Supporting Information

Exploring the Self-Assembly of Glycopeptides using Diphenylalanine Scaffold

Rinat Roytman[1],[#] Lihi Adler-Abramovich[2],[#] K. S. Ajish Kumar[1], Ting-Chun Kuan[3],
Chun-Cheng Lin[3], Ehud Gazit*[2] and Ashraf Brik*[1]

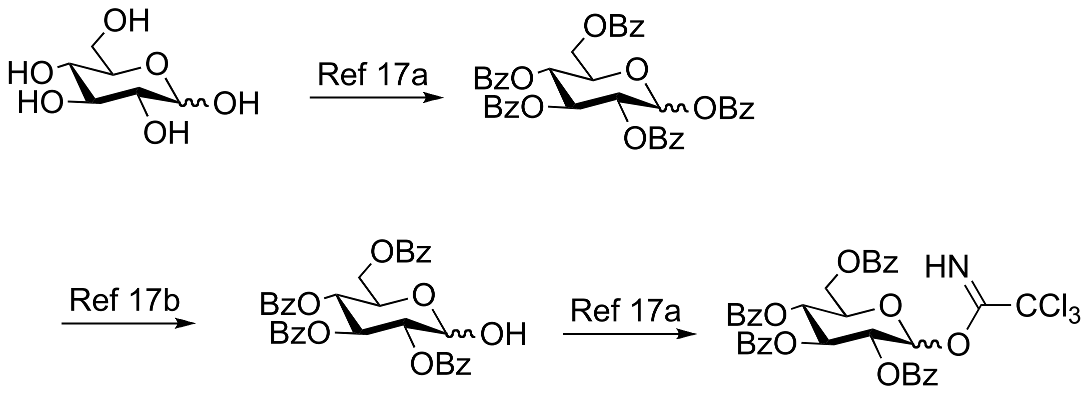
[1]*Department of Chemistry, Ben-Gurion University of Negev, Israel*
[2]*Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Israel*
[3]*Department of [3]Chemistry, National Tsing Hua University, Hsinchu, Taiwan*.

[#] contributed equally

E-mail: abrik@bgu.ac.il

E-mail: ehudg@post.tau.ac.il

Synthesis of compound 14

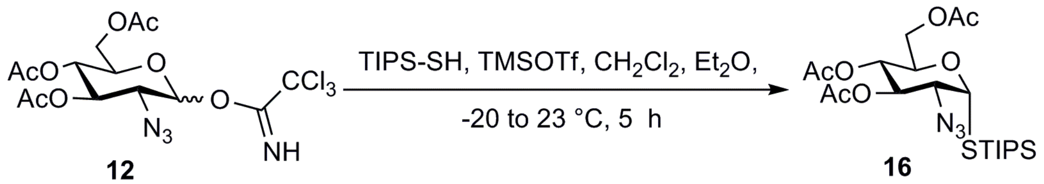


14

Compound **14** was synthesized according to reference 17.

Synthesis of α-glucopyranosyl-(1→3)-D-glucosamine thiol (10):

2-Azido-2-deoxy-3,4,6-tri-*O*-acetyl-*α*-D-glucopyranosyl-triisopropylsilylthiol (16):



12 **16**

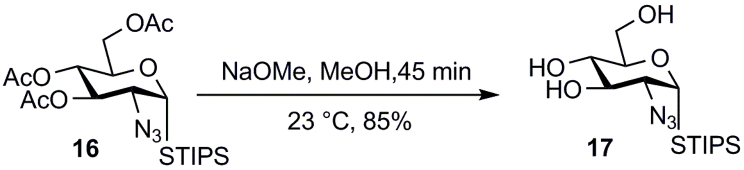
Trimethylsilyltrifluoromethane sulfonate (1.52 mL, 8.39 mmol) was added to a solution of trichloroacetimidate **12** (7.98 g, 16.77 mmol) and triisopropylsilane thiol (9.60 mL, 44.79 mmol) in 1: 1 mixture of dichloromethane and diethylether (120 mL) at -20 °C. The resulting mixture was stirred at this temperature for 3 h, then at 23 °C for 2 h. The reaction mixture was neutralized with N,N-diisopropylethylamine (0.74 mL), diluted with dichloromethane (650 mL), and washed with saturated aqueous sodium chloride solution (250 mL). The organic layer was dried (magnesium sulfate), filtered, and concentrated. The residue was purified by silica-gel flash chromatography (20% ethyl acetate in hexane) to afford **16** (6.55 g, 78%) as white solid. R*f* = 3.80 (40% ethyl acetate in hexane);

^{1}H NMR (500 MHz, $CDCl_3$) δ 5.46 (d, *J* = 5.0 Hz, 1H), 5.43 (dd, *J* = 10.6, 9.2 Hz, 1H), 5.01 (dd, *J* = 10.2, 9.1 Hz, 1H), 4.50 (ddd, *J* = 10.0, 4.2, 2.0 Hz, 1H), 4.26 – 4.23 (m, 1H), 3.95 (dd, *J* = 12.4, 2.1 Hz, 1H), 3.77 (dd, *J* = 10.6, 5.0 Hz, 1H), 2.02 (s, 3H), 2.00 (s, 3H), 1.97 (s, 3H), 1.26 – 1.21 (m, 3H), 1.07 (d, *J* = 7.4 Hz, 18H);

^{13}C NMR (126 MHz, $CDCl_3$) δ 170.8, 170.0, 169.9, 80.0, 71.5, 68.9, 68.6, 62.7, 62.2, 20.9, 20.8, 20.8, 18.5, 18.3, 12.9;

ESI-MS: Calculated for $C_{21}H_{37}N_3O_7SSi$: 503.21 Da [M^+], Observed: 525.20 Da [M+Na].

2-Azido-2-deoxy -*α*-D-glucopyranosyl-triisopropylsilylthiol (17):



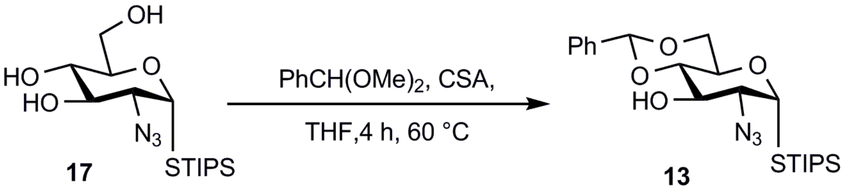
Fresh solution of sodium methoxide, prepared from sodium (120 mg, 5.50 mmol) in dry methanol (40 mL) was added to a pre cooled 0 °C solution of monosaccharide **16** (6.55 g, 13.04 mmol) in methanol (110 mL). The solution was stirred at this temperature for 45 min, and then concentrated in vacuo. The residue was purified by silica-gel flash chromatography (7% methanol in chloroform) to afford triol **17** (4.00 g, 81%) as colorless oil. R*f* = 0.32 (8% methanol in chloroform);

^{1}H NMR (500 MHz, CD_3OD) δ 5.52 (d, *J* = 4.9 Hz, 1H), 4.02 (dt, *J* = 10.4, 3.0 Hz, 1H), 3.86 (dd, *J* = 10.4, 8.8 Hz, 1H), 3.81 (dd, *J* = 12.0, 3.0 Hz, 1H), 3.71 (dd, *J* = 12.1, 3.0 Hz, 1H), 3.5-3.46 (m, 2H), 1.34 – 1.29 (m, 3H), 1.16 (d, *J* = 7.3 Hz, 18H);

^{13}C NMR (126 MHz, CD_3OD) δ 81.9, 74.5, 73.2, 71.6, 65.8, 61.8, 18.9, 18.8, 14.0;

ESI-MS: Calculated for $C_{15}H_{31}N_3O_4SSi$: 377.18 Da [M^+], Observed: 400.18 Da [M+Na].

2-Azo-4,6-*O*-benzylidene-2-deoxy-α-D-glucopyranosyl triisopropylsilane (13):



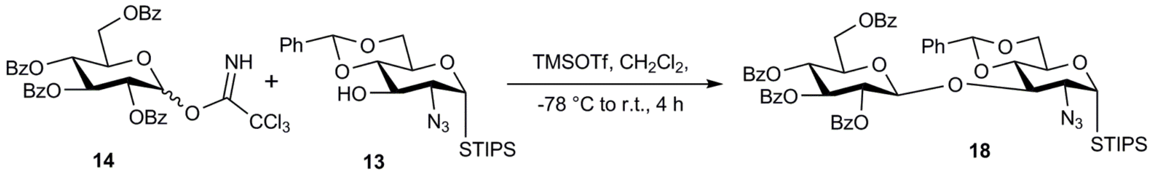
2-Azido-2-deoxy-*α*-D-glucopyranosyl-triisopropylsilanethiol (**17**) (2.83 g, 7.49 mmol) was dissolved in THF (30 mL) and treated with $PhCH(OMe)_2$ (2.26 mL, 14.99 mmol). The reaction mixture was refluxed for 4 h in the presence of CSA (cat). The reaction was quenched by the addition of saturated aqueous $NaHCO_3$ (50 mL), and the aqueous layer was extracted with EtOAc (3 x 20 mL). The combined organic layers were washed with brine and dried over $MgSO_4$, then concentrated. The residue obtained was purified by silica-gel flash chromatography (20% ethyl acetate in hexane) to give **13** as white solid (2.33 g, 67%). R*f* = 0.56 (20% ethyl acetate in hexane);

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.46 – 7.42 (m, 2H), 7.35 – 7.28 (m, 3H), 5.48 (s, 1H), 5.44 (d, *J* = 5.2 Hz, 1H), 4.33 (td, *J* = 10.0, 4.9 Hz, 1H), 4.18 – 4.13 (m, 2H), 3.67 (t, *J* = 9.5 Hz, 1H), 3.63 (dd, *J* = 10.0, 5.2 Hz, 1H), 3.45 (t, *J* = 9.4 Hz, 1H), 2.61 (d, *J* = 2.6 Hz,1H), 1.26 – 1.20 (m, 3H), 1.07 (dd, *J* = 7.4, 2.3 Hz, 18H);

^{13}C NMR (126 MHz, $CDCl_3$) δ 137.0, 129.5, 128.5, 126.4, 102.3, 82.2, 81.0, 69.6, 68.8, 64.9, 63.4, 18.6, 18.3, 12.9;

ESI-MS: Calculated for $C_{22}H_{35}N_3O_4SSi$: 465.21 Da [M^+], Observed: 487.2 Da [M+Na].

Disaccharide (18):



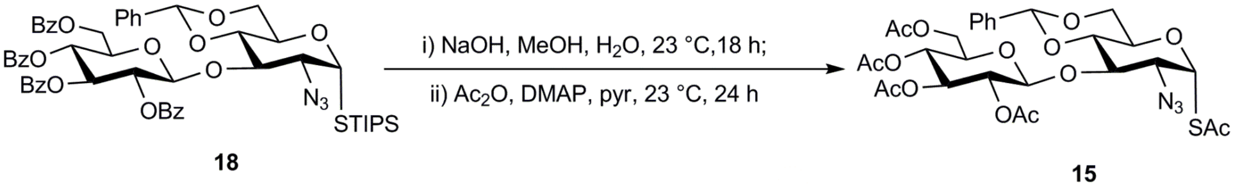
To a solution of compound **13** (0.80 g, 1.72 mmol) and **14** (1.53 g, 2.57 mmol) in anhydrous CH_2Cl_2 (20 mL) containing 4Å MS was added TMSOTf (19.75 µL) at -78 °C and stirred for 2 h at this temperature. Another portion of TMSOTf (39.51 µL) was added and the solution was stirred for additional 2 h at room temperature before the reaction was quenched by the addition of Et_3N. The mixture was filtered and washed with CH_2Cl_2. The filtrate was concentrated to give a residue that was purified by silica gel flash column (gradient elution: 10% to 30% ethyl acetate in hexane) to afford **18** (1.64 g, 91%) as a white foamy solid. R*f* = 0.40 (20% ethyl acetate in hexane);

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.92 –789 (m, 4H), 7.79 – 7.59 (m, 4H), 7.46 – 7.38 (m, 5H), 7.36 – 7.19 (m, 12H), 5.78 (t, *J* = 9.6 Hz, 1H), 5.60 (t, *J* = 9.7 Hz, 1H), 5.53 (s, 1H), 5.52 (dd, *J* = 8.0, 1.8 Hz, 1H), 5.37 (d, *J* = 5.3 Hz, 1H), 5.09 (d, *J* = 7.9 Hz, 1H), 4.39 (dd, *J* = 12.1, 3.4 Hz, 1H), 4.33 (td, *J* = 9.8, 4.5 Hz, 1H), 4.23 (dd, *J* = 12.1, 4.5 Hz, 1H), 4.14 – 4.10 (m, 2H), 3.87 (dt, *J* = 9.8, 4.0 Hz, 1H), 3.70 – 3.62 (m, 3H), 1.19 – 1.16 (m, 3H), 1.01 (dd, *J* = 7.4, 2.4 Hz, 18H);

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.2, 165.9, 165.4, 165.3, 137.2, 133.5, 133.3, 133.2, 133.1, 129.9, 129.9 (strong), 129.9, 129.7, 129.5, 129.3, 129.0, 128.5, 128.4 (strong), 128.3, 126.2, 101.7, 101.3, 80.8, 80.5, 80.4, 73.3, 72.4, 72.3, 69.8, 68.9, 64.7, 63.9, 63.1, 18.6, 18.4, 12.9;

ESI-MS: Calculated for $C_{56}H_{61}N_3O_1SSi$: 1044.25 Da [M^+], Observed: 1065.6 Da [M+Na].

Disaccharide (15):



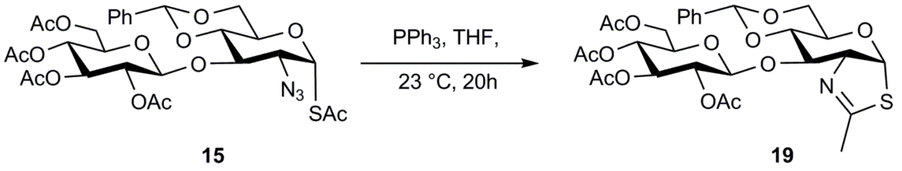
Freshly prepared sodium methoxide from sodium (46 mg, 2.02 mmol,) in dry methanol (10 mL) was added to a solution of disaccharide **18** (1.76 g, 1.68 mmol) in methanol (31 mL). The solution was deoxygenated, and stirred at 23 °C for 18 h and then concentrated in vacuo. The crude product was then suspended in pyridine (15 mL) and acetic anhydride (1.10 mL, 11.87 mmol) and deoxygenated. A catalytic amount of DMAP was added, and the resulting solution stirred at 23 °C for 24 h and then concentrated in vacuo. The residue was purified by silica-gel flash chromatography (gradient elution: 10% to 40% ethyl acetate in hexane) to afford the disaccharide **15** as major product (α : β = 4:1 (From ^{1}H-NMR)) (0.59 g, 51%). R*f* = 0.69 (50% ethyl acetate in hexane). Only peaks due to α product are shown;

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.37 (m, 2H), 7.32 – 7.30 (m, 3H), 6.04 (d, *J* = 5.7 Hz, 1H), 5.49 (s, 1H), 5.07 (t, *J* = 9.4 Hz, 1H), 5.01 (t, *J* = 9.5 Hz, 1H), 4.96 (dd, *J* = 9.2, 8.2 Hz, 1H), 4.74 (d, *J* = 8.0 Hz, 1H), 4.18 (dd, *J* = 6.9, 2.9 Hz, 1H), 4.00 (dd, *J* = 12.4, 4.0 Hz, 1H), 3.93 – 3.90 (m, 1H), 3.81 (dd, *J* = 12.4, 2.3 Hz, 1H), 3.69 – 3.65 (m, 4H), 3.37 – 3.33 (m, 1H), 2.39 (s, 3H), 2.01 (s, 3H),1.93 (s, 3H), 1.92 (s, 3H), 1.91 (s, 3H);

^{13}C NMR (126 MHz, $CDCl_3$) δ 190.8, 169.8, 169.5, 168.5, 168.4, 135.9, 128.5, 127.5, 125.1, 125.2, 100.8, 99.8, 81.6, 79.2, 78.2, 72.1, 71.2, 71.1, 67.5, 66.9, 66.1, 62.0, 60.6, 30.6, 19.9, 19.8, 19.7, 19.6;

ESI-MS: Calculated for $C_{29}H_{35}N_3O_{14}S$: 681.66 Da [M$^+$], Observed: 704.09 Da [M+Na].

Disaccharide (19):



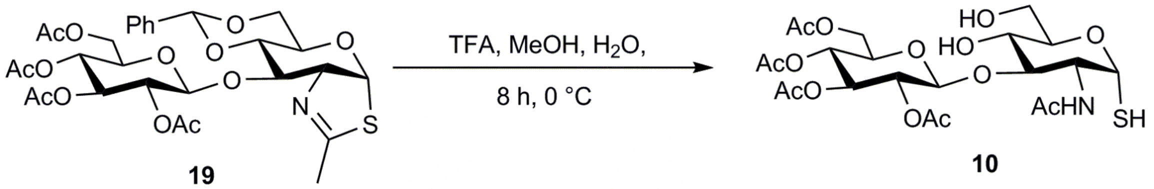
Triphenylphosphine (390 mg, 1.49 mmol) was added to a solution of disaccharide **15** (676 mg, 0.99 mmol) in THF (20 mL). The reaction mixture was heated and maintained at 50 °C for 20 h, and then concentrated. The residue was purified by silica gel flash column chromatography (gradient elution: 20% to 40% ethyl acetate in hexane) to afford the thiazoline **19** (90 mg, 53%) as colorless syrup. R*f* = 0.37 (50% ethyl acetate in hexane);

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.42 – 7.39 (m, 2H), 7.29 – 7.26 (m, 3H), 6.31 (d, *J* = 7.3 Hz, 1H), 5.52 (s, 1H), 5.13 (t, *J* = 9.5 Hz, 1H), 5.03 (t, *J* = 9.7 Hz, 1H), 4.97 (dd, *J* = 9.5, 8.0 Hz, 1H), 4.71 (d, *J* = 8.0 Hz, 1H), 4.26 – 4.20 (m, 2H), 4.15 – 4.13 (m, 2H), 3.87 (dd, *J* = 12.3, 2.2 Hz, 1H), 3.74 (dd, *J* = 9.3, 7.3 Hz, 1H), 3.68 (dt, *J* = 7.8, 4.0 Hz, 1H), 3.62 (t, *J* = 9.8 Hz, 1H), 3.49 – 3.45 (m, 1H), 2.17 (d, *J* = 1.2 Hz, 3H), 2.01 (s, 3H), 1.94 (s, 3H), 1.93 (s, 3H), 1.92 (s, 3H);

^{13}C NMR (126 MHz, CDCl$_3$) δ 170.8, 170.4, 169.7, 169.5, 167.7, 137.3, 129.3, 128.4, 126.2, 101.6, 101.4, 92.1, 81.8, 79.1, 77.9, 72.9, 71.9, 71.7, 68.9, 68.3, 64.3, 61.8, 21.9, 21.0, 20.9, 20.8, 20.7;

ESI-MS: Calculated for $C_{29}H_{35}N_3O_{13}S$: 637.19 Da [M$^+$], Observed: 659. 09 Da [M+Na].

α -glucopyranosyl-(1→3)-D-glucosamine thiol (10):



19 **10**

Trifluoroacetic acid (0.67 mL, 9.11 mmol) was added to a solution of thiazoline **19** (306 mg, 0.48 mmol) in methanol (15 mL) containing water (1.29 mL). The resulting solution was deoxygenated, and stirred at 0 °C for 8 h. The reaction mixture was diluted with toluene (15 mL) and then concentrated. Residual TFA was removed from the product by azeotropic coevaporation (2 x 15 mL methanol; 1 x 15 mL toluene), followed by lyophylization (30 mL water). The residue was purified by silica gel flash column chromatography (10% methanol in chloroform) to afford **10** (239 mg, 88%), as a white solid. R*f* = 0.83 (30% methanol in chloroform);

^{1}H NMR (500 MHz, CDCl$_3$) δ 5.64 (d, *J* = 4.8 Hz, 1H), 5.63 – 5.56 (m, 1H), 5.16 (t, *J* = 9.5 Hz, 1H), 5.03 – 4.97 (m 1H), 4.91 (dd, *J* = 9.6, 7.9 Hz, 1H), 4.63 (d, *J* = 7.9 Hz, 1H), 4.33 – 4.29 (m, 1H), 4.18 (dd, *J* = 12.3, 2.5 Hz, 1H), 3.91 – 3.88 (m, 1H), 3.83 (d, *J* = 12.1 Hz, 1H), 3.77 (m, 1H), 3.58 – 3.53 (m, 3H), 3.45 (dd, *J* = 10.6, 2.3 Hz, 1H), 2.04 (s, *J* = 2.2 Hz, 3H), 2.03 (d, *J* = 3.3 Hz, 1H), 2.02 (s, 1H), 2.00 (s, 3H), 1.98 (s, 3H), 1.97 (s 1H), 1.96 (s, 3H), 1.95 (s, 3H);

^{13}C NMR (126 MHz, $CDCl_3$) δ 170.7, 170.4, 170.0, 169.5, 169.4, 101.0, 82.7, 79.2, 73.0, 72.5, 72.2, 71.6, 69.5, 68.3, 62.5, 61.9, 52.2, 23.6, 20.8, 20.7, 20.6 (strong);

ESI-MS: Calculated for $C_{22}H_{33}NO_{14}S$: 567.17 Da [M^+], Observed: 589.2 Da [M+Na].

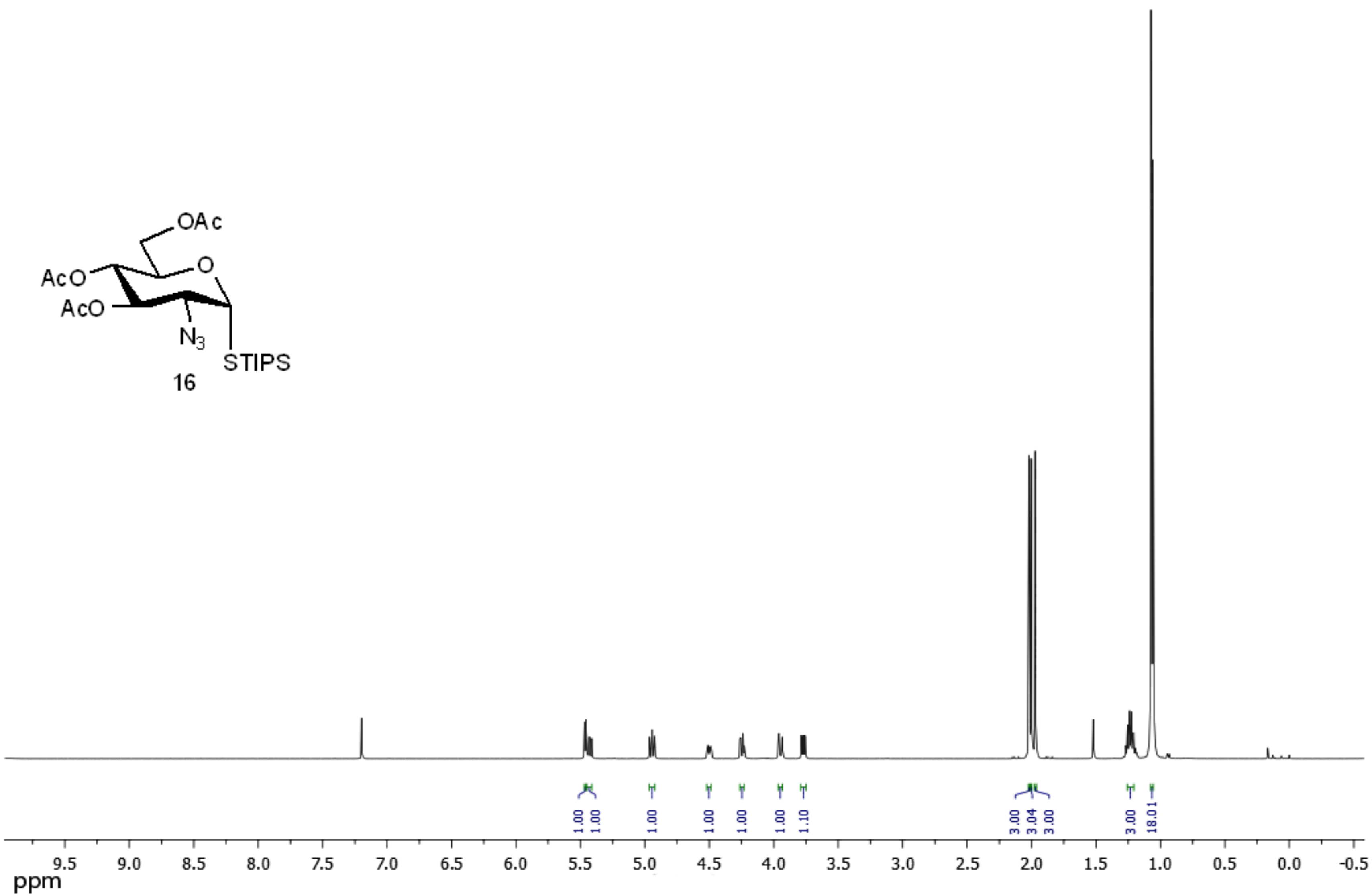


Figure1:^{1}H NMR (500 MHz, $CDCl_3$) spectrum of compound 16

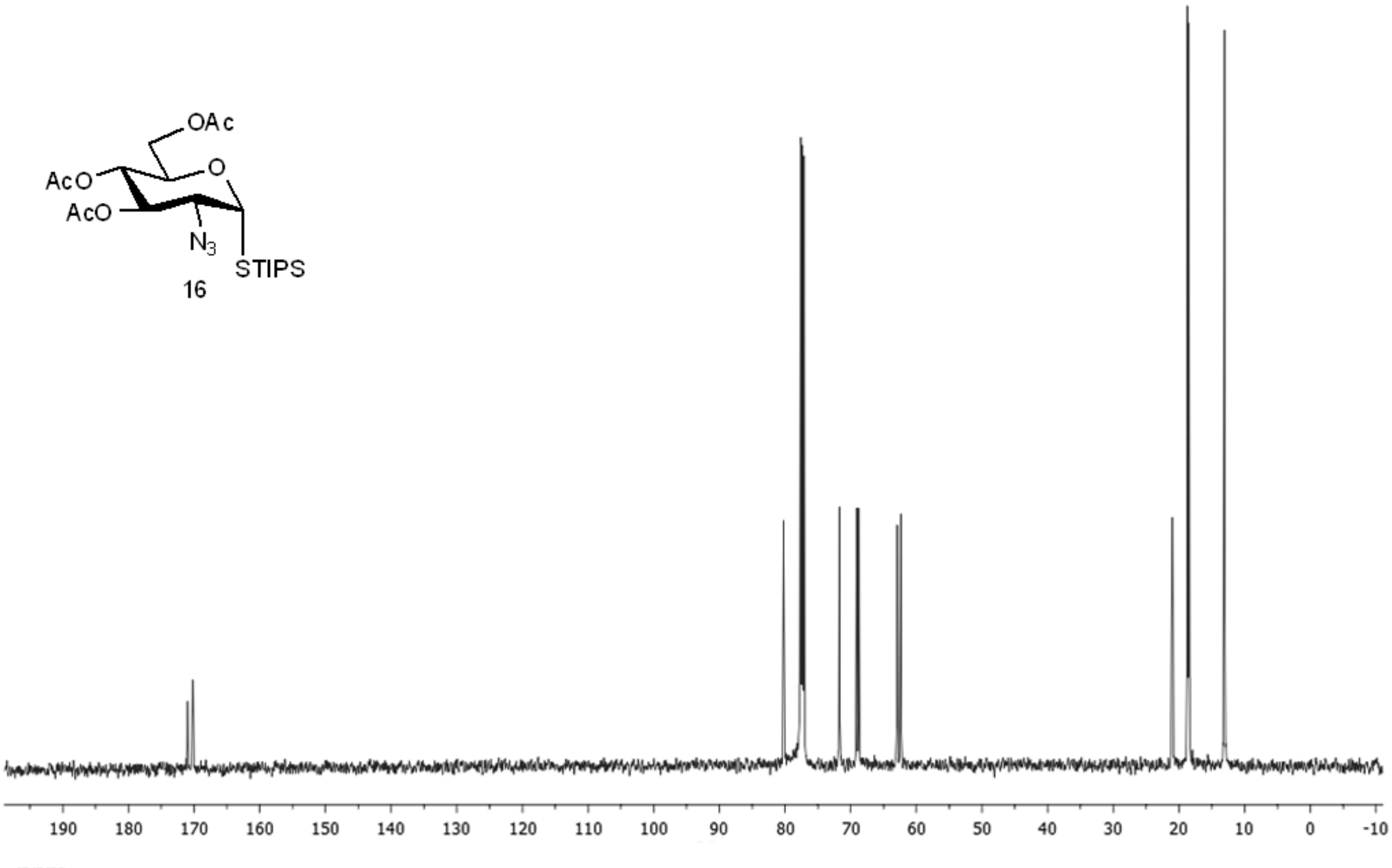


Figure2: ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 16

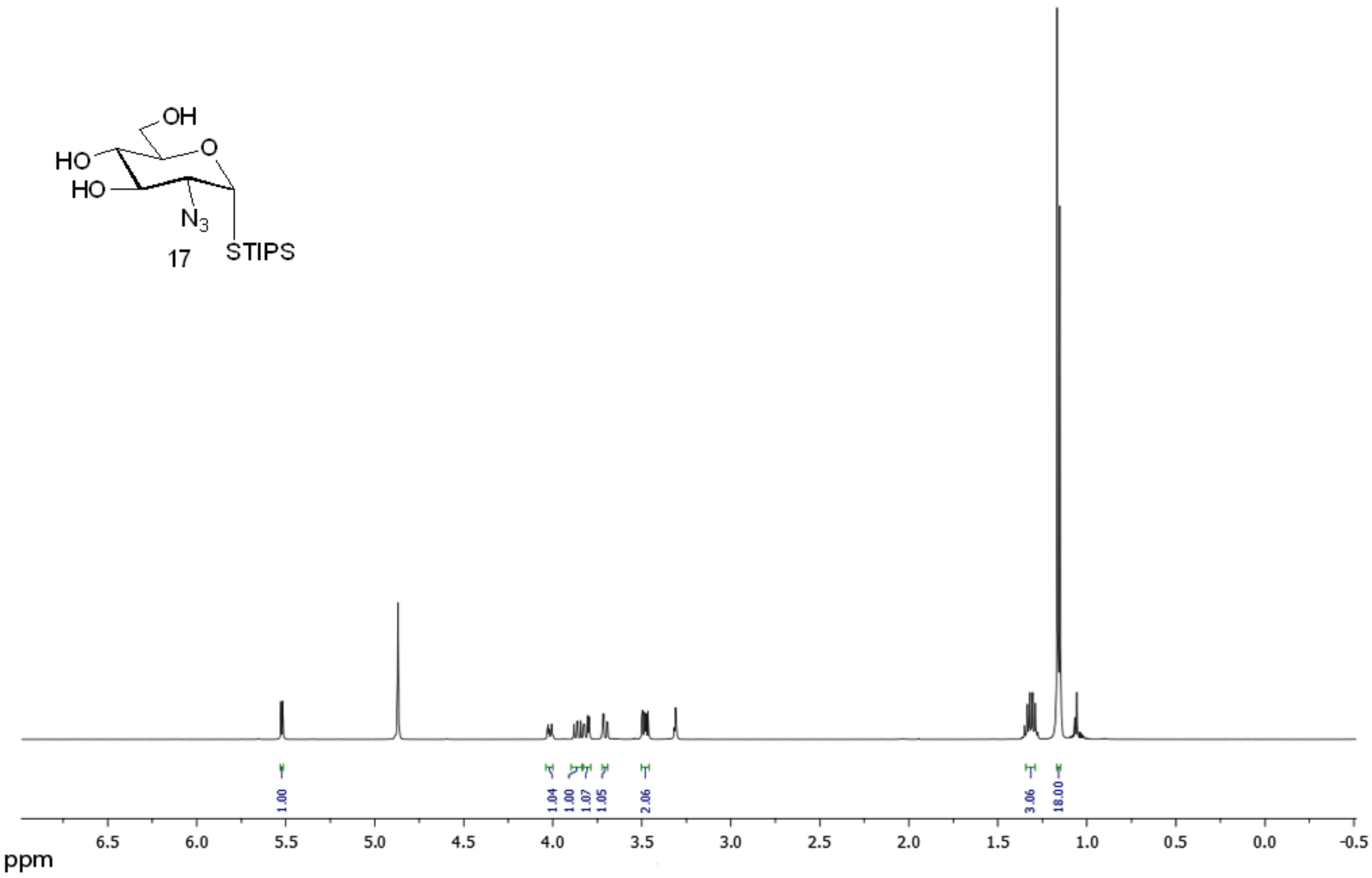


Figure3: ^{1}H NMR (500 MHz, CD_3OD) spectrum of compound 17

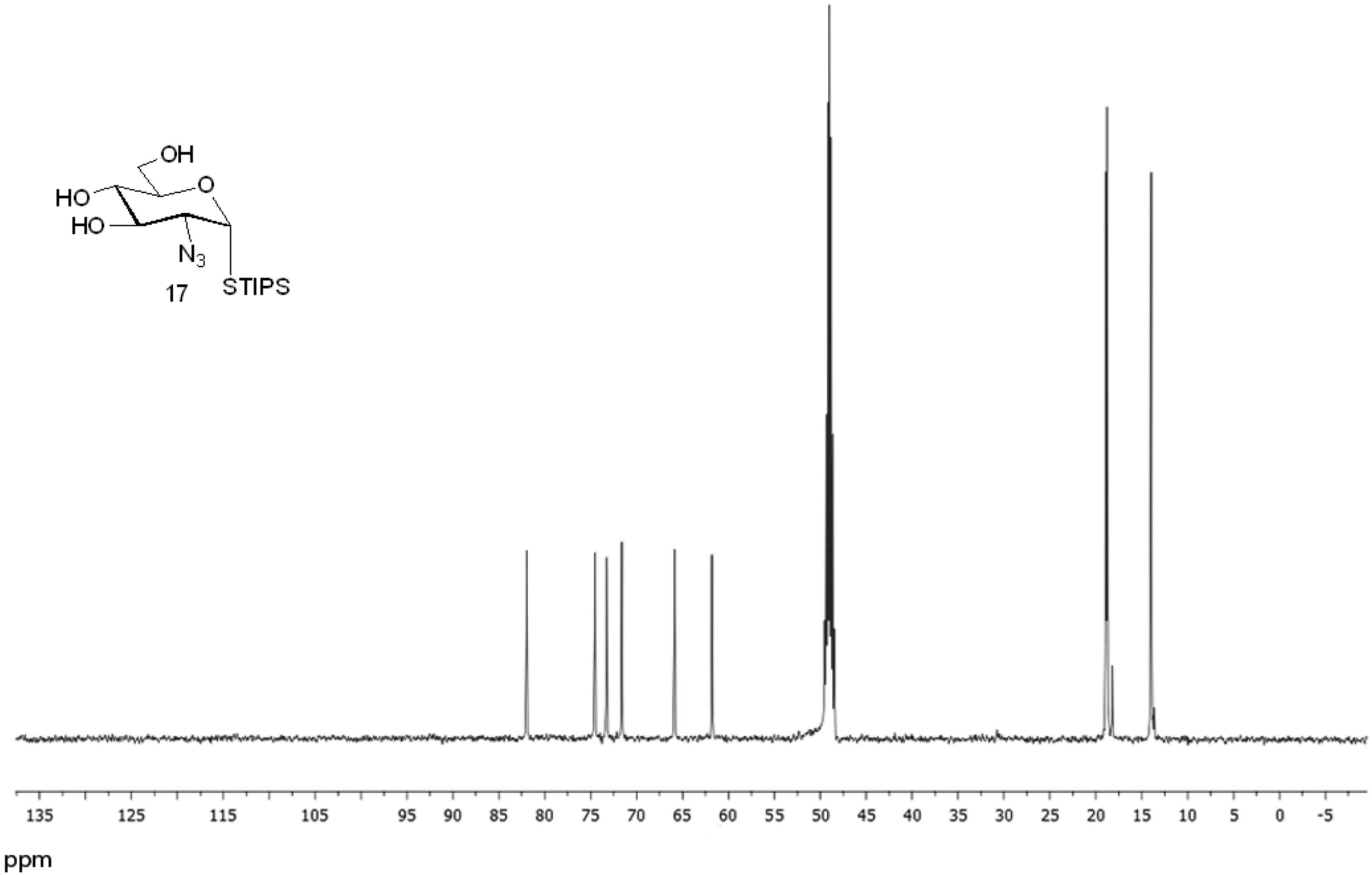


Figure4: ^{13}C NMR (126 MHz, CD_3OD) spectrum of compound 17

13

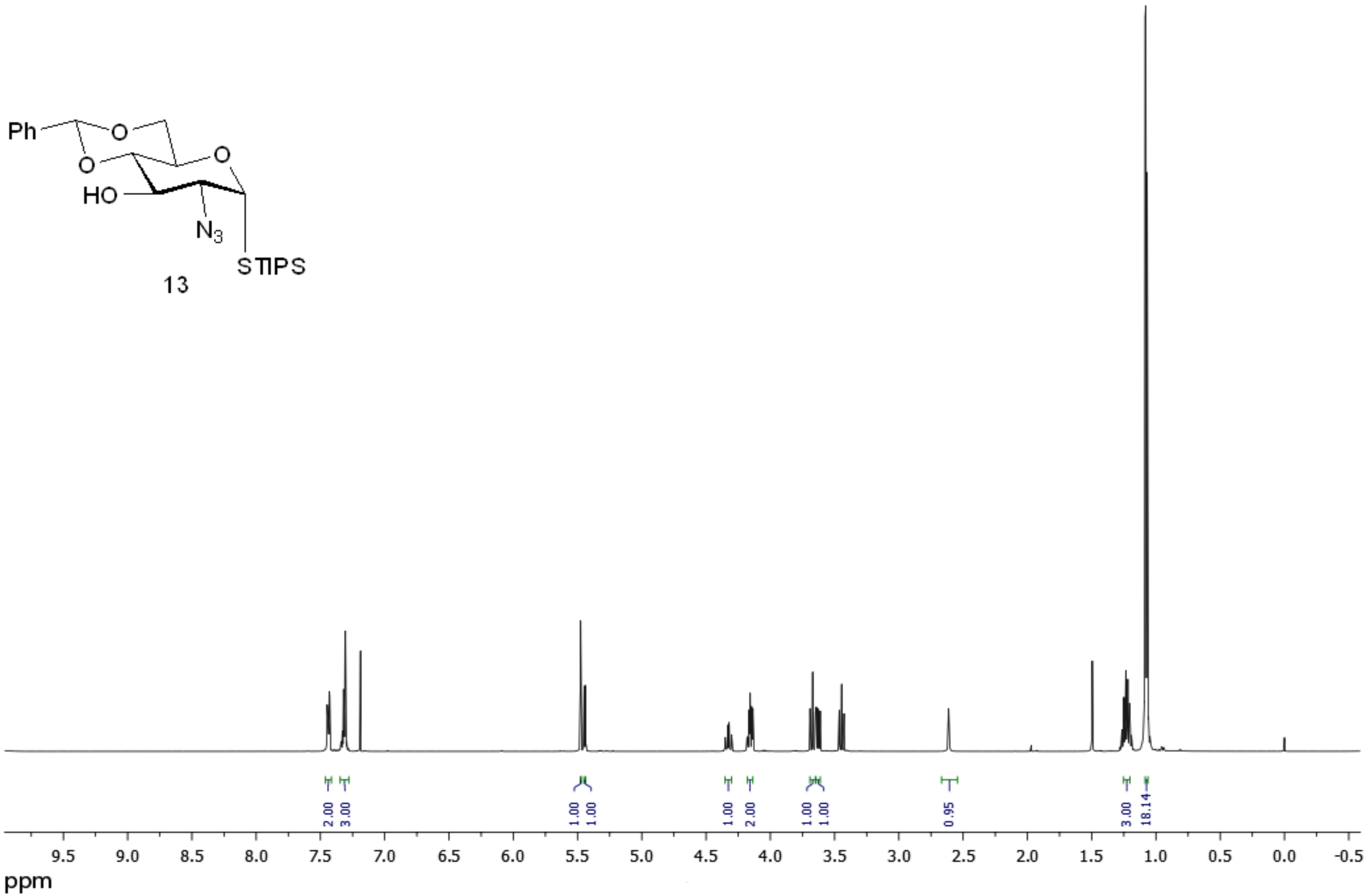


Figure 5: ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of compound 13

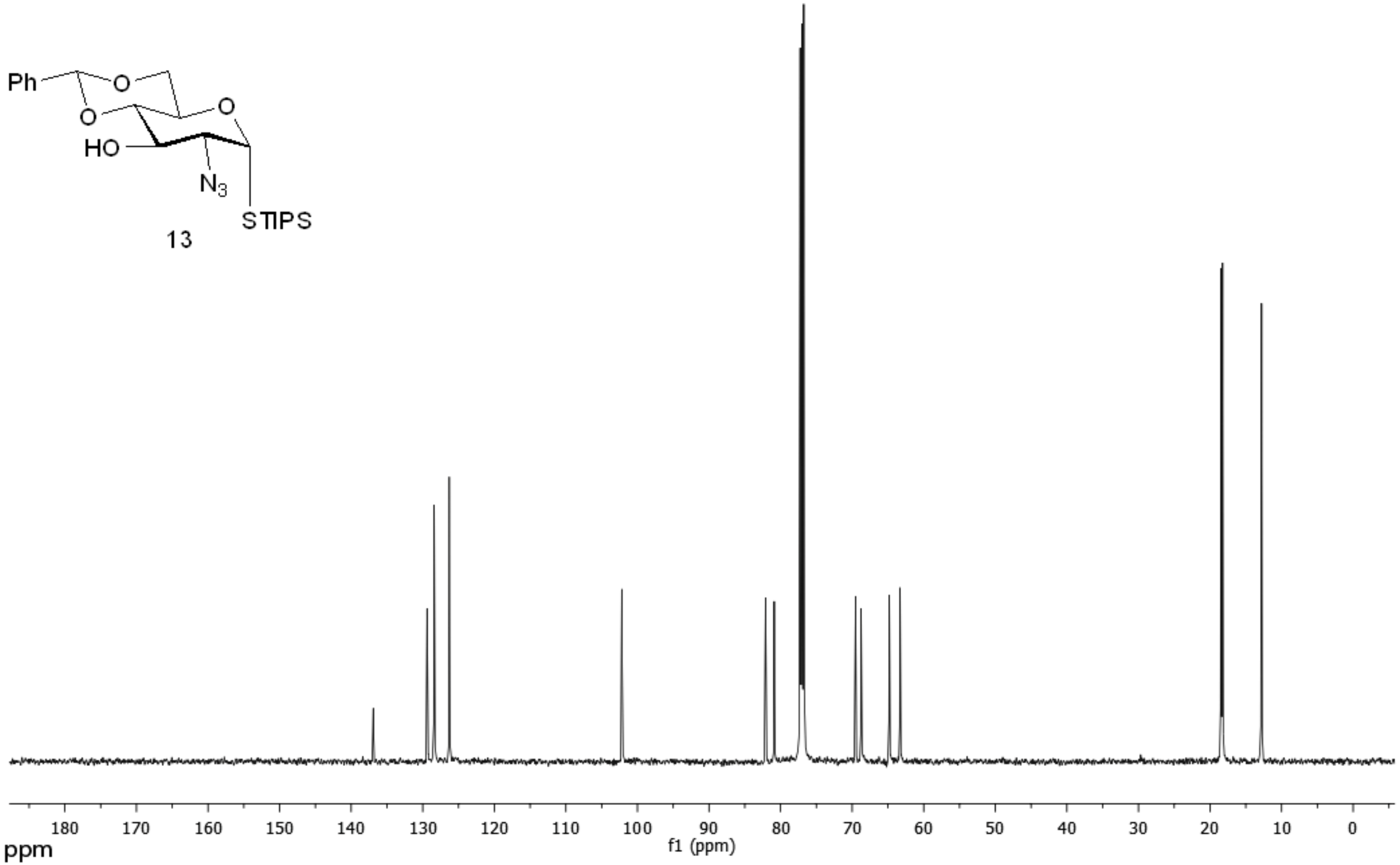


Figure 6: ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 13

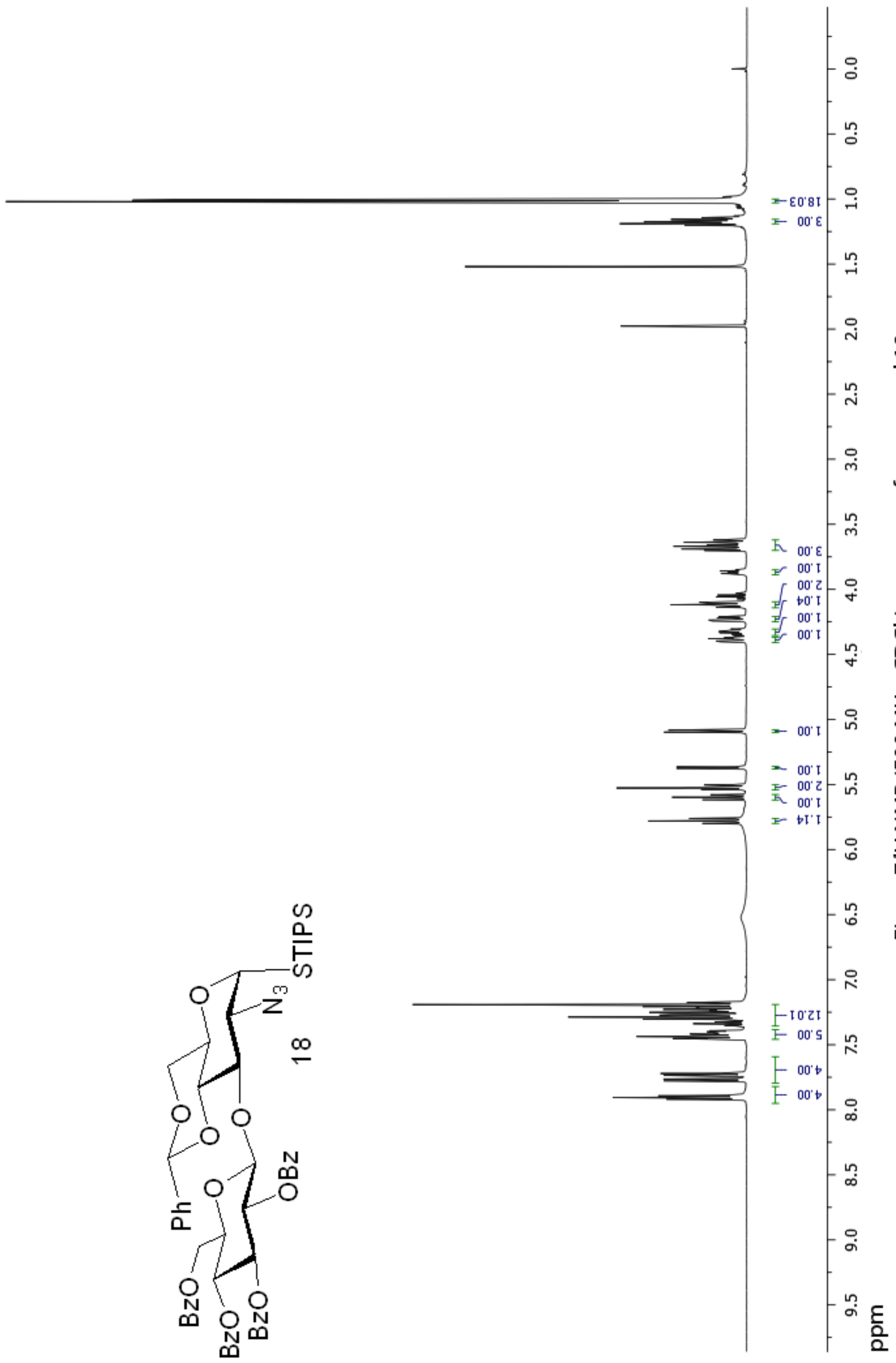


Figure7:^{1}H NMR (500 MHz, CDCl$_3$) spectrum of compound 18

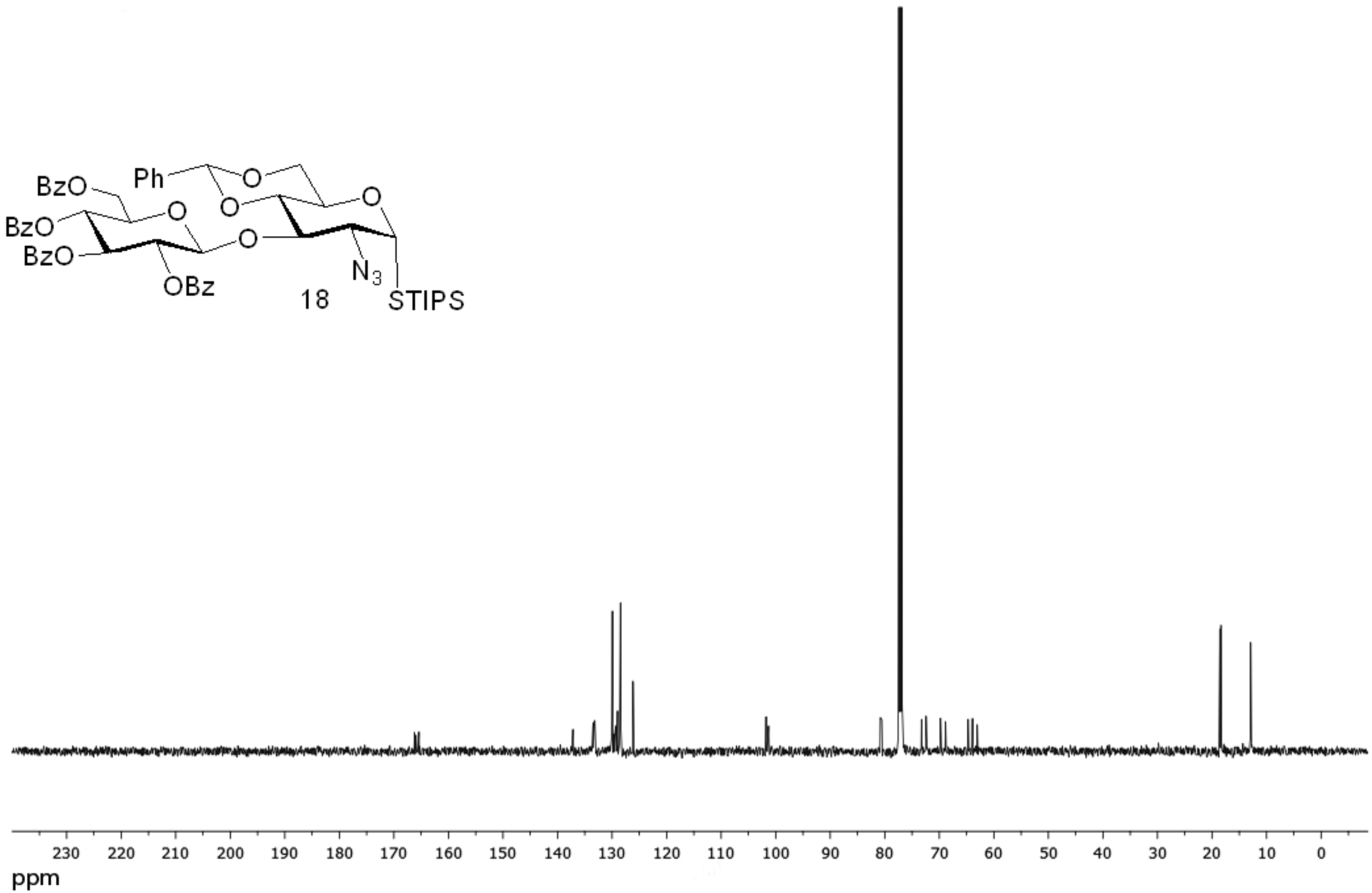


Figure8: ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 18

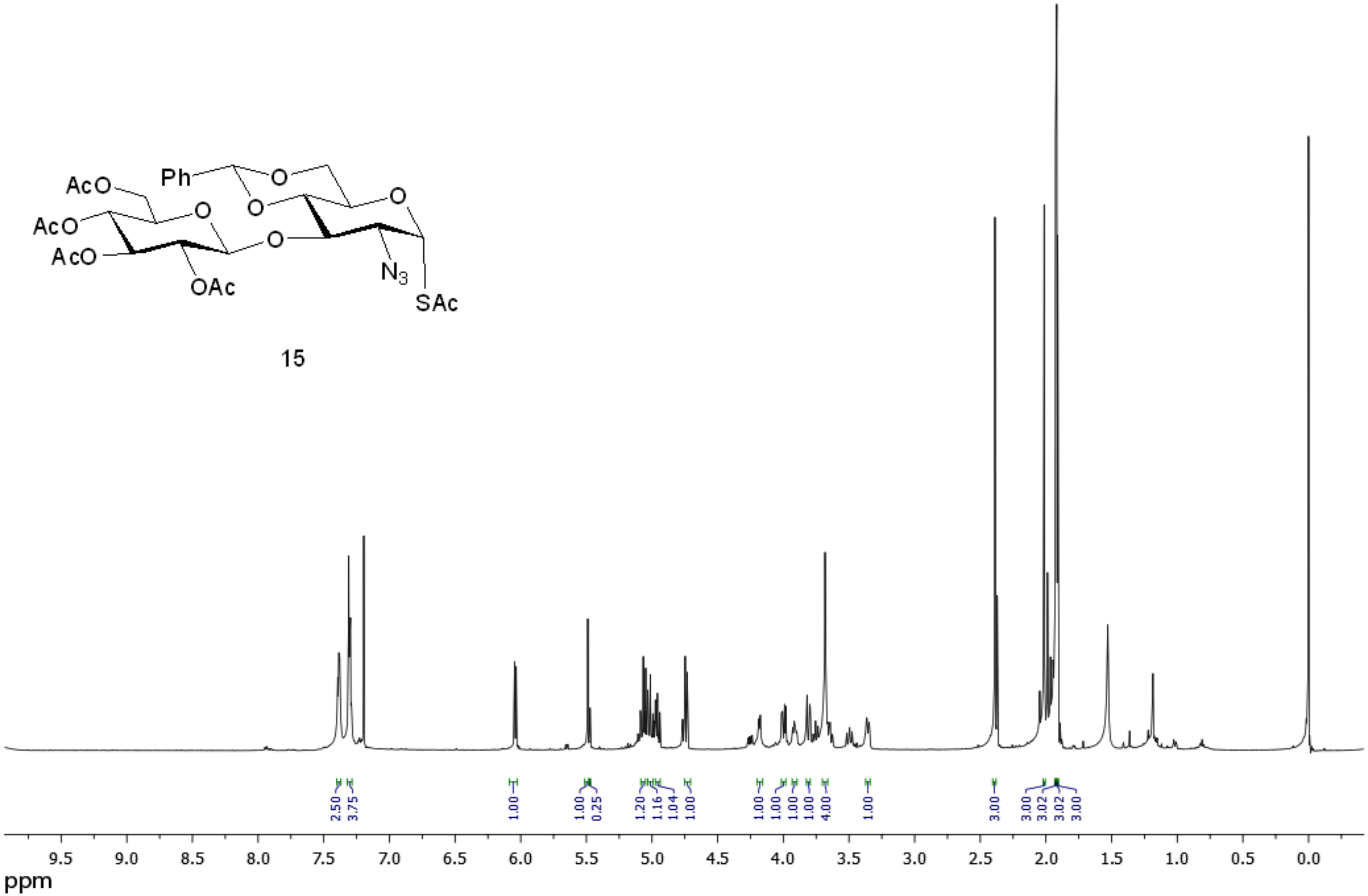
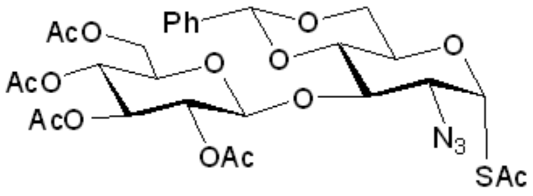


Figure9:^{1}H NMR (500 MHz, $CDCl_3$,) spectrum of compound 15



15

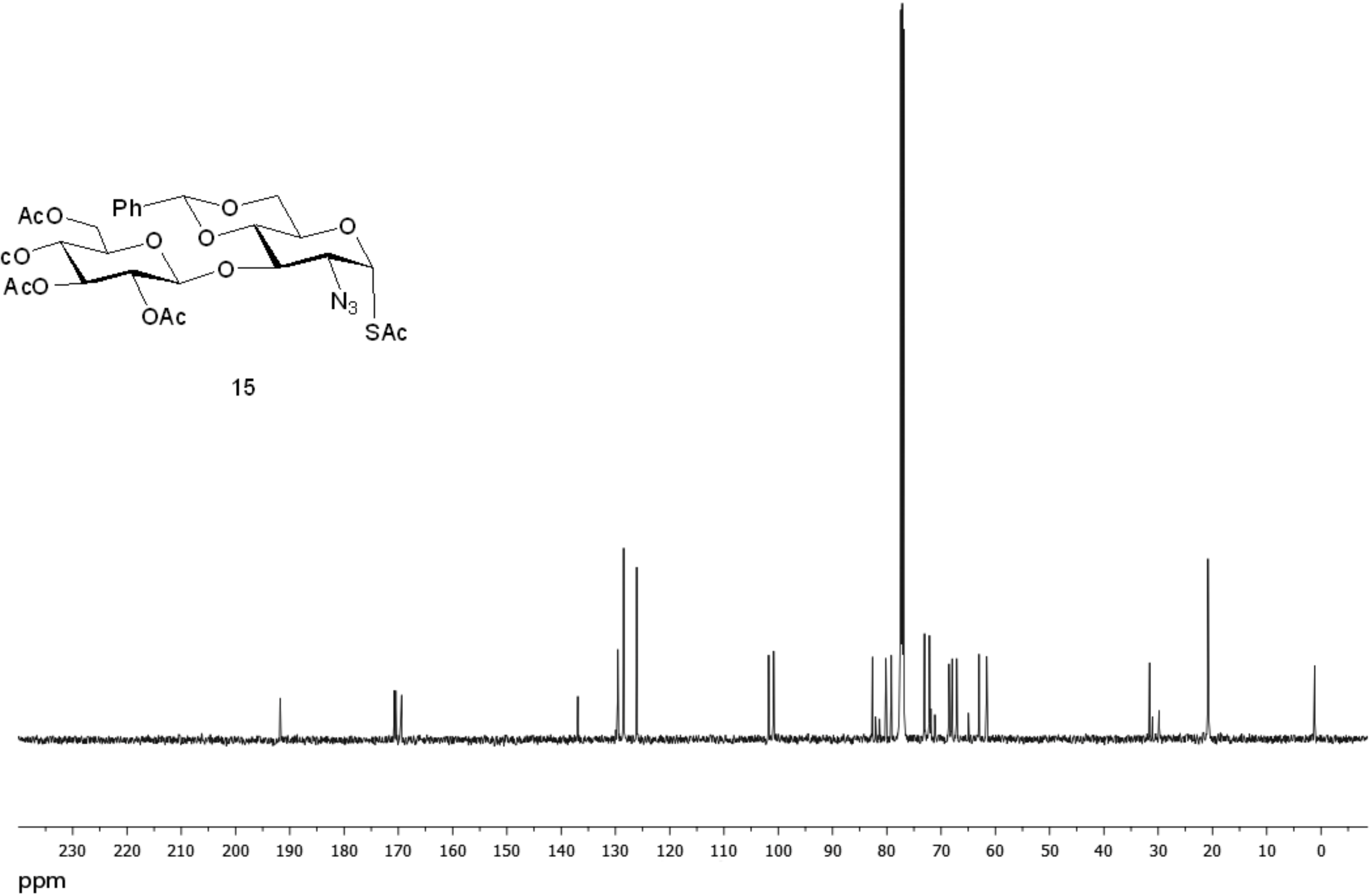
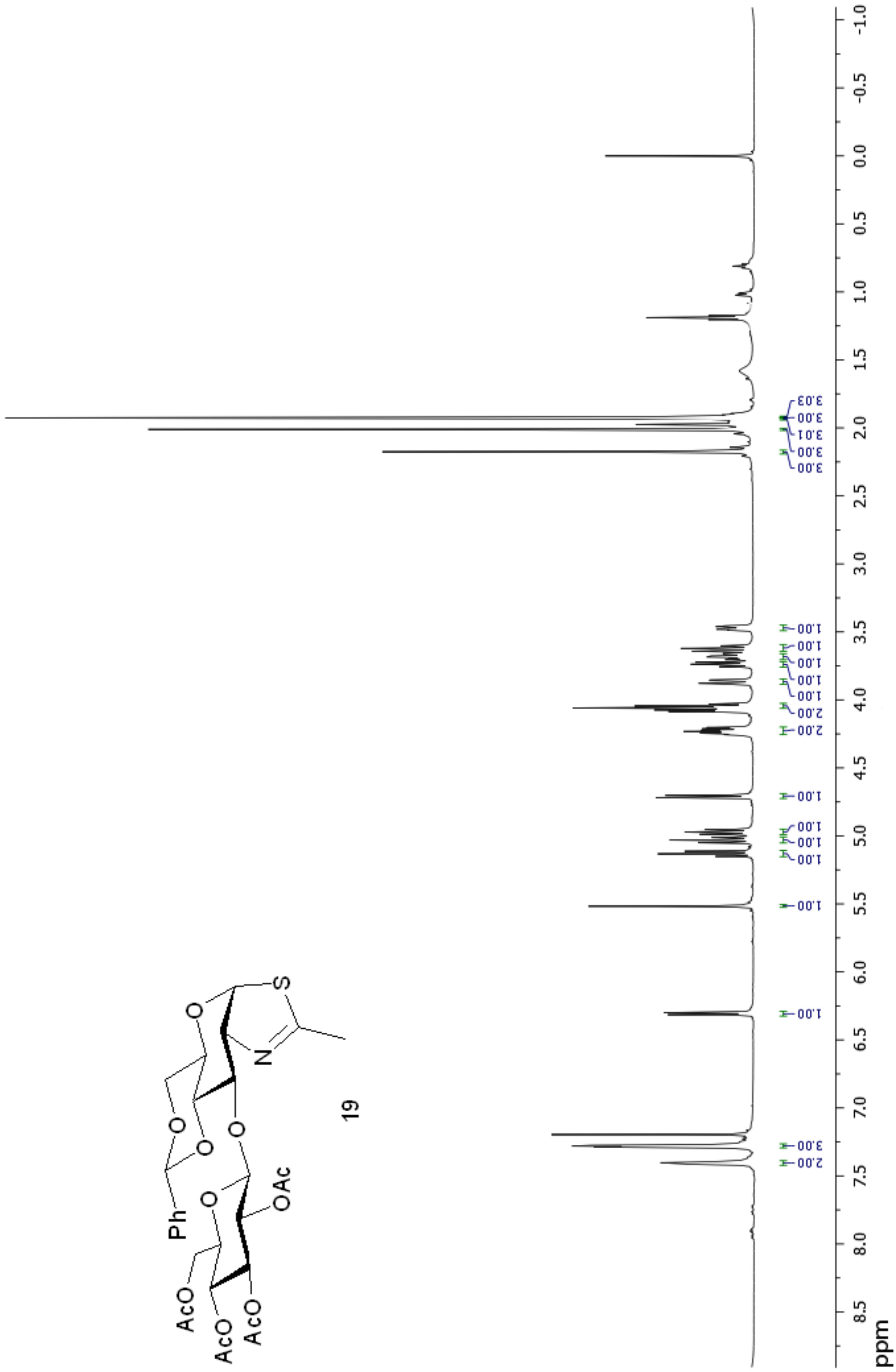


Figure10 : ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 15



19

Figure11: ¹H NMR (500 MHz, CDCl₃) spectrum of compound 19

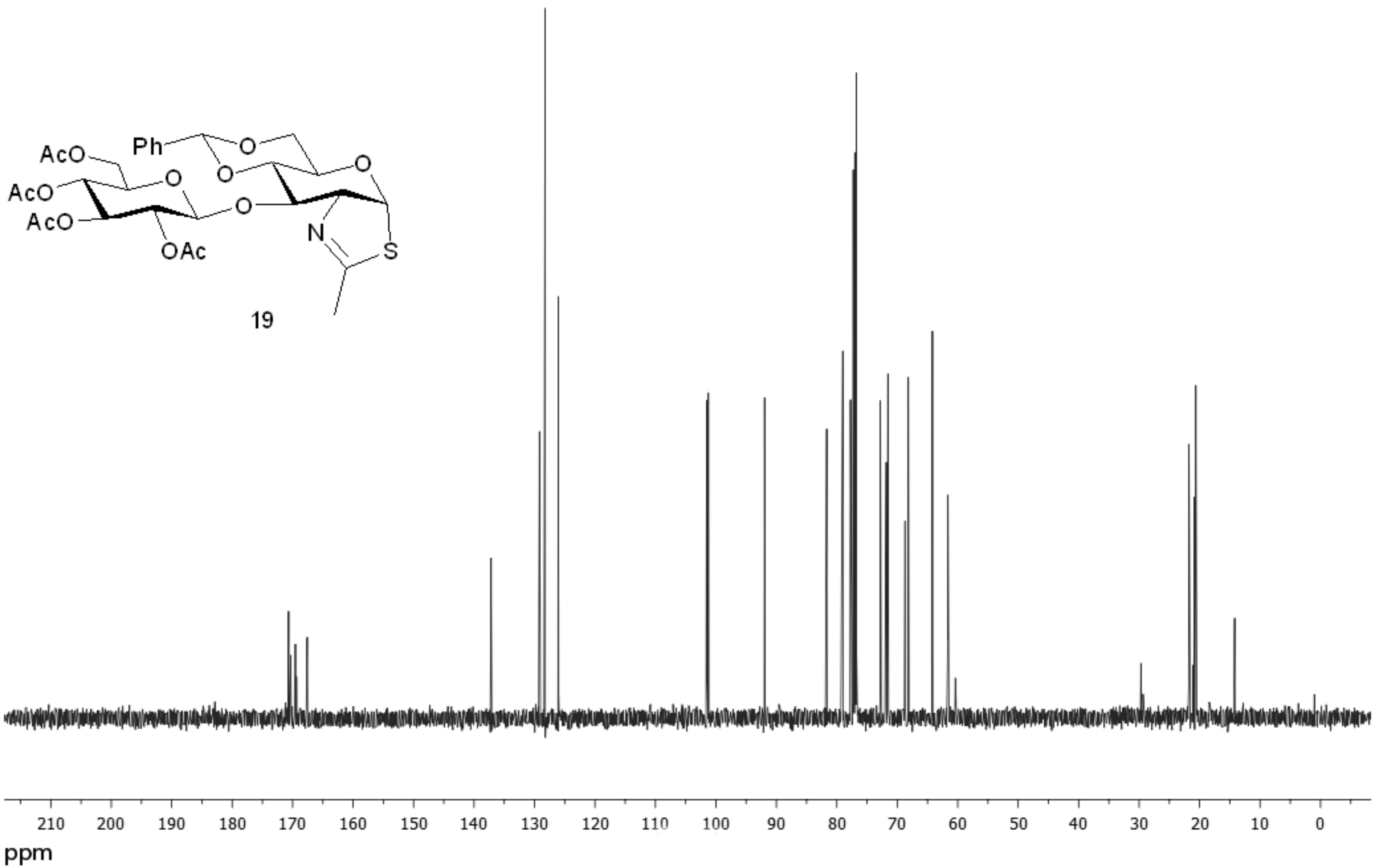


Figure12:^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 19

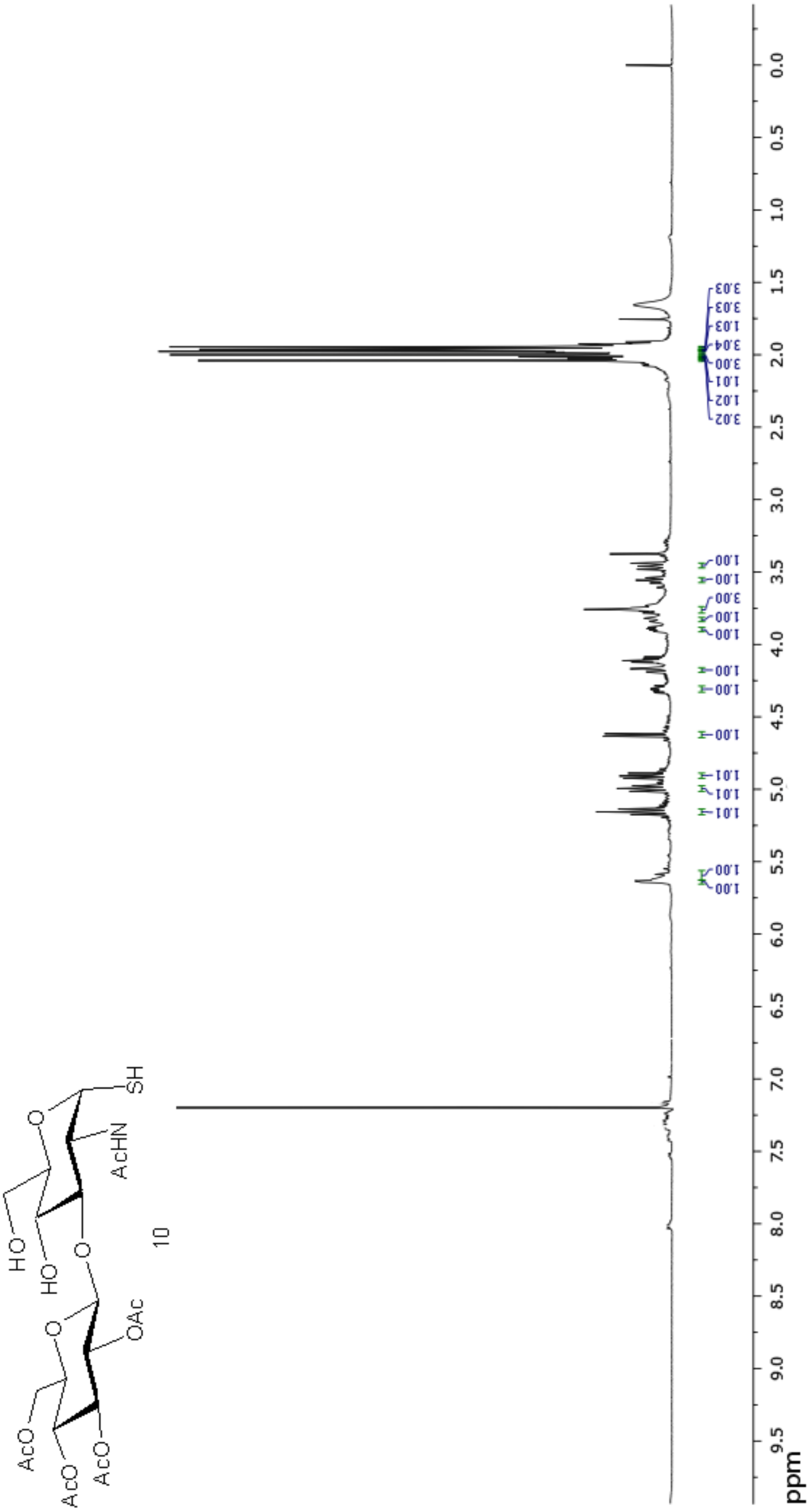


Figure13: ^1H NMR (500 MHz, $CDCl_3$) spectrum of compound 10

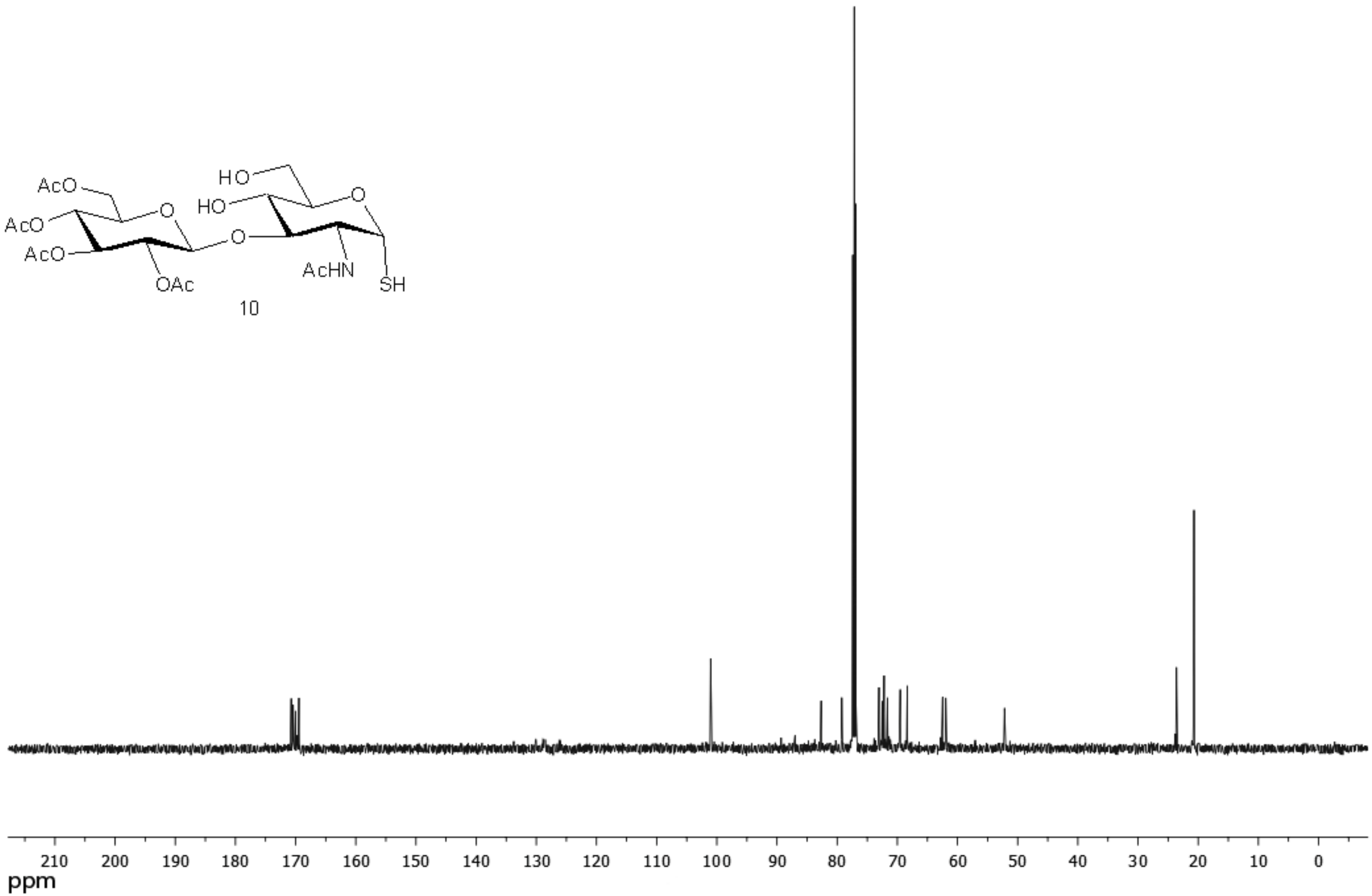


Figure14:^{13}C NMR (126 MHz, $CDCl_3$) spectrum of compound 10

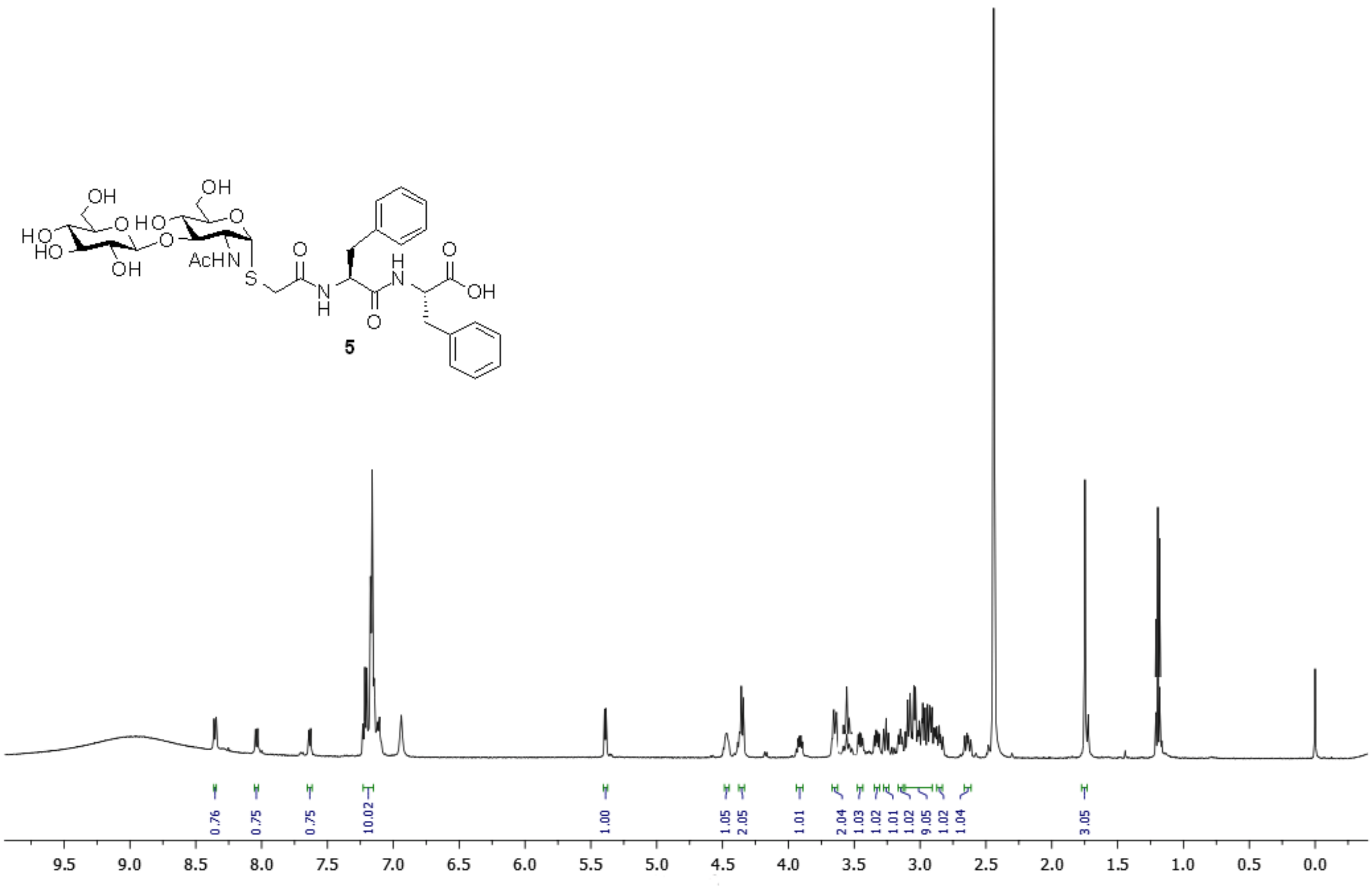


Figure15: ¹H NMR (500 MHz, DMSO-D6, 10% TFA-D) spectrum of compound 5

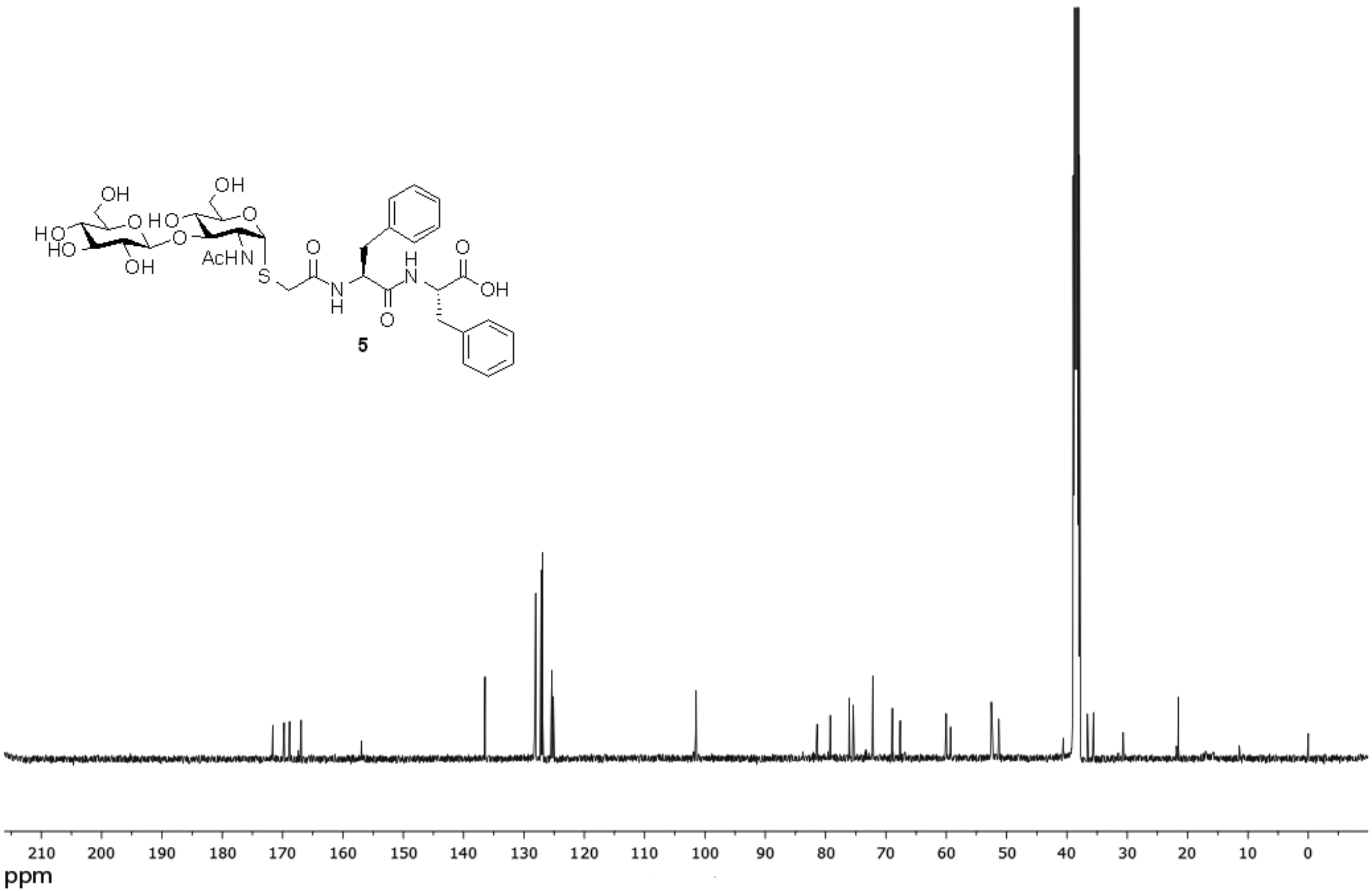
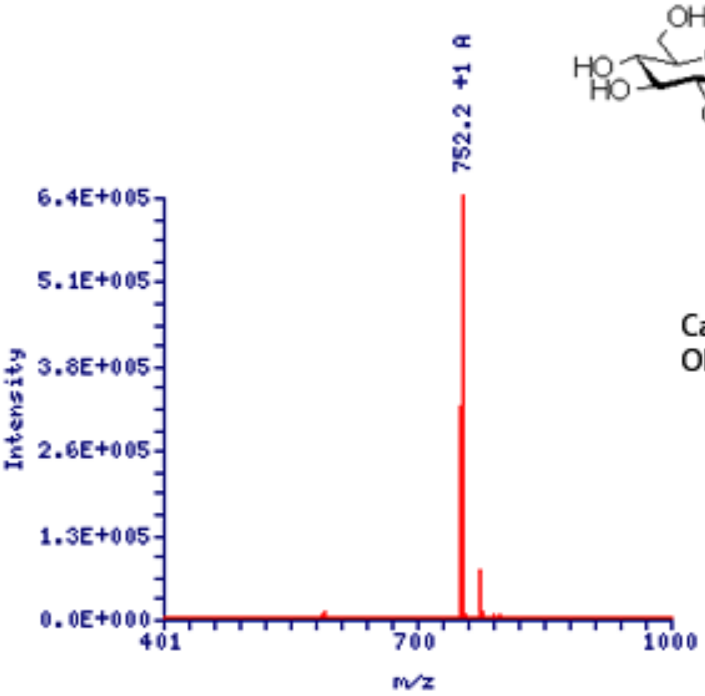
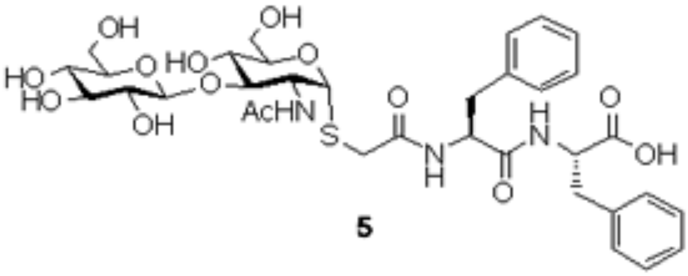


Figure16: ^{13}C NMR (126 MHz, DMSO-D$_6$) spectrum of compound 5





5

Calculated Mass [$C_{29}H_{35}N_{3}O_{13}S$}: 751.8 Da
Observed: 752.2 Da

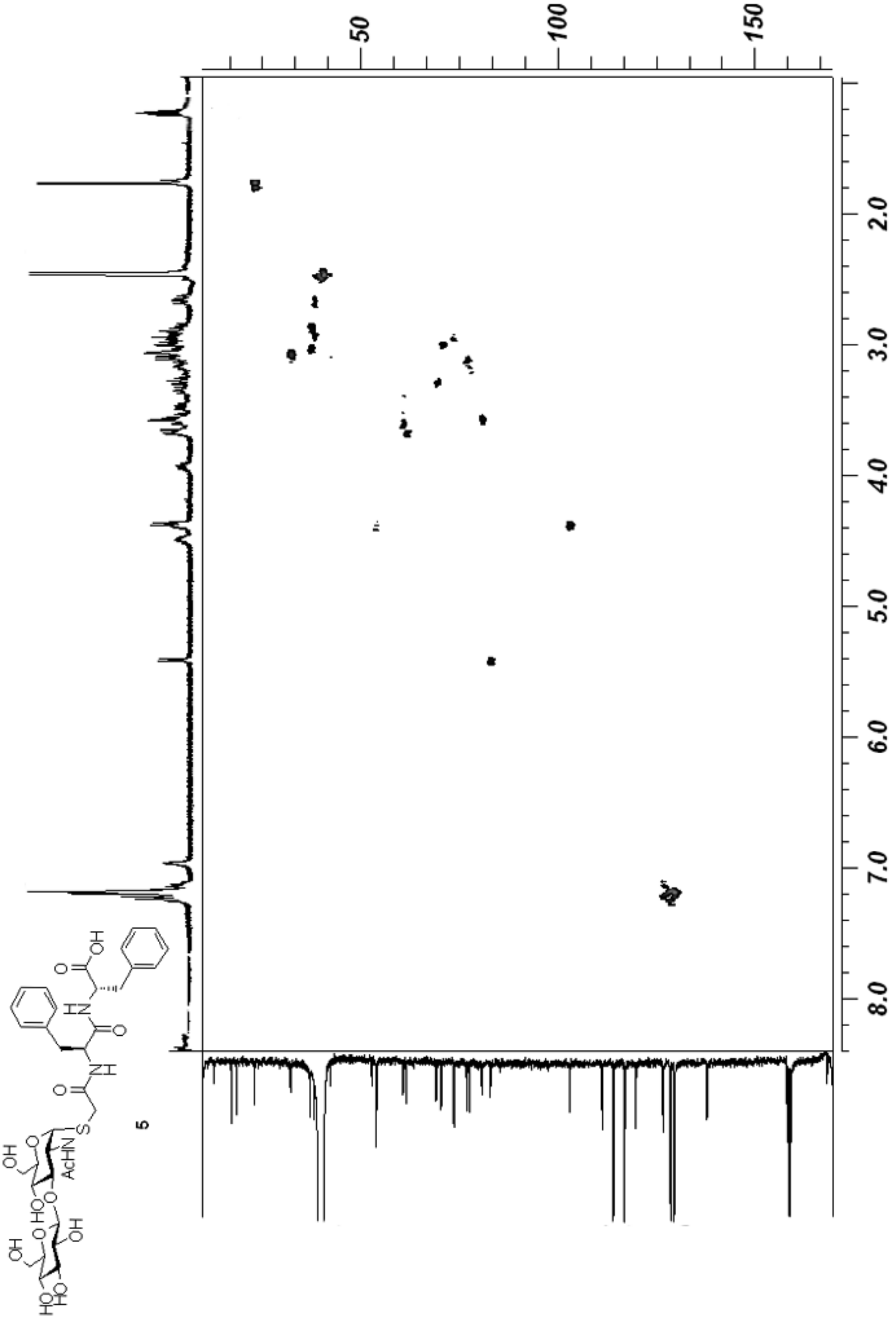


Figure17: HMQC NMR spectrum of compound 5

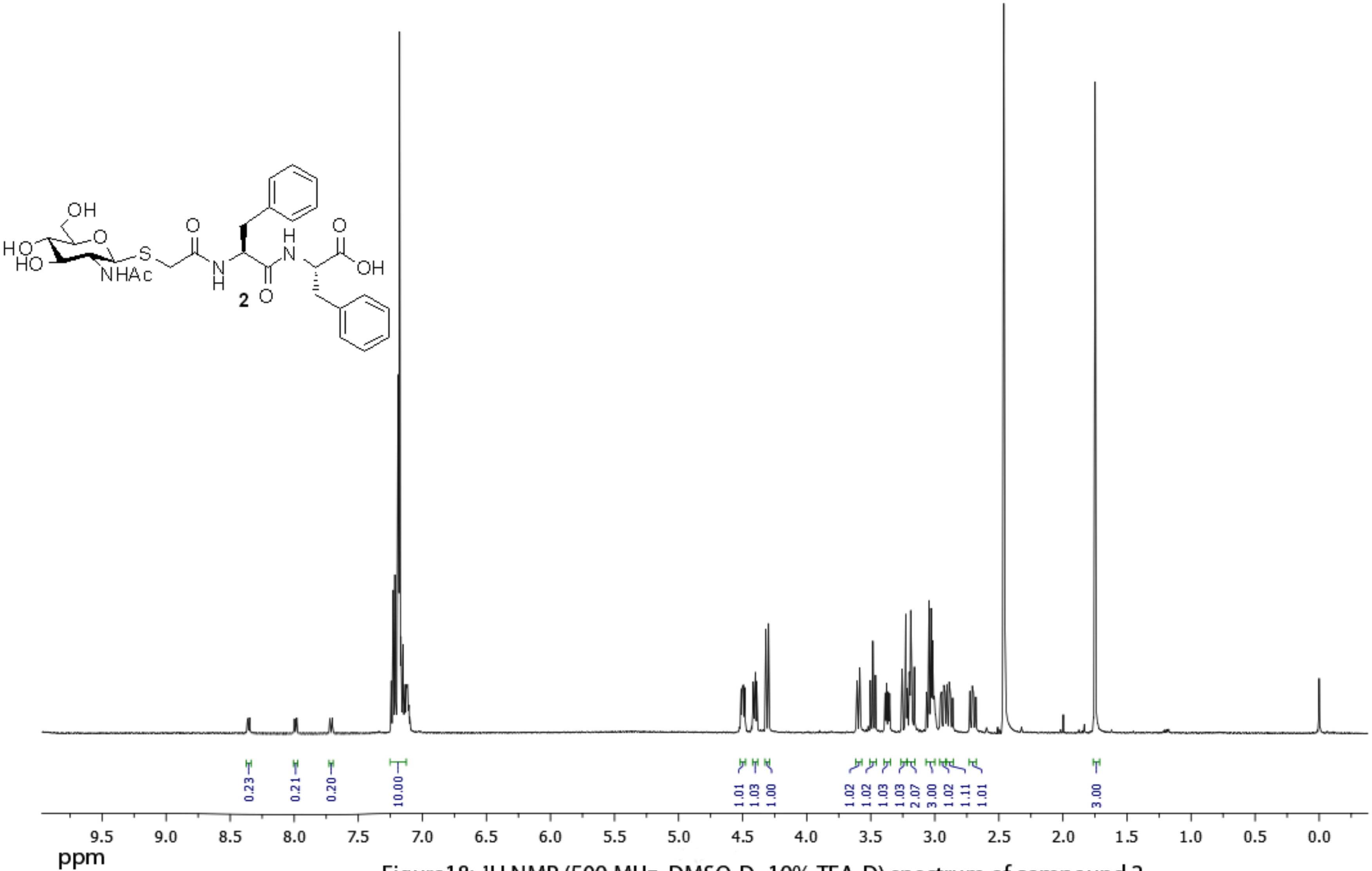


Figure18: ^{1}H NMR (500 MHz, DMSO-D$_6$,10% TFA-D) spectrum of compound 2

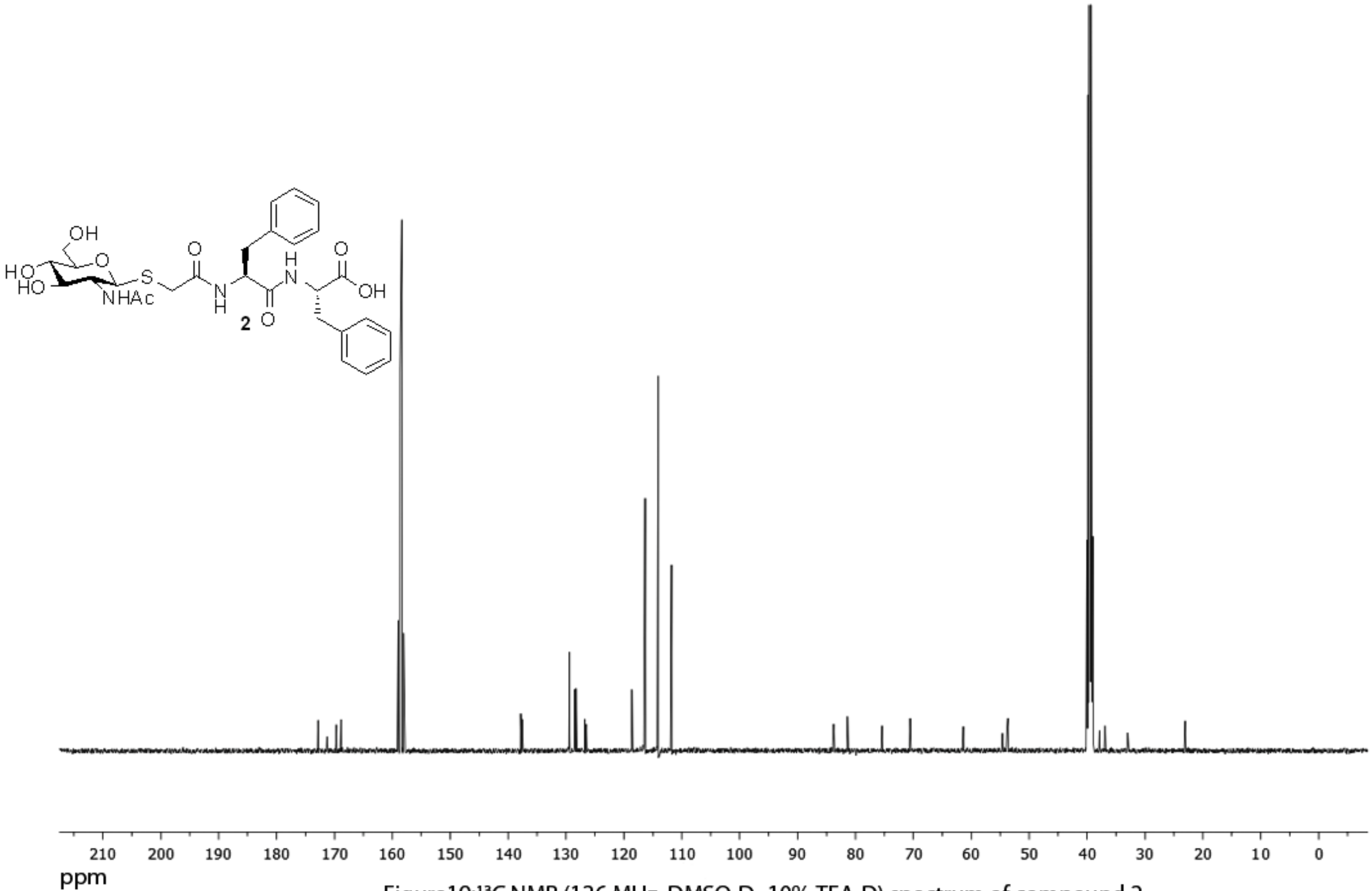
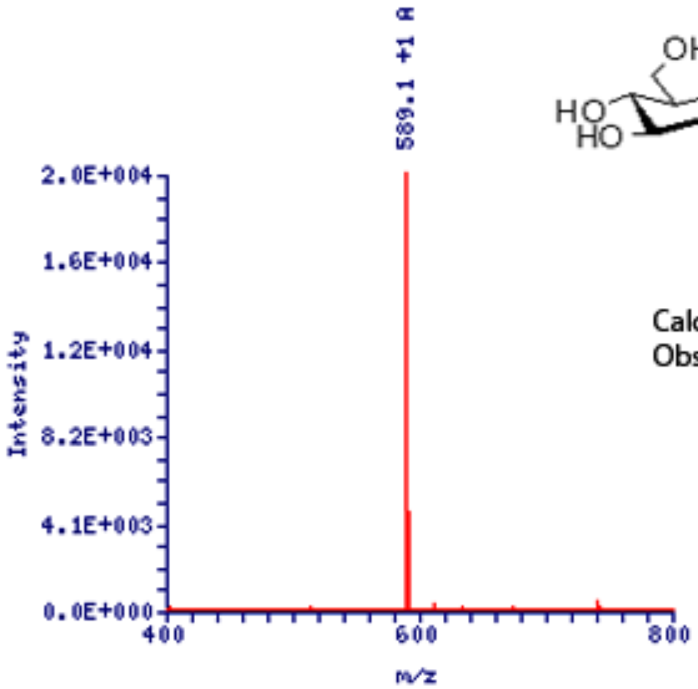
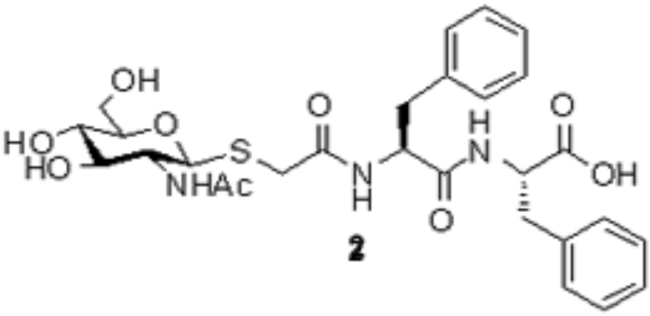


Figure19:^{13}C NMR (126 MHz, DMSO-D$_6$,10% TFA-D) spectrum of compound 2





2

Calculated Mass [$C_{28}H_{35}N_3O_9S$]: 589.21 Da
Observed: 589.1 Da

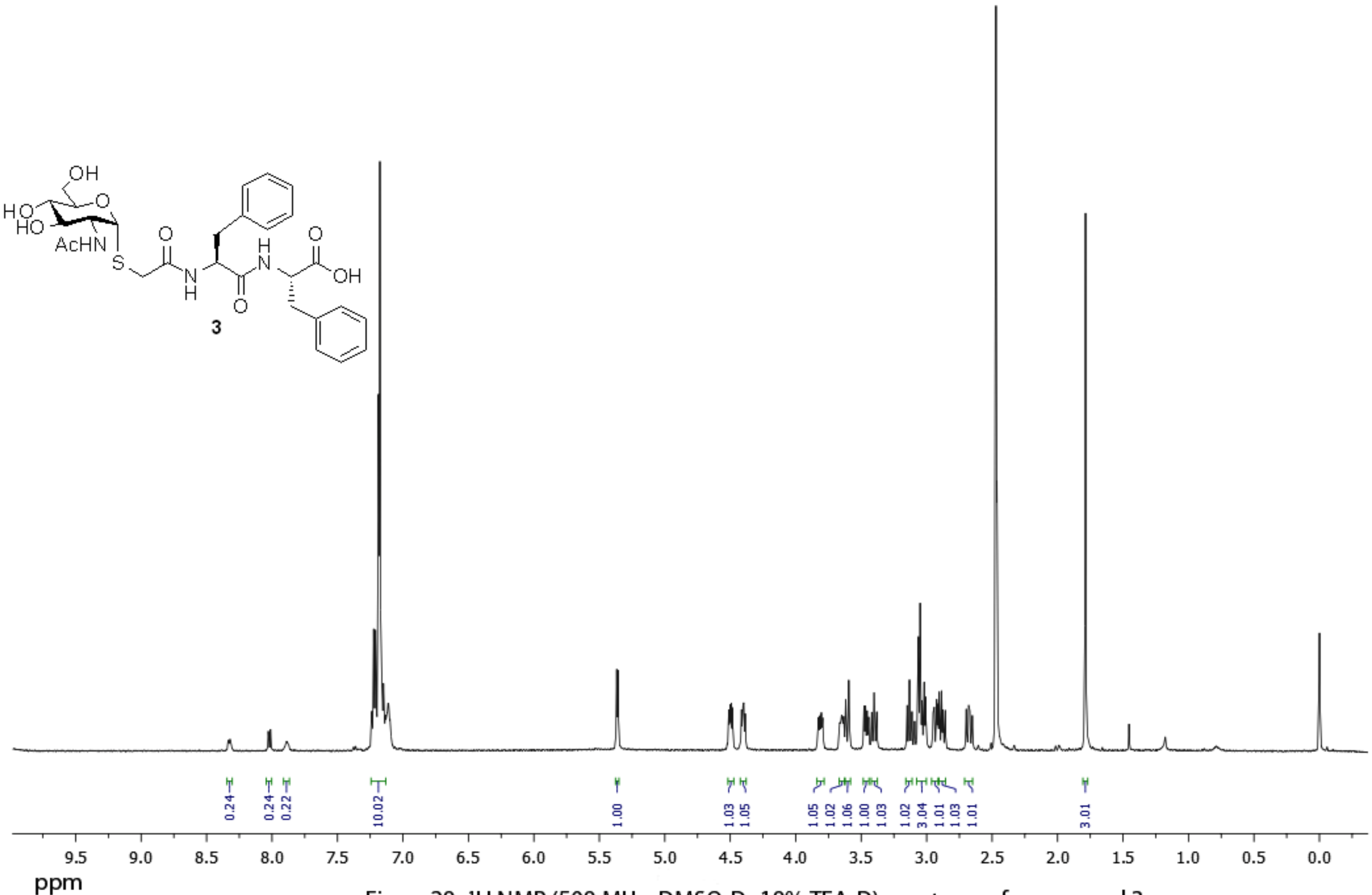


Figure20: ^{1}H NMR (500 MHz, DMSO-D$_6$,10% TFA-D) spectrum of compound 3

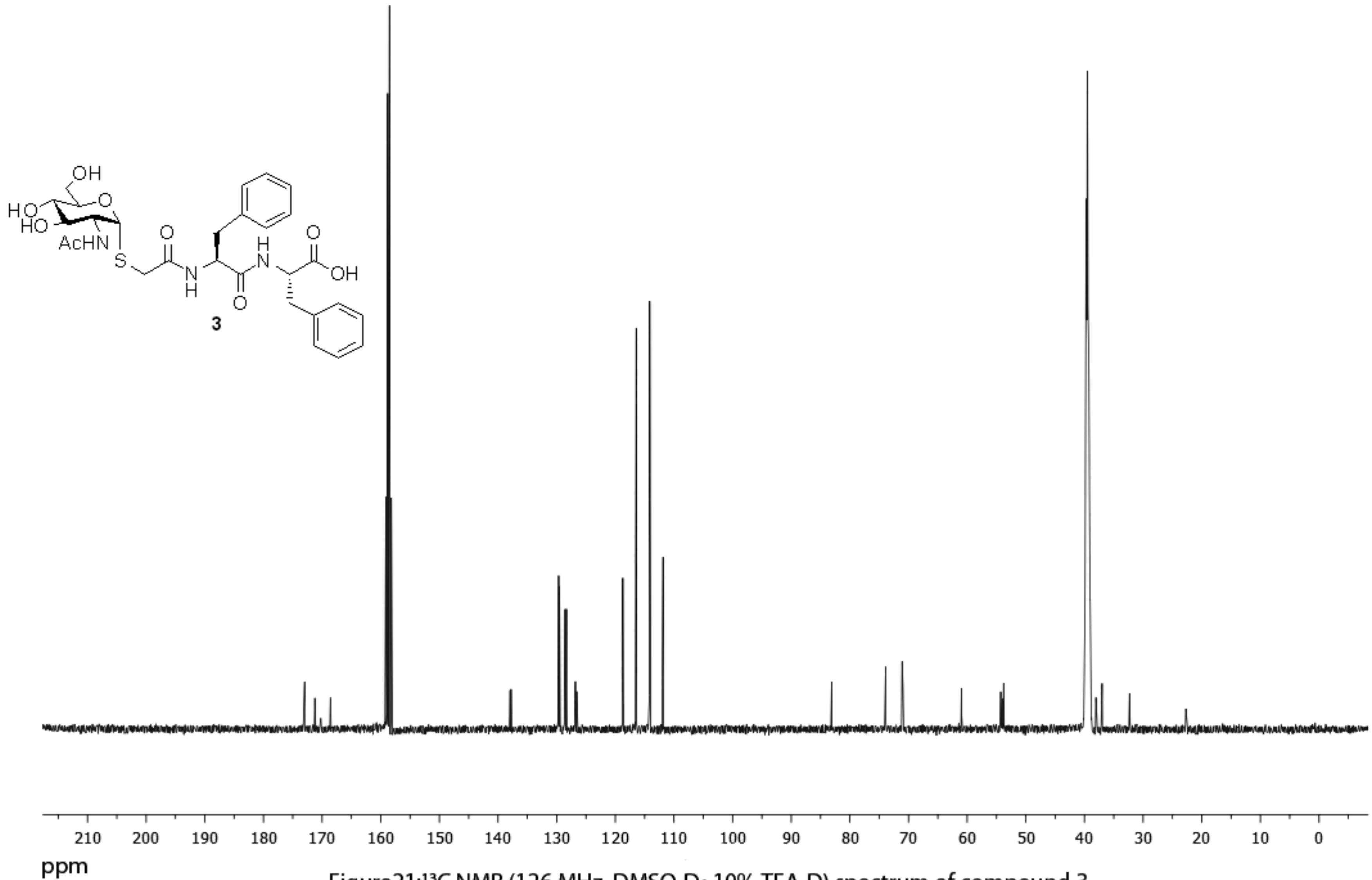
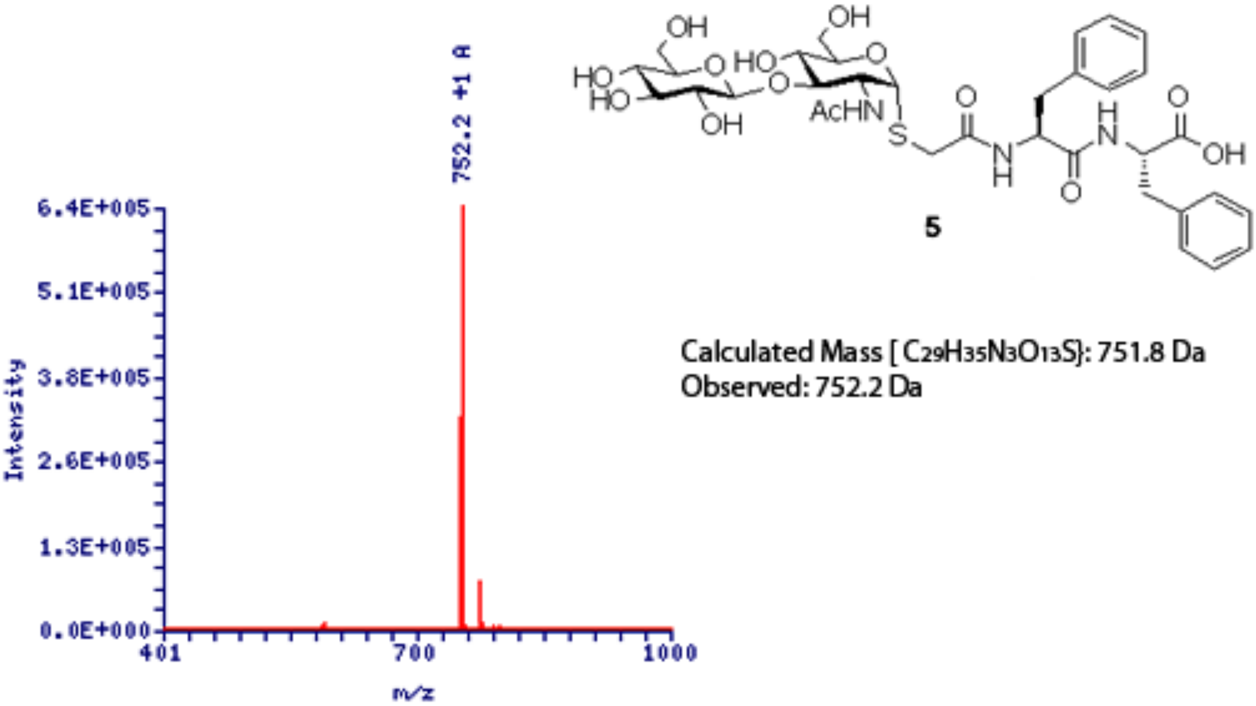


Figure21: ^{13}C NMR (126 MHz, DMSO-D_6,10% TFA-D) spectrum of compound 3



5

Calculated Mass [$C_{29}H_{35}N_3O_{13}S$}: 751.8 Da
Observed: 752.2 Da

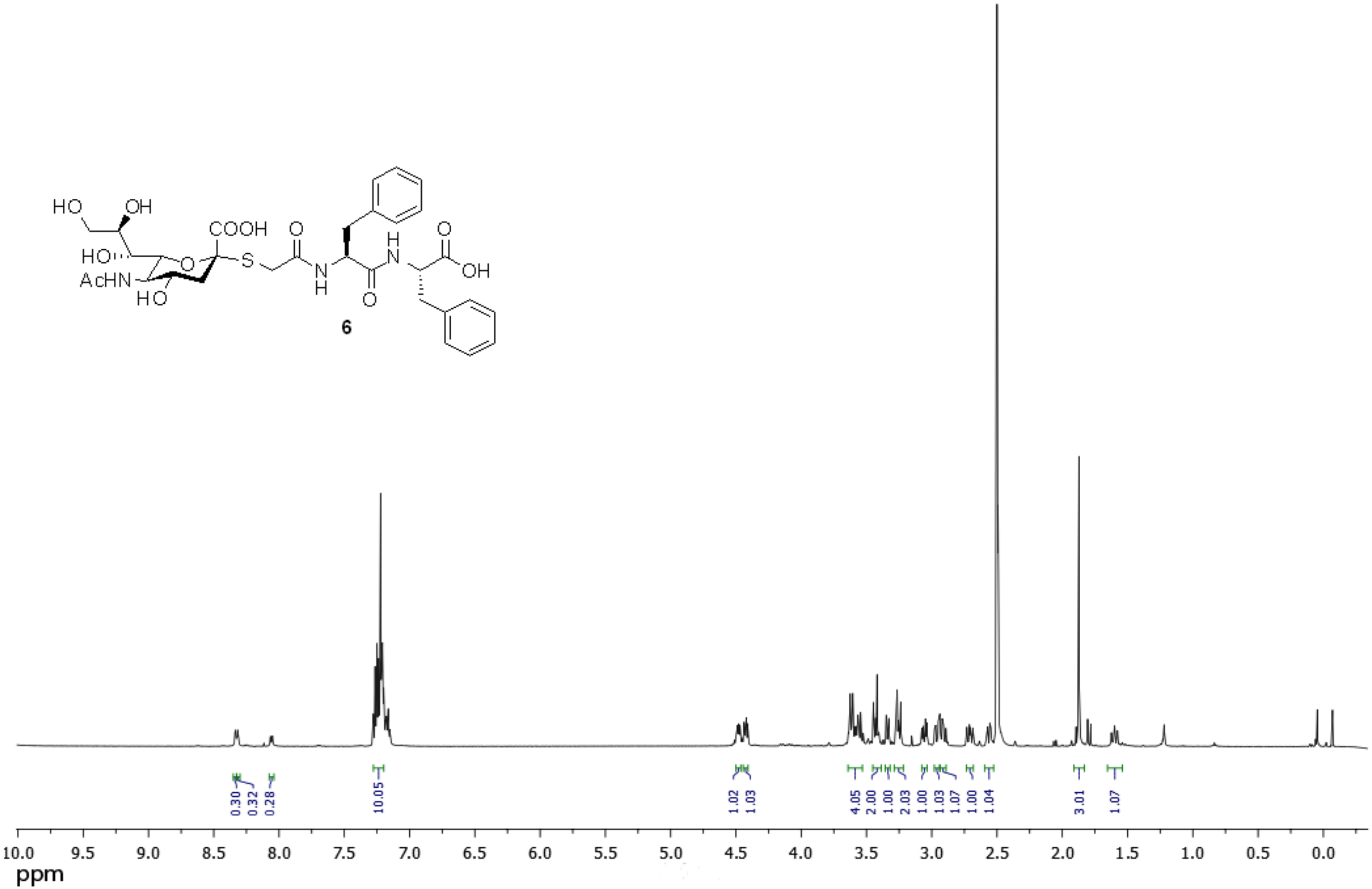


Figure22: ^{1}H NMR (500 MHz, DMSO-D$_6$, 10% TFA-D) spectrum of compound 6

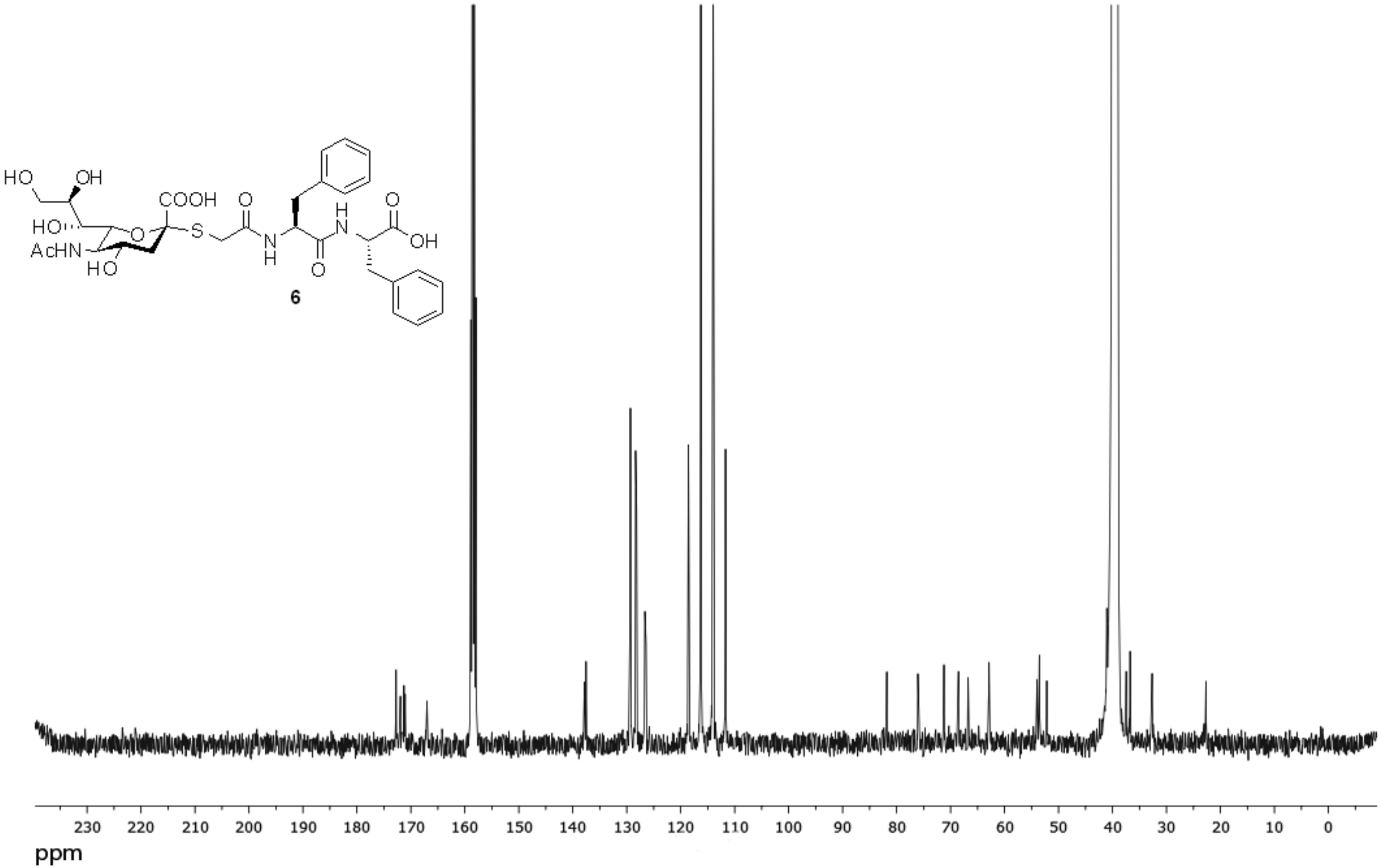
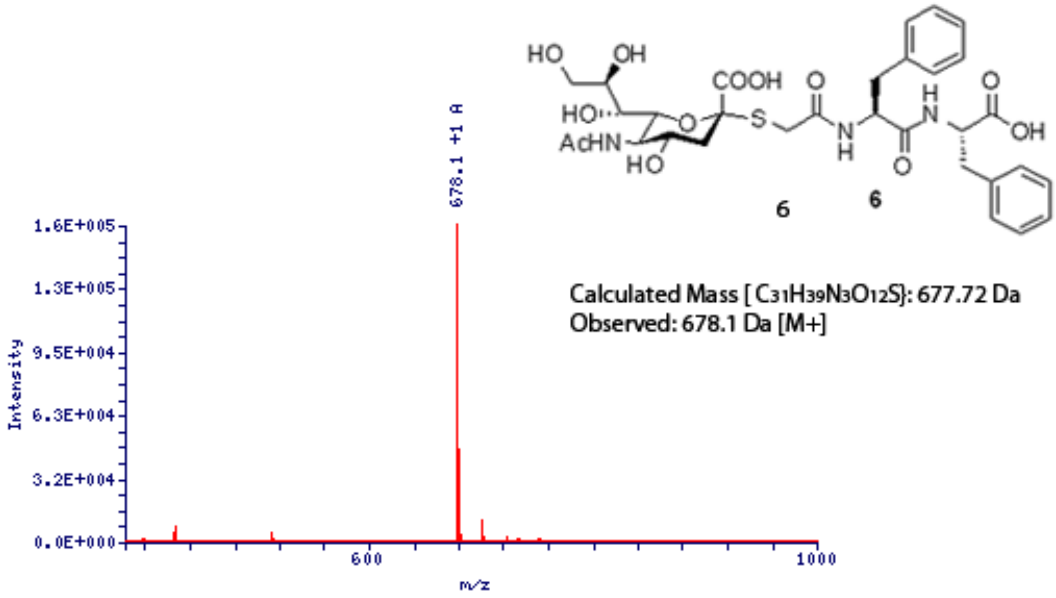


Figure23:^{13}C NMR (126 MHz, DMSO-D$_6$, 10% TFA-D) spectrum of compound 6



6

6

Calculated Mass [$C_{31}H_{39}N_3O_{12}S$]: 677.72 Da
Observed: 678.1 Da [M+]

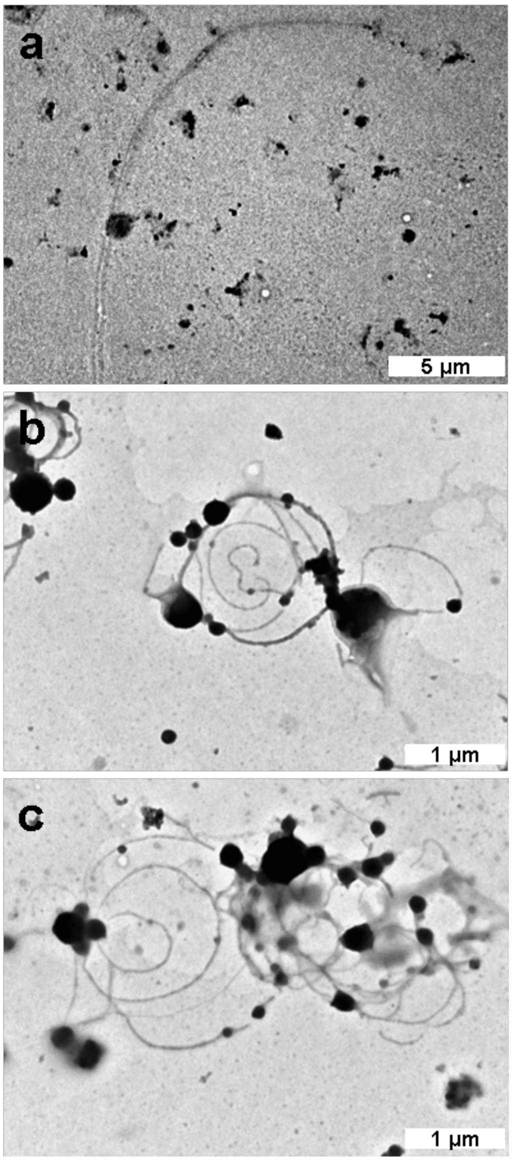
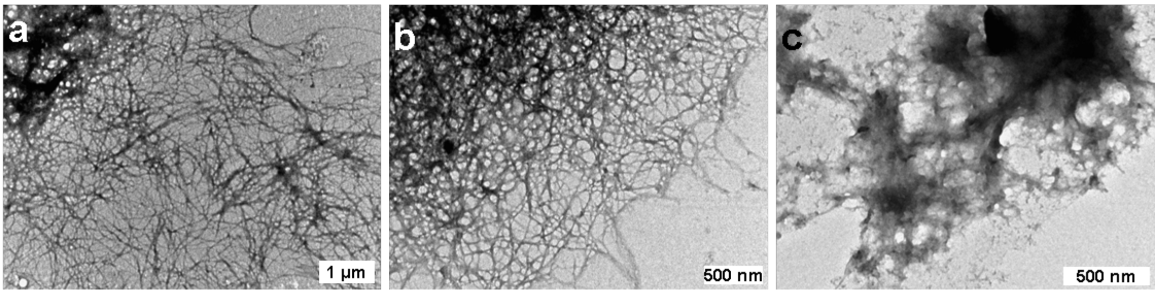


Fig. S1. a-c TEM micrograph of glycosylated diphenylalanine **6** self-assembles into ordered nano-structures after two hours of incubation at concentration of 6 mM.

Additional note

We have also studied the influence of glycosylated diphenylalanine **5** concentrations on the aggregation behavior. We found that at concentrations of 6 mM and **3** mM the glycosylated diphenylalanine self-assembled into ordered structures. However, at a concentration of 1.5 mM only few structures were observed and their morphology was less ordered than at higher concentrations. We added this as a note to the SI.



TEM micrograph of Glycosylated diphenylalanine **5** *assemblies at concentration of (a) 6 mM. (b) 3 mM and (c) 1.5 mM.*