Supplementary material (ESI) for Chemical Communications

Condensing the information in DNA with double-headed nucleotides

Mick Hornum,[a] Pawan K. Sharma,[b] Charlotte Reslow-Jacobsen,[a] Pawan Kumar,[a] Michael Petersen,[a] and Poul Nielsen[a],†

a. Nucleic Acid Center, Department of Physics, Chemistry & Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark.

b. Department of Chemistry, Kurukshetra University, Kurukshetra-136 119, India.

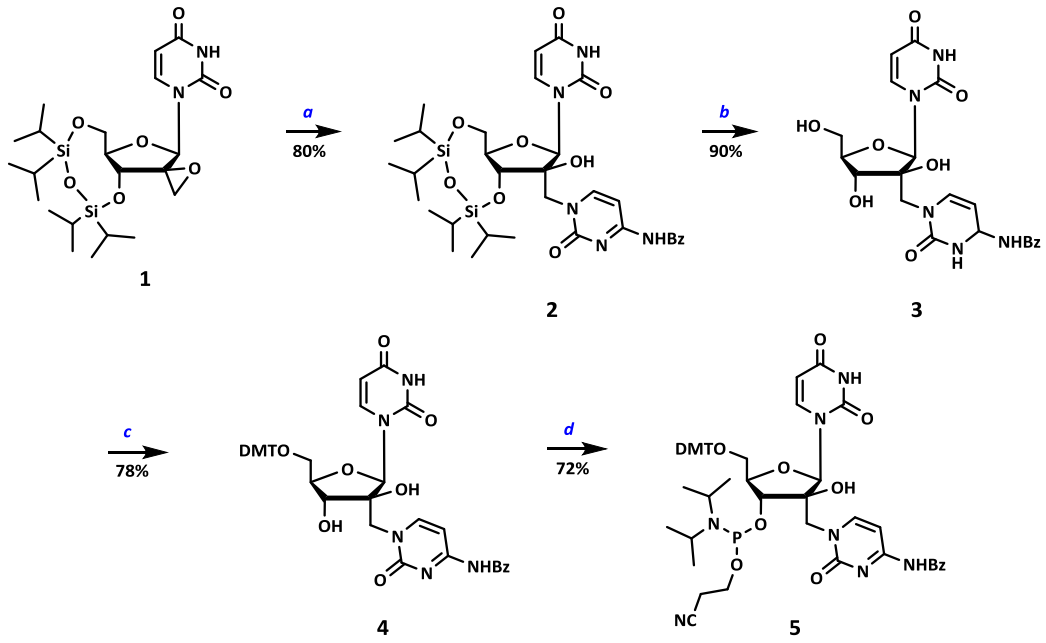
†Corresponding author. Tel.: +45 65502565; email: pouln@sdu.dk.

Electronic Supplementary Information

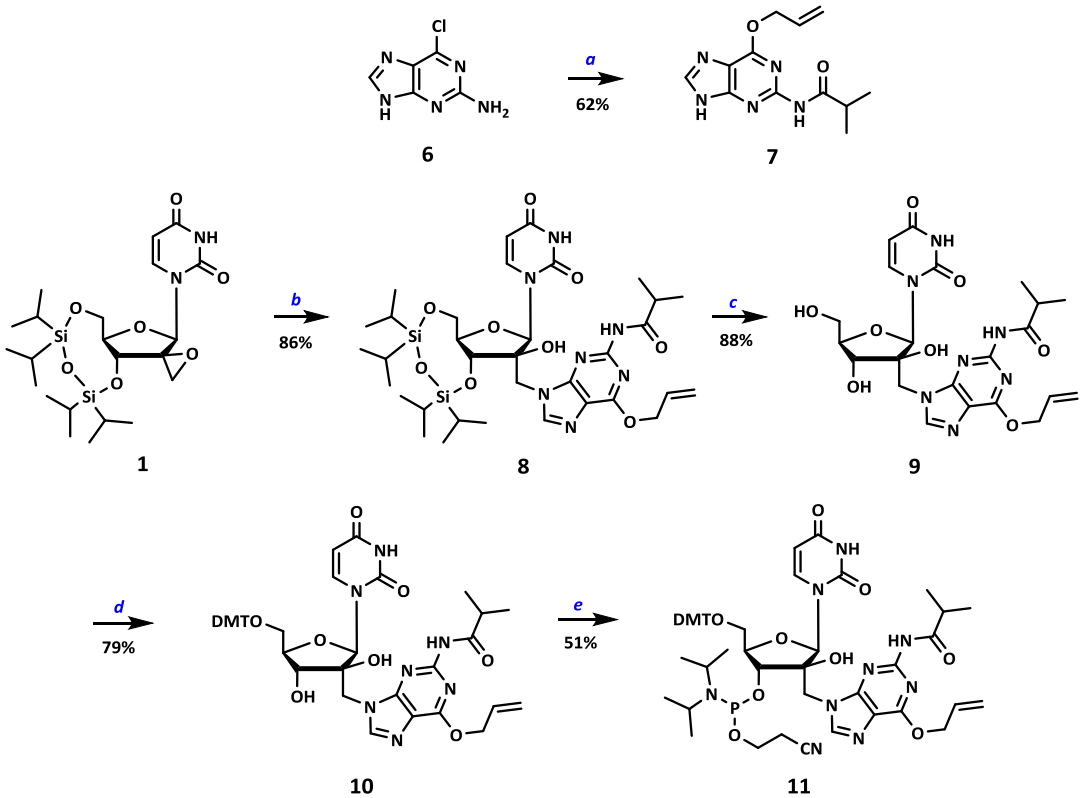
1. Experimental details

1.1 General synthetic details

All chemicals were used as supplied except CH_2Cl_2 (which was distilled) and DMTCl (which was treated with acetyl chloride in anhydrous cyclohexane, degassed and recrystallized from Et_2O). Dry chemicals were prepared using a selection of methods: CH_3OH was freshly distilled from Mg and stored over 3 Å molecular sieves. THF was freshly distilled from Na (+ benzophenone indicator) and stored over 4 Å sieves. CH_2Cl_2 was freshly distilled and stored over 4 Å sieves. Triethylamine and toluene were passed over a column of silica and stored over 4 Å sieves. Petroleum ether was dried over NaH, filtered and stored over 4 Å sieves. DCE was dried over Al_2O_3, filtered and stored over 4 Å sieves. CH_3CN was dried over 3 Å sieves, and pyridine, xylenes and DIPEA were dried over 4 Å sieves. DMF and DMSO were bought in anhydrous form. All reactions were monitored by TLC using silica gel 60 F_{254} coated aluminum-backed plates, which were visualized at 254 nm and/or by staining with 5% H_2SO_4 in EtOH. Flash chromatography was performed using silica gel 60 (particle size 0.040–0.063 mm). After chromatography, appropriate fractions were combined and concentrated. All products were dried under high vacuum (<1 mmHg) overnight to give products in high analytical purity. HRMS-ESI was recorded on a Bruker micrOTOF-Q II (ESI) quadrupole-time of flight instrument in positive ion mode with an accuracy of ±5 ppm. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a Bruker Avance III 400 instrument at 400.12 MHz, 100.62 MHz and 161.97 MHz, respectively. Chemical shifts are reported in ppm relative to tetramethylsilane ($\delta_{H,C}$ 0 ppm) or the deuterated solvents (DMSO-d_6 δ_c 39.5 ppm, $CDCl_3$ δ_c 77.2 ppm, CD_3OD δ_c 49.0). For ^{31}P NMR spectra, 85% H_3PO_4 was used as external standard. 2D spectra (HSQC, COSY and HMBC) have been used in assigning ^{1}H and ^{13}C NMR signals. NOESY or ROESY have been used in establishing stereochemistry whenever appropriate. Nucleoside phosphoramidites were treated as reactive intermediates, and their identities are based solely on ^{31}P NMR spectroscopy and HRMS-ESI.



Scheme A1. *Reagents and conditions:* (a) NaHMDS, N^{Bz}-cytosine, THF, 55 °C, 30 h. (b) TBAF, THF, rt, 4 h. (c) DMTCl, pyridine, 35 °C, 48 h. (d) 2-Cyanoethyl-*N*,*N*,*N*',*N*'-tetraisopropylphosphordiamidite, diisopropyl tetrazolide, CH_2Cl_2, rt, 18 h.



Scheme A2. *Reagents and conditions:* (*a*) *i.* DABCO, DMSO, rt, 6 h. *ii.* allyl alcohol, DMSO, rt, 24 h. *iii.* isobutyric anhydride, DMA, 130 °C, 3 h; (*b*) NaH, O^6-allyl-2-isobutyramido-9*H*-purine, DMF, 35 °C, 18 h. (*c*) TBAF, THF, rt, 2 h. (*d*) DMTCl, pyridine, rt, 36 h. (*e*) 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, rt, 12 h.

1-(2′-*C*-(4-*N*-Benzoylcytosin-1-yl)methyl-3′,5′-*O*-(1,1,3,3-tetraisopropyl-1,3-disiloxane-1,3-diyl)-β-D-arabinofuranosyl)uracil (2)

A solution of NaHMDS in THF (1 M, 1.5 mL, 1.50 mmol) was added dropwise to a suspension of 4-*N*-benzoylcytosine (0.38 g, 1.80 mmol) in anhydrous THF (10 mL). The mixture was stirred at 55 °C for 1 h. A solution of 3′,5′-*O*-(tetraisopropyldisiloxane-1,3-diyl)uridine-2′-spiroepoxide **1** (0.49 g, 1.00 mmol) in anhydrous THF (10 mL) was added dropwise, and the mixture was stirred at 55 °C for 30 h. After cooling to rt, a saturated aqueous solution of NH_4Cl (30 mL) was added followed by H_2O (30 mL) and CH_2Cl_2 (100 mL). The aqueous phase was separated and extracted with CH_2Cl_2 (2 × 30 mL). The combined organic phases were dried ($MgSO_4$), and concentrated under reduced pressure. The crude material was purified by flash chromatography (0—5 % CH_3OH in CH_2Cl_2) to give compound **2** (0.56 g, 0.78 mmol) as a white foam. Yield: 80%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 9.71 (br, 1H, NH), 9.53 (br, 1H, NH), 7.92 (d, J = 7.5 Hz, 2H, Bz), 7.80 (d, J = 5.0 Hz, 1H, ^{C}H6), 7.56 (t, J = 7.5 Hz, 1H, Bz), 7.55 (d, J = 5.0 Hz, 1H, ^{C}H5), 7.52 (d, J = 8.2 Hz, 1H, ^{U}H6), 7.46 (t, J = 7.5 Hz, 2H, Bz), 6.20 (br, 1H, 2′-OH), 5.92 (s, 1H, H1′), 5.59 (d, 1H, J = 8.2, ^{U}H5), 4.41 (d, J = 14.5 Hz, 1H, 2′-CH_2a), 4.29 (d, J = 14.5 Hz, 1H, 2′-CH_2b), 4.25 (d, J = 5.2 Hz, 1H, H3′), 4.12 (dd, 1H, J = 13.0 Hz, 1.8 Hz, H5′), 4.00 (dd, 1H, J = 13.0 Hz, 2.3 Hz, H5′), 3.88—3.79 (m, 1H, H4′), 1.13—1.02 (m, 28H, TIPDS). ^{13}C NMR (101 MHz, $CDCl_3$): δ 163.2, 163.1 (^{C}C4, ^{U}C4), 151.1 (^{U}C2, ^{C}C2), 140.8 (^{U}C6, ^{C}C6), 133.0, 132.9, 128.8, 128.0 (Bz), 101.8 (^{U}C5), 98.0 (^{C}C5), 85.6 (C1′), 80.2 (C2′, C4′), 77.3 (C3′), 60.2 (C5′), 55.9 (2′-CH_2), 17.4, 17.31, 17.30, 17.22, 17.17, 17.0, 16.9, 13.4, 13.0, 12.8, 12.4 (TIPDS). HRMS-ESI: calcd. for $C_{33}H_{48}N_5O_9Si_2^+$ m/z 714.2985, found 714.3008.

1-(2′-*C*-(4-*N*-benzoylcytosin-1-yl)methyl-β-D-arabinofuranosyl)uracil (3)

To a solution of compound **2** (0.72 g, 1.01 mmol) in anhydrous THF (25 mL) was added TBAF (1 M solution in THF, 1.50 mL, 1.50 mmol), and the reaction mixture was stirred at rt for 4 h. The mixture was concentrated under reduced pressure, and the crude material was purified by flash chromatography (0—20 % CH_3OH in CH_2Cl_2) to give compound **3** (0.43 g, 0.91 mmol) as a colourless oil. Yield: 90%. ^{1}H NMR (400 MHz, DMSO-d_6): δ 11.34 (br, 2H, NH), 8.13 (d, 1H, J = 7.3 Hz, ^{C}H6), 8.02 (d, 2H, J = 7.5 Hz, Bz), 7.70 (d,

1H, *J* = 8.2 Hz, ^{U}H6), 7.64 (t, 1H, *J* = 7.5 Hz, Bz), 7.52 (t, 2H, *J* = 7.5 Hz, Bz), 7.40 (d, 1H, *J* = 7.3 Hz, ^{C}H5), 6.26 (s, 1H, 2′-OH), 6.08 (s, 1H, 3′-OH), 6.00 (s, 1H, H1′), 5.63 (dd, 1H, *J* = 8.1, 2.0 Hz, ^{U}H5), 5.46 (t, 1H, *J* = 5.1 Hz, 5′-OH), 4.28 (d, 1H, *J* = 14.3 Hz, 2′-CH_2*a*), 3.95—3.92 (m, 1H, H4′), 3.90 (d, *J* = 14.3 Hz, 2′-CH_2*b*), 3.82 (t, *J* = 2.8 Hz, 1H, H3′), 3.71—3.61 (m, 2H, H5′). ^{13}C NMR (101 MHz, DMSO-d_6): *δ* 167.2 (Bz), 163.3, 163.1 (^{C}C4, ^{U}C4), 157.4 (^{C}C2), 152.4 (^{C}C6), 150.6 (^{U}C2), 141.9 (^{U}C6), 132.9, 132.7, 128.4, 128.3 (Bz), 100.5 (^{U}C5), 96.5 (^{C}C5), 85.0 (C1′), 84.6 (C4′) 80.8 (C2′), 75.1 (C3′), 60.8 (C5′), 50.0 (2′-CH_2). HRMS-ESI: calcd. for $C_{21}H_{22}N_5O_8^+$ *m/z* 472.1463, found 472.1439.

1-(2′-*C*-(4-*N*-Benzoylcytosin-1-yl)methyl-5′-*O*-(4,4′-dimethoxytrityl)-β-D-arabinofuranosyl)uracil (4)

Compound **3** (0.47 g, 1.00 mmol) was azeotropically dried with anhydrous pyridine (3 × 10 mL) and dissolved in anhydrous pyridine (15 mL). DMTCl (0.425 g, 1.25 mmol) was added to the magnetically stirred solution under argon, and the reaction mixture was stirred at 35 °C for 48 h. Ethanol (99.9%, 0.5 mL) was added and the solvent was removed under reduced pressure. The crude material was redissolved in CH_2Cl_2 (45 mL) and washed with a saturated aqueous solution of $NaHCO_3$ (3 × 20 mL). The organic phase was dried ($MgSO_4$) and concentrated under reduced pressure, and the crude material was purified by flash chromatography (0—8 % CH_3OH in CH_2Cl_2) to give compound **4** (0.60 g, 0.78 mmol) as a slightly yellow solid. Yield: 78%. ^{1}H NMR (400 MHz, $CDCl_3$): *δ* 9.42 (br, 1H, NH), 9.07 (br, 1H, NH), 7.90 (d, 2H, *J* = 7.5, Bz), 7.82 (d, *J* = 6.9 Hz, 1H, ^{C}H6), 7.61 (t, *J* = 7.5 Hz, 1H, Bz), 7.60 (d, *J* = 6.9 Hz, 1H, ^{C}H5), 7.54 (d, *J* = 8.0 Hz, 1H, ^{U}H6), 7.51 (t, *J* = 7.5 Hz, 2H, Bz), 7.32—7.15 (m, 13 H, DMT), 6.79 (d, *J* = 8.7 Hz, 4H, DMT), 6.23 (br, 1H, 2′-OH), 6.14 (s, 1H, H1′), 5.60 (d, 1H, *J* = 8.0, ^{U}H5), 5.34 (br, 1H, 3′-OH), 4.76 (d, *J* = 14.6 Hz, 1H, 2′-CH_2*a*), 4.25 (s, 1H, H4′), 3.77 (s, 1H, C3′), 3.76 (s, 6H, DMT), 3.67 (d, *J* = 10.3 Hz, 1H, H5′*a*), 3.58—3.49 (m, 2H, H5′*b*, 2′-CH_2*b*). ^{13}C NMR (101 MHz, $CDCl_3$): *δ* 163.3 (^{C}C4), 163.0 (^{U}C4), 158.9 (DMT), 150.7 (^{U}C2), 143.3, 142.1 (^{U}C6, DMT), 137.9, 134.4, 133.4, 132.8, 130.0, 129.1, 129.0, 128.2, 128.1, 128.0, 127.7, 127.4, 125.3 (^{C}C2, ^{C}C6, Bz), 113.4 (DMT), 101.2 (^{U}C5), 97.0 (^{C}C5) , 88.6 (DMT), 86.1 (C1′), 84.3 (C4′), 81.7 (C2′), 75.8 (C3′), 63.7 (C5′), 55.3 (DMT), 50.1 (2′-CH_2). HRMS-ESI: calcd. for $C_{42}H_{40}N_5O_{10}^+$ *m/z* 774.2778, found 774.2758.

1-(2′-*C*-(4-*N*-Benzoylcytosin-1-yl)methyl-3′-*O*-(*P*-(2-cyanoethyl)-*N*,*N*-diisopropylaminophosphinyl)-5′-*O*-(4,4′-dimethoxytrityl)-β-D-arabinofuranosyl)uracil (5)

Compound **4** (100 mg, 130 μmol) was azeotropically dried with anhydrous DCE (3 × 5 mL) and dissolved in CH_2Cl_2 (3 mL). 2-Cyanoethyl-*N*,*N*,*N*',*N*'-tetraisopropylphosphordiamidite (0.83 mL, 0.26 mmol) and diisopropyl tetrazolide (67 mg, 0.39 mmol) were added under argon to the magnetically stirred solution. The reaction mixture was stirred at rt for 18 h. Ethanol (99.9%, 0.5 mL) was added and the solution was concentrated under reduced pressure. The crude material was purified by flash chromatography (0—10 % CH_3OH in CH_2Cl_2). The purified material was precipitated from CH_2Cl_2—petroleum ether to give compound **5** (91 mg, 93 μmol) as a white solid. Yield: 72% (contains small amounts of a H-phosphonate impurity at 14 ppm). ^{31}P NMR (162 MHz, $CDCl_3$): δ 151.81, 150.99. HRMS-ESI: calcd. for $C_{51}H_{56}N_7O_{11}PNa^+$ *m*/*z* 996.3698, found 996.3402.

6-*O*-Allyl-2-isobutyramido-9*H*-purine (7)

To a magnetically stirred solution of 2-amino-6-chloropurine **6** (500 mg, 2.95 mmol) in anhydrous DMSO (3 mL) was added DABCO (1.82 g, 16.22 mmol). The reaction mixture was stirred for 6 h at rt, upon which the reaction mixture was filtered and rinsed with 5 mL DMSO to give a white solid. The solid was suspended in anhydrous DMSO (6 mL), and to this stirred suspension was slowly added a solution of allyl alcohol (1.10 mL, 16.22 mmol) in DMSO (3 mL), which had been pretreated with NaH (60% suspension in mineral oil, 236 mg, 5.90) for 10 min. The reaction mixture was stirred for 24 h at rt. Hereafter, H_2O (10 mL) was slowly added, and the mixture was extracted with EtOAc (5 × 20 mL). The combined organic phases were dried over Na_2SO_4 and concentrated to near dryness under reduced pressure. The crude material was purified by flash chromatography (0—15% CH_3OH in CH_2Cl_2) to obtain the O^6-allyl protected intermediate as a white solid (465 mg, 2.43 mmol). Yield: 82%. ^1H NMR (400 MHz, DMSO-d_6): δ 12.41 (br, 1H, NH), 7.81 (s, 1H, H8), 6.25 (br, 2H, NH_2), 6.18 (ddt, *J* = 17.2, 10.5, 5.6 Hz, 1H, allyl C*H*=CH_2), 5.42 (d, *J* = 17.2 Hz, 1H, allyl

CH=C*H*$_2$), 5.27 (dd, *J* = 10.5, 1.5 Hz, 1H, allyl CH=C*H*$_2$), 4.94 (d, *J* = 5.6 Hz, 1H, allyl *O*-C*H*$_2$). ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.6 (C6, C2), 155.1 (C4), 137.7 (C8), 133.4 (allyl *C*H=*C*H$_2$), 117.8 (allyl CH=*C*H$_2$), 113.4 (C5), 65.7 (allyl O-*C*H$_2$). The 6-*O*-allyl protected intermediate (250 mg, 1.31 mmol) was dissolved in anhydrous DMA (5 mL), and isobutyric anhydride (0.65 mL, 3.92 mmol) was added to the magnetically stirred solution. The reaction mixture was heated to 130 °C for 3 h and then concentrated under reduced pressure. The crude material was purified by flash chromatography (0—100% EtOAc in petroleum ether) to obtain compound **7** (261 mg, 1.00 mmol) as a white solid. Yield: 76%. ^{1}H NMR (400 MHz, DMSO-d_6): δ 13.18 (br, 1H, NH), 10.25 (br, 1H, NH), 8.18 (s, 1H, H8), 6.18 (ddd, *J* = 16.5, 10.5, 5.7 Hz, 1H, allyl C*H*=CH$_2$), 5.47 (d, *J* = 16.5 Hz, 1H, allyl CH=C*H*$_2$), 5.31 (d, *J* = 10.5 Hz, 1H, allyl CH=C*H*$_2$), 5.06 (d, *J* = 5.7 Hz, 1H, allyl *O*-C*H*$_2$), 2.88 (sept, *J* = 6.5 Hz, 1H, iBu), 1.09 (d, *J* = 6.5 Hz, 6H, iBu). ^{13}C NMR (101 MHz, $CDCl_3$): δ 174.9 (iBu), 159.3 (C6), 153.8 (C4), 151.8 (C2), 140.9 (C8), 133.1 (allyl *C*H=*C*H$_2$), 118.4 (allyl CH=*C*H$_2$), 116.8 (C5), 66.7 (allyl *O*-C*H*$_2$), 34.2 (iBu), 19.3 (iBu). HRMS-ESI: calcd. for $C_{12}H_{15}N_5NaO_2^+$ *m/z* 284.1118, found 284.1128.

1-(2′-*C*-(6-*O*-Allyl-2-isobutyramido-9*H*-purin-9-yl)methyl-3′,5′-*O*-(1,1,3,3-tetraisopropyl-1,3-disiloxane-1,3-diyl)-β-D-arabinofuranosyl)uracil (8)

A mixture of compound **7** (63 mg, 241 μmol) and NaH (a 60% suspension in mineral oil, 12 mg, 300 μmol) in anhydrous DMF (1 mL) was stirred under argon at rt for 30 min. To this mixture was added a solution of 3′,5′-*O*-(tetraisopropyldisiloxane-1,3-diyl)uridine-2′-spiroepoxide **1** (100 mg, 200 μmol) in anhydrous DMF (1 mL). The mixture was heated to 35 °C and stirred for 18 h. After cooling to rt, saturated aqueous NH_4Cl (1 mL) was added followed by H_2O (1 mL) and EtOAc (2 mL). The layers were separated, and the aqueous phase was extracted with EtOAc (3 × 2 mL). The combined organic phases were dried over Na_2SO_4 and concentrated under reduced pressure. The crude material was purified by flash chromatography (0—90% EtOAc in petroleum ether) to give compound **8** (131 mg, 172 μmol) as a white solid. Yield: 86%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H, ^{U}NH), 8.17 (s, 1H, ^{G}NH), 7.83 (s, 1H, ^{G}H8), 7.63 (d, *J* = 8.2 Hz, 1H, ^{U}H6), 6.11—5.98 (m, 1H, allyl C*H*=CH$_2$), 5.96 (s, 1H, H1′), 5.52 (d, *J* = 8.2 Hz, 1H, ^{U}H5), 5.38 (dd, *J* = 17.2, 1.2 Hz, 1H, allyl CH=C*H*$_2$),

5.23 (dd, *J* = 10.5, 1.2 Hz, 1H, allyl CH=C*H*$_2$), 4.98—4.93 (m, 1H, allyl *O*-C*H*$_2$), 4.67 (d, *J* = 15.2 Hz, 1H, 2′-CH*a*), 4.49—4.39 (m, 2H, 2′-CH*b*, H4′), 4.09—4.05 (m, 2H, H5′), 3.85—3.79 (m, 1H, H3′), 2.76—2.61 (m, 1H, iBu), 2.13 (s, 1H, 2′-OH), 1.26 (d, *J* = 6.8 Hz, 3H, iBu), 1.25 (d, *J* = 6.8 Hz, 3H, iBu), 1.14—0.97 (m, 28H, TIPDS). ^{13}C NMR (101 MHz, $CDCl_3$): *δ* 174.9 (iBu), 162.5 (^{U}C4), 160.9 (^{G}C6), 152.9 (^{G}C2), 151.1, 149.5 (^{U}C2, ^{G}C4), 143.2 (^{G}C8), 142.7 (^{U}C6), 132.0 (allyl *C*H=*C*H$_2$), 118.8, 118.7 (^{G}C5, allyl CH=*C*H$_2$), 100.8 (^{U}C5), 84.8 (C1′), 82.2 (C3′), 80.6 (C4′), 80.2 (C2′), 68.4 (allyl O-*C*H$_2$), 62.8 (C5′), 50.5 (2′-CH$_2$), 36.8 (iBu), 19.4 (iBu), 19.3 (iBu), 17.5, 17.4, 17.3, 17.2, 17.1, 17.0, 13.4, 13.2, 13.1, 12.6 (TIPDS). HRMS-ESI: calcd. for $C_{34}H_{53}N_7NaO_9Si_2^+$ *m/z* 782.3336, found 782.3314.

1-(2′-*C*-(6-*O*-Allyl-2-isobutyramido-9*H*-purin-9-yl)methyl-β-D-arabinofuranosyl)uracil (9)

To a solution of compound **8** (198 mg, 0.26 mmol) in anhydrous THF (5 mL) was added TBAF (1 M solution in THF, 0.65 mL, 0.65 mmol), and the solution was stirred at rt for 2 h. The solution was concentrated under reduced pressure, and co-evaporated with anhydrous toluene (10 mL). The crude material was purified by flash chromatography (10—20% CH_3OH in CH_2Cl_2) to give a rotameric mixture of compound **9** (119 mg, 230 μmol) as a white solid. Yield: 88%. *Rotamer I*: ^{1}H NMR (400 MHz, CD_3OD): *δ* 7.93 (s, 1H, ^{G}H8), 7.73 (d, *J* = 8.2 Hz, 1H, ^{U}H6), 6.26 (s, 1H, H1′), 6.18 (ddd, *J* = 17.2, 10.5, 5.7 Hz, 1H, allyl C*H*=CH$_2$), 5.59 (d, *J* = 8.2 Hz, 1H, ^{U}H5), 5.50 (dd, *J* = 17.2, 1.4 Hz, 1H, allyl CH=C*H*$_2$*a*), 5.32 (dd, *J* = 10.5, 1.4 Hz, 1H, allyl CH=C*H*$_2$*b*), 5.06 (d, *J* = 5.7 Hz, 2H, allyl *O*-CH$_2$), 4.83 (d, *J* = 15.1 Hz, 1H, 2′-CH*a*), 4.40 (d, *J* = 15.1 Hz, 1H, 2′-CH*b*), 4.26—4.20 (m, 1H, H4′), 3.89—3.84 (m, 1H, H3′), 3.84—3.72 (m, 2H, H5′), 2.76 (sept, *J* = 6.8 Hz, 1H, iBu), 1.24 (d, *J* = 5.0 Hz, 3H, iBu), 1.22 (d, *J* = 5.0 Hz, 3H, iBu). ^{13}C NMR (101 MHz, CD_3OD): *δ* 178.3 (iBu), 164.0 (^{U}C4), 161.7 (^{G}C6), 153.8, 153.6, 152.3 (^{U}C2, ^{G}C4, ^{G}C2), 146.2 (^{G}C8), 142.9 (^{U}C6), 133.8 (allyl *C*H=CH$_2$), 119.1 (allyl CH=*C*H$_2$), 118.6 (^{G}C5), 100.7 (^{U}C5), 87.9, 87.5 (C1′, C4′), 82.1 (C2′), 77.6 (C3′), 69.0 (allyl *O*-CH$_2$), 62.9 (C5′), 45.7 (2′-CH$_2$), 37.0 (iBu), 19.8 (iBu), 19.7 (iBu). HRMS-ESI: calcd. for $C_{22}H_{27}N_7O_8Na$ *m/z* 540.1813, found 540.1837. *Rotamer II*: ^{1}H NMR (400 MHz, CD_3OD): *δ* 8.09 (s, 1H, ^{G}H8), 7.84 (d, *J* = 8.2 Hz, 1H, ^{U}H6), 6.21 (s, 1H, H1′), 6.23—6.12 (m, 1H, allyl C*H*=CH$_2$), 5.67 (d, *J* = 8.2 Hz, 1H, ^{U}H5), 5.50 (dd, *J* = 17.2, 1.5 Hz, 1H, allyl CH=C*H*$_2$*a*), 5.31 (dd, *J* = 10.5,

1.5 Hz, 1H, allyl CH=C$\underline{H_2}$*b*), 5.06 (dt, *J* = 5.6 Hz, 2H, allyl *O*-CH_2), 4.75 (d, *J* = 14.8 Hz, 1H, 2′-CH*a*), 4.28 (d, *J* = 14.8 Hz, 1H, 2′-CH*b*), 4.17 (dt, *J* = 5.2, 2.6 Hz, 1H, H4′), 3.85—3.71 (m, 3H, H5′, H3′), 2.76 (sept, 1H, iBu), 1.22 (d, *J* = 6.9 Hz, 6H, iBu). HRMS-ESI: calcd. for $C_{22}H_{27}N_7O_8Na^+$ *m/z* 540.1813, found 540.1816.

1-(2′-*C*-(6-*O*-Allyl-2-isobutyramido-9*H*-purin-9-yl)methyl-5′-*O*-(4,4′-dimethoxytrityl)-β-D-arabinofuranosyl)uracil (10)

Compound **9** (94 mg, 182 μmol) was azeotropically dried with anhydrous pyridine (2 × 5 mL) and dissolved in anhydrous pyridine (5 mL). DMTCl (111 mg, 0.33 mmol) was added in one portion and the solution was stirred at rt for 36 h. EtOH (99.9%, 1 mL) was added, and the mixture was concentrated and co-evaporated with anhydrous toluene (3 mL). The residue was taken up in CH_2Cl_2 (10 mL) and washed with saturated aqueous $NaHCO_3$ (2 × 5 mL), dried over Na_2SO_4, and concentrated under reduced pressure. The crude material was purified by flash chromatography (0—10% CH_3OH in CH_2Cl_2) to give compound **10** (118 mg, 144 μmol) as a white foam. Yield: 79%. ^{1}H NMR (400 MHz, $CDCl_3$): *δ* 9.21 (br, 1H, ^{U}NH), 8.24 (br, 1H, ^{G}NH), 7.71 (s, 1H, ^{G}H8), 7.60 (d, *J* = 8.0 Hz, 1H, ^{U}H6), 7.42—7.04 (m, 9H, DMT), 6.79 (br, 1H, 3′-OH), 6.71 (dd, *J* = 8.9, 4.2 Hz, 4H, DMT), 6.34 (s, 1H, H1′), 6.11—5.91 (m, 1H, allyl C$\underline{H}$=CH_2), 5.62 (d, *J* = 8.0 Hz, 1H, ^{U}H5), 5.38 (d, *J* = 17.1 Hz, 1H, allyl CH=C$\underline{H_2}$), 5.27 (d, *J* = 10.5 Hz, 1H, allyl CH=C$\underline{H_2}$), 4.91 (s, 2H, 2′-CH*a*, 2′-CH*b*), 4.85 (d, *J* = 14.6 Hz, 1H, allyl *O*-CH*a*), 4.44—4.41 (m, 1H, H4′), 4.04 (d, *J* = 14.6 Hz, 1H, allyl *O*-CH*b*), 3.71 (s, 3H, DMT), 3.70 (s, 3H, DMT), 3.63—3.49 (m, 2H, H3′, H5′*a*), 3.43 (dd, *J* = 10.9, 4.7 Hz, 1H, H5′*b*), 2.85—2.69 (m, 1H, iBu), 1.29 (d, *J* = 6.9 Hz, 6H, iBu). ^{13}C NMR (101 MHz, $CDCl_3$): *δ* 163.1 (^{U}C4), 160.4 (^{G}C6), 158.7 (DMT), 152.4, 151.6, 151.0 (^{G}C2, ^{G}C4, ^{U}C2), 144.7, 143.7, 142.4, 134.9, 134.8, 131.7, 130.1, 128.1, 128.0, 127.2, 119.6 (DMT), 117.4 (^{G}C5), 113.3 (DMT), 101.2 (^{U}C5), 86.4 (C1′), 85.3 (C4′), 81.2 (C2′), 76.0 (C3′), 68.4 (2′-CH_2), 64.1 (C5′), 55.2 (DMT), 44.2 (allyl *O*-CH_2), 36.6 (iBu), 19.3 (iBu). HRMS-ESI: calcd. for $C_{43}H_{45}N_7NaO_{10}^+$ *m/z* 842.3120, found 842.3091.

1-(2′-*C*-(6-*O*-Allyl-2-isobutyramido-9*H*-purin-9-yl)methyl-3′-*O*-(*P*-(2-cyanoethyl)-*N*,*N*-diisopropyl-aminophosphinyl)-5′-*O*-(4,4′-dimethoxytrityl)-β-D-arabinofuranosyl)uracil (11)

Compound **10** (98 mg, 120 μmol) was azeotropically dried with anhydrous DCE (2 × 5 mL), and dissolved in anhydrous CH_2Cl_2 (3 mL). DIPEA (83 μL, 0.48 mmol) and 2-cyanoethyl *N*,*N*-diisopropylchloro-phosphoramidite (80 μL, 0.36 mmol) were added under argon to the magnetically stirred solution. The reaction mixture was stirred at rt for 12 h. Ethanol (99.9%, 0.5 mL) was added and the solution was concentrated under reduced pressure. The crude material was purified by flash chromatography (0—100% EtOAc in petroleum ether). The purified material was precipitated from CH_2Cl_2—petroleum ether to give compound **11** (62 mg, 61 μmol) as a white solid. Yield: 51% (contains small amounts of a H-phosphonate impurity at 14 ppm). ^{31}P NMR (162 MHz, $CDCl_3$): δ 151.18, 151.07. HRMS-ESI: calcd. for $C_{52}H_{62}N_9O_{11}PNa^+$ *m/z* 1042.4199, found 1042.4164.

1.2 Oligonucleotides

Oligonucleotides were synthesized on an Expedite™ Nucleic Acid Synthesis System (Model 8909)[1] fully-automated DNA synthesizer in ~0.2 μmol scale loaded on 500 Å controlled-pore glass (CPG) supports using the phosphoramidite approach and following the manufacturer's protocol. Double coupling (2 × 5 min) cycles were used for commercial phosphoramidites and prolonged coupling times (20 minutes) were used for the modified phosphoramidites. The phosphoramidites were activated using 1*H*-tetrazole and incorporated into oligonucleotides via manual couplings: 10 μmol of the modified phosphoramidite was dissolved in anhydrous CH_3CN (2 mL), treated with a solution of the activator, and infused into the reaction compartment. The stepwise coupling efficiencies were monitored by measuring the absorbance of the trityl cation at 495 nm, which in in general were between 95—100% for the commercial phosphoramidites, and 90—100% for the modified phosphoramidites. The final 5′-terminal DMT group in the oligonucleotides was kept on for purification purposes.

The final crude oligonucleotides on solid support were treated with aq. NH_3 (28%, 1 mL) at rt for 24 h. For oligonucleotides containing $\mathbf{U_G}$, the oligonucleotide was pretreated with a mixture of $Et_2NH{\cdot}HCO_3$ (18 mg), $Pd(PPh_3)_4$ (18 mg) and PPh_3 (18 mg) in CH_2Cl_2 (1 mL) at 50 °C for 36 h, followed by decantation and thorough washings with CH_2Cl_2, acetone, and H_2O. After cleavage from the support, the oligonucleotides were filtered and the filtrates were evaporated to dryness at 45 °C by a steady N_2 flow, and dissolved in an aqueous triethylammonium acetate buffer (500 μL, 0.05 M, pH 7.4). Analytically pure oligonucleotides were obtained by reversed-phase HPLC purification on a Waters 600 system using Xterra MS C18 10 μm (7.8 × 50 mm) columns and Xterra MS C18 10 μm (7.8 × 10 mm) precolumns. Elution was performed with 100% eluent A over 2 min, followed by a linear gradient down to 30% eluent A over 38 min, then a wash with 100% eluent B over 10 min, 100% eluent A over 10 min. (Eluent A = triethylammonium acetate (0.05 M, pH 7.4). Eluent B = 75% CH_3CN/H_2O (3:1, v/v)). The pure fractions were pooled and evaporated at 45 °C. The 5′-terminal DMT group was removed by treatment with acetic acid (80% in H_2O, 100 μL) for 30 min, upon which an aqueous solution of NaOAc (15 μL, 3 M), an aqueous solution of $NaClO_4$ (15 μL, 5 M), and pure

[1] PerSeptive Biosystems Inc., 500 Old Connecticut Path Framingham, MA, United States.

acetone (1 mL) were added. The oligonucleotides precipitated overnight at –20 °C. The supernatant was removed from the sedimented solid (centrifugation, 12,000 rpm, 10 min at 2 °C), and the remaining pellet was washed with cold acetone (3 × 1 mL) and dissolved in 500 µL Milli-Q® H_2O.

Mass spectra of the oligonucleotides were recorded on a MALDI-TOF mass spectrometry instrument in ES^+ mode using a MSP AnchorChip 600/96 microScout Target (Bruker Daltonics) with a matrix consisting of 3-hydroxypicolinic acid (3HPA) and diammonium hydrogen citrate. All oligonucleotides were obtained in >85% purity and the determined *m*/*z* values were within ±4 amu (i.e. ~1‰) of the calculated masses. The concentrations of the purified oligonucleotides were determined by the optical density at 260 nm, assuming the molar absorptivities of the oligonucleotides equal the sum of each constituent nucleotide monomers. The extinctions coefficients ε_{260} of the double-headed nucleotides were approximated as the sum of the two constituent nucleobases.

1.3 T_m measurements

The duplex samples consisted of 1.5 µM concentrations of each oligonucleotide in medium salt buffer (Na_2HPO_4 (2.5 mM), NaH_2PO_4 (5 mM), NaCl (100 mM), EDTA (0.1 mM), pH 7). The oligonucleotides were annealed by heating the sample to 80 °C followed by a slow cooling to 10 °C. The increase in absorbance at 260 nm as a function of temperature from 10 °C to 75 °C or 80 °C (1 °C/min) was recorded on a PerkinElmer Lambda 35 UV/Vis spectrometer using a PTP-6 Peltier Temperature Programmer. The listed T_m values are averages of at least duplicate readings within 0.5 °C, as determined from the local maximum of the first derivatives of the absorbance vs temperature curves. All melting curves were sharp and well-defined, and found to be reversible upon cooling.

1.4 Modelling

MD simulations were carried out using the *pmemd* module within the AMBER suite[2] using the ff99-bsc0[3] parameters for nucleic acids and ions. The $\mathbf{U_A}$ nucleoside was built in *xleap* and geometry optimized at the HF/6-31G(d) level using Gaussian.[4] The backbone of the optimized nucleoside remained in a duplex-like conformation throughout the geometry optimization. Atomic charges were calculated using the RESP methodology[5] keeping the native charges from the force field of Cornell et al. for backbone and sugar atoms except for C1′, C2′, H1′ and H2′[6] to keep consistency with the charge derivation for the native atomic charges. In the building of modified duplexes, all starting coordinates were generated using idealized B-DNA geometries. In *leap*, net-neutralizing Na^+ ions were added and the whole system was surrounded by a truncated octahedron of TIP3P waters[7] with a minimum distance of 10 Å from the helix to the edges of the box. For modified systems, an initial energy minimization was performed with only the $\mathbf{U_A}$ residue and the 3′-following nucleotide free to move. Harmonic positional restraints with a force constant of 500 kcal mol^{-1} $Å^{-2}$ were applied to the remainder of the system. This energy minimization relieved the strain on the modified backbone. Two further energy minimizations were performed, the first of solvent and ions using harmonic positional restraints with a force constant of 500 kcal mol^{-1} $Å^{-2}$ on the entire duplex, and the second with the entire system free to move. In the MD equilibration, the SHAKE algorithm[8] was applied with a 1 fs time step. The nonbonding cutoff was set to 9 Å and the particle mesh Ewald method[9] with default parameters was used to calculate long-range electrostatic interactions. The temperature of the system was raised from 0 to 300 K over 20 ps at constant volume using the Berendsen thermostat with a 0.2 ps coupling parameter and applying harmonic positional restraints with a force constant of 10 kcal mol^{-1} $Å^{-2}$ to

---

[2] Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D. M.; Seibel, G. L.; Kollman, P. A. *Comp. Phys. Commun.* **1995**, *91*, 1.

[3] Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

[4] Frisch, M. J. et al. Gaussian 03, Revision D.02 **2004**, Gaussian, Inc., Wallingford, CT.

[5] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

[6] (a) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179. (b) P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comp. Chem.* **1995**, *16*, 1357.

[7] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[8] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327

[9] Cheatham, T. E. III.; Miller, J. L.; Fox, T. Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.

the DNA molecules. Centre of mass movement was removed every 1000th step. Subsequently, 10 ps MD was carried out at 300 K at constant volume and temperature with parameters as in the first round of equilibration. The final step of equilibration consisted of 200 ps unrestrained isothermal–isobaric simulation at 300 K with a time step of 2 ps. The reference pressure was 1 atm and the barostat coupling parameter 1.0 ps and the thermostat coupling parameter 1.0 ps. MD production runs were carried out for 140 ns with parameters as for the last step of the equilibration except that a Langevin thermostat was used with a collision frequency of 1.0 ps^{-1}.

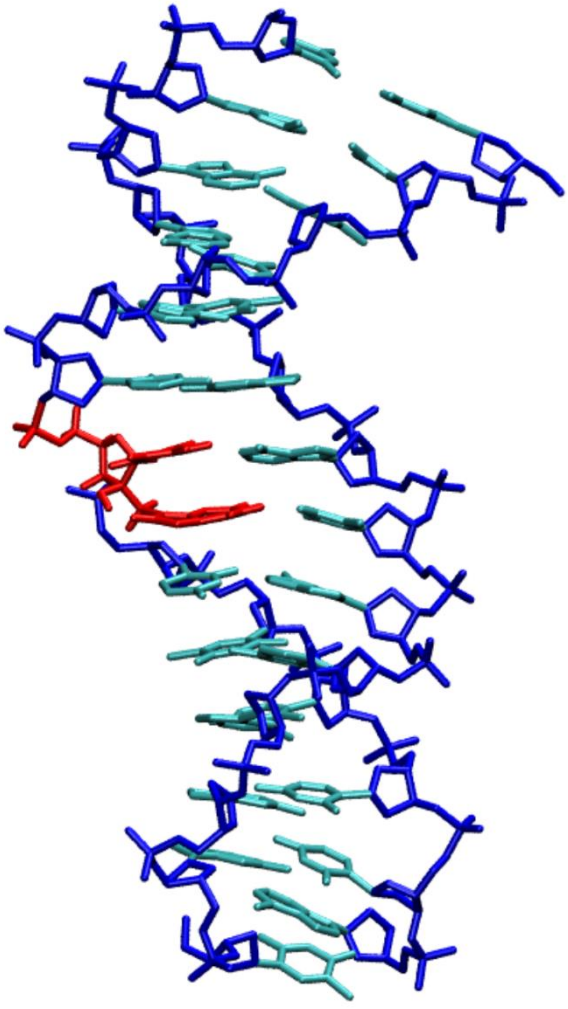


Fig. A1. Full-view snapshot of a DNA duplex from an MD simulation showing nucleobases (cyan), sugar-phosphate backbone (blue) and the $\mathbf{U_A}$ residue (red).

2. MALDI-TOF

Table A.1. Calculated and found MALDI-TOF *m*/*z* values for the synthesized oligonucleotides containing $\mathbf{U_C}$ and $\mathbf{U_G}$.

Sequence	**Calcd. *m*/*z***	**Found *m*/*z***
5′-dGCTCAC $\mathbf{U_C}$ CTCCCA	3957.6	3953.1
5′-dCGCA $\mathbf{U_C}$ ATTCGC	3416.3	3416.0
5′-dGCGAA $\mathbf{U_C}$ ATGCG	3508.3	3505.5
5′-dGCTCAC $\mathbf{U_G}$ CTCCCA	3994.7	3996.0
5′-dCGCA $\mathbf{U_G}$ ATTCGC	3457.2	3455.7
5′-dGCGAA $\mathbf{U_G}$ ATGCG	3546.3	3548.0

The corresponding *m*/*z* values for oligonucleotides containing $\mathbf{U_T}$ and $\mathbf{U_A}$ can be found in our preceding paper.[10]

[10] Kumar, P.; Kumar, P. K.; Nielsen, P. *J. Org. Chem.* **2014**, *79*, 11534.

3. IE-HPLC profiles

Analytical anion-exchange HPLC (IE-HPLC) was performed using a Merck-Hitachi Lachrom system equipped with a Dionex DNAPac® PA100 Analytical Oligonucleotide column (13 µm, 250 mm × 4 mm) heated to 60 °C. Elution was performed with an 10% isocratic hold of C, starting with a fixed 2% solution of A in B, followed by a linear gradient to a 30% solution of A in B in 23 min at a flow rate of 2.0 mL/min.

A = H_2O; B = $NaClO_4$ (1.0 M); C = TrisCl (0.25 M), pH 8.0

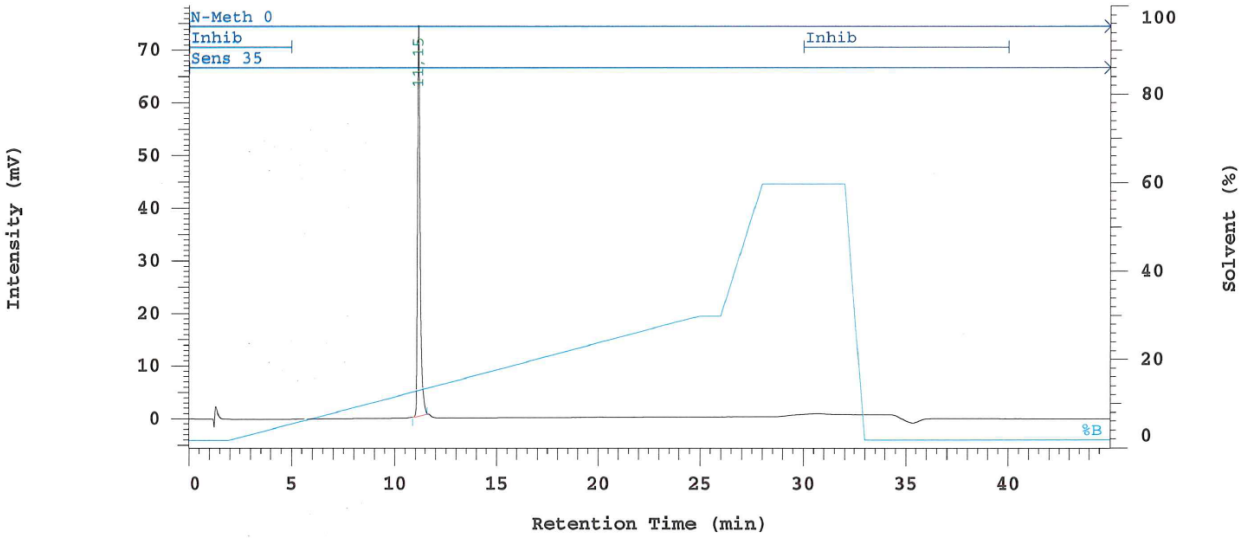


Fig A.2. HPLC profile of the DNA sequence 5′-dGCTCAC **U_C** CTCCCA

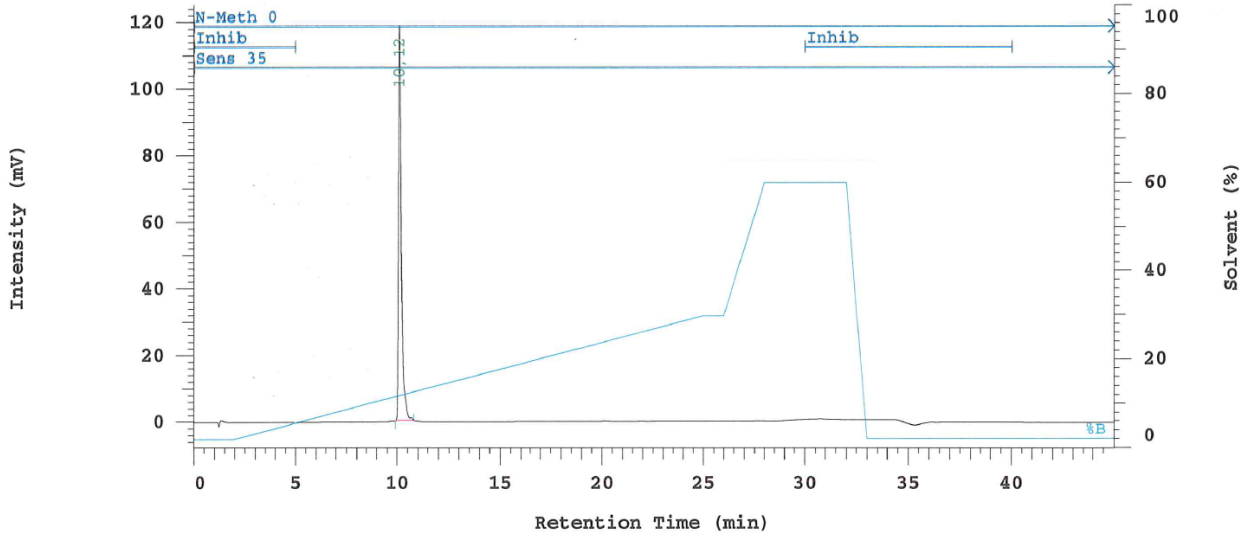


Fig A.3. HPLC profile of the DNA sequence 5′-dCGCA **U_C** ATTCGC

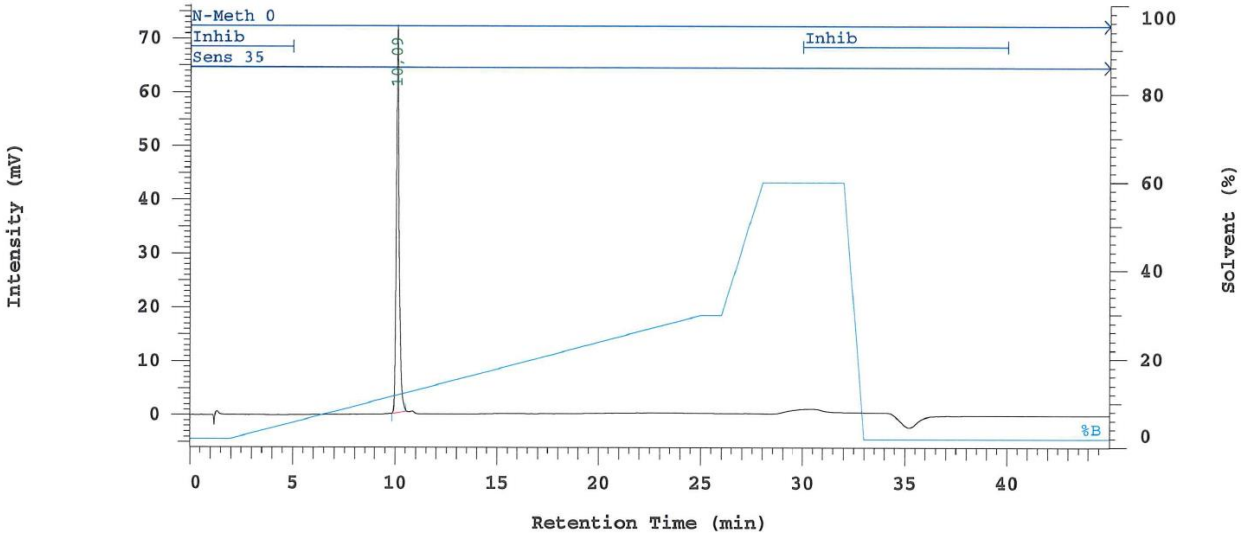


Fig A.4. HPLC profile of the DNA sequence 5′-dGCGAA $\mathbf{U_C}$ ATGCG

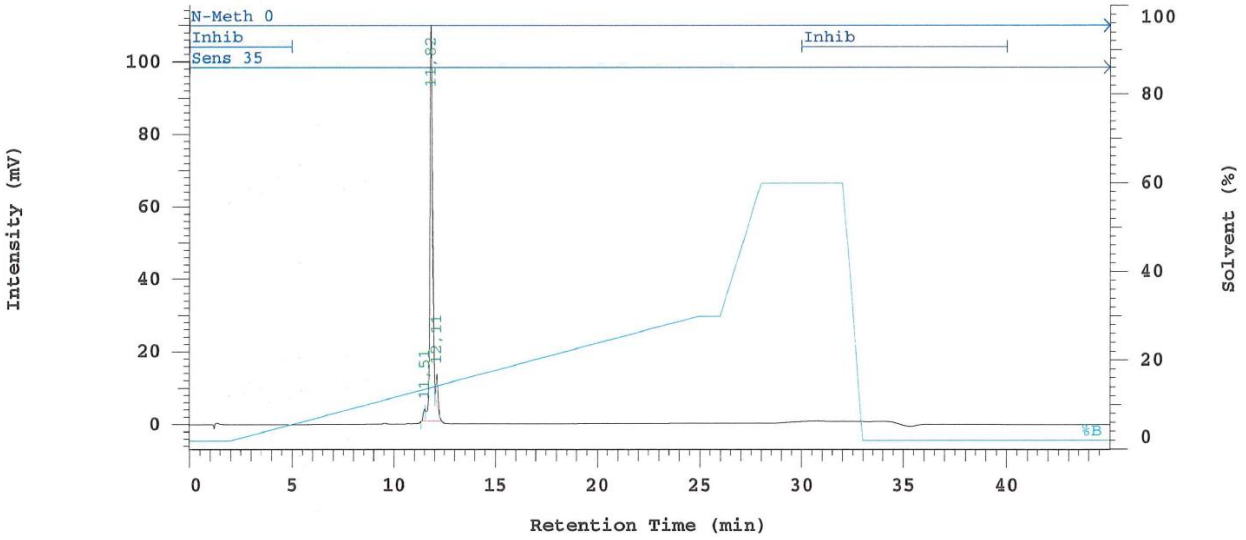


Fig A.5. HPLC profile of the DNA sequence 5′-dGCTCAC $\mathbf{U_G}$ CTCCCA

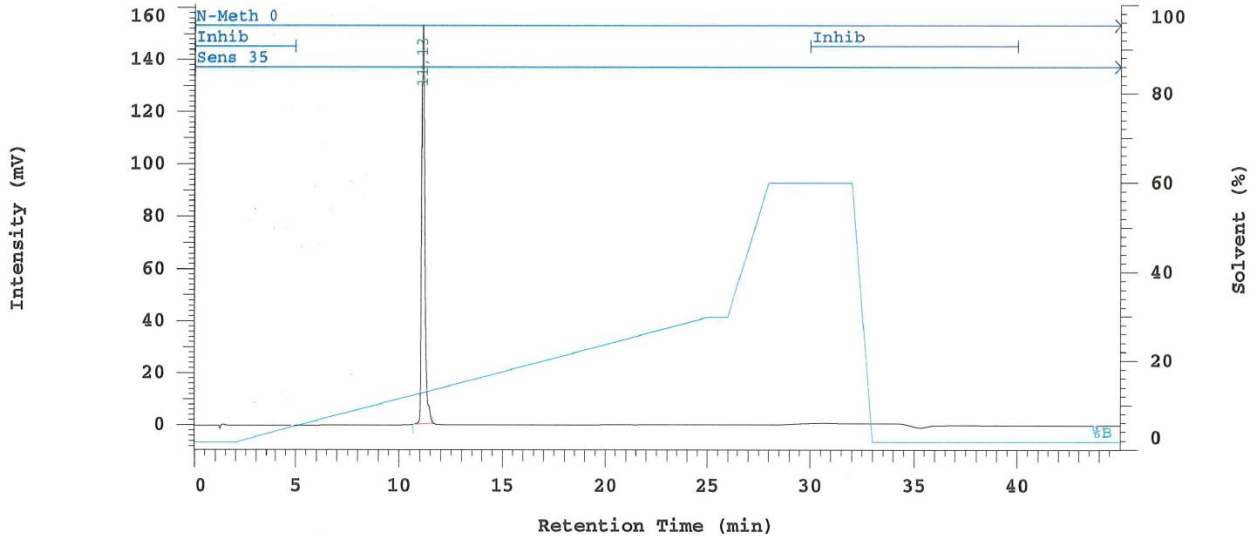


Fig A.6. HPLC profile of 5′-dCGCA $\mathbf{U_G}$ ATTCGC

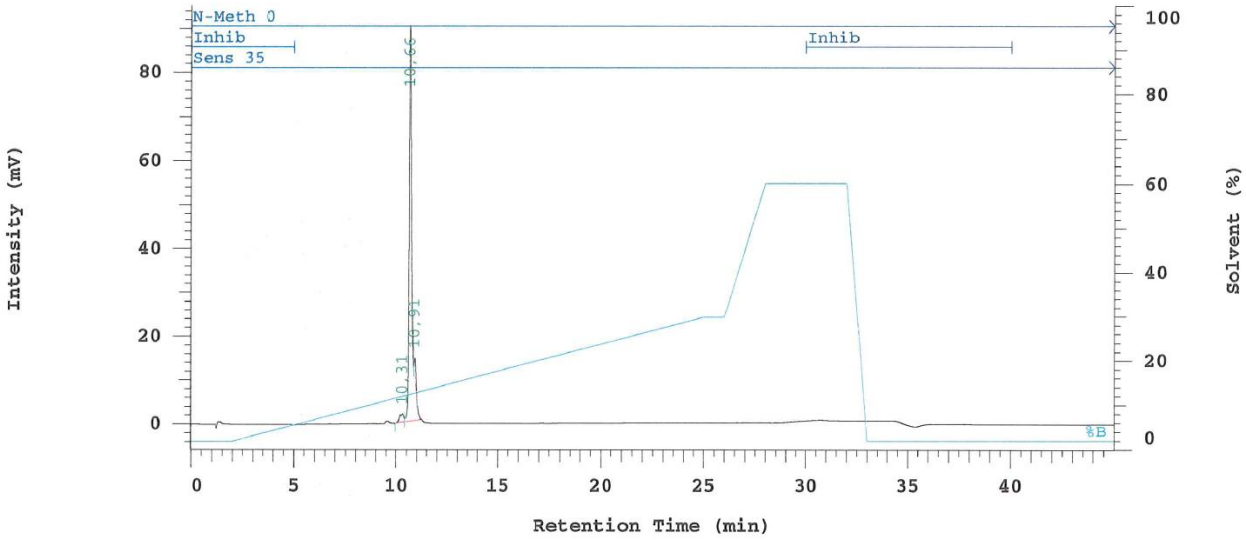
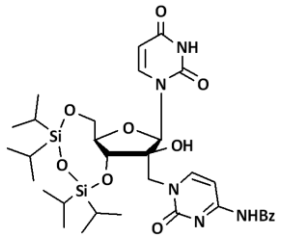


Fig A.7. HPLC profile of 5′-dGCGAA $\mathbf{U_G}$ ATGCG

4. NMR spectra



2

Fig A.8. 400 MHz ^{1}H NMR spectrum of compound **2** in $CDCl_3$.

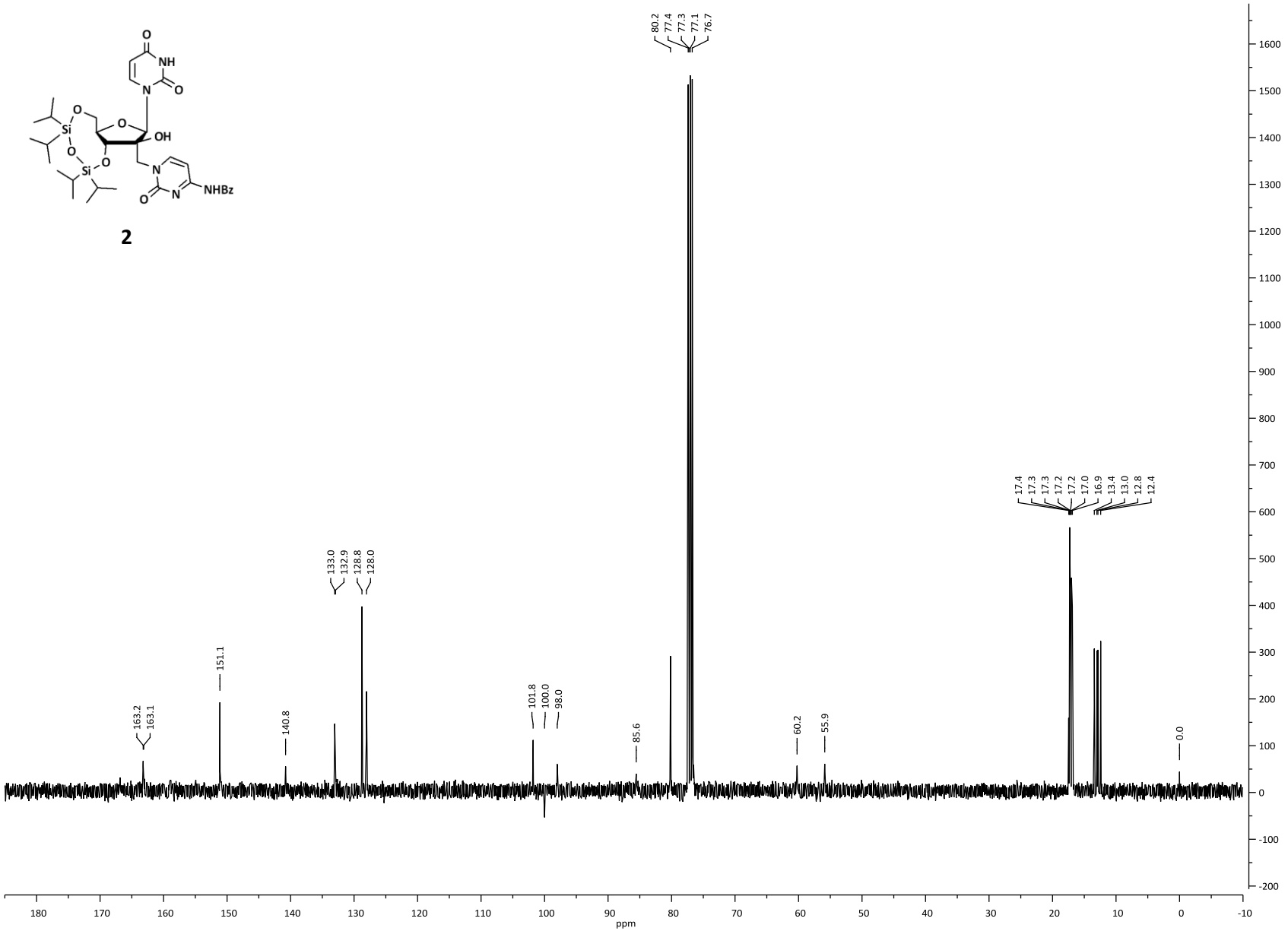


Fig A.9. 101 MHz ^{13}C NMR spectrum of compound **2** in $CDCl_3$.

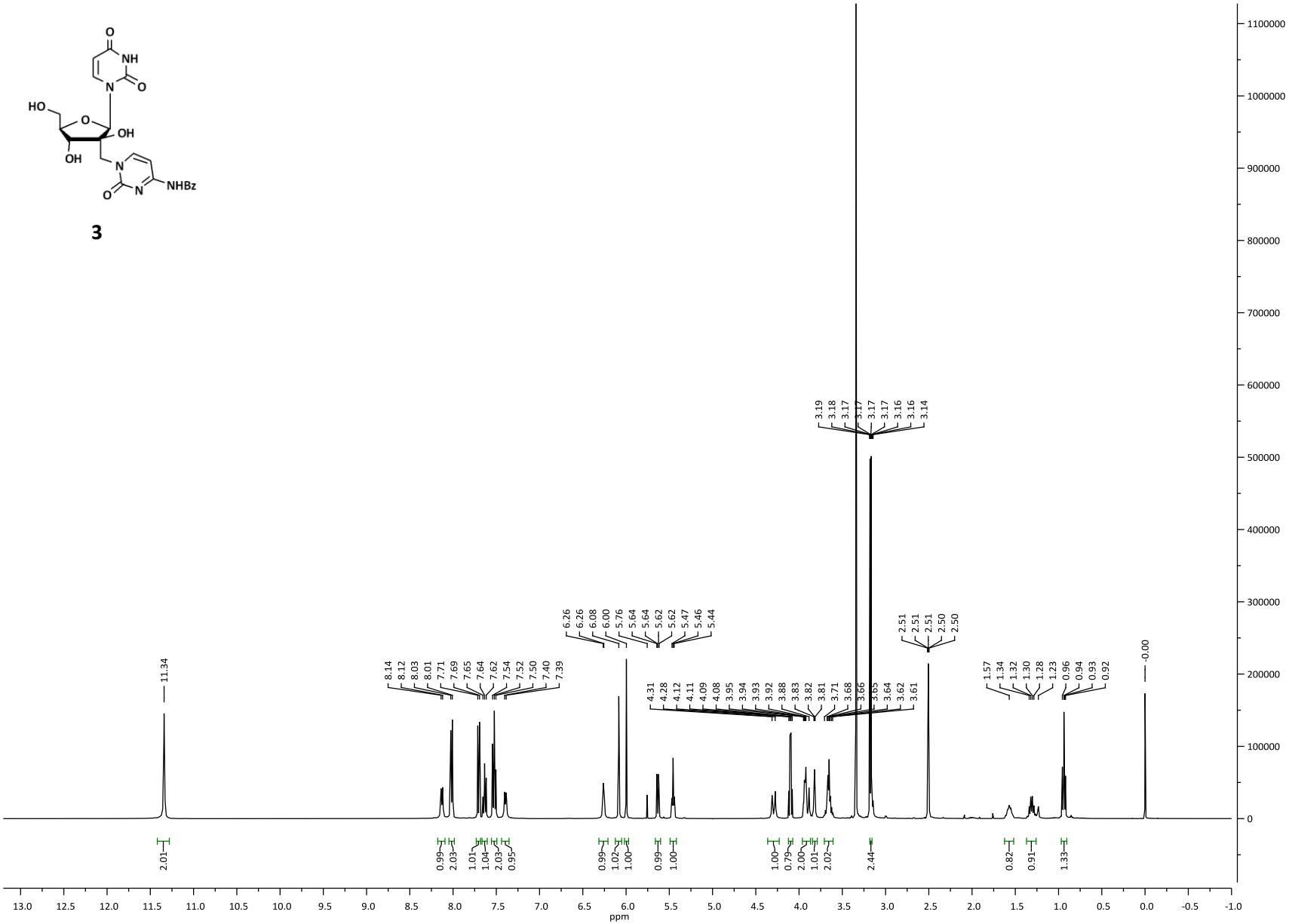


Fig A.10. 400 MHz ^{1}H NMR spectrum of compound **3** in DMSO-d_6.

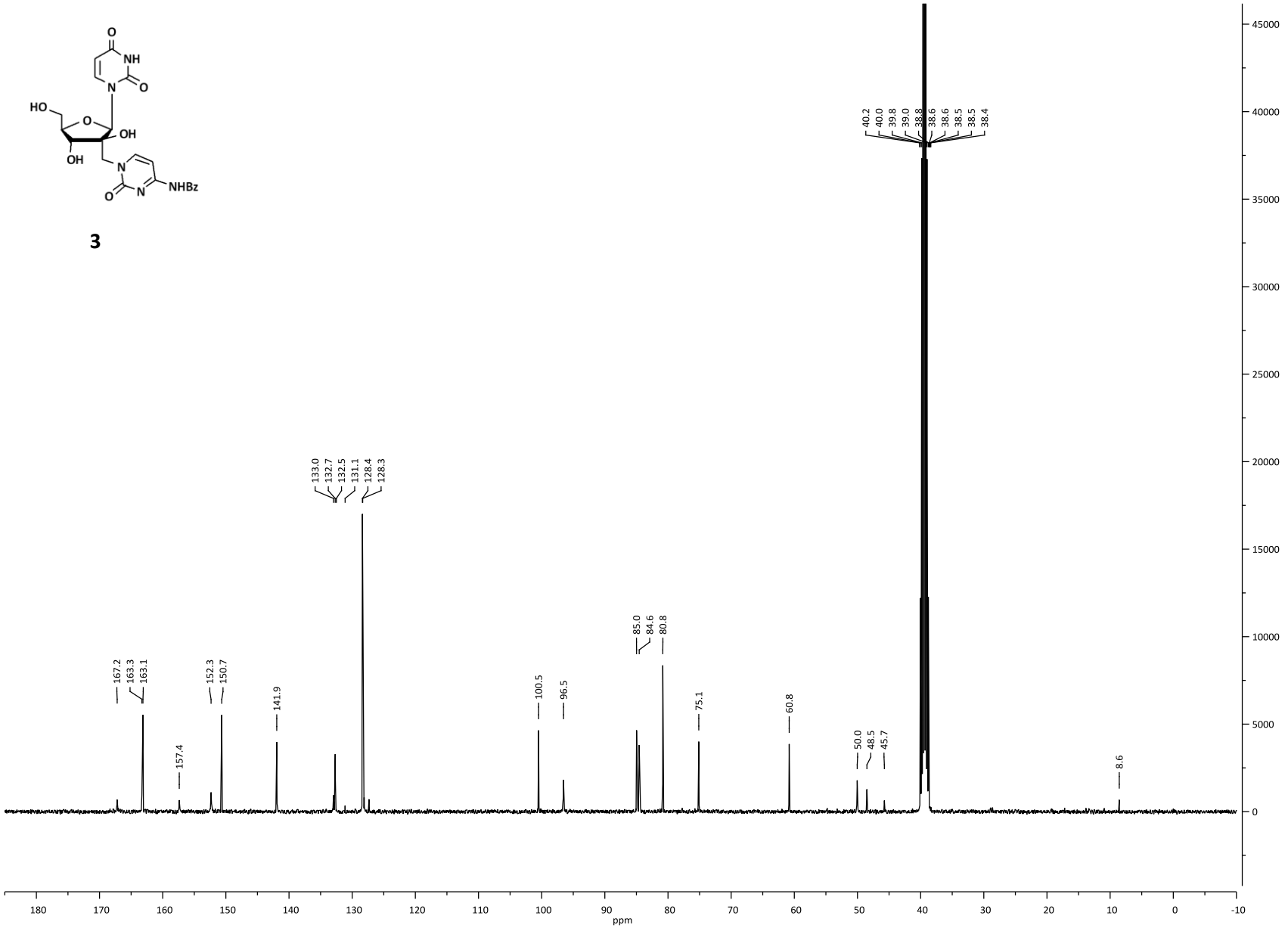


Fig A.11. 101 MHz ^{13}C NMR spectrum of compound **3** in DMSO-d_6.

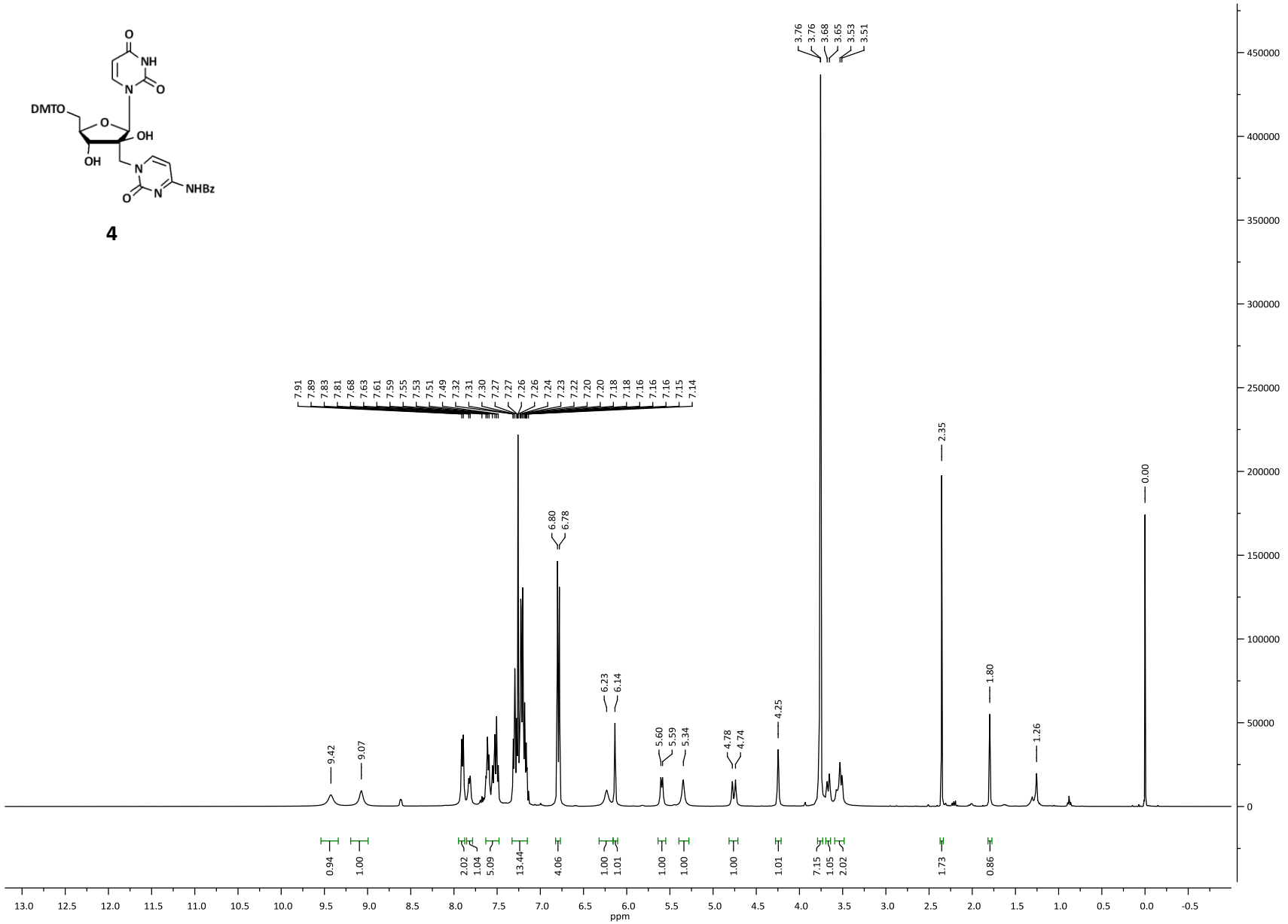


Fig A.12. 400 MHz ^{1}H NMR spectrum of compound **4** in $CDCl_3$.

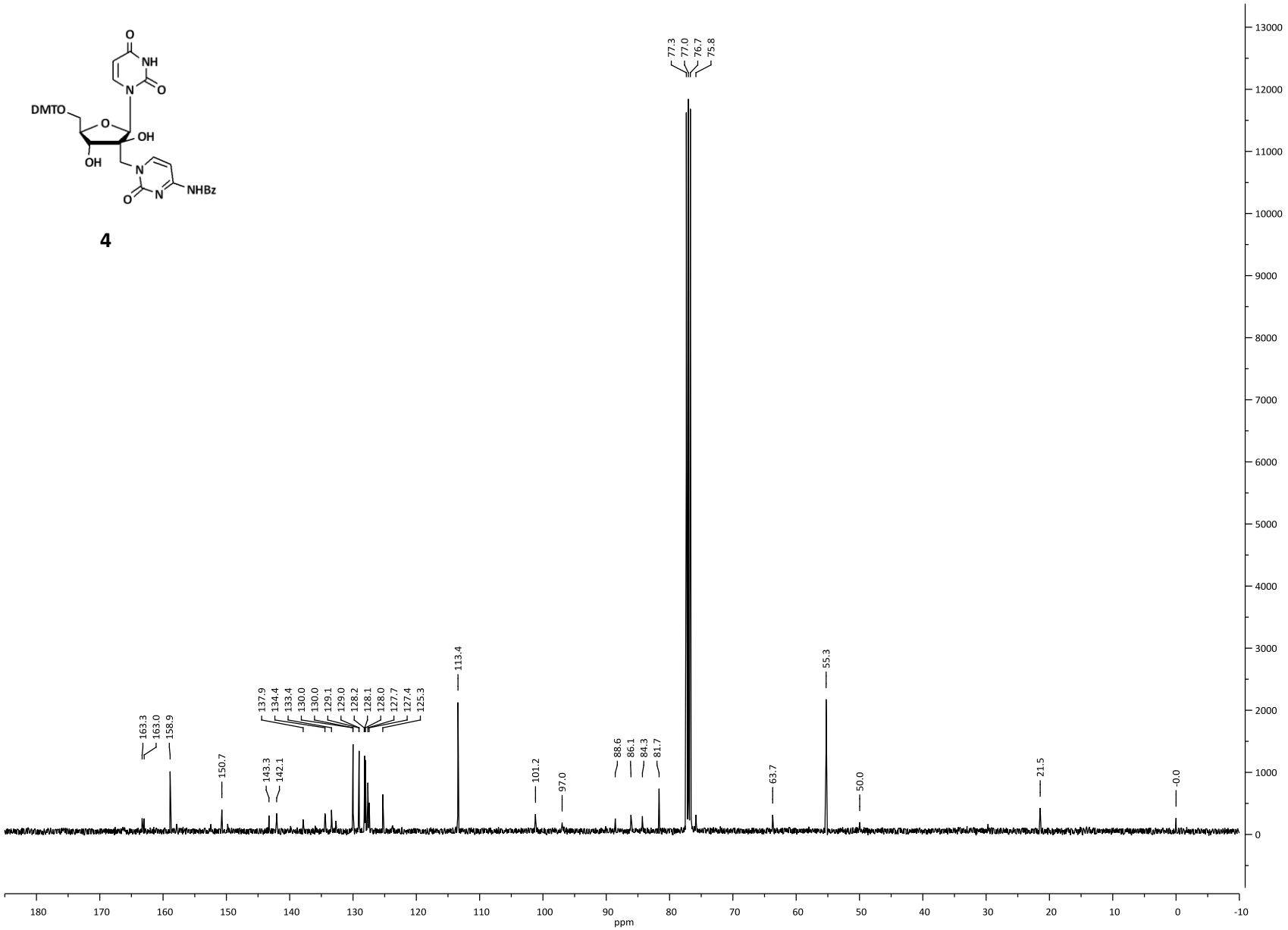


Fig A.13. 101 MHz ^{13}C NMR spectrum of compound **4** in $CDCl_3$.

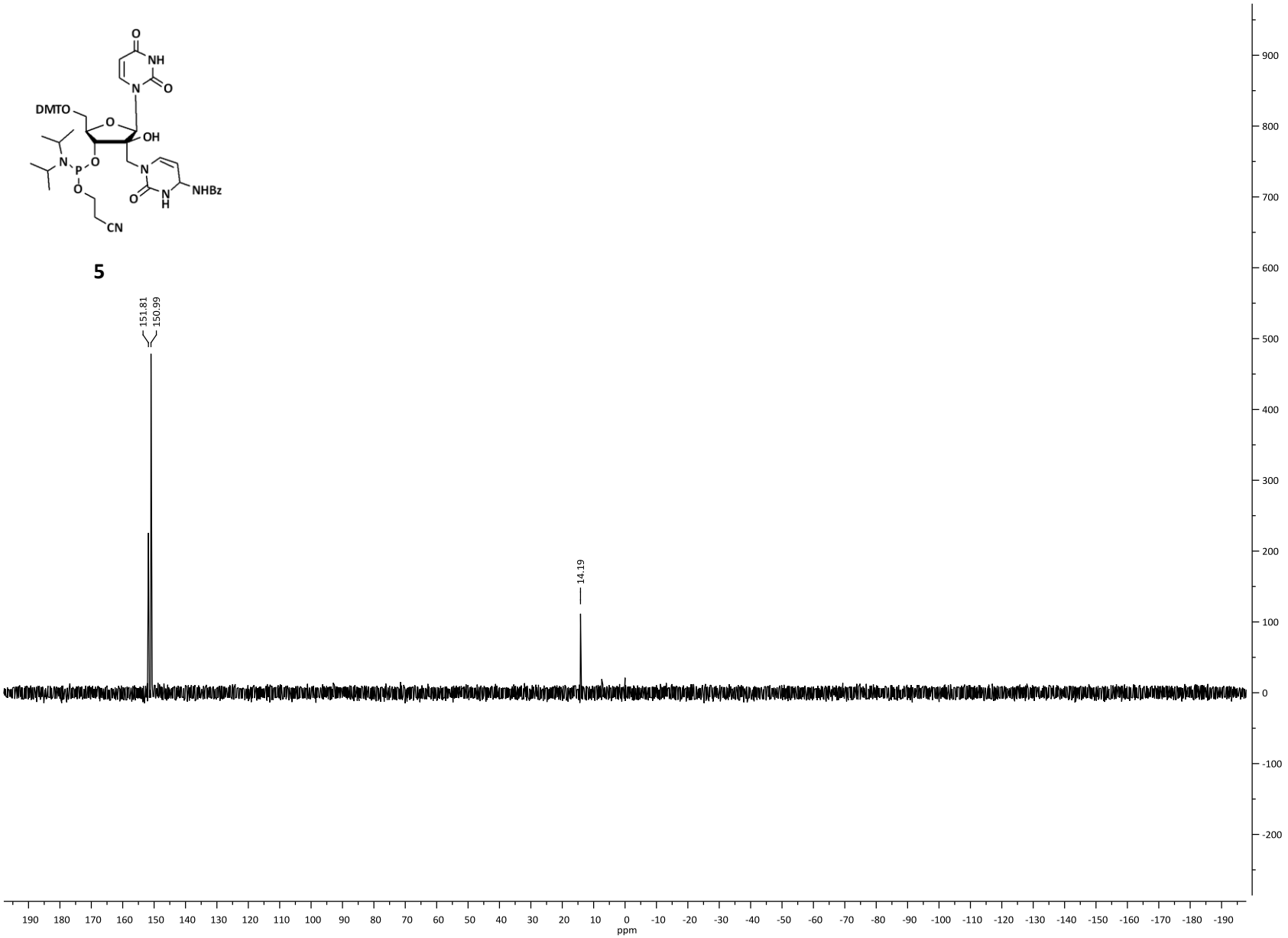


Fig A.14. 162 MHz ^{31}P NMR spectrum of compound **5** in $CDCl_3$.

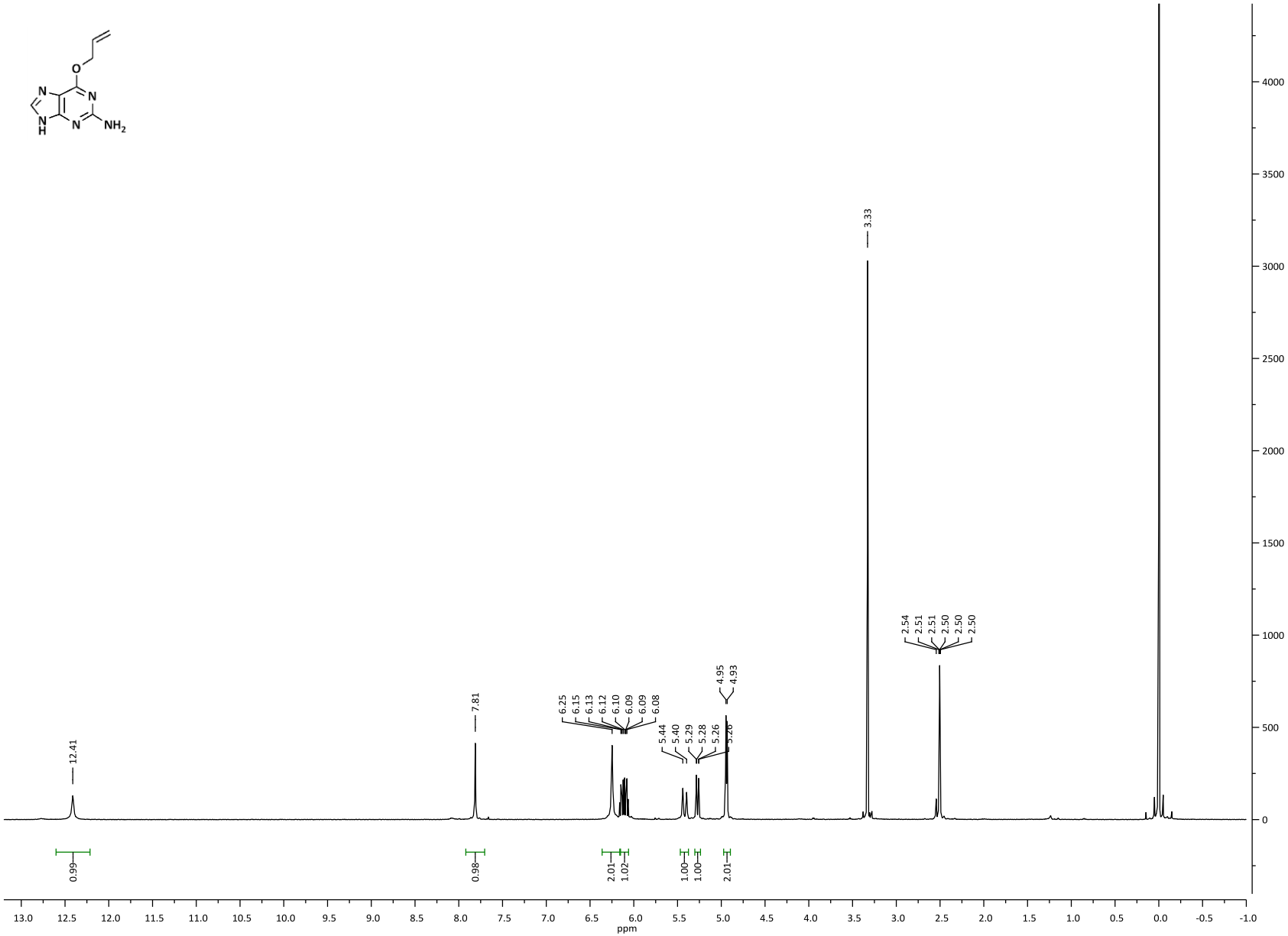


Fig A.15. 400 MHz ^{1}H NMR spectrum of O^{6}-allylguanine in DMSO-d_6.

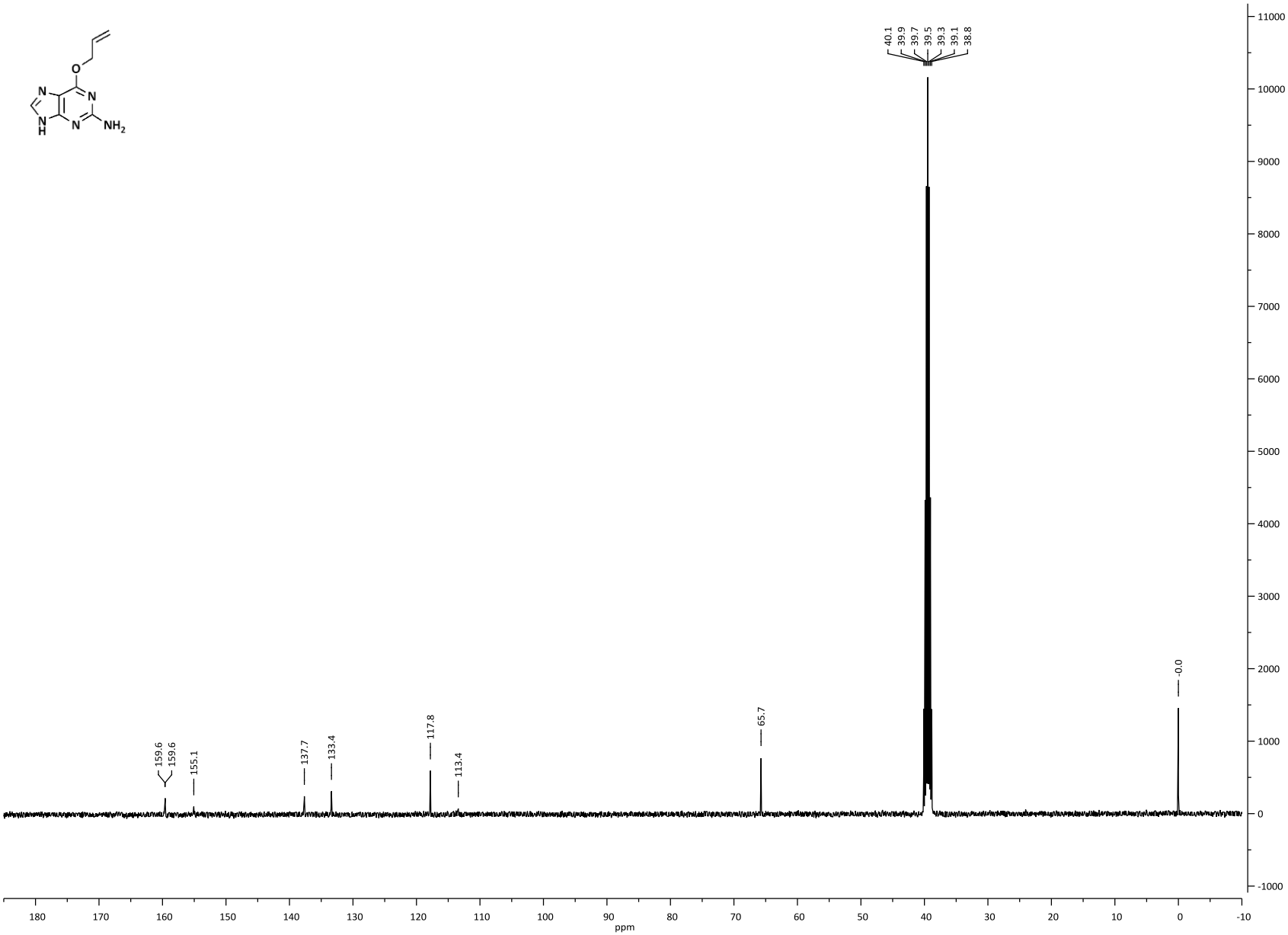


Fig A.16. 101 MHz ^{13}C NMR spectrum of O^6-allylguanine in DMSO-d_6.

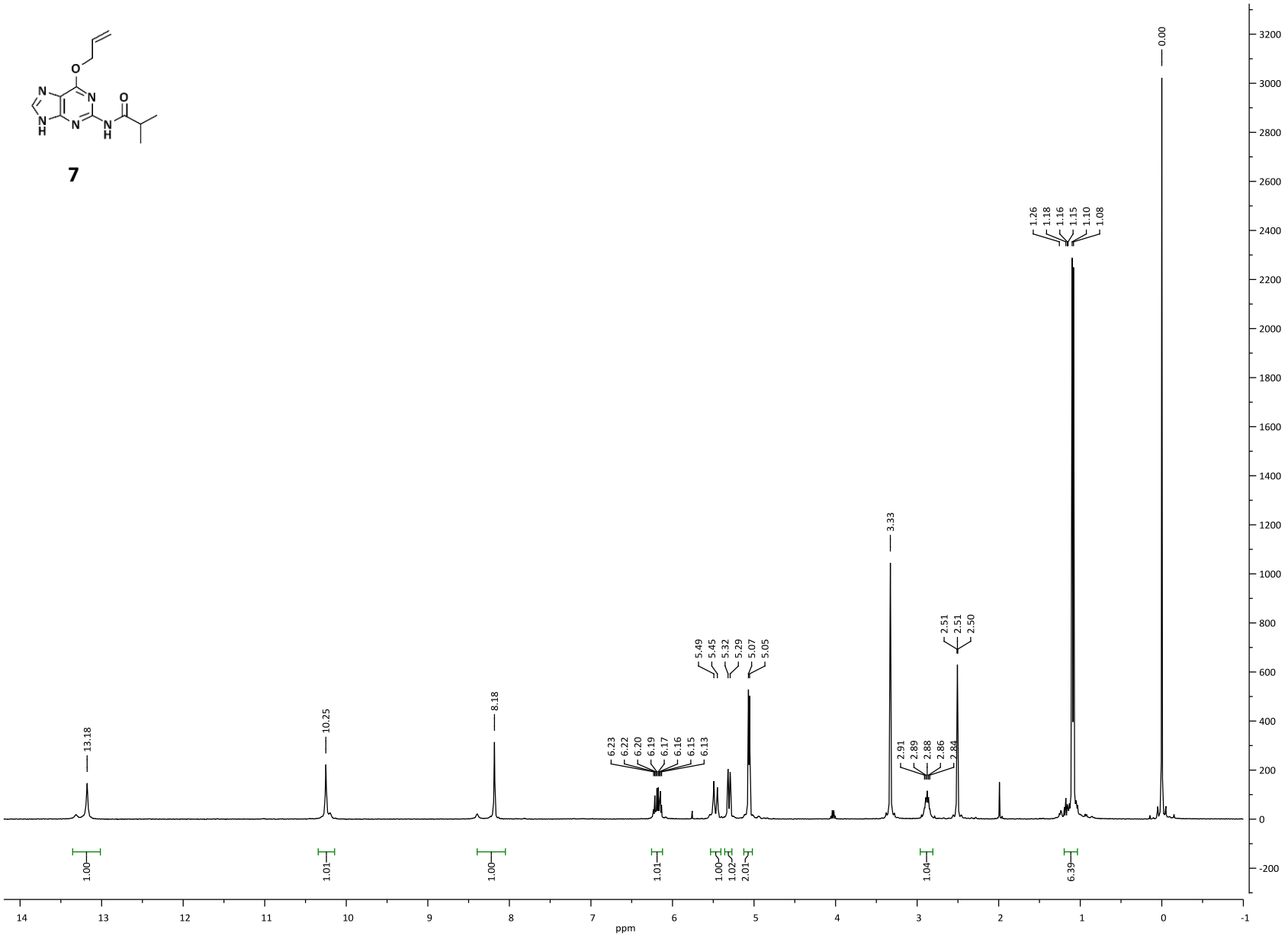


Fig A.17. 400 MHz ^{1}H NMR spectrum of compound **7** in DMSO-d_6.

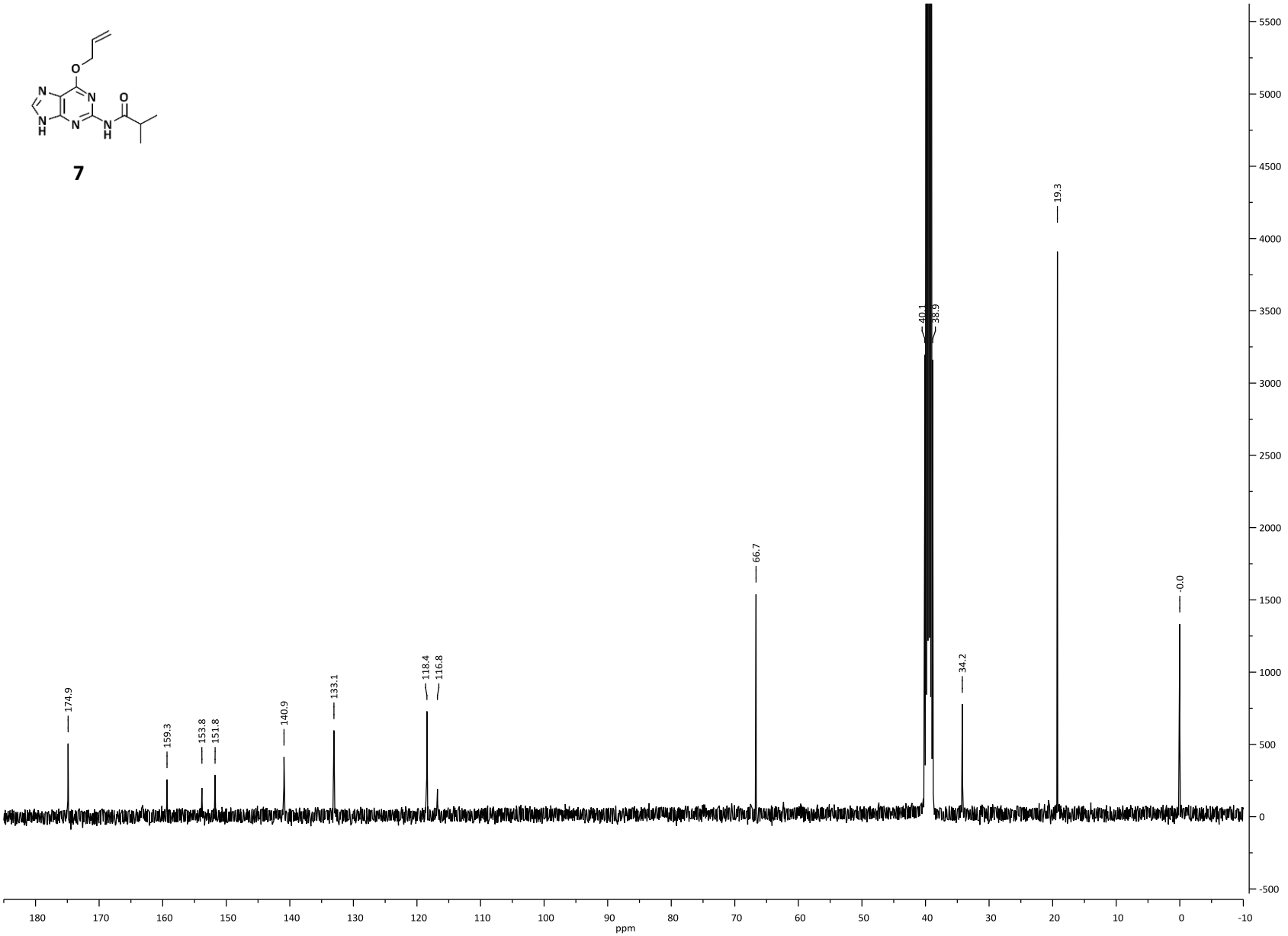


Fig A.18. 101 MHz ^{13}C NMR spectrum of compound **7** in DMSO-d_6.

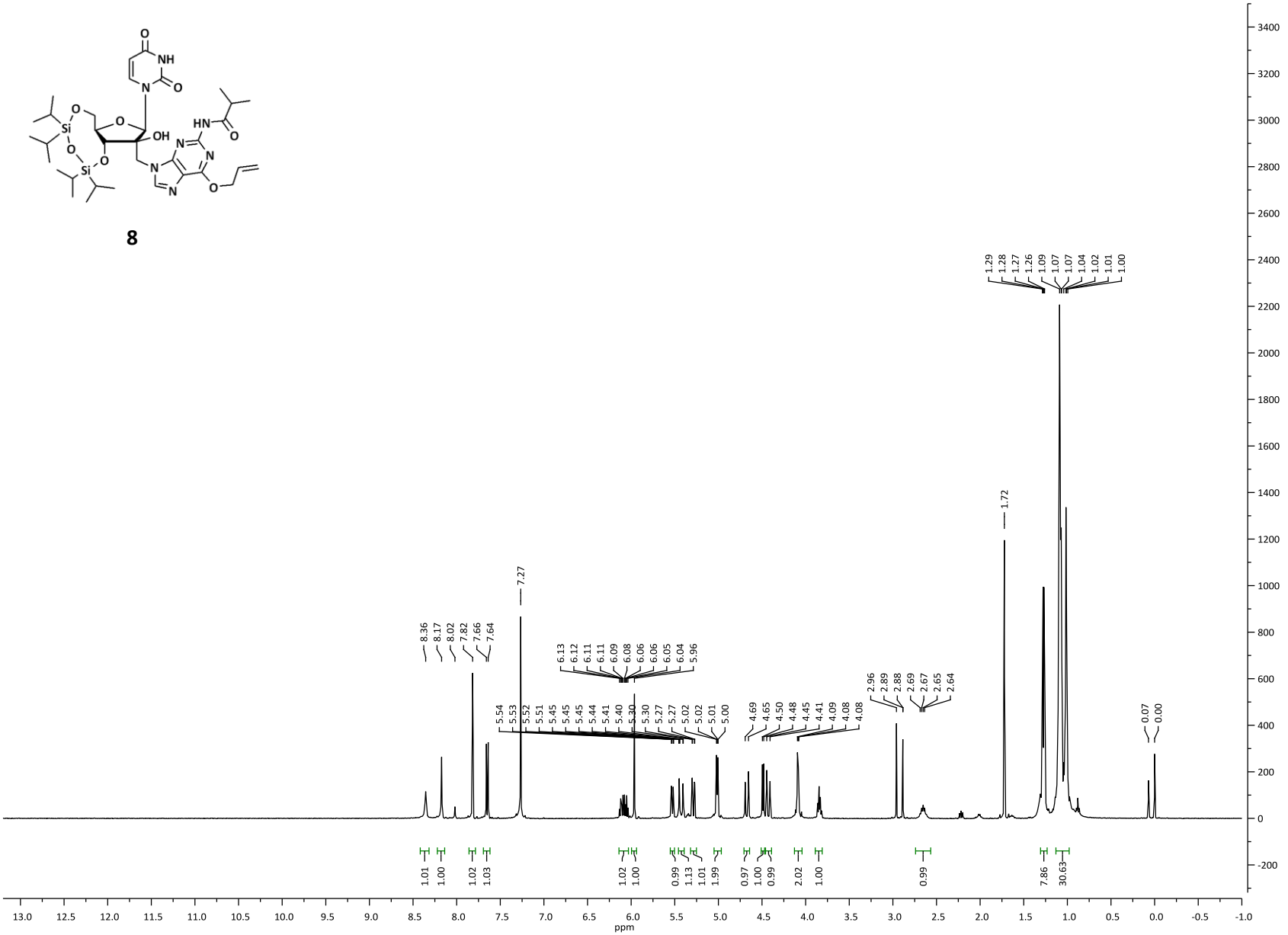


Fig A.19. 400 MHz ^{1}H NMR spectrum of compound **8** in $CDCl_3$.

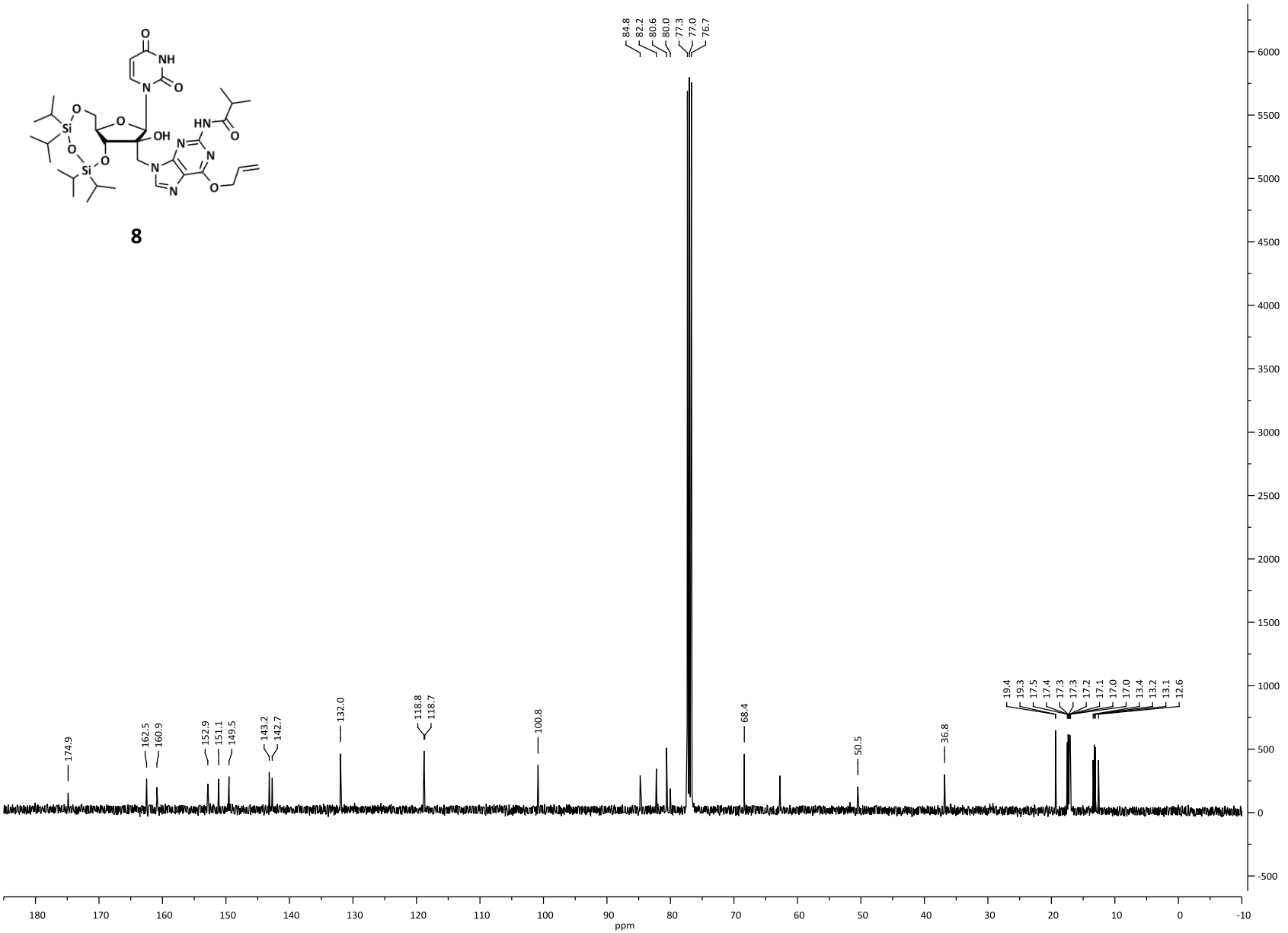


Fig A.20. 101 MHz ^{13}C NMR spectrum of compound **8** in $CDCl_3$.

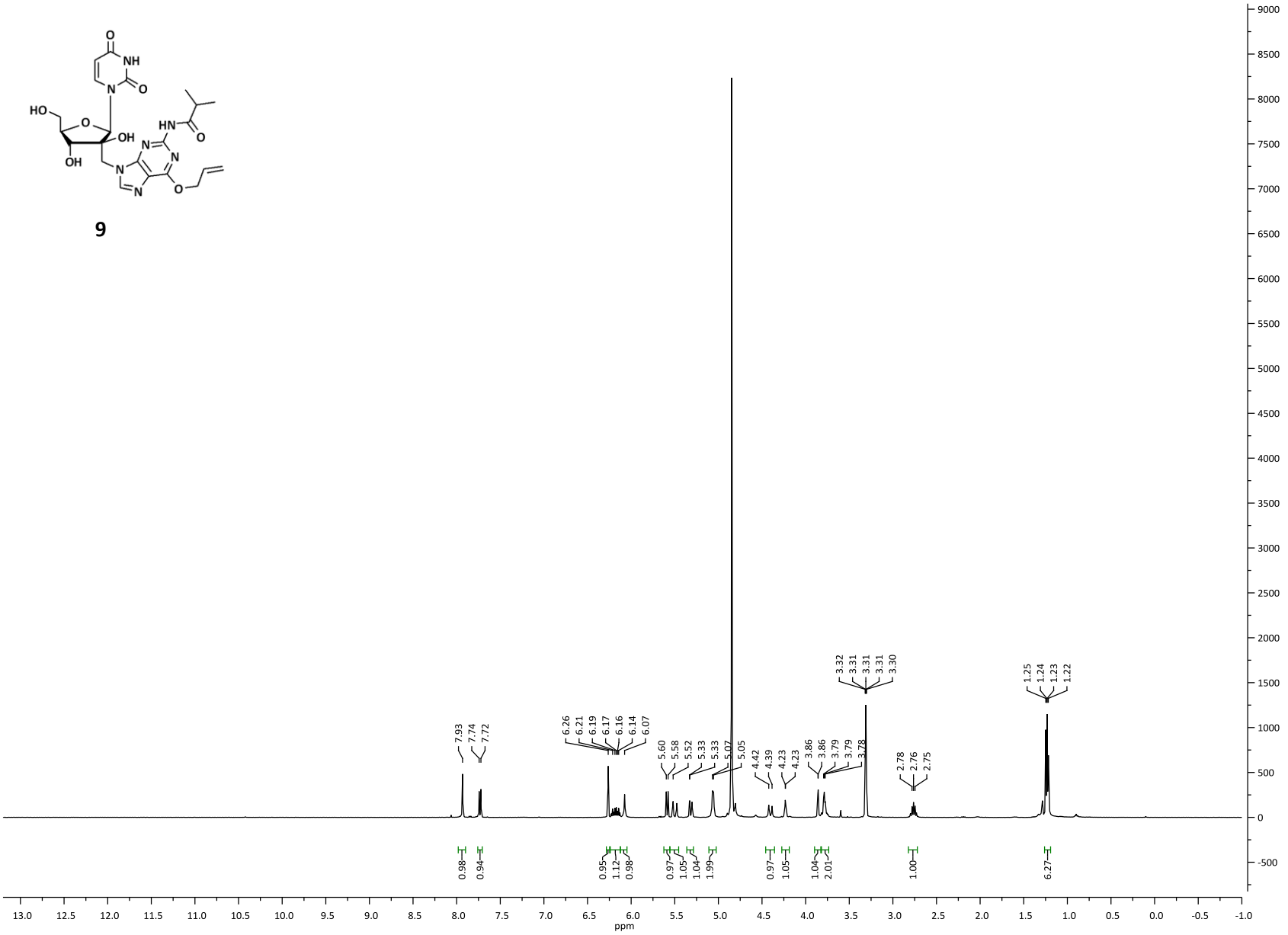


Fig A.21. 400 MHz ^{1}H NMR spectrum of compound **9** in DMSO-d_6.

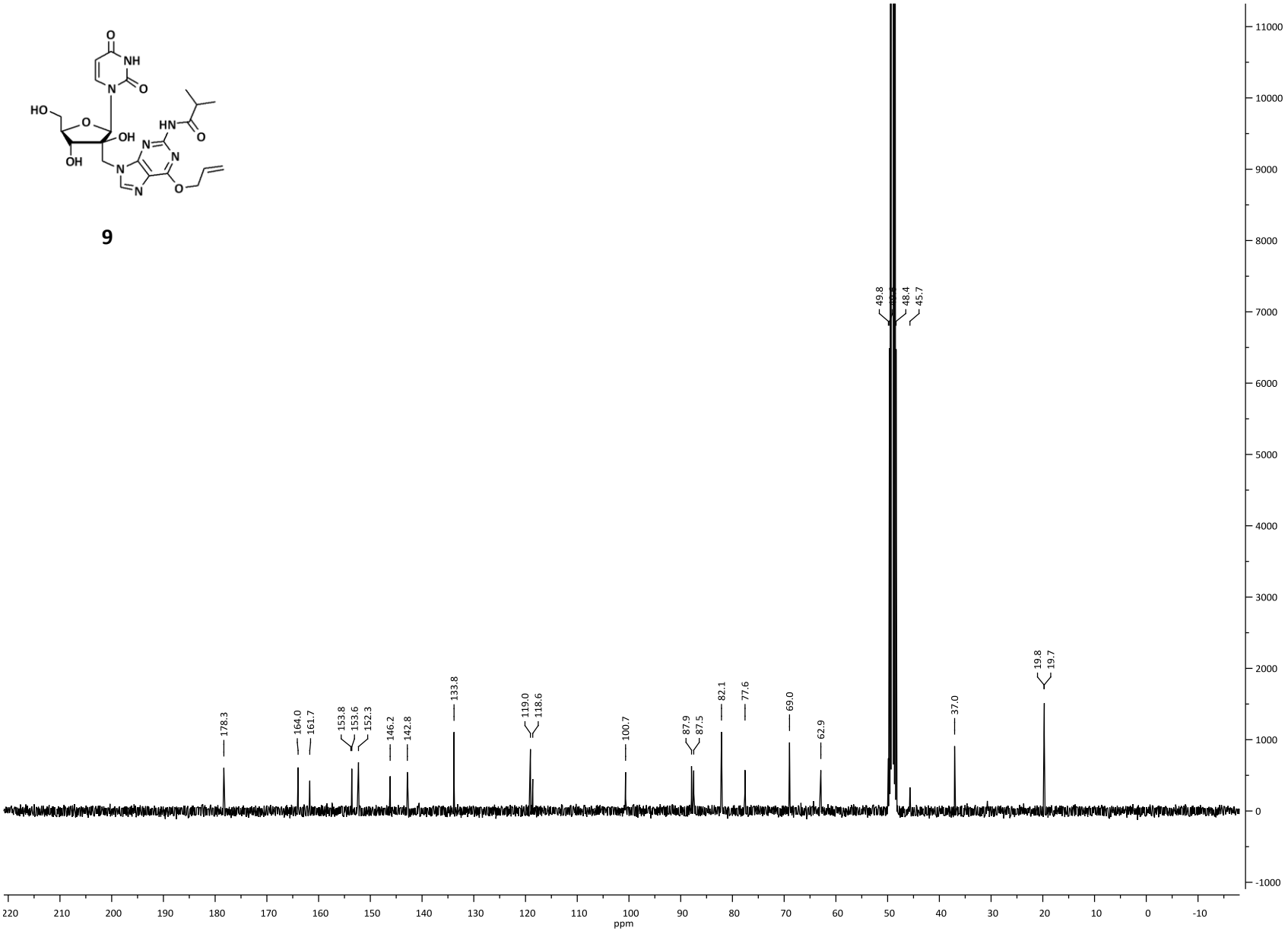


Fig A.22. 101 MHz ^{13}C NMR spectrum of compound **9** in DMSO-d_6.

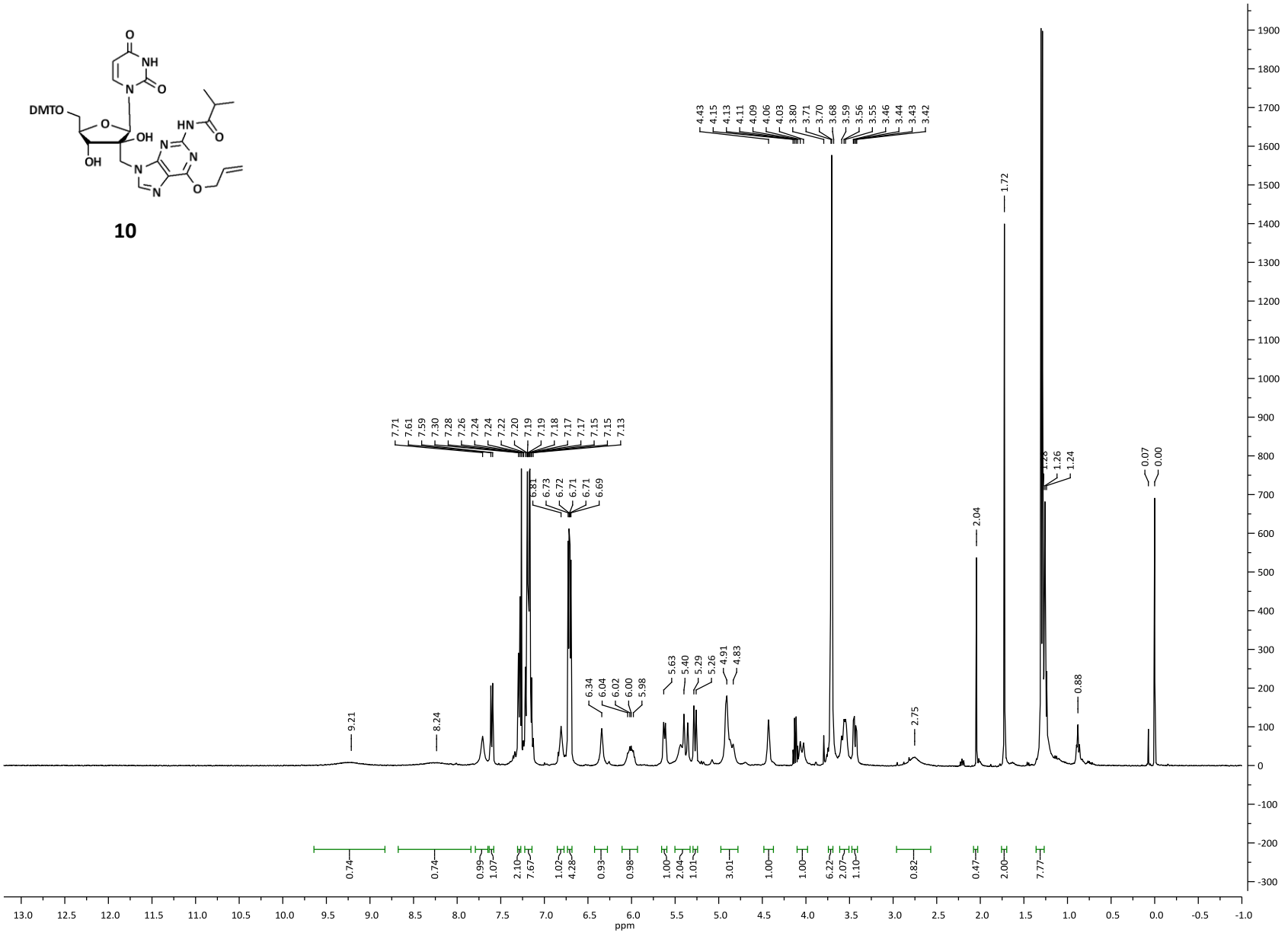


Fig A.23. 400 MHz ^{1}H NMR spectrum of compound **10** in $CDCl_3$.

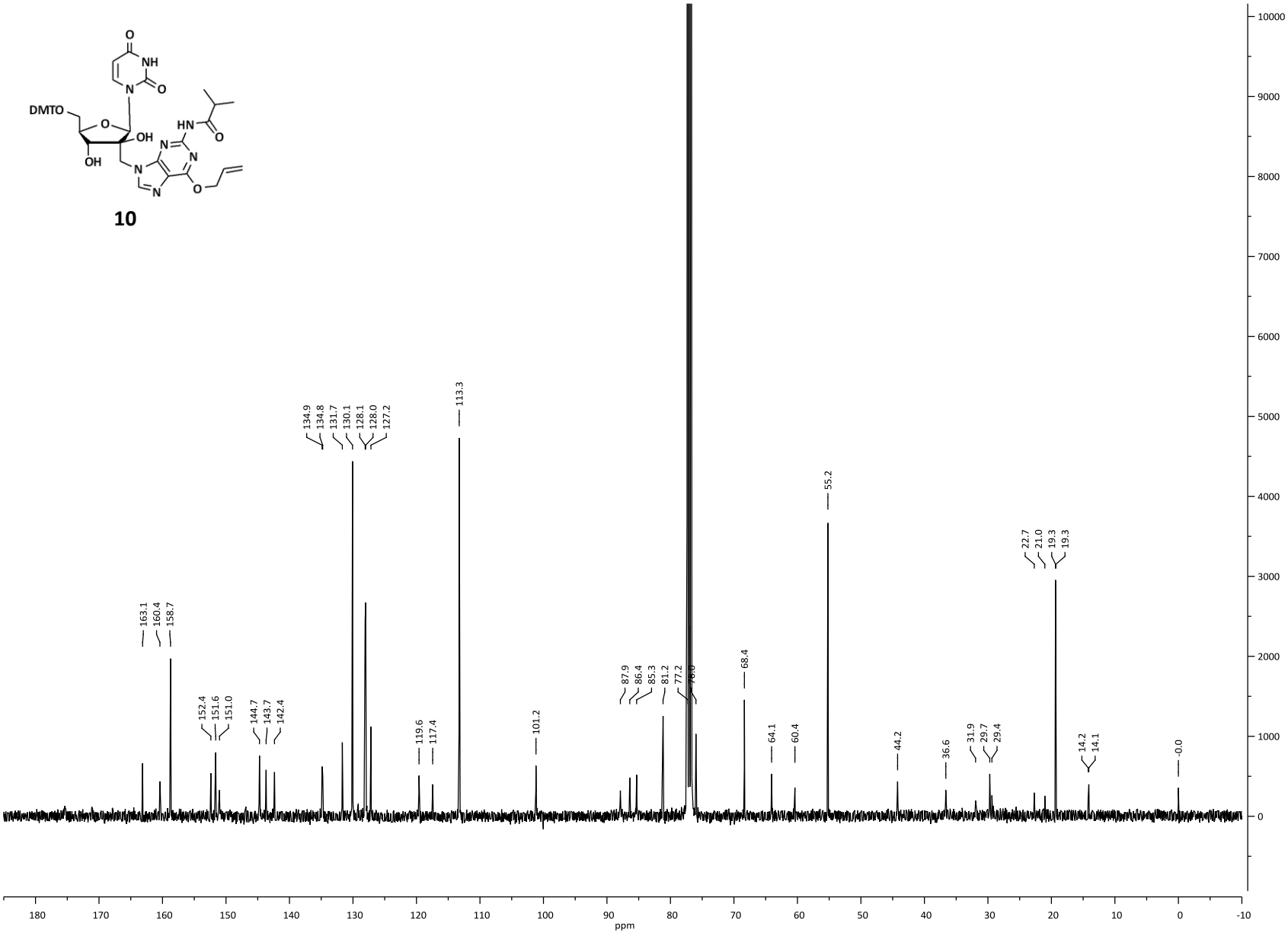


Fig A.24. 101 MHz ^{13}C NMR spectrum of compound **10** in $CDCl_3$.

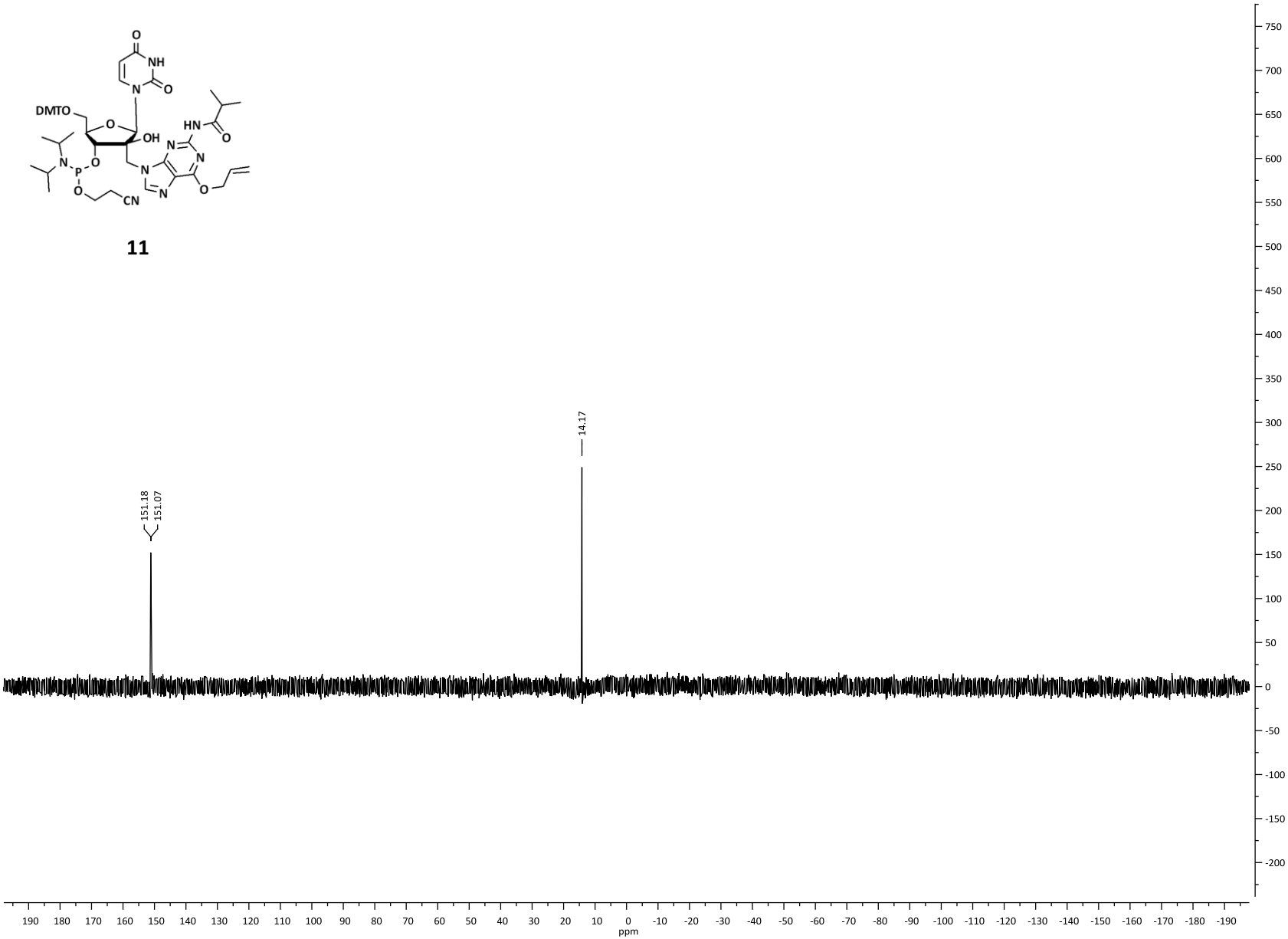


Fig A.25. 162 MHz ^{31}P NMR spectrum of compound **11** in $CDCl_3$.