Supporting Information for:

Synthetic Polyprenol-Pyrophosphate Linked Oligosaccharides Are Efficient Substrates for Mycobacterial Galactan Biosynthetic Enzymes

Xiaochao Xue, Ruixiang Blake Zheng, Akihiko Koizumi, Ling Han,

John S. Klassen, and Todd L. Lowary*

Alberta Glycomics Centre and Department of Chemistry, University of Alberta,

11227 Saskatchewan Drive, Edmonton, Alberta T6G 2G2, Canada

tlowary@ualberta.ca

Table of Contents

Production and purification of TB-GlfT1 and Ms-GlfT1

PCR and cloning procedure

TB-GlfT1. The GlfT1 genes for *Rv3782* from the *M. tuberculosis* H37Rv genome were amplified using TB genomic DNA as a template by PCR with Q5® High-Fidelity DNA Polymerase (New England BioLab Inc.). The sense strand 5′-TTATTAGGATCCATGACTGAATCGGTCTTCGCCG-3′ contained a BamHI site at the 5′-end and the antisense strand 5′-TTATTAGATATCTTATCATTTTTCGAACTGCG GGTGGCTCCAAGCGCTTGCAGATCCTCCAGGCTTGCCA-3′ contained TGATAA as the stop codon, a strep tag and an EcoRV site at the 3′-end. Following the manufacturer's instruction, the PCR profile was 98 ℃ for 3 s (initial denaturation) and then 30 cycles of 98 ℃ for 10 s; 72 ℃ for 30 s; and 72 ℃ for 30s. The PCR product was digested with BamHI and EcoRv and separated from reaction impurities by agarose gel electrophoresis. The digested DNA was excised from the gel and purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). The purified *Rv3782* cDNA was ligated into the pMF406 vector (a gift from Professor Xao-Yong Fan, Laboratory of Tuberculosis & Mycobacteria Shanghai Public Health Clinical Center, Fudan University, Shanghai China) that had been previously digested with BamHI and EcoRv, yielding an expression construct for TB-GlfT1. Ca^{2+} competent DH5α *E. coli* cells were transformed with the ligation mix and plated on LB-agar plates containing kanamycin (25 mg/L). After growing at 37 ℃ overnight, a single colony was picked to inoculate 5 mL LB

medium containing kanamycin (25 mg/L) for another overnight growth. Plasmid DNA was purified using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and sequenced for verification. Sequence verified pMF406 expression plasmids for GlfT1 were electrotransformed into competent *Mycobacterium smegmatis* mc^2 *155* cells and plated on Middlebrook 7H10 agar plate containing kanamycin (15 mg/L) for growth. A fresh single colony was used to inoculate 3.5 mL 7H9 broth containing ADC enrichment cultures and kanamycin (15 mg/L) for 24 h. Sterile glycerol was added into the cultures, aliquoted and stored at – 80 ℃ freezer for further use.

Ms-GlfT1. The genomic DNA was extracted from the fresh culture of *M. smegmatis* mc^2155. Similar cloning procedures to that used for TB-GlfT1 provided expression constructs using the genomic DNA as the template. The sense and antisense sequences were 5′-TTATTAGGATCCATGACGCACACTGAGGTCGTCTG-3′ and 5′-TTATTAGA ATTCTCAGTGGTGGTGGTGGTGGTGCTCGAGTCGCTGGAACCTTTCGCGTCT-3′, respectively.

Bacterial protein expression

Single colonies of TB-GlfT1 and Ms-GlfT1 were used to inoculate 100 mL LB cultures containing kanamycin (15 mg/L), glycerol (2%) and 0.2% Tween 80. The cultures were incubated at 37 ℃ with shaking at 200 rpm for 2 days. Fresh LB medium was inoculated with turbid cultures at a dilution of 1:50. Protein expression was induced by addition of

0.02% acetamide when cell density reached 1.2-1.4 at 600nm. Bacterial cultures were further grown at 37 ℃ with shaking at 200 rpm for 24 h. Cell suspensions were centrifuged at 6500 × g at 4 ℃ for 15 min and cell pellets were purified immediately or stored at − 80 ℃ freezer for further use.

Purification

Purification of TB-GlfT1. To purify the Strep-tagged TB-GlfT1, cell pellets (20 g/L culture) were resuspended in 100 mM Tris-Cl, containing 1 mM EDTA, 150 mM NaCl, 10% Glycerol and 0.1% Triton X-100 (pH 8.0) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche). The suspension was passed through a cell disruptor at 35000 psi twice and centrifuged at 100500 g for 60 min. The supernatant was loaded onto the Strep-Tactin Sepharose column (2.5 mL resin) equilibrated with 100 mM Tris-Cl, 150 mM NaCl, 10% Glycerol (pH 8.0) at a flow rate of 1 mL/min and washed with the same buffer until the OD_{280} was at the baseline. The enzyme was eluted with 50 mM MOPS, 0.15 M NaCl, 10% Glycerol and 5 mM Desthiobiotin (pH 7.6) at a flow rate of 1 mL/min. Fractions were frozen immediately for further use at − 80 °C.

Purification of Ms-GlfT1. To purify the His-tagged Ms-GlfT1, the cell pellet (20 g /L culture) was resuspended in 50 mM sodium phosphate buffer, pH 7.6, containing 400 mM NaCl, 20 mM imidazole, 10% glycerol and 0.1% Triton X-100 supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche). The suspension was passed

through a cell disruptor through at 35000 psi twice. The cell lysate was centrifuged at 100500 g for 60 min at 4 ℃. The Ms-GlfT1 enzyme was purified from the resulting supernatant by applying directly to a Ni-NTA Superflow column that has been treated with an equilibrating buffer of 50 mM sodium phosphate buffer, pH 7.6, containing 400 mM NaCl, 20 mM imidazole, 10% glycerol. The column was washed with equilibrating buffer until the A280 nm returned to background levels. At this point the column was treated with the second wash buffer (50 mM sodium phosphate buffer, pH 7.6, containing 400 mM NaCl, 50 mM imidazole, 10% glycerol). The protein was eluted with 50 mM sodium phosphate buffer, pH 7.6, containing 400 mM NaCl, 250 mM imidazole, 10% glycerol. A flow rate of 1 mL/min was used through the column equilibration, column washing and protein elution steps. Fractions were frozen immediately for further use at − 80 ℃.

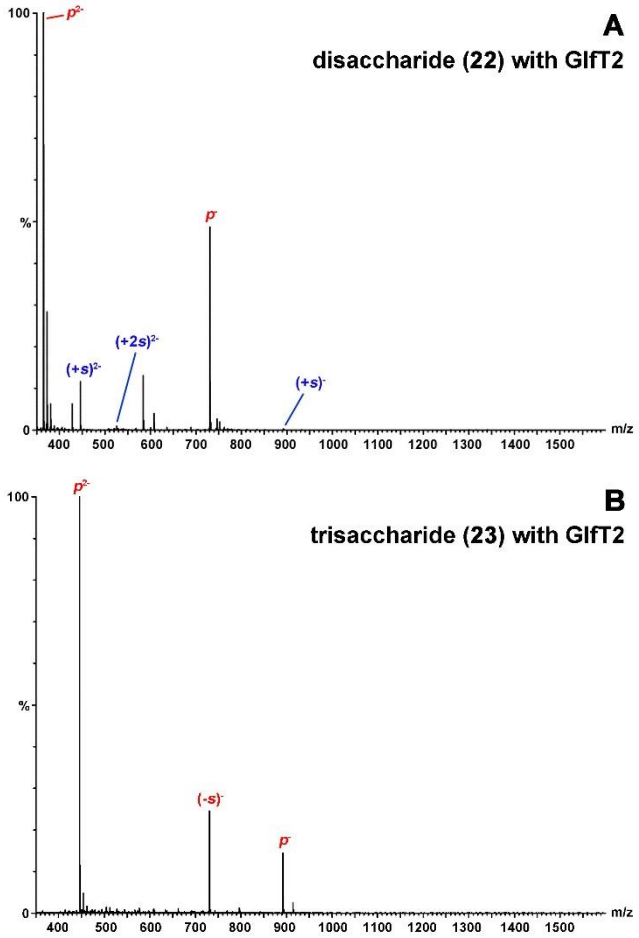
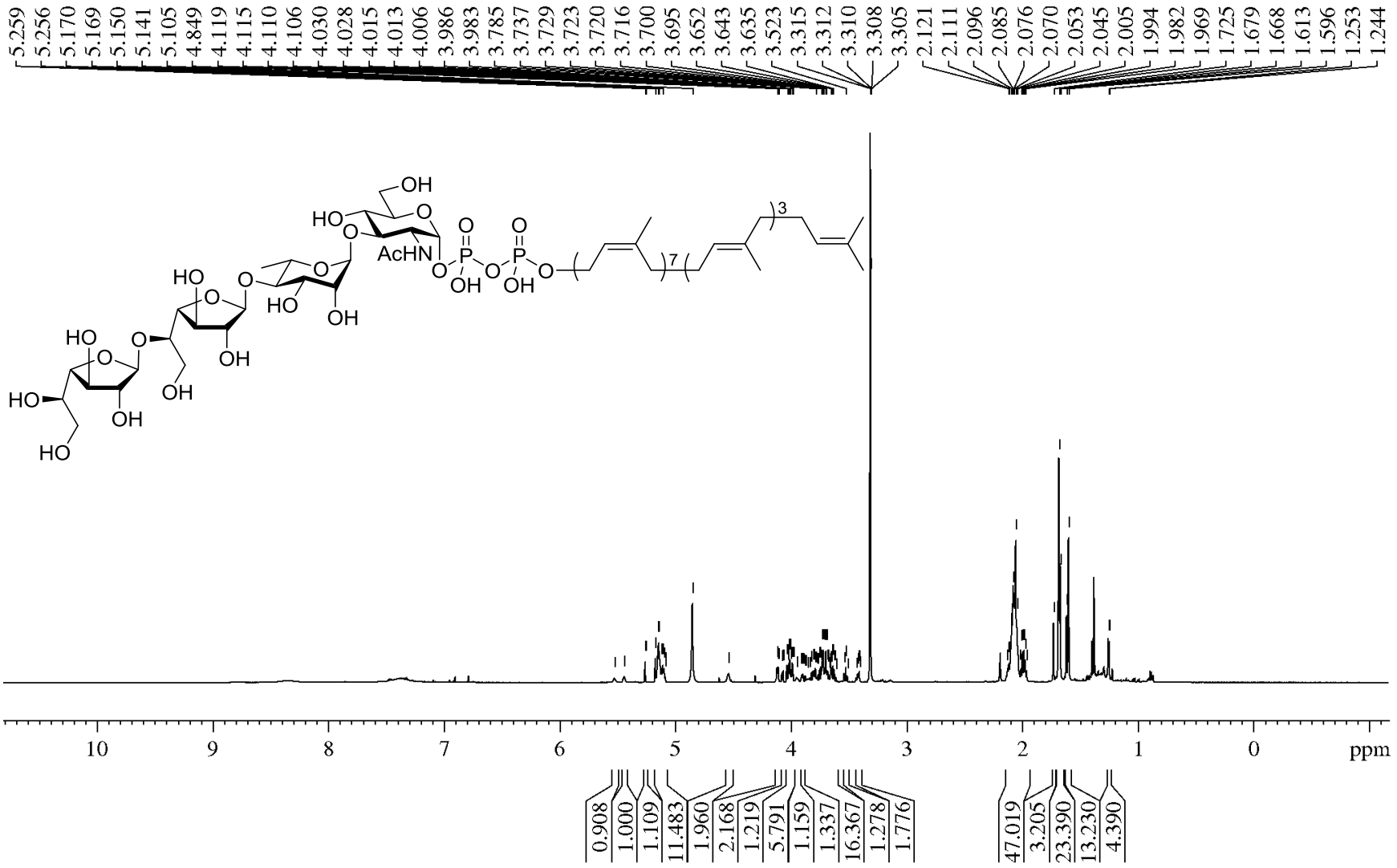
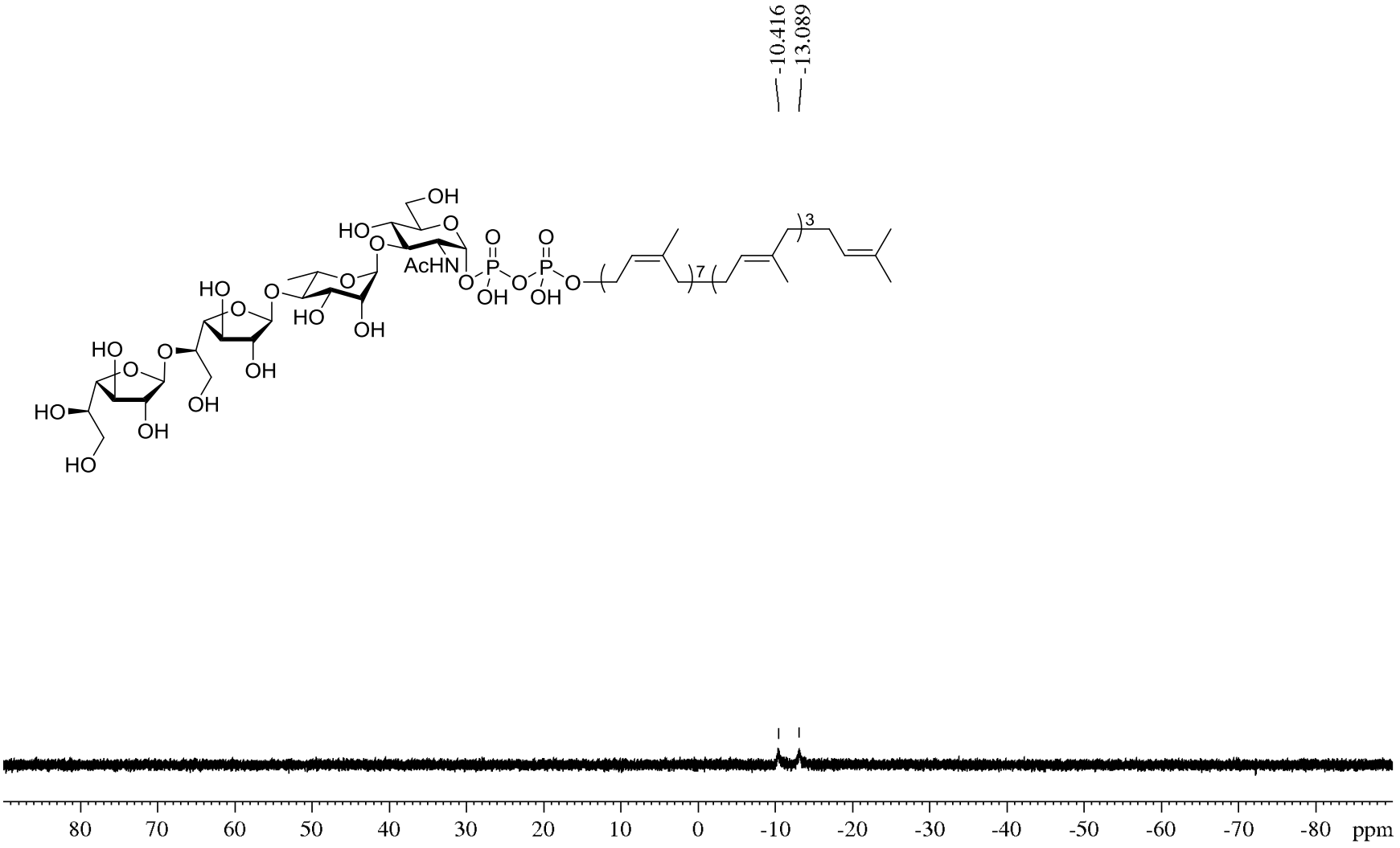


Figure S1. Negative ion mode ESI mass spectra of GlfT2-catalyzed elongation reactions with **22** and **23**. Peaks are labeled with the number of additional Gal*f* residues (s) in the product and the charge state of the ion (superscripted number). The peaks labelled p correspond to the starting material (**22** or **23**).

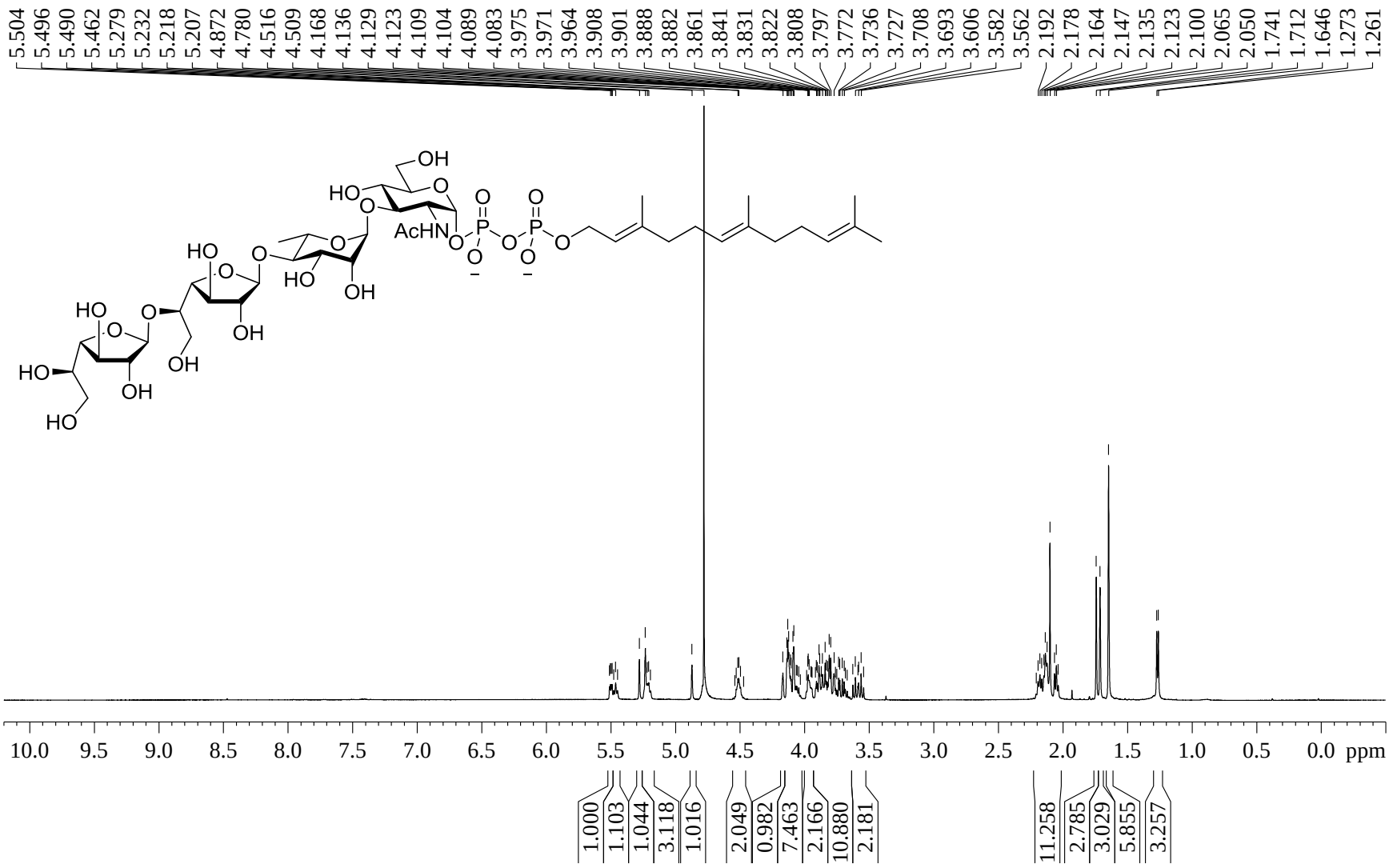
^{1}H NMR of Compound 3 (700 MHz, CD_3OD)



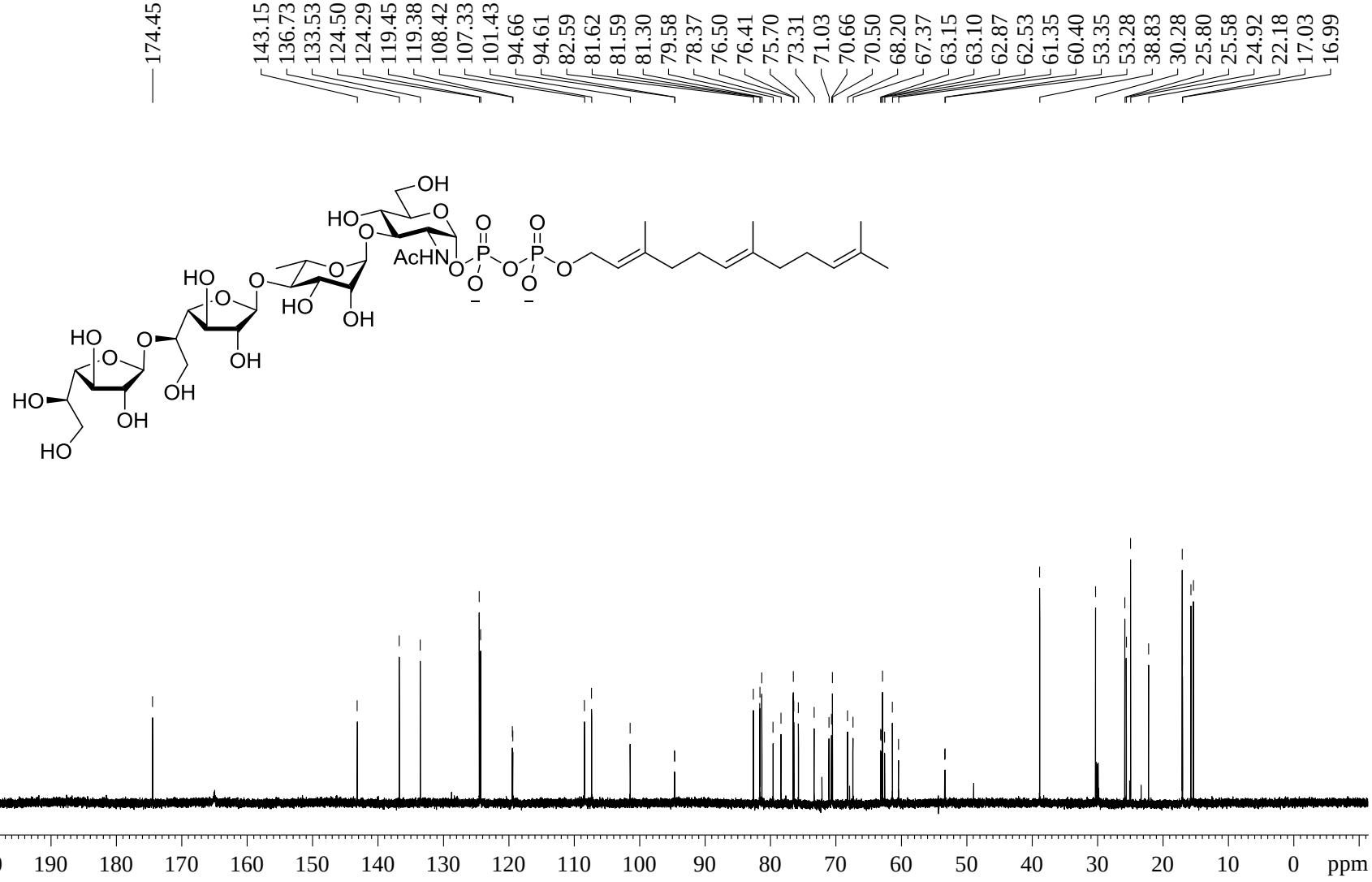
^{31}P NMR of Compound 3 (162 MHz, CD_3OD)



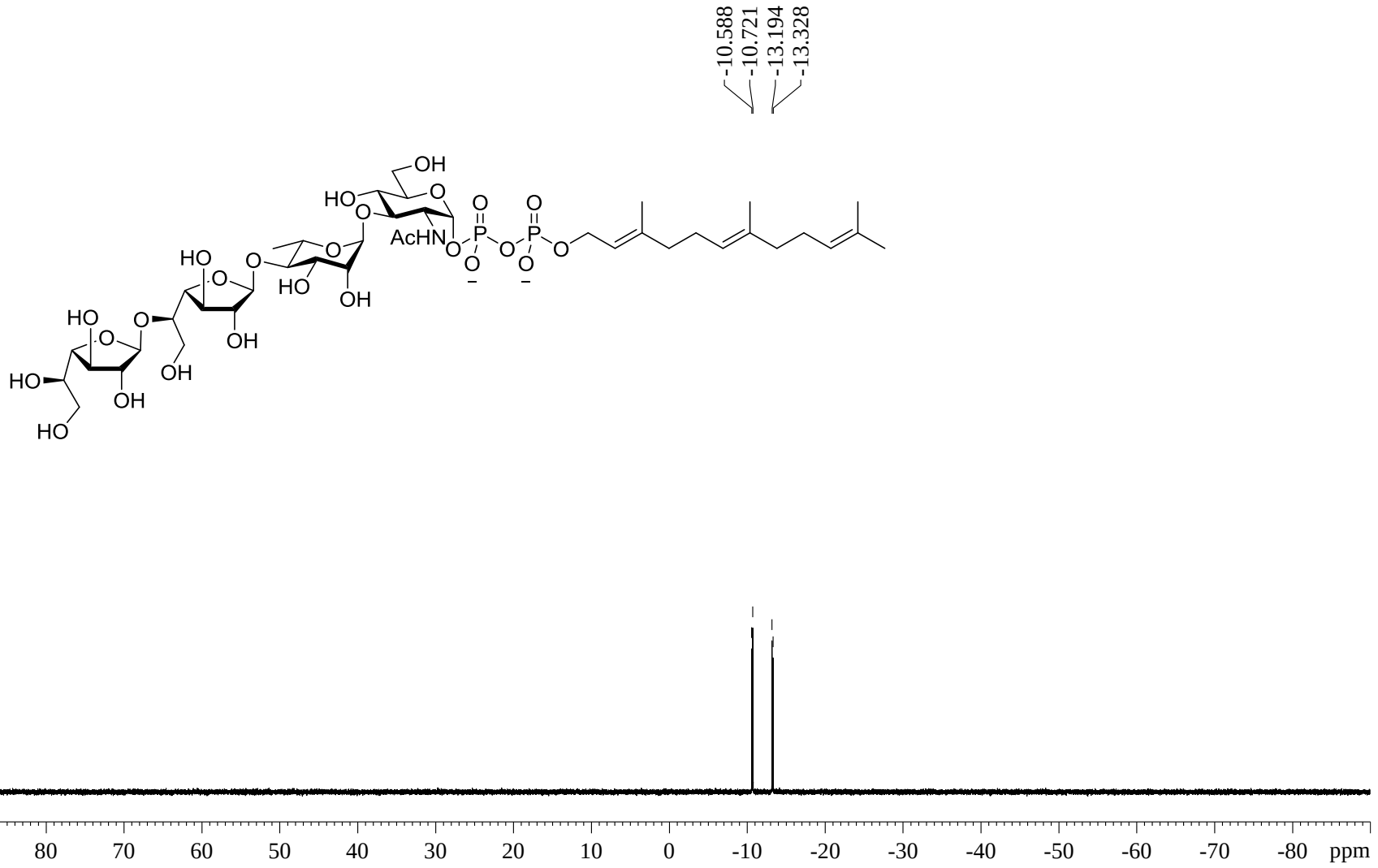
^{1}H NMR of Compound 4 (500 MHz, D_2O)



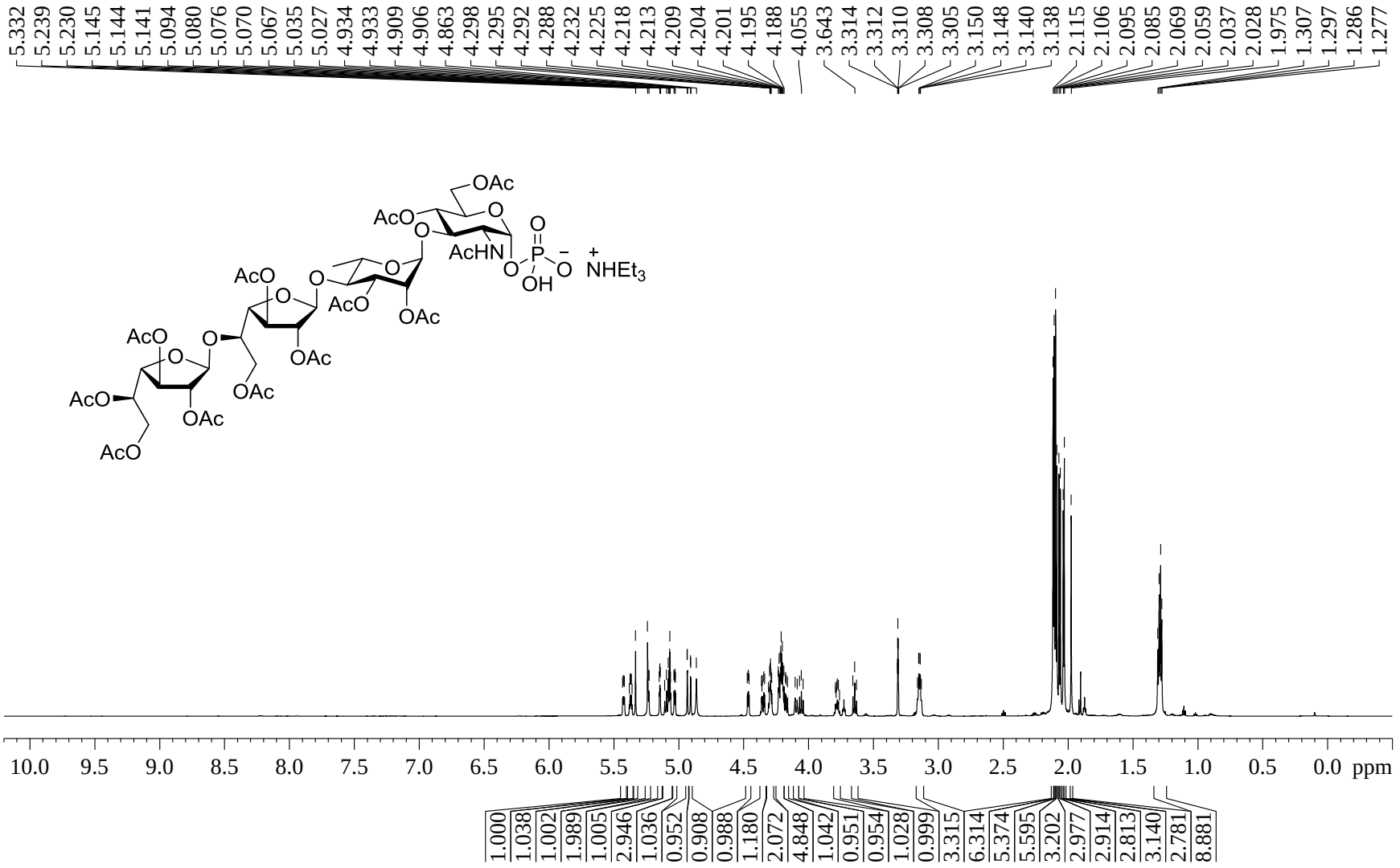
^{13}C NMR of Compound 4 (125 MHz, D_2O)



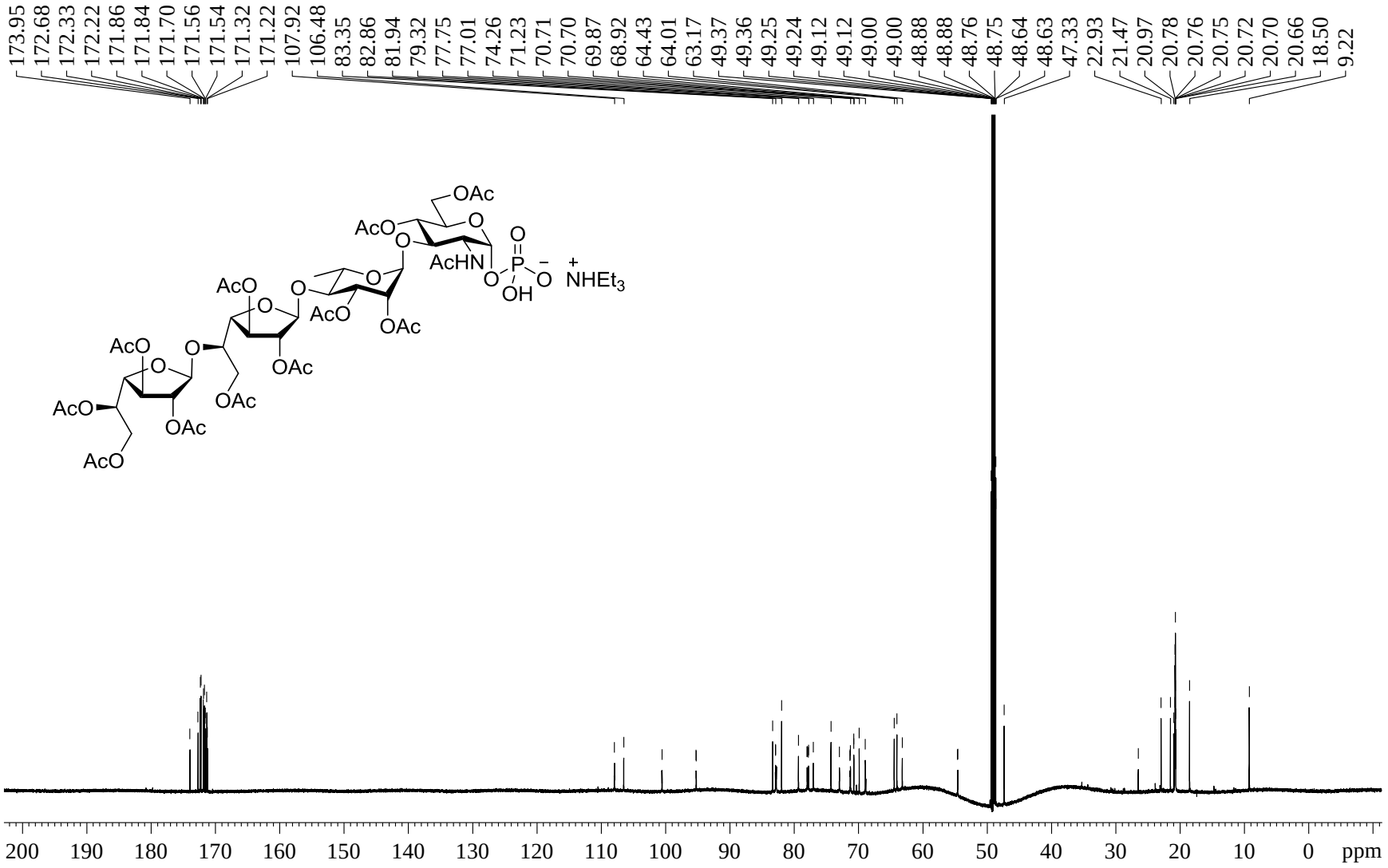
^{31}P NMR of Compound 4 (162 MHz, D_2O)



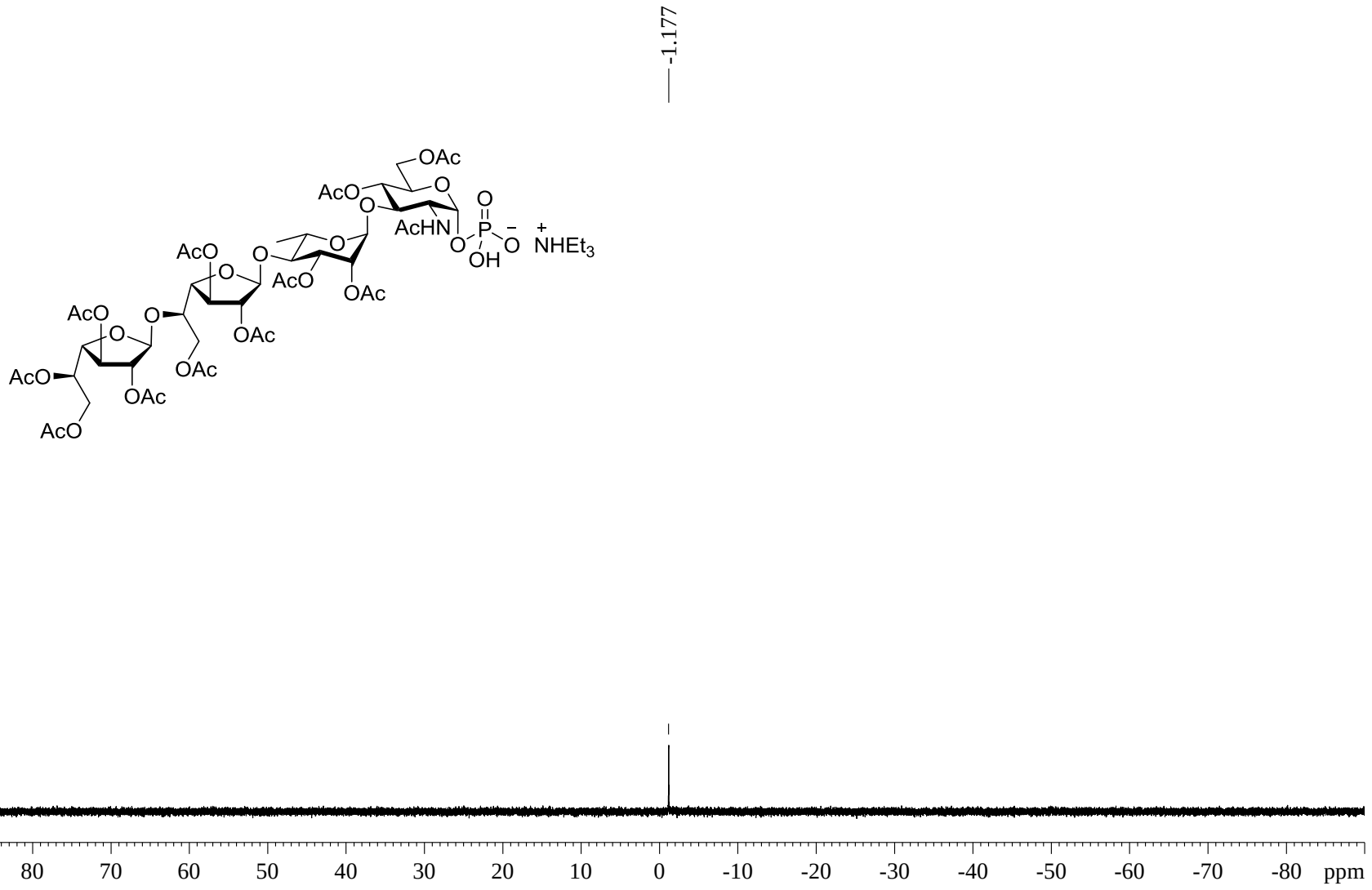
^{1}H NMR of Compound 5 (700 MHz, CD_3OD)



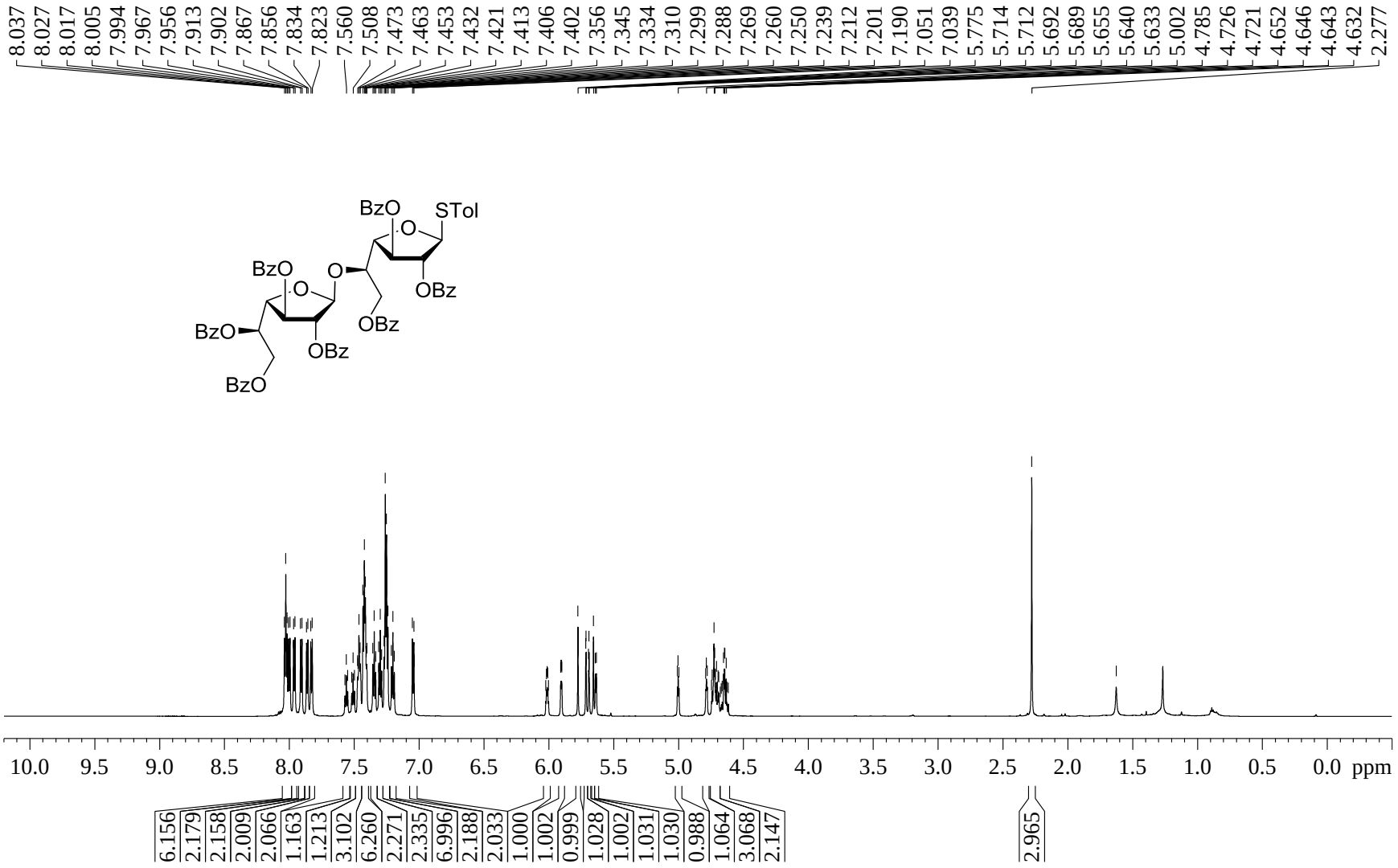
^{13}C NMR of Compound 5 (175 MHz, CD_3OD)



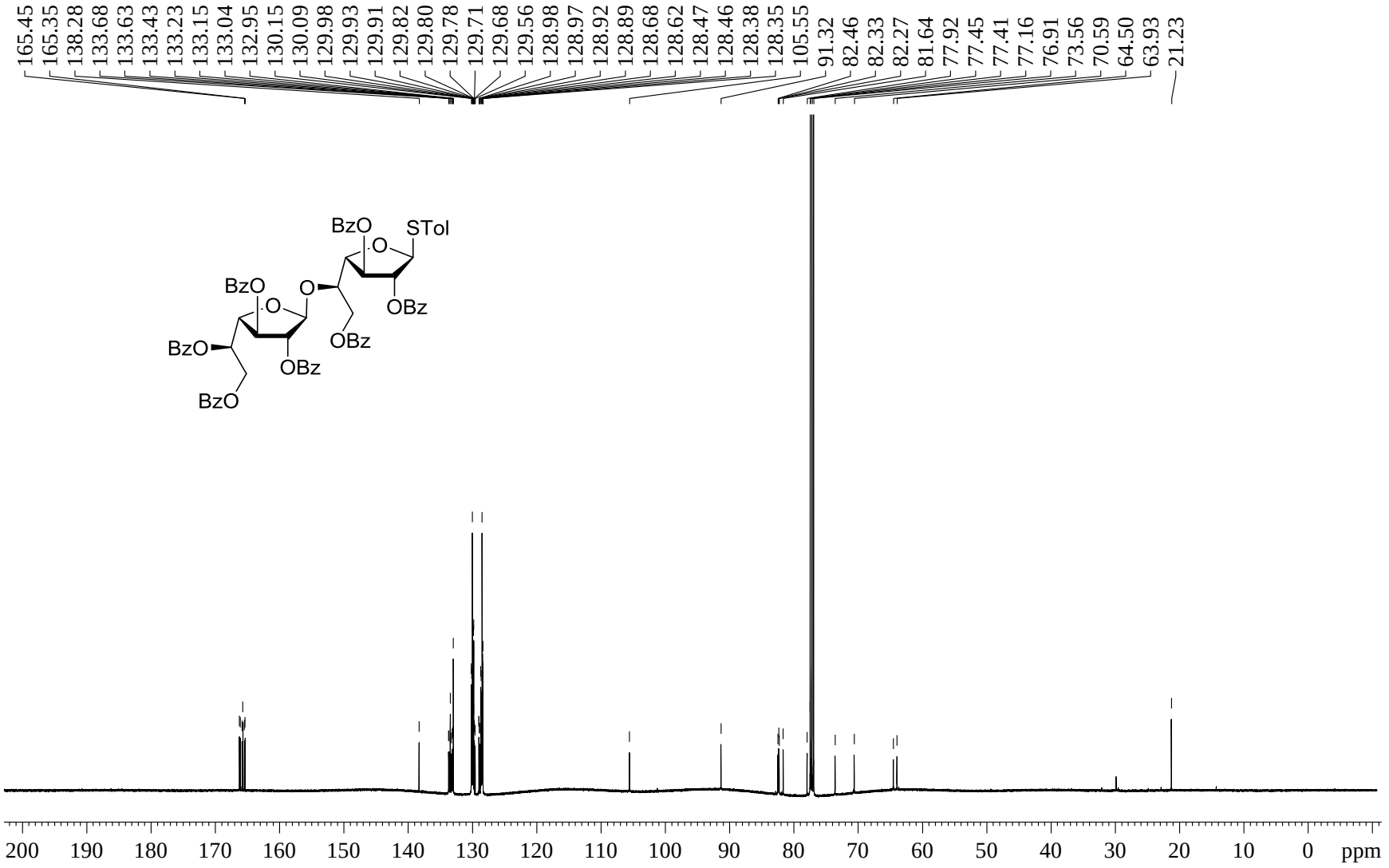
^{31}P NMR of Compound 5 (162 MHz, CD_3OD)



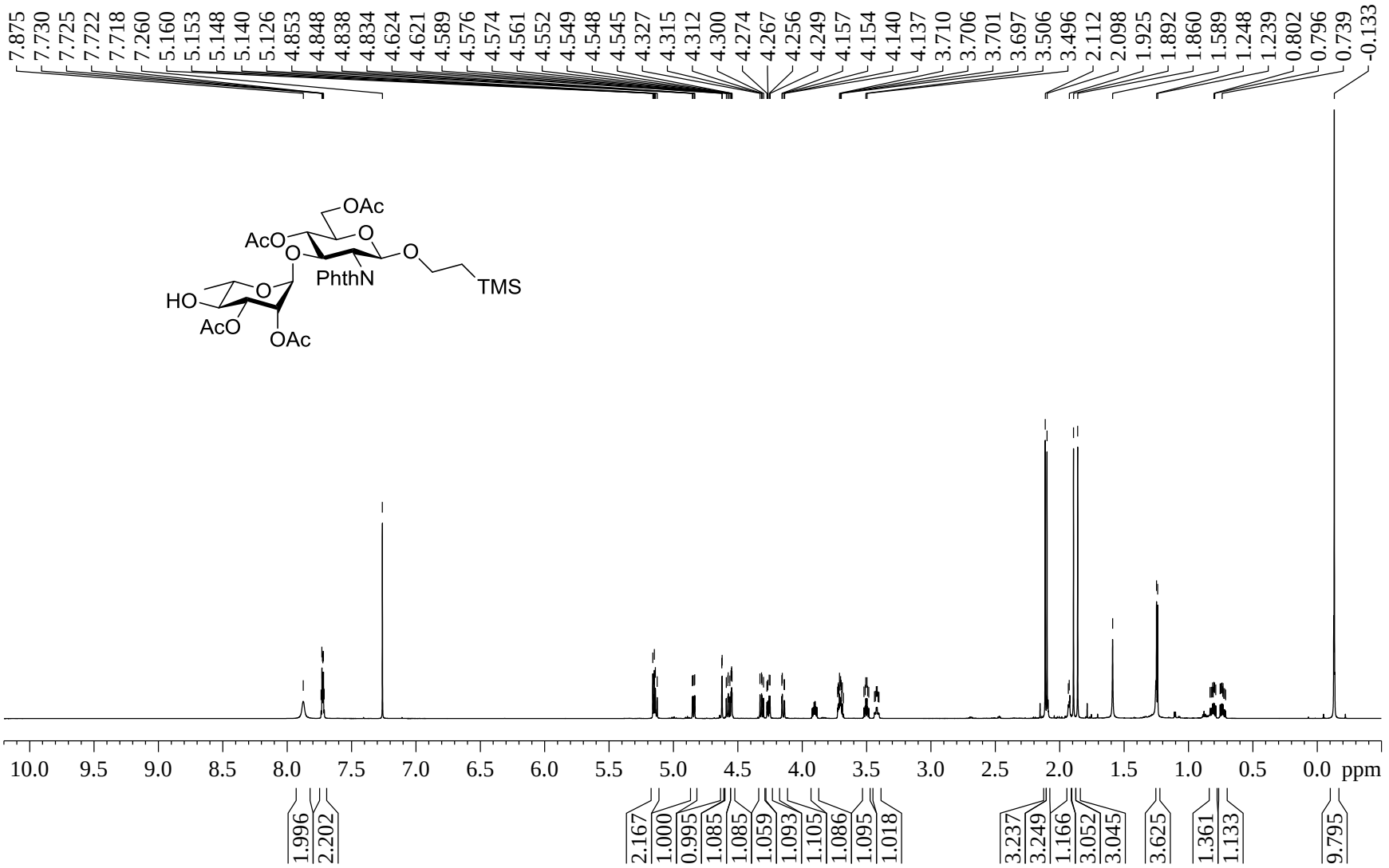
^{1}H NMR of Compound 7 (700 MHz, $CDCl_3$)



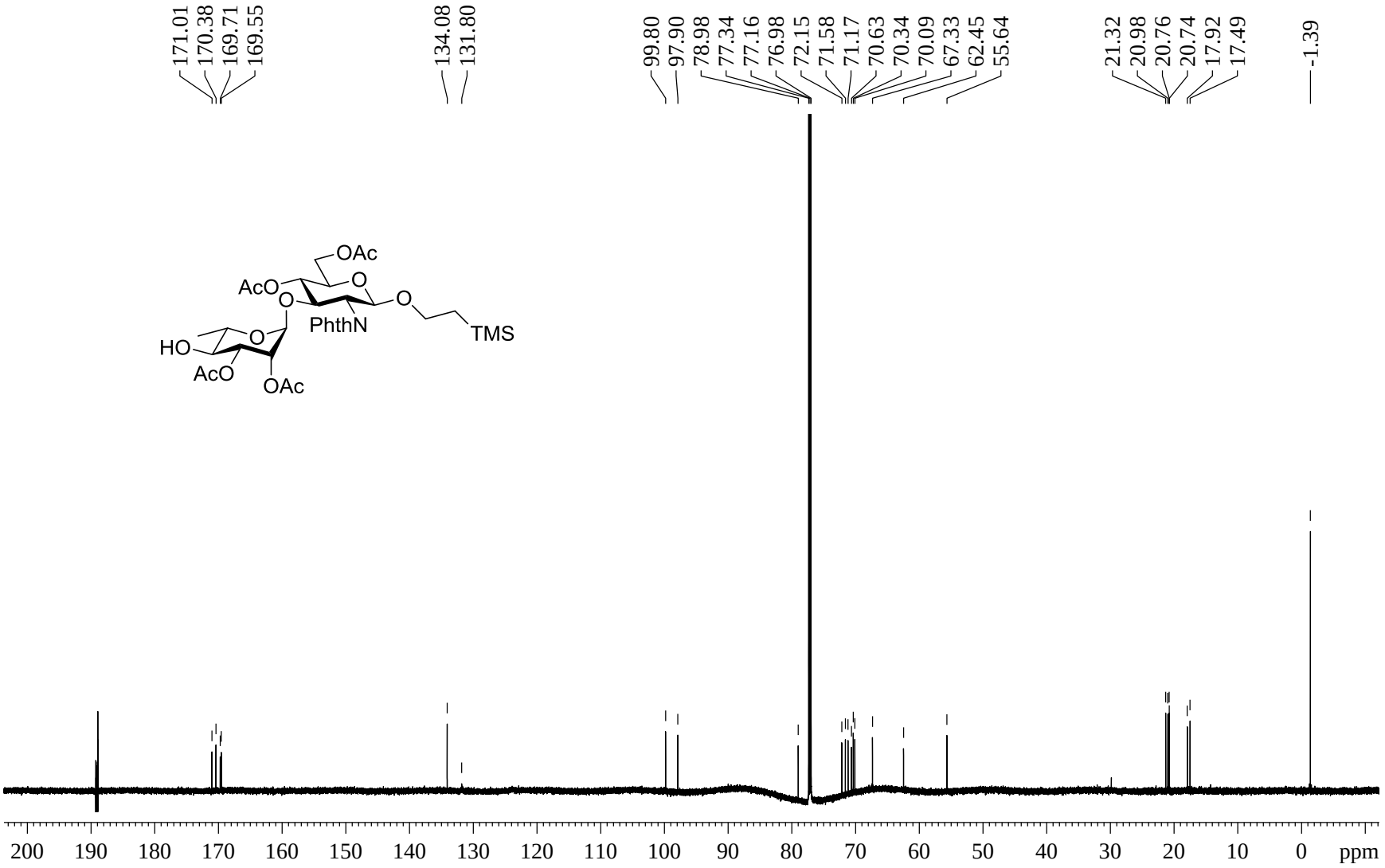
^{13}C NMR of Compound 7 (125 MHz, $CDCl_3$)



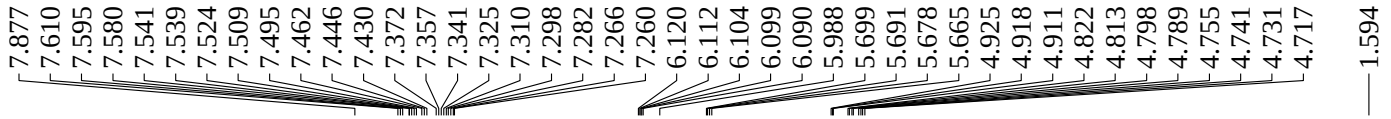
^{1}H NMR of Compound 8 (700 MHz, $CDCl_3$)

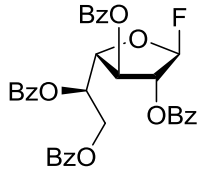


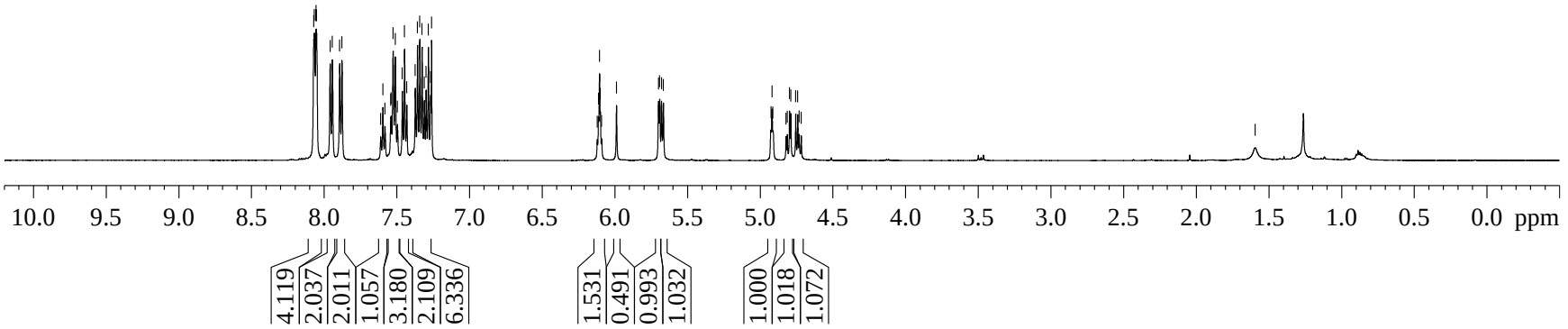
^{13}C NMR of Compound 8 (175 MHz, $CDCl_3$)



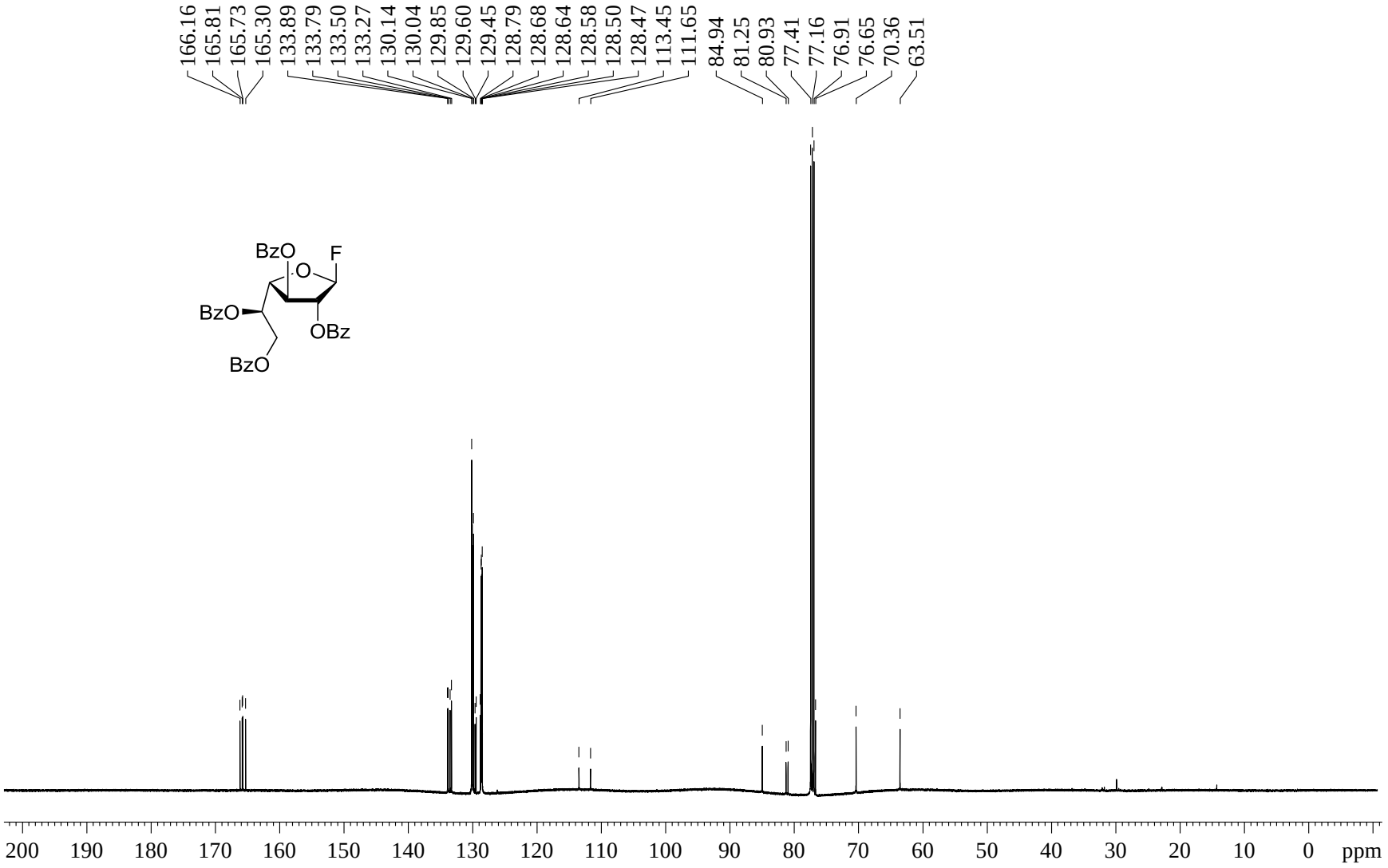
^{1}H NMR of Compound 9 (500 MHz, $CDCl_3$)



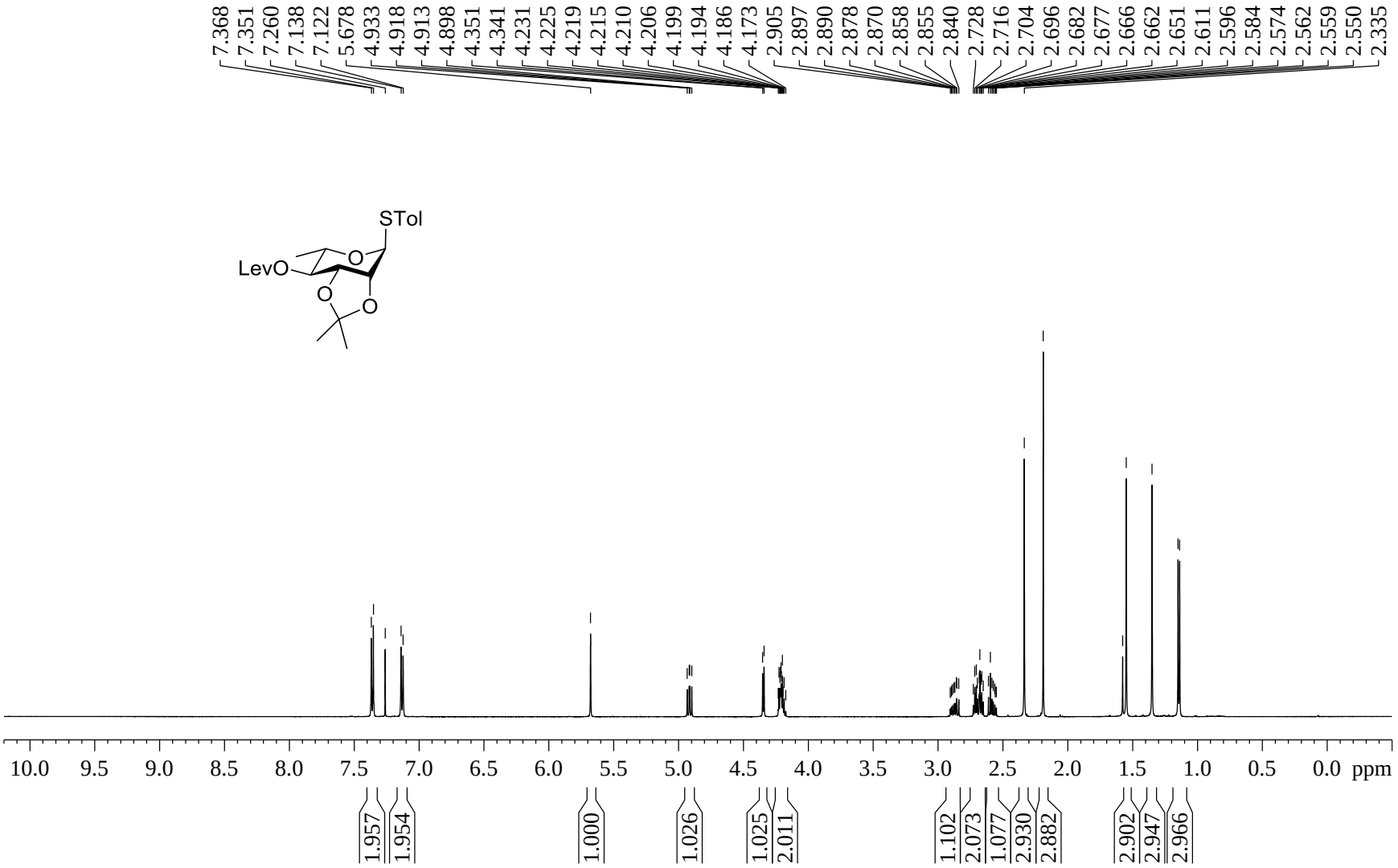




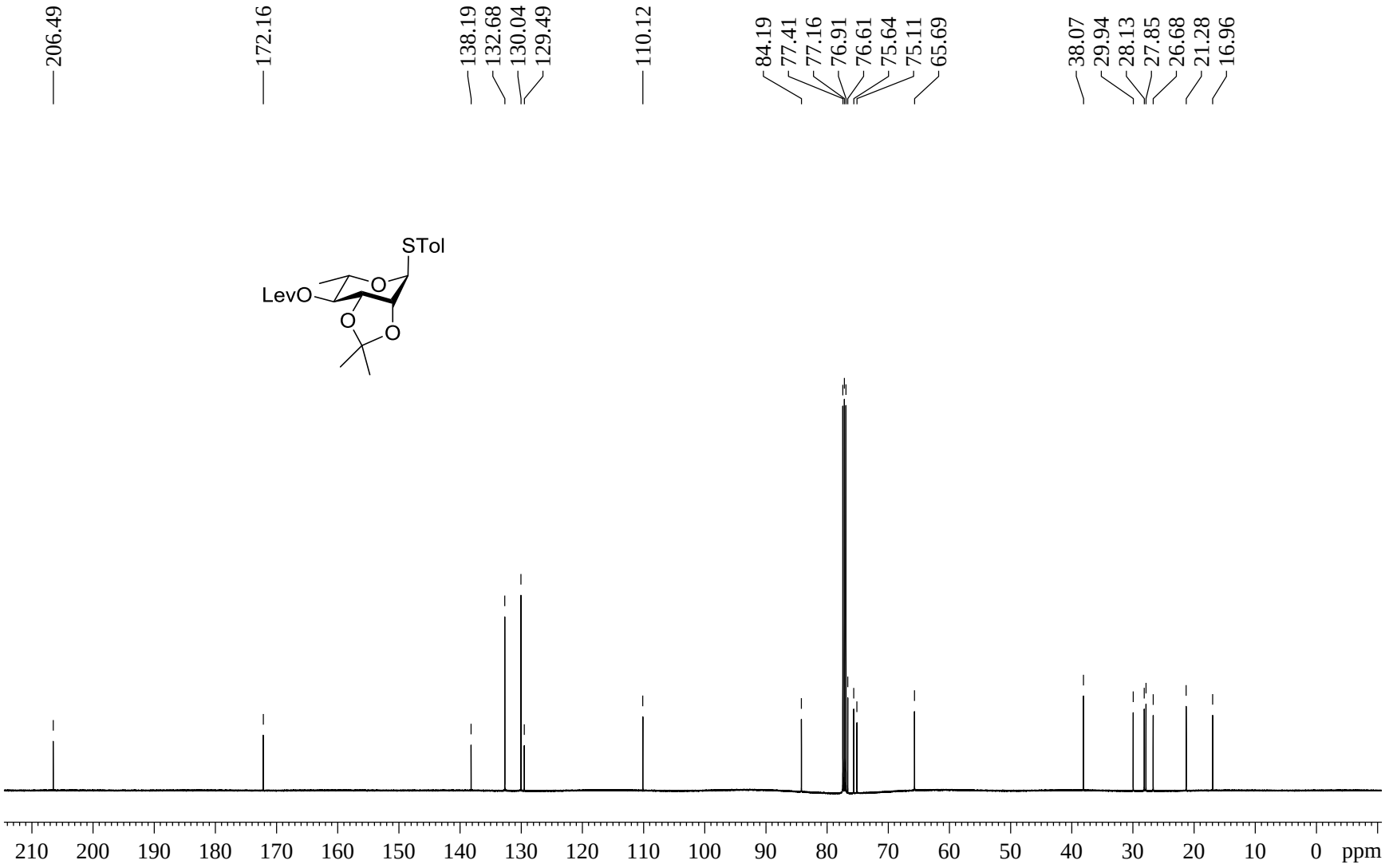
^{13}C NMR of Compound 9 (125 MHz, $CDCl_3$)



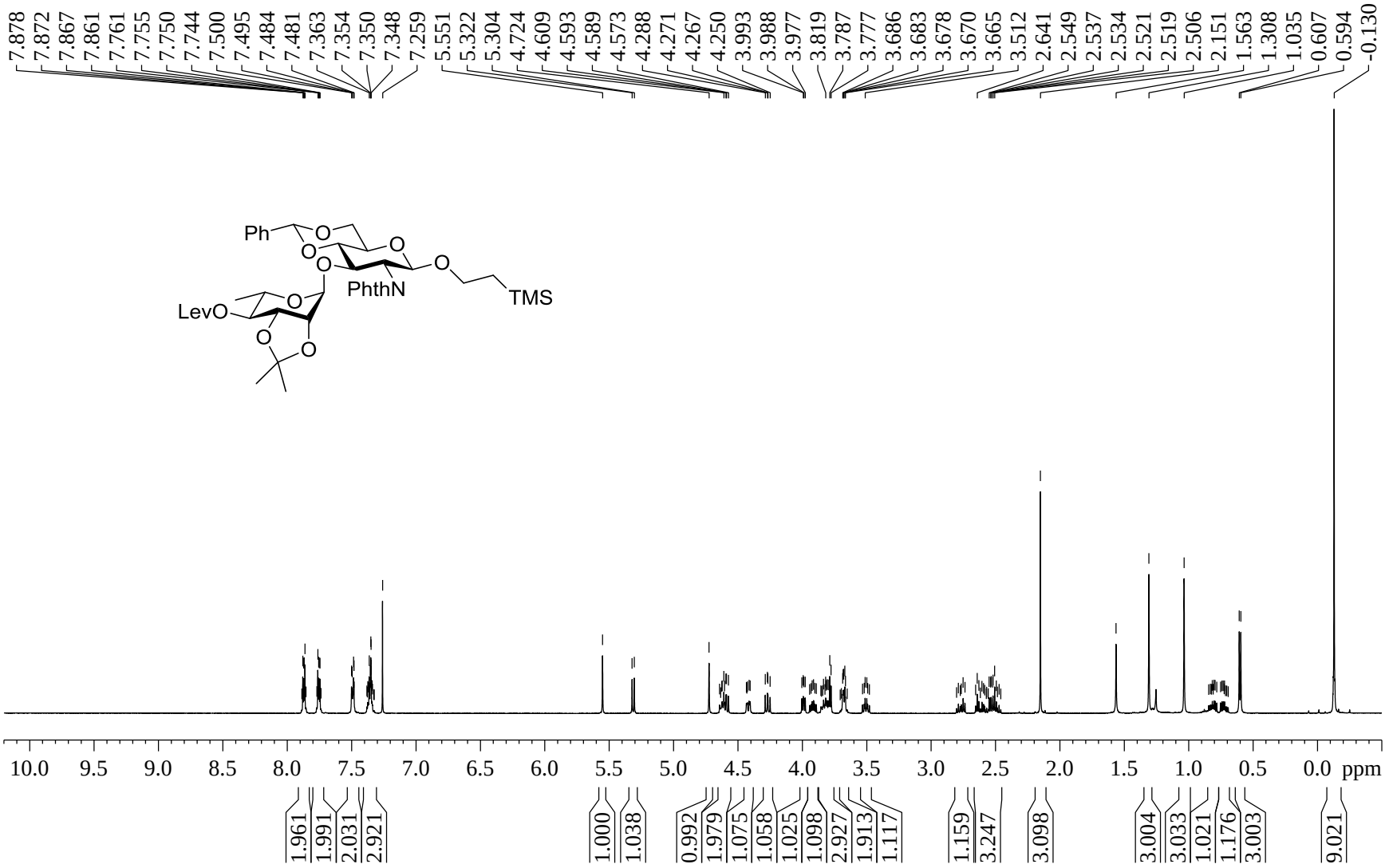
^{1}H NMR of Compound 11 (500 MHz, $CDCl_3$)



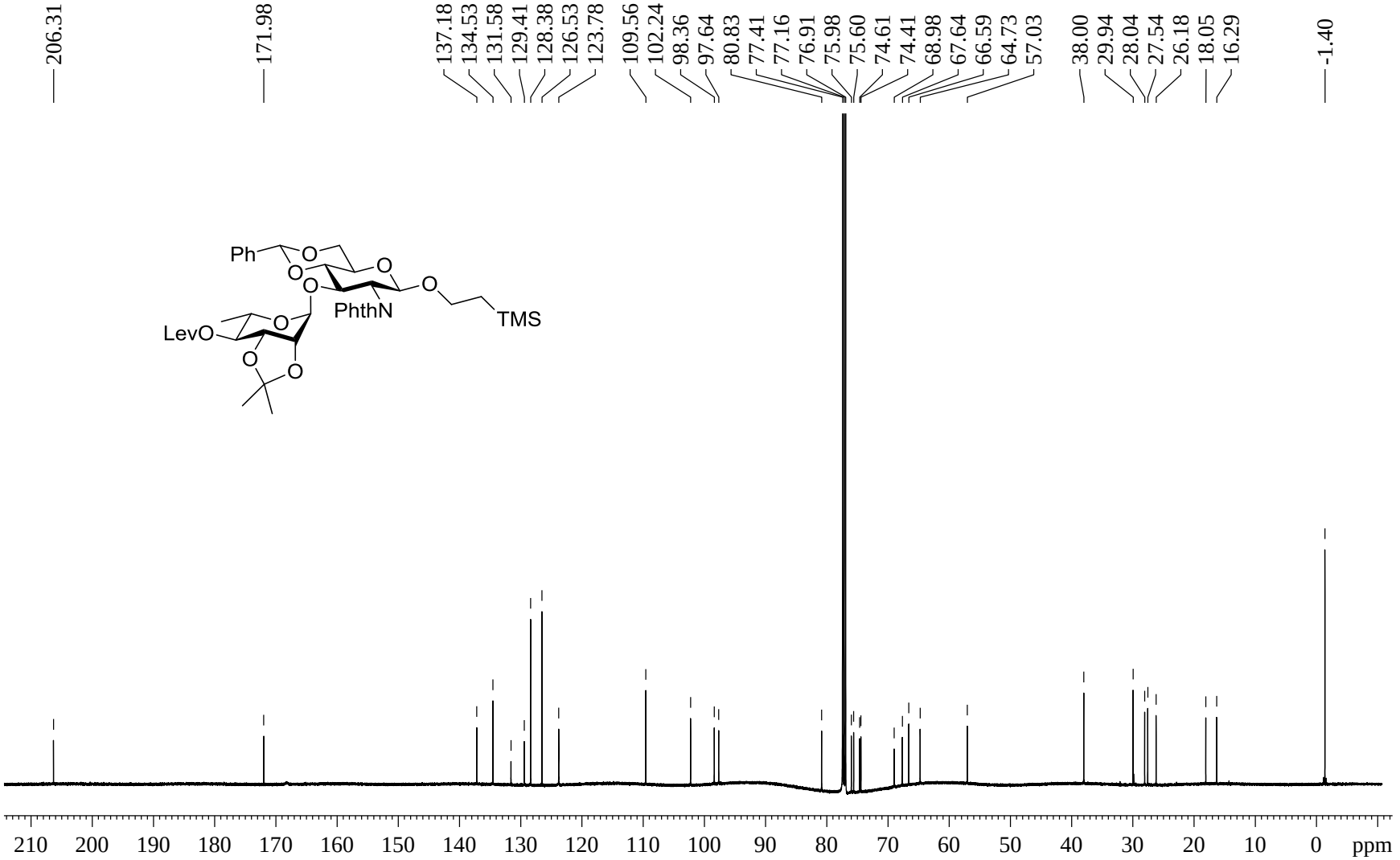
^{13}C NMR of Compound 11 (125 MHz, $CDCl_3$)



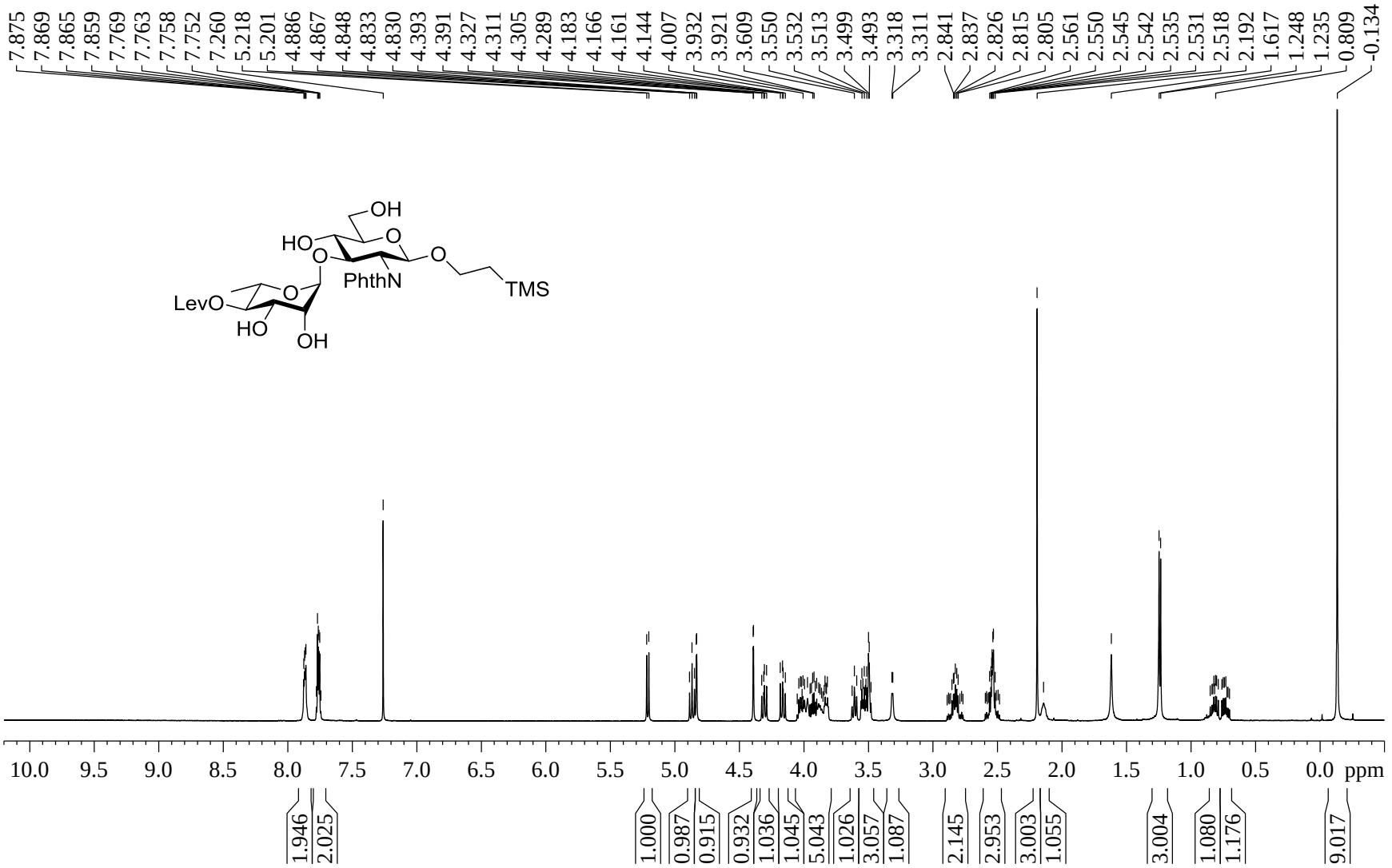
^{1}H NMR of Compound 15 (500 MHz, $CDCl_3$)



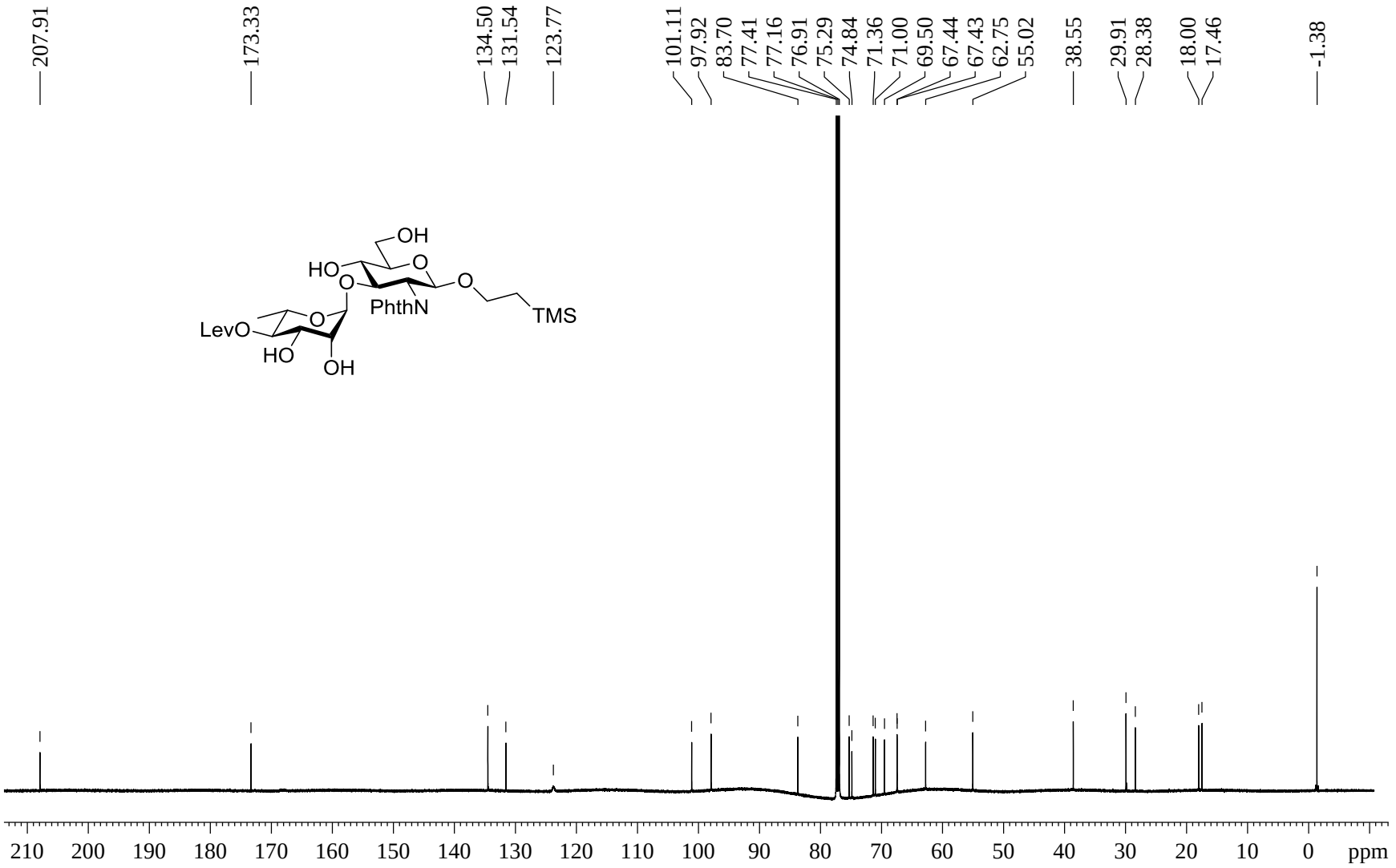
^{13}C NMR of Compound 15 (125 MHz, $CDCl_3$)



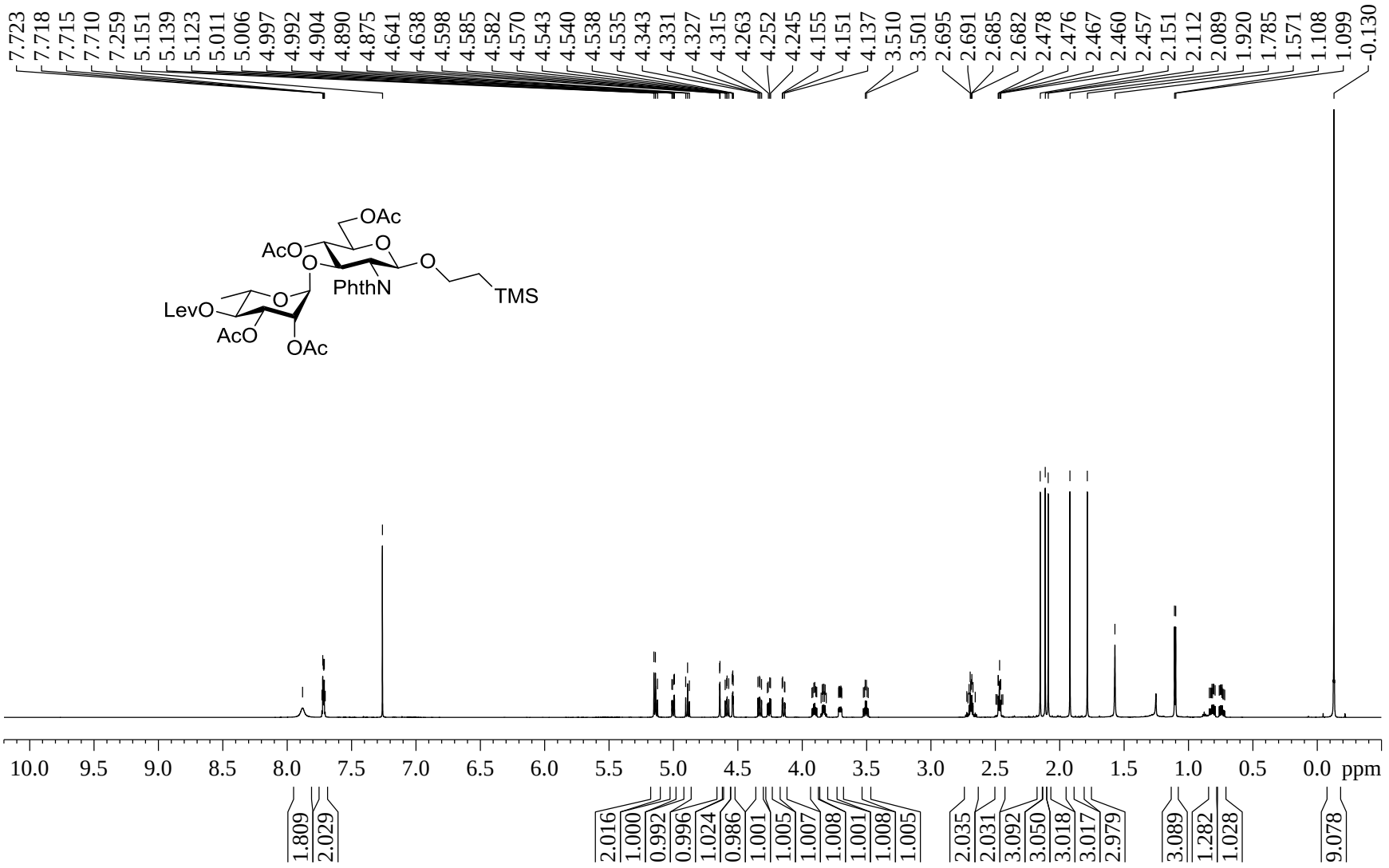
^{1}H NMR of Compound 16 (500 MHz, $CDCl_3$)



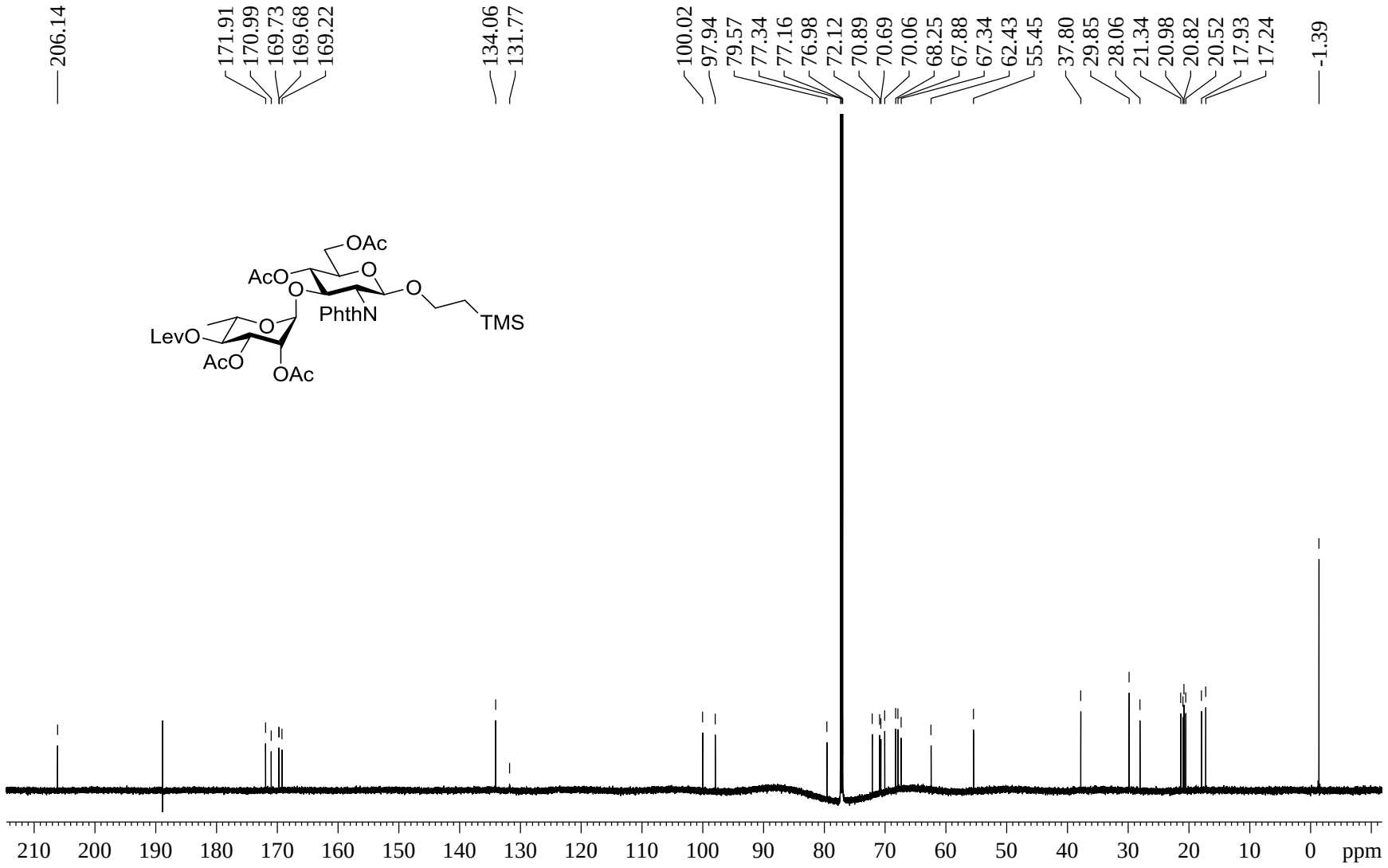
^{13}C NMR of Compound 16 (125 MHz, $CDCl_3$)



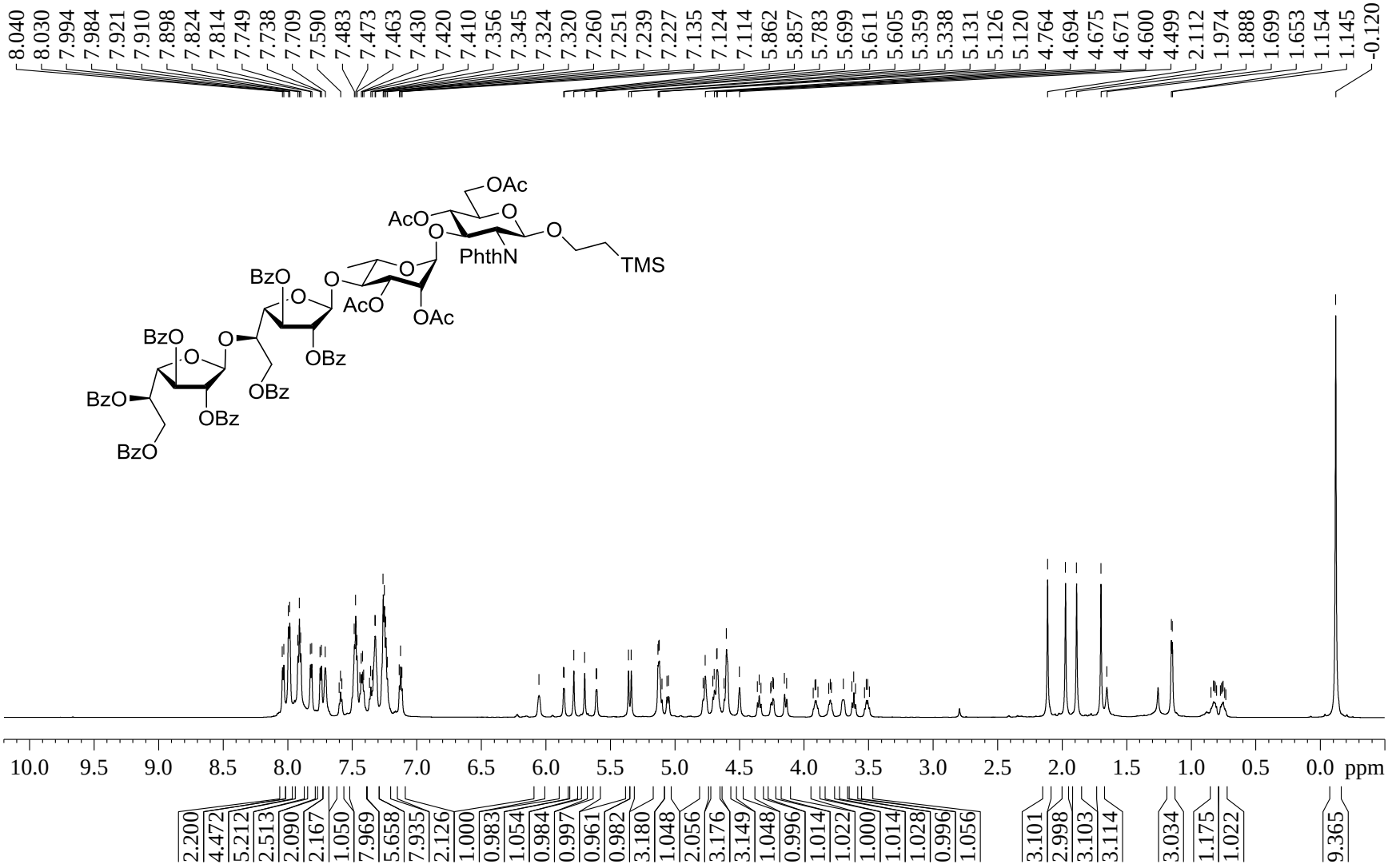
^{1}H NMR of Compound 17 (700 MHz, $CDCl_3$)



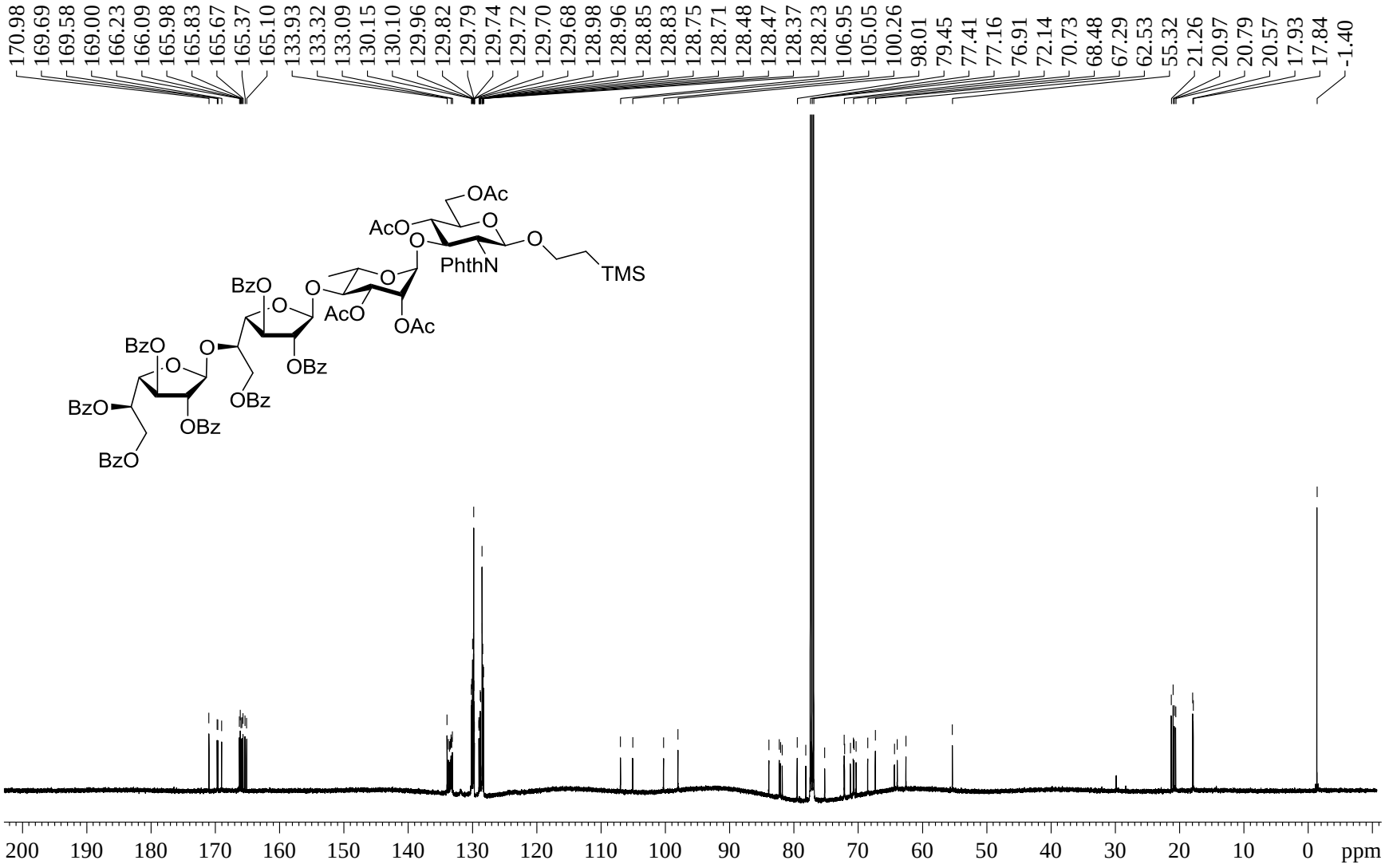
^{13}C NMR of Compound 17 (175 MHz, $CDCl_3$)



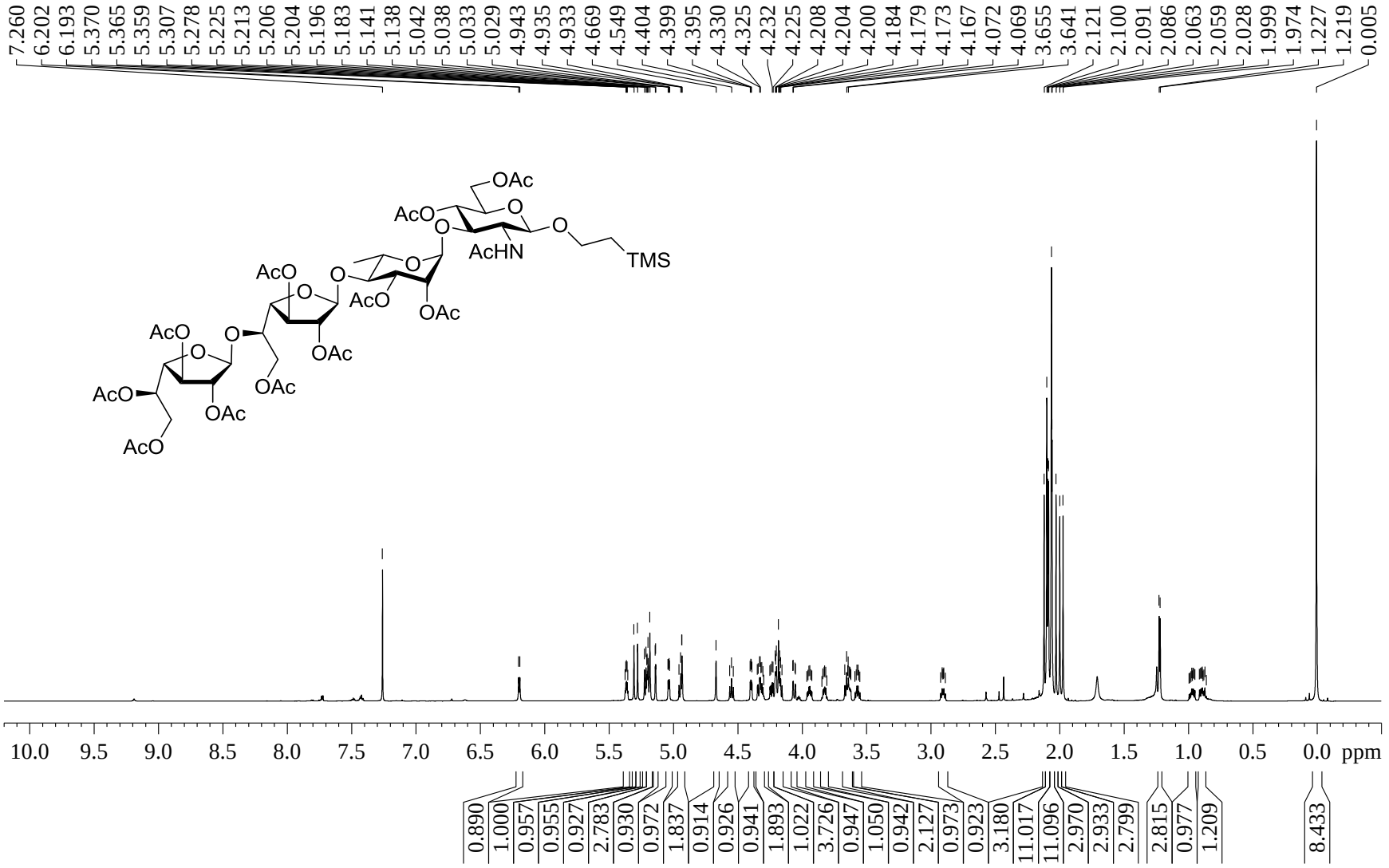
^{1}H NMR of Compound 18 (700 MHz, $CDCl_3$)



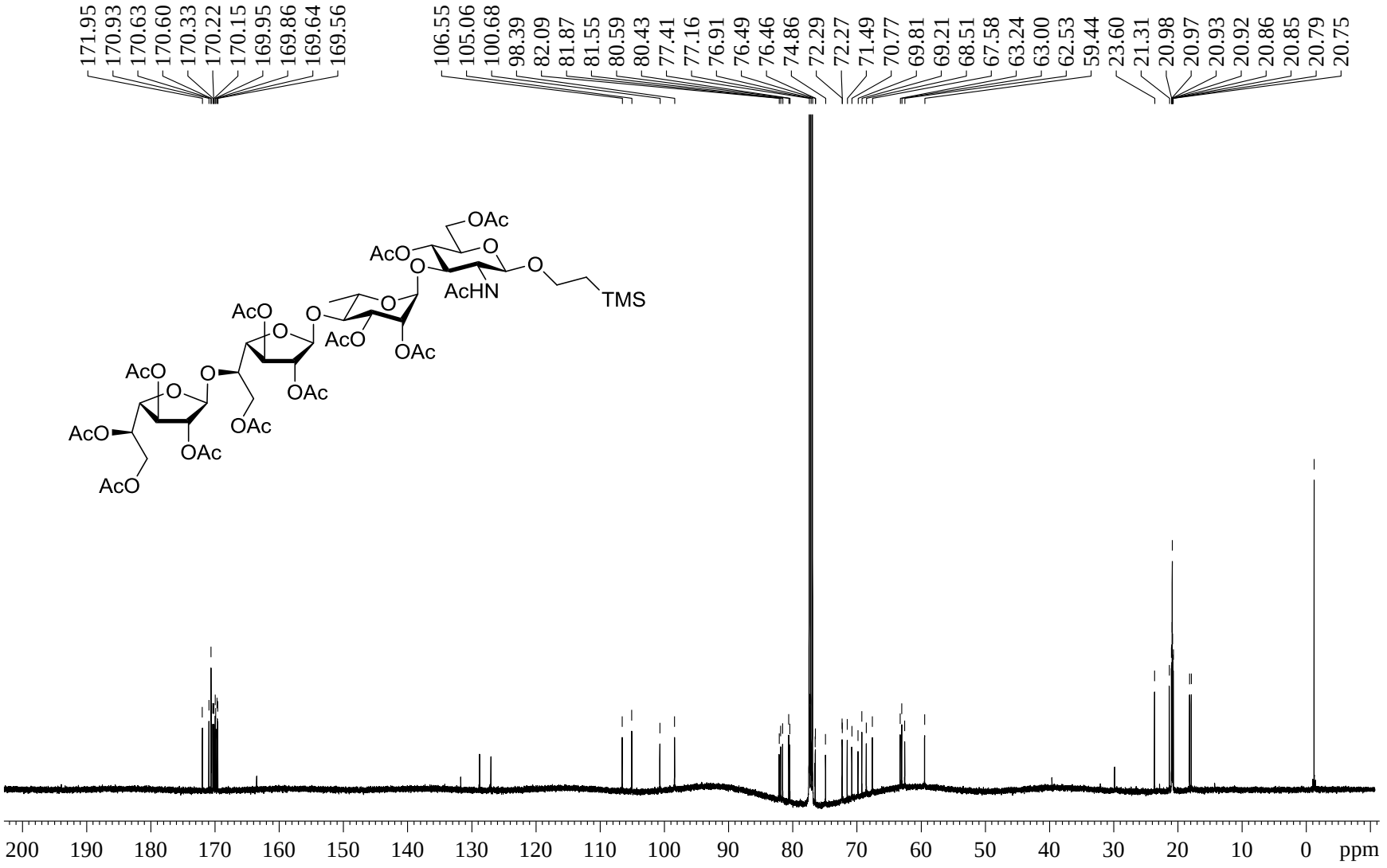
^{13}C NMR of Compound 18 (125 MHz, $CDCl_3$)



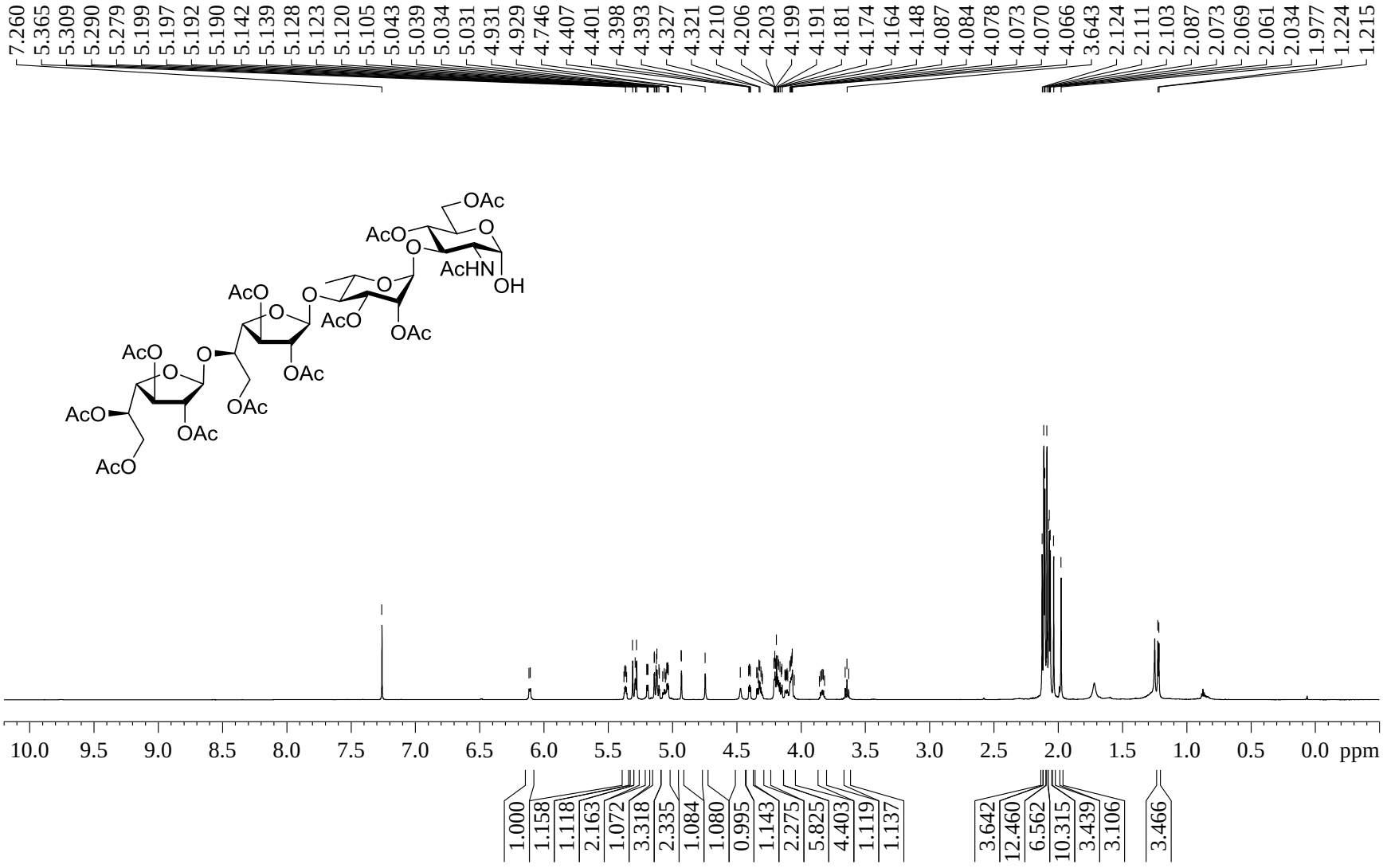
^{1}H NMR of Compound 19 (700 MHz, $CDCl_3$)



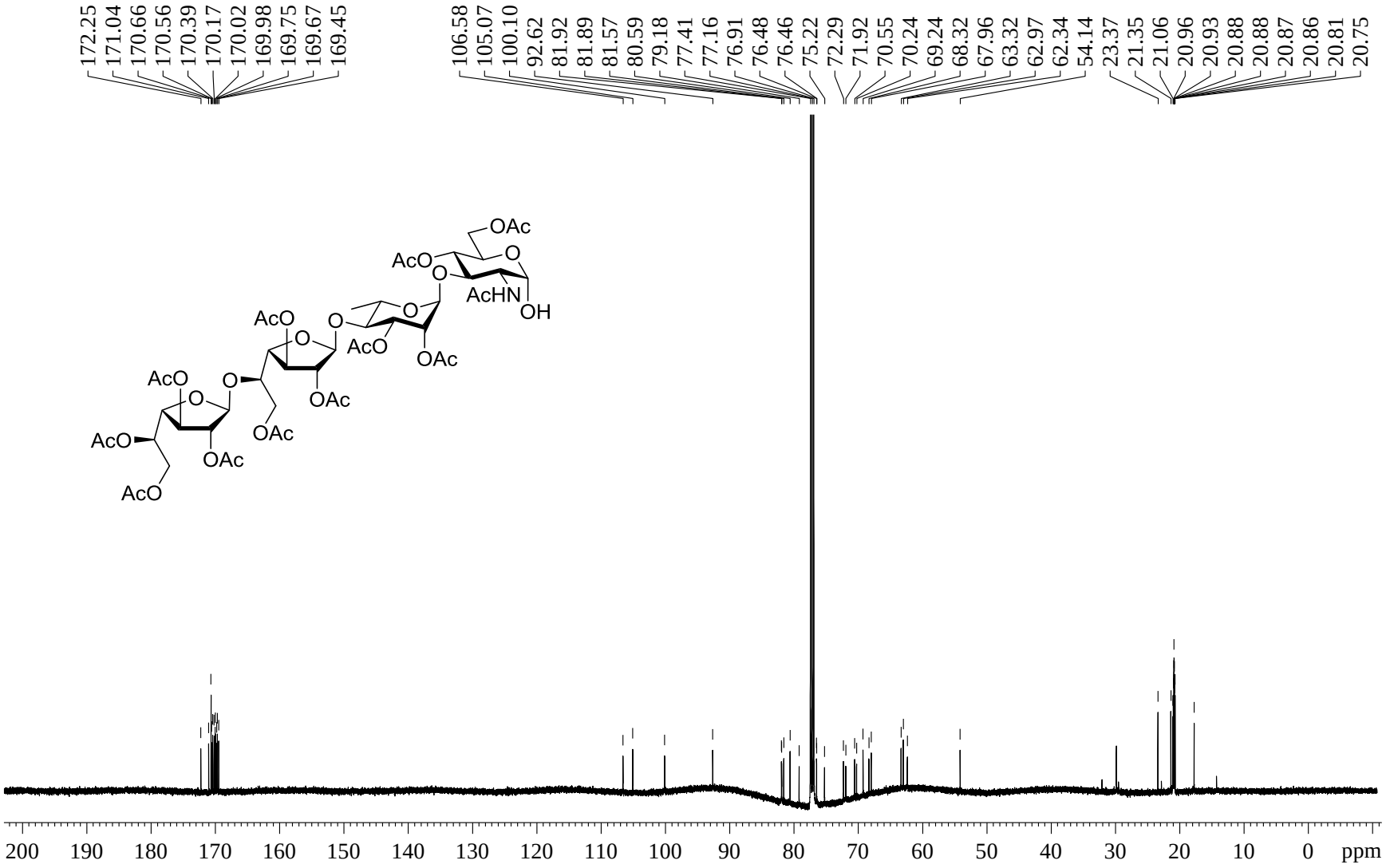
^{13}C NMR of Compound 19 (125 MHz, $CDCl_3$)



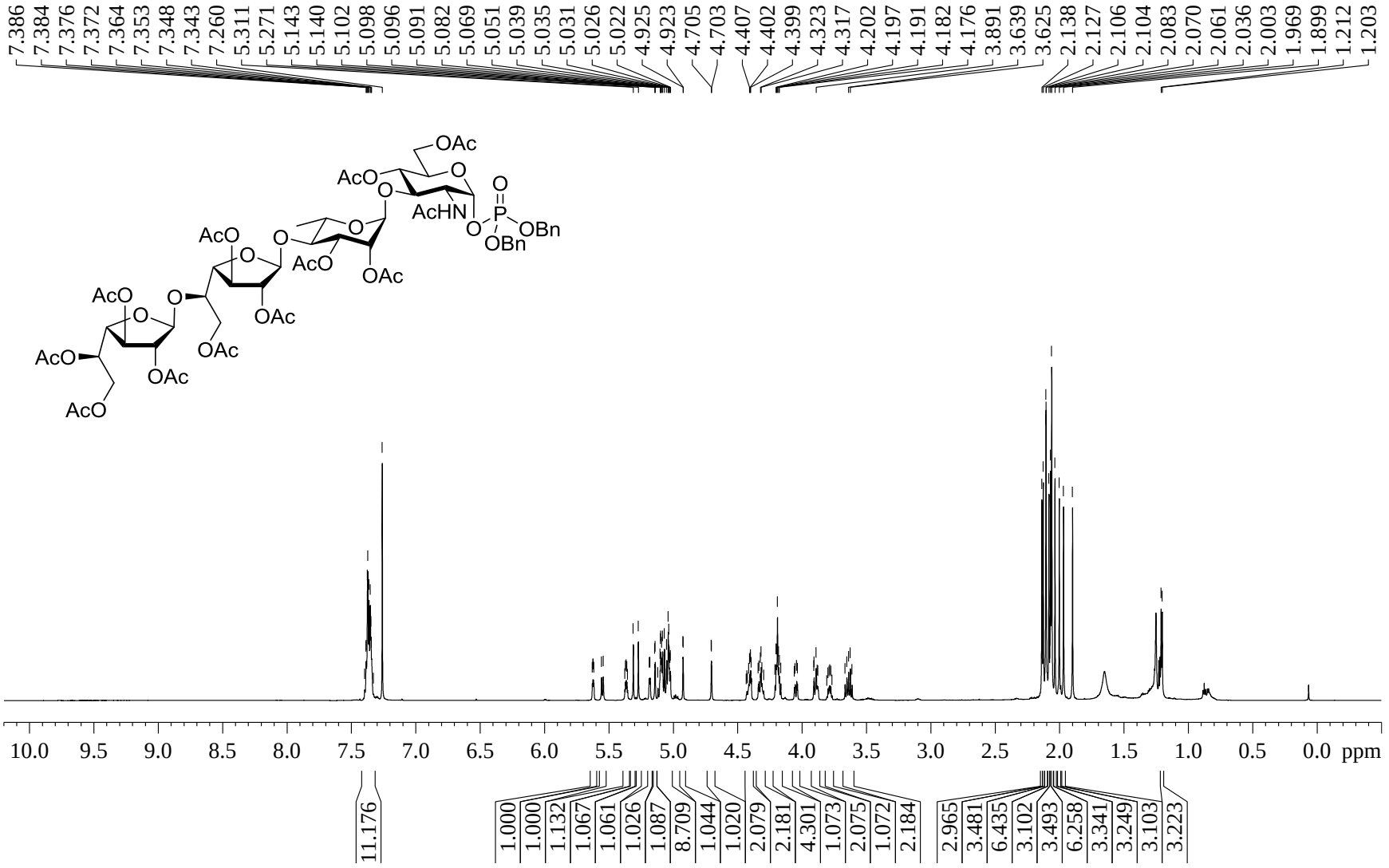
^{1}H NMR of Compound 20 (700 MHz, $CDCl_3$)



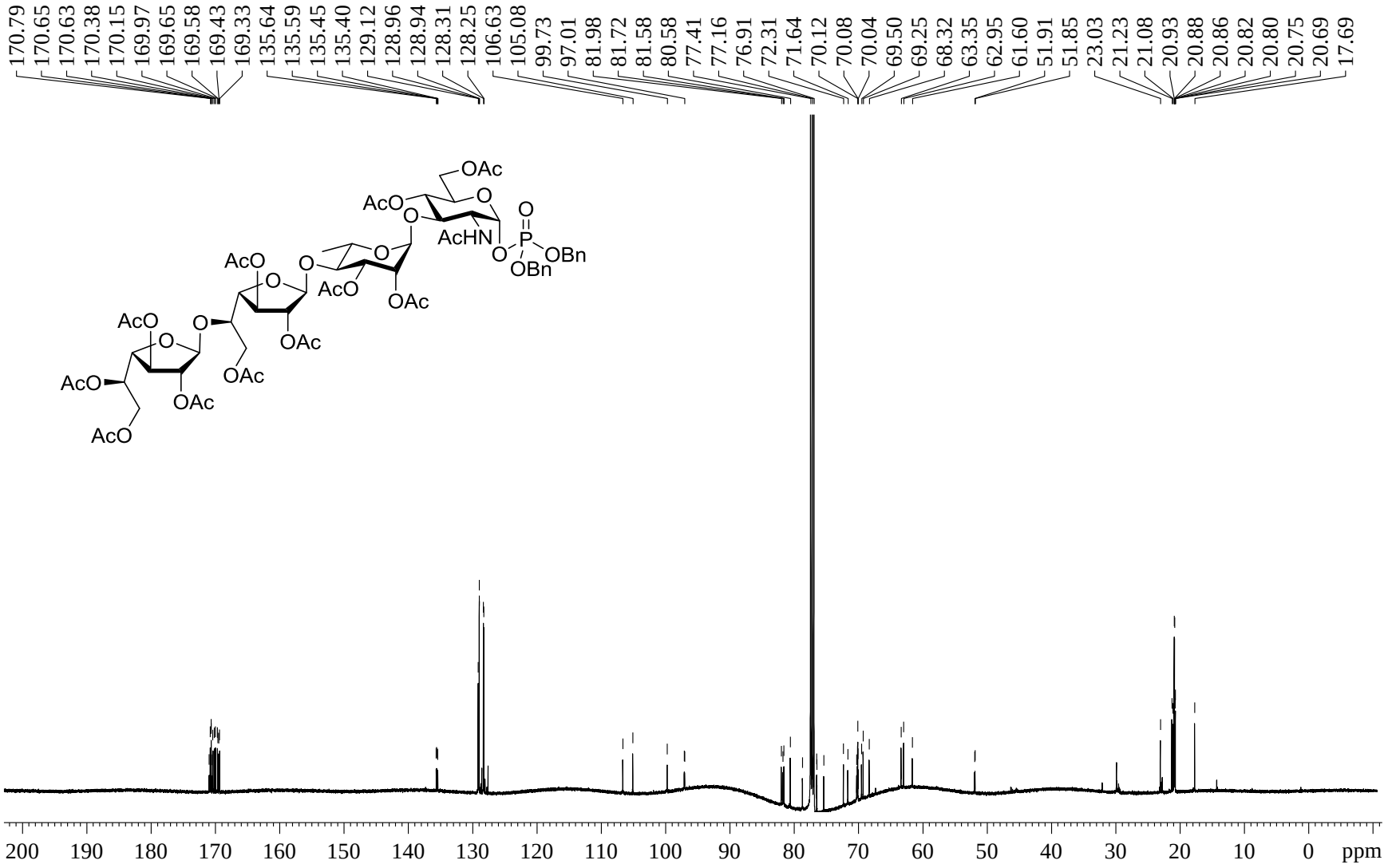
^{13}C NMR of Compound 20 (125 MHz, $CDCl_3$)



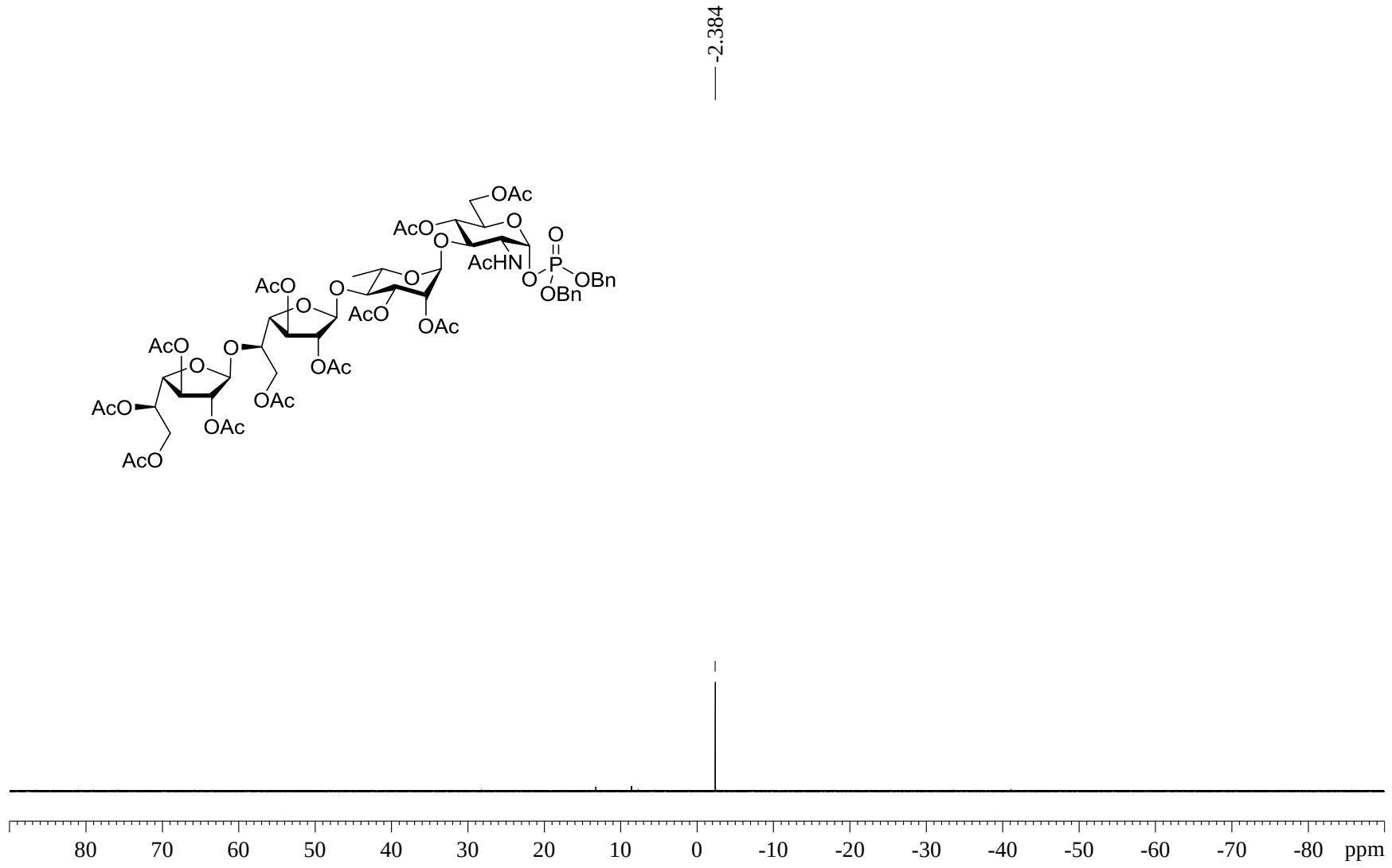
^{1}H NMR of Compound 21 (700 MHz, $CDCl_3$)



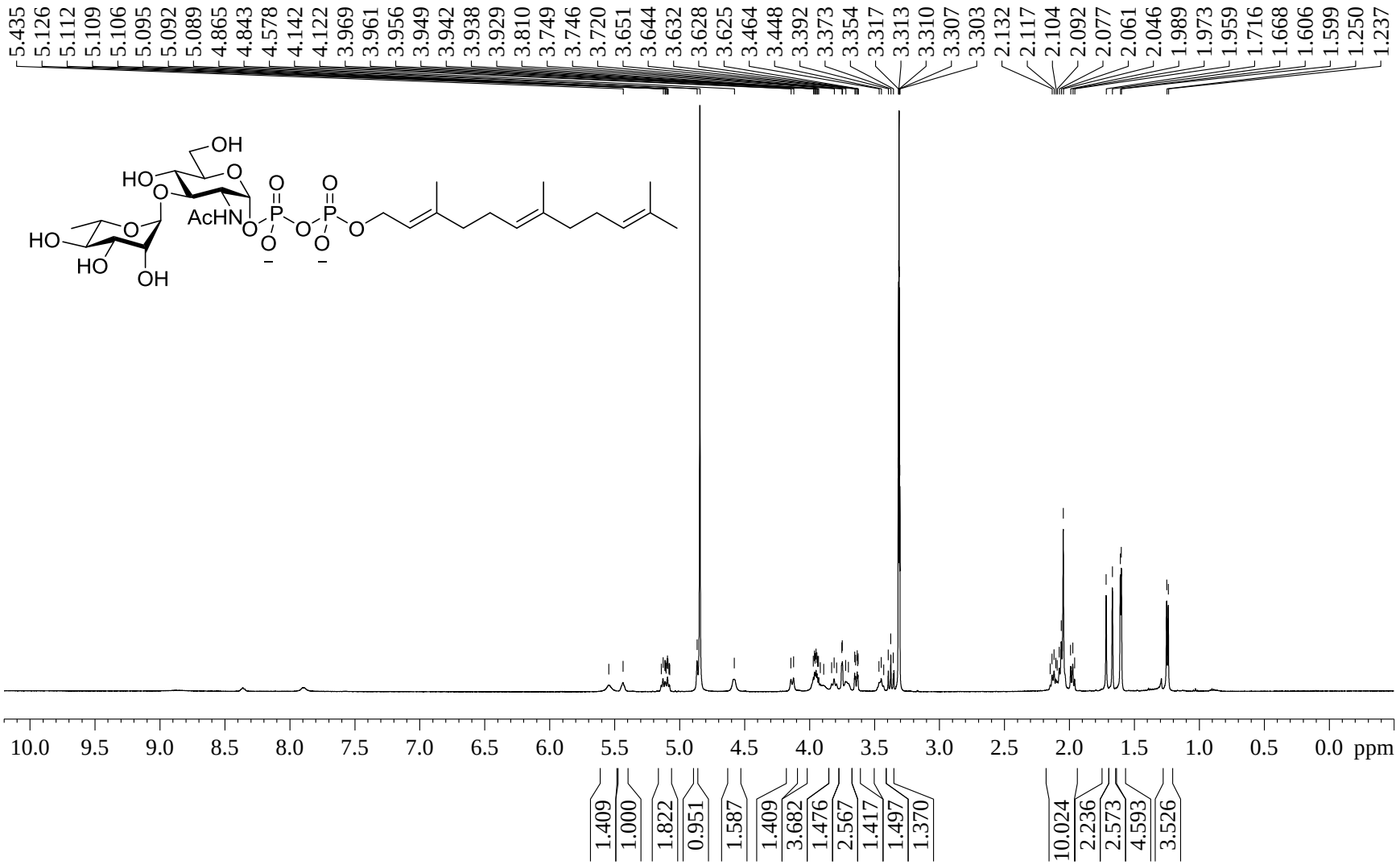
^{13}C NMR of Compound 21 (125 MHz, $CDCl_3$)



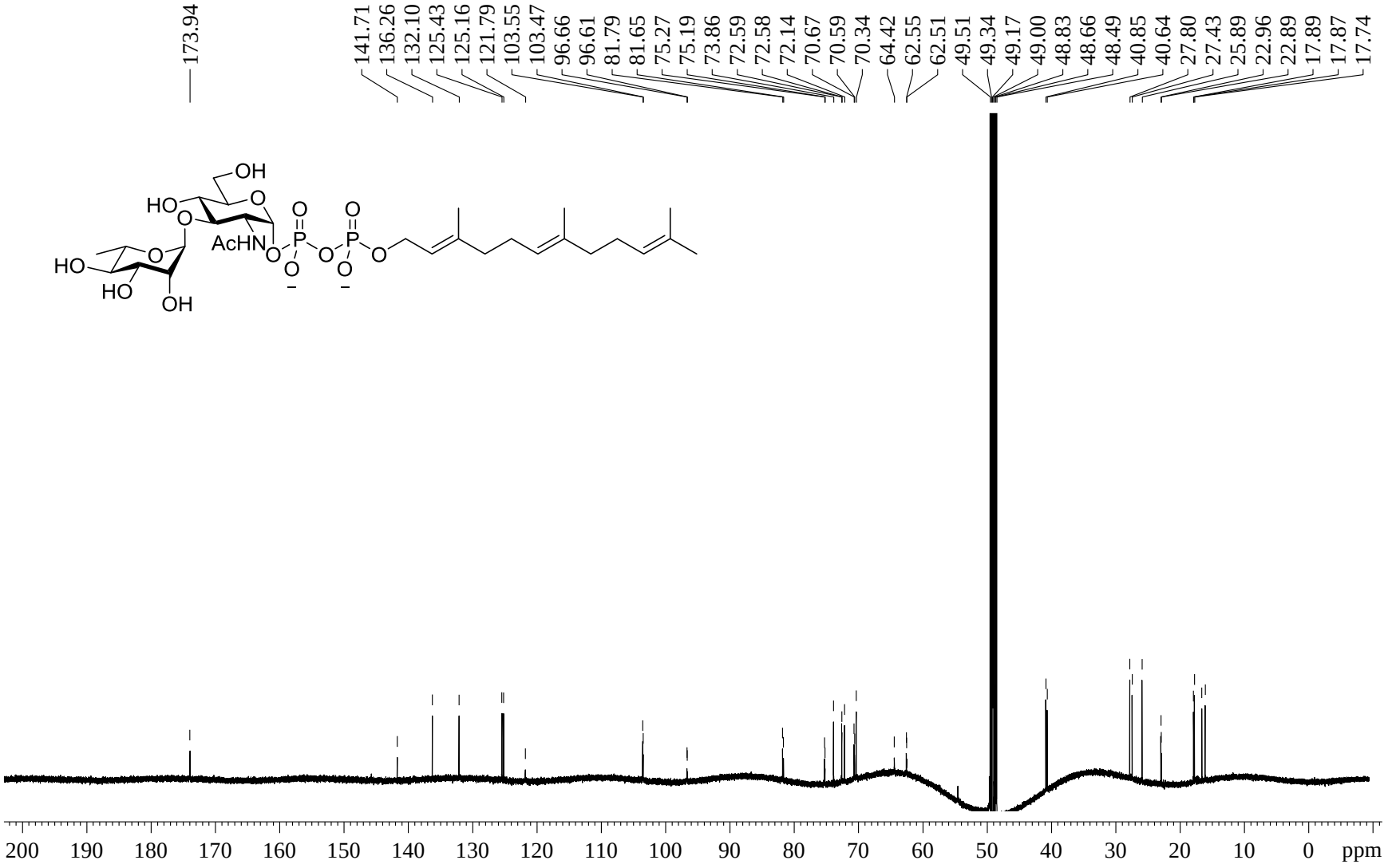
^{31}P NMR of Compound 21 (162 MHz, $CDCl_3$)



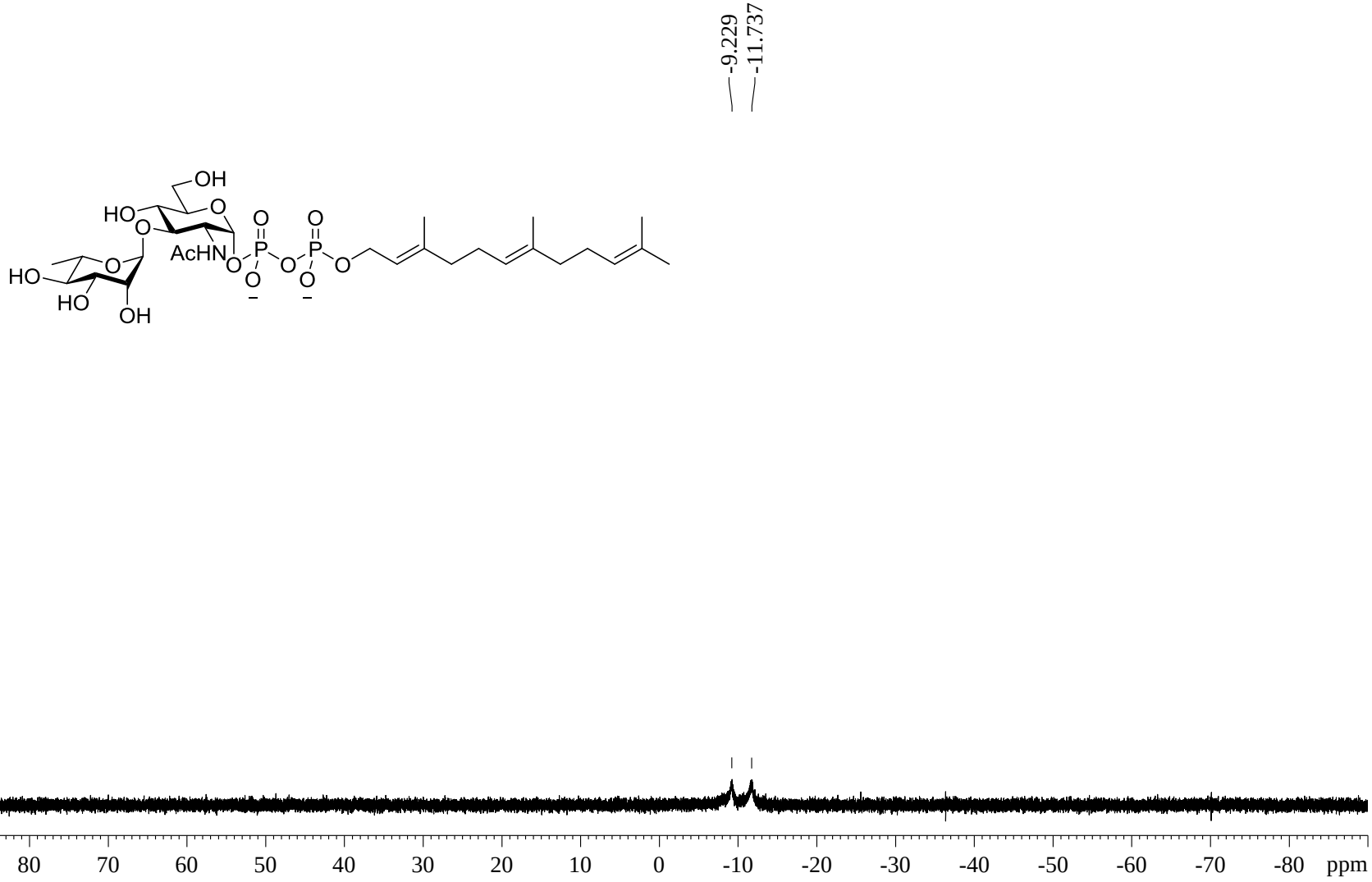
^{1}H NMR of Compound 22 (500 MHz, CD_3OD)



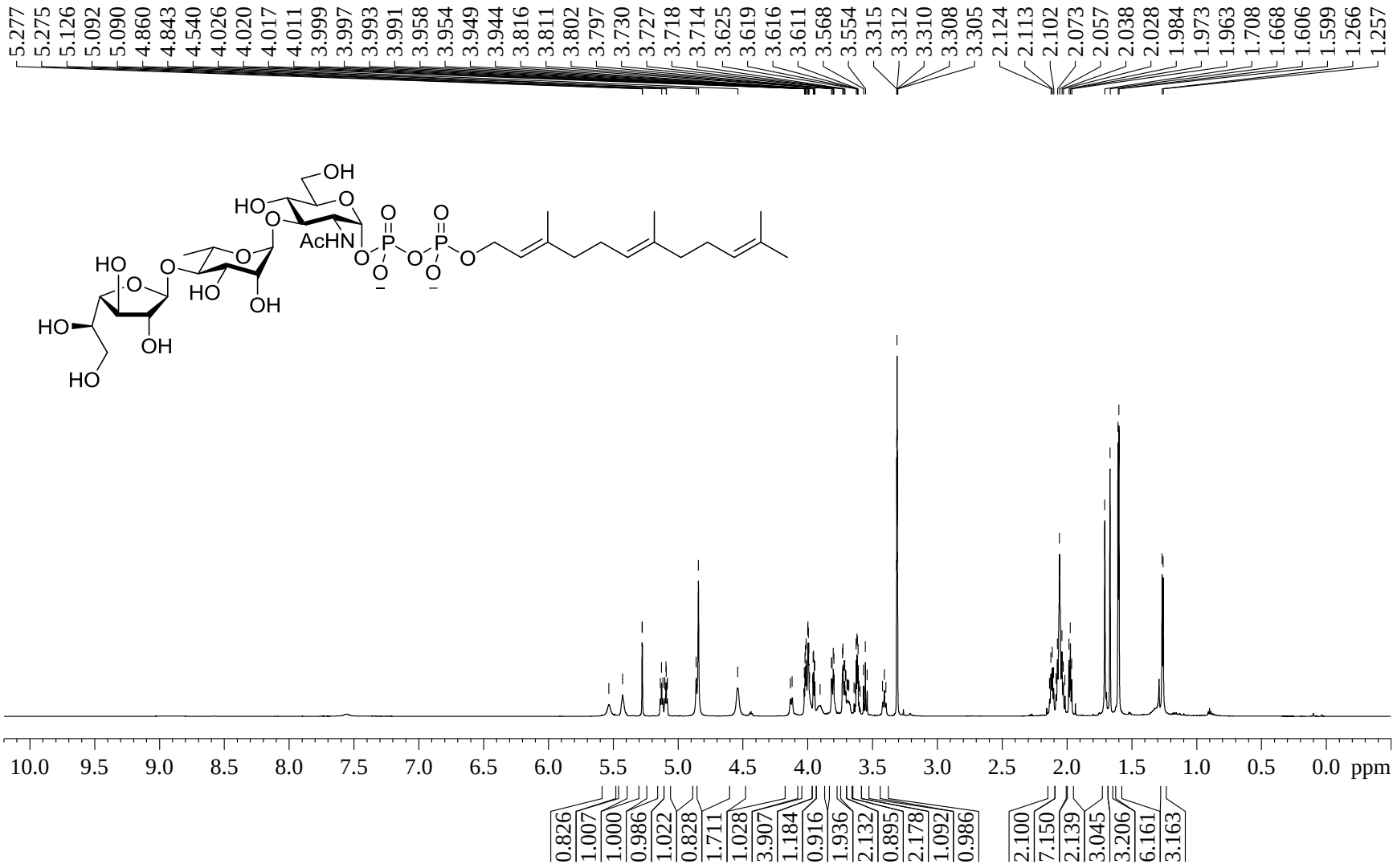
^{13}C NMR of Compound 22 (125 MHz, CD_3OD)



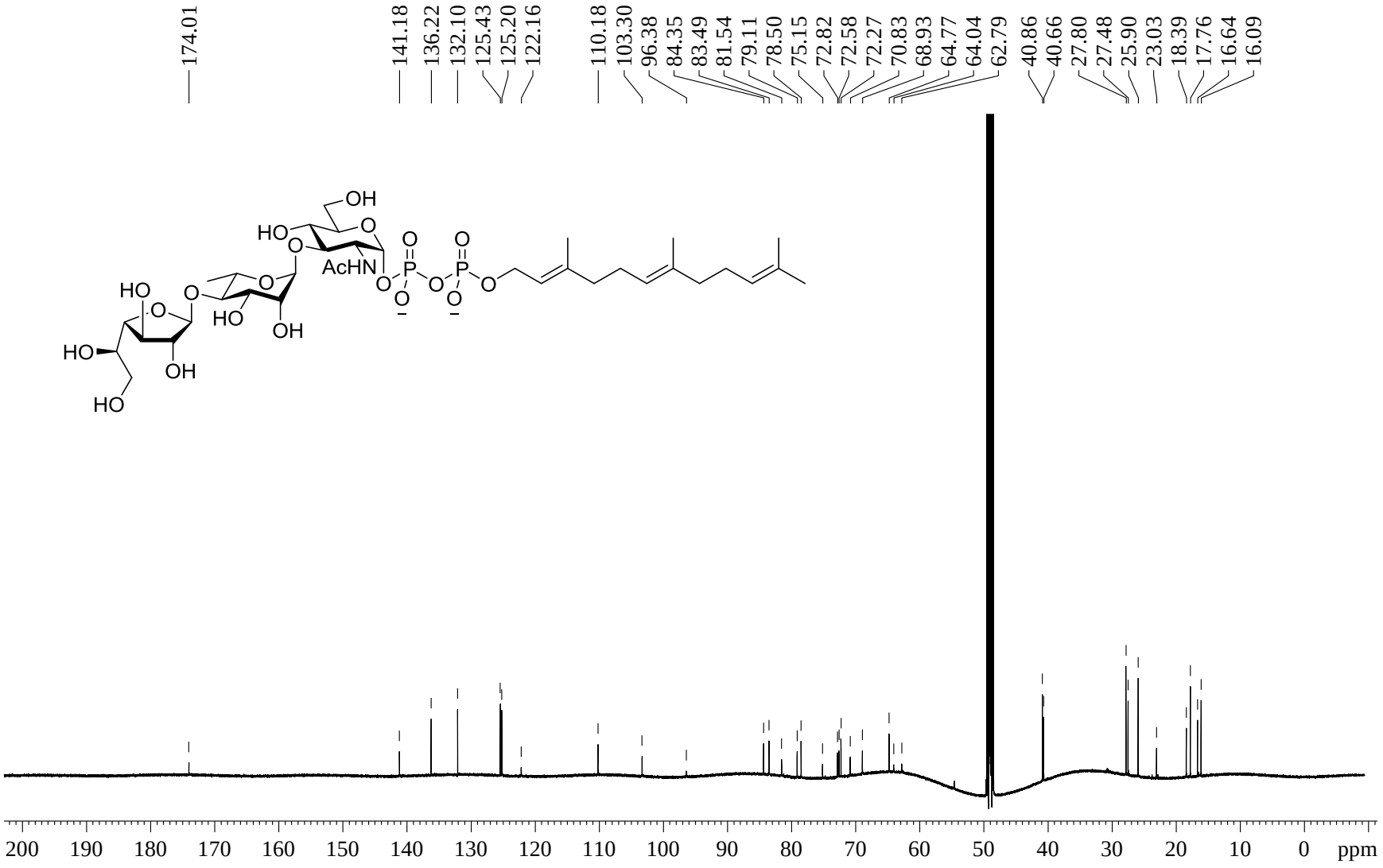
^{31}P NMR of Compound 22 (162 MHz, CD_3OD)



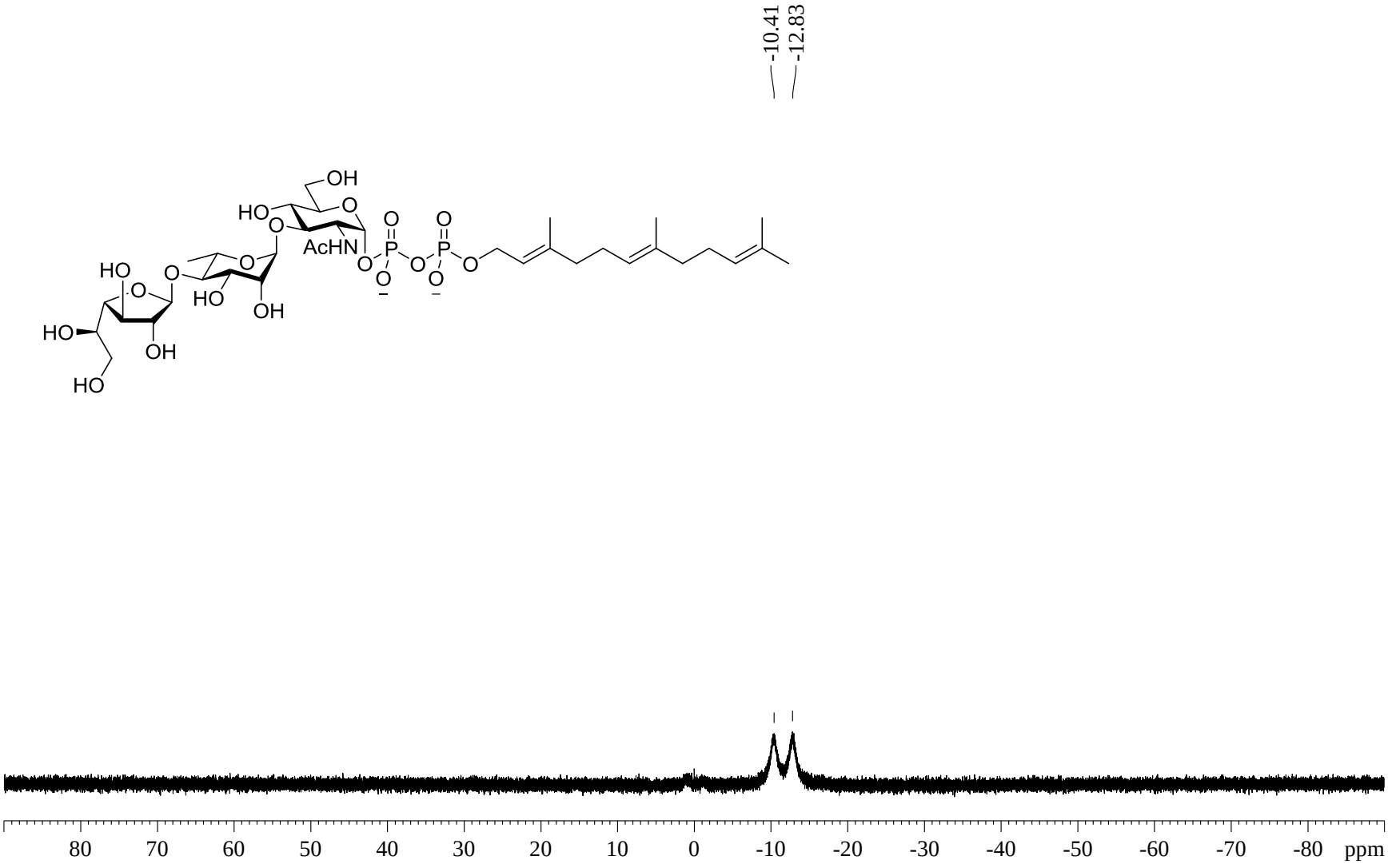
^{1}H NMR of Compound 23 (700 MHz, CD_3OD)



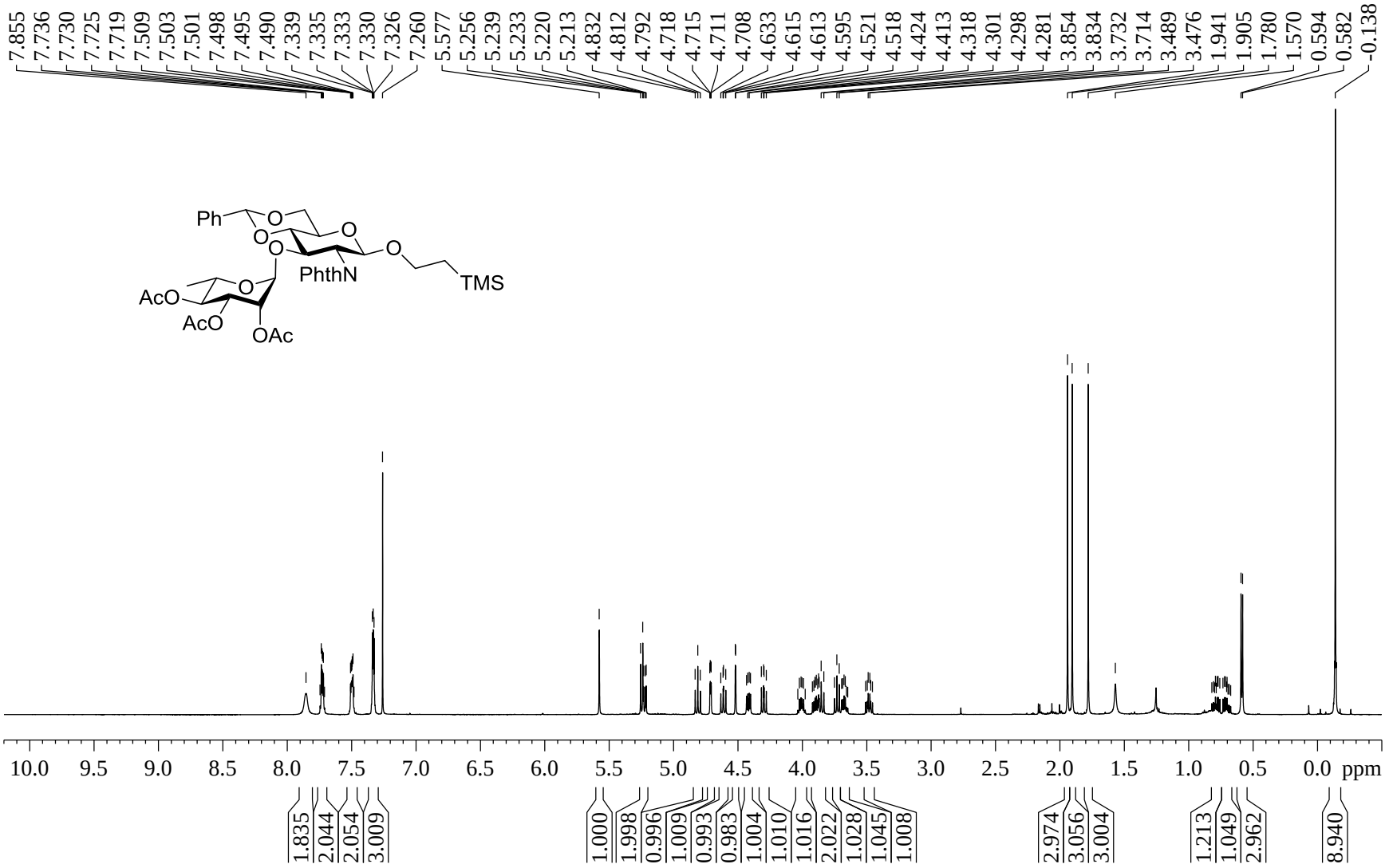
^{13}C NMR of Compound 23 (125 MHz, CD_3OD)



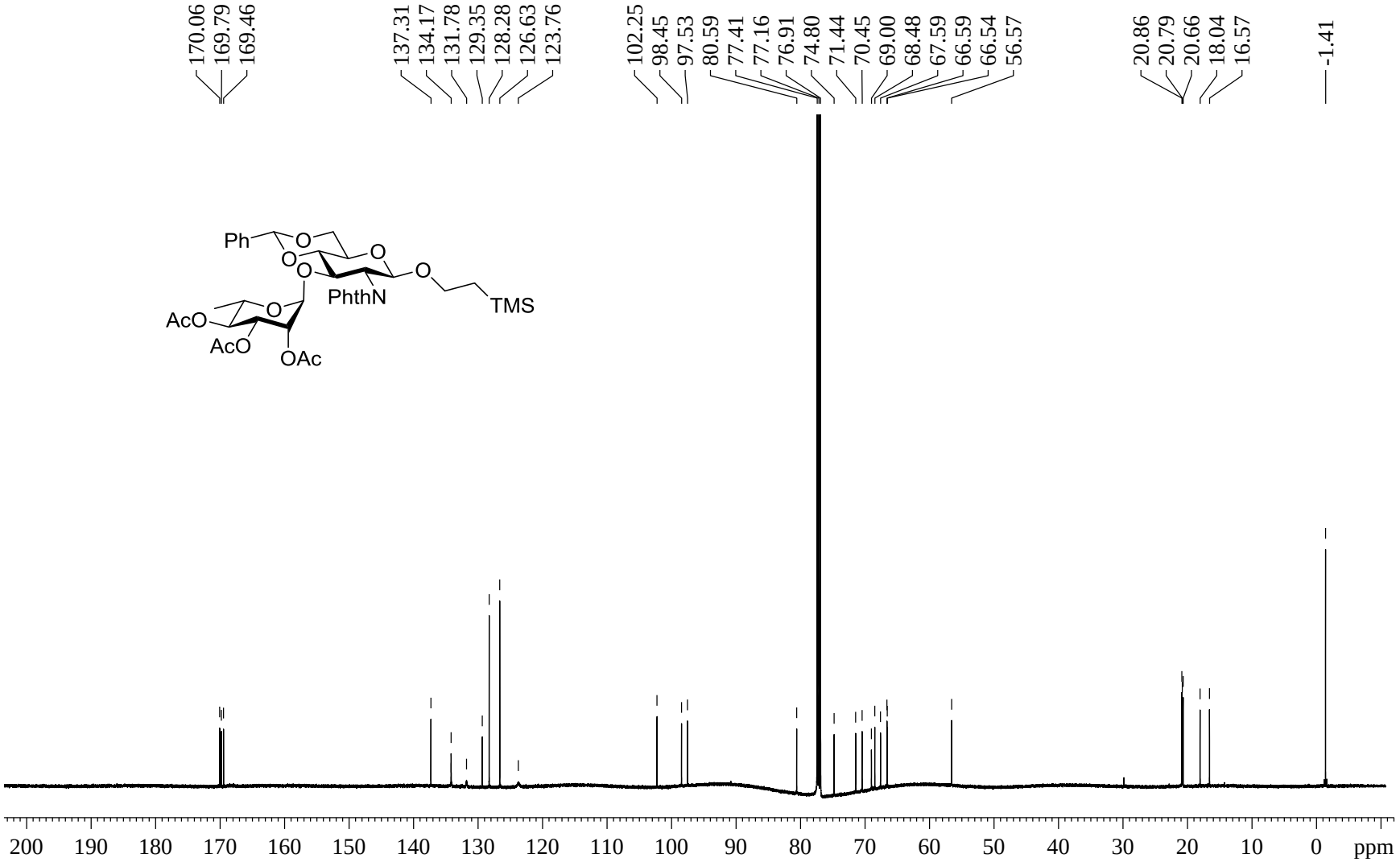
^{31}P NMR of Compound 23 (202 MHz, CD_3OD)



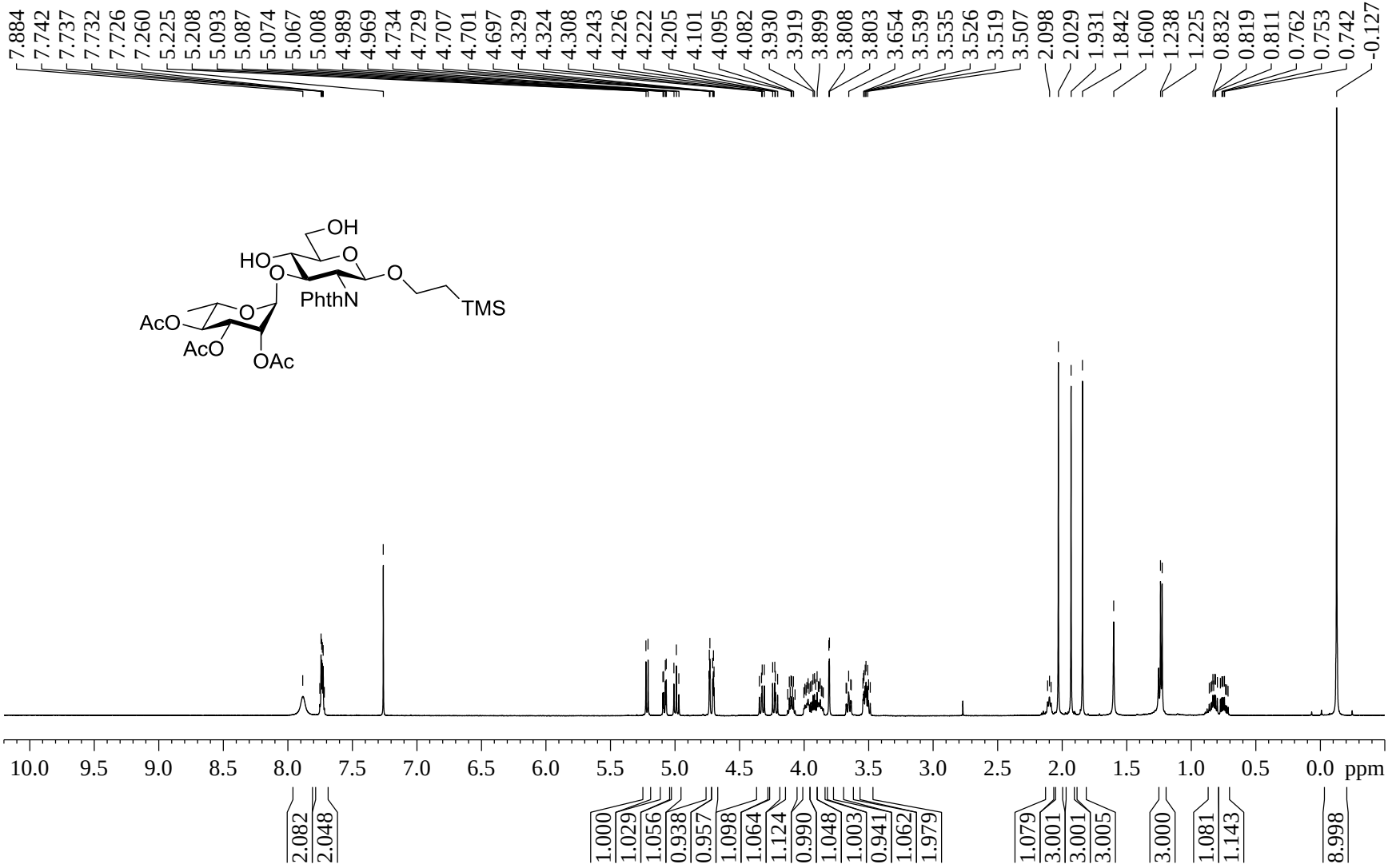
^{1}H NMR of Compound 25 (500 MHz, $CDCl_3$)



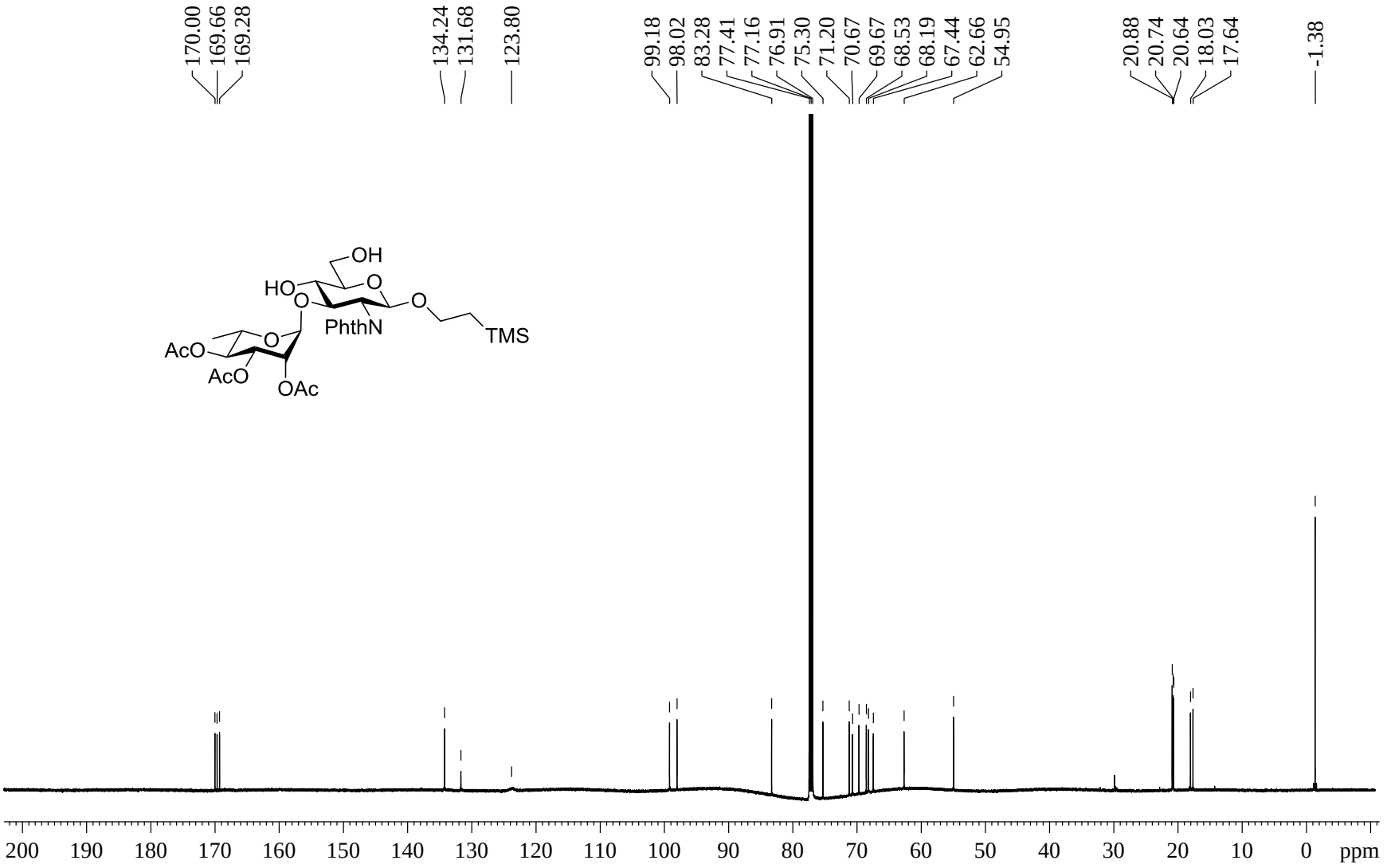
^{13}C NMR of Compound 25 (125 MHz, $CDCl_3$)



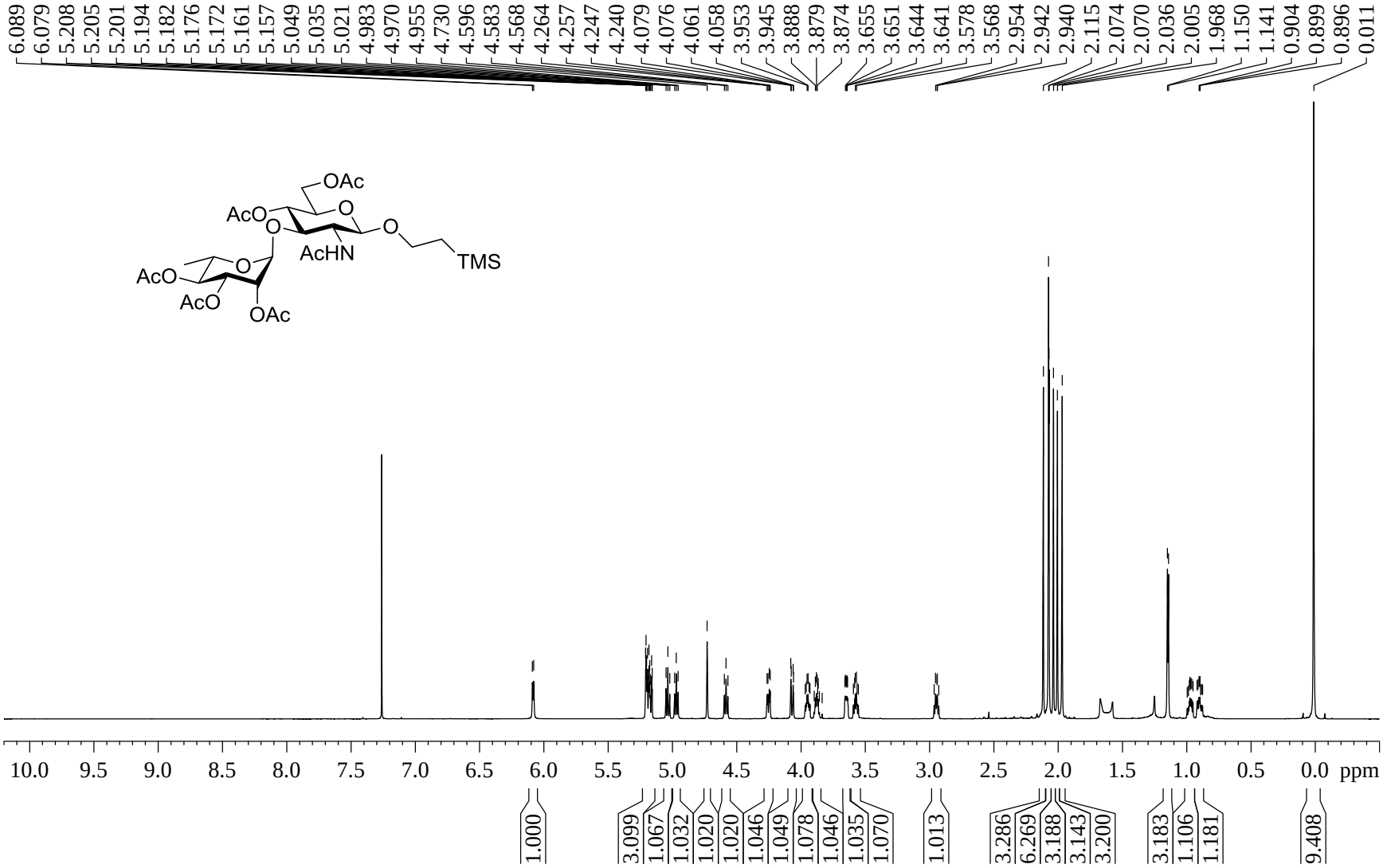
^{1}H NMR of Compound 26 (500 MHz, $CDCl_3$)



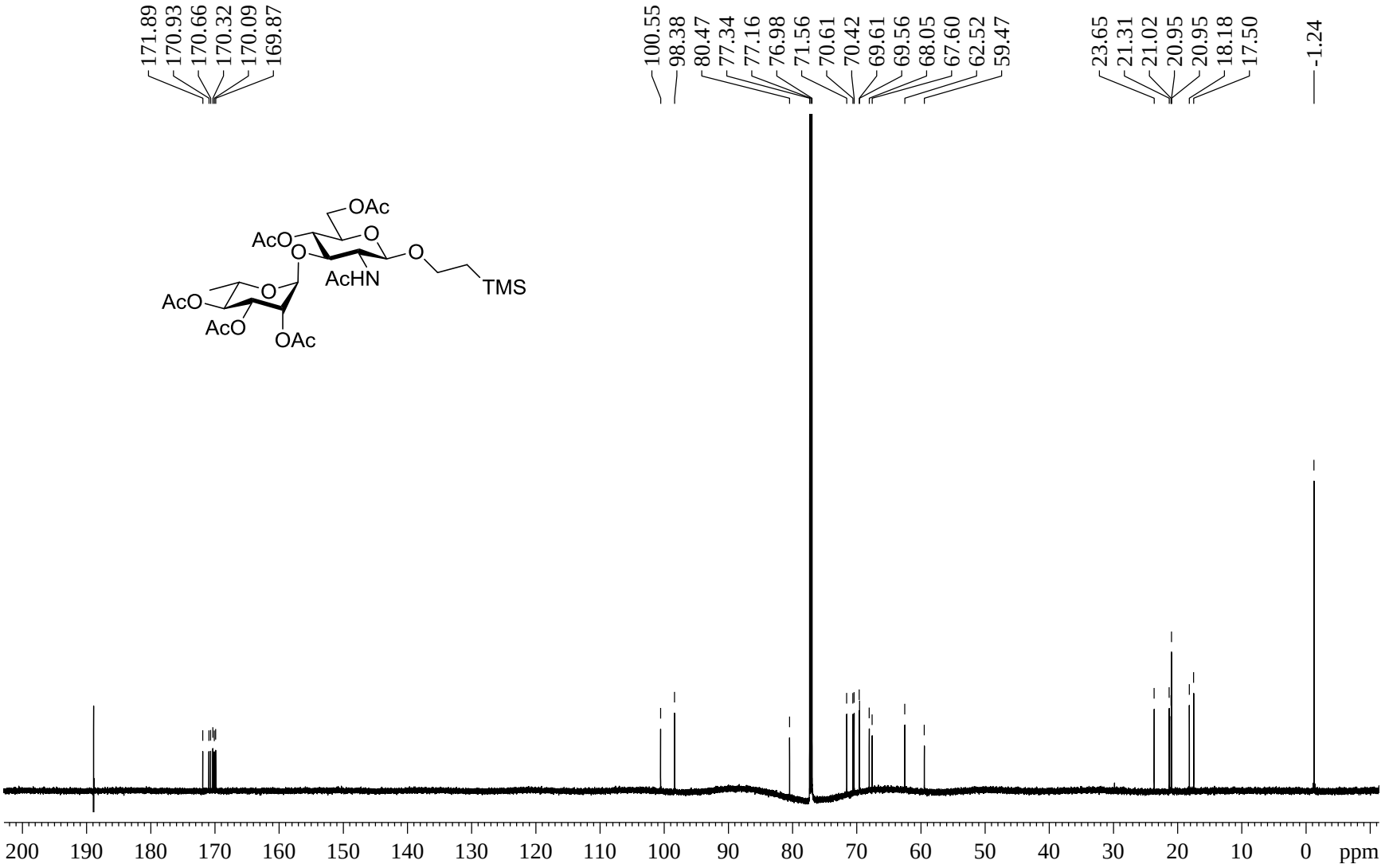
^{13}C NMR of Compound 26 (125 MHz, $CDCl_3$)



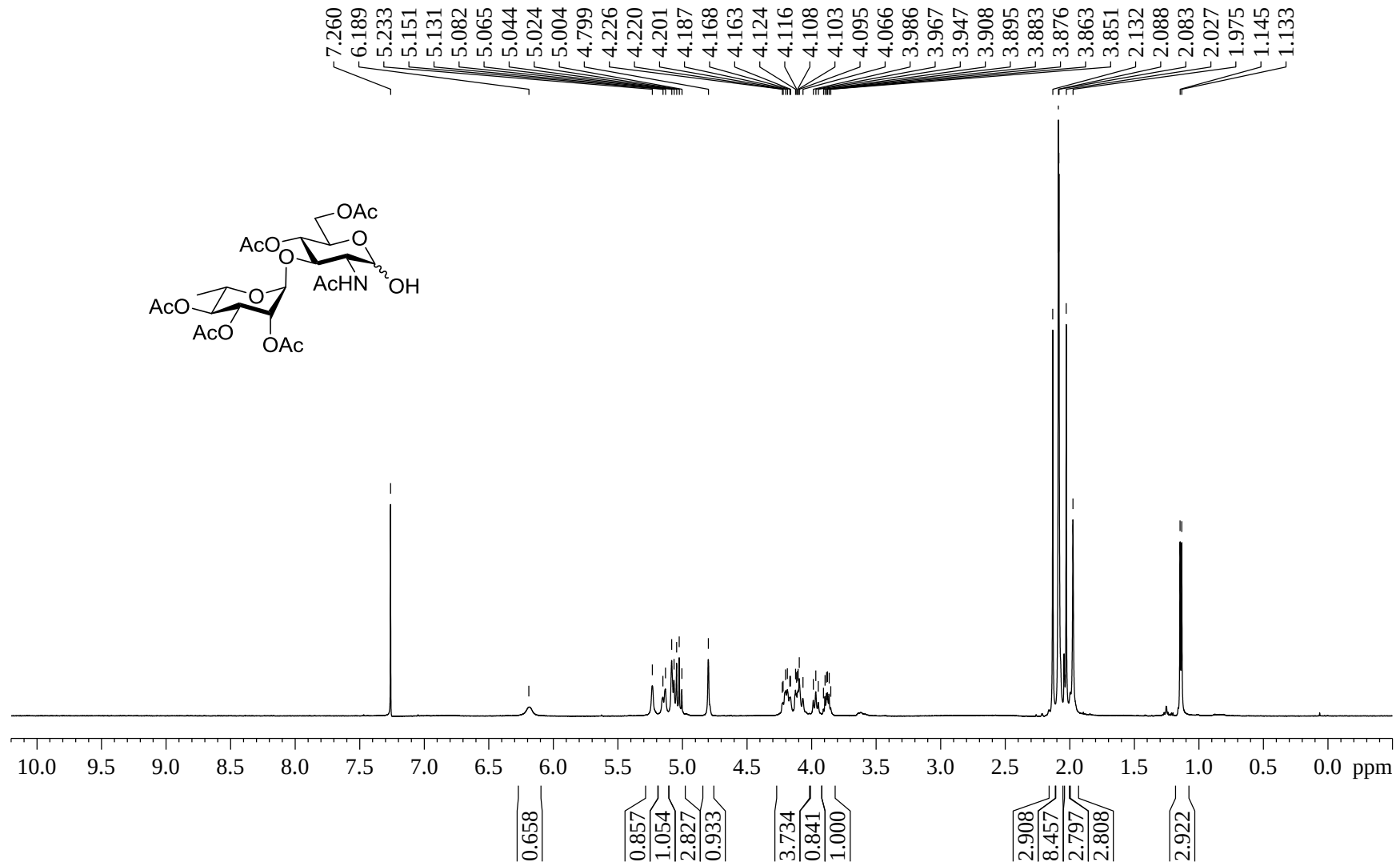
^{1}H NMR of Compound 27 (700 MHz, $CDCl_3$)



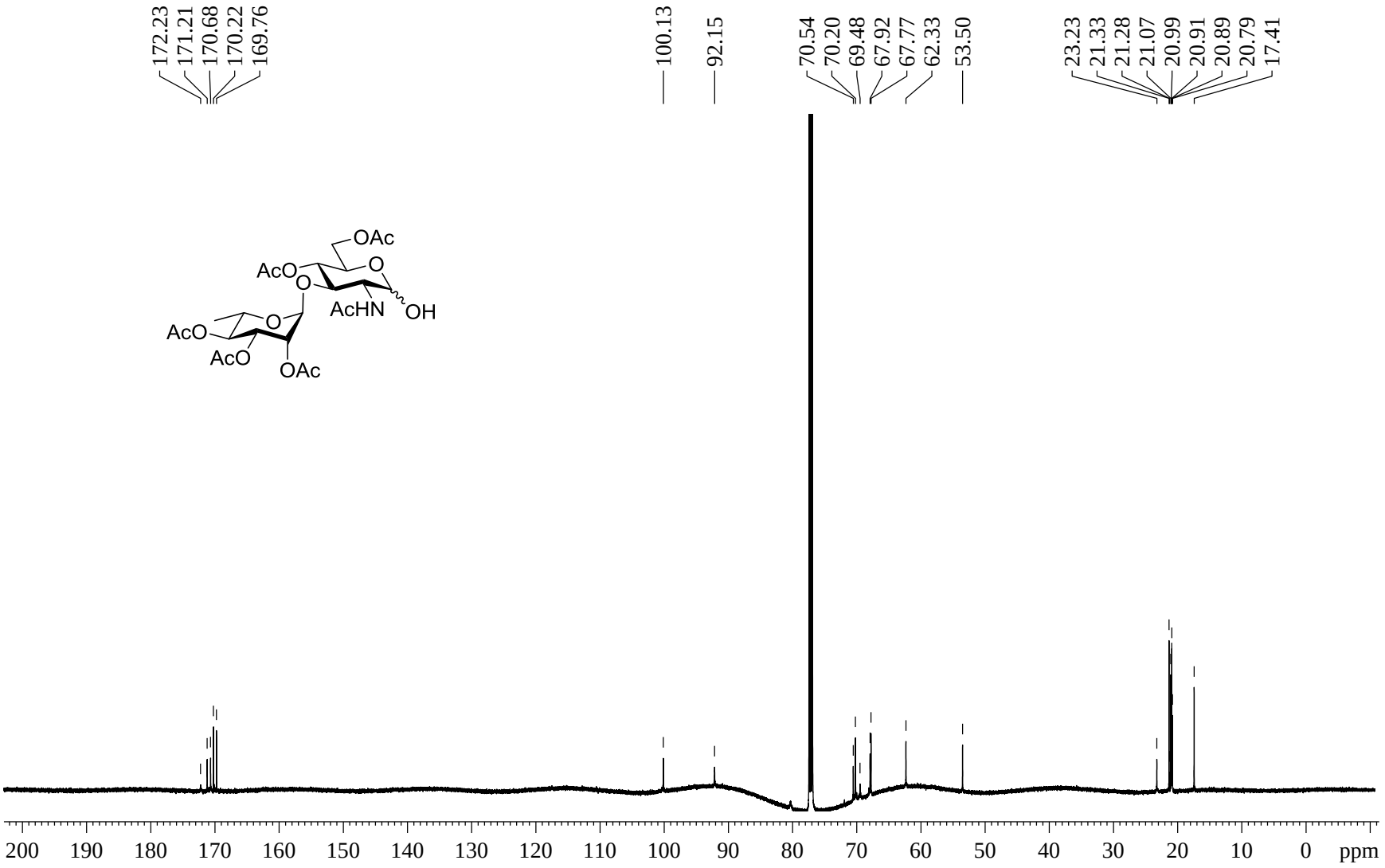
^{13}C NMR of Compound 27 (175 MHz, $CDCl_3$)



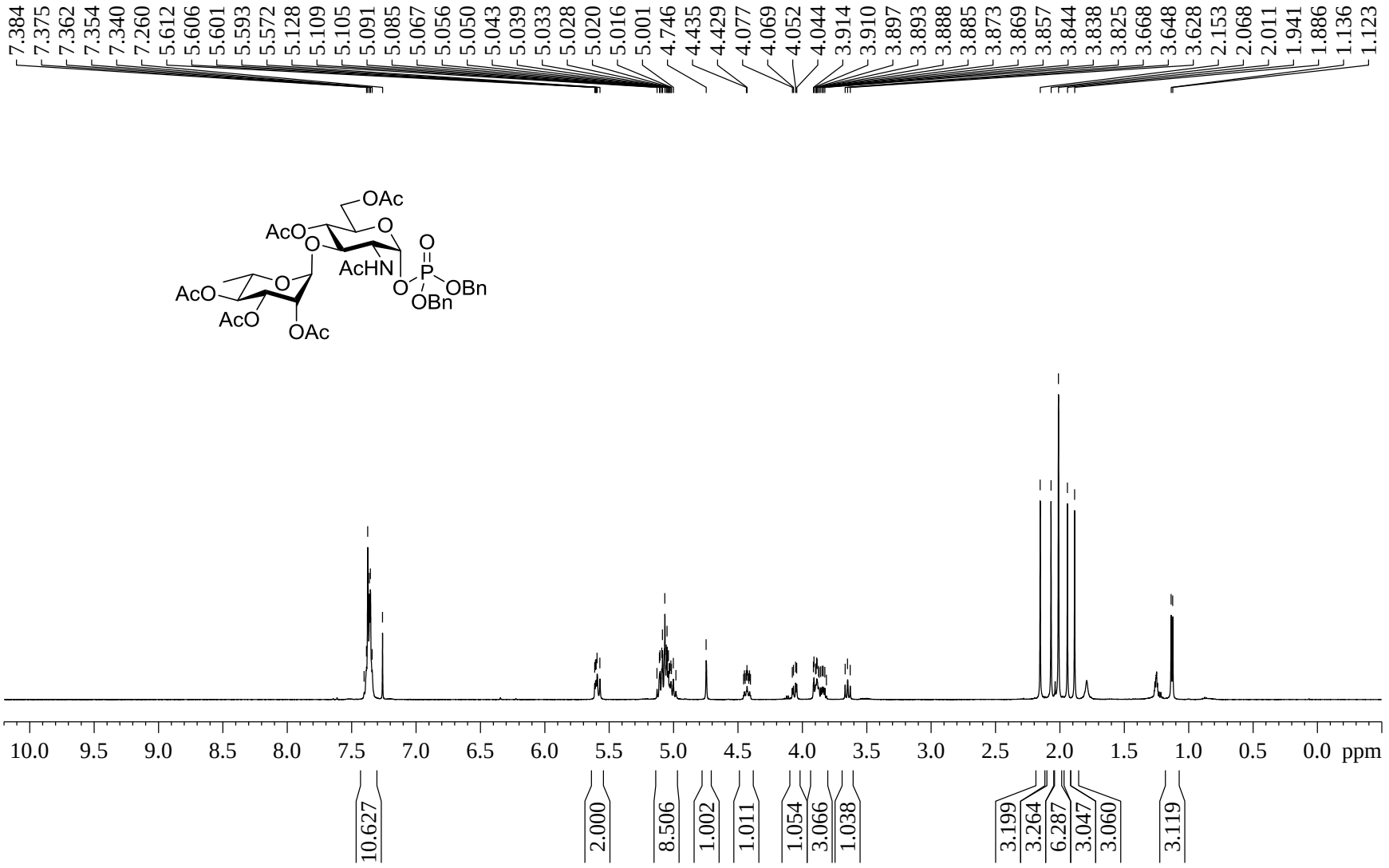
^{1}H NMR of Compound 28 (500 MHz, $CDCl_3$)



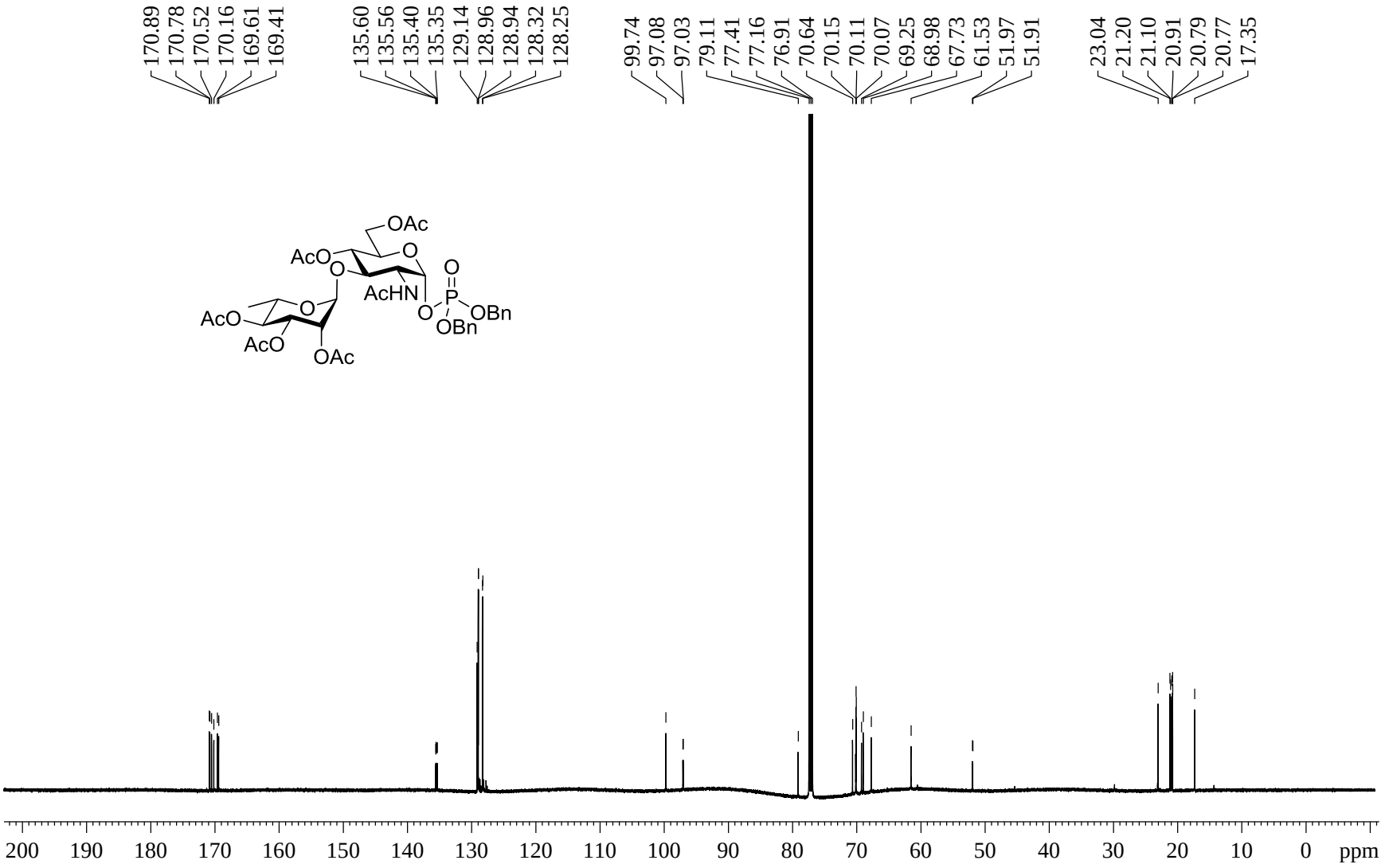
^{13}C NMR of Compound 28 (125 MHz, $CDCl_3$)



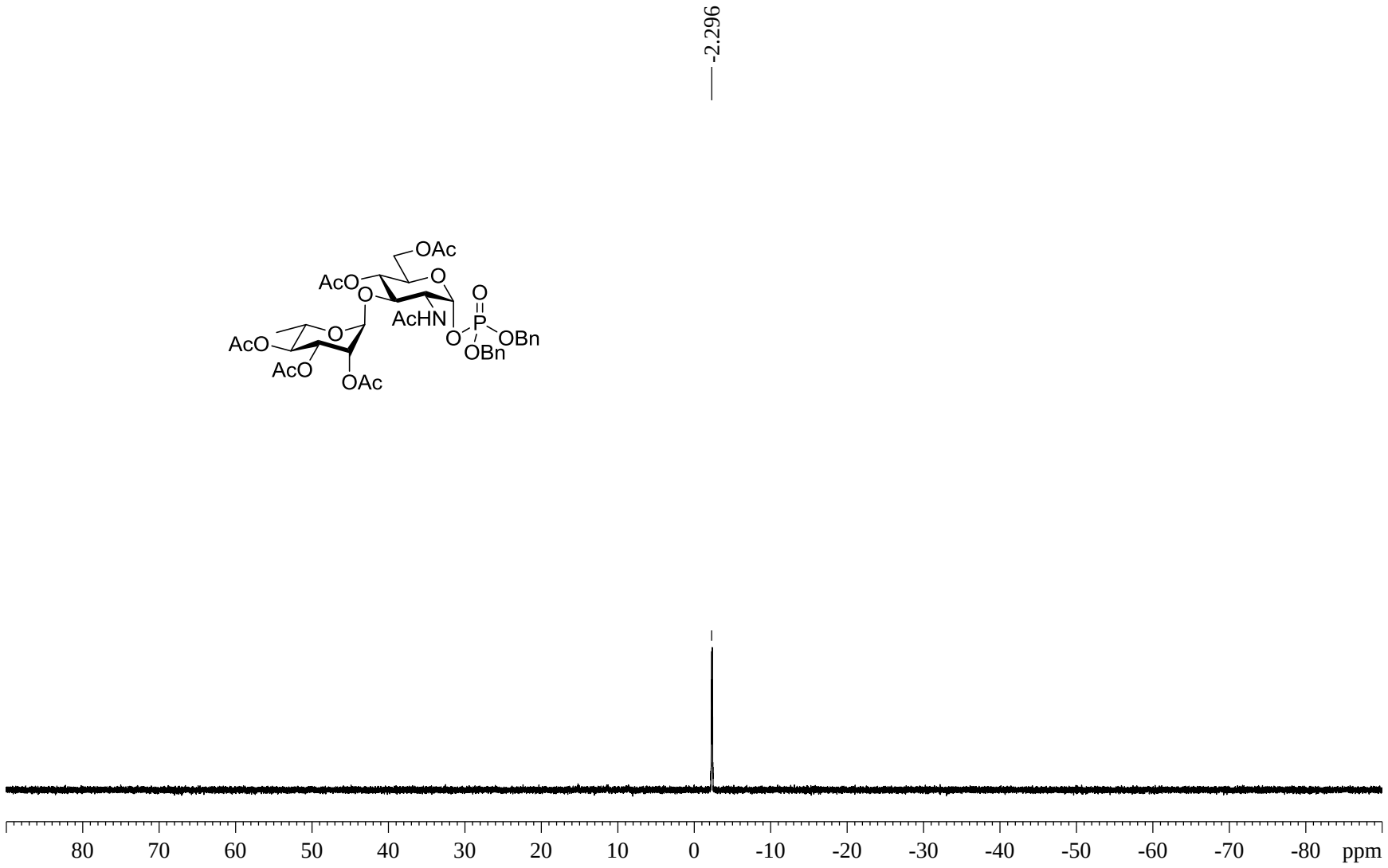
^{1}H NMR of Compound 29 (500 MHz, $CDCl_3$)



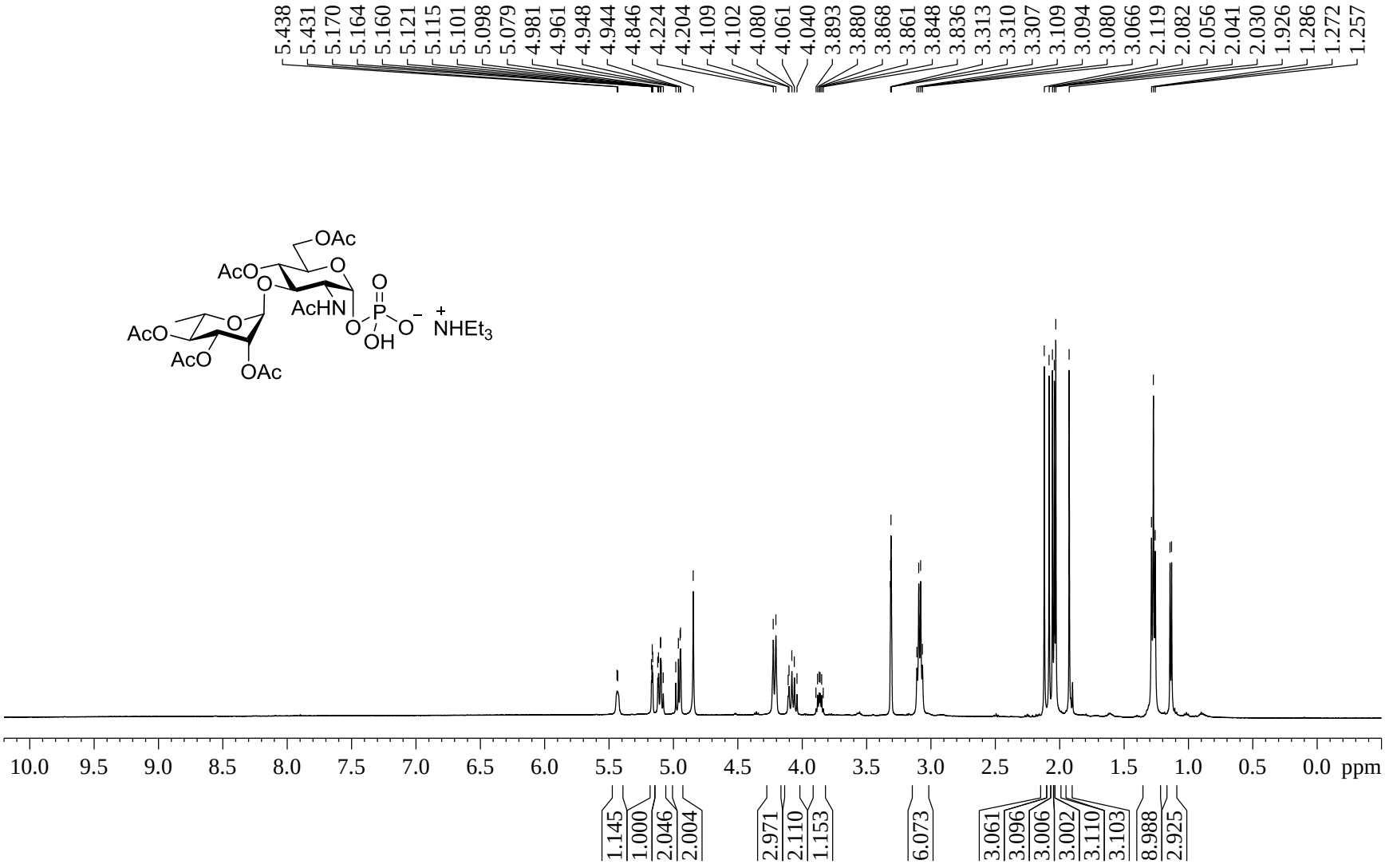
^{13}C NMR of Compound 29 (125 MHz, $CDCl_3$)



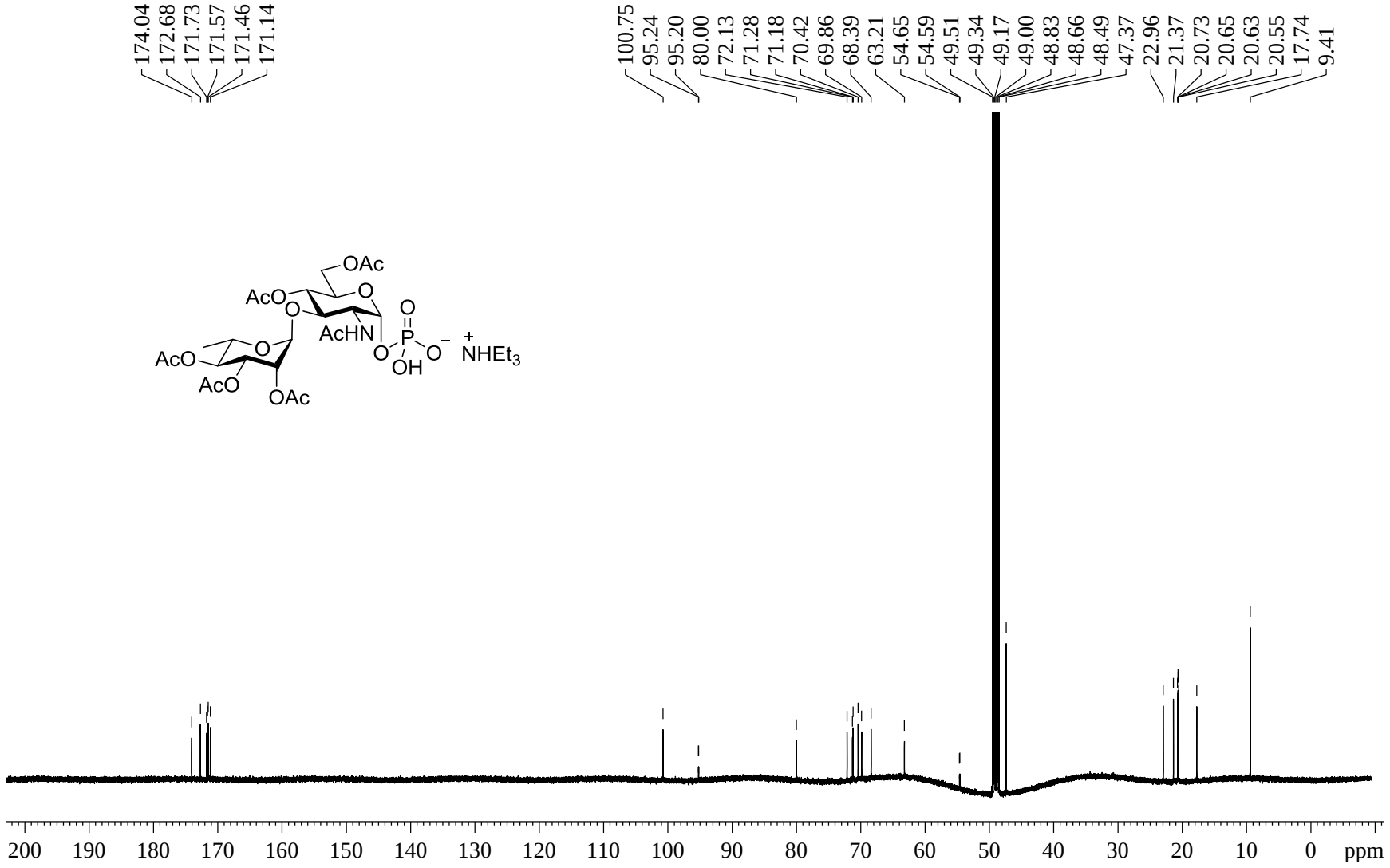
^{31}P NMR of Compound 29 (162 MHz, $CDCl_3$)



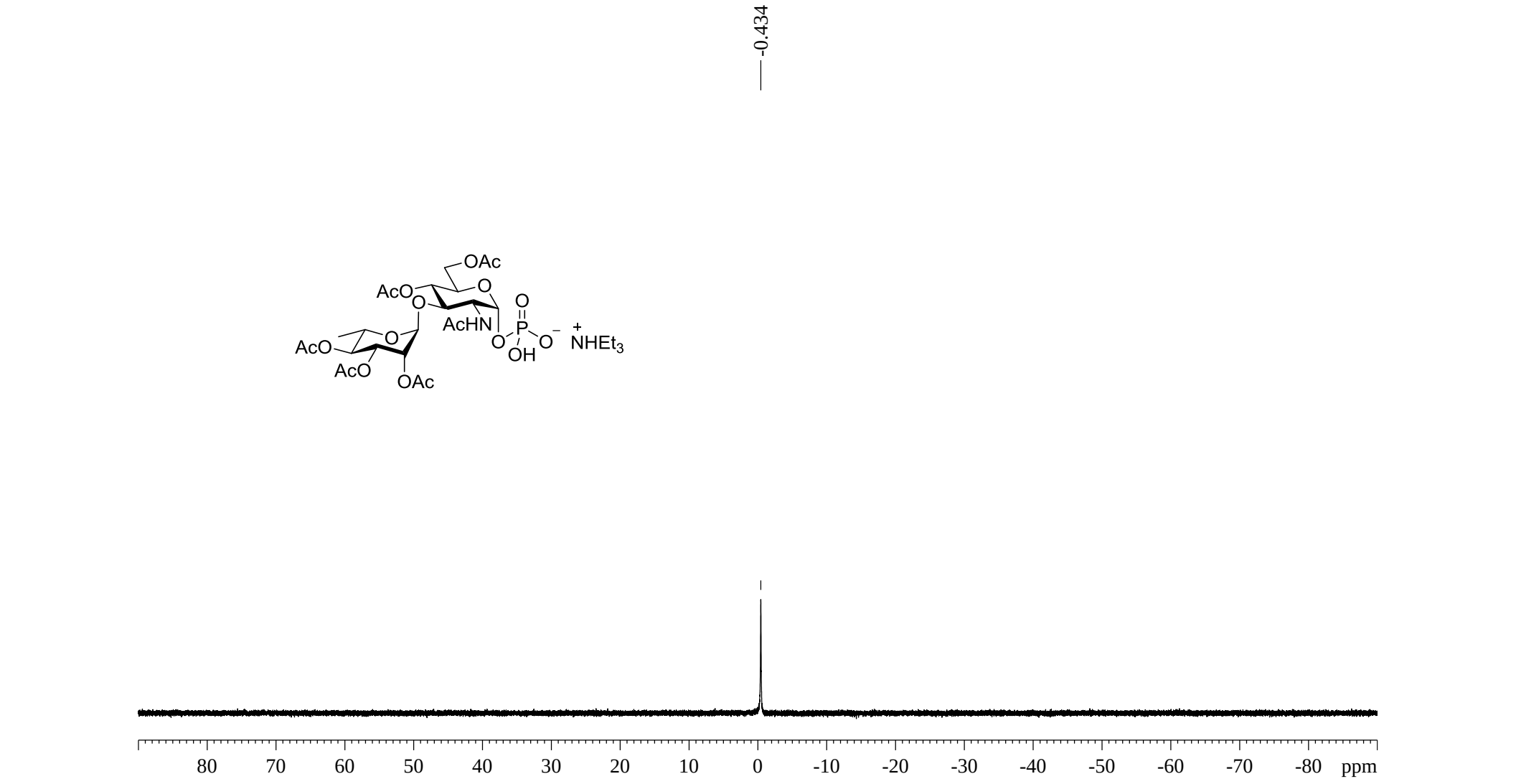
^{1}H NMR of Compound 30 (500 MHz, $CDCl_3$)



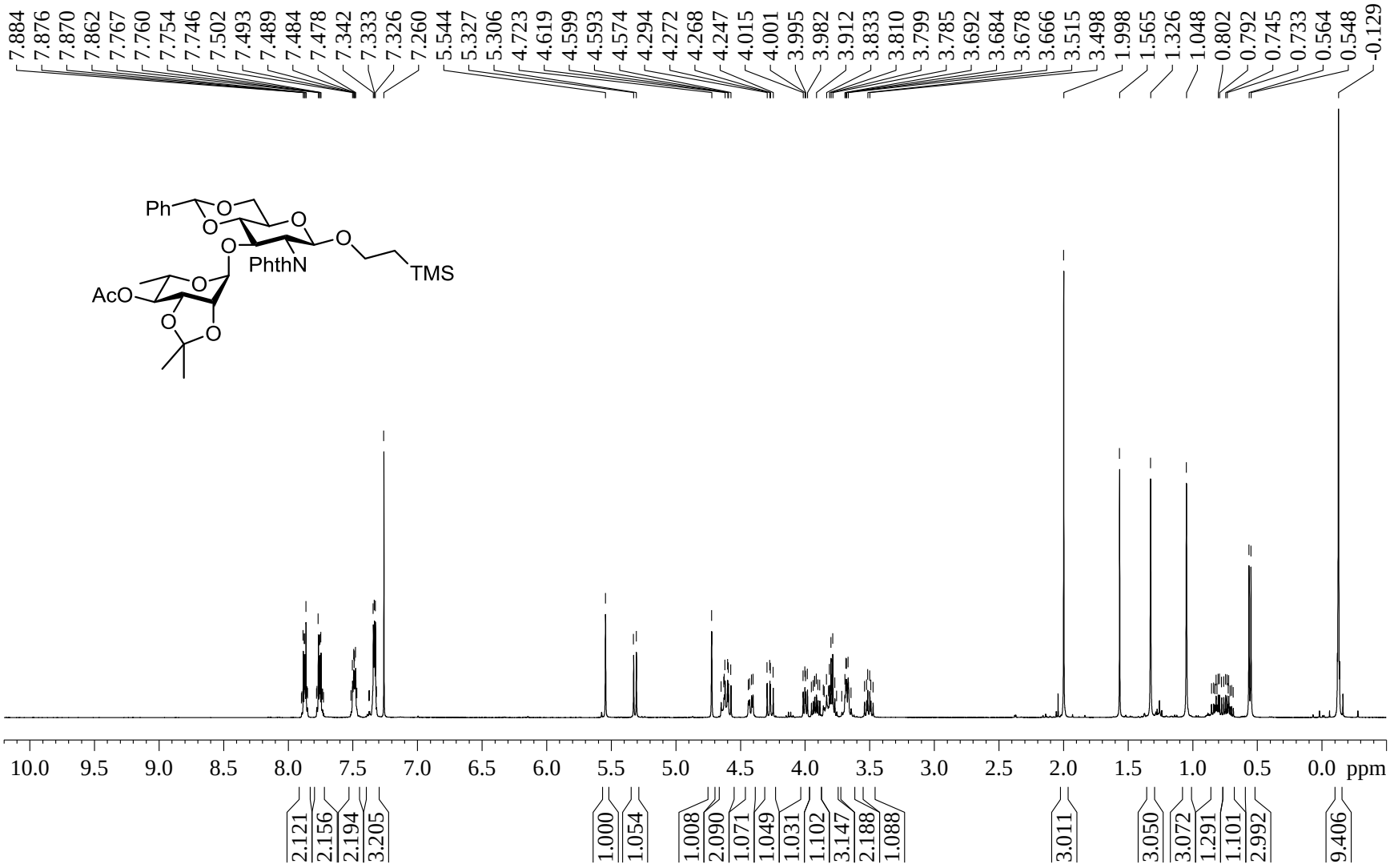
^{13}C NMR of Compound 30 (125 MHz, $CDCl_3$)



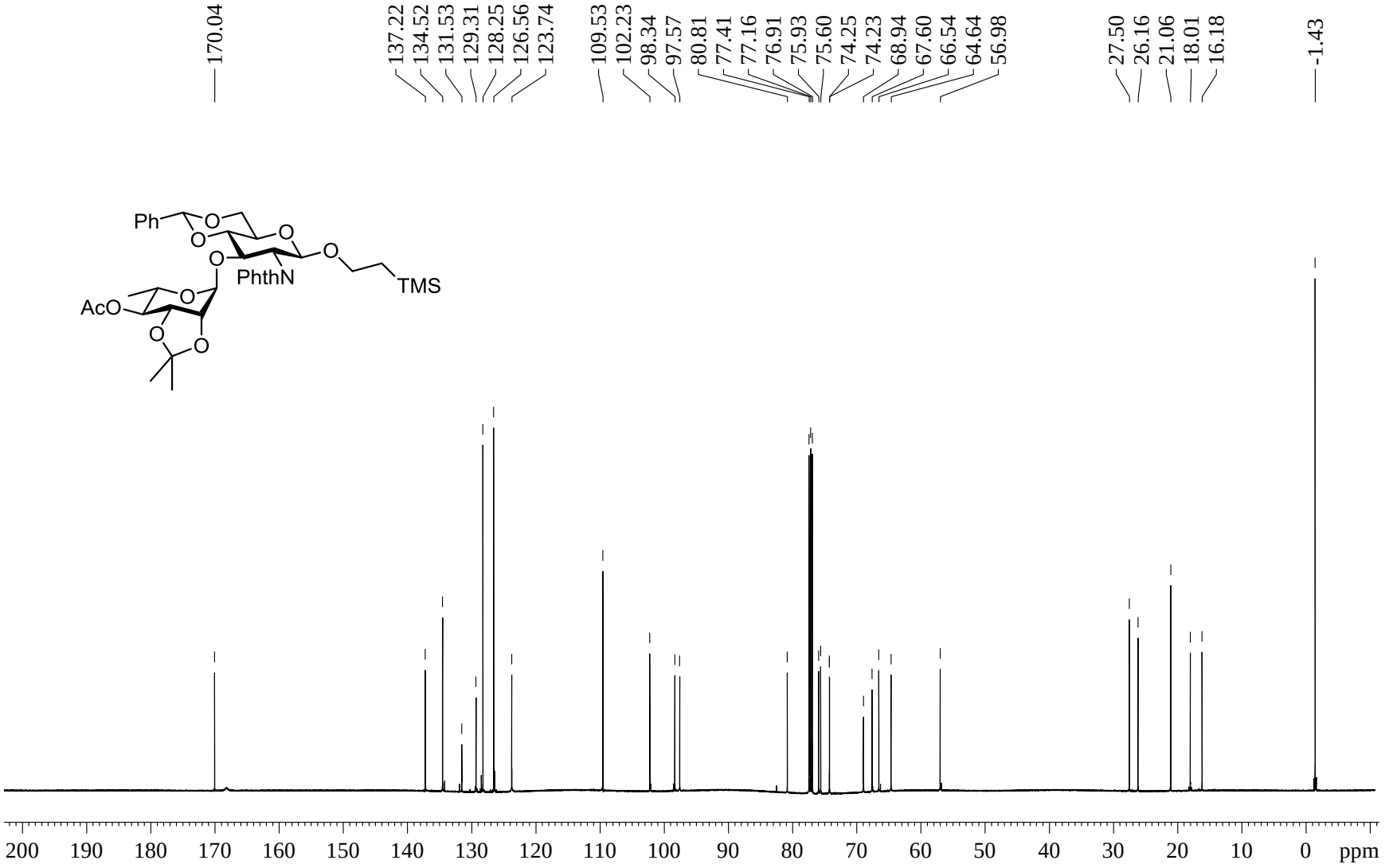
^{31}P NMR of Compound 30 (162 MHz, $CDCl_3$)



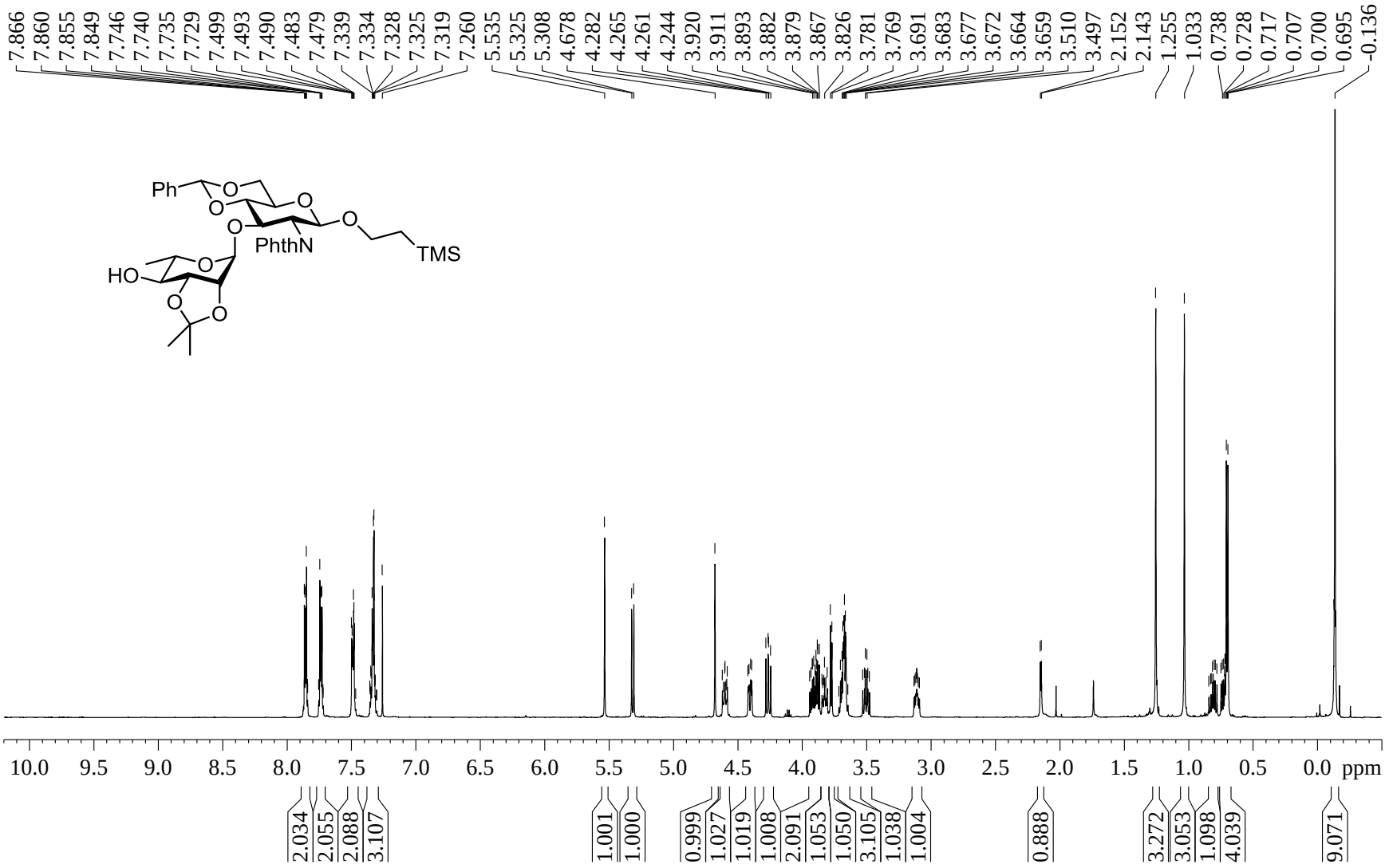
^{1}H NMR of Compound 32 (400 MHz, $CDCl_3$)



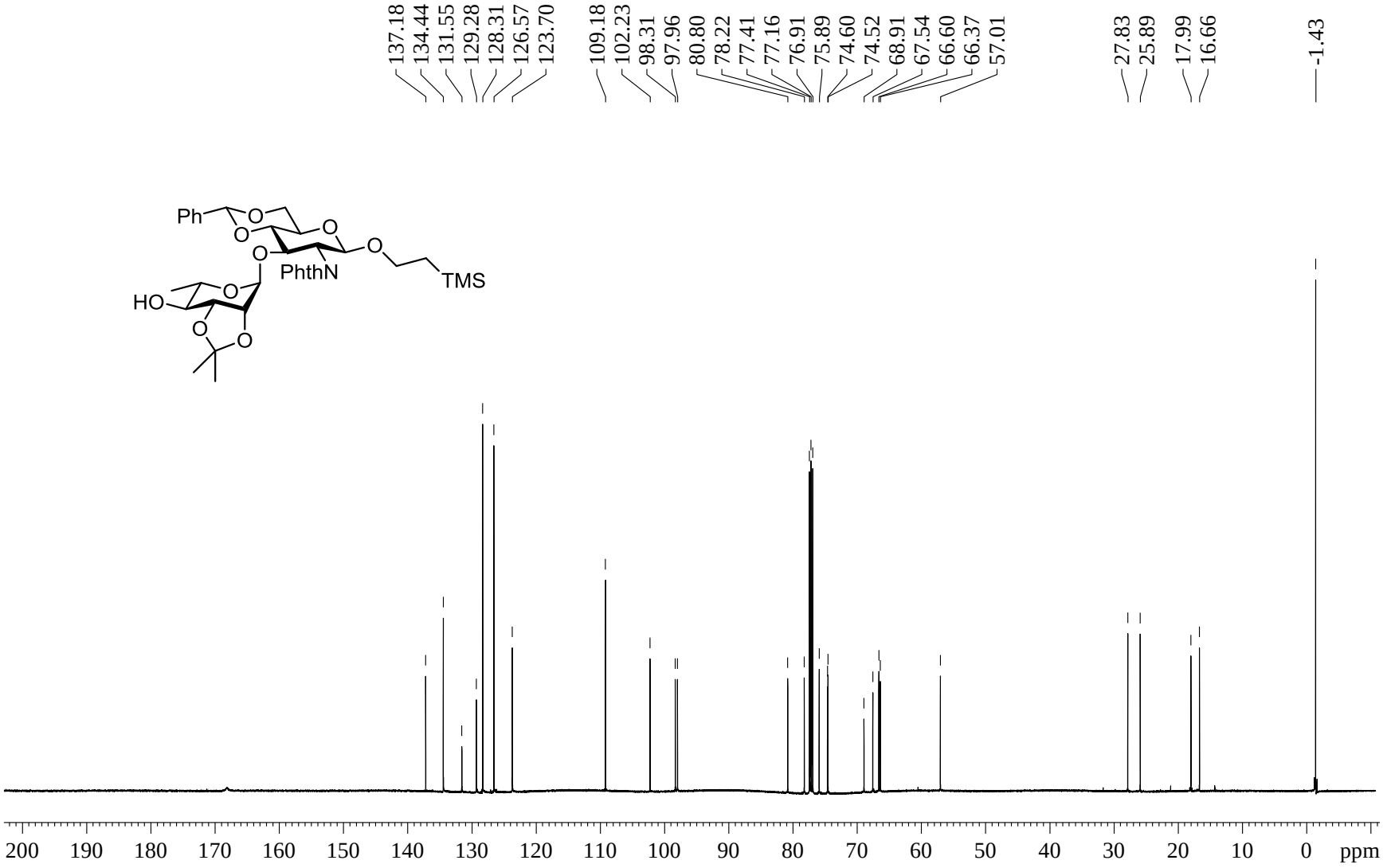
^{13}C NMR of Compound 32 (125 MHz, $CDCl_3$)



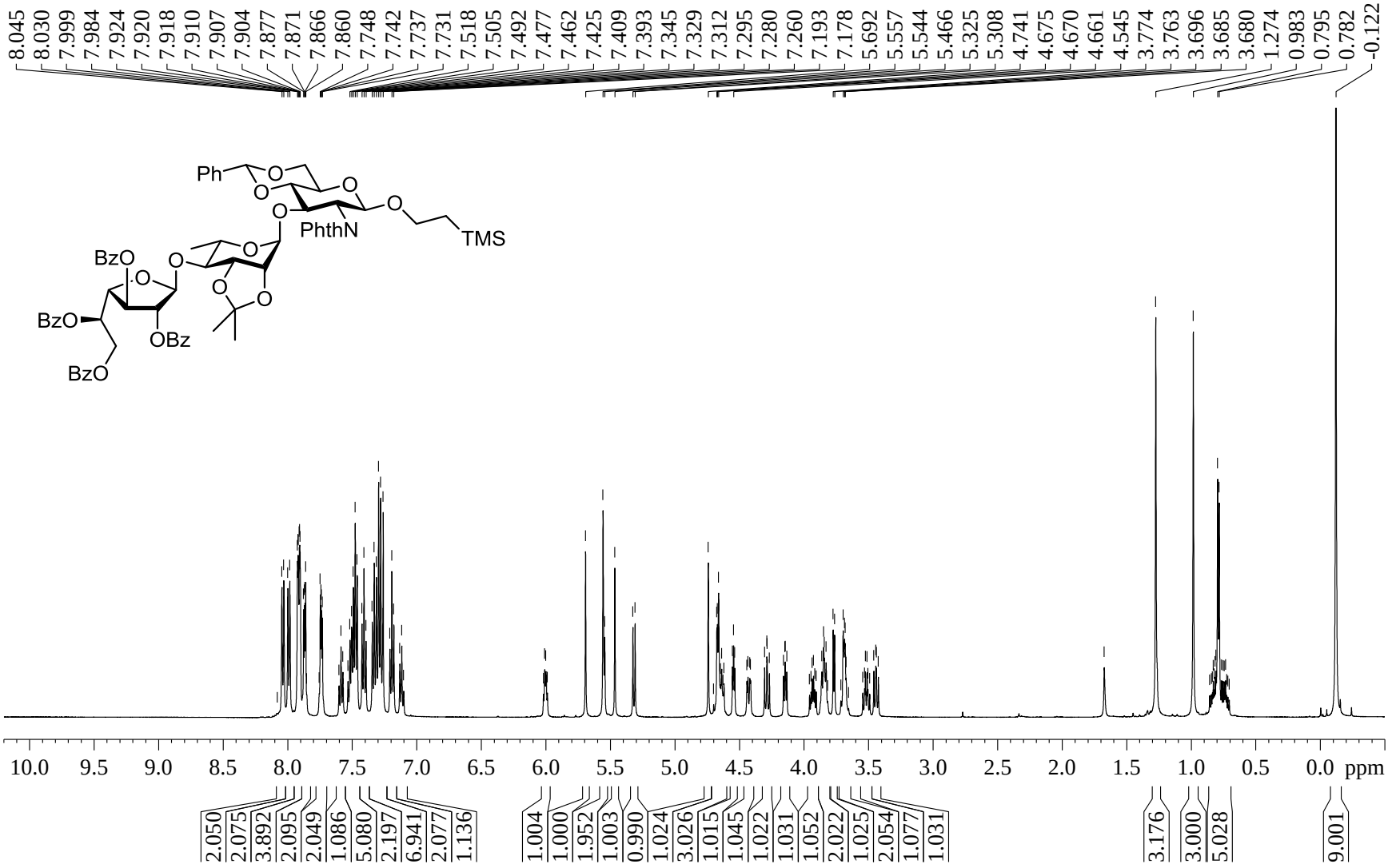
^{1}H NMR of Compound 33 (500 MHz, $CDCl_3$)



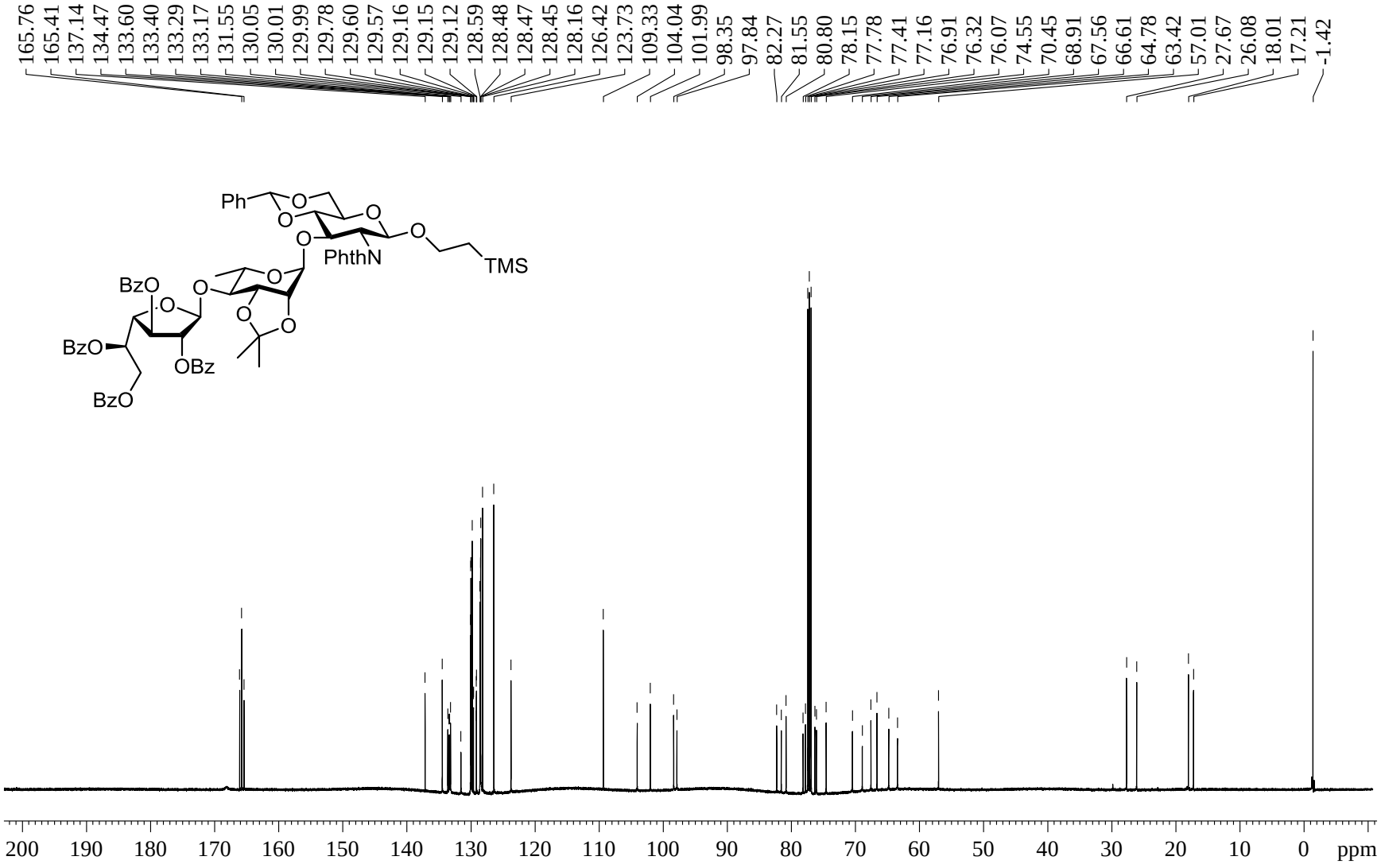
^{13}C NMR of Compound 33 (125 MHz, $CDCl_3$)



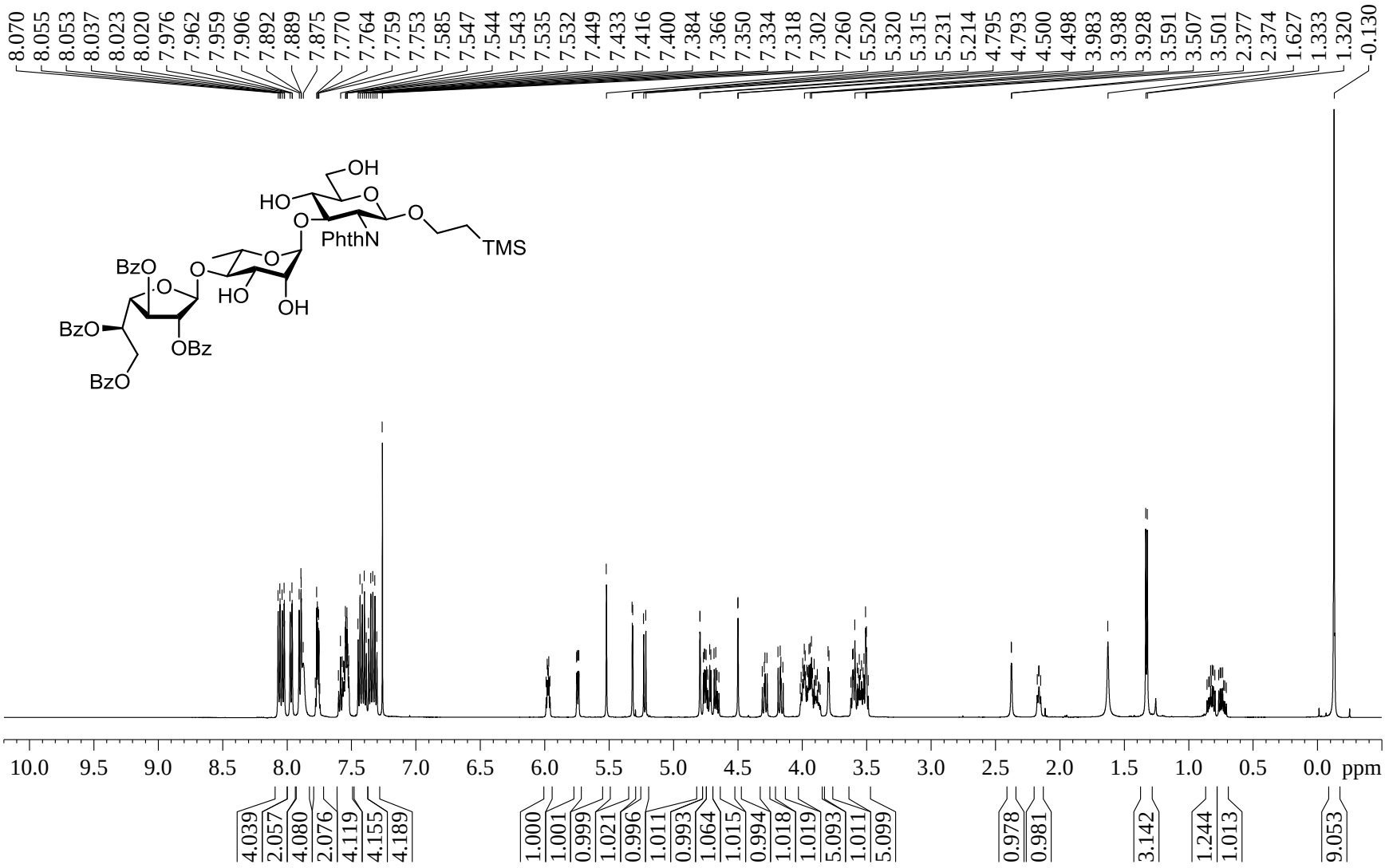
^{1}H NMR of Compound 34 (500 MHz, $CDCl_3$)



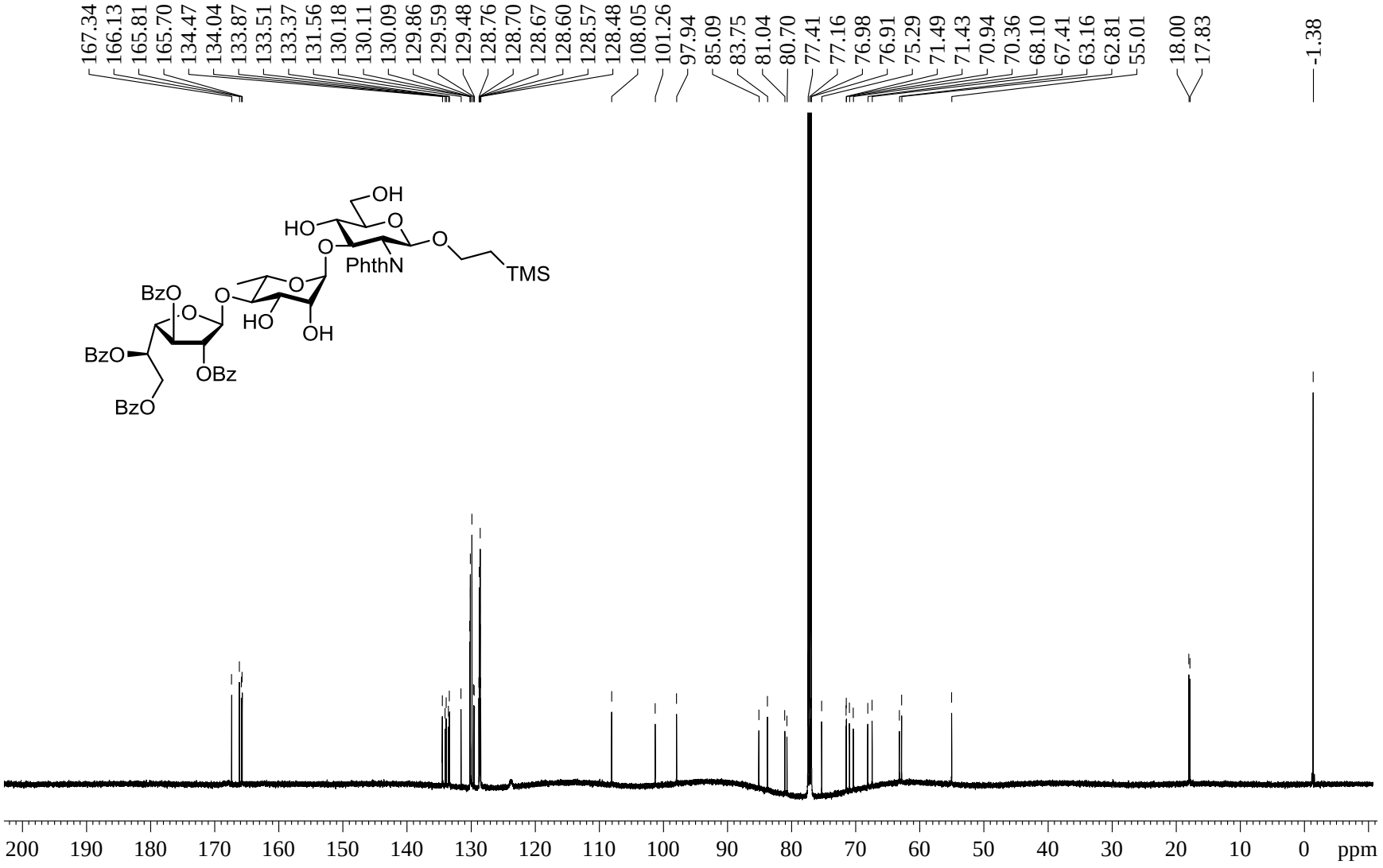
^{13}C NMR of Compound 34 (125 MHz, $CDCl_3$)



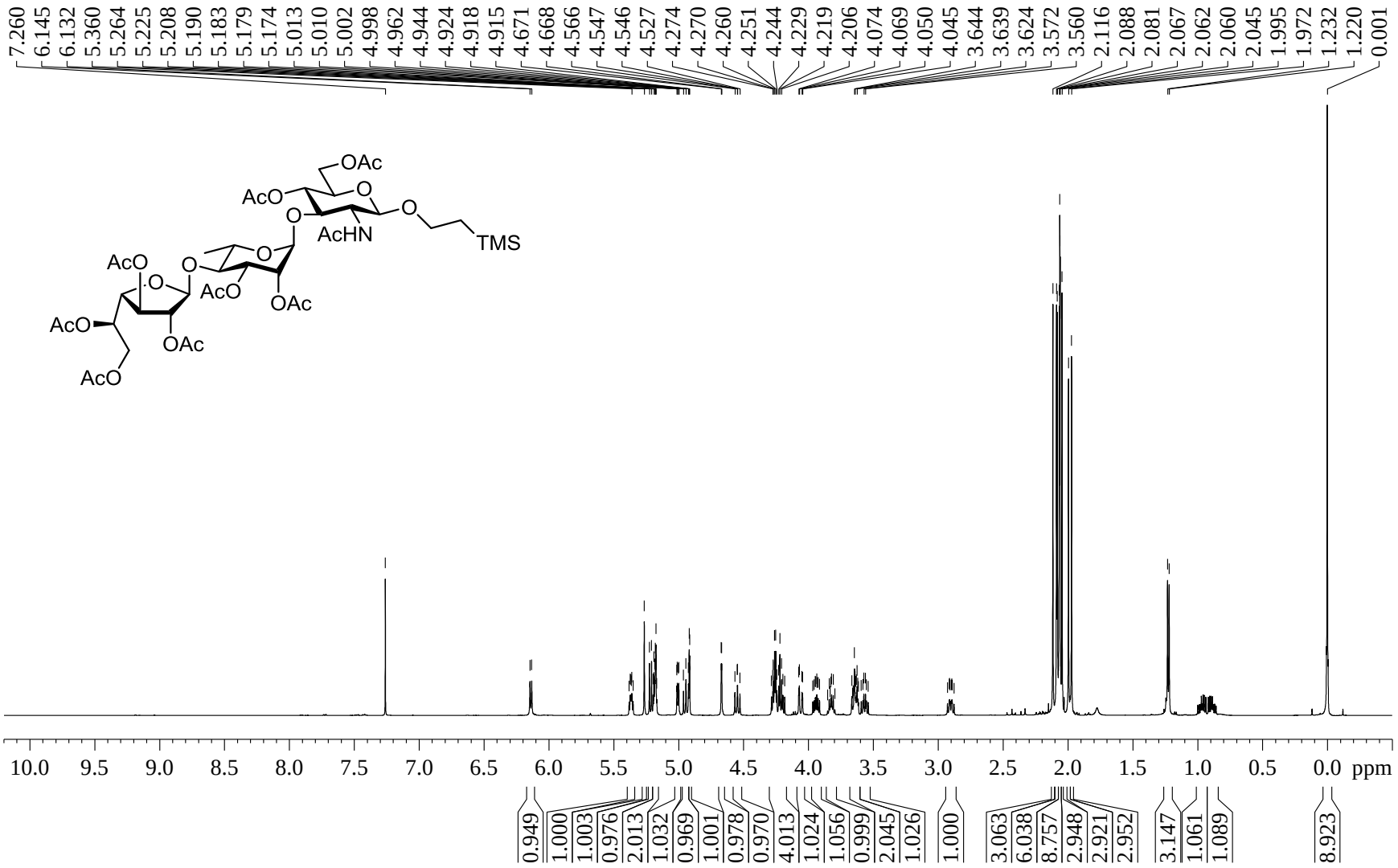
^{1}H NMR of Compound 35 (500 MHz, $CDCl_3$)



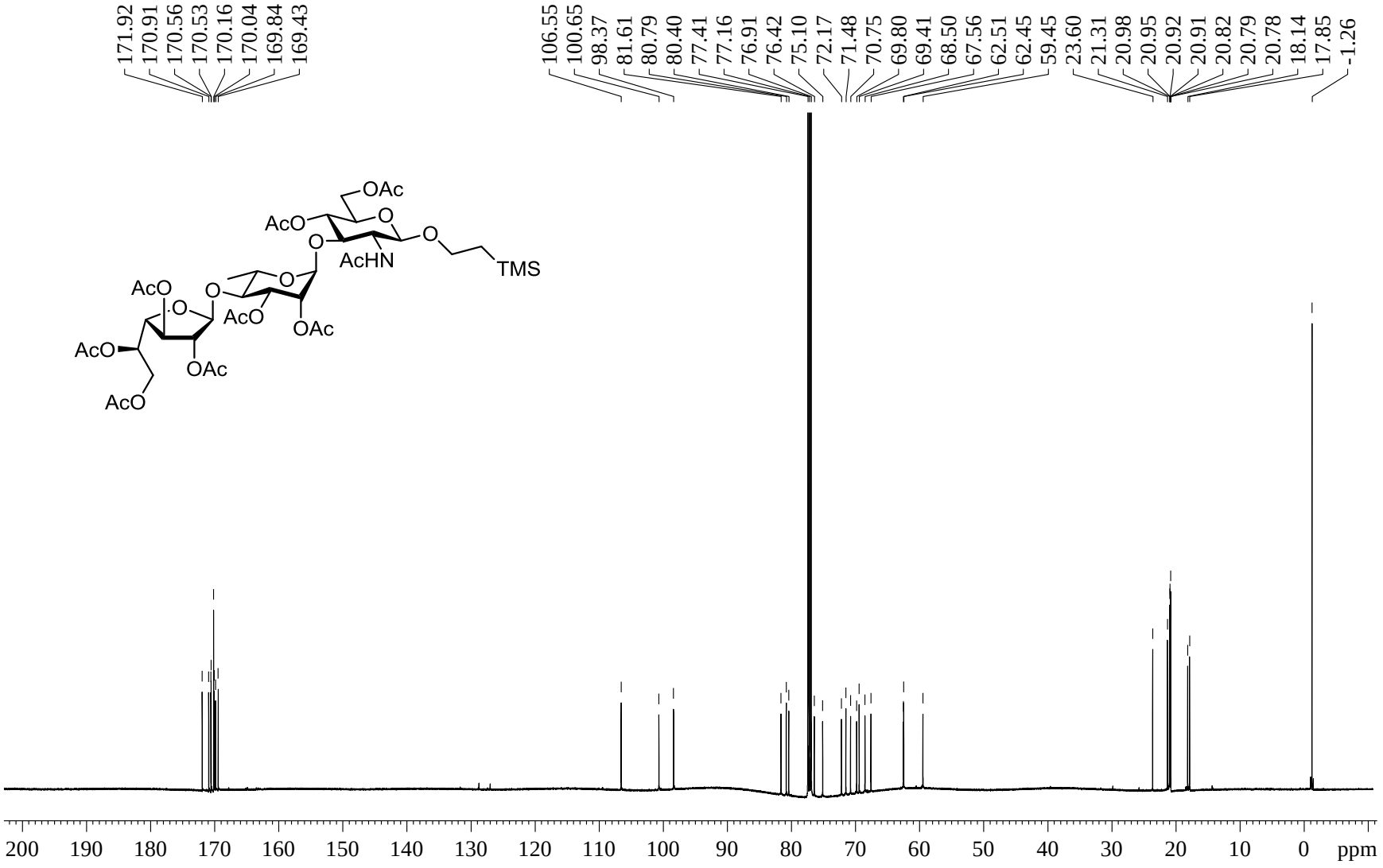
^{13}C NMR of Compound 35 (125 MHz, $CDCl_3$)



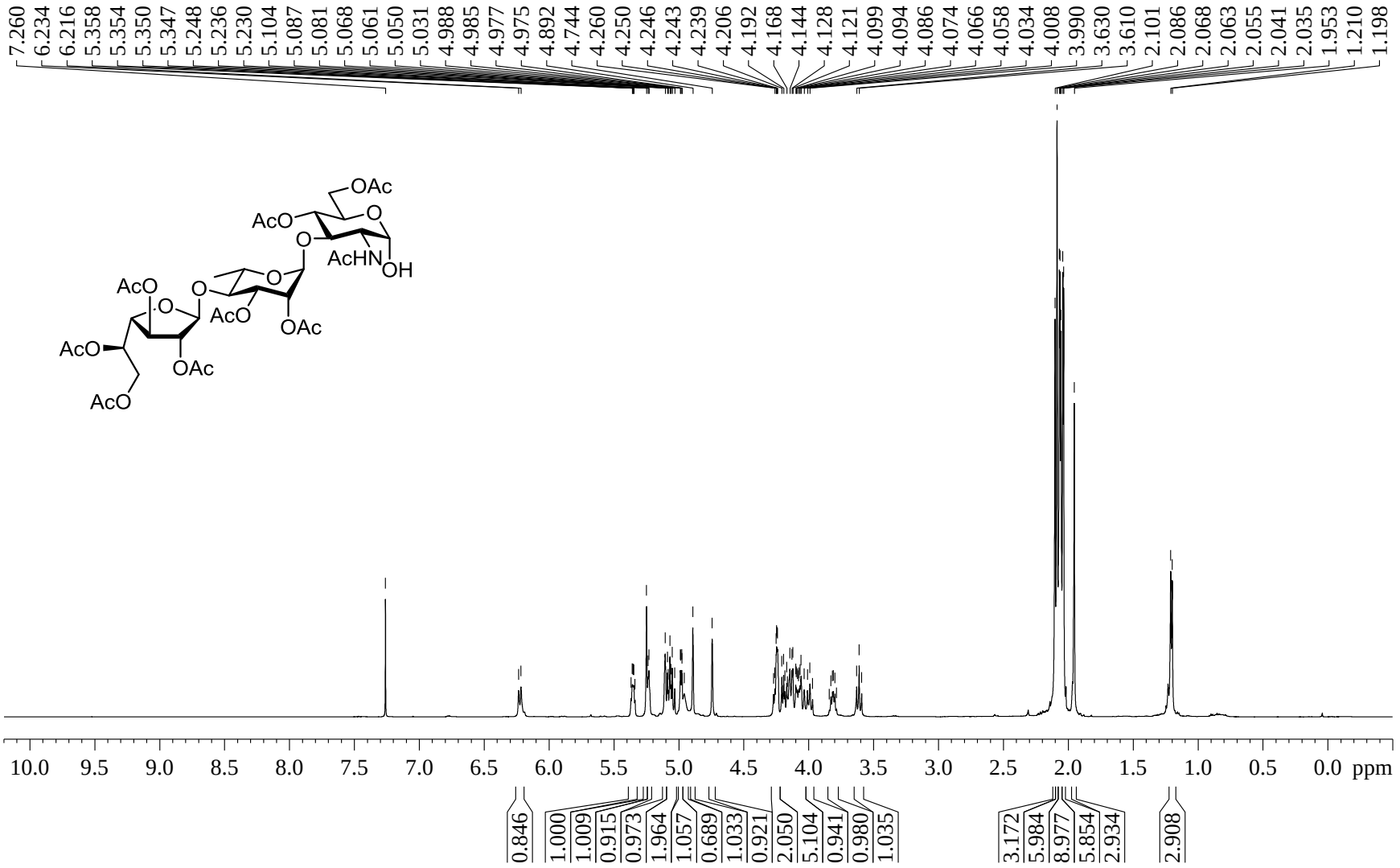
^{1}H NMR of Compound 36 (500 MHz, $CDCl_3$)



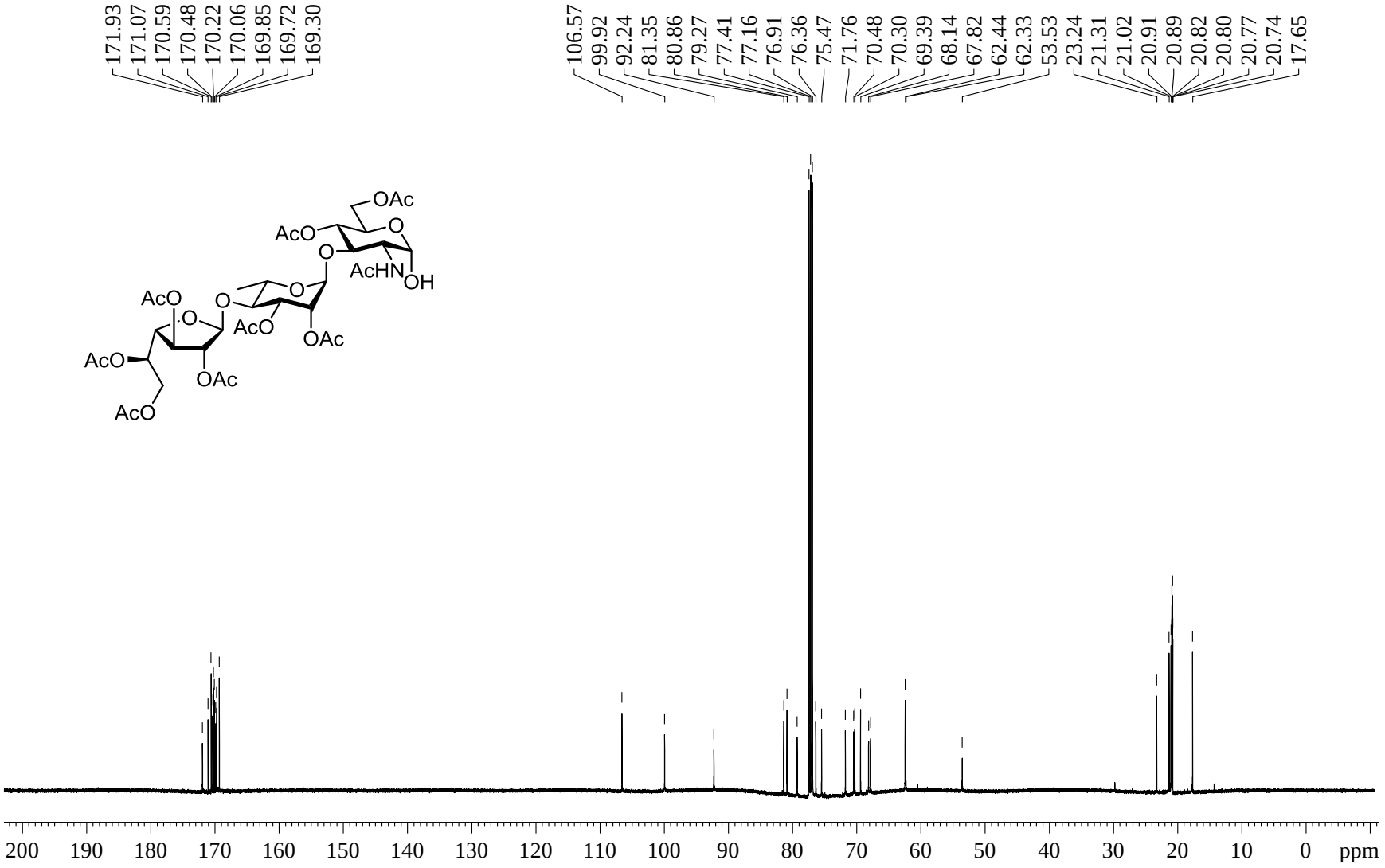
^{13}C NMR of Compound 36 (125 MHz, $CDCl_3$)



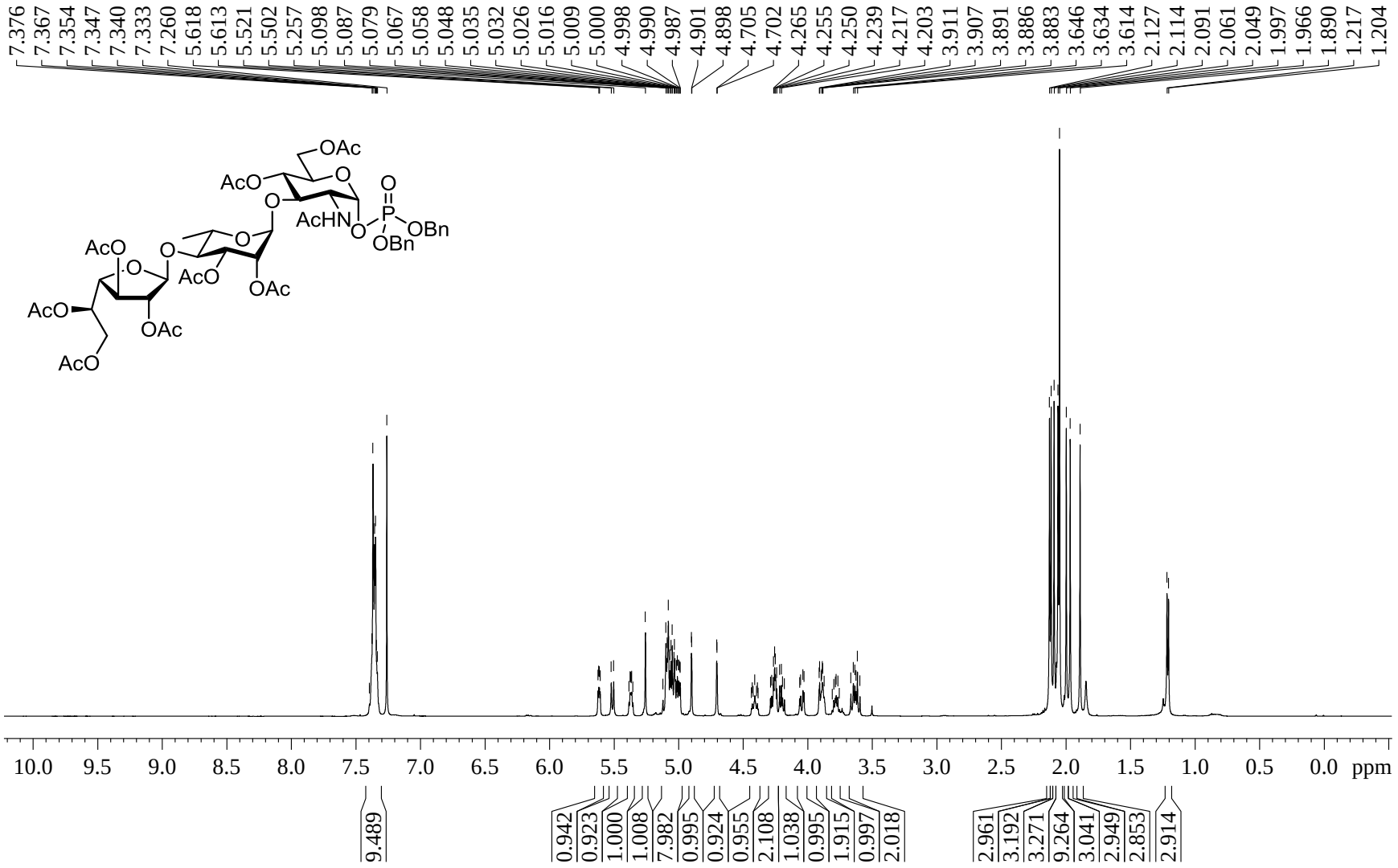
^{1}H NMR of Compound 37 (500 MHz, $CDCl_3$)



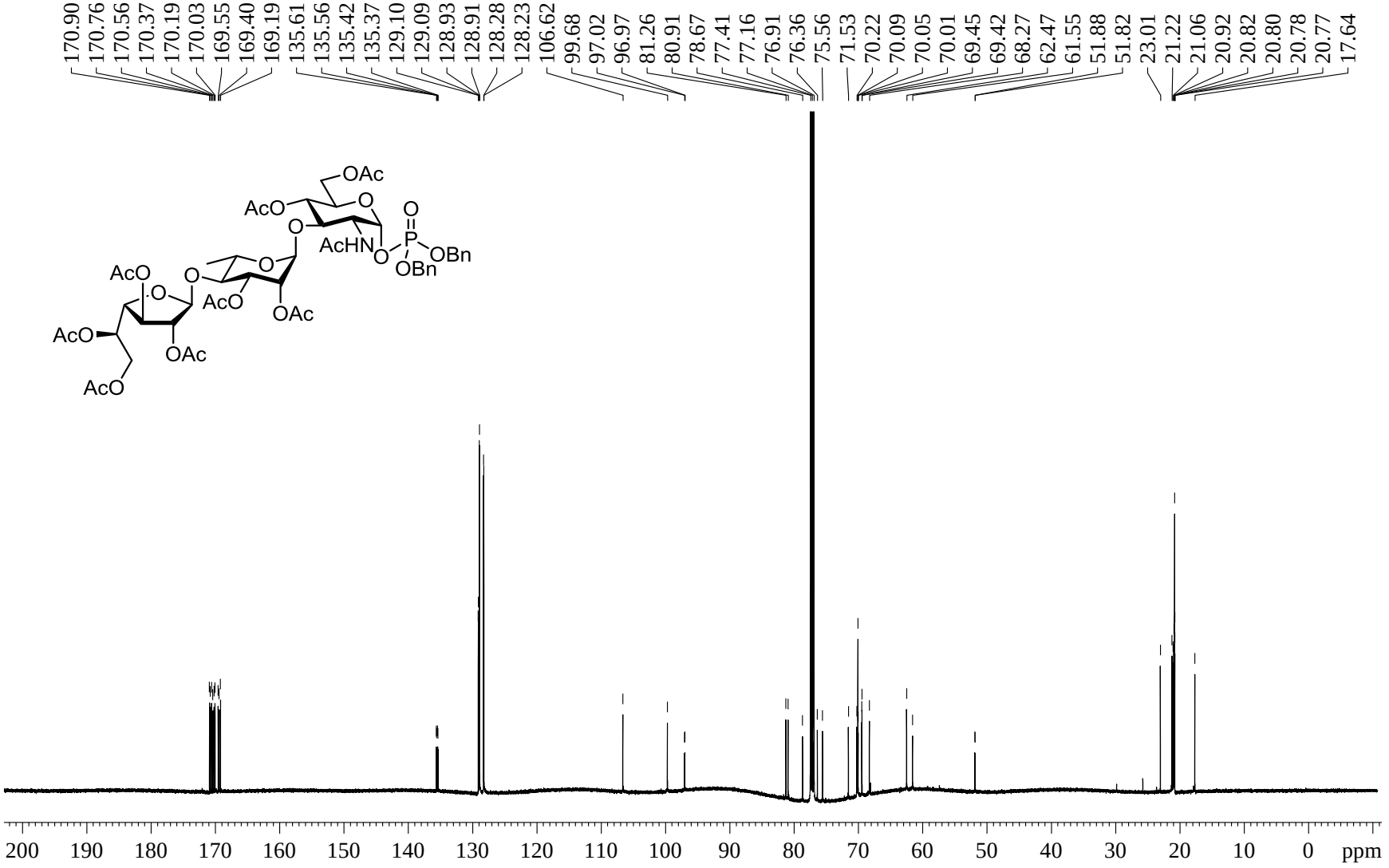
^{13}C NMR of Compound 37 (125 MHz, $CDCl_3$)



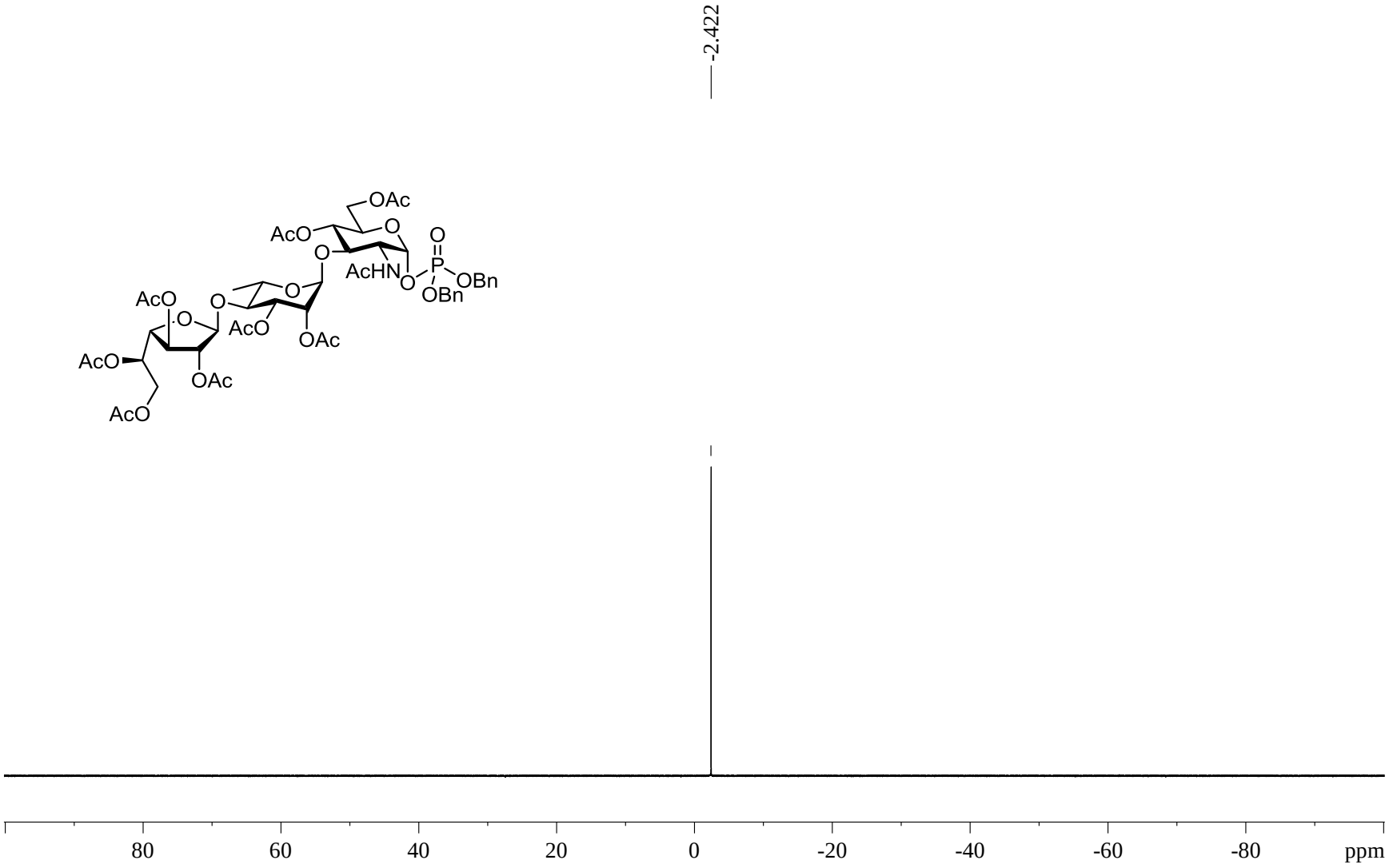
^{1}H NMR of Compound 38 (500 MHz, $CDCl_3$)



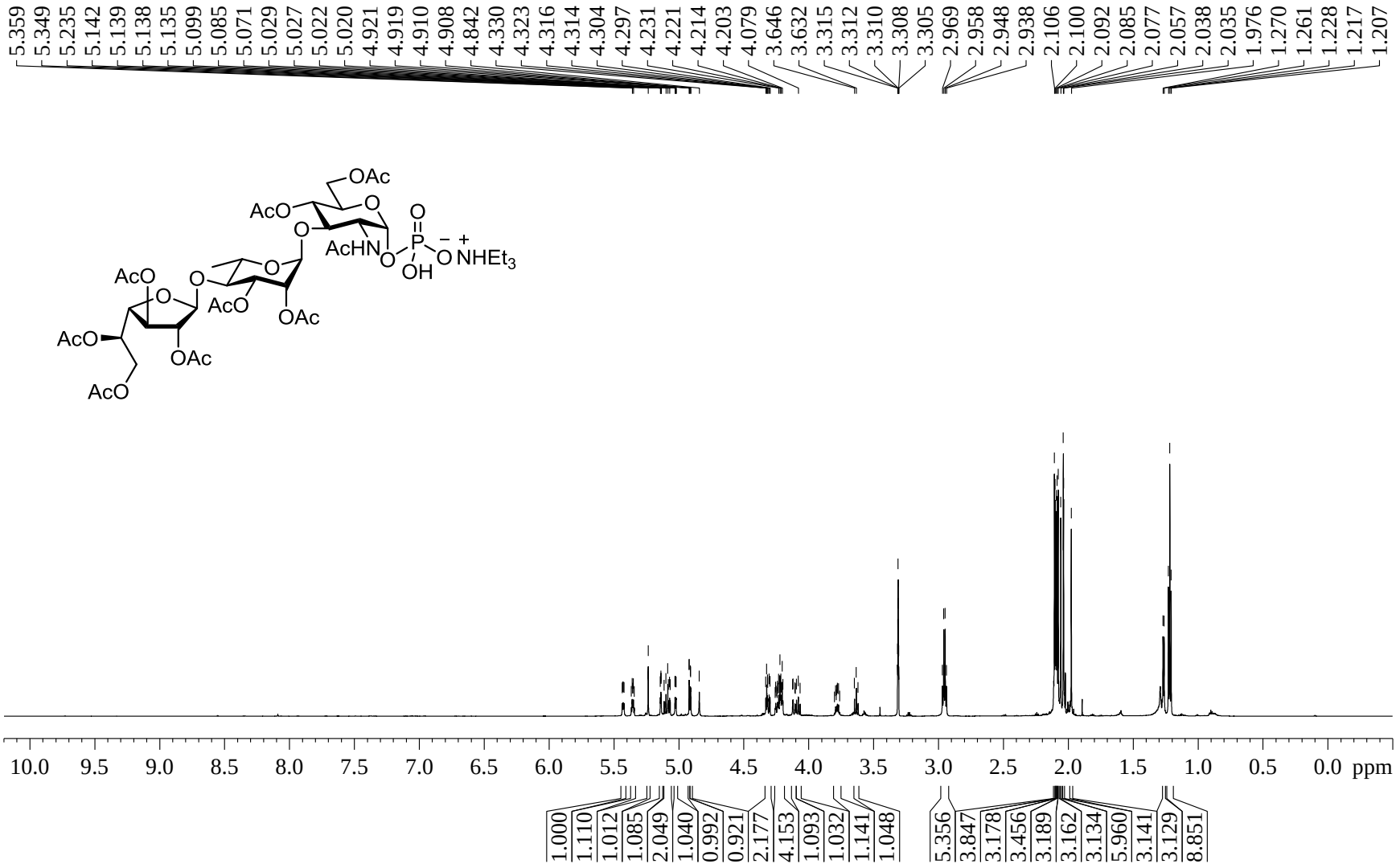
^{13}C NMR of Compound 38 (125 MHz, $CDCl_3$)



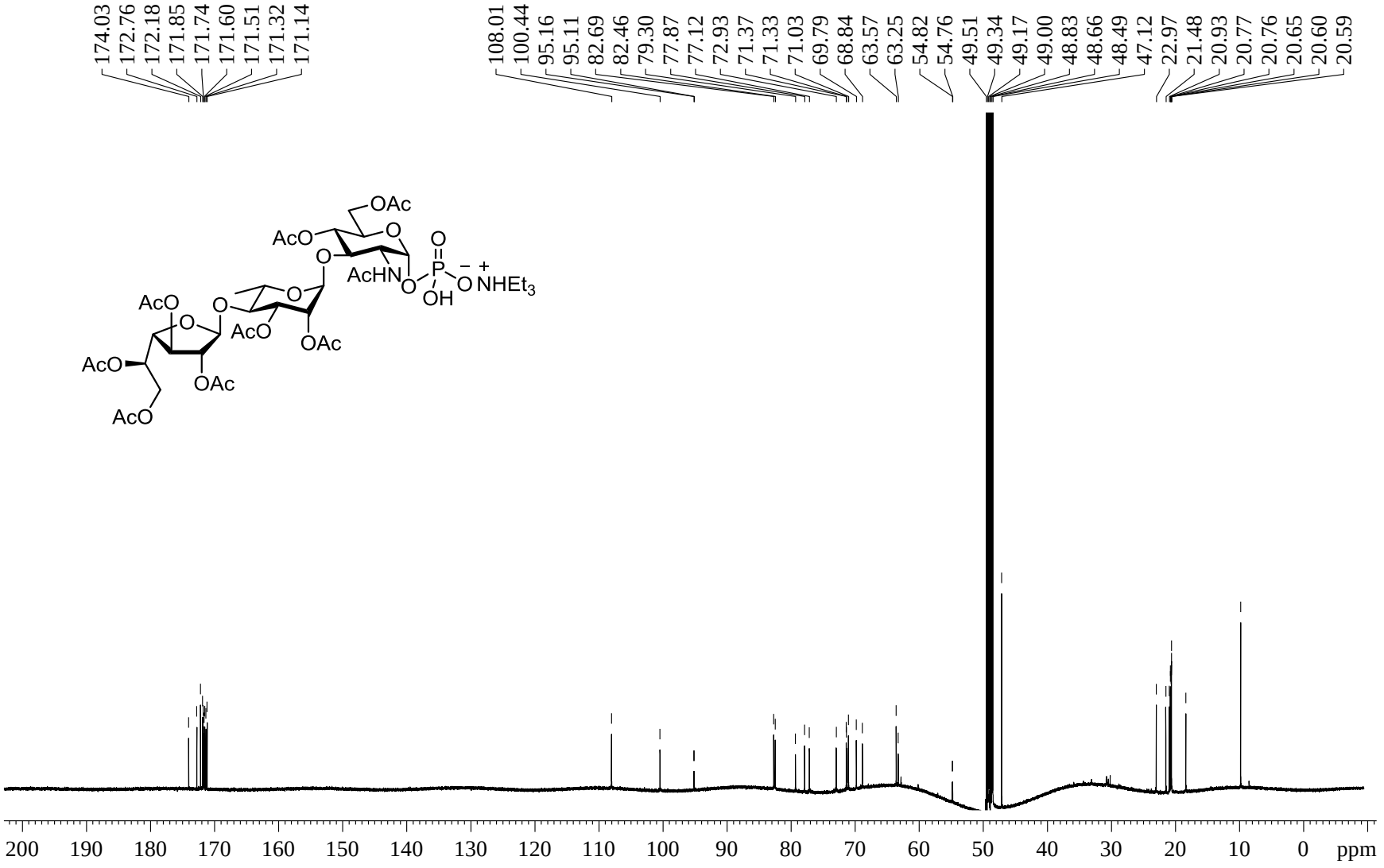
^{31}P NMR of Compound 38 (202 MHz, $CDCl_3$)



^{1}H NMRof Compound 39 (700 MHz, CD_3OD)



^{13}C NMR of Compound 39 (125 MHz, CD_3OD)



^{31}P NMR of Compound 39 (202MHz, CD_3OD)