Supporting information

Controlling an organic synthesis robot with machine learning to search for new reactivity

Jarosław M. Granda, Liva Donina, Vincenza Dragone, De-Liang Long, and Leroy Cronin*

School of Chemistry, University of Glasgow, Glasgow G12 8QQ (UK)
Email: Lee.Cronin@Glasgow.ac.uk

Contents:

1. Hardware specification and robot assembly guide

Table S1. Robot components and their specification.

32 x Adapter, Luer (Male) to 1/4"-28 Flat Bottom (Female), ETFE/Polypropylene

P-675 (Kinesis)

1 x Block Connector, 3-Way, 1/4"-28 (Flat Bottom), PEEK, Mounting holes Part No. 001057 (Kinesis)

94 x Flangeless Fitting, for 1/16" OD Tubing, 1/4"-28 Flat Bottom, PEEK™/ETFE, Natural/Blue Part No. XP-283 (Kinesis)

Needles, Sterican, Long Length 120mm

2 x Tubing, PTFE, 1/16" (1.6mm) OD x 0.8mm ID, 20M Part No. 008T16-080-20 (Kinesis)

27 x C-Series Syringe Pumps Tricontinent C3000 with 5.0 mL syringes and 4-way distribution valves

Duran® laboratory bottles, with caps capacity 100 mL, blue PP screw cap and pouring ring Sigma-Aldrich Z305170-10EA and Z305200-10EA

18 x 50-250 mL for reagents

2 x 1L for solvent and waste

7 x Suba-Seal® septa red rubber, Suba Seal, 33, neck I.D., 17.5 mm, Z124613-100EA

7 x canonical flask 25 mL for reactors and mixer

Benchtop MS: expression^{L} CMS from Advion

- **Flow rate range:** 10 µL/min to 2 mL/min APCI
- **Polarity:** Positive & Negative ion switching in single analysis
- **Polarity Switching Speed:** 50 ms
- *m/z* **Range:**; -L model ***m/z*** 10 to ***m/z*** 2,000
- **Resolution:** 0.5-0.7 ***m/z*** units (FWHM) at 1000 ***m/z*** units sec^{-1} over entire acquisition range
- **Accuracy:** ± 0.1 ***m/z*** units over entire acquisition range
- **Dynamic Range:** 4.5 orders of magnitude
- **Ion Source Parameters**
- Capillary temperature(V) 250.0
- Capillary Voltage(V) 180.0
- Source Voltage(V) 20
- Source Voltage Dynamics (V) 30
- APCI Source Temperature 350
- Corona Discharge (uA) 5.0

Benchtop NMR: Spinsolve 43 Carbon from Magritek

• Frequency: 42.5 MHz Proton, 10.8 MHz Carbon
• Resolution: 50% linewidth < 0.5 Hz (12 ppb)
• Lineshape: 0.55% linewidth < 20 Hz
• Dimensions: 58 x 43 x 40 cm
• Weight: 55 kg
• Magnet: Permanent and cryogen free
• Stray field: < 2 G all around system

In house made flow NMR cell:

Benchtop IR: Thermo Scientific™ Nicolet™iS™5 FT-IR equipped with a ZnSe Golden Gate Attenuated Total Reflectance Infrared **(**ATR-IR) flow cell

1.1 Robot assembly guide

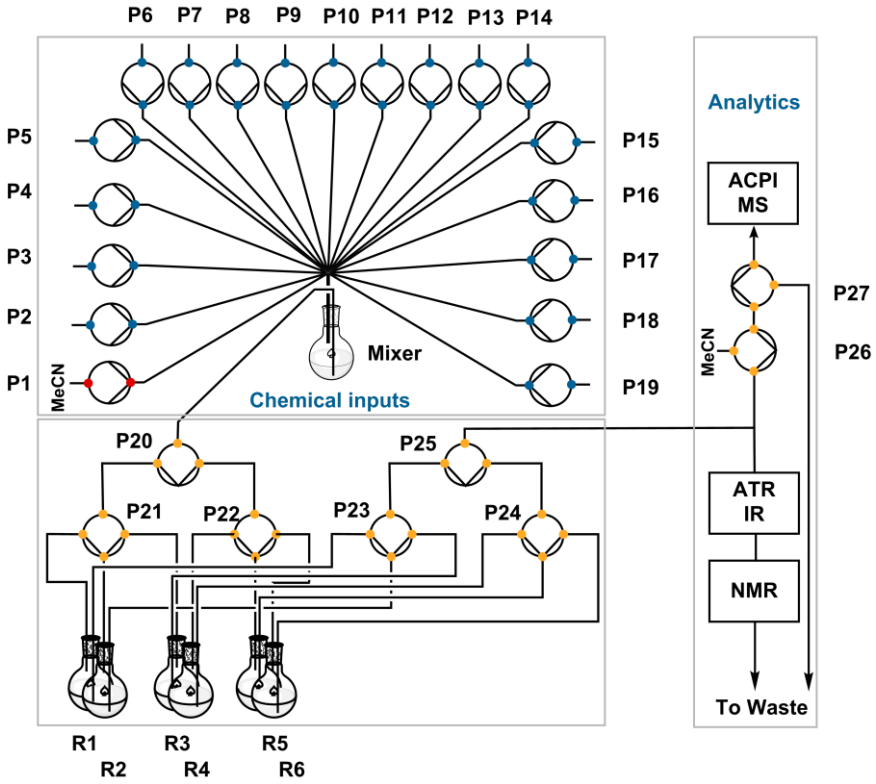


Figure S1. Platform scheme with pump numeration.

Solvent pump: Connect pump **P1** input valve position to the bottle with solvent (acetonitrile) and output valve position to the mixer using needle and Luer to 1/4"-28 Flat Bottom adapter (**P1** on Scheme S1). **Starting materials pumps:** For each of eighteen pumps (**P2-P19**): Connect the input valve position of the pump to the bottle containing starting material with PTFE tubing. Connect the output valve position of the pump to the mixer flask equipped with septa using Luer to 1/4"-28 Flat Bottom adapter and a long needle with PTFE tubing. **Reactors' pumps:** 1. For pump (**P20**) (moving the reaction mixtures) connect the input valve position to the mixer using Luer to 1/4"-28 Flat Bottom adapter and a needle (the needle should touch the bottom of the flask to ensure complete transfer of reaction mixtures. Connect the output and extra valve position of this pump (**P20**) to the next two pumps (**P21-22**) to the S valve position. Connect to the *I,O,E* positions of pumps **P21-P22** to reactors **R1-R6**. For pumps, **P23-P24** connect the *I,O,E* valve positions to the respective reactors. Connect the S valve positions of pumps **23** and **24** to the *I, E* and valve positions of the valve **P25**. Connect the output valve position of the pump **P25** to a 3-way block connector. **Analytics:** Connect the ATR-IR flow cell to the 3-way block connector. Connect the second end of the ATR-IR flow cell to the NMR flow cell. Connect the output of the NMR flow cell

to the waste bottle. **Dilution pump and ms pump:** Connect the input valve position of the pump **P26** (equipped with 0.5 mL syringe) to the 3-way block connector. Connect the *E* valve position with the solvent bottle. Connect the *O* valve position with the *S* position of the pump **P27**. Connect the input valve of **P27** to the Advion Mass spectrometer ACPI source. Connect the output valve position of **P27** to the waste bottle. In total three RS232 connections were utilized to connect the pumps to the computer. The pumps can be conveniently connected to the USB via RS232 to USB converter cable (see Figures S2-S5 for cabling diagram, and a photo of connections).

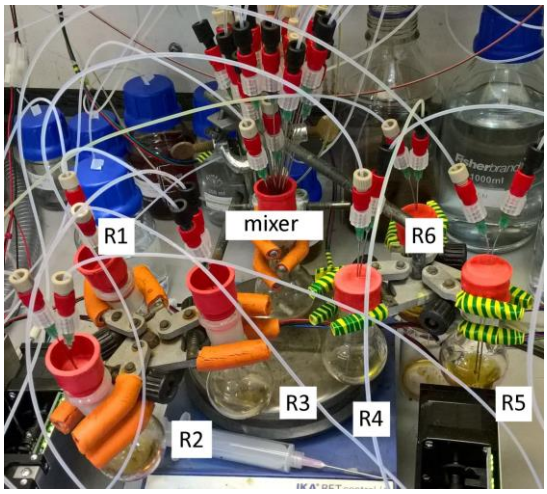


Figure S2. A picture of mixer and reactors of assembled robot.

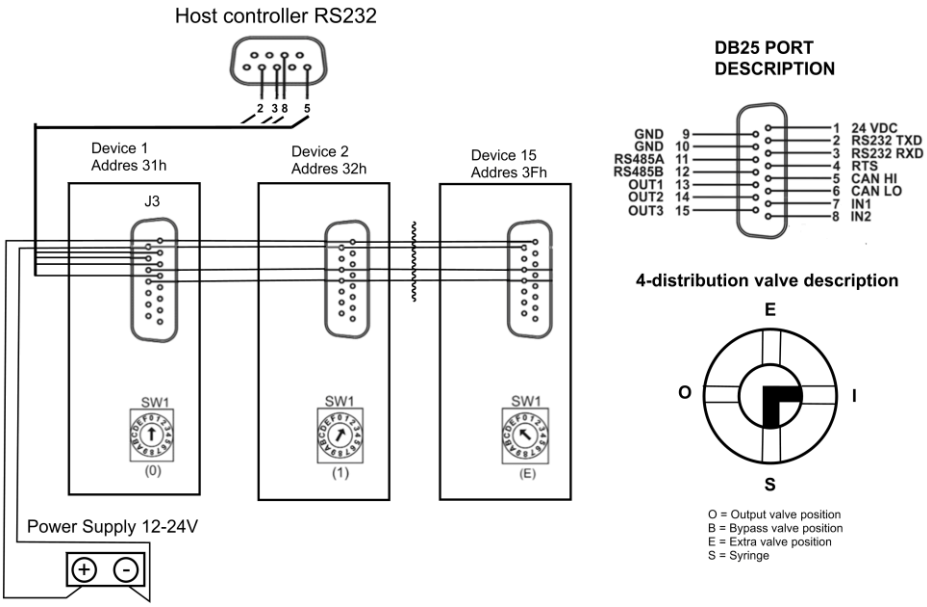


Figure S3. Cabling diagram for connecting the syringe pumps to the RS232 port and daisy chain, selecting pump address and 4-way distribution valve description.

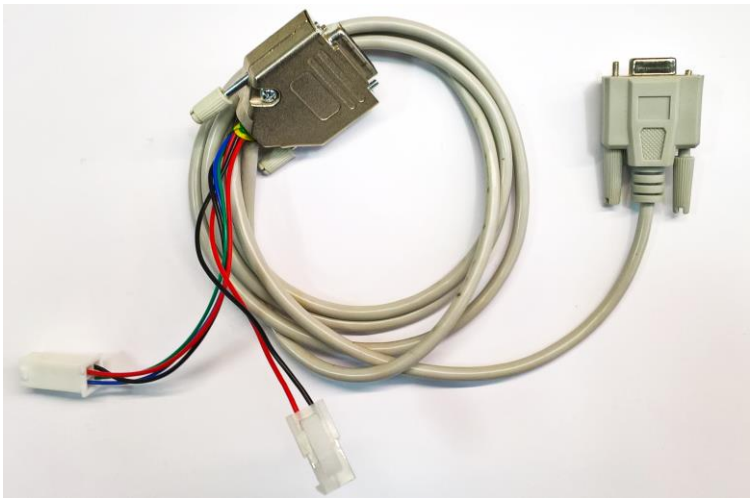


Figure S4. A connector for conecting the syringe pumps to RS232 port.

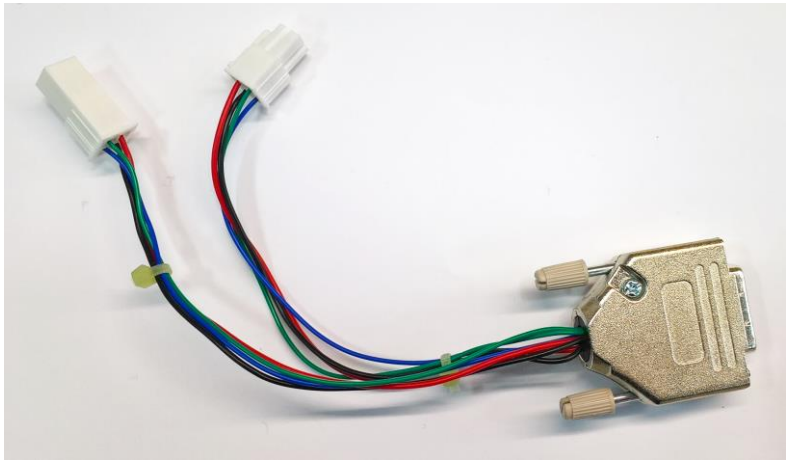


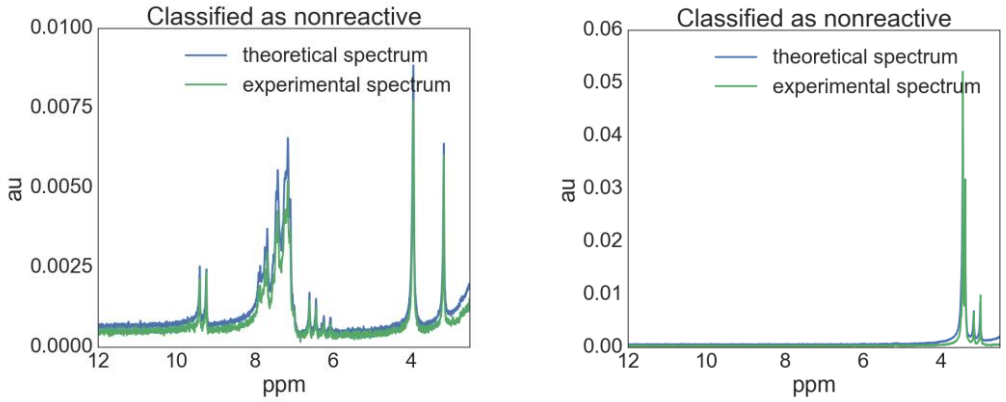
Figure S5. An example of daisy-chain DB25 connector.

2. Reaction detection algorithm details

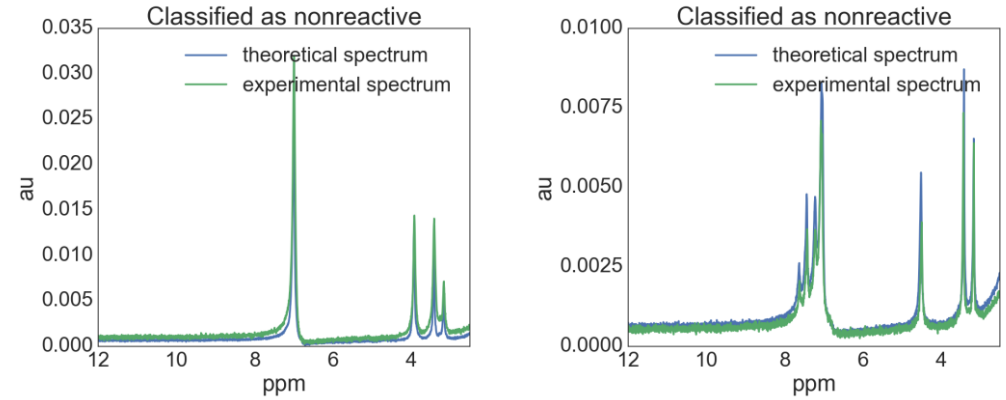
The reaction detection problem has been reformulated as a binary classification problem given the spectra of the starting materials and reaction mixture (X) estimate the category of the reactivity Y = 0 (unreactive) Y=1 (reactive). Training the SVM machine learning classifier for reaction prediction based on NMR/IR: The training set constituted 72 reactions. The category of reactivity has been assigned for each experiment by an expert chemist. Processing of NMR spectra: a) Fourier transform of FID b) Auto phase the spectrum c) Reference the solvent d) Normalize the intensity of the solvent peak to 1.0 e) Cut the spectrum to the region between 2.5 and 12.0 ppm. The IR spectra were used without any pre-processing. 1. For each experiment in the training set: a) Load the nmr/ir spectrum of the reaction mixture b) Calculate the theoretical spectrum of the reaction mixture (sum of the starting materials) c) Read from file the expected reaction category y = 0 (unreactive) y =1 (reactive) (the reactivity of the reactions has been labelled by an expert chemist) d) Combine theoretical and experimental spectra and add them to the training set. 2. Train SVM model on the reactivity labels (*Y*= 0 or 1) and the spectra (*X*). The SVM classifiers for reaction detection were validate using leave one out cross validation achieving an accuracy of 86 % percent for classifying the reaction as reactive or unreactive.

To automatically detect a chemical reaction by a robot the following algorithm has been implemented: 1. Load and process the NMR/IR spectrum of the reaction mixture. 2. Load reaction configuration file with volumes and concentration of the reagents 3. Load and process the spectra of starting materials. 4. Create the theoretical spectrum of the reaction mixture by adding spectra of starting materials taking into account concentrations and volumes of each starting materials. 5. Feed the reaction mixture and theoretical IR/NMR spectrum to the trained SVM classifier. 6. Predict reactivity of the reaction mixture y = 0 (unreactive) or y = 1 (reactive) For examples of reactive and unreactive examples of reaction mixtures classified by NMR and IR see figure S6-S9.

a) **b)**



c) **d)**



e) **f)**

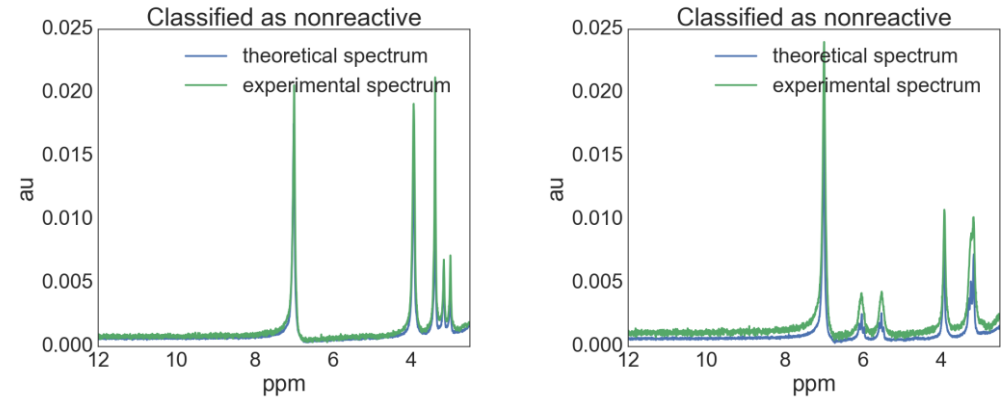


Figure S6. **Examples of ^{1}H NMR (43 MHz) of unreactive reaction mixtures as classified by machine learning: a)** benzoylcyanide + cinnamaldehyde + nitromethane ***b)*** dimethylacetylene-dicarboxylate + methylpropiolate **c)** malononitrile + phenylacetylchloride **d)** TosMic + benzofuroxan + malononitrile **e)** methylpropiolate+ nitromethane + phenylacetylchloride **f)** itaconicanhydride + phenylacetylchloride + trichloroacetonitrile

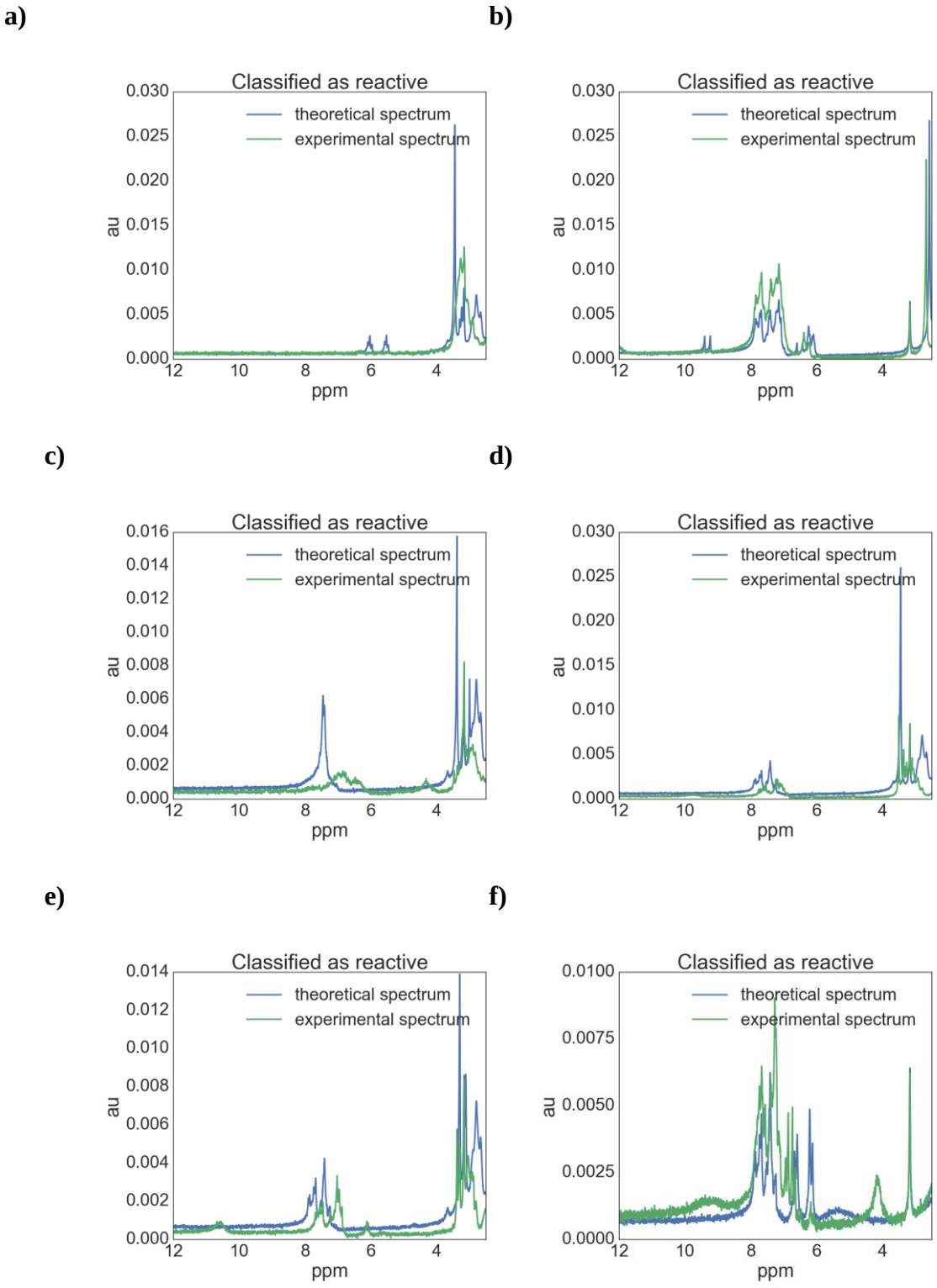


Figure S7. Examples of ^{1}H NMR (43 MHz) of reactive reaction mixtures as classified by machine learning: **a)** 2-aminthiazole + benzoylcyanide **b)** DMAP + benzoylcyanide + cinnamaldehyde **c)** DBU + methylpropiolate + nitrosobenzene **d)** DBU + benzoylcyanide + dimethylacetylene-dicarboxylate **e)** DBU+ benzoylcyanide + methylacetoacetate **f)** 2aminothiazole + benzoylcyanide

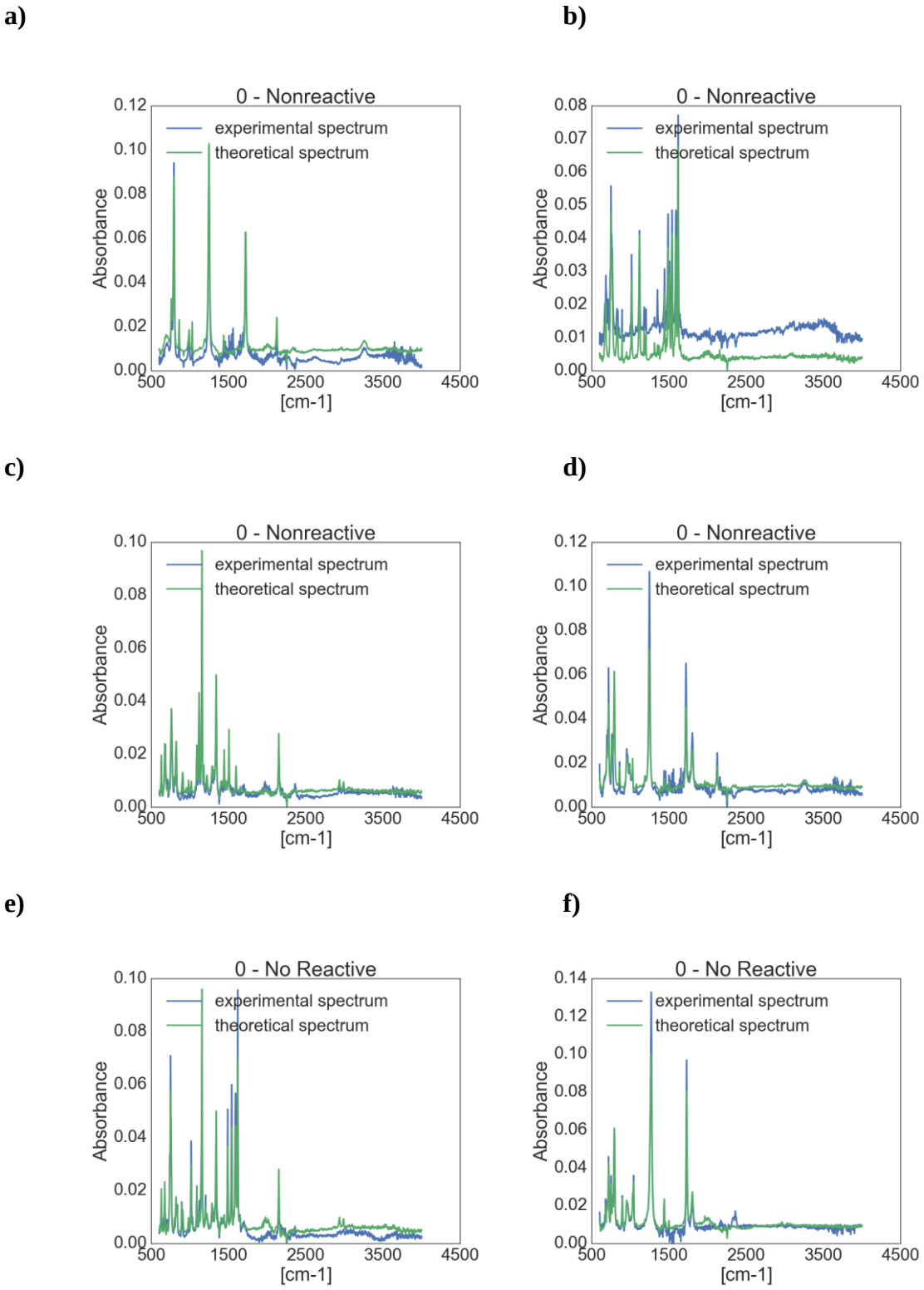
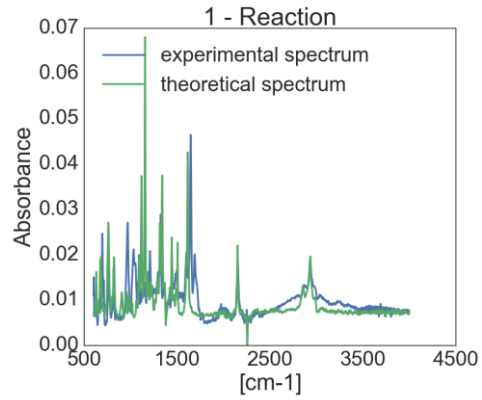
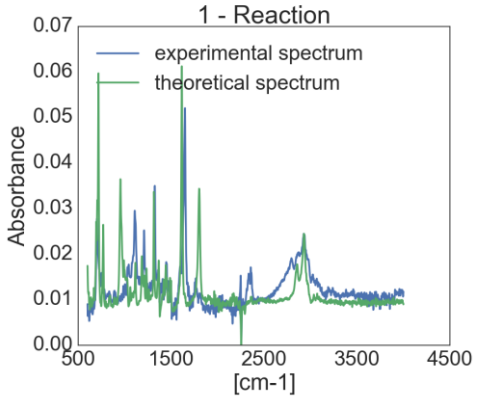


Figure S8. Examples of ATR-IR spectra of unreactive reaction mixtures as classified by machine learning**: a)** methyl propiolate + trichloroacetonitrile **b)** TosMic + nitrosobenzene **c)** benzofuroxan + nitrosobenzene **d)** methylpropiolate + phenylacetylchloride + trichloroacetonitrile **e)** TosMic benzofuroxan **f)** dimethylacetylenedicarboxylate + phenylacetylchloride + trichloroacetonitrile

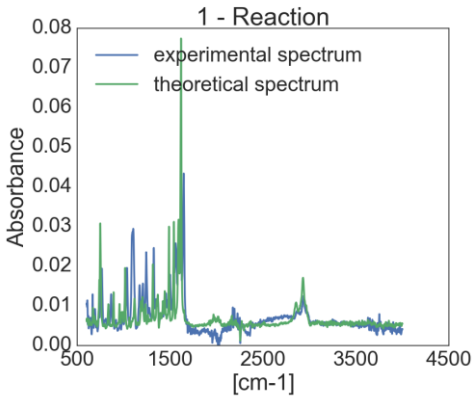
a)



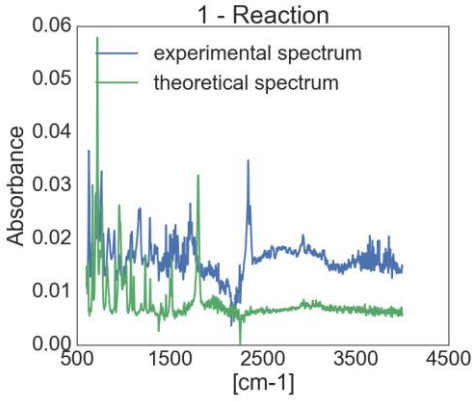
b)



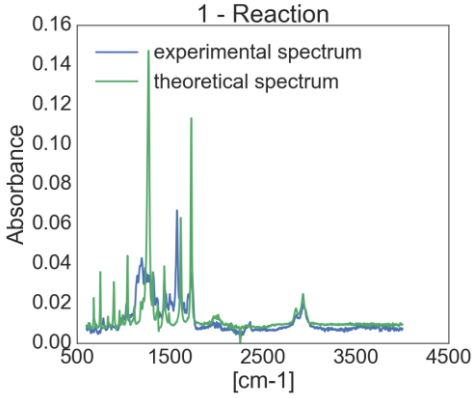
c)



d)



e)



f)

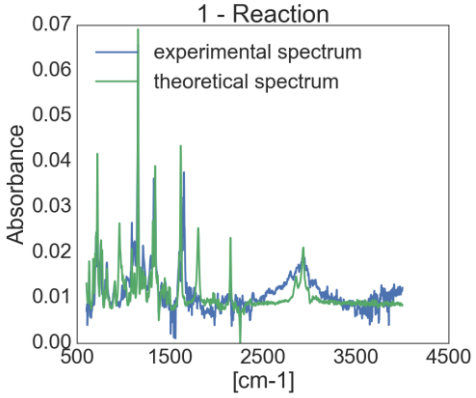


Figure S9. Examples of ATR-IR spectra of reactive reaction mixtures as classified by machine learning**: a)** DBU + TosMic + nitrosobenzene **b)** DBU + phenylacetyl chloride **c)** methyl imidazole phenylacetyl chloride **d)** DBU+DMAD **e)** DBU+benzofuroxan + malononitrile **f)** DBU+ TosMic + phenylacetyl chloride.

2.1 Chemical space exploration algorithm

The machine learning algorithm for exploration of chemical space has been reformulated as follows:

Start with a pool of unlabelled data (pool of chemical reactions from a given chemical space). Choose a few reactions at random and perform them in the robotic platform (measure reaction outcomes using in-line analytics) assign them to one of the two categories (reactive or unreactive) using SVM reactivity classifier. For all the reaction performed so far, fit the LDA classifier - build the model of chemical space. The goal which was set to the robotic platform was to explore the most reactive parts of chemical space by querying the model to assess the reactivity of all unlabelled (unperformed) reactions and perform the most reactive reactions in the robot. Update the model of chemical space with all the reactions performed in the platform. Iteratively explore the chemical space searching for maximum reactivity;

In the initial phase, the chemical space is being randomly explored collecting the initial data required for building model of it. The robot performs the reaction by addition of starting materials to the mixer and then the reaction mixture is transferred to the proper reaction flask. After reaction time the reaction mixture is analysed with NMR and IR. The reaction is then encoded as a reaction vector X *e.g.* $X = [1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0]$ and also is being classified as reactive or non-reactive $Y = 1$ or $Y = 0$ and this information is then added to the database performed reactions. Then the Linear Discriminant Analysis classifier is trained on the obtained data. All the unperformed reaction from the pool of the starting materials (two and three component) are then scored according to the probability of reaction obtained from the LDA. The reactions with the highest probability are then performed and analysed in the robotic system. Then the outcome of the reaction is added to the existing database of the reactions and the LDA classifier is retrained on the obtained data. In this way, it was possible to implement a feedback loop allowing the robot to explore the chemical space in iterative fashion gaining knowledge of chemical space as the chemical space is being explored. We were interested if this approach to the exploration of chemical space can lead to its efficient navigation and the robot can learn the reactivity patterns. The algorithm has been validated using 5-fold cross validation giving accuracy 86 % of predicting reactivity and ROC area = 0.89 (see Figure S10).

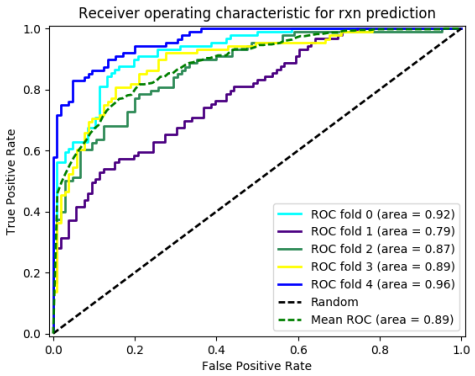


Figure S10. Receiver operating characteristic for LDA classifier in predicting reactivity of unknown reactions.

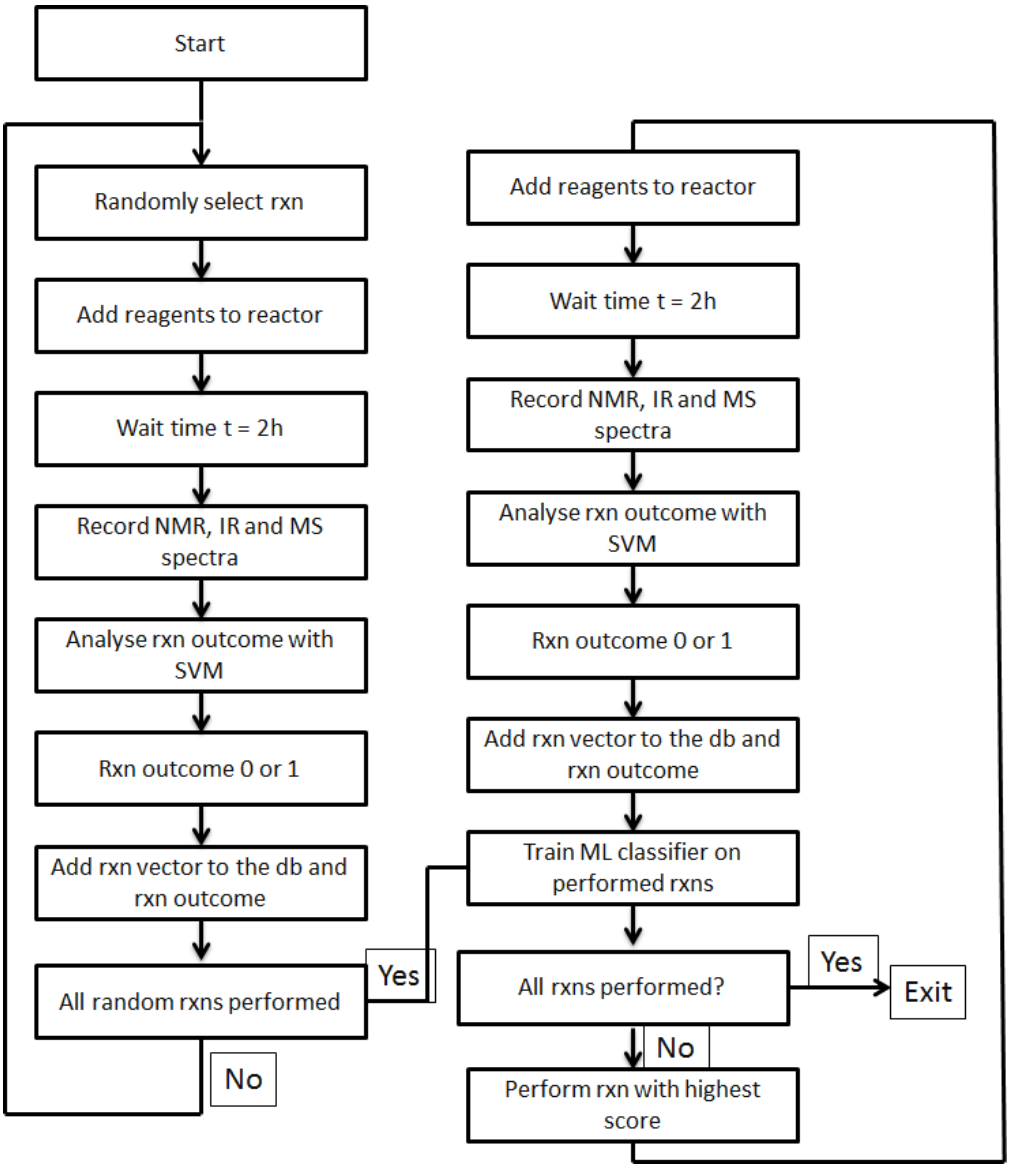


Figure S11. The detailed algorithm for the exploration of the chemical space.

3. Real-time spectra recorded in the platform for discovered reactions indicating new reactions

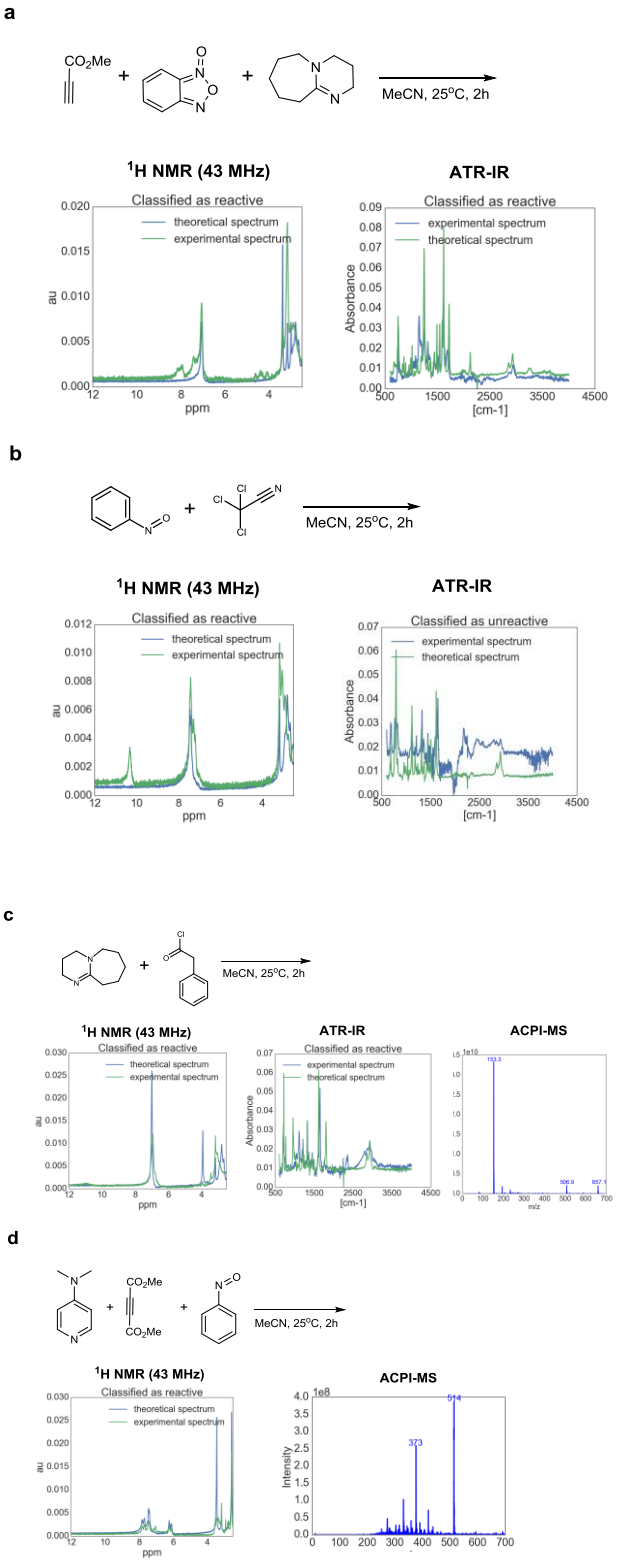
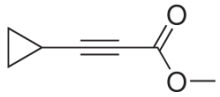


Figure S12. Real time spectra recorded for the new transformations.

4. Precursors Syntheses (S1-S6)

A solution of 2.0 M *n*-butyllithium in cyclohexane (11.2 mL, 22.3 mmol) was slowly added to a solution of terminal alkyne (22.3 mmol) in dry THF at -80 °C. After an hour methyl chloroformate (5.264 g, 55.7 mmol) was added at -80 °C. After one hour the reaction mixture was slowly allowed to warm up. The reaction was then quenched in a mixture of ice and water and extracted with ethyl acetate. The combined organic phase was dried on Na_2SO_4 and after filtration, the solvent was removed under vacuum. Each product was obtained by column chromatography using a mixture of ethyl acetate/hexane of 20/80 (*v*/*v*).

Methyl cyclopropyl-propynoate (S1)



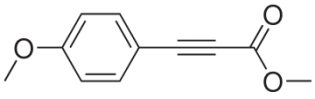
Yield: 1.4 g corresponding to 51.3% yield

Physical state: colorless liquid.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 3.73 (s, 3H, CH_3), 1.41-1.33 (m, 1H, CH), 0.97-0.88 (m, 4H, CH_2) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 154.9 (C=O), 94.3 (C≡CC=O), 68.9 (C≡CC=O), 53.2 (OCH_3), 9.9 (CH_2, 2C), 0.1 (CH) ppm. Ref peak at 77.75

Methyl (4-methoxy-phenyl)-propynoate (S2)



Yield: 3.4 g corresponding to 80.0% yield

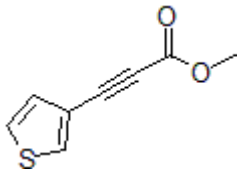
Physical state: viscous yellow liquid

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 7.55-7.52 (m, 2H, PhCH close to OMe), 6.90-6.86 (m, 2H, PhCH close to alkyne), 3.83 (s, 3H, $PhOCH_3$), 3.82 (s, 3H, CO_2CH_3), ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 161.6 (PhC_q close to OMe), 154.7 (C=O), 135.0 (PhCH close to OMe), 114.3 (PhCH close to alkyne), 111.3 (PhC_q close to alkyne), 87.4 (C≡CC=O), 79.8 (C≡CC=O), 55.4 ($PhOCH_3$), 52.7 (OCH_3) ppm.

HRESI-MS: $[C_{11}H_{10}NaO_3]^+$ calculated 213.05, found 213.0516.

Methyl thiophen-3-yl-propynoate (S3)



Yield: 3.52 g corresponding to 95.0% yield

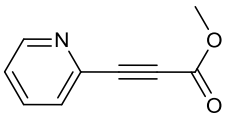
Physical state: yellow viscous liquid

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 7.75 (dd, 1H, J = 1.2 *Hz*, J = 3.0 *Hz*), 7.32 (dd, 1H, J = 3.0 *Hz*, J = 5.0 *Hz*), 7.23 (dd, 1H, J = 1.2 *Hz*, J = 5.0 *Hz*), 3.83 (s, 3H, CO_2CH_3), ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 154.5 (C=O), 133.8 (ArCH, C2), 130.2 (ArCH, C4), 130.2 (ArCH, C5), 111.3 (ArC_q), 81.9 (C≡CC=O), 80.4 (C≡CC=O), 52.8 (OCH_3) ppm.

HRESI-MS: $[C_8H_6NaO_2S]^+$ calculated 189.00, found 188.9976.

Methyl 3-(pyridin-2-yl)propiolate (S4)



Yield: 2.822 g corresponding to 78% yield

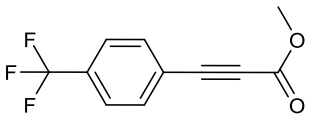
Physical state: off white solid

^{1}H-NMR ($CDCl_3$, 400 MHz): ^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.64 (ddd, *J* = 4.8, 1.8, 1.0 Hz, 1H), 7.71 (tdd, *J* = 7.7, 1.8, 0.8 Hz, 1H), 7.58 (dq, *J* = 7.8, 1.1 Hz, 1H), 7.34 (ddt, *J* = 7.7, 4.8, 1.0 Hz, 1H), 3.84 (s, 3H).ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): ^{13}C NMR (101 MHz, Chloroform-*d*) δ 153.86 , 150.51 , 140.47 , 136.28, 128.54 , 124.63 , 84.17 , 78.79 , 52.93 ppm.

HRESI-MS: $[M+Na]^+$ calculated 184.0369, found 184.0367.

Methyl 3-(4-(trifluoromethyl)phenyl)propiolate (S5)



Yield: 994 mg corresponding to 58% yield

Physical state: yellow solid

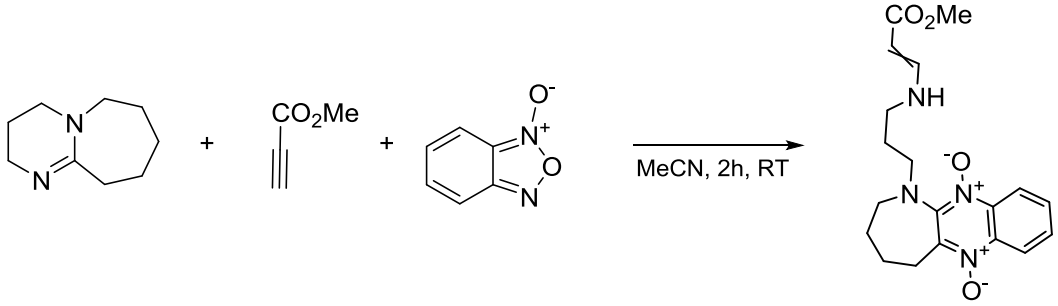
^{1}H-NMR ($CDCl_3$, 400 MHz) δ 7.69 (d, *J* = 8.2 Hz, 2H), 7.64 (d, *J* = 8.3 Hz, 2H), 3.85 (s, 3H) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): ^{13}C NMR (101 MHz, $CDCl_3$) δ 153.95 (s), 133.14 (s), 132.24 (q, *J* = 33.0 Hz), 125.53 (q, *J* = 3.8 Hz), 123.50 (q, *J* = 272.6 Hz), 123.37 (q, *J* = 1.6 Hz), 84.18 (s), 81.93 (s), 52.95 (s) ppm.

HREI-MS: $[M]^+$ calculated 228.0348, found 228.0398.

5. New reactions discovered using the platform

5.1 Multicomponent reaction of methyl propiolate, benzofuroxan, DBU



To a round bottom flask were added benzofuroxan (1.67 mmol), methyl propiolate (1.67 mmol) and DBU (1.67 mmol) and the reaction was stirred for 2 h at room temperature. Then silica gel (5.0 g) was added and the solvent was evaporated. The product was purified by column chromatography AcOEt/Acetone (1/1) (*v*/*v*) yielding products as yellow sticky solid 192 mg inseparable mixture of isomers (Z/E 1.4:1).

Note: The structure of the product has been determined for compound **19b**, being single Z isomer. The identity of isomers has been assign from ^{1}H-^{1}H NOESY correlation and ratio of the isomers from the integration of the respective signals.

Z/E-1-(3-((3-methoxy-3-oxoprop-1-en-1-yl)amino)propyl)-2,3,4,5-tetrahydro-1H-azepino[2,3-b]quinoxaline 6,11-dioxide (19)

Yield: 192 mg corresponding to 31% yield

Physical state: yellow sticky solid.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ ^{1}H NMR (600 MHz, Chloroform-*d*) δ 8.47 (qd, *J* = 8.8, 1.3 Hz, 2H), 7.93 – 7.87 (m, 0.41H), 7.72 (q, *J* = 8.5, 7.9 Hz, 1H), 7.65 – 7.61 (m, 1H), 7.60 – 7.55 (m, 0.55H), 6.62 (dd, *J* = 13.2, 8.0 Hz, 0.42H), 6.20 (bs, 0.60H), 4.70 (d, *J* = 13.2 Hz, 0.58H), 4.42 (d, *J* = 8.0 Hz, 0.42H), 3.64 – 3.50 (m, 5H), 3.43 (t, *J* = 5.7 Hz, 2H), 3.35 – 3.17 (m, 4H), 2.03 – 1.85 (m, 6H).

^{13}C-NMR ($CDCl_3$, 100 MHz): δ ^{13}C NMR (151 MHz, Chloroform-*d*) δ 170.56 , 169.74 , 151.66 , 148.97 , 146.86 , 146.17 , 141.74 , 141.67 , 135.80 , 135.53 , 134.23 , 134.19 , 130.98 , 130.78 , 128.70 , 128.60 , 119.39 , 119.32 , 118.61 , 118.60 , 83.59 , 81.18 , 49.96 ,

49.67 , 48.11 , 46.73 , 46.00 , 45.81 , 44.58 , 29.65 , 26.20 , 25.94 , 23.74 , 23.12 , 20.95 , 20.78.

HRESI-MS: $[C_{19}H_{24}N_4NaO_4]^+$ calculated 395.1690, found 395.1680.

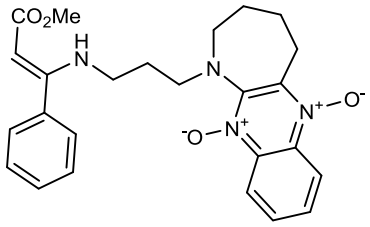
5.2 Synthesis of small library of molecules by multicomponent reaction of DBU, benzofuroxan and methyl propiolates

3-substituted-3-[3-(6,11-dioxy-2,3,4,5-tetrahydro-azepino[2,3-*b*]quinoxalin-1-yl)-propylamino]-acrylic acid esters syntheses

General Procedure

2 mmol of benzofuroxan were dissolved in 5 mL of DCM, then the alkyne ester (2 mmol, 272.22 mg) and DBU (2 mmol, 304.48 mg) were added. The reaction was stirred at room temperature for 2 hours, before quenching it with 1 mL of NH_4Cl (1M) and 3 ml of water. The organic phase was collected; the aqueous phase was then washed with DCM (2 x 3 mL). The combined organic phase was dried on Na_2SO_4 and after filtration the solvent was removed under vacuum. The crude reaction was then purified by column chromatography using a mixture of ethyl acetate/acetone of 35/65 (*v*/*v*).

Z-3-(phenyl)-3-[3-(6,11-dioxy-2,3,4,5-tetrahydro-azepino[2,3-*b*]quinoxalin-1-yl)-propylamino]-acrylic acid methyl esters (19b)



Yield: 425 mg corresponding to 47.4% yields

Physical state: brown sticky solid.

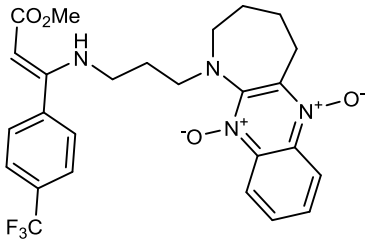
^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.58 (bb, 1H, NH overlap with other peaks), 8.55 (dd, J = 1.1 *Hz*, J = 8.5 *Hz*, 1H, ArCH C9-cycle), 8.49 (dd, J = 1.1 *Hz*, J = 8.5 *Hz*, 1H, ArCH C7-cycle), 7.75 (ddd, J = 1.4 *Hz*, J = 7.0 *Hz*, 1H, J = 8.5 *Hz*, ArCH C8-cycle), 7.66 (ddd, J = 1.4 *Hz*, J = 7.0 *Hz*, J = 8.5 *Hz*, 1H, ArCH C10-cycle), 7.38-7.30 (m, 5H, ArCH), 4.60 (s, 1H,

CH), 3.65 (s, 3H, CH_3), 3.59-3.56 (m, 2H, CH_2N), 3.37-3.34 (m, 2H, CH_2N C2-cycle), 3.27-3.18 (m, 4H, $\underline{CH_2}NH$ and C5-cycle), 1.93-1.80 (m, 6H, $\underline{CH_2}CH_2NH$ and C3,4-cycle) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 170.6 (C=O), 164.8 ($\underline{C}$=CH), 146.3 (C_q between N1,N11-cycle), 142.2 (C_q between C5,N6-cycle), 136.4 (C_q between C10,N11-cycle), 136.4 (C_q Ph), 134.8 (C_q between C10,N11-cycle), 131.2 (CH, C8-cycle), 129.3 (CH, C4-phenyl), 129.0 (CH, C10-cycle), 128.4 (CH, C2,6-phenyl), 127.8 (CH, C3,5-phenyl), 119.9 (CH, C9-cycle), 119.1 (CH, C7-cycle), 85.2 (CH), 50.2 (CH_3), 49.5 (CH_2), 47.2 (CH_2), 42.5 (CH_2), 30.1 (CH_2), 24.6 (CH_2), 21.6 (CH_2), 21.6 (CH_2) ppm.

HRESI-MS: $[C_{25}H_{28}N_4NaO_4]^+$ calculated 471.2008, found 471.1985.

(Z)-1-(3-((3-methoxy-3-oxo-1-(4-(trifluoromethyl)phenyl)prop-1-en-1-yl)amino)propyl)-2,3,4,5-tetrahydro-1H-azepino[2,3-b]quinoxaline 6,11-dioxide (19c)



Yield: 348 mg corresponding to 34% yields

Physical state: yellow sticky solid.

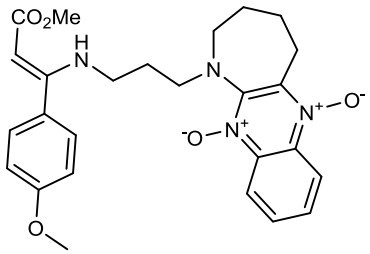
^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.58 (s, 1H), 8.54 (dd, *J* = 8.5, 1.4 Hz, 1H), 8.48 (dd, *J* = 8.6, 1.4 Hz, 1H), 7.74 (dd, *J* = 8.5, 1.5 Hz, 1H), 7.67 (dq, *J* = 7.2, 3.3, 2.4 Hz, 3H), 7.49 (d, *J* = 8.0 Hz, 2H), 4.60 (s, 1H), 3.65 (s, 3H), 3.59 – 3.50 (m, 2H), 3.38 (t, *J* = 5.6 Hz, 2H), 3.31 (dd, *J* = 7.6, 4.1 Hz, 2H), 3.19 (q, *J* = 6.7 Hz, 2H), 2.05 – 1.77 (m, 6H).

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

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.40, 162.98, 146.43, 142.09, 139.65, 136.35, 134.80, 131.41 (q, *J* = 128 Hz), 131.23, 129.09, 128.33, 125.52 (q, *J* = 3.6 Hz), 123.79 (q, *J* = 272.1 Hz), 119.95, 119.10, 86.23, 50.37, 49.01, 46.69, 42.52, 30.04, 26.24, 24.43, 21.47.

HRESI-MS: $[C_{26}H_{27}F_3N_4O_4+Na]^+$ calculated 539.1877, found 539.1858

(Z)-3-(4-methoxy-phenyl)-3-[3-(6,11-dioxy-2,3,4,5-tetrahydro-azepino[2,3-*b*]quinoxalin-1-yl)-propylamino]-acrylic acid methyl esters (19d)



Yield: 107 mg corresponding to 11.2 % yields

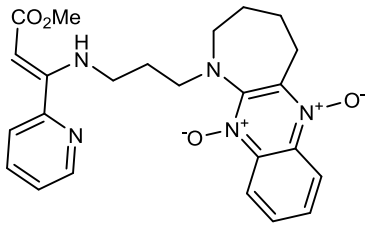
Physical state: brown sticky solid.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.55 (bb, 1H, NH overlap with other peaks), 8.52 (dd, J = 1.0 *Hz*, J = 8.5 *Hz*, 1H, ArCH C9-cycle), 8.47 (dd, J = 1.0 *Hz*, J = 8.5 *Hz*, 1H, ArCH C7-cycle), 7.73 (ddd, J = 1.4 *Hz*, J = 7.0 *Hz*, 1H, J = 8.5 *Hz*, ArCH C8-cycle), 7.64 (ddd, J = 1.4 *Hz*, J = 7.0 *Hz*, J = 8.5 *Hz*, 1H, ArCH C10-cycle), 7.25 (d, J = 8.8 *Hz*, 2H, ArCH C3,5-phenyl), 6.85 (d, J = 8.8 *Hz*, 2H, ArCH C2,4-phenyl), 4.58 (s, 1H, CH), 3.82 (s, 3H, $PhOCH_3$), 3.62 (s, 3H, $C(O)OCH_3$), 3.59-3.54 (m, 2H, CH_2N), 3.38-3.34 (m, 2H, CH_2N C2-cycle), 3.28-3.19 (m, 4H, CH_2NH and C5-cycle), 1.93-1.79 (m, 6H, CH_2CH_2NH and C3,4-cycle) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 170.7 (C=O), 164.8 (PhCOMe), 160.5 (C=CH), 146.3 (C_q between N1,N11-cycle), 142.2 (C_q between C5,N6-cycle), 136.4 (C_q between C10,N11-cycle), 134.8 (C_q between C10,N11-cycle), 131.2 (CH, C8-cycle), 129.2 (CH, C3,5-phenyl), 129.0 (CH, C10-cycle), 128.2 (C_q Ph), 119.9 (CH, C9-cycle), 119.1 (CH, C7-cycle), 113.8 (CH, C2,6-phenyl), 85.1 (CH), 55.3 (CH_3), 50.2 (CH_3), 49.6 (CH_2), 47.2 (CH_2), 42.5 (CH_2), 30.2 (CH_2), 26.1 (CH_2), 24.7 (CH_2), 21.5 (CH_2) ppm.

HRESI-MS: $[C_{26}H_{30}N_4NaO_5]^+$ calculated 501.2113, found 501.2172.

(Z)-1-(3-((3-methoxy-3-oxo-1-(pyridin-2-yl)prop-1-en-1-yl)amino)propyl)-2,3,4,5-tetrahydro-*1H*-azepino[2,3-b]quinoxaline 6,11-dioxide (19e)



Yield: 232 mg corresponding to 26% yields

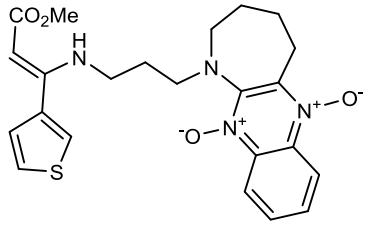
Physical state: yellow sticky solid.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.54 (ddd, *J* = 4.8, 1.8, 0.9 *Hz*, 1H), 8.51 – 8.44 (m, *2H*), 8.45 – 8.38 (m, 1H), 7.71 – 7.55 (m, 3H), 7.34 (dt, *J* = 7.8, 1.1 *Hz*, 1H), 7.23 (ddd, *J* = 7.6, 4.8, 1.2 Hz, 1H), 4.63 (s, 1H), 3.58 (s, 3H), 3.56 – 3.51 (m, 2H), 3.35 (t, *J* = 5.5 Hz, 2H), 3.27 – 3.19 (m, 4H), 1.93 – 1.75 (m, 6H).

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 170.65 , 162.15 , 154.65 , 149.16 , 146.31 , 142.21 , 136.78 , 136.39 , 134.79 , 131.15 , 128.99 , 123.87 , 123.31 , 119.95 , 119.10 , 85.74 , 50.31 , 49.57 , 47.18 , 42.38 , 30.14 , 26.14 , 24.62 , 21.52 .

HRESI-MS: $[C_{24}H_{27}N_5NaO_4]^+$ calculated 472.1955, found 472.1939.

(Z)-3-(thiophen-3-yl)-3-[3-(6,11-dioxy-2,3,4,5-tetrahydro-azepino[2,3-*b*]quinoxalin-1-yl)-propylamino]-acrylic acid methyl esters (19f)



Yield: 236 mg corresponding to 26.0% yields

Physical state: brown sticky solid.

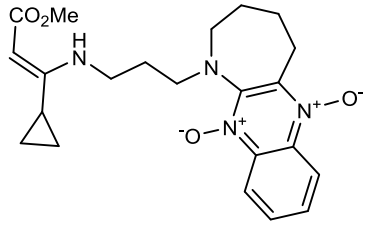
^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.56 (bb, 1H, NH overlap with other peaks), 8.55 (dd, J = 1.1 *Hz*, J = 8.5 *Hz*, 1H, ArCH C9-cycle), 8.50 (dd, J = 1.1 *Hz*, J = 8.5 *Hz*, 1H, ArCH C7-cycle), 7.75 (ddd, J = 1.4 *Hz*, J = 7.1 *Hz*, 1H, J = 8.5 *Hz*, ArCH C8-cycle), 7.67 (ddd, J = 1.4 *Hz*, J = 7.1 *Hz*, J = 8.5 *Hz*, 1H, ArCH C10-cycle), 7.38 (dd, J = 1.3 *Hz*, J = 3.0 *Hz*, 1H, ArCH

C2-thiophen), 7.32 (dd, J = 3.0 *Hz*, J = 5.0 *Hz*, 1H, ArCH C5-thiphen), 7.07 (dd, J = 1.3 *Hz*, J = 5.0 *Hz*, 1H, ArCH C4-thiphen), 4.68 (s, 1H, CH), 3.64 (s, 3H, CH_3), 3.62-3.56 (m, 2H, CH_2N), 3.40-3.37 (m, 2H, CH_2N C2-cycle), 3.33-3.28 (m, 4H, CH_2NH and C5-cycle), 1.98-1.82 (m, 6H, CH_2CH_2NH and C3,4-cycle) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 170.7 (C=O), 159.5 (C=CH), 146.3 (C_q between N1,N11-cycle), 142.2 (C_q between C5,N6-cycle), 136.7 (C_q between C10,N11-cycle), 136.4 (C_q thiophene), 134.8 (C_q between C10,N11-cycle), 131.2 (CH, C8-cycle), 129.1 (CH, C10-cycle), 127.2 (CH, C4-thiphene), 126.1 (CH, C5-thiphene), 125.3 (CH, C2-thiphene), 120.0 (CH, C9-cycle), 119.1 (CH, C7-cycle), 85.0 (CH), 50.2 (CH_3), 49.5 (CH_2), 47.0 (CH_2), 42.6 (CH_2), 30.2 (CH_2), 26.2 (CH_2), 24.6 (CH_2), 21.5 (CH_2) ppm.

HRESI-MS: $[C_{23}H_{26}N_4NaO_5S]^+$ calculated 477.1572, found 477.1549.

(Z)-3-cyclopropyl-3-[3-(6,11-dioxy-2,3,4,5-tetrahydro-azepino[2,3-*b*]quinoxalin-1-yl)-propylamino]-acrylic acid methyl esters (19g)



Yield: 184 mg corresponding to 22.0% yields

Physical state: brown sticky solid.

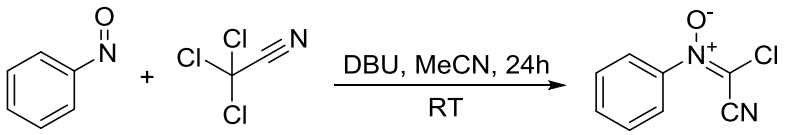
^{1}H-NMR ($CDCl_3$, 400 MHz): δ 8.70 (bb, 1H, NH), 8.54 (dd, J = 1.1 *Hz*, J = 8.6 *Hz*, 1H, ArCH C8-cycle), 8.52 (dd, J = 1.1 *Hz*, J = 8.6 *Hz*, 1H, ArCH C9-cycle), 7.75 (ddd, J = 1.3 *Hz*, J = 7.0 *Hz*, J = 8.6 *Hz*, 1H, ArCH C8-cycle), 7.66 (ddd, J = 1.3 *Hz*, J = 7.0 *Hz*, J = 8.6 *Hz*, 1H, ArCH C10-cycle), 4.26 (s, 1H, CH=C), 3.69-3.64 (m, 2H, CH_2CH_2N), 3.58 (s, 3H, CH_3), 3.56 (dt, J_d = 5.1 *Hz*, J_t = 6.8 *Hz*, 2H, $NHCH_2CH_2$), 3.49-3.47 (m, 2H, CH_2 C2-cycle), 3.40-3.37 (m, 2H, CH_2 C5-cycle), 2.12-2.05 (m, 2H, $NHCH_2CH_2CH_2$), 2.00-1.87 (m, 4H, CH_2 C3,4 cycle), 1.52 (tt, J = 5.4 *Hz*, J = 8.5 *Hz*, 1H, $CHCH_2$), 0.88-0.83 (m, 4H, $CHCH_2$) ppm.

^{13}C-NMR ($CDCl_3$, 100 MHz): δ 171.5 (C=O), 166.5 (C=CH), 146.5 (C_q between N1-N11-cycle), 142.2 (C_q between C5-N6-cycle), 136.4 (C_q between C10-N11-cycle), 134.8

(C_q between N6-C7-cycle), 131.2 (CH C7-cycle), 129.1 (CH C10-cycle), 119.9 (CH C8-cycle), 119.1 (CH C9-cycle), 77.6 (<u>CH</u>=C), 50.0 (OCH_3), 49.3 (CH_2 C2-cycle), 47.1 ($CH_2\underline{CH_2}N$), 40.6 ($NH\underline{CH_2}CH_2$), 29.7 ($NHCH_2\underline{CH_2}CH_2$), 26.3 ($CH_2$ C5-cycle), 24.8 (CH_2 C3-cycle), 21.5 (CH_2 C4-cycle), 12.4 ($\underline{CH}CH_2$), 6.8 (2C, $CH\underline{CH_2}$), ppm.

<u>HRESI-MS</u>: $[C_{22}H_{28}N_4NaO_4]^+$ calculated 435.2008, found 435.1994.

5.3 Synthesis of Z-1-chloro-1-cyano-*N*-phenylmethanimine (21) oxide from nitrosobenzene and trichloroacetonitrile in the presence of DBU



To a round bottom flask (10 mL) were added acetonitrile (5.0 mL), trichloroacetonitrile (2.0 mmol, 0.20 mL), and nitrosobenzene (2.0 mmol, 214 mg). Then DBU (2.0 mmol, 0.3 mL) was added dropwise. The reaction mixture was stirred for 24 h. After this time, the reaction mixture was poured into sat. aq. NH_4Cl (30 mL) and quickly extracted ethyl acetate (30 mL). The organic layer was dried with Na_2SO_4. The silica gel (4 g) was added and the solvent was removed *in vacuo*. The title product was purified by column chromatography hexane/ AcOEt 10/1 (*v/v*) giving pale white crystals (100 mg, 27 %). **Note:** Product should be stored in the fridge otherwise decomposition is visible;

<u>Yield</u>: 100 mg corresponding to 28.0% yields

<u>Rf</u>: 0.33 EtOAC/Hexane 10%

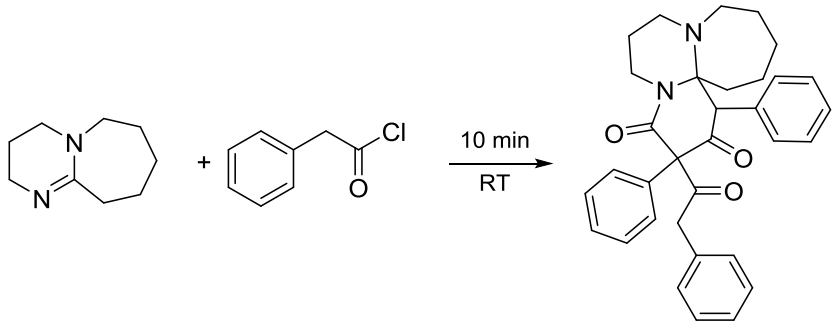
<u>Physical state</u>: white solid.

<u>^{1}H-NMR ($CDCl_3$, 600 MHz)</u>: δ 7.64-7.61 (m, 3H, PhCH), 7.57-7.54 (m, 2H, PhCH) ppm.

<u>^{13}C-NMR ($CDCl_3$, 150 MHz)</u>: δ 144.5 (C=Cl), 142.8 (<u>C</u>N=O), 132.2 (PhCH), 129.4 (2xPCH), 123.1 (2xPhCH), 111.0 (C≡N) ppm.

<u>HRESI-MS</u>: EI$^+$ (m/z): $[C_8H_5ClN_2O]^+$ calculated 180.01, found 180.00/APCI (m/z): calculated $[C_{11}H_8ClN_3O]^+$ 222.04, found 222.00

5.4 New reactivity of ketenes with DBU 1,3-diphenyl-3-(2-phenylacetyl) hexahydro-6*H*,10*H*-pyrido[2',1':2,3]pyrimido[1,2-a]azepine-2,4(1*H*,3*H*)-dione(22)



To a round bottom flask (10 mL) was added acetonitrile (5.0 mL) and phenylacetyl chloride (3.0 mmol, 0.40 mL) and dropwise DBU (3.0 mmol, 0.45mL). The reaction mixture was stirred for 10 min. and then poured into sat. NH_4Cl and extracted with Et_2O (2x50 mL). The organic phase was washed twice with sat. aq. $NaHCO_3$ and dried with Na_2SO_4. Then silica was added (4.0 g) and solvent was removed under reduced pressure. Column chromatography hexanes/AcOEt (5/1)(*v*/*v*) gave title product as a colourless solid (70 mg, 18%)

Yield: 70 mg corresponding to 18% yield

Physical state: colorless sticky solid.

^{1}H-NMR ($CDCl_3$, 400 MHz): ^{1}H NMR (600 MHz, $CDCl_3$-*d*) δ 7.34 – 7.13 (m, 13H), 6.92 – 6.87 (m, 2H), 4.56 (ddd, *J* = 13.6, 6.8, 2.0 Hz, 1H), 4.09 (s, 1H), 3.38 (q, *J* = 15.3 Hz, 2H), 2.97 – 2.80 (m, 2H), 2.68 (ddd, *J* = 14.9, 5.2, 3.1 Hz, 1H), 2.55 – 2.50 (m, 1H), 2.48 – 2.37 (m, 2H), 1.96 (dt, *J* = 12.4, 6.2 Hz, 1H), 1.83 – 1.72 (m, 2H), 1.70 – 1.59 (m, 3H), 1.57 – 1.38 (m, 3H).

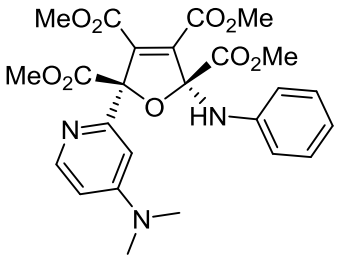
^{13}C-NMR ($CDCl_3$, 100 MHz): ^{13}C NMR (151 MHz, $CDCl_3$-*d*) δ 169.45 , 163.85 , 156.64 , 136.20 , 132.55 , 132.01 , 129.88 (s, 2C), 129.65 (s, 2C), 129.05(s, 2C) , 128.55(s, 2C) , 128.03 (s, 2C) , 127.71(s, 2C) , 127.60 , 127.58 , 127.15 , 124.57 , 80.02 , 51.24 , 49.46 , 45.53 , 40.82 , 38.16 , 36.09 , 28.51 , 28.40 , 23.46 , 22.21 .

HRESI-MS: $[C_{33}H_{34}N_2NaO_3]^+$ calculated 529.2462, found 529.2447.

5.5 Multicomponent reaction of DMAD, DMAP and nitrosobenzene

Nitrosobenzene (1.0 mL, 1M in MeCN) were combined with DMAD (2.0 mL, 1M in MeCN) followed by dropwise addition of DMAP (1.0 mL, 1M in MeCN). The reaction was stirred at room temperature for 2h. The reaction mixture was evaporated and the products were isolated with Reveleris X2 flash column chromatography by setting the gradient between 3:1 EtOAc/Hexane and 1:1 EtOAc/Hexane 0-1.9min, 79 % EA 1.9-2.9min 79-> 92% EA. flow 40 mL/min. **Note**: the configuration of diastereo-isomers has been assigned on the basis of X-ray analysis.

***trans*-2,3,4,5-tetramethyl-2-[4-(dimethylamino)pyridin-2-yl]-5-(phenylamino)-2,5-dihydrofuran -2,3,4,5-tetracarboxylate (20)**



Yield: 16.8% (0.086g)

Physical state: brown sticky solid

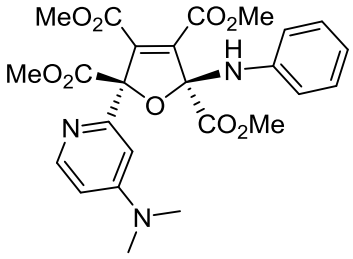
^{1}H NMR (600 MHz, $CDCl_3$) δ 8.33 (d, *J* = 5.9 Hz, 1H), 7.34 – 7.30 (m, 2H), 7.08 (d, *J* = 7.8 Hz, 2H), 7.03 (t, *J* = 7.3 Hz, 1H), 6.84 (d, *J* = 2.6 Hz, 1H), 6.57 (dd, *J* = 5.9, 2.6 Hz, 1H), 5.97 (s, 1H), 4.08 (s, 3H), 4.05 (s, 3H), 4.03 (s, 3H), 3.95 (s, 3H), 2.93 (s, 6H)

^{13}C NMR (151 MHz, $CDCl_3$) δ 168.69, 167.19, 162.01, 161.02, 155.11, 154.37, 148.21, 145.00, 141.48, 133.16, 128.65(2C), 120.44, 117.64(2C), 105.73, 103.39, 97.53, 93.16, 53.24, 52.45, 52.38, 52.28, 38.40(2C).

HRESI-MS: [$C_{25}H_{27}N_3O_9 + H^+$] calculated: 514.1820; found: 514.1821

Rf = 0.20 ethyl acetate/hexane (2/1)(*v*/*v*)

***cis*-2,3,4,5-tetramethyl 2-[4-(dimethylamino)pyridin-2-yl]-5-(phenylamino)-2,5-dihydrofuran-2,3,4,5-tetracarboxylate (20)**



Yield: 7.0% (0.036g)

Physical state: brown sticky solid

^{1}H NMR (600 MHz, $CDCl_3$) δ 8.08 (d, *J* = 5.9 Hz, 1H), 7.20 (t, *J* = 7.9 Hz, 2H), 7.03 (d, *J* = 7.8 Hz, 2H), 6.92 (t, *J* = 7.4 Hz, 1H), 6.85 (d, *J* = 2.5 Hz, 1H), 6.40 (dd, *J* = 5.9, 2.6 Hz, 1H), 5.54 (s, 1H), 3.82 (s, 3H), 3.72 (s, 3H), 3.57 (s, 3H), 3.55 (s, 3H), 2.98 (s, 6H)

^{13}C NMR (151 MHz, $CDCl_3$) δ 168.11, 167.42, 161.91, 161.29, 156.81, 154.46, 148.06, 144.75, 141.19, 133.85, 128.47(2C), 121.35, 119.19(2C), 105.76, 103.49, 98.98, 93.51, 52.83, 52.45, 52.33, 52.27, 38.70(2C).

Rf = 0.28 2/1 ethyl acetate/hexane (1/1)(*v*/*v*)

HRESI-MS: [$C_{25}H_{27}N_3O_9 + H^+$] calculated: 514.1820; found: 514.1819

6. Copies of ^{1}H and ^{13}C NMR spectra

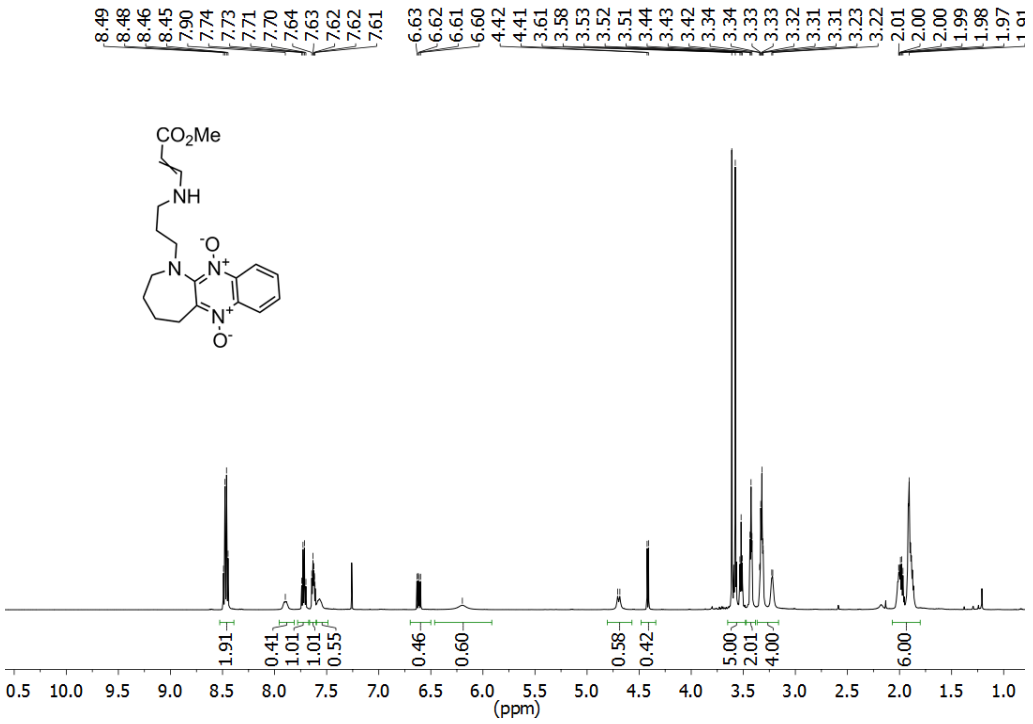


Figure S13. ^{1}H NMR spectrum of compound **19**.

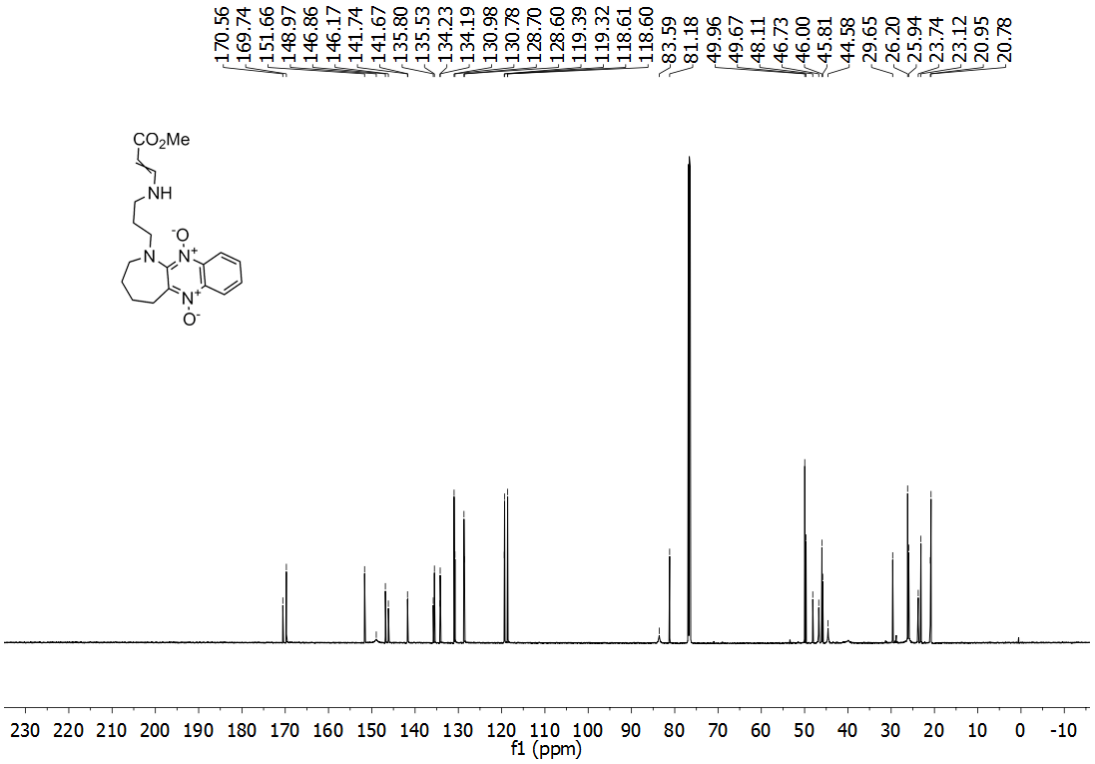


Figure S14. ^{13} C NMR spectrum of compound **19**.

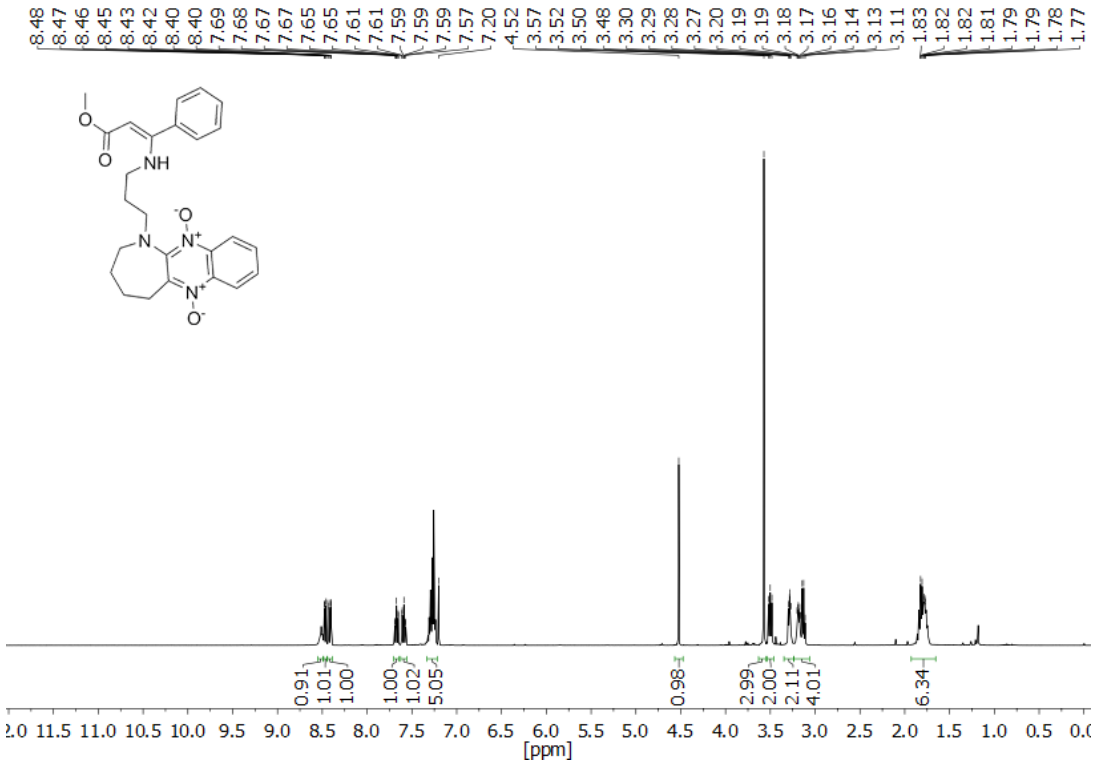


Figure S15. [1]H NMR spectrum of compound **19b**.

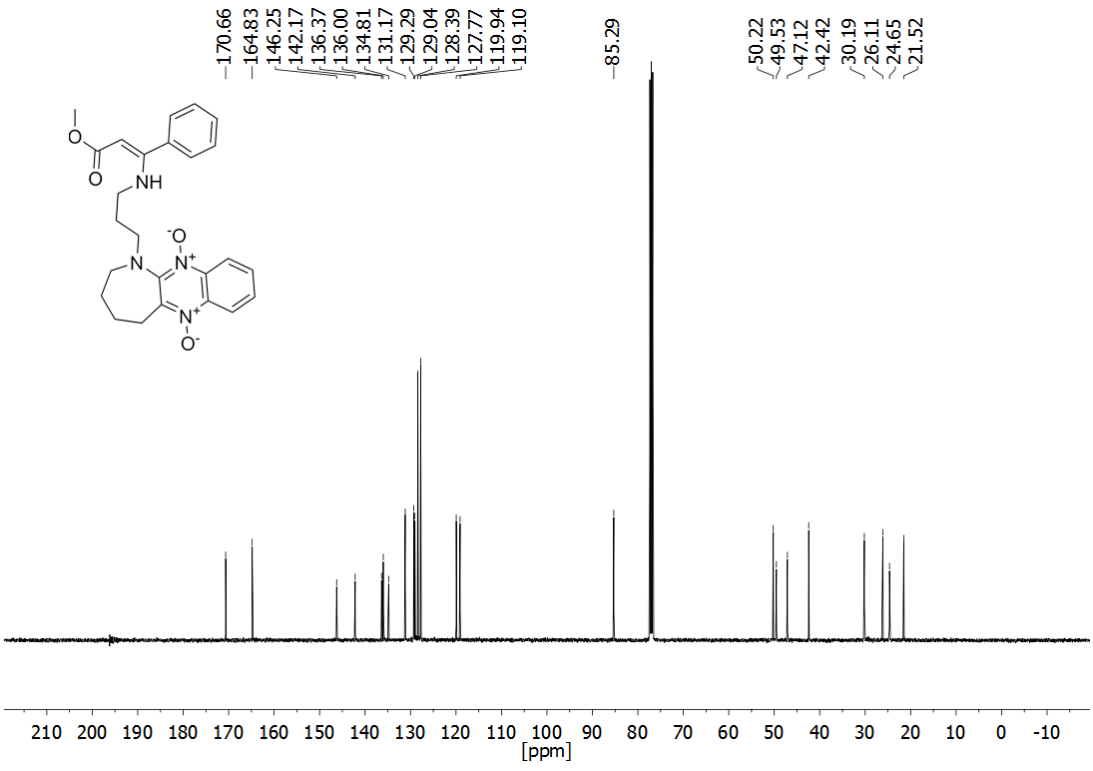


Figure S16. [13] C NMR spectrum of compound **19b**.

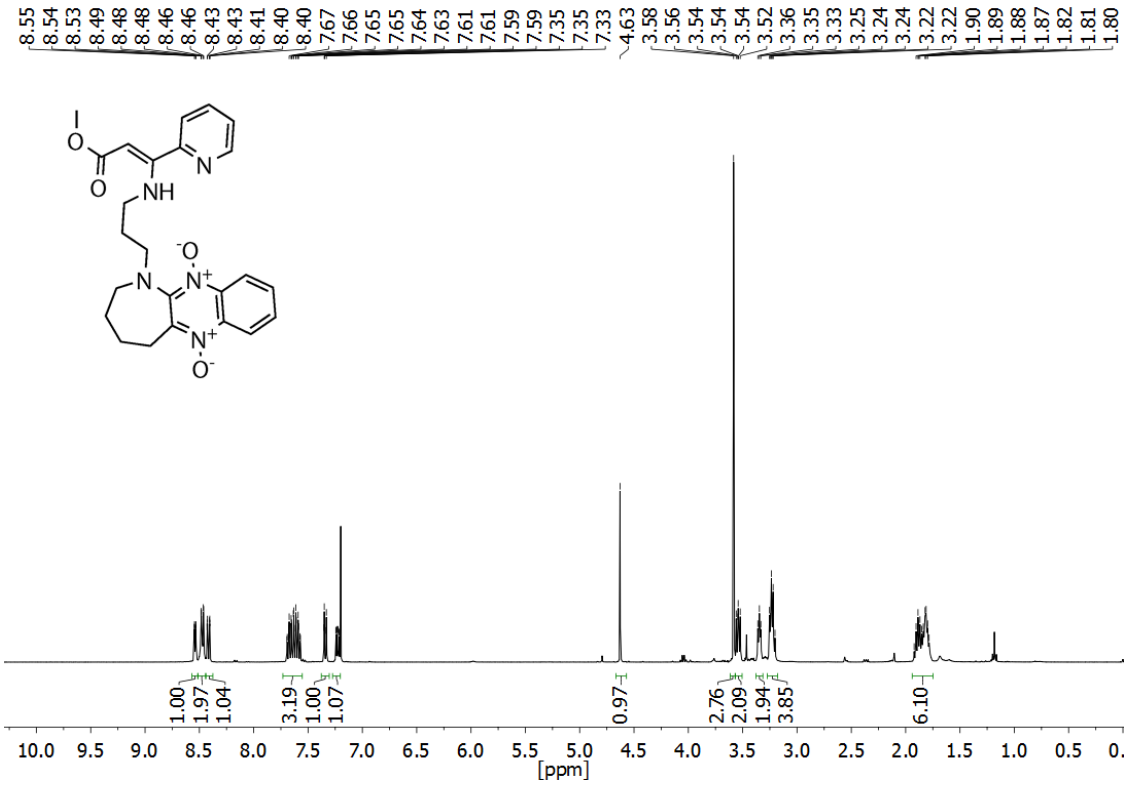


Figure S17. [1]H NMR spectrum of compound **19e**.

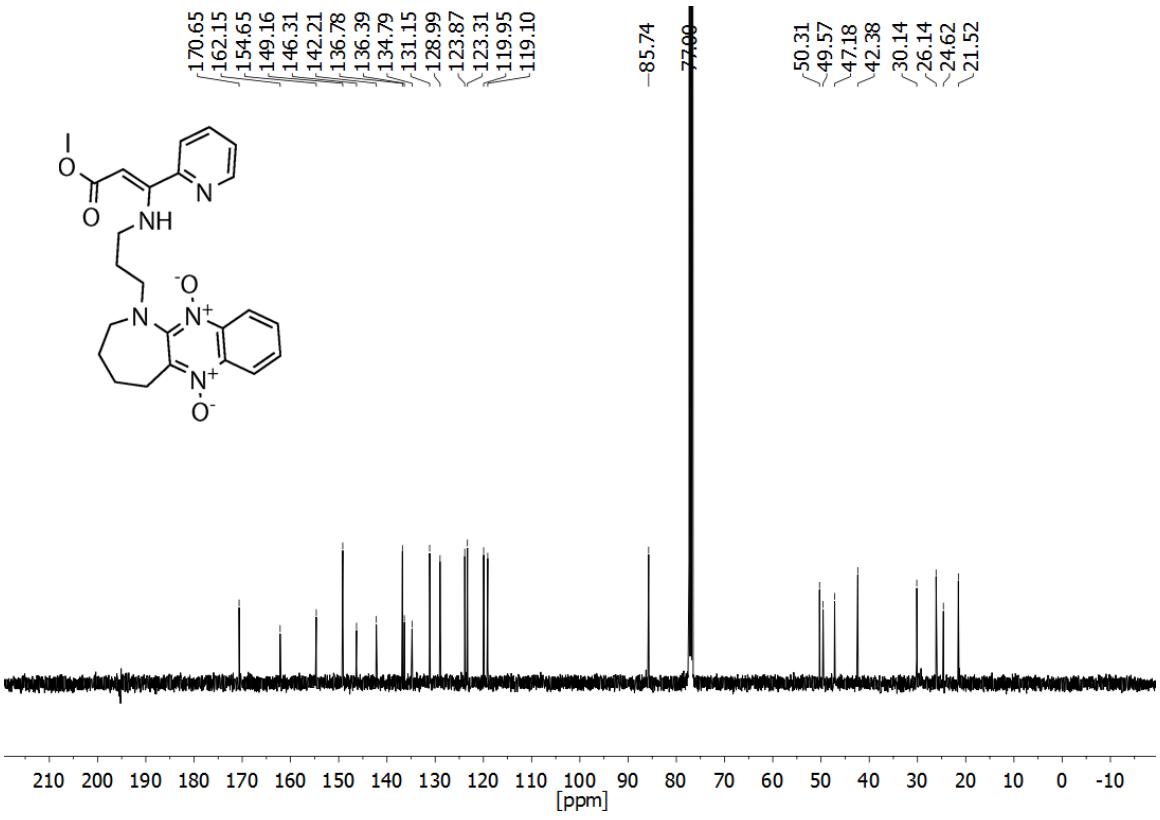


Figure S18. [13] C NMR spectrum of compound **19e**.

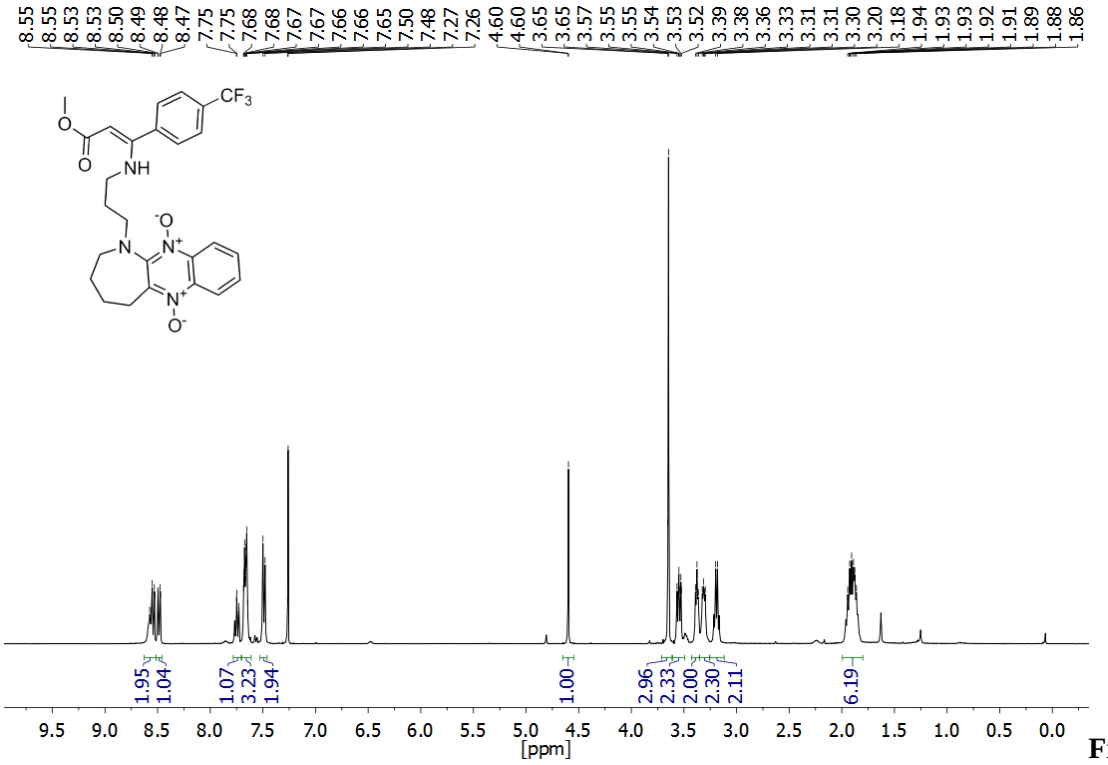


Figure S19. ^{1}H NMR spectrum of compound **19c**.

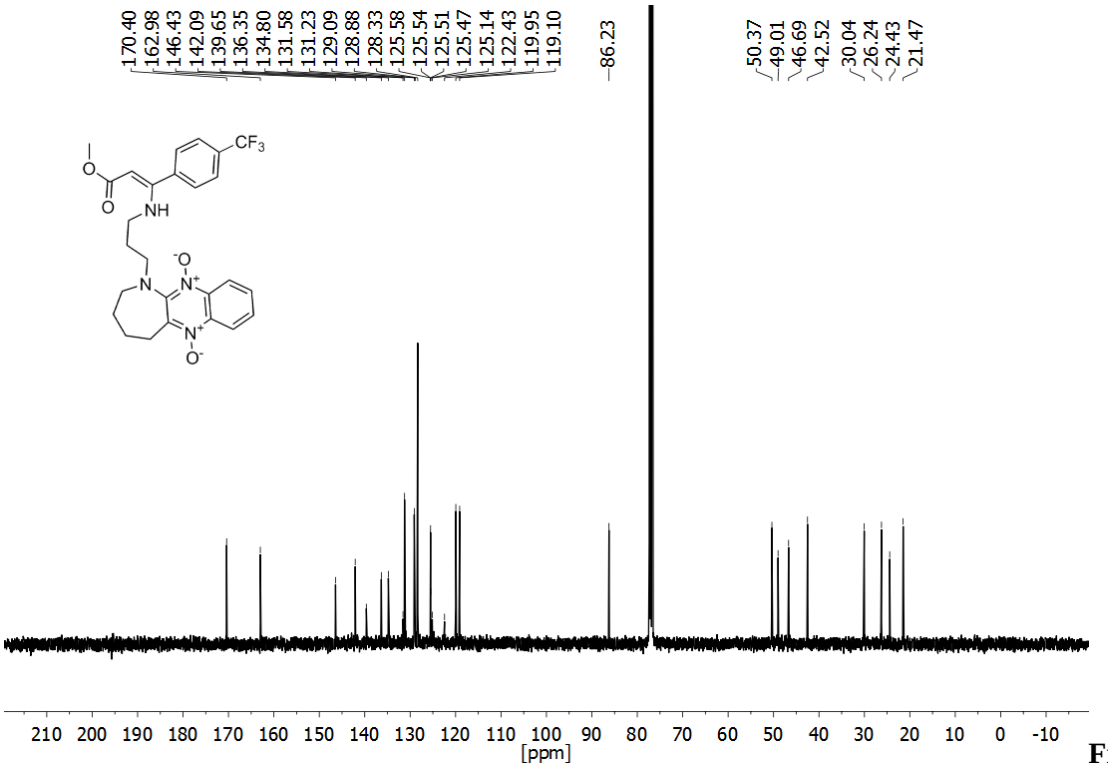


Figure S20. ^{13} C NMR spectrum of compound **19c**.

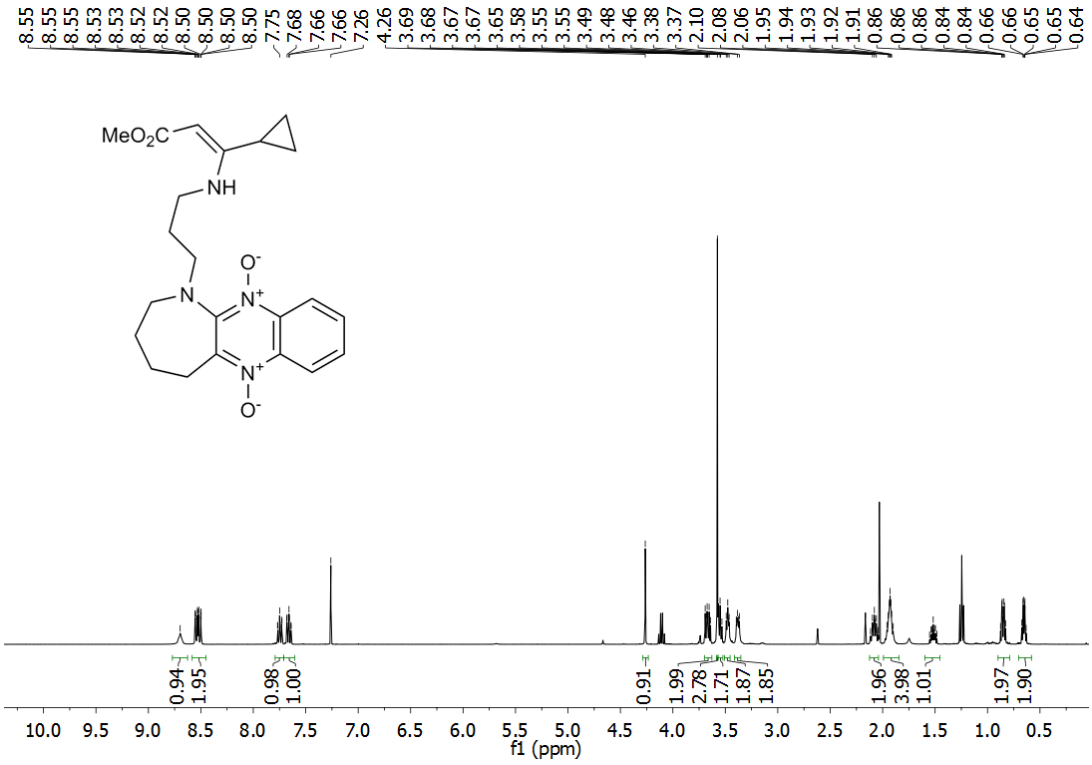


Figure S31. [1]H NMR spectrum of compound **19g**.

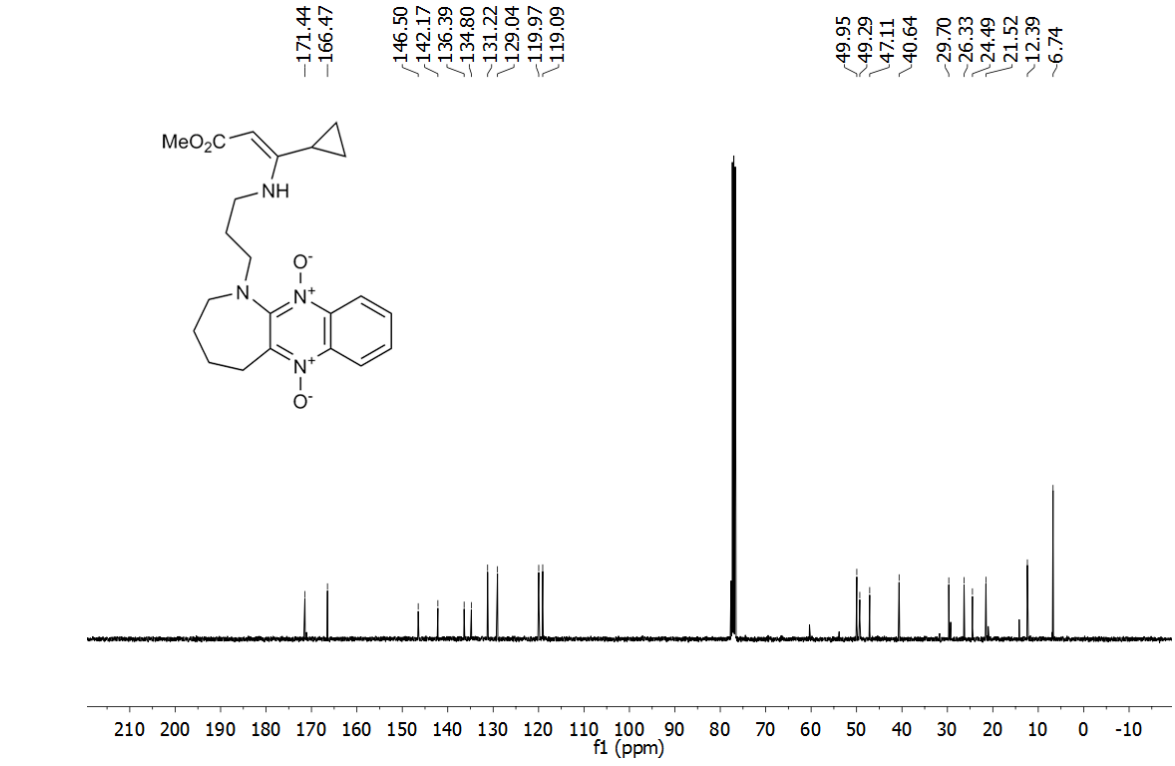


Figure S32. [13] C NMR spectrum of compound **19g**.

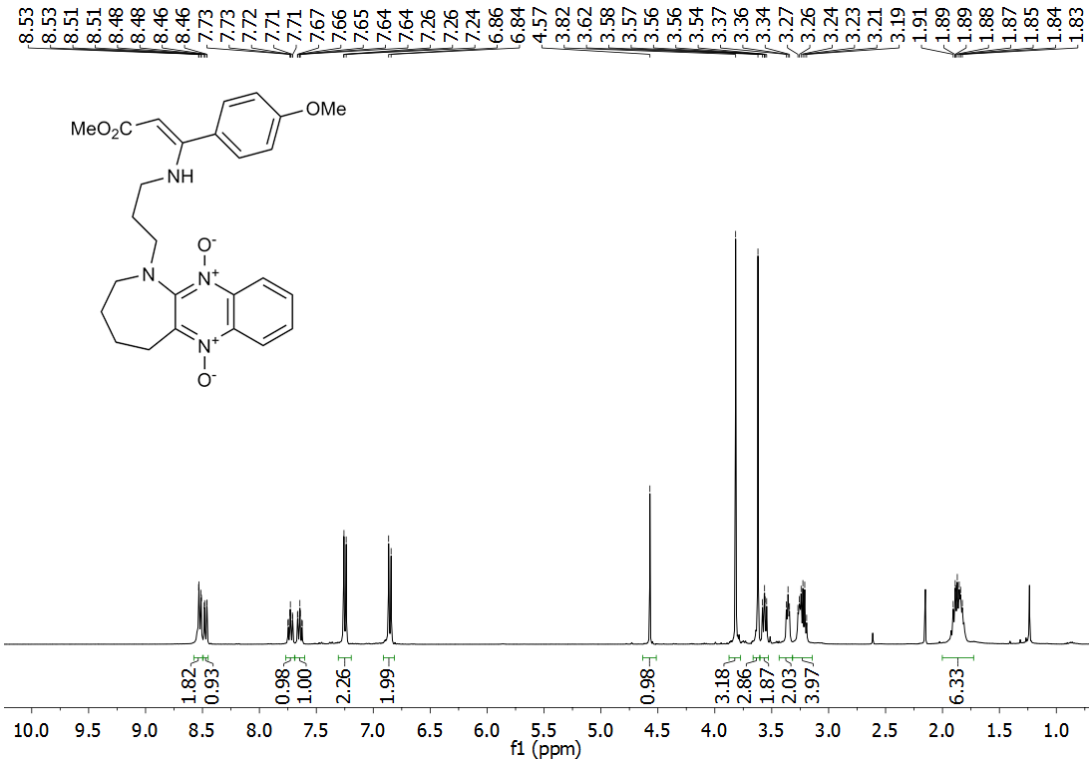


Figure S33. [1]H NMR spectrum of compound **19d**.

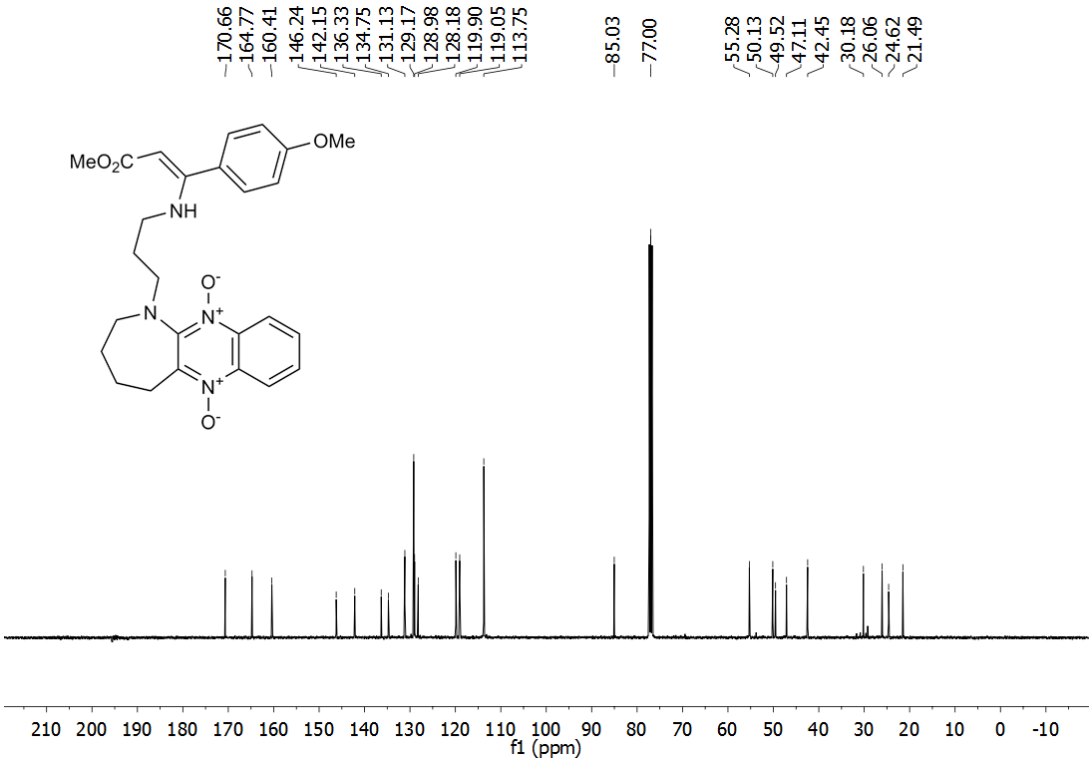


Figure S34. [13] C NMR spectrum of compound **19d**.

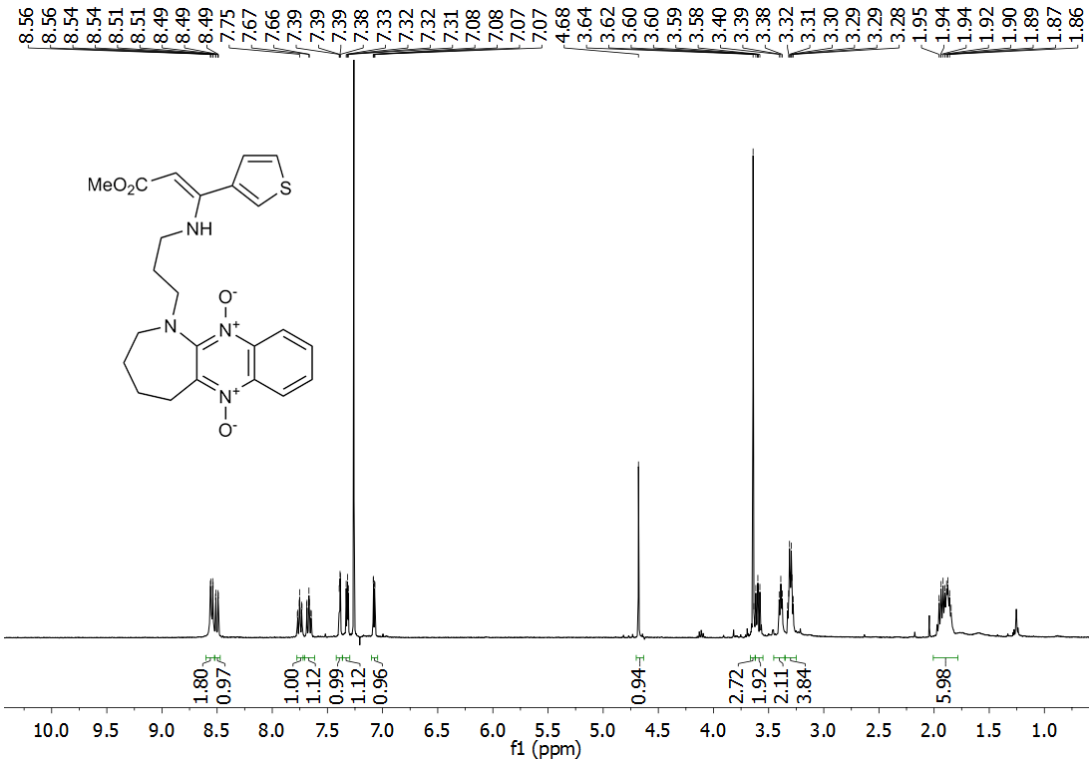


Figure S35. [1]H NMR spectrum of compound **19f**.

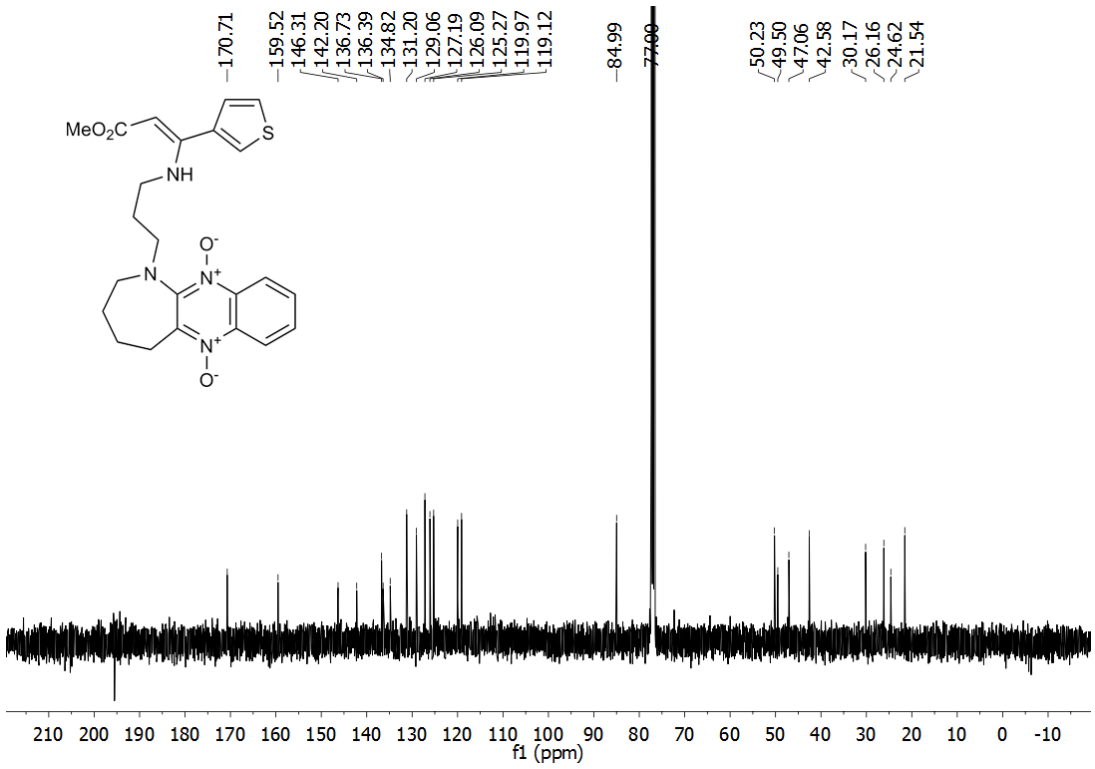


Figure S36. [13] C NMR spectrum of compound **19f**.

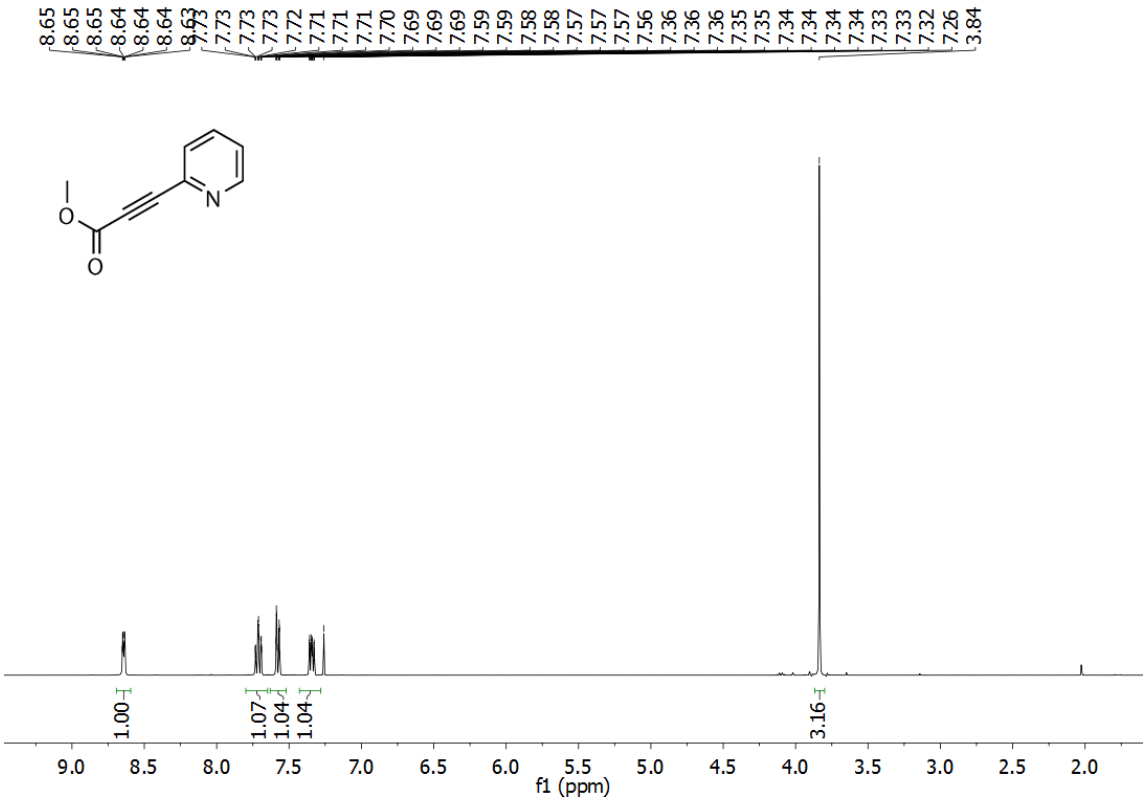
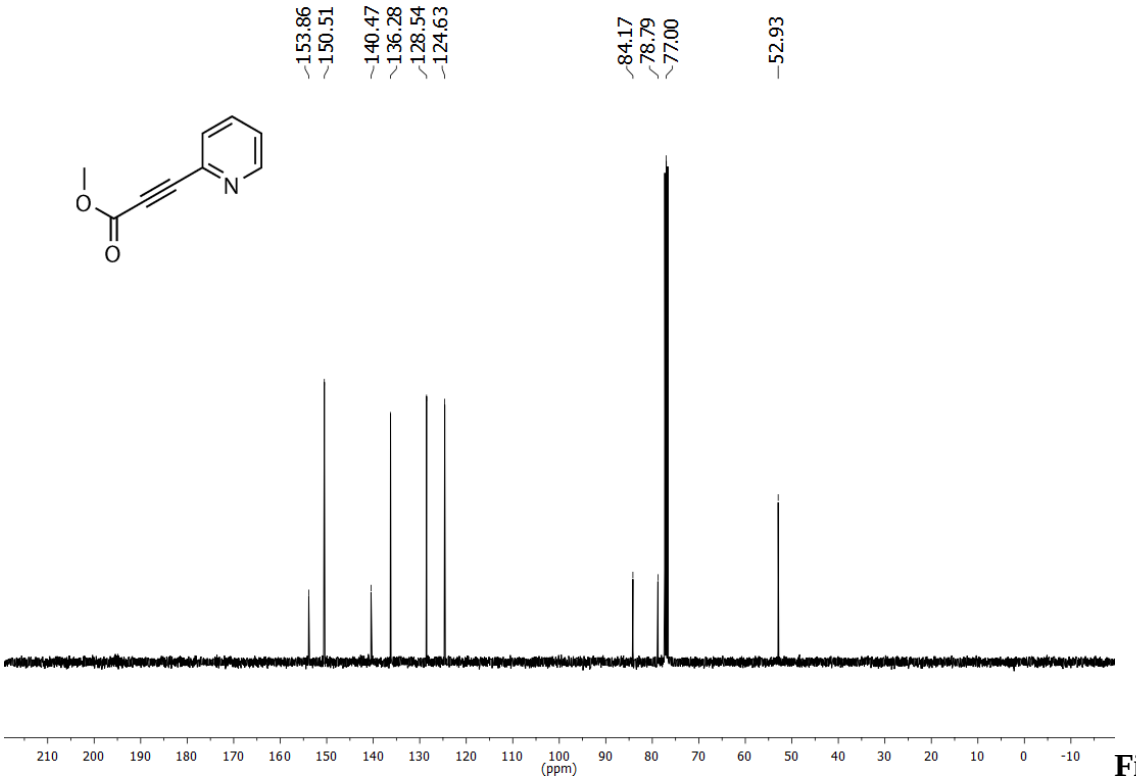


Figure S21. [1]H NMR spectrum of compound **S4**.



Figure

Figure S22. [13] C NMR spectrum of compound **S4**.

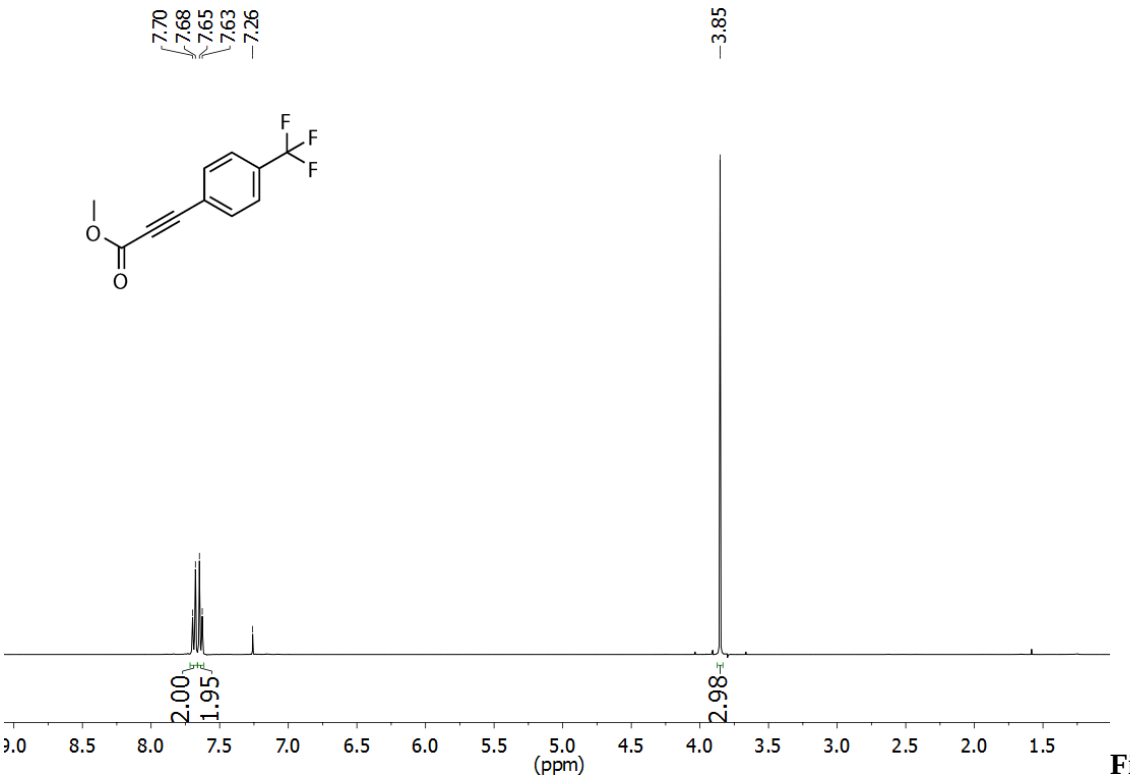


Figure S23. [1]H NMR spectrum of compound **S5**.

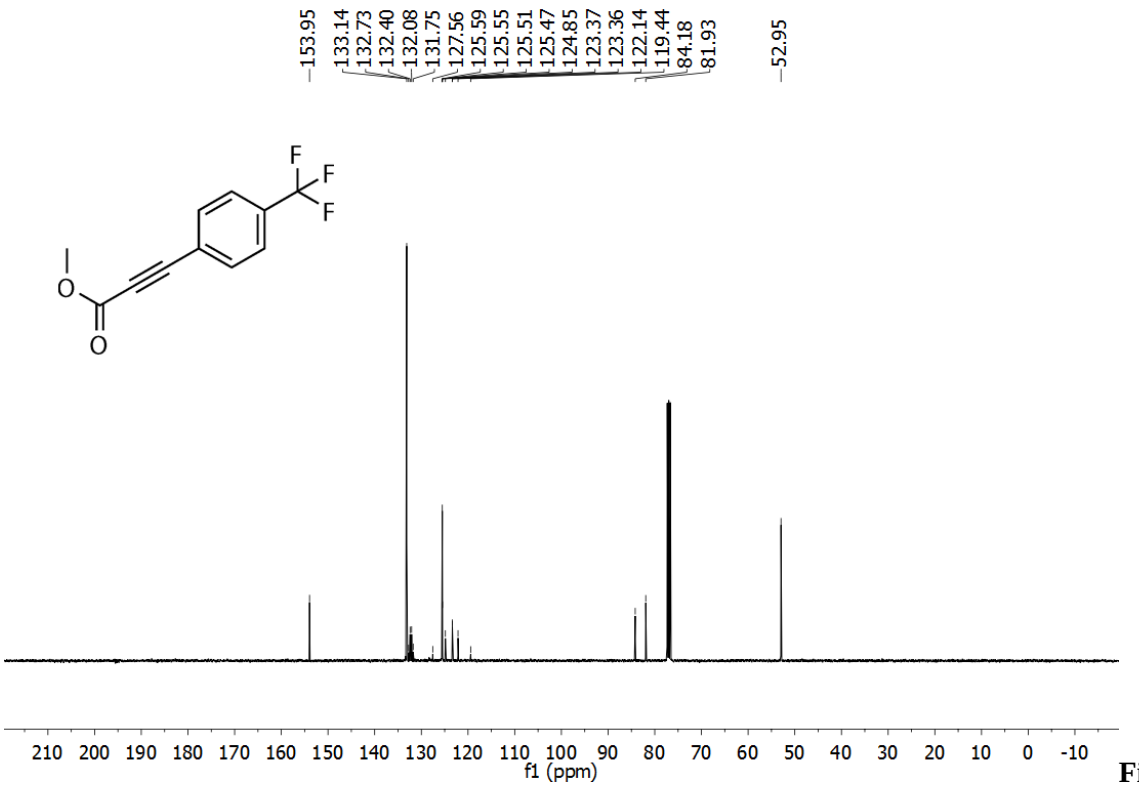


Figure S24. [13] C NMR spectrum of compound **S5**.

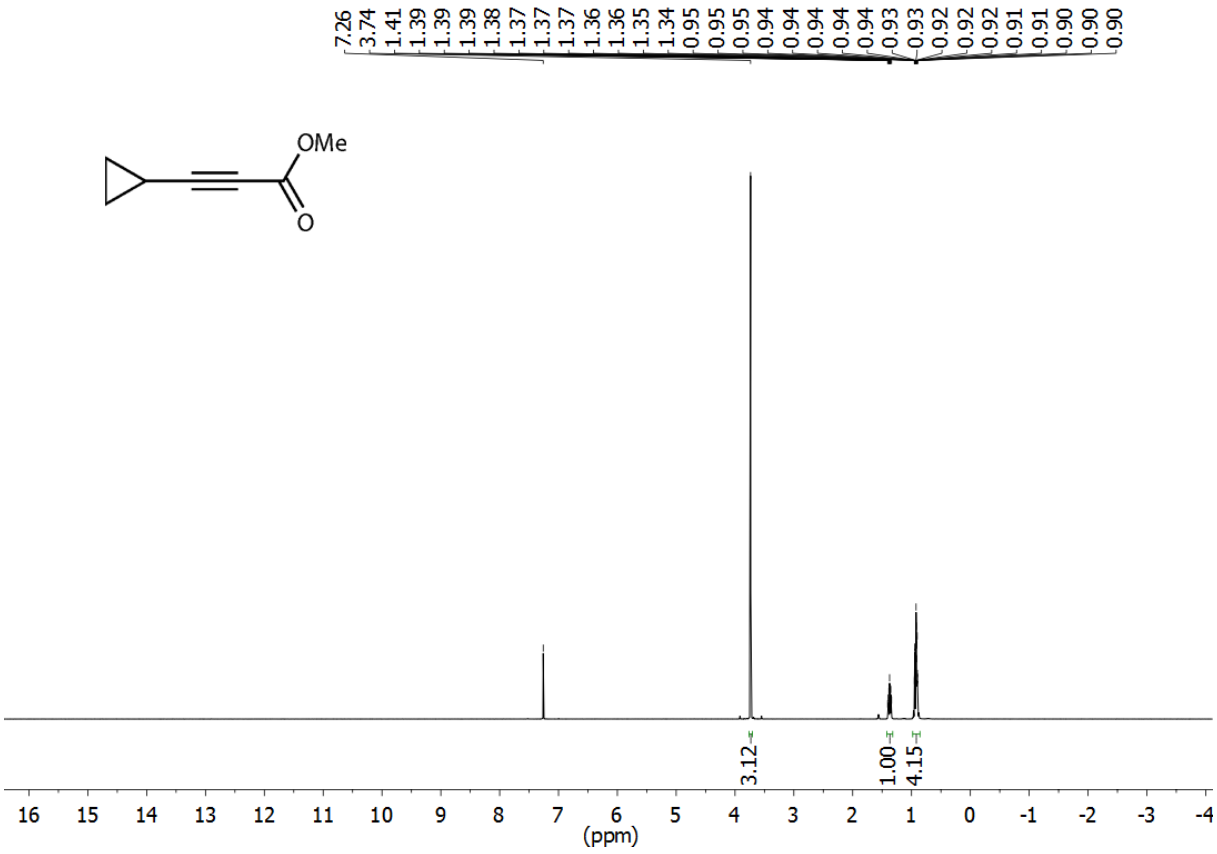


Figure S25. [1]H NMR spectrum of compound **S1**.

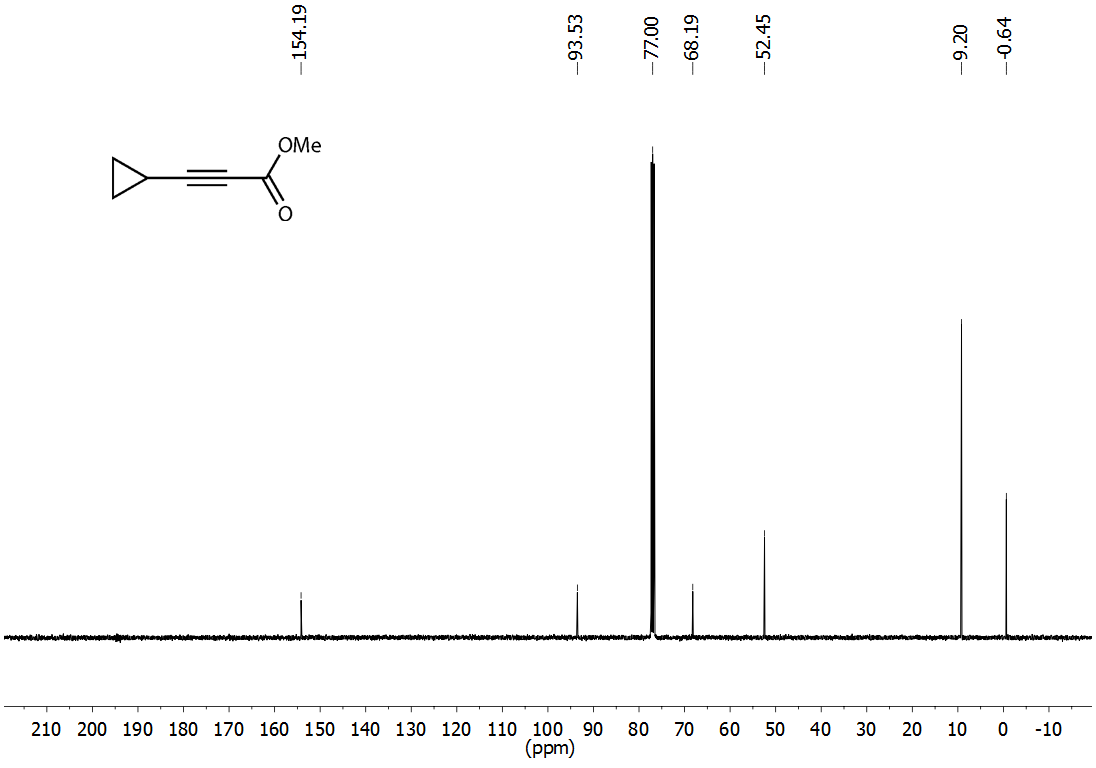


Figure S26. [13] C NMR spectrum of compound **S1**.

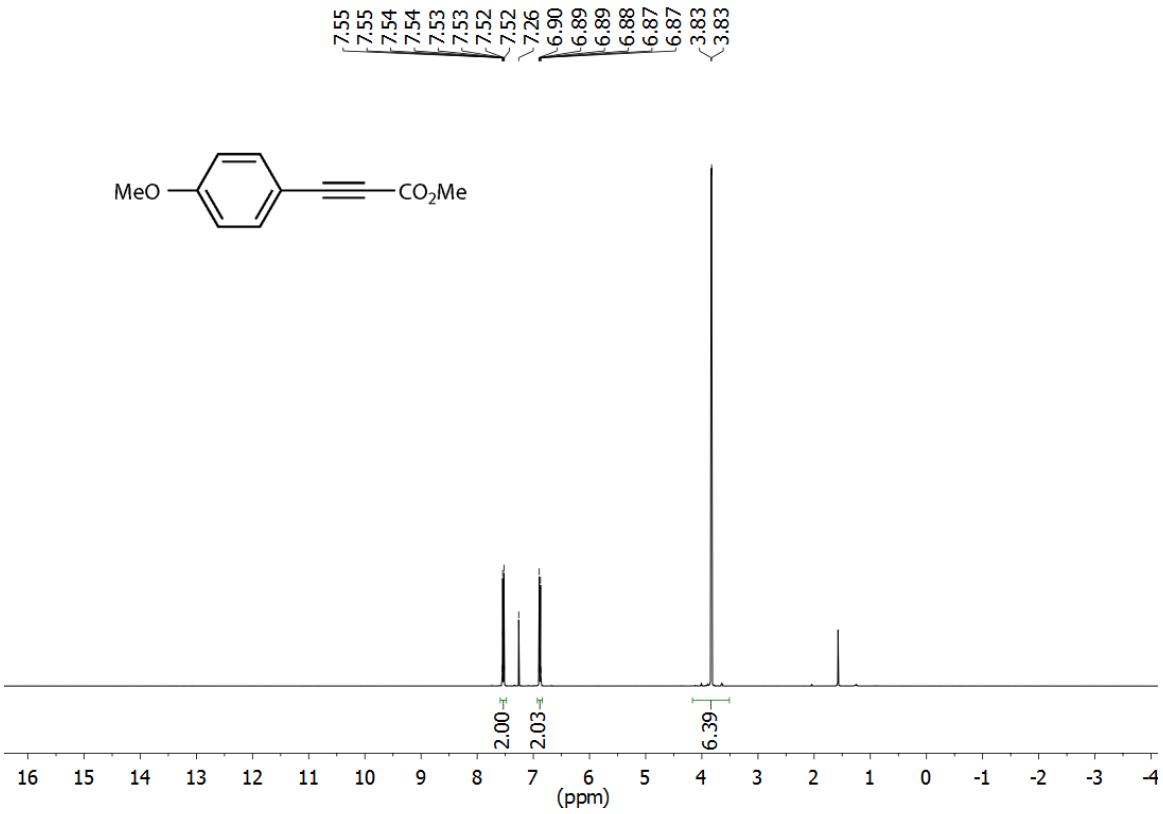


Figure S27. [1]H NMR spectrum of compound **S2**.

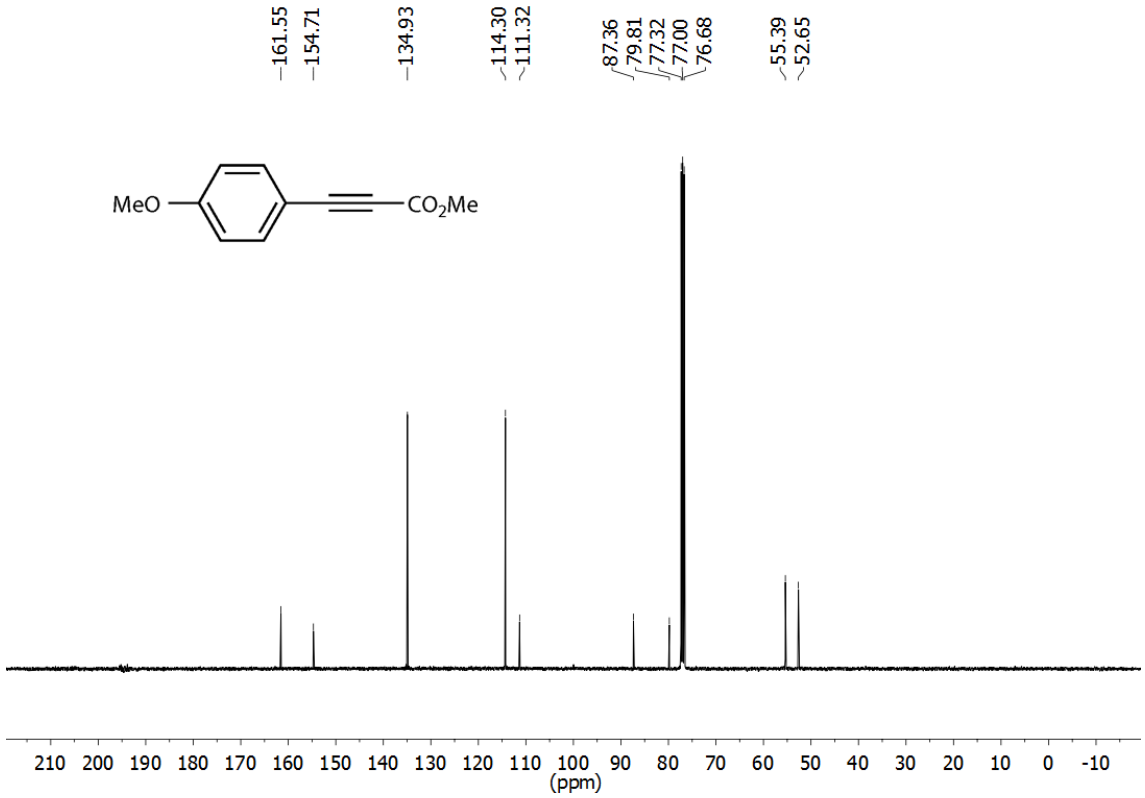


Figure S28. [13] C NMR spectrum of compound **S2**.

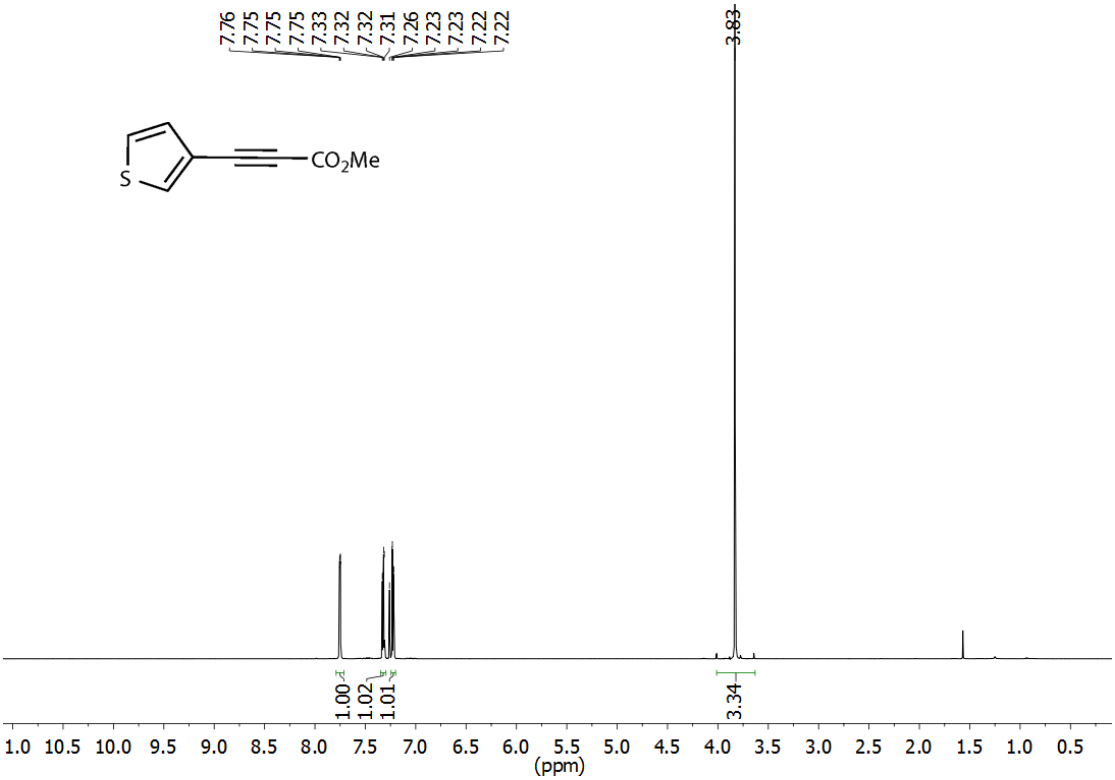


Figure S29. [1]H NMR spectrum of compound **S3**.

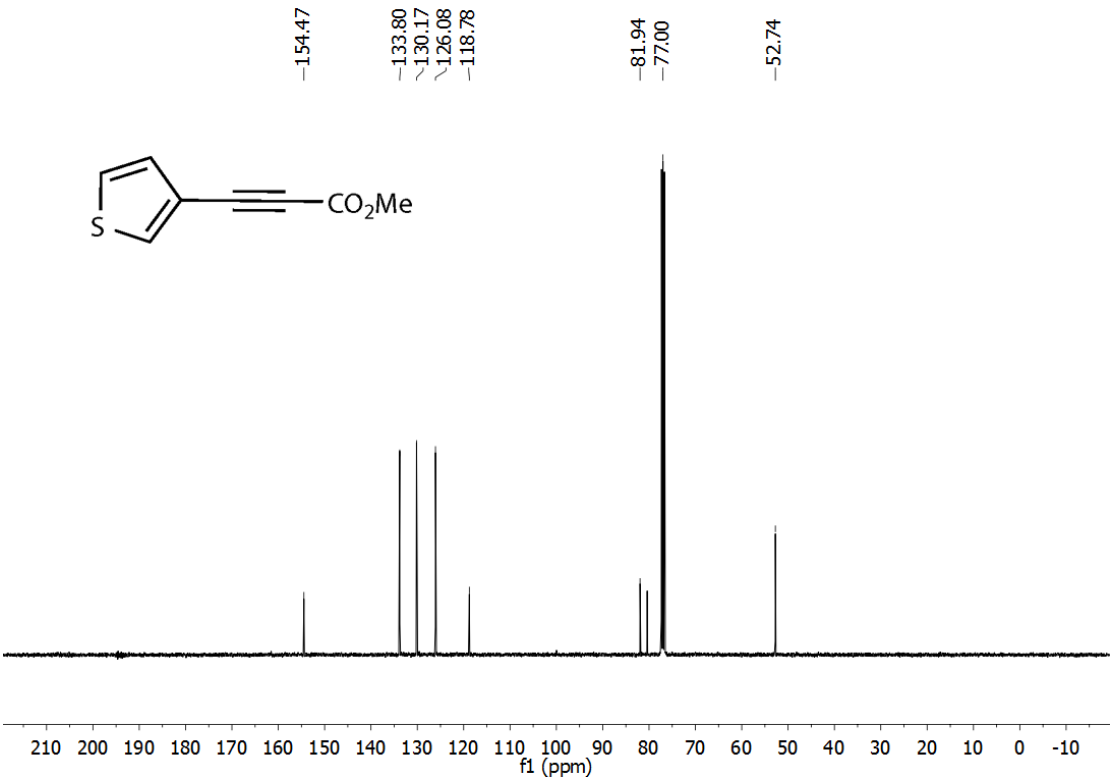


Figure S30. [13] C NMR spectrum of compound **S3**.

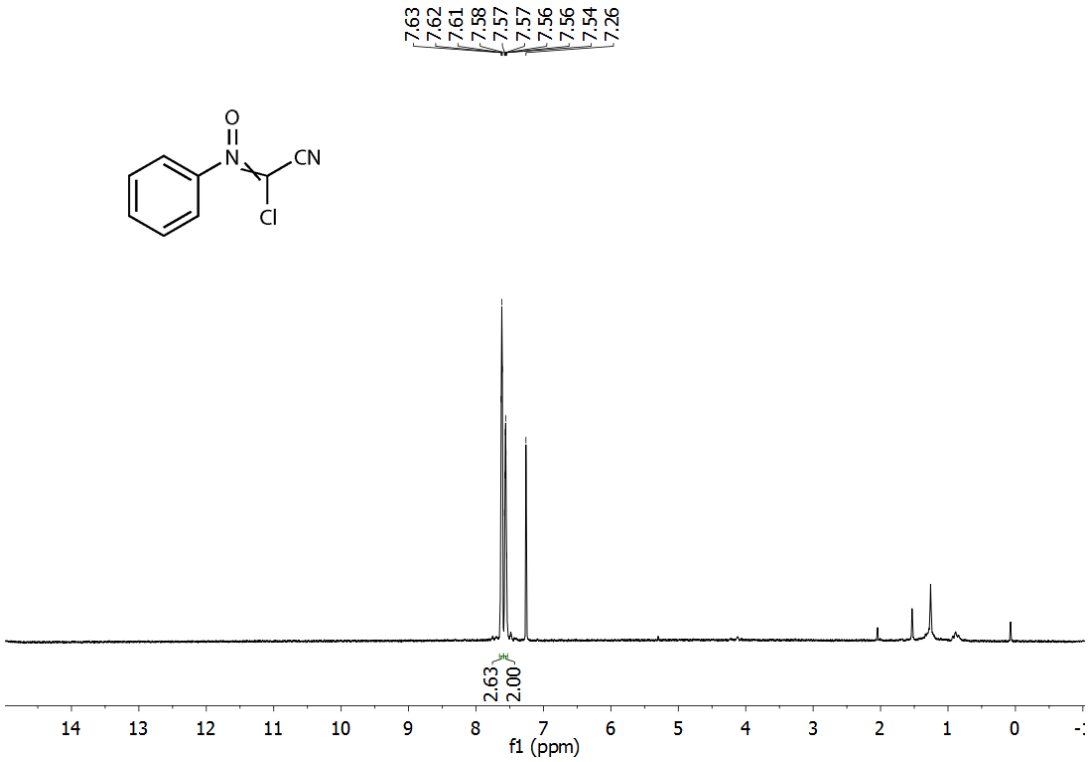


Figure S37. [1]H NMR spectrum of compound **21**.

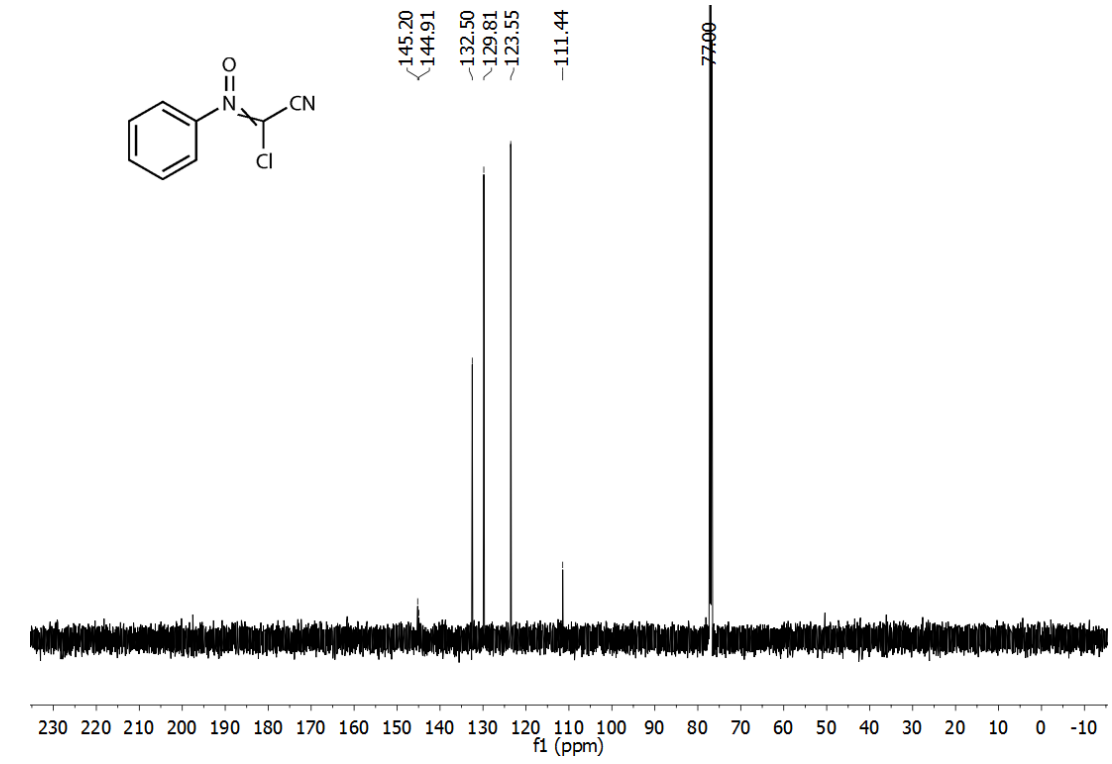


Figure S38. [13] C NMR spectrum of compound **21**.

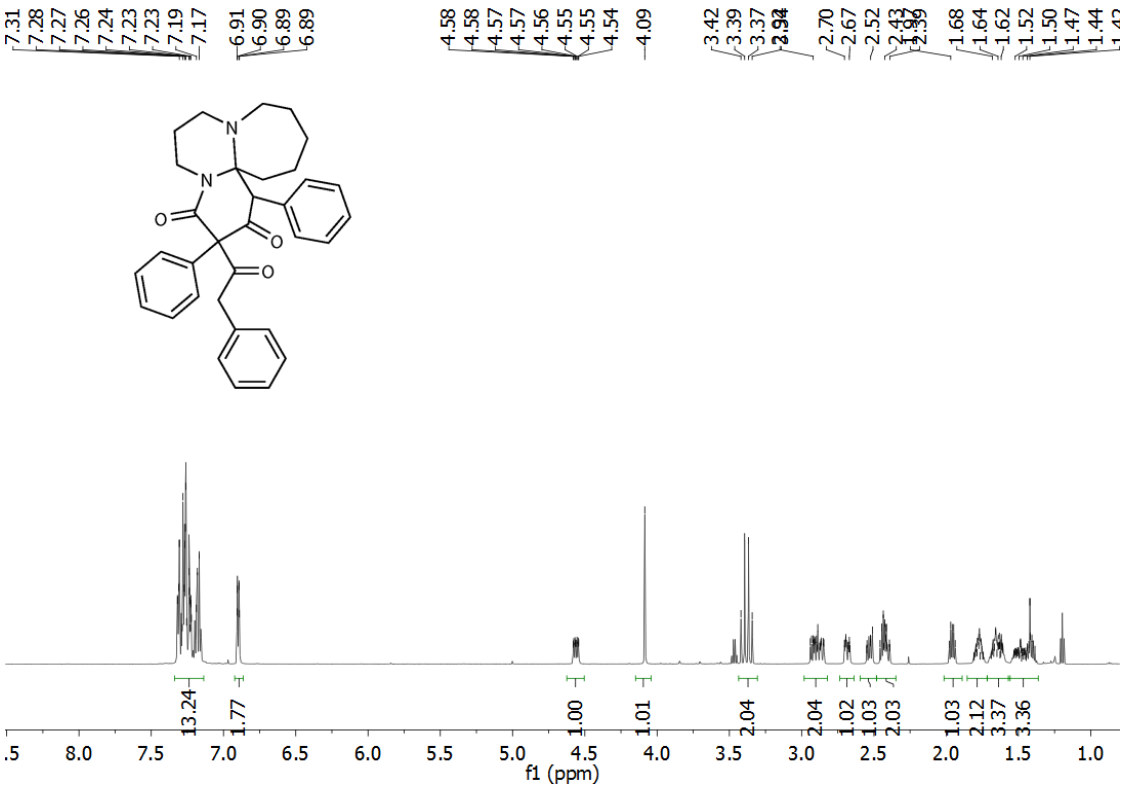


Figure S39. [1]H NMR spectrum of compound **22**.

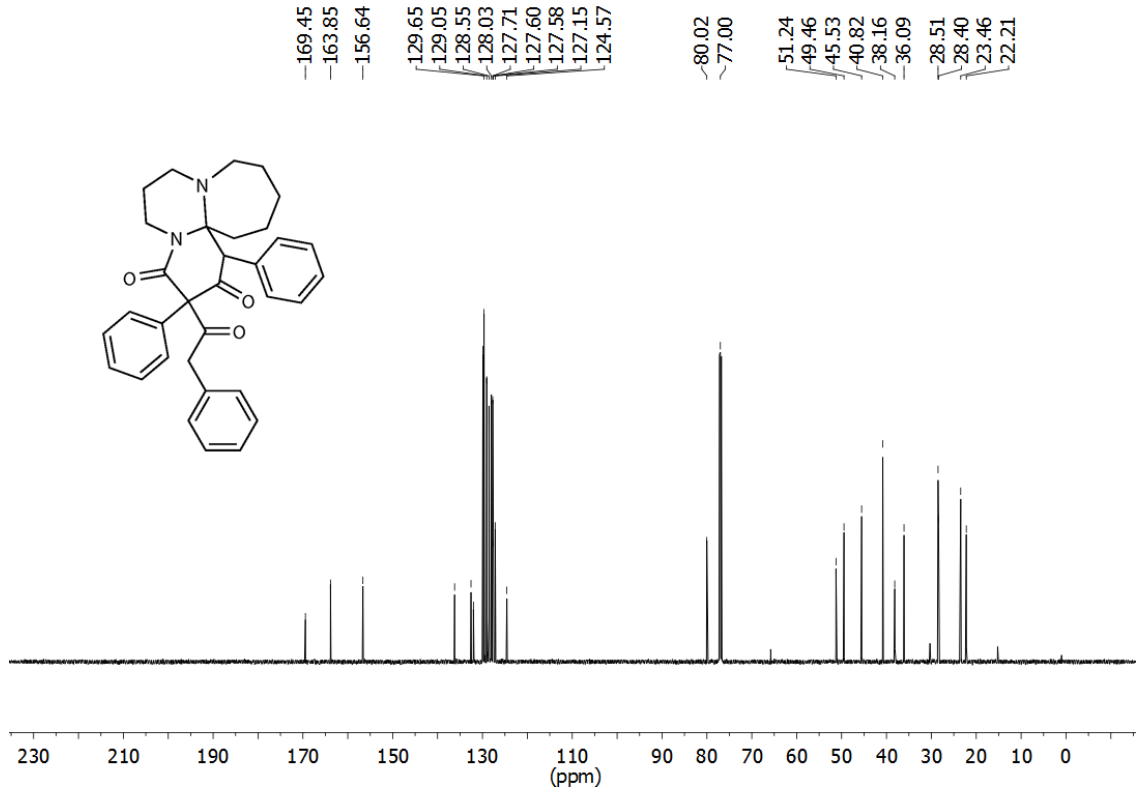


Figure S40. [13] C NMR spectrum of compound **22**.

7. Structural assignments.

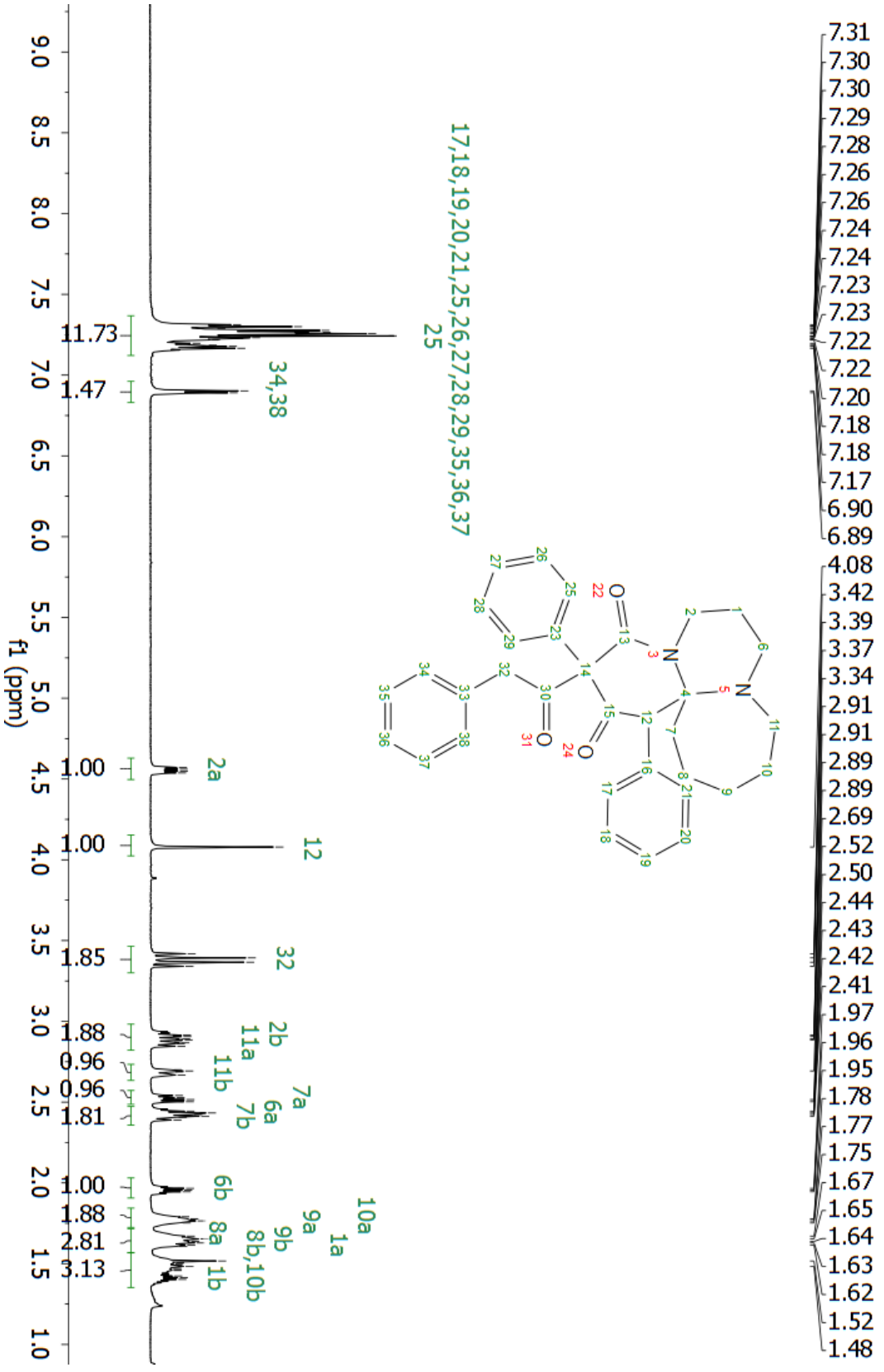


Figure S43. Proton assignments for compound **22** and atom numeration.

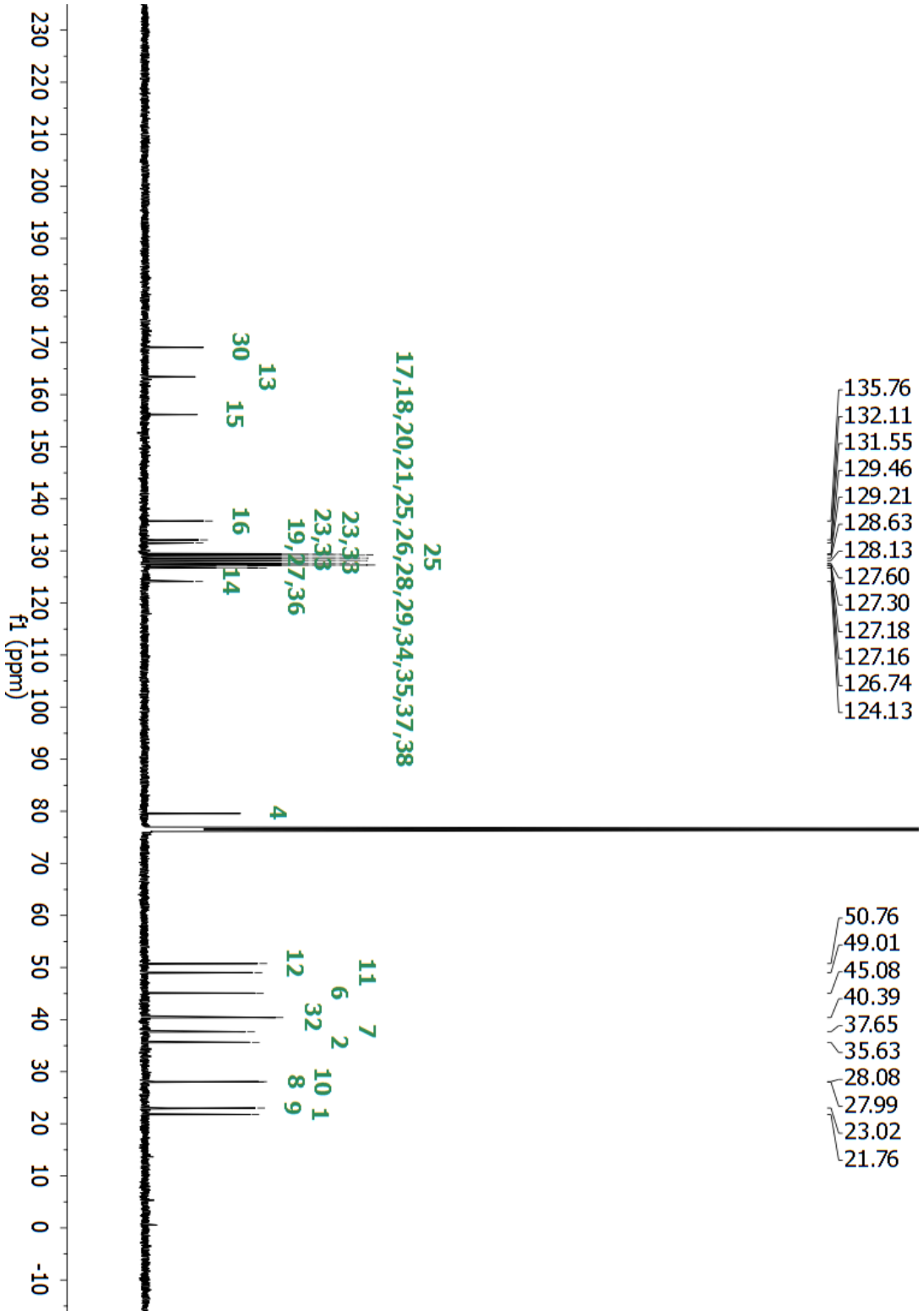


Figure S44. Carbon assignments for compound **22**.

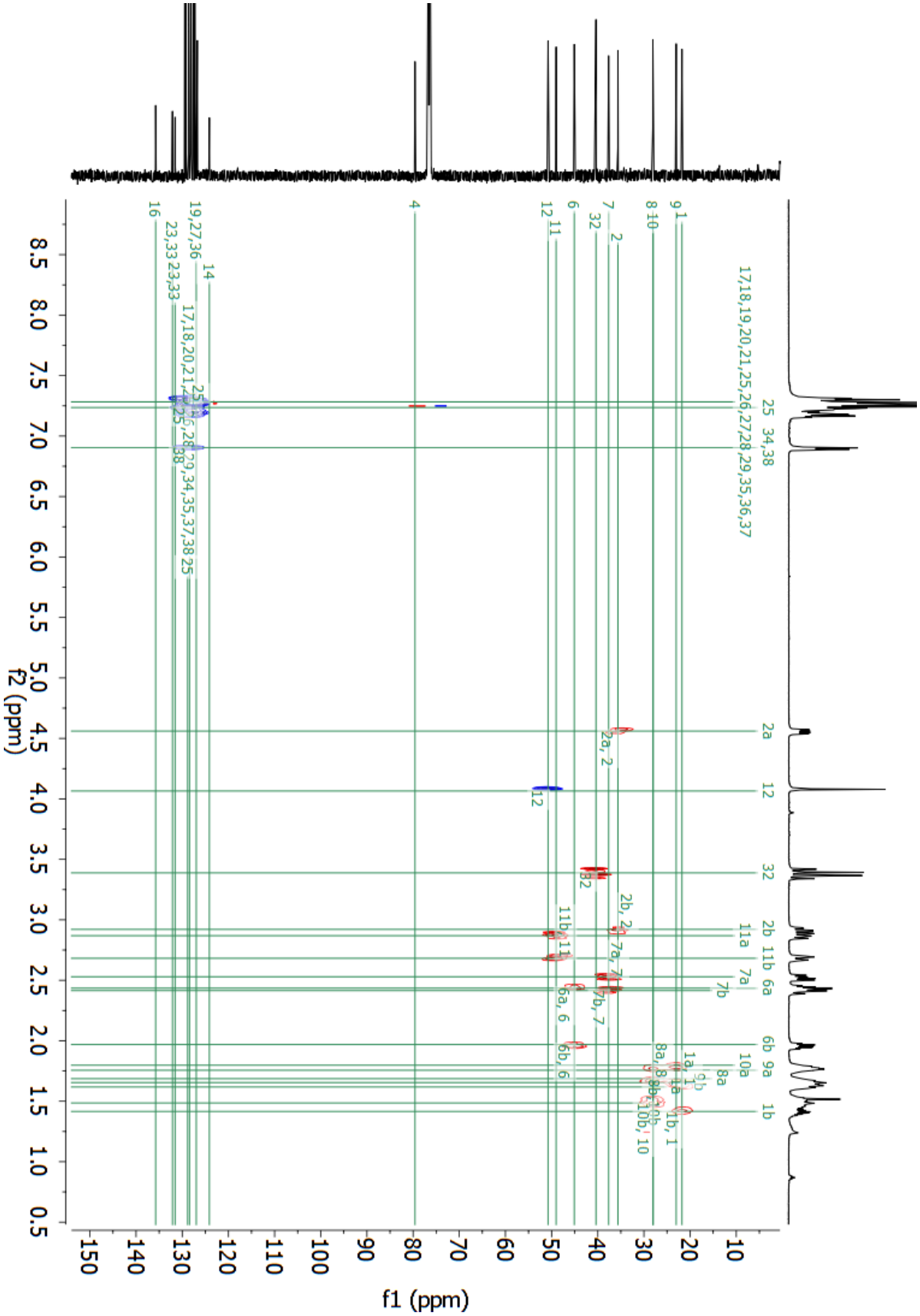


Figure S45. Phase sensitive HSQC spectra of compound **22**.

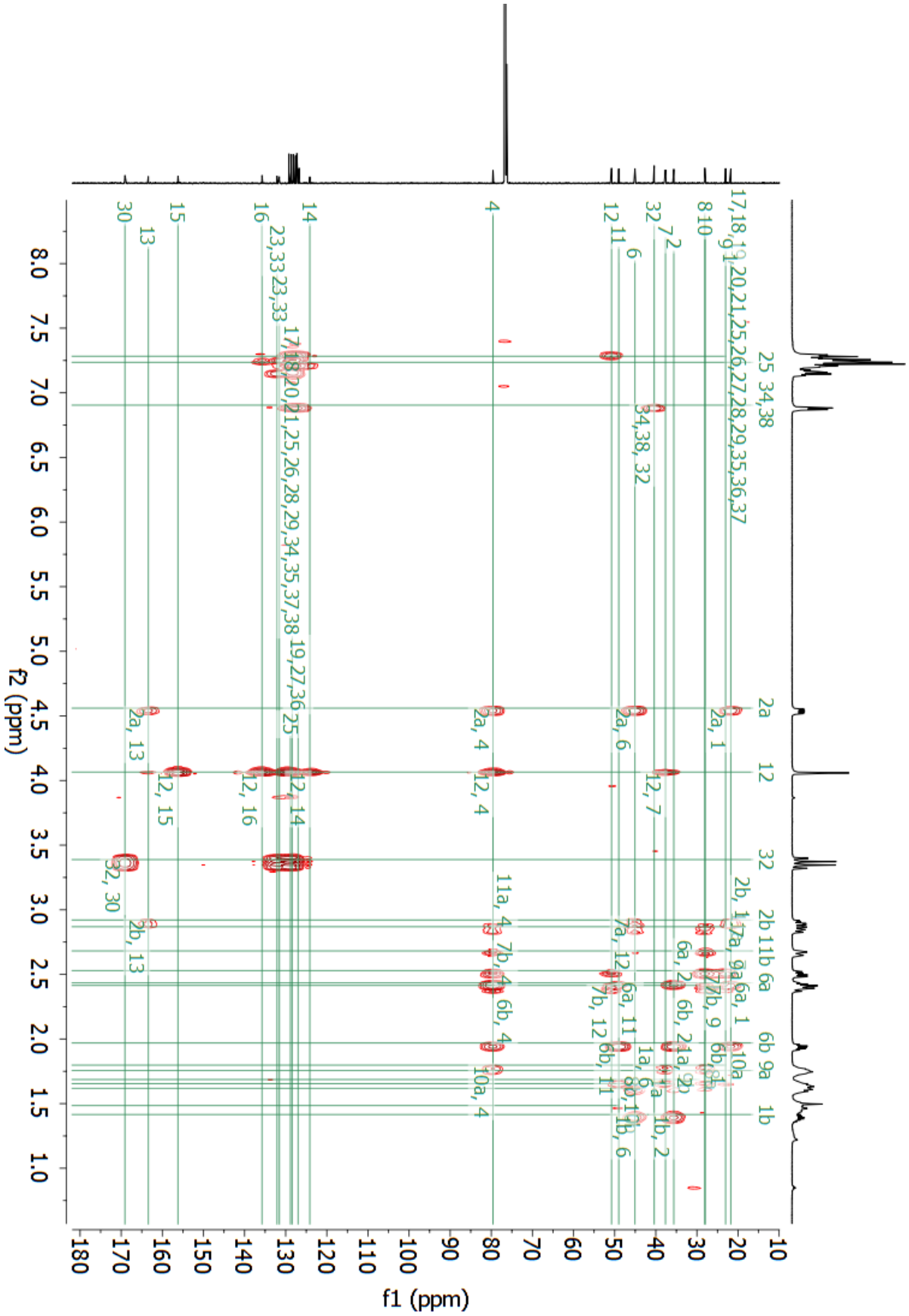


Figure S46. HMBC spectrum of compound **22**.

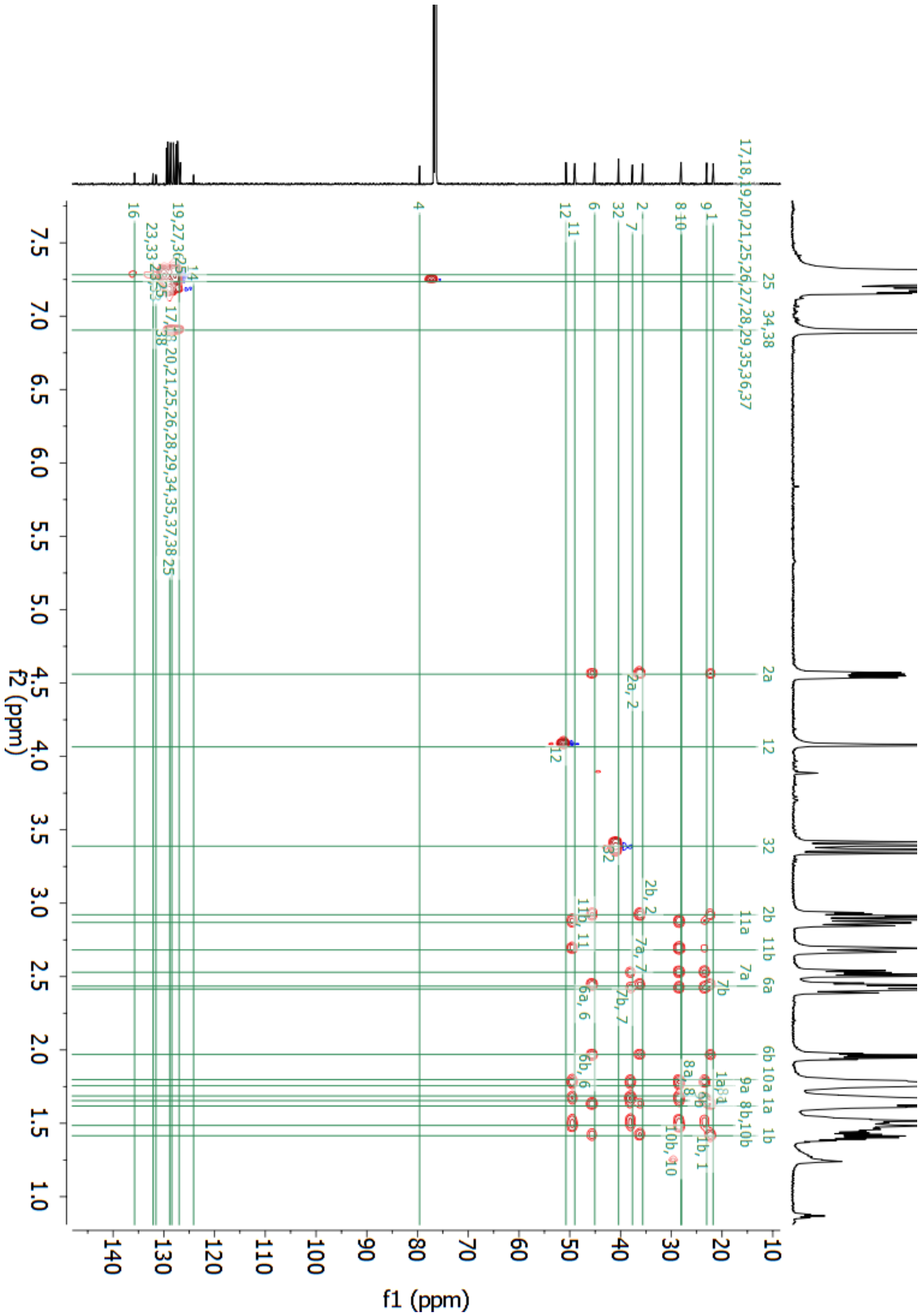


Figure S47. HSQC-TOCSY spectrum of compound **22**.

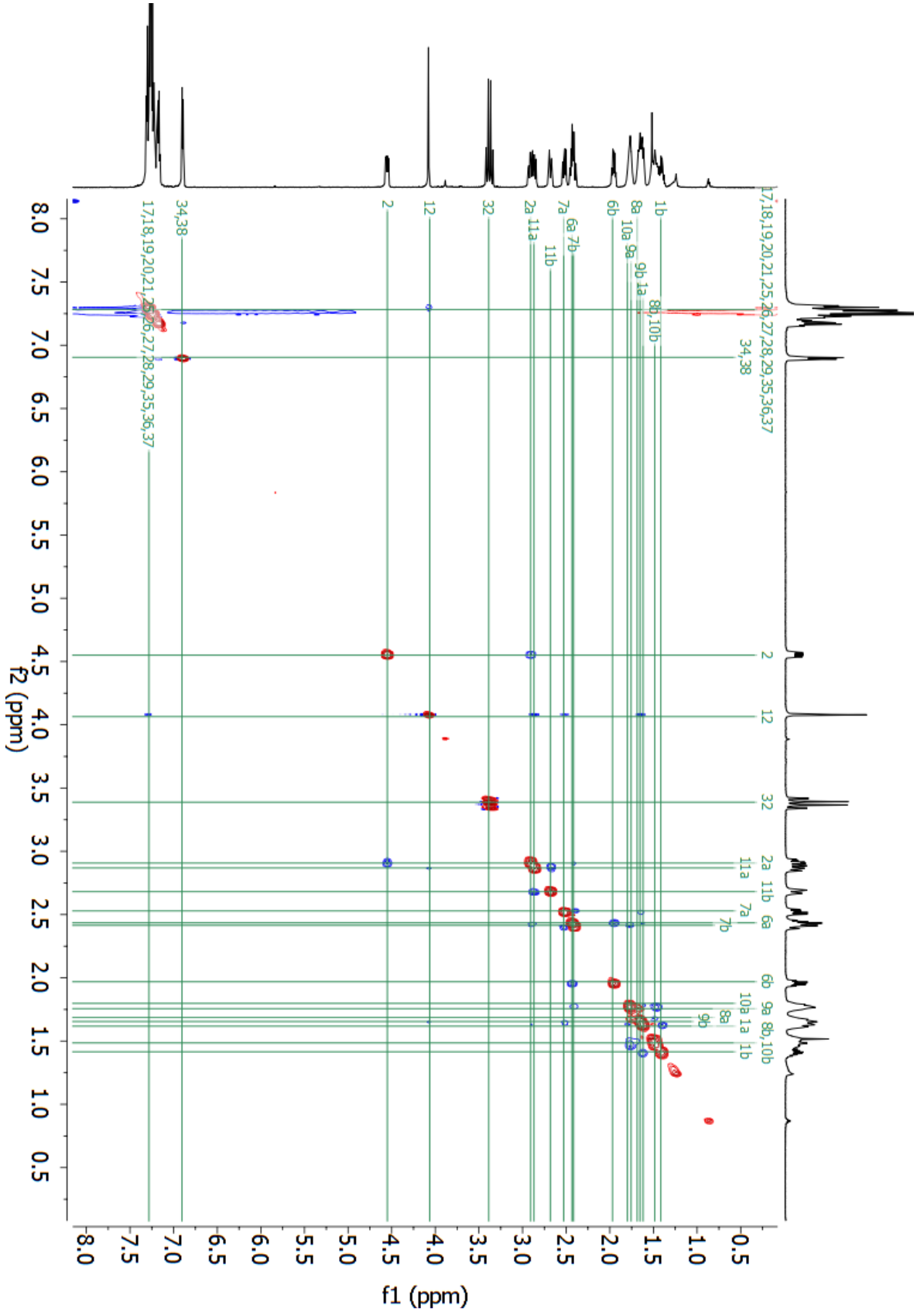


Figure S48. ^{1}H –^{1}H ROESY spectrum of compound **22**.

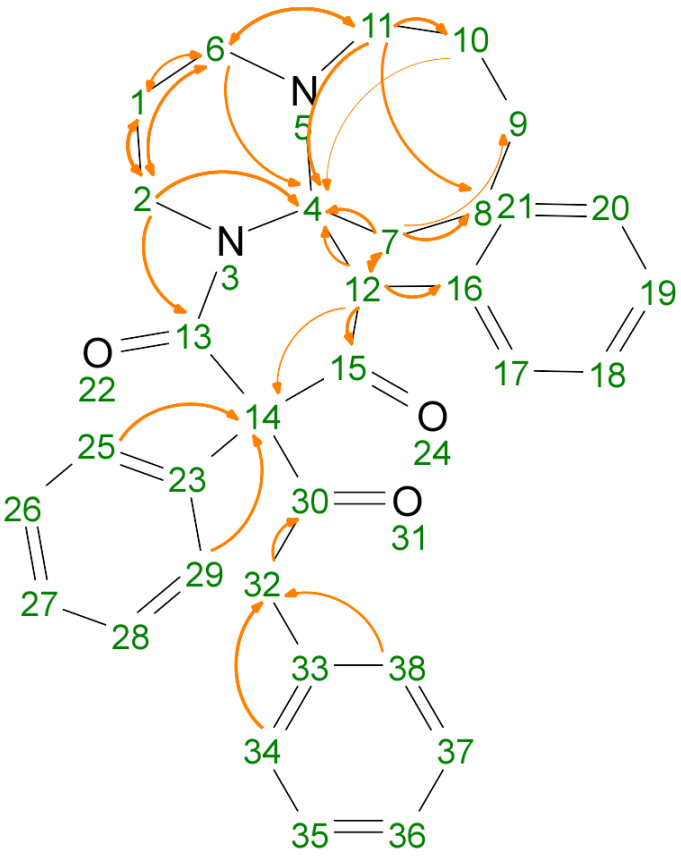


Figure S49. Key ^{1}H-^{13}C correlations from HMBC spectrum confirming structure of **22**.

Table S2. Summary of structural assignments for compound **22**.

Atom	Chemical Shift	HSQC correlations	HMBC correlations
1 C	21.76	1a, 1b	6a, 2a, 2b, 6b
Ha	1.62	1	2, 6
Hb	1.42	1	2, 6
2 C	35.63	2a, 2b	6a, 1b, 1a, 6b
Ha	4.56	2	1, 6, 4, 13
Hb	2.90..2.95	2	1, 6, 4, 13
3 N			
4 C	79.59		11a, 11b, 7a, 7b, 6a, 2a, 2b, 6b, 10a, 12
5 N			
6 C	45.08	6a, 6b	11a, 1b, 1a, 2a, 2b
Ha	2.43	6	1, 2, 4, 11
Hb	1.94..2.00	6	1, 2, 11, 4
7 C	37.65	7a, 7b	12
Ha	2.53	7	8, 9, 4, 12
Hb	2.41	7	8, 9, 4, 12
8 C	28.08	8a, 8b	11a, 11b, 7a, 7b
Ha	1.67..1.70	8	
Hb	1.49	8	
9 C	23.02	9a, 9b	7a, 7b
Ha	1.74..1.77	9	
Hb	1.65	9	
10 C	27.99	10a, 10b	11a, 11b
Ha	1.79..1.81	10	4
Hb	1.49	10	
11 C	49.01	11a, 11b	6a, 6b
Ha	2.87	11	8, 10, 6, 4
Hb	2.66..2.70	11	8, 10, 4
12 C	50.76	12	7a, 7b
H	4.06	12	7, 4, 15, 14, 16
13 C	163.41		2a, 2b
14 C	124.13		12
15 C	156.19		12
16 C	135.76		12
17 C	127.26..129.58		
H	7.03..7.53		
18 C	127.26..129.58		
H	7.03..7.53		
19 C	126.72..127.22		
H	7.03..7.53		
20 C	127.26..129.58		
H	7.03..7.53		
20 C	127.26..129.58		
H	7.03..7.53		
21 C	127.26..129.58		
H	7.03..7.53		

22 O			
23 C	131.55, 132.11		
24 O			
25 C	122.93..134.75, 127.26..129.58	25	
H	7.03..7.53, 7.10..7.37	25	
26 C	127.26..129.58		
H	7.03..7.53		
27 C	126.72..127.22		
H	7.03..7.53		
28 C	127.26..129.58		
H	7.03..7.53		
29 C	127.26..129.58		
H	7.03..7.53		
30 C	169.08		32
31 O			
32 C	40.39	32	34, 38
H2	3.32..3.45	32	30
33 C	131.55, 132.11		
34 C	127.26..129.58		
H	6.9		32
35 C	127.26..129.58		
H	7.03..7.53		
36 C	126.72..127.22		
H	7.03..7.53		
37 C	127.26..129.58		
H	7.03..7.53		
38 C	127.26..129.58	38	
H	6.9	38	32

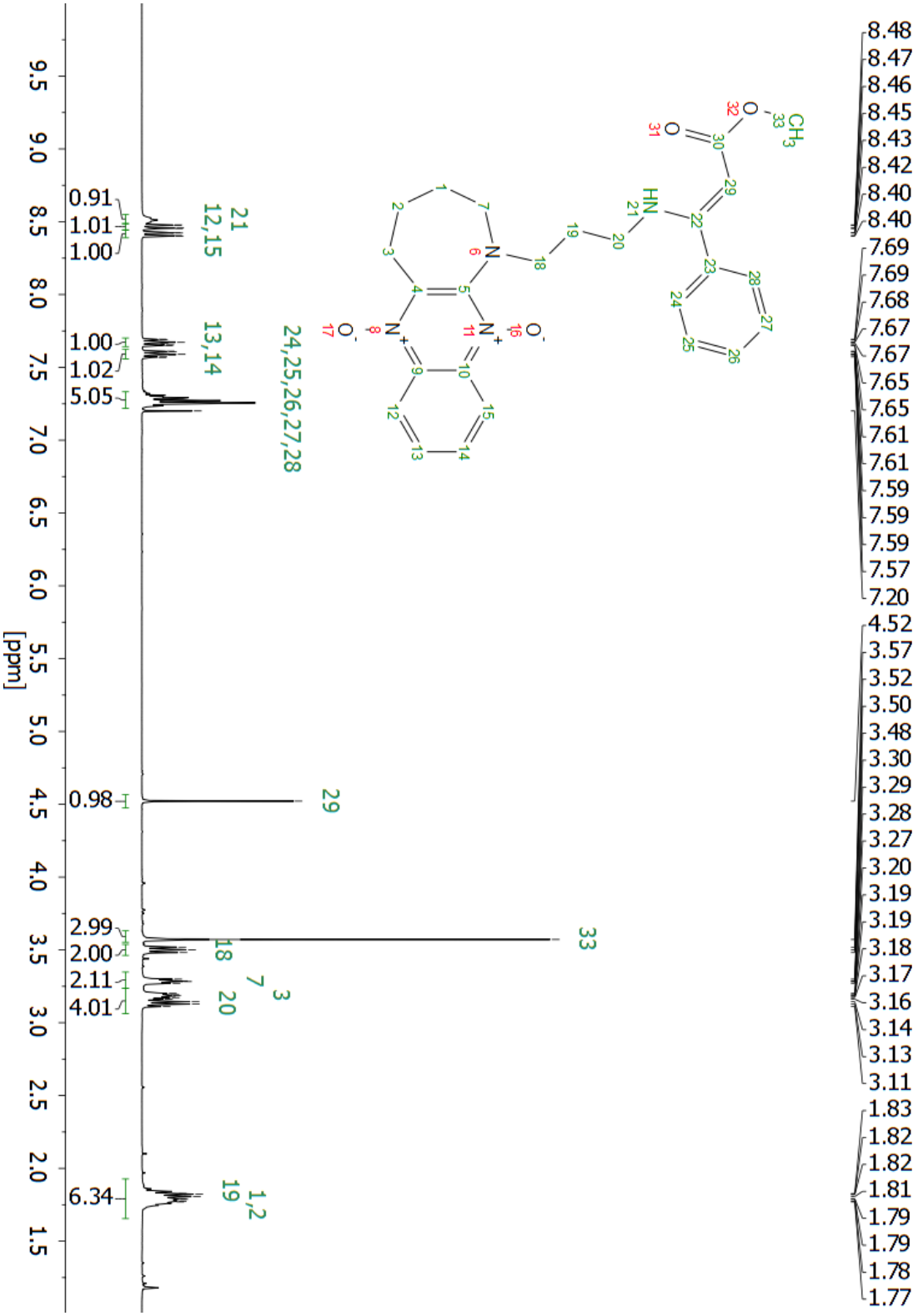


Figure S50. Assignments of ^{1}H NMR signals for compound **19b** and numeration of atoms.

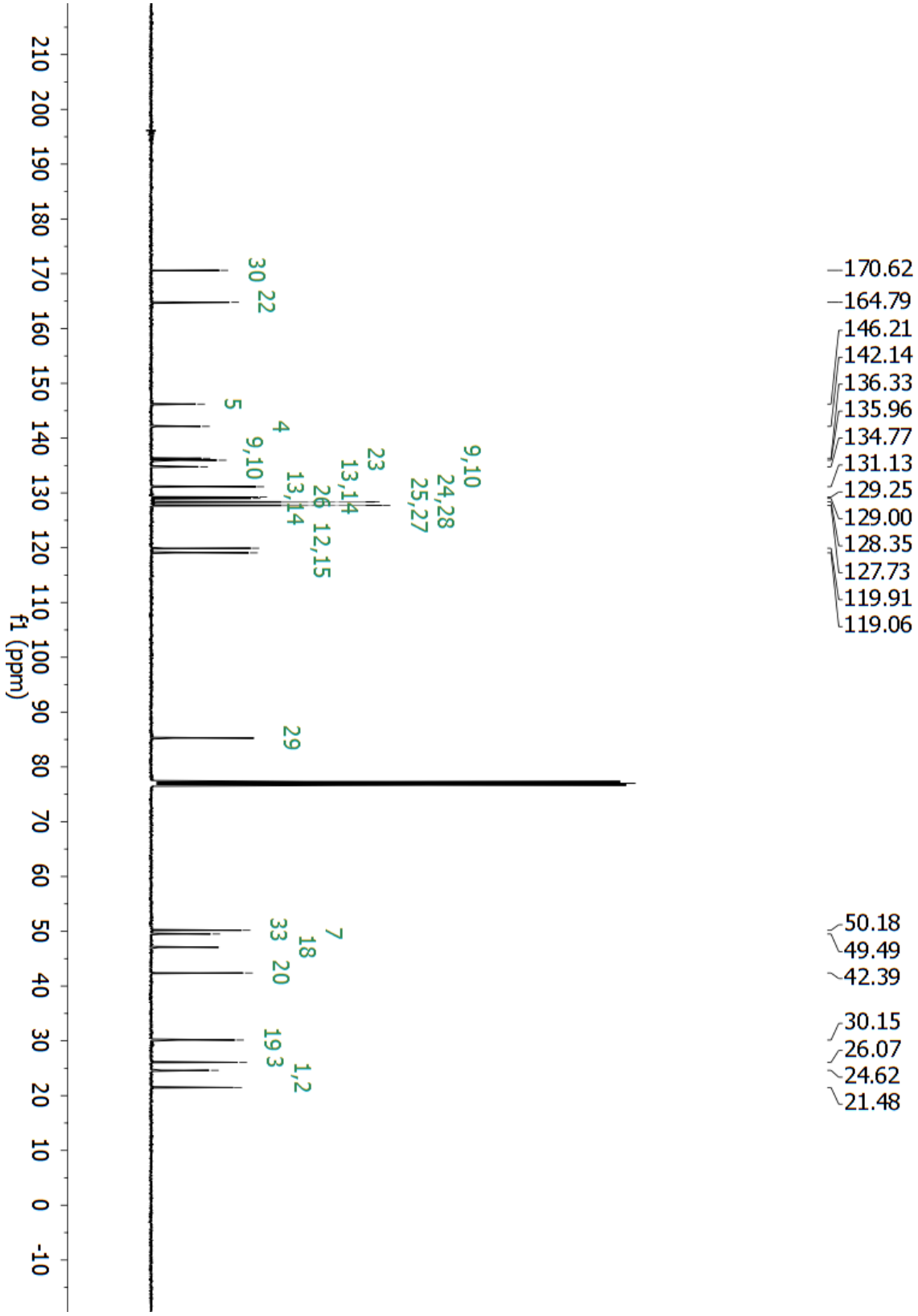


Figure S51. Assignments of ^{13}C NMR signals for compound **19b**.

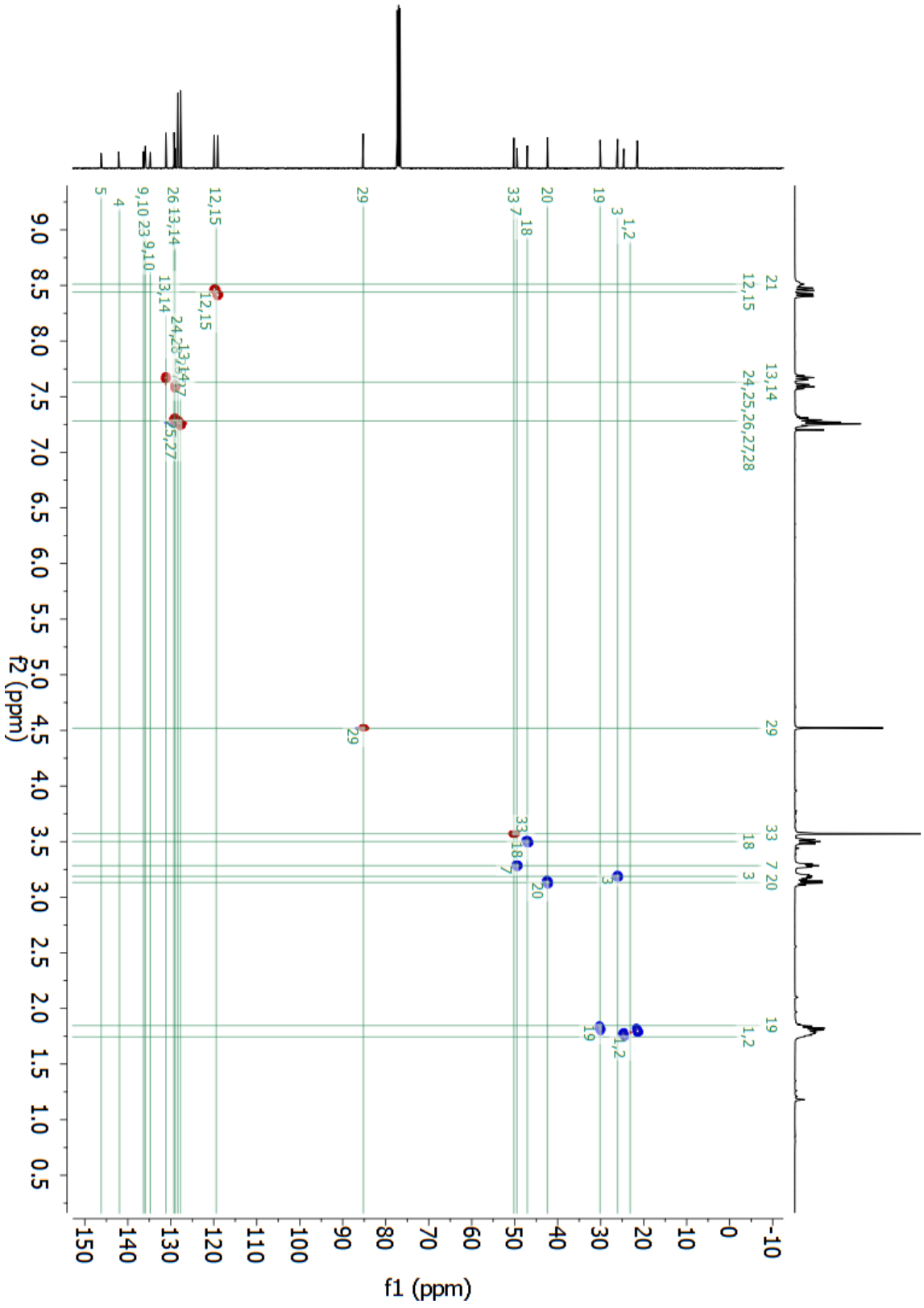


Figure S52. ^{1}H-^{13}C HSQC spectrum of compound **19b** and signal assignment.

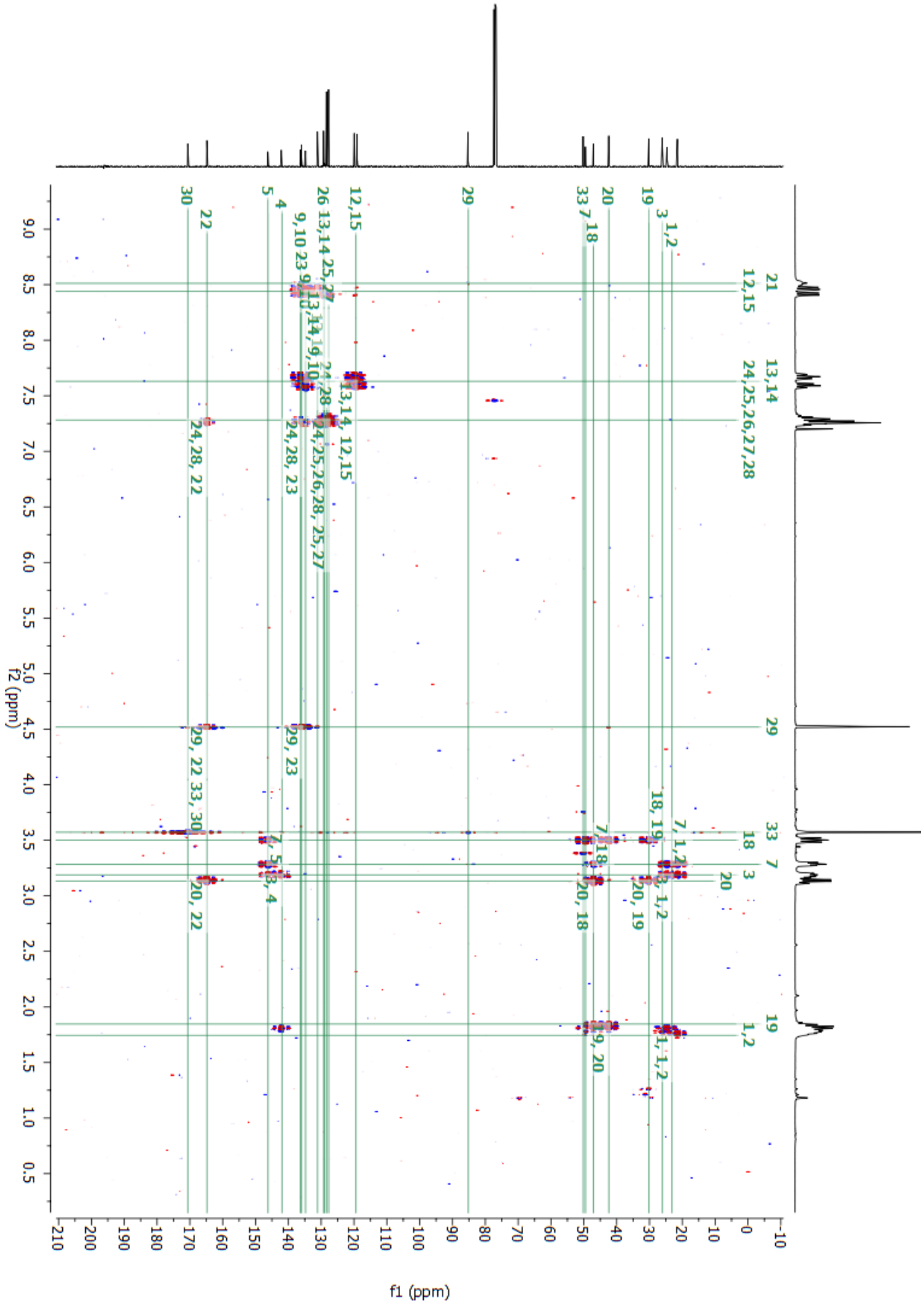


Figure S53. ^{1}H-^{13}C HMBC spectrum of compound **19b**.

a) **b)**

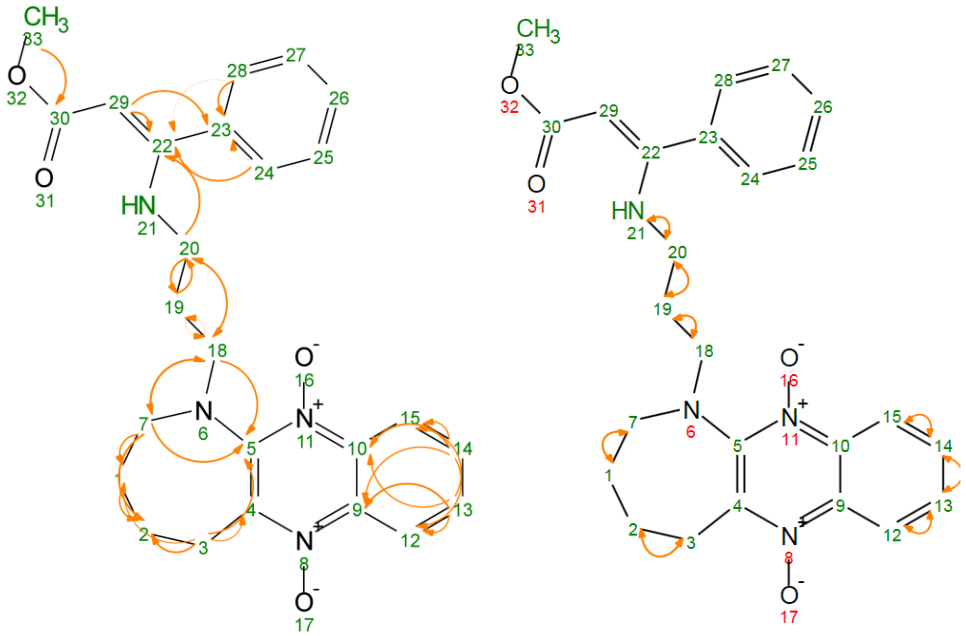


Figure S54.a) Key ^{1}H-^{13}C correlations from HMBC spectrum b) 1H-1H correlation from COSY confirming structure of **19b**.

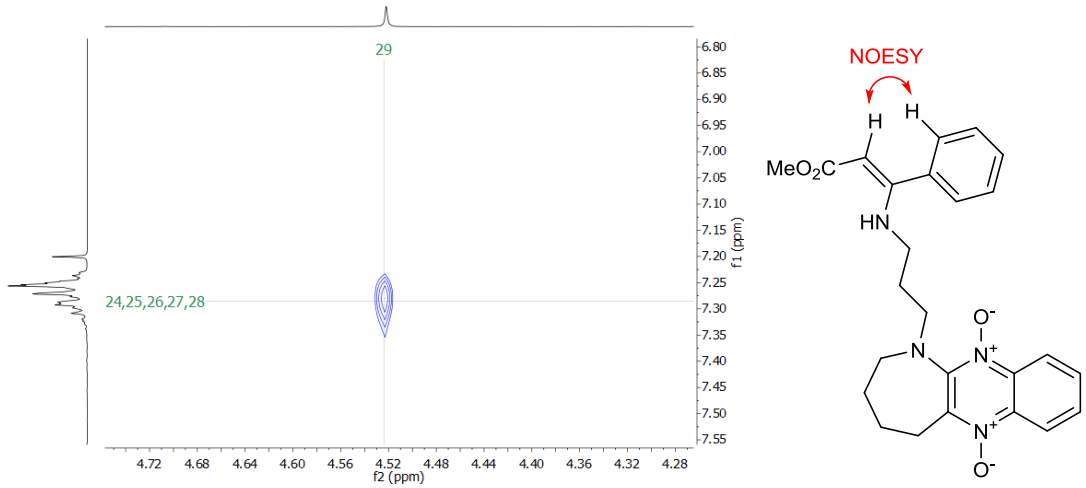


Figure S55. 2D NOESY cross-peak confirming the *Z* configuration of compound **19b**.

Table S3. Summary of structural assignments for compound **19b**.

Atom	Chemical Shift	HSQC correlations	HMBC correlations
1 C	21.23..25.07	1	2,3, 7
H2	1.69..1.80	1	2,3
2 C	21.23..25.07	2	1, 3, 7
H2	1.69..1.80	2	1,3
3 C	26.07	3	
H2	3.19	3	1, 2, 4, 5
4 C	141.29..142.58		3
5 C	145.67..146.74		3, 7, 18
6 N			
7 C	49.49	7	18
H2	3.28	7	1, 2, 18, 5
8 N			
9 C	134.77, 136.33		13, 14
10 C	134.77, 136.33		13, 14
11 N			
12 C	118.94..119.89	12	13, 14
H	8.39..8.49	12	
13 C	129.00, 131.13	13	
H	7.56..7.70	13	12, 15, 9, 10
14 C	129.00, 131.13	14	
H	7.56..7.70	14	12, 15, 9, 10
15 C	118.94..119.89	15	13, 14
H	8.39..8.49	15	
16 O			
17 O			
18 C	47.08	18	19, 20, 7
H2	3.50	18	19, 20, 7, 5
19 C	30.15	19	20, 18
H2	1.81..1.88	19	20, 18
20 C	42.39	20	19, 18
H2	3.10..3.16	20	19, 18, 22
21 N			
H	8.51		
22 C	164.79		20, 29, 24, 28
23 C	135.96		29, 24, 28
24 C	128.35	24	25, 28
H	7.22..7.34	24	22, 23, 25, 26
25 C	127.73	25	24, 26, 28

H	7.22..7.34	25	24, 26, 27
26 C	129.25	26	24, 25, 27
H	7.22..7.34	26	25, 27
27 C	127.73	27	25, 26
H	7.22..7.34	27	26, 28
28 C	128.35	28	27
H	7.22..7.34	28	22, 23, 24, 25
29 C	85.21	29	
H	4.52	29	22, 23
30 C	170.62		33
31 O			
32 O			
33 C	50.18	33	
H3	3.57	33	30

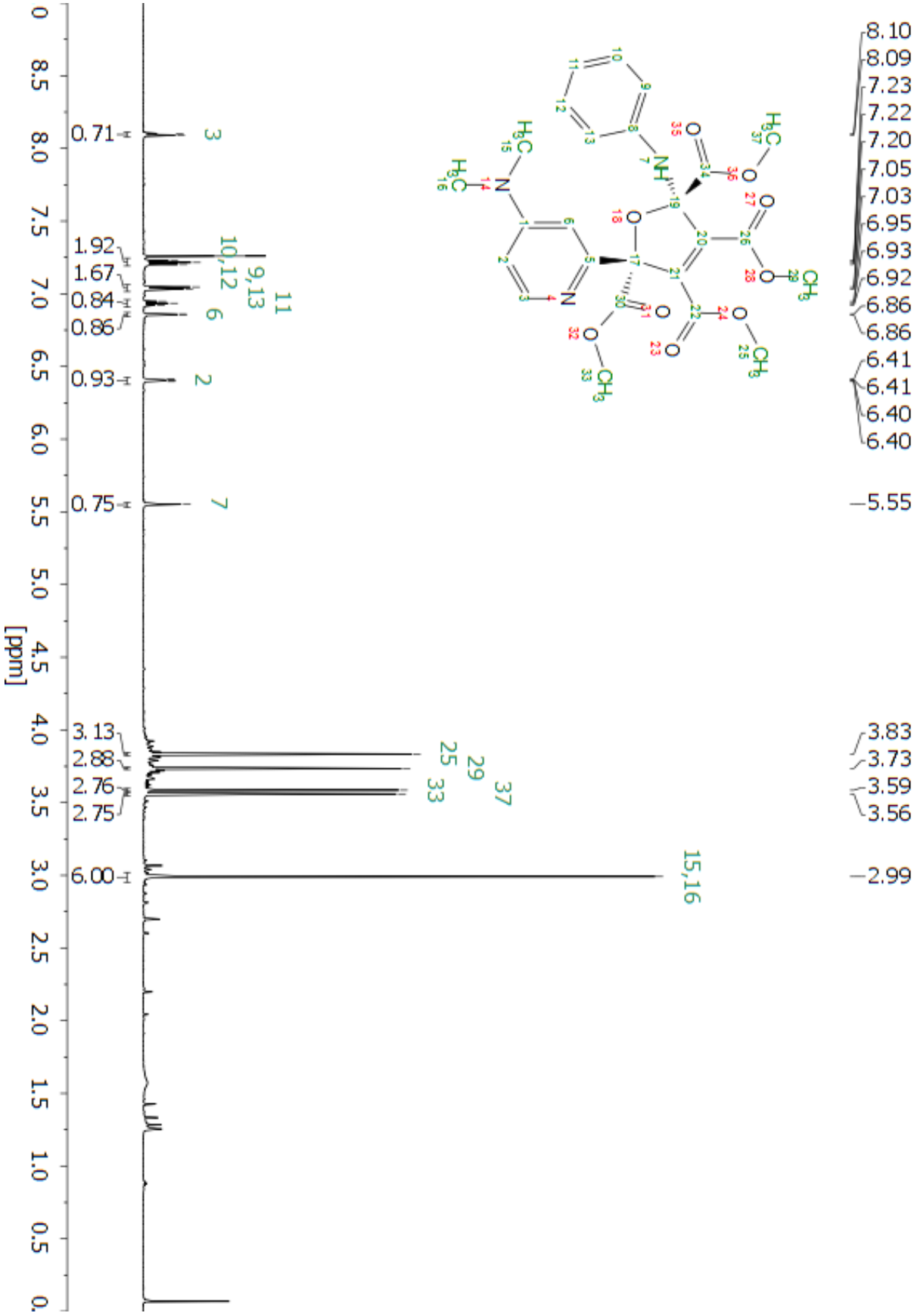


Figure S56. Proton assignments for compound *cis*-**20** and atom numeration.

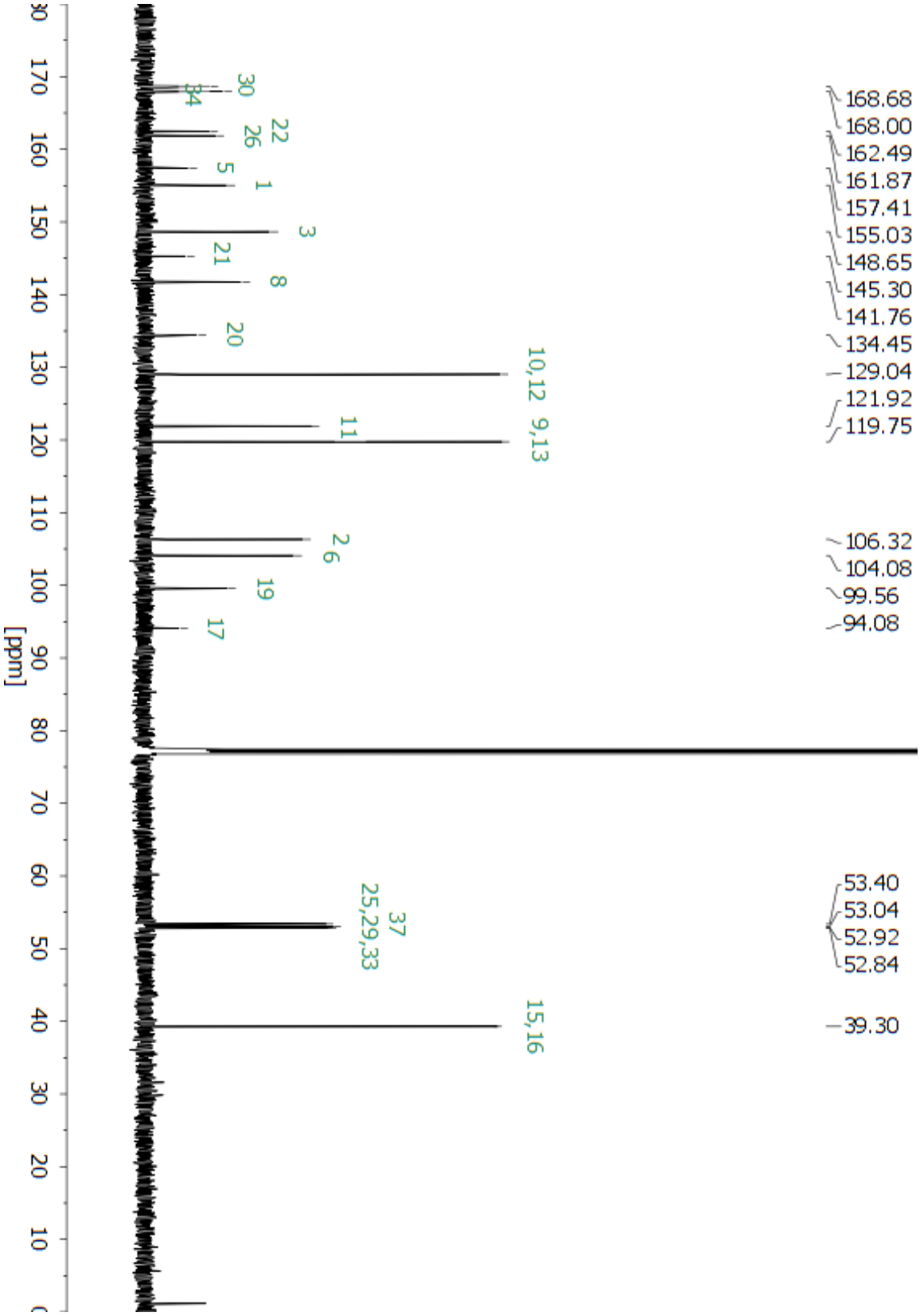


Figure S57. Carbon assignments for compound *cis*-**20** and atom numeration.

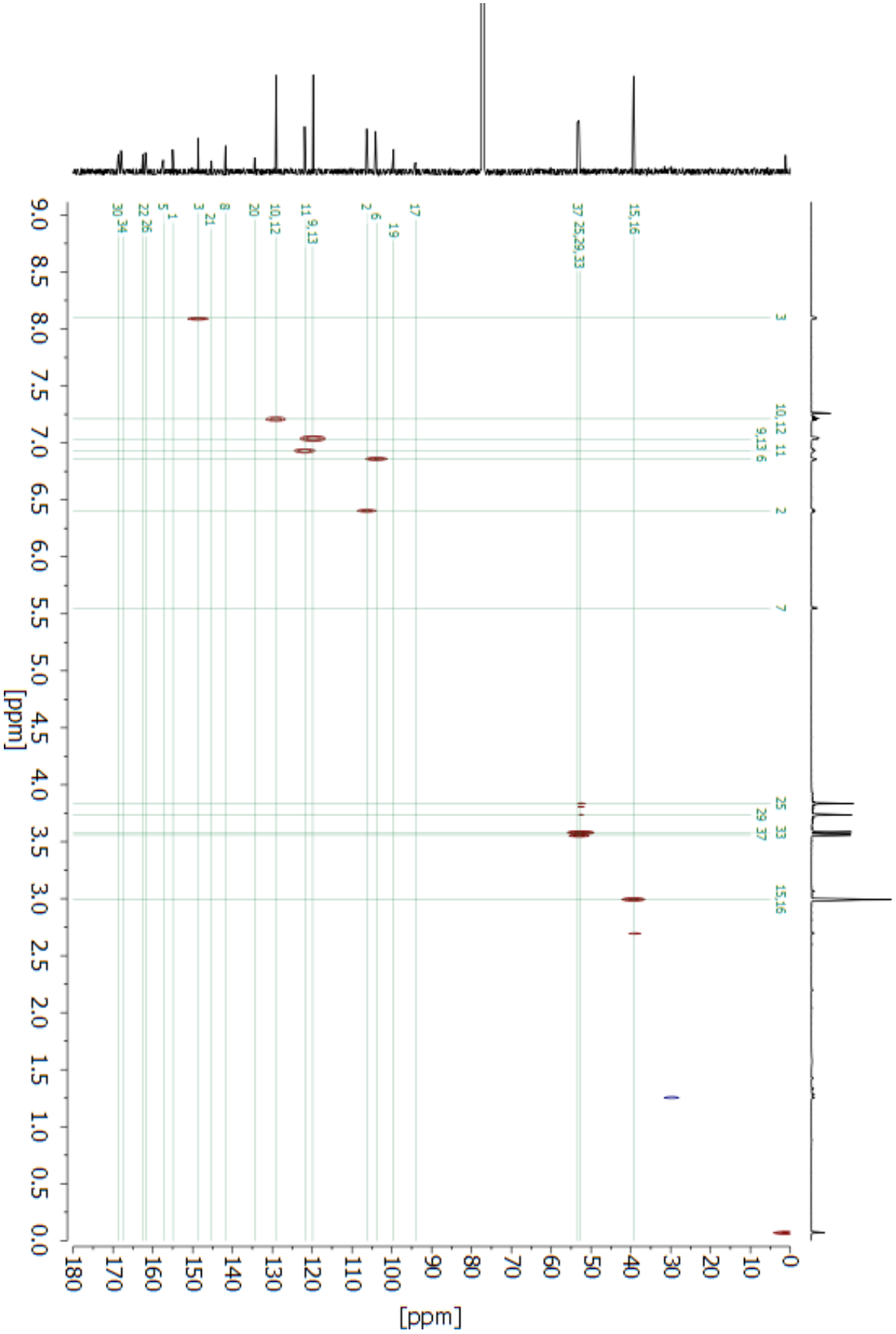


Figure S58. ^{1}H-^{13}C HSQC correlation for compound *cis*-**20** and signal assignment.

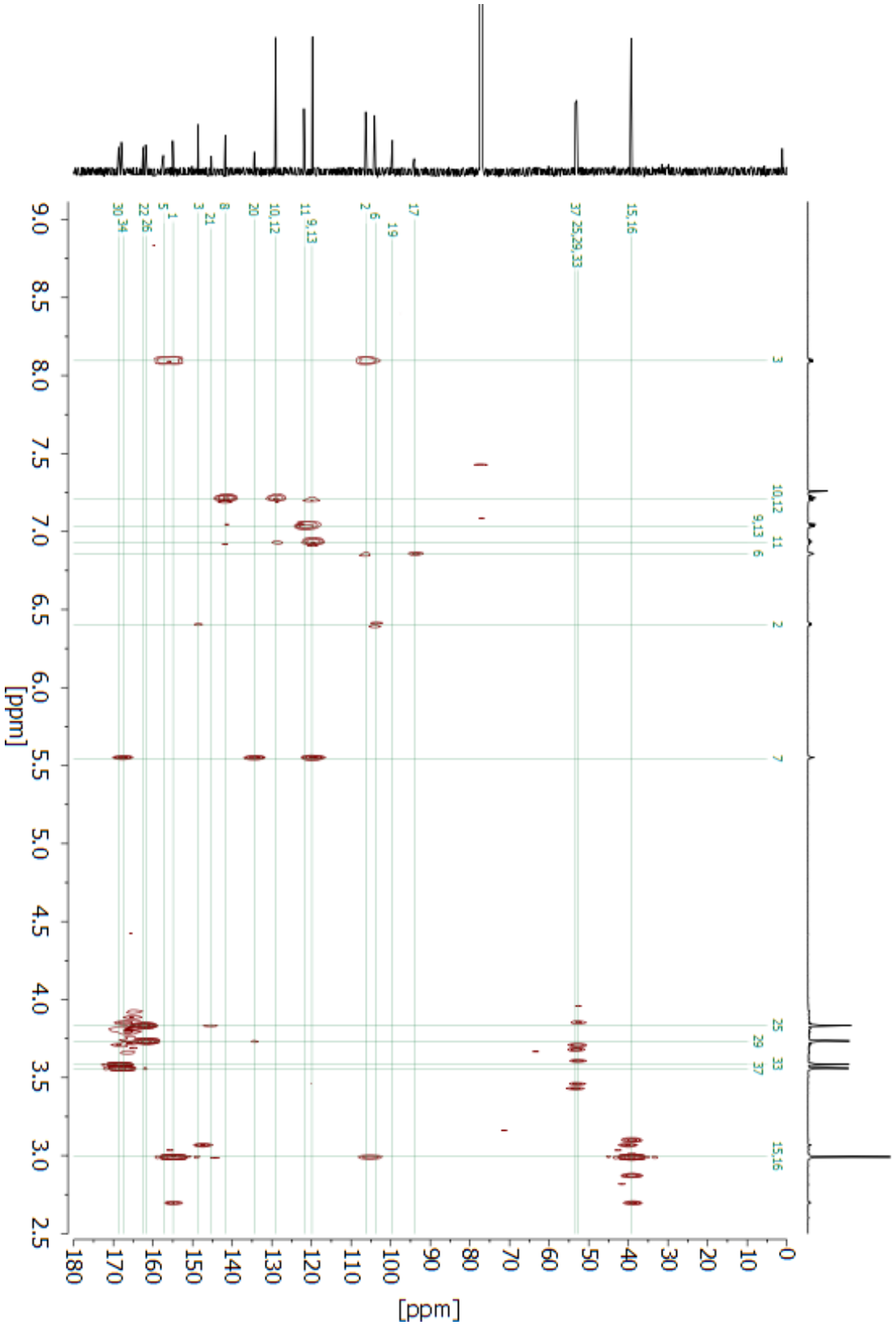


Figure S59. ^{1}H-^{13}C HMBC correlation for compound *cis*-**20** and signal assignment.

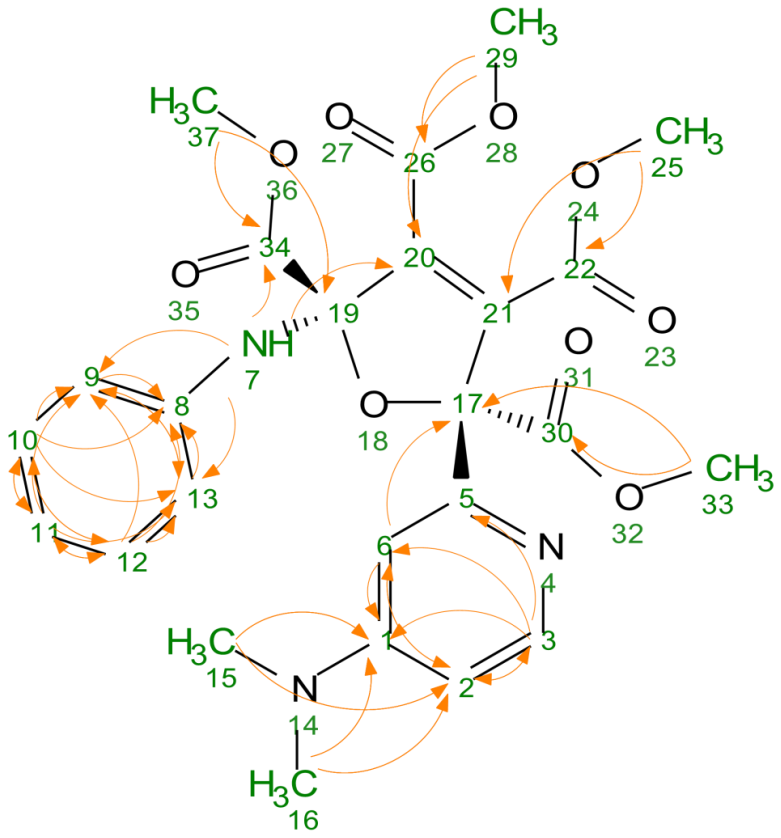


Figure S60. Key ^{1}H-^{13}C correlations from HMBC spectrum confirming structure of *cis*-**20**.

Table S4. Summary of structural assignments for compound *cis*-**20**.

Atom		Chemical Shift	HSQC	HMBC
1 C		154.95		3, 15, 16
2 C		106.27	2	3, 6, 15, 16
	H	6.40	2	3, 6
3 C		148.49	3	2, 15, 16
	H	8.09	3	1, 2, 5, 6
5 C		157.36		3
6 C		103.85	6	2, 3, 15, 16
	H	6.86	6	2, 17
7 N				
	H	5.55		9, 13, 20, 34
8 C		141.76		9, 10, 12, 13
9 C		119.87	9, 13	7, 9, 10, 11, 12,
	H	7.03	9, 13	8, 9, 13
10 C		129.06	10, 12	11, 12
	H	7.21	10, 12	8, 9, 11, 12, 13
11 C		121.70	11	10, 12
	H	6.93	11	9, 10, 12, 13
12 C		129.06	10, 12	10, 11
	H	7.21	10, 12	8, 9, 10, 11, 13
13 C		119.87	9, 13	7, 9, 10, 11, 12
	H	7.03	9, 13	8, 9
15 C		39.23	15, 16	16
	H3	2.99	15, 16	1, 2, 3, 16
16 C		39.23	15, 16	15
	H3	2.99	15, 16	1, 2, 3, 16
17 C		94.02		6
19 C		99.55		37
20 C		134.46		7, 29
21 C		145.47		25
22 C		162.50		25
25 C		52.87	25, 29, 33	
	H3	3.83	25, 29, 33	21, 22
26 C		161.70		29
29 C		52.87	25, 29, 33	
	H3	3.73	25, 29, 33	20, 26
30 C		168.75		33
33 C		52.87	25, 29, 33	
	H3	3.58	25, 29, 33	30
34 C		167.46		7
37 C		53.36	37	
	H3	3.56	37	19

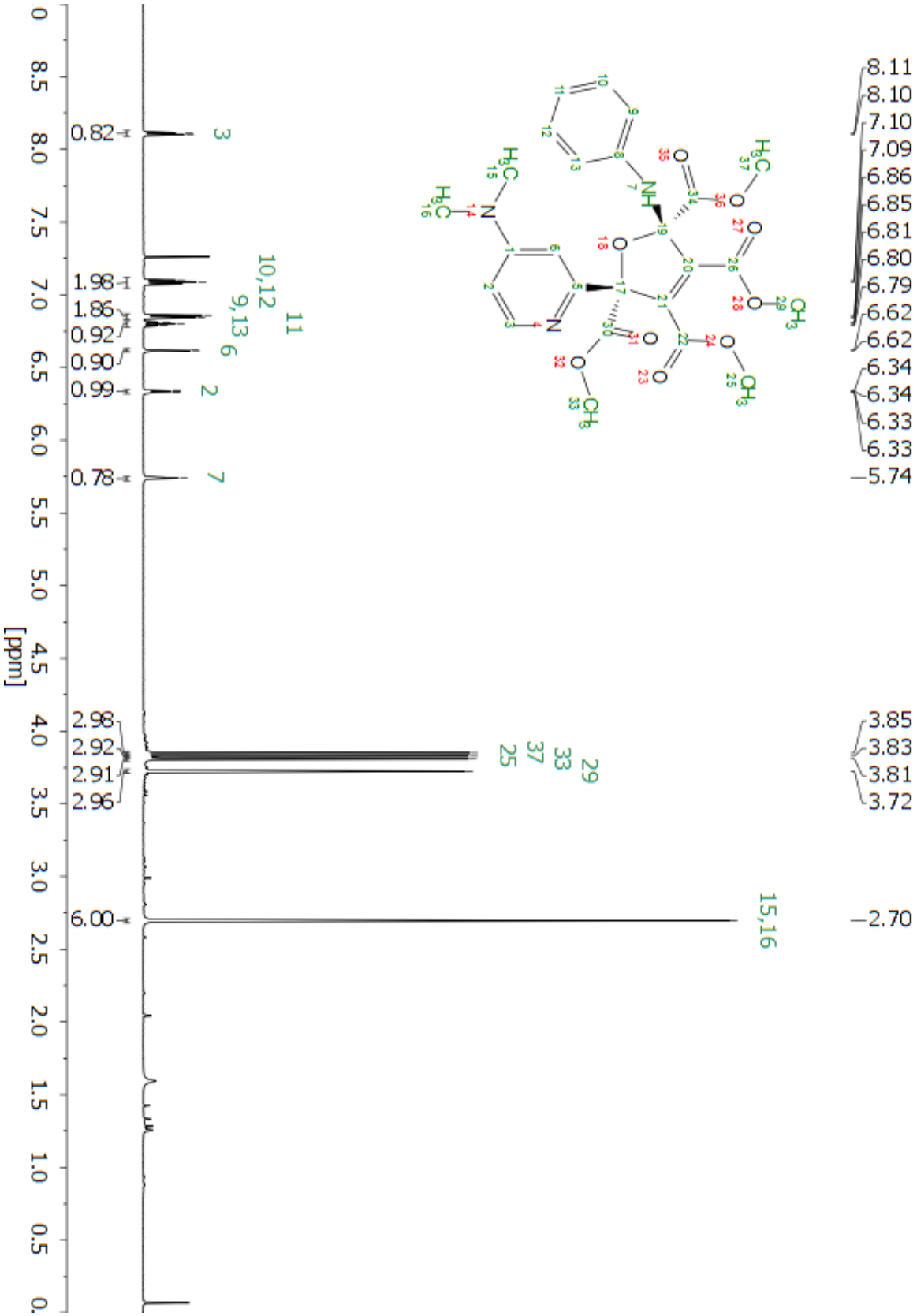


Figure S61. Proton assignments for compound *trans*-**20** and atom numeration.

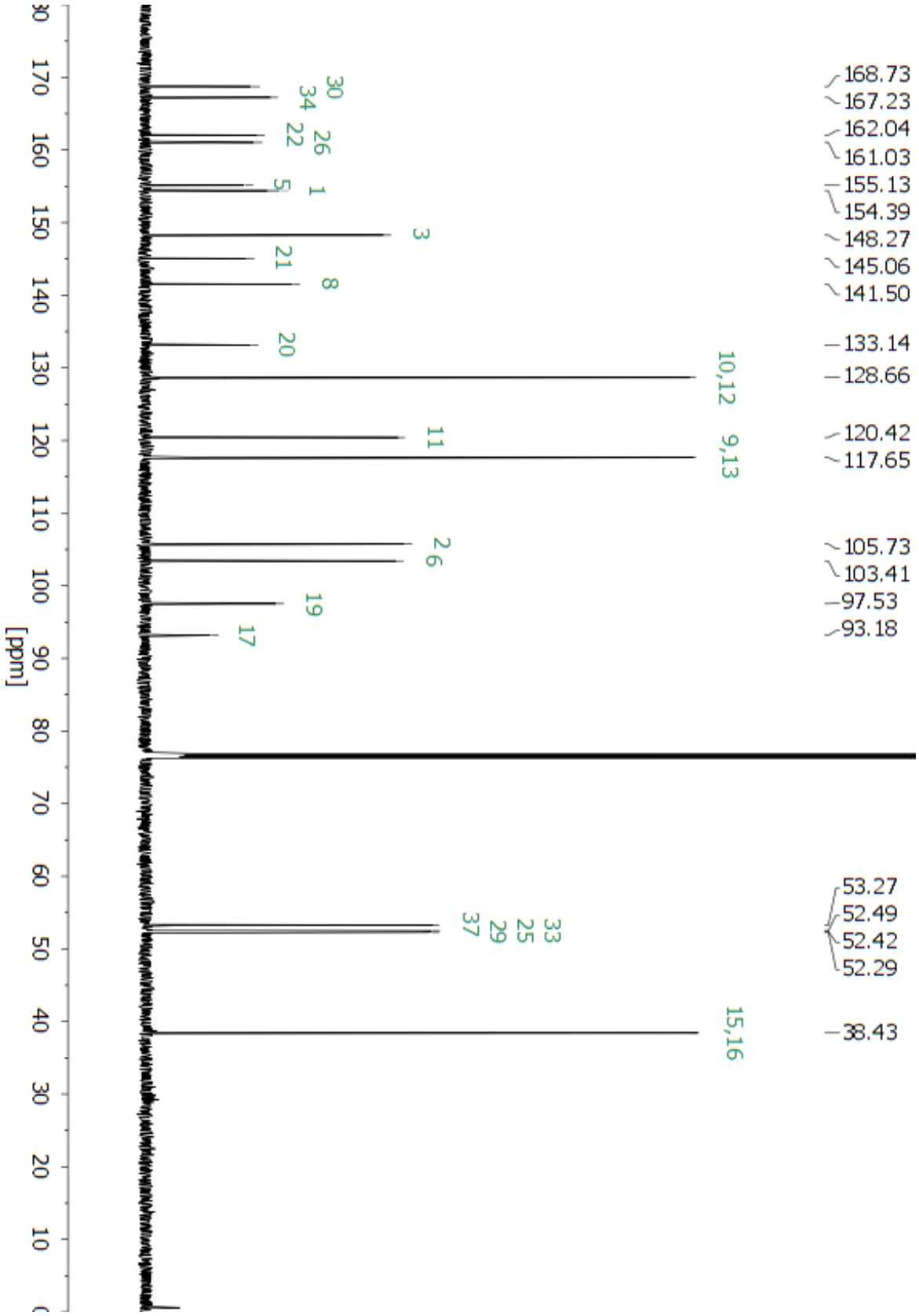


Figure S62. Carbon assignments for compound *trans*-**20** and atom numeration.

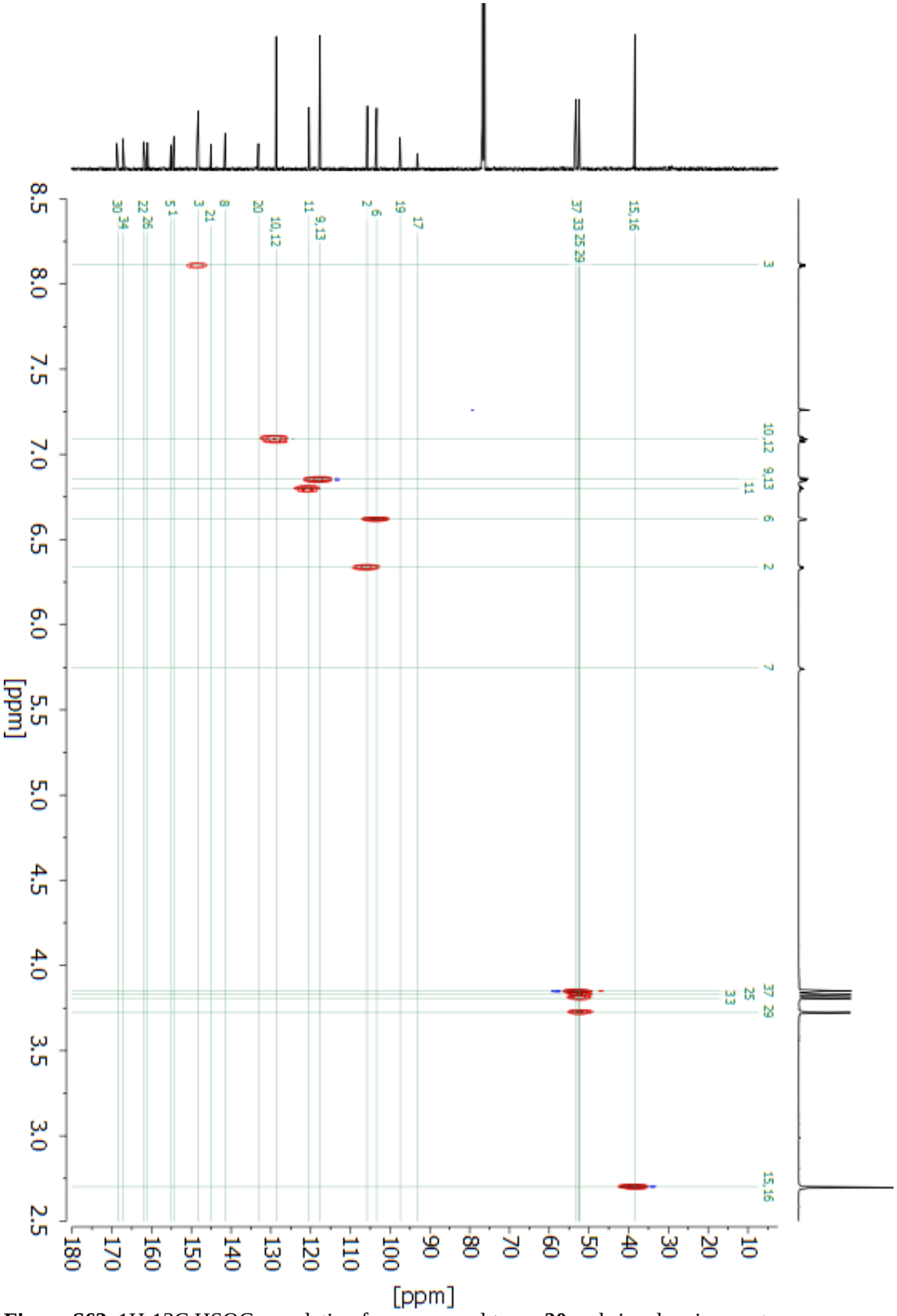


Figure S63. 1H-13C HSQC correlation for compound *trans*-**20** and signal assignment.

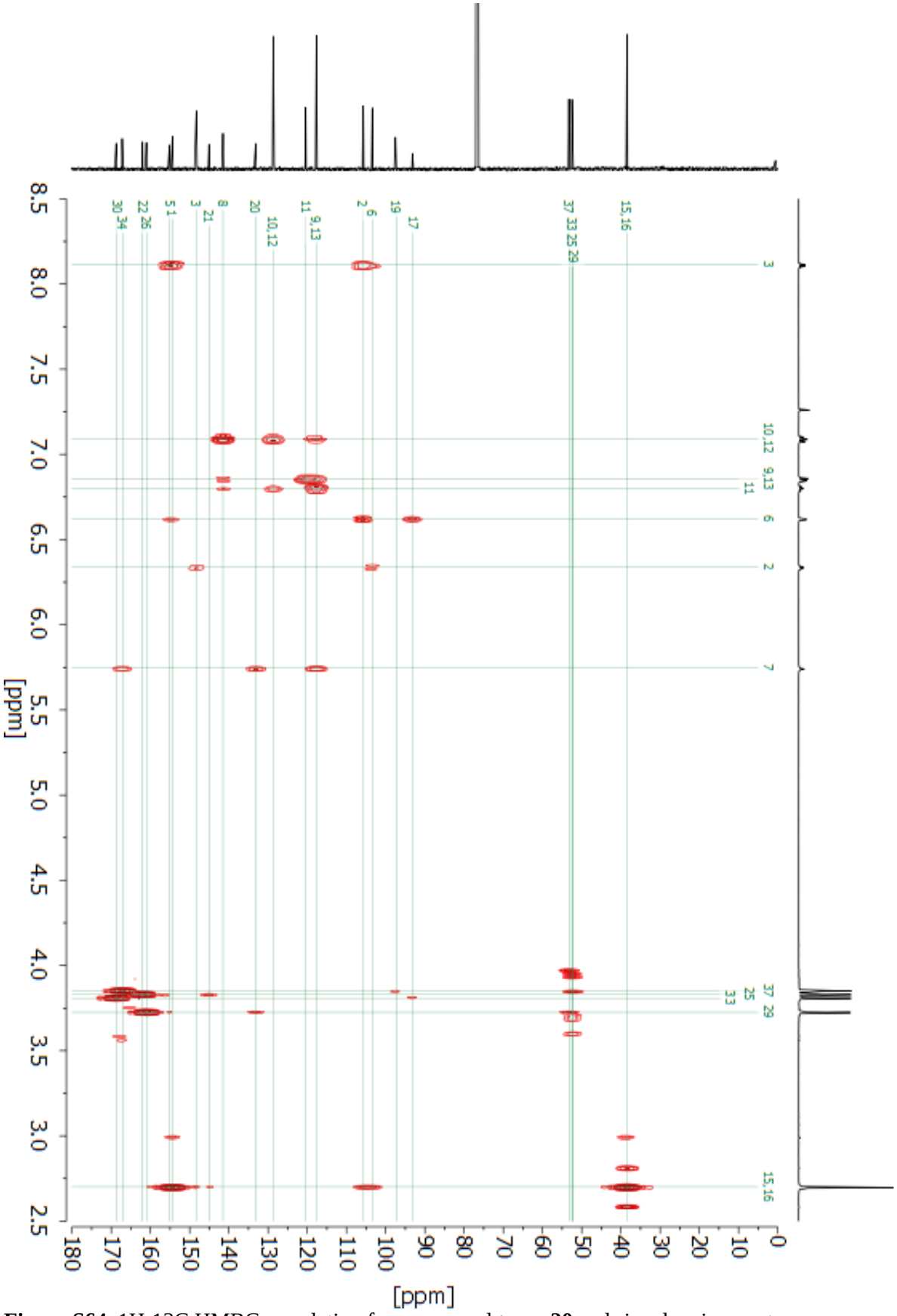


Figure S64. 1H-13C HMBC correlation for compound *trans*-**20** and signal assignment.

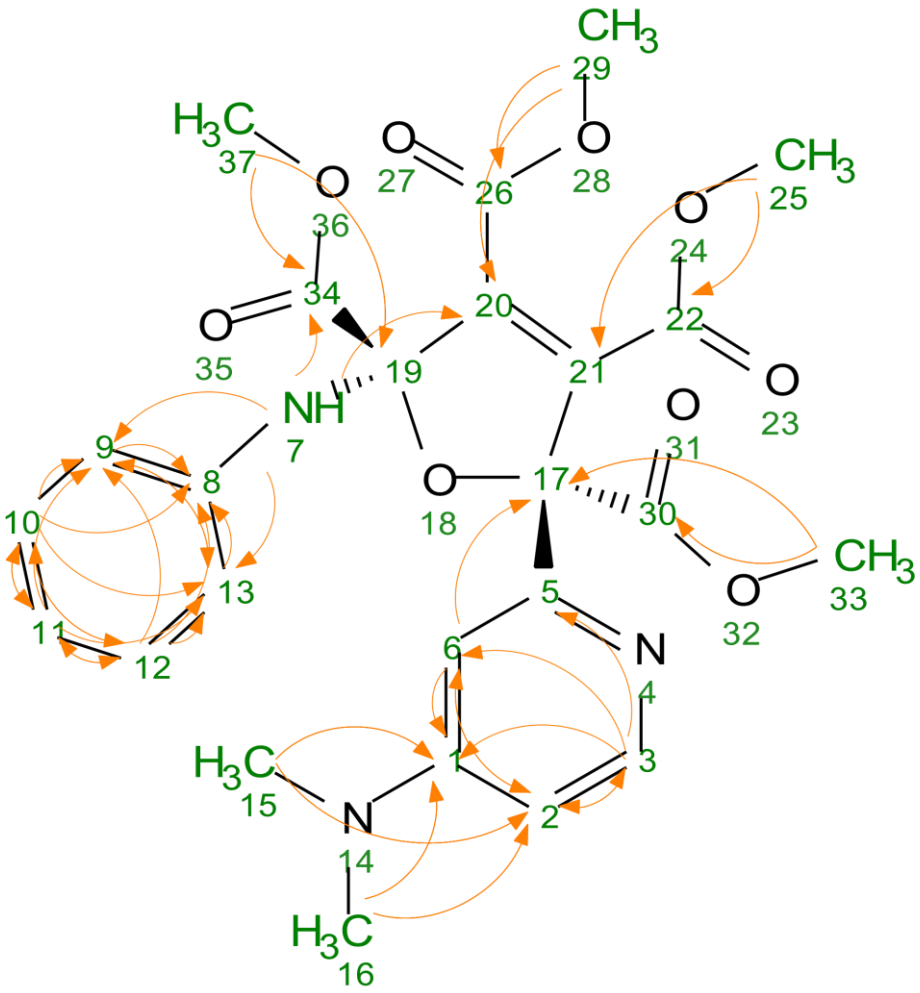


Figure S65. Key ^{1}H-^{13}C correlations from HMBC spectrum confirming structure of *trans*-**20**.

Table S5. Summary of structural assignments for compound *trans*-**20**.

Atom		Chemical Shift	HSQC	HMBC
1 C		154.39		3, 15, 16
2 C		105.81	2	3, 6, 15, 16
	H	6.34	2	3, 6
3 C		148.27	3	2, 15, 16
	H	8.11	3	1, 2, 5, 6
5 C		155.13		3, 6
6 C		103.38	6	2, 3, 15, 16
	H	6.62	6	1, 2, 5, 17
7 N				
	H	5.75		9, 13, 20, 30, 34
8 C		141.50		9, 10, 11, 12, 13
9 C		117.65	9, 13	7, 10, 11, 12, 13
	H	6.86	9, 13	8, 11, 13
10 C		128.66	10, 12	11, 12
	H	7.09	10, 12	8, 9, 11, 12, 13
11 C		120.42	11	10, 12
	H	6.80	11	8, 9, 10, 12, 13
12 C		128.66	10, 12	10, 11
	H	7.09	10, 12	8, 9, 10, 11, 13
13 C		117.65	9, 13	7, 9, 10, 11, 12
	H	6.86	9, 13	8, 9, 11
15 C		38.43	15, 16	16
	H3	2.70	15, 16	1, 2, 3, 5, 6, 16
16 C		38.43	15, 16	15
	H3	2.70	15, 16	1, 2, 3, 5, 6, 15
17 C		93.11		6, 33
19 C		97.36		37
20 C		133.14		7, 29
21 C		145.06		25
22 C		162.05		25
25 C		52.42	25	
	H3	3.83	25	21, 22
26 C		161.03		29
29 C		52.29	29	
	H3	3.72	29	20, 26
30 C		168.54		33
33 C		52.49	33	
	H3	3.81	33	30
34 C		167.15		7
37 C		53.27	37	
	H3	3.85	37	34

8. Crystal structure details

Single crystal X-Ray diffraction: Single crystal datasets and unit cells were collected at 150 K unless stated otherwise on the following instrument: Bruker Apex II Quasar X-ray Diffractometer with enhanced X-ray beam (λMo-Kα = 0.71073 Å, T = 150 K, monochromated graphite). Structure solution and refinement were carried out with SHELXS-97[1] and SHELXL-97[2] *via* WinGX[3]. The X-ray crystallographic data for structures reported in this article have been deposited at the Cambridge Crystallographic Data Centre, under deposition number CCDC-1810787-1810788. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

[1] G. Sheldrick, *Acta Crystallogr. A* **1990**, *46*, 467-473.

[2] G. Sheldrick, *Acta Crystallogr. A* **2008**, *64*, 112-122.

[3] L. Farrugia, *J. Appl. Crystallogr.* **1999**, *32*, 837-838.

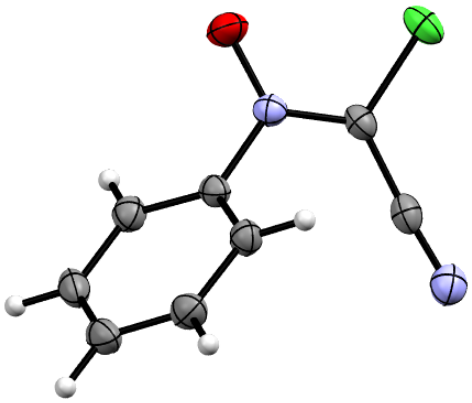


Figure S66. X-ray structure of compound **21** (50 % probabilty level)**.**

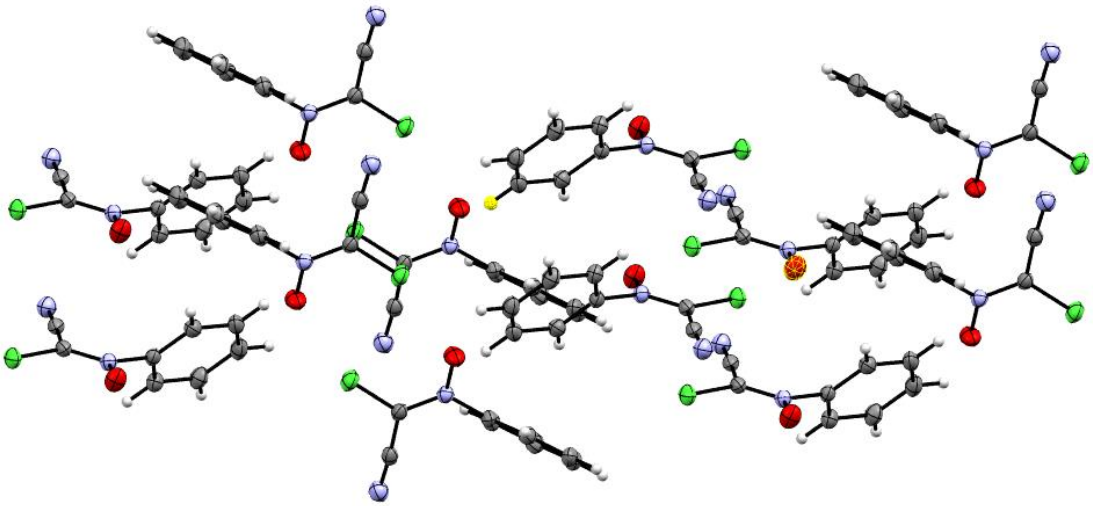


Figure S67. Crystal packing of compound **21**.

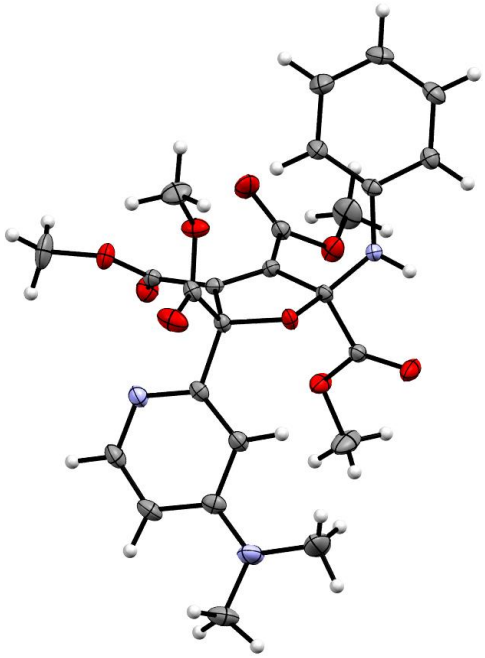


Figure S68. Crystal structure of compound **20** (50 % probability level)

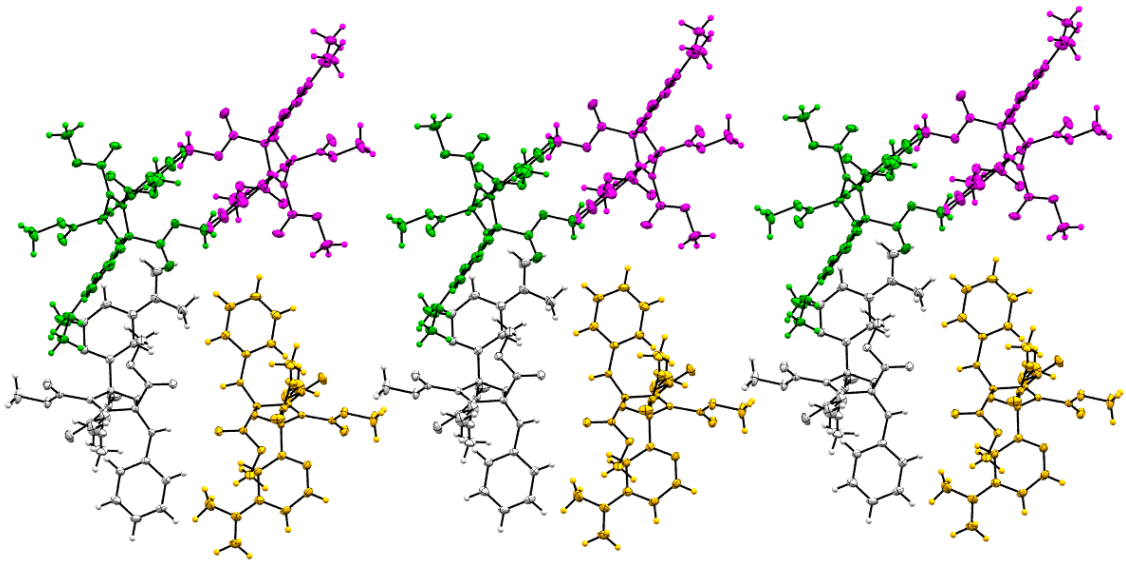


Figure S69. Crystal packing of compound **20** (coloured by symmetry equivalence).

Table S5. Crystal data and structure refinement for **21**.

Identification code	**CCDC 1810788**	
Empirical formula	C8 H5 Cl N2 O	
Formula weight	180.59	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	C 2/c	
Unit cell dimensions	a = 24.182(2) Å	α= 90°.
	b = 4.0877(3) Å	β= 117.415(10)°.
	c = 17.9047(18) Å	γ = 90°.
Volume	1571.1(3) $Å^3$	
Z	8	
Density (calculated)	1.527 Mg/m^3	
Absorption coefficient	0.430 mm^{-1}	
F(000)	736	
Crystal size	0.100 x 0.100 x 0.060 mm^3	
Theta range for data collection	2.563 to 25.997°.	
Index ranges	-29<=h<=29, -5<=k<=5, -22<=l<=22	
Reflections collected	18038	
Independent reflections	1496 [R(int) = 0.0471]	
Completeness to theta = 25.242°	97.7 %	

Absorption correction	Empirical
Max. and min. transmission	0.746 and 0.651
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	1496 / 0 / 109
Goodness-of-fit on F^2	1.058
Final R indices [I>2sigma(I)]	R1 = 0.0321, wR2 = 0.0754
R indices (all data)	R1 = 0.0400, wR2 = 0.0800
Extinction coefficient	n/a
Largest diff. peak and hole	0.21 and -0.23 e.$Å^{-3}$

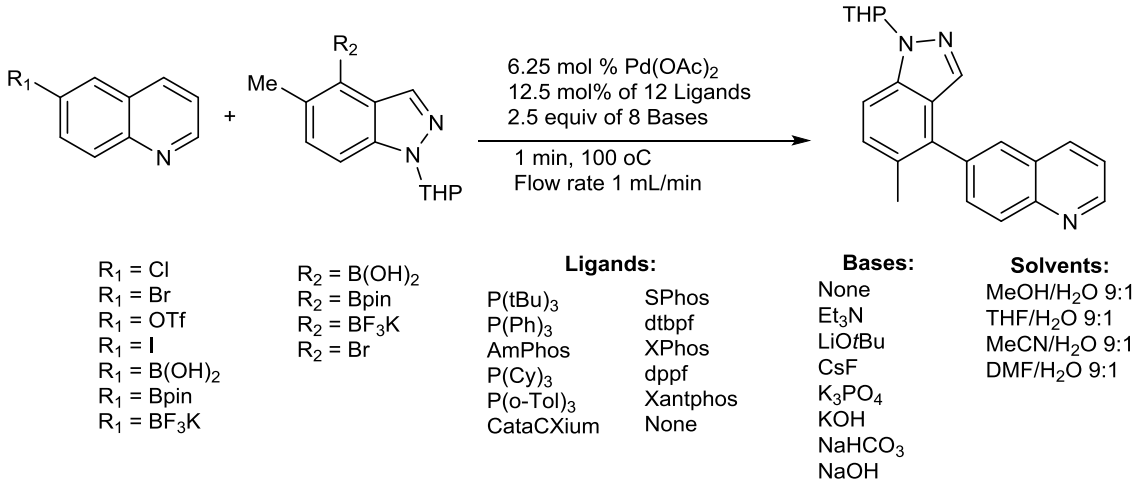
Table S6. Crystal data and structure refinement for *cis*-**20**.

Identification code	**CCDC 1810787**	
Empirical formula	C25 H27 N3 O9	
Formula weight	513.49	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P 21/n	
Unit cell dimensions	a = 14.9853(9) Å	α= 90°.
	b = 9.5292(6) Å	β= 110.852(3)°.
	c = 18.5901(12) Å	γ = 90°.
Volume	2480.8(3) $Å^3$	
Z	4	
Density (calculated)	1.375 Mg/m^3	
Absorption coefficient	0.106 mm^{-1}	
F(000)	1080	
Crystal size	0.100 x 0.100 x 0.100 mm^3	
Theta range for data collection	2.169 to 25.997°.	
Index ranges	-18<=h<=18, -11<=k<=11, -22<=l<=22	
Reflections collected	38770	
Independent reflections	4872 [R(int) = 0.0283]	
Completeness to theta = 25.242°	100.0 %	
Absorption correction	Empirical	
Max. and min. transmission	0.746 and 0.627	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	4872 / 0 / 343	

Goodness-of-fit on F^2	1.059
Final R indices [I>2sigma(I)]	R1 = 0.0342, wR2 = 0.0899
R indices (all data)	R1 = 0.0382, wR2 = 0.0927
Extinction coefficient	n/a
Largest diff. peak and hole	0.29 and -0.22 e.$Å^{-3}$

9. Machine learning exploration of Suzuki-Miyaura reaction

The reaction space (Perera *et al.*, Science 359, 429–434, **2018**) contains 5760 experiments of Suzuki-Miyaura coupling reaction (Scheme S1).



Scheme S1. Chemical space of Suzuki-Miayura coupling.

One-hot encoding of reactions and data processing

Each reaction has been one-hot encoded as a vector of length [1x37](Figure S70). This representation doesn't require any chemical knowledge about chemical system being investigated. The yields were scaled to range 0.0 – 1.0. The constant parameters for all reactions haven't been encoded e.g. amount of palladium acetate, temperature, and flow rate.

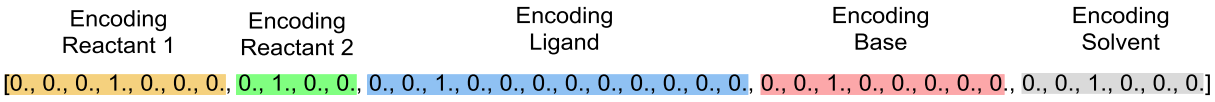


Figure S70. An example of encoding one of the Suzuki-Miyaura reactions.

Neural network architecture

The neural network (NN) (Figure S71) comprised two layers: 50 neurons in the first fully connected layer with sigmoid activation function and dropout probability 0.8 for training. The

second layer comprised 7 neurons in the fully connected layer also with sigmoid activation. The final prediction of yield was obtained as a linear regression of the output from the second layer. Mean squared error between predicted and experimental yield was implemented as a loss function to train NN. The NN was implemented in Tensorflow.

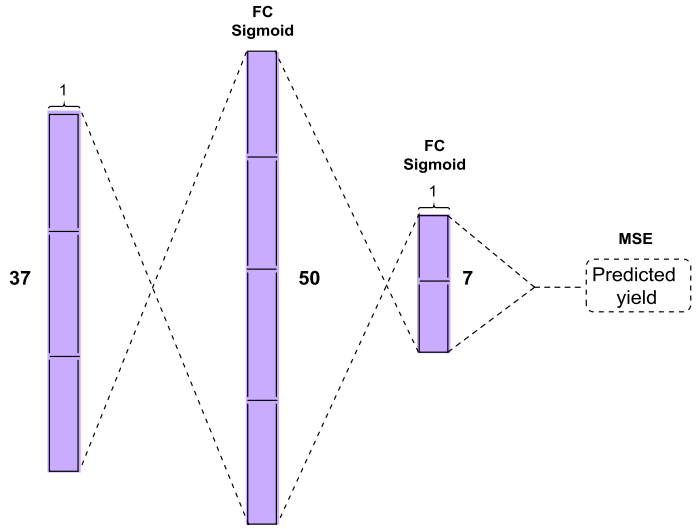


Figure S71. The structure of the neural network used for yield prediction.

Validation of the neural network for yield prediction

The reaction data has been randomly splited into training / validation / test sets 60/10/30 % and neural network has been trained and validated using training and validation set. The neural network was trained using Adam Optimizer(D.P. Kingma, J. Ba, CoRR, abs/1412.6980, 2014) (learning rate = 0.005) and Mean Squared Error as a loss function with minibatch size of 100. To minimize overfitting the early stopping have been implemented. Figure 72a shows the training process for 300 epochs. The mean squared error was MSE = 0.01208 for 1728 reactions in the test set (see Figure S72b).

a **b**

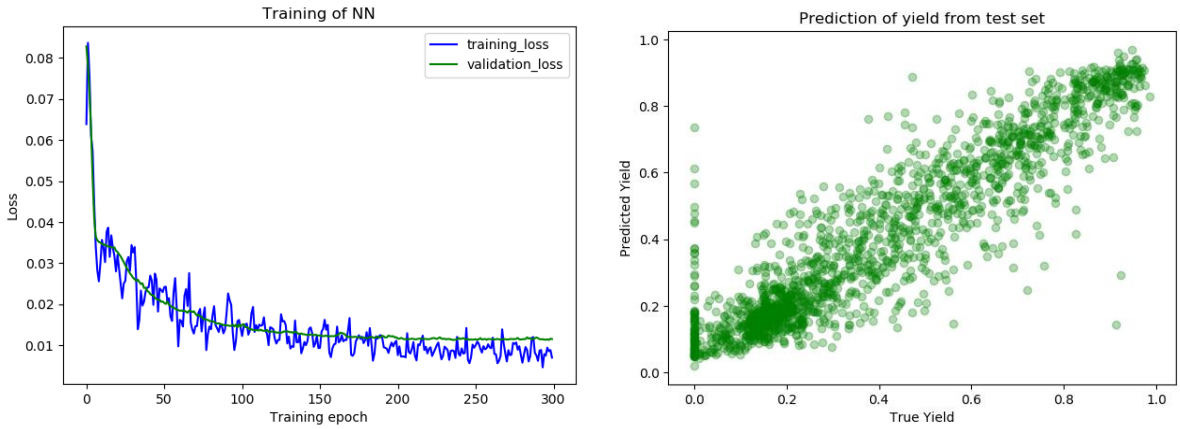
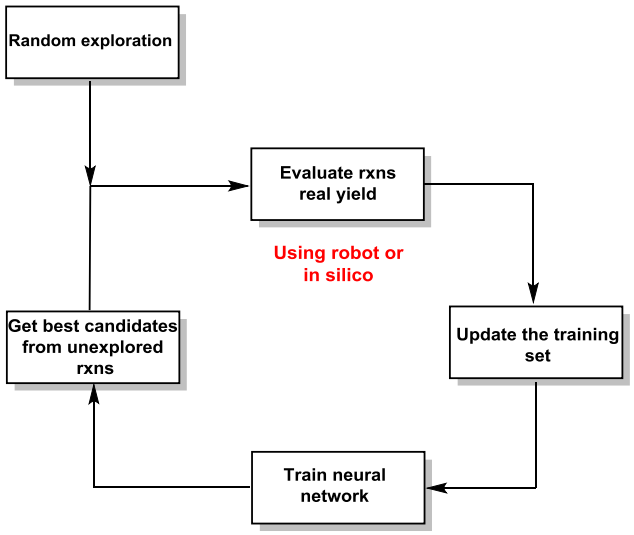


Figure S72. a. Loss for training and validation sets during training. **b**. Correlation for test set between predicted and experimental yield for the test set.

Simulating exploration of the reaction space using machine learning

The main goal of the simulation was to show that our approach is able to help in design and development of organic reactions including high yielding transition metal catalysed transformations. Analogously to simulation in our paper the algorithm starts with random screening of chemical space by selecting 10% of reactions (576 reactions), then the neural net is trained on this data and all the remaining reactions are then scored by the model. The candidates with the highest predicted yield are then added to the performed reactions, and the performance of the NN is evaluated by calculating the mean of true yield and standard deviation (SD) of the yield. The NN is then retrained and the whole cycle is repeated until the whole space is explored. (Figure 5a) The space was explored in a batch of 100 to demonstrate compatibility with high throughput screening, as well as to evaluate the performance of NN. Exploratory phase consists of 576 reactions and gave on average yield about 39 % with SD = 27 % (Figure 5b). After training the NN, for the next batch of 100 reactions selected the average yield was typically 86 % with SD = 11%.

a



b

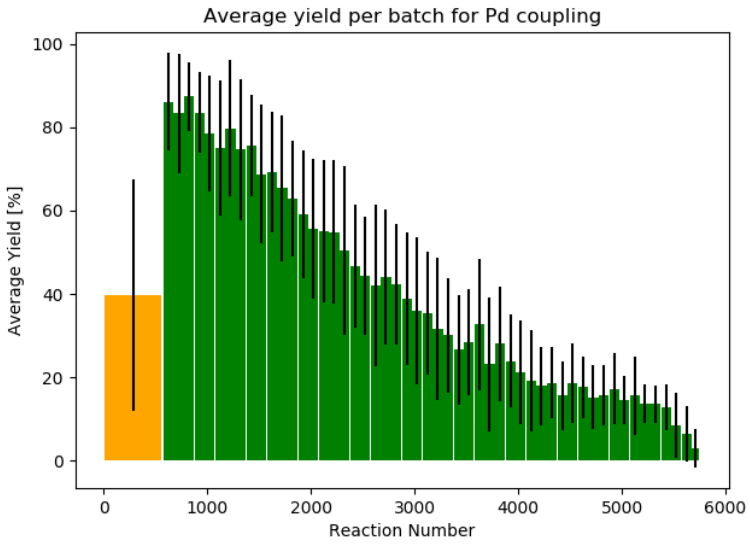


Figure S73. a. Closed loop exploration of chemical space of Suzuki-Miyaura reaction searching for maximum yield. **b**. Average yields and standard deviation of the yield during exploration of the chemical space. The yellow bar represents the random search containing 10 % of chemical space (576 reactions) and the green bars show subsequent batches of 100 reaction chosen by neural network.

The subsequent batches contained less and less reactive starting materials, reaching finally unreactive parts of the reaction space. These results prove that our machine learning approach can successfully applied to such important field as transition metal catalysis, helping to design high yielding transformations and as well as search for reactivity.

10. Tanimoto similarity analysis of reactions

The Tanimoto similarity index uses the fingerprints of molecules to compare how similar they are to each other. To compare reactions we calculate the similarity between each reagent and the product and calculate the mean from the obtained values. For reactions where the reagents undergo a slight modification to reach the product this reaction similarity value would be close to 1. Conversely, if the reagents change substantially so that the product is very different then the result would be close to 0. Starting from a database of over 40 million reactions we exclude non-organic reactions and any reactions that do not have the same number of reagents and products as our discovered reactions. We further filter according to reactions that have the complete structural data in the database. This filtering has still left us with over ~3.5 million reactions to compare to. The script used to perform these calculations was written in Python using the RDKit library. Running over the millions of reactions in addition to the discovered reactions allows for a comparison of the statistical features within the whole reaction similarity distribution.

11. Supplementary files

Supplementary file: **Space_Exploration.csv** shows exemplary run of LDA algorithm exploring the chemical space. The first ninety experiments were chosen randomly and the next subsequent experiments were chosen using the LDA classifier. The name column contains the identity of the reaction composed from the names of the starting materials. The reactivity column contains the assignment of reactivity from SVM classifier for a given reaction mixture.

Supplementary file: **LDA_reactivity.csv** contains the LDA scores for all 969 reaction formed from chemical space shown in Figure 3. The name column contains the identity of the reaction composed from the names of the starting materials. The LDA_reactivity column contains scores from LDA and reactivity column contains the assignment of reactivity from SVM classifier for a given reaction mixture.