Supporting Information

Structure-Based Design of Highly Potent HIV-1 Protease Inhibitors Containing New Tricyclic Ring P2-Ligands : Design, Synthesis, Biological and X-ray Structural Studies

Arun K. Ghosh,[*,†] *Satish Kovela,*[†] *Heather L. Osswald,*[†] *Masayuki Amano,*[∫,∧,] *Manabu Aoki,*[∫,∧,§]

Johnson Agniswamy,[‡] *Yuan-Fang Wang,*[‡] *Irene T. Weber,*[‡] *and Hiroaki Mitsuya*[∫,#,§]

Table of Contents

General Methods.

All chemicals and reagents were purchased from commercial suppliers and used without further purification unless otherwise noted. The following reaction solvents were distilled prior to use: dichloromethane from calcium hydride, diethyl ether and tetrahydrofuran from Na/benzophenone, methanol and ethanol from activated magnesium under argon. All reactions

were carried out under an argon atmosphere in either flame or oven-dried (120 ^{o}C) glassware. TLC analysis was conducted using glass-backed Thin-Layer Silica Gel Chromatography Plates (60 Å, 250 µm thickness, F-254 indicator). Column chromatography was performed using 230-400 mesh, 60 Å pore diameter silica gel. ^{1}H and ^{13}C NMR spectra were recorded at room temperature on a Bruker AV800, DRX-500 and ARX-400. Chemical shifts (δ values) are reported in parts per million, and are referenced to the deuterated residual solvent peak. NMR data is reported as: δ value (chemical shift, *J*-value (Hz), integration, where s = singlet, d = doublet, t = triplet, q = quartet, brs = broad singlet). Optical rotations were recorded on a Perkin Elmer 341 polarimeter. HRMS and LRMS spectra were recorded at the Purdue University Department of Chemistry Mass Spectrometry Center. HPLC analysis and purification was done an on Agilent 1100 series instrument using a YMC Pack ODS-A column of 4.6 mm ID for analysis and either 10 mm ID or 20 mm ID for purification. The purity of all test compounds was determined by HPLC analysis to be ≥95% pure.

Expression and purification of protease species

Expression and purification of protease were carried out as previously described [1]. Briefly, Rosetta (DE3) pLysS strain (Novagen) was transformed with an expression vector (pET-30a), which contained the genes of wild-type HIV-1^{NL4-3}-PR (PR^{WT}) using heat-shock. The culture was grown in a shake flask containing 30 mL of Luria broth plus kanamycin and chloramphenicol (LB$^{Km+/Cp+}$) at 37°C overnight. In the expression of PR^{WT}, twenty milliliter of the grown culture was added to 1 L of ZYM-10052 [1.0% N-Z amine, 0.5% yeast extract, 25 mM disodium hydrogenphosphate, 25 mM potassium dihydrogenphosphate, 50 mM ammonium chloride, 5 mM sodium sulfate, 1.0% glycerol, 0.05% glucose, 0.2% α-lactose, 2 mM magnesium sulphate] plus kanamycin and chloramphenicol (ZYM-10052$^{Km+/Cp+}$). The ZYM-10052$^{Km+/Cp+}$ culture was further continued at 37°C for 20~22 hours. Then the culture was spun down for pellet collection, and thus-obtained pellets were stored at −80 °C until use. For purification of PR^{WT}, the pellet was resuspended in buffer A [20 mM Tris, 1 mM EDTA, and 1 mM DTT] and lysed with sonication. The cell lysates were separated into a supernatant fraction and an inclusion body fraction with centrifugation. PR^{WT} was confirmed to be present in the inclusion body fraction, which was washed five times with buffer A containing 2 M urea and then with buffer A without urea. The twice-washed pellet was solubilized and PRs were unfolded with 100 mM formic acid (pH 2.8). The unfolded PRs were purified using the fast protein liquid chromatography system (ÄKTA pure 25; GE Healthcare) and separated using the reverse phase chromatography column (RESOURS RPC 3 mL; GE Healthcare) using the gradient of buffer B [1.0% formic acid, 2.0% acetonitrile] and buffer C [1% formic acid, 70% acetonitrile]. The flow rate was set to 1.0 mL min^{-1} and the column was equilibrated with 75% buffer B and 25% buffer C. Then, the amount of buffer C was increased to 75% over a 30 min period (10-time the column volume). PR^{WT} was eluted with 35~50% buffer C. After the elution, buffer C amount was increased to 100% in 6 min and returned to the starting condition over the next 6 min. The peak fractions including PR^{WT} were collected and three-time diluted with buffer B. The diluted PR^{WT} solution was injected into the ÄKTA pure 25 again and the targeted PR^{WT} was purified using the same purification step as described above. The collected fractions containing PR^{WT} were subjected to desalting (HiTrap Desalting; GE Healthcare) and the eluted solution was equilibrated using 100 mM formic acid and stored at -80 °C until use.

The unfolded PR^{WT} was refolded with the addition of a neutralizing buffer A [100 mM ammonium acetate pH 6.0, 0.005% Tween-20], making the final pH 5.0 to 5.2. The PR^{WT}-containing solution was run through Amicon Ultra-15 10K centrifugal filter units (Millipore), giving a solution containing PR (5~8 mg/ml) in 10 mM ammonium acetate pH 5.0 and 0.005% Tween-20. Occasionally, twice greater concentrations of a test compound were used for crystalization. After centrifugation, the supernatants were collected and subjected to crystallization using the hanging-drop vapor diffusion method. Nextal Tubes ProComplex Suite (QIAGEN) and Wizard Crystallization Screen Series (Emerald BioSystems) were used for the first screening to determine the optimum crystallization condition.

Determination of X-ray structure of HIV-1 protease-inhibitor complex

The HIV-1 protease was expressed and purified as described previously [2]. The PR/GRL-19-17 (compound **5c**) complex was crystallized by the hanging drop vapor diffusion method with well solution of 1.25 M NaCl, 0.1 M Sodium Acetate, pH 4.8 while PR/GRL-52-16 (compound **4a**) crystals were grown with a reservoir solution of 0.65 M NaCl, 0.1M Sodium Acetate, pH 6.0. Diffraction data were collected on a single crystal cooled to 90 K at SER-CAT (22-BM beamline), Advanced Photon Source, Argonne National Lab (Chicago, USA) with X-ray wavelength of 1.0 Å. The two-diffraction data were processed by HKL-2000 [3] to a Rmerge of 5.6% and 7.9%. The complex structures were solved by PHASER [4] in CCP4i Suite [5, 6,7] using previously determined isomorphous structure with PDB code 3NU3[8] as start model. The protease and inhibitors **5c** and **4a** complexes were refined by SHELX-2014 [9, 10] up to 1.3 Å resolution and by REFMAC5 [11] to 1.25 Å resolution respectively. PRODRG-2 [12] and Jligand [13] were used to construct inhibitors and the restraints for refinement. COOT [14,15] was used for model building. Anisotropic atomic displacement parameters (B factors) were applied for all atoms including solvent molecules. The final refined solvent structure comprised one Na^+ ion, two Cl^- ions, two acetate ion and 200 water molecules for protease/**5c** and one Na^+ ion, two Cl^- ions, one glycerol, one formic acid and 222 water molecules for protease/**4a**, respectively. The crystallographic statistics are listed in Table 1. The coordinates and structure factors of protease/**4a** and protease/**4c** were deposited in the Protein Data Bank [16] with code 6VOD and 6VOE.

References

(1) Mahalingam, B.; Louis, J. M.; Hung, J.; Harrison, R. W.; Weber, I. T.; Structural implications of drug-resistant mutants of HIV-1 protease: High-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins* **2001**, *43*, 455–464.

(2) Agniswamy, J.; Kneller, D. W.; Brothers, R.; Wang, Y-F.; Harrison, R. W.; Weber, I. T. Highly Drug-Resistant HIV-1 Protease Mutant PRS17 Shows Enhanced Binding to Substrate Analogues. ACS Omega 2019, 4, 8707-8719.

(3) Otwinowski, Z.; Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology, 276: Macromolecular Crystallography, Part A; Carter, C.W., Jr., Sweet, R. M., Eds.; Academic Press: New York, 1997; pp 307-326.

(4) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. J. Appl. Crystallogr. 2007, 40, 658-674.

(5) M. D. Winn et al. Overview of the CCP4 suite and current developments. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2011, 67, 235-242
(6) Collaborative Computational Project, Number 4 The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr., Sect. D: Biol. Crystallogr., 1994, 50, 760-763.
(7) Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E. A graphical user interface to the CCP4 program suite. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2003, 59, 1131-1137.
(8) Shen, C.-H.; Wang, Y.-F., Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. Febs J. 2010, 277, 3699-3714
(9) Sheldrick, G. M. A short history of SHELX. Acta Crystallogr., Sect. A: Found. Crystallogr. 2008, 64, 112-122.
(10) Sheldrick, G.M. Crystal structure refinement with SHELXL. Acta Crystallogr. C71 (2015) 3-8.
(11) Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.
(12) Schuettelkopf, A. W.; van Aalten, D. M. F. PRODRG;a tool for high-throughput crystallography of protein-ligand complexes." Acta Crystallogr., Sect. D: Biol. Crystallogr. 2004, 60, 1355-1363.
(13) Lebedev, A. A., P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov (2012). "JLigand: a graphical tool for the CCP4 template-restraint library." Acta Crystallogr D Biol Crystallogr **68**(Pt 4): 431-440.
(14) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2010, 66, 486-501.
(15) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2004, 60, 2126-2132.
(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. Nucleic Acids Res. 2000, 28, 235-242.

Table 1: Crystallographic Data Collection and Refinement Statistics

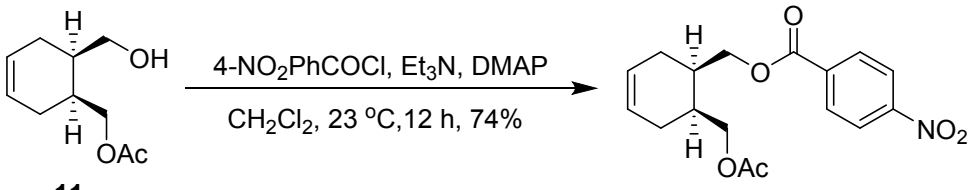
Caomplex	PR/GRL-19-17 (**5c**)	PR/GRL-52-16 (**4a**)
Space group	$P2_12_12$	$P2_12_12$
Unit cell dimensions: (Å)		
a	58.49	58.78
b	86.24	86.21
c	45.96	46.02
Resolution range (Å)	50-1.30 (1.35-1.30)	50-1.25 (1.29-1.25)
Unique reflections	54,691 (3,776)	58388 (3202)
R_{merge} (%) overall (final shell)	5.6 (49.5)	7.9 (51.8)
I/σ(I) overall (final shell)	28.1 (3.2)	16.9 (2.1)
Completeness (%) overall (final shell)	94.4 (66.4)	89.4 (49.9)
Redundancy overall (final shell)	6.1 (3.8)	6.4 (3.2)
Refinement		
R (%)	16.0	15.9

R_{free} (%)	20.2	18.5
No. of solvent atoms	200	234
RMS deviation from ideality		
Bonds (Å)	0.01	0.01
Angle distance	0.03 (Å)	0.03 degree
Average B-factors (Å^2)		
Wilson Plot B factor	13.5	12.5
Main-chain atoms	15.5	12.7
Side-chain atoms	21.7	17.6
Whole chain atoms	18.5	15.2
Inhibitor	14.2	10.1
Solvent	26.1	24.1

Cells, viruses, and antiviral agents. Human CD4$^+$ MT-2 cells were grown in RPMI-1640-based culture medium supplemented with 10% fetal calf serum (FCS: JRH Biosciences, Lenexa, MD), 50 unit/mL penicillin, and 100 □g/mL of kanamycin. The following HIV-1 viruses were employed for the drug susceptibility assay (see below): a laboratory HIV-1strain (HIV-1_{LAI}), a clinical HIV-1 strain isolated from drug-naive patients with AIDS (HIV-$1_{ERS104pre}$) (1), and six HIV-1 clinical isolates which were originally isolated from patients with AIDS, who had received 9 to 11 anti-HIV-1 drugs over the past 32 to 83 months, and were genotypically and phenotypically characterized as multi-PI-resistant HIV-1 variants (1, 2). All such primary HIV-1 strains were passaged once or twice in 3-day old phytohemagglutinin-activated peripheral blood mononuclear cells (PHA-PBM), and the culture supernatants were stored at –80 °C until use. Amprenavir (APV) was received as a gift from Glaxo-Wellcome, Research Triangle Park, NC. Darunavir (DRV) was synthesized as previously described (3).

1. Yoshimura, K., et al. *Proc. Natl. Acad. Sci.* USA **96**, 8675-8680 (1999).
2. Koh, Y., et al. *Antimicrob. Agents Chemother.* **53**, 987-996 (2009).
3. Koh Y, et al *J Mol Biol* **282**, 28709-28720 (2007)

((1*R*,6*S*)-6-(Acetoxymethyl)cyclohex-3-en-1-yl)methyl 4-nitrobenzoate (S2):

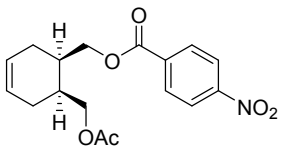


11 **S2**

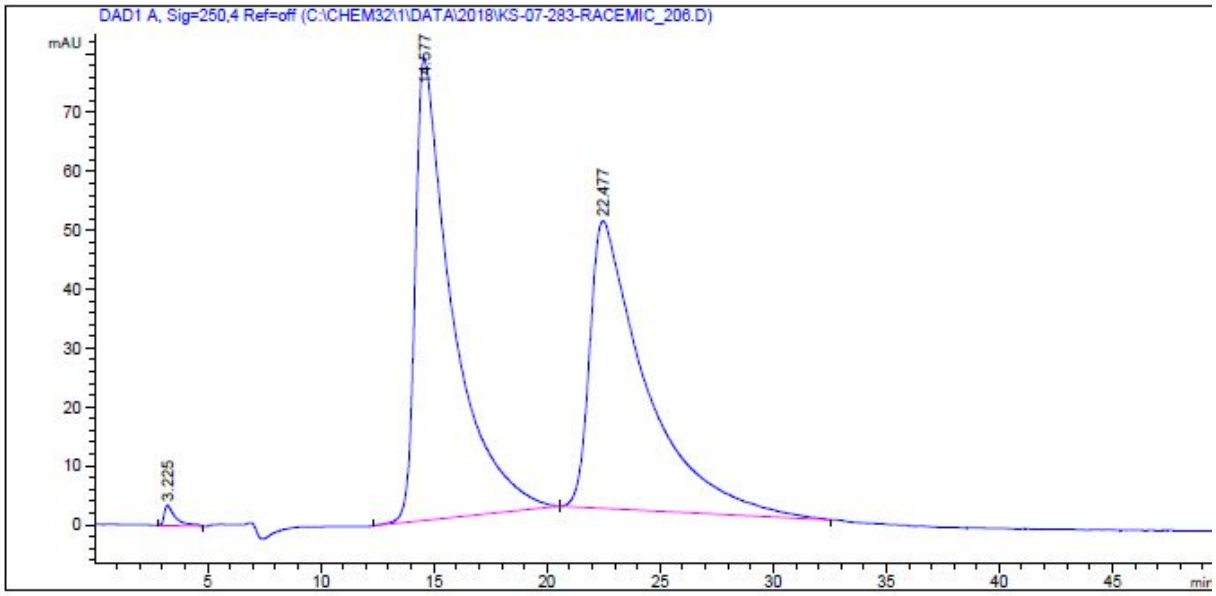
To a stirred solution of alcohol **11** (15 mg, 0.08 mmol) in CH_2Cl_2 were added Et_3N (17 µL, 0.12 mmol), 4-Nitrobenzoyl chloride (16 mg, 0.08 mmol) and DMAP (1 mg, 0.008 mmol) at 0 °C. The reaction mixture was stirred at 23 °C for 12 h. The reaction mixture was quenched with water and extracted with EtOAc (2×15 mL). The combined organic layer was washed with brine solution, dried over Na2SO4, filtered and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (10% EtOAc in hexane) to afford **S2** (20 mg, 74%) as an oil. R_f = 0.7 (30% EtOAc/hexanes). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.31 – 8.27 (m, 2H), 8.23 – 8.18 (m, 2H), 5.72 – 5.64 (m, 2H), 4.46 – 4.31 (m, 2H), 4.20 – 4.06 (m, 2H), 2.46 – 2.38 (m, 1H), 2.35 – 2.19 (m, 3H), 2.10 – 1.95 (m, 5H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.9, 164.5, 150.4, 135.5, 130.6, 125.1, 124.9, 123.5, 66.5, 65.0, 33.7, 33.6, 26.5, 26.5, 20.9. LRMS-ESI (m/z): 334.1 $[M+H]^+$.

HPLC data of racemic (±)-S2:

Sample Info : CHIRALCEL OZ-3
10% IPA/HEX
5 uL inj
1.0 mL/min
254 nm
1.0 mg/mL



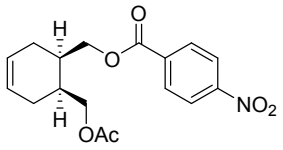
(±)-**S2**




```
=====================================================================
                        Area Percent Report
=====================================================================

Sorted By             :      Signal
Multiplier            :      1.0000
Dilution              :      1.0000
Sample Amount:        :      5.00000  [ng/ul]   (not used in calc.)
Use Multiplier & Dilution Factor with ISTDs


Signal 1: DAD1 A, Sig=250,4 Ref=off

Peak RetTime Type  Width     Area      Height     Area
  #   [min]        [min]   [mAU*s]     [mAU]       %
----|-------|----|-------|----------|----------|--------|
   1   3.225 BB    0.4712  117.68350    3.40532   0.6729
   2  14.577 BB    1.5301 8975.32617   78.49078  51.3179
   3  22.477 BB    2.2486 8396.65918   48.86779  48.0093

Totals :                  1.74897e4   130.76389

=====================================================================
                        *** End of Report ***
```

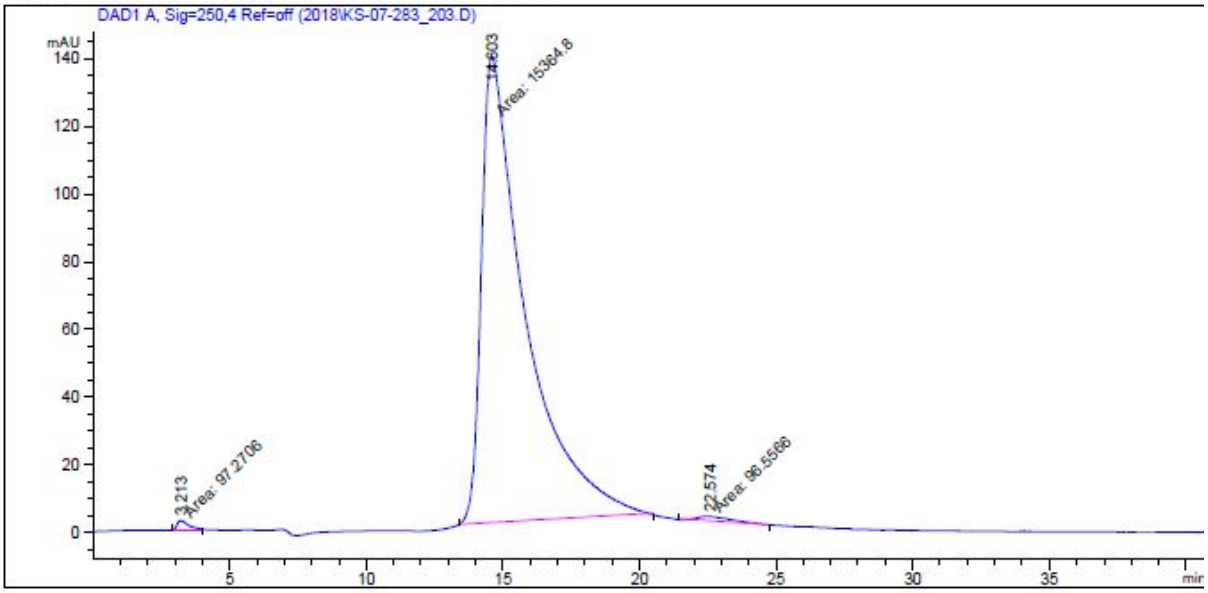
HPLC data of racemic (+) S2:

Sample Info : CHIRALCEL OZ-3
10% IPA/HEX
5 uL inj
1.0 mL/min
254 nm
1.0 mg/mL

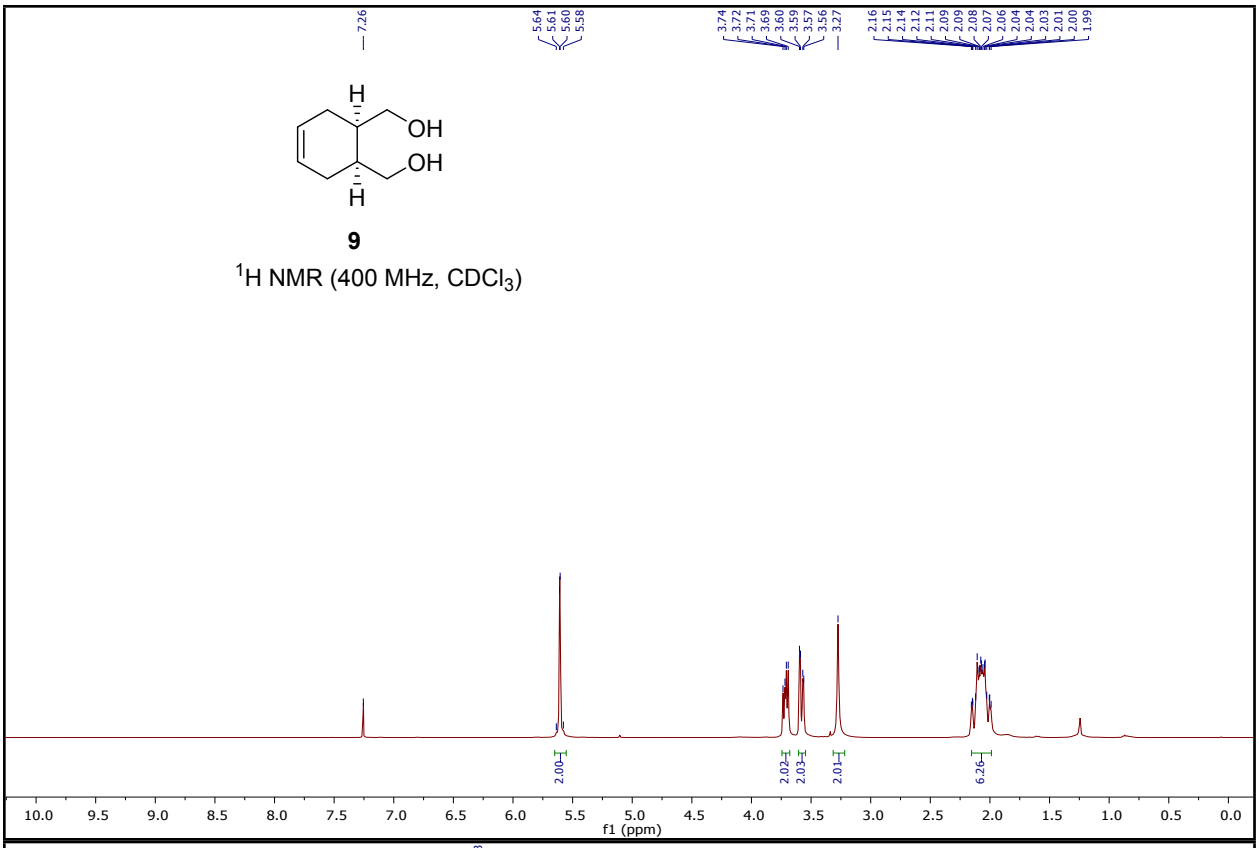


(+)-S2

Additional Info : Peak(s) manually integrated

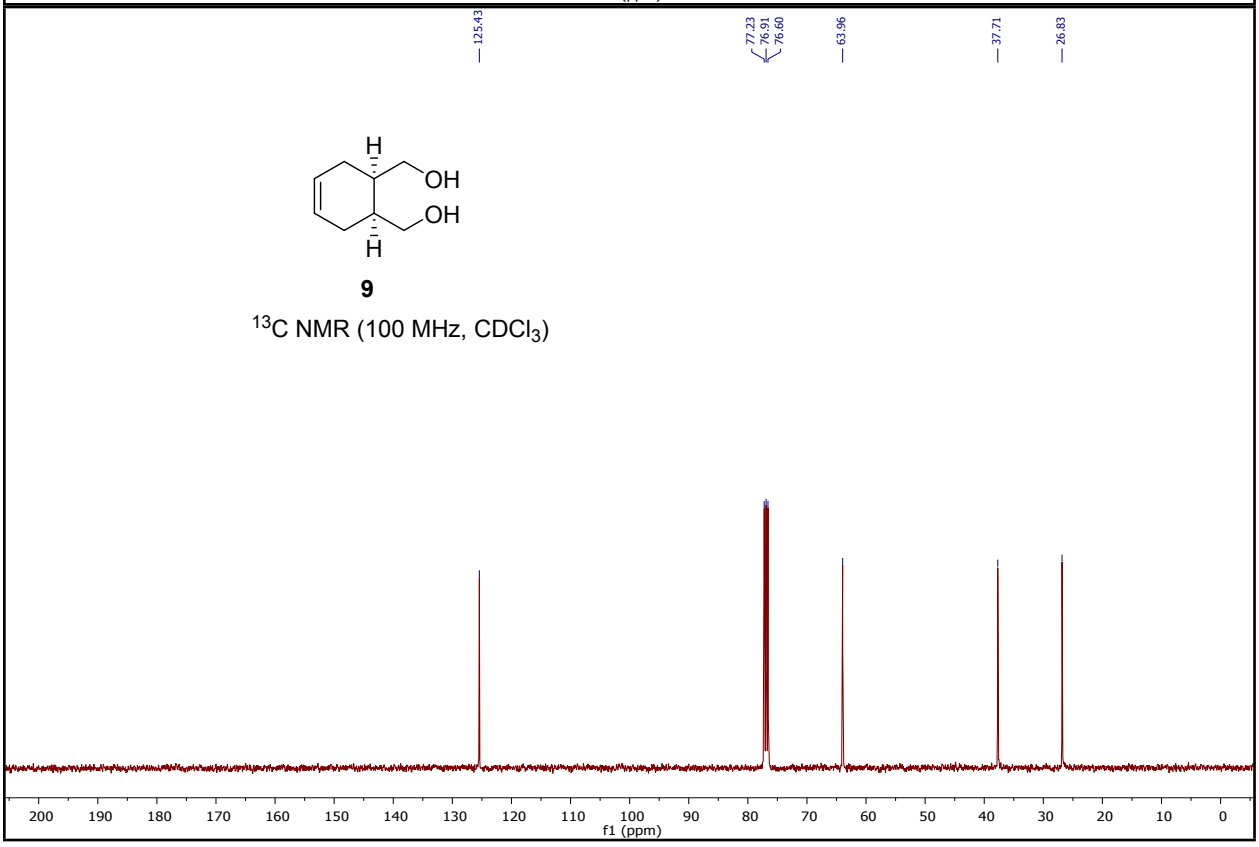



```
==================================================================================
                          Area Percent Report
==================================================================================

Sorted By             :      Signal
Multiplier            :      1.0000
Dilution              :      1.0000
Sample Amount:        :      5.00000  [ng/ul]    (not used in calc.)
Use Multiplier & Dilution Factor with ISTDs


Signal 1: DAD1 A, Sig=250,4 Ref=off

Peak RetTime Type  Width     Area      Height     Area
  #   [min]        [min]   [mAU*s]     [mAU]       %
----|-------|----|-------|----------|----------|--------|
   1   3.213 MM    0.5342   97.27065    3.03488   0.6252
   2  14.603 MM    1.8633 1.53648e4   137.43466  98.7542
   3  22.574 MM    1.3626   96.55656    1.18103   0.6206

Totals :                  1.55586e4   141.65057

==================================================================================
                          *** End of Report ***
```



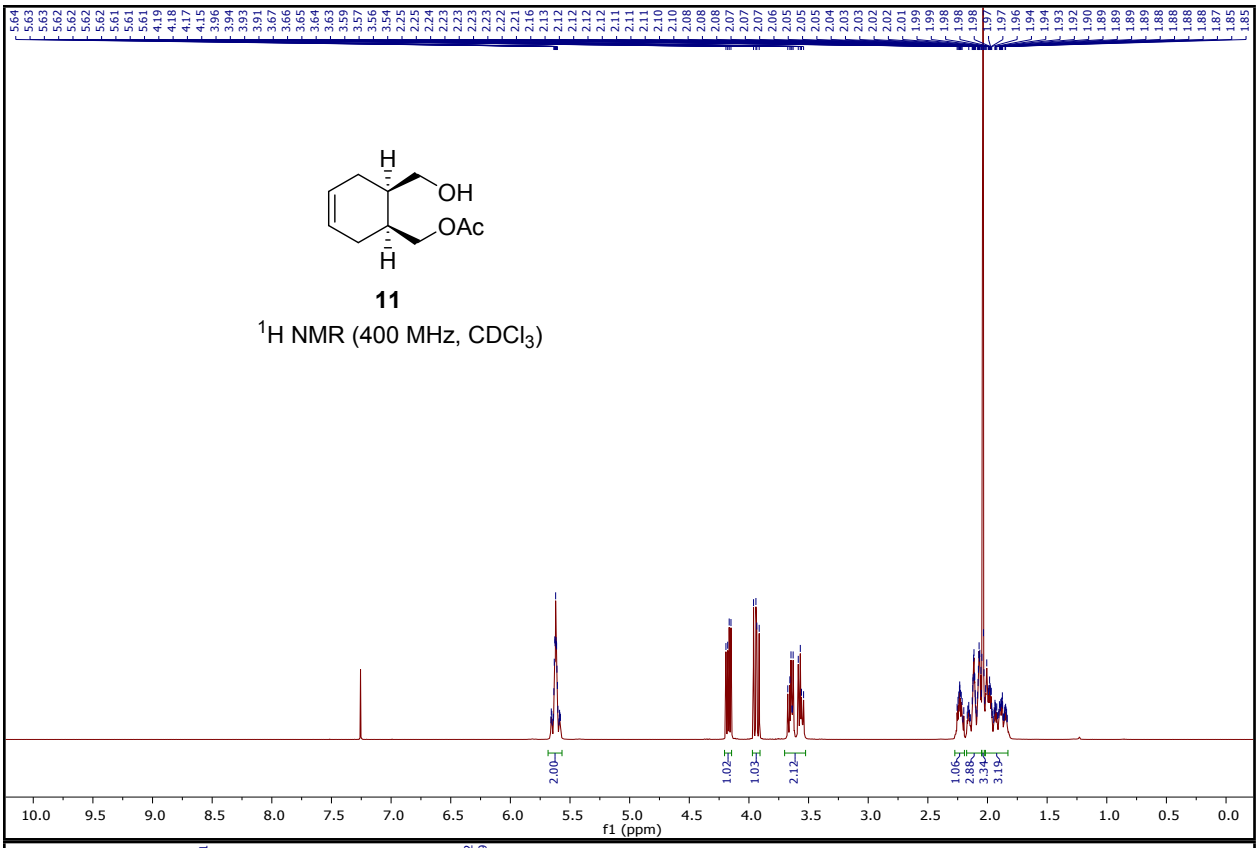
9

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



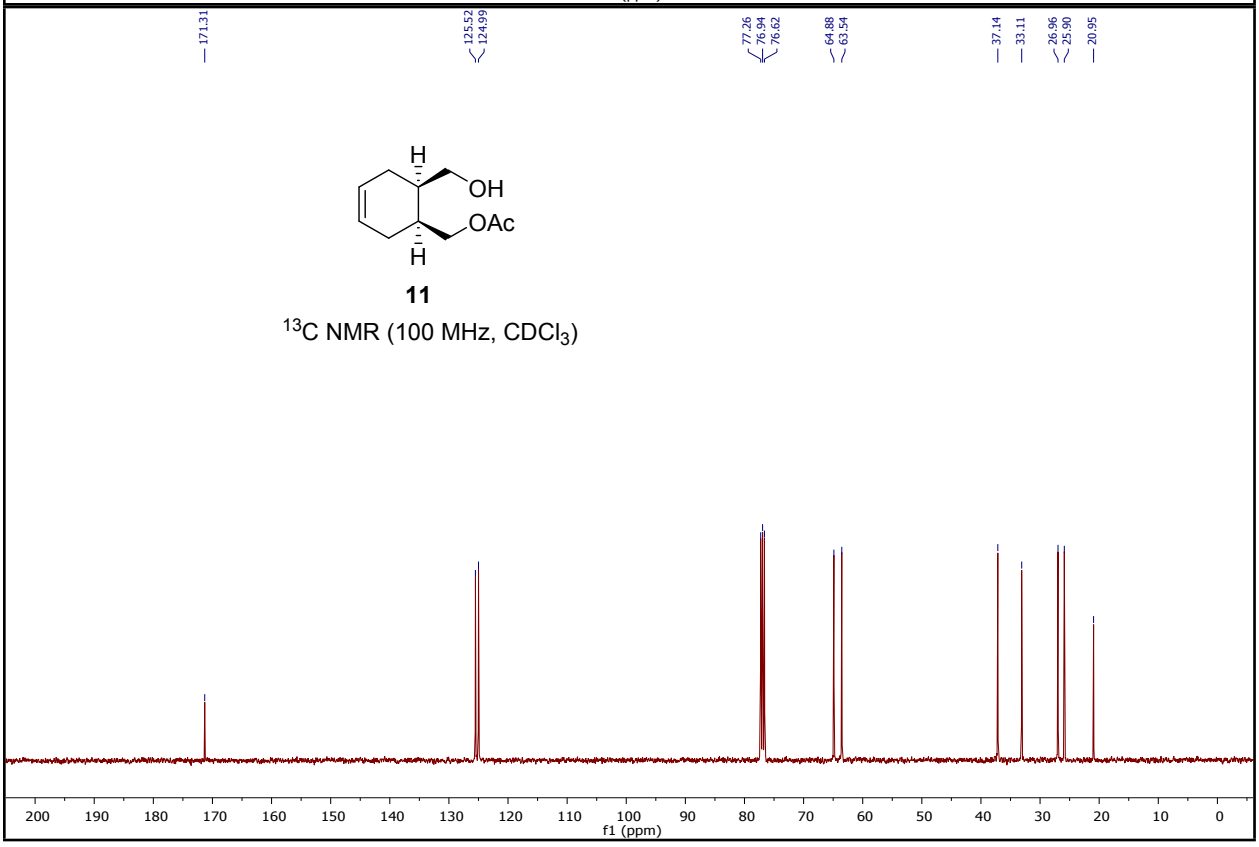
9

^{13}C NMR (100 MHz, $CDCl_3$)



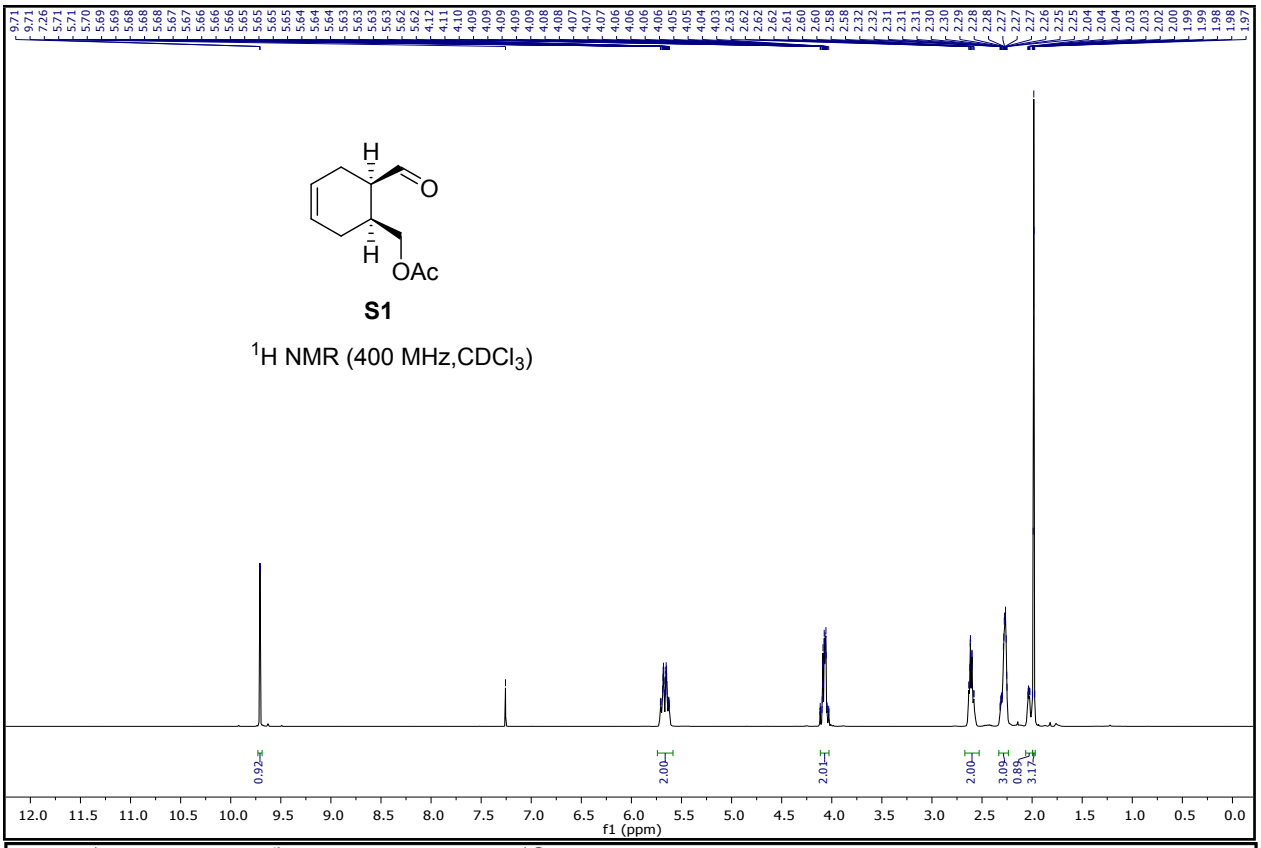
11

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



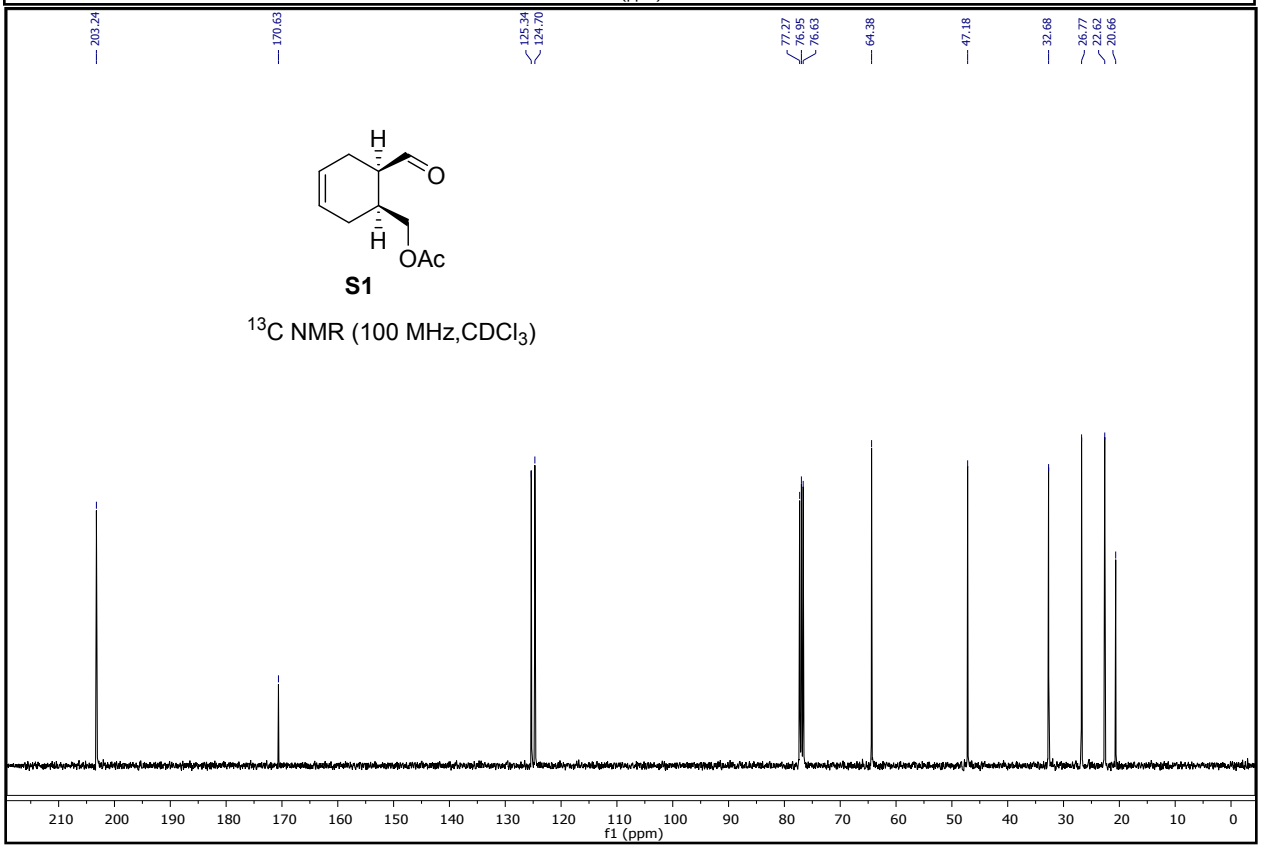
11

^{13}C NMR (100 MHz, $CDCl_3$)



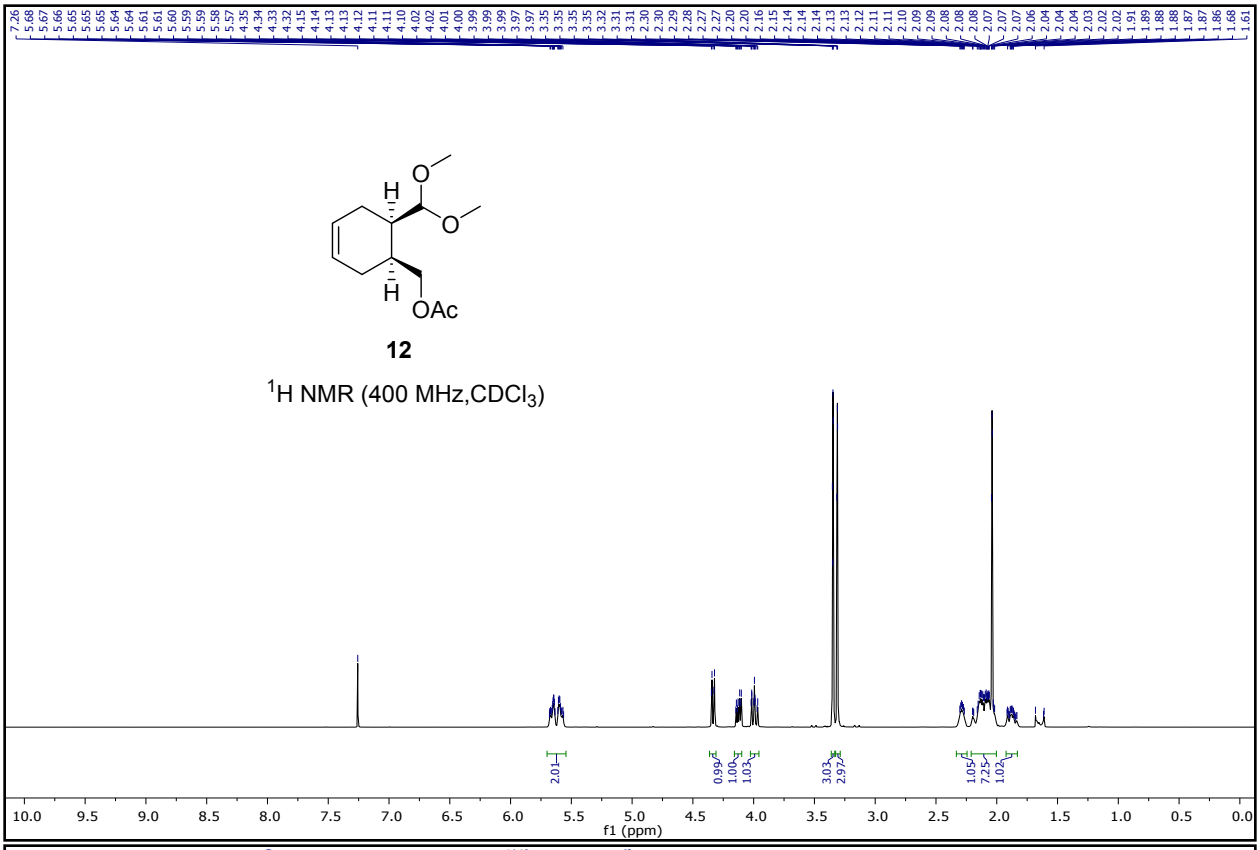
S1

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



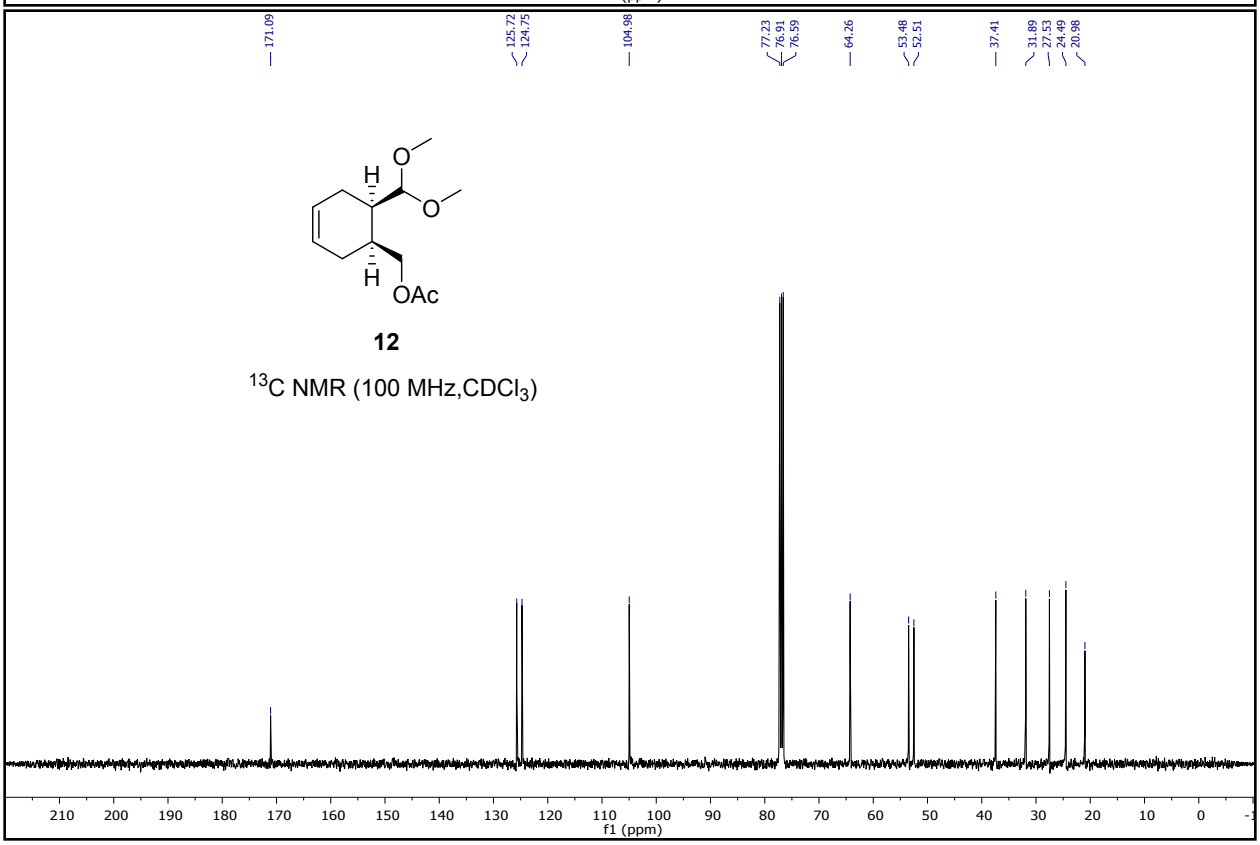
S1

^{13}C NMR (100 MHz,$CDCl_3$)



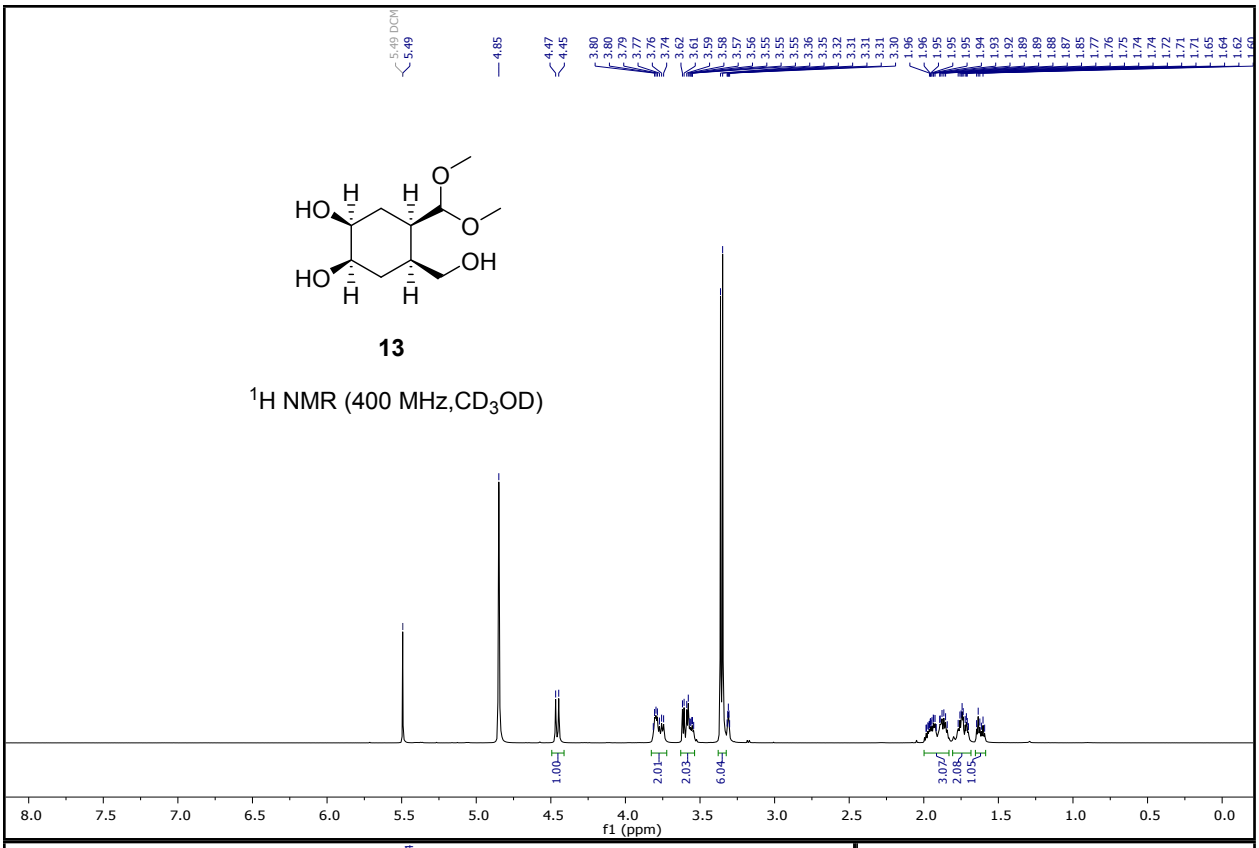
12

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



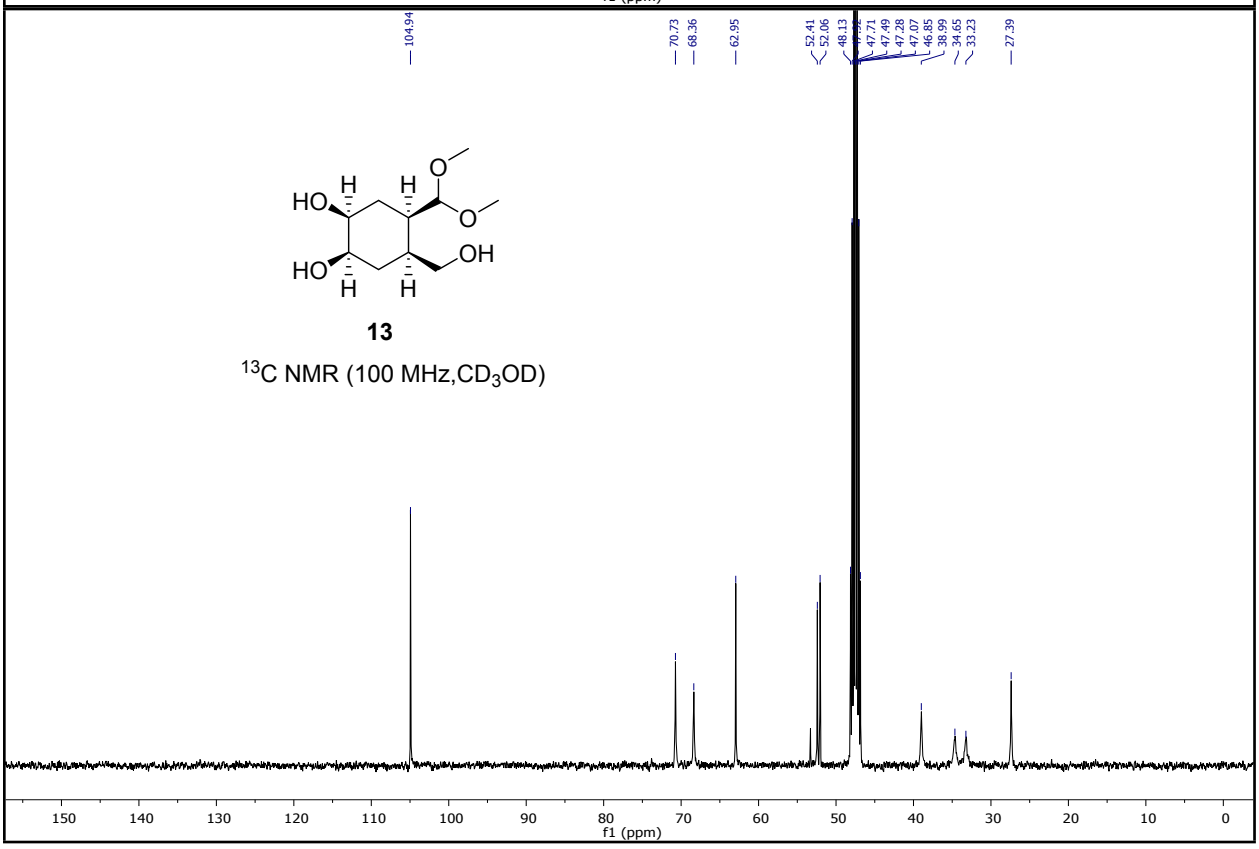
12

^{13}C NMR (100 MHz,$CDCl_3$)



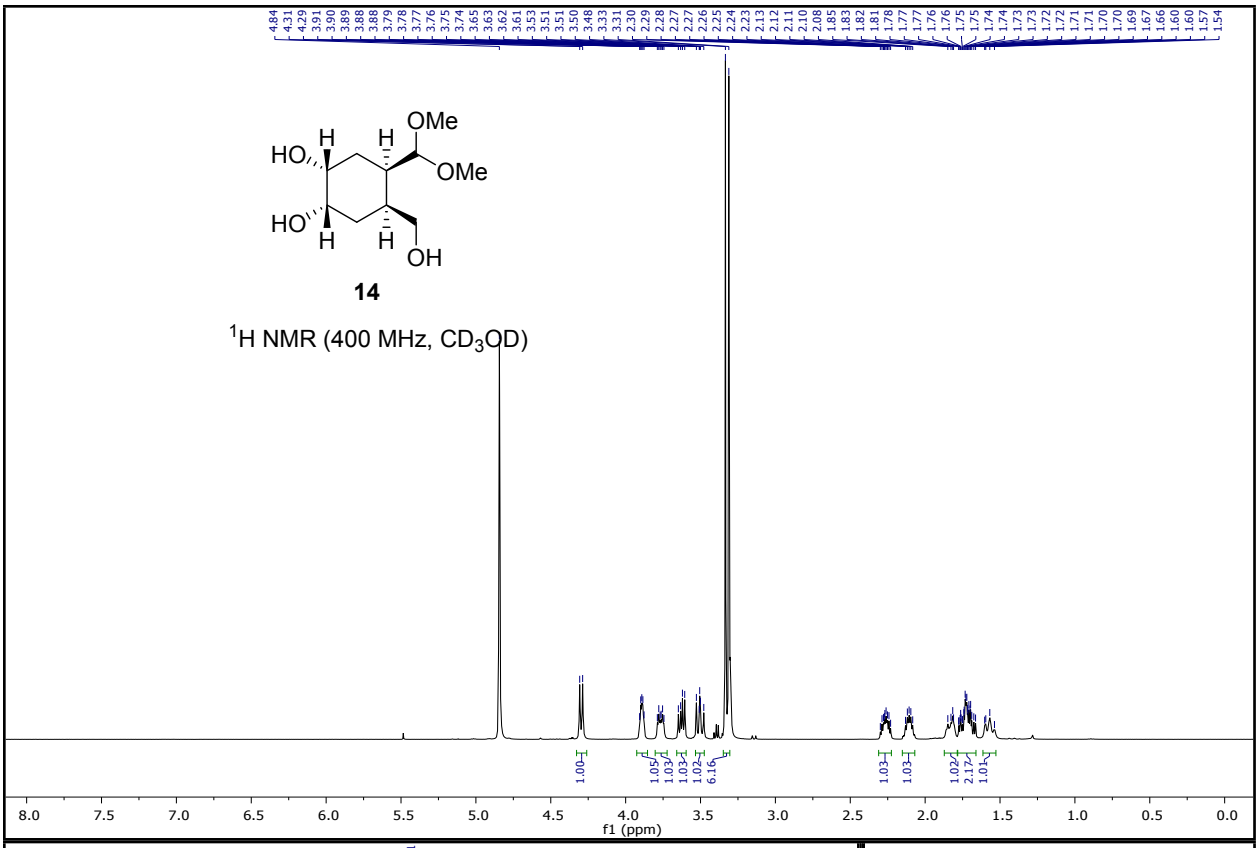
13

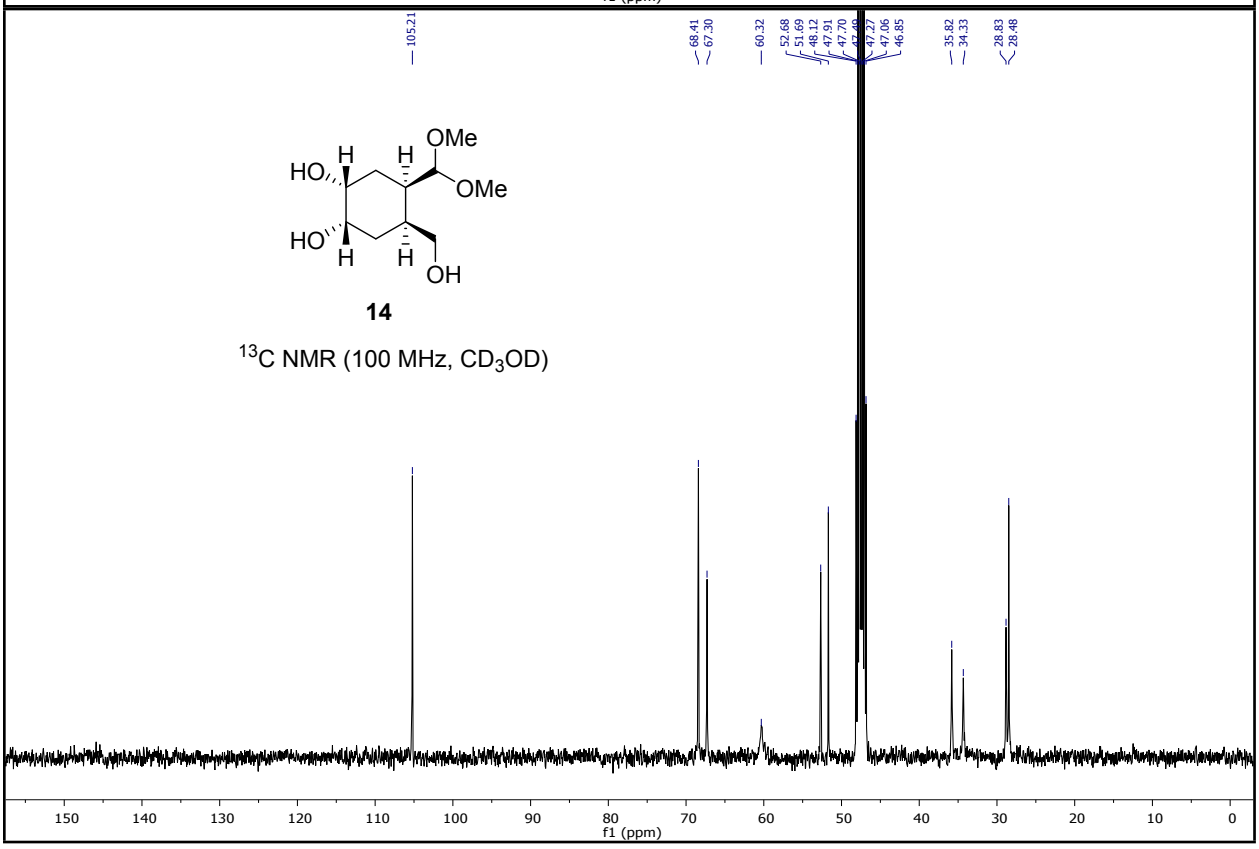
^{1}H NMR (400 MHz,CD_3OD)

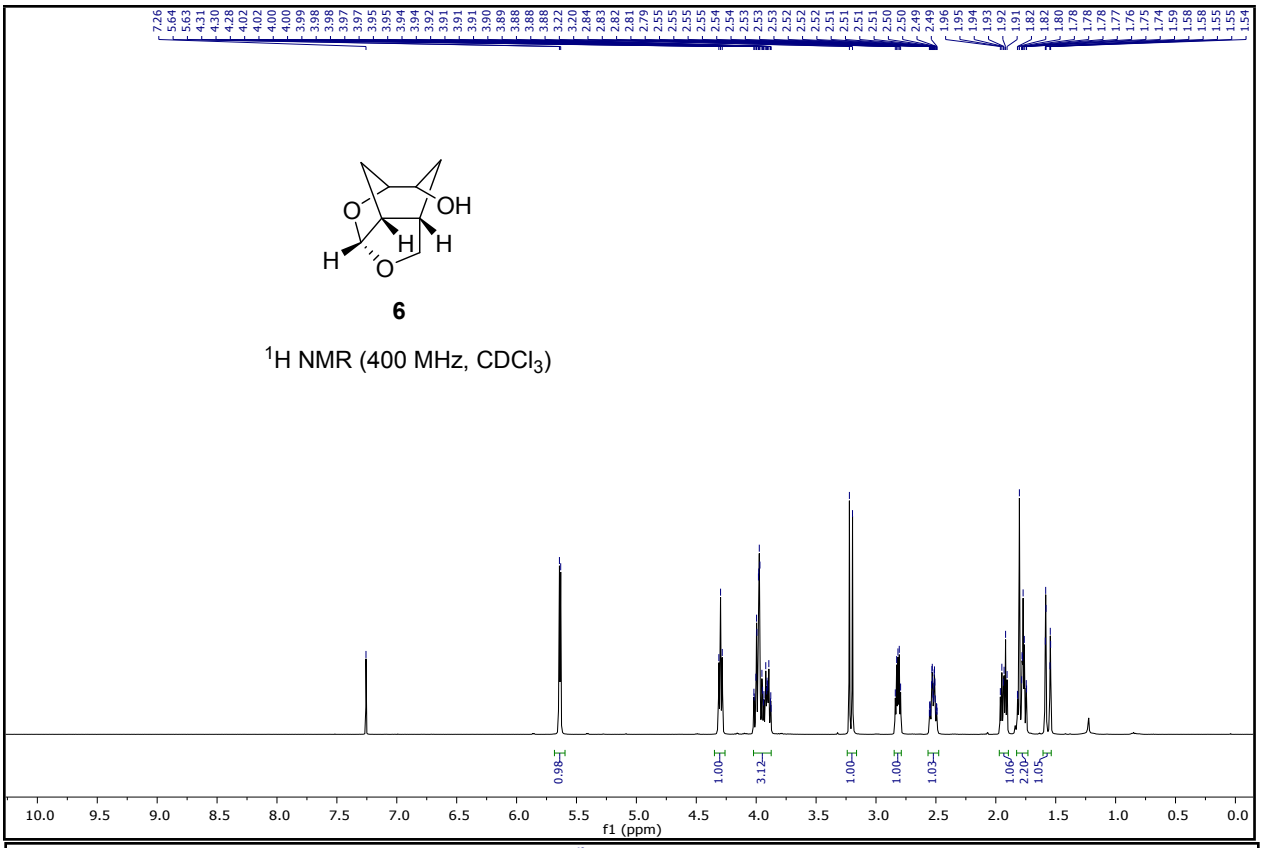


13

^{13}C NMR (100 MHz,CD_3OD)

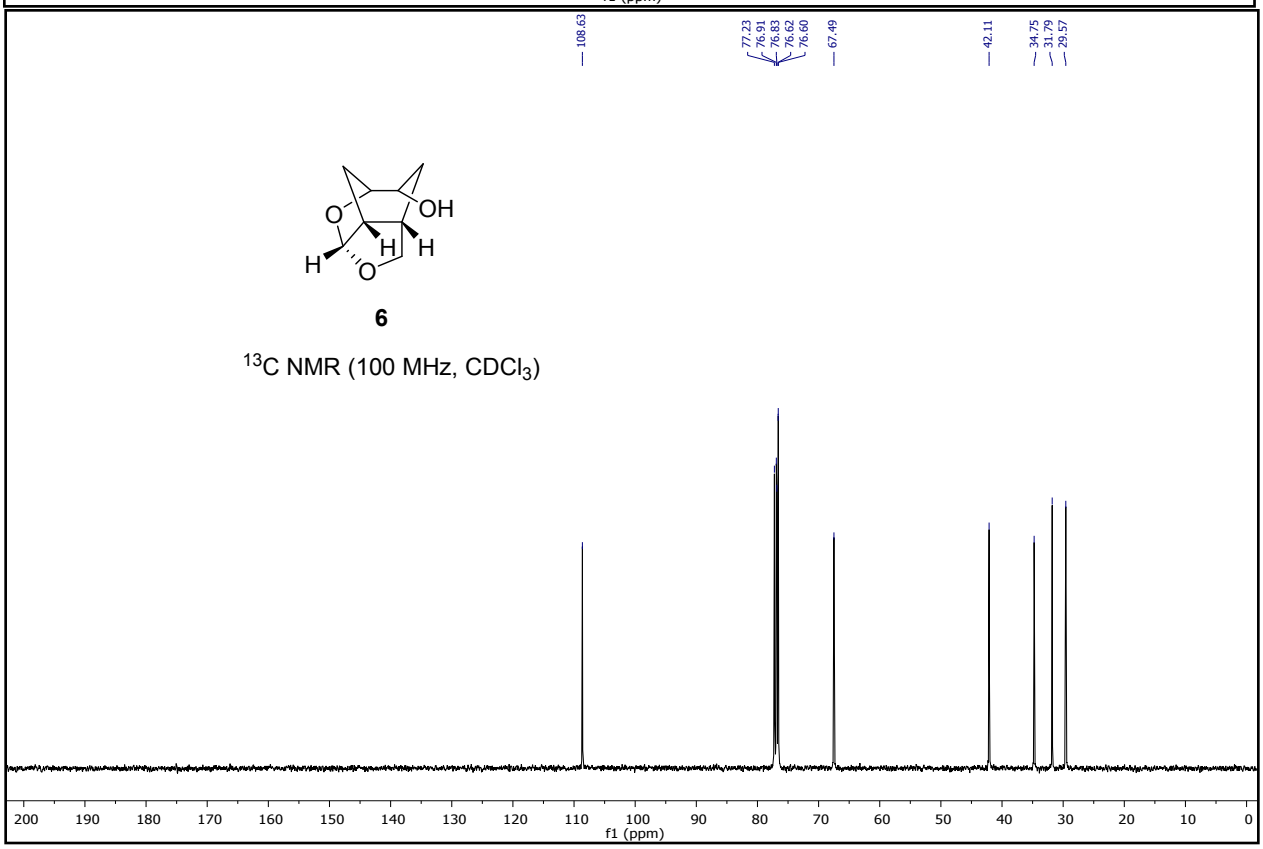

OMe
OMe
HO
HO
OH
14
1H NMR (400 MHz, CD3OD)
f1 (ppm)


OMe
OMe
HO
HO
OH
14
13C NMR (100 MHz, CD3OD)
f1 (ppm)



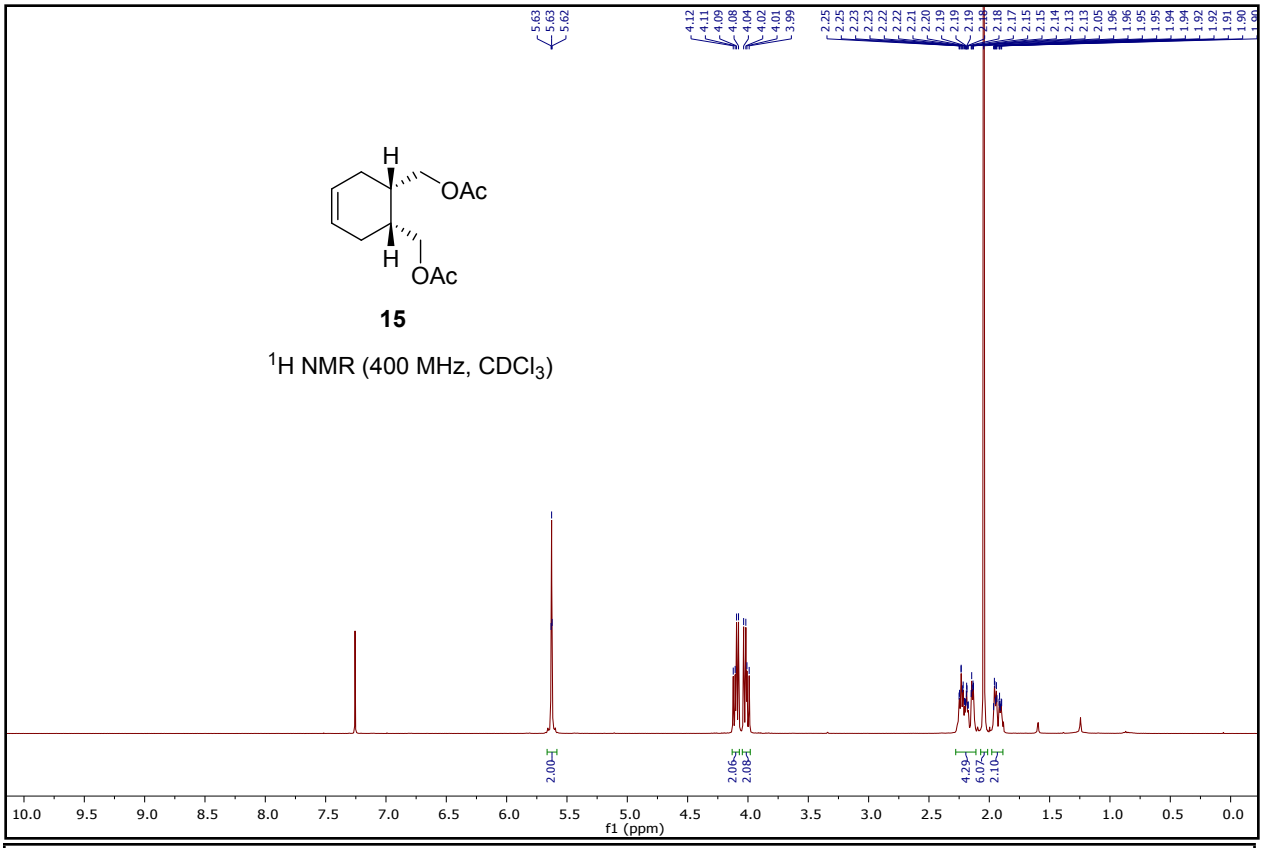
6

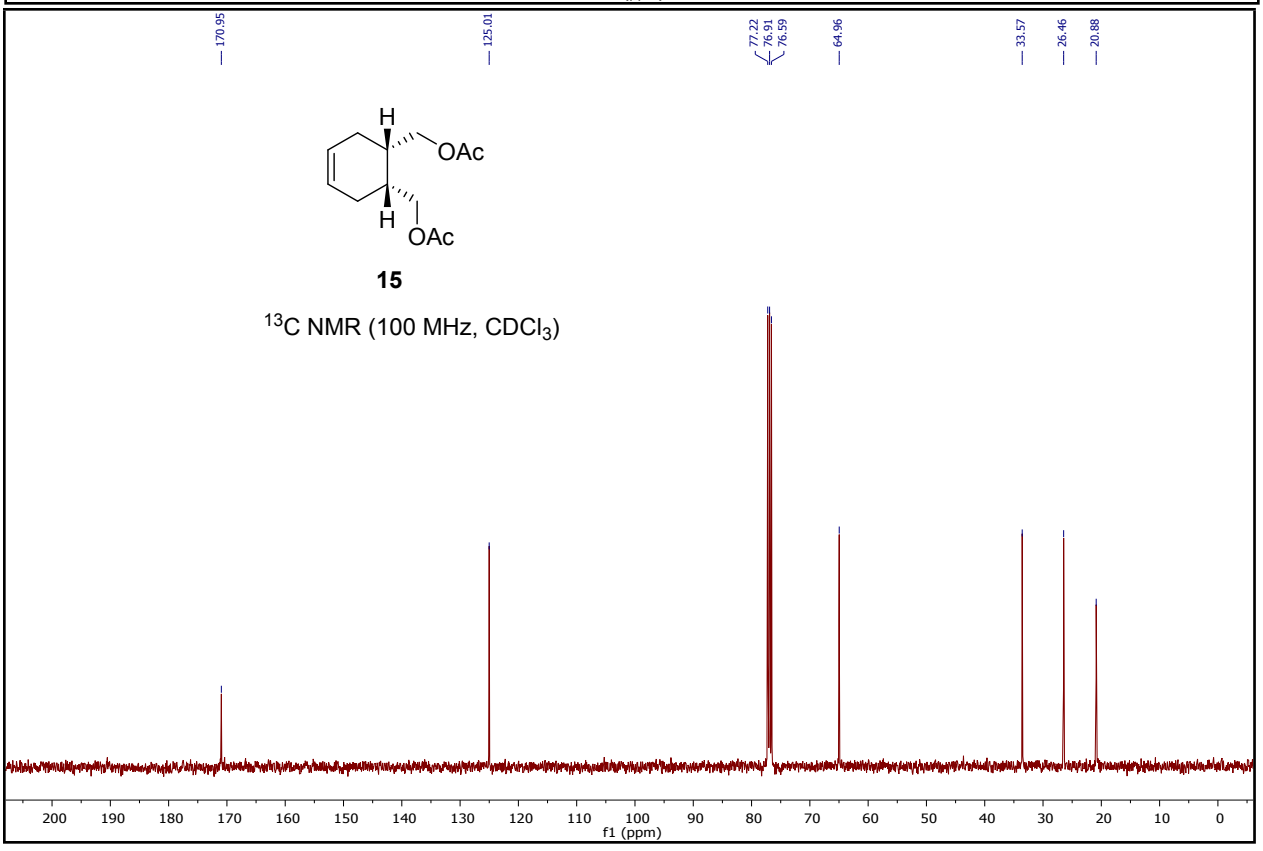
^{1}H NMR (400 MHz, $CDCl_3$)

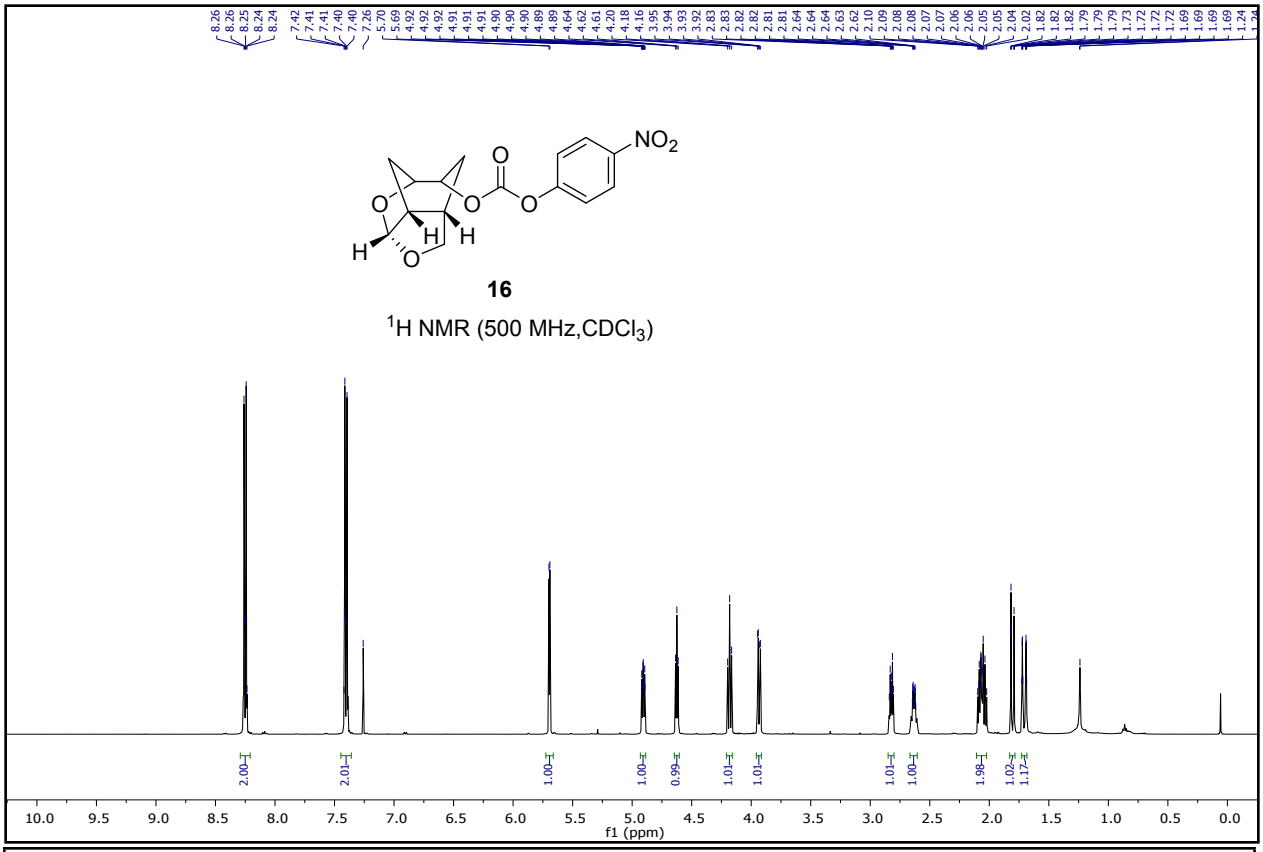


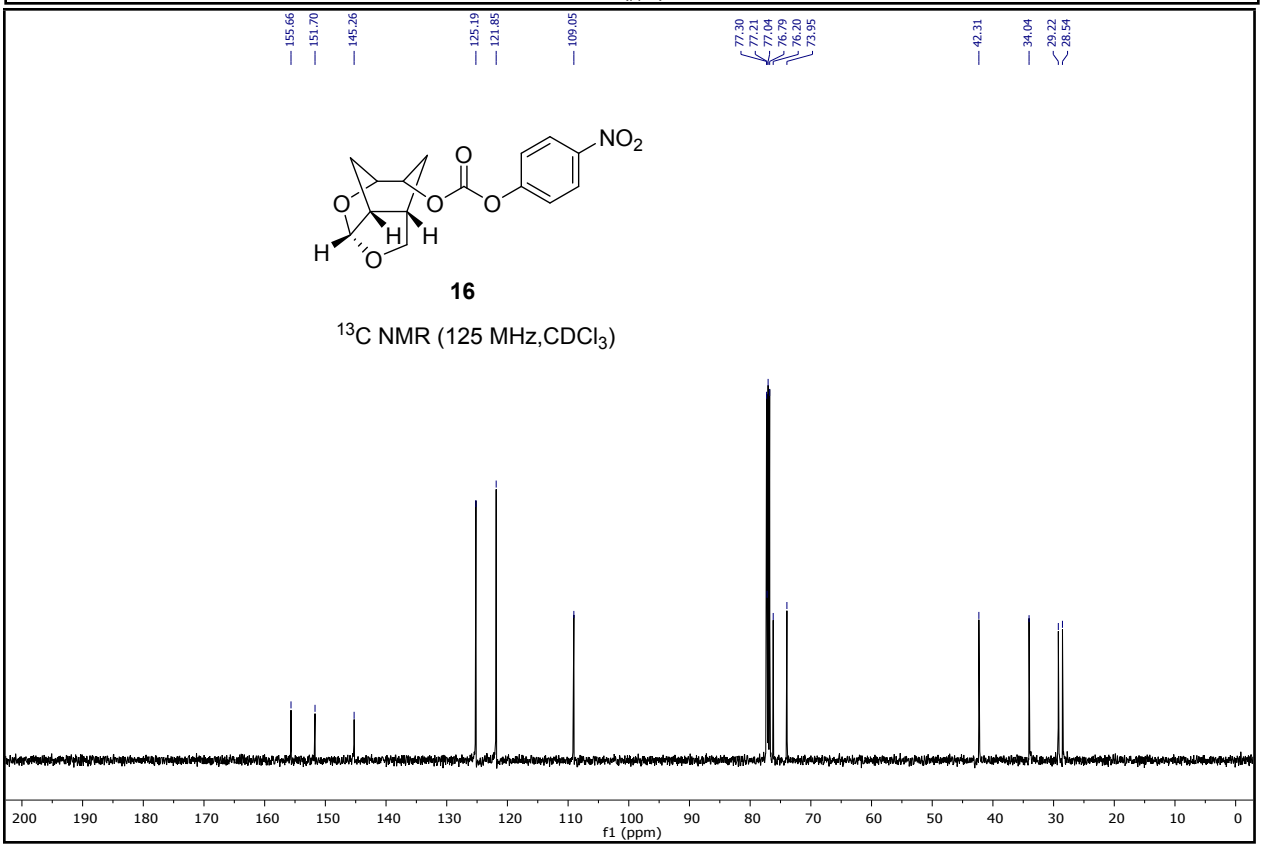
6

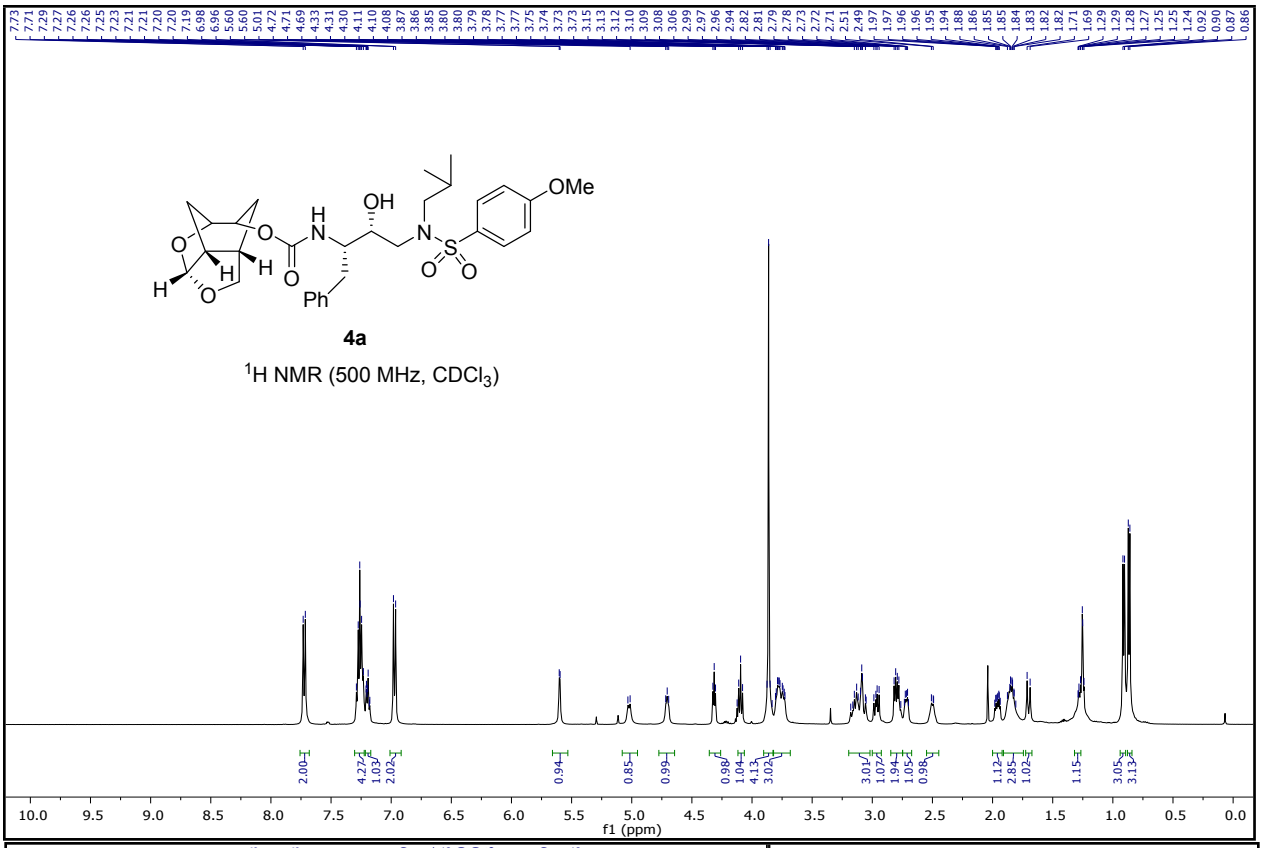
^{13}C NMR (100 MHz, $CDCl_3$)


H
OAc
H
OAc
15
1H NMR (400 MHz, CDCl3)
f1 (ppm)


H
OAc
H
OAc
15
13C NMR (100 MHz, CDCl3)
f1 (ppm)

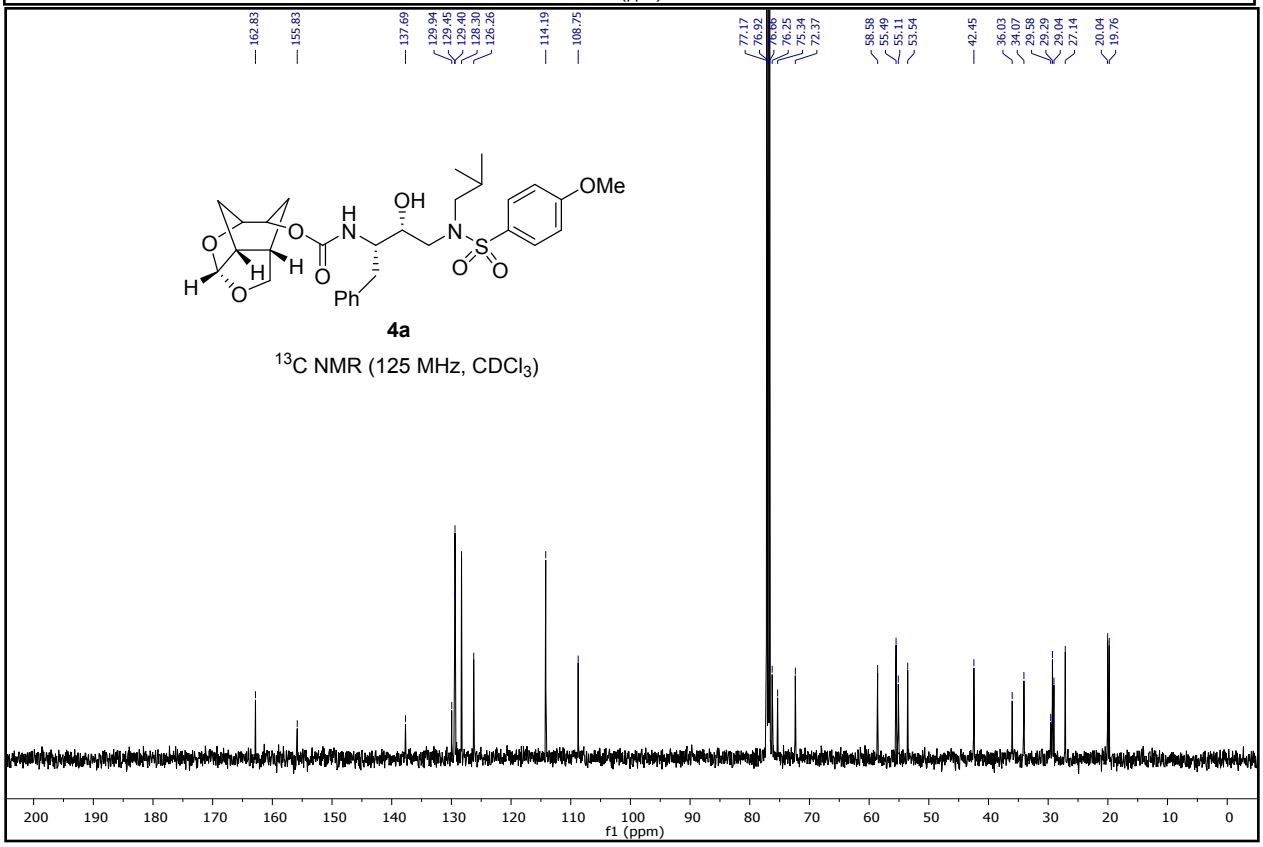

NO2
O
O
O
O
H
H
H
O
16
1H NMR (500 MHz,CDCl3)
f1 (ppm)


NO2
O
O
O
O
H
H
H
O
16
13C NMR (125 MHz,CDCl3)
f1 (ppm)



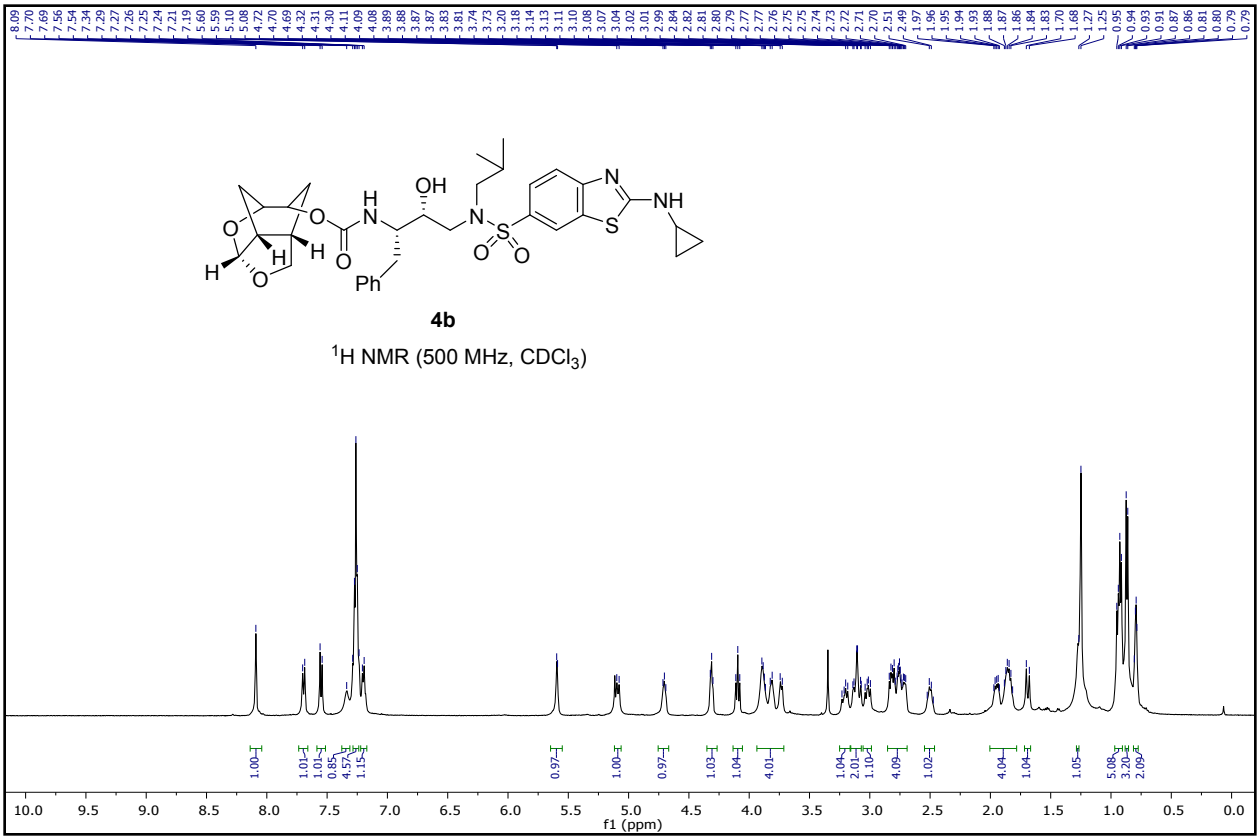
4a

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



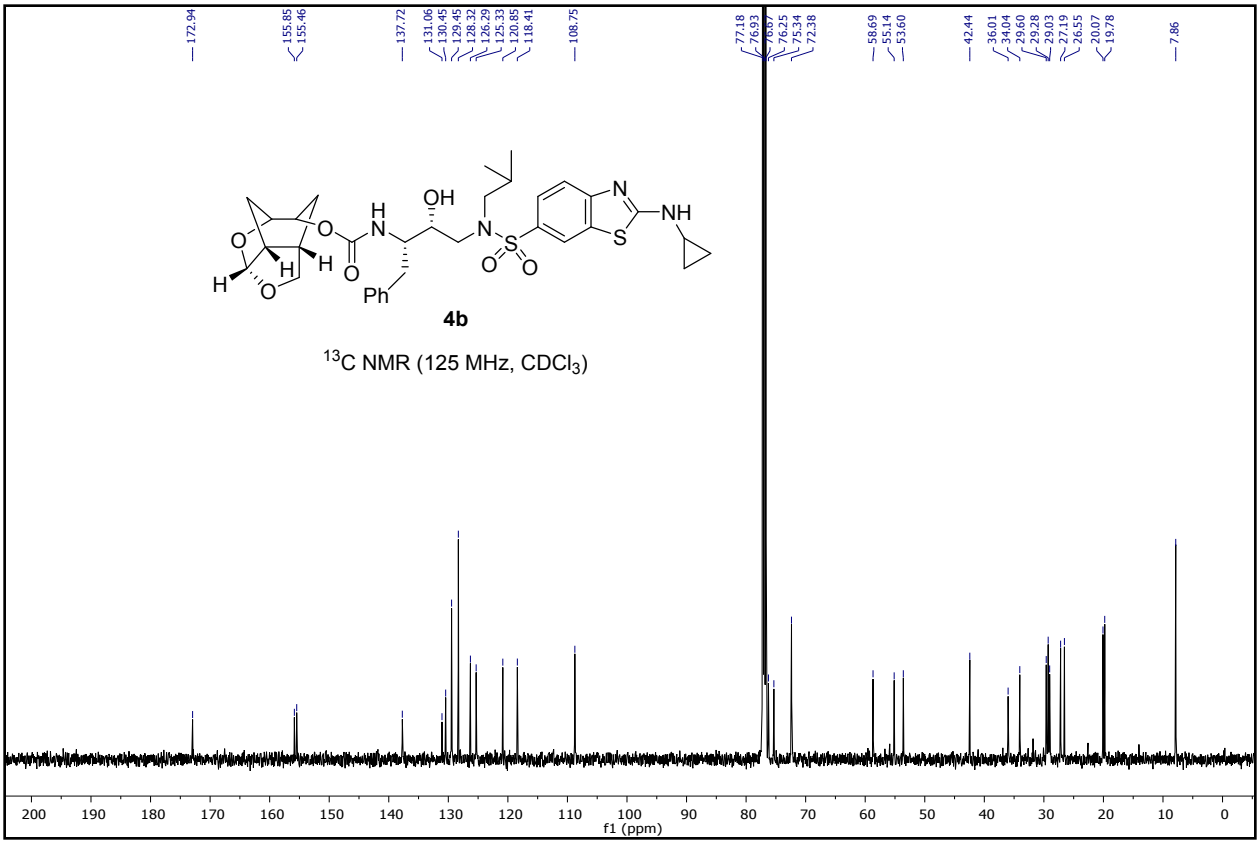
4a

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



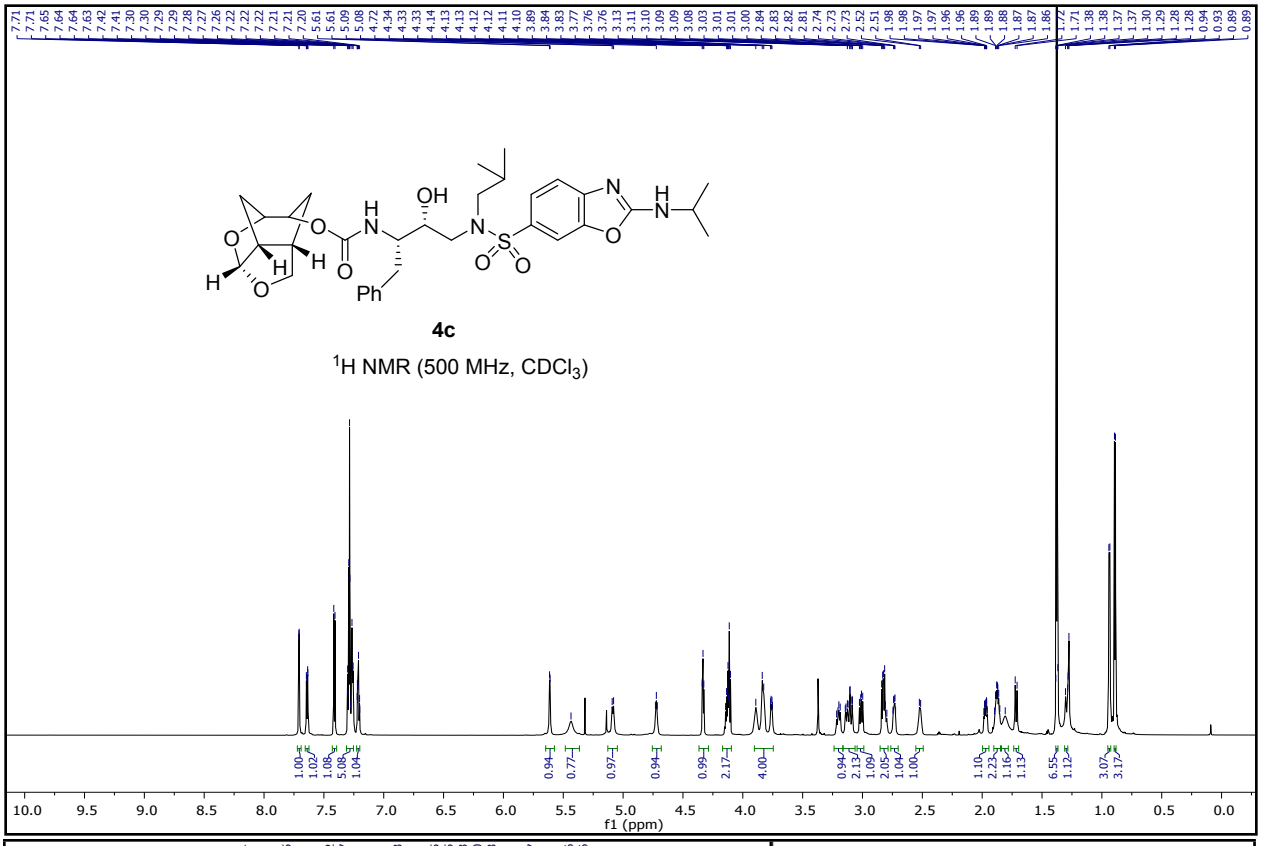
4b

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



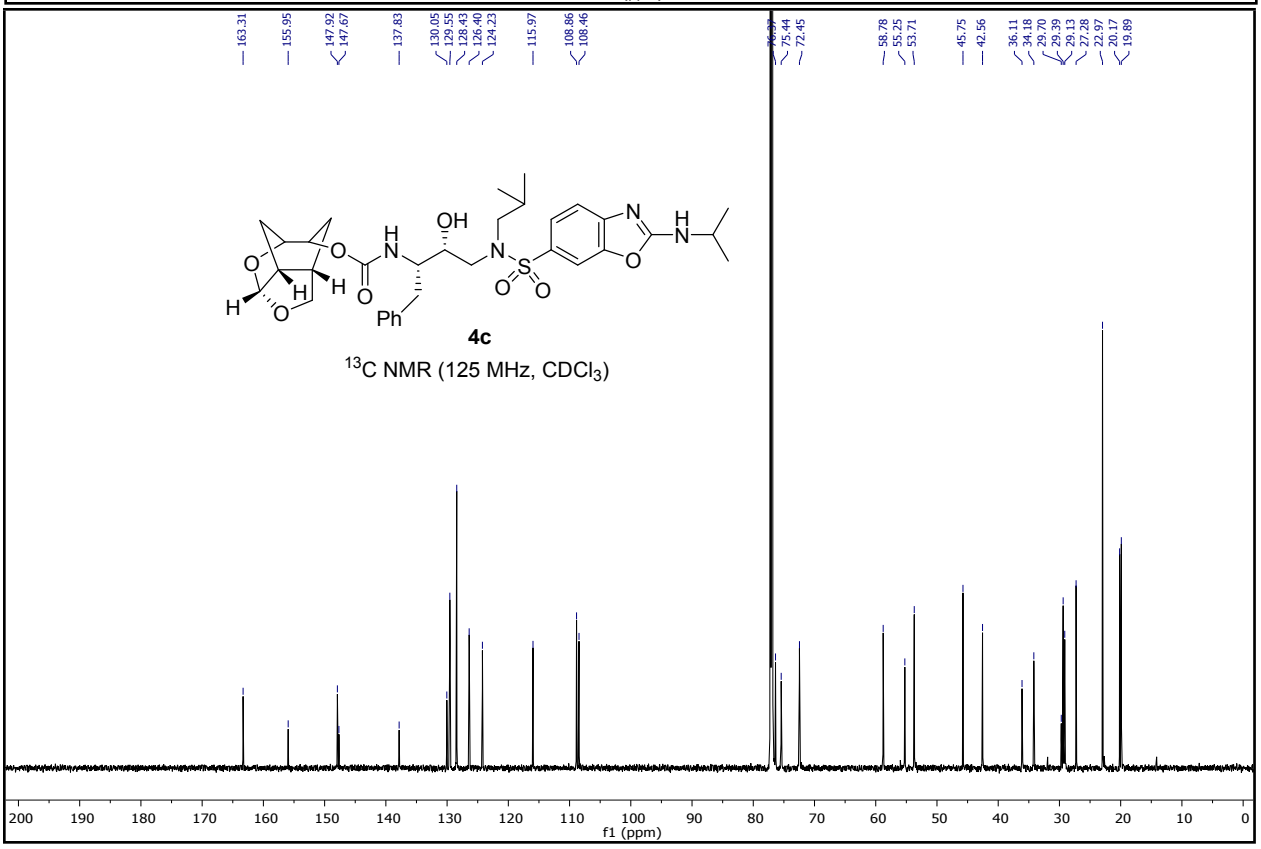
4b

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



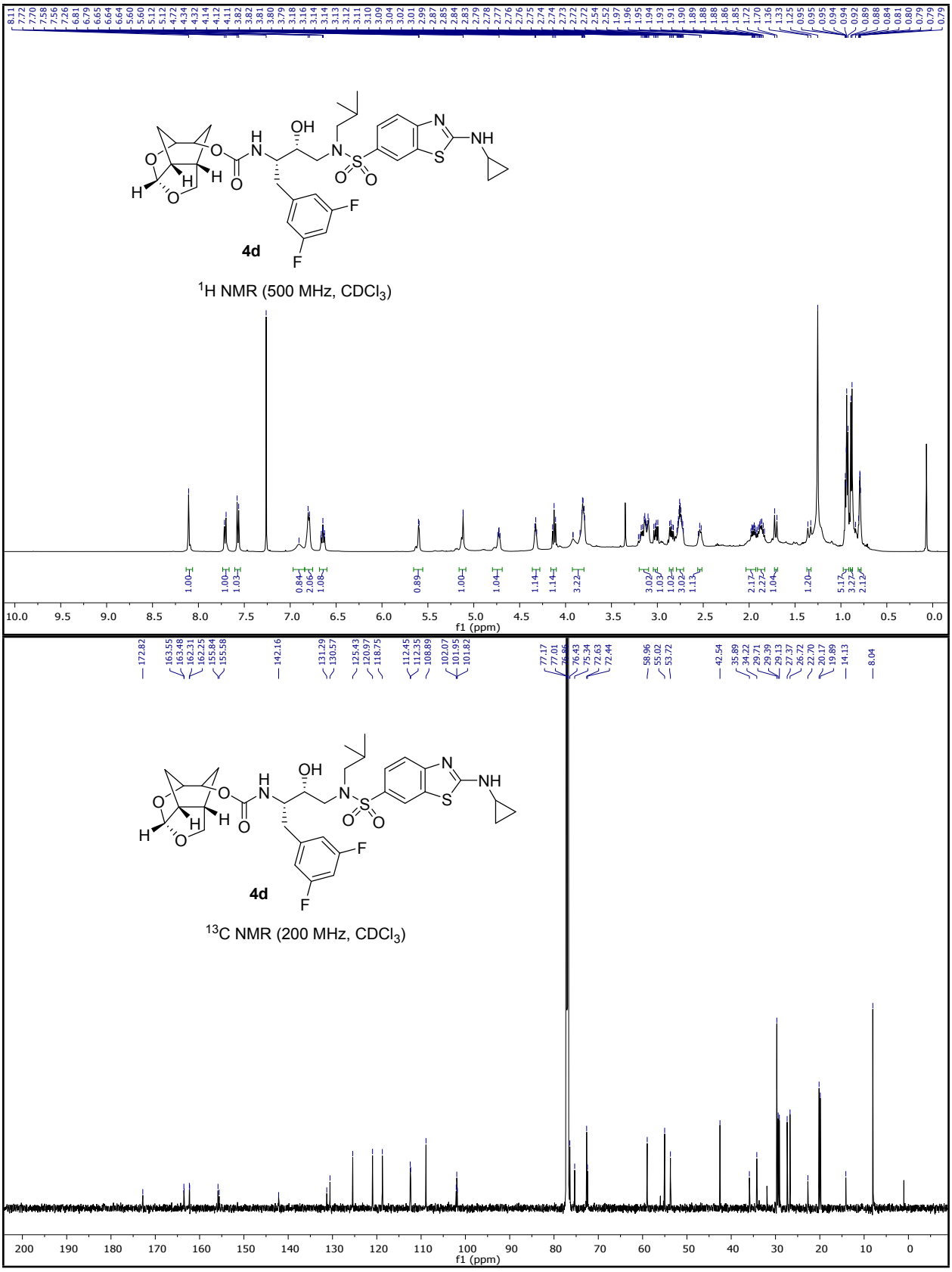
4c

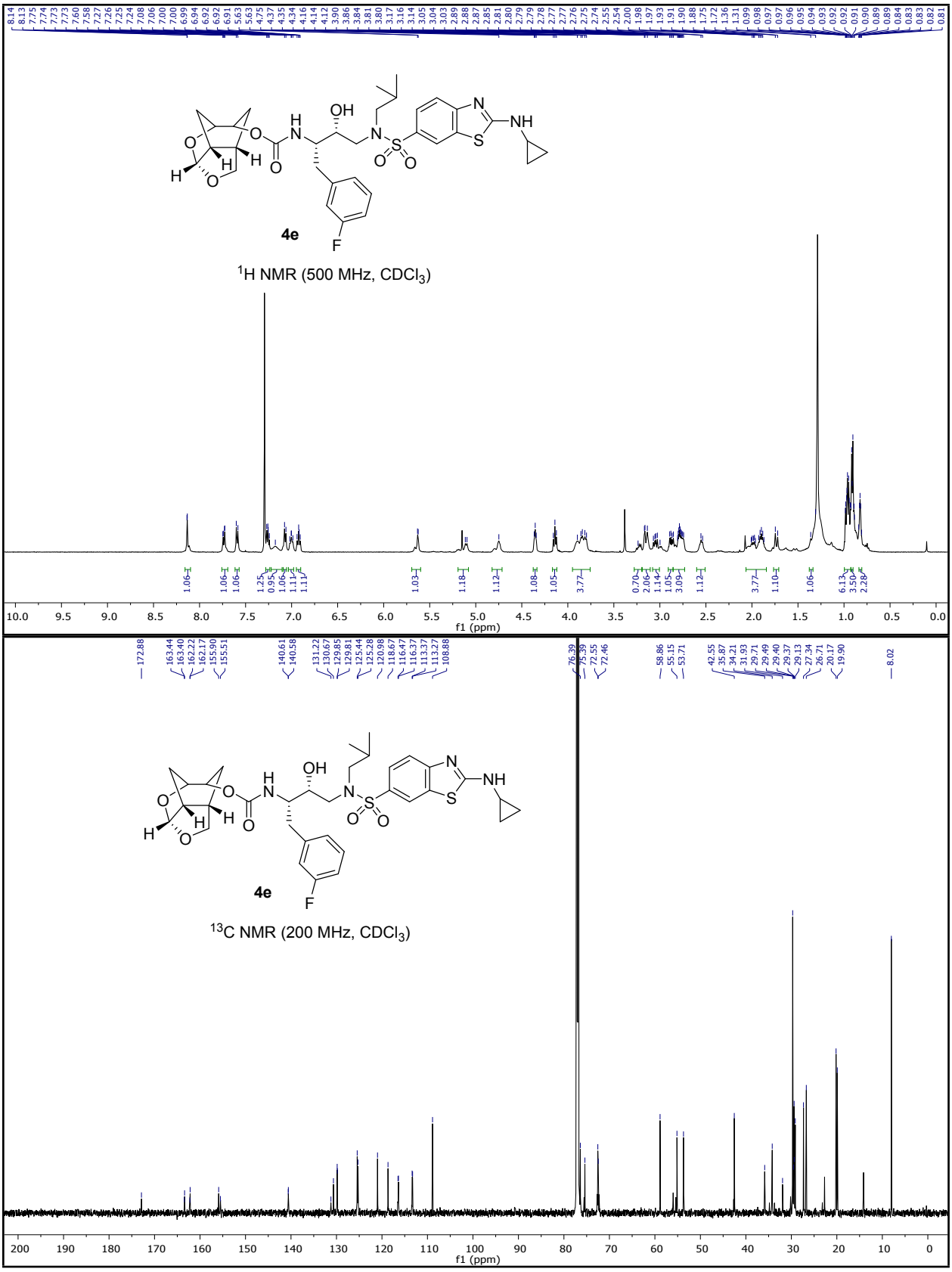
^{1}H NMR (500 MHz, $CDCl_3$)

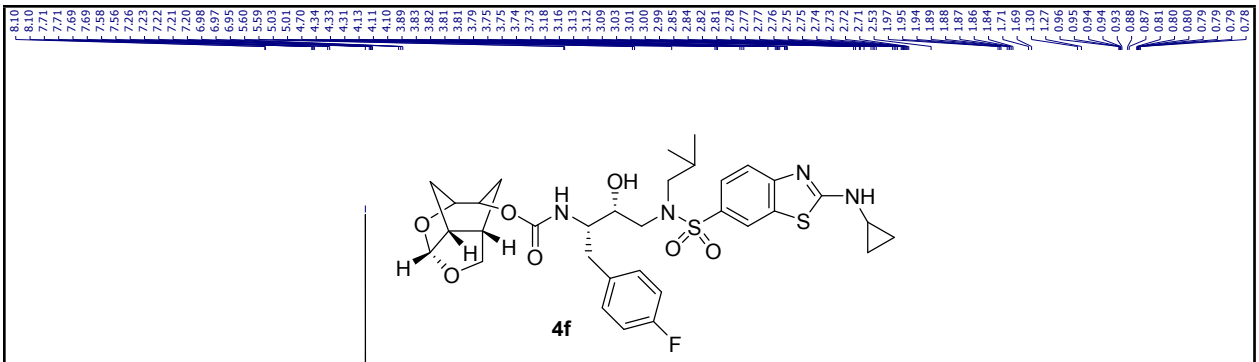


4c

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

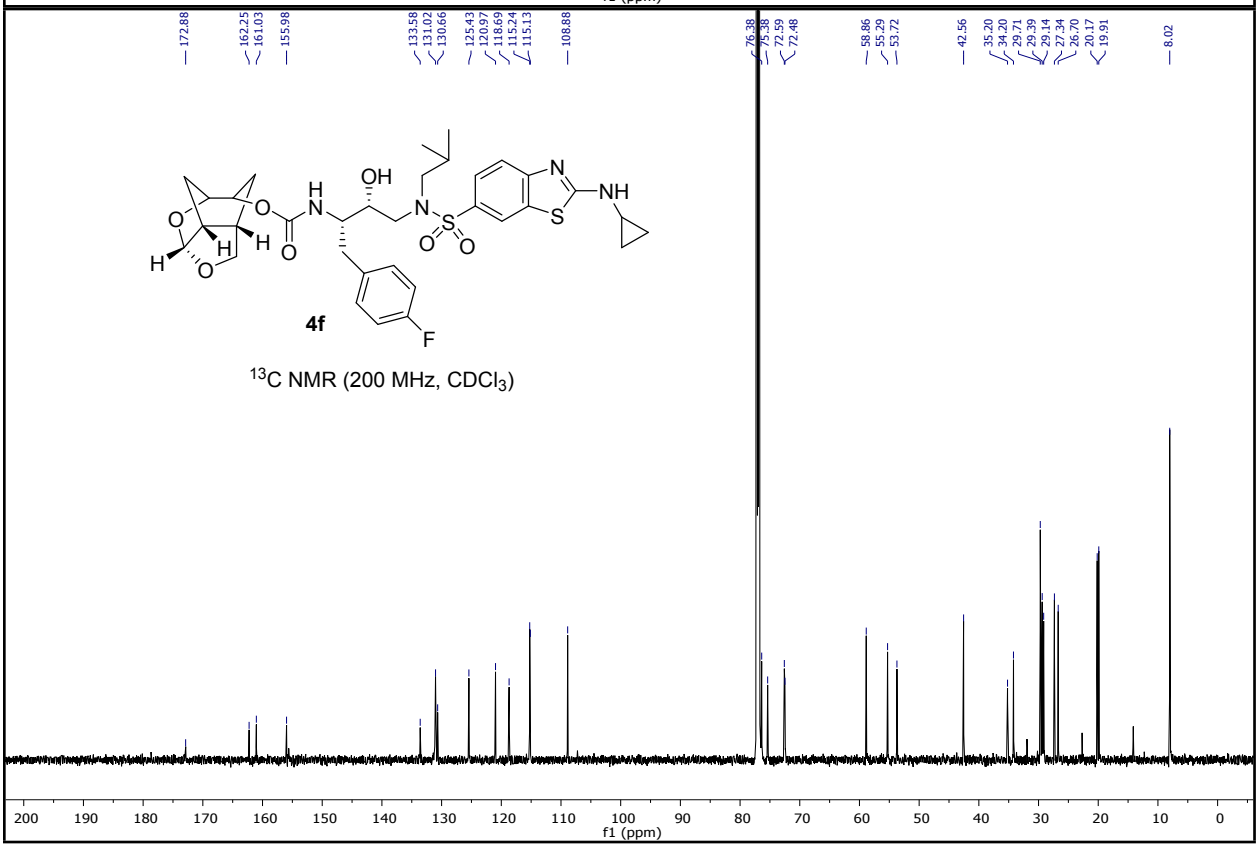

4d
1H NMR (500 MHz, CDCl3)
f1 (ppm)
4d
13C NMR (200 MHz, CDCl3)
f1 (ppm)


4e
^{1}H NMR (500 MHz, CDCl$_3$)
f1 (ppm)
4e
^{13}C NMR (200 MHz, CDCl$_3$)
f1 (ppm)



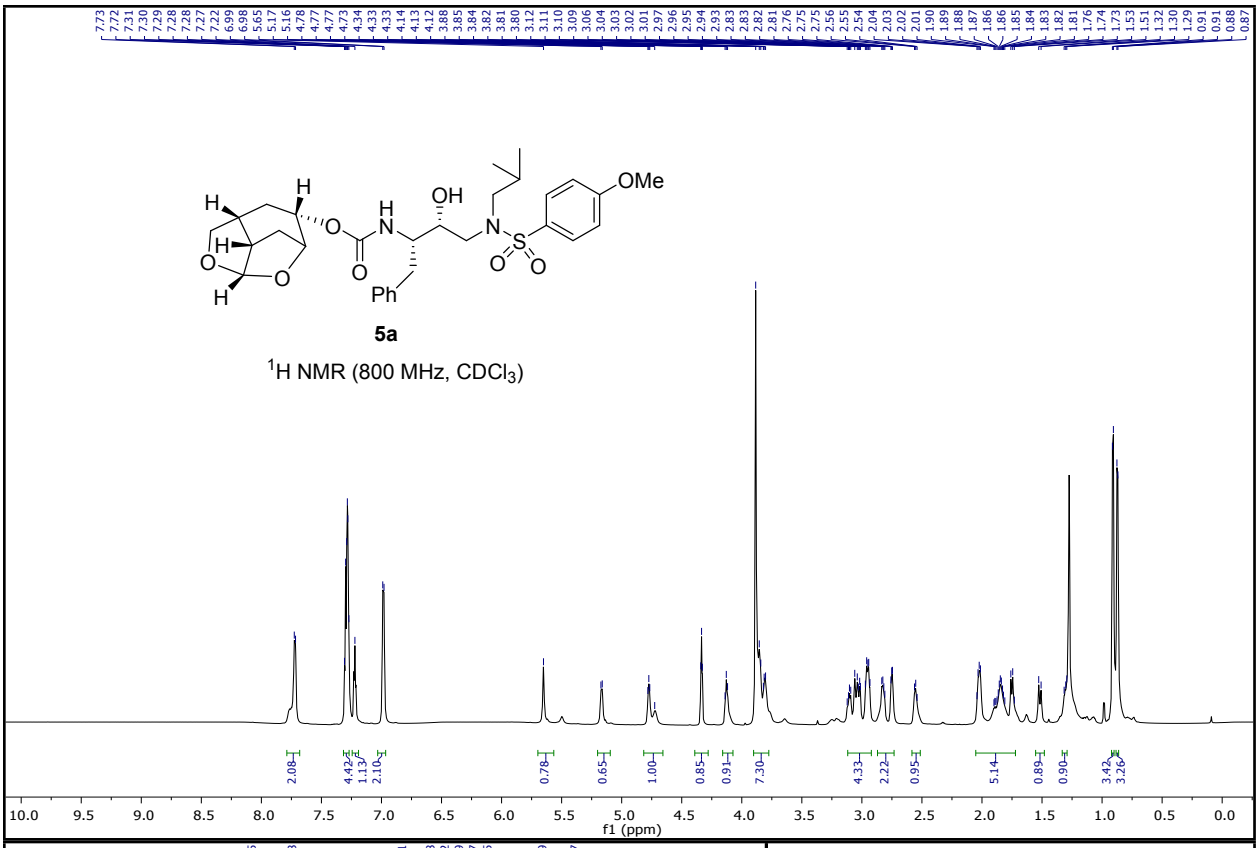
4f

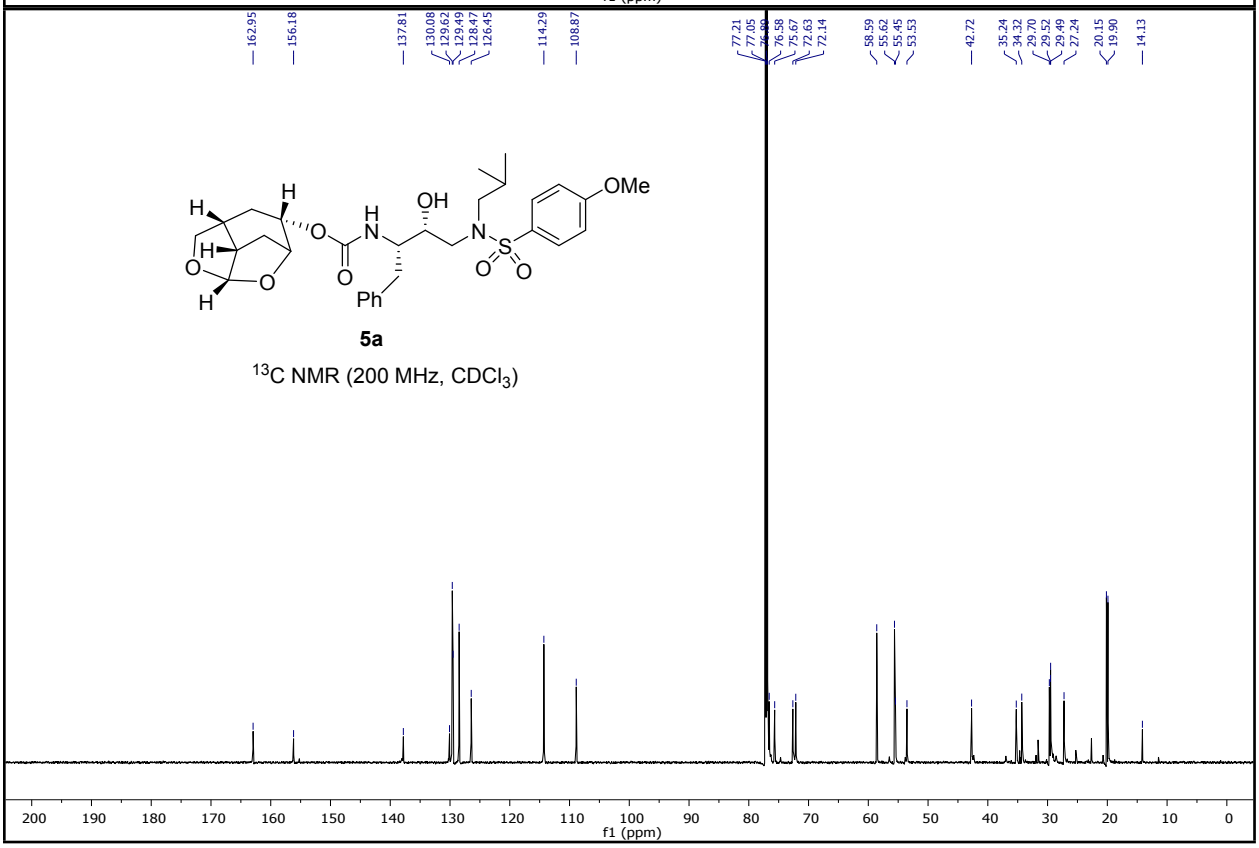
^{1}H NMR (500 MHz, $CDCl_3$)

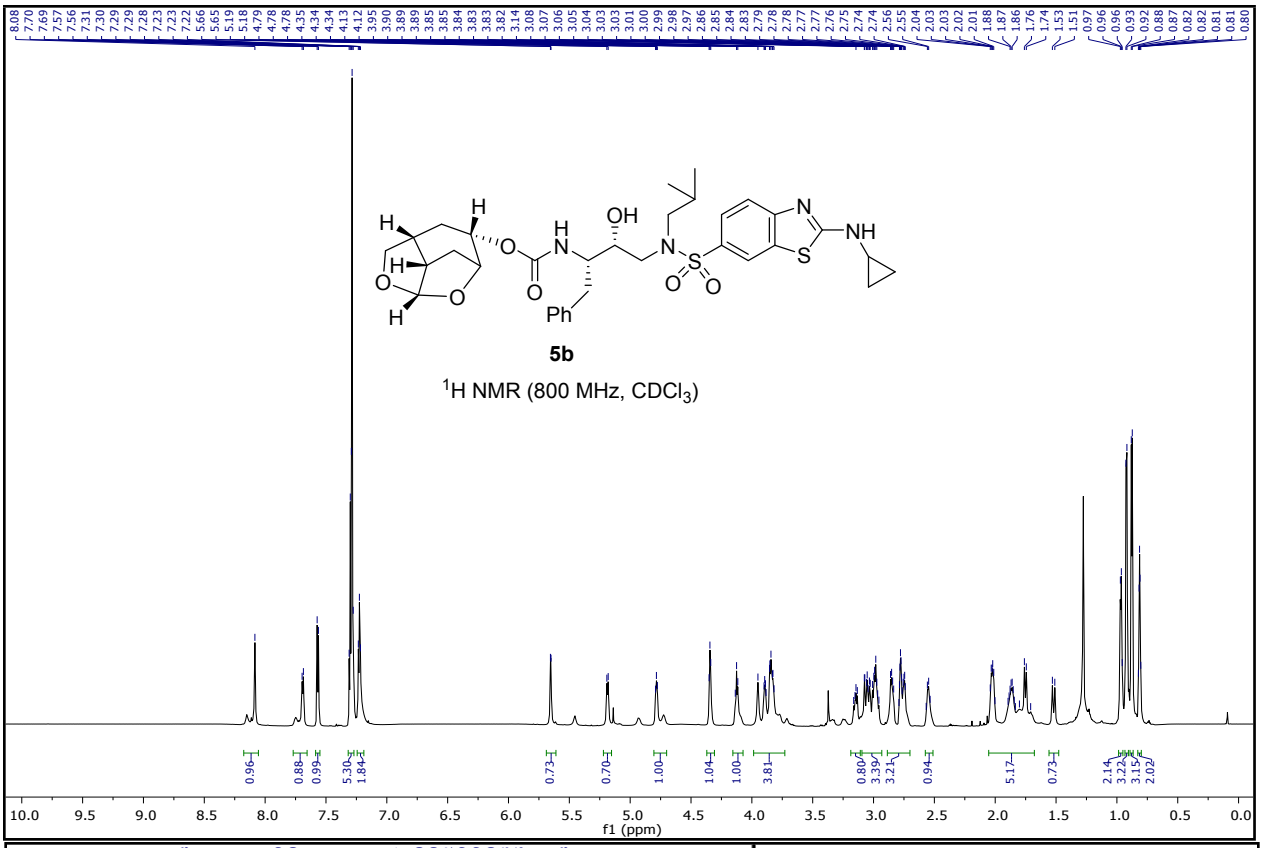


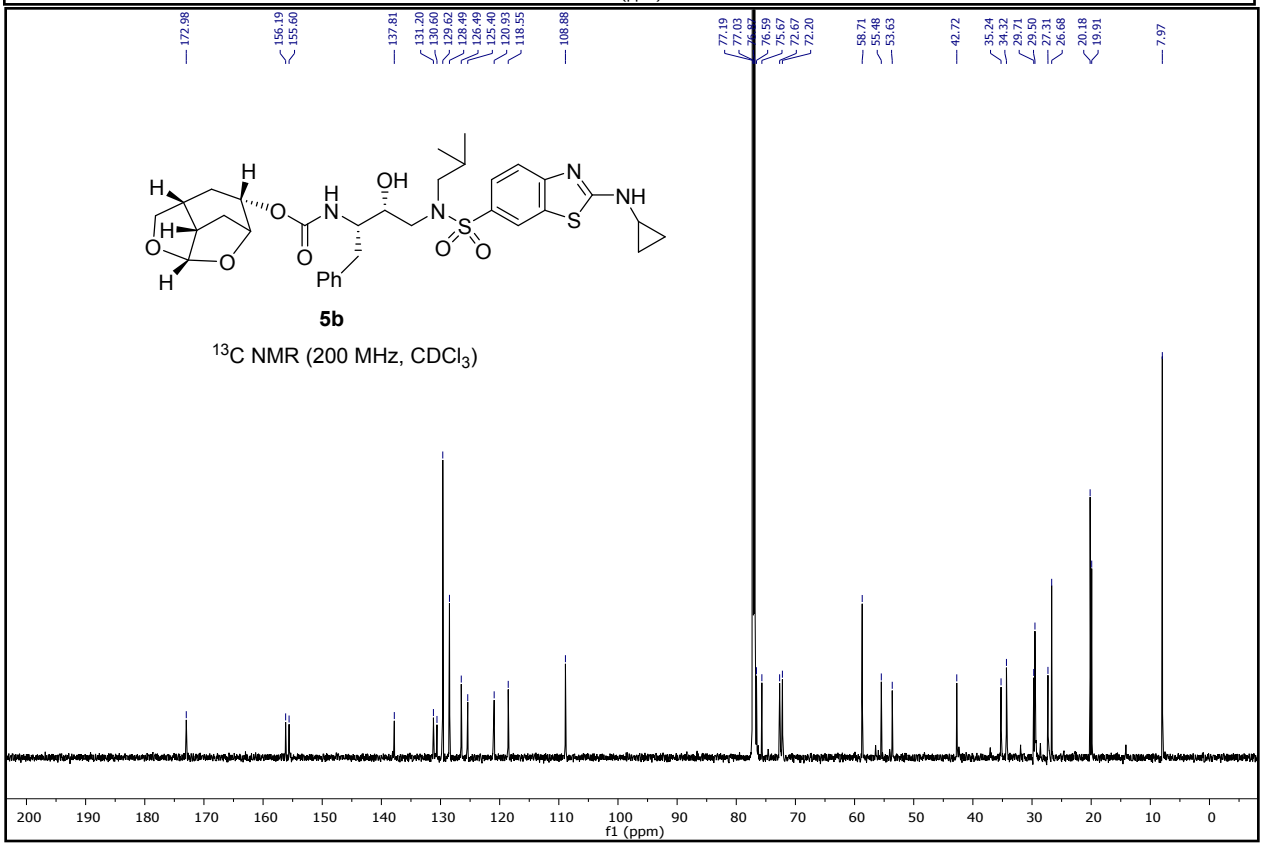
4f

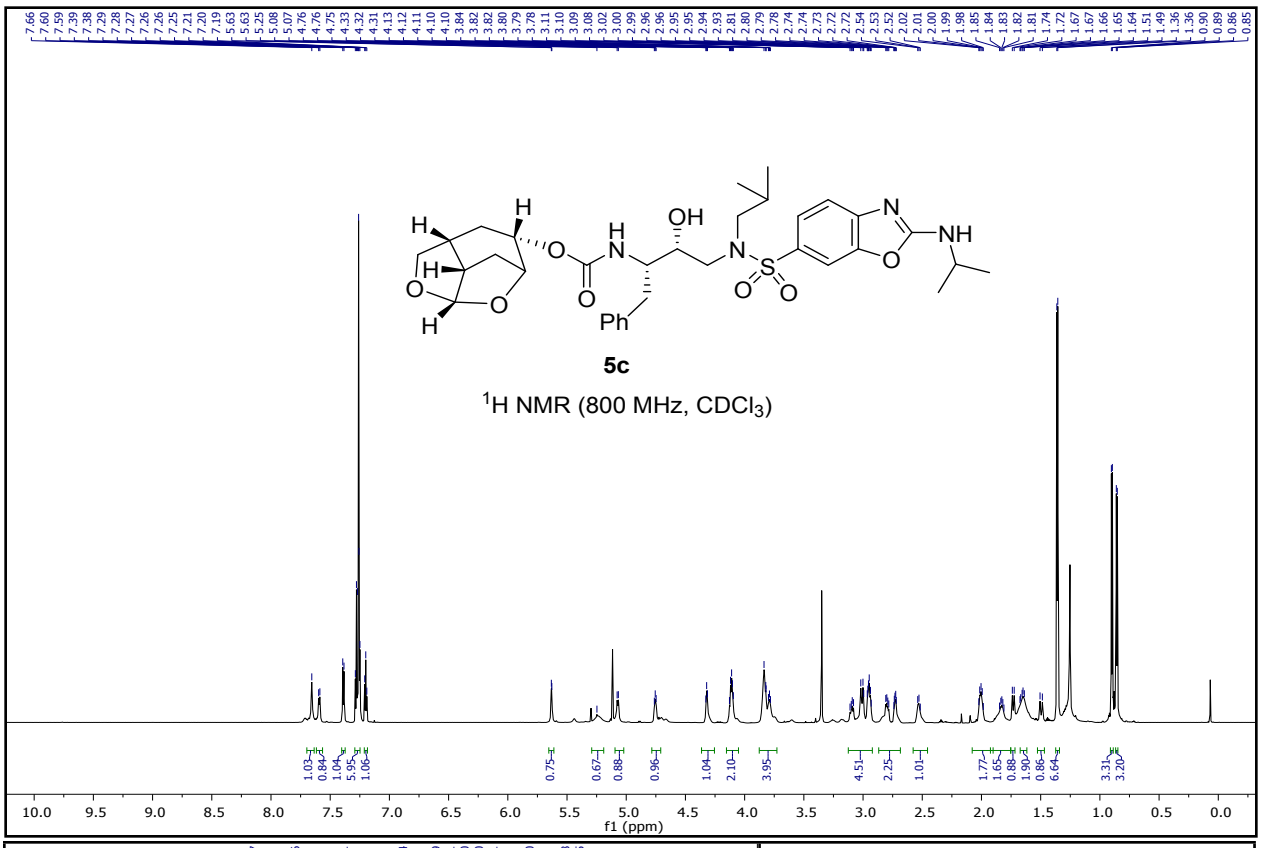
^{13}C NMR (200 MHz, $CDCl_3$)

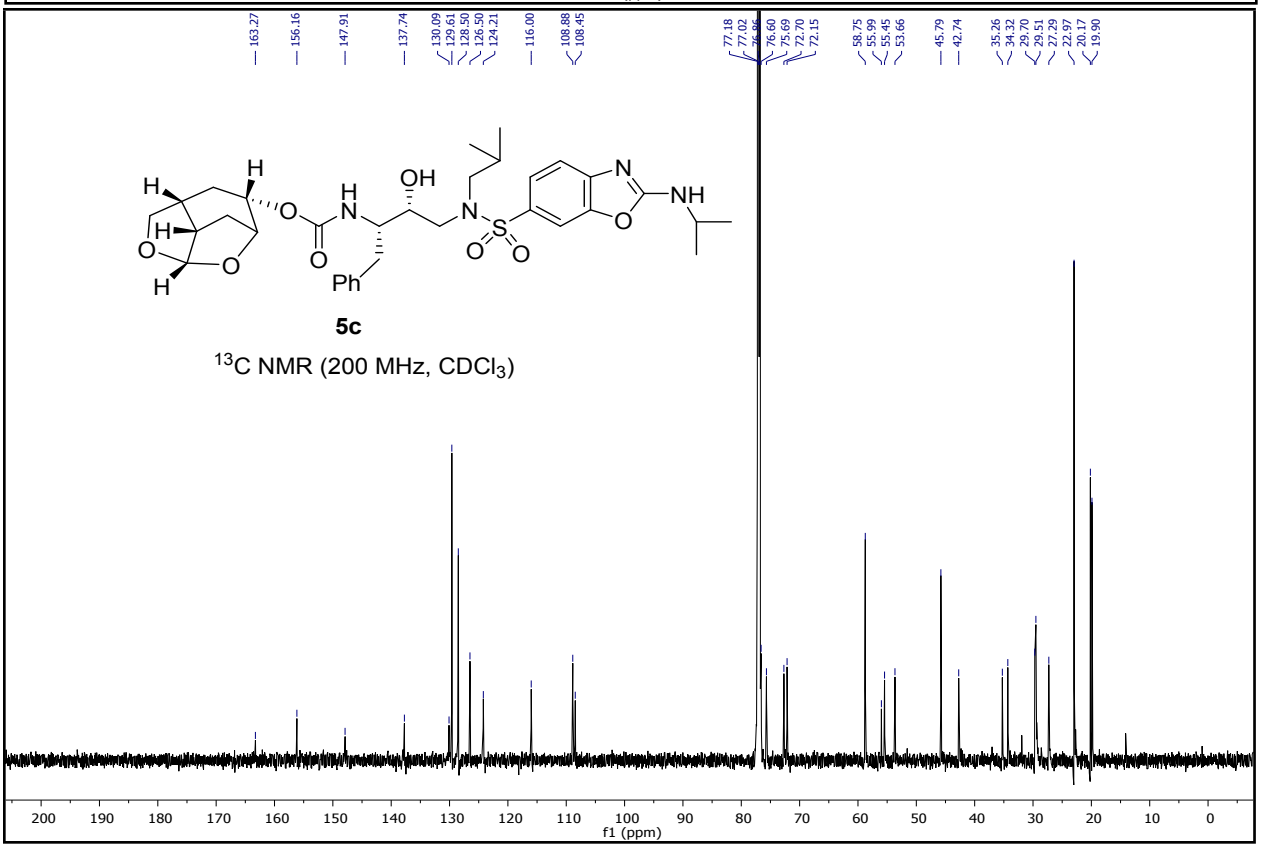

OMe
OH
Ph
5a
1H NMR (800 MHz, CDCl3)
f1 (ppm)

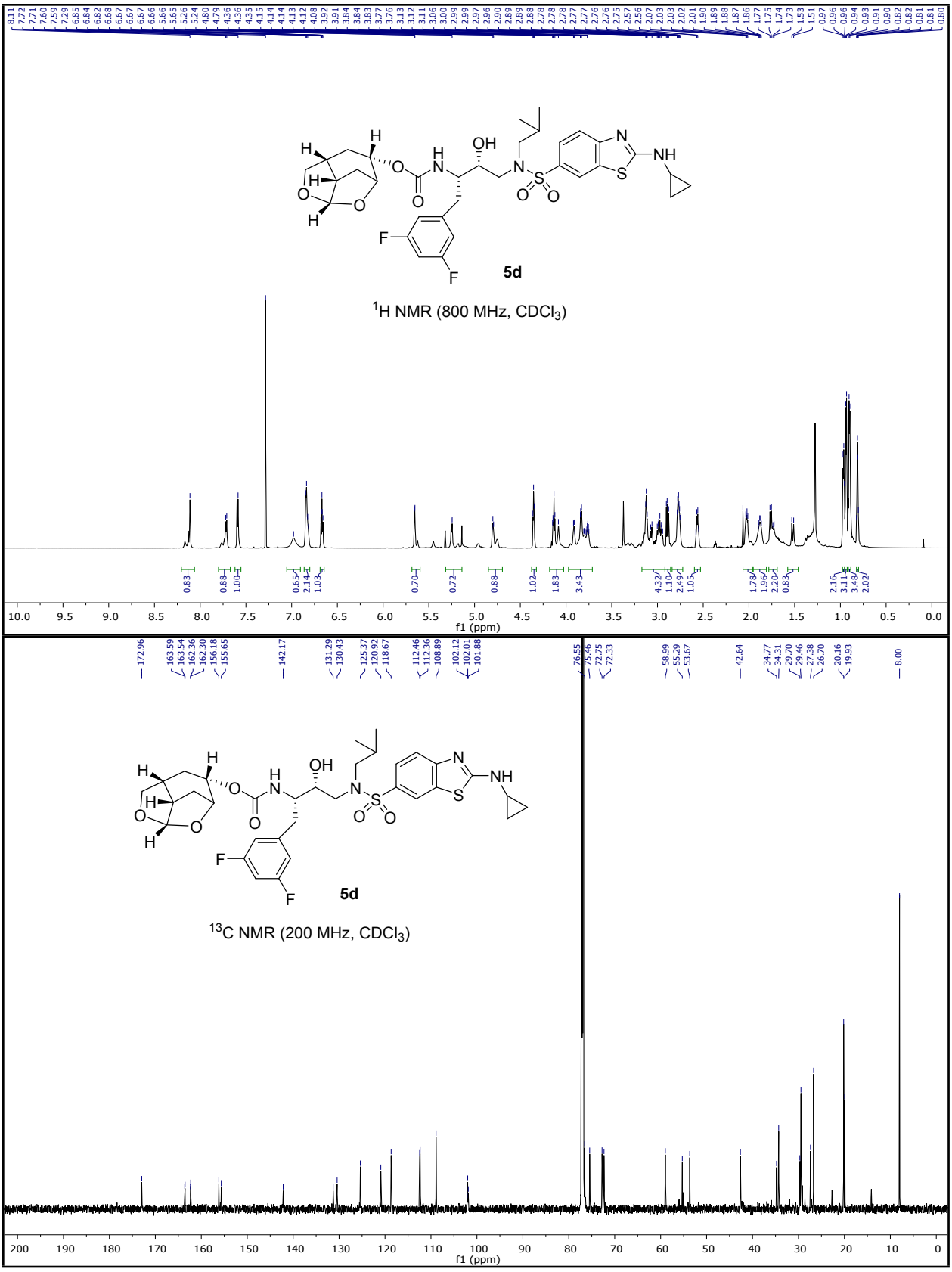

OMe
OH
Ph
5a
13C NMR (200 MHz, CDCl3)
f1 (ppm)

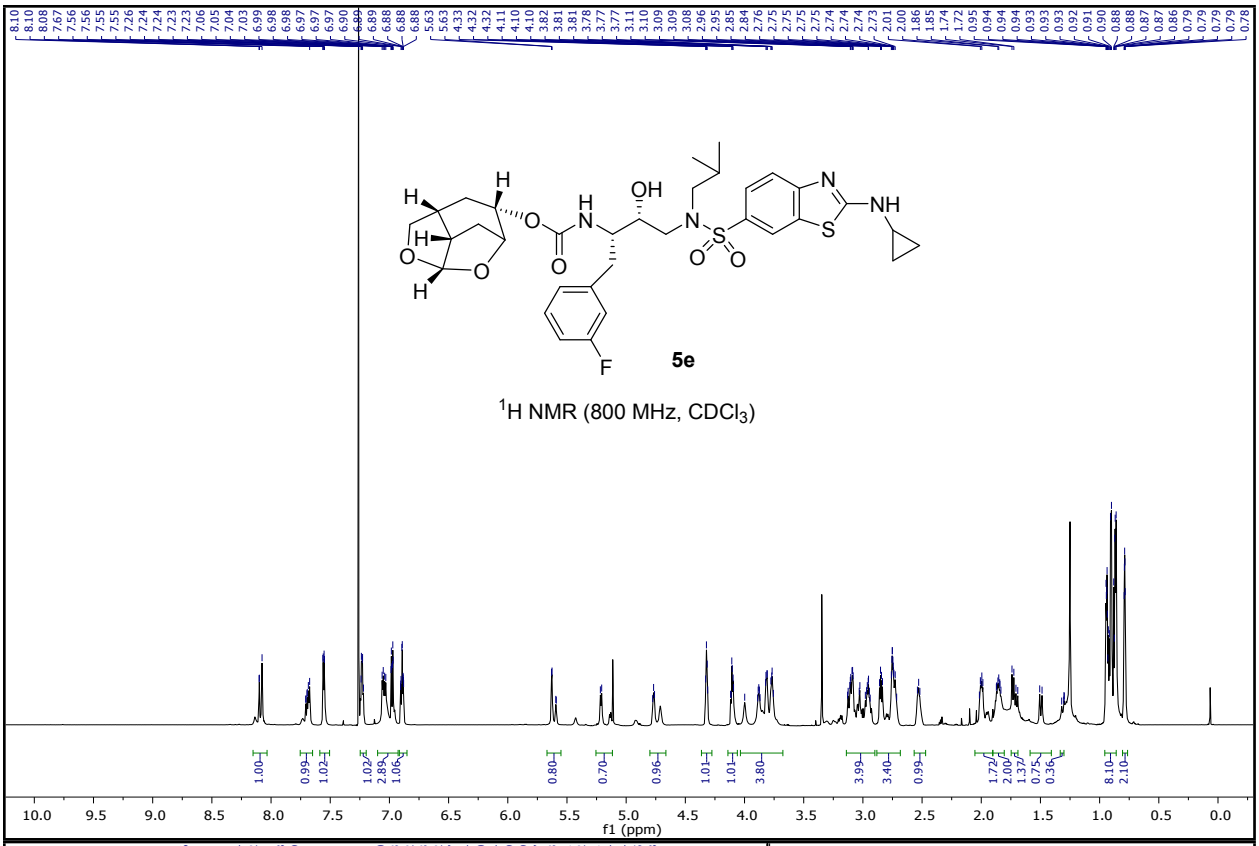

5b
1H NMR (800 MHz, CDCl3)

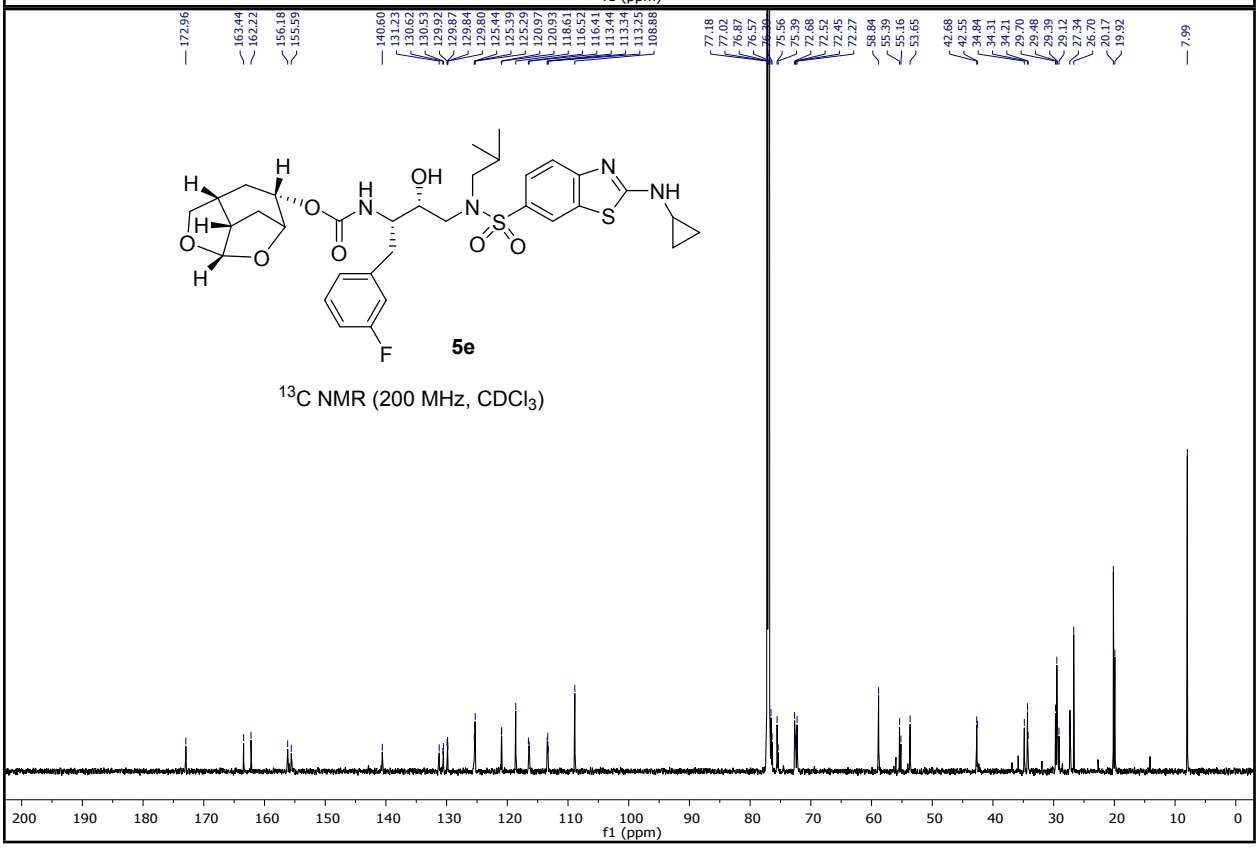

5b
13C NMR (200 MHz, CDCl3)

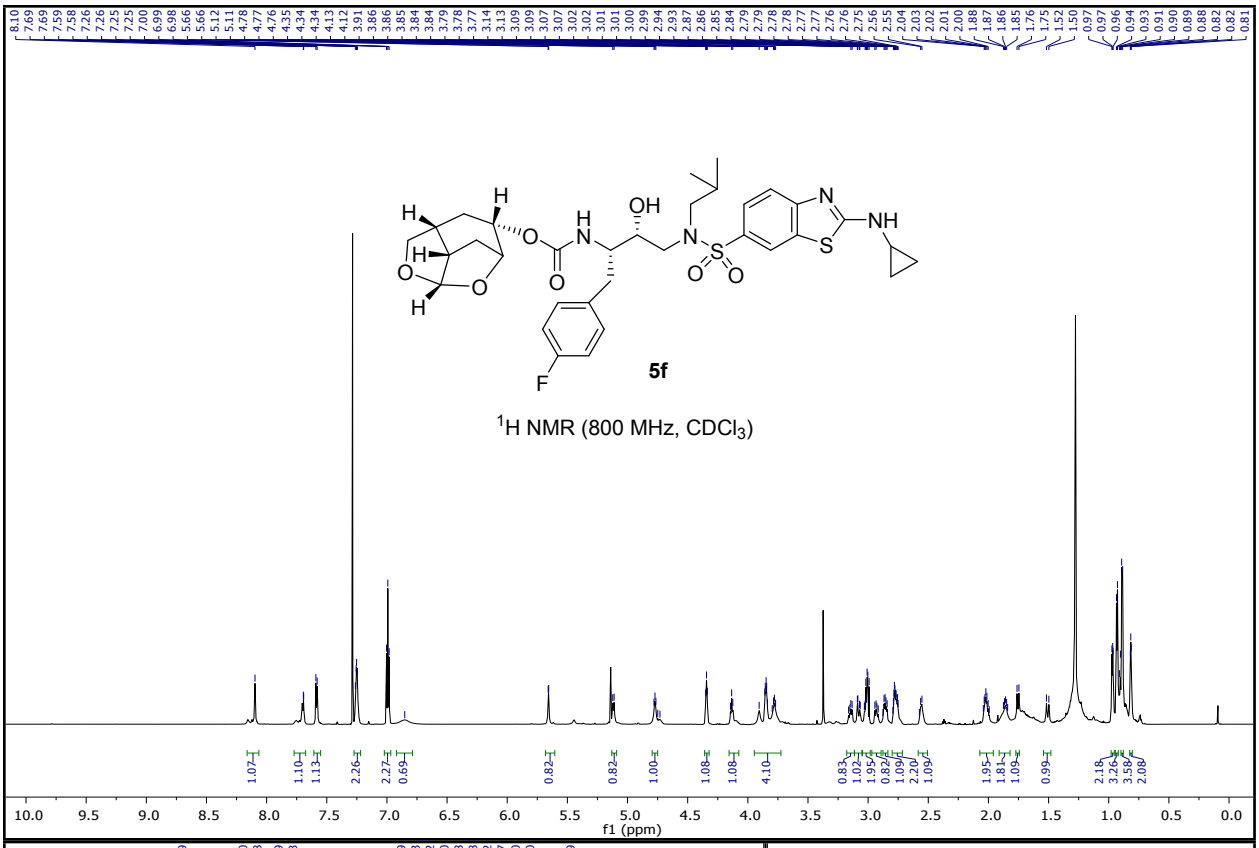

5c
1H NMR (800 MHz, CDCl3)
f1 (ppm)

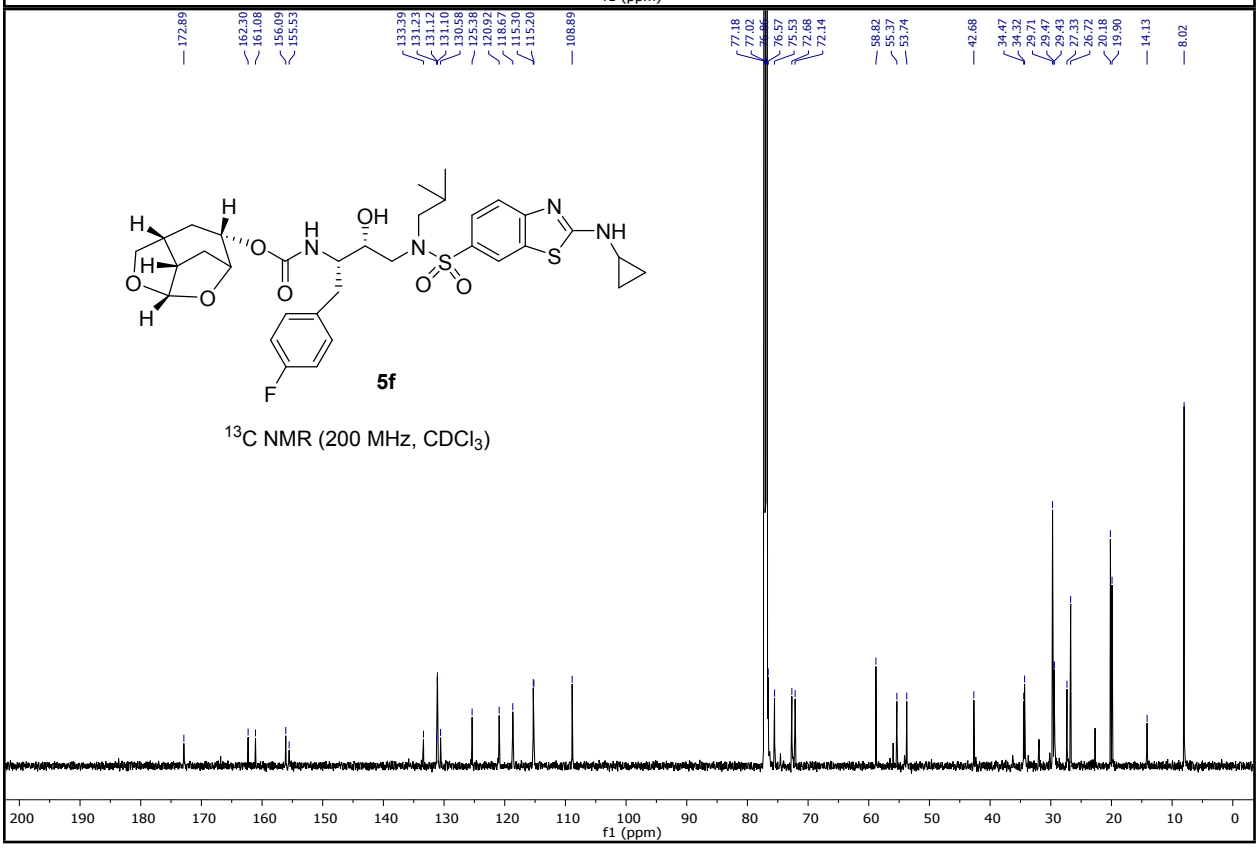

5c
13C NMR (200 MHz, CDCl3)
f1 (ppm)


5d
1H NMR (800 MHz, CDCl3)
f1 (ppm)
5d
13C NMR (200 MHz, CDCl3)
f1 (ppm)


5e
1H NMR (800 MHz, CDCl3)
f1 (ppm)


5e
13C NMR (200 MHz, CDCl3)
f1 (ppm)


5f
1H NMR (800 MHz, CDCl3)
f1 (ppm)


5f
13C NMR (200 MHz, CDCl3)
f1 (ppm)