Supporting Information

Maytansinol Functionalization: Towards Useful Probes for Studying Microtubule Dynamics

Zlata Boiarska,[a]¥ Helena Pérez-Peña,[a]¥ Anne-Catherine Abel,[b]¥ Paola Marzullo,[a]¥ Beatriz Álvarez-Bernad,[c] Francesca Bonato,[c] Benedetta Santini,[a] Dragos Horvath,[d] Daniel Lucena-Agell,[c] Francesca Vasile,[a] Maurizio Sironi,[a] J. Fernando Díaz,[c] Andrea E. Prota,[b] Stefano Pieraccini,*[a] and Daniele Passarella*[a]

[a] Z. Boiarska, H. Pérez-Peña, Dr. P. Marzullo, Dr. B. Santini, Prof. Dr. F. Vasile, Prof. Dr. M. Sironi, Prof. Dr. S. Pieraccini, Prof. Dr. D. Passarella
Department of Chemistry, Università degli Studi di Milano
Via Golgi 19, 20133 Milan (Italy)
E-mail: daniele.passarella@unimi.it
Homepage: https://sites.unimi.it/passalab/

[b] A-C. Abel, Dr. A.E. Prota,
Laboratory of Biomolecular Research, Paul Scherrer Institute
Forschungsstrasse 111, 5232 Villigen PSI (Switzerland)

[c] B. Álvarez-Bernad, F. Bonato, Dr. D. Lucena-Agell, Dr. J. F. Díaz
Centro de Investigaciones Biológicas Margarita Salas
Consejo Superior de Investigaciones Científicas
Ramiro de Maeztu 9
28040 Madrid (Spain)

[c] Dr. D. Horvath
Laboratory of Chemoinformatics, Faculty of Chemistry
University of Strasbourg
67081 Strasbourg (France)

¥ The authors contributed equally

SUMMARY

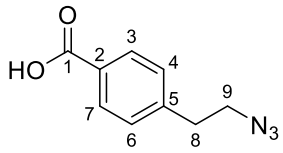
CHEMISTRY

General Experimental Procedures

Unless otherwise stated, reagents were purchased from general suppliers (Sigma Aldrich and Fluorochem) and used without further purification. All solvents were of reagent grade or HPLC grade. All reactions were carried out in oven-dried glassware and dry solvents, under nitrogen atmosphere and were monitored by glasses or aluminium TLC on silica gel (Merck precoated 60F254 plates), with detection by UV light (254 nm), or by TLC stains as permanganate, or by HPLC Agilent 1100. Analytical HPLC was performed on Agilent 1100 Series System RP column ZORBAX SB-C8 (3.5µm x 4.6 x 150 mm). The pressure was about 85 bar, with a constant flow rate of 1 mL/min. UV spectra were recorded at 254 nm and 210 nm with DAD detection. The mobile phase consisted of a mixture of H_2O/ACN and the gradient was programmed using the following method: isocratic for 1 min at 50% ACN, then gradient for 10 min to 90% ACN.

Products were purified using Biotage Isolera™ One System and Biotage® Sfär C18 6 g D Duo 30 µm as cartridges (BIOTAGE).

Synthesis of 4-(2-azidoethyl)benzoic acid

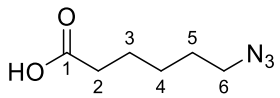


To a solution of 4-(2-bromoethyl)benzoic acid (80 mg, 0.35 mmol) in DMSO (1.4 ml), was added NaN_3 (24.2 mg, 0.38 mmol). The mixture was stirred at room temperature for 20 hours. Then, water (2.8 mL) was added, and the solution was extracted with EtOAc (4 × 6 mL). The organic layer was washed with brine, and the combined organic layers were dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was not purified to provide the product (57.5 mg, 0.30 mmol, 86% yield) as white solid.

^{1}H NMR (400 MHz, chloroform-*d*) δ 8.0 (d, *J* = 8.2 Hz, 2H, 3, 7), 7.3 (d, *J* = 8.1 Hz, 2H, 4, 6), 3.5 (t, *J* = 7.1 Hz, 2H, 9), 2.9 (t, *J* = 7.1 Hz, 2H, 8).

^{13}C NMR (101 MHz, chloroform-*d*) δ 170.4 (1), 144.3 (5), 131.0 (3, 7), 130.8 (2), 129.5 (4, 6), 52.6 (9), 36.0 (8).

Synthesis of 6-azidohexanoic acid



To a solution of 6-bromohexanoic acid (195 mg, 1 mmol) in DMSO (4 ml), was added NaN_3 (69 mg, 1.1 mmol). The mixture was stirred at room temperature for 20 hours. Then, water (4 mL) was added, and the solution was extracted with EtOAc (4 × 8 mL). The organic layer was washed with brine, and the combined organic layers were dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was not purified to provide the product (142 mg, 0.91 mmol, 91% yield) as white solid.

^{1}H NMR (400 MHz, chloroform-*d*) δ 3.3 (t, *J* = 6.9 Hz, 2H, 6), 2.4 (t, *J* = 7.4 Hz, 2H, 2), 1.7 (m, 4H, 3, 5), 1.5 (m, 2H, 4).

^{13}C NMR (101 MHz, chloroform-*d*) δ 180.2 (1), 51.9 (6), 34.5 (2), 29.2 (5), 26.8 (4), 24.8 (3).

Other synthetic procedures have been described in the Experimental Section of the main text.

General maytansinol acylation with carboxylic acid

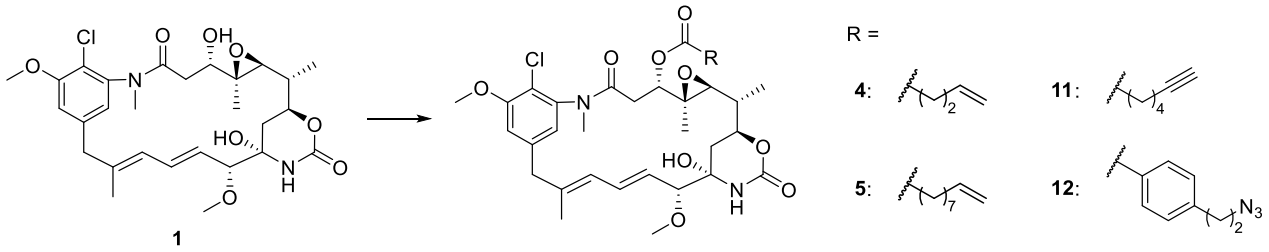
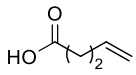
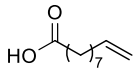
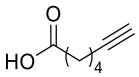
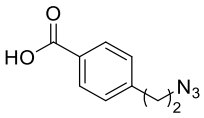


Table S1. Details of reaction condition for esterification of maytansinol

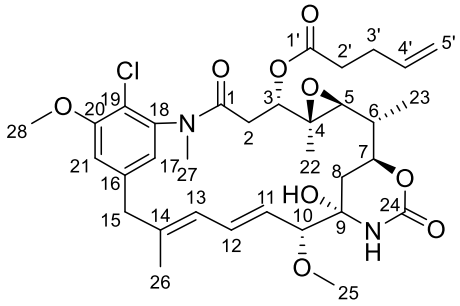
Carboxylic acid	Product	Yield %
	4	68
	5	57
	11	66
	12	68

Characterization data:

^{1}H NMR and ^{13}C-NMR spectra were recorded on a Bruker Avance Spectrometer 400 MHz using commercially available deuterated solvents (chloroform-d, methanol-d4, acetone-d6, DMSO-d6, dichloromethane-d2) at room temperature. Chemical shifts (δ) are reported in parts per million (ppm) and are reported relative to TMS, used as an internal standard. Data for ^{1}H NMR are reported as follows: chemical shift (δ/ppm), multiplicity, coupling constants (Hz). Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, m = multiplet, br s= broad singlet. Data for ^{13}C NMR are reported in terms of chemical shift (δ/ppm).

High resolution mass spectra (HR-MS) were recorded on a Water QToF Premier high resolution UPLC ES MS/MS.

Compound 4



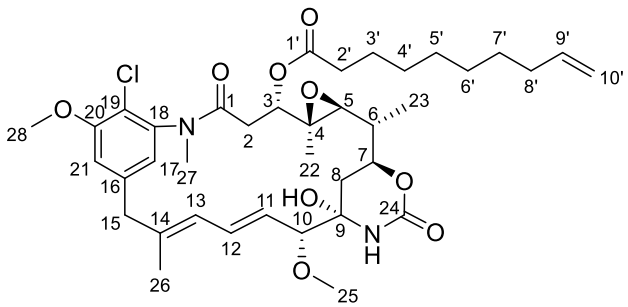
4

^{1}H NMR (400 MHz, Acetone-d6) δ 7.23 (d, J = 1.9 Hz, 1H, 21), 6.96 (d, J = 1.9 Hz, 1H, 17), 6.73 (dd, J = 15.5, 11.1 Hz, 1H, 12), 6.62 (d, J = 5.7 Hz, 1H, OH), 6.40 (s, 1H, NH), 6.32 (d, J = 11.1 Hz, 1H, 13), 5.91 (ddt, J = 16.9, 10.3, 6.5 Hz, 1H, 4'), 5.62 (dd, J = 15.5, 8.9 Hz, 1H, 11), 5.19 – 5.07 (m, 1H, 5'a), 5.02 (d, J = 11.1 Hz, 1H, 5'b), 4.83 (dd, J = 12.0, 2.7 Hz, 1H, 3), 4.25 – 4.09 (m, 1H, 7), 4.00 (s, 3H, 28), 3.62 (m, 2H, 10, 15a), 3.39 – 3.28 (m, 4H, 15b, 25), 3.11 (s, 3H, 25), 2.76 (d, J = 9.9 Hz, 1H, 5), 2.74 – 2.54 (m, 4H, 2, 2'), 2.48 – 2.31 (m, 2H, 3'), 1.75 (s, 3H, 26), 1.70-1.58 (m, 2H, 8), 1.58 – 1.45 (m, 1H, 6), 1.20 (d, J = 6.3 Hz, 3H, 23), 0.93 (s, 3H, 22).

^{13}C NMR (101 MHz, CD2Cl2) δ 171.6 (1), 168.6 (1'), 156.2 (20), 151.8 (24), 142.6 (18), 140.4 (16), 140.4 (14), 137.0 (4'), 132.6 (12), 128.0 (11), 124.4 (13), 122.3 (17), 115.5 (19), 115.1 (5'), 113.3 (21) , 88.2 (10), 81.2 (9), 76.9 (3), 74.3 (7), 66.3 (5), 60.6 (4), 56.7 (25), 56.6 (28), 47.1 (15), 38.5 (6), 35.8 (8), 35.4 (27), 33.4 (2'), 32.9 (2), 28.7 (3'), 15.6 (26), 14.3 (23), 12.1 (22).

HRMS (ESI) *m/z* $[M+Na]^{+}$ 669.2553 (calcd for $C_{33}H_{43}ClN_2O_9Na$, 669.2555)

Compound 5



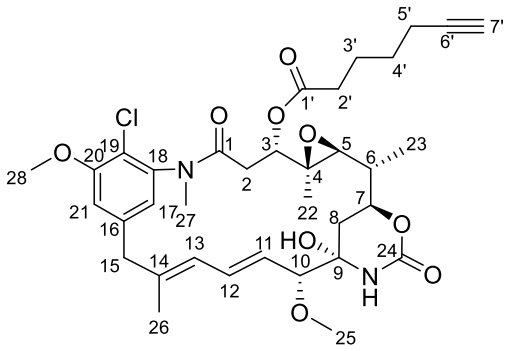
5

^{1}H NMR (400 MHz, Acetone-d6) δ 7.23 (d, J = 1.9 Hz, 1H, 21), 6.96 (d, J = 1.9 Hz, 1H, 17), 6.73 (dd, J = 15.5, 11.0 Hz, 1H, 12), 6.40 (s, 1H, NH), 6.33 (d, J = 11.1 Hz, 1H, 13), 5.80 (ddt, J = 17.0, 10.3, 6.7 Hz, 1H, 9'), 5.62 (dd, J = 15.5, 8.9 Hz, 1H, 11), 5.08 – 4.74 (m, 3H, 10', 3), 4.22 – 4.12 (m, 1H, 7), 4.01 (s, 3H, 28), 3.70 – 3.54 (m, 2H, 10, 15a), 3.40 – 3.31 (m, 4H, 25, 15b), 3.11 (s, 3H, 27), 2.76 (d, J = 9.7 Hz, 1H, 5), 2.65 – 2.45 (m, 3H, 2b, 2'), 2.09 – 1.99 (m, 2H, 8', overlap with (CD3)2CO)), 1.70 – 1.59 (m, 3H, 3', 8a), 1.55 – 1.25 (m, 10H, 8b, 6, 4', 5', 6', 7'), 1.19 (d, J = 6.4 Hz, 3H, 23), 0.93 (s, 3H, 22).

^{13}C NMR (101 MHz, Acetone) δ 172.8 (1), 169.0 (1'), 156.9 (20), 152.0 (24), 143.3 (18), 142.0 (16), 140.5 (14), 139.8 (9'), 133.3 (12), 129.7 (11), 125.8 (13), 123.3 (17), 119.5 (19), 114.7 (10'), 114.5 (21), 89.4 (10), 81.7 (9), 77.5 (3), 74.8 (7), 67.1 (5), 61.5 (4), 57.0 (25), 56.8 (28), 47.2 (15), 39.1 (6), 36.9 (8), 35.6 (27), 34.7 (2'), 34.5 (8'), 33.5 (2), 30.2 (6'), 29.8 (7'), 29.7 (5'), 29.7 (4'), 25.7 (3'), 15.8 (26), 14.8 (23), 12.5 (22).

HRMS (ESI) *m/z* [M+Na]$^{+}$ 739.3320 (calcd for $C_{38}H_{53}ClN_2O_9Na$, 739.3337)

Compound 11



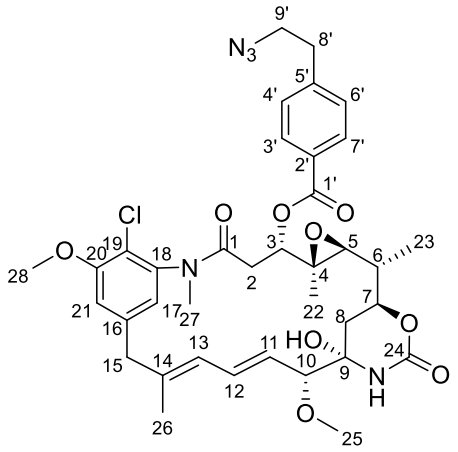
11

^{1}H NMR (400 MHz, chloroform-*d*) δ 6.83 (d, *J* = 2.0 Hz, 1H, 21), 6.79 (d, *J* = 1.8 Hz, 1H, 17), 6.44 (dd, *J* = 15.5, 11.0 Hz, 1H, 12), 6.31 (s, 1H, NH), 6.16 (d, *J* = 10.9 Hz, 1H, 13), 5.49 (dd, *J* = 15.5, 8.9 Hz, 1H, 11), 4.89 (dd, *J* = 11.9, 3.0 Hz, 1H, 3), 4.25 (ddd, *J* = 12.4, 10.6, 2.0 Hz, 1H, 7), 3.99 (s, 3H, 28), 3.56 – 3.46 (m, 2H, 10, 15a), 3.36 (s, 3H, 25), 3.21 (d, *J* = 13.0 Hz, 1H, 15b), 3.17 (s, 3H, 27), 2.89 (d, *J* = 9.7 Hz, 1H, 5), 2.57 – 2.44 (m, 2H, 2'a, 2b), 2.44 – 2.31 (m, 1H, 2'b), 2.26 (td, *J* = 6.9, 2.7 Hz, 2H, 5'), 2.23 – 2.15 (m, 1H, 2a), 1.95 (t, *J* = 2.6 Hz, 1H, 7'), 1.81 (p, *J* = 7.6 Hz, 2H, 3'), 1.68 (s, 3H, 26), 1.66 – 1.54 (m, 3H, 4', 8b), 1.54 – 1.42 (m, 1H, 6), 1.28 (d, *J* = 6.4 Hz, 3H, 23), 1.26 – 1.18 (m, 1H, 8a), 0.83 (s, 3H, 22).

^{13}C NMR (101 MHz, chloroform-*d*) δ 171.9 (1), 168.8 (1'), 156.2 (20), 152.4 (24), 142.7 (18), 140.2 (16), 140.1 (14), 132.5 (12), 128.2 (11), 124.6 (13), 122.3 (17), 119.6 (19), 113.1 (21), 88.3 (10), 84.2 (6'), 81.1 (9), 77.0 (3), 74.4 (7), 69.0 (7'), 66.5 (5), 60.4 (4), 56.9 (25), 56.7 (28), 47.3 (15), 38.6 (6), 35.9 (8), 35.7 (27), 33.7 (2'), 32.9 (2), 27.9 (4'), 23.9 (3'), 18.4 (5'), 15.9 (26), 14.6 (23), 12.2 (22).

HRMS (ESI) *m/z* [M+Na]$^{+}$ 695.2710 (calcd for $C_{35}H_{45}ClN_2O_9Na$, 695.2711)

Compound 12



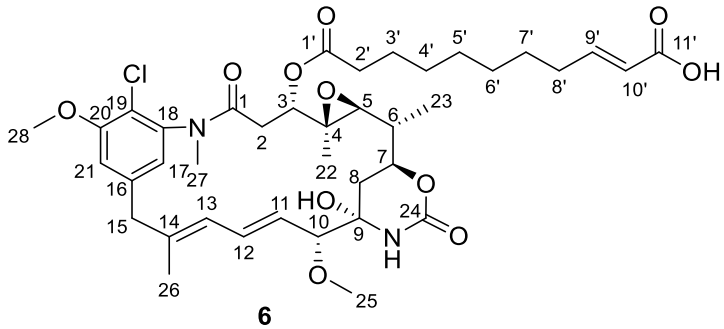
12

^{1}H NMR (400 MHz, acetone-d_6) δ 8.1 (d, *J* = 8.2 Hz, 2H, 3', 7'), 7.6 (d, *J* = 8.0 Hz, 2H, 4', 6'), 7.3 (d, *J* = 1.9 Hz, 1H, 21), 7.2 (d, *J* = 1.8 Hz, 1H, 17), 6.6 (dd, *J* = 15.4, 11.0 Hz, 1H, 12), 6.3 (s, 1H, NH), 6.1 (d, *J* = 11.0 Hz, 1H, 13), 5.1 – 4.9 (m, 2H, 3, 11), 4.4 (d, *J* = 2.1 Hz, 1H, OH), 4.2 (ddd, *J* = 12.5, 10.5, 2.3 Hz, 1H, 7), 4.0 (s, 3H, 28), 3.7 (td, *J* = 7.0, 2.1 Hz, 2H, 9'), 3.6 (d, *J* = 12.6 Hz, 1H, 15a), 3.5 (d, *J* = 9.1 Hz, 1H, 10), 3.4 (d, *J* = 12.6 Hz, 1H, 15b), 3.2 (s, 3H, 25), 3.1 (s, 3H, 27), 3.1 – 3.0 (m, 3H, 5, 8'), 2.8 (d, *J* = 14.4 Hz, 1H, 2a), 2.3 (dd, *J* = 14.4, 3.2 Hz, 1H, 2b), 1.8 (s, 3H, 26), 1.6 (dt, *J* = 13.5, 2.0 Hz, 1H, 8a), 1.5 – 1.4 (m, 2H, 6, 8b), 1.2 (d, *J* = 6.4 Hz, 3H, 23), 1.0 (s, 3H, 22).

^{13}C NMR (101 MHz, acetone-d_6) δ 168.7 (1), 166.4 (1'), 156.9 (20), 151.7 (24), 145,0 (5'), 143.1 (18), 142.0 (16), 140.1 (14), 133.0 (12), 130.8 (3', 7'), 129.6 (2'), 129.5 (4', 6'), 129.3 (11), 125.5 (13), 122.8 (17), 119.3 (19), 114.2 (21), 89.5 (10), 81.2 (9), 78.0 (3), 74.4 (7), 67.0 (5), 61.2 (4), 56.8 (28), 56.4 (25), 52.5 (9'), 46.9 (15), 39.4 (6), 36.8 (8), 35.6 (8'), 35.2 (27), 33.4 (2), 15.5 (26), 14.7 (23), 12.8 (22).

HRMS (ESI) *m/z* $[M+Na]^+$ 760.2732 (calcd for $C_{37}H_{44}ClN_5O_9Na$, 760.2725).

Compound 6



6

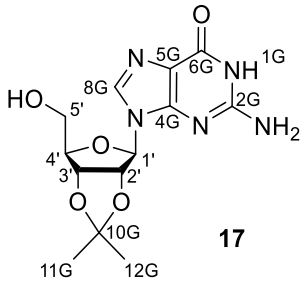
^{1}H NMR (400 MHz, Acetone-d6) δ 7.23 (d, J = 1.8 Hz, 1H, 21), 6.96 (d, J = 1.8 Hz, 1H, 17), 6.95 – 6.84 (m, 1H, 9'), 6.73 (dd, J = 15.4, 11.1 Hz, 1H, 12), 6.43 (bs, 1H, NH), 6.33 (d, J = 11.1 Hz, 1H, 13), 5.82 (d, J = 15.6 Hz, 1H, 10'), 5.62 (dd, J = 15.4, 8.9 Hz, 1H, 11), 4.83 (dd, J = 11.9, 2.7 Hz, 1H, 3), 4.18 (ddd, J = 12.5, 10.4, 2.2 Hz, 1H, 7), 4.01 (s, 3H, 28), 3.68 – 3.55 (m, 2H, 10, 15a), 3.39 – 3.25 (m, 4H, 25, 15b), 3.11 (s, 3H, 27), 2.76 (d, J = 9.7 Hz, 1H, 5), 2.67 – 2.40 (m, 4H, 2, 2'), 2.27 – 2.18 (m, 2H, 8'), 2.05 – 2.02 (m, 2H, 8'), 1.75 (s, 3H, 26), 1.71 – 1.60 (m, 2H, 3'), 1.52 – 1.46 (m, 3H, 7', 6), 1.45 – 1.35 (m, 6H, 4',5', 6'), 1.20 (d, J = 6.4 Hz, 3H, 23), 0.93 (s, 3H,22).

^{13}C NMR (101 MHz, Acetone) δ 172.8 (1), 169.1 (1'), 167.7 (11'), 156.9 (20), 152.1 (24), 149.8 (9'), 143.3 (18), 142.0 (16), 140.5 (14), 133.3 (12), 129.7 (11), 125.8 (13), 123.3 (17), 122.6 (10'), 119.5 (19), 114.5 (21), 89.5 (10), 81.7 (9), 77.5 (3),

74.8 (7), 67.1 (5), 61.5 (4), 57.0 (28), 56.8 (25), 47.2 (15), 39.1 (6), 36.9 (8), 35.6 (27), 34.7 (2'), 33.5 (6'), 32.6 (2), 28.9 (5'), 26.5 (7'), 25.8 (4'), 25.7 (3'), 15.8 (26), 14.8 (23), 12.5 (22).

HRMS (ESI) *m/z* $[M+Na]^+$ 783.3235 (calcd for $C_{39}H_{53}ClN_2O_{11}Na$, 783.3236).

Compound 17

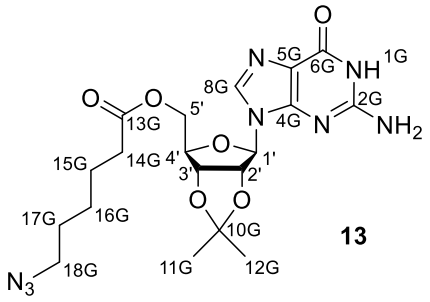


17

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.9 (s, 1H, 8G), 6.4 (s, 2H, NH_2), 5.9 (d, *J* = 2.8 Hz, 1H, 1'), 5.2 (dd, *J* = 6.4, 2.8 Hz, 1H, 2'), 5.0 (dd, *J* = 6.4, 2.5 Hz, 1H, 3'), 4.1 (dt, *J* = 4.7, 2.5 Hz, 1H, 4'), 3.6 (t, *J* = 4.7 Hz, 2H, 5'), 2.1 (s, 1H, OH), 1.5 (s, 3H, 11G), 1.3 (s, 3H, 12G).

^{13}C NMR (101 MHz, DMSO) δ 157.2 (6G), 154.2 (2G), 151.2 (4G), 136.3 (8G), 117.4 (5G), 113.6 (10G), 89.1 (1'), 87.1 (4'), 84.1 (2'), 81.7 (3'), 62.2 (5'), 27.6 (11G), 25.8 (12G).

Compound 13



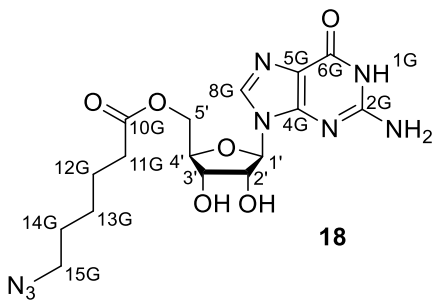
13

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.7 (s, 1H, NH_2), 7.9 (s, 1H, 8G), 6.6 (s, 1H), 6.2 – 5.9 (m, 1H, 1'), 5.5 – 5.2 (m, 1H, 2'), 5.1 (dd, *J* = 5.9, 3.4 Hz, 1H, 3'), 4.2 (dd, *J* = 8.0, 4.0 Hz, 2H, 4', 5'a), 4.1 (dd, *J* = 12.6, 8.0 Hz, 1H, 5'b), 3.3 (t, *J* = 6.9 Hz, 2H, 18G), 2.3 (dt, *J* = 7.0, 3.4 Hz, 2H, 14G), 1.6 – 1.4 (m, 7H, 11G, 15G, 17G), 1.3 (s, 5H, 12G, 16G).

^{13}C NMR (101 MHz, DMSO-d_6) δ 173.7 (13G), 157.8 (6G), 154.8, 151.6 (4G), 137.3, 118.1 (5G), 114.4 (10G), 89.4 (1'), 85.3 (4'), 84.8 (2'), 82.2 (3'), 65.1 (5'), 52.0 (18G), 34.2 (14G), 29.0 (17G), 28.1 (11G), 26.7 (15G), 26.4 (12G), 25.0 (16G).

HRMS (ESI) *m/z* $[M+Na]^+$ 485.1881 (calcd for $C_{19}H_{26}N_8O_6Na$, 485.1873)

Compound 18



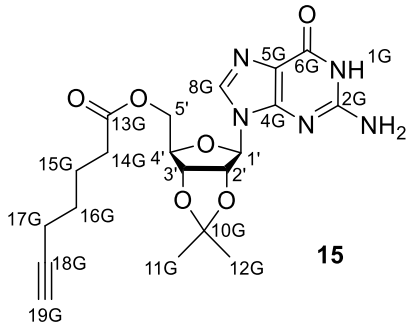
18

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.8 (s, 1H, 1G), 7.9 (s, 1H), 6.6 (s, 2H, NH_2), 5.7 (d, *J* = 5.1 Hz, 1H, 1'), 4.4 (t, *J* = 5.2 Hz, 1H), 4.3 (dd, *J* = 12.0, 3.7 Hz, 1H, 5'a), 4.2 (dt, *J* = 9.8, 5.3 Hz, 2H, 3', 5'b), 4.0 (q, *J* = 4.8 Hz, 1H), 3.3 (d, *J* = 7.2 Hz, 2H, 15G), 2.3 (t, *J* = 7.4 Hz, 2H, 11G), 1.5 (h, *J* = 6.8 Hz, 4H, 12G, 14G), 1.3 (p, *J* = 4.8, 4.2 Hz, 2H, 13G).

^{13}C NMR (101 MHz, DMSO-d_6) δ 173.8 (10G), 157.8 (6G), 154.9 (2G), 152.4 (4G), 136.5 (8G), 117.8 (5G), 87.8 (1'), 82.5 (4'), 74.2 (2'), 71.4 (3'), 65.0 (5'), 51.6 (15G), 34.3 (11G), 29.0 (14G), 26.7 (13G), 25.0 (12G).

HRMS (ESI) *m/z* $[M+Na]^+$ 445.1569 (calcd for $C_{16}H_{22}N_8O_6Na$, 445.1560)

Compound 15



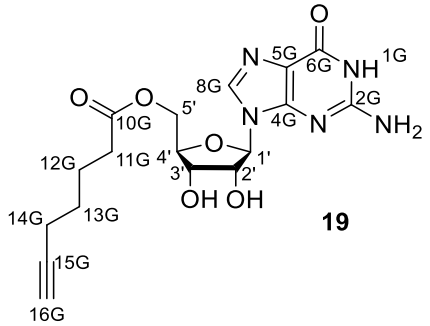
15

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.7 (s, 1H, 1G), 7.9 (s, 1H, 8G), 6.5 (s, 2H, NH_2), 6.0 (d, *J* = 2.0 Hz, 1H, 1'), 5.3 (dd, *J* = 6.3, 2.0 Hz, 1H, 2'), 5.1 (dd, *J* = 6.3, 3.5 Hz, 1H, 3'), 4.3 (dt, *J* = 6.1, 4.3 Hz, 2H, 4', 5'a), 4.1 (dd, *J* = 12.8, 7.9 Hz, 1H, 5'b), 2.7 (t, *J* = 2.6 Hz, 1H, 19G), 2.3 (td, *J* = 7.4, 3.2 Hz, 2H, 14G), 2.1 (dq, *J* = 7.0, 3.1 Hz, 2H, 17G), 1.6 – 1.5 (m, 2H, 15G), 1.5 (s, 3H, 11G), 1.4 (p, *J* = 7.2 Hz, 2H, 16G), 1.3 (s, 3H, 12G).

^{13}C NMR (101 MHz, DMSO-d_6) δ 173.6 (13G), 157.8 (6G), 154.8 (2G), 151.6 (4G), 137.3 (8G), 118.1 (5G), 114.4 (10G), 89.4 (1'), 85.3 (18G), 85.3 (4'), 84.8 (2'), 82.2 (3'), 72.4 (19G), 65.1 (5'), 33.8 (14G), 28.4 (16G), 28.1 (11G), 26.4 (12G), 24.6 (15G), 18.5 (17G).

HRMS (ESI) *m/z* $[M+Na]^+$ 454.1710 (calcd for $C_{20}H_{25}N_5O_6Na$, 454.1703)

Compound 19



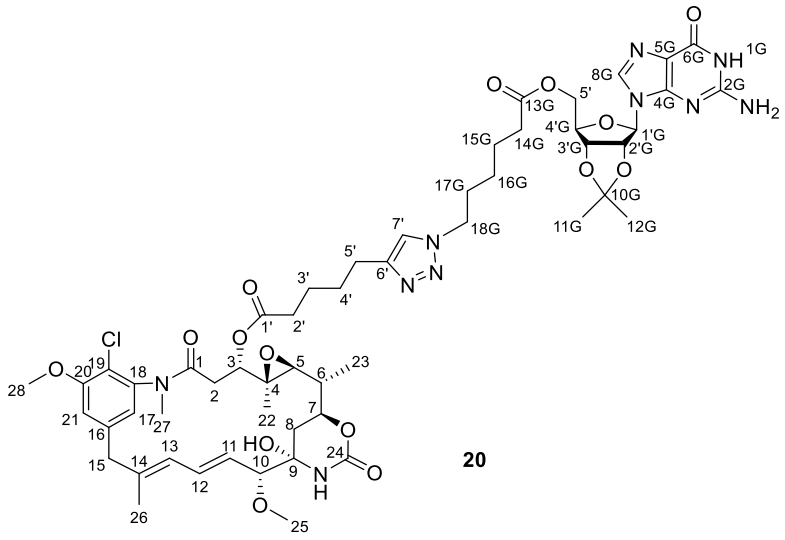
19

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.7 (s, 1H, 1G), 7.9 (s, 1H, 8G), 6.5 (s, 2H, NH_2), 5.8 – 5.7 (m, 1H, 1'), 5.5 (d, *J* = 5.4 Hz, 1H, 2'OH), 5.3 (d, *J* = 4.8 Hz, 1H, 3'OH), 4.4 (d, *J* = 5.0 Hz, 1H, 2'), 4.3 (dd, *J* = 11.9, 3.5 Hz, 1H, 5'a), 4.2 – 4.1 (m, 2H, 3', 5'b), 4.0 (q, *J* = 5.9, 4.5 Hz, 1H, 4'), 2.7 (d, *J* = 2.4 Hz, 1H, 16G), 2.3 (t, *J* = 7.2 Hz, 2H, 11G), 2.2 (td, *J* = 6.9, 2.4 Hz, 2H, 14G), 1.6 (q, *J* = 7.4 Hz, 2H, 12G), 1.5 (q, *J* = 7.2 Hz, 2H, 13G).

^{13}C NMR (101 MHz, DMSO-d_6) δ 173.7 (10G), 157.9 (6G), 154.8 (2G), 152.4 (4G), 136.6 (8G), 117.8 (5G), 87.8 (1'), 85.3 (15G), 82.5 (4'), 74.2 (2'), 72.4 (16G), 71.4 (3'), 65.0 (5'), 33.9 (11G), 28.4 (13G), 24.7 (12G), 18.5 (14G).

HRMS (ESI) *m/z* $[M+Na]^+$ 414.1396 (calcd for $C_{17}H_{21}N_5O_6Na$, 414.1390)

Compound 20



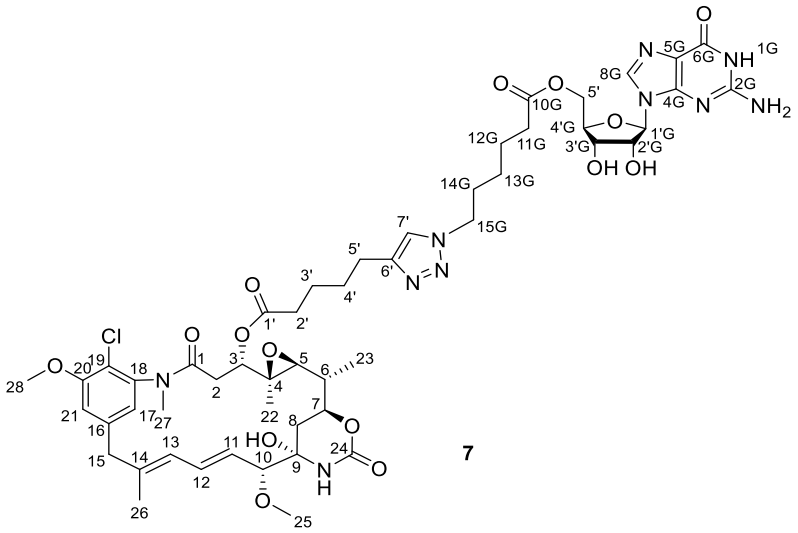
20

^{1}H (600 MHz, DMSO-d_6) δ 7.9 (s, 1H, 7'), 7.8 (d, *J* = 2.0 Hz, 1H, 8G), 7.2 (d, *J* = 1.8 Hz, 1H, 21), 7.0 (s, 2H, NH_2), 6.9 (s, 1H, NH), 6.8 (d, *J* = 1.8 Hz, 1H, 17), 6.6 (dd, *J* = 15.4, 11.1 Hz, 1H, 12), 6.2 (d, *J* = 11.1 Hz, 1H, 13), 6.0 (d, *J* = 2.0 Hz, 1H, 1'G), 5.4 (dd, *J* = 15.4, 8.9 Hz, 1H, 11), 5.3 (dd, *J* = 6.3, 2.1 Hz, 1H, 2'G), 5.1 (dd, *J* = 6.3, 3.4 Hz, 1H, 3'G), 4.6 (dd, *J* = 11.9, 2.7 Hz, 1H, 3), 4.3 – 4.2 (m, 4H, 4'G, 5'Ga, 18G), 4.2 – 4.0 (m, 2H, 5'Gb, 7), 3.9 (s, 3H, 28), 3.6 – 3.5 (m, 2H, 10, 15a), 3.3 (s, 1H, 15b), 3.2 (s, 3H, 25), 2.9 (s, 3H, 27), 2.7 (dt, *J* = 13.6, 7.4 Hz, 2H, 5'), 2.6 (d, *J* = 9.8 Hz, 1H, 5), 2.5 – 2.4 (m, 3H, 2', 2a), 2.3 (td, *J* = 7.6, 4.9 Hz, 2H, 14G), 2.0 (dd, *J* = 13.5, 2.3 Hz, 1H, 2b), 1.7 (dt, *J* = 13.3, 6.8 Hz, 4H, 3', 17G), 1.7 – 1.6 (m, 5H, 4', 26), 1.5 (s, 3H, 11G), 1.5 – 1.4 (m, 4H, 6, 8a, 15G), 1.4 – 1.3 (m, 1H, 8b), 1.3 (s, 3H, 12G), 1.2 – 1.1 (m, 2H, 16G), 1.1 (d, *J* = 6.4 Hz, 3H, 23), 0.8 (s, 3H, 22).

^{13}C NMR (151 MHz, DMSO-d_6) δ 173.0 (13G), 172.3 (1'), 168.5 (1), 155.7 (20), 154.9 (6G), 151.8 (24), 151.7 (2G), 151.1 (4G), 146.9 (6'), 141.9 (18), 141.5 (16), 139.5 (14), 136.3 (8G), 132.3 (12), 129.2 (11), 125.0 (13), 122.3 (17), 122.1 (7'), 117.9 (19), 117.4 (5G), 114.4 (21), 113.7 (10G), 88.8 (1'G), 88.5 (10), 84.6 (4'G), 84.1 (2'G), 81.6 (3'G), 80.7 (9), 76.5 (3), 73.9 (7), 66.2 (5), 64.4 (5'G), 61.1 (4), 57.0 (28), 56.7 (25), 49.4 (18G), 46.1 (15), 37.8 (6), 36.5 (8), 35.5 (27), 33.5 (2'), 33.5 (2), 32.7 (14G), 30.2 (4'), 29.9 (17G), 28.8 (5'), 28.8 (11G), 28.7 (12G), 25.4 (3'), 24.6 (15G), 15.7 (26), 14.8 (23), 11.8 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1157.4691 (calcd for $C_{54}H_{71}ClN_{10}O_{15}Na$ 1157.4687);

Compound 7



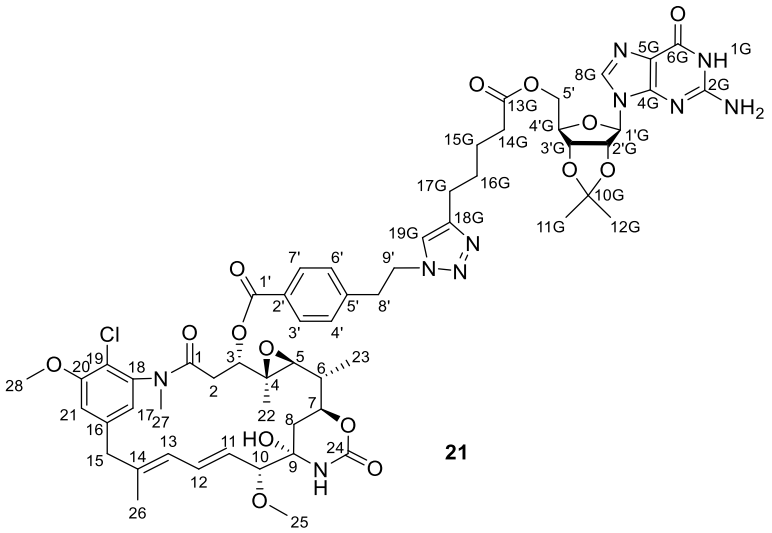
7

^{1}H (600 MHz, DMSO-d_6) δ 7.9 (s, 2H, 7', 8G), 7.3 – 7.2 (m, 1H, 21), 6.9 (s, 1H, NH), 6.8 – 6.7 (m, 1H, 17), 6.6 (s, 2H, NH_2), 6.6 (dd, *J* = 15.3, 11.2 Hz, 1H, 12), 6.2 (d, *J* = 11.1 Hz, 1H, 13), 5.7 (d, *J* = 5.1 Hz, 1H, 1'G), 5.4 (dd, *J* = 15.3, 8.9 Hz, 1H, 11), 4.6 (dd, *J* = 11.9, 2.3 Hz, 1H, 3), 4.4 (t, *J* = 5.1 Hz, 1H, 2'G), 4.3 (dd, *J* = 11.9, 3.7 Hz, 1H, 5'Ga), 4.2 (t, *J* = 7.1 Hz, 2H, 15G), 4.2 (dt, *J* = 12.8, 5.5 Hz, 2H, 3'G, 5'Gb), 4.1 (t, *J* = 11.9 Hz, 1H, 7), 4.0 (q, *J* = 4.8 Hz, 1H, 4'G), 3.9 (s, 3H, 28), 3.5 (d, *J* = 12.6 Hz, 1H, 15a), 3.5 (d, *J* = 8.8 Hz, 1H, 10), 3.3 (d, *J* = 12.7 Hz, 1H, 15b), 3.2 (s, 3H, 25), 2.9 (s, 3H, 27), 2.6 (q, *J* = 6.8 Hz, 2H, 5'), 2.6 (d, *J* = 9.7 Hz, 1H, 5), 2.5 – 2.4 (m, 3H, 2', 2b), 2.3 (dt, *J* = 7.8, 3.8 Hz, 2H, 11G), 2.0 (d, *J* = 11.9 Hz, 1H, 2a), 1.8 – 1.7 (m, 4H, 3', 14G), 1.6 (s, 5H, 4', 26), 1.6 – 1.4 (m, 4H, 6, 8b, 12G), 1.3 (d, *J* = 12.5 Hz, 1H, 8a), 1.2 (p, *J* = 7.6 Hz, 2H, 13G), 1.1 (d, *J* = 6.3 Hz, 3H, 23), 0.8 (s, 3H, 22).

^{13}C NMR (151 MHz, DMSO-d_6) δ 173.1 (10G), 172.3 (1'), 168.5 (1), 157.4 (6G), 155.7 (20), 154.4 (2G), 151.8 (4G), 151.7 (24), 146.9 (6'), 141.9 (18), 141.5 (16), 139.5 (14), 135.4 (8G), 132.3 (12), 129.1 (11), 125.2 (13), 122.3 (17), 122.1 (7'), 117.9 (19), 117.2 (5G), 114.4 (21), 88.5 (10), 87.1 (1'G), 81.8 (4'G), 80.6 (9), 76.5 (3), 73.9 (7), 73.6 (2'G), 70.7 (3'G), 66.2 (5), 64.3 (5'G), 61.1 (4), 57.0 (28), 56.7 (25), 49.4 (15G), 46.1 (15), 37.8 (6), 36.7 (8), 35.5 (27), 33.6 (2'), 33.5 (11G), 32.7 (2), 29.9 (14G), 28.8 (4'), 25.8 (5'), 25.4 (13G), 24.6 (12G), 24.2 (3'), 15.7 (26), 14.8 (23), 11.9 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1117.4380 (calcd for $C_{51}H_{67}ClN_{10}O_{15}Na$ 1117.4374);

Compound 21



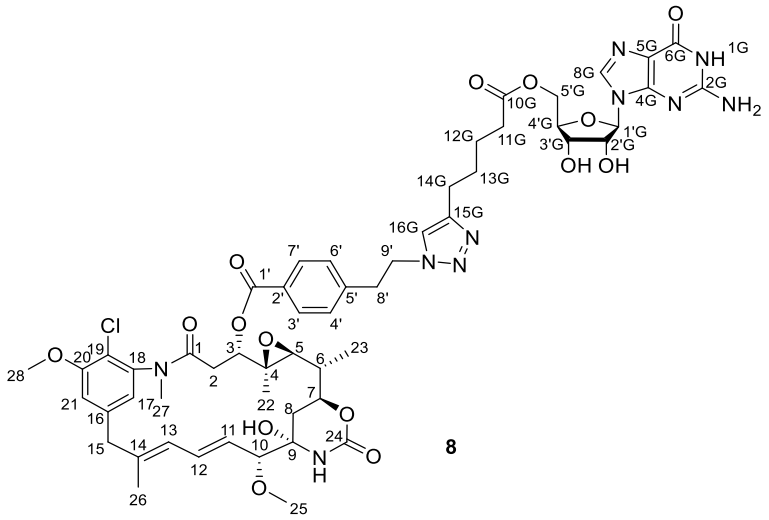
21

^{1}H (600 MHz, DMSO-d_6) δ 7.9 – 7.8 (m, 3H, 3', 7', 8G), 7.8 (s, 1H, 19G), 7.5 (d, *J* = 8.1 Hz, 2H, 4', 6'), 7.2 (s, 1H, 21), 6.9 (s, 1H, 17), 6.8 (s, 1H, NH), 6.7 (s, 2H, NH_2), 6.5 (dd, *J* = 15.4, 11.1 Hz, 1H, 12), 6.0 – 6.0 (m, 1H, 1'G), 5.9 (d, *J* = 11.1 Hz, 1H, 13), 5.3 (dd, *J* = 6.2, 1.6 Hz, 1H, 2'G), 5.1 (dd, *J* = 6.0, 3.4 Hz, 1H, 3'G), 4.9 (dd, *J* = 15.4, 9.2 Hz, 1H, 11), 4.7 (dd, *J* = 12.0, 2.6 Hz, 1H, 3), 4.6 (t, *J* = 7.2 Hz, 2H, 9'), 4.3 – 4.2 (m, 2H, 4'G, 5'Gb), 4.2 – 4.1 (m, 2H, 5'Ga, 7), 4.0 (s, 3H, 28), 3.4 (d, *J* = 12.6 Hz, 1H, 15a), 3.4 (d, *J* = 9.2 Hz, 1H, 10), 3.3 (d, *J* = 7.7 Hz, 3H, 8', 15b), 3.2 (s, 3H, 25), 3.0 (s, 3H, 27), 2.8 (d, *J* = 9.7 Hz, 1H, 5), 2.6 – 2.6 (m, 1H, 2a), 2.5 (d, *J* = 6.7 Hz, 2H, 17G), 2.3 – 2.2 (m, 3H, 2b, 14G), 1.6 (s, 3H, 26), 1.5 (s, 3H, 12G), 1.5 – 1.4 (m, 6H, 6, 8b, 15G, 16G), 1.3 (s, 4H, 8a, 11G), 1.1 (d, *J* = 6.3 Hz, 3H, 23), 0.8 (s, 3H, 22).

^{13}C NMR (151 MHz, DMSO-d_6) δ 173.0 (13G), 168.4 (1), 165.7 (1'), 157.4 (6G), 155.9 (20), 154.4 (2G), 151.7 (24), 151.0 (4G), 146.8 (18G), 144.0 (5'), 141.9 (18), 141.6 (16), 139.5 (14), 136.1 (8G), 132.3 (12), 129.9 (3', 7'), 129.2 (4', 6'), 129.0 (11), 128.4 (2'), 125.0 (13), 122.2 (17), 122.0 (19G), 118.0 (19), 117.4 (5G), 114.2 (21), 113.8 (10G), 88.9 (10), 88.8 (1'G), 84.6 (4'G), 84.1 (2'G), 81.5 (3'G), 80.4 (9), 76.9 (3), 73.7 (7), 65.8 (5), 64.4 (5'G), 61.1 (4), 57.1 (28), 56.5 (25), 50.2 (9'), 46.1 (15), 38.4 (6), 36.7 (8), 36.1 (8'), 35.5 (27), 33.4 (2, 14G), 28.7 (15G), 27.5 (11G, 12G), 25.1 (17G), 24.3 (16G), 15.8 (26), 15.0 (23), 12.5 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1191.4546 (calcd for $C_{57}H_{69}ClN_{10}O_{15}Na$ 1191.4530).

Compound 8



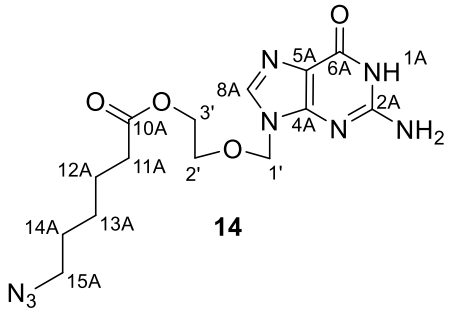
8

^{1}H (600 MHz, DMSO-d_6) δ 7.8 (d, *J* = 7.1 Hz, 3H, 3', 7', 8G), 7.8 (s, 1H, 16G), 7.5 (d, *J* = 8.1 Hz, 2H, 4', 6'), 7.2 (s, 1H, 21), 6.9 (s, 1H, 17), 6.8 (s, 1H, 1G), 6.7 (s, 2H, NH_2), 6.5 (dd, *J* = 15.3, 11.3 Hz, 1H, 12), 5.9 (d, *J* = 11.3 Hz, 1H, 13), 5.7 (d, *J* = 5.1 Hz, 1H, 1'G), 4.9 (dd, *J* = 15.3, 9.1 Hz, 1H, 11), 4.7 (dd, *J* = 12.0, 2.7 Hz, 1H, 3), 4.6 (t, *J* = 7.1 Hz, 2H, 9'), 4.4 (t, *J* = 5.0 Hz, 1H, 2'G), 4.3 (dd, *J* = 11.8, 3.3 Hz, 1H, 5'Ga), 4.2 – 4.1 (m, 3H, 3'G, 5'Gb, 7), 4.0 (q, *J* = 4.5 Hz, 1H, 4'G), 4.0 (s, 3H, 28), 3.4 (d, *J* = 12.5 Hz, 1H, 15b), 3.4 (d, *J* = 9.1 Hz, 1H, 10), 3.3 – 3.2 (m, 3H, 8', 15a), 3.2 (s, 3H, 25), 3.0 (s, 3H, 27), 2.8 (d, *J* = 9.7 Hz, 1H, 5), 2.6 (t, *J* = 13.9 Hz, 1H, 2a), 2.6 – 2.5 (m, 2H, 14G), 2.3 (q, *J* = 7.4 Hz, 2H, 11G), 2.2 (d, *J* = 12.1 Hz, 1H, 2b), 1.6 (s, 3H, 26), 1.6 – 1.5 (m, 4H, 12G, 13G), 1.5 – 1.4 (m, 2H, 6, 8a), 1.3 (d, *J* = 12.6 Hz, 1H, 8b), 1.1 (d, *J* = 6.3 Hz, 3H, 23), 0.8 (s, 3H, 22).

^{13}C NMR (151 MHz, DMSO-d_6) δ 173.1 (10G), 168.5 (1), 165.7 (1'), 157.4 (6G), 155.9 (20), 154.4 (2G), 151.8 (4G), 151.7 (24), 146.9 (15G), 144.0 (5'), 141.9 (18), 141.7 (16), 139.5 (14), 136.2 (8G), 132.3 (12), 129.9 (3', 7'), 129.2 (4', 6'), 129.1 (11), 128.5 (2'), 125.0 (13), 122.2 (17), 122.0 (16G), 118.0 (19), 117.2 (5G), 114.3 (21), 88.9 (10), 87.1 (1'G), 81.8 (4'G), 80.4 (9), 76.9 (3), 73.7 (2'G), 73.6 (7), 70.8 (3'G), 65.9 (5), 64.3 (5'G), 61.1 (4), 57.1 (28), 56.6 (25), 50.2 (9'), 46.1 (15), 39.6, 38.4 (6), 37.8 (8), 36.1 (8'), 35.6 (27), 33.5 (2), 33.5 (11G), 28.7 (12G), 25.1 (14G), 24.4 (13G), 15.8 (26), 15.0 (23), 12.5 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1151.4226 (calcd for $C_{54}H_{65}ClN_{10}O_{15}Na$ 1151.4217);

Compound 14

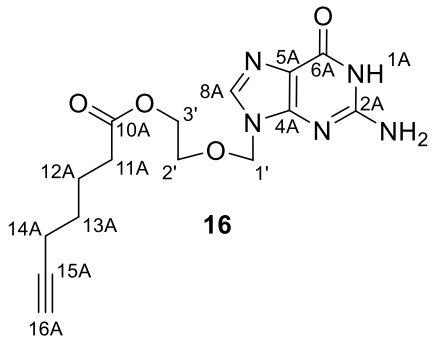


^{1}H NMR (400 MHz, DMSO-d6) δ 10.63 (s, 1H, NH), 7.81 (s, 1H, 8A), 6.51 (s, 2H, NH_2), 5.35 (s, 2H, 1'), 4.09 (t, J = 4.5 Hz 1H, 3'), 3.66 (t, 1H, J = 4.5 Hz, 2'), 3.31 (t, J = 6.8 Hz, 2H, 15A), 2.24 (t, J = 7.4 Hz, 2H, 11A), 1.57 – 1.43 (m, 4H, 12A, 14A), 1.35 – 1.23 (m, 2H, 13A).

13C NMR (101 MHz, DMSO) δ 172.7 (10A), 156.9 (6A), 154.0 (2A), 151.5 (4A), 116.5 (5A), 71.8 (1'), 66.6 (2'), 62.6 (3'), 50.5 (15A), 33.2 (11A), 27.9 (14A), 25.6 (13A), 23.9 (12A).

HRMS (ESI) m/z $[M+Na]^+$ 387.1522 (calcd for $C_{14}H_{20}N_8O_4Na$ 387.1505)

Compound 16

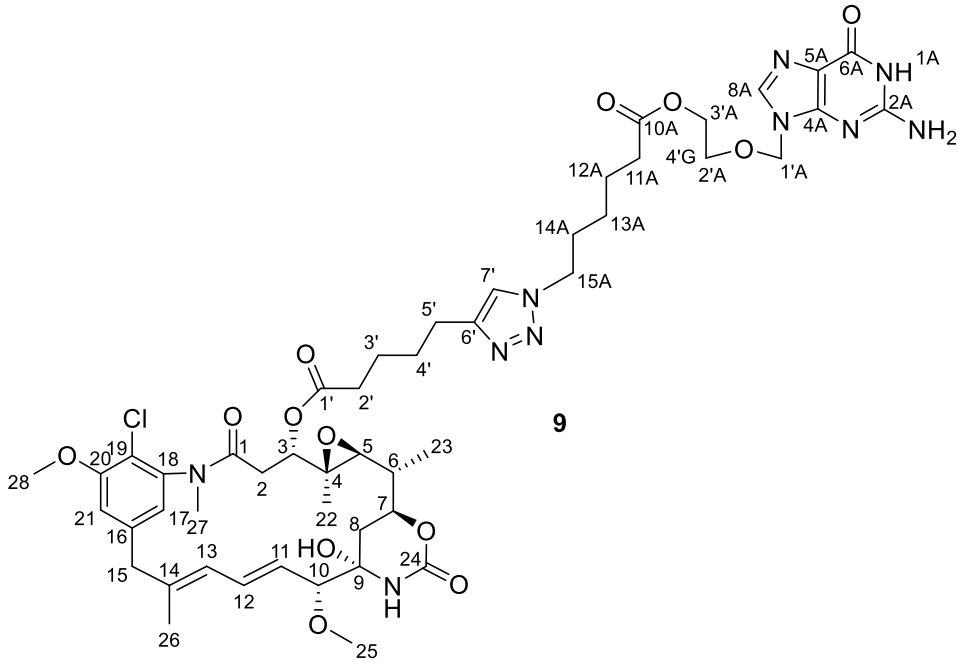


^{1}H NMR (400 MHz, DMSO-d_6) δ 7.81 (s, 1H, 8A), 6.49 (s, 2H, NH_2), 5.35 (s, 2H, 1'), 4.12 – 4.06 (t, *J* = 4.9 Hz, 2H, 3'), 3.70 – 3.63 (t, J = 4.9 Hz 2H, 2'), 2.74 (t, *J* = 2.7 Hz, 1H, 16A), 2.25 (t, *J* = 7.4 Hz, 2H, 11A), 2.15 (td, *J* = 7.0, 2.7 Hz, 2H, 14A), 1.62 – 1.50 (m, 2H, 12A), 1.42 (p, *J* = 7.0 Hz, 2H, 13A).

13C NMR (101 MHz, DMSO-d6) δ 172.6 (10A), 156.8 (6A), 153.9 (2A), 151.4 (4A), 137.7 (8A), 116.5 (5A), 84.2 (15A), 71.8 (1'), 71.3 (16A), 66.5 (2'), 62.6 (3'), 32.8 (11A), 27.2 (13A), 23.5 (12A), 17.4 (14A).

HRMS (ESI) m/z $[M+Na]^+$ 356.1330 (calcd for $C_{15}H_{19}N_5O_4Na$ 356.1335)

Compound 9



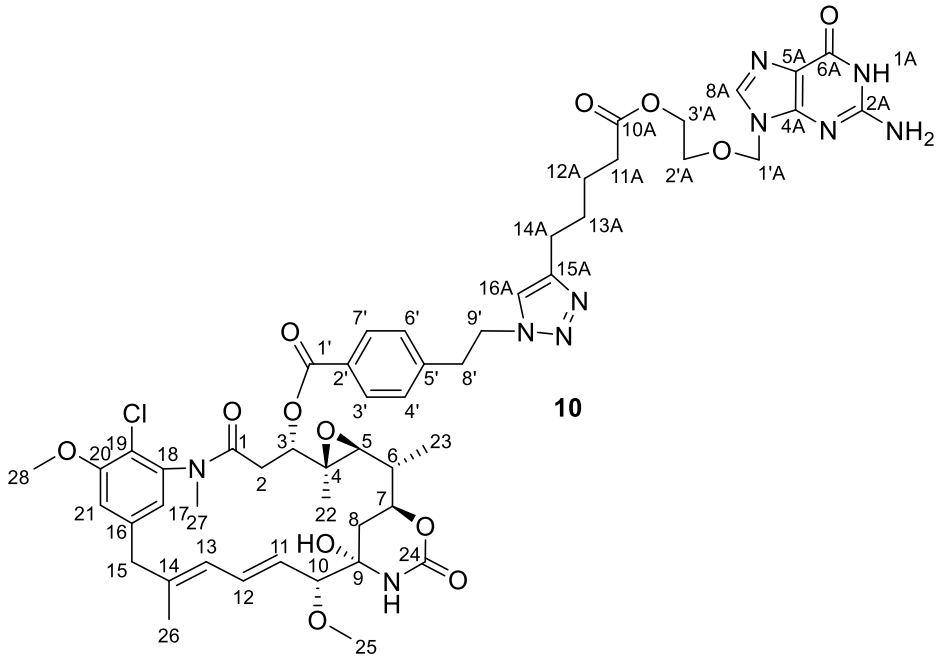
9

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.87 (d, J = 4.9 Hz, 1H,7), 7.80 (s, 1H, 8A), 7.23 (d, J = 1.7 Hz, 1H, 21), 6.94 – 6.87 (bs, 1H, NH), 6.77 (d, J = 1.7 Hz, 1H, 17), 6.65 (s, 2H, NH_2), 6.58 (dd, J = 15.4, 11.1 Hz, 1H, 12), 6.22 (d, J = 11.1 Hz, 1H, 13), 5.98 (s, 1H, 1A, NH), 5.45 (dd, J = 15.4, 8.9 Hz, 1H, 11), 5.35 (s, 2H, 1'A), 4.64 (dd, J = 11.9, 2.6 Hz, 1H, 3), 4.25 (t, J = 7.1 Hz, 2H, 3'A), 4.13 – 4.02 (m, 4H, 2'A, 15A'), 3.95 (s, 3H, 28), 3.70 – 3.64 (m, 2H, 7, 10), 3.51 (d, J = 9.3 Hz, 1H, 15a), 3.31-3.23 (m, 4H, 15b, 25), 2.95 (s, 3H, 27), 2.67 – 2.64 (m, 2H, 5'), 2.61 (d, J = 9.7 Hz, 1H, 5), 2.44 – 2.38 (m, 1H, 2b), 2.22 (t, J = 7.2 Hz, 2H, 11A'), 2.15 – 2.10 (m, 2H, 2'), 2.01 (d, J = 12.9 Hz, 1H, 2a), 1.79 – 1.69 (m, 4H, 3', 14A'), 1.65-1.60 (m, 5H, 26, 4'), 1.56 – 1.43 (m, 4H, 6, 12A', 8b), 1.39 – 1.33 (m, 1H, 8a), 1.22-1.18 (m, 2H, 13A'), 1.12 (d, J = 6.3 Hz, 3H, 23), 0.82 (s, 3H, 22).

^{13}C NMR (101 MHz, DMSO-d6) δ 172.6 (10A), 171.4 (1'), 168.0 (1), 156.7 (6A), 155.2 (20), 154.0 (2A), 151.5 (4A), 151.2 (24), 146.4 (6'), 141.4 (18), 141.0 (16), 139.0 (14), 131.9 (12), 128.7 (11), 124.7 (13), 121.8 (17), 121.7 (7'), 117.4 (19), 116.5 (5A), 113.9 (21), 88.1 (10), 80.2 (9), 76.1 (3), 73.4 (7), 71.8 (1A'), 66.5 (2'A), 65.7 (5), 62.6 (3'A), 60.6 (4), 56.6 (28), 56.2 (25), 50.5 (15A), 48.9 (15), 35.8 (6), 35.0 (27), 33.3 (8), 33.1 (2'), 29.4 (11A), 28.3 (2), 25.3 (14A), 25.3 (4'), 24.9 (5'), 24.5 (13A), 24.1 (12A), 23.7 (3'), 15.3 (26), 14.4 (23), 11.4 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1059.4302 (calcd for $C_{49}H_{65}ClN_{10}O_{13}Na$ 1059.4319)

Compound 10



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.83 (m, 4H, 8A, 3’, 7’, 16A), 7.46 (d, J = 8.1 Hz, 2H, 4’, 6’), 7.26 (d, J = 1.8 Hz, 1H, 21), 6.93 (d, J = 1.8 Hz, 1H, 17), 6.80 (s, 1H, NH), 6.55 (s, 2H, NH_2), 6.48 (dd, J = 15.3, 11.0 Hz, 1H, 12), 5.93 (d, J = 11.0 Hz, 1H, 13), 5.53 (s, 1H, 1A), 5.35 (s, 2H, 1’A), 4.94 (dd, J = 15.3, 9.0 Hz, 1H, 11), 4.75 (dd, J = 12.0, 3.0 Hz, 1H, 3), 4.64 (t, J = 7.2 Hz, 2H, 9’), 4.17 – 4.06 (m, 3H, 3’A, 7), 3.96 (s, 3H, 28), 3.70 – 3.63 (m, 2H, 2’A), 3.45 (d, J = 12.5 Hz, 1H, 15a), 3.41 – 3.22 (m, 4H, 10, 8’, 15b), 3.17 (s, 3H, 25), 2.99 (s, 3H, 27), 2.83 (d, J = 9.6 Hz, 1H, 5), 2.67 – 2.52 (m, 3H, 2a, 14A), 2.31 – 2.2 (m, 3H, 2b, 11A), 1.64 (s, 3H, 26), 1.56 – 1.41 (m, 5H, 8b, 13A, 12A), 1.38 – 1.18 (m, 2H, 8a, 6), 1.13 (d, J = 6.4 Hz, 3H, 23), 0.83 (s, 3H, 22).

^{13}C NMR (101 MHz, DMSO-d6) 13C δ 172.9 (10A), 168.2 (1), 165.4 (1’), 157.2 (6A), 155.7 (20), 154.2 (2A), 151.6 (4A), 151.4 (24), 147.0 (15A), 143.8 (5’), 141.7 (18), 141.4 (16), 139.3 (14), 135.0 (8A), 132.0 (12), 129.7 (3’,7’), 129.0 (4’,6’), 128.8 (11), 128.2 (2’), 124.7 (13), 122.0 (17), 121.7 (16A), 117.7 (19), 116.7 (5A), 114.0 (21), 88.7 (10), 80.2 (9), 76.5 (3), 73.5 (7), 72.0 (1’A), 66.7 (2’A), 65.5 (5), 62.8 (3’A), 60.9 (4), 56.9 (28), 56.3 (25), 50.0 (9’) , 45.9 (15), 38.2 (6), 36.5 (8) 35.9 (8’), 35.3 (27), 33.6 (2), 33.2 (11A), 28.4 (12A), 24.8 (14A), 24.1 (13A), 15.6 (26), 14.8 (23), 12.3 (22).

HRMS (ESI) m/z $[M+Na]^+$ 1093.4210 (calcd for $C_{52}H_{63}ClN_{10}O_{13}Na$ 1093.4162)

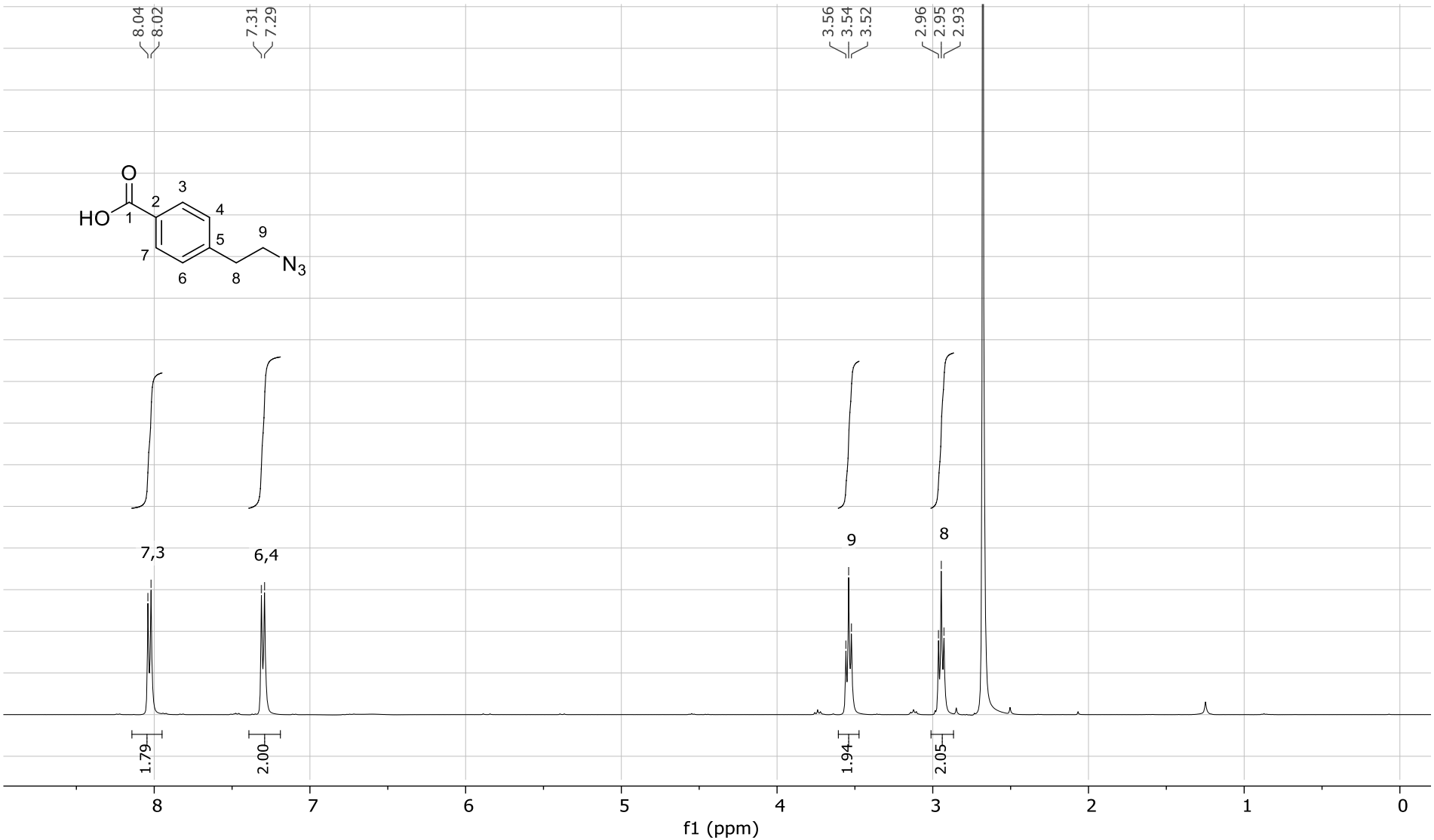


Figure S1. ^{1}H NMR spectrum (400 MHz, chloroform-d) of 4-(2-azidoethyl)benzoic acid**.**

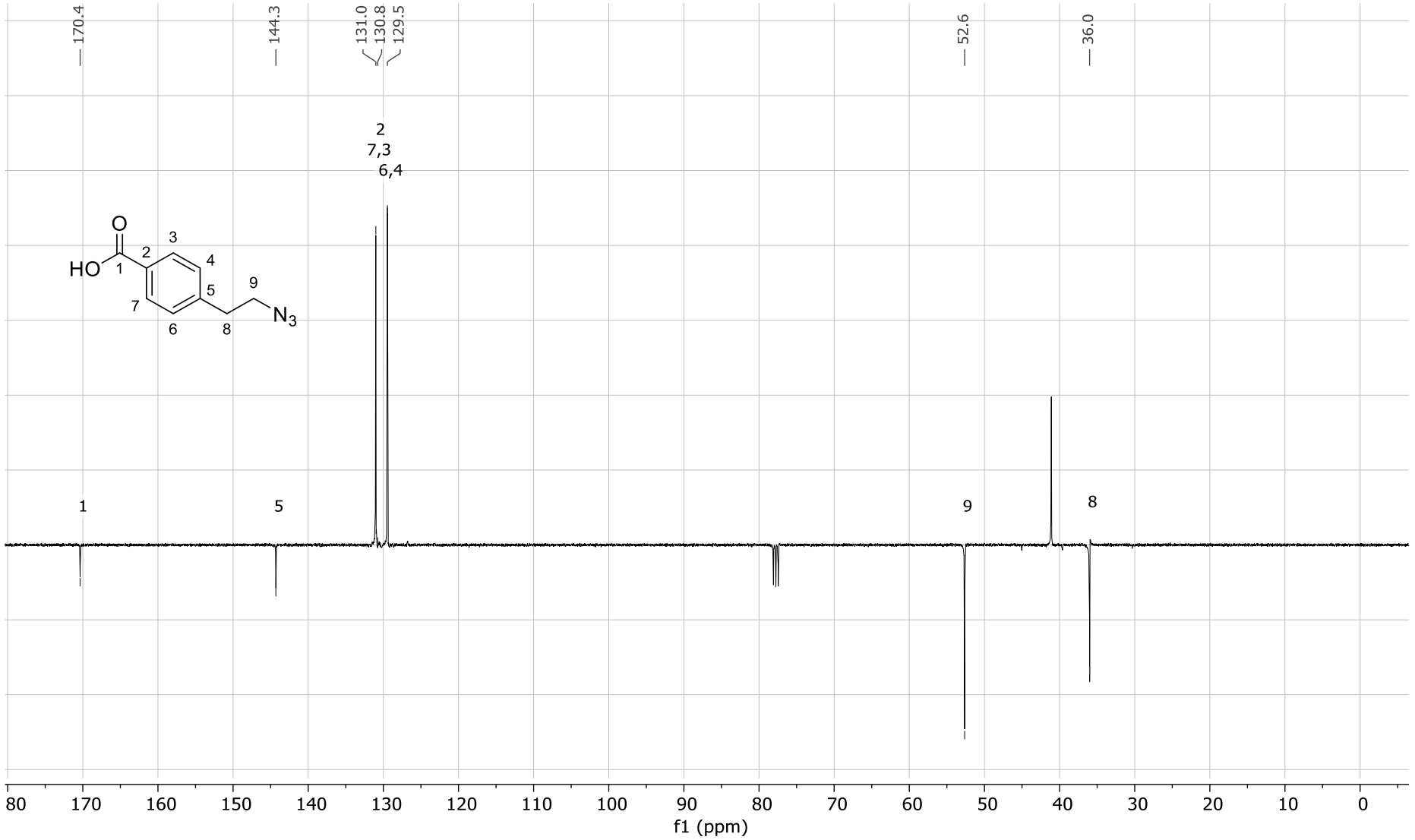


Figure S2. ^{13}C NMR APT spectrum (101 MHz, chloroform-d) of 4-(2-azidoethyl)benzoic acid.

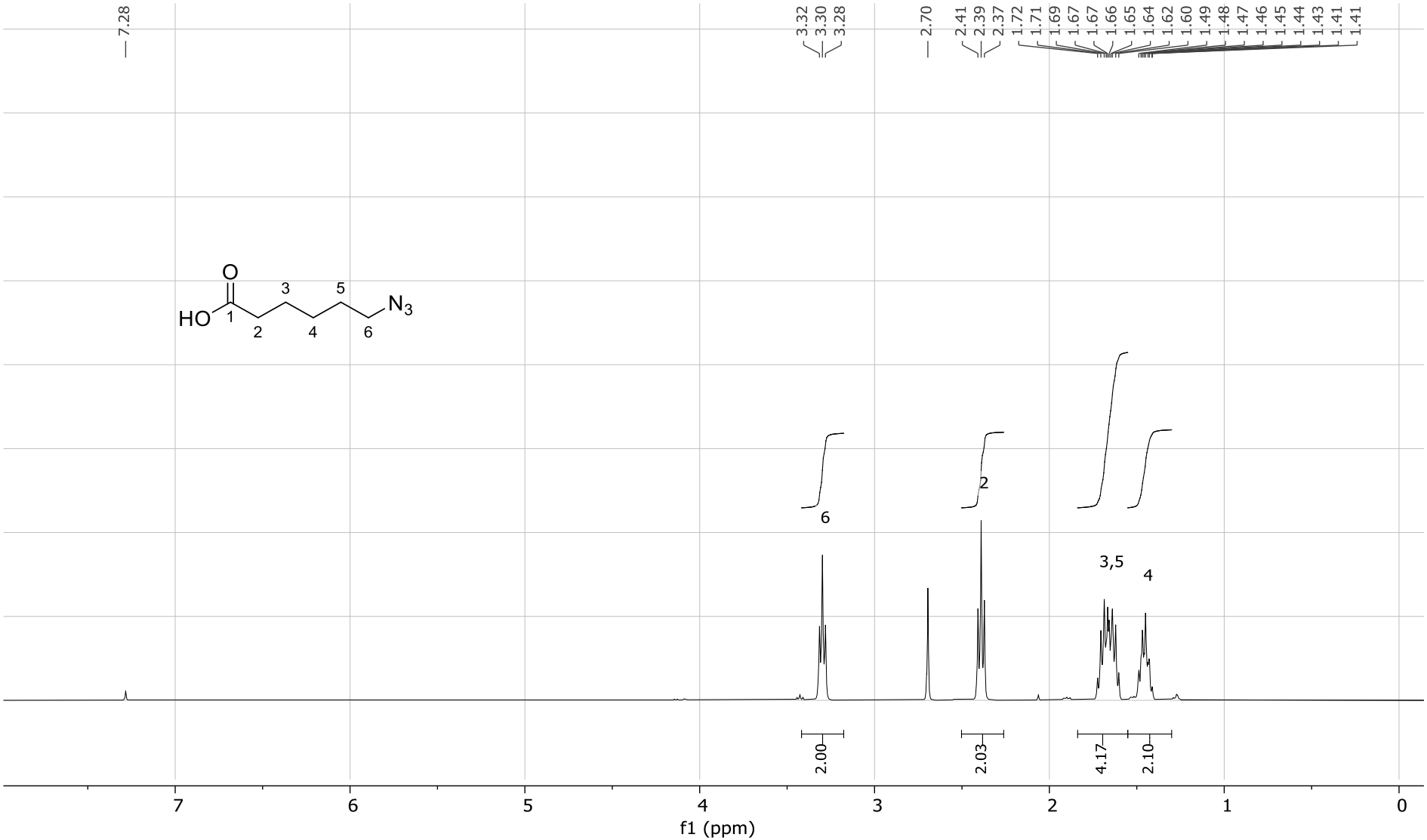


Figure S3. ^{1}H NMR spectrum (400 MHz, chloroform-d) of 6-azidohexanoic acid.

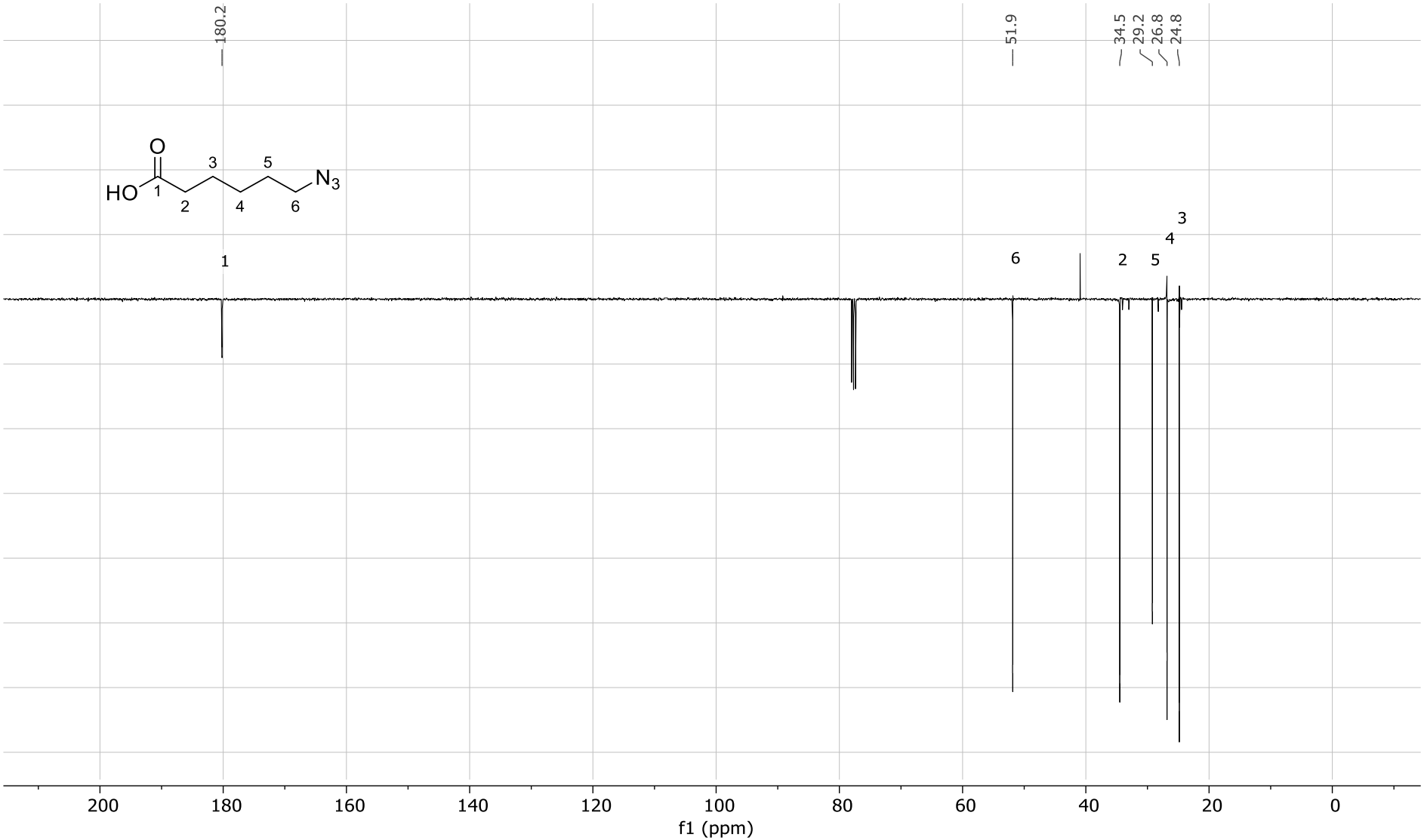


Figure S4. ^{13}C NMR APT spectrum (101 MHz, chloroform-d) of 6-azidohexanoic acid.

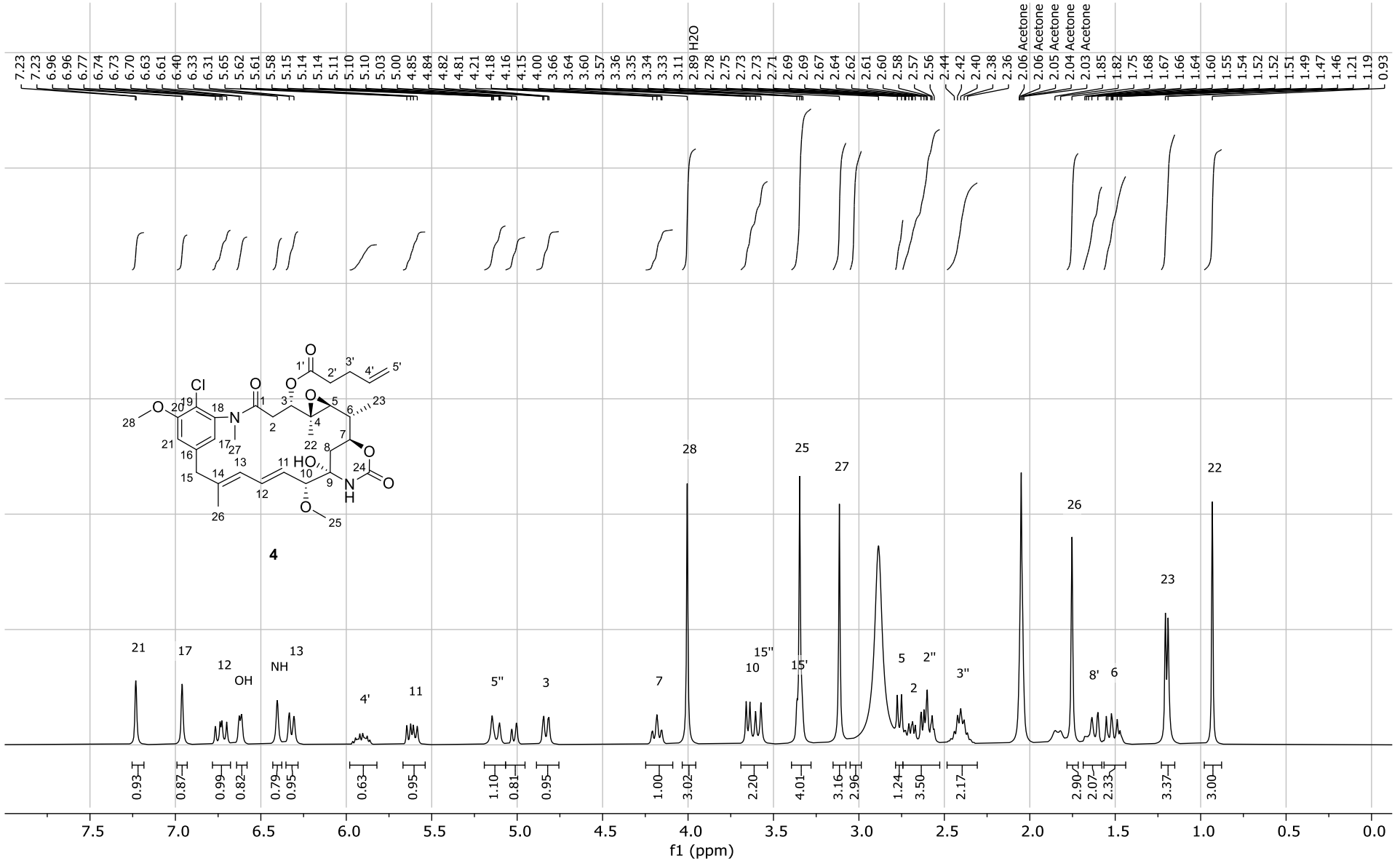


Figure S5. ^{1}H NMR spectrum (400 MHz, acetone-d6) of **4.**

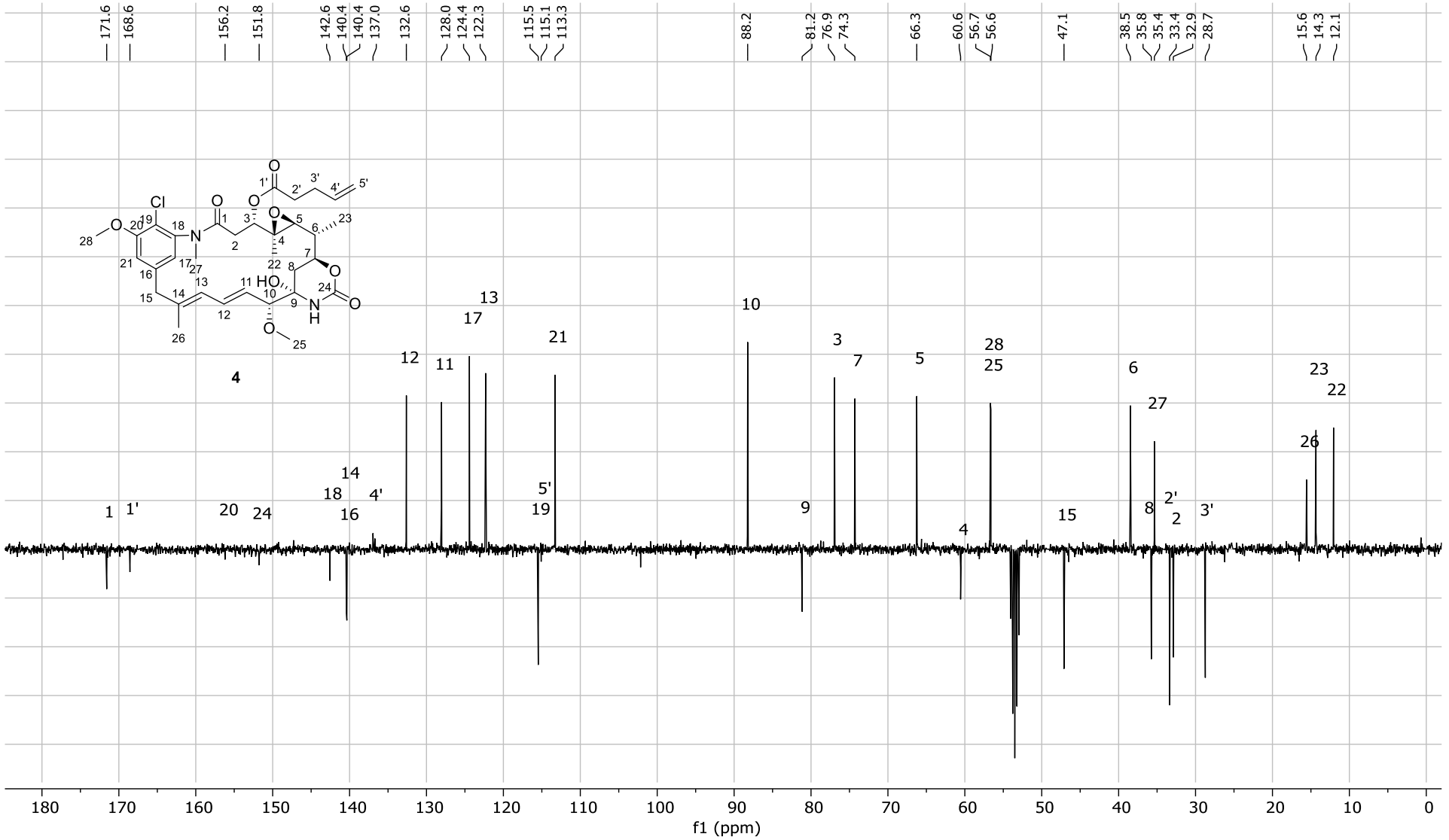


Figure S6. ^{13}C NMR APT spectrum (101 MHz, dichloromethane-d2) of **4**.

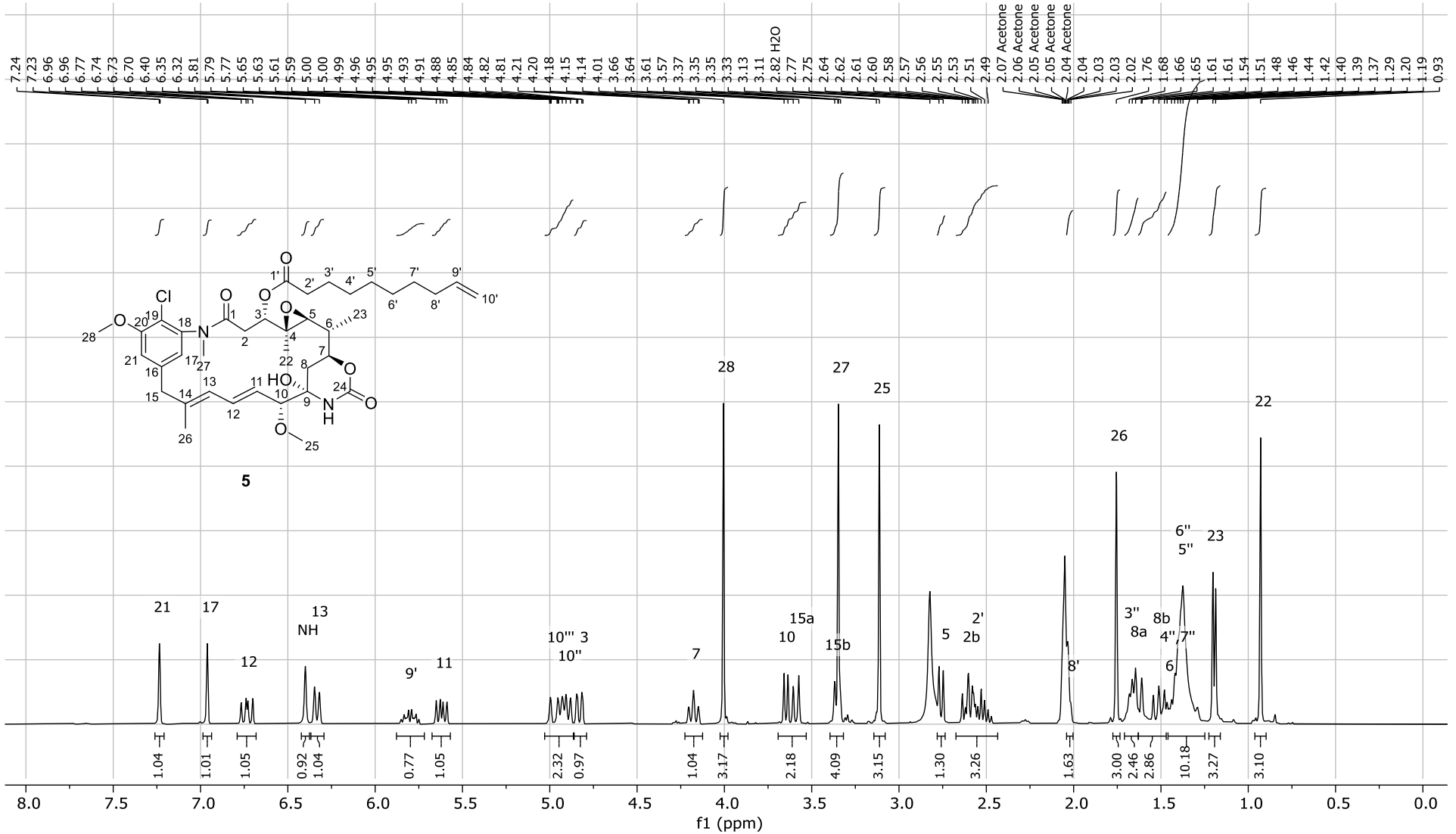


Figure S7. ^{1}H NMR spectrum (400 MHz, acetone-d6) of **5.**

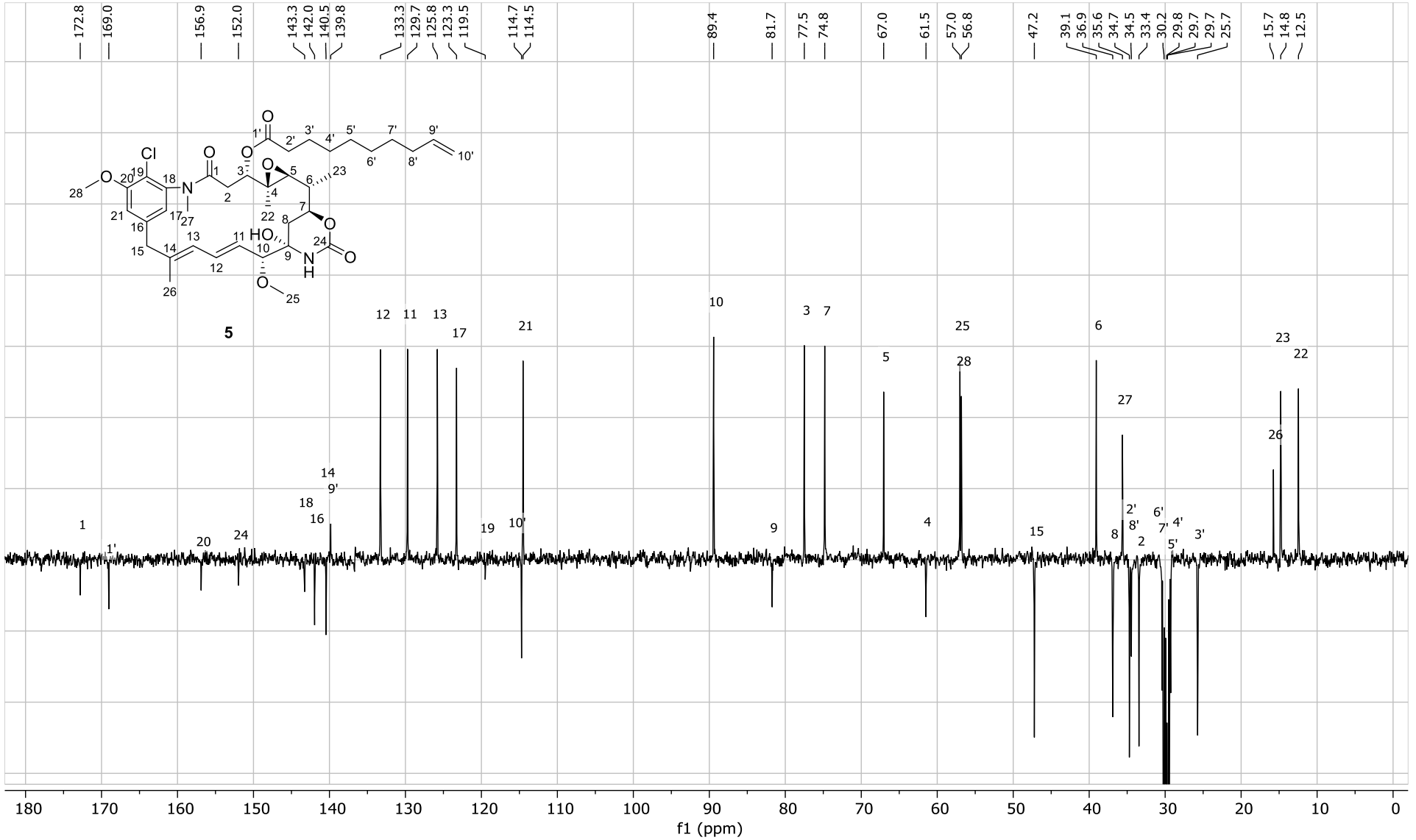


Figure S8. ^{13}C NMR APT spectrum (101 MHz, acetone-d6) of **5**.

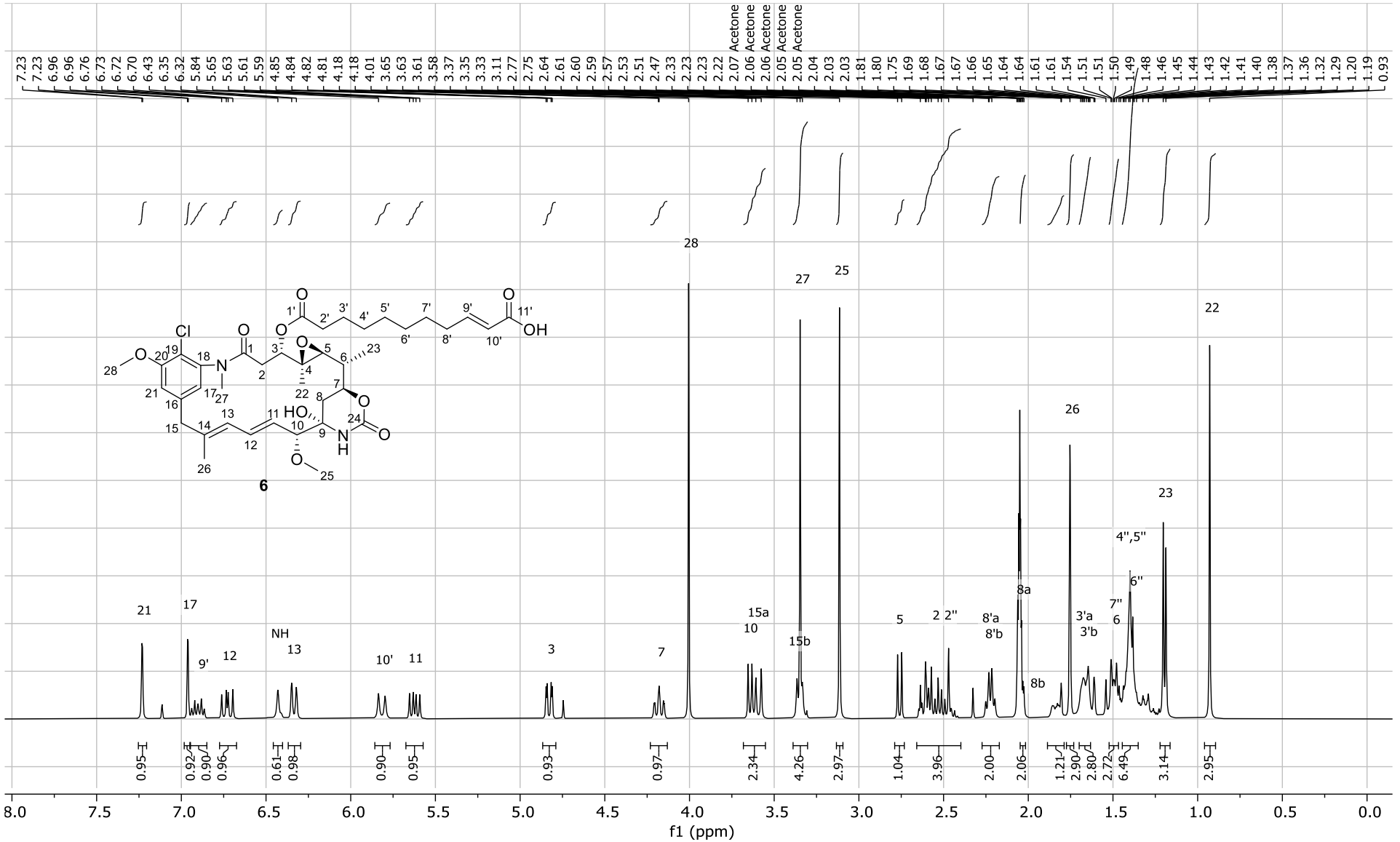


Figure S9. ^{1}H NMR spectrum (400 MHz, acetone-d6) of **6.**

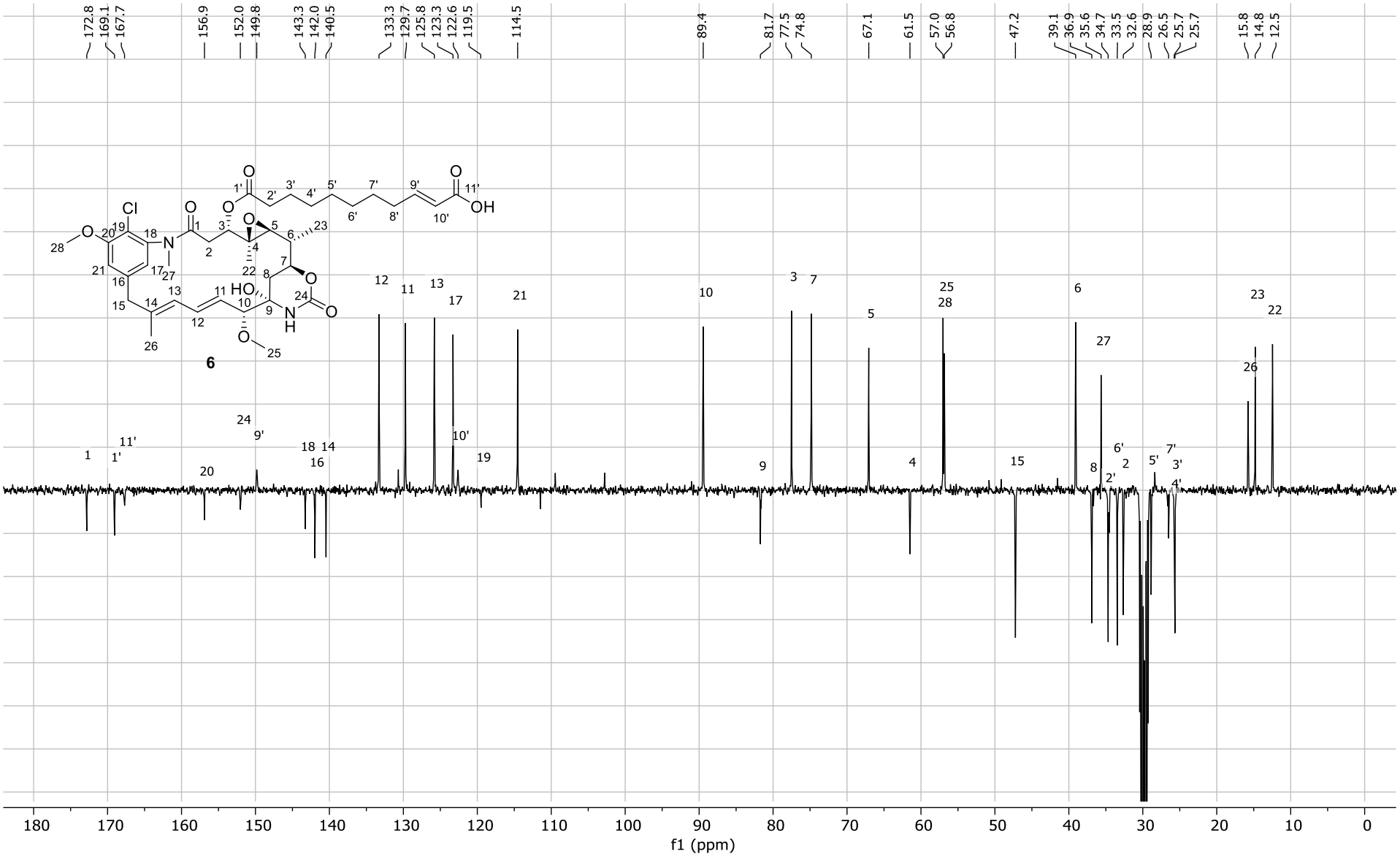


Figure S10. ^{13}C NMR APT spectrum (101 MHz, acetone-d6) of **6**.

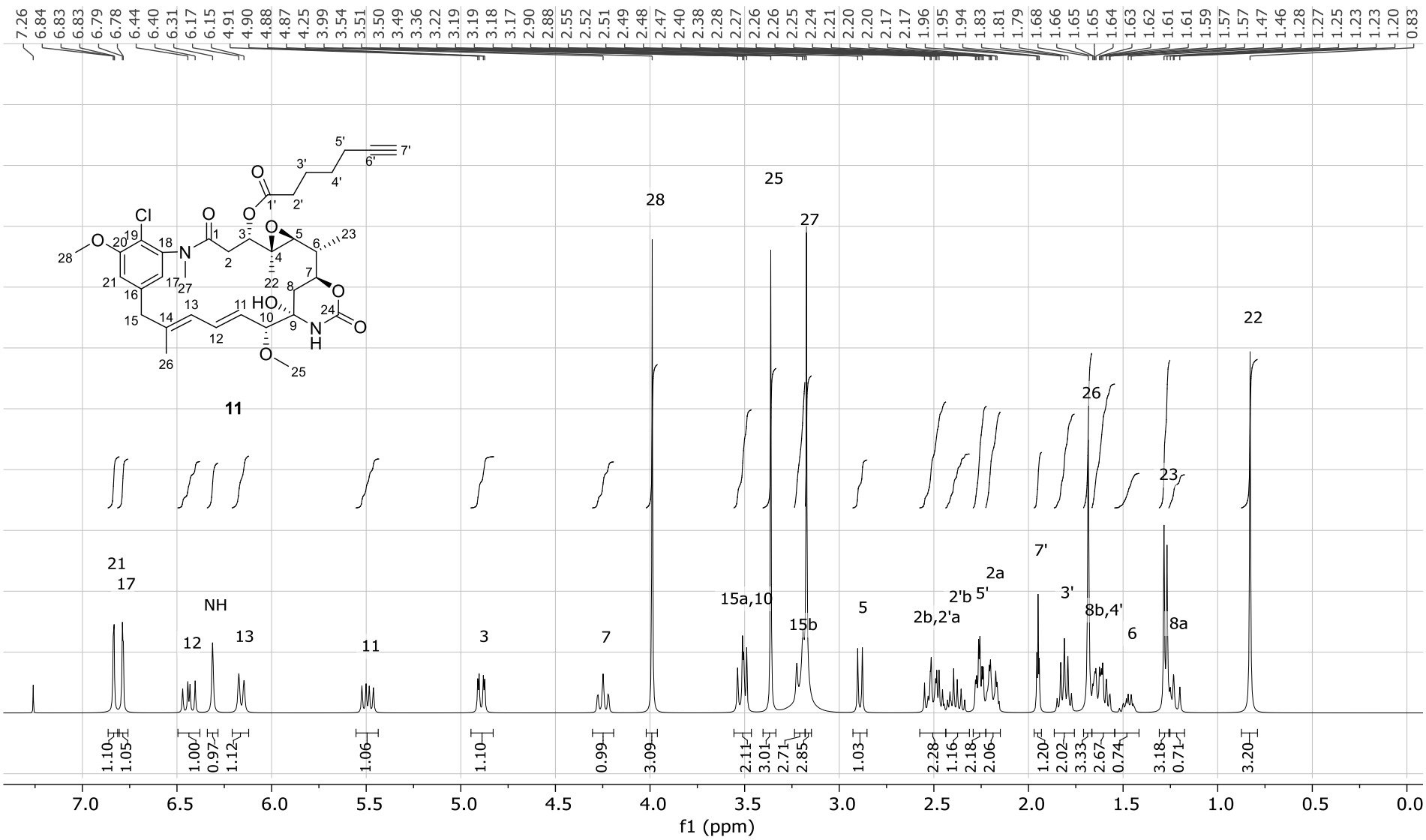


Figure S11. ^{1}H NMR spectrum (400 MHz, chloroform-d) of **11.**

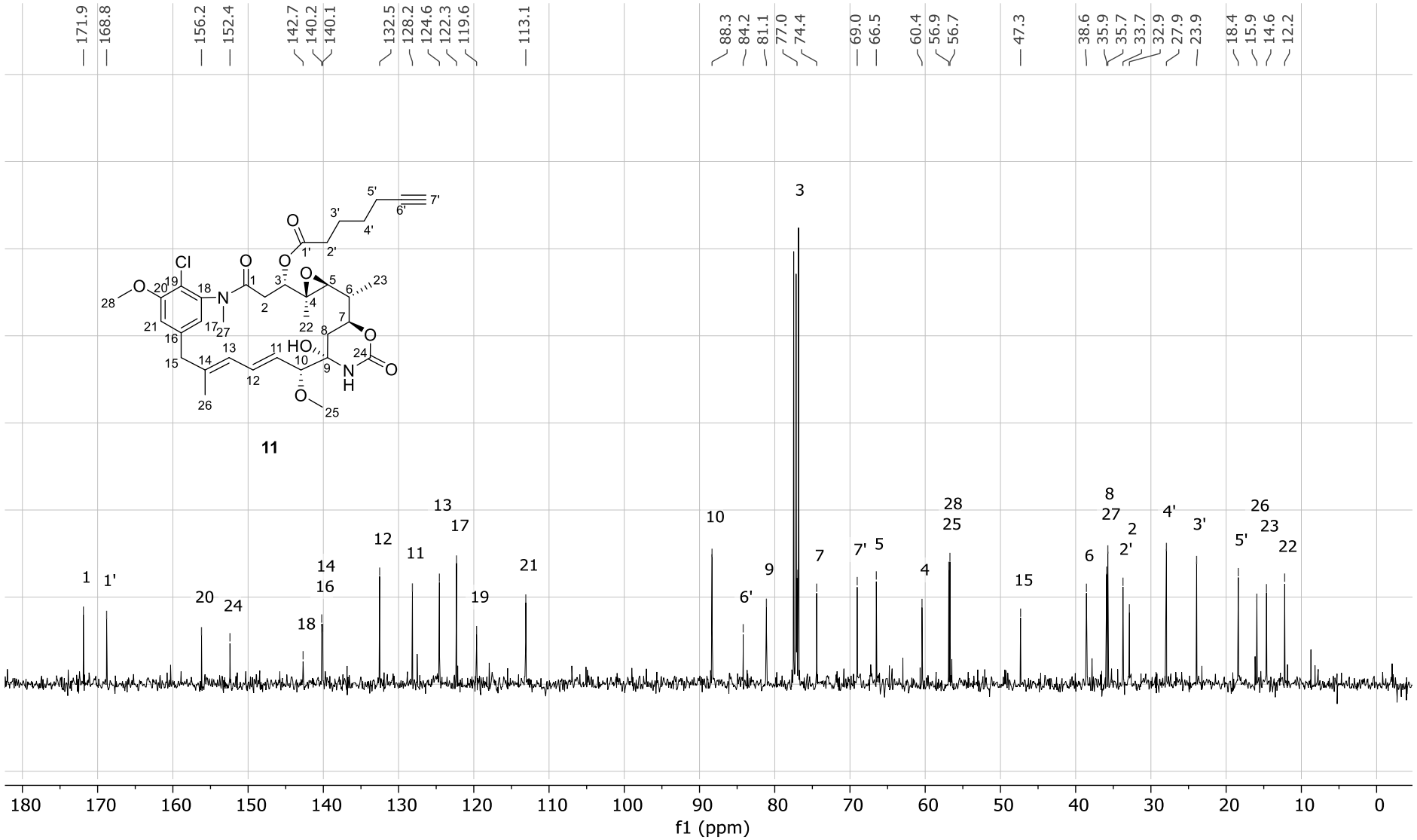


Figure S12. ^{13}C NMR spectrum (101 MHz, chloroform-d) of **11**

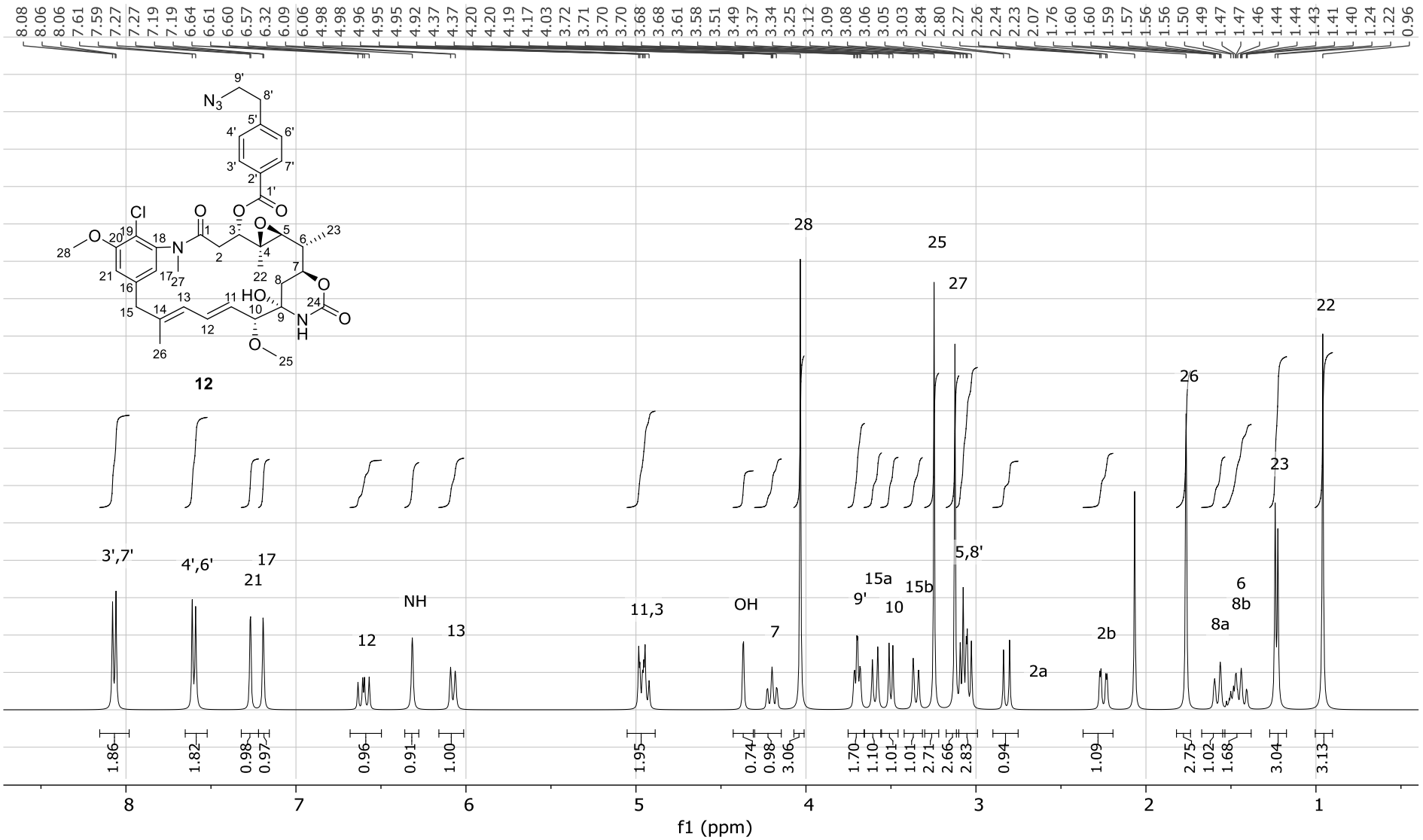


Figure S13. ^{1}H NMR spectrum (400 MHz, chloform-d) of **12**.

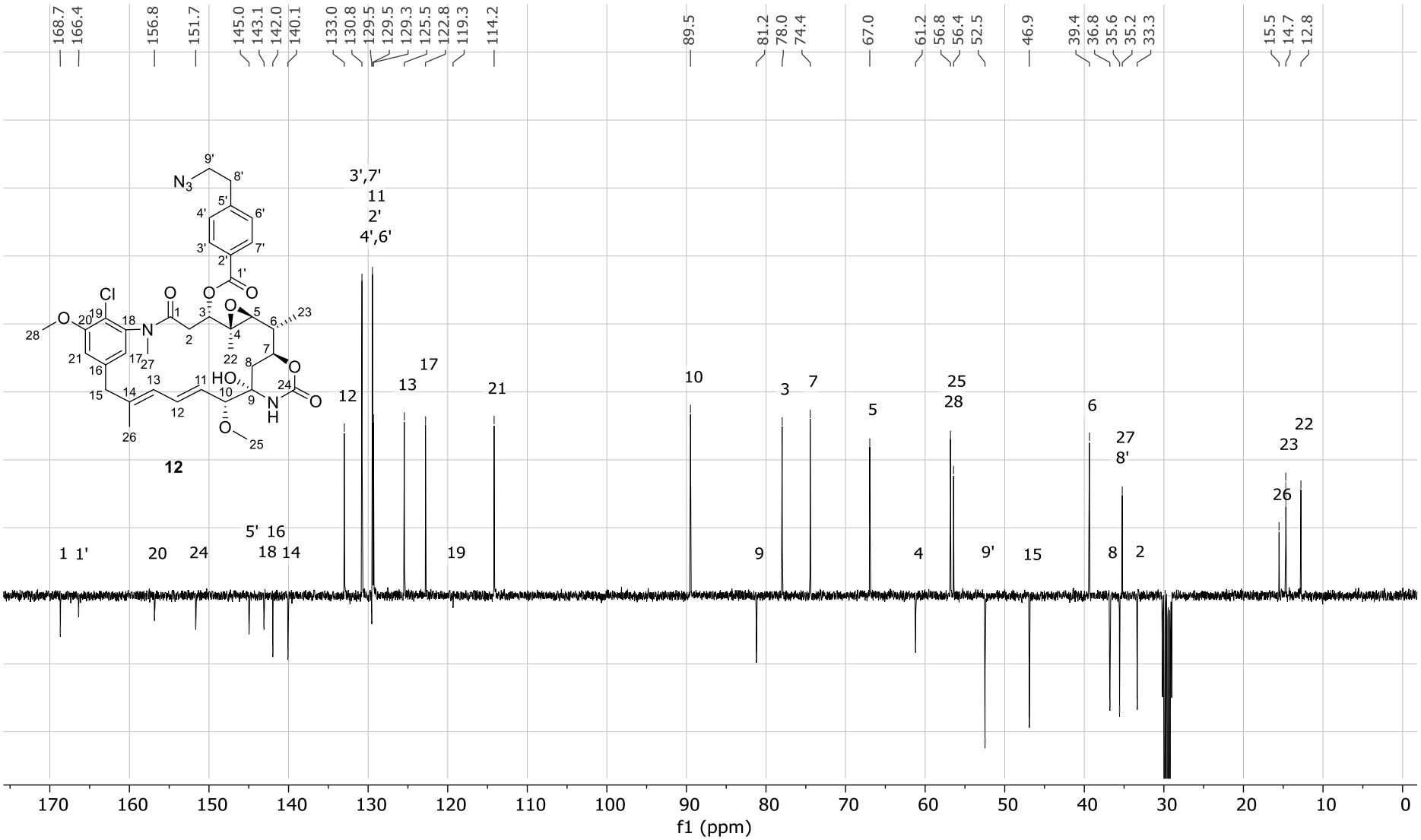


Figure S14. ^{13}C NMR spectrum (101 MHz, chlorofom-d) of **12**.

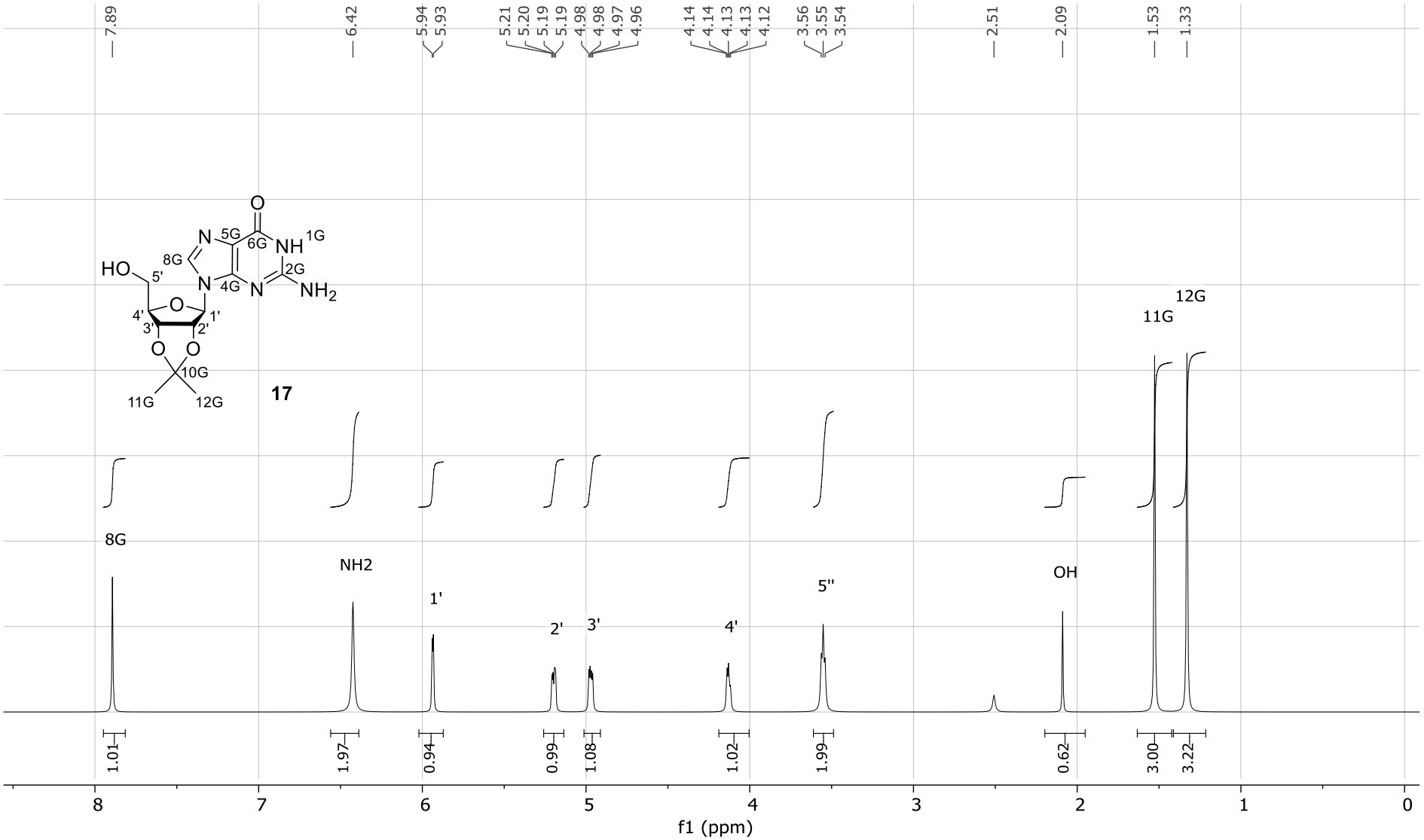


Figure S15. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **17**.

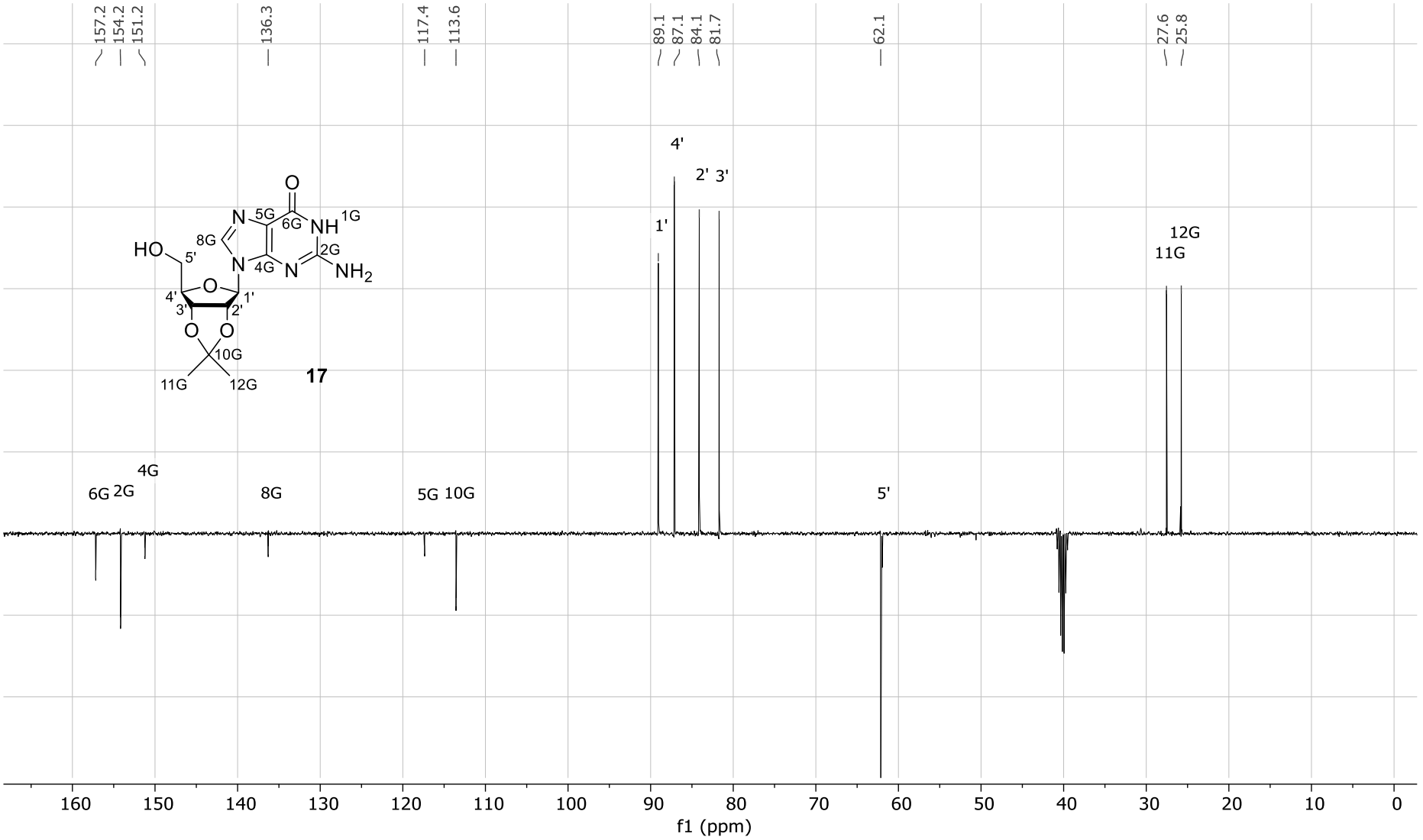


Figure S16. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **17**.

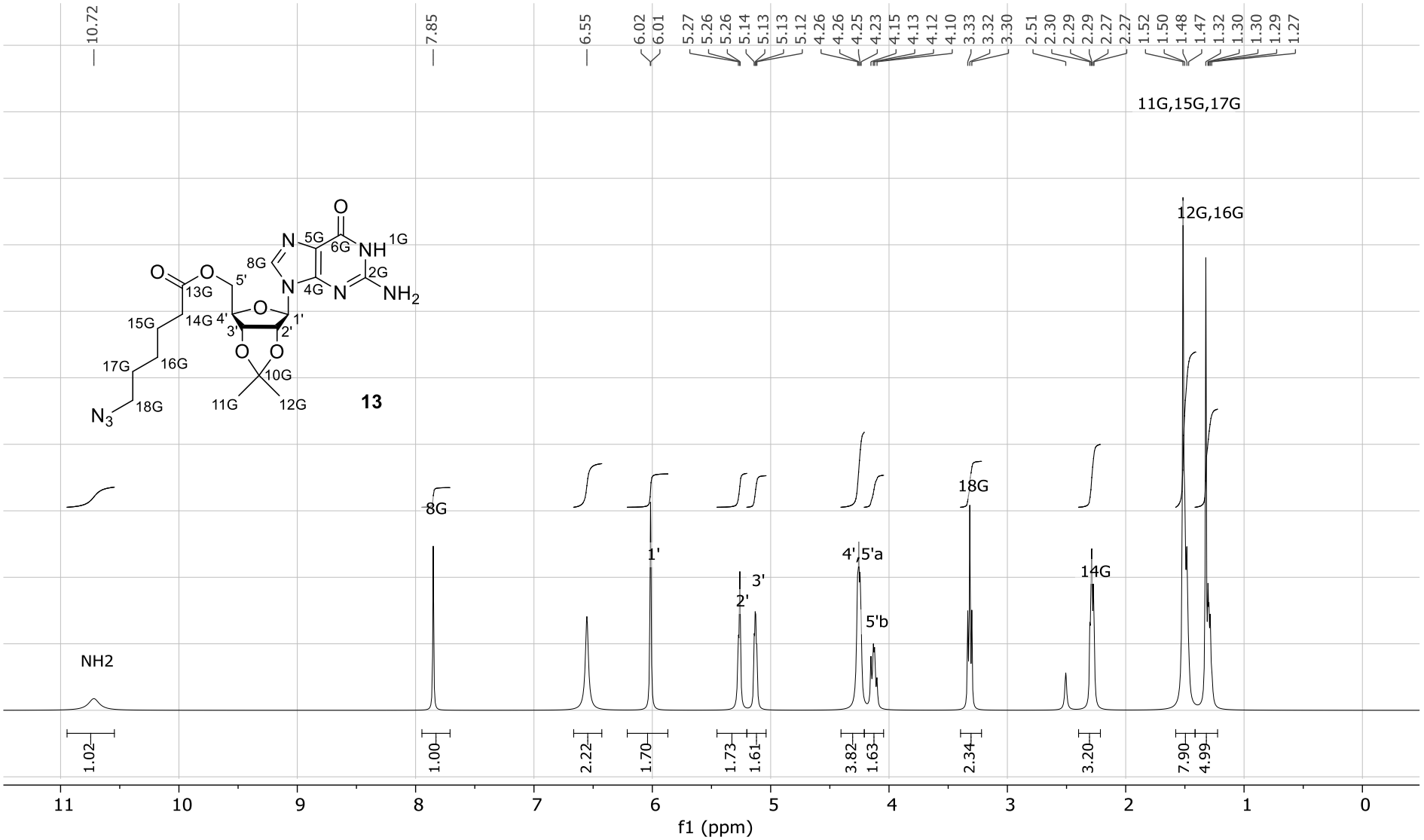


Figure S17. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **13**.

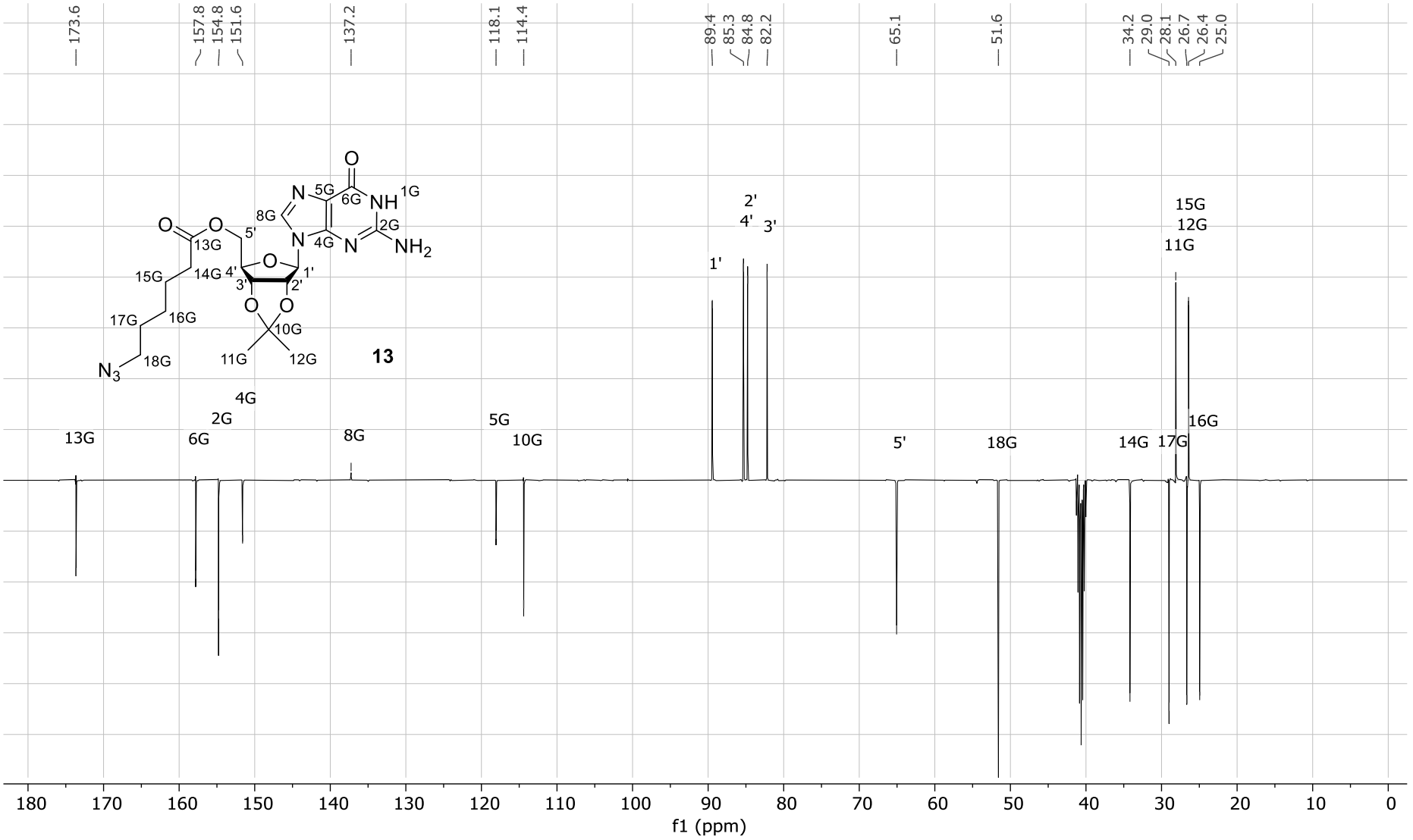


Figure S18. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **13**.

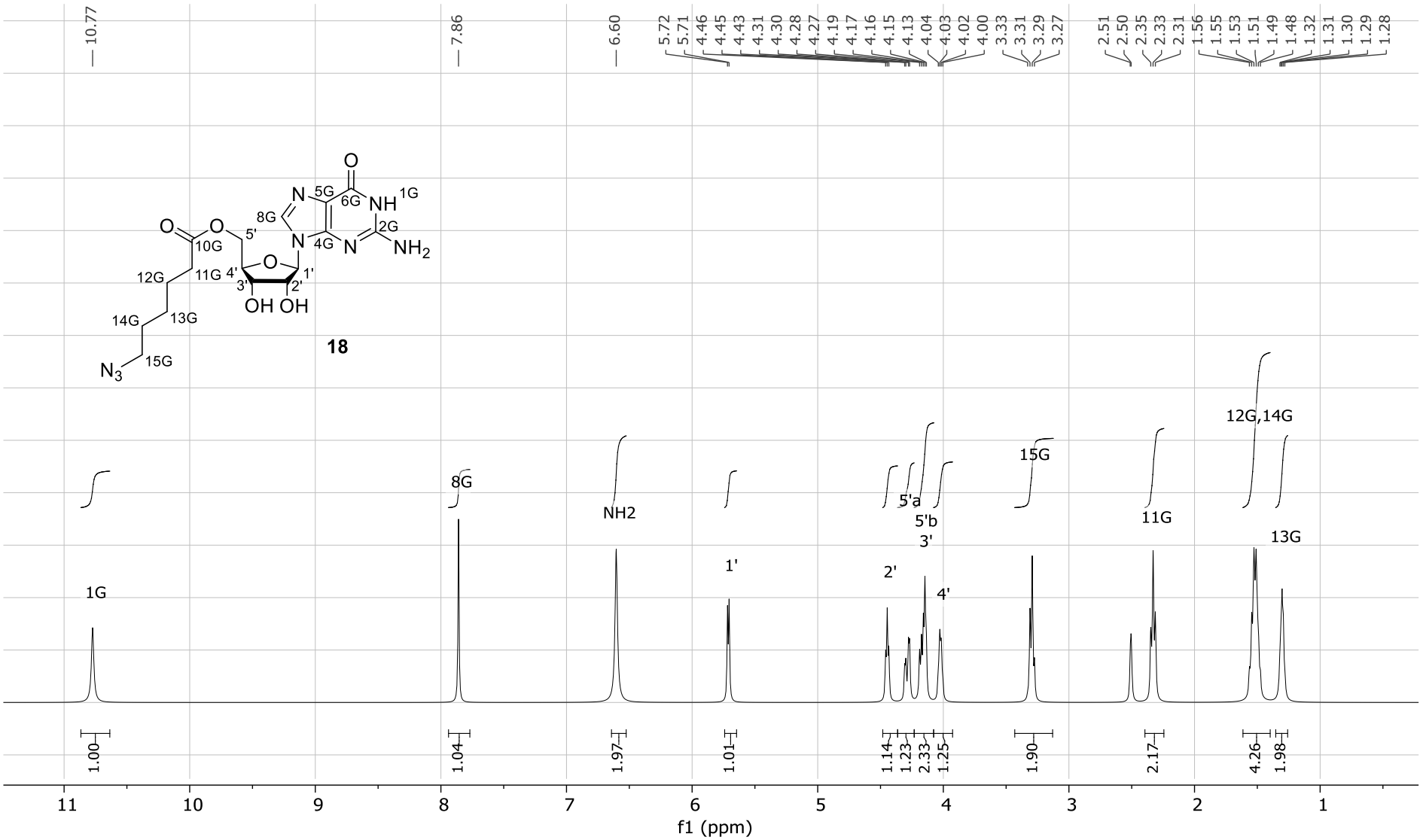


Figure S19. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **18**.

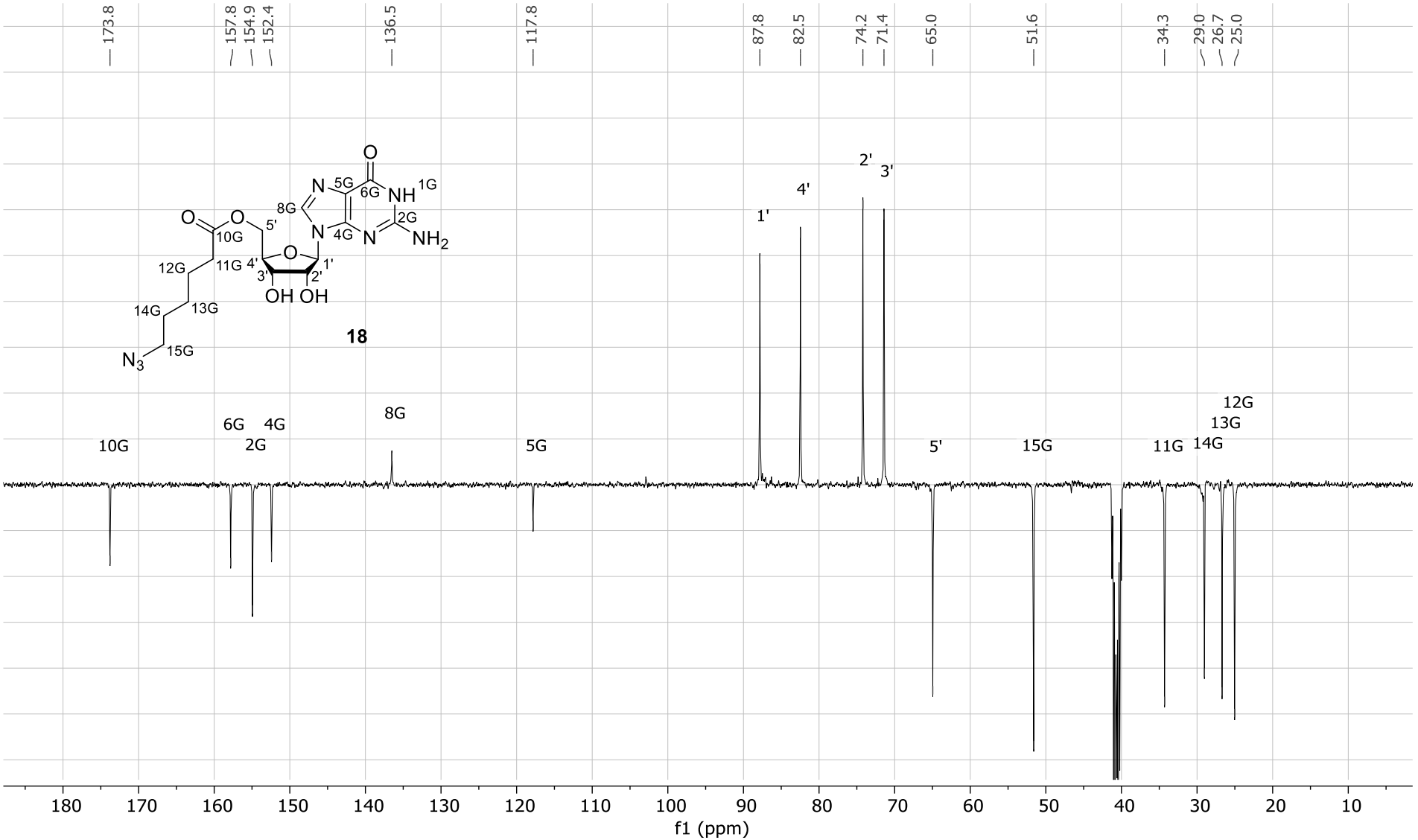


Figure S20. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **18**.

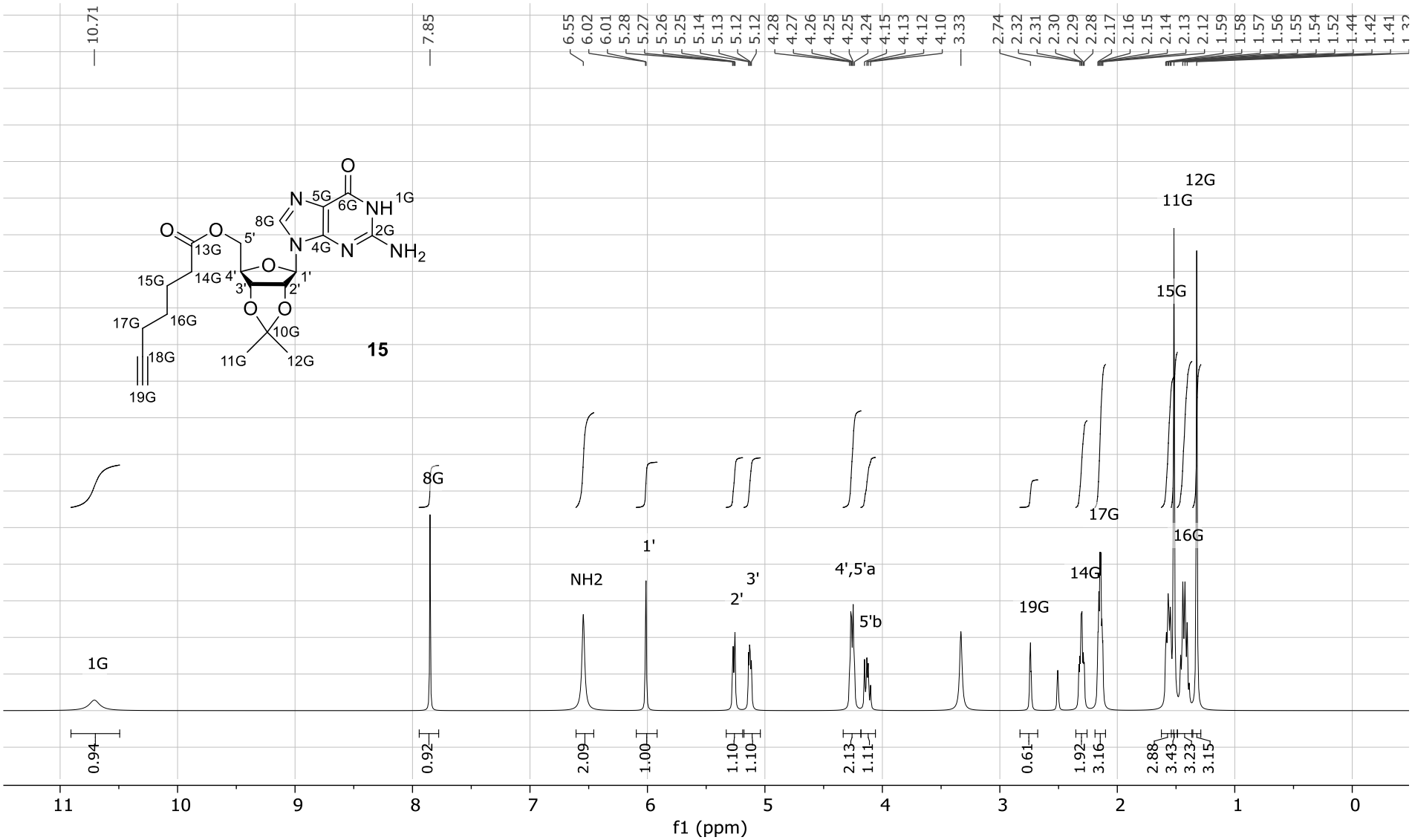


Figure S21. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **15**.

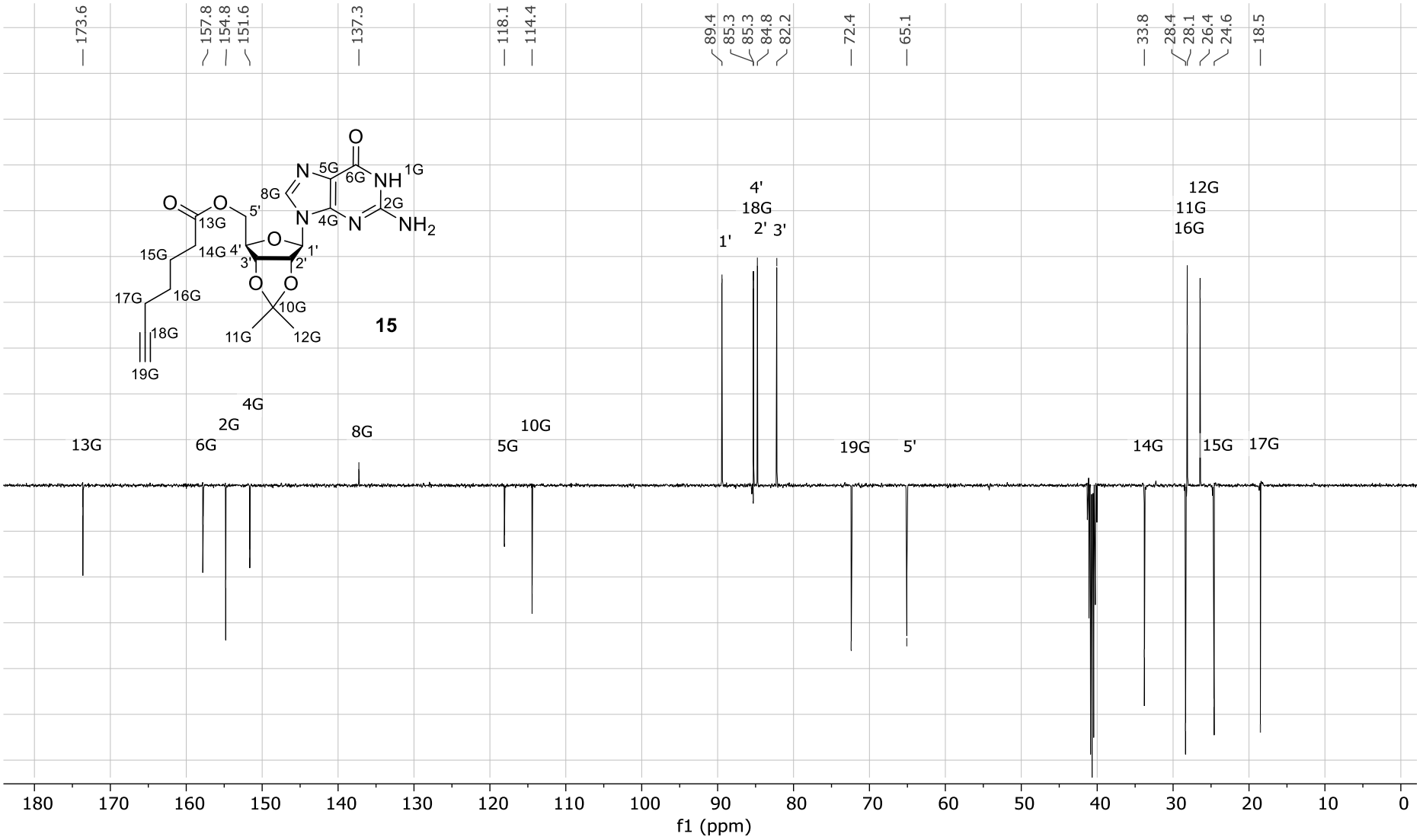


Figure S22. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **15**.

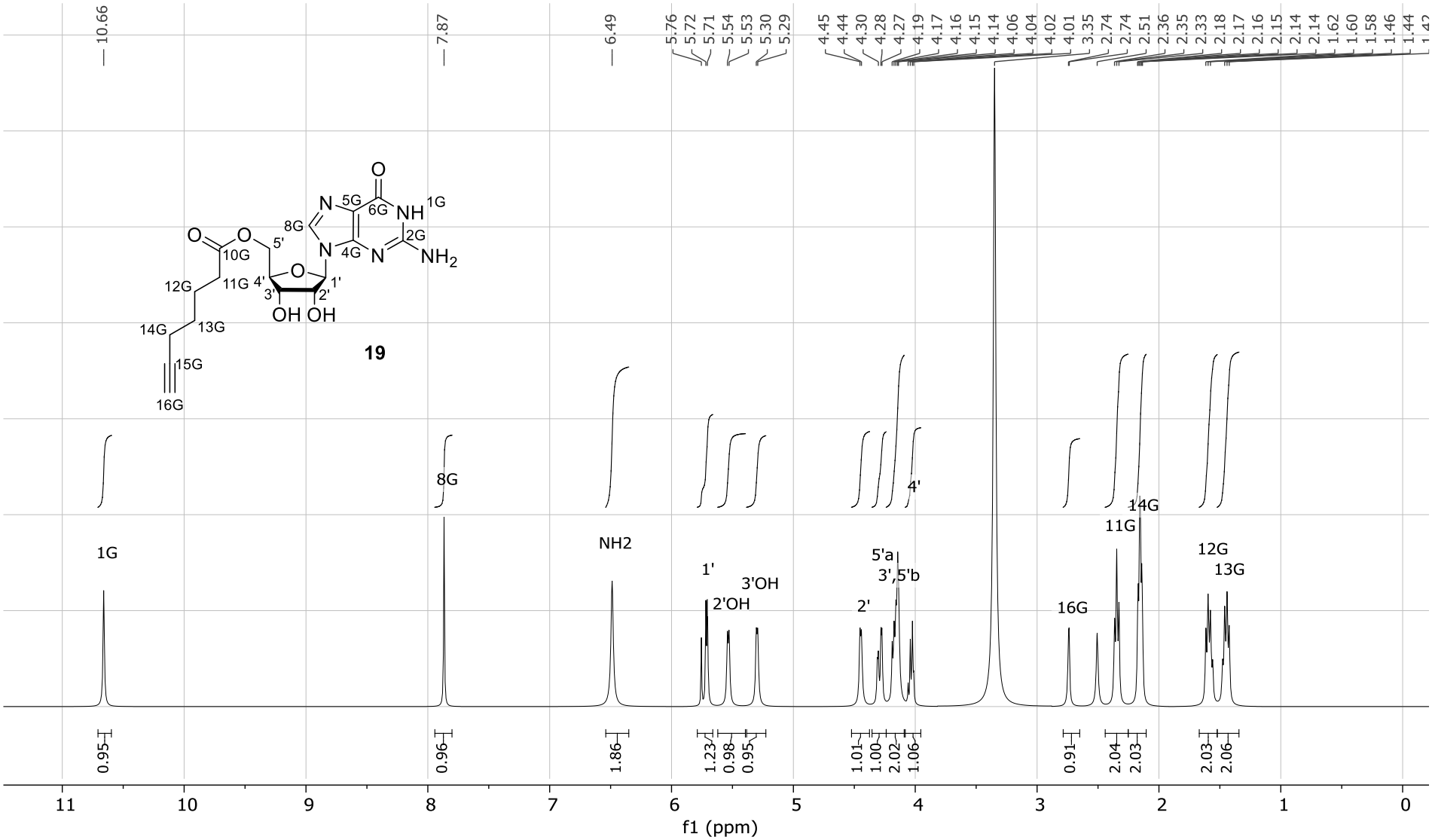


Figure S23. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **19**.

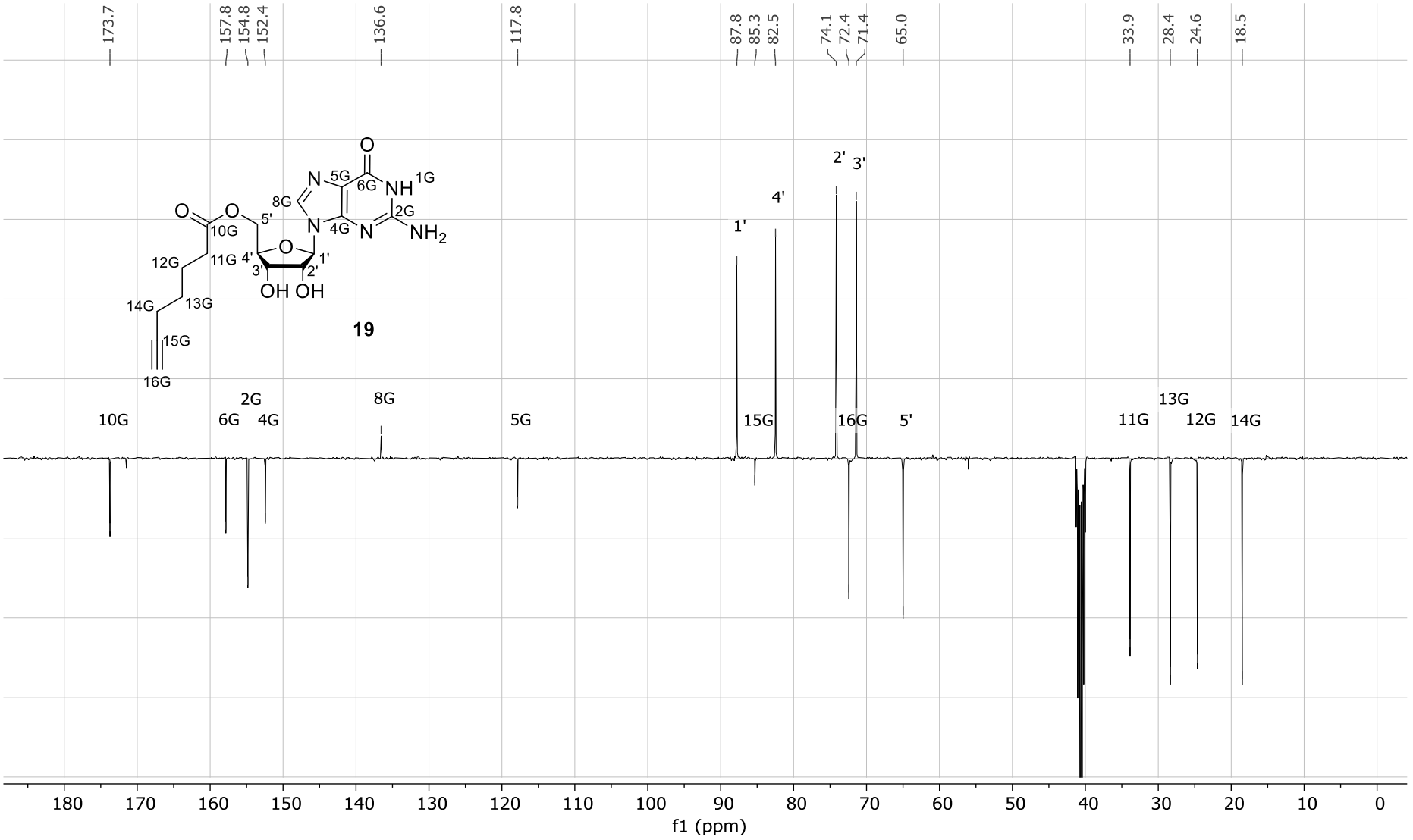


Figure S24. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **19**.

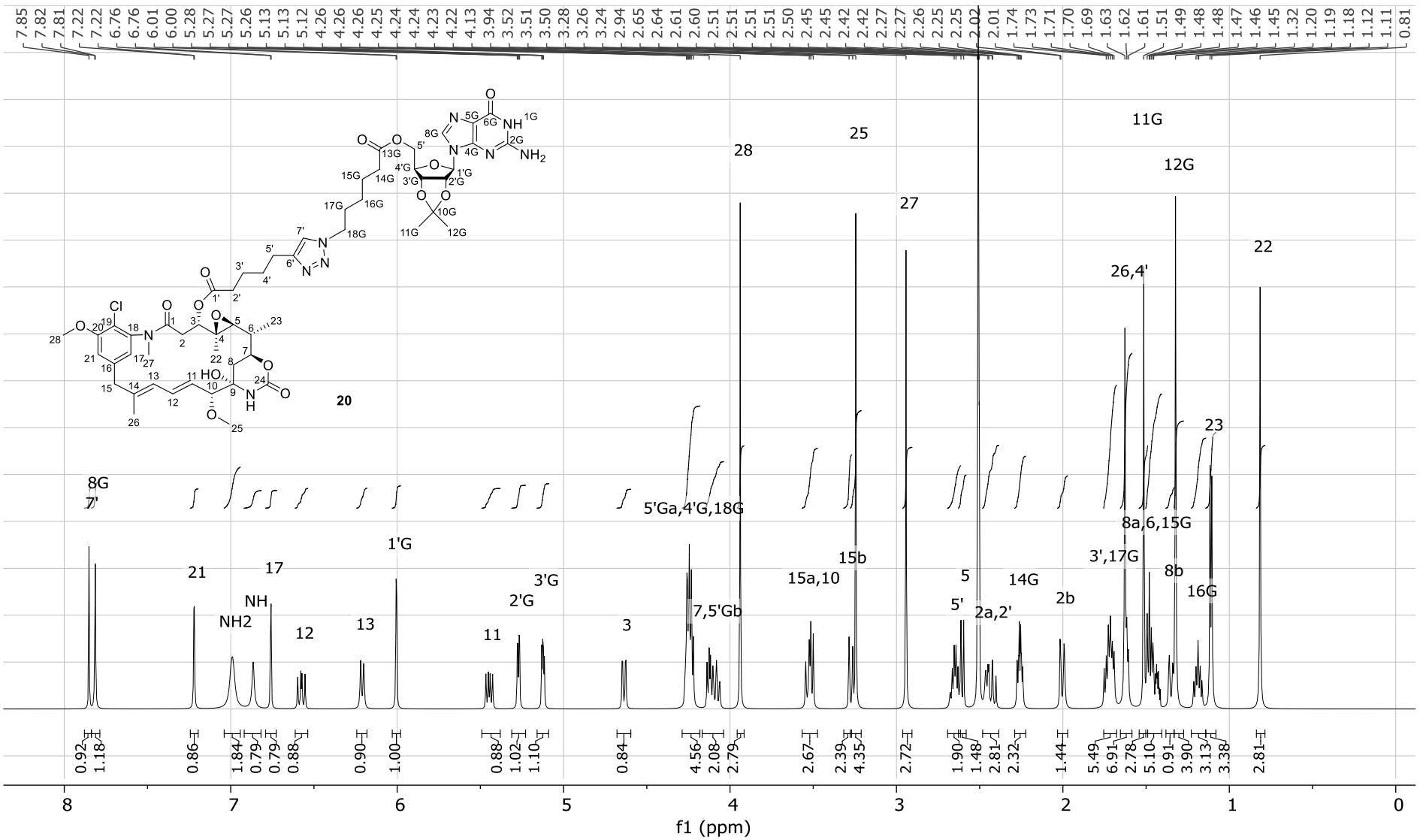


Figure S25. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **20**.

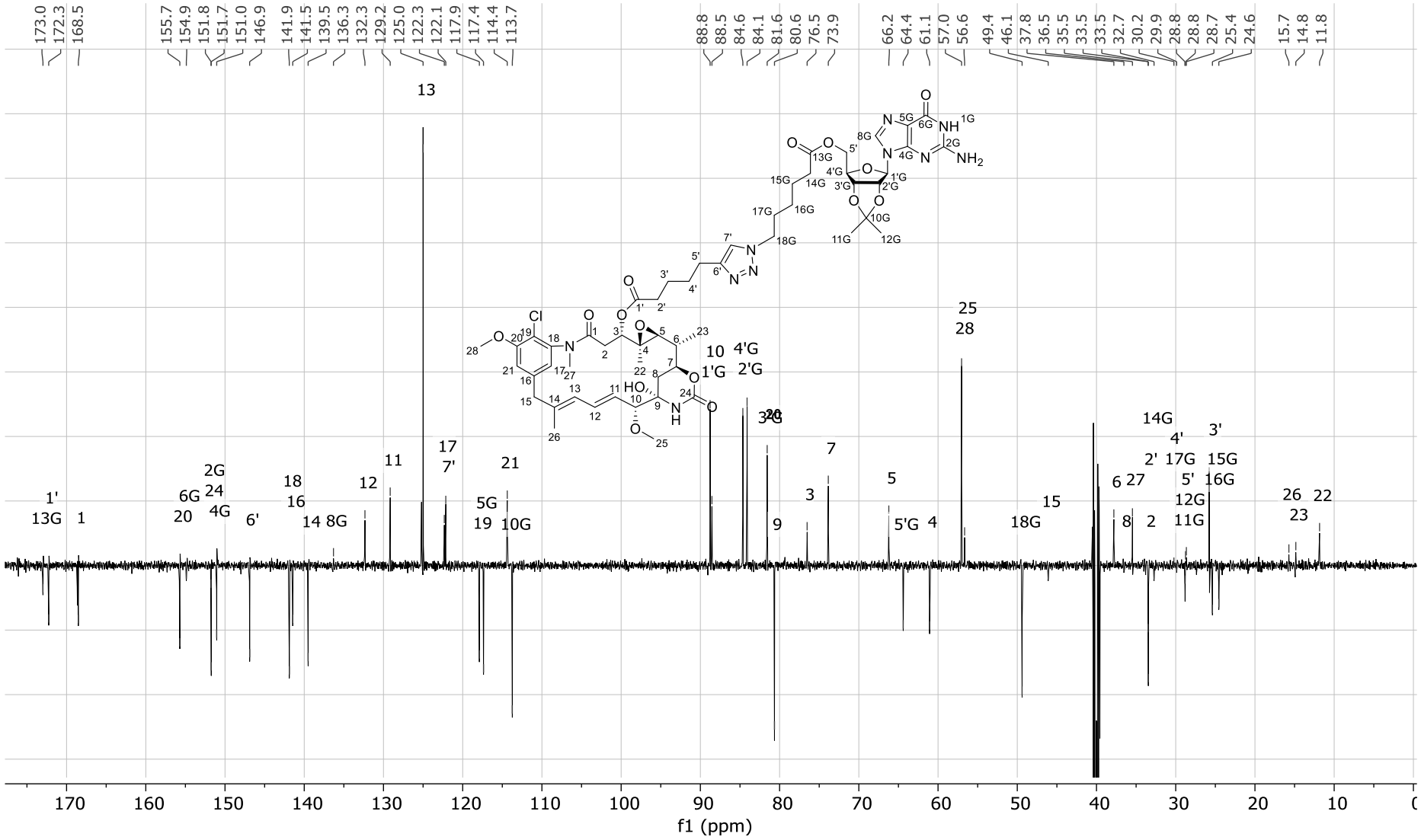


Figure S26. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **20**.

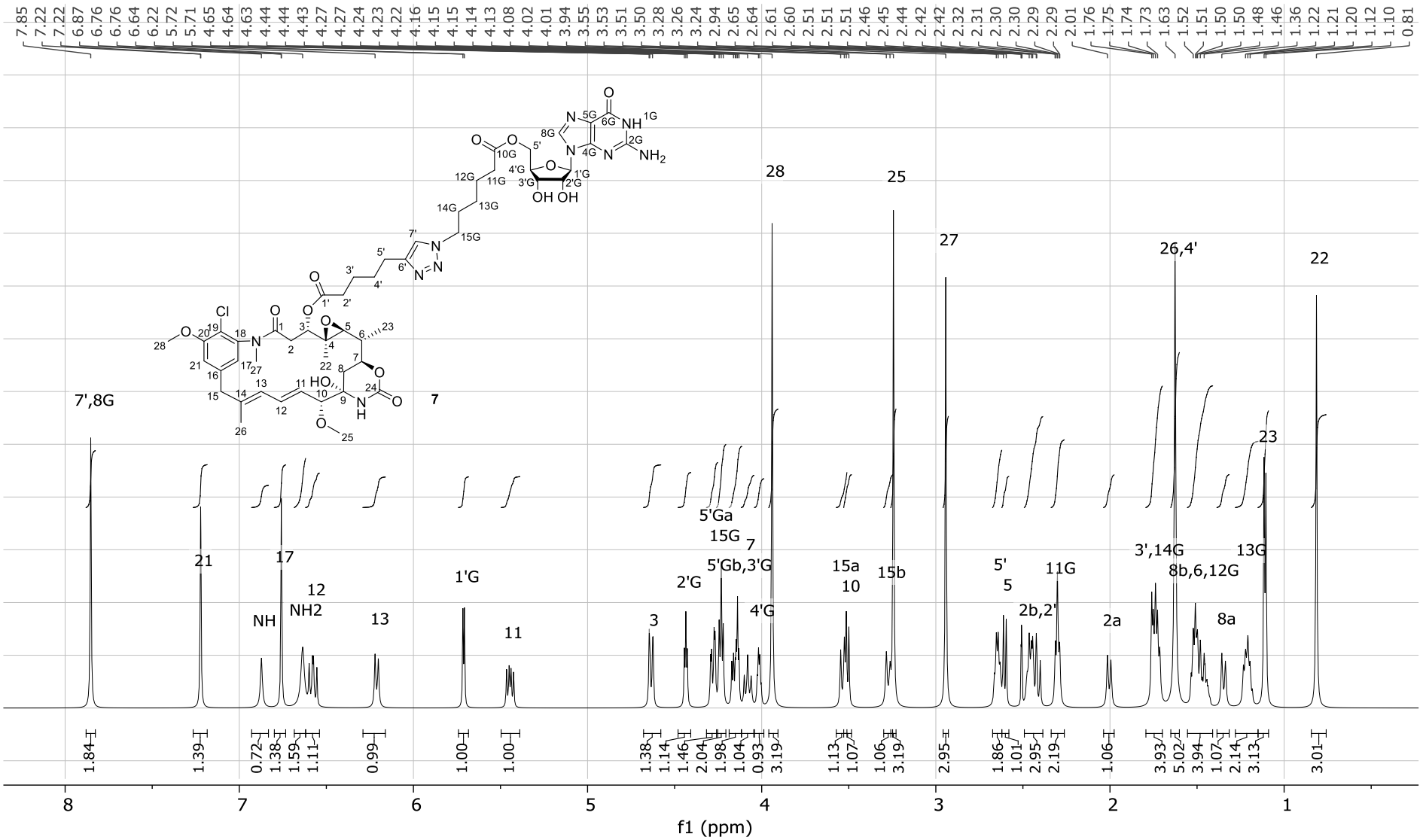


Figure S27. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **7**.

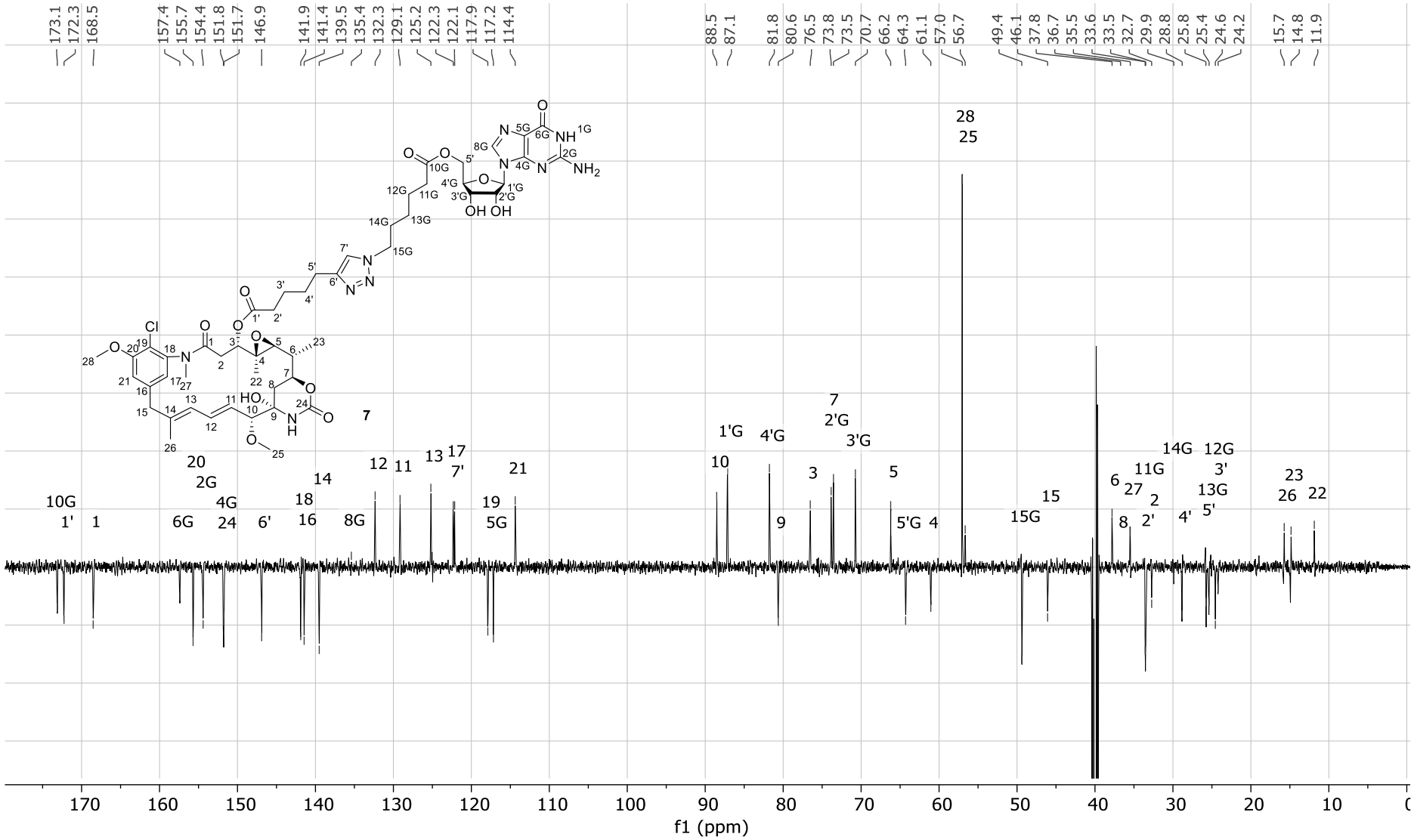


Figure S38. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **7**.

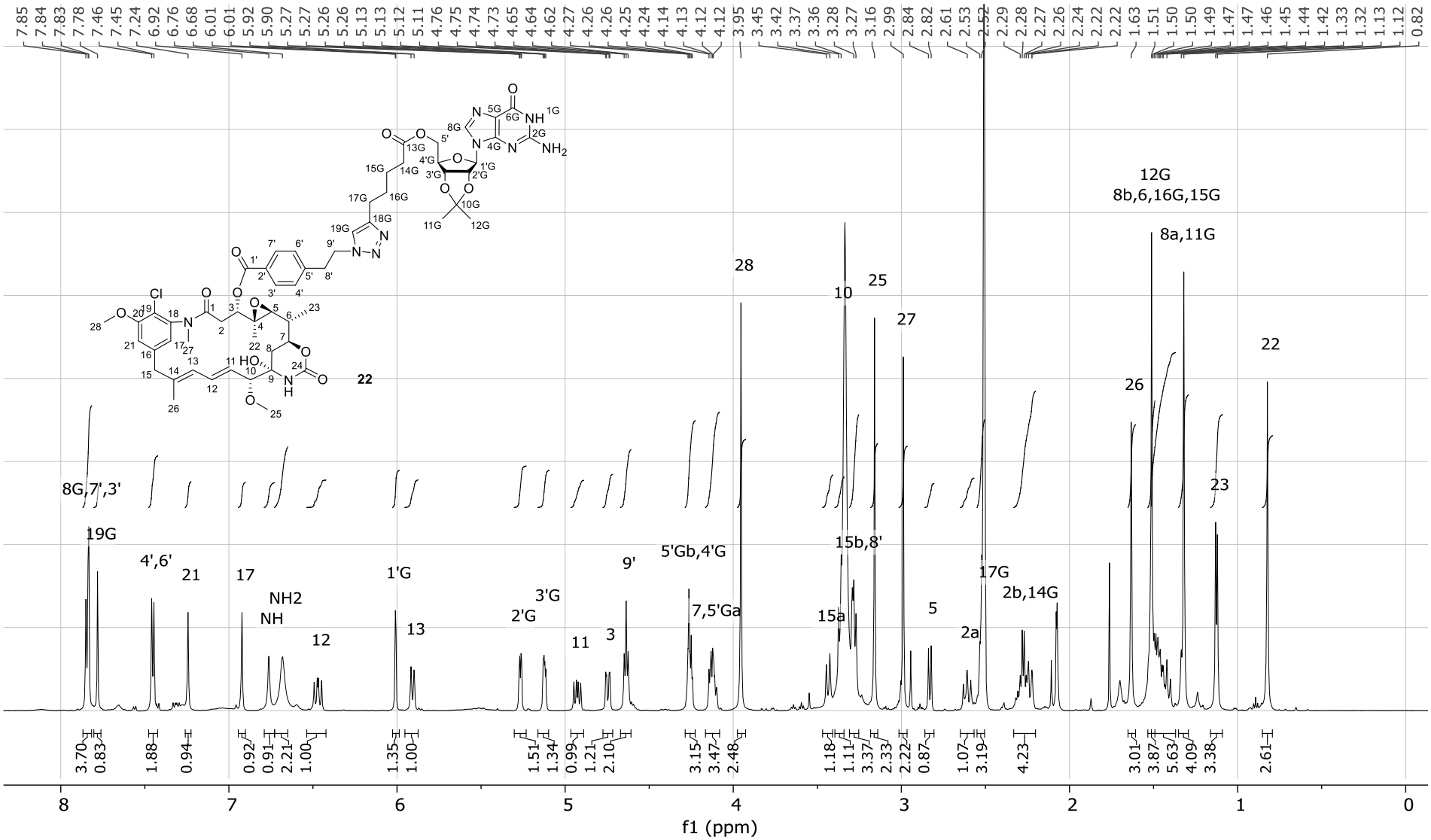


Figure S29. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **22**.

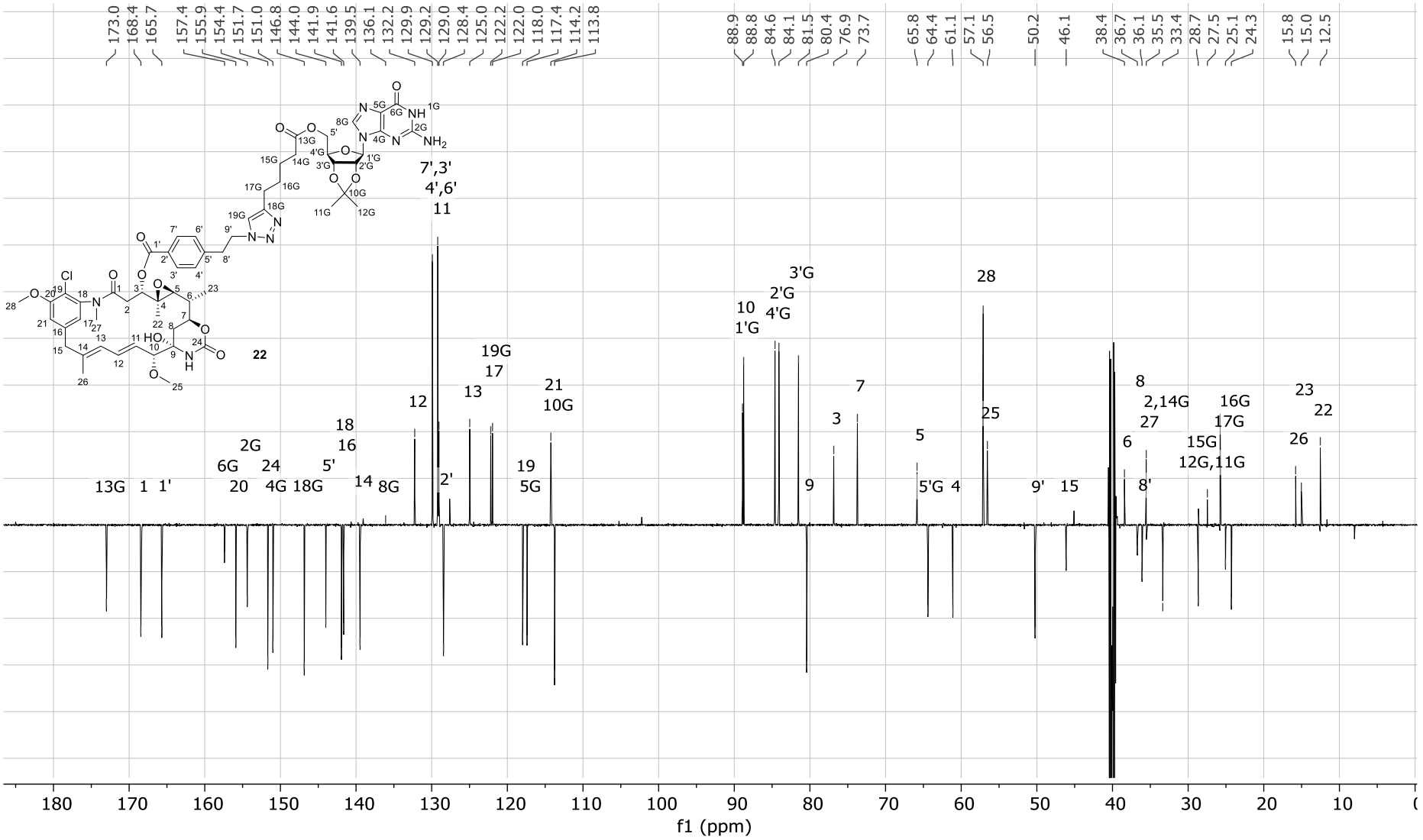


Figure S30. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **22**.

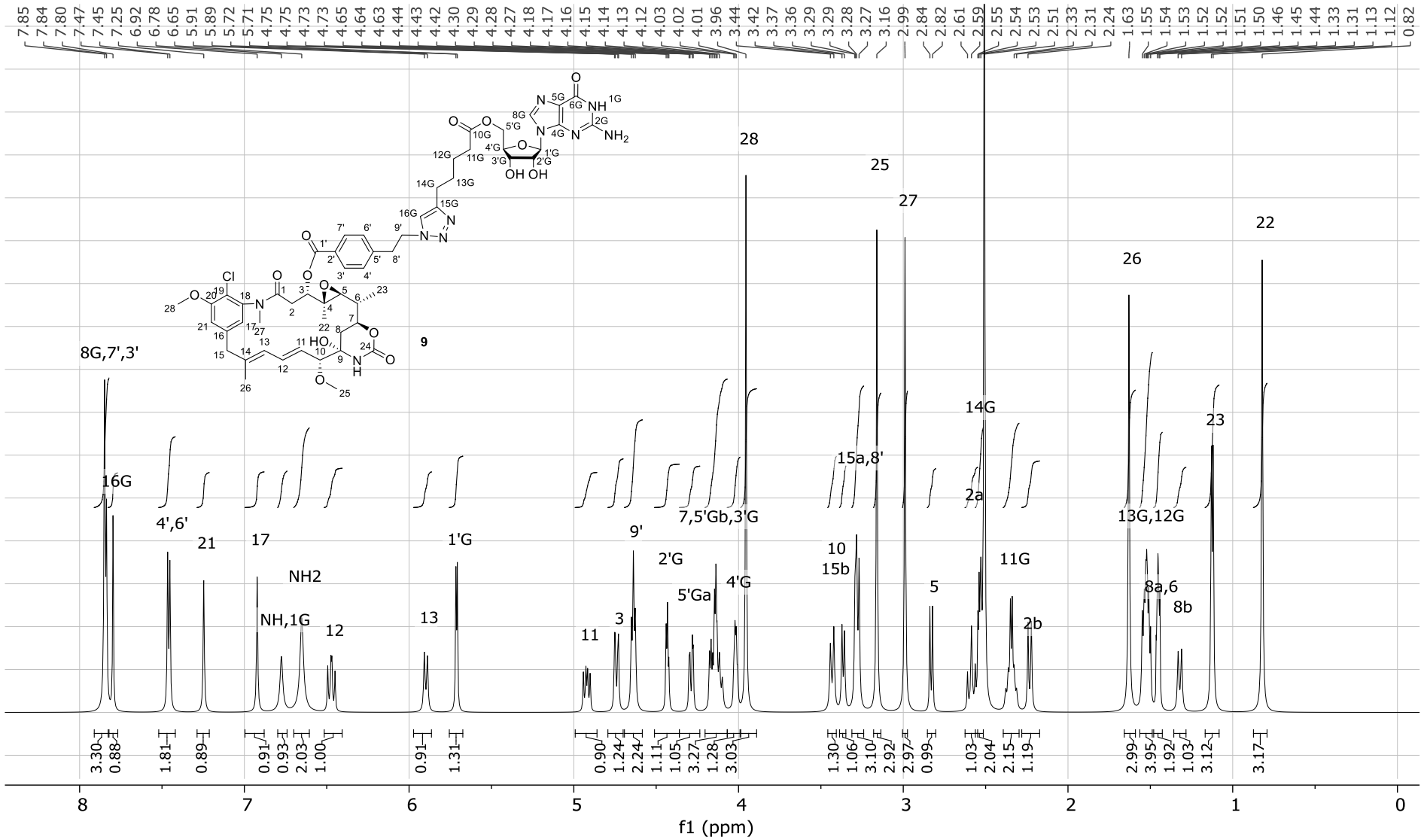


Figure S31. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **9**.

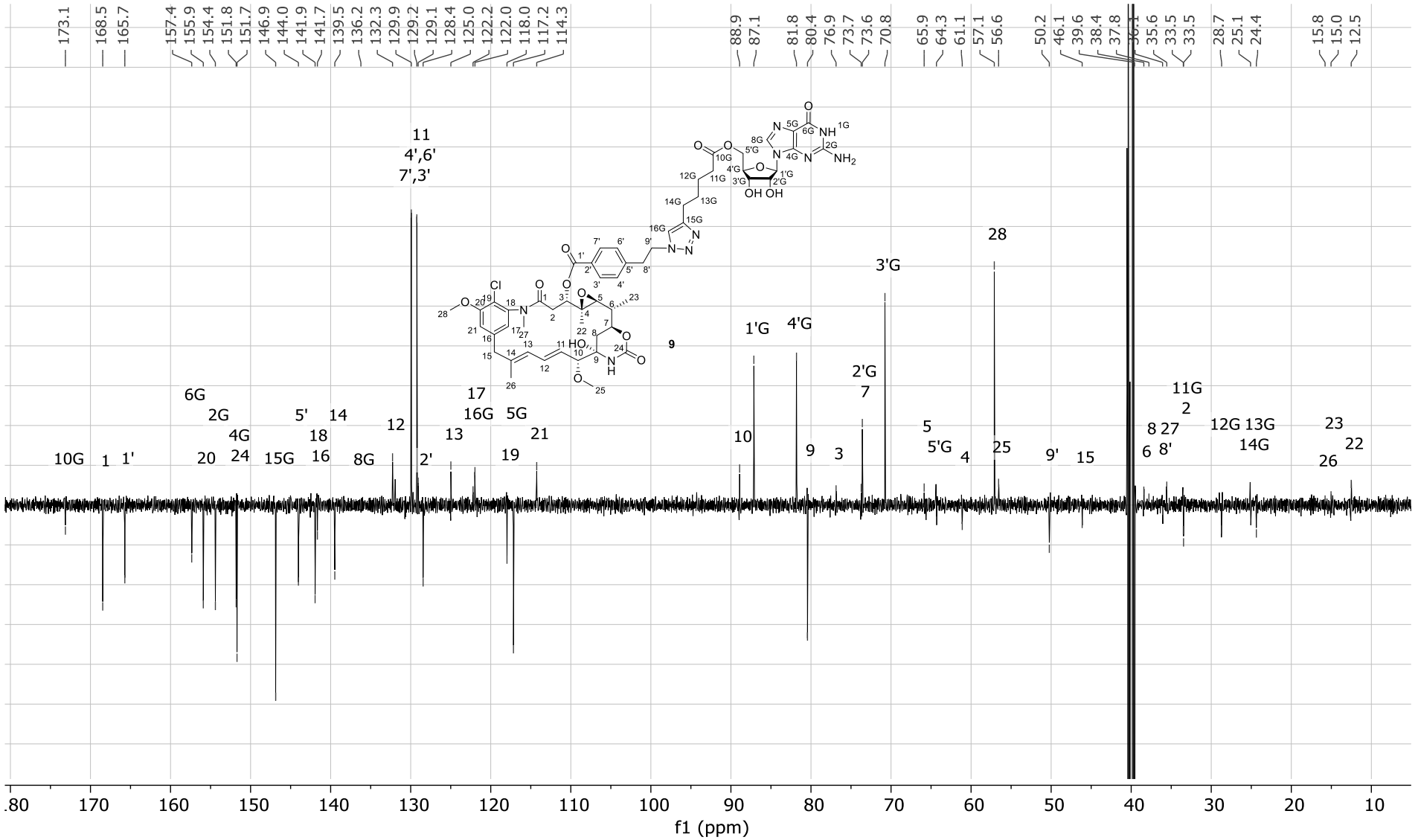


Figure S32. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **9**.

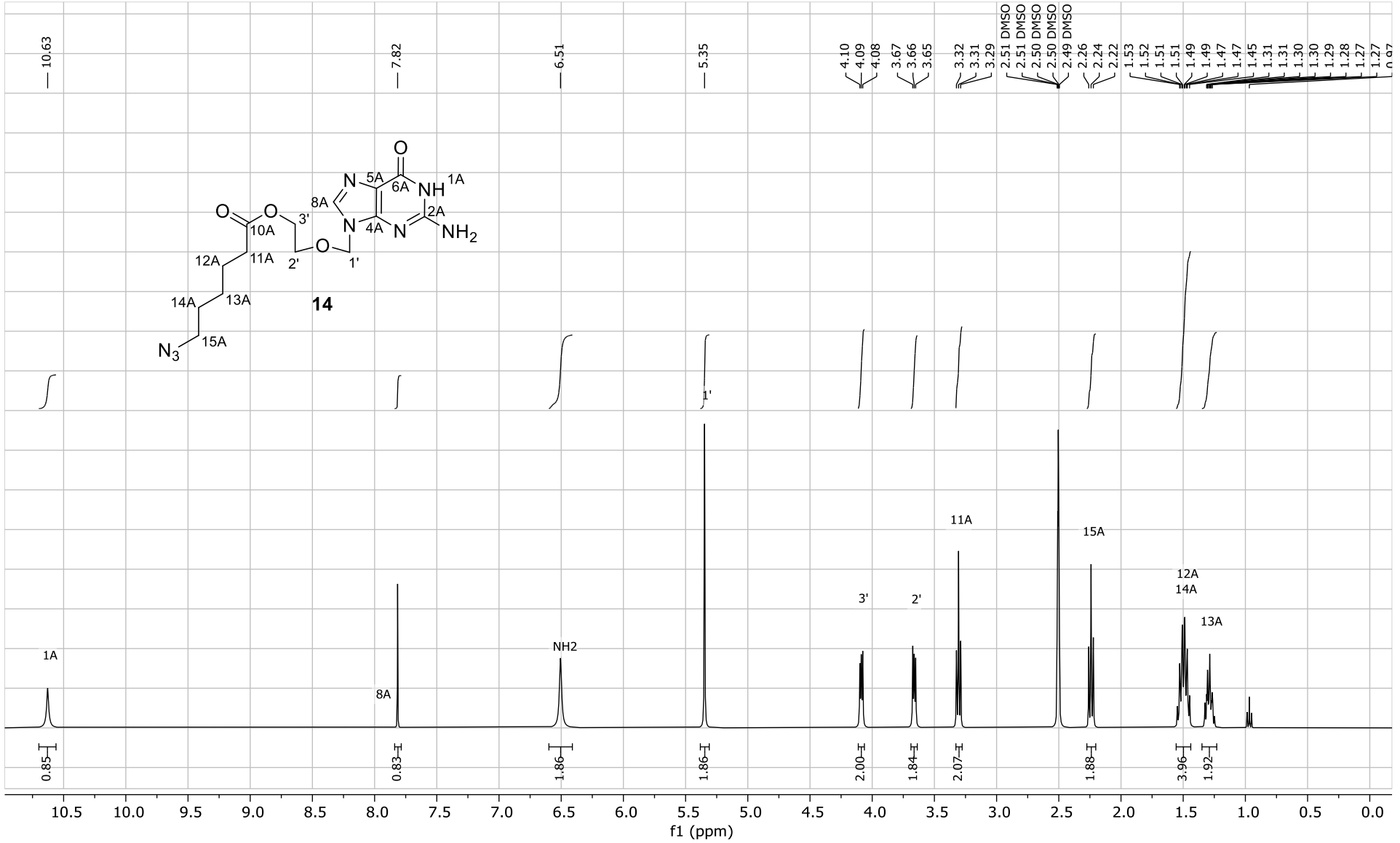


Figure S33. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **14**.

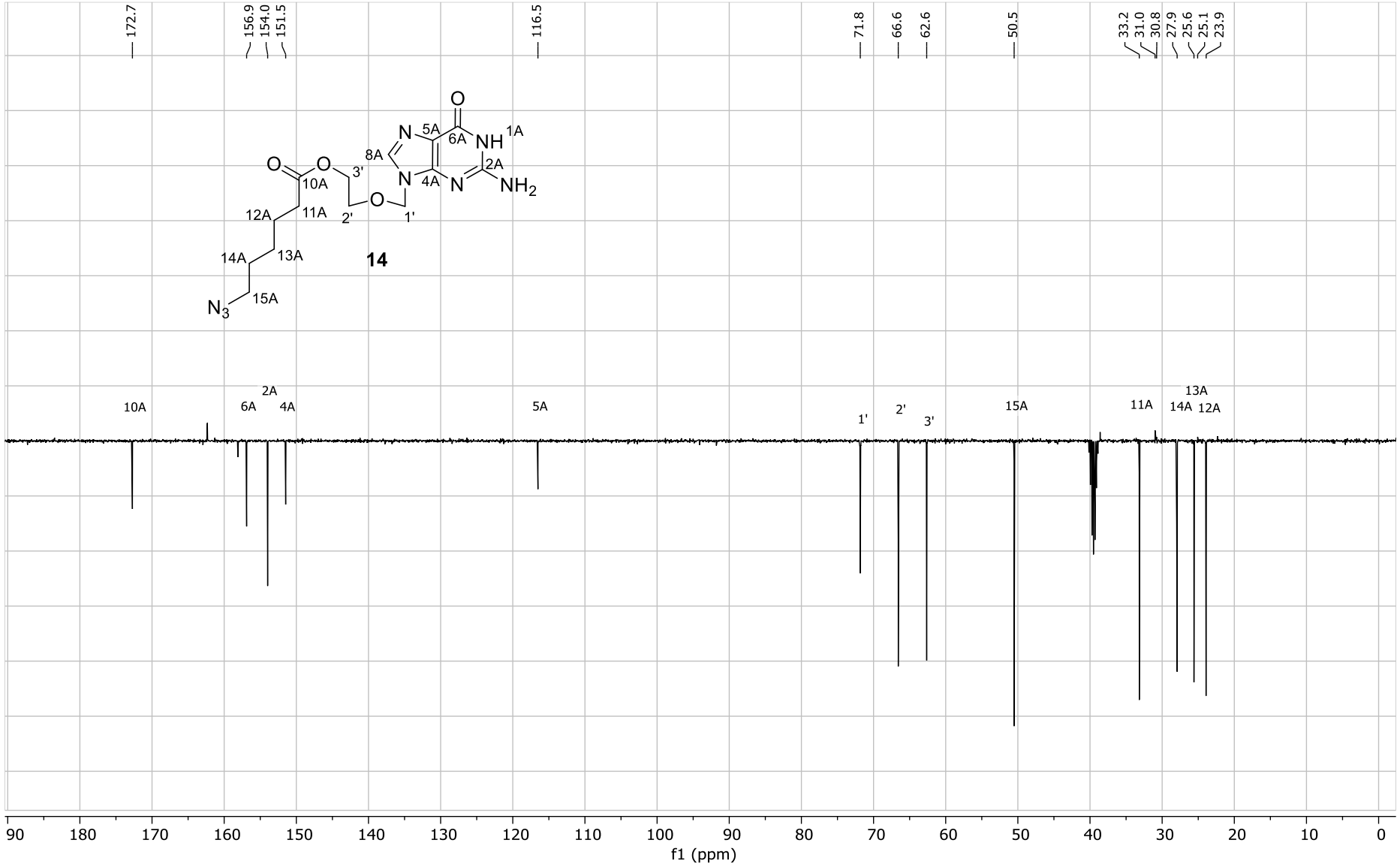


Figure S34. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **14**.

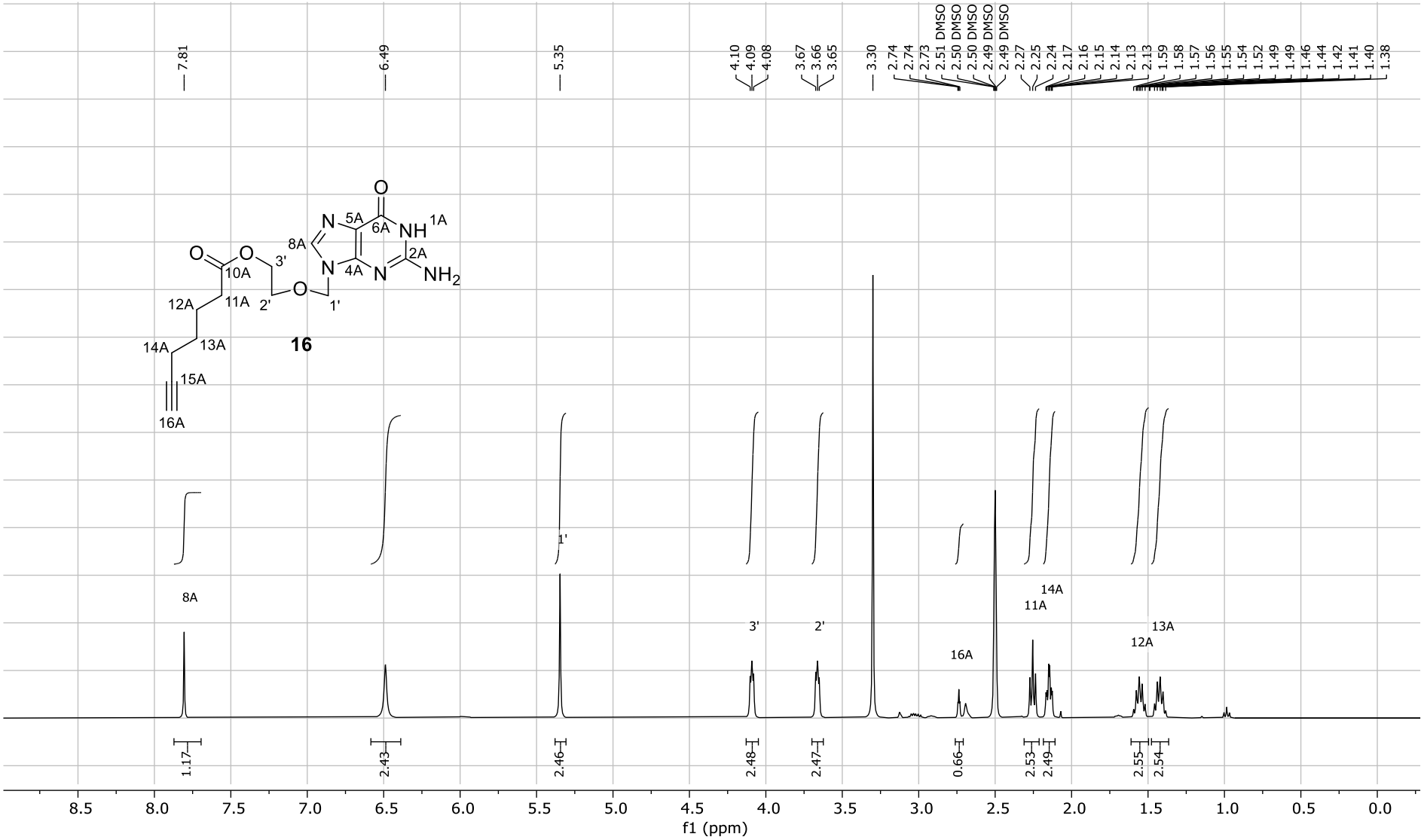


Figure S35. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **16**.

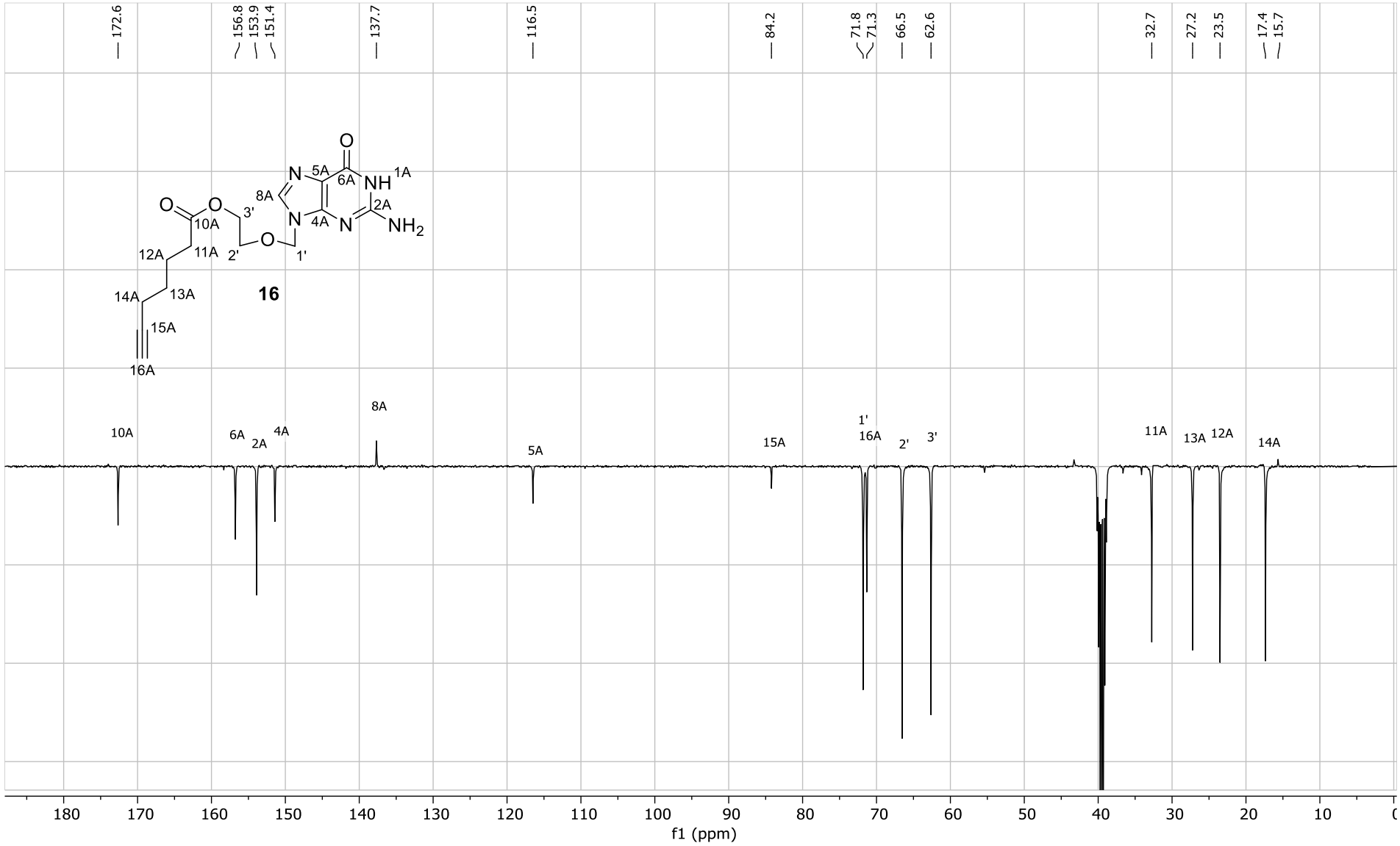


Figure S36. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **16**.

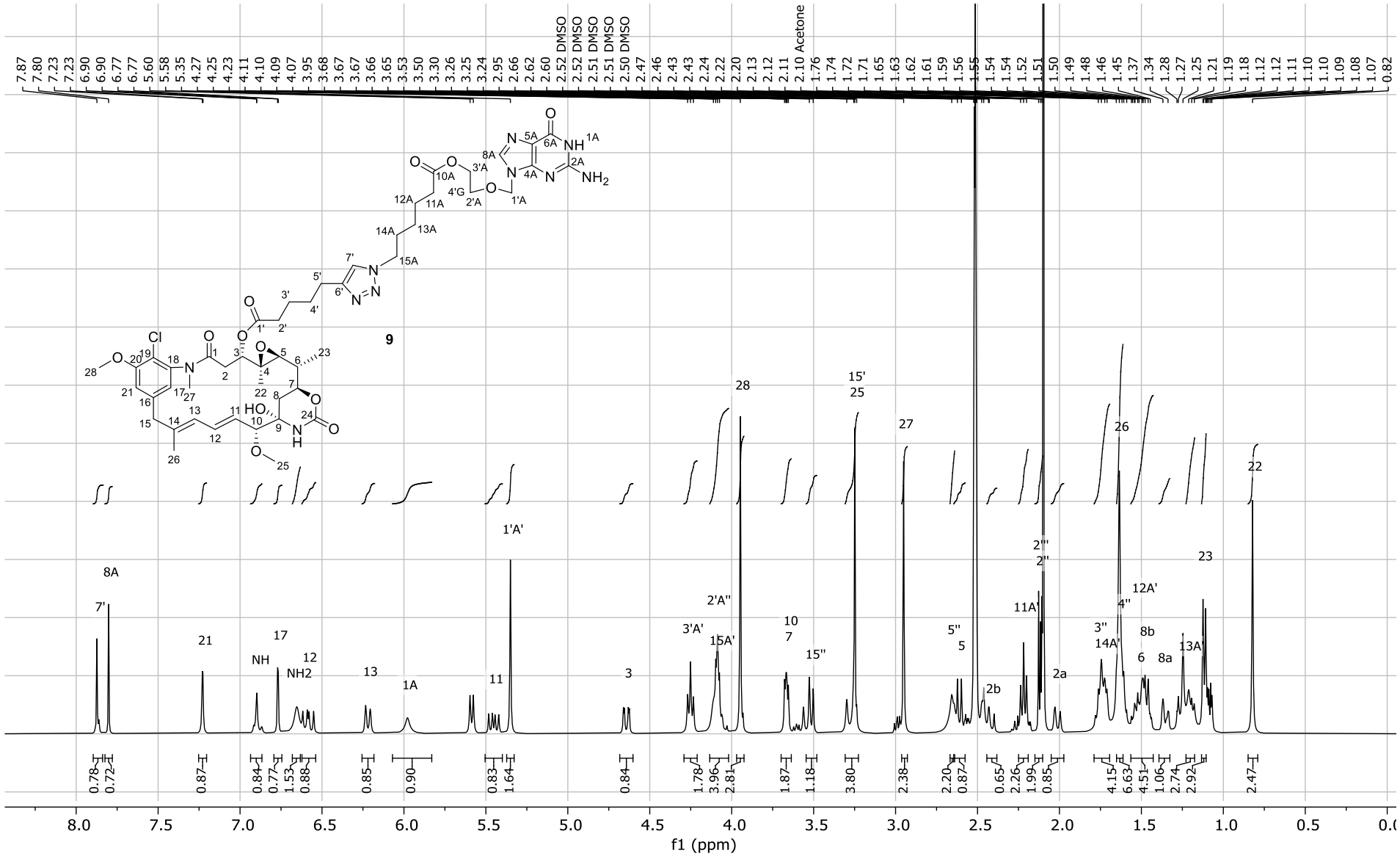


Figure S37. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **9**.

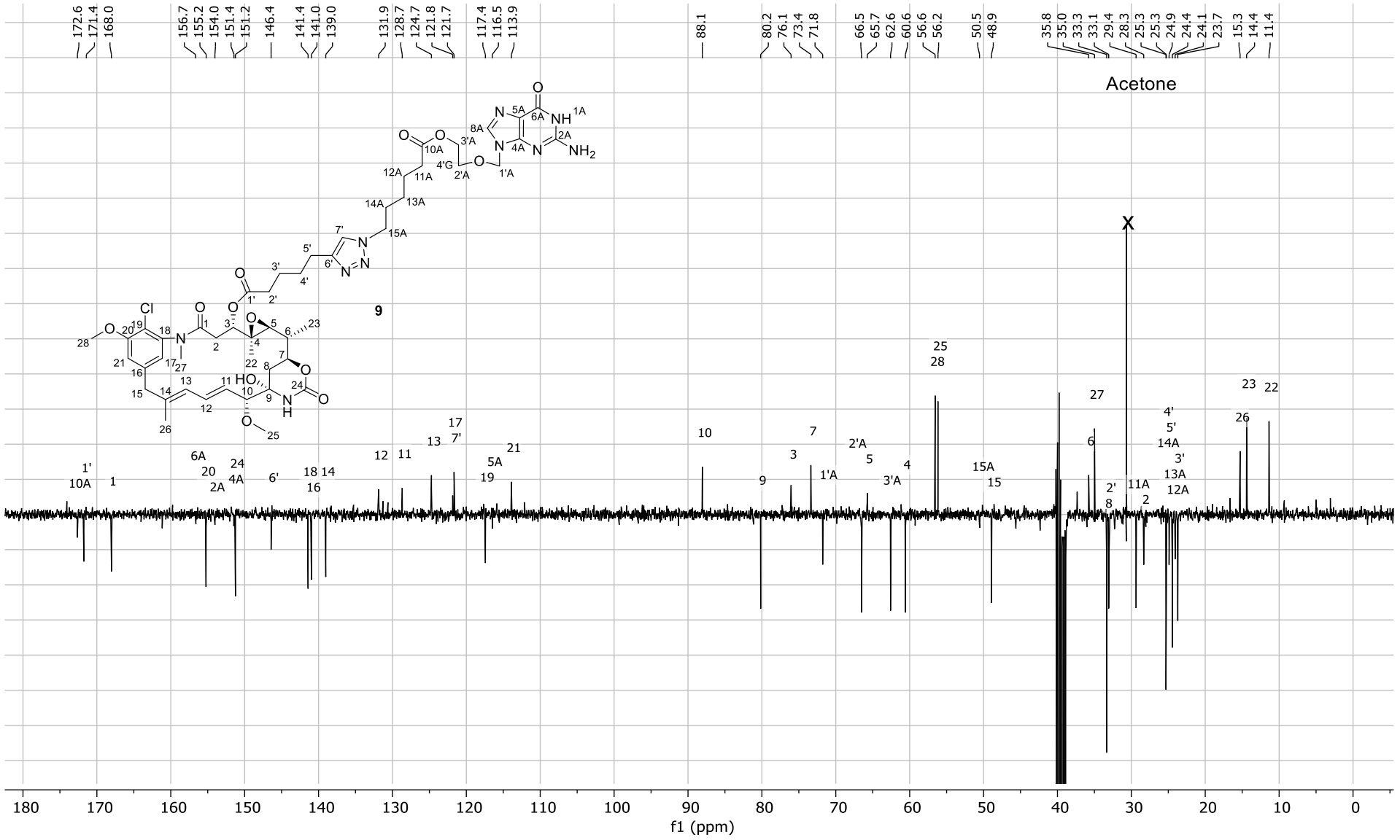


Figure S38. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **9**.

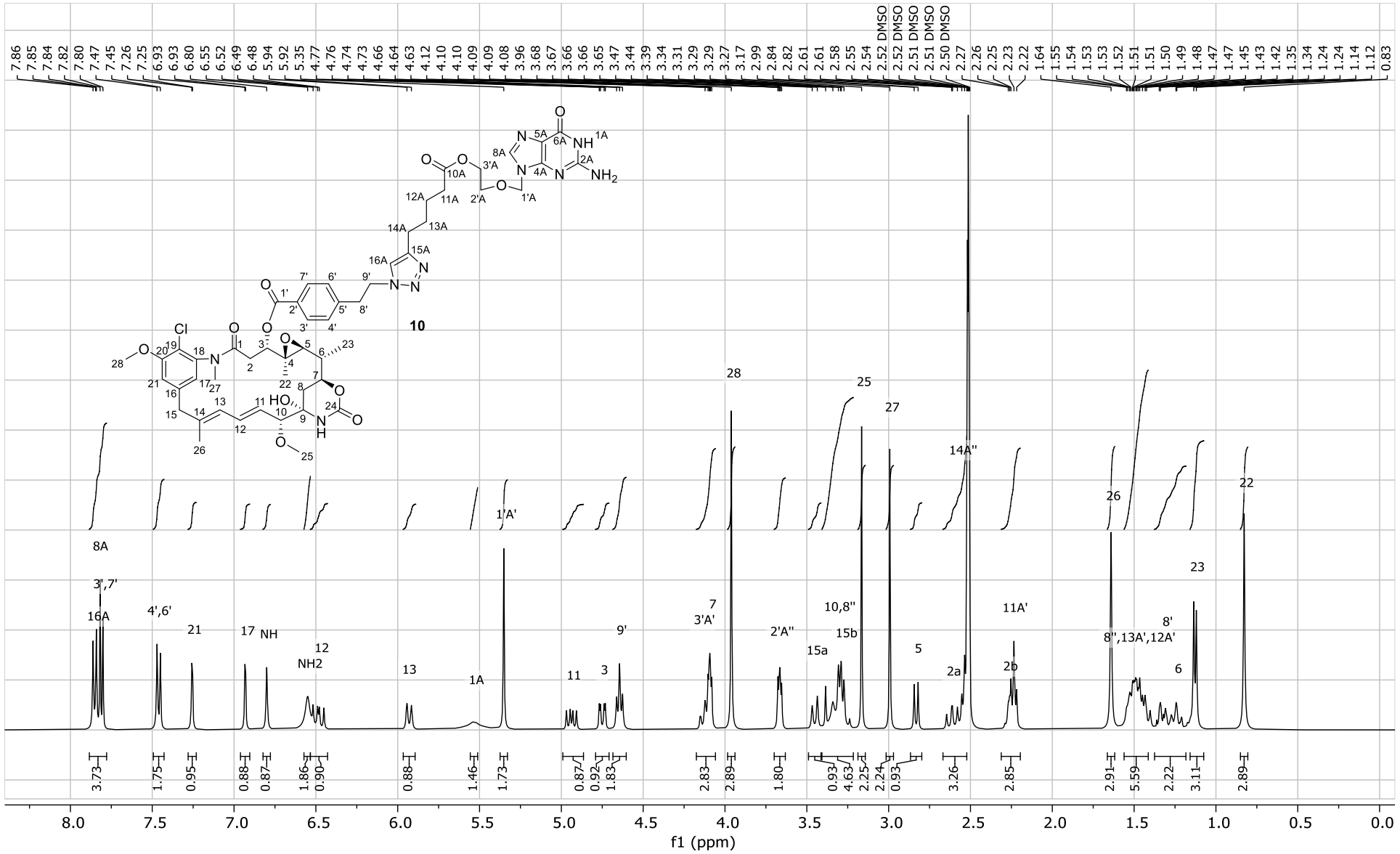


Figure S39. ^{1}H NMR spectrum (400 MHz, DMSO-d6) of **10**.

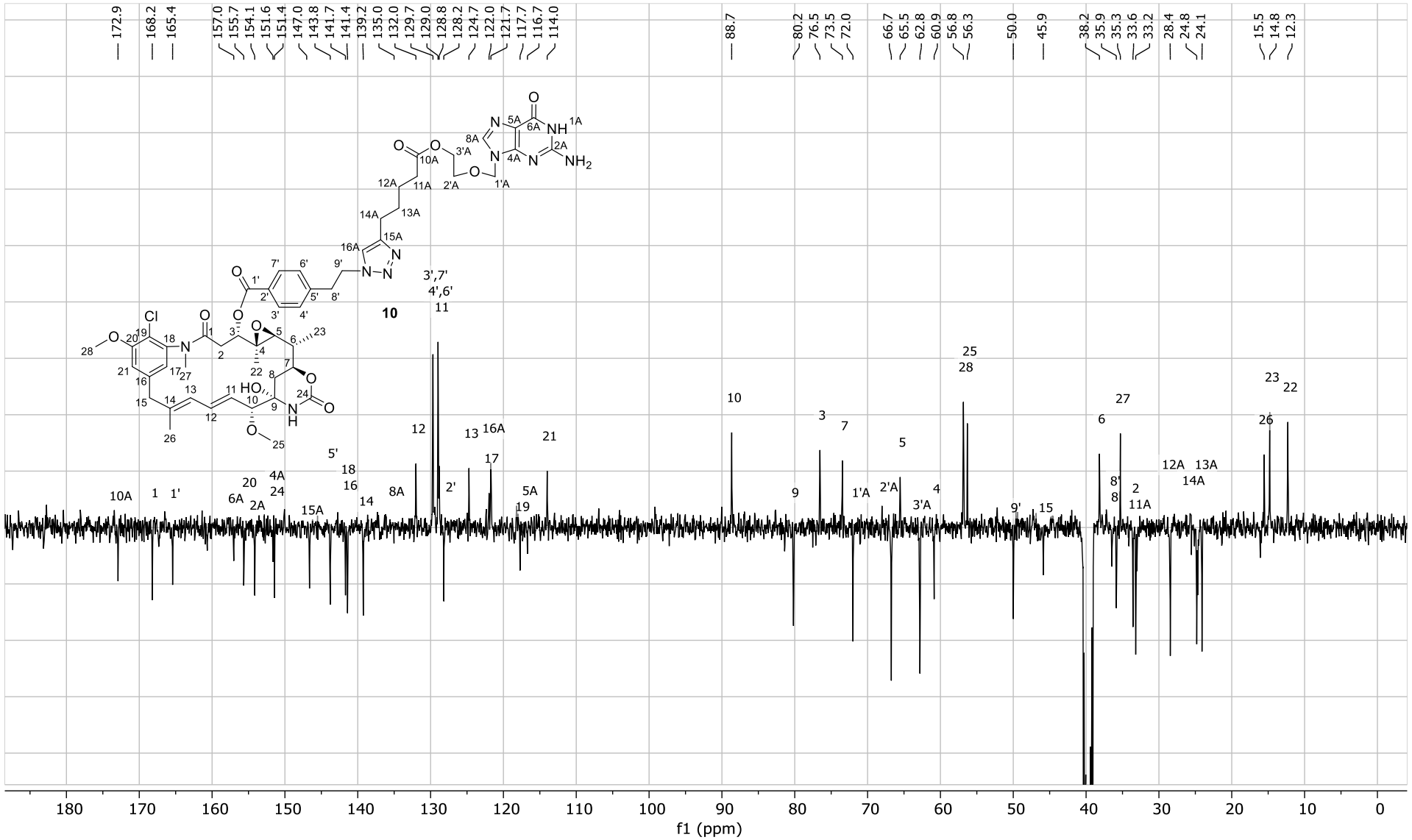


Figure S40. ^{13}C NMR APT spectrum (101 MHz, DMSO-d6) of **10**.

MOLECULAR MODELLING BY DOCKING ANALYSIS

1.1 Modeling of Long-chain Maytansinoids via Molecular Growth with S4MPLE

The docking tool used to model the long-chain maytansinoids via molecular growth present in this manuscript was the Sampler for Multiple Protein-Ligand Entities (S4MPLE), which is an in-house developed docking program.[1]

Methodology and Rigid-Site Docking Benchmarking

The high-resolution crystallographic structure used for the docking studies of the long-chain maytansinoids was obtained from the Protein Data Bank (PDB) entry with ID 4TV8.[2]

1.1.1 Fragment-based Docking Protocol of S4MPLE

S4MPLE has been developed with the aim of a versatile management of degrees of freedom, notably allowing docking scenarios in which a part of the ligand – the "anchor" fragment(s) – can be fixed, all while chosen degrees of freedom in the protein may be actively sampled. Anchors can be fragments in their X-ray determined binding pockets, as in Fragment-Based Drug Design, which can be "linked" or "grown" using S4MPLE. Here, anchors were maytansinol (in view of "growth" by "implants" targeted to reach the guanosine site) or maytansinol and guanosine (in view of "linking"). No flexibility of the protein was allowed. The steps of this growth/linking protocol are similar to the previously reported[1], with some minor amendments:

1.1.1.1 Target Site Preparation

In S4MPLE, the binding site is defined by generating a cut-out zone around the ligands. Two different input target sites were generated. For the docking of molecules **4**-**9**, we considered all the atoms of the target site present in residues within a radius of 10 Å around the maytansine molecule present in the structure 4TV8 including the cation Mg^{2+}. For the docking of molecules **10**-**13**, we considered all the atoms of the target site present in all residues within a radius of 10 Å around the maytansine and GDP molecules. In this case, the cation Mg^{2+} was excluded. Docking input files for the target protein β-tubulin were generated using PyMol 2.3.4 (the PyMOL Molecular Graphics System Version 2.3.4 Schrödinger, LLC). The protein site file was stored at "ref.mol2" in Sybyl MOL2 format in a site directory, together with the fixed_atom list (here, all protein atoms were considered fixed).

1.1.1.2 Preparation of Anchors

Anchor fragments must be specified in SDF format, with the coordinates taken from the fragment pose in the protein site. If connection to the linker involves a chemical reaction altering the geometry of the connecting atoms, then the estimated geometry of the fragment after reaction must be used (not relevant here). The anchor fragment must be fully hydrogenated in agreement of the assumed protonation status of putatively ionizable groups, with one notable exception: the H atom(s) yielding the free valency expected to be used for connecting to the linkers must be deleted, while the concerned connecting heavy atom(s) must be marked using the atom mapper functionality of structure editors such as Marvin or ChemDraw. In a growth scenario, the only connecting heavy atom in the single anchor fragment must be mapped to **1**. For linking, both anchors must be given in a same SDF, with connecting heavy atoms mapped as **1** and **2**, respectively. At least some of the atoms in both anchor fragments must be defined as fixed. A specific "anchor" directory containing the above-mentioned SDF (to be named ref.sdf) and a list of fixed atoms (by numbers as in ref.sdf, in file fixed_atoms) must be provided by the user.

1.1.1.3 Preparation of linker fragments

Mono- (for growth) or bidentate fragments (for linking) must be provided in SMILES format, with respective connecting atoms mapped "**1**" or "**1**" and "**2**", such as to match the marks of associated anchor connectors. They can be drawn manually (as the case here) or prepared by chemoinformatics standardization/reaction transformation tools and mapped with atom mapper protocols.

1.1.1.4 Preparation of the Actual Ligand Structures

The SMILES string of each linker fragment is first converted to 3D, with explicit hydrogens as expected at physiological pH, using a the same ChemAxon API-based tool serving to pre-process regular ligands. The resulting fragment SDF is merged with the anchor ref.sdf, and the list of anchor fixed atoms is imported. Then, S4MPLE is invoked with the "connector" directive, herewith creating covalent bonds between marked atoms of same map number. Note that at this point, neither the required GAFF atom types nor the partial charges are assigned. However, S4MPLE will proceed to a coarse fitting of the linker to the (fixed) anchor atoms, based on "default" force field parameters, based on a few tens of evolutionary algorithm generations. This will coarsely preposition the linker in the reference frame of the anchor, all while roughly ensuring reasonable bond lengths and valence angle values (as far as possible) and avoiding anchor-linker atom clashes. The resulting ligand geometry is stored together with the inherited list of fixed anchor atoms. It is next subjected to GAFF force field typing, Gasteiger partial charge calculation and detection of saturated rings requiring explicit sampling, and then added to a tar archive as ready for docking.

1.1.1.5 Docking Protocol

S4MPLE is used with the "fitted" parameterization of the AMBER/GAFF engine, including an implicit desolvation model, as previously reported.[1]

1.1.1.5.1 Free-ligand Sampling

First, the completely free ligand (anchors no longer fixed) is sampled for 200 generations, obtaining its lowest free-state energy value Efree. This is a key component of the S4MPLE docking score Ebound-Efree (vide infra).

1.1.1.5.2 In-Site Linker Sampling

Given the above-prepared protein site directory and the tar archive of pre-processed ligands, the ligand growth protocol will iteratively extract files pertaining to a given ligand, merge them with the protein information and proceed to an evolutionary sampling of the non-fixed moieties in the resulting construct. First, this involves a calibration step of the contact fingerprint-based redundancy threshold: random geometries are generated and the mean, minimal and maximal dissimilarity scores of their contact fingerprints is monitored. Based on these observations, a dissimilarity threshold below which two conformers are considered redundant is derived.[1]

Then, the non-fixed linker is sampled for 500 evolutionary generations, in search for its optimal orientation with favorable contacts to the protein site, all while respecting the covalent constraints imposed by bonds to fixed anchor fragments. The best obtained energy level of this simulation is stored into a "local best energy" list. This 500-generation simulation is then repeated, leading to a presumably different local "best" energy, concatenated to the local best energy list, which is sorted ascendingly. As soon as the first three top energy values in this list are within 1 kcal/mol, it is considered that the simulation "reproducibly" found the so-far best energy optimum. Otherwise, up to 500-generation simulations are repeated, after which the process is force-stopped in spite of failure to "converge" in the above sense.

1.1.1.5.3 Bound Ligand Energy

Best energies above are considered with fixed anchor fragments, and are thus not directly comparable with the ligand lowest free-state energy value Efree. Therefore, the 20 top poses from the above sampling step are resubmitted to S4MPLE for energy minimization of the fully free ligand (anchoring fixed_atoms are ignored). However, if this relaxation step triggers a complete repositioning of the ligand, at the cost of significantly dragging anchor fragments away from their "native" positions, the resulting pose must be rejected. Therefore, S4MPLE is used to calculate the heavy-atom RMSD values between initial (fixed-anchor) and energy-minimized ligand poses. The RMSD is calculated within the fixed reference frame of the protein site (ligand structures are not re-overlaid to minimize it). If energy minimization causes the ligand to drift away by a RMSD>1.5Å, the pose is discarded. The lowest minimized energy of non-discarded poses is then taken to represent Ebound.

1.1.1.5.4 Molecular Superimposition

The structure comparison tool Match Maker in the software UCSF Chimera 1.14[3] was used for the superposition of molecular structures.

1.1.1.5.5 Figures Preparation

The structural renderings were obtained using PyMol 2.3.4.

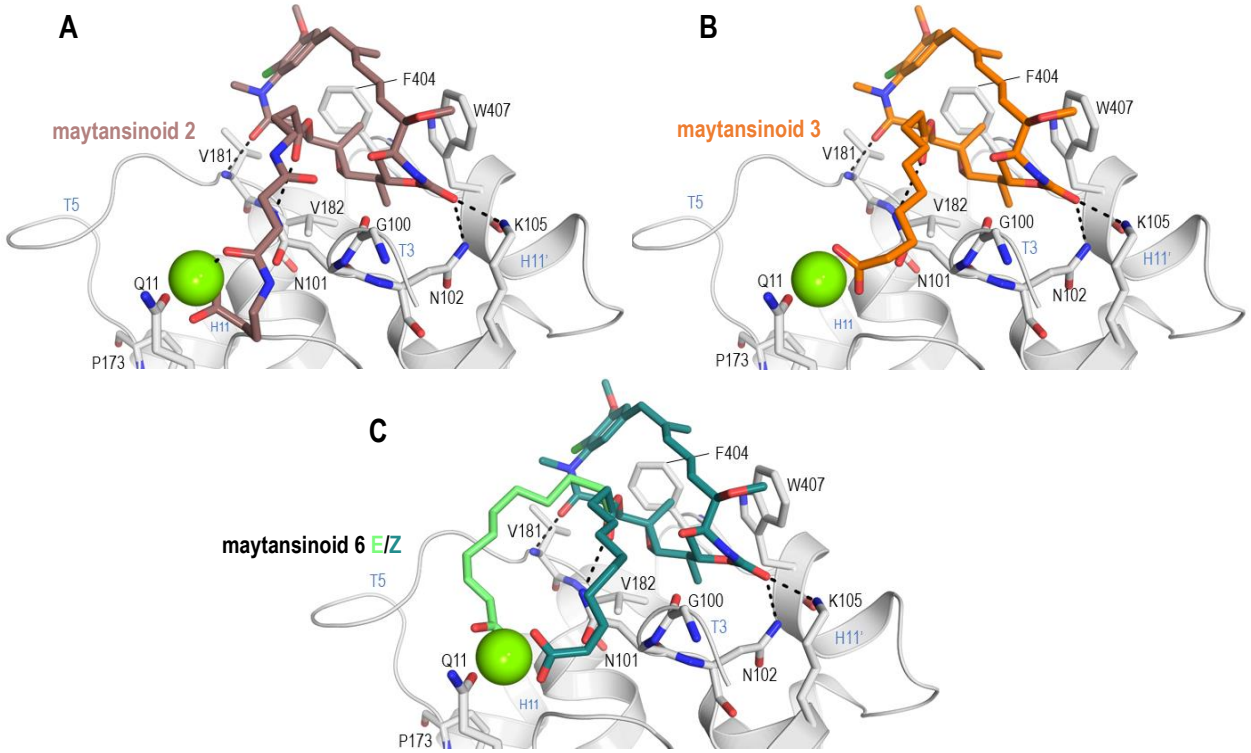


Figure S41. Docking results of the compounds (A) **2**, (B) **3** and (C) the E/Z isomers of the maytansinoid **6** when bound to β-tubulin (grey) in the presence of the Mg^{2+} (green ball).

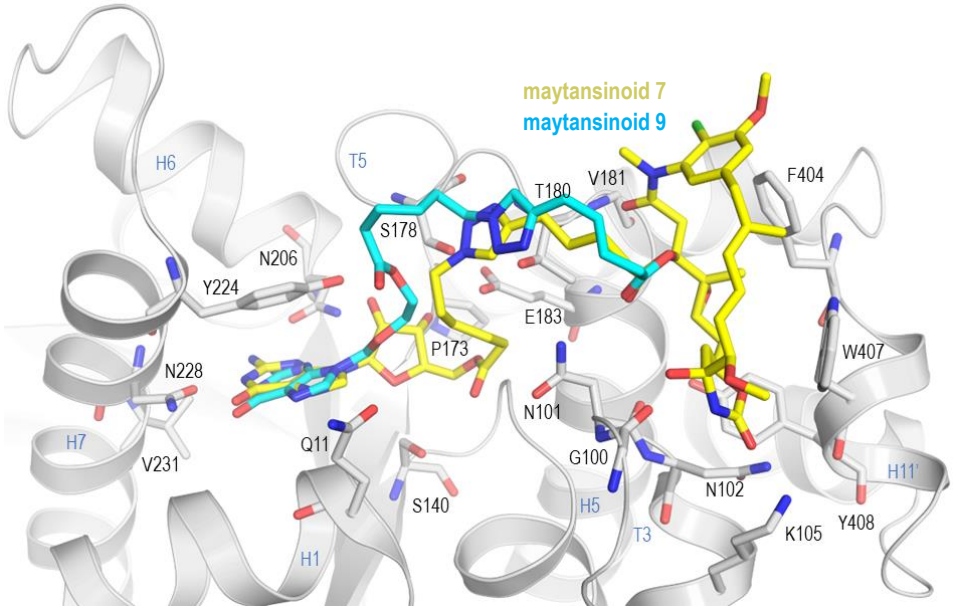


Figure S41. Docking result of the maytansinoid-guanosine conjugate **7** (yellow) superimposed to the docking result of the maytansinoid-acyclovir conjugate **9** (cyan) when bound to tubulin (grey).

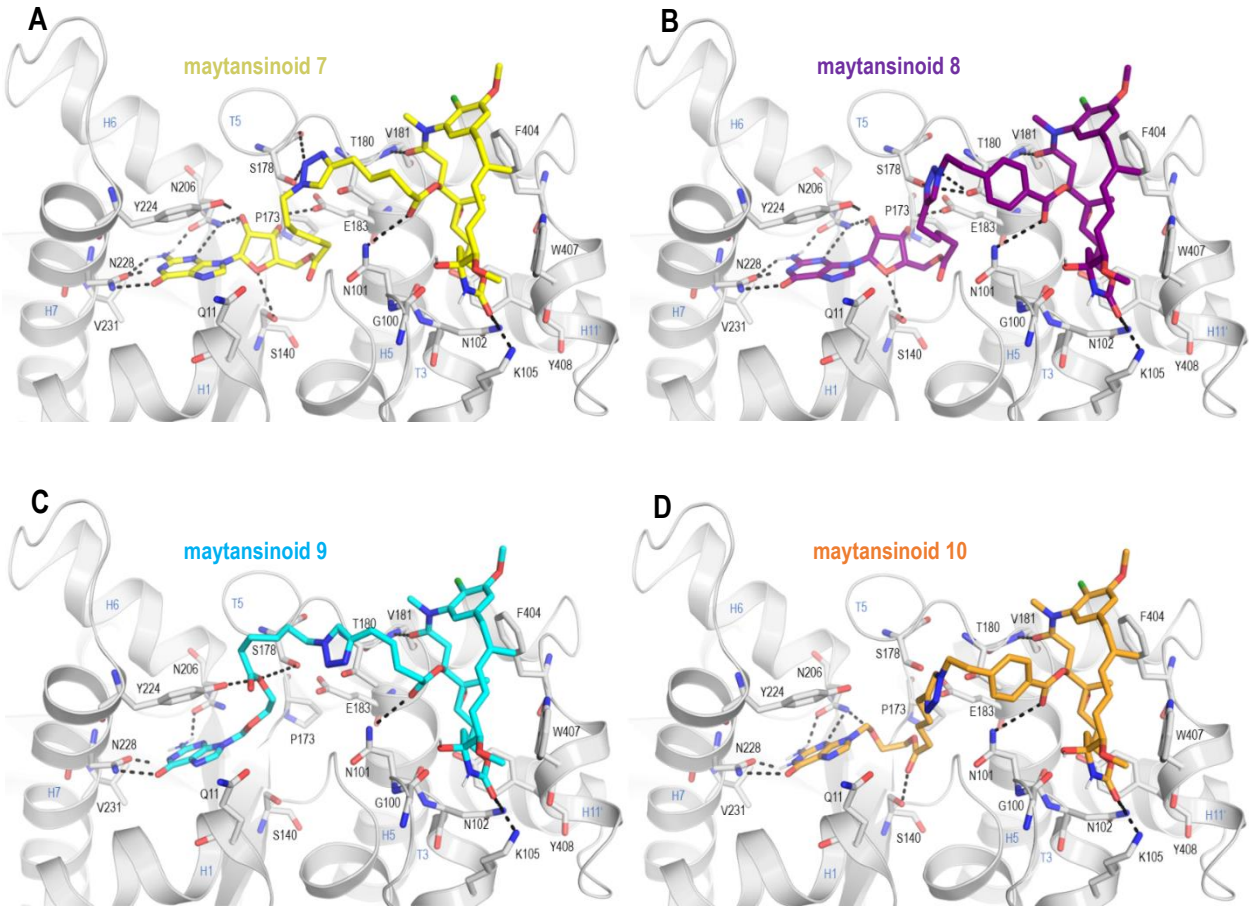


Figure S43. Docking results of the maytansinoid-guanosine conjugates (A) **7** and (B) **8** and, of the maytansinoid-acyclovir conjugates (C) **9** and (D) **10**. The black spotted lines represent the polar interactions between the ligands and β-tubulin (grey).

CRYSTALLIZATION, DATA COLLECTION AND STRUCTURE DETERMINATION

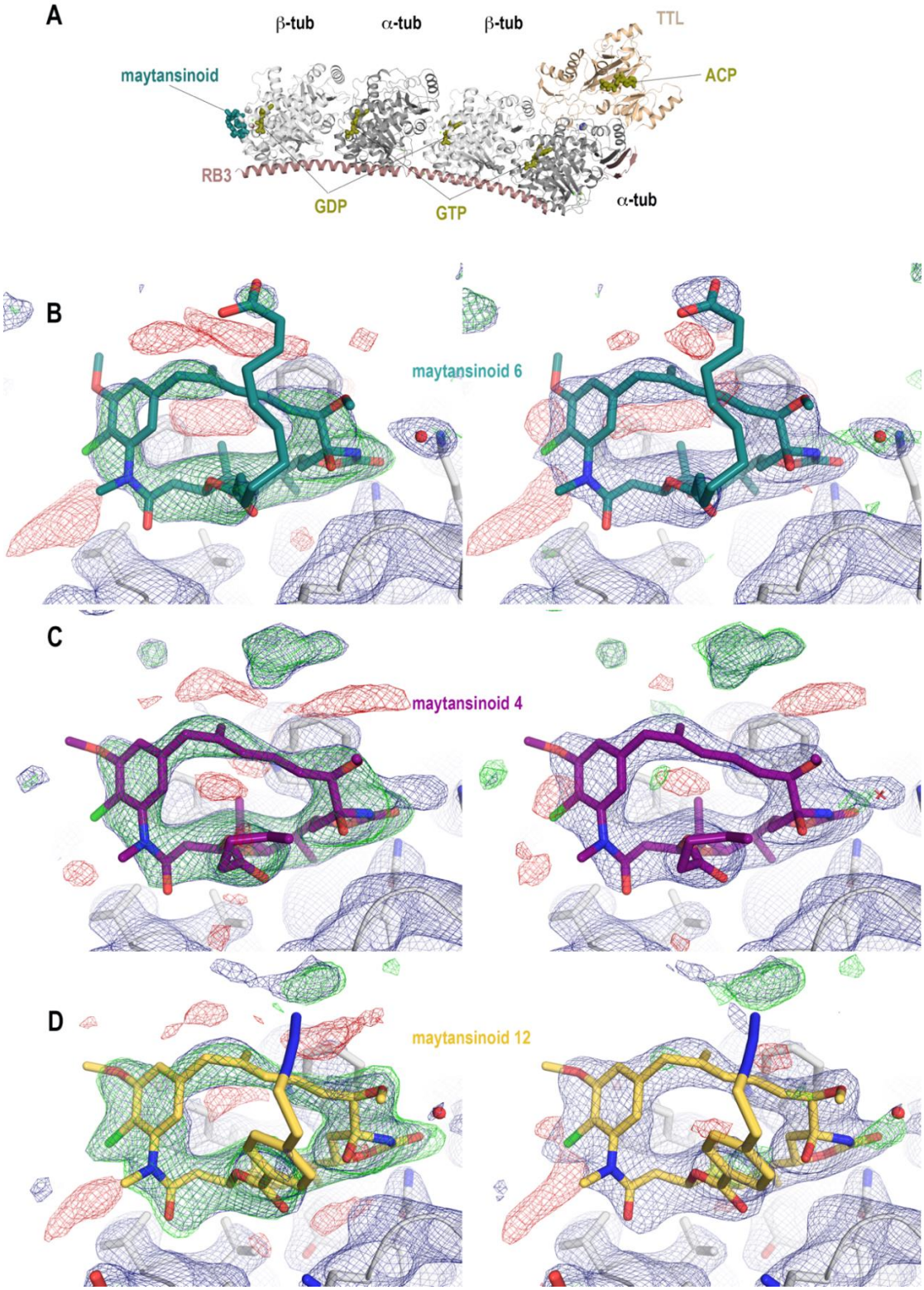


Figure S44. Display of the T_2R-TTL-maytansinoid complex structure and the fit of the maytansinoids **4**, **6** and **12** into the electron density

In A the protein structure of the T_2R-TLL complex composed of 2 tubulin dimers (T), the stathmin-like protein RB3 (R) and the tubulin tyrosine ligase (TTL) with a bound maytainsinoid is displayed. The protein backbone is shown in ribbon representation (α-tub dark grey, β-tub light grey), while nucleotides (olive) and maytansinoid (deep teal) are shown in stick representation. B, C and D display the electron density fit of maytansinoids **4**, **6** and **12**, (PDB IDs: 8B7A, 8B7B and 8B7C). Maytansinoids, as well as close tubulin residues, are shown in stick representation, the SigmaA-weighted 2mFo-DFc maps are contoured at + 1.0 σ (blue mesh) and the mFo-DFc maps at $\pm$ 3.0 σ (green/red mesh). For each maytansinoid, both a superposition of the ligand with the corresponding omit map (left) and the fit into the refined map (right) are shown.

Table S2. Data collection and refinement statistics.

	T_2R – TTL - 4 complex	**T_2R-TTL - 6 - complex**	**T_2R - TTL - 12 complex**
PDB ID	8B7A	8B7B	8B7C
Data collection			
Wavelength	1.000029	0.999995	1.000029
Resolution range	47.23 - 2.25 (2.33 - 2.25)	47.83 - 2.25 (2.33 - 2.25)	49.57 - 1.9 (1.968 - 1.9)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	105.271 158.55 181.846 90 90 90	104.085 155.637 181.924 90 90 90	104.866 157.939 180.57 90 90 90
Total reflections	1965496 (183962)	1913067 (182121)	3181903 (307017)
Unique reflections	144448 (14292)	139947 (13689)	23535013.5 (23217)
Multiplicity	13.6 (12.9)	13.7 (13.3)	13.5 (13.2)
Completeness (%)	99.93 (99.73)	99.36 (98.84)	99.93 (99.81)
Mean I/sigma (I)	14.23 (0.61)	16.76 (0.86)	15.92 (0.58)
Wilson B-factor	50.24	49.70	39.51
R_{merge}	0.1852 (3.992)	0.1489 (2.913)	0.1195 (3.954)
R_{meas}	0.1924 (4.158)	0.1546 (3.029)	0.1242 (4.111)
R_{pim}	0.0518 (1.152)	0.04133 (0.8236)	0.03355 (1.118)
$CC^{1/2}$	0.999 (0.246)	0.999 (0.357)	0.999 (0.238)
CC*	1 (0.628)	1 (0.725)	1 (0.62)
Refinement			
Reflections in refinement	144457 (14298)	139929 (13680)	235205 (23223)
Reflections for R_{free}	7225 (714)	6996 (684)	11762 (1162)
R_{work}	0.1809 (0.3484)	0.1837 (0.3206)	0.1865 (0.3738)
R_{free}	0.2230 (0.3854)	0.2243 (0.3433)	0.2142 (0.3903)
CC_{work}	0.955 (0.536)	0.962 (0.660)	0.965 (0.557)
CC_{free}	0.922 (0.511)	0.951 (0.620)	0.955 (0.565)
Number of non-hydrogen atoms	18429	18314	18822
macromolecules	17316	17349	17417
ligands	224	232	229
solvent	889	733	1176
Protein residues	2193	2180	2192
RMS(bonds)	0.003	0.002	0.003
RMS(angles)	0.57	0.52	0.57
Ramachandran statistics			
favored (%)	97.86	97.35	98.11
allowed (%)	2.10	2.56	1.89
outliers (%)	0.05	0.09	0.00
Rotamer outliers (%)	0.58	0.37	0.26
Clashscore	3.04	9.28	2.65
Average B-factor	67.09	68.23	54.86
macromolecules	67.35	68.53	54.98
ligands	67.35	69.02	50.81
solvent	61.92	60.97	53.83
Number of TLS groups	27	22	20

Statistics for the highest-resolution shell are shown in parentheses.

Supplemental References:

[1] L. Hoffer, C. Chira, G. Marcou, A. Varnek, D. Horvath, *S4MPLE-Sampler for Multiple Protein-Ligand Entities: Methodology and Rigid-Site Docking Benchmarking*, **2015**.

[2] A. E. Prota, K. Bargsten, J. F. Diaz, M. Marsh, C. Cuevas, M. Liniger, C. Neuhaus, J. M. Andreu, K. H. Altmann, M. O. Steinmetz, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13817–13821.

[3] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.