Supplementary information

for

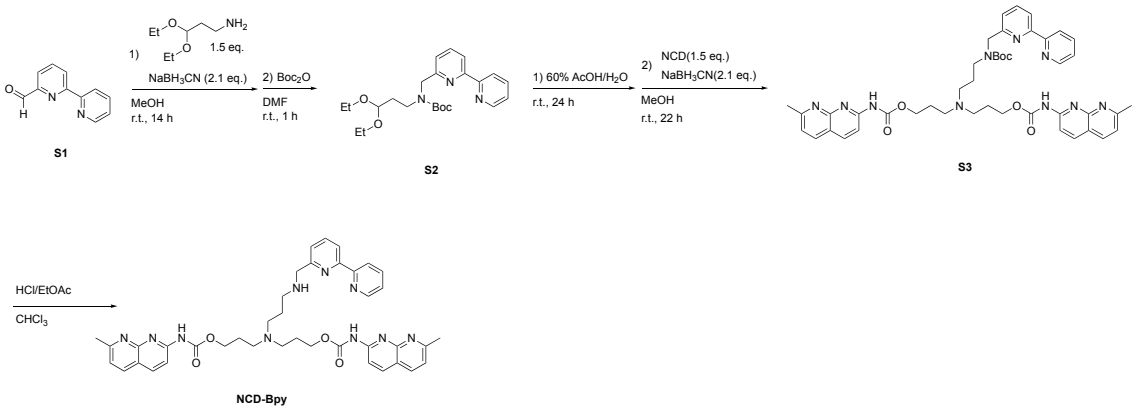
Chemical Probing of Thymine in the TGG/CGG Triad exploring the Deamination of 5-Methylcytosine in the CGG Repeat

Yuki Yagi, Takeshi Yamada, and Kazuhiko Nakatani*

Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (ISIR), Osaka University, Mihoga-oka, Ibaraki 567-0047, Japan

Contents:

Scheme S1. An overview of synthetic scheme of NCD-Bpy:



***tert*-Butyl ([2,2'-bipyridin]-6-ylmethyl)(3,3-diethoxypropyl)carbamate S2**

Synthesis of the starting material 2,2'-bipyridine-6-carbaldehyde **S1** was prepared according to the previously reported procedure.[1] To a solution of **S1** (124 mg, 0.673 mmol) in methanol (15 mL), was added 3,3-diethoxypropane-1-amine (149 mg, 1.01 mmol) and sodium cyanoborohydride (88.8 mg, 1.41 mmol), then the mixture was stirred at room temperature for 14 h. Solvent was evaporated to dryness. The resulting material was used in next reaction without purification. To a solution of the crude product obtained in the previous step in *N,N*-dimethylformamide (DMF) (10 mL) was added di-*tert*-butyl dicarbonate (Boc_2O) (150 mg), then the mixture was stirred at room temperature for 1 h. Water (10mL) and ethyl acetate (EtOAc) (30 mL) were added and the aqueous phase was washed with EtOAc (3x20 mL) and the combined organic phase was dried over magnesium sulfate ($MgSO_4$) and concentrated. The crude product was purified by silica gel column chromatography (EtOAc/Hexane = 1/4) to give **S2** (95 mg, 0.23 mmol, 34% in 2 steps) as a colorless oil.

^{1}H-NMR (600 MHz, $CDCl_3$): δ 8.67 (d, J = 4.8 Hz, 1H, Ar-H), 8.43 (d, J = 8.2 Hz, 1H, Ar-H), 8.27 (d, J = 7.6 Hz, 1H, Ar-H), 7.82-7.77 (2H, Ar-H), 7.30 (t, 1H, Ar-H), 7.24 (br, 1H, Ar-H), 4.65 and 4.59 (brs, CH_2-Ar), 4.57 and 4.49 (brt, C*H*CH_2CH_2), 3.62 and 3.47(brm, CH_3CH_2), 3.47 and 3.36 (brt, $CHCH_2CH_2$), 1.96 and 1.91 (brm, $CHCH_2CH_2$), 1.52 and 1.37 (brs, $C(CH_3)_3$), 1.17 (t, 6H, CH_3CH_2).

^{13}C-NMR (151 MHz, $CDCl_3$) δ 158.20, 157.84, 156.25, 156.02, 155.66, 155.49, 149.11, 137.58, 137.41, 136.97, 136.79, 123.68, 123.62, 121.58, 121.21, 120.76, 119.31, 119.14, 101.12, 100.88, 79.84, 79.73, 61.07, 60.91, 53.43, 53.21, 52.64, 44.20, 43.88, 32.49, 32.29, 29.71, 28.48, 28.35, 15.30

FT-MS calculated for $C_{23}H_{33}N_3O_4$ [$(M + Na)^+$] 438.2363, found 438.2352

((3-(([2,2'-Bipyridin]-6-ylmethyl)(*tert*-butoxycarbonyl)amino)propyl)azanediyl)bis(propane-3,1-diyl) bis((7-methyl-1,8-naphthyridin-2-yl)carbamate) S3

S2 (60 mg, 0.144 mmol) was dissolved in 60% aqueous acetic acid (3 mL) and the mixture was stirred at room temperature for 23 h. Then the solution was neutralized with saturated aqueous sodium hydrogen carbonate solution (aq. $NaHCO_3$) (20 mL) and $CHCl_3$ (20 mL) was added. The organic phase was separated

and the aqueous phase was washed with $CHCl_3$ (3x20 mL). The combined organic phase was dried over $MgSO_4$ and concentrated. The crude product was used in the next step without further purification. The crude product (50 mg) was dissolved in MeOH (5 mL). To this solution, azanediylbis(propane-3,1-diyl) bis((7-methyl-1,8-naphthyridin-2-yl)carbamate) (naphthyridine carbamate dimer = **NCD)** (110.7 mg, 0.220 mmol)[2] and sodium cyanoborohydride ($NaBH_3CN$) (19.3 mg, 0.308 mmol) were added. After the reaction mixture was stirred at room temperature for 22 h, the mixture was concentrated and purified using silica gel column chromatography (MeOH/$CHCl_3$ = 1/9) and amino silica gel chromatography (100% $CHCl_3$) successively to give **S3** (30.5 mg, 25% in two steps).

^{1}H-NMR (600 MHz, $CDCl_3$) δ 8.66 (d, J = 4.8 Hz, 1H, Ar-H), 8.43 (d, J = 8.2 Hz, 1H, Ar-H), 8.27 (3H, Ar-H), 8.11 (d, J = 8.2 Hz, 2H, Ar-H), 7.89 (br, 2H, N*H*), 7.76-7.81 (2H, Ar-H), 7.21-7.26 (Integral is likely to be 2H, Ar-H, peaks were identified by 2D-NMR), 7.24 (d, J = 8.2 Hz, 2H, Ar-H), 4.66 and 4.60 (brs, C*H*$_2$-Ar, 2H), 4.25 (br, 4H, OC*H*$_2$CH$_2$CH$_2$N), 3.44 and 3.31 (brt, 2H, NCH$_2$CH$_2$C*H*$_2$NCH$_2$), 2.75 (s, 6H, C*H*$_3$), 2.50 (br, 4H, OCH$_2$CH$_2$C*H*$_2$N), 2.45 and 2.41 (brt, 2H, NC*H*$_2$CH$_2$CH$_2$NCH$_2$), 1.81 (br, 4H, OCH$_2$C*H*$_2$CH$_2$N), 1.76 and 1.70 (br, 2H, NCH$_2$C*H*$_2$CH$_2$NCH$_2$), 1.52 and 1.38 (brs, 9H, C(C*H*$_3$)$_3$). ^{13}C-NMR (176 MHz, $CDCl_3$) δ 163.02, 158.19, 157.92, 156.14, 156.02, 155.74, 155.42, 155.35, 154.64, 153.47, 153.20, 149.09, 139.00, 137.60, 137.43, 136.92, 136.83, 136.38, 123.69, 121.19, 120.81, 119.40, 119.20, 117.94, 112.74, 79.77, 77.23, 77.04, 76.86, 64.08, 52.93, 52.59, 51.49, 51.42, 50.01, 46.21, 46.14, 28.53, 28.34, 26.55, 25.59

FT-MS calculated for $C_{45}H_{52}N_{10}O_6$ [(M + H)$^+$] 829.4144, found 829.4139

((3-(([2,2'-Bipyridin]-6-ylmethyl)amino)propyl)azanediyl)bis(propane-3,1-diyl) bis((7-methyl-1,8-naphthyridin-2-yl)carbamate) NCD-Bpy

S3 (30.5 mg, 0.0368 mmol) was dissolved in chloroform ($CHCl_3$) (1.5 mL) then 4 N-hydrogen chloride in ethyl acetate (HCl/EtOAc) was slowly added and the reaction mixture was stirred at room temperature for 30 min. Solvent was removed by evaporator and the resulting residue was dried in vacuo to give **NCD-Bpy** (20.1 mg, 66%) as a white solid.

^{1}H-NMR (600 MHz, D_2O) δ 8.88 (d, J = 5.5 Hz, 1H, Ar-H), 8.71 (m, 2H, Ar-H), 8.64 (d, J = 8.2 Hz, 2H, Ar-H), 8.32-8.35 (m, 3H, Ar-H), 8.16 (t, J = 7.9 Hz, 1H, Ar-H), 8.09-8.11 (m, 3H, Ar-H), 7.72 (d, J = 8.2 Hz, 1H, Ar-H), 7.66 (d, J = 8.2 Hz, 2H, Ar-H), 4.65 (s, 2H, NC*H*$_2$), 4.40 (t, J = 5.8 Hz, 4H, OC*H*$_2$CH$_2$CH$_2$N), 3.51 (m, 4H, OCH$_2$CH$_2$C*H*$_2$N), 3.45 (t, J = 7.9 Hz, 2H, NC*H*$_2$CH$_2$CH$_2$NCH$_2$), 3.41 (t, J=7.9 Hz, 2H, NCH$_2$CH$_2$C*H*$_2$NCH$_2$) 2.89 (s, 6H, C*H*$_3$), 2.42 (m, 2H, NCH$_2$C*H*$_2$CH$_2$NCH$_2$), 2.28 (m, 4H, OCH$_2$C*H*$_2$CH$_2$N). ^{13}C-NMR (176 MHz, D_2O) δ 176.55, 159.14, 156.64, 153.49, 151.70, 147.73, 147.00, 146.51, 145.97, 145.45, 141.46, 140.18, 139.90, 127.24, 125.90, 124.25, 122.37, 122.00, 119.03, 116.01, 62.82, 50.44, 50.40, 49.60, 44.28, 22.69, 20.85, 19.91, 19.87

FT-MS calculated for $C_{40}H_{44}N_{10}O_4$ [(M + 2H)$^{2+}$] 365.1852, found 365.1844

(1) Nakatani, K.; Hagihara, S.; Sando, S.; Miyazaki, H.; Tanabe, K.; Saito, I. *J. Am. Chem. Soc.* **2000**, *122*, 6309–6310.

(2) Nakatani K.; He H.; Uno, S.; Yamamoto T.; Dohno, C. *Curr. Protoc. Nucleic Acid Chem.* **2008**, unit 8.6, 1–21

Sequence list of oligo deoxy nucleotides

T1: 5′-d(CGCA TGG TCGC)-3′

C1: 5′-d(GCGA CGG TGCG)-3′

^{5m}C1: 5′-d(AAG CGA ^{5m}CGG TGC GAA)-3′

C2: 5′-d(CGCA CGG TCGC)-3′

T3: 5′-d(CTAA GGT AATG)-3′

C4: 5′-d(CATT GGC TTAT)-3′

CGG1: 5′-d(CTA CGG CGG CGG CGG CGG CGG CGG **T**GG CGG CGG TAG)-3′

CGG2: 5′-d(CTA CGG CGG CGG CGG CGG CGG **T**GG CGG CGG CGG TAG)-3′

M1: 5′-d(CTA CGG CGG CGG CGG CGG CGG CGG)-3′

M2: 5′-d(GG CGG CGG TAG)-3′

M3: 5′-d(CTA CGG CGG CGG CGG CGG CGG)-3′

M4: 5′-d(GG CGG CGG CGG TAG)-3′

T_m measurements:

Thermal denaturation profiles were recorded on a UV-2700 spectrophotometer (Shimadzu) equipped with a TMSPC-8 temperature controller and a 10 mm path-length cell. The absorbance of DNA duplex (5 μM) with ligand (20 μM) in sodium cacodylate (10 mM, pH 7.0) containing sodium chloride (100 mM) was monitored at 260 nm from 2 to 100 °C (1 °C·min^{-1}). T_m was calculated by using the median method.

General procedure of sample preparation was as follows. To **NCD-Bpy** dissolved in distilled water in tube, oligo DNA(s) was added and mixed vigorously. To the mixture, pH7 sodium cacodylate buffer and NaCl were added and mixed vigorously. The final total volume was 130 μL.

RP-HPLC analysis:

The reaction of **NCD-Bpy** with origo DNA was monitored by reversed phase HPLC equipped with a Cosmosil 5C_{18}-MS-II column (150 mm), jasco PU-4180 pump, jasco CO-4060 column oven and jasco MD-4014 PDA detector by a gradient elution of 0.1 M triethylamine acetate solution (TEAA) (pH 7.0) with increasing acetonitrile (2–12% over 20 min then 12%–32% over 20 min). The column oven temperature was set at 40 ºC with the elution rate at 1 mL/min, monitored by UV absorbance at 260 nm.

Time course oxidation of thymine in T1/C1 duplex:

The selective oxidation reaction of thymine in **T1**/**C1** duplex was performed at 4 ºC in a solution containing tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH7.5, 10 mM), sodium chloride (NaCl) (100 mM), 5'-d(CGCA TGG TCGC)-3' (**T1**) (5 μM), 5'-d(GCGA CGG TGCG)-3' (**C1**) (5 μM), **NCD-Bpy (**20 μM), aqueous osmium tetroxide (OsO_4) (1 mM) and deoxyadenosine (dA) as internal standard. 10 μL aliquot was analyzed by reversed phase HPLC (RP-HPLC) after 10 min, 90 min, 5 h, 24 h incubation. As a control experiment, a mixture of **NCD** (20 μM) and 2,2'-bipyridine (**Bpy**) (20 μM) were used instead of **NCD-Bpy (**20 μM).

Cleavage reaction at oxidation site and removal of phosphoric termini:

The solution after 24 h incubation for site-selective oxidation by OsO_4 was heated at 90 ºC for 30 min. The resulting product was isolated by RP-HPLC. (Figure 4a) The area indicated by bracket in Figure 4a was isolated and lyophilized. The lyophilized mixture was treated with piperidine (10% v/v, 90 ºC, 30 min), then an aliquot was analyzed by RP-HPLC. (Fig. 4b) Fragment peaks were isolated by RP-HPLC, lyophilized and analyzed using MALDI-TOFMS. (ultraflex III, BRUKER) 3-hydroxypicolinic acid and diammonium hydrogen citrate (50:1) dissolved in 50/50 acetonitrile/water were used as a matrix.

Another aliquot of piperidine treated fragment oligo DNA was lyophilized and treated by alkaline phosphatase (rSAP, New England Biolabs) to remove phosphoric termini according to the instruction by the manufacturer, then analyzed by RP-HPLC. (Fig. 4c) Obtained fragment oligo DNA peaks were isolated and lyophilized, then analyzed using MALDI-TOFMS. The identity of the products was further confirmed by co-elution of the products with the authentic oligo DNA (5'-d(CGCA)-3' and 5'-d(GG TCGC)-3'). (Fig 4d) Overall scheme was summarized in scheme 2.

The Fraction of T1 strand remained (Figure 2c)

time	CGG+Bpy+NCD	S.D.	CGG+NCD-Bpy	S.D.
0	1	0	1	0
0.166	1.25	0.019	1.02	0.035
1.5	1.18	0.048	0.92	0.031
5	1.12	0.030	0.76	0.043
24	1.35	0.120	0.48	0.043

Full range chart for the site selective oxidation of T1/C1 duplex with NCD-Bpy+OsO4 and NCD+Bpy+OsO4

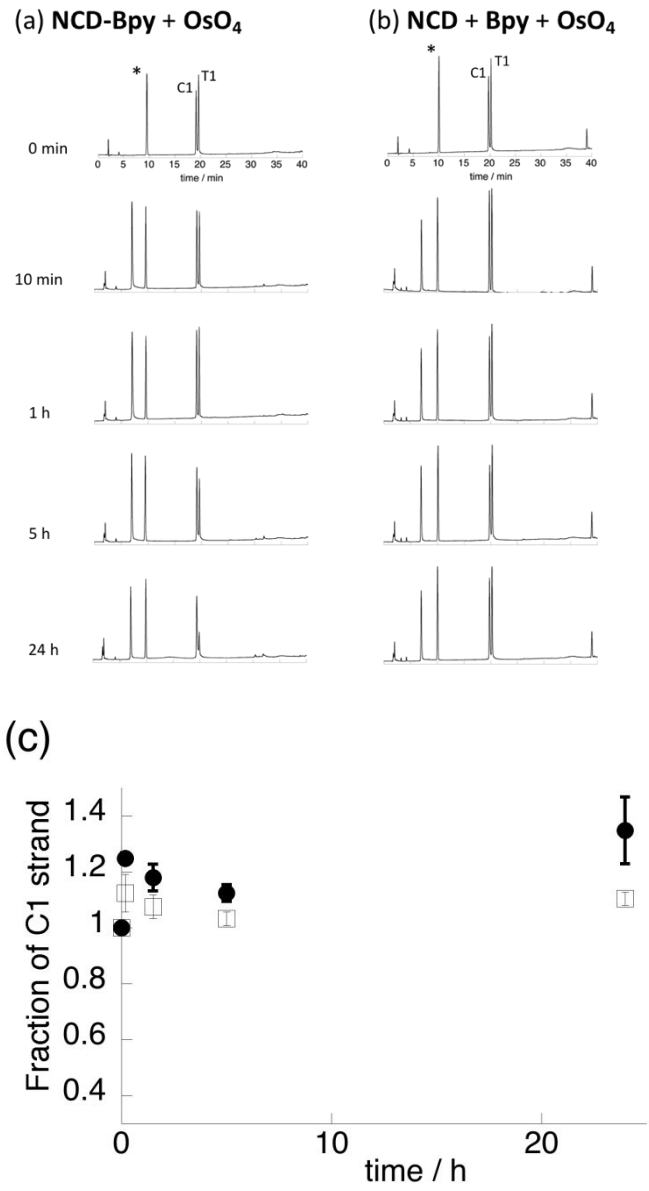


Figure S1. RP-HPLC profiles with a full range (0–40 min) for the oxidation of a duplex of **T1**/**C1** (5 μM each) after 0 min, 10 min, 90 min, 5 h, and 24 h in the presence of (a) **NCD-Bpy** (20 μM) and OsO_4 (1 mM) and (b) **NCD** (20 μM), 2,2'-bipyridine (**Bpy**) (20μM) and OsO_4 (1 mM), in Tris-HCl (pH 7.5, 10 mM) and NaCl (100 mM). dA marked as asterisk (*) was added as an internal standard. (c) Time course for the fraction of **C1** strand with **NCD-Bpy** (20 μM) (filled circle) and a mixture of **NCD** (20 μM) and 2,2'-bipyridine (20 μM) (open square), and OsO_4 (1 mM). The fraction of remaining **C1** strand was normalized to the amount at 0 min. Error bars represent 1 s.d. of the mean. The peak at around 5 – 7 min is from OsO_4.

Attempt of oxidation of T in the G-T mismatch within GGT/GGC triad

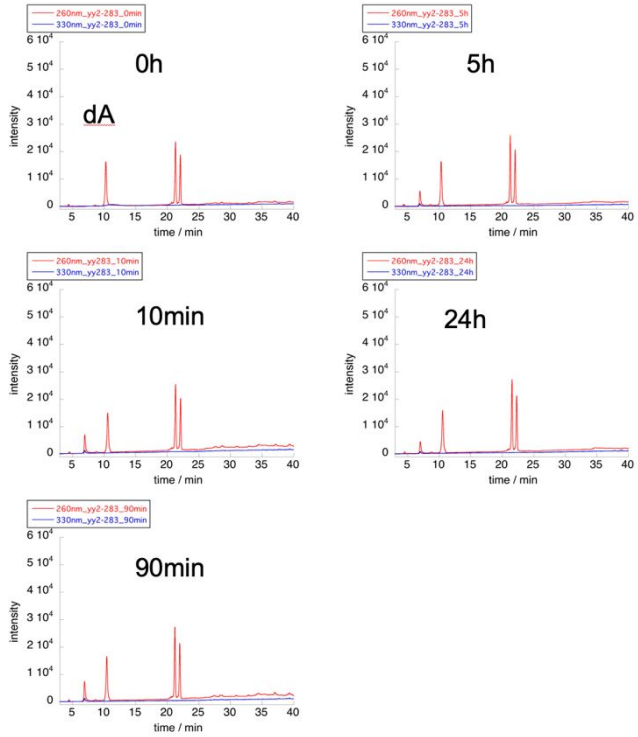


Figure S2. RP-HPLC profiles with a full range (0–40 min) for the oxidation of a duplex of **T3**/**C4** (5 μM each) after 0 min, 10 min, 90 min, 5 h, and 24 h in the presence of **NCD-Bpy** (20 μM) and OsO_4 (1 mM) in Tris-HCl (pH 7.5, 10 mM) and NaCl (100 mM). dA marked was added as an internal standard.

The amount of both strands did not change after 24 h incubation.

Thermal melting profile of ^{5m}C1/C2 duplex:

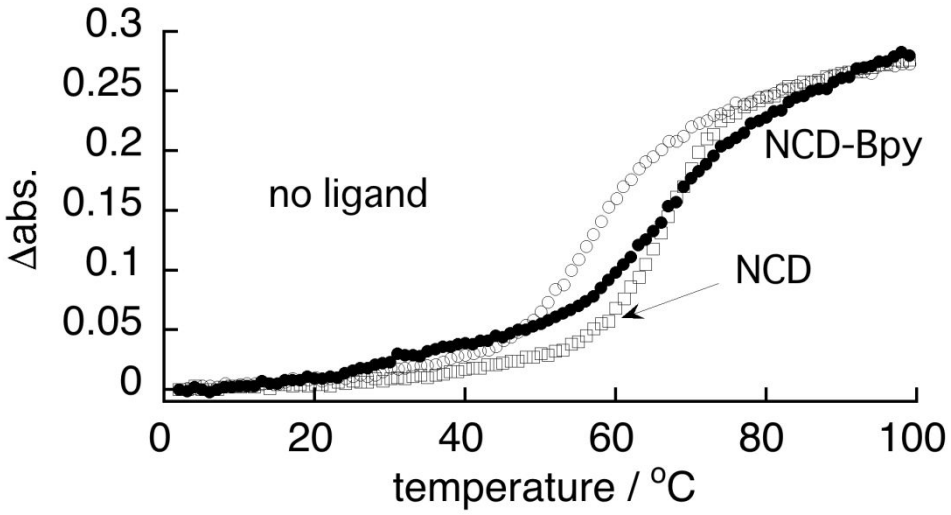


Figure S3. Thermal melting profiles of a DNA duplex of 5'-d(AAG CGA ^{5m}CGG TGC GAA)-3' /5'-d(CGCA CGG TCGC)-3' (**^{5m}C1/C2**) containing ^{5m}CGG/CGG site in the absence (open circle), presence of **NCD-Bpy** (filled circle), and **NCD** (open square). Concentrations of DNA and the ligand were 5 μM and 20μM, respectively. The absorbance at 260 nm was measured in sodium cacodylate buffer (pH 7, 10 mM) containing 0.1 M NaCl.

NCD-Bpy provided T_m (64.6 ºC, ΔT_m = 8.5 ºC) and **NCD** provided T_m (65.6 ºC, ΔT_m = 9.5 ºC).

Although a protracted thermal melting profile with **NCD-Bpy** suggests that **NCD-Bpy** binds to **^{5m}C1/C2** duplex in a less cooperative manner than **NCD**, comparable T_m with **NCD-Bpy** to that with **NCD** suggests that Bpy moiety introduced to **NCD** doesn't strongly interfere the binding of **NCD** moiety to **^{5m}CGG/CGG** site.

Time course reaction of DNA duplex containing ^{5m}CGG/CGG with NCD-Bpy+OsO4

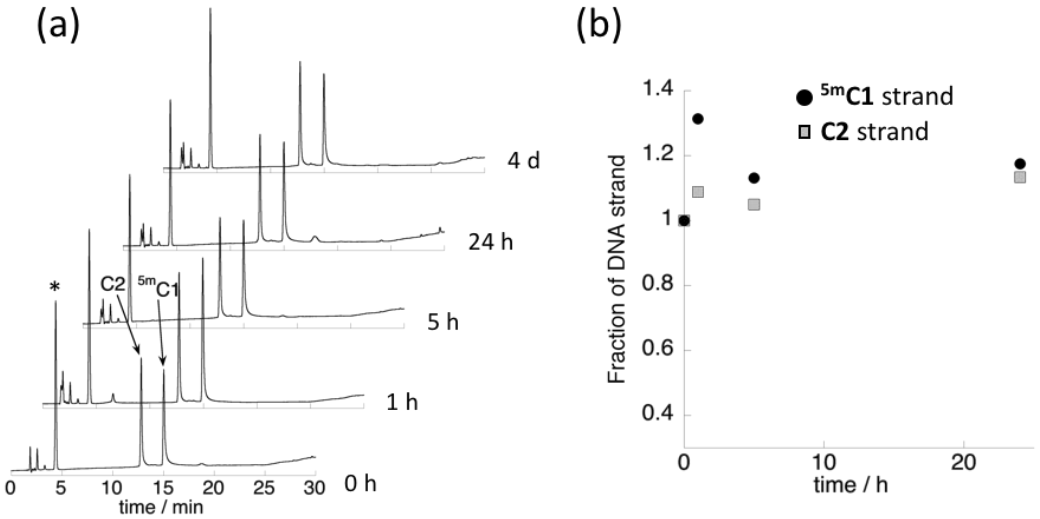


Figure S4. (a) RP-HPLC profiles for the oxidation of a duplex of 5'-d(AAG CGA ^{5m}CGG TGC GAA)-3'/5'-d(CGCA CGG TCGC)-3' (**^{5m}C1/C2**) (5 μM each) after 0 min, 60 min, 5 h, 24 h, and 4 d in the presence of **NCD-Bpy** (20 μM) and OsO_4 (1 mM) in Tris-HCl buffer (pH 7.5, 10 mM) and NaCl (100 mM). dA marked as an asterisk (*) was added as an internal standard.; incubation temperature was 37 ºC.
(b) Time course of the fraction of **^{5m}C1** strand (filled circle) and **C2** strand (gray square), with **NCD-Bpy** (20 μM) and OsO_4 (1 mM). The fraction of remaining **^{5m}C1** and **C2** were normalized to the amount at 0 min.

In this experiment, RP-HPLC analysis was performed with a gradient elution of 0.1 M TEAA (pH 7.0) with increasing acetonitrile (5–10% over 10 min, 10% over 4 min, 10–13% over 6 min, then 13–33% over 10 min). The column was heated at 55 ºC by column oven due to the higher melting temperature of **^{5m}C1**/**C2** duplex.

Site selective oxidation of thymine in CGG repeat:

Selective oxidation of T in the CGG repeat DNA was investigated two DNA **CGG1** and **CGG2**, which contained T in the different position as described **Figure S5** shown below.

Thymine containing CGG repeat DNA 5'-d(CTA CGG CGG CGG CGG CGG CGG CGG **T**GG CGG CGG TAG)-3'(**CGG1)** or 5'-d(CTA CGG CGG CGG CGG CGG CGG **T**GG CGG CGG CGG TAG)-3' (**CGG2**) (5 μM) was incubated with **NCD-Bpy** (40 μM) or a mixture of **NCD** (40 μM) and 2,2'-bipyridine (40 μM) at 4 ºC for 24 h in the presence of OsO_4 (1 mM) in Tris-HCl (pH 7.5, 10 mM) and NaCl (100 mM). After 24 h incubation, an aliquot was treated by hot piperidine (10% v/v, 90 ºC 30min). After oxidation reaction and piperidine treatment as above, resulting solution was lyophilized to remove remaining piperidine and phosphoric termini of the DNA fragments were removed by alkaline phosphatase (rSAP, New England Biolabs) according to the instruction by the manufacturer.

Figure S5

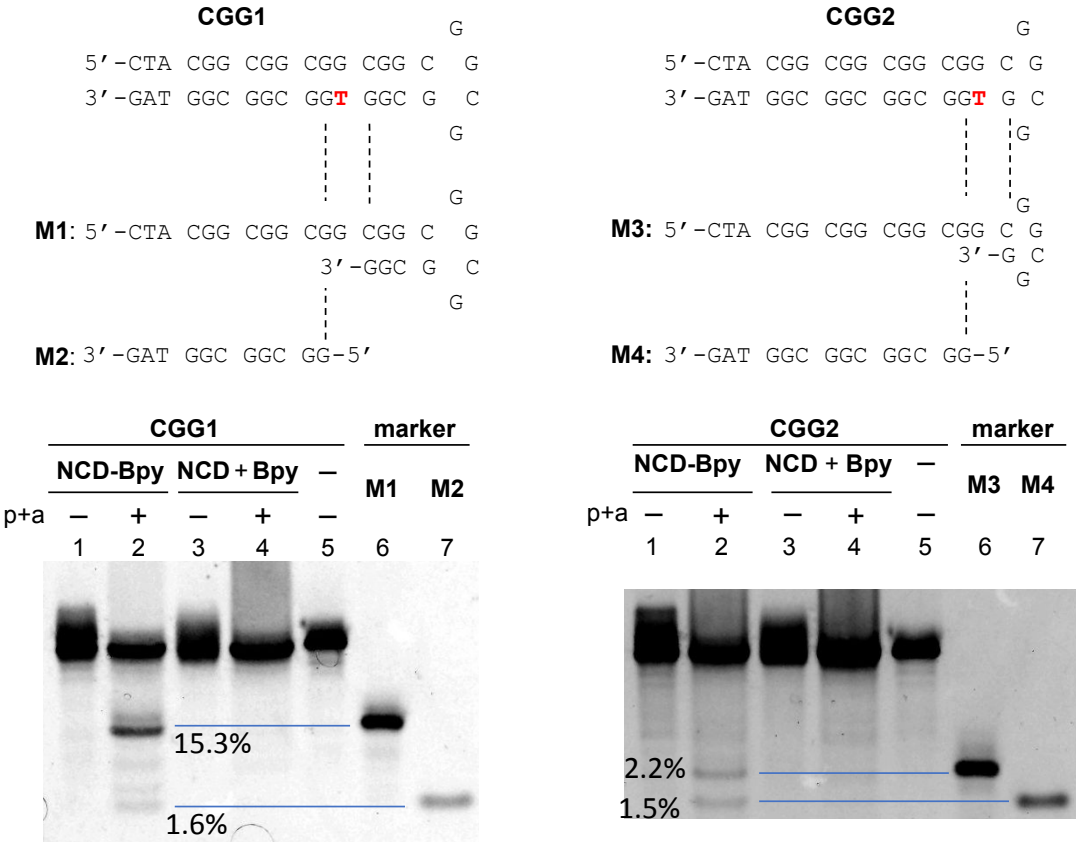


Figure 5 in the main text

Detailed explanation is in the caption of **Figure S6**

The number in the gel represent the fraction of the band determined by densitometry (NIH Image J) in the designated lane, and does not mean the cleavage efficiency.

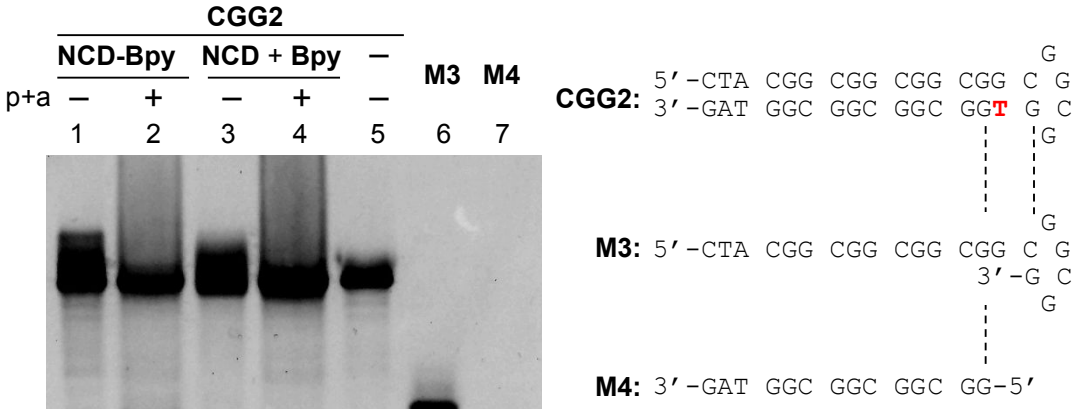
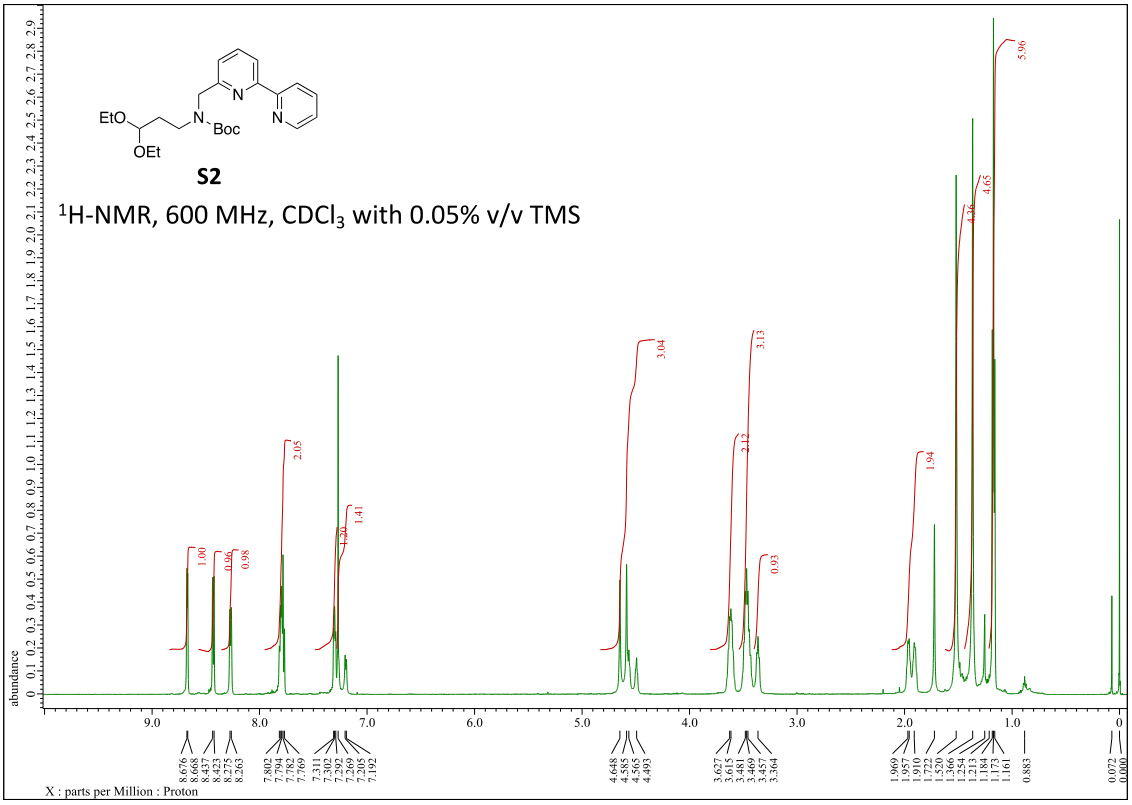
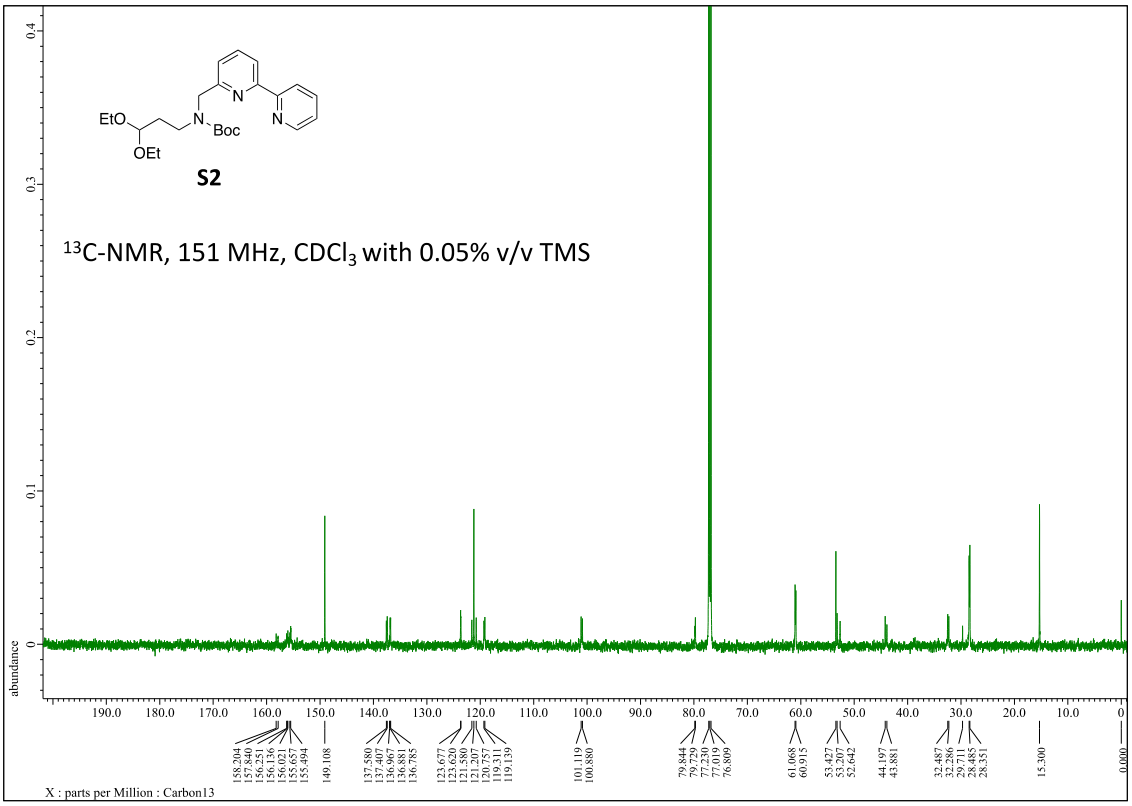


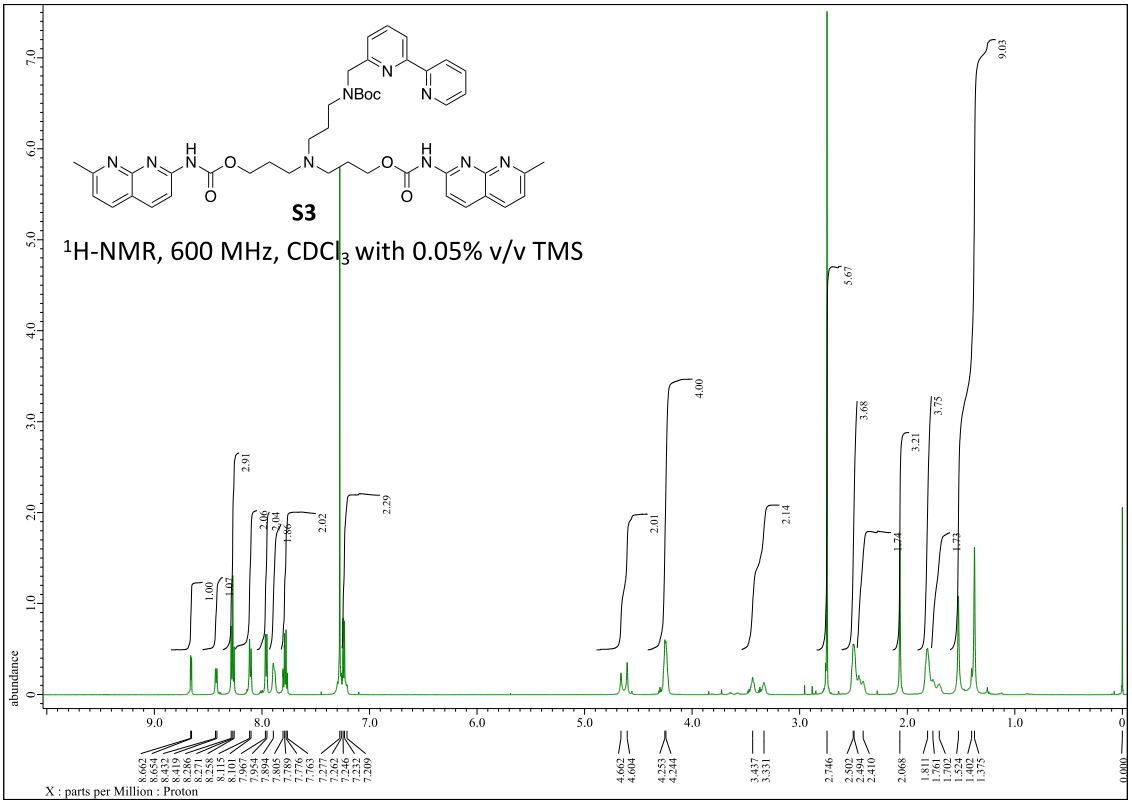
Figure S6. Denaturing PAGE (15% polyacrylamide/6M urea gel) to analyze the chemical modification of thymine in CGG repeat hairpin DNA. **CGG2** (5 μM) was incubated at 4 ºC for 24 h in the presence of **NCD-Bpy** (40 μM) and OsO_4 (1 mM) (lanes 1 and 2) or a mixture of **NCD** (40 μM) and 2,2'-bipyridine (40 μM) (lanes 3 and 4) in Tris-HCl (pH 7.5, 10 mM) and NaCl (100 mM), and then treated with hot piperidine (90 ºC, 30 min) and alkaline phosphatase (p+a) (lanes 2 and 4). All DNAs were stained by SYBR Gold.

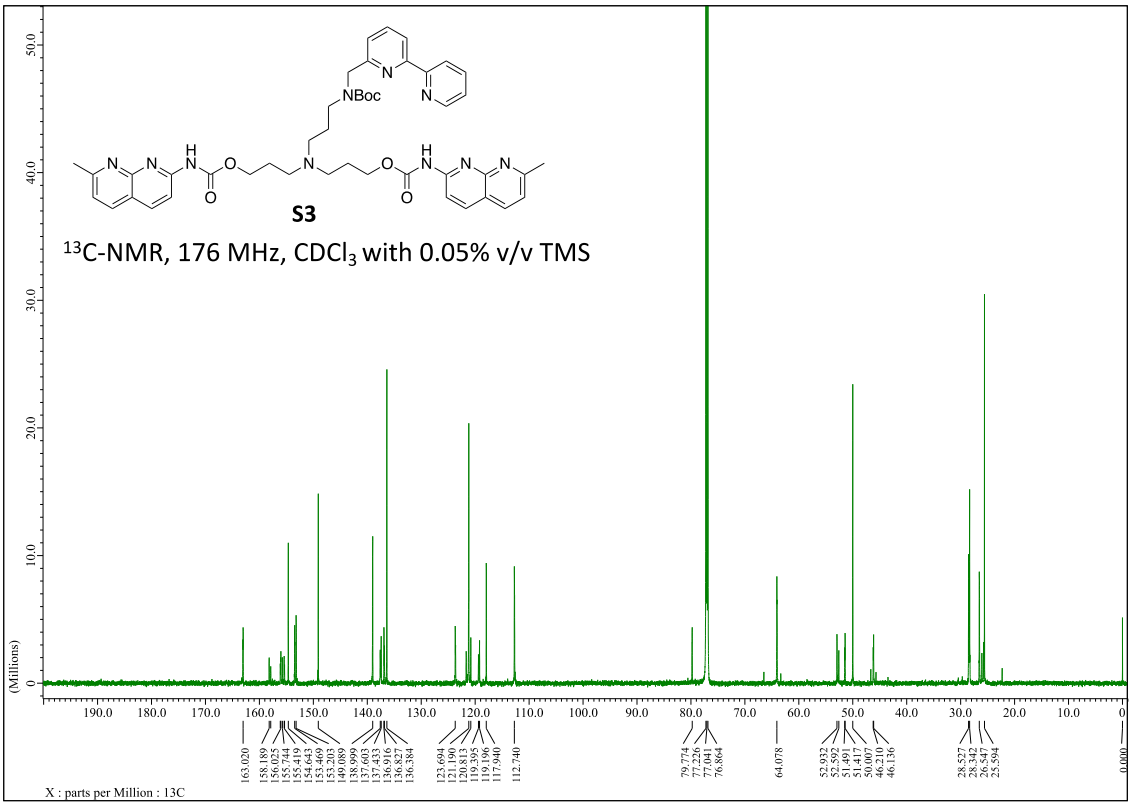
Lane 1, **CGG2**, **NCD-Bpy**, OsO_4; lane 2, **CGG2**, **NCD-Bpy**, OsO_4, then p+a; lane 3, **CGG2**, **NCD** + **Bpy**, OsO_4, lane 4, **CGG2**, **NCD** + **Bpy**, OsO_4, then p+a; lane 5, **CGG2**; lane 6, **M3**; lane7, **M4**.

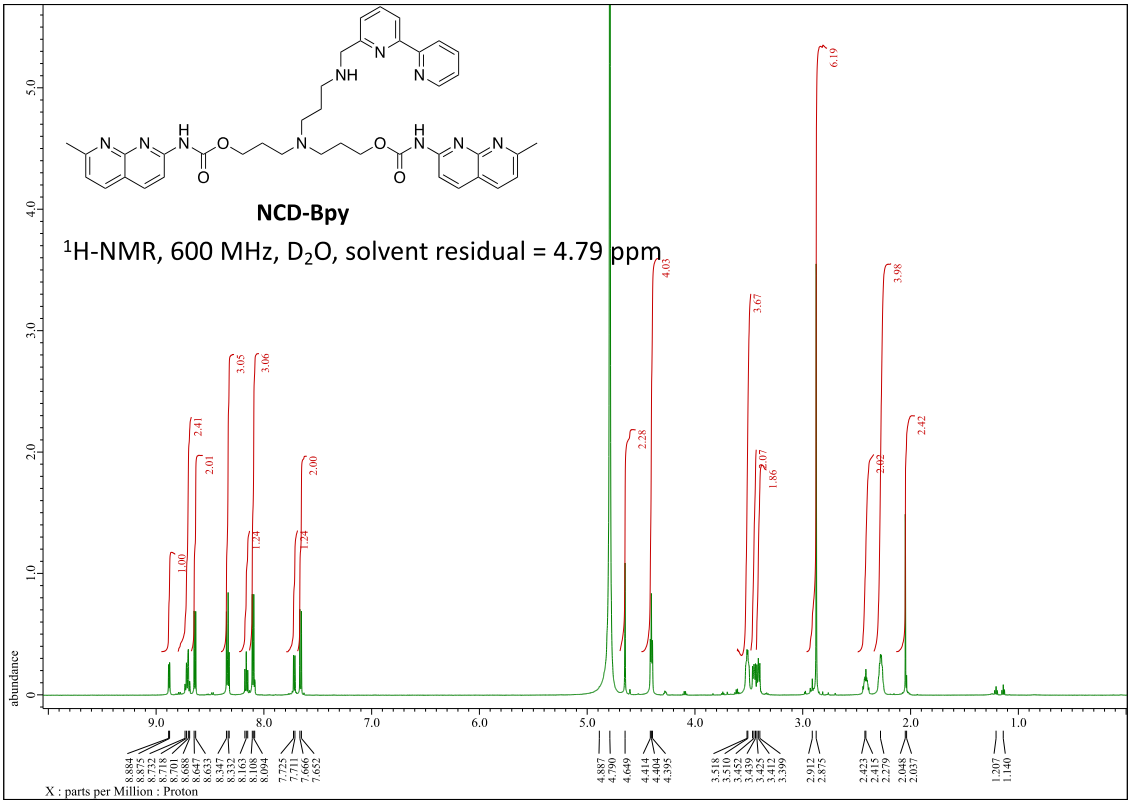
Fragment bands in lane 2 have the same mobility as **M3** and **M4**, which are produced by the cleavage at T in **CGG2**.

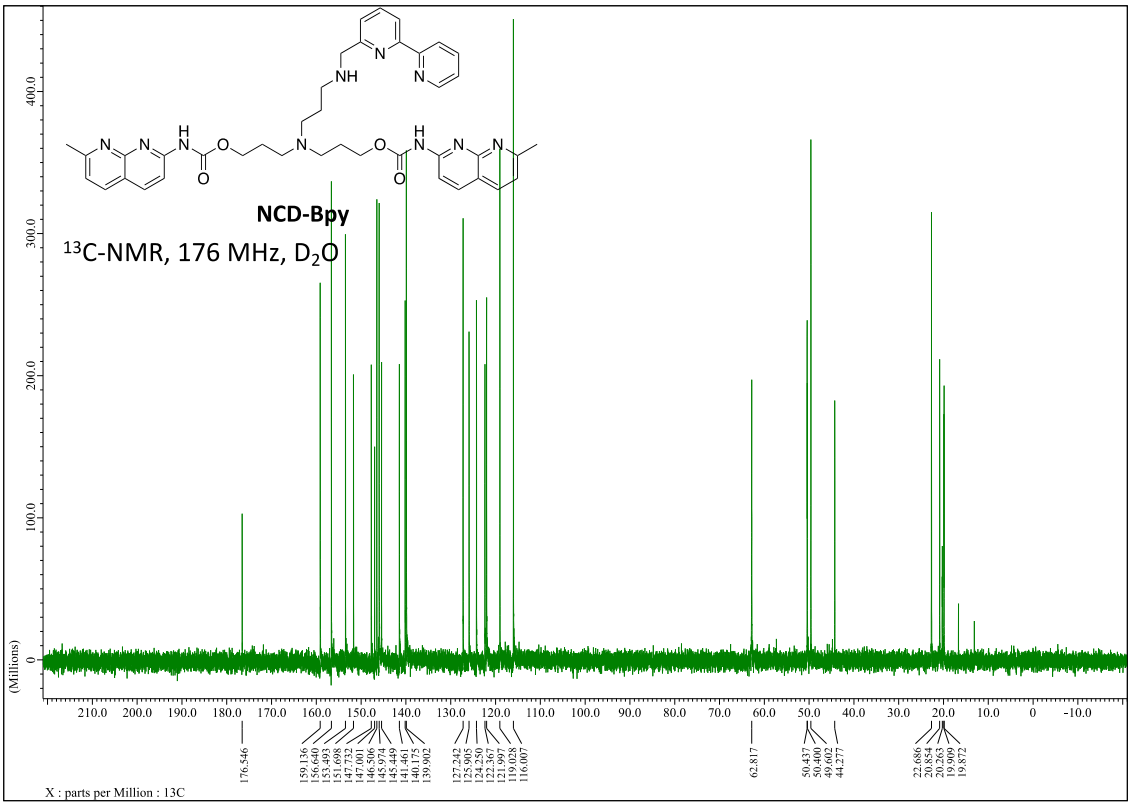

EtO
OEt
N
Boc
N
N
S2
1H-NMR, 600 MHz, CDCl3 with 0.05% v/v TMS
abundance
X : parts per Million : Proton


EtO
OEt
N
Boc
N
N
S2
13C-NMR, 151 MHz, CDCl3 with 0.05% v/v TMS
abundance
X : parts per Million : Carbon13


NBoc
S3
1H-NMR, 600 MHz, CDCl3 with 0.05% v/v TMS
abundance
X : parts per Million : Proton


NBoc
S3
13C-NMR, 176 MHz, CDCl3 with 0.05% v/v TMS
(Millions)
X : parts per Million : 13C


NCD-Bpy
^{1}H-NMR, 600 MHz, D_2O, solvent residual = 4.79 ppm
abundance
X : parts per Million : Proton


NCD-Bpy
^{13}C-NMR, 176 MHz, D_2O
(Millions)
X : parts per Million : 13C