Electronic Supplementary Material (ESI) for Chemical Science.

Site- and Degree-Specific C-H Oxidation on 5-Methylcytosine Homologues for Probing Active DNA Demethylation

Sam Kavoosi,[ab] Babu Sudhamalla,[ab] Debasis Dey,[ab] Kirsten Shriver,[a] Simran Arora,[a] Sushma Sappa,[a] Kabirul Islam*[a]

[a]Department of Chemistry, Pittsburgh, Pittsburgh PA 15260

([b]Equal contribution)

1. General materials, methods and equipment S2
2. Synthesis and characterization of phosphoramidites **1**-**7** S3
3. Synthesis and purification of oligonucleotides S22
4. Expression and purification of proteins S23
5. Protocols for biochemical assays S26
6. Acquisition and analysis of MALDI data S30
7. Supplementary figures and tables S30
8. References S79

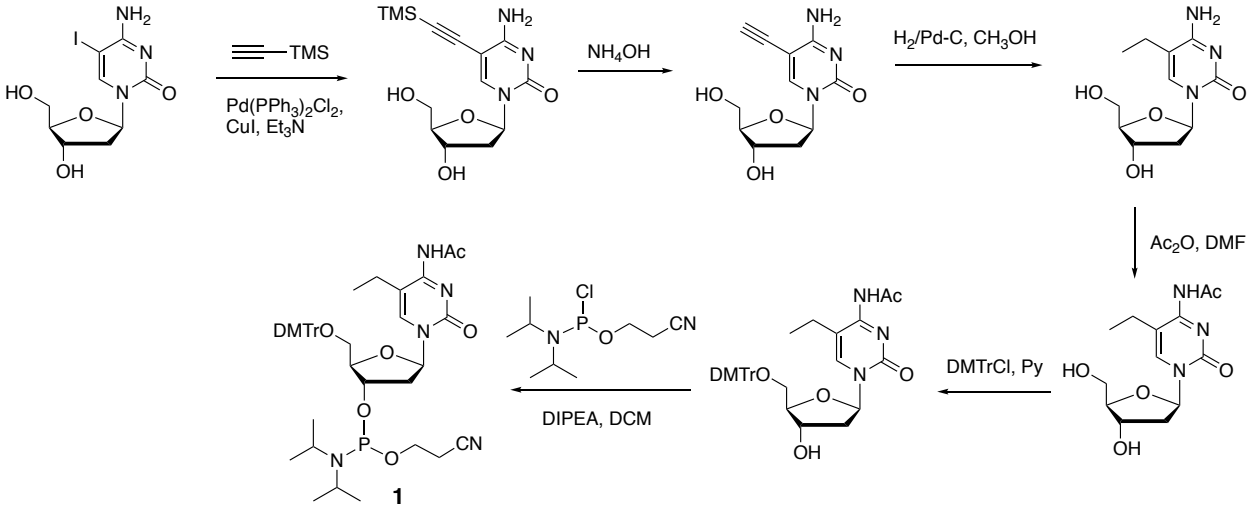
1. General materials, methods and equipment

Chemicals: All chemicals were purchased from established vendors (e.g. Sigma-Aldrich, Acros Organics) and used without purification unless otherwise noted. Optima grade acetonitrile was obtained from Fisher Scientific and degassed under vacuum prior to use during HPLC purification. All reactions to prepare analogues of 5mC and 5hmC were carried out in round bottom flasks and stirred with Teflon®-coated magnetic stir bars under inert atmosphere when needed. Analytical thin layer chromatography (TLC) was performed using EMD 250micron flexible aluminum backed, UV F_{254} pre-coated silica gel plates and visualized under UV light (254 nm) or by staining with phosphomolybdic acid, ninhydrin or anisaldehyde. Reaction solvents were removed by a Büchi rotary evaporator equipped with a dry ice-acetone condenser. Analytical and preparative HPLC was carried out on an Agilent 1220 Infinity HPLC with diode array detector. Concentration and lyophilization of aqueous samples were performed using Savant Sc210A SpeedVac Concentrator (Thermo), followed by Labconco Freeze-Dryer system.

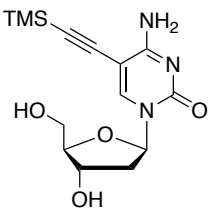
Proton nuclear magnetic resonance spectra (1H NMR) were recorded on Bruker Ultrashield™ Plus 600/500/400/300 MHz instruments at 24°C. Chemical shifts of ^{1}H and ^{13}C NMR spectra are reported as δ in units of parts per million (ppm) relative to tetramethylsilane (δ 0.0) or residual solvent signals: chloroform-d (δ 7.26, singlet), methanol-d_4 (δ 3.30, quintet), and deuterium oxide-d_2 (δ 4.80, singlet). Coupling constants are expressed in Hz. MALDI mass spectra were collected at ultraFlextreme (Bruker) and the data was analyzed using flexAnalysis software. The ESI-MS were recorded on a Q-Exactive™ Thermo Scientific LC-MS with electron spray ionization (ESI) probe.

Plasmids, mutagenic primers and cell lines: All the plasmids are for bacterial expression are obtained as gifts from individual laboratories or purchased from Addgene. Details of these constructs are given in **Table S2**. Mutagenic primers are obtained from Integrated DNA Technologies (**Table S3**). Competent bacterial cells used for protein expression and mutagenesis are given in **Table S4**.

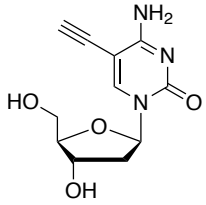
2. Synthesis and characterization of phosphoramidites 1-7



Scheme S1. Synthetic scheme for 5-ethylcytidine (5eC) phosphoramidite **1**.

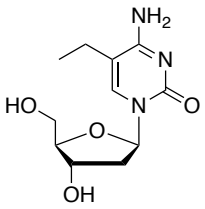


5-Iodo-2'-deoxycytidine (1.54 g, 4.36 mmol) was dissolved in DMF (25 mL) and triethylamine (15.22 mL, 109 mmol), the solution was degassed under vacuum and flushed with N_2.[1,2,3] Then TMS-acetylene (4.00 mL, 28.91 mmol), Bis(triphenylphosphine)palladium(II) dichloride (0.306 g, 0.436 mmol), and copper iodide (83.11 mg, 0.436 mmol) were all added. The reaction was left stirring for 1 hour, after which the black solution was concentrated and chromatographed (0-15% MeOH in DCM) to yield the title TMS compound. ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.38 (s, 1 H), 6.20 (t, *J*=6.26 Hz, 1 H), 4.34-4.39 (m, 1 H), 3.92-3.96 (m, 1 H), 3.83 (dd, *J*=12.04, 2.84 Hz, 1 H), 3.73 (dd, *J*=12.08, 3.48 Hz, 1 H), 2.35-2.42 (m, 1 H), 2.10-2.18 (m, 1 H), 0.238 (s, 9 H), ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 166.13, 156.55, 146.47, 101.61, 96.60, 92.66, 89.02, 87.96, 71.56, 62.33, 42.40, 0.17 ppm; HRMS (ESI) calcd. for $C_{14}H_{22}O_4N_3Si$ $[M+H]^+$: 324.13741; found: 324.13537.

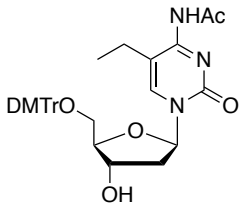


To a solution of starting TMS compound in MeOH (5 mL) was added 28% aqueous ammonia (30 mL) and the reaction was left stirring for 2 hours. The solution was then concentrated and chromatographed (30% MeOH in DCM) to yield the unprotected alkyne (0.819 g, 74.7% over two steps). ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.42 (s, 1 H), 6.20 (t, *J*=6.28 Hz, 1 H), 4.34-4.40 (m, 1 H), 3.95 (q, *J*=3.45 Hz, 1 H), 3.83 (dd, *J*=11.98, 3.14 Hz, 1 H), 3. 82 (s, 1 H), 3.74 (dd, *J*=12.04, 3.56 Hz, 1 H), 2.39 (ddd, *J*=13.59, 6.11, 4.15 Hz, 1 H), 2.10-2.19 (m, 1 H), ppm; ^{13}C-

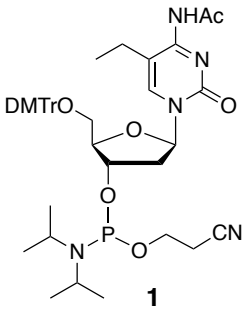
NMR (Methanol-d_4, 100 MHz): δ = 163.43, 156.68, 146.74, 91.80, 89.07, 87.99, 85.22, 75.57, 71.64, 62.37, 42.42 ppm; HRMS (ESI) calcd. for $C_{11}H_{14}O_4N_3$ $[M+H]^+$: 252.09788; found: 252.09809.



To a solution of alkyne compound (0.819 g, 3.26 mmol) in MeOH (15 mL) was added Pd/C (0.694 g of 20% Pd/C 50% wet, 0.652 mmol of Pd, 20 mol%). The reaction mixture was degassed under vacuum and then stirred at rt in a hydrogen atmosphere overnight. The solution was filtered through celite to yield the product (0.654 g, 78.6%). ^{1}H-NMR (Methanol-d_4, 500 MHz): δ = 7.89 (s, 1 H), 6.28 (t, *J*=6.47 Hz, 1 H), 4.37-4.40 (m, 1 H), 3.93 (q, *J*=3.43 Hz, 1 H), 3.82 (dd, *J*=11.98, 3.13 Hz, 1 H), 3.74 (dd, *J*=11.98, 3.58 Hz), 2.31-2.39 (m, 3 H), 2.13-2.19 (m, 1 H), 1.18 (t, *J*=7.43 Hz, 1 H), ppm; ^{13}C-NMR (Methanol-d_4, 125 MHz): δ = 166.65, 158.04, 139.30, 110.05, 88.83, 87.50, 71.94, 62.65, 42.07, 21.29, 12.80 ppm; HRMS (ESI) calcd. for $C_{11}H_{18}O_4N_3$ $[M+H]^+$: 256.12918; found: 256.12922.

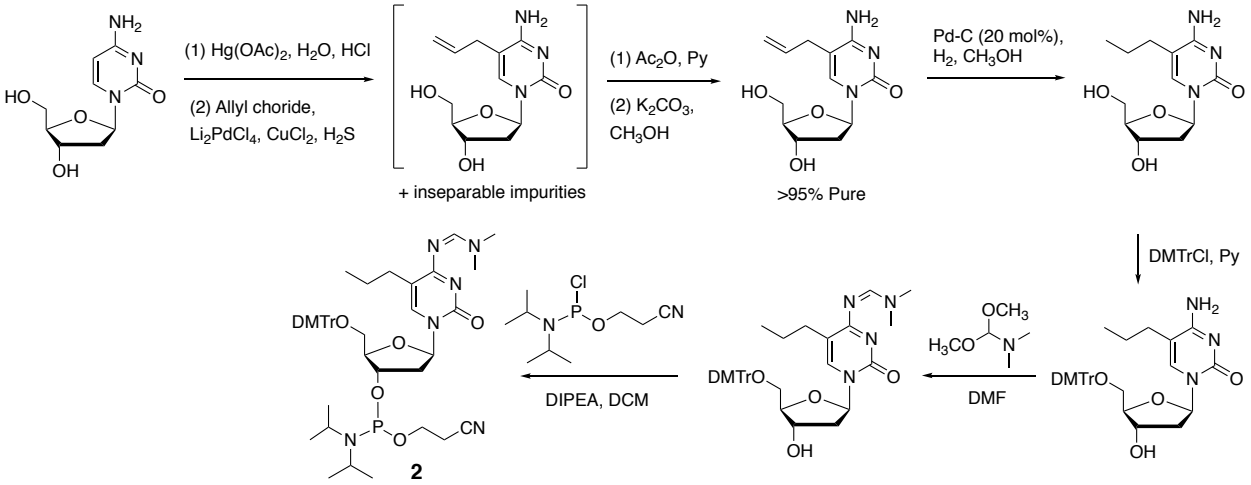


To a solution of starting compound (0.565 g, 2.21 mmol) in DMF (10 mL) was added acetic anhydride (0.199 mL, 2.11 mmol) and was left stirring overnight. The reaction was quenched with MeOH and concentrated. The residue was dissolved in pyridine (10 mL), then DMTrCl (0.825 g, 2.44 mmol) was added and the reaction was left stirring overnight. The reaction was quenched with MeOH, concentrated, and chromatographed (1-5% MeOH in DCM) to yield the required product (0.362 g, 27% over two steps). ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 7.74 (bs, 1 H), 7.35-7.40 (m, 2 H), 7.17-7.33 (m, 7 H), 6.82 (d, *J*=8.84 Hz, 4 H), 6.37 (t, *J*=6.54 Hz, 1 H), 4.53 (quin, *J*=3.13 Hz, 1 H,), 4.08-4.17 (m, 1 H), 3.79 (s, 6 H), 3.48 (dd, *J*=10.42, 3.46 Hz, 1 H), 3.37 (dd, *J*=10.52, 3.28 Hz, 1 H), 2.45 (bs, 3 H), 2.25 (quin, *J*=2.25 Hz, 1 H), 1.99-2.12 (m, 2 H), 1.88-1.98 (m, 1 H), 0.875 (t, *J*=7.46 Hz, 3 H), ppm; ^{13}C-NMR (Chloroform-d, 100 MHz): δ = 158.85, 144.42, 135.53, 130.21, 128.28, 128.10, 127.28, 113.39, 86.97, 86.41, 72.31, 63.52, 55.38, 20.76, 13.45 ppm; HRMS (ESI) calcd. for $C_{34}H_{38}O_7N_3$ $[M+H]^+$: 600.27043; found: 600.27004.



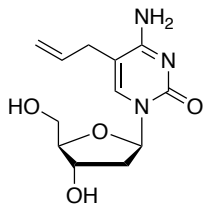
1

A solution of precursor compound (0.22 g, 0.367 mmol, 1equiv.) and DIPEA (0.479 mL, 2.75 mmol, 7.5 equiv.) in freshly distilled DCM (8 mL) was added chlorophosphoramidite (0.118 mL, 0.55 mmol, 1.5 equiv.) in dropwise fashion. The reaction mixture was stirred for 1.5 h and then concentrated *in vacuo* under inert atmosphere. The residue was chromatographed (30-80% EtOAc in Hexanes+0.1 % pyridine) to obtain **1** (0.18 g, 61%) white foam. ^{1}H NMR (Acetonitrile-d_3, 400MHz): δ = 7.71 (br. s., 1 H), 7.45 - 7.50 (m, 2 H), 7.30 - 7.38 (m, 6 H), 7.24 - 7.30 (m, 1 H), 6.87 - 6.93 (m, 4 H), 6.26 (q, *J*=6.8 Hz, 1 H), 4.57 - 4.68 (m, 1 H), 4.17 (dd, *J*=15.1, 3.5 Hz, 1 H), 3.79 (d, *J*=2.0 Hz, 7 H), 3.71 - 3.77 (m, 1 H), 3.54 - 3.71 (m, 3 H), 3.40 - 3.50 (m, 1 H), 3.33 (ddd, *J*=10.7, 8.7, 3.9 Hz, 1 H), 2.67 (t, *J*=5.9 Hz, 1 H), 2.54 (t, *J*=6.0 Hz, 2 H), 2.26 - 2.41 (m, 3 H), 2.06 - 2.20 (m, 2 H), 1.15 - 1.22 (m, 9 H), 1.07 (d, *J*=6.8 Hz, 3 H), 0.91 (t, *J*=7.4 Hz, 3 H) ppm. ^{13}C NMR (Acetonitrile-d_3, 100MHz): δ = 171.7, 159.8, 145.9, 136.7, 136.7, 136.7, 136.7, 131.1, 131.1, 129.1, 129.1, 129.0, 128.9, 128.1, 128.0, 119.6, 119.4, 114.2, 87.5, 87.5, 86.4, 86.4, 86.2, 74.2, 74.1, 73.7, 73.6, 64.1, 63.9, 61.0, 59.6, 59.6, 59.4, 59.4, 58.2, 56.0, 56.0, 44.1, 44.0, 25.0, 24.9, 24.9, 21.3, 21.2, 21.1, 21.0, 21.0, 14.6, 14.0, 13.9 ppm. HRMS (ESI) calcd. for $C_{43}H_{55}O_8N_5P$ [M+H]+: 800.37828; found: 800.37379.



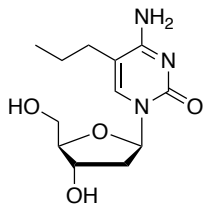
Scheme S2. Synthetic scheme for 5-propylcytidine (5pC) phosphoramidite **2**

A round-bottomed flask was charged with 2-deoxycytidine (1.0g, 4.4 mmol, 1 equiv.), water (5 mL) was added to dissolve followed by slow addition of 1M HCl (4.4 mL).[4,5] After 5 min $Hg(OAc)_2$ (1.54 g, 4.84 mmol, 1.1 equiv.) was added cautiously. The reaction mixture was heated to 75°C for 4.5 h in which time 3/4th volume of solvent would evaporate to leave a white precipitate. To which 15 mL of methanol was added followed by allyl chloride (4.3 mL). $CuCl_2$ (0.651 g, 4.84

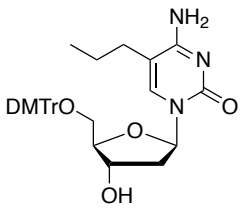


mmol, 1.1 equiv.) was added at once followed by Li_2PdCl_4 (0.231 g, 0.88 mmol, 20 mol% in 15mL MeOH). The reaction mixture was stirred at room temperature for 4.0 h. the solution turned red after 15 min. after 4 h, H_2S gas was passed through the reaction mixture in a fume hood for 1 min. The precipitates were filtered off over Celite pad and the filtrate was concentrated *in vacuo.* The residue was chromatographed (10-35% MeOH, in $CHCl_3$ + 0.1% Et_3N) to get 5-allylcytidine (0.31 g, 23 %) as white solids. The reaction often produced an inseparable mixture of product, starting material (deoxycytidine) and substantial amount of side products. We employed an acetylation and deacetylation maneuver as described below to obtain 5-allylcytidine in greater than 95% purity.

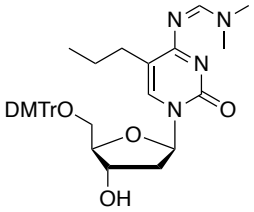
To a stirred solution of mixture 5-allylcytidine with starting compound and other bi-products (0.2g, approx. 0.75 mmol, 1equiv.) in pyridine (5 mL) was added acetic anhydride (0.73 mL, 7.5 mmol, 10 equiv.). The reaction mixture was stirred at room temperature for 2h, after which the reaction mixture was concentrated under reduced pressure. The crude product was purified by flash chromatography using methanol/chloroform (0:20-1:19) to obtain acetylated allyl-dC and acetylated dC. The mixture was then dissolved in methanol (10 mL) and K_2CO_3 (2.0 g) was added at once and stirred at rt for 6h. The reaction mixture was concentrated and chromatographed using methanol/chloroform (1:19-1:5) to obtain 5-allylcytidine with >95% purity. ^{1}H NMR (Deuterium Oxide, 300MHz): δ = 7.66 (s, 1 H), 6.24 (t, *J*=6.5 Hz, 1 H), 5.90 (ddt, *J*=16.9, 10.7, 5.9 Hz, 1 H), 5.03 - 5.22 (m, 2 H), 4.40 (dt, *J*=6.5, 4.4 Hz, 1 H), 4.00 (q, *J*=4.2 Hz, 1 H), 3.81 (dd, *J*=12.5, 3.4 Hz, 1 H), 3.63 - 3.76 (m, 1 H), 3.10 (d, *J*=5.9 Hz, 2 H), 2.39 (ddd, *J*=14.1, 6.5, 4.5 Hz, 1 H), 2.15 - 2.33 (m, 1 H) ppm.



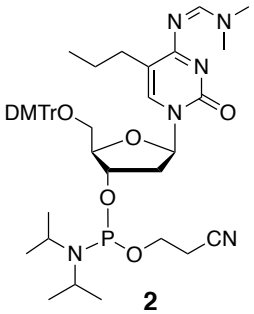
To a solution of 5-allylcytidine (0.1 g, 0.796 mmol, 1equiv.) in MeOH (3 mL) was added Pd in charcoal (0.094 g, 0.160 mmol, 20 mol%). The reaction mixture was degassed under vacuum for 5 min and then stirred 12h at rt in hydrogen atmosphere. The reaction mixture was then filtered with celite and concentrated under reduced pressure to obtain 5-propylcytidine (0.04 g, 63%) white foam. ^{1}H NMR (Methanol-d_4, 400MHz): δ = 7.88 (s, 1 H), 6.26 (t, *J*=6.4 Hz, 1 H), 4.39 (dt, *J*=6.3, 3.8 Hz, 1 H), 3.95 (q, *J*=3.3 Hz, 1 H), 3.82 (dd, *J*=12.2, 3.1 Hz, 1 H), 3.74 (dd, *J*=12.2, 3.6 Hz, 1 H), 2.29 - 2.41 (m, 3 H), 2.16 (dd, *J*=13.4, 6.7 Hz, 1 H), 1.48 - 1.61 (m, 2 H), 0.97 (t, *J*=7.4 Hz, 3 H) ppm. ^{13}C NMR (Methanol-d_4, 100MHz): δ = 166.7, 158.0, 140.2, 108.6, 88.8, 87.5, 72.0, 62.7, 30.2, 22.3, 14.0 ppm. HRMS-ESI m/z $[M+H]^+$calcd. for $C_{12}H_{20}N_3O_4$: 270.1448, found 270.1443(-1.78 ppm)



A round bottomed flask was charged with 5-propylcytidine (0.06g, 0.22 mmol, 1 equiv.) and DMTrCl (0.089 g, 0.264 mmol, 1 equiv.), which was placed in hi-vacuum for 40 min. To this 3 mL of pyridine was added. The reaction mixture was stirred at room temperature for 14h. The reaction was quenched with the addition of 0.5 ml of methanol. The volatiles were evaporated, and the residue was chromatographed (0-5% MeOH, in $CHCl_3$ + 0.1% Et_3N) to get the title compound in (0.1 g, 79%) yield as white solids. ^{1}H NMR (Methanol-d_4, 600 MHz): δ = 7.73 (s, 1 H), 7.45 (d, *J*=7.5 Hz, 2 H), 7.28 - 7.35 (m, 6 H), 7.21 - 7.28 (m, 1 H), 6.88 (d, *J*=8.3 Hz, 4 H), 6.34 (t, *J*=6.4 Hz, 1 H), 4.51 (dt, *J*=6.2, 3.3 Hz, 1 H), 4.05 (q, *J*=3.0 Hz, 1 H), 3.79 (s, 6 H), 3.42 - 3.51 (m, 1 H), 3.36 (dd, *J*=10.5, 3.4 Hz, 1 H), 2.43 (ddd, *J*=13.5, 5.9, 3.6 Hz, 1 H), 2.28 (dt, *J*=13.6, 6.8 Hz, 1 H), 1.97 (td, *J*=10.0, 4.9 Hz, 1 H), 1.80 (td, *J*=10.0, 4.9 Hz, 1 H), 1.18 - 1.31 (m, 3 H), 0.70 (t, *J*=7.3 Hz, 3 H) ppm. ^{13}C NMR (Methanol-d_4, 150 MHz): δ = 164.7, 158.9, 155.8, 144.5, 138.4, 135.5, 130.0, 129.9, 128.1, 127.5, 126.7, 112.8, 107.5, 86.6, 86.4, 85.9, 71.1, 63.2, 54.3, 40.9, 28.6, 21.5, 12.5 ppm.

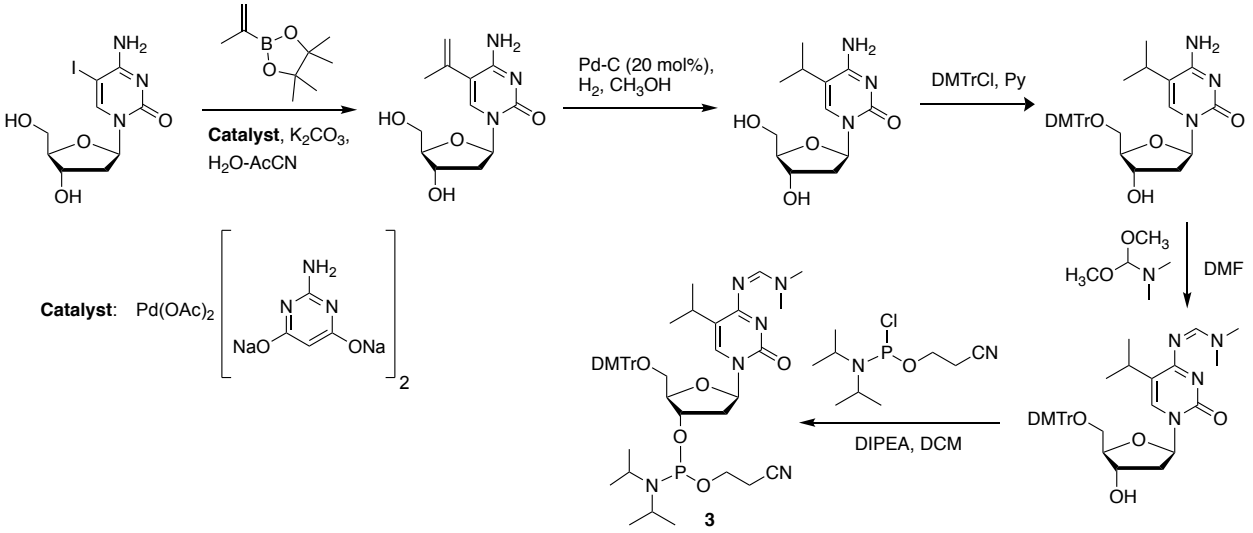


To a stirring solution of starting amine (0.06g, 0.104 mmol, 1equiv.) in anhydrous DMF (2 mL) was added DMF dimethylacetal (0.07 mL, 0.522 mmol, 5 equiv.). The reaction mixture was stirred at rt for 1h. The reaction mixture was concentrated and chromatographed (0-4% MeOH in $CHCl_3$+0.1% Et_3N). The required product was isolated as light-yellow floppy solid (0.065g, 99%). ^{1}H NMR (Methanol-d_4, 500 MHz): δ = 8.60 (s, 1 H), 7.84 (s, 1 H), 7.39 - 7.47 (m, 2 H), 7.26 - 7.36 (m, 6 H), 7.18 - 7.26 (m, 1 H), 6.80 - 6.92 (m, 4 H), 6.36 (t, *J*=6.5 Hz, 1 H), 4.46 - 4.57 (m, 1 H), 4.07 (q, *J*=3.2 Hz, 1 H), 3.78 (s, 6 H), 3.46 (dd, *J*=10.6, 2.8 Hz, 1 H), 3.37 (dd, *J*=10.5, 3.5 Hz, 1 H), 2.48 (ddd, *J*=13.5, 6.1, 3.6 Hz, 1 H), 2.27 (dt, *J*=13.5, 6.7 Hz, 1 H), 2.14 (ddd, *J*=13.7, 9.4, 6.0 Hz, 1 H), 1.92 (ddd, *J*=13.7, 9.3, 6.3 Hz, 1 H), 1.25 - 1.41 (m, 2 H), 0.69 (t, *J*=7.3 Hz, 3 H) ppm. ^{13}C NMR (Methanol-d_4, 125 MHz): δ = 171.0, 158.9, 157.8, 156.9, 144.6, 138.5, 135.6, 135.6, 130.0, 130.0, 128.1, 127.5, 126.7, 115.8, 112.8, 86.6, 86.5, 86.2, 78.1, 71.1, 63.2, 54.3, 41.1, 40.0, 33.9, 29.9, 22.4, 12.9 ppm.

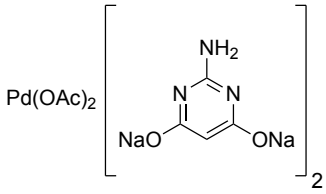


2

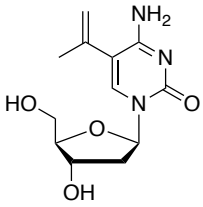
A solution of starting compound (0.137 g, 0.217 mmol, 1equiv.) and DIPEA (0.284 mL, 1.63 mmol, 7.5 equiv.) in freshly distilled DCM (10 mL) was added chlorophosphoramidite (0.073 mL, 0.326 mmol, 1.5 equiv.) in drop-wise fashion. The reaction mixture was stirred for 1.5 h and then concentrated *in vacuo* under inert atmosphere. The residue was chromatographed (50-75% CH_3CN in CH_2Cl_2+0.1 % pyridine) to obtain **2** (0.10 g, 55%) as colorless compound. ^{1}H NMR (Acetonitrile-d_3, 500MHz): δ = 8.66 (s, 1 H), 7.63 (d, *J*=16.5 Hz, 1 H), 7.53 (s, 1 H), 7.45 - 7.49 (m, 2 H), 7.31 - 7.37 (m, 6 H), 7.24 - 7.29 (m, 1 H), 6.85 - 6.92 (m, 4 H), 6.27 - 6.36 (m, 1 H), 4.13 - 4.17 (m, 2 H), 4.06 - 4.09 (m, 1 H), 3.77 - 3.80 (m, 7 H), 3.66 - 3.74 (m, 1 H), 3.58 - 3.66 (m, 2 H), 3.36 - 3.43 (m, 2 H), 3.26 - 3.34 (m, 1 H), 3.17 (s, 3 H), 3.10 (s, 3 H), 2.98 (d, *J*=7.2 Hz, 1 H), 2.66 (t, *J*=6.0 Hz, 1 H), 2.51 - 2.58 (m, 1 H), 2.30 (d, *J*=6.9 Hz, 1 H), 2.14 - 2.23 (m, 1 H), 1.10 - 1.21 (m, 10 H), 1.06 (d, *J*=6.9 Hz, 3 H), 0.69 - 0.77 (m, 3 H) ppm. ^{13}C NMR (Acetonitrile-d_3, 125 MHz): δ = 171.2, 159.3, 158.1, 156.3, 145.5, 139.0, 138.9, 136.4, 136.3, 130.7, 130.6, 128.7, 128.7, 128.6, 128.5, 128.5, 127.5, 127.5, 114.8, 114.8, 113.8, 113.7, 87.0, 86.2, 85.4, 85.4, 74.1, 74.0, 63.8, 63.6, 59.1, 59.1, 59.0, 59.0, 58.8, 58.7, 55.5, 55.5, 47.3, 46.4, 43.7, 43.6, 41.2, 40.7, 40.6, 34.9, 30.7, 24.5, 24.5, 24.4, 23.2, 23.2, 20.6, 20.6, 20.5, 20.2, 20.2, 13.8, 13.8 ppm.



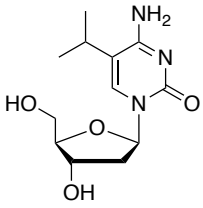
Scheme S3. Synthetic scheme for 5-isopropylcytidine (5iC) phosphoramidite **3**



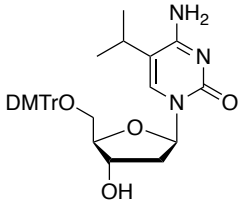
Catalysis prep: 2-Amino-4,6-dihydroxypyrimidine was dissolved in water (3 mL) and NaOH (0.125 mL from 8 M stock) for 5 minutes at room temperature.[6] Then palladium(II) acetate (55 mg, 0.25 mmol) was added and stirred at 65°C for 60 minutes. The solution was diluted with water (5 mL) for a final concentration of 50 mM.



5-iodo-2'-deoxycytidine (1.05 g, 2.97 mmol), isopropenylboronic acid pinacol ester (0.800 mL, 4.25 mmol), and potassium carbonate (0.589 g, 0.426 mmol) were dissolved in water/acetonitrile (10:2.5 mL). Then the catalyst (2.13 mL from 50 mM stock) was added and the reaction was refluxed at 80°C overnight. The reaction was concentrated and chromatographed from (10-20% MeOH in DCM) to yield the required compound. ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.04 (s, 1 H), 6.26 (1H, t, *J*=6.30 Hz), 5.30-5.31 (m, 1 H), 5.10 (bs, 1 H), 4.38 (td, *J*=7.92 , 3.14 Hz, 1 H), 3.93 (q, *J*=3.45 Hz, 1 H), 3.82 (dd, *J*=11.96, 3.04 Hz, 1 H), 3.73 (dd, *J*=11.92 ,3.44 Hz, 1 H), 2.37 (ddd, *J*=13.56, 6.20, 4.20 Hz, 1 H), 2.13-2.20 (m, 1 H), 2.01 (bs, 3 H), ppm; ^{13}C-NMR (Methanol-d_4, 125 MHz): δ = 164.93, 157.29, 140.32, 138.69, 118.56, 111.90, 88.94, 87.60, 71.76, 62.44, 42.17, 23.50 ppm; HRMS (ESI) calcd. for $C_{12}H_{18}O_4N_3$ $[M+H]^+$: 268.12918; found: 268.12838.

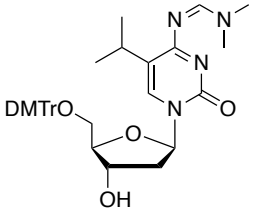


Compound starting alkene was dissolved in methanol (30 mL) then Pd-C (0.071 g of Pd, 0.67 mmol) was added. The reaction vessel was degassed, purged with hydrogen gas and the solution was left for overnight stirring. The reaction was filtered through celite, concentrated and chromatographed from (0-12% MeOH in DCM) to yield 5-isopropylcytidine. ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.07 (s, 1 H), 6.30 (t, *J*=6.28 Hz, 1 H), 4.39-4.42 (m, 1 H), 3.95-3.96 (m, 1 H), 3.83 (dd, *J*=11.86, 2.62 Hz, 1 H), 3.75 (dd, *J*=11.86, 2.90, Hz, 1 H), 2.76 (sep, *J*=6.48 Hz, 1 H), 2.33-2.38 (m, 1 H), 2.16-2.22 (m, 1 H), 1.21 (d, *J*=6.72 Hz, 6 H), ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 165.26, 156.57, 138.87, 114.70, 89.02, 87.67, 71.89, 62.44, 42.23, 26.67, 22.14, 22.08 ppm; HRMS (ESI) calcd. for $C_{12}H_{20}O_4N_3$ $[M+H]^+$: 270.14483 found: 270.14628.

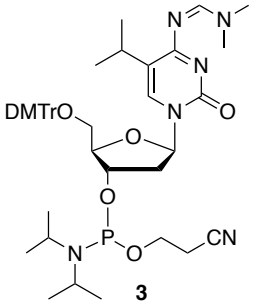


5-isopropylcytidine and DMTrCl (0.533, 1.57 mmol) were dried on vacuum for 1 hour. To the dried material was added pyridine (7 mL) and the reaction was stirred overnight. The reaction was quenched with MeOH (0.5 mL) and concentrated. The residue was chromatographed (1-8% MeOH in DCM w/ 0.1% TEA) to yield the required product. ^{1}H-NMR (Methanol-d_4, 400 MHz): δ

= 7.41-7.47 (m, 3 H), 7.25-7.33 (m, 6 H), 7.19-7.24 (m, 1 H), 6.83-6.88 (m, 4 H), 6.28-6.33 (m, 1 H), 4.37 (quin, *J*=3.11 Hz, 1 H), 4.05 (q, *J*=3.63 Hz, 1 H), 3.77 (s, 6 H), 3.42 (dd, *J*=10.28, 3.64 Hz, 1 H), 3.27-3.32 (m, 1 H), 2.59 (sep, *J*=6.78 Hz, 1 H), 2.38 (ddd, *J*=13.52, 5.80, 3.04 Hz, 1 H), 2.13 (quin, *J*=6.89 Hz, 1 H), 0.948 (d, *J*=6.84 Hz, 3 H), 0.925 (d, *J*=6.84 Hz, 3 H), ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 166.20, 160.24, 157.79, 146.05, 136.99, 136.89, 136.80, 131.40, 131.34, 129.43, 128.86, 128.01, 114.86, 114.16, 87.71, 87.57, 87.25, 72.21, 64.89, 55.72, 41.95, 26.82, 22.05, 21.98 ppm; HRMS (ESI) calcd. for $C_{33}H_{38}O_6N_3$ $[M+H]^+$: 572.27551; found: 572.27458.



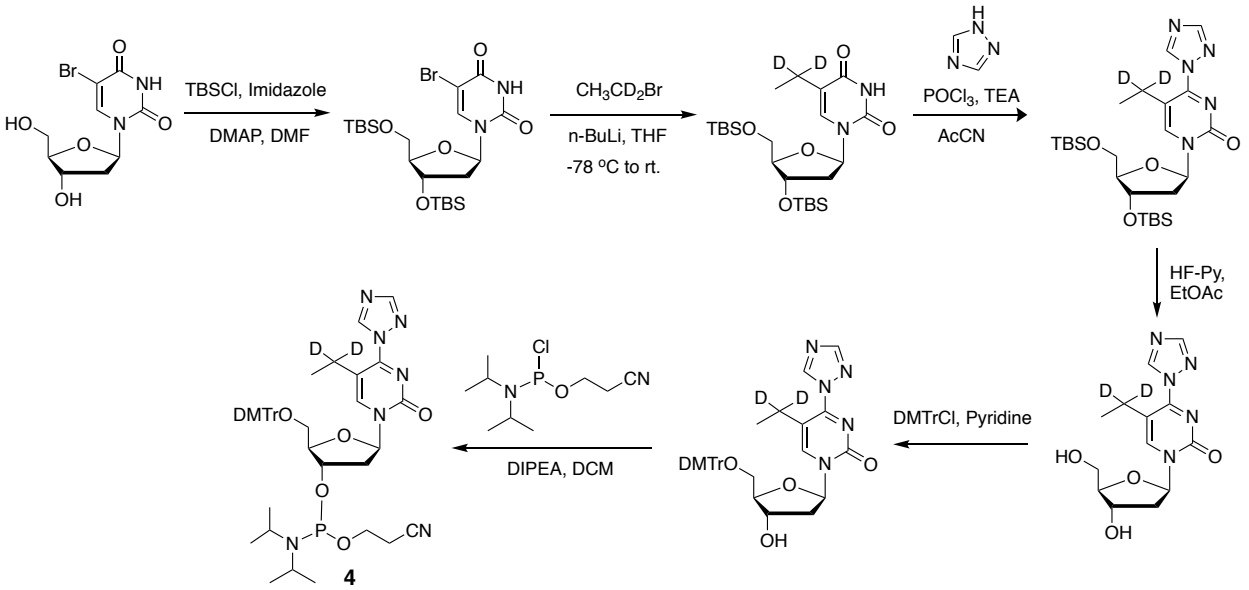
The free amine precursor and N,N-Dimethylformamide dimethyl acetal (1 mL, 7.53 mmol) were dissolved in DMF (10 mL) and stirred at 60°C for 1 hr. Reaction was concentrated and resuspended in DCM, the organic layer was washed with $NaHCO_3$/Water/Brine. The organic extract was concentrated and chromatographed from (1-8% MeOH in DCM w/ 0.1% TEA) to yield the required imine product (0.600 g, 32.3% over 4 steps). ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.58 (s, 1 H), 7.63 (s, 1 H), 7.43 (d, *J*=7.36 Hz, 2 H), 7.31 (dd, *J*=8.80, 1.40 Hz, 4 H), 7.25 (t, *J*=7.44 Hz, 2 H), 7.15-7.21 (m, 1 H), 6.82 (dd, *J*=8.84, 1.12 Hz, 4 H), 6.26-6.39 (m, 1 H), 4.40 (quin, *J*=3.13 Hz, 1 H), 4.07 (q, *J*=3.51 Hz, 1 H), 3.73 (s, 6 H), 3.43 (dd, *J*=10.36, 3.52 Hz, 1 H), 3.29-3.34 (m, 1 H), 3.17 (s, 3 H), 3.12 (s, 3 H), 2.83 (sep, *J*=6.87 Hz, 1 H), 2.44 (ddd, *J*=13.51, 5.89, 3.17 Hz, 1 H), 2.12-2.20 (m, 1 H), 0.965 (t, *J*=6.88 Hz, 6 H), ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 171.89, 160.18, 159.10, 157.98, 146.01, 137.68, 131.37, 131.32, 129.42, 128.86, 128.00, 122.71, 114.16, 87.78, 87.71, 87.59, 72.68, 64.81, 55.69, 42.22, 41.41, 35.38, 28.33, 22.51, 22.40 ppm; HRMS (ESI) calcd. for $C_{36}H_{43}O_6N_4$ $[M+H]^+$: 627.31771; found: 627.31576.



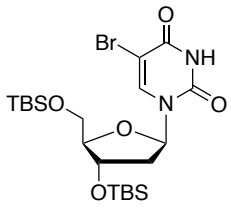
3

The starting alcohol (0.072 g, 0.169 mmol) was dissolved in DCM (7mL) and DIPEA (0.161 mL, 0.933 mmol), and immediately the solution was degassed and flushed with N_2. Then 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.042 mL, 0.187 mmol) was added. After 2 hours the reaction was concentrated and chromatographed (50-70% Acetonitrile in DCM w/ 0.1% TEA) to yield **3** (0.101 g, 56.4%) as a mixture of diastereomers.^{1}H-NMR (Acetonitrile-d_3, 400 MHz): δ = 8.63 (s, 1 H), 7.40-7.52 (m, 3 H), 7.26-7.36 (m, 6 H), 7.19-7.25 (s, 1 H), 6.81-6.90 (m, 4 H), 6.24-6.35 (m, 1 H), 4.48-4.62 (m, 1 H), 4.08-4.17 (m, 1 H), 3.69-3.80 (m, 7 H), 3.49-3.66 (m, 3 H), 3.35-3.45 (m, 1 H), 3.21-3.28 (m, 1 H), 3.14 (s, 3 H), 3.08 (s, 3 H), 2.84-2.98 (m, 1 H), 2.63 (t, *J*=5.94 Hz, 1 H), 2.36-2.54

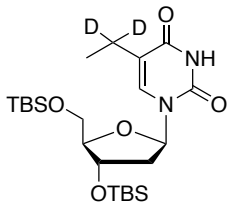
(m, 2 H), 2.18-2.30 (m, 1 H), 1.11-1.18 (m, 9 H), 0.926-1.08 (m, 9 H), ppm; ^{13}C-NMR (Acetonitrile-d_3, 100 MHz): δ = 171.15, 159.66, 158.40, 156.38, 145.87, 137.24, 137.22, 136.77, 136.71, 136.66, 136.60, 131.01, 130.98, 129.03, 128.96, 128.84, 127.85, 120.95 120.93, 119.45, 119.30, 114.06, 87.15, 86.62, 64.25, 64.08, 59.49, 59.30, 55.86, 43.99, 43.87, 41.53, 40.73, 40.68, 35.36, 27.72, 27.69, 24.89, 24.82, 24.75, 22.50, 22.48, 22.41, 21.00, 20.93, 20.85 ppm; ^{31}P-NMR (Acetonitrile-d_3, 162 MHz): δ = 147.80, 147.75, ppm; HRMS (ESI) calcd. for $C_{45}H_{60}O_7N_6P$ $[M+H]^+$: 827.42556; found: 827.42297.



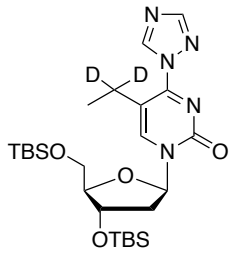
Scheme S4. Synthetic scheme for phosphoramidite **4**



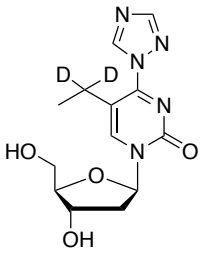
5-Bromouridine (2.00 g 6.50 mmol), TBS-Cl (2.16g, 14.33 mmol), Imidazole (1.99 g, 29.3 mmol), and DMAP (catalytic) were dissolved in DMF (20 mL) at 0°C. Then the reaction was stirred at room temperature overnight, it was quenched with sat. $NaHCO_3$ and extracted x3 with DCM. The organic extract was concentrated and chromatographed (20% EtOAc/Hex) to yield TBS protected product. ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 8.89 (bs, 1 H), 8.07 (s, 1H), 6.29 (dd, *J*=7.70, 5.78 Hz, 1 H), 4.37-4.43 (m, 1 H), 3.98 (q, *J*=2.15 Hz, 1 H), 3.90 (dd, *J*=11.50, 2.26 Hz, 1 H), 3.76 (dd, *J*=11.52, 2.12 Hz, 1 H), 2.31 (ddd, *J*=13.15, 5.77, 2.49 Hz, 1 H), 1.96-2.05 (m, 1 H), 0.934 (s, 9 H), 0.890 (s, 9 H), 0.144 (s, 3 H), 0.134 (s, 3 H), 0.081 (s, 3 H), 0.072 (s, 3H), ppm; ^{13}C-NMR (Chloroform-d, 100 MHz): δ = 159.03, 149.58, 139.59, 96.98, 88.60, 88.02, 72.59, 63.17, 42.17, 26.21, 25.88, 18.62, 18.14, -4.50, -4.70, -5.15, -5.23 ppm; HRMS (ESI) calcd. for $C_{21}H_{40}O_5N_2{}^{79}BrSi_2$ $[M+H]^+$: 535.16536; found: 535.16322.



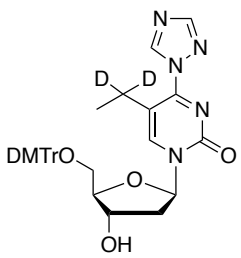
To a solution of TBS-protected 5-bromouridine (0.769 g, 1.43 mmol) in THF (5 mL) at -78°C was added n-BuLi (2.24 mL, 3.58 mmol, 1.6 M) and the reaction was stirred at -78°C for 1 hr. Bromoethane-1,1-d2 (0.320 mL, 4.29 mmol) was then added and the reaction was warmed up to room temperature and stirred overnight. This reaction was also performed on a 1.36 (0.725 g) and 1.53 (0.815 g) mmol scale separately. After the reactions were complete, they were combined and concentrated. The crude product was chromatographed from (0-20% EtOAc in Hex) to yield the deuterated product (0.254 g, 12.1%). ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 8.91 (s, 1 H), 7.39 (s, 1 H), 6.33 (dd, *J*=8.04, 5.76 Hz, 1 H), 4.38-4.40 (m, 1 H), 3.92 (q, *J*=2.59 Hz, 1 H), 3.84 (dd, *J*=11.34, 2.82 Hz, 1 H), 3.75 (dd, *J*=11.34, 2.70 Hz, 1 H), 2.24 (ddd, *J*=13.10, 5.76, 2.44 Hz, 1 H), 1.95-2.02 (m, 1 H), 1.09 (s, 3 H), 0.92 (9H, s), 0.89 (s, 9H), 0.096 (s, 6H), 0.073 (s, 3 H), 0.066 (s, 3 H) ppm; ^{13}C-NMR (Chloroform-d, 100 MHz): δ = 163.51 150.41, 135.07, 116.79, 87.90, 84.97, 72.39, 63.17, 41.35, 26.06, 25.88, 18.54, 18.14, 13.43, -4.52, -4.71, -5.30, -5.31 ppm; HRMS (ESI) calcd. for $C_{23}H_{43}{}^{2}H_{2}N_{2}O_{5}Si_{2}$ $[M+H]^{+}$: 487.29871; found: 487.29877.



To a solution of the deuterated starting material (0.254 g, 0.522 mmol), 1,2,4-triazole (0.724 g, 10.4 mmol), and Et_3N (1.82 mL, 13.0 mmol) in acetonitrile (10 mL) at 0°C was added $POCl_3$ (0.127 mL, 1.36 mmol). After 30 mins the reaction was placed at room temperature and more $POCl_3$ (0.063 mL, 0.68 mmol) was added, the reaction was stirred overnight. TLC had indicated the formation of a blue fluorescent spot and consumption of starting material. The solution was concentrated, and chromatographed (25-60% EtOAc in Hex) to yield the required compound (0.277 g, 99.0%). ^{1}H-NMR (Chloroform-d, 500 MHz): δ = 9.26 (s, 1 H), 8.21 (s, 1 H), 8.11 (s, 1 H), 6.28 (t, *J*=6.38 Hz, 1 H), 4.37-4.40 (m, 1 H), 4.06 (q, *J*=2.75 Hz, 1 H), 3.93 (dd, *J*=11.50, 2.75 Hz, 1H), 3.79 (dd, *J*=11.45, 2.75 Hz, 1 H), 2.65 (ddd, *J*=13.48, 6.05, 3.43 Hz, 1 H), 2.02-2.08 (m, 1 H), 1.12 (s, 3 H), 0.91 (s, 9 H), 0.90 (s, 9 H), 0.11 (s, 3 H), 0.10 (s, 3 H), 0.086 (s, 3 H), 0.073 (s, 3 H) ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 158.03, 153.97, 153.74, 146.27, 145.40, 111.56, 88.92, 88.01, 71.98, 62.91, 42.73, 18.58, 18.16, 14.96, -4.41, -4.73, -5.20, -5.26 ppm; HRMS (ESI) calcd. for $C_{25}H_{44}{}^{2}H_{2}N_{5}O_{4}Si_{2}$ $[M+H]^{+}$: 538.32084; found: 538.32115.

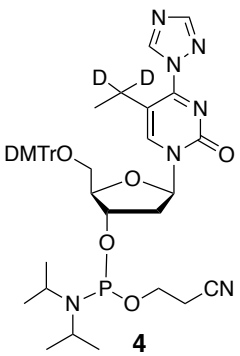


To a solution of starting material (0.277 g, 0.515 mmol) and pyridine (0.202 mL, 2.51 mmol) in EtOAc (15 mL) in a polypropylene tube was added HF-pyridine (0.216 mL, 2.40 mmol). The reaction was stirred overnight, then more HF-pyridine (0.1 mL, 1.11 mmol) was added. After 3 hours the reaction was quenched with 1 mL of methoxytrimethylsilane, and a white precipitate had formed. The reaction was concentrated and chromatographed (5-10% MeOH in DCM) to yield the required TBS-deprotected compound (0.145 g, 91%). ^{1}H-NMR (Methanol-d_4, 500 MHz): δ 9.32 (s, 1 H), 8.82 (s, 1 H), 8.25 (s, 1 H), 6.25 (t, *J*=5.78 Hz, 1 H), 4.42 (q, *J*=5.01 Hz, 1 H), 4.05-4.07 (m, 1 H), 3.92 (dd, *J*=12.10, 2.42 Hz, 1H), 3.79 (dd, *J*=12.10, 3.02 Hz, 1 H), 2.55-2.62 (m, 1 H), 2.26-2.33 (m, 1 H), 1.16 (s, 3 H) ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 159.33, 156.13, 154.22, 149.38, 146.56, 113.63, 89.72, 89.42, 70.95, 61.89, 42.61, 14.66 ppm; HRMS (ESI) calcd. for $C_{13}H_{16}{}^{2}H_2O_4N_5$ $[M+H]^+$: 310.14788; found: 310.14937.



The diol compound (0.145 g, 0.466 mmol) and DMTrCl (217 mg, 0.641 mmol) were dried on vacuum for 1 hour. To the dried material was added pyridine (4 mL) and the reaction was stirred overnight. The reaction was quenched with MeOH (0.5 mL) and concentrated. The residue was chromatographed (1-5% MeOH in DCM + 0.2% Et_3N) to yield the required product (0.194 g, 68%). ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 9.25 (s, 1 H), 8.27 (s, 1 H), 8.09 (s, 1 H), 7.38 (d, *J*=7.24 Hz, 2 H), 7.19-7.29 (m, 7 H), 6.82 (d, *J*=8.76 Hz, 4 H), 6.37 (t, *J*=6.30 Hz, 1 H), 4.58 (bs, 1 H), 4.23-4.26 (m, 1 H), 3.51 (dd, *J*=10.54, 3.10 Hz,1 H), 3.40 (dd, *J*=10.58, 3.22 Hz, 1 H), 3.33 (s, 1 H), 2.83 (ddd, *J*=13.64, 5.70, 3.62, Hz, 1 H), 2.27-2.34 (m, 1 H), 0.86 (s, 3 H) ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 158.83, 158.05, 154.13, 153.72, 146.45, 145.32, 144.37, 135.46, 130.15, 128.19, 128.12, 127.29, 113.40, 112.18, 87.84, 87.02, 86.99, 71.95, 63.34, 55.36, 42.36, 14.83 ppm; HRMS (ESI) calcd. for $C_{34}H_{34}{}^{2}H_2O_6N_5$ $[M+H]^+$: 612.27856; found: 612.27898.

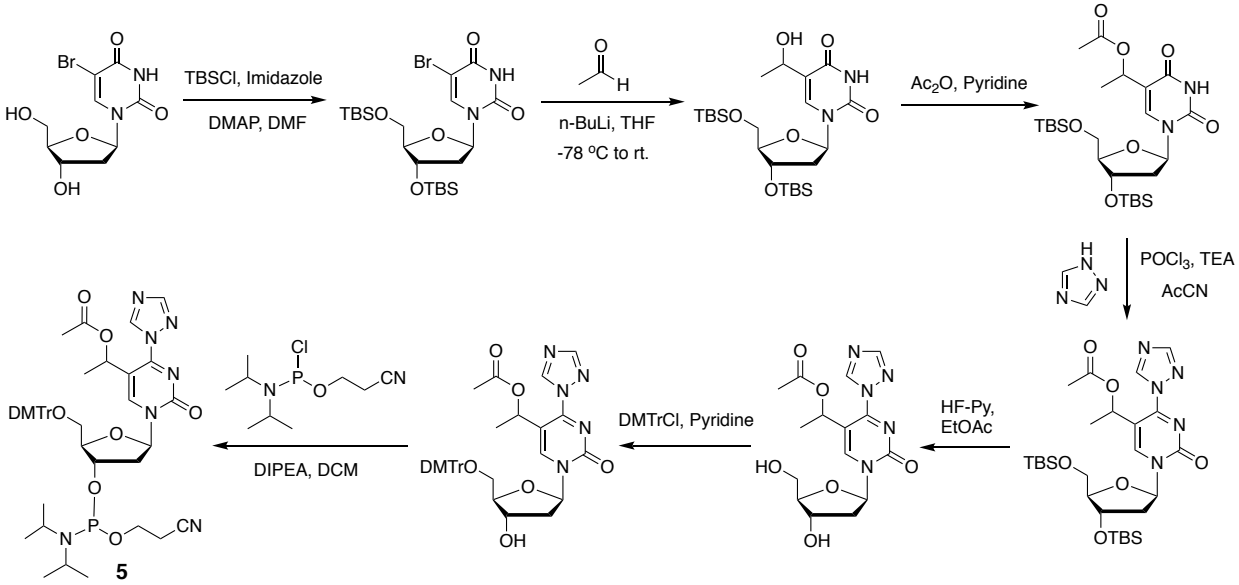
The precursor compound (0.100 g, 0.164 mmol) was dissolved in DCM (4 mL) and DIPEA (0.200 mL, 1.15 mmol), and immediately the solution was degassed and flushed with N_2. Then 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.073 mL, 0.329 mmol) was added, after 1 hour the reaction was diluted with DCM (60 mL) and washed with 5% $NaHCO_3$ and brine. The



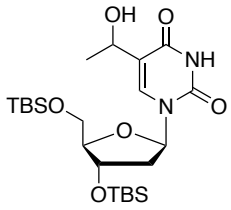
4

organic phase was concentrated and chromatographed (50-80% EtOAc in Hex w/ 0.1% TEA) to yield **4** (0.098 g, 73.8%) as a mixture of diastereomers. ^{1}H-NMR (Acetone-d_6, 400 MHz): δ = 9.15 (s, 1 H), [8.31 (s, 0.5 H), 8.28 (s, 0.5 H)], 8.18 (s, 1 H), 7.44-7.51 (m, 2 H), 7.27-7.37 (m, 6 H), 7.19-7.26 (m, 1 H), 6.84-6.92 (m, 4 H), 6.25 (q, *J*=6.45 Hz, 1 H) 4.67-4.77 (m, 1 H), [4.23-4.28 (m, 0.5 H), 4.28-4.31 (m, 0.5 H)], 3.77-3.91 (m, 1 H), [3.75 (s, 3 H), 3.74 (s, 3 H)], 3.55-3.73 (m, 3 H), 3.52 (dd, *J*=10.94, 3.10 Hz, 1 H), 3.42 (dt, *J*=10.38, 3.87 Hz, 1 H), 2.65-2.76 (m , 2 H), 2.60 (t, *J*=6.00 Hz, 1 H), 2.41-2.52 (m, 1 H), 1.12-1.17 (m, 9 H), 1.06 (d, *J* =6.80, 3 H), [0.83 (s, 1.5 H), 0.82 (s, 1.5 H)] ppm; ^{13}C-NMR (Acetone-d_6, 100 MHz): δ = 159.95, 159.10, 154.62, 154.07, 147.69, 147.65, 146.08, 145.87, 136.63, 136.61, 136.59, 131.21, 129.26, 129.22, 128.92, 128.00, 127.98, 119.13, 118.99, 114.19, 111.91, 111.86, 88.43, 88.34, 87.72, 87.69, 63.87, 63.76, 59.75, 59.56, 55.70, 55.68, 44.17, 44.05, 41.62, 41.59, 41.44, 41.40, 25.08, 25.01, 24.94, 20.95, 20.88, 20.82, 15.29, 15.26 ppm; ^{31}P-NMR (Acetone-d_6, 162 MHz): δ = 148.45, 148.21, ppm; HRMS (ESI) calcd. for $C_{43}H_{51}{}^{2}H_2O_7N_7P$ $[M+H]^+$: 812.38641; found: 812.37804.

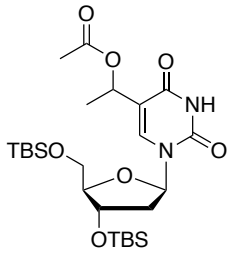
Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has the diastereomers reported together.



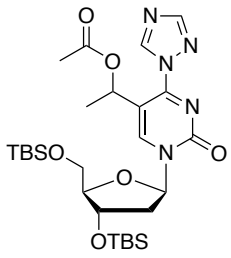
Scheme S5. Synthetic scheme for phosphoramidite **5**



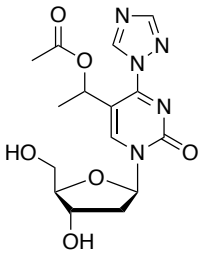
To a solution of TBS protected 5-bromouridine (2.04 g, 3.81 mmol) in THF (15 mL) at -78°C was added n-BuLi (6.19 mL, 9.9 mmol, 1.6 M) and the reaction was stirred at -78°C for 1 hr. Acetaldehyde (0.56 mL, 9.9 mmol) in THF (1 mL) was then added and the reaction was stirred at -78°C. After 1 hour the reaction was placed at room temperature and quenched with water and extracted x3 EtOAc. It was then chromatographed (18-50% EtOAc in Hexane) to yield the required product as a mixture of diastereomers (0.797 g, 41.8%). ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 8.71 (bs, 1 H), [7.55 (s, 1 H) minor 38%, 7.54 (s, 1 H), major 62%], 6.26-6.30 (m, 1 H), [4.69 (q, *J*=6.44 Hz, 1H) major 62%, 4.64 (q, *J*=6.47 Hz, 1H)], 4.37-4.40 (m, 1 H), 3.93-3.96 (m, 1 H), 3.74-3.82 (m, 2 H), 3.11 (bs, 1 H), 2.25-2.30 (m, 1 H), 1.95-2.03 (m, 1 H), 1.45-1.48 (m, 3 H), 0.91 (s, 9 H), 0.89 (s, 9 H), 0.10 (s, 6 H), 0.08 (s, 3 H), 0.07 (s, 3 H), ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 163.29, 163.22, 149.83, 149.81, 135.45, 135.40, 117.67, 117.62, 88.14, 85.51, 85.47, 72.50, 72.39, 65.74, 65.09, 63.29, 63.19, 41.42, 41.35, 26.13, 25.90, 22.51, 22.07, 18.59, 18.15, -4.51, -4.67, -5.20, -5.23, -5.25 ppm; HRMS (ESI) calcd. for $C_{23}H_{45}O_6N_2Si_2$ $[M+H]^+$: 501.28107; found: 501.28034. **Note**: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals and relative abundance for each isomer. Peaks at 8.71 and 3.11 denote exchangeable protons N-H, and O-H respectively. ^{13}C has both diastereomers reported together.



To a solution of precursor alcohol (0.689 g, 1.38 mmol) in pyridine (8 mL) was added acetic anhydride (0.389 mL, 4.13 mmol). The reaction was stirred at r.t. for 16 hours after which it was concentrated and chromatographed (15-20% EtOAc in Hex) to yield the acetyl protected product as a mixture of diastereomers (0.625 g, 83.7%). ^{1}H-NMR (Chloroform-d, 400 MHz): δ = 8.36 (bs, 1 H), [7.59 (s, 0.5 H), 7.58 (s, 0.5 H)], 6.23-6.26 (m, 1 H), [5.77 (q, *J*=6.62 Hz, 0.5 H), 5.71 (q, *J*=6.60 Hz, 0.5 H)], 4.38-4.41 (m, 1 H), 3.95-3.96 (m, 1 H), 3.78 (m, 2 H), 2.27-2.33 (m, 1 H), 2.05 (s, 3 H), 1.96-2.02 (m, 1 H), [1.51 (d, *J*=6.60 Hz, 1.5 H), 1.50 (d, *J*=6.55 Hz, 1.5 H)], 0.92 (s, 9 H), 0.89 (s, 9 H), 0.10 (s, 6 H), 0.09 (s, 3 H), 0.08 (s, 3 H), ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 170.01, 169.94, 161.46, 161.41, 149.74, 137.51, 137.46, 114.49, 114.43, 88.20, 88.13, 85.73, 72.57, 72.50, 67.14, 66.71, 63.35, 63.25, 41.35, 41.32, 26.12, 25.89, 21.34, 19.58, 19.47, 18.57, 18.15, -4.52, -4.68, -5.17, -5.19, -5.26 ppm; HRMS (ESI) calcd. for $C_{25}H_{47}O_7N_2Si_2$ $[M+H]^+$: 543.29163; found: 543.29053. Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. The peak at 8.36 denote the exchangeable proton N-H. ^{13}C has both diastereomers reported together.

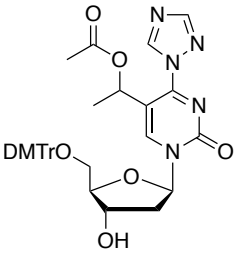


To a solution of precursor compound (0.535 g, 0.901 mmol), 1,2,4-triazole (1.25 g, 18.0 mmol), and Et_3N (3.14 mL, 22.5 mmol) in acetonitrile (10 mL) at 0°C was added $POCl_3$ (0.218 mL, 2.34 mmol). TLC had indicated the formation of a blue fluorescent spot. After 1 hour at 0°C the reaction was quenched with methanol, concentrated, and chromatographed (20-50% EtOAc in Hex) to yield the required product as a mixture of diastereomers (0.253 g, 47.2%). ^{1}H-NMR (Dichloromethane-d_2, 500 MHz): δ = [9.19 (s, 0.5 H), 9.18 (s, 0.5 H)], [8.45 (s, 0.5 H), s (8.41 (s, 0.5 H)], [8.12 (s, 0.5 H), 8.11 (s, 0.5 H)], 6.63-6.74 (m, 1 H), 6.18 (q, *J*=6.89 Hz, 1 H), 4.38-4.43 (m, 1 H), 4.11-4.17 (m, 1 H), 3.76-3.86 (m, 2 H), 2.65-2.72 (m, 1 H), 1.97-2.06 (m, 4 H), [1.56 (d, *J*=6.35 Hz, 1.5 H), 1.55 (d, *J*=6.35 Hz, 1.5 H)], 0.902-0.943 (m, 18 H), 0.091-0.144 (m, 12 H)), ppm; ^{13}C-NMR (Dichloromethane-d_2, 125 MHz): δ = 169.84, 169.76, 156.87, 156.74, 154.23, 154.20, 153.57, 153.56, 146.03, 145.86, 145.53, 145.49, 111.27, 110.97, 89.83, 89.53, 89.14, 89.02, 73.38, 73.14, 66.73, 66.66, 63.87, 63.83, 42.85, 42.50, 26.23, 26.21, 26.00, 21.85, 21.41, 21.33, 21.16, 18.73, 18.35, 18.35, -4.45, -4.65, -4.95, -5.10, -5.11, -5.13, -5.17, ppm; HRMS (ESI) calcd. for $C_{27}H_{48}O_6N_5Si_2$ $[M+H]^+$: 594.31376; found: 594.31306. It should be noted that this reaction yields some undesired elimination product. Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has diastereomers reported together.

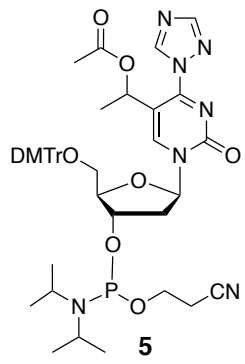


To a solution of TBS-protected compound (0.253 g, 0.426 mmol) and pyridine (0.158 mL, 1.96 mmol) in EtOAc (10 mL) in a polypropylene tube was added HF-pyridine (0.169 mL, 1.87 mmol). The reaction was stirred for six hours, then more HF-pyridine (0.169 mL, 1.87 mmol) was added. The solution was stirred overnight and then quenched with 1 mL of methoxytrimethylsilane, and a white precipitate had formed. The reaction was concentrated and chromatographed (1-10% MeOH in DCM) to yield as a diastereomeric mixture of required products (0.140 g, 89.9%). ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = [9.34 (s, 0.5 H), 9.33 (s, 0.5 H)], [9.16 (s, 0.5 H), 9.14 (s, 0.5 H)], 8.23-9.29 (m, 1 H), 6.58-6.67 (m, 1 H), 6.25 (q, *J*=5.03 Hz, 1 H), 4.38-4.47 (m, 1 H), 4.06-4.13 (m, 1 H), 3.89-3.99 (m, 1 H), 3.76-3.83 (m, 1 H), 2.56-2.65 (m, 1 H), 2.24-2.40 (m, 1 H), 1.96-2.07 (s, 3 H), [1.59 (d, *J*=6.00 Hz, 1.5 H), 1.58 (d, *J*=6.16 Hz, 1.5 H)], ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 171.88, 171.81, 157.74, 157.56, 155.75, 155.73, 154.46, 154.41, 149.15, 148.85, 146.61, 146.58, 113.02, 112.74, 89.98, 89.91, 89.88, 89.84, 71.02, 67.76, 67.73, 61.81, 61.74, 42.81, 42.66, 21.30, 21.03, 21.01, ppm; HRMS (ESI) calcd. for $C_{15}H_{20}O_6N_5$ $[M+H]^+$: 366.14081; found: 366.14271. It should be noted that this contains some undesired elimination

product. Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has diastereomers reported together.



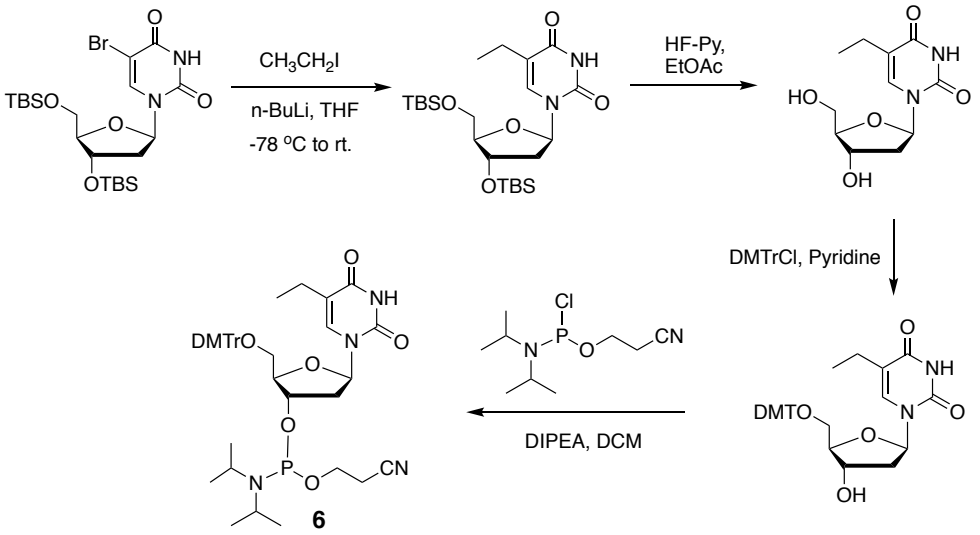
Precursor compound (0.139 g, 0.383 mmol) and DMTrCl (0.143 g, 0.421 mmol) were dried on vacuum for 1 hour. To the dried material was added pyridine (2 mL) and the reaction was stirred overnight. The reaction was quenched with MeOH (0.5 mL) and concentrated. The residue was chromatographed (1-5% MeOH in DCM) to yield DMTr-protected product as a mixture of diastereomers (0.115 g, 45.2%). ^{1}H-NMR (Dichloromethane-d_2, 500 MHz): δ = 9.17-9.22 (m, 1 H), [8.34 (s, 0.55 H), 8.32 (s, 0.45 H)], 8.09-8.15 (m, 1 H), 7.39-7.45 (m, 2 H), 7.27-7.33 (m, 6 H), 7.20-7.25 (m, 1 H), 6.82-6.86 (m, 4 H), [6.65 (q, *J*=6.36 Hz, 0.45 H), 6.59 (q, *J*=6.36 Hz, 0.55 H)], 6.14-6.20 (m, 1 H), 4.37 (bs, 1 H), [4.24 (q, *J*=4.12 Hz, 0.55 H), 4.20 (q, *J*=4.37 Hz, 0.45 H), 3.76 (s, 6 H), 3.42-3.46 (m, 1 H), 3.36-3.41 (m, 1 H), 2.71-2.81 (m, 1 H), 2.10-2.21 (m, 1 H), [1.84 (s, 1.6 H), 1.57 (s, 1.4 H)], [1.45 (d, *J*=6.35 Hz, 1.3 H), 1.34, (d, *J*=6.40 Hz, 1.7 H)], ppm; ^{13}C-NMR (Dichloromethane-d_2, 125 MHz): δ = 169.90. 169.88, 159.28, 156.85, 156.82, 154.26, 154.22, 153.56, 153.52, 150.18, 145.65, 145.59, 145.55, 145.59, 145.07, 145.03, 136.00, 135.95, 135.90, 130.45, 130.42, 128.43, 127.46, 127.43, 113.70, 111.40, 111.16, 88.83, 88.56, 87.26, 87.25, 87.14, 86.99, 72.90, 72.63, 66.62, 66.49, 64.12, 64.05, 55.65, 41.93, 41.90, 21.49, 21.28, 21.21, 20.95, ppm; HRMS (ESI) calcd. for $C_{36}H_{38}O_8N_5$ $[M+H]^+$: 668.27149; found: 668.27392. It should be noted that this contains some undesired elimination product. Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has diastereomers reported together.



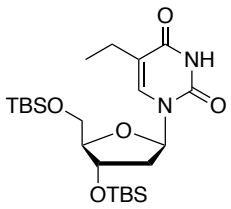
5

The precursor compound (0.112 g, 0.169 mmol) was dissolved in DCM (5mL) and DIPEA (0.206 mL, 1.18 mmol), and immediately the solution was degassed and flushed with N_2. Then 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.075 mL, 0.337 mmol) was added. After 1 hour the reaction was concentrated and chromatographed (30-80% EtOAc in Hex w/ 0.1% TEA) to yield **5** (0.101 g, 56.4%) as a mixture of diastereomers.^{1}H-NMR (Dichloromethane-d_2, 500 MHz): δ = 9.20 (s, 1 H), 8.31-8.41 (m, 1 H), 8.12 (s, 1 H), 7.21-7.47 (m, 9 H), 6.81-6.86 (m, 4 H), 6.56-6.68 (m, 1 H), 6.15-6.21 (m, 1 H), 4.43-4.52 (m, 1 H), 4.29-4.40 (m, 1 H), 3.74-3.88 (m, 7 H), 3.53-3.72 (m, 3 H), 3.35-3.50 (m, 2 H), 2.78-2.95 (m, 1 H), 2.59-2.66 (m, 1 H), 2.42-2.50 (m, 1 H), 2.10-2.26 (m, 1 H), [1.79-1.83, 1.53-1.57 (m, 3 H)], [1.42-1.47, 1.32-1.37 (m, 3 H)], 1.15-1.19 (m, 9 H),

1.05-1.12 (m, 3 H), ppm; ^{13}C-NMR (Dichloromethane-d_2, 125 MHz): δ = 169.92, 159.31, 154.26, 154.23, 153.51, 153.45, 145.74, 145.69, 145.57, 145.51, 145.10, 145.08, 136.07, 136.02, 135.96, 135.93, 130.57, 130.53, 130.49, 128.58, 128.52, 128.40, 127.48, 127.45, 127.42, 118.15, 118.00, 113.72, 111.40, 111.38, 111.24, 111.17, 88.96, 88.82, 88.64, 88.59, 66.66, 66.60, 66.50, 63.78, 63.71, 63.61, 58.93, 58.90, 58.87, 58.78, 58.75, 58.72, 55.68, 55.65, 43.91, 43.87, 43.81, 43.77, 43.73, 41.14, 41.07, 24.83, 24.78, 21.44, 21.37, 21.32, 21.19, 21.07, 20.94, 20.86, 20.81, 20.71, 20.69, 20.63, ppm; ^{31}P-NMR (Dichloromethane-d_2, 202 MHz): δ = 149.18, 149.07, 148.87, 148.73 ppm; HRMS (ESI) calcd. for $C_{45}H_{55}O_9N_7P$ $[M+H]^+$: 868.37934; found: 868.37496. It should be noted that this contains some undesired elimination product. Note: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has diastereomers reported together.

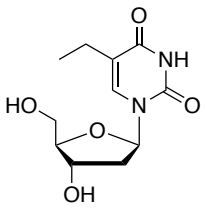


Scheme S6. Synthetic scheme for phosphoramidite **6**

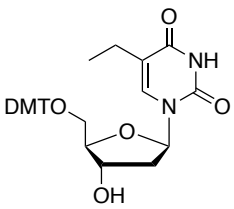


To a solution of TBS-protected 5-bromouridine (0.700 g, 1.31 mmol) in THF (15 mL) at -78°C was added n-BuLi (2.04 mL, 3.27 mmol, 1.6 M) and the reaction was stirred at -78°C for 1 hr. Iodoethane (0.316 mL, 3.93 mmol) was then added and the reaction was warmed up to room temperature and stirred overnight, and then concentrated. The crude product was chromatographed from (10-15% EtOAc in Hex) to yield the alkylated product (71.8 mg, 11.3%). ^{1}H-NMR (Chloroform-d, 500 MHz): δ = 8.56 (s, 1 H), 7.40 (s, 1 H), 6.32 (dd, *J*=8.10, 5.75 Hz, 1 H), 4.38-4.41 (m, 1 H), 3.93 (q, *J*=2.67 Hz, 1 H), 3.84 (dd, *J*=11.35, 2.85 Hz, 1 H), 3.75 (dd, *J*=11.35, 2.75 Hz, 1 H), 2.29-2.38 (m, 2 H), 2.24 (ddd, *J*=13.10, 5.75, 2.45 Hz, 1 H), 1.96-2.03 (m, 1 H), 1.11 (t, *J*=7.45 Hz, 3 H), 0.92 (s, 9 H), 0.89 (s, 9 H), 0.10 (s, 6 H), 0.08 (s, 3 H), 0.07 (s, 3 H) ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 163.32, 150.29, 135.08, 116.86, 87.94, 85.01, 72.43,

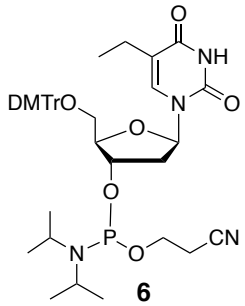
63.20, 41.36, 20.77, 18.56, 18.15, 13.61, -4.51, -4.69, -5.28, -5.29, ppm; HRMS (ESI) calcd. for $C_{23}H_{45}N_2O_5Si2$ $[M+H]^+$: 485.2867; found: 485.2836.



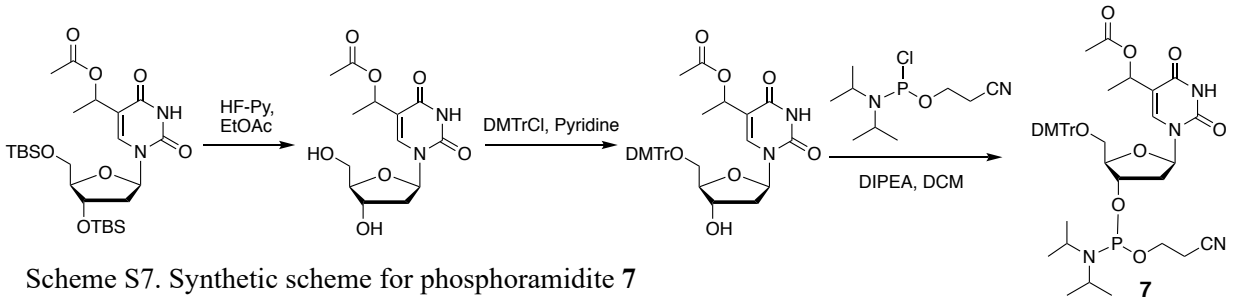
To a solution of the TBS-protected precursor compound (68.4 mg, 0.141 mmol) and pyridine (0.053 mL, 0.64 mmol) in EtOAc (10 mL) in a polypropylene tube was added HF-pyridine (0.056 mL, 0.62 mmol). The reaction was stirred for six hours, then more HF-pyridine (0.056 mL, 0.62 mmol) was added. The solution was stirred overnight and then quenched with 0.5 mL of methoxytrimethylsilane. The reaction was concentrated and chromatographed (1-10% MeOH in DCM) the required diol (33.0 mg, 91.3%). ^{1}H-NMR (Methanol-d_4, 500 MHz): δ = 7.82 (s, 1 H), 6.29 (t, *J*=6.75 Hz, 1 H), 4.39-4.43 (m ,1 H), 3.92 (q, *J*=3.23 Hz, 1 H), 3.80 (dd, *J*=11.98, 3.08 Hz, 1 H), 3.73 (dd, *J*=11.95, 3.45 Hz, 1 H), 2.297-2.36 (m, 2 H), 2.19-2.291 (m, 2 H), 1.12 (t, *J*=7.47 Hz, 3 H), ppm; ^{13}C-NMR (Methanol-d_4, 125 MHz): δ = 165.94, 152.25, 137.58, 117.38, 88.83, 86.36, 72.22, 62.76, 41.25, 21.05, 13.25, ppm; (ESI) calcd. for $C_{11}H_{17}N_2O_5$ $[M+H]^+$: 257.11320; found: 257.11399.



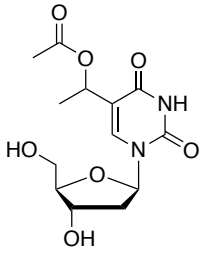
5-ethyl-2'-deoxy-uridine (32 mg, 0.125 mmol) and DMTrCl (45.6, 0.137 mmol) were dried on vacuum for 1 hour. To the dried material was added pyridine (1 mL) and the reaction was stirred overnight. The reaction was quenched with MeOH (0.5 mL) and concentrated. The residue was chromatographed (1-5% MeOH in DCM w/ 0.1% pyridine) to yield the required product (37 mg, 53%). ^{1}H-NMR (Chloroform-d, 500 MHz): δ = 7.43 (s, 1 H), 7.34-7.39 (m, 2 H), 7.23-7.30 (m, 6 H), 7.18-7.23 (m, 1 H), 6.80 (d, *J*=8.41 Hz, 4 H), 6.40 (dd, *J*=7.92, 5.89 Hz, 1 H), 4.53 (quin, *J*=2.83 Hz, 1 H), 4.05 (q, *J*=3.00 Hz, 1 H), 3.76 (s, 6 H), 3.43 (dd, *J*=10.41, 3.41 Hz, 1 H), 3.33 (dd, *J*=10.37, 3.10 Hz, 1 H), 2.39 (ddd, *J*=13.53, 5.79, 2.67 Hz, 1 H), 2.23-2.29 (m, 1 H), 1.81-1.97 (m, 2 H), 0.816 (t, *J*=7.44 Hz, 1 H), ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 163.47, 158.82, 150.59, 149.64, 144.40, 136.39, 135.53, 135.52, 135.10, 130.22, 129.27, 128.29, 128.08, 127.26, 117.30, 113.37, 86.96, 86.22, 84.78, 72.59, 63.69, 55.37, 40.90, 20.39 13.38 , ppm; HRMS (ESI) calcd. for $C_{32}H_{34}N_2O_7Na$ $[M+Na]^+$: 581.2264; found: 581.2267.



The precursor compound (36.0 mg, 0.0644 mmol) was dissolved in DCM (2mL) and DIPEA (0.112 mL, 0.644 mmol), and immediately the solution was degassed and flushed with N_2. Then 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.036 mL, 0.161 mmol) was added. After 1 hour the reaction was concentrated and chromatographed (30-50% EtOAc in Hex w/ 0.1% TEA) to yield **6** (29.1 mg, 59.4%) as a mixture of diastereomers.^{1}H-NMR (Acetonitrile-d_3, 300 MHz): δ = 7.39-7.52 (m, 2 H), 7.19-7.38 (m, 8 H), 6.79-6.92 (m, 4 H), 6.22-6.31 (m, 1 H), 4.53-4.70 (m, 1 H), 4.03-4.12 (m, 1 H), 3.72-3.86 (m, 7 H), 3.53-3.67 (m, 3 H), 3.32-3.41 (m, 1 H), 3.22-3.31 (m, 1 H), 2.63 (t, *J*=5.99 Hz, 1 H), 2.51 (t, *J*=6.01 Hz, 1 H), 2.07-2.44 (m, 4 H), 1.11-1.17 (m, 9 H), 1.04 (d, *J*=6.78 Hz, 3 H), 0.839 (t, *J*=7.43 Hz, 3 H), ppm; ^{13}C-NMR (Acetonitrile-d_3, 75 MHz): δ = 164.18, 159.77, 151.33, 145.82, 136.69, 136.65, 136.60, 136.16, 136.10, 131.08, 129.09, 129.03, 128.89, 127.99, 119.53, 119.37, 117.30, 117.25, 114.10, 87.39, 87.36, 85.97, 85.91, 85.72, 85.64, 85.24, 85.20, 74.40, 74.17, 74.01, 73.79, 64.20, 64.08, 59.57, 59.31, 55.92, 44.08, 43.92, 39.96, 39.92, 39.84, 39.76, 24.93, 24.87, 24.83, 24.77, 21.05, 20.99, 20.90, 13.80, 13.77, ppm; ^{31}P-NMR (Acetonitrile-d_3, 121 MHz): δ = 147.97, 147.93, ppm; HRMS (ESI) calcd. for $C_{41}H_{54}N_4O_9P$ $[M+H_3O]^+$: 777.36229; found: 777.36207.

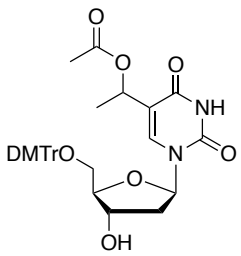


Scheme S7. Synthetic scheme for phosphoramidite **7**

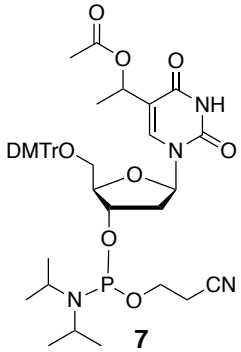


To a solution of the TBS-protected precursor compound (0.316 g, 0.426 mmol) and pyridine (0.217 mL, 2.69 mmol) in EtOAc (10 mL) in a polypropylene tube was added HF-pyridine (0.232 mL, 2.57 mmol). The reaction was stirred for six hours, then more HF-pyridine (0.232 mL, 2.57 mmol) was added. The solution was stirred overnight and then quenched with 1 mL of methoxytrimethylsilane, and a white precipitate had formed. The reaction was concentrated and chromatographed (1-10% MeOH in DCM) to yield as a mixture of diastereomers the required diol (0.137 g, 75.0%). ^{1}H-NMR (Methanol-d_4, 400 MHz): δ = 8.15-8.20 (m, 1 H), 6.27-6.33 (m, 1 H), 5.75 (q, *J*=6.55 Hz, 1 H), 4.38-4.45 (m, 1 H), 3.94 (q, *J*=3.00 Hz, 1 H), 3.78-3.85 (m, 1 H), 3.75

(dd, *J*=11.90, 2.90, 1 H), 2.26-2.35 (m, 1 H), 2.16-2.26 (m, 1 H), 2.05 (s, 3 H), [1.46 (d, *J*=6.56 Hz, 1.5 H), 1.45 (d, *J*=6.56 Hz, 1.5 H)], ppm; ^{13}C-NMR (Methanol-d_4, 100 MHz): δ = 171.98, 171.94, 164.31, 151.93, 151.90, 139.50, 139.08, 115.67, 115.48, 89.13, 86.82, 86.75, 72.25, 72.20, 67.41, 67.35, 62.67, 62.63, 41.77, 41.64, 21.04, 19.64, 19.50 ppm; HRMS (ESI) calcd. for $C_{13}H_{19}O_7N_2$ $[M+H]^+$: 315.11868; found: 315.12002. **Note**: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. ^{13}C has both diastereomers reported together.



The precursor diol (0.137 g, 0.438 mmol) and DMTrCl (0.163 g, 0.482 mmol) were dried on vacuum for 1 hour. To the dried material was added pyridine (2 mL) and the reaction was stirred overnight. The reaction was quenched with MeOH (0.5 mL) and concentrated. The residue was chromatographed (1-5% MeOH in DCM) to yield the DMTr-protected product as a mixture of diastereomers (0.146 g, 54.1%). ^{1}H-NMR (Chloroform-d, 500 MHz): δ = 8.79 (bs, 1 H), [7.60 (s, 0.5 H), 7.53 (s, 0.5 H)], 7.40-7.44 (m, 2 H), 7.20-7.33 (m, 7 H), 6.82-6.86 (m, 4 H), 6.26-6.34 (m, 1 H), 5.50-5.85 (sextet, *J*=6.16 Hz, 1 H), 4.43-4.50 (m, 1 H), 4.02-4.07 (m, 1 H), 3.79 (s, 1 H), 3.44 (dd, *J*=10.28, 4.33, 1 H), 3.35-3.40 (m, 1 H), 2.40-2.49 (m, 1 H), 2.18-2.25 (m, 1 H), [(s, 1.5 H), (s, 1.5 H)], [1.33 (d, *J*=6.55 Hz, 1.5 H), 1.32 (d, *J*=6.55 Hz, 1.5 H)], ppm; ^{13}C-NMR (Chloroform-d, 125 MHz): δ = 170.07, 169.92, 161.49, 161.42, 158.88, 149.89, 149.86, 144.57, 144.53, 137.16, 136.91, 135.76, 135.67, 135.63, 130.20, 128.28, 128.26, 128.12, 128.11, 127.20, 114.84, 113.46, 86.92, 86.85, 86.04, 85.78, 85.52, 85.15, 72.63, 72.35, 66.78, 66.57, 63.67, 55.36, 41.00, 40.71, 21.00, 20.97, 19.45, 19.39 ppm; HRMS (ESI) calcd. for $C_{34}H_{35}O_9N_2$ $[M-H]^-$: 615.23371; found: 615.23529. **Note**: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. The peak at 8.79 denotes the exchangeable proton N-H. ^{13}C has both diastereomers reported together.



7

The precursor compound (0.100 g, 0.162 mmol) was dissolved in DCM (5mL) and DIPEA (0.198 mL, 1.14 mmol), and immediately the solution was degassed and flushed with N_2. Then 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.072 mL, 0.325 mmol) was added. After 1 hour the reaction was concentrated and chromatographed (30-80% EtOAc in Hex w/ 0.1% TEA) to yield **7** (0.101 g, 76.3%) as a mixture of diastereomers. ^{1}H-NMR (Acetonitrile-d_3, 400 MHz): δ = 9.10 (bs, 1 H), 7.42-7.52 (m, 3 H), 7.26-7.36 (m, 6 H), 7.19-7.26 (m, 1 H), 6.83-6.90 (m, 4 H), 6.11-6.24 (m, 1 H), 5.46-5.60 (m, 1 H), 4.42-4.57 (m, 1 H), 4.05-4.13 (m, 1 H), 3.68-3.82 (m, 7 H), 3.48-3.67 (m, 3 H), 3.30-

3.42 (m, 1 H), 3.21-3.29 (m, 1 H), 2.63 (t, *J*=5.96 Hz, 1 H), 2.50 (t, *J*=5.98 Hz, 1 H), 2.24-2.47 (m, 2 H), [1.81 (s, 1.5 H), 1.69 (s, 1.5 H)], 1.25-1.29 (m, 3 H), 1.11-1.17 (m, 9 H), 0.987-1.06 (d, *J*=6.80 Hz, 3 H), ppm; ^{13}C-NMR (Acetonitrile-d_3, 100 MHz): δ = 170.61, 170.58, 162.60, 159.67, 150.75, 145.89, 138.10, 138.04, 136.77, 136.67, 136.63, 136.61, 130.98, 128.97, 128.92, 128.87, 127.86, 127.84, 119.46, 119.31, 115.17, 115.06, 114.98, 114.08, 114.07, 87.22, 87.20, 87.17, 87.15, 86.01, 85.86, 67.02, 66.98, 66.82, 66.79, 64.28, 64.12, 59.52, 59.49, 59.33, 59.30, 55.85, 55.84, 44.00, 43.88, 39.93, 39.88, 39.83, 39.72, 39.62, 39.58, 24.87, 24.84, 24.79, 24.78, 24.12, 21.19, 21.06, 20.99, 20.91, 20.84, 19.64, 19.61, 19.47, ppm; ^{31}P-NMR (Acetonitrile-d_3, 162 MHz): δ = 148.06, 148.01, ppm; HRMS (ESI) calcd. for $C_{43}H_{54}O_{10}N_4P$ $[M+H]^+$: 817.35721; found: 817.35316. **Note**: ^{1}H Peaks marked with [ ] denote observed diastereotopic signals. The peak at 9.10 denotes the exchangeable proton N-H. ^{13}C has diastereomers reported together.

3. Synthesis and purification of oligonucleotides

DNA Oligonucleotides (Table S1) were synthesized using standard DNA phosphoramidite monomers (Glen Research) in an EXPIDITE Nucleic Acid Synthesis System (PerSeptive Biosystems). The modified 5-alkyl and 5-hydroxyalkyl (5eC **1**, 5pC **2**, 5iC **3**, 5-CH_3CD_2-C **4**, 5heC **5**, 5eU **6** and 5heU **7**) nucleoside phoshoramidites were synthesized. 5mC, 5hmC, 5hmU (Glen Research) and 6-Fluorescein (FAM) Phosphoramidite (Millipore-Sigma) were purchased from indicated vendors. Standard solid-phase DNA synthesis protocol was followed. To ensure good coupling elongated (4 min 30 sec) coupling times were applied for the coupling of modified bases and for standard bases (2 min) normal coupling was applied. FAM group was added at the 5'position of the oligonucleotides. Both FAM-labeled and unlabeled oligomers were removed from the resin and deprotected with ammonium hydroxide (28% v/v) at room temperature for overnight except for 5hmC which was deprotected at 75°C overnight. DMTr containing DNA was purified initially through a Poly Pak II purification cartridge (Glen Research). FAM-labeled and Poly Pak purified DNA were then purified through HPLC on a C-18 column with the following gradient: Solvent A: 0.1 M TEAA pH 7, Solvent B: Acetonitrile; 0 min 5% B, 10 min 40% B, 15 min 100% B with a flow rate of 4 mL/min. The appropriate fractions were collected, for FAM labeled DNA only the bands containing absorbances at both 260 and 495 nm were collected. All fractions were concentrated down with a SpeedVac concentrator, and if required the DNA was also gel purified through agarose and extracted with QIAEX II Gel Extraction Kit. All oligomers were re-dissolved in either nuclease free water or duplex buffer (100 mM KOAc and 30 mM HEPES pH 7.5,

Integrated DNA Technology) and were confirmed either through high resolution MALDI-TOF-MS or LCMS.

4. Expression and purification of proteins

(A) Expression and purification of TET2 in E. coli. The catalytic domain of human TET2 is comprised of a Cys-rich domain and a DSβH (also known as jelly-roll motif) domain with a large low-complexity insert. It has been reported by Yanhui Xu group that this unstructured insert poses significant challenge for crystallization.[7] By screening a series of deletion mutants, Xu group found TET2 (1099–1936 with residues 1481–1843 replaced by a 15-residue GS-linker GGGGSGGGGSGGGGS) to be the minimum catalytically active fragment suitable for bacterial expression. A 6xHis tag was introduced at the N-terminus for the purpose of initial purification. The TET2 expression construct used in the current study was obtained from the Xu group at Fudan University.

The wild type TET2 plasmid was transformed into *Escherichia coli* strain BL21 star (DE3) competent cells (Invitrogen) using pET-28b kanamycin-resistant vector.[8] A single colony was picked up and grown overnight at 37°C in 10 mL of Luria-Bertani (LB) broth in presence of 50 μg/mL kanamycin. The culture was diluted 100-fold and allowed to grow at 37°C to an optical density (OD_{600}) of 0.8, and protein expression was induced overnight at 17°C with 0.5 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). Proteins were purified as follows: harvested cells were resuspended in 15 mL lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 25 mM imidazole, Lysozyme, DNase, and Roche protease inhibitor cocktail). The cells were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 40 min at 4°C. The soluble extracts were subjected to Ni-NTA agarose resin (Thermo) according to manufacturer's instructions. After passing 20 volumes of washing buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 25 mM imidazole), proteins were eluted with a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 400 mM imidazole. Proteins were further purified by gel filtration chromatography (Superdex-200) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 10% glycerol. Purified proteins were concentrated using Amicon Ultra-10k centrifugal filter device (Merck Millipore Ltd.). The protein concentration was determined using Bradford assay kit (BioRad

Laboratories) with BSA as a standard. The concentrated proteins were stored at -80°C before use.

(B) Expression and purification of UHRF2 in E. coli. N-terminus 6xHis-tagged human UHRF2-SRA domain (residues 419-648) expression vector pET28a-LIC (Addgene ID: 28158) were obtained from Addgene. The plasmids were transformed into *Escherichia coli* strain BL21 star (DE3) competent cells (Invitrogen) using pET-28b kanamycin-resistant vector. A single colony was picked up and grown overnight at 37°C in 10 mL of Luria-Bertani (LB) broth in presence of 50 μg/mL kanamycin. The culture was diluted 100-fold and allowed to grow at 37°C to an optical density (OD_{600}) of 0.8, and protein expression was induced overnight at 17°C with 0.5 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). The cells were harvested by centrifugation at 4000 rpm for 30 minutes. Following bacterial cell lysis, expressed proteins were purified by nickel affinity chromatography and followed by size exclusion chromatography using a superdex-75pg FPLC column.

(C) Expression and purification of AID in E. coli. The C-terminus 6xHis-tagged human AID expression construct pET28b (Addgene ID: 15923) was obtained from Addgene.[9] The plasmid was transformed into *Escherichia coli* strain Rosetta2 (DE3) competent cells (Invitrogen) using pET-28b kanamycin-resistant vector. A single colony was picked up and grown overnight at 37°C in 10 mL of Luria-Bertani (LB) broth in presence of 50 μg/mL kanamycin and 35 μg/mL chloramphenicol. The culture was diluted 100-fold and allowed to grow at 37°C to an optical density (OD_{600}) of 0.8, and protein expression was induced overnight at 17°C with 1 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). The cells were harvested by centrifugation at 4000 rpm for 30 minutes. Proteins were purified as follows: 1 L harvested cells were resuspended in 15 mL buffer A (50 mM MES pH 6.5, 500 mM KCl, 1 mM PMSF, and Roche protease inhibitor cocktail). The resuspended cells were frozen at -80°C for 1hr, cells were thawed and lysed by pulsed sonication (Qsonica-Q700), centrifuged at 13,000 rpm for 40 min at 4°C. The soluble extracts were subjected to Ni-NTA agarose resin (Thermo) according to manufacturer's instructions. After passing 20 volumes of washing buffer (buffer A + 30 mM imidazole), proteins were eluted with elution buffer (buffer A + 150mM Imidazole). Proteins were further purified by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM MES pH 6.5, 500 mM KCl, 5% glycerol. Purified proteins were concentrated using Amicon Ultra-10k centrifugal filter device (Merck Millipore Ltd.). The protein concentration was determined using Bradford assay kit (BioRad Laboratories) with BSA as a

standard. The concentrated proteins were stored at -80°C before use. AID-Y116A mutant reported in the current work was generated using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) using appropriate mutagenic primers (Table S3). The mutant plasmid was transformed into XL10-Gold Ultracompetent Cells and extracted using GeneJET Plasmid MiniPrep Kit from Thermo Scientific and confirmed by DNA sequencing. The mutant was expressed and purified as stated for the wild type protein.

(D) Expression and purification of APOBEC3A in E. coli. Wild type (Accession number NM_145699) APOBEC-3A containing a C-terminus 6xHis-tag inserted into a pET21 ampicillin-resistant vector was provided by provided by Prof. Gronenborn from the University of Pittsburgh.[10] The plasmid was transformed into *Escherichia coli* strain BL21 star (DE3) competent cells (Invitrogen). A single colony was picked up and grown overnight at 37°C in 10 mL of Luria-Bertani (LB) broth in presence of 100 μg/mL ampicillin. The inoculated media was centrifuged at 4000 g for 10 mins at 4°C. The LB broth was discarded and resuspended in 1 mL of 1% glycerol minimal media supplemented with 0.3 mM leucine (GMML). The suspended cells were then diluted 500-fold and allowed to grow at 37°C to an optical density (OD600) of 0.8 and protein expression was induced for 20 hours at 17°C with 0.4 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). Proteins were purified as follows: 1 L of harvested cells were resuspended in 15 mL lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 25 mM imidazole, Lysozyme, DNase, and Roche protease inhibitor cocktail). The cells were lysed by pulsed sonication (Qsonica-Q700) and centrifuged at 13,000 rpm for 40 min at 4°C. The soluble extracts were subjected to Ni-NTA agarose resin (Thermo) according to manufacturer's instructions. After passing 20 volumes of washing buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 25 mM imidazole), proteins were eluted with a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 400 mM imidazole. Proteins were further purified by gel filtration chromatography (Superdex-75) using AKTA pure FPLC system (GE healthcare) with buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 10% glycerol. Purified proteins were concentrated using Amicon Ultra-10k centrifugal filter device (Merck Millipore Ltd.). The protein concentration was determined using Bradford assay kit (BioRad Laboratories) with BSA as a standard. The concentrated proteins were stored at -80°C before use. APOBEC 3A-Y130A mutant reported in the current work was generated using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) using appropriate mutagenic primers (Table S3). The mutant plasmid was transformed into XL10-Gold Ultracompetent Cells and extracted using

GeneJET Plasmid MiniPrep Kit from Thermo Scientific and confirmed by DNA sequencing. The mutant was expressed and purified as stated for the wild type protein.

(E) Expression and purification of TDG in E. coli. The N-terminus 6xHis-tagged human TDG bacterial expression construct pET28 kanamycin-resistant vector was obtained from Alexander C. Drohat laboratory, University of Maryland.[11,12] The wild type TDG plasmid was transformed into *Escherichia coli* BL21 star (DE3) competent cells (Invitrogen) using pET28 kanamycin-resistant vector. A single colony was picked up and grown overnight at 37°C in 10 mL of Luria-Bertani (LB) broth in the presence of 50 μg/mL kanamycin. The culture was diluted 100-fold and allowed to grow at 37°C to an optical density (OD_{600}) of 0.8, and protein expression was induced overnight at 17°C with 0.25 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific).

(F) Expression and purification of β-GT in E. coli. The N-terminus 6xHis-tagged T4 *β*-GT (*β*-glucosyltransferase) plasmid was a kind gift of Prof. Heinrich Leonhardt at the Ludwig Maximilians University, Munich.[13] The pET28b kanamycin-resistant vector was used to transform *E. Coli* BL21 codon plus (DE3) RIPL competent cells (Invitrogen). A single colony was picked up and grown overnight at 37 ^{0}C in 10 mL of Luria-Bertani (LB) broth in presence of 50 μg/mL kanamycin and 35 μg/mL chloramphenicol. The culture was diluted 100-fold and allowed to grow at 37°C to an optical density (OD_{600}) of 0.8, and protein expression was induced overnight at 18°C with 0.5 mM IPTG in an Innova 44® Incubator shaker (New Brunswick Scientific). Protein purification protocol was followed as described above.

5. Protocols for Biochemical Assays

(A) TET2 activity on synthetic DNAs. The substrate DNAs were first duplexed using BioRad T100 Thermal Cycler with the following parameters in duplex buffer (100 mM KOAc, and 30 mM HEPES pH 7.5): (1) 95°C for 5 min, (2) -1°C from 95°C every min over 40 min to 55°C, (3) 55°C for 30 min, (4) -1°C from 55°C every min over 20 min to 35°C. Duplexed DNAs were stored at -20°C. 10 μM of various double-stranded DNA substrates were incubated with 5-10 μM of TET2 (1099–1936 del-insert) in a 25 μL assay containing 50 mM HEPES (pH 8.0), 100 mM NaCl, 100 μM $Fe(NH_4)_2(SO_4)_2$, 2 mM ascorbate, 1 mM DTT, 1 mM ATP, and 1 mM 2-KG.[7,14] The samples were incubated for 3 hr. at 37°C. The product DNA was denatured at 100°C for 10 min and subsequently, the DNA was desalted by adding 5 μL of AG® 50W-X8 Cation Exchange Resin (BioRad, Cat # 143-5441) directly into the biochemical mixture and agitated followed by incubation

for 5 min at room temperature. The samples were centrifuged at 10,000 rpm for 2 min. The oxidized products were analyzed by MALDI-TOF mass spectrometry (Bruker-ultrafleXtreme™ MALDI-TOF/TOF spectrometer) by spotting 1 μL of sample and then mixed with 1 μL of 3-Hydroxypicolinic Acid (3-HPA) matrix on MALDI plate.

(B) Fluorescence Polarization-based binding Assay. Fluorescence polarization assay was performed according to a published protocol to measure binding of UHRF2-SRA (aa 419-648) domains to the 6-carboxy-fluorescein (FAM)-labeled duplexed DNAs (See Table S1 for sequence).[15,16,17] The purified proteins were dialyzed against a binding assay buffer overnight at 4°C. The experiments were performed in 384-well small volume black/clear microtiter plates (Falcon) with 10 nM FAM-labeled DNA and varying concentrations of proteins (0.005 to 100 μM) in a buffer containing (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, 14 mM β-mercaptoethanol, and 5% glycerol) at room temperature for 20 min. Fluorescence anisotropy was measured in a Tecan M1000 plate reader using its FP module and at excitation/emission wavelengths of 470 and 512 nm, respectively. For determinations of the dissociation constants (K_d), the background corrected polarization values were converted to fraction bound before plotted against the concentrations of proteins. The data were fitted to a single-site binding equation $Y = B_{max} X / (K_d + X)$, where Y is the specific binding, B_{max} is the maximal binding and X is the concentration of the ligand, using the SigmaPlot software.

(C) hTDG glycosylase assay of thymine, 5-ethyluridine (5eU), 5-hydroxymethyluridine (5hmU) and 5-hydroxyethyluridine (5heU) DNAs. For *in vitro* enzymatic activity assays, 0.5 μM of double-stranded fluorescein labeled 32-nt DNA containing either T or 5eU or 5hmU or 5heU modification was incubated with 10 μM hTDG in a 20 μL assay containing 10 mM Tris-HCl pH 8.0, 1 mM EDTA, and 0.1% BSA (See Table S1 for sequence of the DNAs).[18] The samples were incubated for 1 hr. at 37°C, then an additional 10 μM hTDG was added for a total of 20 μM in the assay solution and left at 37°C for another 1 hr. The reactions were quenched, and the abasic sites cleaved with 2 μL of 1 M NaOH that was heated to 95°C for 10 min. Then the loading buffer was added which contained 30 μL of formamide, 5 μL of 100 mM EDTA, and 4 μL of Novex™ TBE-Urea Sample Buffer (2X) was added and heated to 95°C for 10 mins. 20 μL of the final solution was loaded onto a 15%-acrylamide/24%-formamide urea gel and ran at 200 V in TBE (1X) buffer for 1 hr. Gels were immediately imaged on Bio-Rad ChemiDoc™ XRS+ platform.

(D) Deaminase Assay with wild type APOBEC-3A and Y130A mutant on 5hmC and 5heC DNAs. For *in vitro* enzymatic activity assays, 0.25 μM of single stranded fluorescein labeled 21-mer DNA containing either 5hmC or 5heC modification was incubated with 2 μM APOBEC A3A or its Y130A mutant in a 20 μL containing 20 mM Tris-HCl pH 8.0, 1 mM DTT, and 5mM EDTA.[19] See Table S2 for sequence of these DNAs. The samples were incubated for 16 hrs. at 37°C, then an additional 3 μM enzyme was added for a total of 5 μM in the assay solution and left at 37°C for another 2 hrs. Reactions were quenched by heating the solutions to 95°C for 10 mins. Then 0.5 μL of 132 μM compliment DNA (in duplex buffer) was added and duplexed according to the duplex protocol (described in section 8A). Then 0.55 μL of 2 mg/mL BSA was added along with 1 μL of 127 μM hTDG and left at 37°C for 16 hrs. followed by an additional 1 μL of 127 μM hTDG added and left at 37°C for 2 hrs. The reactions were quenched, and the abasic site was cleaved with 2 μL of 1 M NaOH upon heating at 95°C for 10 min. A loading buffer was added that contained 30 μL of formamide and 5 μL of 100 mM EDTA and heated to 95°C for 10 min. The samples were processed as described above for the hTDG assay.

(E) Deaminase Assay with wild type AID and Y116A mutant on 5hmC and 5heC DNAs. For *in vitro* enzymatic activity assays, 0.25 μM of single stranded fluorescein labeled 21-mer DNA containing either 5hmC or 5heC and was incubated with 20 μM wild type AID or Y116A mutant in a 20 μL containing 20 mM Tris-HCl pH 8.0, 1 mM DTT, and 5mM EDTA.[20] See Table S1 for sequence of these DNAs. The samples were incubated for 1 hr. at 37°C, then an additional 20 μM enzyme was added for a total of 40 μM in the assay solution at 37°C for another 1 hr. Then 0.5 μL of 132 μM compliment DNA (in duplex buffer) was added and duplexed according to the duplex protocol (described in section 8A). Then 0.55 μL of 2 mg/mL BSA was added along with 1 μL of 127 μM hTDG and left at 37°C for 1 hr, followed by an additional 1 μL of 127 μM hTDG and left at 37°C for 1 hr. The remaining steps were performed exactly the same as described for APOBEC 3A in the preceding section.

(F) Oxidation of 5hmC and 5heC DNAs with $KRuO_4$ followed by hTDG glycosylase Assay. Single stranded fluorescein labeled 32-mer DNA containing either 5hmC or 5heC modification was diluted to 10 μM DNA with 24 μL of 0.05 M NaOH, and 1 μL of $KRuO_4$ solution (15 mM in 0.05 M NaOH). Incubation conditions varied by substrate, 5hmC oxidation was done on ice for 1 hr, 5heC oxidation was carried out at rt. overnight. After oxidation, the DNA was desalted using QIAEX II Gel Extraction Kit and was concentrated with a SpeedVac concentrator. The DNA

residue was dissolved in 10 μL duplex buffer and 10 μL of 48.7 μM compliment DNA (in duplex buffer) was added. The DNA was duplexed according to the duplex protocol and the hTDG glycosylase protocol was followed as described above.

(G) β-glucosyltransferase (β-GT)-mediated glucosylation of 5hmC and 5heC. For 5hmC and 5heC glucosylation, first a demethylase assay was carried out using wild type TET2 on 8-nt 5mC (or 5eC) containing double-stranded DNA substrate to obtain 5hmC (5heC) DNA. A 60 μL of demethylase assay containing 5 μM TET2, 10 μM 8-nt 5mC (or 5eC) DNA and 100 μM 2-KG **1** was incubated at 37°C for 1 hr. in buffer containing 50 mM HEPES (pH 8.0), 100 mM NaCl, 100 μM $Fe(NH_4)_2(SO_4)_2$, 2 mM ascorbate, 1 mM DTT and 1 mM ATP. After 1 hr. of incubation, 20 μL of assay mixture was withdrawn to confirm 5hmC (or 5heC) formation. 15 μM of T4-β-GT, 100 μM UDP-glucose (Sigma-Aldrich, Cat # U4625) or UDP-N_3-glucose (MP Biomedicals, LLC. Cat # 101208) were added to each 20 μL of assay and the reaction carried out at 37°C for 2 hr.[21] After 2 hr. of incubation, 20 μL of assay mixture was purified using QIAquick Nucleotide Removal Kit (QIAGEN) following manufacturer's instructions. The oligonucleotides were further concentrated using speedvac concentrator and analyzed by MALDI-TOF mass spectrometry.

(H) Labeling of 5hmC and 5heC by $KRuO_4$ oxidation followed by biotinylation. 25 μL assay containing 50 μM single-stranded DNA containing 5hmC or 5heC was incubated with 1 μL of $KRuO_4$ solution for 15 mM in 0.05 M NaOH, the rest of the volume was made up with 0.05 M NaOH. Incubation conditions varied by substrate, 5hmC oxidation was done on ice for 1 hr., 5heC oxidation was carried out at room temperature overnight. After oxidation, the DNA was desalted using QIAEX II Gel Extraction Kit following manufacturer's protocol and was concentrated with a SpeedVac concentrator. The DNA was then dissolved in 10 μL of 75 mM MES buffer pH 6.0, then 1.1 μL of 100 mM of biotin-oxime based aldehyde-reactive probe trifluoroacetate salt (Cayman Chemical, cat #10009350) was added. The solution was agitated at 1400 rpm for 2hr. at 40°C. Reaction product was analyzed by dot-blot using anti-biotin antibody. First, ¼ volume of 2 M NaOH–50 mM EDTA was added to the DNA which was then denatured for 10 min at 95°C and transferred quickly to ice, followed by addition of 1:1 ice cold 2 M ammonium acetate. Immobilin-P PVDF membranes (Merck Millipore Ltd.) were cut to size, wet with MeOH for 20 sec, and equilibrated in TE buffer for 5min, then assembled into a 96-well Bio-Dot microfiltration apparatus (Bio-Rad).[22] Each well was washed with 400 μl TE drawn through with gentle vacuum, and the denatured DNA was loaded, followed by another TE wash. The membrane was blocked for 2 hr.

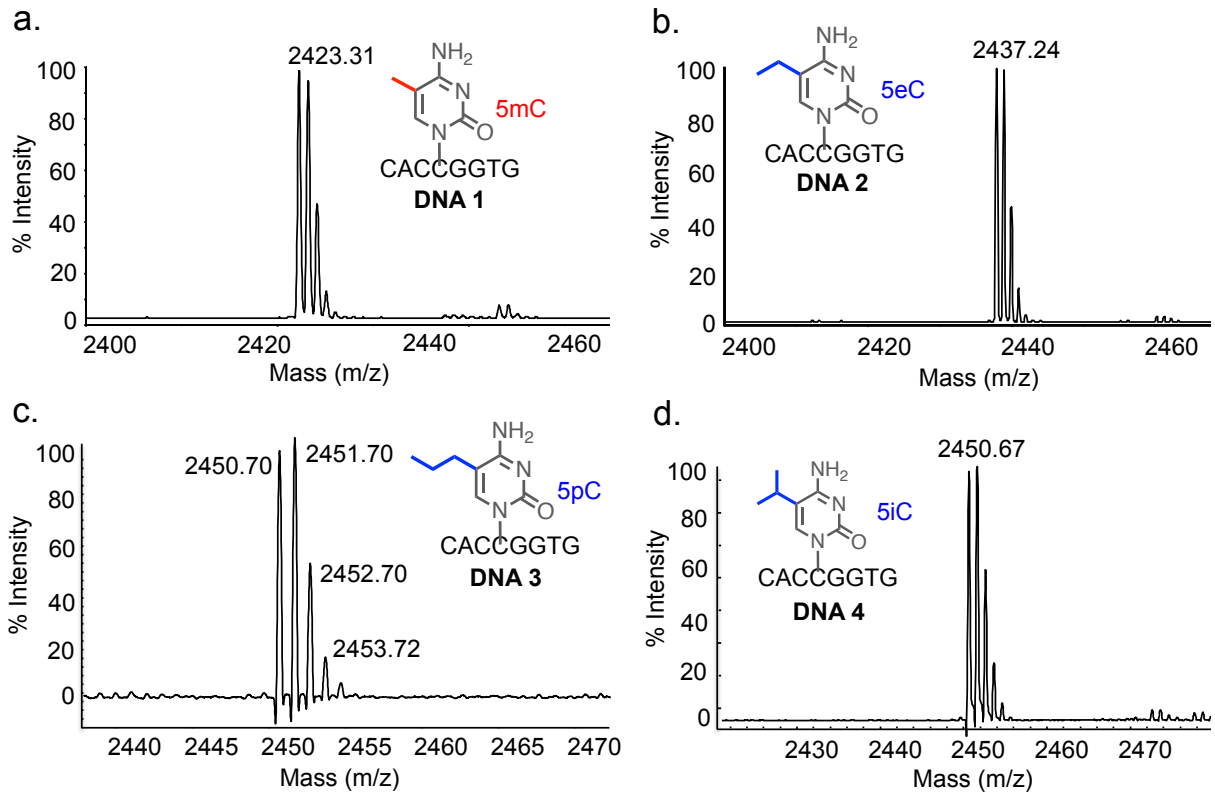
in 5% milk–TBST, washed 3× with TBST, and blotted at 4°C overnight with anti-biotin primary antibody (1:500 dilution, Biotin pAb, catalog no. PA1-26792, Invitrogen) for 16 hr. at 4°C. The antibody solutions were removed, and membrane was washed three times with TBST buffer at room temperature. The blots were then incubated with HRP-conjugated secondary antibody donkey anti-goat IgG (H+L) pAb (catalog no. NBP1-74815, Novus) with 5% nonfat dry milk (1:500 dilution) in TBST for 2 hr. at room temperature. After similar washing, imaged by chemiluminescence using VISIGLO HRP Chemiluminescent substrates A and B (cat. no. N252-120ML and N253-120ML, aMReSCO) following manufacturer's protocol.

6. **Acquisition and analysis of MALDI data.** The MALDI data were collected on Bruker-ultrafleXtreme™ MALDI-TOF/TOF mass spectrometers. The experiments were conducted in reflectron negative TOF mode and the ion intensities were normalized to the highest peak. The background noise was corrected for calculating signal/noise ratios using AB Sciex Data Explorer.

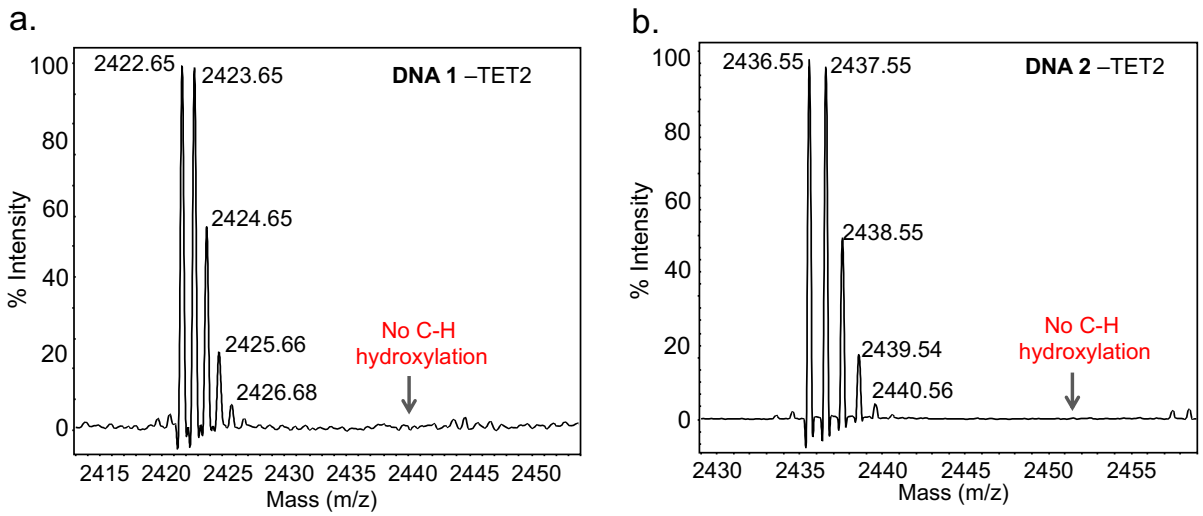
7. Supplementary figures and tables

DNAs	Sequence	Therotical Mass	Observed Mass	MS Figure	Purpose	Complementarity
DNA 1	5'-CAC5mCGGTG-3'	2423.46	2423.31	Figure S1	TET2 activity	Self
DNA 2	5'-CAC5eCGGTG-3'	2437.48	2437.24	Figure S1	TET2 activity	Self
DNA 3	5'-CAC5pCGGTG-3'	2451.49	2451.7	Figure S1	TET2 activity	Self
DNA 4	5'-CAC5iCGGTG-3'	2451.49	2451.67	Figure S1	TET2 activity	Self
DNA 5	5'-CAC5CH3CD2CGGTG-3'	2439.49	2439.35	Figure S5	TET2 activity	Self
DNA 6	5'-CTGGTC5hmCGGATG-3'	3705.65	3705.5	Figure S6	UHRF2 binding	With **DNA 8**
DNA 7	5'-CTGGTC5heCGGATG-3'	3719.67	3719.3	Figure S6	UHRF2 binding	With **DNA 8**
DNA 8	5'- FAM-CCATCCGGACCA-3'	4110.77	4110.75	Figure S6	UHRF2 binding	With **DNA 6 and 7**
DNA 9	5'- FAM-TCGGATGTTGTGGGTCAG**T**GCATGATAGTGTA-3'	10513.78	10513.84	Figure S8	TDG activity	With **DNA 13**
DNA 10	5'- FAM-TCGGATGTTGTGGGTCAG5hmUGCATGATAGTGTA-3'	10529.78	10529.92 - 10541.90	Figure S8	TDG activity	With **DNA 13**
DNA 11	5'- FAM-TCGGATGTTGTGGGTCAG5eUGCATGATAGTGTA-3'	10527.8	10527.82	Figure S8	TDG activity	With **DNA 13**
DNA 12	5'- FAM-TCGGATGTTGTGGGTCAG5heUGCATGATAGTGTA-3'	10543.8	10543.90 - 10553.85	Figure S8	TDG activity	With **DNA 13**
DNA 13	5'–TACACTATCATGCGCTGACCCACAACATCCGA–3'	9677.66	9677.69 - 9688.69	Figure S8	TDG activity	With **DNAs 9, 10, 11** and **12**
DNA 14	5' - FAM-AAAAAAAAAT5mCGGGAAAAAAA - 3'	7078.35	7077.35 - 7086.39	Figure S9	Coupled AID/A3A-TDG activity	With **DNA 18**
DNA 15	5'- FAM-AAAAAAAAAT5hmCGGGAAAAAAA-3'	7094.35	7093.35 - 7101.36	Figure S9	Coupled AID/A3A-TDG activity	With **DNA 18**
DNA 16	5' - FAM-AAAAAAAAAT5eCGGGAAAAAAA - 3'	7092.366	7091.36 - 7101.39	Figure S9	Coupled AID/A3A-TDG activity	With **DNA 18**
DNA 17	5'- FAM-AAAAAAAAAT5heCGGGAAAAAAA-3'	7108.36	7106.38 - 7115.38	Figure S9	Coupled AID/A3A-TDG activity	With **DNA 18**
DNA 18	5'-TTTTTTTCCCGATTTTTTTTT–3'	6312.03	6311.05 - 6318.03	Figure S9	Coupled AID/A3A-TDG activity	With **DNA 14, 15, 16** and **17**
DNA 19	5'-FAM-TCGGATGTTGTGGGTCAG5hmCGCATGATAGTGTA-3'	10528.8	10528.86 - 10539.90	Figure S11	KRuO4 oxidation-TDG activity	With **DNA 13**
DNA 20	5'- FAM-TCGGATGTTGTGGGTCAG5heCGCATGATAGTGTA-3'	10542.88	10541.83 - 10551.85	Figure S11	KRuO4 oxidation-TDG activity	With **DNA 13**

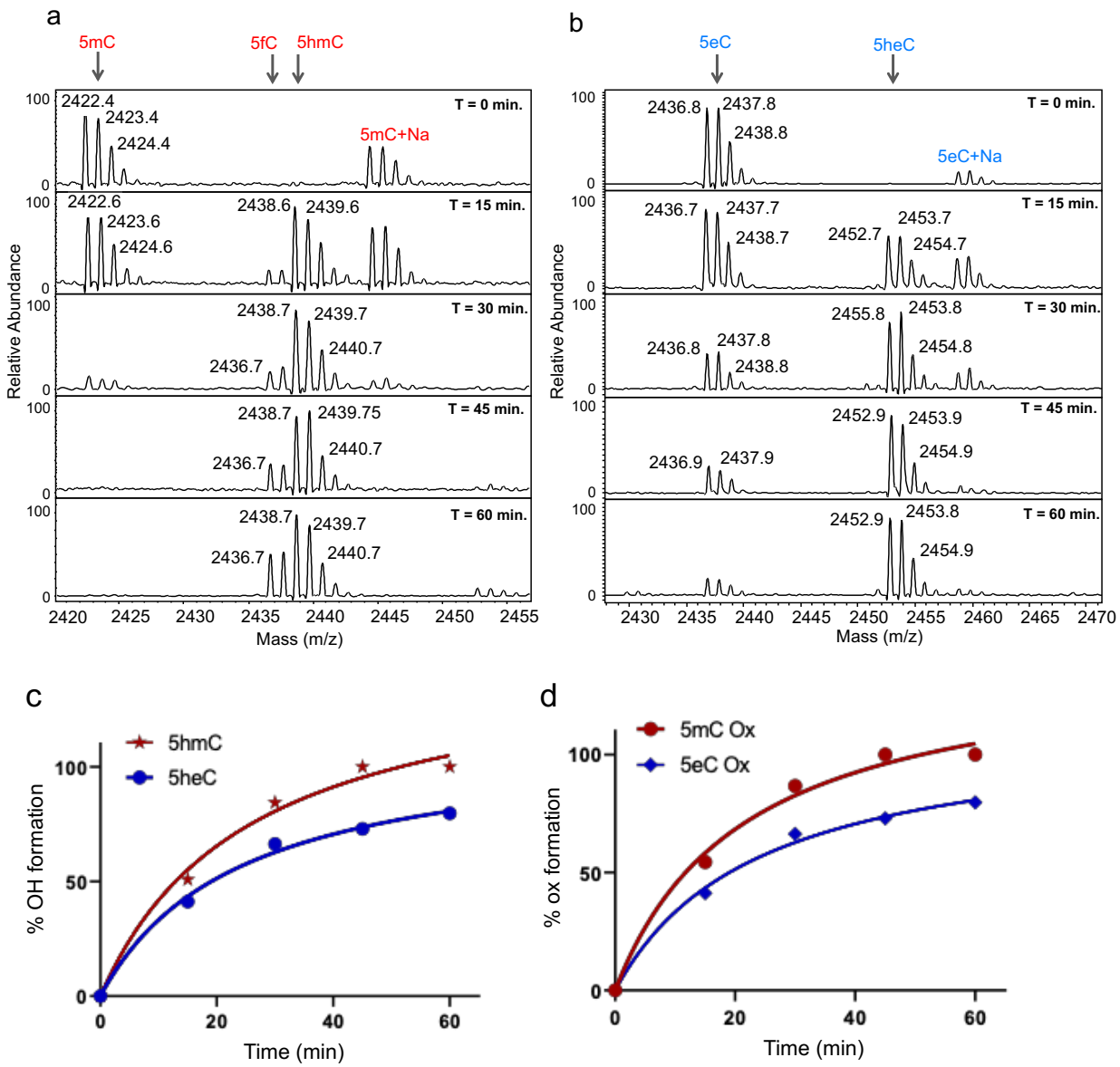
Supplementary Table S1. List of modified oligonucleotides used in the current study. A total of 20 DNAs are synthesized and characterized either by MALDI-HRMS or ESI-HRMS. Supplementary figure numbers are provided for the corresponding high-resolution mass spectra.



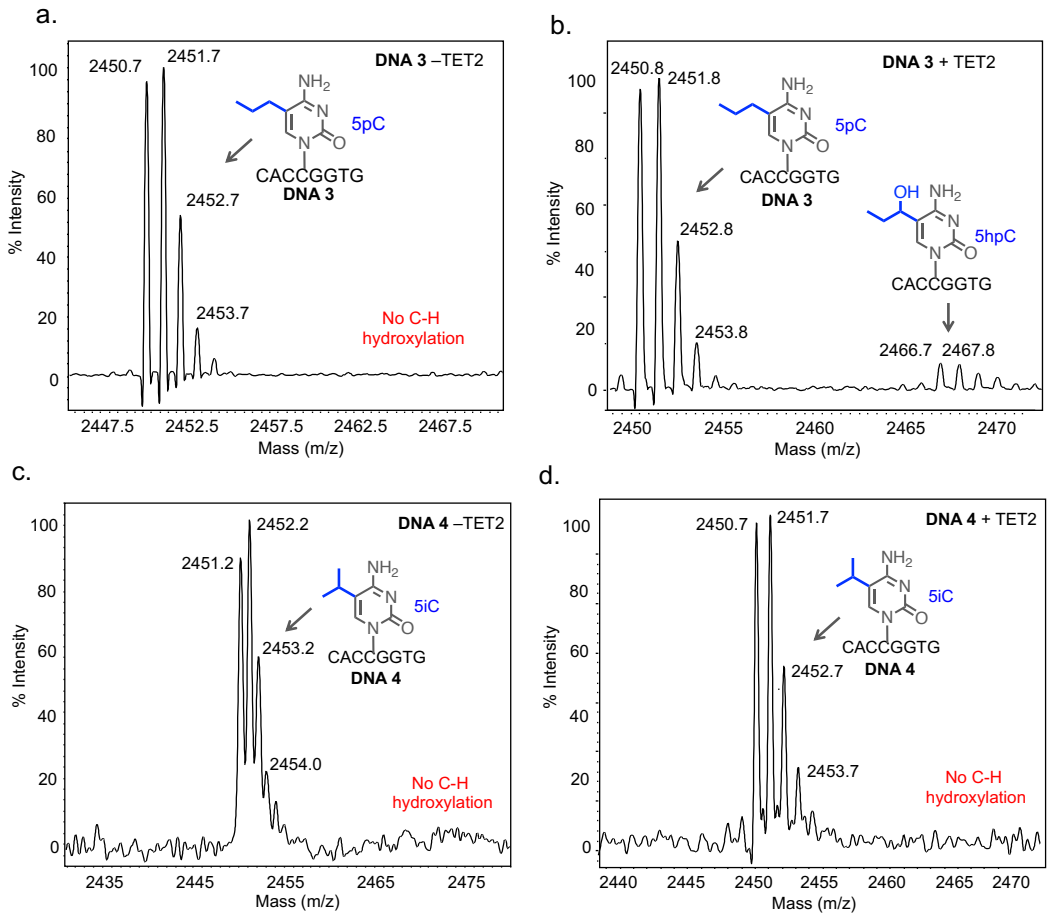
Supplementary Figure S1. MALDI mass spectra of DNAs **1** – **4** containing 5mC (a), 5eC (b), 5pC (c) and 5iC (d), respectively.



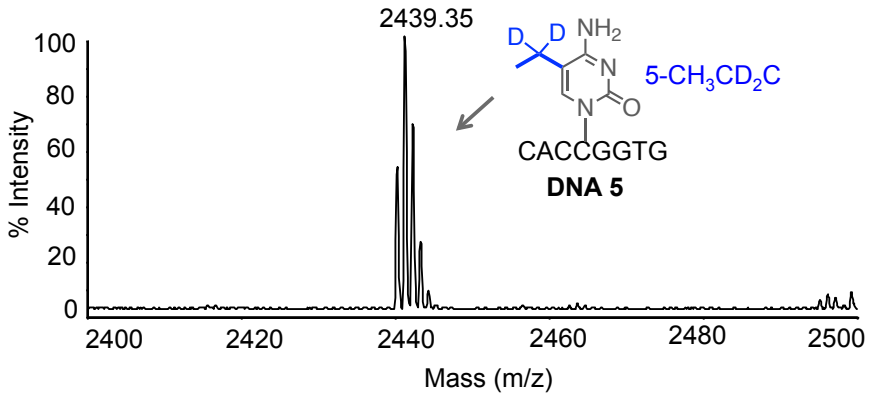
Supplementary Figure S2. MALDI mass spectra showing no C-H hydroxylation of **DNA 1** (a) and **DNA 2** (b) in the absence of TET2.



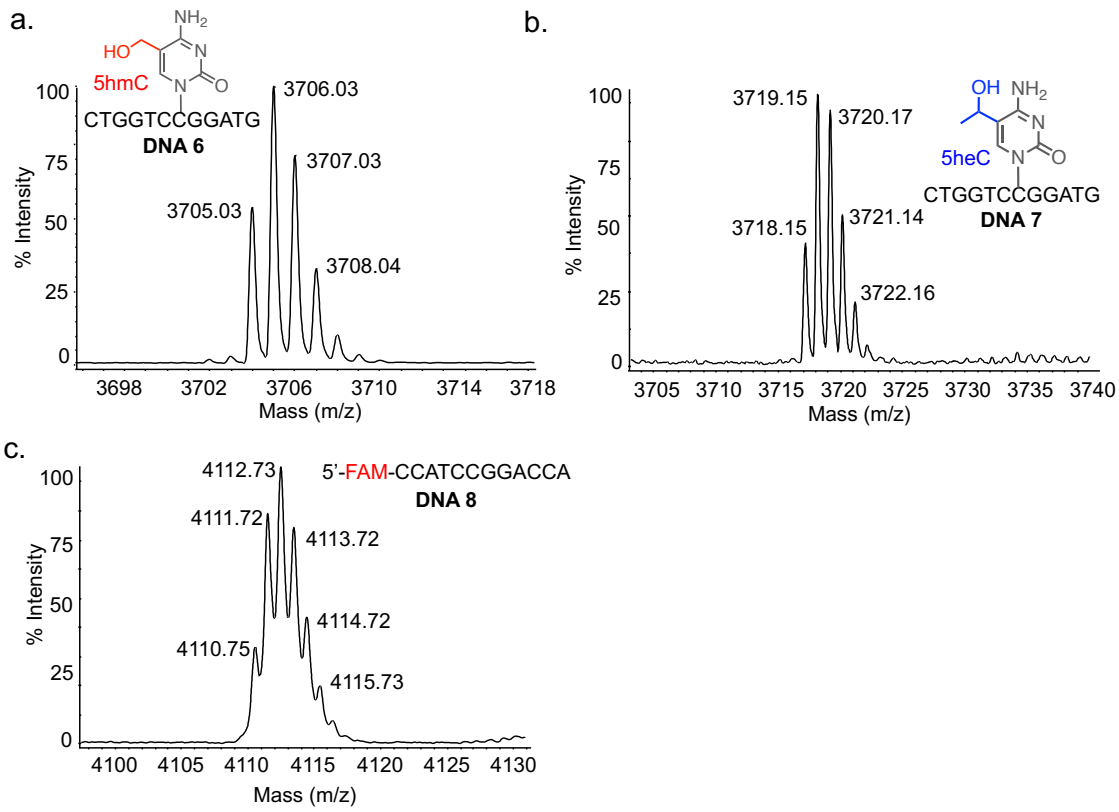
Supplementary Figure S3. Time-dependent activity of TET2 on 5mC (a) and 5eC (b) as determined by MALDI-MS. (c) Plot of % 5hmC and 5heC formation as the function of time (based on the results provided in (a) and (b). (d) Plot of % of total oxidized product (5hmC and 5fC from 5mC and 5heC from 5eC) formation as the function of time (based on the results provided in (a) and (b).



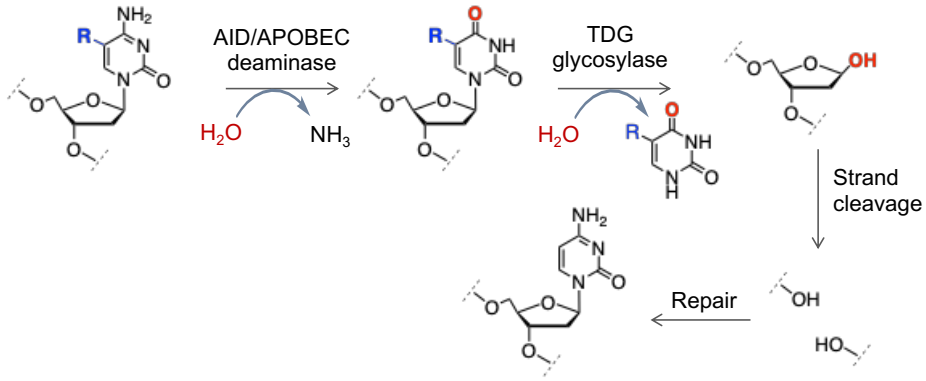
Supplementary Figure S4. MALDI spectra showing activity of TET2 on DNA **3** (b) and DNA **4** (d). (a) and (c) are corresponding no enzyme controls. DNA **3** is marginally oxidized to give 5-hydroxypropylcytidine (5hpC). **DNA 4** was not oxidized by TET2. These assays were performed by incubating 10 μM of TET2 with 10 μM of DNA substrates for 3 hr. at 37°C.



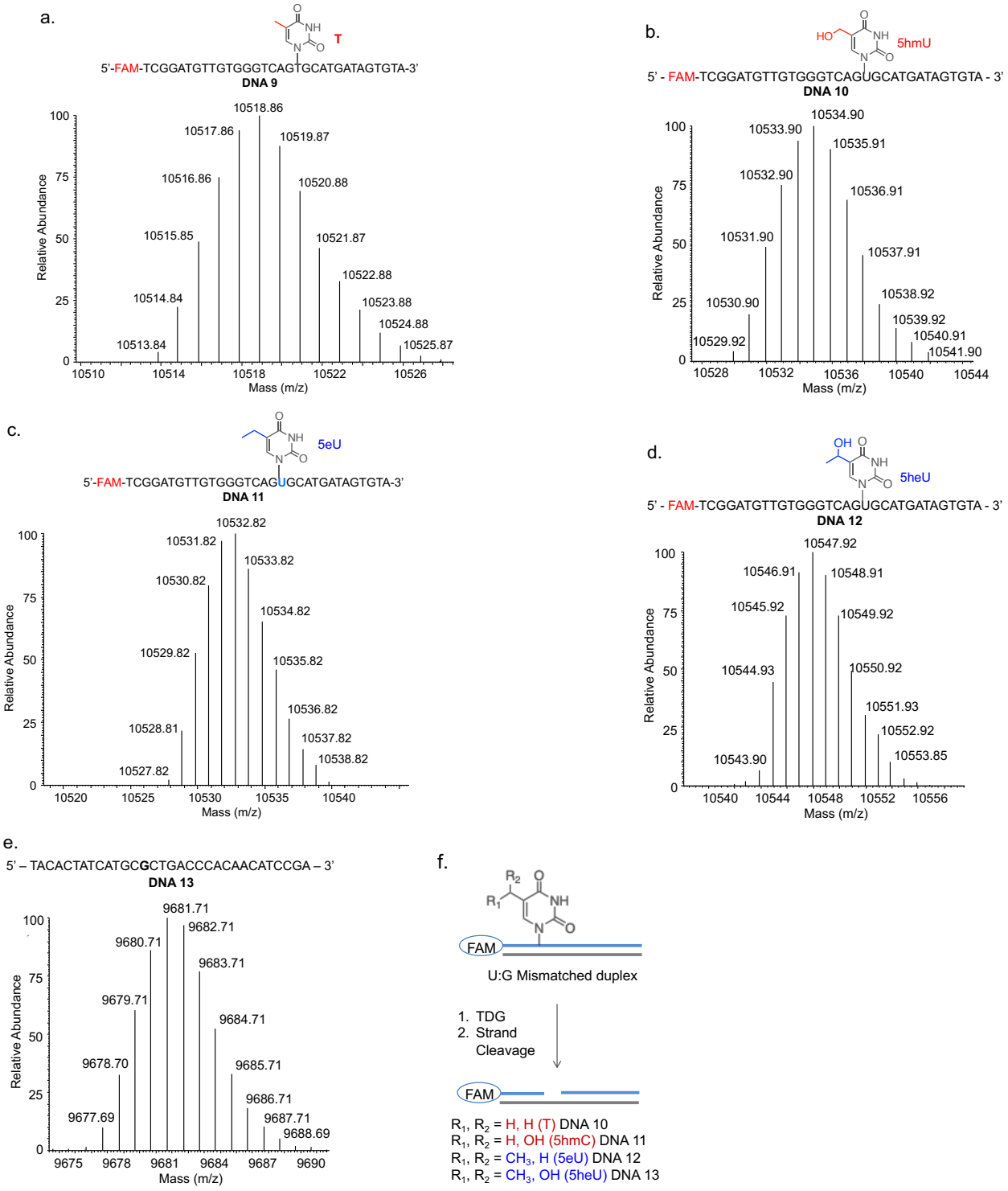
Supplementary Figure S5. MALDI spectrum of DNA **5** carrying 5-CH_3CD_2-cytidine.



Supplementary Figure S6. MALDI MS spectra of DNAs **6** – **8** containing 5hmC (a), 5heC (b) and FAM-attached DNA (c), respectively. The DNAs are employed to examine the binding to UHRF2-SRA domain.

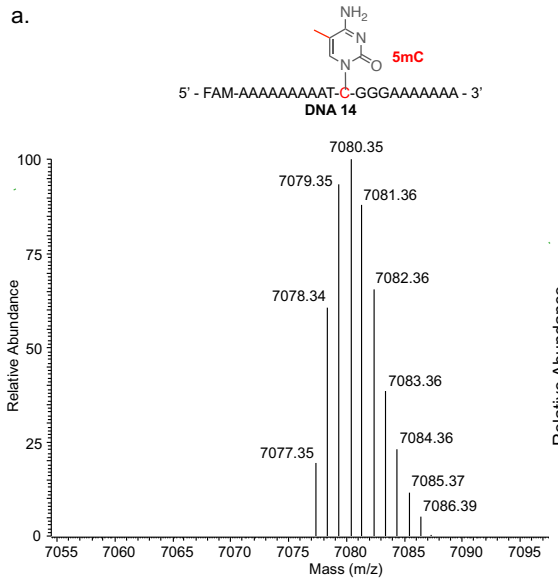


Supplementary Figure S7. Schematic showing active demethylation of 5hmC (R= -CH_2OH) by AID/APOBEC mediated deamination followed by TDG catalyzed base excision, strand cleavage and repair.

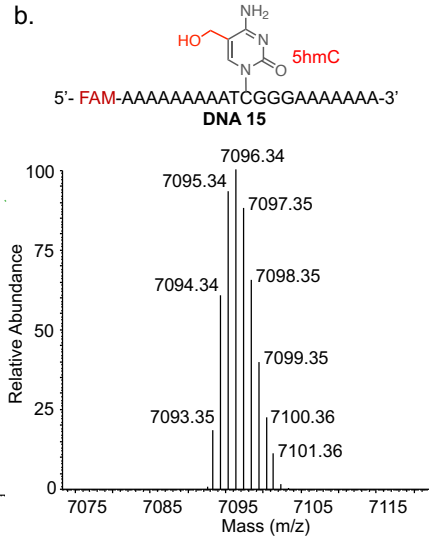


Supplementary Figure S8. (a-e) Deconvoluted ESI-HRMS spectra of DNAs **9** – **13.** DNA **13** was duplexed with DNAs **9, 10, 11** and **12** to generate U:G mismatches prior to TDG-mediated base excision. (f) Schematic representation of TDG activity on duplexed DNAs followed by alkaline strand cleavage. The cleaved DNA product carrying a 5'-FAM group is expected to appear as lower band in denaturing gel.

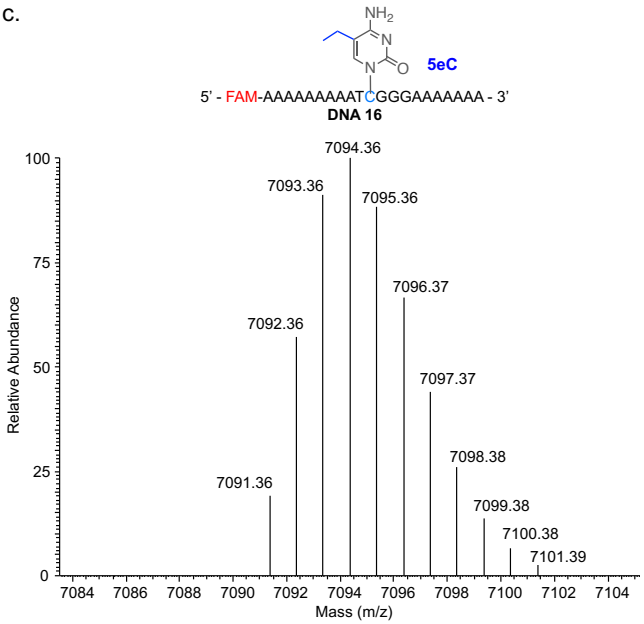
a.


5mC
5' - FAM-AAAAAAAAAT-C-GGGAAAAAAA - 3'
DNA 14
Relative Abundance
Mass (m/z)
7080.35
7079.35
7081.36
7078.34
7082.36
7083.36
7084.36
7077.35
7085.37
7086.39

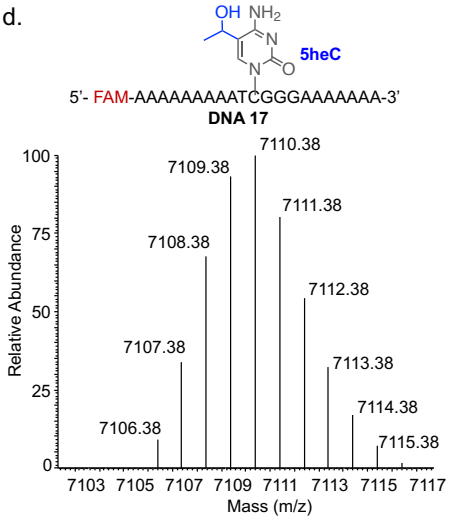
b.


5hmC
5'- FAM-AAAAAAAAATCGGGAAAAAAA-3'
DNA 15
Relative Abundance
Mass (m/z)
7096.34
7095.34
7097.35
7094.34
7098.35
7099.35
7093.35
7100.36
7101.36

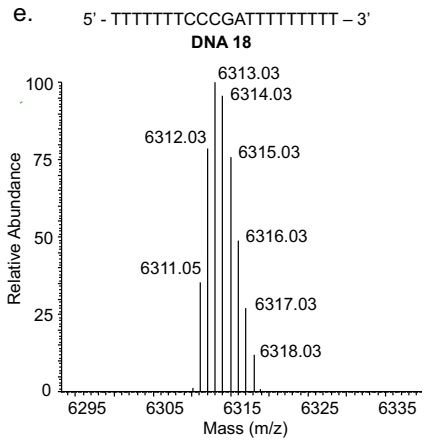
c.


5eC
5' - FAM-AAAAAAAAATCGGGAAAAAAA - 3'
DNA 16
Relative Abundance
Mass (m/z)
7094.36
7093.36
7095.36
7096.37
7092.36
7097.37
7098.38
7091.36
7099.38
7100.38
7101.39

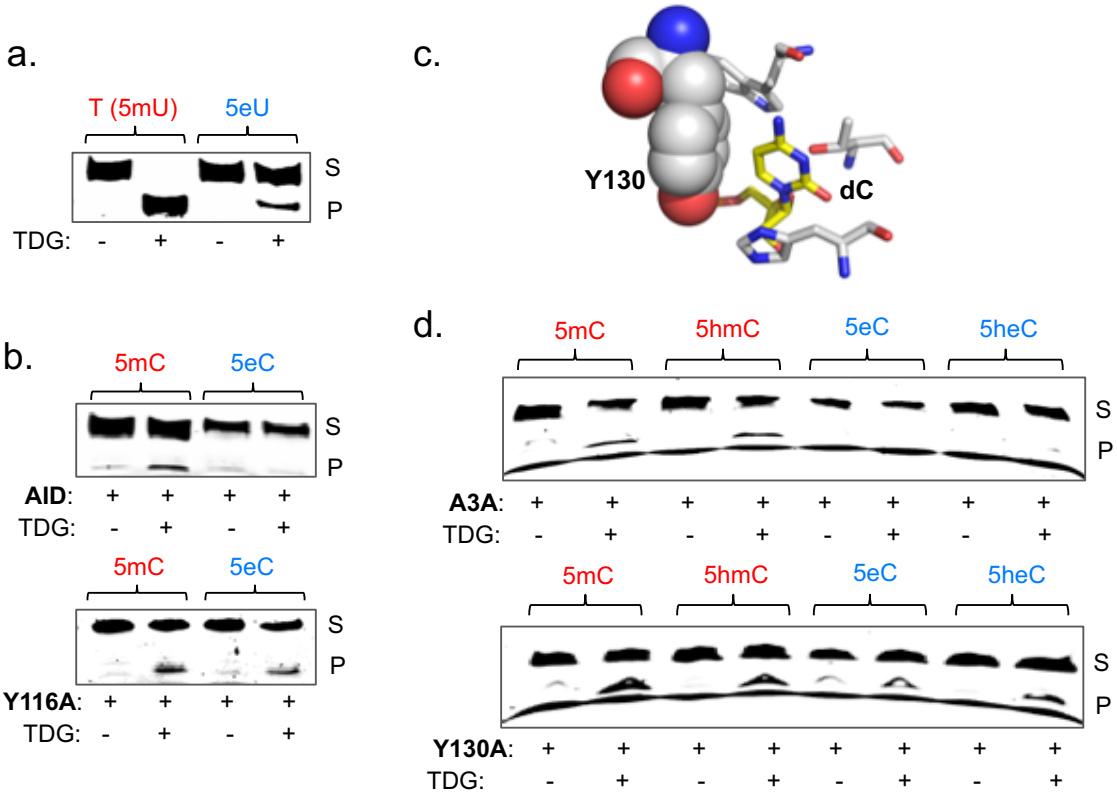
d.


5heC
5'- FAM-AAAAAAAAATCGGGAAAAAAA-3'
DNA 17
Relative Abundance
Mass (m/z)
7110.38
7109.38
7111.38
7108.38
7112.38
7107.38
7113.38
7114.38
7106.38
7115.38

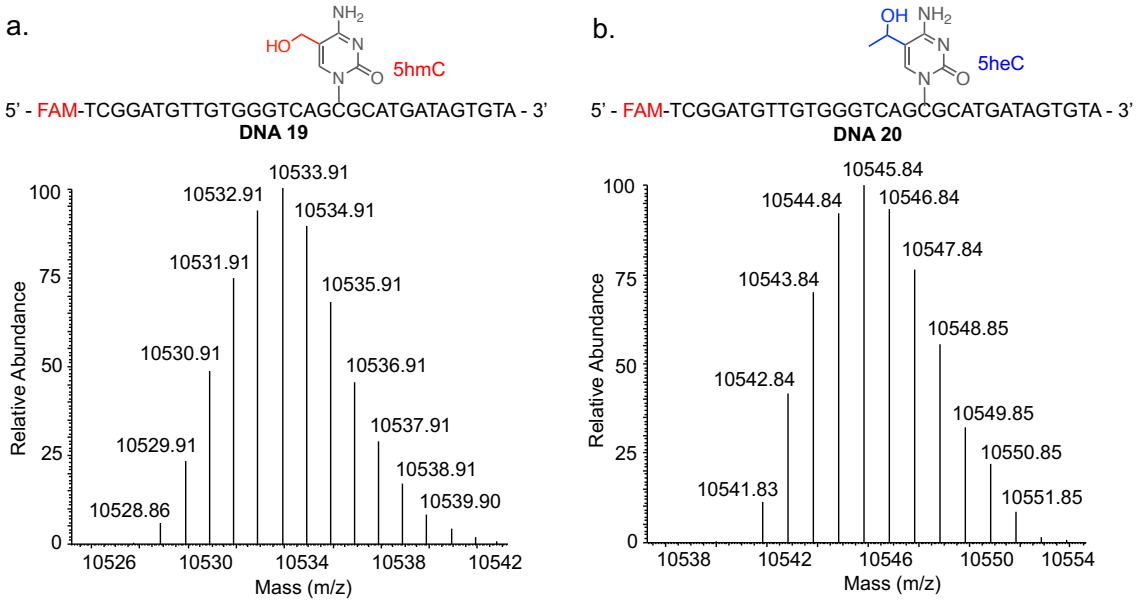
e.


5' - TTTTTTTCCCGATTTTTTTTT – 3'
DNA 18
Relative Abundance
Mass (m/z)
6313.03
6314.03
6312.03
6315.03
6316.03
6311.05
6317.03
6318.03

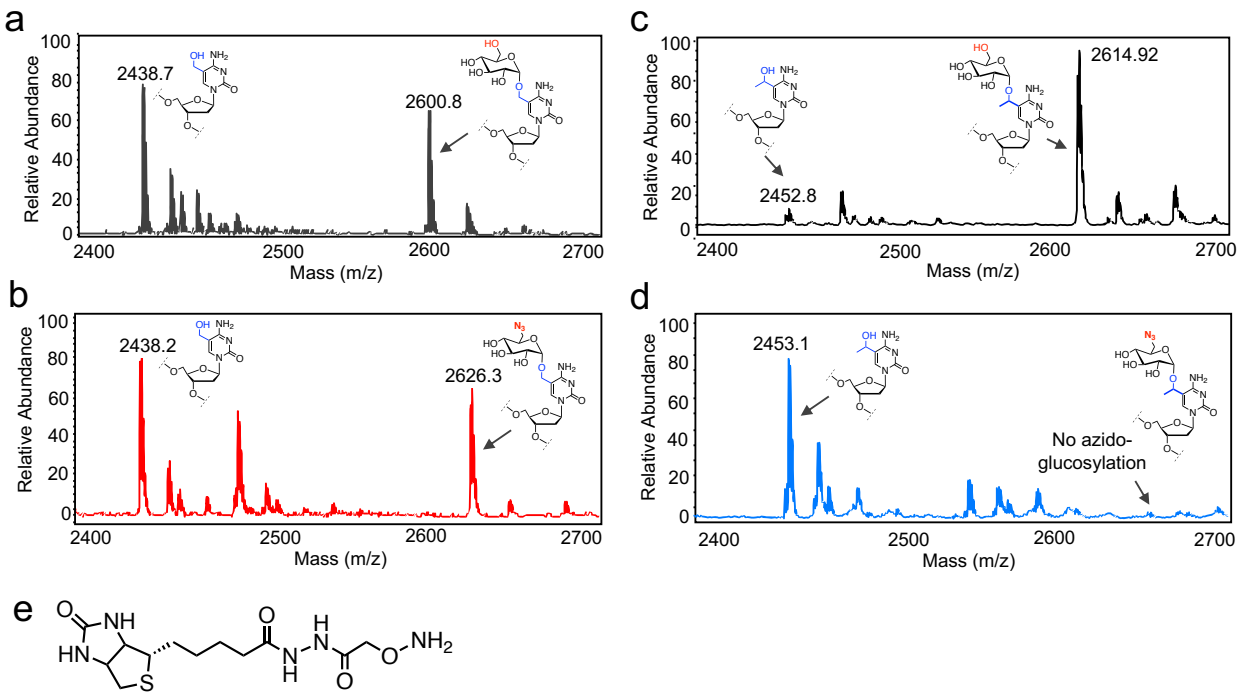
Supplementary Figure S9. Deconvoluted ESI-HRMS spectra of DNA **14** (a), DNA **15** (b), DNA **16** (c), DNA **17** (d) and DNA **18** (e). DNA **18** was duplexed with each of DNA **14**, **15**, **16** and **17** and subjected to cytidine deamination by wild type AID/APOBEC 3A and their mutants. The deaminated products were subjected to TDG-mediated base excision and strand cleavage.



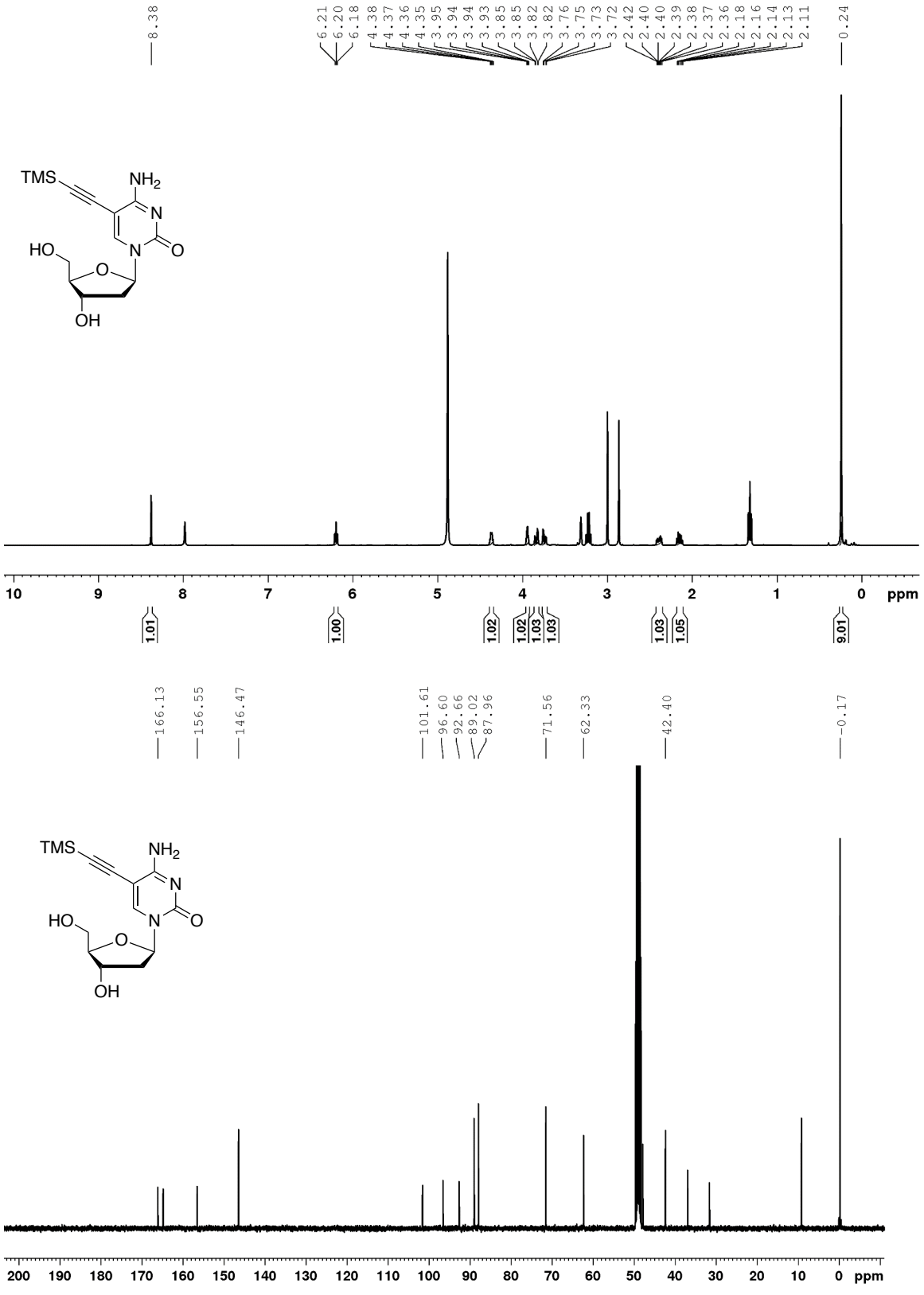
Supplementary Figure S10. (a) TDG can excise both T (5mU) and 5eU duplexed DNAs carrying a T:G and 5eU:G mismatches, respectively, as confirmed by in-gel fluorescence. (b) Activity of wild type AID and its Y116A mutant towards DNA strands carrying either 5mC or 5eC modifications. Deaminated DNAs were duplexed prior to TDG-mediated base excision and strand cleavage. (c) Crystal structures of human A3A (PDB 5keg). Conserved tyrosine Y130 resides closely to substrate cytidine in the active site. (d) Activity of wild type A3A and its Y130A mutant towards DNA strands carrying either 5mC or 5hmC or 5eC or 5heC modification. Deaminated DNAs were duplexed prior to TDG-mediated base excision and strand cleavage.



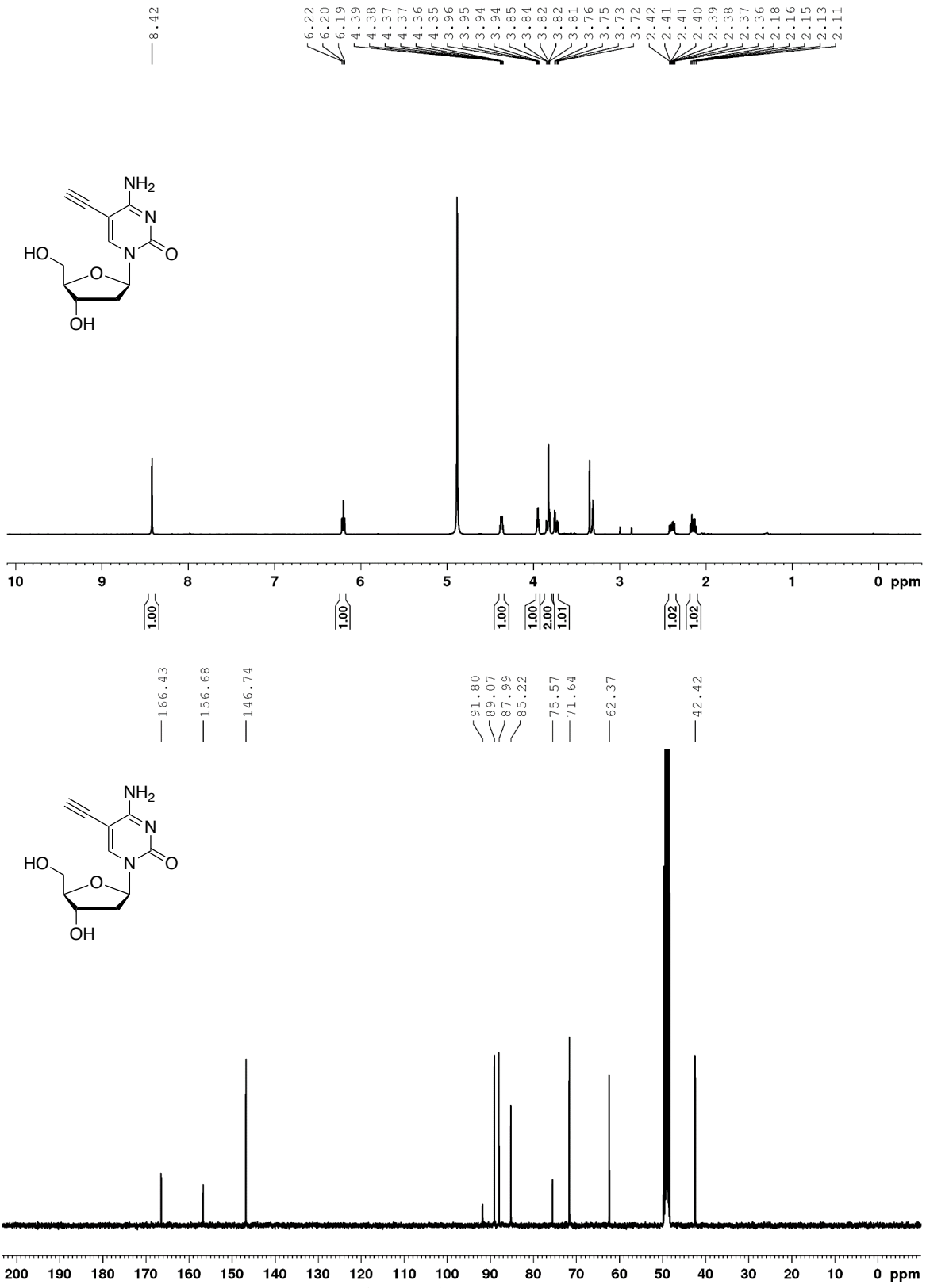
Supplementary Figure S11. (a) Deconvoluted ESI-HRMS spectra of DNAs **19** (a) and **20** (b) carrying 5hmC and 5heC, respectively. These DNAs were duplexed with DNA **13** (Figure S8) and subjected to $KRuO_4$ oxidation followed by TDG-mediated base excision and strand cleavage as shown in Figure 6 in the manuscript.



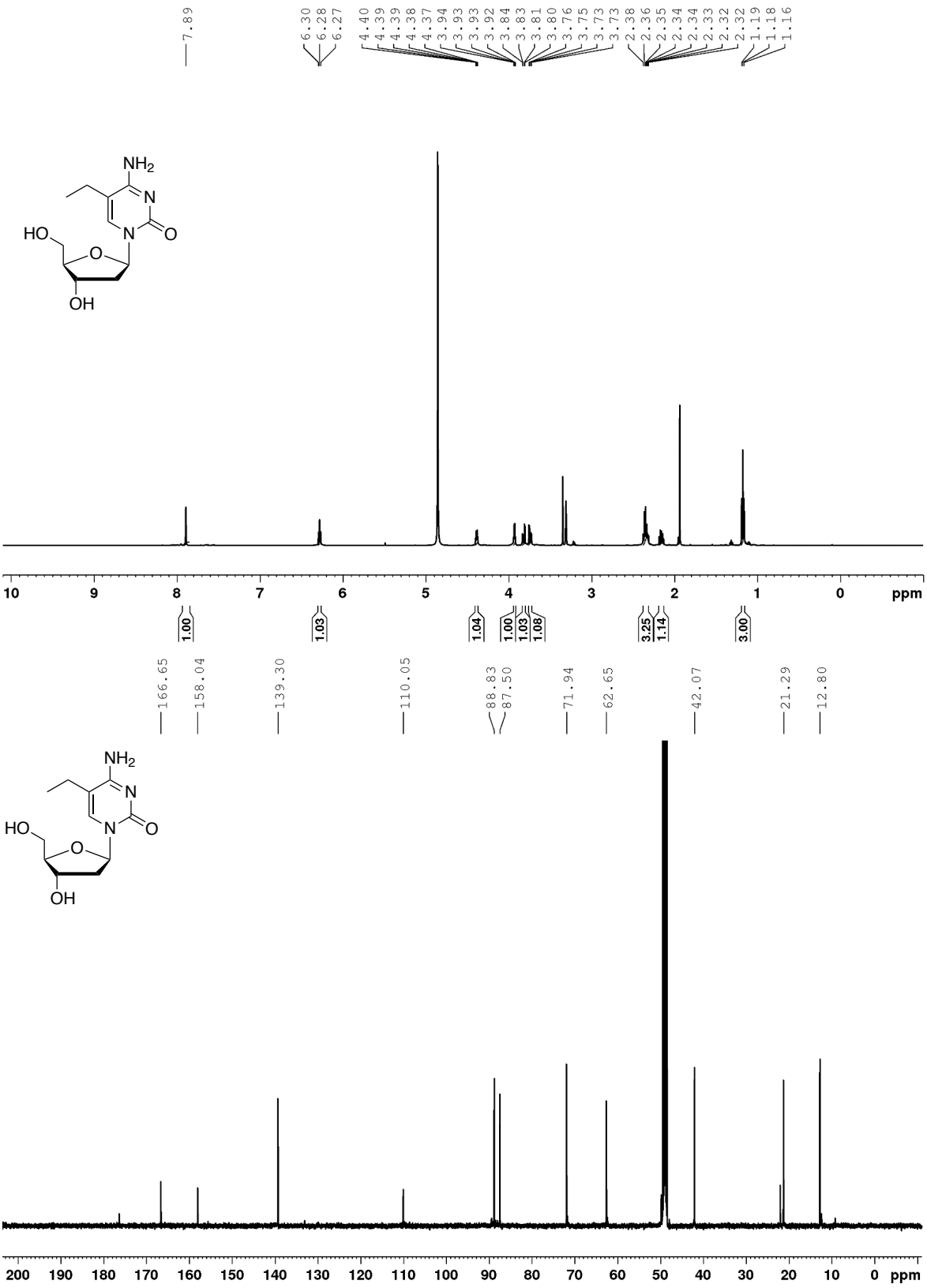
Supplementary Figure S12. MALDI-MS showing: (a) β-GT mediated glucosylation of 5hmC which was obtained by TET2-mediated oxidation of DNA **1** (Supplementary Table 1); (b) β-GT mediated 6-azido-glucosylation of 5hmC which was obtained by TET2-mediated oxidation of DNA **1** (Supplementary Table 1); (c) β-GT mediated glucosylation of 5heC which was obtained by TET2-mediated oxidation of DNA **2** (Supplementary Table 1); and (d) β-GT cannot transfer 6-azido-glucose unit to 5heC which was obtained by TET2-mediated oxidation of DNA **2** (Supplementary Table 1). e) Exact structure of the biotin-hydroxylamine ARP reagent used in the current study.



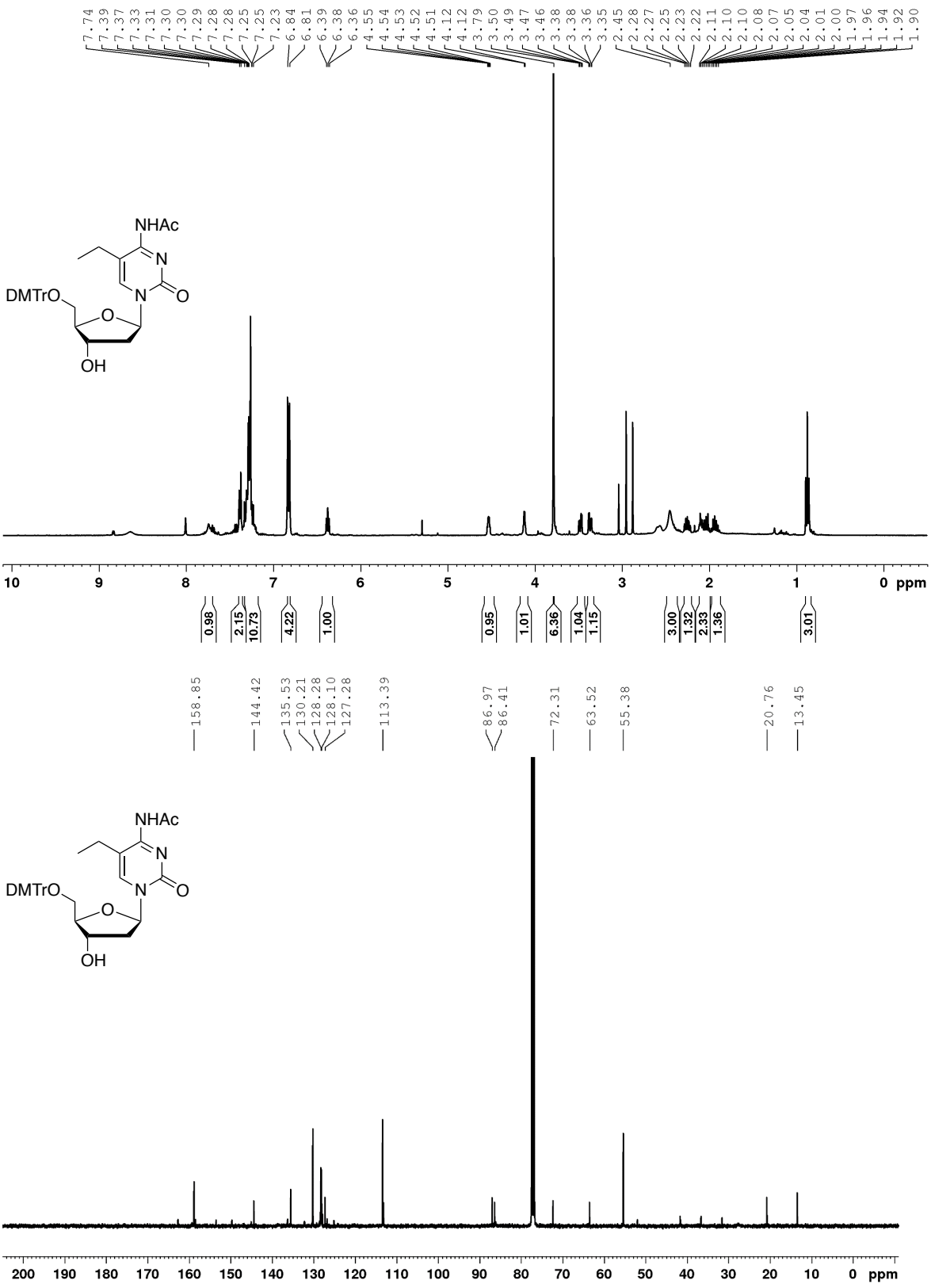
Supplementary Figure S13. ^{1}H and ^{13}C NMR spectra of the indicated compound in CD_3OD.



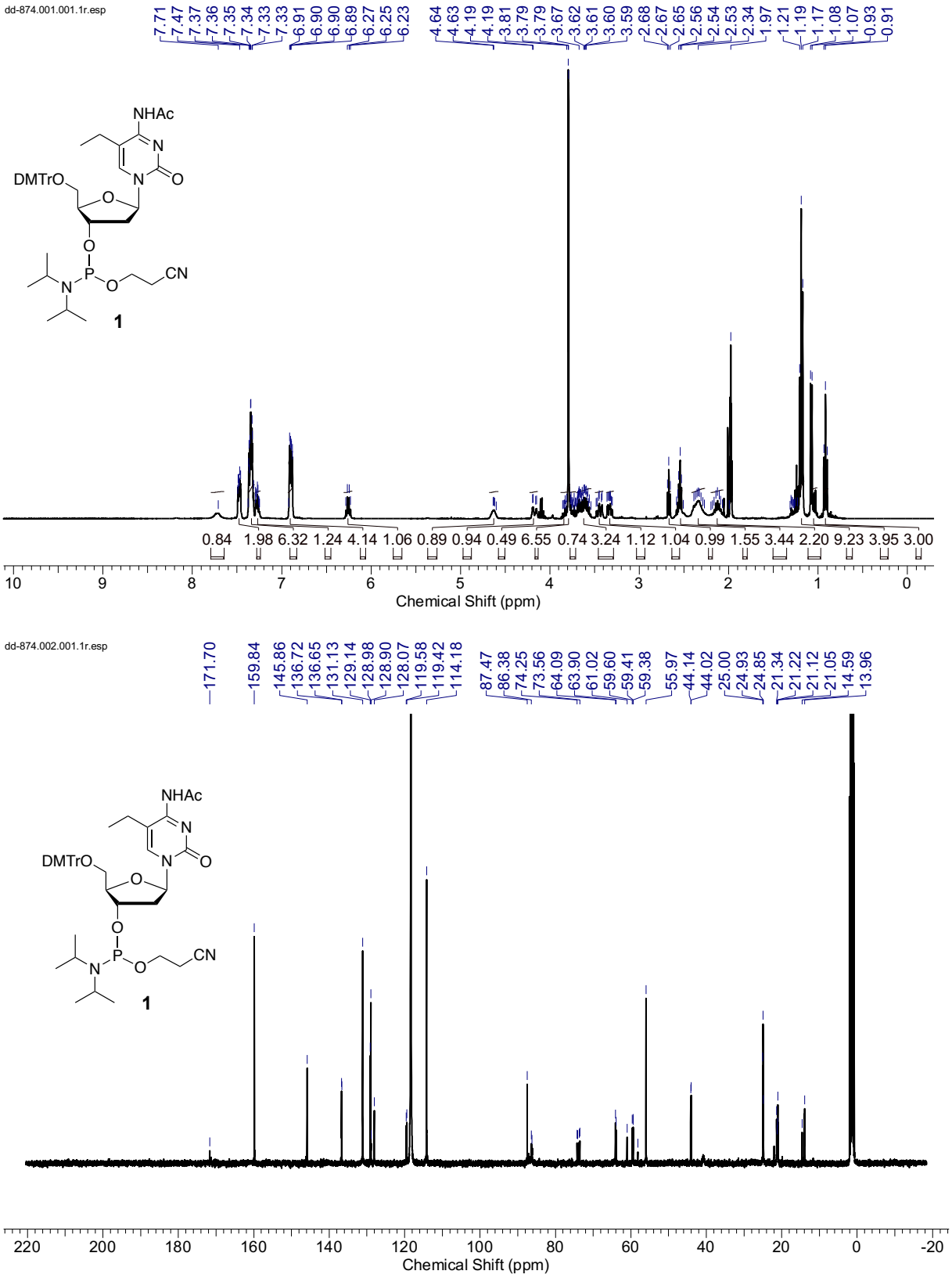
Supplementary Figure S14. ^{1}H and ^{13}C NMR spectra of the indicated compound in CD_3OD.



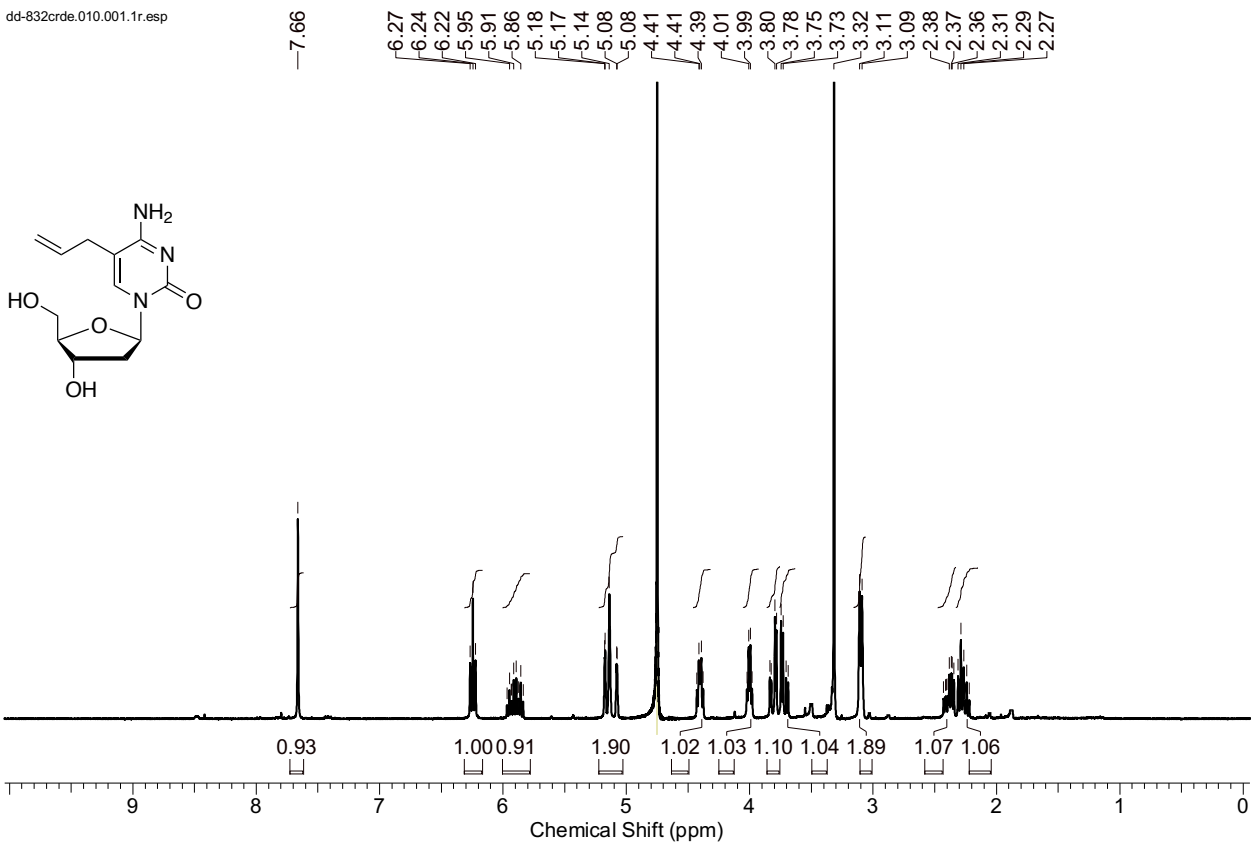
Supplementary Figure S15. ^{1}H and ^{13}C NMR spectra of the indicated compound in $CDCl_3$.



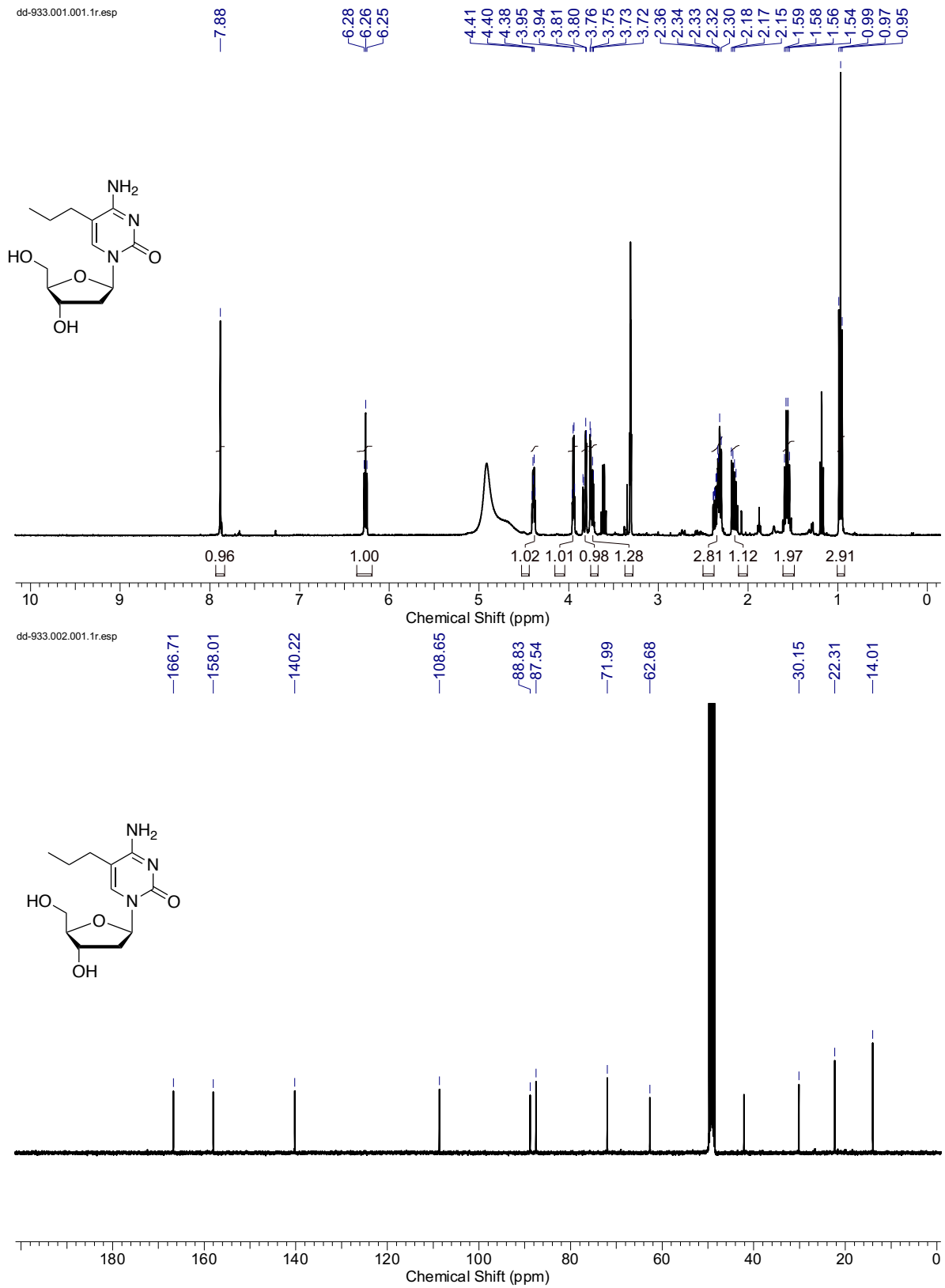
Supplementary Figure S16. ^{1}H and ^{13}C NMR spectra of the indicated compound in $CDCl_3$.



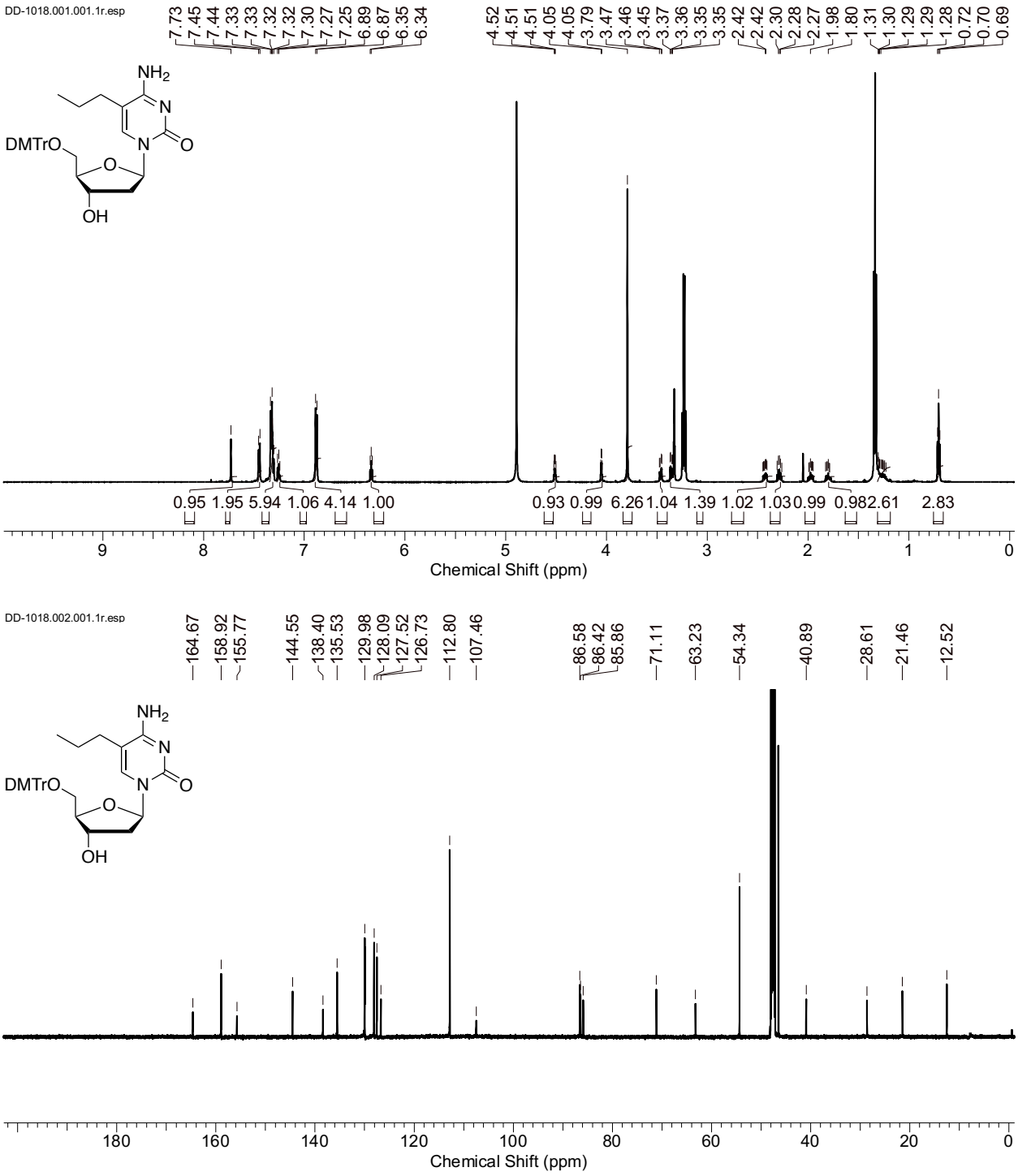
Supplementary Figure S17. ^{1}H and ^{13}C NMR spectra of **1** in Acetonitrile-d_3.



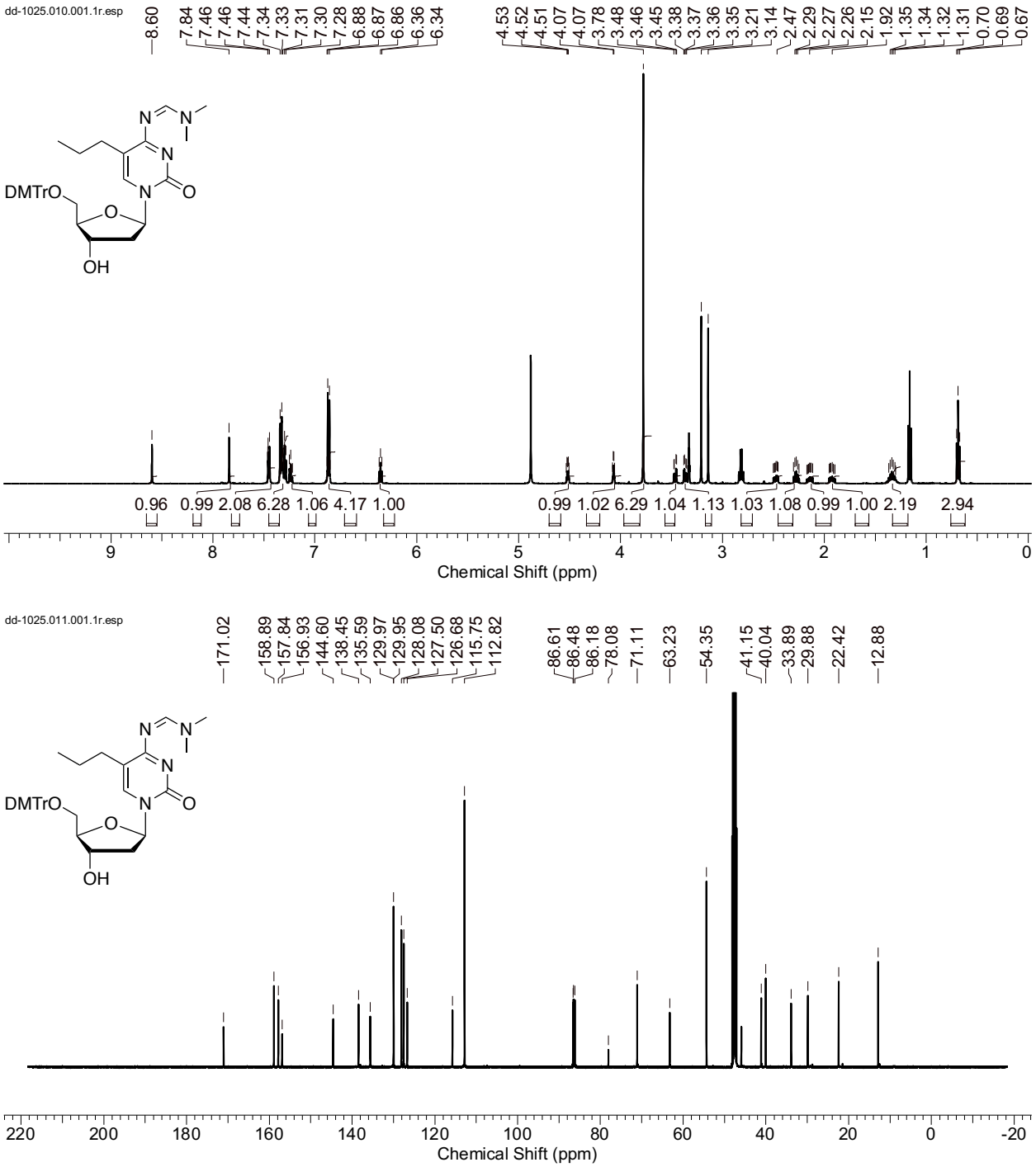
Supplementary Figure S18. ^{1}H and ^{13}C NMR spectra of the indicated compound in D_2O.



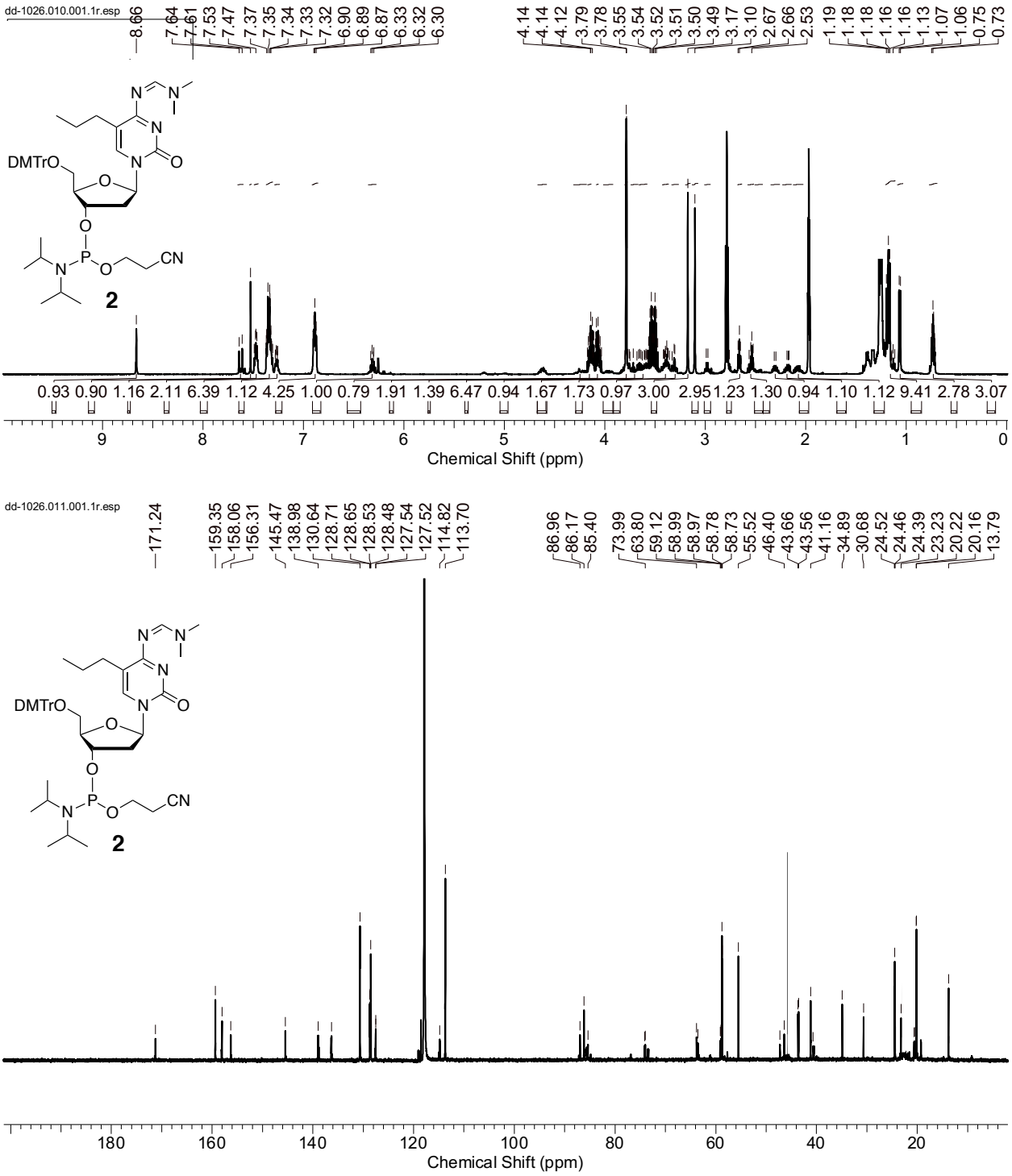
Supplementary Figure S19. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



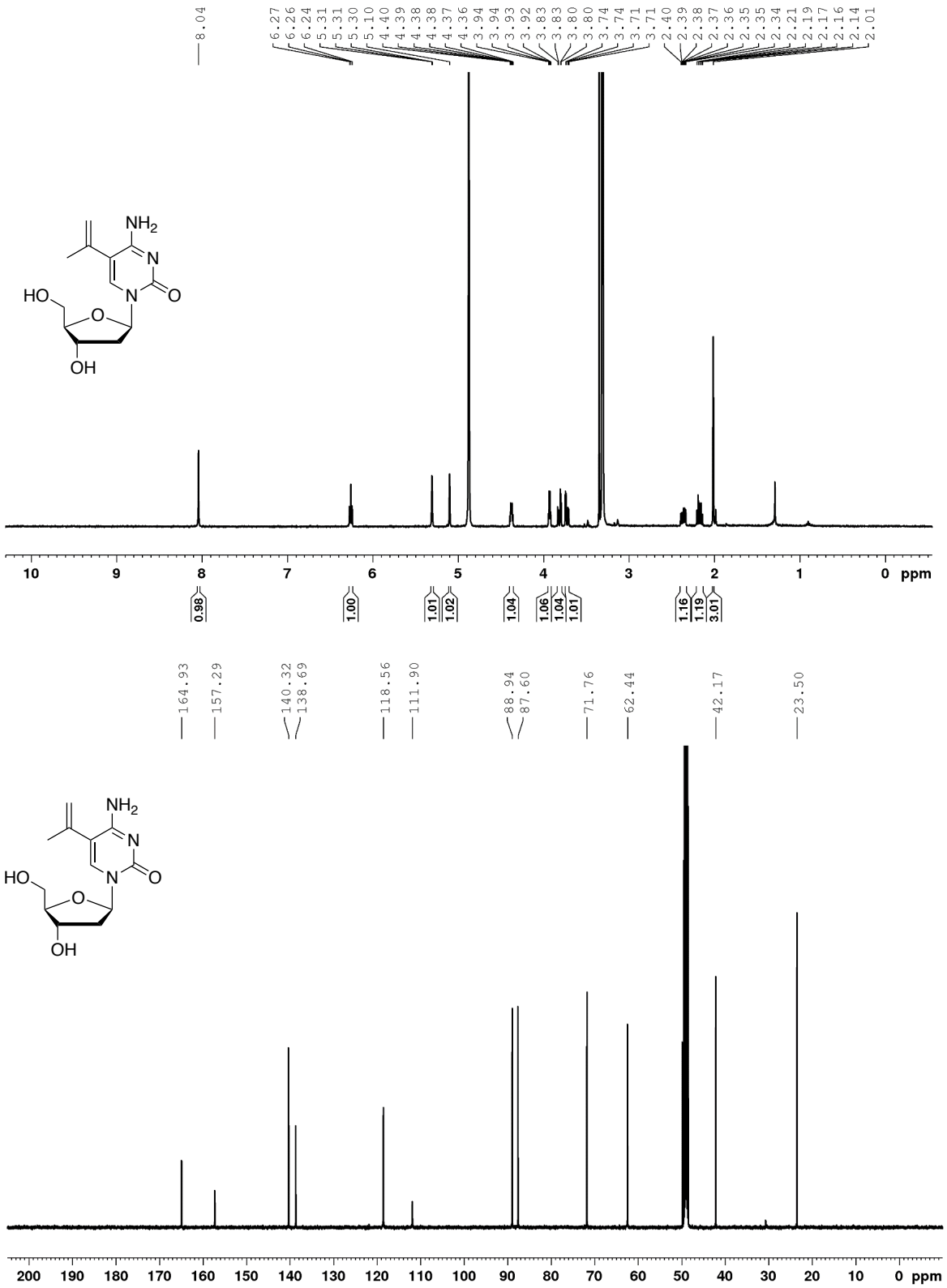
Supplementary Figure S20. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



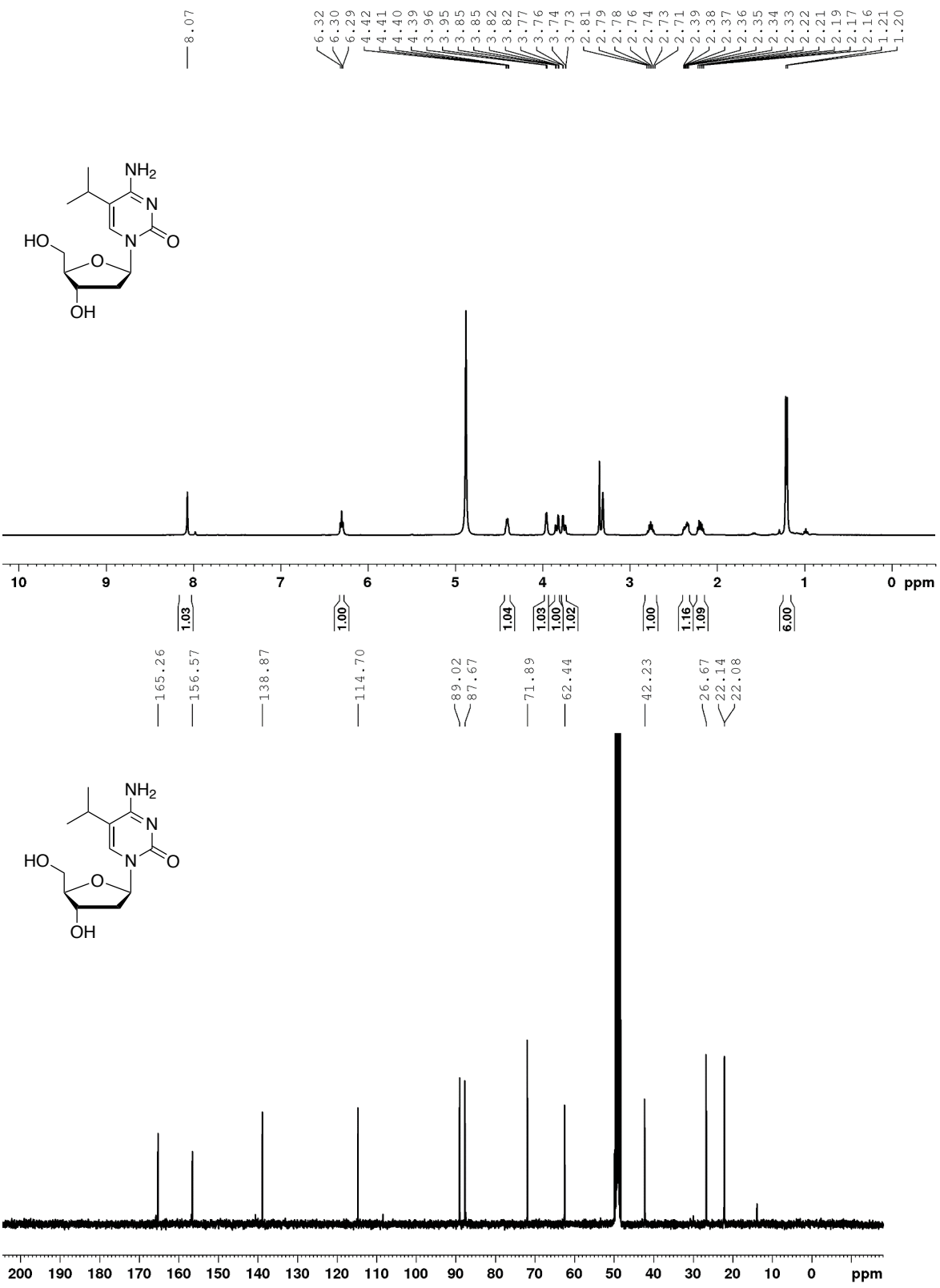
Supplementary Figure S21. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



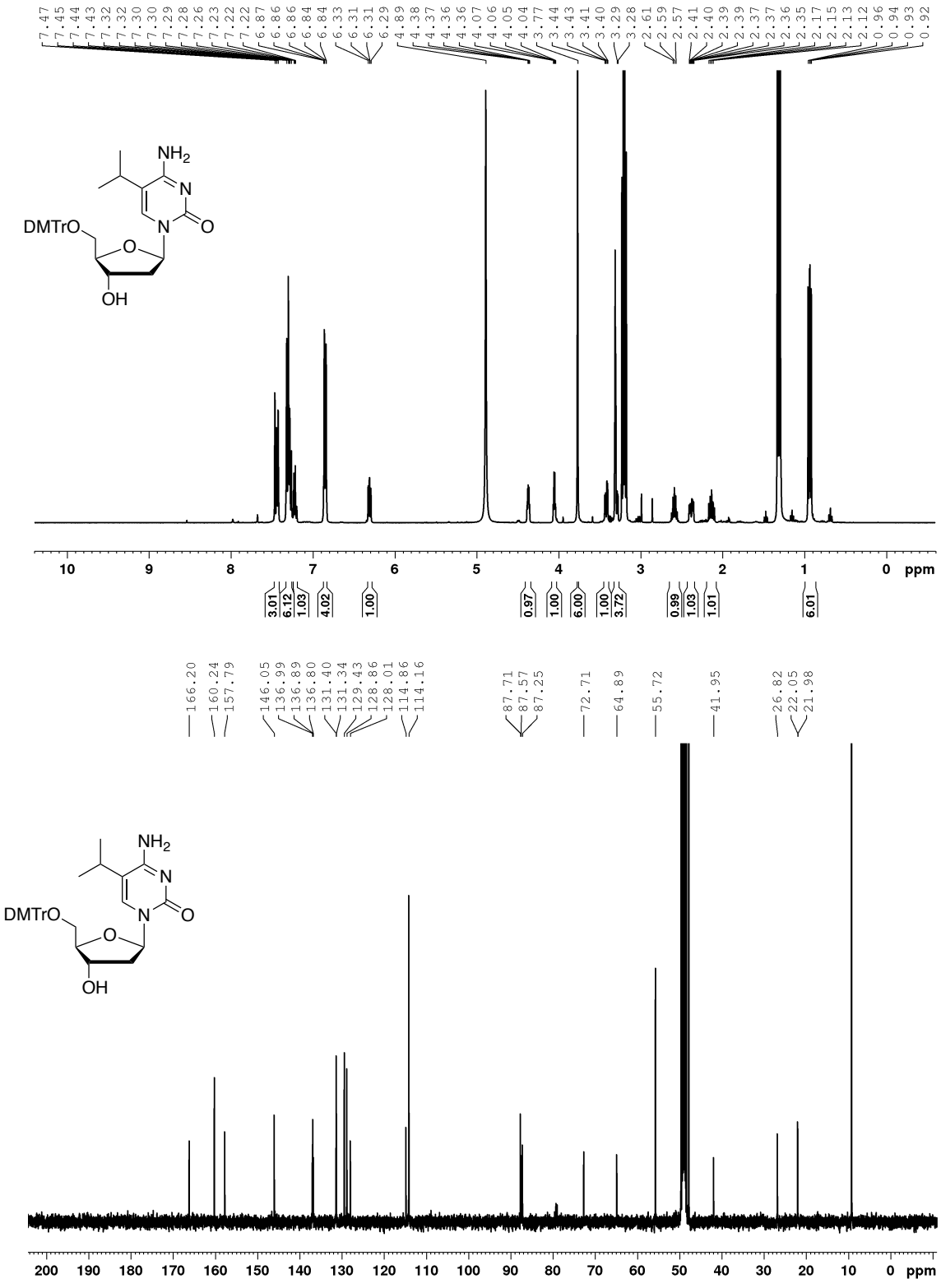
Supplementary Figure S22. ^{1}H and ^{13}C NMR spectra of **2** in acetonitrile-d_3.



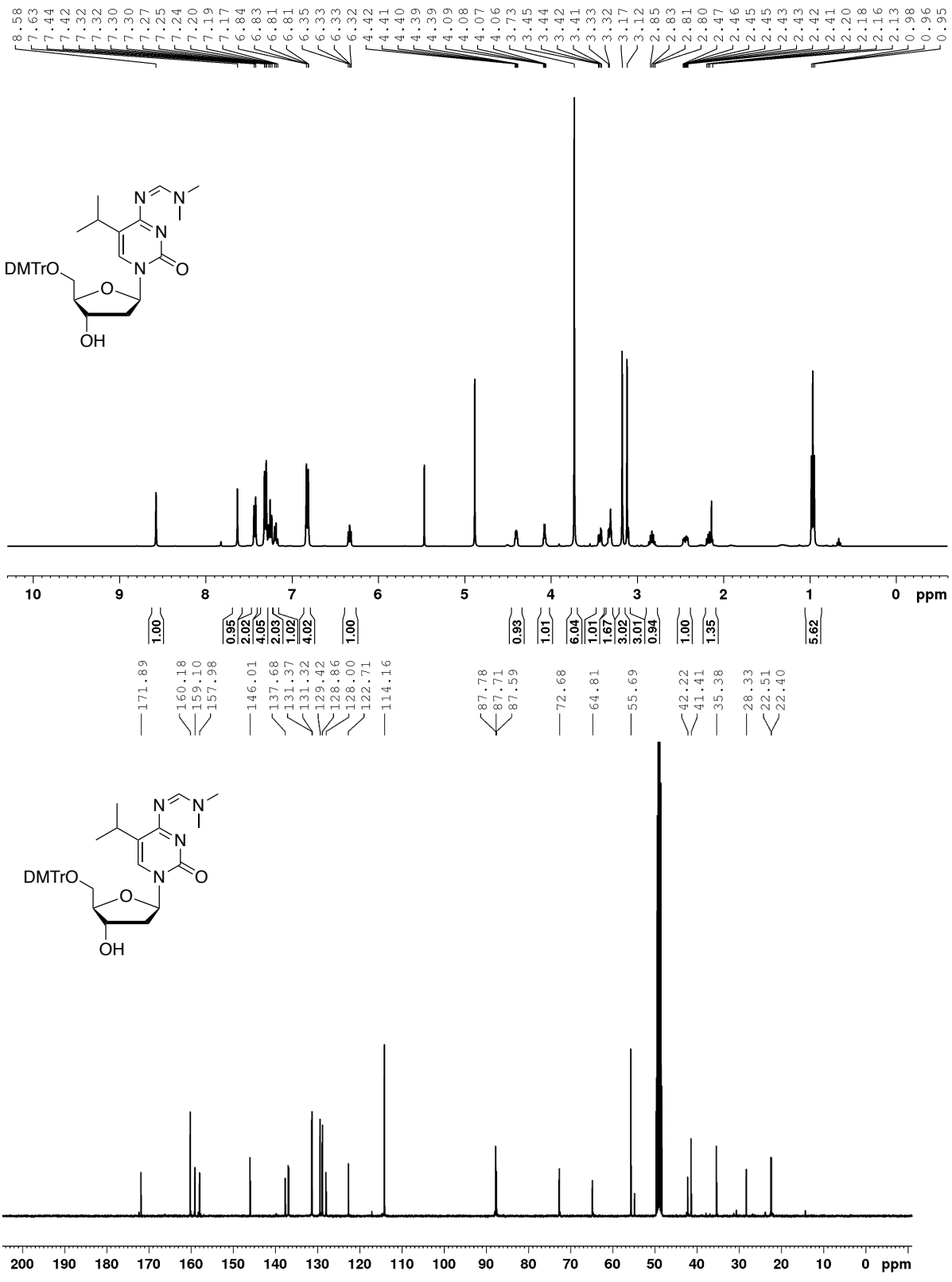
Supplementary Figure S23. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



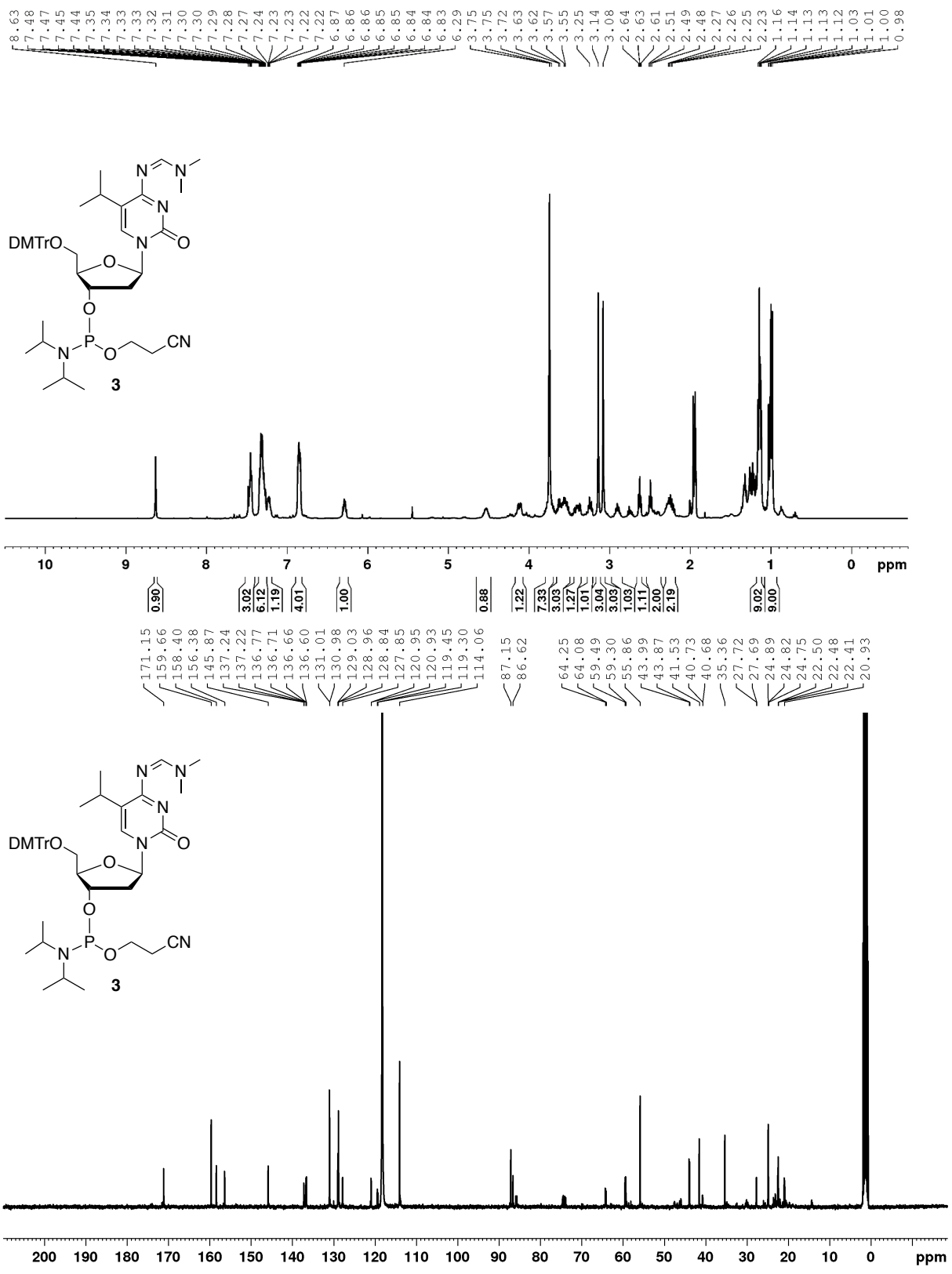
Supplementary Figure S24. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



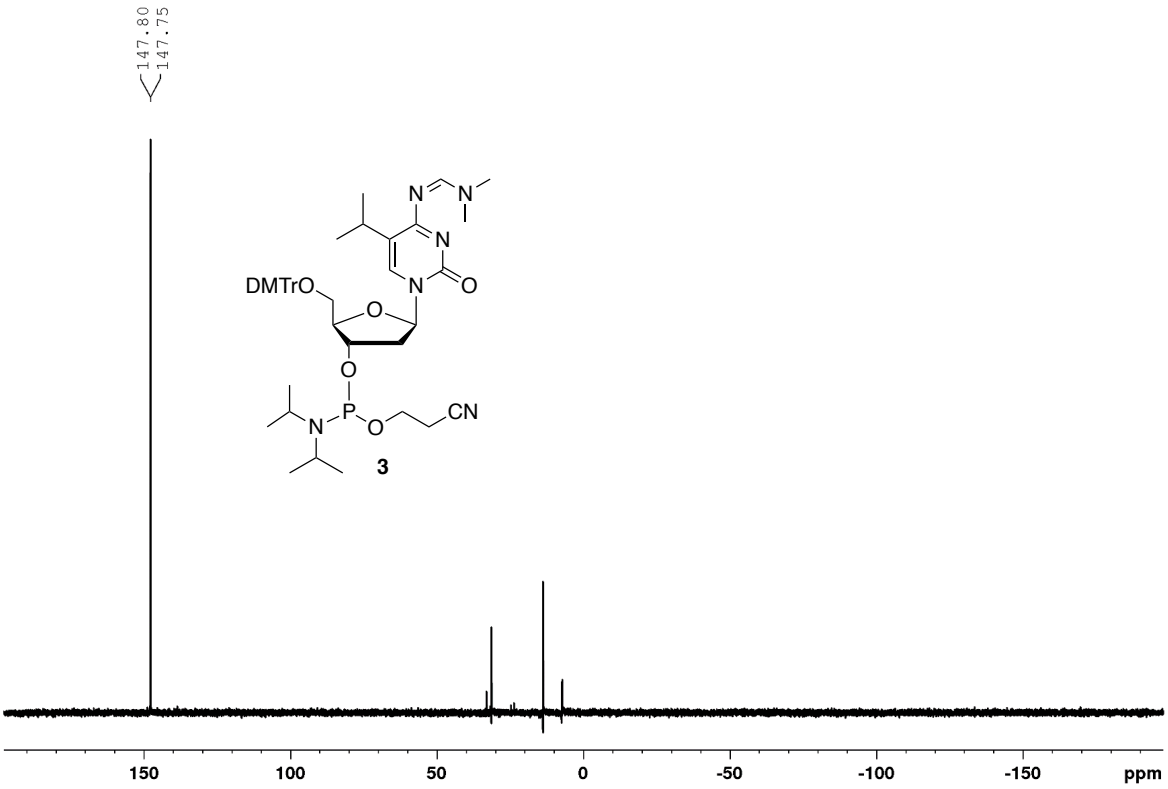
Supplementary Figure S25. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



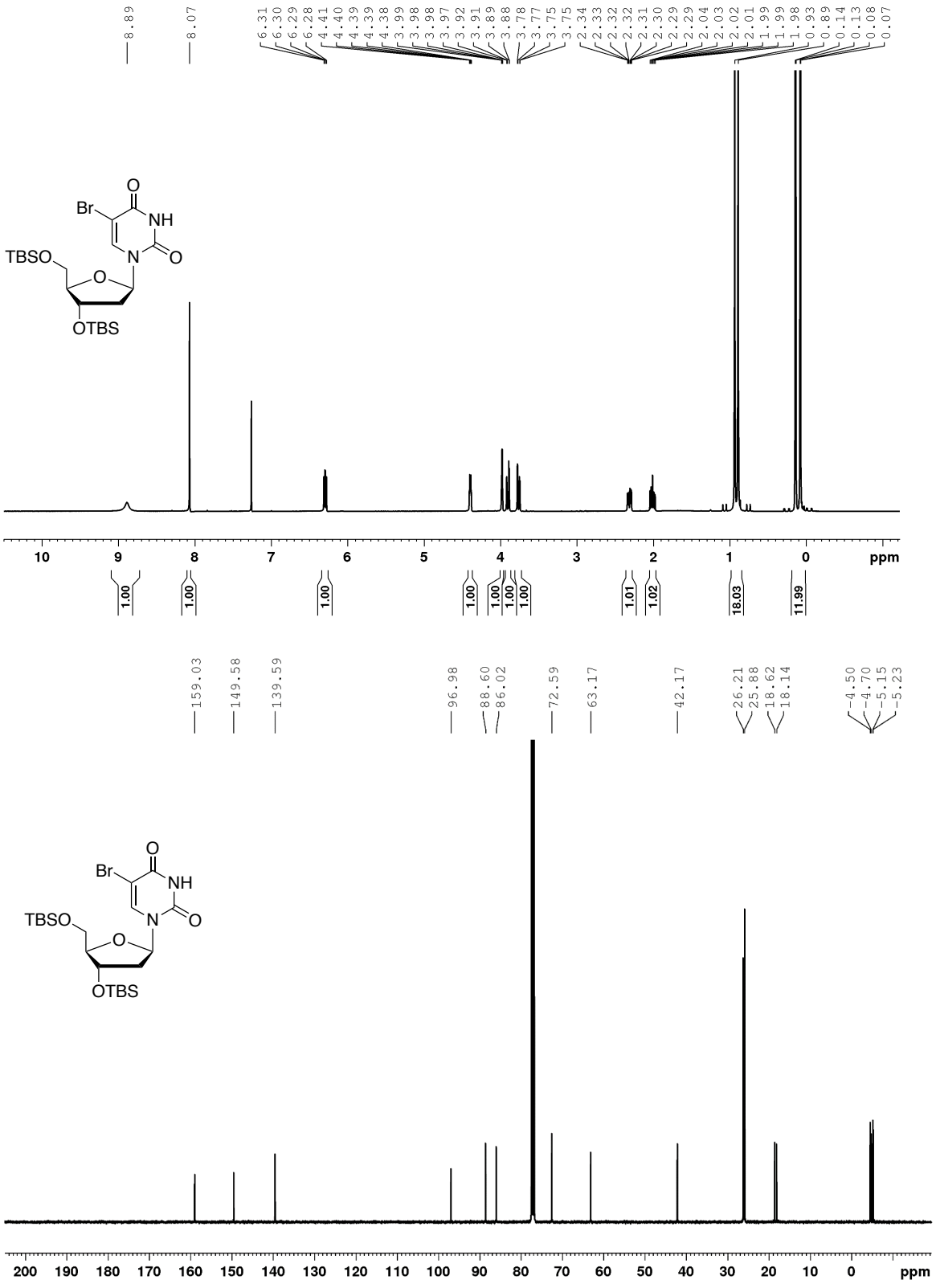
Supplementary Figure S26. ^{1}H and ^{13}C NMR spectra of the indicated compound in methanol-d_4.



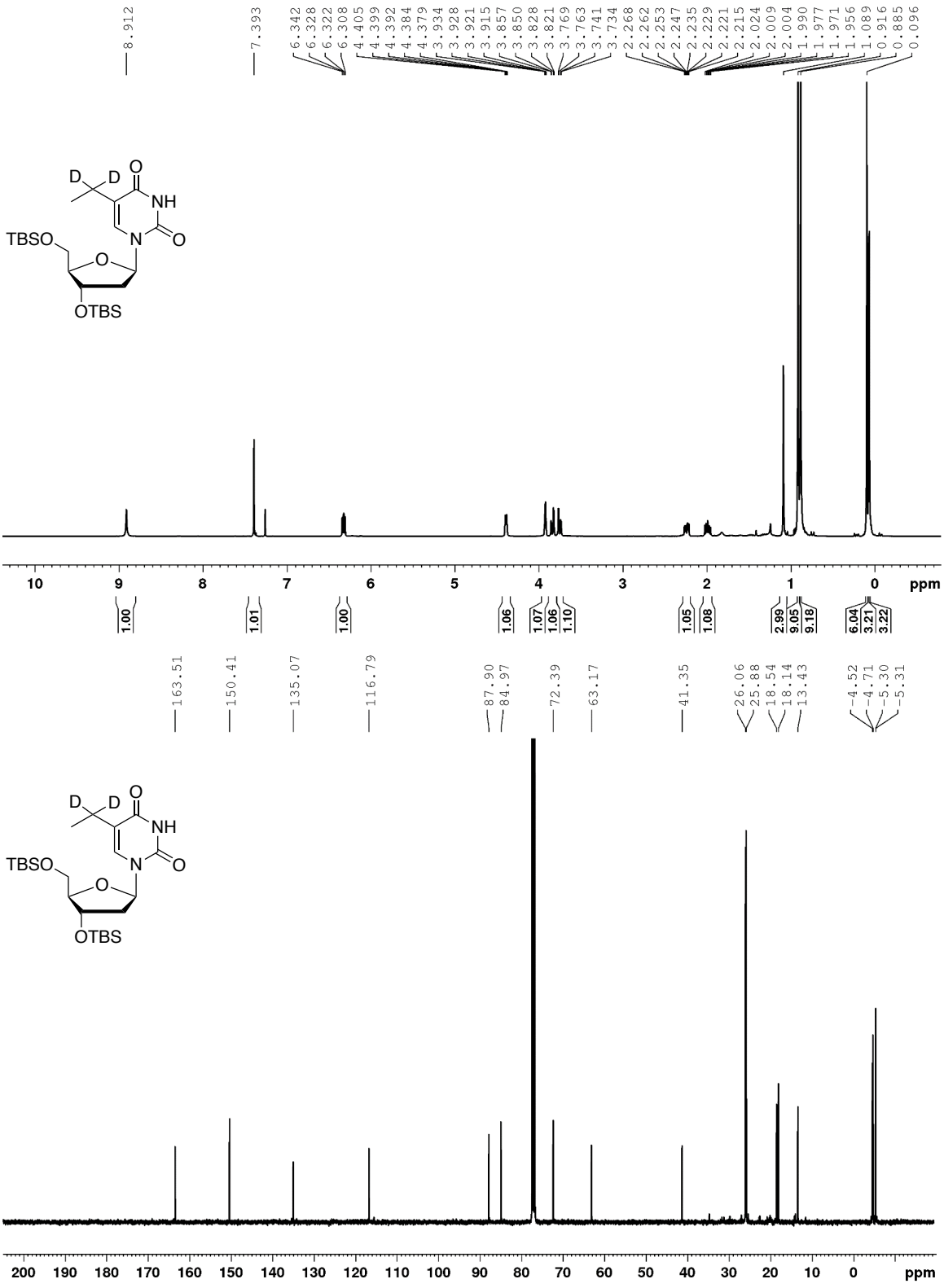
Supplementary Figure S27. ^{1}H and ^{13}C NMR spectra of compound **3** in acetonitrile-d_3.



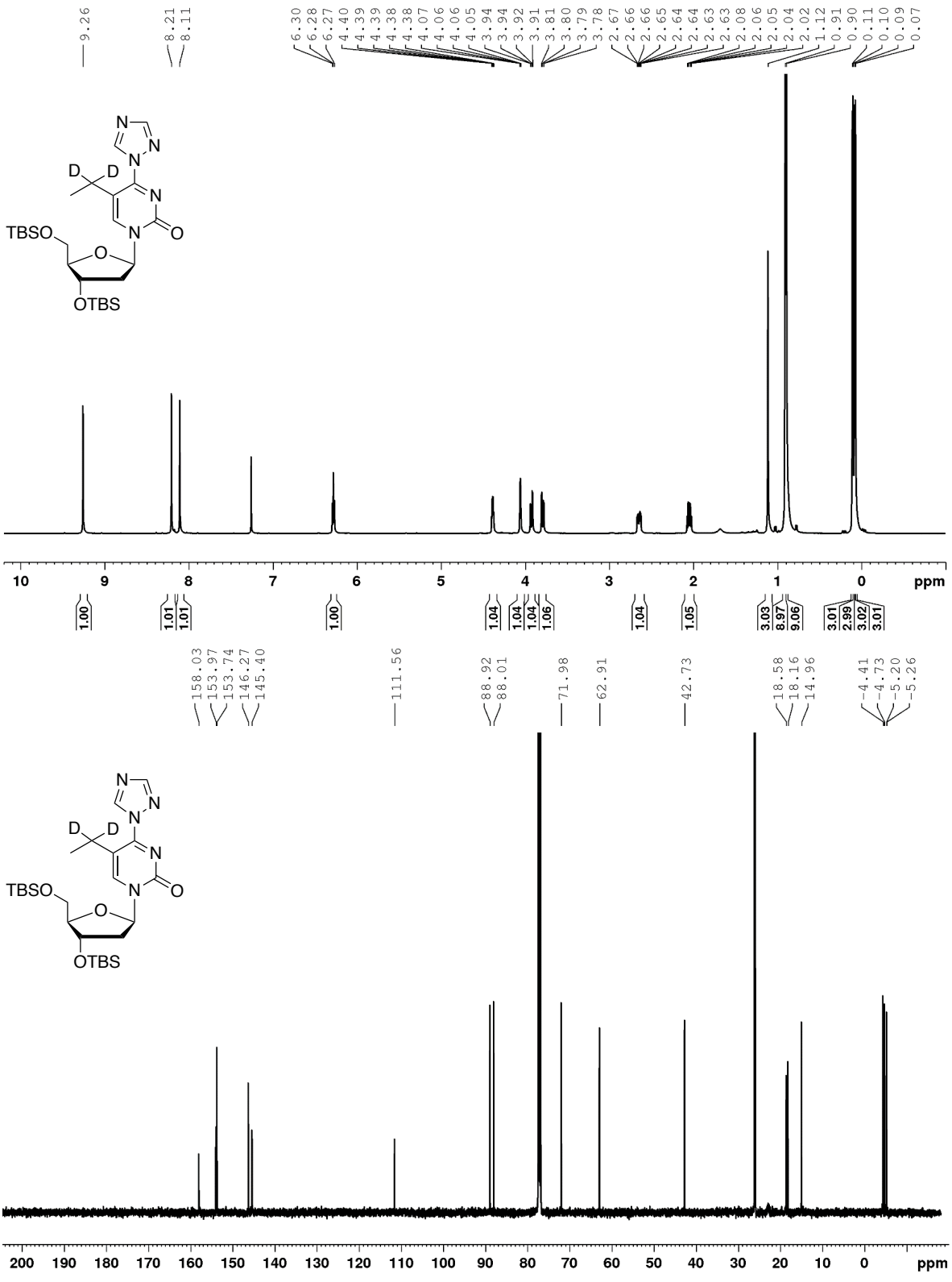
Supplementary Figure S28. ^{31}P NMR spectrum of compound **3** in acetonitrile-d_3.



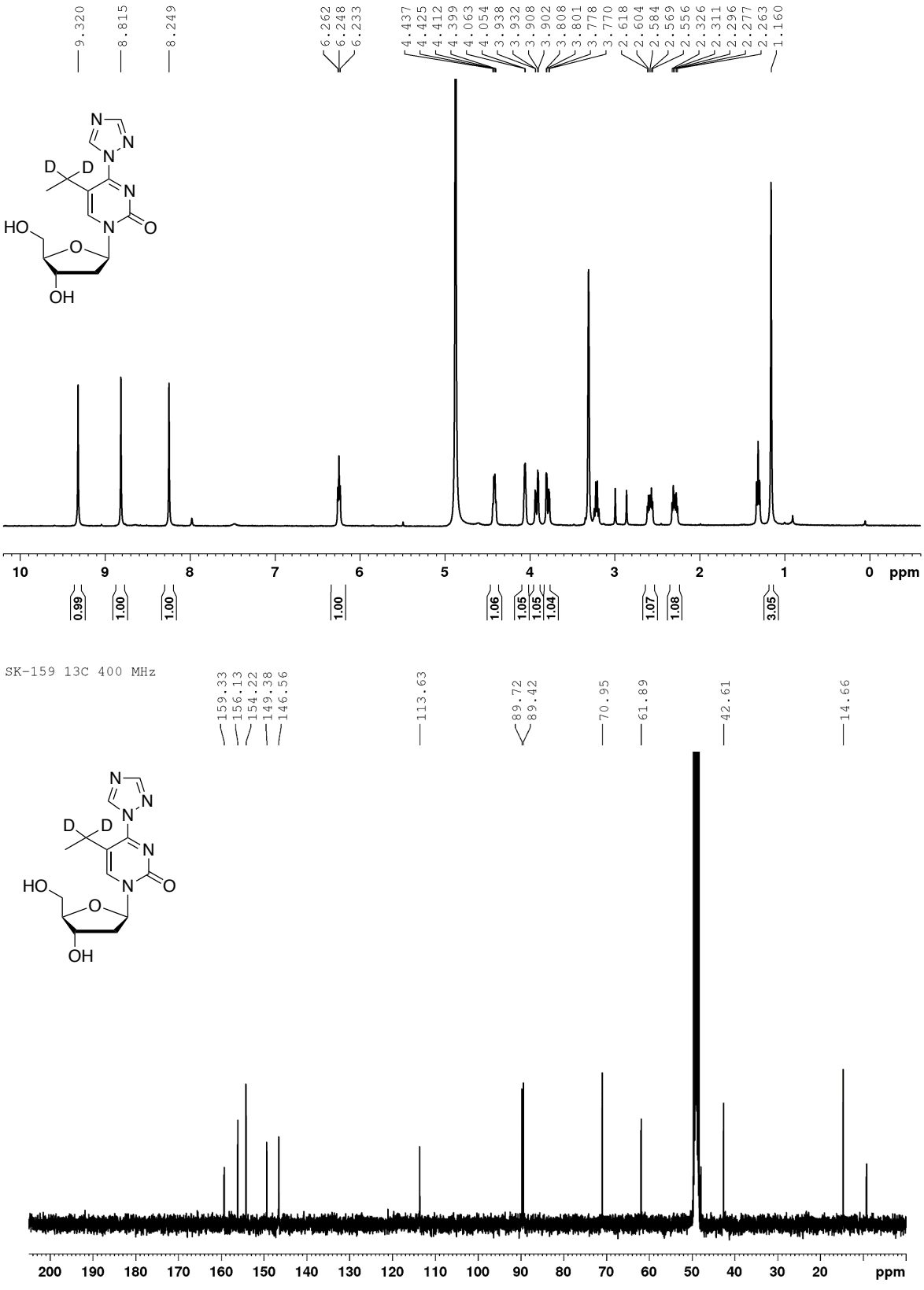
Supplementary Figure S29. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.



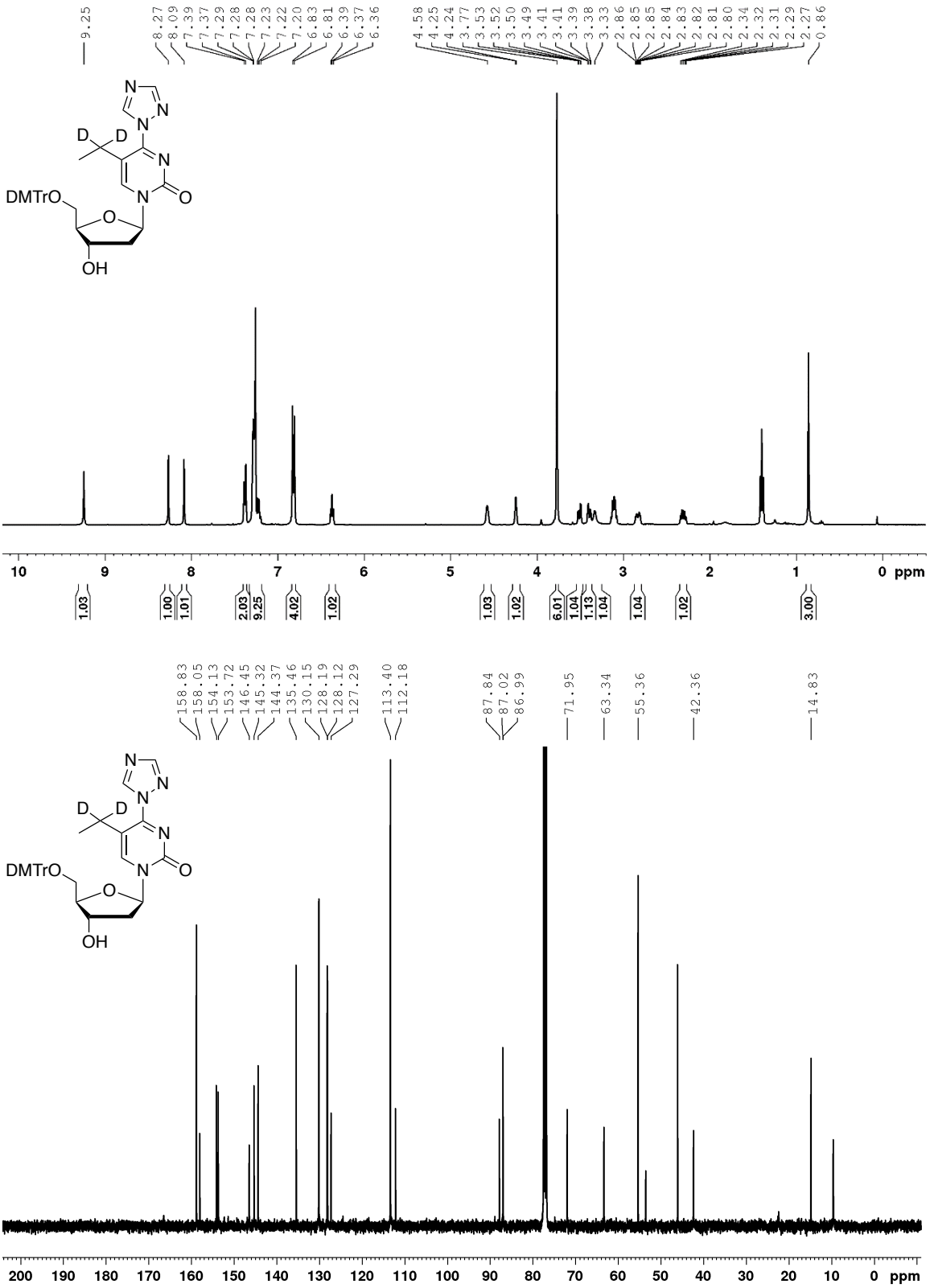
Supplementary Figure S30. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.



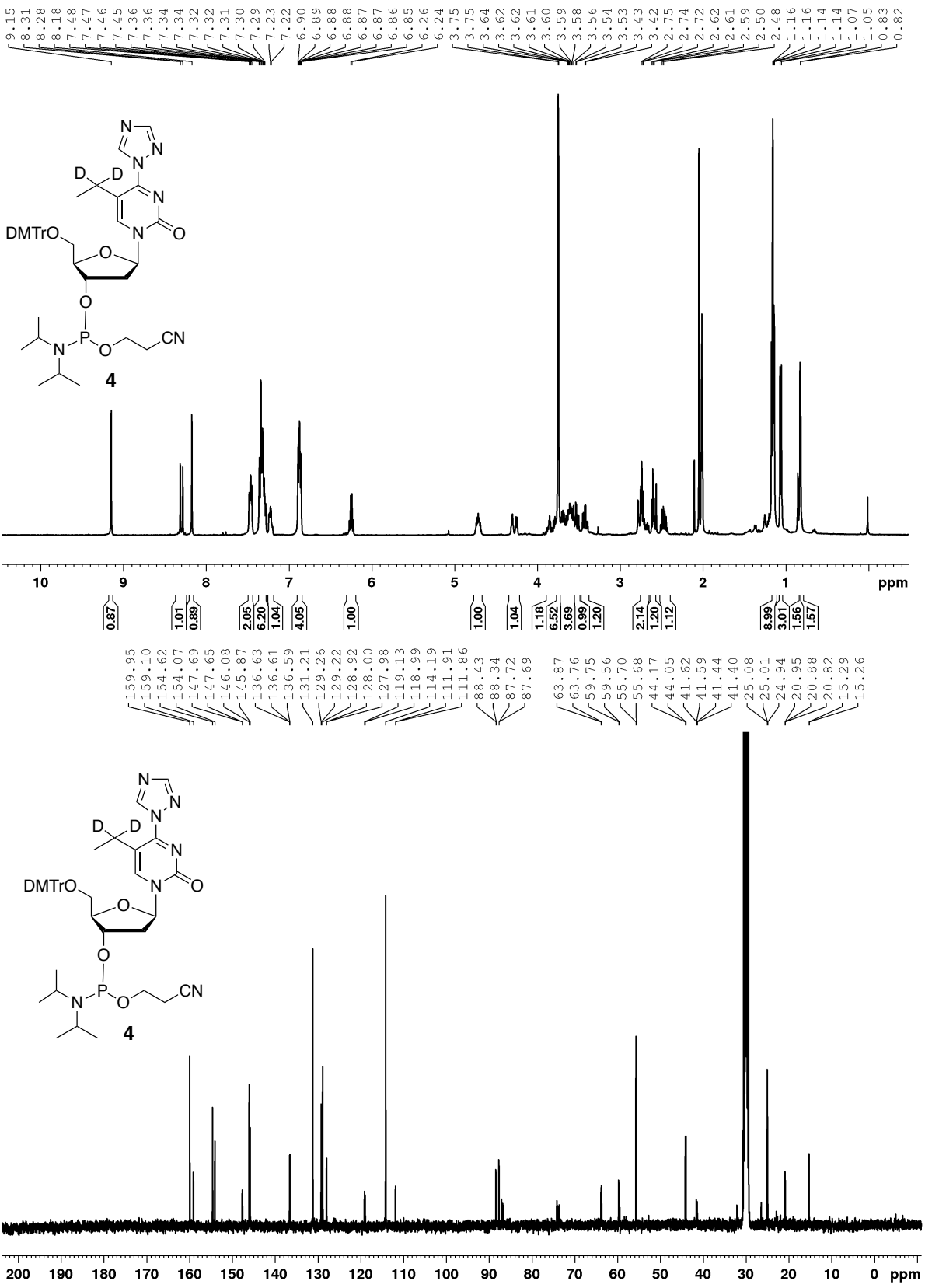
Supplementary Figure S31. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.



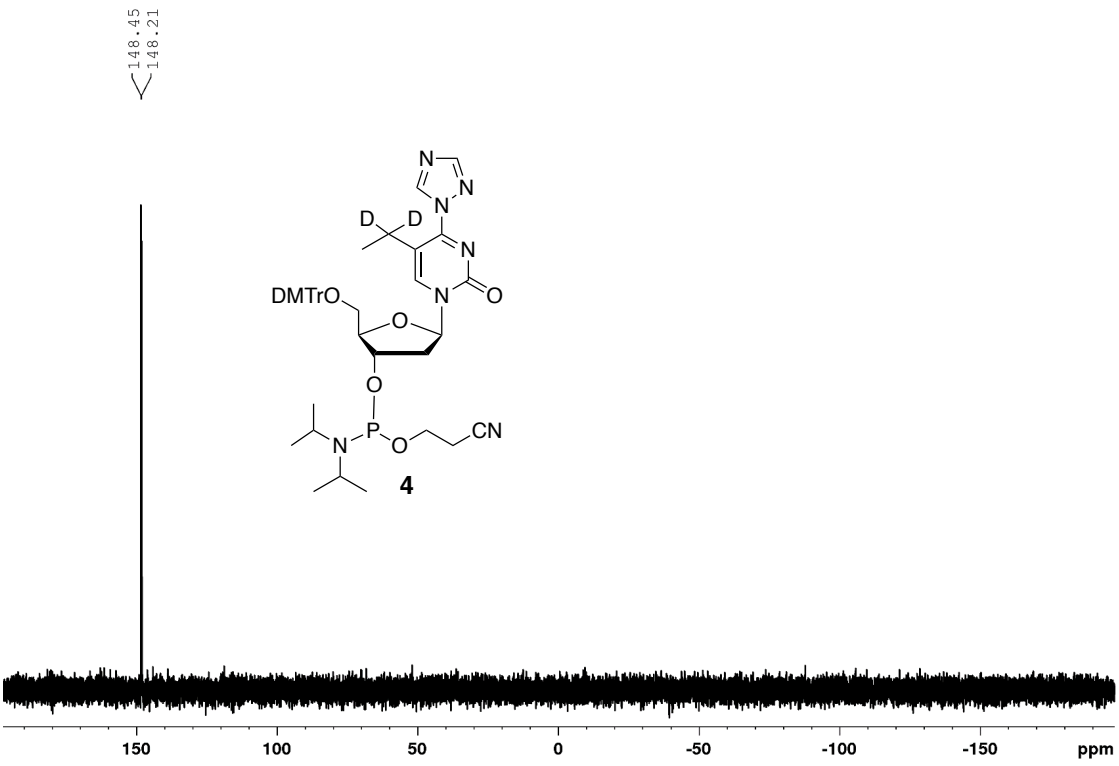
Supplementary Figure S32. ^{1}H and ^{13}C NMR spectra of the title compound in methanol-d_3.



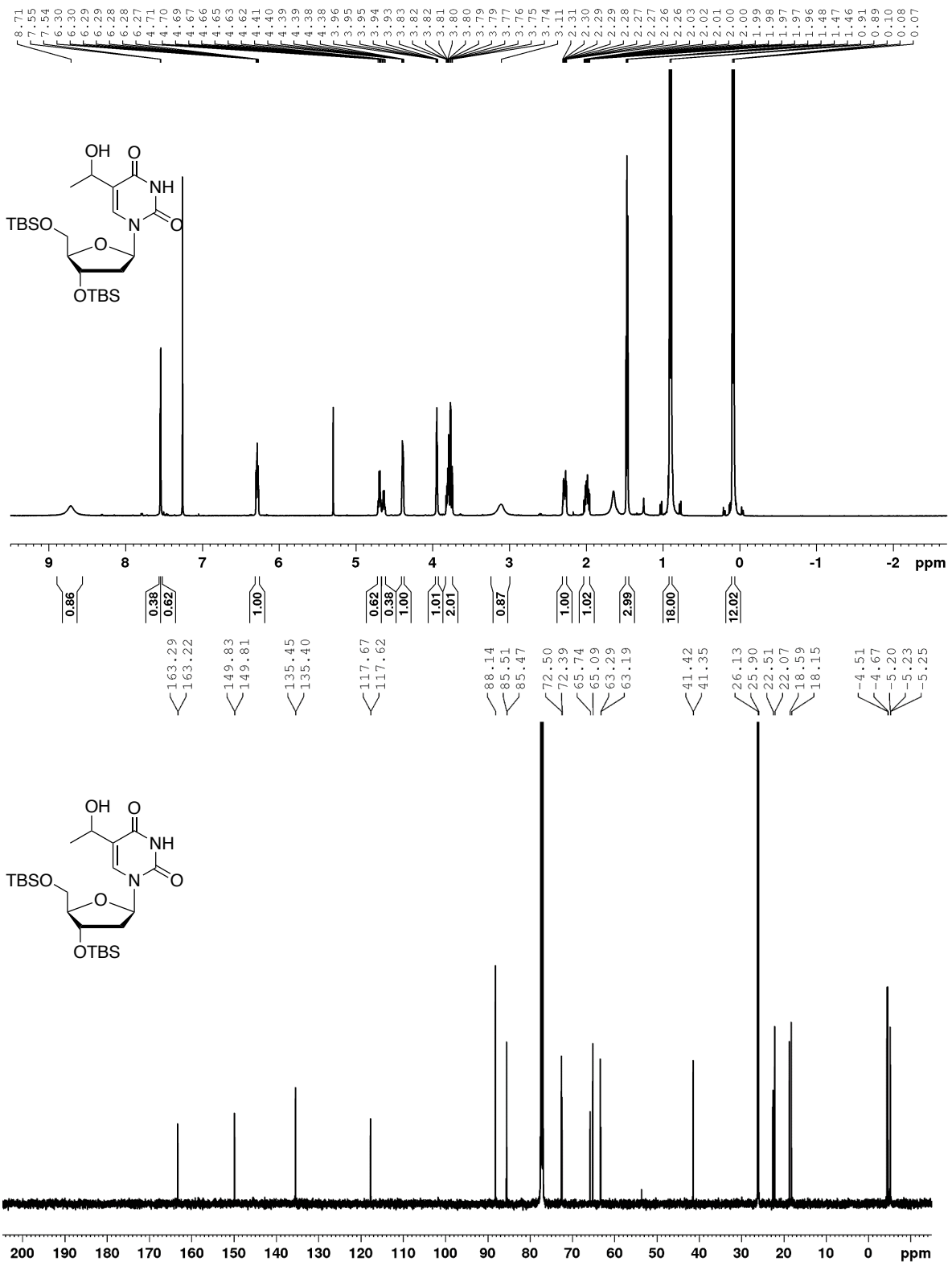
Supplementary Figure S33. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.



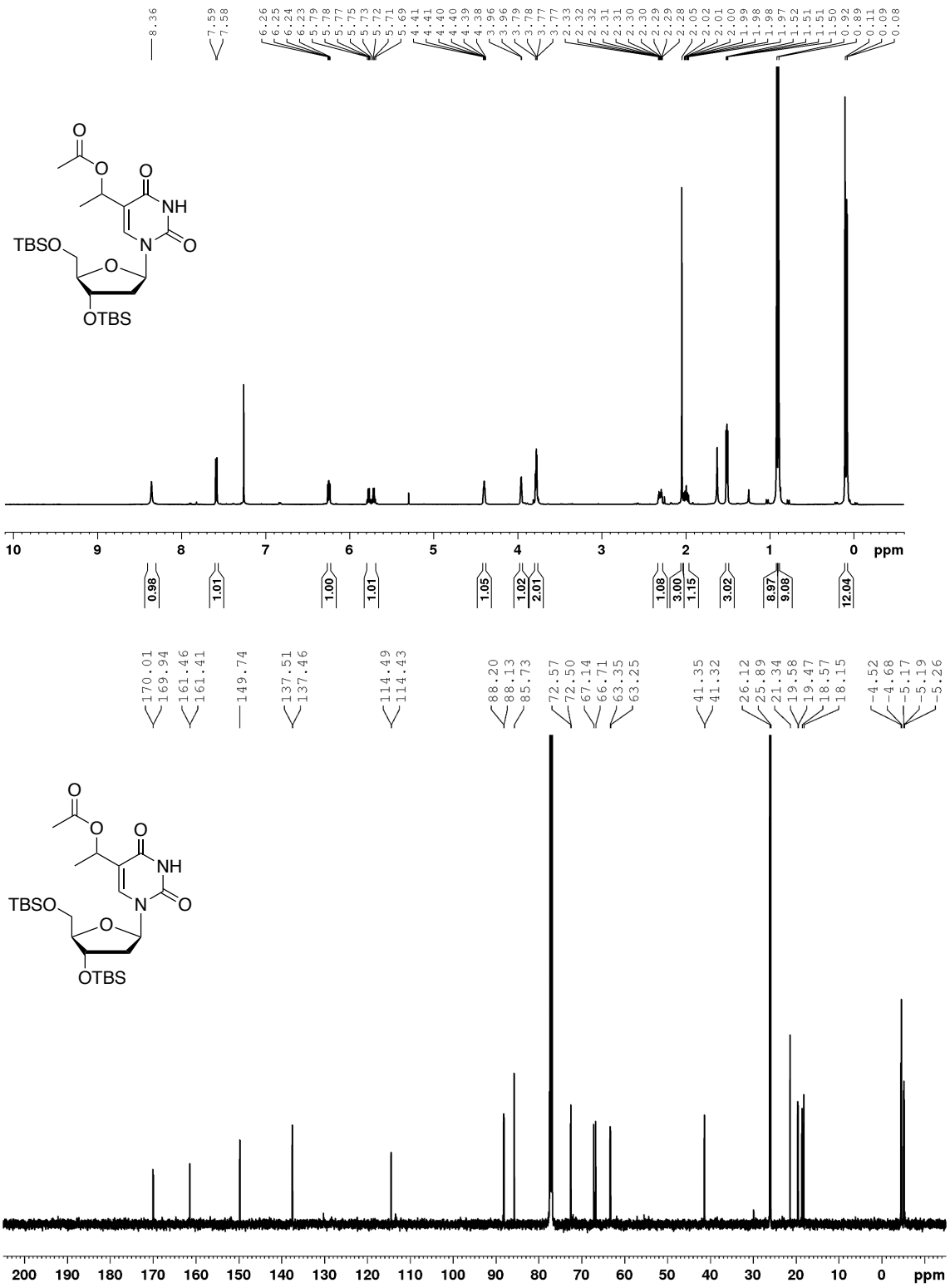
Supplementary Figure S34. ^{1}H and ^{13}C NMR spectra of compound **4** in acetone-d_6.



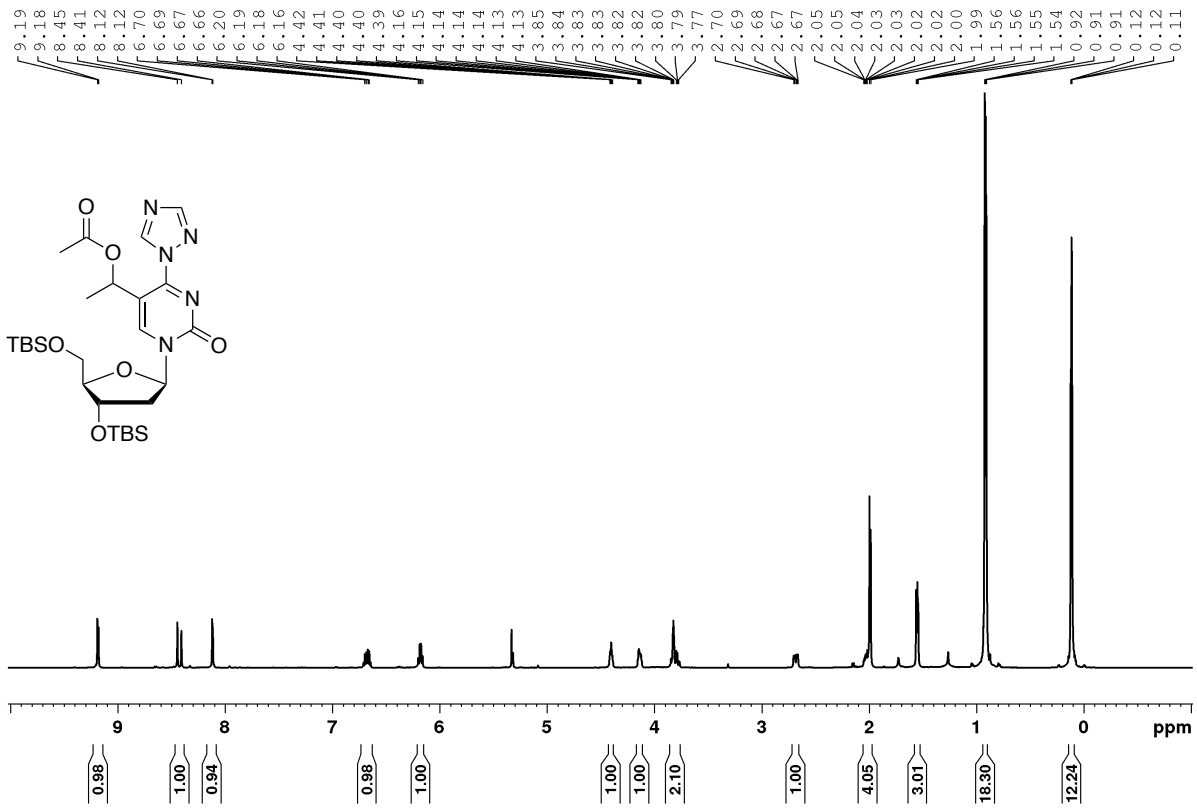
Supplementary Figure S35. ^{31}P NMR spectrum of compound **4** in acetone-d_6.

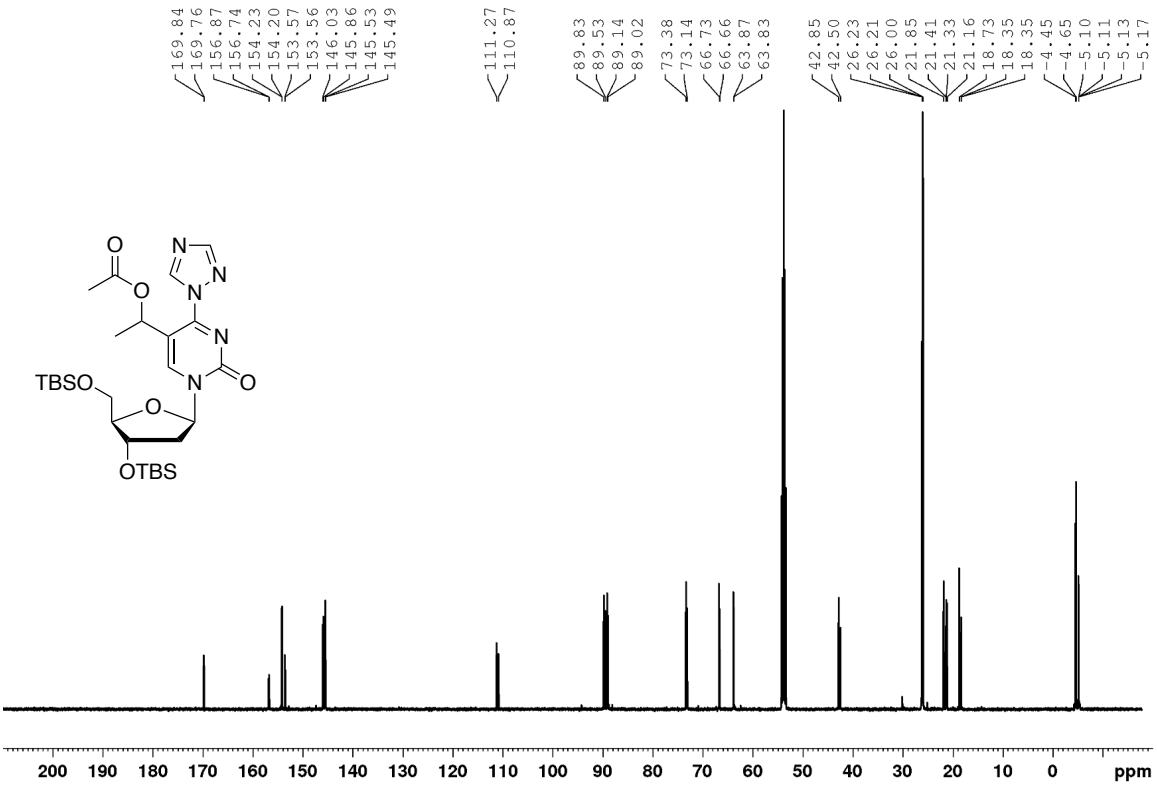


Supplementary Figure S36. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.

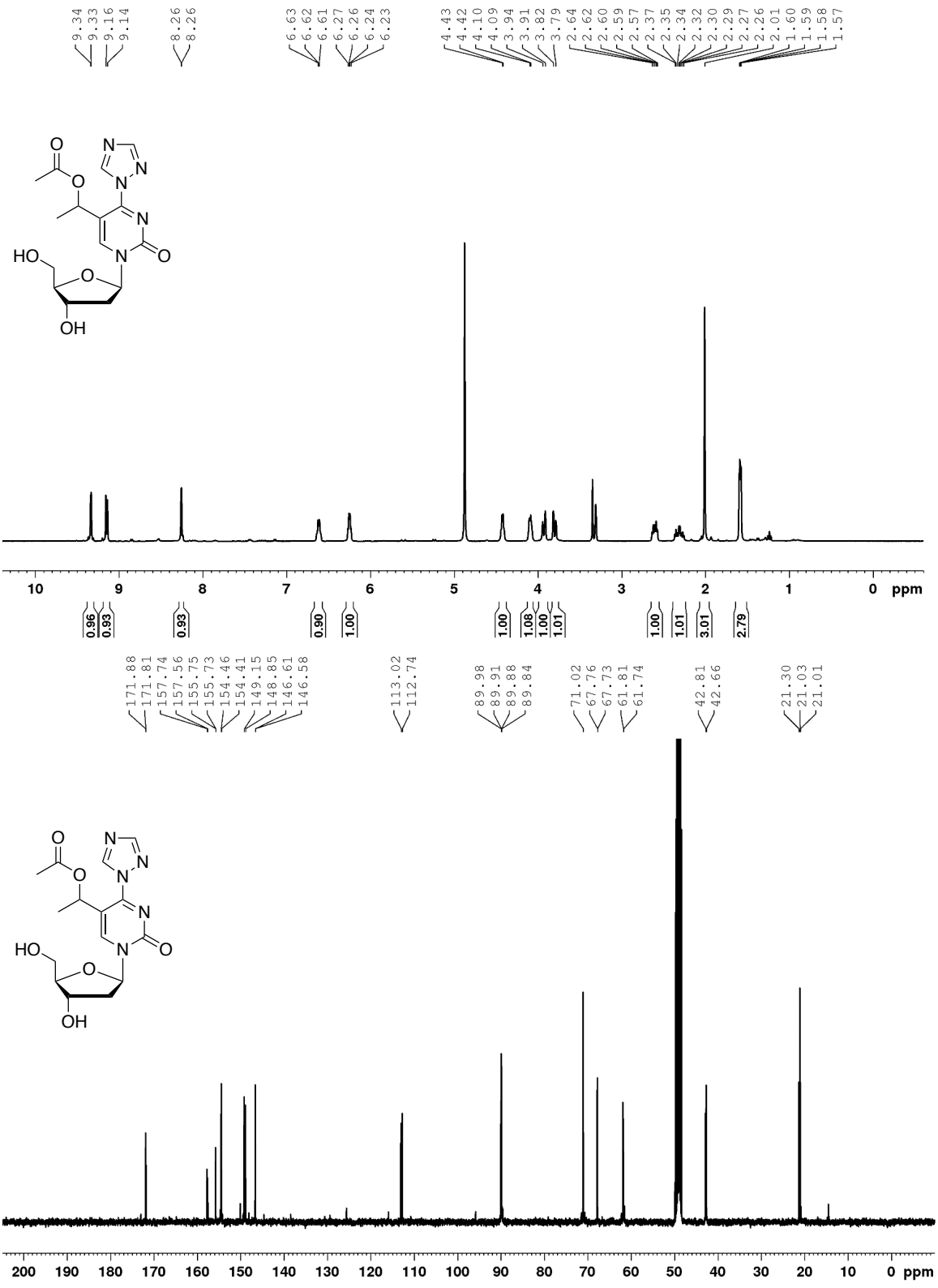


Supplementary Figure S37. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.

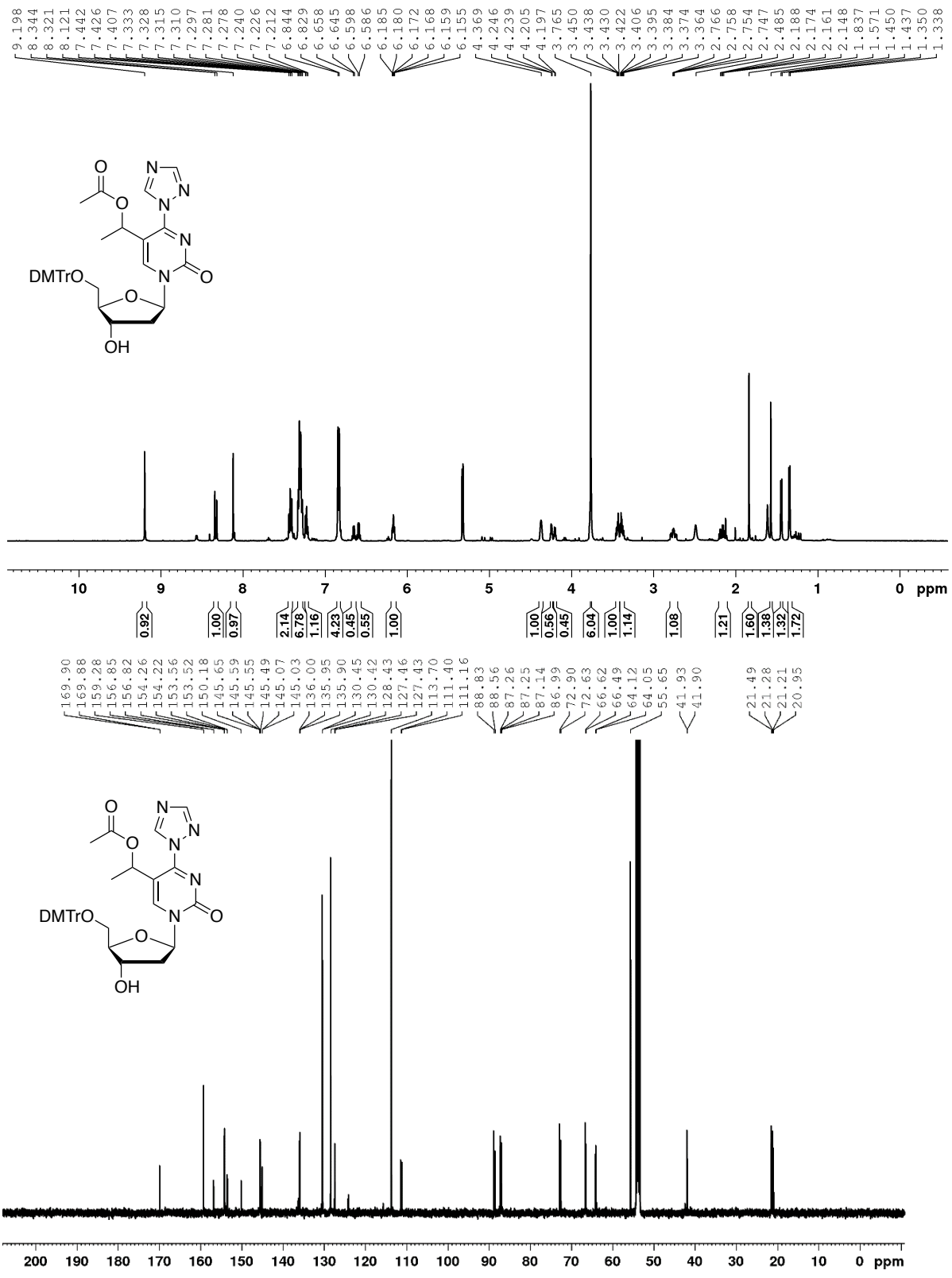




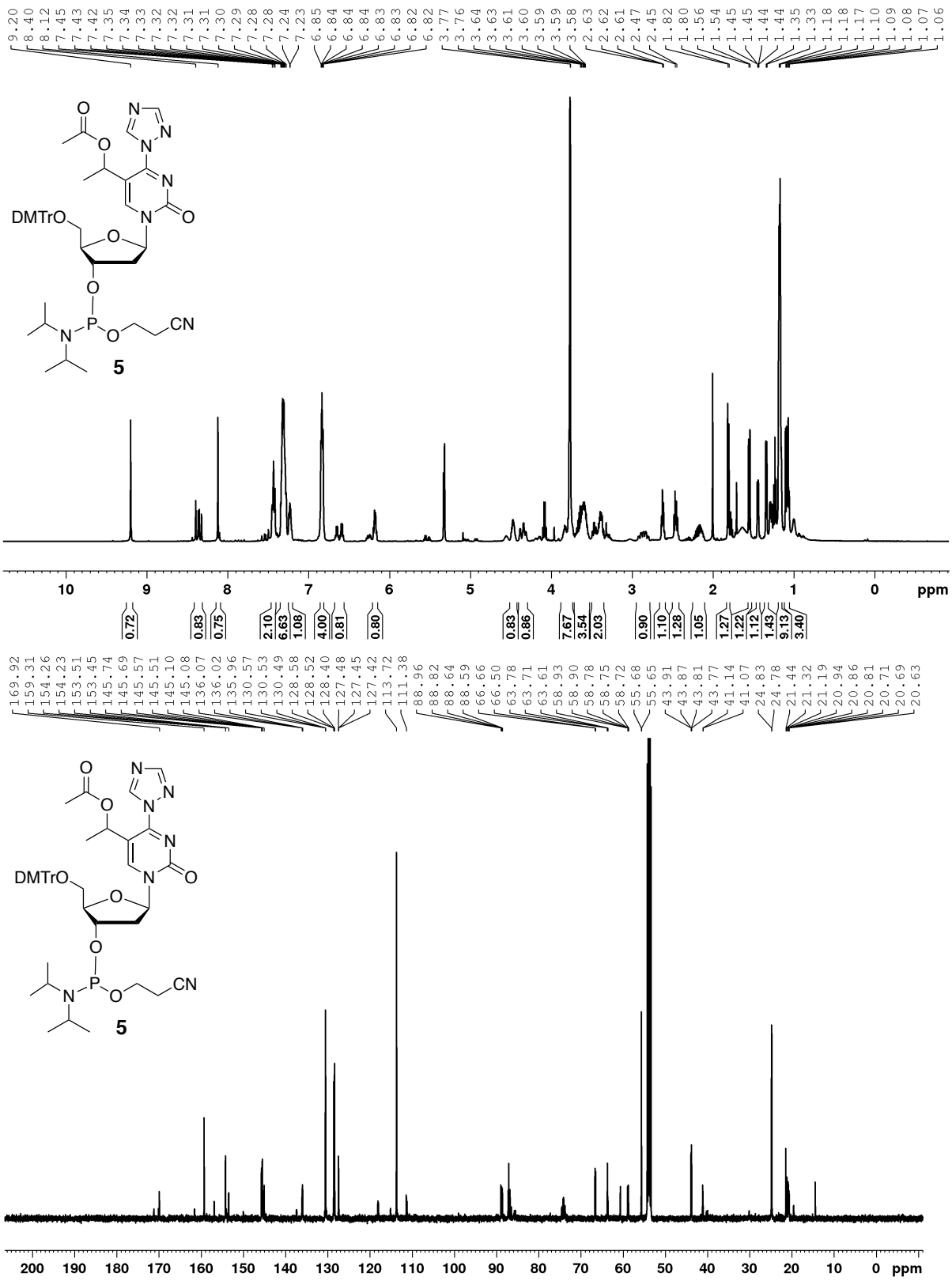
Supplementary Figure S38. ^{1}H and ^{13}C NMR spectra of the title compound in CD_2Cl_2.



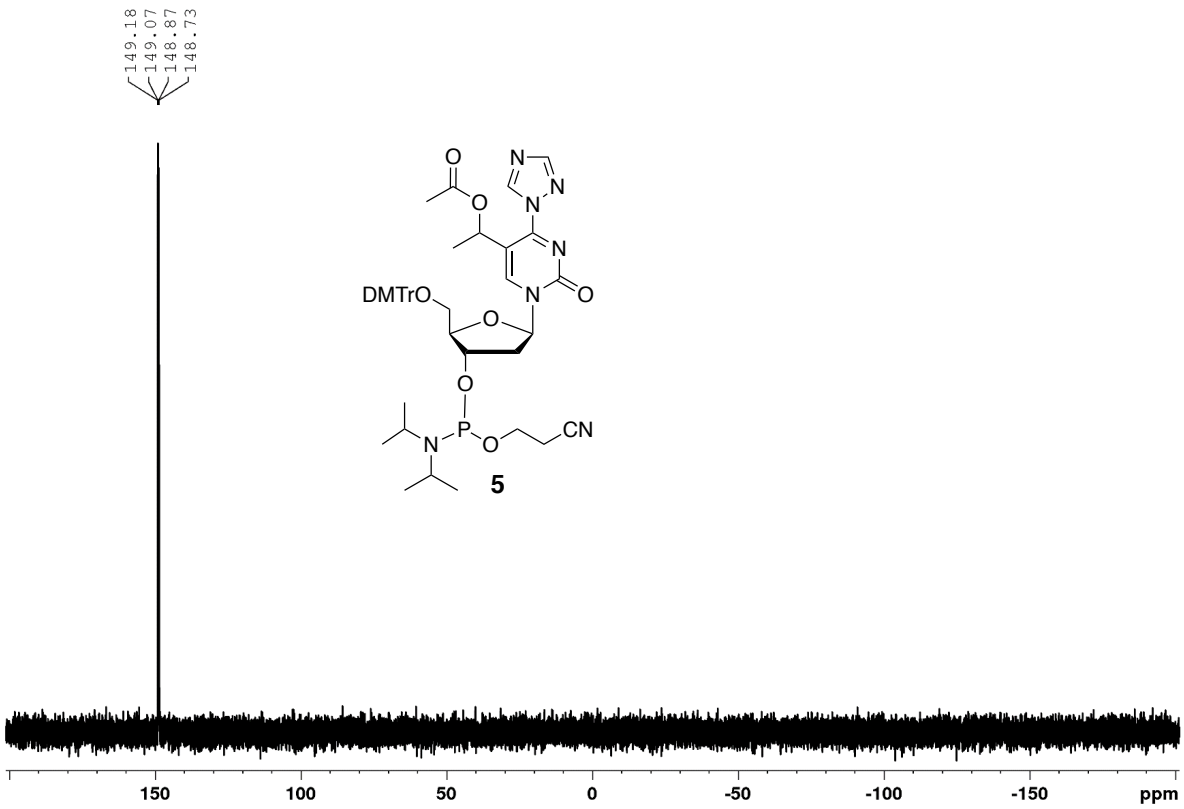
Supplementary Figure S39. ^{1}H and ^{13}C NMR spectra of the title compound in methanol-d_4.



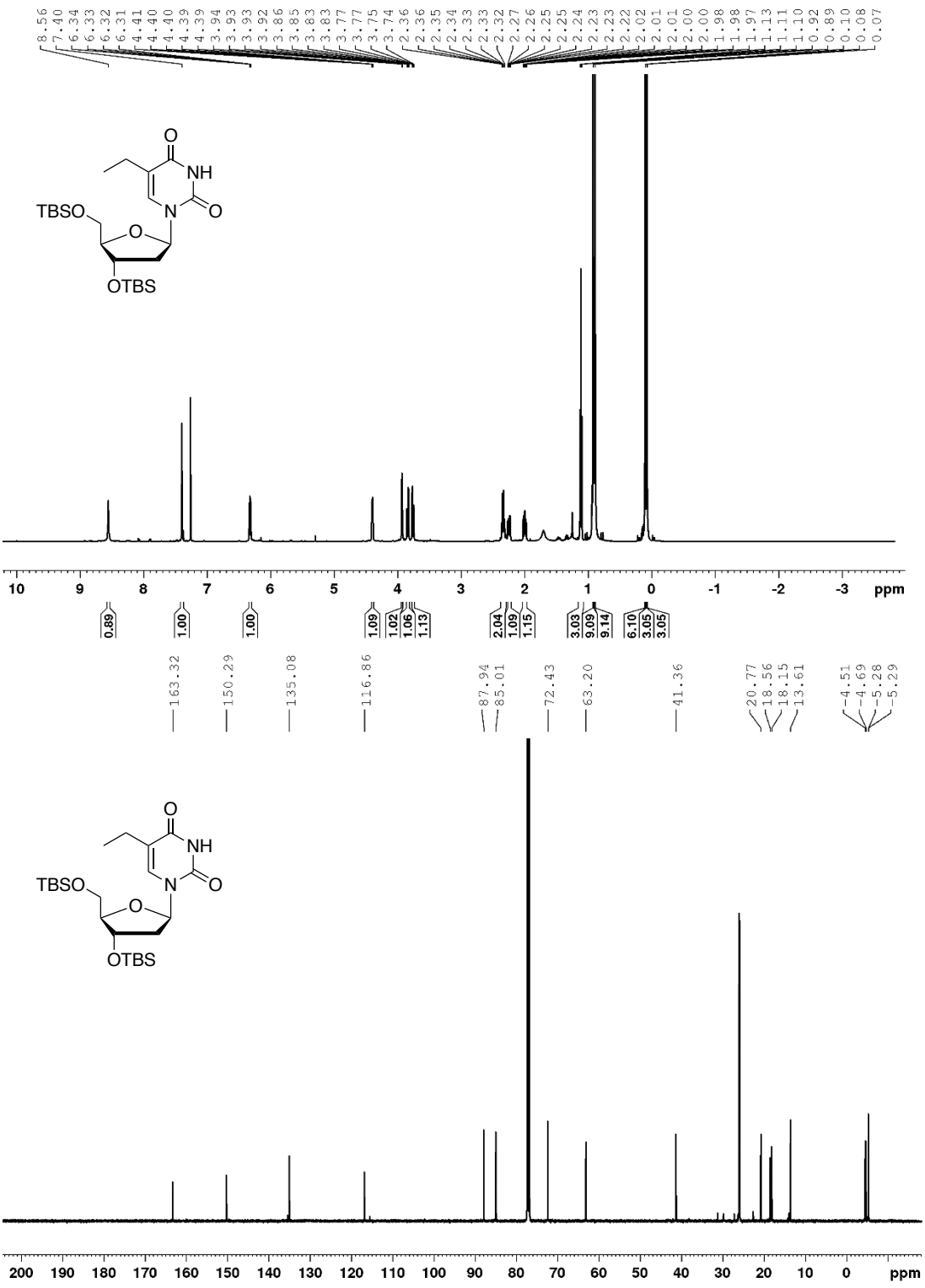
Supplementary Figure S40. ^{1}H and ^{13}C NMR spectra of the title compound in CD_2Cl_2.



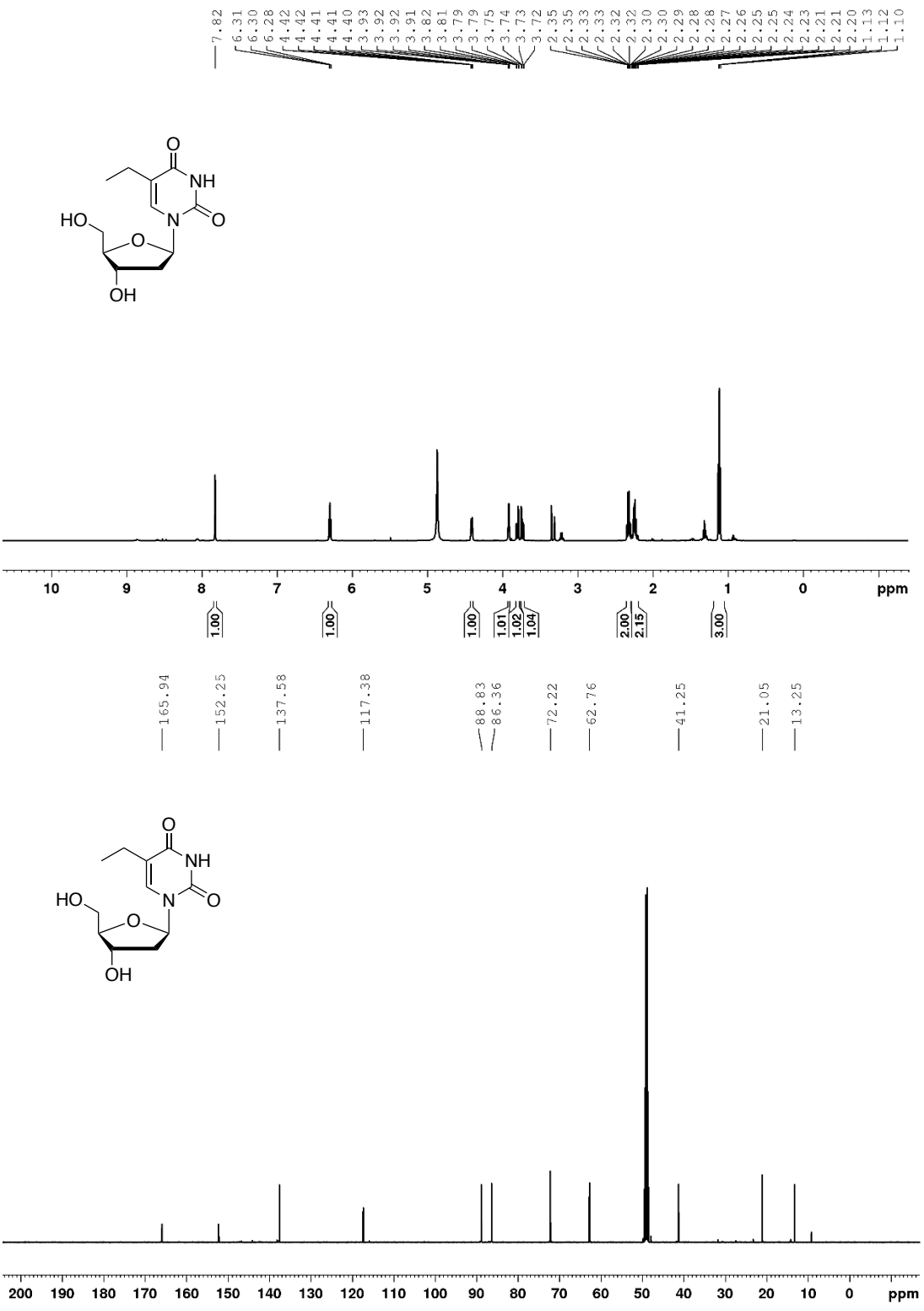
Supplementary Figure S41. ^{1}H and ^{13}C NMR spectra of compound **5** in CD_2Cl_2.



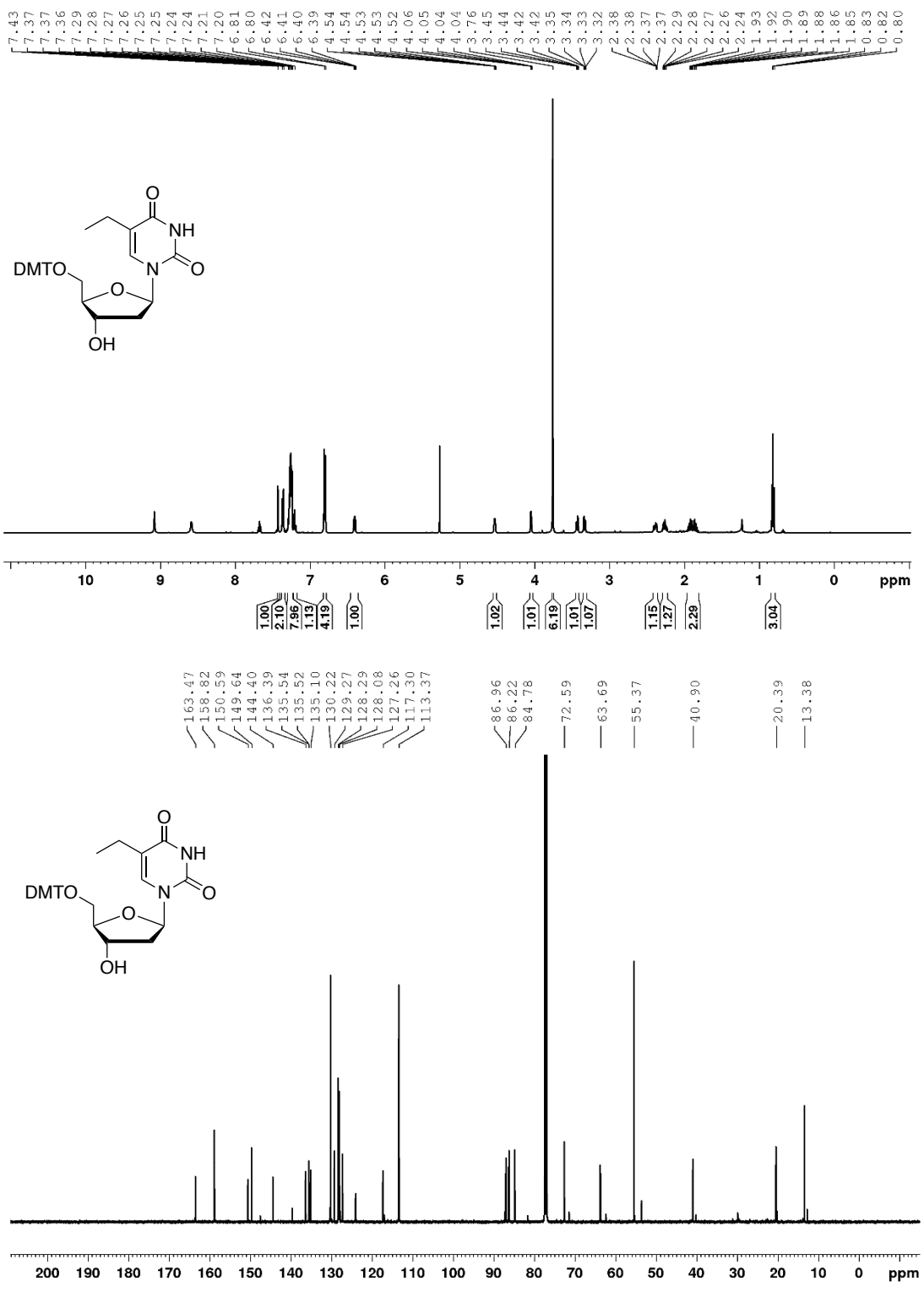
Supplementary Figure S42. ^{31}P NMR spectrum of compound **5** in CD_2Cl_2.



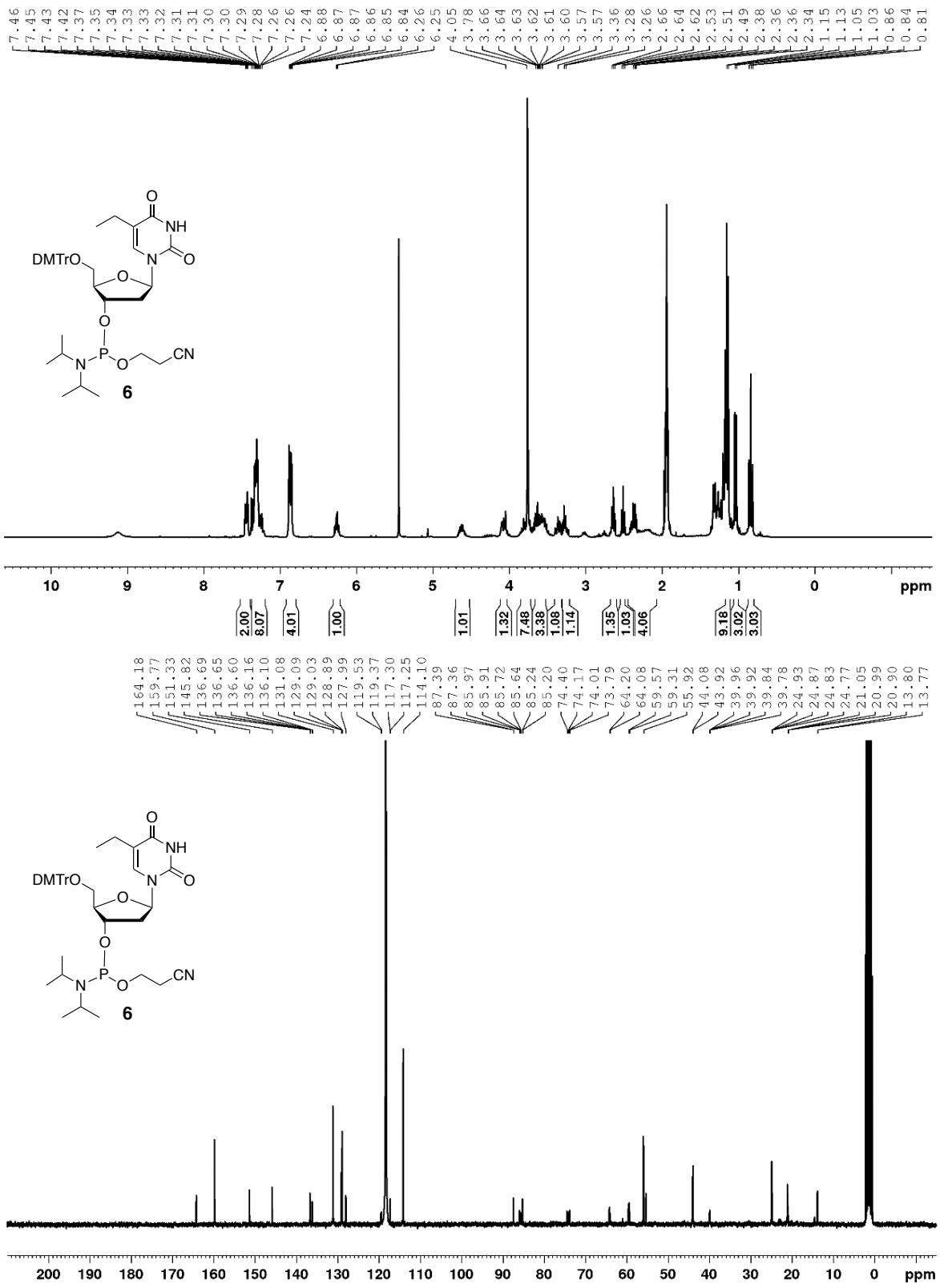
Supplementary Figure S43. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$



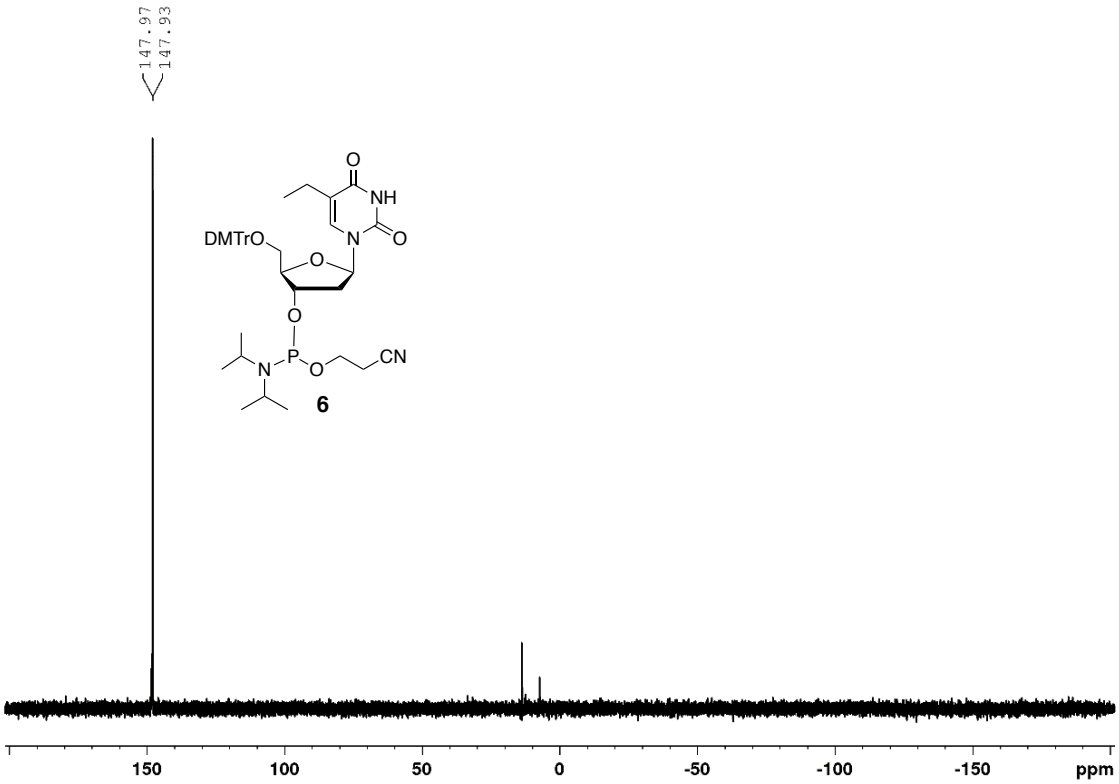
Supplementary Figure S44. ^{1}H and ^{13}C NMR spectra of the title compound in CD_3OD



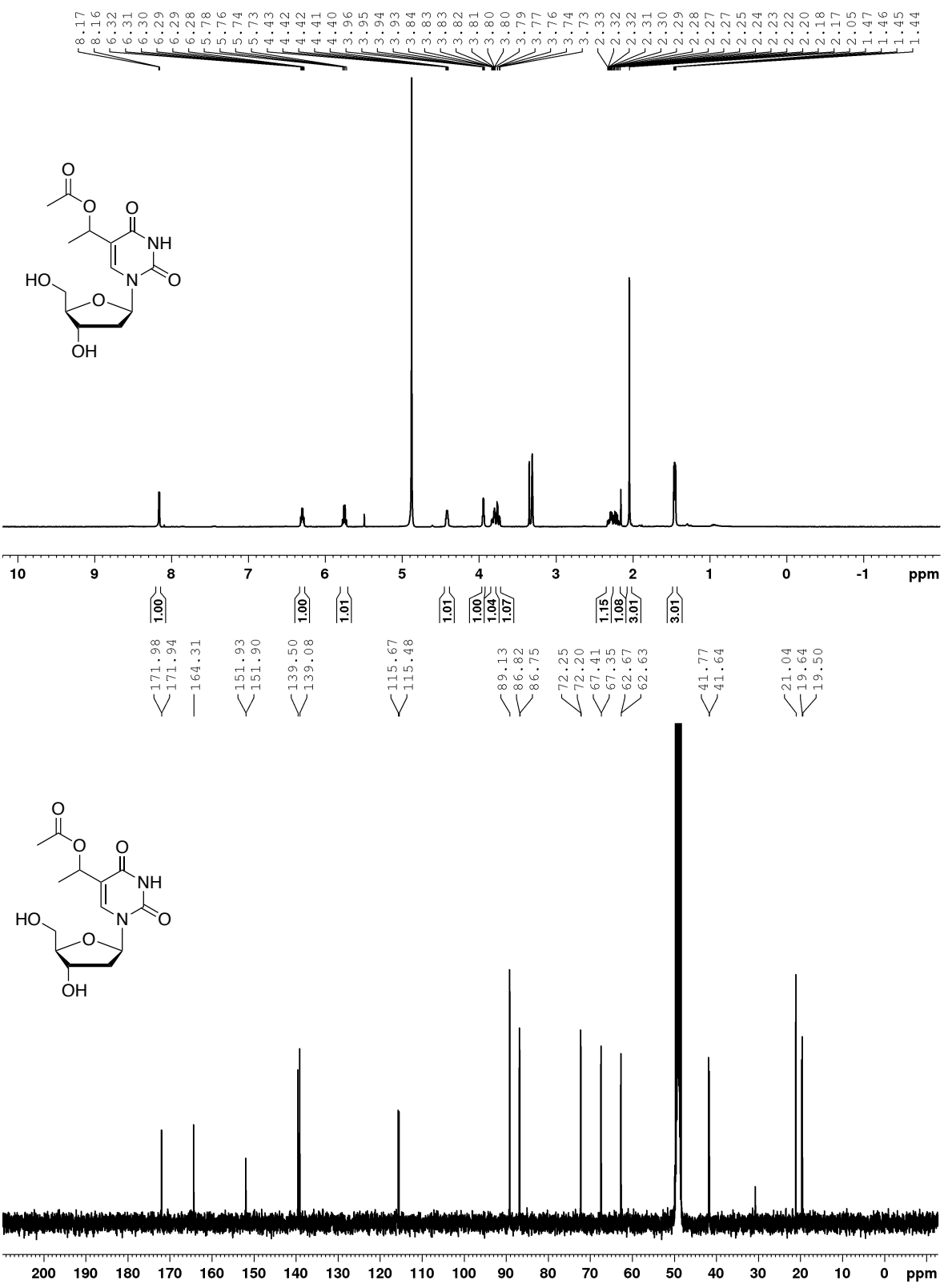
Supplementary Figure S45. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$



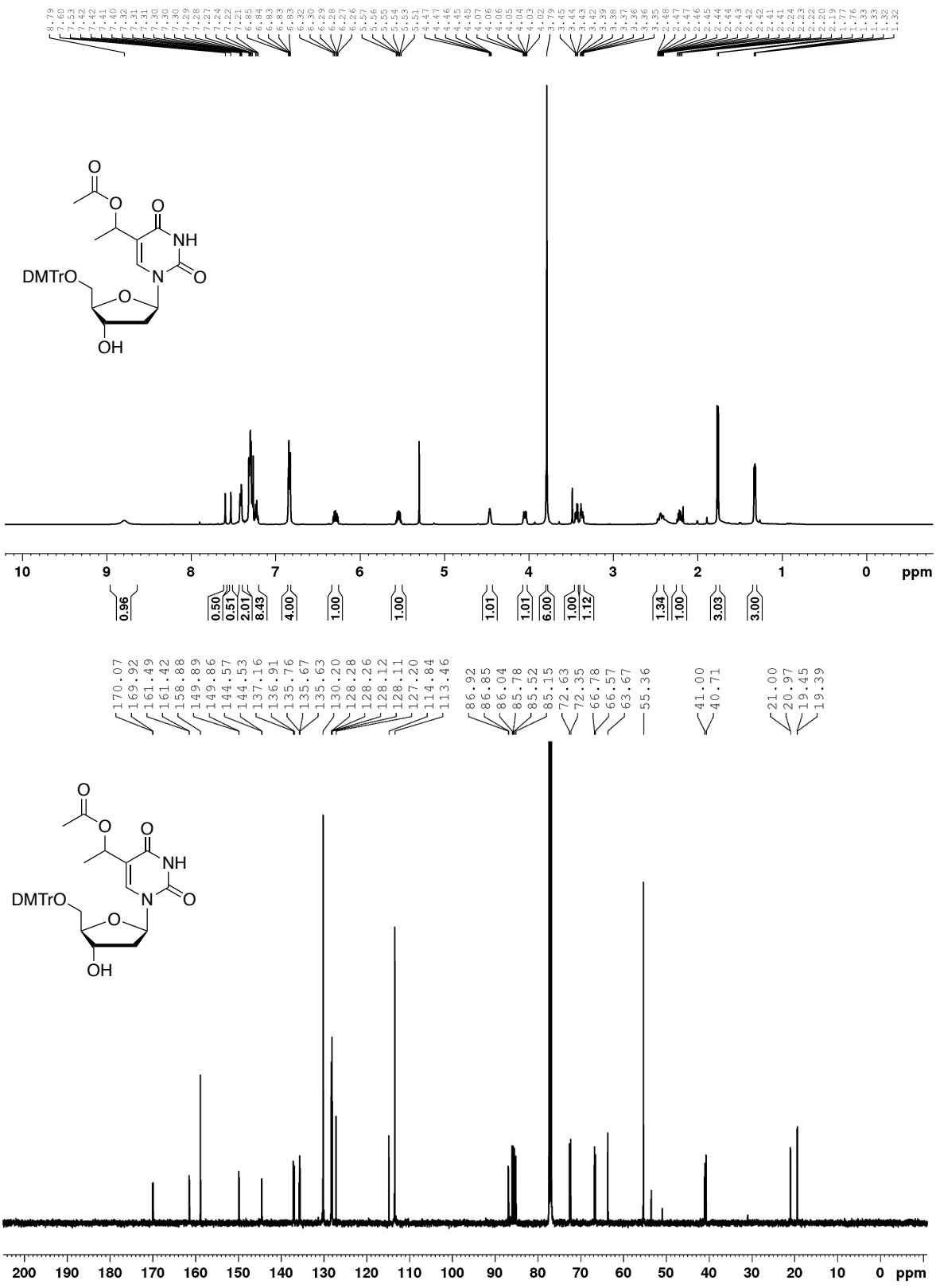
Supplementary Figure S46. ^{1}H and ^{13}C NMR spectra of compound **6** in acetonitrile-d_3.



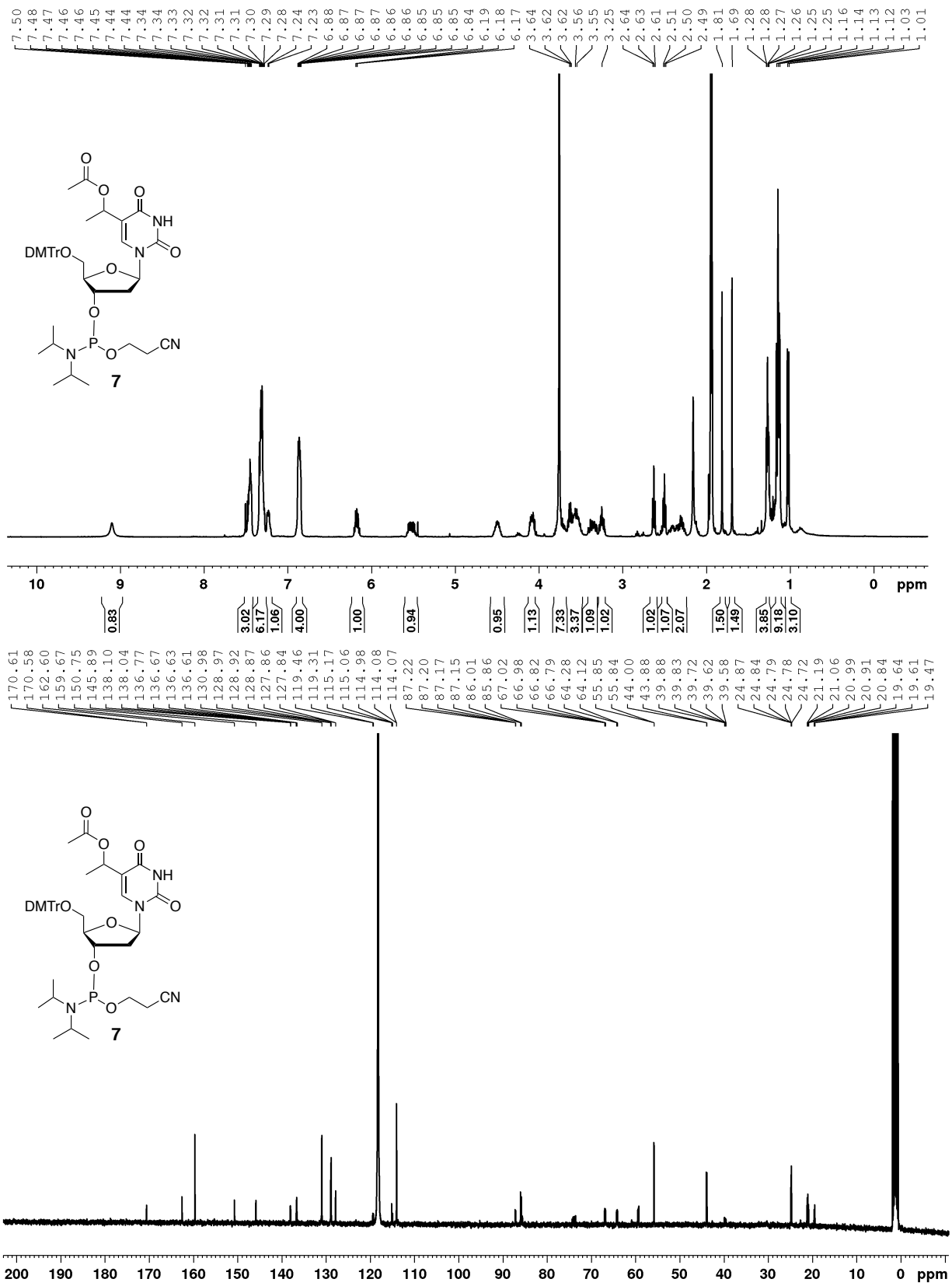
Supplementary Figure S47. ^{31}P NMR spectrum of compound **6** in acetonitrile-d_3.



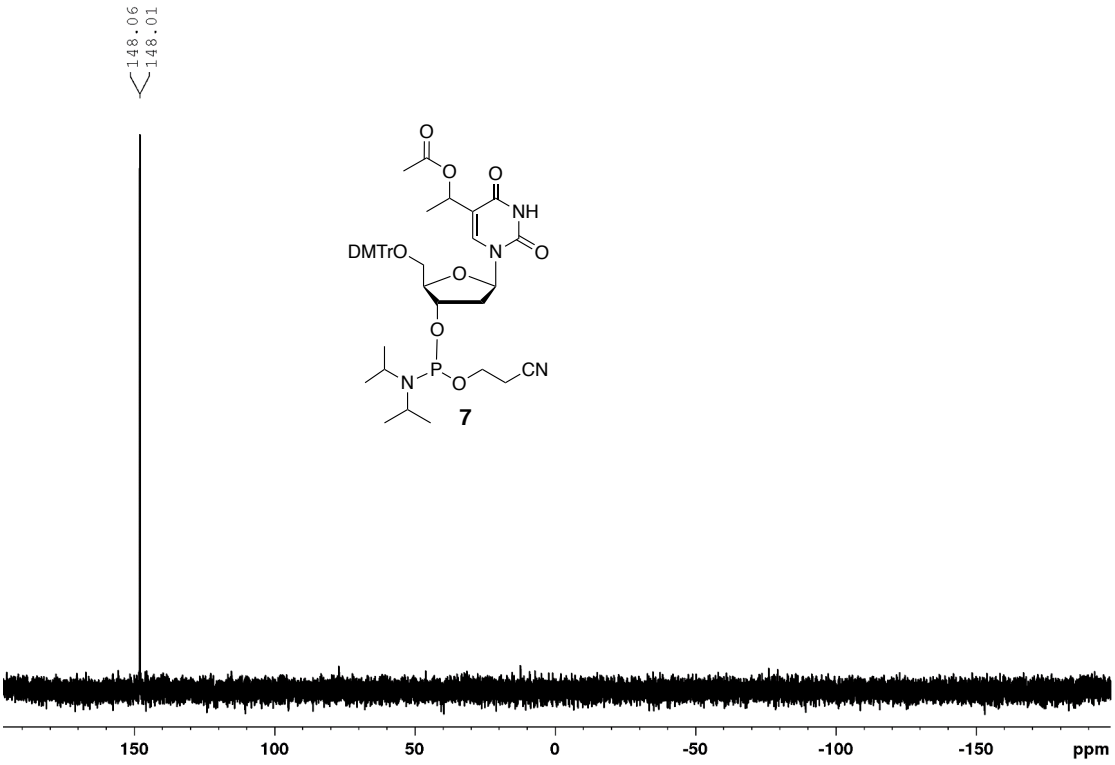
Supplementary Figure S48. ^{1}H and ^{13}C NMR spectra of the title compound in methanol-d_4.



Supplementary Figure S49. ^{1}H and ^{13}C NMR spectra of the title compound in $CDCl_3$.



Supplementary Figure S50. ^{1}H and ^{13}C NMR spectra of compound **7** in acetonitrile-d_3.



Supplementary Figure S51. ^{31}P NMR spectrum of compound **7** in acetonitrile-d_3.

GENE	VECTOR	AFFINITY TAG	RESISTANCE
TET2	pPEI	N-6xHis	Kanamycin
UHRF2	pET28a-LIC	N-6xHis	Kanamycin
TDG	pET28b	N-6xHis	Kanamycin
AID	pET28b	C-6xHis	Kanamycin
APOBEC3A	pET21	N-6xHis	Ampicillin
β-GT	pET28b	N-6xHis	Kanamycin

Supplementary Table S2. List of the genes used in the current study. The expression vector, antibiotic resistance and the affinity tag present for protein purification are provided.

Mutations	Mutagenic Primers
AID_Y116A	**FWD:** 5'-GAT TTT CAC CGC GCG CCT CGC CTT CTG TGA AGA CC-3' **REV:** 5'-GGT CTT CAC AGA AGG CGA GGC GCG CGG TGA AAA TC-3'
APOBEC3A_Y130A	**FWD:** 5'-CGC TGC CCG CAT CGC AGA TTA CGA CCC CCT ATA TAA-3' **REV:** 5'-TTA TAT AGG GGG TCG TAA TCT GCG ATG CGG GCA GCG-3'

Supplementary Table S3. List of primers designed for site-directed mutagenesis.

Proteins	Bacterial cells for protein expression
TET2	*E. coli* BL21 [DE3] Star
UHRF2	*E. coli* BL21 [DE3] Star
TDG	*E. coli* BL21 [DE3] Star
AID	*E. coli* BL21 [DE3] Rosetta2
AID-Y116A	*E. coli* BL21 [DE3] Rosetta2
APOBEC3A	*E. coli* BL21 [DE3] Star
APOBEC3A_Y130A	*E. coli* BL21 [DE3] Star
β-GT	*E. Coli* BL21 codon plus (DE3) RIPL

Supplementary Table S4. List of competent bacterial cells used in the current study for the expression of the indicated proteins.

8. References

1. M. J. Robins and P. J. Barr, *J. Org. Chem.* 1983, **48**, 1854-1862.
2. R. Guza, D. Kotandeniya, K. Murphy, T. Dissanayake, C. Lin, G. M. Giambasu, R. R. Lad, F. Wojciechowski, S. Amin, S. J. Sturla, R. H. Hudson, D. M. York, R. Jankowiak, R. Jones and N. Y. Tretyakova, *Nucleic Acids Res.* 2011, **39**, 3988-4006.
3. S. A. Ingale, H. Mei, P. Leonard and F. Seela, *J. Org. Chem.*2013, **78**, 11271-11282.
4. J. L. Ruth and D. E. Bergstrom, *J. Org. Chem.*1978, **43**, 2870-2876.
5. D. Kotandeniya, C. L. Seiler, J. Fernandez, S. S. Pujari, L. Curwick, K. Murphy, S. Wickramaratne, S. Yan, D. Murphy, Y. Y. Sham and N. Y. Tretyakova, *Chem. Commun.* 2018, **54**, 1061-1064.
6. L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield and B. G. Davis, *Angew. Chem. Int. Ed.* 2013, **52**, 10553-10558.

7. L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang and Y. Xu, *Cell*, 2013, **155**, 1545-1555.
8. B. Sudhamalla, S. Wang, V. Snyder, S. Kavoosi, S. Arora and K. Islam, *J. Am. Chem. Soc.* 2018, **140**, 10263-10269.
9. E. Besmer, E. Market and F. N. Papavasiliou, *Mol. Cell. Biol.* 2006, **26**, 4378-4385.
10. I. J. Byeon, J. Ahn, M. Mitra, C. H. Byeon, K. Hercik, J. Hritz, L. M. Charlton, J. G. Levin and A. M. Gronenborn, *Nat. Commun.* 2013, **4**, 1890.
11. L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg and A. C. Drohat, *Biochem.* 2016, **55**, 6205-6208.
12. L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang and C. He, *Nat. Chem. Biol.* 2012, **8**, 328-330.
13. A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada and H. Leonhardt, *Nucleic Acids Res.* 2010, **38**, e181.
14. B. Sudhamalla, D. Dey, M. Breski and K. Islam, *Anal. Biochem.* 2017, **534**, 28-35.
15. H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang and X. Cheng, *Nucleic Acids Res.* 2012, **40**, 4841-4849.
16. T. Zhou, J. Xiong, M. Wang, N. Yang, J. Wong, B. Zhu and R. M. Xu, *Mol. Cell*, 2014, **54**, 879-886.
17. M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, *Cell*, 2012, **151**, 1417-1430.
18. A. Maiti and A. C. Drohat, *J. Bio. Chem.* 2011, **286**, 35334-35338.
19. E. K. Schutsky, C. S. Nabel, A. K. F. Davis, J. E. DeNizio and R. M. Kohli, *Nucleic acids research*, 2017, **45**, 7655-7665.
20. Q. Qiao, L. Wang, F. L. Meng, J. K. Hwang, F. W. Alt and H. Wu, *Molecular Cell*, 2017, **67**, 361-373.e364.
21. C. X. Song, T. A. Clark, X. Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He and J. Korlach, *Nat. Methods*, 2011, **9**, 75-77.
22. M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X. J. Cao, B. A. Garcia, G. A. Cisneros and R. M. Kohli, *Nat. Chem. Biol.* 2017, **13**, 181-187.