Supporting Information

2-Amino-5-Methylene-Pyrimidine-4,6-dione-Based Janus G-C Nucleobase as a Versatile Building Block for Self-Assembly

Mahendra A. Wagh,[a,b] Dinesh Shinde,[a,b] Gamidi Rama Krishna,[a] and Gangadhar J. Sanjayan*[,a,b]

[a]*Organic Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Rd. Pashan, Pune 411008, India,*

[b]*Academy of Scientific and Innovative Research (AcSIR), Sector-19, Kamla Nehru Nagar, Ghaziabad, Uttar Pradesh-201 002, India.*

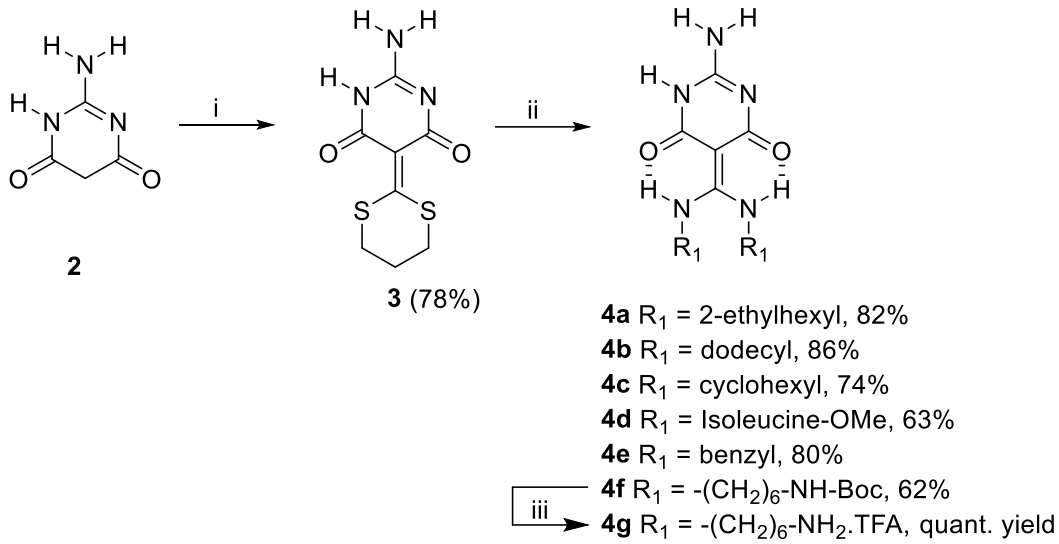
Email: gj.sanjayan@ncl.res.in

Table of Contents:

General Methods

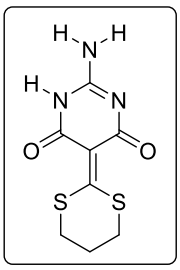
Unless otherwise stated, all chemicals and reagents were obtained commercially. Dry solvents were prepared by standard procedures. Analytical Thin Layer Chromatography was done on precoated silica gel plates (Kieselgel $60F_{254}$, Merck). Column chromatographic purifications were done with 100-200 and 230-400 mesh silica gel. NMR spectra were recorded in $CDCl_3$, DMSO-*d*$_6$ and D_2O on AV 400 and AV 500 MHz, respectively, on Bruker NMR spectrometers. All chemical shifts are reported in δ ppm downfield to TMS and peak multiplicities are referred to as singlet (s), doublet (d), quartet (q), broad singlet (bs), and multiplet (m). Melting points were recorded on MEL-TEMP. HRMS (ESI) data were recorded on a Thermo Scientific Q-Exactive, Accela 1250 pump.

Scheme 1:



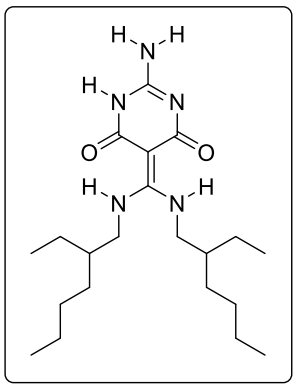
Reaction conditions: i) KOH, CS_2, 1,3-dibromopropane, H_2O, rt, 6h; ii) R_1-NH_2, DMSO, rt, 5h; iii) TFA:DCM (50:50), 0 °C-rt, 30 min.

Compound 3



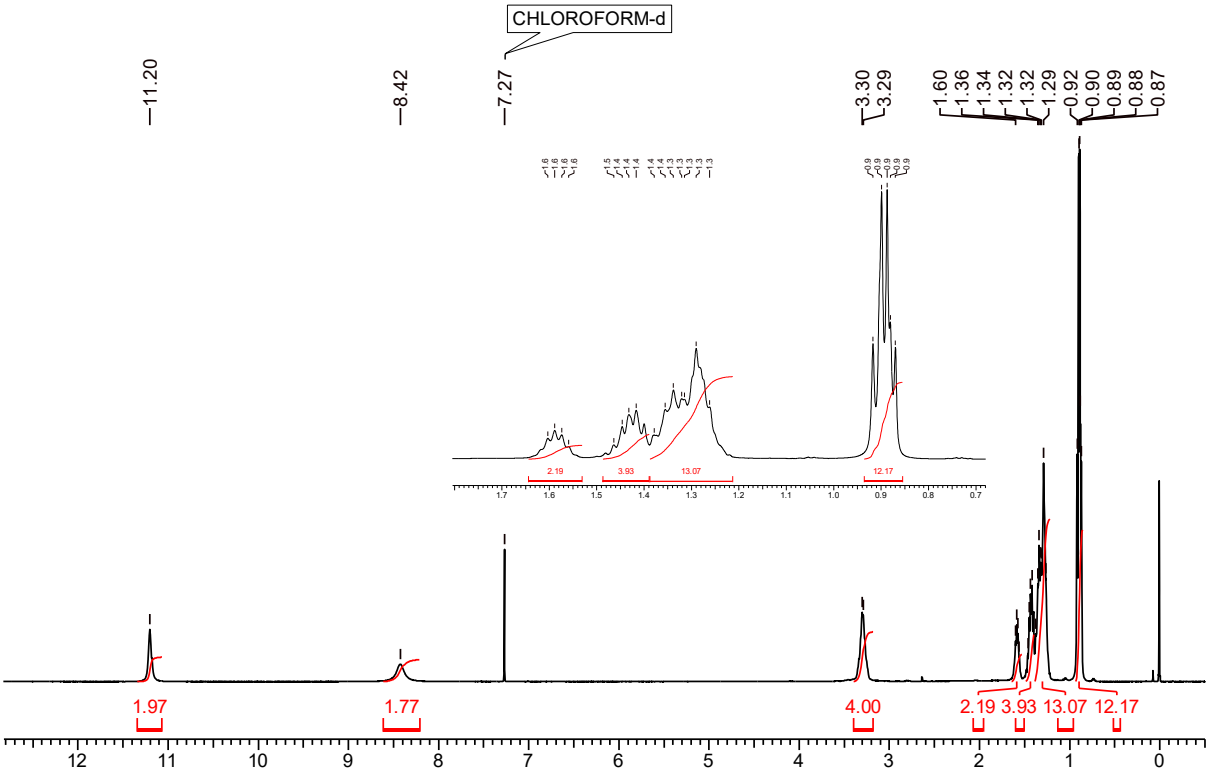
To a suspension of **2** (5 g, 39.34 mmol, 1 equiv.) in 20 ml distilled water at 0°C, KOH (5.52 g, 98.35 mmol, 2.5 equiv.) dissolved in 10 ml distilled water and CS_2 (8.24 ml, 129.82 mmol, 3.3 equiv.) were added dropwise and the reaction mixture was stirred for 30 minutes at room temperature. Finally, 1,3-dibromopropane (4.81 ml, 47.21 mmol, 1.2 equiv.) was added to the above reaction mixture. The reaction mixture was stirred for 6 h at room temperature and then poured into a citric acid-containing brine solution. The resulting solid was filtered and washed with water followed by diethyl ether and dried in a P_2O_5 desiccator giving **3** (7.5 g, 78%) as a yellow solid. *Note:* Owing to its poor solubility in common organic solvents, **3** was carried forward for the next reactions, without further purification and characterization.

Compound 4a

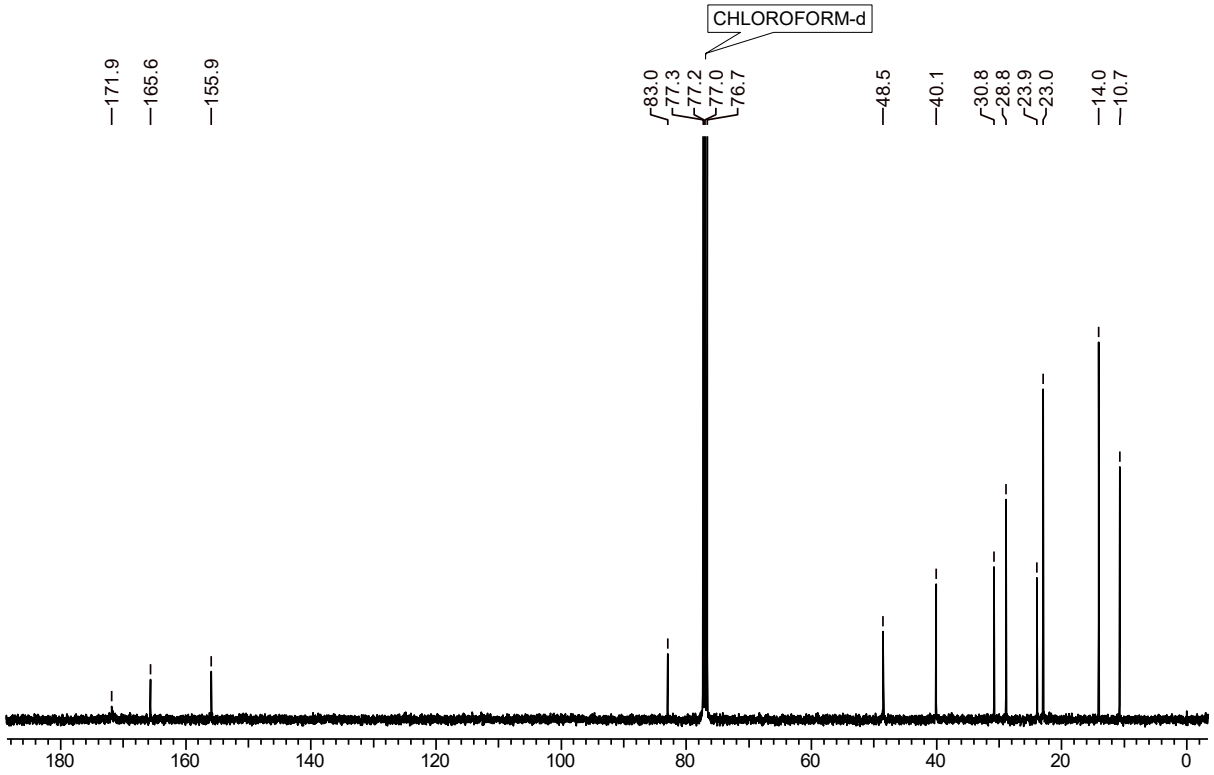


To a suspension of **3** (0.50 g, 2.06 mmol, 1 equiv.) in DMSO, 2-ethylhexylamine (1.01 ml, 6.17 mmol, 3 equiv.) was added and stirred at room temperature for 6 h. After completion of the reaction, the reaction mixture was poured into a citric acid-containing brine solution and extracted with DCM and sequentially washed with water and brine solution. The organic layer was dried over Na_2SO_4 and evaporated *in vacuo.* The residue obtained was purified by column chromatography (eluent: 3% MeOH/DCM, R_f: 0.5) to afford **4a** (0.66 g, 82%) as a white solid: mp: >250 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 11.20 (bs, 2H), 8.42 (bs, 2H), 3.30-3.29 (m, 4H), 1.60-1.56 (m, 2H), 1.48-1.40 (m, 4H), 1.38-1.26 (m, 13H), 0.92-0.87 (q, J = 6.9 Hz, 12H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ: 171.9, 165.6, 155.9, 83.0, 48.5, 40.1, 30.8, 28.8,

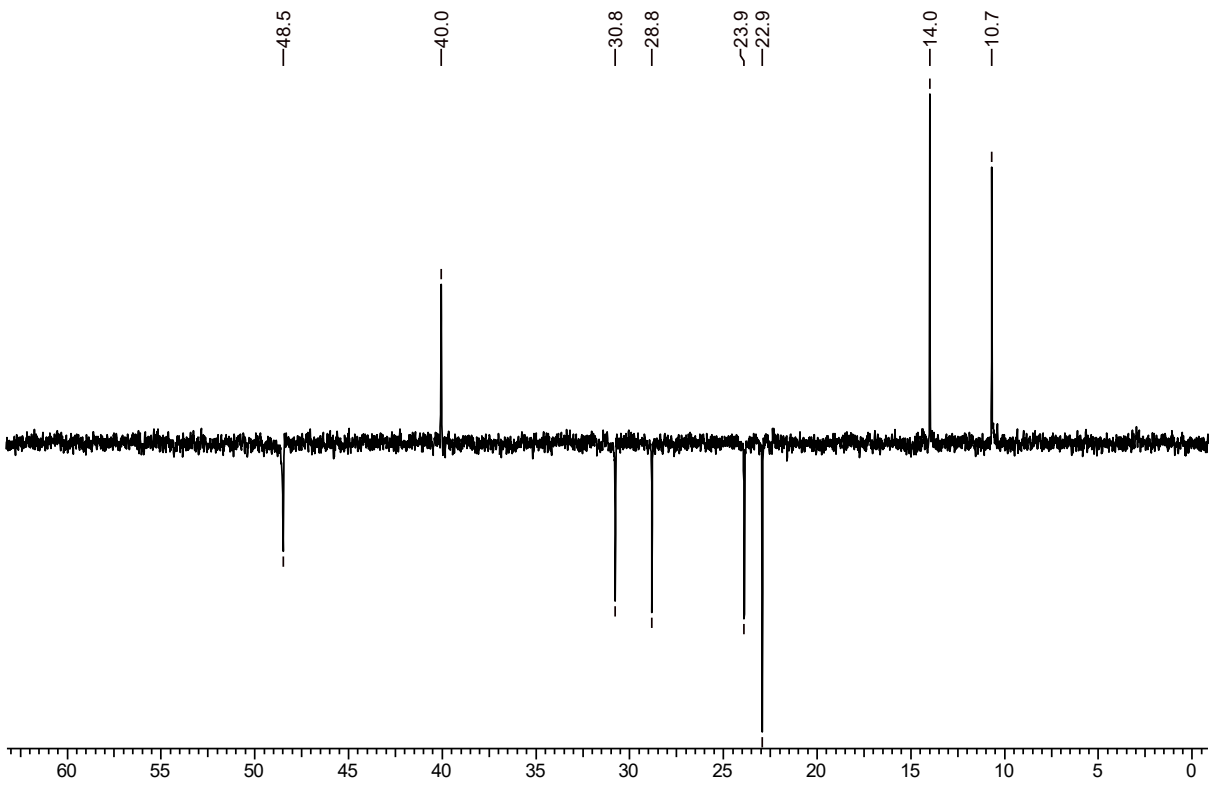
23.9, 23.0, 14.0, 10.7; HRMS (ESI) calculated $[M+H]^+$ for $C_{21}H_{40}N_5O_2$: 394.3177, found 394.3173 $[M+H]^+$.



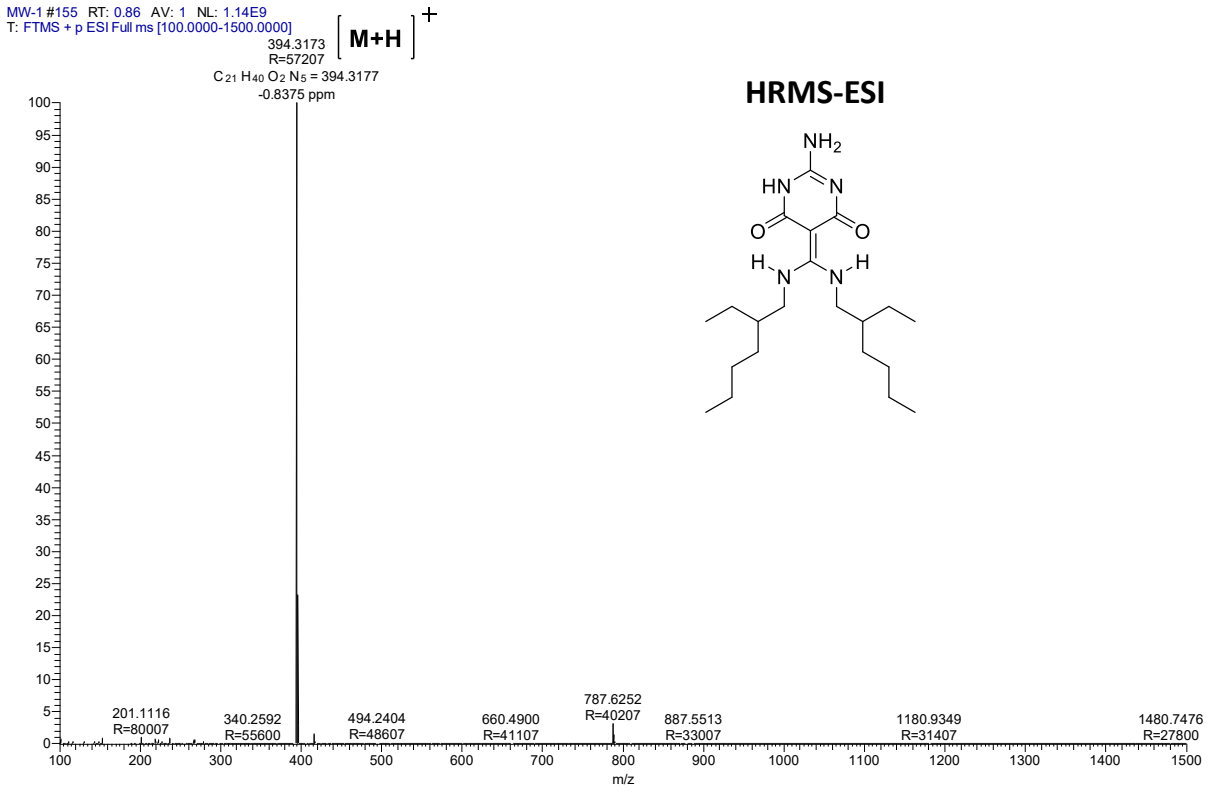
^{1}H NMR spectrum of compound **4a** ($CDCl_3$, 400 MHz, 298 K)



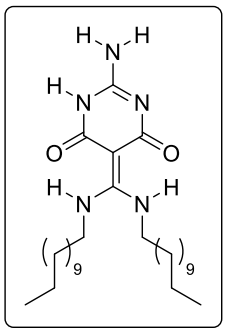
^{13}C{^{1}H} NMR spectrum of compound **4a** ($CDCl_3$, 100 MHz, 298 K)



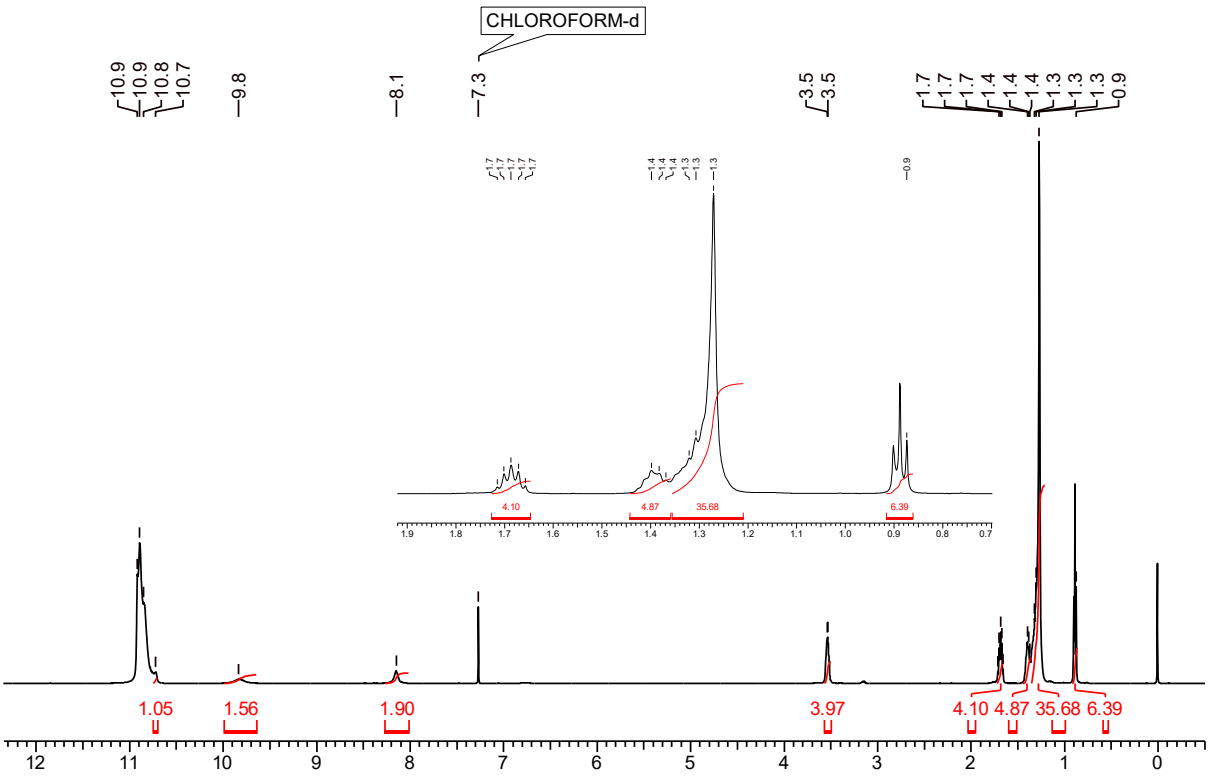
DEPT-135 NMR spectrum of compound **4a** ($CDCl_3$, 100 MHz, 298 K)



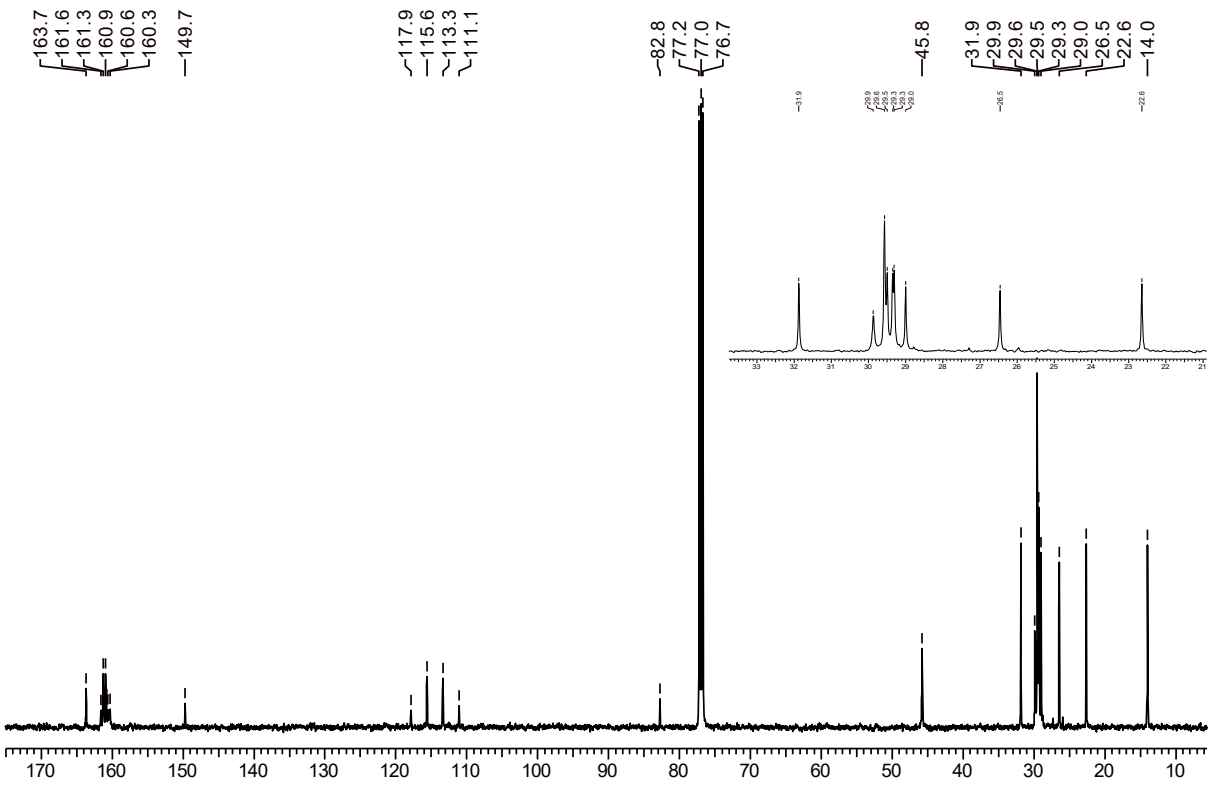
Compound 4b



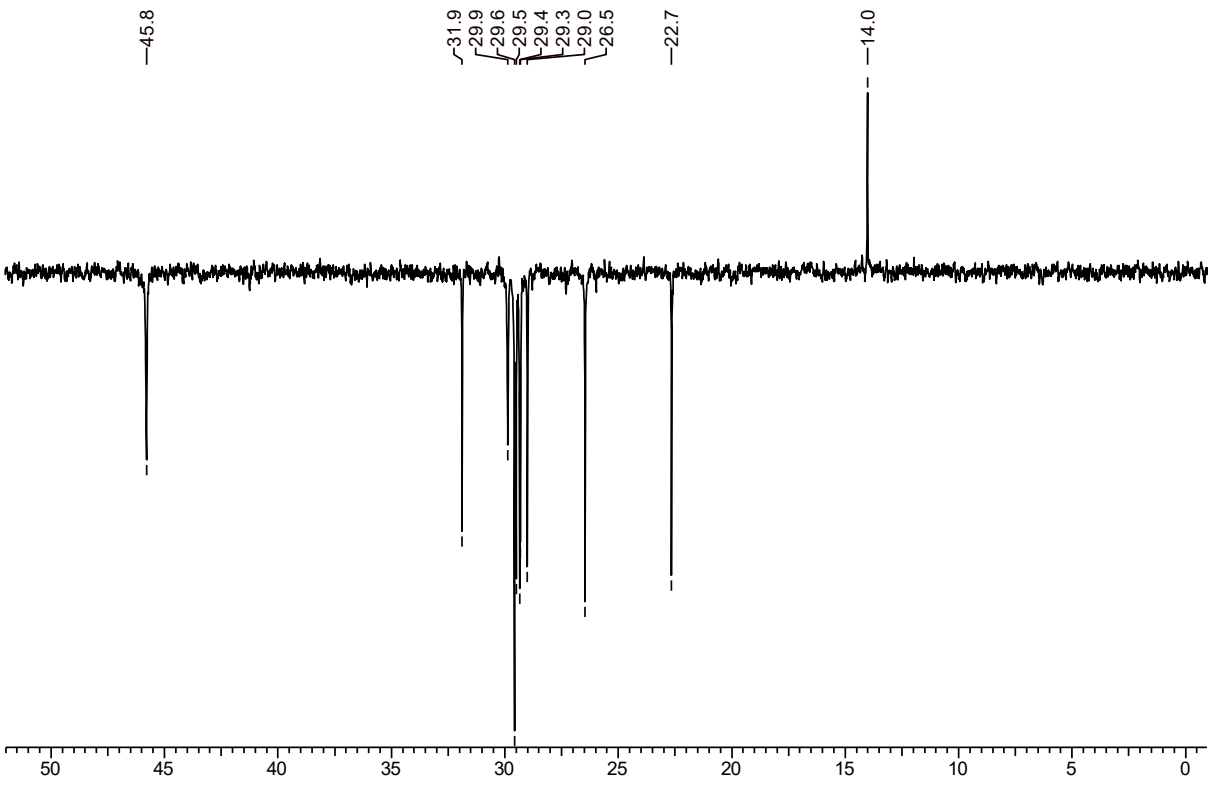
4b was obtained following the similar procedure of **4a**. Yield: 86%, white solid. *Note:* We could not purify compound **4b** by column chromatography as it was not soluble in organic solvents. But, before submitting for NMR, compound **4b** was sequentially washed with DMSO, MeOH and diethyl ether. The NMR was taken in $CDCl_3$ containing 1 drop of TFA. mp: compound decomposes above >240 °C; ^{1}H NMR (500 MHz, $CDCl_3$ +1 drop TFA) δ: 10.92-10.85 (residual TFA peak and three proton peaks), 10.72 (bs, 1H), 9.83 (bs, 2H), 8.15 (bs, 2H), 3.54-3.53 (m, 4H), 1.71-1.66 (m, 4H), 1.40-1.37 (m, 5H), 1.32-1.27 (m, 36H), 0.90-0.87 (t, J = 6.9 Hz, 6H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$ +1 drop TFA) δ: 163.7, 161.6-160.3 (two carbon peaks merges into the TFA signal); 149.7, 117.9-111.1 (residual TFA peak), 82.8, 45.8, 31.9, 29.9, 29.6, 29.4, 29.3, 29.0, 26.5, 22.7, 14.0; HRMS (ESI) calculated $[M+H]^+$ for $C_{29}H_{56}N_5O_2$: 506.4429, found 506.4422 $[M+H]^+$.



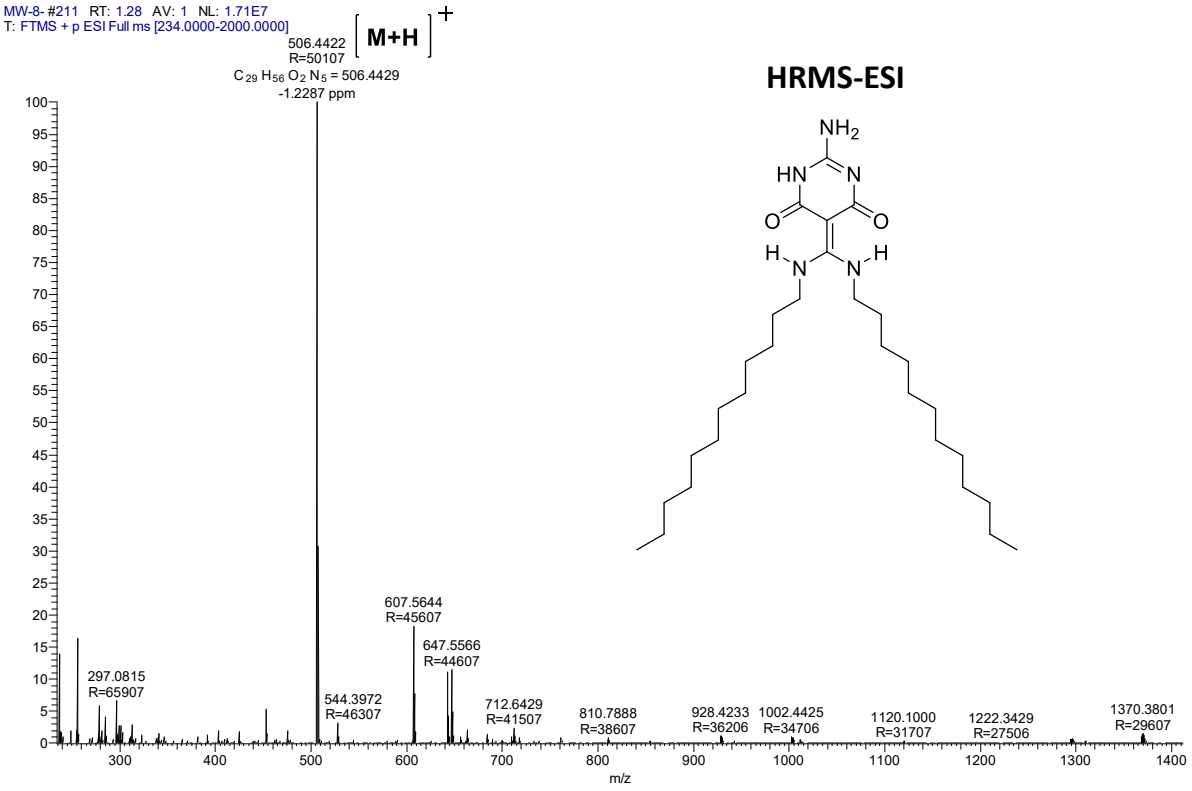
^{1}H NMR spectrum of compound **4b** ($CDCl_3$ +1 drop TFA, 500 MHz, 298 K)



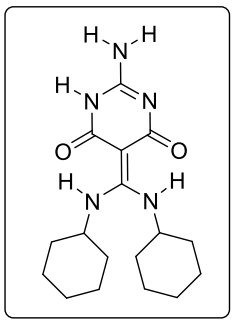
$^{13}C\{^{1}H\}$ NMR spectrum of compound **4b** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



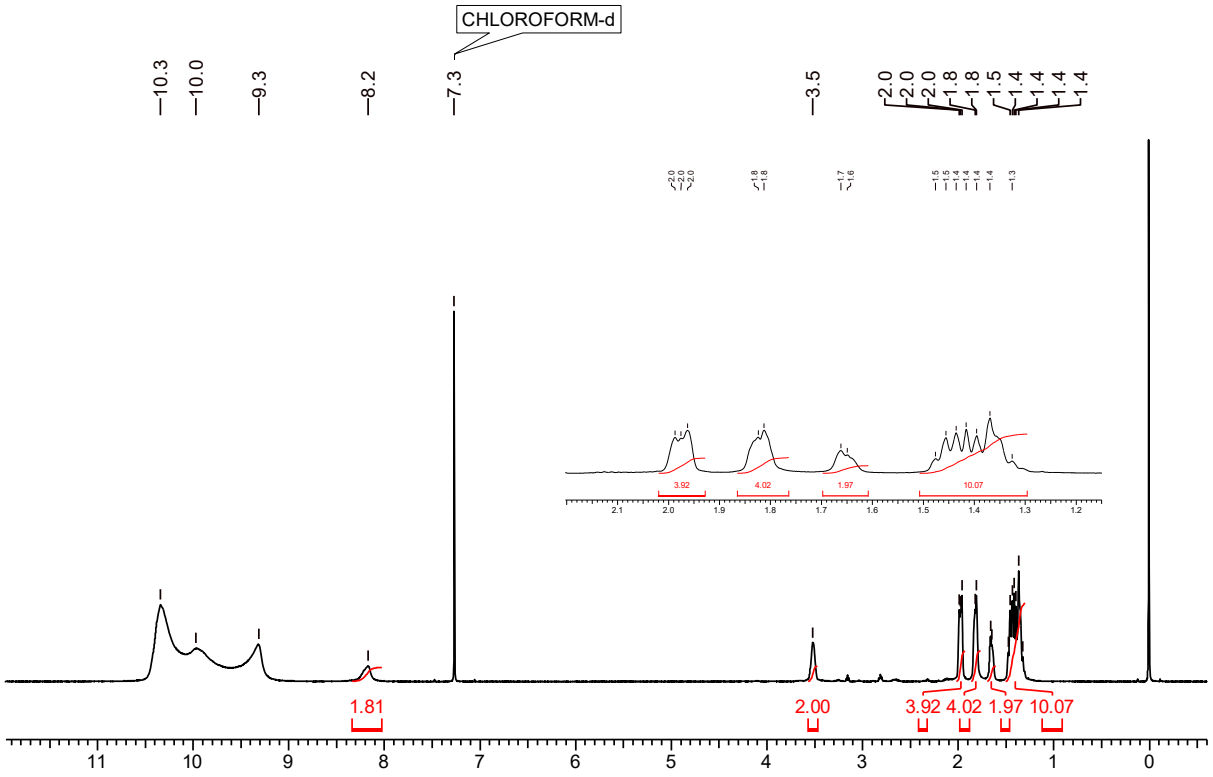
DEPT-135 NMR spectrum of compound **4b** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



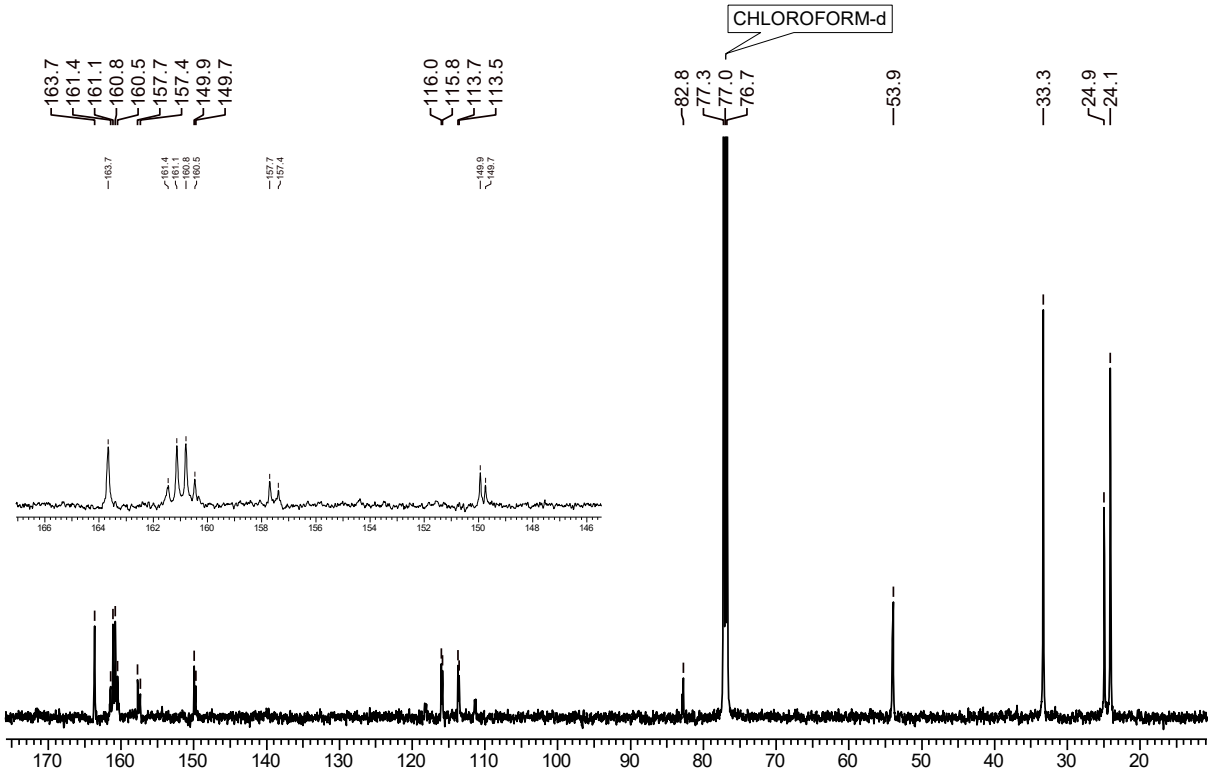
Compound 4c



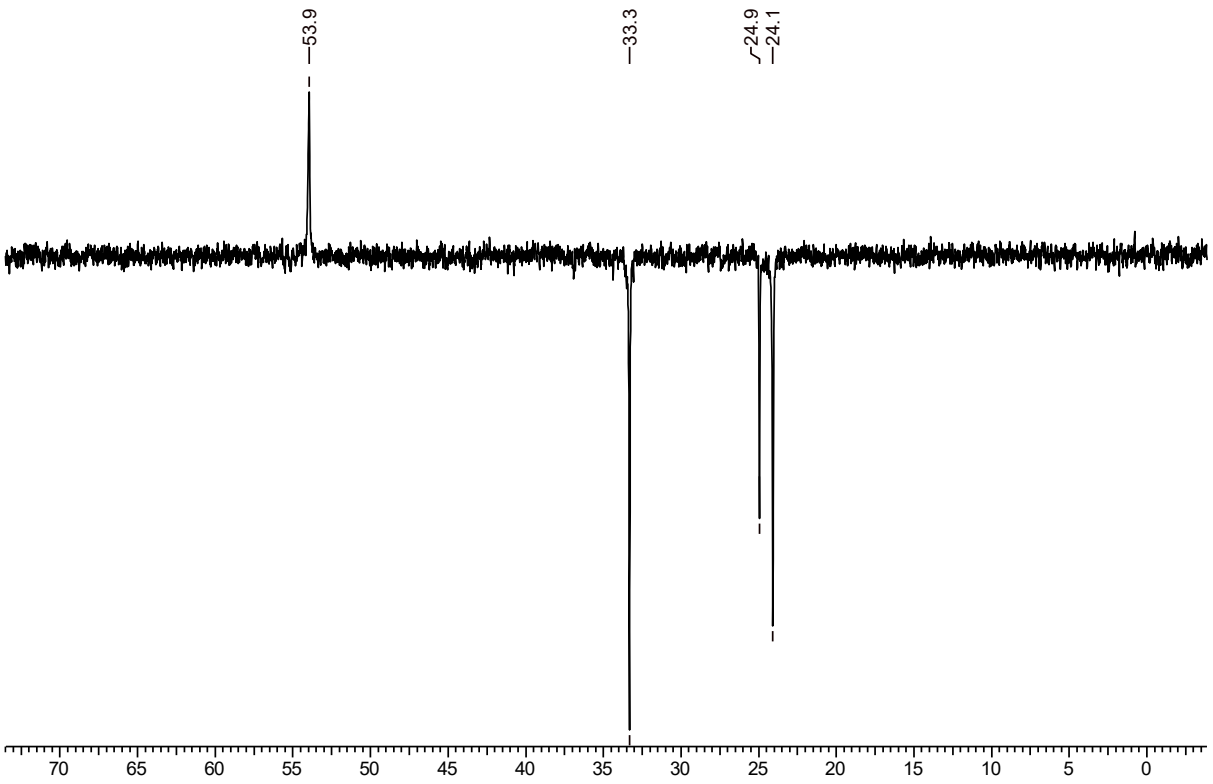
4c was obtained following the similar procedure of **4a**. Yield: 74%, white solid. *Note:* We could not purify compound **4c** by column chromatography as compound **4c** was not soluble in organic solvents. Before submitting for NMR, the compound **4c** was washed with MeOH (3 times) and diethyl ether. The NMR was taken in $CDCl_3$ containing 1 drop of TFA. mp: >250 °C; ^{1}H NMR (500 MHz, $CDCl_3$ +1 drop TFA) δ: 10.34-9.32 (residual TFA peak and three proton peaks merges into the TFA signal), 8.17 (bs, 2H), 3.52 (bs, 2H), 1.99-1.96 (m, 4H), 1.82-1.81 (m, 4H), 1.66-1.65 (m, 2H), 1.48-1.33 (m, 10H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$ +1 drop TFA) δ: 163.7, 161.4-160.5 (residual TFA signal); 157.7, 157.4, 149.9, 149.7, 116.0-113.5 (residual TFA peak), 82.8, 53.9, 33.3, 24.9, 24.1; HRMS (ESI) calculated $[M+H]^+$ for $C_{17}H_{28}N_5O_2$: 334. 2238, found 334.2228 $[M+H]^+$.



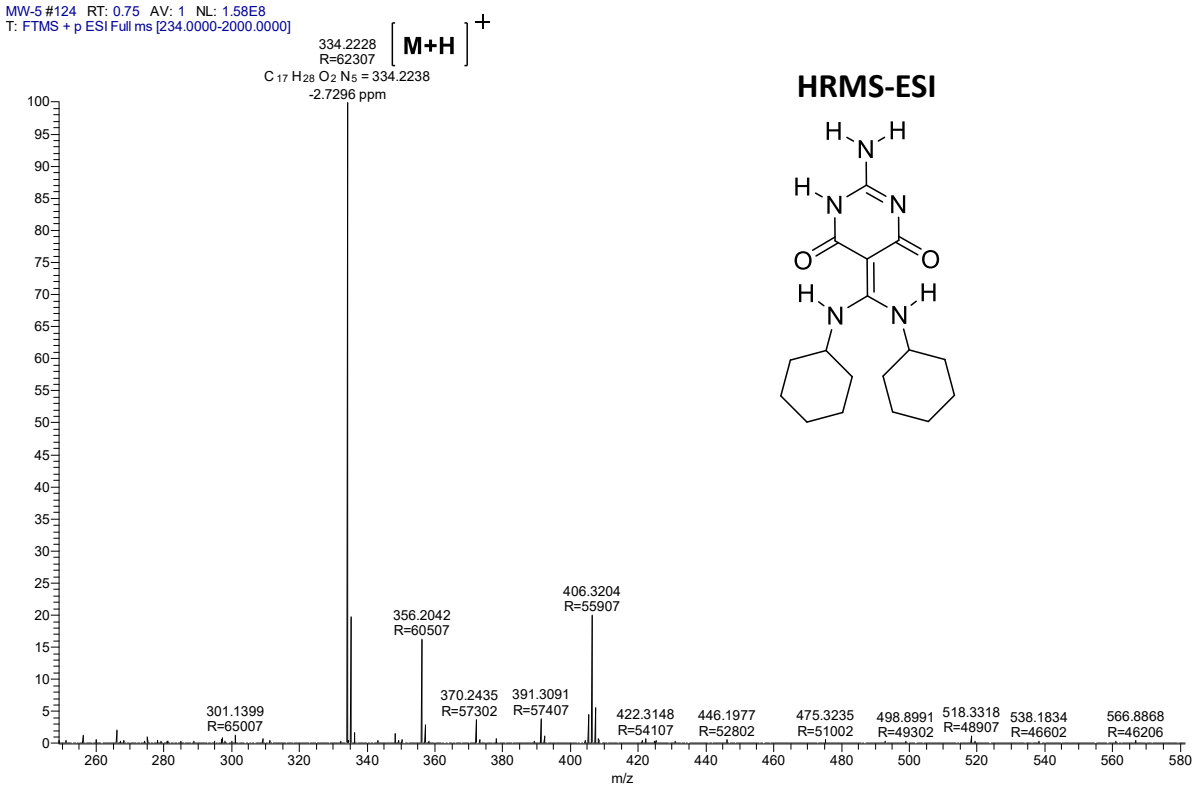
^{1}H NMR spectrum of compound **4c** ($CDCl_3$ +1 drop TFA, 500 MHz, 298 K)



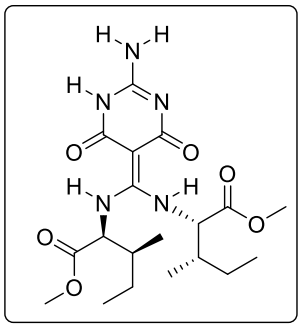
^{13}C{^{1}H} NMR spectrum of compound **4c** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



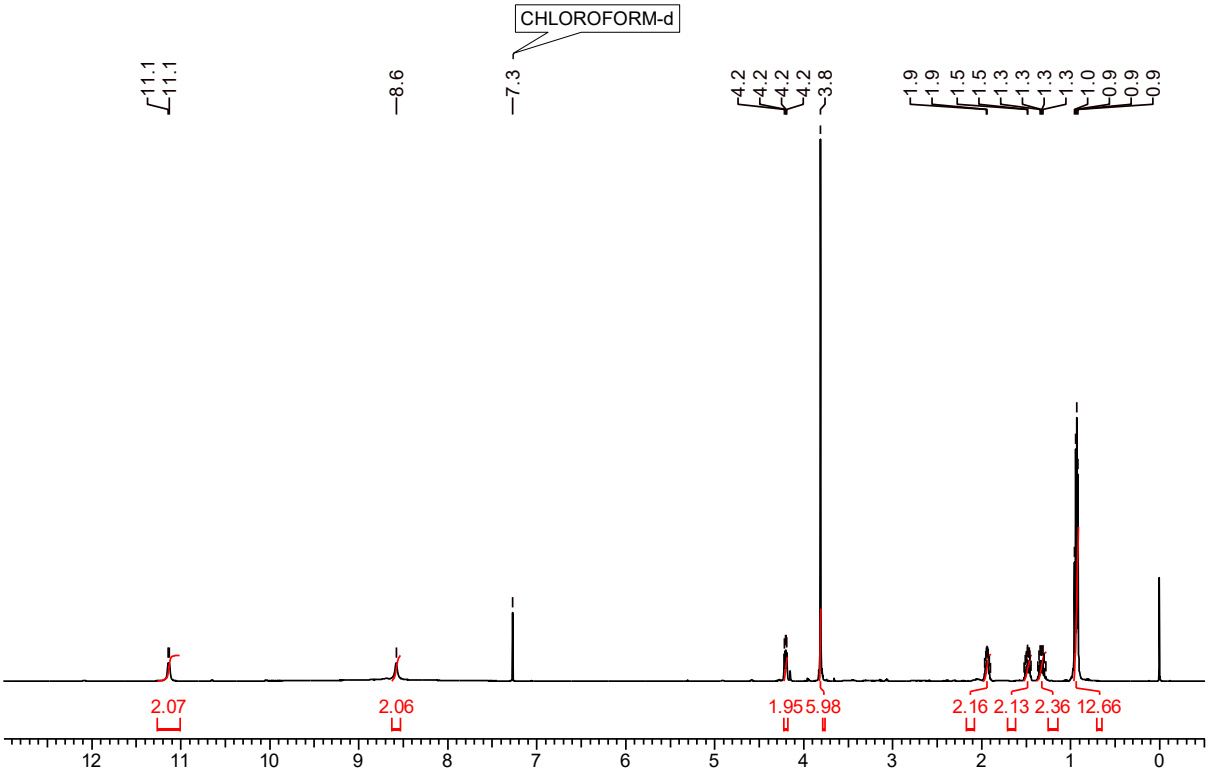
DEPT-135 NMR spectrum of compound **4c** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



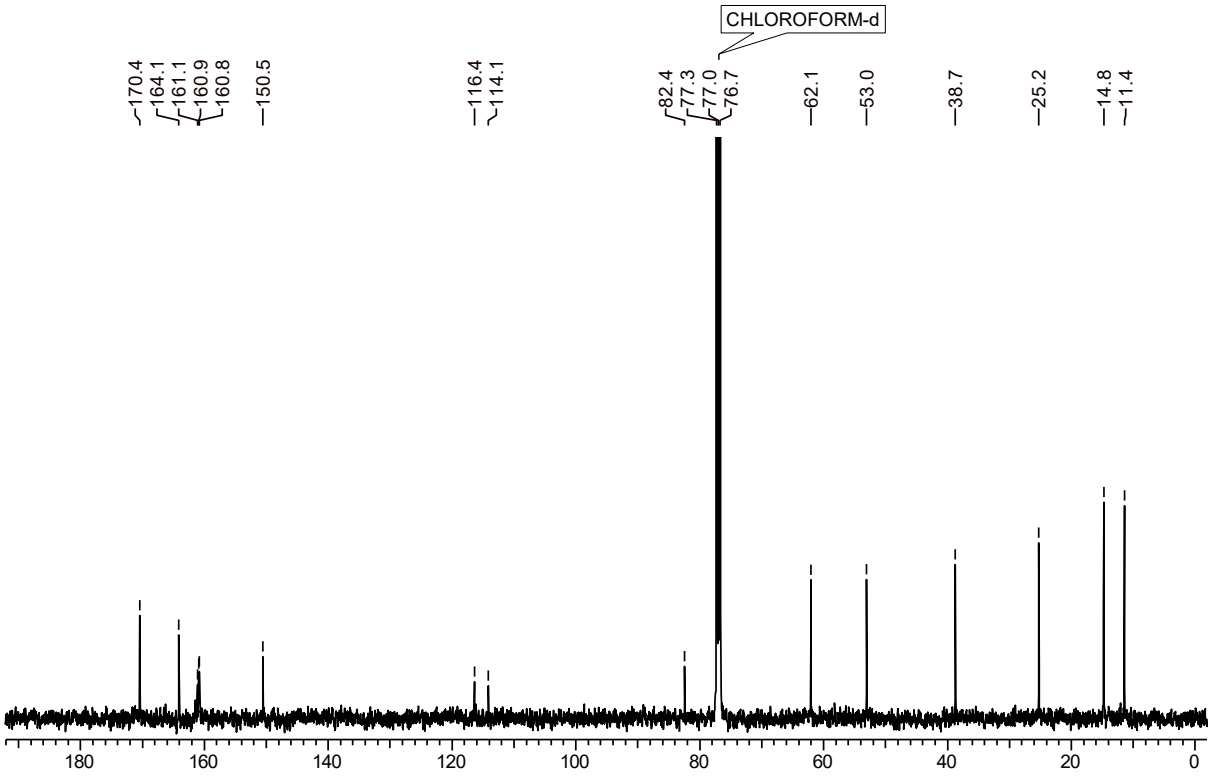
Compound 4d



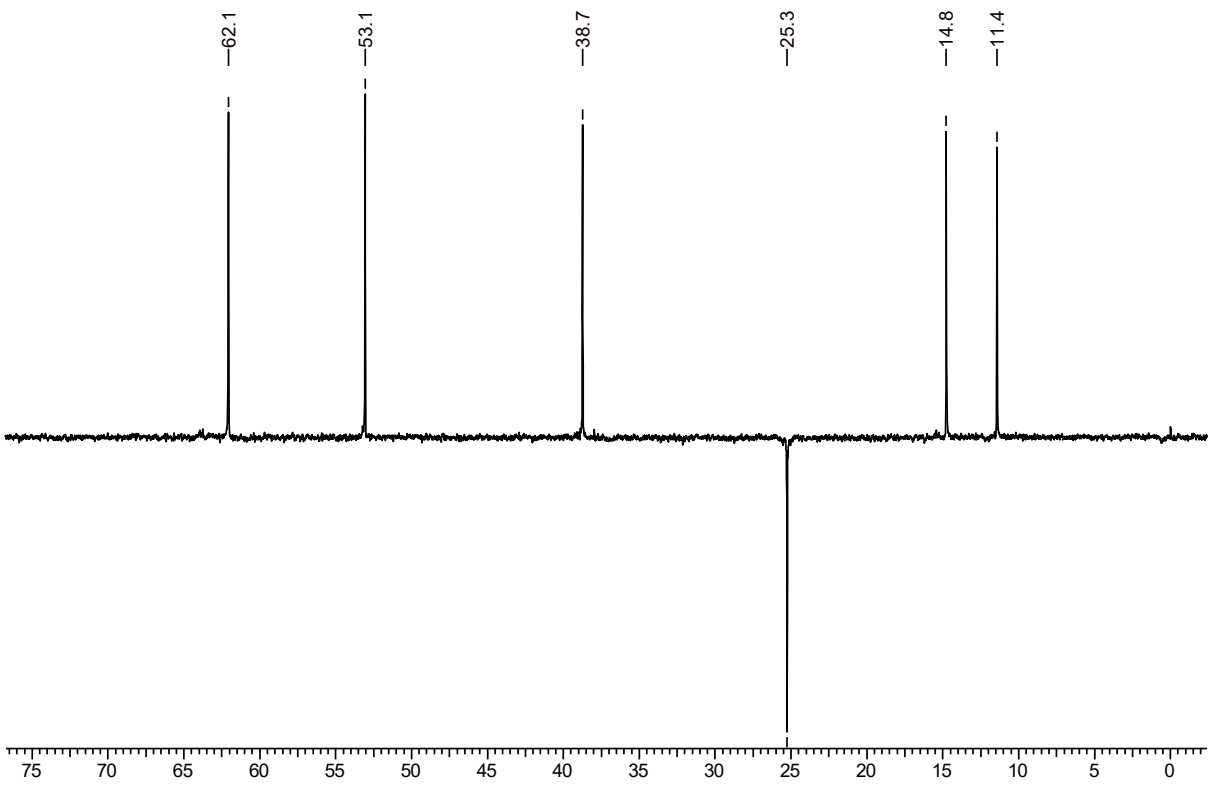
4d was obtained following the similar procedure of **4a**. The residue obtained was purified by column chromatography (eluent: 3% MeOH/DCM, R_f: 0.5) to afford **4d** (0.33 g, 63%) as a white solid. After column chromatography, the compound was not soluble in $CDCl_3$. Therefore, the NMR was taken in $CDCl_3$ containing 1 drop of TFA. mp: >250 °C; ^{1}H NMR (500 MHz, $CDCl_3$ +1 drop TFA) δ: 11.14-11.13 (d, J = 6.6 Hz, 2H), 8.58 (bs, 2H), 4.22-4.21 (d, J = 5.0 Hz, 1 H), 4.20-4.19 (d, J = 5.0 Hz, 2H), 3.81 (s, 6H), 1.97-1.90 (m, 2H), 1.52-1.44 (m, 2H), 1.37-1.28 (m, 2H), 0.96-0.92 (m, 12H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$+1 drop TFA) δ: 170.4, 164.1, 161.4-160.8 (residual TFA signal), 150.5, 116.4-114.1 (residual TFA signal), 82.4, 62.1, 53.0, 38.7, 25.2, 14.8, 11.4; HRMS (ESI) calculated $[M+H]^+$ for $C_{19}H_{32}N_5O_6$: 426. 2347, found 426.2339 $[M+H]^+$.



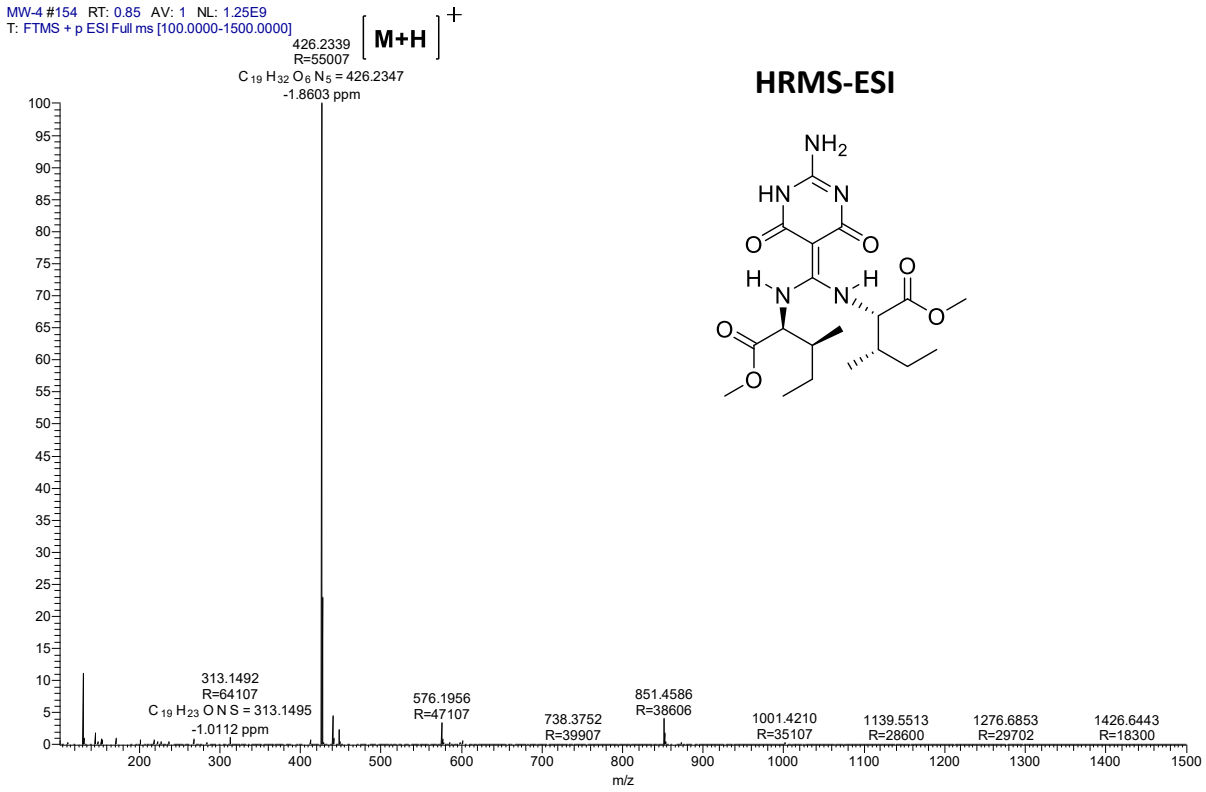
^{1}H NMR spectrum of compound **4d** ($CDCl_3$ +1 drop TFA, 500 MHz, 298 K)



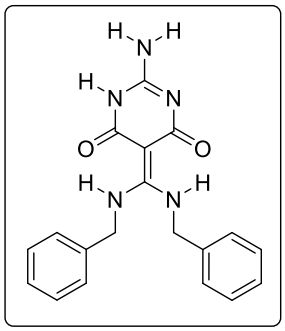
$^{13}C\{^{1}H\}$ NMR spectrum of compound **4d** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



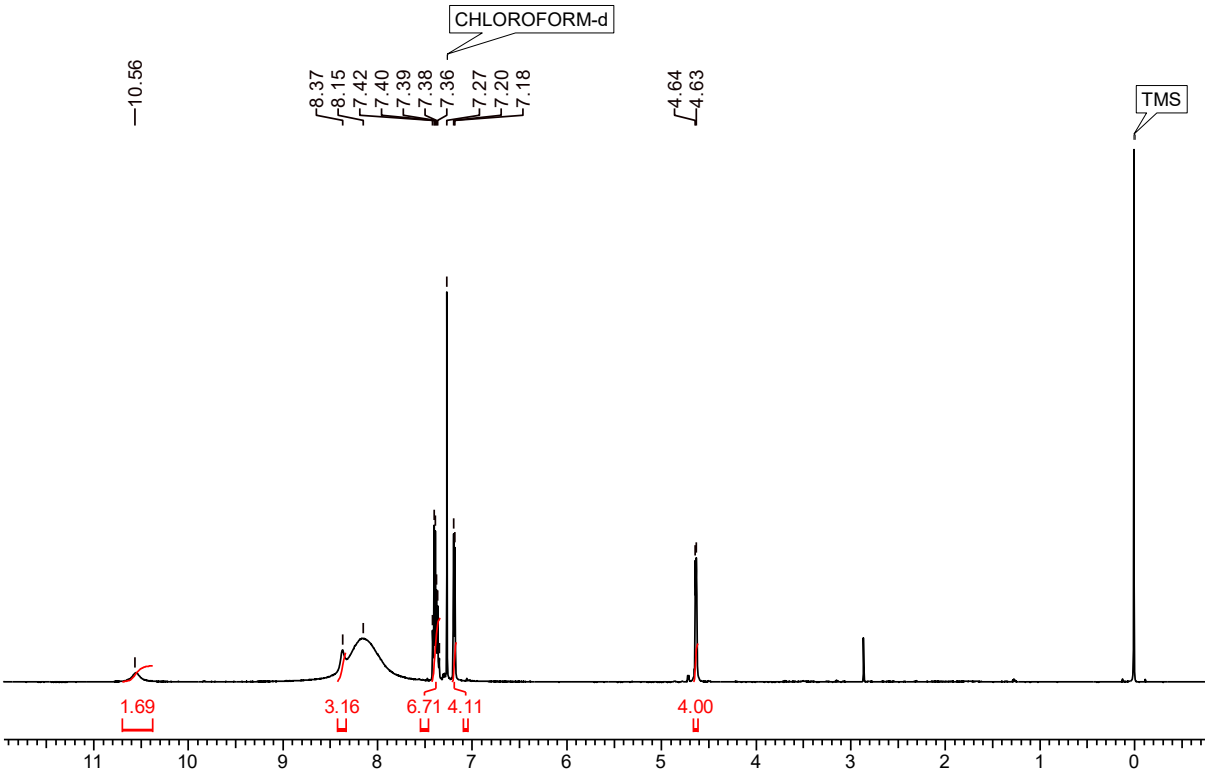
DEPT-135 NMR spectrum of compound **4d** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



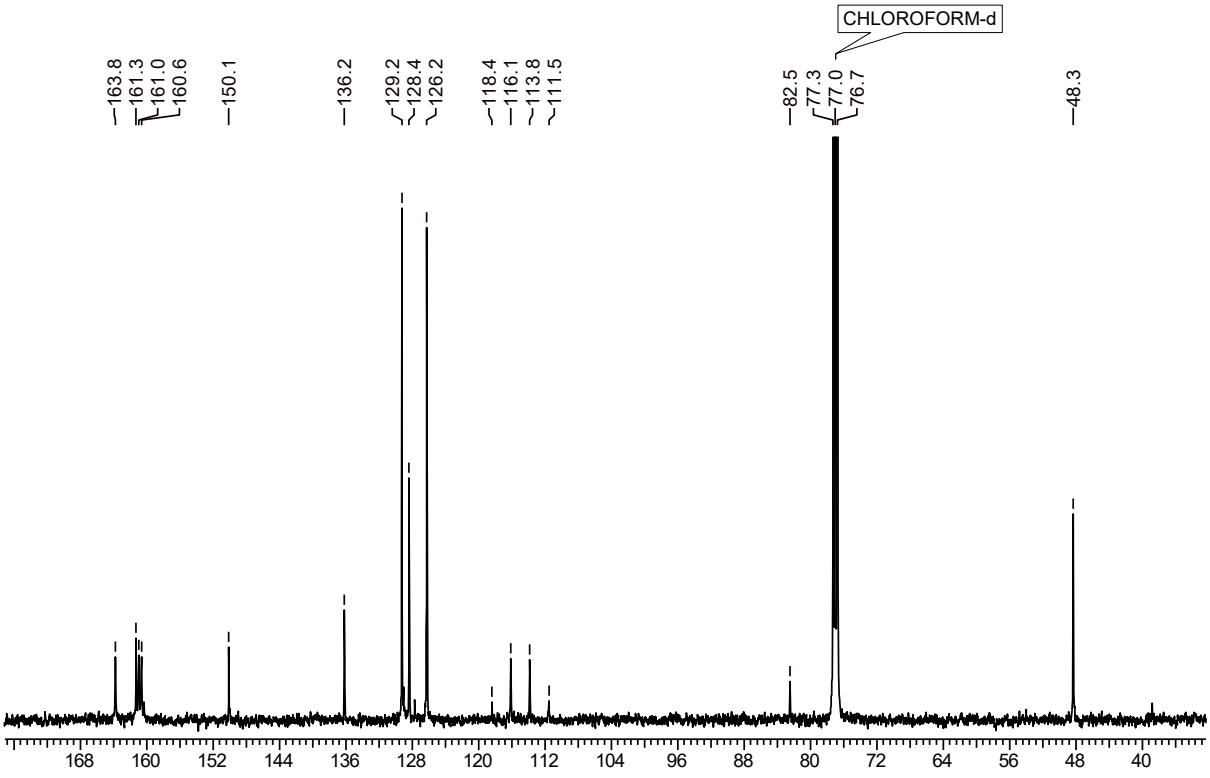
Compound 4e



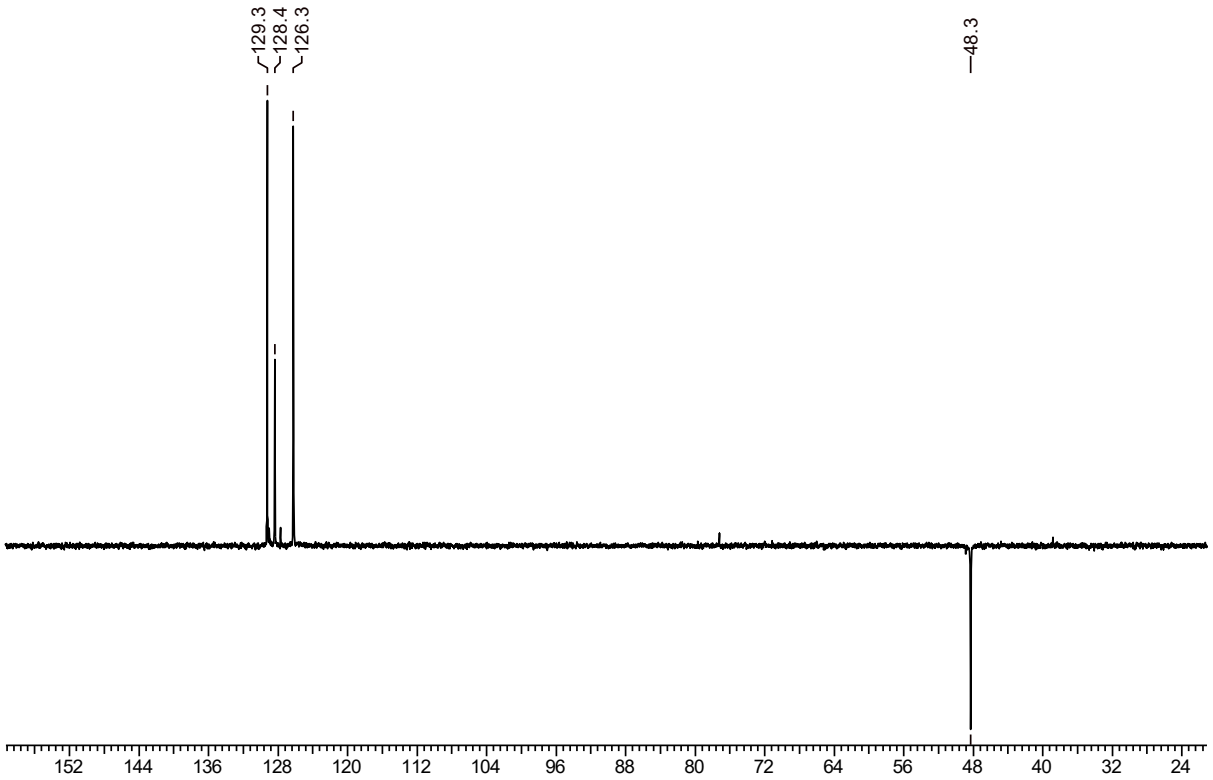
4e was obtained following the similar procedure of **4a**. Yield: 80%, white solid. *Note:* We could not purify this compound by column chromatography as it was not soluble in organic solvents. But before submitting for NMR, the solid compound was washed with MeOH (3 times) and diethyl ether. The NMR was taken in $CDCl_3$ containing 1 drop of TFA. mp: >250 °C; ^{1}H NMR (500 MHz, $CDCl_3$ +1 drop TFA) δ: 10.56 (bs, 2H), 8.37 (bs, 3H), 8.15 (residual TFA signal), 7.42-7.35 (m, 6H), 7.20-7.18 (d, J = 7.1 Hz, 4H), 4.64-4.63 (d, J = 5.5 Hz, 4H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$ +1 drop TFA) δ: 163.8, 161.3-160.3 (residual TFA signal), 150.1, 136.2, 129.2, 128.4, 126.2, 118.4-111.5 (residual TFA signal), 82.5, 48.3; HRMS (ESI) calculated $[M+H]^+$ for $C_{19}H_{20}N_5O_2$: 350. 1612, found 350.1615 $[M+H]^+$.



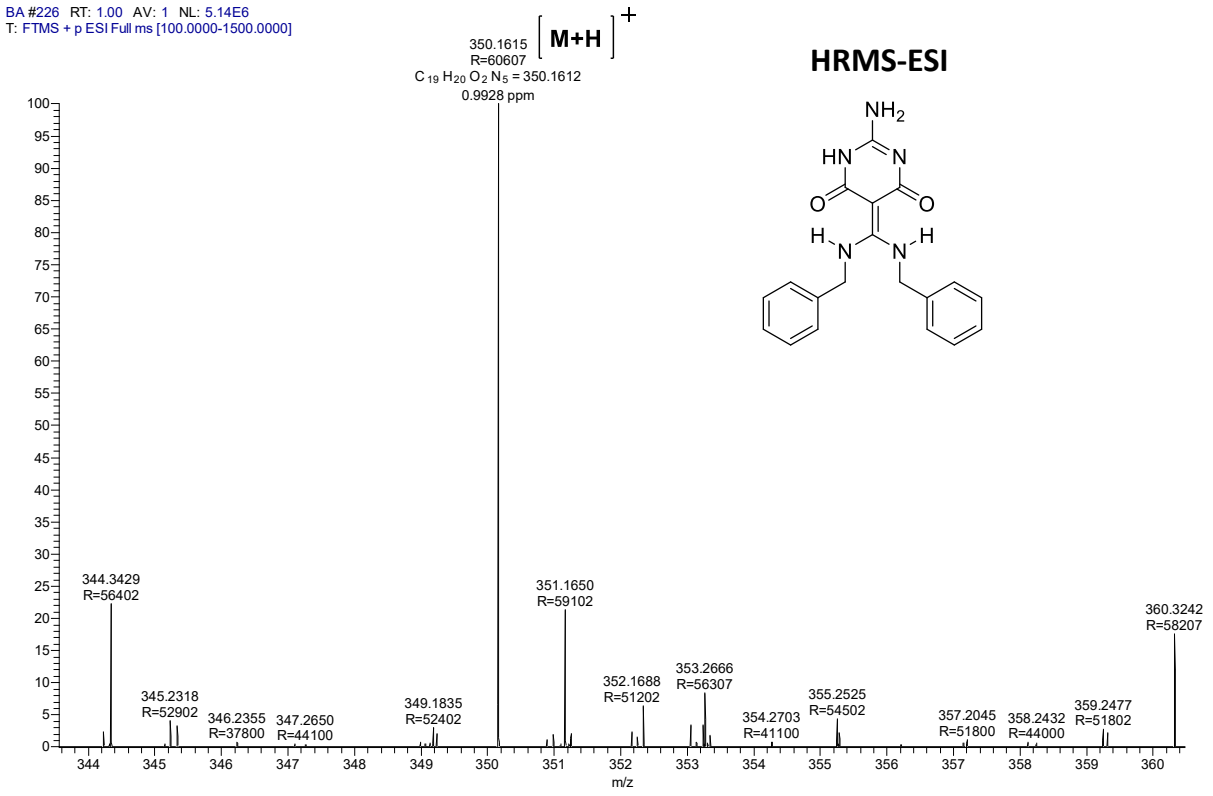
^{1}H NMR spectrum of compound **4e** ($CDCl_3$ +1 drop TFA, 500 MHz, 298 K)



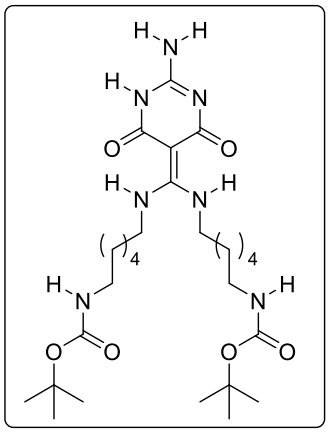
^{13}C{^{1}H} NMR spectrum of compound **4e** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



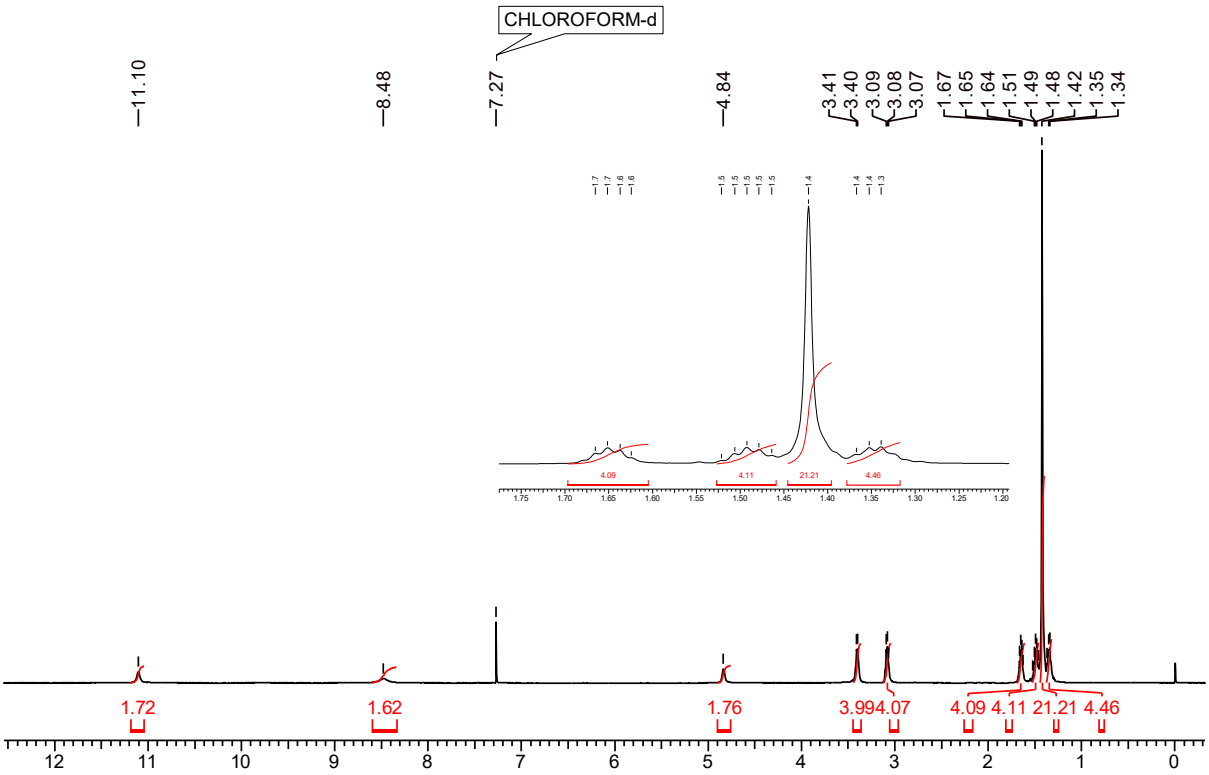
DEPT-135 NMR spectrum of compound **4e** ($CDCl_3$ +1 drop TFA, 125 MHz, 298 K)



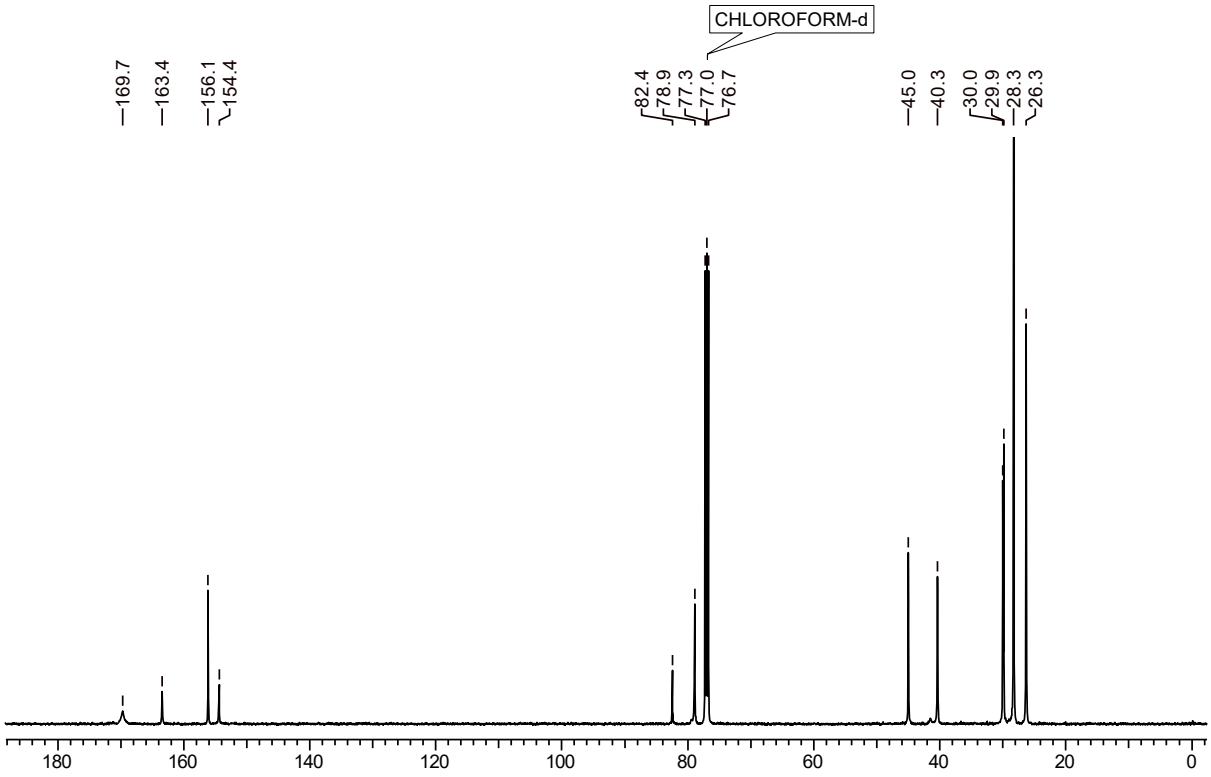
Compound 4f



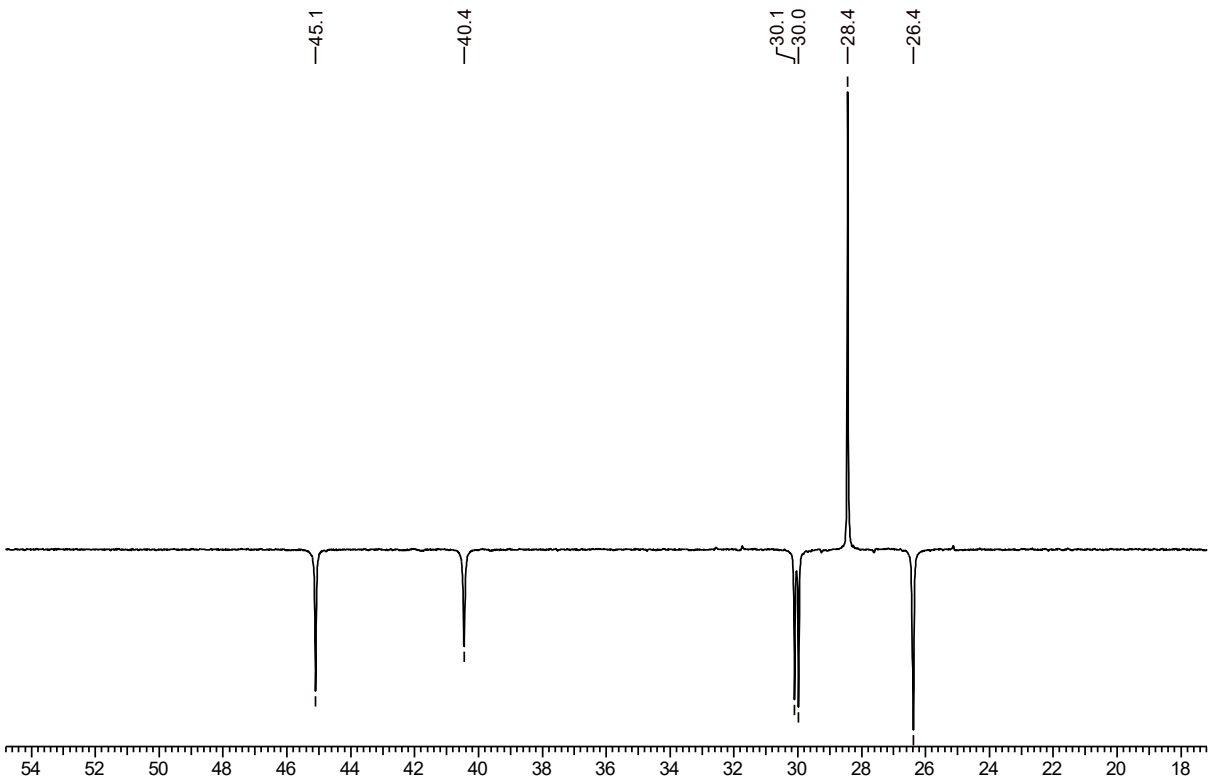
4f was obtained following the similar procedure of **4a**. The residue obtained was purified by column chromatography (eluent: 3% MeOH/DCM, R_f: 0.5) to afford **4f** (7.25 g, 62%) as a white solid: mp: 195-200 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 11.10 (bs, 2H), 8.48 (bs, 2H), 4.84 (bs, 2H), 3.41-3.40 (m, 4H), 3.10-3.07 (m, 4H), 1.67-1.62 (m, 4H), 1.52-1.46 (m, 4H), 1.42 (bs, 21H), 1.37-1.34 (m, 4H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$) δ: 169.7, 163.4, 156.1, 154.4, 82.4, 78.9, 45.0, 40.3, 30.0, 29.9, 28.3, 26.3; HRMS (ESI) calculated $[M+H]^+$ for $C_{27}H_{50}N_7O_6$: 568. 3817, found 568.3809 $[M+H]^+$.



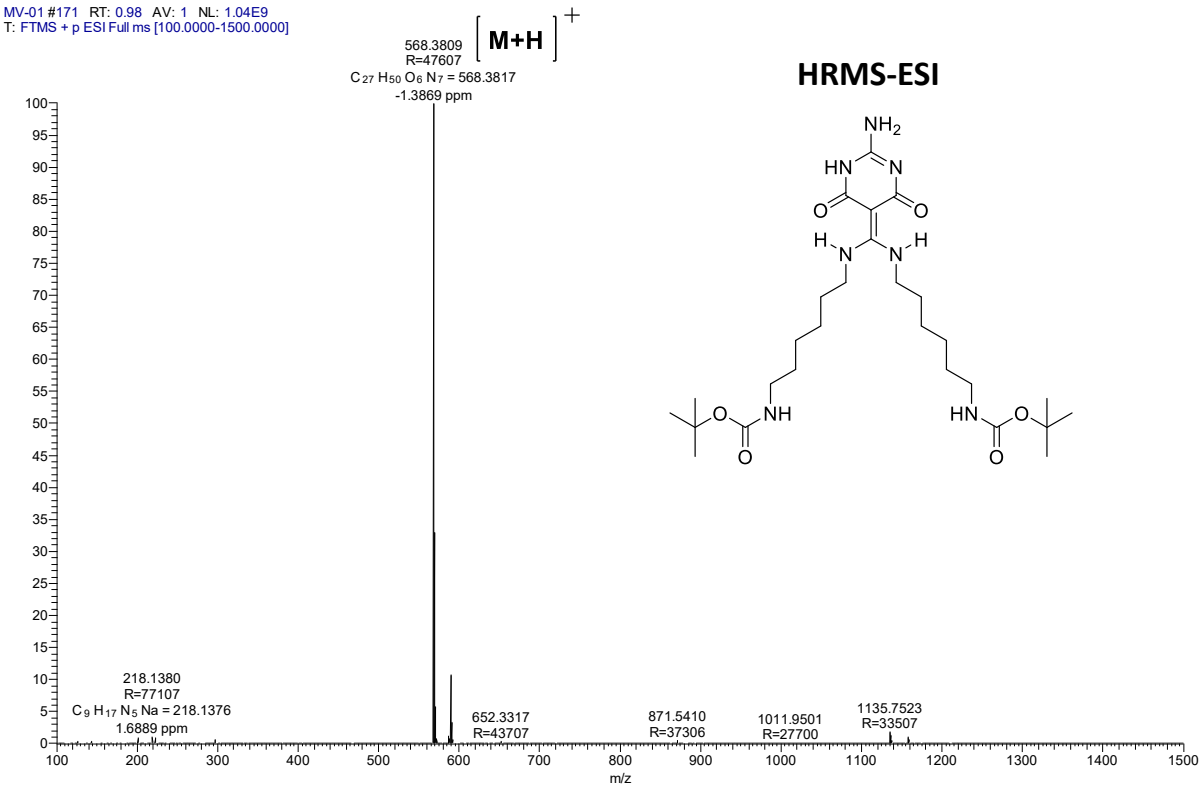
^{1}H NMR spectrum of compound **4f** ($CDCl_3$, 500 MHz, 298 K)



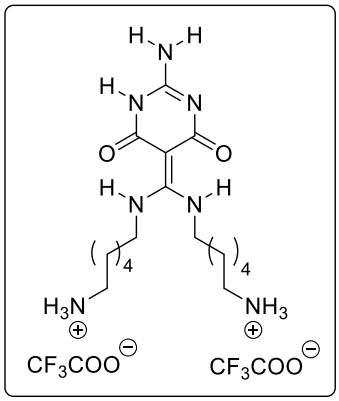
$^{13}C\{^{1}H\}$ NMR spectrum of compound **4f** ($CDCl_3$, 125 MHz, 298 K)



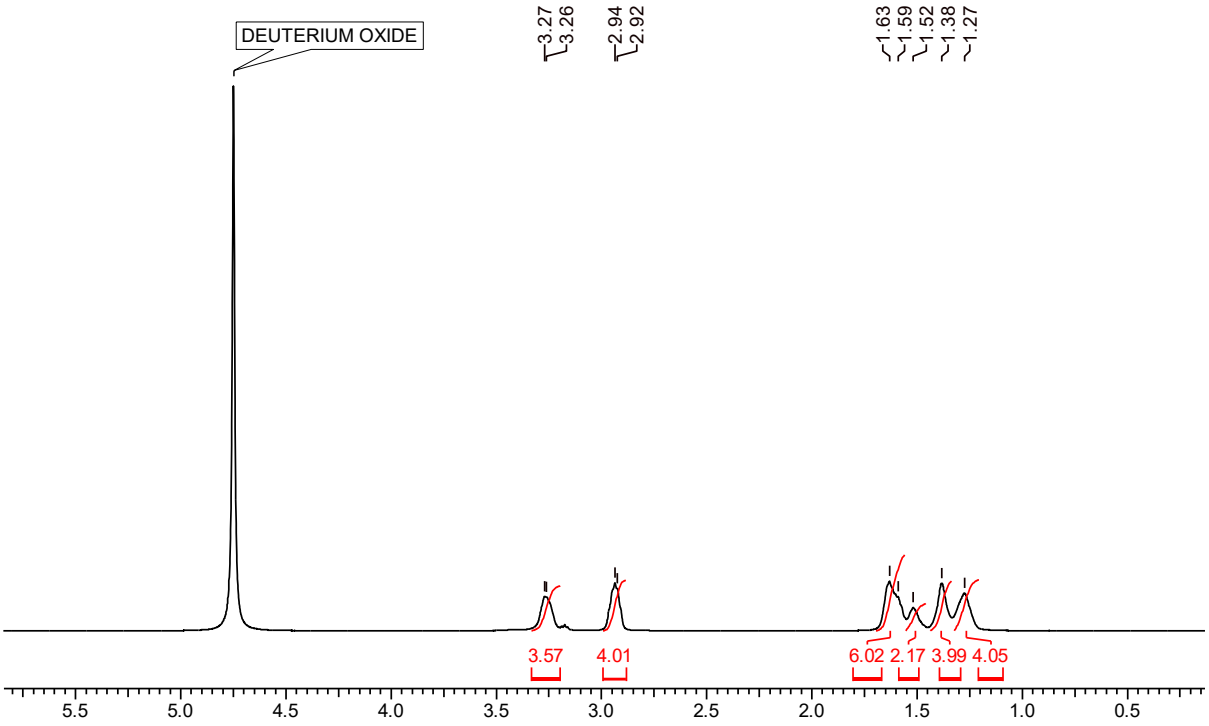
DEPT-135 NMR spectrum of compound **4f** ($CDCl_3$, 125 MHz, 298 K)



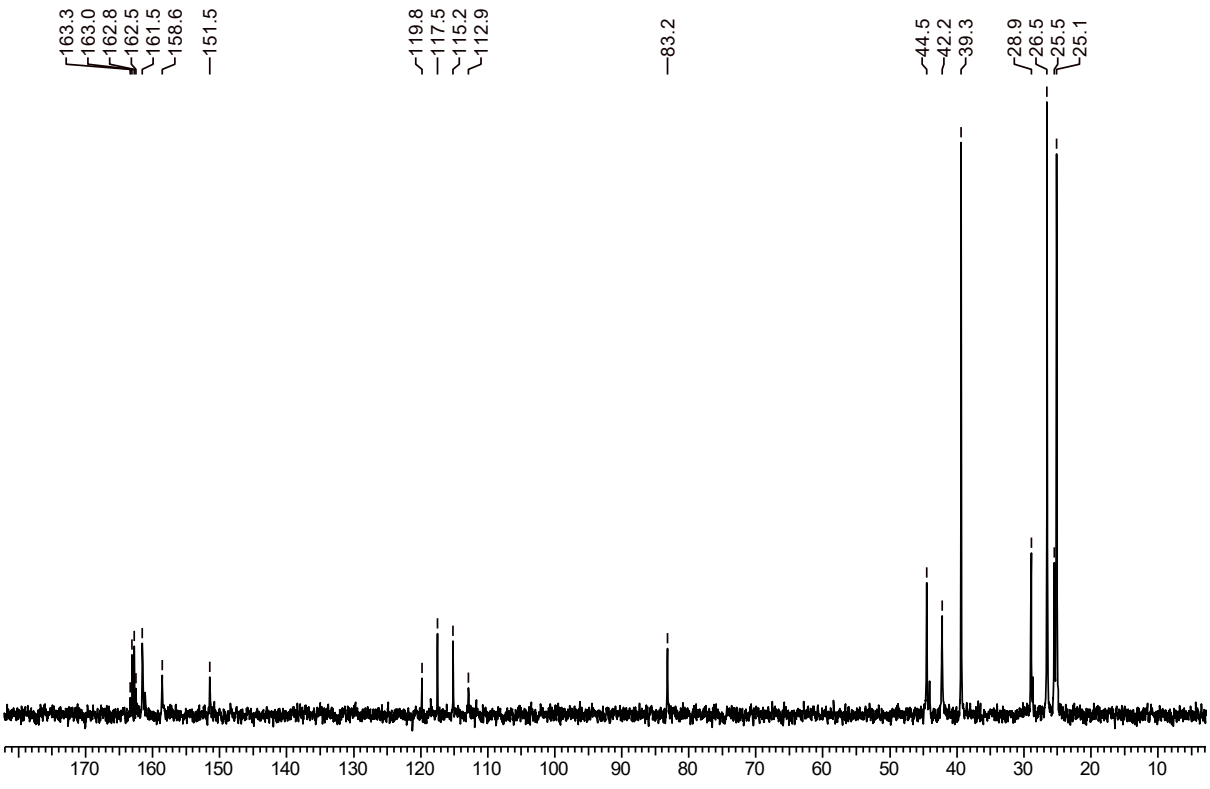
Compound 4g



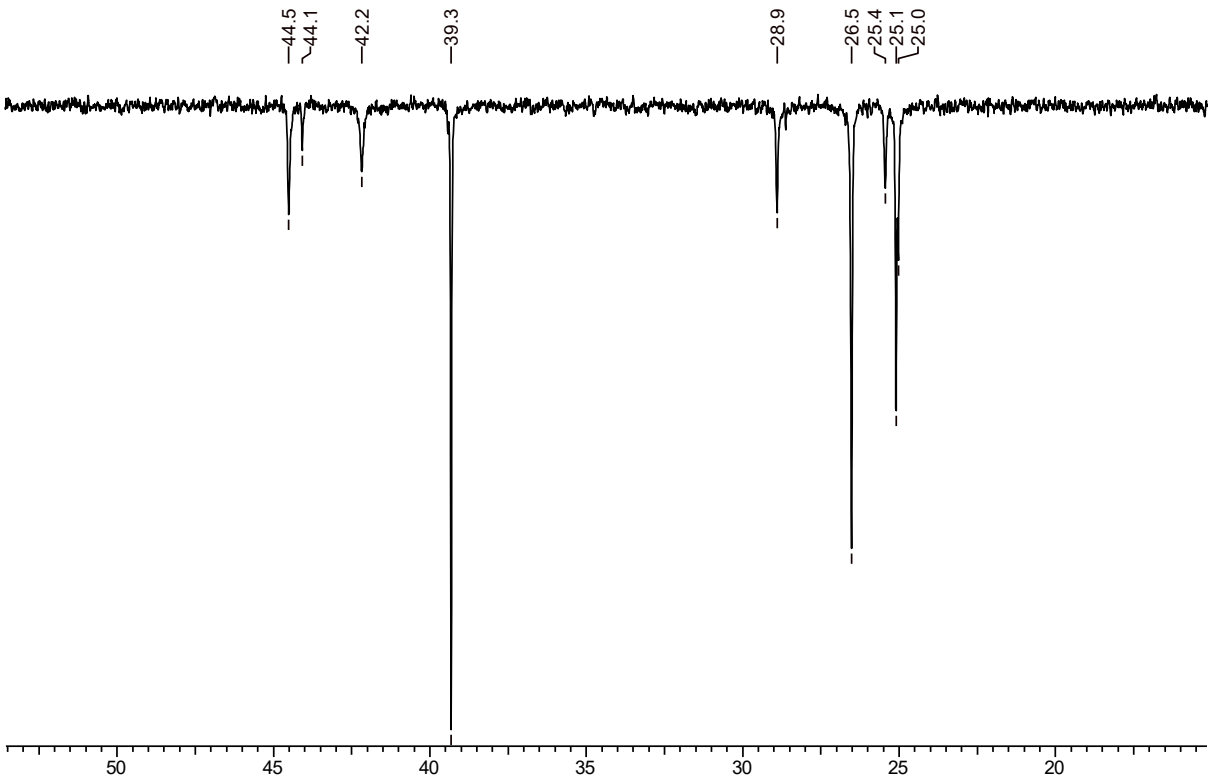
The Boc-protected compound **4f** was subjected to deprotection by using TFA:DCM (1:1) mixture for 30 minutes at 0°C. After completion of the reaction, the mixture was stripped off solvent and co-evaporated with DCM (3 times). Further, diethyl ether (Et_2O) was added to the reaction mixture and the resultant solid was washed thrice with Et_2O by decantation and dried under vacuum giving **4g** in quantitative yield, as a white solid: mp: 205-210 °C; ^{1}H NMR (500 MHz, D_2O) δ: 3.27-3.26 (m, 4H), 2.94-2.92 (m, 4H), 1.63-1.59 (m, 6H), 1.52 (bs, 2H), 1.38 (bs, 4H), 1.27(bs, 4H); $^{13}C\{^1H\}$ NMR (125 MHz, D_2O) δ: 163.3-162.5 (residual TFA signal), 161.5, 158.6, 151.5, 119.8-112.9 (residual TFA signal), 83.2, 44.5, 42.2, 39.3, 28.9, 26.5, 25.5, 25.1; HRMS (ESI) calculated $[M+H]^+$ for $C_{17}H_{34}N_7O_2$: 368. 2768, found 368.2758 $[M+H]^+$.



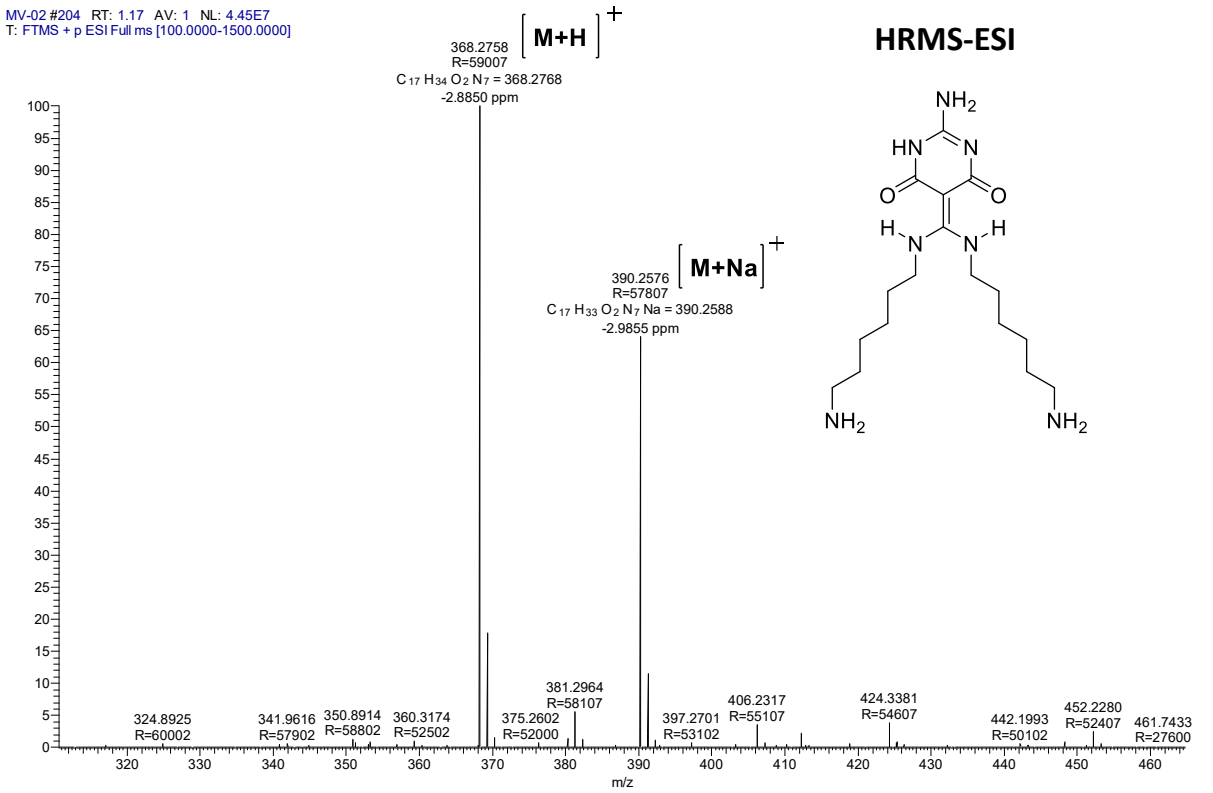
^{1}H NMR spectrum of compound **4g** (D_2O, 500 MHz, 298 K)



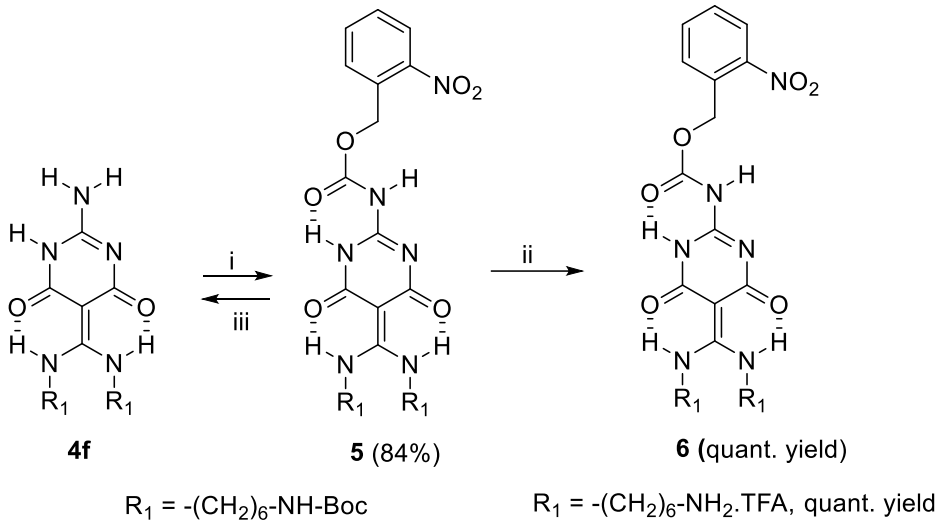
^{13}C{^{1}H} NMR spectrum of compound **4g** (D_2O, 125 MHz, 298 K)



DEPT-135 NMR spectrum of compound **4g** (D_2O, 125 MHz, 298 K)

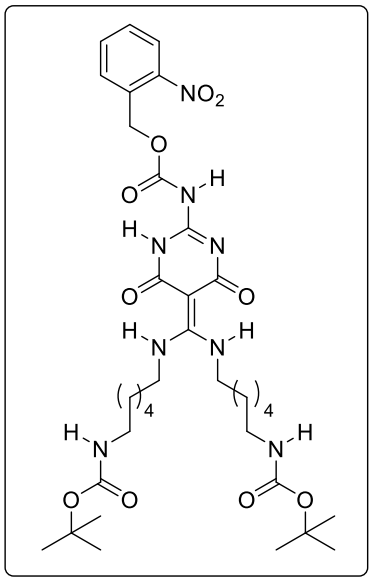


Scheme 2:

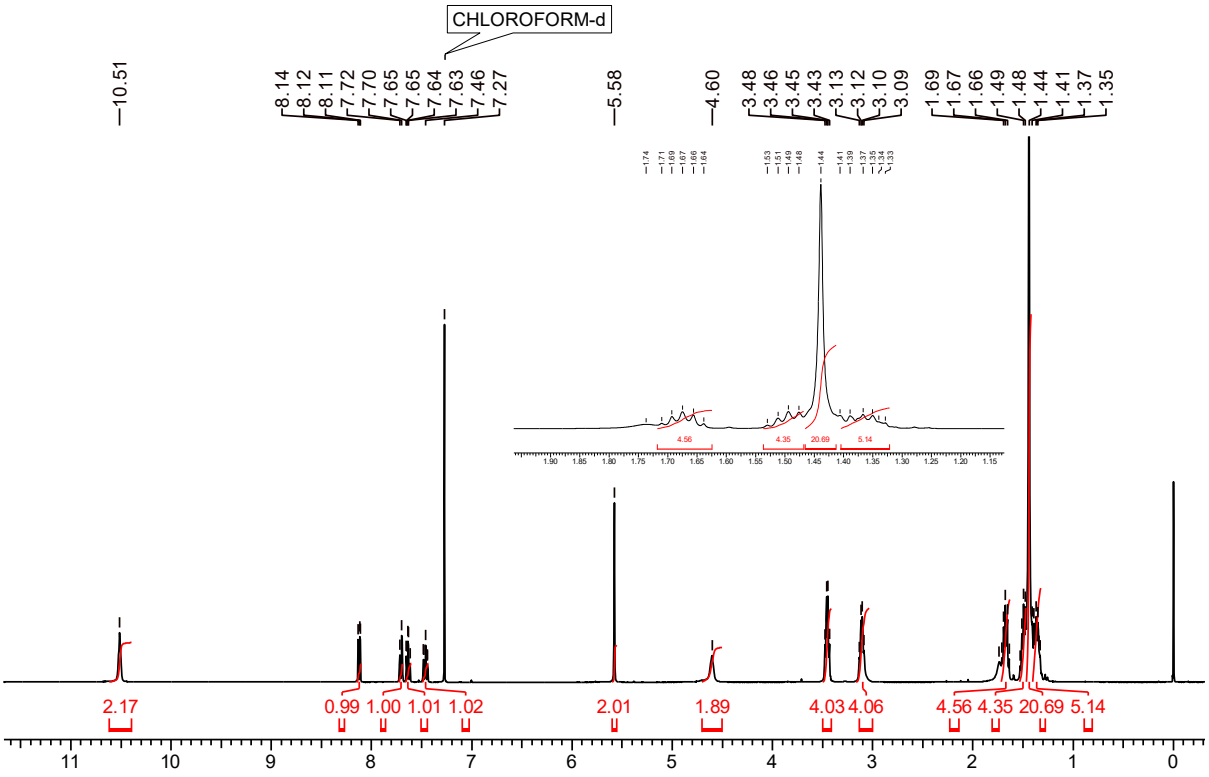


Reaction conditions: i) 2-nitrobenzyl chloroformate, DIPEA, THF, 0 °C-rt, 4h; ii) TFA:DCM (50:50), 0 °C-rt, 30 min; iii) THF (10 mg mL^{-1}), 360-370 nm UV light, rt, 4h, 82%.

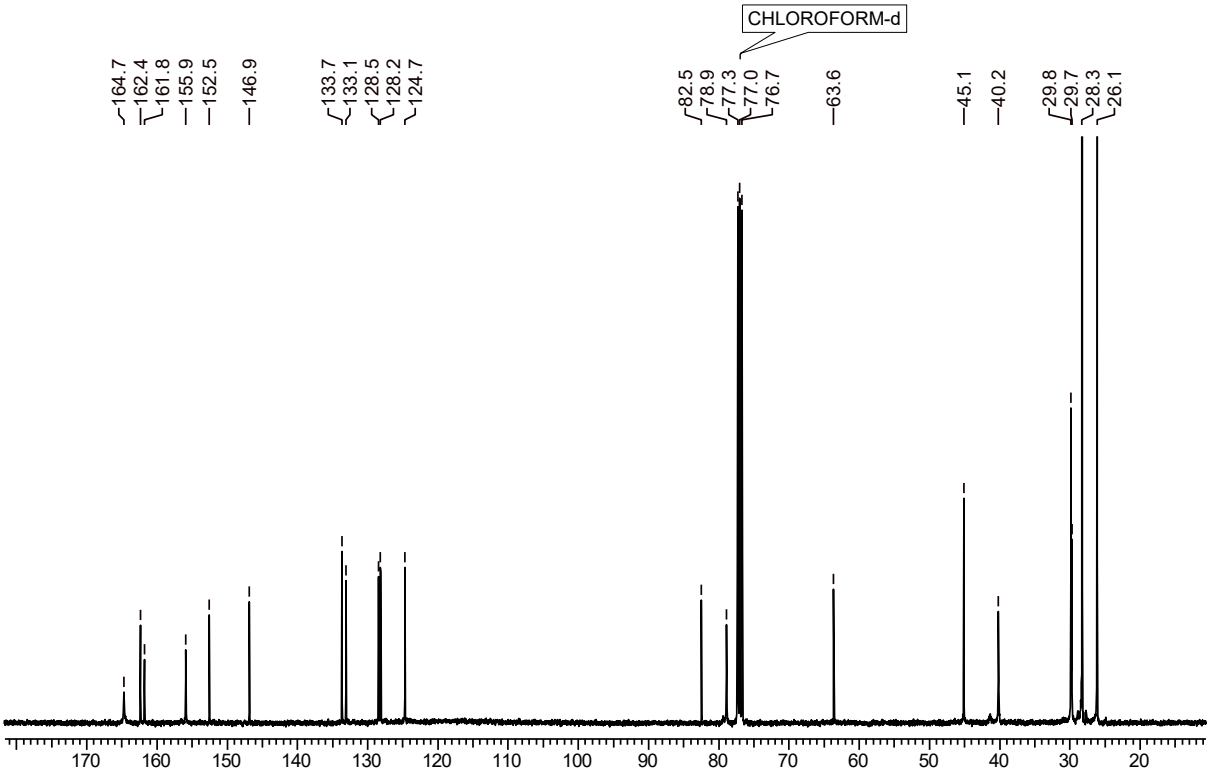
Compound 5



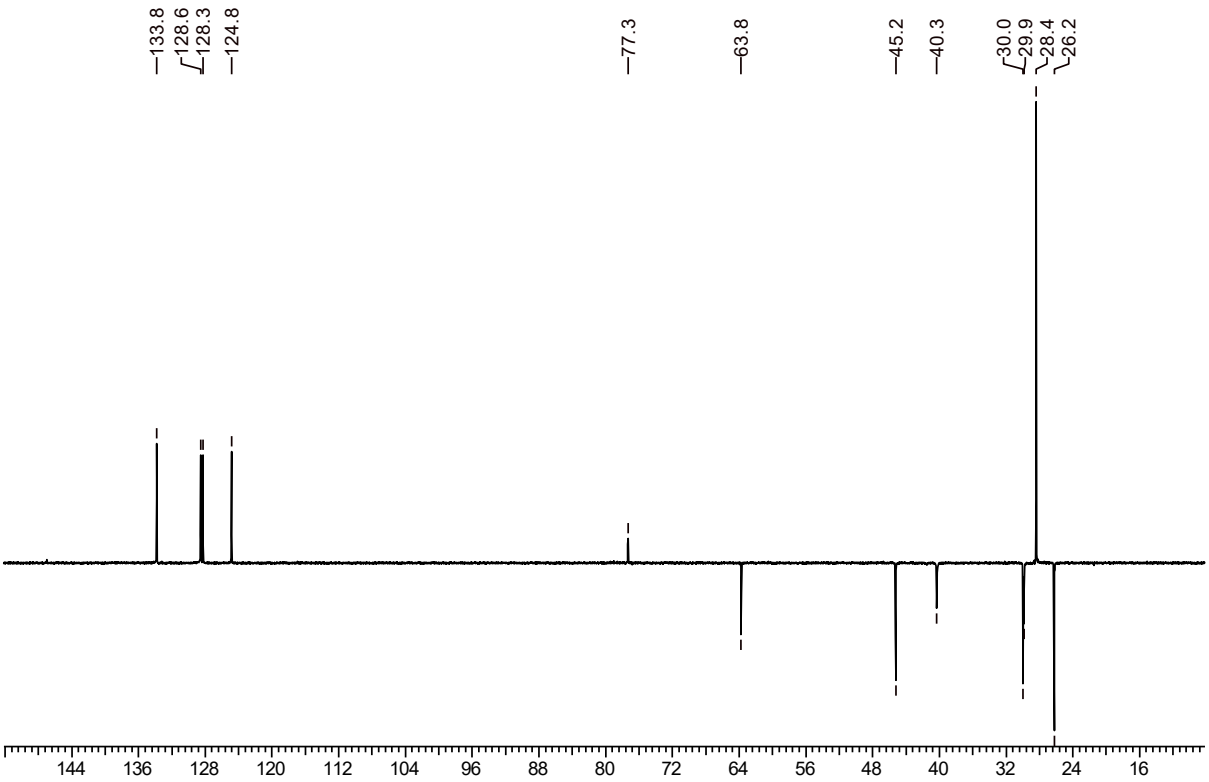
To a solution of **4f** (1 g, 1.76 mmol, 1 equiv.) in dry THF at 0 °C, DIPEA (0.61 ml, 3.52 mmol, 2 equiv.) and 2-nitrobenzyl chloroformate (0.49 g, 2.29 mmol, 1.3 equiv.) were added and the reaction mixture stirred at room temperature until the starting material completely disappeared. After completion of the reaction, the mixture was evaporated on a rotary evaporator. The resultant reaction mixture was partitioned between DCM and water. The organic layer was separated, washed with water, aq. citric acid and brine solution, dried over Na_2SO_4, filtered and concentrated *in vacuo*. Purification by column chromatography (eluent: 2% MeOH/DCM, R_f: 0.5) afforded **5** (1.11 g, 84%) as a sticky solid; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.51 (bs, 2H), 8.14-8.11 (dd, *J* = 1.1, 8.2 Hz, 1H), 7.72-7.70 (m, 1H), 7.65-7.61 (m, 1H), 7.48-7.44 (m, 1H), 5.58 (s, 2H), 4.60 (bs, 2H), 3.48-3.43 (m, 4H), 3.13-3.09 (m, 4H), 1.71-1.64 (m, 4H), 1.53-1.48 (m, 4H), 1.44 (bs, 21H), 1.41-1.33 (m, 5H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ: 164.7, 162.4, 161.8, 155.9, 152.5, 146.9, 133.7, 133.1, 128.5, 128.2, 124.7, 82.5, 78.9, 63.6, 45.1, 40.2, 29.8, 29.7, 28.3, 26.1; HRMS (ESI) calculated $[M+H]^+$ for $C_{35}H_{55}N_8O_{10}$: 747.4036, found 747.4026 $[M+H]^+$.



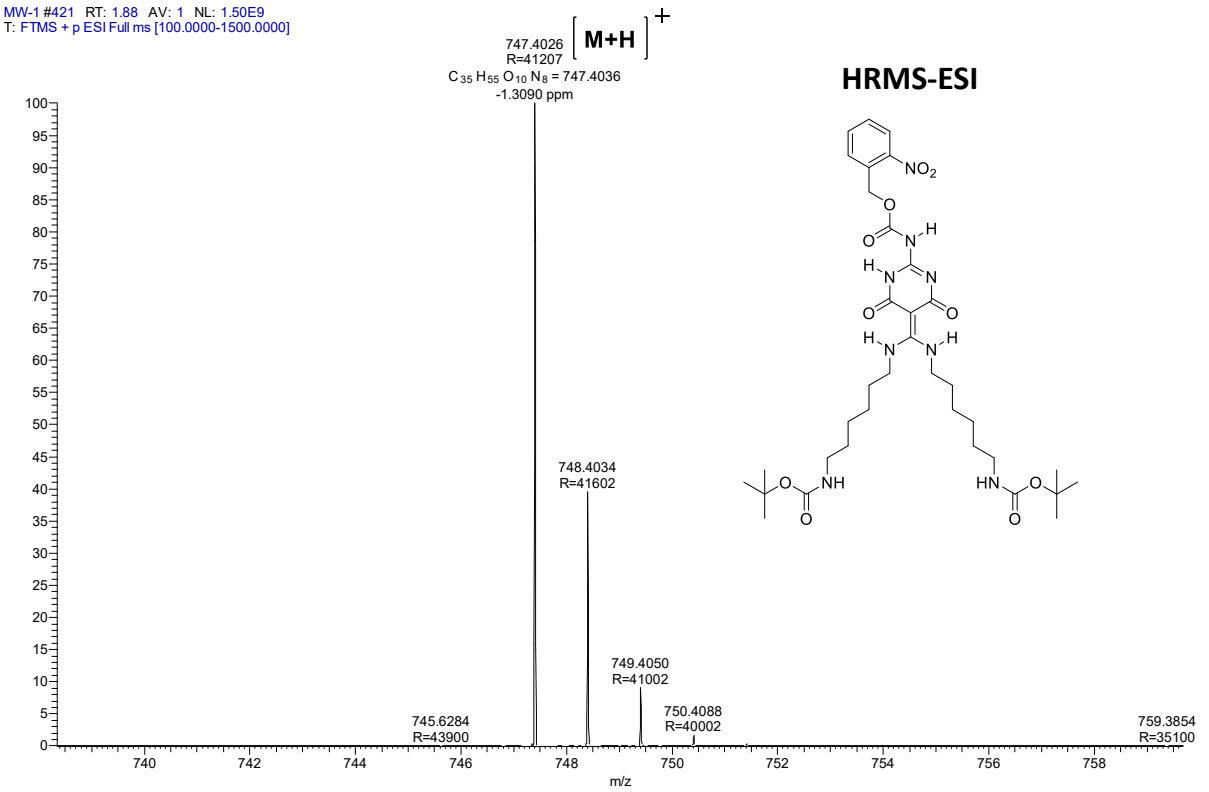
^{1}H NMR spectrum of compound **5** ($CDCl_3$, 400 MHz, 298 K)



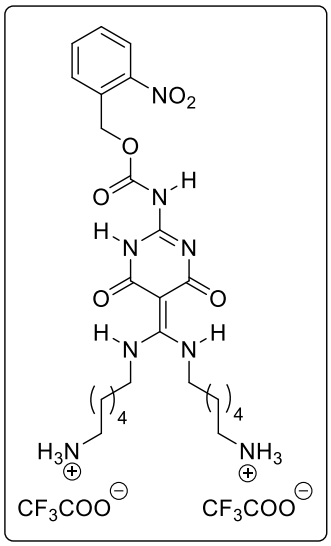
^{13}C{^{1}H} NMR spectrum of compound **5** ($CDCl_3$, 100 MHz, 298 K)



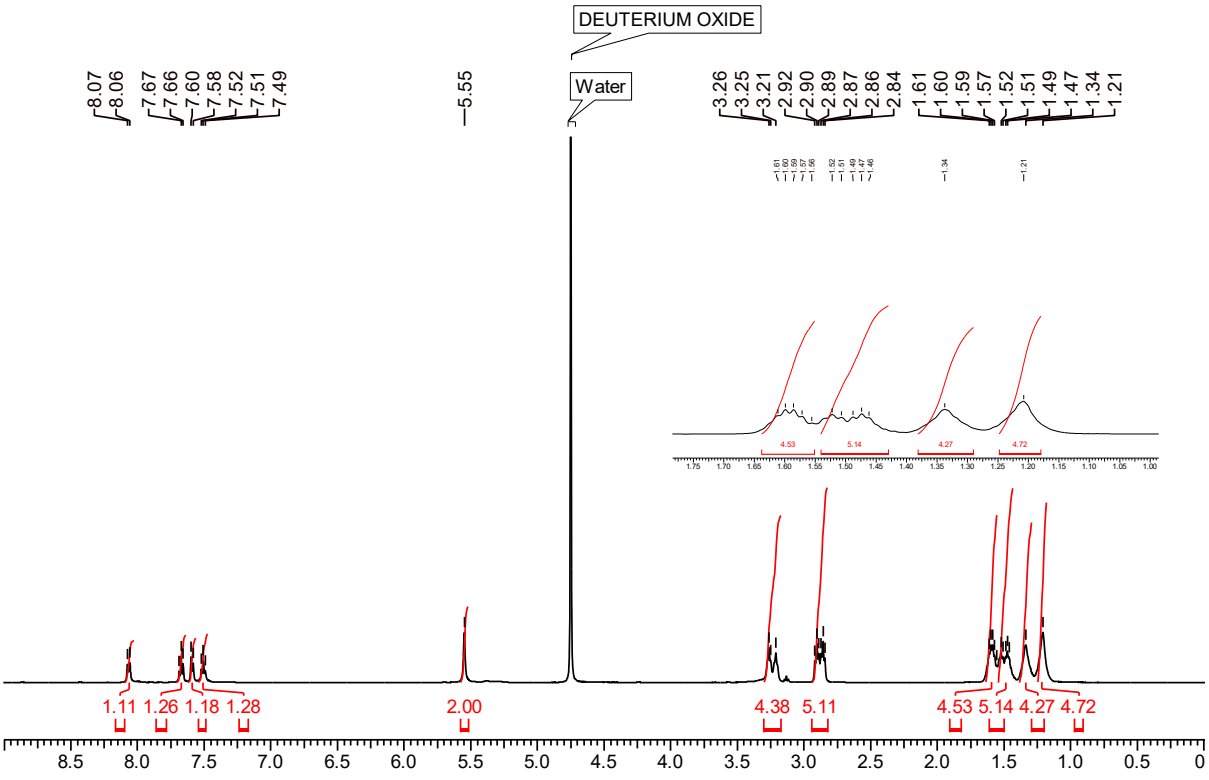
DEPT-135 NMR spectrum of compound **5** ($CDCl_3$, 100 MHz, 298 K)



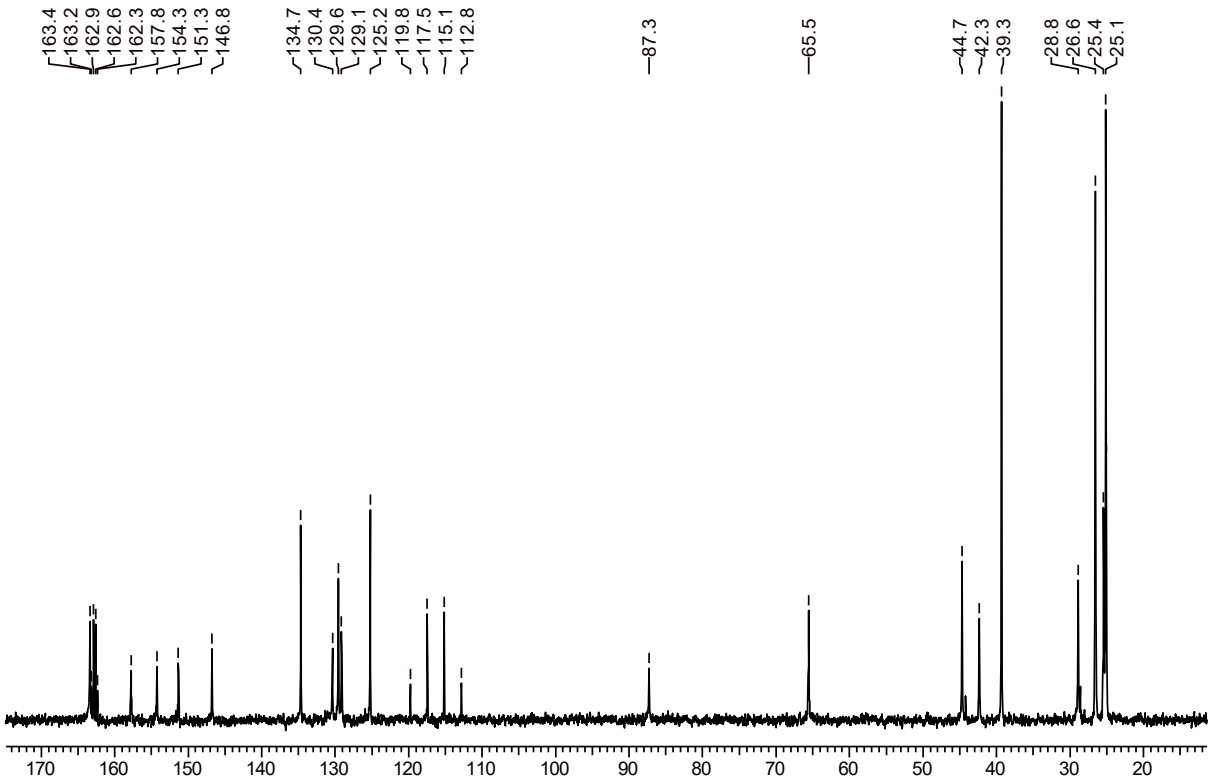
Compound 6



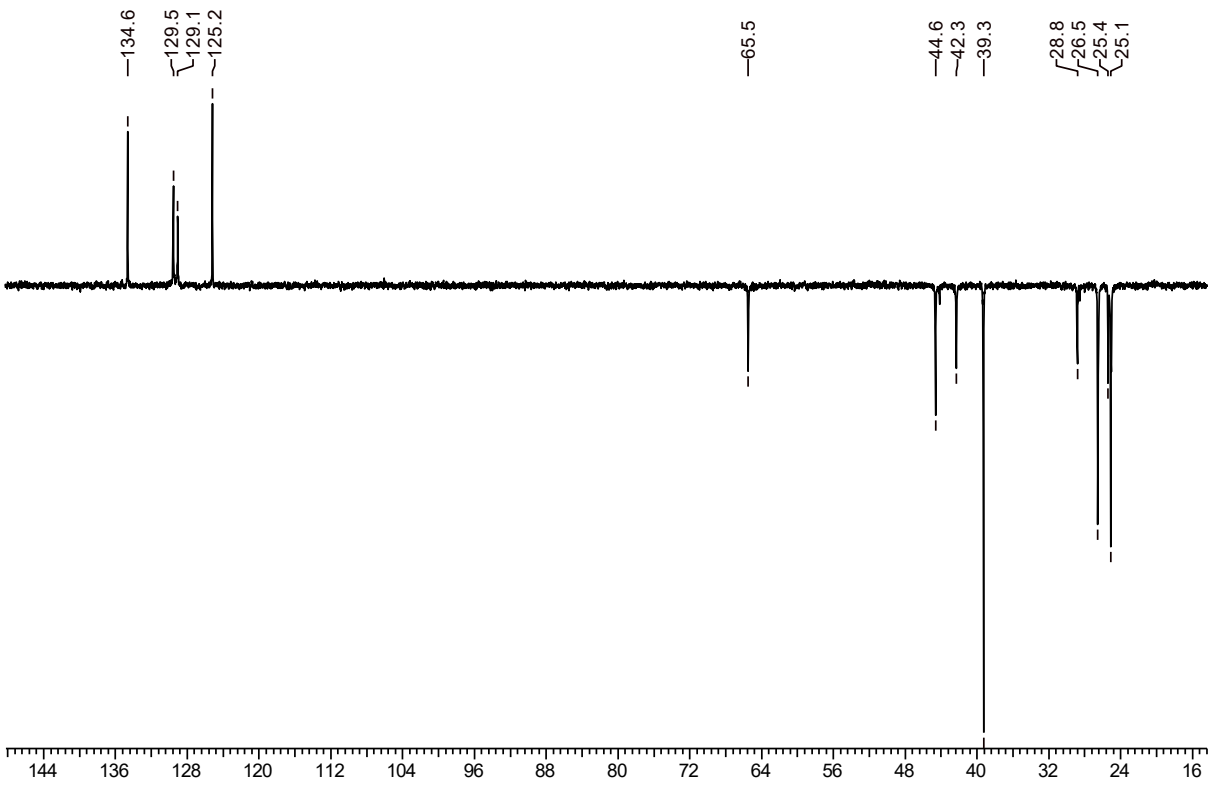
The Boc deprotection of **5** was carried out using the same procedure employed for **4g** to afford **6** in quantitative yield, as a white solid. ^{1}H NMR (500 MHz, D_2O) δ: 8.07-8.06 (d, J = 7.9 Hz, 1H), 7.69-7.66 (m, 1H), 7.60-7.58 (m, 1H), 7.52-7.49 (m, 1H), 5.55 (bs, 2H), 3.26-3.21 (m, 4H), 2.92-2.84 (m, 5H), 1.61-1.56 (m, 4H), 1.52-1.46 (m, 5H), 1.34 (bs, 4H), 1.21 (bs, 5H); ^{13}C{^{1}H} NMR (125 MHz, D_2O) δ: 163.4-162.3 (residual TFA signal), 157.8, 154.3, 151.3, 146.8, 134.7, 130.4, 129.6, 129.1, 125.2, 119.8-112.8 (residual TFA signal), 87.3, 65.5, 44.7, 42.3, 39.3, 28.8, 26.6, 25.4, 25.1, 25.1; HRMS (ESI) calculated $[M+H]^+$ for $C_{25}H_{39}N_8O_6$: 547. 2987, found 547.3002 $[M+H]^+$.



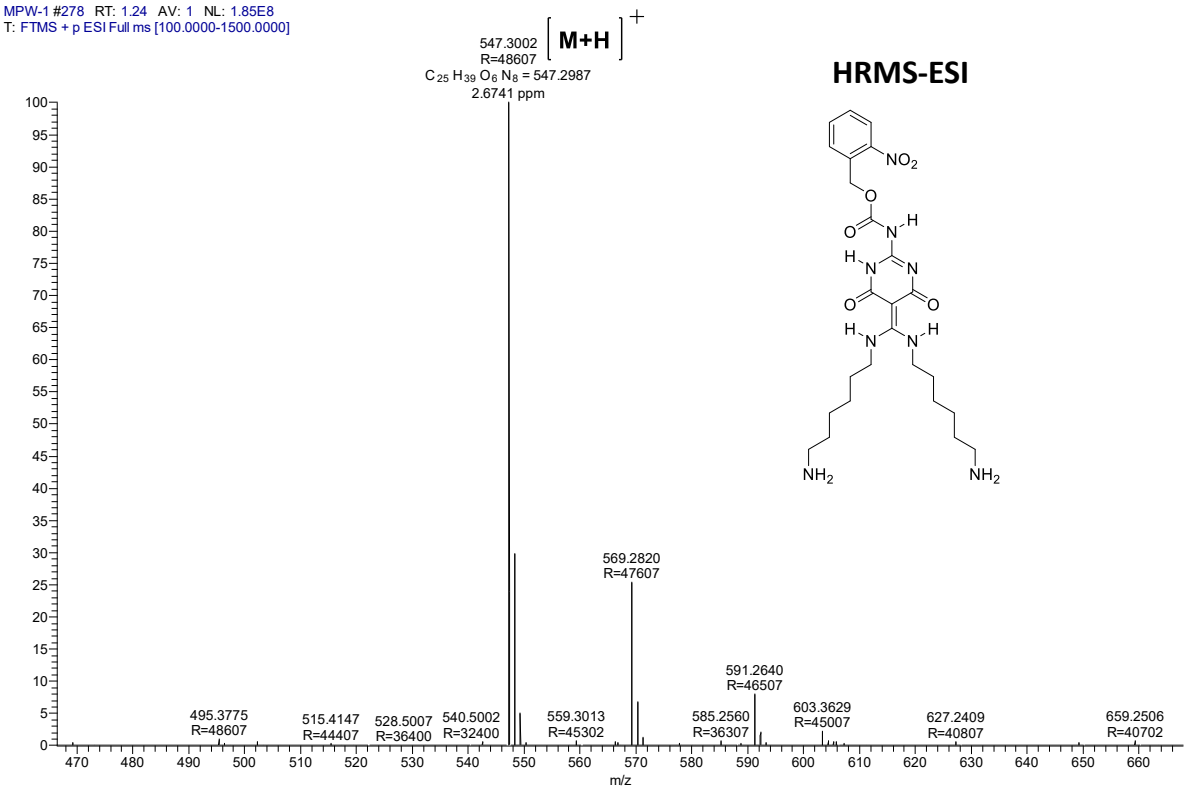
^{1}H NMR spectrum of compound **6** (D_2O, 500 MHz, 298 K)



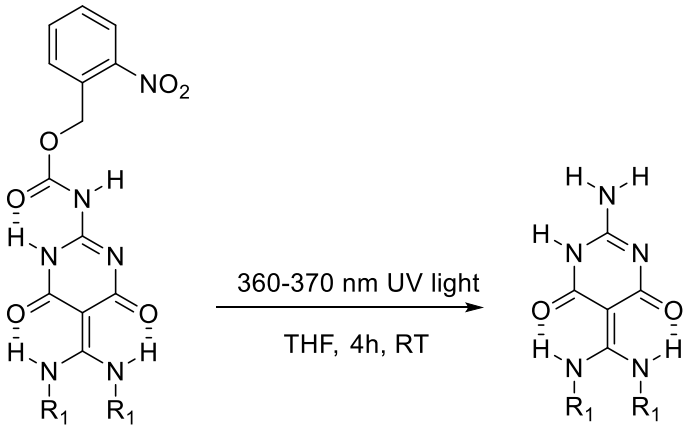
$^{13}C\{^{1}H\}$ NMR spectrum of compound **6** (D_2O, 125 MHz, 298 K)



DEPT-135 NMR spectrum of compound **6** (D_2O, 125 MHz, 298 K)



Deprotection of *O*-NB group of compound 5

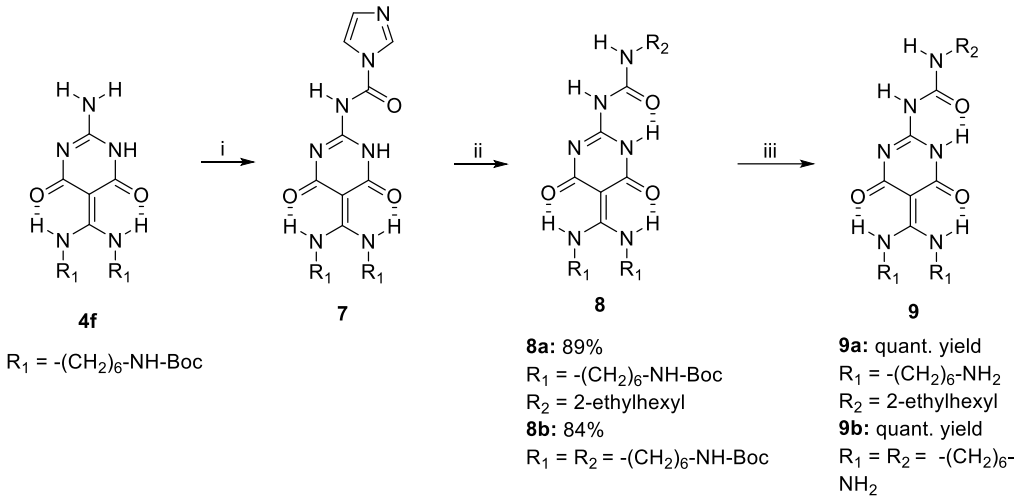


5 R_1 = -$(CH_2)_6$-NH-Boc **4f** (82%)

Compound **5** (10 mg) was dissolved in THF (1 ml) and kept under 360-370 nm UV light while vigorously stirring for 4h. After completion of the reaction, the mixture was evaporated on a rotary evaporator and purified by column chromatography (eluent: 3% MeOH/DCM, R_f: 0.5) to afford **4f** (8.2 mg, 82%) as a white solid. Its authenticity was confirmed by comparing its TLC and ^{1}H NMR with that **4f** obtained by an alternate route (scheme-1).

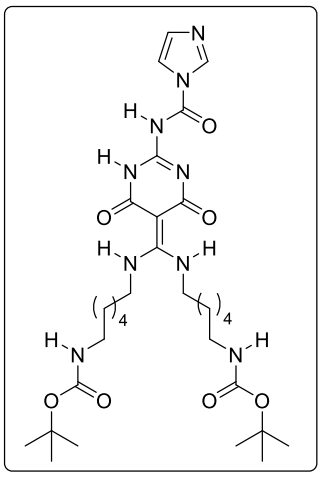
Note: Characterization data of compound **4f** is available in S16-S18.

Scheme 3:



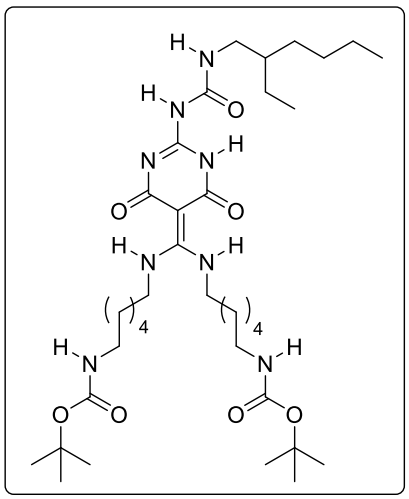
Reaction conditions: i) CDI, $CHCl_3$, rt, 4h; ii) R_2-NH_2, $CHCl_3$, rt, 5h; iii) TFA:DCM (50:50), 0 °C-rt, 30 min.

Compound 7



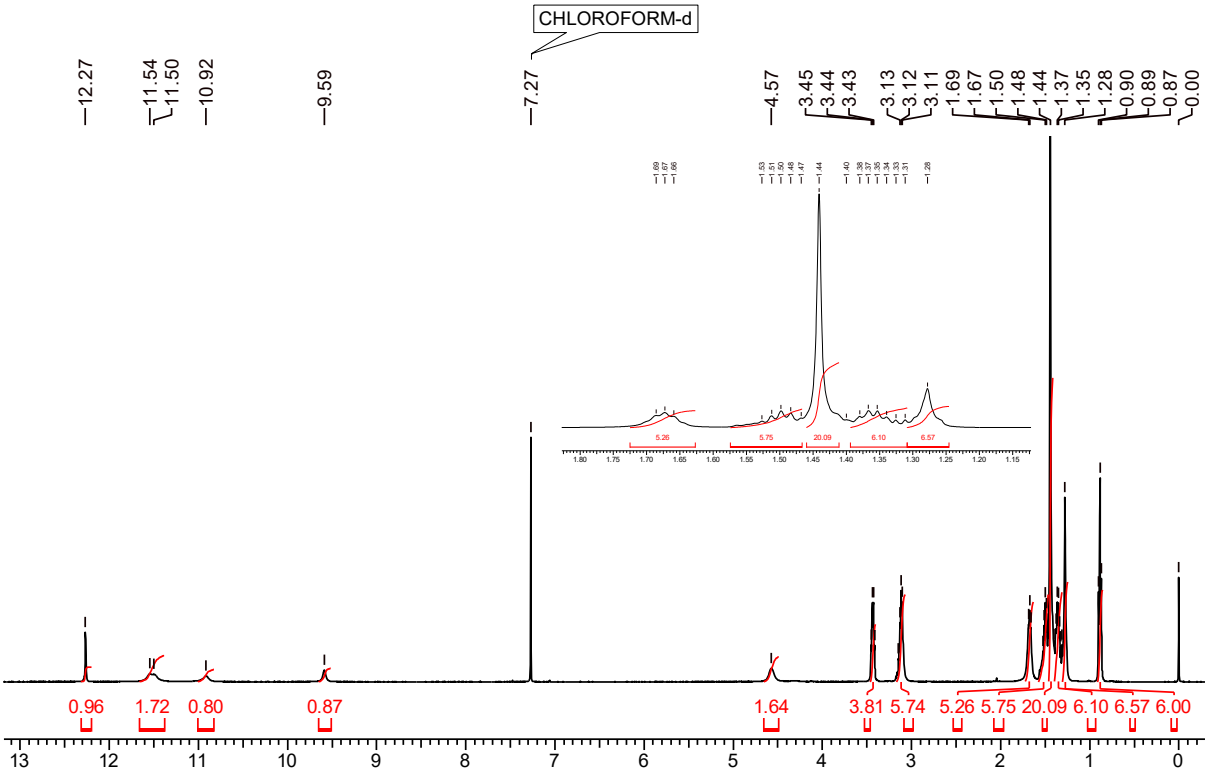
To a solution of **4f** (1 g, 1.76 mmol, 1 equiv.) in dry $CHCl_3$, carbonyldiimidazole (0.37 g, 2.29 mmol, 1.3 equiv.) was added at 0 °C and the reaction mixture stirred at room temperature until the starting material completely disappeared. After completion of the reaction, the reaction mixture was directly used (without any purifications) for further reactions.

Compound 8a

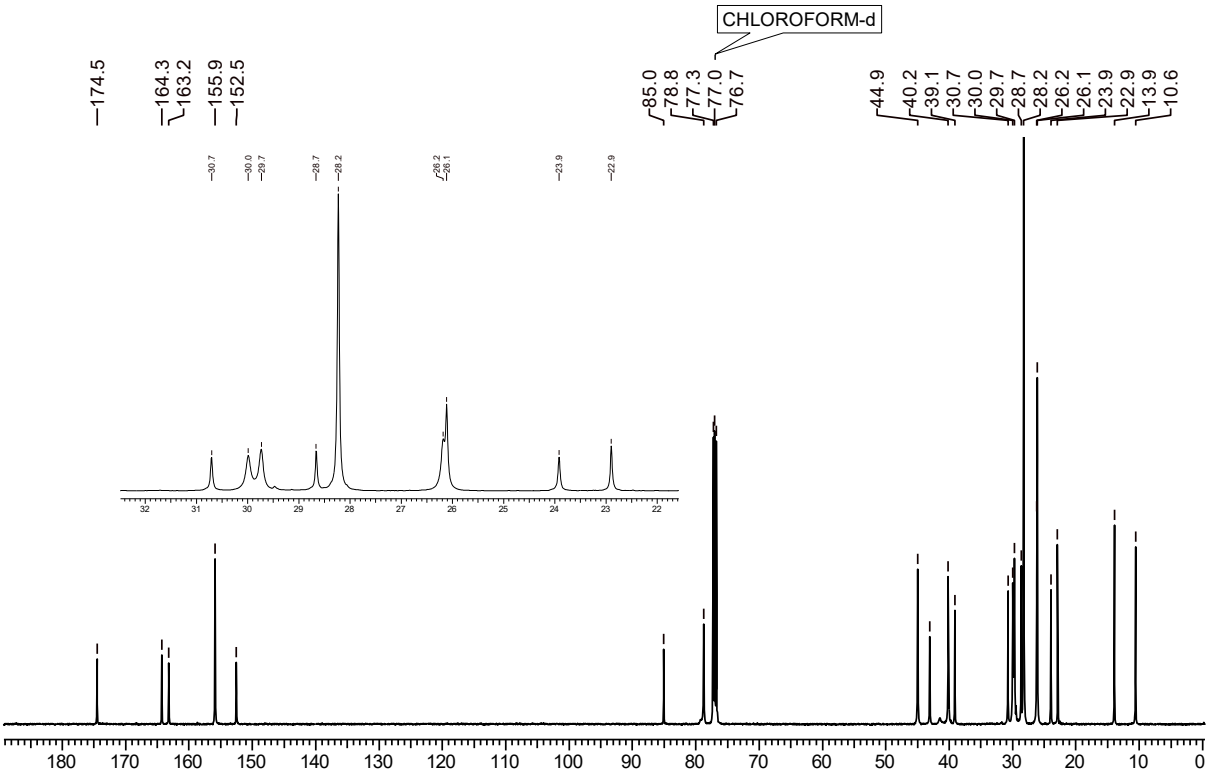


To a solution of freshly prepared **7** (0.50 g, 0.755 mmol, 1 equiv.) in dry $CHCl_3$, 2-ethylhexylamine (0.12 g, 0.982 mmol, 1.3 equiv.) was added and the reaction mixture stirred at room temperature until the starting material completely disappeared. After completion of the reaction, the mixture was evaporated on a rotary evaporator. The resultant reaction mixture was partitioned between DCM and water. The organic layer separated, washed with water, aq. citric acid, and brine solution, dried over Na_2SO_4, filtered and concentrated *in vacuo*. Purification by column chromatography (eluent: 2% MeOH/DCM, R_f: 0.5) afforded **8a** (0.49 g, 89%) as a white solid: mp: 75-80 °C; ^{1}H NMR

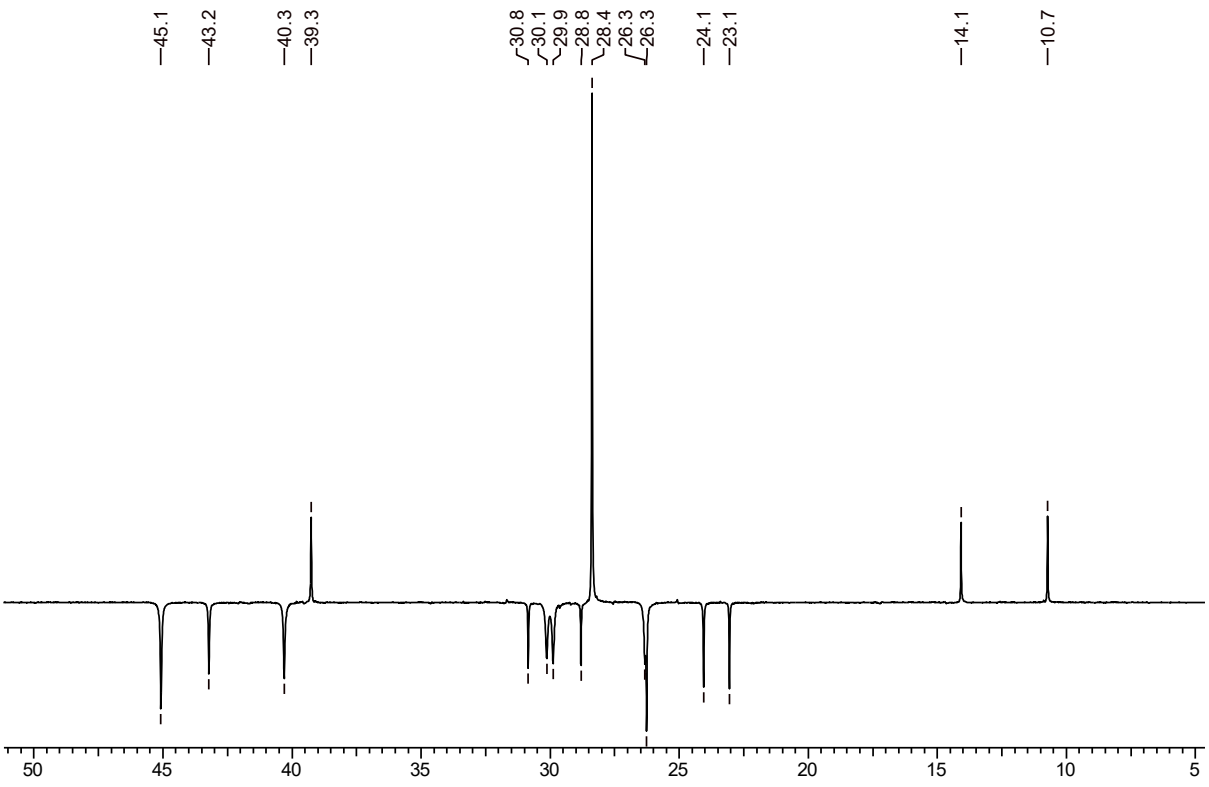
(500 MHz, $CDCl_3$) δ: 12.27 (S, 1H), 11.54-11.50 (bs, 2H), 10.92 (bs, 1H), 9.59 (bs, 1H), 4.57 (bs, 1H), 3.45-3.41 (m, 4H), 3.17-3.11 (m, 6H), 1.69-1.66 (m, 5H), 1.56-1.47 (m, 6H), 1.44 (bs, 20H), 1.40-1.31 (m, 6H), 1.28 (bs, 6H), 0.90-0.87 (m, 6H); $^{13}C\{^1H\}$ NMR (125 MHz, $CDCl_3$) δ: 174.5, 164.3, 163.2, 155.9, 152.5, 85.0, 78.8, 44.9, 43.1, 40.2, 39.1, 30.7, 30.0, 29.7, 28.7, 28.2, 26.2, 26.1, 23.9, 22.9, 13.9, 10.6; HRMS (ESI) calculated $[M+H]^+$ for $C_{36}H_{67}N_8O_7$: 723.5127, found 723.5108 $[M+H]^+$.



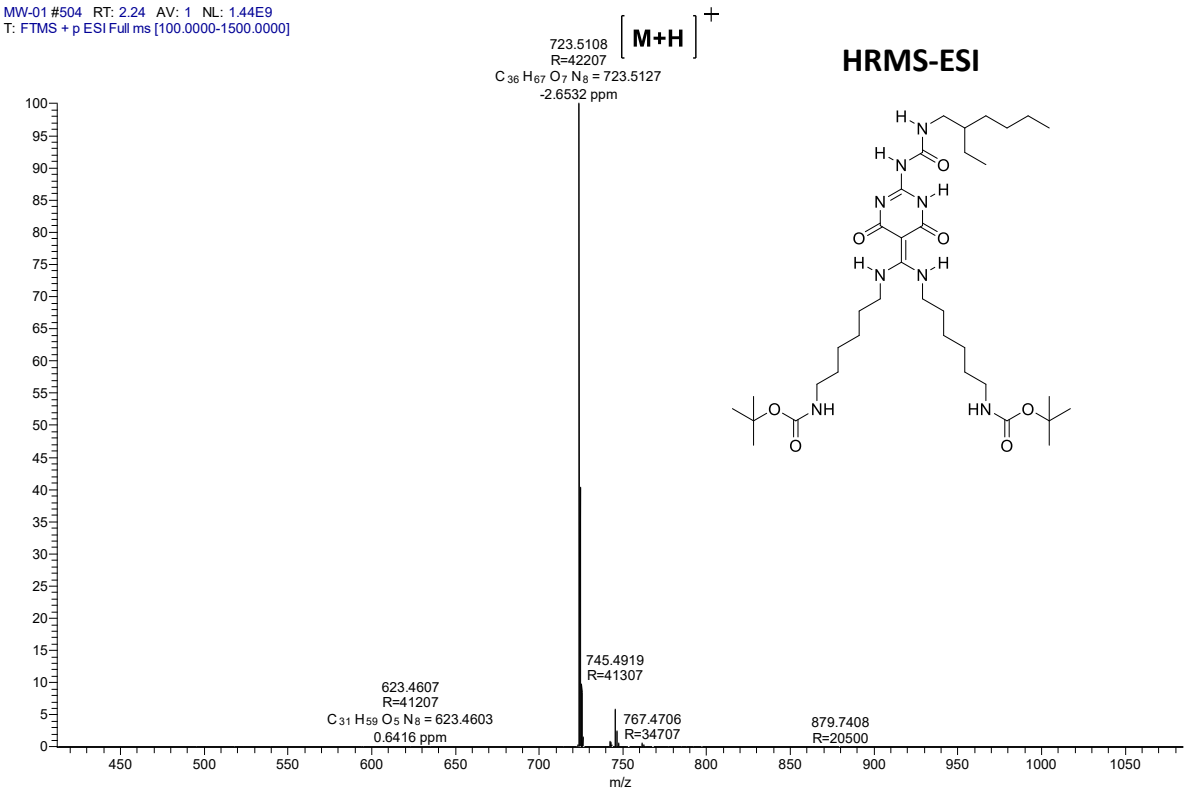
^1H NMR spectrum of compound **8a** ($CDCl_3$, 500 MHz, 298 K)



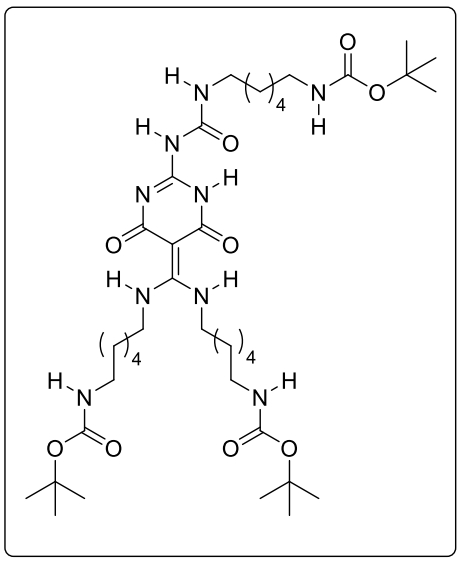
$^{13}C\{^{1}H\}$ NMR spectrum of compound **8a** ($CDCl_3$, 125 MHz, 298 K)



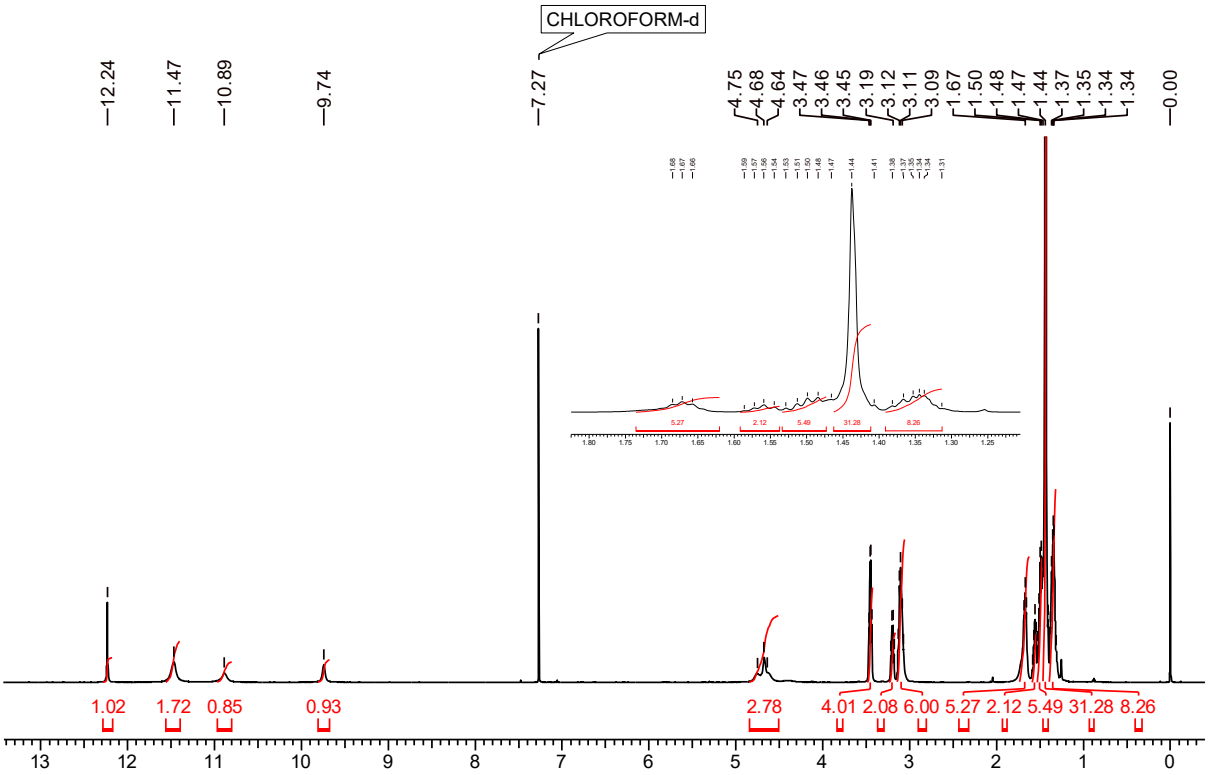
DEPT-135 NMR spectrum of compound **8a** ($CDCl_3$, 125 MHz, 298 K)



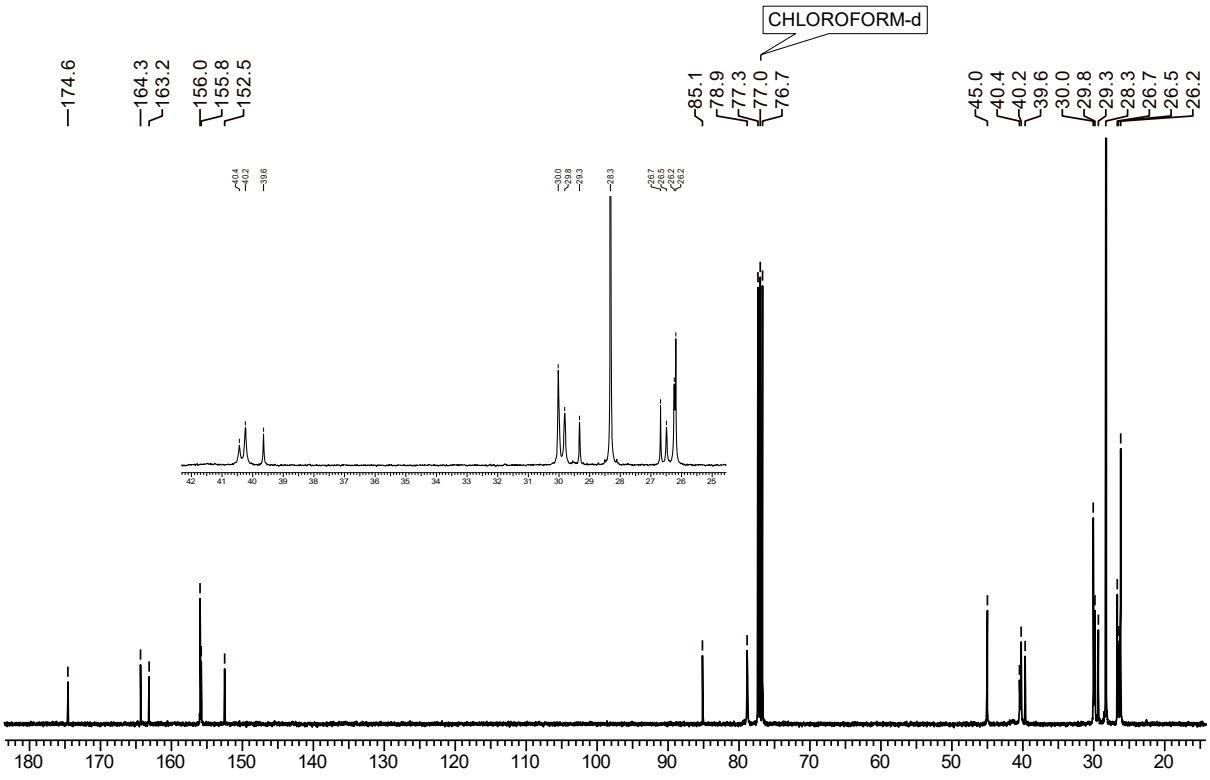
Compound 8b



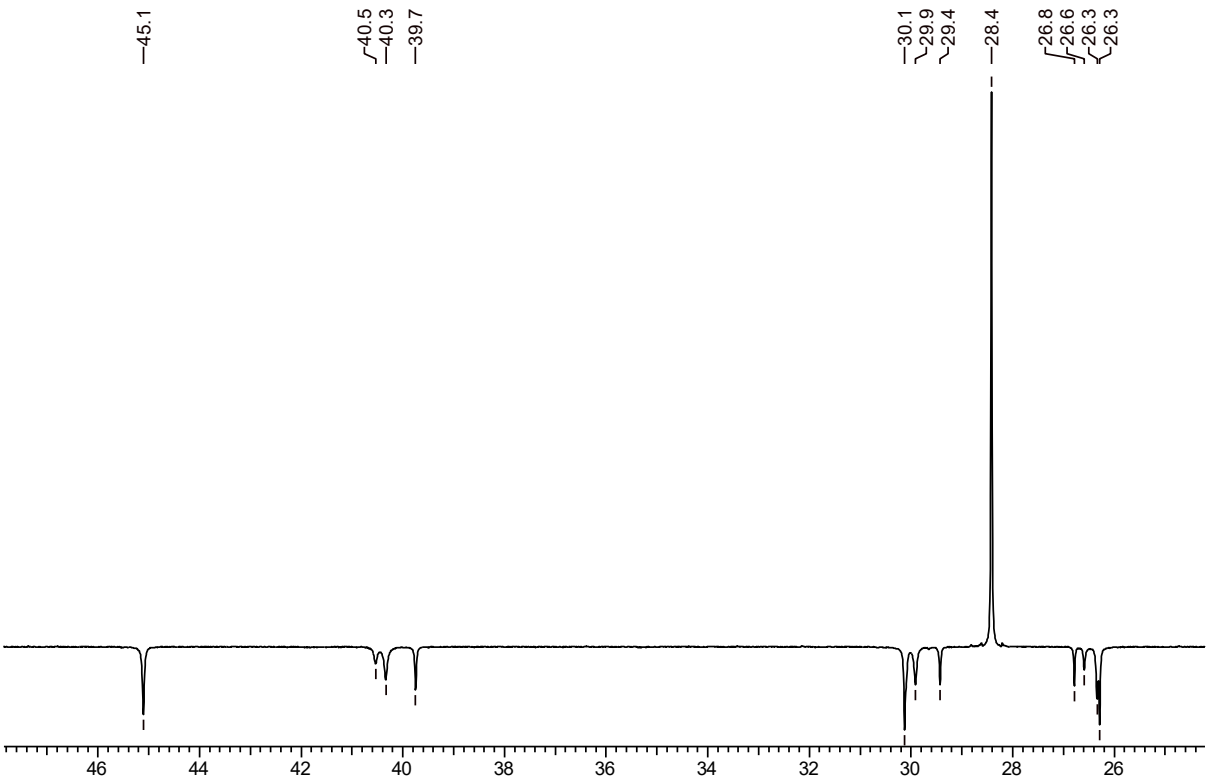
To a solution of **7** (0.50 g, 0.755 mmol, 1 equiv.) in dry $CHCl_3$, *tert*-butyl (6-aminohexyl) carbamate (0.21, 0.982 mmol,1.3 equiv.) was added and the reaction mixture stirred at room temperature until the starting material completely disappeared. After completion of the reaction, the mixture was evaporated on a rotary evaporator. The resultant reaction mixture was partitioned between DCM and water. The organic layer was separated, washed with water, aq. citric acid, and brine solution, dried over Na_2SO_4, filtered and concentrated *in vacuo*. Purification by column chromatography (eluent: 2% MeOH/DCM, R_f: 0.5) afforded **8b** (0.51 g, 84%) as a white solid: mp: 115-120 °C;^{1}H NMR (500 MHz, $CDCl_3$) δ: 12.24 (s, 1H), 11.47 (bs, 2H), 10.89 (bs, 1H), 9.74 (bs, 1H), 4.75-4.62 (m, 3H), 3.47-3.43 (m, 4H), 3.22-3.18 (m, 2H), 3.13-3.07 (m, 6H), 1.68-1.66 (m, 5H), 1.59-1.54 (m, 2H), 1.53-1.48 (m, 5H), 1.44 (bs, 31H), 1.38-1.34 (m, 8H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$) δ: 174.6, 164.3, 163.2, 156.0, 155.8, 152.5, 85.1, 78.9, 45.0, 40.4, 40.2, 39.6, 30.0, 29.8, 29.3, 28.3, 26.7, 26.5, 26.2, 26.2; HRMS (ESI) calculated $[M+H]^+$ for $C_{39}H_{72}N_9O_9$: 810.5448, found 810.5440 $[M+H]^+$.



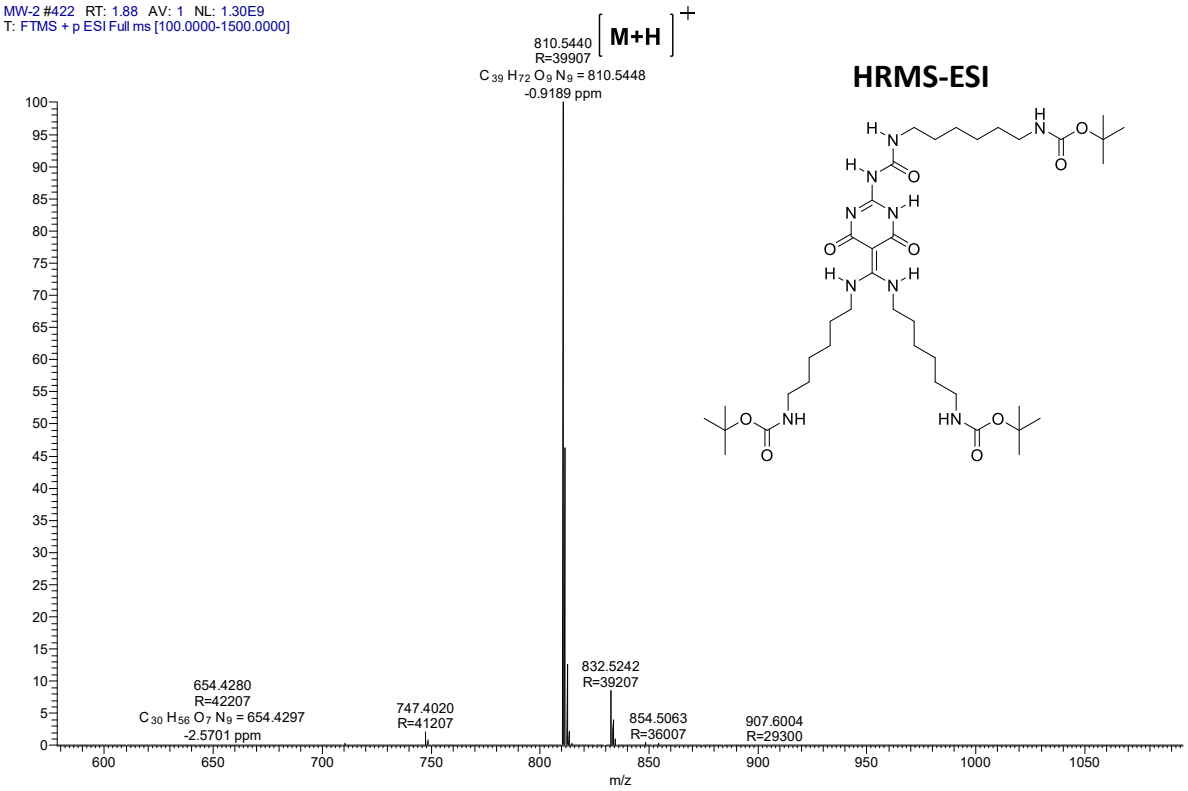
^{1}H NMR spectrum of compound **8b** ($CDCl_3$, 500 MHz, 298 K)



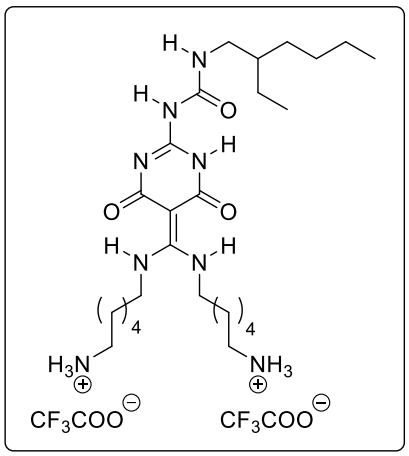
^{13}C{^{1}H} NMR spectrum of compound **8b** ($CDCl_3$, 125 MHz, 298 K)



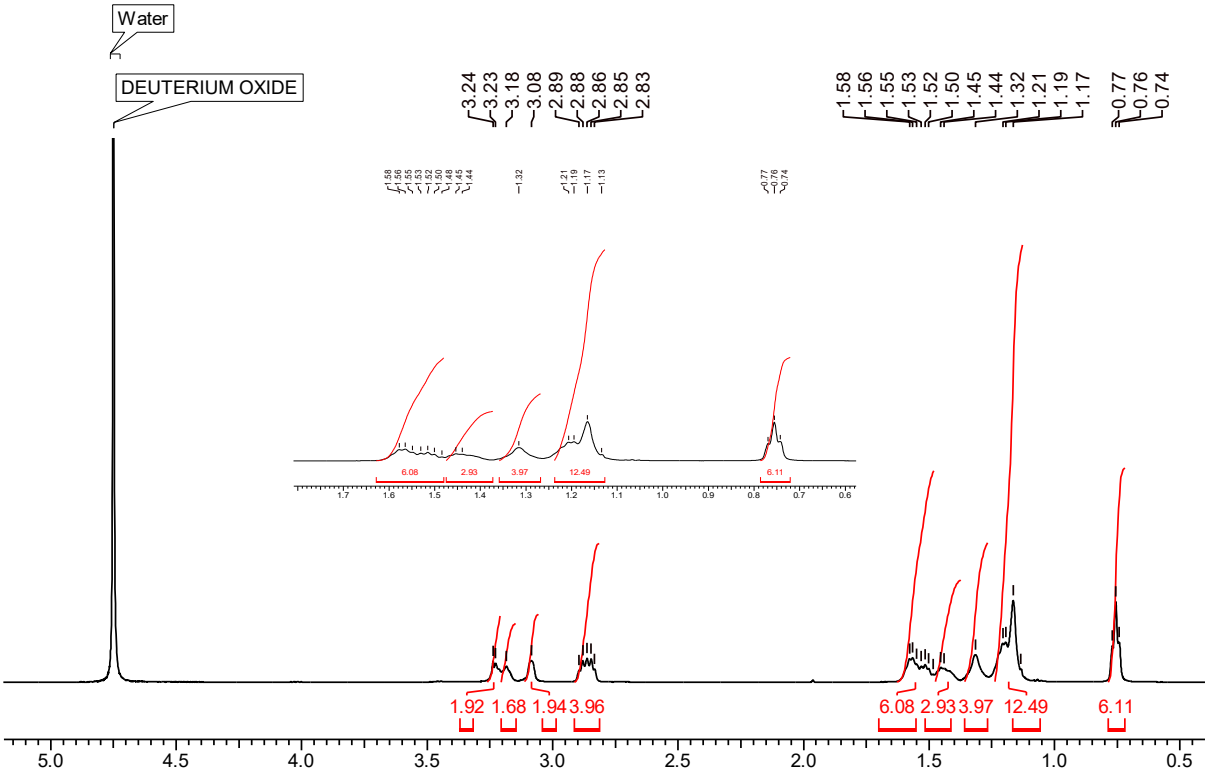
DEPT-135 NMR spectrum of compound **8b** ($CDCl_3$, 125 MHz, 298 K)



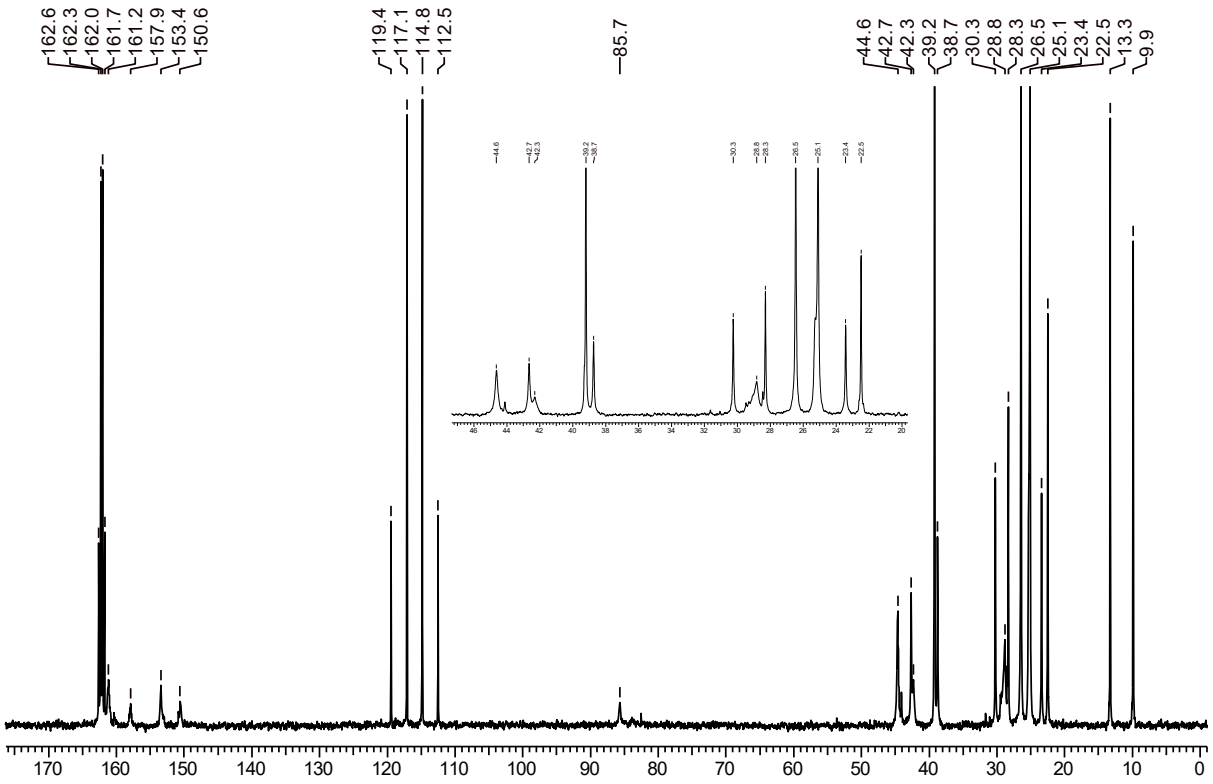
Compound 9a



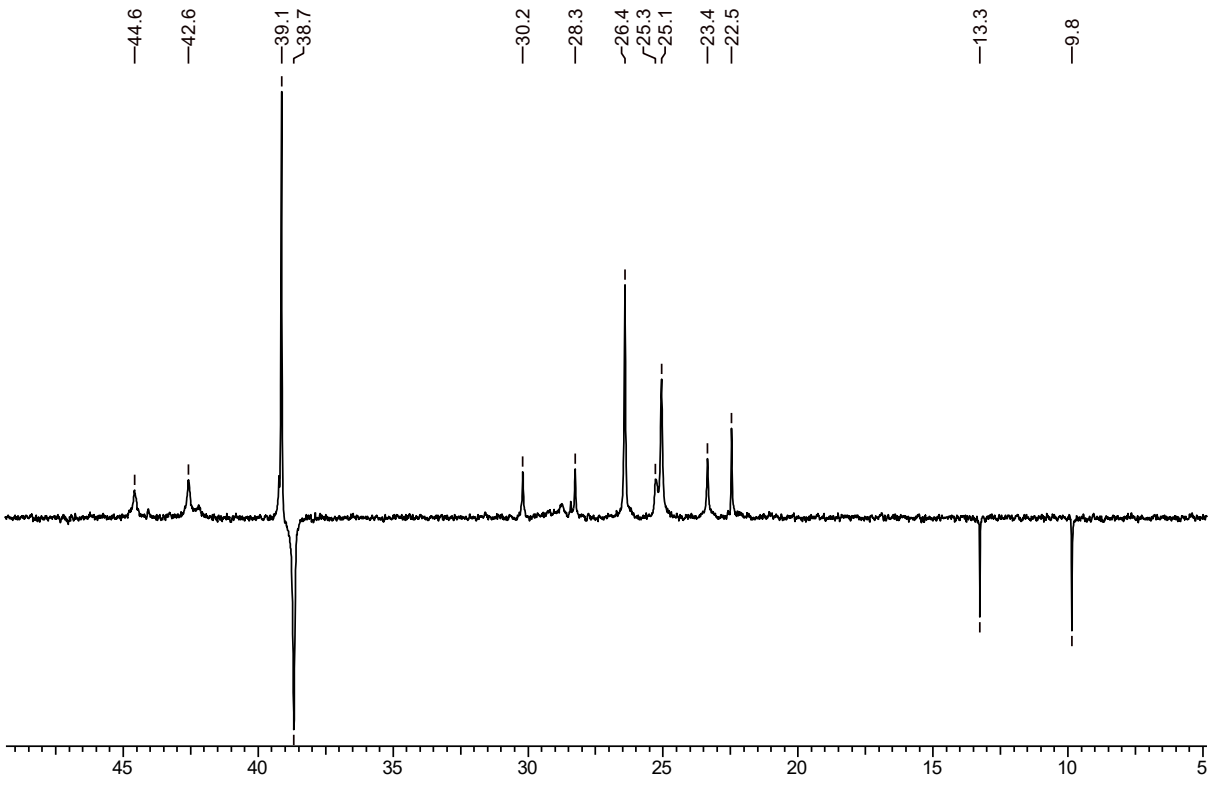
The Boc deprotection of **8a** was carried out using the same procedure employed for **4g** to afford **9a** in quantitative yield, as white solid. mp: 107-112 °C; ^{1}H NMR (500 MHz, D_2O) δ: 3.24-3.23 (m, 2H), 3.18 (bs, 2H), 3.08 (bs, 2H), 2.89-2.83 (m, 4H), 1.58-1.50 (m, 6H), 1.45-1.42 (m, 3H), 1.32 (bs, 4H), 1.21-1.17 (m, 12H), 0.77-0.74 (m, 6H); ^{13}C{^{1}H} NMR (125 MHz, $CDCl_3$) δ: 163.6-162.7 (residual TFA signal), 161.1, 157.9, 153.4, 150.6, 119.4-112.5 (residual TFA signal), 85.7, 44.6, 42.7, 42.3, 39.2, 38.7, 30.3, 28.8, 28.3, 26.5, 25.3, 25.1, 23.4, 22.5, 13.3, 9.9; HRMS (ESI) calculated $[M+H]^+$ for $C_{26}H_{51}N_8O_3$: 523.4079, found 523.4074 $[M+H]^+$.



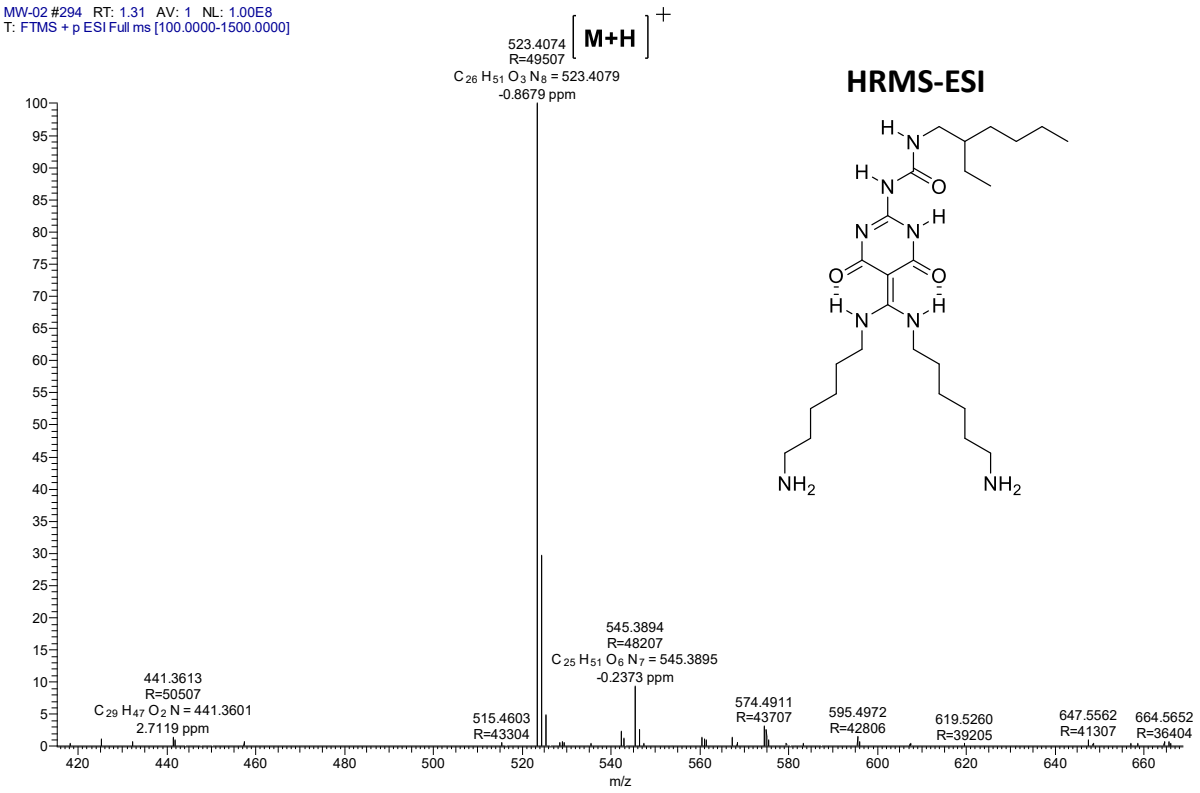
^{1}H NMR spectrum of compound **9a** (D_2O, 500 MHz, 298 K)



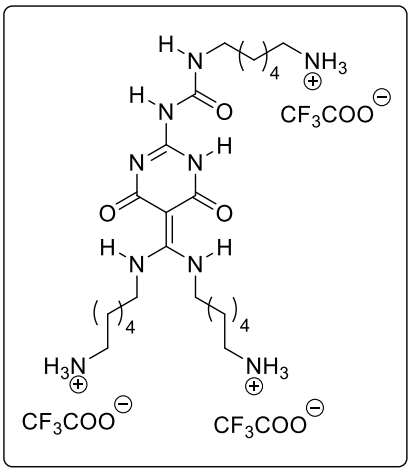
$^{13}C\{^{1}H\}$ NMR spectrum of compound **9a** (D_2O, 125 MHz, 298 K)



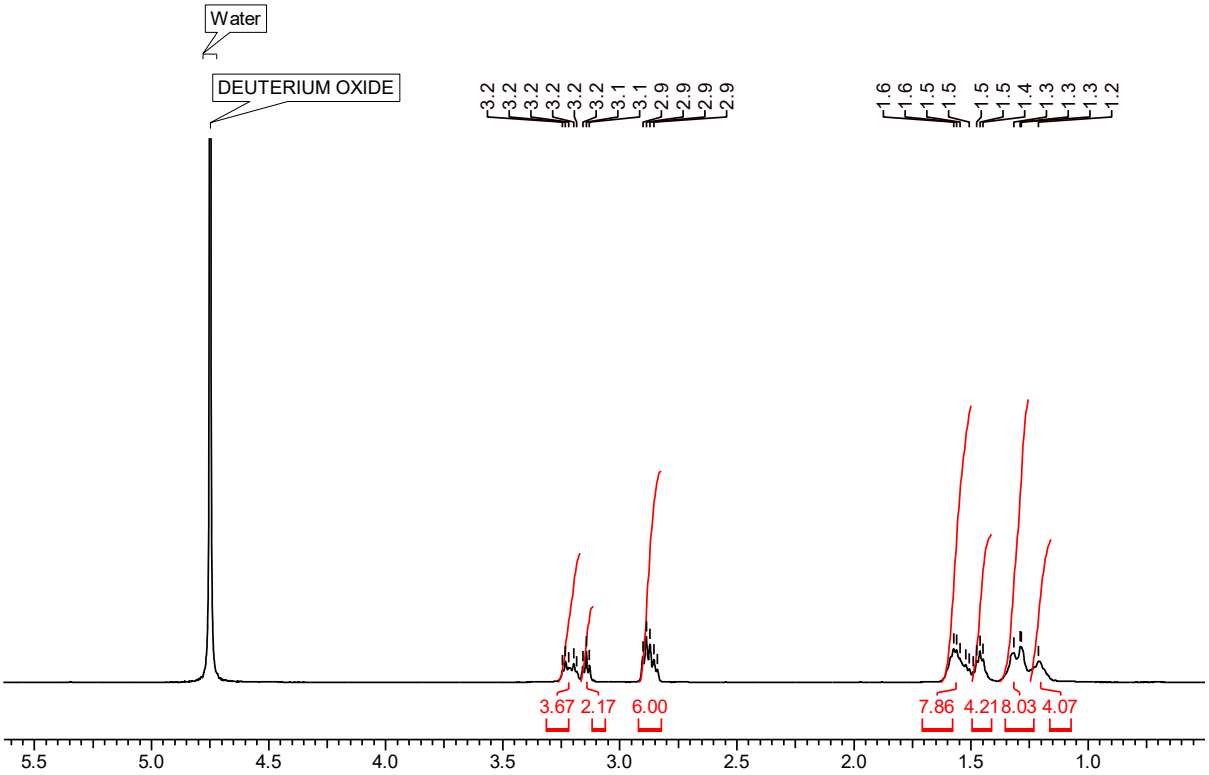
DEPT-135 NMR spectrum of compound **9a** (D_2O, 125 MHz, 298 K)



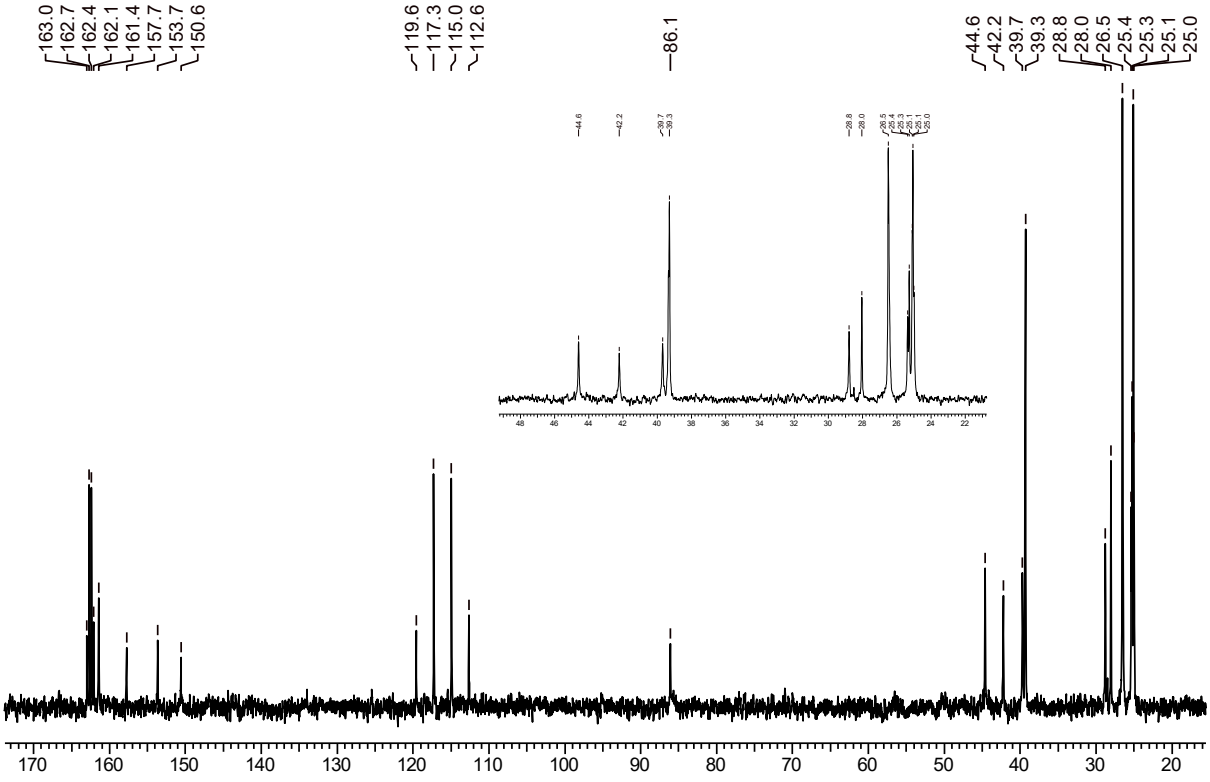
Compound 9b



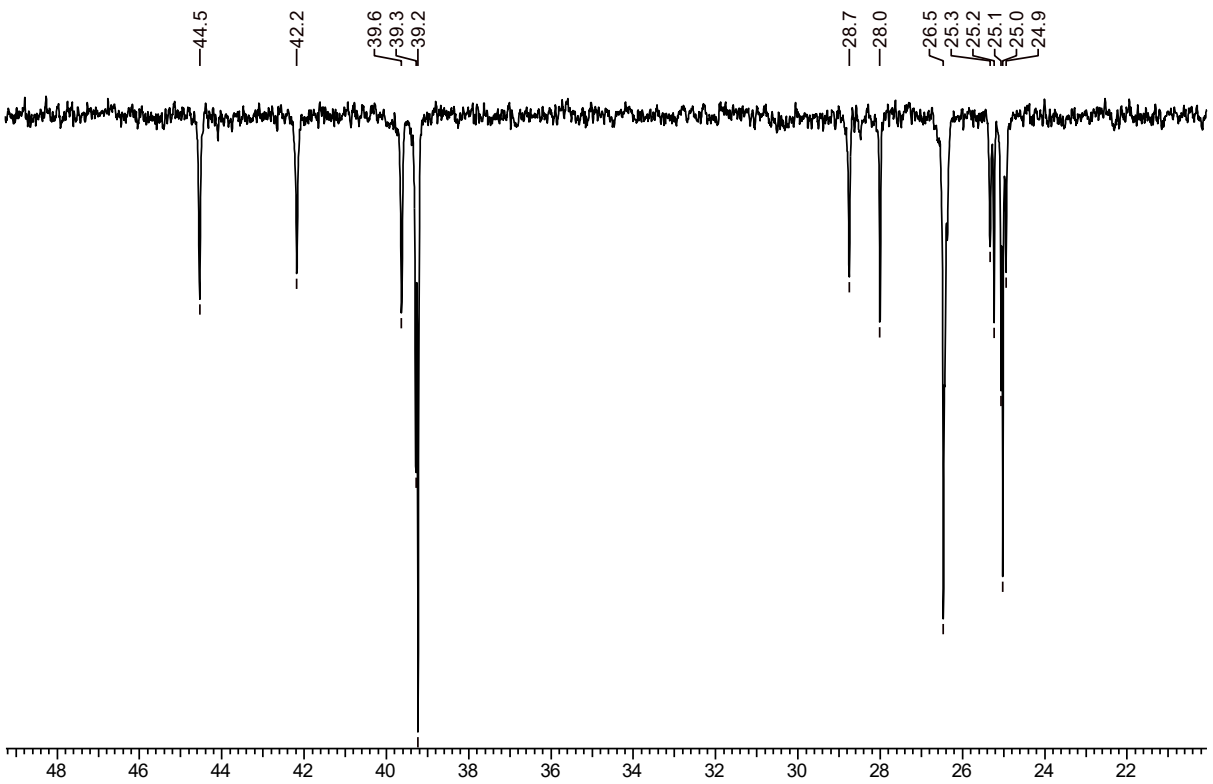
The Boc deprotection of **8b** was carried out using the same procedure employed for **4g** to afford **9b** in quantitative yield, as a white sticky solid. ^{1}H NMR (500 MHz, D_2O δ: 3.24-3.18 (m, 4H), 3.16-3.13 (m, 2H), 2.90-2.84 (m, 6H), 1.57-1.51 (m, 8H), 1.49-1.45 (m, 4H), 1.32-1.28 (m, 8H), 1.25-1.21 (m, 4H); ^{13}C{^{1}H} NMR (125 MHz, D_2O) δ: 163.0-162.1 (residual TFA signal), 161.4, 157.7, 153.7, 150.6, 119.6-112.6 (residual TFA signal), 86.1, 44.6, 42.2, 39.7, 39.3, 28.8, 28.0, 26.5, 25.4, 25.3, 25.1, 25.0; HRMS (ESI) calculated $[M+H]^+$ for $C_{24}H_{48}N_9O_3$: 510.3875, found 510.3872 $[M+H]^+$.



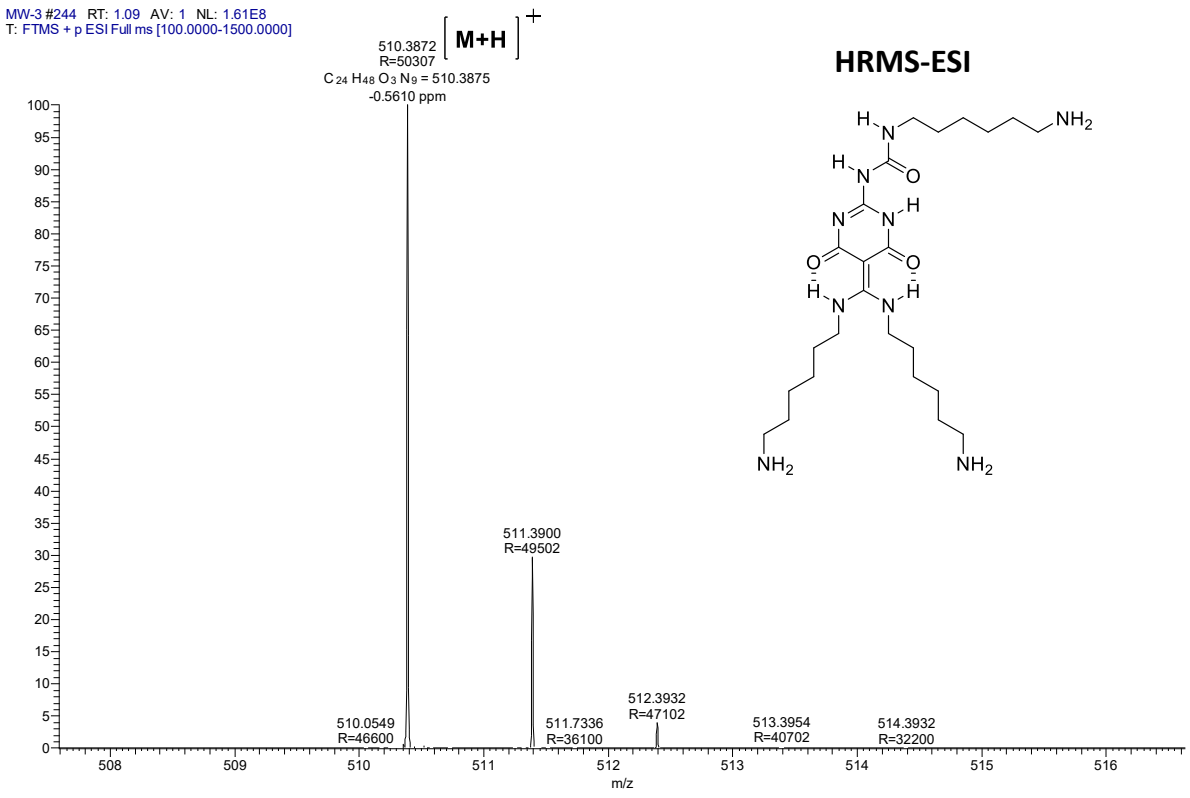
^{1}H NMR spectrum of compound **9b** (D_2O, 500 MHz, 298 K)



^{13}C{^{1}H} NMR spectrum of compound **9b** (D_2O, 125 MHz, 298 K)



DEPT-135 NMR spectrum of compound **9b** (D_2O, 125 MHz, 298 K)



NMR dilution studies

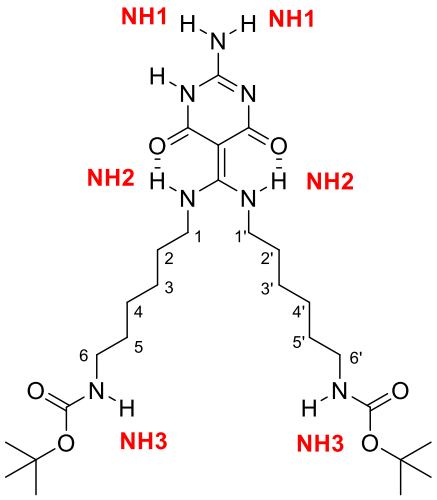


Fig. S1 Chemical structure of compound **4f**. *Note:* The NH1, NH2 and NH3 were assigned by using 2D NMR experiments (see below).

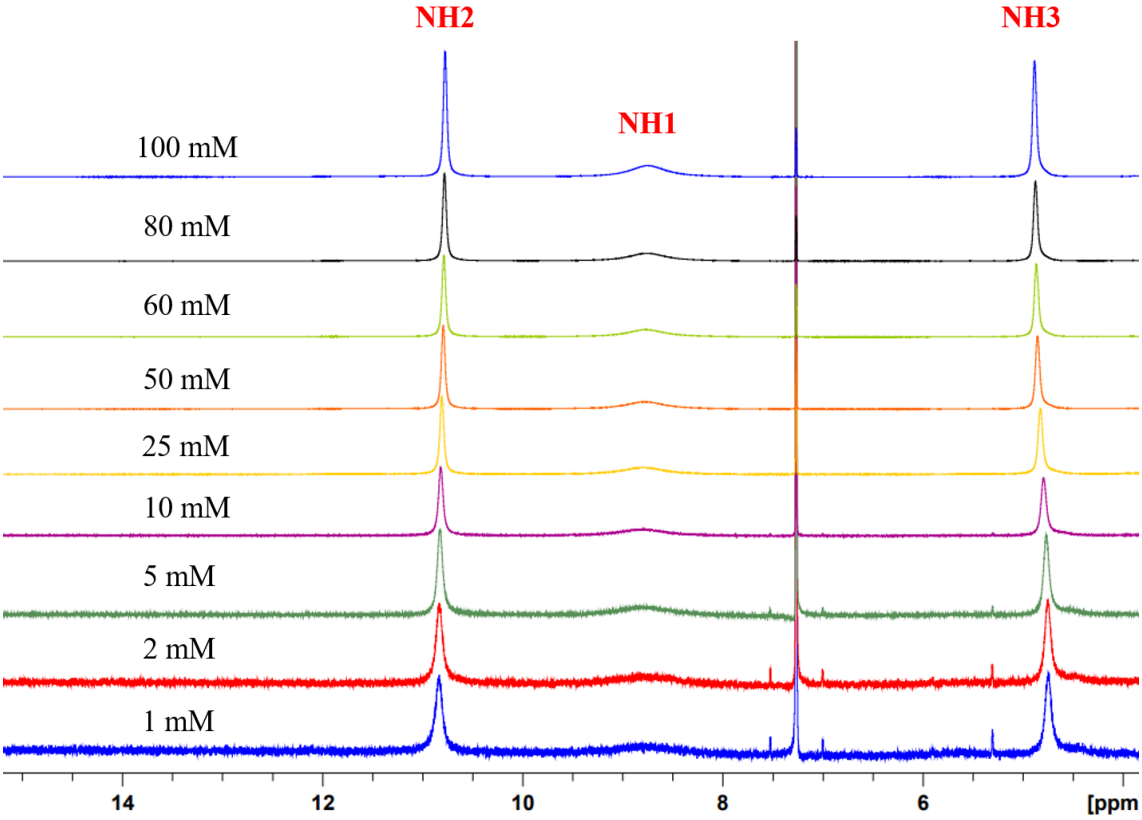


Fig. S2 Stacked partial ^{1}H NMR spectra of compound **4f** at different concentrations in $CDCl_3$ (400 MHz, 298 K).

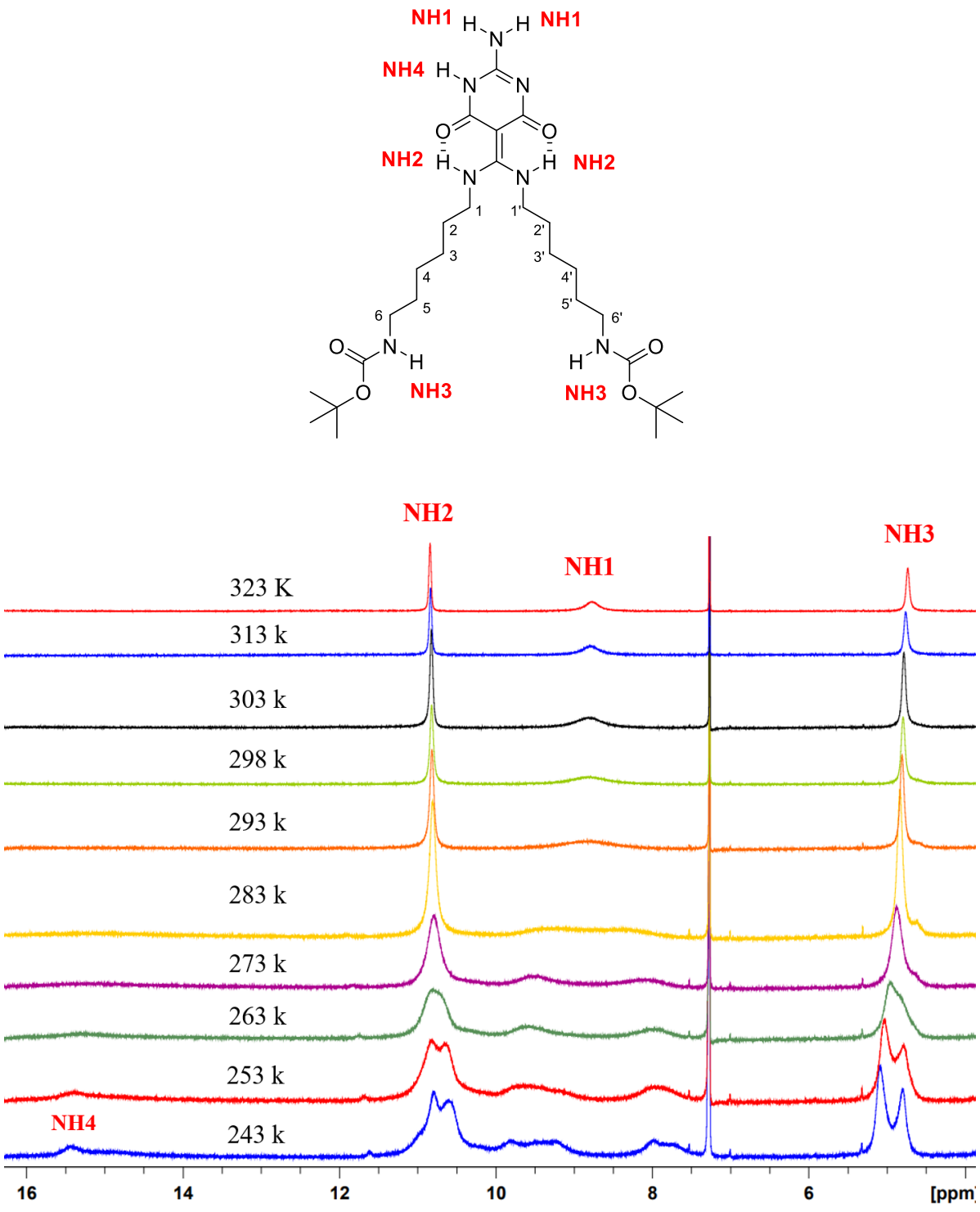


Fig. S3 Stacked partial (NH region) ^{1}H NMR spectra (400 MHz) of compound **4f** at different temperatures (10 mM in $CDCl_3$).

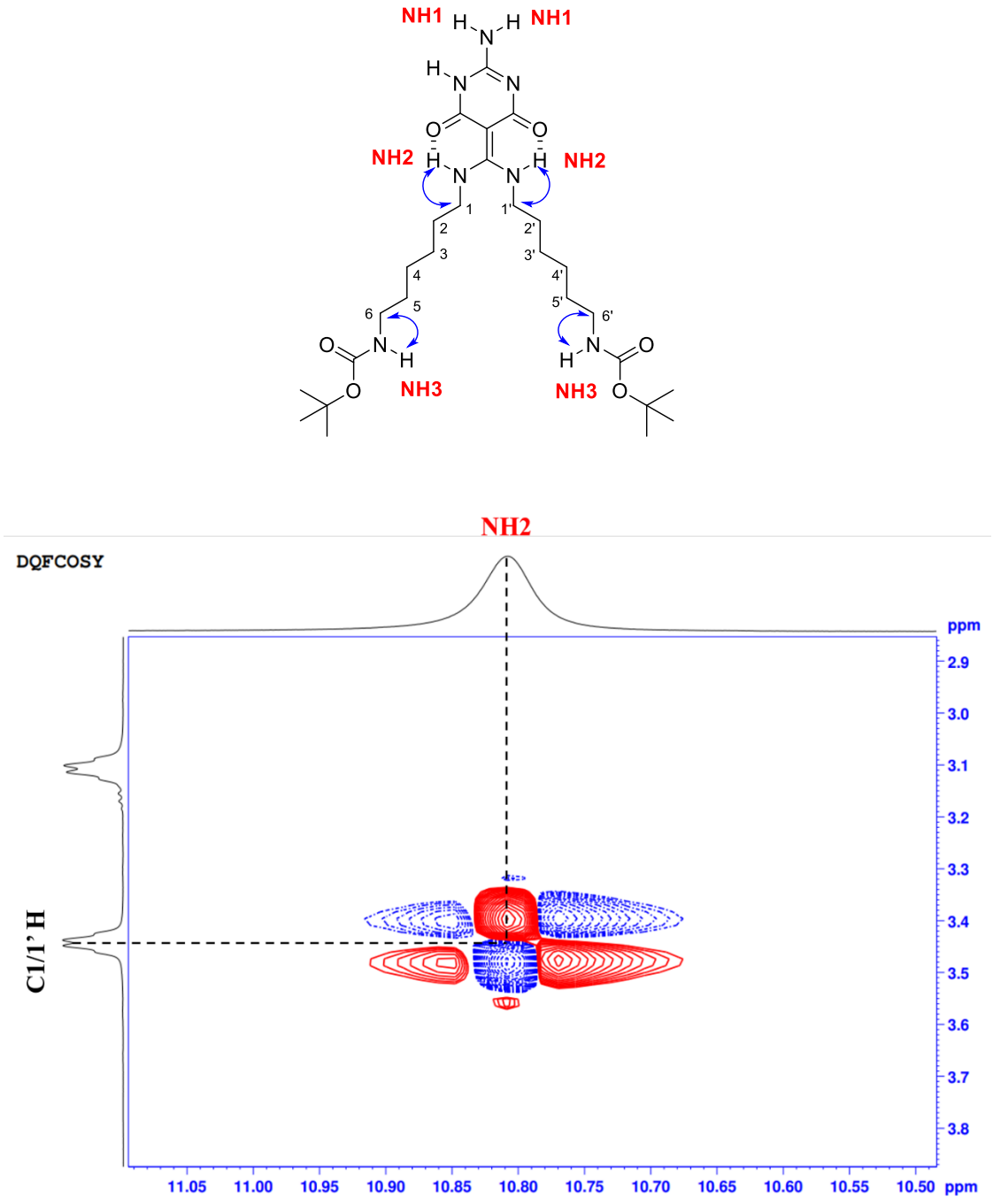


Fig. S4 2D COSY extracts of **4f** ($CDCl_3$, 500 MHz).

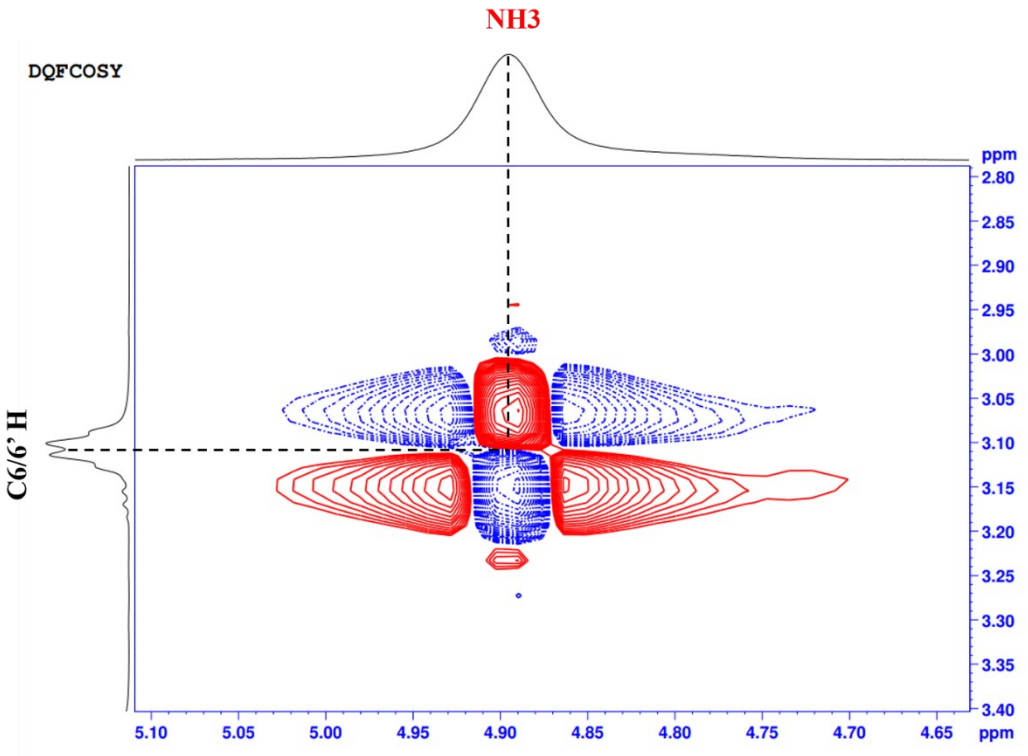


Fig. S5 2D COSY extracts of **4f** ($CDCl_3$, 500 MHz)

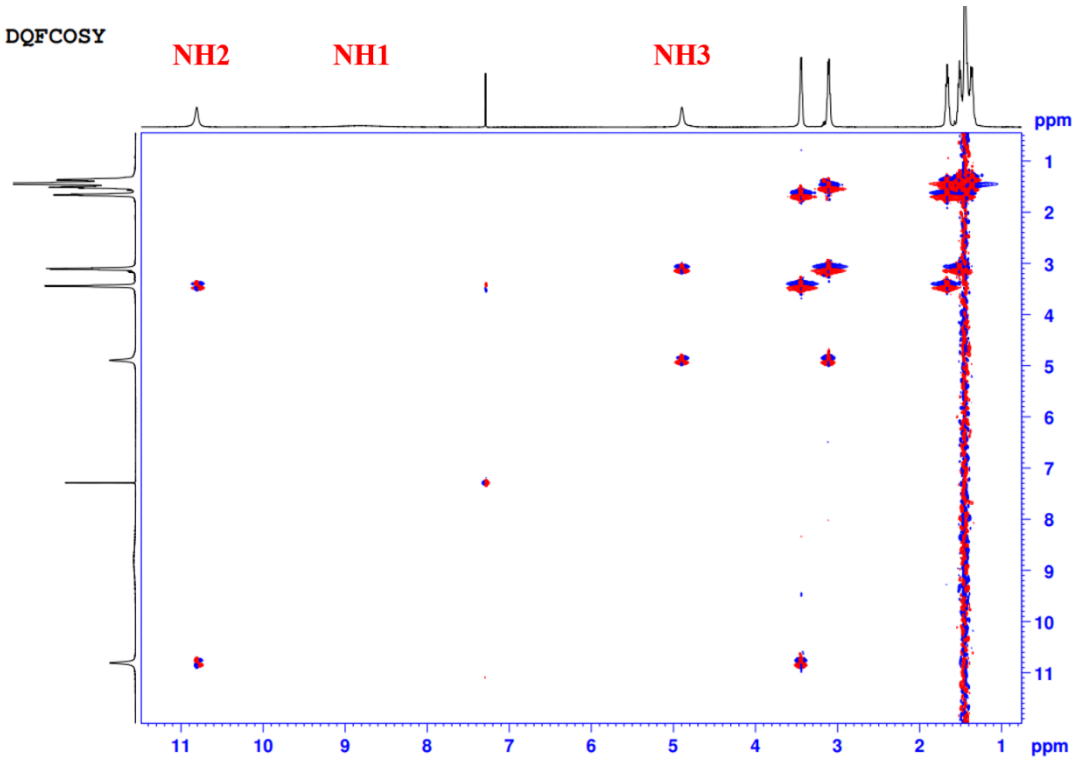


Fig. S6 2D COSY spectrum of **7f** ($CDCl_3$, 500 MHz)

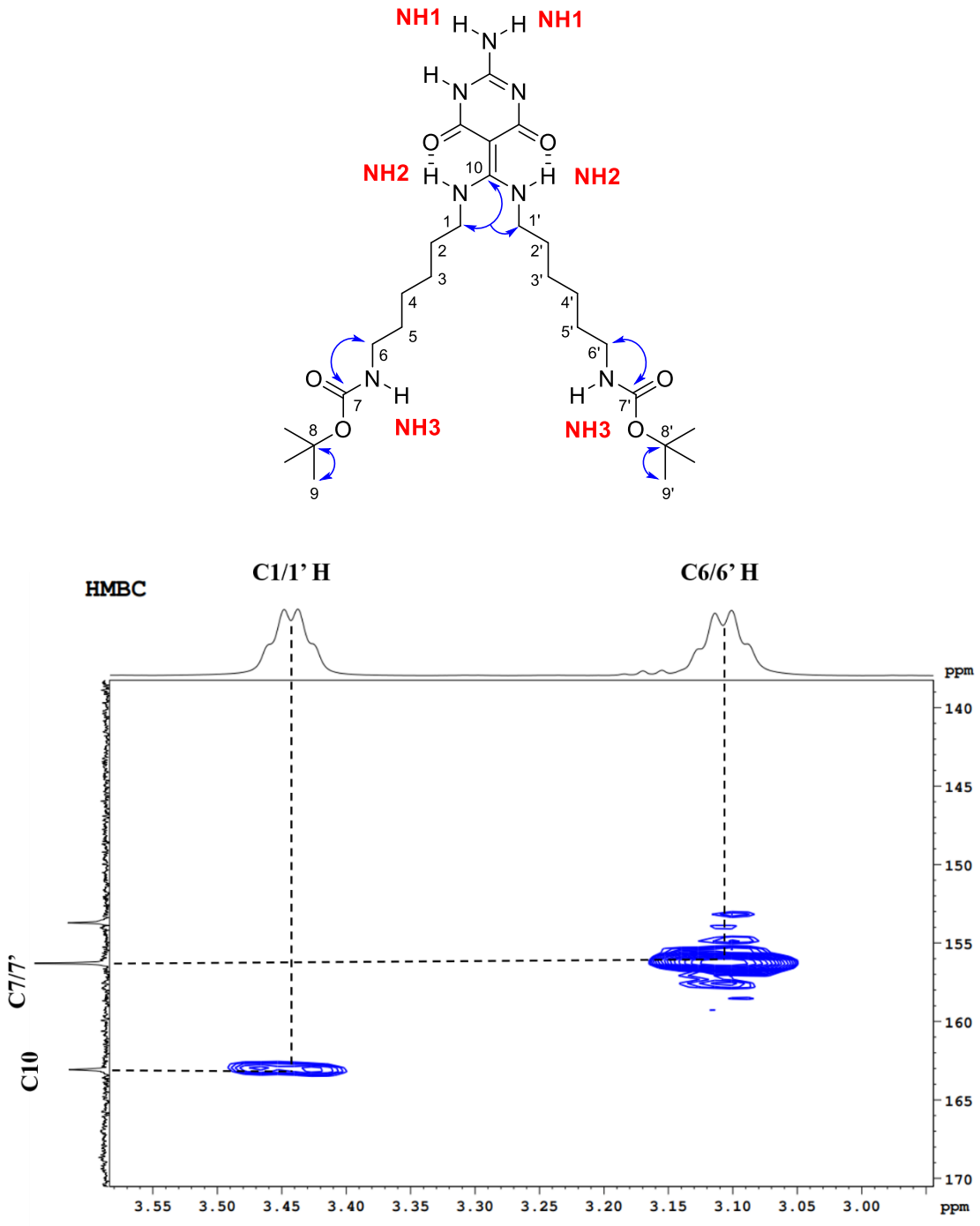


Fig. S7 2D HMBC extracts of **4f** ($CDCl_3$, 500 MHz)

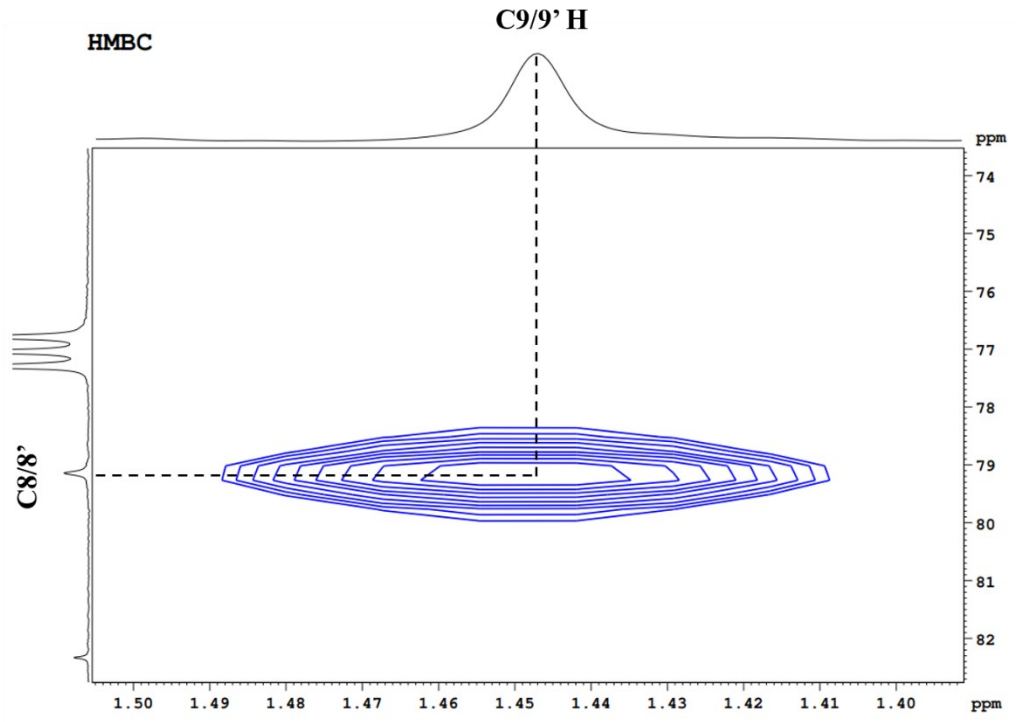


Fig. S8 2D HMBC extracts of **4f** ($CDCl_3$, 500 MHz)

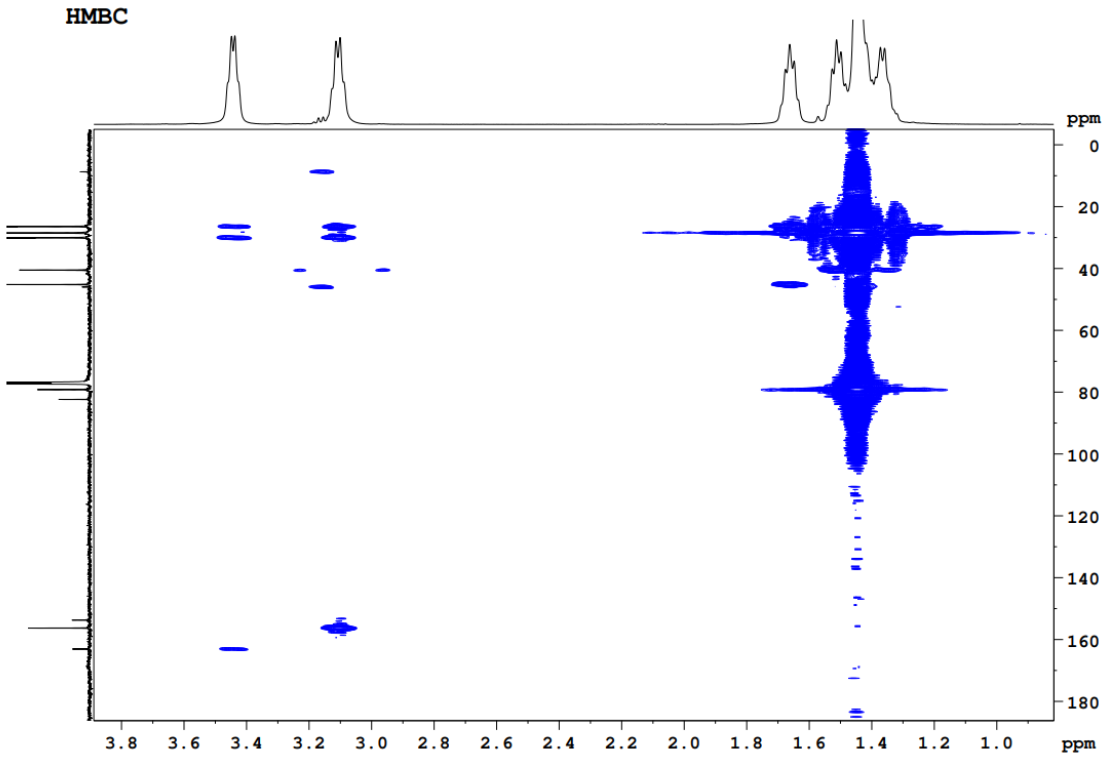


Fig. S9 2D HMBC spectrum of **4f** ($CDCl_3$, 500 MHz)

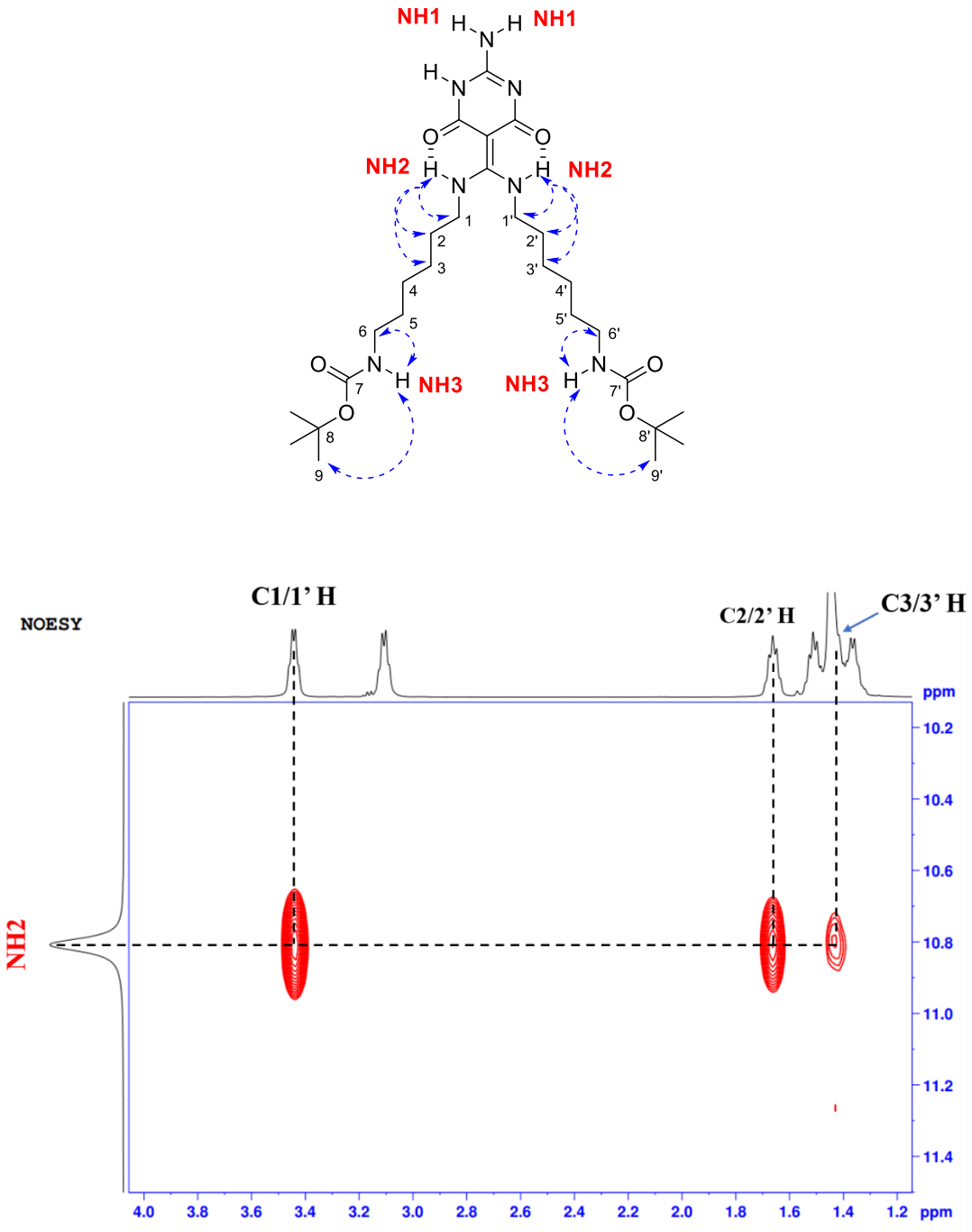


Fig. S10 2D NOESY extracts of **4f** ($CDCl_3$, 500 MHz)

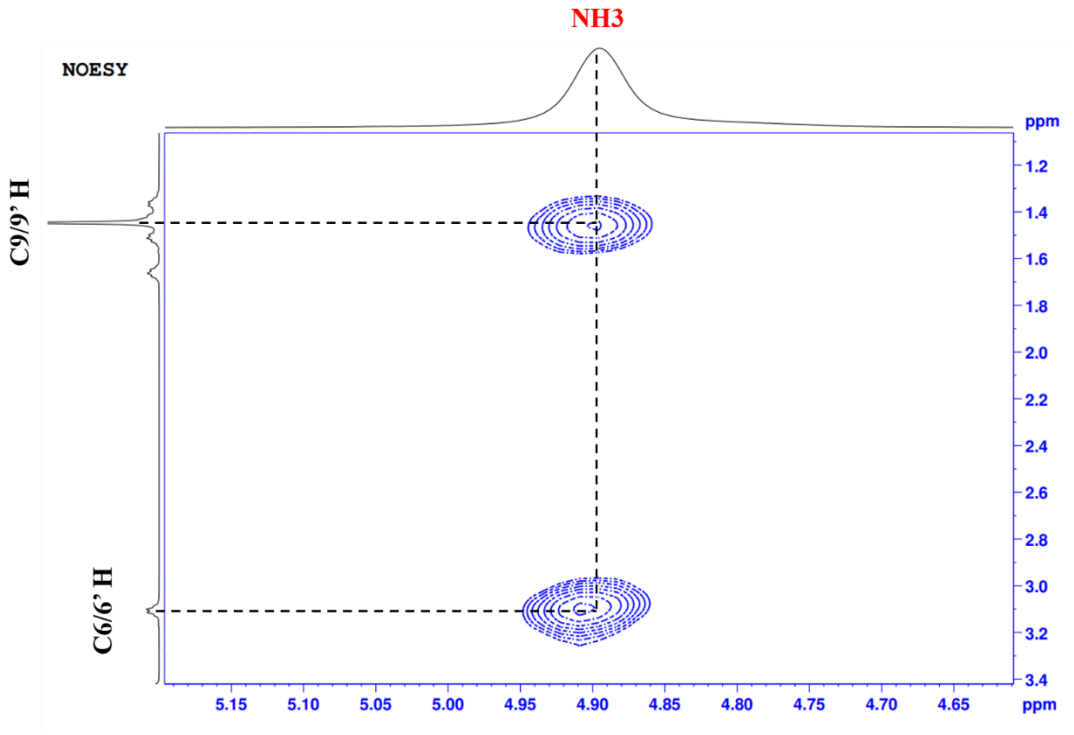


Fig. S11 2D NOESY extracts of **4f** ($CDCl_3$, 500 MHz)

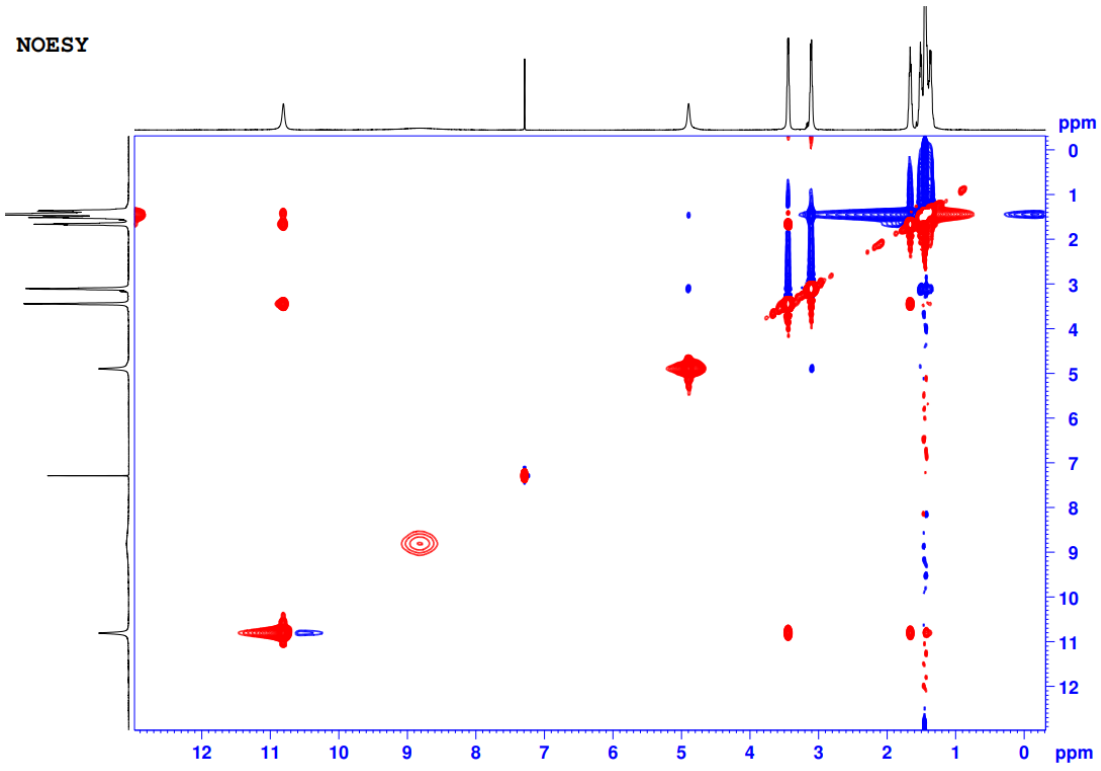


Fig. S12 2D NOESY spectrum of **4f** ($CDCl_3$, 500 MHz)

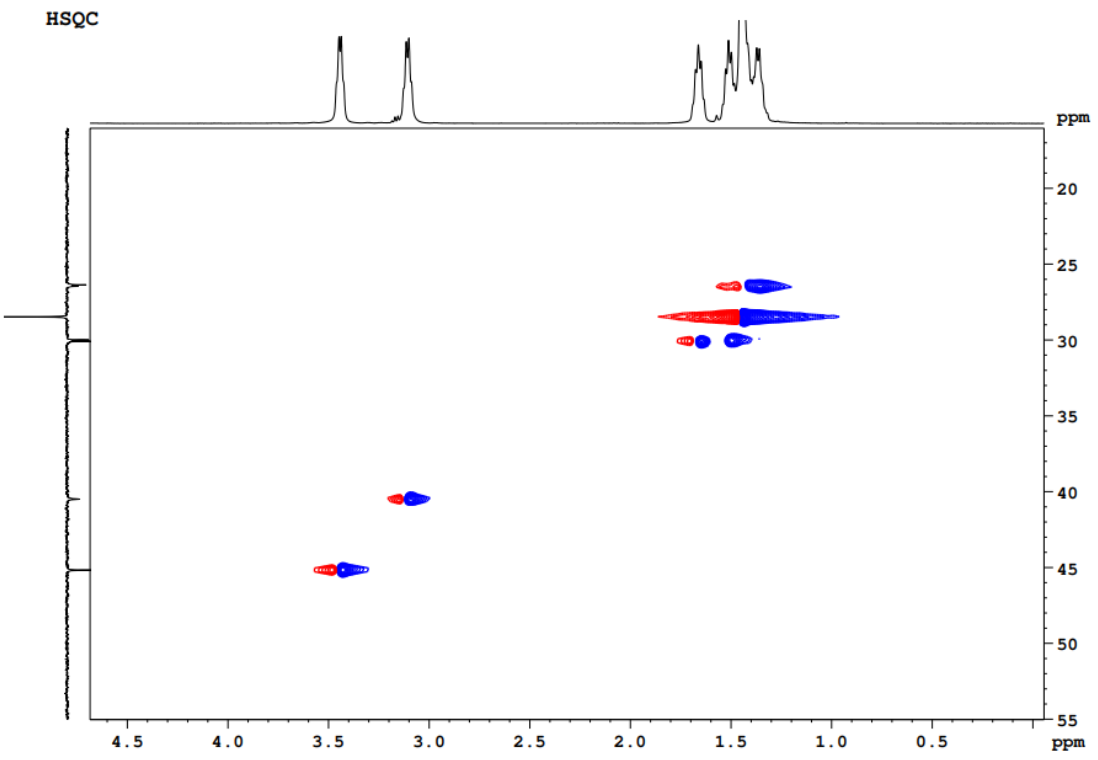


Fig. S13 2D HSQC spectrum of **4f** ($CDCl_3$, 500 MHz)

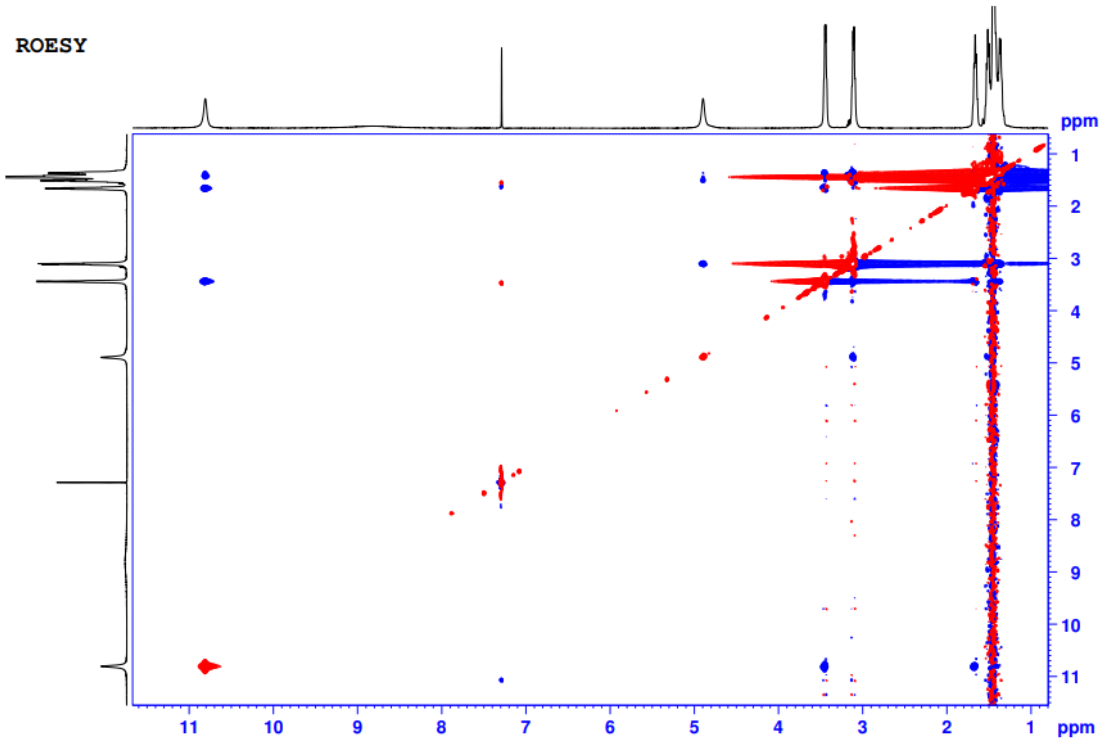


Fig. S14 2D ROESY spectrum of **4f** ($CDCl_3$, 500 MHz)

^{1}H NMR (DOSY) experiment

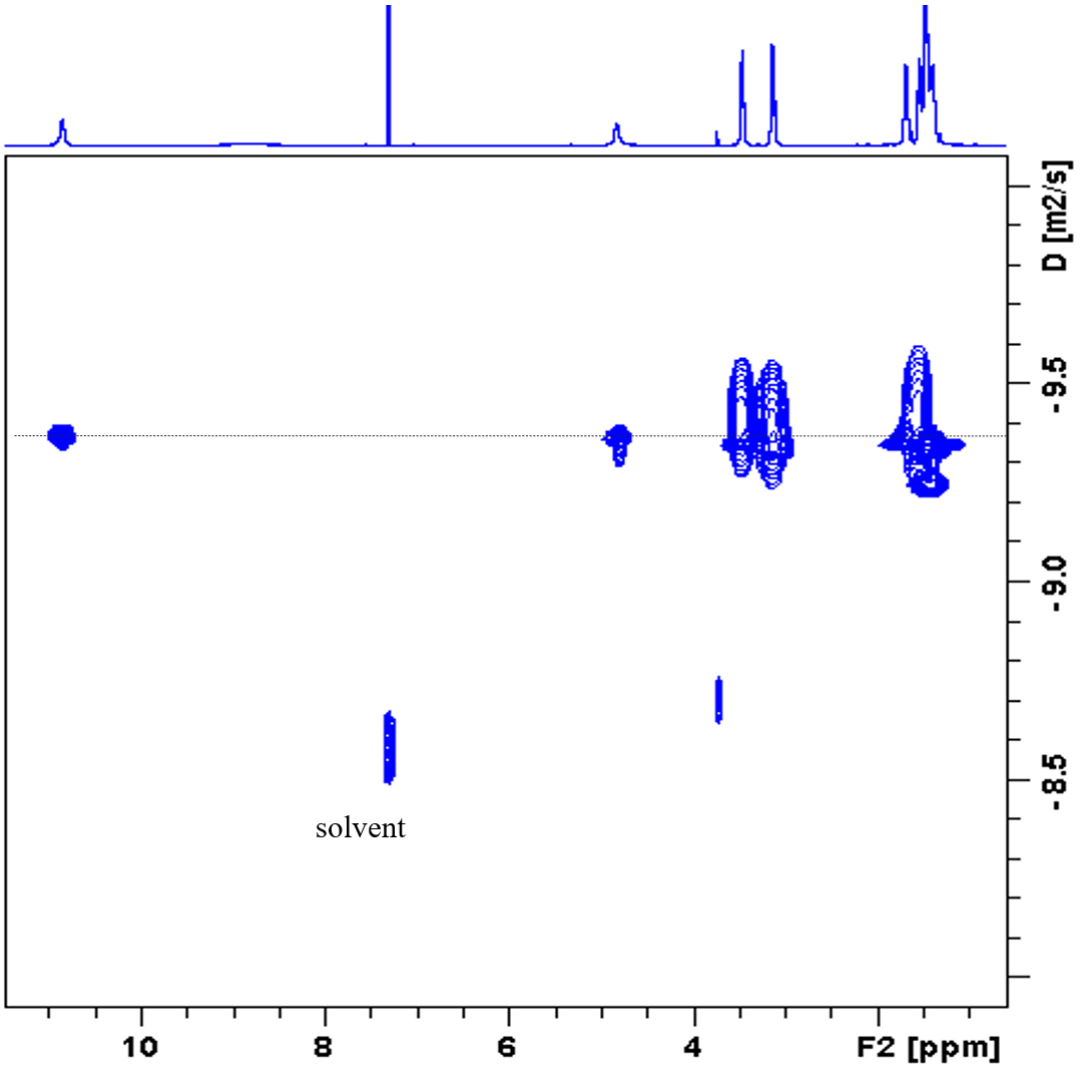


Fig. S15 DOSY spectrum of **4f** ($CDCl_3$, 400 MHz)

The diffusion coefficient of compound **4g** is 4.2 e^{-10} m^2/s which is about six times higher than that of the solvent chloroform.

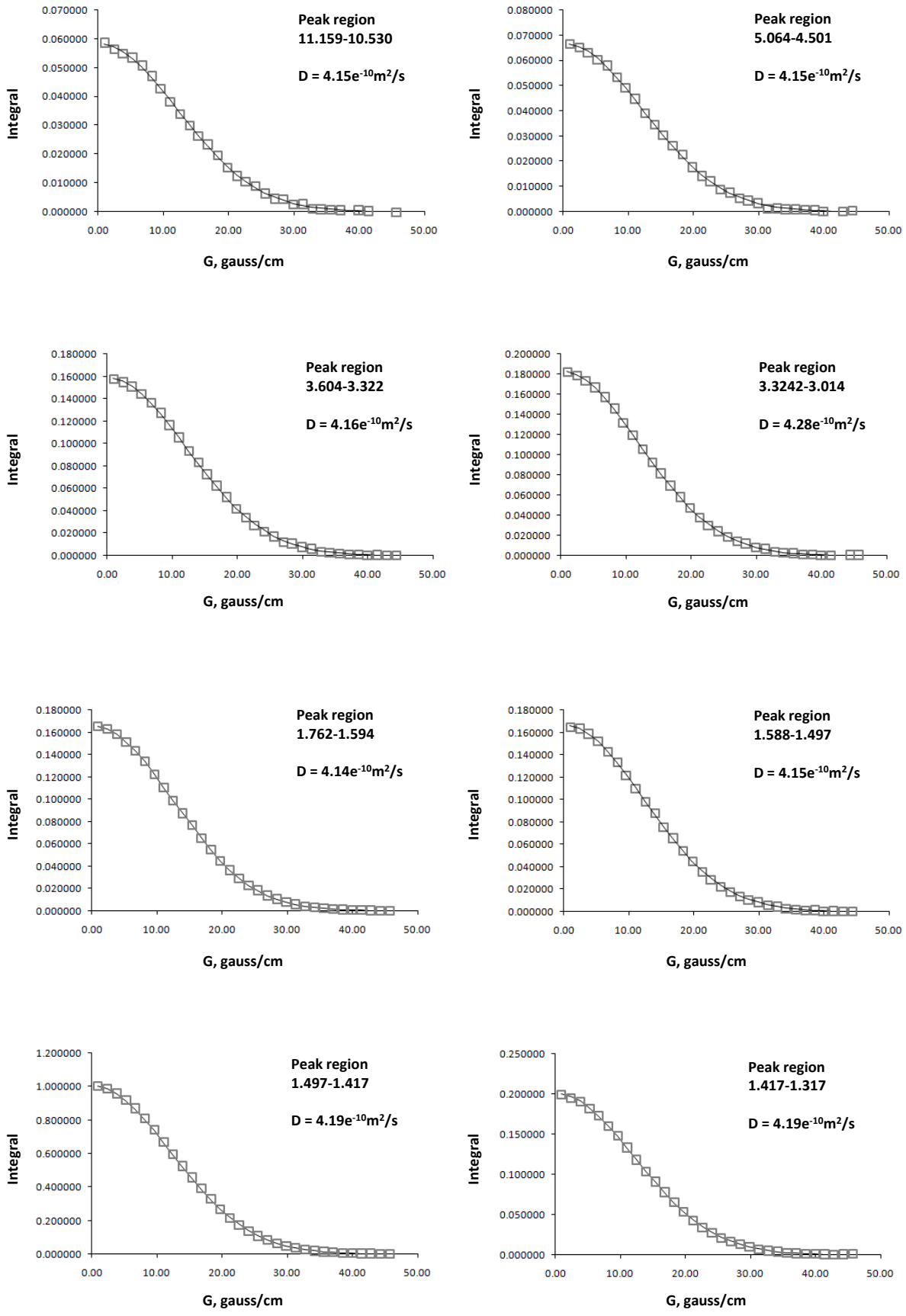


Fig. S16 Fits to peak integrals as a function of gradient strength measured in the DOSY experiment.

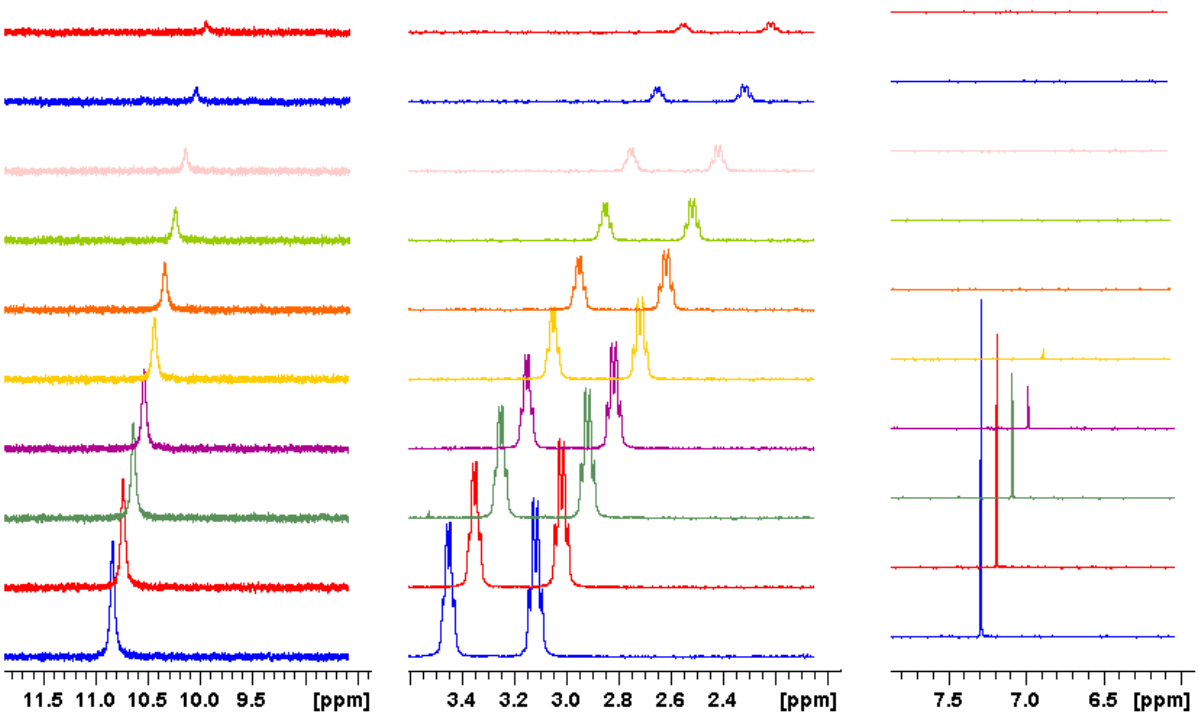


Fig. S17 Alternate traces from the DOSY spectrum indicating signal attenuation with increasing gradient strength in some representative spectral regions. Fast diffusion of the solvent is evident in the right panel.

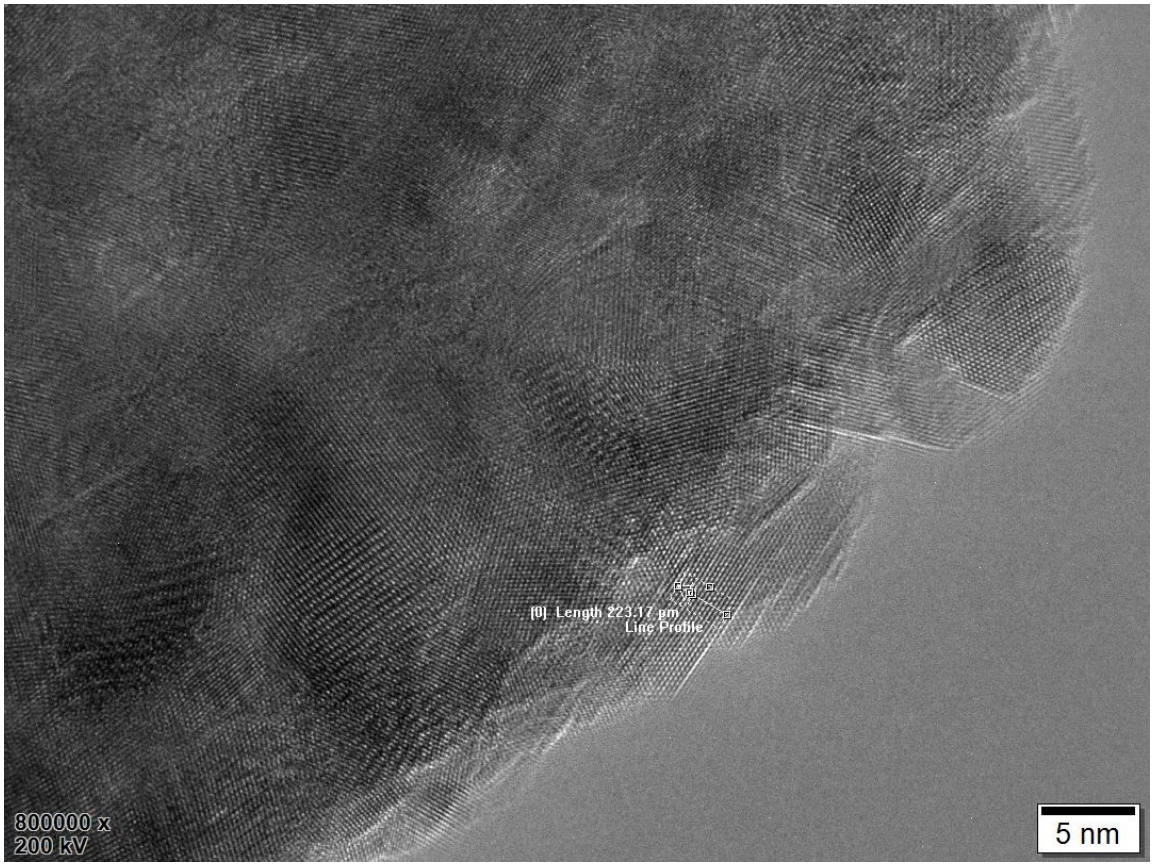


Fig. S18 HRTEM image of **4f**

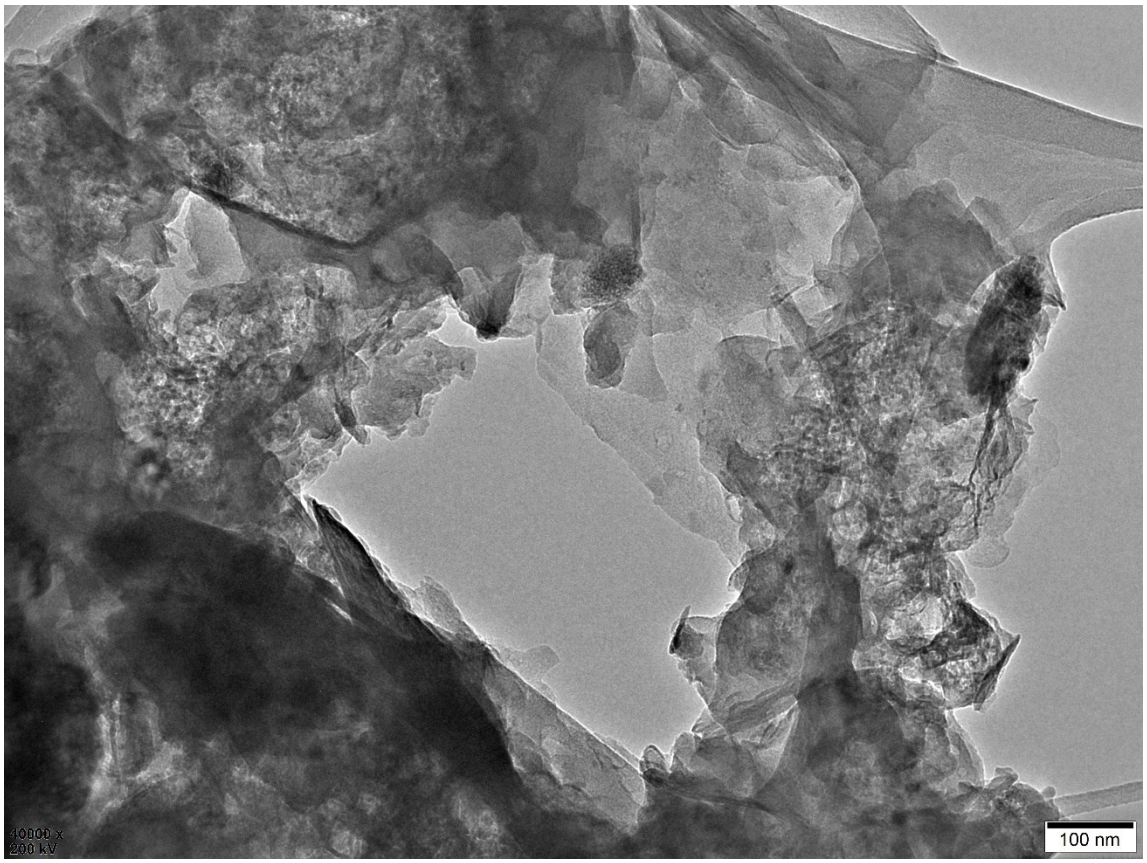


Fig. S19 HRTEM images of **4f**

Fig. S20 SAED images of **4f**

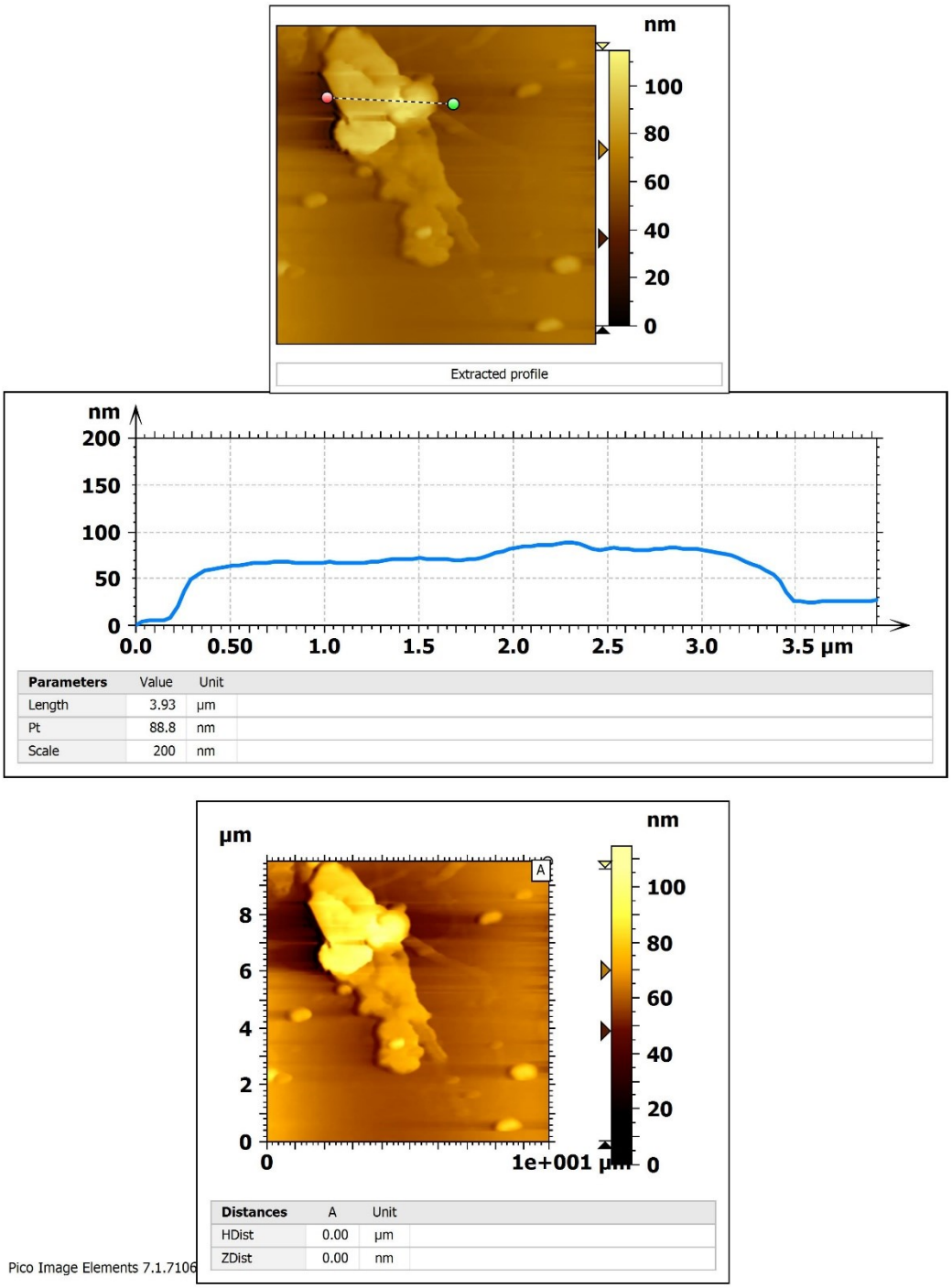


Fig. S21 AFM images of compound **4f**

The HRTEM (See fig. S18-S20) and AFM (See fig. S21) images revealed that compound **4f** is highly crystalline in nature as there is lattice fringes with d-spacing 0.2 nm and forming sheet-like morphology, as anticipated.

General X-ray Crystallography Analysis:

Single crystals which are suitable for single-crystal X-ray diffraction measurements were obtained from dichloromethane and methanol solvent mixture by slow evaporation method. The single crystal X-ray diffraction measurements were performed to determine the crystal structure of compound **4a** at 100 K using APEX3 (Bruker, 2016; Bruker D8 VENTURE Kappa Duo PHOTON II CPAD) diffractometer having graphite-monochromatized (MoKα = 0.71073). The X-ray generator was operated at 50 kV and 30 mA. A preliminary set of unit cell parameters and an orientation matrix were calculated from 36 frames, and the cell refinement was performed by SAINT-Plus (Bruker, 2016). An optimized strategy used for data collection consisted of different sets of φ and ω scans with 0.5^{o} steps φ/ω. The data were collected within a time frame of 10 sec by setting the sample to a detector distance fixed at 40 cm. All the data points were corrected for Lorentzian, polarization, and absorption effects using SAINT-Plus and SADABS programs (Bruker, 2016). SHELXS-97 (Sheldrick, 2008) was used for structure solution, and full-matrix least-squares refinement on F^2.[1,2] The molecular graphics of ORTEP diagrams was performed by Mercury software. The crystal symmetry of the components was cross-checked by running the cif files through PLATON (Spek, 2020) software and notified that no additional symmetry was observed. The Encifer software was used to correct the cif files.

X-ray crystal structure analysis of compound 4a:

Crystallization: About 0.020 g of compound **4a** was dissolved in 1 ml of DCM and 4-5 drops of methanol were added. This solution was kept at room temperature for 3 days when needle-shaped colourless crystals were obtained.

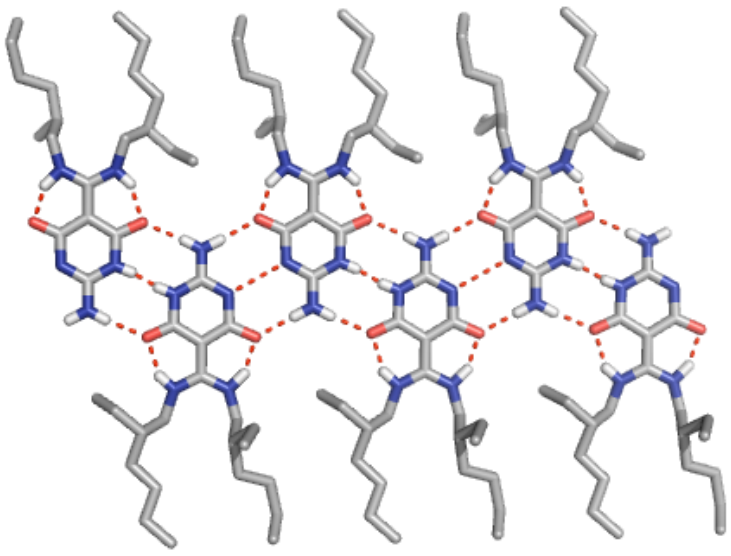


Fig. 22 Single-crystal X-ray of Janus G-C nucleobase **4a** showing tape-like supramolecular self-assembly.

Crystal data	Compound **4a**
Chemical formula	$C_{21}H_{39}N_5O_2$
Formula weight (M_r)	393.58
Crystal system	Triclinic
Space group	*P*-1
Temperature T (K)	100
a (Å)	8.4072 (12)
b (Å)	9.8706 (14)
c (Å)	13.707 (2)
α (°)	106.490 (5)
β (°)	93.080 (5)
γ (°)	94.165 (5)
Z	2
Volume ($Å^3$)	1084.5 (3)
Source of radiation	Mo*K*α (0.71073)
D_{calc} (g cm^{-3})	1.199
Crystal size (mm)	0.12×0.1×0.05
μ (mm^{-1})	0.08
Data collection	
Diffractometer	Bruker D8 VENTURE Kappa Duo PHOTON II CPAD
Absorption correction	Multi-scan (SADABS; Bruker, 2016)
T_{min}, T_{max}	
No. of measured, independent and observed [I > 2σ(I)] reflections	56681, 5174, 3678
Theta range (°)	2.27- 19.70
R_{int}	0.163
Refinement	
R[F^2> 2σ (F^2)], wR(F^2)	0.089, 0.255
GOF on F^2	1.07
No. of independent reflections	5174
No. of parameters	365
F_000	428
No. of restraints	0
H-atom treatment	Constr
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e $A^{\circ -3}$)	0.68, −0.39
CCDC number	2235072

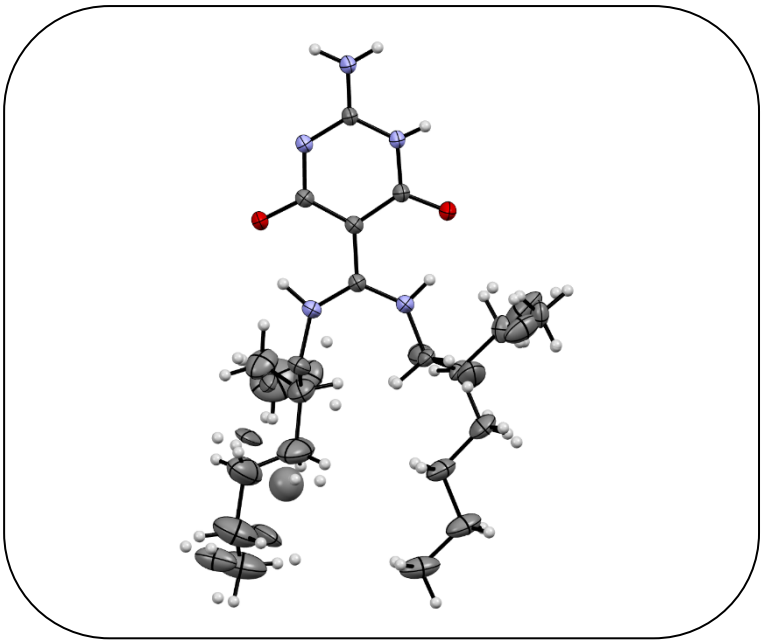


Fig. 23 ORTEP diagram of compound **4a** (ellipsoid probability = 50%)

References:

1. G. M. Sheldrick, Crystal structure refinement with SHELXL, *Acta Cryst*., **2015**, *C71*, 3–8.
2. G. M. Sheldrick, SHELXT - Integrated space-group and crystal-structure determination, *Acta Cryst*., **2015**, *A71*, 3–8.