

A Reactivity First Approach to Autonomous Discovery of New Chemistry

Dario Caramelli, Jarosław M. Granda, S. Hessam M. Mehr, Dario Cambié, Alon B. Henson and Leroy Cronin* School of Chemistry, the University of Glasgow, University Avenue, Glasgow G12 8QQ, UK.

Email: Lee.Cronin@glasgow.ac.uk

Abstract

We present a robotic chemical discovery system capable of learning the generalized notion of reactivity using a neural network model that can abstract the reactivity from the identity of the reagents. The system is controlled using an algorithm that works in conjunction with this learned knowledge, the robot was able to autonomously explore a large number of potential reactions and assess the reactivity of mixtures, including unknown datasets, regardless the identity of the starting materials. The system identified a range of chemical reactions and products, some of which were well-known, some new but predictable from known pathways, but also some unpredictable reactions that yielded new molecules. The search was done within a budget of 15 inputs combined in 1018 reactions, which allowed us not only to discover a new photochemical reaction, but also a new reactivity mode for a well-known reagent (*p*-toluenesulfonylmethyl isocyanide, TosMIC). This involved the reaction of six equivalents of TosMIC in a ‘multi-step, single-substrate’ cascade reaction yielding a trimeric product in high yield (47% unoptimized) with formation of five new C-C bonds involving *sp-sp*² and *sp-sp*³ carbon centres. Analysis reveals that this transformation is intrinsically unpredictable, demonstrating the possibility of reactivity-first robotic discovery of unknown reaction methodologies without requiring human input.

Many discoveries in the chemistry laboratory are the result of chance observations, and it is hard to know ahead of time where a new reaction or molecule will be found. (1) We can explore chemical space mathematically using rule-based generation methods, (2) or by mapping chemical reaction databases, (3) but much of the search is done through traditional approaches using cheminformatics, (4,5) artificial intelligence (6-8) or computation (9,10). Reactivity-first approaches (11-12) have only recently been tentatively explored, but the vast majority of organic synthesis is target-oriented, (13) which means that the discovery of new reactions is a chance event, or results from the need to access a new transformation (14). The development of new transformations and methodologies (15) however is a complex problem requiring a high degree of expert knowledge (16, 17). Furthermore, the current approaches to reaction or method discovery are generally constrained to known heuristics and the discovery of novel reactions is rare. The search for unexpected results can be accelerated with automated systems and in the last decade high-throughput experimentation (18) has shown its potential in speeding up reaction preparation and analysis (typically applied in reaction optimization and combinatorial chemistry) (19-21). However, an increase of reaction throughput does not automatically lead to the serendipitous discovery of entirely new transformations while, on the other hand, the discovery of new reaction pathways from first principle (i.e., *in silico*, based on quantum mechanics) is hard both due to the combinatorial explosion of possible reaction pathways and the computational cost of accurate modelling of the energy hypersurface. To overcome these limitations, an increasing number

of approaches are starting to involve a feedback loop from the on-line analytics and a decision-making algorithm to perform only a fraction of the possible combinations, considered interesting (20). In such a “closed-loop” (21-25) approach, the system automatically explores a chemical space in a trial-and-error fashion mimicking a human experimenter. The system requires three main parts: a chemical robot to perform and analyse the reactions, a program for interpretation of analytical data, and an algorithm that correlates the outcome of the reaction with the input and process parameters. This last part closes the loop by suggesting the predicted optimal parameters for the next reactions (Figure 1).

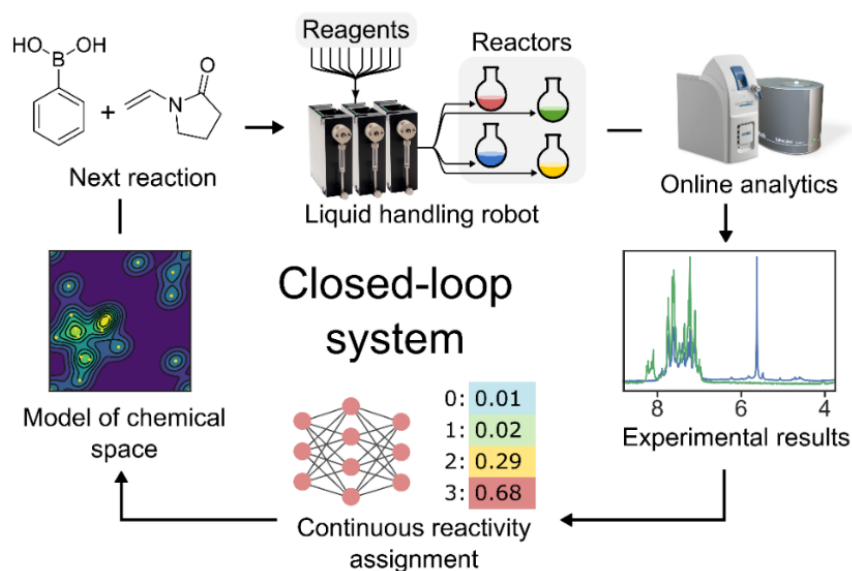


Figure 1: **Closed-loop framework for chemical space exploration.** A liquid handling robot performs an experiment and collects NMR and MS spectra. This data is processed to assess reactivity and create a model of the chemical space that is queried to formulate the next experiment to be performed.

Although closed-loop approaches have proved effective in reaction optimization, their application towards the discovery of new reactions remains underexplored. This is because assessing the reactivity of an unknown reaction with unpredictable products is harder than using metrics meant for optimization of a known target compound, such as yield or selectivity. For autonomous reactivity-first discovery to become feasible, both analytical method as well as reactivity detection algorithm need to be general-purpose and robust. Proton NMR spectroscopy is applicable to a wide range of samples and inexpensive benchtop instruments amenable to online analysis are available. Still, there are no general-purpose automated algorithms to detect reactivity solely relying on the NMR spectra of the starting materials and that of the reaction mixture. Many reactions are known to lead to subtle changes in the spectrum. In contrast, many trivial phenomena, e.g. proton exchange, can manifest as in visually distinct spectra, requiring an expert to discern whether or not a significant has occurred. In response to this challenge, we devised a convolutional neural network called *Reactify* that can automatically assess the reactivity of NMR spectra it has never seen before, having captured the expertise of the human chemist during training. Whilst binary reactive/non-reactive classification of reactions has led to progress (12, 26), we hypothesised that a continuous measure would allow the system to abstract the notion of reactivity, rather than restrict it to a given fixed set of reagents. To exploit this, we designed a system

to explore reactivity of an experimental space using a liquid handling platform to prepare and analyse the reactions, a convolutional neural network for the reactivity assessment and a linear regressor to correlate the starting materials to the reactivity.

The chemical robot

To emulate a human chemist, experiments were performed in conventional round bottom flasks that were automatically cleaned after each reaction by flushing them with clean solvent. Starting materials were stored as 1M stock solutions in dimethyl sulfoxide (DMSO) and the platform used 30 syringe pumps with integrated valves to mix them into six parallel reactors. As optional expansion of the chemical space, three of the reactors were also equipped with visible-light Light Emitting Diodes (LED) to promote photochemical reactions. During the exploration of Chemical space 2 (Figure 3c) the reactions performed in these reactors were prepared adding 2.5% mol of a molecule known to act as photocatalysts, associated to the LED wavelength: 2,4,6-triphenylpyrylium tetrafluoroborate (**PC1**, 405 nm), tris(2,2'-bipyridyl)dichlororuthenium(II) hexahydrate (**PC2**, 450 nm) and rose bengal (**PC3**, 565 nm). After three hours the mixtures were analysed automatically with a benchtop NMR and MS. The software managed the preparation and analysis of the reactions. It was designed to run them in parallel by shifting each experiment starting time to efficiently share the on-line analytics and cleaning cycles. Through this optimized schedule it was possible to perform up to 36 reactions per day, each with a reaction time of three hours giving a total of over 100 reaction hours per day (Figure 2).

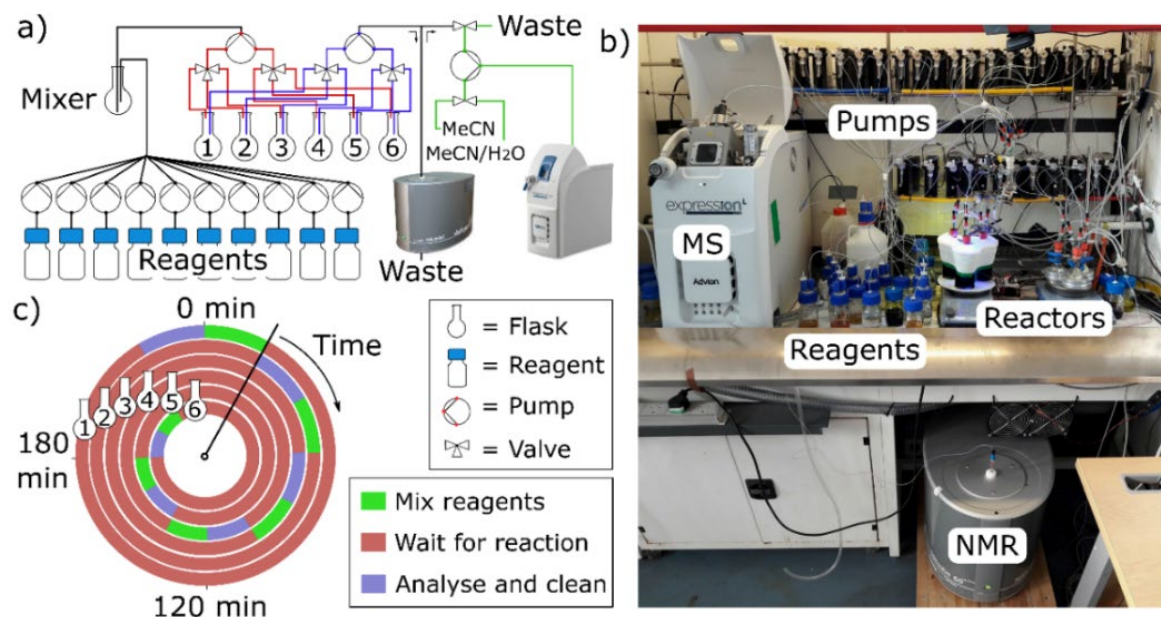


Figure 2: **Liquid handling platform** (a) Schematic of the platform, a series of reagents are added by dedicated pumps to a mixer flask. A pump expanded with two extra valves (obtained by removing the syringe from a normal pump) is used to transfer the reaction mixture in one of the six reactors (1-6) (red); another pump with the same set-up (blue) is used to connect the reactors with the benchtop NMR. Finally, a third expanded pump is used for an in-line dilution prior to injection in the MS (green). (b) Picture of the platform. The pumps are visible on the shelves, on two lines. At the bottom there is the NMR instrument equipped with a flow probe, The MS is on the left, the reactors are in the centre while the reagents, the solvent drum and the waste container are on the left. (c) Six parallel reactions were started with a time-offset to allow the platform to continuously perform physical operations.

The reactivity assignment algorithm

Initially, we explored the chemical space of six simple molecules (Chemical space 1, Figure 3b) mixed in binary and tertiary combinations yielding 440 reactions. To assess the reactivity of the reaction mixtures by ^1H NMR we decided to use a convolutional neural network (CNN, Figure 3a) trained to mimic the assignments made by a human experimenter. As with any other supervised learning algorithm, we needed to provide the true reactivity values associated with inputs in the training set. The reactions were therefore manually scored by an expert organic chemist using four classes of reactivity (0 – non-reactive, 3 – very reactive) to describe the difference between the reaction mixture and the superimposed ^1H -NMR spectra of the starting materials. High values were assigned to mixtures with several new peaks and disappearance of the starting material signals. Experiments showing little or chemically insignificant spectral changes were assigned a low reactivity class (Figure 3e).

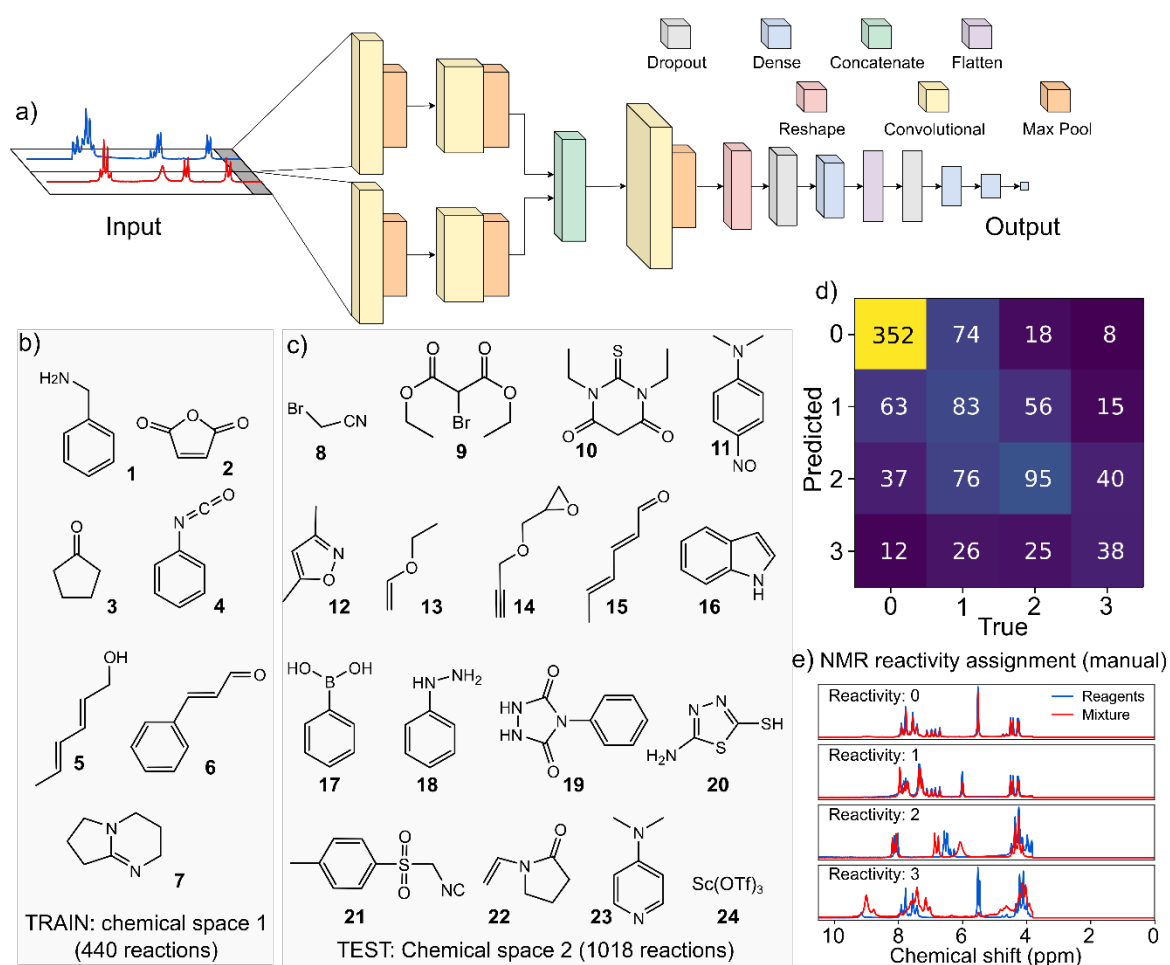


Figure 3. CNN assessment of reactivity in two different chemical spaces (a) Structure of the Neural network used to assign the reactivity to the NMR spectra. Data of the mixture and the sum of starting materials are used as input. The network is trained using 440 reactions from a chemical space (b) and tested on 1018 reactions performed from combinations of 15 different molecules (c). (d) The accuracy on the Test set plotted as a confusion matrix shows that the network successfully learned to generalize the reactivity beyond reagents in the training set. (e) All data are manually classified into four classes.

The reactivity assignment network was designed purposely without any information about the chemical structures of the materials and trained to detect reactive reactions by correlating the raw spectroscopic

data to the values assigned by a chemist. The network was built using a combination of convolutional and dense layers was used, taking as input a pair of NMR signals corresponding to the mixture before and after the reaction, the former estimated by superimposing the spectra of the starting materials. The output was designed to vary continuously between 0 and 1 and binned into 0.25 intervals corresponding to the reactivity classes described above. After training the network on data from 440 initial reactions, its performance was evaluated on 1018 reactions between fifteen different starting materials (chemical space 2, Figure 3c). The results are shown in Figure 3d, where the confusion matrix compares the class assignments made by the neural network (predicted) versus manual assignment (true values). The network predicted the correct reactivity class 56% of the time (the random baseline being 25%). The encouraging results for an unseen set of molecules suggest that the CNN was capable of generalizing a notion of reactivity in NMR spectra independent of the reagents used.

Algorithm for the exploration of the chemical space

To close the loop and drive the exploration of the chemical space we needed a representation of chemical space that could accommodate many different classes of molecules and correlate the presence of various structural motifs with the observed reactivity. To this end, we used the *junction tree variational autoencoder* algorithm (27) to translate the molecular structure of the reagents into fixed-length fingerprint vectors. The autoencoder combines of a tree-structured scaffold generated over chemical substructures with a graph message passing network. The resulting 56-dimensional vectors are then used as input to the reactivity estimation neural network we developed (Figure 4a). The idea behind the algorithm is to train the model on a small fraction of the chemical space, explored at random, and use the knowledge acquired to predict the reactivity of the remaining possible combinations. The reactant combinations are then sorted by predicted reactivity and the best candidates are reacted in the platform. After each reaction the model is re-trained using the newly obtained reactivity information. By guiding the robot with such a reactivity-driven algorithm it will be possible to perform the reactive combinations first, meaning that only a fraction of the chemical space will need to be explored. Furthermore, thanks to the structure encoding the network structure is independent from the identity of the molecules involved. It is therefore possible to update and use the model on any organic reaction involving three reagents, meaning that this method is easily scalable to vast chemical spaces with a large number of starting materials. We imagine that by training the model on bigger datasets the scope of the predictions will also expand. After the first run of 1018 reactions, the dataset was used to simulate further explorations in order to test the ability of the system to navigate the chemical space.

Initially, we tested the model on the full dataset of 1018 combinations, obtaining a mean squared error in the prediction of 0.17 for the test set (train/test ratio: 0.8, Figure 4c). In a different experiment aimed at simulating exploration, we trained the model on a random batch comprising 10% of the dataset (ca. 100 reactions) and predicted the reactivity values of the remaining ca. 900 combinations. We then simulated performing the most reactive combinations by revealing the outcome of 50 combinations predicted to have the highest reactivity. The model was retrained on the expanded dataset of 150 reactions and the process was repeated until all the 1018 reactions were explored (Figure 4d). The

results of the simulation are shown in Figure 4d. The initial random batch of experiments had an average reactivity of 0.40 ± 0.29 . Following training, the first generation of 50 reactions suggested by the model gave an average reactivity of 0.60 ± 0.23 . Over time the algorithm is trained on more data but at the same time the reactive combinations are taken out of the dataset, leaving only the unreactive ones, as evidenced by the decline in reactivity of subsequent generations (Figure 4e).

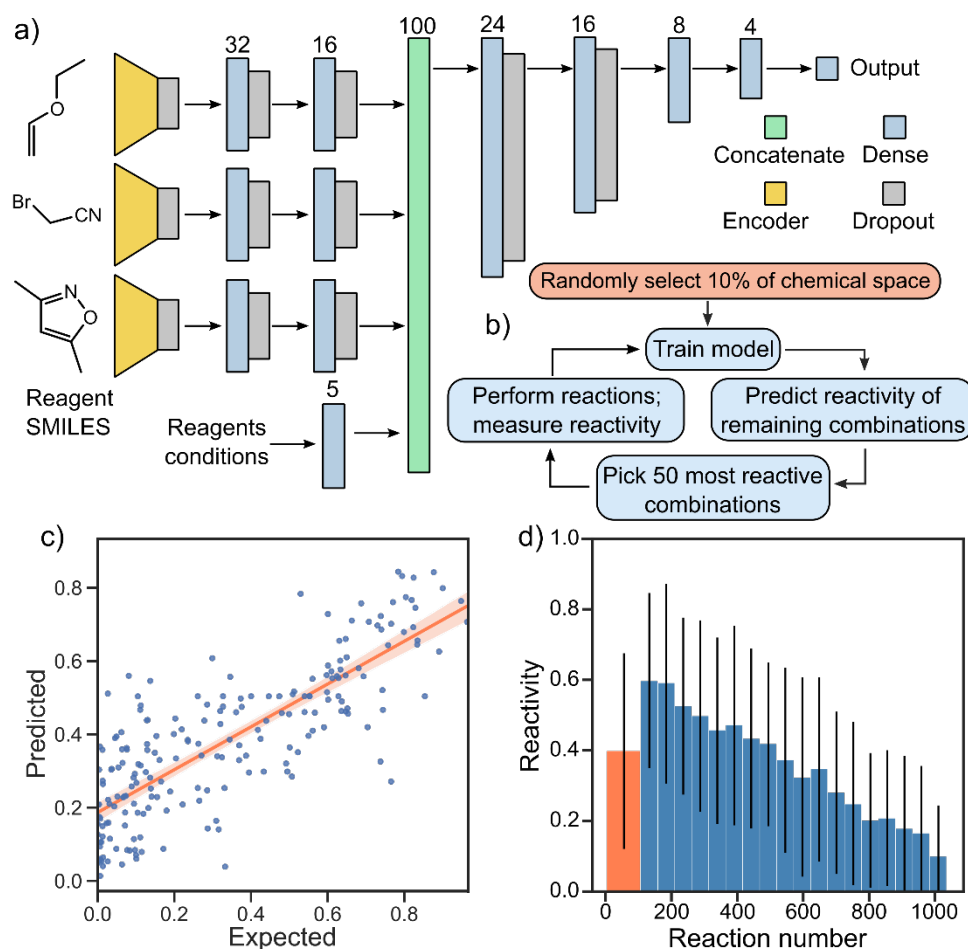


Figure 4. **Aspects of autonomous chemical space exploration algorithm.** (a) Structure of the neural network used for reactivity prediction. A junction-tree neural network encodes the molecular structure of each reactant into a 56 dimensional vector. (b) Scheme of the algorithm used to simulate the chemical space exploration. (c) Correlation of predicted versus observed reactivity for the test set. The correlation is demonstrated by fitting a linear regression model with the shaded area representing the 99% confidence interval obtained using bootstrap. (d) Results of the chemical space exploration simulation. After the initial selection of 100 random reactions (orange), the algorithms start to create a model that correlates parameters to reactivity. By prioritizing combinations that are predicted to be reactive, the space is explored in a more efficient way. The error bars show standard deviation.

Discovery of a multi-step-one-substrate reaction cascade

Based on the data obtained from the benchtop instruments, a small selection of five combinations showing high reactivity were repeated in the platform and the products manually isolated (Figure 5). Two of these are known molecules already reported in literature (28-30); one is a new product for which the generic procedure is known (31); and the last two are 'novel' as they lead to unknown molecules

through unknown procedures, making them genuine discoveries. Reaction (a) is a C-H functionalization made through photoredox catalysis. It has been already described in 2012 (28) where the authors used the same reagents, $[\text{Ir}(\text{ppy})_2(\text{dtbbpy})]\text{PF}_6$ as photocatalyst, two equivalents of Na_2HPO_4 and acetonitrile as solvent. Reaction (b) has been reported as method for synthesis of N-aryl-C,C-dimethoxycarbonylnitrones (29) and originally involved sodium hydroxide and THF as solvent.

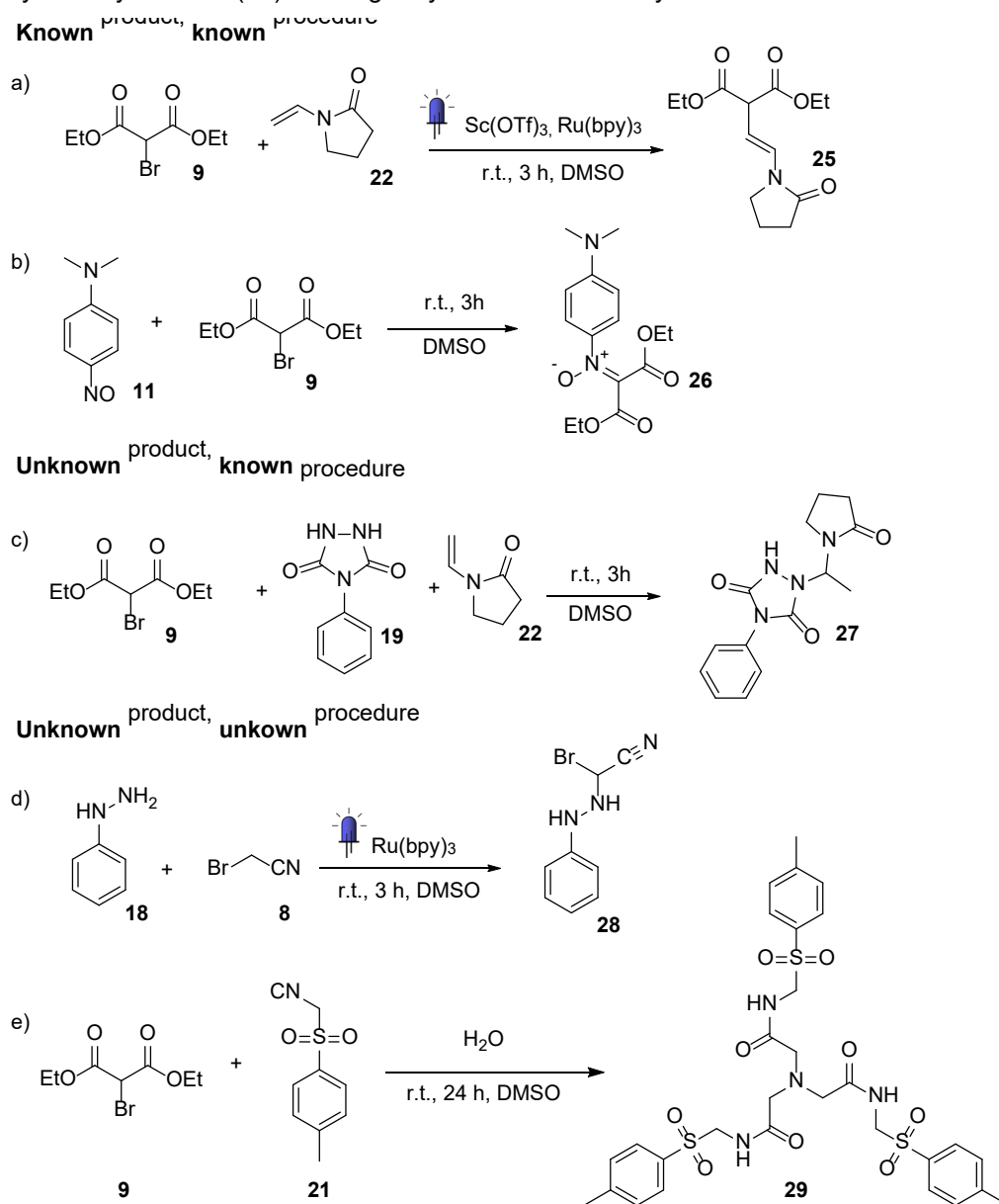


Figure 5: Five reactions showing high reactivity have been found and characterised. Reactions (a) and (b) have been previously reported in the literature (27-29) with the exact same product. Reaction (c) is known in literature but has never been used to make **27** (30). Reactions (d) and (e) are unreported in literature.

Formation of molecule **86** following reaction (b) is known (29). Reaction (c) is a hydroamidation (31) promoted by diethyl Br-malonate and the product is unreported in literature. Reaction (d) is a photochemical reaction involving the addition of phenylhydrazine and bromoacetonitrile in presence of

tris(2,2'-bipyridyl)dichlororuthenium(II) hexahydrate and 450 nm irradiation. In this reaction a new C-N bond is formed while the bromide, usually a leaving group, is kept in its place. It is unreported in literature. Reaction (e) was discovered during the analysis of the mixture of *p*-toluenesulfonylmethyl isocyanide (TosMIC) and diethyl bromomalonate. X-ray analysis of the isolated product confirmed that the trimeric product **29** was formed. The molecular structure of the product showed an unusual increase on complexity and a non-trivial mechanism of formation, especially taking in consideration the three central methylene carbons. Interestingly, the XRD showed a tubular supramolecular assembly composed of six molecules packed as a ring and multiple rings stacked together leaving void space with an average diameter of 12 Å (Figure. 6a) (details in supplementary information SI-6).

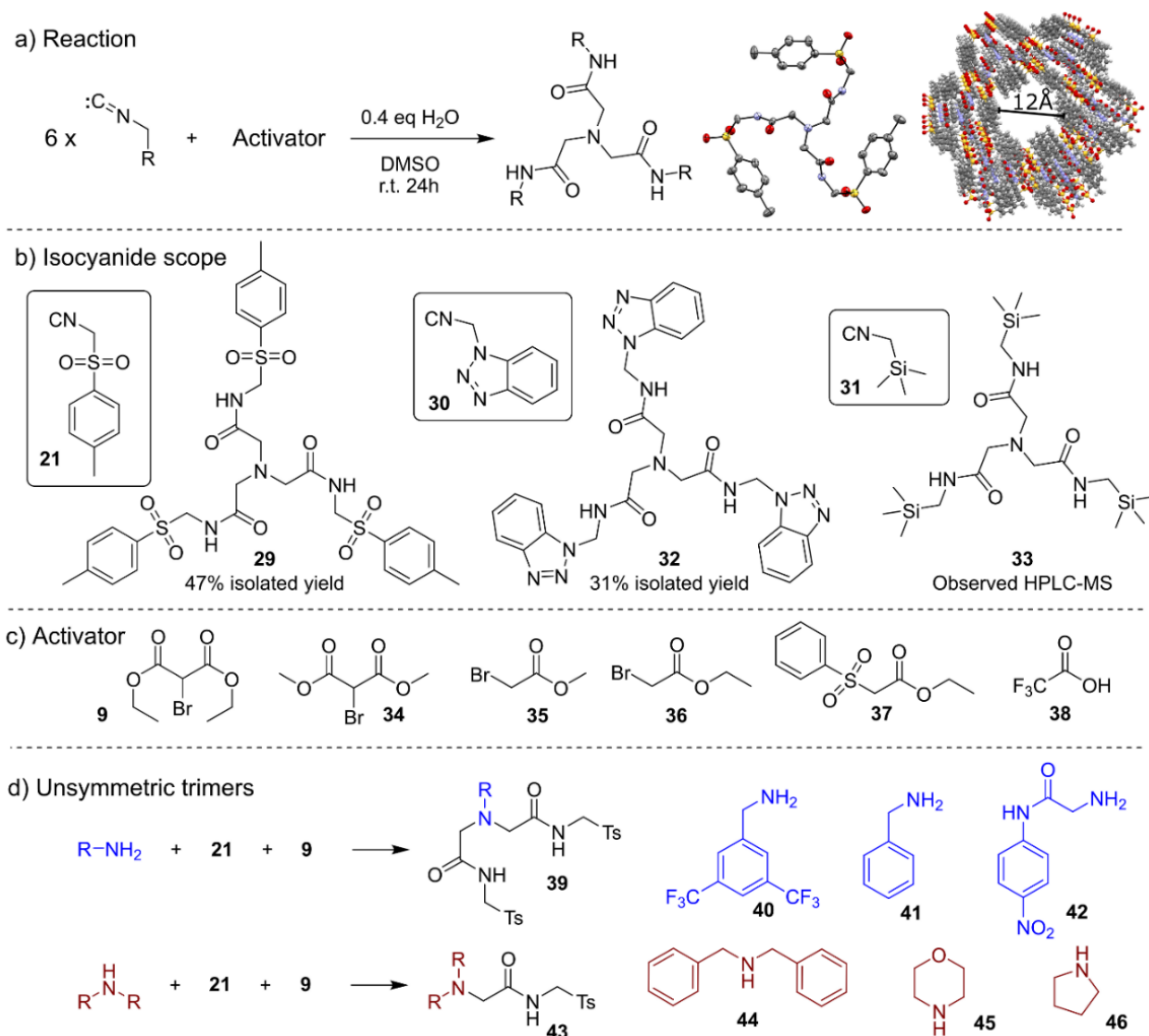


Figure. 6. **The reaction of diethyl bromomalonate and TosMIC discovered with the automated platform.** (a) General scheme of the reaction. Six equivalents of isocyanide are consumed in presence of an activator, water and DMSO **29**, on the right-hand side the X-ray structure of **29** and its tube-shaped supramolecular structure. (b) Analogous products obtained with variations of the isocyanide. (c) The reaction has been carried out using variations of diethyl bromomalonate, yielding the same product. (d) By performing the reaction in the presence of an amine we observed variations of the product, suggesting the mechanism reported in the next figure.

To explore this transformation further we decided to perform the reaction with a range of isocyanides to elucidate a possible mechanism. From the seven isocyanides (SI-3.6) tested, a similar product was isolated in the reaction with the 1*H*-benzotriazol-1-ylmethyl isocyanide **30** while traces of the reaction with (trimethylsilyl)methyl isocyanide **31** were detected by LC-MS (Figure. 6b). Variations of diethyl bromomalonate have also been explored finding working alternatives in several similar molecules including trifluoroacetic acid (TFA, **38**) (Figure. 6c). All these variations yielded the same product, suggesting that the second reagent is not directly involved in the product formation but rather acts as some kind of a promoter; this is because the reaction does not give the product in any detectable amount in the absence of it. In order to confirm the involvement of the hypothesised intermediates presented in the mechanism below the reaction has also been carried out in the presence of various amines (Figure. 6d) and the presence of the relative products **39** and **43** has been established by LC-MS in all cases. These correspond to an asymmetric version of the product **29**, where one or two branches have been replaced with the amine R-group. The possibility of tuning the branches in this way gave us precious information about the mechanism and increased the flexibility and the possible applications of this reaction.

To understand the formation of the core of the molecule we prepared two isotopically enriched versions of the TosMIC substrate, labelling the isocyanide carbon and the methylene carbon. To confirm the source of the three oxygen atoms found in the product we performed the reaction in both synthesised ¹⁸O-DMSO and anhydrous DMSO with small amounts of H₂¹⁸O. These experiments revealed that all the three central methylene carbons come from the CH₂ carbons of the TosMIC, while the oxygen atoms come from DMSO (Figure 7b). This means that the product is obtained using at least six equivalents of isocyanide and that DMSO is also taking part in the reaction, in fact the reaction has been repeated in DMF and MeCN without success (SI-4.4). While testing the reaction in different conditions we also noticed its poor reproducibility, with significant yield fluctuations even under apparently identical conditions. Upon further investigation, it was found that the amount of trace water present in the solvent has a marked effect on the reaction profile. By testing different amounts of water, we found that the reaction does not yield any product under strictly anhydrous conditions and in presence of more than 2 equivalents of water. The best yields were obtained with 0.4 equivalents (SI-4.3). The reaction kinetics have been investigated with online HPLC, showing the formation of an intermediate after 2 hours that eventually disappears after 34 hours with the simultaneous formation of product **29** (SI-4.2). MS analysis of the corresponding peak showed a mass consistent with compound **55**: the two-branches imine analogous to the product **29**. The chromatogram also showed the presence of high amounts of the single-branched amine **52**. This was in accordance with the data reported in Figure 7d and supported the hypothesis of a mechanism involving the formation of a central amine group that undergoes two identical semi-reactions to build the other two branches. Given this information we propose the mechanism reported in Figure 7a. The role of diethyl bromomalonate (and the other activators) is to promote the oxidation of the isocyanide group to isocyanate (32). The formation of the central methylene carbons can be explained with the formation of imine **51** by oxidation of the amine

50 (32). **51** is then reduced to form the single branched amine **52** and attacked by it to form **53**, which undergoes an elimination of ammonia followed by a reduction to produce **55** (34).

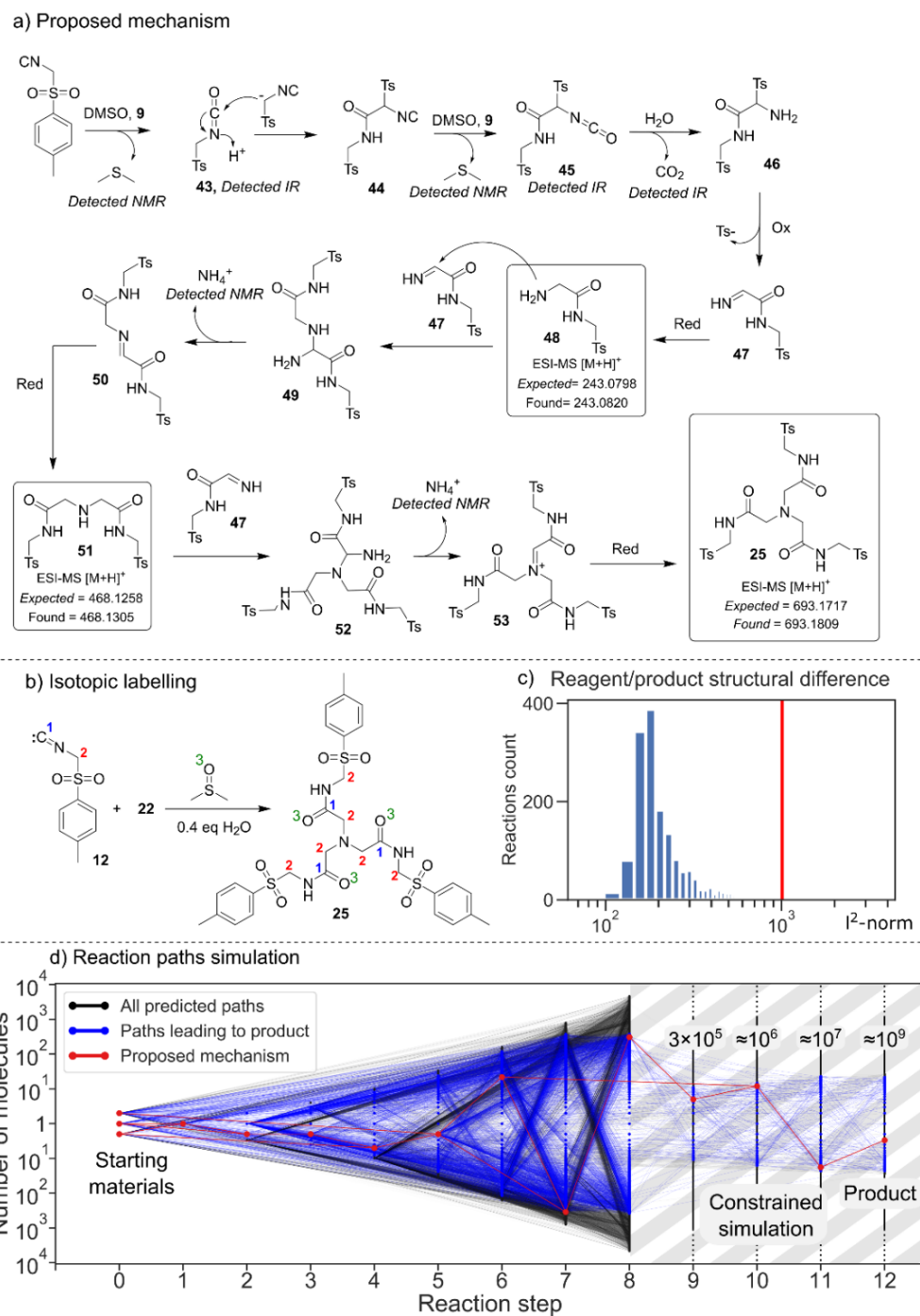


Figure 7: **Reaction mechanism and cheminformatic analysis.** (a) Scheme of the proposed mechanism. Two of the intermediates have been found by HPLC-MS analysis while the isocyanate group, CO₂, DMS and ammonia have been detected at IR and NMR, respectively. (b) Isotopic labelling of carbon atoms C¹ and C² of TosMIC and the DMSO oxygen helped determine the source of core atoms in the product molecule. (c) Comparison of structural change between reagents and products among known reactions of TosMIC (red line indicates discovered reaction) (d) Estimation of reaction unpredictability by estimating the size of the relevant reaction network. A full simulation could only be carried out for the first eight steps of the simulation as the combinatorial explosion produces a vastly greater number of molecules than could be feasibly analyzed.

The mechanism is then repeated with the consumption of another equivalent of **51** to yield the final product **29**. In addition, this mechanistic hypothesis is supported by an online IR experiment showing the presence of an isocyanate group (**47** and **49**) and CO₂ as well as the observation of dimethyl sulphide and ammonium signals in the NMR analysis of the mixture (SI-4.6-4.7). To compare this new reaction with known transformations in the literature, we compiled a set of 1656 known reactions found in the *Reaxys* database (35) involving TosMIC. The RDKit (36) python library was used to extract a fingerprint difference between starting materials and products, indicating the extent to which each reaction transforms the structure of its input reagents.

To compare fingerprint differences among reactions, the ℓ^2 -norm of the fingerprint-difference vectors was calculated. Low values of the ℓ^2 -norm indicate a high structural similarity between reagents and products while high values correspond to complex reactions involving several transformations. The results are shown in Figure 7c, where the ℓ^2 -norm values of literature reactions are grouped into bins on a logarithmic x-axis. The ℓ^2 -norm of the reaction discovered is indicated with the red line, signalling an unusual degree of structural change compared to known reactions. To gauge the serendipitous nature of our discovery quantitatively, we carried out a simulation to assess the size of the chemical reaction network. This network was generated by repeatedly applying a set of common reaction templates, including the ones invoked in our proposed mechanism, to generate an expanding pool of chemicals. Comparing the total size of the resulting network to the subset leading to the product gives an indication of the unpredictability for the particular product obtained. The results are shown in Figure 7d, where blue lines indicate the reaction network relevant to product formation. The overall network is vastly larger in size than the subset potentially leading to product (10^{10} chemicals vs. 10^5 chemicals), indicating that the observed pathway is highly improbable to predict *a priori*.

In summary, we showed that closed-loop approaches combining automatic reaction execution and reactivity assessment using machine learning can play a crucial role in the discovery of novel reactions in unexplored parts of chemical space. Our neural network model can abstract the reactivity from the identity of the reagents, and we expect that this type of algorithm will also progressively improve in accuracy when presented with reactivity data for subsequent chemical spaces. The continuous reactivities provided by the CNN are correlated with reagent structural features, showing that it is possible to explore chemical space intelligently and discover unpredictable reactions, thanks to the unbiased nature of our system. Our results demonstrate the possibility of NMR-driven universal reactivity detection as a key enabler of autonomous discovery in a closed loop. Within this framework, we also show the potential of reactivity-first chemical space search and its suitability to the discovery of novel molecules and mechanisms.

Materials and Methods

General experimental remarks. Chemicals and solvents were supplied by *Fisher Chemicals*, *Sigma Aldrich*, *Lancaster Chemicals Ltd* and *Tokyo chemicals industry*, used as received. Deuterated solvents

were obtained from *Goss Scientific Instruments Ltd.* and *Cambridge Isotope Laboratories Inc.* All commercial starting materials were used as supplied, without further purification. Off-line NMR data was recorded on a Bruker Advance 600 MHz or a Bruker Advance 400 MHz, in deuterated solvent, at T = 298 K, using TMS as the scale reference. Chemical shifts are reported using the δ -scale, referenced to the residual solvent protons in the deuterated solvent for ^1H and ^{13}C NMR (i. e. ^1H : δ (CDCl_3) = 7.26; ^{13}C : δ (CDCl_3) = 77.16). All chemical shifts are given in ppm and all coupling constants (J) are given in Hz (J) as absolute values. Characterization of spin multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = double doublet, dt = double triplet, dq = double quartet, and ddt = double doublet of triplets. 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. The MS apparatus was a Bruker MaXis Impact instrument, acquisition range at 50–2000 m/z .

Liquid handling platform. The control over the fluids was performed using C3000 model, TriContinent™ pumps (Tricontinent Ltd, CA, USA). They were equipped with distribution (3-way) and 90°/120° (2-way) valves. 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. The liquid connectivity was assured using PTFE tubing (1/16" 1.6mm OD x 0.8mm ID) cut to the desired length and PEEK/PTFE flangeless fittings. To perform a reaction the robot mixed 2 ml of the selected starting materials (from 1 M stock solutions) into the mixer flask and then moved the mixture into one of the six round bottom flasks (25 ml). They were placed on top of two hotplates for magnetic stirring and three of them were irradiated with visible-light LED (Thorlabs, USA). After three hours the mixture was analysed and the flask was washed with 5 ml of DMSO for three times. The software to control the platform was written in Python and was optimised to continuously run six reactions in parallel.

Benchtop NMR spectroscopy. The online NMR spectra were recorded using a Spinsolve benchtop NMR from Magritek (60 MHz). Shimming was performed before each experiment directly on the sample. The instrument was equipped with a flow-cell to allow online analysis. The cell was designed to go through the instrument and its location placed the thicker part (5 mm diameter) at the centre of the magnets. Both inlet and outlet were connected to normal PTFE tubing with screw caps (Figure S1). The flow cell allowed automatic reaction monitoring in real time by pumping 3 ml of solution from the reaction mixture. The instrument was controlled with Python through a TCP connection with the API exposed by Spinsolve software.

Benchtop MS spectroscopy. The spectra were recorded using an Advion Expression CMS equipped with an ESI (electrospray ionization) module. The mass spectrometer was controlled using a Python library created to wrap around the binary libraries supplied by Advion. Before injection the mixture was diluted by taking 0.1 ml of reaction mixture into the syringe and adding with 0.4 ml of acetonitrile. 0.4 ml of the diluted solution was pumped into waste and the process repeated five times to obtain a 10^{-4}M

solution. After each injection the instrument was cleaned by flushing it with acetonitrile and a water/acetonitrile 1:1 solution.

Automatic reactivity assignment. NMR data was checked manually and a reactivity value was assigned between 0 and 3. The mixture spectrum was compared with the superimposition of the starting materials spectra and the criteria for the assignment are 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. Before the training of the neural network the 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. A 2x271 matrix obtained by the processed spectra of the mixture and the superimposition of the starting materials was used as input for the neural network. Details about the neural network architecture can be found in the SI-2.3. The network was trained on 440 reactions obtained from combinations of Chemical space 1 (Figure 3b) and used to assess its accuracy on 1018 reactions from Chemical space 2 (figure 3c).

Reactivity predictions and automatic space exploration. The reactivity data assigned by the neural network was correlated with the starting materials represented as 56 dimensional fingerprint vectors. The fingerprints were calculated from the reagents SMILES using the encoding part of the junction tree autoencoder developed by Jin et al. (27). The other reaction conditions (presence of acid, base and photocatalyst) were fed into the model as a one-hot encoded vector. The software for training and testing the neural network was written using the Tensorflow library for Python. The network was used to run a simulation of the chemical space exploration where data from the full dataset was progressively (in batches of 50 reactions) accessed following the reactivity predictions generated by the linear regressor. The simulation was implemented in Python.

General procedure for synthesis of products 25, 26, 27, 28, 29. Diethyl 2-bromomalonate (2 mmol, 0.41 ml), *p*-toluenesulfonylmethyl isocyanide (2 mmol, 0.39 g) and water (0.8 mmol, 15 μ l) are mixed in 4 ml of anhydrous DMSO and stirred for 24 hours at 30° C. The reaction mixture is diluted with water (20:1) and extracted with ethyl acetate. The organic phase is separated and washed with brine. Mg₂SO₄ is then added to the reaction mixture and after filtration the solvent is removed under reduced pressure. Products **27** and **29** precipitated as white solids during the evaporation of ethyl acetate, they are isolated by filtration and washed with ethyl acetate. Products **25**, **26** and **28** were purified with chromatographic column.

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/Reactify> and the code is provided under the GPLv3 license.

Supplementary information is available in the online version of the paper.

Acknowledgments. The authors gratefully acknowledge financial support from the EPSRC (Grant Nos EP/S030603/1, EP/S019472/1, EP/S017046/1, EP/L015668/1, EP/L023652/1), the ERC (project

670467 SMART-POM). JMG acknowledge financial support from the Polish Ministry of Science and Higher Education grant no. 1295/MOB/IV/2015/0. We would also like to thank Prof. Bartosz Grzybowski and his team for comments on the reaction mechanism involving TOSMIC.

Author Contributions

LC conceived the concept, the abstraction, algorithm, and the project and coordinated the efforts of the research team. DCaramelli built and coded the platform with help from AH, developed the algorithms for data analysis and manually characterised the reactions with help from JMG and gathered data for the proposed mechanism with help from DCambie. HM and JMG did the cheminformatics analysis. LC, DCaramelli, JMG, and HM wrote the manuscript with input from all the authors.

Keywords

Chemical Automation, Reaction Discovery, NMR, Organic Synthesis, Artificial Intelligence

ORCID

Dario Caramelli (0000-0002-3567-7068)
Jarosław M. Granda (0000-0002-5058-7669)
Dario Cambié (0000-0003-2722-5106)
S. Hessam M. Mehr (0000-0001-7710-3102)
Alon B. Henson (0000-0001-6028-439X)
Leroy Cronin (0000-0001-8035-5757)

References

1. B. Grzybowski, A., K. J. M. Bishop, B. Kowalczyk, C.E. Wilmer, The 'wired' universe of organic chemistry. *Nat. Chem.* **1**, 31–36 doi: 10.1038/nchem.136 (2009)
2. R. Oeschger. *et al.*, Diverse functionalization of strong alkyl C – H bonds by undirected borylation, *Science*, **3**, 736–741 doi: 10.1126/science.aba6146 (2020).
3. J. L. Reymond, L. Ruddigkeit, L. Blum, R. van Deursen, The enumeration of chemical space. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 717–733 doi: 10.1002/wcms.1104 (2012).
4. R. Herges, Reaction planning: prediction of new organic reactions. *J. Chem. Inf. Comput. Sci.* **30**, 377–383 doi: 10.1021/ci00068a006 (1990)
5. Y. V. Suleimanov, W. H. Green, Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.* **11**, 4248–4259 doi: 10.1021/acs.jctc.5b00407 (2015)
6. M. H. S. Segler, T. Kogej, C. Tyrchan, M. P. Waller, Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **4**, 120–131 doi: 10.1021/acscentsci.7b00512 (2018)

7. I. I. Baskin, T. I. Madzhidov, I. S. Antipin, A. A. Varnek, Artificial intelligence in synthetic chemistry: achievements and prospects. *Russ. Chem. Rev.* **86**, 1127–1156 doi: 10.1070/rcr4746 (2017)
8. C. W. Coley, W. H. Green, K. F. Jensen, Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **51**, 1281–1289 doi: 10.1021/acs.accounts.8b00087 (2018)
9. T. Katsila, G. A. Spyroulias, G. P. Patrinos, M. T. Matsoukas, Computational approaches in target identification and drug discovery. *Comput. Struct. Biotechnol. J.* **14**, 177–184 doi: 10.1016/j.csbj.2016.04.004 (2016).
10. D. P. Tabor, *et al.*, Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 doi: 10.1038/s41578-018-0005-z (2018).
11. P. S. Gromski, A. B. Henson, J. M. Granda, L. Cronin, How to explore chemical space using algorithms and automation, *Nat. Rev. Chem.*, **3**, 119–128, doi: 10.1038/s41570-018-0066-y (2019)
12. J. M. Granda, L. Donina, V. Dragone, D. L. Long, L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, *Nature*, **559** (7714), 377–381, doi: 10.1038/s41586-018-0307-8 (2018)
13. M. D. Burke, G. Lalic, Teaching target-oriented and diversity-oriented organic synthesis at Harvard University. *Chem. Biol.* **9**, 535–541 doi: 10.1016/S1074-5521(02)00143-6 (2002).
14. A. McNally, C. K. Prier, D. W. C. Macmillan, Discovery of an α -Amino C–H Arylation Reaction Using the Strategy of Accelerated Serendipity, *Science*, **334**, 6059, 1114, doi: 10.1126/science.1213920 (2011)
15. C. Houben, A. A. Lapkin, Automatic discovery and optimization of chemical processes. *Curr. Opin. Chem. Eng.*, **9**, 1, doi: 10.1016/j.coche.2015.07.001 (2015)
16. E. P. Gajewska, *et al.*, Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses. *Chem* **6**, 280–293 doi: 10.1016/j.chempr.2019.11.016 (2020).
17. P. Schwaller, *et al.*, Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **11**, 3316–3325 doi: 10.1039/c9sc05704h (2020).
18. S. M. Mennen, *et al.*, The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **23**, 1213–1242 doi: 10.1021/acs.oprd.9b00140 (2019).

19. K. Poschary, *et al.*, Machine assisted reaction optimization: A self-optimizing reactor system for continuous-flow photochemical reactions. *Tetrahedron* **74**, 3171–3175 doi: 10.1016/j.tet.2018.04.019 (2018)
20. F. Häse, L. M. Roch, C. Kreisbeck, A. Aspuru-Guzik, Phoenix: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **4**, 1134–1145 doi: 10.1021/acscentsci.8b00307 (2018).
21. J. Montgomery, High-Throughput Discovery of New Chemical Reactions, *Science*, **333**, 1387–1389, doi: 10.1126/science.1210735 (2011)
22. D. Perera, *et al.*, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science*. **359**, 429–434 doi: 10.1126/science.aap9112 (2018)
23. C. J. Richmond, *et al.*, A flow-system array for the discovery and scale up of inorganic clusters. *Nat. Chem.* **4**, 1037–1043 doi: 10.1038/nchem.1489 (2012)
24. A. M. Schweidtmann, *et al.*, Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277–282, doi: 10.1016/j.cej.2018.07.031 (2018)
25. E. V. Rebrov, A. J. Expósito, OpenFlowChem – a platform for quick, robust and flexible automation and self-optimisation of flow chemistry, *React. Chem. Eng.* **3**, 769–780, doi: 10.1039/c8re00046h (2018)
26. V. Dragone, V. Sans, A. B. Henson, J. M. Granda, L. Cronin, An autonomous organic reaction search engine for chemical reactivity, *Nat. Commun.*, **8**, 15733, doi: 10.1038/ncomms15733 (2017)
27. <https://arxiv.org/pdf/1802.04364.pdf> (accessed on March 30, 2021)
28. 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).
29. 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).
30. I. El Hassan, R. Lauricella, B. Tuccio, Formation of β -fluorinated aminoxyl radicals from N-arylnitrones. *Mendeleev Commun.* **16**, 149–151 (2006) doi: 10.1070/MC2006v016n03ABEH002308

31. W. Ma, *et al.*, Iron-Catalyzed Anti-Markovnikov Hydroamination and Hydroamidation of Allylic Alcohols. *J. Am. Chem. Soc.* **141**, 13506–13515 (2019) doi: 10.1021/jacs.9b05221
32. H. V. Le, B. Ganem, Trifluoroacetic anhydride-catalyzed oxidation of isonitriles by DMSO: A rapid, convenient synthesis of isocyanates, *Org. Lett.*, **13** (10), 2584-2585, doi: 10.1021/ol200695y (2011).
33. A. M. van Leusen, J. Wildeman, O. H. Oldenziel, Base-Induced Cycloaddition of Sulfonylmethyl Isocyanides to C, N Double Bonds. Synthesis of 1, 5-Disubstituted and 1, 4, 5-Trisubstituted Imidazoles from Aldimines and Imidoyl Chlorides. *J. Org. Chem.* **42**, 1153–1159 doi: 10.1021/jo00427a012 (1977).
34. C. Guérin, V. Bellosta, G. Guillaumot, J. Cossy, Mild nonpimerizing N -alkylation of amines by alcohols without transition metals, *J. Org. Lett.*, **13** (13), 3534-3537, doi: 10.1021/ol201351a (2011)
35. Reaxys. <https://new.reaxys.com/> (accessed on March 30, 2021)
36. www.rdkit.org (accessed on March 30, 2021)

Supplementary information

A Reactivity First Approach to Autonomous Discovery of New Chemistry

Dario Caramelli, Jaroslaw M. Granda, S. Hessam M. Mehr, Dario Cambie, Alon Henson, 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 GPLv3 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 specification	4
1.1	Benchtop MS	6
1.2	Benchtop NMR	7
1.3	Syringe pumps	8
1.4	Building the platform	9
1.5	The chemical space	10
1.6	The light shield	11
2	Neural network for reactivity assignment	13
2.1	Manual reactivity assignment	13
2.2	NMR spectra pre-processing	15
2.3	Neural network structure	16
3	Chemical space exploration network	16
4	Reaction discovered with the platform and variations	17
4.1	Diethyl 2-bromomalonate and p-toluenesulfonylmethyl isocyanide	17
4.2	1H-benzotriazol-1-ylmethyl isocyanide and diethyl 2-bromomalonate	18
4.3	Silyl isocyanide and diethyl 2-bromomalonate	19
4.4	Variations of activators	20
4.5	Asymmetric variations of product 25	22
4.6	Other Isocyanide variations	24
5	Reaction mechanism data	25
5.1	HPLC-analysis	25
5.2	Time resolved HPLC analysis	28
5.3	Water influence	31
5.4	Reaction performed in different solvents	32
5.5	Synthesis of isotopically substituted starting materials	34
5.5.1	Synthesis of (¹³ C)TosMIC	34
5.5.2	Isocyanide Substitution	34
5.5.3	Methylene Labelling	37
5.5.4	DMSO Labelling	40
5.5.5	Reaction in ¹⁸ O H ₂ O	42

5.6	IR reaction monitoring	44
5.7	Online NMR monitoring	46
6	Cheminformatics simulation	48
6.1	Reaction network	48
6.2	Similarity index with known TosMIC reactions	49
7	Other reactions discovered and re-discovered	51
7.1	Phenylhydrazine and bromoacetonitrile under 450 nm irradiation.....	52
7.2	Diethyl 2-bromomalonate and 1-vinyl-2-pyrrolidinone under 450 nm irradiation	53
7.3	<i>N,N</i> -dimethyl-4-nitrosoaniline, bromoacetonitrile and diethyl 2-bromomalonate	54
7.4	Diethyl 2-bromomalonate, 4-phenylurazole and 1-vinyl-2-pyrrolidinone.....	55
8	Crystal structure details.....	56
9	¹ H and ¹³ C NMR spectra	59
10	References	87

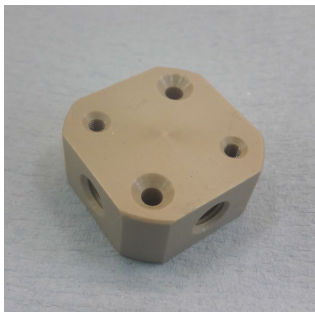
1 Hardware specification



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)



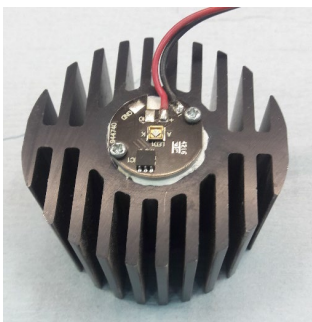
Duran® laboratory bottles, with caps capacity 100 mL, blue PP screw cap and pouring ring Sigma-Aldrich Z305170-10EA and Z305200-10EA 18 x 50-250 mL for reagents 2 x 1 L for solvent and waste



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 μ L/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 5×10^3

ESI parameters

- Capillary temperature(V) 250.0
- Capillary Voltage(V) 180.0
- ESI Gas Temperature: 250 $^{\circ}\text{C}$
- 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
- ¹H 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.

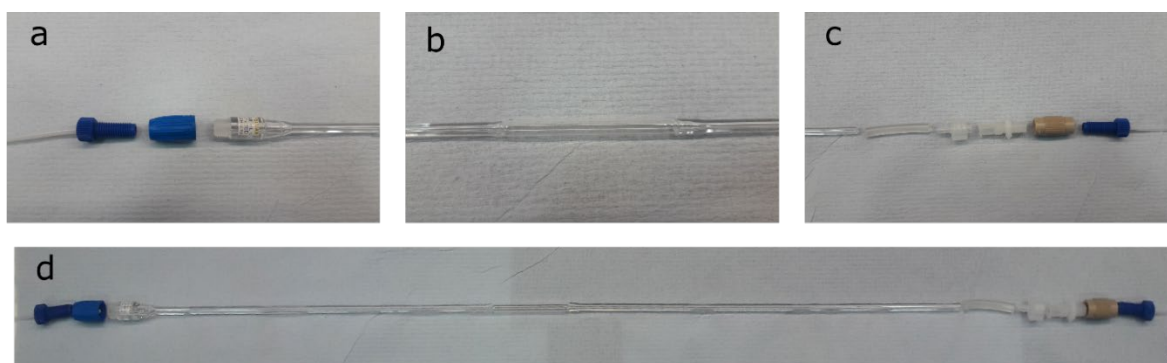


Figure S1: a) top connector, the screw connector has been welded to the glass cell by the glassblower. b) Central part, it has been made by welding a normal 5 mm NMR tube to the rest of the cell. c) bottom connection, the glass could not have the threading because the whole cell needs to fit through 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.

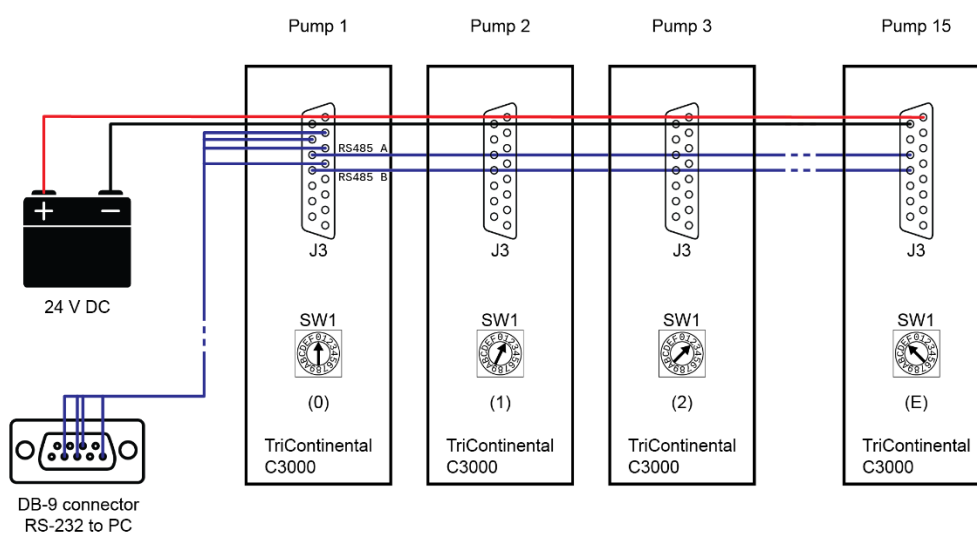


Figure S2: Schematic representation of the daisy chain connection used to control and power the pumps. On a single line up to 15 pumps can be connected, the individual addressed are defined with a physical switch on their back with positions between 0 and E. F position is used for debugging.

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**).

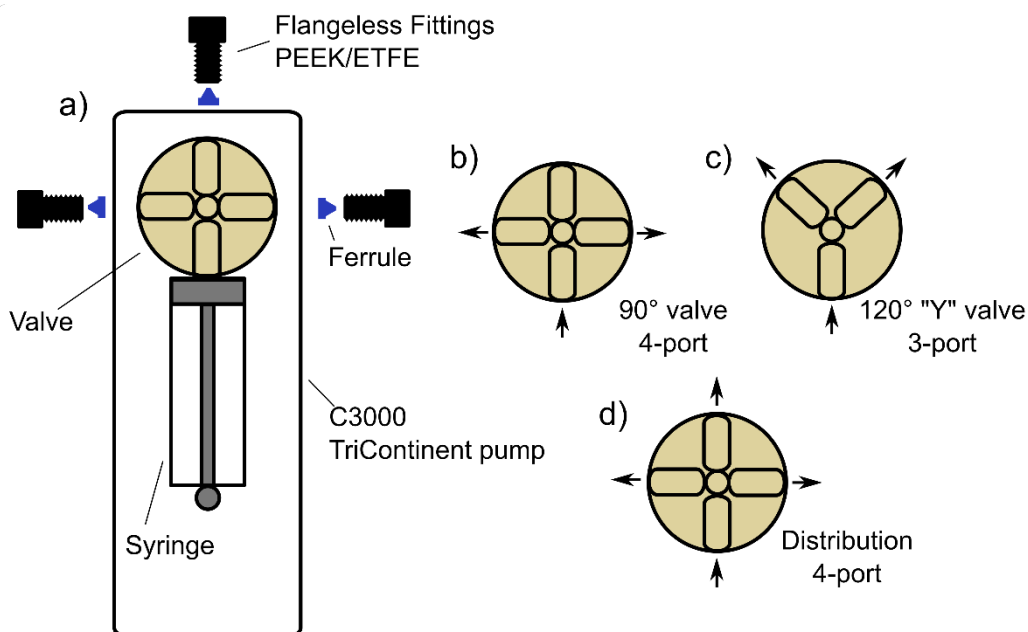


Figure S3: a) Schematic view of the front of a C3000 TriContinent pump. b) Scheme of a 90° valve, the top port is used to bypass the pump, therefore it is possible to pump from the syringe only in two directions. c) Scheme of the 120° valve, it can pump in two directions. d) Scheme of the Distribution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

1.4 Building the platform

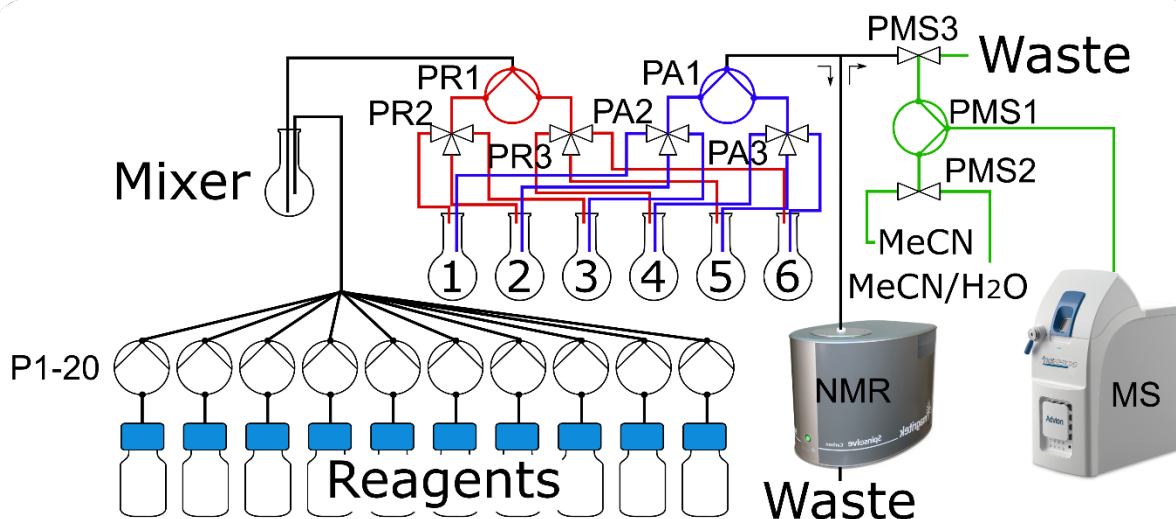


Figure S4: Scheme of the platform including pumps labels.

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**).

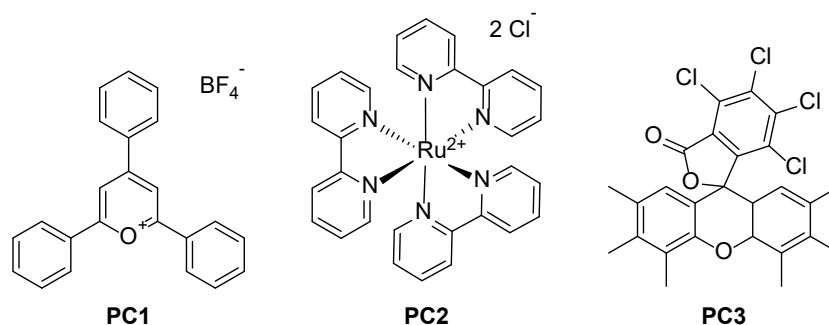


Figure S5: Photocatalysts used as additives to expand the chemical space.

1.6 The light shield

A shielding apparatus was designed and built 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**).

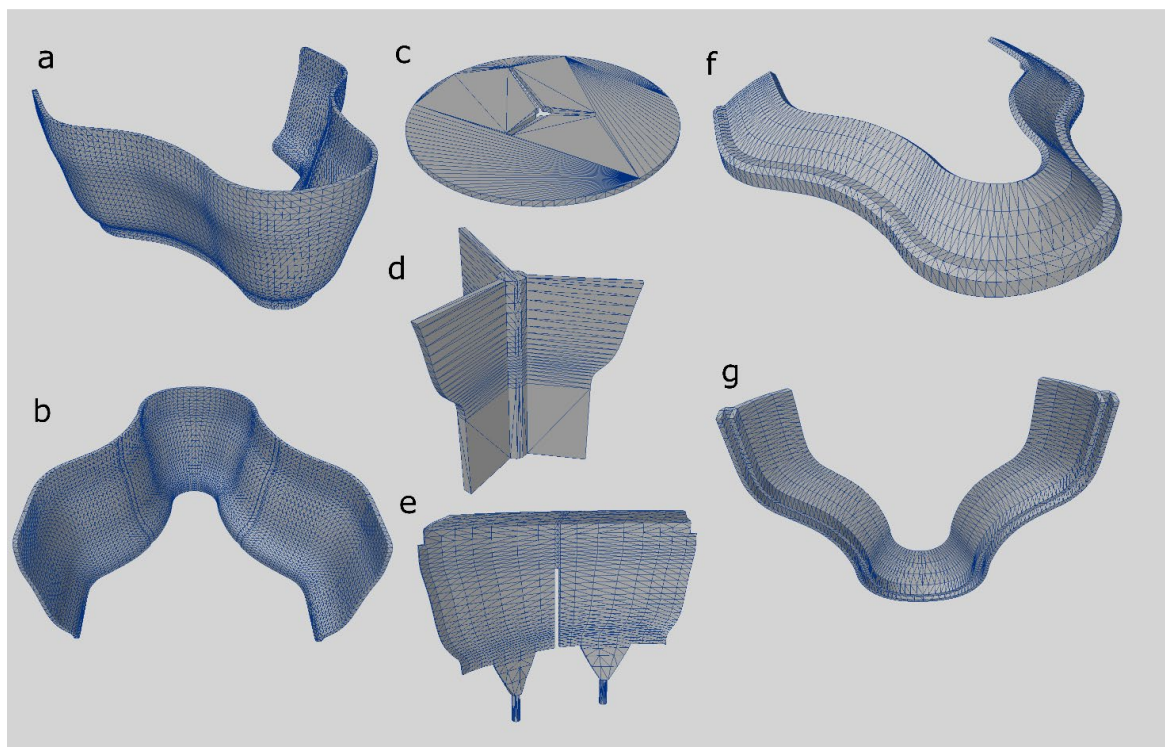


Figure S6: **a,b:** front shield, the internal part was covered with aluminium foil. **c:** base of the apparatus, a narrow gap in the centre allows to slot in the internal walls. **d:** internal walls, they isolate the three reactors from the unwanted LEDs light. **e:** backwall, it confines the light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

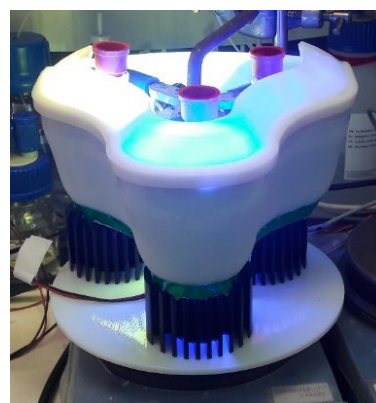
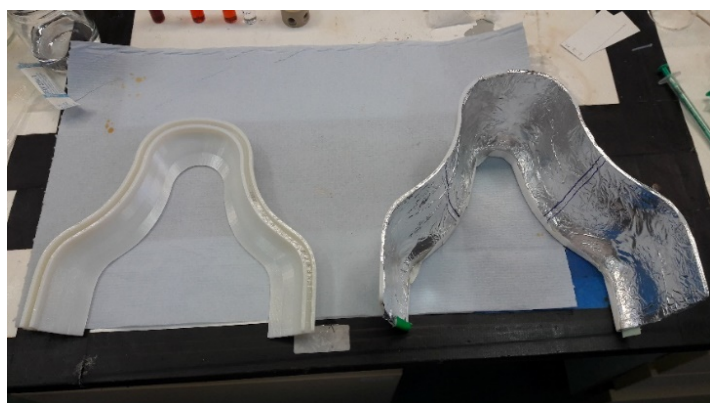


Figure S7: *Left:* front shield covered with aluminium foil and its cover. *Right:* Full setup. The LEDs are mounted on the heatsinks and placed on a stirring plate. Reactors are held on top and isolated by the 3D printed shield.

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**.

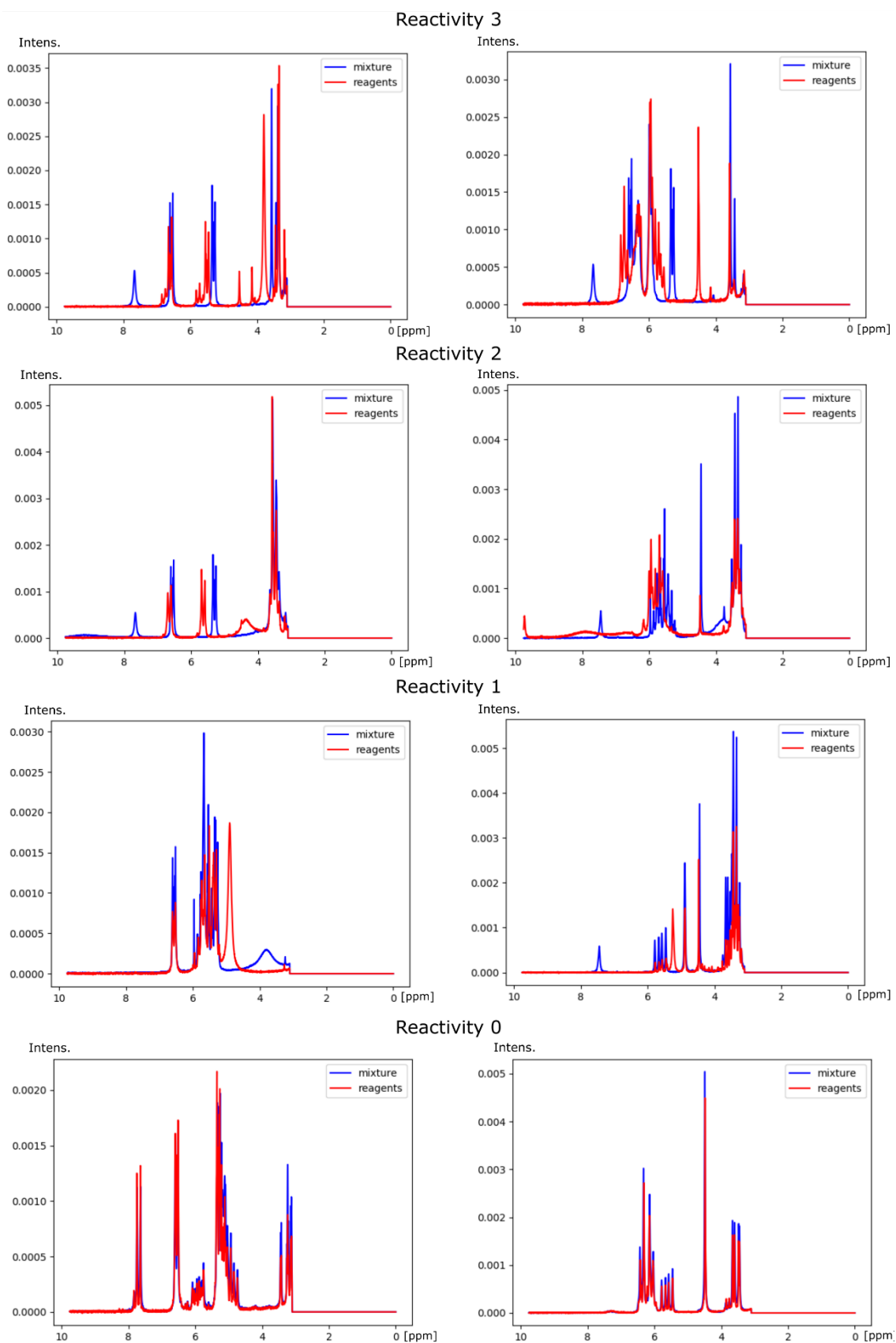


Figure S8: Examples of NMR spectra evaluated manually. By looking at the new peaks appearing and the reagents peaks disappearing a reactivity between 0 and 3 is assigned.

2.2 NMR spectra pre-processing

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.

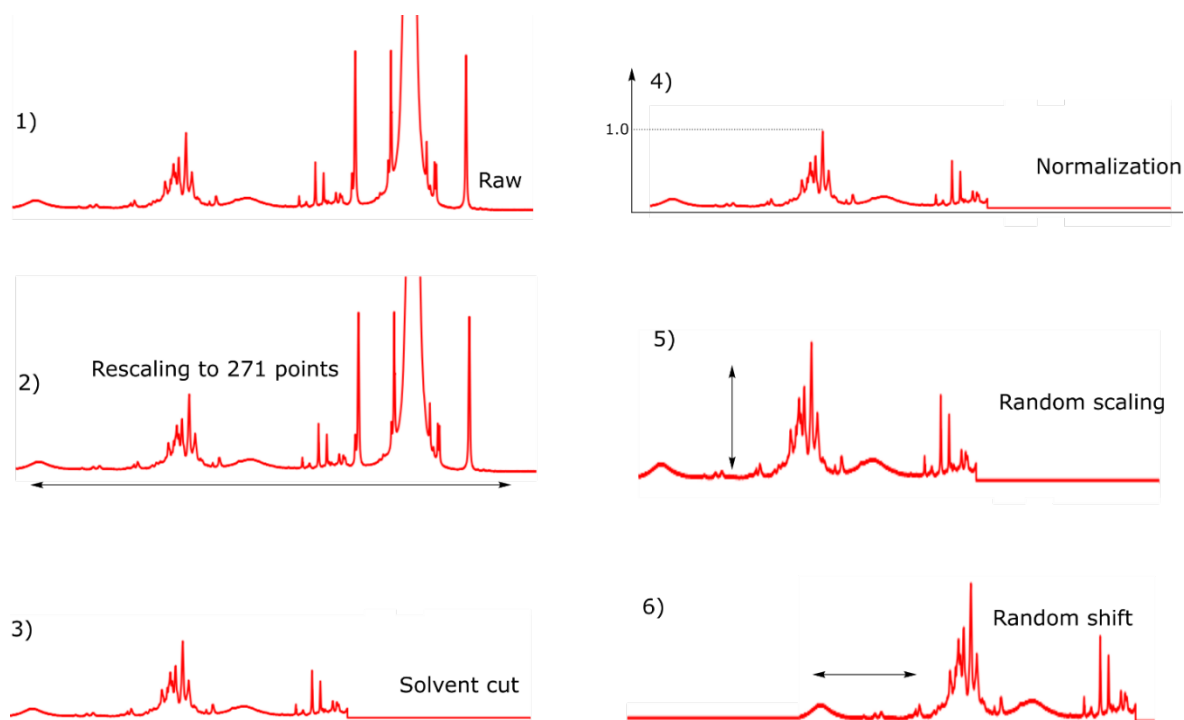


Figure S9: Pre-processing performed on the NMR spectra before using them as input. The raw spectrum (1) is rescaled to 271 points (2) and the solvent is cut out (3). It is then normalized to the highest peak (4) and a random factor on both axes is introduced during the network training to avoid overfitting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

The dataset was fed into the network by encoding the reagents involved into SMILES. The auto encoder developed by Jin et al 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 based on Tensorflow (2) python library.

4 Reaction discovered with the platform and variations

4.1 Diethyl 2-bromomalonate and p-toluenesulfonylmethyl isocyanide.

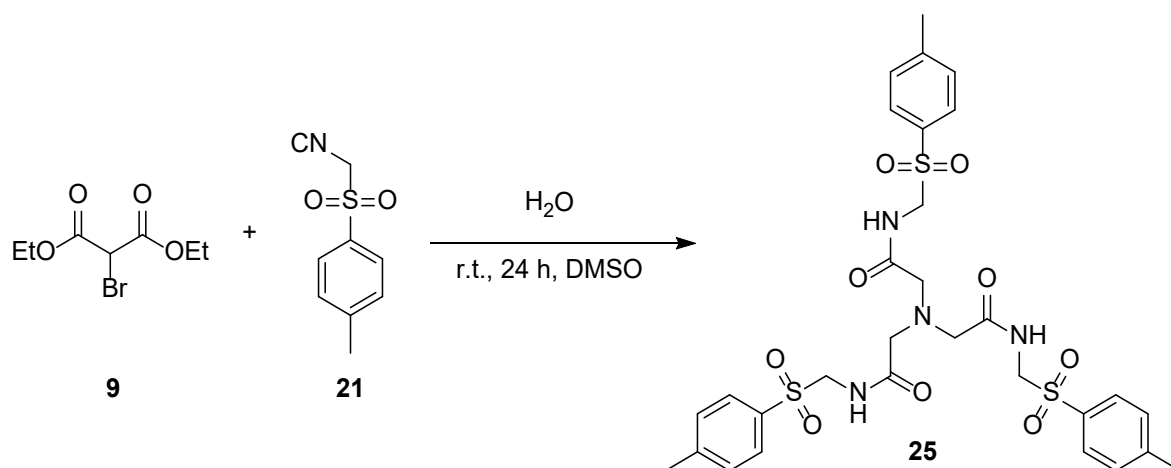


Figure S10: Scheme of the reaction yielding product **25**

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)

^1H NMR (600 MHz, $\text{DMSO}-d_6$) 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: $[\text{C}_{30}\text{H}_{36}\text{N}_4\text{O}_9\text{S}_3\text{Na}]^+$ Calculated 715.1537 m/z , measured 715.1503 m/z

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

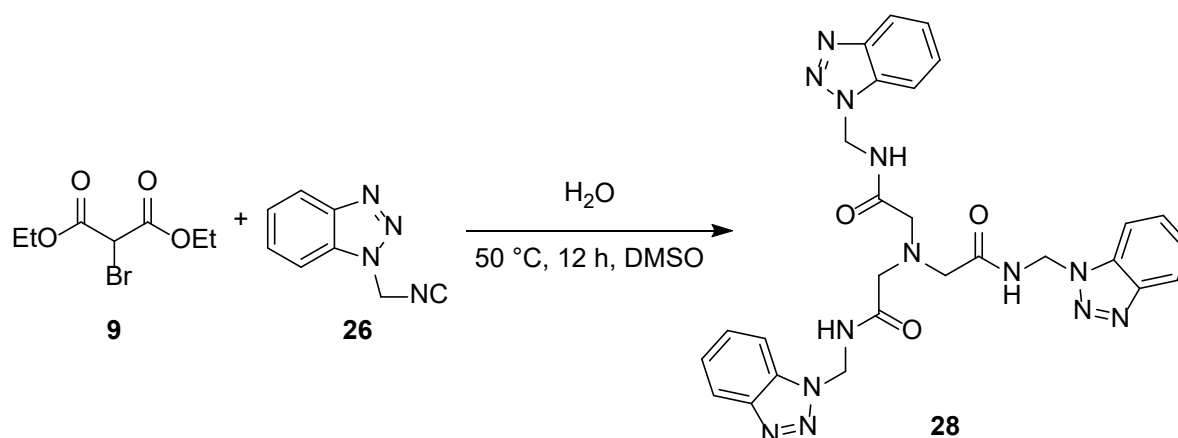


Figure S11: Scheme of the reaction yielding product 28

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 μ l)) 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. Mg₂SO₄ 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)

¹H NMR (600 MHz, CDCl₃) 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).

¹³C NMR (151 MHz, CDCl₃) δ 133.79, 107.78, 94.54, 90.27, 86.75, 81.52, 72.87, 21.40, 12.93.

ESI-HR-MS: [C₂₇H₂₇N₁₃NaO₃]⁺ calculated 604.2252 m/z, measured 604.2232 m/z.

4.3 Silyl isocyanide and diethyl 2-bromomalonate

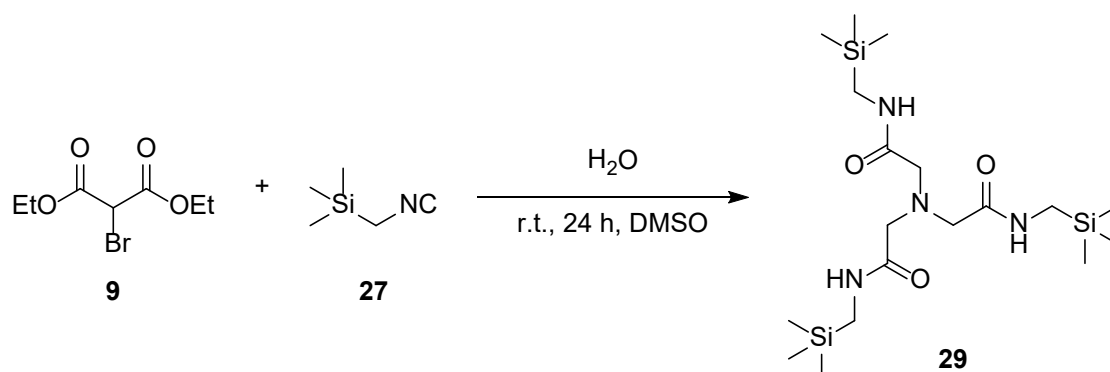


Figure S12: Scheme of the reaction yielding product **29**. The product is only observed in traces in HPLC-MS.

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

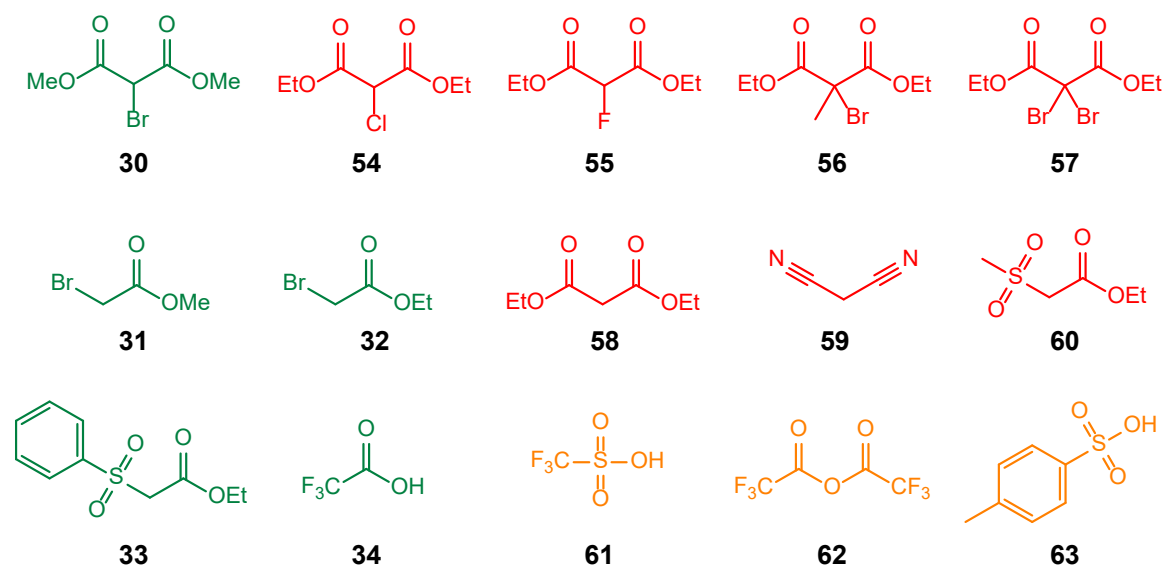


Figure S13: List of molecules tried as activator replacing diethyl bromo malonate.

The reaction has been repeated according to the conditions indicated in section 3.1, replacing the diethyl bromomalonate with various analogous molecules and organic acids. The molecules coloured in green (**30**, **31**, **32**, **33** and **34**), yielded product **25** as white precipitate. The molecules coloured in red (**54**, **55**, **56**, **57**, **58**, **59**, **60**) did not produce the desired product. In the reaction with molecule **61**, **62** and **63** the product was observed in the HPLC analysis but did not precipitate. The reaction with molecule **63** produced molecule **64**, while molecule **34** produced a mixture of product **25** and **65** (**Figure S14**). The synthesis of molecules **69** and **70** from TosMIC as already been reported in literature (3).

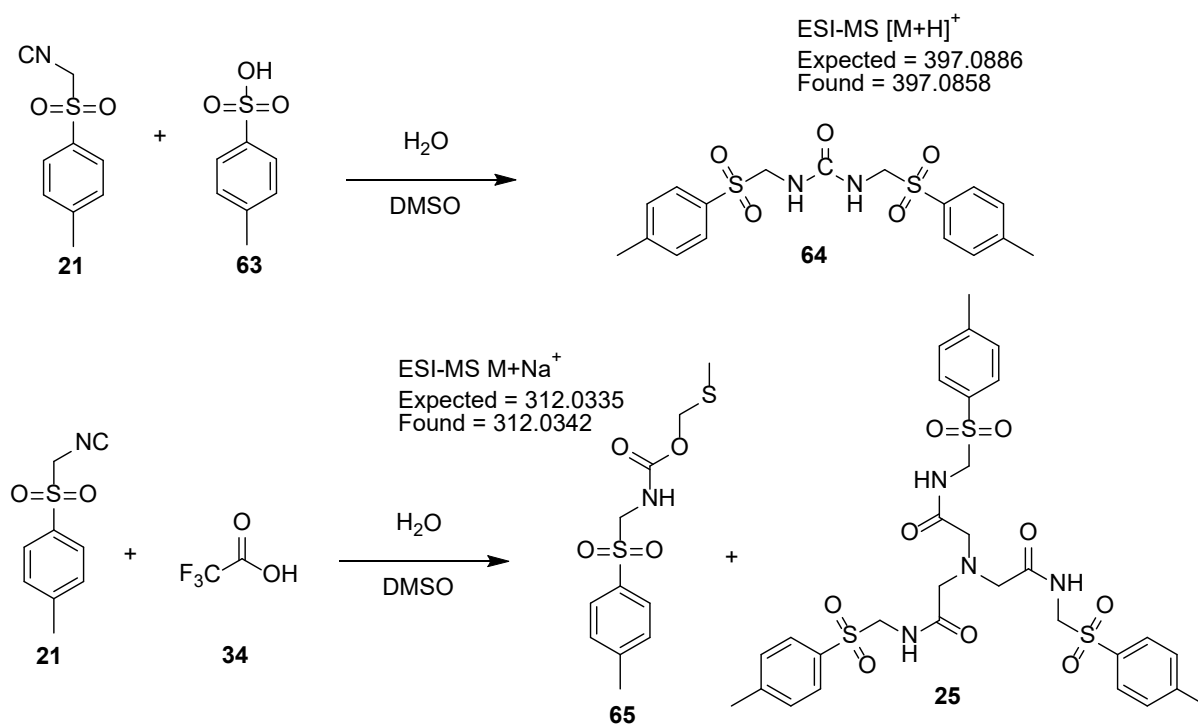


Figure S14: Side products obtained while testing PTSA (**63**) and TFA (**34**) as activators. The reactions yielding products **64** and **65** are known in literature.

Compound 64

$^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$) 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).

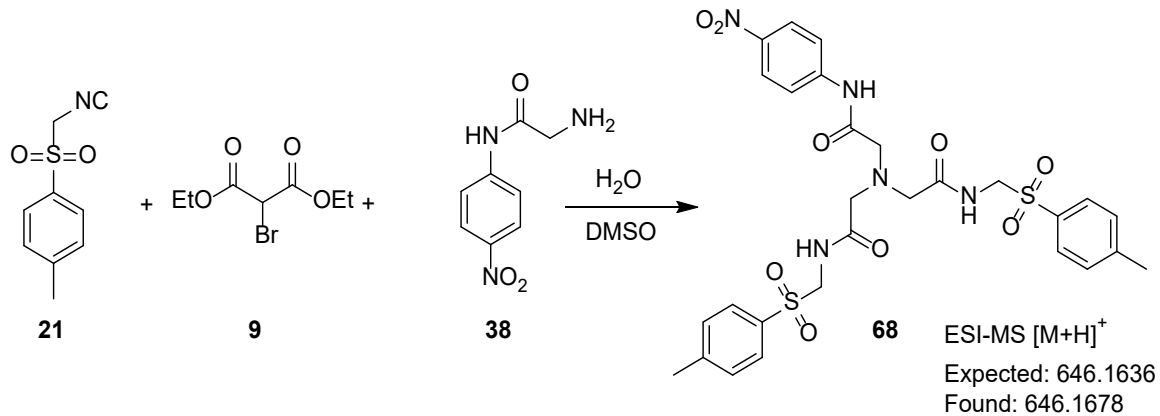
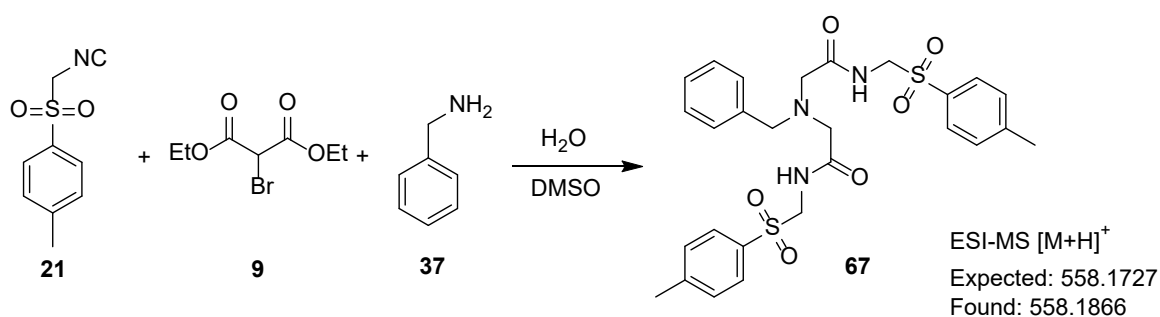
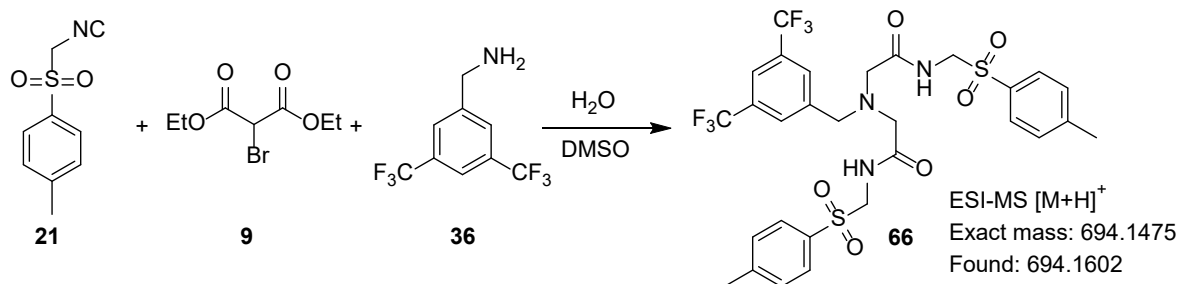
Compound 65

$^1\text{H NMR}$ (600 MHz, $\text{DMSO-}d_6$) 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).

$^{13}\text{C NMR}$ (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 3.1 adding one equivalent of amine (**36**, **37**, **38**, **40**). 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 S15**.



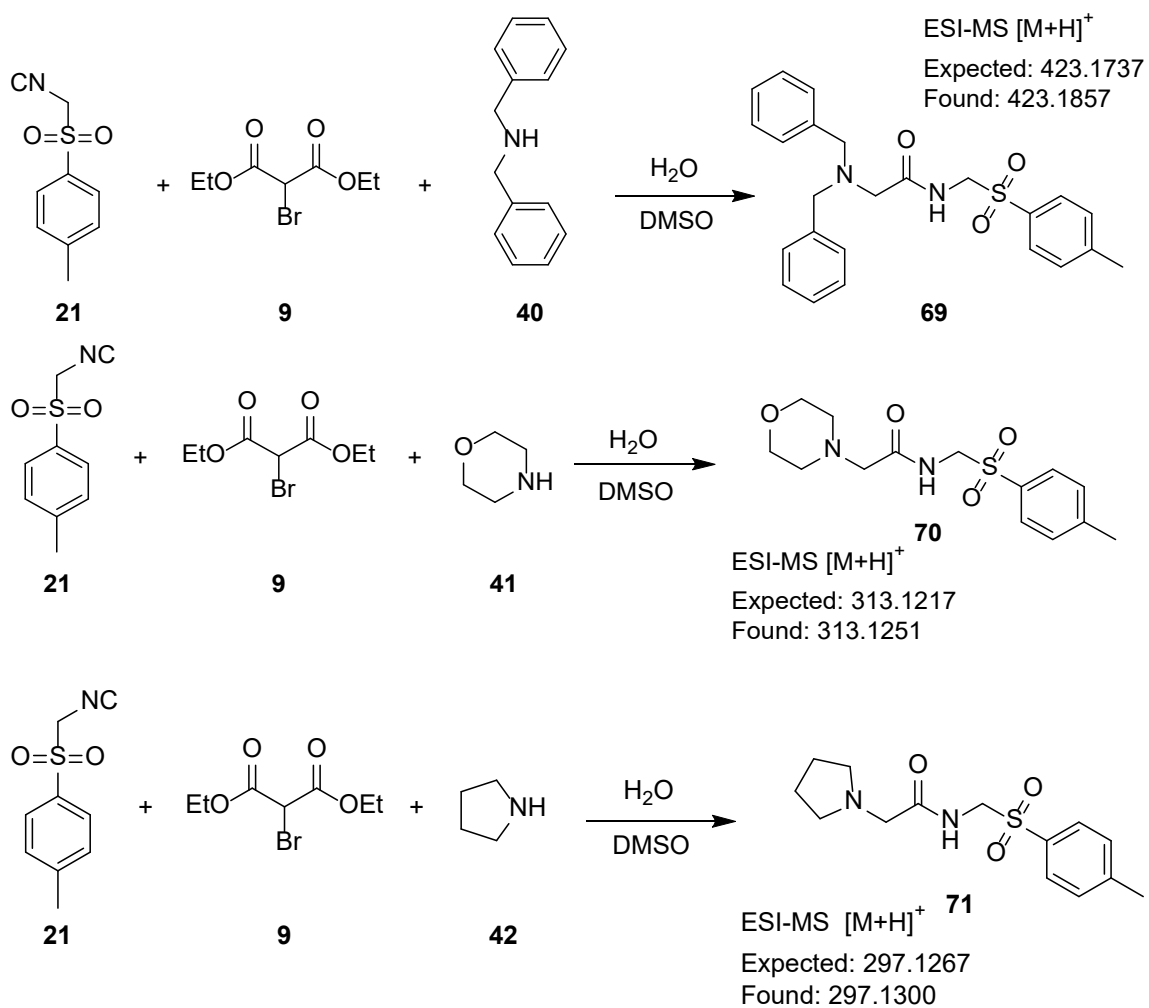


Figure S15: Scheme of the reaction variations involving the presence of a primary or secondary amine. The analogs of the mechanism intermediates have been detected by HPLC-MS.

4.6 Other Isocyanide variations

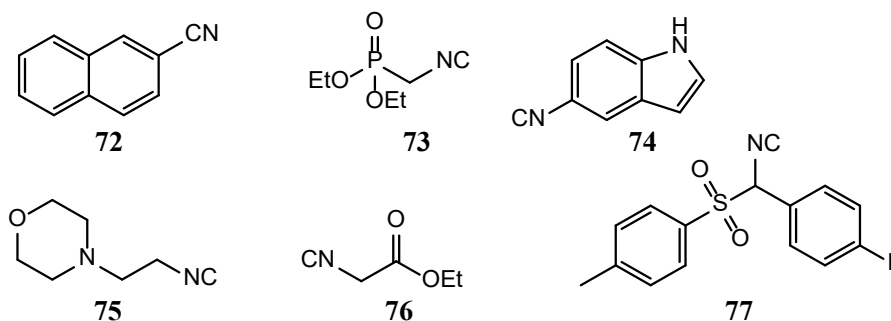


Figure S16: Other isocyanide variations tried on the reaction. None of them yielded product **25**.

The reaction has also been tried on the isocyanides reported in **Figure S16** (2-Naphthyl isocyanide **72**, Diethyl isocyanomethylphosphonate **73**, 1H-Indol-5-yl isocyanide **74**, 2-Morpholinoethyl isocyanide **75**, Ethyl isocyanoacetate **76**, α -(p-Toluenesulfonyl)-4-fluorobenzylisocyanide **77**). The reactions were prepared according to the procedure reported in section 3.1 and repeated at 50°C. HPLC-MS analysis showed no traces of the analogues of product **25**. The analysis of molecule **77** showed the presence of the one-branched (**78**) and two-branched (**79**) analogues suggesting that the intermediate is too hindered to form the third branch (**Figure S17**).

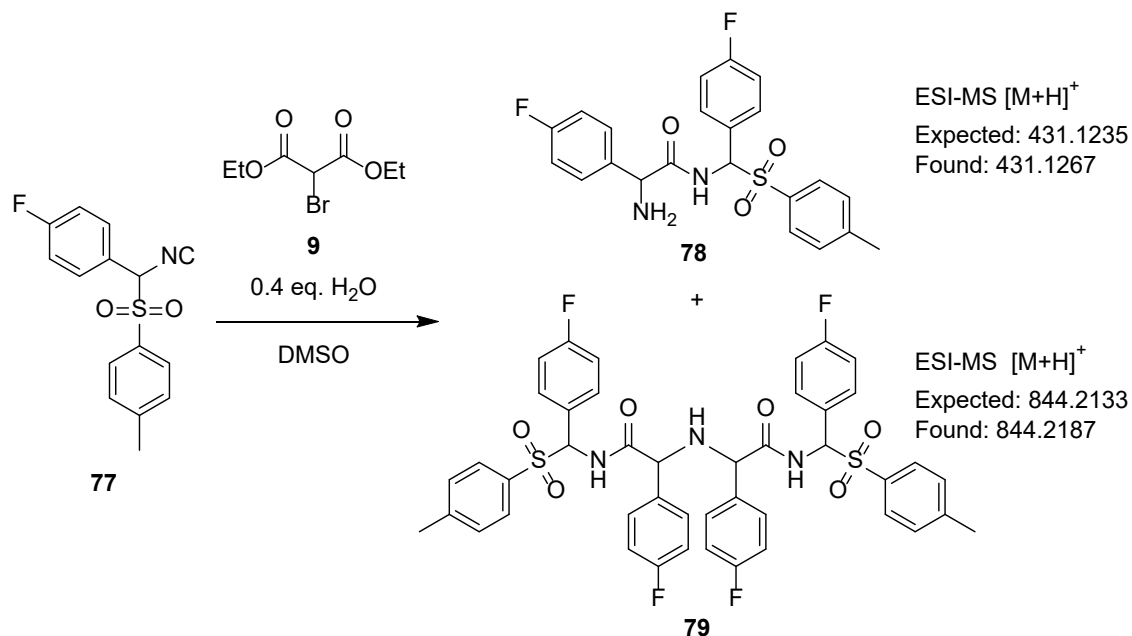


Figure S17: Isocyanide **77** did not produce the expected analogue of **25**. However, the one-branched (**78**) and two-branched (**79**) versions were detected.

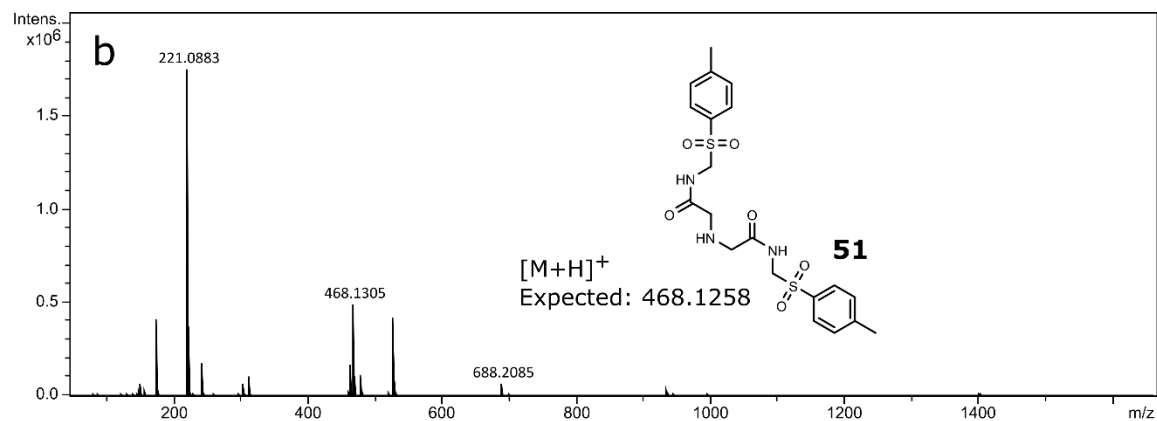
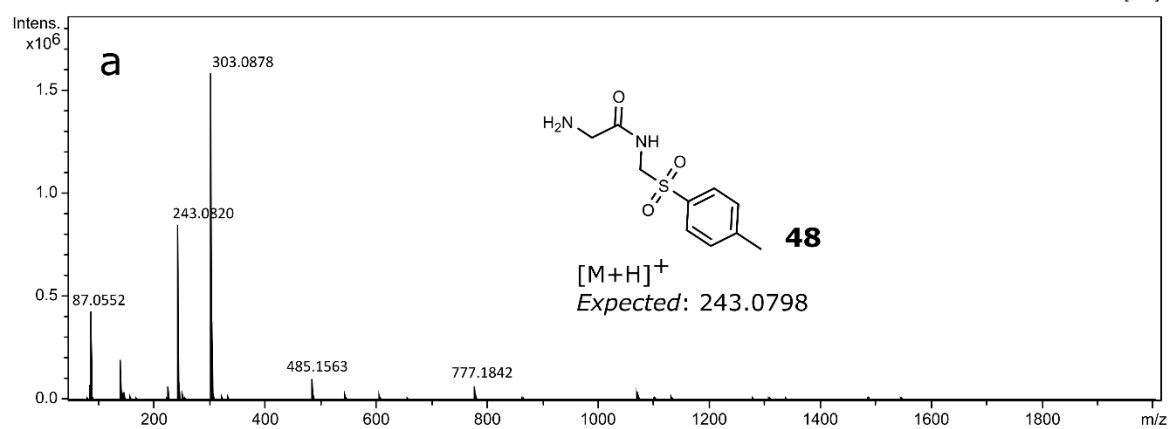
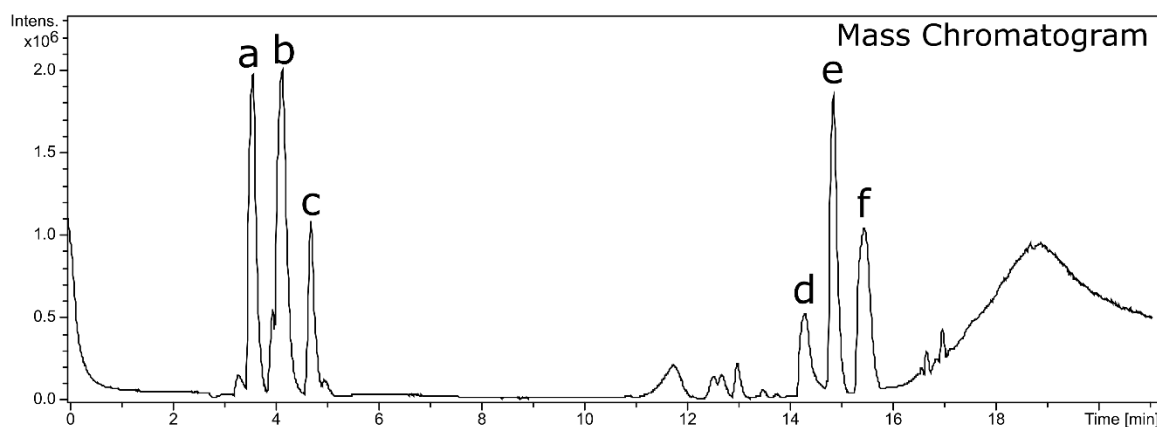
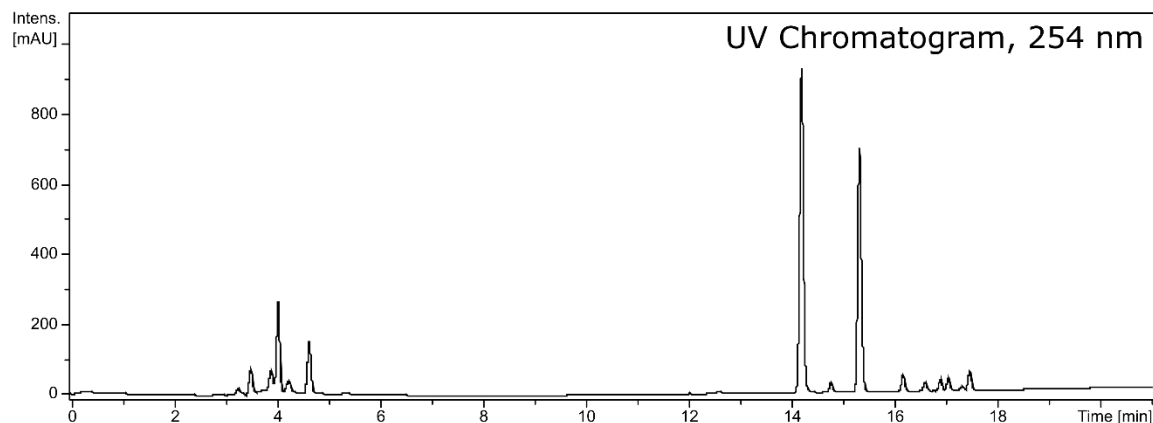
5 Reaction mechanism data

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 S1**. The column compartment was maintained at 30 °C. Results after 12h of reaction are showed in **Figure S18**. 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 S1: 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



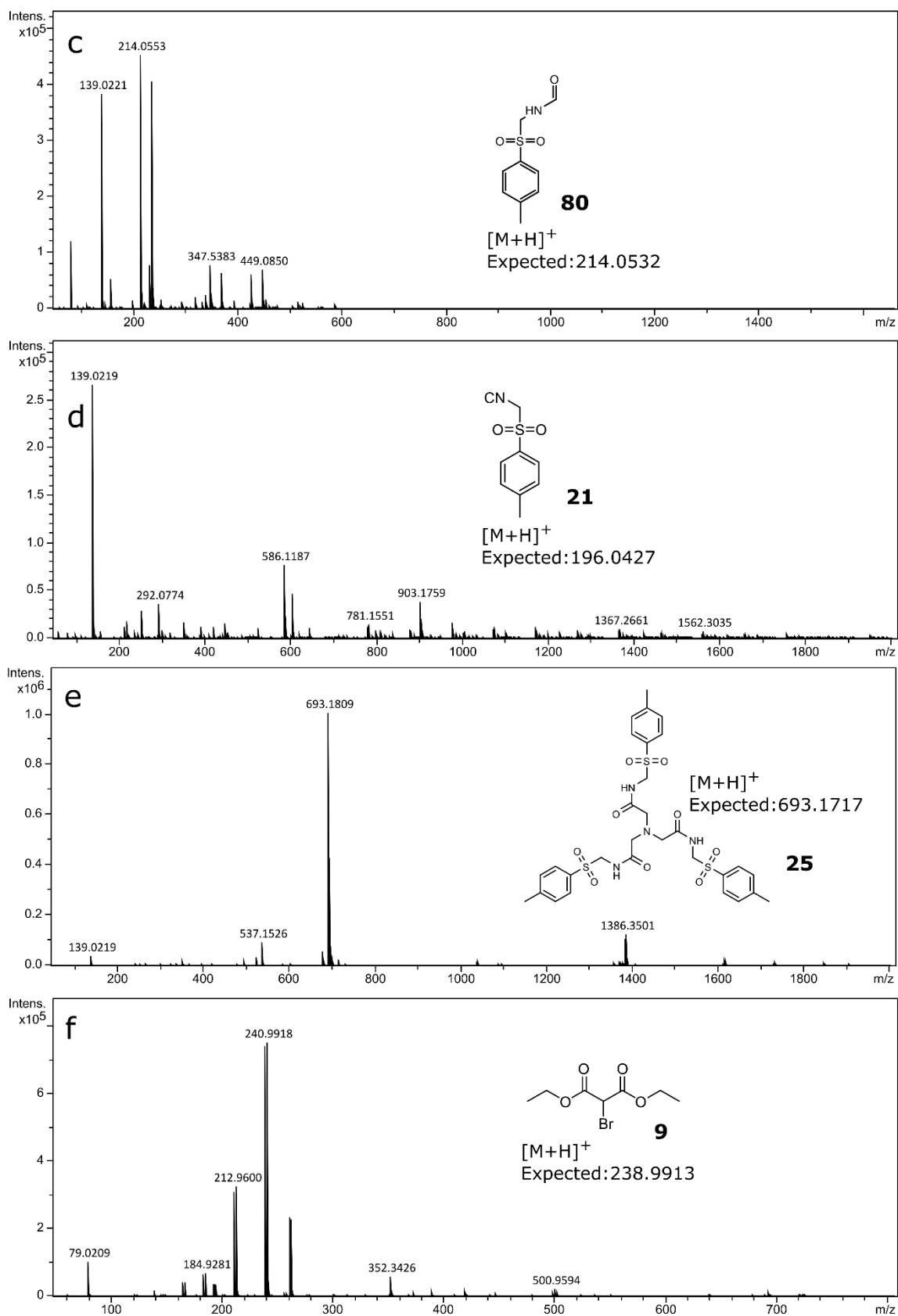
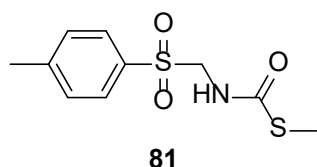


Figure S18: HPLC-MS analysis of the reaction mixture. The MS data of the major peaks are reported with the assigned molecules

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 μ l 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 S20**. 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 **51** forms reaching its max at 20 h and then gets consumed while the integrals of the final product **25** 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 **81** in **Figure S19**, an adduct of DMSO and TosMIC not involved in the proposed mechanism.



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

Figure S19: Structure of the side product corresponding to the peak 3 in **Figure S20**. It is believed to not be involved in the mechanism.

^1H 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.

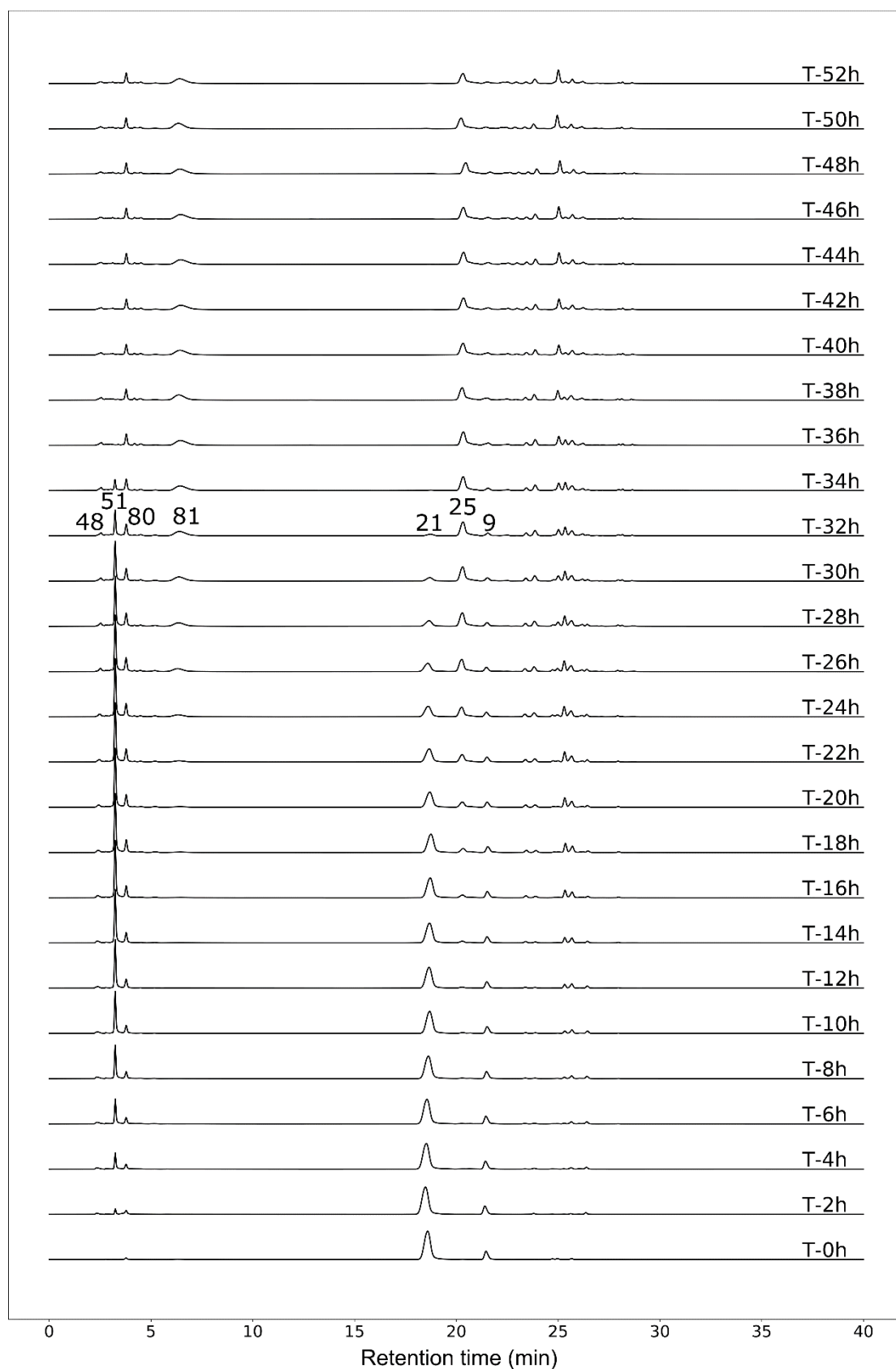


Figure S20: Online HPLC analysis of the reaction mixture. The known peaks are marked with the molecule number.

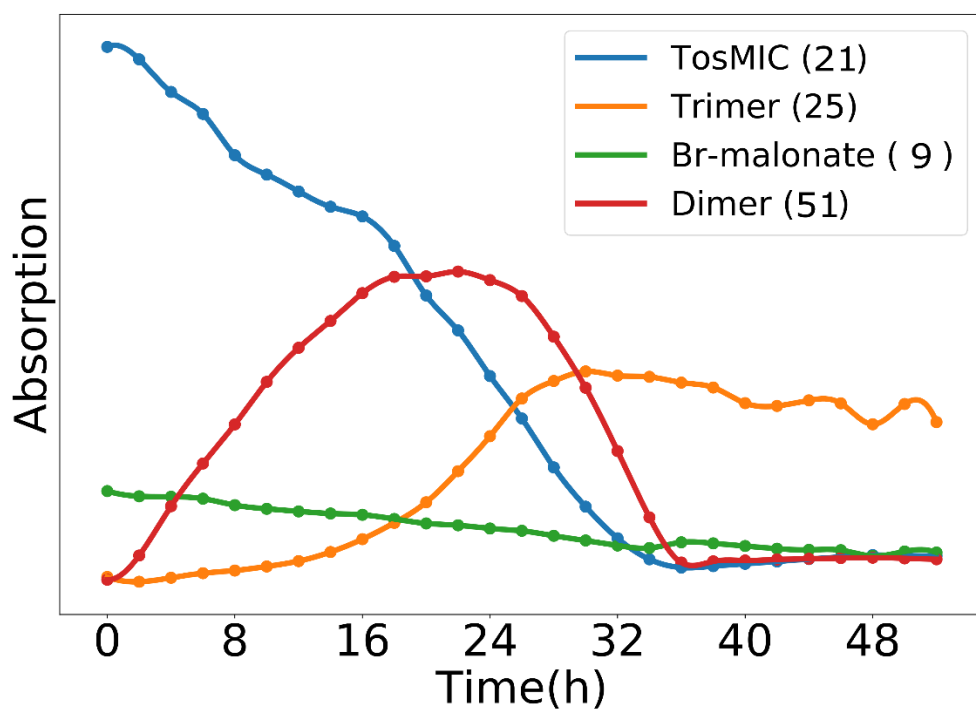


Figure S21: Visualisation of the time-resolved integrals in Figure S20 of the reagents **12** and **22**, the intermediary **50** and the product **25**.

5.3 Water influence

The calibration curve of the product **25** absorption is acquired with multiple injections of product **25** and an internal standard (hexafluorobenzene) at different ratios (**Figure S22**), integrals were integrated from the 210 nm wavelength data.

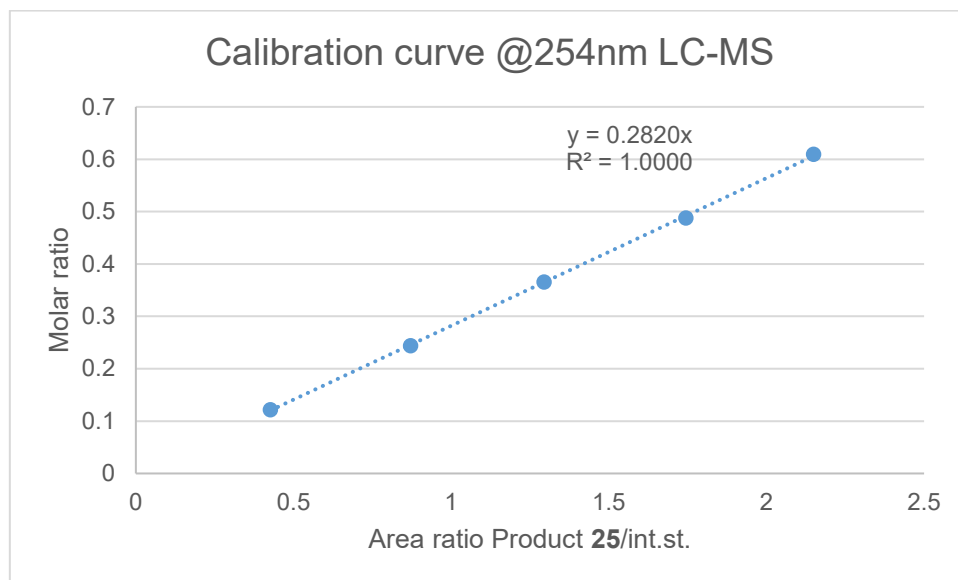


Figure S22: Calibration curve for the product **25**. Samples at different ratios with the internal standard were injected.

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 S2**)

Table S2: The yield of the reaction performed in presence of different amounts of water was determined through HPLC analysis.

Water equivalentents	Area product 25 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%

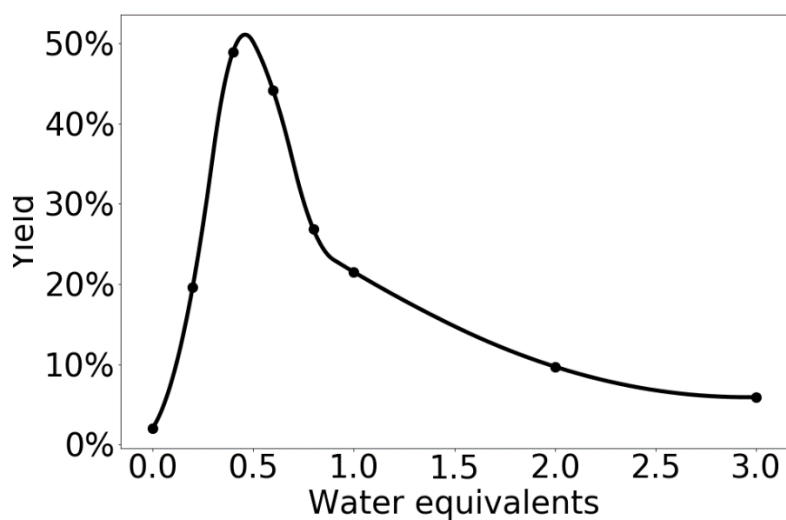


Figure S23: Visualized data points of **Table S2**.

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 S24**). As further test it was also repeated in different ratio of DMSO/DMF showing the direct influence of DMSO in the yield (**Figure S25**).

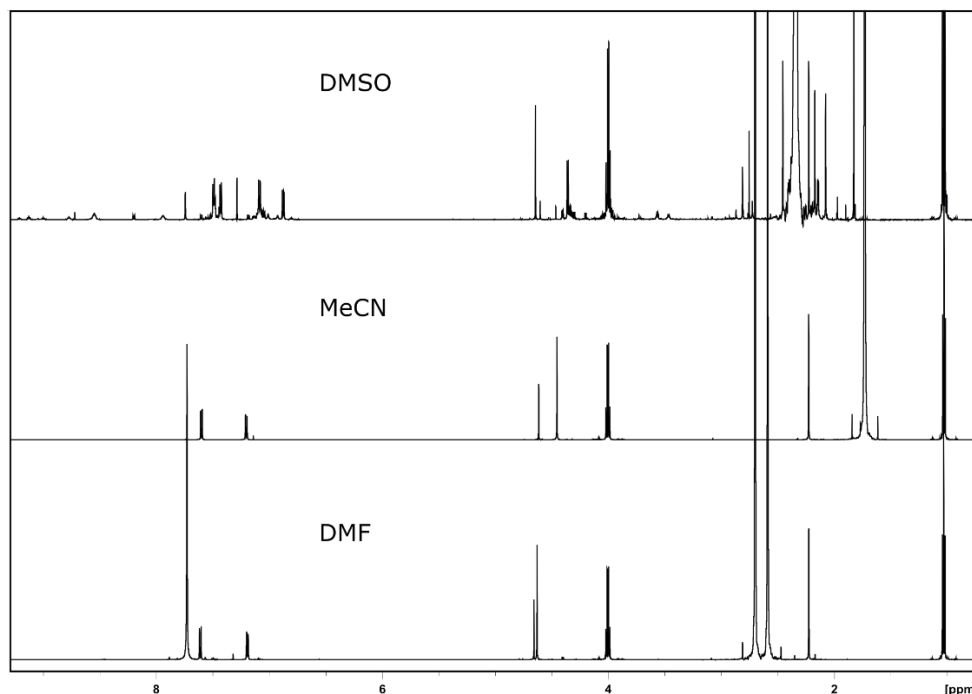


Figure S24: ^1H -NMR spectra of the reaction performed in different solvents. Acetonitrile and DMF show no reactivity.

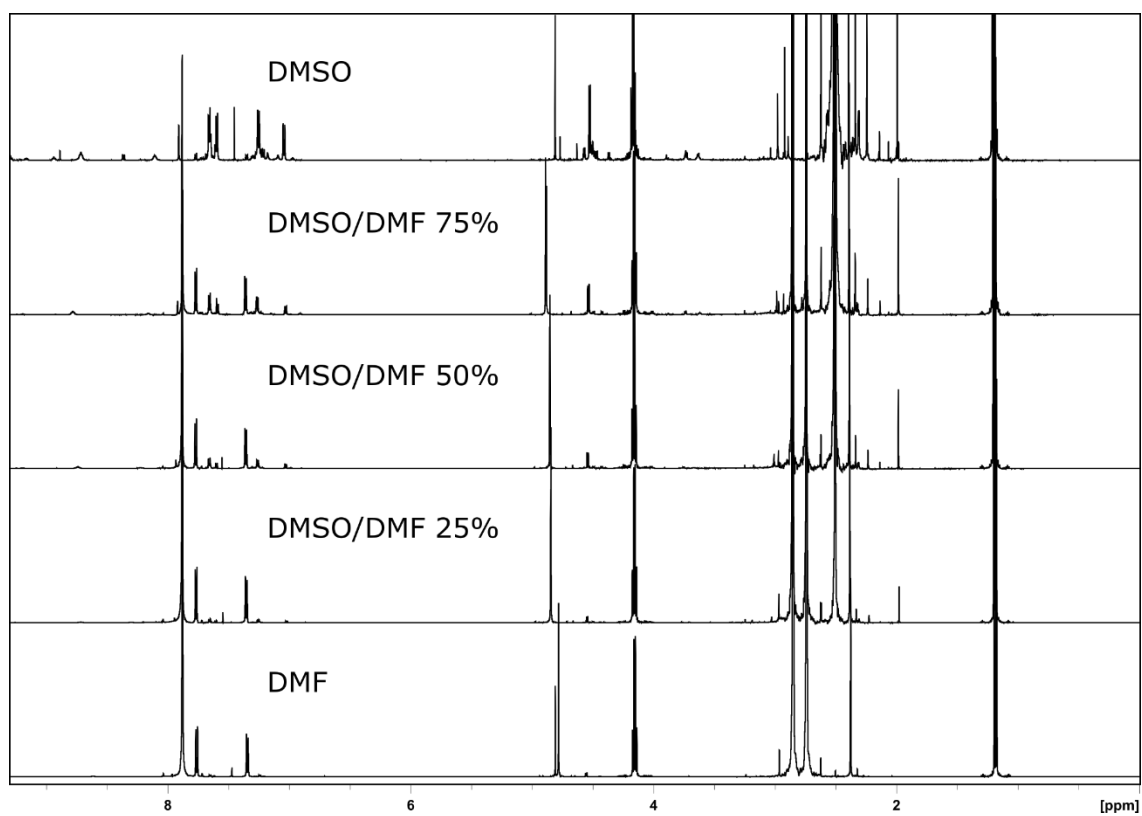


Figure S25: ¹H-NMR spectra of the reaction performed in a mixture of DMF and DMSO at different ratios.

5.5 Synthesis of isotopically substituted starting materials

5.5.1 Synthesis of (¹³C)TosMIC

¹³C-isotopically substituted TosMIC was prepared, based on literature procedures (4, 5), using either ¹³C formamide (to substitute the isocyanide carbon) or a solution of ¹³C-formaldehyde (to substitute the methylene carbon).

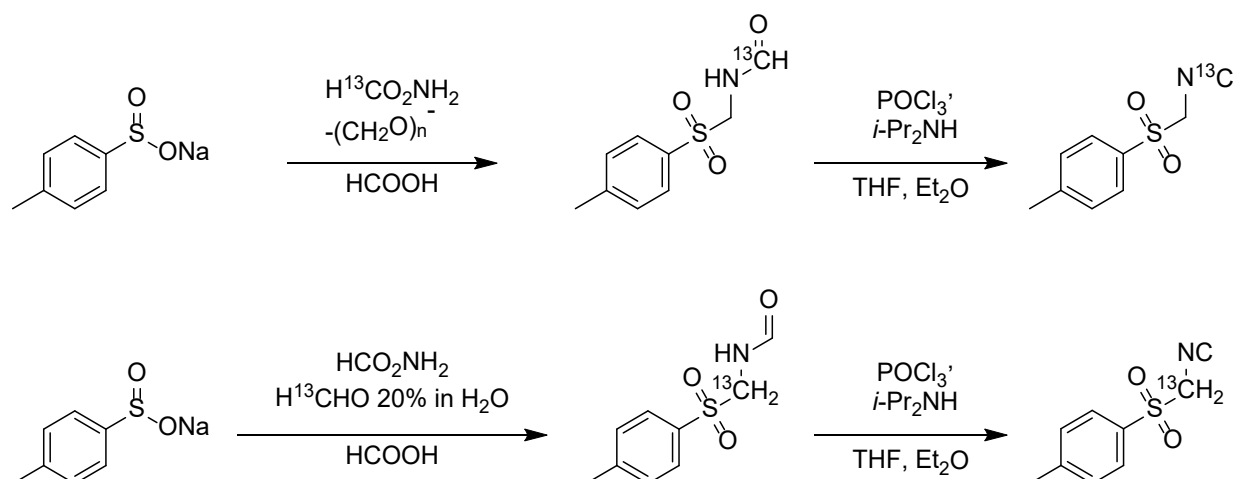


Figure S26: Synthetic approach to ¹³C-isotopically substituted TosMIC.

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-toluenesulfonate (1.00 g, 5.6 mmol), paraformaldehyde (Sigma-Aldrich, 505 mg, 16.8 mmol, 3 eq.), formamide-¹³C (Sigma-Aldrich, 99 atom% ¹³C, 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.

¹H 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).

¹³C 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.

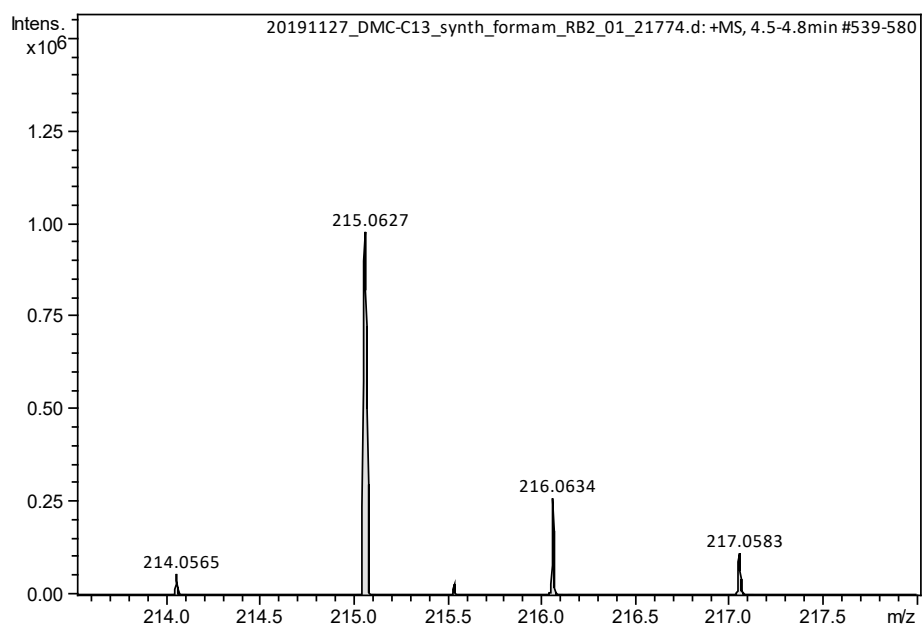


Figure S27: MS peak corresponding to $(1-^{13}C)N$ -(tosylmethyl)-formamide.

5.5.2.2 TosCH₂N¹³C

A 25 ml pear-shaped flask, equipped with magnetic stir bar was charged with ¹³C-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 (6, 7, 8).

¹H 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).

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

LC-MS [M – H]⁻ calculated for C₈¹³CH₈NO₂S 195.031; found 194.9928.

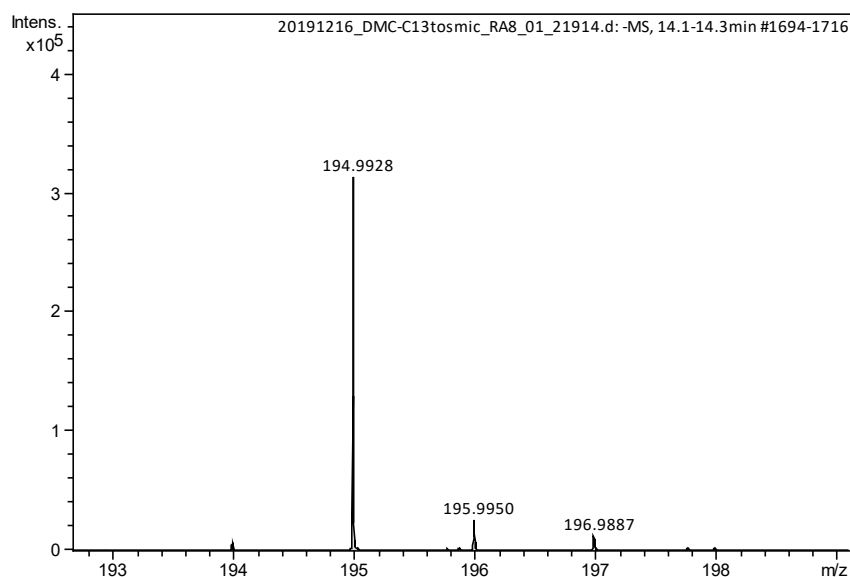


Figure S28: MS peak corresponding to TosCH₂N¹³C

5.5.2.3 Product 25 - ¹³C-1 variation

Synthesized according to standard procedure 3.1.

¹H 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, $\text{DMSO-}d_6$) δ 170.05 (^{13}C), 144.71, 134.39, 129.74, 128.49, 60.00, 56.47 (d, $J = 52.6$ Hz), 21.08.

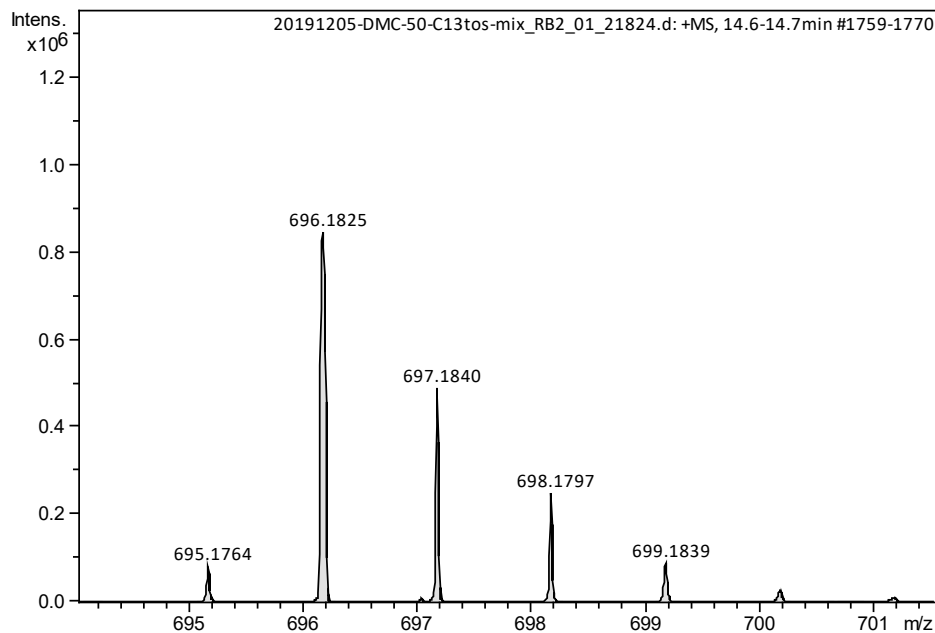


Figure S29: MS peak corresponding to Product 25 - ^{13}C -1 variation

5.5.3 Methylene Labelling

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

Synthesized according to a modified literature procedure (4, 5). A three-necked, round-bottomed flask equipped with a magnetic stirrer is charged with 766 mg (4.3 mmol, 1 eq.) of sodium p-toluenesulfonate, 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 μl (692 mg, 15 mmol, 3.5 eq.) of formic acid. The stirred reaction mixture is heated at 90 $^\circ\text{C}$. The sodium p-toluenesulfonate dissolves during heating, and the solution is kept at 90–95 $^\circ\text{C}$ 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.

^1H NMR (400 MHz, $\text{DMSO-}d_6$) 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, $\text{DMSO-}d_6$) 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 [$\text{M} + \text{H}$] $^+$ calculated for $\text{C}_8^{13}\text{CH}_{12}\text{NO}_3\text{S}$, 215.057; found, 215.0576.

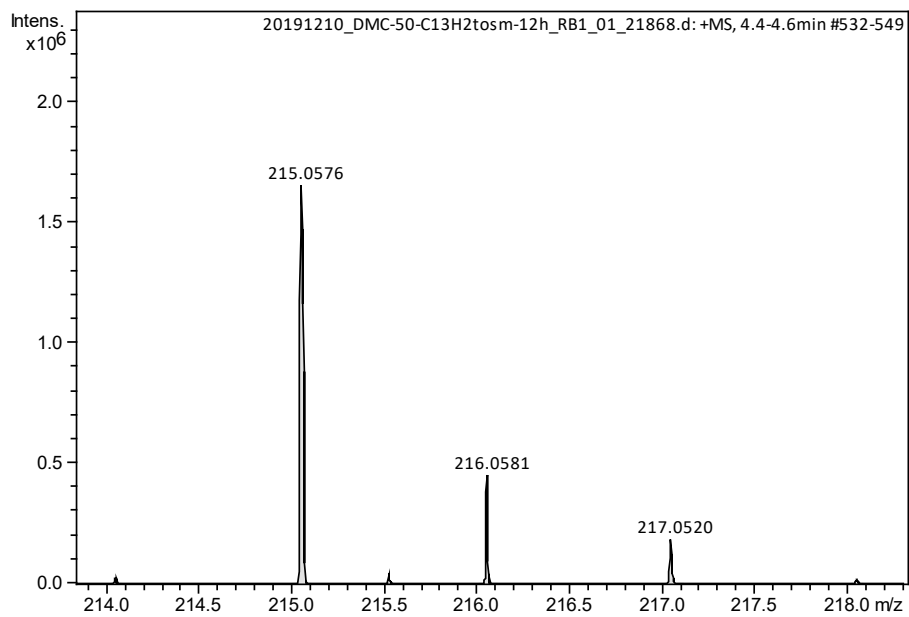


Figure S30: MS peak corresponding to N-(tosyl(¹³C)methyl)-formamide

5.5.3.2 Tos¹³CH₂NC

Same procedure as 4.5.2.2.

¹H 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).

¹³C 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 C₈¹³CH₈NO₂S 195.031; found 194.9984.

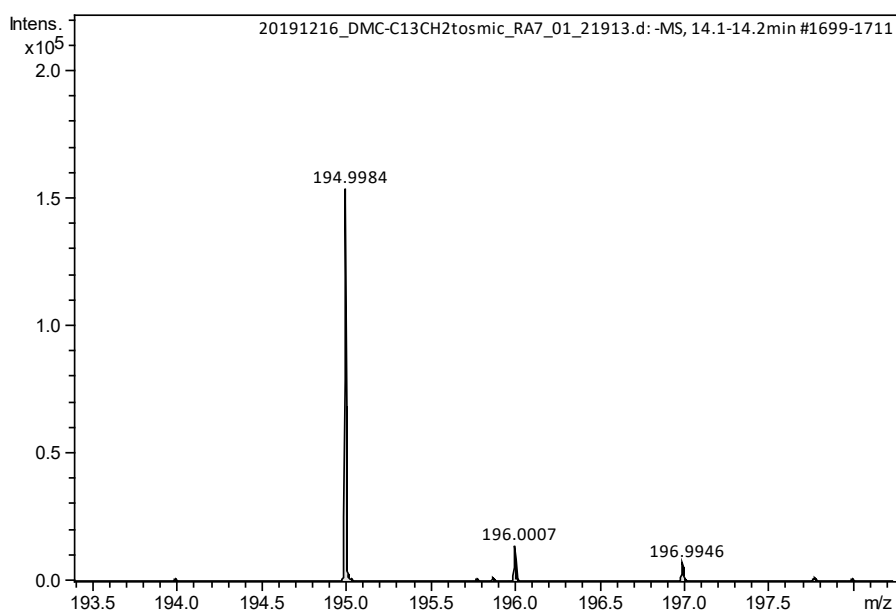


Figure S31: MS peak corresponding to Tos¹³CH₂NC

5.5.3.3 Product 25 - ¹³C-2 variation

Synthesized according to standard procedure 3.1.

¹H 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).

¹³C 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.

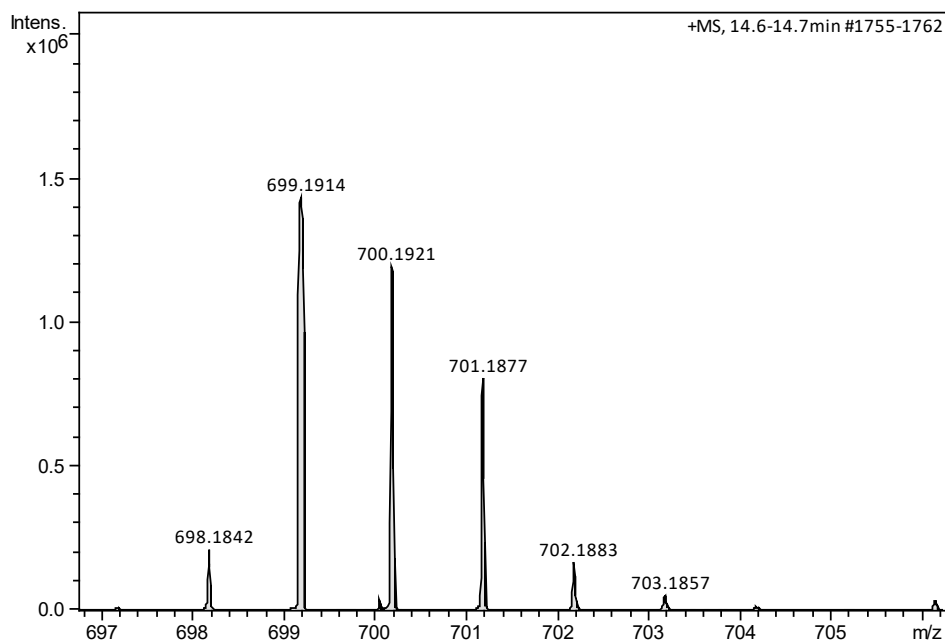


Figure S32: MS peak corresponding to Product 25 – ¹³C-2 variation

5.5.4 DMSO Labelling

5.5.4.1 Synthesis of ¹⁸O-DMSO

Method for the preparation of ¹⁸O 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 ¹⁸O-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 ¹⁸O-DMSO, pure enough for use as solvent in the subsequent reaction.

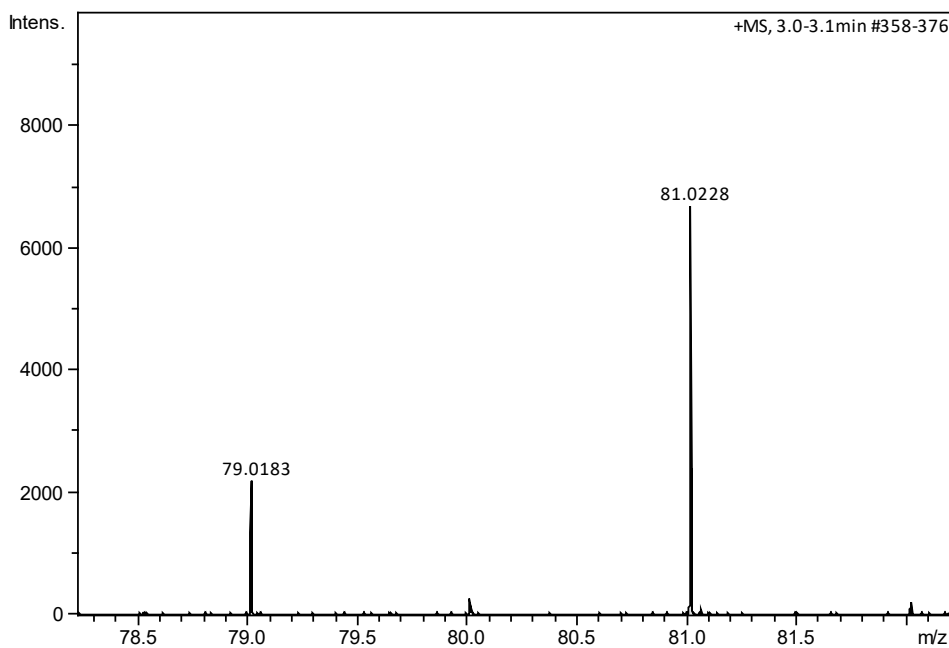


Figure S33: MS spectrum of the isotopically labelled DMSO, from the peaks intensity is it visible a 3:1 ratio between ^{18}O -DMSO and ^{16}O -DMSO.

5.5.4.2 Product 25 – ^{18}O -DMSO Labelled

The reaction procedure followed the standard conditions presented in 3.1. 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.

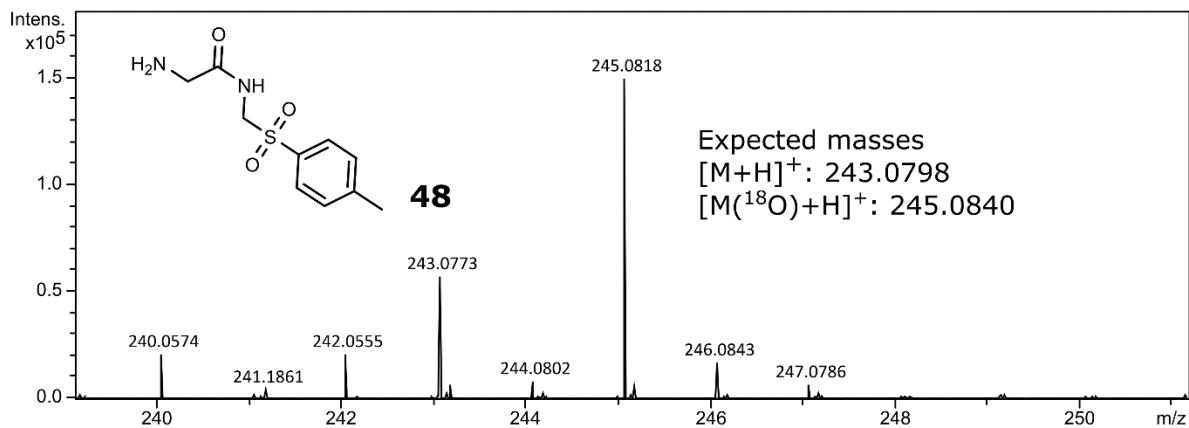
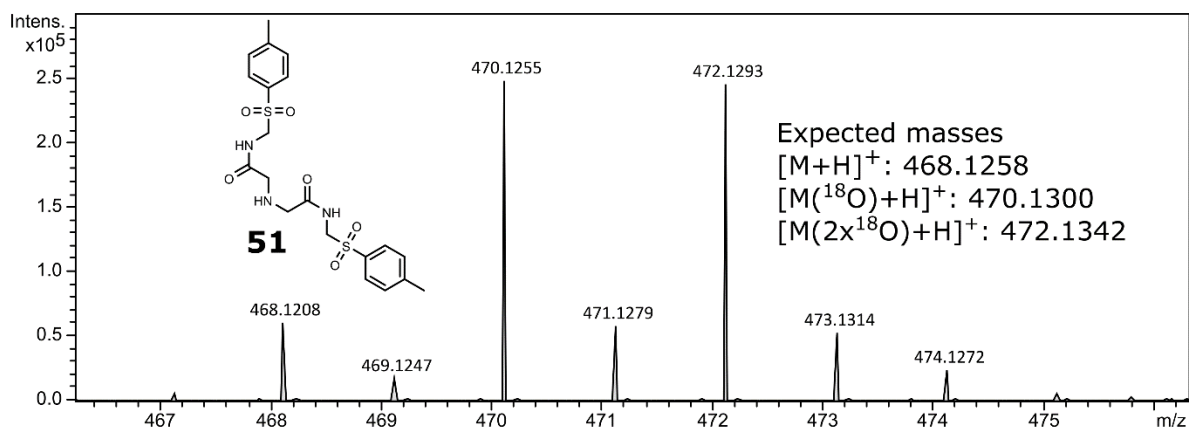
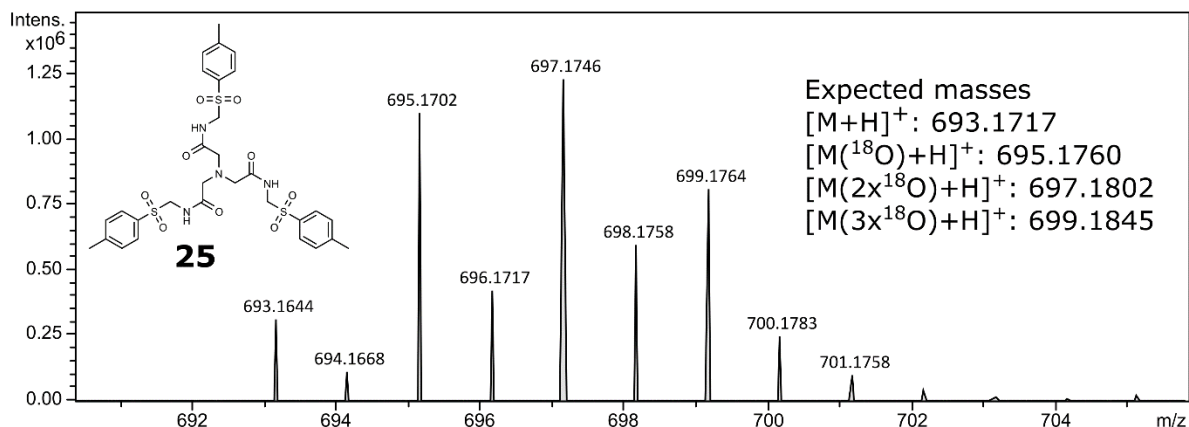


Figure S34: MS spectrum of products **25**, **51** and **48** prepared in ^{18}O -DMSO. The isotopic pattern of all molecules shows incorporation in accordance with the 3:1 labelling of the solvent, suggesting that all the oxygens in the product come from DMSO.

5.5.5 Reaction in $^{18}\text{OH}_2$

The synthesis of the trimer (**25**) was undertaken according to the procedure reported in SI-3.1, 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 **25** and the two intermediates **48** and **51** (**Figure S35**). The most intense peaks

of all compounds correspond to the non-labelled versions, indicating that the oxygens do not come from the water. However, **25** and **51** 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.

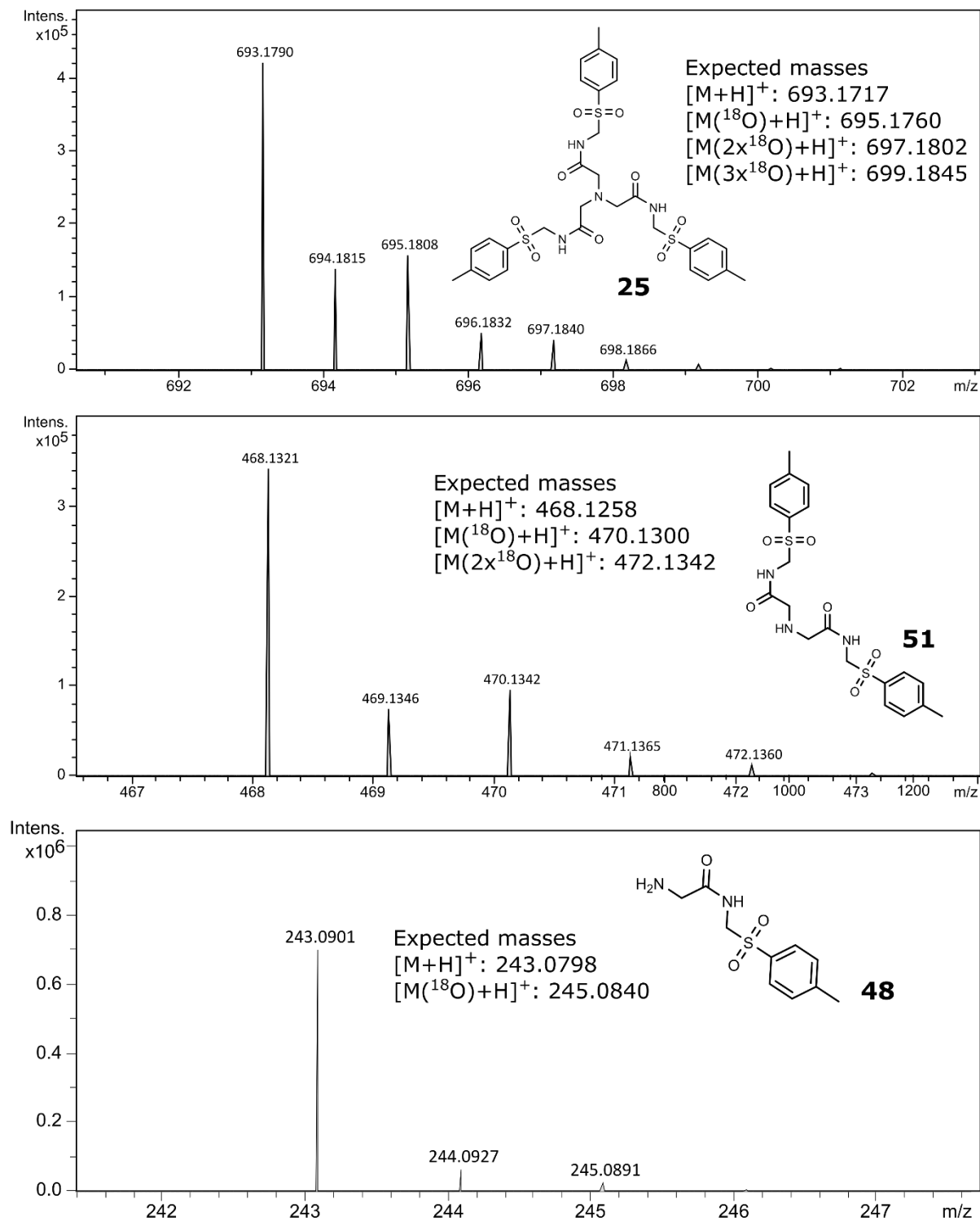


Figure S35: In order: mass isotope patterns of the product **25** and the intermediates **51** and **48**. The reaction was performed in presence of labelled water and shows no incorporation. The small increase in the +2 isotopes of **25** and **51** is explained with the imine equilibrium of the carbonyls.

5.6 IR reaction monitoring.

The reaction prepared according to the standard recipe 3.1 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 S36** and **Figure S37**. The signal of the CN bond from TosMIC at 2150 cm^{-1} disappears after 10 h while the formation of isocyanate (10) (2257 cm^{-1}) and CO_2 (2349 cm^{-1}) are observed.

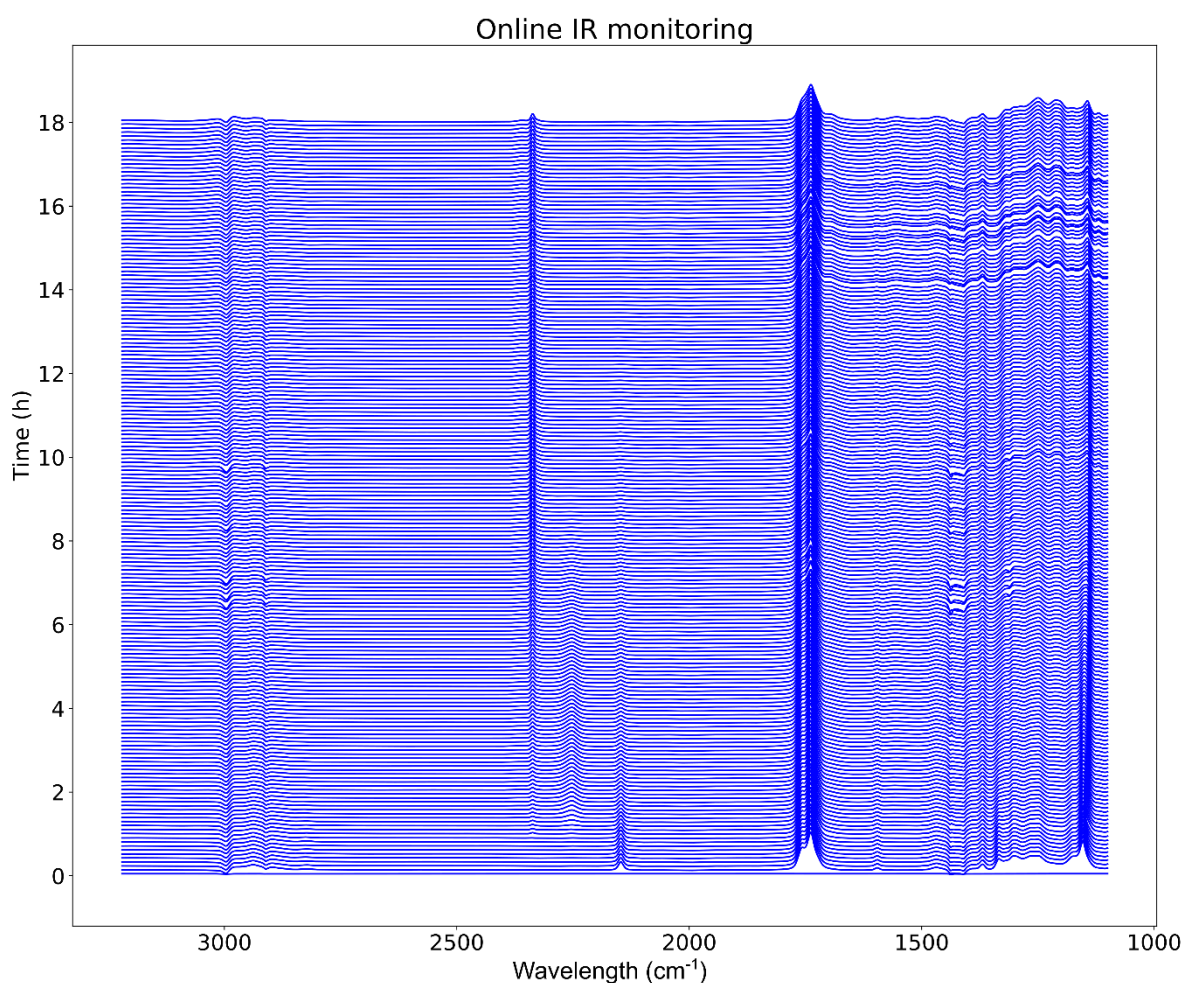


Figure S36: Online FT-IR analysis of the reaction mixture. A new spectrum is acquired every 5 minutes. For 18 hours, the wavelength are plotted vs the time of acquisition. The disappearing of TosMIC and appearing of an isocyanate bond and CO_2 are visible around 2250 cm^{-1} . Expansion of this area in **Figure S37**.

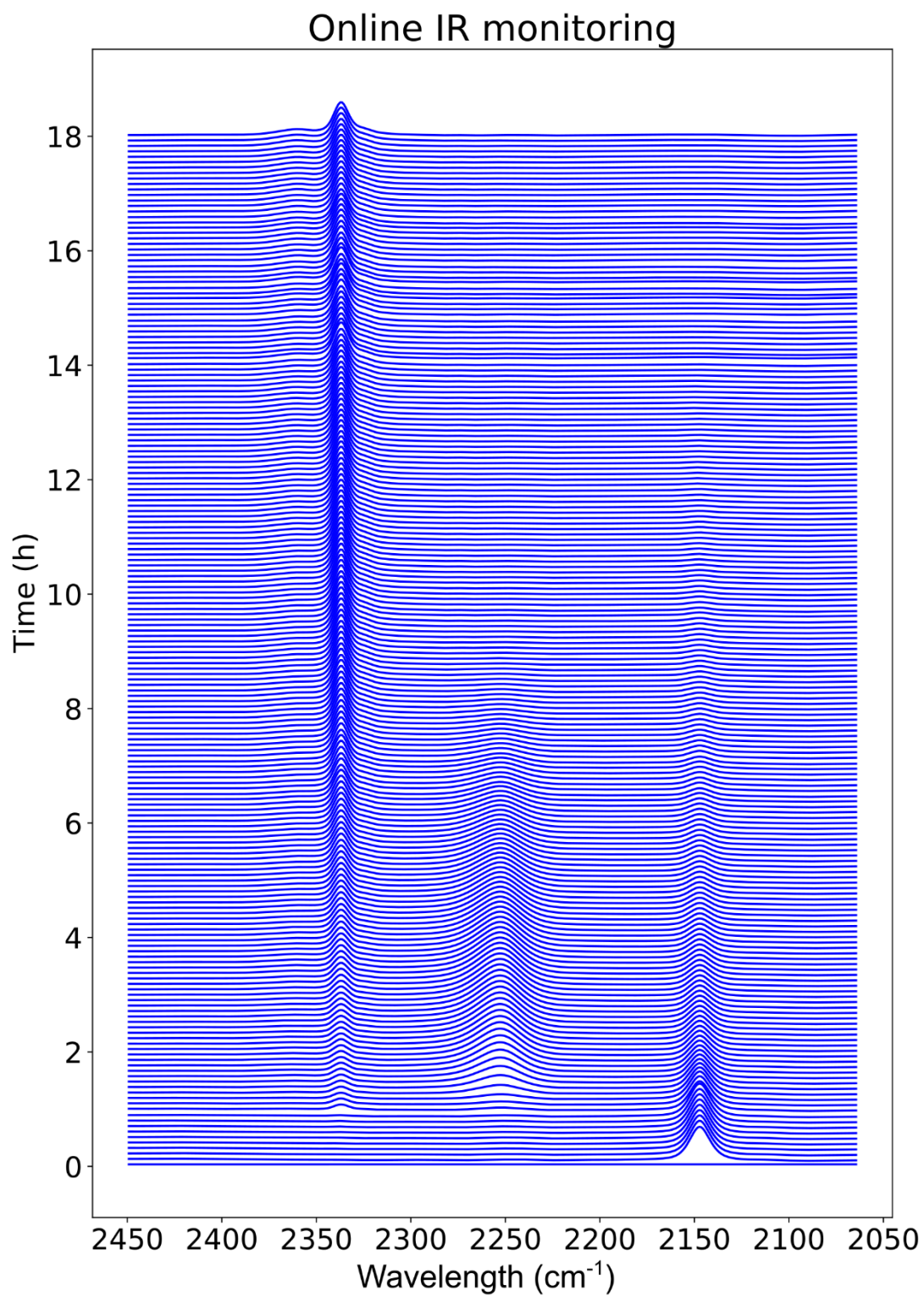


Figure S37: Detail of the online IR monitoring (**Figure S36**) showing the disappearing of the CN bond from TosMIC at 2150 cm^{-1} and the formation of isocyanate (2257 cm^{-1}) and CO_2 (2349 cm^{-1}).

5.7 Online NMR monitoring

Diethyl 2-bromomalonate (0.5 mmol, 0.1 ml) and p-toluenesulfonylmethyl isocyanide (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 (11) (**Figure S38**). The formation of dimethyl sulphide is also observed as a quintuplet at 2 ppm (12) (**Figure S39**). Data was processed using MestReNova software suite.

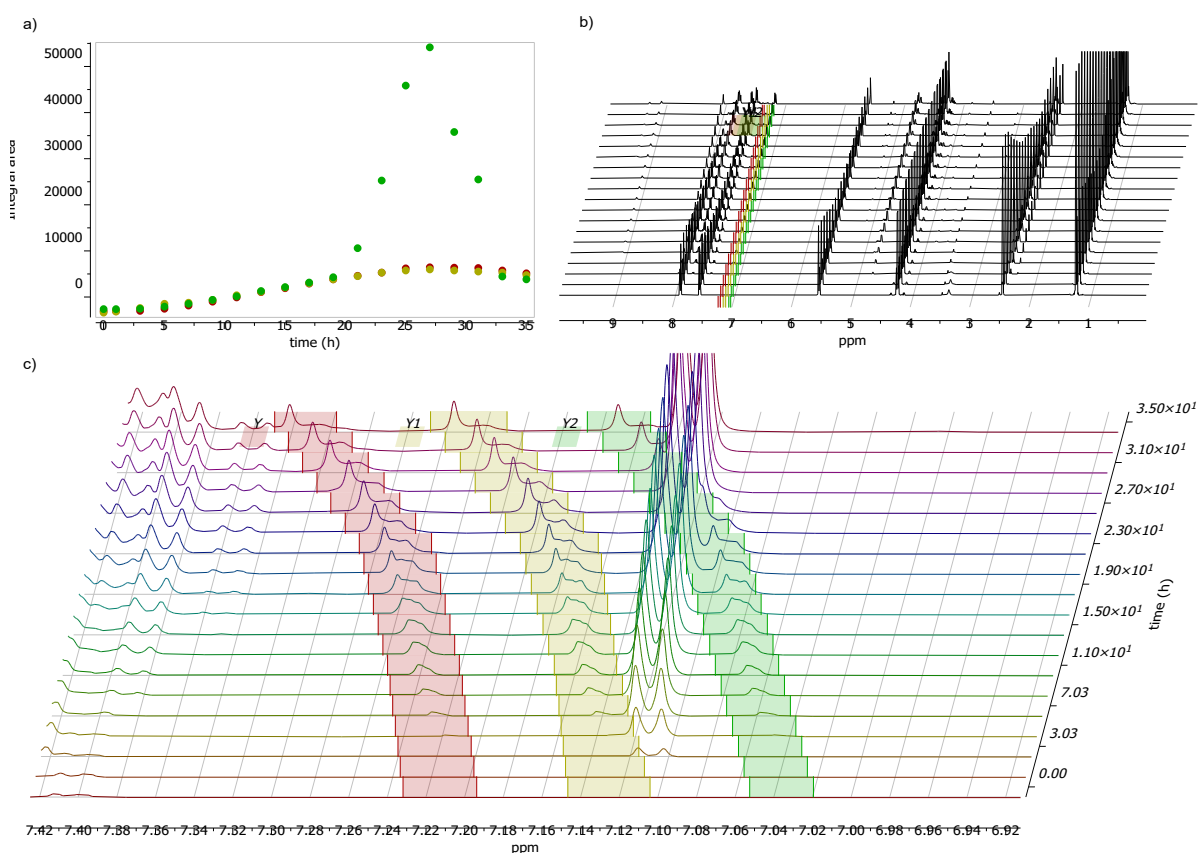


Figure S38: Online NMR monitoring of the reaction. The formation of the ammonium peaks is observed as three peaks (red, yellow and green) with 1:1:1 ratio around 7 ppm. a) Area of the three peaks over time. During the shifting peak Y2 partially overlaps with another product. b) full spectra. c) zoomed spectra showing the presence of ammonium peaks.

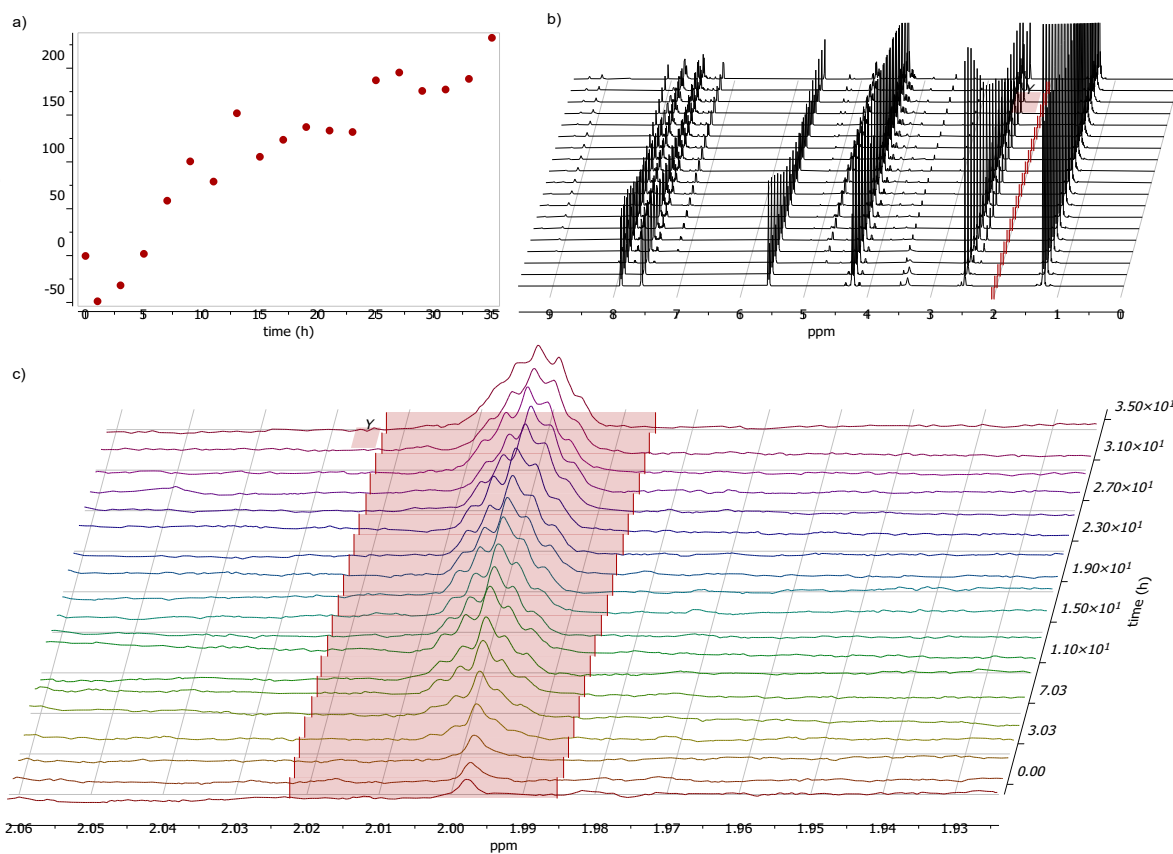


Figure S39: Online NMR monitoring of the reaction. The formation of dimethyl sulfide is observed at 2 ppm. a) Area of the peak over time. b) full spectra. c) zoomed spectra showing the presence of dimethyl sulfide peaks.

6 Cheminformatics simulation

6.1 Reaction network

A list of 21 reaction templates was encoded as SMARTS (**Figure S40**). 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 **25**. The two resulting networks were then merged by keeping only the molecules simultaneously connected to TosMIC and **25**. 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 S40: List of reaction SMARTS used to create the chemical space network.

6.2 Similarity index with known TosMIC reactions

1656 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 S41**. They are variations of polymerization reactions and Ugi four-component reactions.

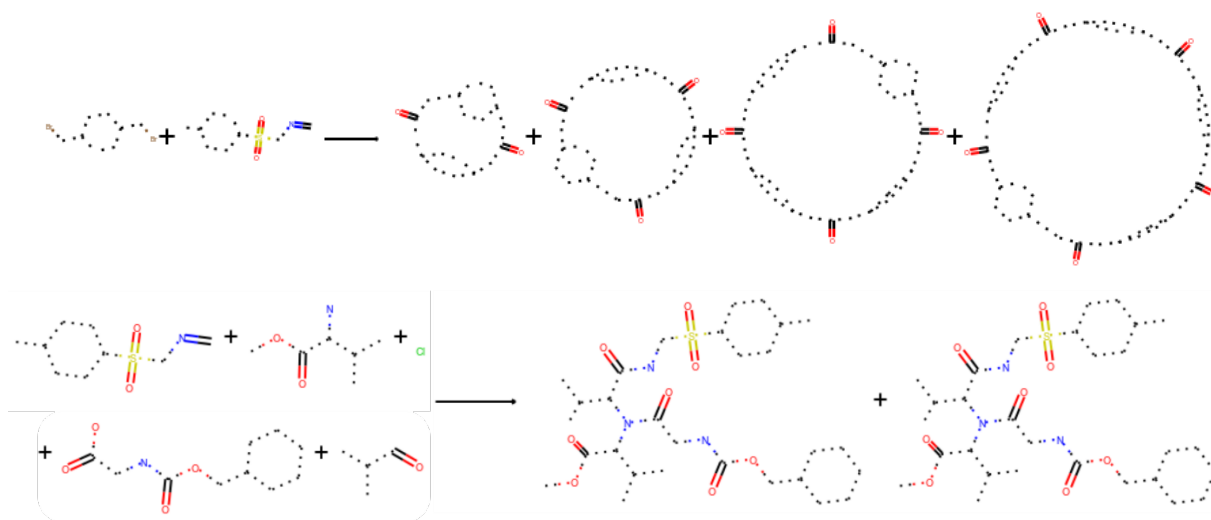


Figure S41: Templates of the two main types of reactions presenting a higher difference generation.

Next, the dataset has also been used to compare the discovered reaction with known TosMIC reactions using the Tanimoto similarity index (13). A matrix of cross correlation between the reaction fingerprints is showed in **Figure S42**. The Tanimoto index between reaction leading to product **25** 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 S43**. As reference the reactions with highest similarity to the discovered reaction are showed in **Figure S44**. The script used to perform these calculations was written in Python using the RDKit library.

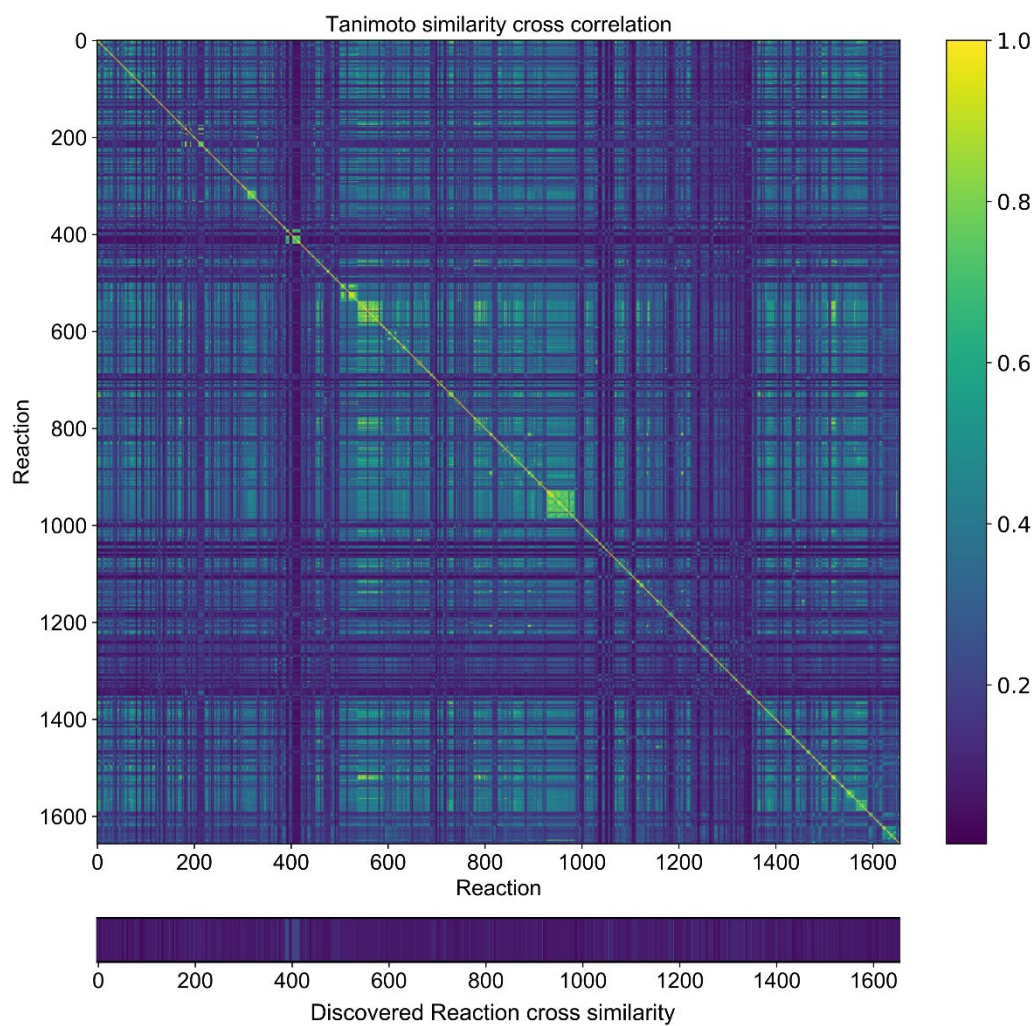


Figure S42: Tanimoto similarity index between reaction fingerprints of the literature reactions of TosMIC. In the bottom the similarity index between literature reactions and discovered reaction giving product **25**.

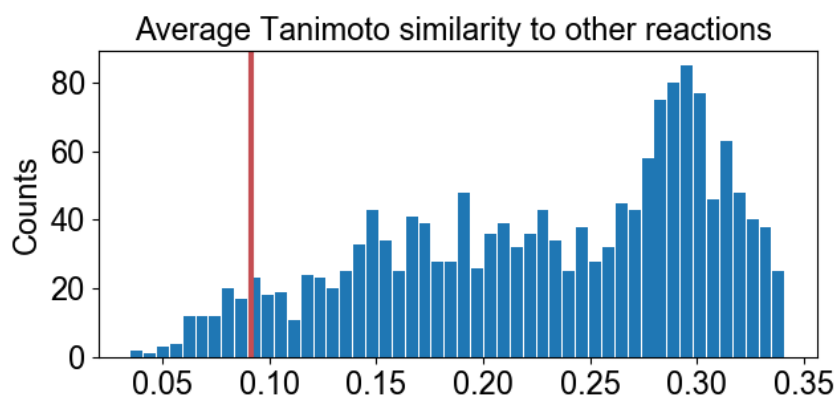


Figure S43: Histogram comparing the average similarity for each reaction to other reactions in **Figure S42**.

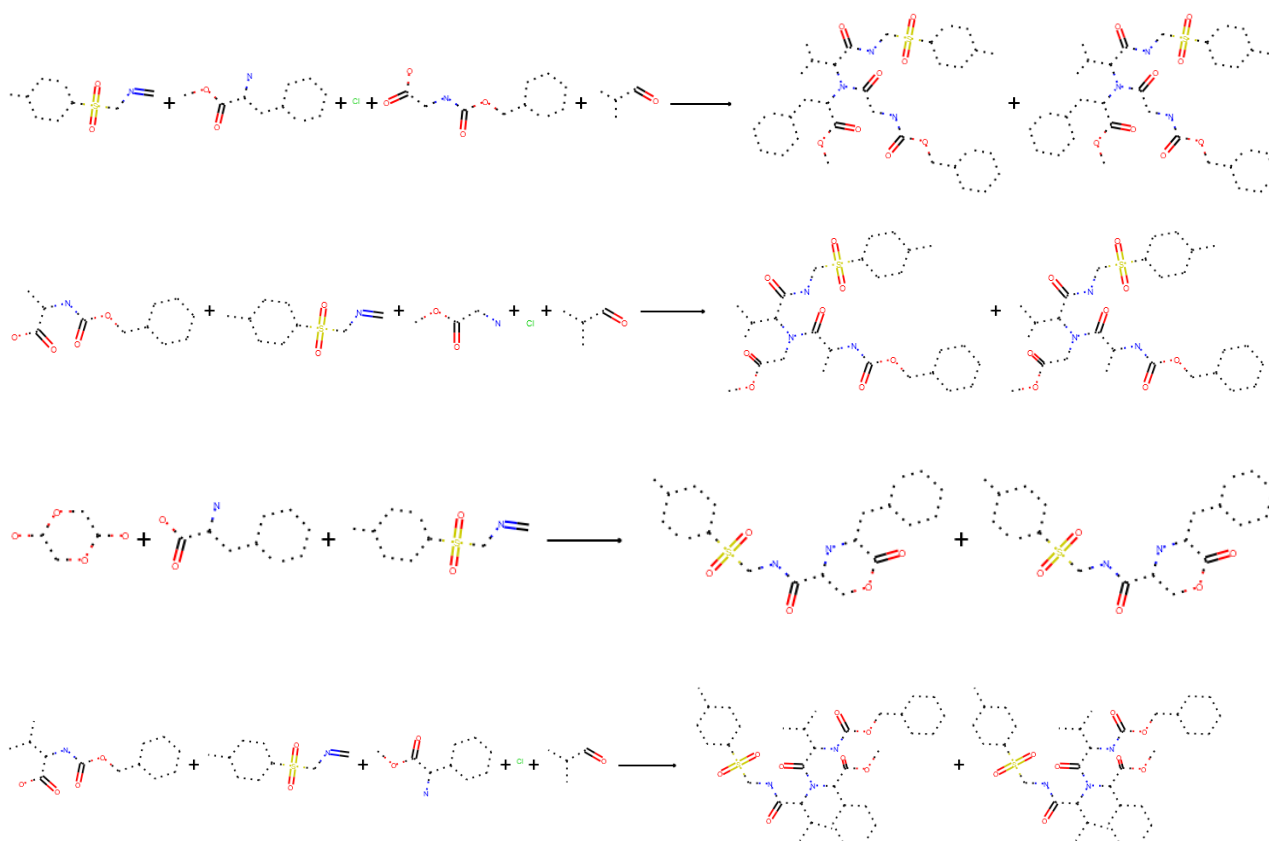


Figure S44: Reactions having the most similar reaction fingerprints to discovered reaction, according to Tanimoto similarity index.

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

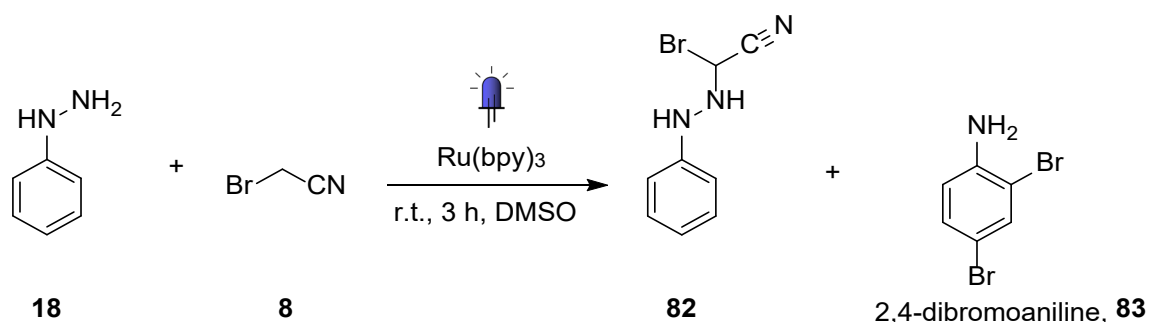


Figure S45: Scheme of the reaction between phenylhydrazine and bromoacetonitrile, under 450 nm irradiation and the Ru(bpy)₃ photocatalyst. The first product is unreported in literature.

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. Mg₂SO₄ 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 **82** and 2,4-dibromoaniline **83**. **82** 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 **82**

Yield: 7% (31 mg)

IUPAC name: Bromo-(N'-phenyl-hydrazino)-acetonitrile

¹H NMR (600 MHz, CDCl₃) 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)

¹³C NMR (151 MHz, CDCl₃) δ 141.22, 129.14, 123.22, 113.77, 110.68, 99.62.

Crystallographic data: See section 7

2,4-dibromoaniline, **83**

Yield: 11% (55 mg)

¹H NMR (600 MHz, CDCl₃) 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

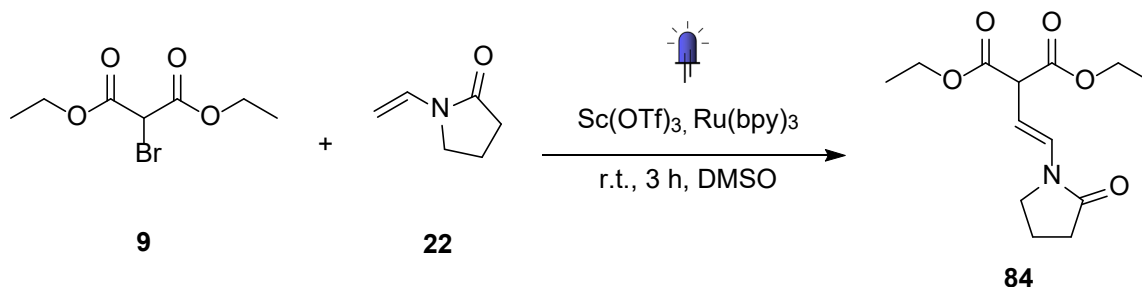


Figure S46: Scheme of the reaction between diethyl 2-bromomalonate and 1-vinyl-2-pyrrolidinone under 450 nm irradiation and the Ru(bpy)₃ photocatalyst.

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. Mg₂SO₄ 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 (14) where the authors used the same reagents, [Ir(ppy)₂(dtbbpy)]PF₆ as photocatalyst, 2 equivalents of Na₂HPO₄ and acetonitrile as solvent.

Yield: 53% (285 mg)

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

¹H NMR (600 MHz, CDCl₃) 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: [C₁₃H₁₉NNaO₅]⁺ Calculated 292.1155 *m/z*, measured 292.1147 *m/z*

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

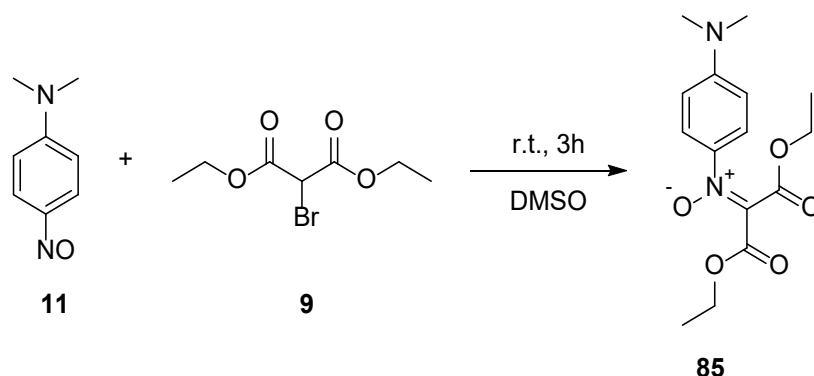


Figure S47: Reaction reported between *N,N*-dimethyl-4-nitrosoaniline and diethyl-2-bromomalonate.

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 nitron. It is already known in literature (15) 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 (16).

Yield: 18% (111 mg)

IUPAC name: *N*-(4-(dimethylamino)phenyl)-*C,C*-diethoxy-carbonylnitron

^1H 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: $[\text{C}_{15}\text{H}_{20}\text{N}_2\text{NaO}_5]^+$ Calculated 331.1264 m/z , measured 331.1248 m/z .

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

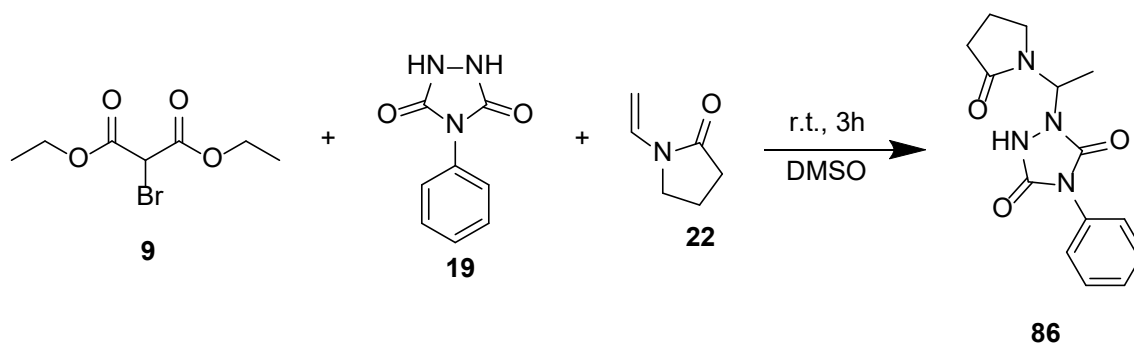


Figure S48: Scheme of the reaction between 4-phenylurazole, 1-vinyl-2-pyrrolidinone, and diethyl-2-bromomalonate.

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 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 (17) *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

1H 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)

^{13}C 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 7

8 Crystal structure details

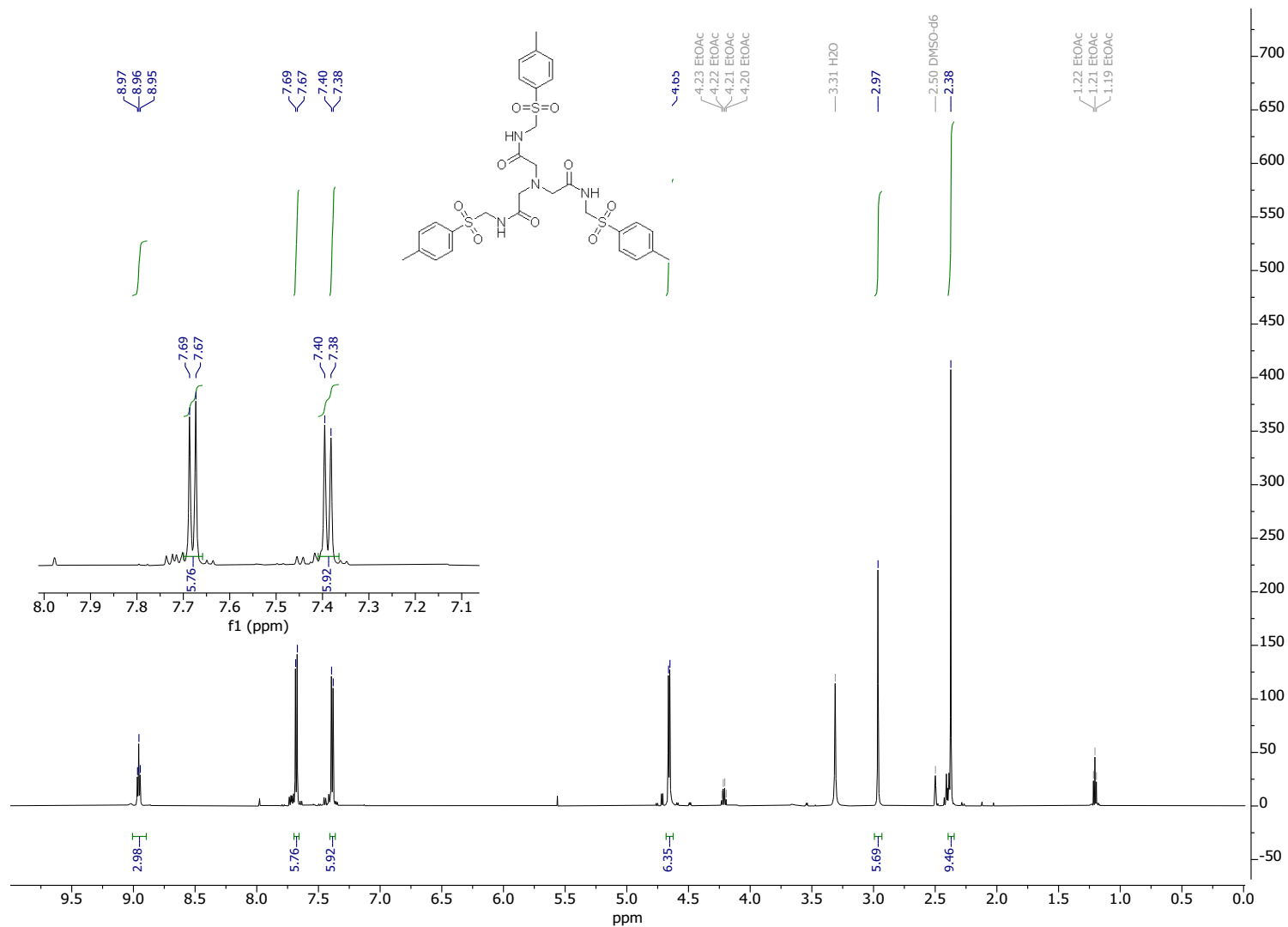
Single Crystal X-ray Diffraction: Suitable single crystals were selected and mounted by using the MiTeGen MicroMounts™ kit with Fomblin oil. X-ray diffraction intensity data were measured at 150(2) K on Bruker Apex II Quasar diffractometer using Mo K α [$\lambda = 0.71073$ Å] radiation. Structure solution and refinement were carried out with SHELXT (18) and SHELXL-2018 (19) via WinGX (20). Corrections for incident and diffracted beam absorption effects were applied using empirical methods (21).

Name	Compound 25
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 $\alpha = 90^\circ$ b = 22.260(8) Å $\beta = 90^\circ$ c = 9.328(4) Å $\gamma = 120^\circ$
Volume	4003(3) Å ³
Z	4
Density (calculated)	1.214 Mg/m ³
Absorption coefficient	0.263 mm ⁻¹
F(000)	1540
Crystal size	0.232 x 0.030 x 0.016 mm ³
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 F ²
Data / restraints / parameters	1937 / 0 / 140
Goodness-of-fit on F ²	1.043
Final R indices [$I > 2\sigma(I)$]	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

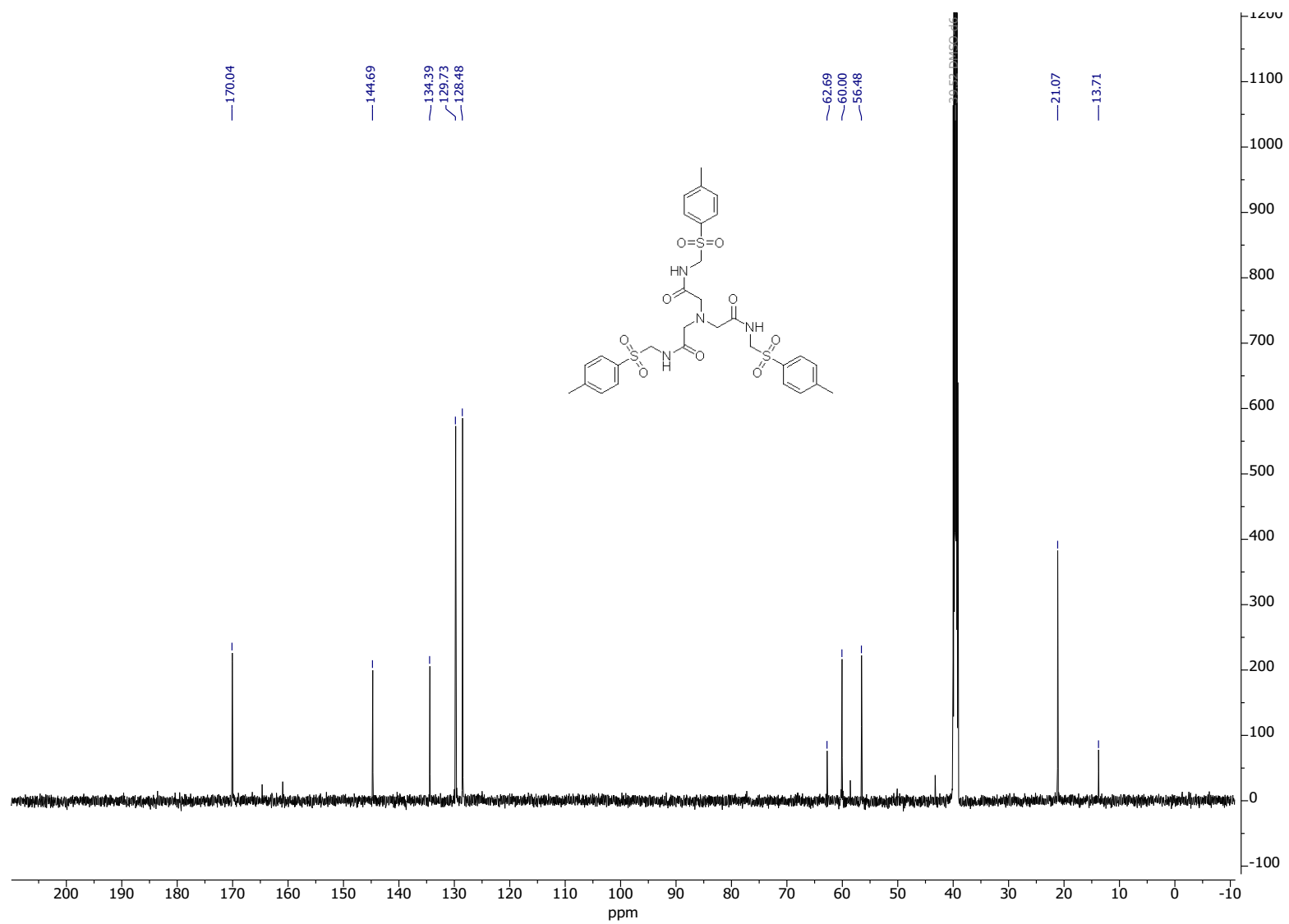
Name	Compound 82	
Empirical formula	C ₈ H ₅ BrN ₃	
Formula weight	223.06	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	Pna21	
Unit cell dimensions	a = 12.777(4) Å	α = 90°
	b = 4.5611(14) Å	β = 90°
	c = 14.595(5) Å	γ = 90°
Volume	850.6(5) Å ³	
Z	4	
Density (calculated)	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 > 2σ(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 86
Empirical formula	C ₁₄ H ₁₆ N ₄ O 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 > 2σ(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

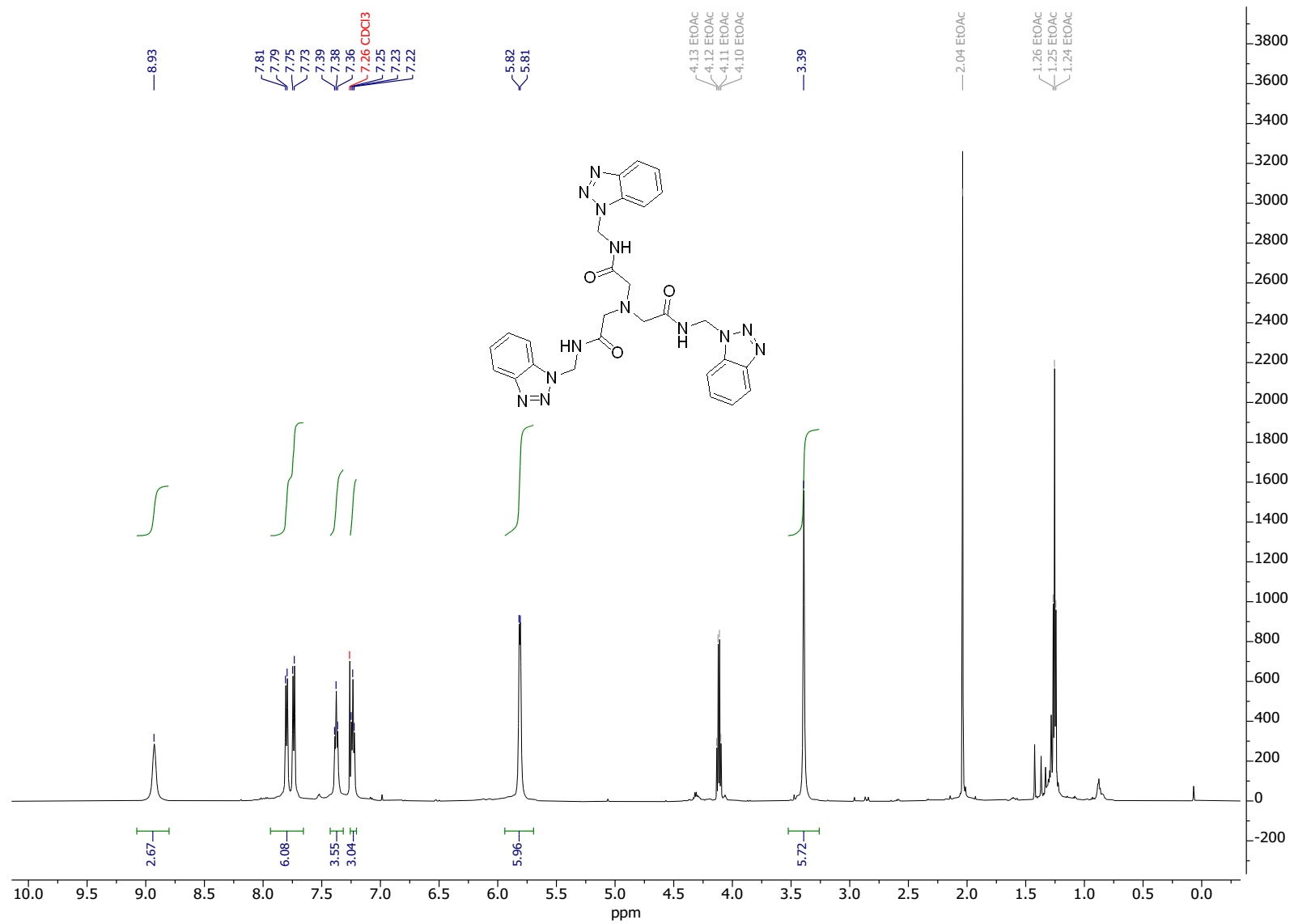
9 ¹H and ¹³C NMR spectra



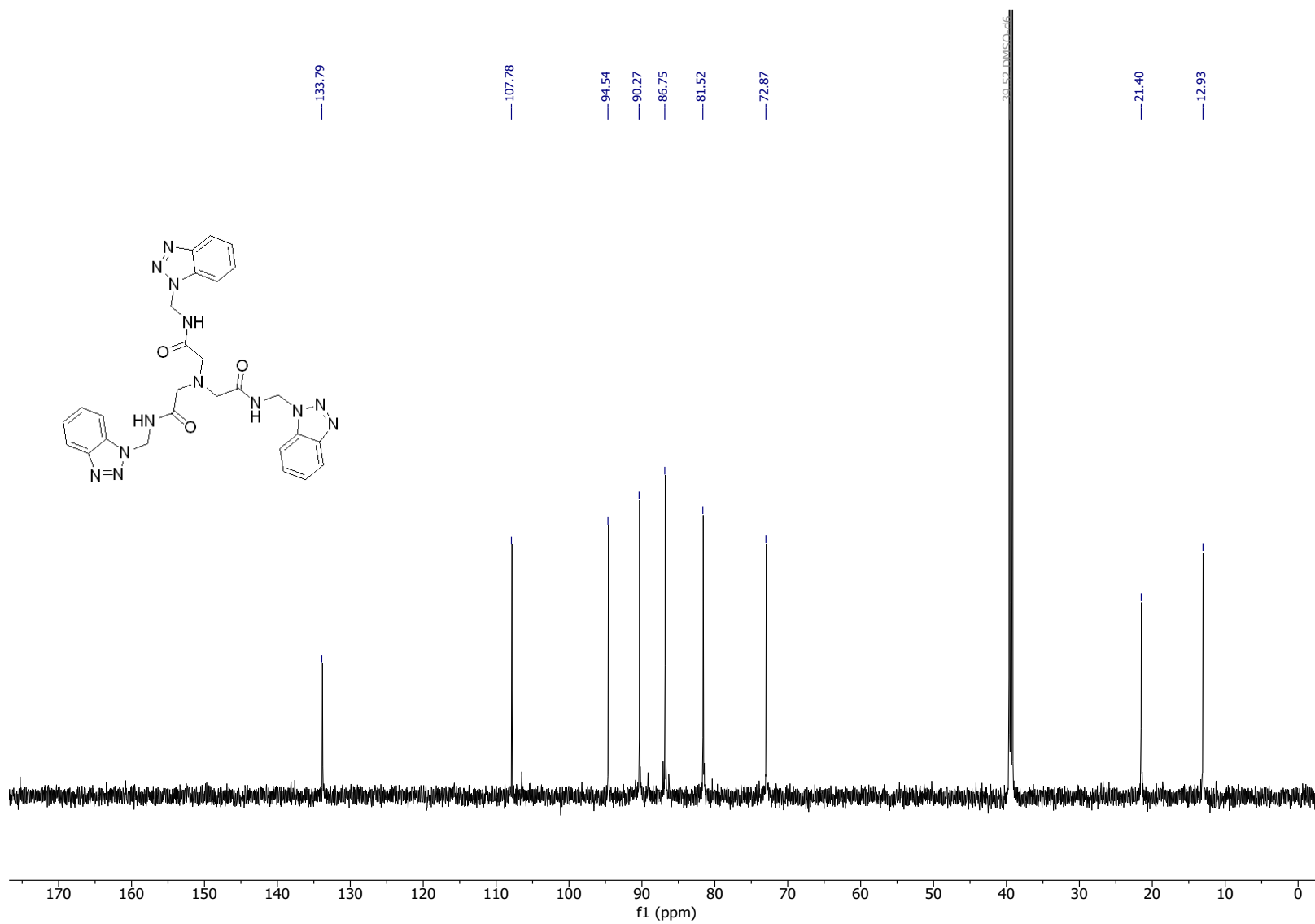
Spectrum S1: ¹H-NMR of Product 25



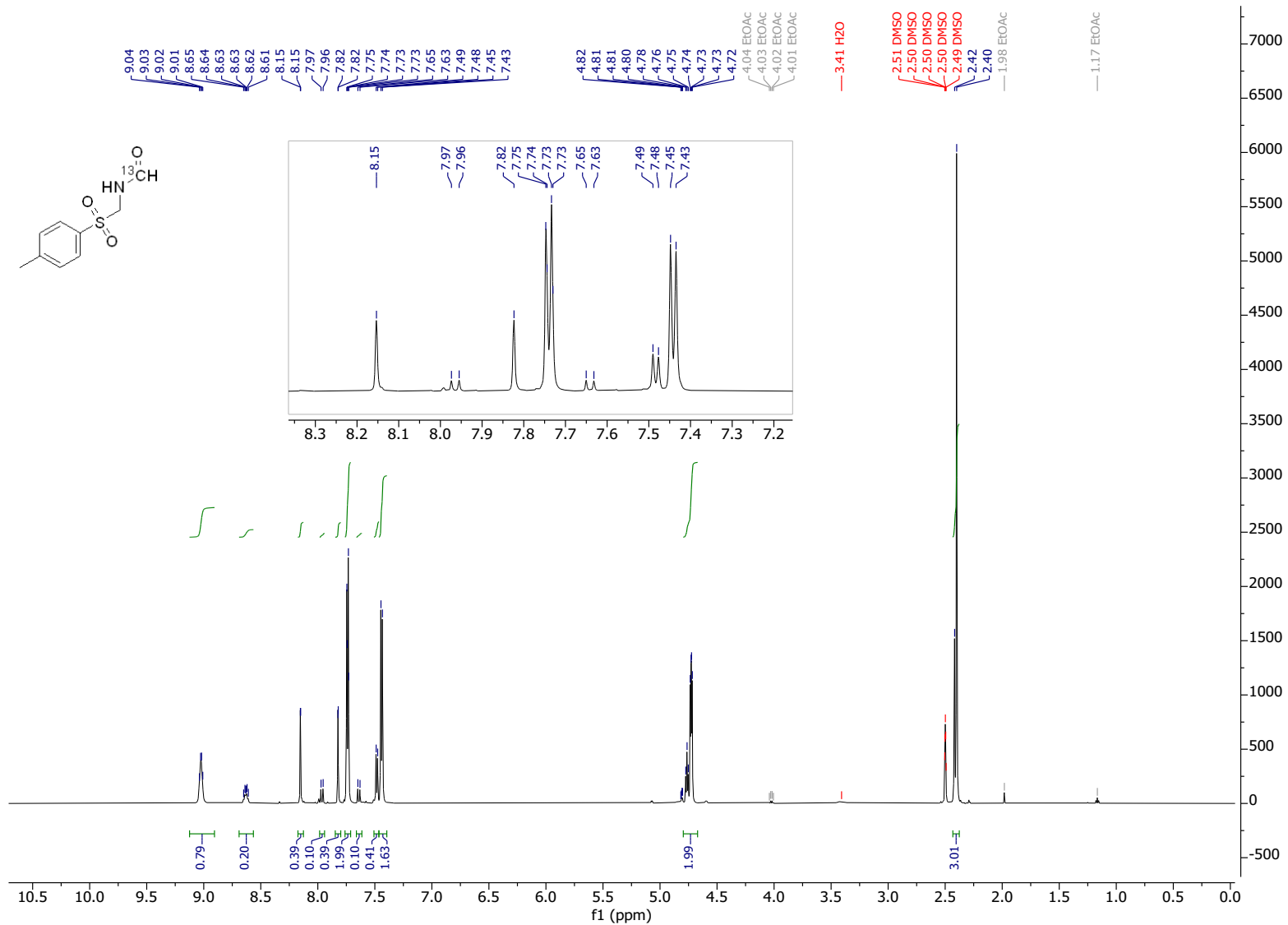
Spectrum S2: ¹³C-NMR of Product 25



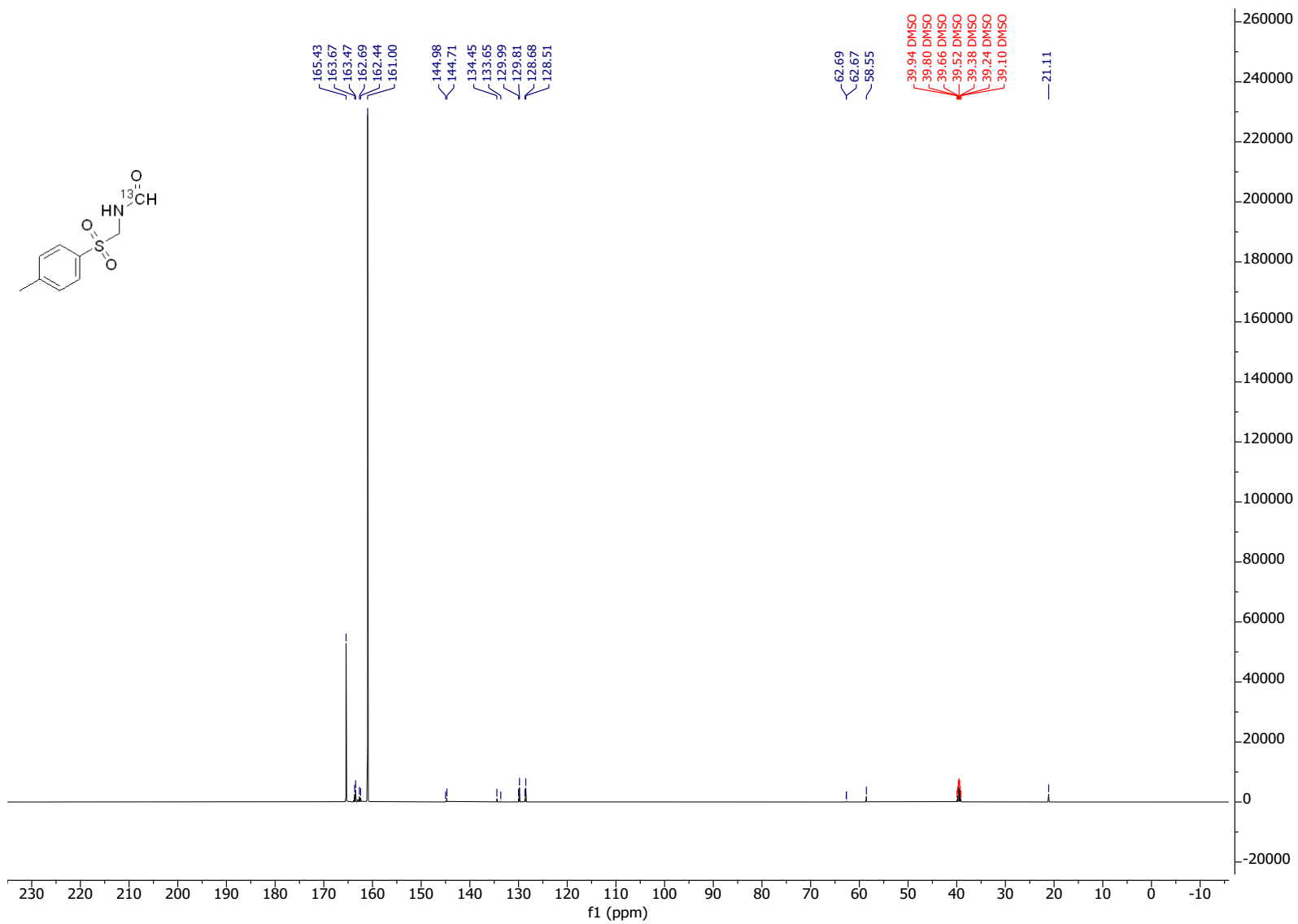
Spectrum S3: ¹H-NMR of Product 28



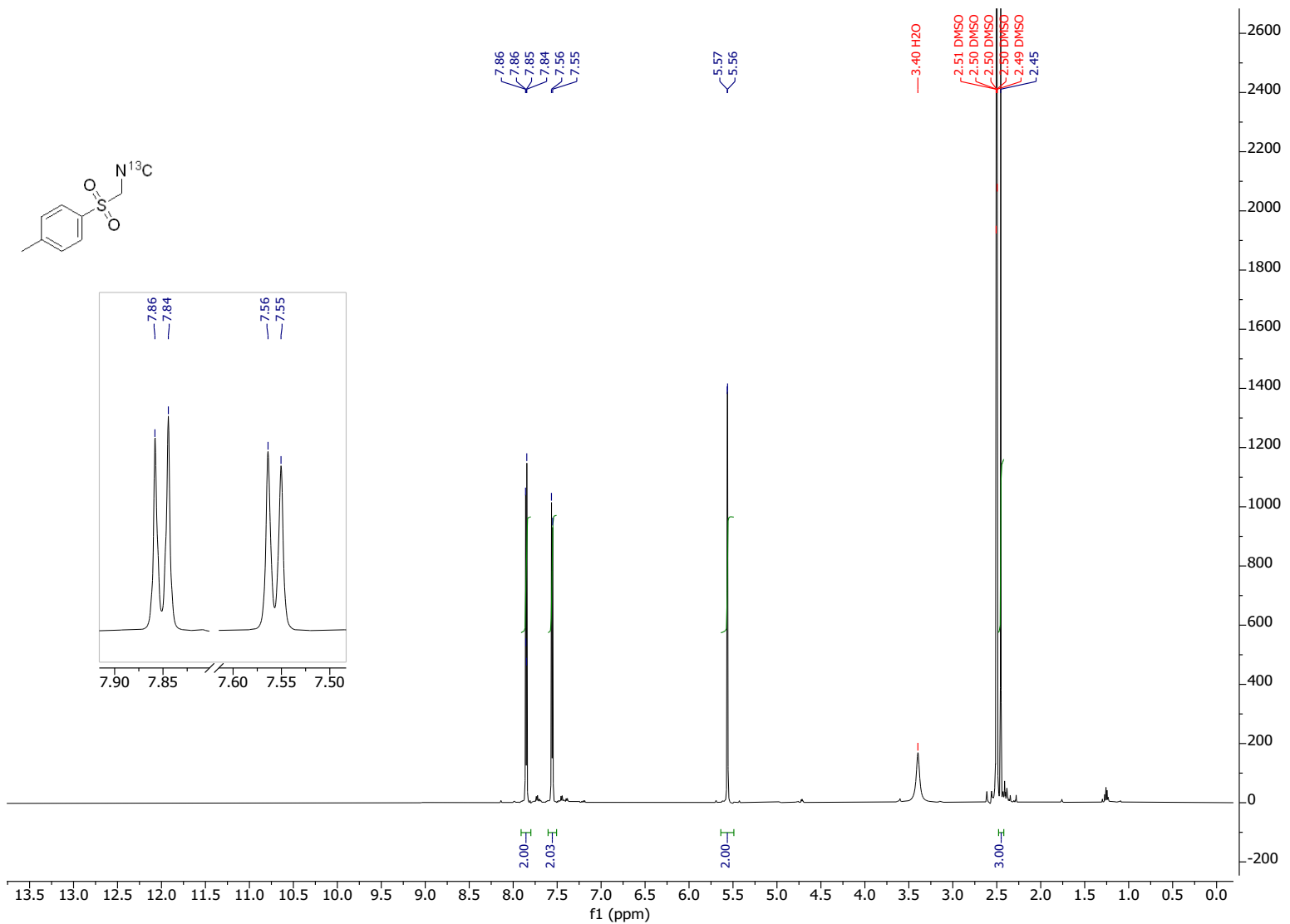
Spectrum S4: ¹³C-NMR of Product 28



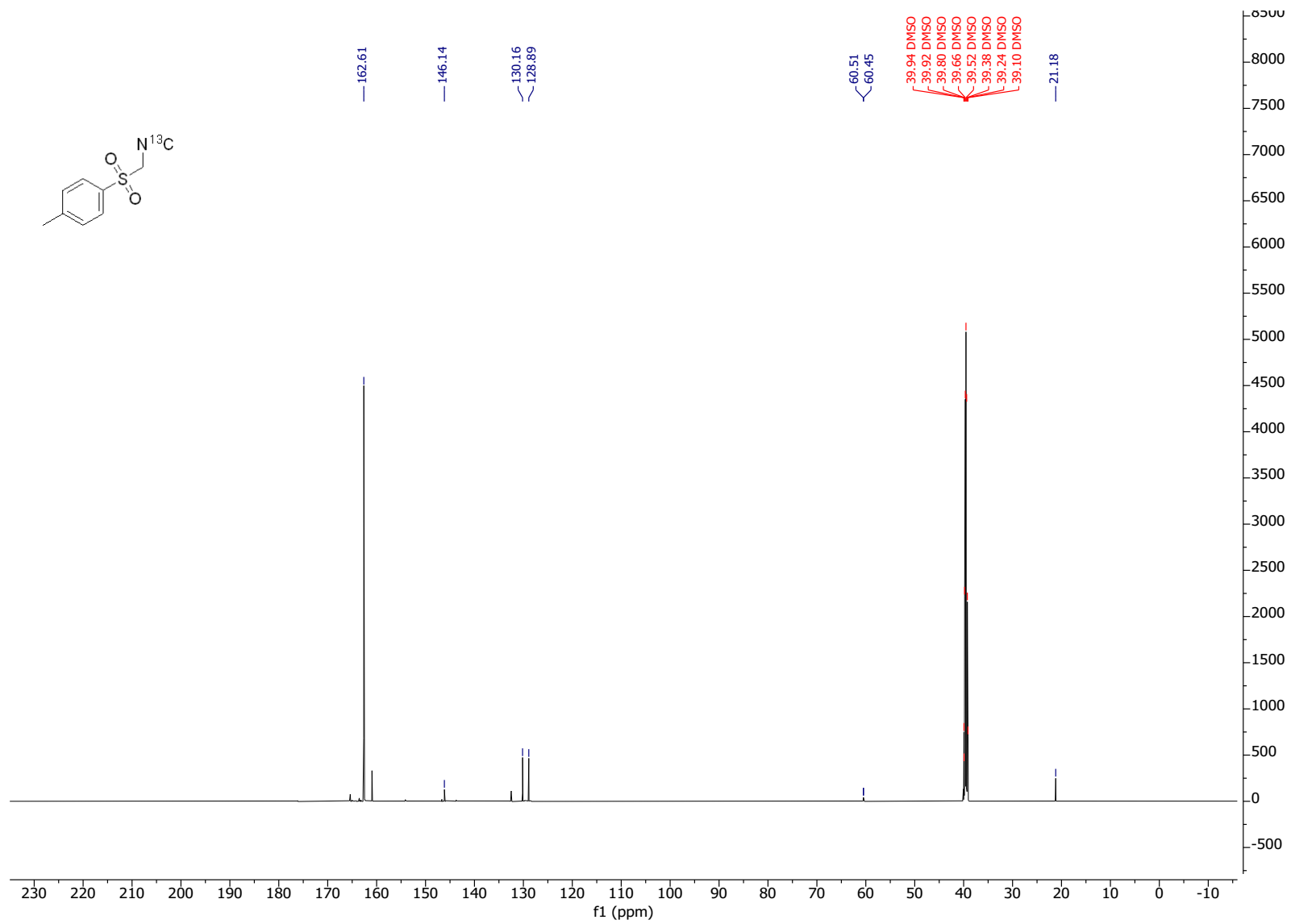
Spectrum S5: ¹H-NMR of (1-¹³C)N-(tosylmethyl)-formamide



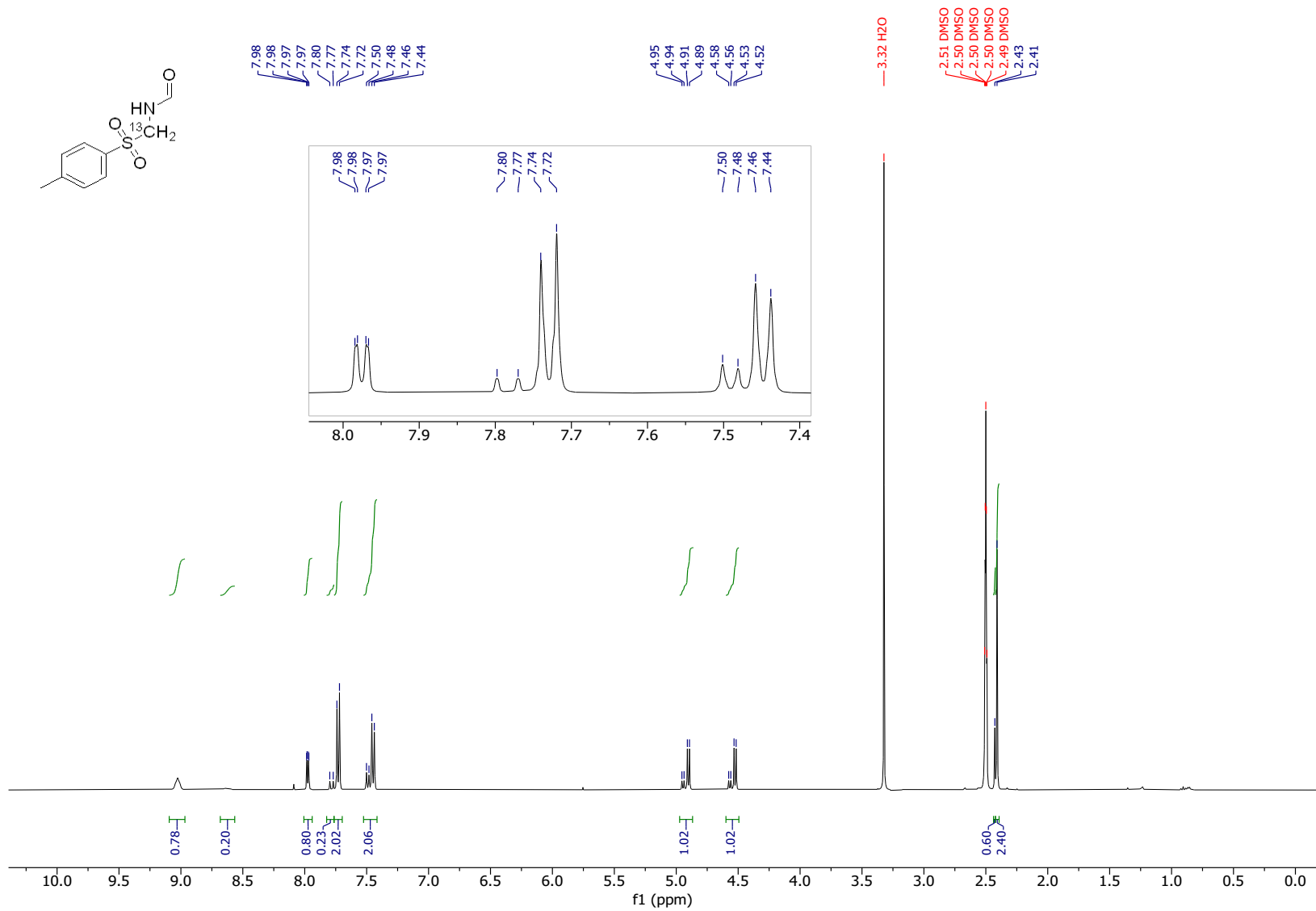
Spectrum S6: ¹³C-NMR of (1-¹³C)N-(tosylmethyl)-formamide



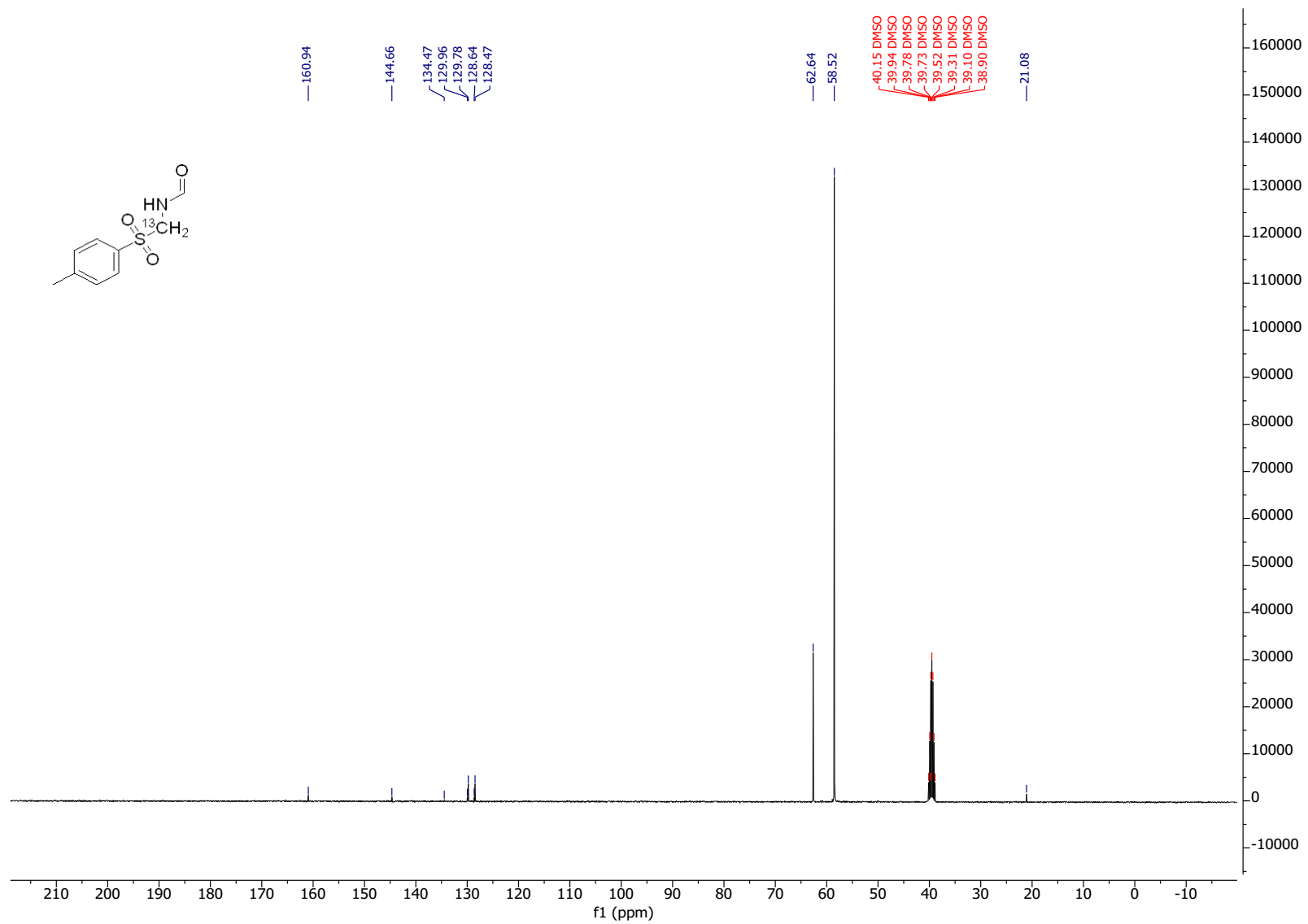
Spectrum S7: ¹H-NMR of TosCH₂N¹³C



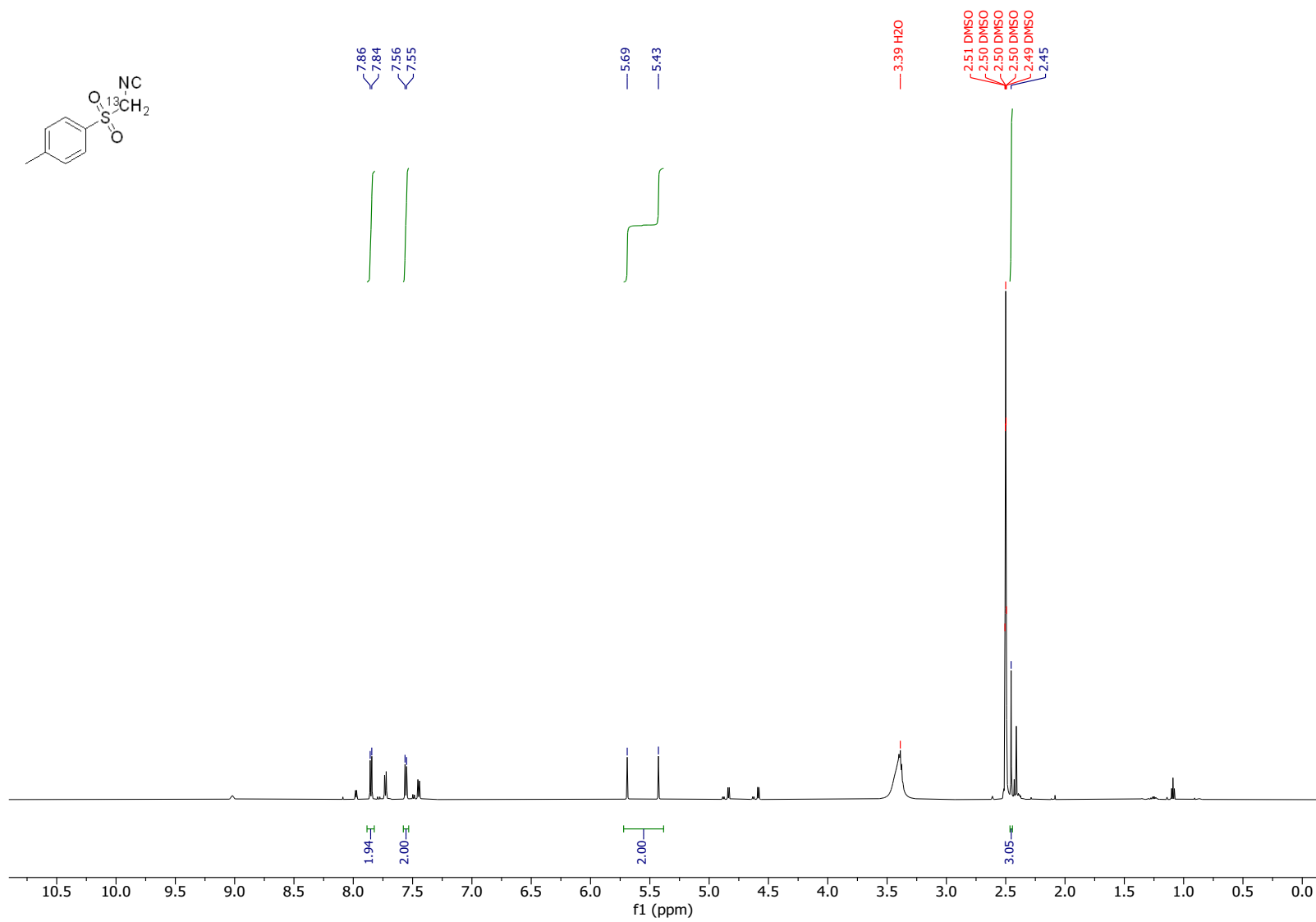
Spectrum S8: ^{13}C -NMR of $\text{TosCH}_2\text{N}^{13}\text{C}$



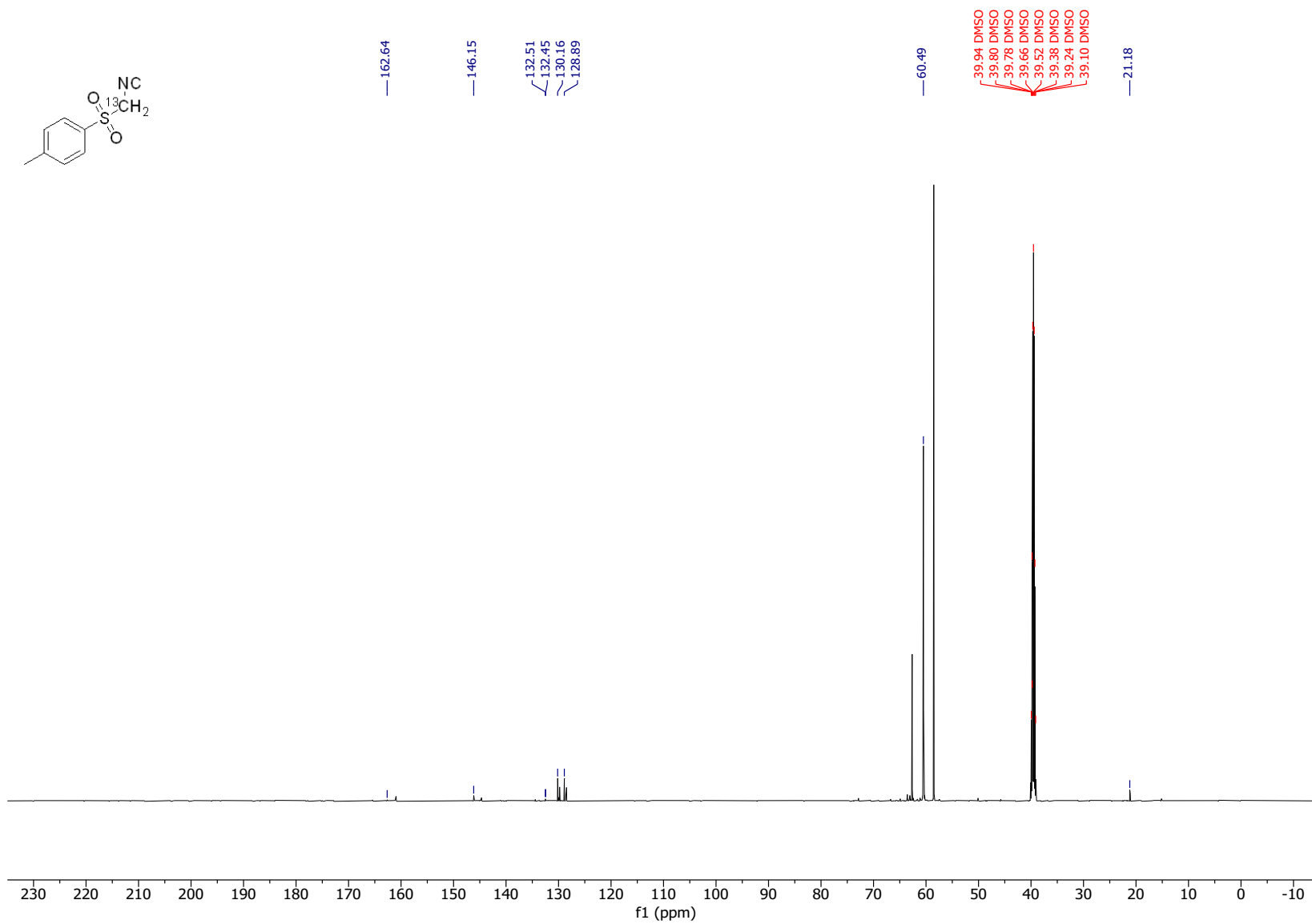
Spectrum S9: ¹H-NMR of N-(tosyl(¹³C)methyl)-formamide



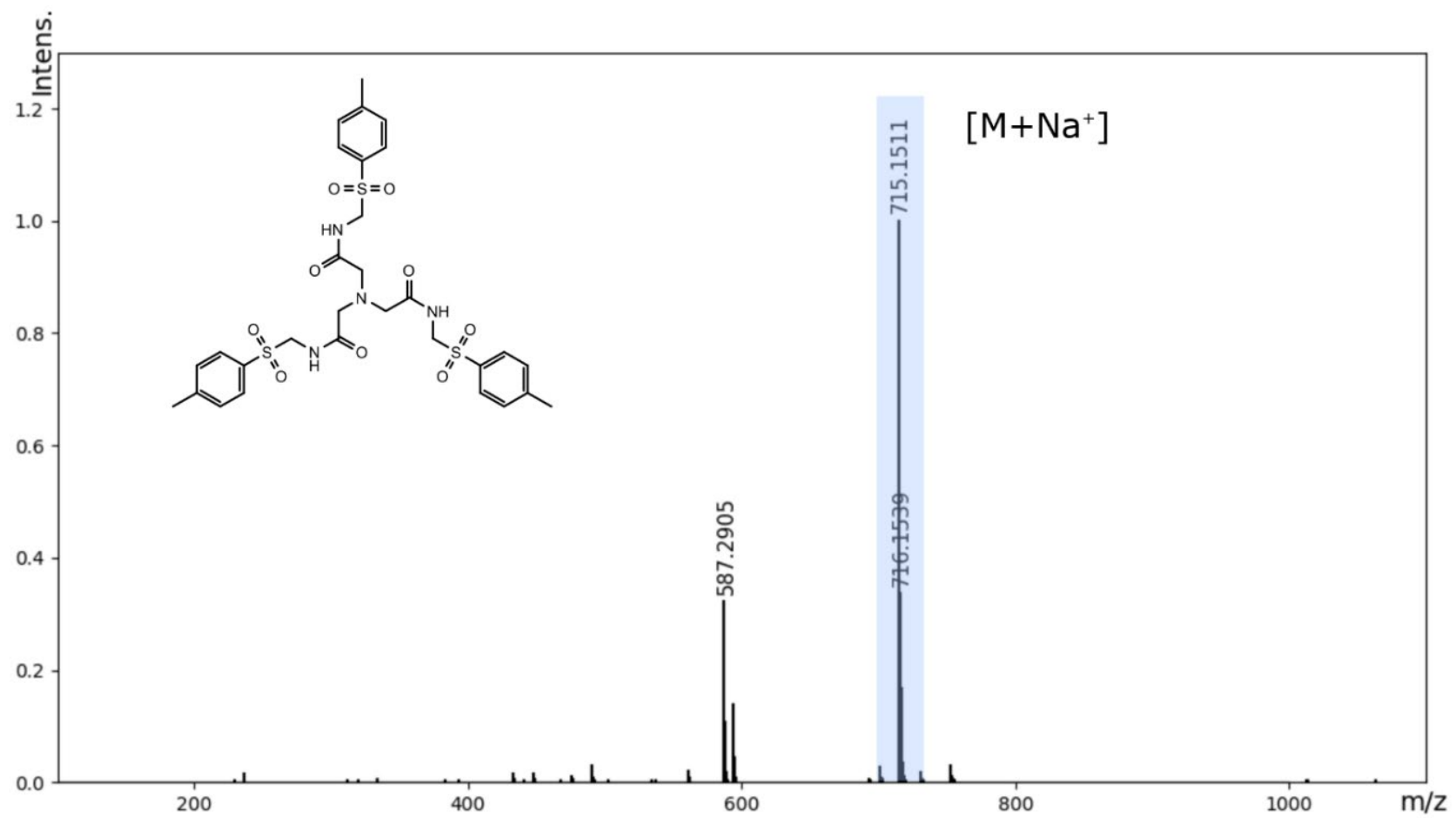
Spectrum S10: ¹³C-NMR of N-(tosyl(¹³C)methyl)-formamide



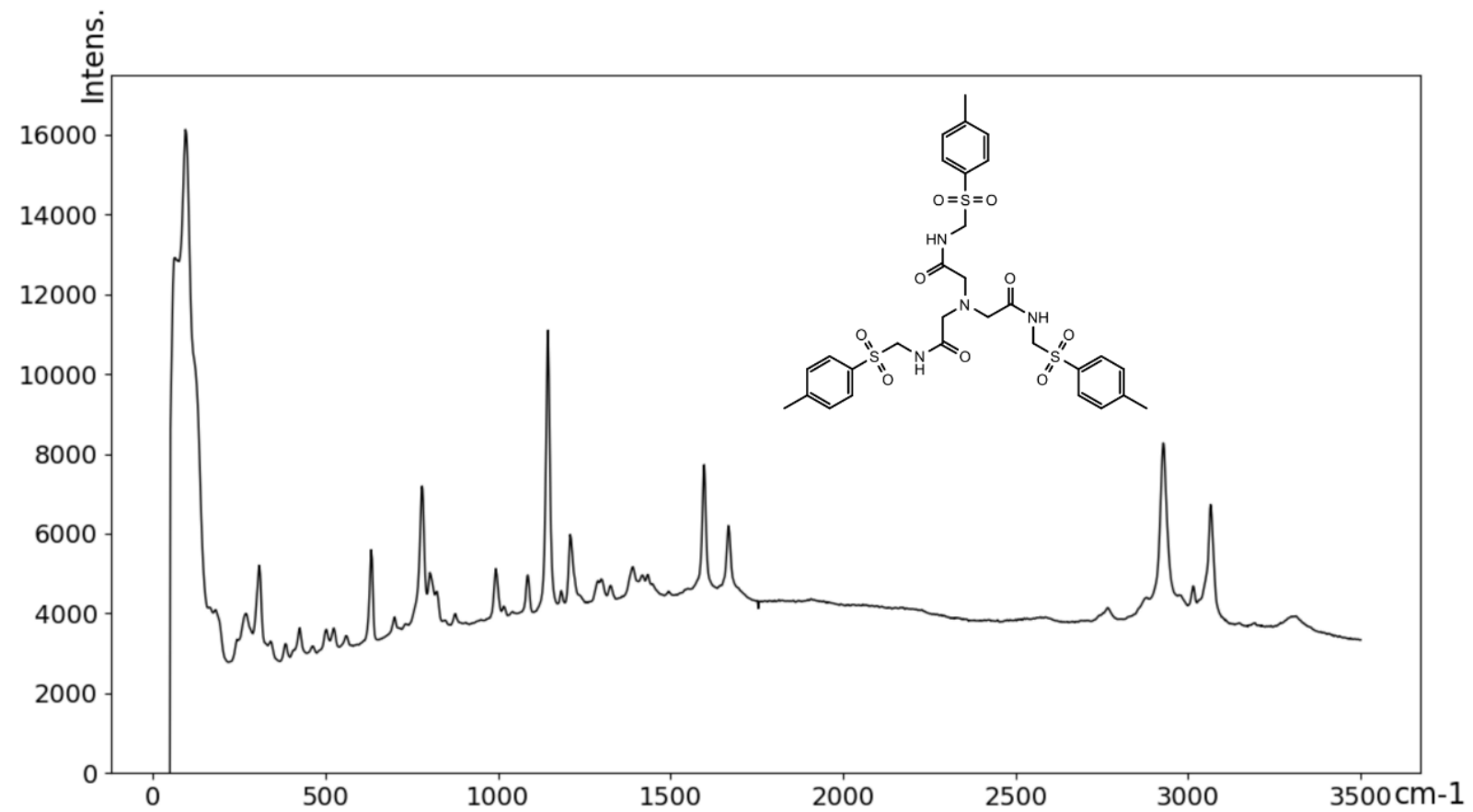
Spectrum S11: ¹H-NMR of Tos¹³CH₂NC



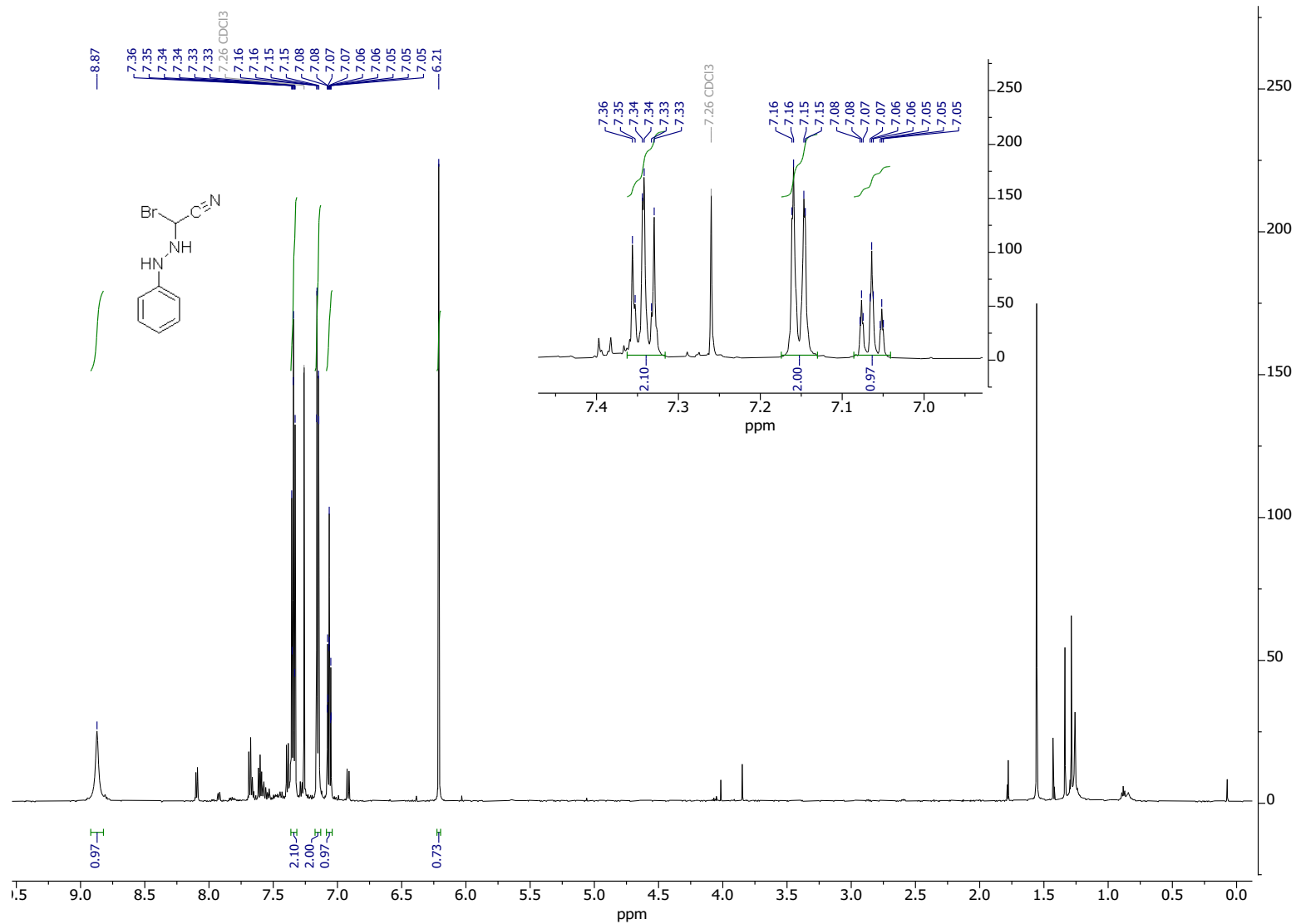
Spectrum S12: ¹³C-NMR of Tos¹³CH₂NC



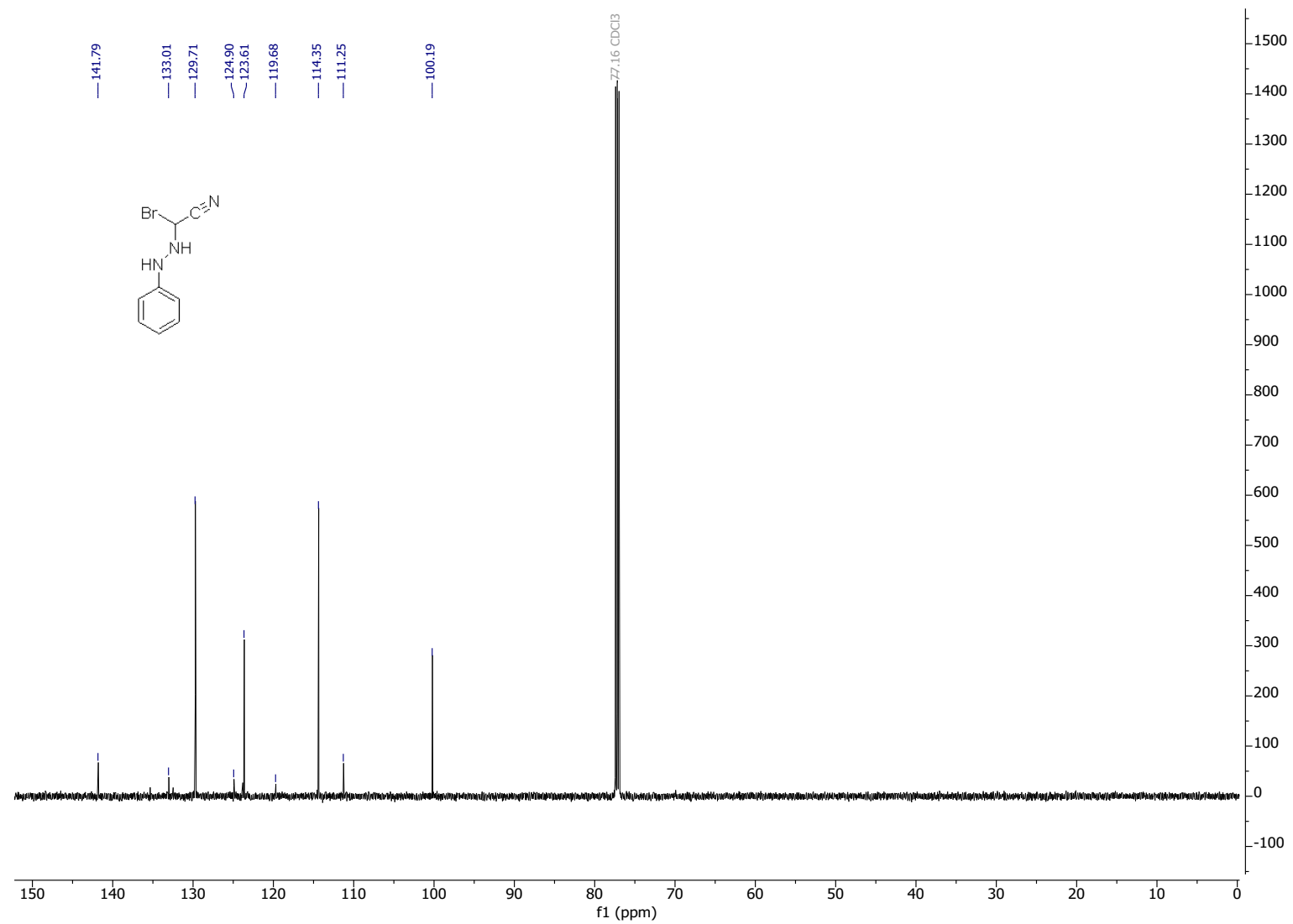
Spectrum S13: MS spectrum of compound 25



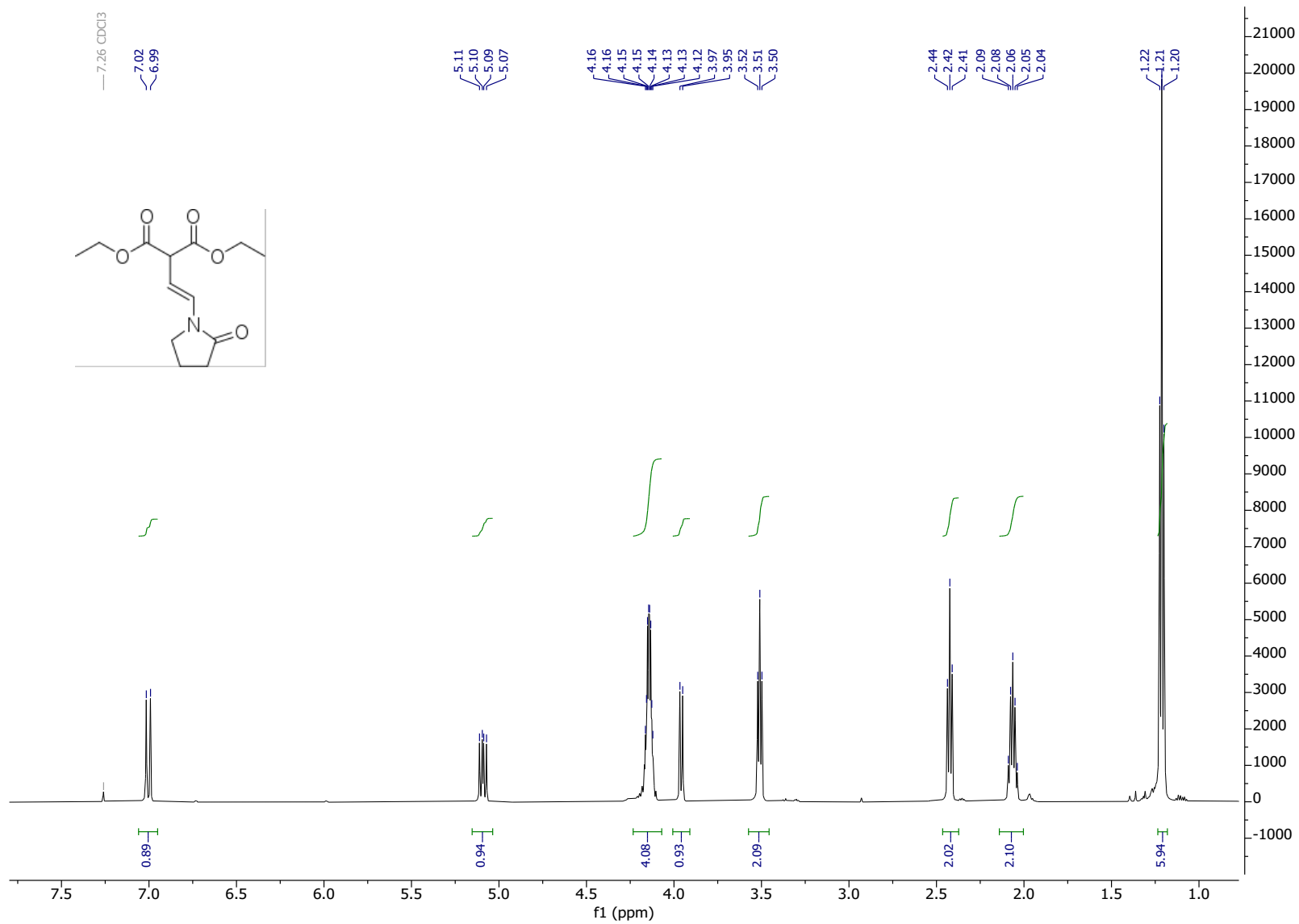
Spectrum S14: Raman spectrum of compound **25**



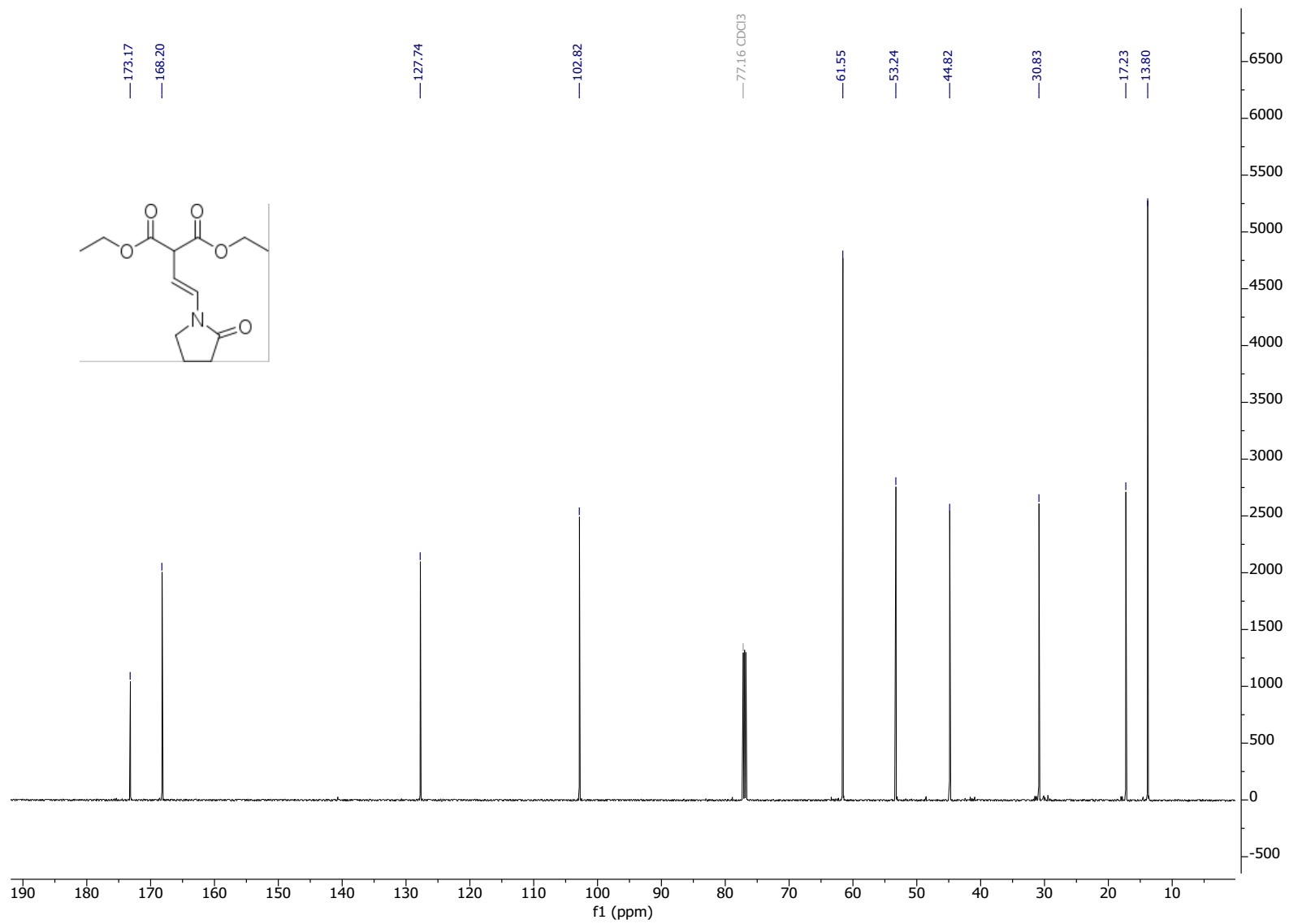
Spectrum S15: ¹H-NMR of compound **82**



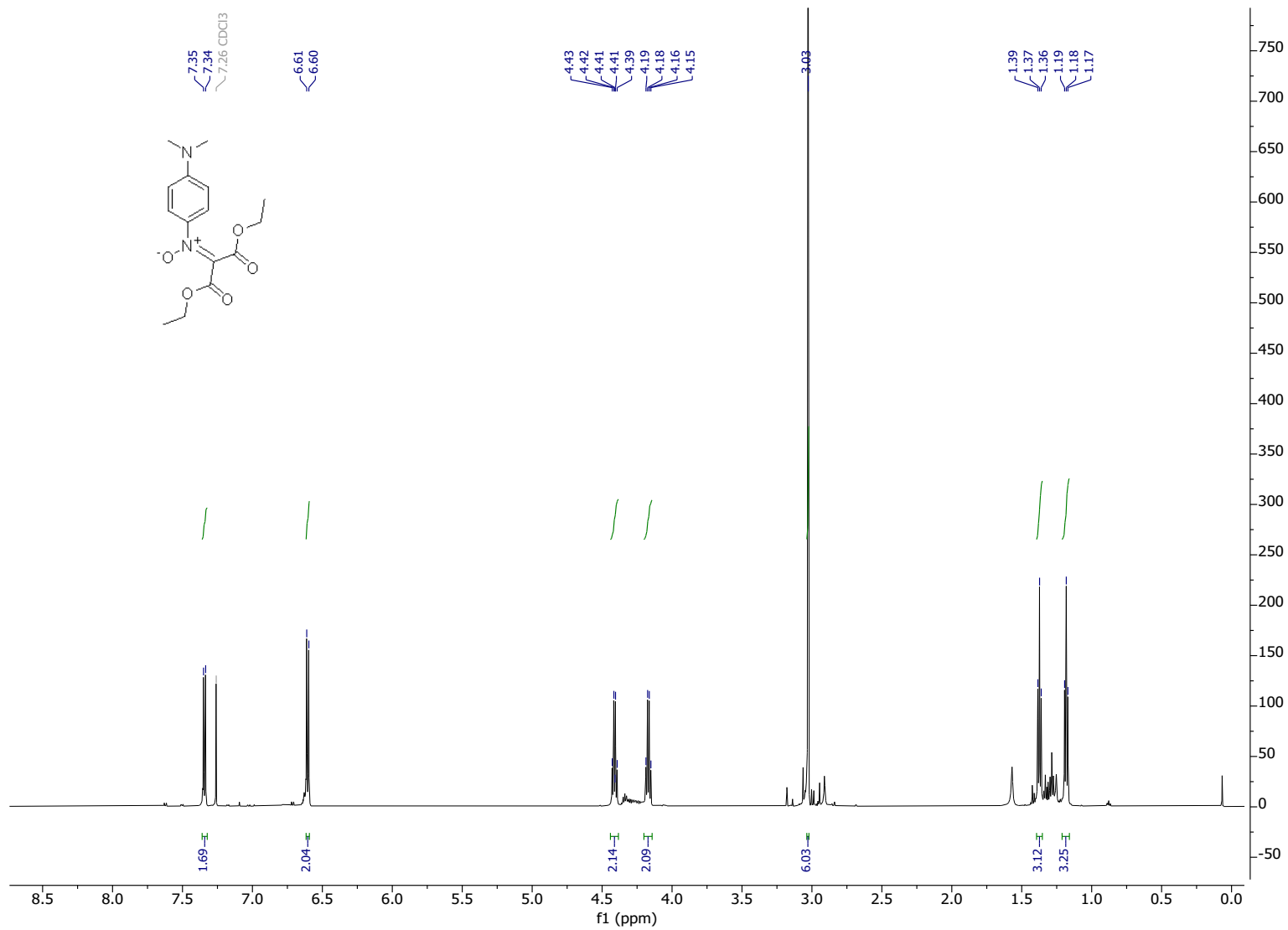
Spectrum S16: ^{13}C -NMR of compound 82



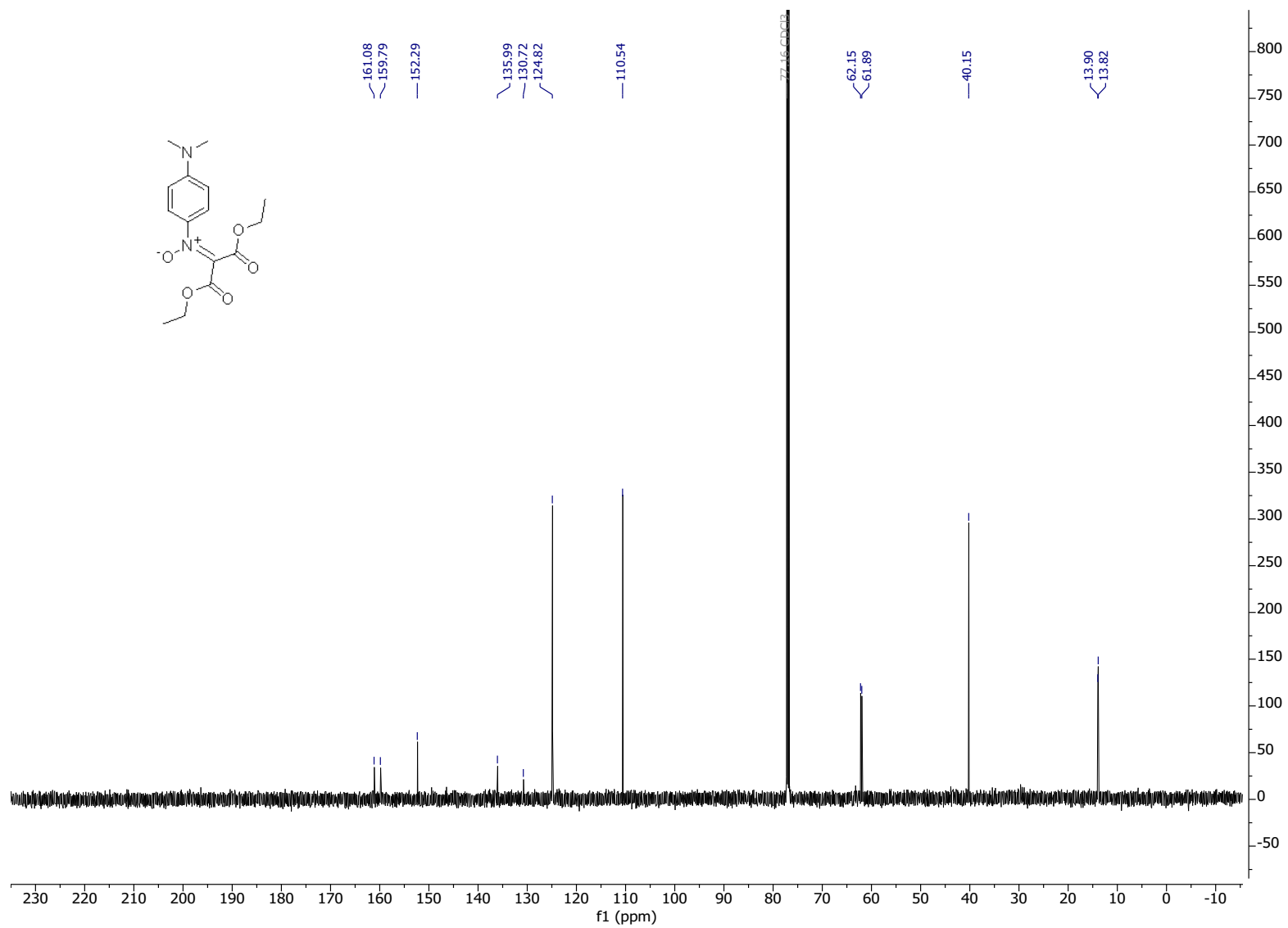
Spectrum S17: ¹H-NMR of compound **84**



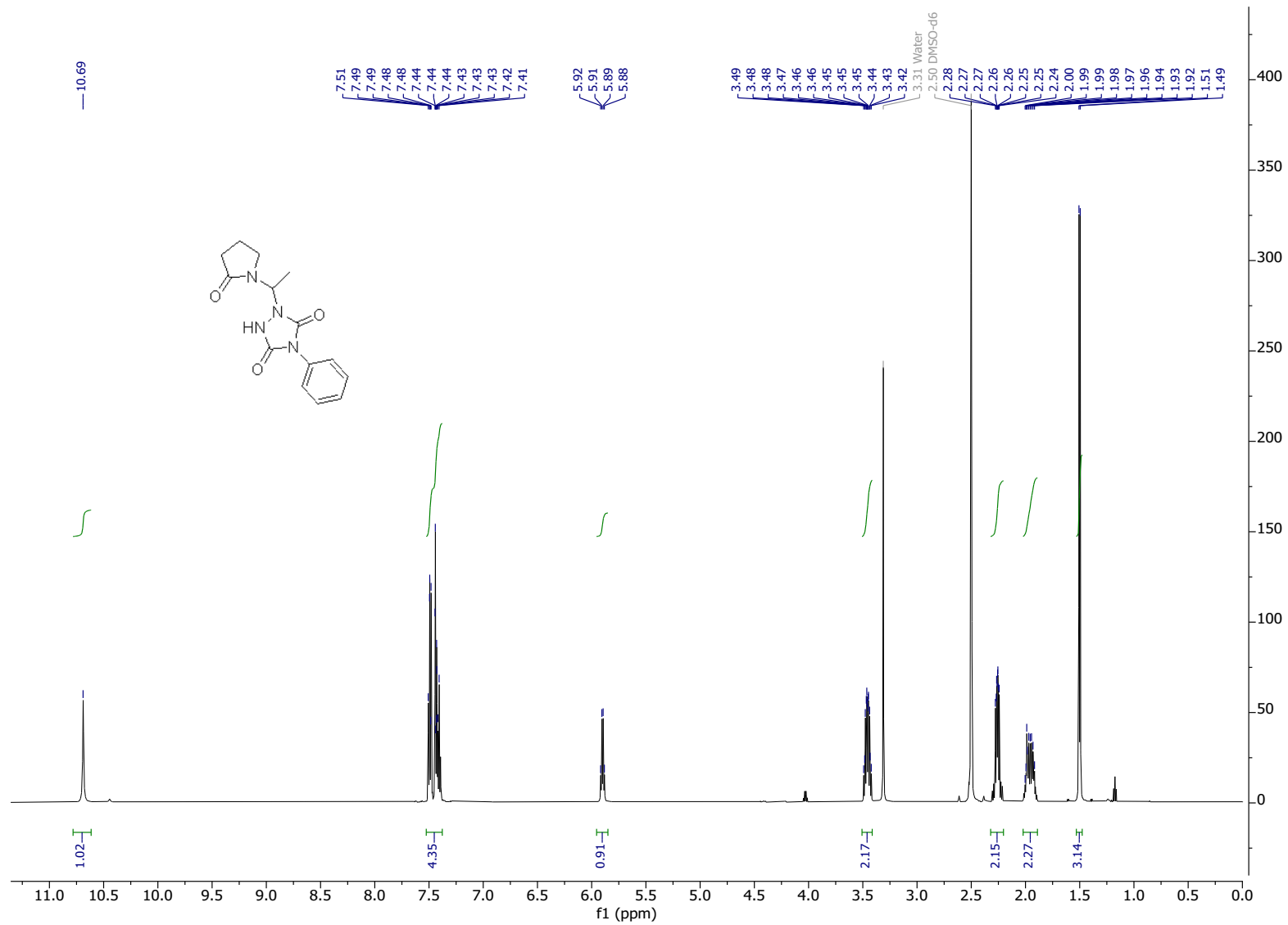
Spectrum S18: ¹³C-NMR of compound 85



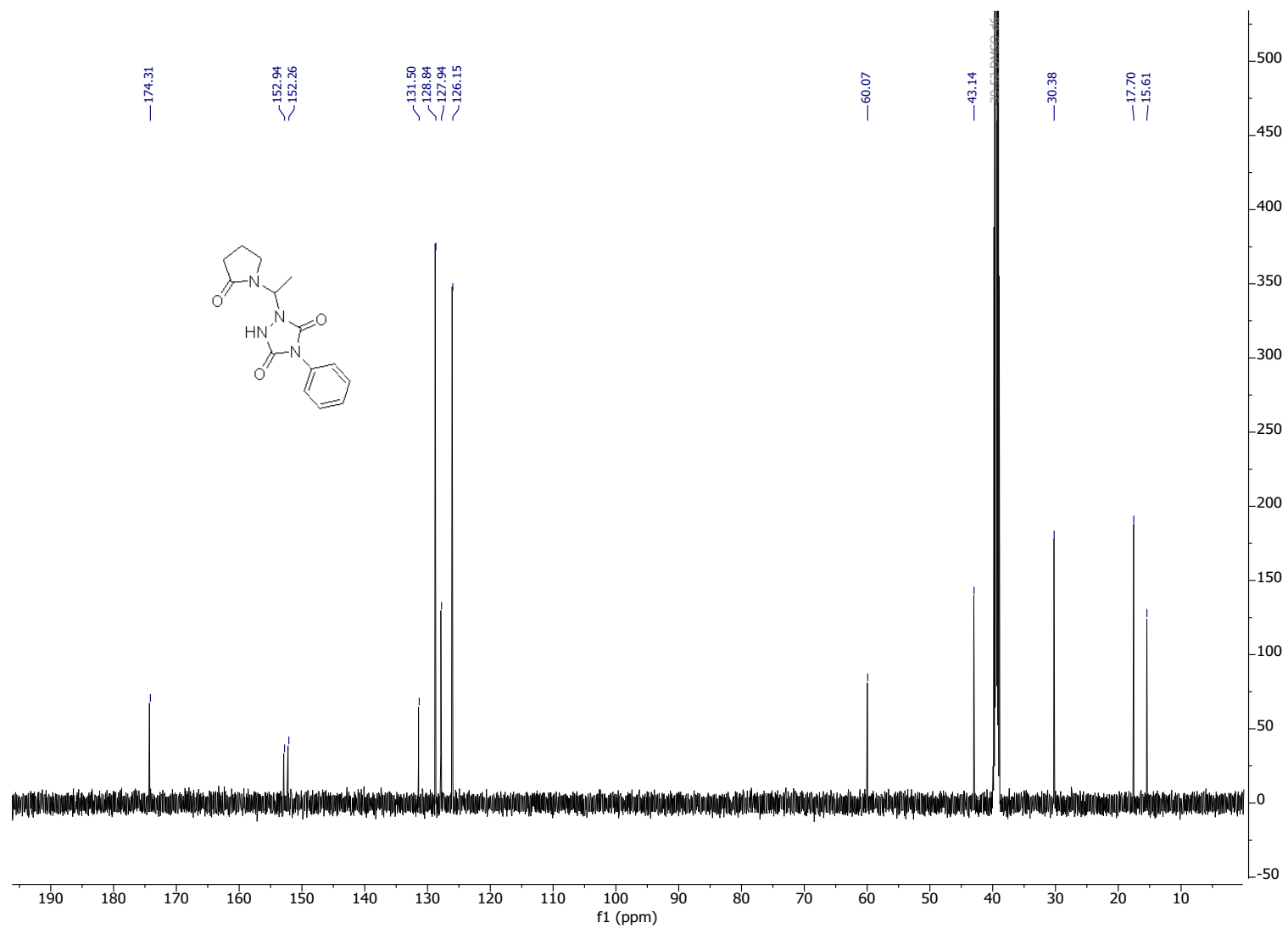
Spectrum S19: ¹H-NMR of compound 85



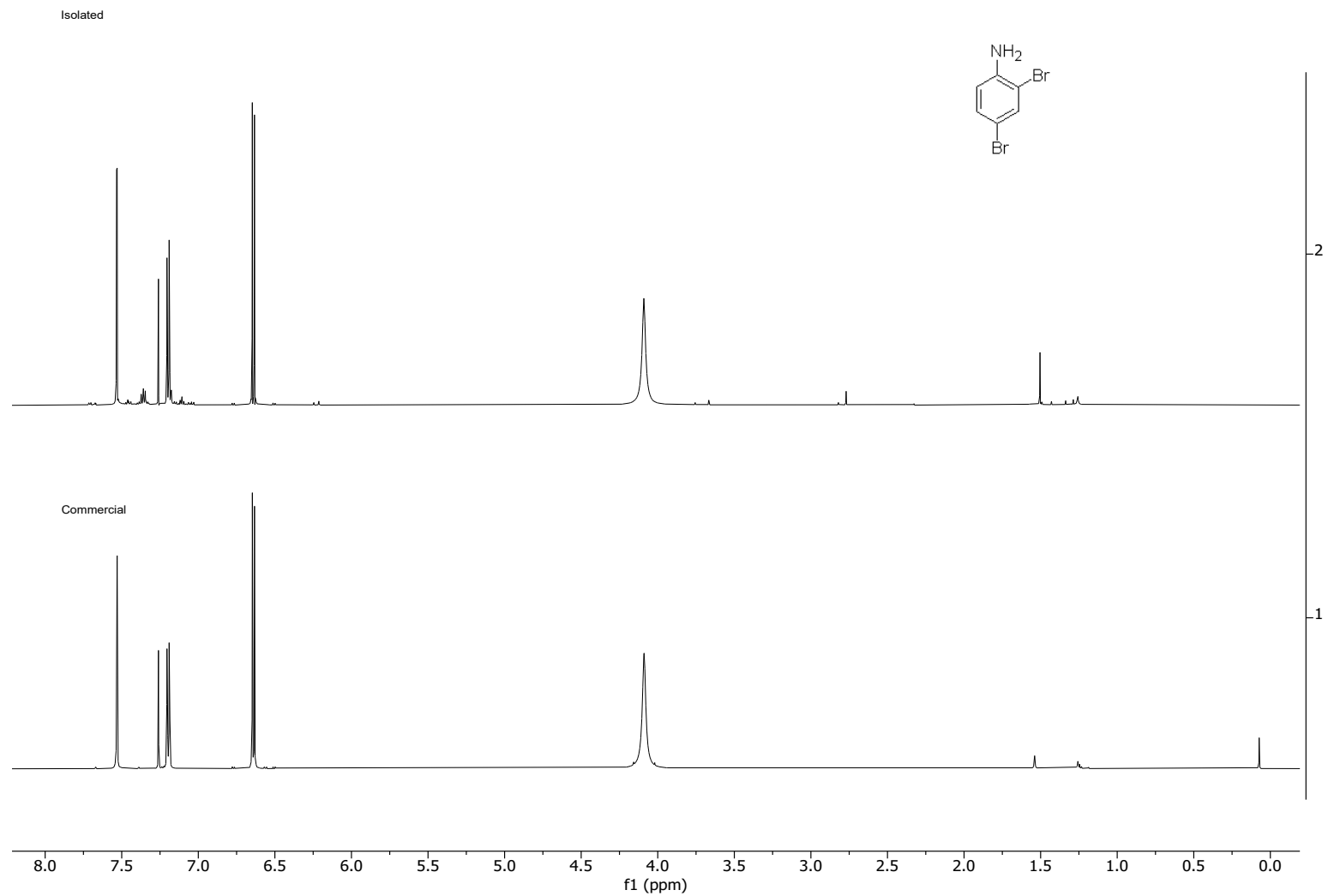
Spectrum S20: ¹³C-NMR of compound 85



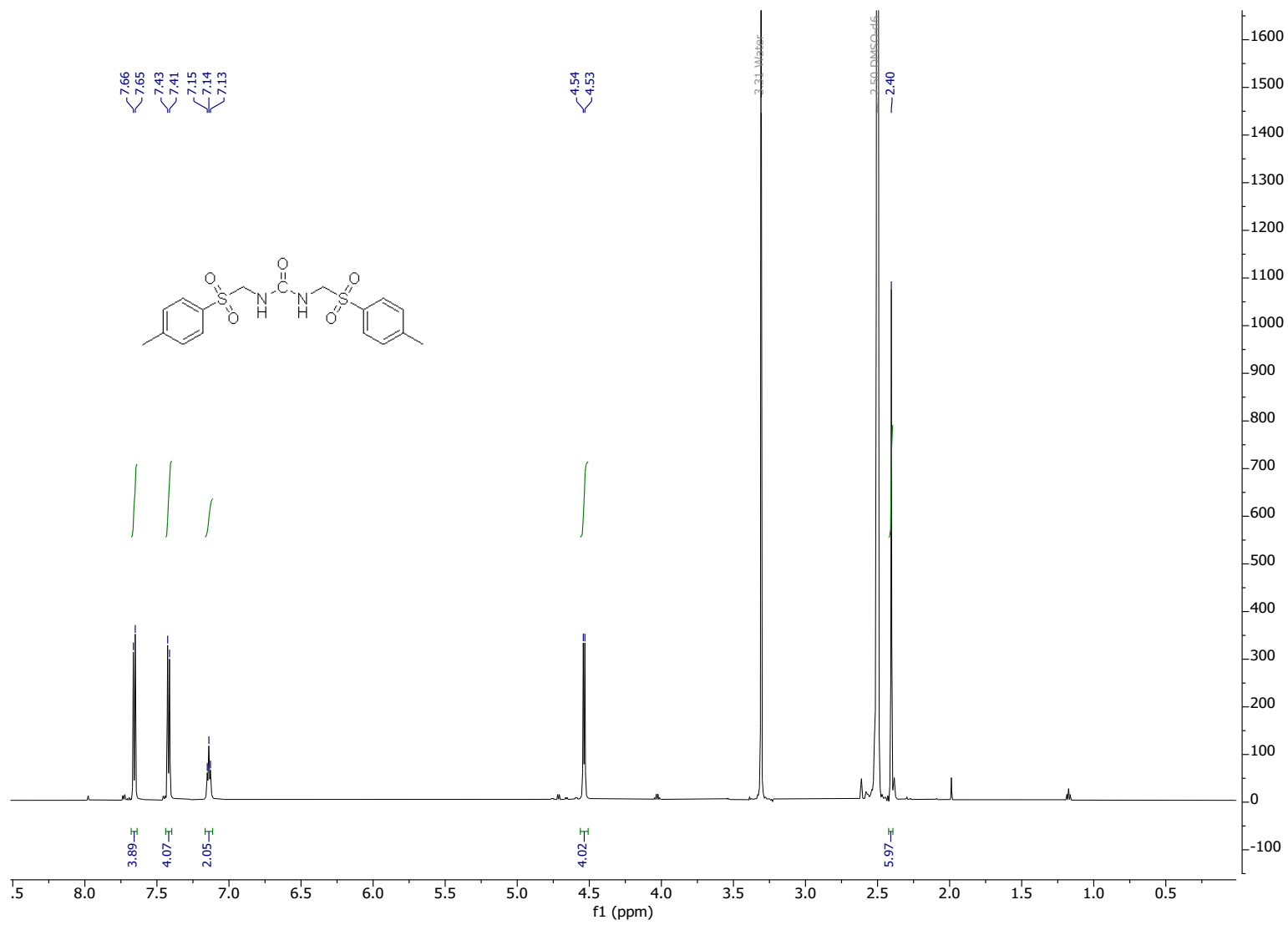
Spectrum S21: ¹H-NMR of compound 86



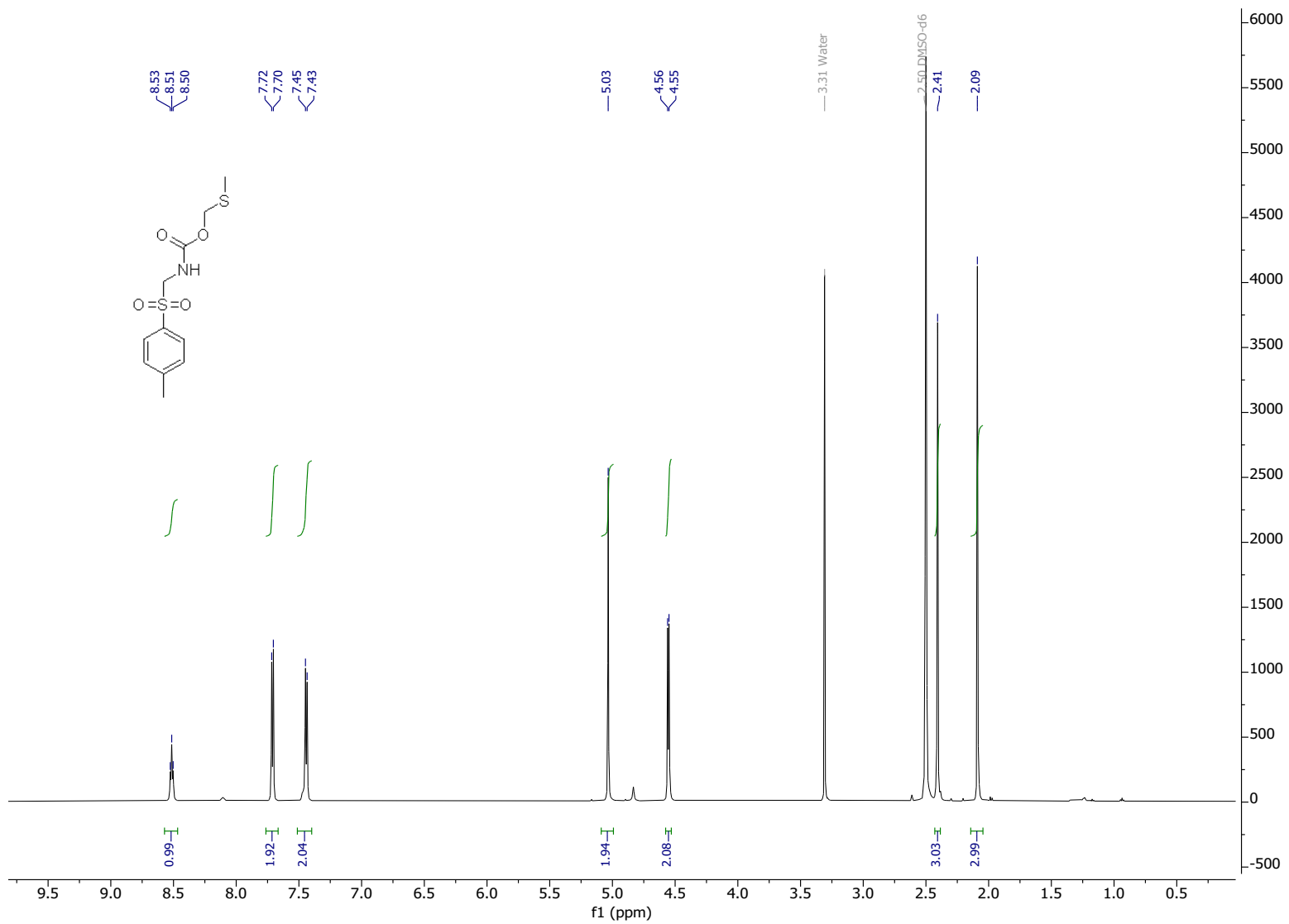
Spectrum S22: ^{13}C -NMR of compound 86



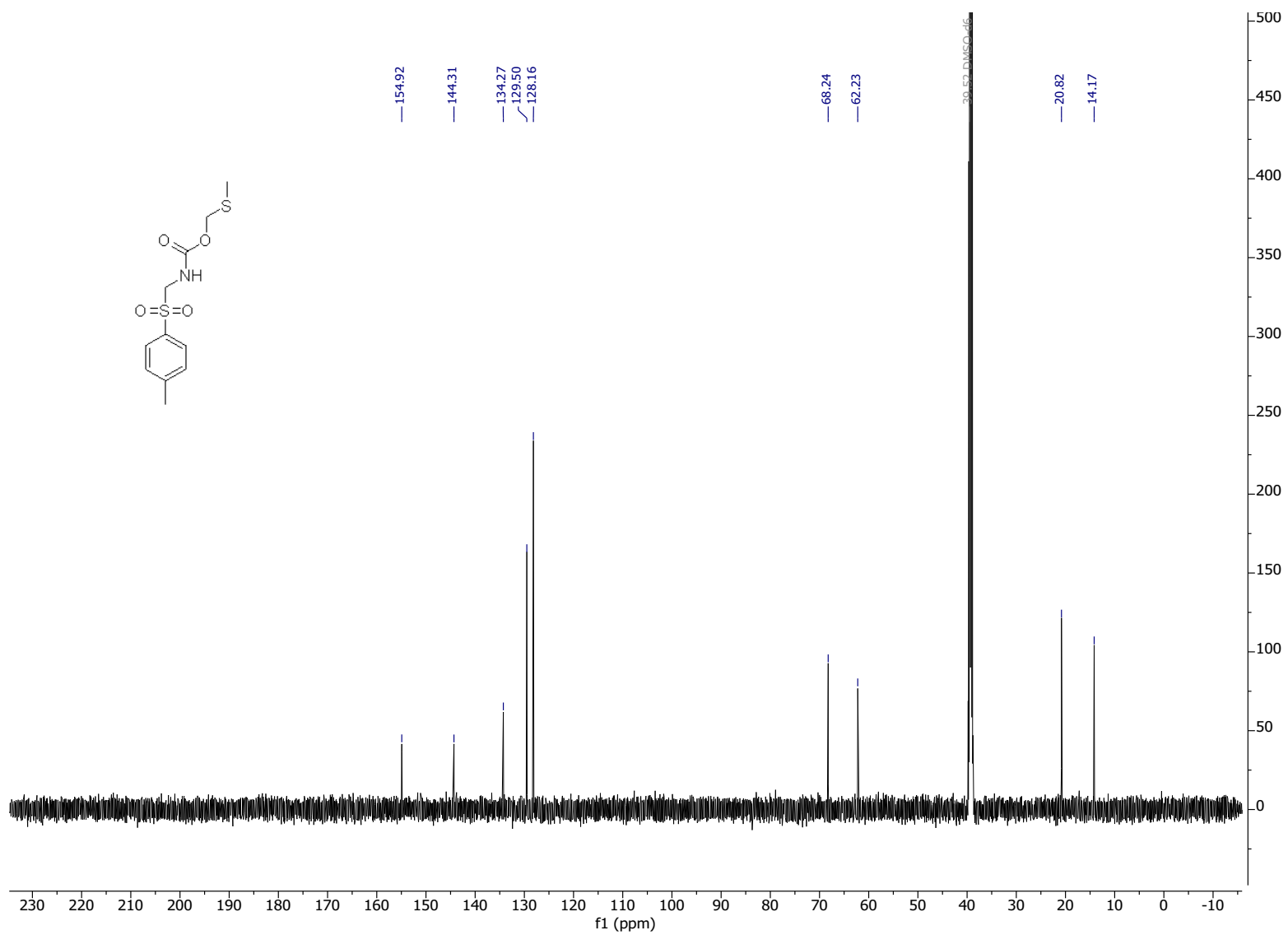
Spectrum S23: $^1\text{H-NMR}$ of compound **83**, compared with the commercially available 2,4-dibromoaniline



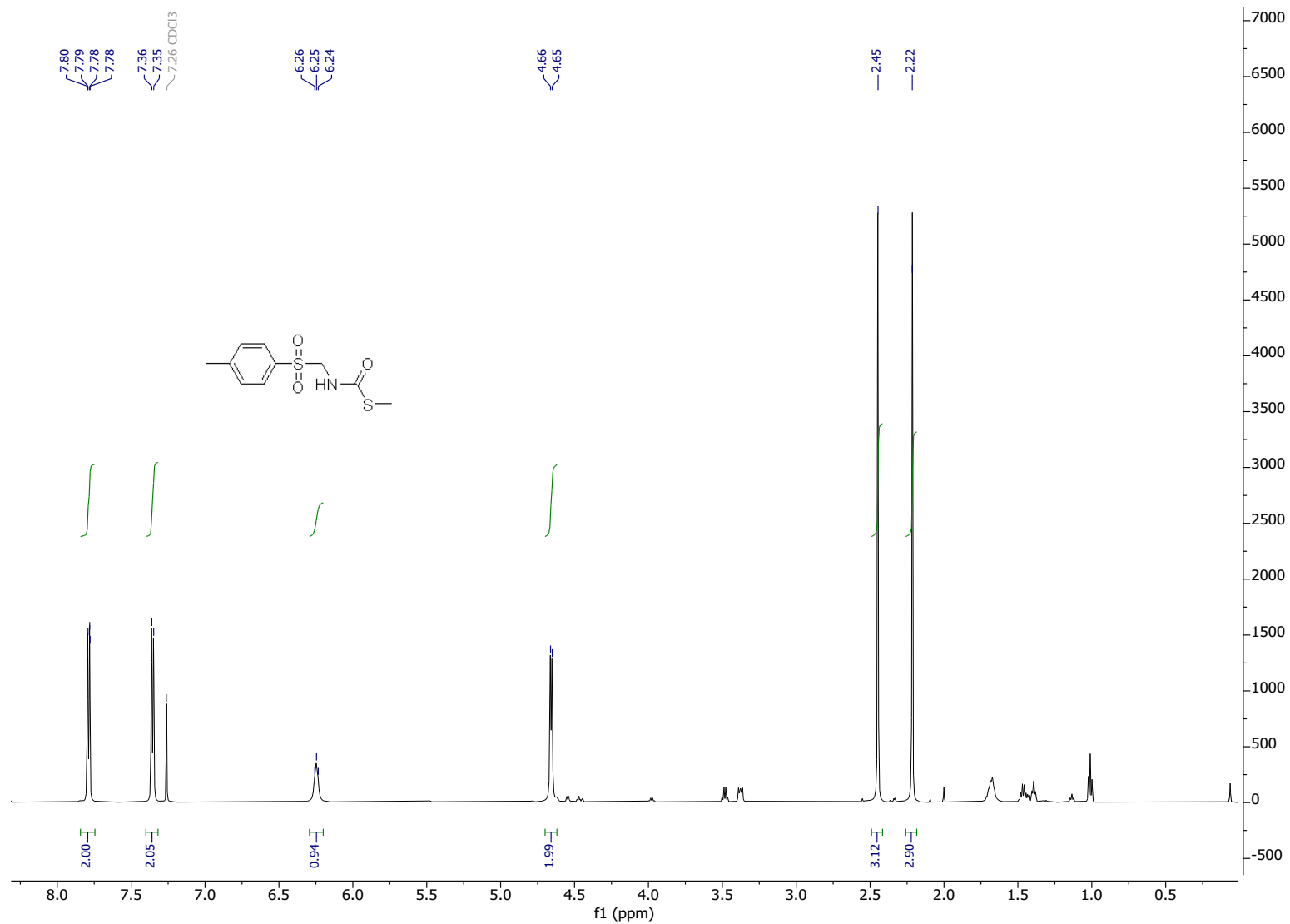
Spectrum S24: ¹H-NMR of compound 64



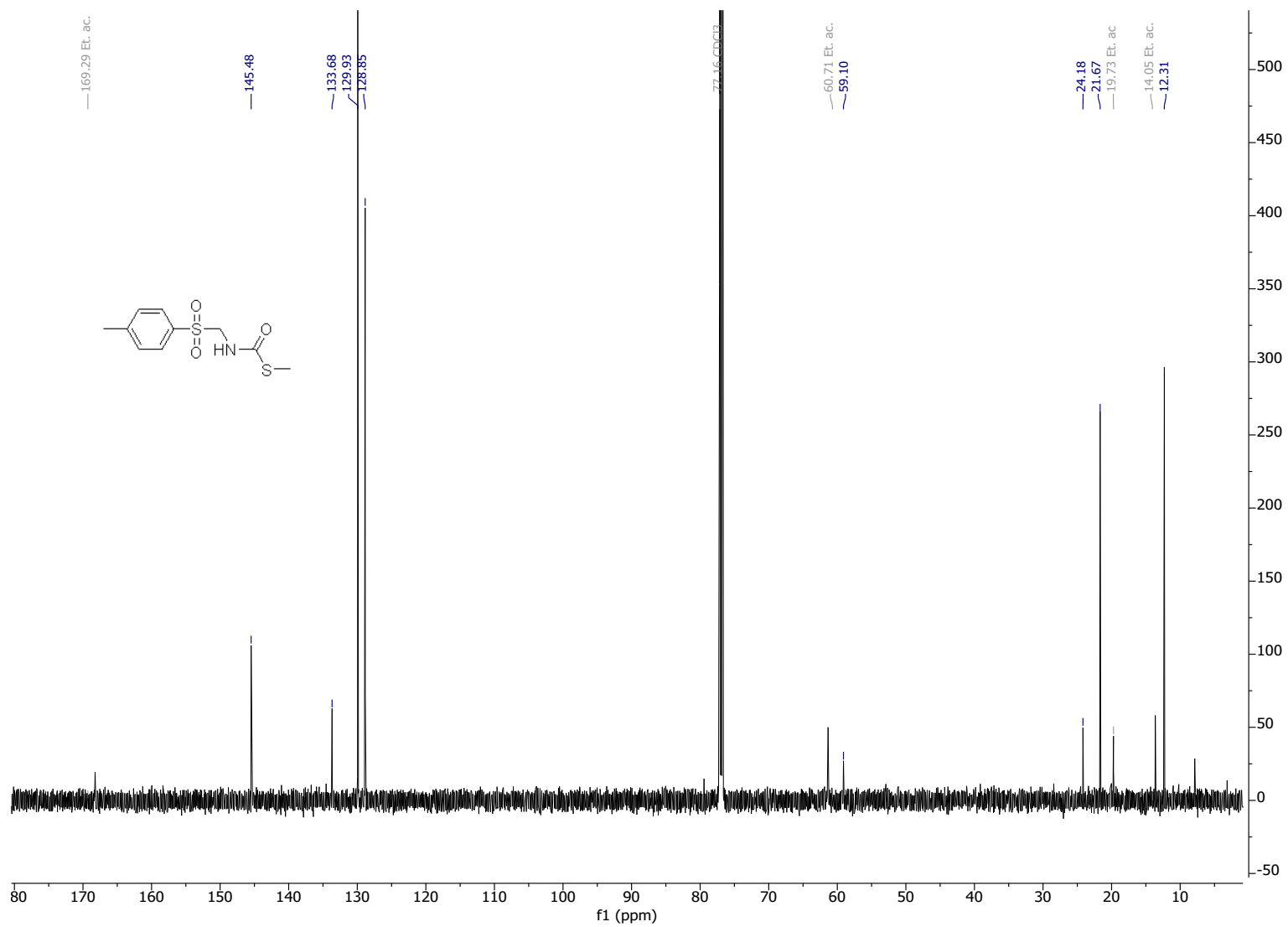
Spectrum S25: ¹H-NMR of compound **65**



Spectrum S26: ¹³C-NMR of compound 65



Spectrum S27: ¹H-NMR of compound 81



Spectrum S28: ^{13}C -NMR of compound 8

10 References

1. D.P. Kingma, J. Ba, CoRR, abs/1412.6980, 2014
2. <https://github.com/tensorflow/tensorflow> (accessed 11/03/19)
3. 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)
4. 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)
5. B. E. Hoogenboom, O. H. Oldenzien, A. M. Leusen, p-TOLYLSULFONYLMETHYL ISOCYANIDE. *Org. Synth.* **57**, 102 doi: 10.15227/orgsyn.057.0102 (1977).
6. 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).
7. 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).
8. A. Gossauer, K. Suhl, Totalsynthese des Verrucarins E sowie ihre Anwendung zur Herstellung eines ¹³C-markierten Derivates desselben. *Helvetica Chimica Acta*, **59** (5), 1698-1704 doi: 10.1002/hlca.19760590530 (1976)
9. A. H. Fenselau, J. G. Moffatt, Sulfoxide-Carbodiimide Reactions. III. Mechanism of the Oxidation Reaction. *J. Am. Chem. Soc.* **88**, 1762-1765 doi: 10.1021/ja00960a033 (1966)
10. 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)
11. W. Kemp, Nuclear Magnetic Resonance in Chemistry: A Multinuclear Introduction, 75 (Macmillan education ltd, 1986)
12. 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)
13. 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)
14. 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)

15. 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).
16. I. El Hassan, R. Lauricella, B. Tuccio, Formation of β -fluorinated aminoxyl radicals from N-arylketonitrones. *Mendeleev Commun.* **16**, 149–151 (2006).
17. E. Senogles, R. A. Thomas, The kinetics and mechanism of the acid-catalysed hydrolysis of N-vinylpyrrolidin-2-one. *J. Chem. Soc. Perkin Trans.* **2** 825–828 doi:10.1039/P29800000825 (1980)
18. G. M. Sheldrick, SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr. Sect. A Found. Crystallogr.* **71**, 3–8 doi: 10.1107/S2053273314026370 (2015).
19. G. M. Sheldrick, Crystal structure refinement with SHELXL. *Acta Crystallogr. Sect. C Struct. Chem.* **71**, 3–8 doi: 10.1107/S2053229614024218 (2015).
20. L. J. Farrugia, WinGX suite for small-molecule single-crystal crystallography. *J. Appl. Crystallogr.* **32**, 837–838 doi: 10.1107/S0021889899006020 (1999).
21. R. H. Blessing, An empirical correction for absorption anisotropy *Acta Crystallogr.* **51**, 33-8 doi: 10.1107/s0108767394005726 (1995)