), 6.91 (d, *J* = 8.9 Hz, 4H), 6.26 (t, *J* = 6.6 Hz, 1H), 5.33 (dt, *J* = 12.5, 6.2 Hz, 1H), 4.59 (s, 1H), 4.09 (d, *J* = 3.4 Hz, 1H), 3.80 (s, 6H), 3.39 (d, *J* = 3.6 Hz, 2H), 2.44 – 2.36 (m, 2H), 1.52 (s, 3H), 1.36 (dd, *J* = 6.2, 3.8 Hz, 6H). HRMS (ESI) calcd for $C_{34}H_{39}N_2O_7$ $[M+H]^+$ *m/z* 587.2752, found *m/z* 587.2758.

5′-*O*-(4,4′-dimethoxytrityl)-O^2-*n*-butylthymidine (DMTr-O^2-*n*BudT, 3e). Obtained as a white foam (52% yield). ^{1}H NMR (300 MHz, acetone-d_6): δ 7.71 (s, 1H), 7.55 – 7.24 (m, 9H), 6.90 (d, *J* = 8.7 Hz, 4H), 6.28 (t, *J* = 6.6 Hz, 1H), 4.60 (dd, *J* = 8.6, 4.7 Hz, 1H), 4.36 (t, *J* = 6.6 Hz, 2H),

4.10 (q, *J* = 3.4 Hz, 1H), 3.79 (s, 6H), 3.40 (d, *J* = 3.5 Hz, 2H), 2.43 (t, *J* = 6.0 Hz, 2H), 1.82 – 1.69 (m, 2H), 1.52 (s, 3H), 1.50 – 1.41 (m, 2H), 0.97 (t, *J* = 7.4 Hz, 3H). HRMS (ESI) calcd for $C_{35}H_{41}N_2O_7$ $[M+H]^+$ *m/z* 601.2908, found *m/z* 601.2897.

5′-*O*-(4,4′-dimethoxytrityl)-O^2-*sec*-butylthymidine (DMTr-O^2-sBudT, 3f). Obtained as a white foam (46% yield). ^{1}H NMR (300 MHz, acetone-d_6): δ 7.71 (s, 1H), 7.53 – 7.19 (m, 9H), 6.90 (d, *J* = 8.8 Hz, 4H), 6.32 – 6.24 (m, 1H), 5.25 – 5.15 (m, 1H), 4.60 (s, 1H), 4.09 (s, 1H), 3.80 (s, 6H), 3.40 (s, 2H), 2.45 – 2.38 (m, 2H), 1.72 (td, *J* = 14.0, 7.2 Hz, 2H), 1.51 (s, 3H), 1.33 (dd, *J* = 6.2, 4.2 Hz, 3H), 0.96 (t, *J* = 7.4 Hz, 3H). HRMS (ESI) calcd for $C_{35}H_{41}N_2O_7$ $[M+H]^+$ *m/z* 601.2908, found *m/z* 601.2913.

5′-*O*-(4,4′-dimethoxytrityl)-O^2-*iso*-butylthymidine (DMTr-O^2-*i*BudT, 3g). Obtained as a white foam (55% yield). ^{1}H NMR (300 MHz, acetone-d_6): δ 7.72 (s, 1H), 7.54 – 7.27 (m, 9H), 6.90 (d, *J* = 8.9 Hz, 4H), 6.31 (t, *J* = 6.5 Hz, 1H), 4.62 (dd, *J* = 8.5, 4.9 Hz, 1H), 4.17 – 4.07 (m, 3H), 3.80 (s, 6H), 3.41 (d, *J* = 3.5 Hz, 2H), 2.44 (dd, *J* = 6.5, 5.1 Hz, 2H), 2.11 (dd, *J* = 13.5, 6.7 Hz, 1H), 1.52 (d, *J* = 1.1 Hz, 3H), 1.03 (s, 3H), 1.00 (d, *J* = 4.0 Hz, 3H). HRMS (ESI) calcd for $C_{35}H_{41}N_2O_7$ $[M+H]^+$ *m/z* 601.2908, found *m/z* 601.2917.

4c: ^{31}P NMR ($CDCl_3$): δ 150.31, 149.78.

4d: ^{31}P NMR ($CDCl_3$): δ 150.30, 149.75.

4e: ^{31}P NMR ($CDCl_3$): δ 150.42, 149.81.

4f: ^{31}P NMR ($CDCl_3$): δ 150.44, 150.36, 149.75.

4g: ^{31}P NMR ($CDCl_3$): δ 150.46, 149.73.

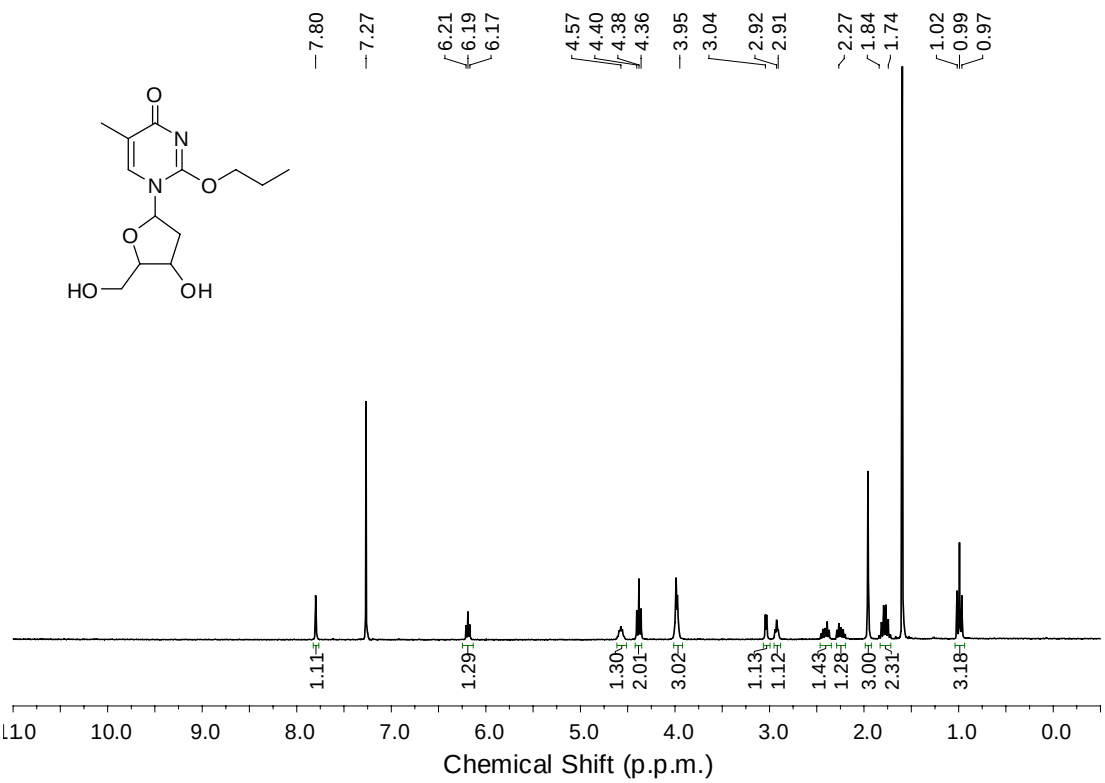


Figure S1. ^{1}H NMR spectrum of O^{2}-*n*PrdT (300 MHz, $CDCl_3$, 25°C).

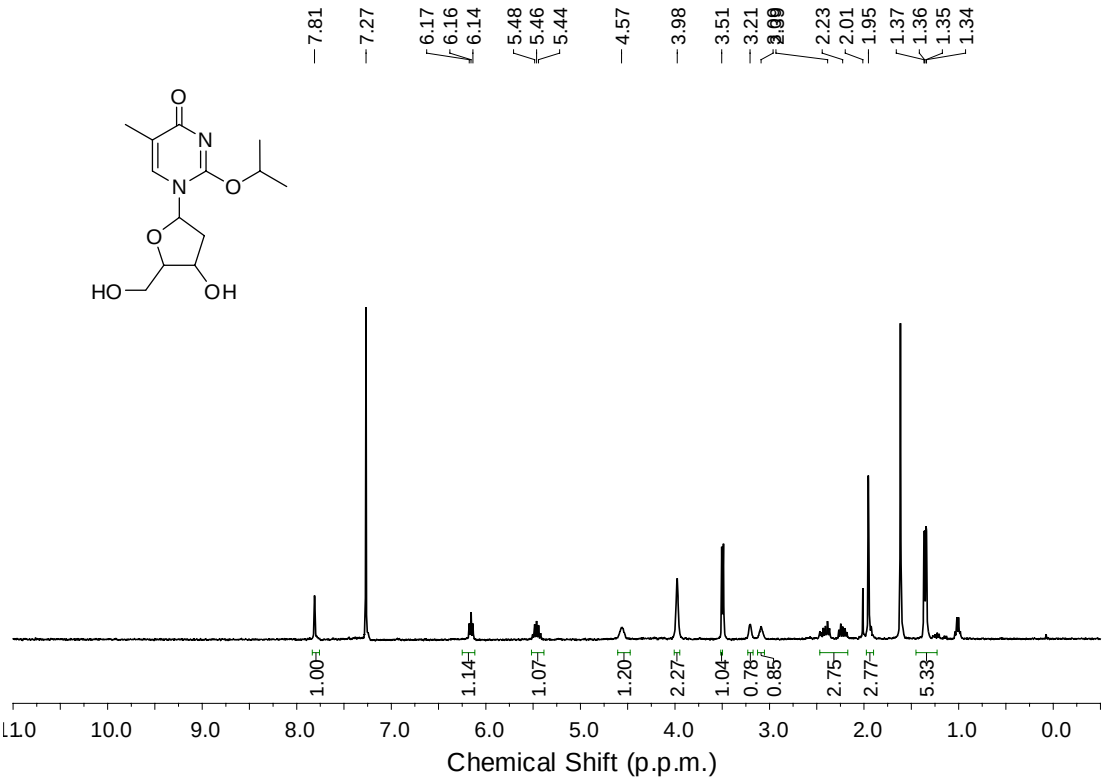


Figure S2. ^{1}H NMR spectrum of O^{2}-*i*PrdT (300 MHz, $CDCl_3$, 25°C).

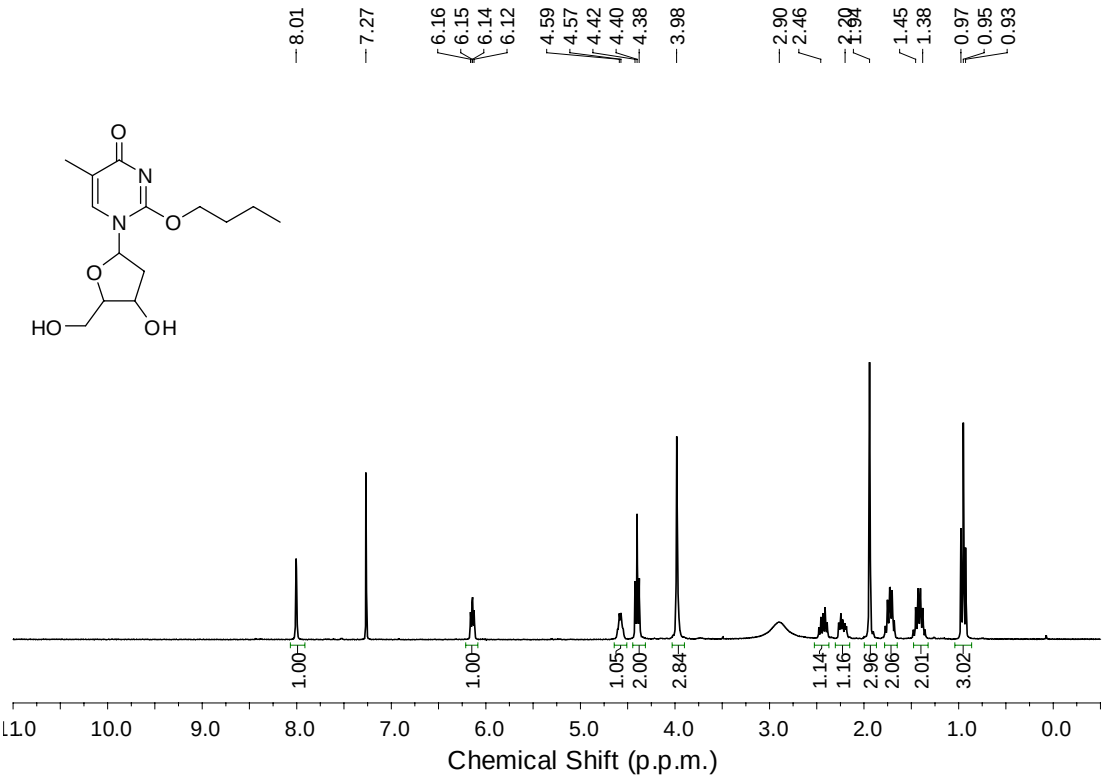


Figure S3. ^{1}H NMR spectrum of O^{2}-*n*BudT (300 MHz, $CDCl_3$, 25°C).

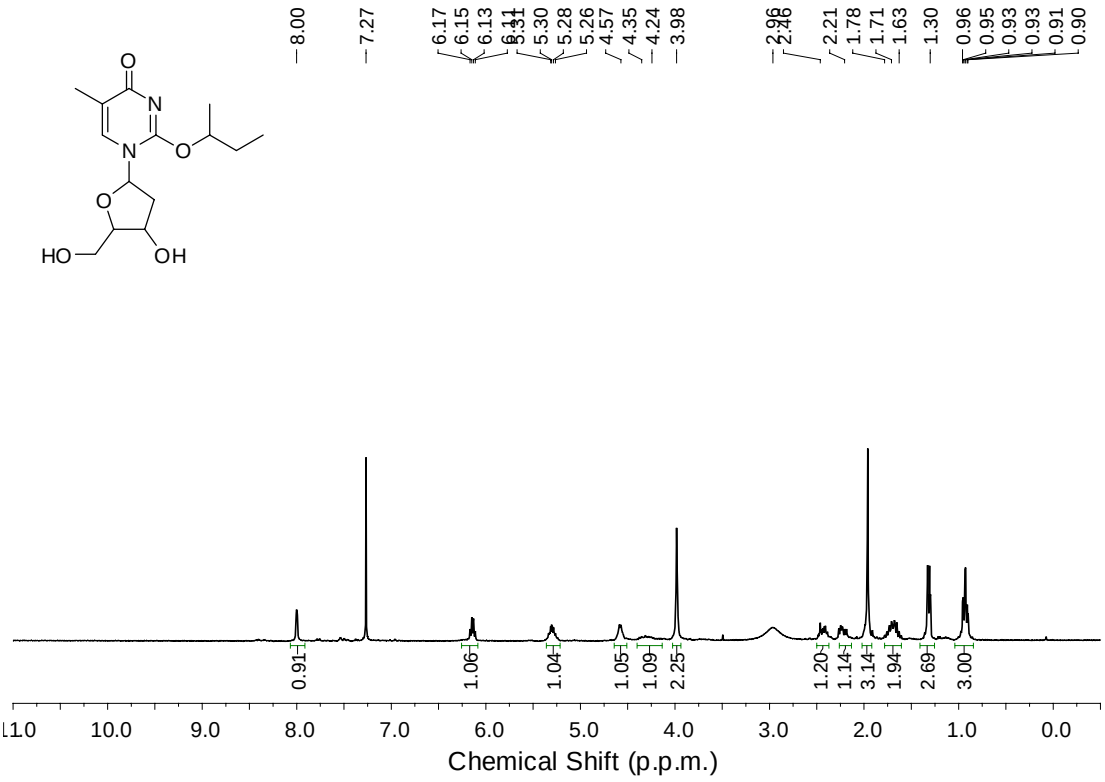


Figure S4. ^{1}H NMR spectrum of O^{2}-*s*BudT (300 MHz, $CDCl_3$, 25°C).

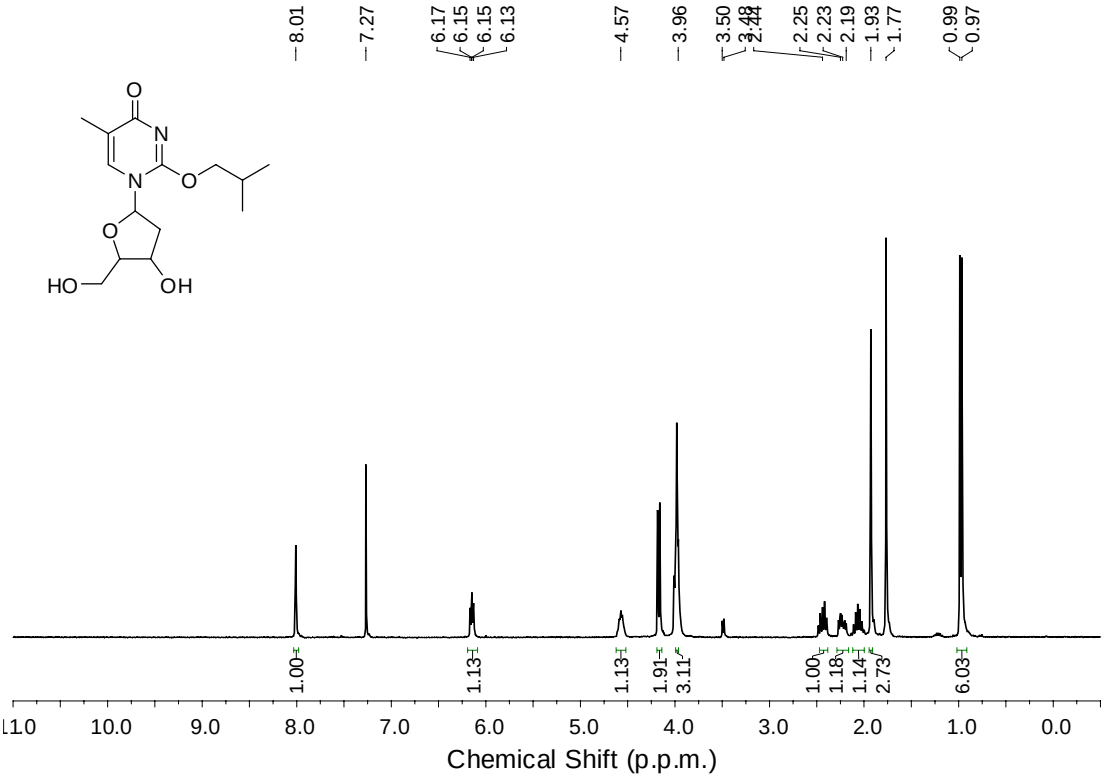


Figure S5. ^{1}H NMR spectrum of O^{2}-*i*BudT (300 MHz, $CDCl_3$, 25°C).

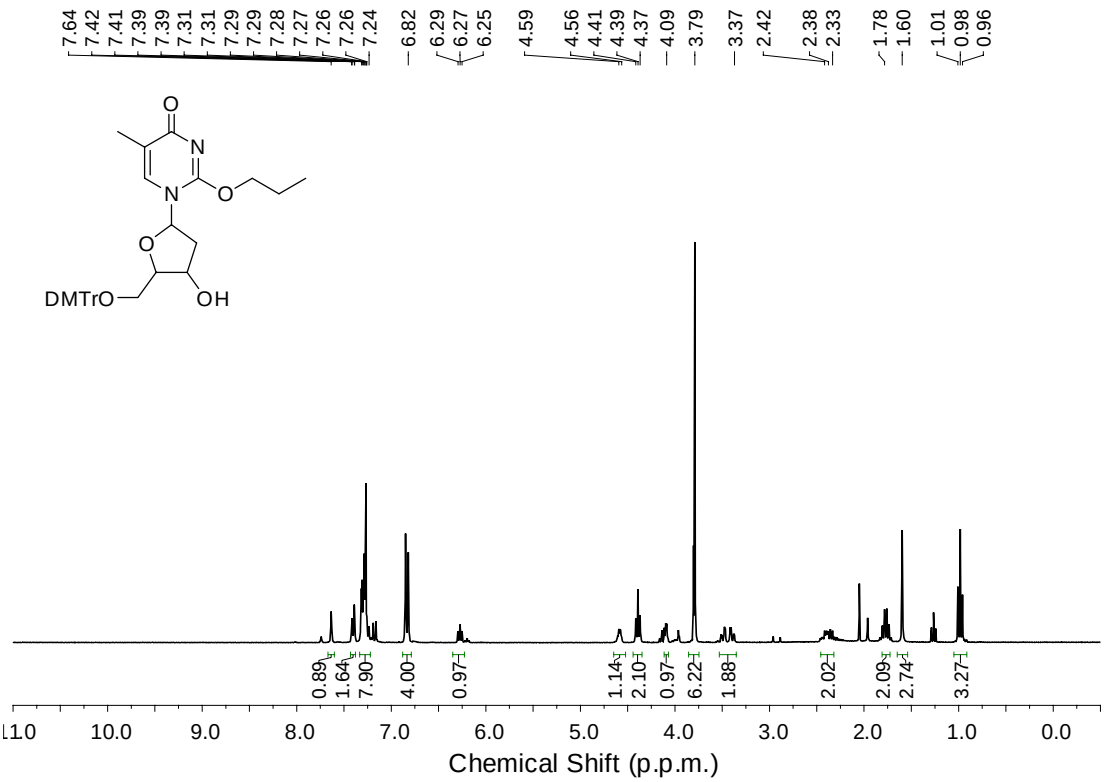


Figure S6. ^{1}H NMR spectrum of 5'-DMTr-O^2-*n*PrdT (300 MHz, $CDCl_3$, 25^{o}C).

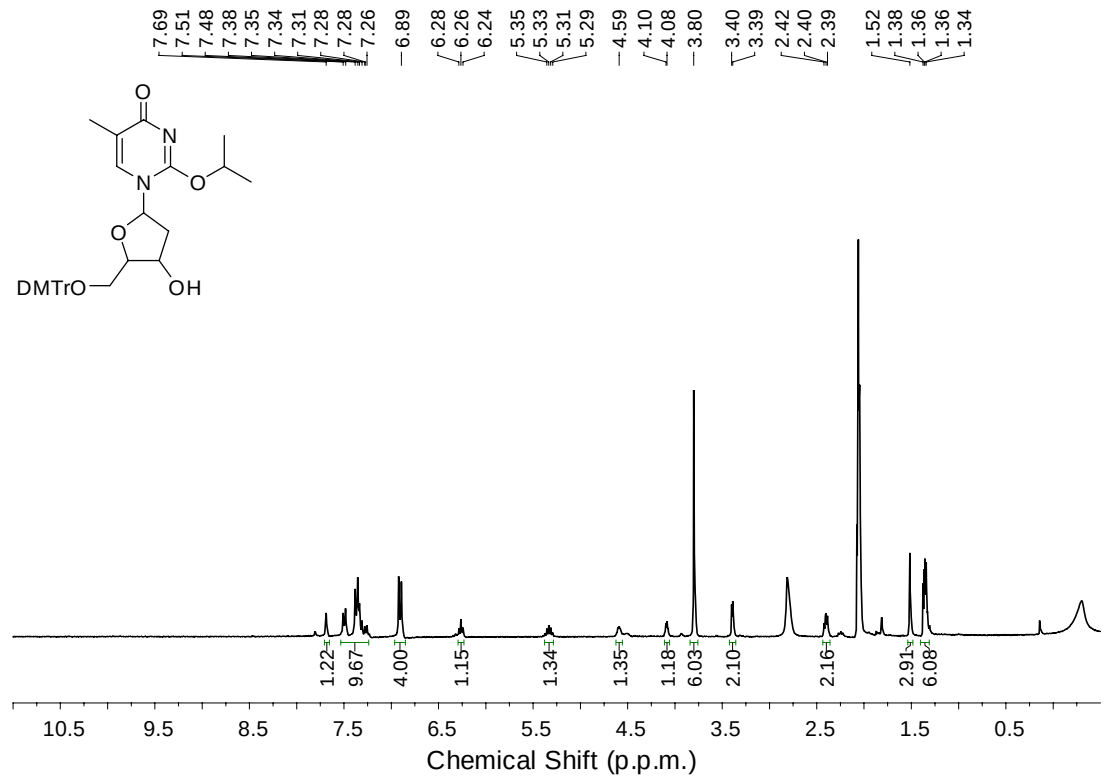


Figure S7. ^{1}H NMR spectrum of 5'-DMTr-O^2-*i*PrdT (300 MHz, acetone-d_6, 25^{o}C).

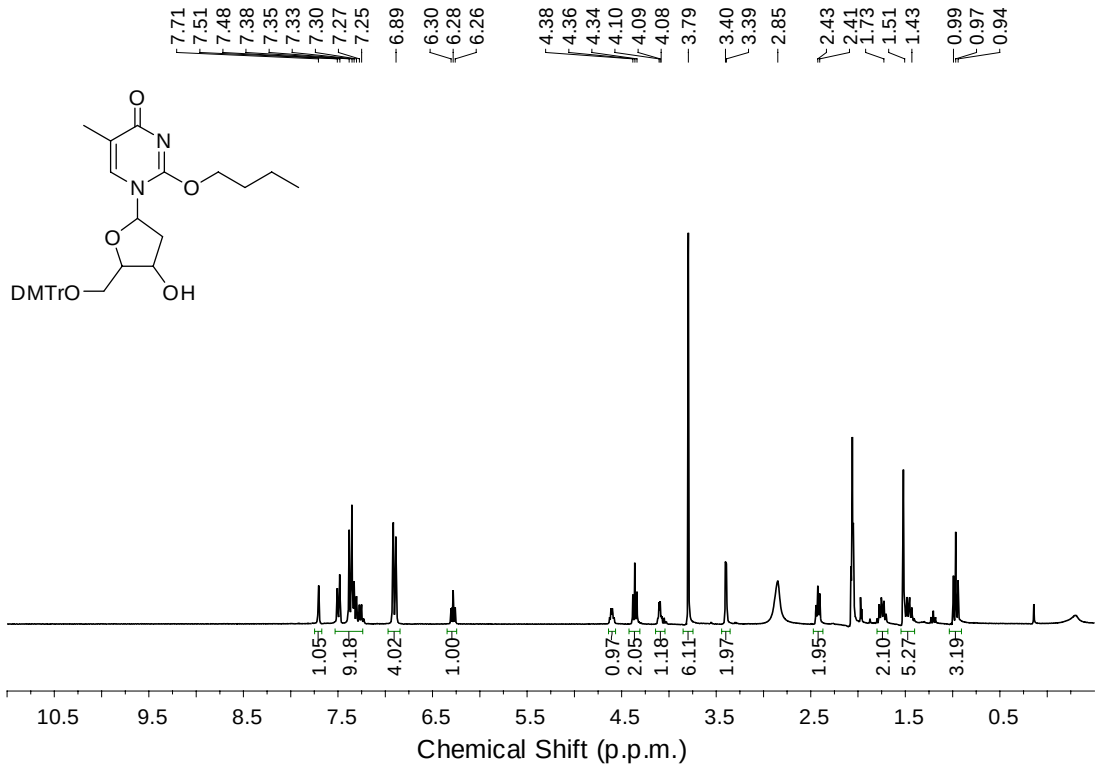


Figure S8. ^{1}H NMR spectrum of 5'-DMTr-O^{2}-*n*BudT (300 MHz, acetone-d_6, 25^{o}C).

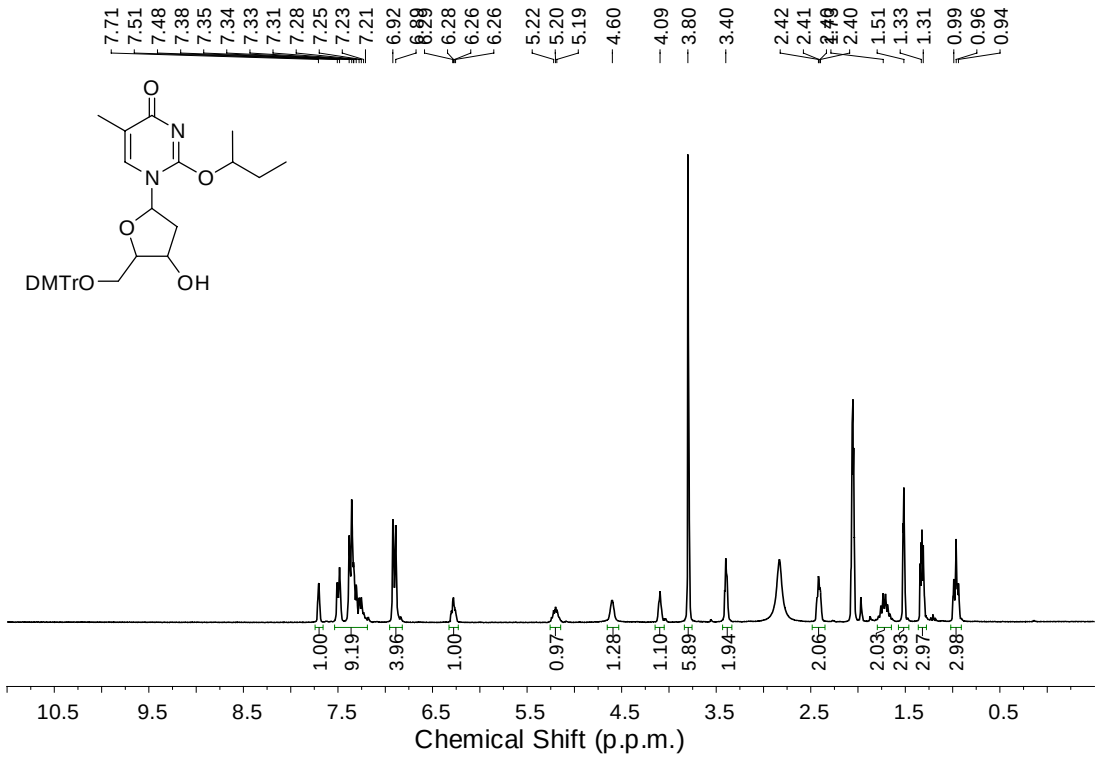


Figure S9. ^{1}H NMR spectrum of 5'-DMTr-O^{2}-sBudT (300 MHz, acetone-d_6, 25^{o}C).

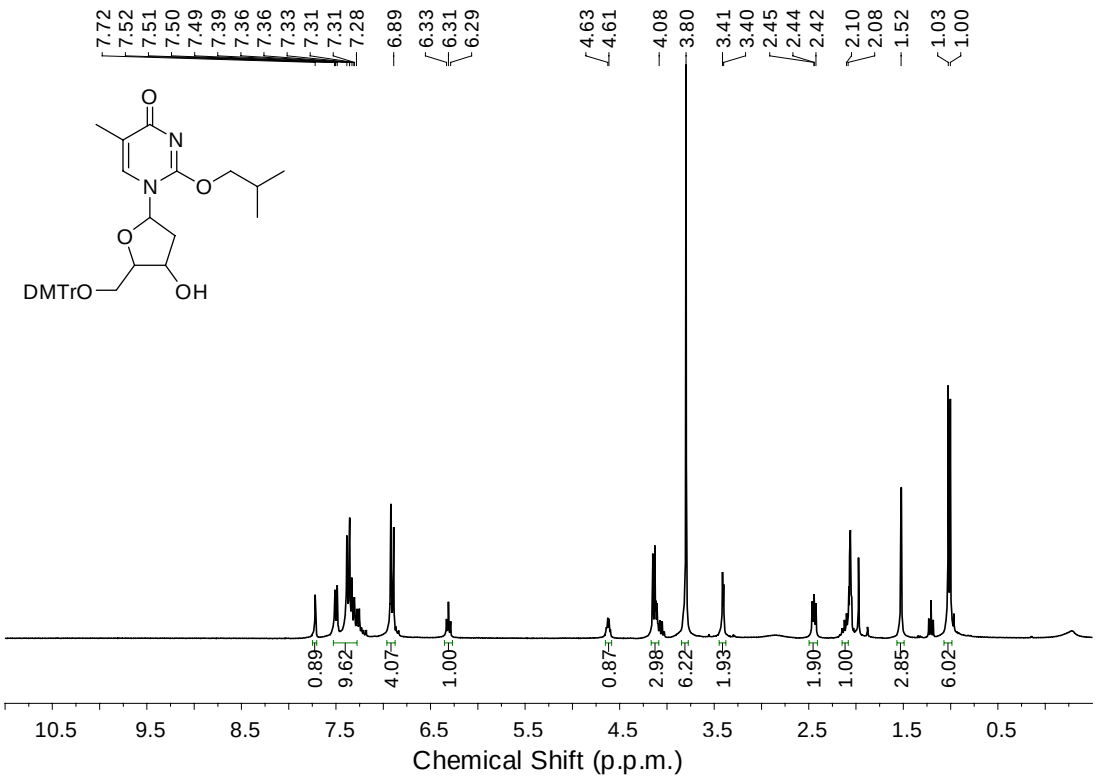


Figure S10. ^{1}H NMR spectrum of 5'-DMTr-O^2-*i*BudT (300 MHz, acetone-d_6, 25^{o}C).

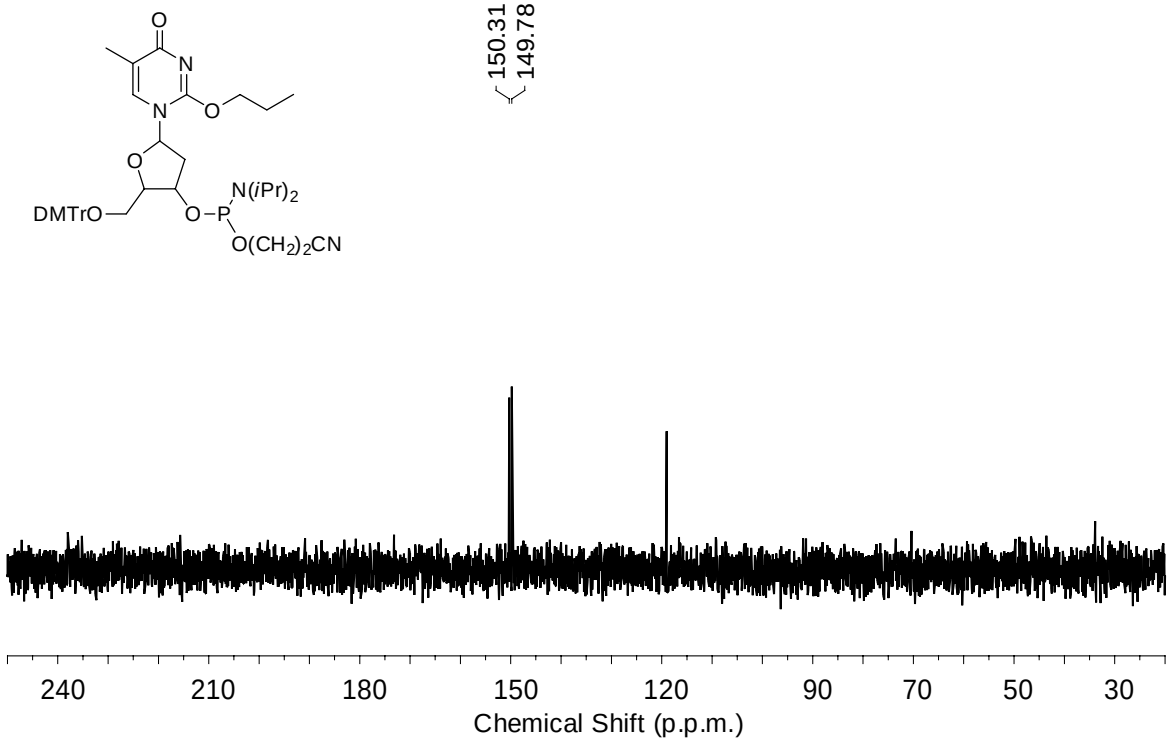


Figure S11. ^{31}P NMR spectrum of phosphoramidite building block of O^2-*n*PrdT (80 MHz, $CDCl_3$, 25ºC).

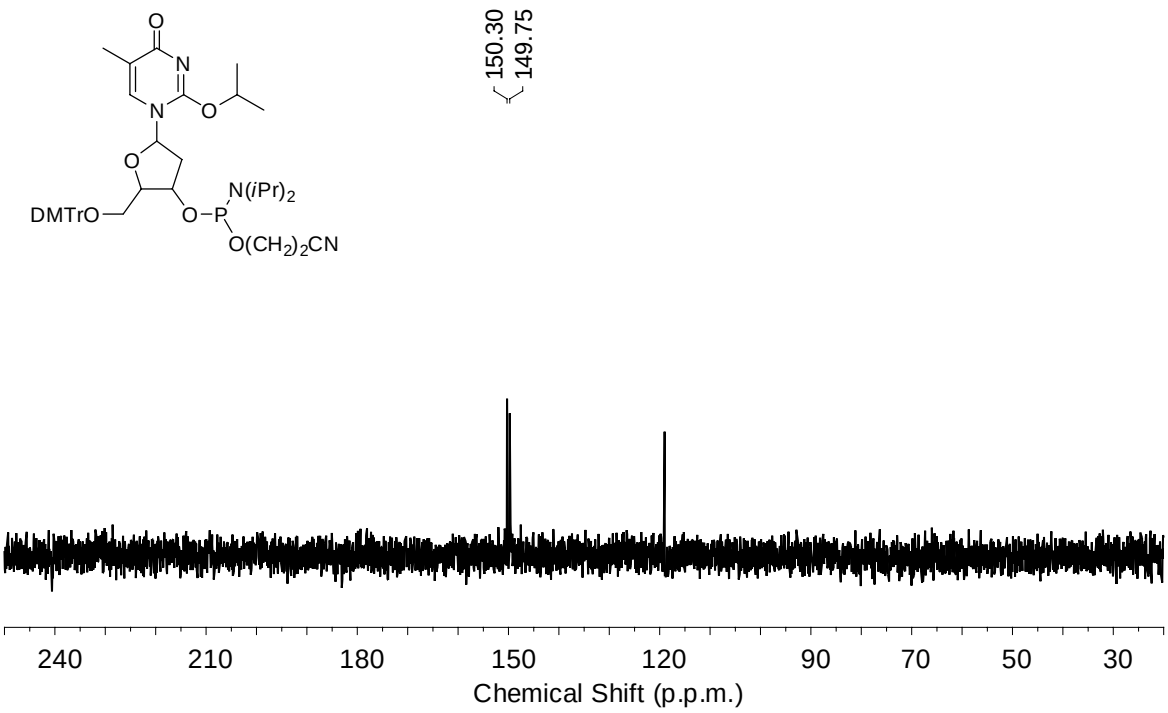


Figure S12. ^{31}P NMR spectrum of phosphoramidite building block of O^2-*i*PrdT (80 MHz, $CDCl_3$, 25ºC).

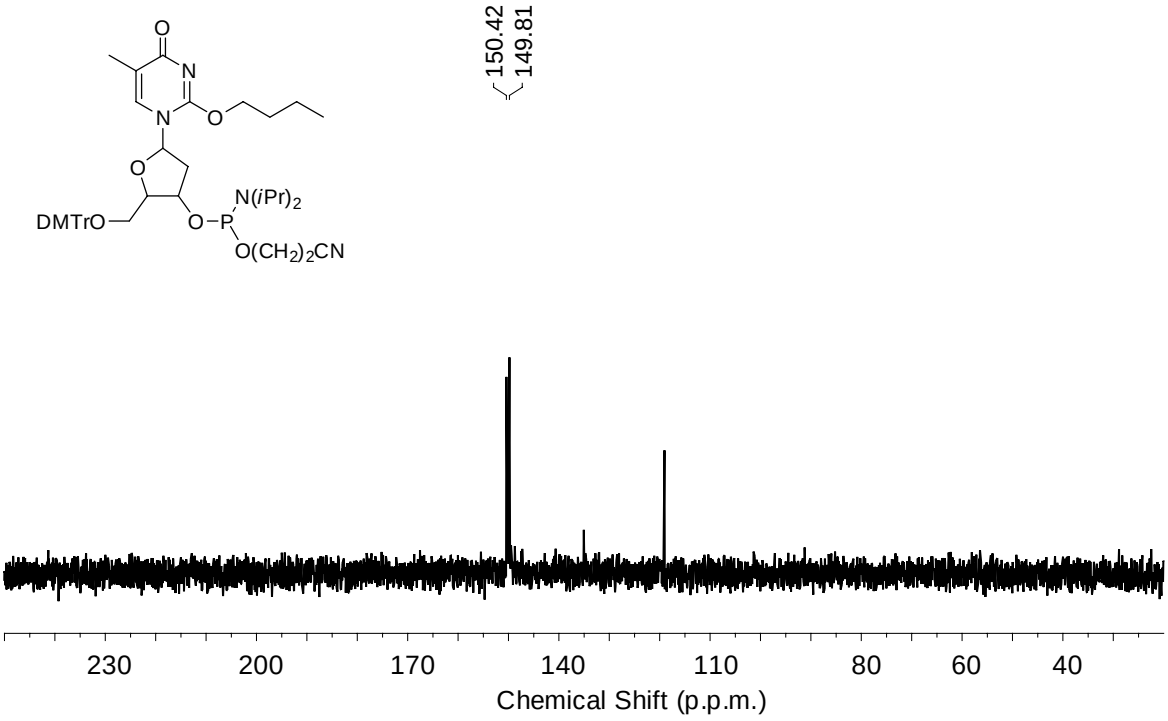


Figure S13. ^{31}P NMR spectrum of phosphoramidite building block of O^2-*n*BudT (80 MHz, $CDCl_3$, 25ºC).

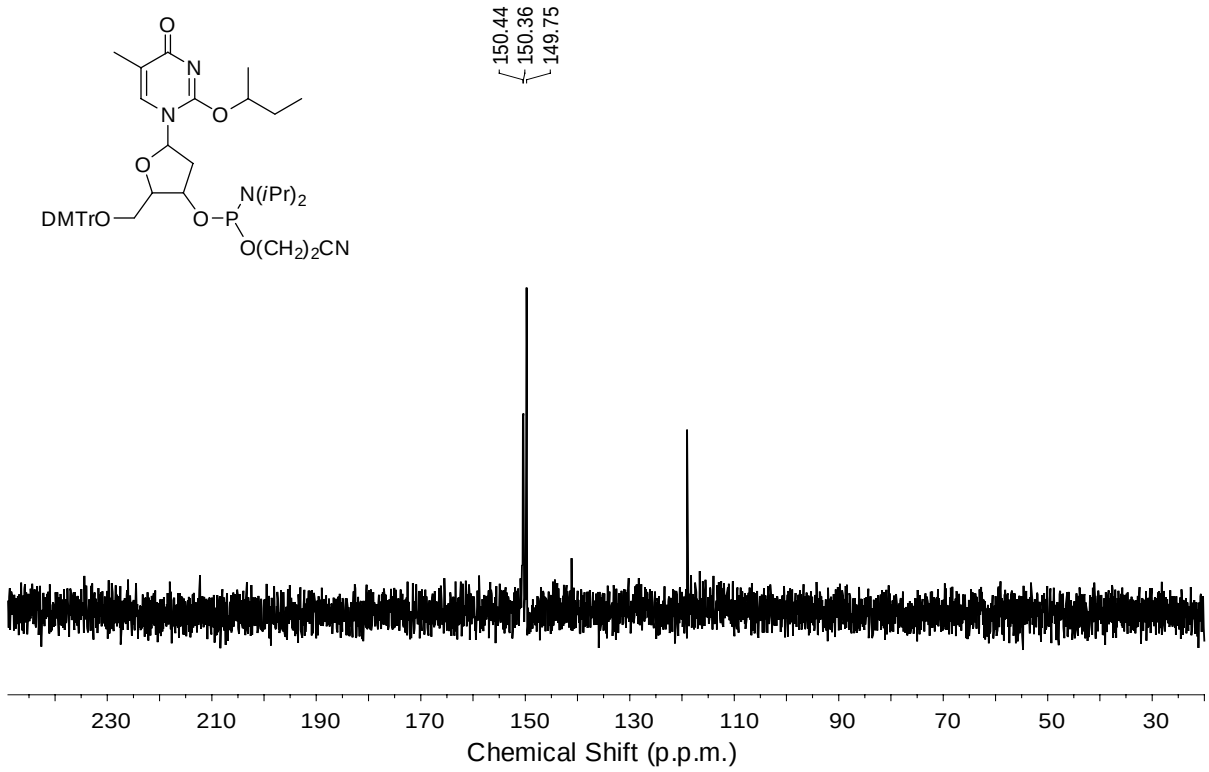


Figure S14. ^{31}P NMR spectrum of phosphoramidite building block of O^2-*s*BudT (80 MHz, $CDCl_3$, 25ºC).

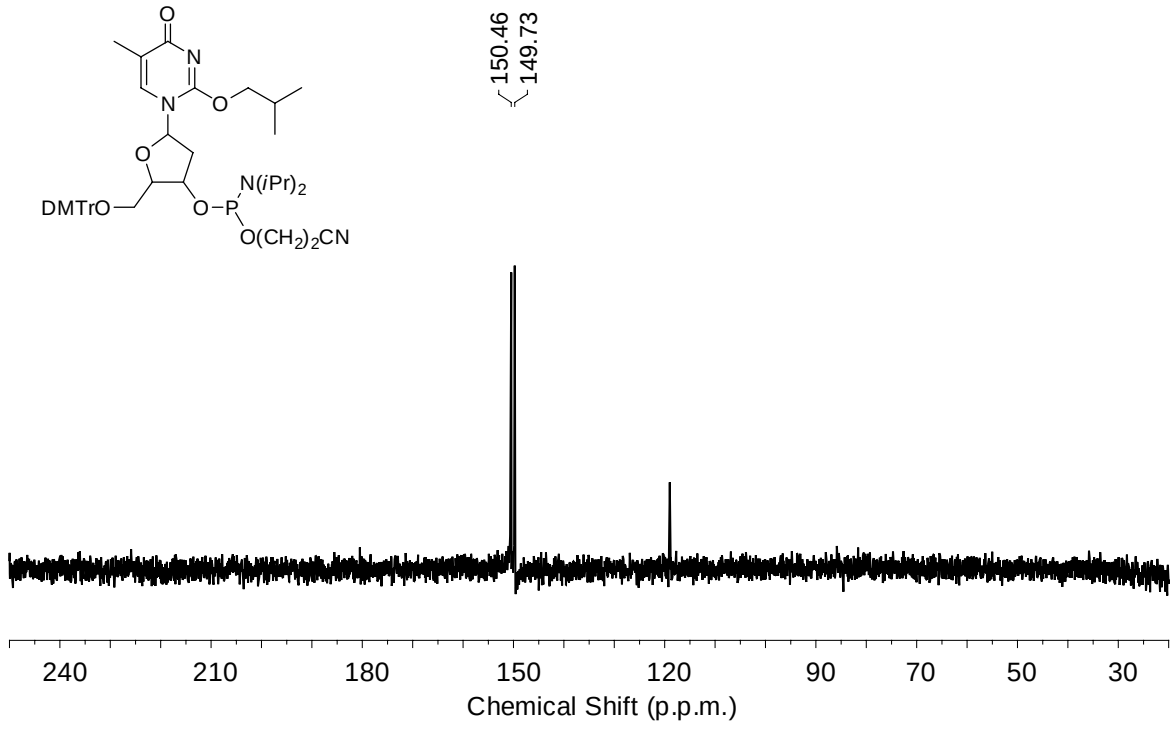


Figure S15. ^{31}P NMR spectrum of phosphoramidite building block of O^2-*i*BudT (80 MHz, $CDCl_3$, 25ºC).

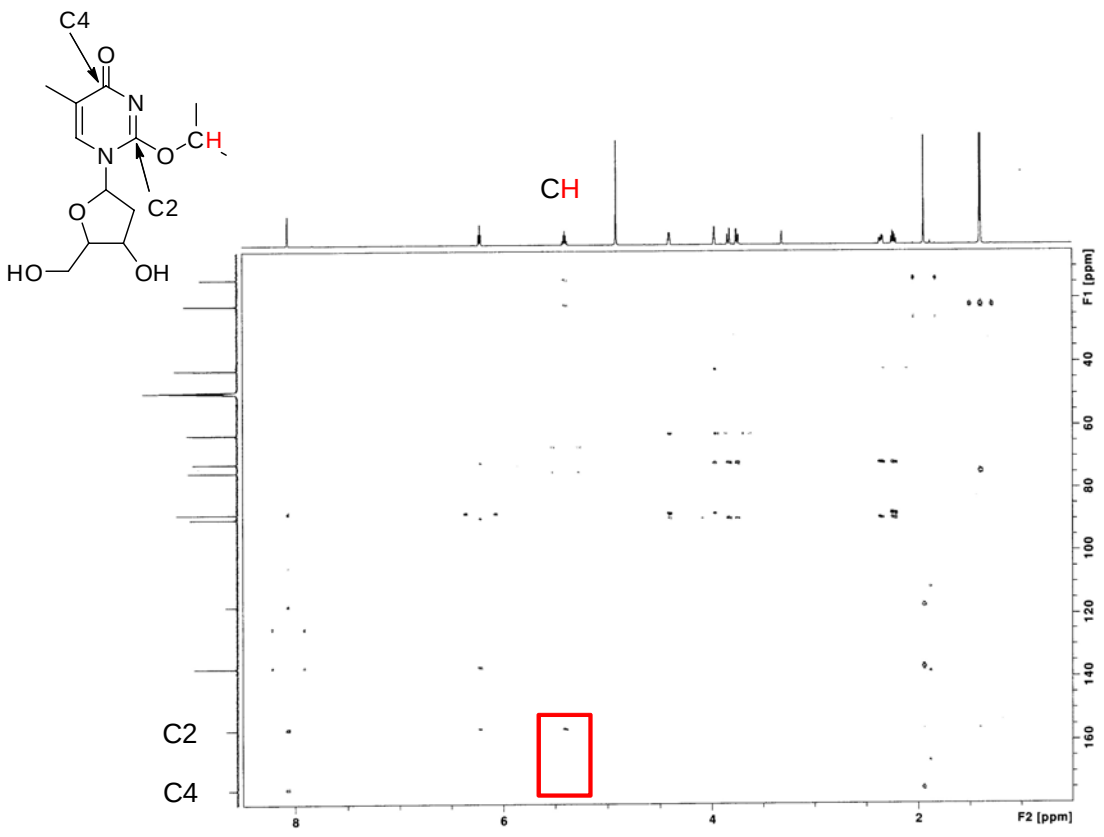


Figure S16. The 2-D HMBC spectrum of O^2-*i*PrdT (600 MHz, CD_3OD, 25°C) showing the correlation between the methine protons of the isopropyl functionality and the C2, but not the C4 carbon of the thymine ring.

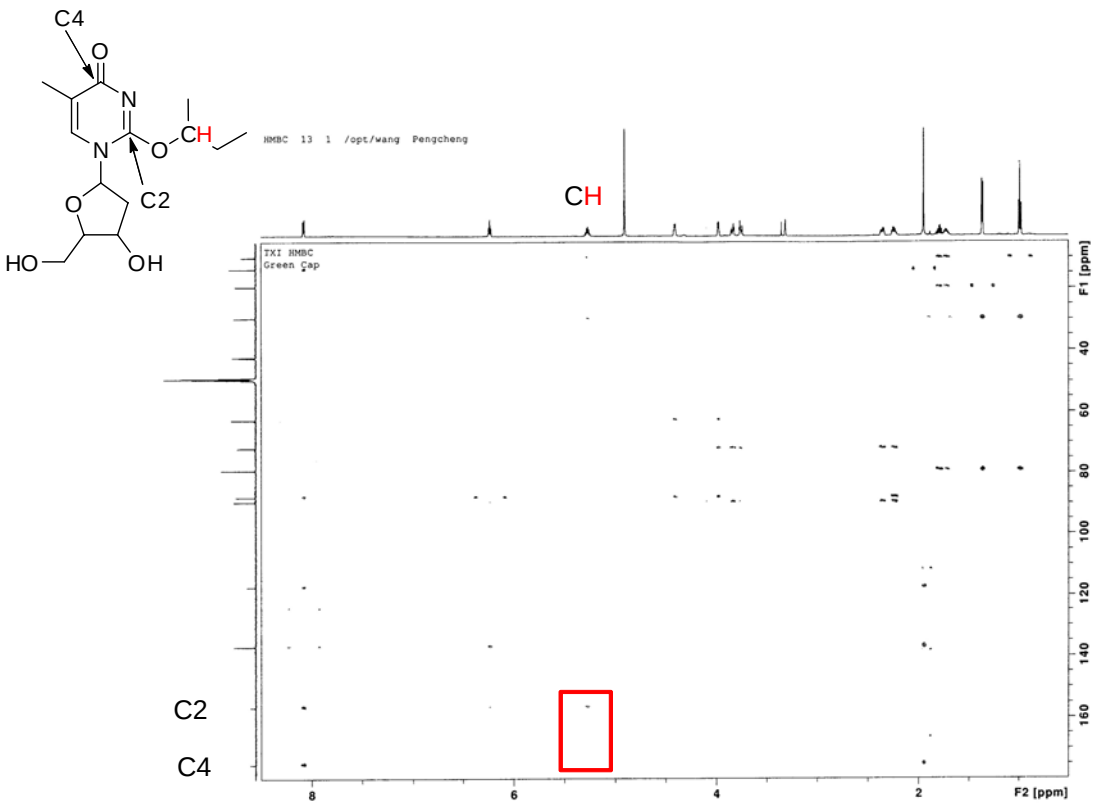


Figure S17. The 2-D HMBC spectrum of O^2-sBudT (600 MHz, CD_3OD, 25°C) showing the correlation between the methine proton of the *sec*-butyl functionality and the C2, but not the C4 carbon of the thymine ring.

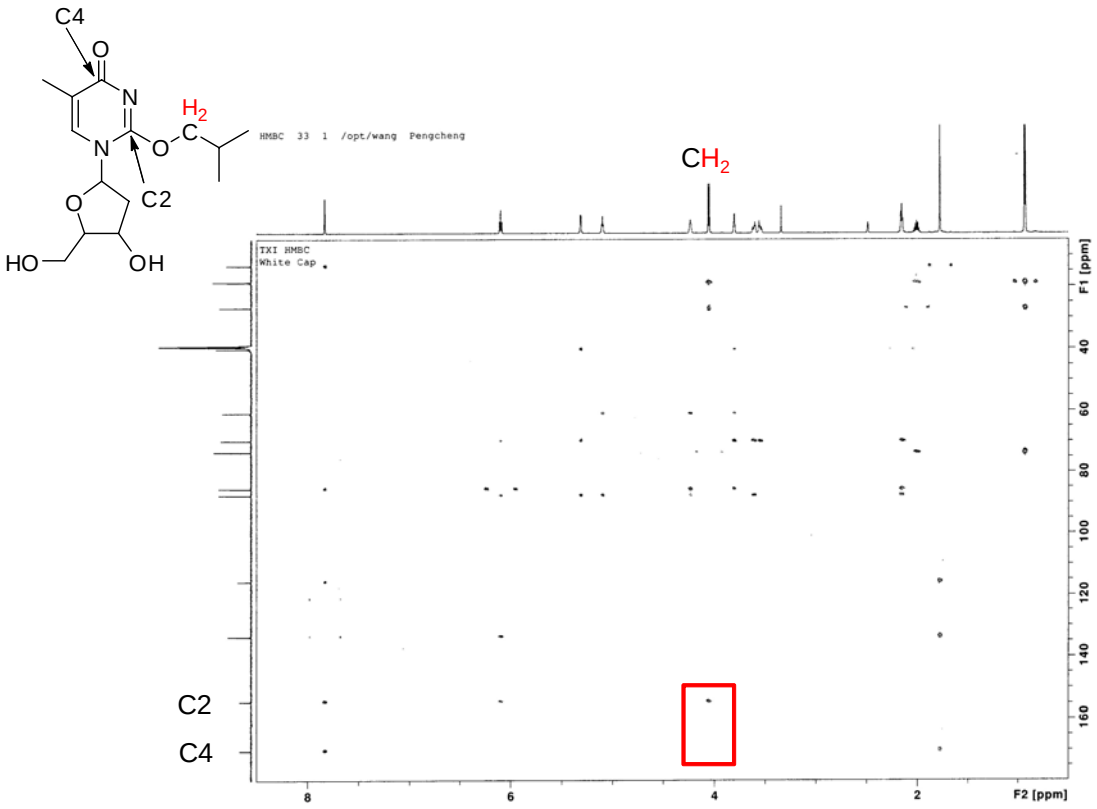
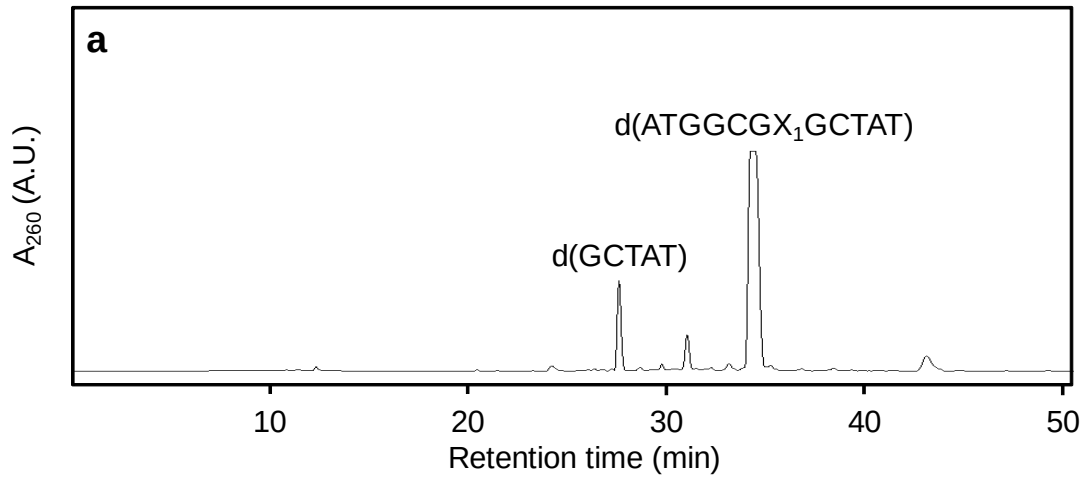
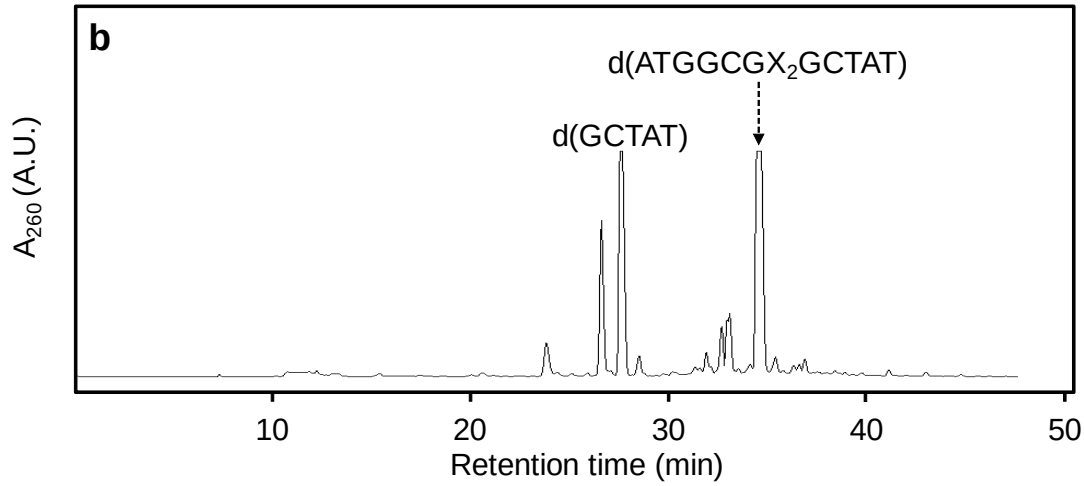
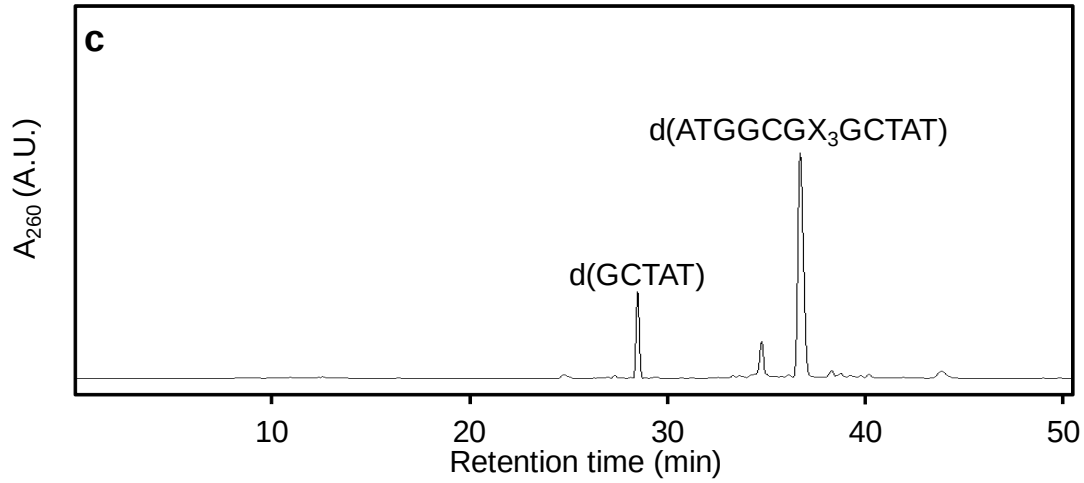


Figure S18. The 2-D HMBC spectrum of O^2-*i*BudT (600 MHz, DMSO-d_6, 25°C) showing the correlation between the methylene protons of the isobutyl functionality and the C2, but not the C4 carbon of the thymine ring.


a
d(ATGGCGX1GCTAT)
d(GCTAT)
A260 (A.U.)
10
20
30
40
50
Retention time (min)


b
d(ATGGCGX2GCTAT)
d(GCTAT)
A260 (A.U.)
10
20
30
40
50
Retention time (min)


c
d(ATGGCGX3GCTAT)
d(GCTAT)
A260 (A.U.)
10
20
30
40
50
Retention time (min)

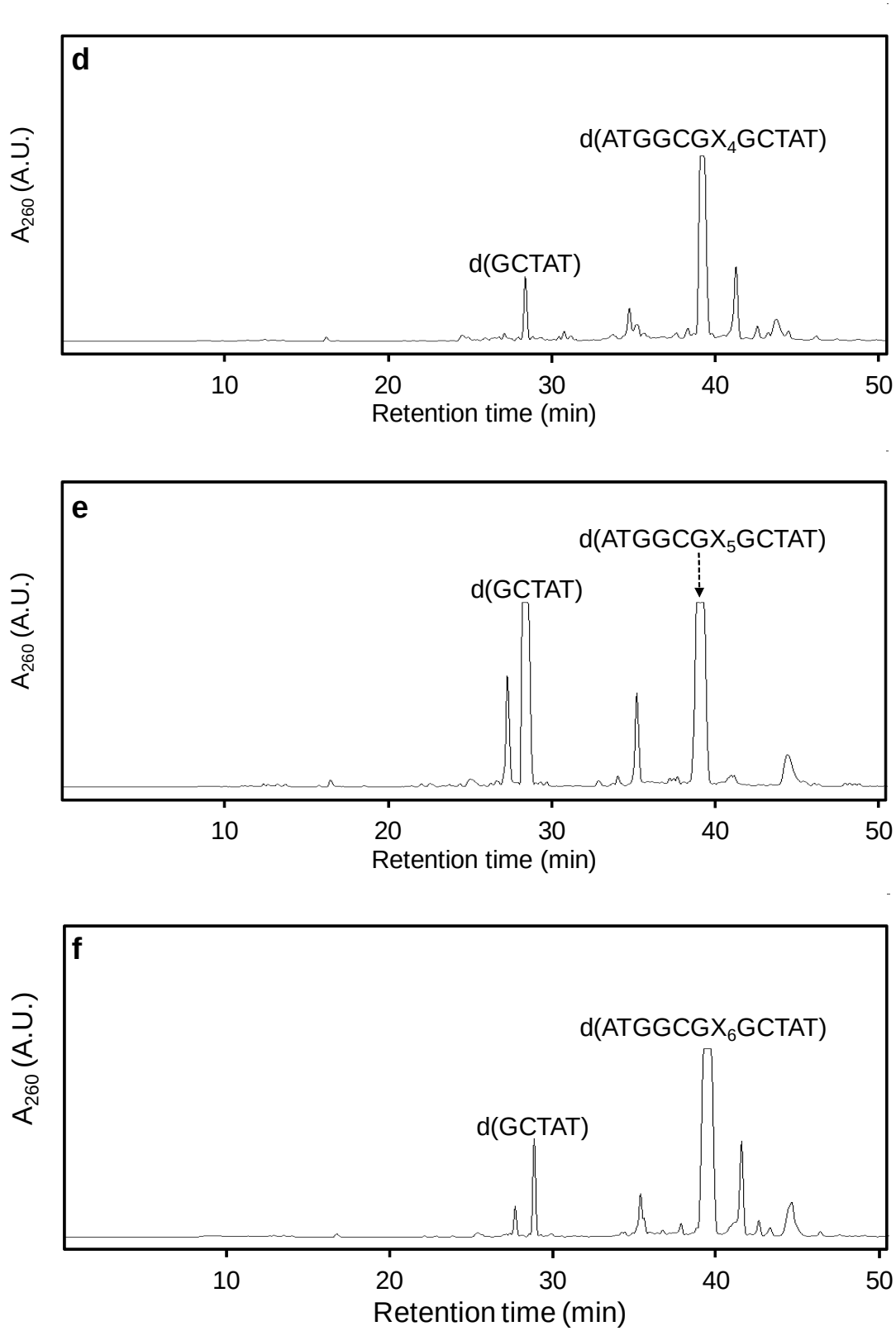


Figure S19. HPLC traces for the separation of the synthesized 12mer ODNs: (a) X_1= O^2-EtdT; (b) X_2 = O^2-*n*PrdT; (c) X_3 = O^2-*i*PrdT; (d) X_4 = O^2-*n*BudT; (e) X_5 = O^2-*s*BudT; (f) X_6 = O^2-*i*BudT.

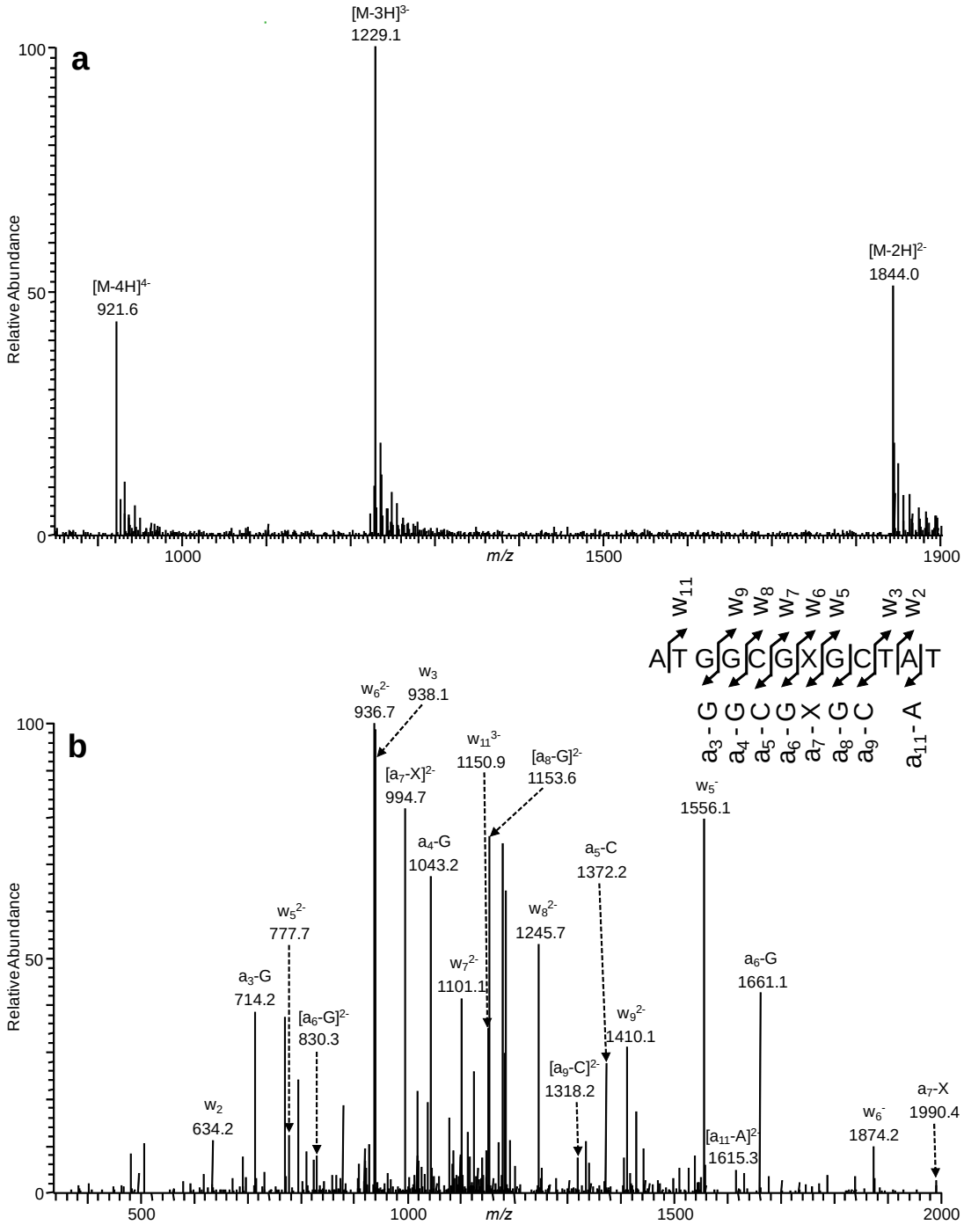


Figure S20. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-MedT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1229.1).

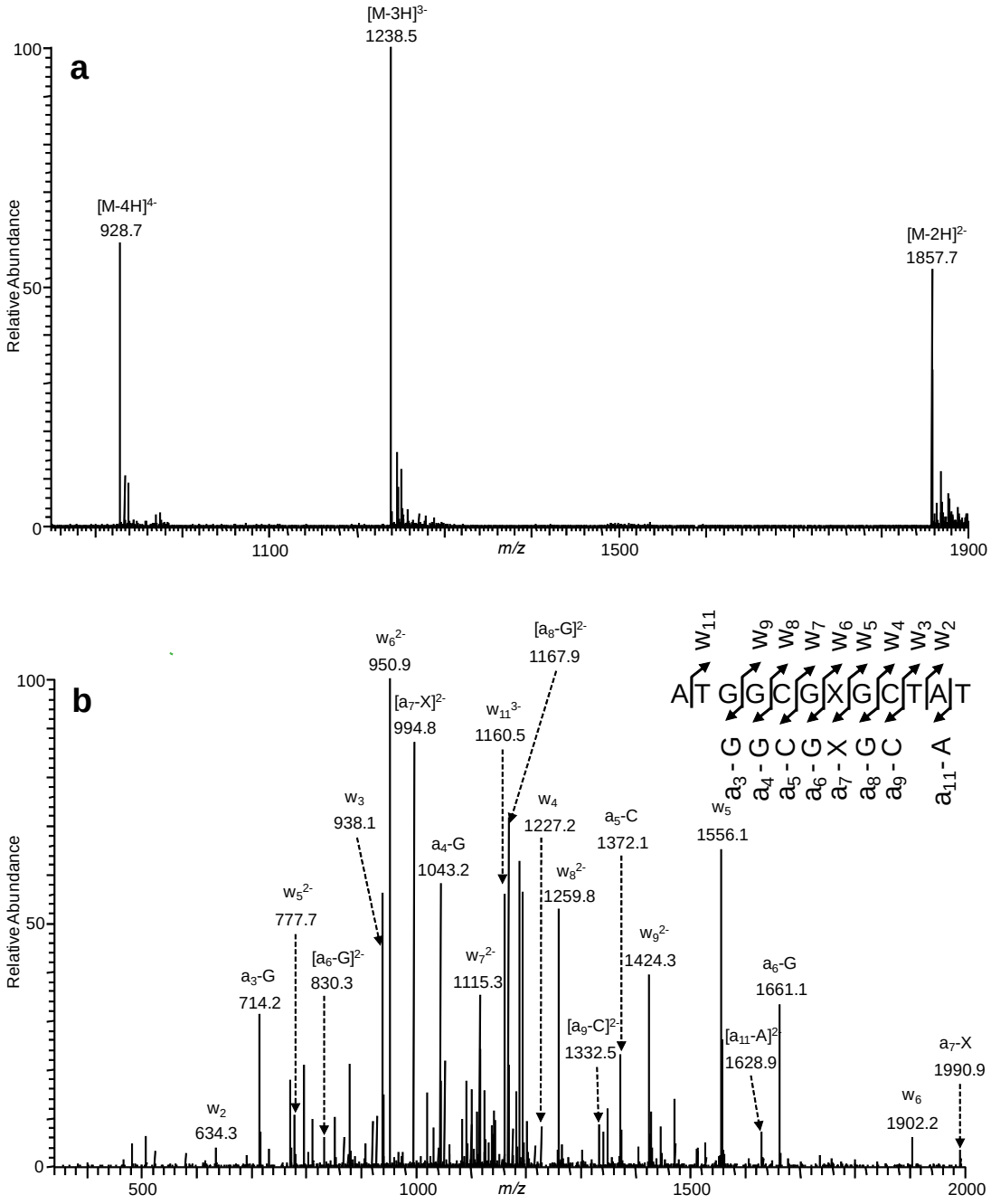


Figure S21. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-*n*PrdT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1238.5).

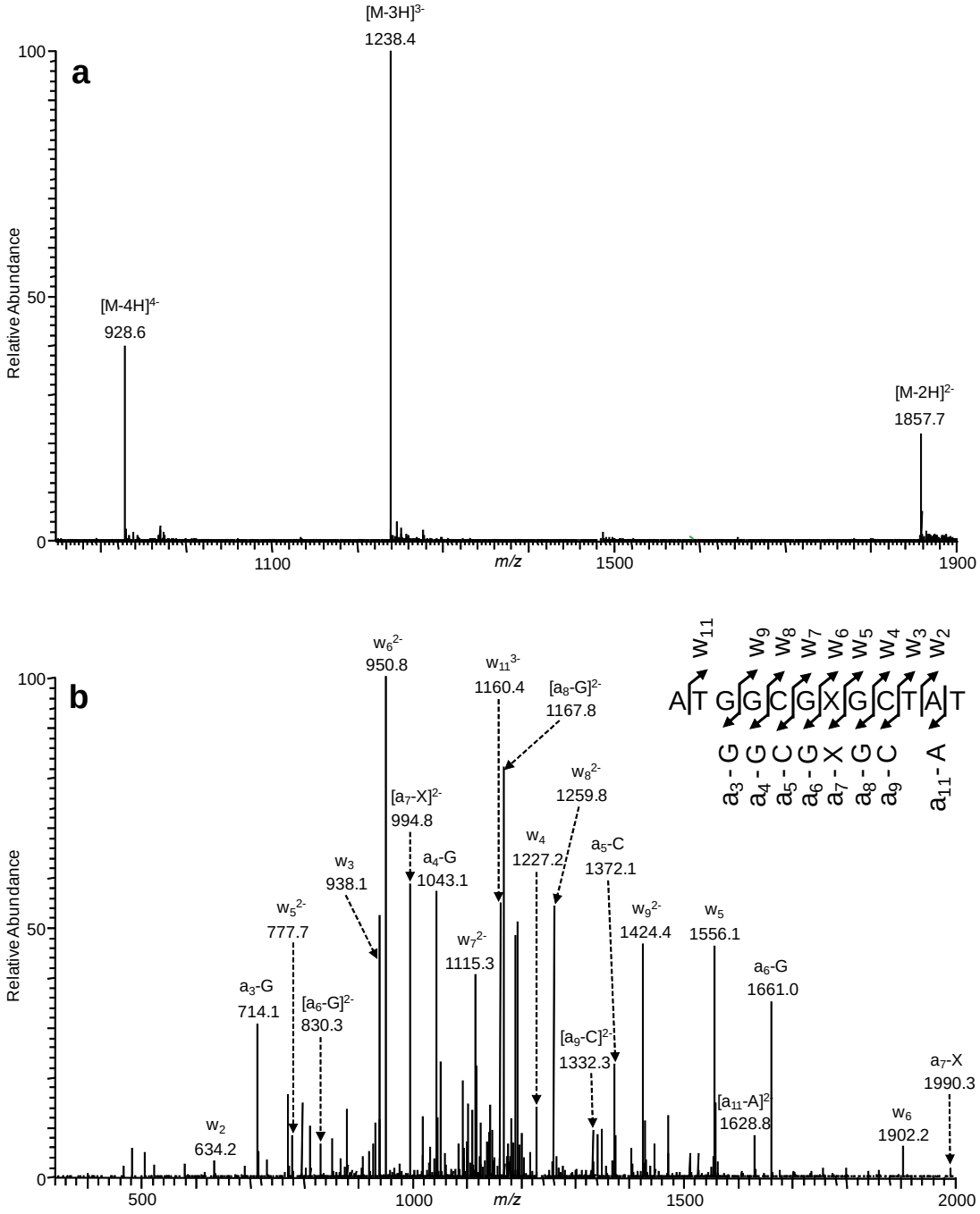


Figure S22. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-*i*PrdT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1238.4).

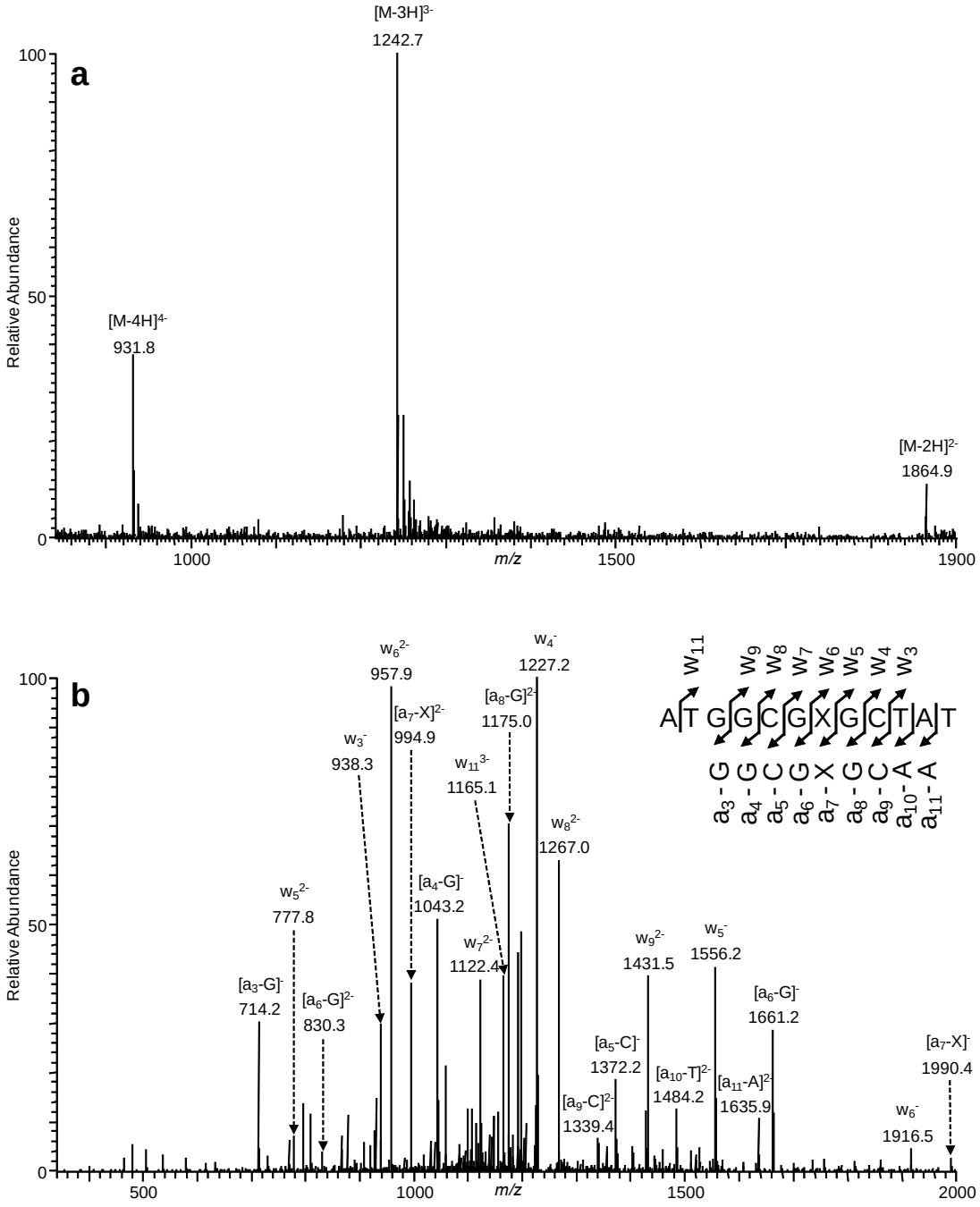


Figure S23. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-*n*BudT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1242.7).

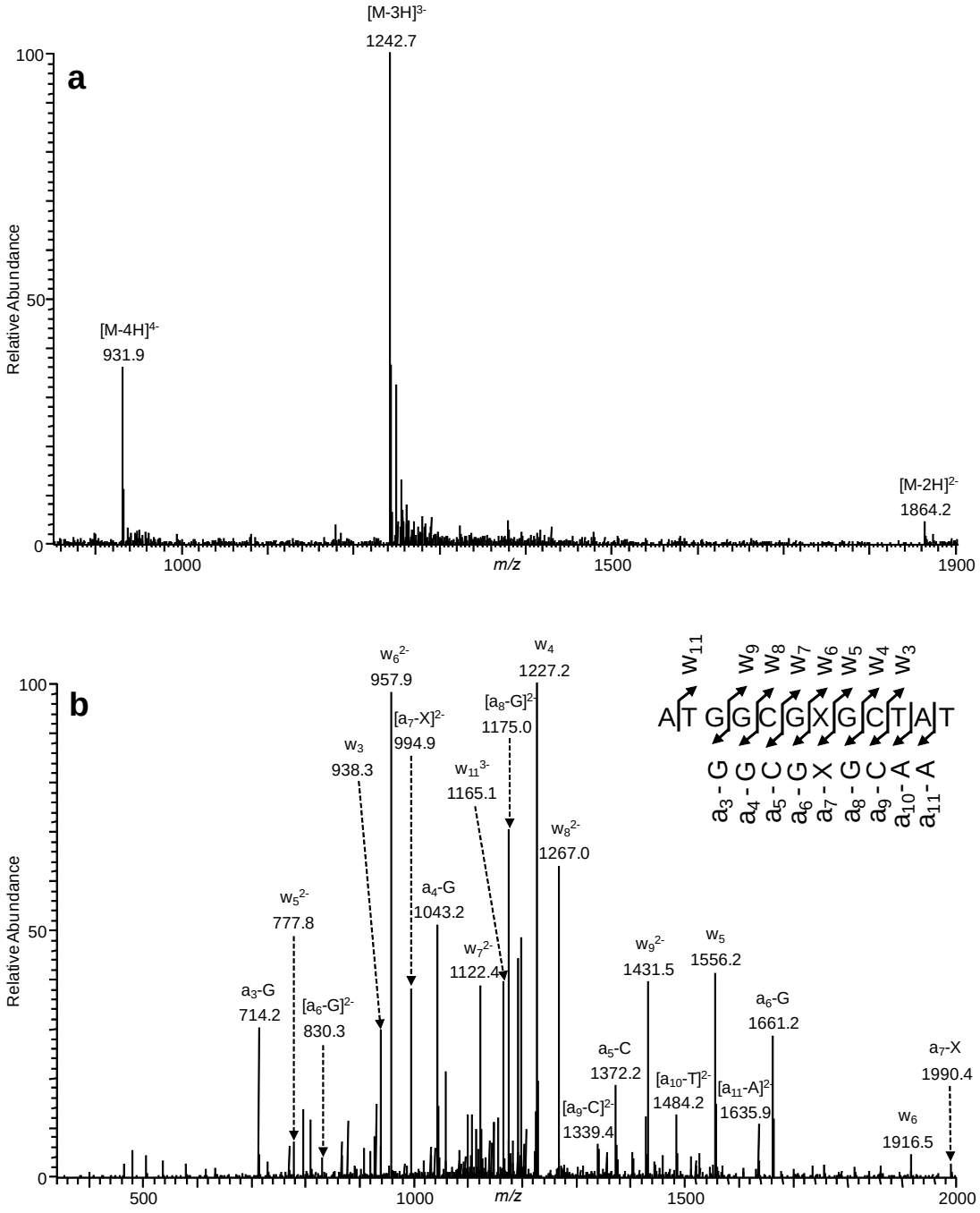


Figure S24. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-*s*BudT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1242.7).

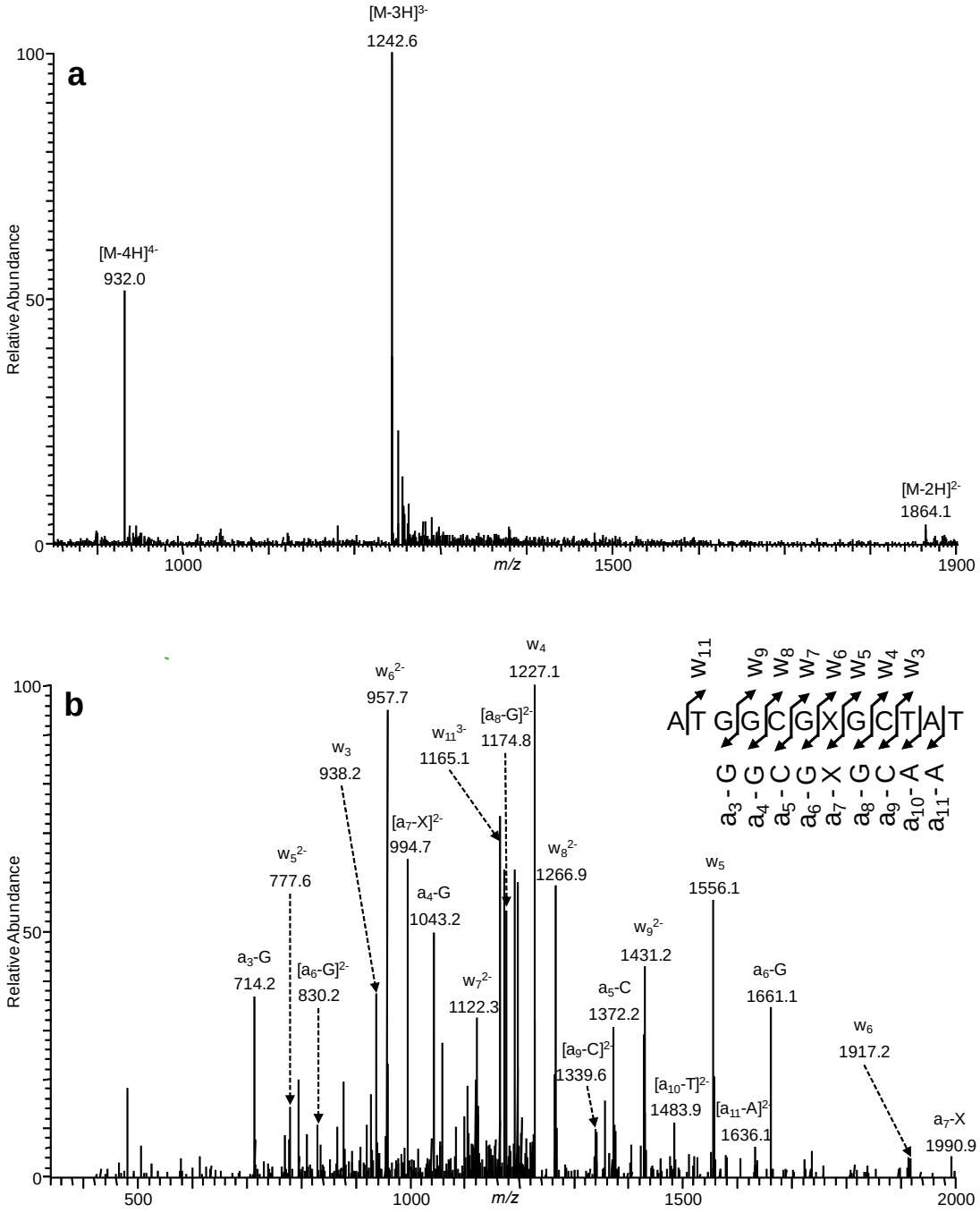
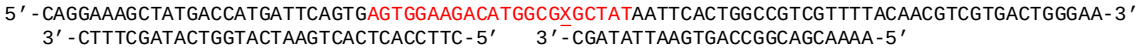


Figure S25. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-*i*BudT (a) Negative-ion ESI-MS; (b) the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1242.6).

a. Construction of Lesion-containing genome:



b. Construction of competitor genome:

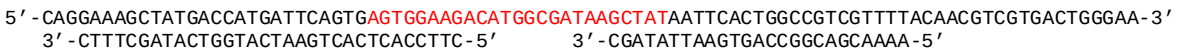


Figure S26. Schematic diagrams showing the construction of the lesion-containing (a) and competitor genomes. Displayed are the partial sequence of the linearized M13, the 22-mer lesion-containing or the 25-mer lesion-free insert (shown in red), and the two scaffolds employed for the ligation reactions. The lesion site is underlined (X).

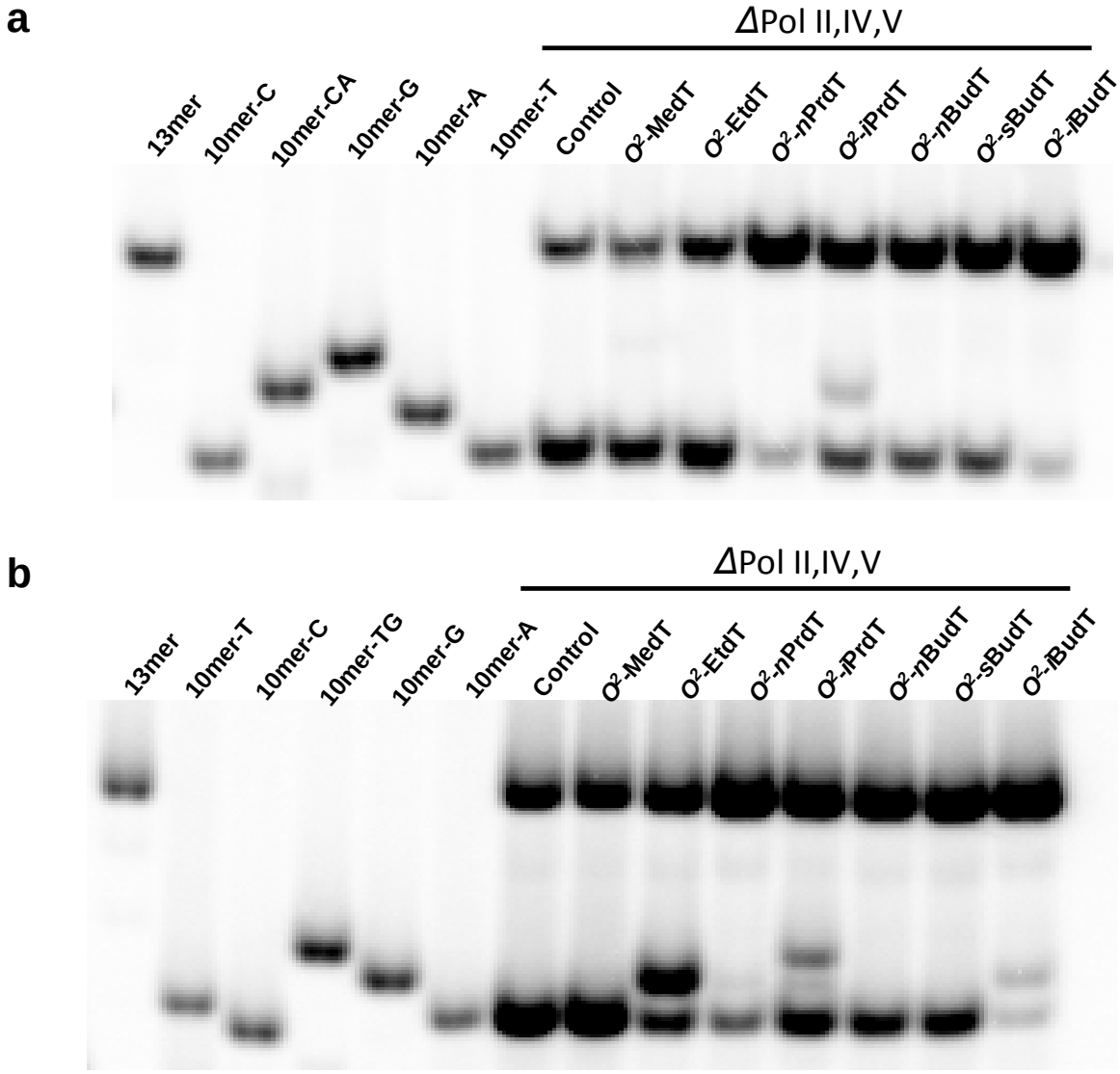


Figure S27. Native PAGE (30%) for monitoring the bypass efficiencies and mutation frequencies of O^2-alkyldT lesions in SOS-induced AB1157 *E. coli* cells that are depleted in Pol II, Pol IV, and Pol V. (a) Gel image showing the 13-mer and 10-mer products released from the top-strand (lesion-containing 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'-GGCGMGCTAT-3', with 'M' being A, C, G, and T, respectively, and 10mer-CA designates TG→CA tandem double mutation (See text). (b) Gel image showing the 13-mer and 10-mer products released from the bottom-strand (opposite to 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. The lesion-containing genomes were mixed individually with the competitor genome at molar ratios of 10:1 or 20:1 for the transfection experiments (See Materials and Methods).

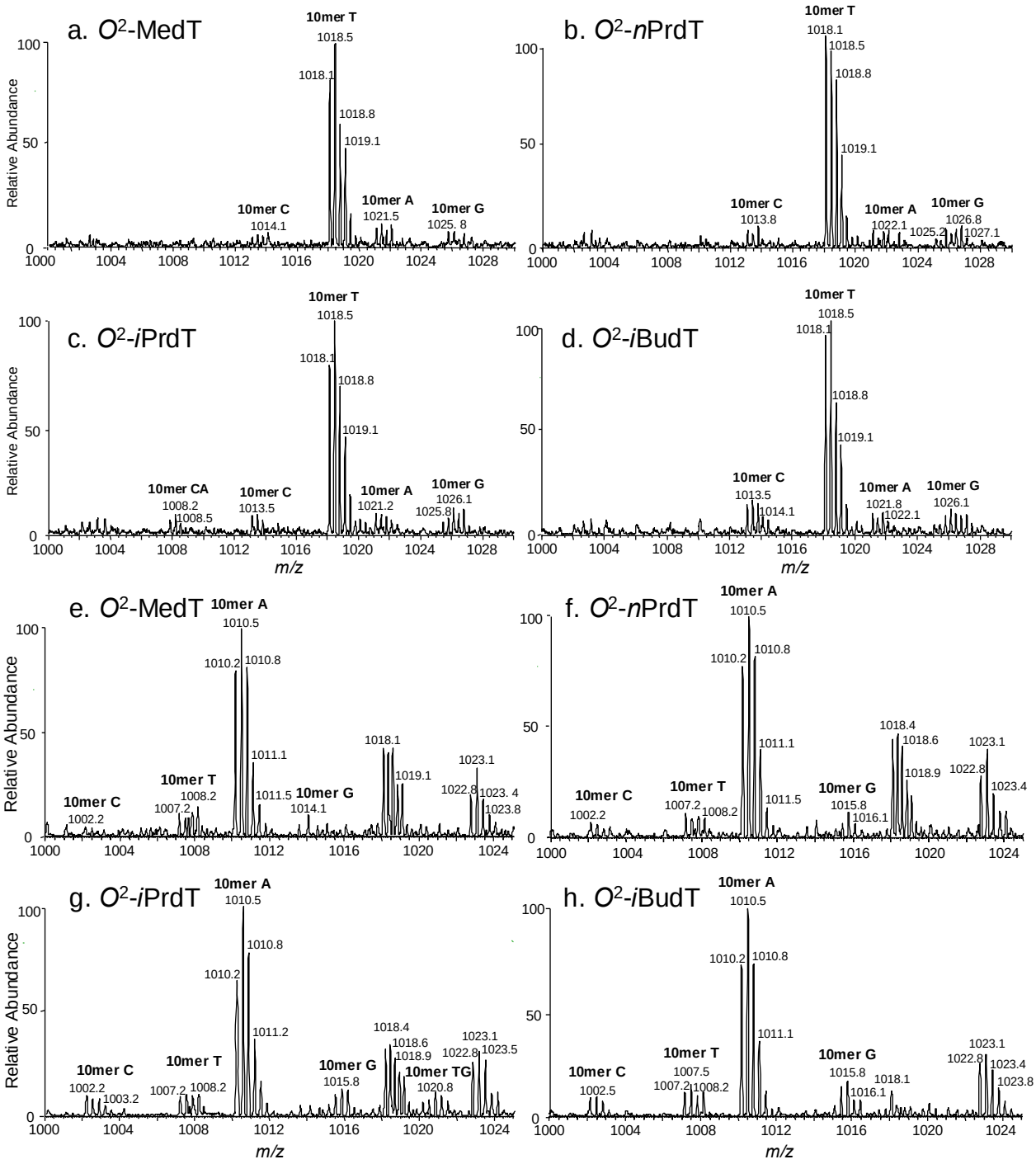


Figure S28. Higher-resolution "ultra zoom" scan ESI-MS for the restriction fragments for the PCR products from the replication of O^2-MedT (a, e), O^2-*n*PrdT (b, f), O^2-*i*PrdT (c, g) and O^2-*i*BudT (d, h)-bearing single-stranded M13 genomes in SOS-induced wild-type AB1157 cells. Displayed in (a)-(d) are the $[M - 3H]^{3-}$ ions for the lesion-containing strand products, and in (e)-(h) are the $[M - 3H]^{3-}$ ions for the complementary strand products. All the mutagenic products were further confirmed by MS/MS analyses, and representative MS/MS results for O^2-iPrdT are shown in Figures S29-S30.

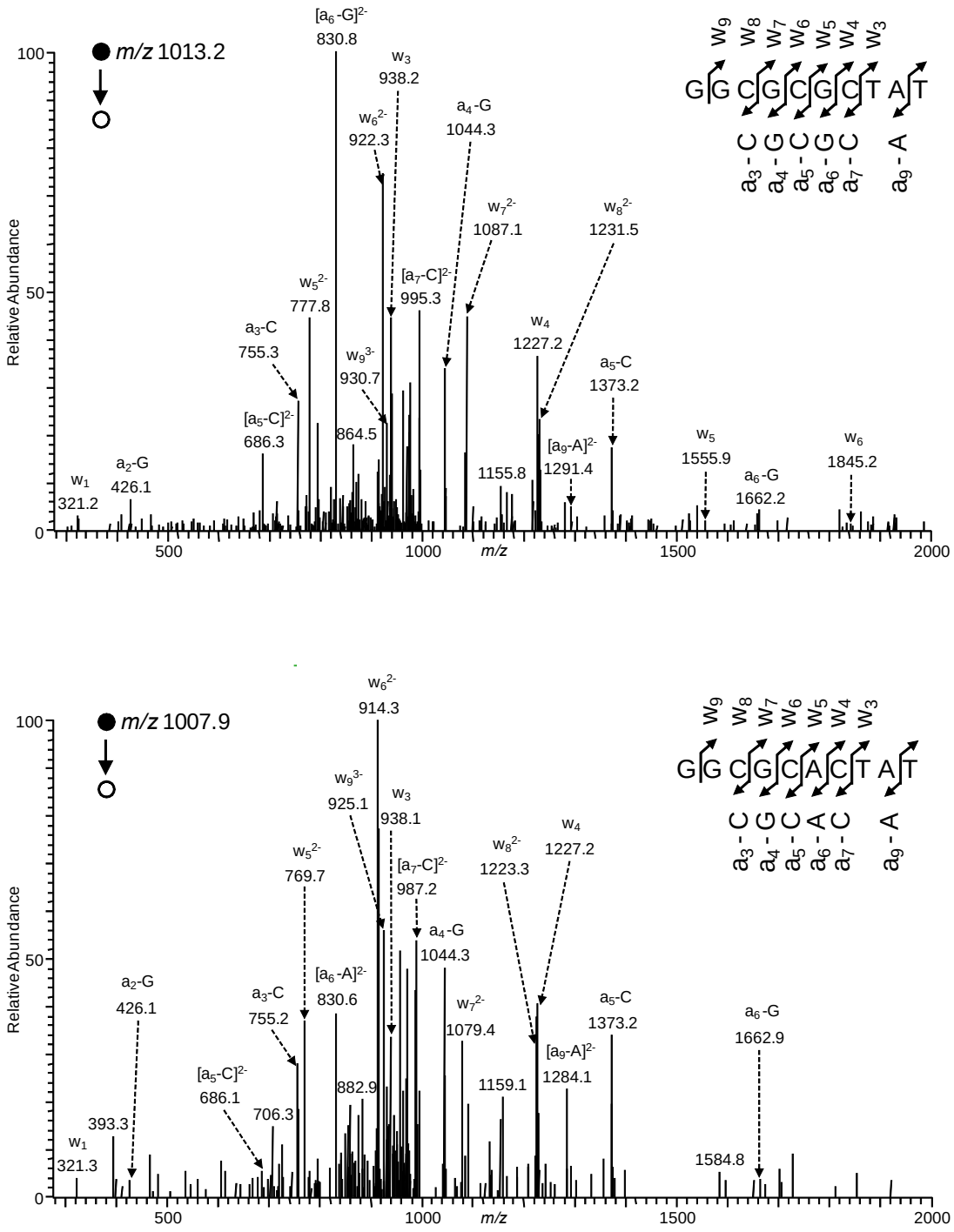


Figure S29. LC-MS/MS for the identification of restriction fragments of PCR products. MS/MS for the $[M - 3H]^{3-}$ ions of 10 mer C (top) and 10 mer CA (bottom) products from for the progeny of O^2-*i*PrdT-containing genome in SOS-induced wild-type AB1157 cells. Because of overlap in signal of the $[M - 3H]^{3-}$ ion of 10 mer A and the $[M - 4H + Na^+]^{3-}$ ion of 10 mer C as well as between the $[M - 3H]^{3-}$ ion of 10 mer G and the $[M - 4H + Na^+]^{3-}$ ion of 10 mer T, the T → A and T → G mutation products were monitored using the complementary strand (Figure S30).

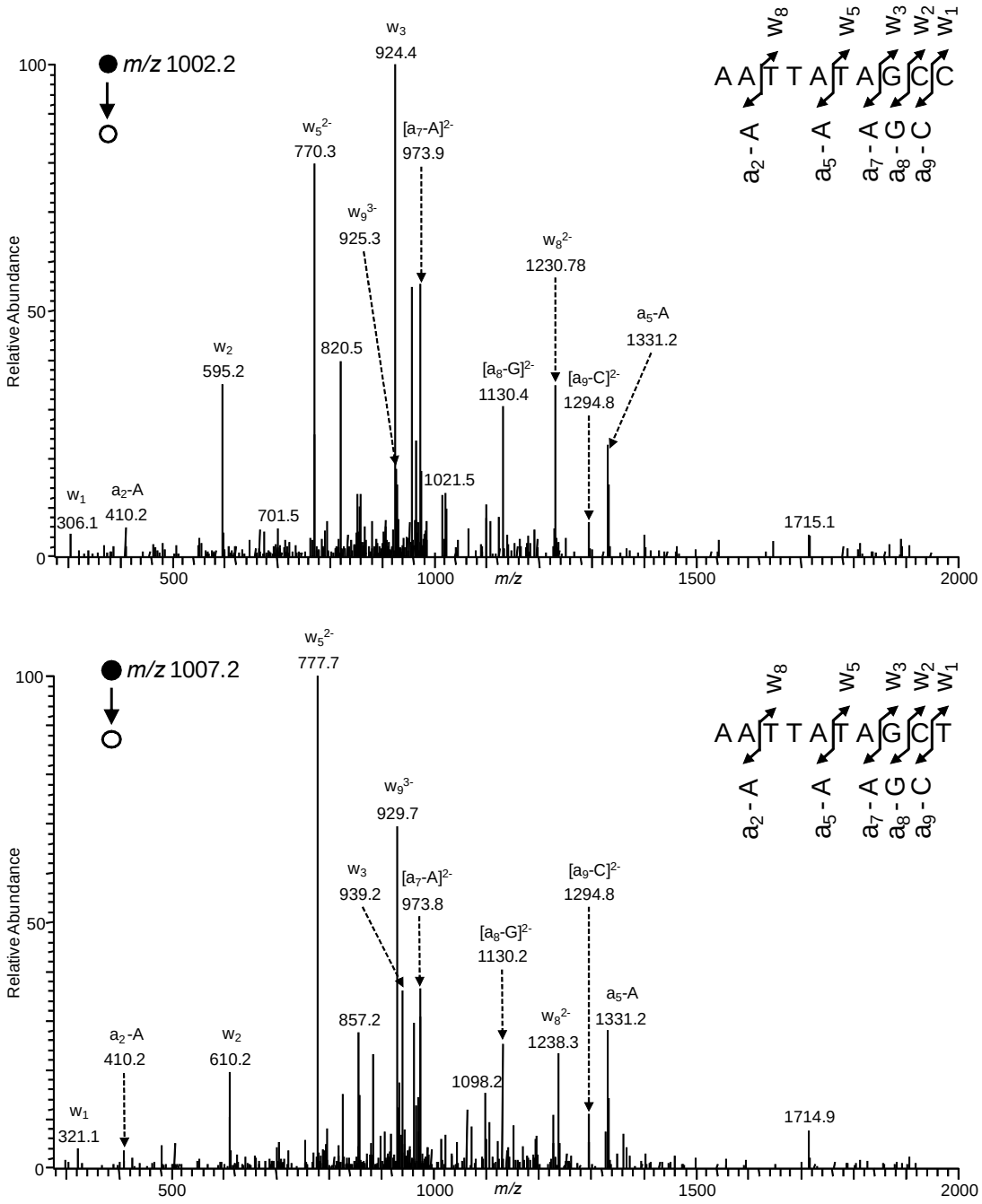


Figure S30. LC-MS/MS for the identification of restriction fragments of PCR products in SOS-induced wild-type AB1157 cells. MS/MS for the $[M - 3H]^{3-}$ ions of 10 mer C (corresponding to T → G mutation at the lesion site, top), 10 mer T (corresponding to T→A mutation at the lesion site, bottom).