Supplementary information

Discovering new chemistry with an autonomous robotic platform driven by a reactivity-seeking neural network

Dario Caramelli, Jarosław M. Granda, S. Hessam M. Mehr, Dario Cambié, Alon B. Henson and Leroy Cronin*

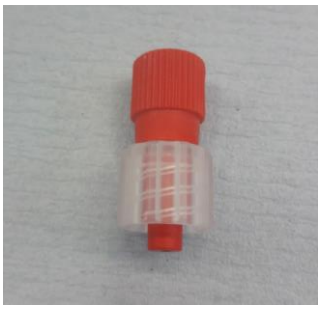
School of Chemistry, The University of Glasgow, Glasgow G12 8QQ (UK)
Web: www.croninlab.com Email: Lee.Cronin@Glasgow.ac.uk

Code availability. The code for the neural network testing, the chemical space exploration and the cheminformatics plots can be found online at https://github.com/croningp/Organic-Finder and the code is provided under the MIT license.

Supporting data set availability. The data and code required to reproduce the figures in the manuscript is here at https://zenodo.org/record/4670997 (doi: 10.5281/zenodo.4670997)

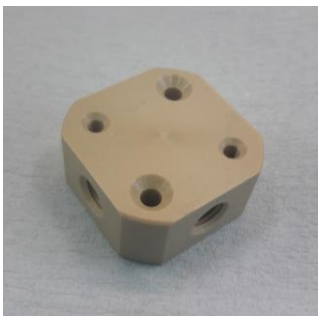
CONTENTS

1 Hardware specifications



Adapter, Luer (Male) to 1/4"-28 Flat Bottom (Female), ETFE/Polypropylene P-675 (IDEX-HS P-675)

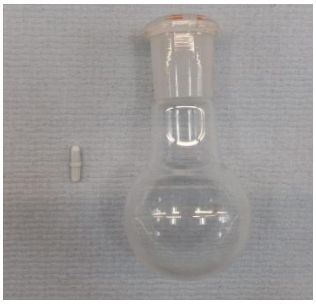
Flangeless Fitting, for 1/16" OD Tubing, 1/4"-28 Flat Bottom, PEEK/ETFE (IDEX-HS XP-218BLK)



Block Connector, 3-Way, 1/4"-28 (Flat Bottom), PEEK, with mounting holes (Diba Dibafit 001057)

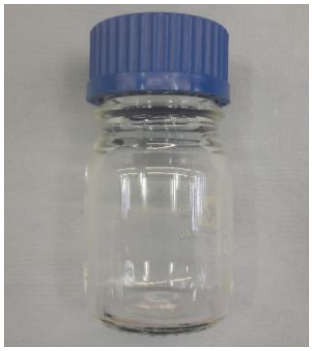
Needles, Sterican, Long Length 120mm

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



Round bottom flasks, 25 ml. (Fisher)

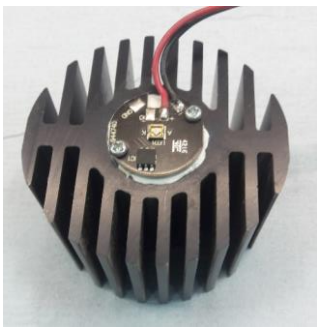
Stirrer bars, PTFE coated, 13x3 mm (Fisher)



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 1 L for solvent and waste

C-Series Syringe Pumps Tricontinent C3000 with 5.0 and 0.5 mL syringes, 4-way non-distribution and 4-way distribution valves.

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



Visible light LED:
Thorlabs
-M565D2, 565nm, 880mW
-M450D3, 450nm, 1850mW
-M405D2, 405nm, 1500mW
Heatsink: 1.8K/W, 60x37.5mm, RS components, 722-6795

1.1 Benchtop MS

Advion expression-CMS

- Dimensions: 66 x 28 x 56 cm
- Weight: 32 kg
- Gas requirements: Nitrogen 98% pure, 4.1 Bar, 8 L/min
- Flow rate range: 10 µL/min to 2 mL/min
- Polarity: Positive & Negative ion switching in single analysis
- *m/z* Range: 10 to 2,000 *m/z*
- 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
- Linear dynamic range of 5 x 10^3

ESI parameters

- Capillary temperature(V) 250.0
- Capillary Voltage(V) 180.0
- ESI Gas Temperature: 250 °C
- ESI Voltage: 3500 V
- Calibration: Agilent ESI Tuning mix G2421A

1.2 Benchtop NMR

Spinsolve 60 Carbon from Magritek

- Frequency: 60 MHz Proton
- Resolution: 50% linewidth < 0.5 Hz
- Lineshape: 0.55% linewidth < 20 Hz
- ^{1}H Sensitivity: >120:1 for 1% ethyl benzene
- Dimensions: 58 x 43 x 40 cm
- Weight: 60 kg
- Magnet: Permanent and cryogen free
- Stray field: < 2 G all around system

The instrument is equipped with a flow-cell to allow online analysis. The cell goes through the instrument and its location places the NMR tube part at the centre of the magnets. Both inlet and outlet are

connected to normal PTFE tubing with screw caps (**Figure S1**). The flow cell allows automatic reaction monitoring in real time by pumping 3 ml of solution from the reaction mixture.

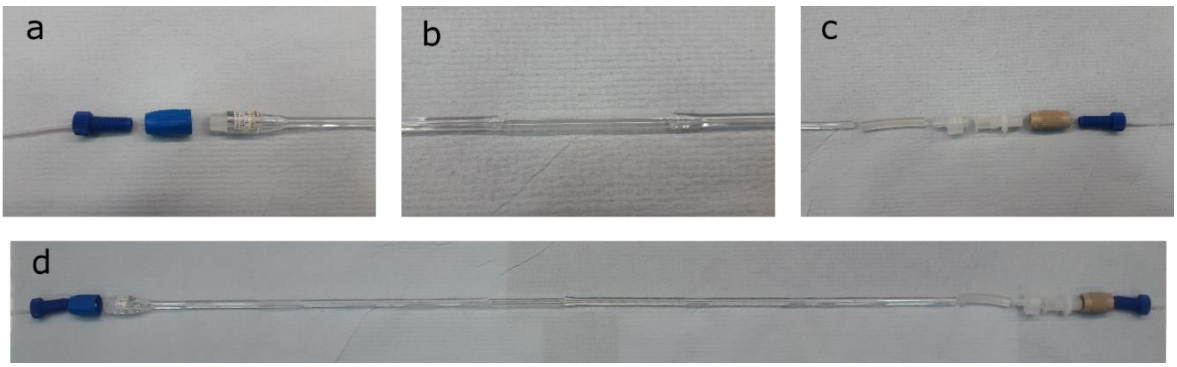


Figure S1: a) top connector, the screw connector has been welded to the glass cell by the glassblower. b) Central part, it has been made by welding a normal 5 mm NMR tube to the rest of the cell. c) bottom connection, the glass could not have the threading because the whole cell needs to fit through the instrument. The connection is made with a chemical resistant tubing followed to a plug to syringe, a syringe to male connector, female to female and finally a normal screwed connector. d) entire NMR flow cell

1.3 Syringe pumps

The control over the fluids was performed using C3000 model, TriContinent™ pumps (Tricontinent Ltd, CA, USA). 5 ml syringes (TriContinent™) were used for all functions except the pumps connected to the MS instrument and the photocatalysts which used a 0.5 ml syringe. The pumps were connected to the computer and each other by a daisy chain with a RS232 serial communication cable and DA-15 connectors. Up to 15 pumps can be connected on the same line and addresses are selected with a physical switch on the back (positions 0 to E, F is used for debugging) (**Figure S2**). Our project used two serial lines for a total of 30 pumps.

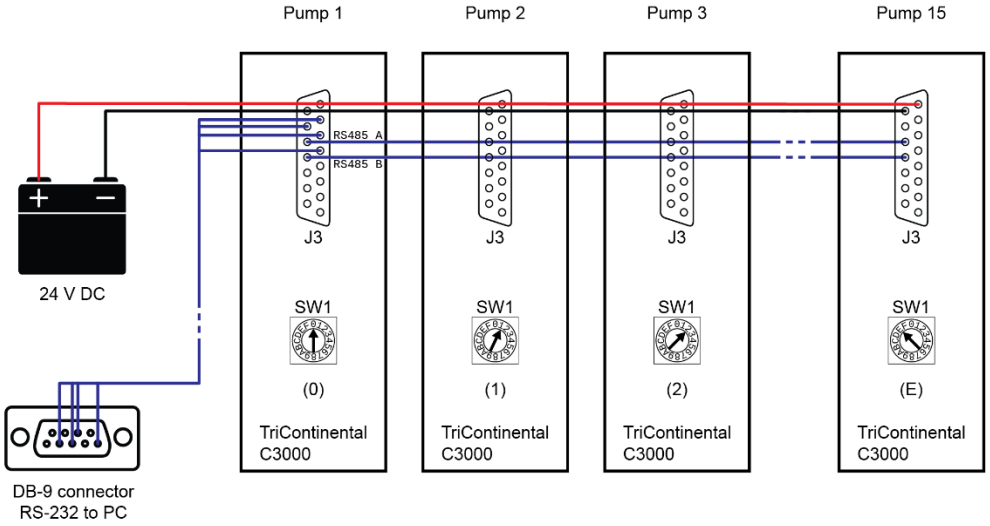


Figure S2: Schematic representation of the daisy chain connection used to control and power the pumps. On a single line up to 15 pumps can be connected, the individual addressed are defined with a physical switch on their back with positions between 0 and E. F position is used for debugging.

The pumps were equipped with distribution (4-port), 90° and 120° (3-port) valves. The distribution valve allows a 3-way connectivity (is it possible to pump a solution from the syringe through one of three ports). 90° and 120° valves allow 2-way connectivity (**Figure S3**).

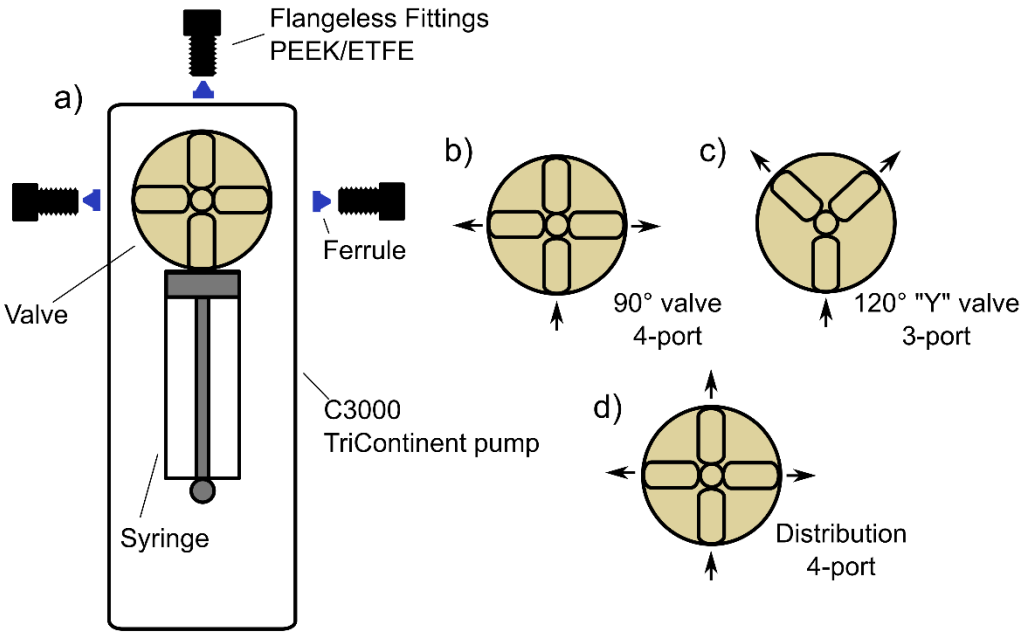


Figure S3: a) Schematic view of the front of a C3000 TriContinent pump. b) Scheme of a 90° valve, the top port is used to bypass the pump, therefore it is possible to pump from the syringe only in two directions. c) Scheme of the 120° valve, it can pump in two directions. d) Scheme of the Distribution valve, it can pump in three directions.

1.4 Building the platform

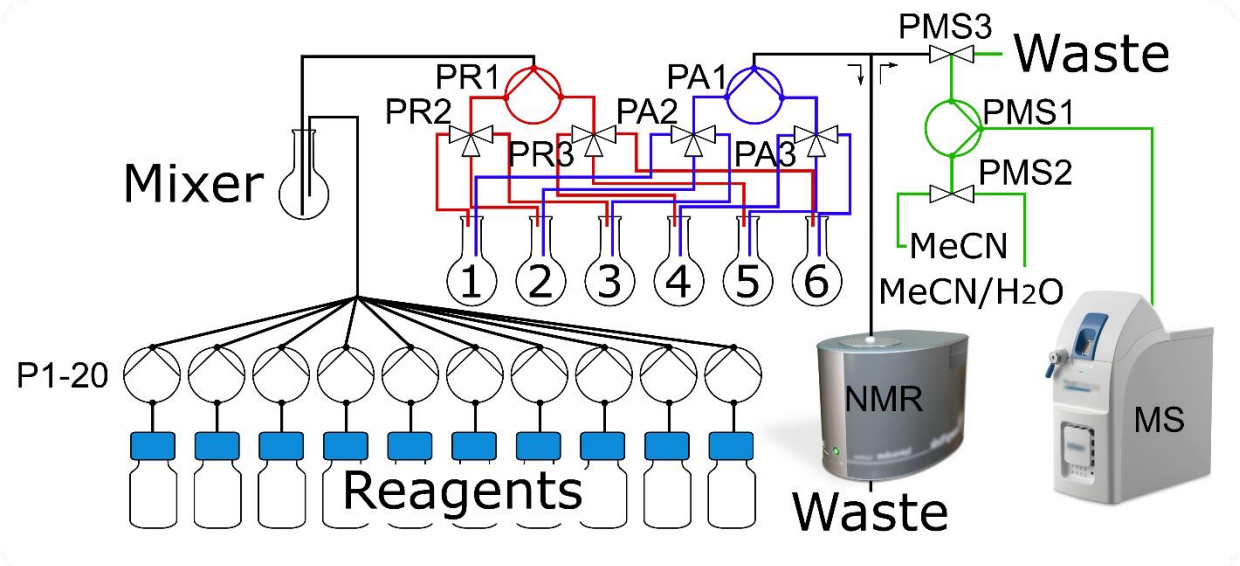


Figure S4: Scheme of the platform.

All fluidic connections to the pumps are made with PTFE, (1/16" OD x 0.8mm ID) tubing and PEEK/ETFE flangeless Fitting (for 1/16" OD Tubing, 1/4"-28 Flat Bottom).

Reagents pumps (P1-20): These pumps are responsible of the addition of reagents, additives and photocatalysts into the mixer flask. Each pump is assigned to a single reagent bottle. They are equipped with 90° or 120° valves. Each pump is connected on one port to a single reagent bottle and on the other to the mixer flask, this connection ends with a Luer to 1/4"-28 Flat Bottom adapter and a long needle in order to go through the septum on the mixer flask.

Reactor pumps (PR1-3): These three pumps manage the transfer of the reaction solution from the mixer flask to one of the six reactors. They are all equipped with 4-way distribution valves and two of them (PR2 and PR3) are used without the syringe to expand the number of ports of the main one (PR1). PR1 is connected from the top port to the mixer flask with a Luer to 1/4"-28 Flat Bottom adapter and a long needle. The other two ports are connected to the syringe ports of PR2 and PR3. The other three ports of PR2 and PR3 are connected to the six reactors with a Luer to 1/4"-28 Flat Bottom adapter and a long needle.

Analysis pumps (PA1-3): These pumps are assembled in the same way of PR1-2-3. They all have 4-way distribution valves. PA2 and PA3 are connected to the six reactors with the three main ports while the syringe ports are connected to the left and right ports of PR1. The top port of PR1 is connected to a 3-way block connector.

3-way block connector: This connects PA1 to the NMR flow cell and PMS3 it is a simple T-splitter.

MS pumps (PMS1-3): They have the same configuration of the reactors pumps. PMS2 and PMS3 have 90° valves and are connected to PMS1 from the syringe ports. PMS3 is connected to the 3-way block connector and to the waste tank while PMS2 is connected to a bottle of acetonitrile (used for dilution) and a bottle containing a mixture of acetonitrile/water 50:50 (used to clean the MS). PMS1 is equipped with a 0.5 ml syringe and is connected through the top port to the MS instrument.

Analysis routine: 4 ml of the reaction mixture is pumped from one of the reactor flasks through the 3-way connector block using PA1 and either PA2 or PA3. Since PMS3 is blocking one of the connector lines this movement results in the solution flowing into the NMR cell. Then PMS1 pumps 0.1 ml of solution from the connector block into its syringe. Again, since PA1 seals the other connection, this movement successfully pumps back the solution from the NMR experiment into PMS1. Finally, the mixture is diluted and pumped to the MS. Both analyses are then started at the same time.

1.5 The chemical space

In the Chemical space 1 explored molecules were selected from a pool of 6 starting materials and mixed in combinations of two and three (**Figure 3b**). 2 ml of each reagent was added from a 1M stock solution. In Chemical space 2 reagents were added in the same way from a pool of 15 compounds (**Figure 3c**). Each reaction had the additional selection of an additive: either a base (4-Dimethylaminopyridine, **23**) or a Lewis acid (scandium triflate, **24**). Furthermore the chemical space involved the presence of one of three molecules known to act as photocatalysts: 2,4,6-triphenylpyrylium tetrafluoroborate (**PC1**), tris(2,2′-bipyridyl)dichlororuthenium(II) hexahydrate (**PC2**) and rose bengal (**PC3**) (**Figure S5**). They were added in 2.5% mol to the reactors irradiated a wavelength corresponding to photocatalyst absorption (405 nm for **PC1**, 450 nm for **PC2** and 565 nm for **PC3**).

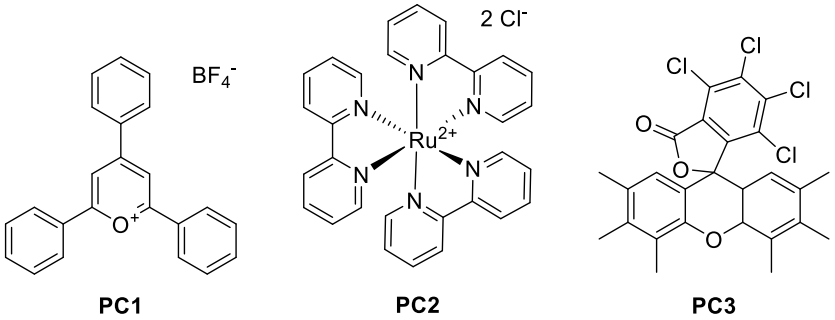


Figure S5: Photocatalysts used as additives to expand the chemical space.

1.6 The light shield

A shielding apparatus was designed and build in order to provide eye protection for the user and to isolate the three reactors from the light of each other. All parts were designed in Lightwave3D and printed on a Stratasys Connex 500 polyjet 3D printer. The apparatus was made of five parts showed as 3D models in **Figure S6**. The base (**c**) is placed on the heating plate for stirring and is furnished with a gap to sledge in the internal walls (**d**). Two holes have also been drilled in the back of the base to accommodate the back wall (**e**). Three heatsinks with the relative LEDs are placed on top of the base separated by the internal walls. The front shield (**a, b**) is then placed on top of the heatsinks. All parts were modelled around the measurements of clamps, heatsinks, and flasks, therefore the shield gently bends around the light emitters avoiding the light to shine outside or irradiate the wrong reactor. A designed cover (**f, g**) is placed on top of the front shield in order to block the light coming out from the top, as a result of a 0.5 cm spline it can sit firmly on top of the front shield.

All parts were printed in a polymer material called *verowhite;* a photosensitive polymer liquid that is solidified by UV light layer by layer. Since the final object presented a slight transparency, the internal part of the front shield was covered with aluminium foil. This solution assures that the LEDs light will reflect back into the reactors (**Figure S7**).

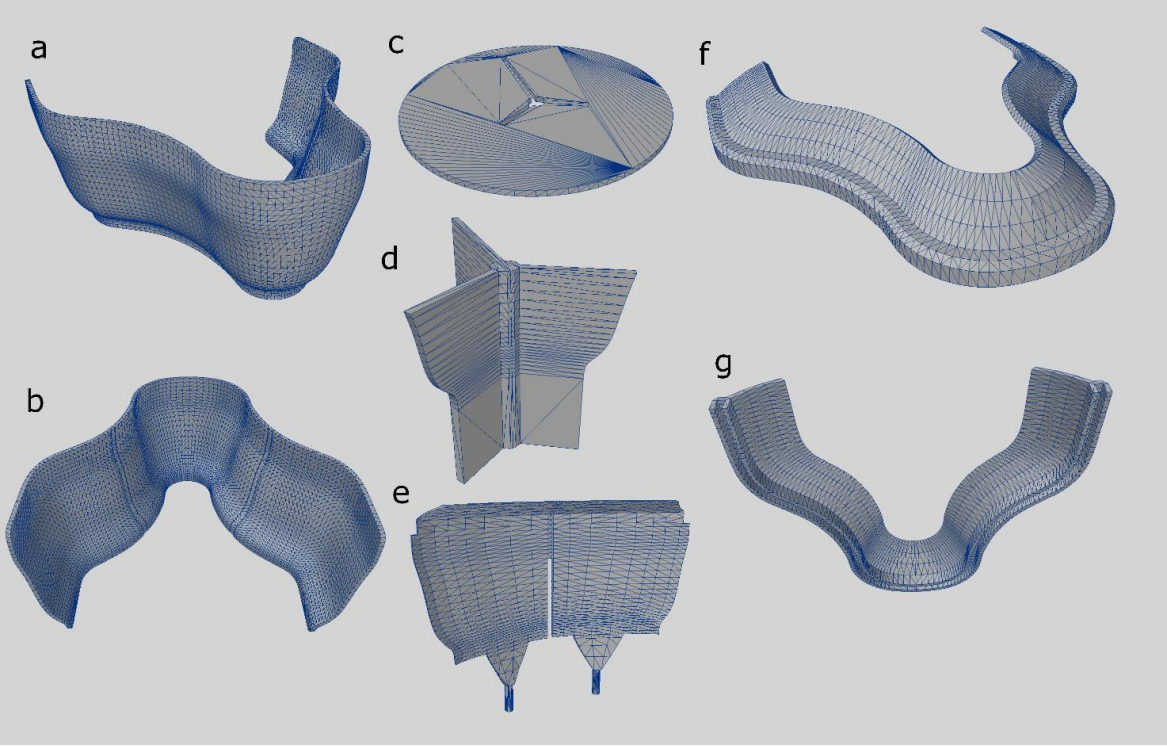


Figure S6: a,b: front shield, the internal part was covered with aluminium foil. **c**: base of the apparatus, a narrow gap in the centre allows to slot in the internal walls. **d**: internal walls, they isolate the three reactors from the unwanted LEDs light. **e**: backwall, it confines the light coming out from the back. **f, g**: cover lid, it blocks the light coming from the top.

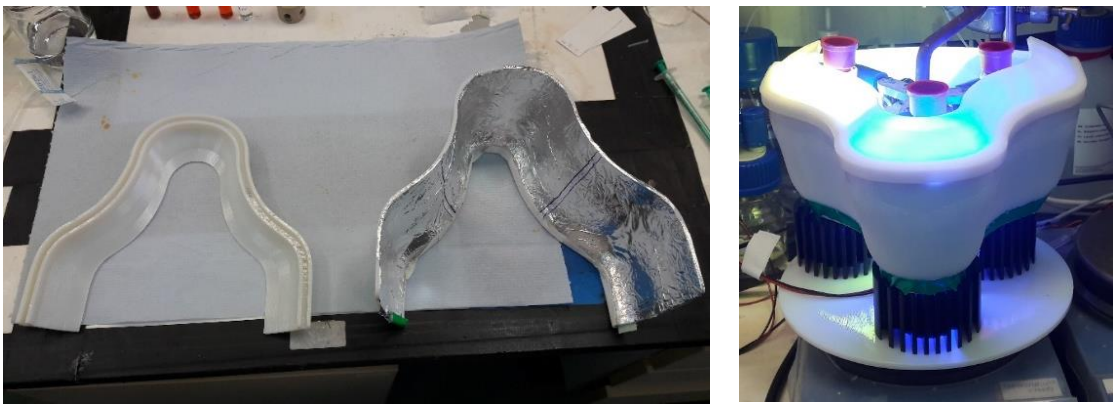


Figure S7: *Left*: front shield covered with aluminium foil and its cover. *Right*: Full setup. The LEDs are mounted on the heatsinks and placed on a stirring plate. Reactors are held on top and isolated by the 3D printed shield.

2 Neural network for reactivity assignment

2.1 Manual reactivity assignment

NMR data was checked manually and a reactivity value was assigned between 0 and 3. To do this the mixture spectrum was compared with the superimposition of the starting materials' spectra and the criteria for the assignment are based on the appearance of new peaks, their intensity, peaks shifting and reagents peaks disappearing. Although there were borderline cases between two values some general guidelines were followed: a) absolutely no difference or a slight shift = 0; b) one peak appearing or a big shift, medium intensities = 1; c) two or three peaks appearing in high intensity = 2 and d) more than three peaks appearing with a high intensity = 3. Examples of real NMR data with their manual evaluation are reported in **Figure S8**.

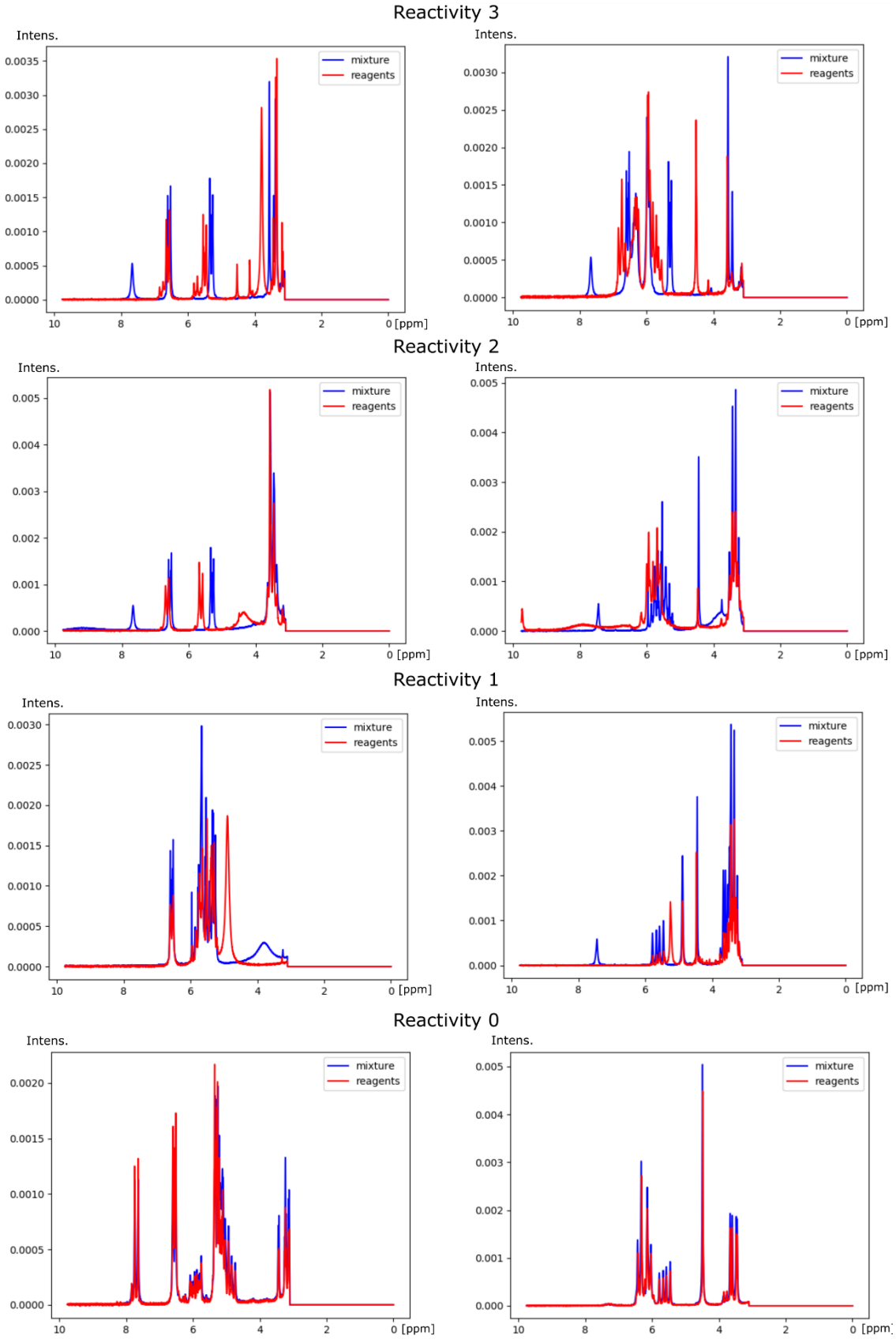


Figure S8: Examples of NMR spectra evaluated manually. By looking at the new peaks appearing and the reagents peaks disappearing a reactivity between 0 and 3 is assigned.

2.2 The Reactify neural network

NMR spectra were resampled to rescale them from 4878 to 271 points. They were then normalized to 1 and the solvent peak was removed by cutting the spectrum at 3 ppm. In order to avoid overfitting a random scaling (y-axis) and shifting (x-axis) was applied on both the mixture spectrum and the reagents superimposition during training (**Figure S9**). True values classes were normalized to 1, meaning that original values from 0 to 3 corresponded to 0, 0.33, 0.66 and 1.

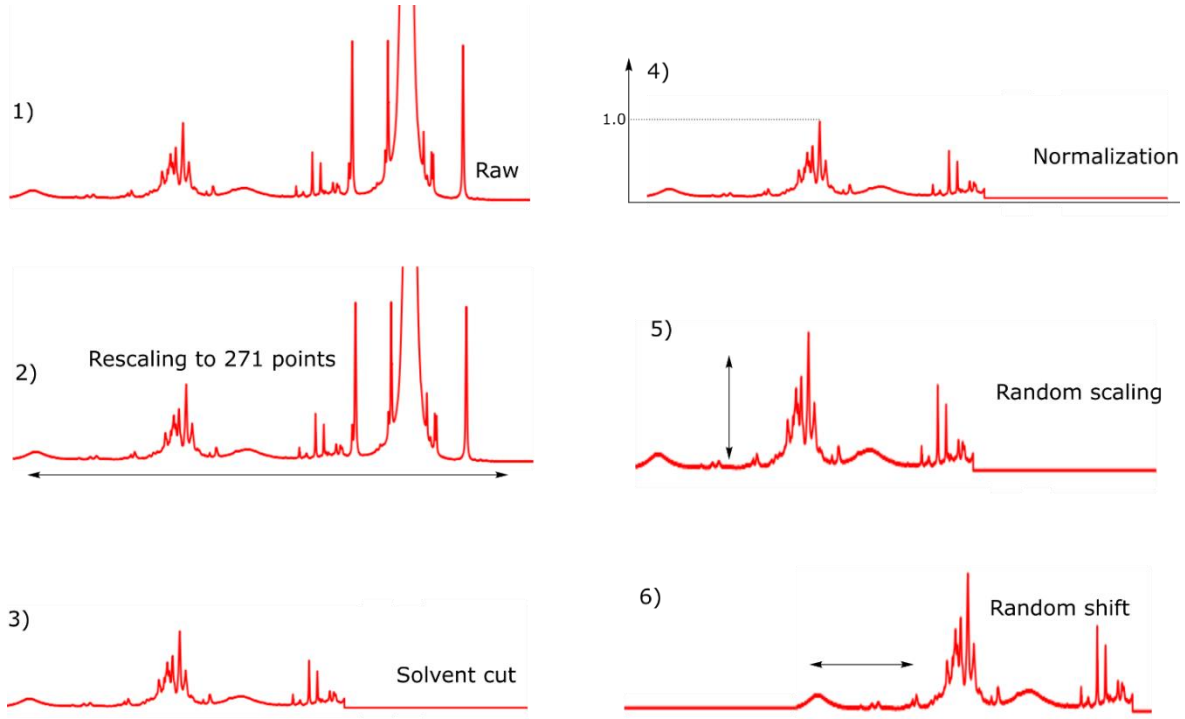


Figure S9: Pre-processing performed on the NMR spectra before using them as input. The raw spectrum (1) is rescaled to 271 points (2) and the solvent is cut out (3). It is then normalized to the highest peak (4) and a random factor on both axes is introduced during the network training to avoid overfitting (5,6)

2.3 Neural network structure

The dataset was fed into the network and each experiment was represented as a 271 x 2 matrix where the first row corresponded to the intensity values of the mixture spectrum while the second one the reconstructed spectrum obtained by superimposition the starting materials spectra. The structure of the network is reported in Figure 3 of the manuscript. The tuning of parameters were achieved with a grid search, where the network is retrained multiple times with different combinations. The network was trained with 250 epochs and early stopping to avoid overfitting. Training data consisted in 440 NMR spectra from chemical space 1, split into training/validation as 0.9/0.1. Loss was calculated with cross entropy function and minimized using the AdamOptimizer algorithm (1). The network code is based on Tensorflow (2) python library.

3 Reactivity prediction network

The dataset was fed into the network by encoding the reagents involved into SMILES. The auto encoder developed by Jin et all converted the SMILES of each reagent into a vector of 56 floating values. The other reaction variables were encoded separately as one-hot encoded vector. The dataset was expanded by repeating the reactions entries and changing the order of reagents. The structure of the network is reported in Figure 4 of the manuscript. Tuning of parameters was achieved with grid search. Training data consisted in 1018 reactions from the chemical space 2, split into training/validation as 0.8/0.1. Loss was calculated with cross entropy function and minimized using the AdamOptimizer algorithm (1). The network code is implemented using Tensorflow (2) library.

3.1 Performance and uncertainty estimation

As only a small part of chemical space is explored in a single experiment, the network uses a large degree of *dropout* to avoid overfitting and favour the recognition of general patterns. Consequently, dropout can be applied during inference to obtain a Bayesian estimate of the network's predictions (3). In this case, we sampled 10 different output values for each given input. Figure S10 compares the reactivities predicted by the network with the expected values. In this case the expected reactivities are not manually assigned and instead are the result of passing the corresponding NMR spectra data through the reactivity assignment network (Reactify, section 2). Figure S11 shows that the uncertainties obtained through Bayesian interpretation of dropout do not correlate strongly with the error in predicted reactivity.

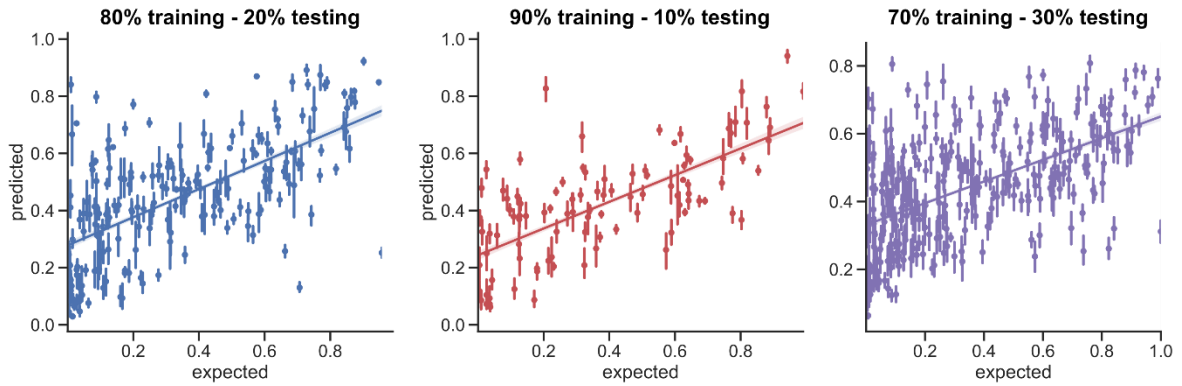


Figure S10: Performance of the reactivity prediction network on chemical space 2 (photochemical) using different train-test ratios. Vertical bars on each point show 99% confidence intervals obtained by bootstrapping of 10 model predictions.

Table S1 compares the accuracy of our model for various train–test ratios. We report mean squared error (MSE) and mean absolute error (MAE) for various train–test ratios.

	Chemical space 1		Chemical space 2	
	MSE	**MAE**	**MSE**	**MAE**
Model (80% train–20% test)	0.031	0.14	0.052	0.19
Model (80%–20%, weighted)	0.034	0.15	0.048	0.18
Model (90%–10%)	0.028	0.12	0.046	0.18
Model (90%–10%, weighted)	0.025	0.11	0.049	0.19
Model (70%–30%)	0.028	0.13	0.058	0.19
Model (70%–30%, weighted)	0.029	0.13	0.055	0.19
Model (50%–50%)	0.025	0.13	0.055	0.19
Model (50%–50%, weighted)	0.029	0.13	0.052	0.19
Model (10%–90%)	0.073	0.24	0.071	0.23
Model (10%–90%, weighted)	0.71	0.23	0.079	0.23
Baseline*	0.069	0.23	0.080	0.24
Baseline (weighted)	0.079	0.24	0.11	0.29
Dummy‡	0.19	0.34	0.22	0.38
Dummy (weighted)	0.26	0.42	0.36	0.54

* The baseline model is defined as one producing a constant prediction equal to the average reactivity of the training set (with 90% of data being used for training).

‡ The dummy model is defined as one producing a constant prediction equal to 0 (unreactive; the most abundant class) for all observations.

Table S1: Performance evaluation of reactivity prediction network for various train-test split ratios compared against a baseline constant prediction equal to the average reactivity as well as a dummy constant prediction equal to 0 (the most abundant class). Chemical space 1 was used for hyperparameter selection.

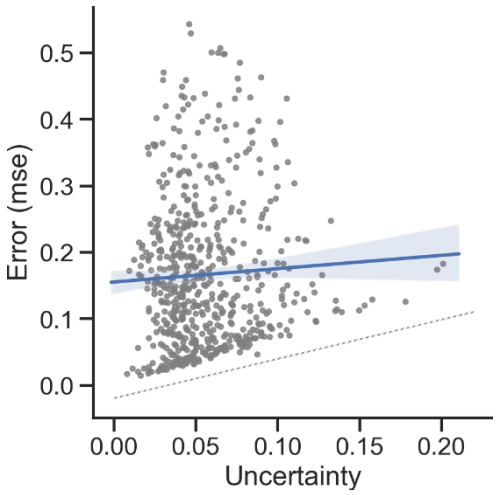


Figure S11: Error-uncertainty correlation for the reactivity prediction network. Uncertainties are calculated as the standard deviation of predictions when dropout layers are activated during inference (see reference (3) for using dropout as Bayesian approximation). Shaded area shows the 95% confidence interval obtained by bootstrapping (1000x) of data points. The amount of uncertainty does not seem to be strongly correlated with prediction error overall (Spearman correlation coefficient 0.05) but is an accurate lower bound estimator of error (dashed line provided as visual aid). Network was trained on 90% of available data in chemical space 2. Predictions for all 6 permutations of 3 inputs are shown. Reaction discovered and variations

4 Reaction discovered and variations

4.1 Diethyl 2-bromomalonate and p-toluenesulfonylmethyl isocyanide.

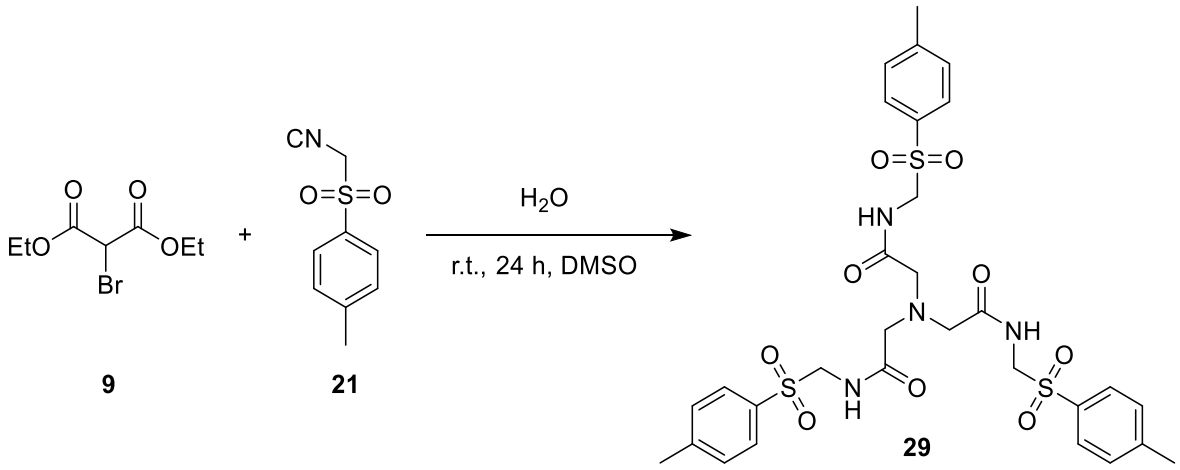


Figure S12: Scheme of the reaction yielding product **29**.

Diethyl 2-bromomalonate (2 mmol, 0.41 ml), p-toluenesulfonylmethyl isocyanide (2 mmol, 0.39 g) and water (0.8 mmol, 15 µl) are mixed in 4 ml of anhydrous DMSO and stirred for 24 hours at room temperature. The reaction mixture is diluted with water (20:1) and extracted with ethyl acetate. The organic phase is separated and washed with brine. Mg_2SO_4 is then added to the reaction mixture and after filtration the solvent is removed under reduced pressure. During the evaporation of ethyl acetate, the product precipitates as a white solid, it is isolated by filtration and washed with ethyl acetate.

IUPAC name: N,N,N-tris(2-(4-methylphenylsulfonyl)acetamide)amine
Yield: 47% (108 mg)
^{1}H NMR (600 MHz, DMSO-d_6) δ 8.96 (t, J = 6.7 Hz, 1H), 7.68 (d, J = 8.2 Hz, 2H), 7.39 (d, J = 8.0 Hz, 2H), 4.66 (d, J = 6.7 Hz, 2H), 2.97 (s, 2H), 2.38 (s, 3H).
^{13}C NMR (151 MHz, DMSO) δ 170.04, 144.69, 134.39, 129.73, 128.48, 60.00, 56.48, 21.07.
ESI-HR-MS: $[C_{30}H_{36}N_4O_9S_3Na]^+$ Calculated 715.1537 *m/z*, measured 715.1503 *m/z*

4.2 1H-benzotriazol-1-ylmethyl isocyanide and diethyl 2-bromomalonate

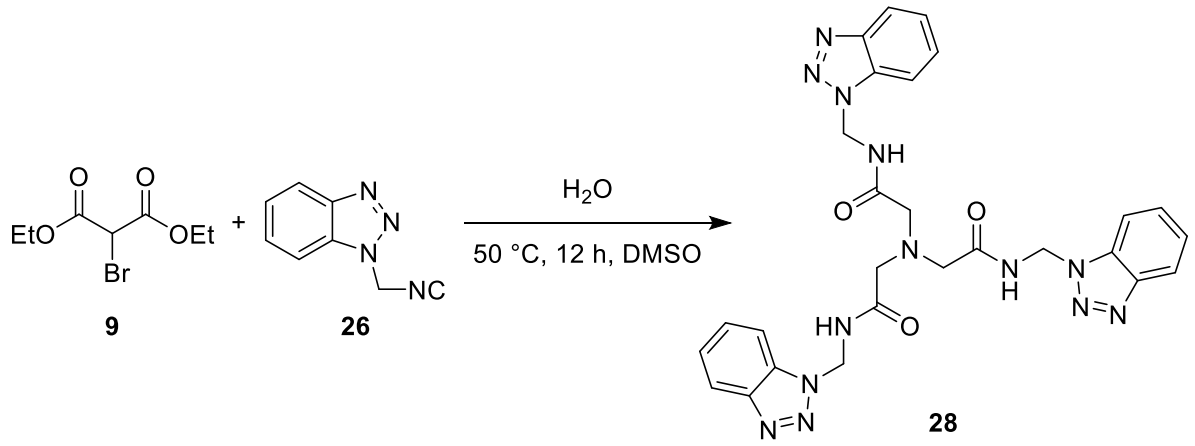


Figure S13: Scheme of the reaction yielding product **28**

Diethyl 2-bromomalonate (1.89 mmol, 0.32 ml), 1H-benzotriazol-1-ylmethyl isocyanide (1.89 mmol, 0.3 g) and water (water (0.76 mmol, 14 µl)) are mixed in 4 ml of anhydrous DMSO and stirred at 50 °C overnight. The reaction mixture is diluted with water (20:1) and extracted with ethyl acetate. The organic phase is separated and washed with brine. Mg_2SO_4 is then added to the reaction mixture and after filtration the solvent is removed under vacuum. The crude is purified with a (silica gel) chromatographic column, gradient elution was used: EtOAc/hexane 1:1 - EtOAc 100% - EtOAc /methanol 24:1.

IUPAC name: N,N,N-tris(N-(Benzotriazol-1-ylmethyl)acetamide)amine
Yield: 31% (57 mg)
^{1}H NMR (600 MHz, $CDCl_3$) δ 8.93 (s, 3H), 7.77 (dd, *J* = 36.8, 8.4 Hz, 6H), 7.38 (t, *J* = 7.7 Hz, 4H), 7.23 (t, *J* = 7.7 Hz, 3H), 5.81 (d, *J* = 6.6 Hz, 6H), 3.39 (s, 6H).
^{13}C NMR (151 MHz, $CDCl_3$) δ 133.79, 107.78, 94.54, 90.27, 86.75, 81.52, 72.87, 21.40, 12.93.
ESI-HR-MS: $[C_{27}H_{27}N_{13}NaO_3]^+$ calculated 604.2252 m/z, measured 604.2232 *m/z*.

4.3 Silyl isocyanide and diethyl 2-bromomalonate

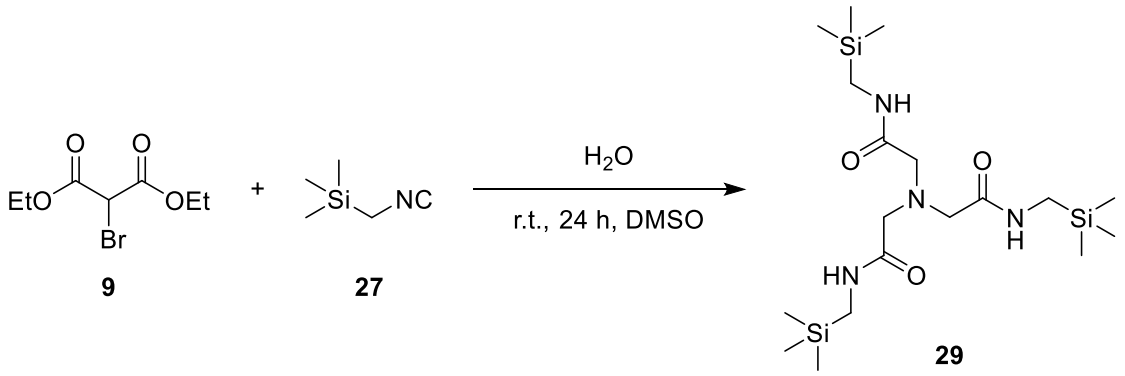


Figure S14: Scheme of the reaction yielding product **29.** The product is only observed in traces in HPLC-MS.

Diethyl 2-bromomalonate (2 mmol, 0.41 ml) and (Trimethylsilyl)methyl isocyanide (2 mmol, 0.39 g) are mixed in 4 ml of DMSO and stirred for 24 hours at room temperature. It was not possible to isolate the product. However HPLC-MS analysis of the mixture showed traces of a molecule with mass 447.2696, the expected mass for $[M+H]^+$ is 447.2623.

4.4 Variations of activators

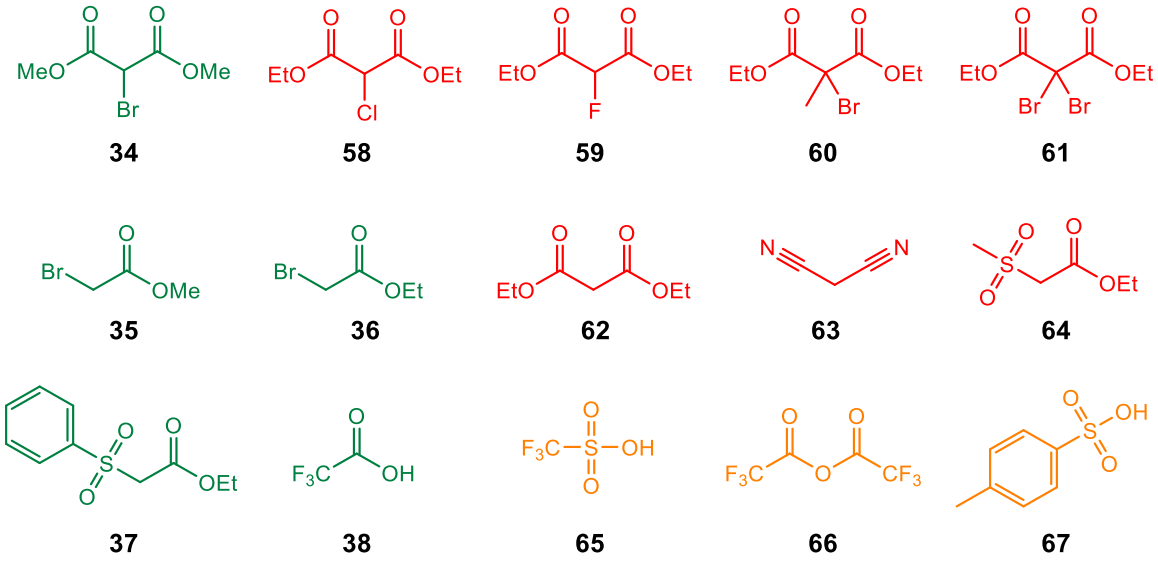


Figure S15: List of molecules tried as activator replacing diethyl bromo malonate.

The reaction has been repeated according to the conditions indicated in section 4.1, replacing the diethyl bromomalonate with various analogous molecules and organic acids. The molecules coloured in green (**34**, **35**, **36**, **37** and **38**), yielded product **29** as white precipitate. The molecules coloured in red (**58, 59, 60, 61, 62, 63, 64**) did not produce the desired product. In the reaction with molecule **65**, **66** and **63** the product was observed in the HPLC analysis but did not precipitate. The reaction with molecule **67** produced molecule **68**, while molecule **38** produced a mixture of product **29** and **69** (**Figure S16**)**.** The synthesis of molecules **68** and **69** from TosMIC as already been reported in literature (4).

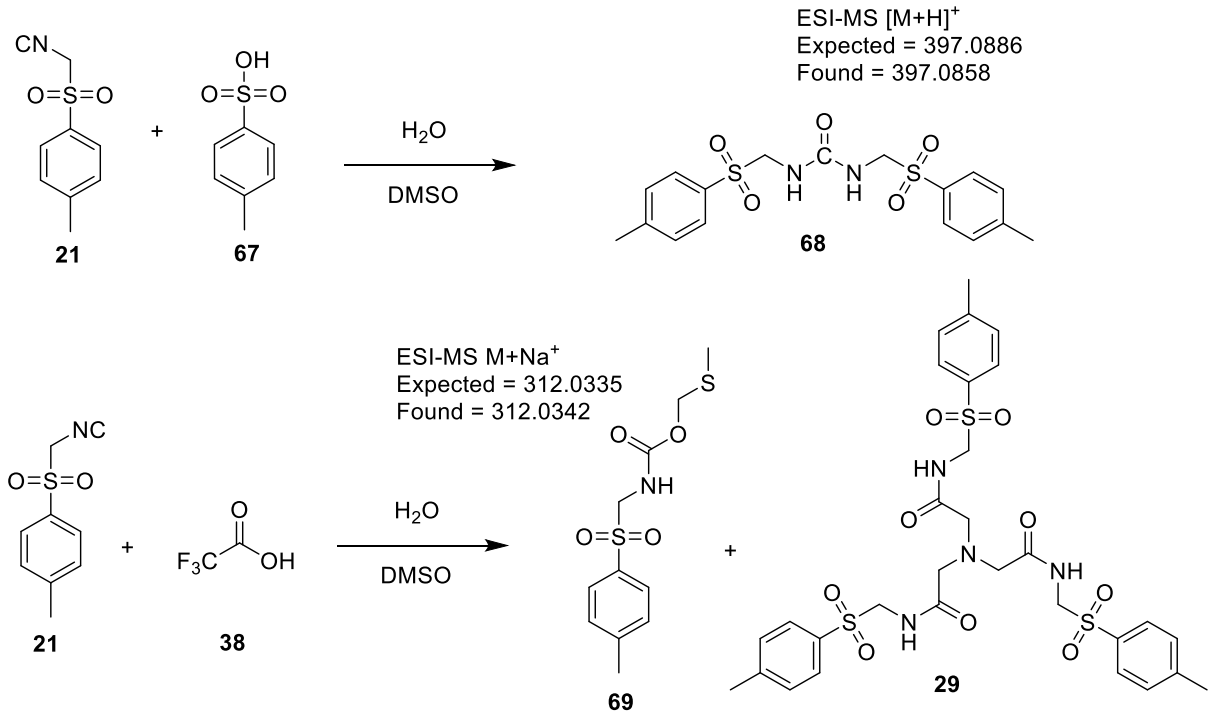


Figure S16: Side products obtained while testing PTSA (**63**) and TFA (**34**) as activators. The reactions yielding products **68** and **69** are known in literature.

Compound **68**

^{1}H NMR (600 MHz, DMSO-d_6) δ 7.66 (d, J = 7.9 Hz, 4H), 7.42 (d, J = 7.9 Hz, 4H), 7.14 (t, J = 6.7 Hz, 2H), 4.54 (d, J = 6.6 Hz, 4H), 2.40 (s, 6H).

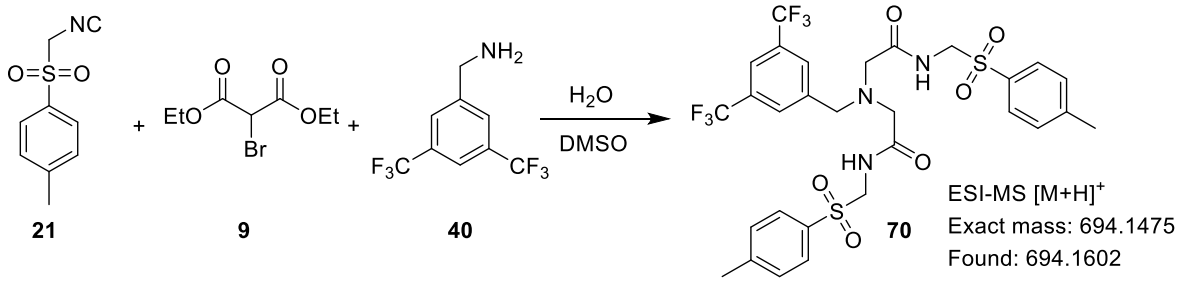
Compound **69**

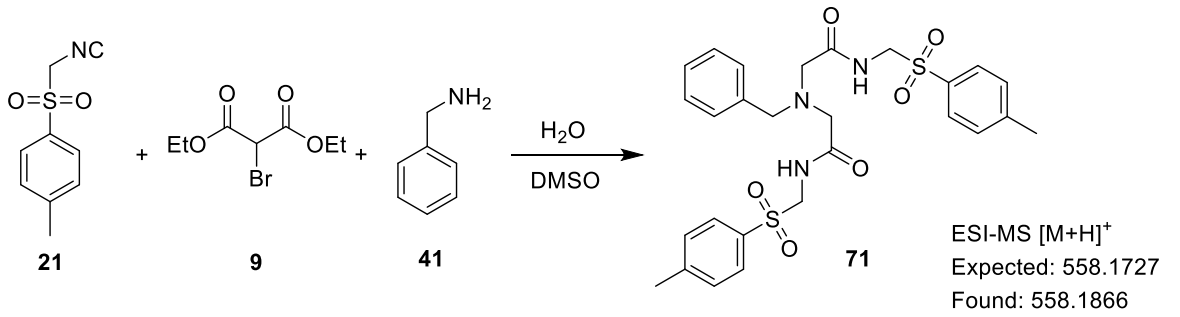
^{1}H NMR (600 MHz, DMSO-d_6) δ 8.51 (t, J = 6.8 Hz, 1H), 7.71 (d, J = 7.9 Hz, 2H), 7.44 (d, J = 7.9 Hz, 2H), 5.03 (s, 2H), 4.55 (d, J = 6.7 Hz, 2H), 2.41 (s, 3H), 2.09 (s, 3H).

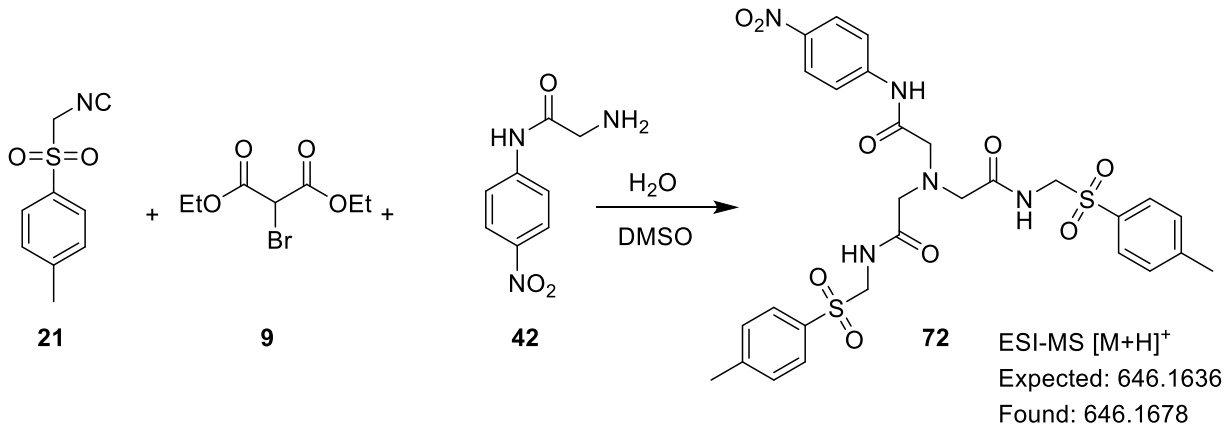
^{13}C NMR (151 MHz, DMSO) δ 154.92, 144.31, 134.27, 129.50, 128.16, 68.24, 62.23, 20.82, 14.17.

4.5 Asymmetric variations of product 25

The reaction has been repeated with the same conditions indicated in section 4.1 adding one equivalent of amine (**40**, **41**, **42**, **44, 45, 46**). The reaction mixtures were sampled after 24 h and analysed in the HPLC-MS. In all cases the expected product was observed. The masses found are reported in **Figure S17**.







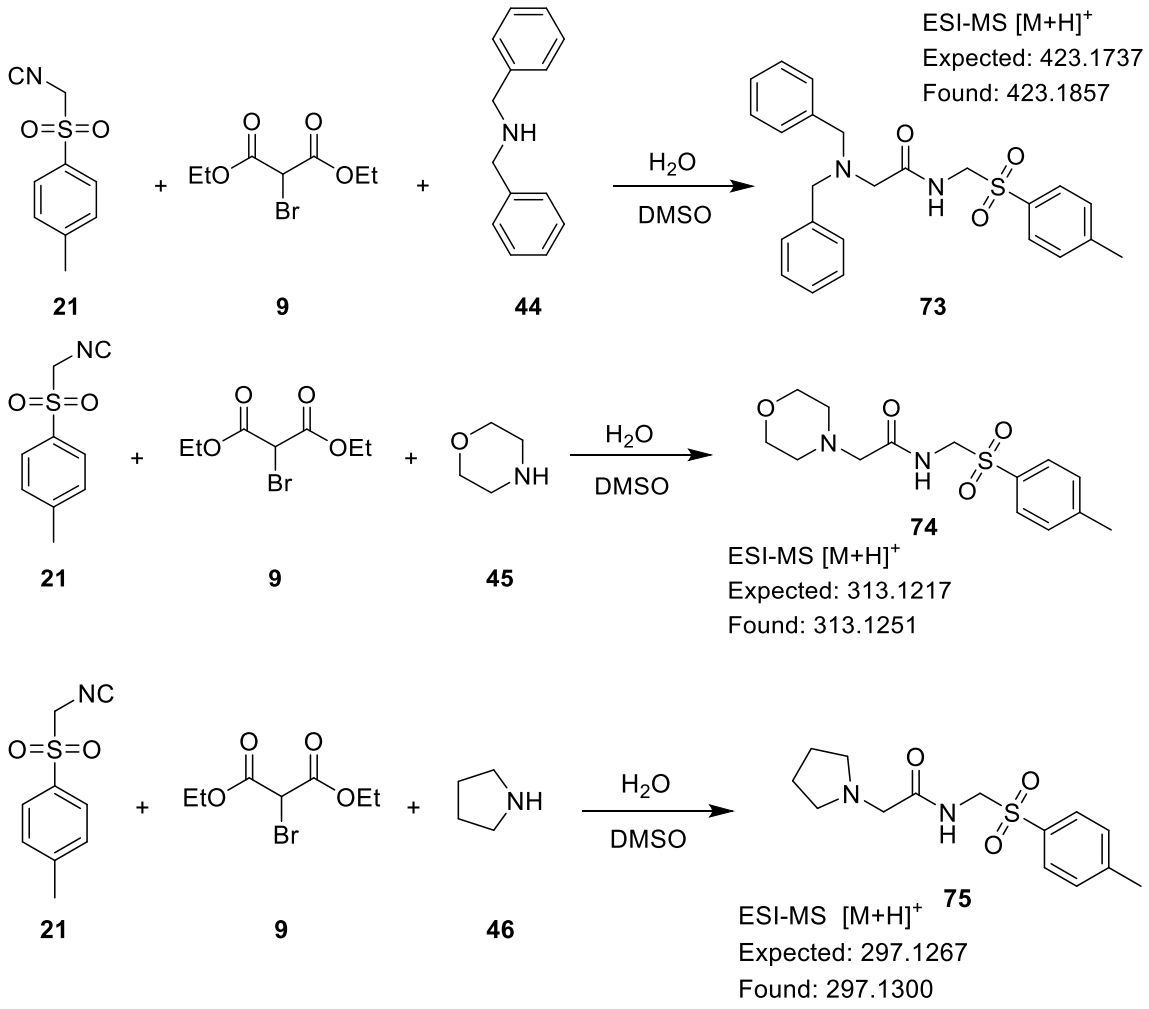


Figure S17: Scheme of the reaction variations involving the presence of a primary or secondary amine. The analogs of the mechanism intermediates have been detected by HPLC-MS.

4.6 Other Isocyanide variations

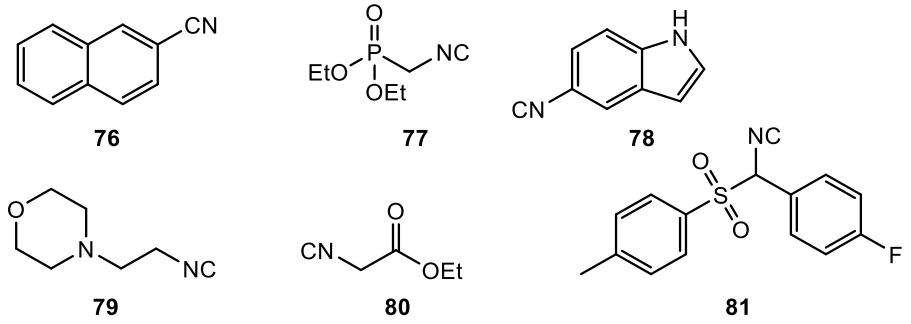


Figure S18: Other isocyanide variations tried on the reaction. None of them yielded product **29.**

The reaction has also been tried on the isocyanides reported in **Figure S18** (2-Naphthyl isocyanide **76**, Diethyl isocyanomethylphosphonate **77**, 1H-Indol-5-yl isocyanide **78**, 2-Morpholinoethyl isocyanide **79**, Ethyl isocyanoacetate **80**, α-(p-Toluenesulfonyl)-4-fluorobenzylisonitrile **81**. The reactions were prepared according to the procedure reported in section 4.1 and repeated at 50°C. HPLC-MS analysis showed no traces of the analogues of product **29**. The analysis of molecule **81** showed the presence of the one-branched (**82**) and two-branched (**83**) analogues suggesting that the intermediate is too hindered to form the third branch (**Figure S19**).

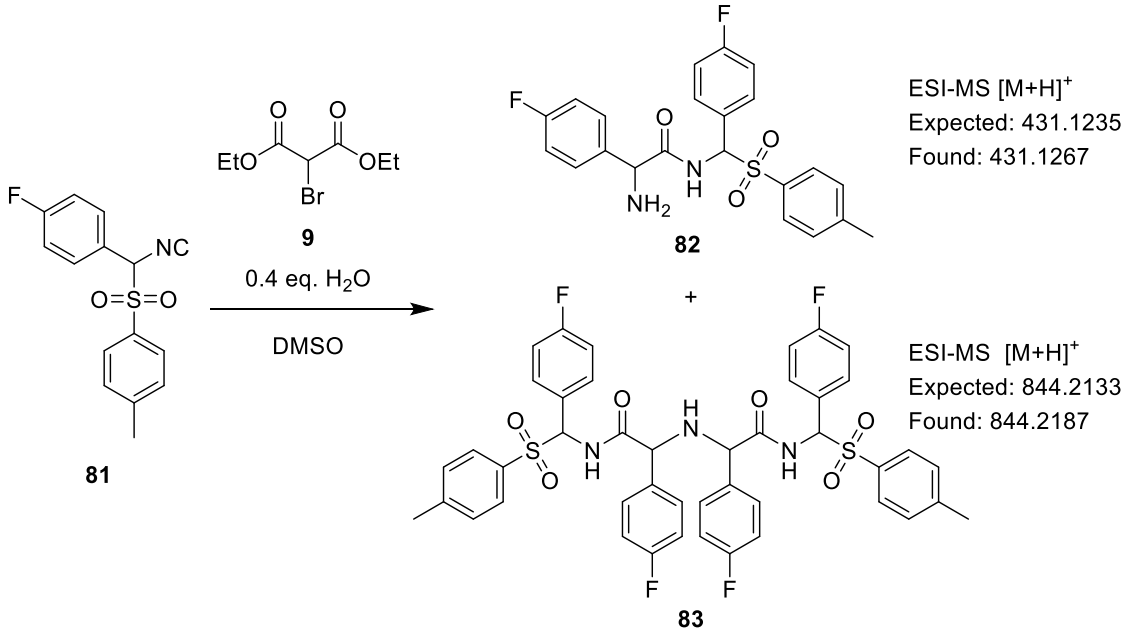


Figure S19: Isocyanide **81** did not produce the expected analogue of **29**. However, the one-branched (**82**) and two-branched (**83**) versions were detected.

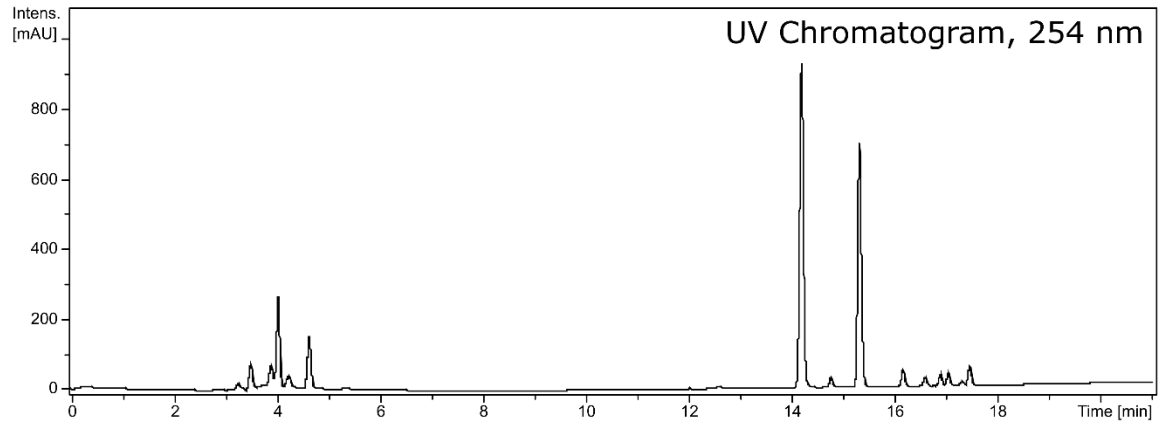
5 Study of reaction mechanism

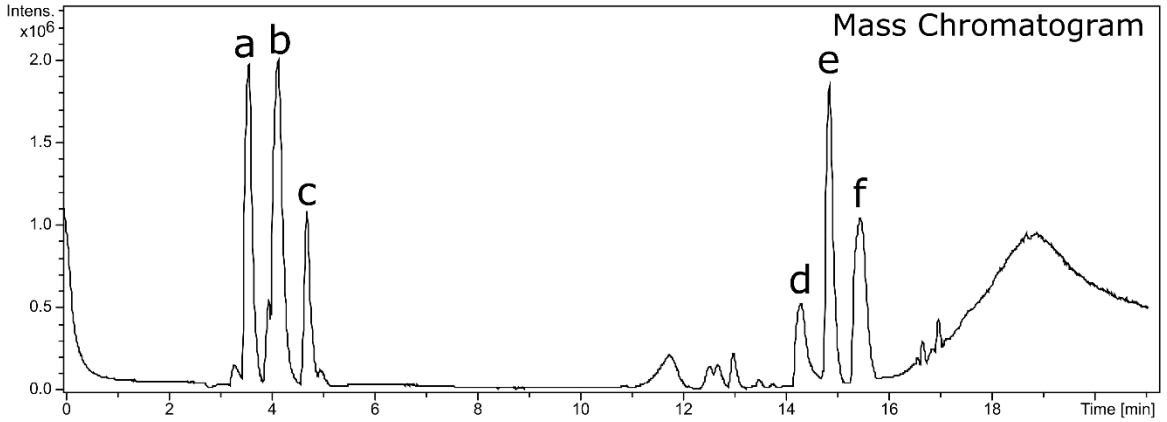
5.1 HPLC-analysis

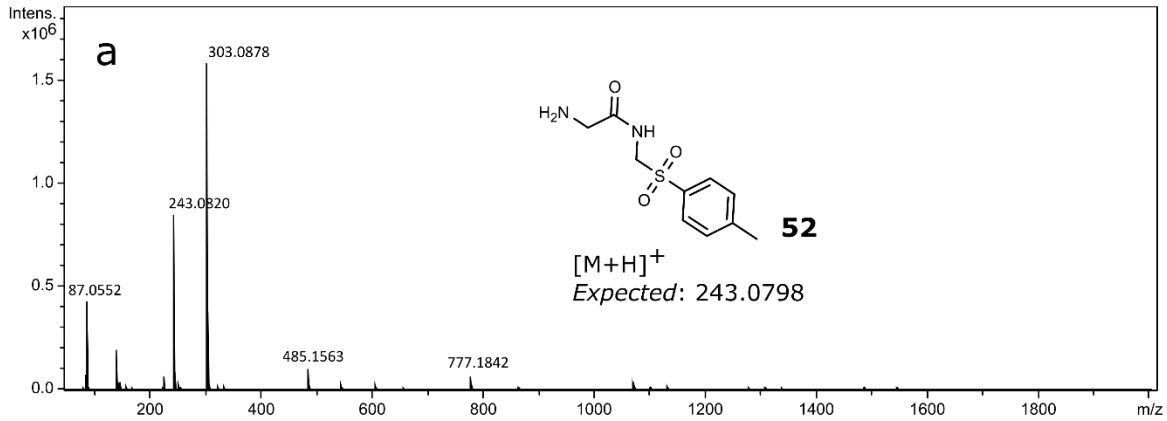
Chromatographic separation of the reaction mixture was achieved with a reverse phase column by Agilent (Poroshell 120 HPH C18, 3.0 x 100 mm, 2.7 µm) on a Thermo Fisher UltiMate 3000 HPLC. From the reaction mixture 45 µL were sampled and diluted in 1ml of acetonitrile (0.022M final concentration). 10 µL of the sample were then injected in the instrument and eluted with a linear gradient mixture of solvents: water w/0.1% v/v formic acid and acetonitrile w/0.1% v/v formic acid at 0.5 mL per minute, over 21 minutes as indicated in **Table S2**. The column compartment was maintained at 30 °C. Results after 12h of reaction are showed in **Figure S20**. As reference the structure and the calculated mass of the compounds are reported. UV detection was performed using a diode array detector (DAD) set on 254 nm. The MS apparatus was a Bruker MaXis Impact instrument, acquisition range at 50–2000 *m/z*. Data was analysed using the Bruker DataAnalysis software suite.

Table S2: HPLC method used for the analysis of the reaction mixture.

Time [min]	Water [%]	Acetonitrile [%]
0	95	5
2	95	5
15	5	95
18	5	95
20	95	5
21	95	5

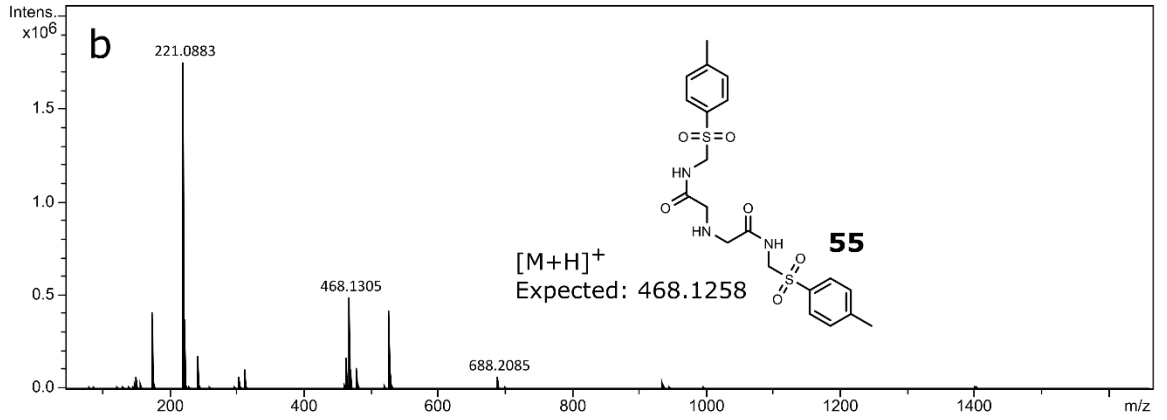






52

$[M+H]^+$
Expected: 243.0798



55

$[M+H]^+$
Expected: 468.1258

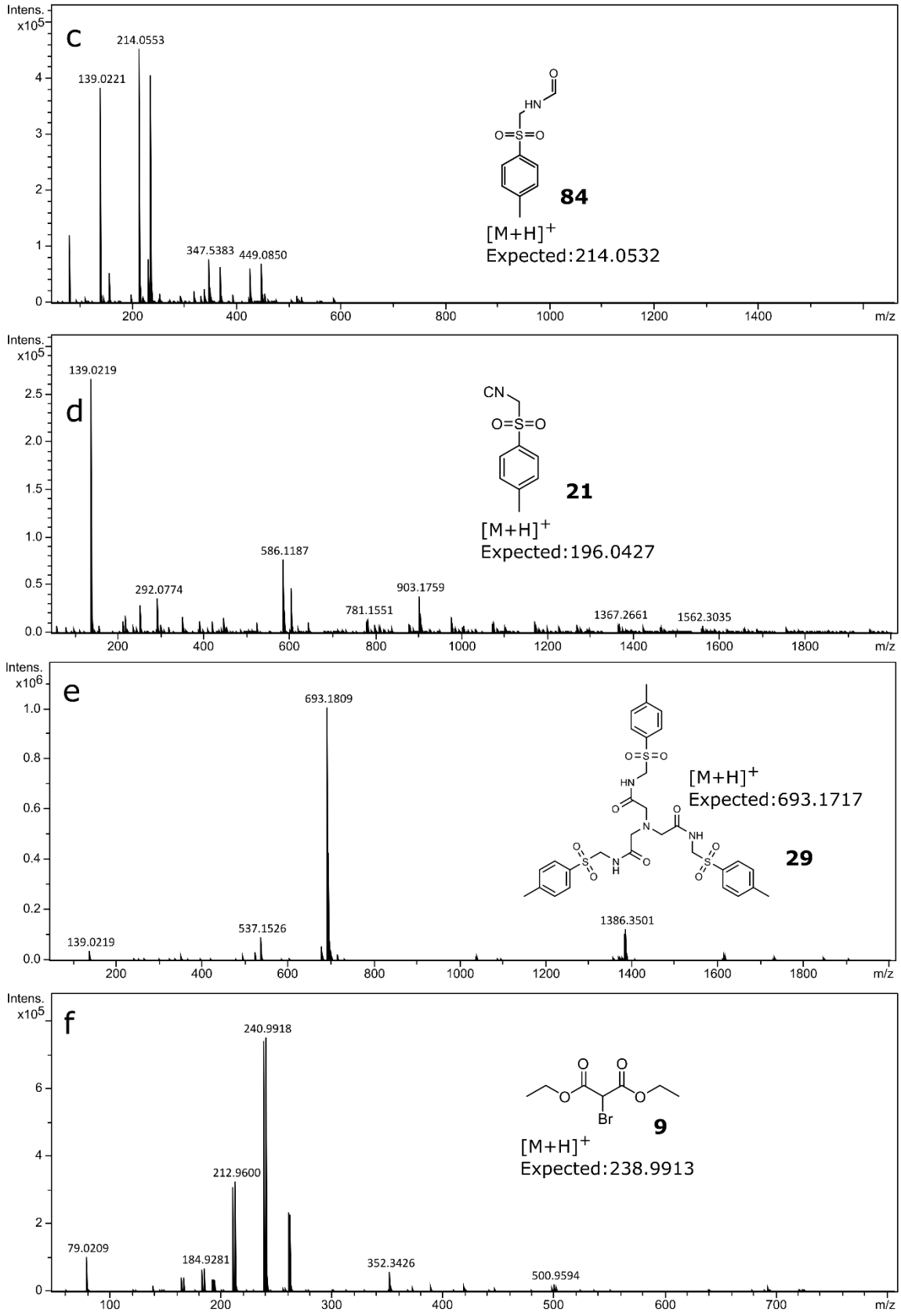
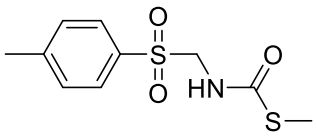


Figure S20: HPLC-MS analysis of the reaction mixture. The MS data of the major peaks are reported with the assigned molecules

5.2 Time resolved HPLC analysis

The reaction was monitored by automatically injecting a sample in the HPLC instrument (agilent 1100 equipped with DAD set on 254 nm) every two hours for 50 h. By using a simple liquid handling platform 0.3 ml were sampled from the reaction mixture and moved to a flask using a Tricontent pump. 6.5 ml of MeCN were added to the flask for dilution and the solution was pumped into a loop valve (Rheodyne, part number: MXP7920-000, equipped with a 10 µl sample loop) connected to the instrument and remotely controlled. Once the valve was switched the HPLC was triggered using a contact closure and the method was run. Results of this experiment are showed in **Figure S22**. The integrals of the main 6 peaks have been extracted and plotted vs time (**Figure 6d** in the manuscript). Results show that over 35 h TosMIC get completely consumed, the dimeric product **55** forms reaching its max at 20 h and then gets consumed while the integrals of the final product **29** and of an unknown product appear (RT: 6 minutes). The molecule corresponding to the peak at 6 minutes has been isolated with flash column chromatography and characterised with NMR and MS. It corresponds to the product **85** in **Figure S21**, an adduct of DMSO and TosMIC not involved in the proposed mechanism.



85

ESI-MS $[M+H]^+$
Expected: 260.0410
Found: 260.0409

Figure S21: Structue of the side product corresponding to the peak 3 in **Figure S22**. It is belived to not be involved in the mechanism.

^{1}H NMR (600 MHz, $CDCl_3$) δ 7.80, 7.79, 7.78, 7.78, 7.36, 7.35, 6.26, 6.25, 6.24, 4.66, 4.65, 2.45, 2.22.
^{13}C NMR (151 MHz, $CDCl_3$) δ 145.48, 133.68, 129.93, 128.85, 59.10, 24.18, 21.67, 12.31.

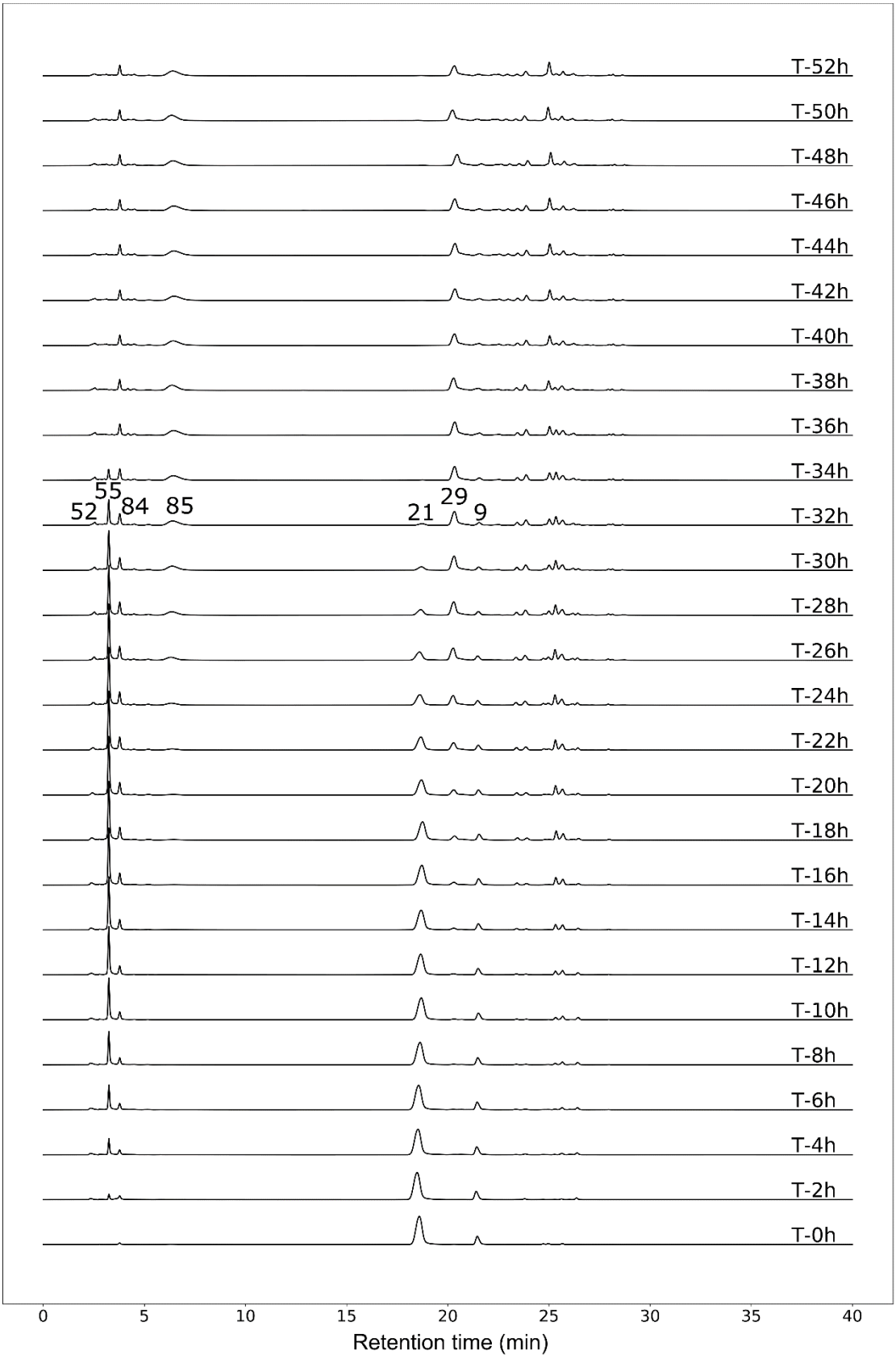


Figure S22: Online HPLC analysis of the reaction mixture. The known peaks are marked with the molecule number.

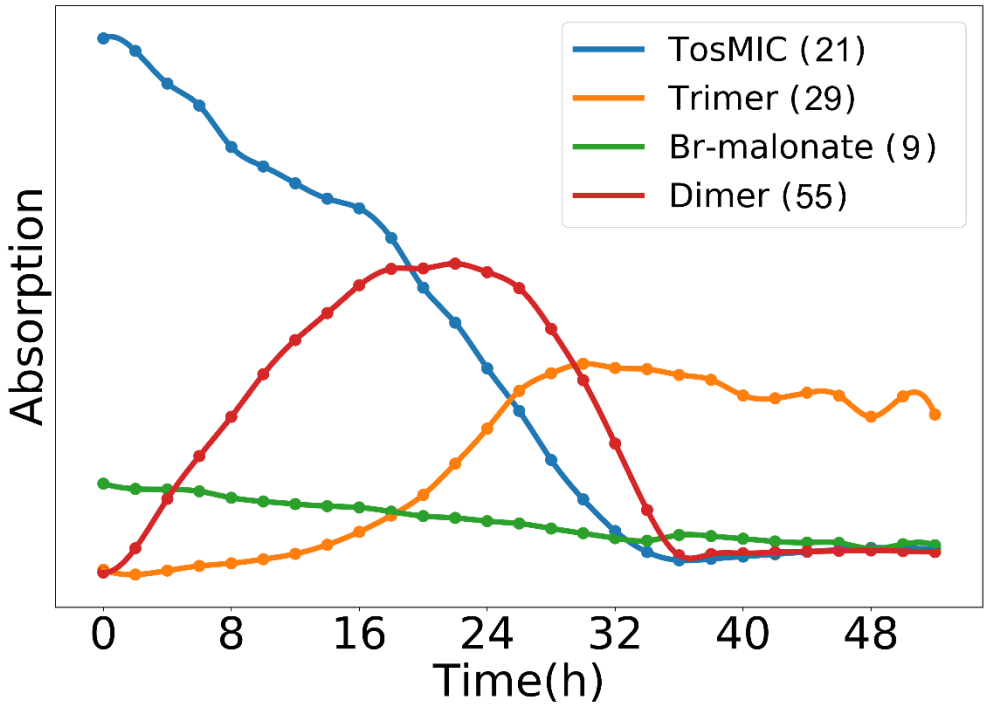


Figure S23: Visualisation of the time-resolved integrals in Figure S22 of the reagents **9** and **21**, the intermediary **55** and the product **29.**

5.3 Water influence

The calibration curve of the product **29** absorption is acquired with multiple injections of product **29** and an internal standard (hexafluorobenzene) at different ratios (**Figure S24**), integrals were integrated from the 210 nm wavelength data.

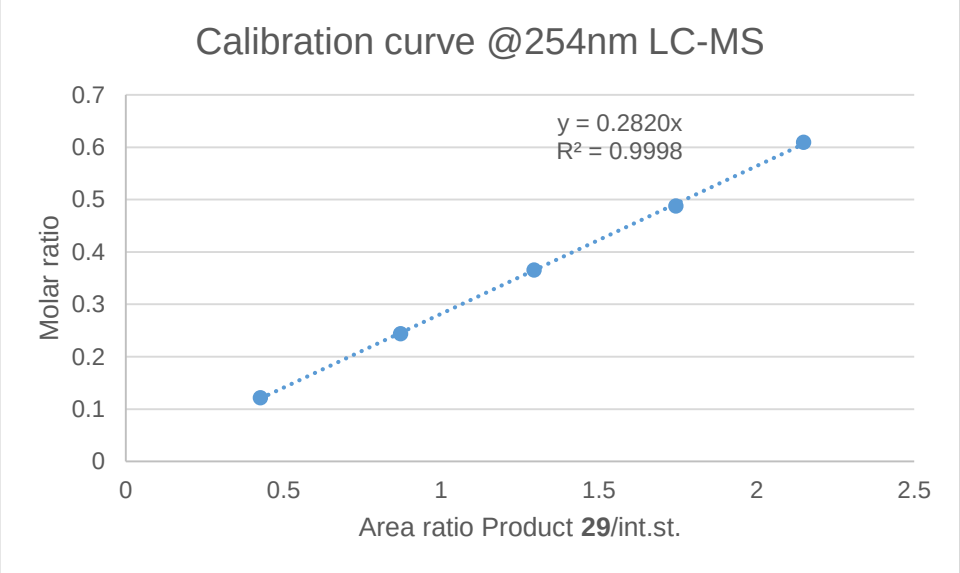


Figure S24: Calibration curve for the product **29.** Samples at different ratios with the internal standard were injected.

The slope of the calibration curve is used to calculate the yield of the product in the mixture in presence of different amount of water (see **Table S3**)

Table S3: The yield of the reaction performed in presence of different amounts of water was determined through HPLC analysis.

Water equivalents	Area product **29** 210nm	Area ISTD 254nm	Yield
0	1803	1865.826	2.2%
0.2	1946	1652.997	19.6%
0.4	4707	1603.213	48.9%
0.6	4189	1584.708	44.1%
0.8	2686	1672.036	26.8%
1	1461	1133.861	21.5%
2	1017	1754.559	9.7%
3	606	1708.202	5.9%

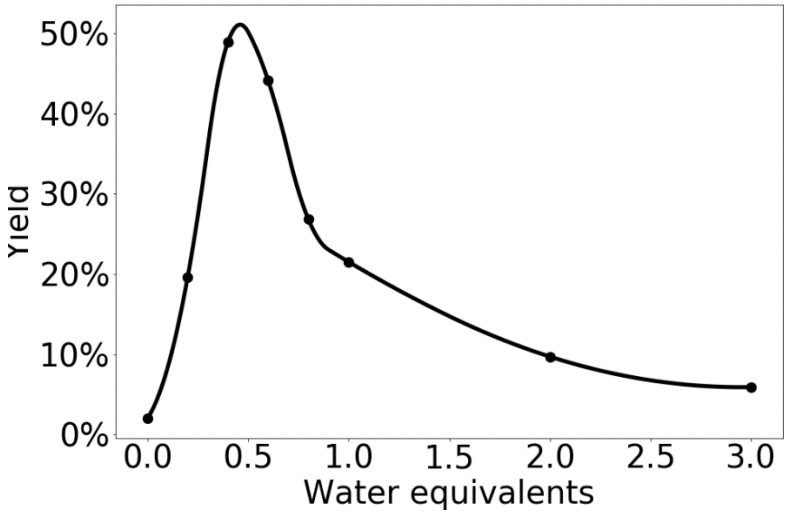


Figure S25: Visualized data points of **Table S2**.

5.4 Reaction performed in different solvents

The reaction was performed in acetonitrile and dimethylformamide. NMR analysis of the mixture showed no sign of reactivity (**Figure S26**). As further test it was also repeated in different ratio of DMSO/DMF showing the direct influence of DMSO in the yield (**Figure S27**).

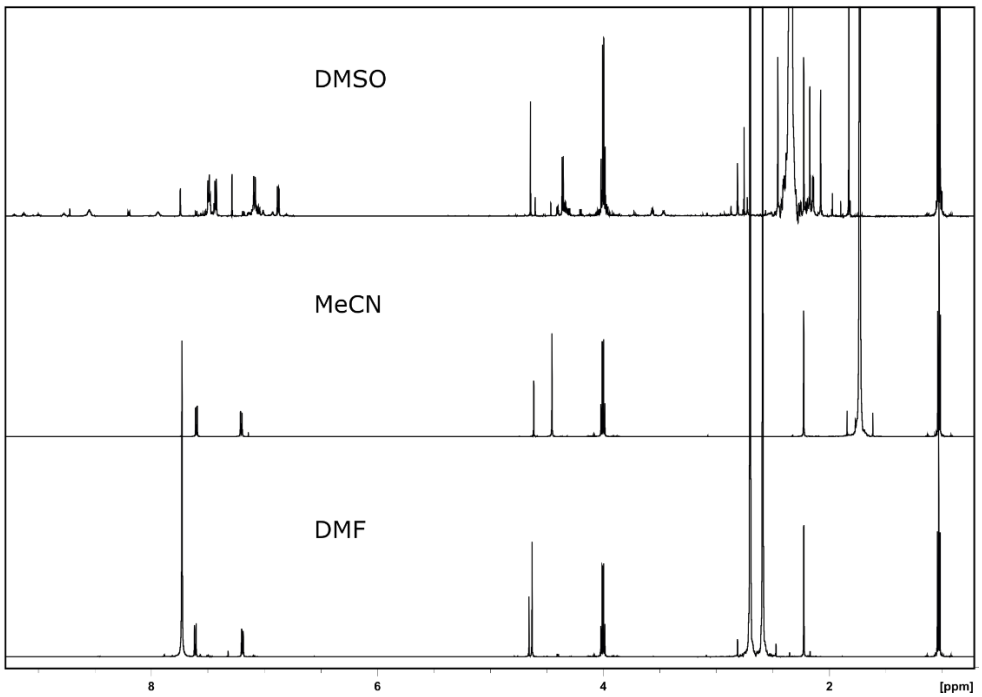


Figure S26: ^{1}H-NMR spectra of the reaction performed in different solvents. Acetonitrile and DMF show no reactivity.

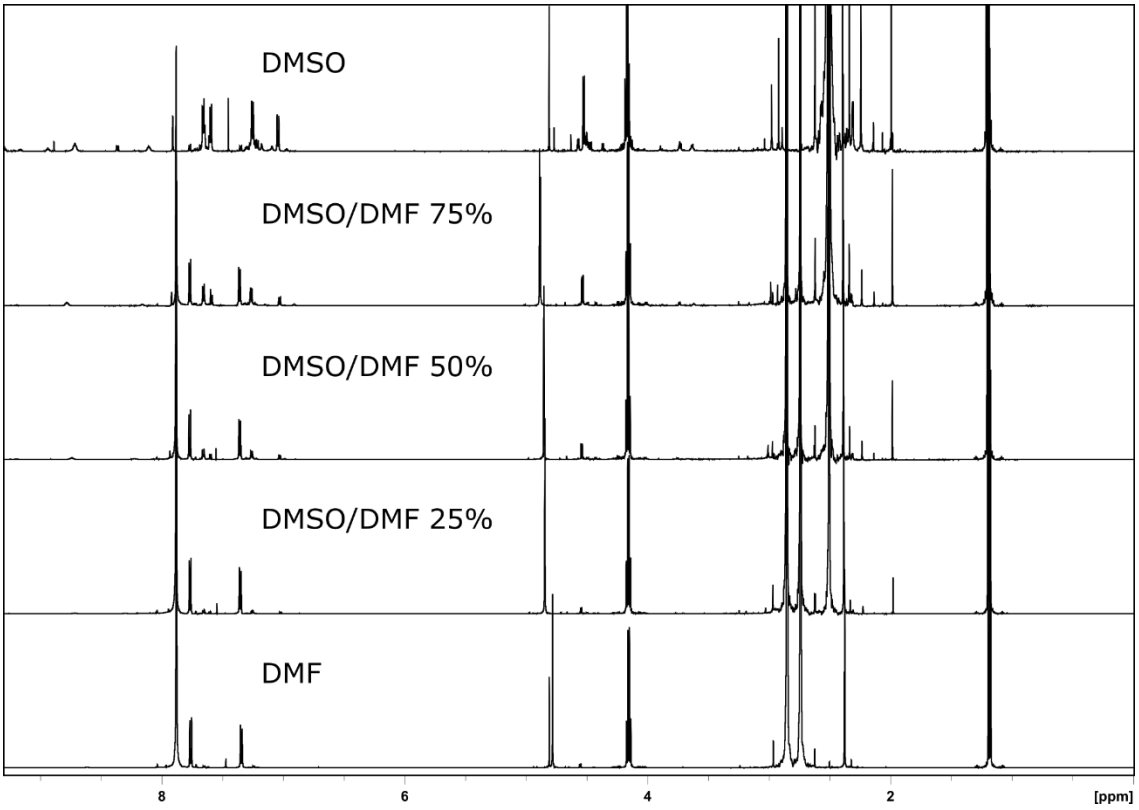


Figure S27: ^{1}H-NMR spectra of the reaction performed in a mixture of DMF and DMSO at different ratios.

5.5 Synthesis of isotopically substituted starting materials

5.5.1 Synthesis of (^{13}C)TosMIC

^{13}C-isotopically substituted TosMIC was prepared, based on literature procedures (5, 6), using either ^{13}C formamide (to substitute the isocyanide carbon) or a solution of ^{13}C-formaldehyde (to substitute the methylene carbon).

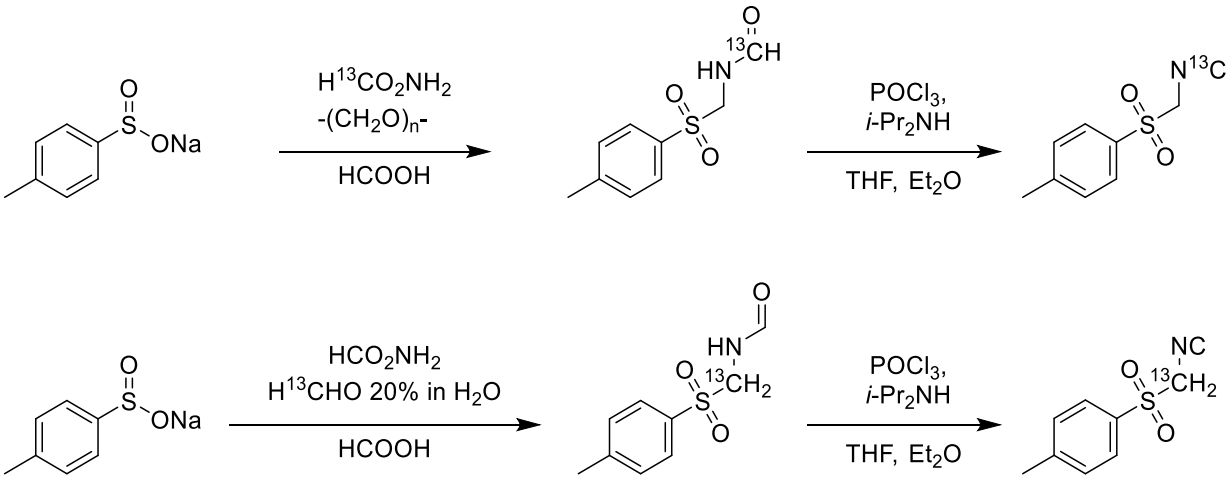


Figure S28: Synthetic approach to ^{13}C-isotopically substituted TosMIC.

5.5.2 Isocyanide Substitution

5.5.2.1 (1-13C)N-(tosylmethyl)-formamide

In a 25 ml round-bottom flask, equipped with a magnetic stir bar and a reflux condenser, were placed sodium p-toluenesulfinate (1.00 g, 5.6 mmol), paraformaldehyde (Sigma-Aldrich, 505 mg, 16.8 mmol, 3 eq.), formamide-^{13}C (Sigma-Aldrich, 99 atom% ^{13}C, 881 µl, 3.9 eq.), and formic acid (Fisher Chemicals, 1.06 ml, 5 eq.). The mixture was heated to 95 °C and stirred for 2 h at the same temperature. After the reaction was cooled to room temperature, it was diluted with cold water and transferred to a separation funnel. The mixture was extracted with ethyl acetate (3 x 40 ml) and the combined organic extracts were washed with brine, dried with $MgSO_4$, filtered and concentrated by rotary evaporation. Upon transfer to a smaller flask, the product spontaneously started to crystallize as small white crystals. The remaining solvent was then removed *in vacuo* yielding 937 mg of white crystals (85%). The pure N-(tosylmethyl)formamide was consistent with the spectral characteristic literature data for the non-isotopically substituted analogue.

^{1}H NMR (600 MHz, DMSO) δ 9.03 (q, *J* = 6.1 Hz, NH), 8.63 (dt, *J* = 11.4, 6.9 Hz, NH), 7.99 (dd, *J* = 197.9, 1.3 Hz, CHO), 7.80 (dd, *J* = 194.0, 11.1 Hz, CHO), 7.76 – 7.72 (m, 3H, Ar), 7.48 (d, *J* = 8.0 Hz, Ar), 7.44 (d, *J* = 8.0 Hz, Ar), 4.76 (t, *J* = 6.6 Hz, CH_2), 4.73 (dd, *J* = 6.8, 4.3 Hz, CH_2), 2.42 (s, CH_3), 2.40 (s, CH_3) (mixture of two rotamers).

^{13}C NMR (151 MHz, DMSO) δ 165.43, 161.00, 144.98, 144.71, 134.45, 133.65, 129.99, 129.81, 128.68, 128.51, 62.68, 58.55, 21.12, 21.11 (mixture of two rotamers).

LC-MS $[M + H]^+$ calculated for $C_8{}^{13}CH_{12}NO_3S$, 215.057; found, 215.0627.

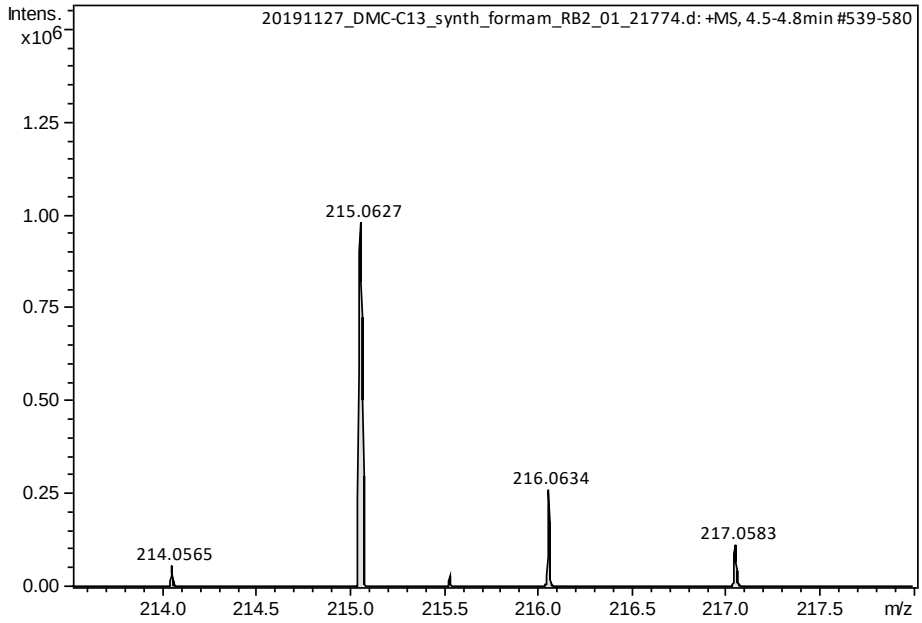


Figure S29: MS peak corresponding to (1-^{13}C)N-(tosylmethyl)-formamide.

5.5.2.2 $TosCH_2N^{13}C$

A 25 ml pear-shaped flask, equipped with magnetic stir bar was charged with ^{13}C-labelled N-(tosylmethyl)formamide (937 mg, 4.39 mmol), dry THF (3 ml), anhydrous diethyl ether (1 ml) and N,N-Diisopropylethylamine (Fluorochem, 2.29 ml, 13.2 mmol, 3 eq.). The stirred suspension was chilled to −5 °C with an ice-salt bath. Under a nitrogen atmosphere, 1.317 ml of a 4 M stock-solution of phosphorous oxychloride (1.2 eq.) in dry THF were slowly added via a syringe pump over the course of 1 hour. Towards the reaction completion the white suspension turned orange. After being stirred for another 30 min at 0 °C, the mixture was diluted with cold water (15 ml) and kept under vigorous agitation. The product then began to separate as a fine, brown crystalline solid that was collected by vacuum filtration, washed with cold water and dried in vacuum (493 mg, 57%). Spectral data in accordance with the literature for labelled-TosMIC (7, 8, 9).

^{1}H NMR (600 MHz, DMSO) δ 7.85 (d, *J* = 8.2 Hz, 2H), 7.56 (d, *J* = 8.2 Hz, 2H), 5.56 (d, *J* = 2.6 Hz, 2H), 2.45 (s, 3H).

^{13}C NMR (151 MHz, DMSO) δ 162.61, 146.14, 130.16, 128.89, 60.48, 21.18.

LC-MS $[M - H]^-$ calculated for $C_8{}^{13}CH_8NO_2S$ 195.031; found 194.9928.

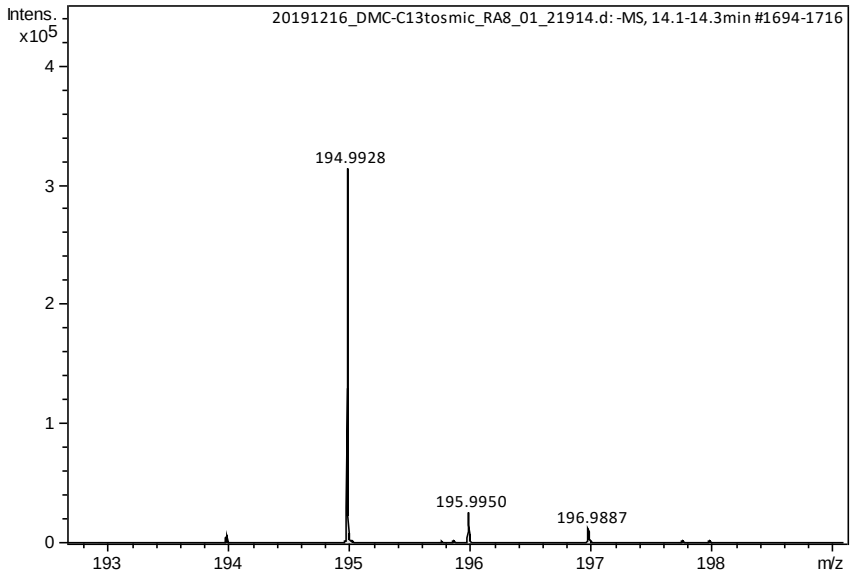


Figure S30: MS peak corresponding to $TosCH_2N^{13}C$

5.5.2.3 Product 25 - ^{13}C-1 variation

Synthesized according to standard procedure 4.1.

^{1}H NMR (600 MHz, DMSO-d_6) δ 8.96 (td, *J* = 6.7, 4.0 Hz, 3H), 7.68 (d, *J* = 8.2 Hz, 6H), 7.39 (d, *J* = 8.2 Hz, 6H), 4.66 (dd, *J* = 6.7, 3.9 Hz, 6H), 2.96 (d, *J* = 4.3 Hz, 6H), 2.37 (s, 9H).

^{13}C NMR (151 MHz, DMSO-d_6) δ 170.05 (^{13}C), 144.71, 134.39, 129.74, 128.49, 60.00, 56.47 (d, *J* = 52.6 Hz), 21.08.

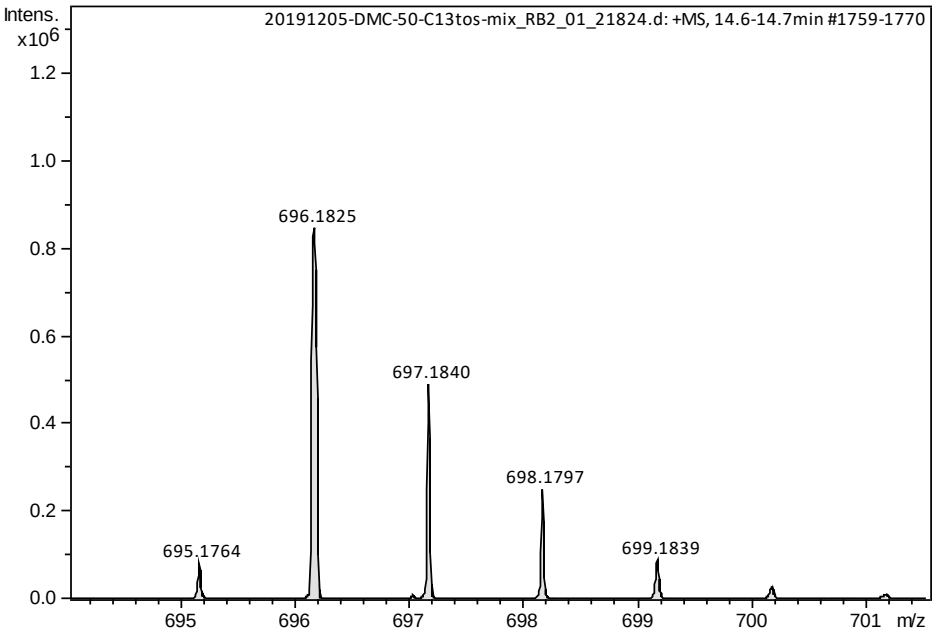


Figure S31: MS peak corresponding to Product 25 - ^{13}C-1 variation

5.5.3 Methylene Labelling

5.5.3.1 N-(tosyl(^{13}C)methyl)-formamide

Synthesized according to a modified literature procedure (5, 6). A three-necked, round-bottomed flask equipped with a magnetic stirrer is charged with 766 mg (4.3 mmol, 1 eq.) of sodium p-toluenesulfinate, 2 ml of water and 1 g (6.45 mmol, 1.5 eq.) of formaldehyde-^{13}C solution (Sigma 20 wt. % in H_2O, 99 atom % ^{13}C, 1.5eq.), 2.322 ml (2.6 g, 43 mmol, 10 eq.) of formamide and 567 ul (692 mg, 15 mmol, 3.5 eq.) of formic acid. The stirred reaction mixture is heated at 90 °C. The sodium p-toluenesulfinate dissolves during heating, and the solution is kept at 90–95 °C for 2 hours. After the reaction was cooled to room temperature, it was diluted with cold water and transferred to a separation funnel. The mixture was extracted with ethyl acetate (3 x 40 ml) and the combined organic extracts were washed with brine, dried with $MgSO_4$, filtered and concentrated by rotary evaporation, yielding the crude N-(tosylmethyl)formamide sufficiently pure for use in the next step.

^{1}H NMR (400 MHz, DMSO-d6) δ 9.10 – 8.99 (m, NH maj), 8.68 – 8.58 (m, NH min), 7.98 (dd, J = 5.8, 1.4 Hz, CHO maj), 7.78 (d, J = 11.1 Hz, CHO min), 7.73 (d, J = 8.2 Hz, Ar, maj + min), 7.49 (d, J = 8.2 Hz, Ar, min), 7.45 (d, J = 8.2 Hz, Ar, maj), 4.76 (dd, J = 151.1, 6.9 Hz, CH_2, min), 4.71 (dd, J = 150.8, 6.8 Hz, CH_2, maj), 2.43 (s, CH_3, min), 2.41 (s, CH_3, maj) (at room temperature, mixture of two rotamers ratio 80:20).

^{13}C NMR (101 MHz, DMSO-d6) δ 160.94, 144.67, 134.47, 129.96, 129.78, 128.64, 128.47, 62.64, 58.52, 21.08 (at room temperature, mixture of two rotamers. Some peaks of the minor component were too weak to be detected).

LC-MS $[M + H]^+$ calculated for $C_8{}^{13}CH_{12}NO_3S$, 215.057; found, 215.0576.

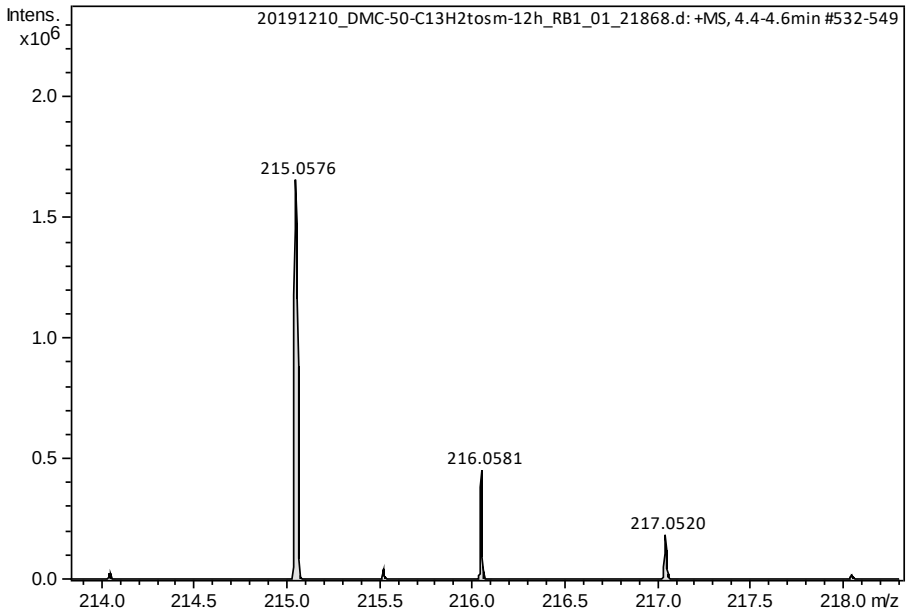


Figure S32: MS peak corresponding to N-(tosyl(^{13}C)methyl)-formamide

5.5.3.2 Tos$^{13}CH_2NC$

Same procedure as 5.5.2.2.

^{1}H NMR (600 MHz, DMSO) δ 7.85 (d, *J* = 8.1 Hz, 1H), 7.56 (d, *J* = 8.0 Hz, 1H), 5.56 (d, *J* = 157.9 Hz, 1H), 2.45 (s, 2H).

^{13}C NMR (151 MHz, DMSO-d_6) δ 162.64, 146.15, 130.16, 128.89, 60.49, 21.18.

LC-MS $[M - H]^-$ calculated for $C_8{}^{13}CH_8NO_2S$ 195.031; found 194.9984.

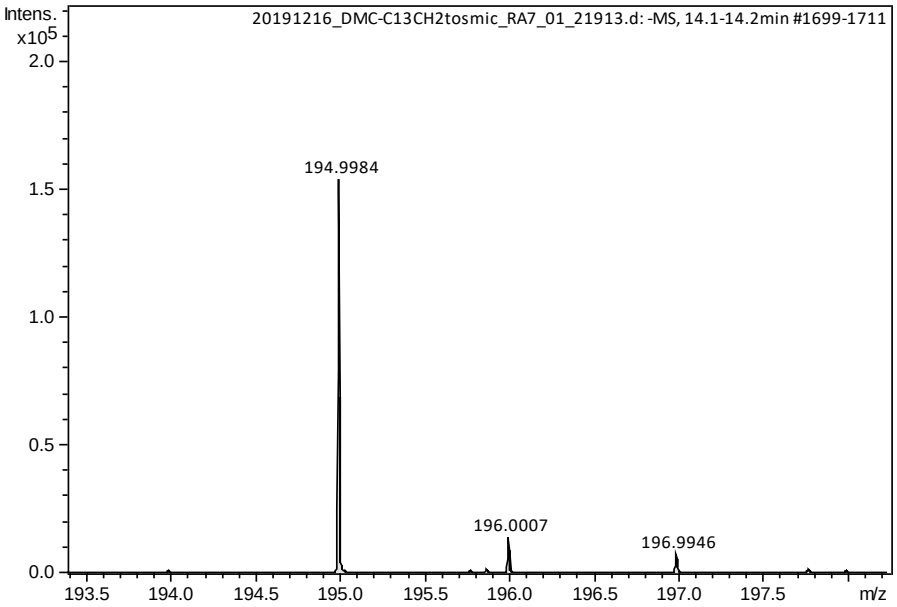


Figure S33: MS peak corresponding to Tos$^{13}CH_2NC$

5.5.3.3 Product 25 - ^{13}C-2 variation

Synthesized according to standard procedure 4.1.

^{1}H NMR (600 MHz, DMSO-d_6) δ 8.95 (t, *J* = 6.1 Hz, 3H), 7.68 (d, *J* = 7.9 Hz, 6H), 7.39 (d, *J* = 7.9 Hz, 6H), 4.65 (dd, *J* = 150.9, 6.7 Hz, 6H), 2.96 (dt, *J* = 136.3, 4.0 Hz, 6H), 2.37 (s, 9H).

^{13}C NMR (151 MHz, DMSO-d_6) δ 170.06 (d, *J* = 52.6 Hz), 144.71, 134.36, 129.74, 128.49, 60.00, 56.48, 21.08.

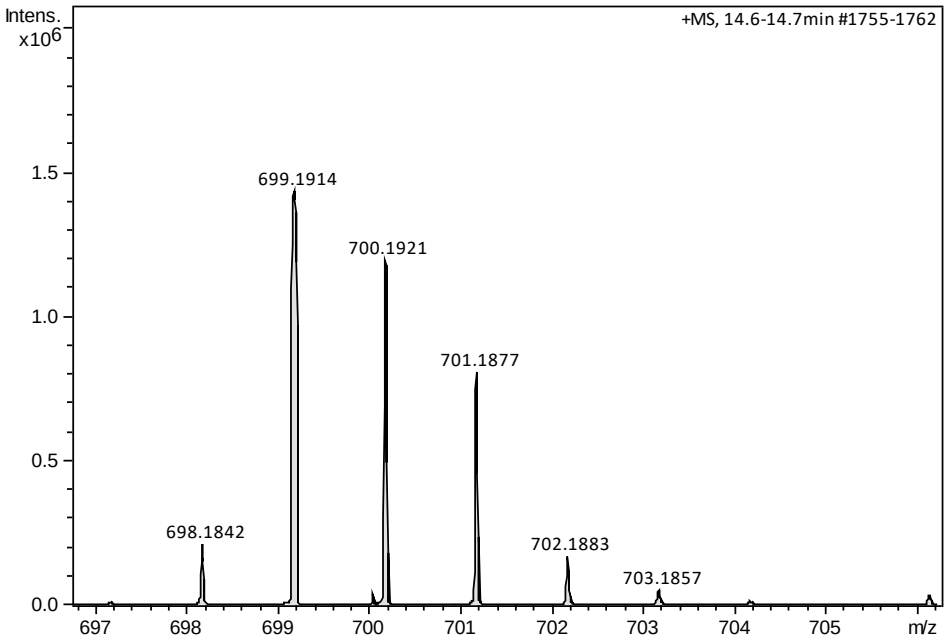


Figure S34: MS peak corresponding to Product 25 – ^{13}C-2 variation

5.5.4 DMSO Labelling

5.5.4.1 Synthesis of ^{18}O-DMSO

Method for the preparation of ^{18}O isotopically substituted DMSO (9): solid dimethylsulfur dibromide (20 g, 90 mmol, TCI Chemicals) was added portion wise over 30 min to a vigorously stirred solution of triethylamine (25.2 ml, 180mmol) and ^{18}O-labeled water (97 atom %, Sigma-Aldrich) (1.0 ml, 50 mmol) in 60 ml of THF (fresh from solvent purification system). An ice bath was used to occasionally cool the reaction. During the addition, the orange reactant slowly dissolved giving a white precipitate of triethylamine hydrobromide. The precipitate was removed by centrifugation (5 min 4000 rpm) and washed twice with ether (same time and speed as before). The combined pale-yellow supernatant and washings were dried by rotary evaporation up to 8 mbar giving 4.6 g of orange oil, composed mostly of ^{18}O-DMSO, pure enough for use as solvent in the subsequent reaction.

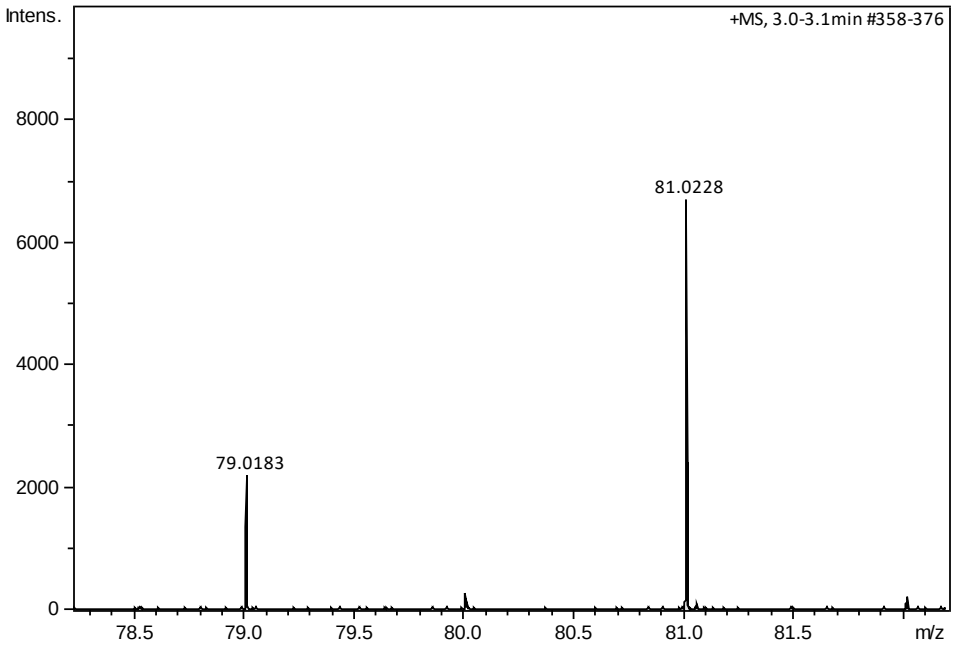


Figure S35: MS spectrum of the isotopically labelled DMSO, from the peaks intensity is it visible a 3:1 ratio between ^{18}O-DMSO and ^{16}O-DMSO.

5.5.4.2 Product 25 – 18O-DMSO Labelled

The reaction procedure followed the standard conditions presented in 4.1. The reaction mixture was analysed with HPLC-MS. The incorporation of oxygen from ^{18}O-DMSO was observed in the reaction, giving a ratio of 0.25:0.89:1:0.65 for non:mono:di:tri-labelled compounds. This is in accordance with the 3:1 ratio of the DMSO labelling and suggests that all the oxygens of the product come from DMSO.

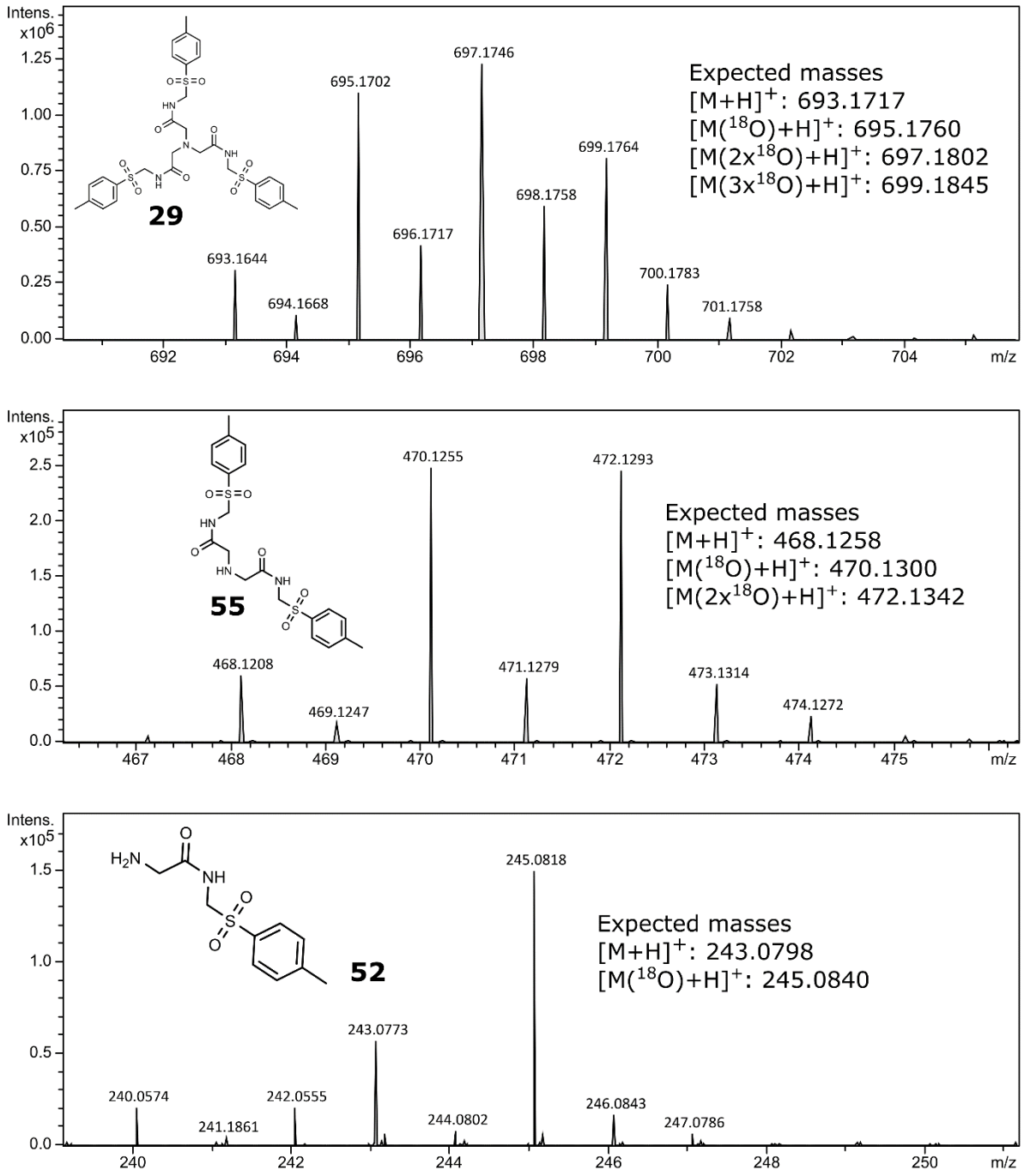


Figure S36: MS spectrum of products **29, 55** and **52** prepared in ^{18}O-DMSO. The isotopic pattern of all molecules shows incorporation in accordance with the 3:1 labelling of the solvent, suggesting that all the oxygens in the product come from DMSO.

5.5.5 Reaction in $^{18}OH_2$

The synthesis of the trimer (**25**) was undertaken according to the procedure reported in SI-4.1, using anhydrous DMSO (dried over activated molecular sieves, 4Å), under an inert atmosphere and with addition of 3 equivalents of $H_2{}^{18}O$. The mixture was analysed by HPLC-MS after 5 h observing the formation of the product **29** and the two intermediates **52** and **55** (**Figure S37**). The most intense peaks

of all compounds correspond to the non-labelled versions, indicating that the oxygens do not come from the water. However, **29** and **55** show an increase in the intensity of the $[M+H+2]^+$ isotope, suggesting the partial incorporation of a water molecule can occur. This can be explained with the presence of amines in solution reversibly forming imine groups with the carbonyls**,** essentially exchanging the oxygens with water.

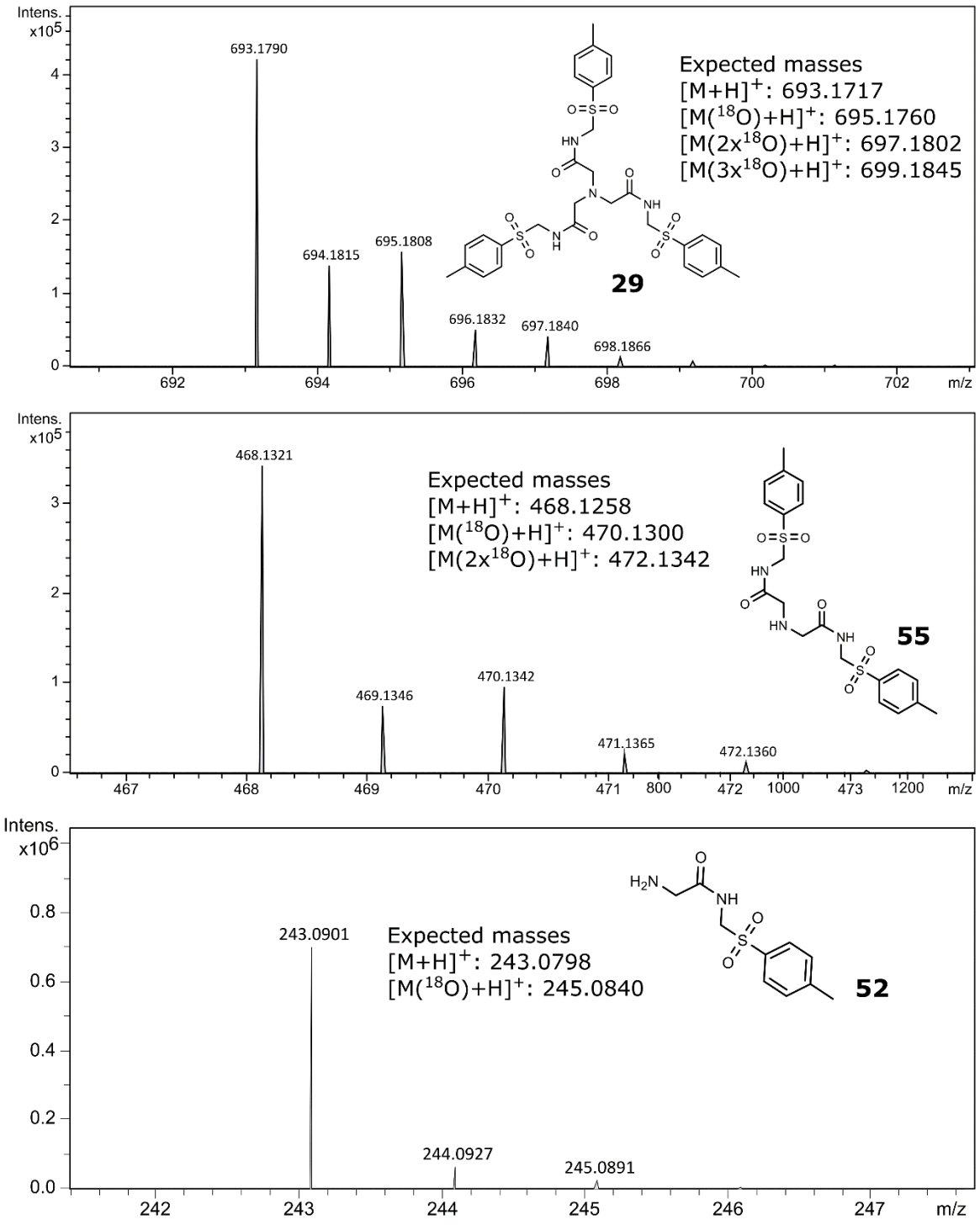


Figure S37: In order: mass isotope patterns of the product **29** and the intermediates **55** and **52**. The reaction was performed in presence of labelled water and shows no incorporation. The small increase in the +2 isotopes of **29** and **55** is explained with the imine equilibrium of the carbonyls.

5.6 IR reaction monitoring.

The reaction prepared according to the standard recipe 4.1 was monitored for 18 hours with an IR instrument (Thermo Scientific™ Nicolet™ iS™5 FTIR) equipped with a flow cell. The reaction was continuously pumped into the flow cell and back into the reactor vessel using a peristaltic pump. The instrument was remotely controlled with python and acquired a new spectrum of the mixture every 5 minutes. Results are showed in pictures **Figure S38** and **Figure S39**. The signal of the CN bond from TosMIC at 2150 cm^{-1} disappears after 10 h while the formation of isocyanate (11) (2257 cm^{-1}) and CO_2 (2349 cm^{-1}) are observed.

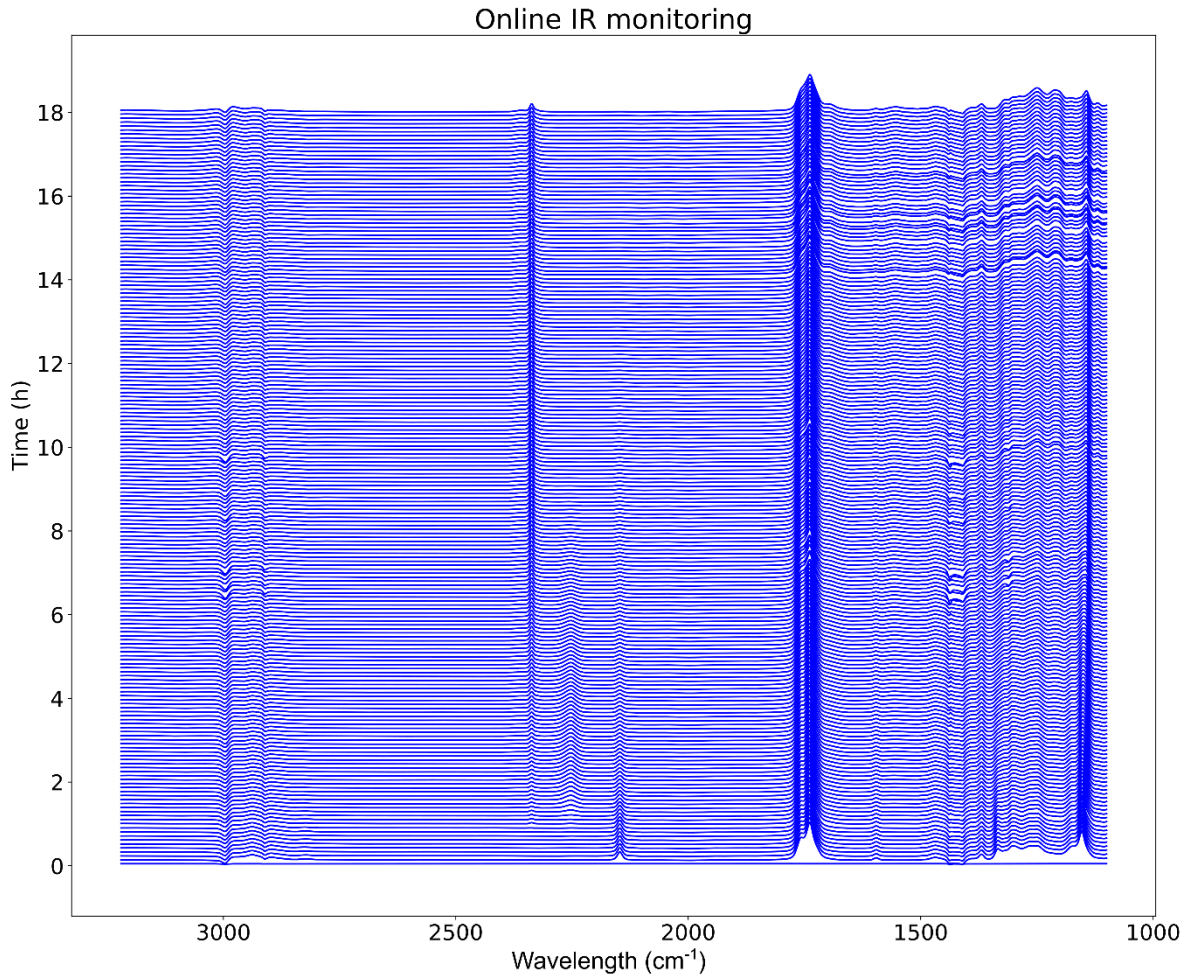


Figure S38: Online FT-IR analysis of the reaction mixture. A new spectrum is acquired every 5 minutes. For 18 hours, the wavelength are plotted vs the time of acquisition. The desappearing of TosMIC and appearing of an isocyanate bond and CO_2 are visible around 2250 cm^{-1}. Espansion of this area in **Figure S39**.

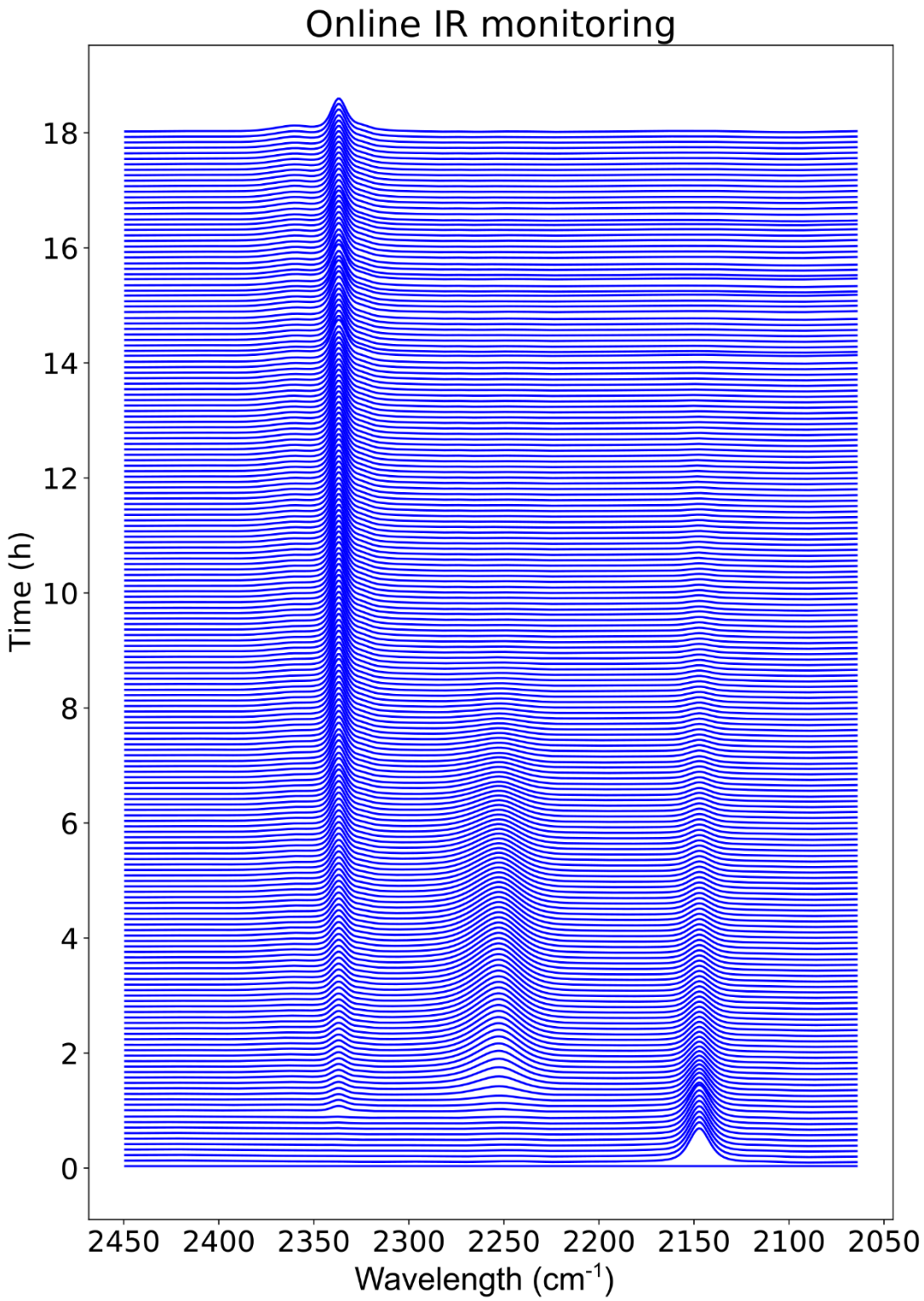


Figure S39: Detail of the online IR monitoring (**Figure S38**) showing the disappearing of the CN bond from TosMIC at 2150 cm^{-1} and the formation of isocyanate (2257 cm^{-1}) and CO_2 (2349 cm^{-1}).

5.7 Online NMR monitoring

Diethyl 2-bromomalonate **9** (0.5 mmol, 0.1 ml) and p-toluenesulfonylmethyl isocyanide **21** (2 mmol, 0.975 g) were mixed in 1 ml of deuterated DMSO inside an NMR tube. The reaction was left inside the NMR spectrometer for 24 h at 298K and the instrument was programmed to acquire a proton and carbon spectra every two hours. Analysis of this data showed the formation of ammonium as three peaks with 1:1:1 intensity ratio between 7.0 and 7.3 ppm (12) (**Figure S40**). The formation of dimethyl sulphide is also observed as a quintuplet at 2 ppm (13) (**Figure S41**). Data was processed using MestReNova software suite.

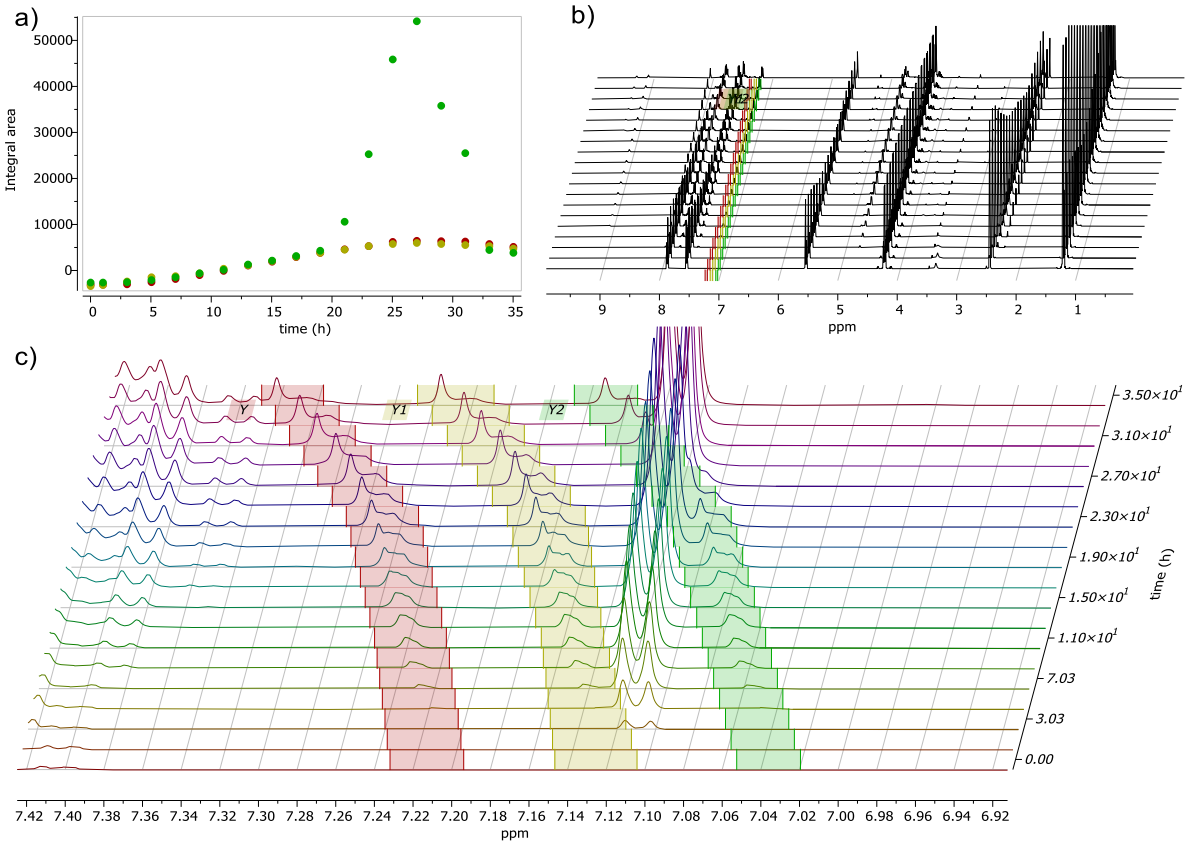


Figure S40: Online NMR monitoring of the reaction. The formation of the ammonium peaks is observed as three peaks (red, yellow and green) with 1:1:1 ratio around 7 ppm. a) Area of the three peaks over time. During the shifting peak Y2 partially overlaps with another product. b) full spectra. c) zoomed spectra showing the presence of ammonium peaks.

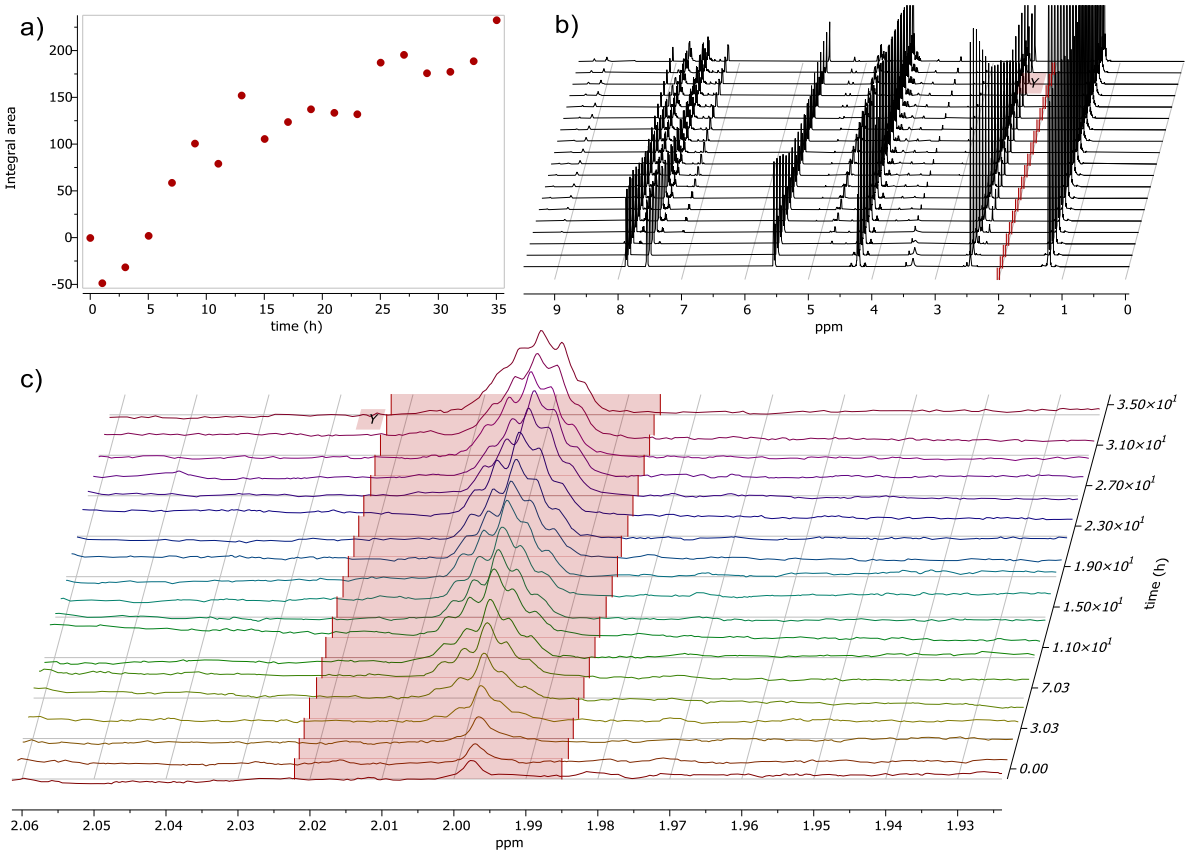


Figure S41: Online NMR monitoring of the reaction. The formation of dimethyl sulfide is observed at 2 ppm. a) Area of the peak over time. b) full spectra. c) zoomed spectra showing the presence of dimethyl sulfide peaks.

6 Cheminformatics simulation

6.1 Reaction network

A list of 21 reaction templates was encoded as SMARTS (**Figure S42**). The simulation started with a pool of three starting materials (TosMIC, water and DMSO) encoded as SMILES. The software applied the reaction templates to the starting materials and gathered the products. These were added to the pool and the process was repeated for the next step. Since the number of molecules progressively increases as result of the combinatorial explosion it was possible to calculate as further as 8 reaction steps. In order to connect the reactions graph to the product the simulation was repeated reversing the reaction SMARTS and starting from product **29**. The two resulting networks were then merged by keeping only the molecules simultaneously connected to TosMIC and **29**. The script used to perform these calculations was written in Python using the RDKit library. An estimate of the number of molecules in the reaction steps between 8 and 12 was made by fitting the natural logarithm of the number of the molecules with a second degree polynomial curve.

```
Reaction_SMARTS = {
  "imine formation 1": [C,#1:4][C:1](=O)[C,#1:5].[CX4:3][Nh2+0:2]>>[*:4][C:1](=[*+:2][*:3])[*:5].O",
  “imine formation 2":
"[C,#1:5][C:1](=O)[C,#1:6].[CX4:4][Nh+0:2][CX4:3]>>[*:5][C:1](=[*+:2]([*:4])[*:3])[*:6].O",
  "primary amine oxidation": "[Nh2+0:1][Ch:2]>>[N:1]=[*:2]",
  "secondary amine oxidation": "[Ch:2][Nh+0:1][C^3:3]>>[*:2]=[N:1][*:3]",
  “primary imine-amine attack": "[N:1]=[C^2:2].[Nh2X3:3][Ch:4]>>[Nh:1][C:2][*:3][*:4]",
  "secondary imine-amine attack":
"[N:1]=[C^2:2].[NhX3:3]([Ch:4])[Ch:5]>>[Nh:1][C:2][*:3]([*:4])[*:5]",
  "amine elimination 1": "[Nh2:1][C:2][ND1X3+0,ND2X3+0:3]>>[C:2]=[*:3].[*:1]",
  "amine elimination 2": "[Nh2:1][C:2][ND3X3+0:3]>>[C:2]=[N+:3].[*:1]",
  "imine reduction": "[N,N+1:1]=[C^2:2]>>[N+0:1][C:2]",
  "imine hydrolysis": "[C^2:1]=[NX2+0:2].[Oh2:3]>>[C:1]=[*:3].[N:2]",
  "isocyanate to urea": "O=[C:1]=[N:2].[Nh2:3]>>O=[C:1]([NH:2])[Nh:3]",
  "isocyanide alpha attack": "[C:1]=[N:2].[Ch:3][N+:4]#[C-:5]>>[C:1]([N:2])[*:3][N+:4]#[C-:5]",
  "carbanion attack": "[C:1]=[N:2].[C-:3]>>[C+0:3][C:1][N:2]",
  "isocyanide oxidation": "[N+:1]#[C-:2].[CH3]S([CH3])=[O:3]>>[N+0:1]=[C+0:2]=[O:3]",
  "isocyanate hydrolysis": "O=C=[N:1].[OH2]>>[N:1].O=C=O",
  "nitrilium hydrolysis": "[C+0:1]#[N+:2].[Oh2]>>[C:1](=O)O.[N+0:2]",
  "Ugi reaction": "[Nh:1][CX4:2][C:3]#[N+:4].O=[C:5][Oh:6]>>O=[C:5][*:1][*:2][C:3](=[*:6])[N+0:4]",
  "imine tautomerism": "[Ch^3:1]-[NX2+0:2]=[C^2:3]>>[*:1]=[*:2]-[*:3]",
  "decarboxylation": "[C:1](C(=O)[Oh])[C:2](=[O:3])>>[C:1][C:2](=[O:3]).O=C=O",
  "Ts elimination": "O=[S:3](=O)[CX4:1][ChX4:2]>>[C:1]=[C:2]",
  "Ts
reduction":“O=[S:3](=O)[CX4:1]([C^2,C^1,N^1:2])[C^2,C^1,N^1:4]>>[C:1]([C^2,C^1,N^1:2])[C^2,C^
1,N^1:4]"
}
```

Figure S42: List of reaction SMARTS used to create the chemical space network.

6.2 Similarity index with known reactions

1652 reactions involving TosMIC as reagent were gathered from the Reaxys database. The reaction fingerprints were calculated by subtracting reagent(s) fingerprints of from product fingerprint(s). Since the fingerprint values were generated from the molecular structure the difference for a similar reaction will be close to a vector of zeroes, while two distinct reactions will produce a vector with high values, positive and negative. In order to quantify and compare the reaction fingerprints we calculated the ℓ-2-norm, the square root of the sum of the squares. Values closed to zero indicate a high structural similarity between reagent(s) and product(s). The examples of the two reactions of TosMIC reactions showing a higher ℓ-2-norm than the discovered one are reported in **Figure S43**. They are variations of polymerization reactions and Ugi four-component reactions.

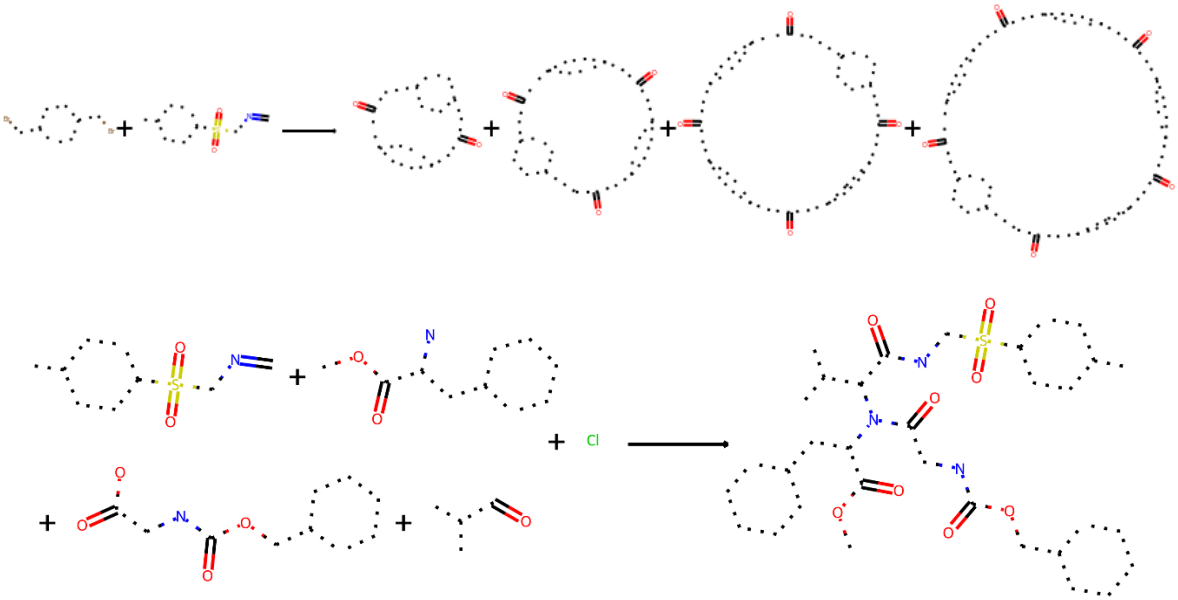


Figure S43: Templates of the two main types of reactions presenting a higher difference generation.

Next, the dataset has also been used to compare the discovered reaction with known TosMIC reactions using the Tanimoto similarity index (14). A matrix of cross correlation between the reaction fingerprints is showed in **Figure S44**. The Tanimoto index between reaction leading to product **29** and the other literature reactions is in the last column/row and expanded in the bottom part of the figure for visualization. The average similarity was 0.091. A histogram comparing this value with the similarity averages of the other reactions is showed in **Figure S45**. As reference the reactions with highest similarity to the discovered reaction are showed in **Figure S46**. The script used to perform these calculations was written in Python using the RDKit library.

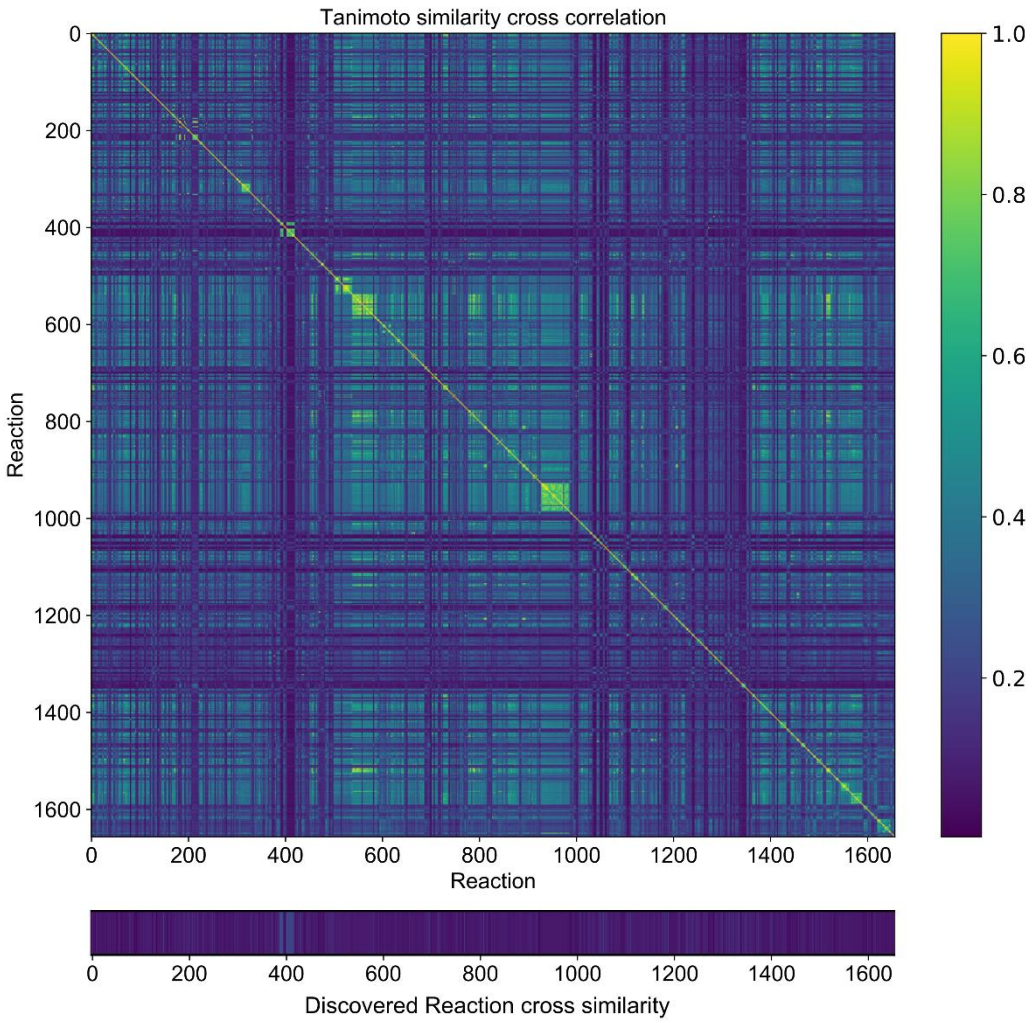


Figure S44: Tanimoto similarity index between reaction fingerprints of the literature reactions of TosMIC. In the bottom the similarity index between literature reactions and discovered reaction giving product **29.**

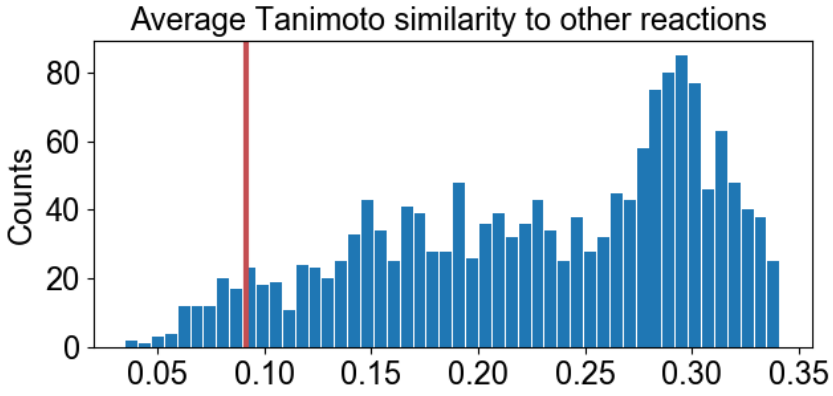


Figure S45: Histogram comparing the average similarity for each reaction to other reactions in **Figure S44**.

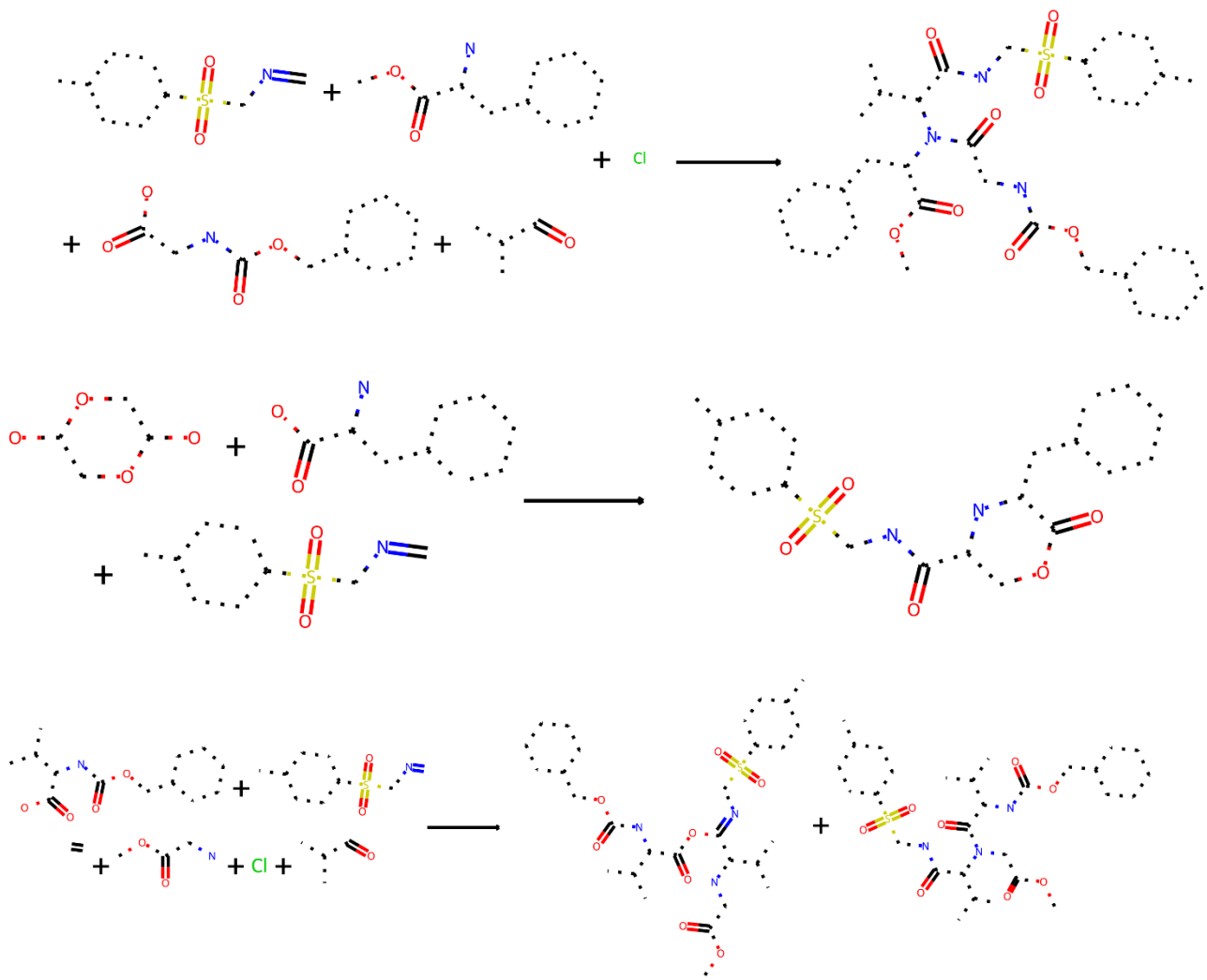


Figure S46: Reactions having the most similar reaction fingerprints to discovered reaction**,** according to Tanimoto similarity index.

The similarity study has also been repeated with known reactions involving the isocyanate **47** (derivate from TosMIC oxidation) since in the mechanism proposed it is the effective active specie. Unfortunately, the Reaxys database contains only 40 reactions involving this molecule (compared to 1652 reactions with TosMIC) therefore there is not much data for comparison. We run the algorithms anyway and reported here the results.

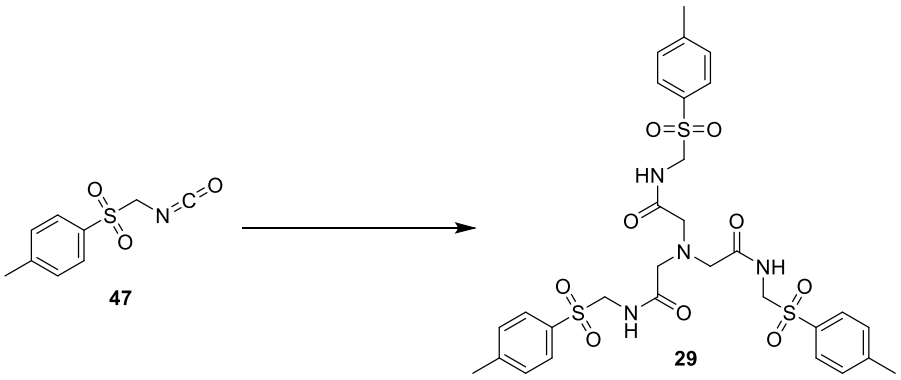


Figure S47: Since our in mechanism the active species is the isocyanate **47** we run the similarity algorithm also with this transformation.

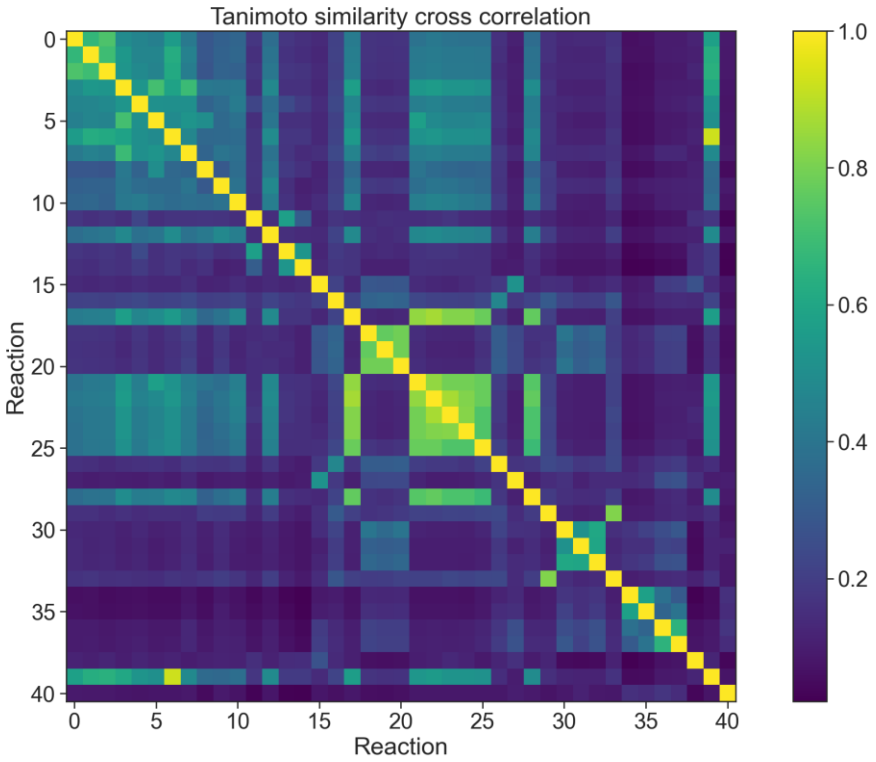


Figure S48: Tanimoto similarity index between reaction fingerprints of the literature reactions of the isocyanate **47**. In the bottom the similarity index between literature reactions and discovered reaction giving product 29 (starting from the isocyanate).

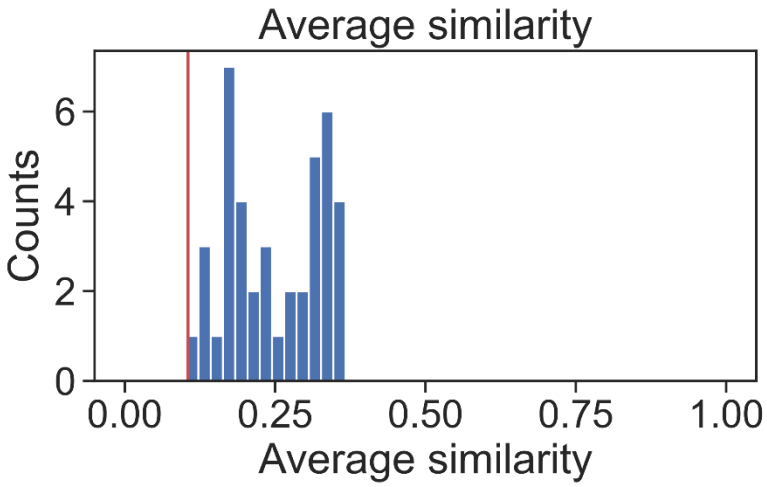


Figure S49: Histogram comparing the average similarity for each reaction to other reactions in Figure S44**,** this is the analogue of Figure S45**.**

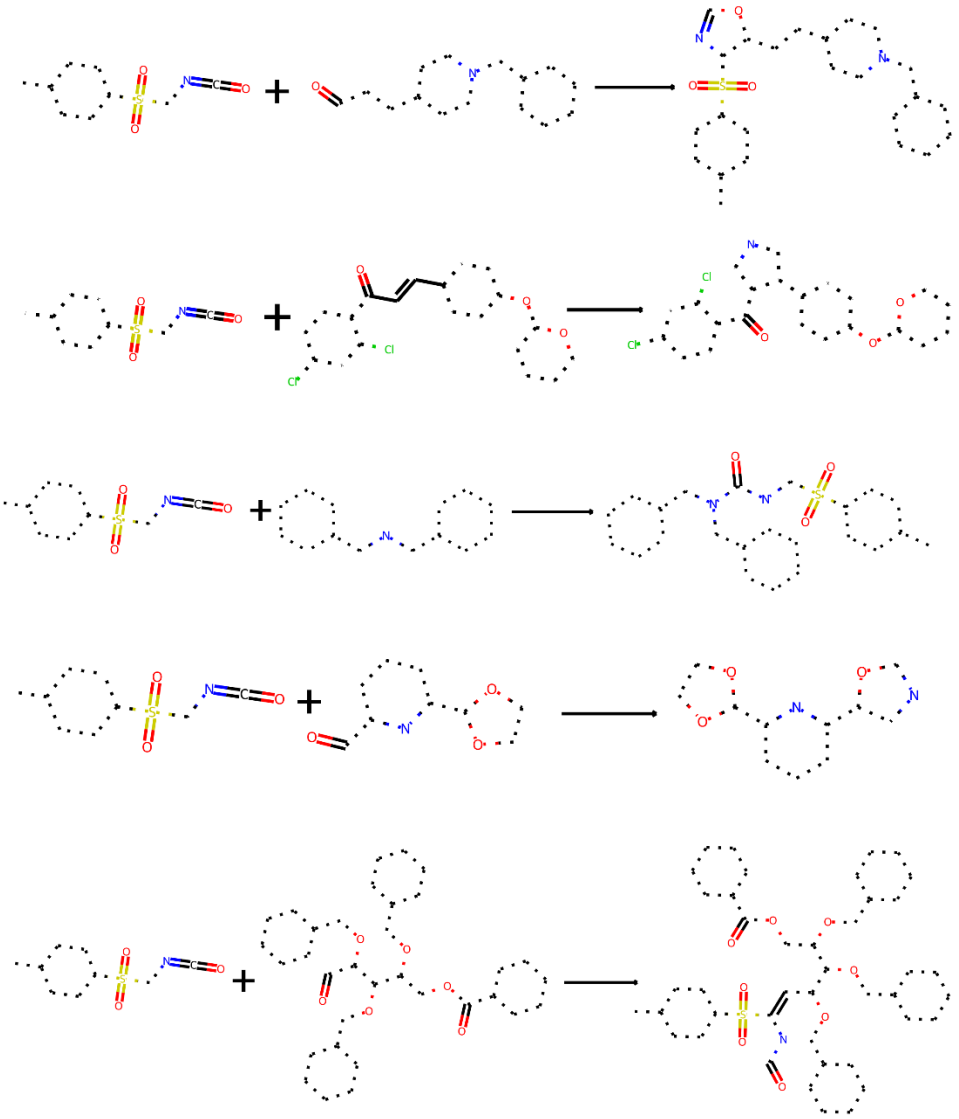


Figure S50: Reactions of isocyanate **47** with the highest similarities to the discovered reaction.

7 Other reactions discovered and re-discovered

A total of 7 reactions have been analysed. One of the yielded product **25** and was further investigated. From two of them despite the evidence of reactivity in NMR analysis and TLC it was not possible to isolate any products in purity high enough for characterisation. The first one was the reaction of 1,3-Diethyl-2-thiobarbituric acid and TosMIC while the second was the reaction of 1,3-Diethyl-2-thiobarbituric acid and phenylhydrazine. Both of them involved the presence of tris(2,2'-bipyridyl)dichlororuthenium(II) hexahydrate and 450 nm irradiation. The remaining four reactions are reported below.

7.1 Phenylhydrazine and bromoacetonitrile under 450 nm irradiation

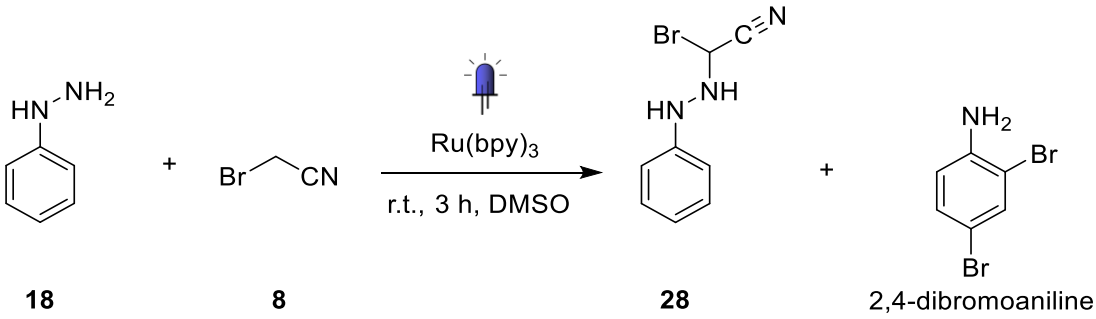


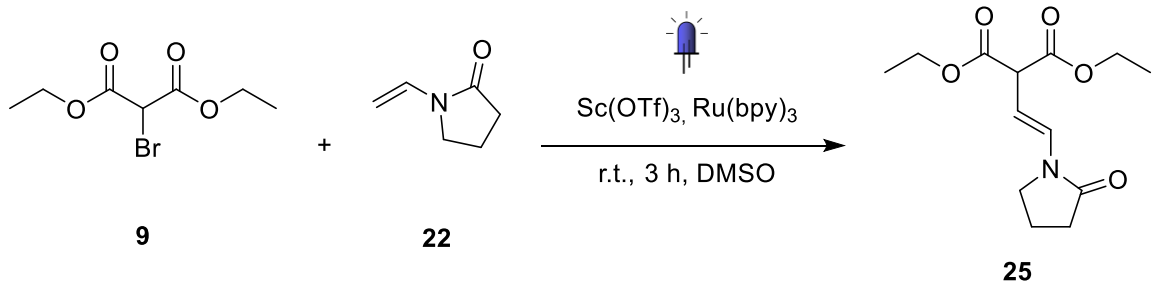
Figure S51: Scheme of the reaction between phenylhydrazine and bromoacetonitrile, under 450 nm irradiation and the $Ru(bpy)_3$ photocatalyst. The first product is unreported in literature.

Phenylhydrazine (2 mmol, 0.23 ml), Bromoacetonitrile (2 mmol, 0.41 ml) and Tris(2,2'-bipyridyl) tris(2,2'-bipyridyl)dichlororuthenium(II) hexahydrate (2.5% mol, 32 mg) are mixed in 4.5 ml of DMSO. The reaction is stirred at room temperature and irradiated with 450 nm LED for 3 hours. The reaction mixture is diluted with water and extracted with ethyl acetate. The organic phase is separated and washed three times with brine. Mg_2SO_4 is then added to the reaction mixture and after filtration the solvent is removed under reduced pressure. The crude is purified with a chromatographic column (silica gel, Hexane/ EtOAc 80:1), isolating both Compound **28** and 2,4-dibromoaniline. **28** is obtained with a photocatalytic addition where a new C-N bond is formed but the bromide, usually a leaving group, is kept in its place. This reaction is unreported in literature. As further tests the reaction was repeated in presence of light but without photocatalysts and in presence of 2,4,6-triphenylpyrylium tetrafluoroborate. Both tests resulted negative as no product formation was detected by TLC.

Compound **28**
Yield: 7% (31 mg)
IUPAC name: Bromo-(N'-phenyl-hydrazino)-acetonitrile
^{1}H NMR (600 MHz, $CDCl_3$) δ 8.90 (s, 1H), 7.37 (d, *J* = 16.0 Hz, 2H), 7.18 (d, *J* = 7.7 Hz, 2H), 7.09 (t, *J* = 7.4 Hz, 1H), 6.24 (s, 1H)
^{13}C NMR (151 MHz, $CDCl_3$) δ 141.22, 129.14, 123.22, 113.77, 110.68, 99.62.
Crystallographic data: See section 8

2,4-dibromoaniline
Yield: 11% (55 mg)
^{1}H NMR (600 MHz, $CDCl_3$) δ 7.56 (d, *J* = 2.2 Hz, 1H), 7.22 (dd, *J* = 2.2, 8.5 Hz, 1H), 6.66 (d, *J* = 8.6 Hz, 1H), 4.11 (s, 1H), validated with commercial material.

7.2 Diethyl 2-bromomalonate and 1-vinyl-2-pyrrolidinone under 450 nm irradiation



9 **22** **25**

Figure S52: Scheme of the reaction between diethyl 2-bromomalonate and 1-vinyl-2-pyrrolidinone under 450 nm irradiation and the $Ru(bpy)_3$ photocatalyst.

Diethyl 2-bromomalonate (2 mmol, 0.34 ml), 1-Vinyl-2-pyrrolidinone (2 mmol, 0.21 ml) and tris(2,2'-bipyridyl)dichlororuthenium(II) hexahydrate (2.5% mol, 32 mg), scandium triflate (2.5% mol, 25 mg) are mixed in 4.5 ml of DMSO. The reaction is stirred at room temperature and irradiated with 450 nm LED for 3 hours. The reaction mixture is diluted with water and extracted with ethyl acetate. The organic phase is separated and washed three times with brine. Mg_2SO_4 is then added to the reaction mixture and after filtration the solvent is removed under vacuum. The crude mixture is purified with chromatographic column (silica gel, hexane/ EtOAc 2:1). The reaction is a C-H functionalization made through photoredox catalysis. A similar reaction has been already described in 2012 (15) where the authors used the same reagents, $[Ir(ppy)_2(dtbbpy)]PF_6$ as photocatalyst, 2 equivalents of Na_2HPO_4 and acetonitrile as solvent.

Yield: 53% (285 mg)

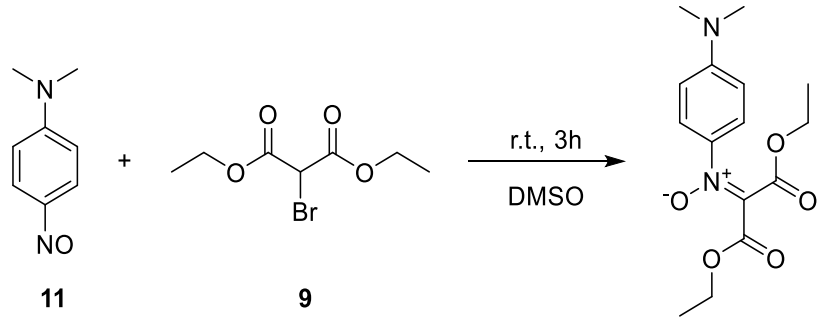
IUPAC name: (E)-Diethyl 2-[2-(2-oxopyrrolidin-1-yl)vinyl]malonate

^{1}H NMR (600 MHz, $CDCl_3$) δ 7.00 (d, J = 14.4 Hz, 1H), 5.09 (dd, J = 14.5, 9.5 Hz, 1H), 4.14 (qd, J = 7.1, 3.6 Hz, 4H), 3.96 (d, J = 9.5 Hz, 1H), 3.51 (t, J = 7.2 Hz, 2H), 2.42 (t, J = 8.1 Hz, 2H), 2.06 (p, J = 7.7 Hz, 2H), 1.21 (t, J = 7.1 Hz, 6H).

13C NMR (151 MHz, CDCl3) δ 173.17, 168.20, 127.74, 102.82, 61.55, 53.24, 44.82, 30.83, 17.23, 13.80.

ESI-HR-MS: $[C_{13}H_{19}NNaO_5]^+$ Calculated 292.1155 *m/z*, measured 292.1147 *m/z*

7.3 *N,N*-dimethyl-4-nitrosoaniline, bromoacetonitrile and diethyl 2-bromomalonate



11 **9** **26**

Figure S53: Reaction reported between *N,N*-dimethyl-4-nitrosoaniline and diethyl-2-bromomalonate.

Bromoacetonitrile (2 mmol, 0.41 ml), N,N-Dimethyl-4-nitrosoaniline (2 mmol, 0.3 g) and Diethyl 2-bromomalonate (2 mmol, 0.34 ml) are mixed in 6 ml of DMSO. The mixture is stirred at room temperature for 3 hours. The reaction is then diluted with water and extracted with ethyl acetate. The organic phase is separated and washed three times with brine. Mg_2SO_4 is then added to the reaction mixture and after filtration the solvent is removed under vacuum. The crude is purified with a chromatographic column (silica gel, Hexane / EtOAc 10:1). The reaction is an addition of nitroso group on the diethyl-2-bromomalonate to form the respective nitrone. It is already known in literature (16) where it involves sodium hydroxide and THF as solvent. Following the same procedure, the exact molecule has been synthesized by El Hassn et al. in 2006 (17).

Yield: 18% (111 mg)
IUPAC name: N-(4-(dimethylamino)phenyl)-C,C-diethoxy-carbonylnitrone
^{1}H NMR (600 MHz, $CDCl_3$) δ 7.37 (d, *J* = 9.1 Hz, 1H), 6.63 (d, *J* = 9.1 Hz, 2H), 4.43 (q, *J* = 7.1 Hz, 2H), 4.19 (q, *J* = 7.1 Hz, 2H), 3.05 (s, 6H), 1.40 (t, *J* = 7.1 Hz, 3H), 1.21 (t, *J* = 7.1 Hz, 3H)
^{13}C NMR (151 MHz, $CDCl_3$) δ 160.72, 159.43, 151.93, 135.63, 130.37, 124.47 110.18, 61.79, 61.53, 39.79, 13.46
ESI-HR-MS: $[C_{15}H_{20}N_2NaO_5]^+$ Calculated 331.1264 *m/z*, measured 331.1248 *m/z*.

7.4 Diethyl 2-bromomalonate, 4-phenylurazole and 1-vinyl-2-pyrrolidinone

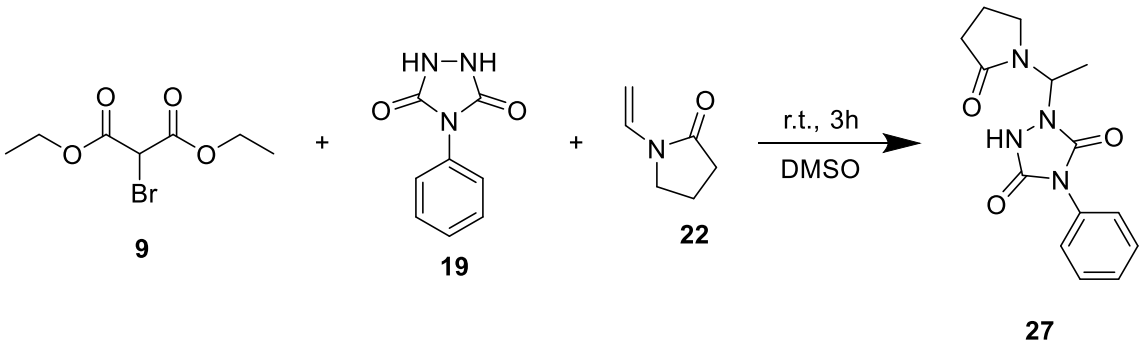


Figure S54: Scheme of the reaction between 4-phenylueazole, 1-vinyl-2-pyrrolidinone, and diethyl-2-bromomalonate.

Diethyl 2-bromomalonate (2 mmol, 0.34 ml), 4-phenylurazole (2 mmol, 0.354 g) and 1-Vinyl-2-pyrrolidinone (2 mmol, 0.23 ml) are mixed in 6 ml of DMSO. The reaction is stirred at room temperature for 3 hours. The mixture is diluted with of water, extracted with ethyl acetate and washed three times with brine. The organic phase is dried with Mg_2SO_4 and the solvent is removed under vacuum. During the evaporation the product precipitates as white crystals, they are filtered and washed with cold ethyl acetate.

The reaction is a nucleophile addition of phenylurazol nitrogen on the pyrrolidinone double bond. A similar reaction has been reported by Senogles (18) *et al.* in 1980 and involved the hydrolysis of 1-vinyl-2-pyrrolidinone in aqueous solutions.

Yield: 16% (92 mg)

IUPAC name: 1-[1-(2-Oxo-pyrrolidin-1-yl)-ethyl]-4-phenyl-[1,2,4]triazolidine-3,5-dione

^{1}H NMR (600 MHz, DMSO) δ 10.69 (s, 1H), 7.42 (m, 5H), 5.90 (s, 1H), 3.46 (m, *J* = 7.0 Hz, 2H), 2.26 (m, *J* = 3.5 Hz, 2H), 1.96 (m, *J* = 8.7 Hz, 2H), 1.50 (s, 3H)

13C NMR (151 MHz, DMSO) δ 174.37, 153.00, 152.32, 131.56, 128.91, 128.00, 126.21, 60.13, 43.20, 30.44, 17.77

ESI-HR-MS: $[C_{14}H_{15}N_4O_3]^-$ Calculated 287.1150 *m/z*, measured 287.1430 *m/z*

Crystallographic data: See section 8

8 Crystal structure details

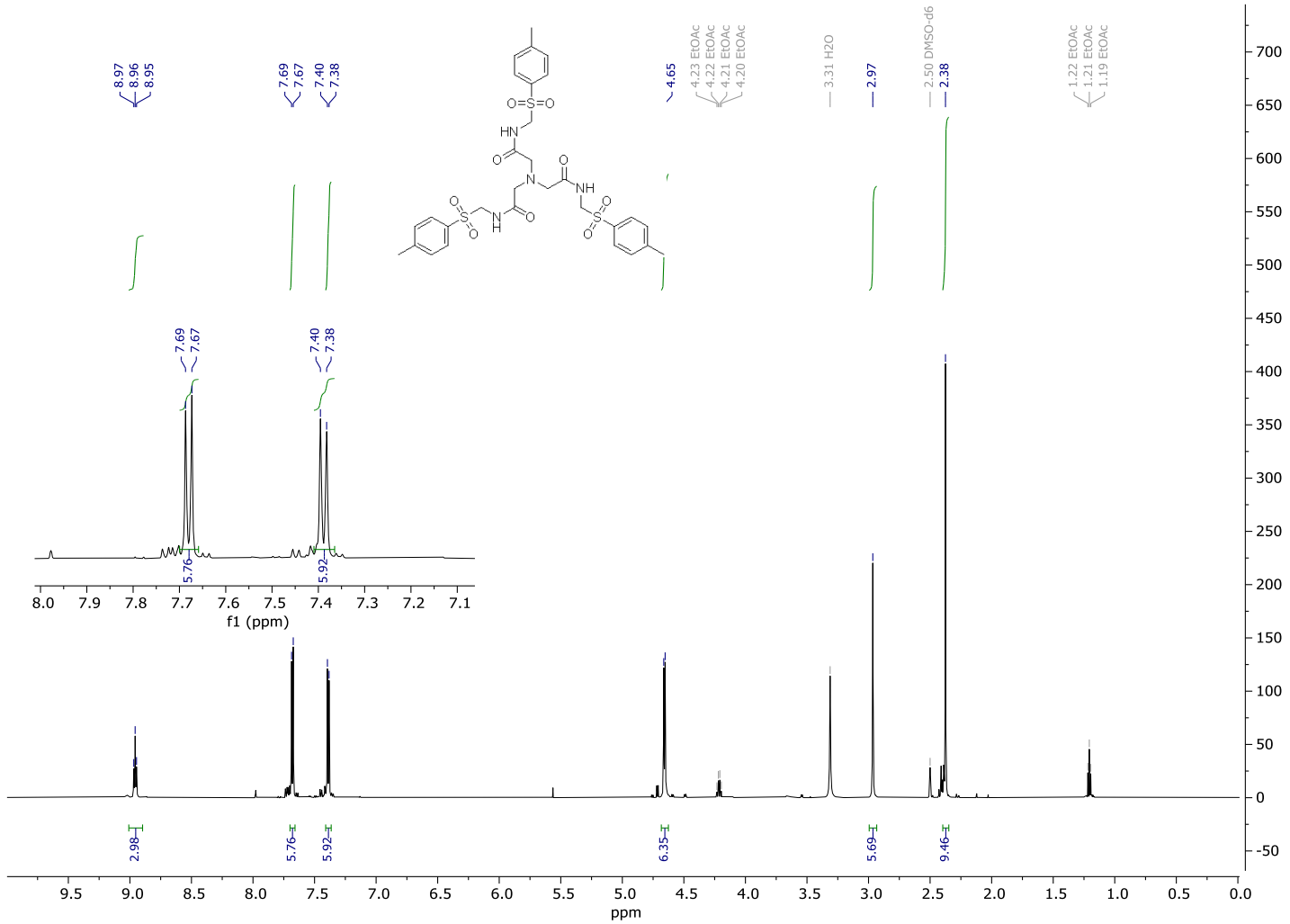
Single Crystal X-ray Diffraction: Suitable single crystals were selected and mounted by using the MiTeGen MicroMountsTM kit with Fomblin oil. X-ray diffraction intensity data were measured at 150(2) K on Bruker Apex II Quasar diffractometer using Mo Kα [λ = 0.71073 Å] radiation. Structure solution and refinement were carried out with SHELXT (19) and SHELXL-2018 (20) via WinGX (22). Corrections for incident and diffracted beam absorption effects were applied using empirical methods (23).

Name	**Compound 29**	
Empirical formula	C31 H39 N4 O9.50 S3.50	
Formula weight	731.87	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Trigonal	
Space group	P -3 c 1	
Unit cell dimensions	a = 22.260(8) Å	α= 90°
	b = 22.260(8) Å	β= 90°
	c = 9.328(4) Å	γ = 120°
Volume	4003(3) Å3	
Z	4	
Density (calculated)	1.214 Mg/m3	
Absorption coefficient	0.263 mm-1	
F(000)	1540	
Crystal size	0.232 x 0.030 x 0.016 mm3	
Theta range for data collection	2.113 to 23.318°	
Index ranges	-24<=h<=24, -24<=k<=24, -10<=l<=10	
Reflections collected	30636	
Independent reflections	1937 [R(int) = 0.2800]	
Completeness to theta = 23.318°	99.70%	
Max. and min. transmission	0.745 and 0.634	
Refinement method	Full-matrix least-squares on F2	
Data / restraints / parameters	1937 / 0 / 140	
Goodness-of-fit on F2	1.043	
Final R indices [I>2sigma(I)]	R1 = 0.0631, wR2 = 0.1369	
R indices (all data)	R1 = 0.1366, wR2 = 0.1760	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.31 and -0.45 e.Å-3	

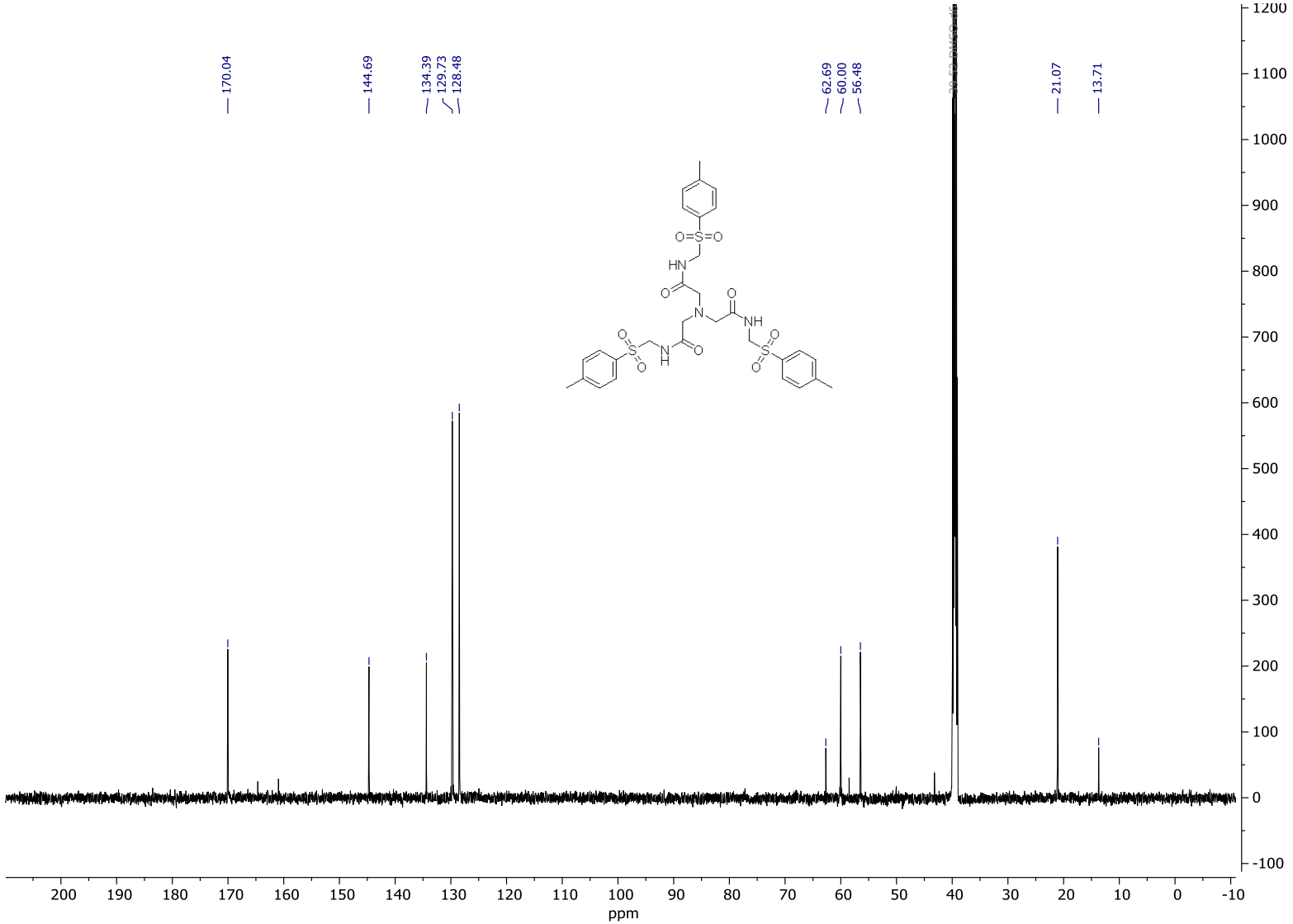
Name	**Compound 28**	
Empirical formula	$C_8H_5BrN_3$	
Formula weight	223.06	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	Pna21	
Unit cell dimensions	a = 12.777(4) Å	α= 90°
	b = 4.5611(14) Å	β= 90°
	c = 14.595(5) Å	γ = 90°
Volume	850.6(5) $Å^3$	
Z	4	
Density (calculated)	1.742 Mg/m^3	
Absorption coefficient	4.775 mm^{-1}	
F(000)	436	
Crystal size	0.157 x 0.052 x 0.040 mm^3	
Theta range for data collection	2.791 to 25.993°	
Index ranges	-15<=h<=15, -5<=k<=5, -18<=l<=18	
Reflections collected	13942	
Independent reflections	1677 [R(int) = 0.0588]	
Completeness to theta = 25.242°	100.00%	
Absorption correction	Empirical	
Max. and min. transmission	0.728 and 0.605	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	1677 / 1 / 109	
Goodness-of-fit on F^2	1.107	
Final R indices [I>2sigma(I)]	R_1 = 0.0520, wR_2 = 0.1558	
R indices (all data)	R_1 = 0.0582, wR_2 = 0.1615	
Absolute structure parameter	0.082(12)	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.75 and -1.17 $e.Å^{-3}$	

Name	**Compound 27**	
Empirical formula	$C_{14}H_{16}N_4O_3$	
Formula weight	288.31	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	Pbca	
Unit cell dimensions	a = 11.972(4) Å	α= 90°
	b = 12.045(5) Å	β= 90°
	c = 18.406(7) Å	γ = 90°
Volume	2654.1(17) $Å^3$	
Z	8	
Density (calculated)	1.443 Mg/m^3	
Absorption coefficient	0.105 mm^{-1}	
F(000)	1216	
Crystal size	0.140 x 0.042 x 0.012 mm^3	
Theta range for data collection	2.213 to 26.000°	
Index ranges	-14<=h<=14, -14<=k<=14, -22<=l<=22	
Reflections collected	30451	
Independent reflections	2600 [R(int) = 0.1131]	
Completeness to theta = 25.242°	100.00%	
Absorption correction	Empirical	
Max. and min. transmission	1.000 and 0.852	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	2600 / 0 / 191	
Goodness-of-fit on F^2	1.032	
Final R indices [I>2sigma(I)]	R_1 = 0.0514, wR_2 = 0.1132	
R indices (all data)	R_1 = 0.0938, wR_2 = 0.1368	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.33 and -0.55 e.$Å^{-3}$	

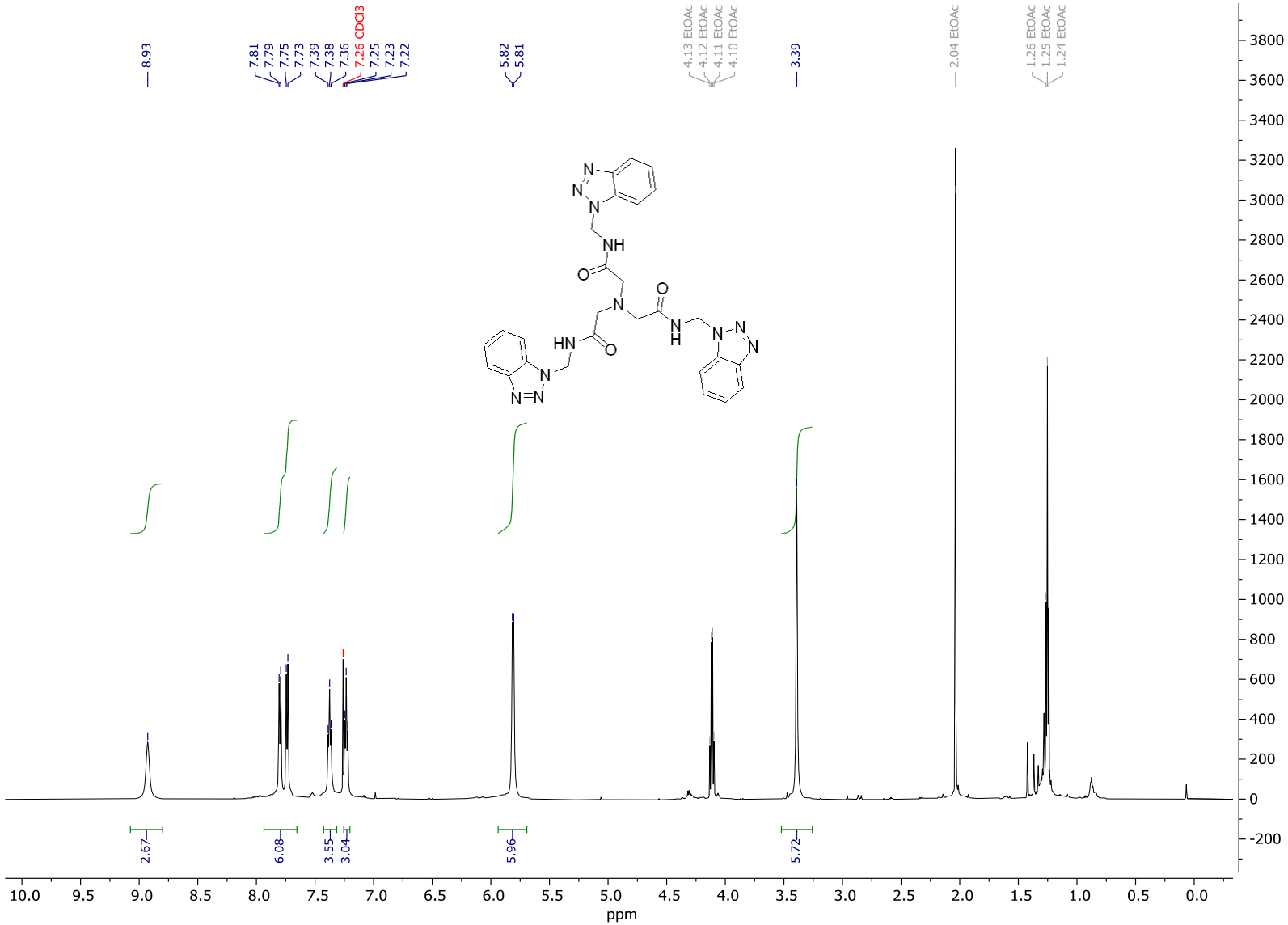
9 ^{1}H and ^{13}C NMR spectra



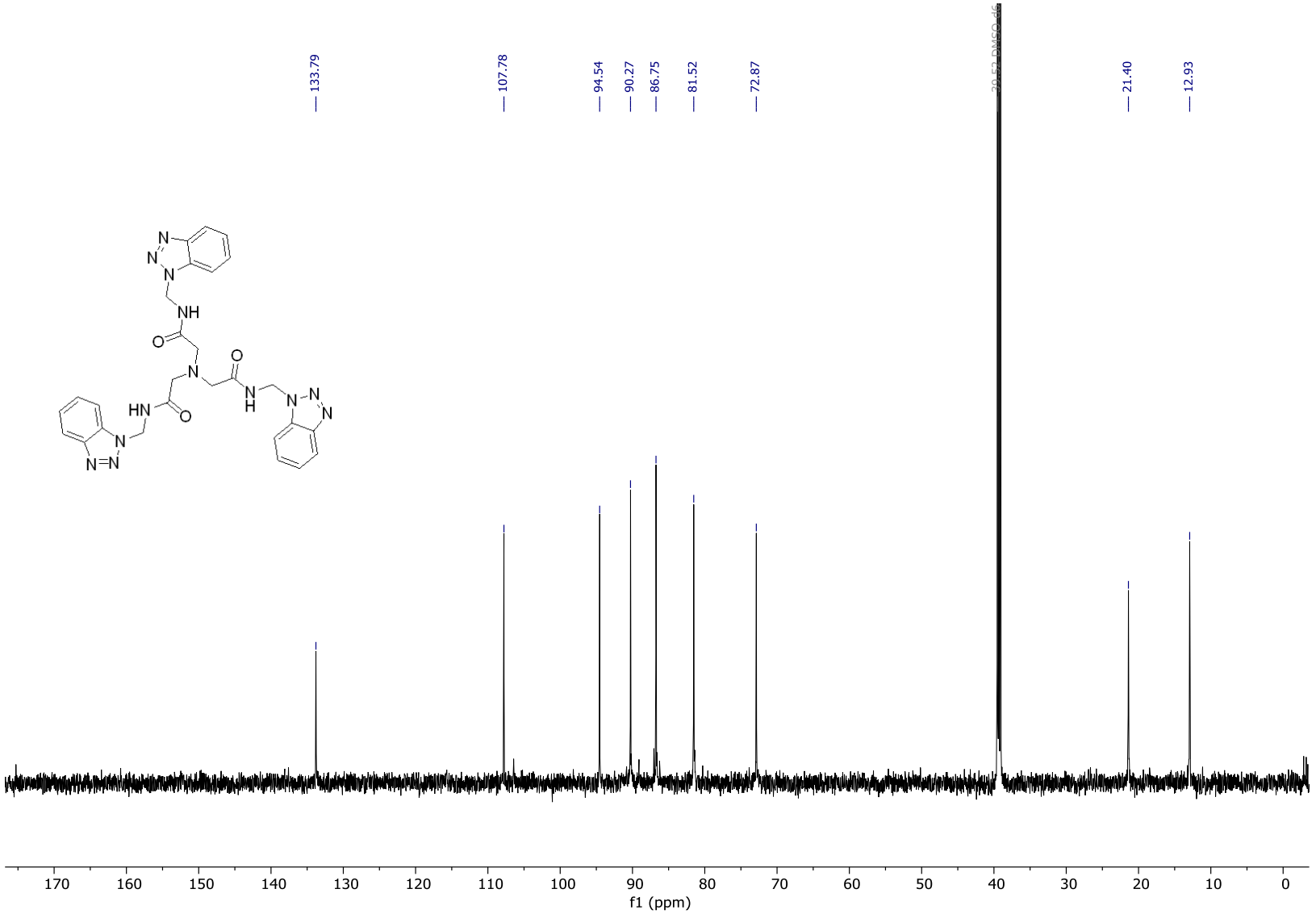
Spectrum S1: ^{1}H-NMR of Product **29**



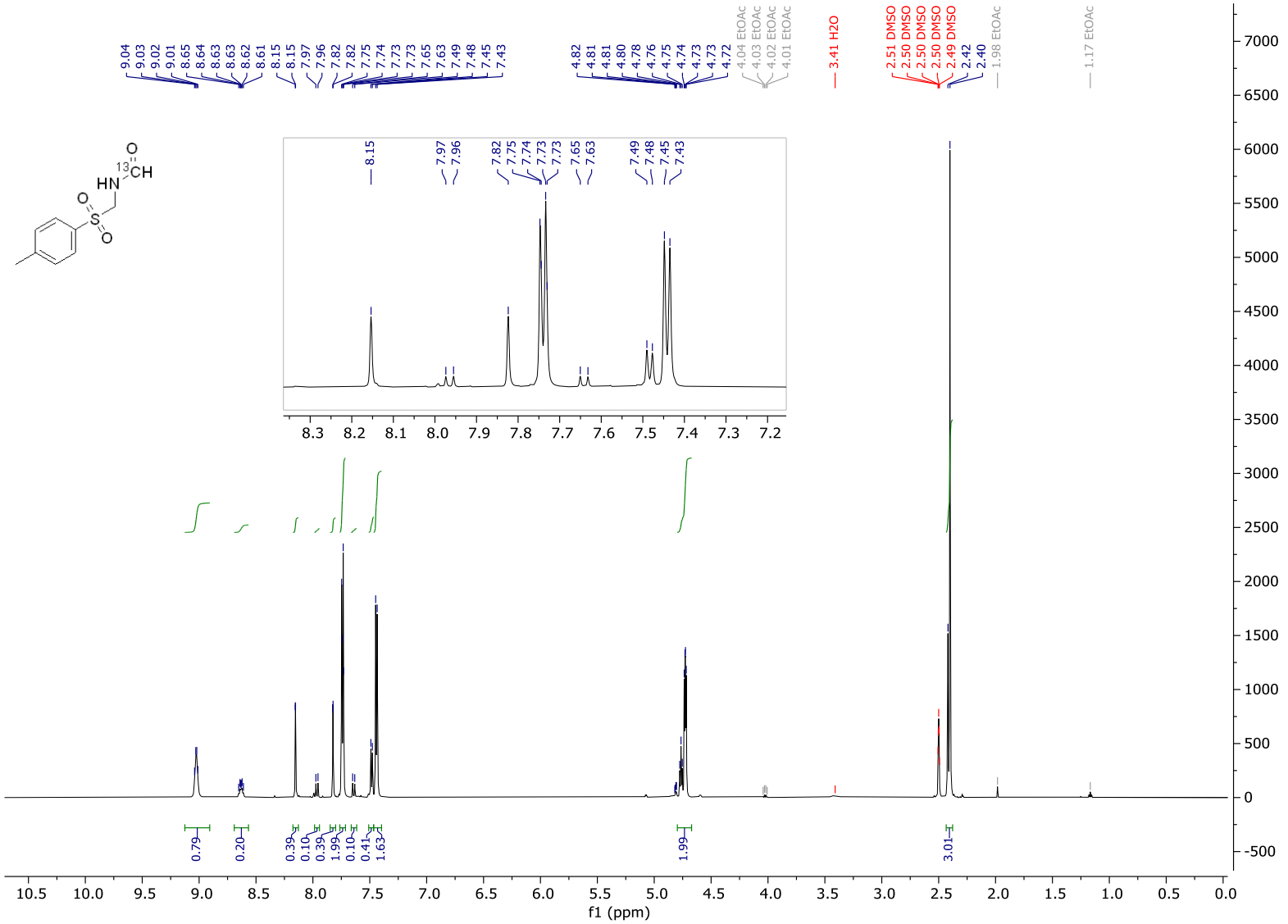
Spectrum S2: ^{13}C-NMR of Product **29**



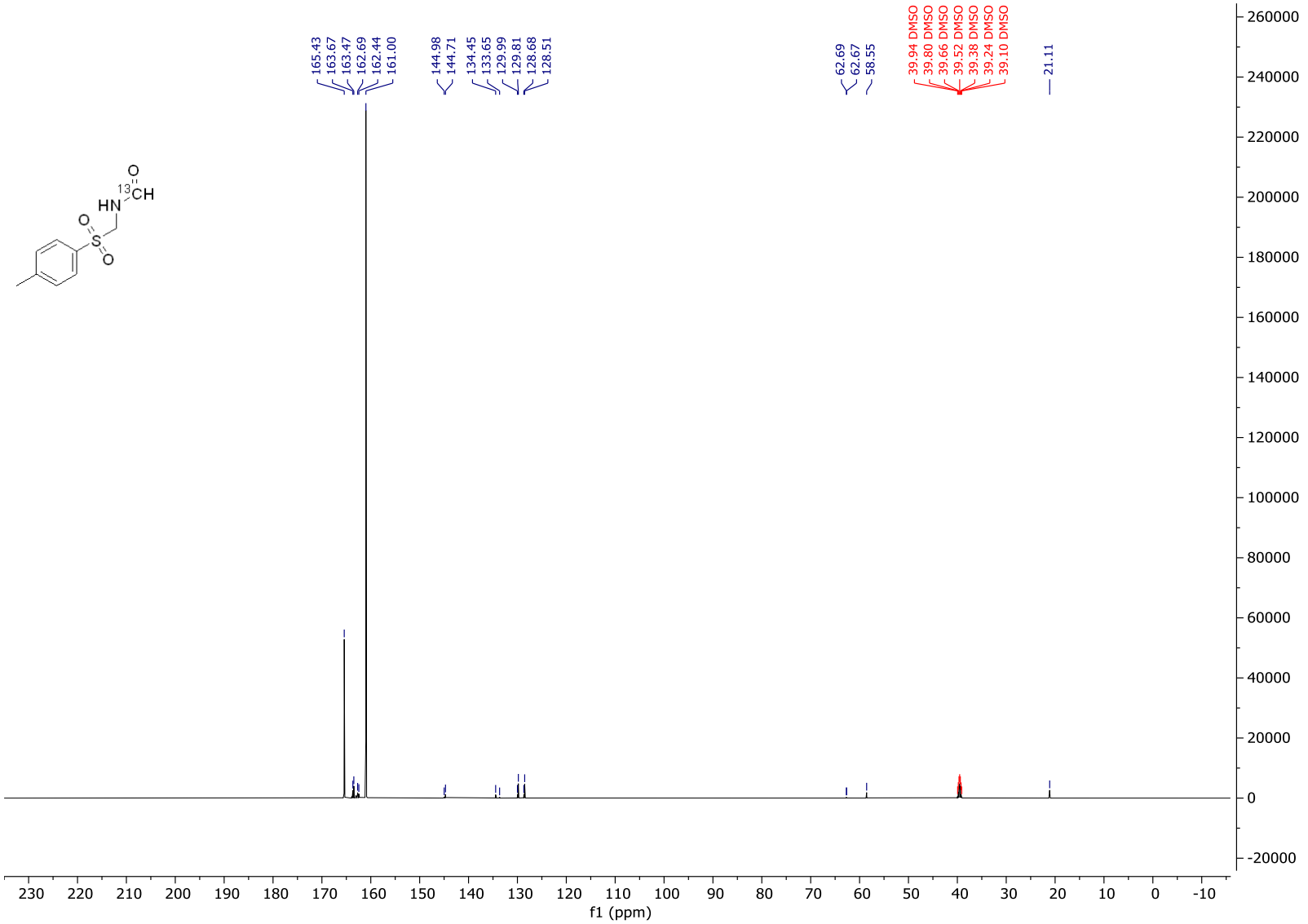
Spectrum S3: ^{1}H-NMR of Product **32**



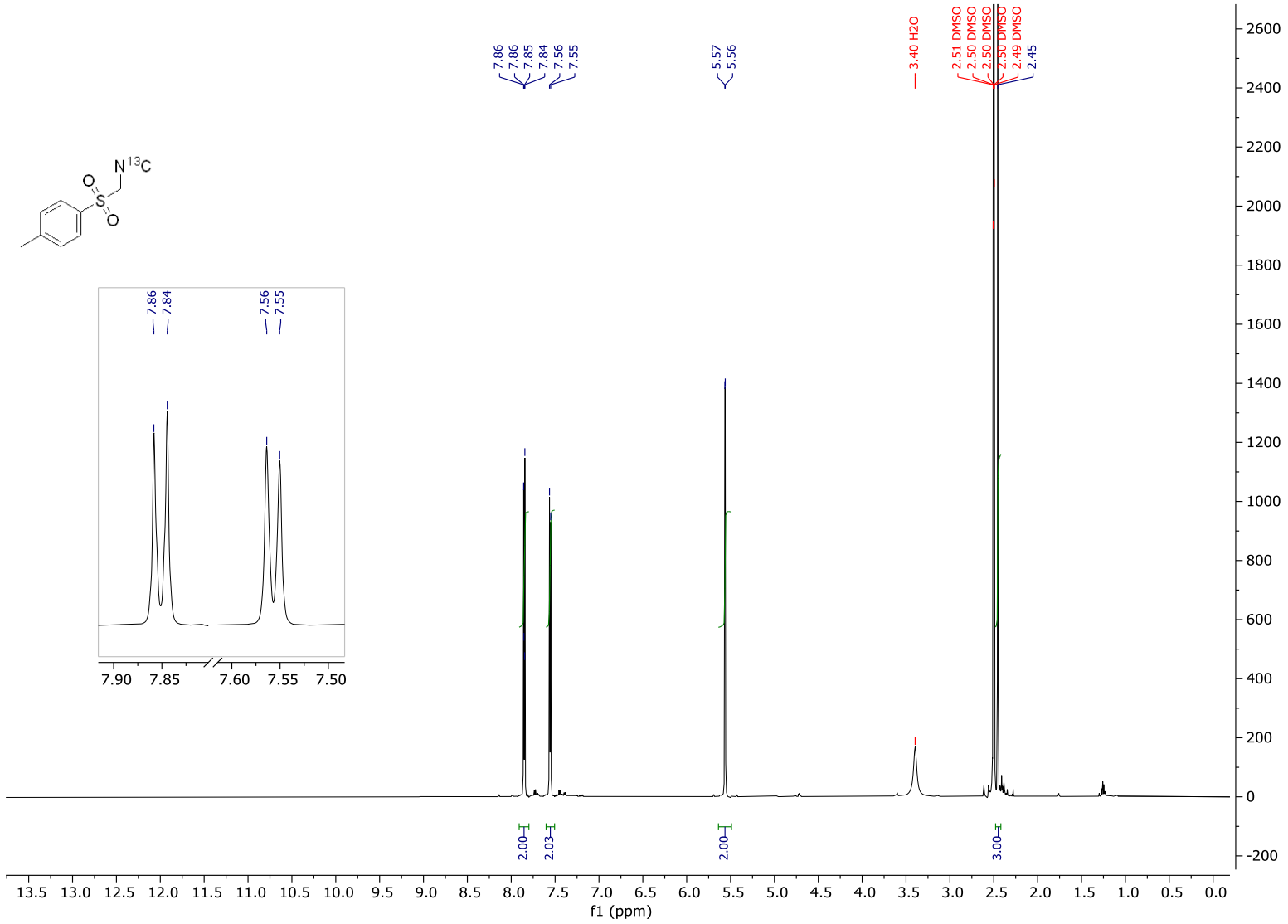
Spectrum S4: ^{13}C-NMR of Product **32**



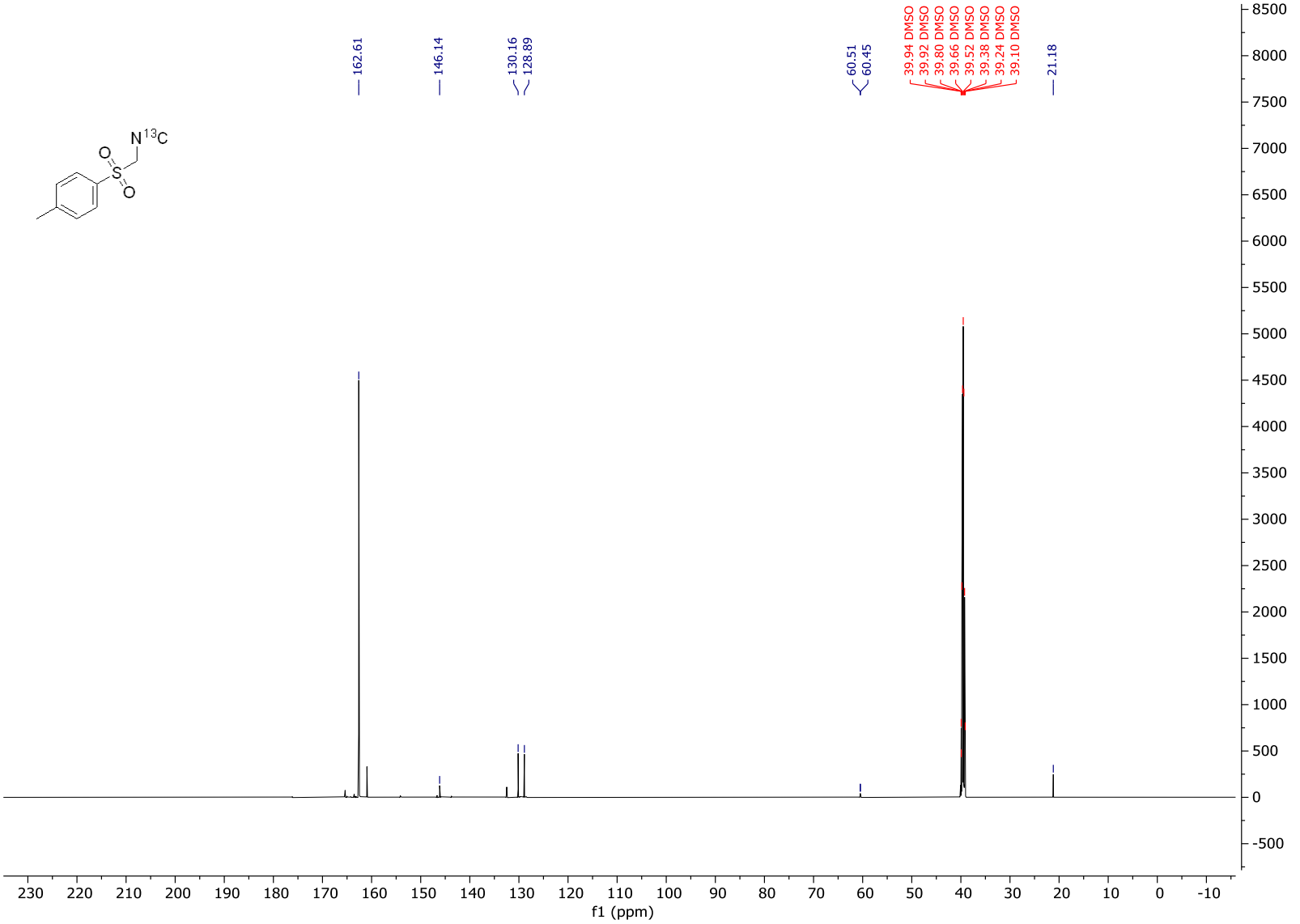
Spectrum S5: ^{1}H-NMR of (1-^{13}C)N-(tosylmethyl)-formamide



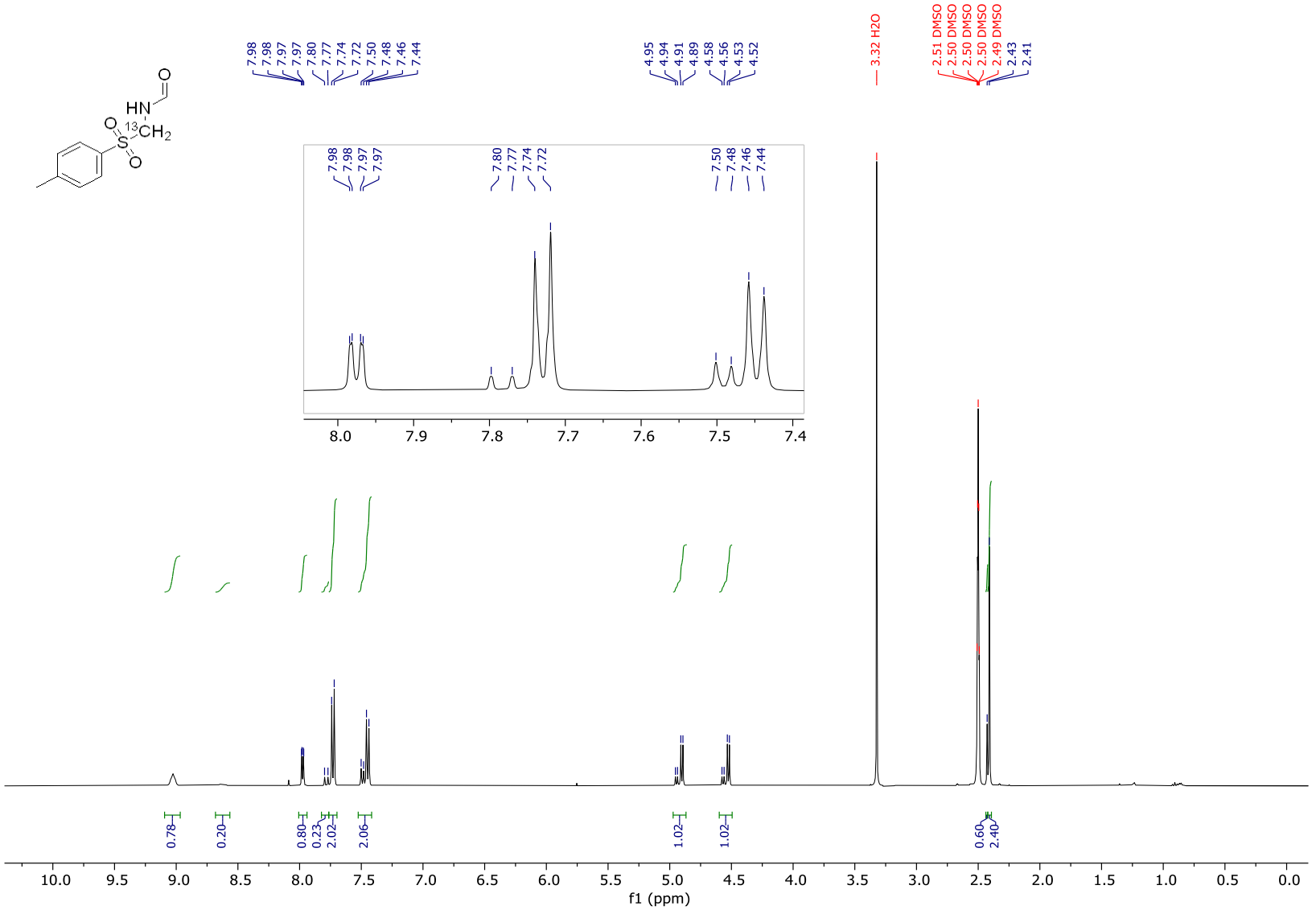
Spectrum S6: ^{13}C-NMR of (1-^{13}C)N-(tosylmethyl)-formamide



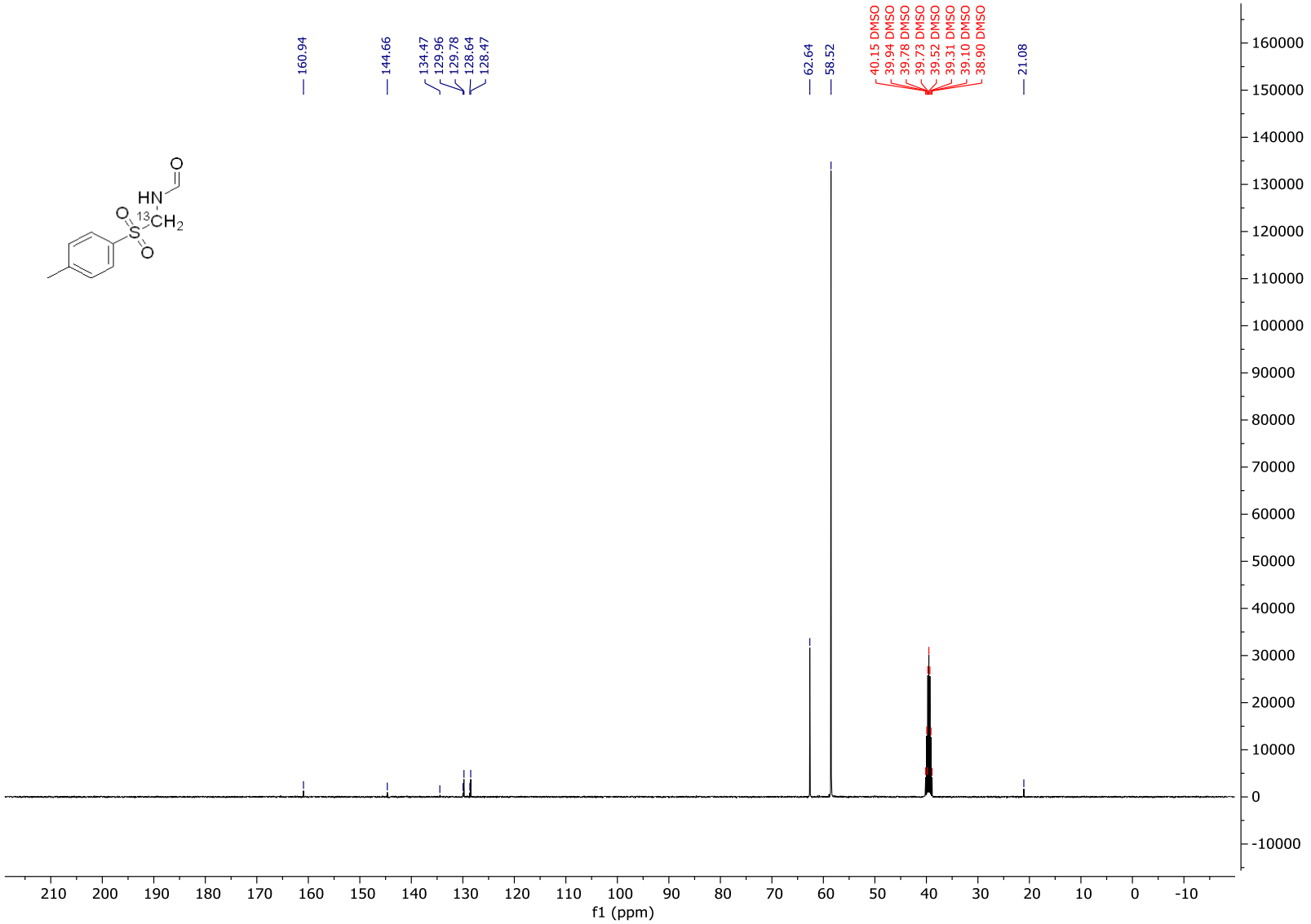
Spectrum S7: ^{1}H-NMR of $TosCH_2N^{13}C$



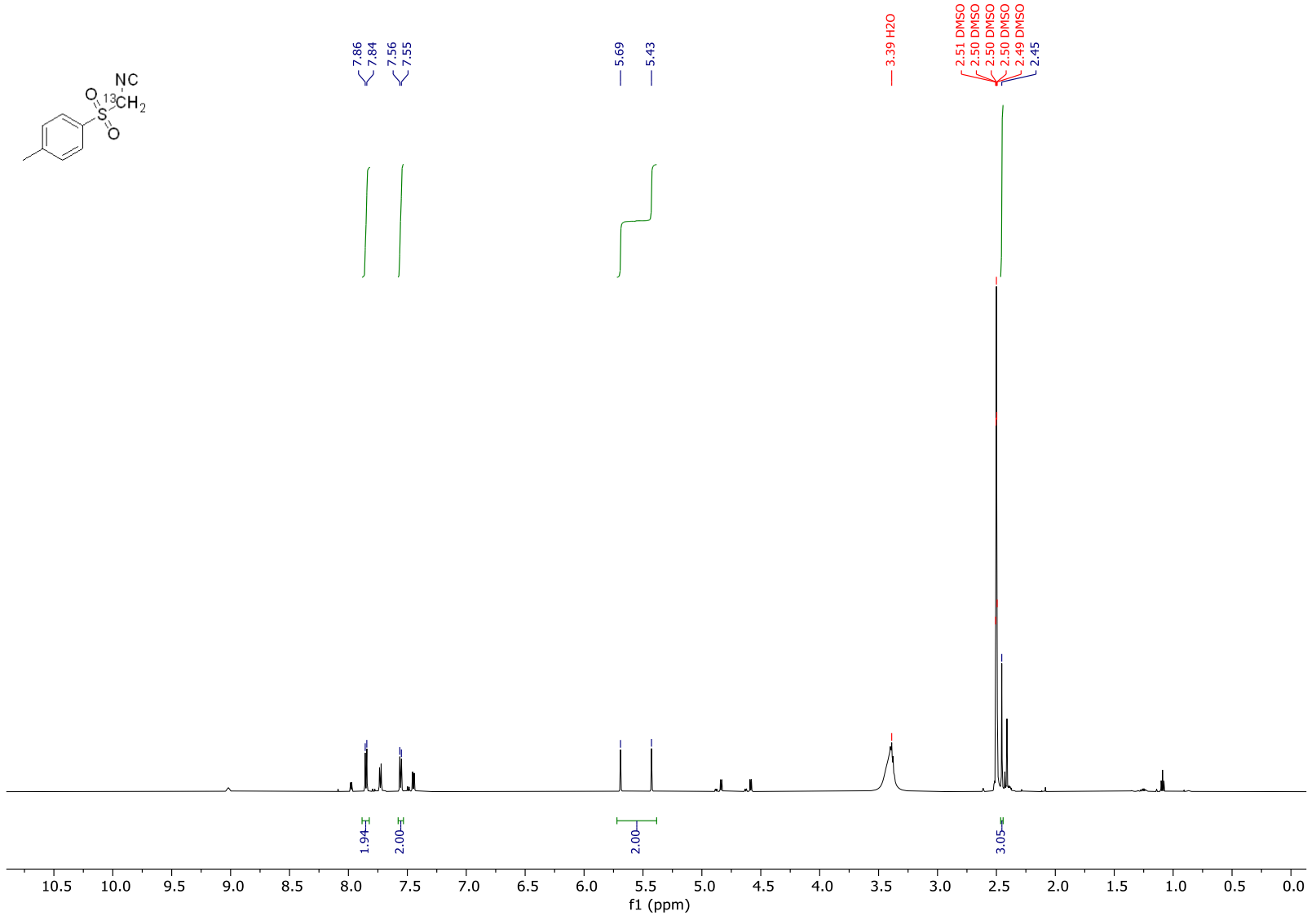
Spectrum S8: ^{13}C-NMR of $TosCH_2N^{13}C$



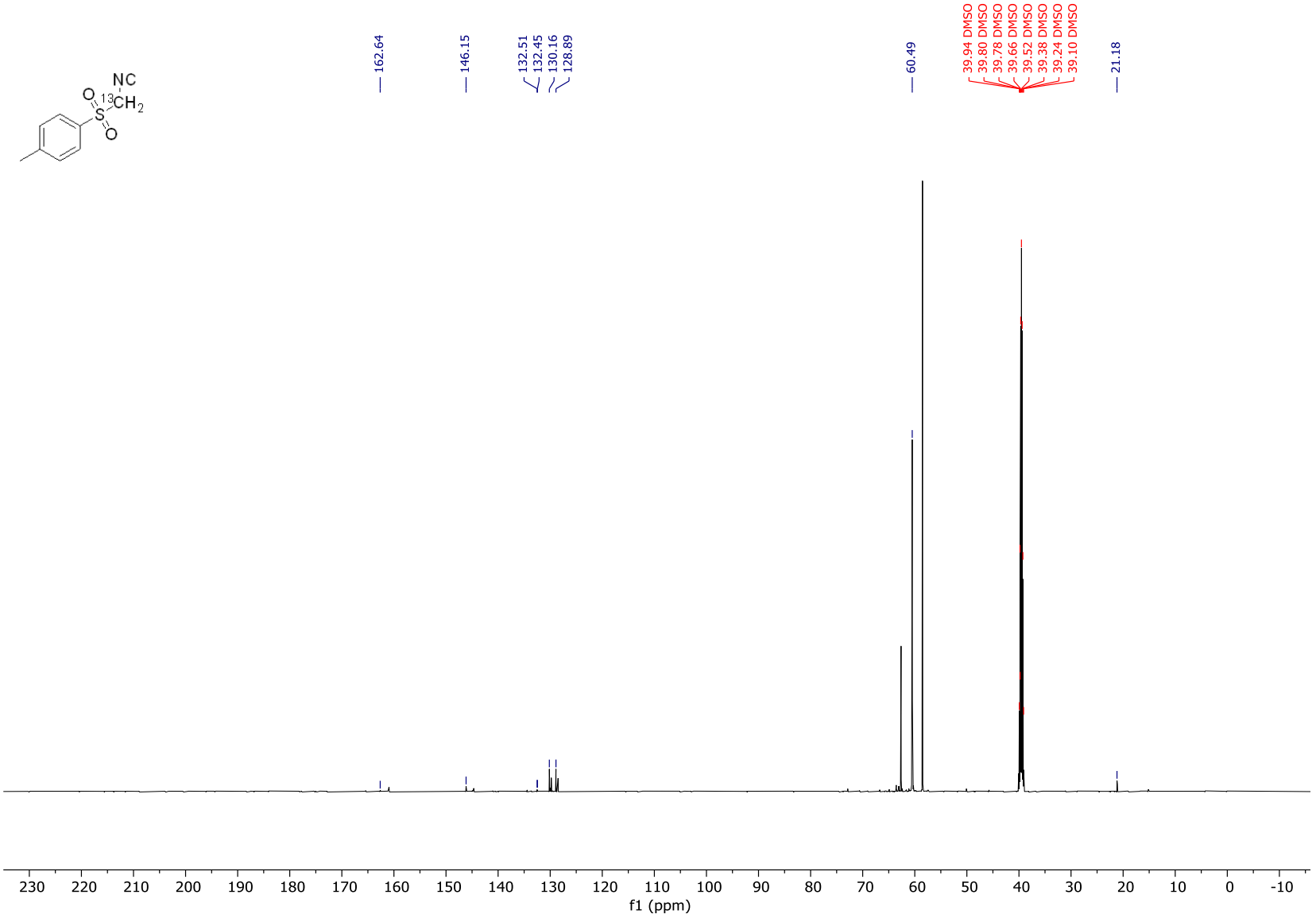
Spectrum S9: ^{1}H-NMR of N-(tosyl(^{13}C)methyl)-formamide



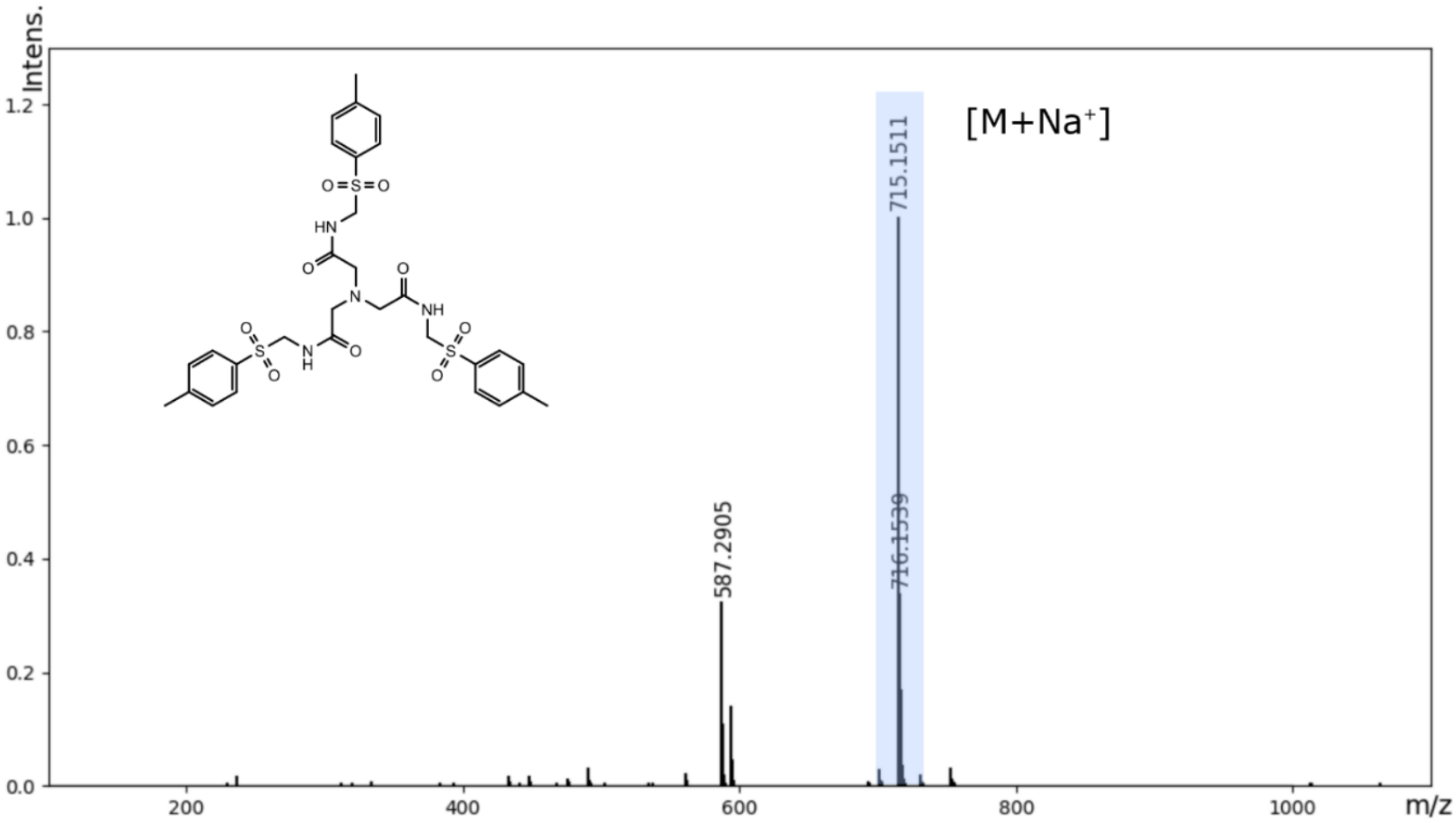
Spectrum S10: ^{13}C-NMR of N-(tosyl(^{13}C)methyl)-formamide



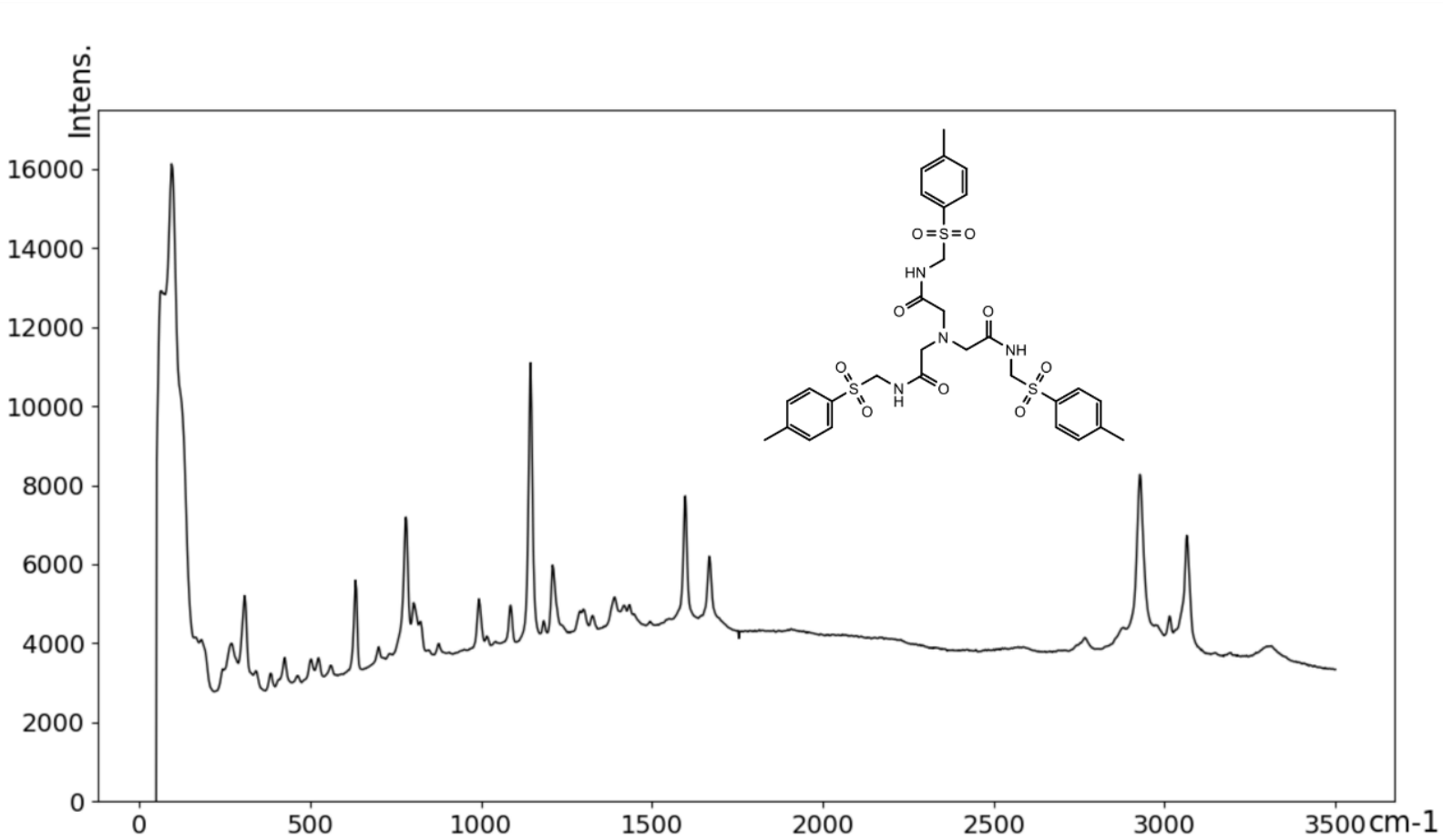
Spectrum S11: ^{1}H-NMR of Tos^{13}CH$_2$NC



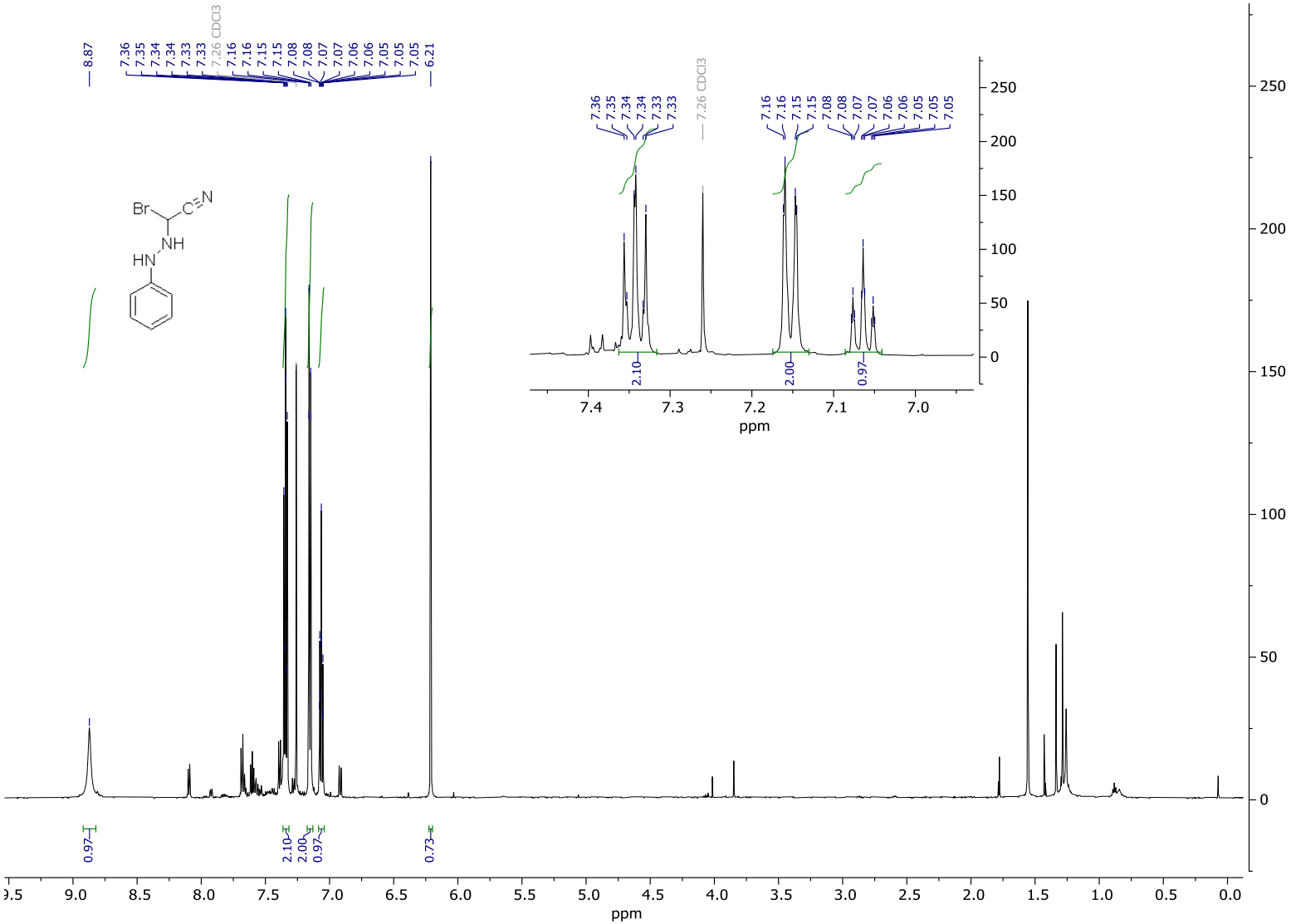
Spectrum S12: ^{13}C-NMR of Tos$^{13}CH_2NC$



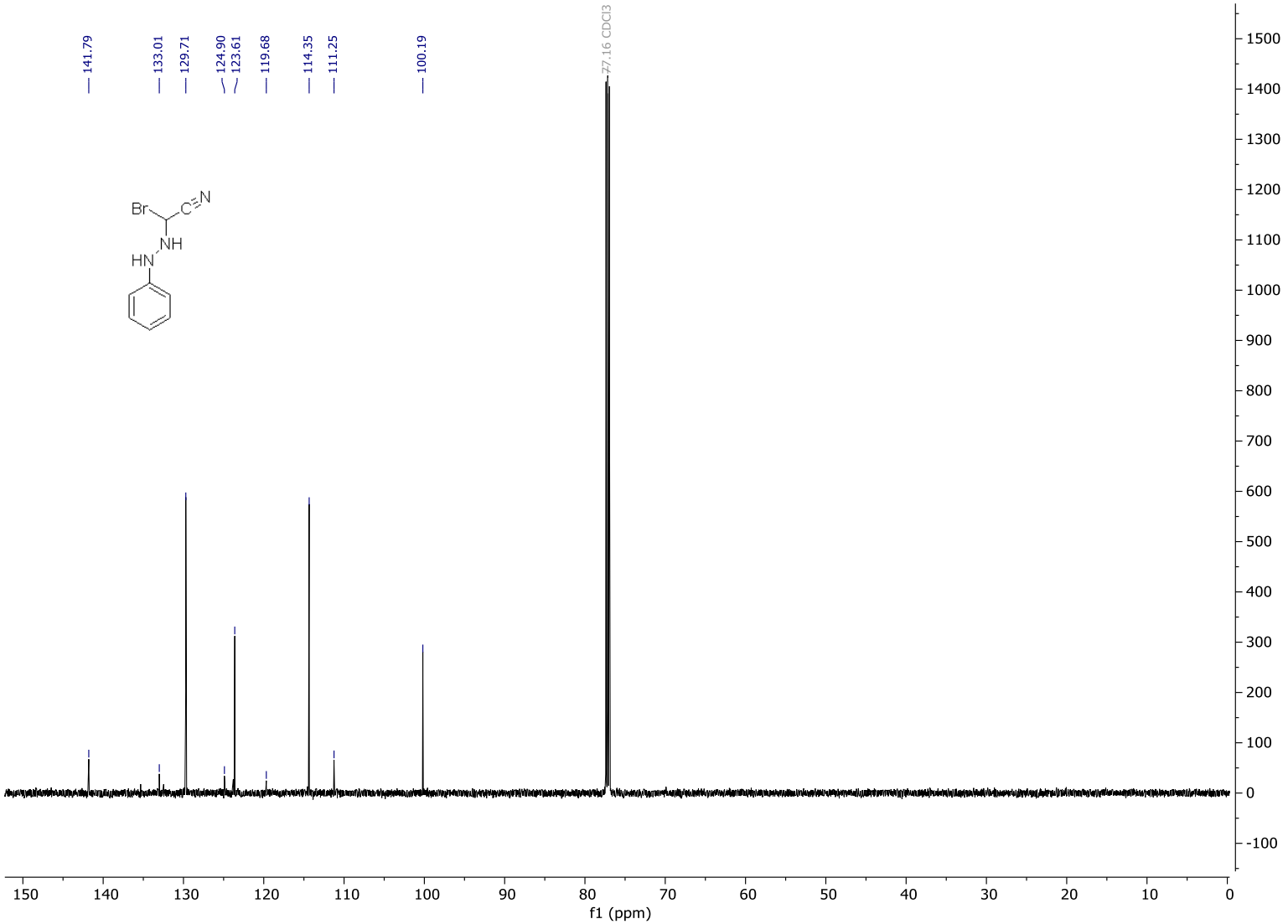
Spectrum S13: MS spectrum of compound **29**



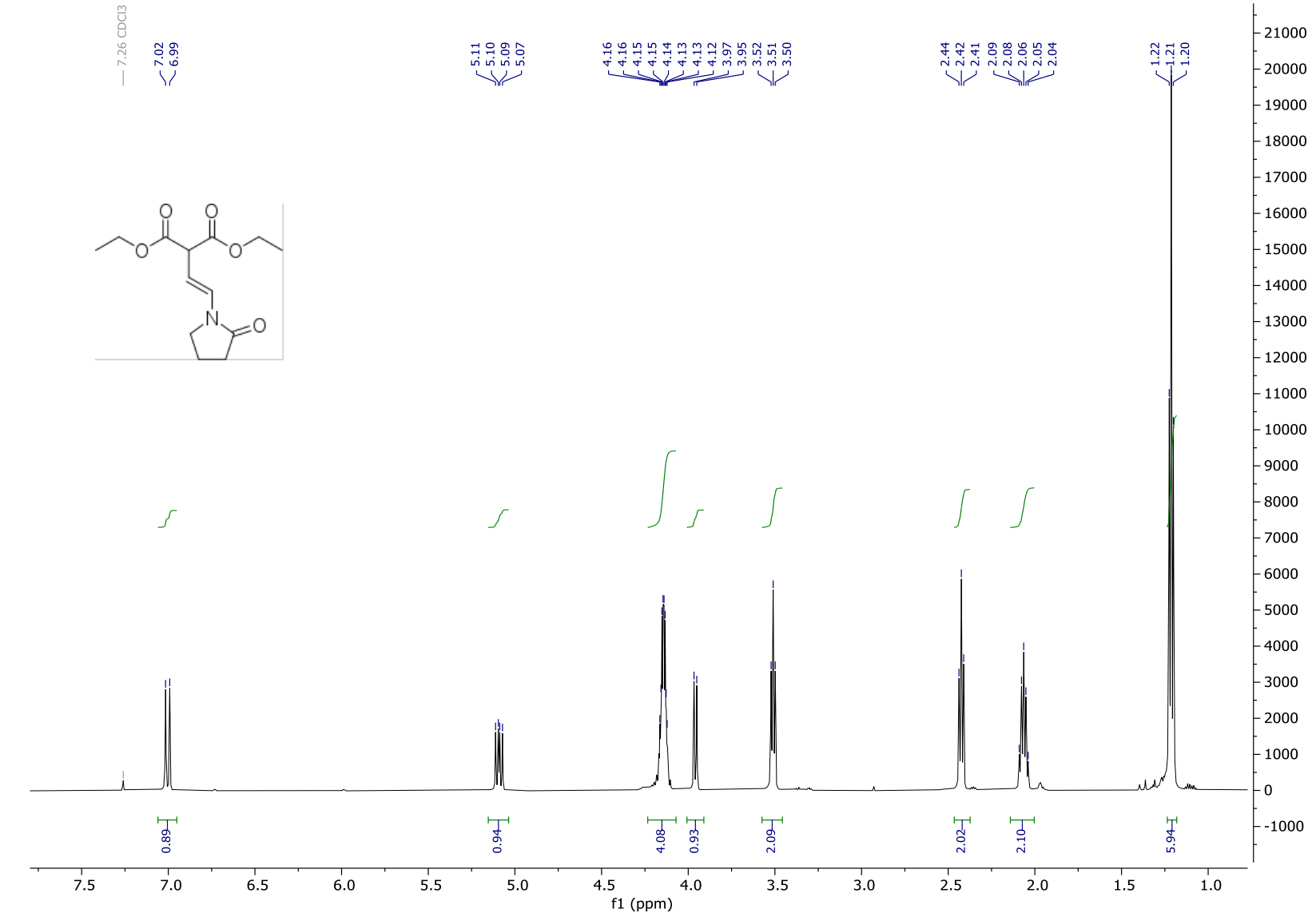
Spectrum S14: Raman spectrum of compound **29**



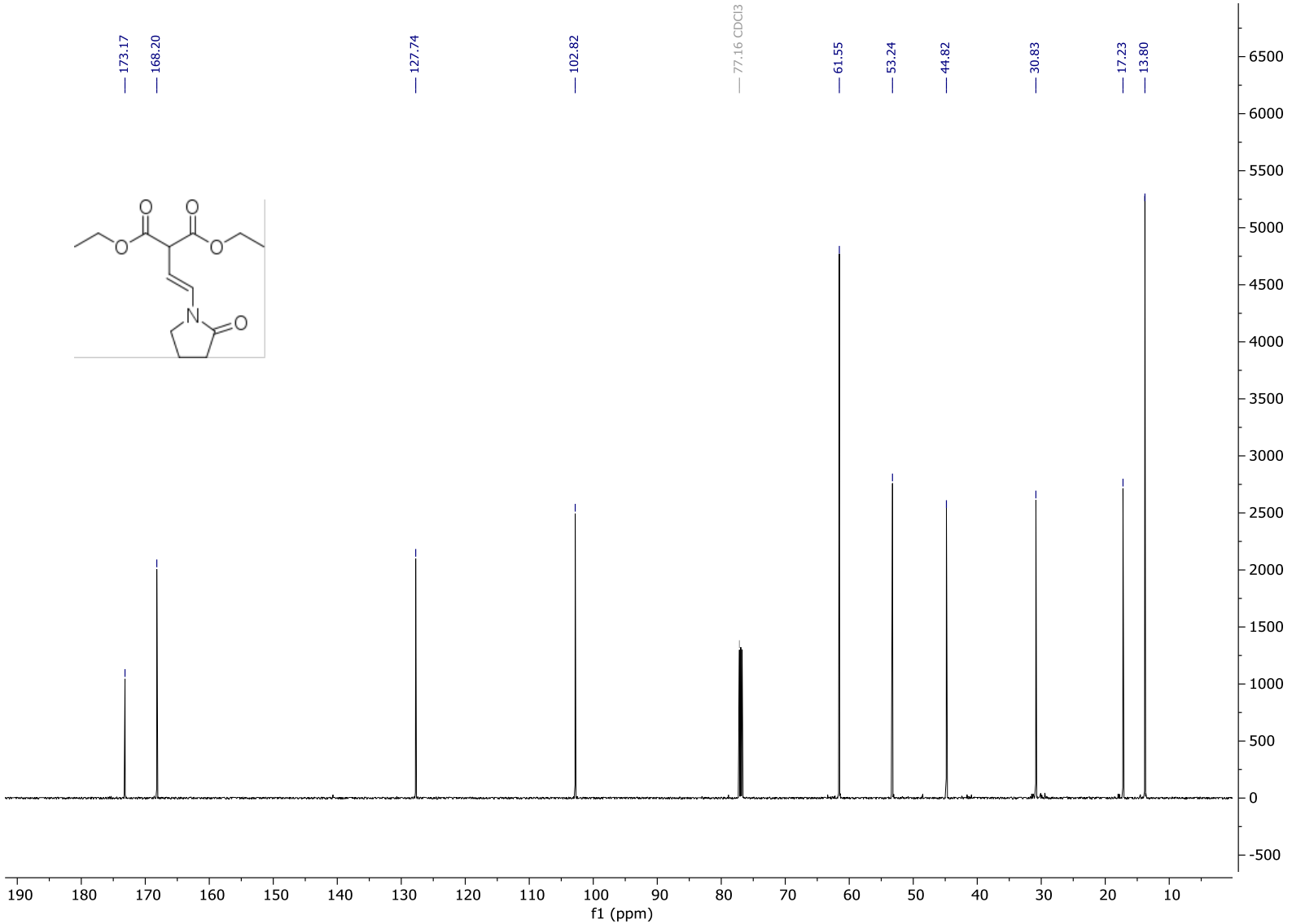
Spectrum S15: ^{1}H-NMR of compound **28**



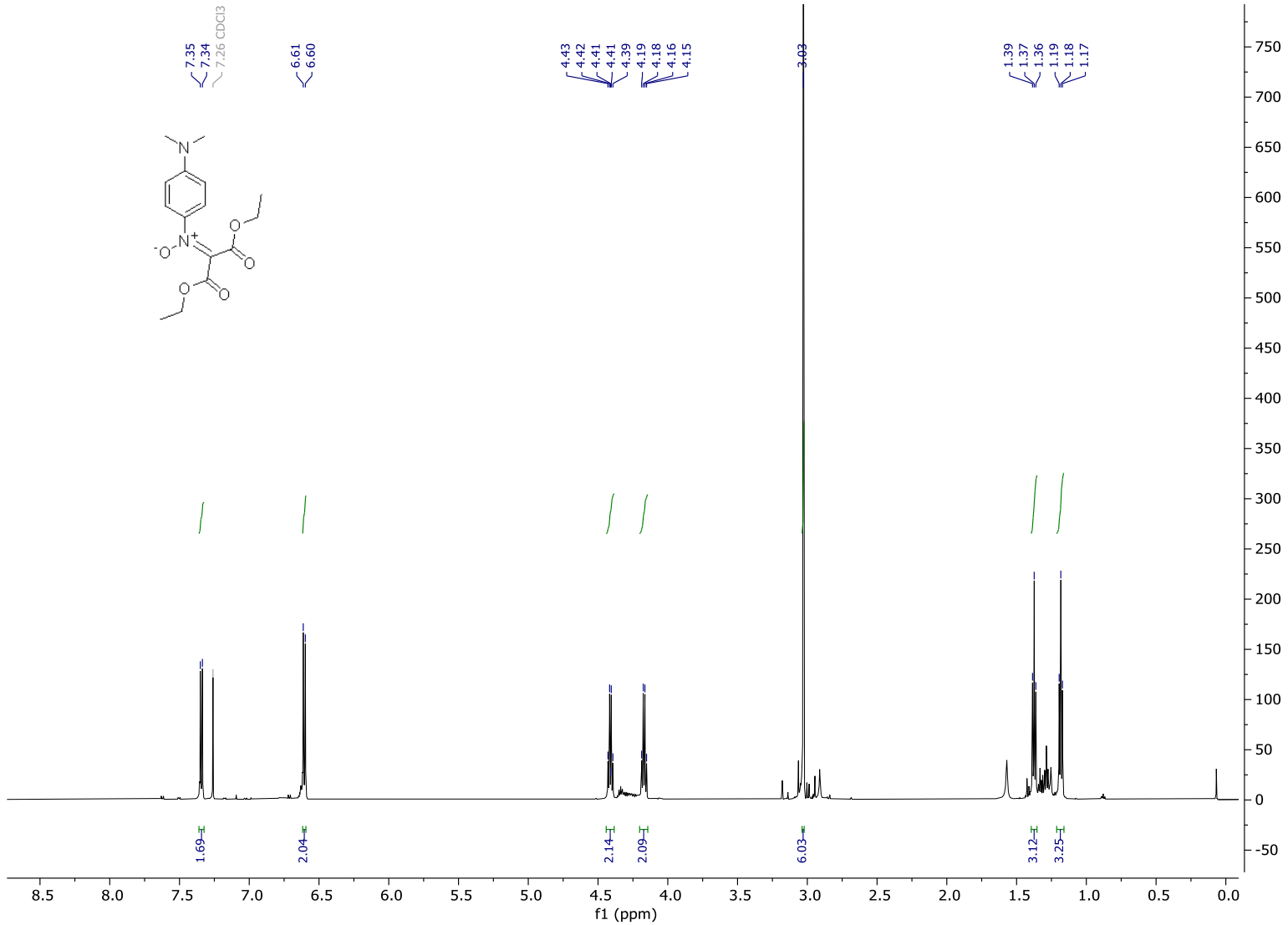
Spectrum S16: ^{13}C-NMR of compound **28**



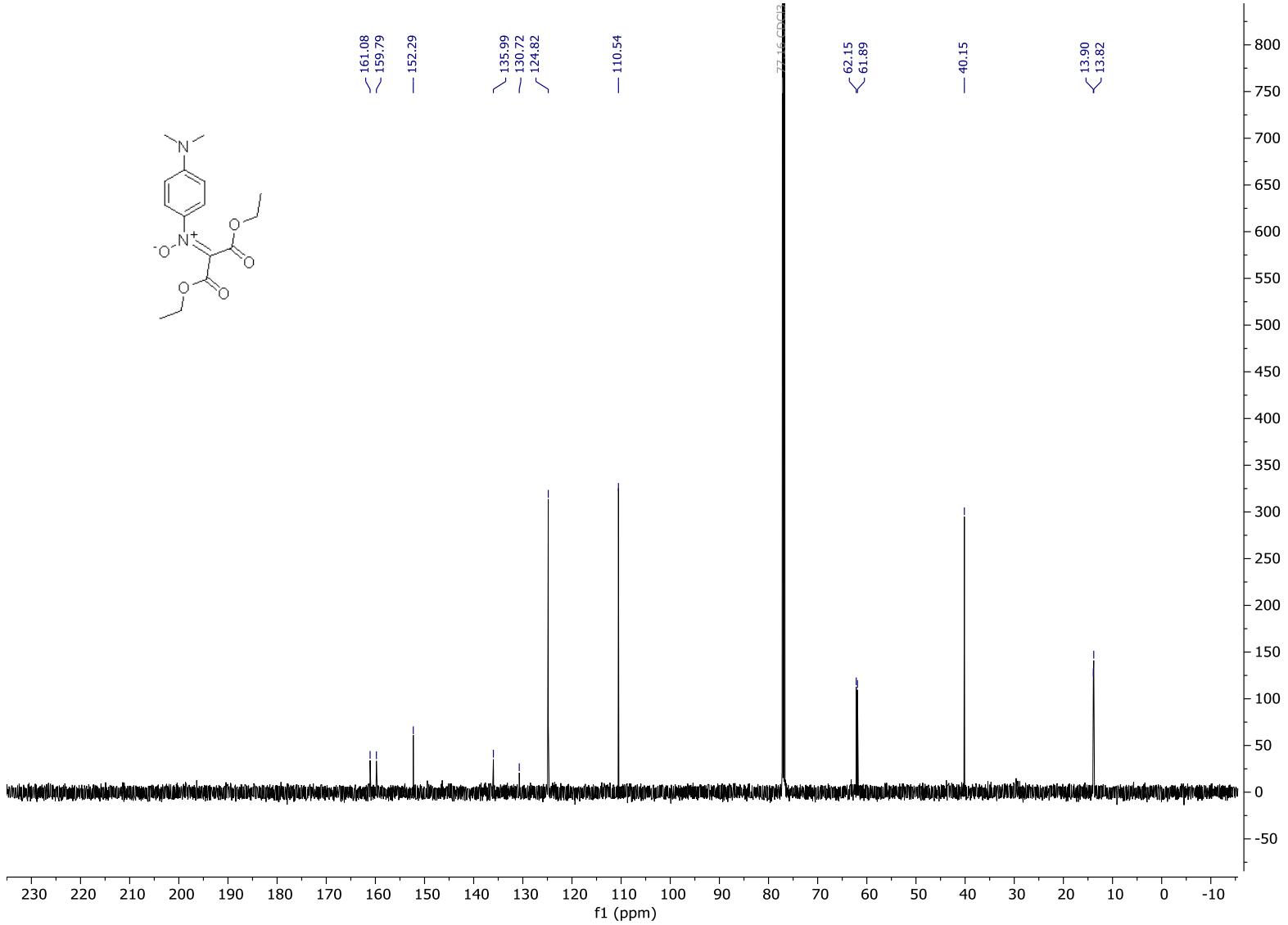
Spectrum S17: ^{1}H-NMR of compound **25**



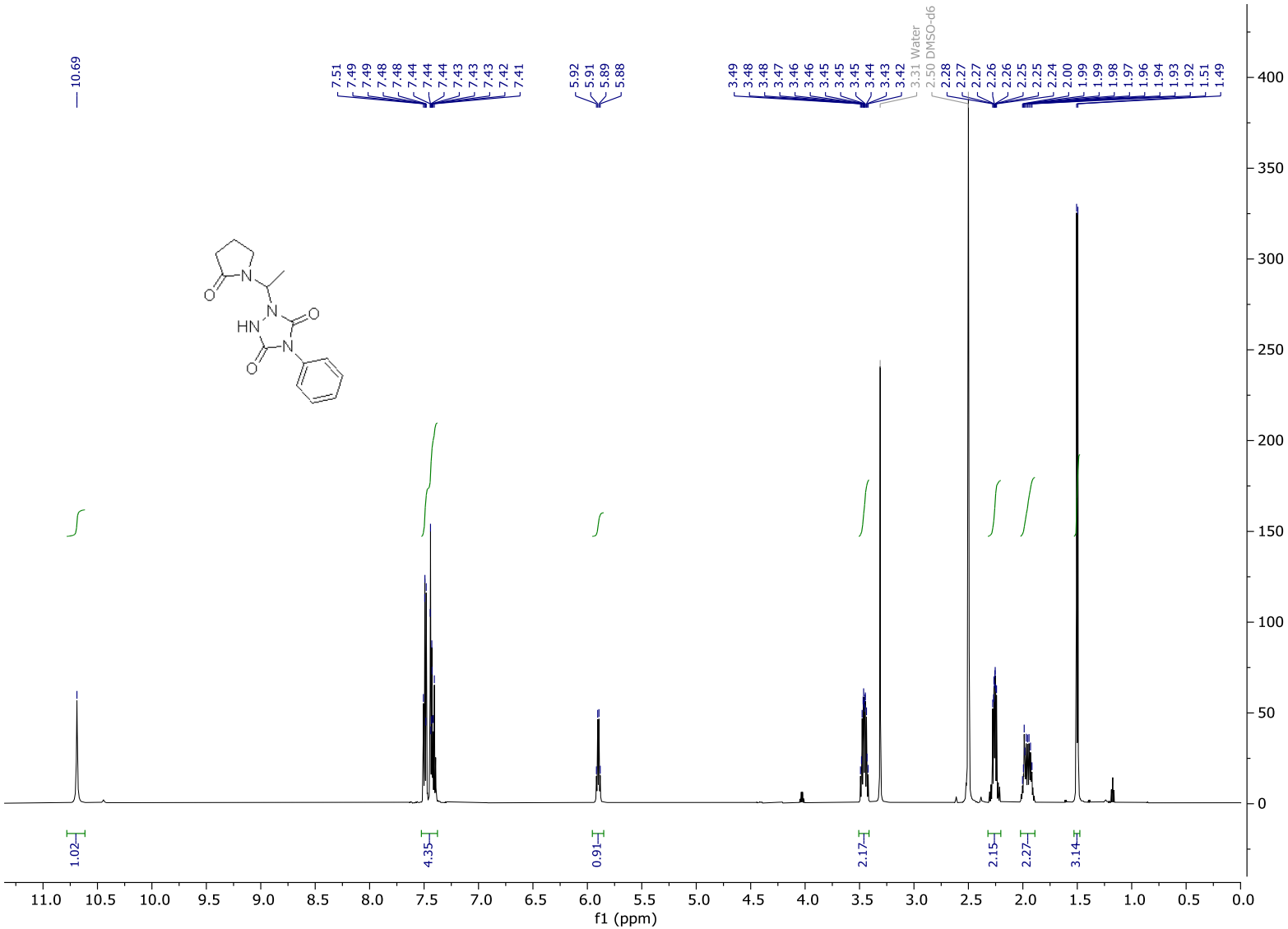
Spectrum S18: ^{13}C-NMR of compound **25**



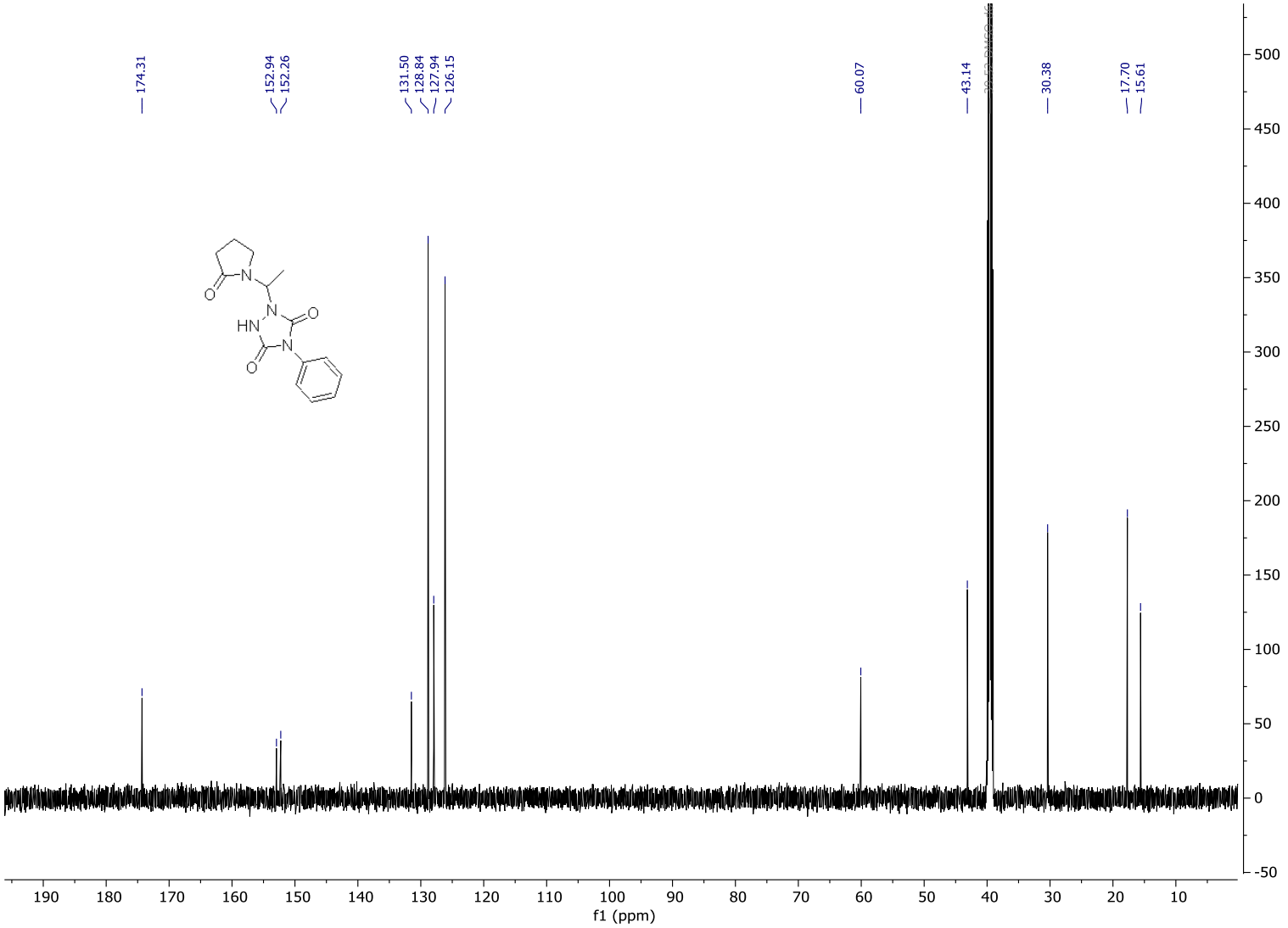
Spectrum S19: ^{1}H-NMR of compound **26**



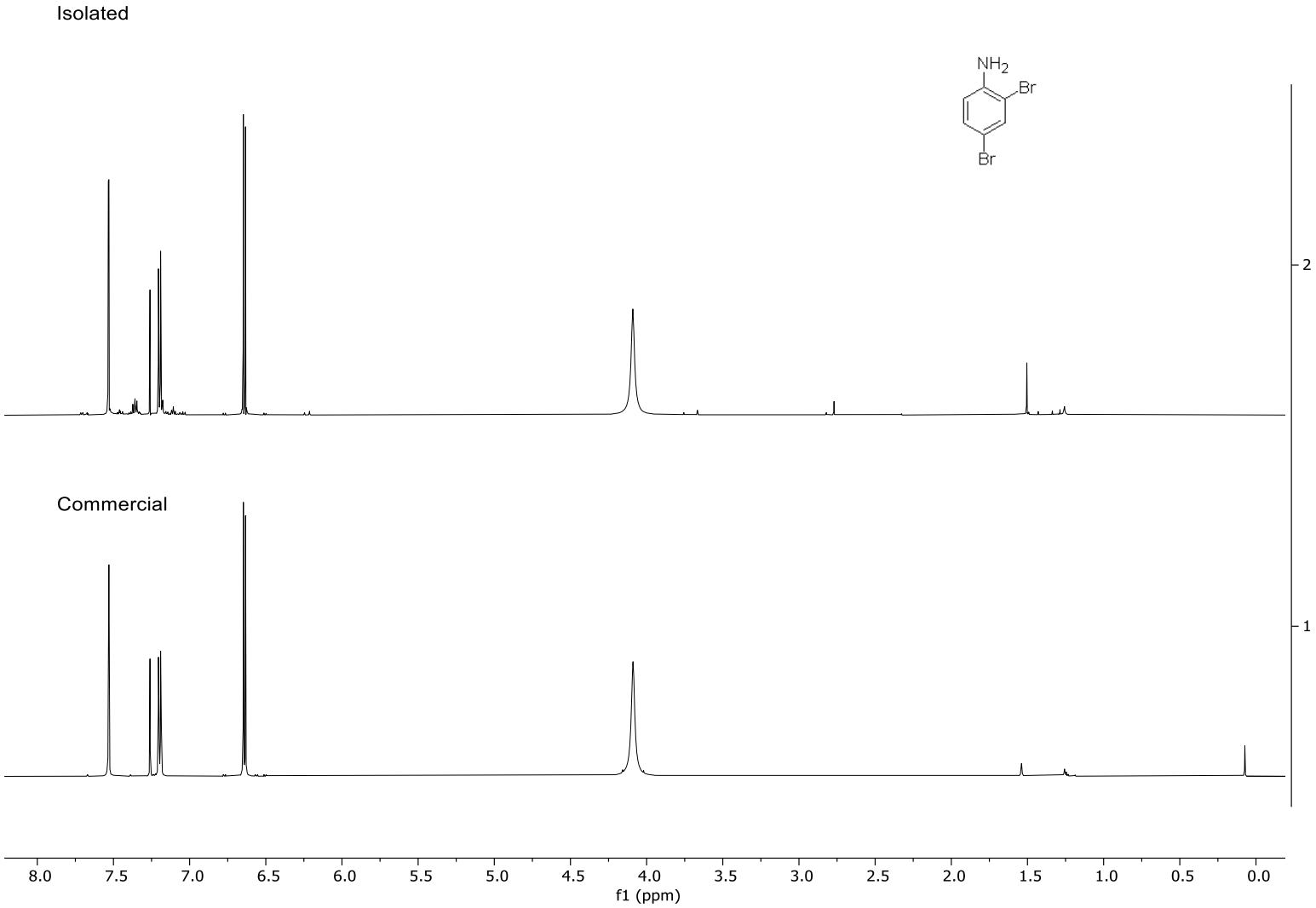
Spectrum S20: ^{13}C-NMR of compound **26**



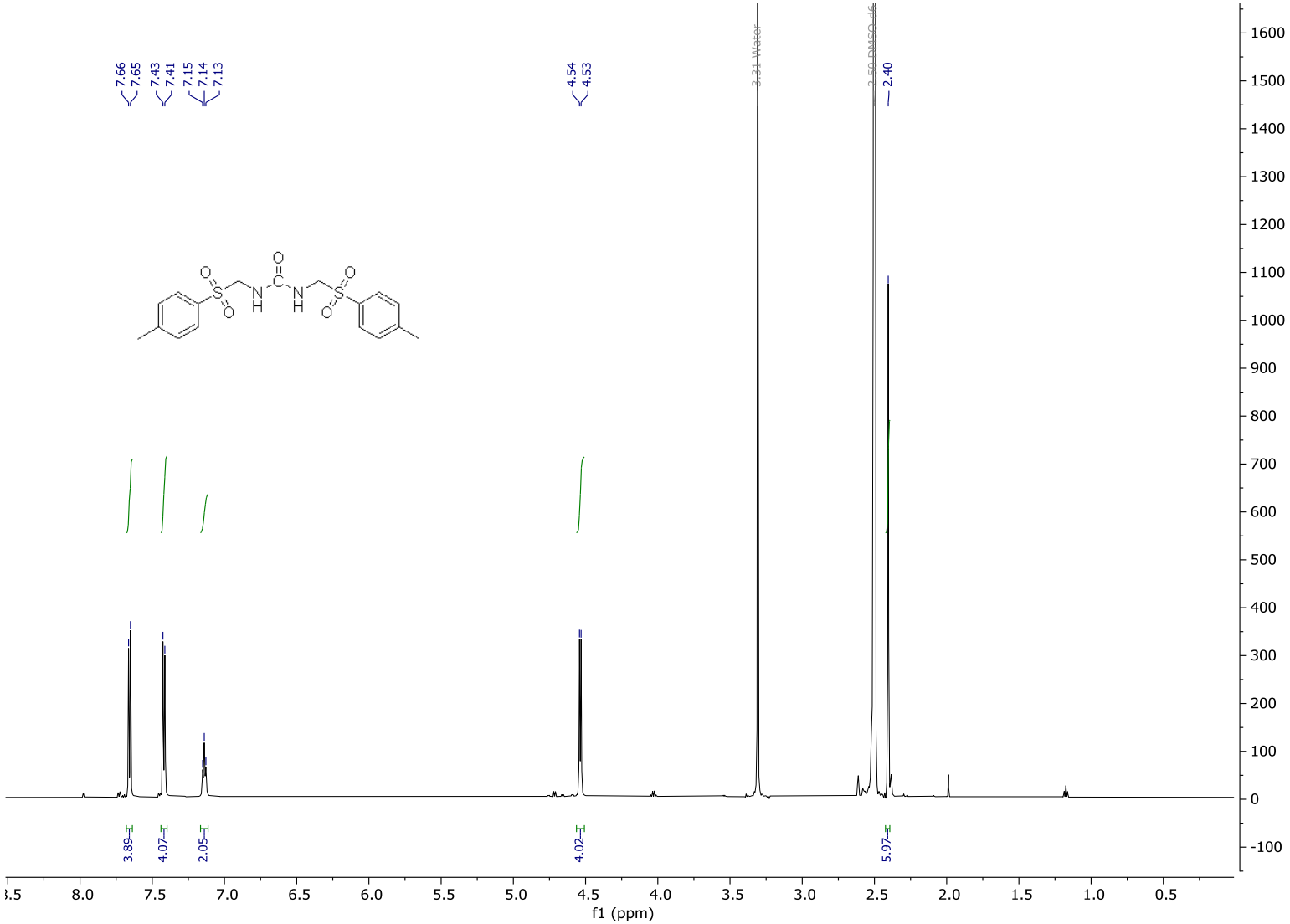
Spectrum S21: ^{1}H-NMR of compound **27**



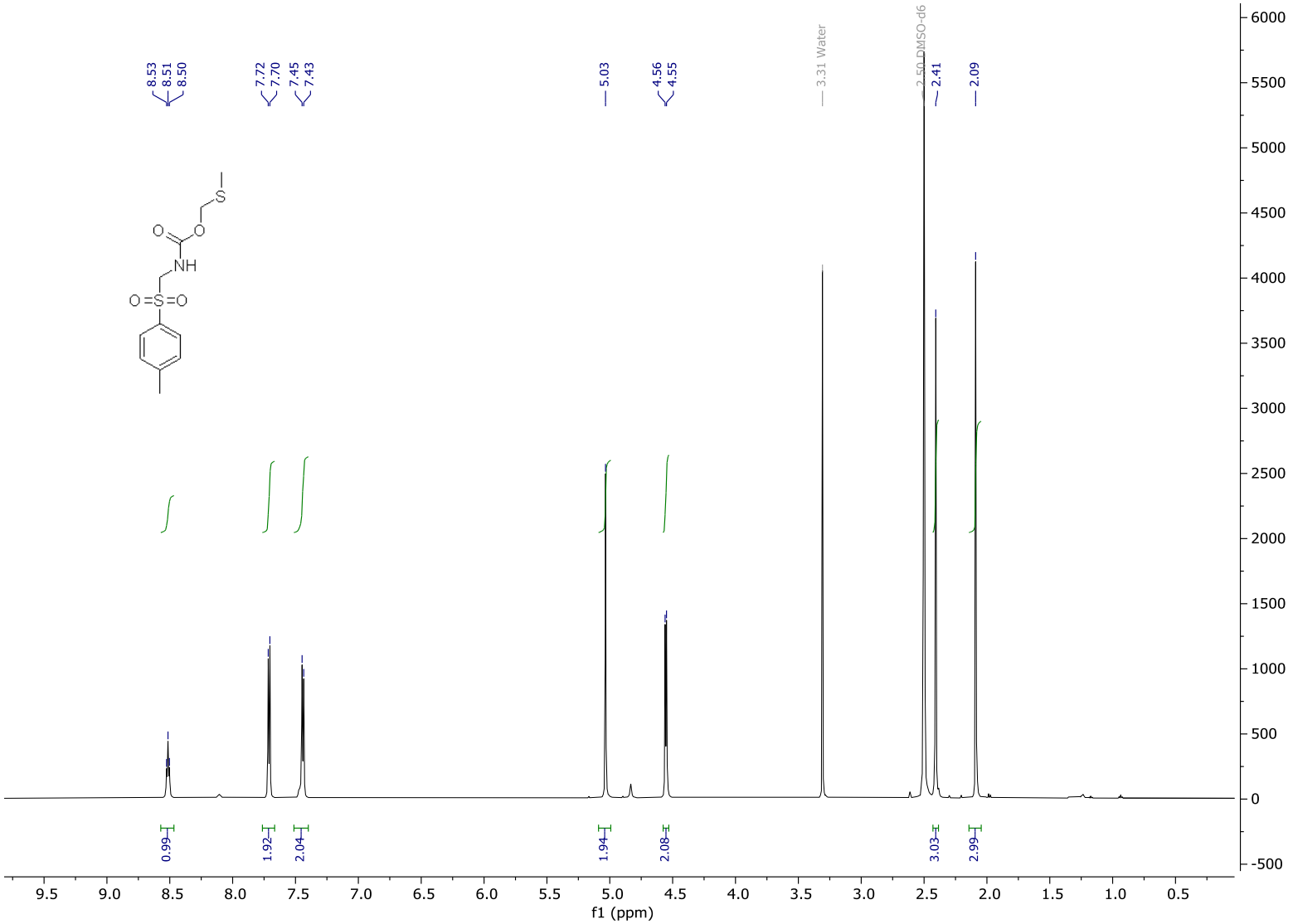
Spectrum S22: ^{13}C-NMR of compound **27**



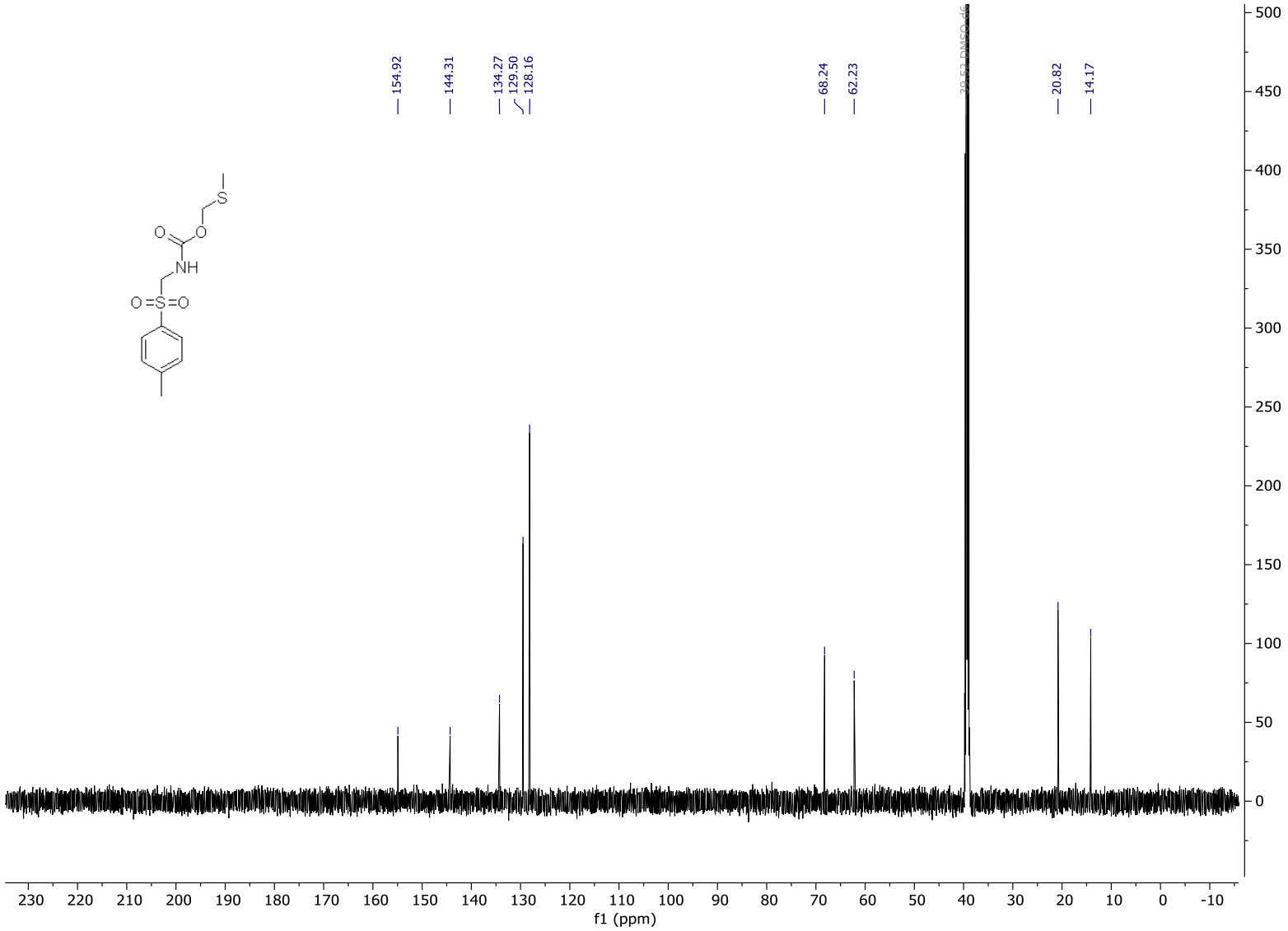
Spectrum S23: ^{1}H-NMR of 2,4-dibromoaniline**,** compared with the same molecule acquired commercially.



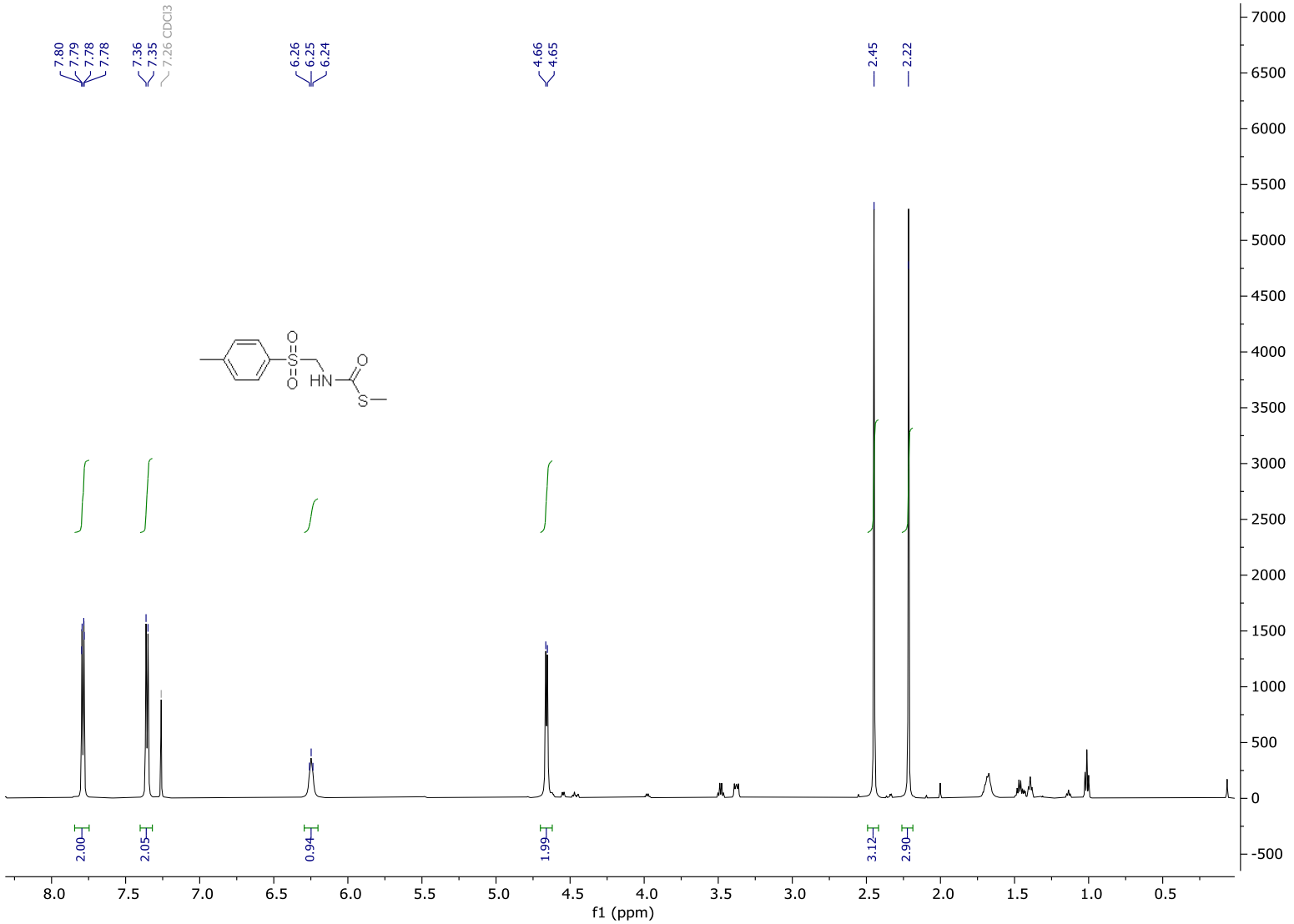
Spectrum S24: ^{1}H-NMR of compound **68**



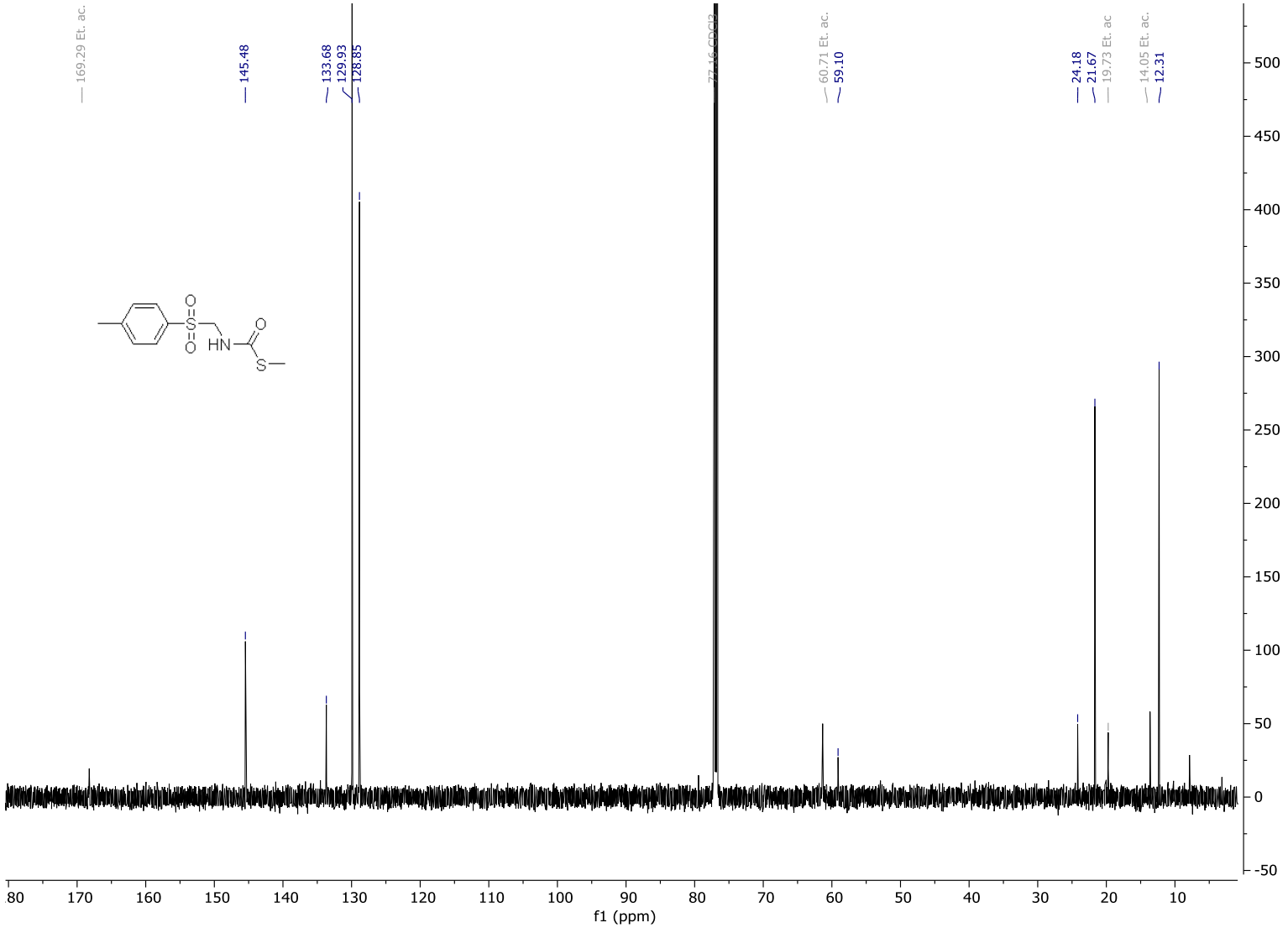
Spectrum S25: ^{1}H-NMR of compound **69**



Spectrum S26: ^{13}C-NMR of compound **69**



Spectrum S27: ^{1}H-NMR of compound **85**



Spectrum S28: ^{13}C-NMR of compound **85**

10 References

1. D.P. Kingma, J. Ba, CoRR, abs/1412.6980, 2014
2. https://github.com/tensorflow/tensorflow (accessed 03/09/21)
3. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *33rd Int. Conf. Mach. Learn. ICML 2016* **3**, 1651–1660 (2016).
4. W. R. Sorenson, Reaction of an Isocyanate and a Carboxylic Acid in Dimethyl Sulfoxide. J. *Org. Chem*. **24**, 978–980 doi: 10.1021/jo01089a024 (1959)
5. Q. Chen, M. T. Huggins, D. A. Lightner, W. Norona, A. F. McDonagh, Synthesis of a 10-oxo-bilirubin: Effects of the oxo group on conformation, transhepatic transport, and glucuronidation. *J. Am. Chem. Soc.* **121**, 9253–9264 doi: 10.1021/ja991814m (1999)
6. B. E. Hoogenboom, O. H. Oldenziel, A. M. Leusen, p-TOLYLSULFONYLMETHYL ISOCYANIDE. *Org. Synth.* **57**, 102 doi: 10.15227/orgsyn.057.0102 (1977).
7. J. J. Cappon, *et al.,* Synthesis of L-histidine specifically labelled with stable isotopes. *Recl. des Trav. Chim. des Pays-Bas* **113**, 318–328 doi: 10.1002/recl.19941130603 (1994).
8. C. Y. Chen, D. F. Bocian, J. S. Lindsey, Synthesis of 24 bacteriochlorin isotopologues, each containing a symmetrical pair of 13C or 15N atoms in the inner core of the macrocycle. *J. Org. Chem.* **79**, 1001–1016 doi: 10.1021/jo402488n (2014).
9. A. Gossauer, K. Suhl, Totalsynthese des Verrucarins E sowie ihre Anwendung zur Herstellung eines ^{13}C-markierten Derivates desselben. *Helvetica Chimica Acta***, 59** (5), 1698-1704 doi: 10.1002/hlca.19760590530 (1976)
10. A. H. Fenselau, J. G. Moffatt, Sulfoxide-Carbodiimide Reactions. III.1Mechanism of the Oxidation Reaction. *J. Am. Chem. Soc.* **88**, 1762-1765 doi: 10.1021/ja00960a033 (1966)
11. H. V. Le, B. Ganem, Trifluoroacetic anhydride-catalyzed oxidation of isonitriles by DMSO: A rapid, convenient synthesis of isocyanates. Org. Lett. 13, 2584–2585 doi: 10.1021/ol200695y (2011)
12. W. Kemp, Nuclear Magnetic Resonance in Chemistry: A Multinuclear Introduction, 75 (Macmillan education ltd, 1986)
13. M. Sattar, V. Rathore, C. D. Prasad, S. Kumar, Transition-metal-free Chemoselective Oxidative C−C Coupling of the sp3 C−H Bond of Oxindoles with Arenes and Addition to Alkene: Synthesis of 3-Aryl Oxindoles, and Benzofuro- and Indoloindoles. *Chem. - An Asian J.* **12**, 734–743 doi:10.1002/asia.201601647 (2017) (supporting information)
14. D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 1–13 doi: 10.1186/s13321-015-0069-3 (2015)
15. H. Jiang, *et al.,* Direct C-H functionalization of enamides and enecarbamates by using visible-light photoredox catalysis. *Chem. - A Eur. J.* **18**, 15158–15166 doi: 10.1002/chem.201201716 (2012)

16. Y. Tomioka, C. Nagahiro, Y. Nomura, H. Maruoka, Synthesis and 1,3-Dipolar Cycloaddition Reactions of N-Aryl-C,C-dimethoxycarbonylnitrones. *J. Heterocyclic Chem.* **40**, 121 doi: 10.1002/chin.200326130 (2003).

17. I. El Hassan, R. Lauricella, B. Tuccio, Formation of β-fluorinated aminoxyl radicals from N-arylketonitrones. *Mendeleev Commun*. **16**, 149–151 (2006).

18. E. Senogles, R. A. Thomas, The kinetics and mechanism of the acid-catalysed hydrolysis of N-vinyl-pyrrolidin-2-one. *J. Chem. Soc. Perkin Trans. **2*** 825–828 doi:10.1039/P29800000825 (1980)

19. G. M. Sheldrick, SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr. Sect. A Found. Crystallogr.* **71**, 3–8 doi: doi: 10.1107/S2053273314026370 (2015).

20. G. M. Sheldrick, Crystal structure refinement with SHELXL. *Acta Crystallogr. Sect. C Struct. Chem.* **71**, 3–8 doi: 10.1107/S2053229614024218 (2015).

21. L. J. Farrugia, WinGX suite for small-molecule single-crystal crystallography. *J. Appl. Crystallogr.* **32**, 837–838 doi: 10.1107/S0021889899006020 (1999).

22. R. H. Blessing, An empirical correction for absorption anisotropy *Acta Crystallogr.* **51**, 33-8 doi: 10.1107/s0108767394005726 (1995)