Supporting Information

Chemical synthesis of 4'-thio and 4'-sulfinyl pyrimidine nucleoside analogues

Mieke Guinan[a], Ningwu Huang[b], Chris Hawes[a], Marcelo Lima[c,d], Mark Smith[b] and Gavin J. Miller[a,d]*

*g.j.miller@keele.ac.uk

[a] Lennard-Jones Laboratory, School of Chemical and Physical Sciences, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom.

[b] Riboscience LLC, 428 Oakmead Pkwy, Sunnyvale, CA 94085, USA.

[c] School of Life Sciences, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom.

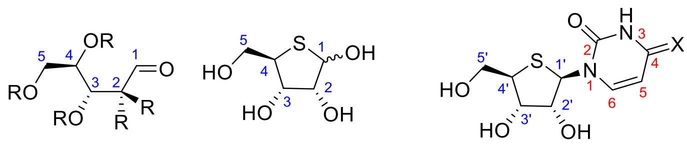
[c] Centre for Glycoscience Research, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom.

The following pages contain representative supporting information and data.

S1. Experimental

S1.1 General Experimental

^1H NMR spectra were recorded on a Bruker Advance 400 (400 MHz) instrument using deuterochloroform (or other indicated solvent) as reference. The chemical shift data for each signal are given as δ in units of parts per million (ppm) relative to tetramethylsilane (TMS) where δ (TMS) = 0.00 ppm. The multiplicity of each signal is indicated by: s (singlet); br s (broad singlet); d (doublet); t (triplet); dd (doublet of doublets); ddd (doublet of doublet of doublets); dddd (doublet of doublet of doublet of doublets); dt (doublet of triplets); ddt (doublet of doublet of triplets); dqd (doublet of quartet of doublets); ddq (doublet of doublet of quartets); sp (septet) or m (multiplet). The multiplicity of each signal may be described as app. (apparent); ov. (overlapping); br. (broad). The number of protons (n) for a given resonance is indicated by nH. Coupling constants (*J*) are quoted in Hz and are recorded to the nearest 0.1 Hz. ^1H NMR resonances were assigned with the aid of gDQCOSY. ^{13}C NMR spectra were recorded on a Bruker Advance 400 (100 MHz) instrument using the PENDANT sequence and internal deuterium lock. The chemical shift data for each signal are given as δ in units of ppm relative to TMS where δ (TMS) = 0.00 ppm. ^{13}C NMR resonances were assigned with the aid of gHSQCAD. ^{19}F NMR were recorded on a Bruker Advance 400 (376 MHz) instrument. ^{31}P NMR were recorded on a Bruker Advance 400 (161 MHz) instrument. NMR data were analysed using Mestrenova software. Analytical thin layer chromatography (TLC) was carried out on pre-coated 0.25 mm ICN Biomedicals GmbH 60 F254 silica gel plates. Visualisation was by absorption of UV light or thermal development after dipping in 5% H_2SO_4 in MeOH. Optical activities were recorded on automatic Rudolph Autopol I or Bellingham and Stanley ADP430 polarimeters (concentration in g/100 mL). HRMS (ESI, NSI) were obtained on Agilent 6530 Q-TOF, LQT Orbitrap XL1 or Waters (Xevo, G2-XS TOF or G2-S ASAP) Micromass LCT spectrometers using a methanol mobile phase in positive/negative ionisation modes as appropriate. Manual column chromatography was carried out on silica gel (Sigma Aldrich 40–63 µm) under a positive pressure of compressed air. Automatic flash chromatography was carried out on silica gel (Reveleris® X2 system) under a positive pressure of compressed N_2. Dry CH_2Cl_2 and DMF was acquired from an Innovative Technology solvent purification system. Anhydrous MeOH, dioxane, EtOH, Et_2O, DMF, acetone was dried over 4 Å molecular sieves. Chemicals were purchased from Acros Organics UK, Aldrich UK, Alfa Aesar UK, Carbosynth, Fisher Scientific, Tokyo Chemical Industry. All solvents and reagents were purified and dried where necessary, by standard techniques. Where appropriate and if not stated otherwise, all non-aqueous reactions were performed under an inert atmosphere of nitrogen, using a vacuum manifold with nitrogen passed through 4 Å molecular sieves and self-indicating silica gel. Brine refers to a saturated aqueous solution of sodium chloride. Hexane refers to n-hexane and petroleum ether to the fraction boiling between 40 and 60 °C. Volumes of less than 0.2 mL were measured and dispensed *via* automatic micropipette or Luer-lock micro-syringe. All reactions requiring heating were conducted using heating blocks atop stirrer hotplates with temperature controlled by an external probe. Reactions requiring lower temperatures were cooled using the following bath compositions: 0 °C (ice/water); -10 °C (acetone/ice). Reactions requiring lower temperature conditions or low temperatures for periods over 3 h were maintained using a Huber chiller unit and an acetone bath. An Agilent preparative HPLC system equipped with variable wavelength detector, auto sampler and 1260 series preparative fraction collector were used. The data was collected and processed using Agilent "Chemstation" 1260 series software. The UV detection wavelength was 254 nm. Assignment of proton and carbon atoms for NMR analysis follow the generic ring numbering systems illustrated below.



S1.2. Synthesis of 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-thioribose 1

2,3,5-tri-*O*-benzoyl-1'-α,β-D-ribofuranose

1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose **2** (100 g, 198 mmol, 1.0 equiv.) was dissolved in MeCN (2.0 L) and H_2O added (10 mL) and the solution cooled to 0 °C. $BF_3{\cdot}OEt_2$ (51 mL, 416 mmol, 1.6 equiv.) was added over 20 minutes and the solution stirred for a further 10 minutes at 0 °C before warming to rt and stirring vigorously for 2.5 h. The reaction was quenched with saturated aqueous $NaHCO_3$ solution (1.2 L) and stirred for 5 minutes. The organic layer was separated and the organic solvent removed *in vacuo* and the crude diluted in EtOAc (500 mL). The aqueous layer was extracted with EtOAc (6 x 300 mL) and the organic layers combined and washed with saturated aqueous $NaHCO_3$ solution (3 x 1 L) and brine (2 x 1 L), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the title compound, crude, as a white foam (83.0 g, 179 mmol, 91%) which was used without further purification. R_f 0.22 (25/75 EtOAc/petroleum ether); 1.0/1.1 ratio anomers; **major anomer:** ^{1}H NMR (400 MHz, $CHCl_3$) δ 8.10 – 7.98 (m, 6H, Ar-H), 7.57 – 7.51 (m, 3H, Ar-H), 7.43 – 7.33 (m, 6H, Ar-H), 5.91 (dd, $J_{H3\text{-}H4}$ = 6.4 Hz, $J_{H3\text{-}H2}$ = 4.8 Hz, 1H, H3), 5.69 (dd, $J_{H2\text{-}H3}$ = 4.9 Hz, $J_{H2\text{-}H1}$ = 1.1 Hz, 1H, H2), 5.64 (dd, $J_{H1\text{-}OH}$ = 3.6 Hz, $J_{H1\text{-}H2}$ = 1.1 Hz, 1H, H1), 4.74 (dd, $J_{H5a\text{-}H5b}$ = 11.4 Hz, $J_{H5a\text{-}H4}$ = 3.4 Hz, 1H, H5a), 4.72 – 4.69 (m, 1H, H4), 4.63 (dd, $J_{H5b\text{-}H5a}$ = 11.0 Hz, $J_{H5b\text{-}H4}$ = 5.2 Hz, 1H, H5b), 3.81 (br s, 1H, OH); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.6 (C=O, Bz), 165.5 (C=O, Bz), 165.4 (C=O, Bz), 133.5 (C_q, Ar-C), 133.4 (C_q, Ar-C), 133.2 (C_q, Ar-C), 129.9 (CH, Ar-C), 129.84 (CH, Ar-C), 129.79 (CH, Ar-C), 129.77 (CH, Ar-C), 128.63 (CH, Ar-C), 128.58 (CH, Ar-C), 128.51 (CH, Ar-C), 128.50 (CH, Ar-C), 128.45 (CH, Ar-C), 128.4 (CH, Ar-C), 100.5 (CH, C1), 79.4 (CH, C4), 76.2 (CH, C2), 72.4 (CH, C3), 65.2 (CH_2, C5); ESI HRMS *m/z* found: $(M+H)^+$ 463.1401 $C_{26}H_{22}N_2O_8$, requires $(M+H)^+$ 463.1387. Data was consistent with literature values.[1]

(2*R*,3*R*,4*S*)-2,3,5-tri-*O*-benzoyl-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*) 3

To a solution of 2,3,5-tri-*O*-benzoyl-1'-α,β-D-ribofuranose (79.7 g, 172 mmol, 1.0 equiv.) in MeOH (115 mL) was added $H_2NOMe{\cdot}HCl$ (21.5 g, 268 mmol, 1.6 equiv.) and the solution cooled to 0 °C. Et_3N (36 mL, 258 mmol, 1.5 equiv.) was added and the solution stirred for a further 15 minutes at 0 °C before warming to rt. After 21 h vigorous stirring, the solvent was removed *in vacuo* and the residue partitioned between EtOAc (1.0 L) and H_2O (1.5 L). The organic layer was separated and the aqueous extracted with EtOAc (2 x 500 mL). The organic layers were combined and washed with H_2O (1 L) and brine (1 L), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the crude **3** as a white foamy syrup (84.9 g, 172 mmol, 90%) which was used without further purification. R_f 0.45 (1/9 acetone/toluene); 3/1 ratio isomers; **major isomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 – 7.98 (m, 6H, Ar-H), 7.62 (d, $J_{H1\text{-}H2}$ = 6.9 Hz, 1H, H1), 7.58 – 7.52 (m, 3H, Ar-H), 7.43 – 7.37 (m, 6H, Ar-H), 6.17 (dd, J_{H2-H1} = 6.9 Hz, J_{H2-H3} = 3.2 Hz, 1H, H2), 5.83 (dd J_{H3-H4} = 8.2 Hz, J_{H3-H2} = 3.2 Hz, 1H, H3), 4.71 – 4.63 (m, 1H, H5a), 4.46 – 4.43 (m, 1H, H5b), 4.42 – 4.39 (m, 1H, H4), 3.84 (s, 3H, OCH_3), 3.13 (d, J_{OH-H4} = 5.8 Hz, 1H, 4-OH); ^{13}C NMR (101 MHz, $CDCl_3$) 166.9 (C=O, Bz), 165.3 (C=O, Bz), 165.1 (C=O, Bz), 145.1 (CH=N, C1), 133.6 (C_q, Ar-C), 133.6 (C_q, Ar-C), 133.4 (C_q, Ar-C), 133.3 (CH, Ar-C), 129.9 (CH, Ar-C), 129.83 (CH, Ar-C), 129.81 (CH, Ar-C), 128.6 (CH, Ar-C), 128.48 (CH, Ar-C), 128.45 (CH, Ar-C), 128.4 (CH, Ar-C), 73.3 (CH, C3), 71.1 (CH, C2), 69.0 (CH, C4), 65.8 (CH_2, C5), 62.3 (OCH_3); ESI HRMS *m/z* found: $(M+Na)^+$ 514.1494 $C_{26}H_{26}NO_8$, requires $(M+Na)^+$ 514.1472. This compound was reported previously,[2] but not fully characterised.

(2R,3R,4S)-2,3,5-tri-*O*-benzoyl-4-*O*-(2',4',5'-trichlorophenylsulfonyl)-1-(methoxyimino)pentane (E/Z) 4

Oxime **3** (70.5 g, 143 mmol, 1.0 equiv.), 2,4,5-trichlorobenzenesulfonyl chloride (44.1 g, 158 mmol, 1.1 equiv.) and *N*-methylimidazole (12.6 mL, 158 mmol, 1.1 equiv.) were dissolved in MeCN (378 mL) and the solution stirred vigorously at rt. After 18 h, reaction completion was observable by TLC (R_f = 0.26 for **3**, R_f = 0.57 for **4** in 50/50 Et_2O/pet. ether) and H_2O (40 mL) was added and the solvent removed *in vacuo*. The residue was diluted in EtOAc (3 L) and washed with saturated aqueous $NaHCO_3$ (1.6 L). The aqueous layer was extracted with EtOAc (2 x 450 mL) and the combined organic layers washed with H_2O (1.6 L). The aqueous layer was extracted with EtOAc (450 mL) and the combined organic phases washed once more with H_2O (1.6 L) and brine (400 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo*. The crude foam was triturated from ice-cold Et_2O (400 mL) and the white solid collected by suction filtration to obtain **4** as a white amorphous solid (43.7 g). The mother liquor was dried *in vacuo* and triturated a second time from ice-cold Et_2O (150 mL) to obtain a further quantity of **4** (32.0 g) (75.7 g total, 103 mmol, 72%). R_f 0.57 (50/50 Et_2O/petroleum ether); 5.7/1.0 ratio isomers, geometries not defined; **major isomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 – 7.89 (m, 8H, Ar-H), 7.62 – 7.55 (m, 3H, Ar-H), 7.47 – 7.41 (m, 6H, Ar-H), 7.48 – 7.39 (m, 6H, Ar-H) 7.44 (d, J_{H1-H2} = 6.2 Hz, 1H, H1), 6.03 (dd, J_{H2-H1} = 6.2 Hz, J_{H2-H3} = 5.2 Hz, 1H, H2), 5.98 (dd, J_{H3-H2} = 5.2 Hz, J_{H3-H4} = 3.8 Hz, 1H, H3), 5.53 (ddd, J_{H4-H5b} = 7.4 Hz, J_{H4-H3} = 3.8 Hz, J_{H4-H5a} = 2.9 Hz, 1H, H4), 4.85 (dd, $J_{H5a-H5b}$ = 12.7 Hz, J_{H5a-H4} = 2.9 Hz, 1H, H5a), 4.67 (dd, $J_{H5b-H5a}$ = 12.7 Hz, J_{H5b-H4} = 7.3 Hz, 1H, H5b), 3.85 (s, 1H, OCH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.7 (C=O, Bz), 164.8 (C=O, Bz), 164.8 (C=O, Bz), 143.9 (CH=N, C1), 139.0 (C_q, Ar-C), 133.9 (C_q, Ar-C), 134.4 (C_q, Ar-C), 133.7 (C_q, Ar-C), 133.5 (C_q, Ar-C), 133.4 (C_q, Ar-C), 132.0 (C_q, Ar-C), 129.9 (CH, Ar-C), 129.7 (CH, Ar-C), 129.6 (CH, Ar-C), 128.8 (CH, Ar-C), 128.7 (CH, Ar-C), 128.64 (CH, Ar-C), 128.61 (CH, Ar-C), 128.59 (CH, Ar-C), 128.5 (CH, Ar-C), 79.6 (CH, C4), 71.6 (CH, C3), 69.7 (CH, C2), 62.4 (OCH_3), 62.3 (CH_2, C5); ESI HRMS *m/z* found: $(M+H)^+$ 734.0422 $C_{33}H_{26}Cl_3NO_{10}S$, requires $(M+H)^+$ 734.0421. This compound was reported previously,[2] but not fully characterised.

(2*S*,3*R*,4*S*)-2,3,5-tri-*O*-benzoyl-4-bromo-1-(methoxyimino)pentane (*E*/*Z*) 5

To a solution of **4** (75.6 g, 103 mmol, 1.0 equiv.) in 2-butanone (270 mL) was added LiBr (40.6 g, 472 mmol, 5.0 equiv.) and the solution stirred at 80 °C. After 18 h, the solution was cooled to rt and the solvent removed *in vacuo*. The residue was partitioned between EtOAc (500 mL) and H_2O (500 mL), the organic layer separated, and the aqueous layer extracted with EtOAc (3 x 300 mL). The combined organic layers were washed with H_2O (400 mL) and brine (400 mL), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the crude **5** as a yellow oil (56.1 g, ~101 mmol, *quant.*). An analytically pure sample of **5** (50.7 g, 91.5 mmol, 89%) was obtained *via* purification on silica gel *via* automated flash chromatography (0 – 32% Et_2O/petroleum ether). R_f 0.36 (1/4, Et_2O/petroleum ether); 3.3/1.0 ratio isomers; **major isomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.12 – 8.04 (m, 6H, Ar-H) 7.62 – 7.57 (m, 3H, Ar-H), 7.48 – 7.43 (m, 6H, Ar-H), 7.50 (d, 1H, J_{H1-H2} = 3.3 Hz, H1), 6.03 (dd, 1H, J_{H2-H3} = 6.5 Hz, J_{H2-H1} = 3.3 Hz, H2), 6.00 (dd, 1H, J_{H3-H2} = 6.5 Hz, J_{H3-H4} = 3.0 Hz, H3), 4.81 – 4.75 (m, 1H, H5a), 4.70 (ddd, J_{H4-H5} = 7.2 Hz, J_{H4-H5} = 6.0 Hz, J_{H4-H3} = 2.9 Hz, 1H, H4), 4.61 – 4.55 (1H, m, H5b), 3.70 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.7 (C=O, Bz), 165.1 (C=O, Bz), 164.7 (C=O, Bz), 144.4 (CH=N, C1), 133.8 (C_q, Ar-C), 133.7 (CH, C1'), 133.5 (C_q, Ar-C), 133.4 (CH, Ar-C), 133.3 (CH, Ar-C), 130.2 (CH, Ar-C), 130.1 (CH, Ar-C), 130.0 (CH, Ar-C), 129.9 (CH, Ar-C), 129.9 (CH, Ar-C), 129.9 (CH, Ar-C), 129.3 (CH, Ar-C), 129.3 (CH, Ar-C), 129.0 (CH, Ar-C), 129.0 (CH, Ar-C), 128.9 (CH, Ar-C), 128.9 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.6 (CH, Ar-C), 128.5 (CH, Ar-C), 128.5 (CH, Ar-C), 128.4 (CH, Ar-C), 71.4 (CH, C2), 70.2 (CH, C3), 64.8 (CH_2, C5), 62.2 (OCH_3), 47.8 (C-Br, C4); ESI HRMS *m/z* found: $(M+H)^+$ 554.0808 $C_{27}H_{24}BrNO_7$, requires $(M+H)^+$ 554.0809. This compound was reported previously,[2] but not fully characterised.

2,3,5-tri-*O*-benzoyl-1-α,β-(4-thio-D-ribofuranose) 6

Glyoxylic acid (35.0 mL, 641 mmol, 7.0 equiv.) was added to a solution of **5** (50.7 g, 91.5 mmol, 1.0 equiv.) in MeCN (183 mL) and the solution heated to 70 °C. After 18 h, the reaction was cooled to rt, poured onto H_2O (1 L) and extracted with EtOAc (4 x 500 mL). The combined organic phases were washed with H_2O (5 x 500 mL) and brine (500 mL), dried over anhydrous Na_2SO_4, filtered and dried *in vacuo* to obtain a mixture of crude aldehyde along with the hydrate form (44.2 g, 84.2 mmol, 92%) which was used immediately without further purification. R_f 0.65 (1/4 Et_2O/petroleum ether); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.72 (d, J_{H1-H2} = 1.0 Hz, 1H, H1), 8.17 – 7.92 (m, 6H, Ar-H), 7.75 – 7.35 (m, 9H, Ar-H), 6.03 (dd, J_{H3-H2} = 7.3 Hz, J_{H3-H4} = 3.3 Hz, 1H, H3), 5.68 (dd, J_{H2-H3} = 7.2 Hz, J_{H2-H1} = 1.0 Hz, 1H, H2), 4.87 – 4.78 (m, 1H, H5a), 4.77 (ddd, J_{H4-H5b} = 7.1 Hz, J_{H4-H5a} = 5.9 Hz, J_{H4-H3} = 3.3 Hz, 1H, H4), 4.60 (dd, $J_{H5b-H5a}$ = 11.2 Hz, J_{H5b-H4} = 7.0 Hz, 1H, H5b); ^{13}C NMR (101 MHz, $CDCl_3$) δ 194.7 (CHO, C1), 165.7 (C=O, Bz), 165.1 (C=O, Bz), 165.0 (C=O, Bz), 134.1 (C_q, Ar-C), 133.5 (C_q, Ar-C), 130.2 (C_q, Ar-C), 130.1 (CH, Ar-C), 130.0 (CH, Ar-C), 129.9 (CH, Ar-C), 129.8 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.5 (CH, Ar-C), 76.9 (CH, C2), 69.4 (CH, C3), 64.4 (CH_2, C5), 47.9 (C-Br, C4). This compound was reported previously,[2] but not fully characterised. The crude aldehdye (44.2 g, 84.2 mmol, 1.0 equiv.) was dissolved in DMF (11 mL) and cooled to 0 °C. NaSH monohydrate (8.11 g, 110 mmol, 1.3 equiv.) dissolved in a minimum volume of H_2O (5 mL) was added and the solution stirred at 0 °C for 30 minutes. The solution was diluted with EtOAc (1.2 L), washed with H_2O (2 x 500 mL) and brine (500 mL), dried over anhydrous Na_2SO_4, filtered and dried *in vacuo* to obtain the crude **6** as an orange syrup (30.7 g, 64.1 mmol, 76%). An analytically pure sample of **6** was obtained *via* purification on silica gel *via* automated flash chromatography (0 – 30% EtOAc/petroleum ether). R_f 0.29 (1/1 Et_2O/petroleum ether); 3/1 ratio anomers; **major anomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 (m, 2H, Ar-H), 7.96 (m, 2H, Ar-H), 7.89 (m, 2H, Ar-H), 7.61 – 7.28 (m, 9H, Ar-H), 6.05 (dd, $J_{H3\text{-}H4}$ = 8.1 Hz, $J_{H3\text{-}H2}$ = 3.6 Hz, 1H, H3), 5.90 (dd, $J_{H2\text{-}H3}$ = 3.6 Hz, $J_{H2\text{-}H1}$ = 2.1 Hz, 1H, H2), 5.51 (d, $J_{H1\text{-}H2}$ = 2.1 Hz, 1H, H1), 4.74 (dd, $J_{H5a\text{-}H5b}$ = 11.4 Hz, $J_{H5a\text{-}H4}$ = 6.4 Hz, 1H, H5a), 4.61 (dd, $J_{H5b\text{-}H5a}$ = 11.4 Hz, $J_{H5b\text{-}H4}$ = 6.1 Hz, 1H, H5b), 4.23 (app. dt, $J_{H4\text{-}H3}$ = 8.0 Hz, $J_{H4\text{-}H5a/b}$ = 6.1 Hz, 1H, H4); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.1 (C=O, Bz), 165.4 (C=O, Bz), 165.4 (C=O, Bz), 133.6 (C_q, Ar-C), 133.4 (C_q, Ar-C), 133.2 (C_q, Ar-C), 129.9 (CH, Ar-C), 129.8 (CH, Ar-C), 129.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.4 (CH, Ar-C), 128.3 (CH, Ar-C), 80.2 (CH, C1), 79.3 (CH, C2), 75.6 (CH, C3), 65.8 (CH_2, C5), 46.3 (CH, C4); ESI HRMS *m/z* found: $(M+Na)^+$ 501.1001 $C_{26}H_{22}O_7S$, requires $(M+Na)^+$ 501.0984. This compound was reported previously,[2] but not fully characterised.

1-β-*O*-acetyl-2,3,5-tri-*O*-benzoyl-1-(4-thio-D-ribofuranose) 1

Ac_2O (6.70 mL, 70.7 mmol, 1.5 equiv.) was added to a solution of the crude anomeric mixture of **6** (22.5 g, 47.1 mmol, 1.0 equiv.) in pyridine (59 mL) and the solution stirred at rt for 30 minutes. The solution was poured onto 1M HCl solution (800 mL) and diluted with EtOAc (1 L). The organic layer was separated and washed with 1M aqueous HCl solution (300 mL), saturated aqueous $NaHCO_3$ solution (3 x 300 mL) and brine (300 mL), dried over anhydrous Na_2SO_4, filtered and dried *in vacuo* to obtain the crude as a yellow foam which was triturated from ice-cold MeOH (70 mL), the precipitate collected by suction filtration and the filtrate washed with ice-cold MeOH (30 mL) to obtain **1** as a white amorphous solid (26.1 g, 50.2 mmol, 72%). R_f 0.67 (EtOAc); $[\alpha]_D^{25.8}$ +14.1 (*c* 1.6, MeCN); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 (m, 2H, Ar-H), 7.97 (m, 2H, Ar-H), 7.89 (m, 2H, Ar-H), 7.66 – 7.54 (m, 1H, Ar-H), 7.54 – 7.42 (m, 4H, Ar-H), 7.38 – 7.28 (m, 4H, Ar-H), 6.06 (d, $J_{H1\text{-}H2}$ = 1.7 Hz, 1H, H1), 5.99 (dd, $J_{H2\text{-}H3}$ = 3.6 Hz, $J_{H2\text{-}H1}$ = 1.7 Hz, 1H, H2), 5.91 (dd, $J_{H3\text{-}H4}$ = 8.6 Hz, $J_{H3\text{-}H2}$ = 3.6 Hz, 1H, H3), 4.73 (dd, $J_{H5a\text{-}H5b}$ = 11.5 Hz, $J_{H5a\text{-}H4}$ = 6.0 Hz, 1H, H5a), 4.54 (dd, $J_{H5b\text{-}H5a}$ = 11.5 Hz, $J_{H5b\text{-}H4}$ = 6.2 Hz, 1H, H5b), 4.25 (app. dt, $J_{H4\text{-}3}$ = 8.6 Hz, $J_{H4\text{-}H5a/b}$ = 6.1 Hz, 1H, H4), 2.12 (s, 3H, Ac-CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.4 (C=O, Ac), 166.0 (C=O, Bz), 165.4 (C=O, Bz), 165.0 (C=O, Bz), 133.7 (C_q, Ar-C), 133.5 (C_q, Ar-C), 133.2 (C_q, Ar-C), 129.9 (CH, Ar-C), 129.8 (CH, Ar-C), 129.7 (CH, Ar-C0, 129.4 (CH, Ar-C), 129.0 (CH, Ar-C0, 128.8 (CH, Ar-C), 128.6 (CH, Ar-C), 128.4 (CH, Ar-C), 128.3 (CH, Ar-C), 79.7 (CH, C1), 76.8 (CH, C2), 75.1 (CH, C3), 65.2 (CH_2, C5), 46.2 (CH, C4), 20.9 (Ac-CH_3); ESI HRMS *m/z* found: $(M+Na)^+$ 543.1079 $C_{28}H_{24}O_8S$, requires $(M+Na)^+$ 543.1084. Data was consistent with literature values.[3]

S1.3. Synthesis of 2-deoxy-2,2-*gem*-difluoro-3,5-tri-*O*-benzoyl-1-*O*-acetyl-thioribose 12

3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-α,β-D-ribofuranose

Commercial lactone **7** (20.0 g, 58.1 mmol, 1.0 equiv.) was dissolved in THF (110 mL) and the solution cooled to 0 °C. Li(O^{t}Bu)AlH (16.2 g, 63.8 mmol, 1.2 equiv.) was added at 0 °C. After 15 minutes, the reaction was quenched with saturated aqueous NH_4Cl solution (200 mL), filtered through a sintered funnel and the organic solvents removed from the mother liquor *in vacuo*. The crude was extracted with EtOAc (3 x 100 mL) and the combined organic layers were washed with H_2O (2 x 150 mL) and brine (150 mL), dried over anhydrous Na_2SO_4, filtered and dried *in vacuo* to obtain the crude as a yellow syrup (19.8 g, ~58.1 mmol, *quant*.). An analytically pure sample of the title compound was obtained *via* purification on silica gel *via* automated flash chromatography (0 – 50% Et_2O/petroleum ether) to obtain the title compound as a yellow syrup (19.1 g, 50.5 mmol, 87%). R_f 0.18 (1/4 Et_2O/petroleum ether); 1.7/1 ratio α/β; **α-anomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.13 – 8.01 (m, 4H, Ar-H), 7.66 – 7.36 (m, 6H, Ar-H), 5.54 – 5.44 (m, 2H, 2 x CH, H1 and H3), 4.80 – 4.73 (m, 1H, CH, H4), 4.67 (app. d (ov), $J_{H5a\text{-}H4}$ = 4.8 Hz, 1H, CH_2, H5a), 4.60 (dd, $J_{H5b\text{-}H5a}$ = 12.0 Hz, $J_{H5b\text{-}H4}$ = 4.4 Hz, 1H, CH_2, H5b), 3.33 (d, $J_{OH\text{-}F}$ = 4.3 Hz, 1H, 1-OH); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.2 (C=O, Bz), 165.2 (C=O, Bz), 133.9 (C_q, Ar-C), 133.3 (C_q, Ar-C), 130.1 (CH, Ar-C), 129.8 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.48 (CH, Ar-C), 128.45 (CH, Ar-C), 121.5 (dd, $J_{C2\text{-}F}$ = 271.9 Hz, $J_{C2\text{-}F}$ = 249.3 Hz, C_q, C2’), 96.1 (dd, $J_{C1\text{-}F}$ = 42.0 Hz, $J_{C1\text{-}F}$ = 23.5 Hz, CH, C1), 79.6 (t, $J_{C4\text{-}F}$ = 3.3 Hz, CH C4), 71.9 (dd, $J_{C3\text{-}F}$ = 36.0 Hz, $J_{C3\text{-}F}$ = 18.0 Hz, CH, C3), 63.2 (CH_2, C5); ^{19}F NMR (376 MHz, $CDCl_3$) δ -109.33 (ddd, $J_{F\text{-}F}$ = 252.0 Hz, J = 16.3 Hz, J = 6.7 Hz), -125.24 (app. d, $J_{F\text{-}F}$ = 251.8 Hz); ESI HRMS *m/z* found: $(M+H)^+$ 379.0970 $C_{19}H_{16}F_2O_4$, requires $(M+H)^+$ 379.0988. Data was consistent with literature values.[4,5]

(2*R*,3*R*,4*S*)-3,5-di-*O*-benzoyl-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*) 8

A solution of 3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-α,β-D-ribofuranose (19.1 g, 50.5 mmol, 1.0 equiv.) and $H_2NOMe{\cdot}HCl$ (6.65 g, 79.7 mmol, 1.5 equiv.) in 3/1 (*v/v*) MeCN/H_2O (410 mL) was cooled to 0 °C. Et_3N (11 mL, 79.7 mmol, 1.5 equiv.) and pyridinium *p*-toluenesulfonate (8.67 g, 34.5 mmol, 0.65 equiv.) was added and the solution stirred for a further 5 minutes at 0 °C and allowed to warm to rt, stirring vigorously. After 4 days, the solvent was removed *in vacuo* and the residue partitioned between EtOAc (1 L) and H_2O (900 mL). The organic layer was separated and washed with H_2O (3 x 900 mL) and brine (900 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain crude **8** as a white foamy syrup (19.4 g, ~47.6 mmol, 90%) which was used without further purification. R_f 0.57 (1/1 Et_2O/petroleum ether); 10/1 isomer ratio; **major isomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.12 – 7.96 (m, 4H, Ar-H), 7.66 – 7.49 (m, 3H, Ar-H and H1), 7.48 – 7.38 (m, 4H, Ar-H), 5.86 (ddd, $J_{H3\text{-}F}$ = 13.1 Hz, $J_{H3\text{-}F}$ = 10.5 Hz, $J_{H3\text{-}H4}$ = 5.9 Hz, 1H, H3), 4.68 (dd, $J_{H5a\text{-}H5b}$ = 11.9 Hz, $J_{H5a\text{-}H4}$ = 2.8 Hz, 1H, H5a), 4.60 – 4.52 (m, 1H, H4), 4.46 (dd, $J_{H5b-H5a}$ = 11.9 Hz, J_{H5b-H4} 5.9 Hz, 1H, H5b), 3.90 (s, 3H, OCH_3), 2.98 (s, 1H, 4-OH); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.9 (C=O, Bz), 165.1 (C=O, Bz), 142.5 (dd, J_{C1-F} = 32.4 Hz, J_{C1-F} = 28.9 Hz, C=N, C1), 133.9 (C_q, Ar-C), 133.3 (C_q, Ar-C), 130.1 (CH, Ar-C), 129.8 (CH, Ar-C), 129.5 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.4 (CH, Ar-C), 116.3 (dd, J_{C2-F} = 246.5 Hz, J_{C2-F} = 244.2 Hz, C_q, C2), 73.0 (dd, J_{C3-F}= 27.1 Hz, J_{C3-F} = 24.6 Hz, CH, C3), 68.3 (CH, C4), 65.6 (CH_2, C5), 63.0 (OCH_3); ^{19}F NMR (376 MHz, $CDCl_3$) δ -104.57 (ddd, J_{F-F} = 277.2 Hz, J_{F-H3} 10.4 Hz, J_{F-H1} = 6.8 Hz), -106.98 (ddd, J_{F-F} = 277.2 Hz, J_{F-H3} = 13.1 Hz, J_{F-H1} = 6.2 Hz); ESI HRMS *m/z* found: $(M+H)^+$ 408.1267 $C_{20}H_{19}F_2NO_6$, requires $(M+H)^+$ 408.1253. This compound was reported previously,[2] but with no analytical data.

(2*S*,3*R*,4*S*)-3,5-di-*O*-benzoyl-4-bromo-4-*O*-(2’,4’,5’-trichlorophenylsulfonyl)-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*)

To a solution of **8** (19.4 g, 47.6 mmol, 1.0 equiv.) in MeCN (125 mL) was added 2,4,5-trichlorobenzenesulfonyl chloride (14.7 g, 52.4 mmol, 1.1 equiv.) and *N*-methylimidazole (4.2 mL, 52.4 mmol, 1.1 equiv.). The resultant suspension was stirred at rt for 3 h, partitioned between H_2O (800 mL) and EtOAc (1.2 L). The organic layer was separated and the aqueous layer extracted with EtOAc (300 mL). The combined organic phases were washed with 5% (*v/v*) brine/H_2O solution (2 x 600 mL) and brine (600 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain the crude as a beige waxy solid which was triturated from rt Et_2O and the white precipitate collected by suction filtration. The mother liquor was dried *in vacuo* and triturated from hot Et_2O and the retentate collected by suction filtration and the solids combined, obtaining the title compound (*E*/*Z*) as a white solid (19.3 g, 29.6 mmol, 62%). R_f 0.51 (1/4 Et_2O/petroleum ether); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 – 8.01 (m, 3H, Ar-H), 7.91 – 7.85 (m, 2H), Ar-H, 7.67 – 7.54 (m, 2H, Ar-H), 7.52 – 7.38 (m, 4H, Ar-H and H1), 6.14 (ddd, J_{H3-F} = 13.2 Hz, J_{H3-F} = 10.3 Hz, J_{H3-H4} = 2.8 Hz, 1H, H3), 5.63 (app. dt, J_{H4-H5b} = 8.5 Hz, $J_{H4-H3/H5a}$ = 2.6 Hz, 1H, H4), 4.81 (dd, $J_{H5a-H5b}$ = 12.8 Hz, J_{H5a-H4} = 2.5 Hz, 1H, H5a), 4.68 (dd, $J_{H5b-H5a}$ = 12.8 Hz, J_{H5b-H4} = 8.5 Hz, 1H, H5b), 3.90 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.8 (C=O, Bz), 164.1 (C=O, Bz), 140.9 (t, J_{H1-F} = 31.8 Hz, C1), 139.1 (C_q Ar-C) 134.3 (C_q, Ar-C), 134.1 (C_q, Ar-C), 133.6 (C_q, Ar-C), 133.3 (C_q, Ar-C), 132.0 (C_q, Ar-C), 131.8 (CH, Ar-C), 131.7 (CH, Ar-C), 130.1 (CH, Ar-C), 129.6 (CH, Ar-C), 128.8 (CH, Ar-C), 128.7 (CH, Ar-C), 128.5 (CH, Ar-C), 128.2 (CH, Ar-C), 115.6 (dd, J_{C2-F} = 247.3 Hz, J_{C2-F} = 243.1 Hz, C_q, C2), 78.5 (CH, C4), 71.3 (dd, J_{C3-F} = 29.8 Hz, J_{C3-F} = 25.3 Hz, C3), 63.3 (OCH_3), 62.2 (CH_2, C5); ^{19}F NMR (376 MHz, $CDCl_3$) δ -102.28 (ddd, J_{F-F} = 283.7 Hz, J_{F-H3} = 10.1, J_{F-H1} = 5.4 Hz), -105.24 (ddd, J_{F-F} = 283.7, J_{F-H3} = 13.3, J_{F-H1} = 6.2 Hz); ESI HRMS *m/z* found: $(M+H)^+$ 649.9996 $C_{26}H_{20}Cl_3F_2NO_8S$, requires $(M+H)^+$ 650.0016.

(2*S*,3*R*,4*S*)-3,5-di-*O*-benzoyl-4-bromo-4-*O*-(2',4',5'-trichlorophenylsulfonyl)-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (E/Z) 9

To a solution of (2*S*,3*R*,4*S*)-3,5-di-*O*-benzoyl-4-bromo-4-*O*-(2',4',5'-trichlorophenylsulfonyl)-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E/Z*) (19.3 g, 29.6 mmol, 1.0 equiv.) in 2-butanone (100 mL) was added LiBr (12.9 g, 148 mmol, 5.0 equiv.). The solution was heated to 80 °C for 18 h, poured onto ice-water (1 L) and extracted with CH_2Cl_2 (5 x 200 mL). The combined organic layers were washed with H_2O (2x 700 mL) and brine (700 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain the crude which was purified on silica gel *via* flash chromatography (0 – 30% Et_2O/petroleum ether) to obtain **9** as a yellow oil (13.5 g, 28.7 mmol, 97%). R_f 0.77 (1/1 Et_2O/petroleum ether); 6.3/1.0 ratio isomers; **major isomer:** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.18 – 8.11 (m, 2H, Ar-H), 8.10 – 8.04 (m, 2H, Ar-H), 7.68 – 7.54 (m, 2H, Ar-H), 7.54 – 7.40 (m, 5H, Ar-H and H1), 6.10 (ddd, $J_{H3\text{-}F}$ = 11.2 Hz, $J_{C3\text{-}F}$ = 9.8 Hz, $J_{H3\text{-}H4}$ = 2.8 Hz, 1H, H3), 4.83 – 4.68 (m, 2H, H5a and H4), 4.48 (dd, $J_{H5;b\text{-}H5a}$ = 13.5 Hz, $J_{H5b\text{-}H4}$ = 9.4 Hz, 1H, H5b), 3.83 (s, 3H, OCH_3); ^{13}C NMR (101 MHz, CDCl3) δ 165.6 (C=O, Bz), 164.6 (C=O, Bz), 141.4 (dd, $J_{C1\text{-}F}$ = 32.0 Hz, $J_{C1\text{-}F}$ = 30.3 Hz, C=N, C1), 134.0 (C_q, Ar-C), 133.4 (C_q, Ar-C), 130.3 (CH, Ar-C), 130.1 (CH, Ar-C), 130.0 (CH, Ar-C), 129.9 (CH, Ar-C), 128.7 (CH, Ar-C), 128.5 (CH, Ar-C), 115.8 (dd, $J_{C2\text{-}F}$ = 246.7 Hz, $J_{C2\text{-}F}$ = 245.7 Hz, C_q, C2), 69.8 (dd, $J_{C3\text{-}F}$ = 28.8 Hz, $J_{C3\text{-}F}$ = 28.1 Hz, CH, C3), 64.7 (CH_2, C5), 63.0 (OCH_3), 44.3 (t, $J_{C4\text{-}F}$ = 2.0 Hz, C-Br, C4); ^{19}F NMR (376 MHz, $CDCl_3$) δ -102.74 (ddd, $J_{F\text{-}F}$ = 279.7 Hz, $J_{F\text{-}H3}$ = 11.1 Hz, $J_{F\text{-}H1}$ = 5.7 Hz), -105.10 (ddd, $J_{F\text{-}F}$ = 279.7 Hz, $J_{F\text{-}H3}$ = 9.4 Hz, $J_{F\text{-}H1}$ = 7.1 Hz); ESI HRMS *m/z* found: $(M+H)^+$ 470.0414 $C_{20}H_{18}BrF_2NO_5$.

3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-α,β-(4-thio-D-ribofuranose) 10

To a solution of **9** (13.5 g, 28.7 mmol, 1.0 equiv.) in MeCN (59.0 mL) was added 50% (*w/v*) glyoxylic acid solution (11.5 mL, 201 mmol, 7.0 equiv.). The solution was heated to 70 °C for 18 h, the solvent removed *in vacuo* and the residue partitioned between EtOAc (600 mL) and H_2O (600 mL). The organic phase was separated and the aqueous extracted with EtOAc (2 x 250 mL) and the combined organic phases washed with H_2O (4 x 400 mL) and brine (400 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain the crude aldehyde as a brown syrup (13.2 g, ~28.7 mmol *quant.*) which was used immediately in the next step without further purification. R_f 0.11 (1/1 Et_2O/petroleum ether); ESI HRMS *m/z* found: 441.0415 $C_{19}H_{15}BrF_2O_5$, requires $(M+H)^+$ 441.0415. Crude aldehyde (8.43 g, 19.1 mmol, 1.0 equiv.) was dissolved in DMF (24 mL) and the solution cooled to 0 °C. $NaSH{\cdot}H_2O$ (1.84 g, 24.8 mmol, 1.3 equiv.) was dissolved in a minimum volume of H_2O (<3 mL), added to the cooled solution and the whole stirred at 0 °C for 1h. The solution was poured onto H_2O (500 mL) and extracted with EtOAc (4 x 125 mL). The organic phases were combined and washed with brine (400 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain crude **10**

(7.32 g, 18.6 mmol, 97%) as a yellow syrup . An analytically pure sample of **10** was obtained *via* purification on silica gel *via* automated flash chromatography (0 – 100% Et_2O/pet. ether). R_f 0.48 (1/1, Et_2O/petroleum ether); 1/3 anomer ratio; **major anomer:** ^1H NMR (400 MHz, $CDCl_3$) δ 8.11 – 8.01 (m, 2H, Ar-H), 7.97 – 7.90 (m, 2H, Ar-H), 7.65 – 7.27 (m, 6H, Ar-H), 6.04 (ddd, $J_{H3\text{-}F}$ = 17.7 Hz, $J_{H3\text{-}F}$ = 7.6 Hz, $J_{H3\text{-}H4}$ = 4.6 Hz, 1H, H3), 5.34 (dd, $J_{H1\text{-}F}$ = 7.0 Hz, $J_{H1\text{-}F}$ 2.4 Hz, 1H, H1), 5.30 (s, 1H, 1-OH), 4.64 (dd, $J_{H5a\text{-}H5b}$ = 11.6 Hz, $J_{H5a\text{-}H4}$ = 6.6 Hz, 1H, H5a), 4.60 – 4.53 (m, 1H, H5b), 3.89 – 3.83 (m, 1H, H4); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.0 (C=O, Bz), 165.0 (C=O, Bz), 133.9 (C_q, Ar-C), 133.3 (C_q, Ar-C), 130.1 (CH, Ar-C), 130.1 (CH, Ar-C), 129.8 (CH, Ar-C), 129.7 (CH, Ar-C), 129.2 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.5 (CH, Ar-C), 128.4 (CH, Ar-C), 128.4 (CH, Ar-C), 123.5 (dd, $J_{C2\text{-}F}$ = 211.4 Hz, $J_{C2\text{-}F}$ = 125.9 Hz, C_q, C2), 76.2 (dd, $J_{C1\text{-}F}$ = 34.5 Hz, $J_{C1\text{-}F}$ = 22.1 Hz, CH, C1), 72.2 (dd, $J_{C3\text{-}F}$ = 27.7 Hz, $J_{C3\text{-}F}$ = 19.0 Hz, CH, C3), 65.0 (CH_2, C5), 41.8 (d, $J_{C4\text{-}F}$ = 5.8 Hz, CH, C4); ^{19}F NMR (376 MHz, $CDCl_3$) δ -119.43 (app. d, $J_{F\text{-}F}$ = 234.1 Hz), -123.72 (ddd, $J_{F\text{-}F}$ = 233.9 Hz, $J_{F\text{-}H3}$ = 17.7 Hz, $J_{F\text{-}H1}$ = 7.0 Hz); ESI HRMS *m/z* found: $(M+H)^+$ 395.0764, $C_{19}H_{15}F_2O_5S$, requires $(M+H)^+$ 395.0765.

3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-*O*-mesyl-1-β-(4-thio-D-ribofuranose) 11

To a solution of **10** (0.695 g, 1.76 mmol, 1.0 equiv.) in CH_2Cl_2 (8.8 mL) was added MsCl (0.20 mL, 2.64 mmol, 1.5 equiv.) and Et_3N (0.40 mL, 2.64 mmol, 1.5 equiv.). The solution was stirred at rt for 3.5 h, diluted with CH_2Cl_2 (100 mL) and washed with H_2O (100 mL), saturated aqueous $NaHCO_3$ solution (100 mL) and brine (100 mL). The organic phase was dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain the crude as a yellow syrup which was purified on silica gel *via* automated flash chromatography (0 – 50% Et_2O/pet/ether) to obtain **11** as a yellow foam (0.744 g, 1.57 mmol, 89%), a mixture of anomers (1/4 ratio α/β). The β-anomer was separated by crystallisation from hot Et_2O (15 mL) to obtain **11-β** as colourless needles. The solvent from the mother liquor was removed *in vacuo* and the residue crystallised by vapour diffusion from CH_2Cl_2/hexane (5 mL each) and the two sets of crystalline solids combined to obtain **11-β** (0.603 g total, 1.28 mmol, 72% total). R_f 0.28 (1/1 Et_2O/petroleum ether); $[\alpha]_D^{25.8}$ -52.6 (*c* 1.4, MeCN); ^1H NMR (400 MHz, $CDCl_3$) δ 8.13 – 8.01 (m, 2H, Ar-H), 8.01 – 7.93 (m, 2H, Ar-H), 7.61 (t, J_{vic} = 7.5 Hz, 1H, Ar-H), 7.57 – 7.40 (m, 3H, Ar-H), 7.41 – 7.29 (m, 2H, Ar-H), 6.04 (br d, $J_{H1\text{-}F}$ = 6.4 Hz, 1H, H1), 6.00 (ddd, $J_{H3\text{-}F}$ = 20.6 Hz, $J_{H3\text{-}H4}$ = 8.5 Hz, $J_{H3\text{-}F}$ = 3.9 Hz, 1H), 4.69 (dd, $J_{H5a\text{-}H5b}$ = 11.8 Hz, $J_{H5a\text{-}H4}$ = 5.5 Hz, 1H, H5a), 4.53 (dd, $J_{H5b\text{-}H5a}$ = 11.8 Hz, $J_{H5b\text{-}H4}$ 5.5 Hz, 1H, H5b), 3.92 (dt, $J_{H4\text{-}H3}$ = 8.5 Hz, $J_{H4\text{-}H5a/h5b}$ = 5.5 Hz, 1H, H4), 3.07 (s, 3H, Ms-CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.8 (C=O, Bz), 164.8 (C=O, Bz), 134.1 (C_q, Ar-C), 133.5 (C_q, Ar-C), 130.2 (CH, Ar-C), 129.7 (CH, Ar-C), 129.1 (CH, Ar-C), 128.7 (CH, Ar-C), 128.5 (CH, Ar-C), 128.1 (CH, Ar-C), 121.7 (dd, $J_{C2\text{-}F}$ = 269.9 Hz, $J_{C2\text{-}F}$ = 253.3 Hz, C2), 81.6 (dd, $J_{C1\text{-}F}$ = 38.0 Hz, $J_{C1\text{-}F}$ = 21.9 Hz, CH, C1), 71.4 (dd, $J_{C3\text{-}F}$ = 25.8 Hz, $J_{C3\text{-}F}$ = 18.2 Hz, CH, C3), 64.0 (CH_2, C5), 42.3 (d, $J_{C4\text{-}F}$ = 6.2 Hz, CH, C4), 40.1 (Ms-CH_3); ^{19}F NMR (376

MHz, $CDCl_3$) δ -116.33 (dd, J_{F-F} = 233.9 Hz, J_{F-H3} = 3.3 Hz), -123.91 (ddd, J_{F-F} = 233.9 Hz, J_{F-H3} = 20.6 Hz, J_{F-H1} = 6.4 Hz); ESI HRMS *m/z* found: $(M+Na)^+$ 495.0372 $C_{20}H_{18}F_2O_7S_2$, requires $(M+Na)^+$ 495.0354.

1-*O*-acetyl-3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-α,β-(4-thio-D-ribofuranose) 12

To a solution of **10** (0.990 g, 2.51 mmol, 1.0 equiv.) in CH_2Cl_2 (13 mL) was added Ac_2O (0.28 mL, 3.01 mmol, 1.2 equiv.) and Et_3N (0.42 mL, 3.01 mmol, 1.2 equiv.). The solution was stirred at rt for 5 h. The solution was diluted with CH_2Cl_2 (100 mL) and washed with saturated aqueous $NaHCO_3$ solution (100 mL) and brine (100 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed *in vacuo* to obtain the crude which was purified on silica gel *via* automated flash chromatography (0 – 30% Et_2O/pet. ether) to obtain **12** as a colourless syrup (0.963 g, 2.21 mmol, 88%). R_f 0.50 (1/1, Et_2O/pet ether); 3/2 ratio anomers; **major anomer:** ^1H NMR (400 MHz, $CDCl_3$) δ 8.14 – 8.00 (m, 2H, Ar-H), 7.96 – 7.90 (m, 2H, Ar-H), 7.65 – 7.52 (m, 1H, Ar-H), 7.54 – 7.39 (m, 4H, Ar-H), 7.34 – 7.27 (m, 1H, Ar-H), 6.04 (d, J_{H1-F} = 7.7 Hz, 1H, H1), 6.05 – 5.95 (ov. m, 1H, H3), 4.66 (dd, $J_{H5a-H5b}$ = 11.6 Hz, J_{H5a-H4} = 6.0 Hz, 1H, H5a), 4.48 (dd, $J_{H5b-H5a}$ = 11.6 Hz, J_{H5b-H4} = 5.7 Hz, 1H, H5b), 3.89 (app. dt, J_{H4-H3} = 8.7 Hz, $J_{H4-H5a/H5b}$ = 5.9 Hz, 1H, H4), 2.13 (s, 3H, Ac-CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 168.9 (C=O, Ac), 165.8 (C=O, Bz), 165.0 (C=O, Bz), 134.0 (C_q, Ar-C), 133.3 (C_q, Ar-C), 130.2 (CH, Ar-C), 129.7 (CH, Ar-C), 129.2 (CH, Ar-C), 128.6 (CH, Ar-C), 128.3 (CH, Ar-C), 123.5 (dd, $J_{C2'-F}$ = 370.5 Hz, $J_{C2'-F}$ = 119.4 Hz, C_q, C2’), 75.0 (dd, J_{C1-F} = 38.2 Hz, J_{C1-F} 20.6 Hz, CH, C1), 72.0 (dd, J_{C3-F} = 25.8 Hz, J_{C3-F} = 18.6 Hz, CH, C3), 64.5 (CH_2, C5), 41.5 (d, J_{C4-F} = 6.2 Hz, CH, C4), 20.8 (Ac-CH_3); ^{19}F NMR (377 MHz, $CDCl_3$) δ -117.94 (dd, J_{F-F} = 236.0 Hz, J_{F-H3} = 3.6 Hz), -123.92 (ddd, J_{F-F} = 235.8 Hz, J_{F-H3} = 20.5 Hz, J_{F-H1} = 7.9 Hz); ESI HRMS *m/z* found: $(M+Na)^+$ 459.0697 $C_{21}H_{18}F_2O_6SNa^+$, requires $(M+Na)^+$ 459.0684.

S1.4. Synthesis of Thio*ribo*uridine/cytidine & Thio*arabino*uridine/cytidine

2',3',5'-tri-*O*-benzoyl,1'-β-(4'-thio-D-ribofuranosyl)uracil 13

Uracil (2.90 g, 26.0 mmol, 1.4 equiv.) was suspended in pyridine (19.4 mL), and the flask charged with hexamethyldisilazane (39.8 mL, 190 mmol, 9.9 equiv.) and the mixture refluxed for 3 h. The solvent was removed *in vacuo* and the flask immediately stoppered and flushed with N_2, to obtain crude 2,4-*O*-silylated uracil as a colourless oil. This was transferred under N_2, rinsing the flask with MeCN (4 x 25 mL), to a flask containing a solution of **1** (10.0 g, 19.2 mmol, 1.0 equiv.) in MeCN (100 mL). The solution was cooled to 0 °C, TMSOTf (2.70 mL, 15.0 mmol, 0.78 equiv.) was added dropwise and the solution stirred at 0 °C for a further 10 minutes before heating to 75 °C for 72 h. The reaction was cooled to 0 °C and quenched with Et_3N (1.8 mL) and stirred at 0 °C for 10 minutes. The solvent was reduced to <10 mL *in vacuo* and the residue re-diluted in EtOAc (600 mL). The organic phase was washed with saturated aqueous $NaHCO_3$ solution (3 x 250 mL) and brine (250 mL), dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to give crude **13** as an orange-brown oil, which was passed through a silica gel plug (3/7 EtOAc/petroleum ether), the filtrate was dried *in vacuo* and then triturated from boiling 2/1 (*v*/*v*) petroleum ether/EtOAc (150 mL) solution, the solid collected by suction filtration and washed with rt petroleum ether (50 mL) to obtain **13** as a white amorphous solid (8.12 g, 14.2 mmol, 74%). R_f 0.80 (1/1, EtOAc/hexane); mp 213 – 215 °C; $[\alpha]_D^{27}$ -80.0 (*c* 1.0, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.16 – 8.11 (m, 2H, Ar-H), 8.08 – 8.02 (m, 2H, Ar-H), 7.97 – 7.90 (m, 2H, Ar-H), 7.74 (d, J_{H6-H5} = 8.2 Hz, 1H, H6), 7.61 – 7.44 (m, 7H, Ar-H), 7.43 – 7.35 (m, 2H, Ar-H), 6.69 (d, $J_{H1'-H2'}$ = 6.8 Hz, 1H, H1'), 5.99 (dd, J_{H2-H3} = 3.6 Hz, J_{H2-H1} = 1.7 Hz, 1H, H2), 5.91 (dd, J_{H3-H4} = 8.6 Hz, J_{H3-H2} = 3.6 Hz, 1H, H3), 5.56 (dd, J_{H5-H6} = 8.2 Hz, J_{H5-NH} = 1.2 Hz, H5), 4.85 (dd, $J_{H5'a-H5'b}$ = 12.0 Hz, $J_{H5'a-H4'}$ = 5.6 Hz, 1H, H5'a), 4.71 (dd, $J_{H5'b-H5'a}$ = 12.0 Hz, $J_{H5'b-H4'}$ = 4.7 Hz, 1H, H5'b), 4.06 (m, 1H, H4'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.1 (C=O, Bz), 165.4 (C=O, Bz), 165.2 (C=O, Bz), 162.1 (C=O, C4), 150.4 (C=O, C2), 133.7 (C_q, Ar-C), 133.5 (C_q, Ar-C), 133.2 (C_q, Ar-C), 129.9 (CH, Ar-C), 129.8 (CH, Ar-C), 129.7 (CH, Ar-C), 129.4 (CH, Ar-C), 129.0 (CH, Ar-C0, 128.8 (CH, Ar-C), 128.6 (CH, Ar-C), 128.4 (CH, Ar-C), 128.3 (CH, Ar-C), 79.7 (CH, C1), 76.8 (CH, C2), 75.1 (CH, C3), 65.2 (CH_2, C5), 46.2 (CH, C4), 20.9 (Ac-CH_3); ESI HRMS *m/z* found: $(M+Na)^+$ 543.1079 $C_{28}H_{24}O_8S$, requires $(M+Na)^+$ 543.1084; Anal. Calcd for $C_{30}H_{24}N_2O_8S$: C, 62.93; H, 4.23; N, 4.89; S, 5.60, Found C, 63.18; H, 4.30; N. 5.09; S, 5.61, as reported.[6]

1'-β-(4'-thio-D-ribofuranosyl)uracil 14

A suspension of **13** (8.12 g, 14.2 mmol, 1.0 equiv.) in MeOH (95 mL) was cooled to 0 °C. The flask was charged with a 7M solution of NH_3 in MeOH (18.2 mL, 128 mmol, 9.0 equiv.) at 0 °C, warmed to 40 °C and stirred for 72 h. The solvent was removed *in vacuo* to give an orange solid which triturated with CH_2Cl_2 (150 mL), filtered through a sintered funnel and the filtrate washed with CH_2Cl_2 (30 mL) and acetone (15 mL) to obtain a beige solid which was then purified on octadecyl modified silica gel *via* automated flash chromatography (H_2O) to afford **14** as a white foam (3.55 g, 13.6 mmol, 96%). R_f 0.28 (15/85 MeOH/CH_2Cl_2); $[\alpha]_D^{24.1}$ +16.00 (*c* 1.0, H_2O);[7,8] ^{1}H NMR (400 MHz, D_2O) δ 8.18 (d, $J_{H6\text{-}H5}$ = 8.1 Hz, 1H H6), 5.95 (d, $J_{H1'\text{-}H2'}$ = 5.7 Hz, 1H), 5.90 (d, $J_{H5\text{-}H6}$ = 8.1 Hz, 1H, H5), 4.38 – 4.33 (m, 1H, 2H), 4.19 (app. t, $J_{H3'\text{-}H2',H3'\text{-}H4'}$ = 4.1 Hz, 1H, H3'), 3.87 (dd, $J_{H5'a\text{-}H5'b}$ = 12.0 Hz, $J_{H5'a\text{-}H4'}$ = 5.3 Hz, 1H, H5'a), 3.82 (dd, $J_{H5'b\text{-}H5'a}$ = 12.0 Hz, $J_{H5'b\text{-}H4'}$ = 5.5 Hz, 1H), 3.47 (m, 1H, H4'); ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.16 (s, 1H, N-H), 8.01 (d, $J_{H6'\text{-}H5'}$ = 8.1 Hz, 1H, H6), 5.91 (d, $J_{H1'\text{-}H2'}$ = 7.4 Hz, 1H), 5.70 (d, $J_{H5'\text{-}H6'}$ = 8.0 Hz, 1H), 4.28 – 4.08 (m, 1H, H2'), 4.04 (s, 1H, H3'), 3.61 (m, 2H), 3.19 (d, *J* = 11.8 Hz, 1H); ^{13}C NMR (101 MHz, D_2O) δ 166.1 (C=O, C4), 152.3 (C=O, C2), 143.0 (CH alkene, C6), 102.4 (CH alkene, C5), 77.4 (CH, C2'), 73.4 (CH, C3'), 64.3 (CH, C1'), 62.4 (CH_2, C5'), 52.2 (CH, C4'); ESI HRMS *m/z* found: $(M+H)^+$ 261.0530 $C_9H_{12}N_2O_5S$, requires $(M+H)^+$ 261.0540. NMR data was consistent with literature values (in DMSO-D_6).[7]

2',3',5'-tri-*O*-benzoyl-4-*C*-(1,2,4-triazole)-1'-β-(4'-thio-D-ribofuranosyl)uracil

A suspension of **13** (2.00 g, 3.49 mmol, 1.0 equiv.) in MeCN (35.0 mL) was cooled to 0 °C. Et_3N (11.2 mL, 80.3 mmol, 23 equiv.), 1,2,4-triazole (5.44 g, 78.5 mmol, 23 equiv.) and $POCl_3$ (0.80 mL, 8.52 mmol, 2.4 equiv.) were added and the solution stirred for a further 10 minutes at 0 °C before warming to rt, stirring vigorously. After 3 h, the solution was poured into an ice-cold saturated aqueous $NaHCO_3$ solution (150 mL) and diluted with EtOAc (150 mL). The organic layer was separated, washed with saturated aqueous $NaHCO_3$ solution (2 x 100 mL) and brine (150 mL), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the crude title compound as a yellow foam (2.29 g, ~3.49 mmol, *quant.*), which was used immediately in the next step without further purification. R_f 0.31 (1/1 EtOAc/petroleum ether); $[\alpha]_D^{24.2}$ -78.4 (*c* 0.6, CH_2Cl_2); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.22 (s, 1H, triazole CH), 8.51 (d, $J_{H6\text{-}H5}$ = 7.4 Hz, 1H, H6), 8.21 – 8.10 (m, 3H, 2 x Ar-H CH and triazole CH), 8.06 – 8.02 (m, 2H, Ar-H), 8.00 – 7.91 (m, 2H, Ar-H), 7.70 – 7.36 (m, 9H, Ar-H), 6.92 (d, $J_{H5\text{-}H6}$ = 7.3 Hz, 1H, H5), 6.85 (d, $J_{H1'\text{-}H2'}$ = 6.5 Hz, 1H, H1'), 6.06 (dd, $J_{H2'\text{-}H1'}$ = 6.5 Hz, $J_{H2'\text{-}H3'}$ = 4.0 Hz, 1H, H2'), 5.95 (app. t, $J_{H3'\text{-}H2'/H4'}$ = 3.9 Hz, 1H, H3'), 4.86 (dd, $J_{H5'a\text{-}H5'b}$ = 12.0 Hz, $J_{H5'a\text{-}H4'}$ = 5.5 Hz, 1H, H5'a), 4.72 (dd, $J_{H5'b\text{-}H5'a}$ = 12.0 Hz, $J_{H5'b\text{-}H4'}$ = 5.0 Hz, 1H, H5'b), 4.21 – 4.14 (m, 1H, H4'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.1 (C=O, Bz), 165.4 (C=O, Bz), 165.2 (C=O, Bz), 159.1 (C=O, C4), 154.8 (C=O, C2), 154.2 (triazole CH), 147.1 (CH alkene, C6), 143.4 (triazole CH), 133.9 (C_q, Ar-C), 133.8 (C_q, Ar-C), 130.1 (C_q, Ar-C), 130.0 (CH, Ar-C), 129.8 (CH, Ar-C), 129.1 (CH, Ar-C), 128.8 (CH, Ar-C), 128.7 (CH, Ar-C), 128.6 (CH, Ar-C), 128.3 (CH, Ar-C), 96.1 (CH alkene, C5), 76.7 (CH, C2'), 74.4 (CH, C3'), 64.3 (CH_2, C5'), 63.9 (CH, C1'), 48.2 (CH, C4'); ESI HRMS *m/z* found: $(M+H)^+$ 624.1567 $C_{32}H_{25}N_5O_7S$, requires $(M+H)^+$ 624.1547.

1'-β-(4'-thio-D-ribofuranosyl)cytosine 15

Crude 2',3',5'-tri-*O*-benzoyl-4-*C*-(1,2,4-triazole)-1'-β-(4'-thio-D-ribofuranosyl)uracil (2.29 g, ~3.49 mmol, 1.0 equiv.) was dissolved in 1,4-dioxane (15 mL) and the solution charged with 25% (*w/v*) NH_4OH solution (15 mL, 107 mmol, 30 equiv.) and the solution stirred at rt in a sealed flask overnight. The solvents were removed *in vacuo* and the crude suspended in MeOH (16 mL) and 7M NH_3/MeOH solution (4.5 mL, 31.4 mmol, 9.0 equiv.) and stirred at 40 °C in a sealed flask for 24 h. The solvent was removed *in vacuo* and the crude purified on silica gel *via* automated flash chromatography (0 – 30% MeOH/$CHCl_3$) to obtain **15** as a yellow foam (0.698 g, 2.69 mmol, 77%). R_f 0.27 (1/9 H_2O/MeCN); $[\alpha]_D^{26.0}$ -20.57 (*c* 0.9, H_2O); ^1H NMR (400 MHz, D_2O) δ 8.18 (d, $J_{H6\text{-}H5}$ = 7.6 Hz, 1H, H6), 6.07 (d, $J_{H5\text{-}H6}$ = 7.6 Hz, 1H, H5), 5.97 (d, $J_{H1'\text{-}H2'}$ = 5.2 Hz, 1H, H1'), 4.39 – 4.26 (m, 1H, H3'), 4.26 – 4.10 (m, 1H, H2'), 3.92 (dd, $J_{H5'a\text{-}H5'b}$ = 12.0 Hz, $J_{H5'a\text{-}H4'}$ = 5.1 Hz, 1H, H5'a), 3.84 (dd, $J_{H5'b\text{-}H5'a}$ = 12.0 Hz, $J_{H5'b\text{-}H4'}$ = 5.6 Hz, 1H, H5b), 3.63 – 3.44 (m, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 165.9 (C-NH_2, C4), 158.2 (C=O, C2), 142.8 (CH alkene, C6), 96.4 (CH alkene, C5), 77.6 (CH, C3'), 73.1 (CH, C2'), 65.0 (CH, C1'), 62.2 (CH_2, C5'), 51.7 (CH, C4'); ESI HRMS *m/z* found: $(M+H)^+$ 260.0714 $C_{13}H_{13}N_3O_4S$, requires 260.0700. NMR data was consistent with literature values.[8]

1'-β-(4'-sulfinyl(S/R)-D-ribofuranosyl)cytosine 16

A solution of **15** (45.0 mg, 0.174 mmol, 1.0 equiv.) in 2/1 (*v*/*v*) H_2O/MeCN solution (0.87 mL) was cooled to 0 °C and *m*-CPBA (43.0 mg, 0.191 mmol, 1.1 equiv.) added. After 18 h the solvents were removed *in vacuo* and the crude purified on octadecyl modified silica gel *via* flash chromatography (0 – 100% MeCN/H_2O) to obtain **16** as a white solid (43.0 mg, 0.156 mmol, 90%) with 20.0 mg purified further *via* preparative HPLC (Table 1, retention time = 4.4 minutes). Sulfoxide **16** was obtained as a white solid, a mixture of diastereoisomers (7.1 mg, 26.5 μmol). R_f 0.57 (1/4, H_2O/MeCN); 1.1/1 diastereoisomer ratio; **major diastereoisomer:** ^1H NMR (400 MHz, D_2O) δ 7.60 (d, $J_{H6\text{-}H5}$ = 7.4 Hz, 1H, H6), 6.00 (ov. d, $J_{H5\text{-}H6}$ = 7.3 Hz, 1H, H5), 5.77 (d, $J_{H1'\text{-}H2'}$ = 9.2 Hz, 1H, H1'), 4.78 (dd, $J_{H2'\text{-}H1'}$ = 9.1 Hz, $J_{H2'\text{-}H3'}$ = 4.9 Hz, 1H, H2'), 4.30 (app. t, $J_{H3'\text{-}H2'/H4'}$ = 4.7 Hz, 1H, H3'), 4.02 (dd, $J_{H5'a\text{-}H5'b}$ = 12.2 Hz, $J_{H5'a\text{-}H4'}$ = 5.2 Hz, 1H, H5'a), 3.99 – 3.95 (ov. m, 1H, H5'b), 3.57 (app. dt, $J_{H4'\text{-}H5'b}$ = 9.5 Hz, $J_{H4'\text{-}H5'a/H3'}$ = 4.9 Hz, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 166.1 (C-NH_2, C4), 157.8 (C=O, C2), 146.4 (CH alkene, C6), 96.4 (CH alkene, C5), 73.1 (CH, C1'), 71.5 (CH, C2'), 70.4 (CH, C3'), 65.7 (CH, C4'), 56.3 (CH_2, C5'); **minor diastereoisomer:** ^1H NMR (400 MHz, D_2O) δ 7.55 (d, $J_{H6\text{-}H5}$ = 7.6 Hz, 1H, H6), 6.01 (d, $J_{H5\text{-}H6}$ = 7.5 Hz, 1H, H5), 5.00 (d, $J_{H1'\text{-}H2'}$ = 8.1 Hz, 1H), 4.68 – 4.64 (ov. m, 1H, H2'), 4.41 (app. t, $J_{H3'\text{-}H2'/H4'}$ = 3.5 Hz, H3'), 3.99 – 3.94 (ov. m, 2H, H5'a and H5'b), 3.39 – 3.30 (m, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 166.8 (C-NH_2, C4), 157.2 (C=O, C2), 142.8 (CH alkene, C6), 123.1, 96.7 (CH alkene, C5), 91.5 (CH, C1'), 74.8 (CH, C4'), 73.3 (CH, C2'), 70.4 (CH, C3'), 58.2 (CH_2, C5'); NSI HRMS *m/z* found: $(M-H)^-$ 274.0510 $C_9H_{12}O_5N_3S$, requires 274.0498.

Time (minutes)	**%A (H_2O)**	**%B (MeOH)**
0.0	100	0
10.0	100	0
12.0	0	100
15.0	0	100
15.1	100	0

Table 1. Preparative HPLC linear gradient system for purification of **16**. A 250 x 21.2 mm column packed 5 µ particle size with Polaris 5 C18-A was employed to load the sample The flow rate was 20 mL/minute. A solution of of **16** in H_2O (100 mg/mL) was prepared and 50 μL injected into the prep HPLC system.

2',2-anhydro-1'-β-(4'-thio-D-ribofuranosyl)uracil 17

A solution of **14** (1.00 g, 3.84 mmol, 1.0 equiv.), $(PhO)_2CO$ (0.910 g, 4.23 mmol, 1.1 equiv.) and $NaHCO_3$ (32.0 mg, 0.384 mmol, 0.10 equiv.) in DMF (1.5 mL) was stirred vigorously at 100 °C for 18 h. The solvent was removed *in vacuo* and the residue triturated from EtOAc (100 mL), filtered through a sintered funnel and the filtrate washed with EtOAc (20 mL) to afford **17** as a tan solid (0.910 g, 3.76 mmol, 98%). R_f 0.35 (1/4 MeOH/CH_2Cl_2); $[\alpha]_D^{25.7}$ -107.8 (*c* 1.2, H_2O); ^{1}H NMR (400 MHz, D_2O) δ 7.73 (d, $J_{H6\text{-}H5}$ = 7.5 Hz, 1H, H6), 6.21 (d, $J_{H1'\text{-}H2'}$ = 7.7 Hz, 1H, H1'), 6.08 (d, $J_{H5\text{-}H6}$ = 7.4 Hz, 1H, H5), 5.46 (d, $J_{H2'\text{-}H1'}$ = 7.7 Hz, 1H, H2'), 4.74 (app. s, 1H, H3'), 3.69 – 3.32 (m, 3H, H5'a, H5'b, H4'); ^{1}H NMR (400 MHz, MeOD) δ 7.79 (d, $J_{H6\text{-}H5}$ = 7.4 Hz, 1H, H6), 6.24 (d, $J_{H1'\text{-}H2'}$ = 7.6 Hz, 1H, H1'), 6.08 (d, $J_{H5\text{-}H6}$ = 7.4 Hz, 1H, H5), 5.46 (dd, $J_{H2'\text{-}H1'}$ = 7.5 Hz, $J_{H2'\text{-}H3'}$ = 0.8 Hz, 1H, H2'), 4.81 – 4.74 (m, 1H, H3'), 3.58 – 3.50 (m, 2H, H4' and H5'a), 3.43 (dd, $J_{H5'b\text{-}H5'a}$ = 9.5 Hz, $J_{H5'b\text{-}H4'}$ = 4.2 Hz, 1H, H5'b); ^{13}C NMR (101 MHz, D_2O) δ 175.4 (C-O, C2), 160.1 (C=O, C4), 138.9 (CH alkene, C6), 109.0 (CH alkene, C5), 92.1 (CH, C2'), 80.5 (CH, C3'), 70.3 (CH, C1'), 62.3 (CH_2, C5'), 59.1 (CH, C4'); ESI HRMS *m/z* found: $(M+H)^+$ 243.0442 $C_9H_{10}N_2O_4S$, requires $(M+H)^+$ 243.0434. NMR data was consistent with literature values (in MeOD).[9]

1'-β-(4'-thio-D-arabinofuranosyl)uracil 18

A suspension of **17** (0.500 g, 2.06 mmol, 1.0 equiv.) and KOH (116 mg, 2.06 mmol 1.0 equiv.) in a 9/1 (*v*/*v*) solution of EtOH/H_2O (10 mL) was stirred vigorously at rt for 18 h. Amberlyst 15(H^+) ion exchange resin was added, the solution stirred for 10 minutes and filtered, washing with H_2O (15 mL) and freeze dried to afford **18** as an orange solid (482 mg, 1.85 mmol, 90%). R_f 0.36 (1/4 MeOH/CH_2Cl_2); $[\alpha]_D^{25.7}$ +46.6 (*c* 0.7, H_2O); ^{1}H NMR (400 MHz, D_2O) δ 8.28 (d, $J_{H6\text{-}H5}$ = 8.1 Hz, 1H, H6), 6.10 (d, $J_{H1'\text{-}H2'}$ = 6.5 Hz, 1H, H1'), 5.84 (d, $J_{H5\text{-}H6}$ = 8.1 Hz, 1H, H5), 4.29 (dd, $J_{H2'\text{-}H3'}$ = 9.1 Hz, $J_{H2'\text{-}H1'}$ = 6.5 Hz, 1H, H2'), 4.03 – 3.81

(m, 3H, H5’a, H5’b, H3’), 3.29 (ddd, $J_{H4'-H3'}$ = 8.7 Hz, $J_{H4'-H5'a}$ 5.2 Hz, $J_{H4'-H5'b}$ = 3.8 Hz, 1H, H4’); ^{13}C NMR (101 MHz, D_2O) δ 166.1 (C=O, C4), 152.4 (C=O, C2), 144.2 (CH alkene, C6), 101.3 (CH alkene, C5), 77.1 (CH, C2’), 74.1 (CH, C3’), 60.7 (CH_2, C5’), 59.3 (CH, C1’), 48.8 (CH, C4’); ESI HRMS *m/z* found: $(M+H)^+$ 261.0545 $C_9H_{12}N_2O_5S$, requires $(M+H)^+$ 261.0540. NMR data was consistent with literature values.[8]

2’,3’,5’-tri-*O*-acetyl-1’-β-(4’-thio-D-arabinofuranosyl)uracil

Ac_2O (3.5 mL, 11.1 mmol, 6.0 equiv.) was added to a solution of **18** (481 mg, 1.84 mmol, 1.0 equiv.) in pyridine (12 mL), and the solution stirred vigorously at rt for 22 h. The solution was poured on 1M aqueous HCl solution (70 mL) and diluted with EtOAc (70 mL). The organic layer was separated and washed with saturated aqueous $NaHCO_3$ (3 x 70 mL) and brine (70 mL), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the crude as an orange foam which was purified on silica gel *via* flash chromatography (50 – 65% EtOAc/hexane) to obtain the title compound as a white foam (0.527 g, 1.36 mmol, 74%). R_f 0.62 (EtOAc); $[\alpha]_D^{25.8}$ +41.2 (*c* 0.7, MeCN); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.99 (s, 1H, N-H), 7.96 (d, J_{H6-H5} = 8.2 Hz, 1H, H6), 6.52 (d, $J_{H1'-H2'}$ = 5.4 Hz, 1H, H1'), 5.78 (dd, J_{H5-H6} = 8.2 Hz, J_{H5-NH} = 2.1 Hz, 1H, H5), 5.58 – 5.51 (m, 1H, H2’), 5.37 (dd, $J_{H3-H2'}$ = 4.8 Hz, $J_{H3'-H4'}$ = 4.0, 1H, H3’), 4.40 (dd, $J_{H5'a-H5'b}$ = 11.6 Hz, $J_{H5'a-H4'}$ = 7.1 Hz, 1H, H5’a), 4.35 (dd, $J_{H5'b-H5'a}$ = 11.6 Hz, $J_{H5'b-H4'}$ = 6.6 Hz, 1H, H5’b), 3.65 (app. td, $J_{H4'-H5'a/b}$ = 6.9 Hz, $J_{H4'-H3'}$ = 4.0 Hz, 1H, H4’), 2.14 (s, 3H, Ac-CH_3), 2.13 (s, 3H, Ac-CH_3), 2.05 (s, 3H, Ac-CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.4 (C=O, Ac), 169.4 (C=O, Ac), 168.6 (C=O, Ac), 162.6 (C=O, C4), 150.6 (C=O, C2), 141.5 (CH alkene, C6), 102.1 (CH alkene, C5), 75.6 (CH, C3’), 75.6 (CH, C2’), 63.7 (CH_2, C5’), 60.5 (CH, C1’), 48.5 (CH, C4’), 20.8 (Ac-CH_3), 20.7 (Ac-CH_3), 20.6 (Ac-CH_3); ESI HRMS *m/z* found: $(M+H)^+$ 387.0873 $C_{15}H_{18}N_2O_8S$, requires $(M+H)^+$ 387.0857.

2’,3’,5’-*O*-tri-acetyl-4-*C*-(1,2,4-triazole)-1’-β-(4’-thio-D-arabinofuranosyl)uracil

A suspension of 2’,3’,5’-tri-*O*-acetyl-1’-β-(4’-thio-D-arabinofuranosyl)uracil (360 mg, 0.299 mmol, 1.0 equiv.) in MeCN (9.3 mL) was cooled to 0 °C. Et_3N (3.0 mL, 21.4 mmol, 23 equiv.), 1,2,4-triazole (1.44 g, 20.9 mmol, 23 equiv.) and $POCl_3$ (0.21 mL, 2.27 mmol, 2.4 equiv.) were added and the solution stirred for a further 5 minutes at 0 °C before warming to rt with vigorous stirring. After 3 h, the solution was poured onto ice-cold saturated aqueous $NaHCO_3$ solution (100 mL) and diluted with EtOAc (100 mL). The organic layer was separated, washed with saturated aqueous $NaHCO_3$ solution (2 x 75 mL) and brine (75 mL), dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo* to obtain the crude title compound as a yellow foam (376 mg, 0.299 mmol, *quant.*), which was used in the next step without further purification. R_f 0.31 (1/1 EtOAc/petroleum ether); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.27 (s, 1H, triazole CH), 8.71 (d, J_{H6-H5} = 7.4 Hz, 1H, H6), 8.14 (s, 1H, triazole-CH), 7.11 (d, J_{H5-H6} = 7.3 Hz, 1H, H5), 6.73 (d, $J_{H1'-H2'}$ = 5.4 Hz, 1H, H1'), 5.70 (app. t, $J_{H2'-H1'/H3'}$ = 5.4 Hz, 1H, H2'), 5.45 –

5.32 (m, 1H, H3'), 4.49 – 4.33 (m, 2H, H5'a and H5'b), 3.80 – 3.61 (m, 1H, H4'), 2.16 (s, 3H, Ac-CH_3), 2.13 (s, 3H, Ac-CH_3), 2.01 (s, 3H, Ac-CH_3); ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.5 (uracil C-N, C4), 159.3 (uracil C=O, C2), 154.2 (triazole-CH), 148.7 (CH alkene, C6), 143.5 (triazole-CH), 94.7, (CH alkene, C5) 75.1 (CH, C3'), 75.1 (CH, C2') 63.5 (CH_2, C5'), 62.3 (CH, C1'), 48.3 (CH, C4'), 20.80 (Ac-CH_3), 20.76 (Ac-CH_3), 20.6 (Ac-CH_3); ESI HRMS *m/z* found $(M+H)^+$ 438.1099 $C_{17}H_{19}N_5O_7S$ requires $(M+H)^+$ 438.1078.

1’-β-(4’-thio-D-arabinofuranosyl)cytosine 19

A solution of 2’,3’,5’-*O*-tri-acetyl-4-*C*-(1,2,4-triazole)-1’-β-(4’-thio-D-arabinofuranosyl)uracil (460 mg, 1.05 mmol, 1.0 equiv.) in neat 7M NH_3 inMeOH solution (3.5 mL, 24.5 mmol, 23 equiv.) was heated to 120 °C in a sealed tube for 24 h. The solvents were removed *in vacuo* and the crude purified on octadecyl modified silica gel *via* automated flash chromatography (0/100, 10/90, 100/0 MeCN/H_2O) to obtain **19** as a white solid (190 mg, 0.733 mmol, 70%). R_f 0.09 (1/4 MeOH/EtOAc); ^1H NMR (400 MHz, D_2O) δ 8.25 (d, $J_{H6\text{-}H5}$ = 7.6 Hz, 1H, H6), 6.26 (d, $J_{H1'\text{-}H2'}$ = 6.5 Hz, 1H, H1’), 6.06 (d, $J_{H5\text{-}H6}$ = 7.5 Hz, 1H, H5), 4.35 (app. dd, $J_{H2'\text{-}H3'}$ = 8.7 Hz, $J_{H2'\text{-}H1'}$ = 6.5 Hz, 1H, H2’), 4.13 – 3.95 (m, 2H, H3’ and H5’a), 3.90 (dd, $J_{H5'b\text{-}H5'a}$ = 12.1 Hz, $J_{H5'b\text{-}H4'}$ = 5.6 Hz, 1H, H5’b), 3.44 – 3.23 (m, 1H, H4’); ^{13}C NMR (101 MHz, D_2O) δ 165.8 (C-NH_2, C4), 158.5 (C=O, C2), 144.1 (CH alkene, C6), 95.6 (CH alkene, C5), 77.2 (CH, C2’), 74.6 (CH, C3’), 61.1 (CH_2, C5’), 59.8 (CH, C1’), 49.1 (CH, C4’). ESI HRMS *m/z* found: $(M+H)^+$ 260.0702 $C_9H_{13}N_3O_4S$, requires $(M+H)^+$ 260.0700. Data was consistent with literature values.[10]

1’-(4’-sulfinyl(*S*,*R*)-D-arabinofuranosyl)cytosine 20

A solution of **19** (116 mg, 0.447 mmol, 1.0 equiv.) in 2/1 (*v/v*) H_2O/MeCN solution (2.2 mL) was cooled to < 5 °C over ice. *m*-CPBA (121 mg, 0.492 mmol, 1.1 equiv.) was added and the solution stirred over ice for 10 minutes and then stirred at rt. After 20 h, The solvents were removed *in vacuo* and the crude purified on octadecyl modified silica gel *via* flash chromatography (0 – 100% MeCN/H_2O) to obtain **20** as a white solid (2.5/1 diastereoisomer ratio, 95.6 mg, 0.347 mmol, 78%. An analytically pure sample (9/1 diastereoisomer ratio) obtained *via* precipitation from a minimum of hot H_2O to obtain a quantity of **20** as a white solid (12 mg, 43.6 μmol, 10%). R_f 0.41 (1/4 H_2O/MeCN); **major diastereoisomer:** ^1H NMR (400 MHz, D_2O) δ 7.57 (d, $J_{H6\text{-}H5}$ = 7.4 Hz, 1H, H6), 5.99 (d, $J_{H5\text{-}H6}$ = 7.4 Hz, 1H, H5), 5.03 (d, $J_{H1'\text{-}H2'}$ = 8.3 Hz, 1H, H1’), 4.77 – 4.73 (ov. m, 1H, H2’), 4.25 (dd, $J_{H3'\text{-}H2'}$ = 8.1 Hz, $J_{H3'\text{-}H4\text{-}}$ = 4.6 Hz, 1H, H3’) 4.19 (ov. dd, $J_{H5'a\text{-}H5'b}$ = 12.8 Hz, $J_{H5'a\text{-}H4'}$ = 4.6 Hz, 1H, H5’a), 4.05 (dd, $J_{H5'b\text{-}H5a}$ = 11.9 Hz, $J_{H5'b\text{-}H4'}$ = 10.1 Hz, 1H, H5’b), 3.05 (app. td, $J_{H4'\text{-}H5'b}$ = 9.4, $J_{H4'\text{-}H5'a/H3'}$ = 4.6 Hz, 1H, H4’); ^{13}C NMR (101 MHz, D_2O) δ 166.8 (C-NH_2, C4), 157.8 (C=O, C2), 146.6 (CH alkene, C6), 96.2 (CH alkene, C5), 82.4 (CH, C1’), 76.9 (CH, C3’), 75.5 (CH, C2’), 75.1 (CH, C4’), 59.1 (CH_2, C5’); NSI HRMS *m/z* found: $(M+Na)^+$ 298.0469 $C_9H_{13}N_3O_5S$, requires $(M+Na)^+$ 298.0468.

S1.5. 4′-Thio and 4′-sulfinylgemcitabine

3′,5′-di-*O*-benzoyl-2′-deoxy-2′-*gem*-difluoro-1′-α,β-(4′-thio-D-ribofuranosyl)-N^4-benzoyl-cytosine 21

N^4-Benzoyl cytosine (0.711 g, 3.30 mmol, 1.4 equiv.) was suspended in pyridine (1.8 mL), the flask charged with hexamethyldisilazane (4.9 mL, 22.7 mmol, 9.9 equiv.), and the mixture refluxed for 2.5 h. The solvent was removed *in vacuo* and the flask immediately stoppered and flushed with N_2, to obtain crude silylated N^4-benzoyl cytosine as a colourless oil, which was suspended in DCE (14 mL). A solution of **12** (1.03 g, 2.36 mmol, 1.0 equiv.) in DCE (10 mL) was transferred under N_2 to the flask and the suspension cooled to 0 °C. $SnCl_4$ (0.83 mL, 7.08 mmol, 3.0 equiv.) was added dropwise and the solution stirred at 0 °C for a further 10 minutes before heating to reflux for 3 h. The reaction was cooled to rt and poured onto ice-cold saturated aqueous $NaHCO_3$ solution (200 mL) and the mixture stirred for 10 minutes, filtered through celite and the filter cake washed with CH_2Cl_2 (50 mL), acetone (50 mL), MeOH (50 mL) and H_2O (50 mL). The organic solvents were removed from the mother *liquor in* vacuo and the mixture diluted with EtOAc (50 mL), extracted with EtOAc (5 x 130 mL) and the organic phase washed with saturated aqueous $NaHCO_3$ solution (2 x 150 mL) and brine (150 mL), dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo* to furnish the crude as a brown syrup, which was purified on silica gel *via* automated flash chromatography (0 – 100% EtOAc/hexanes) to obtain **21** as a yellow syrup (1/1 α/β ratio, 376 mg, 0.636 mmol, 28%), and recovered **12** (262 mg, 0.600 mmol, 25%). The anomeric mixture of **21** was then further purified *via* fractional precipitation from boiling EtOH or boiling EtOAc to furnish **21-α** as a white solid (130 mg, 0.220 mmol, 10%), **21-β** as a white solid (88.0 mg, 149 mmol, 6%) and **21-α/β** as a yellow syrup (1.2/1 α/β ratio, 128 mg, 0.216 mmol, 10%). R_f 0.18 (1/9, acetone/toluene); **β-anomer:** $[\alpha]_D^{24.7}$ -38.9 (*c* 0.6, CH_2Cl_2); ^1H NMR (400 MHz, $CDCl_3$) δ 8.85 (s, 1H, NH), 8.31 (d, $J_{H6\text{-}H5}$ = 7.6 Hz, 1H, H6), 8.12 – 8.05 (m, 4H, Ar-H), 7.96 – 7.87 (m, 2H, Ar-H), 7.67 – 7.59 (m, 3H, Ar-H), 7.56 – 7.44 (m, 7H, Ar-H and H5), 6.99 (app. t, $J_{H1'\text{-}Fa/Fb}$ = 9.7 Hz, 1H, H1′), 5.93 – 5.81 (br m, 1H, H3′), 4.76 (dd, $J_{H5'a\text{-}H5'b}$ = 11.8 Hz, $J_{H5'a\text{-}H4'}$ = 6.6 Hz, 1H), 4.66 (dd, $J_{H5'b\text{-}H5'a}$ = 11.8 Hz, $J_{H5'b\text{-}H4'}$ = 6.1 Hz, 1H, H5′b), 3.97 (br. dt, $J_{H4'\text{-}H3'}$ = 11.6 Hz, $J_{H4'\text{-}H5'a/H5'b}$ = 5.8 Hz, 1H, H4′); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.9 (C=O, Bz), 165.9 (C=O, Bz), 164.6 (C=O, Bz), 162.5 (C-NH, C4), 146.1 (CH alkene, C6), 146.0 (C=O, C2), 134.2 (C_q, Ar-C), 133.7 (C_q, Ar-C), 133.4 (C_q, Ar-C), 132.9 (CH alkene, C5), 130.2 (CH, Ar-C), 129.8 (CH, Ar-C), 129.1 (CH, Ar-C), 128.7 (CH, Ar-C), 128.7 (CH, Ar-C), 128.0 (CH, Ar-C), 127.7 (CH, Ar-C), 123.0 (dd, $J_{C2'\text{-}F}$ = 264.6 Hz, $J_{C2'\text{-}F}$ = 260.0 Hz, C_q, C2′), 72.3 (dd, $J_{C3'\text{-}F}$ = 29.7 Hz, $J_{C3'\text{-}F}$ = 23.5 Hz, CH, C3′), 63.1 (d, $J_{C5'\text{-}F}$ = 1.2 Hz, CH_2, C5′), 59.9 (dd, $J_{C1'\text{-}F}$ = 29.7 Hz, $J_{C1'\text{-}F}$ = 22.9 Hz, CH, C1′), 44.5 (CH, C4′); ^{19}F NMR (376 MHz, $CDCl_3$) δ -114.01 – -114.64 (ov. m); **α-anomer:** $[\alpha]_D^{23.9}$ +15.9 (*c* 0.9, CH_2Cl_2); ^1H NMR (400 MHz, $CDCl_3$) δ 8.93 (s, 1H, NH), 8.32 (dd, $J_{H6\text{-}H5}$ = 7.7 Hz, $J_{H6\text{-}F}$ = 1.6 Hz, 1H, Ar-H), 8.04 – 7.95 (m, 4H, Ar-H), 7.94 – 7.82 (m, 2H, Ar-H), 7.71 – 7.31 (m, 10H, Ar-H and H5), 7.02 (app. t,

$J_{H1'-Fa/Fb}$ = 9.6 Hz, 1H, H1'), 5.94 (app. dt, $J_{H3'-Fa/Fb}$ = 12.8 Hz, $J_{H3'-H4'}$ = 6.5 Hz, 1H, H3'), 4.67 (dd, $J_{H5'a-H5'b}$ = 11.7 Hz, $J_{H5'a-H4'}$ = 6.6 Hz, 1H), 4.55 (dd, $J_{H5;b-H5'a}$ = 11.7 Hz, $J_{H5'b-H4'}$ = 6.4 Hz, 1H), 4.25 (app. q, $J_{H4'-H5'a/H5'b/H3'}$ = 6.4 Hz, 1H, H4'); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.9 (C=O, Bz), 165.9 (cytosine C-NH, C4), 164.5 (C=O, Bz), 162.6 (C=O, Bz), 146.6 (C=O cytosine, C2), 146.6 (CH alkene, C6), 134.2 (C_q, Ar-C), 133.5 (C_q, Ar-C), 133.4 (C_q, Ar-C), 132.9 (CH alkene, C5), 130.1 (CH, Ar-C), 129.8 (CH, Ar-C), 129.1 (CH, Ar-C), 129.0 (CH, Ar-C), 128.7 (CH, Ar-C), 128.5 (CH, Ar-C), 128.0 (CH, Ar-C), 127.7 (CH, Ar-C), 122.7 (app. t, J = 262.8 Hz, C_q, C2'), 73.5 (dd, $J_{C3'-F}$ = 30.9 Hz, $J_{C3'-F}$ = 19.4 Hz, CH, C3'), 63.9 (CH_2, C5'), 59.5 (dd, $J_{C1'-F}$ = 32.4 Hz, $J_{C1'-F}$ = 19.2 Hz, CH, C1'), 45.1 (d, $J_{C4'-F}$ = 1.7 Hz, CH, C4'); ^{19}F NMR (376 MHz, $CDCl_3$) δ -106.35 (app d, J_{F-F} = 239.4 Hz), -120.10 (app. dt, J_{F-F} = 239.5 Hz, $J_{F-H1/H3}$ = 10.6 Hz); NSI HRMS *m/z* found: $(M+H)^+$ 592.1344 $C_{30}H_{24}N_3O_6F_2S$, requires $(M+H)^+$ 592.1348.

2'-deoxy-2'-*gem*-difluoro-1'-α,β-(4'-thio-D-ribofuranosyl)cytosine 22

MeOH (0.50 mL) was added to a suspension of **21** (87.0 mg, 0.148 mmol, 1.0 equiv.) in neat 7M NH_3 in MeOH (0.50 mL, 3.47 mmol, 23 equiv.) until a homogenous solution was obtained. The solution was stirred at rt for 18 h, the solvents removed *in vacuo* and the crude residue purified on octadecyl modified silica gel *via* flash chromatography (0/100, 10/90, 100/0 H_2O/MeOH) to obtain crude **22** as a yellow glass and a mixture of anomers (39.8 mg, 0.142 mmol, 96%). The anomers were separated and purified *via* preparative HPLC (Table 2), retention times α-anomer = 25.3 minutes (18.1 mg, 64.8 μmol, 98% purity, 44% yield), β-anomer = 26.3 minutes (14.1 mg, 50.5 μmol, 34% yield). R_f 0.57 (1/9 H_2O/MeCN); **α-anomer:** $[\alpha]_D^{22.8}$ +5.1 (*c* 1.3, H_2O); ^{1}H NMR (400 MHz, D_2O) δ 8.00 (dd, J_{H6-H5} = 7.6 Hz, J_{H6-F} = 2.6 Hz, 1H, H6), 6.50 (dd, $J_{H1'-F}$ = 12.6 Hz, $J_{H1'-F}$ = 8.9 Hz, 1H, H1'), 5.99 (d, J_{H5-H6} = 7.5 Hz, 1H, H5), 4.31 (ddd, $J_{H3'-F}$ = 16.6 Hz, $J_{H3'-H4'}$ = 8.5 Hz, $J_{H3'-F}$ = 5.7 Hz, 1H, H3'), 3.87 (dd, $J_{H5'a-H5'b}$ = 11.7 Hz, $J_{H5'a-H4'}$ = 3.7 Hz, 1H, H5'a), 3.75 – 3.67 (ov. m, 1H, H5'b), 3.68 – 3.61 (m, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 180.4 (C-NH_2, C4), 157.5 (C=O, C2), 144.1 (d, J_{C6-F} = 3.6 Hz, CH alkene, C6), 123.5 (dd, $J_{C2'-F}$ = 261.1 Hz, $J_{C2'-F}$ = 256.6 Hz, C_q, C2'), 96.2 (CH alkene, C5), 72.2 (dd, $J_{H3'-F}$ = 27.0 Hz, $J_{H3'-F}$ = 20.5 Hz, CH, C3'), 60.9 (CH_2, C5'), 58.4 (dd, $J_{H1'-F}$ = 30.1 Hz, $J_{H1'-F}$ = 19.2 Hz, CH, H1'), 48.5 (d, $J_{C4'-F}$ = 5.1 Hz, CH, C4'); ^{19}F NMR (377 MHz, D_2O) δ -110.10 (app. br d, J_{F-F} = 231.3 Hz), -123.73 – -124.70 (app dt, J_{F-F} = 230.5 Hz, $J_{F-H3'/H1'}$ = 14.5 Hz); ^{19}F NMR {^{1}H} (377 MHz, D_2O) δ -110.10 (app. br dd, J_{F-F} = 232.1 Hz, $J_{F-H1'-H3'}$ = 5.4 Hz), -124.21 (d, J_{F-F} = 231.9 Hz, $J_{F-H1'/H3'}$ = 14.6 Hz); **β-anomer:** ^{1}H NMR (400 MHz, D_2O) δ 8.14 (d, J_{H6-H5} = 7.6 Hz, 1H, H6), 6.39 (dd, $J_{H1'-F}$ = 11.9 Hz, $J_{H1'-F}$ = 2.3 Hz, 1H, H1'), 6.00 (d, J_{H5-H6} = 7.5 Hz, 1H, H5), 4.23 (ddd, $J_{H3'-F}$ = 18.4 Hz, $J_{H3'-H4'}$ = 8.6 Hz, $J_{H3'-F}$ = 6.1 Hz, 1H, H3'), 3.89 (dd, $J_{H5'a-H5'b}$ = 12.3 Hz, $J_{H5'a-H4'}$ = 3.8 Hz, 1H, H5'a), 3.84 (dd, $J_{H5'b-H5'a}$ = 12.3 Hz, $J_{H5'b-H5'a}$ = 5.1 Hz, 1H, H5'b), 3.40 (app dt, $J_{H4'-H3'}$ = 8.7 Hz, $J_{H4'-H5}$ = 4.4 Hz, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 181.0 (C-NH_2, C4), 157.9 (C=O, C2), 142.6 (CH alkene, C6), 123.8 (dd, $J_{C2'-F}$ = 262.3 Hz, $J_{C2'-F}$ = 255.6 Hz, C_q, C2'), 96.6 (CH alkene, C5), 70.5 (dd, $J_{C3'-F}$ = 26.9 Hz, $J_{C3'-F}$ = 21.4 Hz, CH, C3'), 59.8 (CH_2, C5'), 59.7 – 59.1 (m, CH, C1'), 46.3 (d, $J_{C4'-F}$ = 6.1 Hz, CH, C4'); ^{19}F NMR (376 MHz, D_2O) δ -115.37 – -116.20 (m), -117.05 – -119.14 (m); NSI HRMS *m/z* found: $(M+H)^+$ 280.0561 $C_9H_{11}F_2N_3O_3S$, requires $(M+H)^+$ 280.0562. Data was consistent with literature values .[11]

Time (minutes)	%A (10 mM ammonium acetate)	%B (MeOH)
0.0	96	4
17.0	96	4
22.0	60	40
28.0	60	40
30.0	0	100
33.0	0	100
33.1	96	4

Table 2. Preparative HPLC linear gradient system for purification of **22**. A 250 x 21.2 mm column packed 5 µ particle size with Polaris 5 C18-A was employed to load the sample The flow rate was 15 mL/minute. A solution of of **22** in H_2O (100 mg/mL) was prepared and 150 μL injected into the prep HPLC system.

2'-deoxy-2'-*gem*-difluoro-1'-β-(4'-thio-D-ribofuranosyl)cytosine 22-β

MeOH (1.4 mL) was added to a suspension of **21-β** (275 mg, 0.407 mmol, 1.0 equiv.) in neat 7M NH_3 in MeOH (0.70 mL, 4.88 mmol, 12 equiv.) until a homogenous solution was obtained. The solution was stirred at rt for 18 h, the solvents removed *in vacuo* and the crude purified *via* preparative HPLC (Table 3), retention time = 26.3 minutes (66.5 mg, 0.256 mmol, 63%). Data was consistent with literature values .[11]

Time (minutes)	%A (10 mM ammonium acetate)	%B (MeOH)
0.0	96	4
17.0	96	4
22.0	60	40
28.0	60	40
30.0	0	100
33.0	0	100
33.1	96	4

Table 3. Preparative HPLC linear gradient system for purification of **22**. A 250 x 21.2 mm column packed 5 µ particle size with Polaris 5 C18-A was employed to load the sample The flow rate was 15

mL/minute. A solution of of **22** in H_2O (100 mg/mL) was prepared and 150 μL injected into the prep HPLC system.

2'-deoxy-2'-*gem*-difluoro-1'-β-(4'-sulfinyl-D-ribofuranosyl)cytosine 23

A solution of **22-β** (20.0 mg, 71.6 μmol, 1.0 equiv.) in 1/1 (*v*/*v*) H_2O/MeCN solution (0.72 mL) was cooled to 0 °C and *m*-CPBA (14.0 mg, 78.8 μmol, 1.1 equiv.) added. After 18 h, the solvents were removed *in vacuo* and the crude material purified *via* preparative HPLC (Table 4). Retention time = 7.7 minutes (3.2 mg, 10.8 μmol, 15%). R_f 0.41 (1/9, H_2O/MeCN); 4/1 ratio diastereoisomers; **major diastereoisomer:** ^{1}H NMR (400 MHz, D_2O) δ 7.69 (d, J_{H6-H5} = 7.5 Hz, 1H, H6), 6.05 (d, J_{H5-H6} = 7.5 Hz, 1H, H5), 5.11 (app. d, $J_{H1'-F}$ = 18.7 Hz, 1H, H1'), 4.61 (ddd, $J_{H3'-F}$ = 18.5 Hz, $J_{H3'-F}$ = 12.6 Hz, $J_{H3'-H4'}$ = 7.2 Hz, 1H, H3'), 4.25 (dd, $J_{H5'a-H5'b}$ = 12.3 Hz, $J_{H5'a-H5'b}$ = 4.3 Hz, 1H, H5'a), 4.10 (dd, $J_{H5'b-H5'a}$ = 12.3 Hz, $J_{H5'b-H4'}$ = 8.8 Hz, 1H, H5'b), 3.22 – 3.02 (m, 1H, H4'); ^{13}C NMR (101 MHz, D_2O) δ 167.0 (C-NH_2, C4), 156.9 (C=O, C2), 146.5 (CH alkene, C6), 129.3 (dd, $J_{C2'-F}$ = 220.9 Hz, $J_{C2'-F}$ = 127.9 Hz, C_q, C2'), 97.2 (CH alkene, C5), 84.6 (dd, $J_{C1'-F}$ = 61.9 Hz, $J_{C1'-F}$ = 25.0 Hz, CH, C1'), 72.9 (dd, $J_{C3'-F}$ = 12.5 Hz, $J_{C3'-F}$ = 8.4 Hz, CH, C3'), 72.8 (CH, C4'), 58.3 (CH_2, C5'); ^{19}F NMR (377 MHz, D_2O) δ -103.0 (app. dt, J_{F-F} = 241.3 Hz, $J_{F-H1'/H3'}$ = 18.7 Hz), -110.3 (app. dd, J_{F-F} = 241.3 Hz, $J_{F-H3'}$ = 7.4 Hz); NSI HRMS *m/z* found: $(M+H)^+$ 296.0516 $C_9H_{11}F_2N_3O_4S$, requires $(M+H)^+$ 296.0517.

Time (minutes)	**%A (H_2O)**	**%B (MeOH)**
0.0	96	4
5.0	96	4
15.0	0	100
18.0	0	100
18.1	96	4

Table 4. Preparative HPLC linear gradient system for purification of **23**. A 250 x 21.2 mm column packed 5 μ particle size with Polaris 5 C18-A was employed to load the sample The flow rate was 20 mL/minute. A solution of of **23** in H_2O (100 mg/mL) was prepared and 50 μL injected into the prep HPLC system.

S2. Cytotoxicity Assays

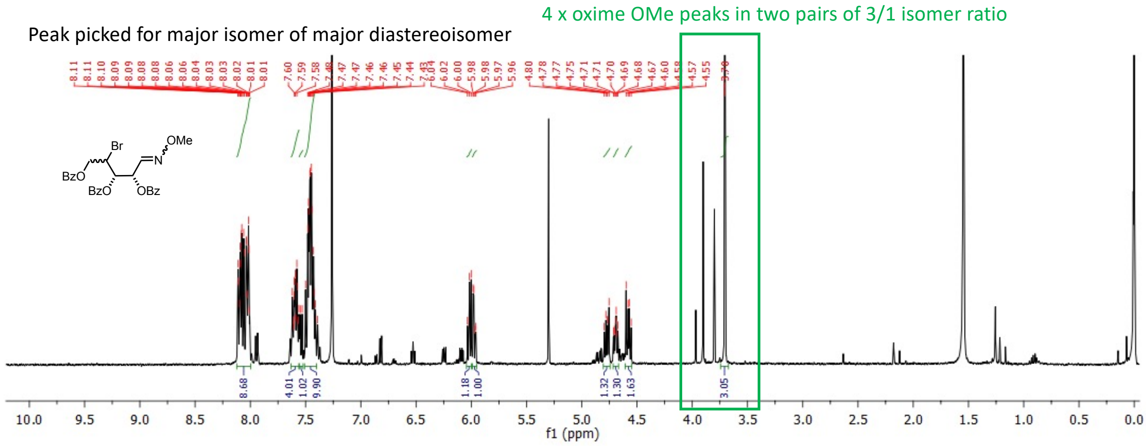
Cell culture: PANC-1 (ATCC, Catalog# CRL-1469) cells were cultured at 37 °C with 5% CO_2 in DMEM (Corning, Catalog# 10-013CV), supplemented with 10% heat-inactivated FBS (Corning, Catalog# 35016CV) and 1X Non-essential amino acids (0.1 mmol each amino acids) (Corning, Catalog# 25025CI) and 1X Penicillin-Streptomycin Solution (Penicillin (100 IU) and Streptomycin (100 µg/mL) (Corning, Catalog# 30002CI). U87-MG (ATCC, Catalog# HTB-14) cells were cultured at 37 °C with 5% CO_2 in EMEM (Lonza, Catalog# 12-611F), supplemented with 10% heat-inactivated FBS (Corning, Catalog# 35016CV) and 1X Non-essential amino acids (0.1 mmol each amino acids) (Corning, Catalog# 25025CI) and 1X Corning™ Penicillin-Streptomycin Solution (Penicillin (100 IU) and Streptomycin (100 μg/mL) (Corning, Catalog# 30002CI). Testing compound stock solution: 40 mM (first experiment) and 5 mM (second experiment), respectively, were in DMSO. Compound DMSO stock solutions were diluted for 20-fold in culture medium, followed by 8 points of 3-fold serial dilutions in medium with 5% DMSO. The tenth point contained no compounds, only medium with 5% DMSO served as DMSO control. Reference compound Gemcitabine was 1 mmol in H_2O. The top concentration for PANC-1 was 5 μmol and for U87-MG was 0.185 μmol. Cytarabine was 2 mmol in H_2O. The top concentration was 10 μmol for both cell lines. The control for these two compounds were cells treated with medium only. The cells were seeded in a density of 4000 cells/well/100 μL for PANC-1 cells and 5000 cells/well/100 ul for U87-MG cells on 96-well white plates and incubated at 37C with 5% CO2 for 24 h. On day two, 10 μL of the serial diluted compounds were added onto the plate with cells, either in duplicate or triplicate. The top concentration of the testing compound was 200 uM (first experiment) and 25 μmol (second experiment). The final DMSO concentration in the assay for all wells was 0.5%. The cells were incubated with the compounds for three days at 37 °C with 5% CO_2. Cell *via*bilities were then determined using CellTiter-Glo, 2.0 (Promega, Catalog# G9243), which quantitated the amount of ATP present, which indicated the presence of metabolically active cells. Briefly, 100 μL of the CellTiter-Glo, 2.0 reagent was added to each well and the luminescent signal was recorded for 0.5 s/well on an EnSpire plate reader. The luminescent signals from 4 wells containing only medium were used as background which was subtracted from all other testing wells. The wells treated with only 0.5% DMSO were DMSO control, was set as 100% of cell *via*bility. All the wells treated with cells will be as % of the Control. Data analysis was performed using GraphPad Prism software.

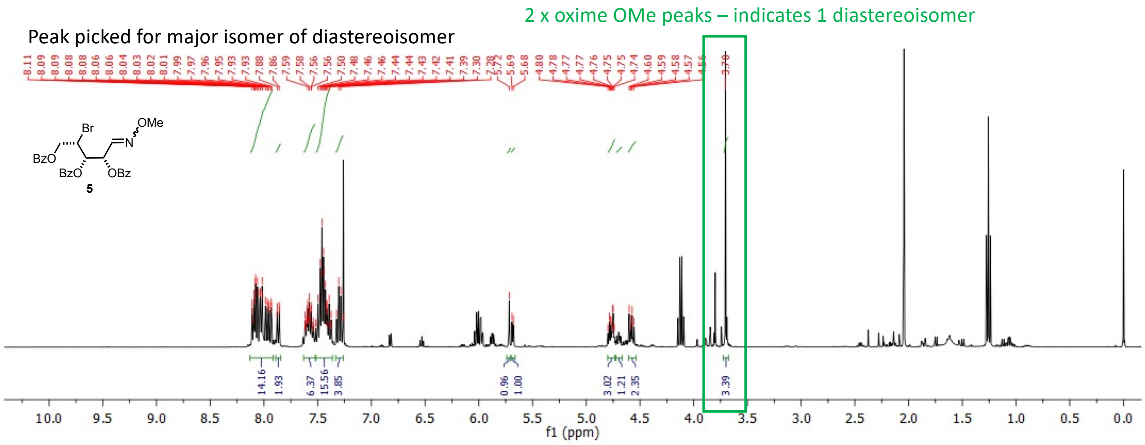
S3. X-Ray Crystallography

Diffraction data were collected on a Bruker D8 Quest ECO diffractometer using graphite-monochromated Mo Kα radiation (λ = 0.71073 Å). Crystals were mounted on Mitegen micromounts in NVH immersion oil, and all collections were carried out at 150 K using an Oxford cryostream. Data collections were carried out using ϕ and ω scans, with collections and data reductions carried out in the Bruker APEX-3 suite of programs.[12] Multi-scan absorption corrections were applied for all datasets using SADABS.[13] The data were solved with the intrinsic phasing routine in SHELXT,[14] and all data were refined on F^2 with full-matrix least squares procedures in SHELXL,[15] operating within the OLEX-2 GUI.[16] All non-hydrogen atoms were refined with anisotropic displacement parameters. Carbon-bound hydrogen atoms were placed in riding positions and refined with isotropic displacement parameters equal to 1.2 or 1.5 times the isotropic equivalent of their carrier atom. Slight positional disorder in one phenyl ring of (β-**11**) was modelled by splitting C7 and C10-C12 over two overlapping orientations with occupancies refined to approximately 0.7:0.3. EADP constraints were applied to the closely overlapping carbon atoms C7/C7A and C10/C10A, and the ring geometry and U_{ij} tensors were restrained with SADI, ISOR and/or RIGU cards where appropriate to maintain sensible geometries.

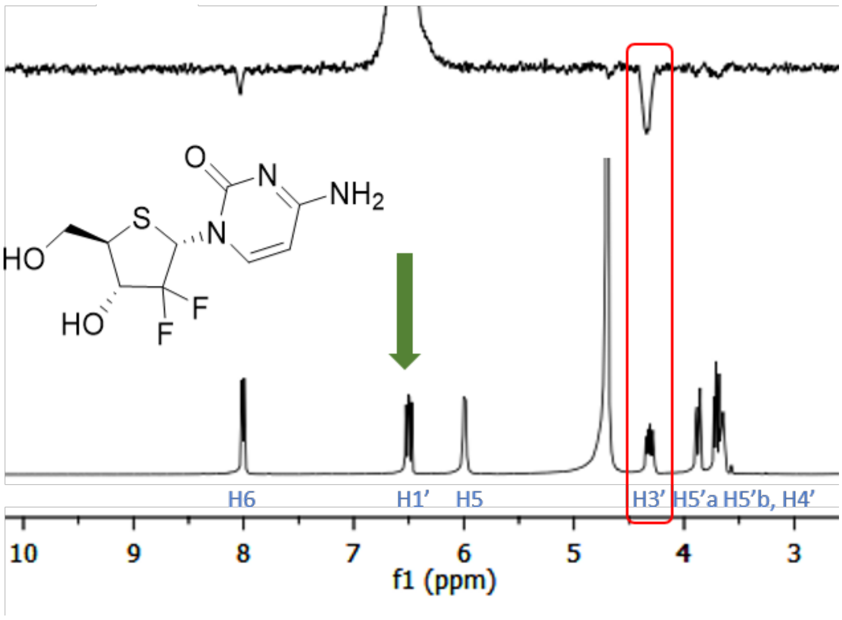
Crystal Data for (β-**11**) $C_{20}H_{18}F_2O_7S_2$ (M =472.46 g/mol): orthorhombic, space group $P2_12_12_1$ (no. 19), a = 5.4744(3) Å, b = 17.6020(11) Å, c = 21.4574(14) Å, V = 2067.6(2) Å^3, Z = 4, T = 150.0 K, μ(MoKα) = 0.316 mm^{-1}, $Dcalc$ = 1.518 g/cm^3, 24500 reflections measured (5.002° ≤ 2Θ ≤ 51.992°), 4070 unique (R_{int} = 0.0783, R_{sigma} = 0.0511) which were used in all calculations. The final R_1 was 0.0620 (I > 2σ(I)) and wR_2 was 0.1412 (all data). CCDC 2115322

S4. [1]H NMR of oxime 5 demonstrating C4 epimers and C4-diastereopure material





S5. NMR n*O*e spectrum for 22-α.



S6. References

1. T. Itoh, H. Takamura, K. Watanabe, Y. Araki and Y. Ishido, *Carbohydr. Res.*, 1986, 156, 241–246.

2. Nakamura, K.; Shimamura, A.; Imoto, J.; Motomasa, T.; Watanabe, K.; Wada, K.; Fujino, Y.; Matsumoto, T.; Takahashi, M.; Okada, H.; et al. Synthetic Intermediate of 1-(2-Deoxy-2-Fluoro-4-Thio-Beta-D-Arabinofuranosyl) Cytosine, Synthetic Intermediate of Thionucleoside and Method for Making the Same. US 2015/0152131 A1, February 12, 2015.

3. Z.-H. Sun and B. Wang, *J. Org. Chem.*, 2008, 73, 2462–2465.

4. T. S. Chou, P. C. Heath, L. E. Patterson, L. M. Poteet, R. E. Lakin and A. H. Hunt, *Synthesis*, 1992, 1992, 565–570.

5. R. Fernández, M. I. Matheu, R. Echarri and S. Castillón, *Tetrahedron*, 1998, 54, 3523–3532.

6. M. Guinan, D. Lynch, M. Smith and G. J. Miller, in *Carbohydrate Chemistry: Proven Synthetic Methods Volume 5*, CRC Press, 2020, pp. 227–232.

7. N. Nishizono, R. Baba, C. Nakamura, K. Oda and M. Machida, *Org. Biomol. Chem.*, 2003, 1, 3692–3697.

8. T. Naka, N. Minakawa, H. Abe, D. Kaga and A. Matsuda, *J. Am. Chem. Soc.*, 2000, 122, 7233–7243.

9. K. Haraguchi, H. Takahashi, N. Shiina, C. Horii, Y. Yoshimura, A. Nishikawa, E. Sasakura, K. T. Nakamura and H. Tanaka, *J. Org. Chem.*, 2002, 67, 5919–5927.

10. B. Urbas and R. L. Whistler, *The Journal of Organic Chemistry*, 1966, 31, 813–816.

11. Y. Yoshimura, K. Kitano, H. Satoh, M. Watanabe, S. Miura, S. Sakata, T. Sasaki and A. Matsuda, *J. Org. Chem.*, 1996, 61, 822–823.

12. *Bruker APEX-3*, Bruker-AXS Inc., Madison, WI, 2016.

13. *SADABS*, Bruker-AXS Inc., Madison, WI, 2016.

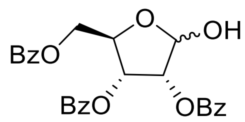
14. G. M. Sheldrick, *Acta Crystallogr., Sect. A: Found. Adv.*, 2015, 71, 3–8.

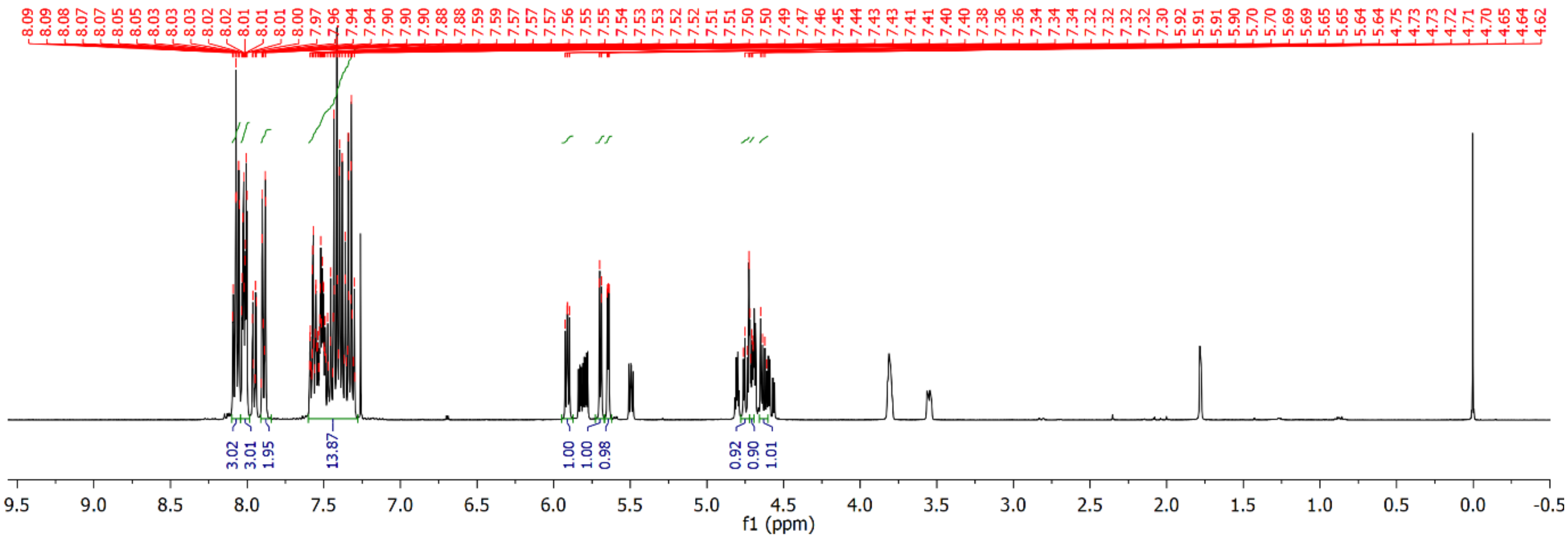
15. G. M. Sheldrick, *Acta Crystallogr., Sect. C: Struct. Chem.*, 2015, 71, 3–8.

16. O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Crystallogr.*, 2009, 42, 339–341.

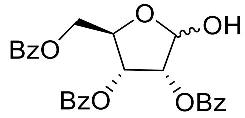
S7. NMR spectra

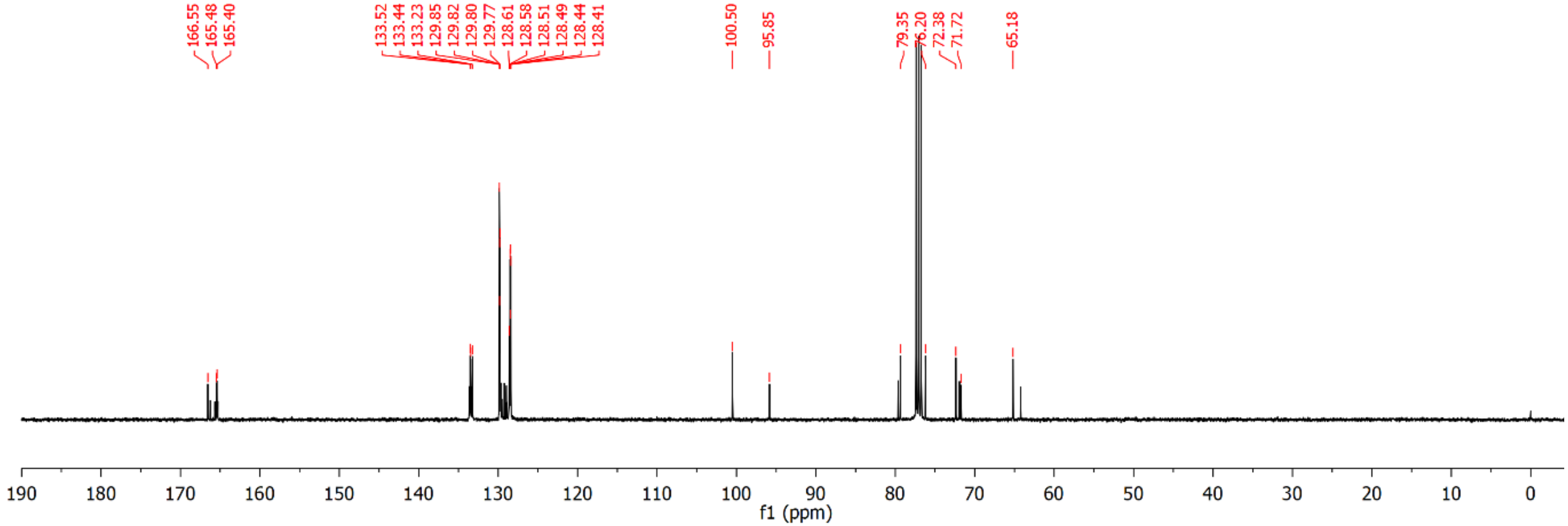
2,3,5-tri-*O*-benzoyl-1’-α,β-D-ribofuranose





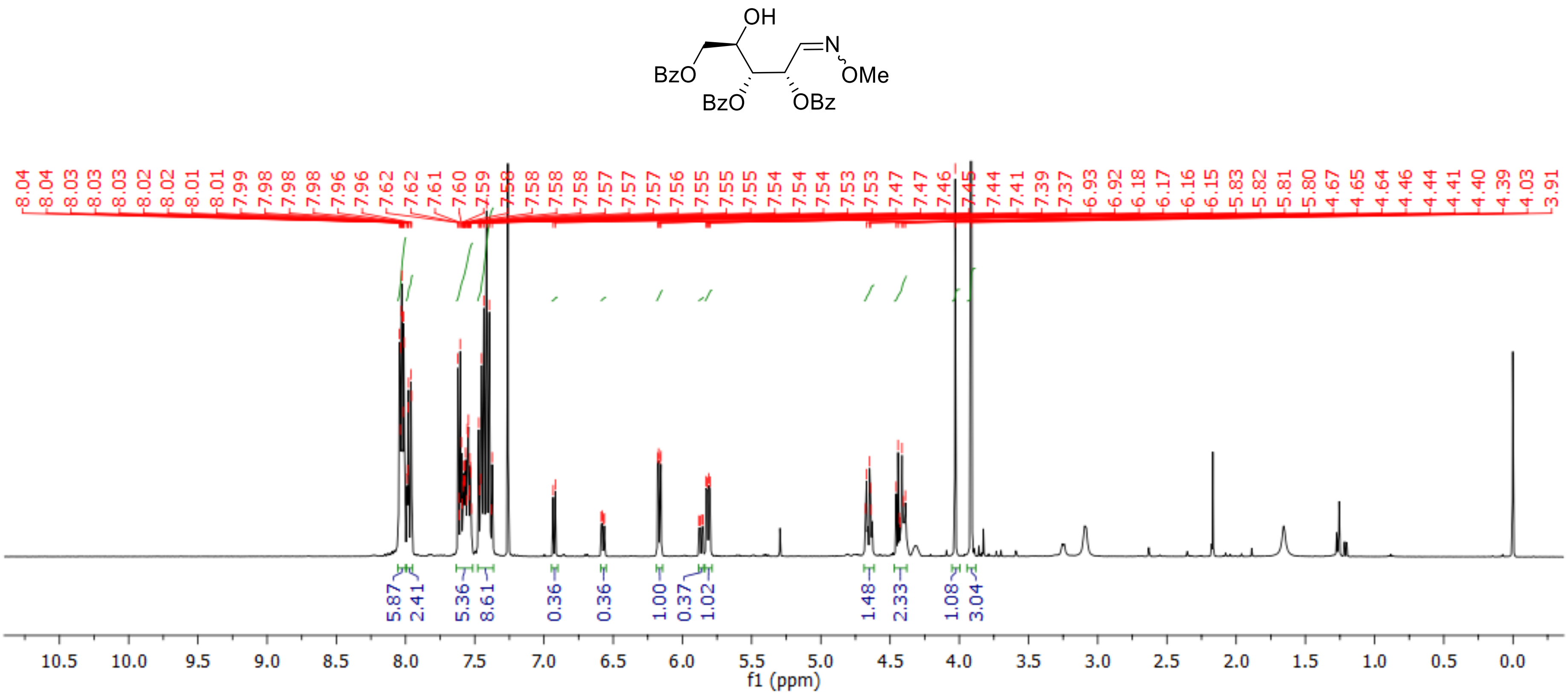
^{1}H NMR (400 MHz, $CDCl_3$)



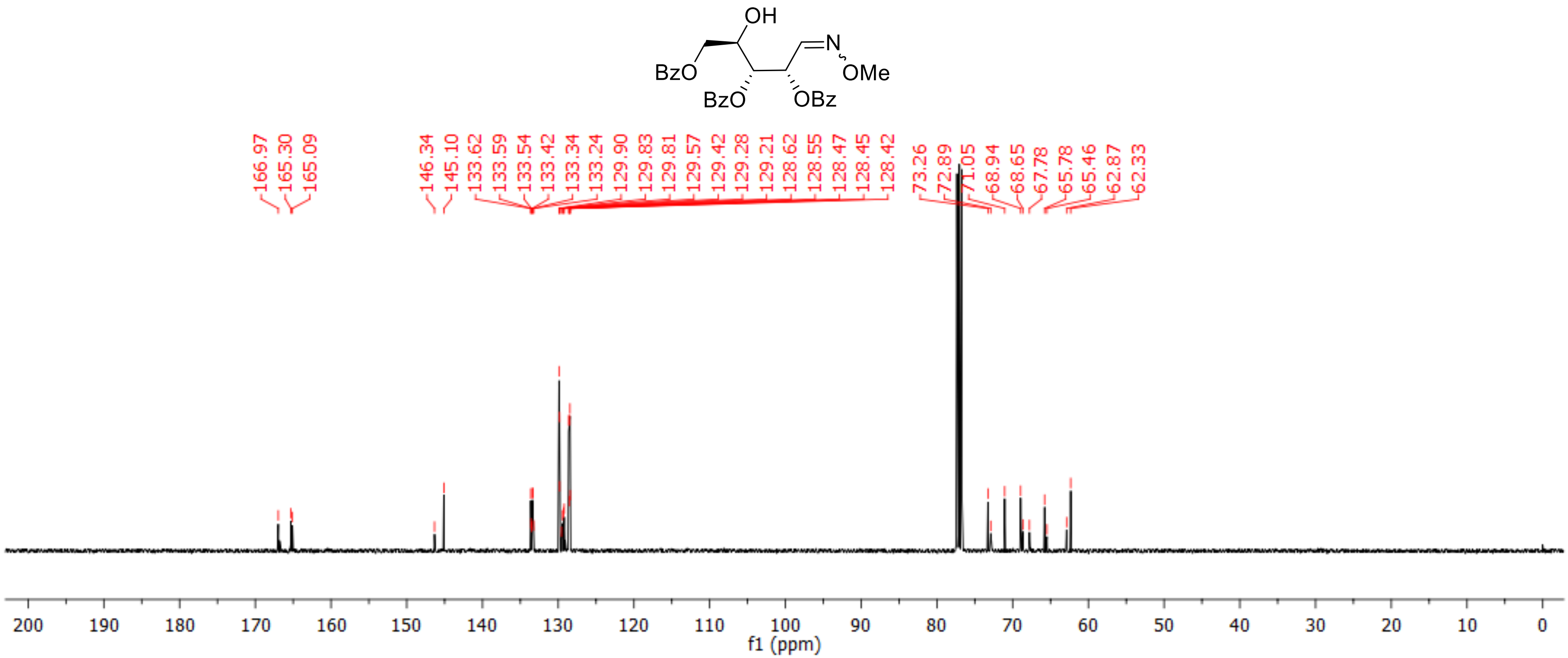


^{13}C NMR (101 MHz, $CDCl_3$)

(2*R*,3*R*,4*S*)-2,3,5-tri-*O*-benzoyl-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*) **3**

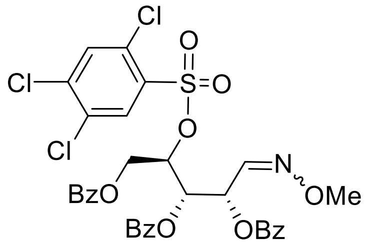


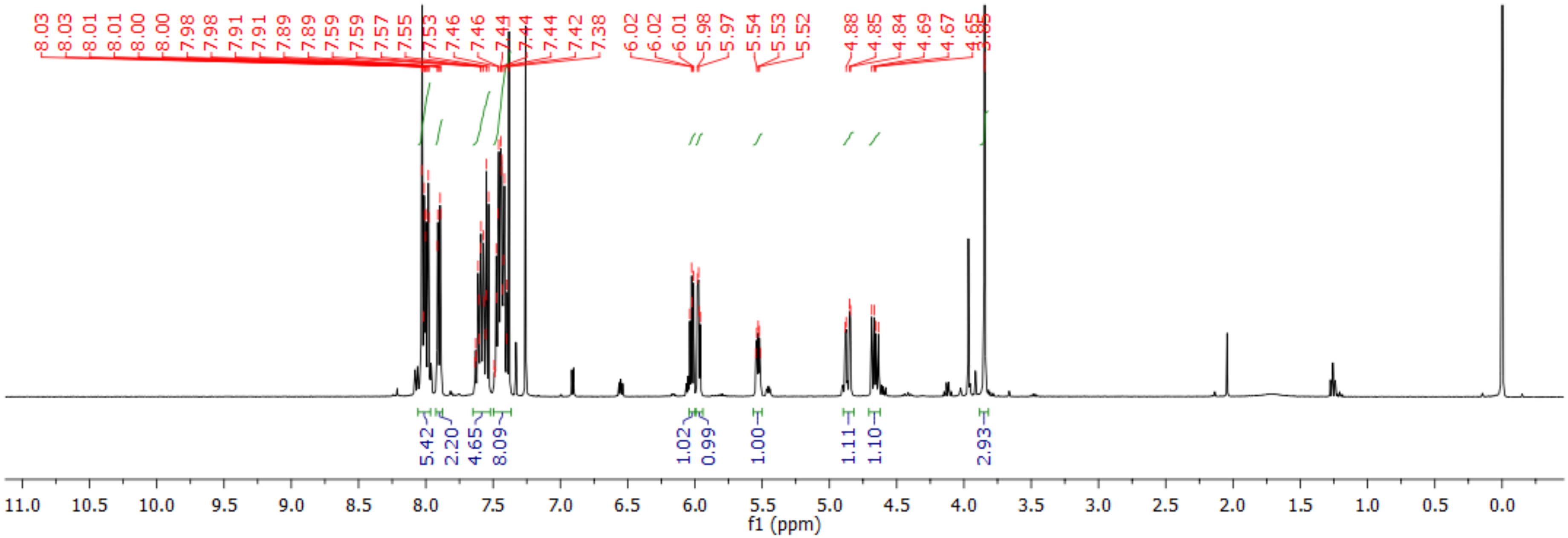
^{1}H NMR (400 MHz, $CDCl_3$)



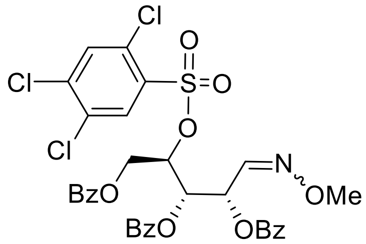
^{13}C NMR (101 MHz, $CDCl_3$)

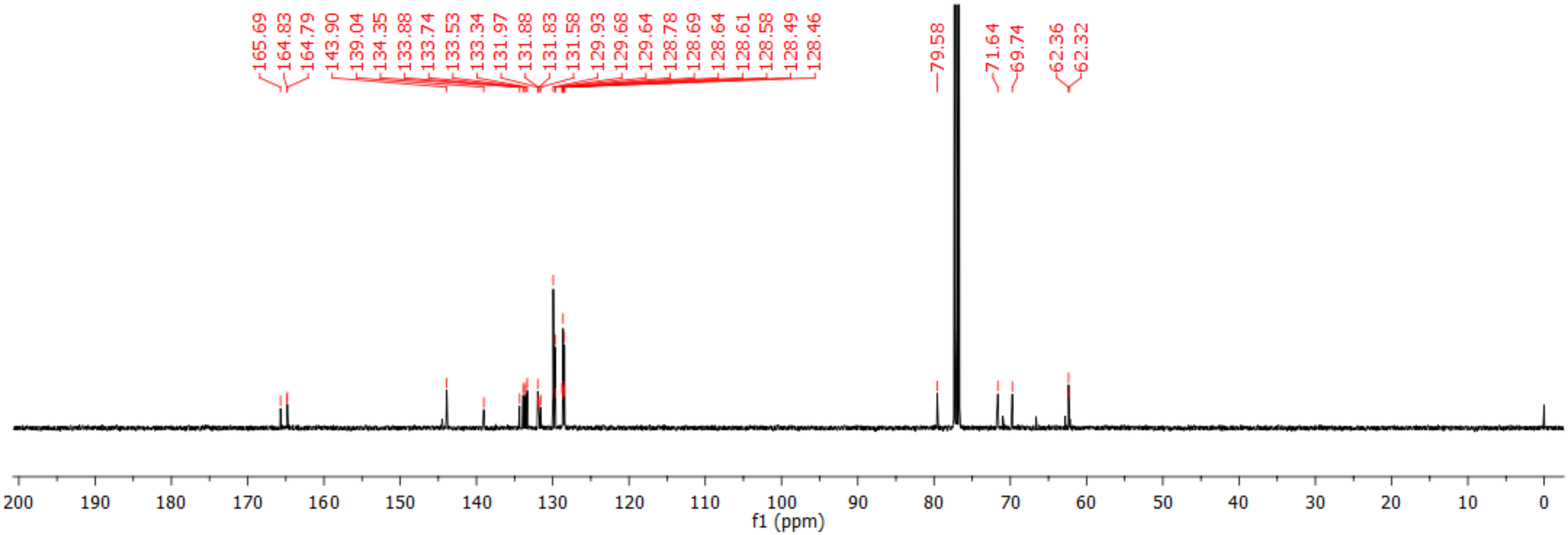
(2R,3R,4S)-2,3,5-tri-O-benzoyl-4-O-(2',4',5'-trichlorophenylsulfonyl)-1-(methoxyimino)pentane (E/Z) **4**





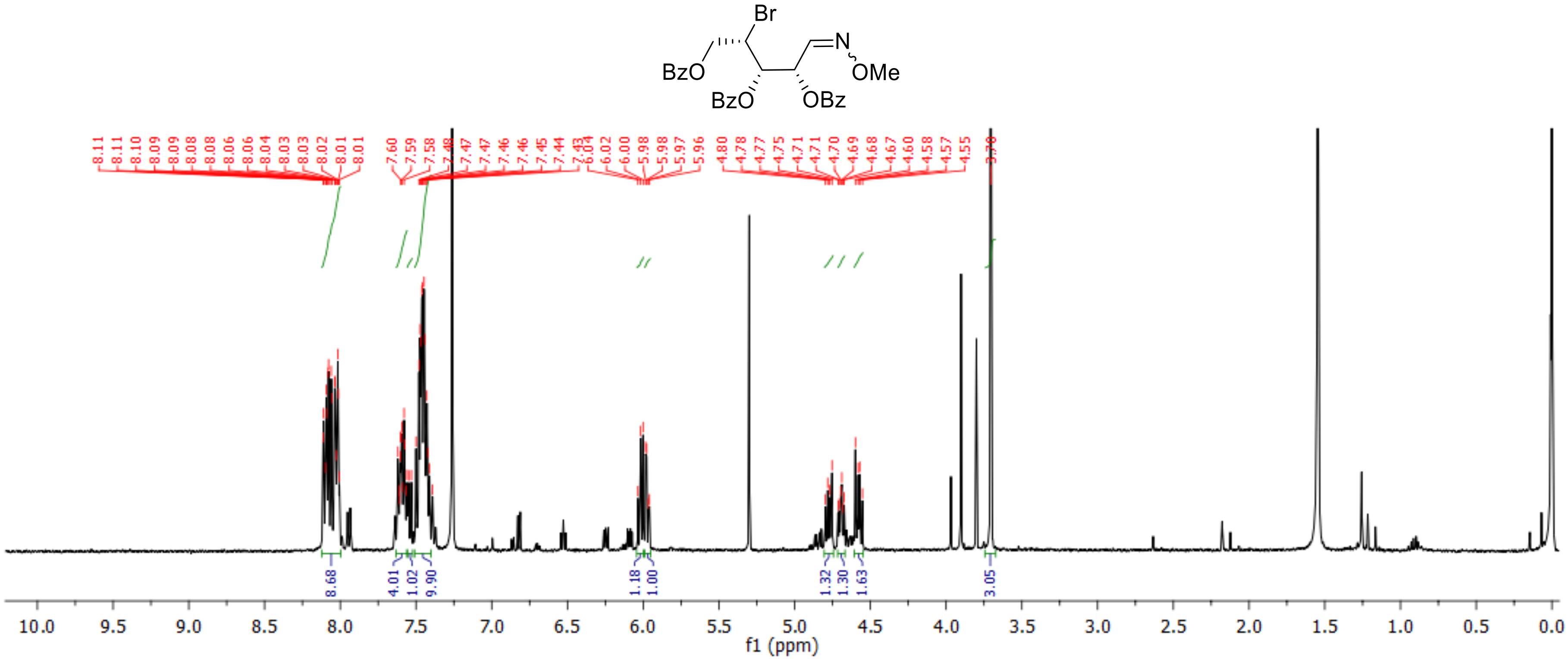
[1]H NMR (400 MHz, $CDCl_3$)



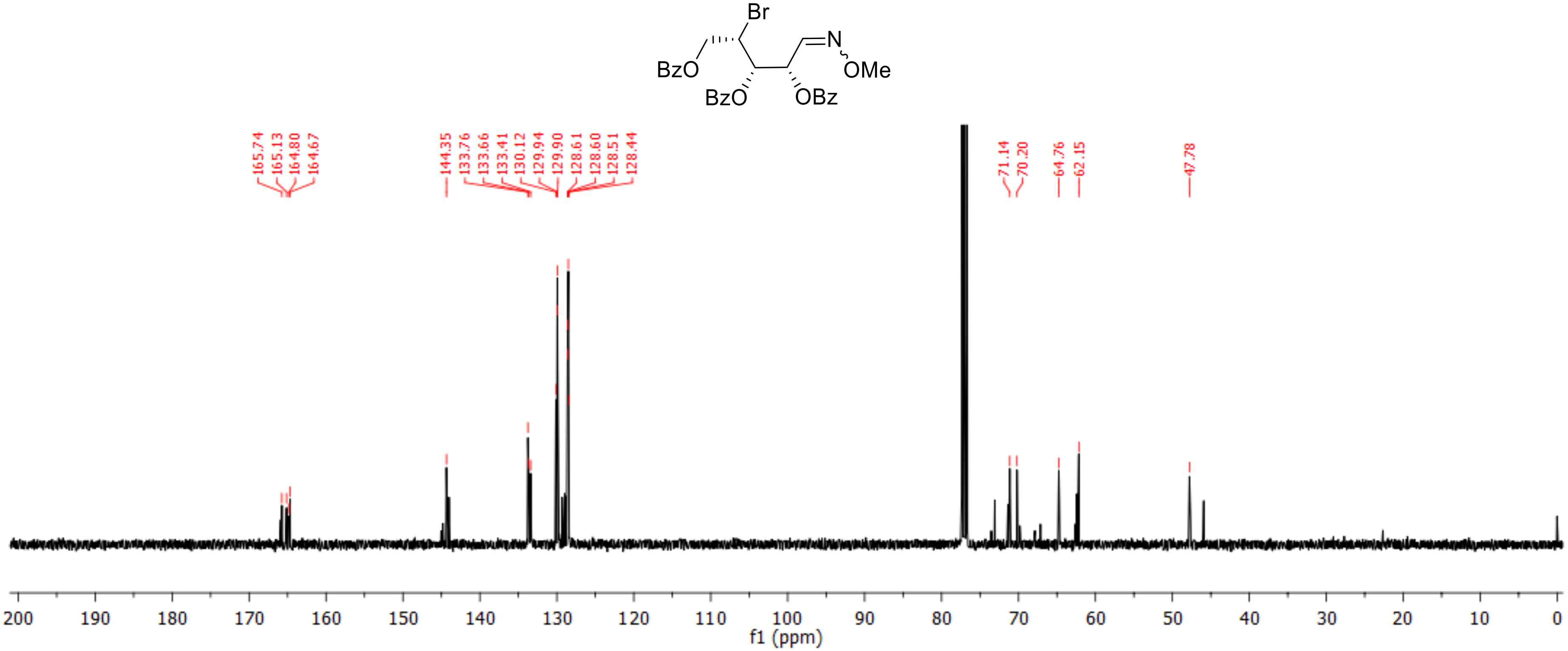


^{13}C NMR (101 MHz, $CDCl_3$)

(2*S*,3*R*,4*S*)-2,3,5-tri-*O*-benzoyl-4-bromo-1-(methoxyimino)pentane (*E*/*Z*) **5**

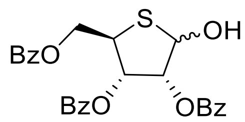


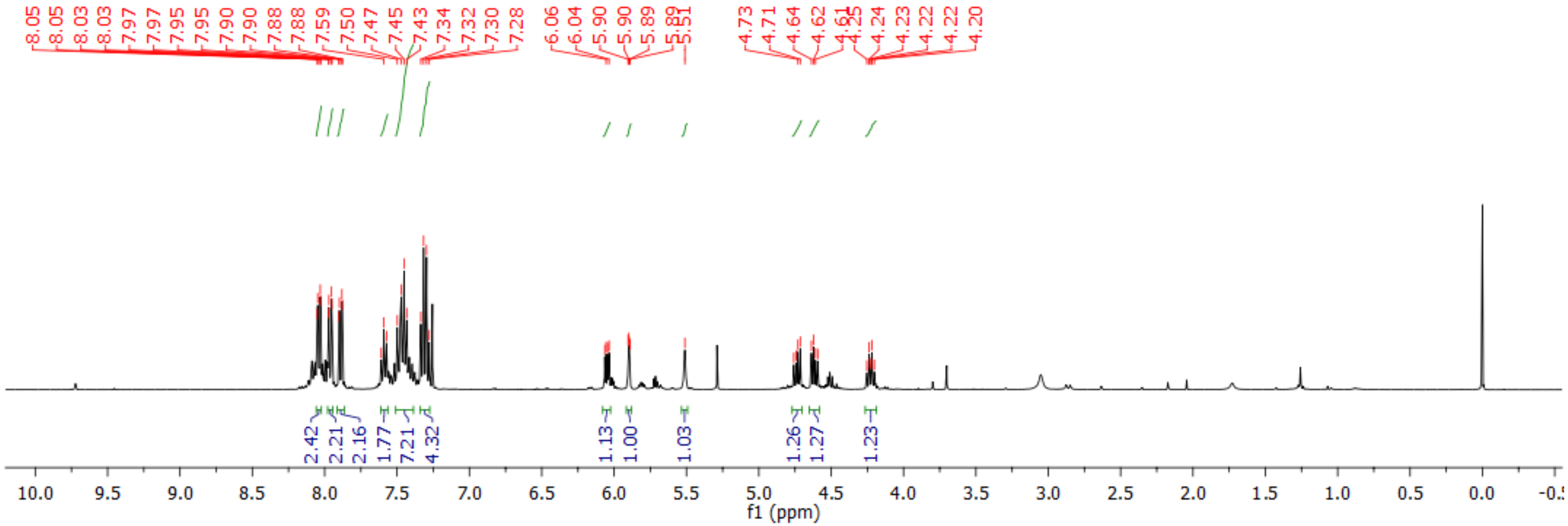
^{1}H NMR (400 MHz, $CDCl_3$)



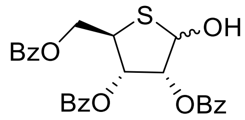
^{13}C NMR (101 MHz, $CDCl_3$)

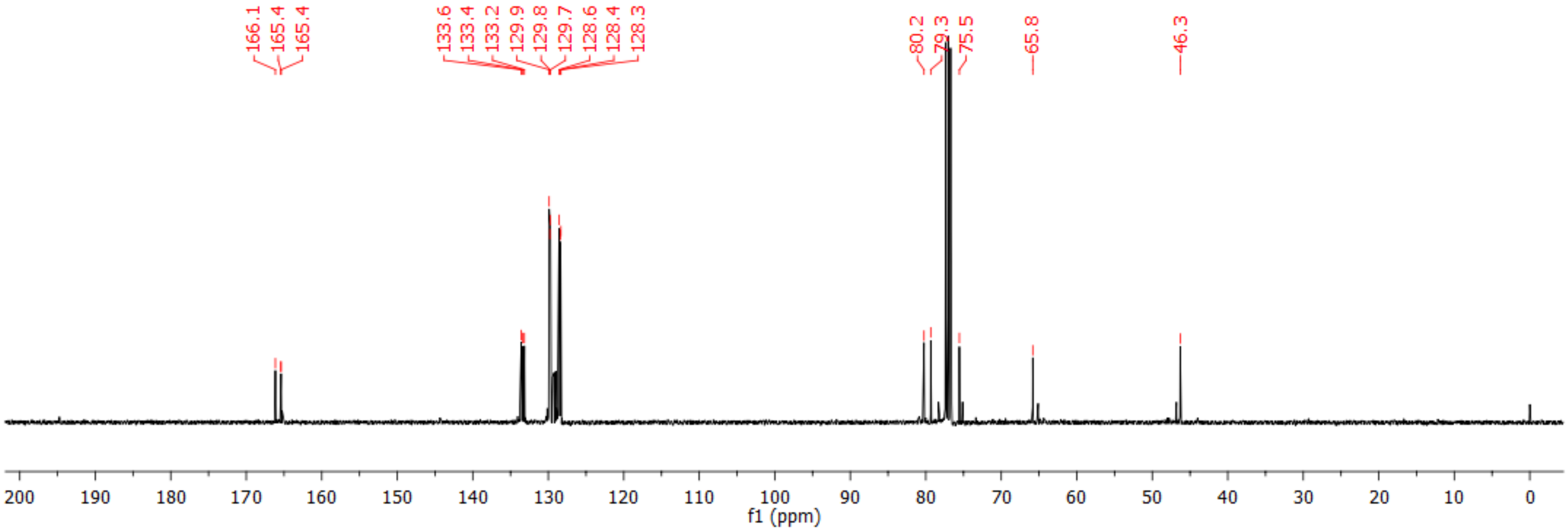
2,3,5-tri-*O*-benzoyl-1-α,β-(4-thio-D-ribofuranose) **6**





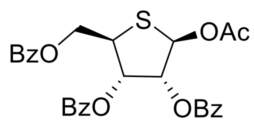
^{1}H NMR (400 MHz, $CDCl_3$)

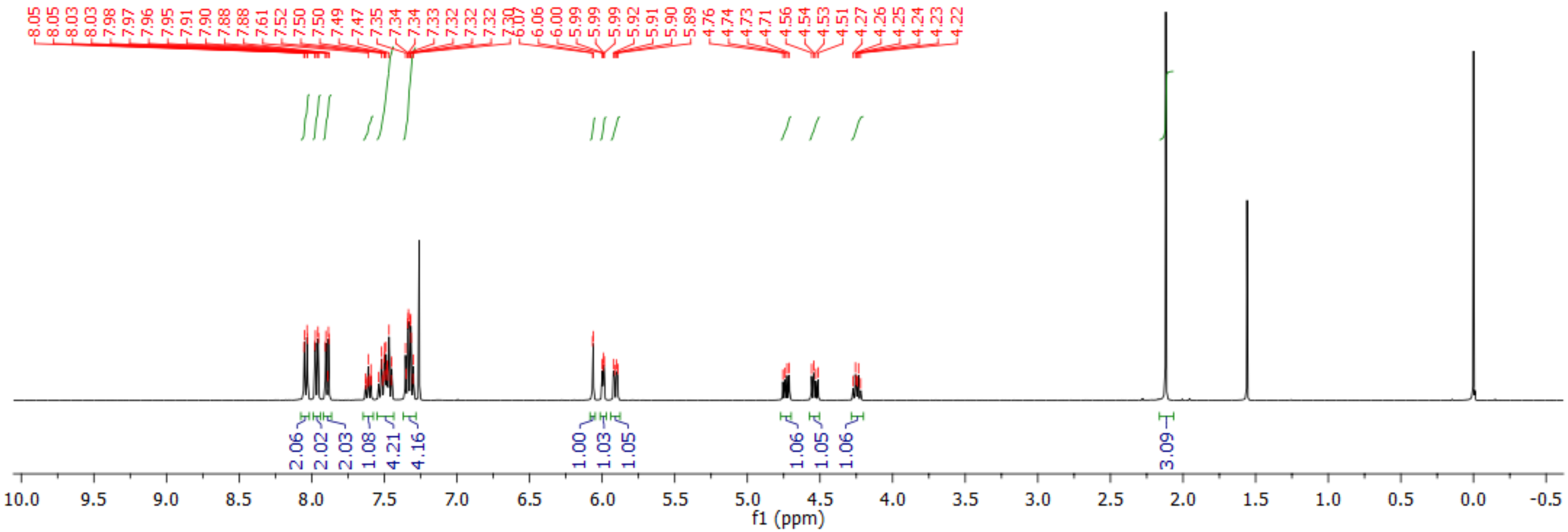




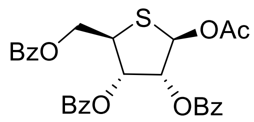
^{13}C NMR (101 MHz, $CDCl_3$)

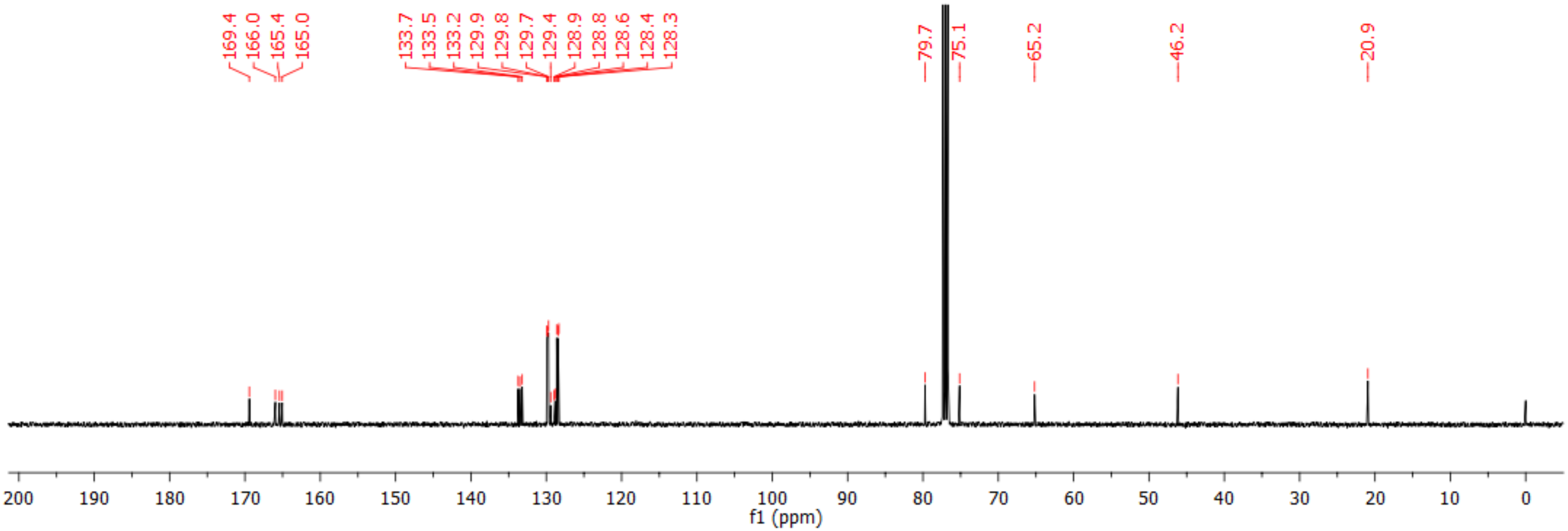
1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-1-β-(4-thio-D-ribofuranose) **1**





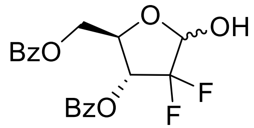
^{1}H NMR (400 MHz, $CDCl_3$)

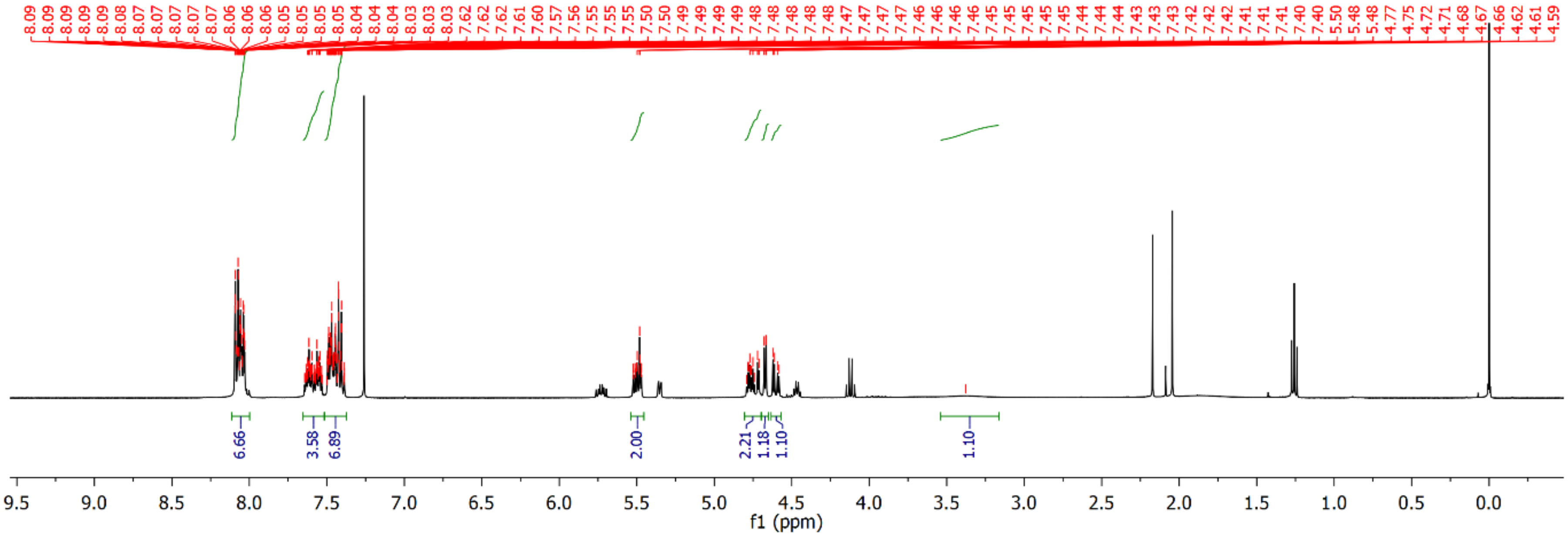




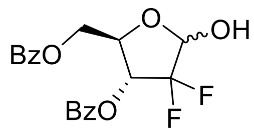
^{13}C NMR (101 MHz, $CDCl_3$)

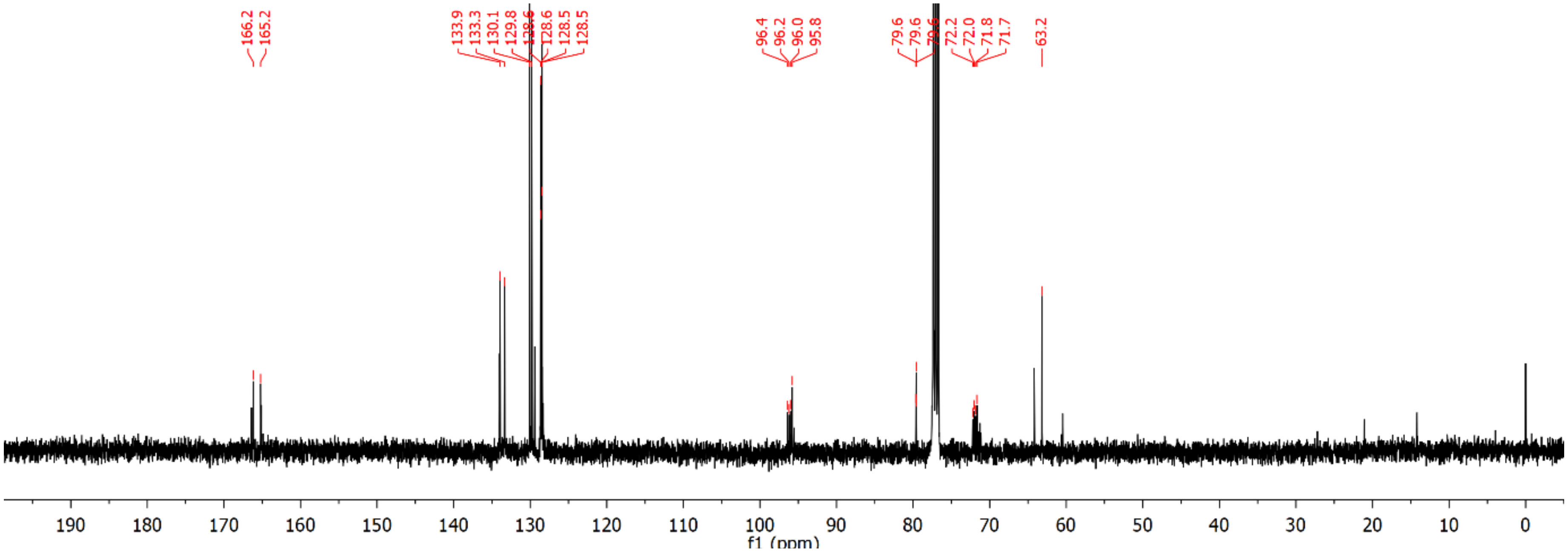
3,5-di-*O*-benzoyl-2-deoxy-2-*gem*-difluoro-1-α,β-D-ribofuranose



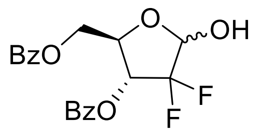


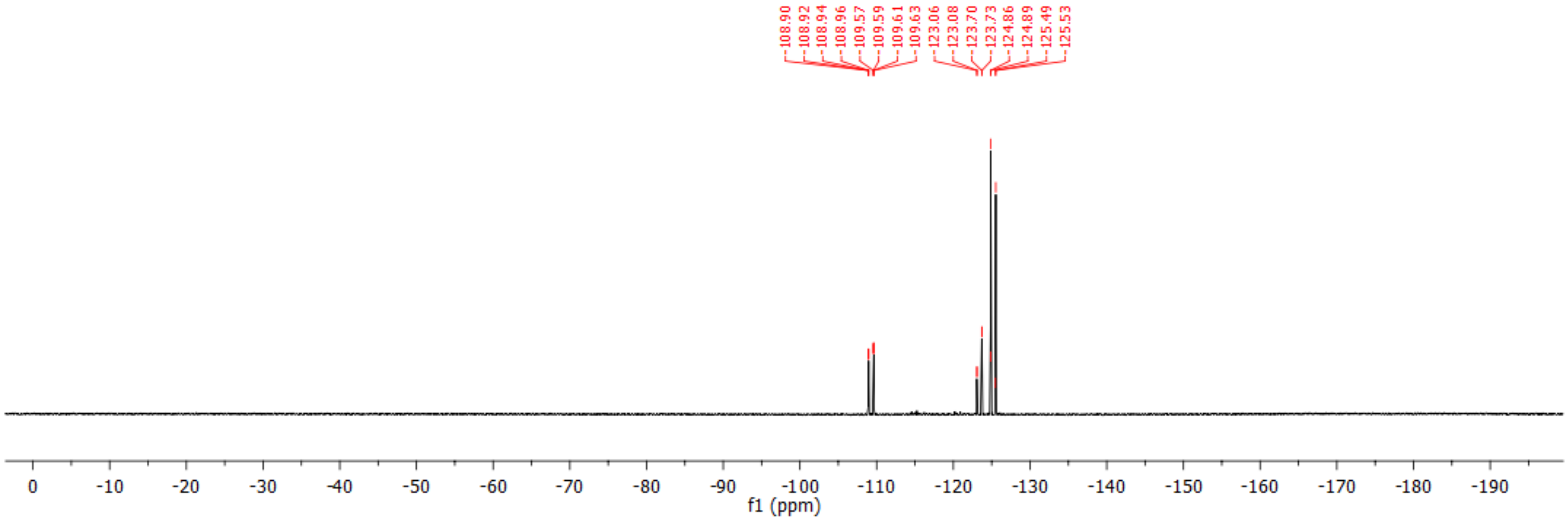
^{1}H NMR (400 MHz, $CDCl_3$)





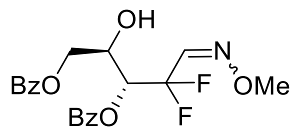
^{13}C NMR (101 MHz, $CDCl_3$)

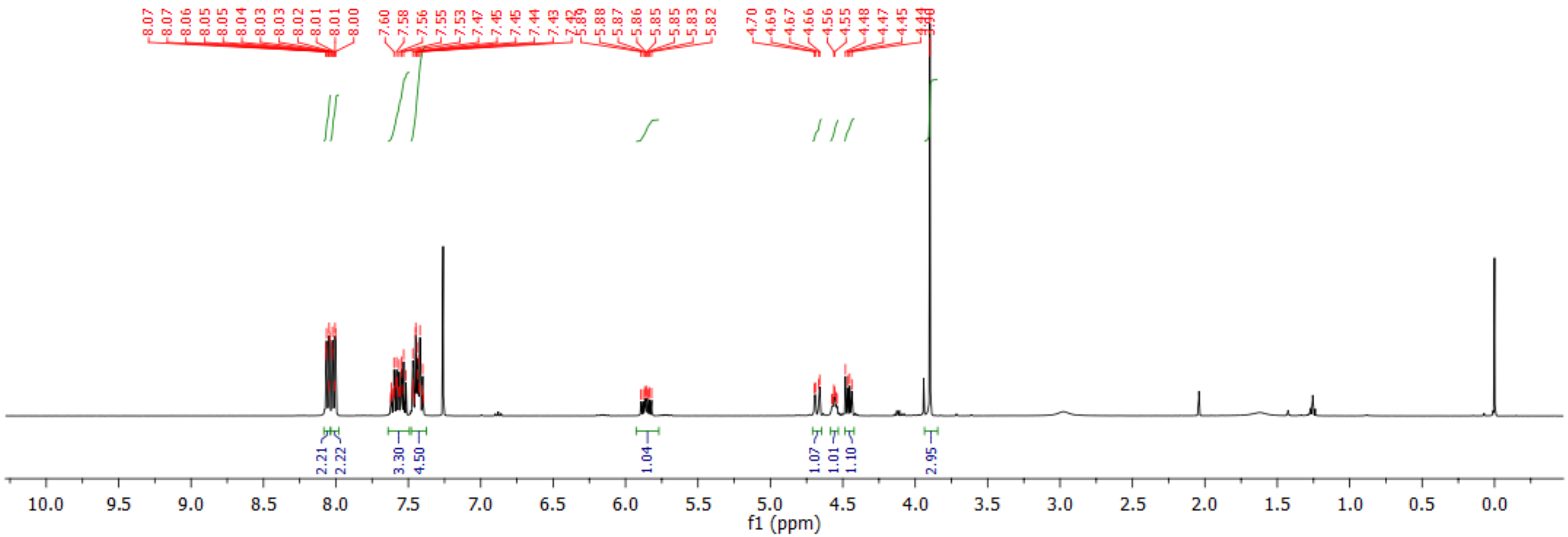




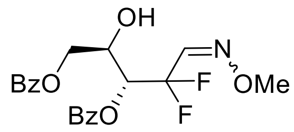
^{19}F NMR (376 MHz, $CDCl_3$)

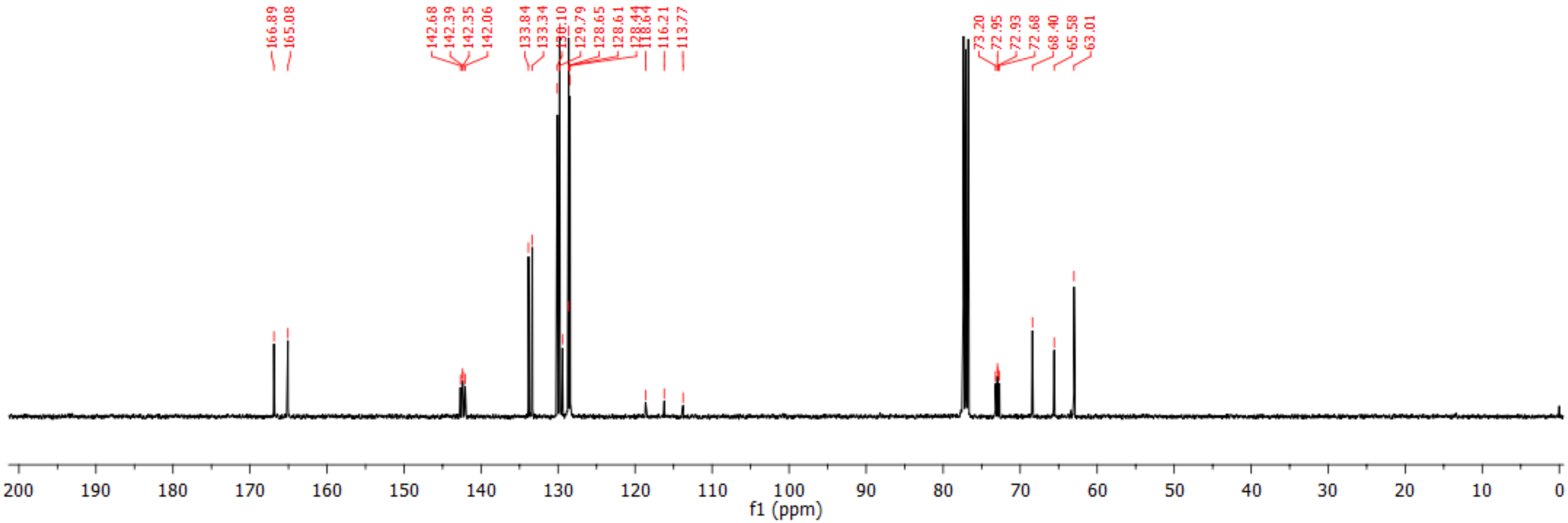
(2*R*,3*R*,4*S*)-3,5-di-*O*-benzoyl-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*) **8**



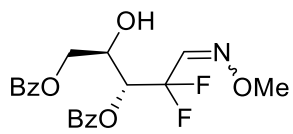


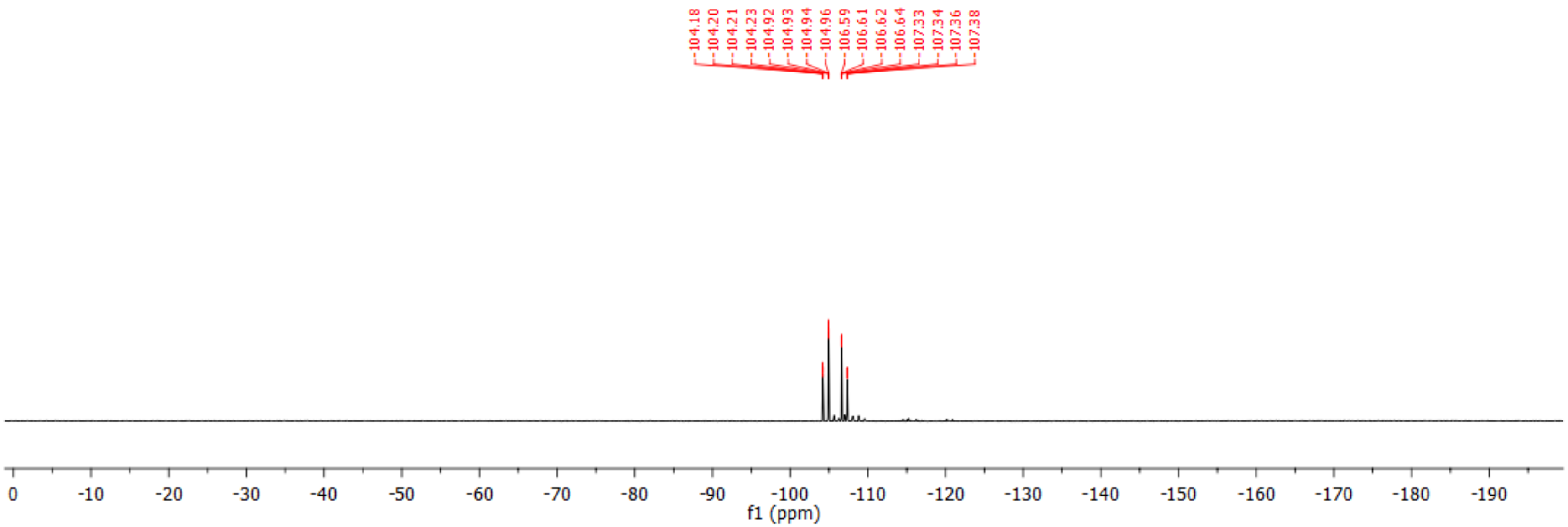
^{1}H NMR (400 MHz, $CDCl_3$)





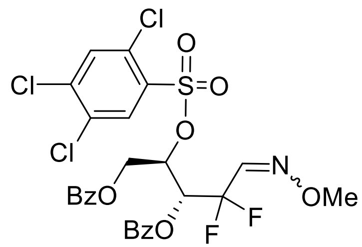
^{13}C NMR (101 MHz, $CDCl_3$)

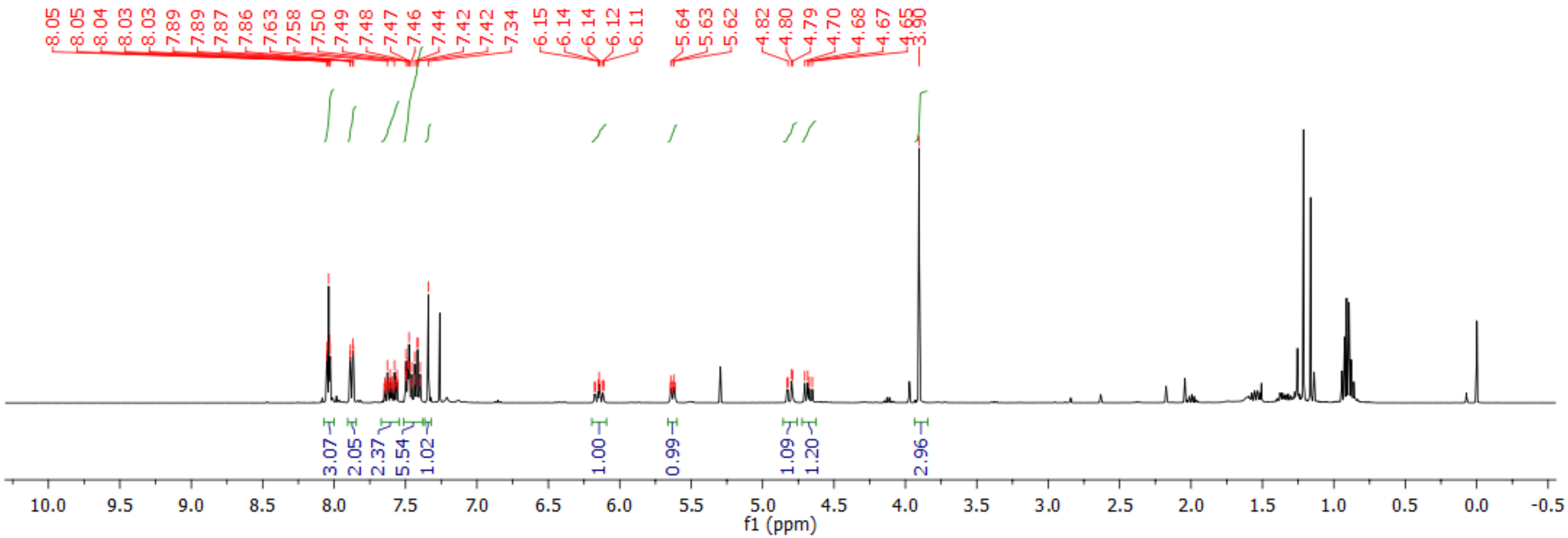




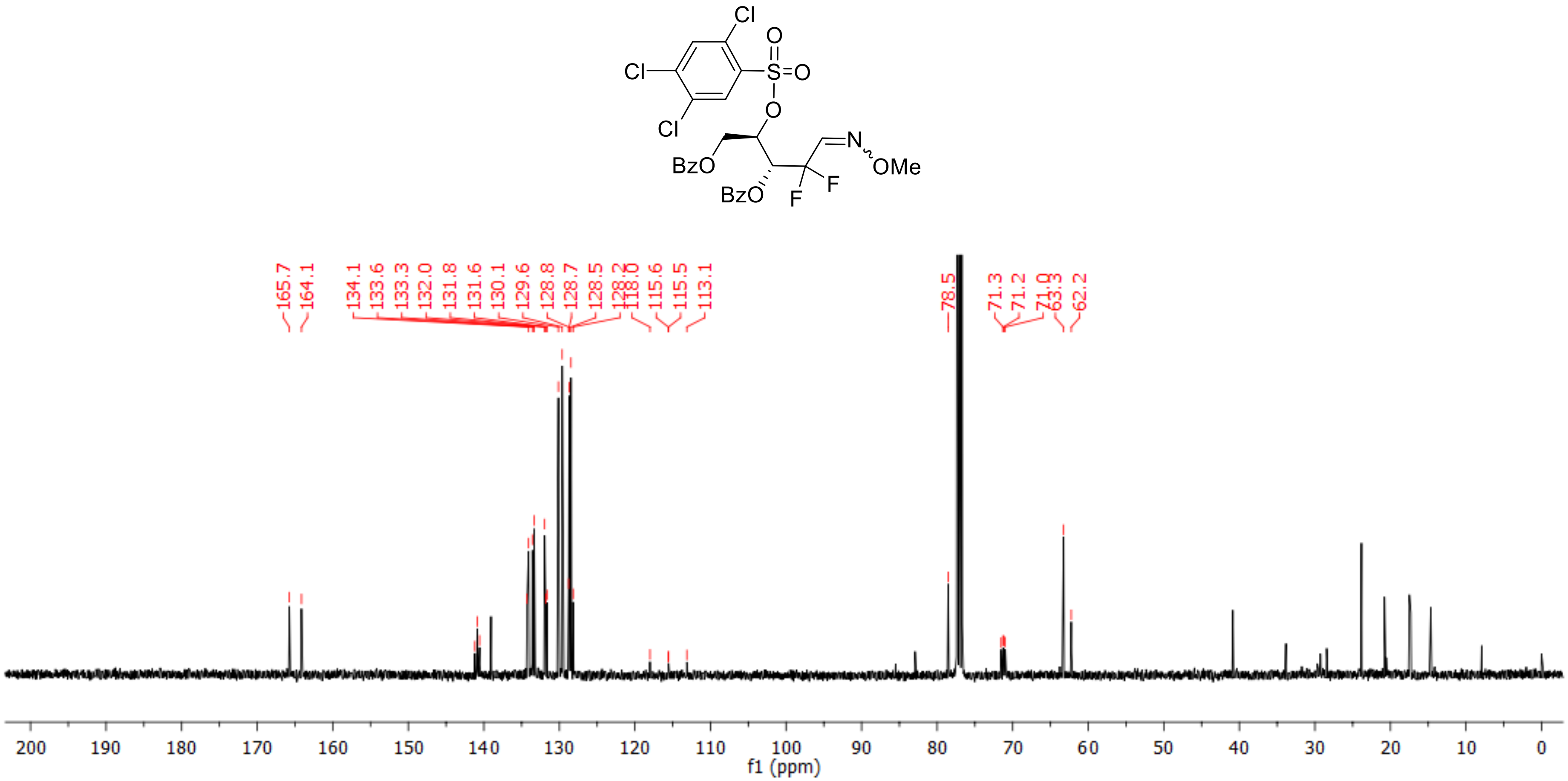
^{19}F NMR (376 MHz, $CDCl_3$)

(2*R*,3*R*,4*S*)-3,5-di-*O*-benzoyl-4-*O*-(2',4',5'-trichlorophenylsulfonyl)-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*)

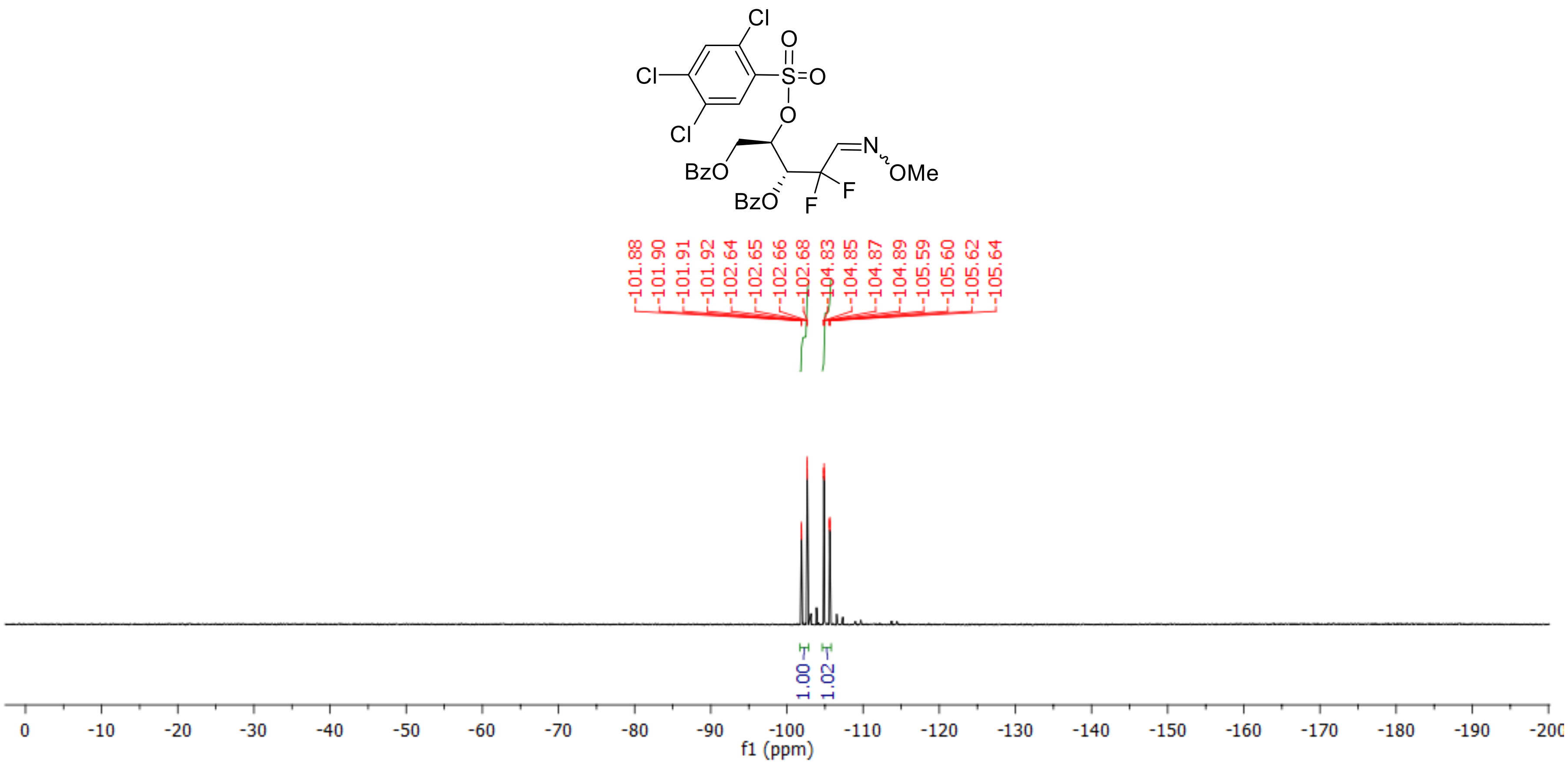




^{1}H NMR (400 MHz, $CDCl_3$)



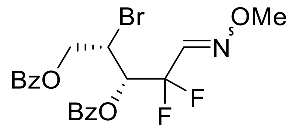
^{13}C NMR (101 MHz, $CDCl_3$)

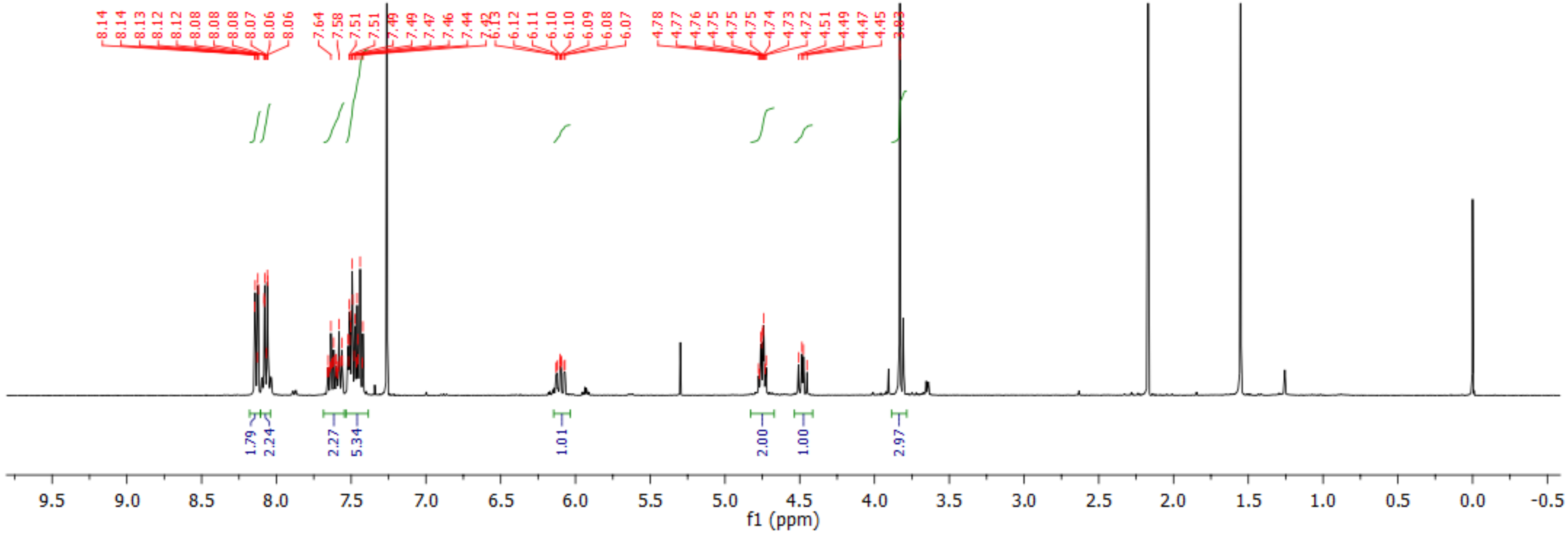


^{19}F NMR (376 MHz, $CDCl_3$)

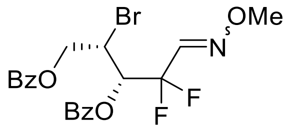
(2*S*,3*R*,4*S*)-3,5-di-*O*-benzoyl-4-bromo-4-*O*-(2',4',5'-trichlorophenylsulfonyl)-2-*gem*-difluoro-4-hydroxy-1-(methoxyimino)pentane (*E*/*Z*)

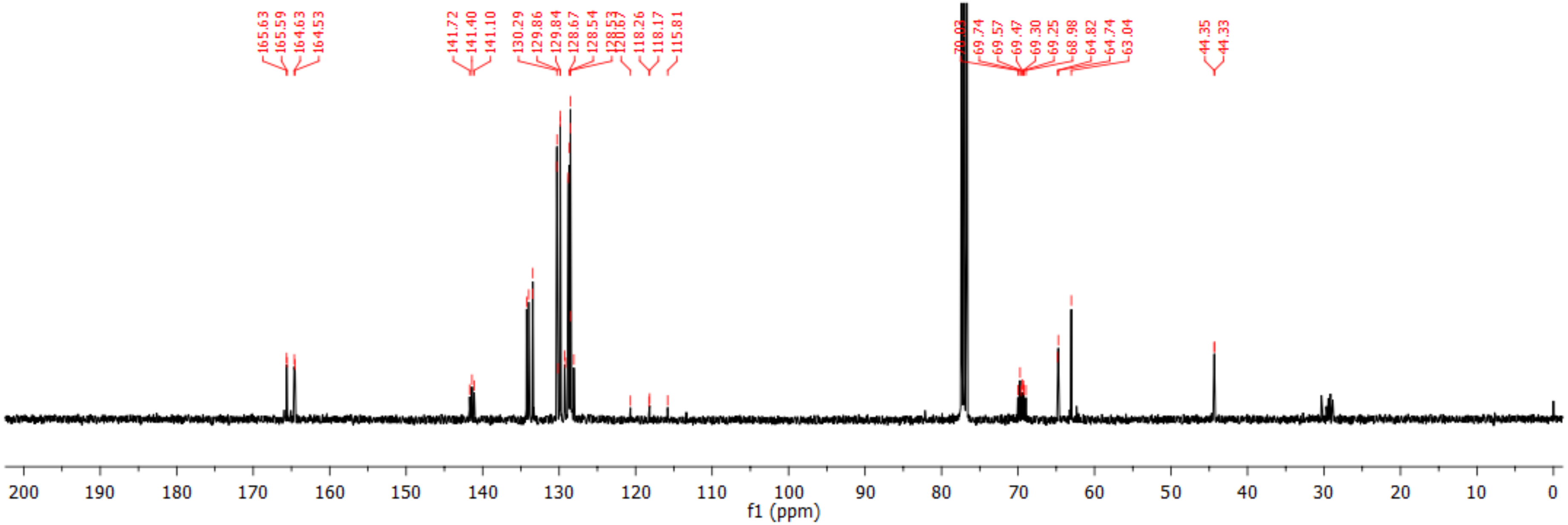
9



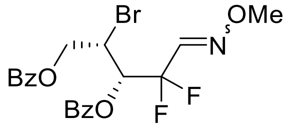


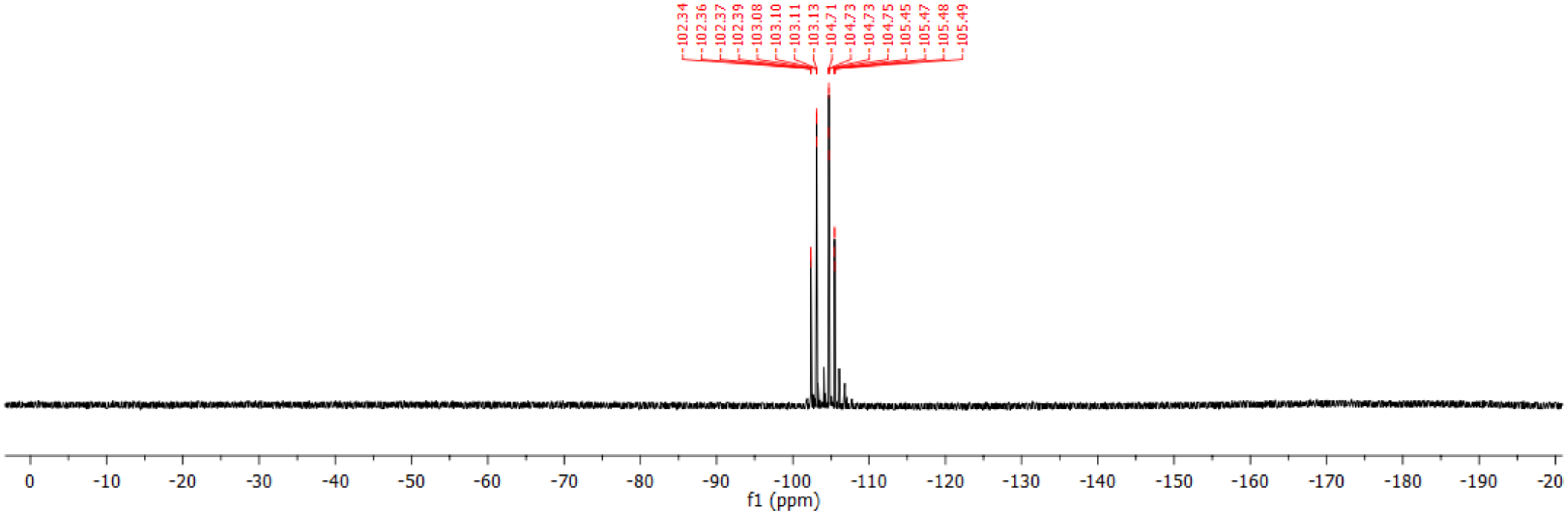
^{1}H NMR (400 MHz, $CDCl_3$)





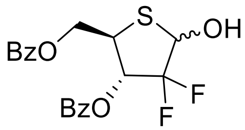
^{13}C NMR (101 MHz, $CDCl_3$)

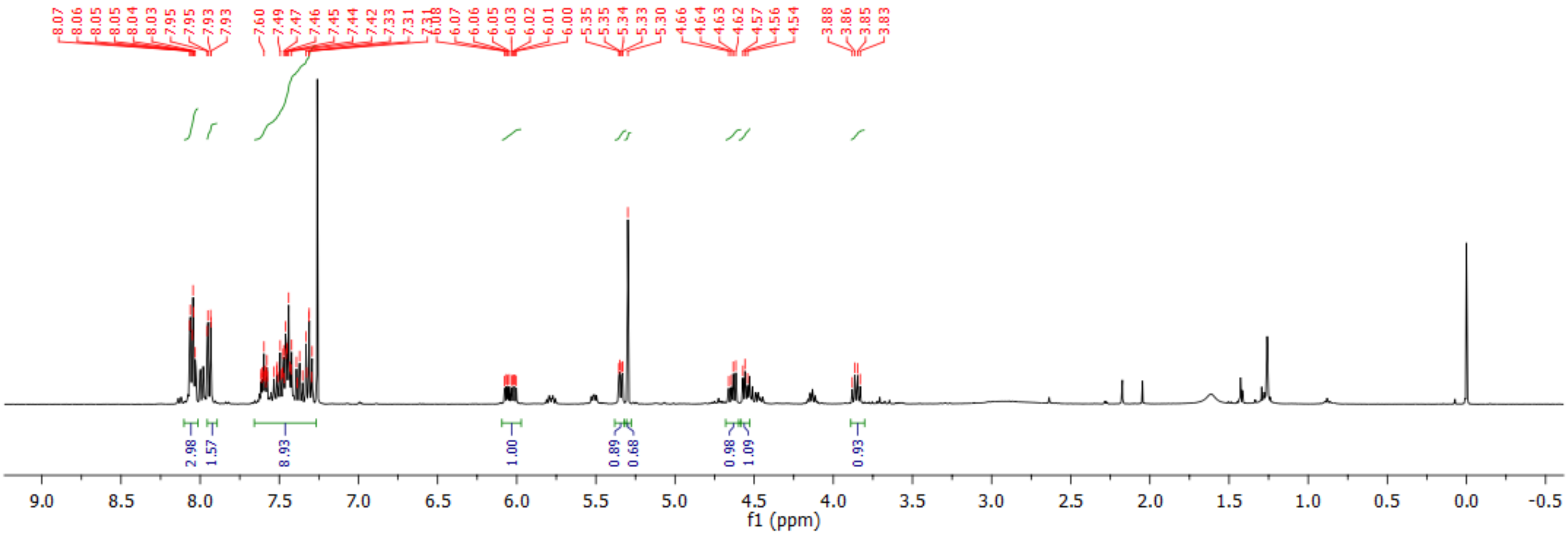




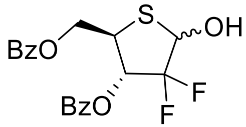
^{19}F NMR (376 MHz, $CDCl_3$)

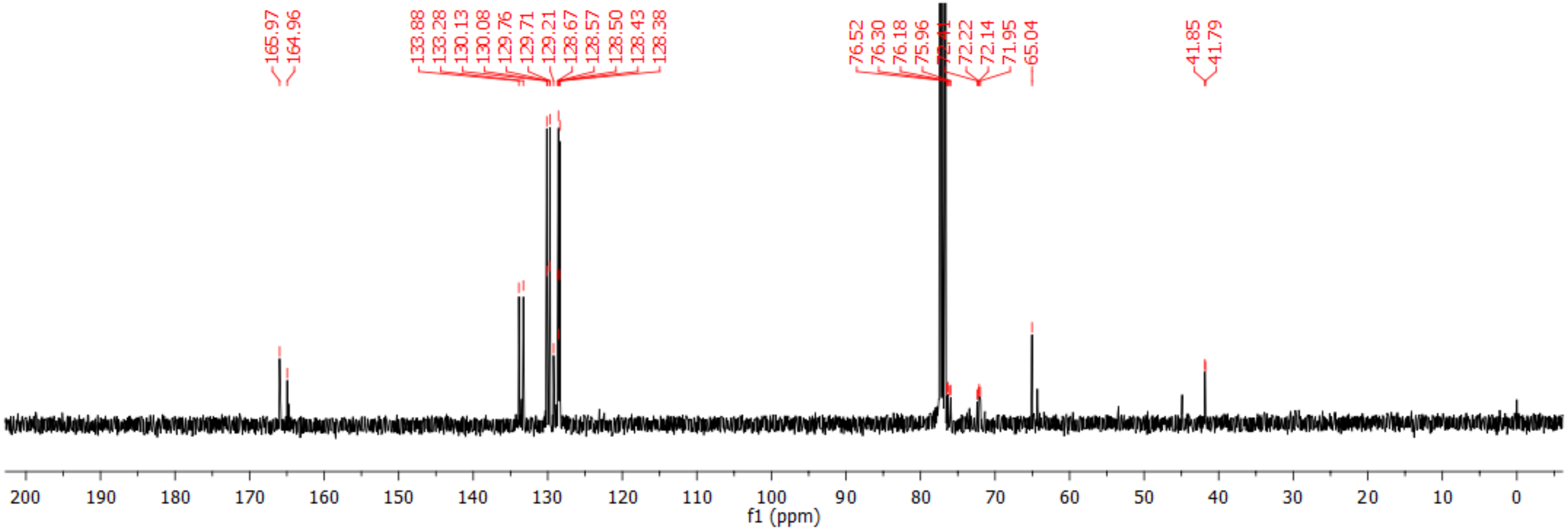
3,5-di-*O*-benzoyl-2-deoxy-2-difluoro-1-α,β-(4-thio-D-ribofuranose) **10**



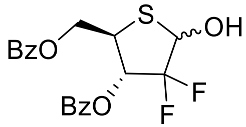


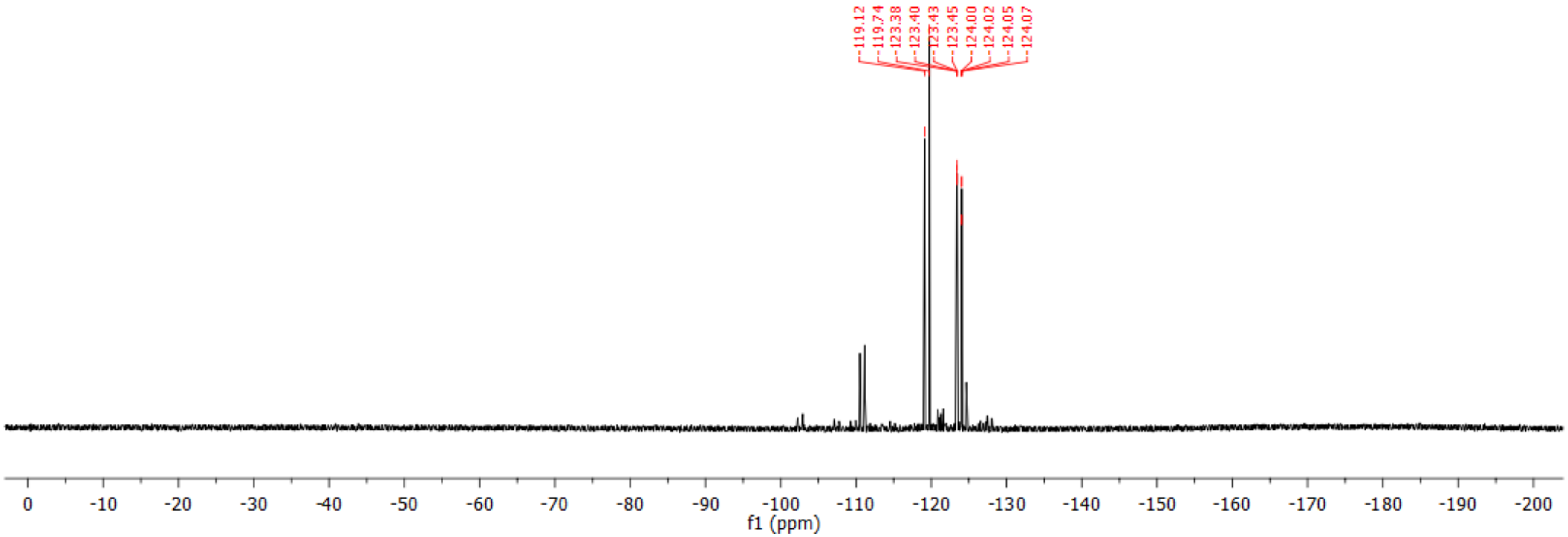
^{1}H NMR (400 MHz, $CDCl_3$)





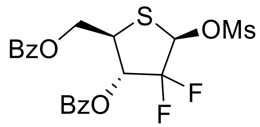
^{13}C NMR (101 MHz, $CDCl_3$)

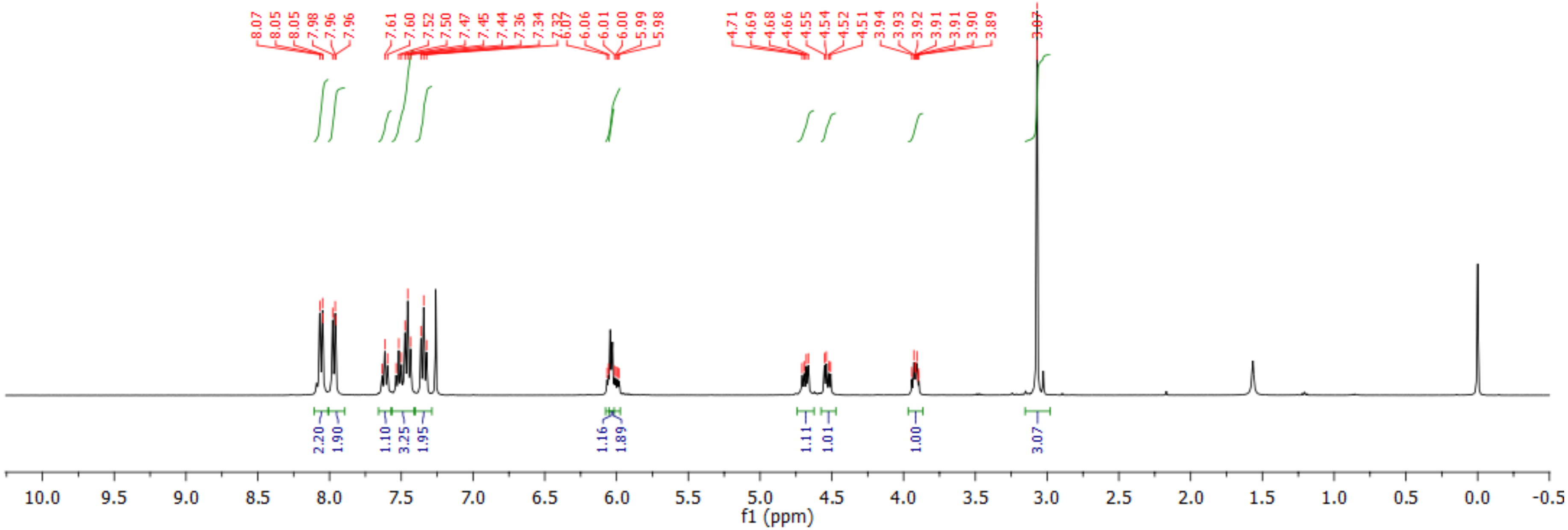




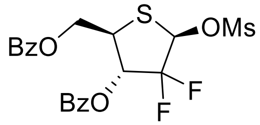
^{19}F NMR (376 MHz, $CDCl_3$)

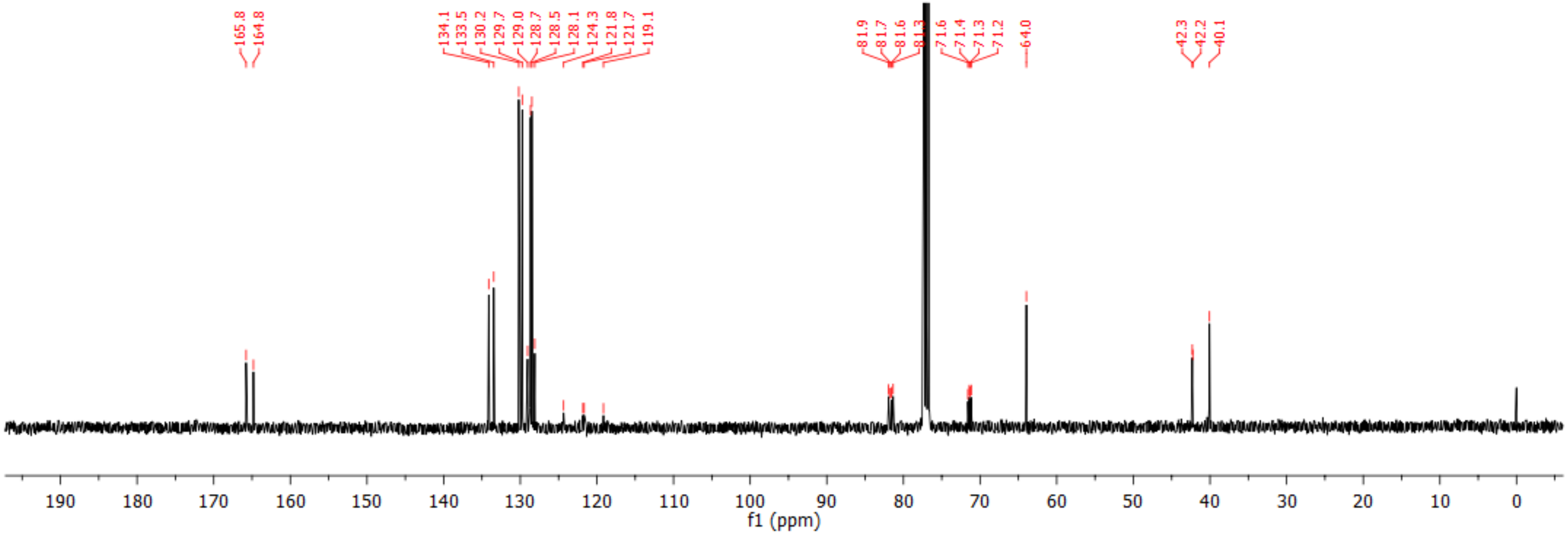
3,5-di-*O*-benzoyl-2-deoxy-2-difluoro-1-*O*-mesyl-1-β-(4-thio-D-ribofuranose) **11**



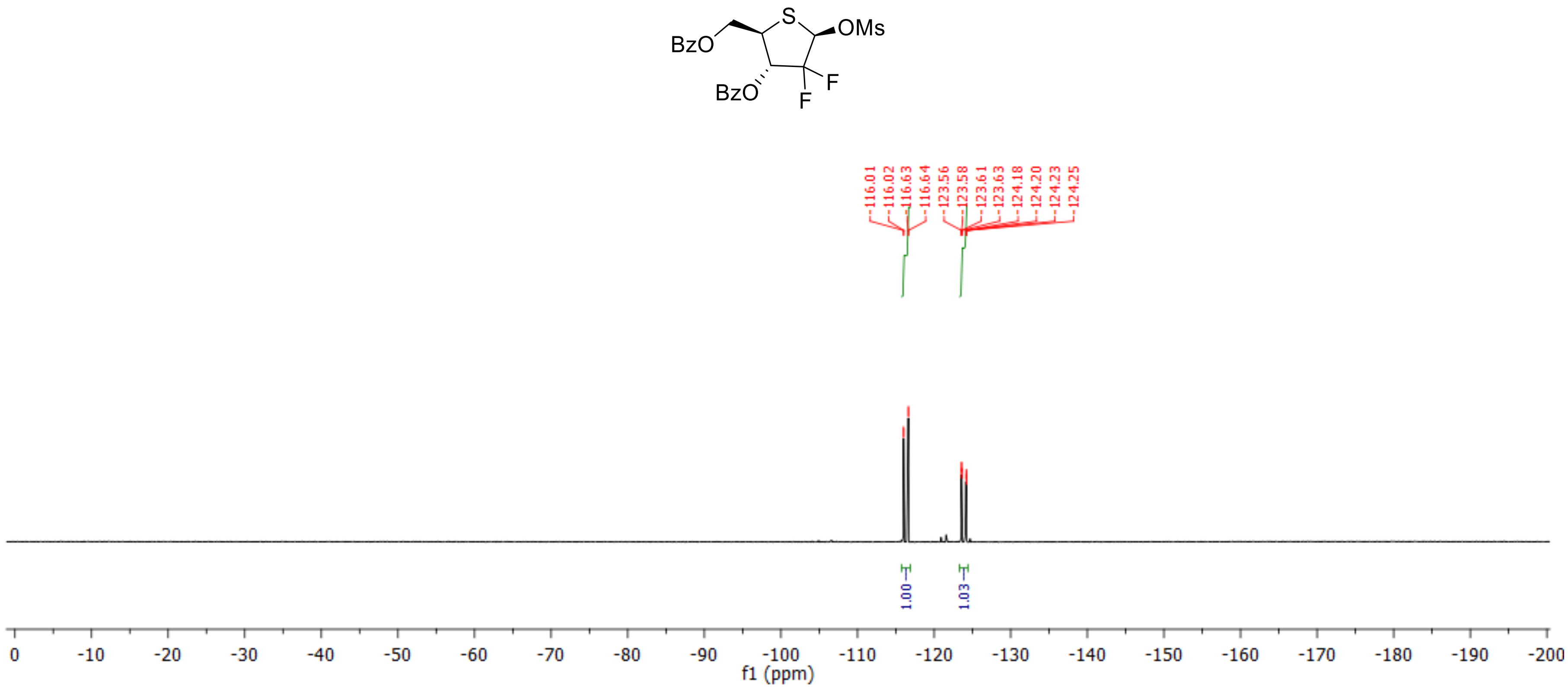


^{1}H NMR (400 MHz, $CDCl_3$)



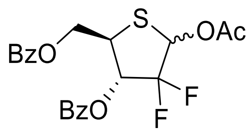


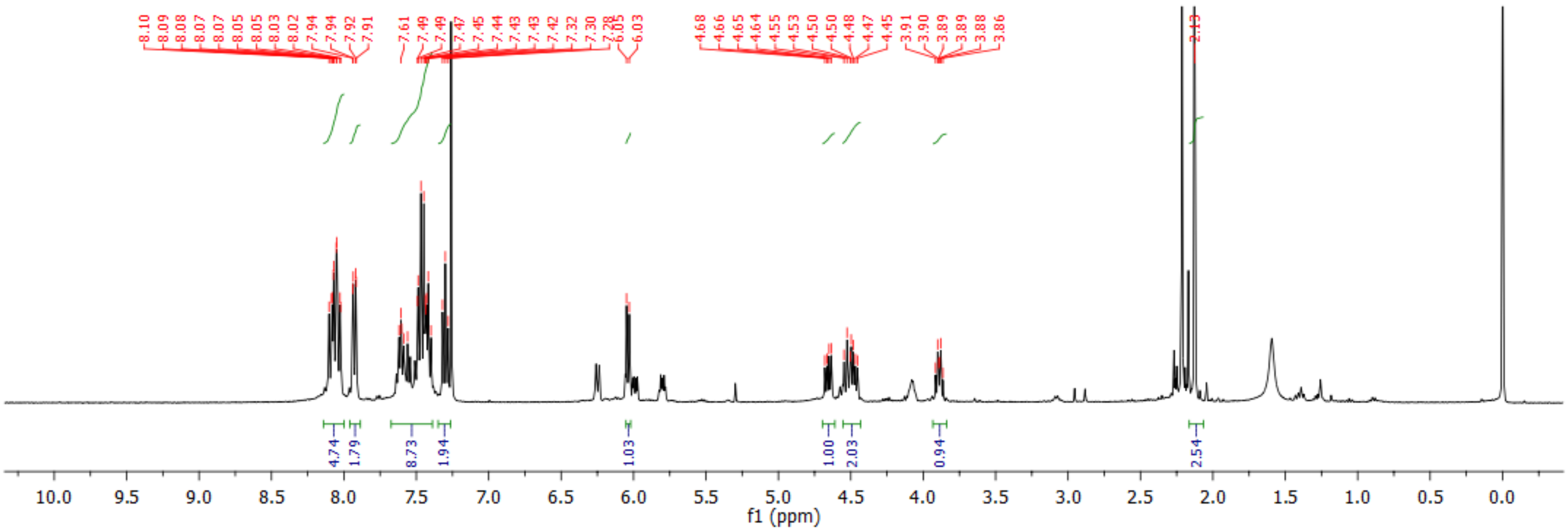
^{13}C NMR (101 MHz, $CDCl_3$)



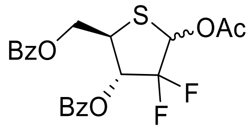
^{19}F NMR (376 MHz, $CDCl_3$)

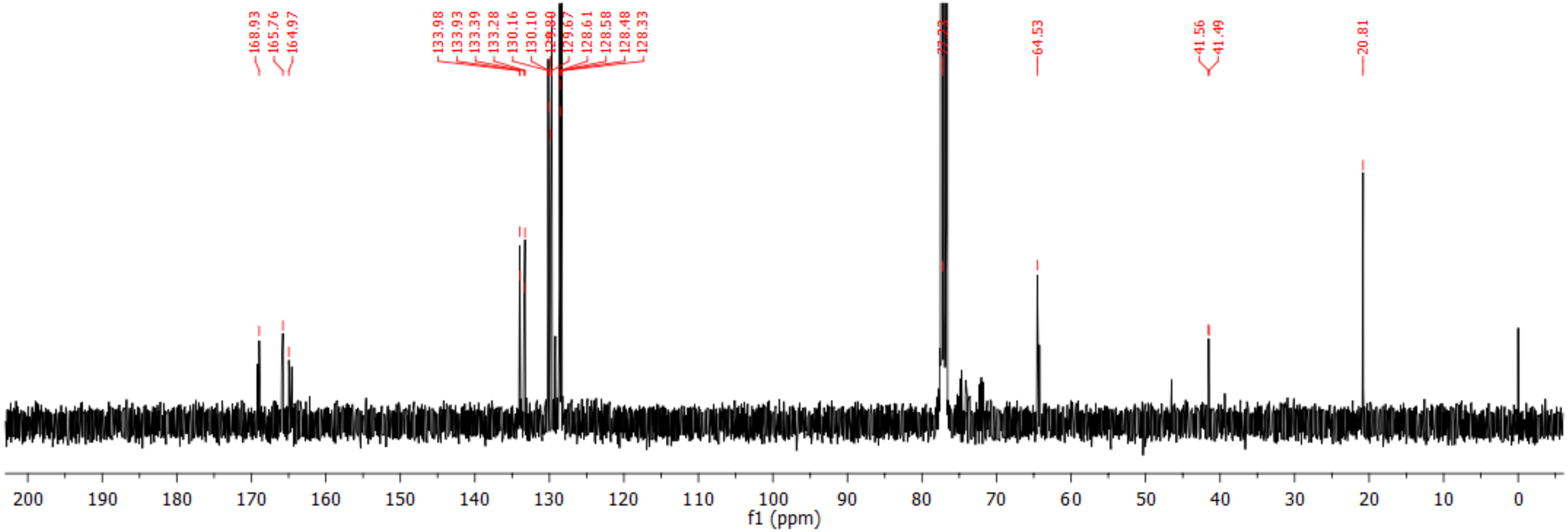
1-*O*-acetyl-3,5-di-*O*-benzoyl-2-deoxy-2-difluoro-1-α,β-(4-thio-D-ribofuranose) **12**



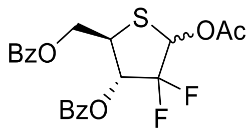


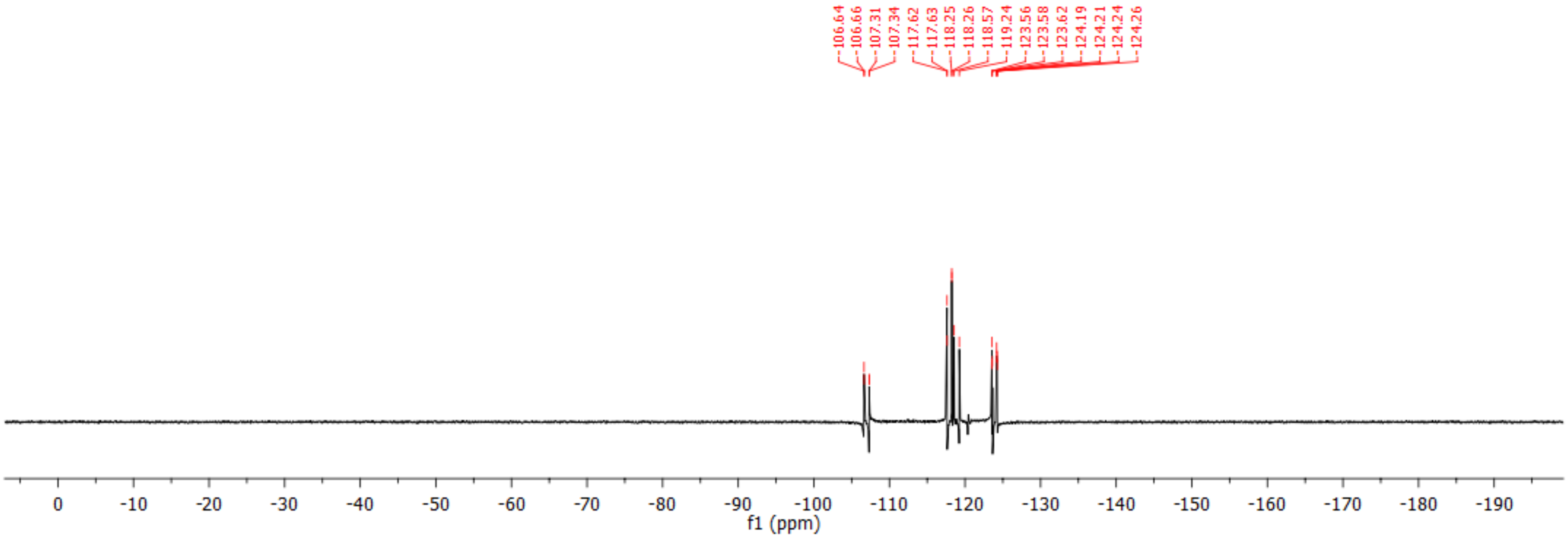
^{1}H NMR (400 MHz, $CDCl_3$)





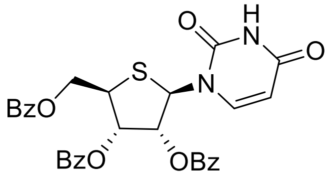
^{13}C NMR (101 MHz, $CDCl_3$)

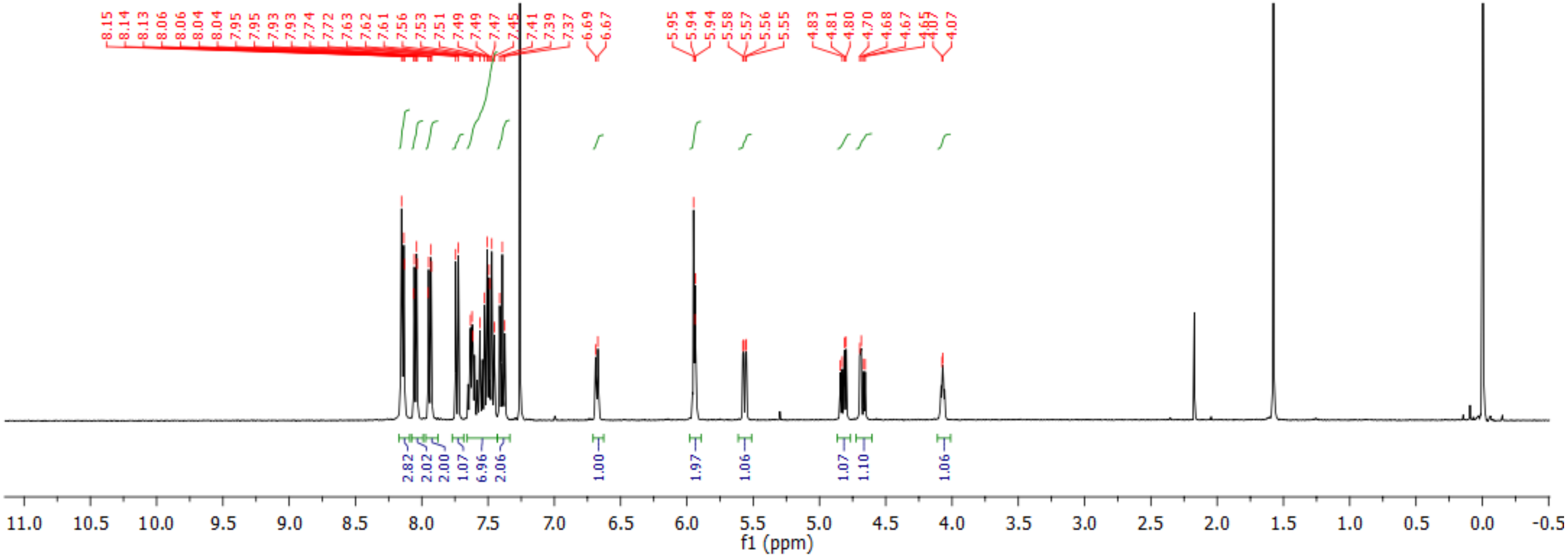




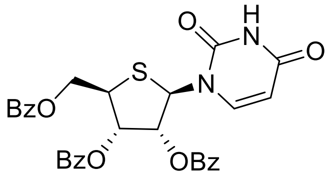
^{19}F NMR (376 MHz, $CDCl_3$)

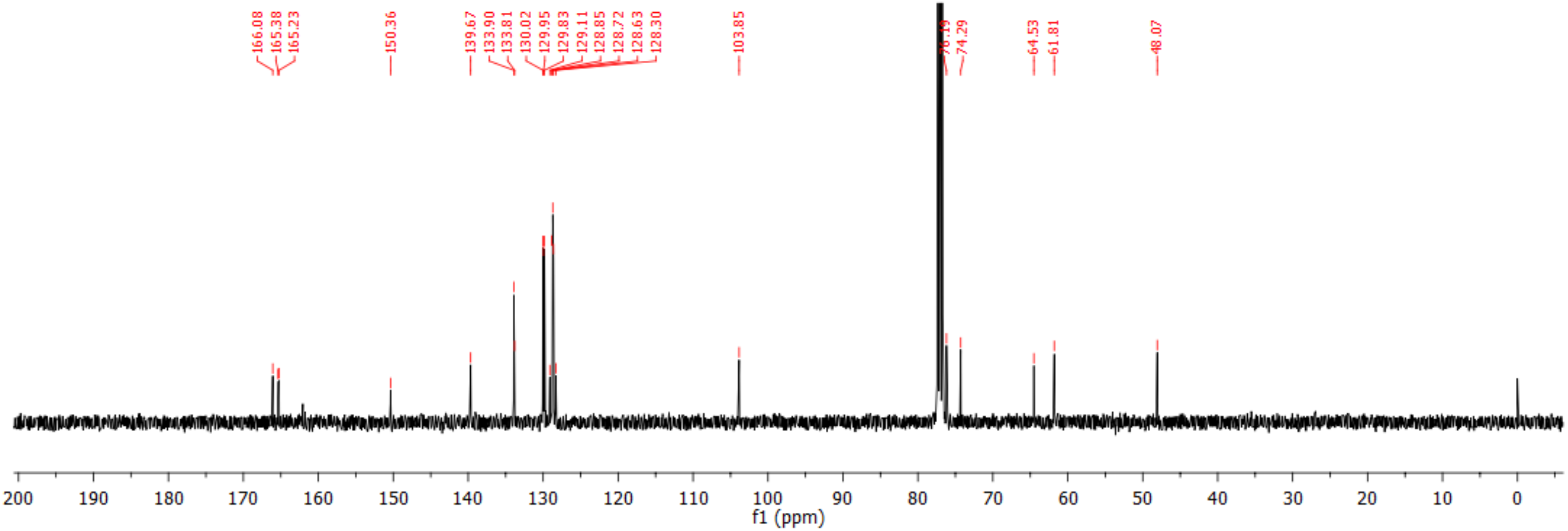
2’,3’,5’-tri-*O*-benzoyl,1’-β-(4’-thio-D-ribofuranosyl)uracil **13**





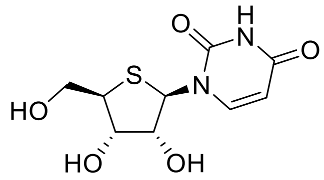
^{1}H NMR (400 MHz, $CDCl_3$)

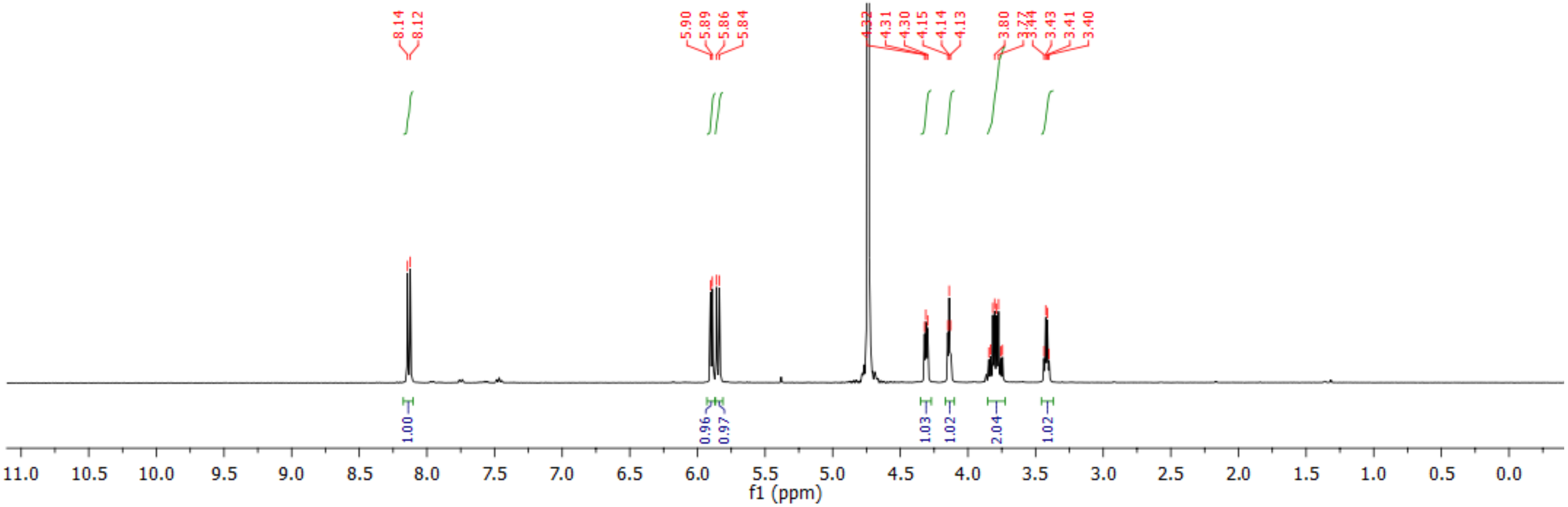




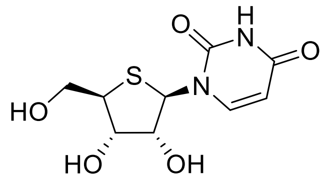
^{13}C NMR (101 MHz, $CDCl_3$)

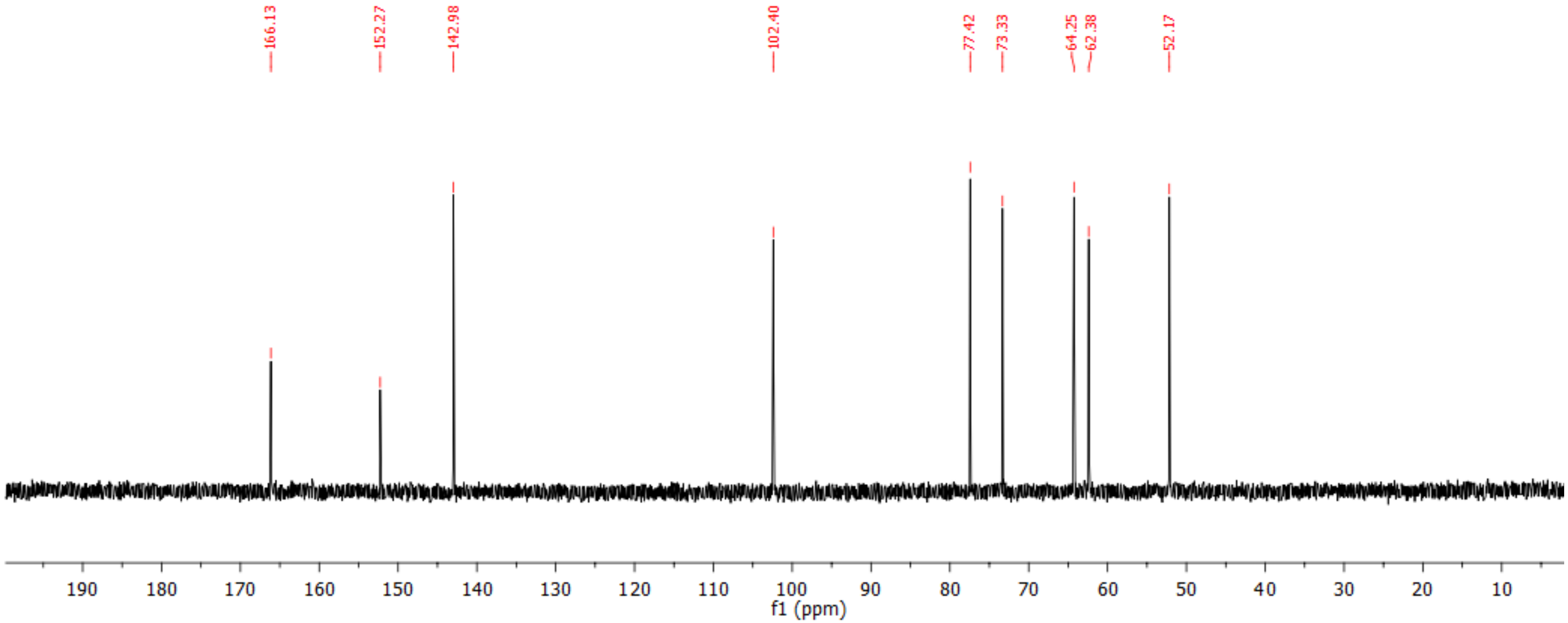
1'-β-(4'-thio-D-ribofuranosyl)uracil 14





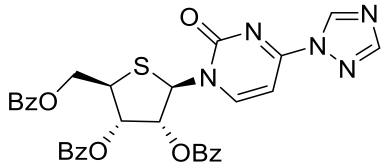
^1H NMR (400 MHz, D_2O)

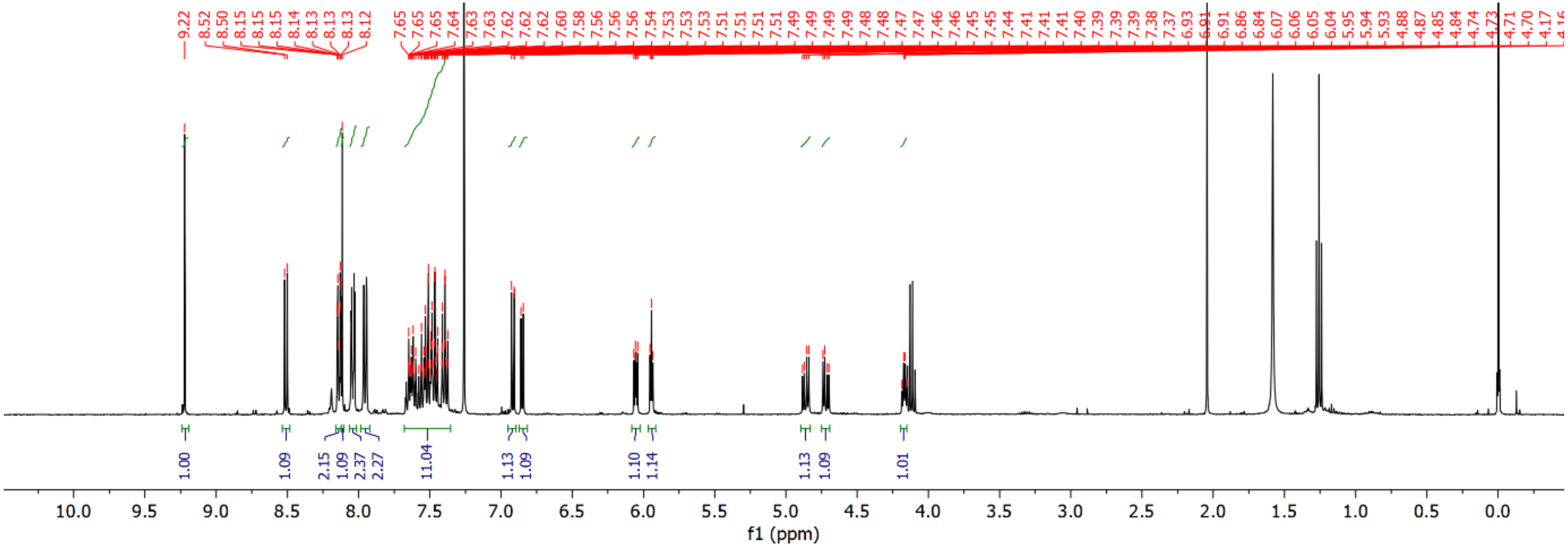




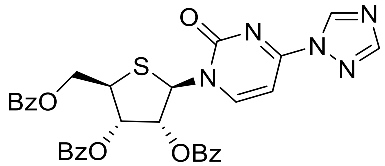
^{13}C NMR (101 MHz, D_2O)

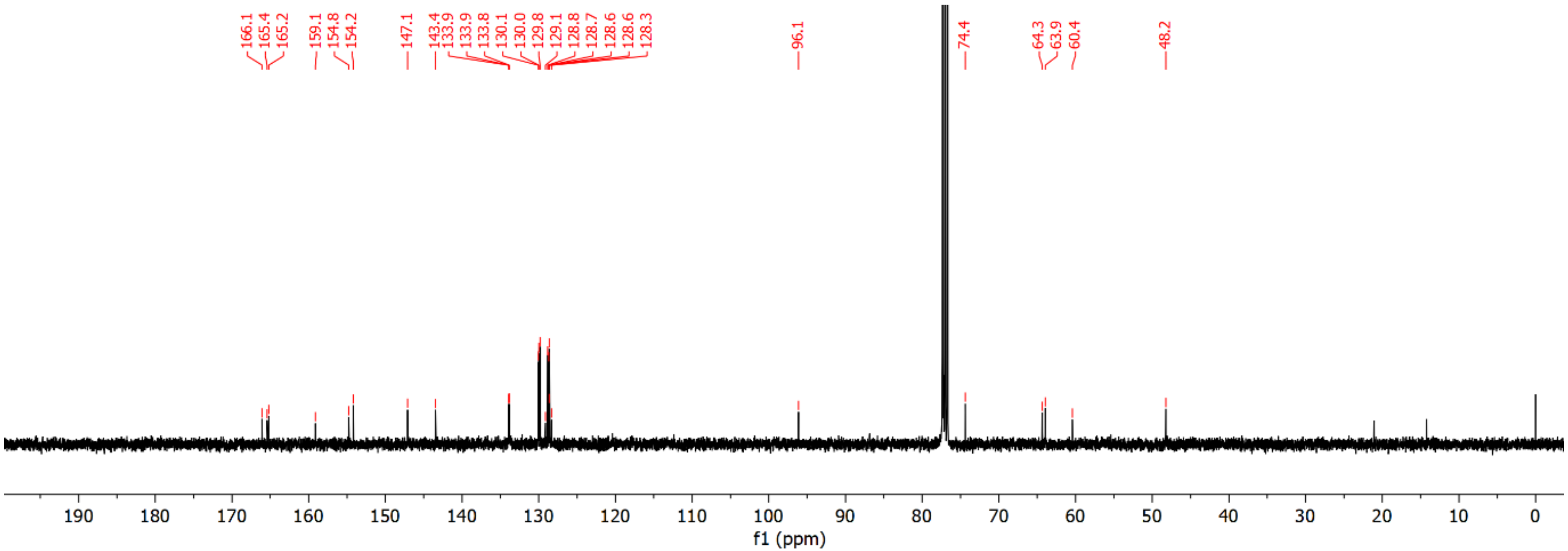
2',3',5'-tri-*O*-benzoyl-4-*C*-(1,2,4-triazole)-1'-β-(4'-thio-D-ribofuranosyl)uracil





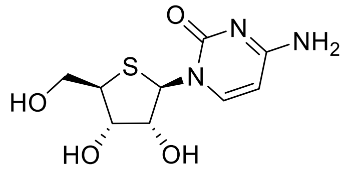
^{1}H NMR (400 MHz, $CDCl_3$)

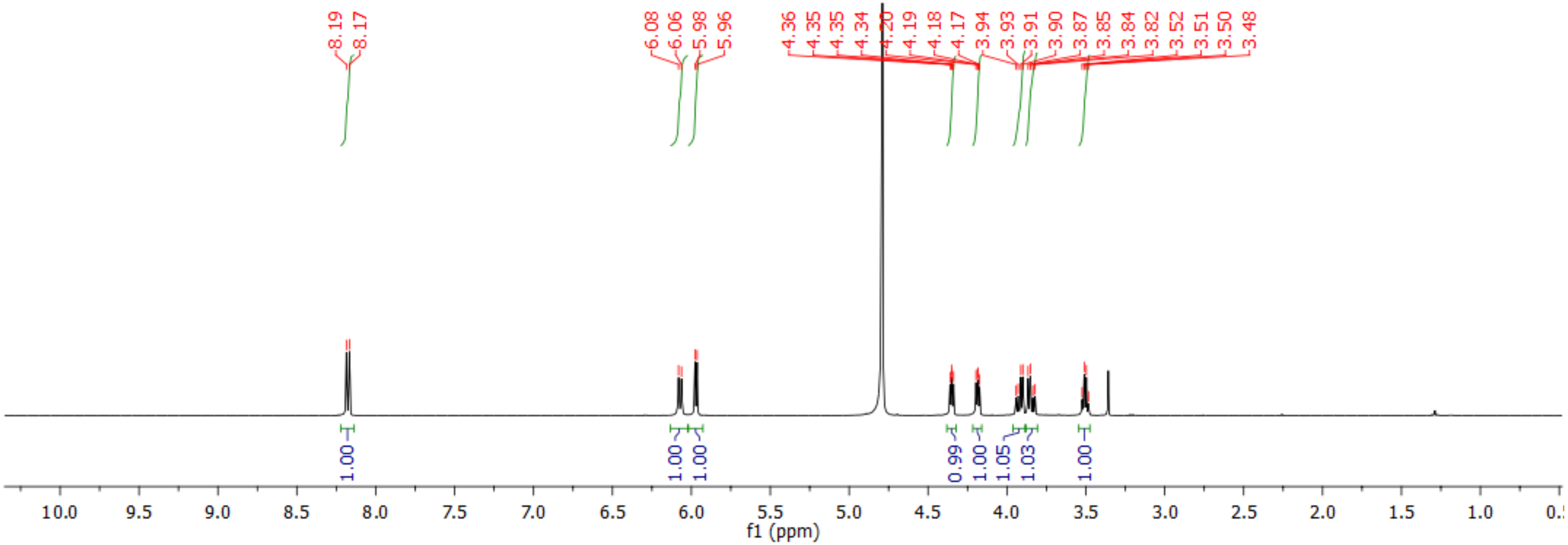




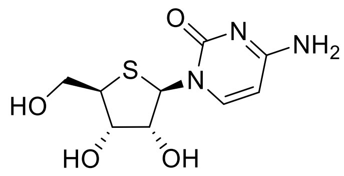
^{13}C NMR (101 MHz, $CDCl_3$)

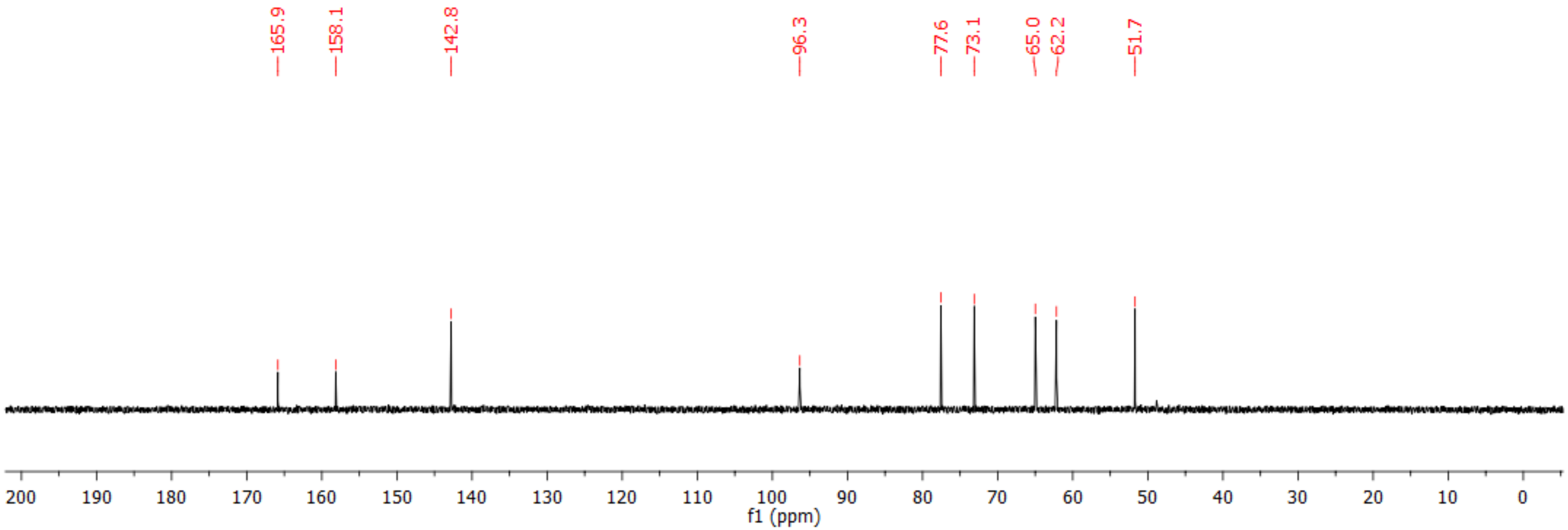
1’-β-(4’-thio-D-ribofuranosyl)cytosine **15**





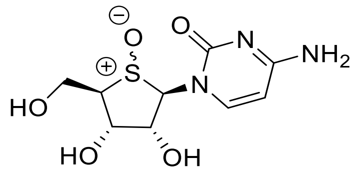
^{1}H NMR (400 MHz, D_2O)

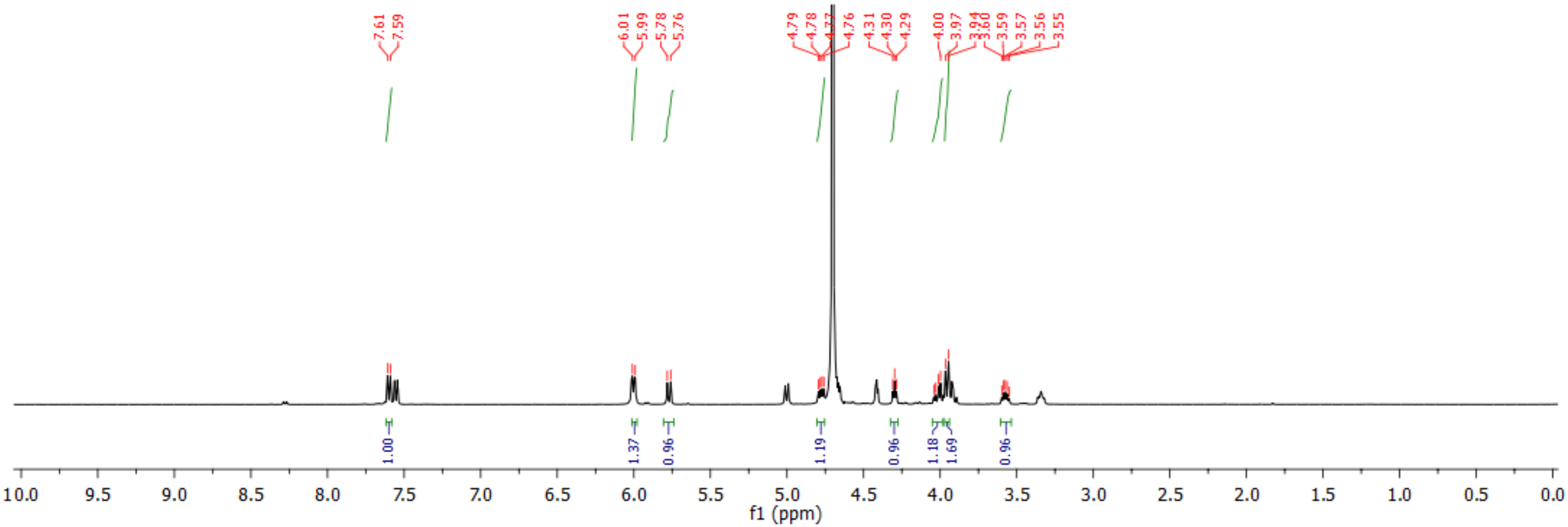




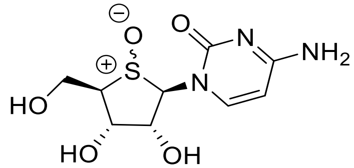
^{13}C NMR (101 MHz, D_2O)

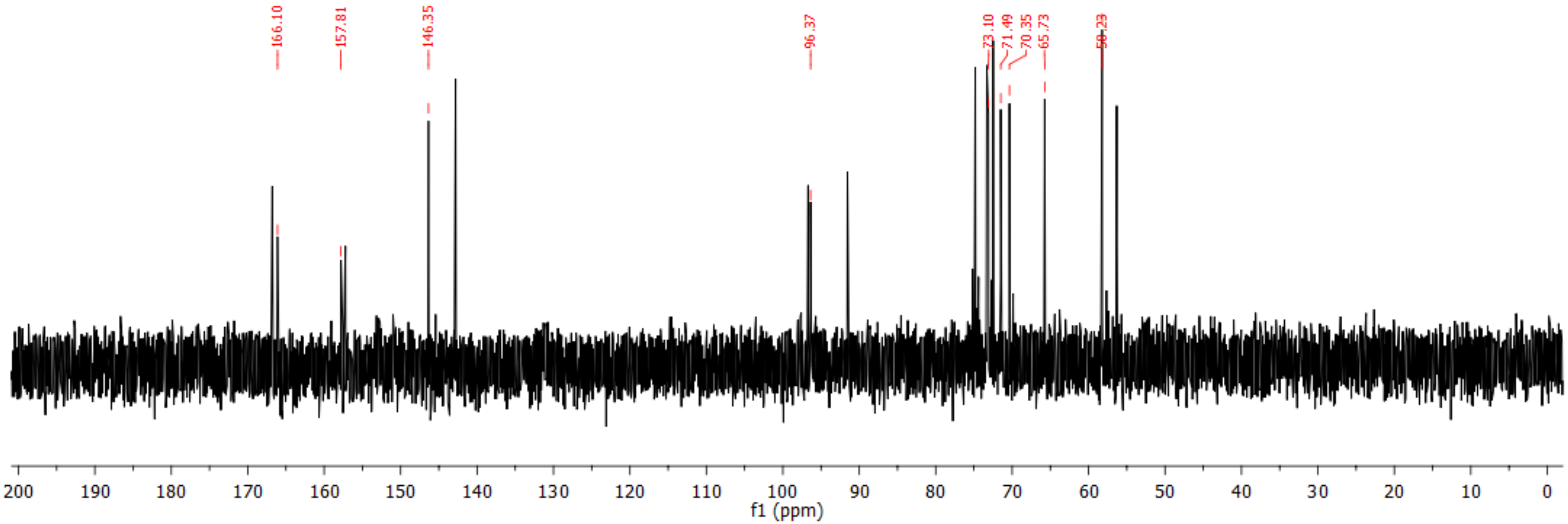
1'-β-(4'-sulfinyl(*S*/*R*)-D-ribofuranosyl)cytosine **16**





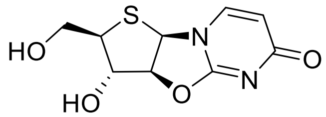
^{1}H NMR (400 MHz, D_2O)

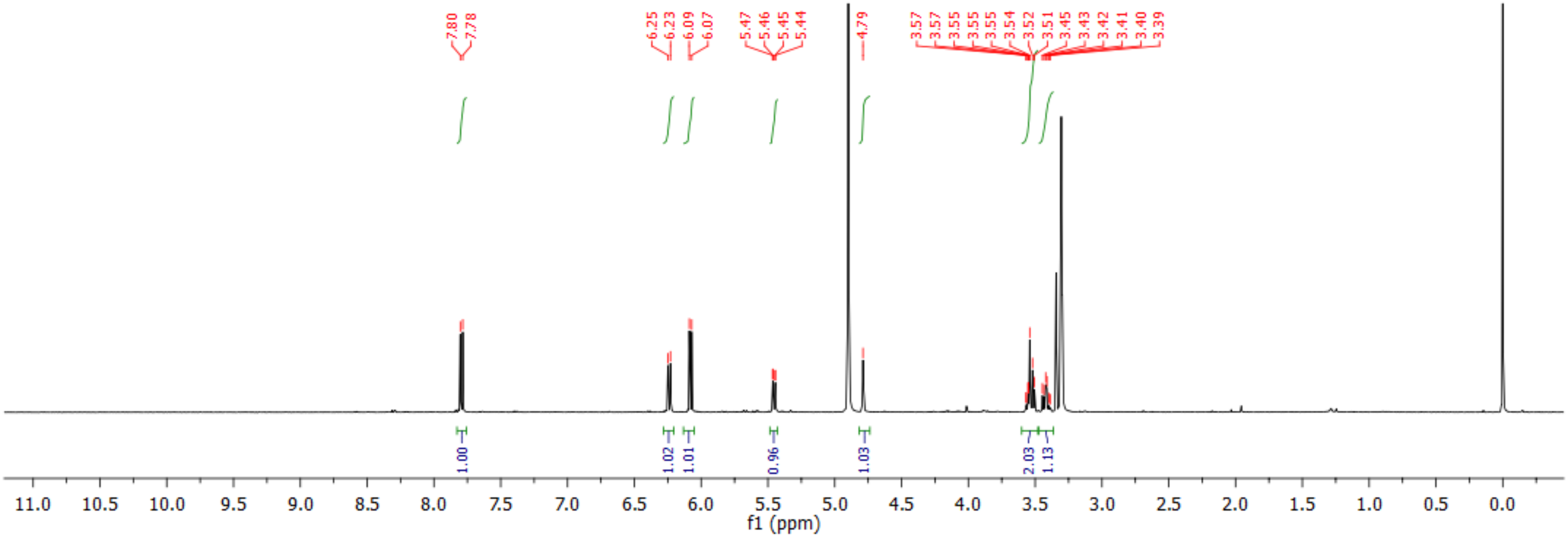




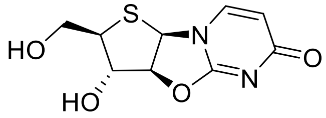
^{13}C NMR (101 MHz, D_2O)

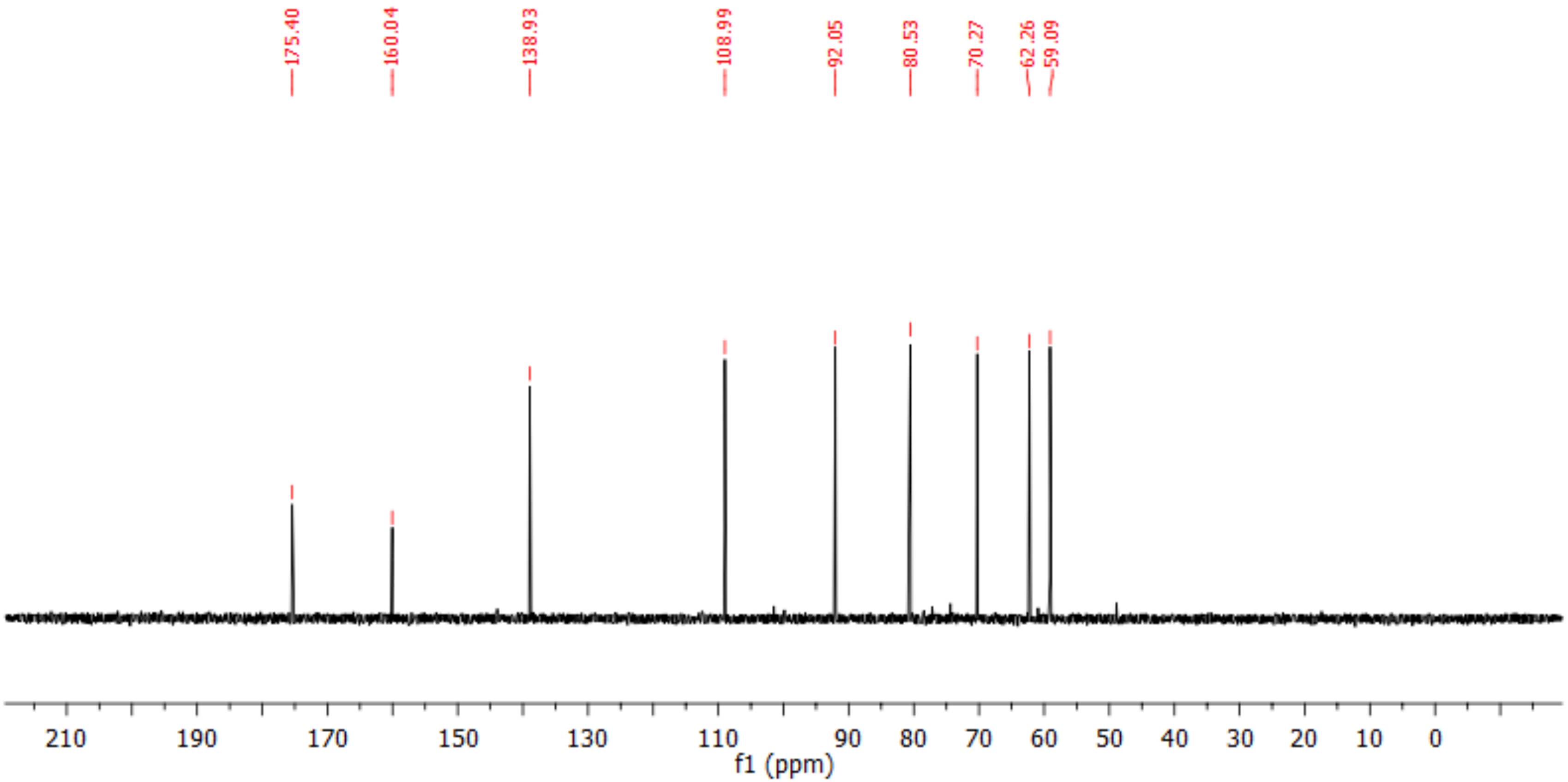
2',2-anhydro-1'-β-(4'-thio-D-ribofuranosyl)uracil **17**





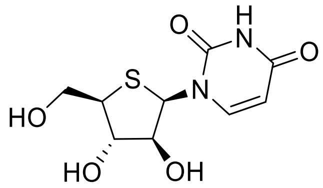
^{1}H NMR (400 MHz, D_2O)

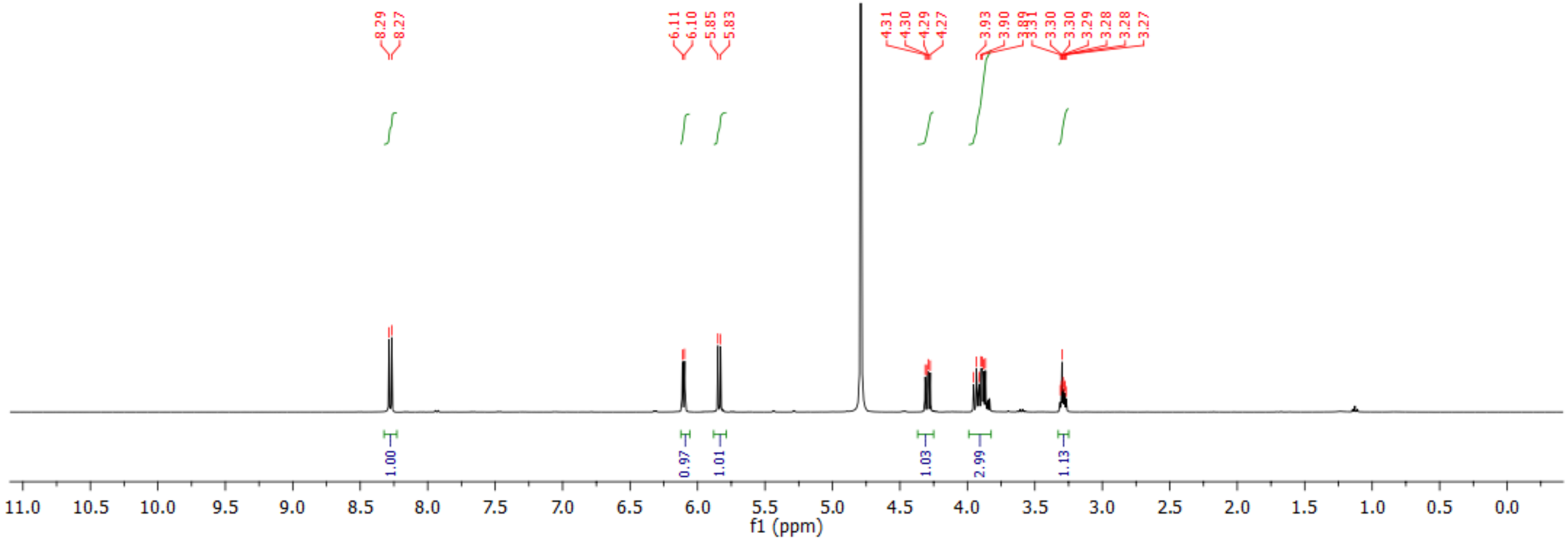




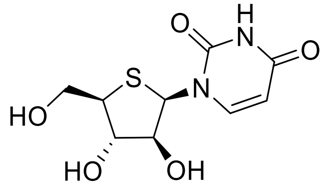
^{13}C NMR (101 MHz, D_2O)

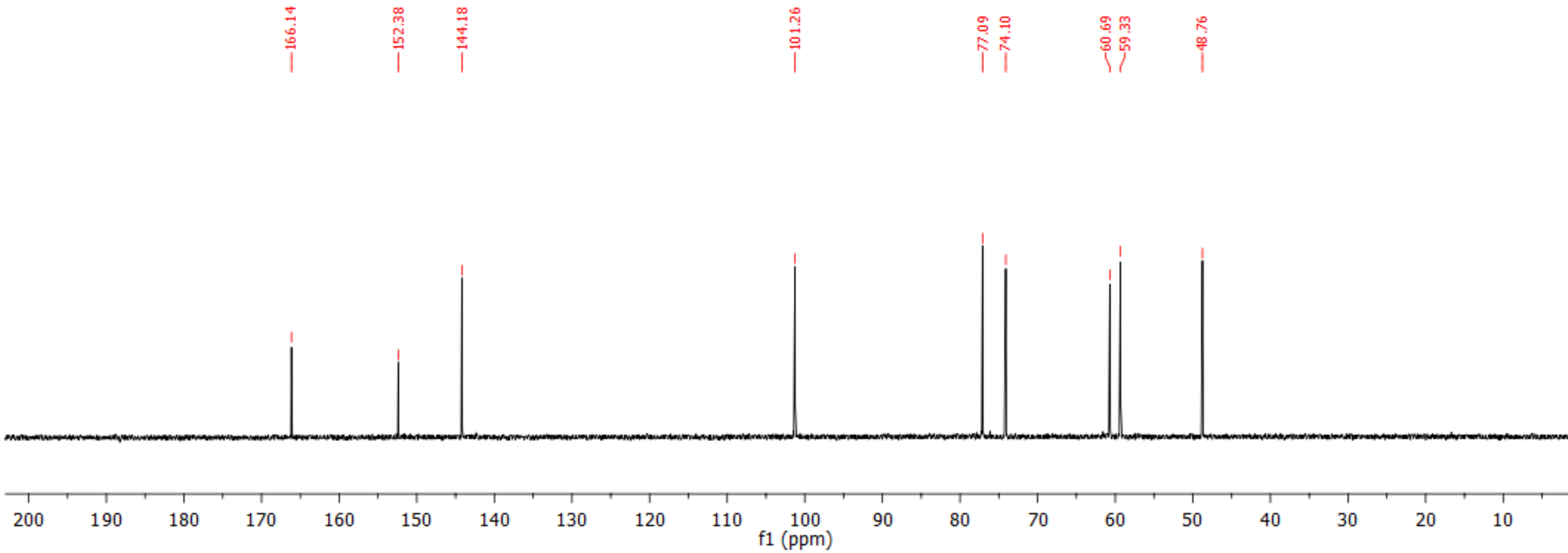
1’-β-(4’-thio-D-arabinofuranosyl)uracil **18**





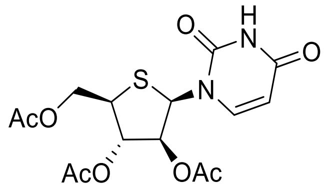
^{1}H NMR (400 MHz, D_2O)

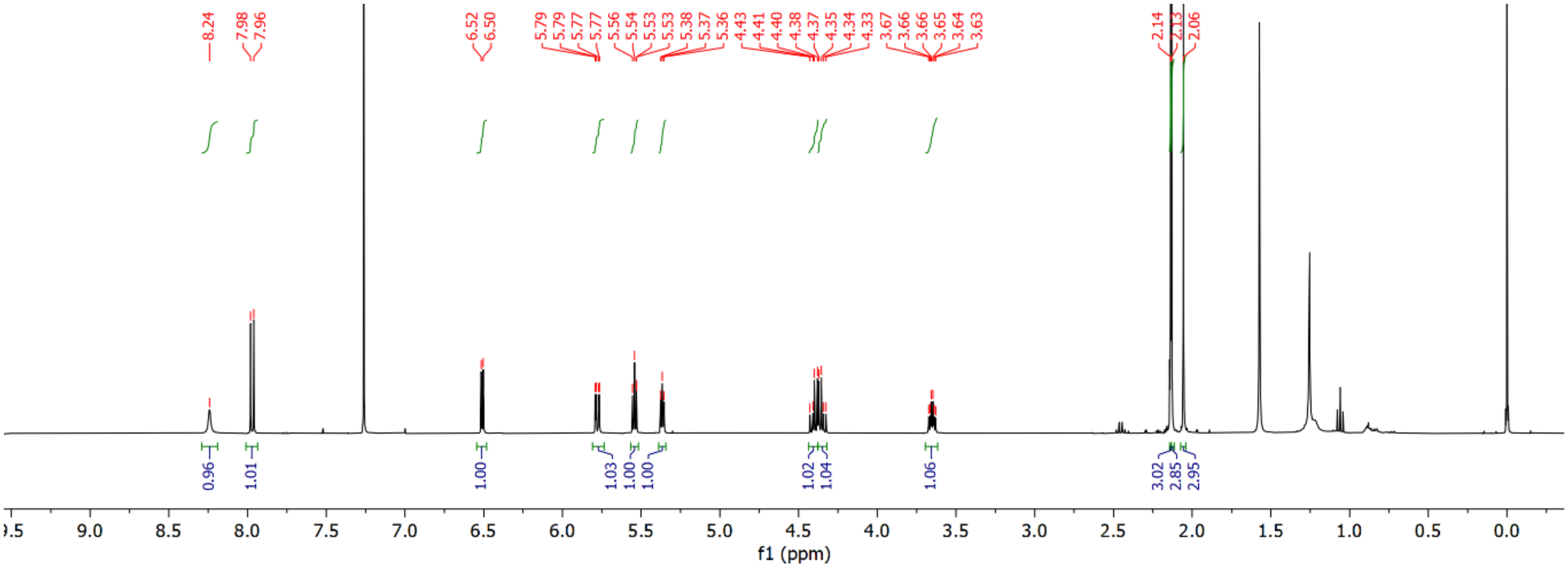




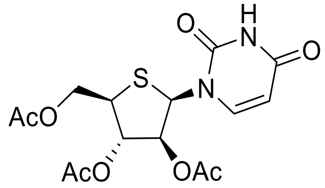
^{13}C NMR (101 MHz, D_2O)

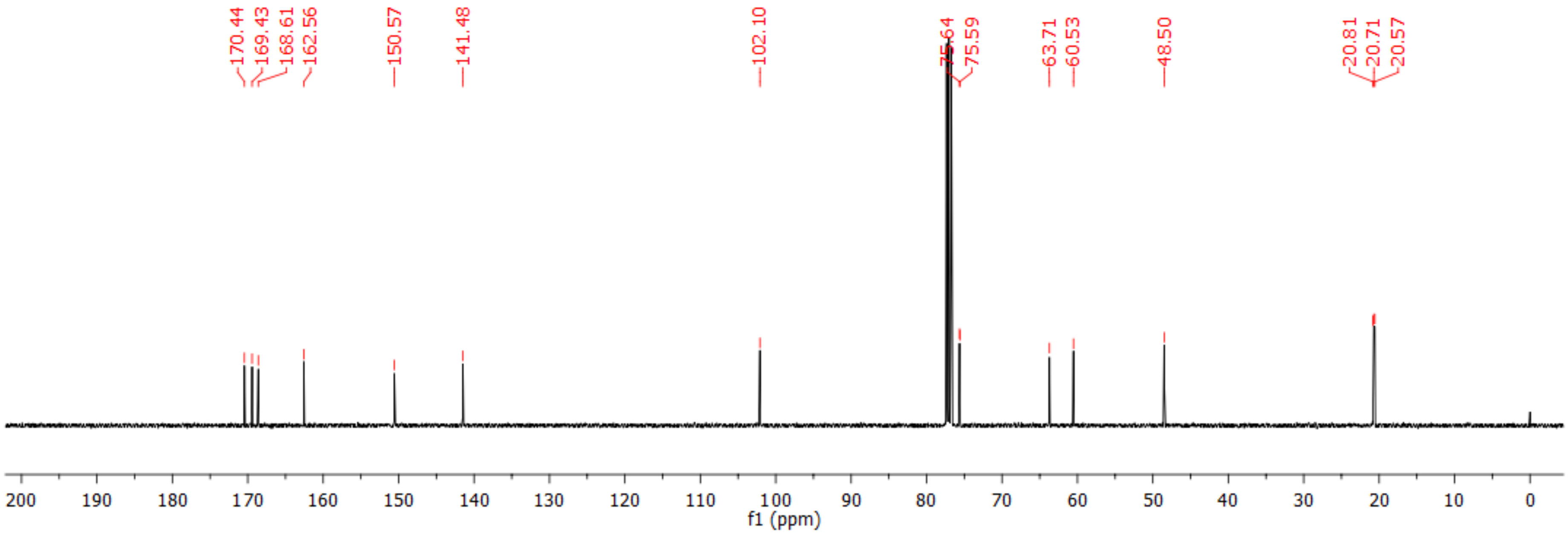
2',3',5'-tri-*O*-acetyl-1'-β-(4'-thio-D-arabinofuranosyl)uracil





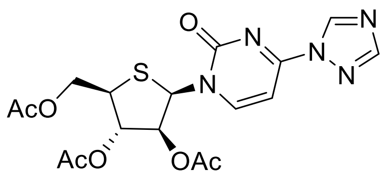
^{1}H NMR (400 MHz, $CDCl_3$)

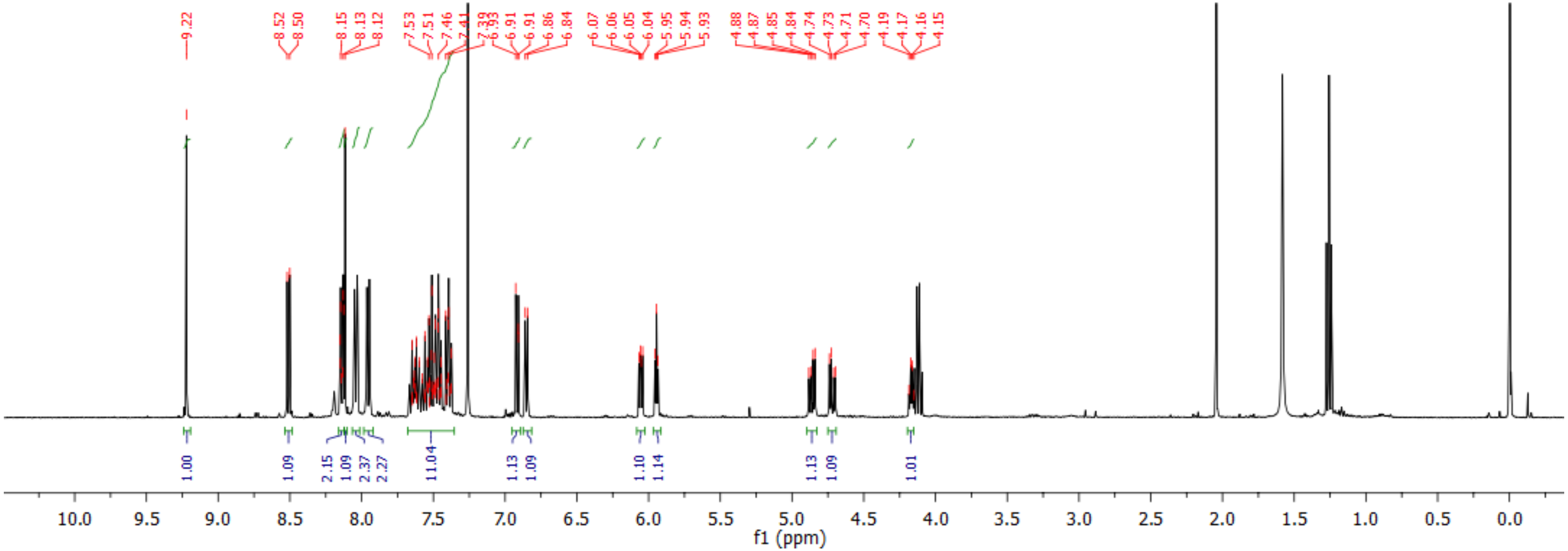




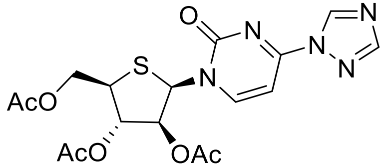
^{13}C NMR (101 MHz, $CDCl_3$)

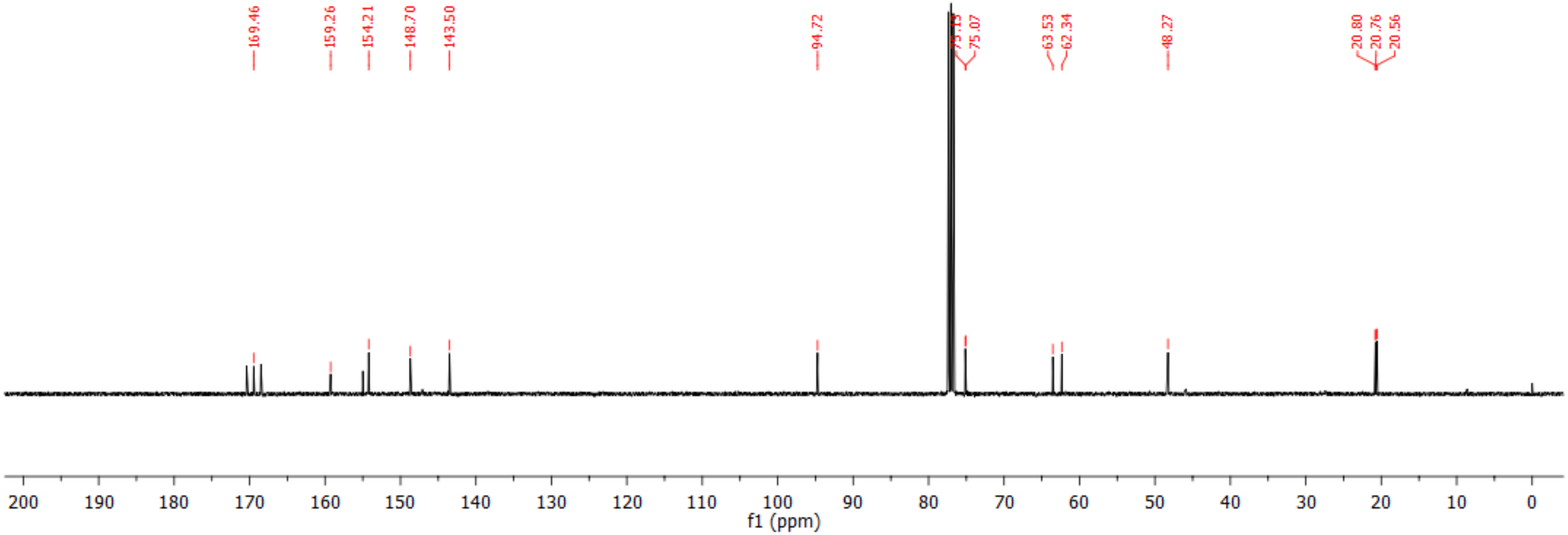
2’,3’,5’-*O*-tri-acetyl-4-*C*-(1,2,4-triazole)-1’-β-(4’-thio-D-arabinofuranosyl)uracil





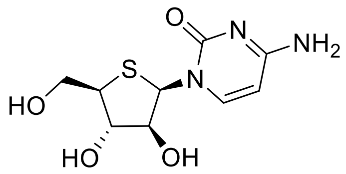
^{1}H NMR (400 MHz, $CDCl_3$)

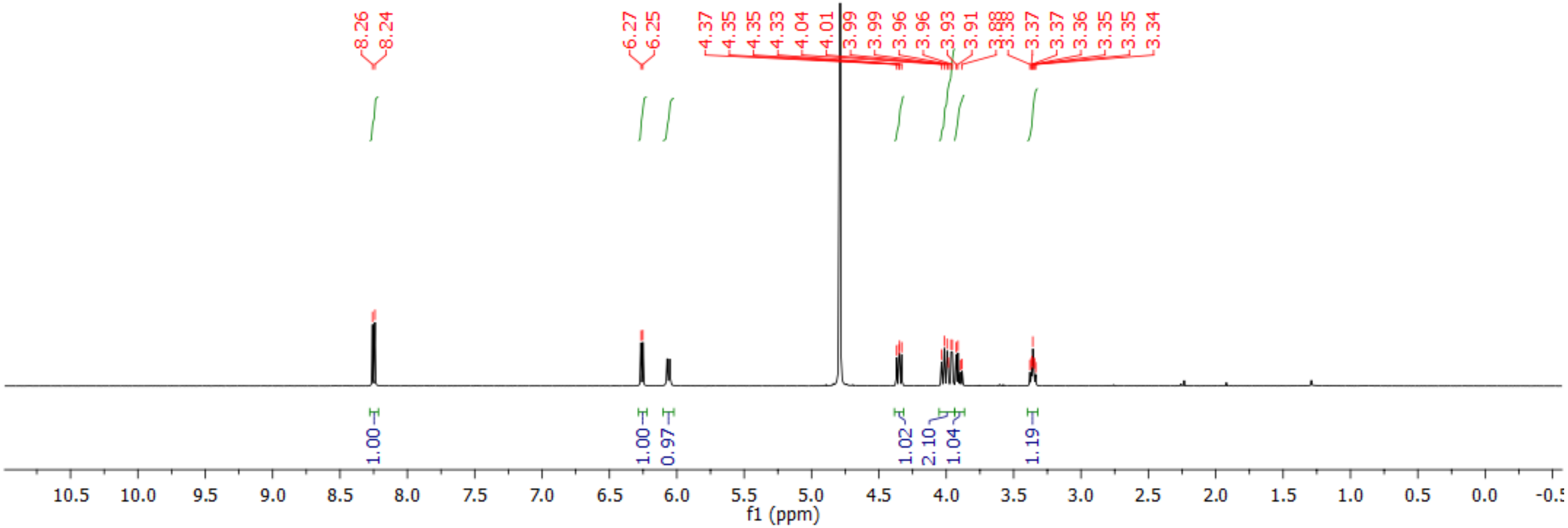




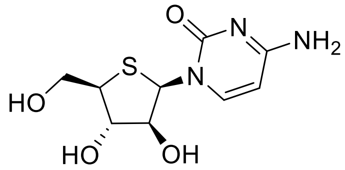
^{13}C NMR (101 MHz, $CDCl_3$)

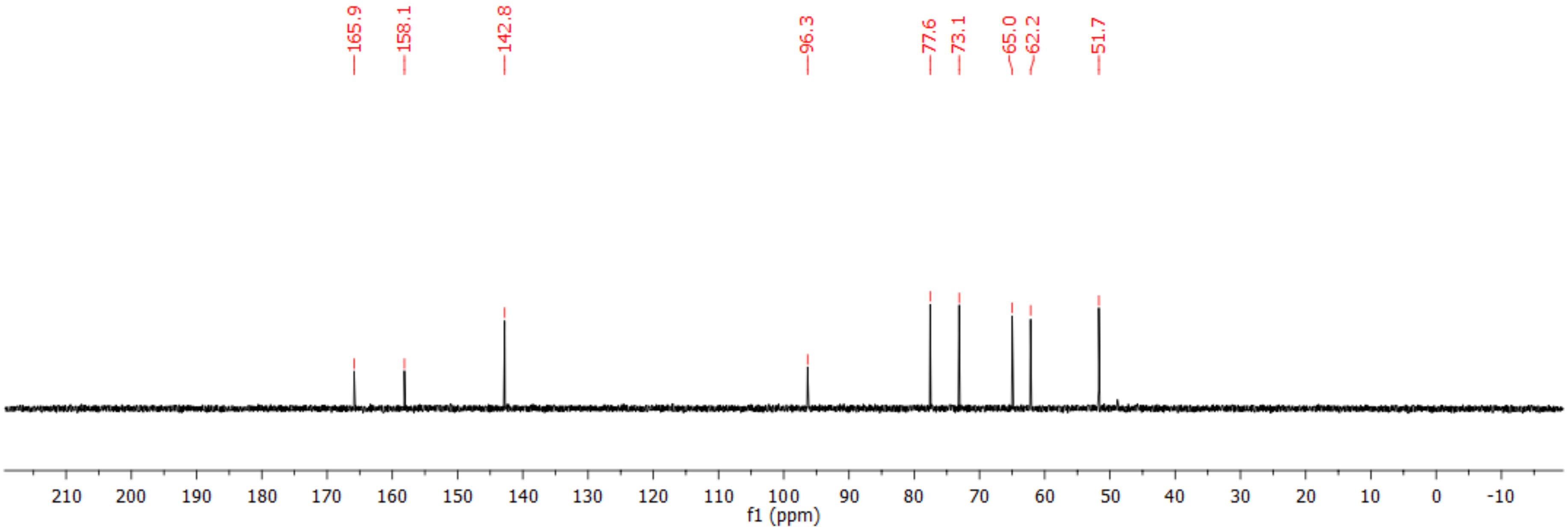
1'-β-(4'-thio-D-arabinofuranosyl)cytosine **19**





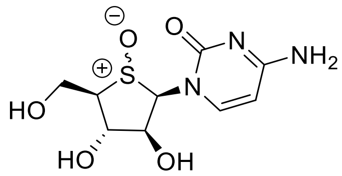
^{1}H NMR (400 MHz, D_2O)

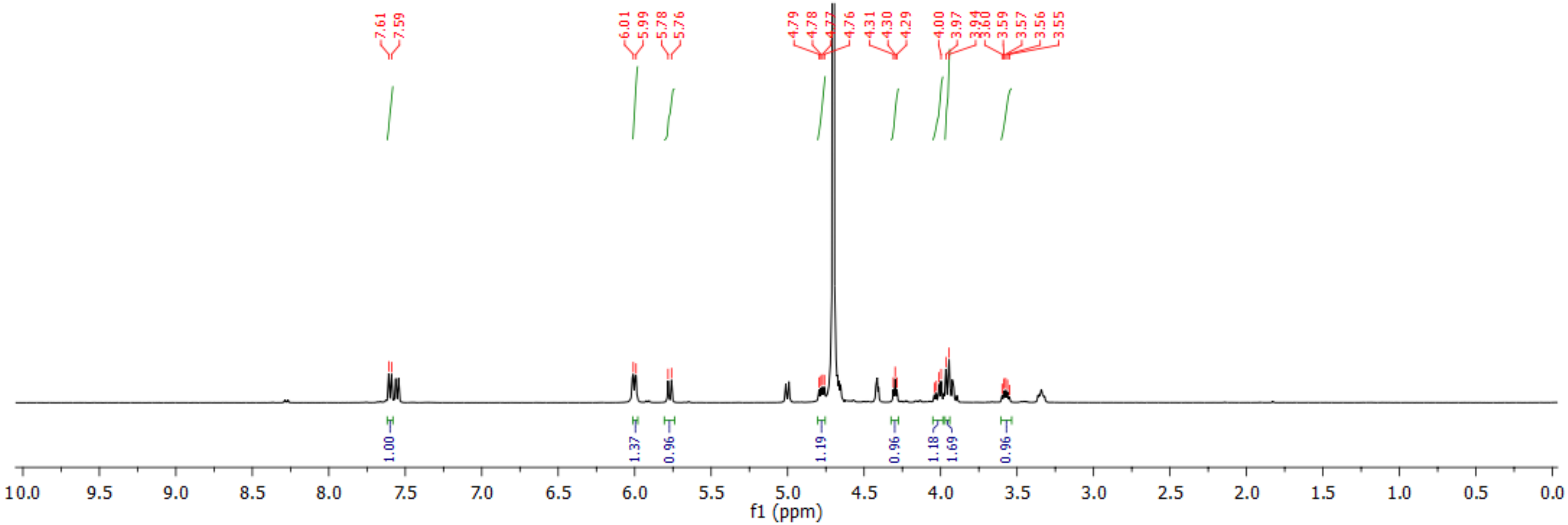




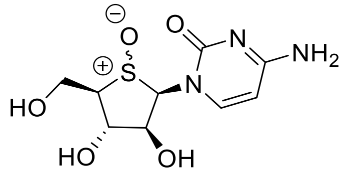
^{13}C NMR (101 MHz, D_2O)

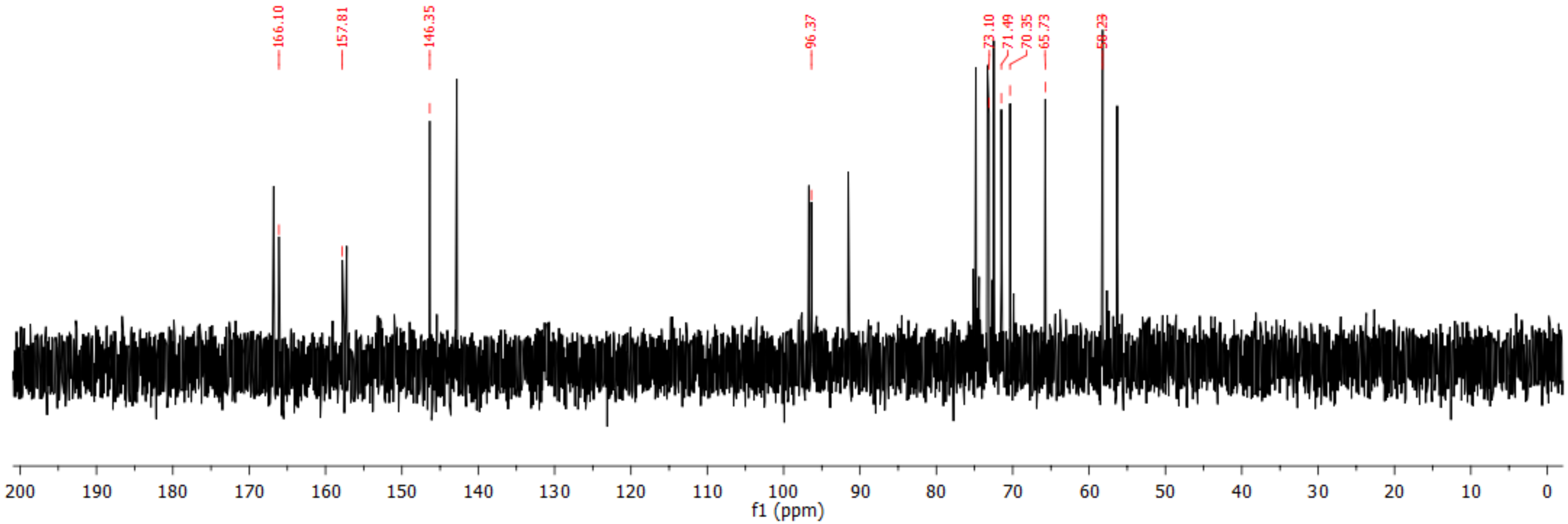
1'-(4'-sulfinyl(*S*/*R*)-D-arabinofuranosyl)cytosine **20**





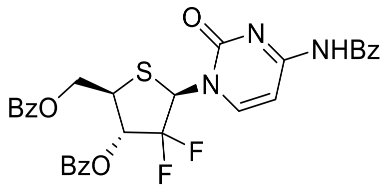
^{1}H NMR (400 MHz, D_2O)

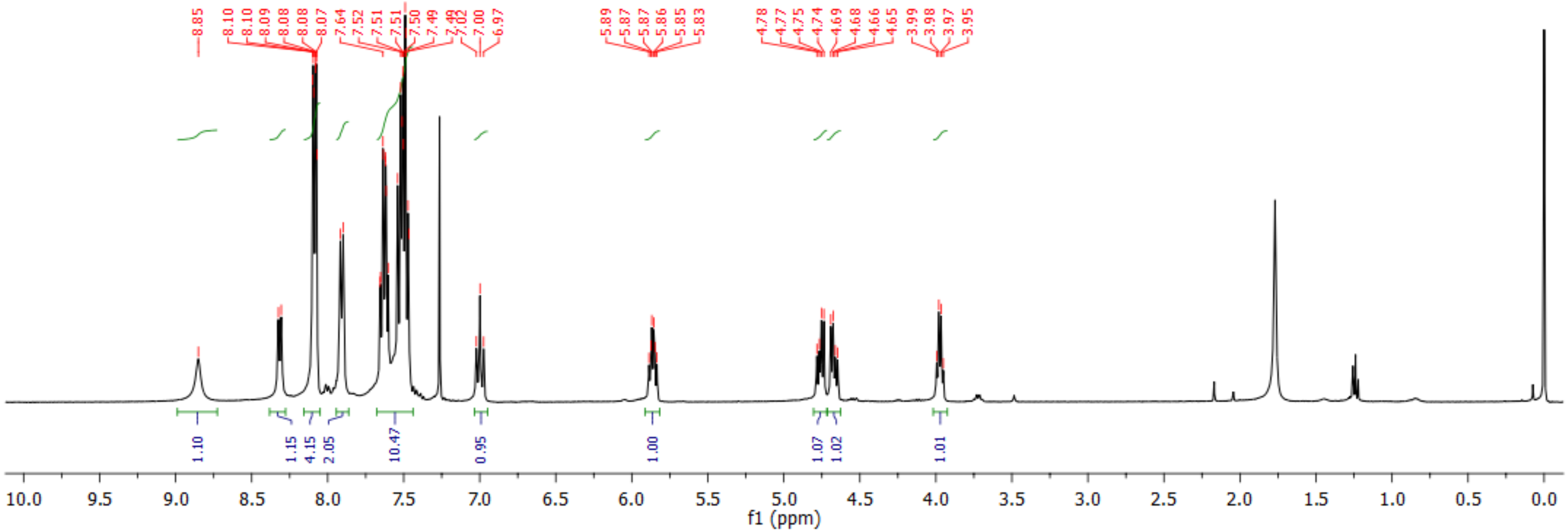


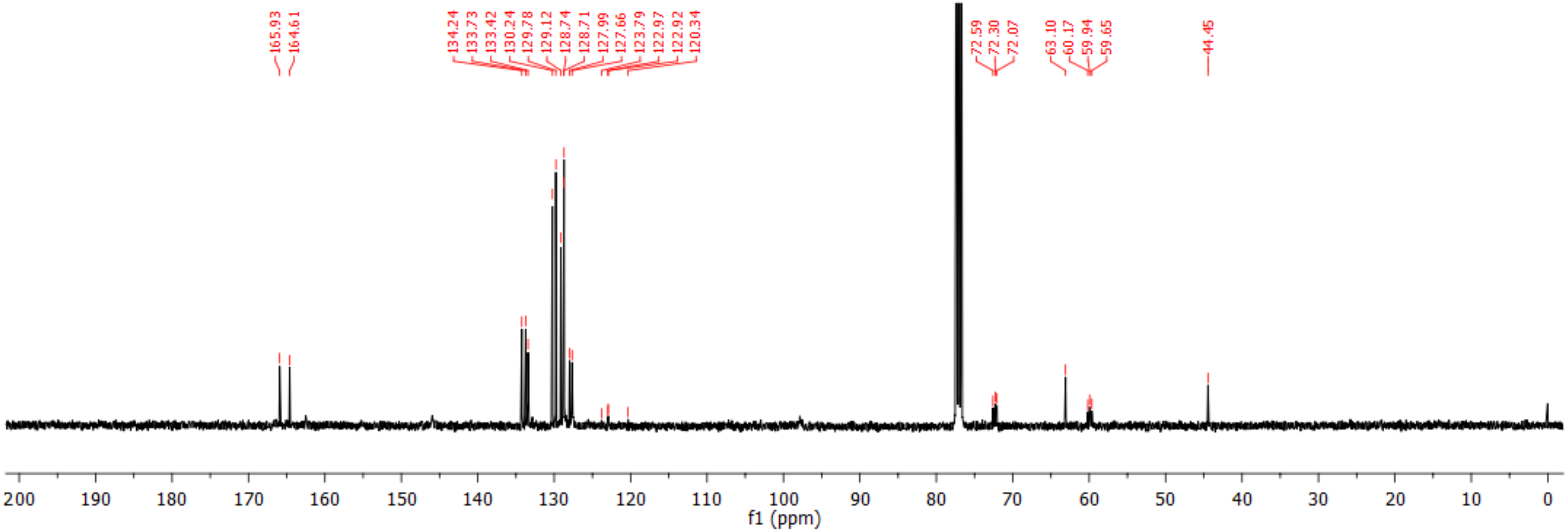


^{13}C NMR (101 MHz, D_2O)

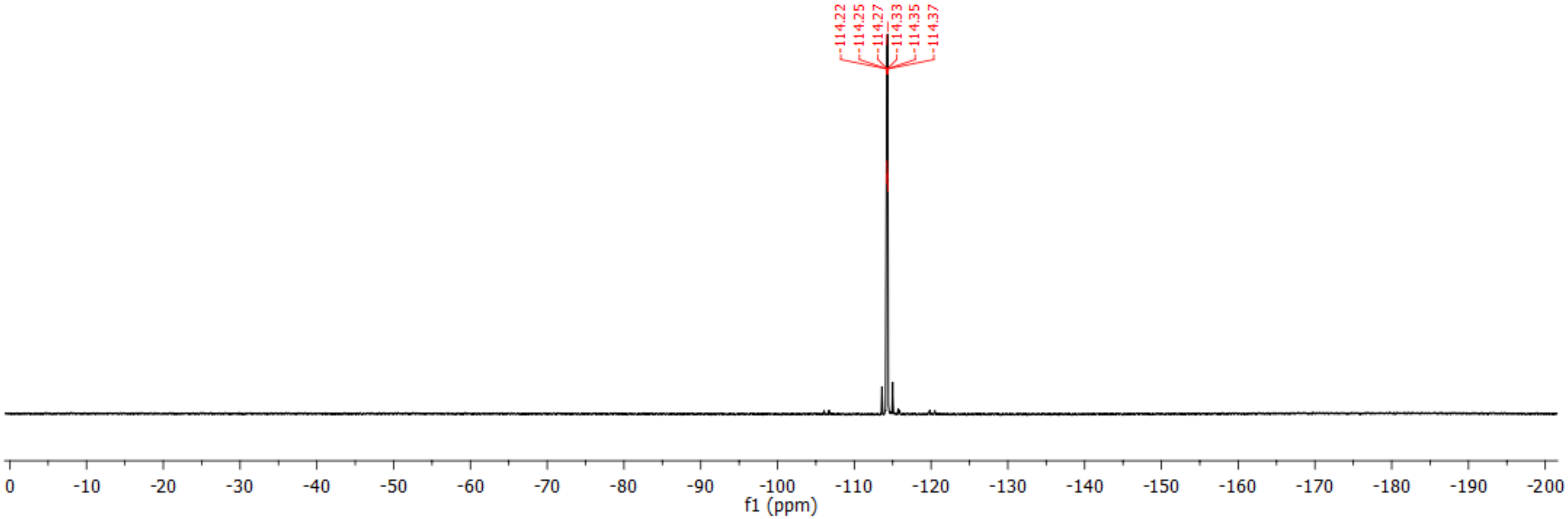
3’,5’-di-*O*-benzoyl-2’-deoxy-2’-*gem*-difluoro-1’-β-(4’-thio-D-ribofuranosyl)-*N*[4]-benzoyl-cytosine **21β**





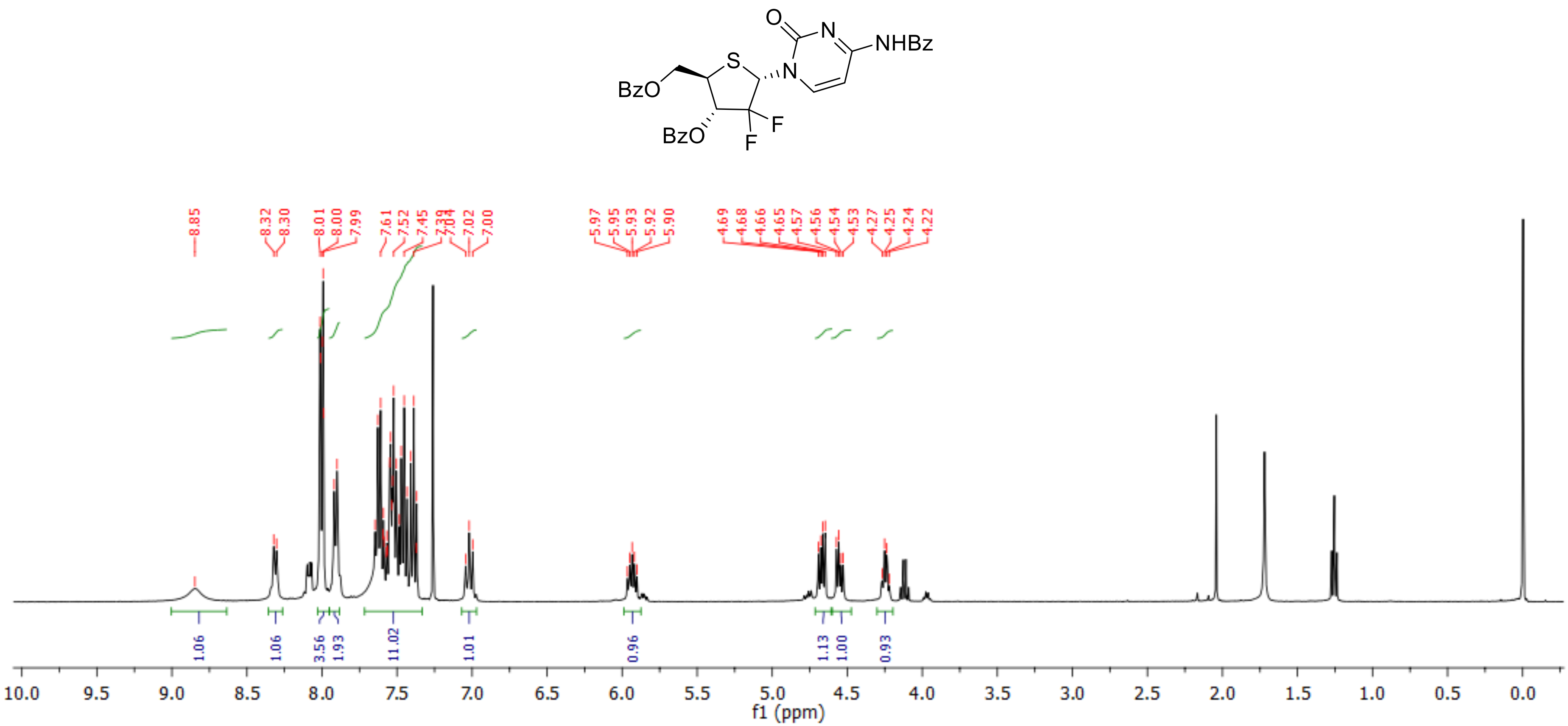


^{13}C NMR (101 MHz, $CDCl_3$)

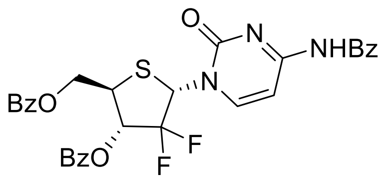


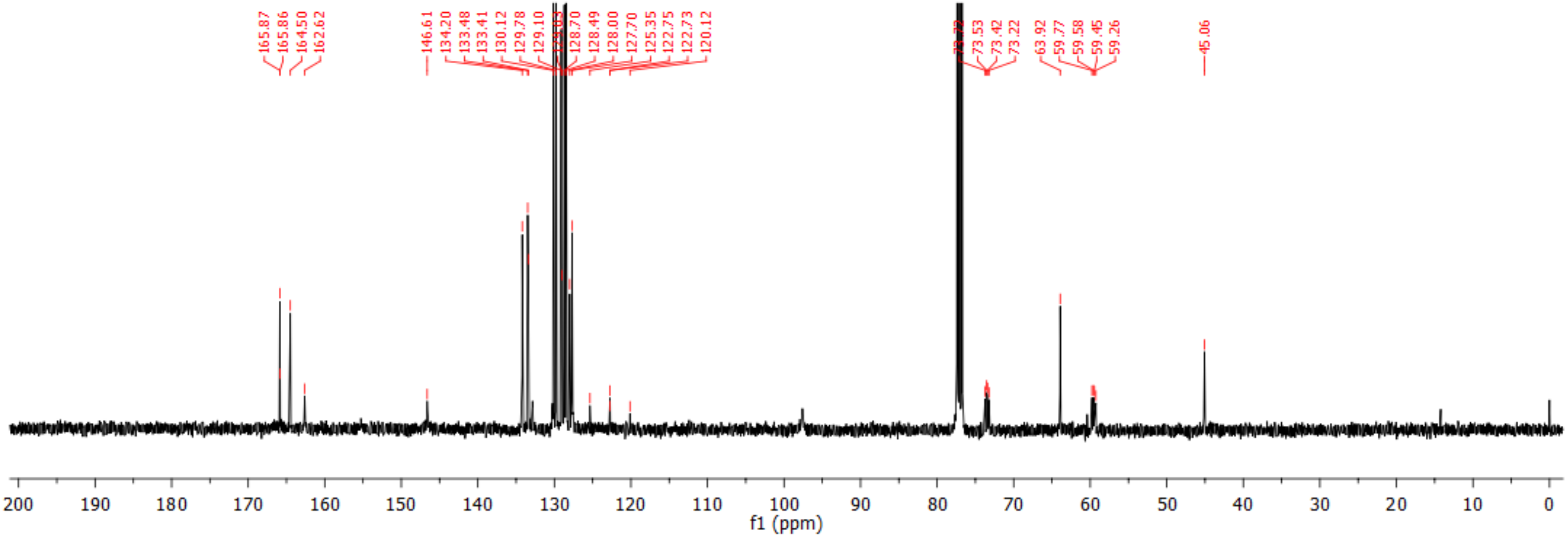
^{19}F NMR (376 MHz, $CDCl_3$)

3',5'-di-*O*-benzoyl-2'-deoxy-2'-*gem*-difluoro-1'-α-(4'-thio-D-ribofuranosyl)-*N*[4]-benzoyl-cytosine **21α**

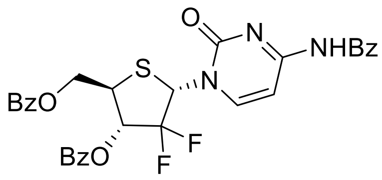


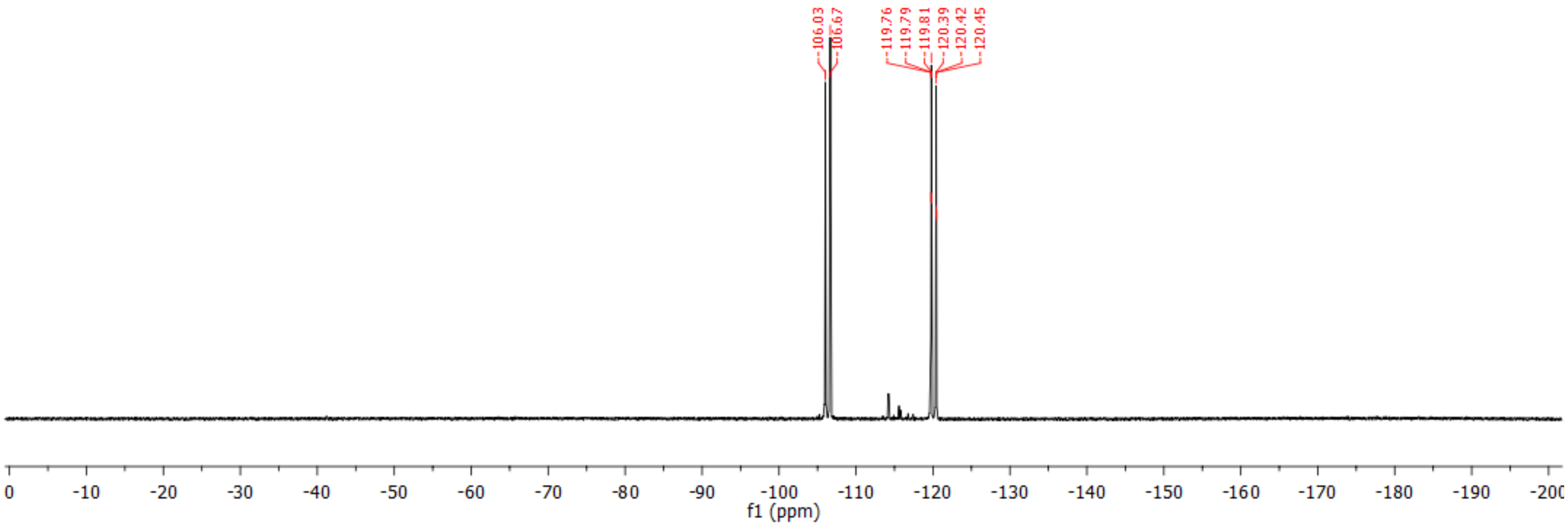
[1]H NMR (400 MHz, $CDCl_3$)





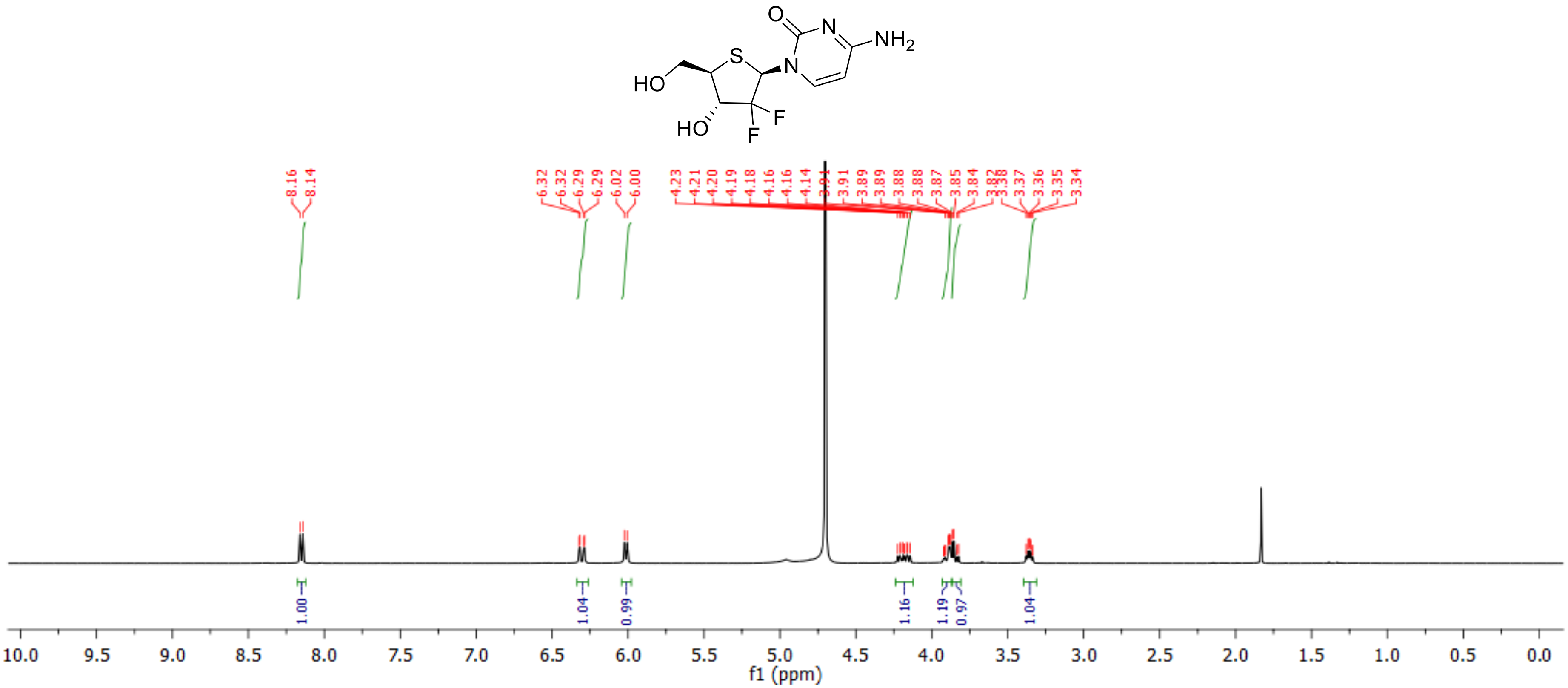
^{13}C NMR (101 MHz, $CDCl_3$)



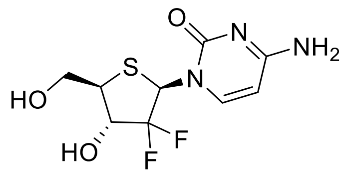


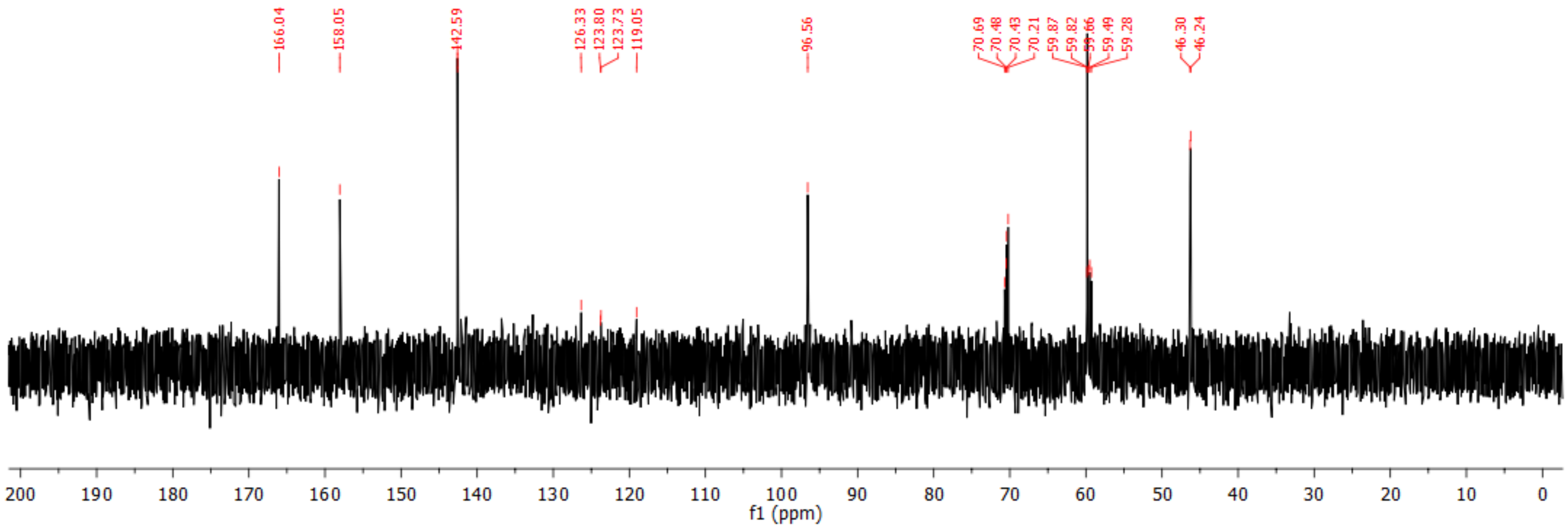
^{19}F NMR (376 MHz, $CDCl_3$)

2’-deoxy-2’-*gem*-difluoro-1’-β-(4’-thio-D-ribofuranosyl)cytosine **22-β**

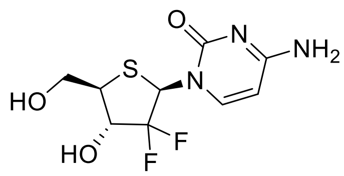


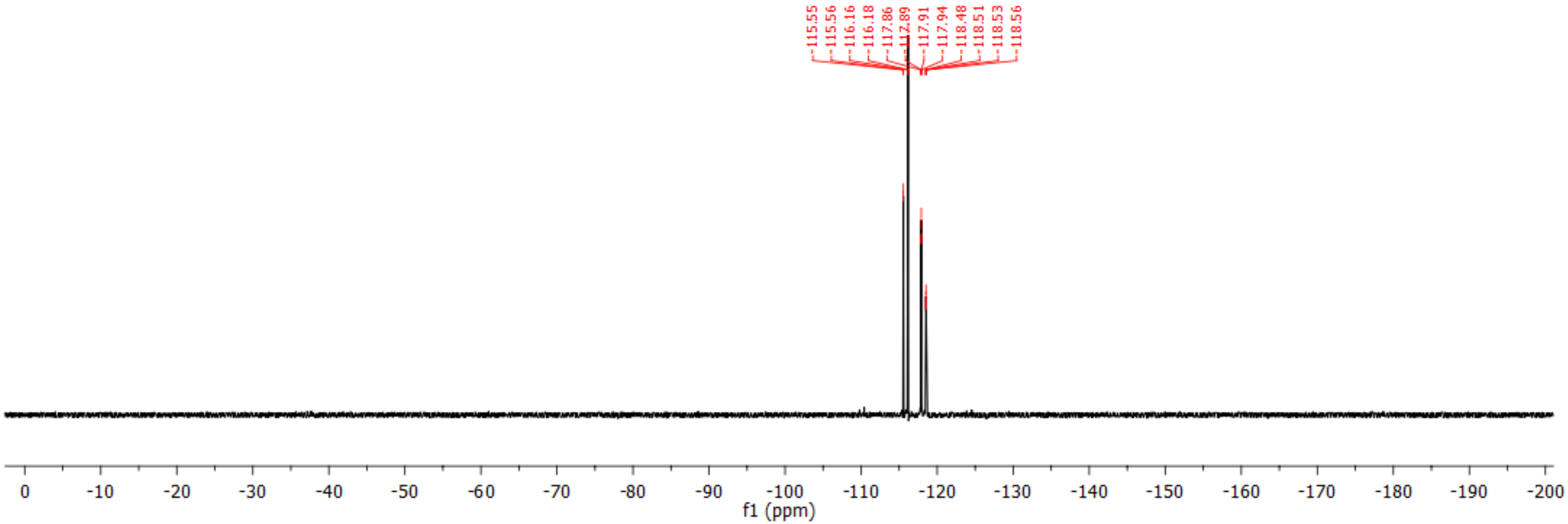
^{1}H NMR (400 MHz, D_2O)





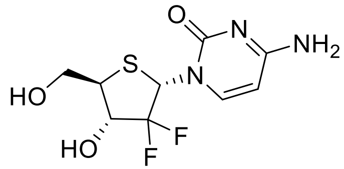
^{13}C NMR (101 MHz, D_2O)

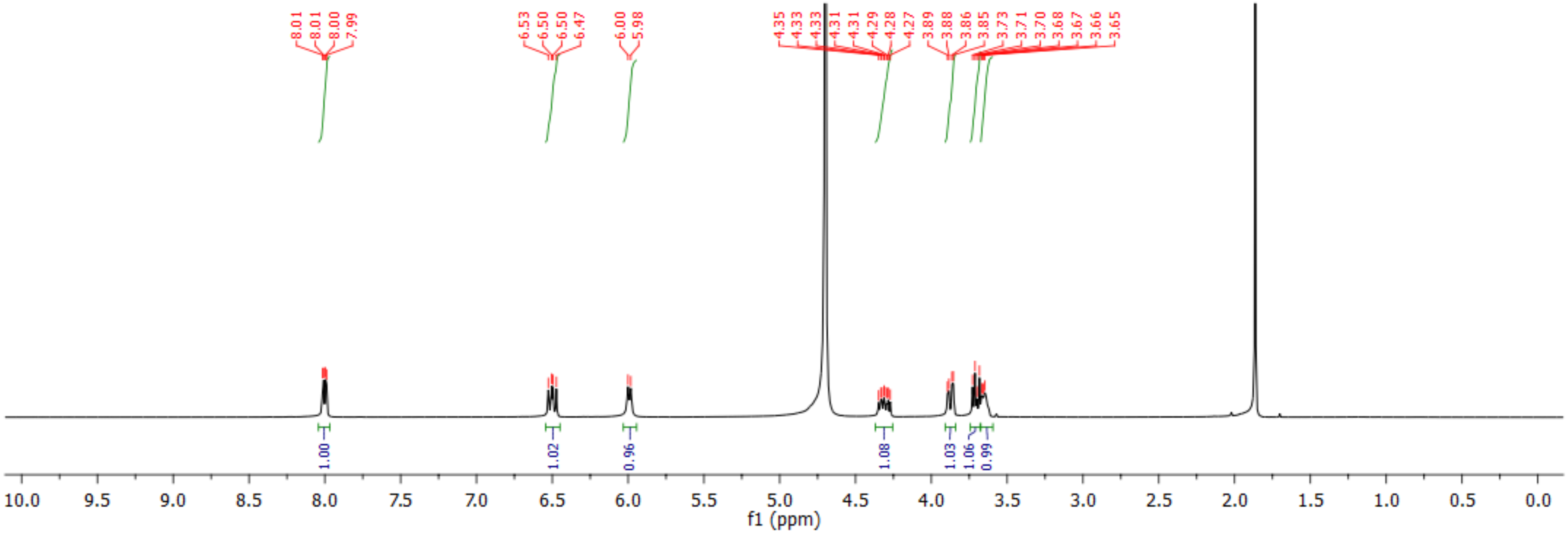




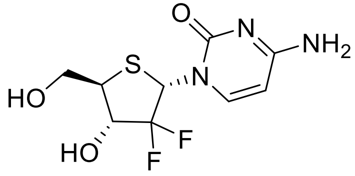
^{19}F NMR (376 MHz, D_2O)

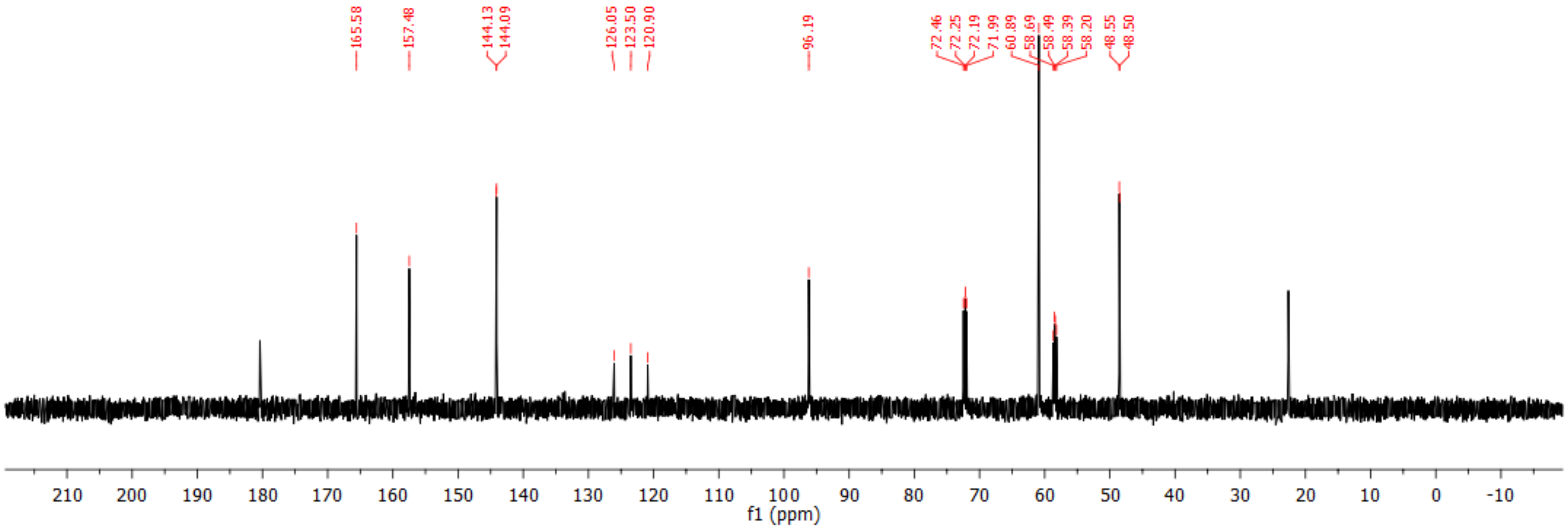
2'-deoxy-2'-*gem*-difluoro-1'-α-(4'-thio-D-ribofuranosyl)cytosine **22-α**



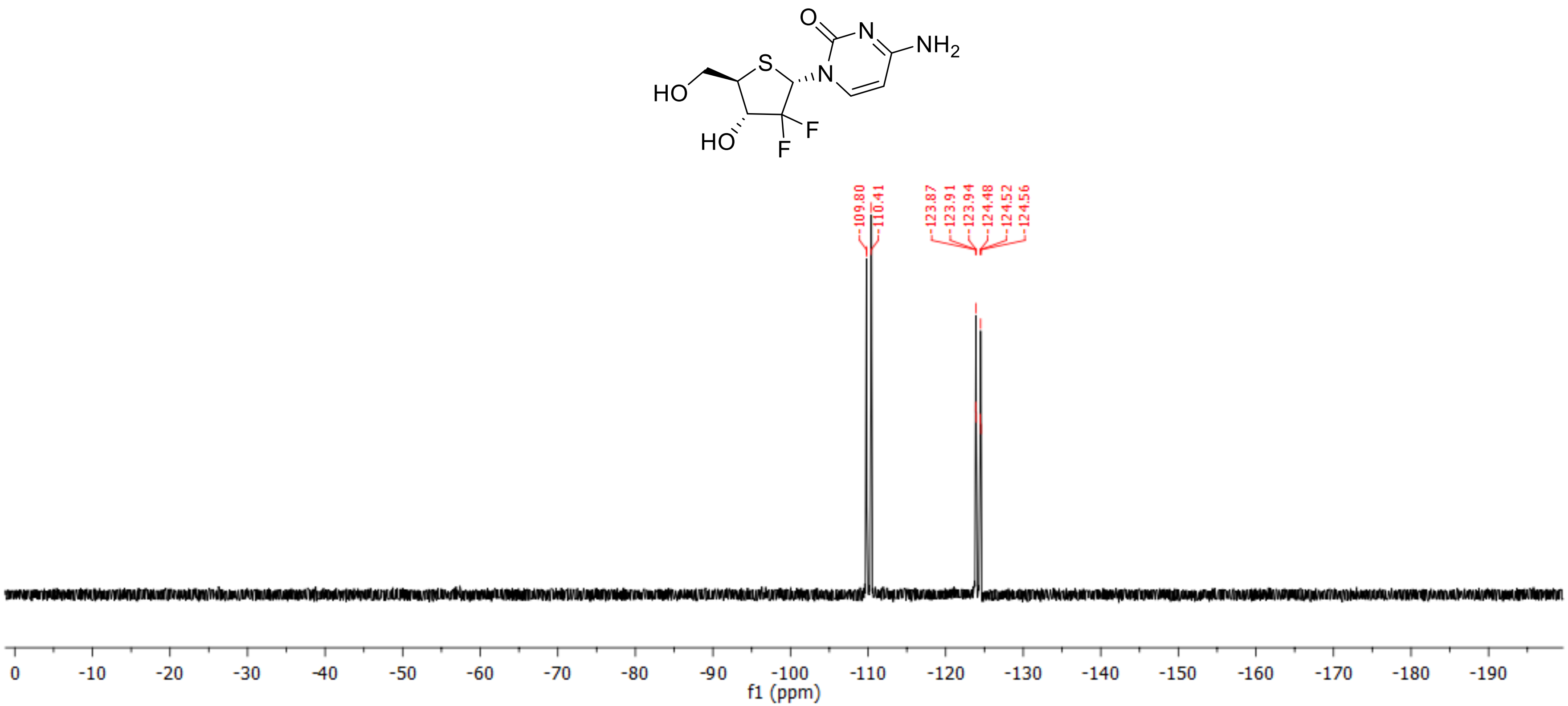


^{1}H NMR (400 MHz, D_2O)



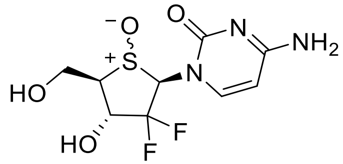


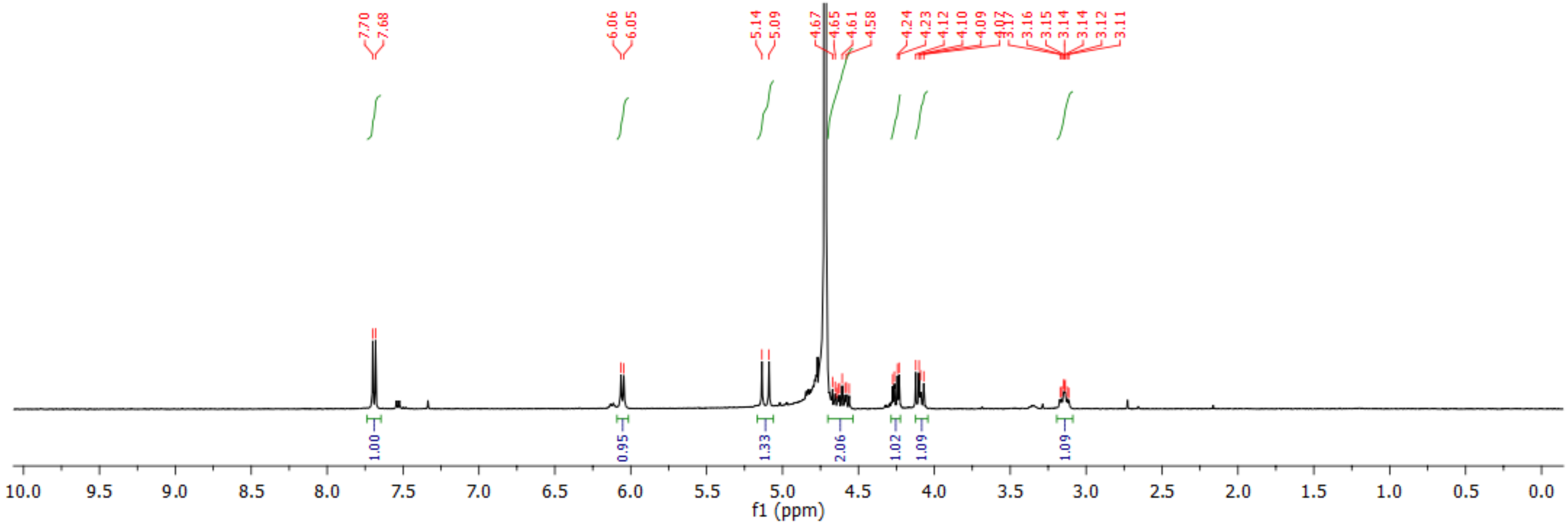
^{13}C NMR (101 MHz, D_2O)



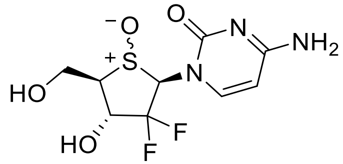
^{19}F NMR (376 MHz, D_2O)

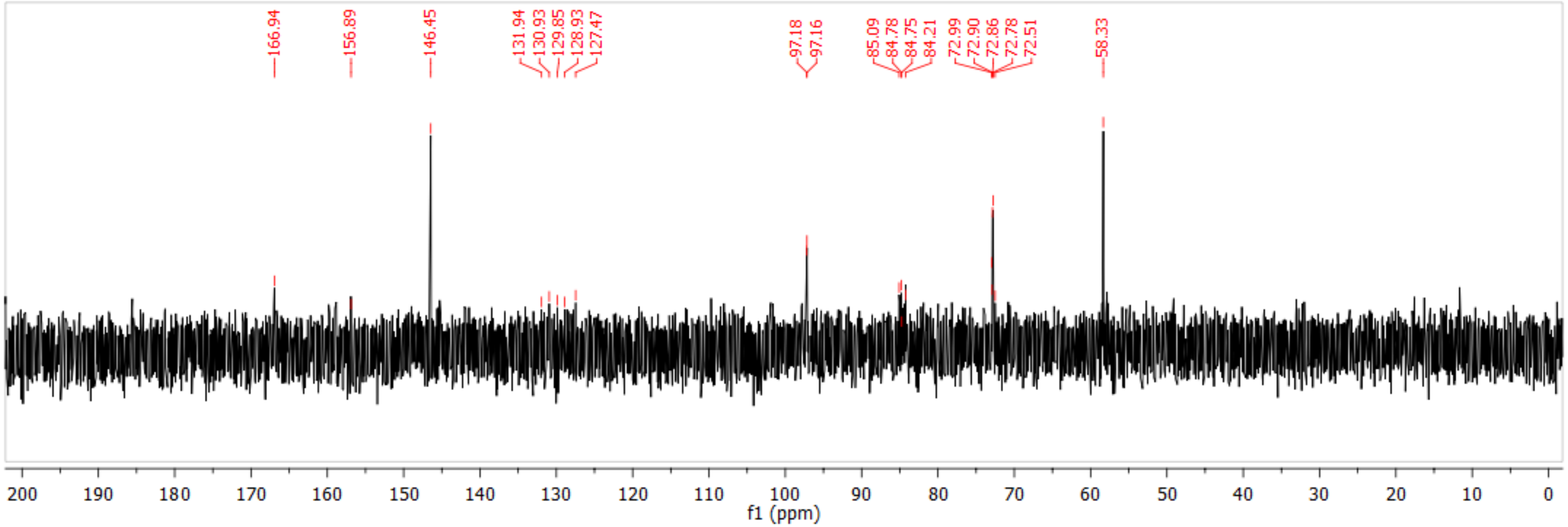
2'-deoxy-2'-*gem*-difluoro-1'-β-(4'-sulfinyl-D-ribofuranosyl)cytosine **23**



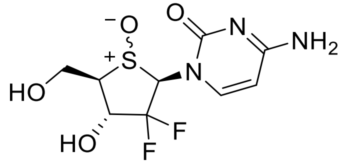


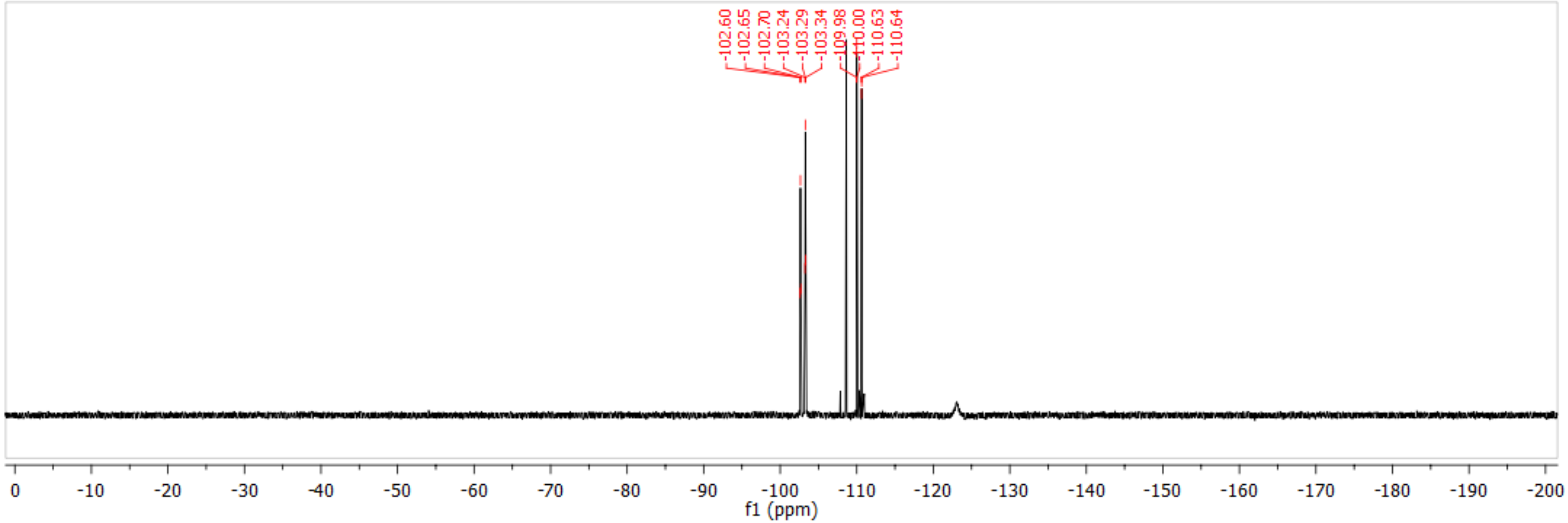
^{1}H NMR (400 MHz, D_2O)





^{13}C NMR (101 MHz, D_2O)





^{19}F NMR (376 MHz, D_2O)