Supporting Information

Structural basis for excision of 5-formylcytosine by Thymine DNA Glycosylase

Lakshmi S. Pidugu[1], Joshua Flowers[2], Christopher T. Coey[1], Edwin Pozharski[1,3,*], Marc M. Greenberg[2,*], and Alexander C. Drohat[1,*]

[1]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201, United States, [2]Department of Chemistry, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, United States, [3]Center for Biomolecular Therapeutics, Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, United States

Contents of Supporting Information:

Experimental Methods

Oligodeoxynucleotides. Standard oligodeoxynucleotides (ODNs) were obtained from IDT and were purified by reverse phase HPLC,[1] exchanged into 0.02 M Tris-HCl pH 7.5, 0.04 M NaCl, and quantified by absorbance as described.[2] ODNs containing 2´-fluoroarabino-5-formyldeoxycytidine (hereafter, fdC^F) were prepared as described below. The DNA used for crystals included a 28mer target strand, 5'-AGC TGT CCA TCG CTC AxG TAC AGA GCT G-3', where x is fdC^F, and its complement, 5'-CAG CTC TGT ACG TGA GCG ATG GAC AGC T-3', such that fdC^F is paired with dG and located in a CpG dinucleotide context.[2,3] The same 28 bp DNA construct was used for glycosylase assays (with x = fC).

Enzymes. Full length human TDG (410 residues), and its N140A variant, were expressed in *E coli* and purified essentially as previously described.[2,4] TDG^{82-308}, a new construct containing residues Ser^{82}-Val^{308} of human TDG was expressed and purified as described.[5] The enzyme preparations were >99% pure, as judged by SDS-PAGE (Coomassie-stained gel), and the enzyme concentration was determined by absorbance at 280 nm.[6,7] The extinction coefficient for TDG^{82-308} is identical to that for $TDG^{111-308}$.[8]

X-ray Crystallography. Samples used for crystallization contained 0.35 mM enzyme (TDG^{82-308}) and 0.42 mM DNA in a buffer of 5 mM Tris-HCl pH 7.5, 0.13 M NaCl, 0.2 mM DTT, 0.2 mM EDTA. Crystals were grown at room temperature (~22 °C) by sitting drop vapor diffusion, using 1 µl of the TDG-DNA sample and 1 or 2 ul of mother liquor, which was 30% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 6.0. Crystals typically appeared within in a few days. Crystals were cryo-protected using mother liquor supplemented with 18% ethylene glycol and flash cooled in liquid nitrogen. X-ray diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL beamline 12-2). Images were processed using XDS [9] and scaled with Aimless [10] from the CCP4 program suite [11] with the help of the autoxds script developed by Ana Gonzalez and Yingssu Tsai (http://smb.slac.stanford.edu/facilities/software/xds). Resolution cutoff was determined based on CC1/2 values.[12] Structures were solved by molecular replacement using Phaser,[13] and a previously reported structure of DNA-bound TDG^{82-308} as the search model (PDBID: 4Z47). Refinement was performed using BUSTER-TNT,[14] or REFMAC5,[15] and model building was performed using Coot.[16] TLS refinement protocol utilized TLSMD

server,[17,18] as described.[1] The structural figures were made with PyMOL (http://www.pymol.org).

Glycosylase Assays. Single turnover kinetics reactions were initiated by adding enzyme (TDG or TDG^{82-308}) to G·fC substrate (0.5 uM) in HEMN.1 buffer (0.02 M HEPES pH 7.5, 0.1 M NaCl, 0.2 mM EDTA, 2.5 mM $MgCl_2$). Aliquots were removed at desired time points, quenched with 50% (v:v) 0.3 M NaOH, 0.03 M EDTA, and heated (3 m, 85 °C) to quantitatively cleave the DNA backbone at abasic sites. The resulting DNA fragments were resolved by HPLC and peak areas were used to determine fraction product.[2] Progress curves (fraction product vs. time) were fitted by non-linear regression to eq. 1:

$$\text{fraction product} = A(1 - \exp(-k_{obs}t)) \quad (1)$$

where A is the amplitude, k_{obs} is the rate constant, and t is the reaction time. Experiments were performed with saturating enzyme ([E] >> K_d; [E] > [S]) such that the observed rate constant reflects the maximal rate of product formation ($k_{obs} \approx k_{max}$) and is not influenced by enzyme-substrate association or by product release or product inhibition.[4]

Acknowledgement

Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office (DOE) by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS; including P41GM103393) and the National Center for Research Resources (NCRR; P41RR001209). The Berkeley Center for Structural Biology is supported in part by the NIH, NIGMS, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. DOE under Contract No. DE-AC02-05CH11231. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, NCRR or NIH.

Table S1. Data collection and refinement statistics

	TDG^{82-308} dG:fdC^{F} (PDB ID: 5T2W)
Data collection	
Space Group	C2
Cell Dimensions	
a, b, c (Å)	83.5,51.9,82.0
β (°)	101.5
Resolution (Å)	39.83-2.20 (2.27-2.20)
R_{pim}	0.170
Mean *I*/σ*I*	4.3 (0.7)
$CC_{1/2}$	0.982 (0.316)
Completeness (%)	99.9 (100.0)
Redundancy	10.1 (10.2)
Wilson B-factor ($Å^2$)	41.5
Refinement	
Program	Buster 2.10.2
Resolution (Å)	39.83-2.20
No. of reflections	17639
R_{work}/R_{free}	0.180/0.226
Number of atoms	
Protein	1541
DNA	1258
Water	156
B-factors ($Å^2$)	
Protein	34.8
DNA	49.7
Water	43.5
Ramachandran Plot	
Favoured (%)	97.4
Allowed (%)	2.6
Outliers (%)	0
RMSD from ideal	
Bond lengths (Å)	0.010
Bond angles (°)	0.99

Values shown in parenthesis are for highest resolution shell. The Ramachandran analysis was performed using Molprobity [19]. Wilson B-factor estimated by phenix.xtriage. Number of atoms includes all atom records explicitly included in the model, including alternate positions.

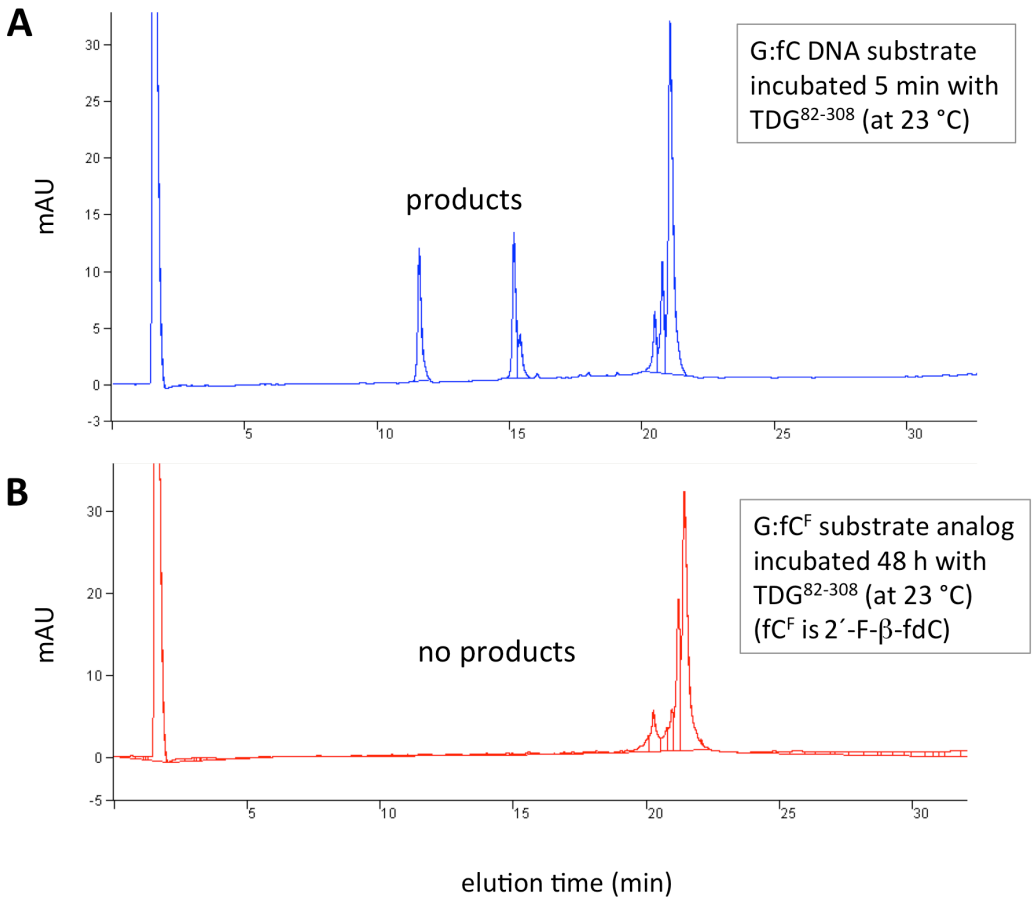


Figure S1. HPLC analysis of TDG^{82-308} activity on fC substrate and non-cleavable fdC^F analog. (A) Reaction of TDG^{82-308} with normal G·fC DNA substrate for 5 min (23 °C) yields a substantial amount of product. (B) Incubation of TDG^{82-308} with DNA containing the fdC^F analog (paired with dG) for 48 h (23 °C) yields no detectable product (fdC^F is 2´-fluoroarabino-5-formyl-dC). The experimental conditions were as described above for glycosylase assays.

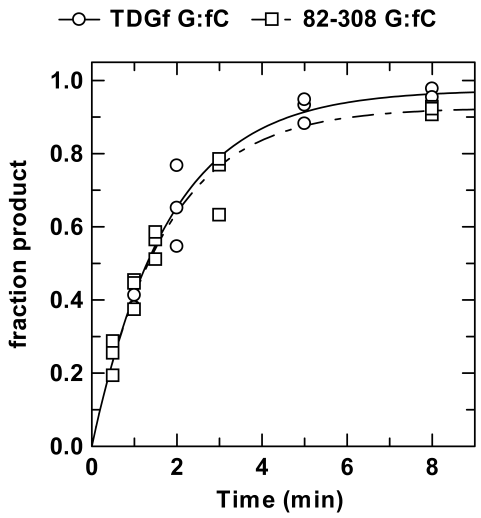


Figure S2. The maximal rate of fC excision by TDG$^{82\text{-}308}$ is the same as that of full-length TDG. Single-turnover kinetics experiments were performed for excision of fC from a G·fC DNA substrate by TDG and TDG$^{82\text{-}308}$ at 23 °C. The data were fitted to a single exponential equation, giving k_{obs} = 0.56 ± 0.05 min^{-1} for TDG and k_{obs} = 0.59 ± 0.04 min^{-1} for TDG$^{82\text{-}308}$. Given the saturating concentration of enzyme (2.5 uM) relative to G·fC substrate (0.5 uM), the rate constants represent the maximal rate of product formation ($k_{obs} \approx k_{max}$).

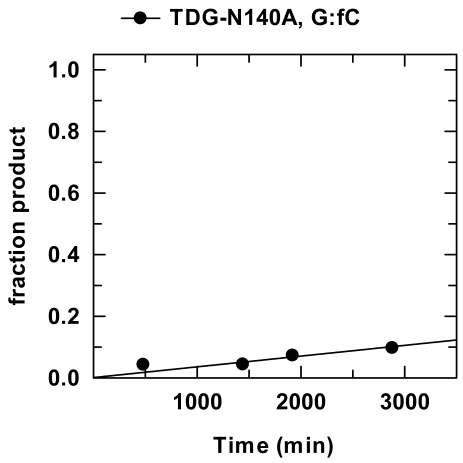
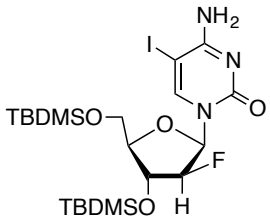
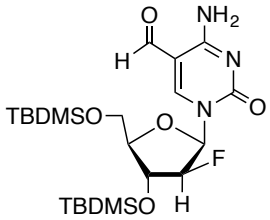


Figure S3. Single turnover kinetics data for excision of fC from G·fC DNA by the N140A variant of full length TDG at 23 °C. Due to the low amount of product formation, the data were fitted by linear regression, giving a rate constant of k_{obs} = (3.5 ± 0.2) × 10^{-5} min^{-1}. Given the saturating concentration of enzyme (2.5 uM) relative to G·fC substrate (0.5 uM), the rate constant represents the maximal rate of product formation ($k_{obs} \approx k_{max}$).

Synthesis of ODNs containing 2′-deoxy-2′-fluoro-D-arabinofuranosyl-5-formyl-cytosine (2'-F-fdC). ODNs containing 2'-F-fdC were prepared via solid-phase oligonucleotide synthesis using phosphoramidite **4**, which was synthesized in a manner similar to previously reported methods.[20-22] The 2'-fluoronucleoside core was prepared using a strategy described by Damha and more recently by He.[20,21] The protecting group scheme developed by Carell was employed to successfully introduce the 5-formylcytosine.[22]

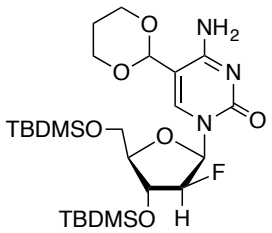


Preparation of S1. 2′-Deoxy-2′-fluoro-D-arabinofuranosyl-5-iodo-cytosine[21] (1.14 g, 3.07 mmol), imidazole (0.941 g, 13.82 mmol), and TBDMSCl (1.39 g, 9.21 mmol) were dissolved in DMF (18.3 mL). The reaction was stirred overnight at 25 °C before being quenched with sat. $NaHCO_3$ and extracted with $CHCl_3$. The organic layer was then washed with water and dried over $MgSO_4$. The solution was filtered, concentrated under reduced pressure, and purified by flash chromatography (5% MeOH/DCM). This yielded 1.56 g (85%) of **S1** as a white solid. ^{1}H-NMR ($CDCl_3$): δ 8.63 (br s, 1H), 7.83 (s, 1H), 6.19 (dd, J = 20, 4 Hz, 1H), 5.63 (br s, 1H), 4.89-5.06 (m, 1H), 4.35-4.43 (m, 1H), 3.92 (m, 1H), 3.79 (m, 2H), 0.93 (s, 9H), 0.89 (s, 9H), 0.11 (m, 12H); ^{13}C-NMR ($CDCl_3$): δ 163.9, 154.3, 147.4, 95.8, 93.9, 85.30, 85.26, 85.1, 75.4, 75.2, 61.5, 56.2, 25.8, 25.5, 18.3, 17.8, -4.9, -5.1, -5.40, -5.42; IR ($CDCl_3$): υ (cm^{-1}) 3450, 2932, 2859, 2247, 1650, 1490, 1391, 1362, 1335, 1289, 1257, 1107, 1006, 910, 838, 779. MS (ESI-TOF) (*m/z*): Calculated: 600.1581 [M+H], Observed: 600.1598.

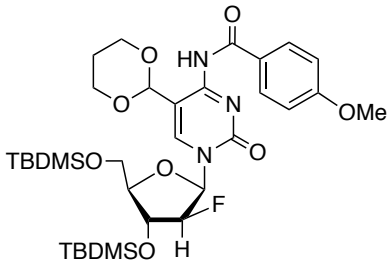


Preparation of S2. Compound **S1** (1.56 g, 2.61 mmol) and $Pd_2(dba)_3$ (230 mg, 0.250 mmol) were added to a pear flask, which was immediately purged with argon for 10 min. Toluene that was sparged for 30 min with argon was added (24.6 mL) to the pear flask. The solution was then sparged for an additional 15 min with argon, followed by 5 min with CO. Triphenylphosphine (394 mg, 1.50 mmol) was dissolved in 1 mL of sparged toluene and added to the pear flask. The solution was then sparged for an additional 5 min with CO. Freshly distilled Bu_3SnH (1.10 mL, 3.76 mmol) dissolved in sparged toluene (37.6 mL) was syringed into the pear flask at 50 °C over a period of 10 h. The reaction was then cooled to 25 °C and filtered through celite before

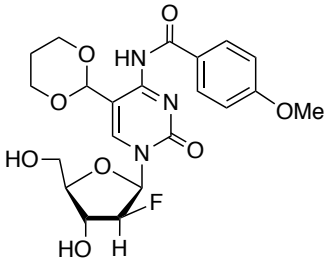
being concentrated under reduced pressure. The crude material was purified by flash chromatography (4:1, 2:1, 1:1 Hex/EtOAc), followed by a second column purification (2% MeOH/DCM). This yielded 858 mg (66%) of **S2** as a slightly yellow solid. ^{1}H-NMR ($CDCl_3$): δ 9.47 (s, 1H), 8.23 (br s, 1H), 8.21 (s, 1H), 7.41 (br s, 0.5 H), 6.22-6.29 (m, 1H), 4.98-5.14 (m, 1H), 4.39-4.46 (m, 1H), 4.00 (m, 1H), 3.82 (m, 2H), 0.91 (s, 9H), 0.90 (s, 9H), 0.11 (m, 12H); ^{13}C-NMR ($CDCl_3$): δ 187.2, 162.8, 154.2, 152.8, 105.3, 95.9, 93.9, 86.1, 86.0, 85.9, 75.6, 75.4, 61.7, 25.8, 25.6, 18.3, 17.9, -4.8, -5.0, -5.4; IR ($CDCl_3$): υ (cm^{-1}) 3395, 2931, 2858, 1672, 1516, 1472, 1418, 1255, 1106, 838, 780. MS (ESI-TOF) (*m/z*): Calculated: 502.2563 [M+H], Observed: 502.2586.



Preparation of 2. Compound **S2** (758 mg, 1.51 mmol) was dissolved in DCM (14.6 mL) at 0 °C under argon. Triethylorthoformate (0.313 mL, 1.82 mmol) and 1,3-propanediol (0.433 mL, 6.04 mmol) were syringed into the flask, followed by the dropwise addition of 1M $TiCl_4$ in DCM (0.604 mL, 0.604 mmol). After the addition of $TiCl_4$ the reaction was warmed up to 25 °C and stirred overnight. The reaction was quenched with ice and sat. $NaHCO_3$. The aqueous layer was extracted with DCM, washed with brine, and dried over $MgSO_4$. The solution was then filtered, concentrated under reduced pressure, and purified by flash chromatography (3-5% MeOH/DCM). This yielded 369 mg of **2** and 352 mg (52%) of the monodesilylated product. The monodesilylated product was dissolved in DMF (3.95 mL) along with imidazole (121 mg, 1.78 mmol) and TBDMSCl (180 mg, 1.19 mmol). The reaction was stirred overnight, quenched with sat. $NaHCO_3$, and extracted with $CHCl_3$. The organic layer was washed with water and dried over $MgSO_4$. The solution was then filtered, concentrated under reduced pressure, and purified by flash chromatography (3-5% MeOH/DCM). This yielded an additional 394 mg of **2** for an overall yield of 763 mg (90%). ^{1}H-NMR ($CDCl_3$): δ 8.48 (br s, 1H), 7.58 (s, 1H), 6.31 (br s, 1H), 6.12 (dd, J = 20, 3 Hz, 1H), 5.20 (s, 1H), 4.83-4.99(m, 1H), 4.29-4.36 (m, 1H), 4.13 (m, 2H), 3.71-3.86 (m, 5H), 2.09 (m, 1H), 1.36 (m, 1H), 0.86 (s, 9H), 0.82 (s, 9H), 0.04 (m, 12H); ^{13}C-NMR ($CDCl_3$): δ 163.6, 154.4, 140.9, 103.4, 98.6, 95.7, 93.8, 85.2, 85.0, 75.5, 75.2, 67.0, 61.5, 25.7, 25.4, 25.2, 18.1, 17.7, -5.1, -5.2, -5.7; IR ($CDCl_3$): υ (cm^{-1}) 3461, 3127, 2596, 2931, 2859, 1671, 1515, 1473, 1361, 1257, 1148, 1101, 1050, 1007, 952, 918, 838, 807, 780; MS (ESI-TOF) (*m/z*): Calculated: 560.2982 [M+H], Observed: 560.3022.

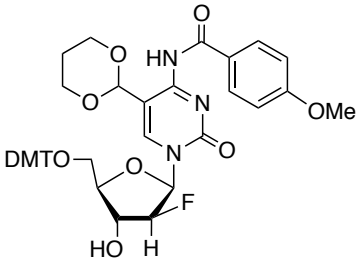


Preparation of S3. Compound **2** (394 mg, 0.704 mmol) was azeotropically dried (3x) with pyridine before being dissolved in pyridine (8.8 mL) under argon at 0 °C. *p*-Methoxybenzoyl chloride (0.104 mL, 0.774 mmol) was syringed in dropwise. The solution was warmed to 25 °C and stirred overnight. The reaction was quenched with ice and sat. $NaHCO_3$. The aqueous layer was then extracted with EtOAc (3x), washed with NH_4Cl, and dried over $MgSO_4$. The solution was filtered and concentrated under reduced pressure before purification by flash chromatography (0-10% Acetone/DCM). This yielded 277 mg (57%) of **2**. ^{1}H-NMR ($CDCl_3$): δ 13.5 (br s, 1H), 8.2 (d, J = 8 Hz, 2H), 7.91 (d, J = 1 Hz, 1H), 6.93 (d, J = 9 Hz, 2H), 6.16 (dd, J = 20, 4 Hz, 1H), 5.81 (br s, 1H), 4.88-5.04 (m, 1H), 4.45 (dd, J = 19, 4 Hz, 1H), 4.24 (dd, J = 11, 5 Hz, 2H), 4.02 (m, 2H), 3.93 (m, 1H), 3.87 (s, 3H), 3.84 (d, J = 5 Hz, 2H), 2.18 (m, 1H), 1.45 (m, 1H), 0.94 (s, 9H), 0.90 (s, 9H), 0.13 (s, 6H), 0.12 (s, 6H); ^{13}C-NMR ($CDCl_3$): δ 163.2, 157.4, 131.9, 129.7, 113.4, 96.1, 95.9, 94.2, 85.5, 85.2, 84.9, 75.5, 75.2, 67.6, 61.4, 55.4, 25.9, 25.64, 25.59, 18.3, 17.9, -4.9, -5.0, -5.45, -5.53; IR ($CDCl_3$): υ (cm^{-1}) 2930, 2857, 1713, 1653, 1572, 1472, 1334, 1281, 1253, 1163, 1104, 837, 780; MS (ESI-TOF) (*m/z*): Calculated: 694.3350 [M+H], Observed: 694.3363.

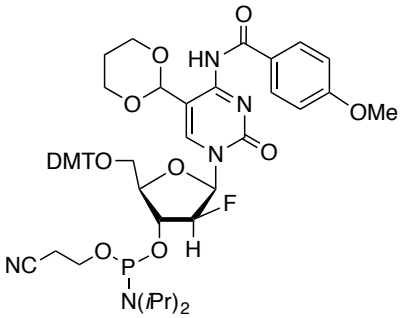


Preparation of 3. Compound **S3** (277 mg, 0.4 mmol) was dissolved in EtOAc (5.70 mL) under argon at 25 °C. Pyridine (0.161 mL, 0.2 mmol) and 70% HF·pyridine (0.105 mL, 4.0 mmol) were then syringed into the solution that was stirred overnight. The reaction was quenched with TMSOMe (0.4 mL) and was stirred for another 30 min before evaporating under reduced pressure and purifying by flash chromatography (6-10% MeOH/DCM) to yield 109 mg (58%) of **3**. ^{1}H-NMR (CD_3OD): δ 8.25 (s, 1H), 7.91 (s, 2H), 7.04 (d, J = 8 Hz, 1H), 6.18 (dd, J = 18, 4 Hz, 1H), 5.61 (s, 1H), 5.03-5.20 (m, 1H), 4.33 (ddd, J = 20, 4, 2 Hz, 1H), 4.25 (dd, J = 12, 4 Hz, 2H), 3.98-4.07 (m, 3H), 3.86 (s, 3H), 3.83 (dd, J = 12, 6 Hz, 1H), 3.77 (dd, J = 12, 6 Hz, 1H), 2.11 (m, 1H), 1.52 (d, J = 14 Hz, 1H); ^{13}C-NMR ($CDCl_3$): δ 165.2, 145.1, 131.2, 115.2, 97.2, 95.3, 87.0, 86.5, 75.1, 74.9, 68.8, 64.3, 61.9, 60.0, 56.1, 55.8, 54.8, 26.8; IR ($CDCl_3$): υ (cm^{-1}) 3368, 3360, 3351, 3337, 3325, 3316, 3309, 3282, 1711, 1671,

1605, 1576, 1484, 1456, 1253, 1177, 1092, 1066, 1046, 1026; MS (ESI-TOF) (*m/z*): Calculated: 466.1620 [M+H], Observed: 466.1643.



Preparation of S4. Compound **3** (109 mg, 0.234 mmol) was azeotropically dried (3x) with distilled pyridine. Pyridine (5.1 mL) was then used to dissolve **3** and DMTCl (95 mg, 0.28 mmol) was added. The reaction stirred overnight at 25 °C before quenching with MeOH and evaporating under reduced pressure. The crude material was redissolved in DCM, washed with $NaHCO_3$ (5% v/v), and brine, before drying over $MgSO_4$. The solution was filtered, evaporated under reduced pressure, and purified by flash chromatography (70% EtOAc/Hex) to yield 98 mg (55%) of **S4**. ^{1}H-NMR ($CDCl_3$): δ 8.22 (d, J = 1 Hz, 1H), 8.09 (s, 2H), 7.50 (m, 2H), 7.18-7.41 (m, 7H), 6.93 (d, J = 9 Hz, 2H), 6.84 (m, 4H), 6.30 (d, J = 20 Hz, 1H), 5.51 (s, 1H), 5.19 (d, J = 52 Hz, 1H), 4.52 (dd, J = 19, 3 Hz, 1H), 4.11 (m, 2H), 3.95 (m, 1H), 3.68-3.90 (m, 13H), 3.51 (d, J = 10 Hz, 1H), 3.30 (dd, J = 11, 5 Hz, 1H), 1.75 (s, 1H), 1.18 (d, J = 14 Hz, 1H); ^{13}C-NMR ($CDCl_3$): δ 163.2, 158.4, 144.8, 136.0, 135.9, 130.1, 128.2, 127.9, 126.7, 113.6, 113.2, 96.0, 94.1, 86.3, 84.0, 75.7, 75.4, 67.3, 67.2, 62.4, 60.4, 55.4, 55.2, 29.7, 25.2, 21.1, 14.2; IR ($CDCl_3$): υ (cm^{-1}) 3308, 2932, 2838, 1709, 1667, 1604, 1568, 1508, 1480, 1305, 1247, 1174, 1093, 1029, 950, 915, 828, 760, 727, 701, 646, 584; MS (ESI-TOF) (*m/z*): Calculated: 768.2927 [M+H], Observed: 768.2950.



Preparation of 4. Compound S**4** (98 mg, 0.128 mmol) was dissolved in a solution of DCM (5.4 mL) and DIPEA (0.15 mL, 0.85 mmol) at 0 °C under argon. Phosphoramidic chloride (0.043 mL, 0.192 mmol) was syringed in dropwise and the reaction was stirred for 2 h. The reaction was diluted with DCM, washed with $NaHCO_3$ (5% v/v) and then brine before drying over $MgSO_4$. The solution was filtered, dried under reduced pressure, and purified by flash chromatography (30-50% EtOAc/Hex). This yielded 85 mg (69%) of **4**. ^{1}H-NMR ($CDCl_3$): δ 13.48 (br s, 1H), 8.22 (m, 3H), 7.51 (m, 2H), 7.19-7.43 (m, 7H), 6.93 (d, J = 9 Hz, 2H), 6.86 (m, 4H), 6.24 (m, 1H), 5.73 (s, 1H), 5.04-5.27 (m, 1H), 4.63 (m, 1H), 4.08 (m, 1H), 3.71-3.96 (m, 14H), 3.56 (m, 4H), 3.29 (dd, J = 11, 6 Hz, 1H), 2.59 (t, J = 6 Hz, 1H), 2.35 (m, 1H), 0.98-1.20

(m, 12H); ^{31}P-NMR ($CDCl_3$): δ 151.50, 150.67; MS (ESI-TOF) (*m/z*): Calculated: 968.4005 [M+H], Observed: 968.3994.

Oligonucleotide synthesis. Oligonucleotide synthesis was carried out on an Applied Biosystems 394 DNA/RNA Synthesizer. Oligonucleotides containing native nucleotides only were prepared via standard conditions. Extended coupling (45 s), capping (25 s), and oxidation (15 s) times were used for coupling unmodified phosphoramidites in oligonucleotides that contained 2'-F-fdC. Phosphoramidite **4** was coupled for 5 min. Oligonucleotides containing 2'-F-fdC were deprotected using concentrated NH_4OH for 24 h at 25 °C. The deprotection was then flash frozen using liquid nitrogen and sublimated on the speed vac. The resin was then redissolved in 80% AcOH/H_2O and was placed in a thermocycler at 20 °C for 24 h. The solution was again flash frozen and sublimated on the speed vac. The resin was then redissolved in 90% formamide loading buffer and purified by gel electrophoresis (20% denaturing PAGE).

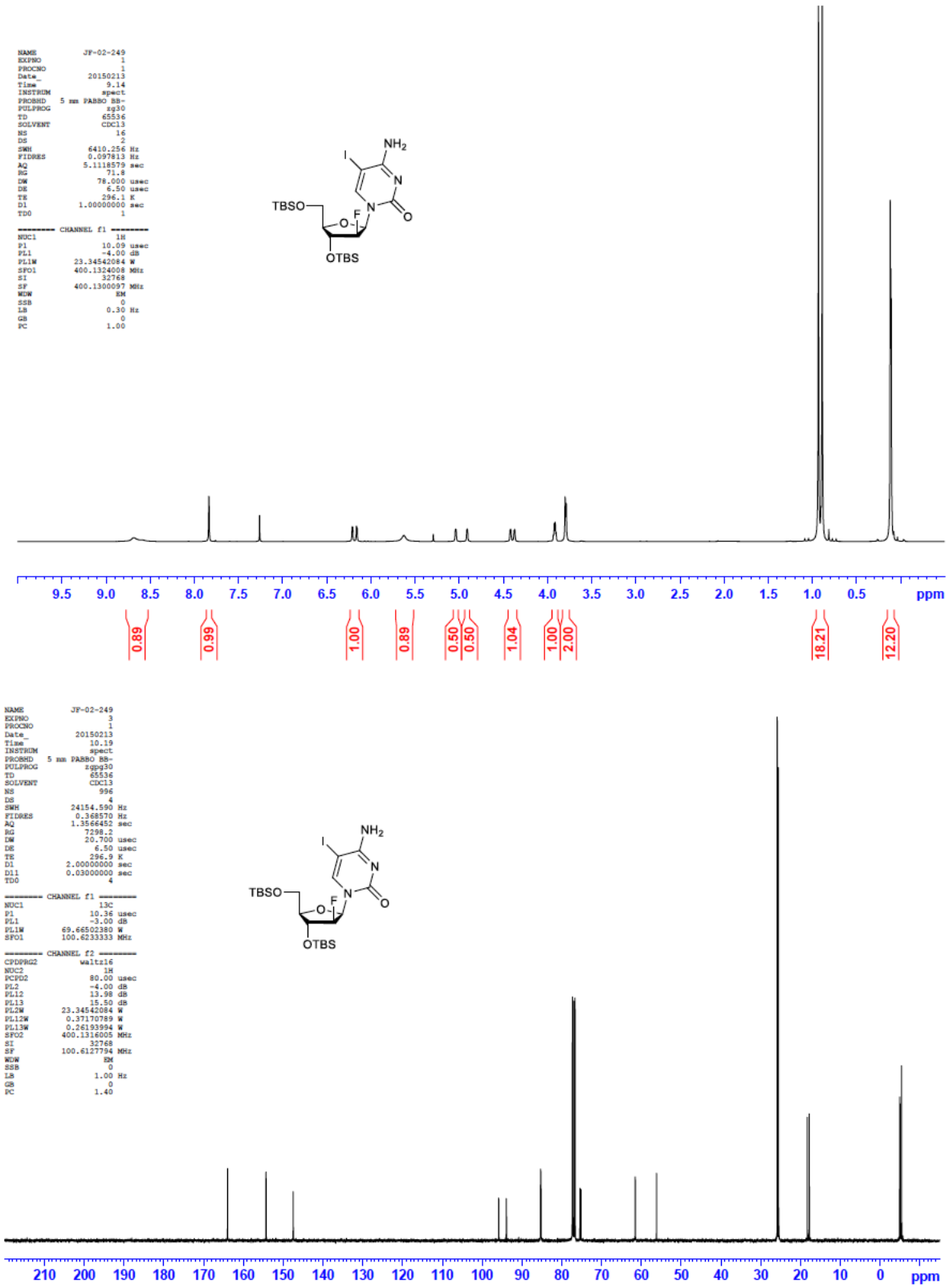


Figure S4. ^{1}H and ^{13}C NMR spectra of **S1**.

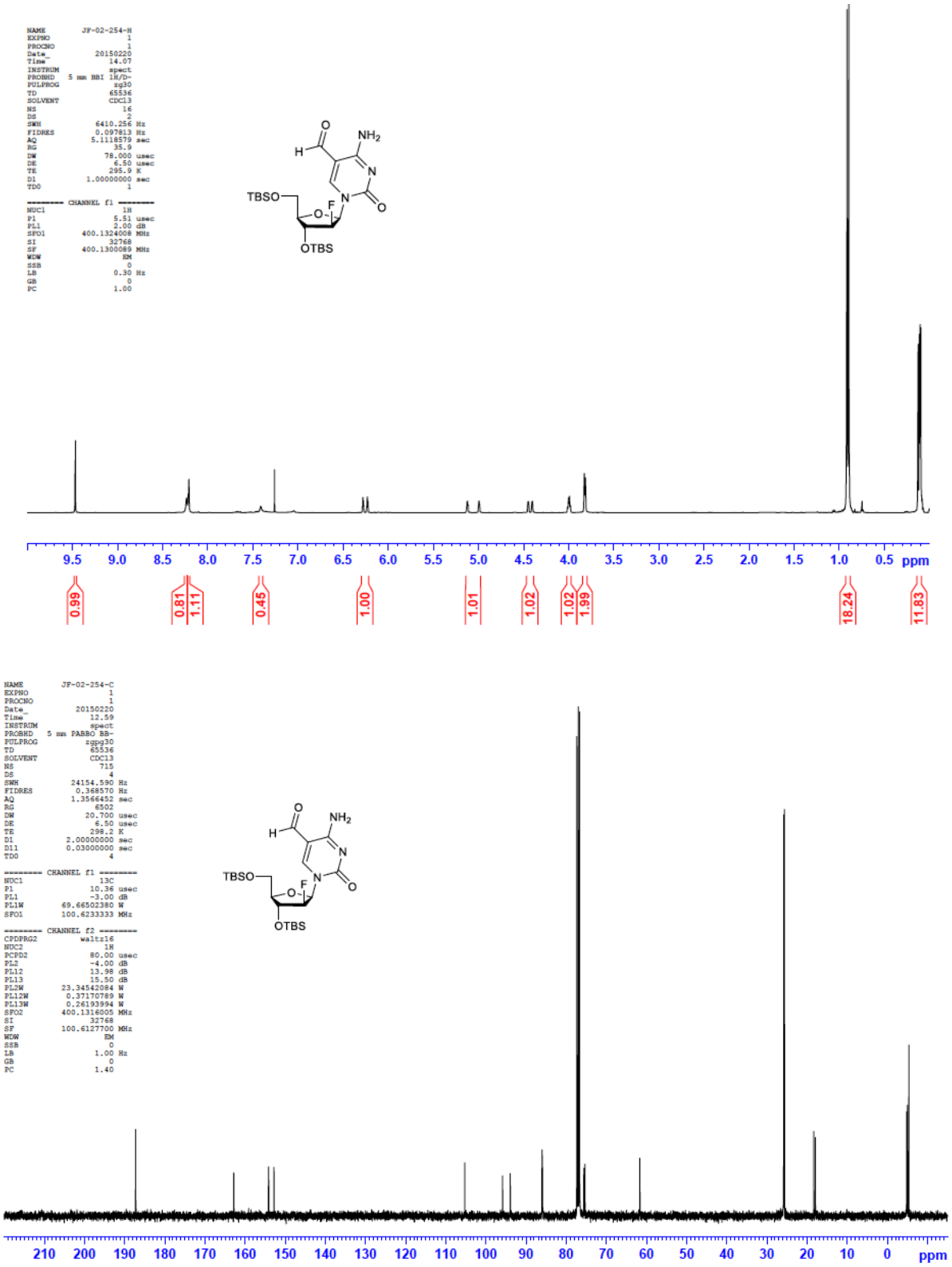


Figure S5. ^{1}H and ^{13}C NMR spectra of **S2**.

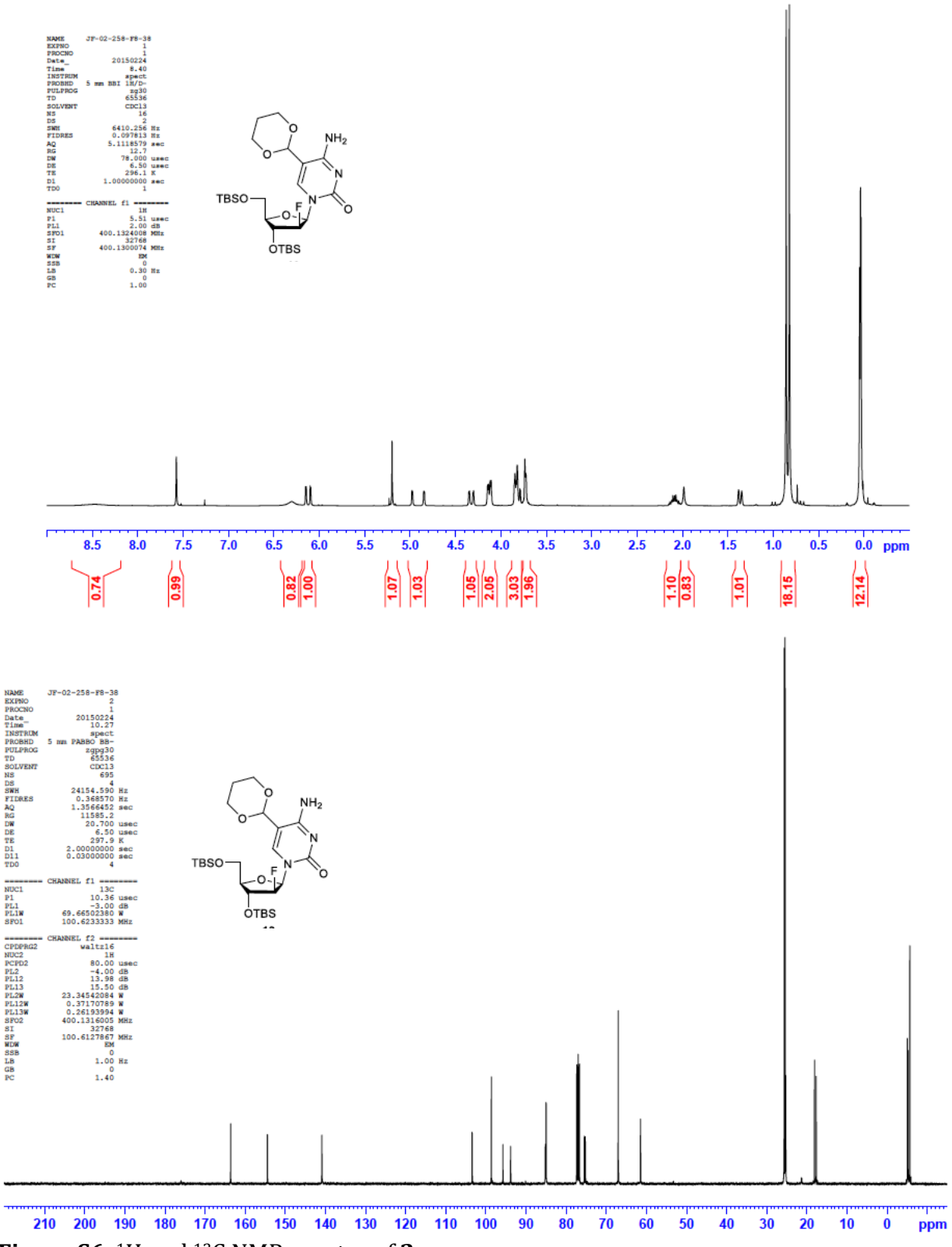


Figure S6. ^{1}H and ^{13}C NMR spectra of **2**.

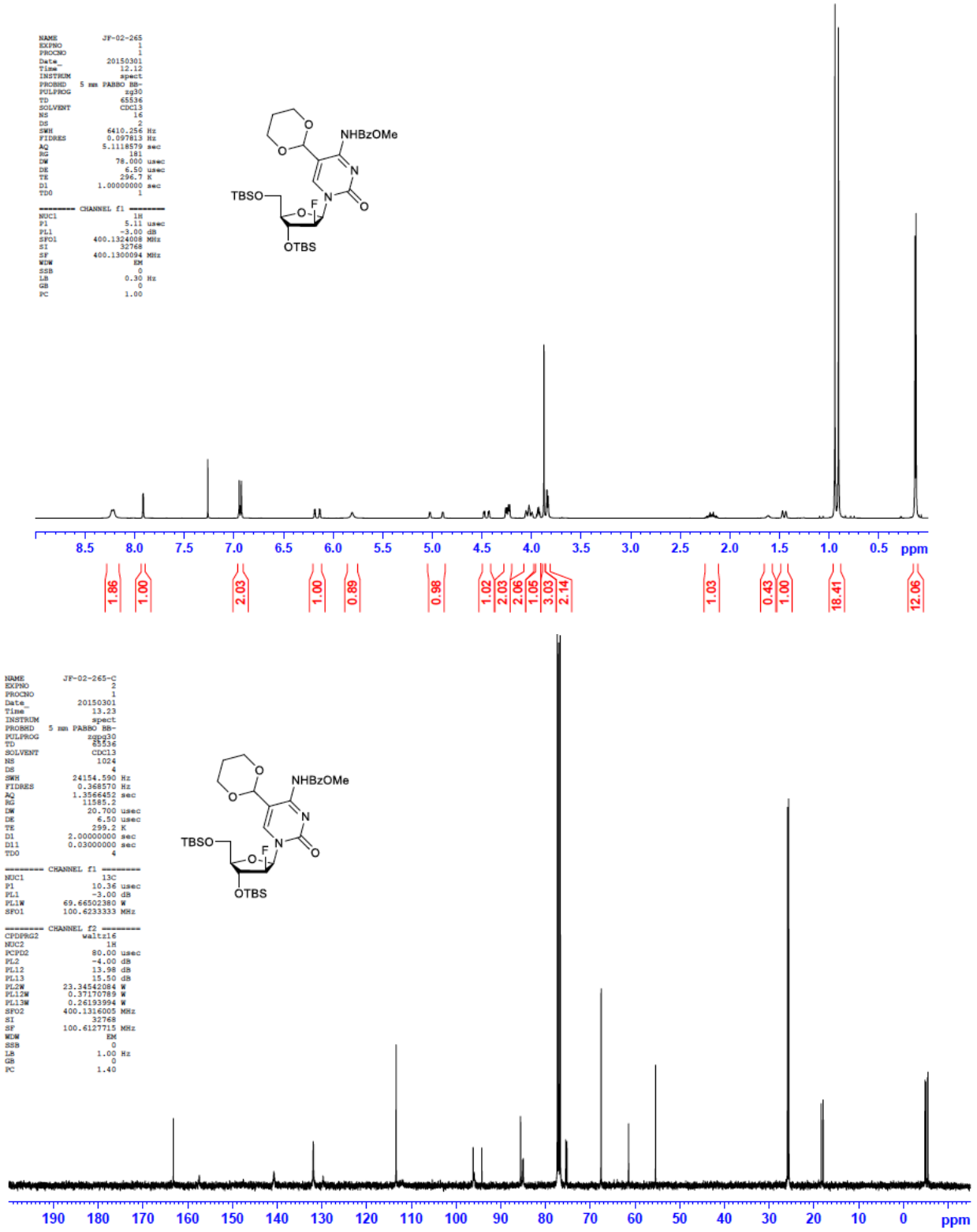


Figure S7. ^{1}H and ^{13}C NMR spectra of **S3**.

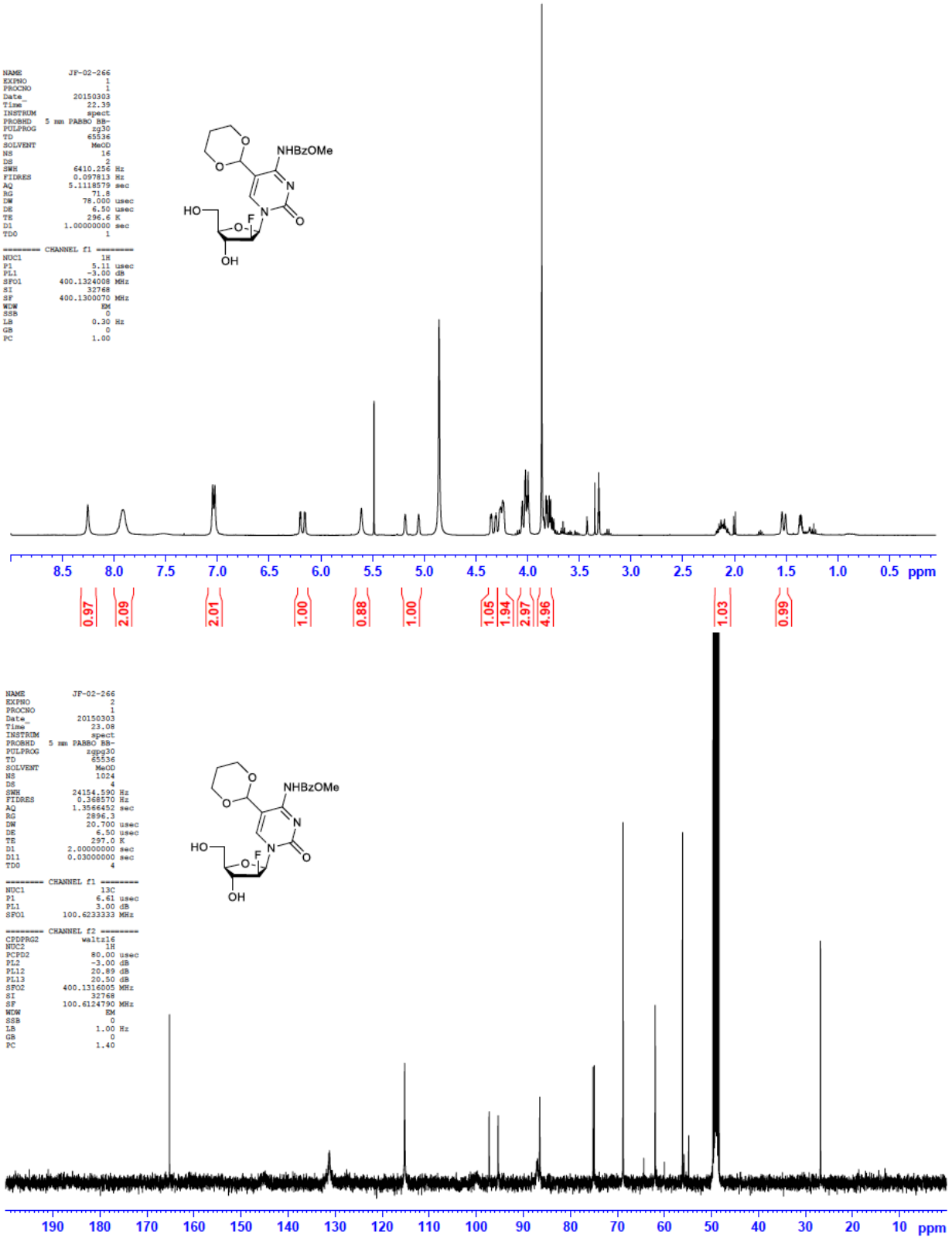


Figure S8. ^{1}H and ^{13}C NMR spectra of **3**.

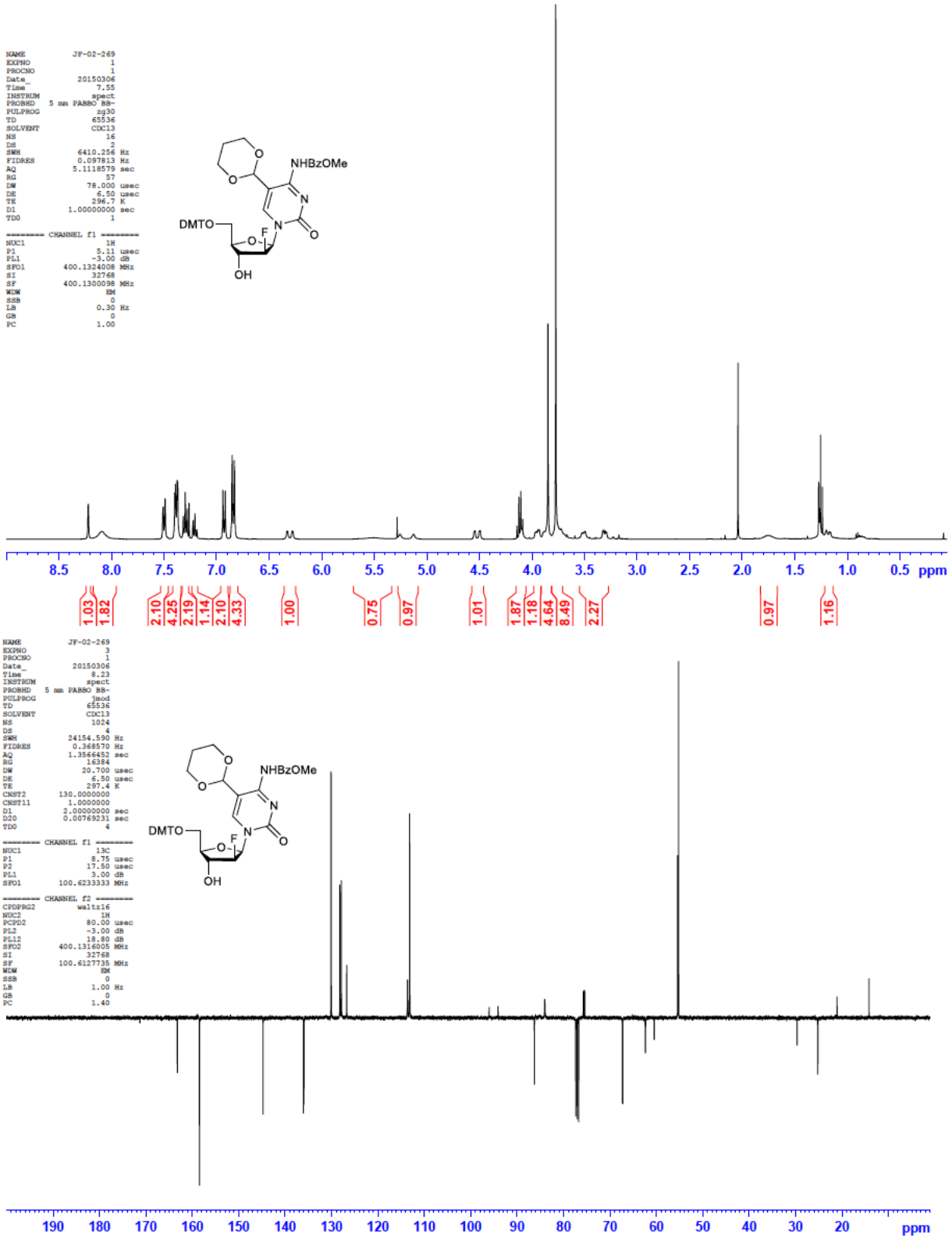


Figure S9. ^{1}H and ^{13}C NMR spectra of **S4**.

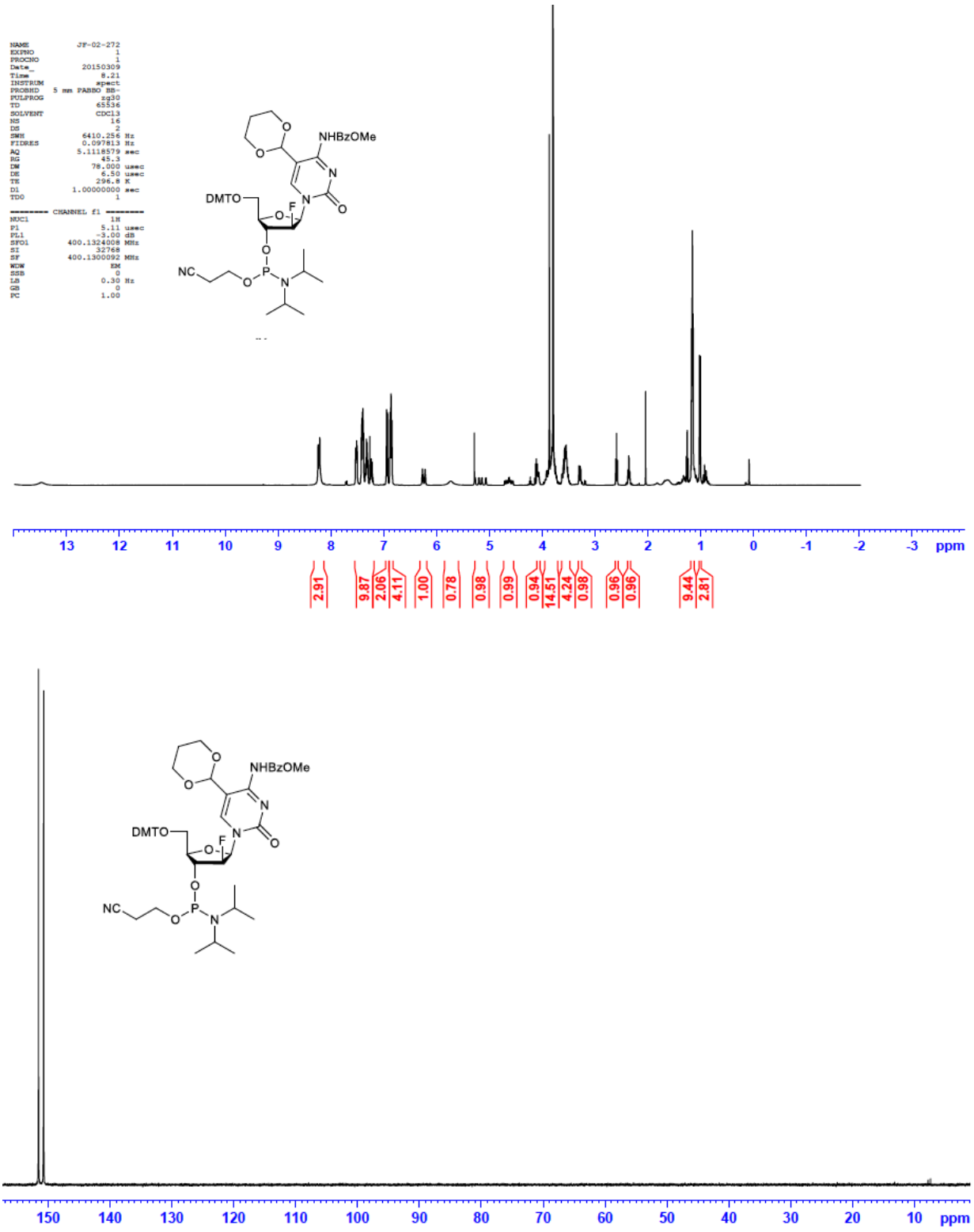


Figure S10. ^{1}H and ^{31}P NMR spectra of **4**.

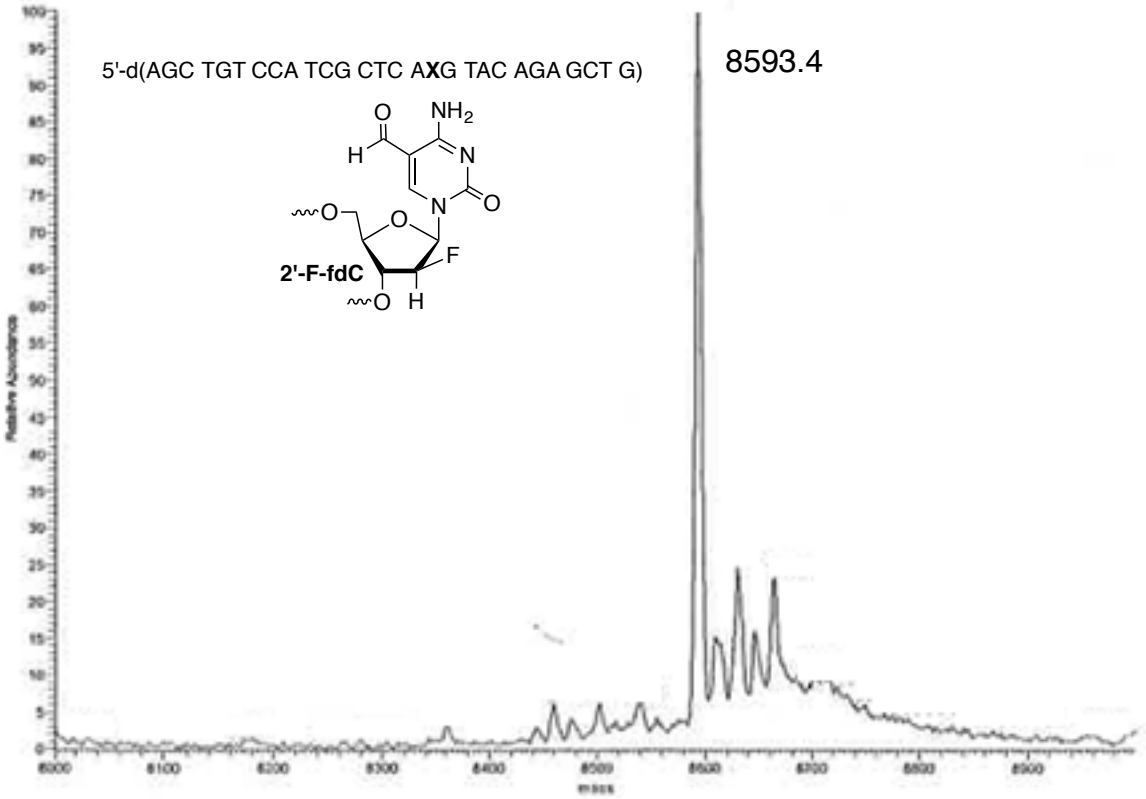


Figure S11. ESI-MS of oligonucleotide containing 2'-F-fdC (Calculated *m/z*: 8594.4).

References

(1) Malik, S. S., Coey, C. T., Varney, K. M., Pozharski, E., and Drohat, A. C. (2015) *Nucleic Acids Res 43*, 9541-9552.

(2) Morgan, M. T., Bennett, M. T., and Drohat, A. C. (2007) *J Biol Chem 282*, 27578-27586.

(3) Waters, T. R., and Swann, P. F. (1998) *J Biol Chem 273*, 20007-20014.

(4) Maiti, A., Morgan, M. T., and Drohat, A. C. (2009) *J Biol Chem 284*, 36680-36688.

(5) Coey, C. T., Malik, S. S., Pidugu, L. S., Varney, K. M., Pozharski, E., and Drohat, A. C. (2016) *Nucleic Acids Res,* doi: 10.1093/nar/gkw768.

(6) Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319-326.

(7) Morgan, M. T., Maiti, A., Fitzgerald, M. E., and Drohat, A. C. (2011) *Nucleic Acids Res 39*, 2319-2329.

(8) Maiti, A., Morgan, M. T., Pozharski, E., and Drohat, A. C. (2008) *Proc Natl Acad Sci USA 105*, 8890-8895.

(9) Kabsch, W. (2010) *Acta Crystallogr D Biol Crystallogr 66*, 125-132.

(10) Evans, P. R. (2011) *Acta Crystallogr D Biol Crystallogr 67*, 282-292.

(11) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) *Acta Crystallogr D Biol Crystallogr 67*, 235-242.

(12) Karplus, P. A., and Diederichs, K. (2012) *Science 336*, 1030-1033.

(13) McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) *Acta Crystallographica Section D-Biological Crystallography 61*, 458-464.

(14) Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Rovers, i. P., Sharff, A., Smart, O. S., Vonrhein, C., and Womack, T. O. (2011) BUSTER version 2.10.2, Global Phasing Ltd., Cambridge, United Kingdom.

(15) Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallographica Section D-Biological Crystallography 57*, 122-133.

(16) Emsley, P., and Cowtan, K. (2004) *Acta Crystallographica Section D-Biological Crystallography 60*, 2126-2132.

(17) Painter, J., and Merritt, E. A. (2006) *Acta Crystallogr D Biol Crystallogr 62*, 439-450.

(18) Painter, J., and Merritt, E. A. (2006) *Journal of Applied Crystallography 39*, 109-111.

(19) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *Acta Crystallographica Section D 66*, 12-21.

(20) Wilds, C. J., and Damha, M. J. (2000) *Nucl. Acids. Res. 28*, 3625-3635.

(21) Dai, Q., Lu, X., Zhang, L., and He, C. (2012) *Tetrahedron 68*, 5145-5151.

(22) Schroder, A. S., Steinbacher, J., Steigenberger, B., Gnerlich, F. A., Schiesser, S., Pfaffeneder, T., and Carell, T. (2014) *Angew Chem Int Ed Engl 53*, 315-318.