Supporting Information

Natural-Like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications.

Lingjun Li,[†] Mélissa Degardin,[†] Thomas Lavergne,[†] Denis A. Malyshev,[†] Kirandeep Dhami,[†] Phillip Ordoukhanian,[‡] and Floyd E. Romesberg[†]*

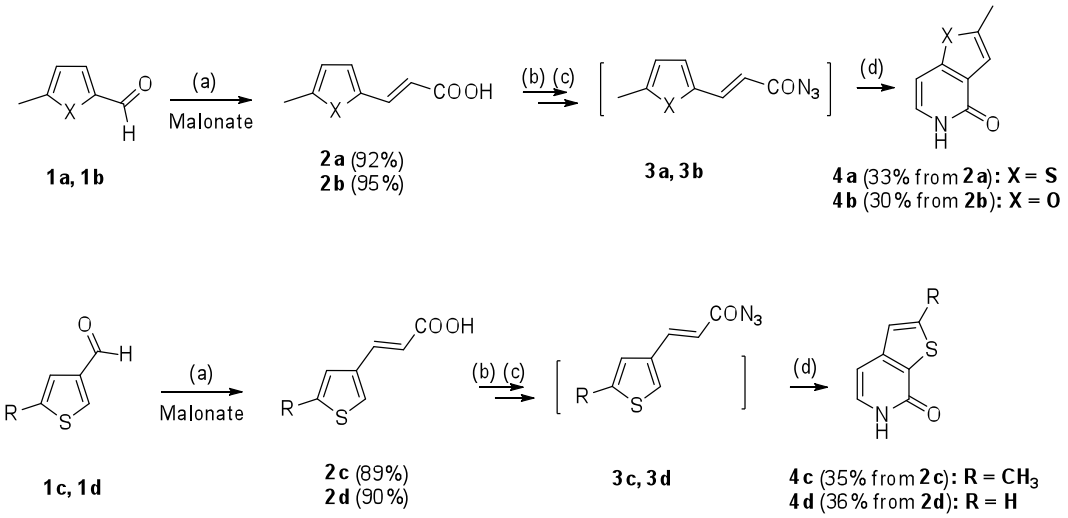
[†]*Department of Chemistry and* [‡]*Center for Protein and Nucleic Acid Research, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, 92037, USA.*

Table of Contents

General Procedure for Triphosphate Synthesis

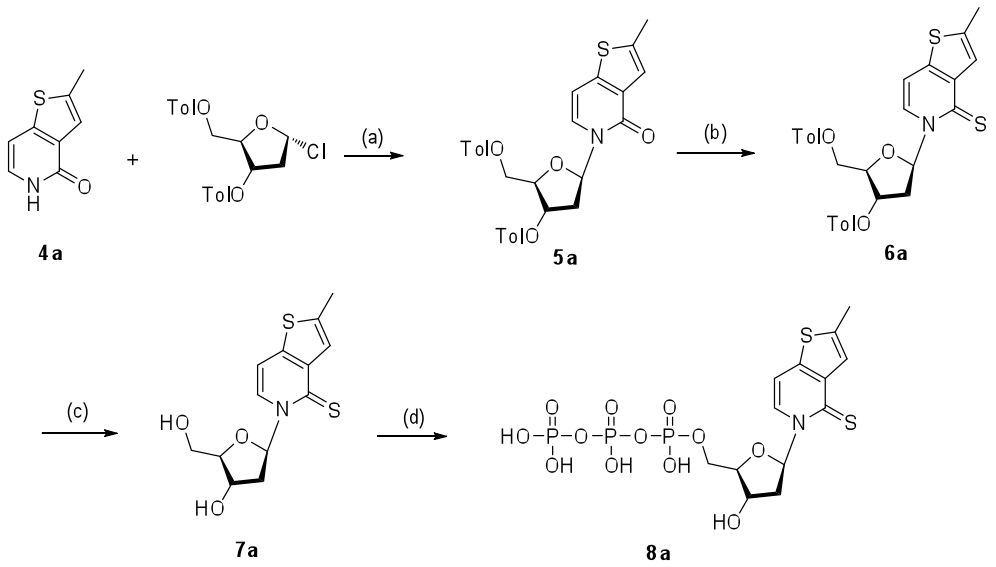
Proton sponge (1.3 equiv) and the free nucleoside derivative (1.0 equiv) were dissolved in dry trimethyl phosphate (40 equiv) and cooled to -15 °C under nitrogen atmosphere. Freshly distilled $POCl_3$ (1.3 equiv) was added dropwise and the resulting mixture was stirred at -10 °C for 2 h. Tributylamine (6.0 equiv) and a solution of tributylammonium pyrophosphate (5.0 eq.) in dimethylformamide (0.5 M) were added. Over 30 min, the reaction was allowed to warm slowly to 0 °C and then was quenched by addition of 0.5 M aqueous $Et_3NH_2CO_3$ (TEAB) pH 7.5 (2 vol-equiv.). The mixture was diluted two-fold with H_2O and the product was isolated on a DEAE Sephadex column (GE Healthcare) with an elution gradient of 0 to 1.2 M TEAB, evaporated, and co-distilled with H_2O (3×). Additional purification by reverse-phase (C18) HPLC (0 - 35% CH_3CN in 0.1 M TEAB, pH 7.5) was performed, (10% – 31% yield).

Synthetic Procedures and Characterizations:



Scheme S1 (a) Piperidine, Py, 100 °C, 12 h, then reflux for 1 h; (b) $SOCl_2$, DMF, $CHCl_3$, reflux, 3 h, (c) NaN_3, 1,4-dioxane, H_2O, 5 °C, 0.5 h; (d) diphenyl ether, 230 °C, 1 h.

The nucleobase analogs **4a**, **4b**, **4c** and **4d** were synthesized based on literature methods[1,2] as shown in **Scheme S1**. Briefly, condensation of the aldehyde (**1a**–**d)** with malonic acid at 100 °C in pyridine as a solvent and piperidine as a catalyst for 12 h, followed by a reflux for 1 h, yielded the corresponding acrylic acid intermediates (**2a**–**d**). Chlorination of these acids with thionyl chloride in chloroform in the presence of DMF afforded the acyl chlorides, which were not purified but could be used directly in the preparation of the azides (**3a**–**d**). Compounds **3a**–**d** were prepared in a biphasic mixture of 1,4-dioxane and water at 5 °C with sodium azide. Crude mixtures of **3a**–**d** in $CHCl_3$ solutions were added portion-wise to diphenyl ether and heated to 230 °C to give the isocyanates that underwent subsequent intramolecular cyclization to the fused 6-5 bicyclic systems **4a**–**d**.



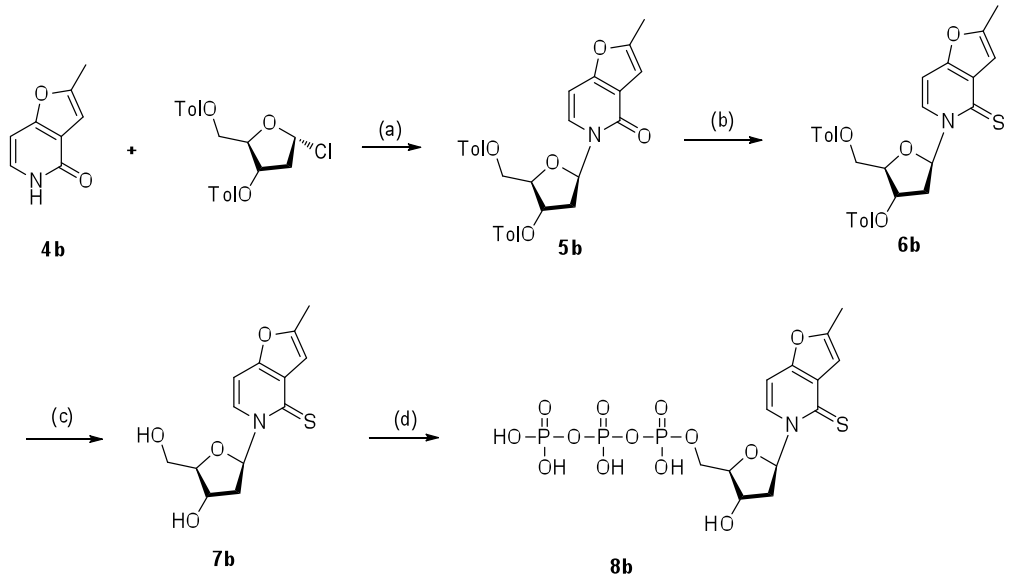
Scheme S2. (a) N,O-bis(TMS)acetamide, $SnCl_4$ (1 M in CH_2Cl_2), CH_2Cl_2, 3 h, 45%; (b) Lawesson's reagent, toluene, reflux, overnight, 58%; (c) 30% NaOMe, MeOH, rt, 1 h, 85%; (d) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 31%.

Compound 5a. To a solution of **4a** (54 mg, 0.33 mmol) in CH_2Cl_2 (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (83 mg, 0.39 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribosyl chloride (196 mg, 0.39 mmol) was added. The reaction mixture was cooled to 0 °C and $SnCl_4$ was added dropwise (1.0 M in CH_2Cl_2, 160 µL, 0.16 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound **5a** as white foam (77 mg, 0.15 mmol, 45%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.97-6.82 (m, 11H, Ar-H), 6.44 (d, J = 7.5 Hz, 1H, H-1'), 5.63 (d, J = 6.5 Hz, 1H, H-3'), 4.76-4.68 (m, 2H, H-5'a, 5'b), 4.59 (d, J = 2.5 Hz, H-4'), 2.89 (dd, J = 1.5, 0.5 Hz, 1H, H-2'a), 2.59 (s, 3H, Ar-CH_3), 2.43 (s, 3H , Ar-CH_3), 2.43 (s, 3H, Ar-CH_3), 2.36-2.30 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 166.5, 158.5, 147.2, 144.8, 144.6, 140.0, 131.1, 130.3, 130.0, 129.9, 129.7, 127.1, 126.9, 125.6, 122.8, 102.7, 85.9, 83.3, 75.6, 64.8, 39.6, 22.1, 16.1. HRMS (ESI$^+$) *m/z* calcd for $C_{29}H_{28}NO_6S$ (M+H$^+$) 518.1632, found 518.1621.

Compound 6a. Compound **5a** (27 mg, 0.052 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (41.5 mg, 0.10 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **6a** as a yellow foam (16 mg, 0.03 mmol, 58%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.00-7.89 (m, 4H, Ar-H), 7.71 (m, 1H, Ar-H), 7.49-7.48 (m, 1H, H-1'), 7.29-7.21 (m, 4H, Ar-H), 7.65-7.62 (m, 1H, Ar-H), 6.90 (d, J = 7.5 Hz, 1H, Ar-H), 5.64-5.62 (m, 1H, H-4'), 4.85-4.74 (m, 2H, H-5'a), 4.68-4.67 (m, 1H, H-5'b), 3.38-3.34 (m, 1H, H-3'), 2.26 (s, 3H), 2.44 (s, 3H), 2.41 (s, 3H), 2.28-2.22 (m, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 174.6, 166.6, 144.9, 144.8, 142.9, 142.7, 142.3, 130.3, 130.0, 129.7, 127.9, 127.0, 126.9, 126.8, 108.3, 100.0, 91.4, 84.0, 74.9, 64.5, 39.3, 22.2, 22.1, 16.3. HRMS (ESI$^+$) *m/z* calcd for $C_{29}H_{28}NO_5S_2$ (M+H$^+$) 534.1403, found 534.1404.

Compound 7a. To a solution of **6a** (20 mg, 0.037 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH_2Cl_2) to afford compound **7a** as yellow foam (9.2 mg, 0.031 mmol, 85%).^{1}H NMR (500 MHz, CD_3OD) δ 8.36 (d, J = 4 Hz, 1H, Ar-H), 7.58 (d, J = 1Hz, 1H, Ar-H), 7.35 (t, J = 4 Hz, 1H, H-1'), 7.22 (d, J = 8 Hz, 1H, Ar-H), 4.07-4.06 (m, 1H, H-4'), 4.07 (d, J = 4 Hz, 1H, H-3'), 3.80 (dd, J = 24, 4 Hz, 2H, H-5'a, b), 2.79-2.76 (m, 1H, H-2'a), 2.13-2.08 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ 173.60, 143.29, 142.26, 142.23, 129.35, 126.11, 108.01, 91.14, 88.44, 70.37, 61.35, 41.59, 14.81. HRMS (ESI$^+$) *m/z* calcd for $C_{13}H_{16}NO_3S_2$ (M+H$^+$) 298.0566, found 298.0569.

Compound 8a. Compound **8a** (11.2 mg, 20.8 μmol, 31%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **7a** (20 mg, 67.3 μmol). ^{31}P NMR (162 MHz, D_2O) δ -10.3 (d, *J* = 19.8 Hz, γ-P), -10.9 (d, *J* = 20.1 Hz, α-P), -22.8 (t, *J* = 19.4 Hz, β-P). MS (MALDI-TOF$^-$, matrix: 9-aminoacridine) (*m/z*): [M-H]$^-$ calcd for $C_{13}H_{17}NO_{12}P_3S_2$, 536.3, found, 536.7.



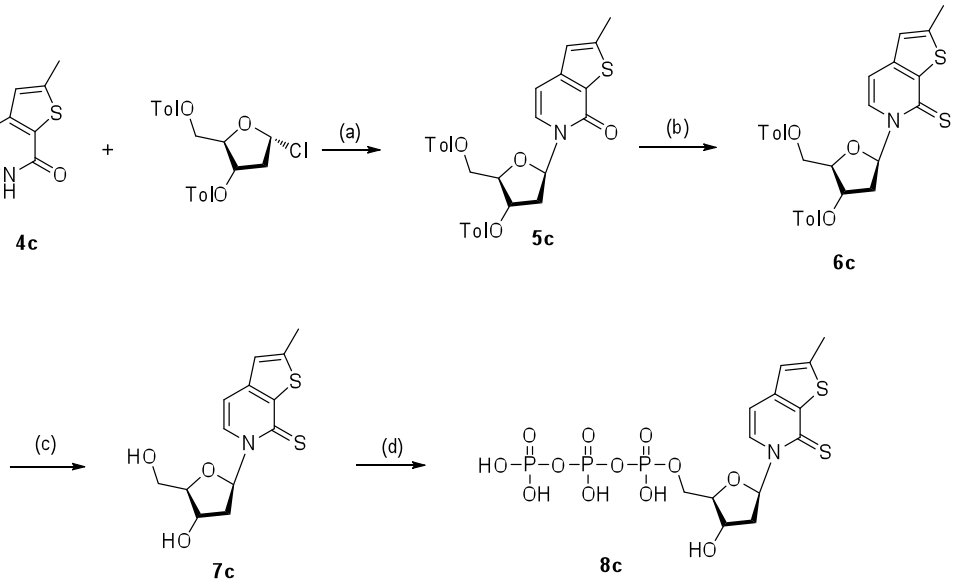
Scheme S3. (a) N,O-bis(TMS)acetamide, $SnCl_4$ (1 M in CH_2Cl_2), CH_2Cl_2, 3 h, 41%; (b) Lawesson's reagent, toluene, reflux, overnight, 52%; (c) 30% NaOMe, MeOH, rt,1 h, 85%; (d) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 21%.

Compound 5b. To a solution of **4b** (100 mg, 0.67 mmol) in CH_2Cl_2 (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (165 mg, 0.81 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribosyl chloride (292 mg, 0.81 mmol) was added. The reaction mixture was cooled to 0 °C and $SnCl_4$ was added dropwise (1.0 M in CH_2Cl_2, 200 μL, 0.2 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound **5b** as white foam (137 mg, 0.27 mmol, 41%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.99 (d, *J* = 8.1 Hz, 2H, Ar-H), 7.93 (d, *J* = 8.1 Hz, 2H, Ar-H), 7.55 (d, *J* = 7.7 Hz, 1H, Ar-H), 7.33 - 7.28 (m, 2H, Ar-H), 7.27 - 7.20 (m, 2H, Ar-H), 6.82 (dd, *J* = 8.3, 5.6 Hz, 1H, Ar-H), 6.57 (d, *J* = 0.9 Hz, 1H, Ar-H), 6.41 (d, *J* = 7.7 Hz, 1H, H-1'), 5.68 - 5.61 (m, 1H, H-4'), 4.75 (dd, *J* = 12.1, 3.4 Hz, 2H, H-5'a, b), 4.62 (q, *J* = 3.1 Hz, 1H, H-3'), 2.94 (ddd, *J* = 14.3, 5.6, 1.7 Hz, 1H, H-2'a), 2.48 - 2.39 (s, 3×3H, Ar-CH_3), 2.36 - 2.26 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 166.5, 159.2, 159.1, 154.5, 144.8, 144.6, 130.2, 130.0, 129.7, 127.5, 127.1, 126.9, 117.5, 103.3, 96.6, 86.0, 83.2, 75.5, 64.7, 39.8, 22.1, 14.1. HRMS (ESI$^+$) *m/z* calcd for $C_{29}H_{28}NO_7$ (M+H$^+$) 502.1860, found 502.1885.

Compound 6b. Compound **5b** (29 mg, 0.056 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (41.5 mg, 0.10 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **6b** as a yellow foam (15 mg, 0.029 mmol, 52%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10-7.89 (m, 5H, Ar-H), 7.52-7.48 (m, 1H, H-1'),7.29-7.22 (m, 4H, Ar-H), 6.8 (d, *J* = 1Hz, 1H, Ar-H), 6.73 (d, *J* = 7.5 Hz, 1H,Ar-H), 5.65-5.62 (m, 1H, H-4'), 4.84-4.74 (m, 2H, H-5'a, b), 4.67-4.65 (m, 1H, H-3'), 3.36-3.32 (m, 1H, H-2'a), 2.44 (s, 3H, Ar-CH_3), 2.43 (s, 3H, s, 3H, Ar-CH_3), 2.41 (s, 3H, s, 3H, Ar-CH_3), 2.27-2.21 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 156.9, 153.9, 144.8, 130.3, 130.0, 129.8, 129.7, 127.9, 106.4, 96.0,83.9, 56.6, 39.5, 22.1, 12.6. HRMS (ESI$^+$) *m/z* calcd for $C_{29}H_{28}NO_6S$ (M+H$^+$) 518.1632, found 518.1638.

Compound 7b. To a solution of **6b** (20 mg, 0.039 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography ($MeOH/CH_2Cl_2$) to afford compound **7b** as yellow foam (9.3 mg, 0.033 mmol, 85%). ^1H NMR (500 MHz, CD_3OD) δ 8.57 (d, *J* = 5 Hz, 1H, Ar-H), 7.42 (t, *J* = 4 Hz, 1H, H-1'), 7.13 (d, *J* = 7.5 Hz, 1H, Ar-H), 6.80 (s, 1H, Ar-H), 4.50-4.47 (m, 1H, H-4'), 4.12 (d, *J* = 3.5 Hz, 1H, H-3'), 3.95 (dd, *J* = 30, 3 Hz, 2H, H-5'a, b), 2.81-2.77 (m, 1H, H-2'a), 2.50 (s, 3H, Ar-CH_3), 2.18-2.14 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ 172.9, 157.2, 154.5, 132.7, 131.3, 105.4, 100.8, 90.9, 88.5, 70.3, 61.3, 41.8, 12.7. HRMS (ESI^+) *m/z* calcd for $C_{13}H_{16}NO_4S$ ($M+H^+$) 282.0795, found 282.0790.

Compound 8b. Compound **8b** (3.7 mg, 7.1 μmol, 10%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **7b** (20 mg, 71.2 μmol). ^{31}P NMR (162 MHz, D_2O) δ -10.4 (d, *J* = 20.0 Hz, γ-P), -10.9 (d, *J* = 19.4 Hz, α-P), -22.8 (t, *J* = 20.0 Hz, β-P). MS (MALDI-TOF^-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{13}H_{17}NO_{13}P_3S$, 520.3, found, 520.1.



Scheme S4. (a) N,O-bis(TMS)acetamide, $SnCl_4$ (1 M in CH_2Cl_2), CH_2Cl_2, 3 h, 40%; (b) Lawesson's reagent, toluene, reflux, overnight, 31%; (c) 30% NaOMe, MeOH, rt,1 h, 81%; (d) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 15%.

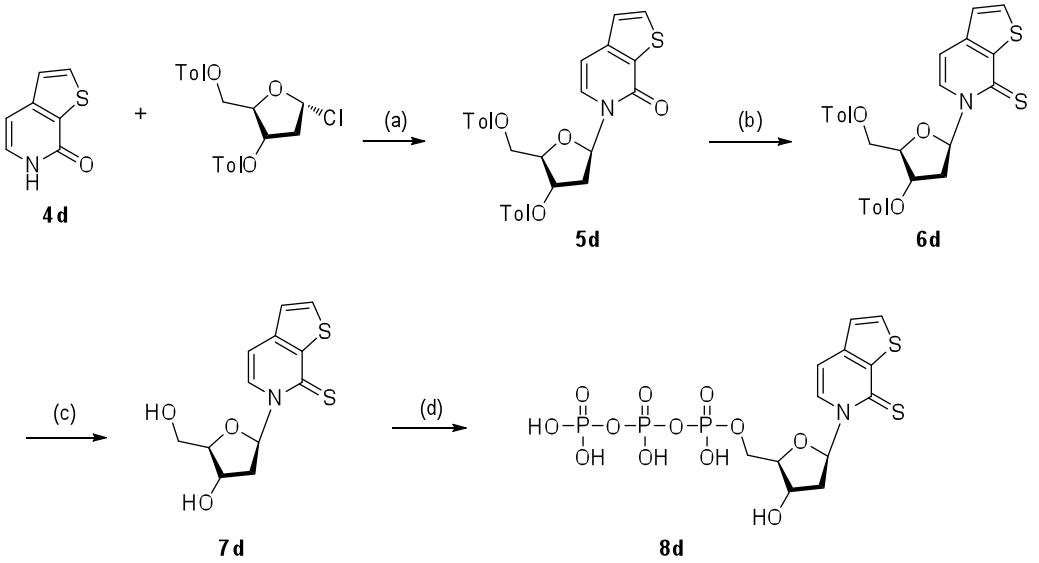
Compound 5c. To a solution of **4c** (46 mg, 0.28 mmol) in CH_2Cl_2 (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (66 mg, 0.33 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribosyl chloride (120 mg, 0.33 mmol) was added. The reaction mixture was cooled to 0 °C and $SnCl_4$ was added dropwise (1.0 M in CH_2Cl_2, 140 μL, 0.14 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound **5c** as white foam (58 mg, 0.11 mmol, 40%). ^1H NMR (500 MHz, $CDCl_3$) δ 7.98-7.90 (m, 4H, Ar-H), 7.53 (d, *J* = 7.4 Hz, 1H, Ar-H), 7.27- 7.21 (m, 4H, Ar-H), 6.83-6.82 (m, 2H, Ar-H), 6.44 (d, *J* = 7.5 Hz, 1H, H-1'), 5.63(d, *J* = 6.5 Hz, 1H, H-4'), 4.76 - 4.60 (m, 2H, H-5'a, b), 4.59 (d, *J* = 2.5 Hz, 1H, H-3'), 2.89 (dd, *J* = 13, 5.5 Hz, H-2'a), 2.59 (s, 3H, Ar-CH_3), 2.43 (s, 3H, Ar-CH_3), 2.40 (s, 3H, Ar-CH_3), 2.37-2.30 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.5, 158.0, 149.9, 146.3, 144.8, 144.6, 130.3, 130.0, 129.7, 128.8, 127.3, 122.8, 103.7, 100.0, 85.8, 83.2, 75.5, 64.8, 39.5, 22.1, 16.7. HRMS (ESI^+) *m/z* calcd for $C_{29}H_{28}NO_6S$ ($M+H^+$) 518.1632, found 518.1631.

Compound 6c. Compound **5c** (50 mg, 0.097 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1.5 mL). Lawesson's reagent (83 mg, 0.20 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude

product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **6c** as a yellow foam (16 mg, 0.03 mmol, 31%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.13-7.97 (m, 5H, Ar-H), 7.52-7.49 (m ,1H, H-1'), 7.37-7.29 (m, 4H, Ar-H), 6.99 (d, *J* = 1Hz, 1H, Ar-H), 6.91 (d, *J* = 7.5 Hz, 1H, Ar-H), 5.73-5.71 (m , 1H, H-4'), 4.91-4.82 (m, 2H, H-5'a, b), 4.76-4.74 (m, 1H, H-3'), 3.41-3.37 (m, 1H, H-2'a), 2.68 (s, 3H, Ar-CH_3), 2.52 (s, 3H, Ar-CH_3), 2.49 (s, 3H, Ar-CH_3), 2.39-2.34 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.1, 166.6, 154.0, 144.9, 144.7, 140.4, 130.3, 130.0, 129.7, 129.6, 127.0, 126.8, 122.7, 109.0, 91.2, 83.9, 75.0, 64.5, 39.4, 22.2, 22.1, 17.0. HRMS (ESI$^+$) *m/z* calcd for $C_{29}H_{28}NO_5S_2$ ($M+H^+$) 534.1403, found 534.1406.

Compound 7c. To a solution of **6c** (20 mg, 0.037 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH_2Cl_2) to afford compound **7c** as yellow foam (8.9 mg, 0.03 mmol, 81%). ^{1}H NMR (500 MHz, CD_3OD) δ 8.48, (d, *J* = 7.5 Hz, 1H, Ar-H), 7.42 (t, *J* = 5 Hz, 1H, H-1'), 7.20 (d, *J* = 5 Hz, 1 H, Ar-H), 7.12 (s, 1H, Ar-H), 4.51-4.48 (m, 1H, H-4'), 4.13 (d, *J* = 5 Hz, 1H, H-3'), 3.95 (dd, *J* = 30, 5 Hz, 2H, H-5'a, b), 2.81-2.78 (m, 1H, H-2'a), 2.67 (s, 3H, Ar-CH_3), 2.21-2.16 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ 171.1, 154.0, 144.1. 141.1, 131.1, 122.7, 108.8, 90.9, 88.5, 70.5, 61.4, 41.7, 15.4. HRMS (ESI$^+$) *m/z* calcd for $C_{13}H_{16}NO_3S_2$ ($M+H^+$) 298.0566, found 298.0566.

Compound 8c. Compound **8c** (10.8 mg, 20.2 μmol, 30%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **7c** (20 mg, 67.3 μmol). ^{31}P NMR (162 MHz, D_2O) δ -10.8 (d, *J* = 19.8 Hz, γ-P), -11.5 (d, *J* = 20.1 Hz, α-P), -23.3 (t, *J* = 20.1 Hz, β-P). MS (MALDI-TOF$^-$, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{13}H_{17}NO_{12}P_3S_2$, 536.3, found, 536.1



Scheme S5. (a) N,O-bis(TMS)acetamide, $SnCl_4$ (1 M in CH_2Cl_2), CH_2Cl_2, 3 h, 39%; (b) Lawesson's reagent, toluene, reflux, overnight, 33%; (c) 30% NaOMe, MeOH, rt, 1 h, 82%; (d) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 30%.

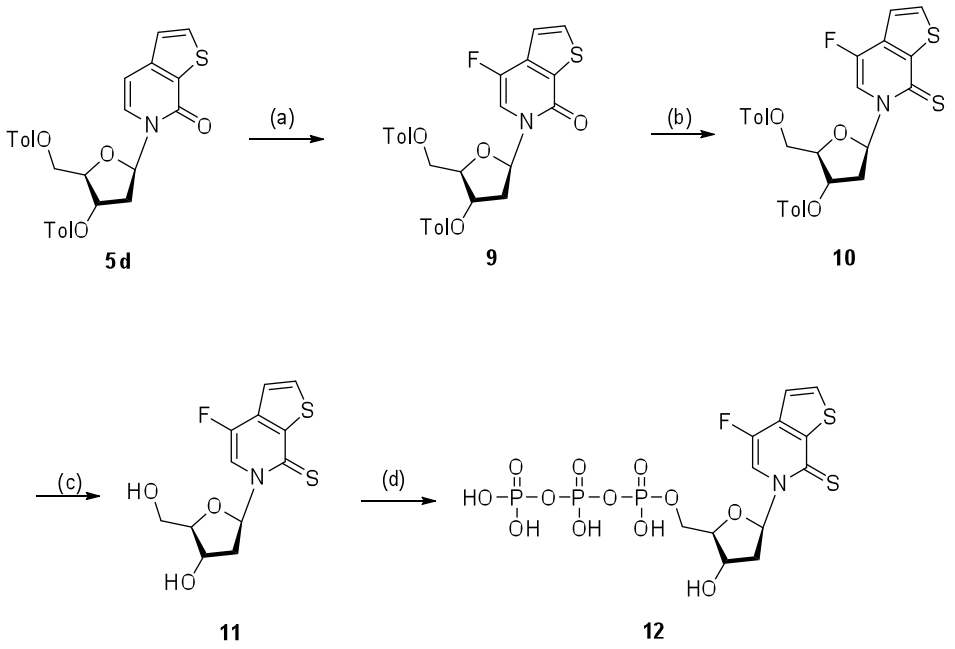
Compound 5d. To a solution of **4d** (200 mg, 1.32 mmol) in CH_2Cl_2 (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (298 mg, 1.46 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribosyl chloride (563 mg, 1.46 mmol) was added. The reaction mixture was cooled to 0 °C and $SnCl_4$ was added dropwise (1.0 M in CH_2Cl_2, 660 μL, 0.66 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound **5d** as a white foam (260 mg, 0.52 mmol, 39%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.98-7.90 (m, 4H, Ar-H), 7.70 (d, *J* = 6 Hz, 1H, Ar-H), 7.55 (d, *J* = 9.5 Hz, 1H, Ar-H), 7.28-7.16 (m, 5H, Ar-H), 6.84-6.85 (m, 1H, Ar-H), 6.57 (d, *J* = 9.5 Hz, H-1'), 5.66-5.64 (m, 1H, H-4'), 4.75-4.72 (m, 2H, H-5'a, b), 4.61 (m, 1H, H-3'), 2.95-2.90 (m, 1H, H-2'a), 2.43 (s, 3H, Ar-CH_3), 2.40 (s, 3H, Ar-CH_3), 2.39-2.31(m, 1H, H-2'b). ^{13}C

NMR (125 MHz, $CDCl_3$) δ 166.2, 166.1, 158.1, 145.1, 144.4, 144.2, 133.8, 129.9, 129.6, 129.3, 129.1, 126.9, 126.5, 124.2, 103.5, 85.5, 82.9, 75.1, 64.4, 39.2, 21.7. HRMS (ESI^+) *m/z* calcd for $C_{20}H_{20}Cl_2N_2O_5S$ ($M+H^+$) 504.1475, found 504.1480.

Compound 6d. Compound **5d** (50 mg, 0.1 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (48 mg, 0.12 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated, and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **6d** as a yellow foam (17 mg, 0.033 mmol, 33%). ^1H NMR (500 MHz, $CDCl_3$) δ 8.14-7.82 (m, 7H, Ar-H), 7.51 (dd, *J* = 7.5, 6.0 Hz, 1H, H-1'), 7.32-7.23 (m, 5H, Ar-H), 6.99 (d, *J* = 7.2 Hz, 1H, Ar-H), 74-5.73 (m, 1H, H-4'), 4.92-4.83 (m, 2H, H-5'a, b), 4.78-4.77 (m, 1H, H-3'), 3.43-3.40 (m, 1H, H-2'a), 2.51 (s, 3H, Ar-CH_3), 2.48 (s, 3H, Ar-CH_3), 2.39-2.36 (m ,1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.5, 166.6, 144.9, 144.8, 139.5, 138.0, 134.5, 130.3, 130.0, 129.7, 129.5, 126.8, 124.7, 109.5, 91.4, 84.0, 75.0, 64.5, 39.4, 22.2, 22.1. HRMS (ESI^+) *m/z* calcd for $C_{28}H_{26}NO_5S_2$ ($M+H^+$) 520.1247, found 520.1241.

Compound 7d. To a solution of **6d** (20 mg, 0.039 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography ($MeOH/CH_2Cl_2$) to afford compound **7d** as yellow foam (9.0 mg, 0.032 mmol, 82%).^1H NMR (500 MHz, CD_3OD) δ 8.48 (d, *J* = 5 Hz, 1H, Ar-H), 8.01 (d, *J* = 5 Hz, 1H, Ar-H), 7.40-7.38 (m, 2 H, Ar-H), 7.29 (d, *J* = 10 Hz, 1H, H-1'), 4.47-4.46 (m, 1H, H-4'), 4.10 (m, 1H, H-3'), 3.94-3.88 (m, 2H, H-5'a ,b), 2.77-2.76 (m, 1H, H-2'a), 2.19-2.14 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ171.2, 144.7, 139.6, 137.6, 130.5, 124.2, 108.8, 90.7, 88.2, 70.1, 61.0, 41.3. HRMS (ESI^+) *m/z* calcd for $C_{12}H_{14}NO_3S_2$ ($M+H^+$) 284.041, found 284.0410.

Compound 8d. Compound **8d** (5.7 mg, 10.9 μmol, 31%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **7d** (10 mg, 35.3 μmol). ^{31}P NMR (162 MHz, D_2O) δ -9.3 (d, *J* = 19.5 Hz, γ-P), -10.8 (d, *J* = 19.8 Hz, α-P), -22.4 (t, *J* = 20.0 Hz, β-P). MS (MALDI-TOF^-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{12}H_{15}NO_{12}P_3S_2^-$, 521.9, found, 521.9.



Scheme S6. (a) i. Selectfluor, $MeOH/CH_3CN$, reflux, 3 h; ii. $TfOH-CH_2Cl_2$ (1:1 v/v), 1 h, 85%; (b) Lawesson's reagent, toluene, reflux, overnight, 32%; (c) 30% NaOMe, MeOH, rt, 1 h, 85%; (d) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 10%.

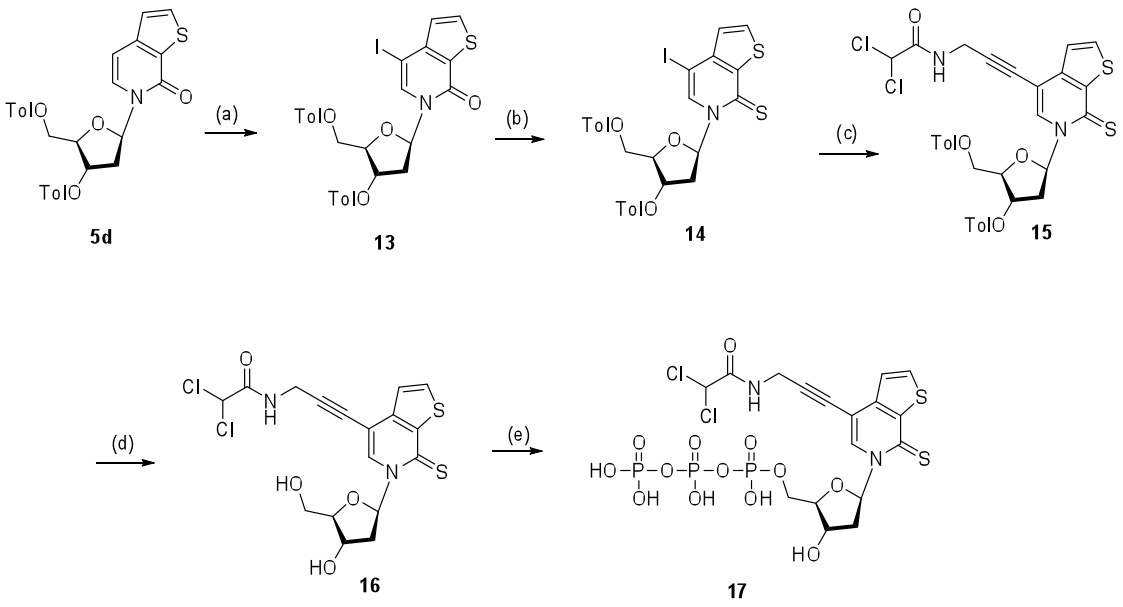
Compound 9. Compound **5d** (55 mg, 0.11 mmol) was dissolved in 1.0 mL $MeOH-CH_3CN$ (1:1 v/v), Selectfluor (42 mg, 0.12 mmol) was added and the mixture was heated at reflux for 3 h, then the solvent was evaporated, the

residue was dissolved in EtOAc (20 mL), the organic phase was washed with water three times. Then the organic solvent was evaporated, and the solid residue was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in 1 mL TfOH-CH_2Cl_2 (1:1 v/v) and the mixture was stirred at room temperature for 1 h, then the mixture was concentrated, and the crude product was subjected to a silica gel column chromatography (hexane/EtOAc) to afford compound **9** as a white solid (49 mg, 0.093 mmol, 85%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.98-7.92 (m, 4H, Ar-H), 7.75 (d, *J* = 5 Hz, 1H, Ar-H), 7.52 (d, *J* = 7.5 Hz, 1H, Ar-H), 7.32-7.21 (m, 5H, Ar-H), 6.82-6.78 (m, 1H, H-1'), 5.64-5.61 (m, 1H, H-4'), 4.80-4.59 (m, 2H, H-5'a, b), 4.62-4.59 (m, 1H, H-3'), 2.93- 2.87 (m, 1H, H-2'a), 2.43 (s, 3H, Ar-CH_3), 2.39 (s, 3H, Ar-CH_3), 2.34-2.27 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.2, 166.1, 156.4, 144.5, 144.3, 137.7, 137.5, 134.6, 129.9, 129.6, 129.3, 126.6, 126.4, 120.2, 112.1, 111.7, 85.5, 83.1, 75.0, 64.2, 39.1, 21.8, 21.7. ^{19}F NMR (376 MHz, $CDCl_3$) δ -151.5. HRMS (ESI$^+$) *m/z* calcd for $C_{28}H_{25}FNO_6S$ (M+H$^+$) 522.1381, found 522.1380.

Compound 10. Compound **9** (20 mg, 0.038 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (18.5 mg, 0.046 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **10** as a yellow foam (6.5 mg, 0.012 mmol, 32%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.11-7.85 (m, 6H, Ar-H), 7.40-7.39(m, 2H, Ar-H, H-1'), 7.28-7.21 (m, 4H, Ar-H), 5.64-5.63 (m, 1H, H-4'), 4.83 (m, 2H, H-5'a, b), 4.69 (m, 1H, H-3'), 3.34-3.29 (m, 1H, H-2'a), 2.44 (s, 1H, Ar-CH_3), 2.40 (s, 3H, Ar-CH_3), 2.30-2.26 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.9, 166.6, 166.5, 144.9, 144.8, 138.6, 130.3, 130.0, 129.9, 129.7, 129.7, 126.9, 126.7, 120.5, 116.3, 116.0, 100.0, 91.6, 84.3, 74.7, 64.3, 39.2, 22.2, 22.1. ^{19}F NMR (376 MHz, $CDCl_3$) δ -142.9. HRMS (ESI$^+$) *m/z* calcd for $C_{28}H_{25}FNO_5S_2$ (M+H$^+$) 538.1153, found 538.1155.

Compound 11. To a solution of **10** (10 mg, 0.019 mmol) in methanol (1.5 mL) was added dropwise 30% NaOMe (4.33 mg, 0.08 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH_2Cl_2) to afford compound **11** as yellow foam (4.9 mg, 0.016 mmol, 85%). ^{1}H NMR (500 MHz, CD_3OD) δ 8.68 (d, *J* = 5 Hz, 1H, Ar-H), 8.12 (d, *J* = 5 Hz, 1H, Ar-H), 7.52 (d, *J* = 5 Hz, 1H, Ar-H), 7.28 (t, *J* = 6.5 Hz, 1H, H-1'), 4.48 (m, 1H, H-4'), 4.10 (m, 1H, H-3'), 3.94 (dd, *J* = 35, 3 Hz, 2H, H-5'a, b), 2.78-2.75 (m, 1H, H-2'a), 2.24-2.19 (m ,1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ 170.2, 150.2, 148.3, 139.0, 131.7, 131.6, 119.8, 117.8, 117.4, 91.5, 88.7, 70.1, 61.0, 41.5. ^{19}F NMR (376 MHz, CD_3OD) δ -145.3. HRMS (ESI$^+$) *m/z* calcd for $C_{12}H_{13}FNO_3S_2$ (M+H$^+$) 302.0315, found 302.0314.

Compound 12. Compound **12** (2.0 mg, 3.7 μmol, 22%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **11** (5 mg, 16.6 μmol). ^{31}P NMR (162 MHz, D_2O) δ -10.9 (d, *J* = 20.0 Hz, γ-P), -11.6 (d, *J* = 21.1 Hz, α-P), -23.3 (t, *J* = 23.1 Hz, β-P).^{19}F NMR (376 MHz, D_2O) δ -138.5 (s). MS (MALDI-TOF$^-$, matrix: 9-aminoacridine) (*m/z*): [M-H]$^-$ calcd for $C_{12}H_{14}FNO_{12}P_3S_2^-$, 539.9, found, 540.1.



Scheme S7. (a) ICl, CH_2Cl_2, 0 °C to rt, overnight, 63%; (b) Lawesson's reagent, toluene, reflux, overnight, 27%; (c) 2,2-dichloro-N-prop-2-yn-1-ylacetamide, $(PPh_3)_4Pd$, CuI, Et_3N, DMF, rt, overnight, 91%; (d) 30% NaOMe, MeOH, rt,1 h, 74%; (e) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 25%.

Compound 13. To a solution of **5d** (73 mg, 0.145 mmol) in CH_2Cl_2 (1 mL) at 0 °C under nitrogen atmosphere was added dropwise iodine monochloride (1.0 M in CH_2Cl_2, 0.15 ml, 0.15 mmol). The resulting mixture was stirred at room temperature overnight. The reaction mixture was quenched with saturated aqueous $NaHCO_3$ and saturated aqueous $Na_2S_2O_3$, extracted with CH_2Cl_2, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound **13** as white foam (57 mg, 0.091 mmol, 63%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.98-7.93 (m, 4H, Ar-H), 7.83 (s, 1H, Ar-H), 7.72 (d, J = 5 Hz, 1H, Ar-H), 7.28-7.17 (m, 5H, Ar-H), 6.78-6.75 (m, 1H, H-1'), 5.65-5.63 (m, 1H, H-4'), 4.76 (m, 2H, H-5'a, b), 4.63-4.62 (m, 1H, H-3'), 2.95-2.91 (m, 1H, H-2'a), 2.43(s, 3H, Ar-CH_3), 2.35 (s, 3H, Ar-CH_3), 2.34-2.29 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 166.5, 157.6, 147.4, 144.9, 144.6, 133.7, 132.7, 130.3, 130.1, 129.8, 129.7, 129.4, 128.5, 127.0, 126.8, 86.1, 83.8, 75.7, 64.7, 39.9, 22.2, 22.1. HRMS (ESI$^+$) *m/z* calcd for $C_{28}H_{25}INO_5S$ ($M+H^+$) 630.0442, found 630.0440.

Compound 14. Compound **13** (30 mg, 0.048 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL), Lawesson's reagent (23 mg, 0.057 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound **14** as a yellow foam (8.4 mg, 0.013 mmol, 27%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.31 (s, 1H, Ar-H), 7.99-7.82 (m, 5H, Ar-H), 7.39-7.36 (m, 1H, H-1'), 7.29-7.20 (m, 5H, Ar-H), 5.65-5.64 (m, 1H, H-4'), 4.83-4.81 (m, 2H, H-5'a, b), 4.71-4.70 (m, 1H, H-3'), 3.35 (dd, J = 15, 5.5 Hz, 1H, H-2'a), 2.44 (s, 3H, Ar-CH_3), 2.39 (s, 3H, Ar-CH_3), 2.27-2.21 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.9, 166.6, 144.9, 144.7, 144.6, 141.8, 137.8, 135.0, 130.3, 130.2, 129.8, 129.7, 128.6, 126.9, 126.7, 91.7, 84.5, 75.3, 64.6, 39.5, 22.2, 22.1. HRMS (ESI$^+$) *m/z* calcd for $C_{28}H_{25}INO_5S_2$ ($M+H^+$) 646.0213, found 646.0219.

Compound 15. To a solution of **14** (10 mg, 0.015 mmol) in DMF (2 mL) under nitrogen atmosphere was added $(PPh_3)_4Pd$ (1.7 mg, 0.0015 mmol), CuI (0.57 mg, 0.011 mmol) and Et_3N (5 µL, 0.030 mmol). The reaction mixture was degassed and a solution of $Cl_2CHCONHCH_2CCH$ (3.8 mg, 0.0225 mmol) in DMF (0.5 mL) was added. The reaction mixture was stirred overnight at room temperature and monitored by TLC. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (MeOH/CH_2Cl_2) to afford compound **15** as yellow foam (9.2 mg, 0.0135 mmol, 91%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.26 (s, 1H, Ar-H), 7.99-7.82 (m, 5H, Ar-H), 7.40-7.37 (m, 2H, Ar-H, H-1'), 7.29-7.21 (m, 4 H, Ar-H), 6.71 (br, 1H, NH),6.95 (s, 1H,

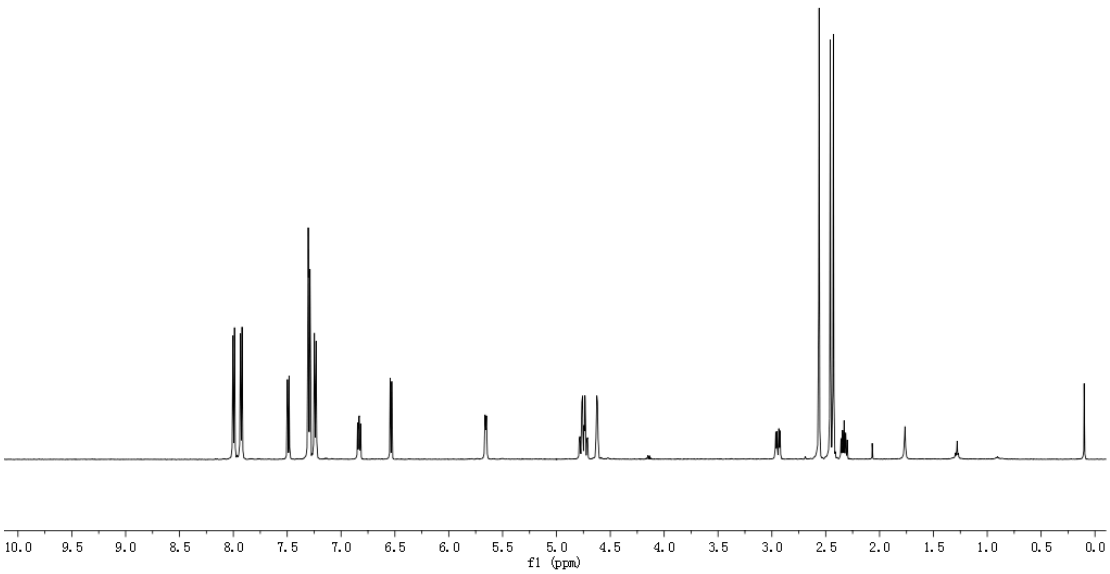
$CHCl_2$), 5.65-5.64 (m, 1H, H-4'), 4.85-4.79 (m, 2H, H-5'a, b), 4.73 (m, 1H, H-3'), 4.26-4.11 (m, 2H, $NHCH_2$), 3.38-3.34 (m, 1H, H-2'a), 2.44 (s, 3H, Ar-CH_3), 2.40 (s, 3H, Ar-CH_3), 2.31-2.25 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.3, 166.6, 164.1, 144.8, 139.0, 138.3, 133.4, 130.3, 130.1, 129.8, 129.7, 126.9, 124.3, 104.8, 91.8, 88.3, 84.5, 78.8, 75.2, 66.4, 64.7, 39.6, 31.3, 22.2. 22.1. HRMS (ESI^+) *m/z* calcd for $C_{33}H_{29}Cl_2N_2O_6S_2$ ($M+H^+$) 683.0839, found 683.0854.

Compound 16. To a solution of **15** (9.2 mg, 0.0135 mmol) in methanol (1.0 ml) was added dropwise 30% NaOMe (2.92 mg, 0.32 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was concentrated and the crude product was subjected to silica gel column chromatography ($MeOH/CH_2Cl_2$) to afford compound **16** as yellow foam (4.5 mg, 0.01 mmol, 74%).^{1}H NMR (500 MHz, CD_3OD) δ 8.69 (s, 1H, Ar-H), 8.06 (d, *J* = 5 Hz, 1H, Ar-H), 7.53 (d, *J* = 5 Hz, 1H, Ar-H), 7.30 (t, *J* = 5 Hz, 1H, H-1'), 6.33 (s, 1H, $CHCl_2$), 4.47-4.46 (m, 1H, H-4'), 4.36 (s, 2H, $NHCH_2$), 4.11-4.08 (m, 1H, H-3'), 3.97 (dd, *J* = 12, 3Hz, 2H, H-5'a, b), 2.79-2.74 (m, 1H, H-2'a), 2.21-2.16 (m, 1H, H-2'b). ^{13}C NMR (125 MHz, CD_3OD) δ 172.9, 138.4, 134.3, 123.9, ,122.8, 104.7, 100.0, 91.2, 88.6, 77.1, 70.3, 66.4, 61.1, 41.7, 30.2. HRMS (ESI^+) *m/z* calcd for $C_{17}H_{17}Cl_2N_2O_4S_2$ ($M+H^+$) 447.0001, found 447.0020.

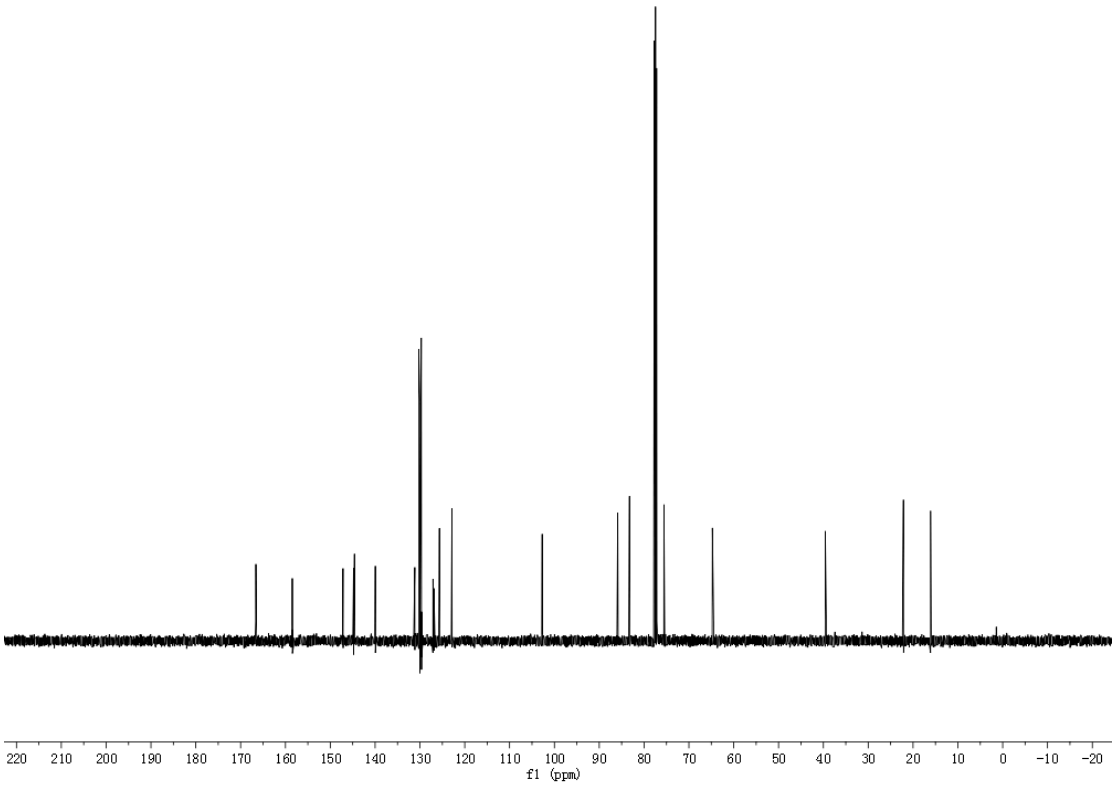
Compound 17. Compound **17** (2.2 mg, 3.1 μmol, 28%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from **16** (5 mg, 11.2 μmol). ^{31}P NMR (162 MHz, D_2O) δ -10.85 (d, *J* = 19.9 Hz, γ-P), -11.63 (d, *J* = 20.0 Hz, α-P), -23.07 (s), -23.26 (t, *J* = 19.7 Hz, β-P). MS (MALDI-TOF^-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{17}H_{18}Cl_2N_2O_{13}P_3S_2^-$, 684.9, found, 685.0.

NMR Spectra

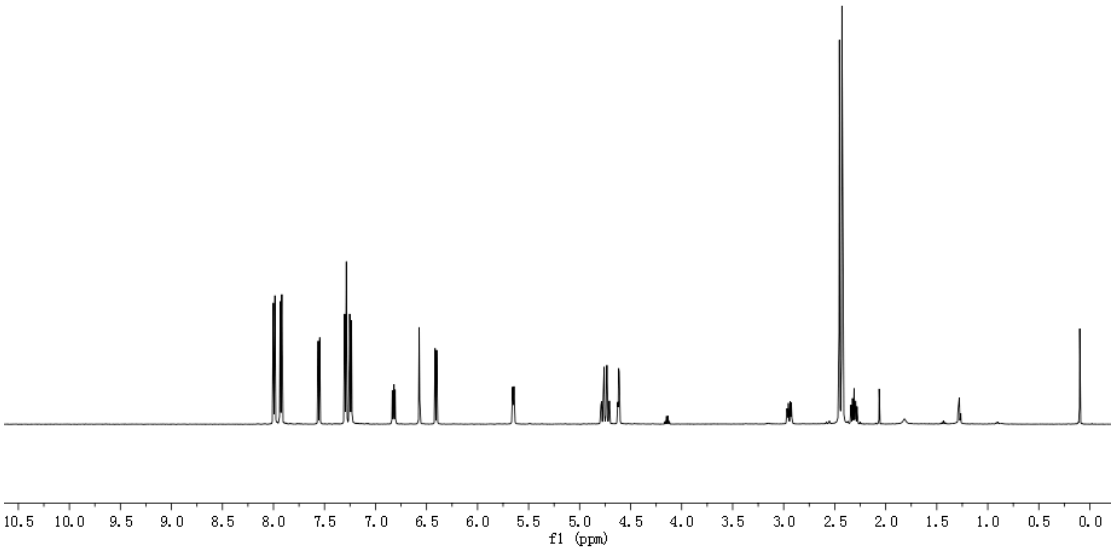
500 MHz ^{1}H NMR spectrum of **5a** ($CDCl_3$)



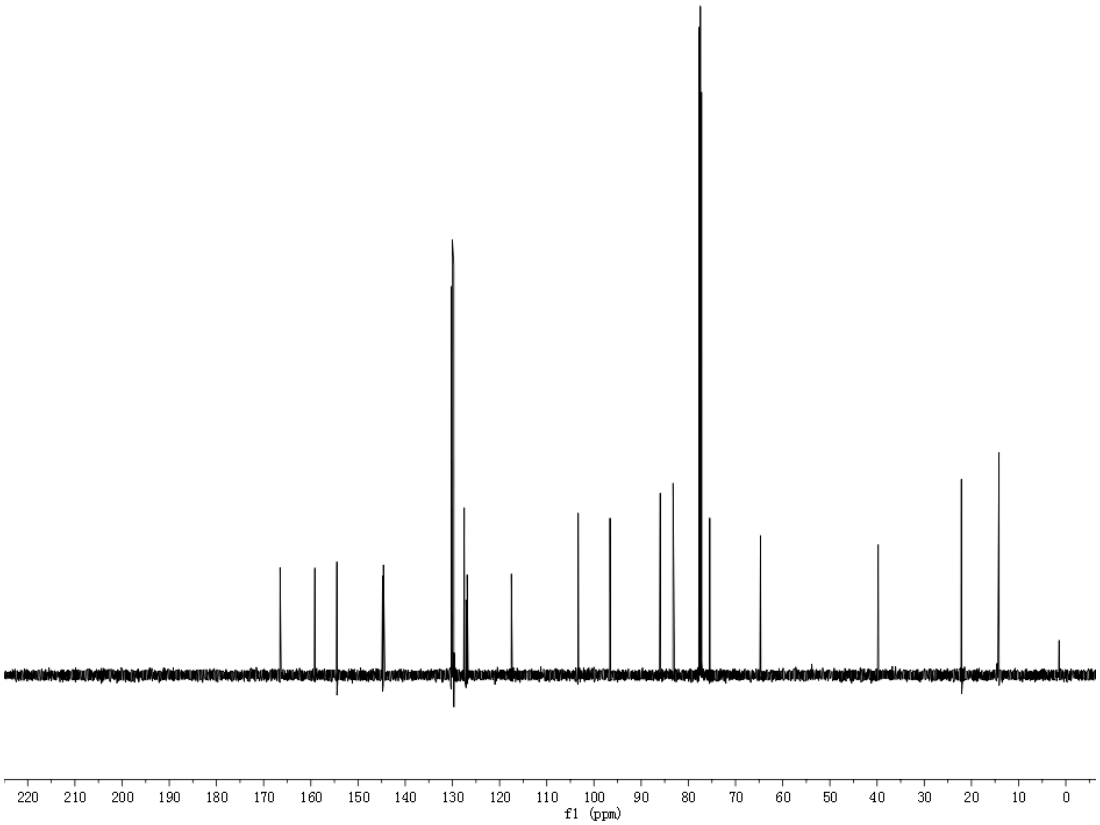
125 MHz ^{13}C NMR spectrum of **5a** ($CDCl_3$)



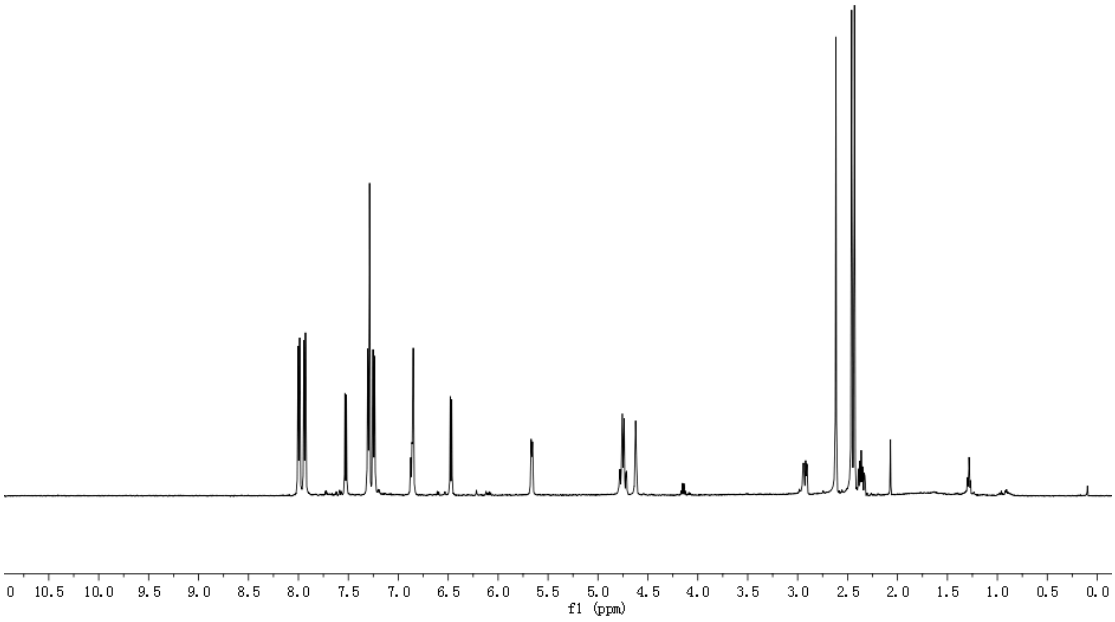
500 MHz ^{1}H NMR spectrum of **5b** ($CDCl_3$)



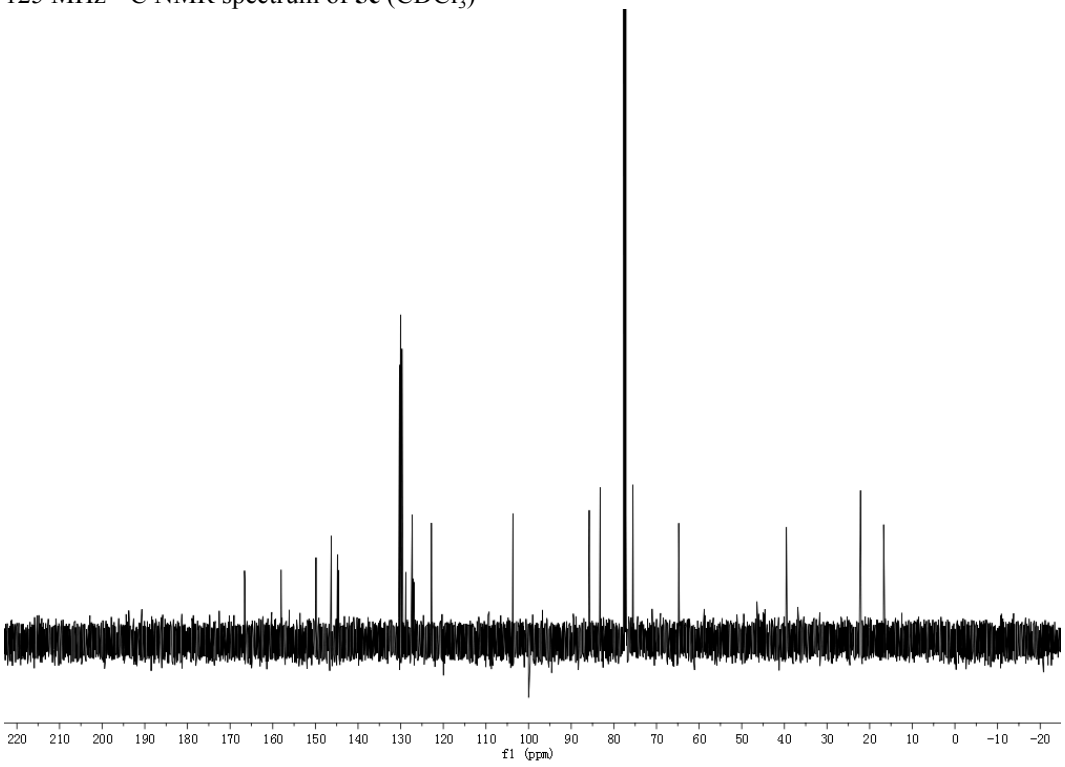
125 MHz ^{13}C NMR spectrum of **5b** ($CDCl_3$)



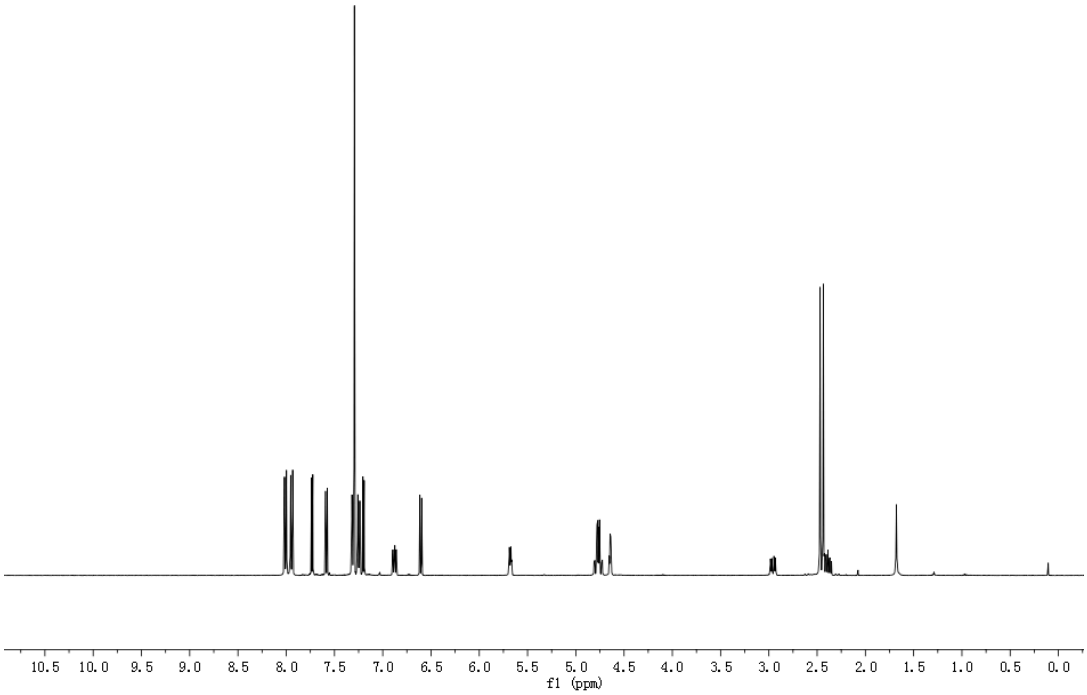
500 MHz ^{1}H NMR spectrum of **5c** ($CDCl_3$)



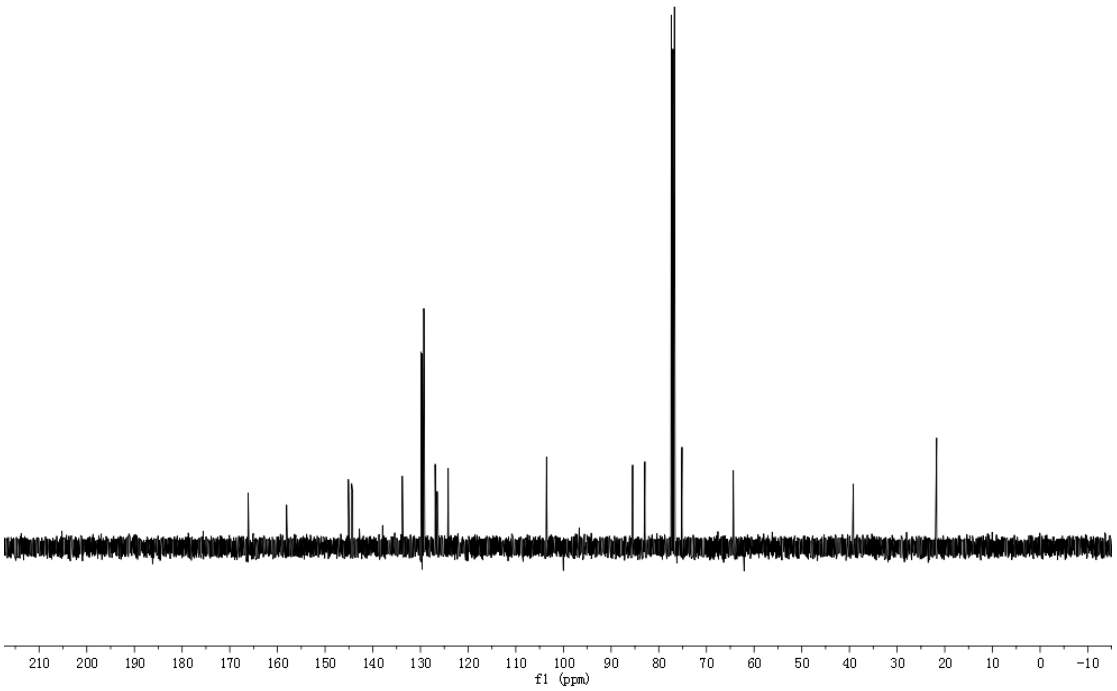
125 MHz ^{13}C NMR spectrum of **5c** ($CDCl_3$)



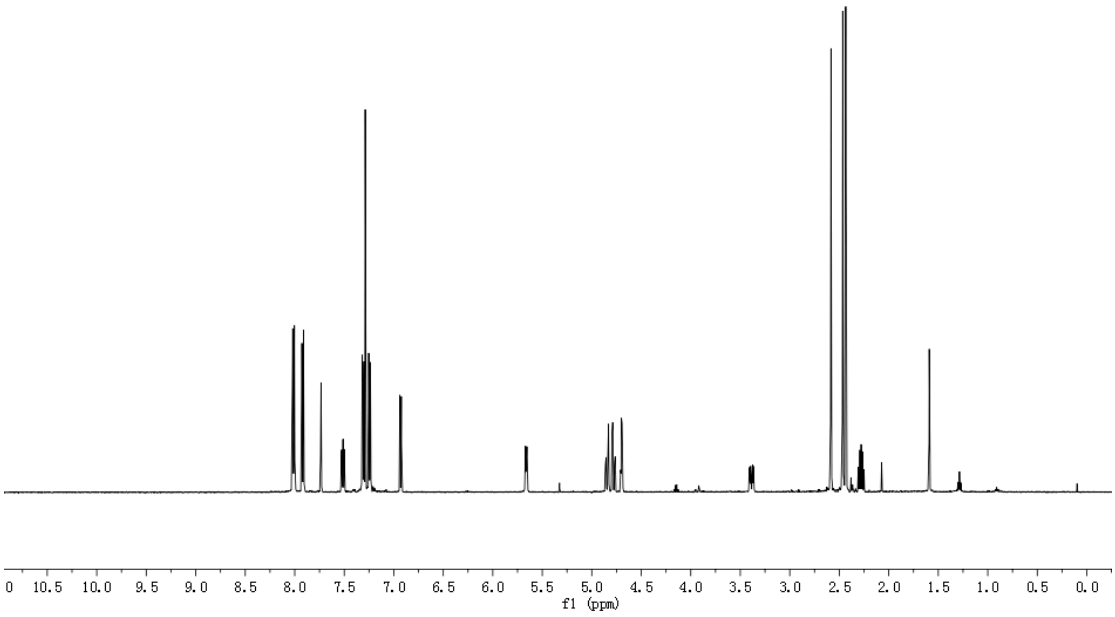
500 MHz ^{1}H NMR spectrum of **5d** ($CDCl_3$)



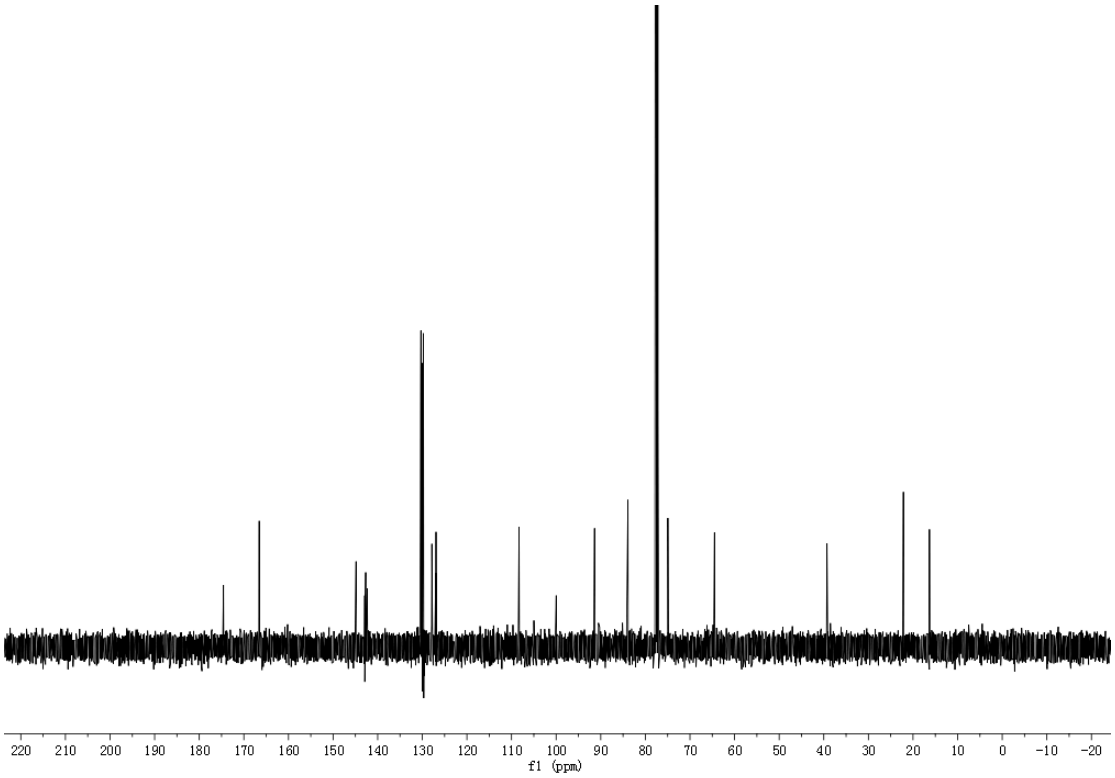
125 MHz ^{13}C NMR spectrum of **5d** ($CDCl_3$)



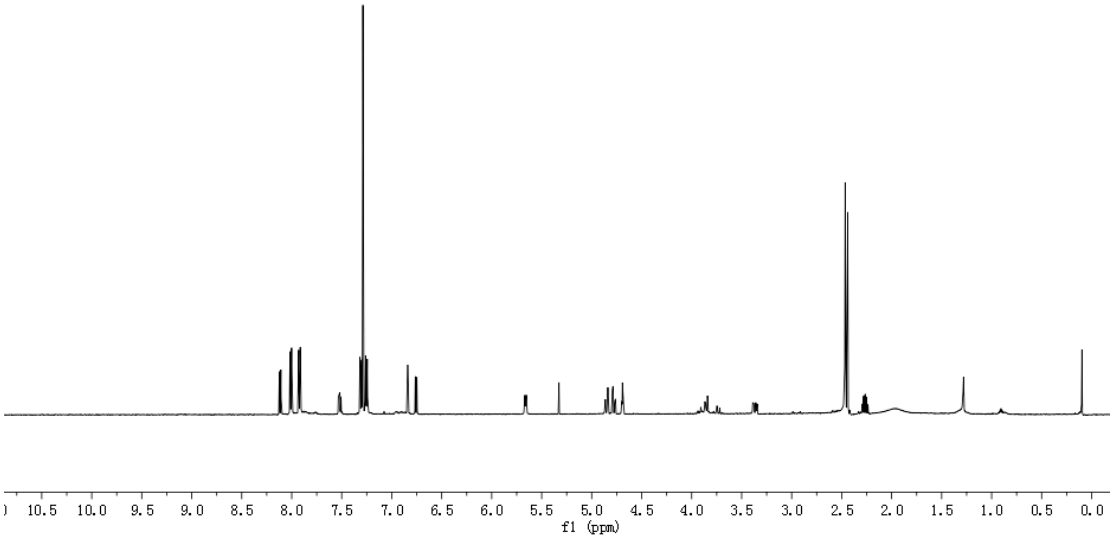
500 MHz ^{1}H NMR spectrum of **6a** ($CDCl_3$)



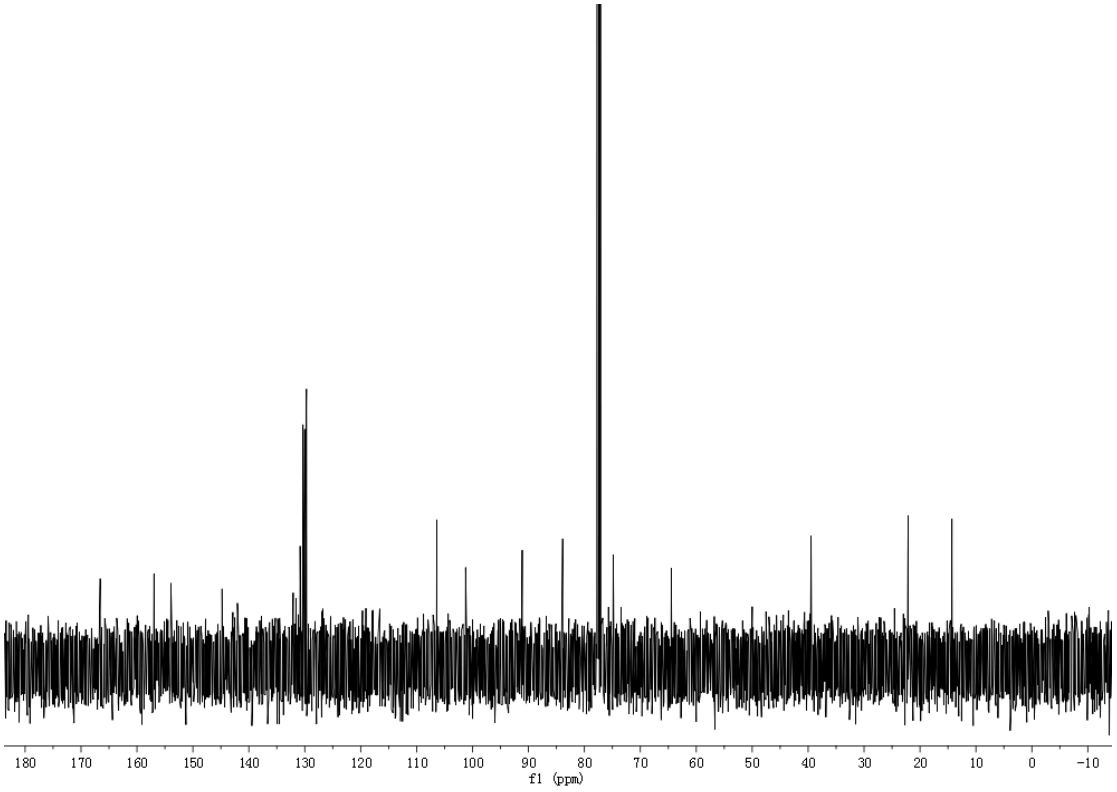
125 MHz ^{13}C NMR spectrum of **6a** ($CDCl_3$)



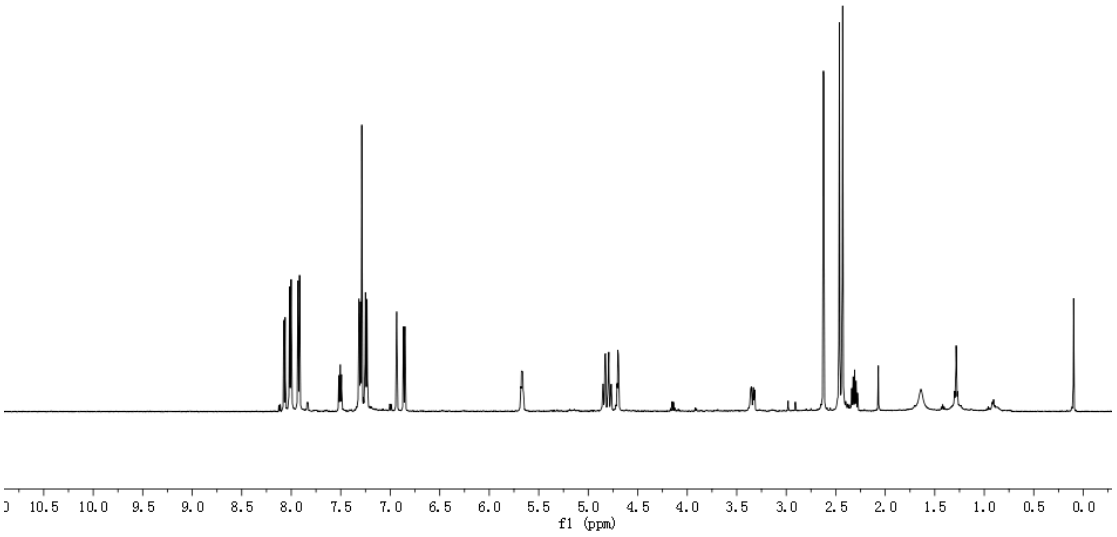
500 MHz ^{1}H NMR spectrum of **6b** ($CDCl_3$)



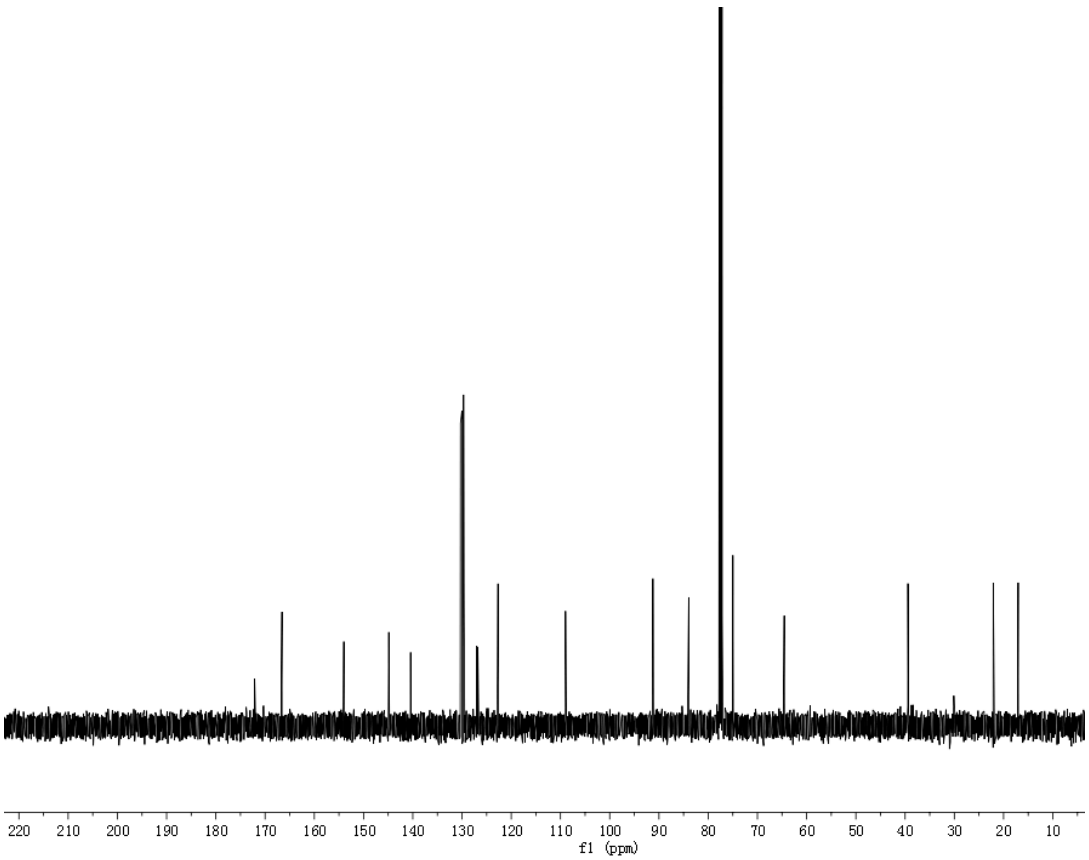
125 MHz ^{13}C NMR spectrum of **6b** ($CDCl_3$)



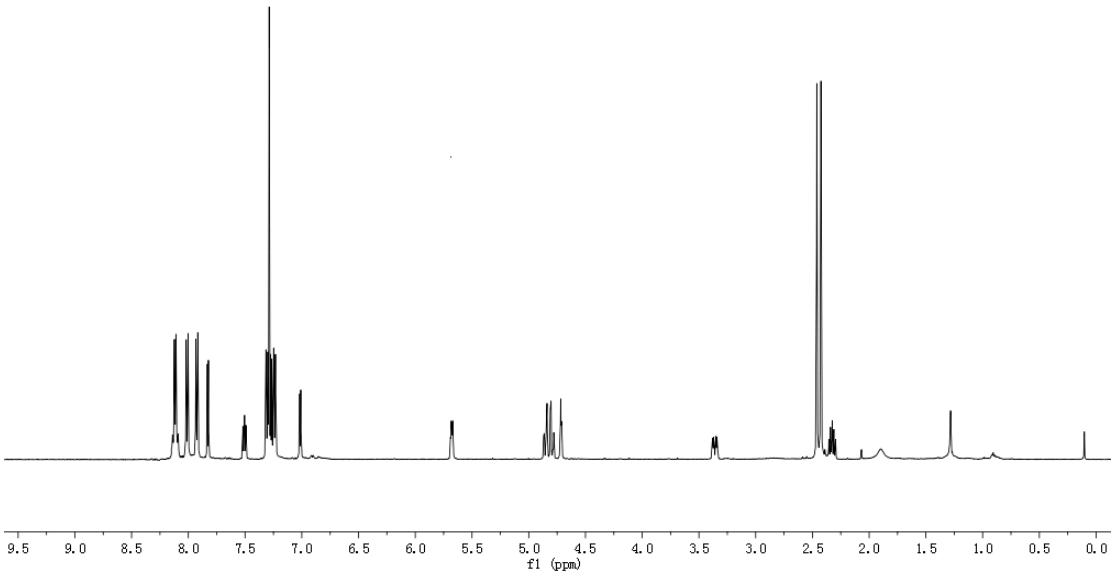
500 MHz ^{1}H NMR spectrum of **6c** ($CDCl_3$)



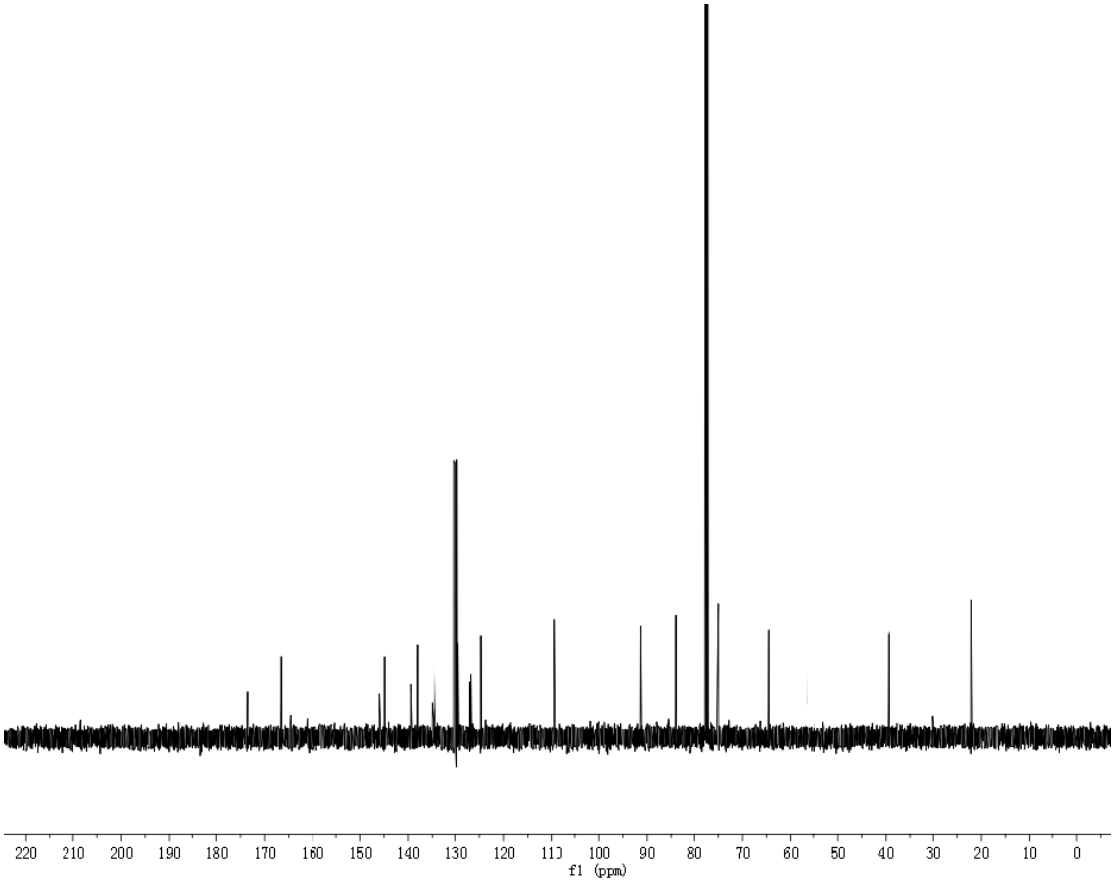
125 MHz ^{13}C NMR spectrum of **6c** ($CDCl_3$)



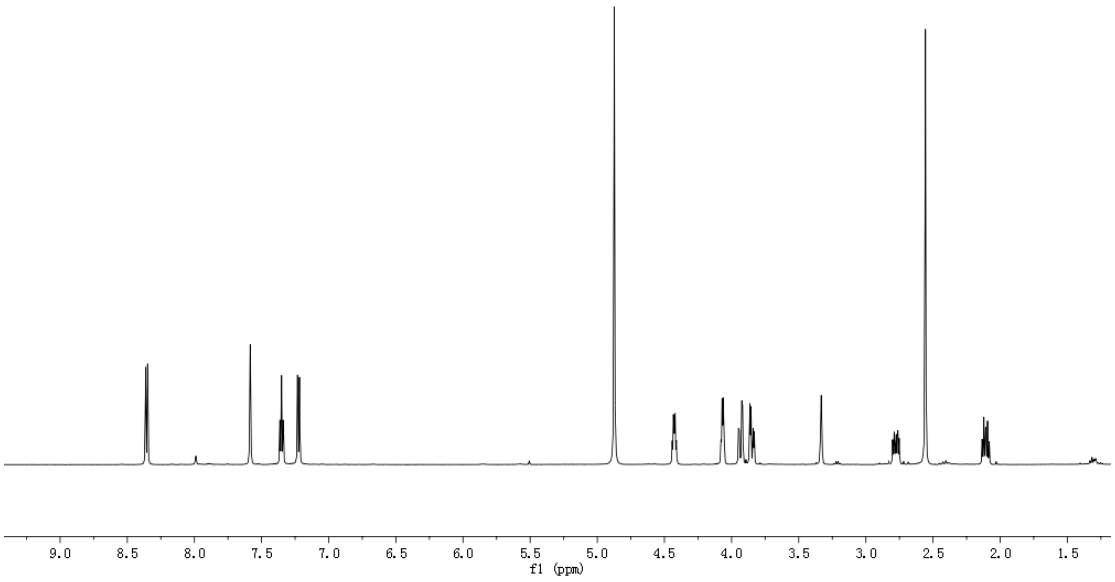
500 MHz ^{1}H NMR spectrum of **6d** ($CDCl_3$)



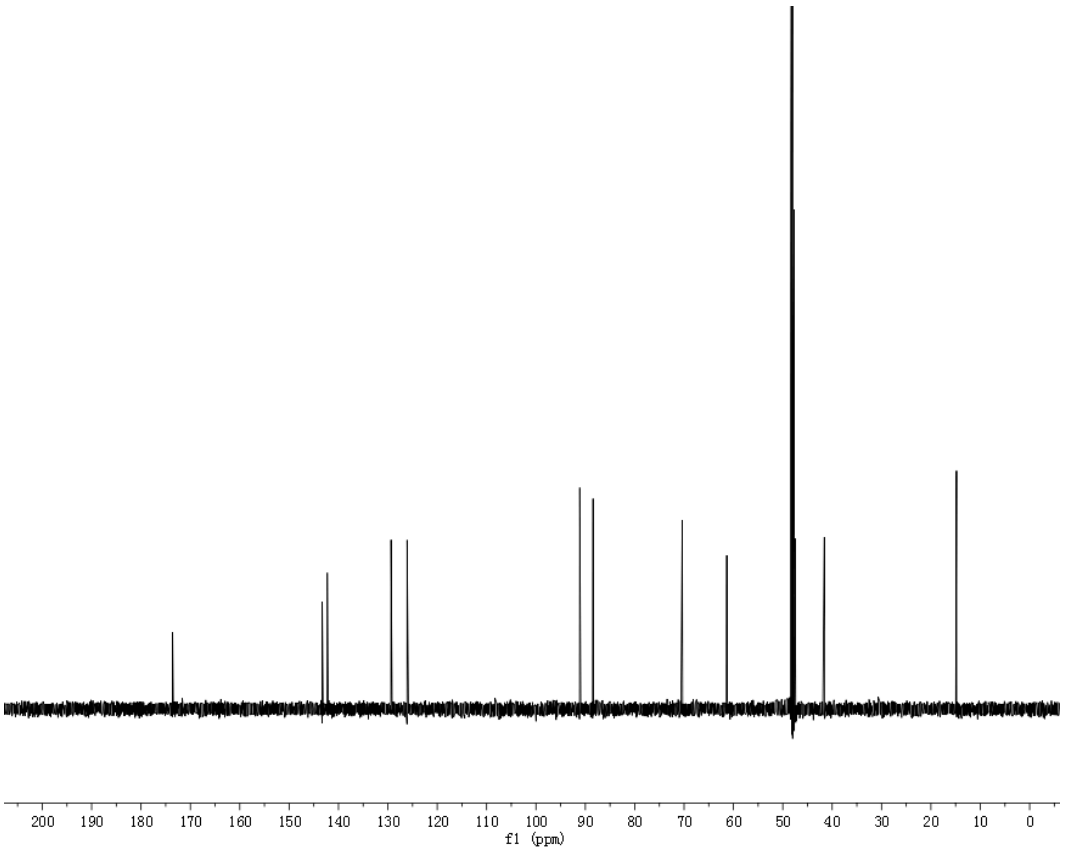
125 MHz ^{13}C NMR spectrum of **6d** ($CDCl_3$)



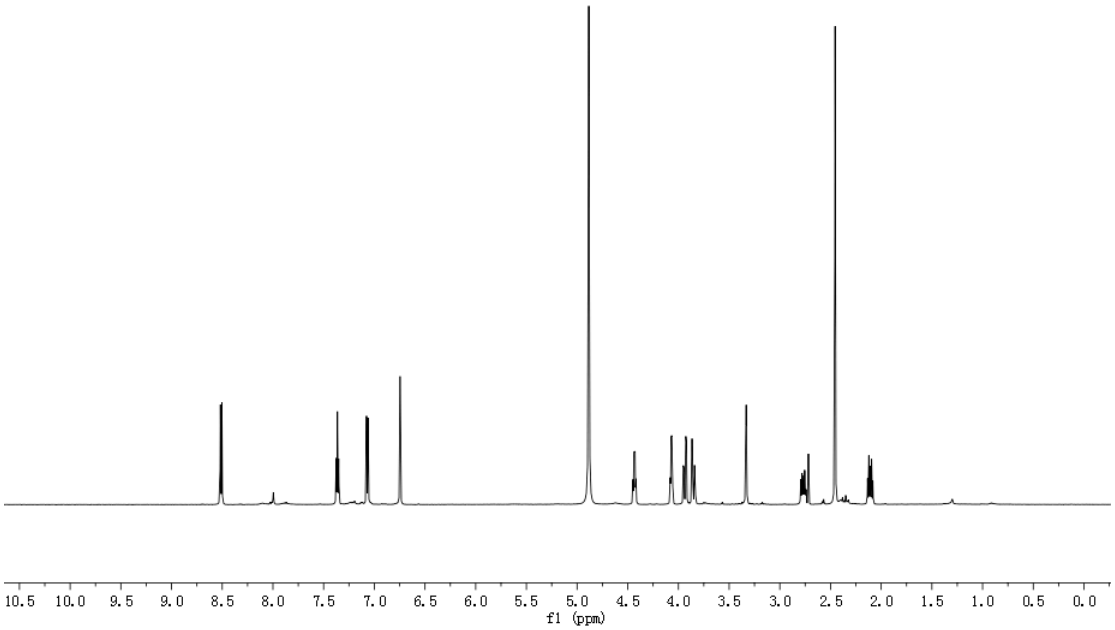
500 MHz ^{1}H NMR spectrum of **7a** (CD_3OD)



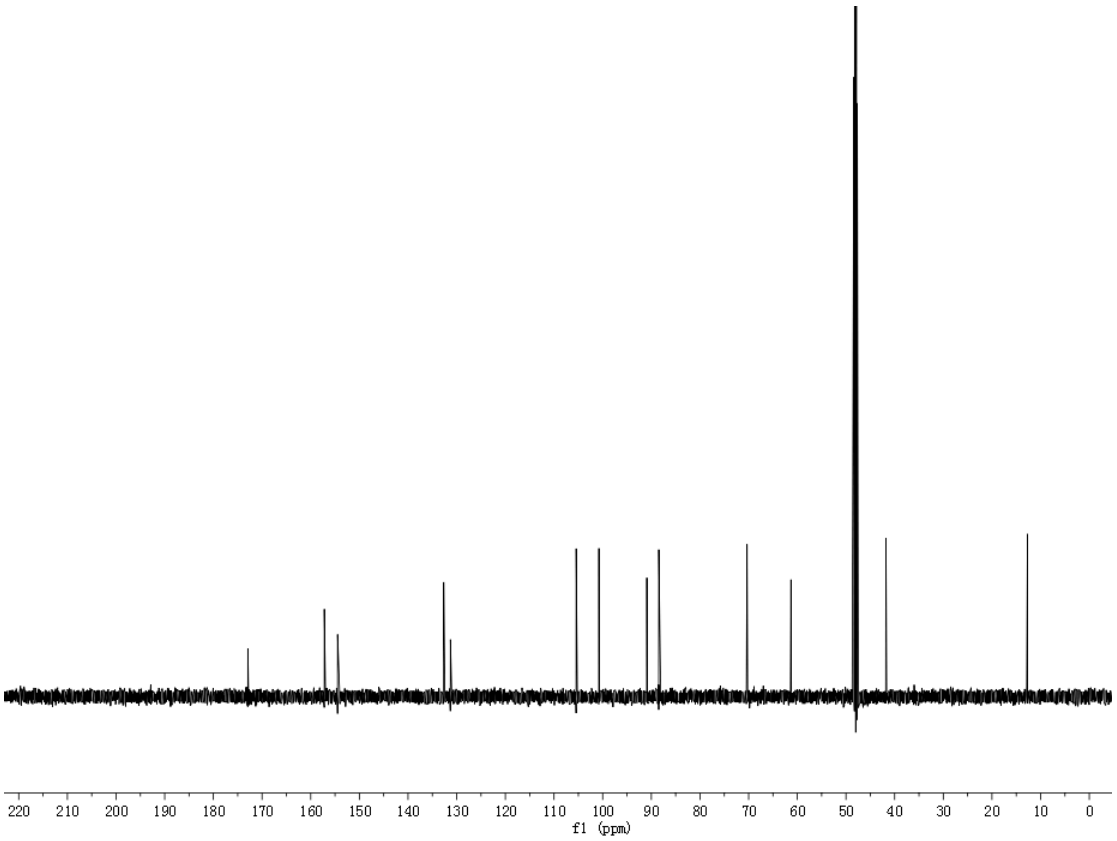
125 MHz ^{13}C NMR spectrum of **7a** (CD_3OD)



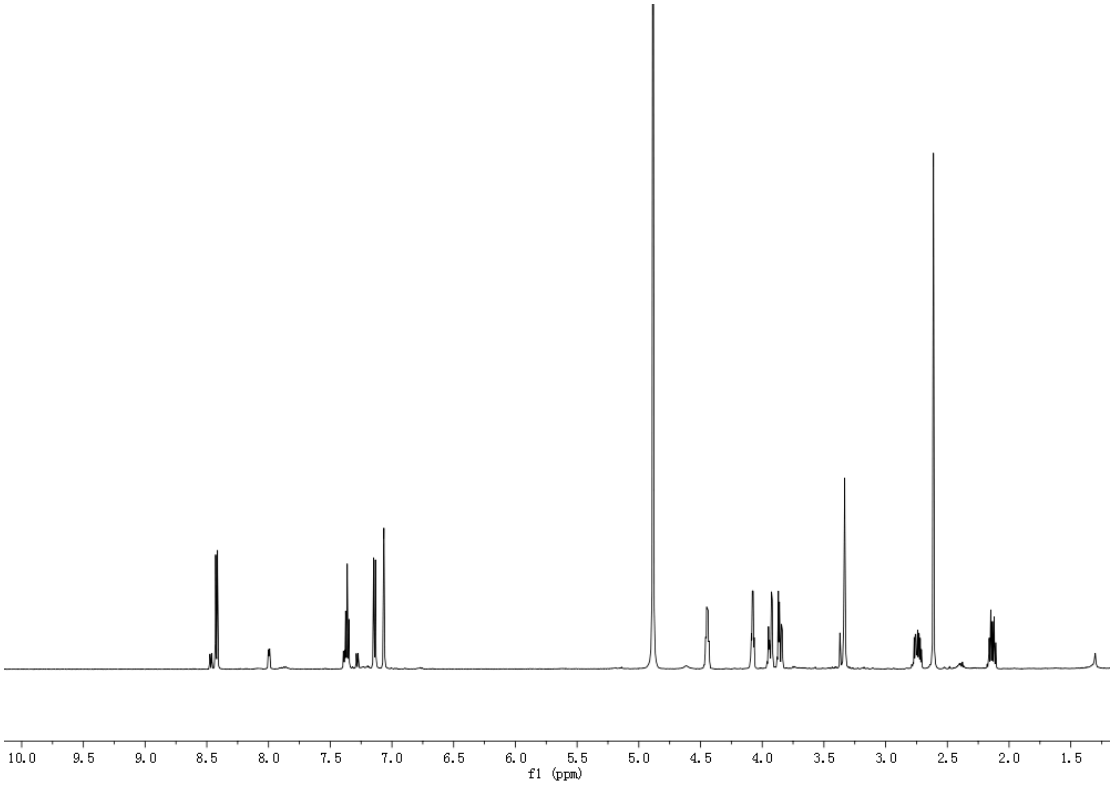
500 MHz ^{1}H NMR spectrum of **7b** (CD_3OD)



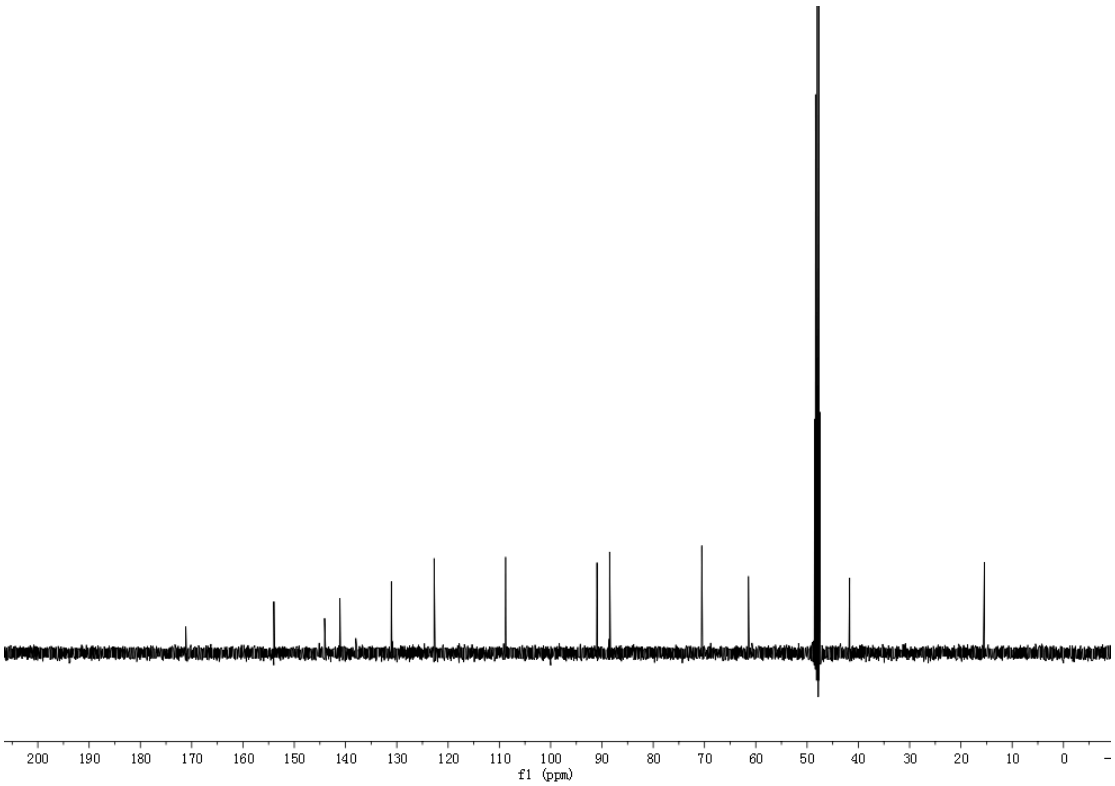
125 MHz ^{13}C NMR spectrum of **7b** (CD_3OD)



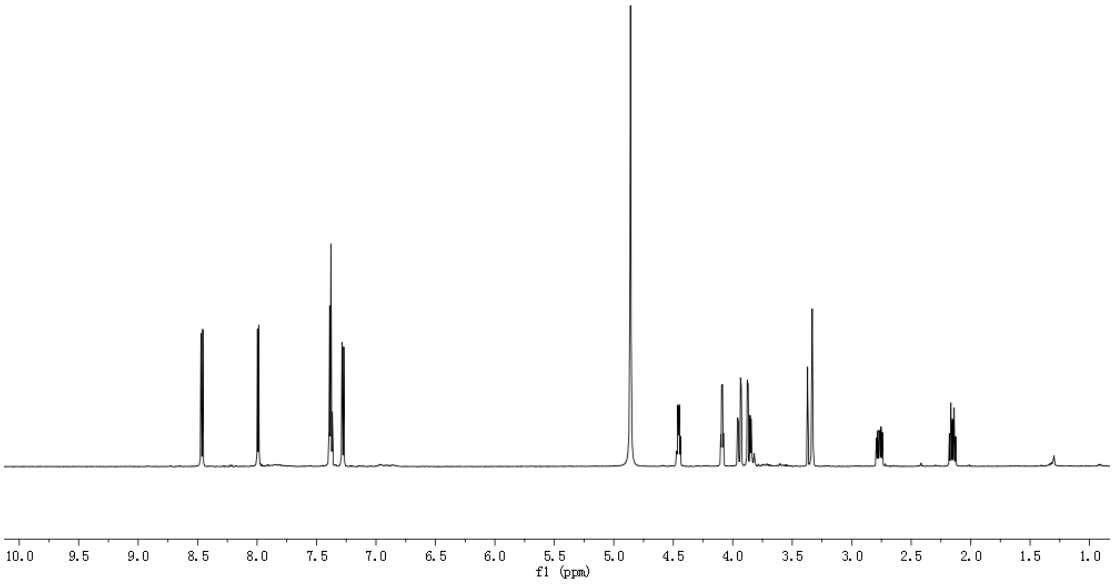
500 MHz ^{1}H NMR spectrum of **7c** (CD_3OD)



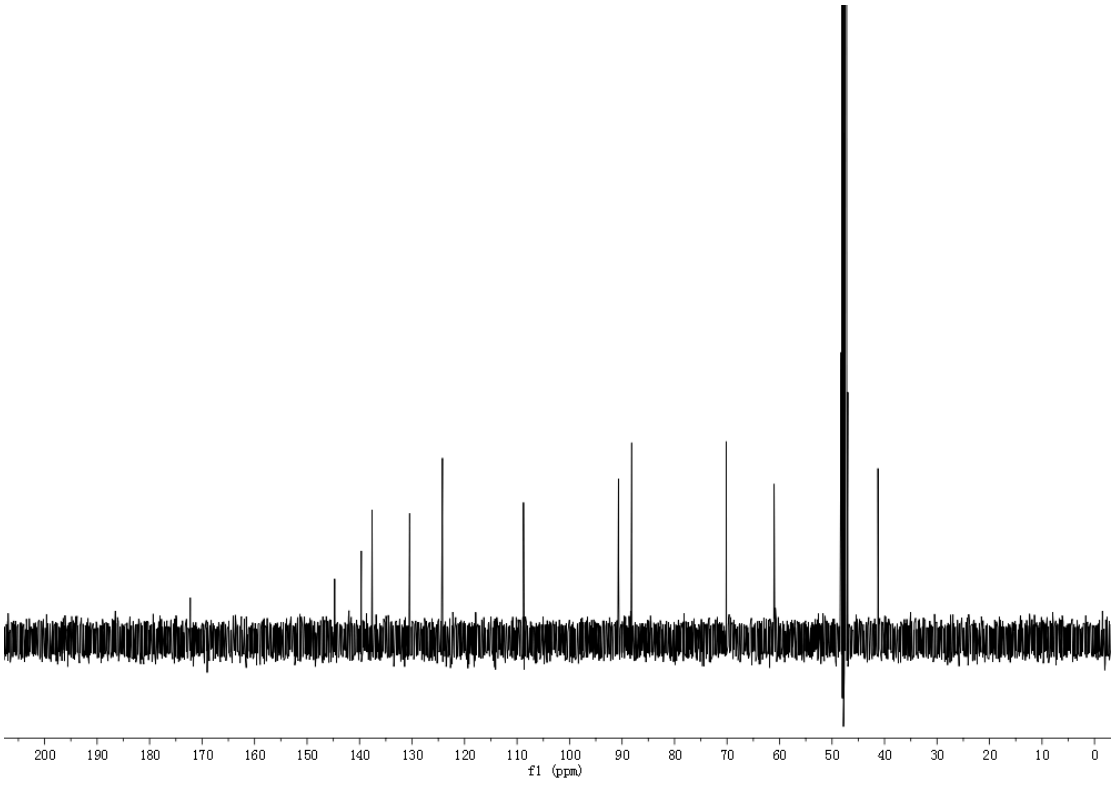
125 MHz ^{13}C NMR spectrum of **7c** (CD_3OD)



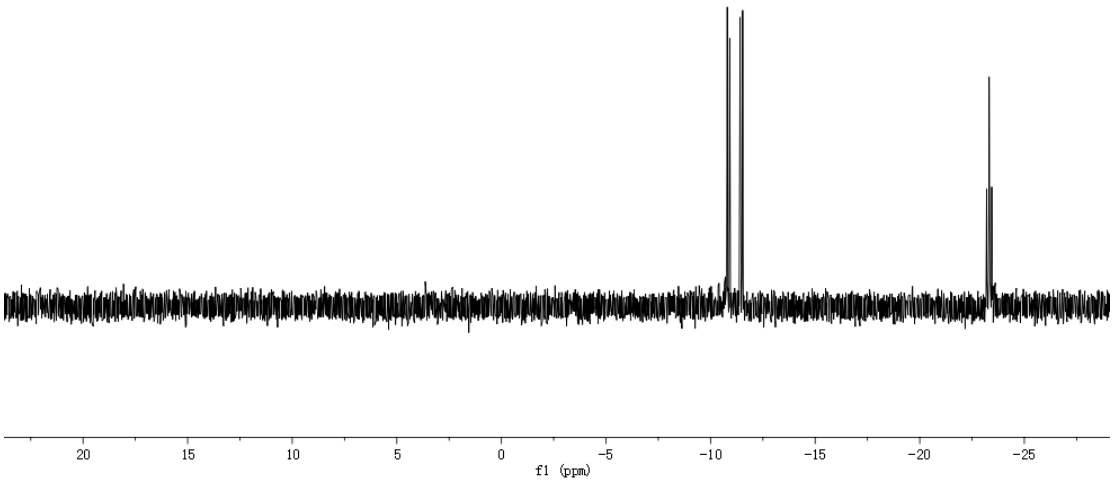
500 MHz ^{1}H NMR spectrum of **7d** (CD_3OD)



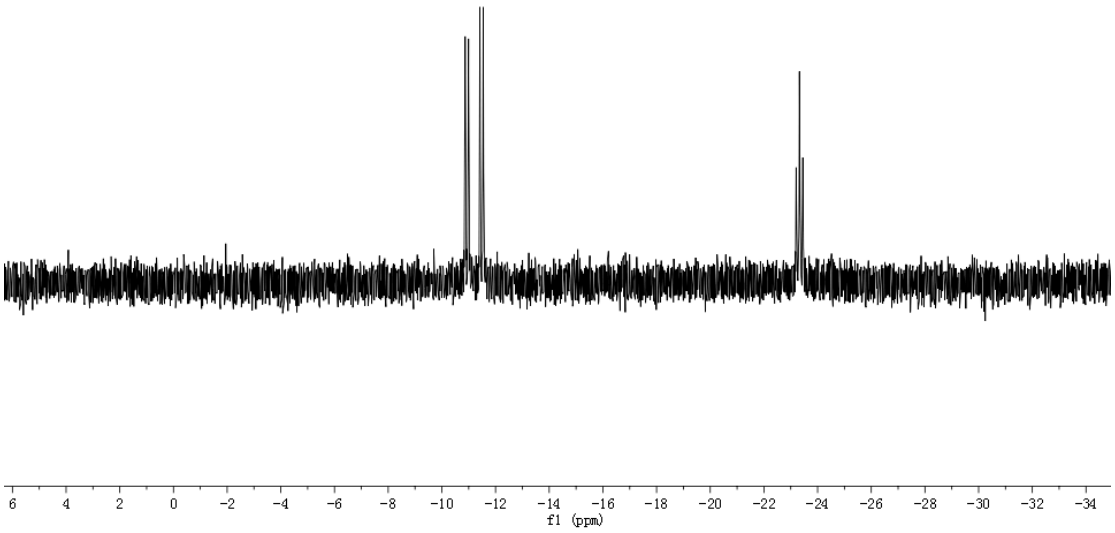
125 MHz ^{13}C NMR spectrum of **7d** (CD_3OD)



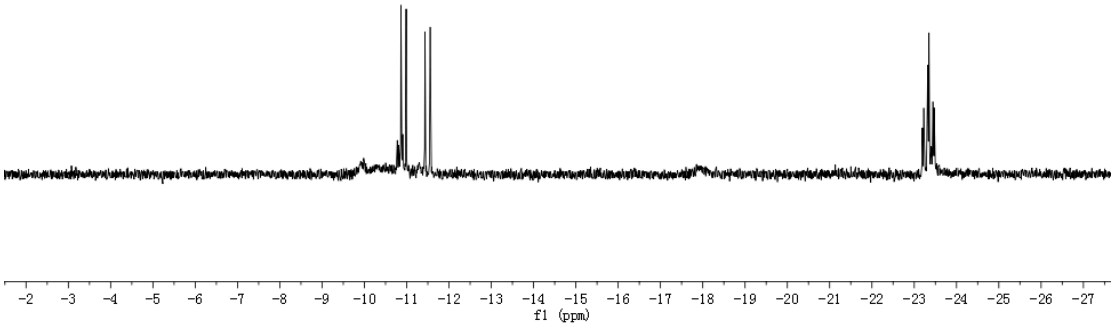
162 MHz ^{31}P NMR spectrum of **8a** (D_2O)



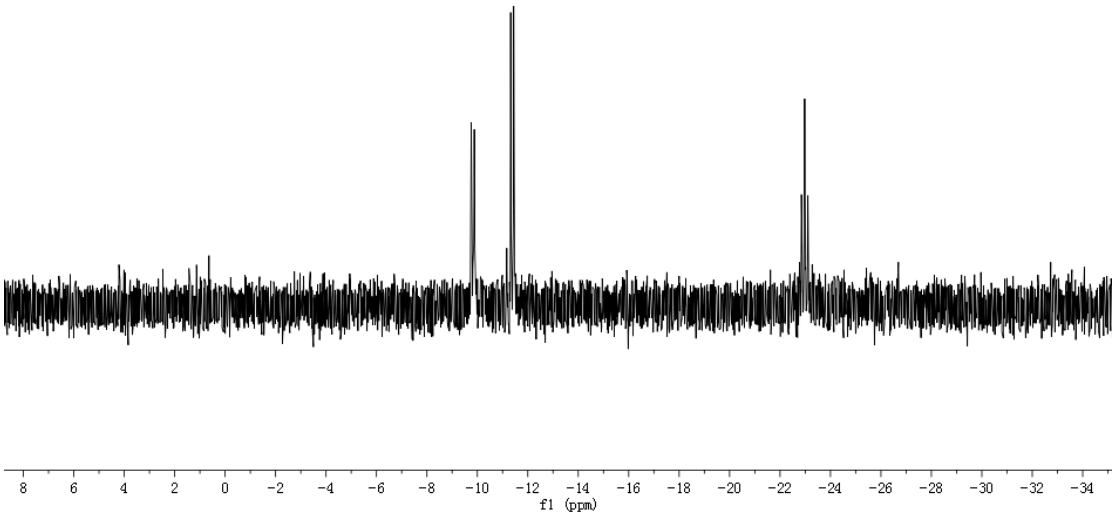
162 MHz ^{31}P NMR spectrum of **8b** (D_2O)



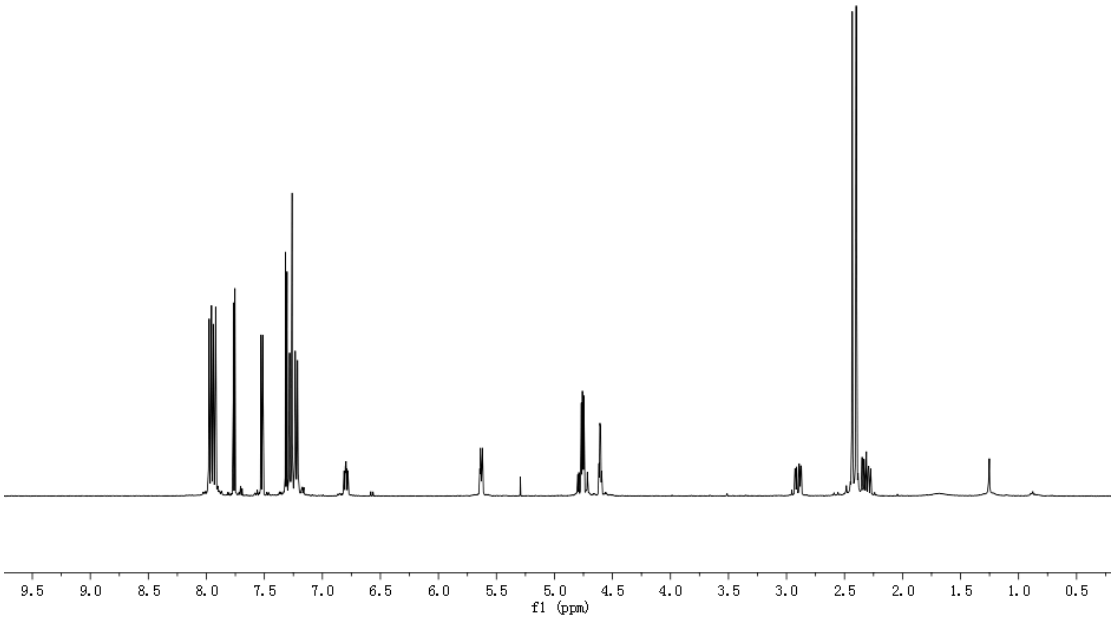
162 MHz ^{31}P NMR spectrum of **8c** (D_2O)



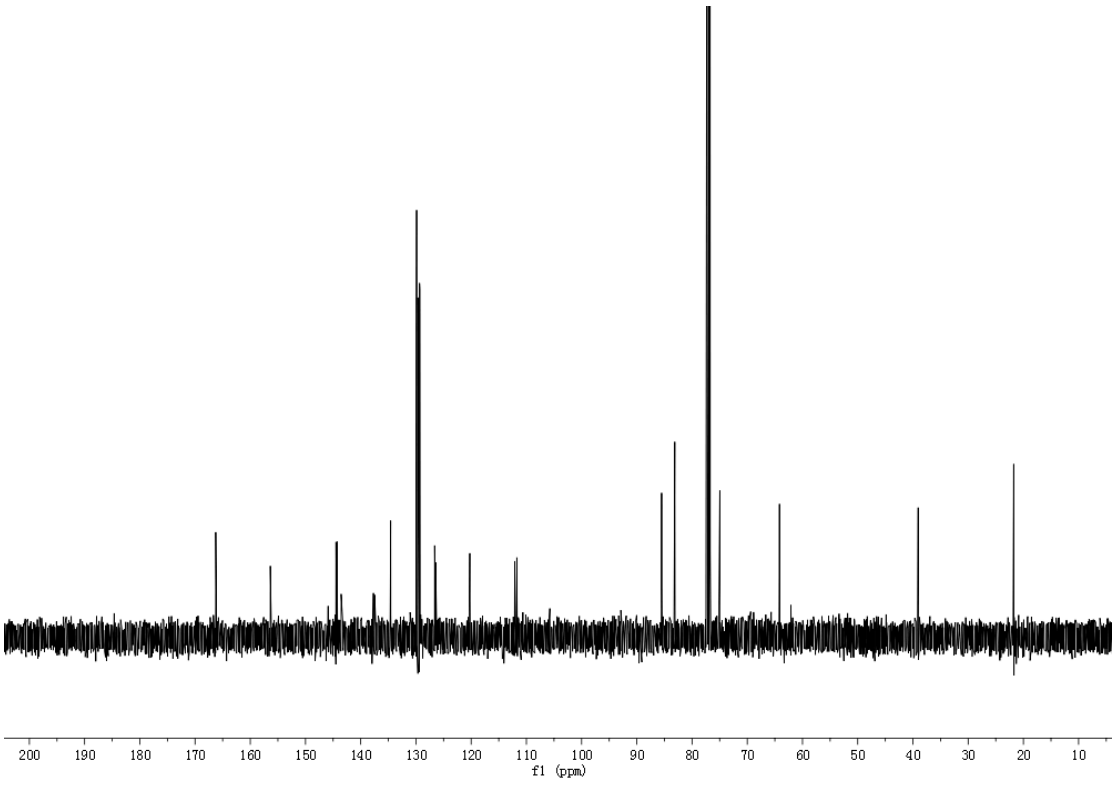
162 MHz ^{31}P NMR spectrum of **8d** (D_2O)



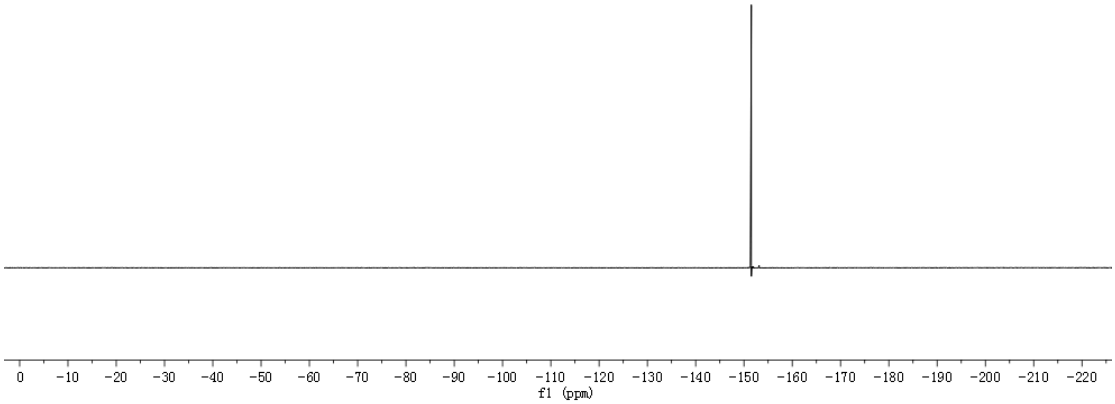
500 MHz ^{1}H NMR spectrum of **9** ($CDCl_3$)



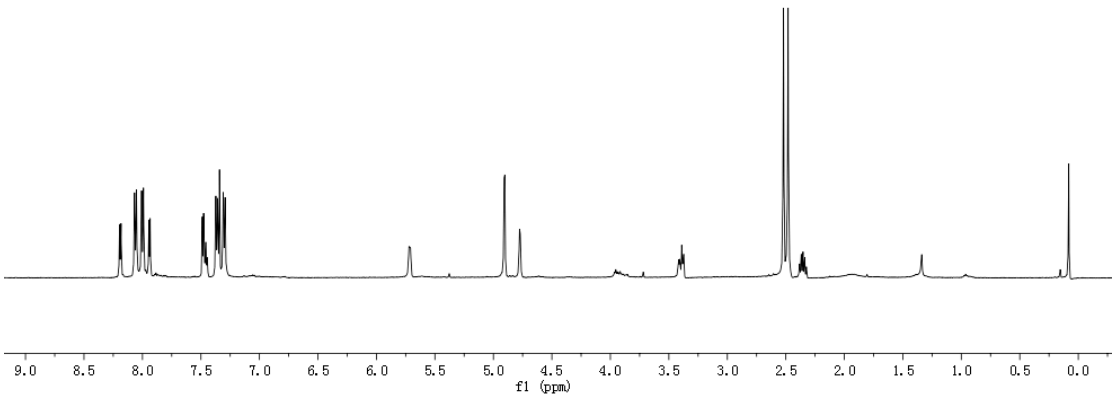
125 MHz ^{13}C NMR spectrum of **9** ($CDCl_3$)



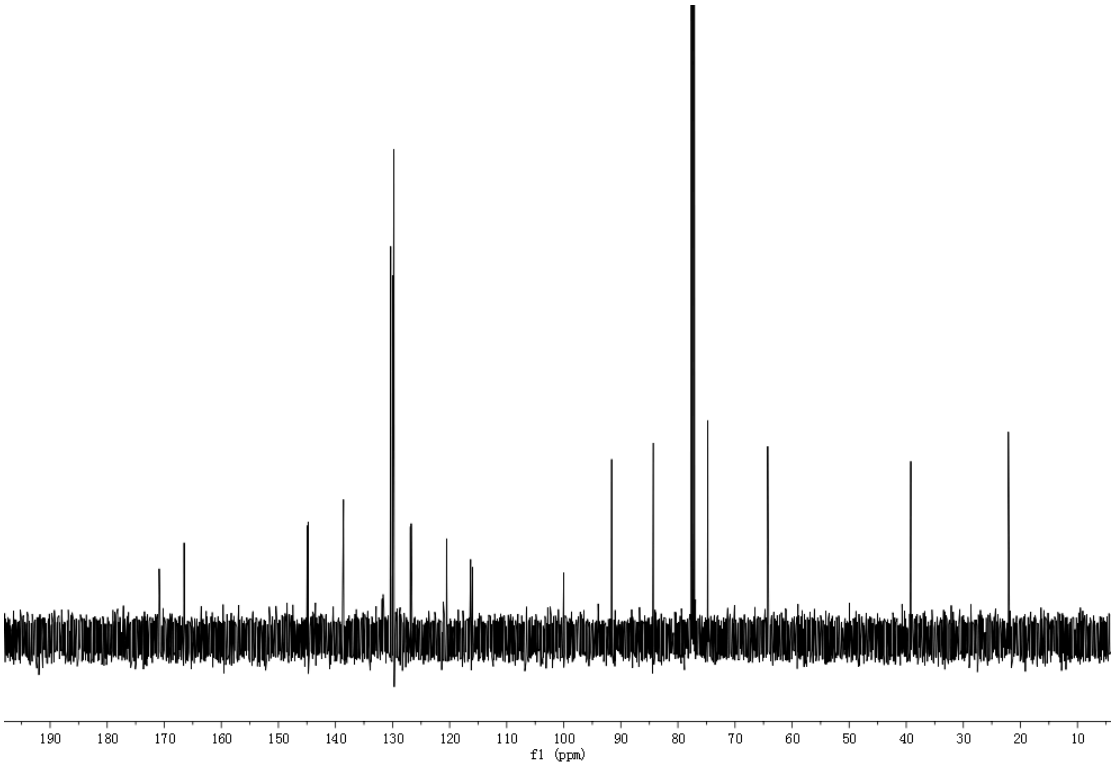
376 MHz ^{19}F NMR spectrum of **9** ($CDCl_3$)



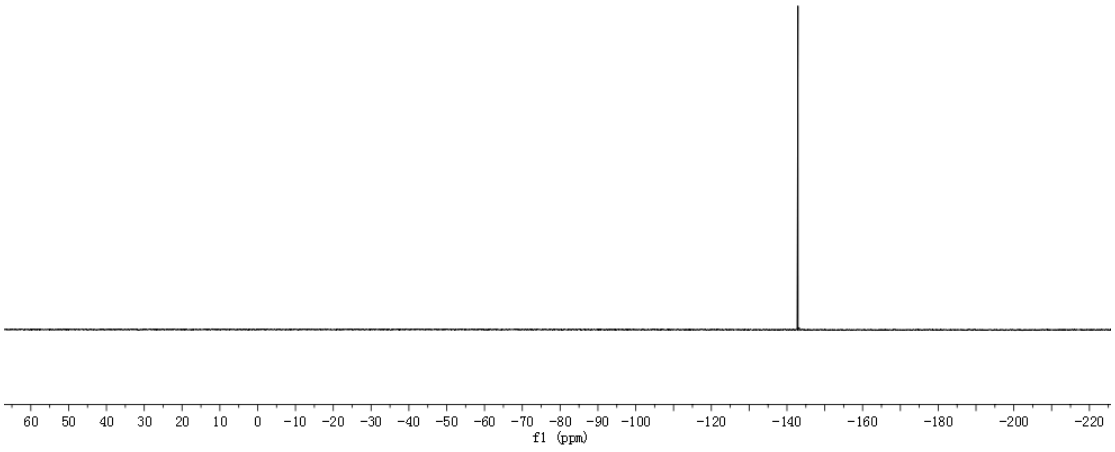
500 MHz ^{1}H NMR spectrum of **10** ($CDCl_3$)



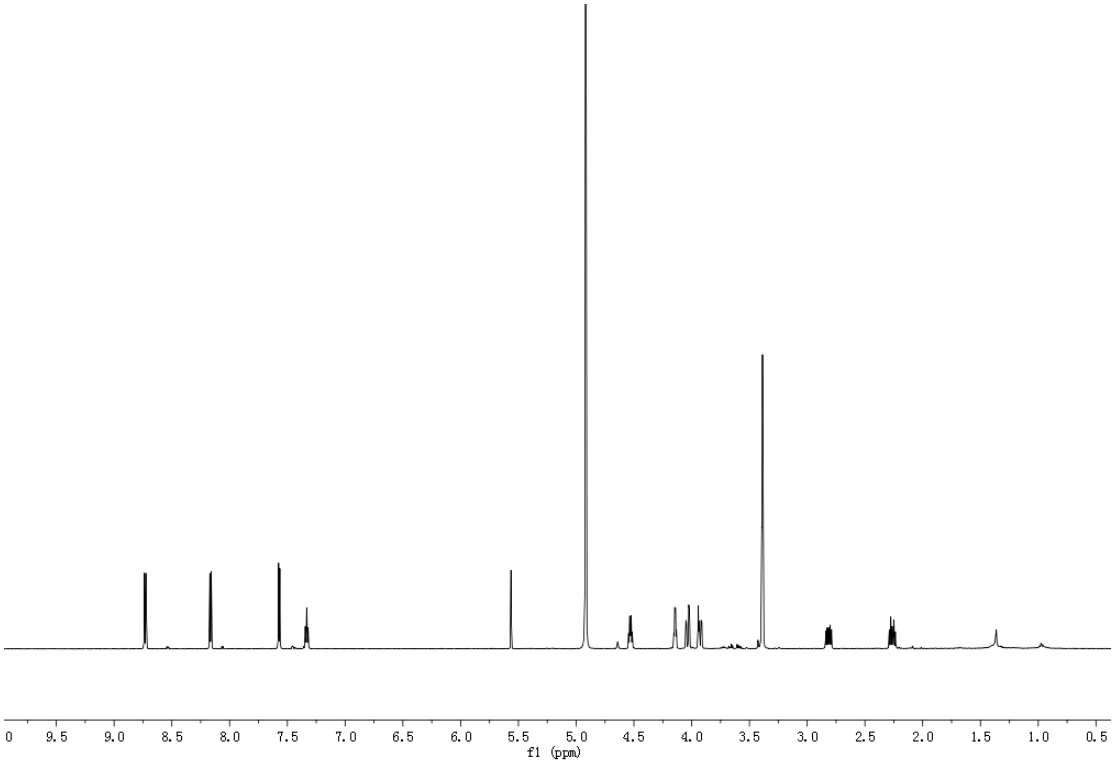
125 MHz ^{13}C NMR spectrum of **10** ($CDCl_3$)



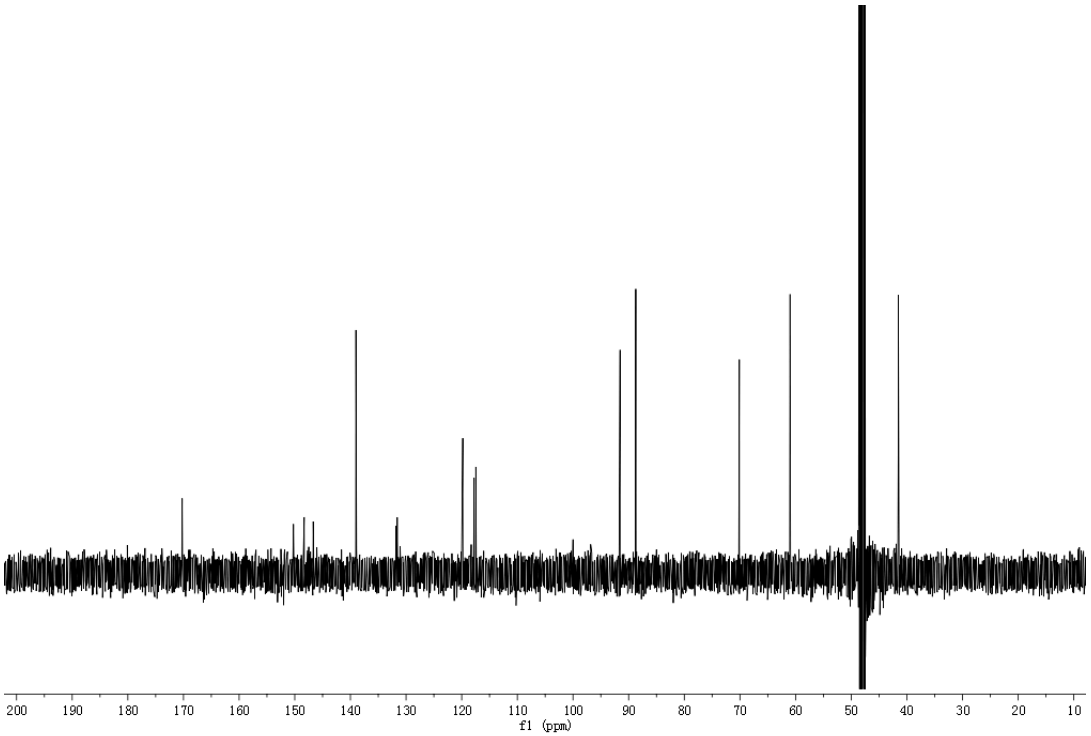
376 MHz ^{19}F NMR spectrum of **10** ($CDCl_3$)



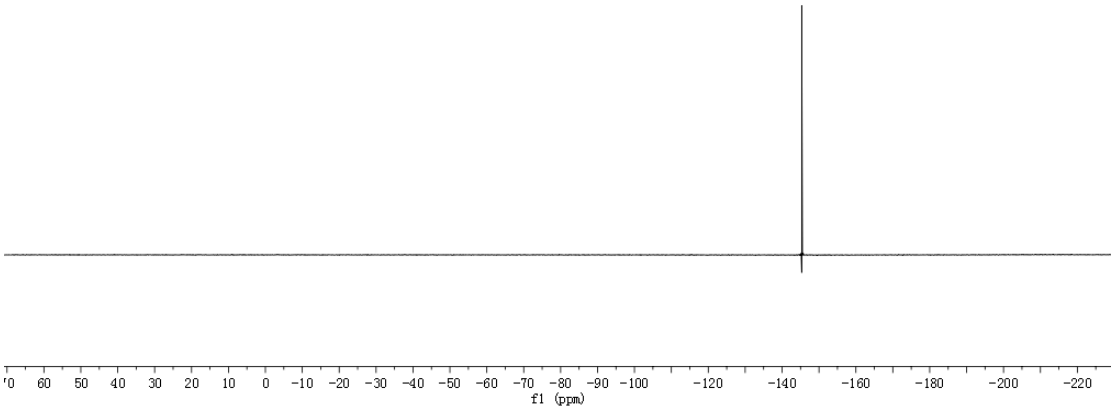
500 MHz ^{1}H NMR spectrum of **11** (CD_3OD)



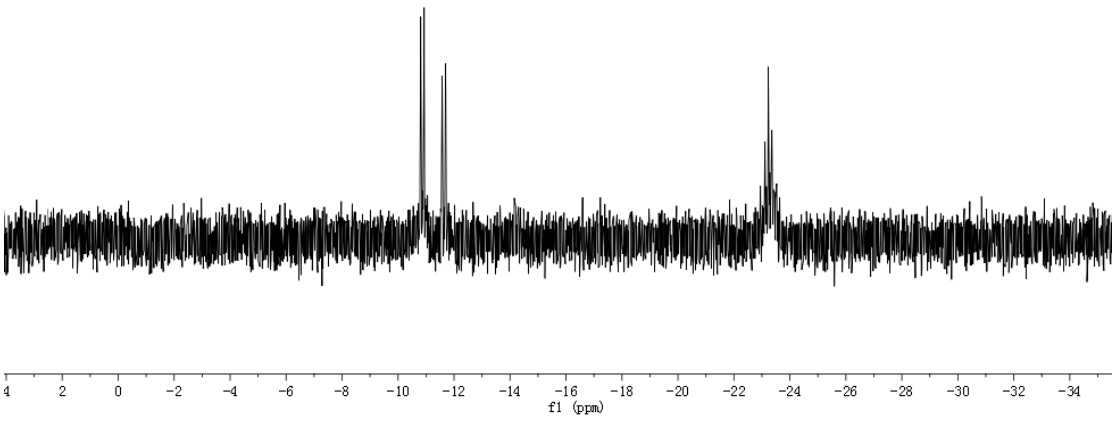
125 MHz ^{13}C NMR spectrum of **11** (CD_3OD)



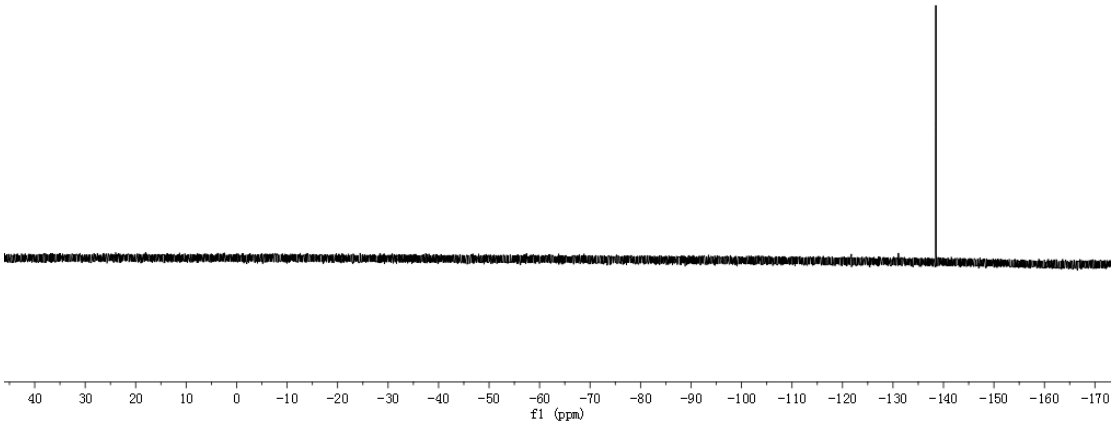
376 MHz ^{19}F NMR spectrum of **11** (CD_3OD)



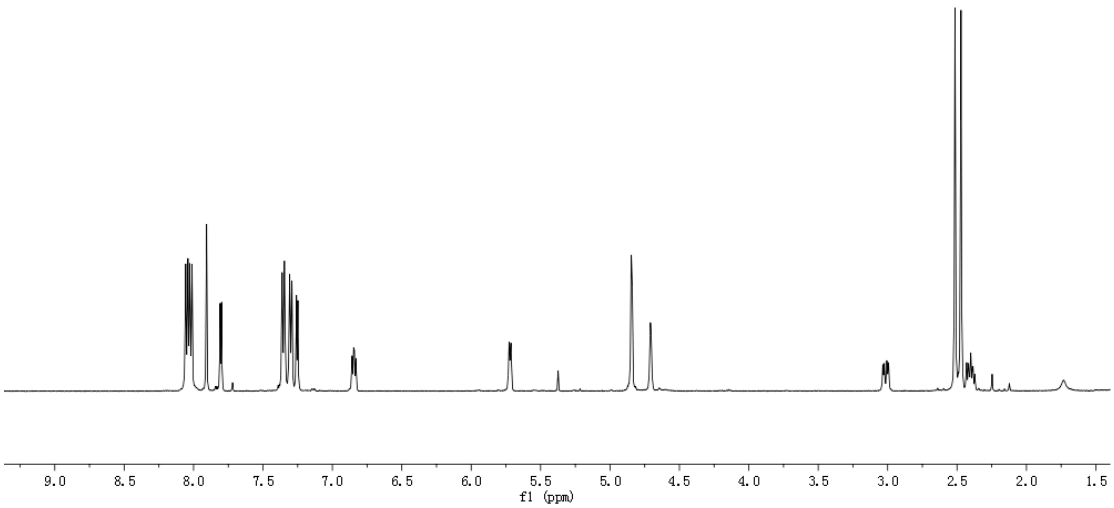
162 MHz ^{31}P NMR spectrum of **12** (D_2O)



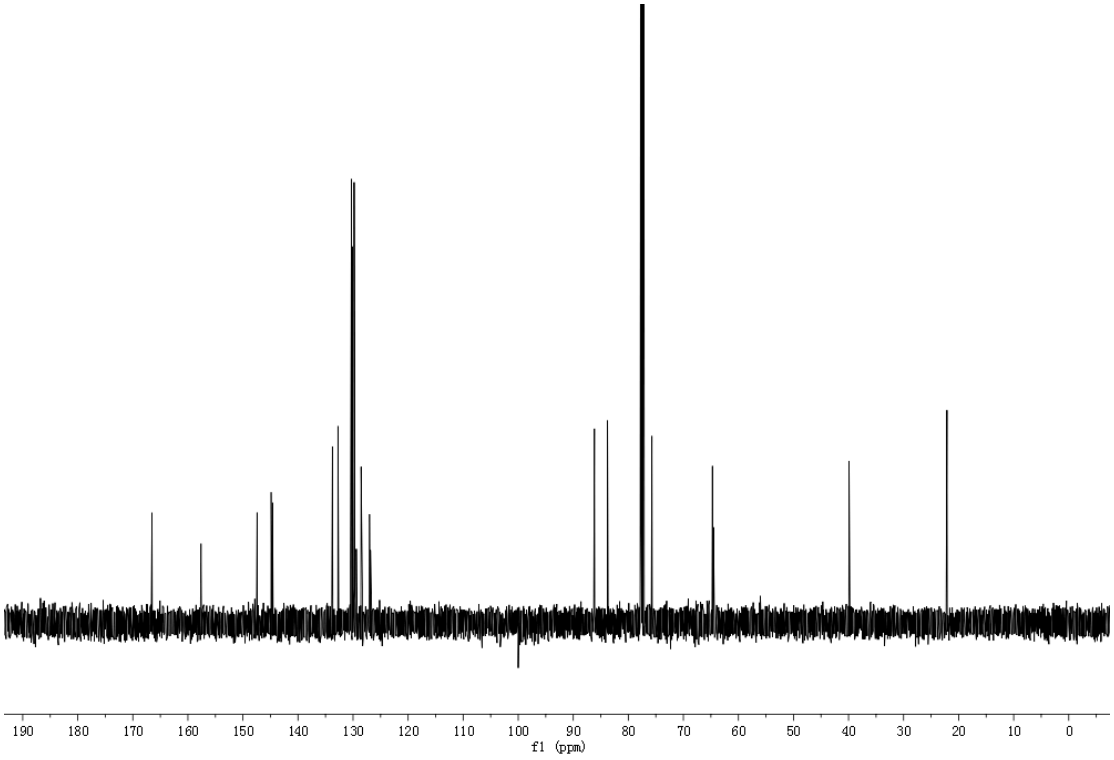
376 MHz ^{19}F NMR spectrum of **12** (D_2O)



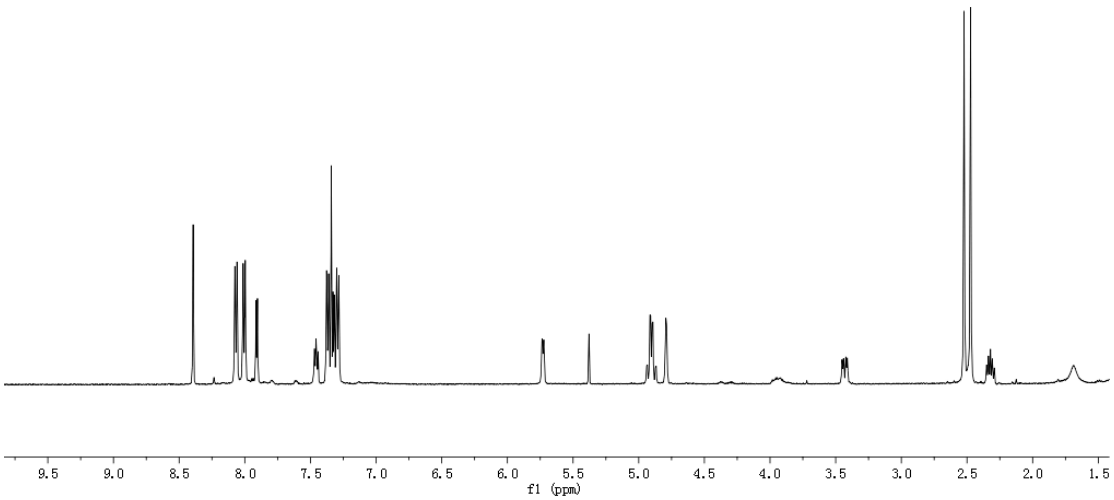
500 MHz ^{1}H NMR spectrum of **13** ($CDCl_3$)



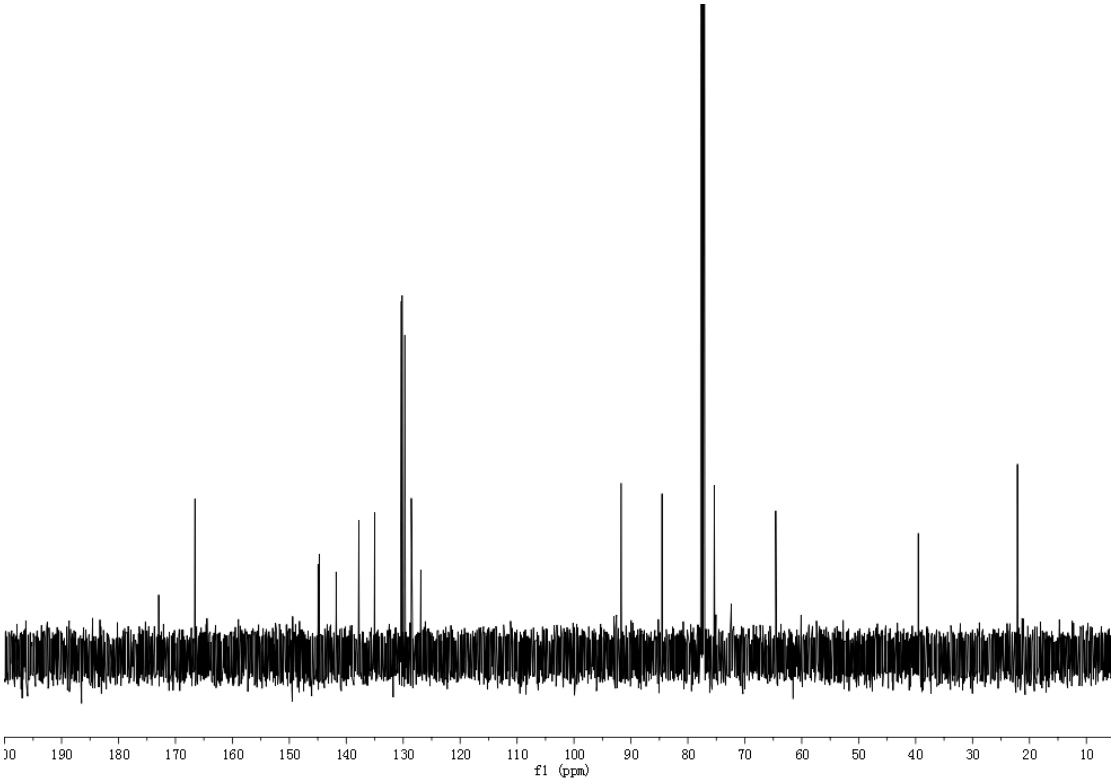
125 MHz ^{13}C NMR spectrum of **13** ($CDCl_3$)



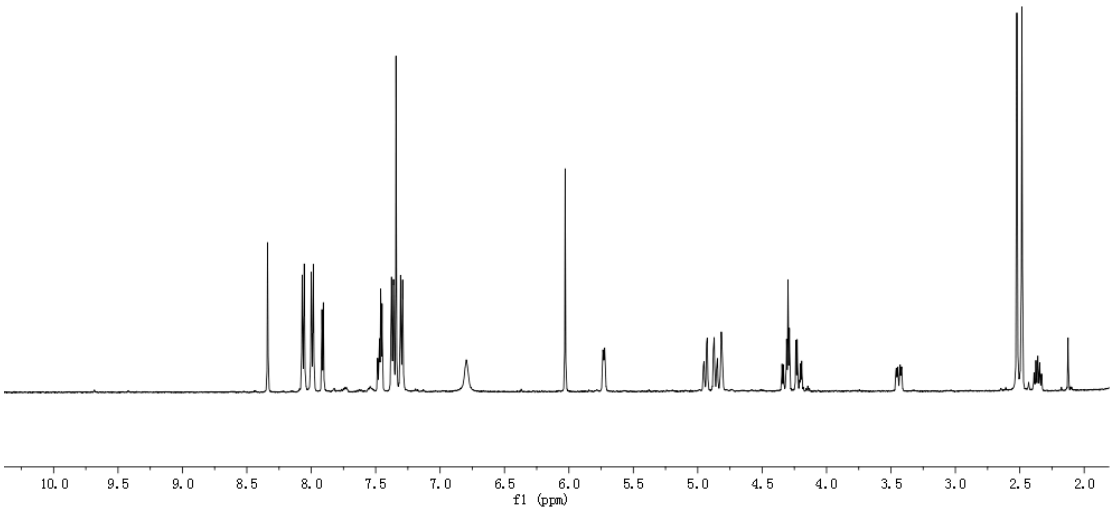
500 MHz ^{1}H NMR spectrum of **14** ($CDCl_3$)



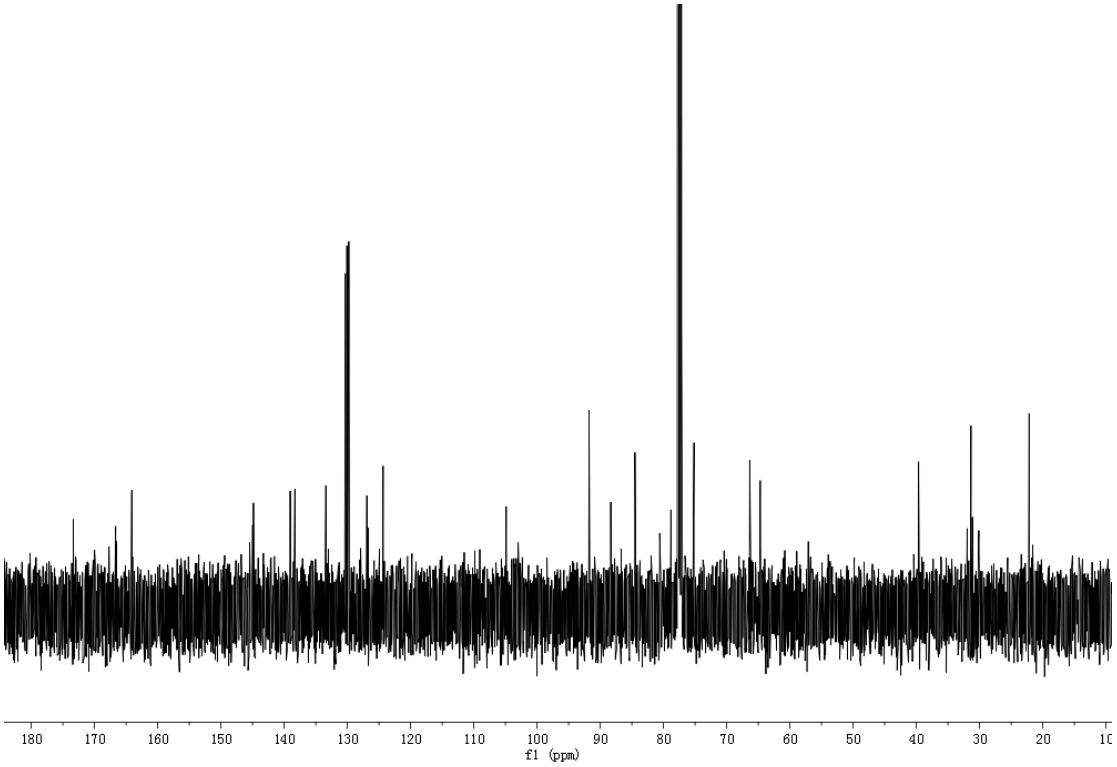
125 MHz ^{13}C NMR spectrum of **14** ($CDCl_3$)



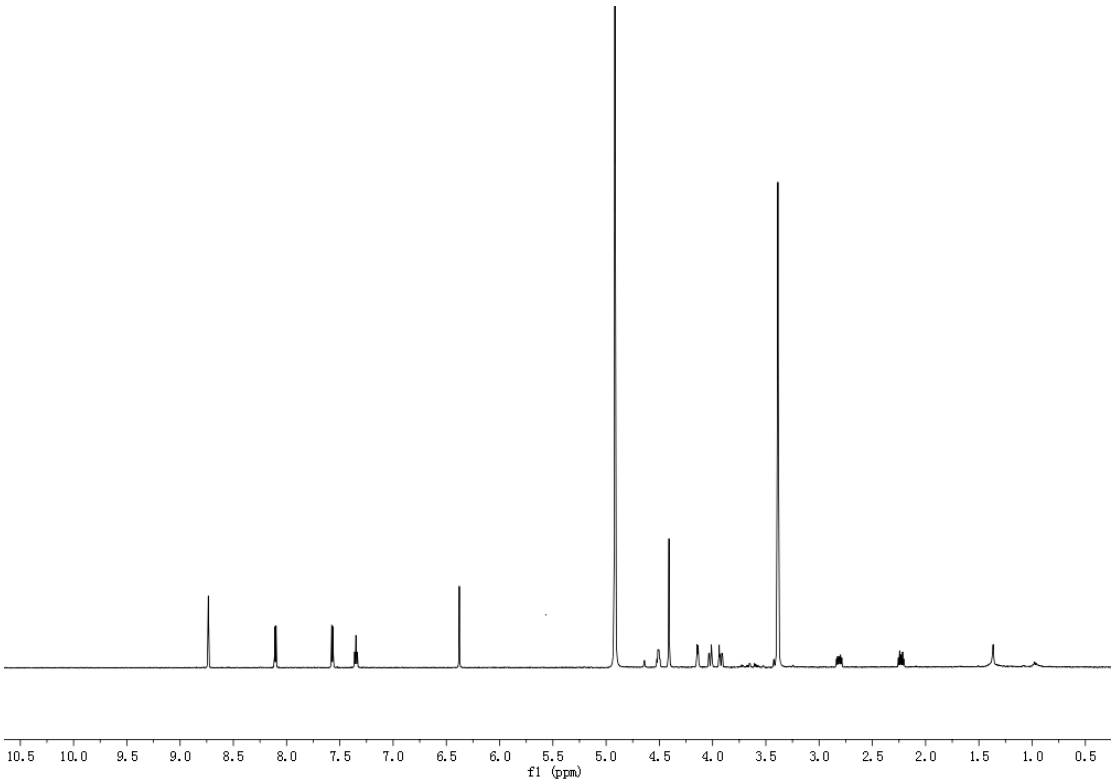
500 MHz ^{1}H NMR spectrum of **15** ($CDCl_3$)



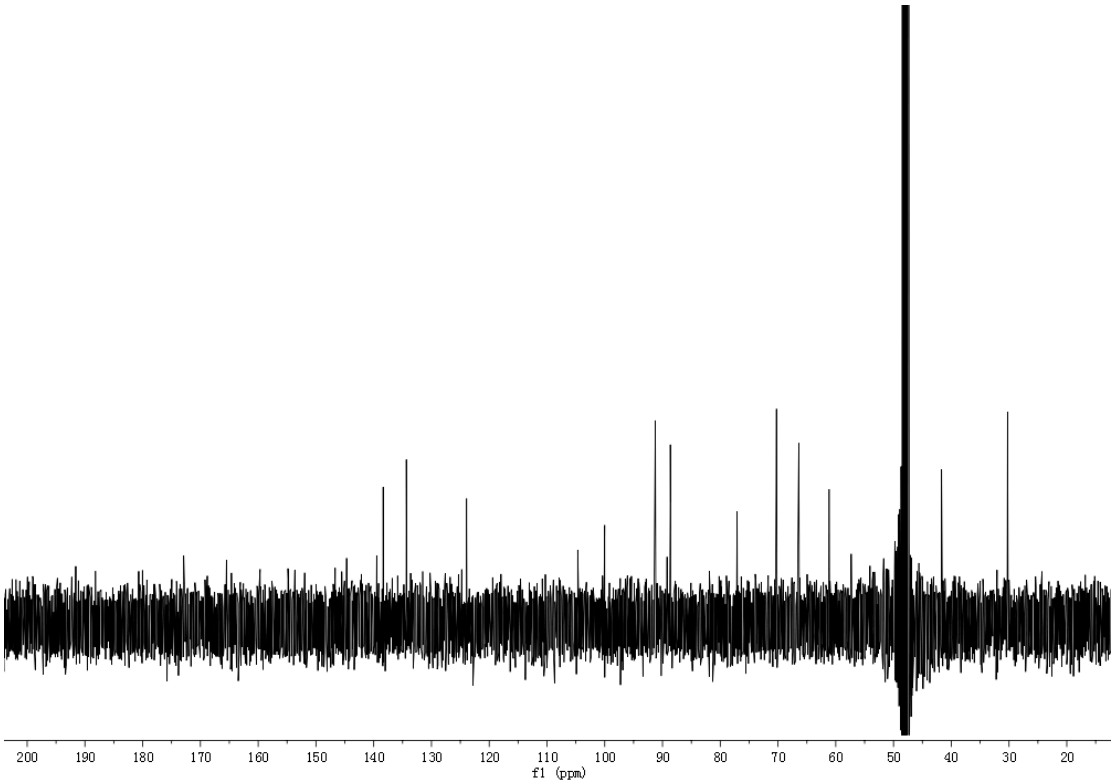
125 MHz ^{13}C NMR spectrum of **15** ($CDCl_3$)



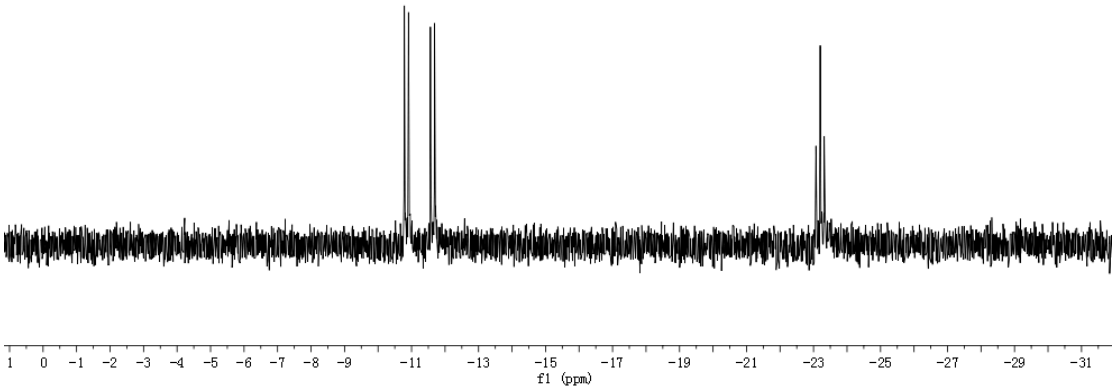
500 MHz ^{1}H NMR spectrum of **16** (CD_3OD)



125 MHz ^{13}C NMR spectrum of **16** (CD_3OD)



162 MHz ^{31}P NMR spectrum of **17** (D_2O)



Gel-Based Incorporation/Extension Assay.

Primer oligonucleotides (Integrated DNA Technologies) were 5'-radiolabeled with T4 polynucleotide kinase (New England Biolabs; Ipswich, MA) and [γ-^{32}P]-ATP (Perkin-Elmer) and annealed to template oligonucleotides (Sequence Context I in Seo *et al.*[3]) by heating to 95 °C followed by slow cooling to room temperature. Reactions were initiated by adding a solution of 2× dNTP and dXTP solution (5 μL) to a solution containing polymerase (73.53 nM) and primer:template (40 nM) in 5 μL Klenow reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM DTT and 50 μg/mL acetylated BSA). After incubation at 25 °C for 5−10 s, reactions were quenched with 20 μL of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanol). Reaction products were resolved by 8 M urea 15% polyacrylamide gel electrophoresis, and gel band intensities corresponding to the extended and unextended primers were quantified by phosphorimaging (Storm Imager, Molecular Dynamics) and Quantity One (Bio-Rad) software. The reported values are the average and standard deviation of three independent determinations.

PCR Amplification

Materials. Taq and OneTaq DNA polymerases were purchased from New England Biolabs (Ipswich, MA). A mixture of dNTPs was purchased from Fermentas (Glen Burnie, MD). SYBR Green I Nucleic Acid Gel Stain (10,000×) was purchased from Life Technologies (Carlsbad, CA).

DNA oligonucleotides. See Table S1 for complete oligonucleotide sequences. The syntheses of the DNA templates D6[4] and 134mer[5] were described previously. Fully natural primers were purchased from Intergrated DNA Technologies (Coralville, Iowa). Reagents and solvents for synthesis of unnatural primers 1-3 were obtained from Glen Research (Sterling, VA) and/or Applied Biosystems (Foster City, CA). The oligonucleotides were prepared using standard automated DNA synthesis with ultra-mild natural phosphoramidites (Glen Research) and d**NaM** phosphoramidite (Berry & Associates, Inc., Dexter, MI) on controlled pore glass supports (0.20 μmol, 1000 Å, Glen Research) and an ABI Expedite 8905 synthesizer. After automated synthesis, the oligonucleotides were cleaved from the support, deprotected by incubation in conc. aqueous ammonia overnight at room temperature, purified by DMT purification (glen-pak™ cartridge, Glen Research), and desalted over Sephadex G-25 (NAP-25 Columns, GE Healthcare). The concentration of single stranded oligonucleotides was determined by UV absorption at 260 nm.

PCR assay. PCR amplifications were performed in a total volume of 25 μL and with conditions specific for each assay as described in Table S2. After amplification, a 5 μL aliquot was analyzed on a 6% non-denaturing PAGE gel ran along with 50bp ladder (Life Technologies) to confirm amplicon size. The remaining solution was purified by spin-column (DNA Clean and Concentrator-5; Zymo Research, Irvine, CA), followed by 4% agarose gel, recovered with Zymoclean Gel DNA Recovery Kit (Zymo Research), quantified by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies), and sequenced on a 3730 DNA Analyzer (Applied Biosystems). Fidelity was determined as the average %retention of the unnatural base pair per doubling as described below.

Determination of fidelity. The percent retention of an unnatural base pair (*F*) was measured using raw sequencing data and normalized to fidelities per doubling, as described.[4,5] Briefly, the presence of an unnatural nucleotide leads to a sharp termination of the sequencing profile, while mutation to a natural nucleotide results in "read-through" (compare sequencing traces of d**NaM**/d**5SICS**$^{\mathbf{PA}}$ and d**NaM**/**dTPT3**$^{\mathbf{PA}}$ amplified with Taq, **Fig. S2**). The extent of the "read-through" is thus inversely correlated with the retention of the unnatural base pair. To use the sequencing data as a quantitative measurement of PCR fidelity, we performed calibration experiments in the range of 50-100% retention of the unnatural base pair (see Malyshev *et al.,* 2009[4] for details). Therefore, low retention (<50%) and high "read-through" make the quantification inaccurate.

Quantification of the high retention (>50%) was performed by adjusting the start and stop points for the Sequencing Analysis software (Applied Biosystems) and then determining the average signal intensity individually for each channel (A, C, G and T) for peaks within those defined points (35-45 nucleotides in length) before (section L) and

after (section R) the unnatural nucleotide. The R/L ratio was normalized using sequencing calibration plots to account for both noise in the sequencing chromatograms and the read-through in the control samples (see Malyshev *et al.,* 2009[4] for details). The R/L ratio of after normalization (R/L_{norm}) corresponds to the percentage of the natural sequences in the pool. Finally, F was calculated as $1 - (R/L)_{norm}$ and the retention of the unnatural base pair per doubling (fidelity, f) was calculated as $F^{\frac{1}{\log_2 A}}$, where A is an amplification and $\log_2 A$ is the number of doublings. Each sample before and PCR amplification was sequenced in triplicate in each direction to minimize sequencing error (Fig. S1-S2).

Gel mobility assay

Corresponding DNA templates (see Fig. 2 and Table S1) were amplified by PCR under the conditions described in Table S2. Upon completion, NaOH (1 M, 12.5 μL) was added directly to PCR samples to a final concentration of 0.2 M and incubated for 5 h at room temperature. After the addition of NaOAc (3 M, pH 5.5, 7.5 μL) and 200 μL of cold ethanol, the samples were mixed, incubated on ice overnight, and DNA was precipitated by centrifugation at 10 000 rfu for 30 min at 4 °C. The supernatant was removed and the pellets were carefully washed with 80% ethanol. The samples were resuspended in 50 μL of the annealing buffer (50 mM Na phosphate, pH 7.5; 100 mM NaCl; 1 mM EDTA), heated to 95 °C and cooled to room temperature over 30 min. NHS-PEG_4-biotin (Thermo Scientific) solution in the annealing buffer (40 mM, 50 μL) was mixed with the DNA samples and incubated overnight at room temperature. The samples were purified by spin-column (DNA Clean and Concentrator-5; Zymo Research, Irvine, CA) and eluted in 10 μL of elution buffer. Half of the sample (5 μL) was mixed with 1 μg of streptavidin (Promega) in annealing buffer, incubated for 30 minutes at 37 °C, mixed with 5× non-denaturing loading buffer (Qiagen), and loaded on 6% non-denaturing PAGE. The other half was mixed with 5× non-denaturing loading buffer, and loaded directly on the gel as a control. After running the gel at 110 V for 30 min, the gel was soaked in 1× Sybr Gold Nucleic Acid Stain (Life Technology, Carlsbad, CA) for 30 min and visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad).

Table S1. Sequences of DNA used in this study. Primer binding regions are underlined.

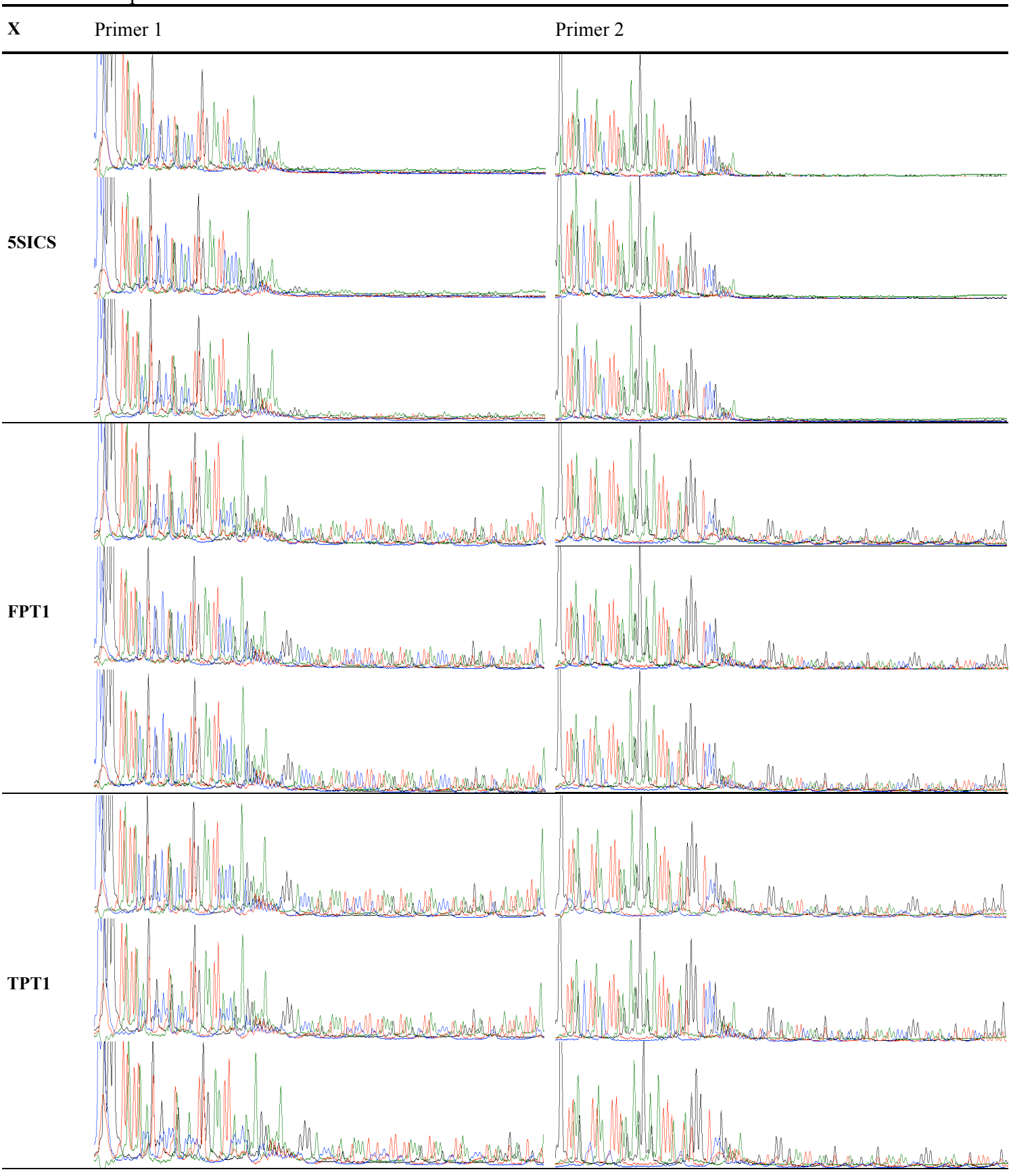
Name	Sequence (5' to 3')	Remarks
Fend1	CACACAGGAAACAGCTATGAC	Primers for PCR (templates D6 and 134mer)
Fend2	GAAATTAATACGACTCACTATAGG	
Fend1-poly-dT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTCACACAG GAAACAGCTATGAC	Primers for Sanger sequencing (templates D6 and 134mer)
Fend2-poly-dT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAAT TAATACGACTCACTATAGG	
D6	CACACAGGAAACAGCTATGACCCGGGTTATTAC ATGCGCTAGCACTTGGAATTCACCAGACGNNN **NaM** NNNCGGGACCCATAGTAAATCTCCTTCTT AAAGTTAAGCTTAACCCTATAGTGAGTCGTATT AATTTC	N = randomized natural nucleotide
134mer	CACACAGGAAACAGCTATGACCCGGGTTATTAC ATGCGCTAGCACTTGGAATTCACAATACT **NaM** TCTTTAAGGAAACCATAGTAAATCTCCTTCTT AAAGTTAAGCTTAACCCTATAGTGAGTCGTATT AATTTC	Used in Fig. 3B
Primer1	**NaM** CCTGCGTCAATGTAATGTTC	Primers for PCR with Temp1-3
Primer2	TTCACGGT **NaM** AGCACGCATAGG	
Primer3	CCAATGTACC **NaM** TGCGTATGTTC	
Primer-rev	CCCTGCGTTTATCTGCTCTC	
Temp1	CCCTGCGTTTATCTGCTCTCTCGGTCGTTCGGC TGCGGCGGAACATTACATTGACGCAGG	60mers used in Fig. 3C. The nucleotides shown in bold form a mispair with d**NaM** in the first round of PCR
Temp2	CCCTGCGTTTATCTGCTCTCTCGGTCGTTCGGC TGCGCGCCTATGCG**T**GCTTACCGTGAA	
Temp3	CCCTGCGTTTATCTGCTCTCTCGGTCGTTCGGC TGCCGGAACATACGCA**T**GGTACATTGG	

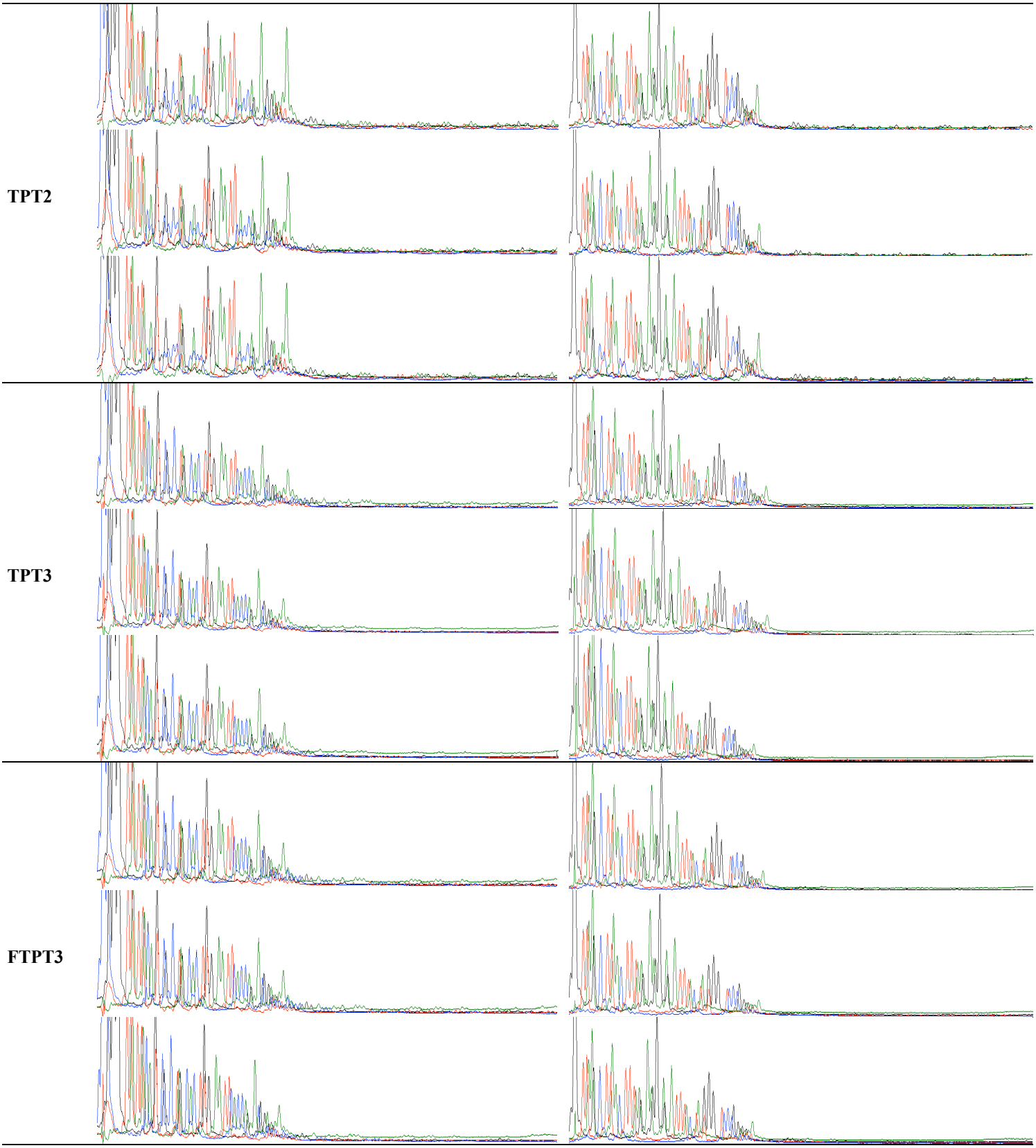
Table S2. PCR conditions.

	OneTaq	Taq	PCR for biotin labeling (Fig. 3B)	PCR for biotin labeling (Fig. 3C)
Buffer	1×OneTaq buffer	1×Taq buffer	1×OneTaq buffer	1×OneTaq buffer
Enzyme concentration, U/μL	OneTaq, 0.02	Taq, 0.02	OneTaq, 0.02	OneTaq, 0.02
Template	D6[1] (0.01 ng)	D6[1] (0.01 ng)	134mer[1] (0.5 ng)	60mer[1] (0.5 ng)
dNTPs, μM	200	200	200	200
d**NaM**TP, μM	100	100	100	100
d**X**TP, μM	100	100	100 of d**5SICS**^{PA}TP or d**TPT3**^{PA}TP	100 of d**5SICS**^{PA}TP or d**TPT3**^{PA}TP
Mg^{2+}, mM	3	3	3	3
Primers, μM	1	1	1	1
SYBR Green I	0.5×	0.5×	0.5×	0.5×
Thermal conditions				
Initial denaturing	-	-	96 °C 1 min	96 °C 1 min
Denaturing	96 °C 10 s	96 °C 10 s	96 °C 15 s	96 °C 15 s
Annealing	60 °C 15 s	60 °C 15 s	60 °C 30 s	64 °C 30 s
Extension	68 °C 60 s	68 °C 15 s	68 °C 2 min	68 °C 2 min
# of cycles	16+16+16 [2]	20	12	12

[1] See Table S1 for sequences of the templates and primers. [2] Initial amount of template was 0.01 ng. PCR mixture was amplified over 16 cycles, diluted 40,000-fold and amplified over another 16 cycles. This dilution/amplification step was repeated resulting in 48 total cycles of amplification.

Figure S1. Raw sequencing data after PCR amplification by OneTaq polymerase in the presence of d**X**TP and d**NaM**TP. Triplicate data shown.




TPT2
TPT3
FTPT3

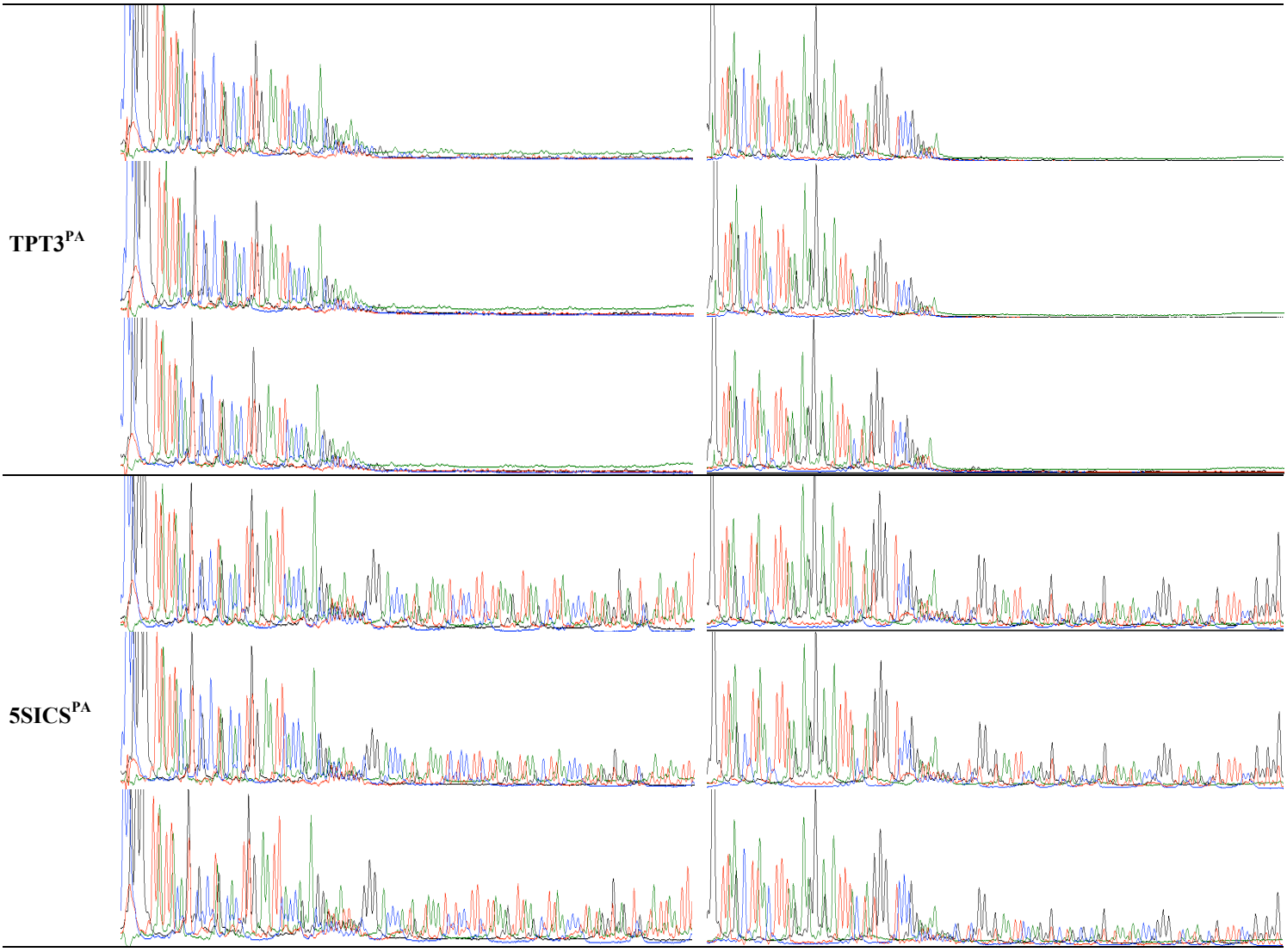
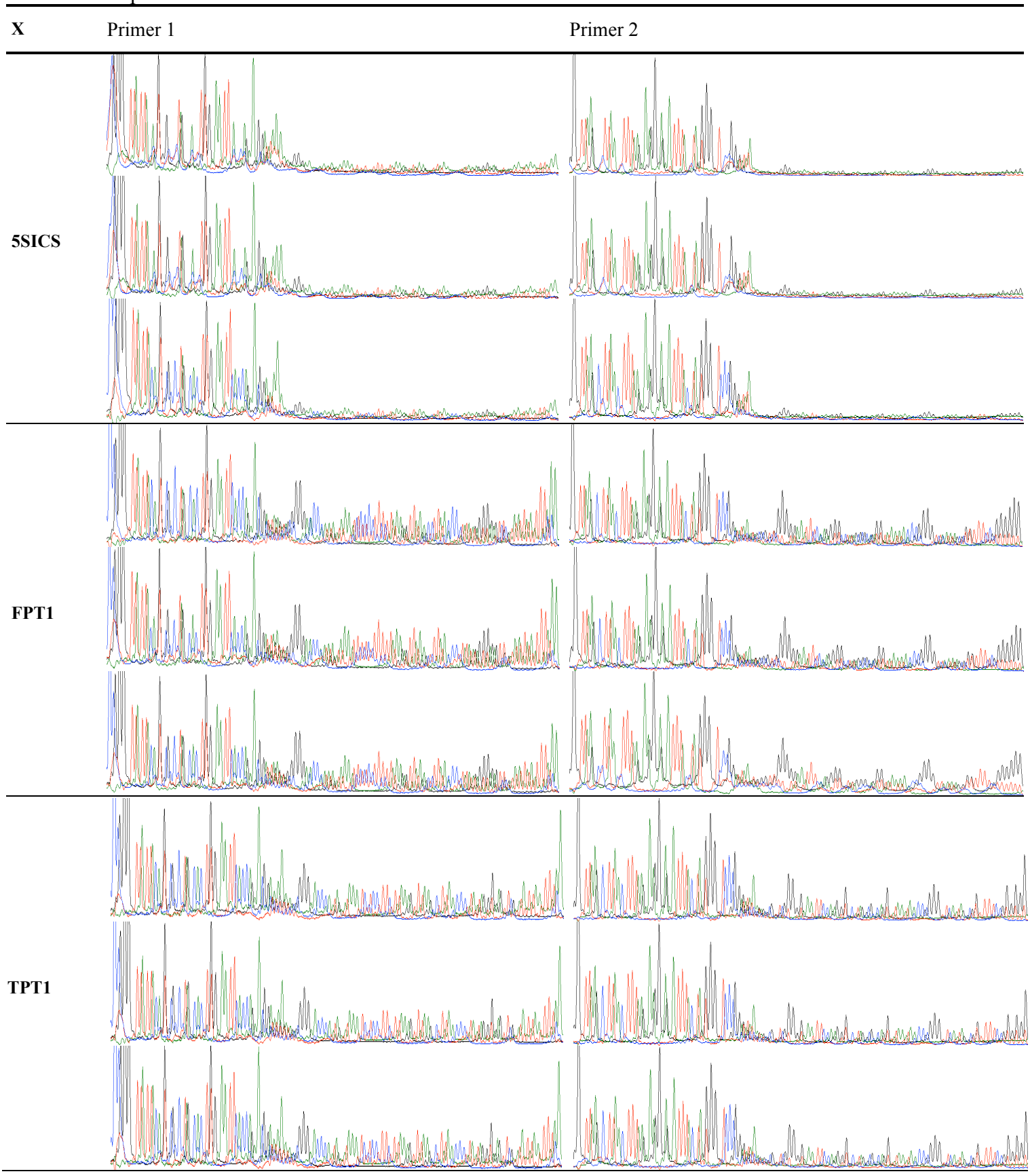
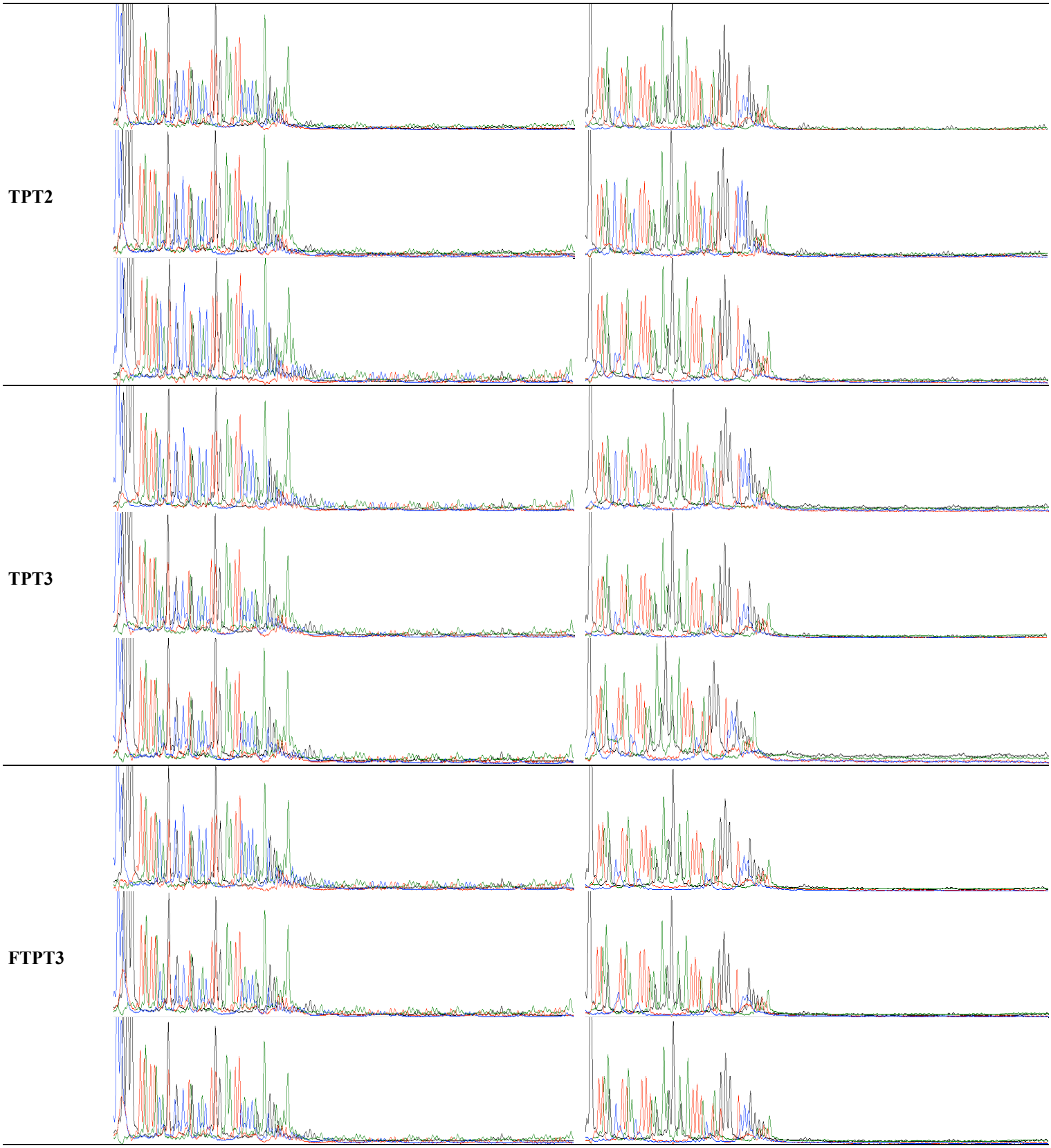
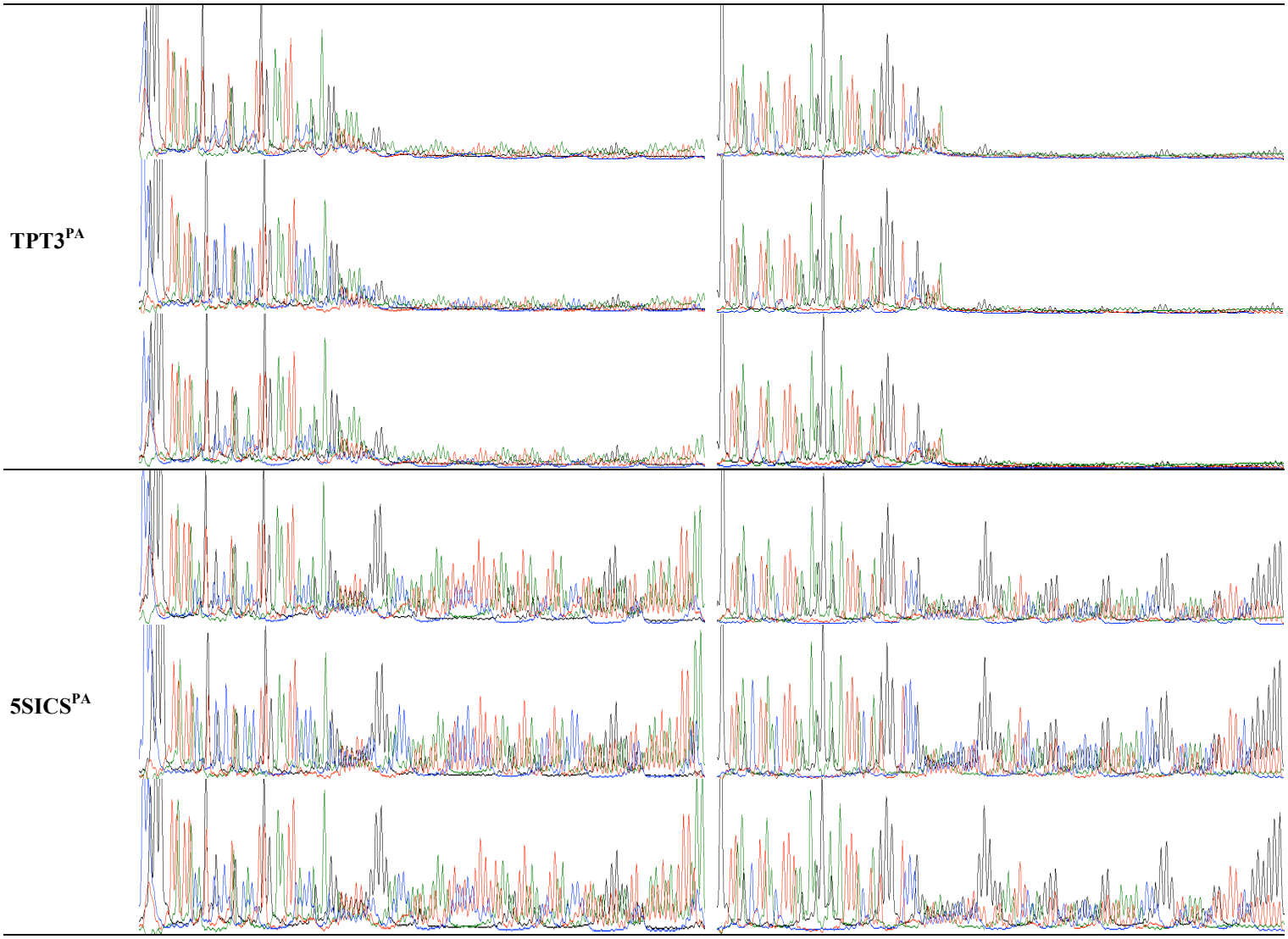

TPT3PA
5SICSPA

Figure S2. Raw sequencing data after PCR amplification by Taq polymerase in the presence of d**X**TP and d**NaM**TP. Triplicate data shown.




TPT2
TPT3
FTPT3


TPT3PA
5SICSPA

References

(1) New, J. S.; Christopher, W. L.; Yevich, J. P.; Butler, R.; Schlemmer, R. F., Jr.; VanderMaelen, C. P.; Cipollina, J. A. *J. Med. Chem.* **1989**, *32*, 1147-1156.

(2) Asagarasu, A.; Matsui, T.; Hayashi, H.; Tamaoki, S.; Yamauchi, Y.; Sato, M. *Chem. Pharm. Bull. (Tokyo)* **2009**, *57*, 34-42.

(3) Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 3246-3252.

(4) Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 14620-14621.

(5) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12005-12010.