

**INSTITUTE OF ORGANIC CHEMISTRY  
POLISH ACADEMY OF SCIENCES**



**DOCTORAL DISSERTATION**

**Computer-Assisted Discovery Of Unprecedented One-Pot Reactions  
And Functional Analogs**

M.Sc. Ahmad Makkawi

A monothematic series of publications with commentary presented to the scientific council of the Institute of Organic Chemistry of the Polish Academy of Sciences in order to obtain a doctoral degree in chemical sciences

Promotor: prof. Bartosz A. Grzybowski

Warsaw, 2025



Research for the doctoral thesis was carried out as part of the following projects:



NARODOWE CENTRUM NAUKI

*“Gated nanozymes: Hybrid, on-nanoparticle catalysts with long-range substrate selectivity”*

NCN MAESTRO (2018/30/A/ST5/00529)

*“Degradation of hazardous chemical compounds into valuable building blocks: from computer-planned routes to experimental validation”*

NCN SONATA (2020/39/D/ST4/01890)

I would like to thank the following:

**Professor Bartosz Grzybowski,**

for showing me the scientific path and all the  
motivation and advice he gave me.

**My all colleagues from team XI,**

for scientific discussions and years of joint work.

My family, for supporting me in these past years.

## TABLE OF CONTENT

<b>1. Publications included in the doctoral dissertation.....</b>	<b>6</b>
<b>2. Abstract.....</b>	<b>7</b>
<b>3. The objective of the work.....</b>	<b>12</b>
<b>4. Literature introduction.....</b>	<b>14</b>
4.1. The role of MCRs and one-pot reactions in modern organic chemistry.....	14
4.2. Drug-likeness and analog design through computational chemistry algorithms.....	17
4.3. Evolution of computer-assisted organic synthesis.....	23
<b>5. Projects' Discussions.....</b>	<b>29</b>
5.1. Validation of the Mech module: With the focus on one-pot reactions.....	29
5.2. Validation of the Analog module: Drug development and analogs generation.....	36
<b>6. Summary.....</b>	<b>43</b>
<b>7. Bibliography.....</b>	<b>44</b>
<b>8. Original publications.....</b>	<b>51</b>
<b>9. Contributors' statements.....</b>	<b>438</b>

## 1. PUBLICATIONS INCLUDED IN THE DOCTORAL DISSERTATION

**P01.** Roszak, R.; Gadina, L.; Wołos, A.; **Makkawi, A.**; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

**P02.** **Makkawi, A.**; Beker, W.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuć, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Źądło-Dobrowolska, A.; and Grzybowski, B. A. Retro-forward synthesis design and experimental validation of potent structural analogs of known drugs. *Chemical Science*, **2025**, *16*, 8383-8393.

## PUBLICATION THAT IS NOT INCLUDED IN THE DISSERTATION

**P03.** Gadina, L.; Baś, S.; **Makkawi, A.**; Bilgi, Y.; Klucznik, T.; Beker, W.; Grzybowski, B. A. Studying the outcomes and mechanisms of carbocationic rearrangements using algorithm-augmented experimentation. *Under Review*.

## 2. ABSTRACT

The integration of computers into chemistry has revolutionized the field, particularly in problem-solving, data management, and AI-assisted synthesis design. Initially used for simple data storage and calculations, modern computational tools now play a critical role in planning complex synthetic routes, enabled by advances in graph theory, reaction-network algorithms, and quantum chemical modeling. Once met with skepticism, AI has now proven capable of mimicking aspects of human reasoning in organic synthesis.

A key example of this progress is the *Allchemy* software, which was an indispensable companion of my doctoral work, providing synthetic ideas and plans that I subsequently set out to test in the laboratory. Unlike older tools that provide only basic recommendations, such as E.J. Corey's LHASA, Allchemy—a comprehensive platform for *in silico*, retro and forward synthesis—takes a more sophisticated approach by combining several specialized tools. Its different modules work together to tackle complex challenges, such as refining potential drug candidates, studying reaction mechanisms, and even exploring the chemical origins of life. It creates clear, logical reaction pathways and simplifies exploring chemical networks, speeding up molecule discovery. By combining vast reaction knowledge base with physical-organic chemistry principles, it designs and optimizes synthetic routes and strategies for even highly complex targets and can also work at the level of mechanistic steps to invent novel transformations, including multicomponent processes and carbocation rearrangements.

In my first study, I initially focused on one of Allchemy's modules (a.k.a **Mech**), a computational tool that demonstrated its ability to design novel multicomponent reactions (MCRs) and one-pot syntheses by integrating reaction mechanisms with physical-organic chemistry principles. Utilizing this **Mech** module, the algorithm predicted mechanistic pathways defining novel MCRs, estimated kinetic rates, and identified potential by-products. My work was primarily focused on validating and optimizing, through experimental synthesis, these predictions. I successfully generated a range of molecular frameworks, including novel tricyclic scaffolds, and compared direct- and conjugate-addition reactions—highlighting the critical role of HMPA in distinguishing these mechanisms. The study also explored competing side reactions, which the **Mech** module systematically identified and tracked. Its predictions for reaction selectivity and mechanistic competition aligned closely with experimental results, affirming the model's

accuracy. However, while the tool accelerates reaction discovery, it does not eliminate the need for hands-on experimental optimization, emphasizing the importance of keeping the human chemist “in the loop”.

Looking ahead, incorporating radical mechanisms and catalytic cycles will enhance the module’s predictive power further. By combining automated mechanistic exploration with rational filtering, the **Mech** module has proven effective in bridging computational and experimental chemistry, driving innovation in synthetic reaction design.

My second project dealt with the use of Allchemy’s **Analog** module for drug development. Recent computational advances have transformed analog drug design by enabling efficient synthesis of structural analogs using optimized pathways. Generative AI tools like Allchemy-**Analog** integrate molecular property predictions with target similarity to guide de novo scaffold design. In this project, I experimentally evaluated the module, which relies on bioisosteric substitutions, retrosynthetic analysis, and guided synthesis to create diverse analogs. The workflow ensured flexibility and addressed functional group limitations through refined protocols and adaptable chemical blocks. My experimental validation focused on designing, synthesizing and testing analogs of Ketoprofen and Donepezil. Seven analogs of Ketoprofen and five of Donepezil were synthesized, most showed micromolar to nanomolar binding affinities. One Ketoprofen analog slightly outperformed the original, and a Donepezil analog achieved a strong 36 nM affinity for AChE, closely matching the parent compound.

These experimental binding affinities aligned with theoretical predictions of Allchemy’s internal neural network (and also with predictions of docking programs) to within an order of magnitude. These results show how computer tools can simplify drug analog design, but reveal we still need better prediction methods for drug development.



## ABSTRACT IN POLISH

Integracja komputerów z chemią zrewolucjonizowała tę dziedzinę, szczególnie w rozwiązywaniu problemów, zarządzaniu danymi i projektowaniu syntezy wspomaganym przez sztuczną inteligencję. Początkowo używane do prostego przechowywania danych i obliczeń, nowoczesne narzędzia obliczeniowe odgrywają obecnie kluczową rolę w planowaniu złożonych ścieżek syntezy, co jest możliwe dzięki postępom w teorii grafów, algorytmach reakcji sieciowych i modelowaniu chemii kwantowej. Kiedyś spotykane ze sceptycyzmem, AI udowodniło, że jest w stanie naśladować aspekty ludzkiego rozumowania w syntezie organicznej.

Kluczowym przykładem tego postępu jest oprogramowanie Allchemy, które było niezastąpionym towarzyszem mojej pracy doktorskiej, dostarczając syntetycznych pomysłów i planów, które następnie postanowiłem przetestować w laboratorium. W przeciwieństwie do starszych narzędzi, które dostarczają tylko podstawowych rekomendacji, takich jak LHASA E.J. Corey'a, Allchemy—kompleksowa platforma do syntezy *in silico*, retro i forward—przyjmuje bardziej wyrafinowane podejście, łącząc kilka wyspecjalizowanych narzędzi. Jego różne moduły współpracują ze sobą, aby sprostać złożonym wyzwaniom, takim jak selekcja potencjalnych kandydatów na leki, badanie mechanizmów reakcji, a nawet eksploracja chemicznego pochodzenia życia. Tworzy jasne, logiczne ścieżki reakcji i upraszcza eksplorację sieci chemicznych, przyspieszając odkrywanie cząsteczek. Łącząc ogromną bazę wiedzy o reakcjach z zasadami chemii fizyczno-organicznej, projektuje i optymalizuje syntetyczne ścieżki i strategie nawet dla bardzo złożonych celów, a także może pracować na poziomie mechanistycznych kroków, aby wymyślać nowe transformacje, w tym procesy wieloskładnikowe i przegrupowania karbokationów.

W moim pierwszym badaniu początkowo skupiłem się na jednym z modułów Allchemy (znanym również jako **Mech**), narzędziu obliczeniowym, które wykazało swoją zdolność do projektowania nowych reakcji wieloskładnikowych (MCR) i syntez w jednym naczyniu poprzez integrację mechanizmów reakcji z zasadami chemii fizyczno-organicznej. Wykorzystując moduł **Mech**, algorytm przewidział ścieżki mechanistyczne definiujące nowe MCR, oszacował szybkości kinetyczne i zidentyfikował potencjalne produkty uboczne. Moja praca koncentrowała się głównie na walidacji i optymalizacji, poprzez syntezę eksperymentalną, tych przewidywań. Udało mi się wygenerować szereg

„ram molekularnych”, w tym nowe związki oparte na trójcyklicznym szkielecie węglowym, i porównać reakcje bezpośredniej i sprzężonej addycji—podkreślając kluczową rolę HMPA w rozróżnianiu tych mechanizmów. W badaniu zbadano również konkurencyjne reakcje uboczne, które moduł **Mech** systematycznie identyfikował i śledził. Jego przewidywania dotyczące selektywności reakcji i konkurencji mechanistycznej były ściśle powiązane z wynikami eksperymentów, co potwierdza dokładność modelu. Należy odnotować, że powyższe narzędzie przyspiesza odkrywanie reakcji, ale nie eliminuje potrzeby praktycznej optymalizacji eksperymentalnej, podkreślając znaczenie utrzymywania chemika-człowieka „w pętli”.

Patrząc w przyszłość, włączenie mechanizmów rodnikowych i cykli katalitycznych jeszcze bardziej zwiększy moc predykcyjną modułu. Łącząc zautomatyzowaną eksplorację mechanistyczną z racjonalnym filtrowaniem, moduł **Mech** okazał się skuteczny w łączeniu chemii obliczeniowej i eksperymentalnej, napędzając innowacje w projektowaniu reakcji syntetycznych.

Mój drugi projekt dotyczył wykorzystania modułu **Analog** firmy Allchemy do opracowywania leków. Ostatnie postępy obliczeniowe przekształciły projektowanie leków analogowych, umożliwiając wydajną syntezę analogów strukturalnych przy użyciu zoptymalizowanych ścieżek. Narzędzia sztucznej inteligencji generatywnej, takie jak Allchemy-**Analog**, integrują przewidywania właściwości molekularnych z podobieństwem docelowym, aby kierować projektowaniem rusztowania *de novo*. W tym projekcie eksperymentalnie oceniłem moduł, który opiera się na podstawieniach bioizosterycznych, analizie retrosyntetycznej i kierowanej syntezy w celu tworzenia różnorodnych analogów. Przepływ pracy zapewniał elastyczność i uwzględniał ograniczenia grup funkcyjnych dzięki udoskonalonym protokołom i adaptowalnym blokom chemicznym. Moja walidacja eksperymentalna skupiała się na projektowaniu, syntezy i testowaniu analogów ketoprofenu i donepezilu. Zsyntetyzowano siedem analogów ketoprofenu i pięć donepezilu, większość wykazała mikromolarne do nanomolowych powinowactwa wiązania. Jeden analog ketoprofenu nieznacznie przewyższył oryginał, a analog donepezilu osiągnął silne powinowactwo 36 nM do AChE (z ang. acetylcholinesterase), ściśle odpowiadające związkowi macierzystemu. Te eksperymentalne powinowactwa wiązania są zgodne z teoretycznymi przewidywaniami wewnętrznej sieci neuronowej Allchemy (a także z przewidywaniami

programów dokujących) z dokładnością rzędu wielkości. Wyniki te pokazują, jak narzędzia komputerowe mogą uprościć projektowanie analogów leków i warunkują potrzebę lepszych metod predykcyjnych w opracowywaniu leków.

### 3. OBJECTIVE OF THE WORK

Modern computational chemistry algorithms<sup>1-3</sup> offer promising capabilities for the systematic generation of retrosynthetic plans, molecular libraries or even suggestions for new reaction mechanisms. Within my doctoral study, I aimed to harness the power of these tools<sup>4,5</sup>, in particular the Allchemy platform, to test and explore their capabilities in synthetic chemistry design and drug discovery, and also to discover unprecedented routes of synthesis towards novel compounds through two different objectives:

#### **Discovery and Validation of one-pot reactions**

The first objective focused on enhancing—in a joint effort with other group members—computer-assisted organic synthesis by utilizing the Allchemy-**Mech** module to systematically predict and design novel multicomponent (MCRs) and one-pot reactions. After years of algorithmic development, this module reached operational readiness in the early 2020s, and by integrating reaction mechanisms, physical-organic chemistry principles, and kinetic modelling, this module was able to predict feasible mechanistic sequences—including previously unexplored transformations. My experimental validation (see **Fig. 10** later in the text) confirmed the module's accuracy in generating moderate to complex molecular frameworks (e.g., tricyclic scaffolds and Spiro compounds) and elucidating mechanistic nuances, such as the role of HMPA in controlling reaction selectivity. This work underscored the synergy between computational prediction and experimental optimization while highlighting areas for further refinement, such as incorporating radical and catalytic mechanisms.

#### **Computational-Experimental Pipeline for Drug Analog Design**

The second objective centered on evaluating the Allchemy-**Analog** module's capability to streamline the design of drug analogs through bioisosteric substitutions, retrosynthetic analysis, and guided forward synthesis. Using Ketoprofen and Donepezil as case studies, I was able to synthesize and characterize a number of analogs (**Figs. 13, 14** later in the text), assessing both synthetic feasibility and binding affinity predictions (via docking and neural networks). Results demonstrated successful analog generation, with several compounds achieving nanomolar affinities for their perspective enzymes. While computational predictions correlated broadly with experimental data, a certain extent of variations emphasized the need for improved affinity differentiation.

Overall, my work supports the claim that modern chemical-AI algorithms can significantly enhance synthesis planning and reaction discovery. The objective to test and validate such tools (here, Allchemy with its different modules of -Mech, -Analog) for predicting and designing organic reactions, drug analogs, highlights the value of “algorithmic synthesis”, and will hopefully help, other synthetic chemists to embrace these powerful tools in their own research.

## 4. LITERATURE INTRODUCTION

### 4.1. The role of MCRs and one-pot reactions in modern organic chemistry

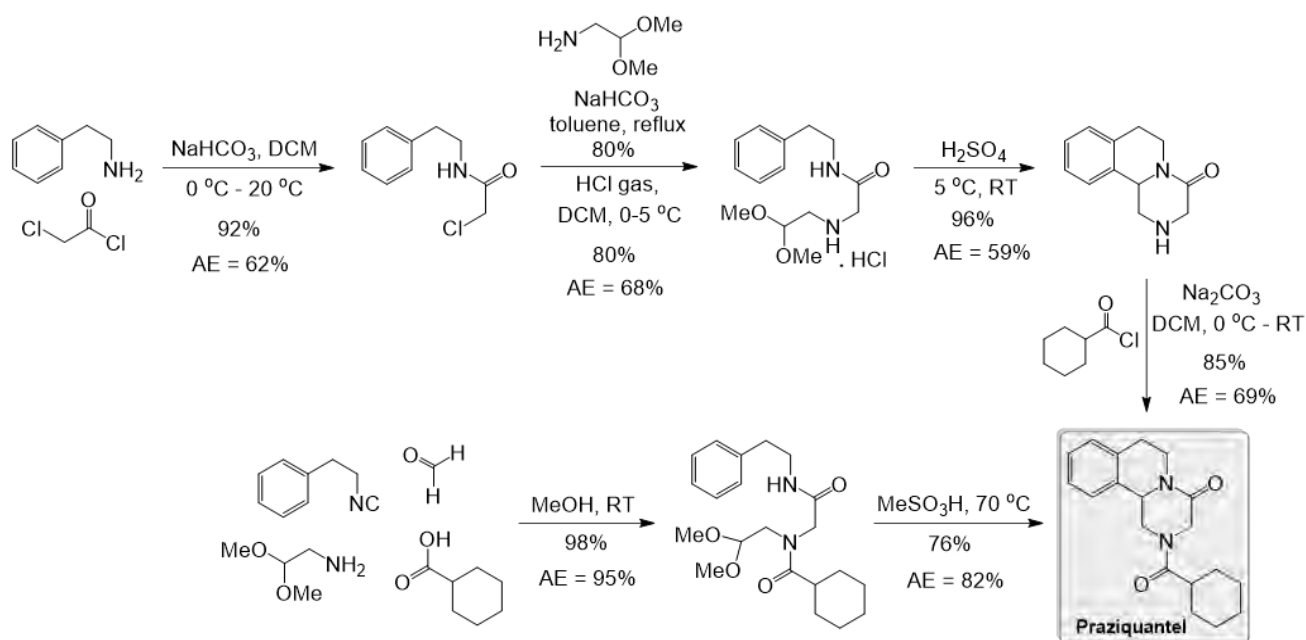
The significance of one-pot reactions was recognized early in the development of organic synthesis<sup>6</sup>. Since their inception, these reactions have evolved along two primary pathways<sup>7,8</sup>. In one approach, multiple orthogonal and irreversible steps are integrated into a single process. In the other, reversible steps are coupled with a final irreversible step, often facilitated by an enzymatic catalyst.

One-pot synthesis presents exciting challenges that fuel innovation in chemistry, paving the way for novel reactions, improved biocatalysts, cutting-edge materials, and deeper understanding of reaction mechanisms. The intricate design and synthetic pathways involved continually captivate researchers, pushing the boundaries in organic chemistry. To boost efficiency, scientists focus on minimizing synthetic steps—particularly in redox reactions<sup>9</sup>—while eliminating unnecessary atoms, streamlining processes for higher yields and more sustainable chemistry. Also, combining multiple reactions in one operation, like one-pot reactions, improves synthesis by consolidating steps such as ring and stereocenter formation without isolating intermediates.

Similar to multicomponent reactions (MCRs)<sup>10</sup>, cascade reactions facilitate efficient assembly of complex molecules from relatively simple starting materials. A cascade reaction simply combines sequential transformations in a single vessel without changing conditions. Hence, reactions performed in this manner allow the desired compound to be obtained in a single reaction stage, eliminating the need for the separation of intermediates. This technique minimizes unwanted by-products and dramatically improves reaction productivity. While designing cascade reactions that allow for precise control over product distributions, maximize molecular efficiency, and exhibit outstanding selectivity remains a significant challenge in the field of chemistry, biological systems have historically faced – and successfully solved through eons of evolution – a similar problem, relying on one-pot reactions to sequentially oxidize fuels and construct a wide range of complex molecules from basic precursors<sup>11</sup>. Nature's example thus teaches us that the design task is doable.

Likewise, catalytic cascade and domino processes enable efficient assembly of many functionalized scaffolds<sup>12</sup>. Instead of step-by-step construction, these transformations elegantly assemble molecules by forming several bonds at once. Also, relying on one-pot cascades and multicomponent reactions (MCRs) offers a smart and sustainable approach to building complex and biologically active molecules while keeping waste to a minimum. Recent studies highlight how these methods deliver two key benefits: they are environmentally friendly while reliably producing enantio-enriched compounds<sup>13</sup>.

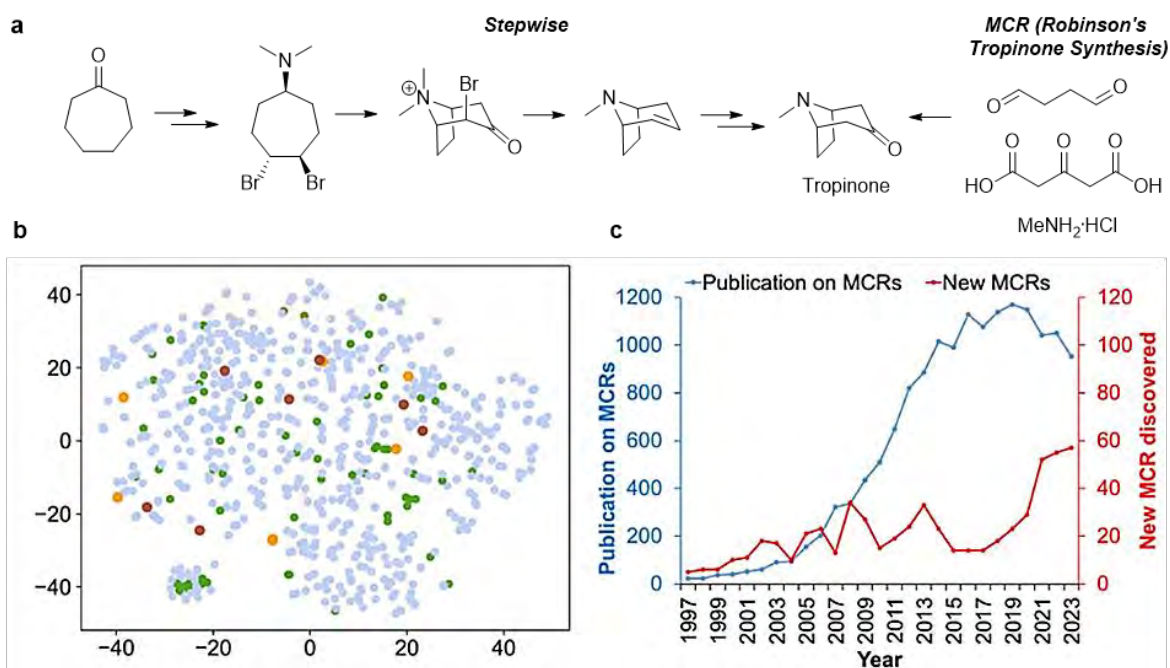
The case of praziquantel (PZQ, **Fig. 1**) serves as an illustrative example. PZQ, an anti-schistosomal drug on WHO's essential medicines list<sup>14</sup>, has been extensively studied, yielding several efficient industrial production methods<sup>15</sup>. The primary industrial process involves five steps, predominantly centered on amide bond formation, with the key transformation being an intramolecular N-acyliminium Pictet–Spengler cyclization. This route is notably robust, delivers high yields, and avoids the use of particularly hazardous reagents. However, a multicomponent approach presents a more concise and efficient alternative. Using a similar cyclization strategy, Dömling significantly reduced the synthetic sequence to just two steps by generating a pre-acylated intermediate through an Ugi four-component reaction prior to cyclization.<sup>16</sup>



**Fig. 1** | Linear vs. multicomponent synthesis of Praziquantel. *The figure was reproduced from ref.<sup>16</sup>*

MCRs and one-pot reactions are also appealing from the standpoint of atom economy (AE), a metric developed by Prof. Barry Trost in 1990<sup>17</sup> and measuring how efficiently a reaction uses atoms from reagents in the final product. This straightforward yet effective and green concept aids in the selection of synthetic routes by highlighting processes that maximize material utilization. Reactions that efficiently convert starting materials into the desired compound inherently minimize the generation of waste, making them highly favorable for sustainable synthesis.

In this context, reactions that enable the construction of complex molecular frameworks from several simple components in a single step, such as multicomponent reactions (MCRs)<sup>18,19</sup> (**Fig. 2a**<sup>5</sup>), and/or those that proceed in a sequential manner as one pot reactions<sup>20</sup>, are particularly valuable for minimizing the need for intermediate isolation and purification. These processes contribute to enhanced step- and atom-economy<sup>21</sup>, while also promoting the sustainability or "greenness"<sup>22</sup> of the overall synthesis. However, despite their potential advantages, the number of recognized MCR families remains limited to a few hundred (**Fig. 2b, c**<sup>5</sup>). This may be due to the fact that the most commonly investigated reactivity patterns—such as isocyanide,  $\beta$ -dicarbonyl, and imine-based MCRs—along with their basic combinations and expansions, have been explored in extensive details<sup>23,24</sup>.



**Fig. 2**<sup>5</sup> | **a** An example highlighting Robinson's one-step MCR synthesis versus a reported fifteen-step synthetic route<sup>25</sup> of Tropinone, for the stepwise synthesis, only the critical steps are illustrated.



**b** t-SNE map shows chemical diversity: small blue/green dots represent 631 known MCRs and 66 one-pot reactions, while larger red/orange markers highlight our newly discovered and experimentally validated reactions. Each dot plots the fingerprint difference between products and their starting materials<sup>26</sup>, capturing overall reaction transformations rather than step-by-step mechanisms. **c** The blue line (left axis) shows yearly MCR publications from a Web of Knowledge search, while the red line (right axis) tracks newly discovered MCR types from the 631-class dataset shown in **b**. Though MCR publications peaked around 2019 before slightly declining, we have seen a promising rebound in novel MCR development in recent years after non-linear trend between 2015-2017. This upward trend now suggests researchers are rediscovering the potential of multicomponent chemistry, focusing less on repeating known reactions and more designing novel ones. *The figure was reproduced from Publication P01.*

Despite over a century of research on multicomponent reactions (MCRs), scientists have only uncovered a few hundred of these valuable chemical processes to date. My research shows that computers, with core reaction mechanisms and organic chemistry principles, can independently create numerous diverse multicomponent and one-pot reactions without human intervention -- it is precisely for this purpose the Allchemy-**Mech** module was developed. By modeling kinetic rates, this state-of-the-art algorithm can also predict yields and identify reactions likely to benefit from organocatalysis—all before experimental testing. Through my own doctoral work and the work of my co-authors, these computational predictions are now validated by experiments, encompassing a range of reactivity types and diverse product frameworks. Without exaggeration, this is a momentous development in the history of chemistry – the first time that computers were able to assist and surpass humans in their ability to rationally design MCRs leading to relatively complex scaffolds.

#### **4.2. Drug-likeness and analog design through computational chemistry algorithms**

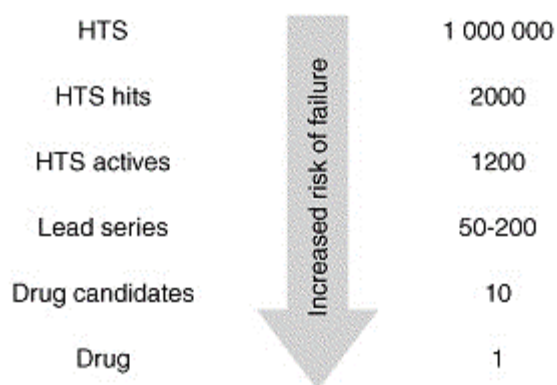
The second topic of my doctoral work focused on an outstanding challenge of drug design – in my own work, not surprisingly, with the help of computers. Predicting prior to synthesis which molecules have the potential to become effective drugs has been a long-standing and grand challenge in medicinal chemistry. If overcome, it would be a revolutionary development, allowing medicinal chemists to focus early on the most promising candidates (as opposed to HTS method; performing high-throughput syntheses

and screens). One of the earliest efforts to address this challenge came in with Lipinski's "rule of five,"<sup>27</sup> which evaluates a molecule's drug-likeness based on factors like size, ability to form hydrogen bonds, and fat-solubility (logP). However, over time, these rules have been questioned, especially since many modern drugs do not meet all of Lipinski's criteria. This concern was notably discussed in Shultz's review<sup>28</sup>.

As a result, better and more precise methods for evaluating drug-likeness have been developed. One advancement is the QED scoring system (Quantitative Estimate of Drug-likeness)<sup>29</sup>, enhancing Lipinski's rules especially in evaluating the physical and chemical features of potential drug candidates. QED offers a more refined way to rank potential drug candidates compared to Lipinski's rules, helping speed up the selection process from large libraries of compounds. While Lipinski's rules are moderately simple and easy to apply, QED gives a broader view of a molecule's drug-likeness by taking more factors into account and spotting possible red flags more effectively (like better identification of unfavourable properties). With this method, researchers can generate ranked lists of candidates, making it easier to identify the ones most likely to succeed, and covering key physicochemical properties, like molecular weight, hydrophobicity, polarity, the count of rotatable bonds (ROTBs), etc.<sup>30-32</sup>

Another important aspect to highlight is the typical route followed to develop potential drug candidates. Medicinal chemists generally adopt two main strategies when building compound libraries in the early phases of drug discovery. The first phase focuses on creating broad sets of compounds without necessarily aiming initially at any specific biological target, whereas the second approach is the opposite. Here we can see both different mentalities when dealing with preclinical research, where the libraries created in the first approach, often used in traditional high-throughput screening (HTS), have been built over years of extensive research. In the second strategy, one can design libraries based on known molecular structures, these libraries are often built by analyzing highly active molecules (HAMs) or studying 3D structures of compounds bound to their targets, resulting in a more selective set of compounds tailored to specific biological goals.

In general, HTS is useful for finding early hits in drug discovery, but improving how it is used is still a major challenge. On average, around a million compounds need to be tested during preclinical development—starting from initial screening to picking a clinical candidate—just to produce one drug that makes it to market. This shows how important it is to make the screening process more efficient. (Fig. 3)<sup>33</sup>



**Fig. 3** | Traditional figures of metric for drug discovery. The numbers provided serve illustrative purposes only. When analyzing (HTS) results, 40% of the initial hits can be false positive. On top of that, only one out of every two to five promising leads usually makes it past the early stages to become a real drug candidate. *The figure was reproduced from ref.*<sup>33</sup>

In theory, testing every possible molecule against a multitude of drug targets would give us the best chance of finding new leads. But in reality, that is simply not possible. The number of drug-like compounds that could be made is estimated to be over  $10^{60}$ —a number so huge it is called the Virtual Chemical Universe.<sup>34</sup> While this offers a valuable concept for researchers, it is impossible to physically create and test all of those compounds. Instead, modern drug discovery uses advanced cheminformatics tools<sup>3,35-37</sup> to explore this vast space. By analyzing key features of molecules in 2D, 3D, and their physical and chemical properties, scientists can focus on the most promising candidates without having to make everything in the lab. Cheminformatics, nowadays, has gained growing emphasis during initial lead discovery phases<sup>35</sup>, this approach helps researchers find good starting points for developing new drugs using methods that are both efficient and affordable.

Within this broad context outlined, my own work focused on a methodical approach in which one does not start a discovery campaign “from scratch”, instead, it focused on generating analogs that are similar to some already existing and promising structures. Of course, on paper, one can easily draw multiple analogs in a very short time – adjust some functional groups, replace some fragments with bioisosteric motifs, etc. In practice, however, even minute structural replacements may complexify the synthesis – in fact, this is a known limitation of *in silico* drug design whereby candidates may be predicted to bind to a certain protein very potently but are hard or even impossible to make.

Fortunately, there are now computational tools available that can rapidly assess the synthesizability of the computer-generated candidates<sup>38</sup>, even in large numbers and also for complex targets. These synthesis-tools come handy in the problem of analog design. In this context, the two primary strategies have emerged (i) one in which the algorithm simulates chemical reactions to generate compounds, which are then filtered according to similarity to a desired “parent” molecule; or (ii) starting with the original drug and work backward, this will generate analogs by substituting the initial reactants with functionally or mechanistically similar components.

These two methods make the design of drug analogs flexible, efficient, and better suited to real-world needs, and both can evaluate how closely a compound fits the target, while also predicting other key characteristics and molecular properties<sup>39</sup>. With these considerations in mind, my work utilized Allchemy’s **Analog** module<sup>4</sup>, that approached generating potential candidates from a different angle. This module follows a two-step process to suggest easily synthesizable analogs. First, the parent compound is modified by replacing its certain parts with matching bioisosteric motifs. This generates a small library of virtual “parent” compounds that are then subjected to retrosynthetic analysis to disconnect them into simple and commercially available building blocks. These generated substrates from retrosyntheses are then used for “forward” syntheses. Importantly, this process also allows a set of popular reagents (a.k.a., “auxiliaries”) that allow the substrates to engage in additional reactions to generate molecules that are similar but not identical to the “parents” from which the substrates were derived. In this way, within minutes, the algorithm can design thousands of diverse structural analogs to the parent molecule(s).

Importantly, because the number of steps in both retro-and forward syntheses are limited and controlled by the user, all these candidates are easily makeable, overcoming one of the major bottlenecks of the so-called “generative models”. Once the algorithm creates the “space” of analogs, these compounds can be evaluated and sorted for various chemical and medicinal-chemical criteria.

**Fig. 4** illustrates some of these sorting and filtering features. For instance, synthetic intermediates as well as compounds featuring functional groups not found in any approved drugs can be filtered out. Candidates not complying with the popular PAINS filters, those predicted to have unfavourable ADME-Tox properties (hERG cardiotoxicity, blood-brain-barrier penetration, high protein plasma binding, etc.) can be triaged. Users can also choose to view only analogs featuring bioisosteric motifs, filter results by various metrics of molecular similarity to the parent(s), and many more. Once a candidate (or candidates) of interest are selected, their properties can be queried in detail, as illustrated in **Fig. 5**. Aside from the specific synthetic routes, Allchemy provides information about the candidates physicochemical and ADME-Tox properties as well as a list of 20 proteins to which a given candidate is predicted to bind most potently (these predictions are made by Allchemy’s neural network models trained on close to 2 million protein binding assays).

The screenshot displays the 'Filters' panel in Allchemy, organized into several sections:

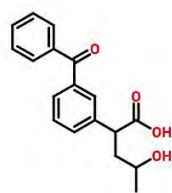
- AI filters:** Includes 'drug-likeness', 'binding', and 'toxicity'.
- ADME filters:** Includes 'human PPB %', 'MDCK-MDR1 efflux ratio', 'HERG binding', and 'Blood-brain barrier penetration'.
- Substructure filters:** Includes 'Filter by PAINS', 'Exclude undesired motifs', and 'Filter by presence of chemical element'.
- Analog-Specific Filters:** Includes 'Exclude intermediates' and 'Show only bioisosteres similarity to lead'.
- Structural filters:** A list of various physicochemical and structural parameters such as 'mass', 'logP', '# H-bond acceptors', 'PSA (polar surface area)', '# of aromatic rings', 'Fraction of heteroatoms', and 'oxygen balance'.

A dropdown menu is open, showing sorting options: 'mass', '# rings', '# stereo centers', 'Balaban index', 'Bertz index', 'N/C ratio', 'oxygen balance', 'best binding', 'similarity to lead Tversky chiral' (highlighted), 'similarity to lead Tversky achiral', 'similarity to lead Tanimoto chiral', and 'similarity to lead Tanimoto achiral'. The main interface shows chemical structures of various molecules, some with highlighted functional groups.

**Fig. 4** | Screenshot from the module illustrating various filtrations available to apply in Allchemy's **Analog** module.

The module includes five categories:

AI-based filters, ADME filters, substructure filters, analog-specific and structural filters. Several of them can be applied simultaneously for more refined results.



Molecular formula	C <sub>18</sub> H <sub>16</sub> O <sub>4</sub>	ΔH <sub>formation</sub>	not calculated
Mass	298.338 g/mol	N/C ratio	0.00
logP	2.857	Oxygen balance	-219.98 %
totalPSA	74,600 Å <sup>2</sup>		
Drug-likeness	0.990 ±0.009		
human PPB %	94.30±1.20	MDCK-MDR1 efflux ratio	2.20±1.90
HERG binding	0.01	Blood-brain barrier penetration	0.46
Tversky similarity chiral	0.86	Tanimoto similarity chiral	0.81
Tversky similarity achiral	0.86	Tanimoto similarity achiral	0.81
Smiles	CC(O)CC(C(=O)O)c1cccc(C(=O)c2ccccc2)c1		

target	description	organism	probability	avg. probability	ligand count	pKi threshold
CHEMBL221	Cyclooxygenase-1	Homo sapiens	0.989 ±0.005	0.173	2352	6.0
CHEMBL2157	Interleukin-8	Homo sapiens	0.981 ±0.009	0.198	208	6.0
CHEMBL2949	Cyclooxygenase-1	Ovis aries	0.969 ±0.014	0.118	2138	6.0
CHEMBL230	Cyclooxygenase-2	Homo sapiens	0.933 ±0.022	0.393	4088	6.0
CHEMBL4042	Cyclooxygenase-1	Rattus norvegicus	0.907 ±0.027	0.191	68	6.0
CHEMBL1293244	Beta-lactamase	Pseudomonas aeruginosa	0.87 ±0.033	0.039	276	6.0
CHEMBL3713095	Galactocerebrosidase	Homo sapiens	0.856 ±0.044	0.392	1701	6.0
CHEMBL3481	Carboxypeptidase A1	Bos taurus	0.839 ±0.064	0.112	95	6.0
CHEMBL1293278	Geminin	Homo sapiens	0.82 ±0.052	0.343	37829	6.0
CHEMBL2176772	E3 ubiquitin-protein ligase TRIM33	Homo sapiens	0.813 ±0.048	0.288	100	6.0
CHEMBL4102	Cyclooxygenase-2	Ovis aries	0.81 ±0.066	0.371	811	6.0
CHEMBL4321	Cyclooxygenase-2	Mus musculus	0.759 ±0.136	0.611	543	6.0
CHEMBL1293232	Survival motor neuron protein	Homo sapiens	0.754 ±0.196	0.274	23846	6.0
CHEMBL1944	Nepriylisin	Homo sapiens	0.754 ±0.071	0.367	497	6.0
CHEMBL2193	Cerebroside-sulfatase	Homo sapiens	0.743 ±0.065	0.376	384	6.0
CHEMBL3713062	Tissue factor pathway inhibitor	Homo sapiens	0.734 ±0.063	0.773	1684	6.0
CHEMBL1907598	Integrin alpha-V/beta-3	Homo sapiens	0.733 ±0.063	0.603	1492	6.0
CHEMBL3622	Cytochrome P450 2C19	Homo sapiens	0.706 ±0.11	0.241	9092	6.0
CHEMBL1255126	Protein Mdm4	Homo sapiens	0.68 ±0.12	0.237	413	6.0
CHEMBL4744	Alpha-2a adrenergic receptor	Bos taurus	0.675 ±0.142	0.364	116	7.0

**Fig. 5** | Screenshot from the Allchemy **Analog** module, summarizing properties of one of the algorithm-generated analogs of Ketoprofen.

### 4.3. Evolution of computer-assisted organic synthesis

Computers, in their early applications, served as rather primitive tools to support the computational capabilities of chemists. In synthetic chemistry, the landscape started to slightly shift with the advent of computer-assisted synthesis design<sup>40</sup>, which aimed to assist us in problem-solving tasks<sup>41</sup>. Of course, these early beginnings were shy compared to nowadays. For example, the inception of early programs like LHASA<sup>40,42</sup> operated by analyzing only a single step at a time, and had remained constrained to addressing relatively straightforward or very simple targets. Many more programs were also designed for similar purposes (some of them are mentioned in this section). While most of them are

no longer in use today, one should recognize their foundational importance even though these innovations did not lead to widely adopted computational tool(s). However, they played a role in advancing the field of computational chemistry, for instance, some laid the groundwork for developing different kinds of molecular notations that computers can read and interpret (like, SMILES)<sup>43</sup>.

Aside from Corey's LHASA, there were several other revolutionary programs for computer-assisted synthesis. For example, SYNCHEM<sup>44</sup>, developed in the late 1960s by Prof. H. Gelernter, aimed to investigate multi-step synthetic pathways using a knowledge base of chemical reactions and practical chemical intuition—it was visionary approach at that time. However, it struggled with stereoselectivity predictions and with complex reactions. Another example appeared in the 1980s, by Prof. P. Y. Johnson called SYNLMA (Synthetic Laboratory Management and Automation)<sup>45</sup>, it took a clever approach by translating chemical reaction knowledge into logical rules, reflecting how experienced chemists actually think about some transformations in organic chemistry. CAMEO<sup>46</sup>, on the other hand, developed in Prof. W. L. Jorgensen's group took a different strategy -- unlike the retrosynthetic tools that were mentioned before, this one focused on predicting the products of organic reactions (in a forward manner) by analyzing starting materials and conditions based on recognized mechanisms. However, it had a few key limitations - it required human intervention to manually input reaction mechanisms, and it could not autonomously learn or improve from new data. In addition, the accuracy was questioned as mechanisms did not necessarily reflect the experimental conditions. Another tool called RAIN<sup>47</sup>, was developed in Prof. Ugi's group in 1987, and their aim was to explore chemical reaction networks—emphasizing both forward and retrosynthetic directions. However, this system did not consider selectivity – meaning it generated many reactions without highlighting the most viable or the most practical options, and these growing reaction networks required heavy computing power, and consequently, was difficult to control or deal with.

One can notice that most of these systems had common limitations – they did not consider some key factors like: solvent choice, temperature, reaction time, or catalyst effects, also they were not reliable in predicting yields or scalability. None of these mentioned tools, despite of their innovative efforts at the time, became a practical tool for chemists, and virtually all were disconnected. The reason might be related to chemists simplifying the models they adopted, or the limitation of the hardware or the computing power which



restricted the progress. Today, our algorithmic understanding of the problem is much more advanced and so is the hardware. Today, it is evident that viable synthesis planning programs must not only incorporate very broad sets of reaction rules, but must also consider potential cross-reactivity conflicts, substituents and functional groups, stereo- and regiochemistry, protecting groups, etc.

To help computers play a more useful role in chemistry, scientists realized the importance of creating solid mathematical models<sup>48</sup> that can reflect how chemical structures and reactions work. One helpful tool, for example, is graph theory, which offers a clear way to represent molecules<sup>49</sup>. Dugundji and Ugi<sup>48,c</sup> were among the earliest to introduce a matrix-based method for modeling molecules that became a key part of this approach. The effectiveness of these systems relies entirely on the size of the reaction database and the quality of the stored information.

Considering all of the above, computers now can demonstrate a distinct advantage due to their substantial memory capacity and rapid data processing speeds, far more powerful than in the past. However, they still remain, to some extent, limited in their ability to perform associative reasoning, a domain where humans excel. Now, one of the fundamental objectives of synthesis design is still to identify and develop optimal synthetic pathways to achieve desired target molecules, But, one can notice that computers have gained the necessary tools over time, and have acquired the abilities to help design synthetic pathways for a wide spectrum of target compounds.

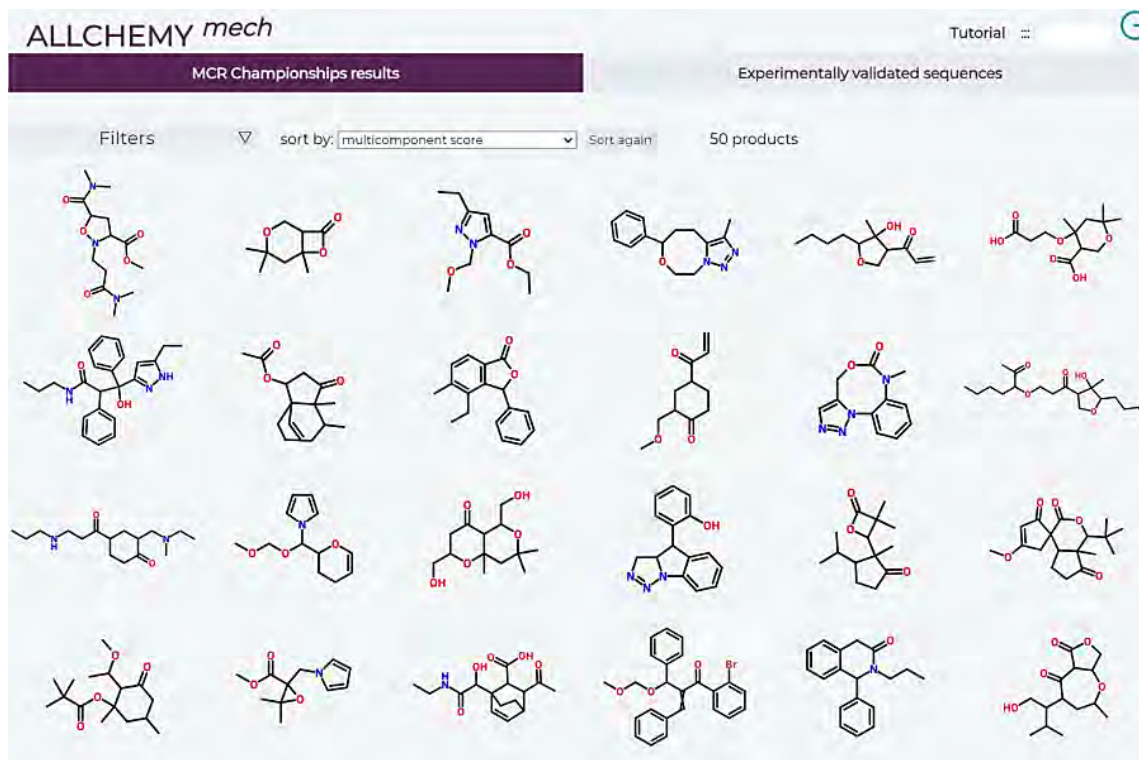
By analyzing vast databases of known reactions and chemical rules, viable synthetic pathways can now be proposed by machines in minutes. While these systems are not flawless for every situation, they have become valuable tools for speeding up discoveries and uncovering new opportunities in synthetic chemistry. However, we should not underestimate the complexity behind the computing systems, even with good data, the process is not straightforward or easy task. One must standardize the different data formats he or she has, and then picking the best way to represent them for the machine to learn from. For such things, there is no universal solution - the right choice depends on both the specific goals and the chemical details the scientist is working with.

Fortunately, the progress in machine learning, network theory, and modern computing systems has encouraged the chemists to explore how computers can be better integrated into research for greater impact. With these advancements, computers can now plan

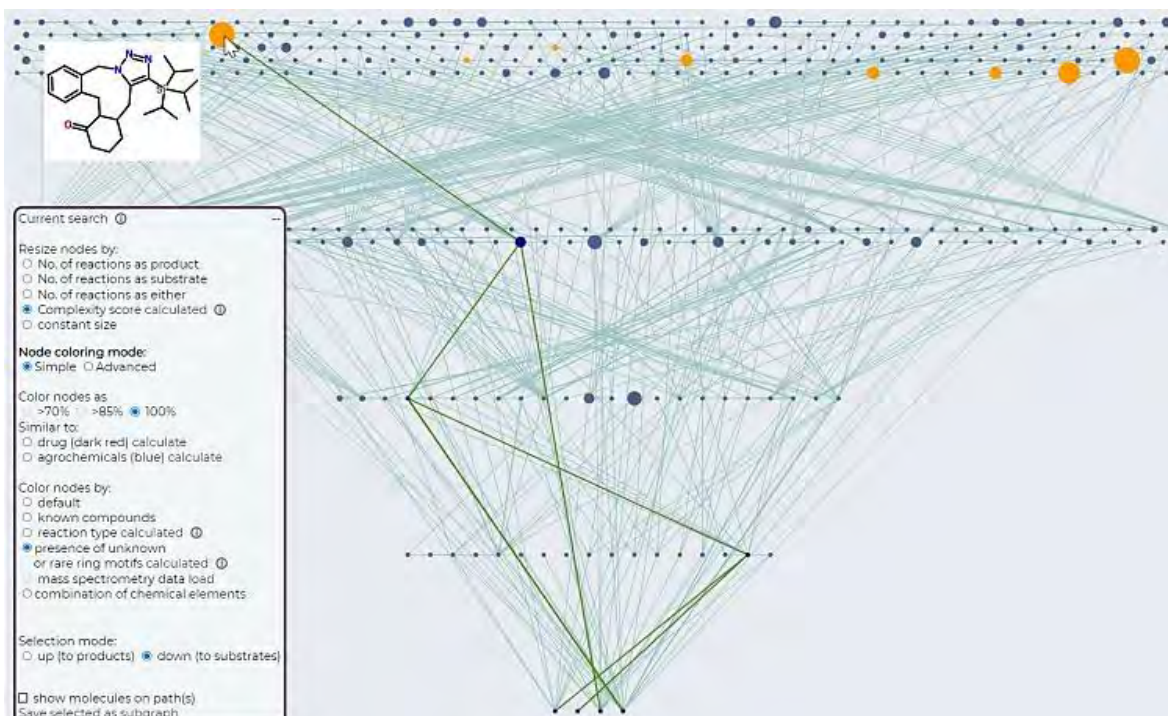
syntheses of complex organic molecules—often in a very short matter of time<sup>50,51</sup>, In doing so, they can recommend optimal conditions (like solvents, temperatures and catalysts, etc.), propose alternative synthesis routes for the same target, and suggest different possible mechanisms.

Hence, my doctoral work relied heavily on the use of one of the modern platforms, Allchemy<sup>4,5,52,53</sup>, developed over the last decade and succeeding the outstanding Chematica/Synthia<sup>1</sup> (also developed by our group and now marketed by Merck/Millipore-Sigma). In addition to traditional retrosynthesis, Allchemy also works in the “forward” direction to explore spaces of synthesizable molecules and to evaluate their properties (notably, affinities to various proteins, see above); in the retro-forward mode to take a “parent” molecule of interest, decompose it into starting materials (and their bioisosteric replacement) and then “reassemble” them into parent’s analogs; or in the mechanistic mode in which it explores sequences of mechanistic steps from which it chooses the mutually compatible ones and those that are unprecedented – thus defining mechanistically novel reactions and, in my own work, one-pot reactions. Each of these tasks is carried out in a separate module. For instance, **Mech** module deals with mechanistic design of new reactions and MCRs, **Analogs** module suggests medicinally-related compounds, **Hopcat** deals with carbocationic rearrangements, **Life** module is related to the origins of life and the C,O,N,S, P-based prebiotic chemistries. In all of these modules, Allchemy provides products along with detailed lists of synthetic pathways enabling their execution (the user manual of each module is presented in different published articles<sup>4,5,52,53</sup>).

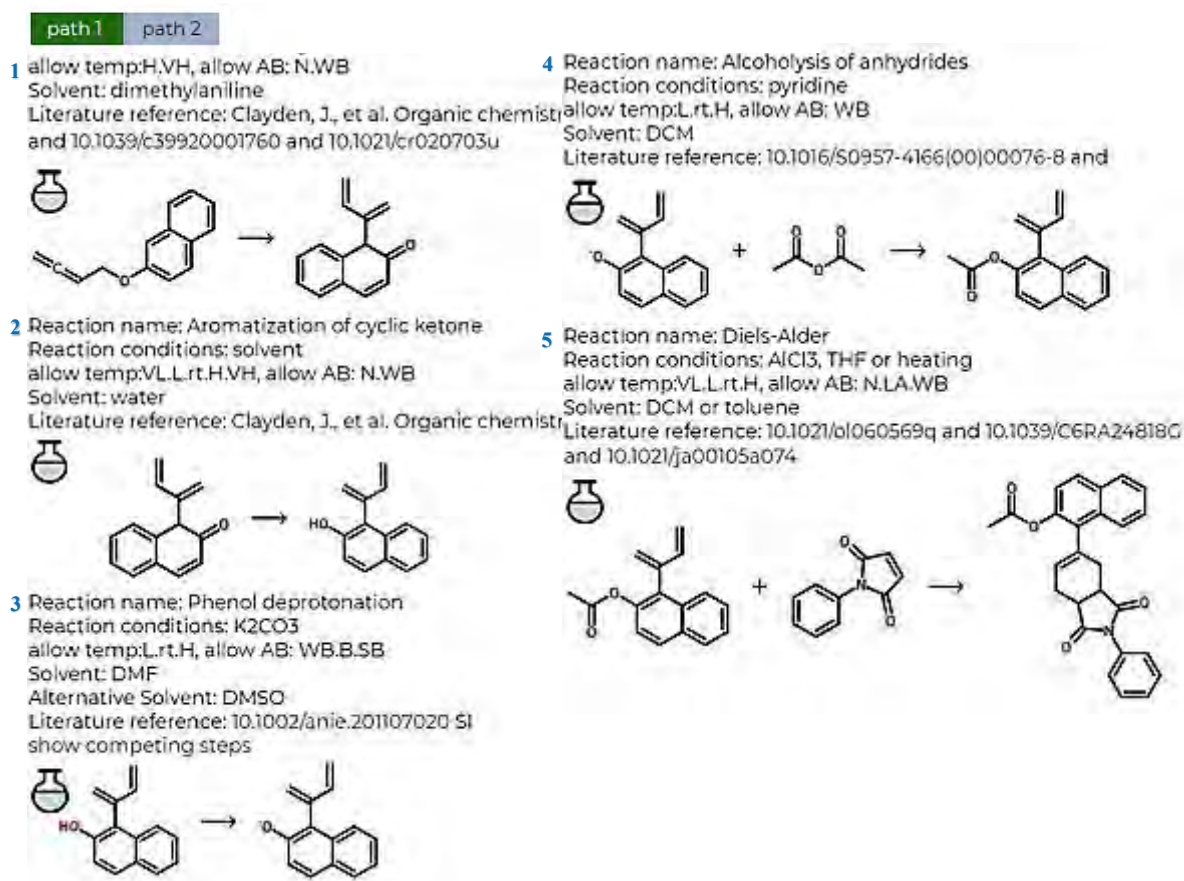
In Allchemy, users can view products and their synthetic routes in two ways: as a list (**Fig. 6**) or as a network tree/diagram (**Fig. 7**). In both formats, clicking on a molecule's icon or node lets users explore its full synthesis path(s). The raw results are usually displayed as in **Fig. 8**, where each step includes standard reaction conditions and references to scientific literature. This helps chemists choose the best synthesis route by combining their own expertise with a wealth of chemical knowledge. What makes Allchemy powerful is its ability to rapidly analyze thousands—even millions—of possible synthetic pathways or mechanistic sequences, redefining the way we tackle complex problems in organic chemistry.



**Fig. 6** | A list of products shown as chemical structures and generated by Allchemy. Synthetic pathway(s) are visualized by clicking on a desired molecule (see **Fig. 8**). *The figure was taken from Publication P01.*



**Fig. 7** | An example of an Allchemy-generated synthetic network with each layer represents a synthetic generation—showing the substrates at the bottom and the products at the top. Users can right-click on any node to choose the shortest synthesis pathway. *The figure was taken from Publication P01.*



**Fig. 8** | An example of a mechanistic pathway created by Allchemy's **Mech** tool. Each reaction step includes details like its name, standard conditions, solvent, and direct references to published examples. *The figure was adapted from Allchemy.*

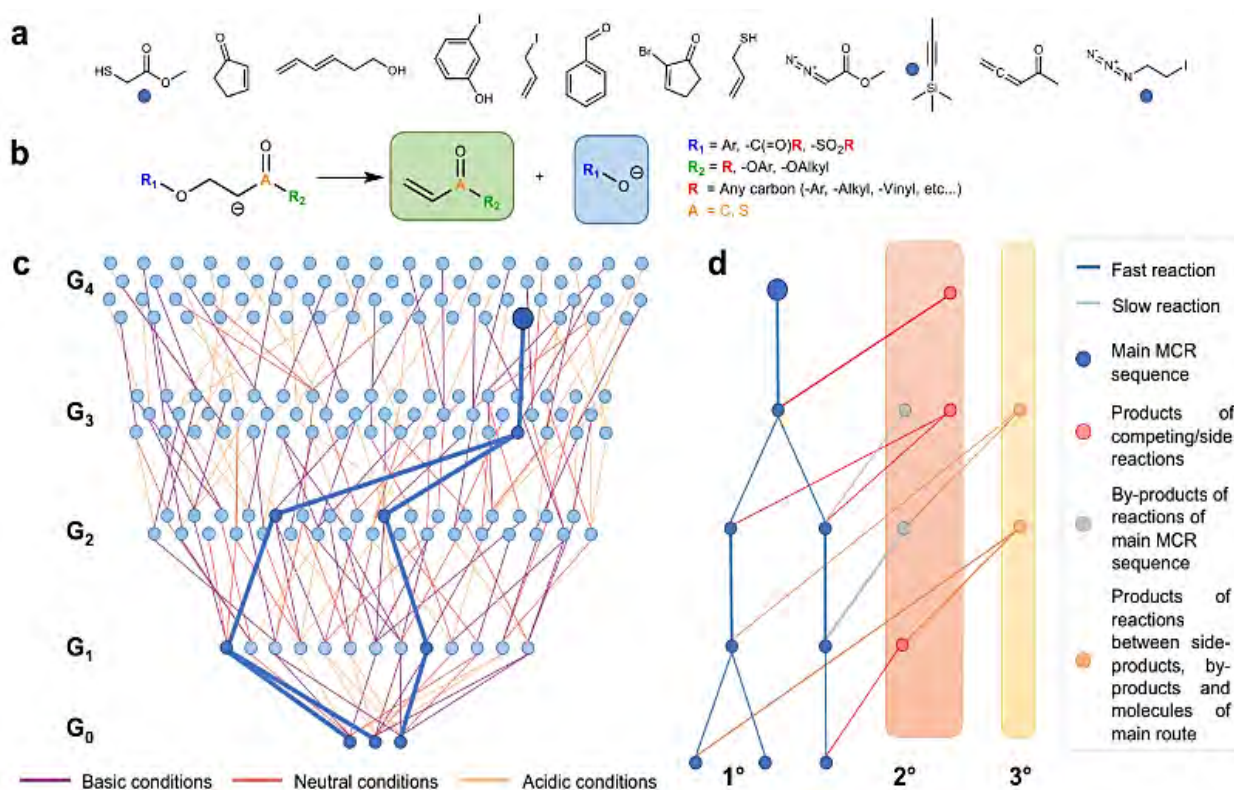
## 5. DISCUSSION

### 5.1. Validation of the Mech Module: With the focus on one-pot reactions

The Substrates' selection:

Although the algorithm of the **Mech** module is capable of processing any user-defined molecular inputs, identifying substrates that facilitate efficient multicomponent reactions (MCRs) or one-pot sequences remains a complex endeavour. To address this, a high-throughput computational analysis of substrate combinations was implemented, drawn from a library of around 2,400 small molecules representing diverse modes of chemical reactivity. These molecules are readily commercially available, each featuring one or two reactive functional groups amenable to various chemical transformations (**Fig. 9a**).

The algorithm starts with initial substrates (G0) and applies mechanistic transformations to generate successive product generations (G1 to Gn) (**Figs. 9c and 10a**), forming expanding reaction networks<sup>52,54,55</sup>. Moreover, atom reuse is limited to twice per substrate to prevent excessive polymerization and control network complexity<sup>5</sup>.

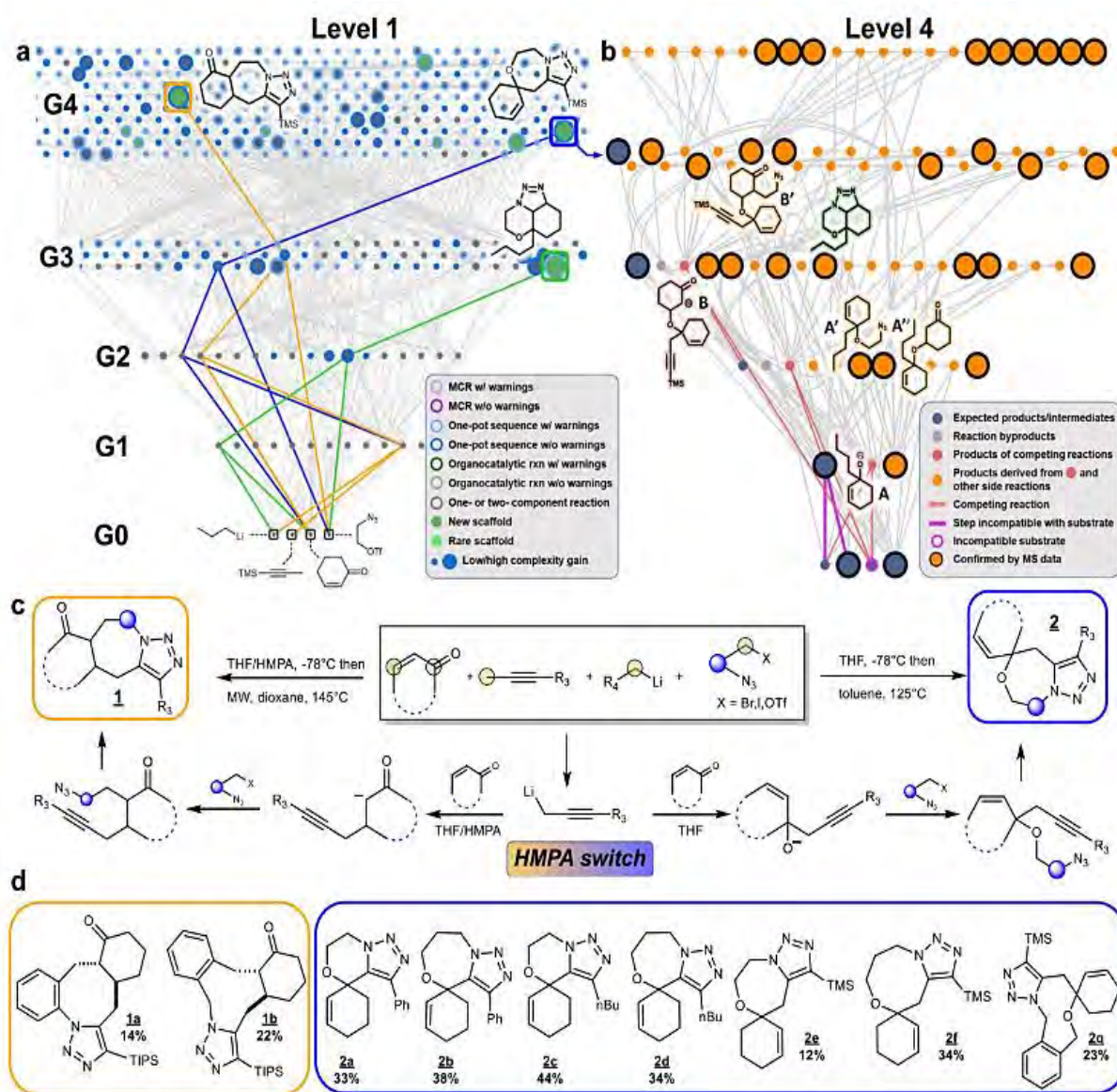


**Fig. 9** | Key elements of the MECH algorithm to discover MCRs. **a** Examples of simple starting materials selected from a collection of around 2,400 compounds. **b** A simplified illustration of one of roughly 8,000 mechanistic transforms, here, an E1cB elimination. The transform is designed to accommodate various substituents at different positions, with examples listed to the right of the reaction scheme. While the transform's record includes classifications of reaction conditions, rates, and other parameters, these details are omitted here for clarity. **c** Mechanistic transformation algorithms iteratively applied to a predefined set of starting materials. In the schematic representation provided, the three circle markers in the bottom row (G0) may be the three molecules from panel a, with the network systematically expanded to G4. Inter-nodal connections are color-coded to distinguish distinct reaction conditions. The corresponding reaction sequences are highlighted in dark blue to denote their synthetic relevance. **d** Such a sequence is expanded sideways, to perform analyses at Levels 2 and higher. Level 2 – branching-out of the main path to include by-products (gray) and products of competing/side reactions possible under the same class of reaction conditions (red); Level 3 – further branching to account for the reactions between side- and by-products. At Level 3 (and higher, not shown here for clarity), undesired reactions of side-/by-products with each other and with the members of the main pathway are also considered and marked in orange. *The figure was adapted from Publication P01.*

To qualify as suitable candidates for MCR or one-pot reactions, mechanistic sequences must satisfy the conditions for each individual step being compatible. The algorithm also carefully evaluates whether any incompatibilities arise between intermediates and unreacted substrates. If all materials contributing to the final product can coexist in the reaction vessel from the outset, the pathway is classified as a feasible MCR. However, if particular intermediates show undesired interactions with certain substrates, the algorithm recommends a one-pot strategy with stepwise addition of the reactive component.

One-pot reactions, non-MCR sequence:

Taking into account all of the above, the one-pot pathways I investigated start with enone, alkyllithium, azidohalide, and alkyne. The mechanistic network expanded to G4 (**Fig. 10a**) encompasses pathways that match the conditions, leading to 391 products with MW<500. Notably, two compounds in G4 represent novel tricyclic structures, 1 and 2, both showing substantial increases in complexity with each step compared to the starting materials. Scaffold 2, in particular, contains a spiro system resembling motifs found in certain drugs and bioactive compounds<sup>56-58</sup>. The routes to these products diverge at the first step.



**Fig. 10** | Example of algorithmically-discovered one-pot sequences and the corresponding mechanistic network expanded to Level 4. **a** A screenshot of the Level 1 network, expanded from cyclohexenone, trimethylsilylpropyne, nBuLi, and azidotriflate substrates, is presented. It maps all mutually compatible reaction pathways feasible under diverse conditions. Node sizes are proportional to complexity increase per mechanistic step. Colors of the halos distinguish MCR/one-pot sequences with or without warnings. Green-filled nodes represent scaffolds not previously reported in the literature. Within this network, two sequences (traced in blue and orange) up to G4 are predicted to be one-pot and leading to novel scaffolds **1** and **2**. A path to another complex scaffold in G3 is also marked (in green). This product is predicted to form via cyclization of the 1,2-adduct derived from nBuLi, cyclohexanone, and azide onto the double bond, and was detected by ESI-MS in the reaction mixture (structure highlighted in green in the L4 network in panel b). **b** A screenshot of the network extending from the blue pathway in panel a

and analyzed at Level 4 is shown. This Level 4 network encompasses various by-products (gray nodes) and side-products (red nodes), along with their subsequent transformations (products in orange), both among themselves and with intermediates from the parent pathway. Larger orange nodes likely represent structural assignments corresponding to peaks detected in the ESI-MS analysis of the crude reaction mixture. Interestingly, although the peaks corresponding to some predicted byproducts (e.g., A, B; structures drawn here with pink highlights) were not manifest in the ESI-MS spectra, their formation is corroborated by further products (A', A'', B'; structures drawn with orange highlights) that can only be derived from these undetected species. Also, the key cross-reactivity mandating sequential addition of reagents rather than MCR (i.e., reaction of alkyllithium with enone during metalation of alkyne) is highlighted by brighter pink connections at the bottom of the network. **c** General scheme depicting intermediates and reaction conditions for the blue and orange one-pot pathways leading to scaffolds 1 and 2 (as shown in panel a) is provided. In the substrates, the available nucleophilic and electrophilic sites are marked yellow and green, respectively, while the dark blue circle and the dotted arcs denote linkers between the azide and (pseudo)halides and a cyclic or acyclic fragment of the enone, respectively. The regioselectivity of addition (1,2- vs 1,4-) of propargyllithium reagent is controlled by the addition of HMPA as co-solvent. **d** Specific derivatives 1a, 1b and 2a–2g synthesized according to the general protocol along with the corresponding isolated yields. Note that the yields are low, as indeed predicted by the algorithm. Compounds 1a and 1b were isolated as single diastereoisomers. THF tetrahydrofuran, HMPA hexamethylphosphoramide, MW microwave, OTf triflate, TMS trimethylsilyl, TIPS triisopropylsilyl. *The figure was adapted from Publication P01.*

In **Fig. 10a**, the pathway traced in blue is predicted to proceed via 1,2-addition, alkoxide intermediate formation, O-alkylation, and a click reaction that closes two rings. In contrast, the pathway traced in orange begins with a 1,4 Michael-type addition, generating a carbanion at the alpha carbon, followed by C-alkylation and a click reaction. The algorithm predicts that: (1) these pathways are feasible only as one-pot reactions, with the enone added after complete consumption of the alkyllithium substrate; (2) the initial steps in both pathways can be carried out using propargyllithium, with HMPA acting as a switch (**Fig. 10c**) to favor the 1,4 addition<sup>59</sup>, and (3) both sequences are expected to result in low yields, approximately 20–40%.

Experimental outcomes confirmed these predictions, with isolated yields of derivatives 1a, 1b, and 2a–2g (**Fig. 10d**) ranging from 12% to 44%. Importantly, the predicted competing reaction modes aligned with ESI-MS data. In **Fig. 10b**, larger orange nodes correspond to side- or by-products with masses matching those observed in the spectra.

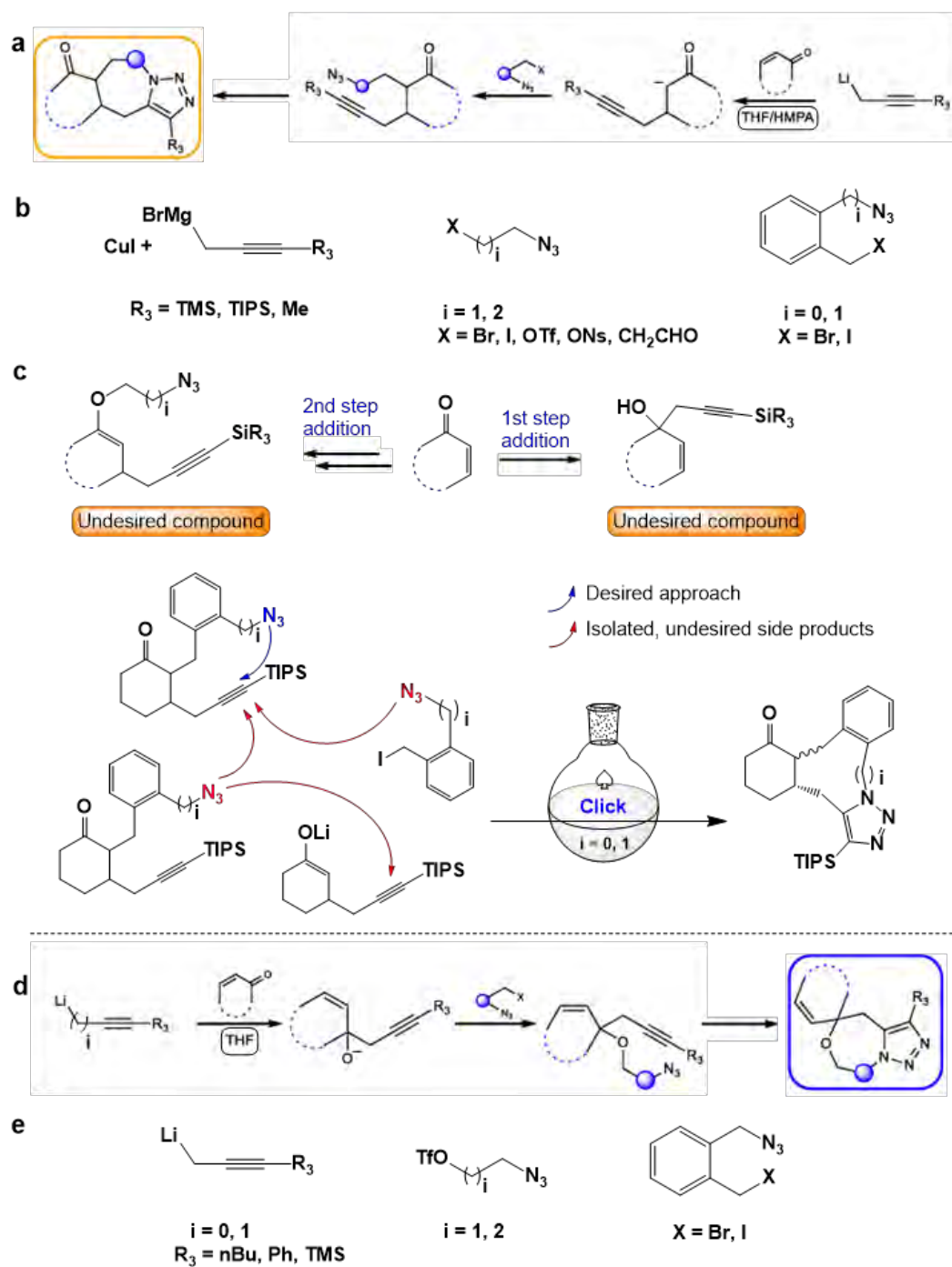


The current version of the **Mech** module encompasses a wide array of acid-base catalyzed steps, including those mediated by Lewis acids, as well as substitutions, eliminations, additions, rearrangements, pericyclic reactions, and fundamental transformations catalyzed by TMs (e.g., mechanistic steps in Suzuki, Buchwald-Hartwig, Heck, and Pauson-Khand reactions). Basic carbocationic chemistry is incorporated, though not exhaustively (a separate different module is built for such kind of chemistry). Radical mechanisms are not yet included, as their accurate application necessitates generalization, along with the potential need for additional heuristics based on thermodynamic and molecular-mechanical considerations (as outlined in the HopCat paper<sup>53</sup>).

Experimental fine-tuning:

Initially, the reaction conditions required careful optimization. Halogenated substrates (e.g., bromide and iodide), as well as functional groups such as triflate (OTf), nosylate (ONs), and even aldehyde moieties present in the azido-substrate (**Fig. 11b**), proved suboptimal. These groups either functioned poorly as leaving groups or led to undesired side products (**Fig. 11c**). For instance, issues arose in the form of either failed formation of the Michael donor in the first step, leading to 1,2-Nu attack instead of 1,4-addition, or undesired products in the second step due to the high reactivity of the resulting enolate species toward aliphatic azides. Consequently, alternative strategies had to be explored.

As illustrated, a key challenge was the effective introduction of the azido group into the substrate. Several experiments were conducted to assess the reactivity of the carbanion generated in the first step. Allyl iodide emerged as the most effective electrophile due to its high electrophilicity. Drawing on this reactivity profile, benzyl bromide was identified as a suitable analog for azido group installation, particularly because ortho-functionalization with an azido group is more synthetically accessible. Utilizing substrates such as 1-azido-2-(iodomethyl)benzene or 1-(azidomethyl)-2-(iodomethyl)benzene resulted in high-yield conversion in the second step. These electrophiles demonstrated excellent reactivity, facilitating efficient cyclization and enabling the completion of the target transformation. Initially, the entire sequence was tested in a stepwise manner. Upon confirming the successful formation of the desired product through this approach, a one-pot cascade strategy was implemented. This ultimately yielded the target compounds via a conjugate addition mechanism, representing a novel outcome.



**Fig. 11** | 1,2- vs. 1,4-Addition Pathways. **b.** The substrates' selection that was tested for the conjugated addition route. **c.** The desired vs. the competing side products that were detected and isolated. **e.** The substrates' selection that was tested for the direct addition route.

The second cascade (**Fig. 11d**) differs fundamentally from the first in the mode of organolithium addition: it proceeds via a 1,2-addition mechanism, in contrast to the 1,4-conjugate addition observed in the first cascade. The appropriate azido-substituted substrate was synthesized through straightforward synthetic procedures. In these 2 routes, HMPA played a vital role in the site of attack. HMPA exerts two distinct effects on the lithium species in conjugate addition reactions: (1) it promotes ion pair dissociation, thereby enhancing 1,4-addition selectivity; and (2) it reduces the Lewis acidity and catalytic activity of the lithium cation, which further favors the 1,4-addition pathway. (**Fig. 11a**)

A critical parameter in this cascade is the temperature at which the organolithium reagent is added. The addition was carried out at  $-20\text{ }^{\circ}\text{C}$ , which was also the optimal temperature for the second step involving the azido-triflate intermediate. At temperatures lower than  $-20\text{ }^{\circ}\text{C}$ , such as  $-78\text{ }^{\circ}\text{C}$ , the conversion to the second-step product was negligible. In contrast, at temperatures above  $-20\text{ }^{\circ}\text{C}$ , undesirable polymerization occurred—likely due to the use of tetrahydrofuran (THF) as solvent, wherein alkyl triflates can catalyze ring-opening polymerization. Therefore, the second step was conducted at  $-20\text{ }^{\circ}\text{C}$  with overnight stirring, affording the desired product in 70% overall yield after two steps. As in the first cascade, a range of starting materials was screened for both steps in order to identify the most reactive substrates and to minimize the formation of side products. Ending up with very similar to exact kind of the first-cascade's substrates. (**Fig. 11e**)

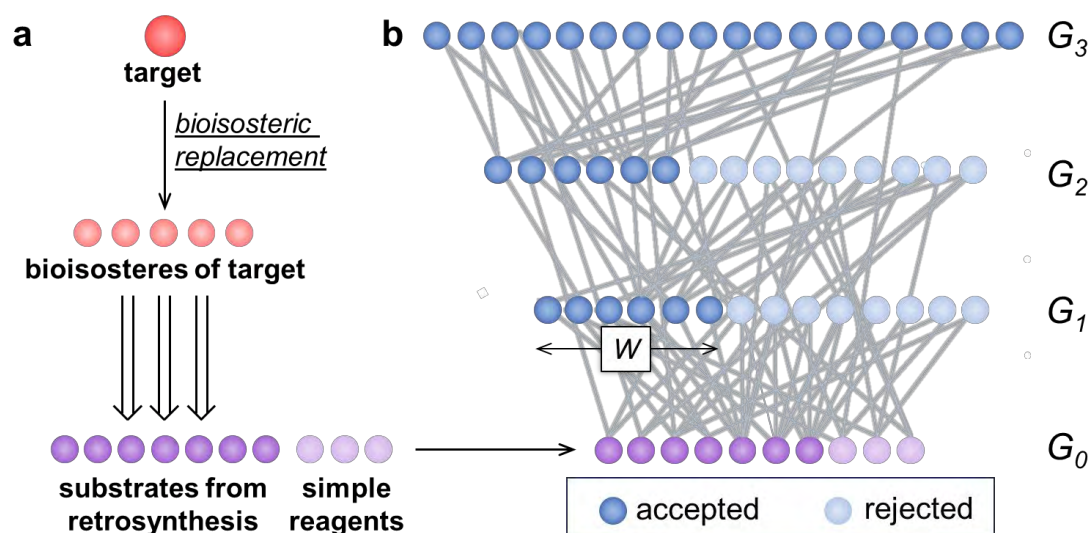
## 5.2. Validation of the Analog Module: Drug development and analogs generation

The primary aim of my study here was not to endorse or critique any specific methodology in drug development but to experimentally evaluate our analog-design pipeline (i.e., **Analog** module<sup>4</sup>), which follows the "classical" *in silico* synthesis paradigm. Using bioisosteric substitutions, we diversify the parent scaffold to potentially enhance biological activity. As discussed earlier in the Introduction, the workflow includes retrosynthetic analysis, followed by guided forward synthesis to generate diverse analogs, and computational binding affinity predictions. The two central questions that I focused on were: the accuracy of computer-designed analog synthesis and the reliability of predicting binding affinities.

The analog-design algorithm relies on guided reaction networks, a concept explored in the group's prior work<sup>52,54,60</sup> and similar to the Gn generations used in the **Mech** module. The second key element is retrosynthesis and substrate selection, focusing on commercially available G0 starting materials. These must be chosen strategically to maximize the generation of structural analogs while avoiding subjective human bias through automation. The substrate set should be small enough to prevent overwhelming the system but must meet two criteria: (i) diversity, incorporating the parent compound's core motifs and related structures, and (ii) synthetic flexibility, ensuring the building blocks can participate in diverse reactions, including functionalization. This approach balances efficiency and versatility in analog design.

Considering these factors, simply disconnecting the parent compound via retrosynthesis may be too limited, as it typically yields substrates that only reconstruct the original molecule, its intermediates, or alternative by-products. Initially, this method was used, but many resulting "similar" compounds had incompatible functional groups, complicating forward synthesis. A more nuanced approach aims to avoid these synthetic challenges by adjusting the retrosynthetic protocol into two important ways (**Fig. 12**).

Specifically, by (i) using bioisosteric replacements to improve biological/physical properties while maintaining functional groups for "mix-and-match" reactivity, enhancing structural diversity; and (ii) by expanding the G0 set of substrates with a minimal set of 23 versatile chemicals to enable versatile molecular modifications in forward networks, selected for their broad synthetic utility.



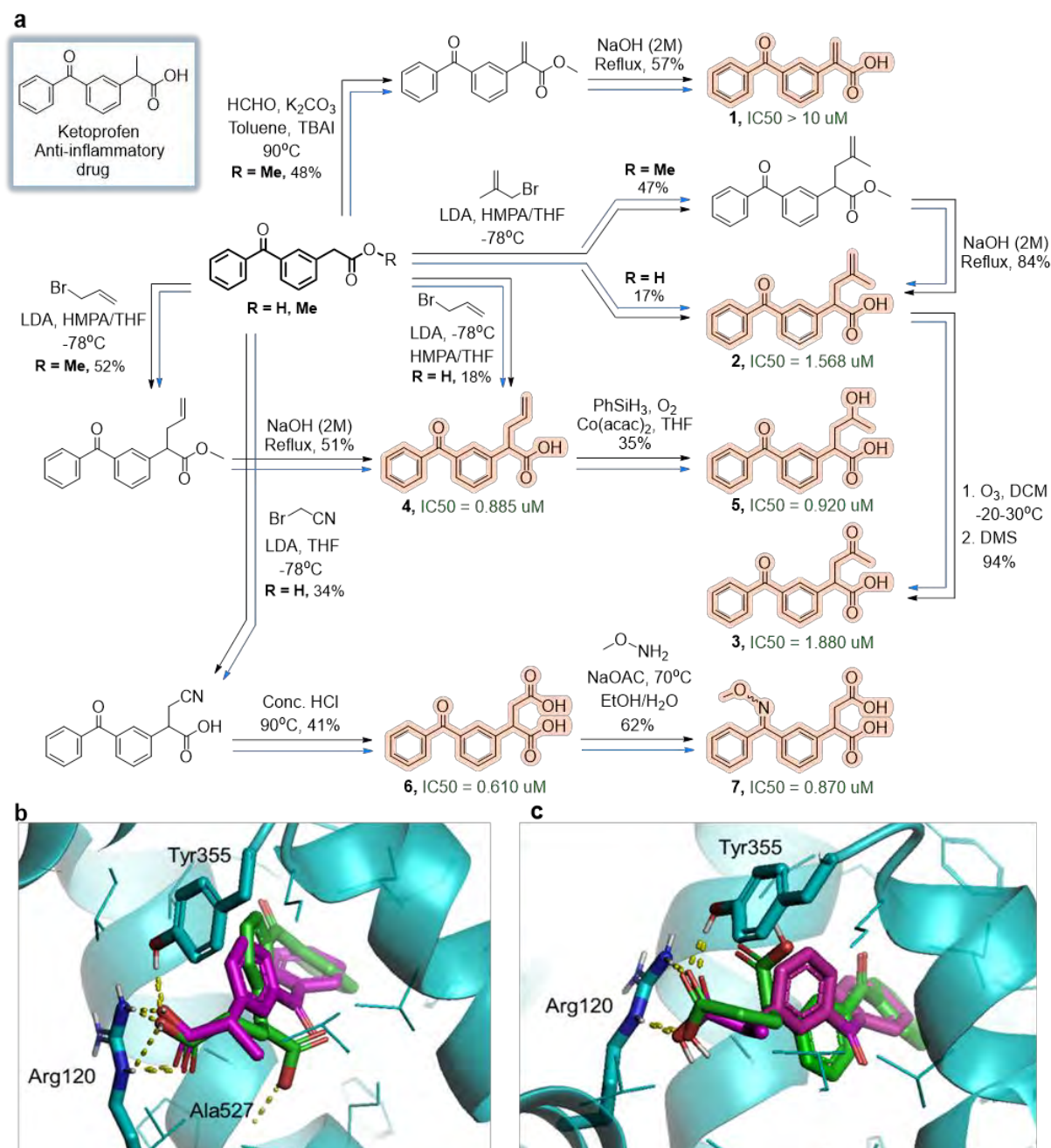
**Fig. 12** | Algorithmic pipeline for substrate selection and subsequent analog generation. **a** For a given molecule (red node), the algorithm identifies modifiable substructures suitable for bioisosteric replacements. It then substitutes these fragments with appropriate bioisosteres, generating multiple parent variants (light red nodes), typically ranging from 10 to 100 for a drug-like compound. Next, retrosynthetic networks are expanded to determine feasible synthetic routes for all generated analogs, retaining only commercially available starting materials. The process halts upon reaching commercially available substrates. Additionally, the set of retrosynthetically derived substrates (violet nodes) is supplemented with 23 simple yet synthetically valuable compounds (light violet nodes). **b** All selected substrates constitute the zeroth generation ( $G_0$ ) for the guided forward search. In this process, after each reaction cycle, only a subset of  $W$  molecules most structurally similar to the parent compound is retained (dark-blue nodes). This iterative refinement directs the synthesis progressively toward close analogs of the parent structure and prevents uncontrolled network expansion, maintaining computational efficiency. *The figure was adapted from Publication P02.*

My work, centered on two parent molecules, i.e., Ketoprofen and Donepezil—Through experimental validation I confirmed the successful execution of concise, computer-designed syntheses for seven Ketoprofen analogs and five Donepezil analogs (with only one synthetic pathway failing). Among the Ketoprofen analogs, six exhibited micromolar binding affinities to human cyclooxygenase-2 (COX-2), including one analog with slightly enhanced binding relative to the parent compound (0.61  $\mu\text{M}$  vs. 0.69  $\mu\text{M}$ ). Similarly, all five Donepezil analogs demonstrated submicromolar binding affinities to acetylcholinesterase (AChE) derived from *Electrophorus electricus*, with one analog

achieving nanomolar affinity comparable to that of the parent compound (36 nM vs. 21 nM).

Nonetheless, binding affinity predictions—generated through three docking programs and a neural network to guide analog selection for synthesis validation—aligned with experimental measurements only within an order of magnitude. Based on these findings, I can conclude that (i) the synthesis-planning components of contemporary computational analog design approaches are highly reliable, but (ii) standard affinity-prediction tools are effective in identifying promising binders yet inadequate for differentiating between moderate ( $\mu\text{M}$ ) and high-affinity (nM) candidates with precision.

**Ketoprofen analogs.** I focused on molecules **1-7**, forming a compact synthetic network depicted in **Fig. 13a**. These analogs are derived from the commercially available 2-(3-benzoylphenyl)acetic acid and had not been previously known or tested for Cyclooxygenase-2 (COX-2) binding (compound **1** was synthesized in this ref.<sup>61</sup> but in the context of hydrocarboxylation methodology, without any biological investigations). The proposed syntheses are straightforward and efficient, with all reactions successfully carried out, yielding compounds as indicated next to the arrows (under conditions identical or closely resembling those suggested by the software, without extensive optimization). In selecting these particular analogs, we were also guided by predictions of their binding affinities to COX-2. To this end, we utilized three established docking programs (AutoDock 4, AutoDock Vina, and Dock 6)<sup>62-65</sup> in addition to Allchemy's neural network. All programs employed the 5IKR PDB structure, and the binding scores for analogs **1-7** (averaged over possible stereoisomers of each racemic analog) were predicted to be comparable to or more favorable than that of Ketoprofen, which was also docked for reference. To validate these predictions, we conducted a spectrofluorometric COX-2 human inhibition assay<sup>66</sup> and measured the IC<sub>50</sub> values, which are indicated in green next to the respective analogs in **Fig. 13a**. As shown, one analog **1** exhibited poor binding ( $> 10 \mu\text{M}$ ), while the remaining six analogs demonstrated micromolar binding, with compound **6** showing slightly better affinity than the parent Ketoprofen ( $0.61 \mu\text{M}$  vs.  $0.69 \mu\text{M}$ ).



**Fig. 13** | Ketoprofen's analogs. **a** A small network of algorithm-guided syntheses for analogs **1-7** is shown, originating from 2-(3-benzoylphenyl)acetic acid and its ester derivatives ( $R = H, Me$ ). Experimentally confirmed pathways (black arrows) align with the algorithm's initial predictions (blue arrows). Reaction conditions and isolated yields are specified for each transformation, with  $IC_{50}$  listed next to the respective analogs. **b & c** All analogs were docked into COX-2 protein. The alignment of docking poses from AutoDock 4 for **b** (*S*)-Ketoprofen and its analog (*S*)-**6**, and **c** (*R*)-Ketoprofen and its analog (*R*)-**6**. Key, protein-ligand hydrogen bonds are depicted as yellow dotted lines. The carboxylic acid moiety of (*S*)-Ketoprofen (colored in *magenta*) forms two hydrogen bonds with Arg120 and one with Tyr355. Its analog (*S*)-**6** (b, colored in *green*) is predicted to be

similarly positioned inside the active site and to form three hydrogen bonds: with Arg120, with Tyr355 and, for the second carboxylic acid moiety, with Ala527. For (*R*)-Ketoprofen (colored in *magenta* in panel c) and its analog (**R**)-**6** (colored in *green*), the carboxylic acid moieties in both compounds form two hydrogen bonds (with Arg120 and with Tyr355). *The figure was adapted from Publication P02.*

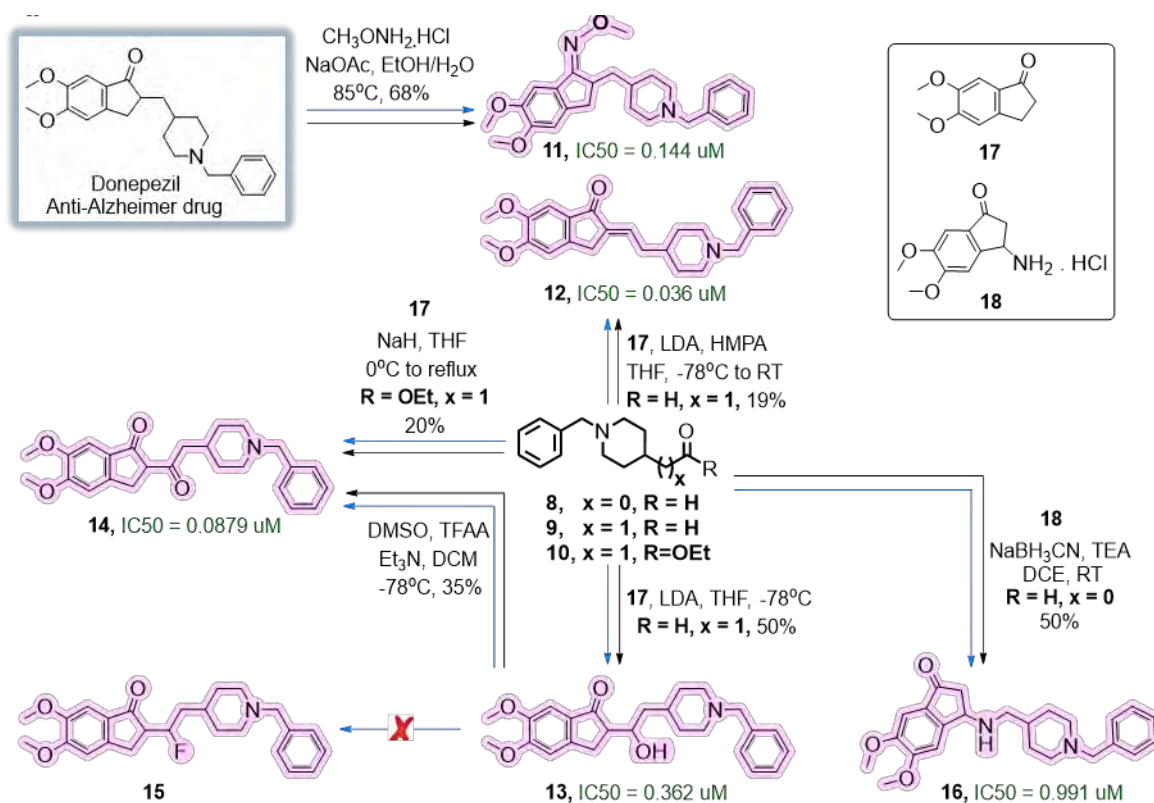
**Donepezil analogs.** A similar protocol to the one used for Ketoprofen was employed. The algorithm identified 18 bioisosteric variants of the Donepezil parent compound. Unlike the Ketoprofen case, disconnections were made more centrally, resulting in substrates such as 5,6-dimethoxy-2,3-dihydro-1H-inden-1-one **17**, as well as 2-(1-benzylpiperidin-4-yl)ethanol and (1-benzylpiperidin-4-yl)acetic acid. These compounds can be readily converted into aldehyde **9**, the central substrate utilized in most of the syntheses, and ester **10**, respectively.

Six compounds were synthesized, as depicted in **Fig. 14**: bioisostere **11** of the parent compound and analogs **12-15**, forming a compact "cluster" that required only six reactions based on the aldehyde **9**, ester **10**, and ketone **17**. Also, the synthesis of analog **16** was produced, identified by an earlier version of the algorithm, using **8** and **18** as starting materials. All reactions were conducted under the conditions recommended by the algorithm, with the exception of the oxidation of **13** to **14**, where Swern oxidation was employed instead of Dess-Martin oxidation. In five out of the six reactions, the expected products were obtained, although the yields were poor to moderate (no optimization of reaction conditions was attempted). In one case, the fluorinated derivative **15** was not formed due to elimination to **12**. While there are reported instances of fluorination occurring on  $\beta$ -hydroxy ketones, Allchemy's pKa model<sup>67</sup> suggests that the C-H position in our cyclic aryl ketone is more acidic, favoring elimination over substitution.

In a manner similar to Ketoprofen, these analogs were chosen primarily because most have not been previously documented (compound **11**, although recognized as an intermediate in the synthesis of acetylcholinesterase (AChE) inhibitors, has not had its biological activity evaluated<sup>68</sup>). Furthermore, their predicted binding affinities (calculated using AutoDock 4, AutoDock Vina, and Dock 6) were either comparable to or exceeded those of the parent compound, Donepezil, which was also docked for reference. Additionally, predictions from Allchemy's neural network suggested that these analogs would display micromolar binding affinities.



The predicted binding affinities were experimentally confirmed using a spectrophotometric AChE inhibition assay<sup>66</sup>, which measured the IC<sub>50</sub> values listed alongside each analog in **Fig. 14**. The bioisostere **11** showed an IC<sub>50</sub> of 144 nM, while two analogs exhibited submicromolar potency: **16** at 991 nM and **13** at 362 nM. In contrast, the other two analogs displayed substantially higher potency, with IC<sub>50</sub> values of 88 nM for **14** and 36 nM for **12**. The latter is particularly noteworthy, as its potency approaches that of Donepezil, which has an IC<sub>50</sub> of 21 nM.



**Fig. 14** | Donepezil's analogs. A small network of algorithm-guided syntheses of analogs **11-16**. Blue arrows highlight Allchemy's suggestions, black arrows highlight the experimentally executed pathways. Conditions and isolated yields are given above reaction arrows. IC<sub>50</sub> values are given next to the analogs. *Note*: As synthesis of **9** and **10** from simpler starting materials was previously described, we followed the literature procedure. All analogs were docked into acetylcholinesterase from *Electrophorus electricus*. The figure was adapted from Publication P02.

Therefore, the findings above lead to three primary conclusions. First, the accuracy of the synthetic predictions is commendable, as 12 out of 13 analogs were successfully synthesized according to the computer-generated pathways. This result aligns with prior validations of prediction tools such as other modules in Allchemy and Chematica/Synthia, which have demonstrated reliability on targets of comparable or greater complexity<sup>1,53</sup>.

In this regard, the computer serves as an effective "calculator," streamlining the design process by facilitating tasks that are otherwise tedious, including bioisostere generation, selection of commercially available starting materials, and planning of the forward synthetic route.

Secondly, while the synthesis process is reliable, accurately predicting analog properties remains difficult. Allchemy's neural network, trained on around 2 million ChEMBL protein assay entries, correctly predicted micromolar potency for most compounds, though not all. For example, compound **1** had a potency over 10  $\mu\text{M}$ , whereas compounds **12** and **14** showed low nanomolar IC<sub>50</sub> values. These outcomes indicate that neither docking simulations nor the neural network can precisely predict binding affinity. However, their estimates are still useful for broadly distinguishing weak binders from those with moderate activity, offering valuable early insights despite limited accuracy.

Finally, it is essential to acknowledge that in the development of potential drug candidates, binding affinity is only one of many critical parameters. Other factors, such as ADME-Tox (Absorption, Distribution, Metabolism, Excretion, and Toxicity), must also be considered. Numerous machine learning models are currently available to assess these properties (e.g., references<sup>69-71</sup> for hERG models, reference<sup>72</sup> for plasma protein binding (PPB), and references<sup>73-75</sup> for blood-brain barrier (BBB) penetration).

## 6. SUMMARY

Computational tools such as the one I used in my doctoral work are not intended to replace the critical role of chemists or to bypass the necessity of optimizing reaction parameters, including temperature control and solvent choice. However, without any doubt, these tools are opening up new horizons for synthesis planning, reaction discovery, and also the development of bioactive substances and, in the fullness of time, drugs. In my own work and in other examples in article *P01*, Allchemy-Mech has provided the first ever set of examples of computer designing mechanistically unprecedented MCRs, there is every chance that this algorithm and other similar to it will enable the identification of multicomponent reactions (MCRs) in sufficient quantities to have a transformative impact on synthetic chemistry. This said, potential future enhancements may involve integrating radical-based reaction pathways, or broadening the scope of catalytic transformations. In addition, the pursuit of one-pot synthesis methodologies will continue to drive innovation across chemistry, enzymology, materials science, and mechanistic studies. Researchers continue to actively explore one-pot synthesis methods, motivated by both cost savings and environmental benefits.

A major challenge in generating analogs is being able to predict their properties early on—ideally before the synthesis takes place. As highlighted in article *P02*, while docking tools cannot provide 100% accurate binding affinity predictions, they still play an important role by quickly filtering out compounds that are unlikely to bind well. Today's machine learning models<sup>69-75</sup> have evolved to assess much more than just binding potential—they are now also used to predict key factors in drug discovery, such as toxicity, absorption, and how a drug is distributed in the body. The Allchemy-Analog module integrates many of these predictive tools to support the process, and that helped saving time and effort in the execution phase. Moreover, while drug-likeness scores can be helpful indicators during early screening, they do not guarantee a compound's success. Many potential drug candidates still fail later in development because of unpredictable biological issues—like unexpected toxicity or insufficient effectiveness—despite having promising molecular features. This underscores why ADME-Tox properties remains indispensable.

In general, the field of synthetic chemistry stands at an interesting point, where incorporating computational tools into routine organic synthesis and medicinal chemistry is not just possible, but it is becoming essential for progress.

## 7. BIBLIOGRAPHY

1. Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; et al. (2018). Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem.* *4*, 522-532.
2. Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. (2020). REINVENT 2.0: An AI Tool for De Novo Drug Design. *J. Chem. Inf. Model* *60*, 5918-5922.
3. Molga, K.; Beker, W.; Roszak, R.; Czerwiński, A.; Grzybowski, B.A. (2025). Hierarchical reaction logic enables computational design of complex peptide syntheses. *J. Am. Chem. Soc.* *147*, 7644–7662.
4. Makkawi, A.; Wiktor, B.; Wołos, A.; et al. (2025). Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chemical Science* *16*, 8383-8393.
5. Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; et al. (2024). Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Commun.* *15*, 10285.
6. Hulme, C.; Gore, V. (2003). Multi-component reactions: emerging chemistry in drug discovery ‘from xylocain to crixivan’. *Curr. Med. Chem.* *10*, 51-80.
7. Lee, J.M.; Na, Y.; Han, H.; Chang, S. (2004). Cooperative multi-catalyst systems for one-pot organic transformations. *Chem. Soc. Rev.* *33*, 302-312.
8. Bruggink, A.; Schoevaart, R.; Kieboom, T. (2003). Concepts of Nature in Organic Synthesis: Cascade Catalysis and Multistep Conversions in Concert. *Organic Process Research & Development* *7*, 622-640.
9. Burns, N.Z.; Baran, P.S.; Hoffmann, R.W. (2009). Redox economy in organic synthesis. *Angew. Chem. Int. Ed.* *48*, 2854-2867.
10. Ugi, I. (2001). Recent progress in the chemistry of multicomponent reactions. *Pure Appl. Chem.* *73*, 187-191.
11. Zhao, W.; Chen, F.E. (2012). One-pot Synthesis and its Practical Application in Pharmaceutical Industry. *Current Organic Synthesis* *9*, 873-897.
12. Wang, W.; Xu, P.F. (2013). *Catalytic Cascade Reactions*, John Wiley and Sons Inc., Hoboken, New Jersey.

13. **(a)** Weber, M.; Frey, W.; Peters, R. (2013). Asymmetric palladium(II)-catalyzed cascade reaction giving quaternary amino succinimides by 1,4-addition and a Nef-type reaction. *Angew. Chem. Int. Ed.* *52*, 13223-13227. **(b)** Qiu, L.; Guo, X.; Ma, C.; et al. (2014). Efficient synthesis of chiral cyclic acetals by metal and Brønsted acid co-catalyzed enantioselective four-component cascade reactions. *Chem. Commun.* *50*, 2196-2198. **(c)** Cala, L.; Mendoza, A.; Fananas, F.J.; Rodriguez, F. (2013). A catalytic multicomponent coupling reaction for the enantioselective synthesis of spiroacetals. *Chem. Commun.* *49*, 2715-2717.
14. W. H. Organization. (2011). WHO model list of essential medicines, 17th edition. Geneva.
15. Dömling, A.; Houry, K. (2010). Praziquantel and Schistosomiasis. *ChemMedChem.* *5*, 1420-1434.
16. Cao, H.; Liu, H.; Dömling, A. (2010). Efficient multicomponent reaction synthesis of the schistosomiasis drug Praziquantel. *Chem. Eur. J.* *16*, 12296-12298.
17. Trost, B.M. (1995). Atom Economy--a challenge for organic synthesis: homogeneous catalysis leads the way. *Angew. Chem. Int. Ed.* *34*, 259-281. **(b)** Trost, B.M. (1991). The Atom Economy--a search for synthetic efficiency. *Science* *254*, 1471-1477.
18. Dömling, A.; Ugi, I. (2000). Multicomponent reactions with isocyanides. *Angew. Chem. Int. Ed.* *39*, 3168-3210.
19. Ganem, B. (2009). Strategies for innovation in multicomponent reaction design. *Acc. Chem. Res.* *42*, 463-472.
20. Wheeldon, I.; Minter, S.D.; Banta, S.; Barton, S.C.; Atanassov, P.; Sigman, M. (2016). Substrate channeling as an approach to cascade reactions. *Nat. Chem.* *8*, 299-309.
21. Hayashi, Y. (2021). Time and pot economy in total synthesis. *Acc. Chem. Res.* *54*, 1385-1398.
22. Cioc, R.C.; Ruijter, E.; Orru, R.V.A. (2014). Multicomponent reactions: advanced tools for sustainable organic synthesis. *Green Chem.* *16*, 2958-2975.
23. Ruijter, E.; Scheffelaar, R.; Orru, R.V.A. (2011). Multicomponent reaction design in the quest for molecular complexity and diversity. *Angew. Chem. Int. Ed.* *50*, 6234-6246.

24. Hayashi, H.; Katsuyama, H.; Takano, H.; Harabuchi, Y.; Maeda, S.; Mita, T. (2022). In silico reaction screening with difluorocarbene for N-difluoroalkylative dearomatization of pyridines. *Nat. Synth. 1*, 804-814.
25. Medley, J.W.; Movassaghi, M. (2013). Robinson's landmark synthesis of tropinone. *Chem. Comm. 49*, 10775-10777.
26. Schneider, N.; Lowe, D.M.; Sayle, R.A.; Landrum, G.A. (2015). Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *J. Chem. Inf. Model. 55*, 39-53.
27. Lipinski, C.A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol. 1*, 337-341.
28. Shultz, M.D. (2018). Two decades under the influence of the rule of five and the changing properties of approved oral drugs. *J. Med. Chem. 62*, 1701-1714.
29. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; Muresan, S.; Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. *Nat. Chem. 4*, 90-98.
30. Ursu, O.; Rayan, A.; Goldblum, A.; Oprea, T.I. (2011). Understanding drug-likeness. *Comp. Mol. Sci. 1*, 760-781.
31. Oprea, T.I. (2000). Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des. 14*, 251-264.
32. Leeson, P.D.; Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov. 6*, 881-890.
33. Hann, M.M.; Oprea, T.I. (2004). Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol 8*, 255-263.
34. Bohacek, R.S.; McMartin, C.; Guida, W.C. (1996). The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev. 16*, 3-50.
35. Das, A.P.; Mathur, P.; Agarwal, S.M. (2024). Machine Learning, Molecular Docking, and Dynamics-Based Computational Identification of Potential Inhibitors against Lung Cancer. *ACS Omega 9*, 4528-4539.
36. Jia, P.; Pei, J.; Wang, G.; Pan, X.; Zhu, Y.; Wu, Y.; Ouyang, L. (2022). The roles of computer-aided drug synthesis in drug development. *Green Synthesis and Catalysis 3*, 11-24.
37. Graff, D.E.; Shakhnovich E.I.; Coley, C.W. (2021). Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci. 12*, 7866-7881.

38. Bradshaw, J.; Paige, B.; Kusner, M.J.; Segler, M.H.S.; Hernández-Lobato, J.M. (2019). A Model to Search for Synthesizable Molecules. arXiv [1906.05221].
39. Fromer, J.C.; Graff, D.E.; Coley, C.W. (2024). Pareto optimization to accelerate multi-objective virtual screening. *Digital Discovery* 3, 467-481.
40. **(a)** Corey, E.J.; Wipke, W.T. (1969). Computer-Assisted Design of Complex Organic Syntheses: pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. *Science* 166, 178-192. **(b)** Corey, E.J.; Wipke, W.T.; Cramer, R.D.; Howe, J. (1972). Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics. *J. Am. Chem. Soc.* 94, 421-430.
41. **(a)** Winter, J.H. (1982). *Chemische Syntheseplanung*. Springer-Verlag: Heidelberg, New York, Berlin. **(b)** Corey E.J.; Cheng, X.M. (1989). *The Logic of Chemical Synthesis*. John Wiley & Sons: New York, Chichester, Brisbane, Toronto, Singapore.
42. Ravitz, O. (2013). Data-driven computer aided synthesis design. *Drug Discov. Today Technol.* 10, 443-449.
43. Weininger, D. (1988). SMILES, a Chemical Language and Information System. *J. Chem. Inf. Comput. Sci.* 28, 31-36.
44. Krebsbach, D.; Gelernter, H.; Sieburth, S.M. (1998). Distributed Heuristic Synthesis Search. *J. Chem. Inf. Comput. Sci.* 38, 595-604.
45. Johnson, P.Y.; Bernstein, I.; Crary, J.; Evans, M.; Wang, T. (1989). Designing an Expert System for Organic Synthesis in Expert Systems Application in Chemistry. ACS Symposium Series, Am. Chem. Soc., Washington.
46. Salatin, T.D.; Jorgensen, W.L. (1979). Computer-Assisted Mechanistic Evaluation of Organic Reactions. *JOC* 45, 2043-2051.
47. Fontain, E.; Reitsam, K. (1990). The Generation of Reaction Networks with RAIN. *J. Chem. Inf. Comput. Sci.* 31, 96-101.
48. **(a)** Gastegger, M.; Schütt, K.T.; Müller, K.R. (2021). Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.* 12, 11473-11483. **(b)** Keith, J.A.; et al. (2021). Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* 121, 9816-9872. **(c)** Dugundji, J.; Ugi, I. (1973). An algebraic model of constitutional chemistry as a basis for chemical computer programs. *Top. Curr. Chem.* 39, 19-64.

49. Kvasničk, V.; Pospíchal, A.J. (1989). Two metrics for a graph-theoretical model of organic chemistry. *Journal of Mathematical Chemistry* 3, 161-191.
50. Szymkuc, S.; Gajewska, E.P.; Klucznik, T.; Molga, K.; et al. (2016). Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* 55, 5904-5937.
51. Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A.A.; Popik, O.; Klucznik, T.; et al. (2020). Computational planning of the synthesis of complex natural products. *Nature* 588, 83-88.
52. Wołos, A.; Roszak, R.; Żądło-Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; et al. (2020). Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. *Science* 369, eaaw1955.
53. Klucznik, T.; Syntrivanis, L.D.; Baś, S.; Mikulak-Klucznik, B.; Moskal, M.; et al. (2024). Computational prediction of complex cationic rearrangement outcomes. *Nature* 625, 508-515.
54. Wołos, A.; Koszelewski, D.; Roszak, R.; Szymkuć, S.; Moskal, M.; et al. (2022). Computer-designed repurposing of chemical wastes into drugs. *Nature* 604, 668-676.
55. Gajewska, E.P.; Szymkuć, S.; Dittwald, P.; Startek, M.; Popik, O.; Mlynarski, J.; Grzybowski, B.A. (2020). Algorithmic discovery of tactical combinations for advanced organic syntheses. *Chem* 6, 280-293.
56. Chavan, S.R.; Gavale, K.S.; Khan, A.; et al. (2017). Iminosugars spiro-linked with morpholine-fused 1,2,3-triazole: Synthesis, conformational analysis, glycosidase inhibitory activity, antifungal assay, and docking studies. *ACS Omega* 2, 7203-7218.
57. Chen, D.; Lu, T.; Yan, Z.; Lu, W.; et al. (2019). Discovery, structural insight, and bioactivities of BY27 as a selective inhibitor of the second bromodomains of BET proteins. *Eur. J. Med. Chem.* 182, 111633.
58. Kimura, T.; et al. (2007). Multi-cyclic cinnamide derivatives. Patent US2007219181A1.
59. Sikorski, W.H.; Reich, H.J. (2001). The Regioselectivity of addition of organolithium reagents to enones and enals: The role of HMPA. *J. Am. Chem. Soc.* 123, 6527-6535.
60. Roszak, R.; Wołos, A.; Benke, M.; Gleń, Ł.; et al. (2024). Emergence of metabolic-like cycles in blockchain-orchestrated reaction networks. *Chem.* 10, 952-970.
61. Liu, D.; Ru, T.; Deng, Z.; Zhang, L.; Ning, Y.; Chen, F.E. (2023). Sulfonate-Modified Picolinamide Diphosphine: A Ligand for Room-Temperature Palladium-Catalyzed



- Hydrocarboxylation in Water with High Branched Selectivity. *ACS Catal.* *13*, 12868-12876.
62. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.;Goodsell, D.S.; Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* *16*, 2785-2791.
63. Trott, O.; Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* *31*, 455-461.
64. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* *61*, 3891-3898.
65. Allen, W.J.; Balius, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. (2015). DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* *36*, 1132-1156.
66. Asghar, S.; Mushtaq, N.; Ahmed, A.; et al. (2024). Potential of Tryptamine Derivatives as Multi-Target Directed Ligands for Alzheimer’s Disease: AChE, MAO-B, and COX-2 as Molecular Targets. *Molecules* *29*, 490.
67. Roszak, R.; Beker, W.; Molga, K.; Grzybowski, B.A. (2019). Rapid and accurate prediction of pKa values of C–H Acids using graph convolutional neural networks. *J. Am. Chem. Soc.* *141*, 17142-17149.
68. Sugimoto, H.; Iimura, Y.; Yamanishi, Y.; Yamatsu, K. (1995). Synthesis and Structure-Activity Relationships of Acetylcholinesterase Inhibitors: 1-Benzyl-4-[(5,6-dimethoxy-1-oxoindan-2-yl)methyl]piperidine Hydrochloride and Related Compounds. *J. Med. Chem.* *38*, 4821-4829.
69. Ogura, K.; Sato, T.; Yuki, H.; Honma, T. (2019). Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. *Sci. Rep.* *9*, 12220.
70. Kim, H.; Nam, H. (2020). hERG-Att: Self-attention-based deep neural network for predicting hERG blockers. *Comput. Biol. Chem.* *87*, 107286.
71. Zhang, X.; Mao, J.; Wei, M.; Qi, Y.; Zhang; J.Z.H. (2022). HergSPred: Accurate Classification of hERG Blockers/Nonblockers with Machine-Learning Models. *J. Chem. Inf. Model.* *62*, 1830-1839.

72. Han, Z.; Xia, Z.; Xia, J.; Tetko I.V.; Wu, S. (2025). The state-of-the-art machine learning model for plasma protein binding prediction: Computational modeling with OCHEM and experimental validation. *Eur. J. Pharm. Sci.* *204*, 106946.
73. Meng, F.; Xi, Y.; Huang, J.; Ayers, P.W. (2021). A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. *Sci. Data* *8*, 289.
74. Kumar, V.; Banerjee, A.; Roy, K. (2024). Breaking the Barriers: Machine-Learning-Based c-RASAR Approach for Accurate Blood–Brain Barrier Permeability Prediction. *J. Chem. Inf. Model.* *64*, 4298-4309.
75. Parakkal, S.C.; Datta, R.; Das, D. (2022). DeepBBBP: High Accuracy Blood-brain-barrier Permeability Prediction with a Mixed Deep Learning Model. *Mol. Inf.* *41*, 2100315.

## **8. ORIGINAL PUBLICATIONS**

# Systematic, computational discovery of multicomponent and one-pot reactions

Received: 15 November 2024

Accepted: 18 November 2024

Published online: 27 November 2024

Check for updates

Rafał Roszak<sup>1,6</sup>, Louis Gadina<sup>2,3,6</sup>, Agnieszka Wołos<sup>1,6</sup>, Ahmad Makkawi<sup>2</sup>, Barbara Mikulak-Klucznik<sup>1</sup>, Yasemin Bilgi<sup>2,3</sup>, Karol Molga<sup>1,2</sup>, Patrycja Gołębiowska<sup>2</sup>, Oskar Popik<sup>2</sup>, Tomasz Klucznik<sup>1</sup>, Sara Szymkuć<sup>1</sup>, Martyna Moskal<sup>1</sup>, Sebastian Baś<sup>2,4</sup>, Rafał Frydrych<sup>2,3</sup>, Jacek Mlynarski<sup>1,2</sup>, Olena Vakuliuk<sup>2</sup>, Daniel T. Gryko<sup>2</sup>✉ & Bartosz A. Grzybowski<sup>2,3,5</sup>✉

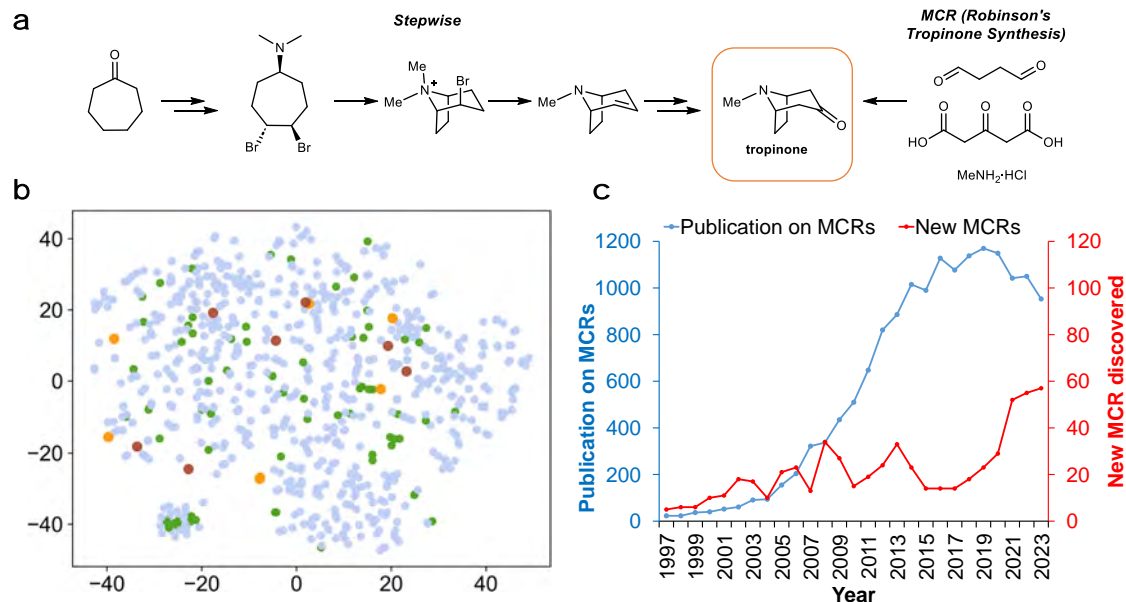
Discovery of new types of reactions is essential to organic chemistry because it expands the scope of accessible molecular scaffolds and can enable more economical syntheses of existing structures. In this context, the so-called multicomponent reactions, MCRs, are of particular interest because they can build complex scaffolds from multiple starting materials in just one step, without purification of intermediates. However, for over a century of active research, MCRs have been discovered rather than designed, and their number remains limited to only several hundred. This work demonstrates that computers taught the essential knowledge of reaction mechanisms and rules of physical-organic chemistry can design – completely autonomously and in large numbers – mechanistically distinct MCRs. Moreover, when supplemented by models to approximate kinetic rates, the algorithm can predict reaction yields and identify reactions that have potential for organocatalysis. These predictions are validated by experiments spanning different modes of reactivity and diverse product scaffolds.

Computational discovery of new reaction classes is one of the holy grails of chemoinformatics, with first efforts by Ivar Ugi<sup>1–4</sup> dating back to 1970s. In this context, reactions that build complex scaffolds from multiple simple components in one step (i.e., multicomponent reactions, MCRs<sup>5–11</sup>; Fig. 1a) and/or proceed sequentially in one pot<sup>12–14</sup> are of particular interest as they minimize separation and purification operations, and increase the overall step- and atom-economy<sup>15</sup> as well as “greenness”<sup>16,17</sup> of synthesis. However, the number of known MCR classes remains limited to several hundred (Fig. 1b, c), perhaps because the most popular reactivity patterns (e.g., isocyanide,  $\beta$ -dicarbonyl, or imine-based MCRs) and their straightforward combinations<sup>18</sup> and extensions<sup>19–21</sup> have been studied in nearly exhaustive detail. Rational discovery of MCRs remains difficult because it entails understanding and analysis of intricate networks of mechanistic steps spanning

multiple substrates, intermediates, and side reactions that can hijack the desired multicomponent sequence. Here, we show that computers equipped with broad knowledge of mechanistic transforms, rules of physical-organic chemistry, and approximations of kinetic rates can perform such network analyses rapidly and in a high-throughput manner, and can guide systematic discovery, ranking, and yield estimation of mechanistically distinct types of MCRs, one-pot sequences and even organocatalytic reactions, several of which we validate by experiment. These results evidence that synthesis-planning algorithms are no longer limited to skillful manipulation of the existing knowledge-base of full reactions<sup>22–28</sup> but can assist in its creative expansion.

Every chemical reaction is a sequence of elementary steps or, at a less precise but very popular representation, of arrow-pushing

<sup>1</sup>Allchemy Inc., Highland, IN, USA. <sup>2</sup>Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland. <sup>3</sup>Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea. <sup>4</sup>Jagiellonian University, Krakow, Poland. <sup>5</sup>Department of Chemistry, Ulsan Institute of Science and Technology, UNIST, Ulsan 44919, Republic of Korea. <sup>6</sup>These authors contributed equally: Rafał Roszak, Louis Gadina, Agnieszka Wołos. ✉e-mail: [daniel.gryko@icho.edu.pl](mailto:daniel.gryko@icho.edu.pl); [nanogrzybowski@gmail.com](mailto:nanogrzybowski@gmail.com)



**Fig. 1 | Significance and current discovery rate of multicomponent reactions.** **a** A classic example illustrating the elegance and efficiency of Robinson's one-step, MCR synthesis of tropinone vs. prior, fifteen-step synthesis<sup>10</sup>. In the latter, only two key steps are shown. **b** t-SNE projection "map" illustrating diversity of 631 known MCR classes/types (smaller blue markers) and 66 one-pot classes (green) vs. the MCRs and one-pots (larger red and orange markers, respectively) discovered in this work and validated by experiment. The known MCR and one-pot classes were curated by our group over the years based on several extensive literature reviews – all this data (631 + 66) is available for download, along with the links to the first publication reporting a given reaction type, as either a .csv or Excel file from <https://doi.org/10.5281/zenodo.10817102>. In the map, each marker corresponds to t-SNE projection of reaction fingerprints, defined as a difference between fingerprint of

the product and fingerprint of the substrates<sup>33</sup> (i.e., difference across a full reaction, not its mechanistic steps). The interactive t-SNE map is deposited at <https://mcmap.allchemy.net>. **c** Blue line and left axis quantify the numbers of papers on MCRs published in a given year (based on "multicomponent reaction" query of the Web of Knowledge database, August 2024). Red line and right axis plot are based on the set of 631 MCR types from **b**. For each year, the number of newly discovered MCR types (i.e., published for the first time in this year) is plotted. The number of publications on MCRs peaked around 2019 and has slightly decreased since. On the other hand, the discovery rate of new MCRs seems to have followed cyclical variations. It should be noted, however, that since the nadir in 2015–2017, it is now increasing perceptibly, perhaps signaling renewed interest in multicomponent reactions.

steps<sup>29</sup>, which has been used in computational chemistry for decades<sup>30–35</sup> (though in most cases to analyze only certain types of chemistries and with limited accuracy, see Supplementary Section S6 in ref. 36 and Supplementary Section S3 here). As we have recently shown for complex carbocationic rearrangements<sup>36</sup>, this level of description is appealing because, compared to quantum methods, it reduces the number of degrees of freedom one needs to consider, while still retaining enough accuracy to rationalize the mechanisms of the vast majority of organic chemical transformations, including the previously unreported reactions<sup>37,38</sup>. In this work, we use a large and diverse collection of arrow-pushing operators to generate networks of mechanistic steps starting from sets of multiple substrates potentially exhibiting different modes of reactivity. We then aim to identify the mechanistic pathways and conditions that would select only some of these modes and would proceed, in one pot, cleanly into products significantly more complex than the starting materials. Uniquely and mindful of various cross-reactivities possible in multicomponent reaction mixtures, we consider possible by-products, products of side reactions, and further reactions of these species as well as their potential interference with the main mechanistic pathway. We scrutinize these processes for kinetics to ensure that side-processes do not hijack the desired sequence, lowering or even nullifying its yield, which we also aim to approximate. Within this general approach, the problem of designing MCRs or one-pot sequences becomes one of selecting the substrates, expanding the mechanistic networks forward and sideways from these substrates, and performing kinetic analysis to trace conflict-free mechanistic routes (Fig. 2).

## Results

### Choice of substrates

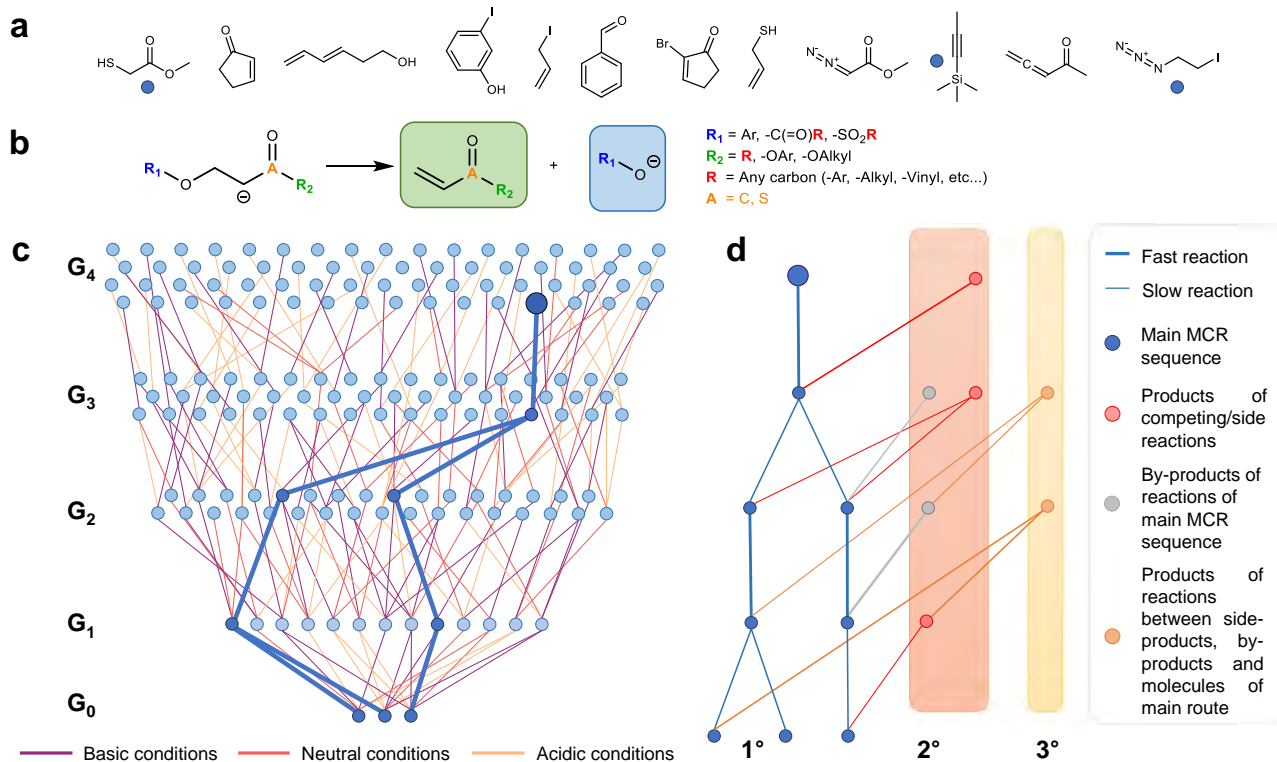
While the algorithm accepts any user-specified molecules as input, guessing the substrates resulting in productive MCRs may be challenging. Instead, we rely on a high-throughput, computational analyses of substrate combinations from a house-curated collection of ca. 2400 simple, diverse and commercially available small molecules featuring one or two groups reactive in various types of transformations (Fig. 2a and, for details, Methods and Supplementary Section S4).

### Mechanistic transforms

To propagate the mechanistic networks, a collection of ~8000 commonly accepted mechanistic transforms was encoded at the aforementioned arrow-pushing level in the SMARTS notation as described before<sup>39,40</sup>. This collection includes a broad range of chemistries although it is certainly not yet without omissions (see Methods). Transforms account for by-products (Fig. 2b) and are categorized according to typical reaction conditions, temperature range and water tolerance, as well as typical speeds (very slow, slow, fast, very fast, and uncertain if conflicting literature data have been reported, VS-S-F-VF-U). Since the focus of the algorithm is to generate scaffolds not yet described in the literature, the algorithm does not consider stereochemistry. For more details on rule coding, see Methods and Supplementary Section S5.

### Forward expansion of mechanistic networks

For a given set of substrates (henceforth, synthetic generation  $G_0$ ), the algorithm applies the mechanistic transforms to create the



**Fig. 2 | Key elements of the MECH algorithm to discover MCRs.** **a** Examples of simple starting materials from the collection of ca. 2400 (see main text and Methods). **b** Abbreviated example of one of ~8000 mechanistic transforms, here E1cB elimination. Different positions can accommodate various substituents, some of which are listed to the right of the reaction scheme. Note that the transform is coded to account for the by-product(s), here a phenoxide, carboxylate or mesylate. Classification of reaction conditions, rates, etc., are also parts of the transform's record but, for clarity, are not shown here. For tutorial of rule coding see Supplementary Section 5 and examples deposited at <https://zenodo.org/records/13381201>. **c** Application of mechanistic transforms to a given set of starting materials iteratively expands the synthetic generations,  $G_n$ , of a network of possible intermediates and immediate by-products. In the schematic miniature drawn, the three circle markers in the bottom row ( $G_0$ ) may be the three molecules from panel a, and the network is expanded to  $G_4$ . Different colors of connections between the

nodes are intended to denote different types of conditions – to emphasize that this “forward” network expansion probes all conditions’ combinations. Within the network thus constructed, the conditions may be matching (i.e., mutually compatible), corresponding to a MCR candidate at the Level 1 of analysis (sequence of steps highlighted in dark blue). **d** Such a sequence is expanded sideways, to perform analyses at Levels 2 and higher. Level 2 – branching-out of the main path to include by-products (gray) and products of competing/side reactions possible under the same class of reaction conditions (red; for condition types, see Methods); Level 3 – further branching to account for the reactions between side- and by-products. At Level 3 (and higher, not shown here for clarity but see Fig. 3b), undesired reactions of side-/by-products with each other and with the members of the main pathway are also considered and marked in orange. Faster reactions are represented schematically by thicker connections and it is essential that, at any junction, the side reactions are not faster than the main-path ones.

first-generation,  $G_1$ , of products and by-products, which are then iteratively reacted<sup>23,25,36</sup> to give generations  $G_2$ ,  $G_3$  (up to some user-specified generation  $n$ ), resulting in rapidly expanding<sup>26</sup> networks of mechanistic steps (Figs. 2c and 3a). At this stage, all classes of reaction conditions are allowed to survey the “synthesizable space” broadly but intermediates containing highly strained scaffolds not known as reaction intermediates (e.g., cyclobutenylene but not benzyne) are eliminated. Molecules can also be checked for the pKa of all C-H bonds<sup>41</sup> to ensure that reactions with electrophiles, such as C-H alkylations, proceed at the most acidic positions. Also, to prevent oligo/polymerization and limit network’s size, each substrate is allowed to contribute atoms to any molecule in the network at most twice (see User Manual).

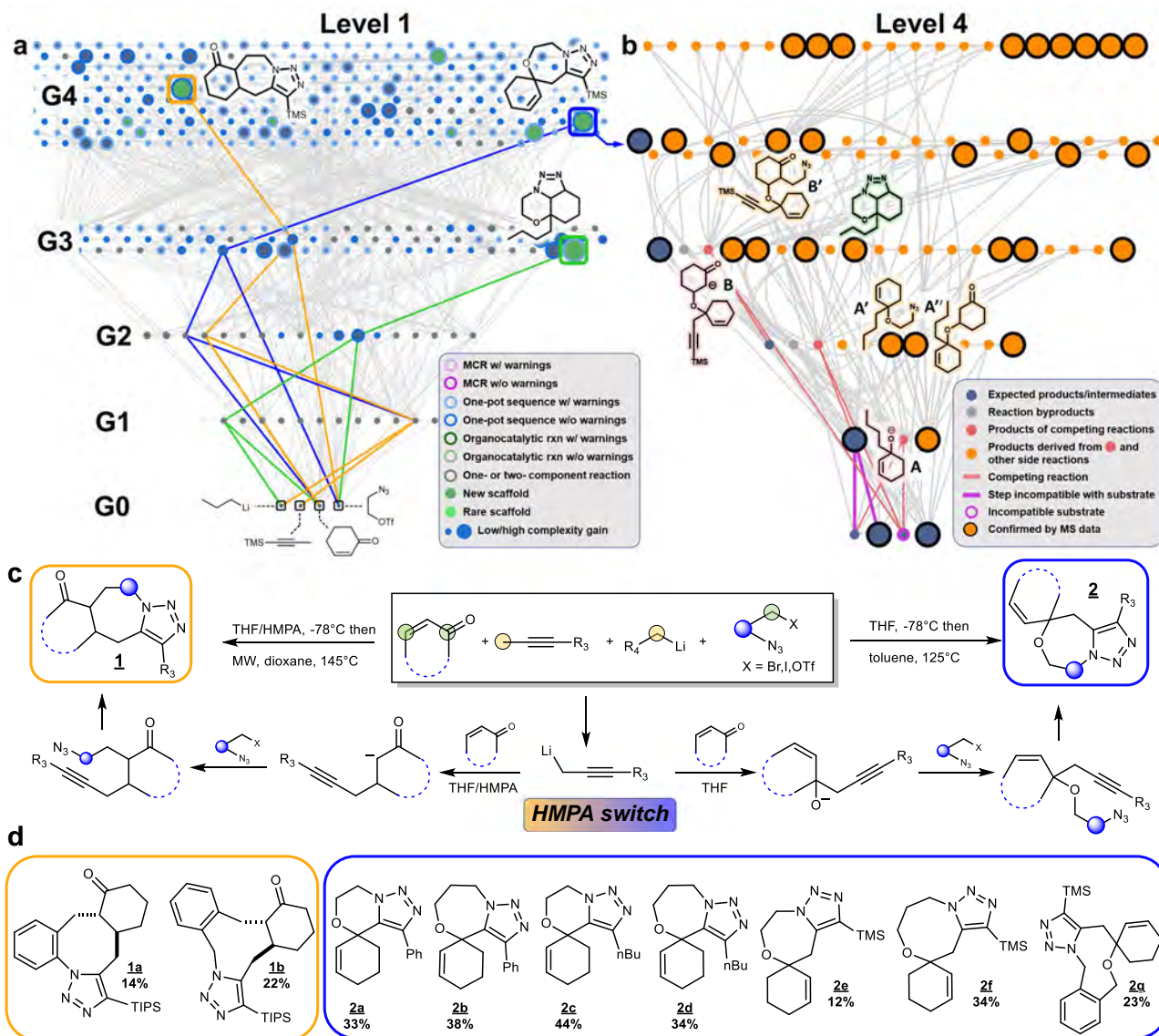
### Selection of mutually-compatible MCR/one-pot sequences

Pathways leading to every neutral molecule within the network thus created are traced by Dijkstra-type algorithm; if multiple routes are detected, they are retrieved and ranked according to length. For any of these mechanistic sequences to be suitable candidates for MCR or one-pot reactions, the conditions specified for individual mechanistic steps must be matching. This is the Level 1 of analysis (Figs. 2c and 3a) and the sequences:

- Cannot combine steps requiring oxidative and reductive conditions, and cannot use water-sensitive steps after water-requiring ones;
- Should use solvents of the same class, although protic solvents are allowed to be added to aprotic ones (but not vice versa);
- Cannot change multiple times between non-overlapping high and low temperature ranges (which would be experimentally impractical);
- Should allow only for monotonic changes in acidity (e.g., basic-acidic-basic changes are not allowed). Additionally, steps proceeding in strongly basic conditions (with, e.g. LDA) are not allowed if earlier steps required acidic conditions.

### Sideways network expansion around main MCR/one-pot routes

If Level 1 analysis identifies a candidate, condition-matching sequence, the aforementioned sideways analysis of potential side reactions is performed (Figs. 2d and 3b). At Level 2, the kinetics of side reactions are examined. Initially, this is done in a rudimentary manner, according to the aforementioned “very slow-slow-fast-very fast-uncertain” categorization of reaction steps (cf. examples in Methods). In particular, warnings are assigned if, for a given reaction of the main path, a side-step possible under the same or similar conditions is faster. Such



**Fig. 3 | Example of algorithmically-discovered one-pot sequences and the corresponding mechanistic network expanded to Level 4.** **a** Screenshot of Level 1 network propagated from cyclohexenone, trimethylsilylpropyne, *n*-butyllithium and azidotriflate substrates to  $n = 4$  generations,  $G_4$ . The network encompasses all mutually-compatible sequences possible under different types of conditions. Node sizes are proportional to complexity increase per mechanistic step,  $\Delta C/n$  (cf. Methods). Colors of the halos define MCR/one-pot sequences with or without warnings. Nodes whose interiors are colored green correspond to scaffolds not described in the literature. Within this network, two sequences (traced in blue and orange) up to  $G_4$  are predicted to be one-pot without warnings and leading to unknown scaffolds **1** and **2** offering marked increase in  $\Delta C/n$  (largest green nodes). A path to another complex scaffold in  $G_3$  is also marked (in green). This product is predicted to form from the 1,2-adduct of *n*BuLi/cyclohexanone/azide cyclizing onto the double bond, and was detected by ESI-MS in the reaction mixture (structure highlighted in green in the L4 network in panel **b**). **b** Screenshot of the network branched-out from the blue pathway in **a** and analyzed at Level 4 (for networks analyzed at Levels 2 and 3, see Supplementary Fig. S158; interactive network expandable to Level 4 is deposited at <https://mcrchampionship.allchemistry.net>). This Level 4 network encompasses various by- and side-products (gray and red nodes, respectively) and their further reactions (products marked in orange) between themselves and with the “parent” pathway. Larger orange nodes are likely

structural assignments of peaks observed in the ESI-MS of the crude-reaction mixture. Interestingly, although the peaks corresponding to some predicted byproducts (e.g., **A**, **B**; structures drawn here with pink highlights) were not manifest in the ESI-MS spectra, their formation is corroborated by further products (**A'**, **A''**, **B'**; structures drawn with orange highlights) that can only be derived from these undetected species. For more structural assignments, see Supplementary Fig. S158. Also, the key cross-reactivity mandating sequential addition of reagents rather than MCR (i.e., reaction of alkyllithium with enone during metalation of alkyne) is highlighted by brighter pink connections at the bottom of the network. **c** General scheme and intermediates of the blue and orange one-pot pathways (leading to scaffolds **1** and **2**, as in **a**) along with reaction conditions. In the substrates, the available nucleophilic and electrophilic sites are marked yellow and green, respectively, while the dark blue circle and the dotted arcs denote linkers between the azide and (pseudo)halides and a cyclic or acyclic fragment of the enone, respectively. The regioselectivity of addition (1,2- vs 1,4-) of propargyllithium reagent is controlled by the addition of HMPA as co-solvent. **d** Specific derivatives **1a**, **1b** and **2a–2g** synthesized according to the general protocol along with the corresponding isolated yields. Note that the yields are low, as indeed predicted by the algorithm (see main text). Compounds **1a** and **1b** were isolated as single diastereoisomers. THF tetrahydrofuran, HMPA hexamethylphosphoramide, MW microwave, OTf triflate, TMS trimethylsilyl, TIPS triisopropylsilyl.

cases are flagged but not permanently removed from the mechanistic network since it is sometimes possible to generate thermodynamic products via a slower reaction (e.g., slow 1,4-addition of cyanide to methyl-vinyl ketone vs. fast 1,2-addition). Additional warnings are assigned if any of the by-products shows cross-reactivity with the main pathway or the reaction mixture becomes too complex (e.g., if three or more metals from catalysts or reagents are present and there is a possibility for unforeseen complexation of active species or deactivation of catalysts by ligand exchange). The by- and side-products from Levels 1 and 2 are allowed to react further, to give species at higher Levels, for which similar cross-reactivity analyses are performed. Importantly, the algorithm also analyzes whether reactivity conflicts between forming intermediates and yet unreacted substrates exist. If all substrates contributing atoms to the final product can be present in the reaction vessel from the beginning, the sequence is categorized as a plausible MCR (with possible condition changes obeying (i)-(iv) above); if, however, some intermediates are found to be cross-reactive with the substrates, then the algorithm suggests a one-pot option with sequential addition of the problematic substrate. In the current work, we focus on MCRs and one-pot sequences that entail no unresolved conflicts or warnings within Level 4 networks (see realistic examples in Fig. 3b and Supplementary Figs. S158–S162).

### Prioritization and post-design evaluation

Because even for small substrate sets, the networks thus constructed may span large numbers of plausible MCR/one-pot products (Fig. 3a), additional analyses are performed to identify those that offer maximal complexification of the scaffold, those producing previously unknown scaffolds, those that are similar to approved drugs, and more (see Methods). The algorithm can also read in the positions of experimentally recorded mass-spectrometric signals and map them onto the Level 2-4 networks, which often facilitates analysis of experimental reaction mixtures (cf. Fig. 3b, Supplementary Fig. S158, and Supplementary Section S1).

### Estimation of yields

Finally, once a desired MCR/one-pot candidate is selected, the algorithm performs a more in-depth kinetic analysis aimed at the estimation of reaction's yield. Since experimental kinetic rate constants for the vast majority of mechanistic steps are not available, we developed a physical-organic model grounded in free-energy linear relationships and approximating the rate constants of mechanistic steps using Mayr's nucleophilicity indices (see refs. 42,43 and Methods).

### Experimental validations

From amongst the multitude of putative MCRs the algorithm has thus far identified, we focused on those that offer mechanistic uniqueness (i.e., substantial difference vs. known MCRs) and high substrate-to-product complexity increase, start from simple (commercially available or easy-to-make) substrates, and produce scaffolds of potential usefulness. Another factor was the conciseness of these protocols vs. traditional retrosynthetic planning that is based on full reactions rather than mechanistic steps and cannot capitalize on the use of reactive intermediates. Accordingly, for all one-pot/MCR products, we also ran the state-of-the-art retrosynthetic program (Chematica/Synthia<sup>22,24</sup>) which either planned multistep routes (on average 4 and up to 11 steps; all deposited at <https://zenodo.org/records/10817102>) or did not suggest any syntheses at all. All sequences are named "Mach" to highlight their machine-driven discovery (and to allude to its speed).

### One-pot, non-MCR sequences

We begin with an example that is, admittedly, simple but serves to illustrate various modalities of the algorithm. Starting from enone, alkyllithium, azidoaldehyde and alkyne, the mechanistic network

propagated to G<sub>4</sub> (Fig. 3a) contains conditions-matched sequences leading to 391 products with MW < 500. Two compounds in G<sub>4</sub> correspond to previously unreported, tricyclic scaffolds **1** and **2**, both characterized by large per-step complexity increase from the substrates, and with **2** featuring a spiro system akin to that in some drugs and bioactive agents<sup>44-47</sup>. The mechanistic sequences to these products diverge at the initial step. The Mach1 route (blue) proceeds via the 1,2 addition, generation of the alkoxide intermediate, O-alkylation, and click reaction closing two rings. The Mach2, orange route starts with 1,4 Michael-type addition creating a carbanion at the alpha carbon, followed by C-alkylation and click reaction. The algorithm predicts that these sequences (1) may be performed only as one-pot rather than MCR (with enone added only after the complete consumption of the alkyllithium substrate); (2) the initial steps in both routes can be carried out using propargyllithium, with HMPA acting as a switch (Fig. 3c) to promote the 1,4 addition<sup>48</sup>; and (3) that both sequences will result in poor yields, ca. 20–40%. All these predictions turned out correct, with the isolated yields of derivatives **1a**, **1b** and **2a–2g** shown in Fig. 3d ranging from 12 to 44%. Of note, the computer-predicted competing reactivity modes were also congruent with ESI-MS analyses – in Fig. 3b and Supplementary Fig. S158, larger orange nodes denote side-/by-products with masses matching the spectra.

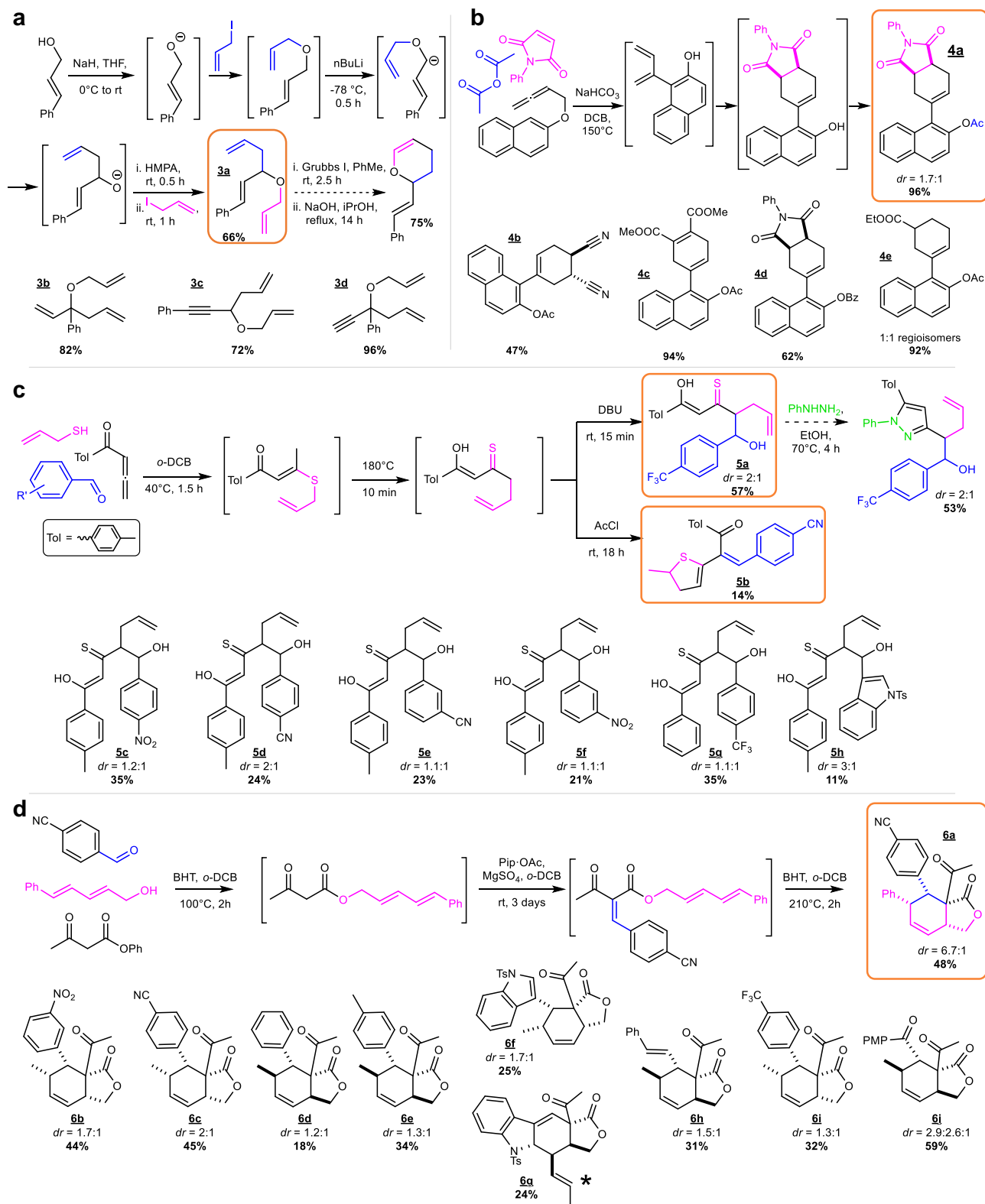
Another prediction for a one-pot, Mach3 sequence relying on a 2,3-Wittig rearrangement and leading to branched diallylic ethers **3a–3d**, is illustrated in Fig. 4a and Supplementary Fig. S159. This sequence was committed to experiment because, after metathesis (not compatible with one-pot conditions and carried out separately, dashed reaction arrow in Fig. 4a), it affords access to cyclic enol ether scaffolds that are used in various medicinal syntheses<sup>49-52</sup>. This sequence was predicted to proceed in good, ~68% yield vs. 66–96% yields we obtained.

### MCR sequences

Turning to MCRs rather than one-pot sequences, Fig. 4b and Supplementary Fig. S160 illustrate a Mach4 sequence, in which an allene, a maleimide derivative, and a carboxylic acid anhydride engage in a sequence of Claisen rearrangement, aromatization, Diels-Alder cycloaddition, deprotonation and acylation to yield a 1-(1-cyclohexenyl)naphthalene, atropisomeric scaffold **4a** familiar from various types of drugs<sup>53,54</sup>. Scaffolds of this type are typically prepared via various multistep protocols<sup>55-59</sup>. The MCR approach shortens these procedures while commencing from substrates of similar complexity and does not require transition metal catalysts or pre-functionalized aryl systems. The experimental yields for **4a** and its analogs **4b–4e** were generally quite satisfactory and in most cases >90% (for the originally predicted sequence, the algorithm predicted 99% vs. 96% in experiment).

Another pair of MCRs using allene as one of the substrates is illustrated in Fig. 4c and begins with a nucleophilic addition of an allyl thiol to the allene and isomerization followed by thio-Claisen rearrangement. Network analysis detailed in Supplementary Figs. S161 and S162 indicates that the sequence can then diverge. In Mach5 MCR, addition of excess base results in straightforward condensation with an aromatic aldehyde occurring at the less acidic methylene group of the thio ketone and leading to **5a** in 57% yield (vs. predicted 43%). This product or its analogs **5c–5h** can further react (outside of the MCR, dashed reaction arrow) with phenyl hydrazine<sup>60</sup> to give substituted pyrazoles which are popular motifs of many drugs. By contrast, in Mach6, addition of acetyl chloride triggers a relatively rare<sup>61</sup> sequence of acetylation of alcohol, acidic elimination of acetic acid catalyzed by the in-situ generated HCl to give the Knoevenagel-type adduct, thio keto-enol tautomerization followed by spontaneous cyclization. The 2,3-dihydrothiophene products **5b** are obtained in significantly lower yields (-10% and up to 14% for the cyano derivative vs. 12% predicted yield, though these experimental values are affected





by partial decomposition of the product during purification), and their applications are less conspicuous<sup>62</sup>.

The sequence underlying Mach7 MCR shown in Fig. 4d – leading to a scaffold akin to oblongolide natural products considered as potential algicide<sup>63</sup> and antiviral<sup>64</sup> agents – is perhaps familiar to a retrosynthetically-trained eye. Indeed, the succession of transesterification of sorbic alcohol, Knoevenagel condensation and Diels-Alder reaction has also been found by Chematica/Synthia.

However, the MECH algorithm correctly predicted that it could be folded-up into a one-step MCR leading to **6a–6j**. The yields of racemic mixtures were up to 59% (compared to 55% predicted by the algorithm and 13–38% for multistep syntheses of similar scaffolds reported in refs. 65,66) and with the procedure readily scalable to gram scale (Supplementary Section S6.7). Also, one less obvious outcome predicted by the algorithm is that for the indole-3-carbaldehyde substrate, the Knoevenagel adduct can engage in a reverse-demand Diels-Alder

**Fig. 4 | Computer-discovered one-pot sequences and MCRs.** For details of mechanistic networks, see Supplementary Figs. S159–S162. **a** Scheme of a one-pot sequence for the synthesis of branched allyl ethers. The sequence is detected as one-pot rather than the MCR because excessive allyl iodide would react with *n*-butyllithium, hampering deprotonation and subsequent Wittig rearrangement (cf. Supplementary Fig. S159 marking this conflict). Non-isolated intermediates are shown in brackets and the isolated product **3a** is framed in orange. This product has been separately cyclized via ring-closing metathesis to afford cyclic enol ether. Additional derivatives **3b–3d** were prepared from allyl iodide and other commercially available  $\beta,\gamma$ -unsaturated alcohols. **b** Scheme of a MCR producing unsaturated  $\beta$ -naphthol esters. Key non-isolated intermediates are shown in brackets and the isolated product **4a** is framed in orange. Additional derivatives **4b–4e** were prepared using different commercially available dienophiles and acylating agents. **c** Scheme of a MCR producing unsaturated hydroxylated monothio- $\beta$ -diketones (existing in the thioenol tautomeric form) under basic conditions (top) or 2,3-dihydrothiophenes under acidic conditions (bottom) applied during the last step. Non-isolated intermediates are shown in brackets and the isolated products (**5a**

originally predicted for the top MCR and highest yielding **5b** for the bottom one are framed in orange. The monothio- $\beta$ -diketone product has been separately reacted (dashed arrow) with phenylhydrazine (green) to afford a substituted pyrazole. Additional products **5c–5h** were prepared by the top MCR. **d** Scheme of the MCR producing unsaturated bicyclic lactones. Key non-isolated intermediates are shown in brackets and the isolated product **6a** is framed in orange. Additional derivatives **6b–6j** were prepared using different commercially available aldehydes and dienes. BHT butylhydroxytoluene, DCB dichlorobenzene, THF tetrahydrofuran, DBU 1,8-diazabicyclo(5.4.0)undec-7-ene, HMPA hexamethylphosphoramide, Pip-OAc piperidinium acetate, dr diastereomeric ratio. Note: **4a** was observed as a 1.7:1 mixture of diastereoisomers with two distinct  $^1\text{H}$  NMR signals (separated by 0.5 ppm) for **Me**-OAc protons. These signals can be attributed to known through-space shielding by **Ph**-N in one of the diastereoisomers. However, no distinct signals allowing for determination of dr's were observed for structurally similar (OBz vs. OAc) **4d**. The product of reverse-demand Diels-Alder cyclization **6g** is marked with a star and was isolated as a single diastereoisomer. Percentage values in all panels are isolated yields.

cycloaddition to give a relatively complex, tetracyclic scaffold **6g** isolated in 24% yield.

### Substrate-reusing and organocatalytic sequences

The next two examples are interesting for the unique ways in which they use or reuse some of the substrates. In the Mach8 sequence shown in Fig. 5a, b, the phenol substrate is first used to form an activated ester that then reacts with 2-allylcyclohexanone to give a spiro  $\beta$ -lactone which, upon addition of  $\text{MgBr}_2$ , undergoes a ring-expanding rearrangement into a substituted hexahydro-2(3H)-benzofuranone **7a** in 31% yield (vs. predicted 48%). Such motifs are found in various natural products and bioactive compounds<sup>67</sup> and the particular scaffold, upon metathesis and reduction, could create a ring system present in lancifonins. However, when iodo-substituted phenols and cyclohexanone (instead of 2-allylcyclohexanone) are used as substrates and the network is propagated to higher generations, iodo-phenol is regenerated as a by-product of the spirocyclization step and then – upon product's decarboxylation – is reused in situ as a substrate in Heck reaction, to complete Mach9 MCR yielding **7b** in up to 35% yield (vs. predicted 35%).

In turn, Fig. 5c–e illustrate a Mach10 reaction that was predicted and then confirmed as organocatalytic. With the initial set of substrates ( $\alpha$ -bromo- $\alpha,\beta$ -unsaturated ester, methyl thioglycolate and sodium azide), the algorithm suggested an MCR that could lead to a dihydrothiophenecarboxylate scaffold **8a** similar to some GABA receptor inactivators<sup>68</sup>. However, the program also indicated that that the C-H pKa of the  $\alpha$ -azidoester be higher than that of the  $\alpha$ -thioester – that is, the deprotonation (either by azide anion<sup>69</sup> or sodium methoxide) at the former locus should be preferred and could lead to rapid elimination (green arrow in Fig. 5c, blue arc connection in the L2 network in Fig. 5d) rather than cyclization. This elimination sets a feedback loop regenerating the thiol (colored pink in Fig. 5d), which effectively acts as an organocatalyst sustaining azide substitution at vinylic  $\alpha$ -position. This was, indeed, verified in experiment with the original reaction to **8b** proceeding under mild conditions in 67% yield (vs. algorithm-predicted 47%), and with the further scope of **8c–8f** illustrated in Fig. 5e. For alkyl ketones, 10 mol% of the thiol is optimal, while for  $\beta$ -aryl ketones, 35 mol% thiol load is necessary due to the trapping of the thiol catalyst in the  $\text{S}_{\text{N}}2$  reaction with the alkyl bromide (obtained after 1,4-addition of thiol to Michael acceptor).

### Discussion

The above experimental examples cover only a tiny fraction of substrate combinations that can give rise to MCRs or one-pot sequences. To broaden and speed up the discovery process, we have automated the choices of substrate triplets and quartets (from the aforementioned set of ca. 2400 reactive molecules) as well as subsequent

network expansion and analysis. With tens of thousands of substrate combinations now probed and with further searches ongoing, the list of the currently 50 top-ranked (by complexity increase per step metric, see Methods) MCR candidates is maintained at <https://mcrchampionship.allchemy.net>. Users of Allchemy's MECH can perform searches with their own substrates of choice, and can opt to “compete” and post their results therein (if the scores place them within top-50), in the world' first “championship” for computerized reaction design.

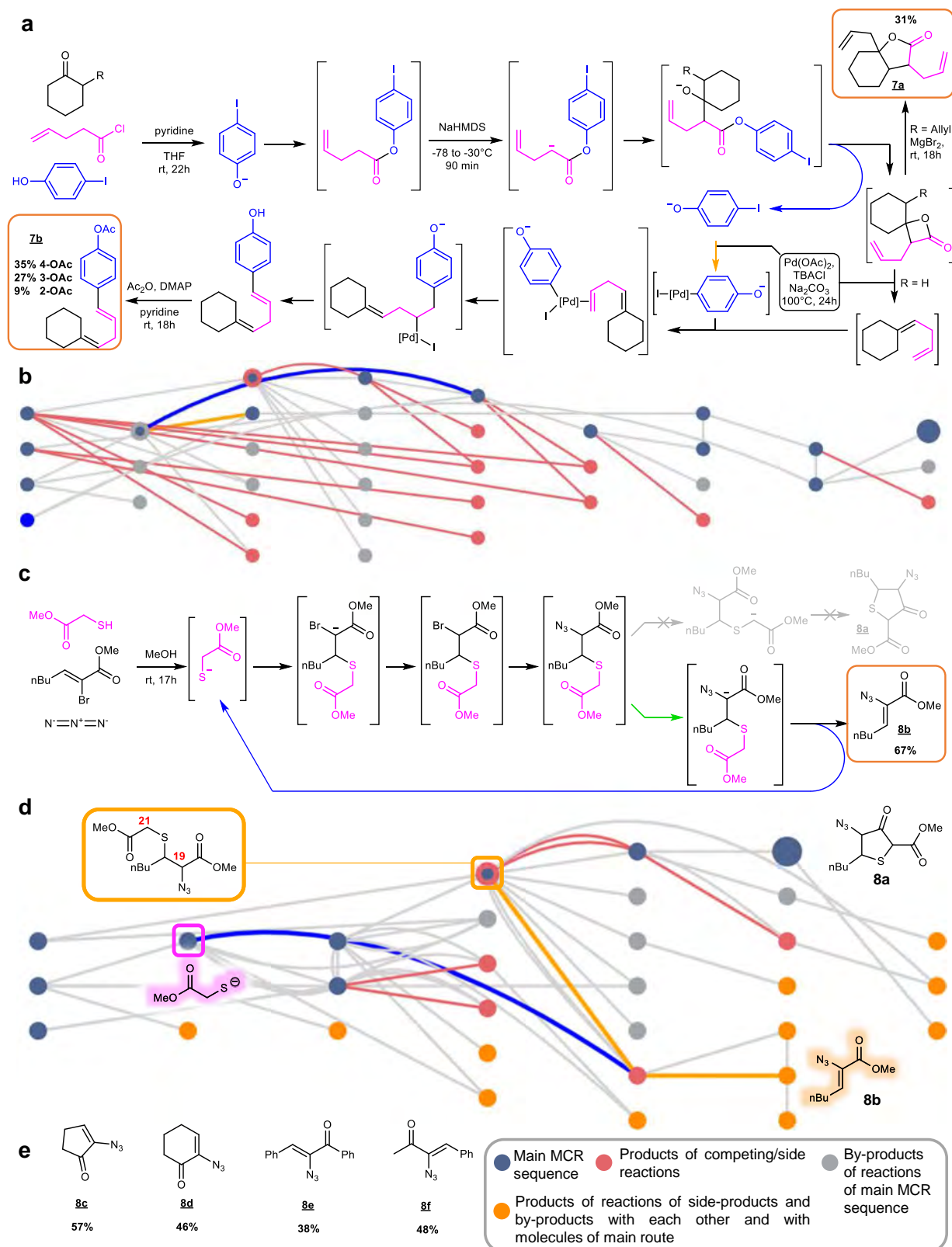
It is our hope that, in the fullness of time, this resource will enable discovery of MCRs in quantities that may have significant impact on the practice of synthetic chemistry. This said, algorithms like ours do not replace all of chemists' insights and the need for conditions' optimization (e.g., in terms of screening for optimal temperatures, solvents, etc.). There is also plenty of room for further improvements (see Supplementary Section S2), for an example of incorrect MCR prediction) and extensions of the algorithm, e.g., to incorporate radical-based mechanisms and additional catalytic transformations, or to adapt the workflow to the retrosynthetic direction (to suggest imaginative disconnections of specific scaffolds).

## Methods

### Mechanistic rules

As outlined in the main text, the mechanistic transforms are encoded in the SMARTS notation and account for by-products (a tutorial on coding the rules is provided in Supplementary Section 5). The templates are generalized – that is, they do not encompass just a single reaction precedent (as in the recently published repository of mechanistic steps for popular radicalic reactions<sup>70</sup>) but each specifies the scope of admissible substituents at various positions of the SMARTS template as well as a list of incompatible groups. These explicitly defined incompatibilities help limit the sizes of the networks and remove from analysis at least the obviously problematic steps, in which two or more motifs would react on commensurate time scales, inevitably leading to undesired complex reaction mixtures and ruining a “clean” MCR.

Furthermore, rules are accompanied by information about reaction conditions that is essential to later wire-up individual mechanistic steps into mutually compatible sequences. In this context, each transform is categorized according to general conditions (basic, neutral, acidic), solvent class (protic/aprotic and polar/non-polar), temperature range (very low =  $\leq -20^\circ\text{C}$ , low =  $-20$  to  $20^\circ\text{C}$ , r.t., high =  $40$  to  $150^\circ\text{C}$ , and very high =  $>150^\circ\text{C}$ ); and water tolerance (yes, no, water is required). One transform can have more than one categorization (e.g., Diels-Alder cycloaddition can be carried out either under neutral conditions at high temperature or at very low, low or room temperatures using a Lewis acid catalyst) – in such cases,



**Fig. 5 | Computer-discovered substrate-reusing MCRs and an organocatalytic reaction.** **a** Scheme of a MCR for the synthesis of arylated skipped dienes. Non-isolated intermediates are shown in brackets and the isolated products are framed in orange. The obtained dienes were separately acetylated for the purpose of purification. The bicyclic lactone **7a** (upper right) was obtained from substituted cyclohexanone (R = allyl) and phenol substrates when MgBr<sub>2</sub>·Et<sub>2</sub>O was used instead of the Pd-catalyst. **b** The Level 2 graph view of the path leading to the arylated diene from **a**. Reuse of iodophenol byproduct in Heck-coupling (with oxidative addition

step marked orange) is marked with the blue arc. **c** Scheme of organocatalytic thiol-catalyzed sp<sup>2</sup>-azidation. Non-isolated intermediates are shown in brackets and the isolated product **8b** is framed in orange. **d** The Level 3 graph view of the path from **c**. Reuse of thiol, acting as an organocatalyst, is marked with the blue arc. **e** Additional vinyl azides **8c–f** prepared by the MCR from **c** using different  $\alpha$ -bromo ketones. Abbreviations: DMAP, 4-dimethylaminopyridine; TBACl, tetrabutylammonium chloride.

multiple conditions are provided and, when considering sequences of compatible steps, are treated as logical alternatives. Each transform also contains specific suggestions for reagents commonly used in reactions involving this mechanistic step (e.g., diethylaluminum chloride in Claisen rearrangement, *n*-butyllithium in [2,3]-Wittig rearrangement, etc.).

Regarding the initial and rough categorization of kinetics, each transform is assigned a typical speed category (very slow, slow, fast, very fast, uncertain). A “very slow” step (conversion time above ca. 24 hrs) is, for example, addition of amines to trisubstituted Michael acceptors. Steps categorized as “slow” (few to ca. 24 h) are, e.g., reaction of a deprotonated nitro compound with a ketone, addition of an alcohol to a protonated nitrile, or 1,3-dipolar cycloaddition of imine and nitrile oxide. Examples of “fast” steps (minutes to few hrs) include deprotonation of alcohols, alcoholysis of anhydrides, or addition of organocuprates to activated alkenes. “Very fast” steps (seconds to minutes) are, for example, decomposition of oxaphosphetanes to alkenes and phosphine oxide, elimination of a chloride anion from an adduct of amine and acyl chloride, tautomerizations leading to aromatic compounds (e.g., 2,4-cyclohexadienone to phenol). “Uncertain” steps are those for which literature provides conflicting reaction data (i.e., wide range of reaction times and/or rates strongly influenced by substrate structures or small changes in reaction conditions) or those for which literature is insufficient to determine the reaction rate of an individual mechanistic step. One example from this category is addition of an imine to phenolic compounds, for which reaction rate strongly depends on the activity/nucleophilicity of phenolic component but even more on reaction conditions, resulting in time spans from 5 minutes to 9 hours for reactions involving the same substrates (see ref. 71 – 9 h<sup>72</sup>; – 7.5 h<sup>73</sup>; – 3 h<sup>74</sup>; – 5 min). Another example is S<sub>N</sub>2 reaction of a secondary bromide with cyanide anion, for which the reaction rate is strongly influenced by the character and size of substituents on the halide component and the type of solvent used, with polar aprotic solvents facilitating the reaction and polar protic solvents impeding it. For instance, reaction of 2-bromo-2-(2-methylphenyl)-1-(morpholin-4-yl)ethanone with sodium cyanide in methanol takes 24 h<sup>75</sup>, while reaction of a similar molecule, methyl 2-(1-bromo-2-methoxy-2-oxoethyl) benzoate, with potassium cyanide in DMF takes only 1 h<sup>76</sup>.

The rules covered in the current version of the MECH module span a broad range of acid-base catalyzed steps (including Lewis acids), substitutions, eliminations, additions, rearrangements, pericyclic reactions as well as basic transformations catalyzed by transition metals (e.g., mechanistic steps of Suzuki, Buchwald-Hartwig, Heck, and Pauson-Khand reactions). Basic carbocationic chemistry is included but not exhaustively (a separate HopCat module dedicated to such rearrangements is available in our recent publication<sup>36</sup>). Also, radical mechanistic steps are not (yet) included since their proper application requires generalization (cf. short discussion in Supplementary Section S3) and likely additional heuristics based on thermodynamic and molecular-mechanical considerations (akin to those we described in the HopCat paper<sup>36</sup>). Some rare types of steps involving  $\pi$ -complexes had to be simplified in notation since they are not properly handled by RDKit (they are encoded as 3-membered rings rather than interaction between metal and multiple bonds, e.g., during Heck reaction).

### Additional details of network expansion

During expansion of mechanistic networks, the program generally uses the individual steps, e.g., imine formation is divided into 1) ketone protonation, 2) imine addition to the protonated ketone, 3) proton transfer from nitrogen to oxygen, 4) formation of an iminium cation via elimination of water, 5) deprotonation of the iminium cation (Supplementary Fig. S163a). However, because the networks expand very rapidly with the number of steps (“synthetic generations”), such step-by-step expansions may be inefficient in exploring longer mechanistic sequences – for instance, the five-step imine formation is

only part of, say, the Ugi multicomponent reaction. To reduce computational cost, we have also encoded some shortcut steps that, for popular transformation types, concatenate individual mechanistic steps (those occurring in a rapid sequence and/or those leading to unstable intermediates; see example in Supplementary Fig. S163b). When executed as one “super-step”, the shortcuts keep all the information about by-products of all individual steps. The network expansions then use both the step-by-step and shortcut strategies. Of note, if a given substrate can engage in a very-fast, VF, mechanistic step (e.g., tautomerization, elimination leading to an aromatic product, etc.), only this rapid step is performed under given reaction conditions. Other competing mechanistic steps can be applied to this substrate only if they proceed under different class of conditions.

### Further details of route prioritization and post-design evaluation

The MECH module offers multiple options to filter, analyze, and prioritize the one-pot/MCR pathways within the mechanistic networks. As described in detail in Supplementary Section S1, the user can filter off those products that are formed via mechanistic steps having non-overlapping “cores” (reactions occurring on disjoint parts of the molecule will likely yield “linear” structures and will not complexify the starting scaffold), or those that do not involve any rearrangements or pericyclic reactions.

To easier identify and prioritize sequences that offer the highest degree of complexification, nodes in the network can be sized in proportion to the increase of structural complexity per step,  $\Delta C/n$ , where  $\Delta C$  is calculated along an atom-mapped path as ( $\alpha$ -#Rearrangements +  $\alpha$ -#RingsFormed + #BondsCreated + #BondsDisconnected), where  $\alpha = 5$  is used here to strongly favor formation of cyclic scaffolds and sequences containing rearrangements. Furthermore, the nodes can be colored as molecules known/unknown in the literature or, more generally, according to whether the scaffold is without precedent in prior literature. The algorithm to determine scaffold uniqueness first defines a scaffold “base” as a set of connected rings, whereby a ring is considered connected if it fulfills either of the two criteria: a) it shares at least one atom with any of the other rings in the base, b) is connected with a double bond to any of the other rings. The final scaffold is obtained from this base by inclusion of atoms connected to the base with double bond (i.e., oxygen from carbonyl group or exomethylene double bond). Note that this definition inherits both elements and bond orders from the parent molecule such that, for instance, cyclohexane, cyclohexene, cyclohexanone and cyclohexanethione are all considered as different scaffolds. Finally, a scaffold is considered without prior precedent if it is not present in the list of 95,191 scaffolds extracted from the Zinc collection<sup>77</sup>. The nodes within the networks can also be colored by similarity to approved drugs, reaction type, hazardous compounds, and more (see User Manual in Supplementary Section S1). Last but not least, the user can input a list of mass-spectrometric signals recorded in experiment and the likely M + 1 and M + 23 nodes will be marked on Level 1–4 trees (Fig. 2b and Supplementary Fig. S127).

### Estimation of yields

To estimate the yields of MCR/one-pot candidates, we developed a physical-organic model grounded in free-energy linear relationships. In this model, to be detailed in a separate publication<sup>78</sup>, the rate constants of mechanistic steps are approximated by using Mayr’s nucleophilicity,  $N$ , and electrophilicity,  $E$ , indices<sup>42,43</sup> as  $\log k_{20\text{deg}} \propto (N + E)$ , which are further fine-tuned by corrections capturing relative reactivities, stoichiometries and amounts of various species in the mechanistic networks,  $\ln k_i = \ln k_i^{\text{Mayr}} + \sum \text{corrections}(r_i)$ . The weights of the individual corrections were trained on the mechanistic networks of 20 diverse MCRs reported before (chosen to represent both low- and high-yielding ones), and the model was then used to predict the yields of the

mechanistically distinct MCRs described in the current publication. For the training set of the known MCRs, the Pearson correlation coefficient ( $\rho^2$ ) between the experimental and modeled yields was 0.80 with mean absolute error of 10.5. For the test set of reactions used in this study,  $\rho^2 = 0.86$  and MAE = 7.3. These metrics compare quite favorably with generally unsatisfactory correlations observed for various machine learning models trained on full, substrate-to-product reactions without any mechanistic knowledge<sup>79–82</sup>.

### Pre-curated collection of substrates available through All-chemistry's user interface

Although arbitrary substrates can be input in Allchemy's MECH module, we have also curated a list of ~2400 simple and commercially available substrates that, in our experience, improve the chances of finding MCR reactions. To begin with, the Zinc collection<sup>77</sup> was pruned to retain only molecules with, at most, 15 heavy atoms. After removing stereochemistry, ~410,000 unique entries were left. Molecules containing either poorly reactive fragments (94 patterns, e.g., heterocycles, polycyclic systems, ethers) or several unfunctionalized carbon atoms were removed, as they only introduced unnecessary structural complexity without desired reactivity. The remaining molecules were queried for the presence of one or two reactive groups defined by experienced synthetic chemists (164 patterns of FGs listed in Supplementary Tables S2, S3) – there were 36,294 such molecules of which 16,631 had one reactive FG and 19,663 had two reactive FGs. In the latter, we only kept molecules in which the FGs were separated by, at most, three atoms – in this way, when these molecules reacted, they were more likely to form smaller rings rather than macrocycles. For some FG combinations, there were many hits (e.g., the algorithm identified 97 commercially available isocyanates and 94 compounds possessing both aryl bromide and secondary amine FGs). In such cases, the compound with the lowest molecular mass was retained.

### Data availability

The list of reactions and literature sources of known MCRs and one-pots is deposited as .csv and Excel files at Zenodo under accession code <https://zenodo.org/records/10817102>. All 3108 unique reactions from all networks are deposited (along with condition classification, rate categorization and optimized rate parameters) at Zenodo under accession code <https://zenodo.org/records/13381381>. Multistep synthesis plans produced by Chematica/Synthia for targets made here via MCRs and one-pots are deposited at Zenodo under accession code <https://zenodo.org/records/10817102> (note: no syntheses were found for Mach6 and for one of the two variants of Mach2). The X-ray crystallographic coordinates for structures reported in this study have been deposited at the Cambridge Crystallographic Data Centre (CCDC), under deposition numbers 2402793. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via [www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif). User manuals are available in Supplementary Section SI. Interactive networks for all examples described in the text and results of MCR Championships are available for analysis at <https://mcrchampionship.allchemy.net> under restricted access. The interactive t-SNE map of known MCRs and one-pots is available under restricted access at <https://mcrmap.allchemy.net>. All searches we described or any other searches one may wish to execute, can be performed under restricted access at <https://mech.allchemy.net>. Access to all restricted services can be obtained by academic users by sending a request to [admin@allchemy.net](mailto:admin@allchemy.net) from an academic address. The restrictions are dictated by server capacity so the access can be provided to twenty concurrent academic users on a rolling basis and two-week slots.

### Code availability

Codes for network expansion and MCR analysis are deposited at <https://zenodo.org/records/13381201>. Codes for the estimation of

kinetic rates and calculation of yields are deposited at <https://zenodo.org/records/13381381>. The same repository (<https://zenodo.org/records/13381381>) houses codes for the optimization of kinetic parameters as well as 30 digitized mechanistic networks on which the rate-prediction model was trained and tested (with details of the model development described in ref. 78). Interactive Allchemy MECH web-app is freely available at <https://mech.allchemy.net> (given server capacity, to twenty concurrent academic users on a rolling basis and two-week slots).

### References

1. Ugi, I. et al. New applications of computers in chemistry. *Angew. Chem. Int. Ed.* **18**, 111–123 (1979).
2. Bauer, J. & Ugi, I. Chemical-reactions and structures without precedent generated by computer-program. *J. Chem. Res.-S* **11**, 298–298 (1982).
3. Bauer, J., Herges, R., Fontain, E. & Ugi, I. IGOR and computer assisted innovation in chemistry. *Chimia* **39**, 43–53 (1985).
4. Ugi, I. K. et al. Computer assistance in the design of syntheses and a new generation of computer programs for the solution of chemical problems by molecular logic. *Pure. Appl. Chem.* **60**, 1573–1586 (1988).
5. Dömling, A. & Ugi, I. Multicomponent reactions with isocyanides. *Angew. Chem. Int. Ed.* **39**, 3168–3210 (2000).
6. Dömling, A., Wang, W. & Wang, K. Chemistry and biology of multicomponent reactions. *Chem. Rev.* **112**, 3083–3135 (2012).
7. Ganem, B. Strategies for innovation in multicomponent reaction design. *Acc. Chem. Res.* **42**, 463–472 (2009).
8. D'Souza, D. M. & Müller, T. J. J. Multi-component syntheses of heterocycles by transition-metal catalysis. *Chem. Soc. Rev.* **36**, 1095–1108 (2007).
9. Phelps, J. M. et al. Multicomponent synthesis of  $\alpha$ -branched amines via a zinc-mediated carbonyl alkylative amination reaction. *J. Am. Chem. Soc.* **146**, 9045–9062 (2024).
10. Medley, J. W. & Movassaghi, M. Robinson's landmark synthesis of tropinone. *Chem. Comm.* **49**, 10775–10777 (2013).
11. Brandner, L. & Müller, T. J. J. Multicomponent synthesis of chromophores – The one-pot approach to functional  $\pi$ -systems. *Front. Chem.* **11**, 1124209 (2023).
12. Zhao, W. Y. & Chen, F. E. One-pot synthesis and its practical application in pharmaceutical industry. *Curr. Org. Synth.* **9**, 873–897 (2012).
13. Broadwater, S. J., Roth, S. L., Price, K. E., Kobašljia, M. & McQuade, D. T. One-pot multi-step synthesis: a challenge spawning innovation. *Org. Biomol. Chem.* **3**, 2899–2906 (2005).
14. Wheeldon, I. et al. Substrate channeling as an approach to cascade reactions. *Nat. Chem.* **8**, 299–309 (2016).
15. Hayashi, Y. Time and pot economy in total synthesis. *Acc. Chem. Res.* **54**, 1385–1398 (2021).
16. Paul, B., Maji, M., Chakrabartia, K. & Kundu, S. Tandem transformations and multicomponent reactions utilizing alcohols following dehydrogenation strategy. *Org. Biomol. Chem.* **18**, 2193–2214 (2020).
17. Cioc, R. C., Ruijter, E. & Orru, R. V. A. Multicomponent reactions: advanced tools for sustainable organic synthesis. *Green Chem* **16**, 2958–2975 (2014).
18. Ruijter, E., Scheffelaar, R. & Orru, R. V. A. Multicomponent reaction design in the quest for molecular complexity and diversity. *Angew. Chem. Int. Ed.* **50**, 6234–6246 (2011).
19. Hayashi, H. et al. In silico reaction screening with difluorocarbene for *N*-difluoroalkylative dearomatization of pyridines. *Nat. Synth.* **1**, 804–814 (2022).
20. Novikov, M. S., Khlebnikov, A. F., Krebs, A. & Kostikov, R. R. Unprecedented 1,3-dipolar cycloaddition of azomethine ylides derived from difluorocarbene and imines to carbonyl compounds –

- Synthesis of oxazolidine derivatives. *Eur. J. Org. Chem.* **1**, 133–137 (1998).
21. Dong, S., Fu, X. & Xu, X. [3+2]-Cycloaddition of catalytically generated pyridinium ylide: A general access to indolizine derivatives. *Asian J. Org. Chem.* **9**, 1133–1143 (2020).
  22. Mikulak-Klucznik, B. et al. Computational planning of the synthesis of complex natural products. *Nature* **588**, 83–88 (2020).
  23. Wołos, A. et al. Computer-designed repurposing of chemical wastes into drugs. *Nature* **604**, 668–676 (2022).
  24. Lin, Y., Zhang, R., Wang, D. & Cernak, T. Computer-aided key step generation in alkaloid total synthesis. *Science* **379**, 453–457 (2023).
  25. Wołos, A. et al. Synthetic connectivity, emergence, and autocatalysis in the network of prebiotic chemistry. *Science* **369**, eaaw1955 (2020).
  26. Gajewska, E. P. et al. Algorithmic discovery of tactical combinations for advanced organic syntheses. *Chem* **6**, 280–293 (2020).
  27. Molga, K. et al. A computer algorithm to discover iterative sequences of organic reactions. *Nat. Synth.* **1**, 49–58 (2022).
  28. Gothard, C. M. et al. Rewiring chemistry: algorithmic discovery and experimental validation of one-pot reactions in the network of organic chemistry. *Angew. Chem. Int. Ed.* **51**, 7922–7927 (2012).
  29. Grossman, R. B. *The Art of Writing Reasonable Organic Reaction Mechanisms* (Springer International Publishing, 2019).
  30. Gund, T. M., Schleyer, P. R., Gund, P. H. & Wipke, W. T. Computer assisted graph theoretical analysis of complex mechanistic problems in polycyclic hydrocarbons. The mechanism of diamantane formation from various pentacyclotetradecanes. *J. Am. Chem. Soc.* **97**, 743–751 (1975).
  31. Marsili, M. *Computer chemistry* (CRC Press, 1990).
  32. Chen, J. H. & Baldi, P. No electron left behind: a rule-based expert system to predict chemical reactions and reaction mechanisms. *J. Chem. Inf. Model.* **49**, 2034–2043 (2009).
  33. Kayala, M. A. & Baldi, P. ReactionPredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning. *J. Chem. Inf. Model.* **51**, 2526–2540 (2012).
  34. Jorgensen, W. L. et al. CAMEO: a program for the logical prediction of the products of organic reactions. *Pure Appl. Chem* **62**, 1921–1932 (1990).
  35. Satoh, H. & Funatsu, K. SOPHIA, a knowledge base-guided reaction prediction system-utilization of a knowledge base derived from a reaction database. *J. Chem. Inf. Comput. Sci.* **35**, 34–44 (1995).
  36. Klucznik, T. et al. Computational prediction of complex cationic rearrangement outcomes. *Nature* **625**, 508–515 (2024).
  37. Roque, J. B., Kuroda, Y., Göttemann, L. T. & Sarpong, R. Deconstructive diversification of cyclic amines. *Nature* **564**, 244–248 (2018).
  38. Kennedy, S. H., Dherange, B. D., Berger, K. J. & Levin, M. D. Skeletal editing through direct nitrogen deletion of secondary amines. *Nature* **593**, 223–227 (2021).
  39. Taitz, Y., Weininger, D. & Delany, J. J. *Daylight Theory: SMARTS - A language for describing molecular patterns*, <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> (1997).
  40. Molga, K., Gajewska, E. P., Szymkuć, S. & Grzybowski, B. A. The logic of translating chemical knowledge into machine-processable forms: a modern playground for physical-organic chemistry. *React. Chem. Eng.* **4**, 1506–1521 (2019).
  41. Roszak, R., Beker, W., Molga, K. & Grzybowski, B. A. Rapid and accurate prediction of pKa values of C–H Acids using graph convolutional neural networks. *J. Am. Chem. Soc.* **141**, 17142–17149 (2019).
  42. Mayr, H. & Patz, M. Scales of nucleophilicity and electrophilicity: A system for ordering polar organic and organometallic reactions. *Angew. Chem. Int. Ed.* **33**, 938–957 (1994).
  43. *Mayr's Database Of Reactivity Parameters - Start page.* (2023) Available at: <https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/> (Accessed: 6th December 2023).
  44. Chavan, S. R. et al. Iminosugars spiro-linked with morpholine-fused 1,2,3-triazole: Synthesis, conformational analysis, glycosidase inhibitory activity, antifungal assay, and docking studies. *ACS Omega* **2**, 7203–7218 (2017).
  45. Tanaka, N. et al. Isolation and structures of attenols A and B. Novel bicyclic triols from the Chinese bivalve *Pinna attenuata*. *Chem Lett* **28**, 1025–1026 (1999).
  46. Chen, D. et al. Discovery, structural insight, and bioactivities of BY27 as a selective inhibitor of the second bromodomains of BET proteins. *Eur. J. Med. Chem.* **182**, 111633 (2019).
  47. Teiji, K. et al. Multi-cyclic cinnamide derivatives. Patent US 2007219181A1 (2007).
  48. Sikorski, W. H. & Reich, H. J. The Regioselectivity of addition of organolithium reagents to enones and enals: The role of HMPA. *J. Am. Chem. Soc.* **123**, 6527–6535 (2001).
  49. Ireland, R. E., Armstrong, I., Lebreton, J. D. J., Meissner, R. S. & Rizzacasa, M. A. Convergent synthesis of polyether ionophore antibiotics: synthesis of the spiroketal and tricyclic glycal segments of monensin. *J. Am. Chem. Soc.* **115**, 7152–7165 (1993).
  50. Danishefsky, S. J., DeNinno, S. & Lartey, P. A concise and stereoselective route to the predominant stereochemical pattern of the tetrahydropyranoid antibiotics: an application to indanomycin. *J. Am. Chem. Soc.* **109**, 2082–2089 (1987).
  51. Parker, K. A. & Georges, A. T. Reductive aromatization of quinols: synthesis of the C-aryl glycoside nucleus of the papulacandins and chaetiacandin. *Org. Lett.* **2**, 497–499 (2000).
  52. Gurjar, M. K., Krishna, L. M., Reddy, B. S. & Chorghade, M. S. A versatile approach to anti-asthmatic compound CMI-977 and its six-membered analogue. *Synthesis* **2000**, 557–560 (2000).
  53. Banwell, M. G. et al. Small molecule glycosaminoglycan mimetics. Patent WO 2006135973A1 (2006).
  54. Mattson, R. J. & Catt, J. D. Piperazinyl-cyclohexanes and cyclohexenes. Patent US 6153611A (2000).
  55. Chongqing, P., Zhu, Z., Zhang, M. & Gu, Z. Palladium-catalyzed enantioselective synthesis of 2-aryl cyclohex-2-enone atropisomers: platform molecules for the divergent synthesis of axially chiral biaryl compounds. *Angew. Chem. Int. Ed.* **56**, 4777–4781 (2017).
  56. Mahecha-Mahecha, C. et al. Sequential Suzuki–Miyaura coupling/Lewis acid-catalyzed cyclization: an entry to functionalized cycloalkane-fused naphthalenes. *Org. Lett.* **22**, 6267–6271 (2020).
  57. Xu, W. & Yoshikai, N. Cobalt-catalyzed directed C–H alkenylation of pivalophenone N–H imine with alkenyl phosphates. *Beilstein J. Org. Chem.* **14**, 709–715 (2018).
  58. Kang, D., Kim, J., Oh, S. & Lee, P. H. Synthesis of naphthalenes via platinum-catalyzed hydroarylation of aryl enynes. *Org. Lett.* **14**, 5636–5639 (2012).
  59. Zhang, X., Sarker, S. & Larock, R. C. Synthesis of naphthalenes and 2-naphthols by the electrophilic cyclization of alkynes. *J. Org. Chem.* **71**, 236–243 (2006).
  60. Kumar, S. V. et al. Cyclocondensation of arylhydrazines with 1, 3-bis (het) arylmonothio-1, 3-diketones and 1, 3-bis (het) aryl-3-(methylthio)-2-propenones: Synthesis of 1-aryl-3, 5-bis (het) arylpyrazoles with complementary regioselectivity. *J. Org. Chem.* **78**, 4960–4973 (2013).
  61. Mohan, C., Singh, P. & Mahajan, M. P. Facile synthesis and regioselective thio-Claisen rearrangements of 5-prop-2-ynyl/enyl-sulfanyl pyrimidinones: transformation to thienopyrimidinones. *Tetrahedron* **61**, 10774–10780 (2005).
  62. Splivallo, R. & Ebeler, S. E. Sulfur volatiles of microbial origin are key contributors to human-sensed truffle aroma. *Appl. Microbiol. Biotechnol.* **99**, 2583–2592 (2015).

63. Dai, J. et al. New oblongolides isolated from the endophytic fungus *Phomopsis* sp. from *Melilotus dentata* from the shores of the Baltic Sea. *Eur. J. Org. Chem.* **2005**, 4009–4016 (2005).
64. Bunyapaiboonsri, T., Yoiprommarat, S., Srikitikulchai, P., Sri-chomthong, K. & Lumyong, S. Oblongolides from the Endophytic Fungus *Phomopsis* sp. BCC 9789. *J. Nat. Prod.* **73**, 55–59 (2010).
65. Shing, T. K. M. & Yang, J. A short synthesis of natural (-)-oblongolide via an intramolecular or a transannular Diels-Alder reaction. *J. Org. Chem.* **60**, 5785–5789 (1995).
66. Magedov, I. V. et al. Reengineered epipodophyllotoxin. *Chem. Commun.* **48**, 10416–10418 (2012).
67. Hur, J., Jang, J. & Sim, J. A Review of the pharmacological activities and recent synthetic advances of  $\gamma$ -butyrolactones. *Int. J. Mol. Sci.* **22**, 2769 (2021).
68. Le, H. V. et al. Design and mechanism of tetrahydrothiophene-based  $\gamma$ -aminobutyric acid aminotransferase inactivators. *J. Am. Chem. Soc.* **137**, 4525–4533 (2015).
69. Kalashnikov, A. I., Sysolyatin, S. V., Sakovich, G. V., Sonina, E. G. & Shchurova, I. A. Facile method for the synthesis of oseltamivir phosphate. *Russ. Chem. Bull.* **62**, 163–170 (2013).
70. Tavakoli, M., Chiu, Y. T. T., Baldi, P., Carlton, A. M. & Van Vranken, D. RMechDB: A public database of elementary radical reaction steps. *J. Chem. Inf. Model.* **63**, 1114–1123 (2023).
71. Tavakoli, H. R., Moosavi, S. M. & Bazgir, A.  $ZrOCl_2 \cdot 8H_2O$  as an efficient catalyst for the synthesis of dibenzo [*b*,*i*]xanthene-tetraones and fluorescent hydroxyl naphthalene-1,4-diones. *Res. Chem. Intermed.* **41**, 3041–3046 (2015).
72. Liu, D., Zhou, S. & Gao, J. Room-temperature synthesis of hydroxynaphthalene-1,4-dione derivative catalyzed by phenylphosphinic acid. *Synth. Commun.* **44**, 1286–1290 (2014).
73. Shaabani, S., Naimi-Jama, M. R. & Maleki, A. Synthesis of 2-hydroxy-1,4-naphthoquinone derivatives via a three-component reaction catalyzed by nanoporous MCM-41. *Dyes Pigment* **122**, 46–49 (2015).
74. Shaterian, H. R. & Mohammadnia, M. Effective preparation of 2-amino-3-cyano-4-aryl-5,10-dioxo-5,10-dihydro-4H-benzo[*g*]chromene and hydroxyl naphthalene-1,4-dione derivatives under ambient and solvent-free conditions. *J. Mol. Liq.* **177**, 353–360 (2013).
75. Tayama, E., Sato, R., Takedachi, K., Iwamoto, H. & Hasegawa, E. A formal method for the de-N,N-dialkylation of Sommelet–Hauser rearrangement products. *Tetrahedron* **68**, 4710–4718 (2012).
76. Kim, S. H., Lee, H. S., Kim, K. H. & Kim, J. N. An expedient synthesis of poly-substituted 1-arylisquinolines from  $\delta$ -ketonitriles via indium-mediated Barbier reaction protocol. *Tetrahedron Lett* **50**, 6476–6479 (2009).
77. Irwin, J. J. et al. A Free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
78. Szymkuć, S., Wołos, A., Roszak, R. & Grzybowski, B. A. Estimation of multicomponent reactions' yields from networks of mechanistic steps. *Nat. Commun.* (2024) In press.
79. Saebi, M. et al. On the use of real-world datasets for reaction yield prediction. *Chem. Sci.* **14**, 4997–5005 (2023).
80. Liu, Z., Moroz, Y. S. & Isayev, O. The challenge of balancing model sensitivity and robustness in predicting yields: a benchmarking study of amide coupling reactions. *Chem. Sci.* **14**, 10835–10846 (2023).
81. Beker, W. et al. Machine learning may sometimes simply capture literature popularity trends: A case study of heterocyclic Suzuki–Miyaura coupling. *J. Am. Chem. Soc.* **144**, 4819–4827 (2022).
82. Skoraczynski, G. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **7**, 3582 (2017).
83. Schneider, N., Lowe, D. M., Sayle, R. A. & Landrum, G. A. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *J. Chem. Inf. Model.* **55**, 39–53 (2015).

## Acknowledgements

Development of the MECH module within the Allchemy platform (by R.R., A.W., K.M., T.K., B.M.-K., S.S., M.M.) was supported by internal funds of Allchemy, Inc. A.M., L.G., Y.B., P.G., O.P. gratefully acknowledge funding from the National Science Centre, Poland (Award 2018/30/A/ST5/00529). S.B. was supported by a grant from the Priority Research Area Anthropocene under the Strategic Programme Excellence Initiative at the Jagiellonian University. J. M. gratefully acknowledge funding from the Foundation for Polish Science (award TEAM/2017-4/38) – these three awards supported part of experimental validations described in this paper. O.V. and D.T.G. gratefully acknowledges support from the Polish National Science Center, Poland (grants OPUS 2020/37/B/ST4/00017) and the Foundation for Polish Science (TEAM POIR.04.04.00-00-3CF4/16-00). The experimental part of this project also received funding from the European Research Council (ERC) under the European Union's or Horizon Europe research and innovation programme (Grant agreement No. 101097337, ARCHIMEDES to D.T.G.). During paper's revision, L.G., Y.B., and R.F. were also generously supported by the Institute for Basic Science, Korea (Project Code IBS-R020-D1). Analysis of pathways and writing of the paper by B.A.G. was also supported by the Institute for Basic Science, Korea (Project Code IBS-R020-D1).

## Author contributions

R.R., A.W., K.M., S.S., M.M. and B.A.G. designed and developed Allchemy platform and performed analyses and calculations described in the paper. A.M. performed syntheses described in Fig. 3, Y.B. and P.G. performed syntheses described in Fig. 4a, B.M.-K. and T.K. performed syntheses described in Fig. 4b, Y.B. performed syntheses described in Fig. 4c, L.G. performed syntheses described in Fig. 4d, O.P. performed syntheses described in Fig. 4a with supervision from J.M., P.G. performed syntheses described in Fig. 4c, S.B. performed syntheses described in Supplementary Section S2 with supervision from J.M. and help from R.F. O.V. and D.T.G. helped with the evaluation of the kinetic networks. B.A.G. conceived and supervised research and wrote the paper with help from other authors.

## Competing interests

The authors declare the following competing interests: R.R., A.W., K.M., B.M.-K., T.K., S.S., M.M. and B.A.G. are consultants and/or stakeholders of Allchemy, Inc. Allchemy software and its MECH module is property of Allchemy Inc., USA. All queries about access options to Allchemy, including academic collaborations, should be sent to [saraszynkuc@allchemy.net](mailto:saraszynkuc@allchemy.net). L.G., A.M., Y.B., P.G., O.P., S.B., R.F., J.M., O.V. and D.T.G. declare no competing interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54611-5>.

**Correspondence** and requests for materials should be addressed to Daniel T. Gryko or Bartosz A. Grzybowski.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024



**Supplementary Information** for Manuscript titled “*Systematic, computational discovery of multicomponent and one-pot reactions.*”

Rafał Roszak<sup>1+</sup>, Louis Gadina<sup>2+</sup>, Agnieszka Wołos<sup>1+</sup>, Ahmad Makkawi<sup>2</sup>, Barbara Mikulak-Klucznik<sup>1</sup>, Yasemin Bilgi<sup>2</sup>, Karol Molga<sup>1,2</sup>, Patrycja Gołębiowska<sup>2</sup>, Oskar Popik<sup>2</sup>, Tomasz Klucznik<sup>1</sup>, Sara Szymkuć<sup>1</sup>, Martyna Moskal<sup>1</sup>, Sebastian Baś<sup>2,3</sup>, Rafał Frydrych<sup>2</sup>, Jacek Mlynarski<sup>2</sup>, Olena Vakuliuk<sup>2</sup>, Daniel T. Gryko<sup>2\*</sup> & Bartosz A. Grzybowski<sup>2,4,5\*</sup>

<sup>1</sup> Allchemy, Inc., Highland, IN, USA.

<sup>2</sup> Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland.

<sup>3</sup> Jagiellonian University, Krakow, Poland

<sup>4</sup> Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea.

<sup>5</sup> Department of Chemistry, Ulsan Institute of Science and Technology, UNIST, Ulsan 44919, Republic of Korea.

+ Authors contributed equally

\*Corresponding authors. E-mail: daniel.gryko@icho.edu.pl (D.T.G.), nanogrzybowski@gmail.com (B. A. G.).

**This PDF file includes:** Supplementary Figures S1 to S163, Supplementary Tables S1 to S20 as well as Supplementary References 1 to 68.

## TABLE OF CONTENTS:

### **Section S1.** User manuals. (p. S-3)

**S1.1.** User manual for the mechanistic WebApp (for dynamic, on-the-fly calculations).

**S1.1.2.** Setting up a new search.

**S1.1.3.** Results analysis.

**S1.1.4.** Sequence analysis.

**S1.1.5.** Network view.

**S1.2.** User manual for the static WebApp (MCR championship and experimentally validated sequences within this publication).

**S1.2.1.** MCR Championship results.

**S1.2.2.** Experimentally validated sequences.

**S1.3.** User manual for interactive t-SNE map.

### **Section S2.** An example of Allchemy's incorrect prediction. (p. S-26)

### **Section S3.** Additional comments and evaluation of other methods for mechanism prediction.. (p. S-29)

**S3.1.** ReactionPredictor: Single-step and pathway predictors.

**S3.2.** RMechDB

**S3.3.** Ab initio molecular dynamics, AIMD, simulations.

**S3.4.** Reproducing reaction mechanisms with Machine-Learning models.

### **Section S4.** List of patterns of reactive groups. (p. S-59)

### **Section S5.** Tutorial on reaction rule coding (p. S-69)

### **Section S6.** Experimental section. (p. S-84)

**S6.1.** Reaction **Mach1** described in main-text **Figure 3** (variant without HMPA).

**S6.2.** Reaction **Mach2** described in main-text **Figure 3** (variant with HMPA).

**S6.3.** Reaction **Mach3** described in main-text **Figure 4a**.

**S6.4.** Reaction **Mach4** described in main-text **Figure 4b**.

**S6.5.** Reaction **Mach5** described in main-text **Figure 4c** (variant with the addition of DBU).

**S6.6.** Reaction **Mach6** described in main-text **Figure 4c** (variant with the addition of acetyl chloride).

**S6.7.** Reaction **Mach7** described in main-text **Figure 4d**.

**S6.8.** Reaction **Mach8** described in main-text **Figure 5a-b** (variant with the ring-expanding rearrangement).

**S6.9.** Reaction **Mach9** described in main-text **Figure 5a-b** (variant with the application of Heck reaction).

**S6.10.** Reaction **Mach10** described in main-text **Figure 5c-e**.

**S6.11.** X-ray crystallographic data for compound **6b'**

**S6.12.** Reaction optimization tables.

**S6.13.** Allchemy screenshots of mechanistic networks and reaction pathways.

### **Section S7.** Supplementary References (p. S273)

## Section S1. User manuals.

### Section S1.1. User manual for the mechanistic WebApp (for dynamic, on-the-fly calculations)

#### *Section S1.1.1. Login and basic information.*

Web Application for the discovery of new reaction types is available for testing by academic users at <https://mech.allchemy.net/> (given server capacity, to 20 concurrent academic users on a rolling basis and two-week slots). To create a new account, send a request to [admin@allchemy.net](mailto:admin@allchemy.net) from your academic e-mail address. For optimal performance, we recommend using web-browsers supporting SVG2, like Google Chrome. Web Application is accessible after logging in using credentials provided by the Allchemy team. After login, a pop-up window containing software recommendations and a short tutorial will appear. Given that calculations performed in Allchemy can easily result in combinatorial “explosion” exceeding the capacity of our servers, the searches are limited to the maximum of 6 synthetic generations, starting from at most 4 substrates and up to 3 common reagents (e.g., ammonia, sulfur, azide). Maximum duration of a single search is set to 15 minutes (soft time limit), and the molecular mass of produced molecules cannot exceed 500 g/mol.

After logging into Allchemy, user will see a *New Search* tab and a top bar consisting of three sections (with *New Search* being active by default). The first tab is used to set up a new calculation (**Supplementary Figure S1a**, see more details in **Section S1.1.2.**), the second tab, *Results*, displays calculated results (**Supplementary Figure S1b**, described in **Section S1.1.3**), or a list of ongoing/enqueued calculations (**Supplementary Figure S1c**). The last tab, *Saved results*, (**Supplementary Figure S1d**) can be used to access previously calculated results. Please note that temporarily saved results are removed when the user logs out of Allchemy – only *Saved calculations* are retained.

**a**

New search

Results

Saved results

## Specify starting materials

Create new collection

Show selected

## Search parameters

Number of synthetic generations: 

Temperature range:

Reaction conditions:

Solvent(s):

## Advanced options

**Search****b**

New search

Results

Saved results

Filters

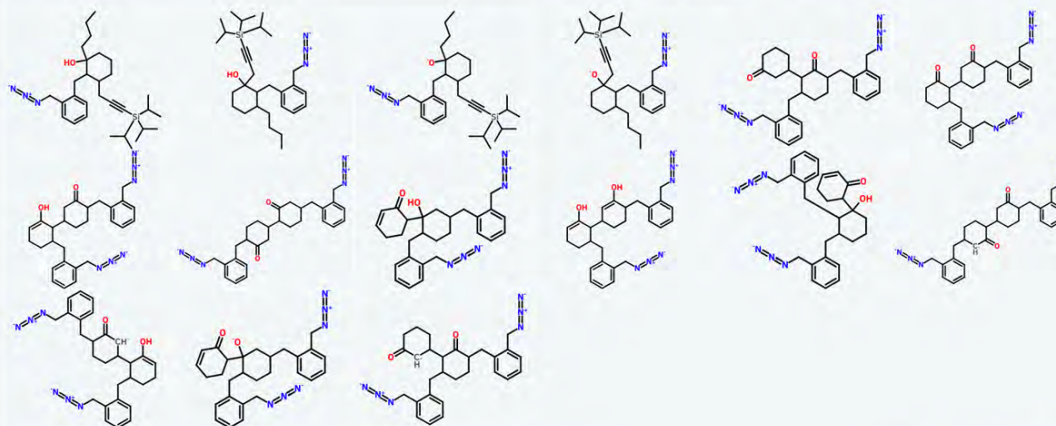
sort by: 

Sort again

375 products

Show as graph

search info



prev page

page 1 of 25

Save

go to page:  Go next page**c**

New search

Results

Saved results

Server is **BUSY**

You have 1 running calculation(s):

1. calculation started at Wed Mar 13 13:07:47 2024 [Terminate](#) [search info](#)

There is one calculation in queue. Your calculation(s) in queue:

1st position in queue [Cancel](#) [search info](#)**d**

New search

Results

Saved results

## Saved calculations

## Temporarily stored calculations

- 2021-06-01T19:32:32.507Z :: (18 products) [show](#)
- 2021-07-12T15:55:44.759Z :: (33 products) [show](#)
- 2021-07-21T16:15:18.786Z :: (296 products) [show](#)
- 2021-08-03T15:07:59.875Z :: (17 products) [show](#)

**Supplementary Figure S1. Overview of Allchemy's mechanistic module.** **a**, *New search* tab allows for setting up new calculations using search settings panel, **b**, *Results* tab is used to display calculated results or **c**, access information about ongoing/enqueued calculations. **d**, *Saved results* tab is used to retrieve temporarily (available until logging out) or permanently saved results.

### ***Section S1.1.2. Setting up a new search.***

To set up a new search, the user needs to 1) define starting materials 2) specify search parameters 3) if necessary, use advanced options to narrow down or expand calculations, 4) press *Search* button to start the calculations.

1) Starting materials can be defined in three ways (**Supplementary Figure S2a**, i): uploading a .txt file (with a list of SMILES separated by dots), using Allchemy's drawing editor, or from text (list of SMILES separated by dots). The user can introduce up to four starting materials. To facilitate selection, a curated set of ~2400 simple and commercially available substrates is available in the software (see also **Section S4**). Additionally, the user can also select up to three common reagents. To use this option, a) select a checkbox (**Supplementary Figure S2a**, ii), b) click on *add common substrates* hyperlink (**Supplementary Figure S2a**, ii), c) select up to three molecules via left-clicking on their panels (**Supplementary Figure S2b**, ix), d) click on the *Keep selected* button (**Supplementary Figure S2b**, x), e) close a window using "x" button placed in the top-right corner (**Supplementary Figure S2b**, xii). The starting materials thus specified can be reviewed in the *Show selected* section (**Supplementary Figure S2a**, iii) by clicking on the *Show molecules* hyperlink.

2) After starting materials are chosen, the user can adjust search parameters, especially the number of synthetic generations (up to 6, **Supplementary Figure S2a**, iv). Additionally, it is possible to narrow down the search by applying only reactions possible under a defined temperature range (e.g., exclude reactions that require very low or very high temperatures), reaction conditions (basic/acidic) and solvents (**Supplementary Figure S2a**, v). In default setting, all temperatures, conditions and solvents are allowed.

3) Advanced options can be used to impose additional limits on the generated molecules (molecular mass, number of heavy atoms, chiral centers or halogens), which will narrow down the search results and shorten calculation time (**Supplementary Figure S2a**, vi). There are also two options that allow for broadening the results (**Supplementary Figure S2a**, vii) – the user can include reactions that are catalyzed by transition metals, by unselecting checkbox *Exclude reactions catalyzed by transition metals*, and/or turn on a “permissive”/exploratory calculation approach (considering a smaller set of incompatible groups and allowing for some level of non-selectivity, assuming that these aspects can be tweaked during experimental reaction optimization). Please mind that application of these options will also increase calculation time.

4) Once the substrates are selected and all search parameters are set to user's preference, calculation can be started by clicking on the *Search button* located at the bottom of the page (**Supplementary Figure S2a**, viii).



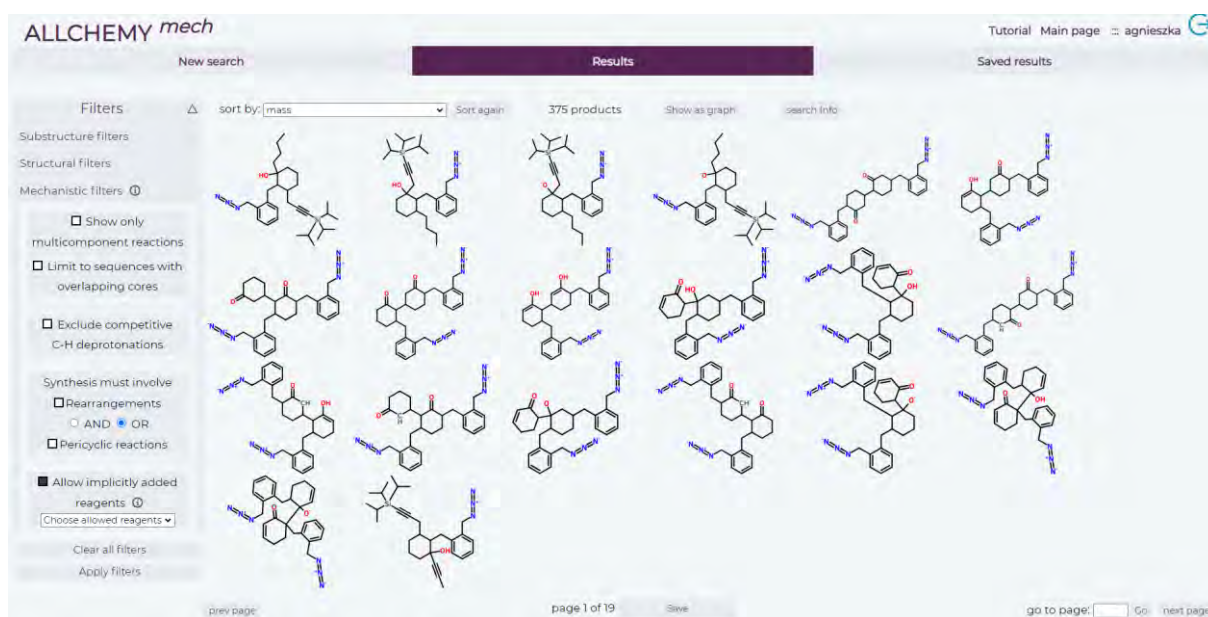


**Supplementary Figure S2. a, New search panel.** (i) User can introduce up to four substrates by uploading .txt file (with a list of SMILES separated by dots), by using Allchemy's drawing editor, or from text (list of SMILES separated by dots). (ii) Additionally, user can add up to three simple reagents after selecting a checkbox and clicking on the *add common substrates* hyperlink. (iii) Selected starting materials can be reviewed upon clicking on the *Show molecules* hyperlink. (iv) Number of synthetic generations can be chosen using numeric input box (up to 6). (v) In order to exclude reactions conducted under undesired conditions, unselect appropriate checkboxes in the *Search parameters* section. (vi) Limits for intermediates can be set in the *Advanced options* section in order to narrow down the search (and respectively reduce its time). (vii) Search can be also expanded by including reactions catalyzed by transition metals or turning on the "permissive" calculation approach. Please mind that use of any of these options will increase calculation time. (viii) Start the calculation by clicking the *Search* button. **b, Add common substrates window** available upon selecting a checkbox and clicking on the corresponding hyperlink. (ix) Select up to three molecules by left-clicking on them. (x) Click on the *Keep selected* button. (xi) In order to restore the full list, click on the *Default collection* button and repeat selection process. (xii) Close the window using "X" button placed in the top-right corner.

### ***Section S.1.1.3. Results analysis.***

After the calculation has been launched, the user is transferred to the *Results* tab with the list of all currently enqueued user's calculations. Allchemy search can take from a couple of seconds to several minutes (with a soft time limit of 15 minutes and hard limit of 45 minutes) and after completion, results will be loaded automatically. If the calculation reaches the time limit, the user will be presented with the results of all synthetic generations that will have been completed by that point.

In Allchemy, there are two modes of results' presentation. The first one, displayed by default, is a paginated view of small molecular structure panels (**Supplementary Figure S3**). Since a large number of molecules can be produced in a single calculations (up to thousands), implemented filtering and sorting functionalities can facilitate navigation through this synthetic space (**Supplementary Figure S3 (i),(iii)** and **Supplementary Figure S4**).



**Supplementary Figure S3. Results page, default panel view.** Search results displayed in a grid as tiles with molecular structures in them. Number of products can be narrowed down using various filtering options (i) that are applied after clicking *Apply filters* button. To change the sorting mode, select a desired option from the drop-down menu (iii) and click *Sort again* button. In order to access information about search settings and the list of starting substrates, click on the *Search info* button (vi). The display mode can be changed from the panel view to the network view via *Show as graph* functionality (v).



Filters are divided into three main categories – substructure, structural and mechanistic. Substructure filters allow to exclude undesired motifs from a predefined list, filter by presence of certain chemical elements or define specific structural motif(s) that should be kept or excluded (**Supplementary Figure S4a**). Structural filters enable the user to filter results according to properties like mass, logP, number of acidic groups, number of halogen atoms, etc. (**Supplementary Figure S4b**). Especially useful filtering option, *Exclude charged compounds*, is placed at the top of this list. Mechanistic filters (**Supplementary Figure S4c**) are geared specifically towards searching for interesting candidates for multicomponent/one-pot reactions. The user can here limit results only to potential multicomponent reactions, filter out products generated via sequences of reactions occurring on disjoint parts of one molecule, exclude sequences with competitive C-H deprotonations or limit to sequences involving interesting transforms like rearrangements or pericyclic reactions. Additionally, the user can filter out molecules, whose syntheses involve implicitly added reagents (all or only selected), like water, alcohols, particular metals. Selected filters are activated by clicking on the *Apply filters* button.

**Supplementary Figure S4. Filtering of the results.** Filters are designed to facilitate navigation through generated synthetic spaces and are divided into three main categories: **a**, Substructure, **b**, Structural and **c**, Mechanistic.

By default, results are sorted by molecular mass, but it is also possible to sort the results according to other properties (e.g., number of rings or stereocenters, as well as a *multicomponent score* defined as the increase of structural complexity per step (see Methods)), by selecting the desired property from a drop-down list (**Supplementary Figure S3** (iii)) and pressing the *Sort again* button (**Supplementary Figure S3** (iv)). The second mode of presenting results can be accessed by clicking the *Show as graph* button (**Supplementary Figure S3** (v)). After a couple of seconds, the results will render in a form of a reaction network. In order to display search settings of the current results, click on the *Search info* button (**Supplementary Figure S3** (vi)).

#### ***Section S1.1.4. Sequence analysis.***

Compound miniatures displayed in the panel view are interactive and clicking on them will open a pop-up window with details of mechanistic sequence(s) leading to this molecule (**Supplementary Figure S5a**). Each sequence tab is color-coded in accordance with the legend (**Supplementary Figure S5a**, i), denoting which sequences are classified as MCRs and which as one-pots. Additionally, for one-pot sequences, the short message indicating the reason why a given sequence was not deemed as multicomponent will be displayed (**Supplementary Figure S5a**, ii). Each mechanistic step is accompanied by its name, typical conditions and solvent and illustrative literature reference in a form of DOI hyperlink.

Some steps can be marked as potentially problematic, by a red exclamation mark placed next to the step name. Hovering over this icon will provide more information about the assigned warning (presence of the by-products showing cross-reactivity with the main sequence, the reaction mixture became too complex, detection of competing reaction(s) faster than reaction on the path).

If competing reaction(s) for a given step has been detected, *Show competing steps* hyperlink will be available under the literature reference (**Supplementary Figure S5a**, iii). Clicking on this hyperlink opens a pop-up window with a list of competing reactions (**Supplementary Figure S5b**).

Additionally, for each step, user can display two types of byproducts – byproducts of the particular transformation, and by-products already present in the reaction mixture (from previous reactions). This can be achieved by clicking on the “flask” icon (**Supplementary Figure S5a**, iv) and selecting desired mode from a pop-up window (**Supplementary Figure S5a**, v). Molecules present in the reaction mixture are displayed as interactive panels that can

be clicked in order to highlight the reaction of origin. Furthermore, compounds showing cross-reactivity are marked with a red frame.

If more than one sequence leading to the molecule has been found, additional tabs will be available in the top panel. By default, sequences are sorted by number of steps, with the shortest one presented as first. Sequences can be also re-ranked by MCR score by clicking on the *Sort by MCR score* hyperlink located under the top panel. Furthermore, if the user finds some sequence with high MCR score, interesting, it can be nominated to the MCR championship by clicking on the button located next to the multicomponent score. Upon clicking on it, the user will be asked to fill a short questionnaire (**Supplementary Figure S6**). If the sequence nominated by the user has MCR score placing it in top-50, it will be added to the list available at <https://mcrchampionship.allchemy.net/>. As the list will be dynamically updated with highest-scored candidates, reaction nominated by the user can also, at some point, fall off the top-50 list. Furthermore, please mind that sequences leading to ionic compounds are scored 0 and cannot be nominated.

Clicking upon *info* tab, located at the leftmost side of the top panel, will display molecular properties of the compound, as well as C-H pKa prediction of Allchemy's Graph Convolutional Neural Network<sup>1</sup> (**Supplementary Figure S7**).

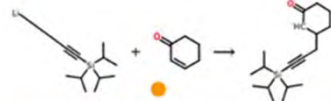
**a** info path 1 path 2 path 3 path 4  
 show path in a graph format  
 Sort by MCR score  
 Complexity score: 3.50 [Nominate MCR champion](#)

- Multicomponent reaction without warnings
- Multicomponent reaction with warnings
- One-pot reaction without warnings
- One-pot reaction with warnings
- Organocatalytic reaction without warnings
- Organocatalytic reaction with warnings
- One- or two- component reaction

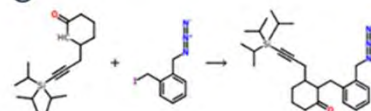
**(i)**  
 Reaction name: Synthesis of organolithium compound  
 Reaction conditions: nBuLi, THF  
 allow temp:V,L,L, allow AB: SB  
 Solvent: THF  
 Alternative Solvent: t-Butyl ethyl ether  
 Literature reference: 10.1016/S0040-4039(00)86930-0  
 show competing steps



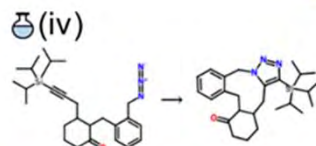
**(ii)**  
 Reaction name: Addition of organolithium compounds to activated alkenes  
 Reaction conditions: THF, HMPA, cooling  
 allow temp:L, allow AB: SB  
 Solvent: THF  
 Alternative Solvent: t-Butyl ethyl ether  
 Literature reference: 10.1016/S0040-4039(00)86930-0  
 show competing steps



**(iii)**  
 Reaction name: Alkylation of enolates (active I)  
 Reaction conditions: THF  
 allow temp:L, allow AB: SB,B  
 Solvent: THF  
 Alternative Solvent: t-Butyl ethyl ether  
 Literature reference: TBD  
 show competing steps



**(iv)**  
 Reaction name: Huisgen cycloaddition  
 Reaction conditions: toluene, heating  
 allow temp:H, allow AB: LA,N,WB  
 Solvent: toluene  
 Literature reference: 10.1016/j.ejmech.2015.08.017 and 10.1016/j.tetlet.2010.12.040  
 show competing steps



Substrates incompatible with the reaction:

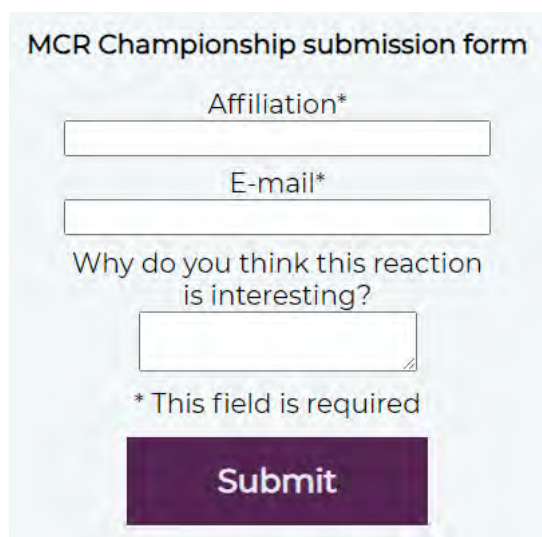
show by-products of this reaction

show molecules already present in reaction mixture

**b** Reaction name: Addition of organometallics to ketones  
 Reaction conditions: THF  
 allow temp:L,V,L, allow AB: SB  
 Solvent: THF  
 Alternative Solvent: t-Butyl ethyl ether  
 Literature reference: 10.1016/j.tetlet.2007.04.109 and 10.1016/j.tetlet.2012.05.050 and 10.1016/j.ejmech.2016.08.067 and 10.1016/j.tetlet.2020.152506

Reaction name: Addition of organometallics to ketones  
 Reaction conditions: THF  
 allow temp:L,V,L, allow AB: SB  
 Solvent: THF  
 Alternative Solvent: t-Butyl ethyl ether  
 Literature reference: 10.1016/j.tetlet.2007.04.109 and 10.1016/j.tetlet.2012.05.050 and 10.1016/j.ejmech.2016.08.067 and 10.1016/j.tetlet.2020.152506

**Supplementary Figure S5. Details of a mechanistic sequence. a,** Each mechanistic step along the sequence is accompanied by reaction name, typical reaction conditions, typical solvent and illustrative literature reference as DOI hyperlink. As multiple sequences can lead to the same molecule, they are available upon clicking on corresponding tabs located in the top panel. Tabs are colored according to the legend (i). If a pathway has been classified as one-pot, hovering over yellow exclamation mark (ii) will display the short message indicating the reason. If side-reactions have been detected, *Show competing steps* hyperlink will appear (iii). Clicking upon it will open a pop-up window with the list of recognized competing reactions, as in the yellow frame in panel (b). Also, for each mechanistic step, by-products can be displayed upon clicking the flask icon (iv). The user can display either by-products of a particular step or by-products that have accumulated in the reaction mixture from the beginning of the sequence (v). Structures of the latter ones are interactive, so clicking on a molecule, will highlight step, in which it was formed.



MCR Championship submission form

Affiliation\*

E-mail\*

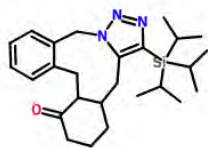
Why do you think this reaction is interesting?

\* This field is required

Submit

**Supplementary Figure S6. Short questionnaire for nominating a pathway for the MCR championship.** It is available upon clicking on the *Nominate MCR champion* button (Supplementary Figure S5). The user is asked to fill in an affiliation and e-mail address, as well as a short explanation why this pathway has been nominated. Evaluation of the submission by our in-house chemists can take up to a week and only pathways with the MCR score placing it in top-50 will be added to the list available on <https://mcrchampionship.allchemy.net/>.

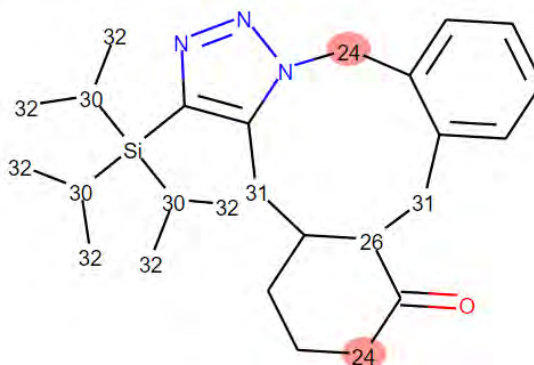
info path 1 path 2 path 3 path 4



Molecular formula	C <sub>26</sub> H <sub>39</sub> N <sub>3</sub> OSi	totalPSA	47.780 Å <sup>2</sup>
Mass	437.704 g/mol	N/C ratio	0.12
logP	5.296	Oxygen balance	-257.71 %

hide C-H pKa info

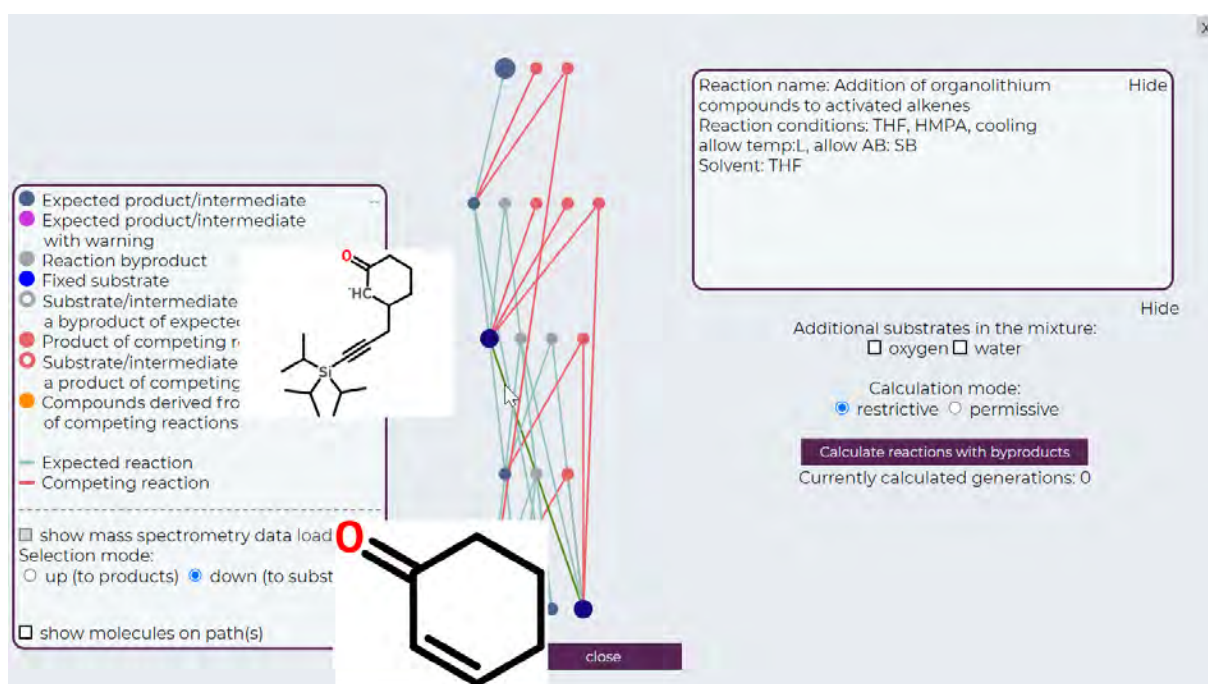
Prediction of Allchemy's Graph Convolutional Neural Network (pKa values in DMSO):



**Supplementary Figure S7.** Information about molecular properties of the compound, as well as C-H pKa prediction of Allchemy's Graph Convolutional Neural Network<sup>1</sup> is available upon clicking on the *Info* tab.

Each individual sequence can be also displayed as a graph. To access this presentation mode, click the *show path in a graph format* hyperlink. The graph displayed constitutes of nodes representing molecules of the main pathway (*blue*), by-products of reactions on pathway (*grey*) and products of competing/side reactions (*red*). Additionally, nodes of molecules that were generated in a step marked by a warning are *violet* while molecules on the main path, that are simultaneously by-products of other reaction(s) or products of competing reaction(s), have additional halo in the corresponding color. Legend describing color-coding of nodes is available in the left side panel. Hovering over a node will display structure of the corresponding molecule next to the node, while hovering over an edge, will display reaction details in the right side panel. Right-clicking on any node will highlight the pathway/sequence leading to the corresponding molecule. A pathway can be deselected by subsequent right-clicking on the node (**Supplementary Figure S8**).

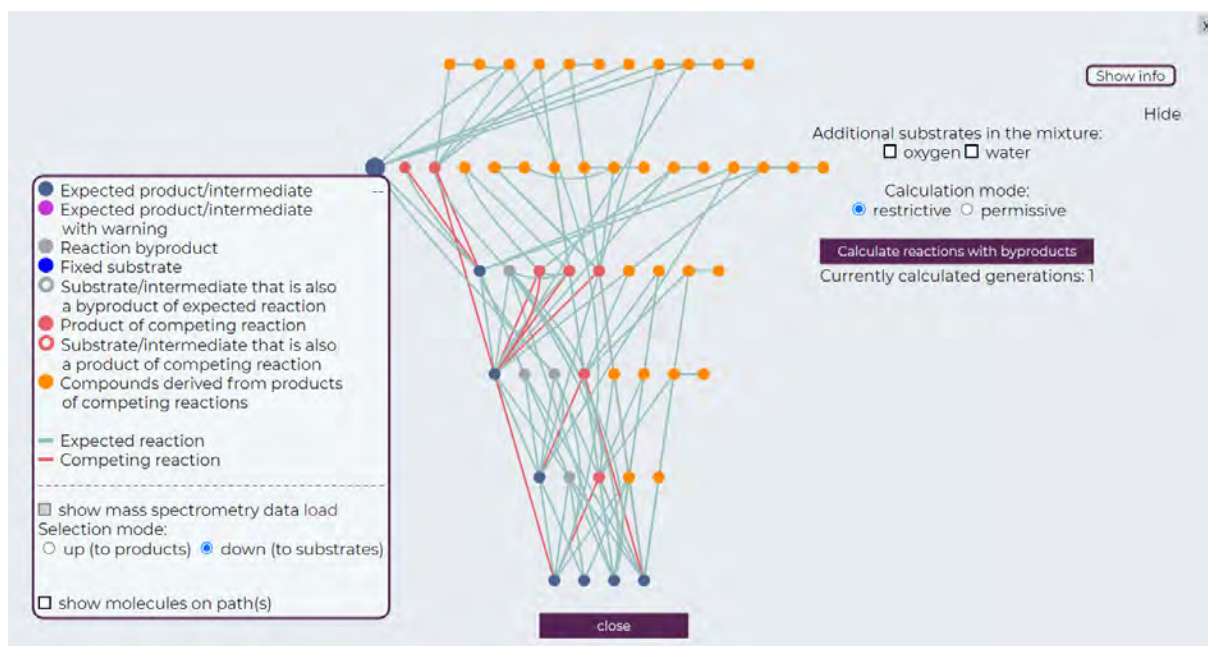




**Supplementary Figure S8. Graph representation of a mechanistic sequence – level 2, L2, analysis.** Blue nodes correspond to substrates and expected products/intermediates of the main pathway, grey nodes – by-products of reactions of the main pathway, red nodes – products of competing reactions involving molecules of the main pathway and transformations possible under similar conditions. Full legend describing color-coding is placed in the left panel. Hovering over a node displays structure of the corresponding molecule and hovering over an edge displays information about transformation in the right-hand panel. In order to expand analysis to level 3, click on the *Calculate reactions with byproducts* button. Calculation will take from few seconds up to few minutes depending on the size of L2 graph. After calculations are finished, a new network will upload, augmented with orange nodes corresponding with products of reactions of side-products and by-products with each other as well as with the molecules on the main pathway (**Supplementary Figure S9**). Analysis can be expanded up to Level 4 in the similar manner. Further expansions, although theoretically possible, have been disabled, as each further level requires more computational effort, simultaneously giving less probable side-products.

Furthermore, if a given mechanistic sequence has been chosen and committed to wet-lab synthesis, it is possible to color nodes according to mass spectroscopic data of a crude reaction mixture (.csv file with masses to the second decimal place in separate rows). In this way, the user can (i) determine which of the predicted by-products have been really formed in experiment the actual synthesis and (ii) trace the mechanistic pathways leading to these

confirmed by-products of the reaction sequence. Additionally, since for some reactions multiple possible reagents are proposed, such an analysis can help with reaction optimization by reducing/avoiding generation of some by-products.



**Supplementary Figure S9. Graph representation of a pathway – level 3, L3, analysis.** Level 2 graph from Supplementary Figure S8 augmented with orange nodes corresponding to products of reactions of side-products (*red nodes*) and by-products (*grey nodes*) with each other as well as with the molecules on the main pathway (*blue nodes*). In order to continue an analysis to Level 4, click on the *Calculate reactions with byproducts* button.

### ***Section S1.1.5. Network view.***

In addition to displaying individual sequences and networks (cf. above), *all* products of any Allchemy search – formed under any conditions allowed when setting up the search – can also be displayed in a graph/network view (**Supplementary Figure S10**) after clicking *Show as graph* button placed in the top panel of the *Results* tab (**Supplementary Figure S3 (v)**). In this view, molecules are represented as nodes; these nodes are connected by edges corresponding to mechanistic transforms. Molecules are placed in “synthetic generations” in which they were first produced, starting from the substrates at the very bottom. When hovering over a node, the miniature of a molecule’s structure will appear next to it. Left-clicking on the node will display a window with synthetic details of the sequence(s) (analogously to clicking on a structure in the panel view, see **Supplementary Figure S5**), and right-clicking will select and highlight the main pathway leading to this molecule. Additionally, if user right-clicks on a node with a pressed “Ctrl” key, alternative pathways leading to the molecule (up to 10 shortest) will be displayed on a graph in a form of colored arcs.

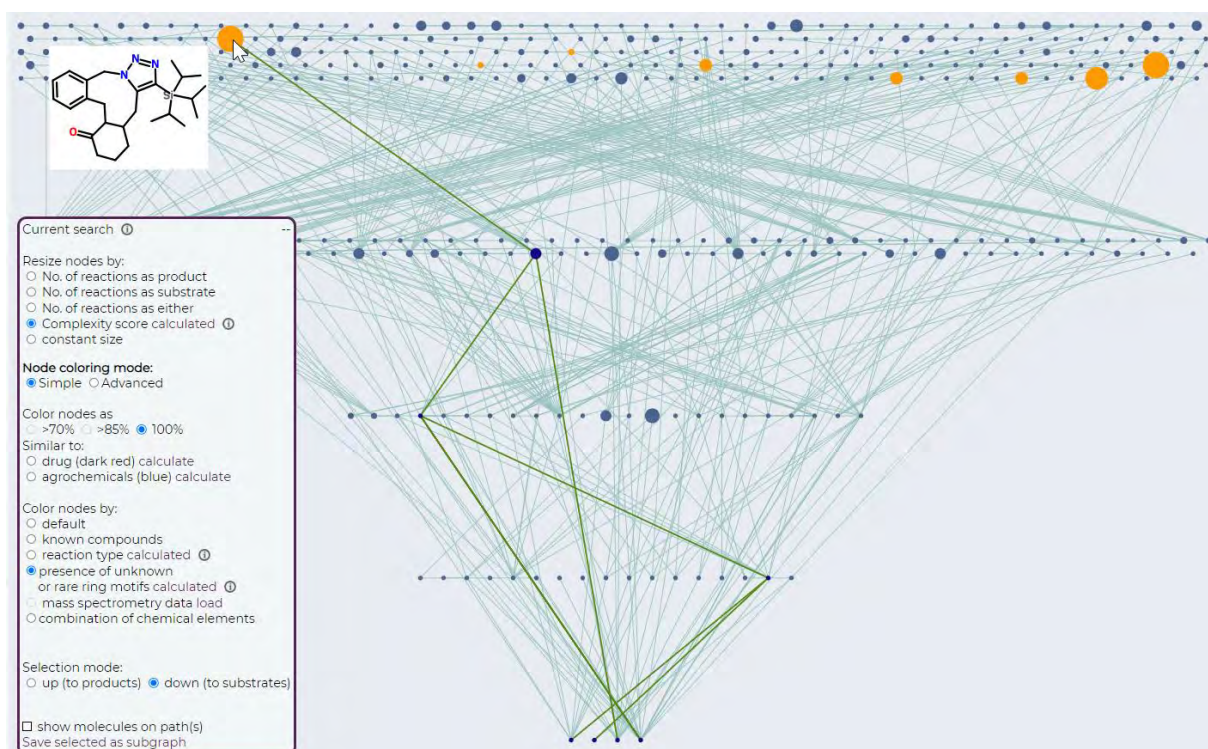
The expandable menu, with information about currently displayed search, as well as resizing/coloring options that can facilitate analysis of the network are located on the left hand-side.

Nodes of the graph can be resized according to their connectivity, or a “multicomponent score” which is a metric describing the increase of structural complexity along the pathway (see main text for more details), as well as colored by properties like similarity to drugs/agrochemicals, reaction type (MCR/one-pot, with/without warnings), presence of particular chemical element(s), presence of rare or unknown ring motifs. The last option can be used to highlight molecules with a ring system that is unprecedented in the literature. Additionally, as in the case of the analysis of a specific mechanistic sequence (**Supplementary Figures S8 and S9**), it is possible to upload mass spectrometry data, and color nodes corresponding to the mass peaks (.csv file with with masses to the second decimal place in separate rows). Furthermore, advanced coloring mode enables to color nodes by two different features at the same time, with one of them presented as a halo. Note that coloring by *reaction type* and *presence of unknown or rare ring motifs* requires prior calculation of these features, by clicking on the *calculate* hyperlink and features that were not calculated will not appear in the drop-down menu in advanced coloring panel.

Another feature that can be helpful in graph analysis is presentation of molecular structures next to the corresponding nodes along the pathways. It can be turned on using *show molecules on*

*path(s)* checkbox and work only on pathways selected after the checkbox has been turned on (molecule structures will not appear along the pathways selected prior to turning on the checkbox). If the number of displayed structures is too large to be clearly visible, a panel with molecular structures will appear above the graph, and hovering over a tile with a given structure will highlight the corresponding node in the network and vice versa, hovering over a node will put corresponding structure into focus.

Additionally, the user can save selected pathway(s) as a separate graph using option *Save selected as subgraph* located at the bottom of the expandable panel.



**Supplementary Figure S10. Network view.** Alternative way of presenting all search results available after clicking on the *Show as graph* button. Each layer represents molecules from a separate synthetic generation with substrates placed at the bottom and the last generation at the top. Selection of the shortest mechanistic pathway is possible by right-clicking on any node. Selection of alternative pathways (up to 10) is possible by right-clicking on the node with “Ctrl” key pressed. Left-side expandable panel was designed to help navigate the graph and facilitate analysis of results. Multiple resizing and coloring options are available, including: resizing by the number of reactions leading to/starting from a given node, multicomponent score, coloring by reaction type, presence of unknown or rare ring motifs, mass spectrometry data. Nodes presented in the figure are resized by multicomponent score and colored by the presence of unknown (*dark-green*) and rare (*light-green*) ring motifs. Choice of selection mode is also

available: *down* (default) tracing the pathway from product to its substrates, and *up* showing products generated using this molecule. Selection of *show molecules on path(s)* checkbox will display structures of the molecules next to the corresponding nodes or in the top panel (if number of molecules on the graph exceeds visibility threshold). Furthermore, it is possible to save selected path(s) as a separate graph using *Save selected as subgraph* feature.

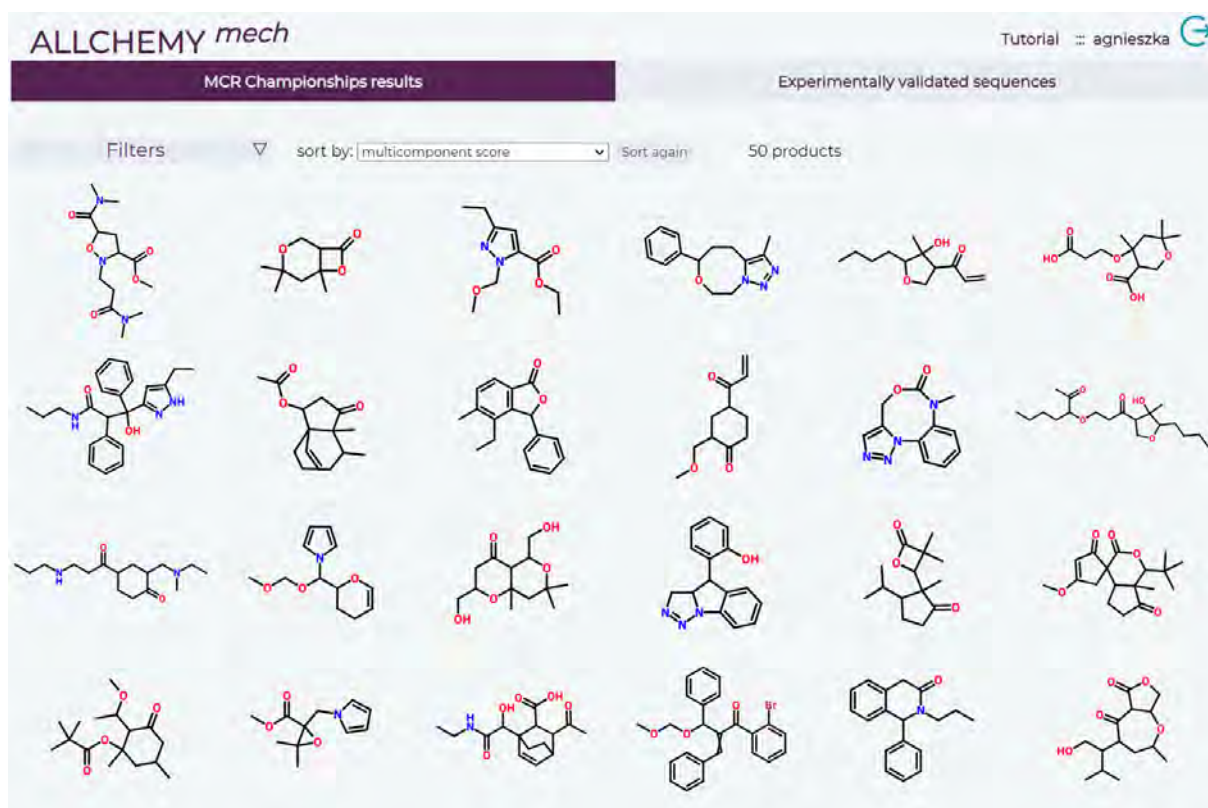
## **Section S1.2. User manual for the static WebApp (MCR championship and experimentally validated sequences within this publication)**

Web Application for browsing pre-calculated results of MCR championship, as well as sequences validated experimentally in this publication, is freely available to academic users at <https://mcrchampionship.allchemy.net/>. In order to create a new account, please send a request to [admin@allchemy.net](mailto:admin@allchemy.net) from your academic address. For optimal performance, we recommend using web-browsers supporting SVG2, like Google Chrome. Web Application is accessible after logging in using credentials provided by the Allchemy team. After login, a pop-up window containing software recommendations and a short tutorial will appear.

### ***Section S1.2.1. MCR Championship results.***

Results of automated discovery of new multicomponent reactions are displayed in the first tab, *MCR championship results* (as default it is an active tab after login). Different triples and quartets of substrates (from our pre-curated set of ~2400 reactive molecules) are systematically scanned, networks are propagated to G9 and 50 top-scoring warning-free candidates for MCRs and one-pots are published at this site. As calculations continue progress, this list of molecules is systematically updated if higher-scoring candidates are found. Furthermore, users of the dynamic WebApp are encouraged to nominate champions from their own calculations.

Page with results is similar to the one described in **Section 1.1.3** for dynamic calculations with exception of *Show as graph* and *Show info* functionalities. Molecules are by-default sorted by MCR-score, but other sorting options (e.g., by mass, number of rings) are also available. All filtering options, except *Exclude charged compounds*, described in **Section 1.1.3** are also available.



**Supplementary Figure S11. MCR championship results.** Results are displayed in a grid as tiles with molecular structures in them. Number of products can be narrowed down using various filtering options that are applied after clicking *Apply filters* button. To change the sorting mode, select a desired option from the drop-down menu and click *Sort again* button. List of molecules will be systematically updated if higher-scored candidates are found.

Left-clicking on any molecule opens a pop-up window with synthetic details of the sequence (like **Supplementary Figure S5**, described in **Section 1.1.4.**), but unlike in the WebApp for dynamic calculations, only one best-scored pathway is presented, hence sorting by MCR-score is not available.

### Section S1.2.2. Experimentally validated sequences.

Reactions that were found by Allchemy's mechanistic module and experimentally validated in this publication are posted in the *Experimentally validated sequences* tab (**Supplementary Figure S12**).

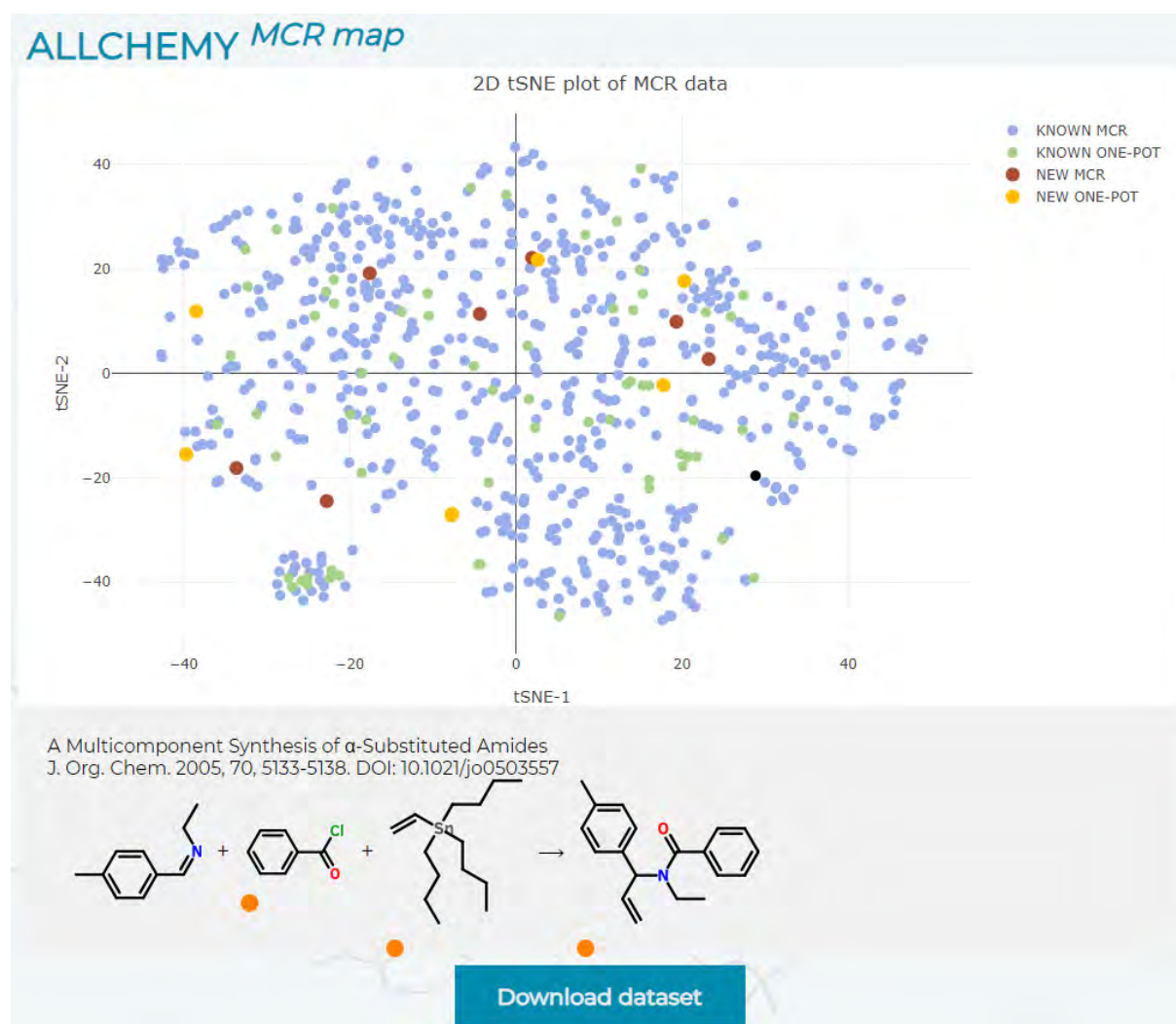
The screenshot displays the ALLCHEMY mech web interface. At the top left, the logo 'ALLCHEMY mech' is visible. To the right, there is a navigation bar with 'Tutorial :: agnieszka' and a refresh icon. Below the logo, the text 'MCR Championships results' is shown. A prominent purple button labeled 'Experimentally validated sequences' is centered at the top. Below this, a control bar includes a 'Filters' dropdown, a 'sort by:' menu set to 'multicomponent score', a 'Sort again' button, and a '10 products' indicator. The main area contains a grid of ten chemical structures. At the bottom, there are navigation controls: 'prev page', 'page 1 of 1', and 'go to page: [input] Go next page'.

**Supplementary Figure S12. Experimentally validated sequences.** Results are displayed in a grid as tiles with molecular structures in them. Molecules can be analyzed using various filtering options that are applied after clicking *Apply filters* button. To change the sorting mode, select a desired option from the drop-down menu and click *Sort again* button.



### Section S1.3. User manual for interactive t-SNE map.

Interactive t-SNE projection “map” illustrating diversity of 422 known classes (not specific examples!) of MCRs (smaller *blue* markers) and 63 one-pot sequences (*green*) vs. the newly discovered MCRs and one-pots (*red* and *yellow* markers, respectively) described in this work is freely available for academic users at <https://mcrmap.allchemistry.net>. In order to create a new account, please send a request to [admin@allchemistry.net](mailto:admin@allchemistry.net) from your academic address. Hovering over any node displays information about the corresponding reaction next to it: t-SNE coordinates, reaction name and literature reference together with DOI. Clicking on the node displays an example of reaction representative to this class, together with reaction name and reference below the map (the selected node changes color to black). The .csv file with the full reaction list can be downloaded by clicking on the *Download dataset* button (see **Supplementary Figure S13**).

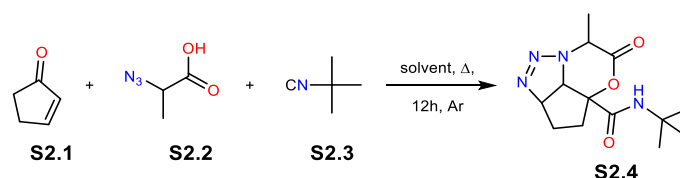


**Supplementary Figure S13.** Interactive t-SNE projection “map” illustrating diversity of known MCRs available at <https://mcrmap.allchemistry.net>. Hovering over any node displays

information about corresponding reaction class next to this node. Clicking on the node displays an example of reaction from a particular class below the map. Blue nodes represent known multicomponent reactions, green nodes – known one-pot reactions, red nodes – newly discovered multicomponent reactions and yellow nodes – newly discovered one-pot reactions. In order to download the dataset, click on *Download dataset* button.

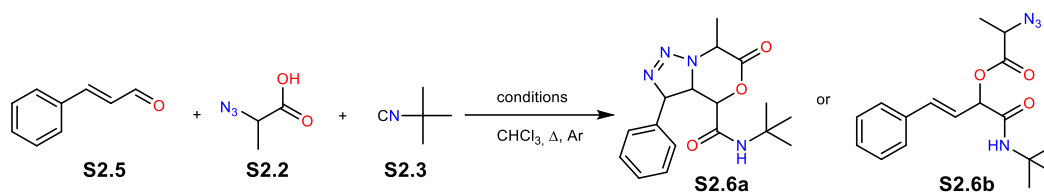
## Section S2. An example of Allchemy's incorrect prediction.

One of the MCR's proposed by Allchemy and subsequently committed to experimental validation involved  $\alpha,\beta$ -unsaturated cyclic ketone (**S2.1**),  $\alpha$ -azido substituted carboxylic acid (**S2.2**) and isocyanide (**S2.3**) resulting in the formation of tricyclic product (**S2.4**) in **Supplementary Figure S14**). The proposed mechanistic sequence entailed activation of the carbonyl groups by an acid, addition of isocyanide to the activated ketone, addition of carboxylic acid to the thus formed nitrilium ion, the Huisgen cycloaddition between the azide group and double bond in the ketone moiety, and finally the Mumm rearrangement leading to the final structure. The multicomponent reaction between carboxylic acids, ketones, or aldehydes and isocyanides is already known as the Passerini reaction<sup>2</sup>, and the algorithm's proposal appeared as an intriguing extension of this classic MCR.



**Supplementary Figure S14.** Allchemy's original proposal for a Passerini-Huisgen MCR.

Unfortunately, all screened conditions (solvents: THF,  $\text{CF}_3\text{CH}_2\text{OH}$ ,  $\text{MeOH}/\text{H}_2\text{O}$ , DMF) did not result in the formation of the desired product **S2.4** or even Passerini or Huisgen partial products. Further literature studies revealed that  $\alpha,\beta$ -unsaturated ketones are poor substrates in the Passerini reaction<sup>2</sup>. In addition,  $\alpha,\beta$ -unsaturated cyclic ketones have been shown to undergo decomposition after Huisgen cycloaddition with azides leading to  $\beta$ -amino-enones<sup>3</sup>. Allchemy's second suggestion was that cinnamic aldehyde **S2.5** instead of **S2.1** should undergo the same transformation, since aldehyde is known to be more reactive than ketones in the Passerini reaction<sup>2</sup>. This time, Passerini product **S2.6b** was obtained in yields up to 37% (**Supplementary Table S1**) but the subsequent cyclization did not take place.

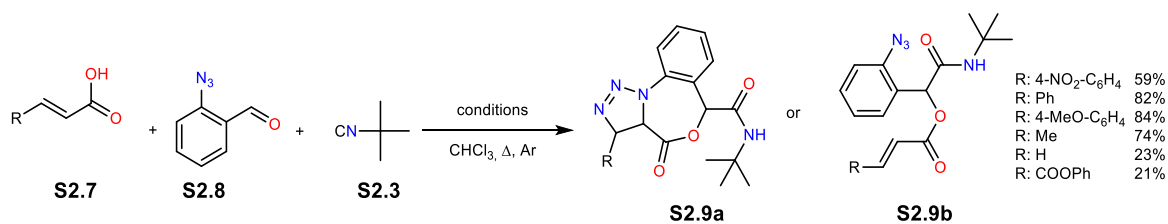


**Supplementary Table S1.** Condition screening for the multicomponent reaction involving aldehyde **S2.5**

	<b>5</b> [mmol]	<b>2</b> [mmol]	<b>3</b> [mmol]	CHCl <sub>3</sub> [ml]	Additives <sup>a</sup>	Time[h]	Yield <sup>b</sup> <b>S2.6a</b> / <b>S2.6b</b>
1	1	1	1	3	none	15	0/29%
2	1	1.5	1	3	none	48	0/30%
3	1	1	1.5	3	none	48	0/28%
4	1	1.5	1.5	3	none	48	0/33%
5	1	1	1	3	2mol% H <sub>2</sub> SO <sub>4</sub>	48	0/32%
6	1	1	1	neat	none	15	0/37%
7	1	1	1	1	none	15	0/30%
8	1	1	1	2	none	15	0/32%
9	1	1	1	3	none	15	0/12%

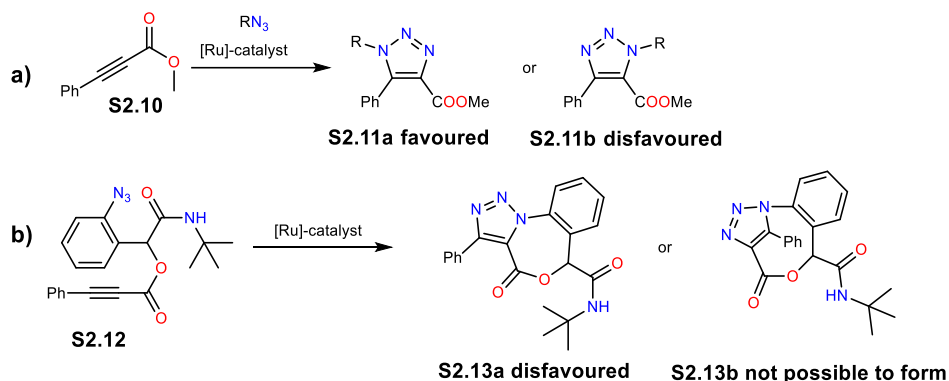
**General** procedure: aldehyde (0.5 mmol), acid (0.5 mmol) and isocyanide (0.5 mmol) in solvent (0.5 mL) were refluxed under argon for defined time, <sup>a</sup> the additive if added together with substrates, <sup>b</sup> isolated yield

Even when pure **S2.6b** was isolated, it did not undergo cyclization at higher temperatures, under microwave irradiation, in diverse solvents, or the presence of catalysts such as acids or metal zinc (zinc powder). The only side reaction observed during this experimental trial was ester hydrolysis causing the loss of the scaffold containing azide. The change from acid **S2.2** to 2-azidobenzoic acid did not solve the cycloaddition problem. The third approach proposed by the software assumed a switch of functional groups between aldehyde **S2.8** and carboxylic acid **S2.7** (**Supplementary Figure S15**). Various investigated  $\alpha,\beta$ -unsaturated acids resulted in a moderate to high yield of product **S2.9b** formation. Several side products were isolated, but none of them was the desired compound **S2.9a**. The change to terminal alkyne with propiolic acid and copper catalyst did not solve the problem, as such conditions were reported to initiate spontaneous decarboxylation during Huisgen cycloaddition<sup>4,5</sup>. This information was confirmed by gas release observed during the reaction and the absence of the desired product.



**Supplementary Figure S15.** The third generation of the investigated transformation.

Upon closer examination of the pertinent literature<sup>4-11</sup>, the above setbacks can be rationalized by group incompatibility and strain arguments – the latter emphasizing (as mentioned in the main text), the need for better evaluation of 3D conformations and near-attack conformers in Allchemy. The specific structural feature in question is the triple bond – as is well known, the use of copper catalysts is limited only to terminal alkynes<sup>7</sup>, whereas ruthenium-based catalysts (capable of activating internal triple bonds) are not compatible with carboxylic acids<sup>8</sup> and enforce high regioselectivity of cycloaddition<sup>9</sup> (**Supplementary Figure S16**). In our case, the cycloadditions were unsuccessful because the correct regioselectivity would necessarily develop high strain of the system (**S2.13b**)<sup>8-10</sup>.



**Supplementary Figure S16. Regioselectivity requirements affecting reaction outcomes. a,** The reported regioselectivity of ester **10** (see ref. <sup>9</sup>). **b,** Illustration why the Huisgen cycloaddition is not possible for the Passerini product **12**.

### Section S3. Additional comments and evaluation of other methods for mechanism prediction.

The vision of computers reasoning at the level of mechanistic steps and ultimately “inventing” new reactions is quite appealing and, as mentioned in the main text, has been pursued for decades. As we have recently discussed in our work on carbocation rearrangements<sup>12</sup>, most efforts to date have relied on quantum mechanical calculations which can be very CPU-intensive but, for at least some classes of problems, can offer valuable predictions (see, for instance, carbocation rearrangement problems<sup>13-14</sup> or pericyclic reactions<sup>15</sup>). In the domain of MCRs, Maeda performed quantum calculations in conjunction with the artificial force induced reaction (AFIR) method<sup>16</sup> to predict the scope of multicomponent *N*-difluoroalkylative dearomatization of pyridines. It should be noted, however, that this type of a MCR had been previously described<sup>17-18</sup> simplifying the analysis of key transition states.

In ref. <sup>12</sup> we argued that quantum approaches may be “too-fine” to predict *de novo* reactions involving long sequences of mechanistic steps – in such cases, the number of degrees of freedom may be too large to consider the problem *a priori*, as the space of possible “next” steps is expanding exponentially (see Extended Figure 3 in ref. <sup>12</sup>) while the inherent calculation inaccuracies (especially of transition-state geometries and energies) propagate. For MCRs, this is further compounded by the presence of multiple substrates and a myriad of possible intermediates. Naturally, we do not discount the possibility that with adequate (though likely immense) computing power, quantum methods will eventually conquer such problems. Nonetheless, in our own work, we decided to pursue an alternative approach inspired by how human experts have been designing (or at least rationalizing) new reactions based on the analysis of familiar “arrow-pushing” steps and their sequences. As described in the main text, the crux of this approach are the constraints deriving from physical-organic chemistry, e.g., those specifying which individual steps can be “wired-up” into sequences that can proceed under the same class of conditions, or those that approximate the rate constants (deriving from Mayr indices and fine-tuned using linear free-energy relationships). The advantage computers offer is that they can implement this type of analysis over entire networks of mechanistic transforms. As we have seen in the main text, such networks for MCRs can be quite complex and, arguably, their generation and analysis by human chemists (by paper-and-pencil) would be an extremely painstaking enterprise (and virtually impossible for screening thousands upon thousands of networks derived from different sets of substrates of which only a small fraction gives rise to valid and interesting MCRs).

Of course, others have attempted different types of algorithms to analyze reaction mechanisms, in conjunction with either lower-end quantum calculations, *ab initio* molecular dynamics (AIMD), or with machine learning. Unfortunately, although these approaches are often very interesting in concept, they have so far been marred by rather serious chemical inconsistencies and/or limitations in extrapolating beyond mechanisms directly taught to the machine. In the following sub-sections, we discuss in some detail the outputs from three recent algorithms: ReactionPredictor from Pierre Baldi's laboratory (**Sections S3.1 and S3.2**, ref. <sup>19</sup>), AIMD from Todd Martinez' laboratory (**Section 3.3**, <sup>20</sup> and <sup>21</sup>) and Machine Learning from MIT's Connor Coley group (**Section 3.4**, <sup>22</sup>).

### **Section S.3.1. ReactionPredictor: Single-step and pathway predictors.**

Only few algorithms for reaction-outcome and mechanism prediction have evolved into a user-friendly tool available to the wide audience. A notable exception is Baldi's ReactionPredictor (available at <http://reactions.ics.uci.edu/>), which covers a broad spectrum of reaction types and is often considered a standard in the field. Upon entering reactants, the program calculates all possible electron sources and electron sinks of the input structures, and filters them to reduce the space of possible reactions. Afterwards, it proposes probable pairings between best-scoring electron sources and sinks, and ranks such predicted transformations using machine learning techniques<sup>19</sup>.

Below, we narrate the performance of the program on various types of reactions.

#### ***1) No support for reactions catalyzed by transition metals***

We first tested the single-step module with starting materials being the typical substrates of the Suzuki coupling. When inputting a metal catalyst as one of the substrates, the program gives the server error (**Supplementary Figure S17a**). On the other hand, when the palladium-source is omitted, the program predicts only highly implausible radical transformations (**Supplementary Figure S17b**). In the pathway (multi-step module) module, the same substrates for the Suzuki reaction were used, and the correct reaction product was also input (this is necessary to run the prediction). Despite having full information about the substrates and the correct product, the program failed to propose any mechanism for the given reaction (calculations resulted in server error).

a

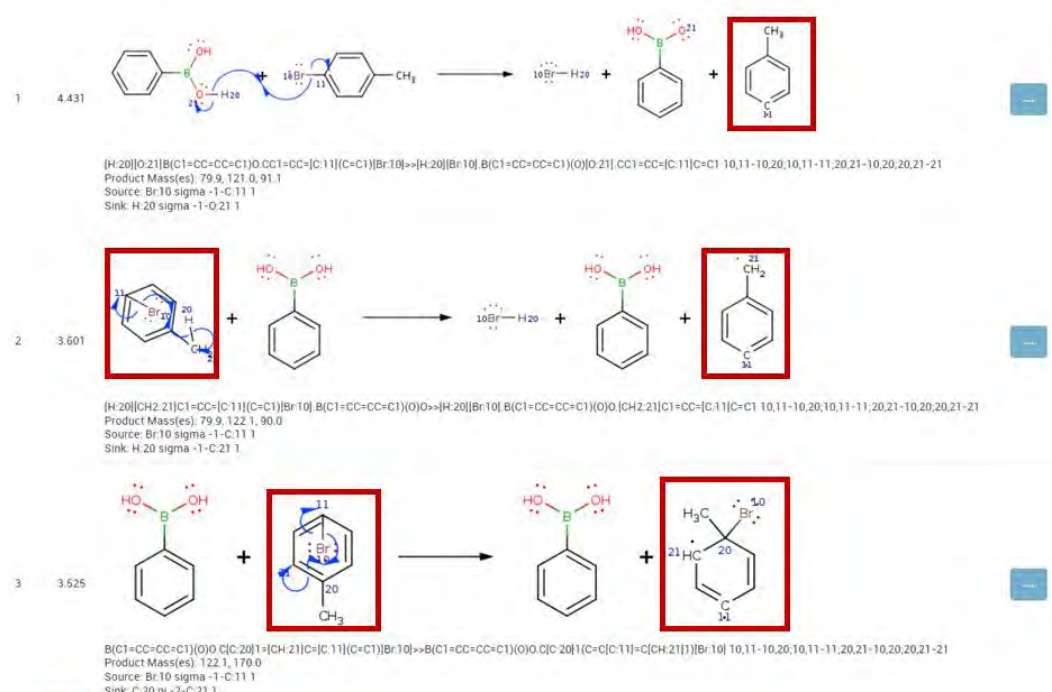
ReactionPredictor Webserver

Home

Server Error

The Server has encountered an unrecoverable error. Please contact the administrator (mohamadt AT uci DOT edu) if this persists.

b



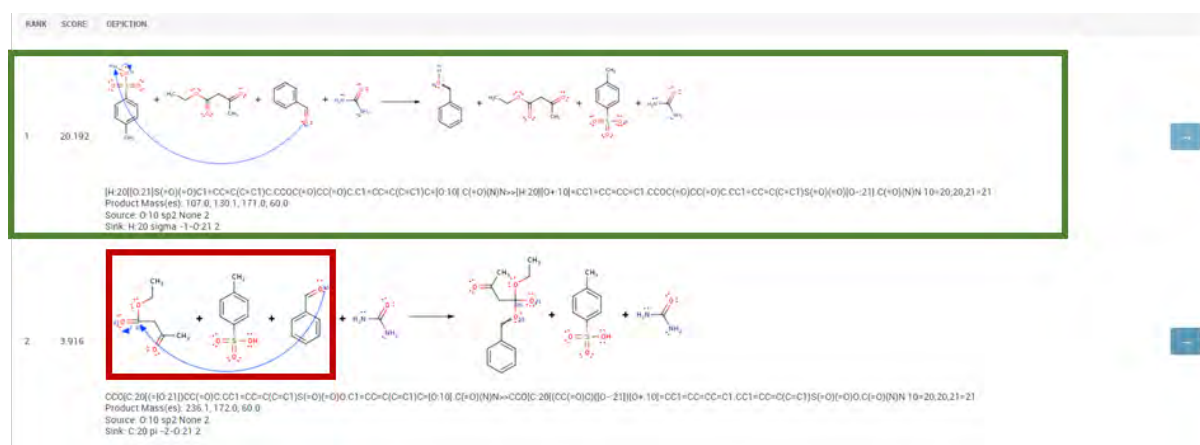
**Supplementary Figure S17. Example of an output generated by single-step ReactionPredictor software.** a, When inputting the transition metal catalyst as a substrate, the software gives ‘server error’. (search conditions: input molecule: Cc1ccc(Br)cc1.CC(=O)O[Pd]OC(C)=O.OB(O)c1ccccc1 or Cc1ccc(Br)cc1.[Pd].OB(O)c1ccccc1; reaction conditions: 298K, THF; chemistry type: predict; min. source: 1; min. sink: 1) b, Top-3 predictions generated by the software after inputting substrates for Suzuki reaction (omitting the Pd-catalyst). The software proposes only chemically implausible radical mechanistic steps. (search conditions: input molecule: Cc1ccc(Br)cc1.OB(O)c1ccccc1; reaction conditions: 298K, THF; chemistry type: predict; min. source: 1; min. sink: 1)

## 2) Problems with multicomponent reactions

Continuing with the single-step module, typical substrates of the Biginelli reaction were input as starting materials. The software is able to reconstruct the first four mechanistic steps (marked with green frames in **Supplementary Figure S18a-d**), although it also gave incorrect alternatives among its top predictions (red frames in **Supplementary Figure 4a,b**). Subsequently, however, it altogether failed to predict elimination of a water molecule (**Supplementary Figure S18e**) to form enamine, which is crucial for Biginelli reaction to proceed, as this enamine is attacked by the reaction's third component, the ketoester. Instead, the software proposed simple proton shifts (**Supplementary Figure S18e**, top-1 and top-2) as well as chemically implausible transformations (e.g., pTSA anion acting as a nucleophile in reactions top-3, top-4, top-7, pTSA anion acting as a base subtracting C-H protons in reactions top-4, top-6, top-9, reactions generating highly-strained molecules, like 4-membered cyclic urea-derivative in top-11, or reaction generating improbable leaving groups, like pTSA-derived  $\text{OH}^-$  in reactions top-12 and top-13 and ammonia in reaction top-15). These problematic outcomes are highlighted by red frames.

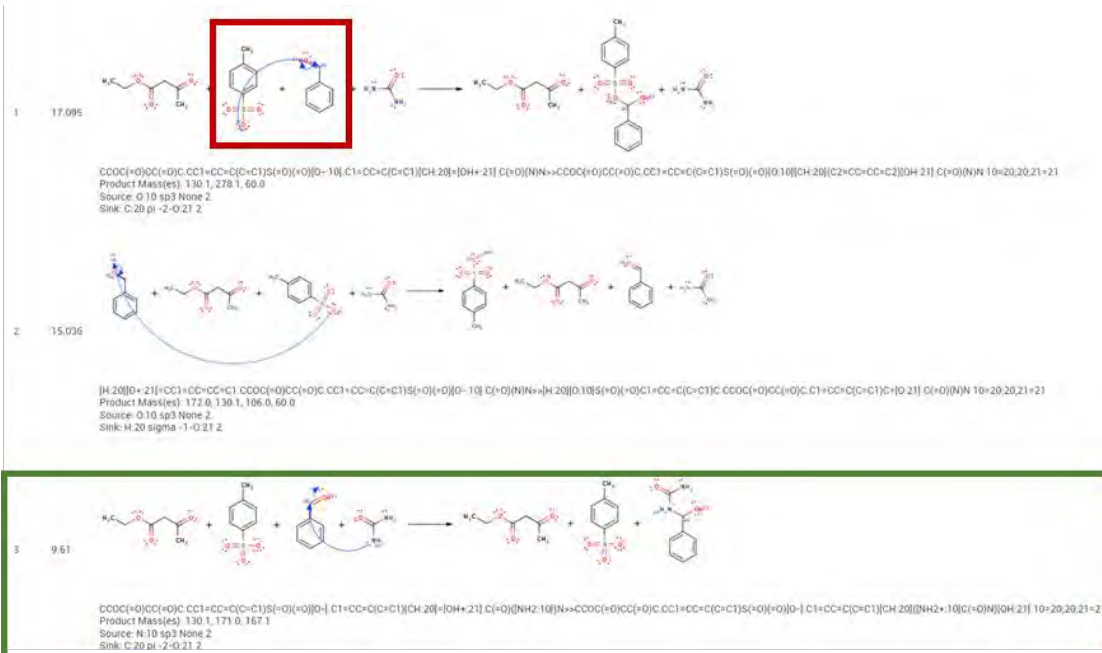
In the pathway (multi-step) module, the same substrates for the Biginelli reaction were used, and the correct reaction product was input (this is necessary to run the mechanism prediction). Despite having full information about the substrates and the correct product, the program failed to propose any mechanism (calculations resulted in server error).

**a**

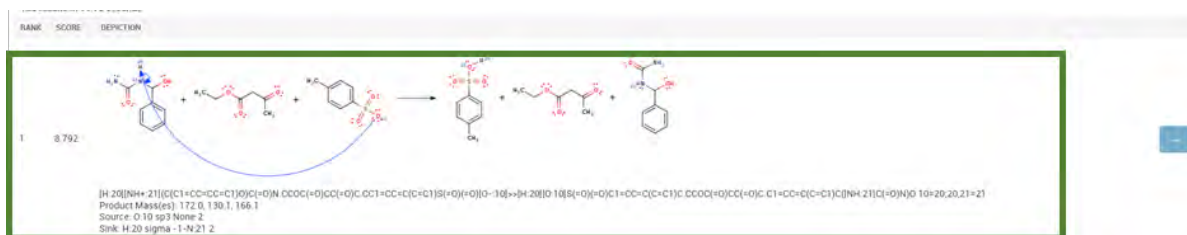


**b**

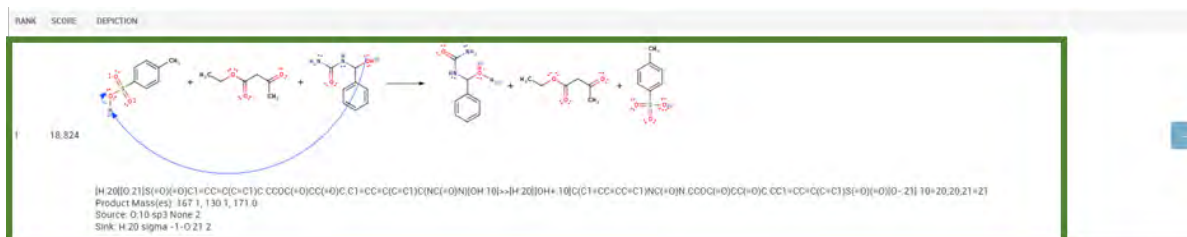




c



d

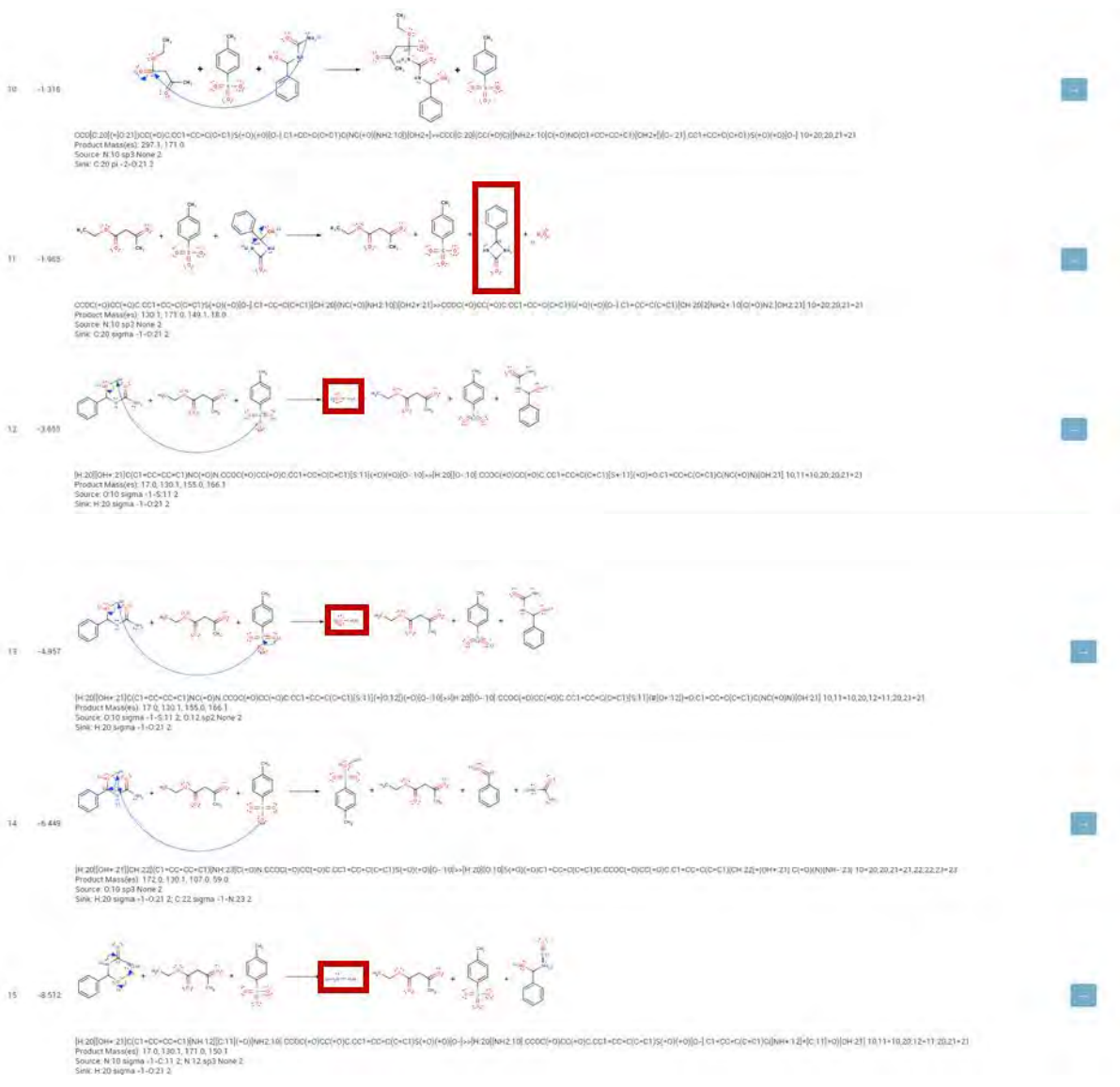


e

144 results in 13.71 seconds

RANK SCORE DEFINITION

1	18.661		<p>[H 20][OH 21][C1-CO-CC-C1]N(C)O(N)CCOC(=O)C(C)C(C)C(C)S(O)O(=O)1[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 172.0, 130.1, 166.1 Source: O 10 sp3 None 2 Sink: H 20 sigma -1-O 21 2</p>
2	7.93		<p>[H 20][OH 21][C1-CO-CC-C1]N(C)O(N)H 21[6]CCOC(=O)C(C)C(C)C(C)S(O)O(=O)1[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 167.0, 130.1, 171.0 Source: N 10 sp3 None 2 Sink: H 20 sigma -1-O 21 2</p>
3	5.146		<p>CCOC(=O)C(C)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 301.1, 167.1 Source: O 10 sp3 None 2 Sink: C 20 pi -2-O 21 3</p>
4	2.589		<p>[H 20][OH 21][C1-CO-CC-C1]N(C)O(N)H 21[6]CCOC(=O)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21,22,22,23*23 Product Mass(es): 301.1, 167.1 Source: O 10 sp3 None 2 Sink: H 20 sigma -1-O 21 2, C 22 pi -3-O 23 2</p>
5	2.459		<p>CCOC(=O)C(C)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)H 21[6]10-20,20,21*21 Product Mass(es): 301.1, 167.1 Source: O 10 sp3 None 2 Sink: C 20 pi -3-O 23 2</p>
6	2.39		<p>[H 20][OH 21][C1-CO-CC-C1]N(C)O(N)H 21[6]CCOC(=O)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21,22,22,23*23 Product Mass(es): 172.0, 129.1, 167.1 Source: O 10 sp3 None 2 Sink: H 20 sigma -1-O 21 2, C 22 pi -2-O 23 2</p>
7	2.274		<p>CCOC(=O)C(C)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 130.1, 129.1, 166.1 Source: O 10 sp3 None 2 Sink: C 20 sigma -1-O 21 2</p>
8	0.855		<p>CCOC(=O)C(C)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 297.1, 171.0 Source: N 10 sp3 None 2 Sink: C 20 pi -2-O 21 2</p>
9	0.491		<p>[H 20][OH 21][C1-CO-CC-C1]N(C)O(N)H 21[6]CCOC(=O)C(C)C(C)C(C)S(O)O(=O)1[6]C1-CC-C(C)C)N(C)O(N)H 21[6]H 20[O 18]S(O)O(C1-CC-C(C)C)C(C)C(C)O(C)C1-CC-O(C)C(N)O(N)[OH 21]10-20,20,21*21 Product Mass(es): 172.0, 129.1, 167.1 Source: O 10 sp3 None 2 Sink: H 20 sigma -1-O 21 2</p>



**Supplementary Figure S18. Example of an output generated by single-step ReactionPredictor software a**, Input molecules (substrates for Biginelli reaction): NC(N)=O. [H]C(=O)c1ccccc1.CCOC(=O)CC(C)=O. CC1=CC=C(C=C1)S(=O)(=O)O; reaction conditions: 343K, water; chemistry type: polar; min. source: 1; min. sink: 1 **b**, Input molecules: NC(N)=O. CCOC(=O)CC(C)=O. [H][O+]=C([H])c1ccccc1. CC1=CC=C(C=C1)S(=O)(=O)[O-]; reaction conditions: 343K, water; chemistry type: polar; min. source: 1; min. sink: 1 **c**, Input molecules: CCOC(=O)CC(=O)C. C1=CC=C(C=C1)C([NH2+])C(=O)N. CC1=CC=C(C=C1)S(=O)(=O)[O-]; reaction conditions: 343K, water; chemistry type: polar; min. source: 1; min. sink: 1 **d**, Input molecules: OS(=O)(=O)C1=CC=C(C=C1)C. CCOC(=O)CC(=O)C. C1=CC=C(C=C1)C(NC(=O)N)O; reaction conditions: 343K, water; chemistry type: polar; min. source: 1; min. sink: 1 **e**. **e**. Input molecules: [OH2+]C(C1=CC=CC=C1)NC(=O)N. CCOC(=O)CC(=O)C.

CC1=CC=C(C=C1)S(=O)(=O)[O-]; reaction conditions: 343K, water; chemistry type: polar; min. source: 1; min. sink: 1 .

### 3) *Problems with reactive groups and reaction core recognition*

In the first two examples in **Supplementary Figure S19**, the input to the program were typical substrates of Curtius rearrangement. In the single-step reaction mode, the software was able to reconstruct the first mechanistic step of the rearrangement (ranked as 6th, marked with green frame in **Supplementary Figure S19a**), but it failed to predict the next step, formation of the carbon-nitrogen bond (**Supplementary Figure S19b**) to form isocyanate, which is crucial for Curtius rearrangement to proceed. Instead, the software proposed reactions in which terminal nitrogen of acyl azide acts as electrophile attacked by a double bond of the aromatic ring system (**Supplementary Figure S19b**, top-1) as well as other chemically implausible transformations (e.g., molecule fragmentation top-3, top-6). None of the other predicted outcomes matched the correct step of the Curtius mechanism.

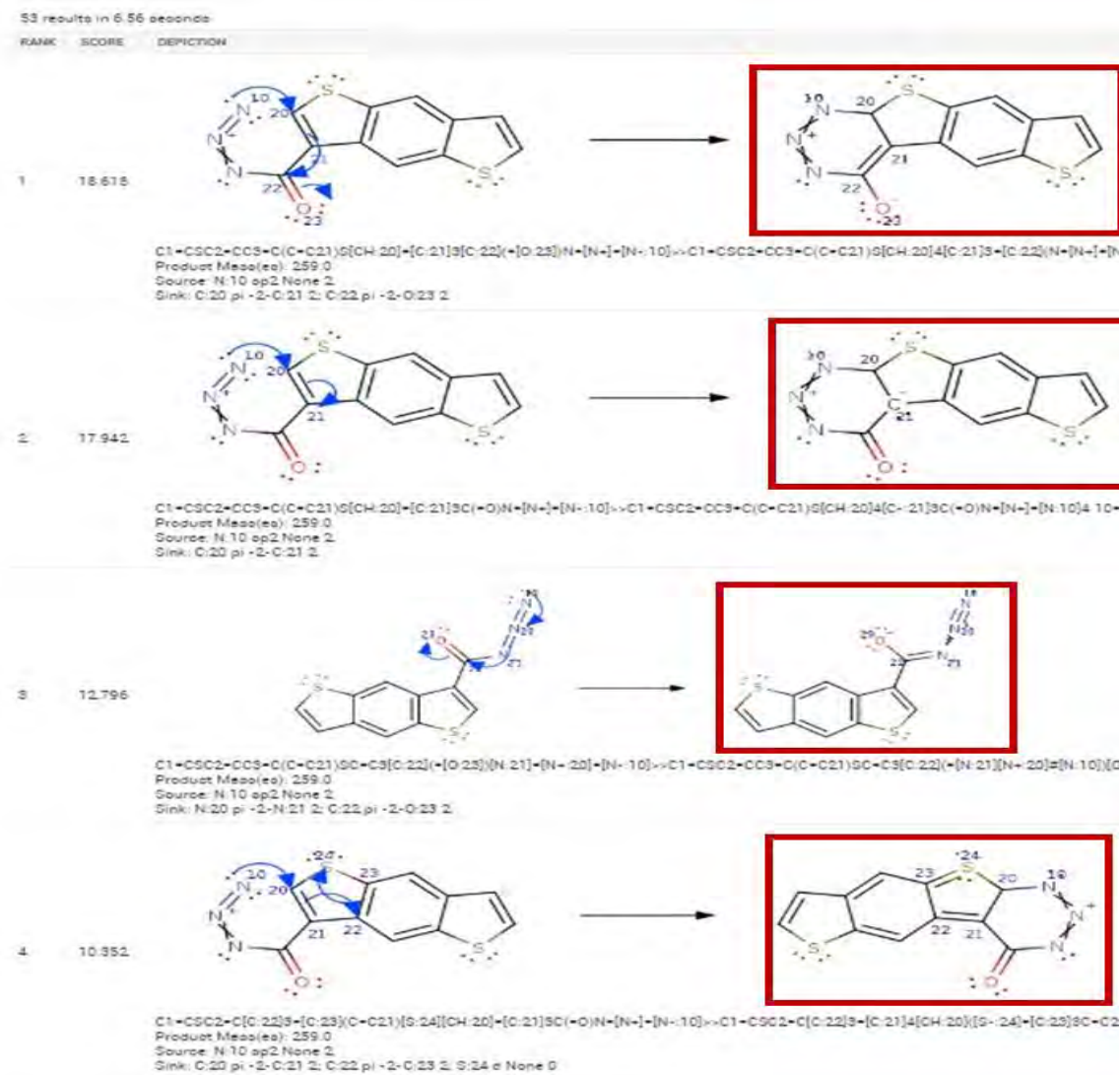
In the pathway (multi-step) module, the same substrates of the Curtius reaction were used, and the correct reaction product was input (again, this is necessary to run the prediction). Despite having full information about the substrates and the correct product, the program failed to propose any mechanism (calculations resulted in server error)

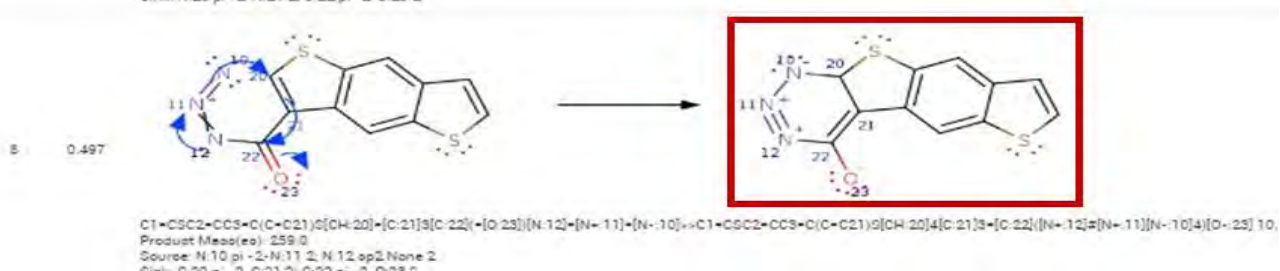
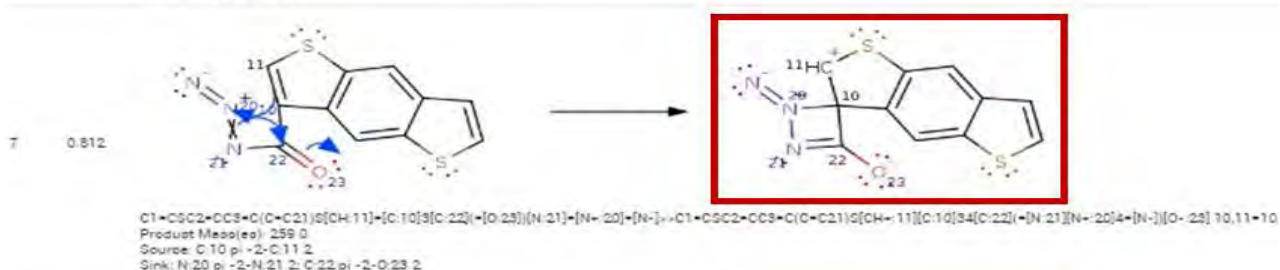
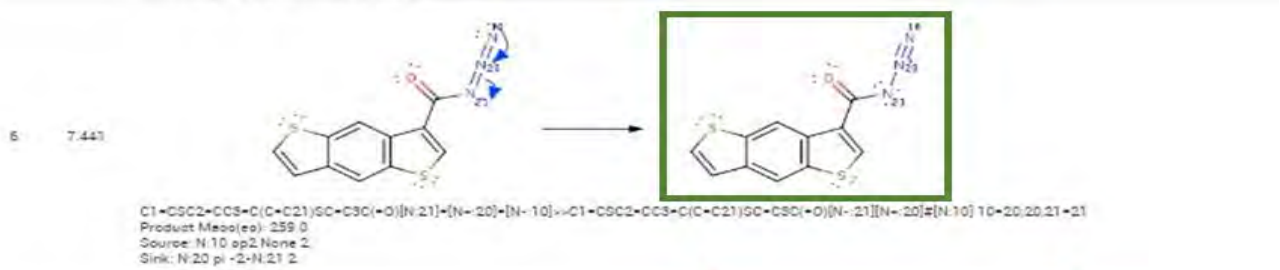
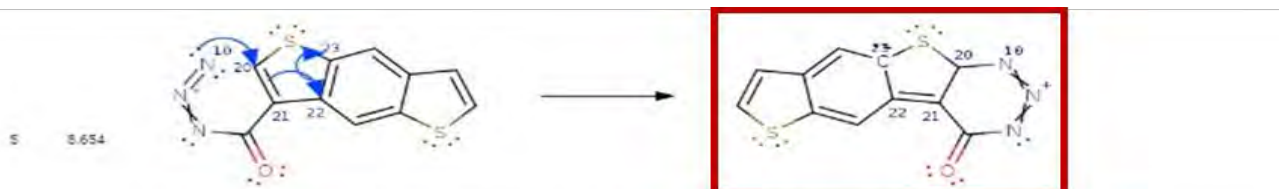
In the remaining examples (**Supplementary Figures S19 – S24**), we chose typical substrates for Chugaev, Hofmann, Cope and Ramberg–Bäcklund eliminations as well as Newman–Kwart rearrangement. In none of these examples was the program able to predict the correct mechanism. Instead, the software proposed many chemically-incorrect reactions, often ignoring important, reactive groups present in the molecules and, instead, performing “reactions” within much more inert aromatic scaffolds (e.g., all top predictions in Hofmann elimination, **Supplementary Figure S21** and Cope elimination, **Supplementary Figure S22**, top-1 to top-3 predictions in Ramberg–Bäcklund elimination, **Supplementary Figure S23**). The algorithm also proposed some nonsensical radical fragmentations (top predictions in Chugaev elimination, **Supplementary Figure S20**, and Newman–Kwart rearrangement, **Supplementary Figure S24**).

In the pathway (multi-step module) module, the same substrates for the above reactions were used, and the correct reaction product was input (this is necessary to run the prediction). Akin to previous examples, calculations resulted in server error for Curtius rearrangement, Chugaev

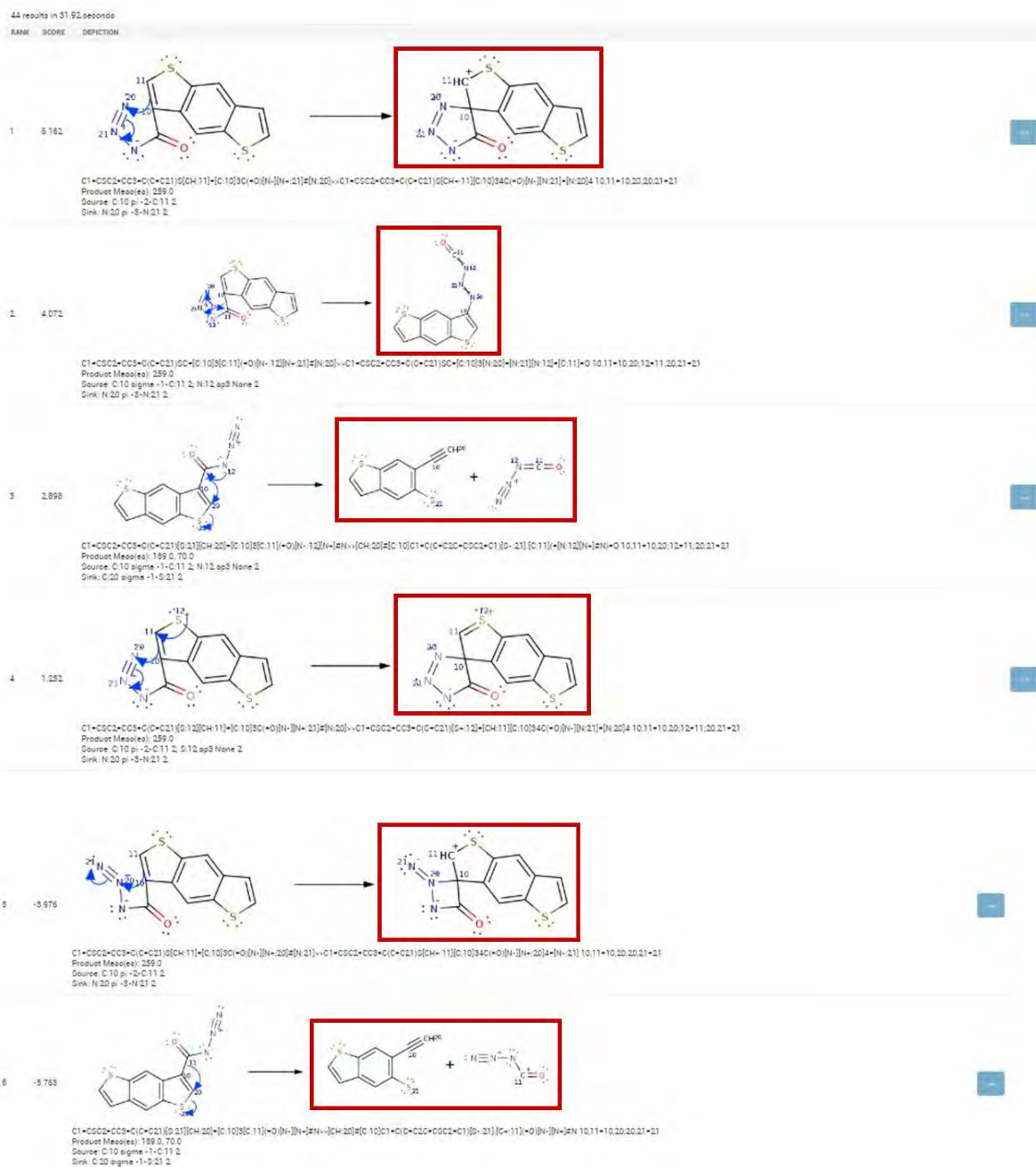
elimination and Newman–Kwart rearrangement. No error but also no results were output for Hofmann elimination, Cope elimination and Ramberg–Bäcklund elimination.

a.





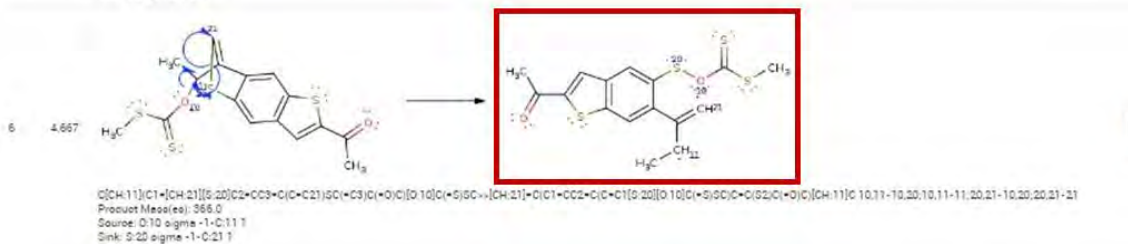
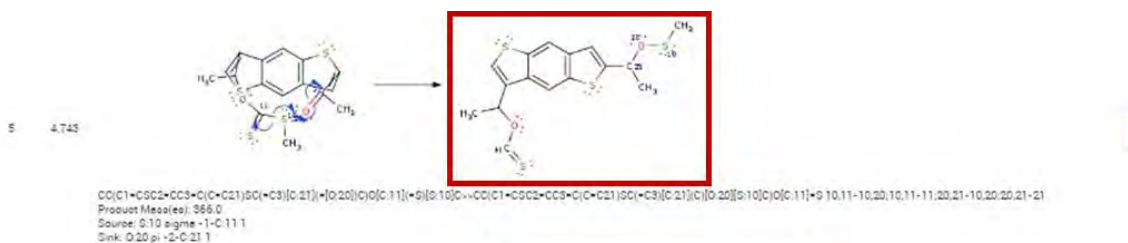
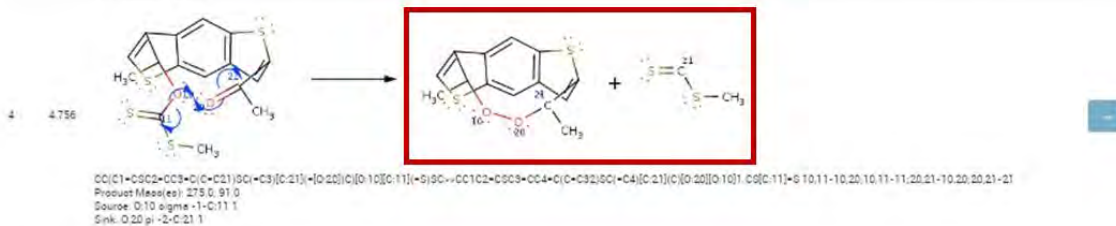
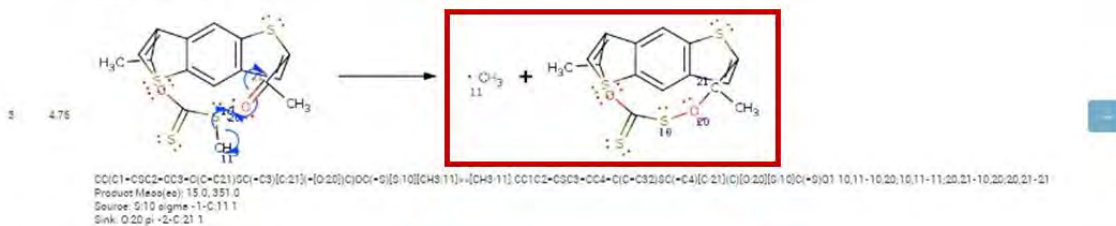
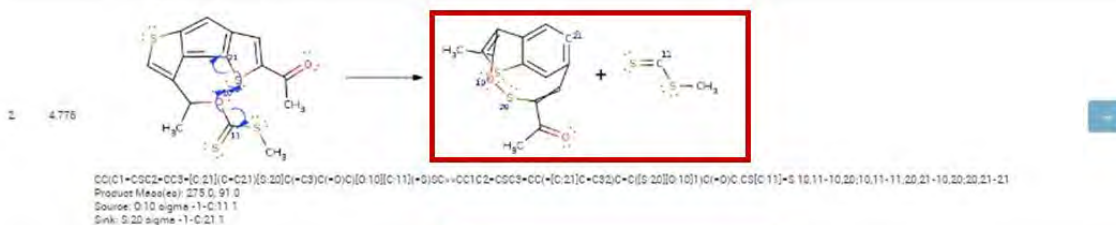
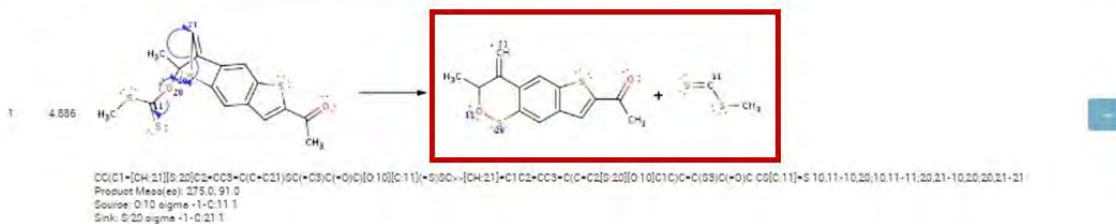
**b.**



**Supplementary Figure S19. Output generated by single-step ReactionPredictor software using substrate prone to Curtius rearrangement. a,** Input molecule: [N-]=[N+]=NC(=O)c1csc2cc3ccsc3cc12; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0; **b,** Input molecule: O=C([N-][N+]#N)c1csc2cc3ccsc3cc12; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.

86 results in 13.63 seconds

RANK SCORE DEPICTION

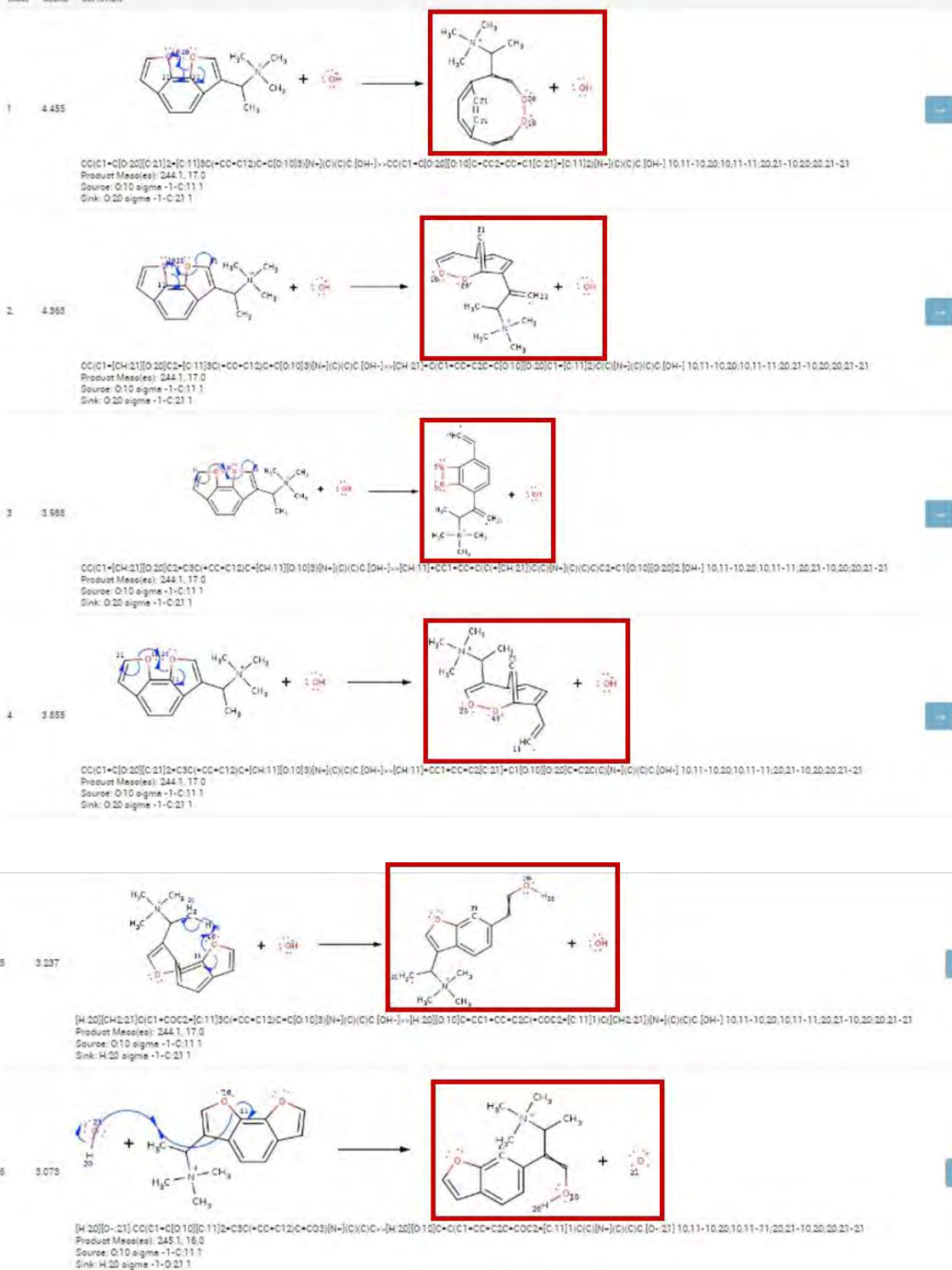


**Supplementary Figure S20. Output generated by single-step ReactionPredictor software using substrate prone to Chugaev elimination.** Input molecule: CSC(=S)OC(C)c1csc2cc3cc(sc3cc12)C(C)=O; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.



26 results in 10.45 seconds

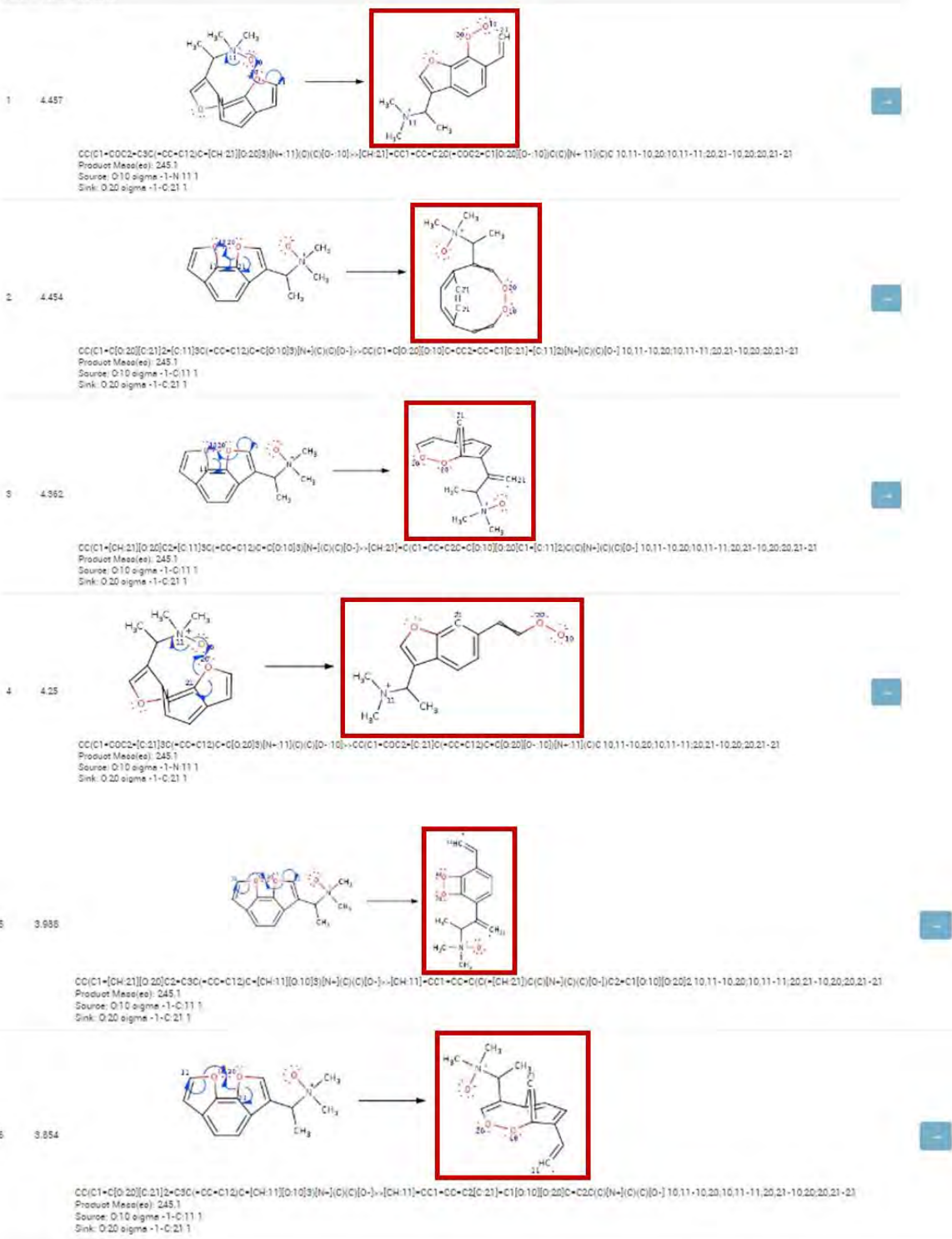
RANK SCORE DEPICTION



**Supplementary Figure S21. Output generated by single-step ReactionPredictor software using substrate prone to Hofmann elimination.** Input molecules: CC(c1coc2c1ccc1ccoc21)[N+](C)(C)C.[OH-]; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.

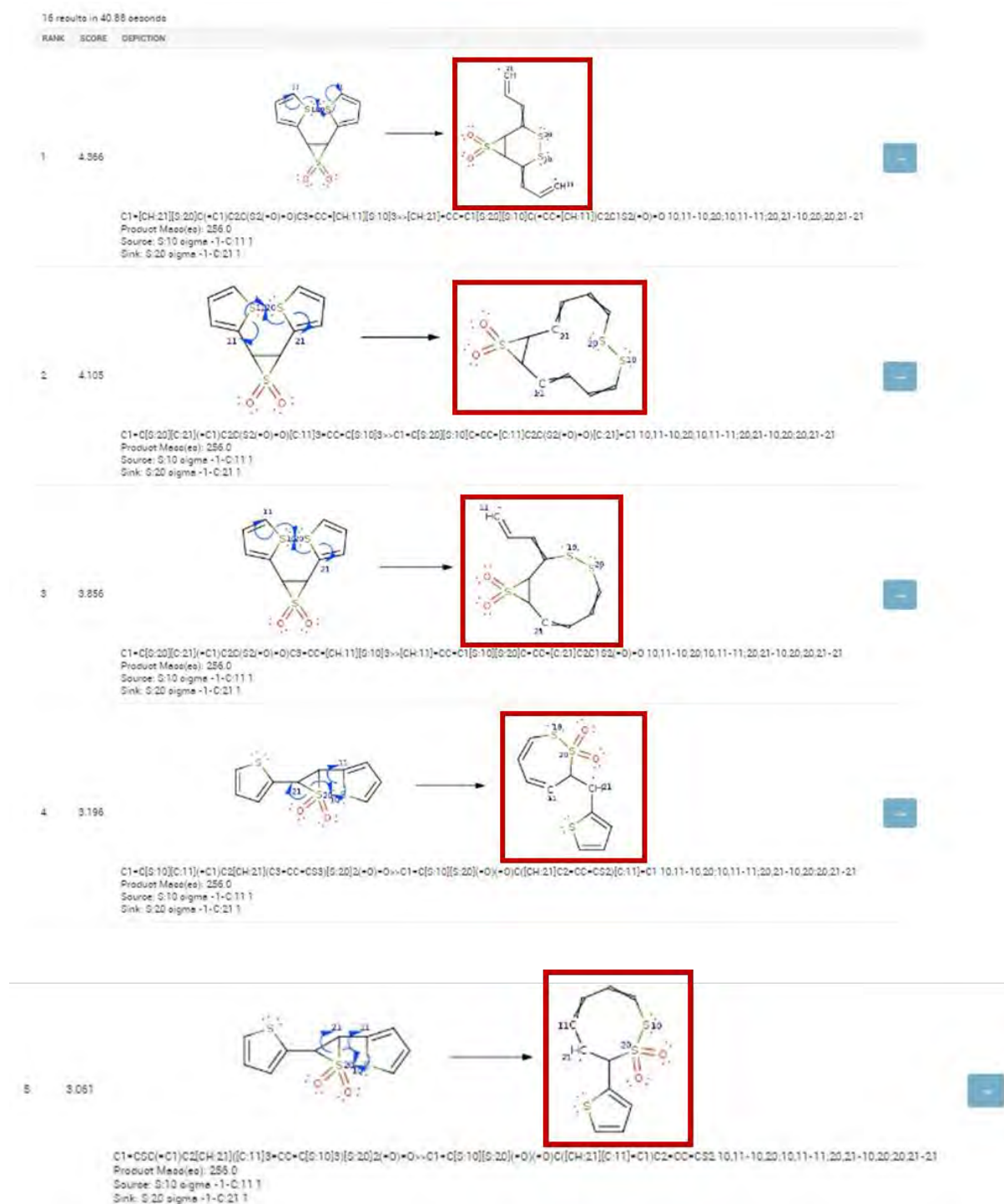
44 results in 25.66 secondi

RANK SCORE DEPICTION



**Supplementary Figure S22. Output generated by single-step ReactionPredictor software using substrate prone to Cope elimination. Input molecule:**

CC(c1coc2c1ccc1ccoc21)[N+](C)(C)[O-]; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.



**Supplementary Figure S23. Output generated by single-step ReactionPredictor software using substrate prone to Ramberg–Bäcklund elimination.** Input molecule: O=S1(=O)C(C1c1cccs1)c1cccs1; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.



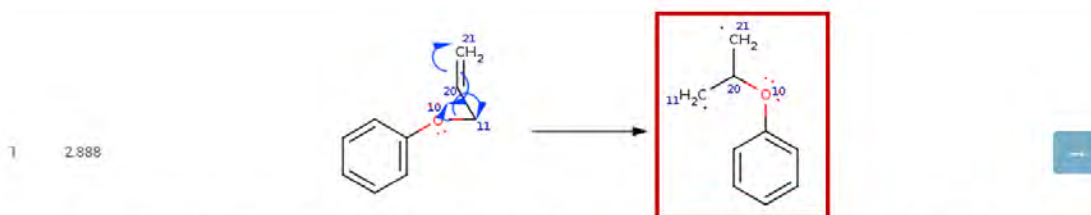
**Supplementary Figure S24. Output generated by single-step ReactionPredictor software using substrate prone to Newman-Kwart rearrangement.** Input molecule: CN(C)C(=S)Oc1ccc2ccccc2c1-c1c(OC(=S)N(C)C)ccc2ccccc12; reaction conditions: 343K, DMSO; chemistry type: predict; min source: 0; min sink: 0.

#### **4) *Problems with prediction of pericyclic reactions***

In the next example, the single-step module was used with typical substrates for the Claisen rearrangement. No correct mechanistic step was proposed – instead, only a series of radical degradations were proposed (**Supplementary Figure S25a**). The correct rearrangement step was proposed only when selecting “Chemistry type: pericyclic” at the beginning of the calculations (**Supplementary Figure S25b**). Still, the program was not able to predict full Claisen rearrangement, as it fails to predict the second step of the process, a simple rearomatization of the benzene ring (**Supplementary Figure S25c**).

In the pathway (multi-step module) module, the same substrates were used along with the correct reaction product. Still, the program failed to propose any mechanism (calculations ended in zero results, **Supplementary Figure S25d**).

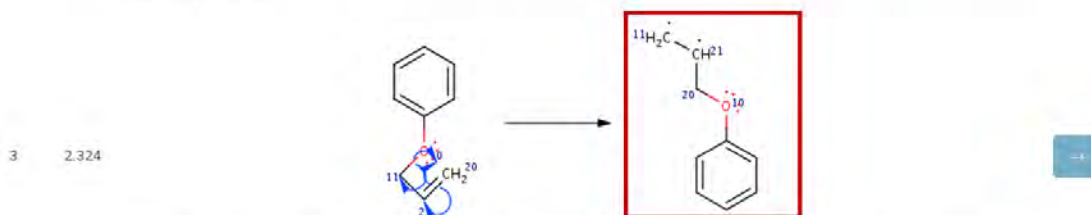
a.



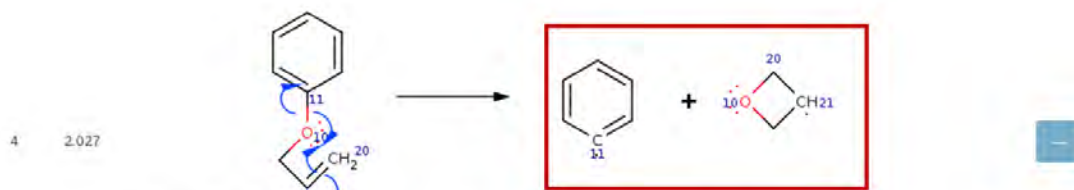
[CH2:21]=[CH:20][CH2:11][O:10]C1=CC=CC=C1>>[CH2:21][CH:20][CH2:11][O:10]C1=CC=CC=C1 10,11-10,20,10,11-11,20,21-10,20,20,21-21  
 Product Mass(es): 134.1  
 Source: O:10 sigma -1-C:11 1  
 Sink: C:20 pi -2-C:21 1



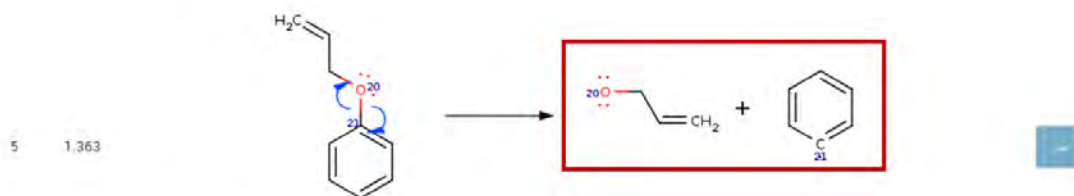
[CH2:21]=[CH:20]C[O:10][C:11]1=CC=CC=C1>>[CH2:21][CH:20]1C[O:10]1.C1=CC=C[C:11]C=C1 10,11-10,20,10,11-11,20,21-10,20,20,21-21  
 Product Mass(es): 57.0, 77.0  
 Source: O:10 sigma -1-C:11 1  
 Sink: C:20 pi -2-C:21 1



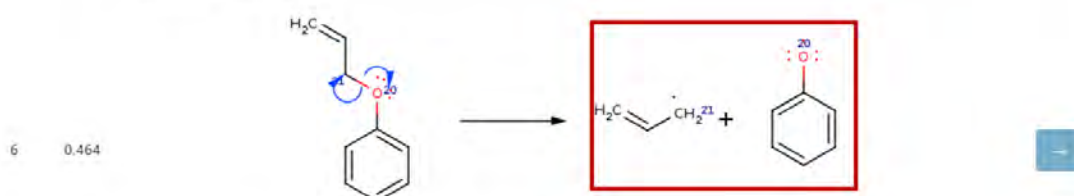
[CH2:20]=[CH:21][CH2:11][O:10]C1=CC=CC=C1>>[CH2:11][CH:21][CH2:20][O:10]C1=CC=CC=C1 10,11-10,20,10,11-11,20,21-10,20,20,21-21  
 Product Mass(es): 134.1  
 Source: O:10 sigma -1-C:11 1  
 Sink: C:20 pi -2-C:21 1



[CH2:20]=[CH:21]C[O:10][C:11]1=CC=CC=C1>>C1=CC=C[C:11]C=C1.[CH2:20]1[CH:21]C[O:10]1 10,11-10,20;10,11-11;20,21-10,20;20,21-21  
 Product Mass(es): 77.0, 57.0  
 Source: O:10 sigma -1-C:11 1  
 Sink: C:20 pi -2-C:21 1

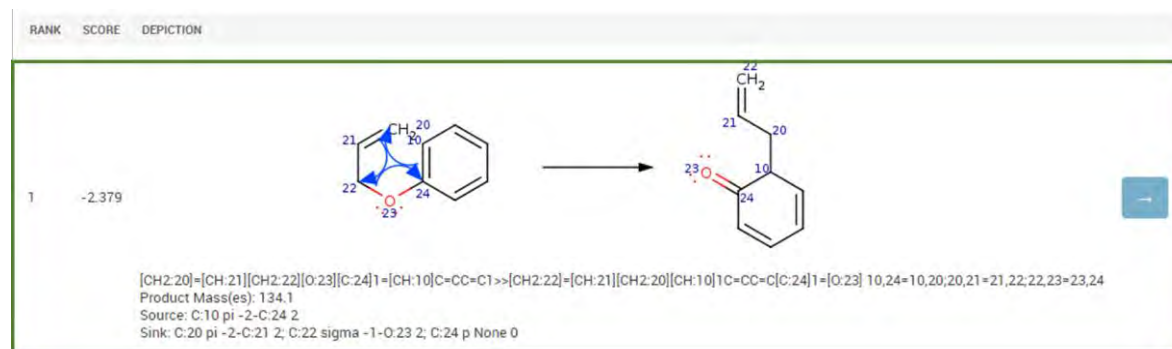


C=CC[O:20][C:21]1=CC=CC=C1>>C=CC[O:20].C1=CC=C[C:21]C=C1 20,21-20,20,21-21  
 Product Mass(es): 57.0, 77.0  
 Source: O:20 sigma -1-C:21 1  
 Sink: O:20 sigma -1-C:21 1

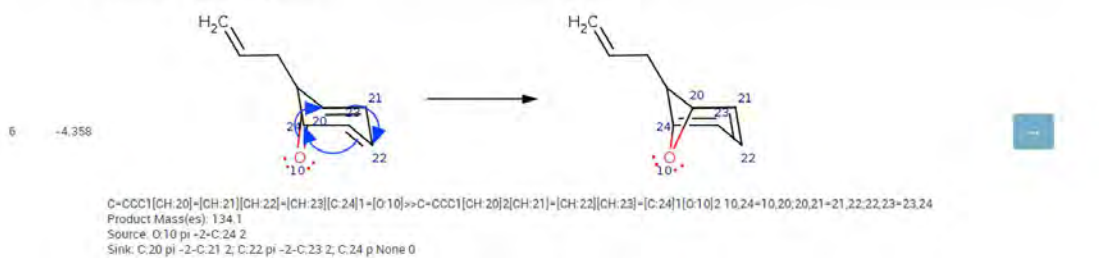
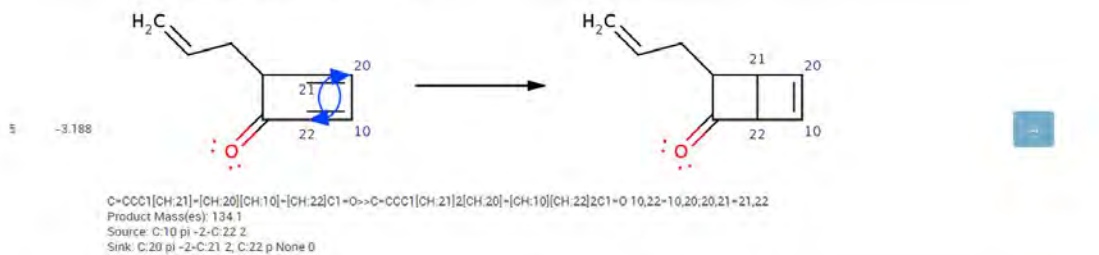
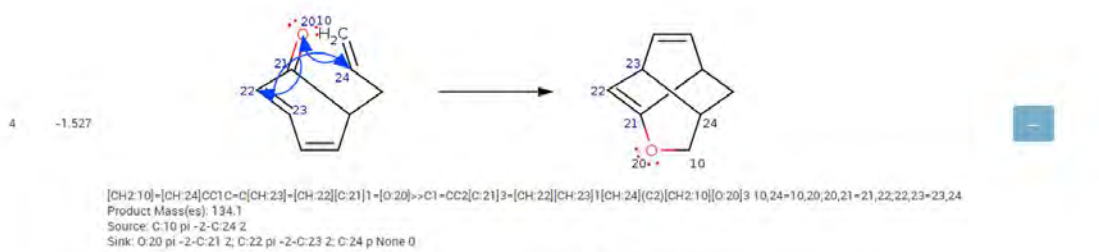
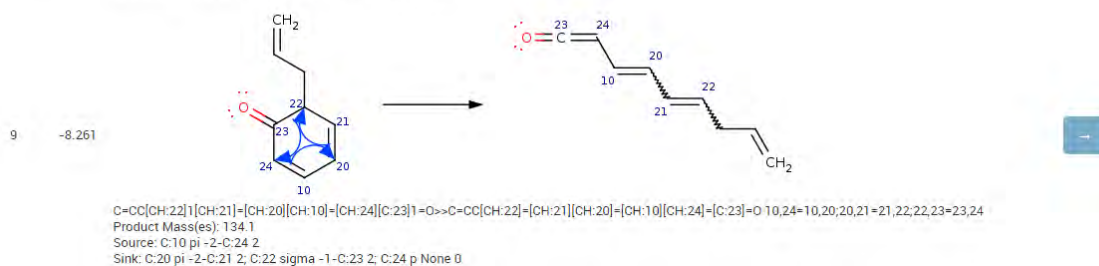
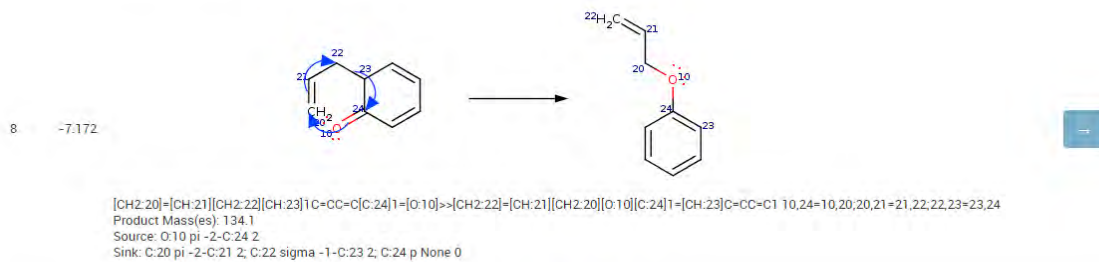
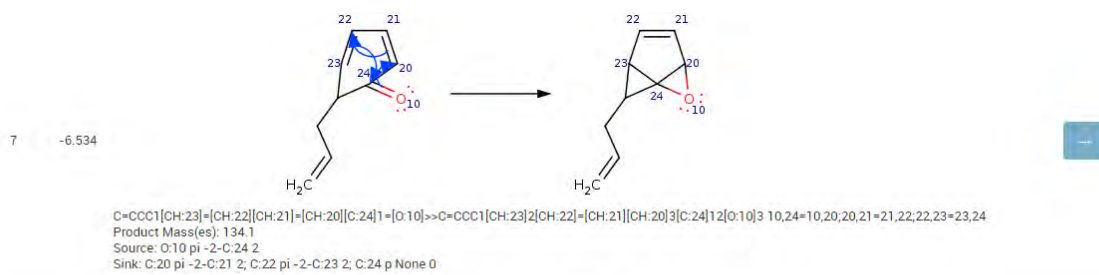


C=C[CH2:21][O:20]C1=CC=CC=C1>>C=C[CH2:21].C1=CC=C[C:21][O:20] 20,21-20,20,21-21  
 Product Mass(es): 41.0, 93.0  
 Source: O:20 sigma -1-C:21 1  
 Sink: O:20 sigma -1-C:21 1

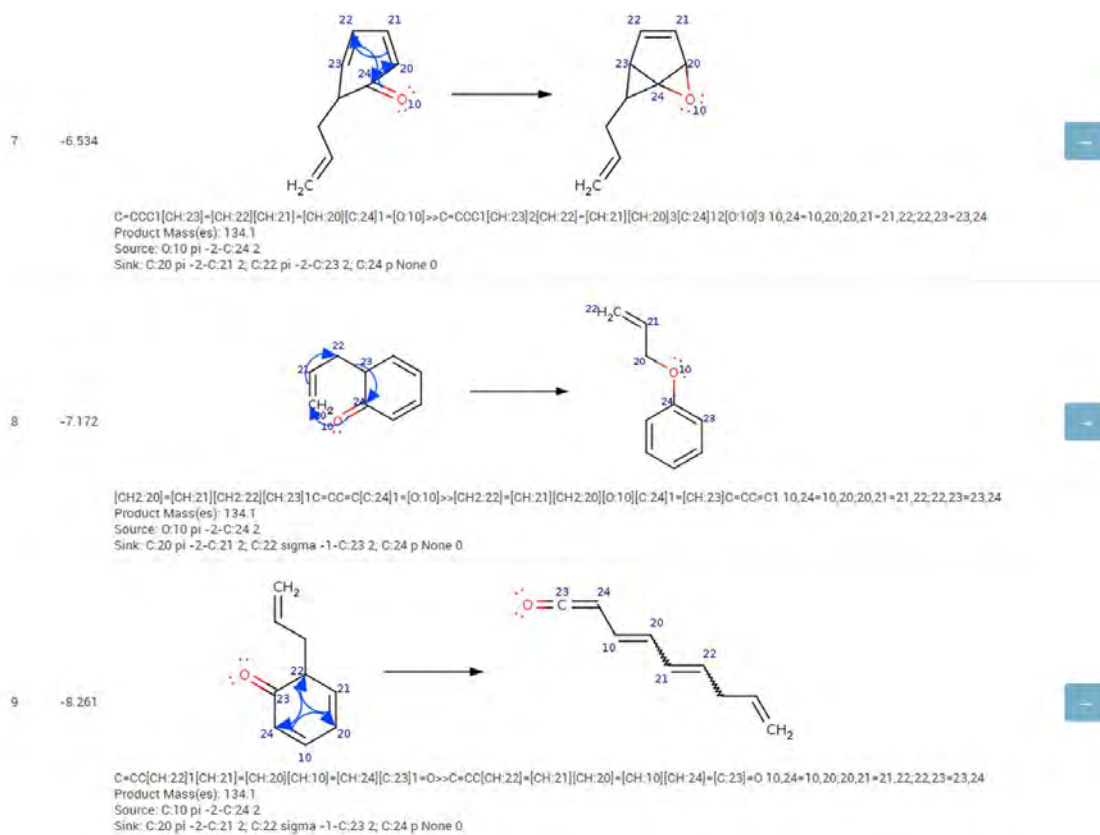
**b.**



C.







d.

### Predicted Pathway

No pathway was found. 0 products checked.

**Supplementary Figure S25. Output generated by ReactionPredictor software using substrate prone to Claisen rearrangement a**, Input molecules: C=CCOc1cccc1; reaction conditions: 343K, DMSO; **chemistry type: predict**; min. source: 1; min. sink: 1 **b**, Input molecules: C=CCOc1cccc1; reaction conditions: 343K, DMSO; **chemistry type: pericyclic**; min. source: 1; min. sink: 1 **c**, Input molecules: Oc1cccc1CC=C; reaction conditions: 343K, DMSO; **chemistry type: predict**; min. source: 1; min. sink: 1; **d**, Output generated by **Multi-Step ReactionPredictor**; input substrate: C=CCOc1cccc1; input product: Oc1cccc1CC=C; input Chemistry Type: pericyclic; other parameters were set to default.

### **Section S.3.2. RMechDB**

Another and more recent development from Baldi's and Van Vranken's laboratories has focused on mechanistic steps of radical reactions. As described in ref. <sup>23</sup> some 5,300 steps were "pedagogically chosen" from published transformations (from various textbooks and primary research articles). This compendium is accompanied by a web application available at <https://deeprxn.ics.uci.edu/rmechdb>.

The examples in **Supplementary Figures S26** and **S27** below demonstrate that an algorithm cannot generalize mechanistic steps beyond those encoded for specific literature precedents. The ability to generalize is an essential part of new reactions' discovery on which we focus in our present work (albeit not using radical-based steps whose generalization is, admittedly, challenging in its own right and will be part of a separate study).

a

RMechDB Search

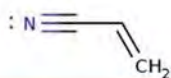
Search by reaction: Search by compound

Choose the compound type  
Molecule

Enter compound SMILES/SMARTS\*

All  Choose a category of radical reactions.  
All  Data sets to be included in the search.  
100  Maximum number of reactions displays.

Search query:



Reaction ID	Initial conditions	Product name	Engineer & UMSIC
123	60.00		
124	121.01		
125	174.1		
126	227.10		
127	142.02		
128	180.21		
129	239.04		
130	188.14		

Reaction 123: C=C[N:] + C=C(C)C >> C=C(C)C[N:]

Reaction 124: C=C[N:] + C=C(C)C(C)C >> C=C(C)C(C)C[N:]

Reaction 125: C=C[N:] + C=C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C[N:]

Reaction 126: C=C[N:] + C=C(C)C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C(C)C[N:]

Reaction 127: C=C[N:] + C=C(C)C(C)C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C(C)C(C)C[N:]

Reaction 128: C=C[N:] + C=C(C)C(C)C(C)C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C(C)C(C)C(C)C[N:]

Reaction 129: C=C[N:] + C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C[N:]

Reaction 130: C=C[N:] + C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C >> C=C(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C[N:]



**Supplementary Figure S26. Examples of outputs generated by Compound search within RMechDB software** (online platform aggregating precedent-based mechanistic steps of radical reactions) accessed at <https://deeprxn.ics.uci.edu/rmechdb> [4.04.2023]. In the Compound search variant, the user inputs the SMILES of the desired molecule and the program displays those elementary steps in the database that contain the given molecule either as a reactant or as a product of the elementary step. **a**, A screenshot of search results obtained by inputting SMILES of a simple acrylonitrile molecule and running search on default settings with the maximum possible number of reactions displayed. The program proposed 8 different radical mechanistic steps that match the input molecule. **b**, A screenshot of search results obtained by inputting a molecule similar to acrylonitrile from (a). A slight change in the molecule's structure

(adding an extra methyl group, which in this case is a non-reactive substituent outside the reaction core) gives no results. c, A screenshot of search results obtained by inputting an acrylonitrile molecule with an additional phenyl group. Also in this case the search gave no results. This example clearly shows that the usage of database with literature reactions encoded verbatim limits the applicability of the software in discovery-oriented tasks (i.e., beyond specific, already published precedents).

a

RMechDB Search

Search by reaction: [Search by compound](#)

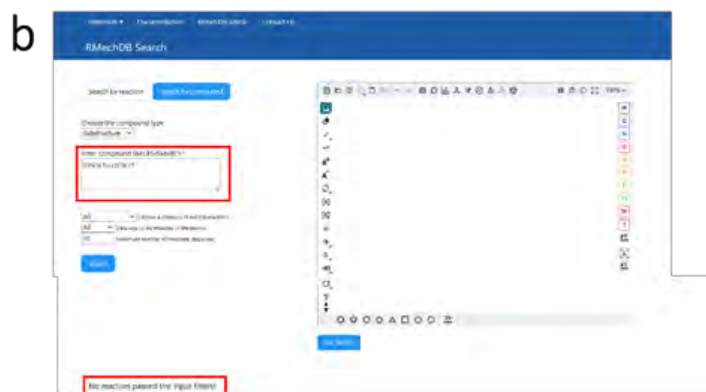
Choose the compound type:  
 Substructure:

Enter compound SMILES/SMART

All  Choose a category of initial reaction:  
 All  (Data sets to be displayed in the search)  
 No  (Show only reaction of the reaction type)

5 elementary steps match your query:

Reaction ID	Initial compound	Product name	Reaction & SMILES
101	101-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
102	102-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
103	103-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
104	104-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
105	105-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
106	106-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
107	107-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
108	108-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
109	109-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>
110	110-1		 <chem>O=C1OC(=O)C1 + ... &gt;&gt; O=C1OC(=O)C1 + ...</chem>



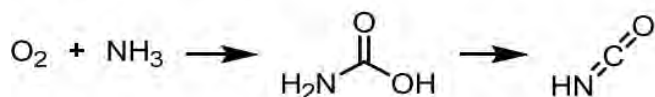
**Supplementary Figure S27. Examples of outputs generated by Compound search within RMechDB software** (online platform aggregating precedent-based mechanistic steps of radical reactions) accessed at <https://deeprxn.ics.uci.edu/rmechdb> [4.04.2023]. **a**, A screenshot of search results obtained by inputting smiles of 4-(OMe)-benzyl radical and running search on default settings with the maximum possible number of reactions displayed. The program proposed 8 different radical mechanistic steps that match the input molecule. **b**, A screenshot of search results obtained by inputting a molecule similar to molecule from example (a). A slight change in the molecule's structure (changing -OMe substituent to -Cl) gives no results. As in **Supplementary Figure S26**, this example indicates that the usage of database with literature reactions encoded verbatim limits the applicability of the software in discovery-oriented tasks (i.e., beyond specific, already published precedents).

### **Section S3.3. *Ab initio* molecular dynamics, AIMD, simulations.**

We now switch gears to methods that aim to reconstruct mechanistic reaction networks based on *ab initio* molecular dynamics, AIMD. These methods were originally developed by Todd Martinez' lab at Stanford. One of the pillars of this approach is to use a “virtual piston” to periodically push molecules towards the center of the “nanoreactor,” which increases the frequency of collisions and barrier crossing to enhance reactivity. As described in detail in <sup>20</sup> an approximate Hartree–Fock (HF) ansatz enables large simulation sizes and long timescales, and putative reaction pathways found are fine-tuned by DFT calculations. Ultimately, the AIMD simulations produce pathways that connect stable reactant(s) and product(s) along minimum-energy paths, MEPs, which give rise to mechanistic networks of a reaction under study. These methods were explored in the context of prebiotic chemistry. As described in our own work on prebiotic chemistry<sup>24</sup>, we evaluated the performance of AIMD as implemented by Das *et al.*<sup>21</sup>. The basic reactions shown below – and reproduced from the Supplementary

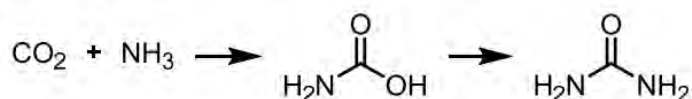
Information of the said ref.<sup>24</sup> – were all predicted by AIMD to proceed under plausible prebiotic conditions, in water (pH 8-10), in the temperature range 80-100°C, and without any metal-based catalysts or high-energy protons. Unfortunately, such outcomes are highly improbable.

Example 1:



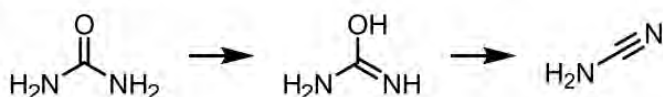
Reaction yielding cyanic acid from carbon dioxide and ammonia requires the usage of either a very strong base (e.g., NaHMDS, ref.<sup>25</sup>) that is incompatible with water or electric discharge (see<sup>26</sup>) in a gas phase. In the conditions described by Das *et al.*, carbon dioxide and ammonia would rather form ammonium bicarbonate and ammonium carbonate<sup>27</sup>.

Example 2:



Reaction leading to urea from carbon dioxide and ammonia are much harsher than claimed by the authors, for example: 182 °C and ~15,5 MPaG<sup>28</sup> or electric discharge on the titanium electrode<sup>29</sup>.

Example 3:



Synthesis of cyanamide from urea requires very high temperatures<sup>30</sup> and anhydrous conditions (preferably in the presence of a dehydrating agent).

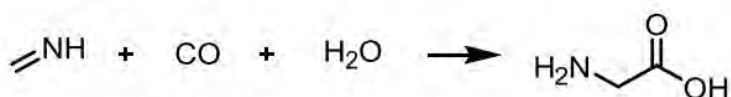
Example 4:



Transformation of formic acid into carbon dioxide requires presence of oxidation catalyst<sup>31</sup>, electrocatalyst<sup>32</sup>, or irradiation<sup>33</sup>.



### Example 5:

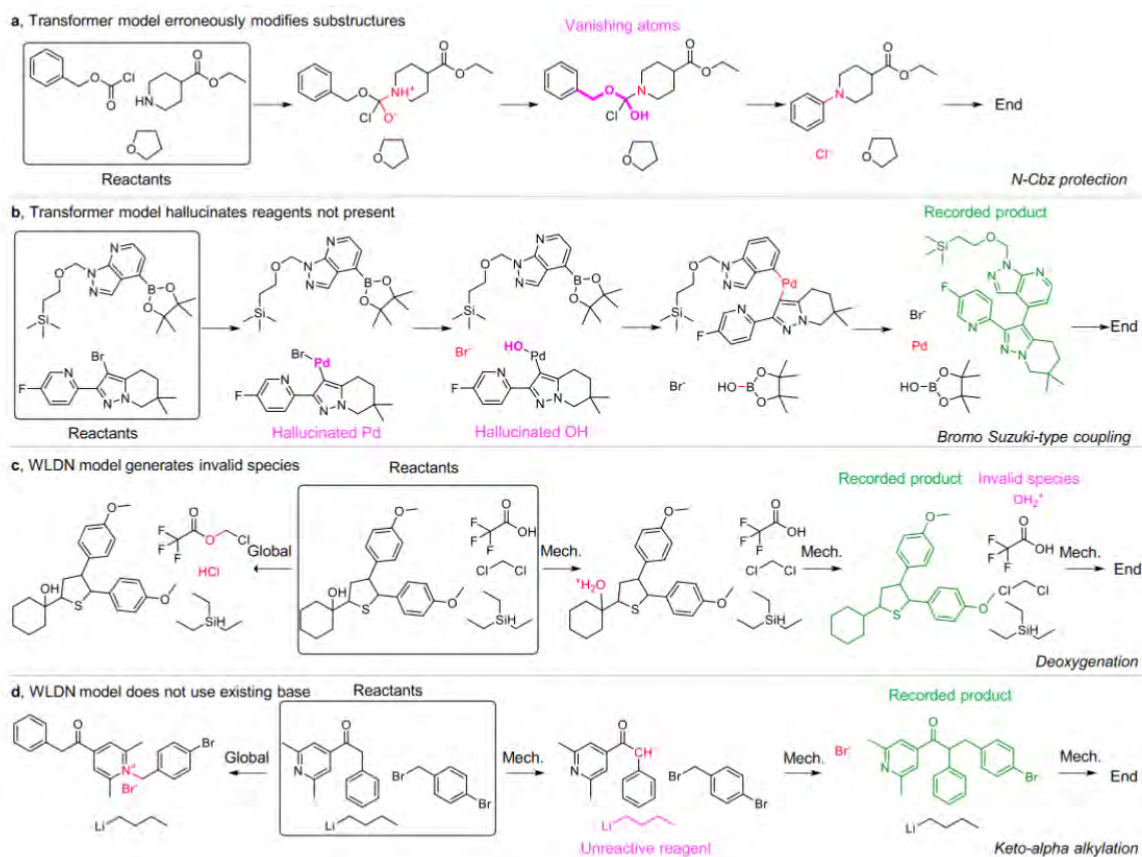


There are no known literature precedents similar to the reaction proposed by the authors. The closest reported example started from carbon monoxide, ammonia and cyanide<sup>34</sup> and required transition metal catalyst and high pressure.

### **Section 3.4. Reproducing reaction mechanisms with Machine-Learning models.**

Finally, we narrate an interesting recent work from MIT's Coley group<sup>22</sup> who trained machine learning models on a mechanistic dataset created by inputting intermediates between experimentally reported reactants and products. To this end, they used expert-coded reaction templates corresponding to 86 most popular reaction types. Subsequently, they applied these mechanistic steps to “connect” the substrates and products of 1.3 million reactions (from Pistachio database) and used this as a training set (using graph-based, sequence-based and graph-to-sequence models).

As the title of the paper itself indicates, the ML models perform well on the reaction classes on which they were trained but, as the authors also admit, there is a substantial proportion of errors in which unphysical species are, per the terminology of LLMs, “hallucinated” (see **Supplementary Figure S28** below). Importantly, the models do not generalize to unseen reaction types and thus cannot be used to discover new reactions which is the main objective of our current work. This, in hindsight, may be expected as the training on “known” mechanistic sequences predisposes the models to produce similar rather than novel mechanisms. We also note that models of this sort are unaware of kinetics which is one of the key elements in our approach, allowing us to detect sequences in which an apparently plausible mechanistic sequence is hijacked by a faster side-reaction (see main-text Figure 1).



**Supplementary Figure S28. Figure reproduced with permission from ref. <sup>22</sup> and illustrating some of the common failure modes of the ML models trained on known mechanistic sequences.** The original caption reads as follows: “Figure 4. Demonstration of typical failure modes observed in mechanistic models. **a**, Failure in prediction by the mechanistic Transformer model due to the deletion of atoms highlighted in magenta. **b**, Successful prediction by the mechanistic Transformer model of the recorded product while generating new atoms highlighted in magenta. **c**, Successful prediction by the mechanistic WLDN model of the recorded product through non-physical mechanisms, but global model fails to predict the product. **d**, Successful prediction by the mechanistic WLDN model of the recorded product while ignoring the organolithium base, but the global mode fails to predict the product.”

## Section S4. List of patterns of reactive groups.

Supplementary Table S2. SMARTS patterns of reactive functional groups used for the construction of pre-curated collection of substrates available through Allchemy's Web-app.

Name	Smarts
Aldehyde Aryl	[CX3H1\$([CX3H1])([c])=[OX1]]
Aldehyde Enolizable	[CX3H1\$([CX3H1])([CX4H3,CX4H2,CX4H1])=[OX1]]
Aldehyde sp3 NonEnolizable	[CX3H1\$([CX3H1])([CX4H0])=[OX1]]
Aldehyde Carbonyl	[CX3H1\$([CX3H1])([CX3]=[OX1])[#6,OX2,NX3,SX2]=[OX1]]
Alcohol Allyl	[CX3]=[CX3][CX4H2\$([CH2]),CX4H1\$([CH])([CX3]=[CX3])[#6],CX4H0\$(C([CX3]=[CX3])([#6])[#6]))][OH]
Alcohol Propargyl	[CX2]#[CX2][CX4H2\$([CH2]),CX4H1\$([CH](C#C)[#6]),CX4H0\$(C([#6])(C#C)[#6]))][OH]
Alcohol	[CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]))][OH]
Phenol	[c][OH]
Carboxylic acid	[CX3\$(CX3)([#6])=[OX1][OH]]
Thiol	[CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]))][SX2H1]
Thiophenol	[c][SX2H1]
Thioacid	[CX3H0\$([CX3)([#6])=[OX1][SX2H1]]
Amine Primary	[CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c][NX3H2]
Amine Secondary	[CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c][NX3H1][CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c]
Hydroxylamine	[NX3H2,NX3H1\$([NX3H1][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c),NX3H0\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c)][CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c)][OX2H1,OX2H0\$(OX2)[CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6])]]
Hydrazine NH2	[NX3H1\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3])),NX3H0\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3]))][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3]))][NX3H2]
Hydrazine NH1	[NX3H1\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3])),NX3H0\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3]))][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3]))][NX3H1\$([NX3][CX4H3,CX4H2\$([CH2])[#6],CX4H1\$([CH])([#6])[#6],CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$([CX3]([NX3][NX3])=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S([NX3][NX3])=[OX1])[#6],SX4\$(S([NX3][NX3])=(O)=(O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX3])=(O)([#6,OX2H0,NX3])[#6,OX2H0,NX3]))]
diCarbonyl CH Acid Acyclic	[CX3H0\$([CX3H0])([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]][CX4!H0R0][CX3H0\$(CX3H0)([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]]
diCarbonyl CH Acid Cyclic	[CX3H0\$([CX3H0])([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]][CX4!H0R1][CX3H0\$(CX3H0)([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]]
CH Acid Carbonyl Nitro	[CX3H0\$([CX3H0])([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]][CX4!H0][N+\$(N+)([CX4!H0])=(O)[O-]]
CH Acid Carbonyl CF3/CN	[CX3H0\$([CX3H0])([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]][CX4!H0][CX2\$(C([CX4!H0])#N,CX4\$(C([CX4!H0])(F)(F)F)]
CH Acid Carbonyl Sulphur/Phosphorus	[CX3H0\$([CX3H0])([CX4!H0])=[OX1])[#6,OX2H0,NX3,SX2]][CX4!H0][SX3\$([SX3]([CX4!H0])=(O)[#6,OX2H0,NX3]),SX4\$([SX4]([CX4!H0])=(O)=(O)[#6,OX2H0,NX3]),PX4\$([PX4]([CX4!H0])=(O)([#6,OX2,NX3,SX2])[#6,OX2,NX3,SX2])]
CH Acid noCarbonyl	[SX3\$([SX3]([CX4!H0])=(O)[#6,OX2H0,NX3]),SX4\$([SX4]([CX4!H0])=(O)=(O)[#6,OX2H0,NX3]),PX4\$([PX4]([CX4!H0])=(O)([#6,OX2,NX3,SX2])[#6,OX2,NX3,SX2]),CX2\$(C([CX4!H0])#N,CX4\$(C([CX4!H0])(F)(F)F),N+\$(N+)([CX4!H0])=(O)[O-]

	)]CX4!H0][SX3\$([SX3]([CX4!H0])(=O)[#6,OX2H0,NX3]),SX4\$([SX4]([CX4!H0])(=O)(=O)[#6,OX2H0,NX3]),PX4\$([PX4]([CX4!H0])(=O)(=O)[#6,OX2,NX3,SX2])[#6,OX2,NX3,SX2]),CX2\$(C([CX4!H0])#N),CX4\$(C([CX4!H0])(F)(F)F),N+\$([N+])([CX4!H0])(=O)[O-])]]
<b>Nitrile Enolizable</b>	[CX4!H0][CX2]#[NX1]
<b>Aldimine</b>	[CX3H1\$([CX3H1])([#6])=NX2\$([NX2][c,CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6])]])]
<b>Ketimine</b>	[CX3H0\$([CX3](=[NX2\$([NX2][SiX4,c,CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6])])))([#6])[#6])]]
<b>Aldimine N-EWG</b>	[CX3H1\$([CX3H1])([#6])=NX2\$([NX2][CX3\$(CX3([NX2])(=[OX1])[#6,OX2,NX3,SX2]),SX3\$(S(=[OX1])[#6]),SX4\$(S(=O)(=O)[#6,OX2H0,NX3]),PX4\$([PX4](=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]])]
<b>Ketimine N-EWG</b>	[CX3H0\$(C(=[NX2\$([NX2][CX3\$(CX3([NX2])(=[OX1])[#6,OX2,NX3,SX2]),SX3\$(S(=[OX1])[#6]),SX4\$(S(=O)(=O)[#6,OX2H0]),PX4\$([PX4](=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])])))([#6])[#6])]]
<b>Oxime</b>	[CX3H1\$([CX3H1](=[NX2\$([NX2][OH])[#6]),CX3H0\$([CX3](=[NX2\$([NX2][OH])])([#6])[#6]),c=[NX2][OH]
<b>Hydrazone</b>	[#6][CX3H1,CX3H0\$([CX3]([#6])[#6])=NX2\$([NX2][NX3H1][CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$(CX3([NX3][NX2])(=[OX1])[#6,OX2H0,NX3]),SX3\$(S([NX3][NX2])(=[OX1])[#6]),SX4\$(S([NX3][NX2])(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX2])(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3]),NX2\$([NX2][NX3H0])([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$(CX3([NX3][NX2])(=[OX1])[#6,OX2H0,NX3]),SX3\$(S([NX3][NX2])(=[OX1])[#6]),SX4\$(S([NX3][NX2])(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX2])(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]])([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4,CX3\$(CX3([NX3][NX2])(=[OX1])[#6,OX2H0,NX3]),SX3\$(S([NX3][NX2])(=[OX1])[#6]),SX4\$(S([NX3][NX2])(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([NX3][NX2])(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]]]
<b>Enamine</b>	[CX3]=[CX3][NX3H2,NX3H1\$([NX3H1][CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4),NX3H0\$([NX3]([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4)[CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4])]
<b>Enamine EWG</b>	[CX3]=[CX3][NX3H1\$([NX3H1][CX3\$(CX3([NX3]C=C)(=[OX1])[#6,OX2H0,NX3]),SX3\$(S([NX3]C=C)(=[OX1])[#6]),SX4\$(S([NX3]C=C)(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([NX3]C=C)(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]),NX3H0\$([NX3]([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4)[CX3\$(CX3([NX3]C=C)(=[OX1])[#6,OX2H0,NX3]),SX3\$(S([NX3]C=C)(=[OX1])[#6]),SX4\$(S([NX3]C=C)(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([NX3]C=C)(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]]]
<b>Enol Ether Alkyl</b>	[CX3]=[CX3][OX2][CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6]),c,SiX4]
<b>Enol Ether EWG</b>	[CX3]=[CX3][OX2][CX3\$(CX3([OX2]C=C)(=[OX1])[#6,OX2H0,NX3]),SX4\$(S([OX2]C=C)(=O)(=O)[#6,OX2H0]),PX4\$([PX4]([OX2]C=C)(=O)(=O)[#6,OX2H0,NX3])[#6,OX2H0,NX3])]
<b>Cyanate</b>	[#6][OX2][CX2]#[NX1]
<b>Isocyanate</b>	[#6][NX2]=[CX2]=[OX1]
<b>Isothiocyanate</b>	[#6][NX2]=[CX2]=[SX1]
<b>Diazonium</b>	[#6\$([#6][NX2+])#[NX1]]
<b>Diazo C-H</b>	[#6][CX3H1]=[N+]=[N-]
<b>Diazo</b>	[#6][CX3H0]([#6])=[N+]=[N-]
<b>Weinreb Amide</b>	[CX3\$(CX3(=[OX1])([NX3H1,NX3H0\$([NX3H0])([CX3\$(CX3(=[OX1])[#6])])][CX3\$(CX3(=[OX1])[#6]),c,CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6])]])]][OX2][c,CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])[#6]),CX4H0\$(C([#6])([#6])[#6])]]]
<b>Oxirane Terminal</b>	[CX4H2,CX4H1\$([CH]([#6])([CX4]3)O3),CX4H0\$(C([#6])([#6])([CX4]3)O3)]2[CX4H2][OX2]2
<b>Oxirane Internal</b>	[CX4H1\$([CH]([#6])([CX4]3)O3),CX4H0\$(C([#6])([#6])([CX4]3)O3)]2[CX4H1\$([CH]([#6])([CX4]3)O3),CX4H0\$(C([#6])([#6])([CX4]3)O3)][OX2]2
<b>Alkyl Halide Primary Cl Activated</b>	[CX4H2\$([CH2][CX3,CX2,c,OX2,SX2])][Cl]
<b>Alkyl Halide Primary Cl</b>	[CX4H3,CX4H2\$([CH2][CX4])][Cl]
<b>Alkyl Halide Primary Br Activated</b>	[CX4H3,CX4H2\$([CH2][CX3,CX2,c,OX2,SX2])][Br]
<b>Alkyl Halide Primary Br</b>	[CX4H2\$([CH2][CX4])][Br]
<b>Alkyl Halide Primary I Activated</b>	[CX4H2\$([CH2][CX3,CX2,c,OX2,SX2])][I]
<b>Alkyl Halide Primary I</b>	[CX4H2\$([CH2][CX4])][I]
<b>Alkyl Halide Secondary Cl</b>	[CX4H1\$([CH]([#6])[#6])][Cl]
<b>Alkyl Halide Secondary Br</b>	[CX4H1\$([CH]([#6])[#6])][Br]
<b>Alkyl Halide Secondary I</b>	[CX4H1\$([CH]([#6])[#6])][I]

Alkyl Halide Tertiary Cl	[CX4H0][Cl]
Alkyl Halide Tertiary Br	[CX4H0][Br]
Alkyl Halide Tertiary I	[CX4H0][I]
Dienone Acyclic	[CX3]=[CX3][CX3R0](=[OX1])[CX3]=[CX3]
Dienone Cyclic	[CX3]=[CX3][CX3R1](=[OX1])[CX3]=[CX3]
Terminal Unsaturated Aldehyde	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H1\$([CX3H1]([CX3]=[CX3])=[OX1])]
Terminal Unsaturated Ketone	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3]([OX1])((CX3]=[CX3])][CX3]=[CX3])\$([CX3H0]([CX3]=[CX3])=[OX1])#6)]
Terminal Unsaturated Amide	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[NX3]]
Terminal Unsaturated Ester	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[OX2\$([OX2][CX4,c]),SX2\$([SX2][CX4,c])]]
Terminal Unsaturated Phosphonate	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][PX4\$([PX4]([CX3]=[CX3])=(O))#6,OX2,NX3,SX2)]#6,OX2,NX3,SX2]]
Terminal Unsaturated Sulfoxide/Sulfone	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][SX3\$([SX3]([CX3]=[CX3])=(O))#6,OX2H0,NX3),SX4\$([SX4]([CX3]=[CX3])=(O)=(O))#6,OX2H0,NX3]]
Terminal Unsaturated CF3	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX4\$(C([CX3]=[CX3])(F)(F)F)]
Terminal Unsaturated Nitrile	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX2\$(C#N)]
Terminal Unsaturated Nitro	[CX3H2]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][N+\$([N+]([CX3]=[CX3])=(O)[O-])]
Internal Unsaturated Aldehyde	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H1\$([CX3H1]([CX3]=[CX3])=[OX1])]
Internal Unsaturated Ketone Acyclic Alpha H	[CX3H1]=@[CX3H1][CX3H0\$([CX3]([OX1])((CX3]=[CX3])][CX3]=[CX3])\$([CX3H0]([CX3]=[CX3])=[OX1])#6)]
Internal Unsaturated Ketone Cyclic Alpha H	[CX3H1]=@[CX3H1][CX3H0\$([CX3]([OX1])((CX3]=[CX3])][CX3]=[CX3])\$([CX3H0]([CX3]=[CX3])=[OX1])#6)]
Internal Unsaturated Amide Alpha H	[CX3H1]=[CX3H1][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[NX3]]
Internal Unsaturated Ester Alpha H	[CX3H1]=[CX3H1][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[OX2\$([OX2][CX4,c]),SX2\$([SX2][CX4,c])]]
Internal Unsaturated Ketone Acyclic Alpha R	[CX3H1]=@[CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3]([OX1])((CX3]=[CX3])][CX3]=[CX3])\$([CX3H0]([CX3]=[CX3])=[OX1])#6)]
Internal Unsaturated Ketone Cyclic Alpha R	[CX3H1]=@[CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3]([OX1])((CX3]=[CX3])][CX3]=[CX3])\$([CX3H0]([CX3]=[CX3])=[OX1])#6)]
Internal Unsaturated Amide Alpha R	[CX3H1]=[CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[NX3]]
Internal Unsaturated Ester Alpha R	[CX3H1]=[CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1])[OX2\$([OX2][CX4,c]),SX2\$([SX2][CX4,c])]]
Internal Unsaturated Phosphonate	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][PX4\$([PX4]([CX3]=[CX3])=(O))#6,OX2,NX3,SX2)]#6,OX2,NX3,SX2]]
Internal Unsaturated Sulfoxide/Sulfone	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][SX3\$([SX3]([CX3]=[CX3])=(O))#6,OX2H0,NX3),SX4\$([SX4]([CX3]=[CX3])=(O)=(O))#6,OX2H0,NX3]]
Internal Unsaturated CF3	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX4\$(C([CX3]=[CX3])(F)(F)F)]
Internal Unsaturated Nitrile	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][CX2\$(C#N)]
Internal Unsaturated Nitro	[CX3H1]=[CX3H1,CX3H0\$([CX3][CX4,c,SiX4,F,Cl,Br,I,SX2H0,OX2H0,NX3])][N+\$([N+]([CX3]=[CX3])=(O)[O-])]
Unsaturated diEWG	[CX3H2,CX3H1]=[CX3]([CX3H1]([CX3H1]([CX3]=[CX3])=[OX1,SX1,NX2]),CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1,SX1,NX2])#6,OX2H0,NX3,SX2),SX3\$([SX3]([CX3]=[CX3])=(O))#6,OX2H0,NX3),SX4\$([SX4]([CX3]=[CX3])=(O)=(O))#6,OX2H0,NX3),PX4\$([PX4]([CX3]=[CX3])=(O))#6,OX2,NX3,SX2)]#6,OX2,NX3,SX2),CX2\$(C([CX3]=[CX3])#N),CX4\$(C([CX3]=[CX3])(F)(F)F),N+\$([N+]([CX3]=[CX3])=(O)[O-]))][CX3H1\$([CX3H1]([CX3]=[CX3])=[OX1,SX1,NX2]),CX3H0\$([CX3H0]([CX3]=[CX3])=[OX1,SX1,NX2])#6,OX2H0,NX3,SX2),SX3\$([SX3]([CX3]=[CX3])=(O))#6,OX2H0,NX3),SX4\$([SX4]([CX3]=[CX3])=(O)=(O))#6,OX2H0,NX3),PX4\$([PX4]([CX3]=[CX3])=(O))#6,OX2,NX3,SX2)]#6,OX2,NX3,SX2),CX2\$(C([CX3]=[CX3])#N),CX4\$(C([CX3]=[CX3])(F)(F)F),N+\$([N+]([CX3]=[CX3])=(O)[O-])]
Alkynyl Aldehyde	[CX2]#[CX2][CX3H1\$([CX3H1]([CX2]#[CX2])=[OX1])]
Alkynyl Ketone	[CX2]#[CX2][CX3H0\$([CX3H0]([CX2]#[CX2])=[OX1])#6)]
Alkynyl Amide	[CX2]#[CX2][CX3H0\$([CX3H0]([CX2]#[CX2])=[OX1])[NX3]]
Alkynyl Ester	[CX2]#[CX2][CX3H0\$([CX3H0]([CX2]#[CX2])=[OX1])[OX2H0]]

<b>Alkynyl Phosphonate</b>	[CX2][CX2][PX4\$(PX4)(CX2)(CX2)(=O)(#6,OX2,NX3,SX2))][#6,OX2,NX3,SX2]]
<b>Alkynyl Sulfoxide/Sulfone</b>	[CX2][CX2][SX3\$(SX3)(CX2)(CX2)(=O)(#6,OX2H0,NX3),SX4\$(SX4)(CX2)(CX2)(=O)(=O)(#6,OX2H0,NX3))]
<b>Alkynyl CF3</b>	[CX2][CX2][CX4\$(C(CX2)(CX2)(F)(F)F)]
<b>Alkynyl Nitrile</b>	[CX2][CX2][CX2\$(C(CX2)(CX2)#N)]
<b>Alkynyl Nitro</b>	[CX2][CX2][N+\$(N+)(CX2)(CX2)(=O)[O-]]
<b>Ketone Enolizable</b>	[CX4H0][CX3](=[OX1])[#6]
<b>Hemiacetal</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6]][OX2][CX4H1,CX4H0\$(CH0)([#6])([#6])([#6])[OH]
<b>Hemiaminal NH0</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6]][NX3](CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c.CX3\$(CX3)(NX3)(=OX1))[#6,OX2H0,NX3],SX3\$(S(NX3)(=OX1))[#6],SX4\$(S(NX3)(=O)(=O)(#6,OX2H0)),PX4\$(PX4)(NX3)(=O)(#6,OX2H0,NX3))[#6,OX2H0,NX3]][CX4H1,CX4H0\$(CH0)([#6])([#6])[#6]][OH]
<b>Hemiaminal NH</b>	[NX3H1](CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c.CX3\$(CX3)(NX3)(=OX1))[#6,OX2H0,NX3],SX3\$(S(NX3)(=OX1))[#6],SX4\$(S(NX3)(=O)(=O)(#6,OX2H0)),PX4\$(PX4)(NX3)(=O)(#6,OX2H0,NX3))[#6,OX2H0,NX3]][CX4H1,CX4H0\$(CH0)([#6])([#6])[#6]][OH]
<b>Acetal Alkyl</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c][OX2](O3CX4O)[CX4,c][CX4,c]3)[CX4H2,CX4H1\$(CH)(O)(O)([#6]),CX4H0\$(CH0)(O)(O)([#6])[#6]][OX2][CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c]
<b>Acetal Acyl</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6]][OX2][CX4H1,CX4H0\$(CH0)([#6])([#6])([#6))[OX2][CX3](=OX1)]#6]
<b>Aminal Alkyl</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],CX3\$(CX3)(=OX1))[#6]][NX3](CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c.CX3\$(CX3)(NX3)(=OX1))[#6,OX2H0,NX3],SX3\$(S(NX3)(=OX1))[#6],SX4\$(S(NX3)(=O)(=O)(#6,OX2H0)),PX4\$(PX4)(NX3)(=O)(#6,OX2H0,NX3))[#6,OX2H0,NX3]][CX4H1,CX4H0\$(CH0)([#6])([#6])[#6]][OX2][CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6]]]
<b>Disulfide</b>	[c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],CX3\$(CX3)(=O)[#6]][SX2][SX2][c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],CX3\$(CX3)(=O)[#6]]]
<b>ThioKetone</b>	[#6][CX3](=[SX1])[#6]
<b>ThioAldehyde</b>	[#6][CX3H1]=[SX1]
<b>Thiirane</b>	[CX4]([SX2]1)[CX4]1
<b>Thiocyanate</b>	[#6][SX2][CX2][NX1]
<b>ThionoEster</b>	[#6][CX3H0](=[SX1])[OX2][c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],CX3\$(CX3)(=O)[#6]]]
<b>ThioEster</b>	[#6][CX3H0](=[OX1])[SX2][c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],CX3\$(CX3)(=O)[#6]]]
<b>ThioAmide</b>	[#6][CX3H0](=[SX1])[NX3H2,NX3H1\$(NX3)[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c],NX3H0\$(NX3)(CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c)[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c]]]
<b>Azide</b>	[#6][NX2]=[N+]=[N-]
<b>Isonitrile</b>	[#6][NX2+][CX1-]
<b>Amide Tertiary</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c][NX3H0](CX3\$(CX3)[#6])(=[OX1])[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c]
<b>Amide Secondary</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c][NX3H1][CX3](=[OX1])[#6]
<b>Amide Primary</b>	[NX3H2][CX3](=[OX1])[#6]
<b>Sulfonamide NH</b>	[NX3H1\$(NX3H1)[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c,SiX4],NX3H2][SX3](=[OX1])[#6]
<b>Sulfonamide NH</b>	[NX3H1\$(NX3H1)[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c,SiX4],NX3H2][SX4](=[OX1])(=[OX1])[#6,OX2H0,NX3H0]
<b>Thioacetal CH</b>	[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c][SX2][CX4H0][SX2][CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)([#6])[#6],CX4H0\$(C)([#6])([#6])[#6],c]
<b>nH Heterocycle r5</b>	[nHr5\$(nHr5)2aaaa2]
<b>nH Heterocycle r6</b>	[nHr6\$(nHr6)2c(=[OX1,SX1])aaaa2,nHr6\$(nHr6)2aac(=[OX1,SX1])aa2)]



<b>Nitrile</b>	[#6!\$(#6)[66CX2]][CX2]#(NX1)
<b>Ester NonEnolizable</b>	[CX3\$(CX3)=(OX1)](OX2)[c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])][CX4H0,c,CX3\$(CX3=O)]
<b>Azirine</b>	[NX2]2=[NX2][CX4H2,CX4H1\$(CH)[#6],CX4H0\$(C(#6)[#6])]2
<b>Cyanamide</b>	[NX2\$(NX2)=CX3,c],NX3H1\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])],NX3H0\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])][CX2]#(NX1)
<b>Amidine</b>	[NX3H2,NX3H1\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])],NX3H0\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[(#6)][#6],CX4H0\$(C(#6)(#6)[#6])][CX3H1,CX3H0\$(CX3)[#6]]=[NX2H1,NX2H0\$(NX2)[#6]]
<b>Carbonate</b>	[OX2\$(OX2)[CX4,c]][CX3](=OX1)[OX2\$(OX2)[CX4,c]]
<b>Thiourea</b>	[NX3][CX3](=SX1)[NX3]
<b>Ureas</b>	[NX3][CX3](=OX1)[NX3]
<b>SemiUreas</b>	[NX3][CX3](=OX1,SX1)[OX2H0,SX2H0]
<b>Chloramines</b>	[F,Cl,Br,I][NX3][CX4,c,CX3,SX4][CX4,c,CX3,SX4]
<b>Guanidine</b>	[NX3]C(=NX2)[NX3]
<b>Allylsilane sp2/sp/allyl</b>	[CX3\$(CX3)=CX3,CX2),CX2\$(CX2)#(CX2),c,CX4\$(CX4)[CX3]=CX3,CX2),CX4\$(CX4)[CX2]#(CX2)][Si](#6)([#6)][#6]
<b>Stannane sp2/sp/allyl</b>	[CX3\$(CX3)=CX3,CX2),CX2\$(CX2)#(CX2),c,CX4\$(CX4)[CX3]=CX3,CX2)][Sn](#6)([#6)][#6]
<b>Nitrate Ester</b>	[#6][OX2][NX2]=[OX1]
<b>Nitroso</b>	[#6,NX3][NX2]=[OX1]
<b>Azo</b>	[#6][NX2]=[NX2][#6]



**Supplementary Table S3.** SMARTS patterns of functional groups forbidden in pre-curved collection of substrates available through Allchemy's Web-app.

Name	Smarts
Sulfonic Acid	[#6,OX2H0,NX3,SX2,SiX4][SX4](=[OX1])(=[OX1])[OH]
Sulfinic Acid	[#6,OX2H0,NX3,SX2,SiX4][SX3](=[OX1])[OH]
Hydrazine NH0	[NX3H0\$(NX3)[CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6],c,SiX4,CX3\$(CX3)[NX3](=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S[NX3][NX3])(=[OX1])[#6],SX4\$(S[NX3][NX3])(=O)(=O)[#6,OX2H0],PX4\$(PX4)[NX3][NX3])(=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S[NX3][NX3])(=[OX1])[#6],SX4\$(S[NX3][NX3])(=O)(=O)[#6,OX2H0],PX4\$(PX4)[NX3][NX3])(=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S[NX3][NX3])(=[OX1])[#6],SX4\$(S[NX3][NX3])(=O)(=O)[#6,OX2H0],PX4\$(PX4)[NX3][NX3])(=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S[NX3][NX3])(=[OX1])[#6],SX4\$(S[NX3][NX3])(=O)(=O)[#6,OX2H0],PX4\$(PX4)[NX3][NX3])(=[OX1,SX1])[#6,OX2H0,NX3],SX3\$(S[NX3][NX3])(=[OX1])[#6],SX4\$(S[NX3][NX3])(=O)(=O)[#6,OX2H0],PX4\$(PX4)[NX3][NX3])(=[OX1,SX1])[#6,OX2H0,NX3]]
Oxime Ether	[#6][CX3H1,CX3H0\$(CX3)([#6])(=[NX2])[#6])=[NX2][OX2][SiX4,c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Oxime O-EWG	[#6][CX3H1,CX3H0\$(CX3)([#6])(=[NX2])[#6])=[NX2][OX2][CX3\$(CX3)([NX2])(=[OX1])[#6,OX2,NX3,SX2],SX3\$(S(=[OX1])[#6],SX4\$(S(=O)(=O)[#6,OX2H0,NX3],PX4\$(PX4)(=O)(=[#6,OX2H0,NX3])[#6,OX2H0,NX3])]
Hydroxamic Acid	[#6][CX3](=[OX1])[NX3H1,NX3H0\$(NX3H0)[c,CX4H3,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]][OX2H1]
Ether	[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][OX2!r3!r4][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Thioether	[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][SX2!r3!r4][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Boronic Acid sp3	[CX4][BX3][OX2H1][OX2H1]
Chloro Fluorinated	[c!\$(c1(Cl)C(-[CX3H1\$(CX3H1)(c)=[OX1,SX1,NX2],CX3H0\$(CX3H0)(c)=[OX1,SX1,NX2])[#6,OX2H0,NX3,SX2],SX3\$(SX3)(c)=[OX1,SX1,NX2],SX4\$(SX4)(c)=[OX1,SX1,NX2])[#6,OX2H0,NX3,SX2],PX4\$(PX4)(c)=[OX1,SX1,NX2])[#6,OX2,NX3,SX2],CX2\$(C(c)F)(F)F),N+\$(N+)(c)=[OX1,SX1,NX2])ccccc1)\$\$(c2(Cl)ccc(-[CX3H1\$(CX3H1)(c)=[OX1,SX1,NX2],CX3H0\$(CX3H0)(c)=[OX1,SX1,NX2])[#6,OX2H0,NX3,SX2],SX3\$(SX3)(c)=[OX1,SX1,NX2],SX4\$(SX4)(c)=[OX1,SX1,NX2])[#6,OX2H0,NX3,SX2],PX4\$(PX4)(c)=[OX1,SX1,NX2])[#6,OX2,NX3,SX2],CX2\$(C(c)F)(F)F),N+\$(N+)(c)=[OX1,SX1,NX2])cc2)\$\$(c3(Cl)nc[c,n]c[c,n]3)\$\$(c3(Cl)[c,n]c[c,n]3)[F,Cl]
Vinyl Fluoride	[CX3]=[CX3]F
Enol	[CX3]=[CX3][OH]
Amine Tertiary	[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][NX3][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Carbamate	[NX3H2,NX3H1\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]],NX3H0\$(NX3)[CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX3](=[OX1])[OX2][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Alkyl Polyhalide	[#6][CX4H1,CX4H0\$(CX4)([#6,F,Cl])([F,Cl])([F,Cl])([F,Cl])([F,Cl])]
Alkyl Fluoride	[CX4]F
Ammonium Salt	[NX4+][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
AzoliumSalt R6	[nX3H0r6+][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
AzoliumSalt R5	[nX3H0r5+][CX4H3,c,CX4H2\$(CH2)[#6],CX4H1\$(CH)[#6][#6],CX4H0\$(C[#6])([#6])[#6]]
Silyl Ether	[CX4.c][OX2][SiX4]([#6])([#6])[#6]
Sulfur Halide	[#6][SX2][Cl,Br,I]
Phosphonic acid	[#6][PX4](=[OX1])[OH][OH]
Phosphinic acid	[#6][PX4](=[OX1])[OH]

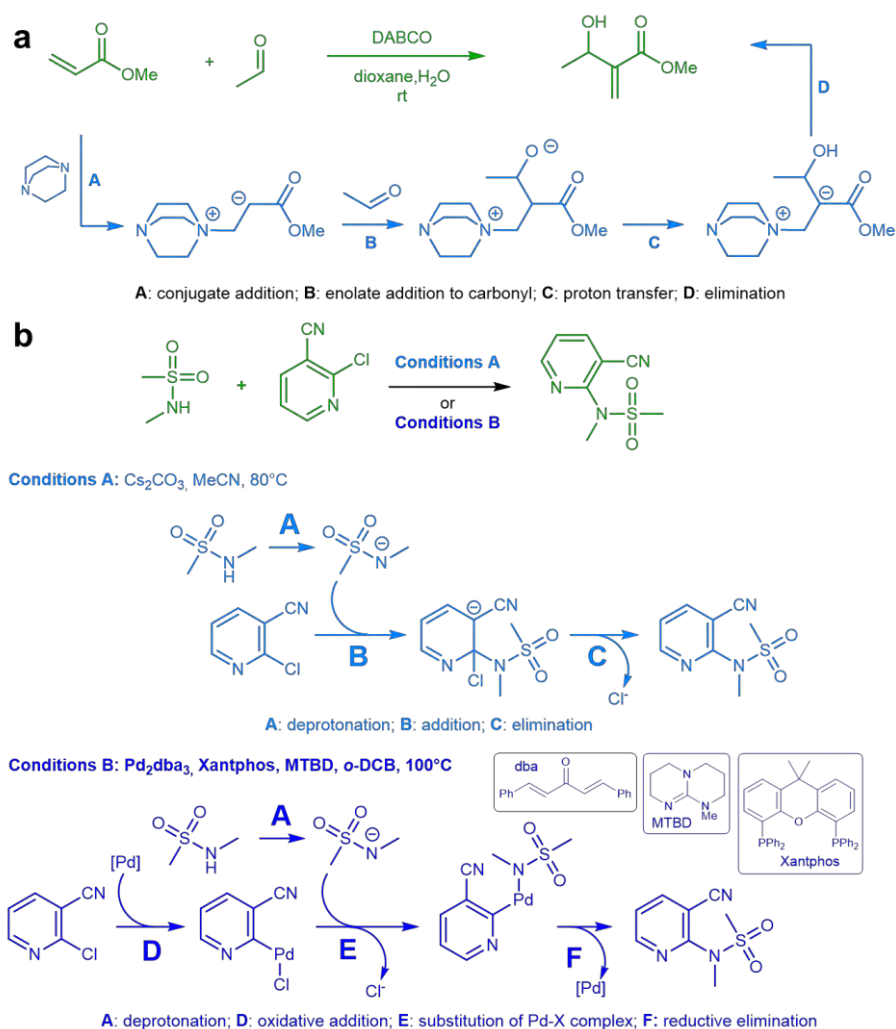
<b>Oxazirine</b>	[NX3]2[OX2][CX4]2
<b>Sulfinate Ester</b>	[CX4,c][OX2][SX3](=O)[OX2][CX4,c]
<b>Sulfonate Ester</b>	[CX4,c][OX2][SX4](=O)(=O)[OX2][CX4,c]
<b>Sulfinylimine</b>	[#6,OX2H0,NX3][SX4](=O)(=[NX2])[#6,OX2,NX3]
<b>Iminoether</b>	[CX4,c][OX2][CX3]([#6])=[NX2H1,NX2H0\$([NX2][CX4,c])]
<b>Boronic Ester sp3</b>	[CX4][BX3]([OX2H0])[OX2H0]
<b>Boronic Hemiester</b>	[#6][BX3]([OX2H0])[OX2H1]
<b>Tropinon</b>	c2cccccc2=O
<b>Peroxide</b>	[#6,CX3,SX4][OX2][OX2][#6,CX3,SX4]
<b>Organophosphate</b>	[CX4H0,c,OX2,SX2,NX3]P(=[OX1,SX1])([OX2,SX2,NX3])[OX2,SX2,NX3]
<b>Organophosphite</b>	[#6,OX2,SX2,NX3][PX3]([OX2,SX2,NX3])[OX2,SX2,NX3]
<b>Sulfonamide NH0</b>	[#6][SX4](=O)(=O)[NX3H0]([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6]),CX4H0\$(C([#6])([#6])([#6]),c))[CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])([#6]),CX4H0\$(C([#6])([#6])([#6]),c)]
<b>Imide SulfonimideH0</b>	[CX3\$([CX3](=[OX1])[#6,OX2\$(O[CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])([#6]),CX4H0\$(C([#6])([#6])([#6]),c),NX3!H2)],SX4\$(S(=O)(=O)[#6,OX2H0,NX3!H2]))][NX3]([CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])([#6]),CX4H0\$(C([#6])([#6])([#6]),c))[CX3\$([CX3](=[OX1])[#6,OX2\$(O[CX4H3,CX4H2\$([CH2][#6]),CX4H1\$([CH]([#6])([#6]),CX4H0\$(C([#6])([#6])([#6]),c),NX3!H2)],SX4\$(S(=O)(=O)[#6,OX2H0,NX3!H2]))]
<b>Furan</b>	c2[o!\$(o2ccc3c2aaaa3)]ccc2
<b>Benzofuran</b>	c2[o\$(o2ccc3c2aaaa3)]ccc2
<b>Tiophen</b>	c2[s!\$(s2ccc3c2aaaa3)]ccc2
<b>Benzotiophen</b>	c2[s\$(s2ccc3c2aaaa3)]ccc2
<b>Pyridine</b>	[nX2H0r6]2cccc2
<b>Pyridazine</b>	[nX2H0r6]2ncccc2
<b>Pyrimidine</b>	[nX2H0r6]2cnccc2
<b>Pyrazine</b>	[nX2H0r6]2ccncc2
<b>1,3,5-Triazine</b>	[nX2H0r6]2c[nX2H0r6]c[nX2H0r6]c2
<b>1,2,4-Triazine</b>	[nX2H0r6]2[nX2H0r6]cc[nX2H0r6]c2
<b>1,2,3-Triazine</b>	[nX2H0r6]2[nX2H0r6][nX2H0r6]ccc2
<b>Azepine</b>	[nX3]3cccccc3
<b>Indole</b>	c2[nX3H0\$(n2ccc3c2aaaa3)]ccc2
<b>Carbazole</b>	c2[nX3H0!\$(n2c(aaaa5)c5c3c2aaaa3)]ccc2
<b>Imidazole</b>	c2[nX3H0]c[nX2]c2
<b>Pyrazole</b>	c2[nX3H0][nX2]cc2
<b>(1H)-1,2,3-Triazole</b>	c2[nX3H0][nX2][nX2]c2
<b>(2H)-1,2,3-Triazole</b>	c2[nX2][nX3H0][nX2]c2
<b>(1H)-1,2,4-Triazole</b>	c2[nX3H0][nX2]c[nX2]2

<b>(4H)-1,2,4-Triazole</b>	c2[nX2][nX2]c[nX3H0]2
<b>(1H)-Tetrazole</b>	n2nn[nX3H0]c2
<b>(2H)-Tetrazole</b>	n2n[nX3H0]nc2
<b>2-Imidazolone</b>	[nX3H0]3c(=[OX1])[nX3H0]cc3
<b>5-Imidazolone</b>	[nX2]3c(=[OX1])c[nX2]c3
<b>Triazolone</b>	[nX3H0]3c(=[OX1])[nX3H0]c[nX2]3
<b>Oxazole</b>	[nX2]3cocc3
<b>Isoxazole</b>	[nX2]3occc3
<b>1,3,4-Oxadiazole</b>	[nX2]3coc[nX2]3
<b>Oxazolone</b>	o2c(=O)[nX3H0]cc2
<b>Thioxazolone</b>	s2c(=O)[nX3H0]cc2
<b>1,2,4-Oxadiazole</b>	[nX2]3co[nX2]c3
<b>1,2,5-Oxadiazole</b>	[nX2]3o[nX2]cc3
<b>Thiazole</b>	[nX2]3csc3
<b>Isothiazole</b>	[nX2]3sccc3
<b>1,3,4-Thiadiazole</b>	s2c[nX2][nX2]c2
<b>1,2,4-Thiadiazole</b>	s2[nX2]c[nX2]c2
<b>1,2,3-Thiadiazole</b>	s2[nX2][nX2]cc2
<b>2-Pyridone</b>	[nH0r6]2c(=[OX1,SX1])aaaa2
<b>4-Pyridone</b>	[nH0r6]2aac(=[OX1,SX1])aa2
<b>Silane sp3</b>	[CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2])][Si]([CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2]))([CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2]))[CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2])
<b>Stannane sp3</b>	[CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2])][Sn]([CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2]))([CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2]))[CX4]!\$([CX4][CX3]=[CX3,CX2])!\$([CX4][CX2]#[CX2])
<b>Sulfone</b>	[#6][SX4](=O)(=O)[#6]
<b>Polycycle [1,1,1]</b>	[CX4]([*]2)([*]3)[*][CX4]23
<b>Polycycle [1,1,2]</b>	[CX4]([*]2)([*]3)[*]-[*][CX4]23
<b>Polycycle [1,1,3]</b>	[CX4]([*]2)([*]3)[*]-[*]-[*][CX4]23
<b>Polycycle [1,2,2]</b>	[CX4]([*]2)([*]-[*]3)[*]-[*][CX4]23
<b>Polycycle [1,2,3]</b>	[CX4]([*]2)([*]-[*]3)[*]-[*]-[*][CX4]23
<b>Polycycle [1,3,3]</b>	[CX4]([*]2)([*]-[*]-[*]3)[*]-[*]-[*][CX4]23
<b>Polycycle [2,2,2]</b>	[CX4]([*]-[*]2)([*]-[*]3)[*]-[*][CX4]23
<b>Polycycle [2,2,3]</b>	[CX4]([*]-[*]2)([*]-[*]3)[*]-[*]-[*][CX4]23
<b>Polycycle [2,3,3]</b>	[CX4]([*]-[*]2)([*]-[*]-[*]3)[*]-[*]-[*][CX4]23
<b>Polycycle [3,3,3]</b>	[CX4]([*]-[*]-[*]2)([*]-[*]-[*]3)[*]-[*]-[*][CX4]23

<b>Phosphine</b>	[#6][PX3]([#6])[#6]
<b>Phosphine Oxide</b>	[#6][PX3](=[OX1])([#6])[#6]
<b>Ketal</b>	[CX4H2,CX4H1\$([CH][#6]),CX4H0\$([CH0][[#6]][#6])]2[OX2][CX4H2,CX4H1\$([CH][[#6]][CX4]),CX4H0\$(C([CX4])([#6])[#6]),c][CX4H2,CX4H1\$([CH][[#6]][CX4]),CX4H0\$(C([CX4])([#6])[#6]),c]O2
<b>Pentafluorosulfanyl</b>	[#6][SX6](F)(F)(F)(F)F
<b>Cyclopropane</b>	[CX4]2[CX4][CX4]2
<b>Cyclohexadiene</b>	[c-]2cccc2

## Section S5. Tutorial on reaction rule coding

Our mechanistic analyses rely on expert-coded reaction rules covering single mechanistic ('arrow-pushing') steps. As noted in the main text, these mechanistic reaction rules cannot be automatically extracted from repositories such as Reaxys or Scifinder because reported reaction entries do not have any information regarding underlying mechanistic steps. For example, a simple two-component Morita-Baylis-Hilman reaction of methyl acrylate and acetaldehyde catalyzed by a tertiary amine is reported as *in green* in **Supplementary Figure S29a** (example taken from ref. <sup>35</sup>) but proceeds via four individual mechanistic steps (*blue*). For multicomponent reactions, the number of mechanistic steps is often higher – a “classic” Ugi four component reaction comprises five major mechanistic steps (shown in **Supplementary Figure S30a**).



**Supplementary Figure S29. a,** Reaction entries reported in repositories (*green*) do not have any information regarding underlying mechanistic steps and cannot therefore be used for automatic “learning” of such steps. Here, reported (*green*) two-component Morita-Baylis-

Hilman reaction proceeds via four separate mechanistic steps (*blue*). **b**, Furthermore, the sequence of mechanistic steps underlying reported reaction entry may depend on the reaction conditions. Substitution of aryl chlorides with sulfonamides (*green*) can proceed either via nucleophilic aromatic substitution when performed under basic *Conditions A* or via Pd-centered catalytic cycle when appropriate complex is used as a catalyst as in *Conditions B*.

The coding of a mechanistic reaction rule begins with careful examination of reported reaction example and identification of plausible sequences of elementary mechanistic steps leading from the substrate(s) to the reported product. Importantly, the reaction conditions and used reagents are carefully evaluated as the same reaction can proceed following substantially different elementary mechanisms when performed under different conditions (e.g., basic or with transition metal catalyst). For example, addition of sulfonamides to aryl chlorides (**Supplementary Figure S29b**) can proceed either via i) nucleophilic aromatic substitution (with deprotonation, addition leading to Meisenheimer complex, and elimination) or via ii) palladium mediated coupling (with deprotonation, oxidative addition, substitution of Pd-X complex with a nucleophile and reductive elimination). In this and other similar examples, assignment of a particular sequence of elementary mechanistic steps is supported by analysis of additional experiments (eg., with deuterated solvents or radical scavengers) or identified byproducts reported in the original manuscripts which allows for confirmation or exclusion of possible mechanisms.

In the next step, for each single identified mechanistic step, the chemist coding the rule must determine the scope of admissible substituents flanking the reaction center to provide appropriate steric (e.g., lack of bulky substituents for S<sub>N</sub>2-related mechanisms) and electronic environment (e.g., presence of appropriate activating electron withdrawing groups for nucleophilic aromatic substitutions or conjugate additions to alkenes). These considerations are coded as a reaction transform written in SMARTS notation<sup>36</sup>, being standard machine-readable representation of molecules and chemical reactions representing fully mapped chemical reaction as an alphanumeric string. In each reaction transform, the substrate set is written on the left side of the reaction arrow while main product and reaction byproducts are written on the right side of the reaction arrow with the main product written as the first one.

After specifying the scope of a given mechanistic step, the list of unequivocally cross-reactive functional groups needs to be specified. Otherwise, the computer will predict implausible reaction outcomes bound to fail in experiment.. In our analyses leading to discovery of MCRs, this list of cross reactive functional groups was used to evaluate if a) a given

substrate/intermediate do not have any cross reactive functional groups and b) if other substrates/intermediates present in the reaction mixture are compatible with given mechanistic step. The latter allowed us to classify discovered step sequences either as MCRs (if all substrates are compatible with each mechanistic steps) or one-pot reactions, when at least one of substrates was incompatible with initial mechanistic steps and had to be added later.

Finally, as detailed in the main text and **Mechanistic rules** part of the **Methods** section, each rule is categorized with respect to:

- i) general conditions (strongly basic, basic, neutral, acidic, strongly acidic, Lewis acid);
- ii) water tolerance (tolerated, not tolerated, required);
- iii) rough kinetics (very fast, fast, slow, very slow, uncertain),
- iv) admissible temperatures (very low, low, rt, high, very high)
- v) solvent class (polar/nonpolar and protic/aprotic)

With these general guidelines, let us first consider coding mechanistic transforms covering some key mechanistic steps of the Ugi reaction. This multicomponent reaction between an amine, an aldehyde, a carboxylic acid and isocyanide yields the  $\alpha$ -aminoacyl amide (with one of the reported examples from <sup>37</sup> shown in **Supplementary Figure S30a**). Importantly, Ugi 4CR comprises five major mechanistic steps<sup>38</sup>: formation of an imine from aldehyde and an amine, proton transfer from carboxylic acid substrate, addition of isocyanide to protonated imine, addition of carboxylate anion to nitrilium cation, and Mumm rearrangement. Each of these mechanistic steps is coded as a separate reaction rule.

As the first example, we discuss coding of the entire reaction rule for “Addition of isocyanide to imine” (step **C** in **Supplementary Figure S30a**). This mechanistic step comprises addition of isocyanide nucleophile with its terminal carbon atom to a protonated imine electrophile to give the nitrilium ion. This is illustrated in **Supplementary Figure S30b** where the reaction rule written in SMARTS notation is shown in the top part. The activated imine substrate is written in SMARTS notation with atoms #1 and #2 representing fragment derived from an aldehyde substrate and atoms #3 and #5 representing fragment derived from an amine substrate of the Ugi reaction. Atom #1 can be either aromatic or aliphatic without any further constrains regarding its hybridization. This is coded as [CX4,c,CX3+0,CX2:1] in SMARTS notation where CX4, c, CX3+0 and CX2 define sp<sup>3</sup>-hybridized, aromatic, uncharged sp<sup>2</sup>-hybridized or sp-hybridized carbon atoms, respectively. Atom #2 written as [CX3H:2] denotes sp<sup>2</sup>-hybridized carbon atom with one hydrogen atom, thus limiting the scope of this particular

reaction transform to aldimines. We note that such addition of isocyanides occurring for protonated ketimines (similar electrophile) is coded as a separate reaction transform because defining protonated ketimine in SMARTS notation requires definition of an additional substituent attached to atom #2. The nitrogen atom of protonated aldimine, #3, is written as [NX3H+:3] in SMARTS notation and defines sp<sup>2</sup>-hybridized nitrogen atom with one hydrogen and single positive charge. Finally, the #5 position denoting aldimine's nitrogen substituent is written as [CX4,c,NX3+0,NX2+0:5] and allows for aryl ('c') or sp<sup>3</sup>-alkyl ('CX4') carbons, uncharged sp<sup>3</sup>-nitrogen ('NX3+0', broadening the scope of this reaction transform with addition of isocyanides to protonated hydrazones; evidenced in ref. <sup>39</sup>) and uncharged sp<sup>2</sup>-nitrogen ('NX2+0', enabling addition to N<sub>2</sub>H<sub>4</sub> derived hydrazides evidenced in ref. <sup>40</sup>).

The isocyanide nucleophile is written in SMARTS notation with atoms #6 and #7 representing isocyanide group ([CX1-:6]#[NX2+:7]). The isocyanide group is attached to atom #8 which can be either aryl ('c'), alkyl ('CX4') or sp<sup>2</sup>-hybridized carbon ('CX3') atom. We note that sp-hybridized atoms are not allowed at this position because addition of such unsaturated isocyanides to carbonyl groups (either imines, aldehydes or ketones) had not been evidenced while their stability may raise concerns – in fact, cyanoisocyanide had been reported only as stabilized Cr-complex<sup>41</sup> while free alkynylisocyanides were prepared via FVP of metal complexes and required cryogenic trapping<sup>42</sup>.

For proper handling of reaction transform in Rdkit package<sup>43</sup>, all atoms which lost their charges (here, #3 and #6) have their neutral charges explicitly specified in the SMARTS string ('[N+0:3]' and '[C+0:6]') of reaction's product (right side of reaction arrow). This reaction transform do not generate any byproducts so the right side of the reaction (product side) contains only one SMARTS defining the main reaction product.

After specifying the reaction scope, groups present outside the reaction 'core' competing with the intended reaction outcome need to be specified. In this particular case, the list of incompatible groups contains

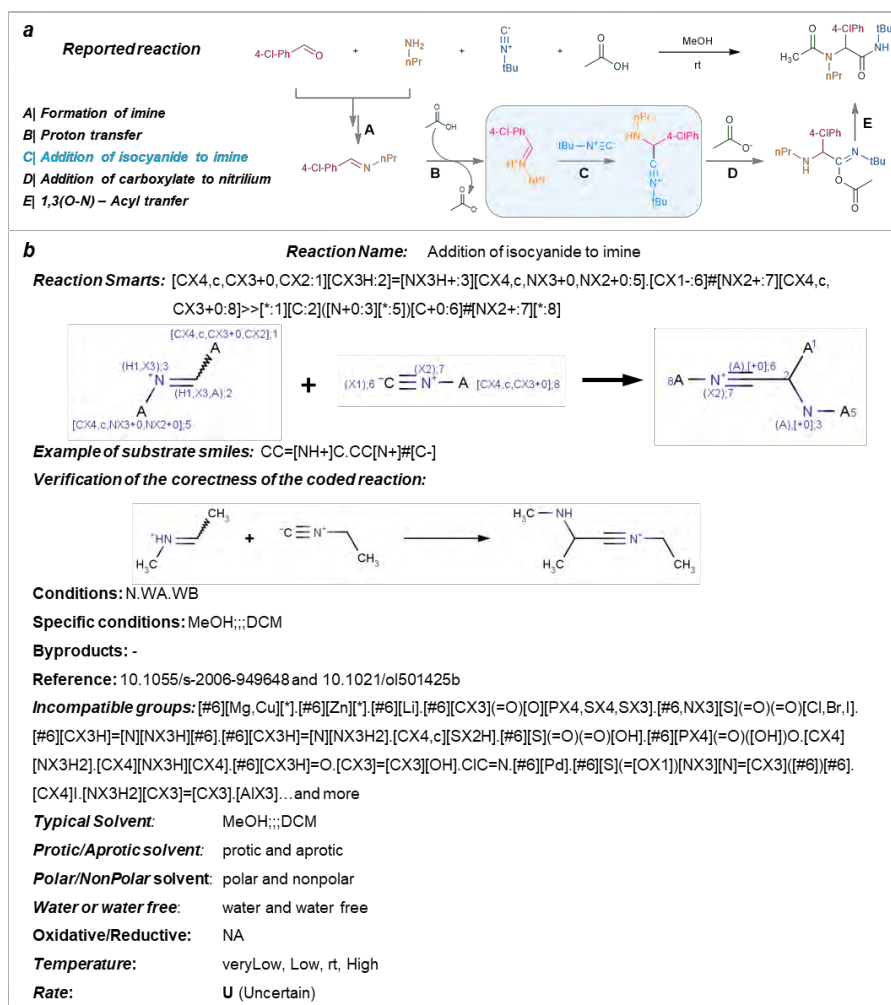
- i) strongly electrophilic groups which may react with the isocyanide nucleophile, e.g., activated esters ('[#6][CX3](=O)[O][PX4,SX4,SX3]'), sulfonyl halides ('[#6,NX3][S](=O)(=O)[Cl,Br,I]'), imidoyl chlorides ('ClC=N') or aldehydes ('[#6][CX3H]=O')



- ii) strongly nucleophilic groups prone to addition to activated imines, e.g., alkyl amines ('[CX4][NH2]' and '[CX4][NX3H][CX4]'), enols ('[CX3]=[CX3][OH]'), enamines ('[CX3]=[CX3][NH2]') or thiols ('[CX4,c][SX2H]');
- iii) groups of reactivity similar to reacting fragments, e.g., hydrazones ('[#6][CX3H]=[N][NX3H][#6]', '[#6][CX3H]=[N][NX3H2]' and '[#6][S](=[OX1])[NX3][N]=[CX3] ([#6])[#6]') and other isocyanides ('[#6][N+]#[C-]');
- iv) strongly basic groups causing deprotonation of iminium cation, e.g., organometallic reagents ('[#6][Mg,Cu][\*]', '[#6][Zn][\*]', '[#6][Li]' and '[AlX3]'). These groups are incompatible also due to their nucleophilicity.
- v) strongly acidic groups (eg., sulfonic and phosphonic acids, '[#6][S](=O)(=O)[OH]' and '[#6][PX4](=O)([OH])O') due to known instability of isocyanides under acidic conditions

Additionally, no palladium-containing species ('[#6][Pd]') should be present in substrates or reaction mixture due to competing formation of complexes and insertion of isocyanides<sup>44</sup>.

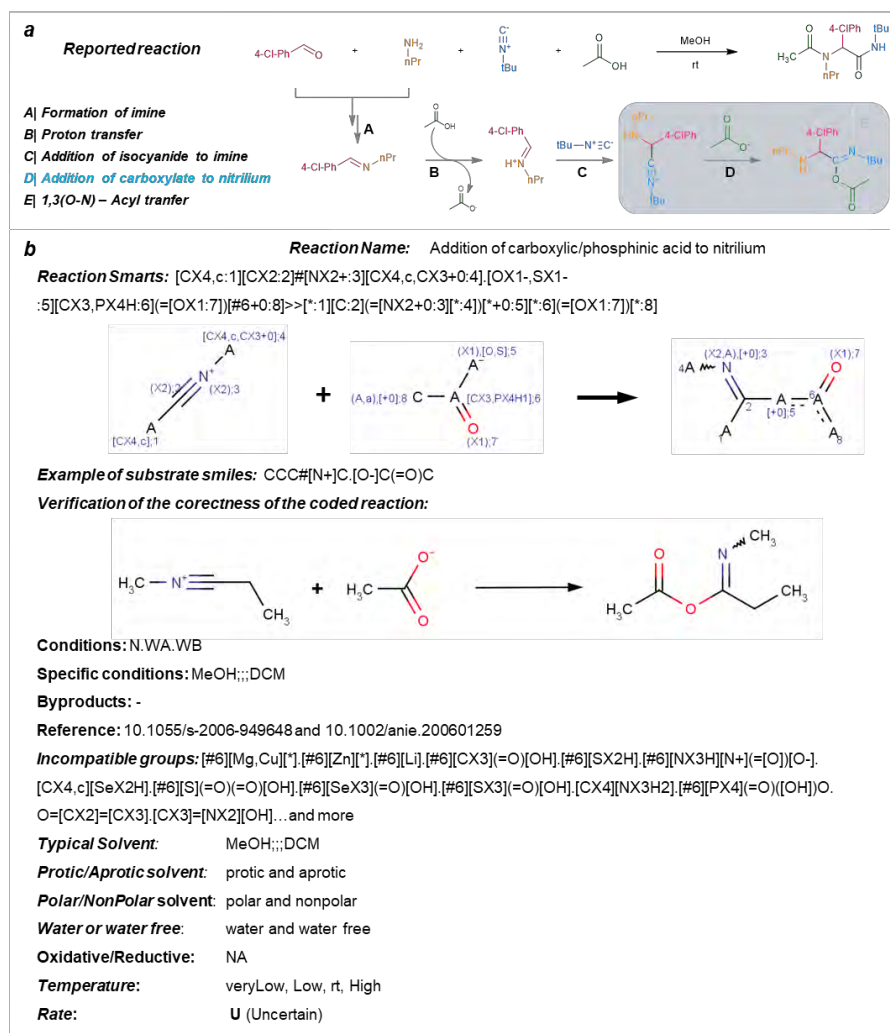
The reaction is then classified according to different criteria important when "wiring-up" sequences from individual mechanistic steps. Addition of isocyanide to protonated imines can proceed under neutral (N), weakly acidic (WA) or weakly basic (WB) condition. Additionally, presence of water is allowed but not required. The reaction can proceed in both polar and nonpolar, protic and aprotic solvents and do not require reductive or oxidative conditions. As the reported reaction rates and temperatures were broadly distributed, the reaction speed is categorized as 'Uncertain' while allowed temperatures range from very low (<-20°C) to high (40 to 150°C).



**Supplementary Figure S30.** Coding of reaction rules covering mechanistic steps. Example of reaction rule (b) for one of the mechanistic steps underlying Ugi four component reaction (a), here: addition of isocyanide to imines.

In the second example (illustrated in **Supplementary Figure S31**), we illustrate how the consecutive step **D** of Ugi four component reaction is coded. In this example, the coded reaction rule covers the addition of deprotonated acid (here, carboxylic or phosphonic) to a nitrilium cation yielding the acyl or phosphonic imidate. First, we define the nitrilium cation in SMARTS notation with atoms #1 (substituent attached to nitrilium's carbon), #2 (nitrilium carbon), #3 (nitrilium nitrogen) and #4 (substituent attached to nitrilium's nitrogen). The substituent attached to nitrilium's carbon mapped as #1 is written as [CX4,c:1] and allows for the presence of either sp<sup>3</sup>-alkyl ('CX4') or aromatic ('c') carbon atom. Atom #2 coded as 'CX2' denotes sp-hybridized carbon atom, atom #3 coded as NX2+ denotes sp-hybridized,

monocharged nitrogen atom while ‘#’ sign between #2 and #3 denotes triple bond between these two atoms. Finally, the substituent attached to nitrilium’s nitrogen, mapped as #5, is allowed to be any sp<sup>3</sup>, sp<sup>2</sup> or aromatic carbon atom, which in SMARTS is written as [CX4,c,CX3+0:4].



**Supplementary Figure S31.** Coding of reaction rules covering mechanistic steps. Example of reaction rule (b) for one of the mechanistic steps underlying Ugi four component reaction, here: ‘Addition of carboxylic/phosphinic acid to nitrilium’.

When the nitrilium cation is defined, we move to the definition of acid nucleophile attacking the nitrilium cation. First, we define the atom attaching to nitrilium cation (#5) as [OX1-,SX1-:5]. This notation represents the deprotonated oxygen and sulfur atoms bound to single (‘X1’) substituent. The nearby atoms, mapped as #6 ([CX3,PX4H:6]) and its neighbouring oxygen atom #7 bound to #6 via double bond (=[OX1:7]) represent the phosphinic acid or carbonyl groups of the (thio)acid nucleophile. We note that SMARTS

notation does not allow for the specification of delocalized bonds of carboxylate anions and this functional group is always coded as -C(=O)[O-] with one single and one double bond between carbon and oxygen atoms. Finally, the last defined atom, [#6+0:8] defines the allowed type of substituent attached to the carboxylic (thio)acid or phosphinic acid and allows for the presence of any type of uncharged carbon ('#6', 6 is atomic number of carbon here) atom – the presence of heteroatoms is not allowed at this position because such deprotonated carbamates or carbonates are extremely prone to decarboxylation.

For proper handling of reaction transform in Rdkit package, all atoms which lost their charges (here, #3 and #5) have their neutral charges explicitly specified in the SMARTS string ('[NX2+0:3]' and '[\*+0:5]') of reaction's product (right side of reaction arrow). This reaction transform do not generate any byproducts so the right side of the reaction (product side) contains only one SMARTS defining the main reaction product.

At this point it is worthwhile to compare the generality of the reaction core of this mechanistic step and the reaction core of the preceding step. The reaction core defined for the "Addition of isocyanide to an imine" comprised seven atoms to define the formed nitrilium cation and contained atoms derived from the imine electrophile (#1, #2, #3, #5). The reaction core covering addition of acids to nitrilium cations does not require the presence of an amine part and comprises only four atoms to define the nitrilium cation. Thus, the latter reaction core is more general, i.e., would be applicable for nitrilium cations obtained via other methods. We note that, in fact, this reaction transform represents also a key step of Passerini three component reaction.

The list of incompatible groups accompanying this reaction transform comprises:

- i) strongly nucleophilic groups prone to addition to iminium cation, e.g., organometallic reagents ('[#6][Mg,Cu][\*]', '[#6][Zn][\*]', '[#6][Li]')
- ii) strongly acidic groups which may neutralise carboxylate/phosphinate anions and/or attack the nitrilium cation, e.g. thiols and selenides ('[#6][SX2H]' and '[CX4,c][SeX2H]'), sulfonic, sulfinic and selenic acids ('[#6][S](=O)(=O)[OH]', '[#6][SeX3](=O)[OH]' and '[#6][SX3](=O)[OH]'), *N*-nitroamines ('[#6][NX3H][N+](=[O])[O-]'), OH-oximes ('[CX3]=[NX2][OH]'), alkyl amines ('[CX4][NX3H2]')
- iii) strongly electrophilic groups prone to addition of of carboxylate anion, e.g., ketenes (O=[CX2]=[CX3]).

The reaction is then classified according to different criteria important when “wiring-up” sequences from individual mechanistic steps. Addition of carboxylate to nitrilium cation can proceed under neutral (N), weakly acidic (WA) or weakly basic (WB) conditions. Additionally, presence of water is allowed but not required (*water* and *water free*). The reaction can proceed in both polar and nonpolar, protic and aprotic solvents and do not require reductive or oxidative conditions. As the reported reaction rates and temperatures were broadly distributed, the reaction speed is categorized as ‘Uncertain’ while allowed temperatures range from very low (<-20°C) to high (40 to 150°C).

In the third example, we switch gears to analyze the coding of one of the mechanistic steps underlying the Fischer indole synthesis with one of the reported examples from ref. <sup>45</sup> shown in **Supplementary Figure S32a** (*green*). This reaction comprises acid-catalyzed reaction of an aldehyde or a ketone with an arylhydrazine to give the indole product. The key mechanistic steps (**Supplementary Figure S32a**, *blue*; several proton shifts are not shown for clarity) underlying this process<sup>46</sup> comprise formation of hydrazone (**A**), isomerization to enamine (**B**), [3,3]-sigmatropic rearrangement (**C**), aromatization (**D**), addition of an amine to iminium cation (**E**), protonation (**F**) and elimination(**G**). The reaction rule covering the [3,3]-sigmatropic rearrangement (Step **C** in **Supplementary Figure S32a**) is illustrated in **Supplementary Figure S32b**.

We begin the coding of the reaction transform with the definition of enehydrazine part with atoms #7, #6, #5 and #4. The first two atoms are limited to sp<sup>2</sup>-hybridized, unsubstituted carbons which is written as [CX3H2:7] and [CX3H:6] in SMARTS notation. We note that for substituted enehydrazines, separate reaction lines with additional defined substituents attached to #6 and #7 are required. The latter two atoms define the remaining 1,2-substituted, protonated hydrazine part of enehydrazine which is written as [NX3H:4][NX4H2+:5].

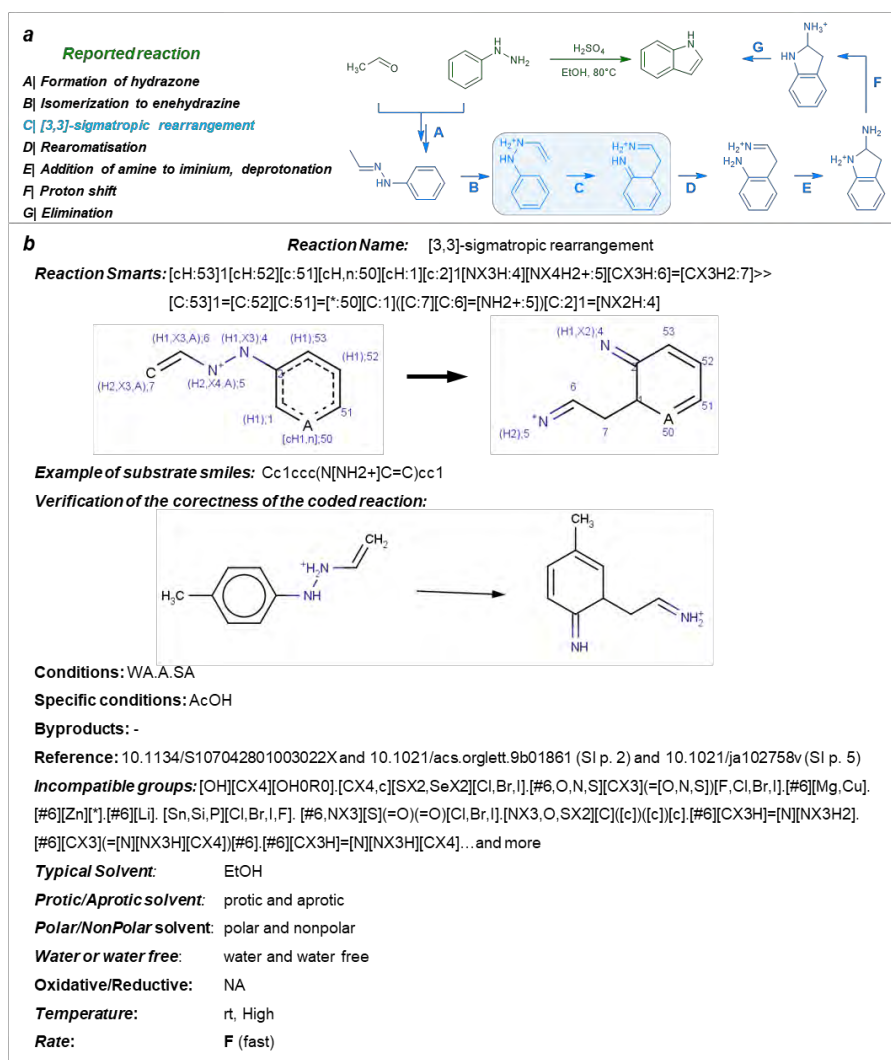
To define an aryl ring participating in [3,3]-sigmatropic rearrangement, the entire six-membered ring is coded with atoms #1,#2,#50-#53. Such coding is necessary to ensure the proposed reaction outcome would not suffer from the formation of regioisomeric mixture of products. Accordingly, both ortho atoms #1 and #53 are coded as ‘cH’ (unsubstituted aromatic carbon atoms), *para*-atom #51 is coded as ‘c’ (substitution is allowed, no additional constraints regarding electronic character of substituent are imposed because entire acid-catalyzed Fischer indolisation was evidenced to proceed equally well for both electron-rich and electron-poor arenes<sup>47</sup>). Finally, *meta* atom #52 is encoded as unsubstituted carbon, ‘cH’, while second *meta* atom #50 allows for the presence of unsubstituted carbon or nitrogen atoms – such combination

allows only for i) symmetric substrates (for which formation of regioisomers is not possible) and ii) rearrangement of 3-hydrazinopyridine derivatives for which selectivity is well evidenced<sup>48,49</sup>. This reaction transform do not generate any byproducts so the right side of the reaction (product side) contains only one SMARTS defining the main reaction product.

The list of incompatible groups accompanying this reaction transform comprises:

- i) highly basic groups prone to reaction with protonated hydrazine, e.g., organometallic reagents ('[#6][Mg,Cu]', '[#6][Zn][\*]', and '[#6][Li]')
- ii) highly electrophilic groups prone to reaction with uncharged nitrogen atom of hydrazine, e.g., acyl, sulfenyl, selenyl, stannyl, silyl or phosphoryl halides ('[CX4,c][SX2,SeX2][Cl,Br,I]', '[#6,O,N,S][CX3](=[O,N,S])[F,Cl,Br,I]', '[Sn,Si,P][Cl,Br,I,F]', '[#6,NX3][S](=O)(=O)[Cl,Br,I]')
- iii) acid labile groups, e.g., acyclic hemiacetals ('[OH][CX4][OH0R0]') or Trt-protected heteroatoms ('[NX3,O,SX2][C]([c])([c])[c]')
- iv) other hydrazones which may interfere with intended reaction outcome due to unfavourable proton-transfer equilibria with protonated enehydrazine ('[#6][CX3H]=[N][NX3H2]', '[#6][CX3](=[N][NX3H][CX4])[#6]', '[#6][CX3H]=[N][NX3H][CX4]')

Reaction is then classified according to different criteria important for “wiring-up” sequences from individual mechanistic steps. This sigmatropic rearrangement can proceed under weakly acidic (WA), acidic (A) or strongly acidic (SA). Additionally, presence of water is allowed but not required (*water* and *water free*). The reaction can proceed in both polar and nonpolar, protic and aprotic solvents and do not require reductive or oxidative conditions. The reaction speed is categorized as ‘fast’ while admissible temperatures are limited to room temperature to high (40 to 150°C).



**Supplementary Figure S32.** Coding of reaction rules covering mechanistic steps. Example of reaction rule (b) for one of the mechanistic steps underlying Fischer indole synthesis (a), here: [3,3]-sigmatropic rearrangement.

In the fourth example, we discuss coding a reaction rule (illustrated in **Supplementary Figure S33**) of  $S_N2$  reaction between phenolate anion and activated secondary alkyl halide. The phenolate coupling partner is coded with atoms mapped #6 and #5. Atom #6, written as 'c' in SMARTS notation denotes aromatic carbon atom. Atom #5, written as [OX1-:5] denotes negatively charged and singularly bound oxygen atom. We then consider alkyl chloride. The secondary carbon atom at which substitution takes place and the attached chloride leaving group are written as [CX4H:2]([Cl:3]) where CX4H denotes  $sp^3$ -hybridized carbon atom with one hydrogen substituent. The other two substituents attached to atom #2 and defined at atoms mapped #1 and #4 need careful specification to ensure lack of steric hindrance around reacting center and presence of activating group(s). The second condition is fulfilled by specifying atom

#4 as [CX3,CX2,c,n:4] representing aryl ('c'), alkynyl and cyano ('CX2'), *N*-bound heterocycle ('n') or sp<sup>2</sup>-hybridized (e.g., alkenyl or acyl) activating substituents. To ensure low-to-moderate steric hindrance around reacting center, the second substituent, #1, is limited to sp<sup>3</sup>-carbon atom with three ('CX4H3') or two ('CX4H2') hydrogen atoms, aryl ('c'), sp<sup>2</sup>-hybridized substituent (CX3, eg., alkenyl or acyl), ethers ('OX2H0') or thioethers('SX2H0').

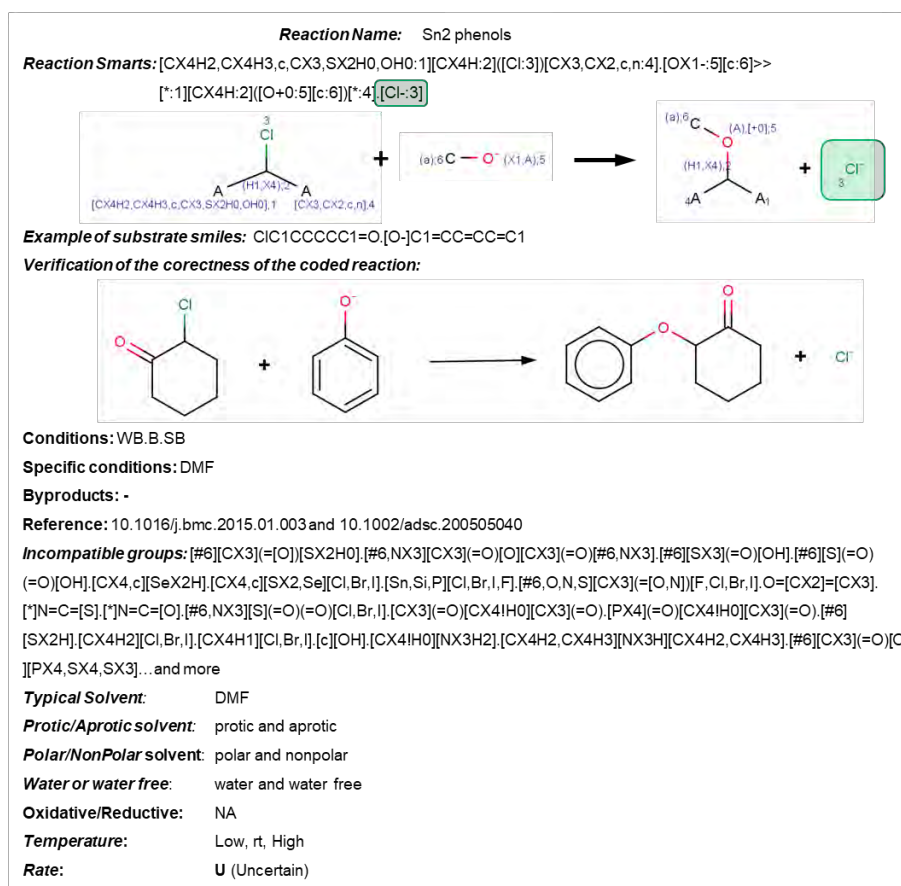
The byproduct generated in this reaction – chloride anion – is included in the SMARTS defining the reaction transform and coded within the right side of the reaction (product side) as [Cl:-3] (*green highlight in Supplementary Figure S33*).

The set of incompatible groups assigned to this reaction rule contains:

- i) highly nucleophilic groups prone to substitution of alkyl halide and potentially competing with the phenolate nucleophile, e.g., thioacids and thiols ('[#6][SX2H]'), selenides ('[CX4,c][SeX2H]'), primary ('[CX4!H0][NX3H2]') and secondary ('[CX4H2,CX4H3][NX3H][CX4H2,CX4H3]') unhindered alkyl amines and other phenols ('[c][OH]')
- ii) highly electrophilic groups potentially competitive with alkyl chloride electrophile, e.g., thioesters ([#6][CX3](=[O])[SX2H0]), anhydrides ('[#6,NX3][CX3](=O)[O][CX3](=O)[#6,NX3]'), sulfenyl, selenyl, stannyl, silyl, phosphoryl and acyl halides ('[CX4,c][SX2,Se][Cl,Br,I]', '[Sn,Si,P][Cl,Br,I,F]', '[#6,O,N,S][CX3](=[O,N])[F,Cl,Br,I]'), ketenes ('O=[CX2]=[CX3]'), iso(thio)cyanates ('[\*]N=C=[S]', '[\*]N=C=[O]'), primary and secondary alkyl chlorides/bromides and iodides ('[CX4H2][Cl,Br,I]', '[CX4H1][Cl,Br,I]'), activated esters ('[#6][CX3](=O)[O][PX4,SX4,SX3]')
- iii) highly acidic groups prone to protonation of phenoxide nucleophile, e.g., sulfinic and sulfonic acids ('[#6][SX3](=O)[OH]', '[#6][S](=O)(=O)[OH]'), *cH*-acids activated with at least two electron withdrawing groups (eg, ketoesters and malonates ('[CX3](=O)[CX4!H0][CX3](=O)') or phosphonoesters ('[PX4](=O)[CX4!H0][CX3](=O)'))

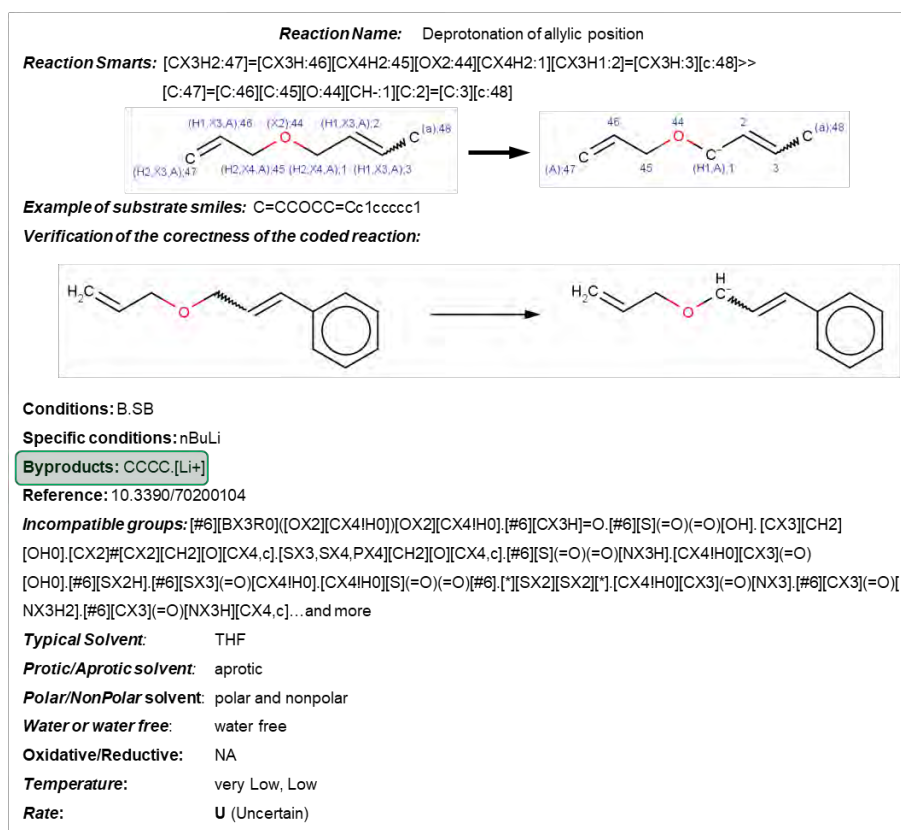
This S<sub>N</sub>2 substitution can proceed under weakly basic (WA), basic (A) or strongly basic (SA) conditions. Additionally, presence of water is allowed but not required (*water* and *water free*). The reaction can proceed in both polar and nonpolar, protic and aprotic solvents and do not require reductive or oxidative conditions. The reaction speed is categorized as 'Uncertain' while range of admissible temperatures is limited to low (-20 °C) to high (40 to 150°C).





**Figure S33.** Coding of reaction rules covering mechanistic steps. Example of reaction rule covering S<sub>N</sub>2 reaction between phenolate and activated secondary alkyl chloride.

In the last example, we discuss coding of the reaction rule (illustrated in **Supplementary Figure S34**) covering deprotonation of an allylic position. In this particular case, the unsymmetrical bisallylic ether is selectively deprotonated to give styryl stabilized carbanion. The coding of the reaction transform begins with defining the bisallylic ether with oxygen atom #44, neighbouring sp<sup>3</sup>-alkyls with two hydrogens ('CX4H2') atoms #1 and #45 and alkenyl substituents comprising #46,#47 and #2,#3 atom pairs. Then, we specify the sets of admissible substituents allowing for selective deprotonation of one of the allylic positions present in the reaction core. To do so, we limit the #46-47 alkenyl to be unsubstituted with three hydrogen atoms attached (-[CX3H:46]=[CX3H2:47]). The stabilizing aryl substituent ('c') at atom #48 is attached to the second alkenyl fragment. Additionally, no other substituents are allowed at positions #2 and #3: these atoms are coded to have one hydrogen atom attached each which is written as ('CX3H1') in SMARTS string.



**Supplementary Figure S34.** Coding of reaction rules covering mechanistic steps. Example of reaction rule covering regioselective deprotonation of bisallylic ether.

The set of incompatible groups accompanying this reaction rule spans:

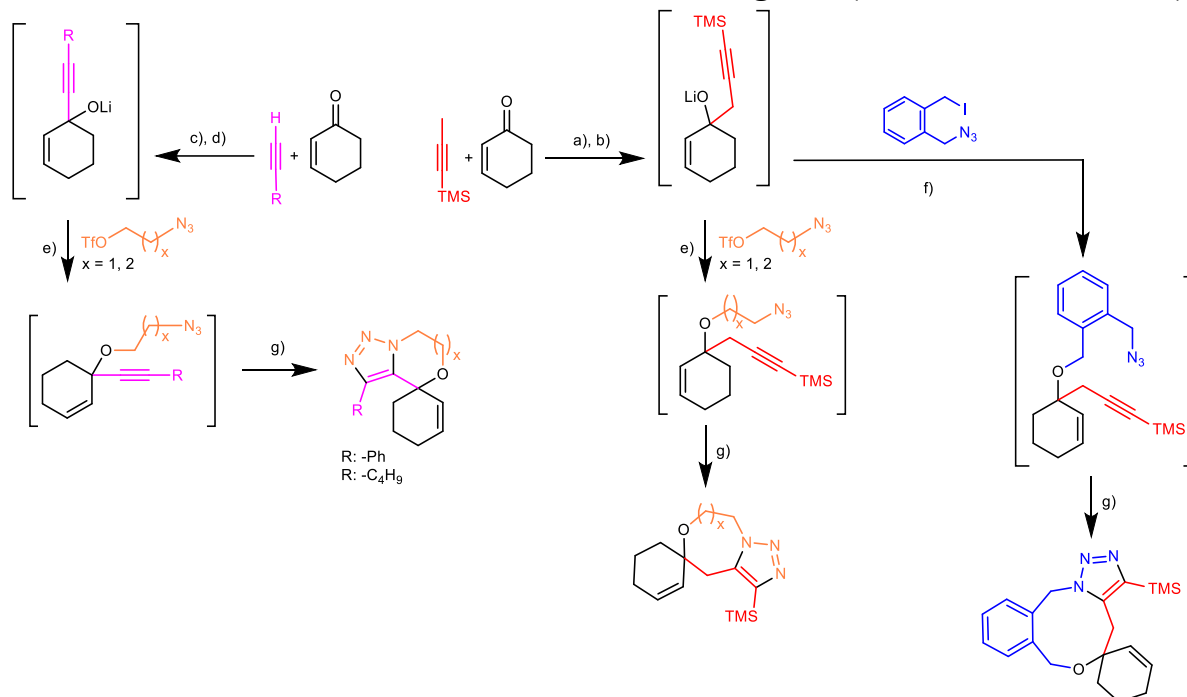
- i) electrophilic groups which may react with BuLi deprotonating agent, e.g., aldehydes ( $[#6][CX3H]=O$ ), disulfides ( $[*][SX2][SX2][*]$ ), boronic esters ( $[#6][BX3R0]([OX2][CX4!H0])[OX2][CX4!H0]$ )
- ii) acidic groups prone to quenching of BuLi deprotonating agent or formed carbanion, e.g., amides ( $[#6][CX3](=O)[NX3H2]$ ,  $[#6][CX3](=O)[NX3H][CX4,c]$ ) and sulfonamides ( $[#6][S](=O)(=O)[NX3H]$ ), enolizable esters ( $[CX4!H0][CX3](=O)[OH0]$ ), sulfoxides, sulfones or amides with at least one acidic hydrogen ( $[#6][SX3](=O)[CX4!H0]$ ,  $[CX4!H0][S](=O)(=O)[#6]$ ,  $[CX4!H0][CX3](=O)[NX3]$ ), sulfonic acids ( $[#6][S](=O)(=O)[OH]$ ), thiols ( $[#6][SX2H]$ );
- iii) acidified allylic or propargylic positions, e.g., allyl ethers or propargyl ethers possessing competitive deprotonation sites, ( $[CX3][CH2][OH0]$ ,  $[CX2]#[CX2][CH2][O][CX4,c]$ )

Finally, reaction is then classified according to different criteria important when “wiring-up” sequences from individual mechanistic steps. This deprotonation of weakly acidic position can proceed under basic (B) or strongly basic (SB) conditions and do not require reductive or oxidative conditions. Additionally, presence of water is not allowed. The reaction can proceed in both polar and nonpolar but only aprotic solvents. The reaction speed is categorized as ‘Uncertain’ while range of admissible temperatures is limited to very Low (<-20 °C) to low (-20 °C). Additionally, for this reaction the formed byproducts (butane and Li+) generated from the reagent used (BuLi) are included in the reaction rule (*green* highlight in **Supplementary Figure S34**).

## Section S6. Experimental section

**General information.** All starting materials and reagents were obtained from commercial sources and used as received unless otherwise noted. All solvents used were freshly distilled prior to use.  $^1\text{H}$  NMR spectra were recorded at 400, 500 or 600 MHz and  $^{13}\text{C}$  NMR spectra were recorded at 100, 125 or 150 MHz with complete proton decoupling. Chemical shifts are given in  $\delta$  relative to the residual signals of the deuterated solvents. High-resolution mass spectra were acquired using electron ionization (EI) or electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) mode with a time-of-flight detector. Infrared (IR) spectra were recorded on a Fourier transform infrared (FT-IR) spectrometer as a thin film on a NaCl plate (film). HPLC analysis were performed on a HPLC system equipped with chiral stationary phase columns, detection at 254 nm. Optical rotations were measured at room temperature with a polarimeter. TLC was performed with aluminum plates coated with 60 F254 silica gel. Plates were visualized with UV light (254 nm) and by treatment with ethanolic *p*-anisaldehyde with sulfuric and glacial acetic acid followed by heating, aqueous cerium(IV) sulfate solution with molybdic and sulfuric acid followed by heating, or aqueous potassium permanganate with sodium hydroxide and potassium carbonate solution followed by heating or ethanolic vanillin with sulfuric acid followed by heating. Reaction products were purified by flash chromatography using silica gel 60 (230-400 mesh). Microwave experiments were conducted in a mono-mode cavity with a microwave power delivery system ranging from 0 to 850 W, allowing pressurized reactions (0–30 bar) to be carried out in sealed glass vials (4–30 mL) equipped with a snap cap and a silicon septum.

### Section S6.1 Reaction Mach1 described in main-text Figure 3 (variant without HMPA).

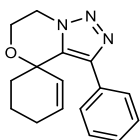


**Supplementary Figure S35. Reaction Mach1 described in main-text Figure 3 (variant without HMPA).** Reagents and conditions: (a) *n*BuLi (1.6 M), THF, -20 °C, 15 min; (b) Cyclohexen-1-one, THF, -20 °C, 15 min; (c) *n*BuLi (1.6 M), THF, -78 °C, 30 min; (d) Cyclohexen-1-one, THF, -78 °C to -20 °C, 30 min; (e) -20 °C, 24 h. (f) NaH, 50 °C, 12 h; (g) THF/Toluene, 90 °C to 125 °C, 12-24 h.

#### One-pot procedure

To a mixture of an alkyne (0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -78 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 30 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -78 °C. The reaction was then allowed to warm up to maximum -20 °C over 30 min, followed by addition of the azido-substrate dropwise at -20 °C. The reaction was left to stir at that temperature for 24 h. After that, anhydrous toluene (5 mL) was added over the mixture at rt, followed by heating the latter in an oil bath of 90 °C, the temperature was then raised gradually to 125 °C and the reaction was stirred for 24 h at that temperature.

The crude mixture was then quenched at rt with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with Et<sub>2</sub>O. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography to afford the target product.



### **3'-phenyl-6',7'-dihydrospiro[cyclohexane-1,4'-[1,2,3]triazolo[5,1-c][1,4]oxazin]-2-ene 2a**

Following the general procedure using (2-azidoethyl trifluoromethanesulfonate) as azido substrate and phenylacetylene as alkyne, the crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the final product **2a** as a white solid (0.053 g, 33%).

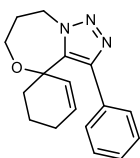
**m.p.** = 160-162 °C;

**<sup>1</sup>H NMR** (600 MHz, CDCl<sub>3</sub>) δ 7.71-7.70 (m, 2H), 7.37-7.35 (m, 2H), 7.33-7.30 (m, 1H), 6.17-6.14 (m, 1H), 6.04-6.01 (m, 1H), 4.48-4.46 (m, 2H), 4.22-4.18 (m, 1H), 4.07-4.03 (m, 1H), 2.21-2.18 (m, 1H), 1.98-1.92 (m, 1H), 1.91-1.88 (m, 1H), 1.87-1.83 (m, 2H), 1.59-1.57 (m, 1H);

**<sup>13</sup>C NMR** (151 MHz, CDCl<sub>3</sub>) δ 142.5, 133.6, 133.4, 131.5, 128.4, 128.2, 128.0, 127.8, 71.7, 58.1, 46.2, 31.2, 24.7, 17.4;

**HRMS** (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. 268.1450; Found 268.1451;

**IR** (film, CDCl<sub>3</sub>) 3016, 2939, 2817, 1948, 1882, 1641, 1358, 1172, 1072, 916, 698 cm<sup>-1</sup>.



### **3'-phenyl-7',8'-dihydro-6'H-spiro[cyclohexane-1,4'-[1,2,3]triazolo[5,1-c][1,4]oxazepin]-2-ene 2b**

To a mixture of phenylacetylene (0.0735 g, 0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -78 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 30 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv), and 1,8-bis(dimethylamino)naphthalene (0.129 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -78 °C. The reaction was then allowed to warm up to maximum -20 °C over 30 min, followed by the dropwise-addition of 3-azidopropyl trifluoromethanesulfonate (0.182 g, 0.78 mmol, 1.3 equiv) at -20 °C. The reaction was left to stir at that temperature for 24 h. After that, 1,4-dioxane (5 mL) was added over the mixture at rt, followed by heating the latter in an oil bath of 90 °C, the temperature was then raised gradually to 125 °C and the reaction was stirred for 72 h at that temperature.

The crude mixture was then quenched at rt with sat. aq. solution of  $\text{NH}_4\text{Cl}$ , and extracted with  $\text{Et}_2\text{O}$ . The combined organic phase was dried over  $\text{MgSO}_4$  and concentrated to dryness. The crude mixture was purified by column chromatography (hexane/ $\text{EtOAc}$  8:2) to afford the final product as white solid **2b** (0.064 g, 38%).

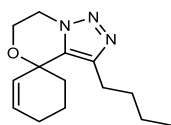
**m.p.** = 119-121 °C;

**$^1\text{H}$  NMR** (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.50-7.48 (dd,  $J$  = 1.92, 7.32 Hz, 2H), 7.37-7.33 (m, 3H), 6.04-6.00 (m, 1H), 5.85-5.83 (d,  $J$  = 10.01 Hz, 1H), 4.85-4.80 (m, 1H), 4.72-4.68 (m, 1H), 3.98-3.94 (m, 1H), 3.83-3.78 (m, 1H), 2.21-2.17 (m, 2H), 2.14-2.09 (m, 1H), 1.92-1.89 (m, 1H), 1.87-1.83 (m, 2H), 1.82-1.77 (m, 1H), 1.54-1.51 (m, 1H);

**$^{13}\text{C}$  NMR** (126 MHz,  $\text{CDCl}_3$ )  $\delta$  145.0, 137.5, 133.3, 132.4, 129.5, 127.9, 127.6, 73.3, 61.1, 47.4, 32.3, 28.2, 24.8, 17.5;

**HRMS** (ESI-HR)  $m/z$ :  $[\text{M}]^+$  Calc. 282.1606; Found 282.1609;

**IR** (film,  $\text{CDCl}_3$ ) 3028, 2935, 2827, 1955, 1894, 1486, 1090, 700  $\text{cm}^{-1}$ .



**3'-butyl-6',7'-dihydrospiro[cyclohexane-1,4'-[1,2,3]triazolo[5,1-c][1,4]oxazin]-2-ene 2c**

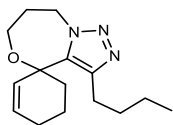
Following the general procedure using (2-azidoethyl trifluoromethanesulfonate) as azido substrate and hex-1-yne as alkyne, the crude mixture was purified by column chromatography (hexane/ $\text{EtOAc}$  8:2) to afford the final product **2c** as a colorless liquid (0.066 g, 44%).

**$^1\text{H}$  NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.13-6.12 (m, 1H), 5.63-5.61 (d,  $J$  = 9.93 Hz, 1H), 4.40-4.32 (m, 2H), 4.10-4.07 (m, 2H), 2.53-2.51 (t,  $J$  = 7.80 Hz, 2H), 2.28-2.07 (m, 3H), 1.93-1.84 (m, 1H), 1.76-1.68 (m, 4H), 1.40-1.36 (m, 2H), 0.95-0.91 (t,  $J$  = 7.35 Hz, 3H);

**$^{13}\text{C}$  NMR** (151 MHz,  $\text{CDCl}_3$ )  $\delta$  142.3, 133.3, 132.8, 127.1, 71.2, 58.4, 45.7, 32.4, 31.4, 30.2, 25.0, 22.5, 17.6, 13.7;

**HRMS** (ESI-HR)  $m/z$ :  $[\text{M}]^+$  Calc. 248.1763; Found 248.1765;

**IR** (film,  $\text{CDCl}_3$ ) 2953, 2870, 1555, 1440, 1354, 1169, 1074, 939, 800, 727  $\text{cm}^{-1}$ .



**3'-butyl-7',8'-dihydro-6'H-spiro[cyclohexane-1,4'-[1,2,3]triazolo[5,1-c][1,4]oxazepin]-2-ene 2d**

Following the general procedure using (3-azidopropyl trifluoromethanesulfonate) as azido substrate and hex-1-yne as alkyne, the crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the final product **2d** as a colorless liquid (0.053 g, 34%).

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 6.08-6.05 (m, 1H), 5.80-5.78 (d, *J* = 10.02 Hz, 1H), 4.75-4.70 (m, 1H), 4.59-4.53 (m, 1H), 3.94-3.90 (m, 1H), 3.78-3.73 (m, 1H), 2.58-2.54 (t, *J* = 7.39 Hz, 2H), 2.20-2.16 (m, 1H), 2.11-2.06 (m, 3H), 2.03-2.00 (m, 1H), 1.86-1.81 (m, 2H), 1.67-1.59 (m, 3H), 1.39-1.31 (m, 2H), 0.92-0.88 (t, *J* = 7.35 Hz, 3H);

**<sup>13</sup>C NMR** (151 MHz, CDCl<sub>3</sub>) 145.0, 135.9, 132.2, 127.3, 73.2, 60.8, 46.7, 34.5, 31.8, 28.0, 25.5, 24.9, 22.5, 17.8, 13.7;

**HRMS** (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. 262.1919; Found 262.1922;

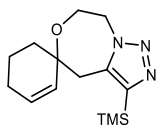
**IR** (film, CDCl<sub>3</sub>) 2953, 2870, 1543, 1435, 1350, 1235, 1170, 1094, 1027, 921, 733 cm<sup>-1</sup>.

**General procedure of the cascade 1,2-addition-alkylation of (2-azidoethyl trifluoromethanesulfonate) and 1,2-addition-alkylation of (3-azidopropyl trifluoromethanesulfonate) :**

To a mixture of 1-(trimethylsilyl)propyne (0.080 g, 0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -20 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -20 °C. Later on, the azido-substrate was added dropwise at that temperature, and the reaction mixture was left to stir at -20 °C for 24 h. After that, anhydrous toluene (5 mL) was added over the mixture at rt, followed by heating the latter in an oil bath of 90 °C, the temperature was then raised gradually to 125 °C and the reaction was stirred for 24 h at that temperature.

The crude mixture was then quenched at rt with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with Et<sub>2</sub>O. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography to afford the target product.





**3'-(trimethylsilyl)-7',8'-dihydro-4'*H*-spiro[cyclohexane-1,5'-[1,2,3]triazolo[1,5-*d*][1,4]oxazepin]-2-ene 2e**

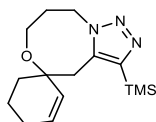
Following the same procedure using (2-azidoethyl trifluoromethanesulfonate) as azido substrate, 1,8-bis(dimethylamino)naphthalene (0.129 g, 0.6 mmol, 1 equiv) was also added to the mixture of the cyclohexenone in THF. The cyclization was performed in THF/1,4-dioxane (5:5mL). The crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the final product **2e** as a colorless liquid (0.02 g, 12%).

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 5.95-5.91 (m, 1H), 5.51-5.48 (dd, *J* = 1.08, 10.20 Hz, 1H), 4.74-4.69 (m, 1H), 4.62-4.57 (m, 1H), 4.01-3.89 (m, 2H), 3.13-3.03 (dd, *J* = 15.54 Hz, 2H), 2.16-2.10 (m, 1H), 2.07-2.01 (m, 1H), 1.91-1.86 (m, 1H), 1.84-1.76 (m, 1H), 1.66-1.62 (m, 1H), 1.55-1.49 (m, 1H), 0.33 (s, 9H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 145.5, 139.6, 132.6, 127.1, 72.4, 61.1, 52.3, 36.4, 33.9, 25.5, 18.2, 0.0;

**HRMS** (EI-HR) *m/z*: [M]<sup>+</sup> Calc. 277.1610; Found 277.1607;

**IR** (film, CDCl<sub>3</sub>) 2951, 2833, 1645, 1465, 1249, 1055, 840, 759 cm<sup>-1</sup>.



**3'-(trimethylsilyl)-8',9'-dihydro-4'*H*,7'*H*-spiro[cyclohexane-1,5'-[1,2,3]triazolo[5,1-*d*][1,5]oxazocin]-2-ene 2f**

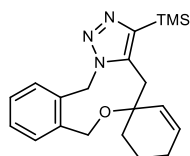
Following the general procedure using (3-azidopropyl trifluoromethanesulfonate) as azido substrate, the crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the final product **2f** as a colorless liquid (0.059 g, 34%).

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 5.93-5.90 (m, 1H), 5.59-5.57 (m, *J* = 10.19 Hz, 1H), 4.67-4.63 (m, 1H), 4.40-4.36 (m, 1H), 3.54-3.50 (m, 1H), 3.37-3.32 (m, 1H), 2.93-2.84 (dd, *J* = 14.91 Hz, 2H), 2.14-2.08 (m, 1H), 2.04-1.98 (m, 3H), 1.91-1.86 (m, 1H), 1.82-1.76 (m, 1H), 1.66-1.56 (m, 2H), 0.32 (s, 9H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 144.2, 138.5, 132.0, 127.9, 73.9, 59.8, 45.5, 34.7, 31.8, 30.3, 25.5, 18.5, 0.0;

**HRMS** (ESI-HR)  $m/z$ :  $[M]^+$  Calc. 292.1845; Found 292.1850;

**IR** (film,  $CDCl_3$ ) 2951, 2834, 1645, 1453, 1248, 1094, 842, 735  $cm^{-1}$ .



**3-(trimethylsilyl)-7,12-dihydro-4H-spiro[benzo[g][1,2,3]triazolo[5,1-d][1,5]oxazonine-5,1'-cyclohexan]-2'-ene 2g**

Following the general procedure using 1-(azidomethyl)-2-(iodomethyl)benzene as azido substrate, the crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the final product 2g as a white solid (0.048 g, 23%).

**m.p.** = 175-177 °C;

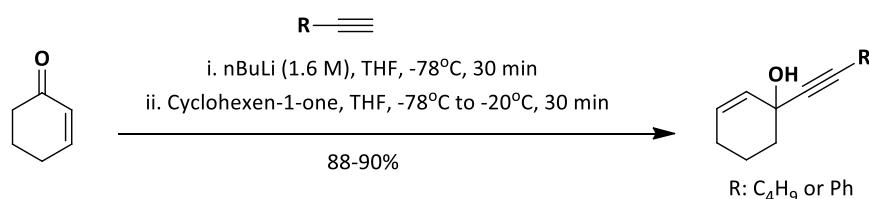
**$^1H$  NMR** (400 MHz,  $CDCl_3$ )  $\delta$  7.51-7.50 (d,  $J$  = 7.15 Hz, 1H), 7.37-7.31 (m, 2H), 7.17-7.15 (d,  $J$  = 7.16 Hz, 1H), 6.06-6.03 (m, 1H), 5.95-5.91 (d,  $J$  = 13.97 Hz, 1H), 5.63-5.59 (m, 2H), 4.54-4.51 (d,  $J$  = 12.65 Hz, 1H), 4.38-4.36 (m, 1H), 2.85-2.76 (dd,  $J$  = 15.04 Hz, 2H), 2.19-2.15 (m, 1H), 2.08-2.01 (m, 1H), 1.91-1.79 (m, 2H), 1.72-1.64 (m, 1H), 1.59-1.52 (m, 1H), 0.34 (s, 9H);

**$^{13}C$  NMR** (151 MHz,  $CDCl_3$ )  $\delta$  143.8, 138.5, 138.2, 133.6, 131.5, 128.7 (2C), 128.6, 128.4, 73.2, 66.2, 51.5 (2C), 35.4, 34.5, 25.5, 18.3, 0.3;

**HRMS** (EI-HR)  $m/z$ :  $[M]^+$  Calc. 353.1923; Found 353.1930;

**IR** (film,  $CDCl_3$ ) 2952, 2871, 1442, 1248, 1071, 841, 753  $cm^{-1}$ .

**Mechanistic validation by stepwise isolation**

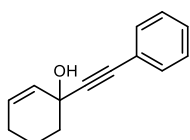


**Supplementary Figure S36.** Reaction scheme of the stepwise process to confirm the first step intermediate of the 1,2-addition-alkylation with terminal alkynes

**General procedure of the cascade 1,2-addition of phenylacetylene and 1,2-addition of 1-hexyne:**

To a mixture of a terminal alkyne (0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -78 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was

stirred for 30 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -78 °C. The reaction was then allowed to warm up to rt over 30 min. The crude mixture was then quenched at rt with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with Et<sub>2</sub>O. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography to afford the first-step product.



### 1-(phenylethynyl)cyclohex-2-en-1-ol S6.1.1

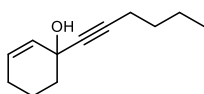
Following the general procedure, phenylacetylene (0.0735 g, 0.72 mmol, 1.2 equiv) was used. The crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the first-step product **S6.1.1** (0.107 g, 90%) as a colorless liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.46-7.44 (m, 2H), 7.31-7.30 (m, 3H), 5.88 (m, 2H), 2.47 (bs, 1H), 2.19-2.13 (m, 1H), 2.09-2.06 (m, 2H), 2.04-2.00 (m, 1H), 1.87-1.81 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 131.70, 130.55, 129.70, 128.26, 128.21, 122.80, 92.78, 83.69, 65.64, 37.97, 24.73, 19.22;

**IR** (film, CDCl<sub>3</sub>) 3342, 2940, 2223, 1489, 1170, 1051, 957, 756, 691 cm<sup>-1</sup>;

**HRMS** (EI-HR) *m/z*: [M]<sup>+</sup> Calc. 197.0966; Found 197.0960.



### 1-(hex-1-yn-1-yl)cyclohex-2-en-1-ol S6.1.2

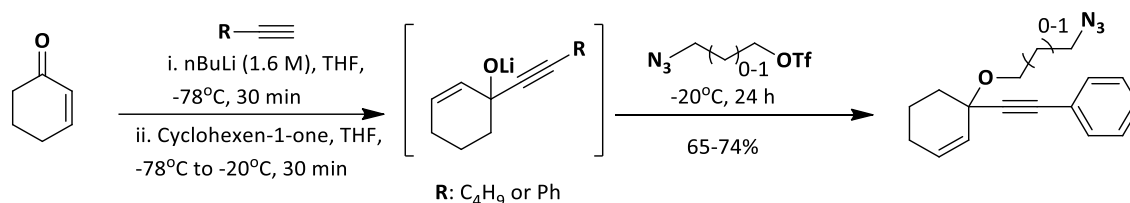
Following the general procedure, 1-hexyne (0.059 g, 0.72 mmol, 1.2 equiv) was used. The crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the first-step product **S6.1.2** (0.095 g, 88%) as a colorless liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 5.83-5.80 (m, 1H), 5.76-5.74 (m, 1H), 2.25-2.21 (t, *J* = 7.03 Hz, 2H), 2.07-1.99 (m, 3H), 1.96 (s, 1H), 1.94-1.88 (m, 1H), 1.79-1.74 (m, 2H), 1.53-1.47 (m, 2H), 1.44-1.39 (m, 2H), 0.94-0.91 (t, *J* = 7.24 Hz, 3H);

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  131.14, 129.09, 84.37, 83.90, 65.26, 38.27, 30.75, 24.69, 21.92, 19.22, 18.43, 13.58;

IR (film,  $\text{CDCl}_3$ ) 3355, 2933, 2869, 2230, 1455, 1182, 1053, 958, 738  $\text{cm}^{-1}$ ;

HRMS (EI-HR)  $m/z$ :  $[\text{M}]^+$  Calc. 177.1279; Found; 177.1283.

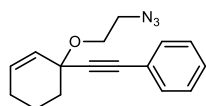


**Supplementary Figure S37.** Reaction scheme of the stepwise process to confirm the second step intermediate of the 1,2-addition-alkylation with terminal alkynes

**General procedure of the cascade 1,2-addition-alkylation of (2-azidoethyl trifluoromethanesulfonate) and 1,2-addition-alkylation of (3-azidopropyl trifluoromethanesulfonate):**

To a mixture of an alkyne (0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at  $-78^\circ\text{C}$ , *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 30 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at  $-78^\circ\text{C}$ . The reaction was then allowed to warm up to maximum  $-20^\circ\text{C}$  over 30 min, followed by addition of the azido-substrate dropwise at  $-20^\circ\text{C}$ . The reaction was left to stir at that temperature for 24 h.

The crude mixture was then quenched at  $0^\circ\text{C}$  with sat. aq. solution of  $\text{NH}_4\text{Cl}$ , and extracted with  $\text{Et}_2\text{O}$ . The combined organic phase was dried over  $\text{MgSO}_4$  and concentrated to dryness. The crude mixture was purified by column chromatography to afford the second-step product.



### **((1-(2-azidoethoxy)cyclohex-2-en-1-yl)ethynyl)benzene S6.1.3**

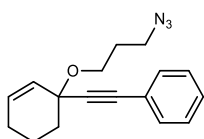
Following the general procedure, phenylacetylene (0.0735 g, 0.72 mmol, 1.2 equiv) and 2-azidoethyl trifluoromethanesulfonate (0.171 g, 0.78 mmol, 1.3 equiv) were used. The crude mixture was purified by column chromatography (hexane/ $\text{EtOAc}$  9:1) to afford the second-step product **S6.1.3** (0.112 g, 70%) as a pale white liquid.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.37-7.35 (m, 2H), 7.24-7.22 (m, 3H), 5.86-5.84 (dt, *J* = 9.97 Hz, 1H), 5.82-5.80 (d, *J* = 10.08 Hz, 1H), 3.81-3.80 (m, 2H), 3.35-3.32 (m, 2H), 2.04-2.01 (m, 1H), 2.00-1.97 (m, *J* = 5.84 Hz, 3H), 1.78-1.77 (m, *J* = Hz, 1H), 1.73-1.71 (m, *J* = Hz, 1H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 131.7, 130.7, 128.3, 128.2, 128.1, 122.7, 90.1, 85.7, 71.0, 62.8, 51.0, 35.0, 24.9, 18.6;

**HRMS** (APCI-HR) *m/z*: [M]<sup>+</sup> Calc. 268.1450; Found 268.1451;

**IR** (film, CDCl<sub>3</sub>) 3032, 2929, 2866, 2105, 1442, 1284, 1079, 757, 691 cm<sup>-1</sup>.



#### **((1-(3-azidopropoxy)cyclohex-2-en-1-yl)ethynyl)benzene S6.1.4**

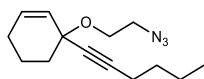
Following the general procedure, phenylacetylene (0.0735 g, 0.72 mmol, 1.2 equiv) and 3-azidopropyl trifluoromethanesulfonate (0.182 g, 0.78 mmol, 1.3 equiv) were used. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.4** (0.125 g, 74%) as a pale white liquid.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.36-7.35 (m, 2H), 7.23-7.22 (m, 3H), 5.84-5.82 (m, 1H), 5.79-5.77 (d, *J* = 9.94 Hz, 1H), 3.69-3.68 (m, *J* = Hz, 2H), 3.36 (t, *J* = 6.76 Hz, 2H), 2.00-1.93 (m, 4H), 1.82-1.79 (quintet, 2H), 1.73-1.70 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 131.7, 130.4, 128.6, 128.2, 128.2, 122.8, 90.5, 85.4, 70.6, 60.2, 48.6, 34.9, 29.7, 24.9, 18.8;

**HRMS** (APCI-HR) *m/z*: [M]<sup>+</sup> Calc. 304.1426; Found 304.1429;

**IR** (film, CDCl<sub>3</sub>) 3032, 2929, 2870, 2095, 1490, 1262, 1081, 756, 691 cm<sup>-1</sup>.



#### **3-(2-azidoethoxy)-3-(hex-1-yn-1-yl)cyclohex-1-ene S6.1.5**

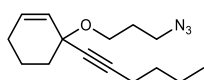
Following the general procedure, 1-hexyne (0.059 g, 0.72 mmol, 1.2 equiv) and 2-azidoethyl trifluoromethanesulfonate (0.171 g, 0.78 mmol, 1.3 equiv) were used. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.5** (0.096 g, 65%) as a colorless liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 5.87-5.85 (m, 1H), 5.80-5.77 (dt, *J* = 10.01 Hz, 1H), 3.83-3.79 (m, 2H), 3.41-3.34 (m, 2H), 2.27-2.23 (t, *J* = 6.99 Hz, 2H), 2.07-2.03 (m, 2H), 1.96-1.93 (m, 2H), 1.83-1.81 (m, 1H), 1.80-1.73 (m, 1H), 1.53-1.50 (m, 2H), 1.45-1.40 (m, 2H), 0.94-0.91 (m, 3H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 130.0, 128.8, 86.3, 81.0, 70.6, 62.4, 51.0, 35.3, 30.8, 24.8, 21.9, 18.6, 18.4, 13.5;

**HRMS** (EI-HR) *m/z*: [M]<sup>+</sup> Calc. 247.1685; Found 247.1686;

**IR** (film, CDCl<sub>3</sub>) 3031, 2933, 2864, 2105, 1456, 1286, 1082, 961, 738 cm<sup>-1</sup>.



### 3-(3-azidopropoxy)-3-(hex-1-yn-1-yl)cyclohex-1-ene S6.1.6

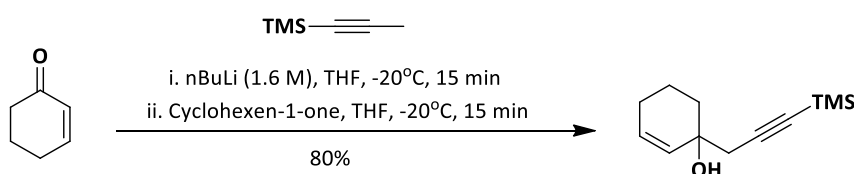
Following the general procedure, 1-hexyne (0.059 g, 0.72 mmol, 1.2 equiv) and 3-azidopropyl trifluoromethanesulfonate (0.182 g, 0.78 mmol, 1.3 equiv) were used. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.6** (0.104 g, 66%) as a colorless liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 5.86-5.84 (m, 1H), 5.78-5.76 (dt, *J* = 9.99 Hz, 1H), 3.71-3.64 (m, 2H), 3.44-3.41 (t, *J* = 6.76 Hz, 2H), 2.27-2.23 (t, *J* = 7.00 Hz, 2H), 2.08-2.02 (m, 2H), 1.93-1.90 (dd, *J* = 5.29, 11.24 Hz, 2H), 1.88-1.84 (dd, *J* = 6.50, 12.40 Hz, 2H), 1.79-1.70 (m, 2H), 1.54-1.50 (m, 2H), 1.46-1.40 (m, 2H), 0.96-0.92 (t, *J* = 7.25 Hz, 3H);

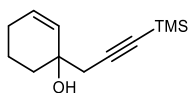
**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 129.7, 129.3, 86.0, 81.4, 70.3, 59.9, 48.7, 35.1, 30.8, 29.7, 24.9, 21.9, 18.8, 18.4, 13.5;

**HRMS** (EI-HR) *m/z*: [M]<sup>+</sup> Calc. 261.1841; Found 261.1832;

**IR** (film, CDCl<sub>3</sub>) 3031, 2932, 2871, 2096, 1456, 1263, 1082, 961, 738 cm<sup>-1</sup>.



**Supplementary Figure S38.** Reaction scheme of the stepwise process to confirm the first step intermediate of the 1,2-addition-alkylation with methylated alkynes



### 1-(3-(trimethylsilyl)prop-2-yn-1-yl)cyclohex-2-en-1-ol **S6.1.7**<sup>50,51</sup>

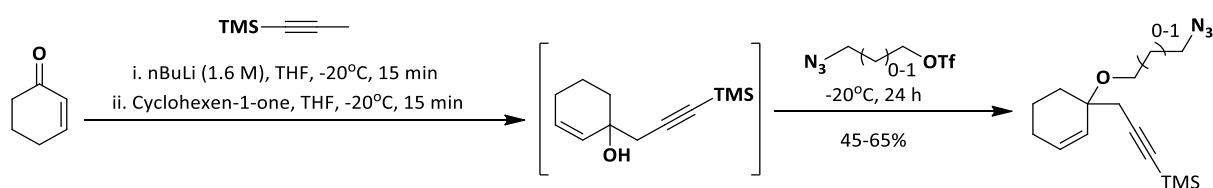
To a mixture of 1-(trimethylsilyl)propyne (0.080 g, 0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -20 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -20 °C. The reaction was then allowed to warm up to rt over 15 min. The crude mixture was then quenched at rt with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with Et<sub>2</sub>O. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography (Et<sub>2</sub>O/pentane 9:1) to afford the first-step product **S6.1.7** as a colorless liquid (0.100 g, 80%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 5.91-5.86 (m, 1H), 5.73-5.71 (d, *J* = 10.04 Hz, 1H), 2.49 (s, 2H), 2.12-2.05 (m, 2H), 2.01-1.93 (m, 1H), 1.86-1.81 (m, 1H), 1.80-1.75 (m, 1H), 1.74-1.65 (m, 1H), 0.19 (s, 9H);

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 130.94, 130.81, 102.70, 88.11, 68.49, 35.32, 34.16, 25.19, 19.13, 0.07;

HRMS (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. 231.1181; Found; 231.1177;

IR (film, CDCl<sub>3</sub>) 3383, 2938, 2175, 1406, 1250, 1030, 843, 760 cm<sup>-1</sup>.



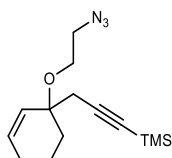
**Supplementary Figure S39.** Reaction scheme of the stepwise process to confirm the second step intermediate of the 1,2-addition-alkylation with methylated alkynes and alkyl triflates

**General procedure of the cascade 1,2-addition-alkylation of (2-azidoethyl trifluoromethanesulfonate) and 1,2-addition-alkylation of (3-azidopropyl trifluoromethanesulfonate):**

To a mixture of 1-(trimethylsilyl)propyne (0.080 g, 0.72 mmol, 1.2 equiv) in anhydrous THF (2 mL) at -20 °C, *n*BuLi (0.49 mL, 1.6 M in THF, 0.78 mmol, 1.3 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of

cyclohexen-1-one (0.0576 g, 0.6 mmol, 1 equiv) in THF (3 mL) at -20 °C. Then, the azido-substrate was added dropwise at that temperature, and the reaction mixture was left to stir at -20 °C for 24 h.

The crude mixture was then quenched at 0 °C with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with Et<sub>2</sub>O. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography to afford the second-step product.



### **(3-(1-(2-azidoethoxy)cyclohex-2-en-1-yl)prop-1-yn-1-yl)trimethylsilane S6.1.8**

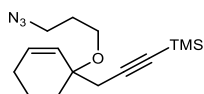
Following the general procedure, 2-azidoethyl trifluoromethanesulfonate (0.171 g, 0.78 mmol, 1.3 equiv) was used. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.8** as a colorless liquid (0.074 g, 45%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 6.04-6.00 (m, 1H), 5.73-5.70 (d, *J* = 10.18 Hz, 1H), 3.68-3.65 (m, 2H), 3.35-3.32 (m, 2H), 2.53-2.51 (d, *J* = 1.60 Hz, 2H), 2.12-2.06 (m, 1H), 2.03-1.97 (m, 1H), 1.94-1.90 (m, 1H), 1.81-1.75 (m, 2H), 1.72-1.66 (m, 1H), 0.17 (s, 9H);

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 132.8, 128.9, 103.3, 86.9, 74.0, 61.9, 51.3, 32.3, 31.4, 25.1, 19.2, 0.0;

HRMS (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. 300.1508; Found 300.1513;

IR (film, CDCl<sub>3</sub>) 3023, 2955, 2868, 2109, 1439, 1250, 1087, 843, 760, 647 cm<sup>-1</sup>.



### **(3-(1-(3-azidopropoxy)cyclohex-2-en-1-yl)prop-1-yn-1-yl)trimethylsilane S6.1.9**

Following the general procedure, 3-azidopropyl trifluoromethanesulfonate (0.182 g, 0.78 mmol, 1.3 equiv) was used. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.9** as a colorless liquid (0.114 g, 65%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 6.01-5.97 (m, 1H), 5.71-5.69 (d, *J* = 10.15 Hz, 1H), 3.53-3.50 (t, *J* = 5.83 Hz, 2H), 3.46-3.43 (t, *J* = 6.73 Hz, 2H), 2.49 (s, *J* = Hz, 2H), 2.11-2.05 (m, 1H),

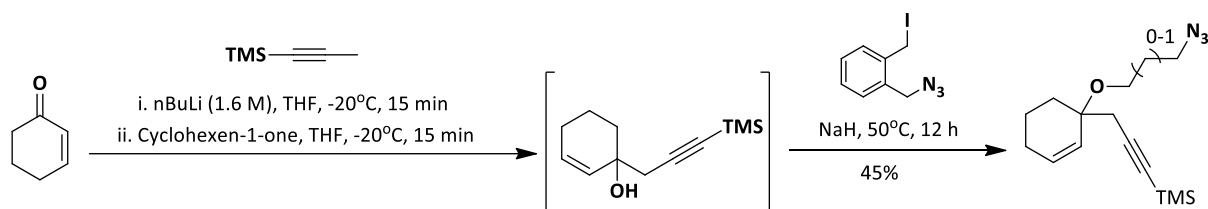


1.99-1.94 (m, 1H), 1.92-1.88 (m, 1H), 1.85-1.80 (m, 2H), 1.77-1.71 (m, 1H), 1.69-1.62 (m, 2H), 0.17 (s, 9H);

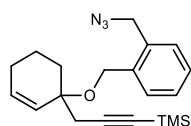
$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  132.2, 129.4, 103.5, 86.7, 73.4, 58.9, 48.6, 32.2, 31.2, 29.8, 25.1, 19.3, 0.0;

HRMS (ESI-HR)  $m/z$ :  $[\text{M}]^+$  Calc. 314.1665; Found 314.1669;

IR (film,  $\text{CDCl}_3$ ) 3023, 2954, 2870, 2096, 1453, 1250, 1088, 843, 760, 647  $\text{cm}^{-1}$ .



**Supplementary Figure S40.** Reaction scheme of the stepwise process to confirm the second step intermediate of the 1,2-addition-alkylation with methylated alkynes and benzyl iodides



### (3-(1-((2-(azidomethyl)benzyl)oxy)cyclohex-2-en-1-yl)prop-1-yn-1-yl)trimethylsilane

#### S6.1.10

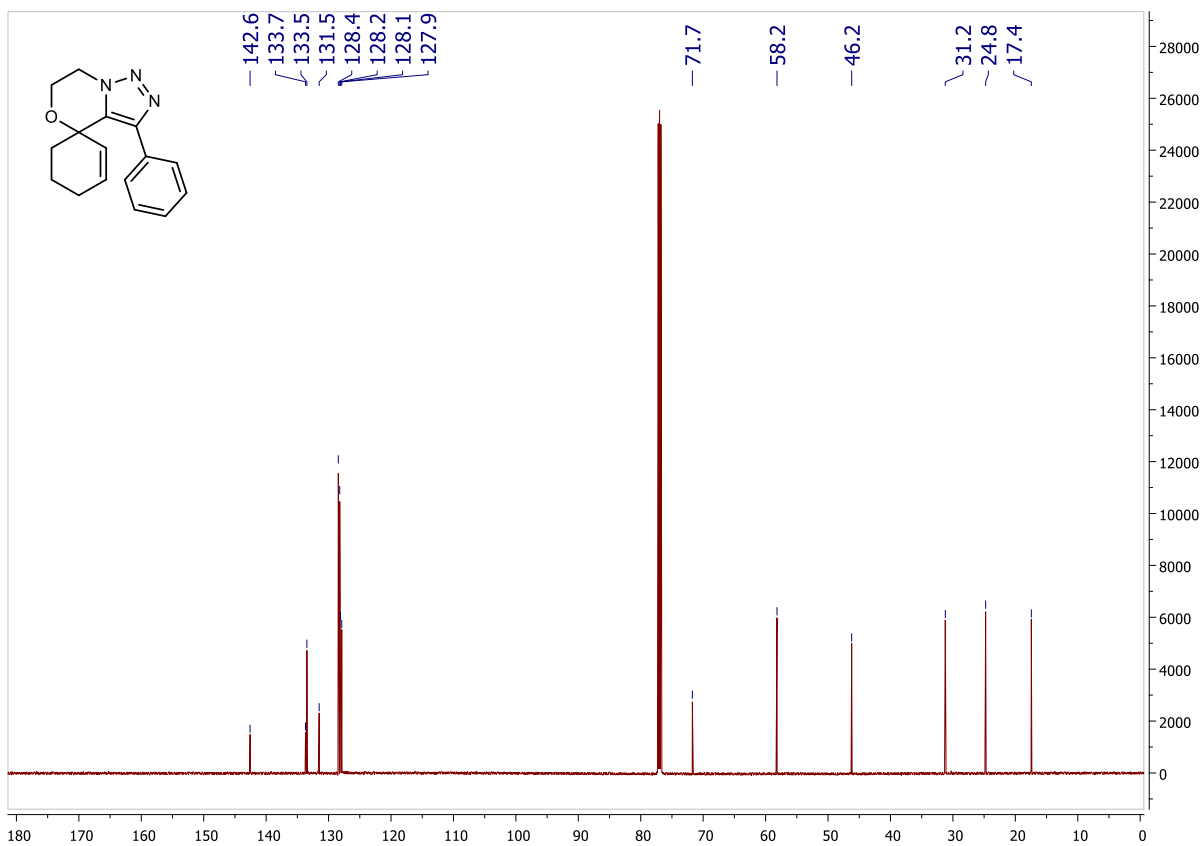
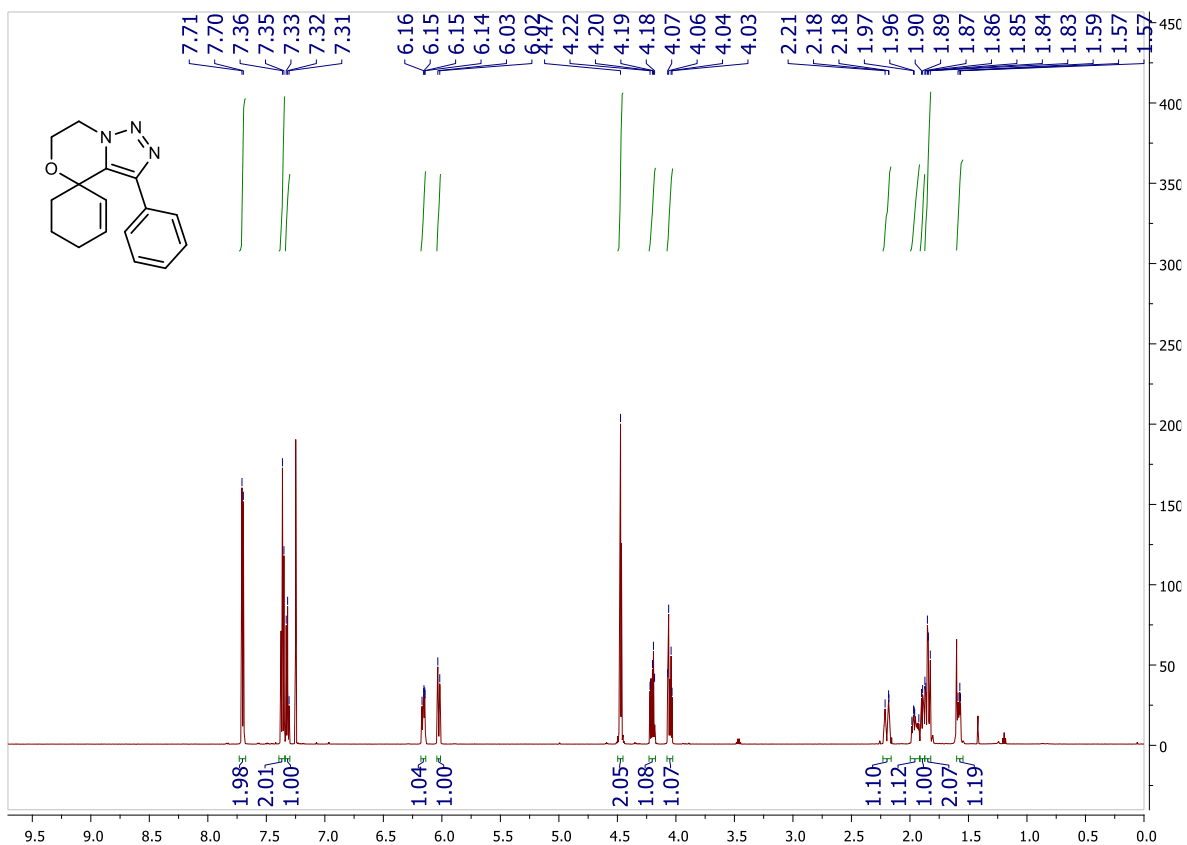
Following the same first step of the general procedure, 1-(azidomethyl)-2-(iodomethyl)benzene (0.213 g, 0.78 mmol, 1.3 equiv) was added dropwise at -20 °C, the reaction was then allowed to warm up to reach rt. NaH (60% dispersion in mineral oil, 40 mg, 1 mmol, 1.67 equiv) was later on added to the mixture, followed by heating it at 50°C for 12 h. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the second-step product **S6.1.10** as off-white liquid (0.095 g, 45%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.50-7.47 (m, 1H), 7.34-7.33 (m, 3H), 6.09-6.04 (m, 1H), 5.84-5.82 (d,  $J = 10.18$  Hz, 1H), 4.58 (s, 2H), 4.50 (s, 2H), 2.60 (s, 2H), 2.15- 2.10 (m, 1H), 2.06-2.01 (m, 2H), 1.84-1.78 (m, 2H), 1.73-1.66 (m, 1H), 0.19 (s, 9H);

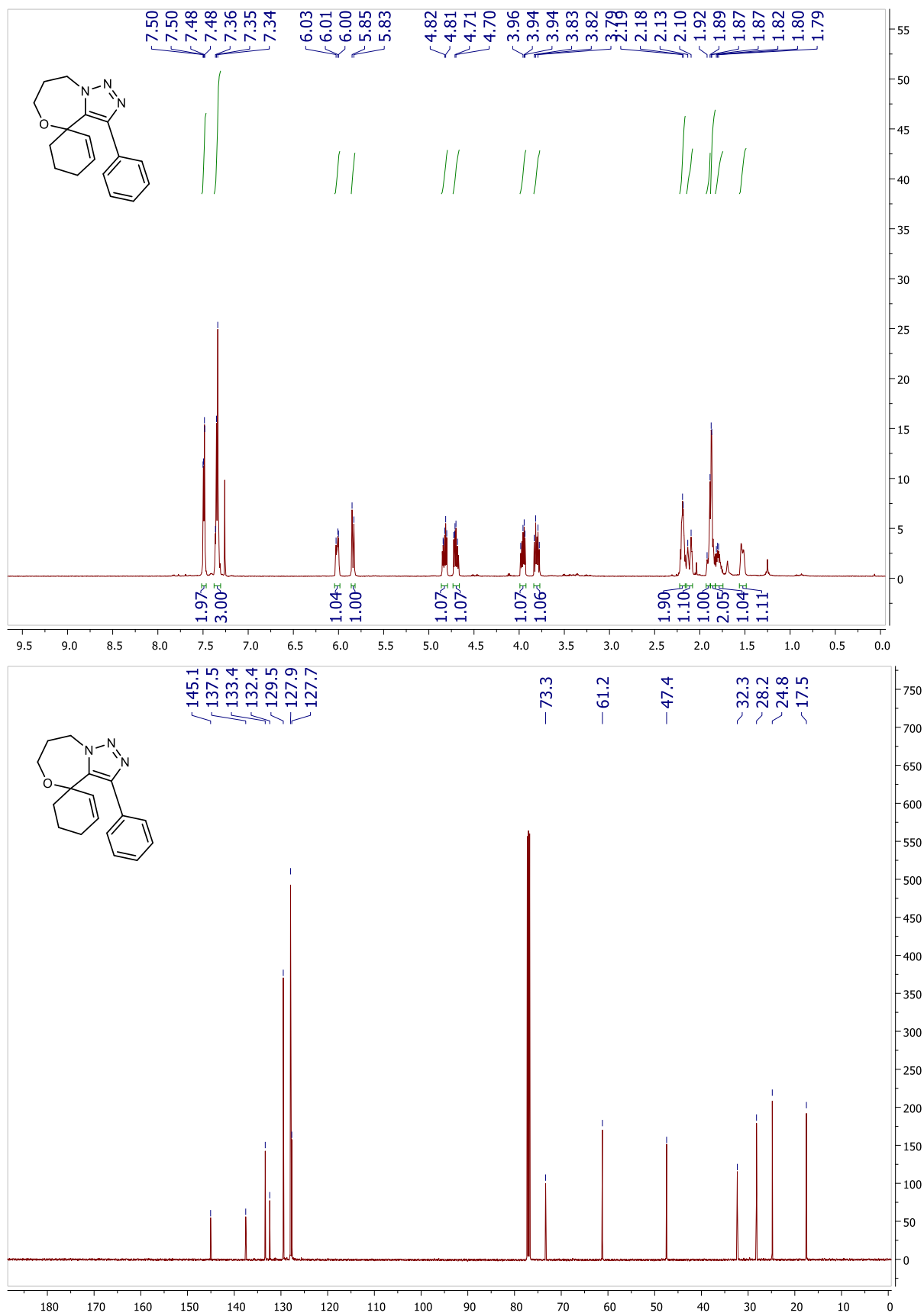
$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ )  $\delta$  137.6, 133.8, 132.9, 129.6, 129.3, 129.2, 128.4, 127.9, 103.5, 86.9, 74.3, 62.7, 52.2, 32.3, 31.5, 25.2, 19.5, 0.0;

HRMS (EI-HR)  $m/z$ :  $[\text{M}]^+$  Calc. 353.1923; Found 353.1928;

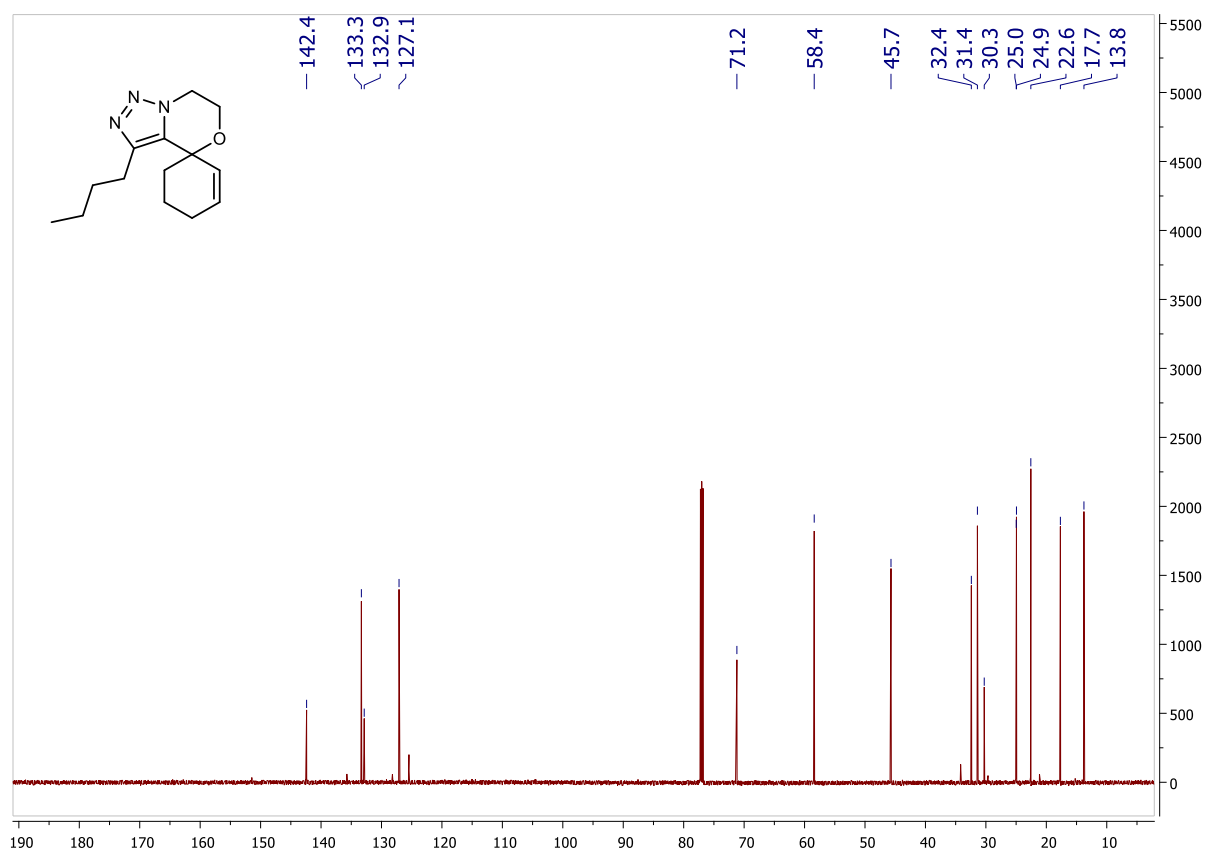
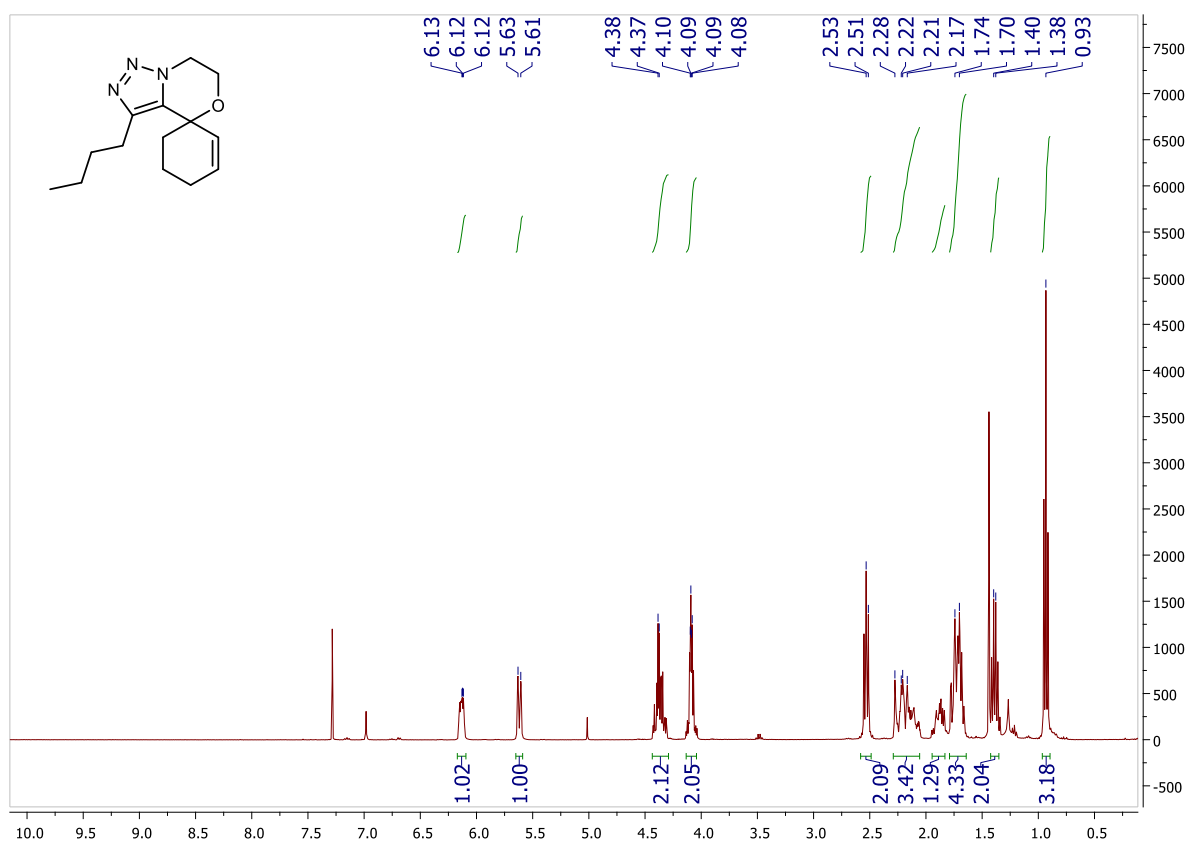
IR (film,  $\text{CDCl}_3$ ) 3032, 2954, 2096, 1455, 1250, 1065, 843, 759  $\text{cm}^{-1}$ .



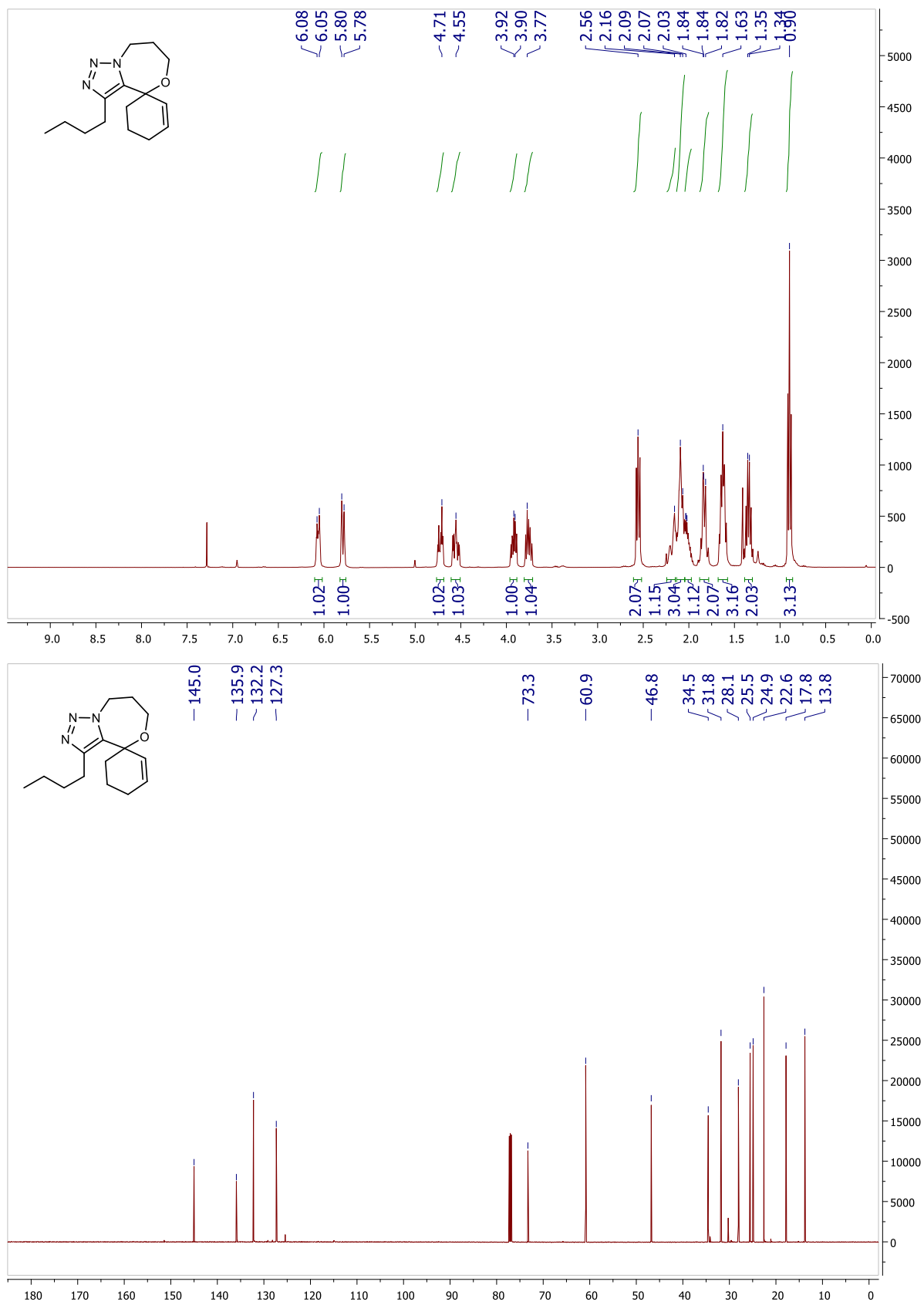
Supplementary Figure S41. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2a**



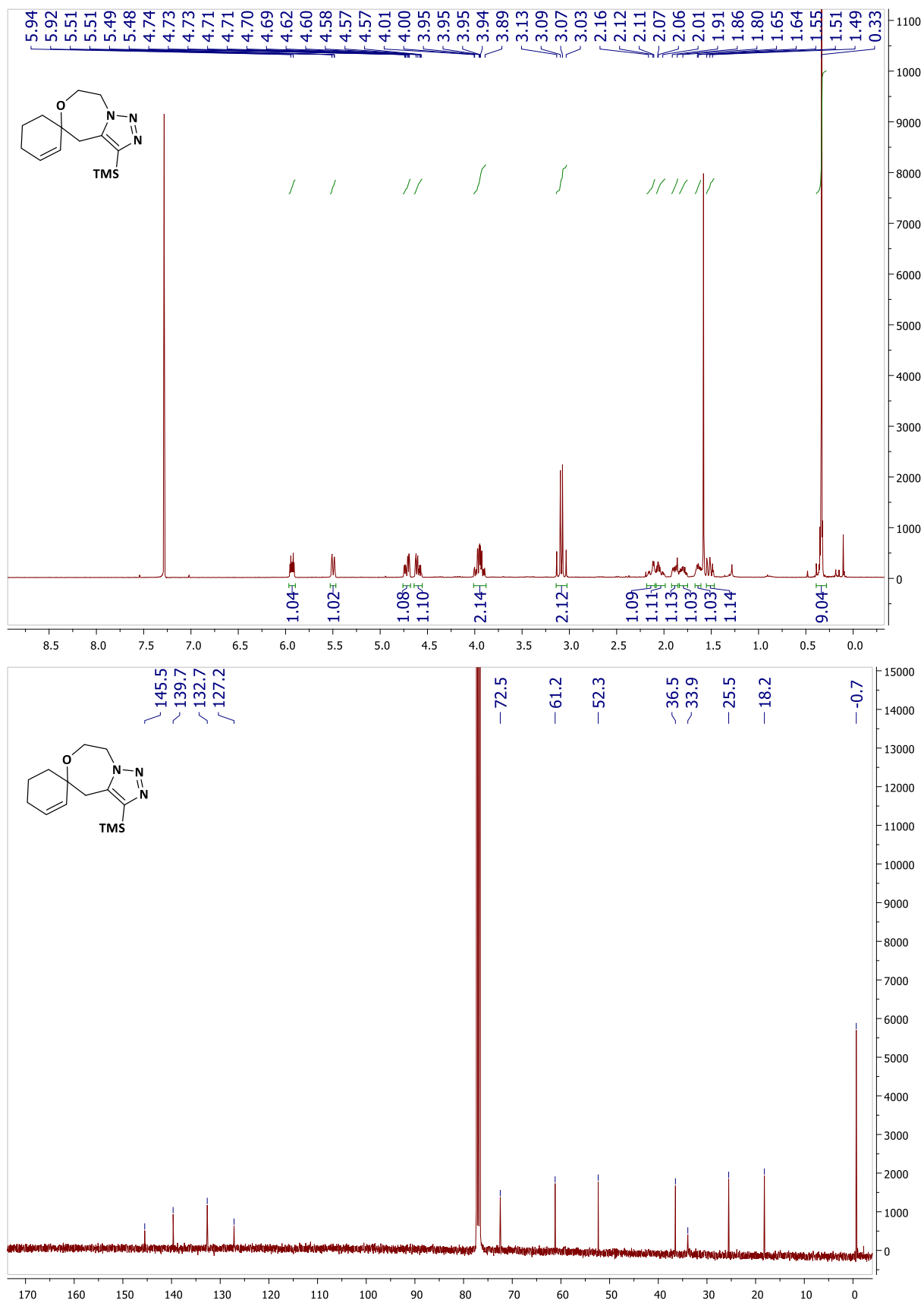
Supplementary Figure S42. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2b**.



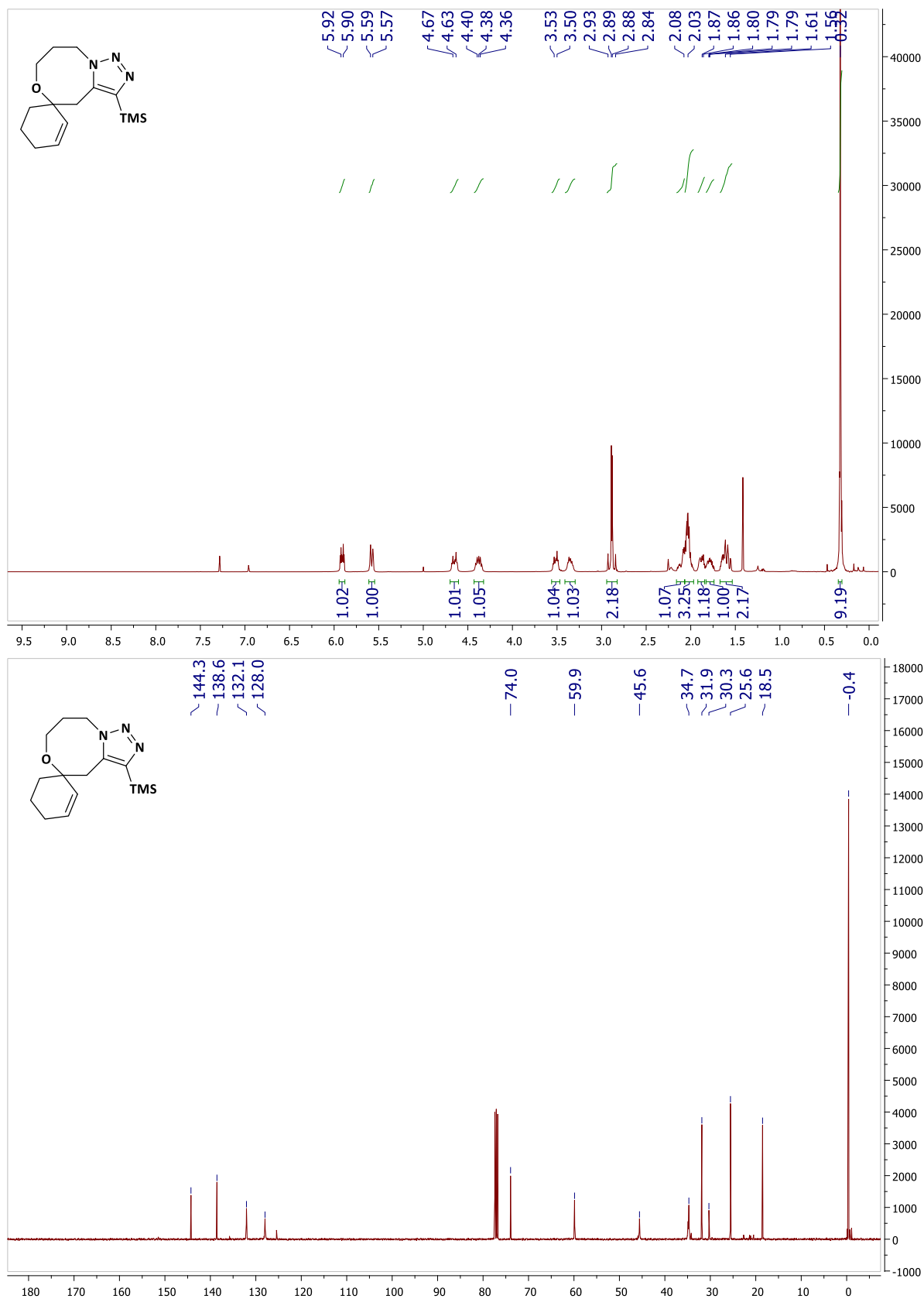
Supplementary Figure S43.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound **2c**.



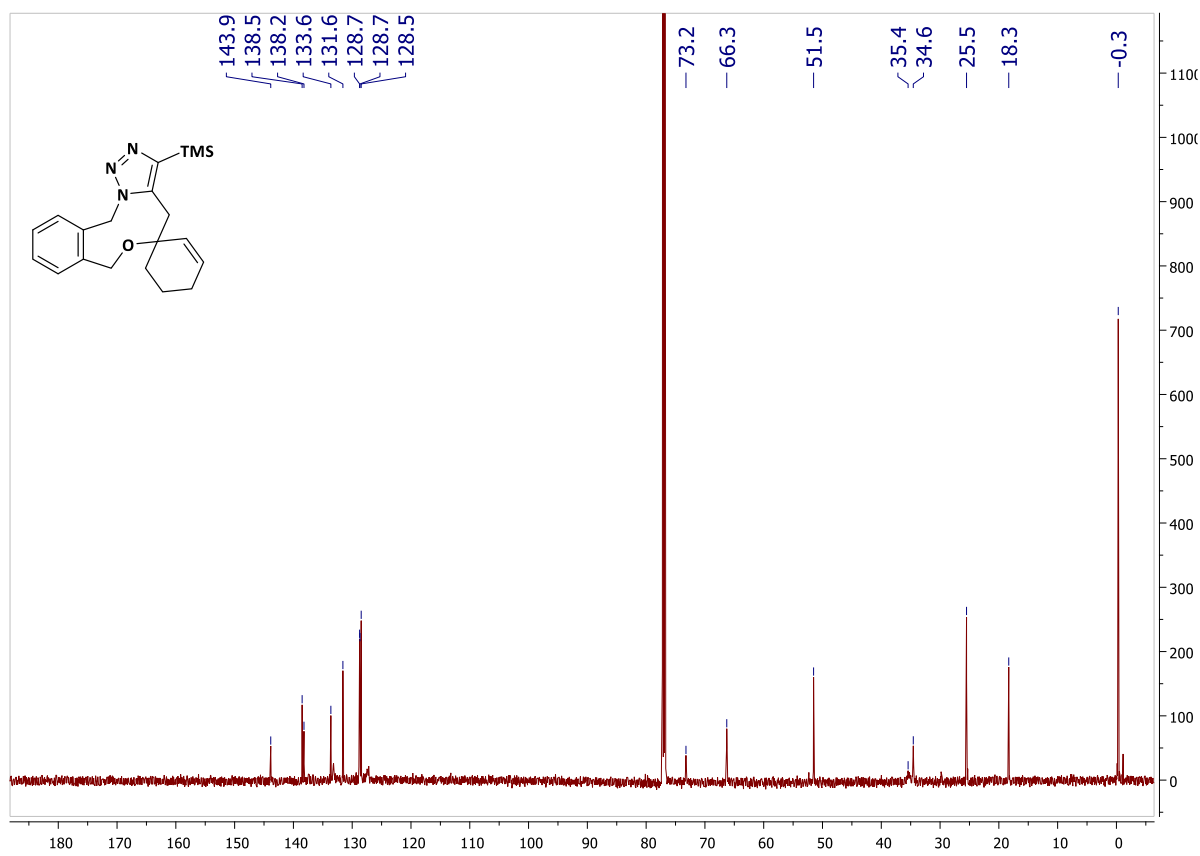
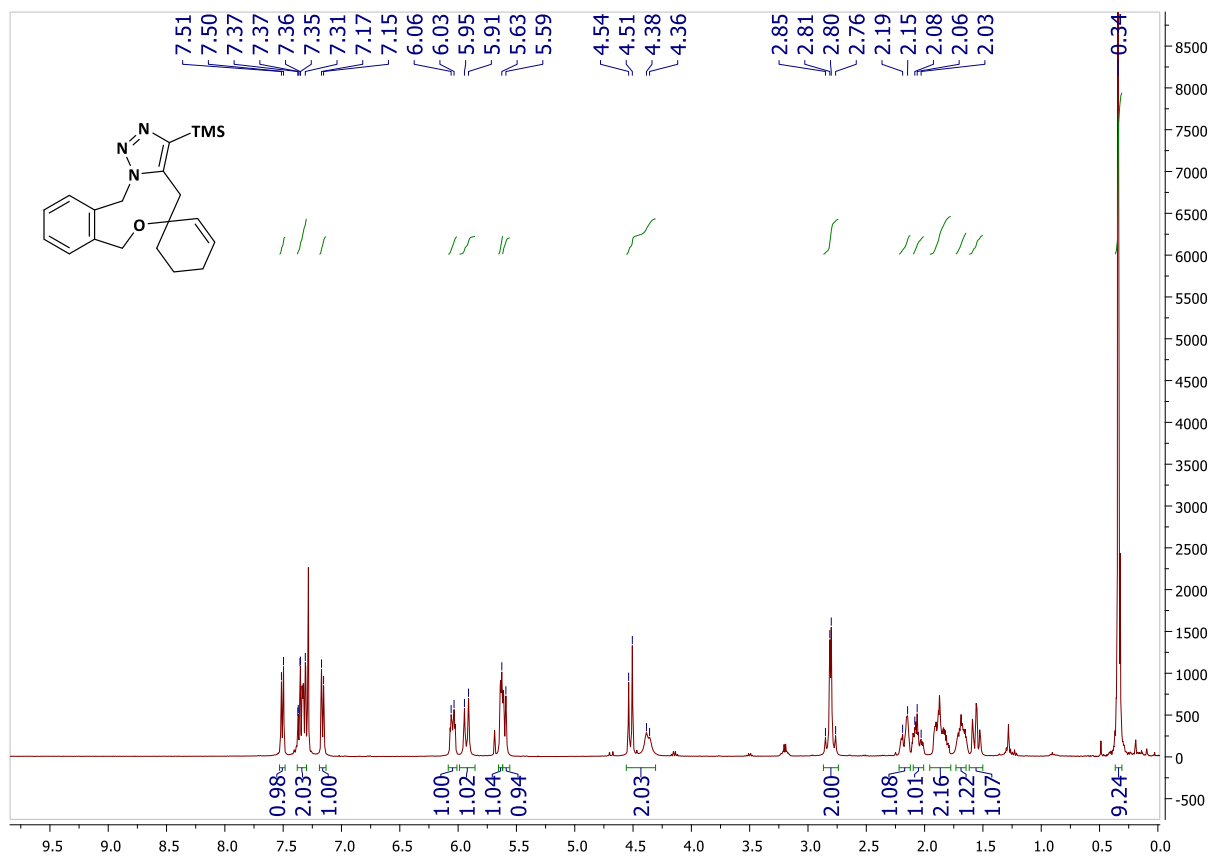
Supplementary Figure S44. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2d**.



Supplementary Figure S45. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2e**.

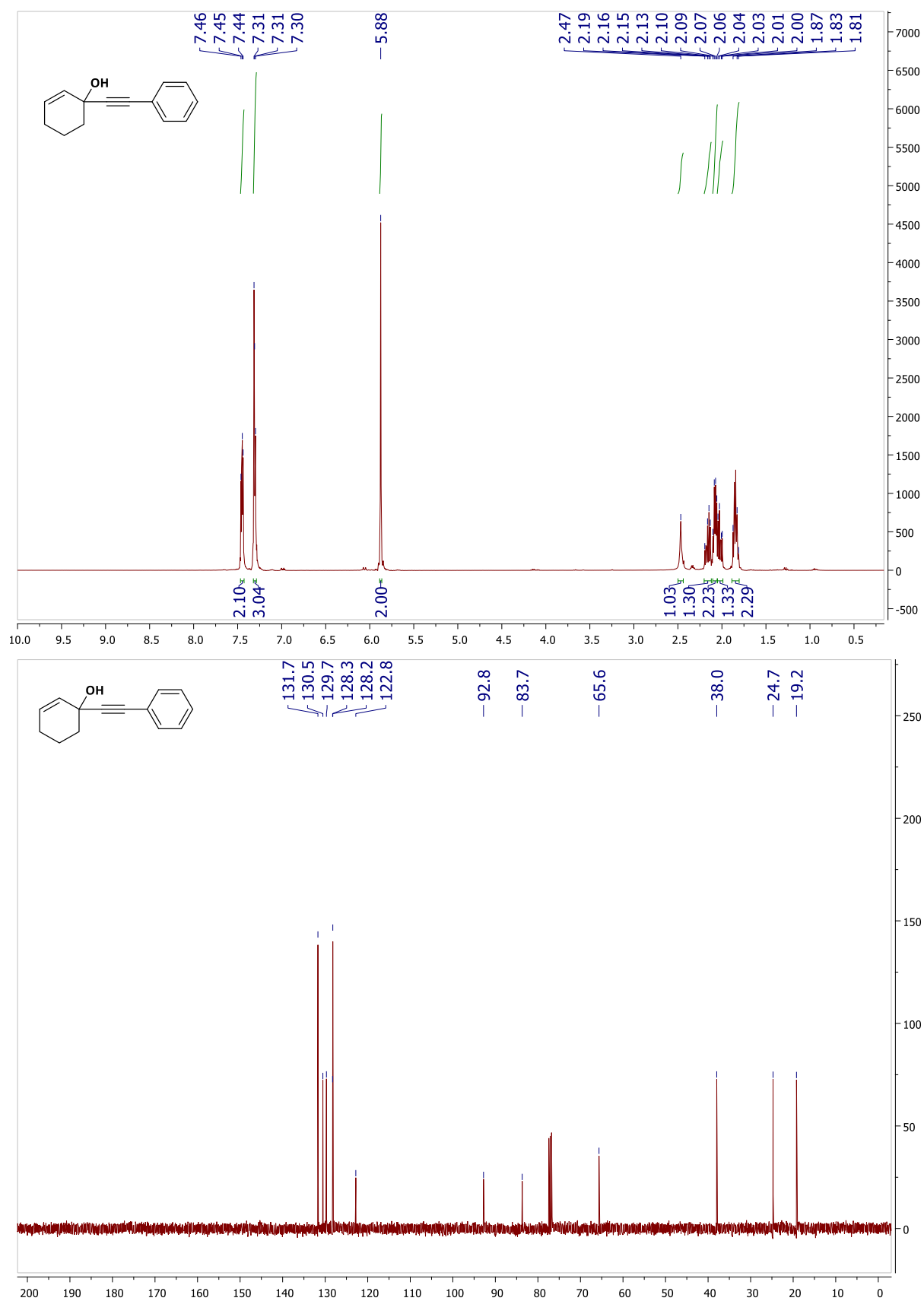


Supplementary Figure S46. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2f**.

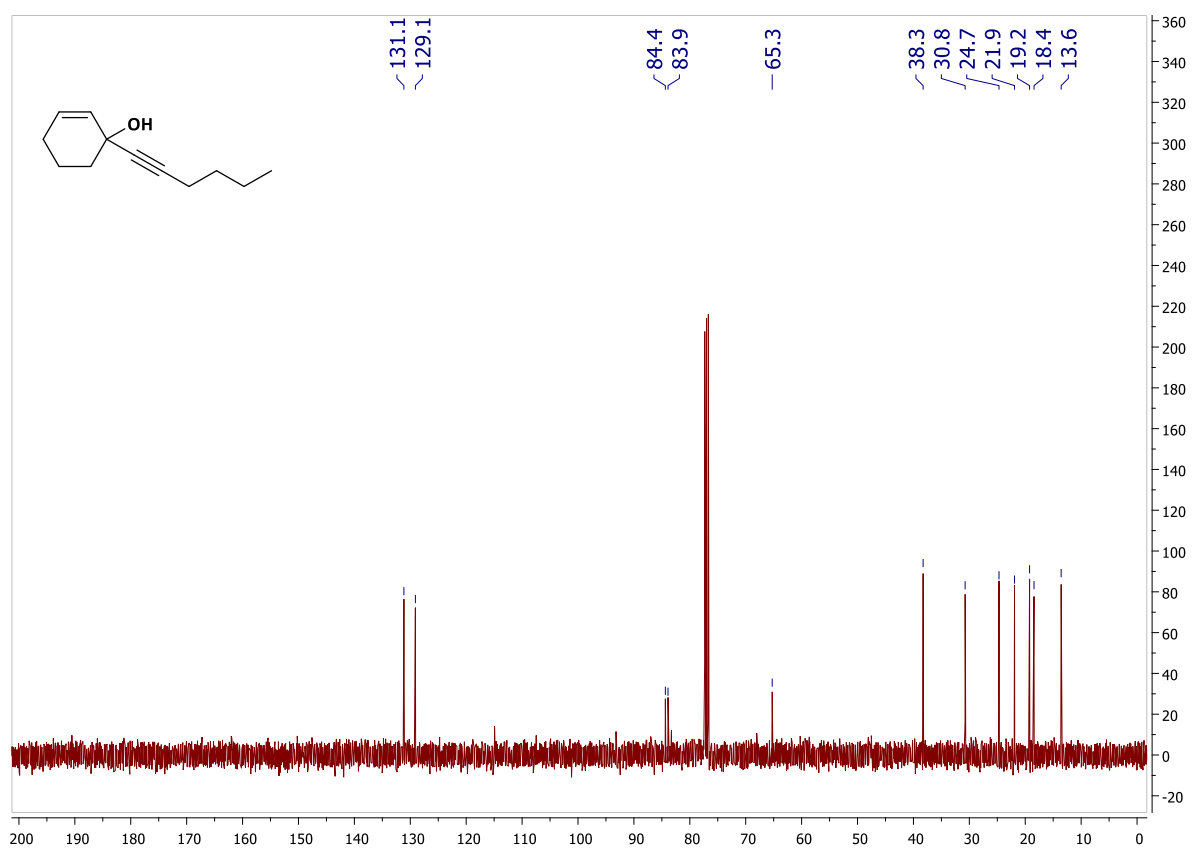
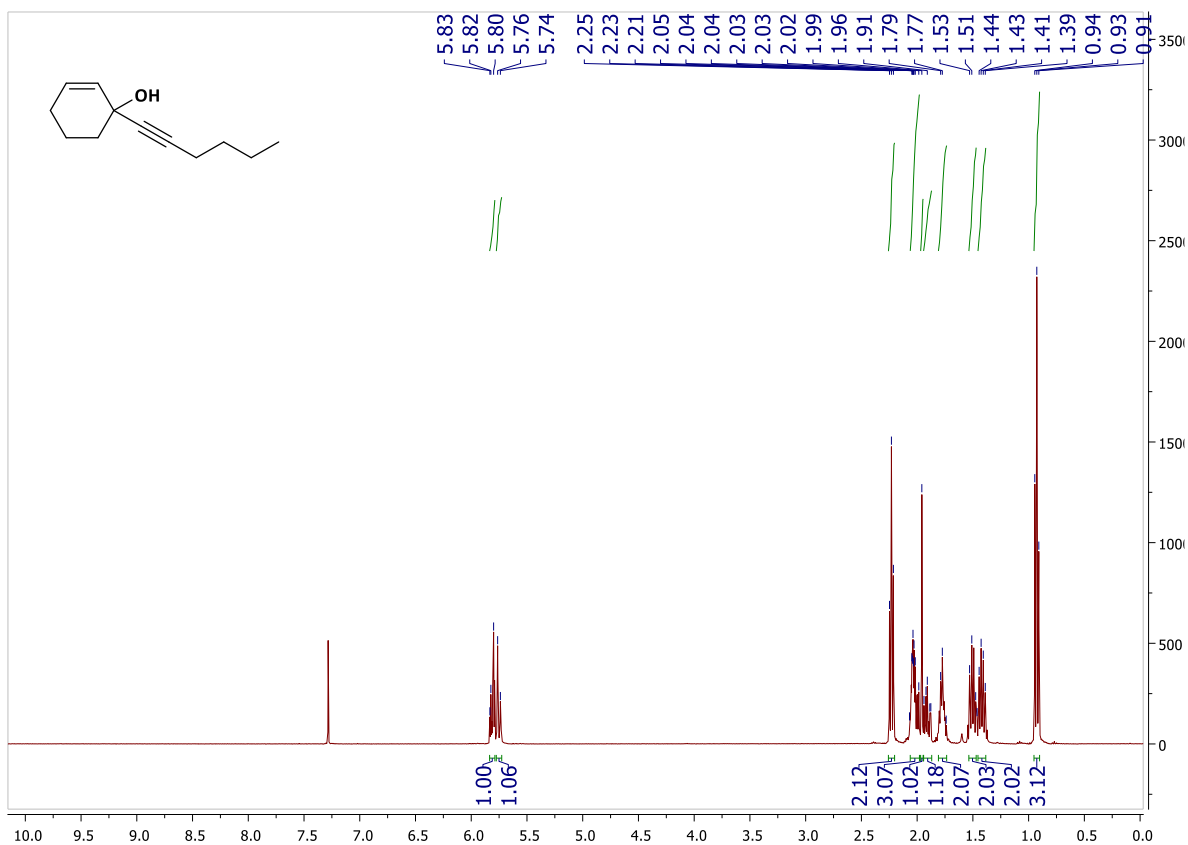


Supplementary Figure S47. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **2g**.

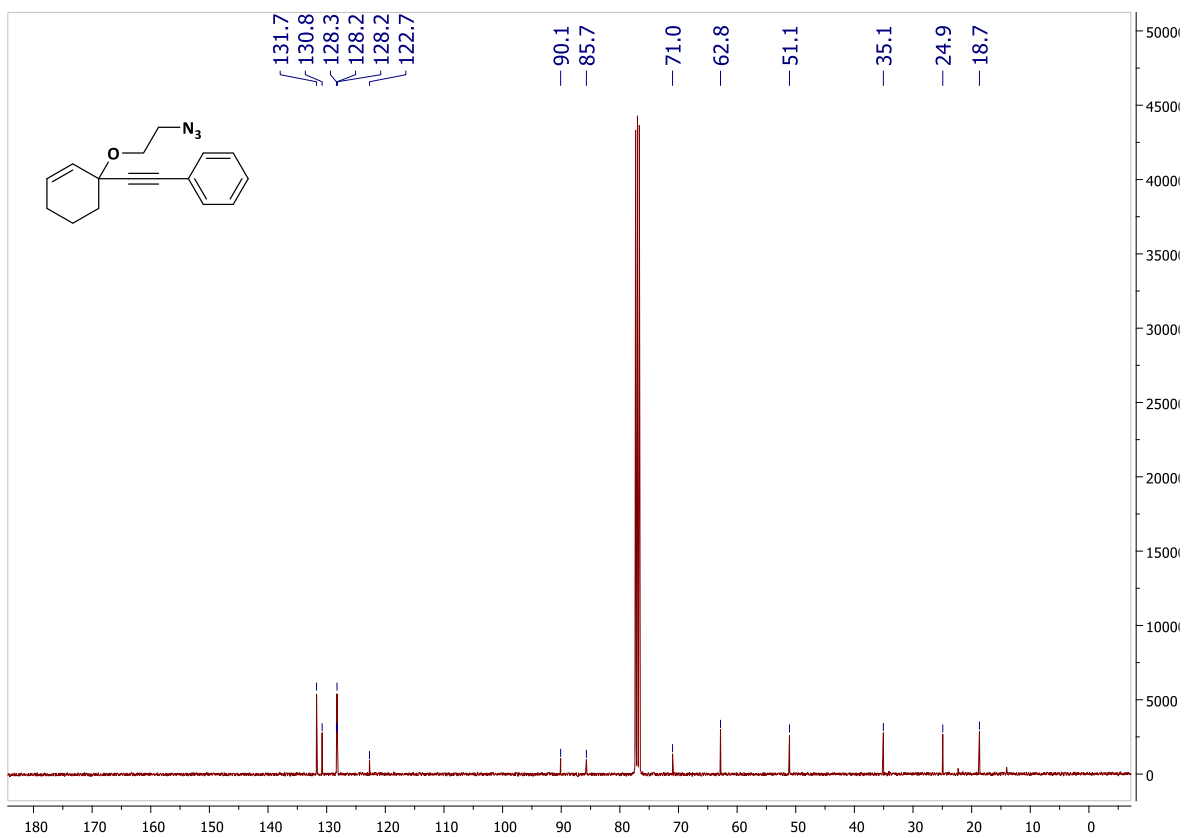
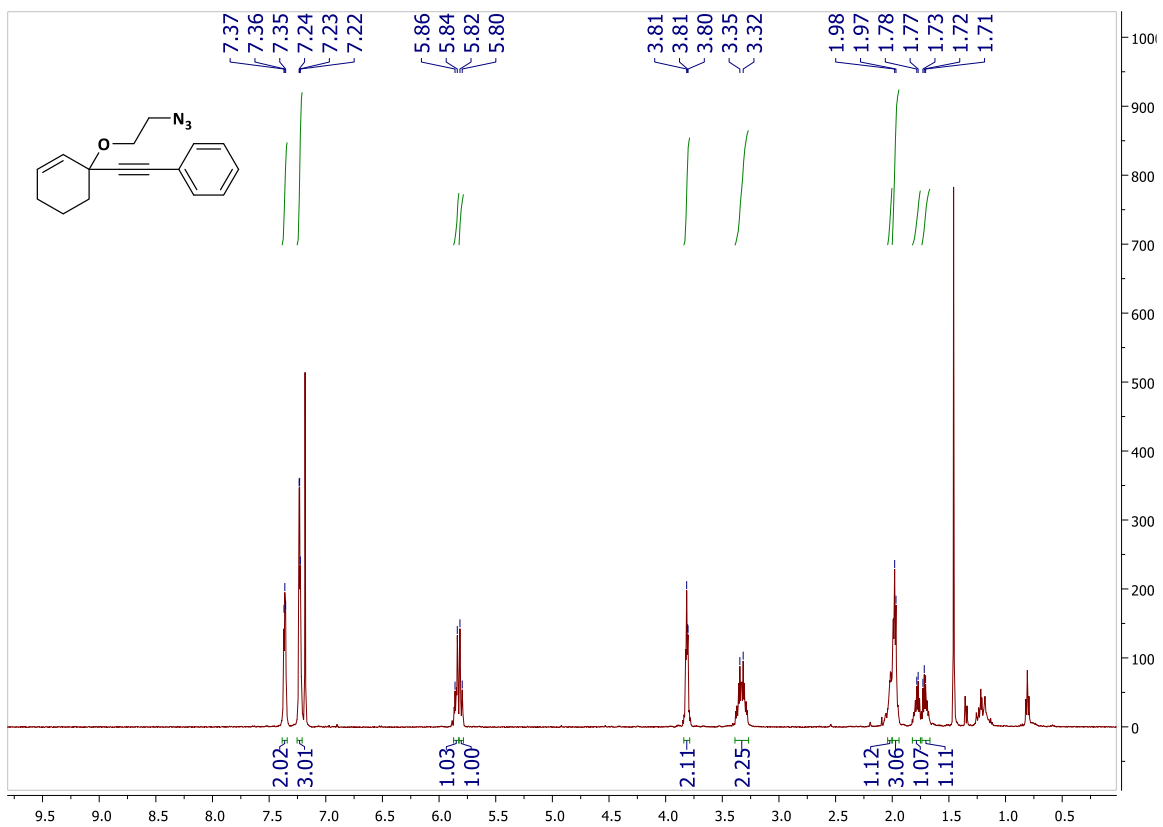




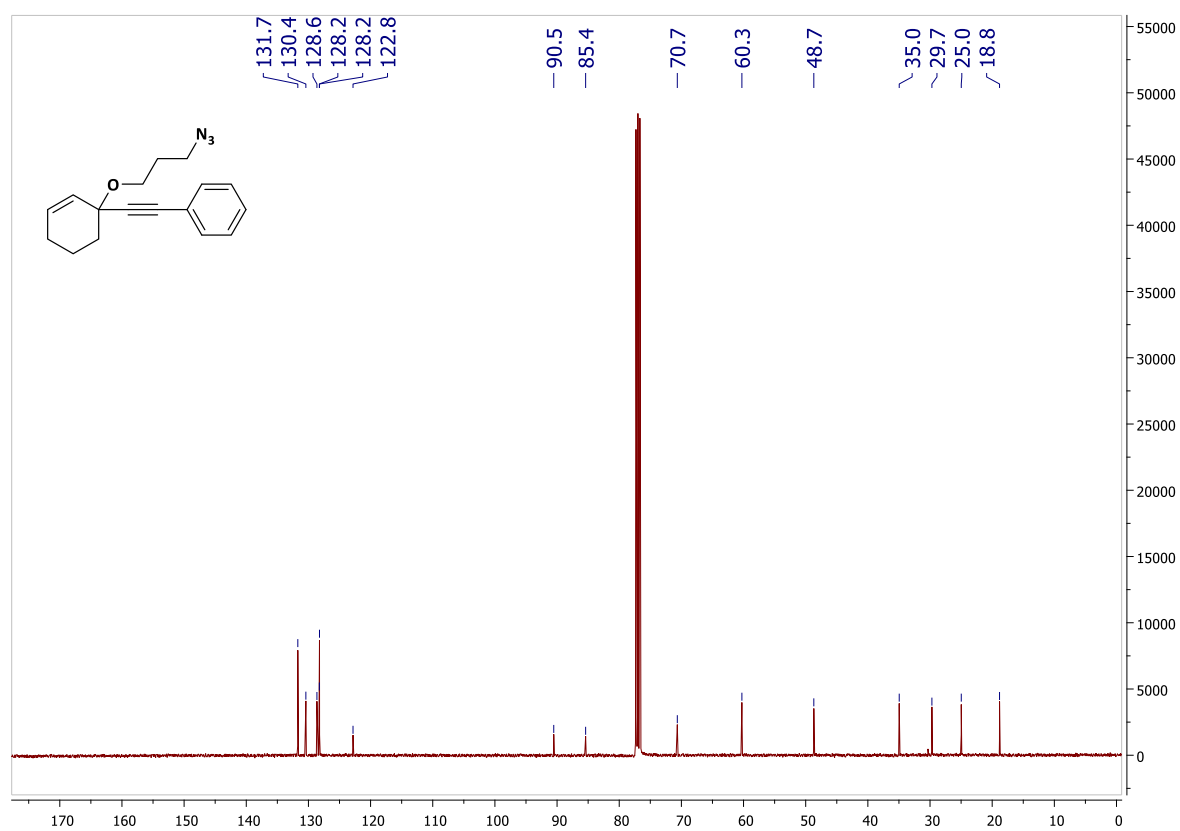
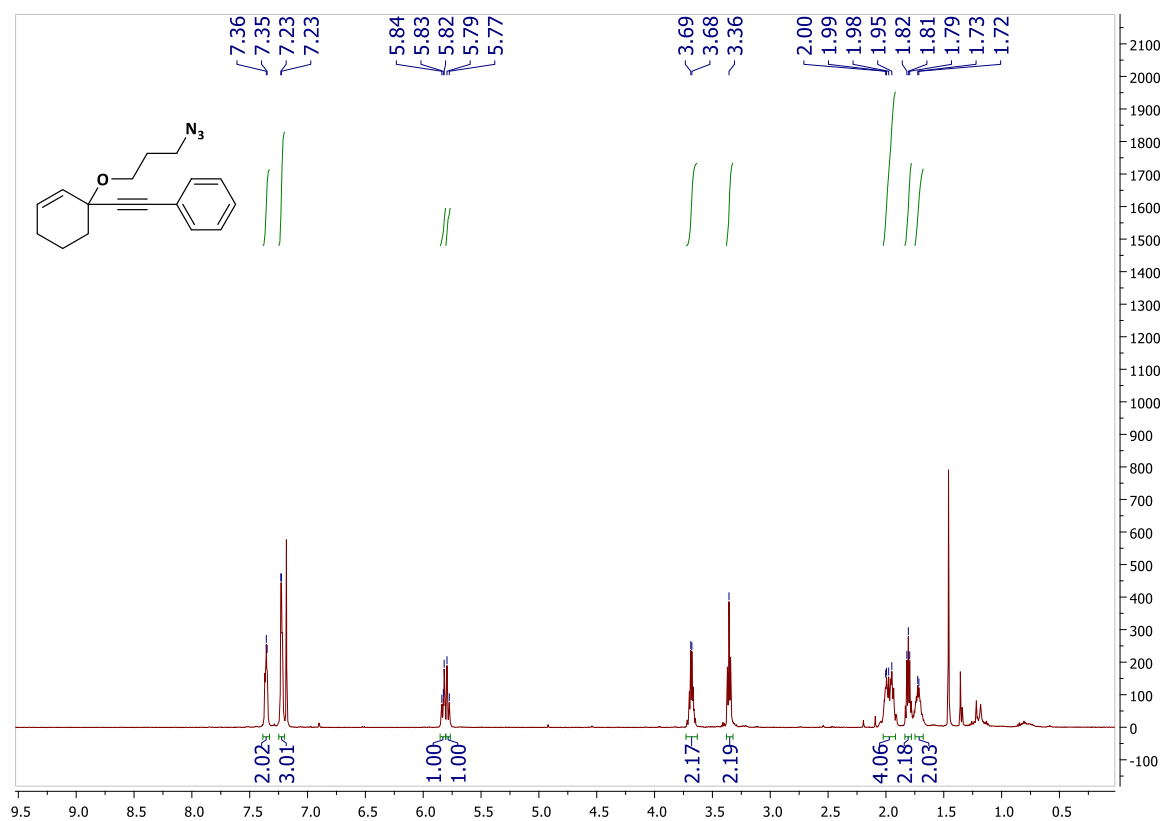
**Supplementary Figure S48.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.1.



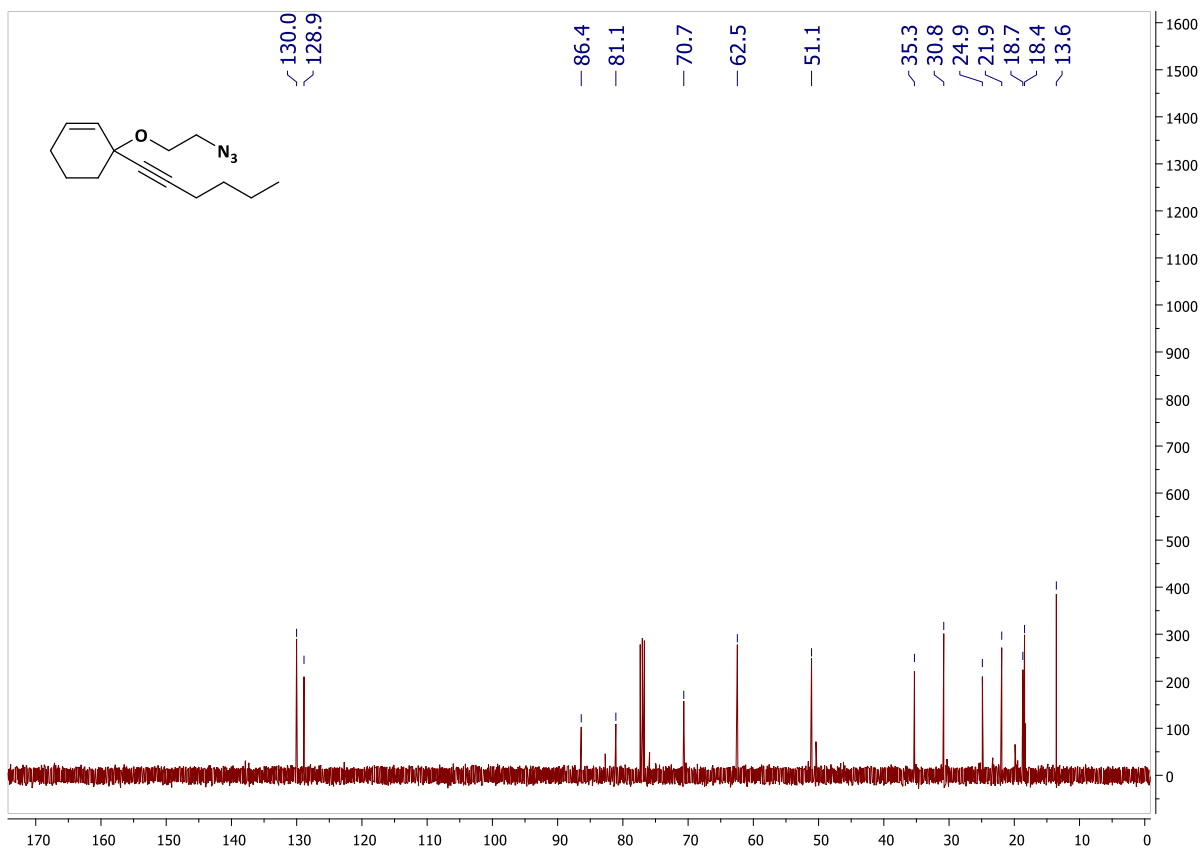
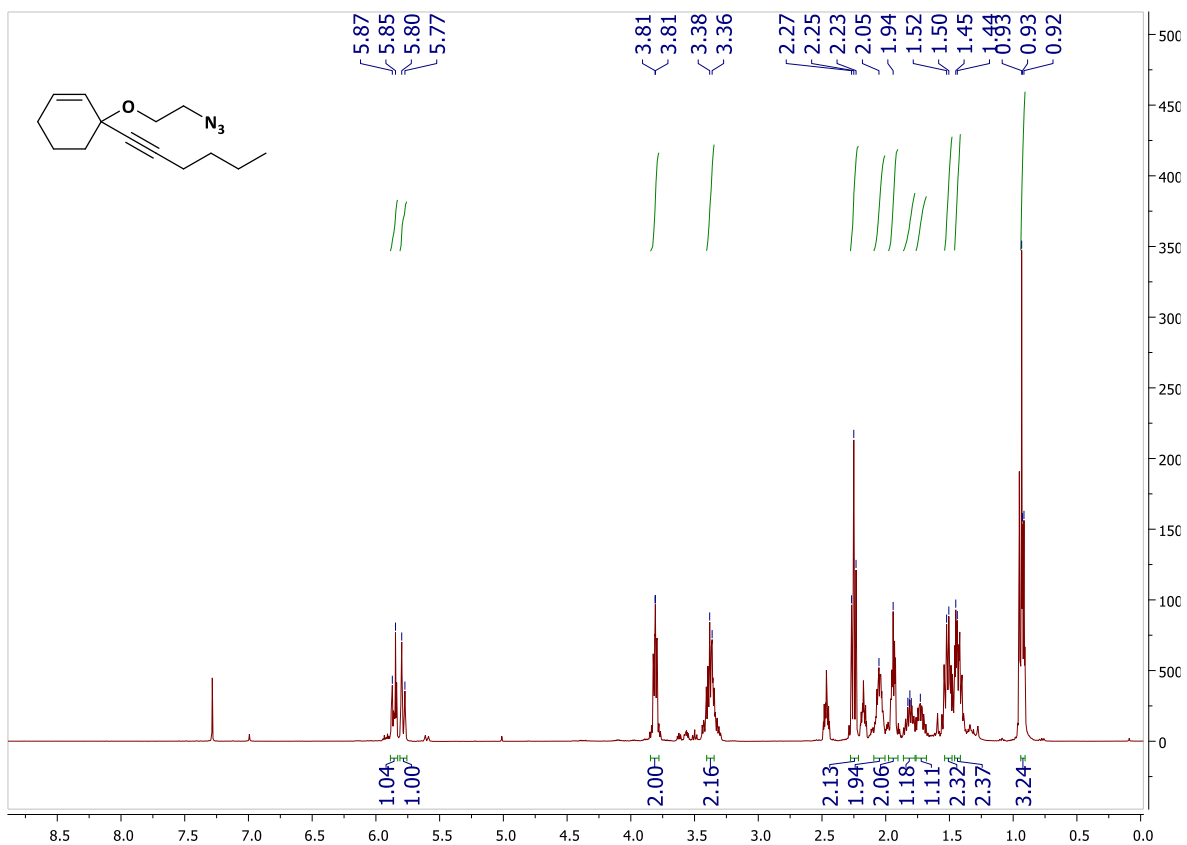
**Supplementary Figure S49.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.2.



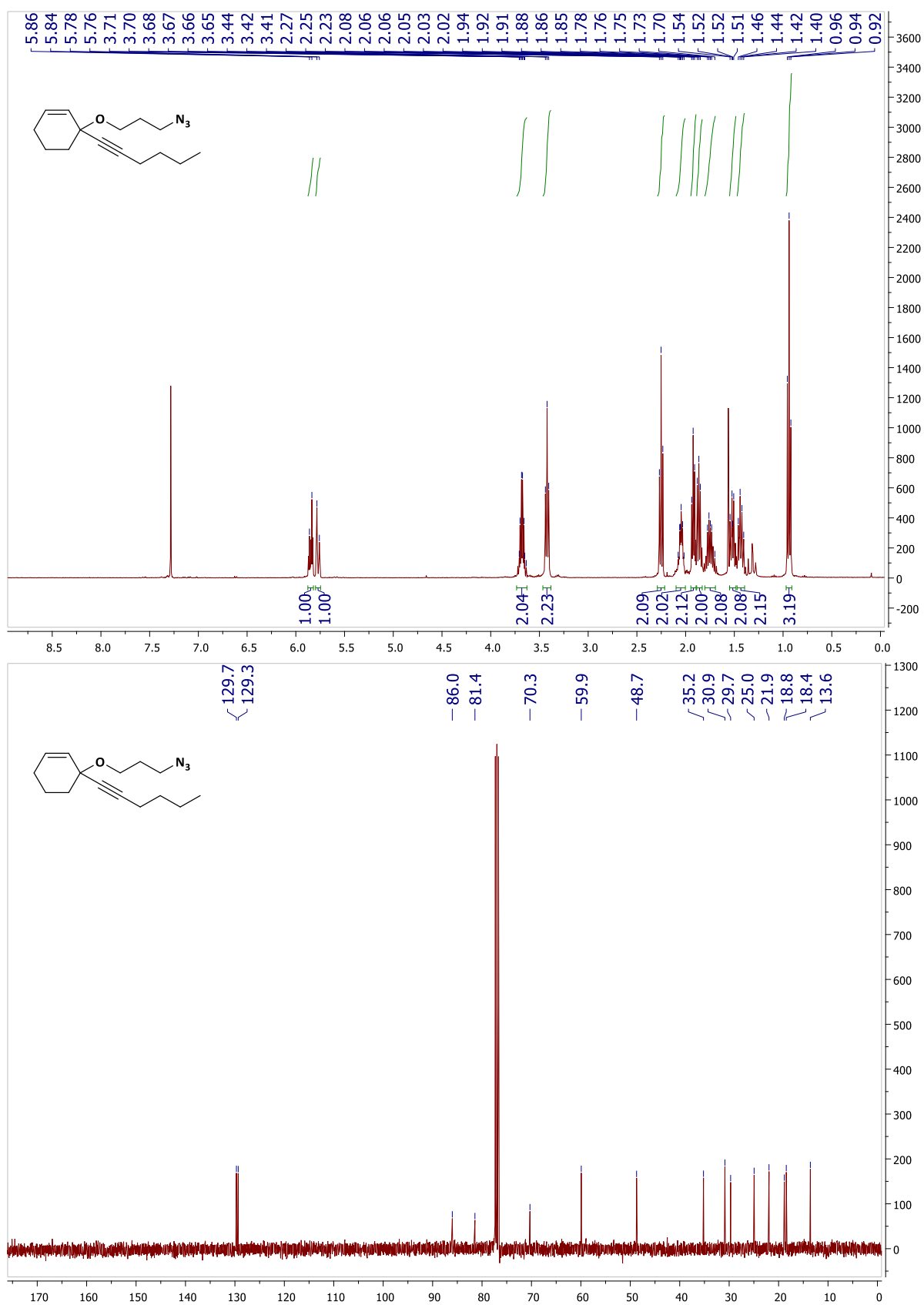
**Supplementary Figure S50.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.3.



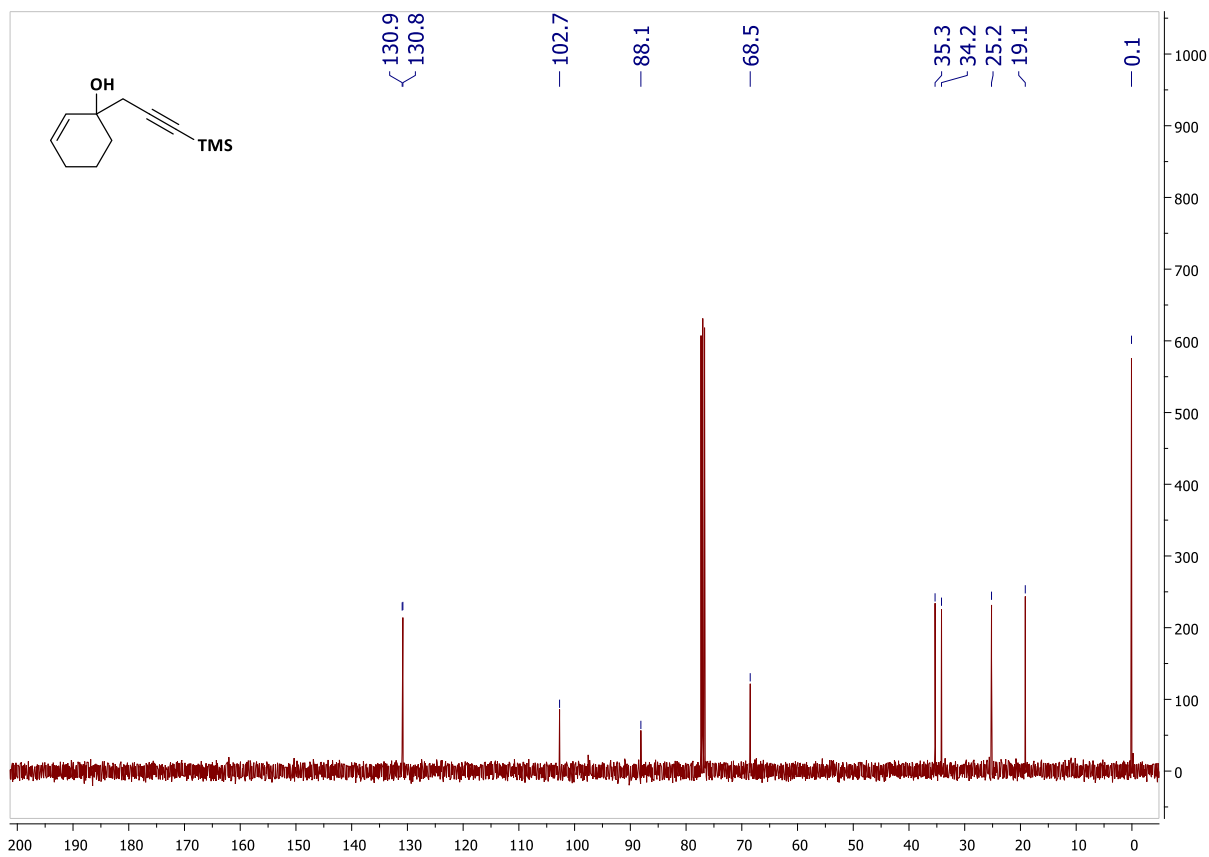
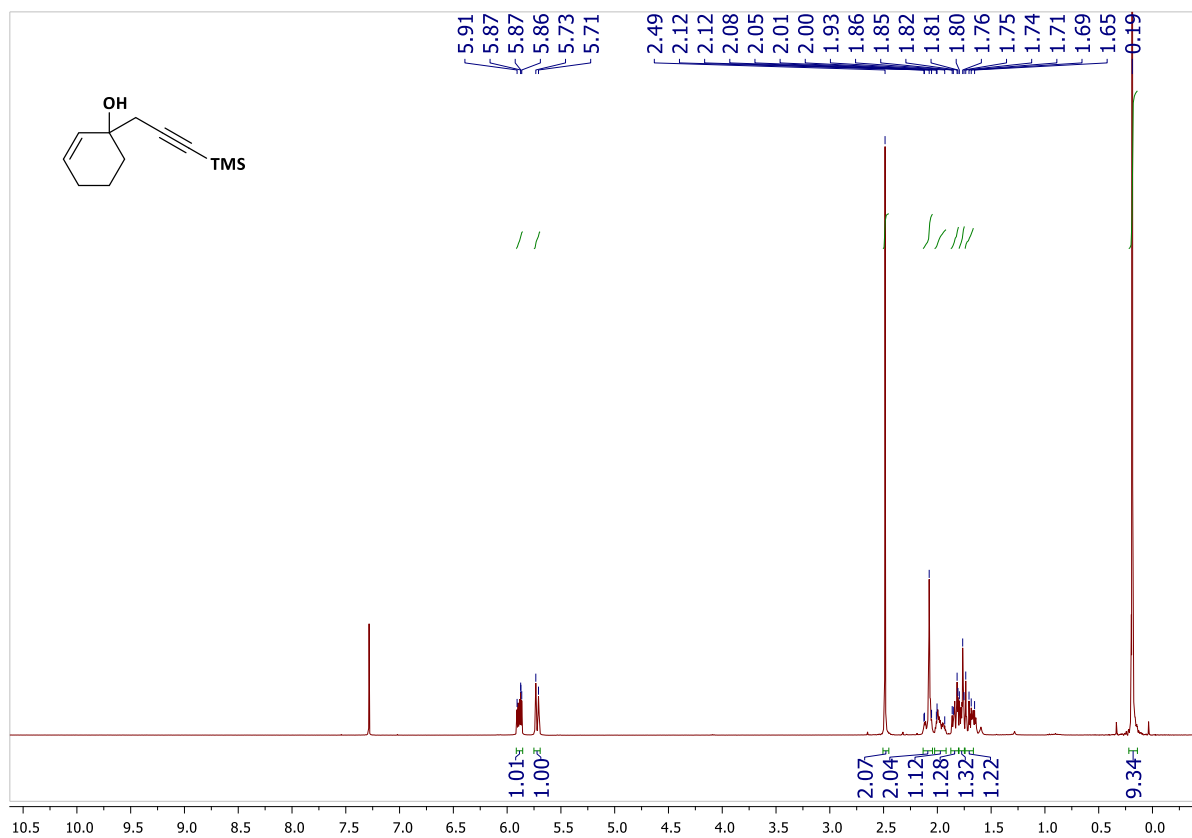
**Supplementary Figure S51:**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.1.4.



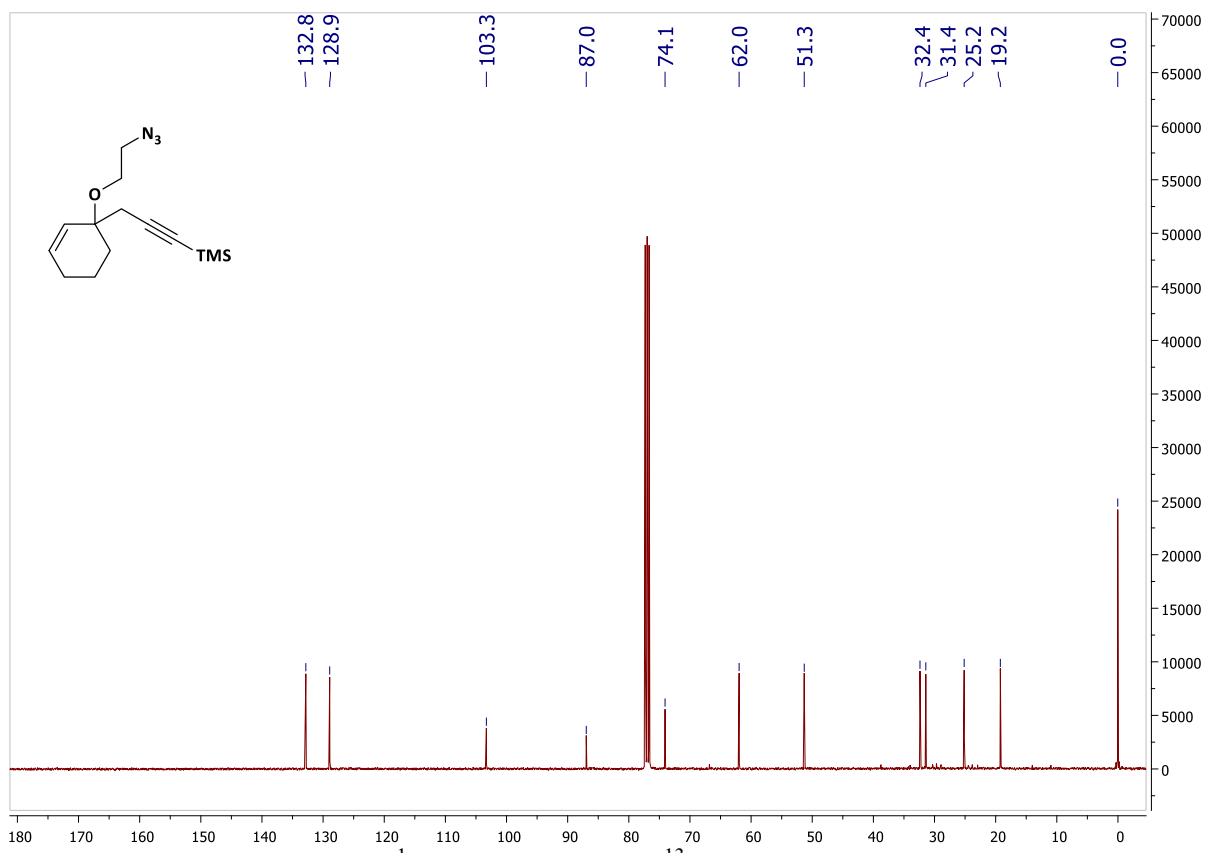
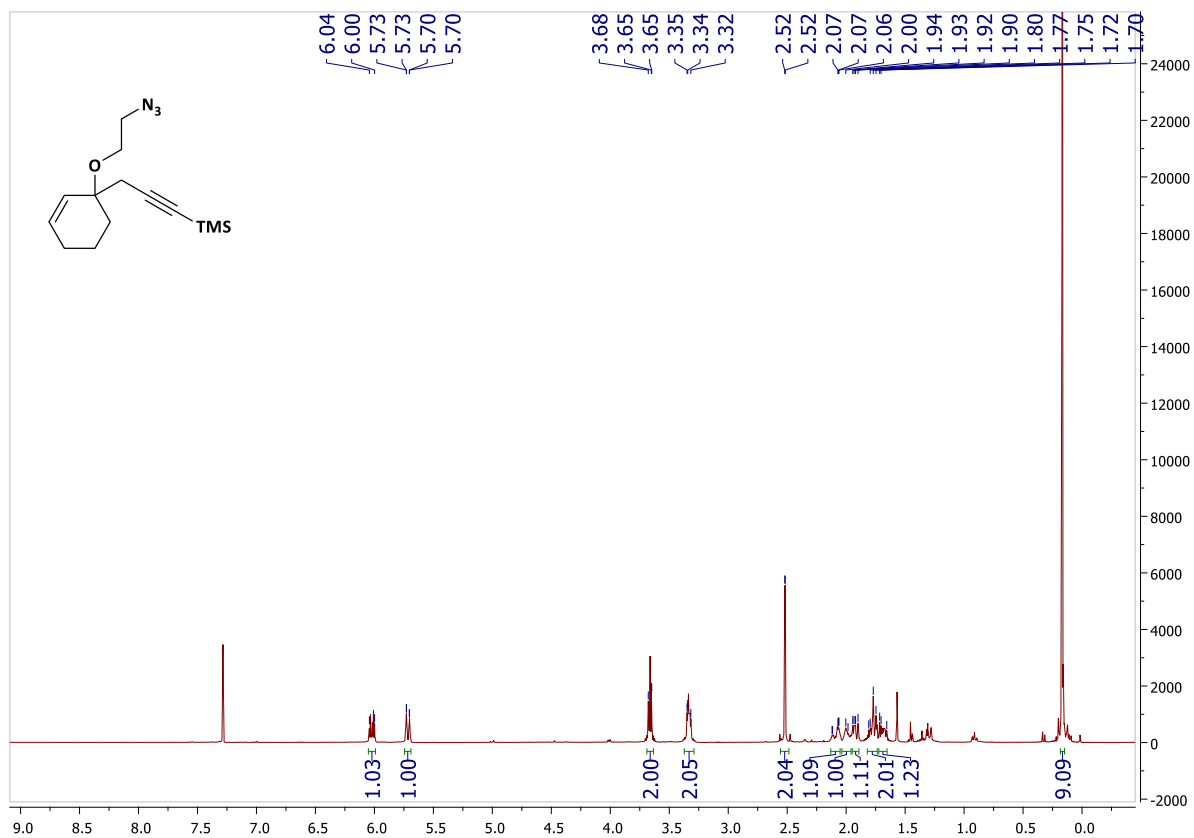
**Supplementary Figure S52.**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.1.5.



Supplementary Figure S53. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.6.

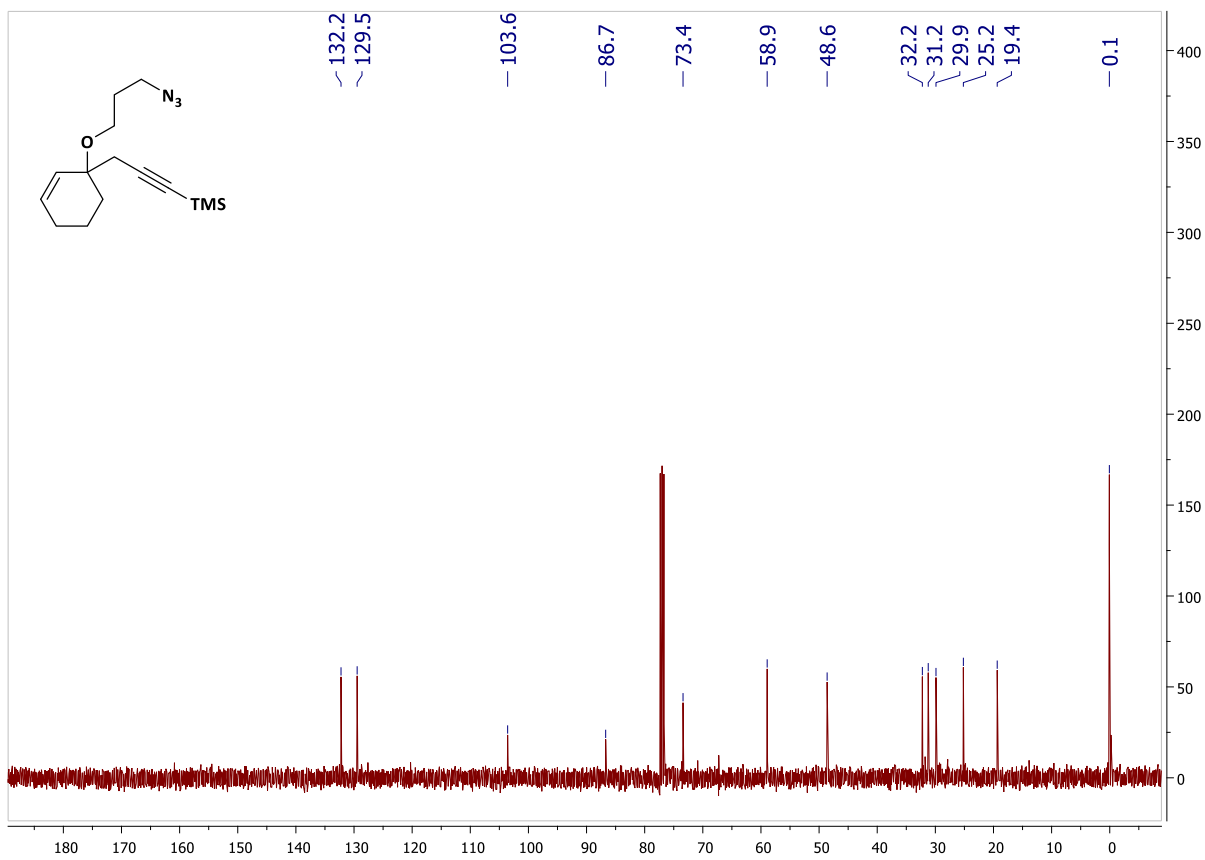
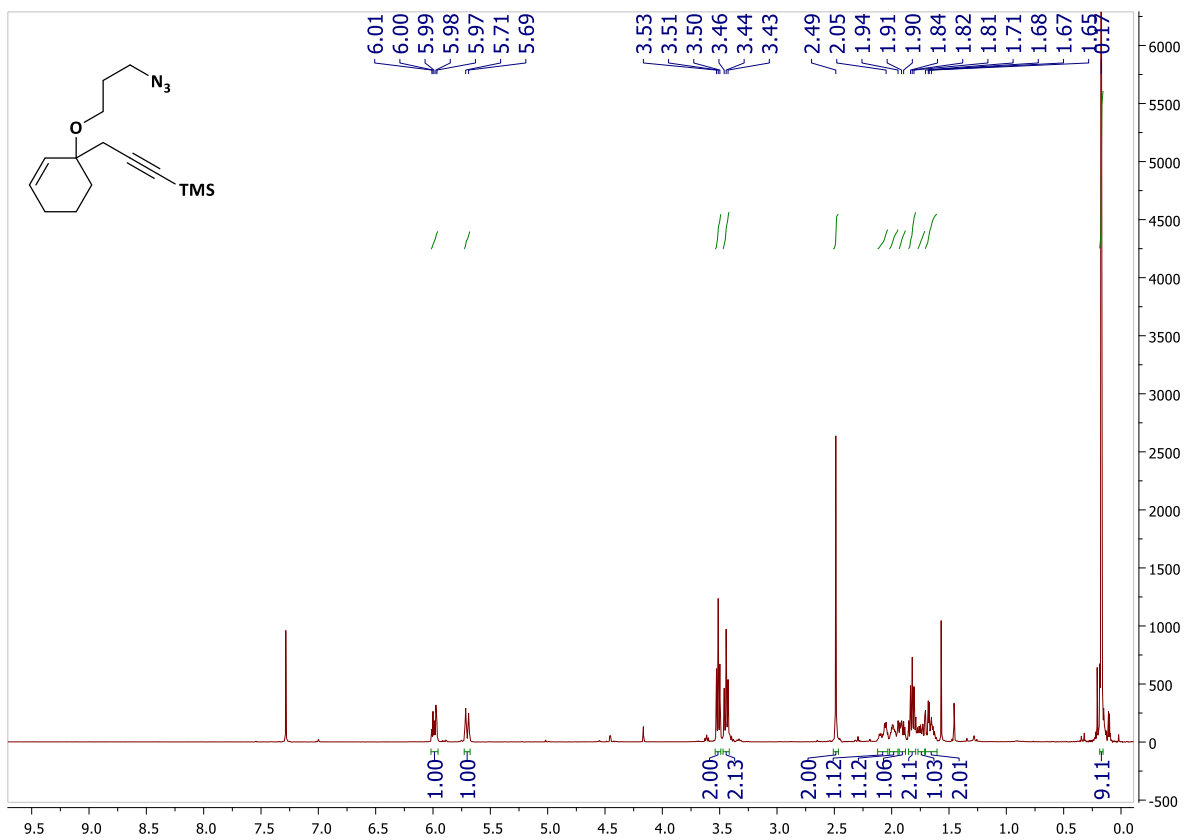


**Supplementary Figure S54.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.7.

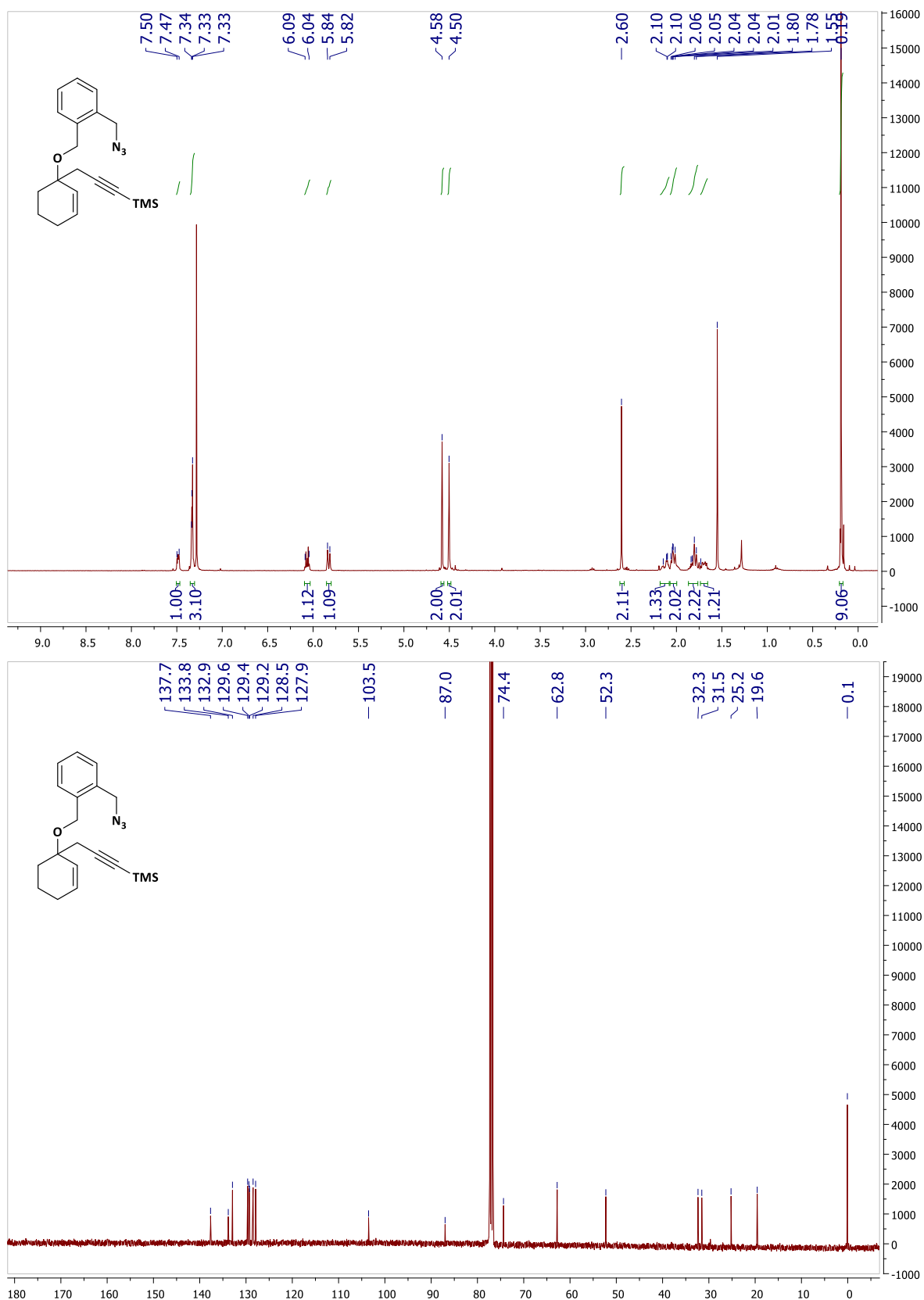


**Supplementary Figure S55.**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.1.8.



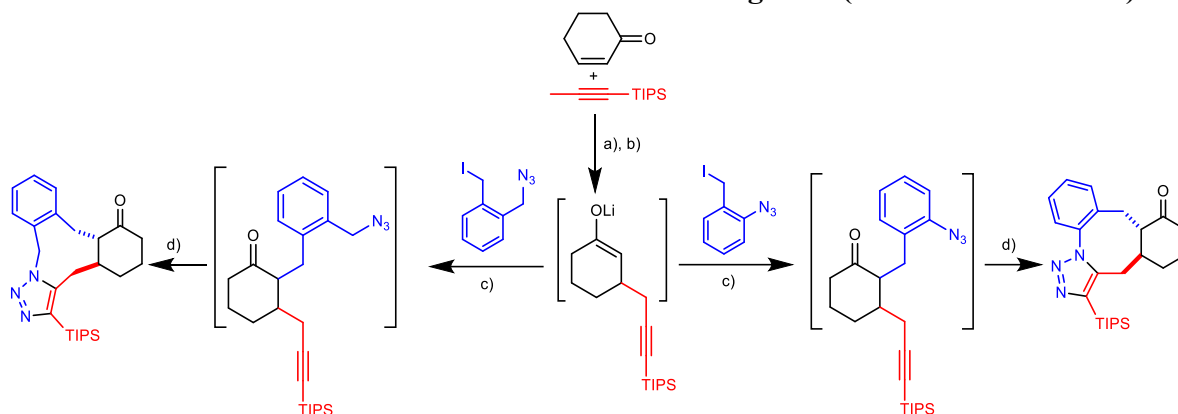


**Supplementary Figure S56.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.9.



**Supplementary Figure S57.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.10.

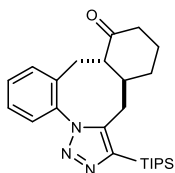
## Section S6.2. Reaction Mach2 described in main-text Figure 3 (variant with HMPA).



**Supplementary Figure S58. Reaction Mach2 described in main-text Figure 3 (variant with HMPA).** Reagents and conditions: (a) *n*BuLi (1.6 M), THF, -20 °C, 15 min; (b) Cyclohexen-1-one, THF/HMPA, -78 °C to -50 °C, 45 min; (c) THF, -78 °C, 24 h; (d) MW, THF/Dioxane, 145 °C, 1 h.

### One-pot procedure

To a mixture of 1-(triisopropylsilyl)-1-propyne (0.137 g, 0.7 mmol, 1.177 equiv) in anhydrous THF (1 mL) at -20 °C, *n*BuLi (0.48 mL, 1.6 M in THF, 0.77 mmol, 1.30 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.067 g, 0.7 mmol, 1.177 equiv) in THF/HMPA (3/1.33 mL) at -78 °C. The reaction mixture was then allowed to warm up to maximum -50 °C over 45 min before it was cooled again to -78 °C, when the azido-substrate (0.595 mmol, 1 equiv) in THF (1.5 mL) was added by using a syringe pump (1 drop/20 sec). After the complete addition, the reaction was left to stir for 24 h at -78 °C, and 1,4-Dioxane (10-15 mL) was added over it. The mixture was then heated in a sealed vial in a microwave reactor for a specified time and at a certain temperature. The mixture was then quenched at 0 °C with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with EtOAc. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography.



**(4aS,8aS)-3-(triisopropylsilyl)-4a,5,6,7,8a,9-hexahydrodibenzo[*d,g*][1,2,3]triazolo[1,5-*a*]azocin-8(4*H*)-one** and **(4aR,8aR)-3-(triisopropylsilyl)-4a,5,6,7,8a,9-hexahydrodibenzo[*d,g*][1,2,3]triazolo[1,5-*a*]azocin-8(4*H*)-one (1a)**

Following the general procedure using 1-azido-2-(iodomethyl)benzene as azido substrate, the MW vial was heated in a microwave reactor for 4 h at 185 °C. The crude mixture was purified

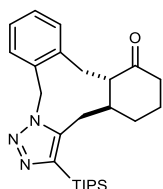
by column chromatography (hexane/EtOAc 9:1) to afford the final product **1a** (35 mg, 14%) as an oily liquid oil, yellow liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.50-7.49 (dd, *J* = 1.27, 7.78 Hz, 1H), 7.44-7.41 (td, *J* = 1.39, 7.52 Hz, 1H), 7.38-7.36 (td, *J* = 1.51, 7.65 Hz, 1H), 7.30-7.28 (dd, *J* = 1.33, 7.63 Hz, 1H), 3.55-3.53 (d, *J* = 14.46 Hz, 1H), 3.02-2.99 (d, *J* = 14.72 Hz, 1H), 2.50-2.48 (m, 1H), 2.39-2.29 (m, 2H), 2.25-2.21 (m, 1H), 2.17-2.10 (m, 1H), 1.92-1.90 (m, 1H), 1.84-1.80 (dd, *J* = 9.42, 14.5 Hz, 1H), 1.76-1.71 (m, 1H), 1.70-1.62 (m, 2H), 1.50-1.45 (m, 3H), 1.14-1.11 (dd, *J* = 7.49, 16.05 Hz, 18H);

**<sup>13</sup>C NMR** (400 MHz, CDCl<sub>3</sub>) δ 209.2, 143.6, 138.8, 137.5, 135.0, 130.2, 130.11, 127.5, 126.5, 58.4, 46.7, 41.1, 34.5, 30.1, 28.5, 25.9, 18.8, 11.7;

**IR** (film, CDCl<sub>3</sub>) 2942, 2865, 1712, 1499, 1463, 1250, 883, 764, 678 cm<sup>-1</sup>;

**HRMS** (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>38</sub>N<sub>3</sub>OSi 424.2784; Found 424.2792.



**(4aS,8aS)-3-(triisopropylsilyl)-4a,6,7,8a,9,14-hexahydro-4H-dibenzo[d,g][1,2,3]triazolo [1,5-a]azonin-8(5H)-one (1b) and (4aR,8aR)-3-(triisopropylsilyl)-4a,6,7,8a,9,14-hexahydro-4H-dibenzo[d,g][1,2,3]triazolo [1,5-a]azonin-8(5H)-one**

Following the general procedure using 1-(azidomethyl)-2-(iodomethyl)benzene as azido substrate, the MW vial was heated in a microwave reactor for 1 h at 145 °C. The crude mixture was purified by column chromatography (hexane/EtOAc 9:1) to afford the final product **1b** (57 mg, 22%) as an oil, yellow liquid.

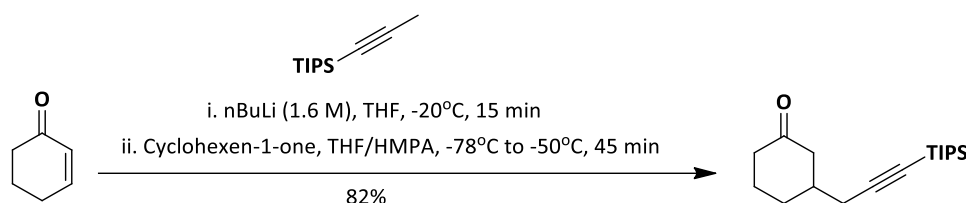
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.39 (m, 1H), 7.25-7.23 (m, 1H), 7.15-7.13 (m, 2H), 5.54-5.45 (q, *J* = 15.27 Hz, 2H), 3.26-3.22 (m, 1H), 3.07-3.04 (m, 1H), 2.76-2.72 (m, 1H), 2.47-2.36 (m, 2H), 2.19-2.10 (m, 2H), 2.01-1.98 (m, 1H), 1.96-1.92 (m, 1H), 1.72 (s, 1H), 1.57-1.46 (m, 2H), 1.40-1.31 (m, 3H), 1.05-1.00 (dd, *J* = 17.87, 7.48 Hz, 18H);

**<sup>13</sup>C NMR** (400 MHz, CDCl<sub>3</sub>) δ 209.2, 141.8, 138.8, 133.3, 131.6, 130.5, 128.9 (2C), 127.0, 56.1 (2C), 50.0, 41.3, 33.8, 31.0, 24.3, 18.9, 18.8, 11.9;

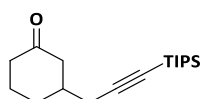
**IR** (film, CDCl<sub>3</sub>) 2944, 2865, 2249, 1708, 1460, 1229, 911, 734, 678, 524 cm<sup>-1</sup>;

**HRMS** (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. for C<sub>26</sub>H<sub>40</sub>N<sub>3</sub>OSi 438.2941; Found 438.2946.

## Mechanistic validation by stepwise isolation



**Supplementary Figure S59.** Reaction scheme of the stepwise process to confirm the first step intermediate.



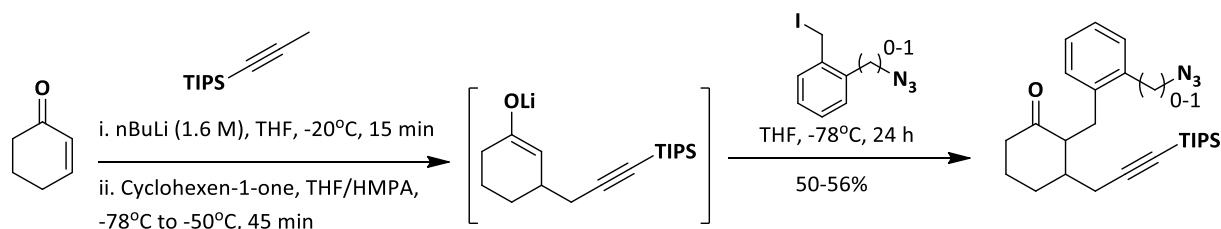
### 3-(3-(triisopropylsilyl)prop-2-yn-1-yl)cyclohexan-1-one **S6.2.1**<sup>50</sup>

To a mixture of 1-(triisopropylsilyl)-1-propyne (0.137 g, 0.7 mmol, 1 equiv) in anhydrous THF (1 mL) at  $-20^\circ\text{C}$ ,  $n\text{BuLi}$  (0.48 mL 1.6 M in THF, 0.77 mmol, 1.1 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.067 g, 0.7 mmol, 1 equiv) in THF/HMPA (3/1.33 mL) at  $-78^\circ\text{C}$ . The reaction was then allowed to warm up to maximum  $-50^\circ\text{C}$  over 45 min.

The crude mixture was quenched at  $0^\circ\text{C}$  with sat. aq. solution of  $\text{NH}_4\text{Cl}$ , and extracted with EtOAc. The combined organic phase was dried over  $\text{MgSO}_4$  and concentrated to dryness. The crude mixture was purified by column chromatography (hexane/EtOAc 8:2) to afford the first-step product<sup>50</sup> **S6.2.1** (0.168 g, 82%) as an oily liquid oil, yellow liquid.

**$^1\text{H}$  NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  2.40-2.36 (m, 1H), 2.32-2.29 (dd,  $J = 4.30, 10.10$  Hz, 1H), 2.27-2.24 (m, 1H), 2.23-2.13 (m, 2H), 2.03-1.99 (m, 1H), 1.96-1.91 (m, 1H), 1.89-1.86 (m, 1H), 1.77-1.69 (m, 1H), 1.64-1.58 (m, 1H), 1.55-1.47 (m, Hz, 1H), 1.00 (m, 21H);

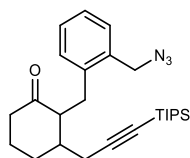
**$^{13}\text{C}$  NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  211.2, 105.4, 82.8, 47.0, 41.1, 38.0, 30.0, 26.9, 24.8, 18.6, 11.2.



**Supplementary Figure S60.** Reaction scheme of the stepwise process to confirm second step intermediate of the 1,4-addition- $\alpha$ -alkylation.

**General procedure of the cascade 1,4-addition- $\alpha$ -alkylation of (2-(azidomethyl)benzyl) and 1,4-addition- $\alpha$ -alkylation of (2-azidobenzyl):**

To a mixture of 1-(triisopropylsilyl)-1-propyne (0.137 g, 0.7 mmol, 1.177 equiv) in anhydrous THF (1 mL) at -20 °C, *n*BuLi (0.48 mL, 1.6 M in THF, 0.77 mmol, 1.30 equiv) was added dropwise. The reaction was stirred for 15 min at that temperature before it was added over a mixture of cyclohexen-1-one (0.067 g, 0.7 mmol, 1.177 equiv) in THF/HMPA (3/1.33 mL) at -78 °C. The reaction mixture was then allowed to warm up to maximum -50 °C over 45 min before it was cooled again to -78 °C, when the azido-substrate (0.595 mmol, 1 equiv) in THF (1.5 mL) was added by using a syringe pump (1 drop/20 sec). After the complete addition, the reaction was left to stir for 24 h at -78 °C. The crude mixture was quenched at 0 °C with sat. aq. solution of NH<sub>4</sub>Cl, and extracted with EtOAc. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated to dryness. The crude mixture was purified by column chromatography.



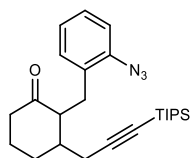
**2-(2-(azidomethyl)benzyl)-3-(3-(triisopropylsilyl)prop-2-yn-1-yl)cyclohexan-1-one S6.2.2**

Following the general procedure using 1-(azidomethyl)-2-(iodomethyl)benzene as azido substrate, the crude mixture was purified by column chromatography (hexane/EtOAc 8.5:1.5) to afford the second-step product **S6.2.2** (146 mg, 56%) as an oily liquid oil, yellow liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$  7.33-7.31 (d, *J* = 7.30 Hz, 1H), 7.28-7.24 (m, 2H), 7.23-7.18 (m, 1H), 4.78-4.75 (d, *J* = 13.54 Hz, 1H), 4.51-4.48 (d, *J* = 13.56 Hz, 1H), 3.13-3.07 (dd, *J* = 9.46, 14.30 Hz, 1H), 2.99-2.94 (m, 1H), 2.75-2.71 (dd, *J* = 2.40, 14.43 Hz, 1H), 2.68-2.62 (dd, *J* = 5.47, 17.34 Hz, 1H), 2.56-2.51 (dd, *J* = 3.39, 17.36 Hz, 1H), 2.42-2.36 (m, 1H), 2.34-2.26 (m, 1H), 2.15-2.08 (m, 1H), 2.02-1.96 (m, 2H), 1.93-1.86 (m, 1H), 1.78-1.68 (m, 1H), 1.11 (s, 21H);

**<sup>13</sup>C NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$  211.9, 139.7, 133.7, 129.8, 129.7, 128.3, 126.1, 104.8, 84.0, 55.9, 53.1, 43.2, 41.9, 30.4, 28.4, 25.8, 24.8, 18.6, 11.3;

**HRMS** (ESI-HR) *m/z*: [M]<sup>+</sup> Calc. for C<sub>26</sub>H<sub>40</sub>N<sub>3</sub>OSi 438.2941; Found 438.2944.



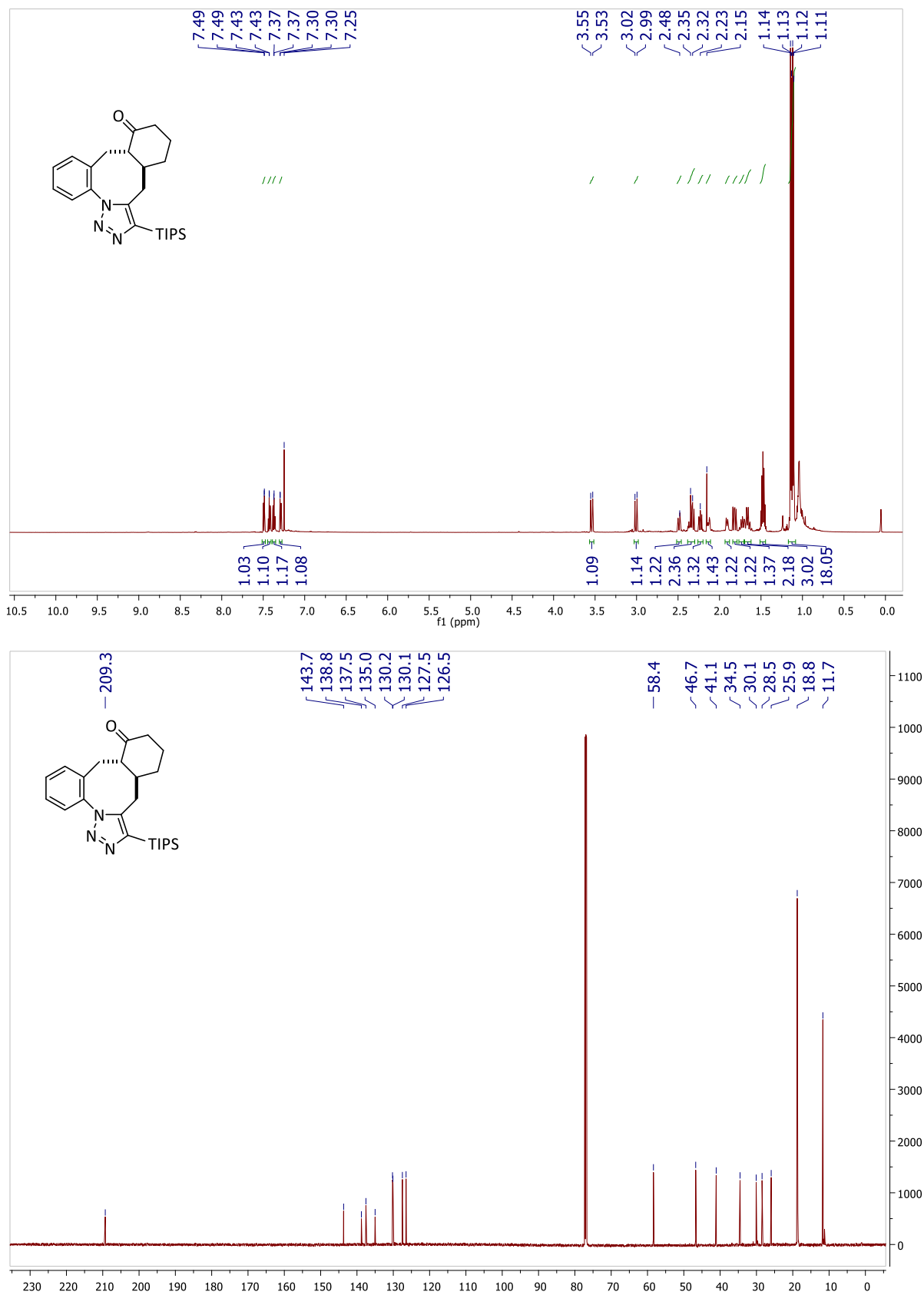
### 2-(2-azidobenzyl)-3-(3-(triisopropylsilyl)prop-2-yn-1-yl)cyclohexan-1-one S6.2.3

Following the general procedure using 1-azido-2-(iodomethyl)benzene as azido substrate, the crude mixture was purified by column chromatography (hexane/EtOAc 8.5:1.5) to afford the second-step product **S6.2.3** (126 mg, 50%) as an oil, yellow liquid.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.54-7.52 (dd,  $J = 1.23, 7.70$  Hz, 1H), 7.48-7.38 (dtd,  $J = 1.47, 7.47, 23.65$  Hz, 2H), 7.34-7.29 (m, 1H), 3.60-3.56 (d,  $J = 14.41$  Hz, 1H), 3.06-3.03 (d,  $J = 14.79$  Hz, 1H), 2.54-2.51 (m, 1H), 2.43-2.33 (m, 2H), 2.29-2.24 (m, 1H), 2.19-2.14 (m, 1H), 1.96-1.91 (m, 1H), 1.88-1.82 (dd,  $J = 9.41, 14.45$  Hz, 1H), 1.78-1.66 (m, 3H), 1.55-1.48 (m, 3H), 1.19-1.14 (dd,  $J = 7.48, 10.42$  Hz, 18H);

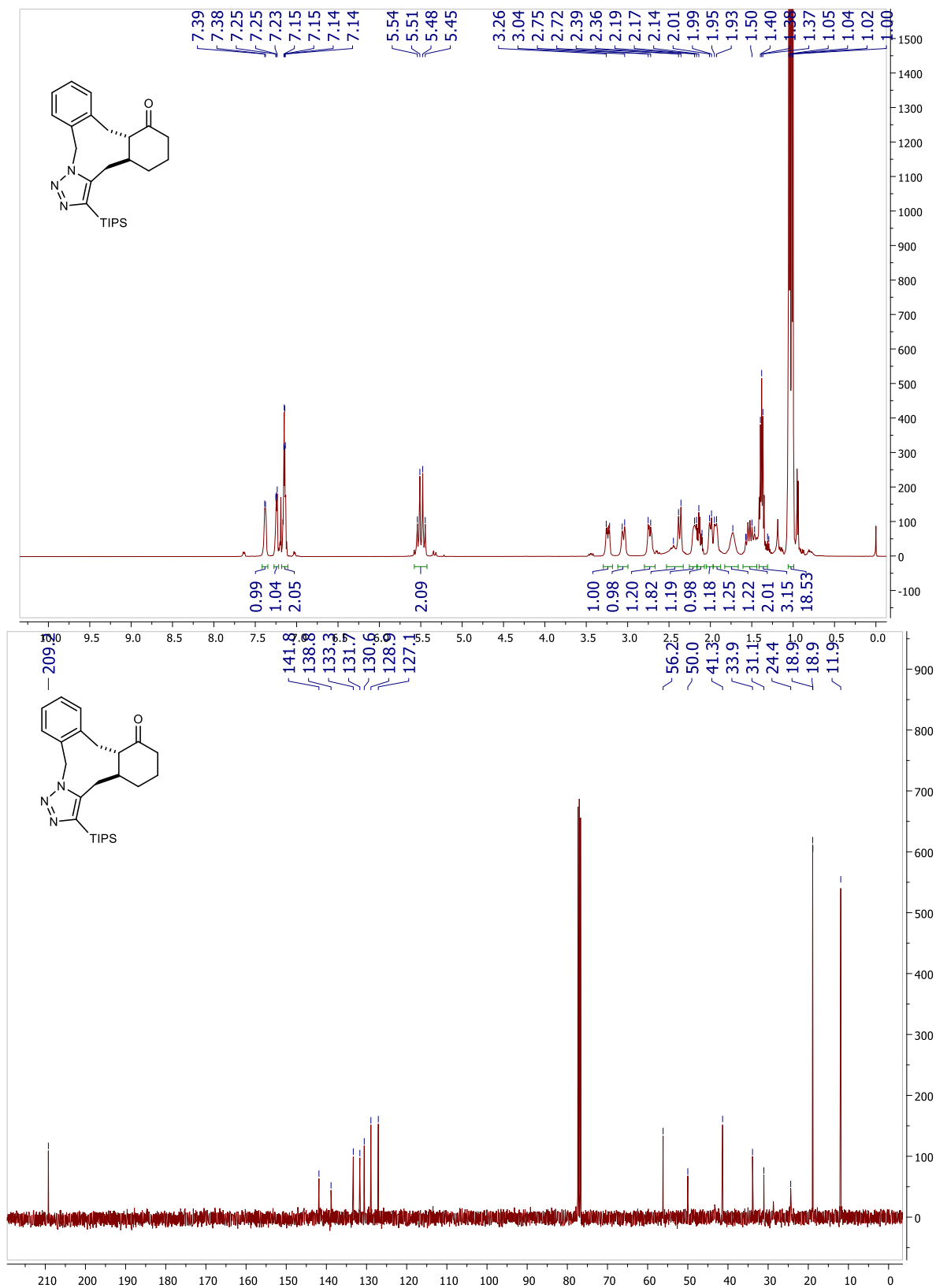
**<sup>13</sup>C NMR**. (400 MHz, CDCl<sub>3</sub>) δ 211.9, 137.9, 132.0, 131.2, 127.4, 124.6, 117.9, 105.7, 83.0, 55.0, 43.0, 41.5, 29.7, 28.3, 25.0, 24.9, 18.6, 11.3;

**HRMS** (ESI-HR)  $m/z$ : [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>38</sub>N<sub>3</sub>OSi 424.2784; Found 424.2792.

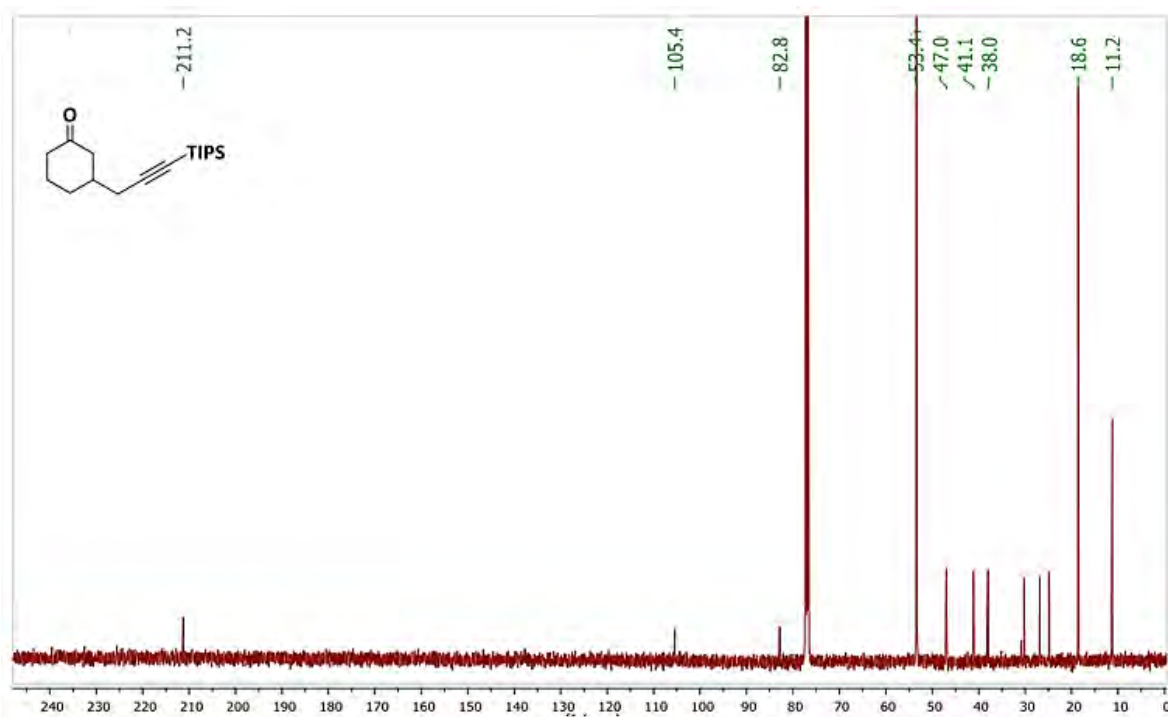
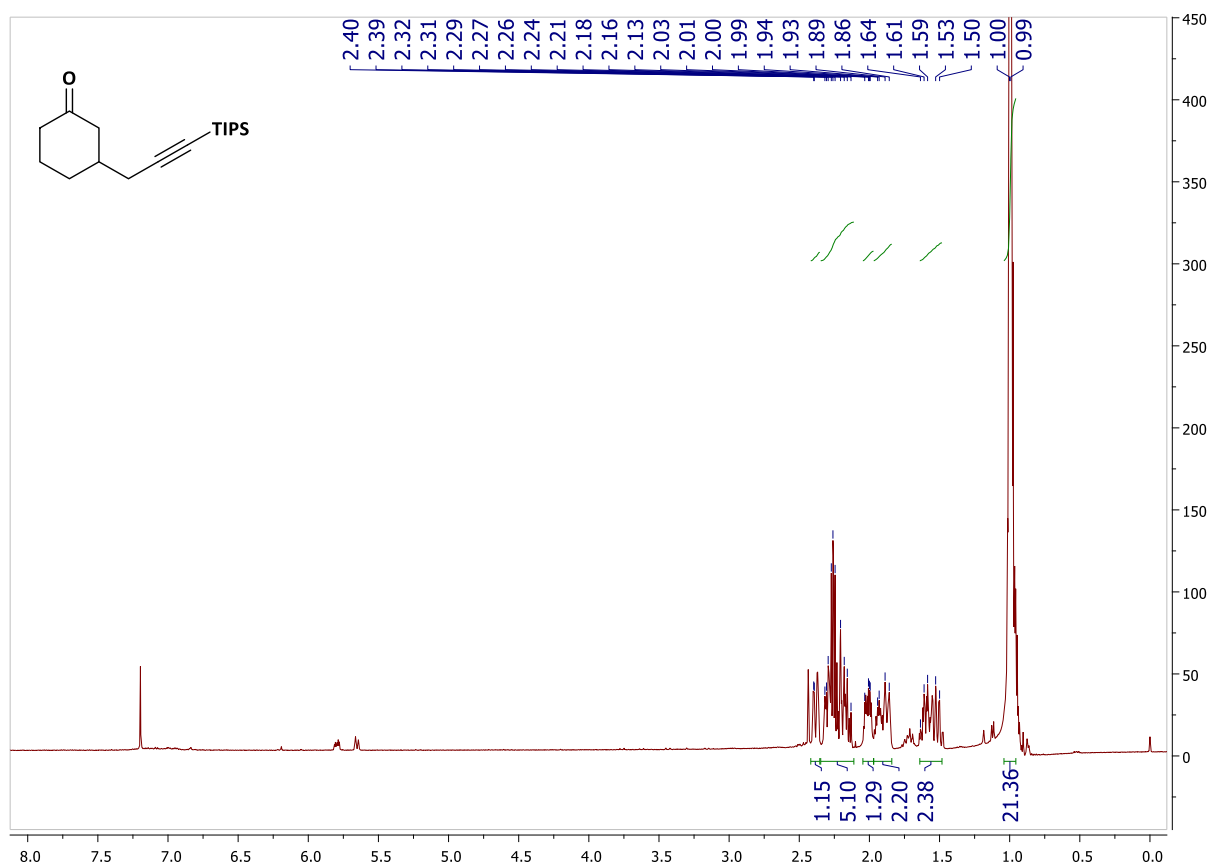


Supplementary Figure S61. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **1a**.

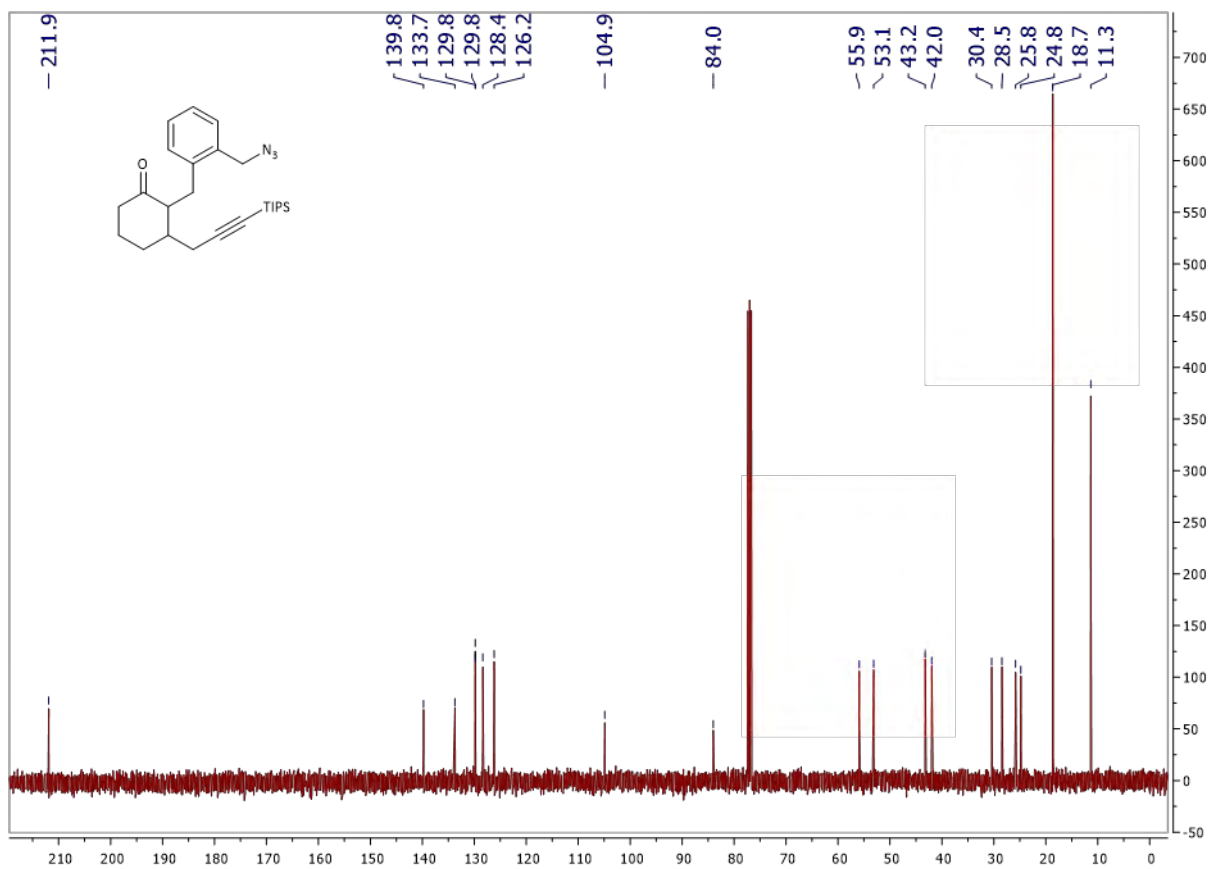
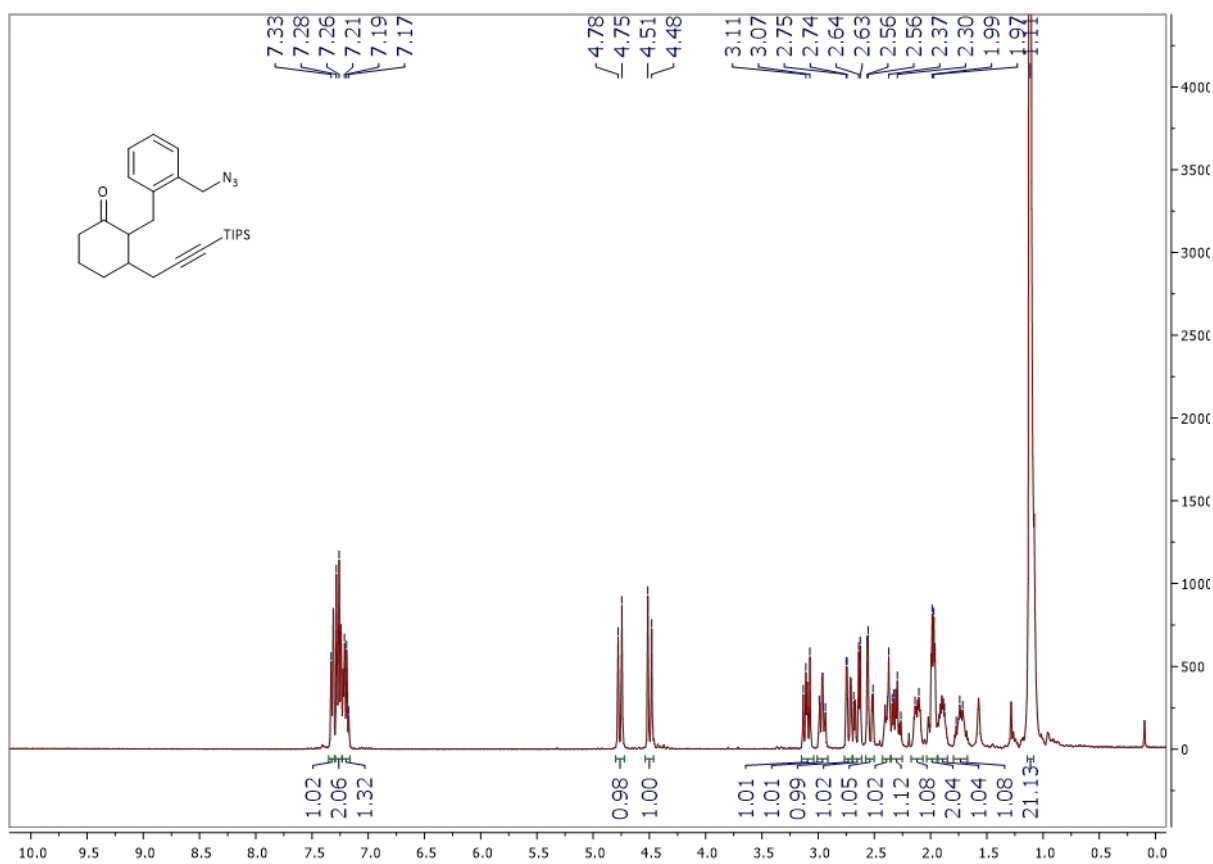




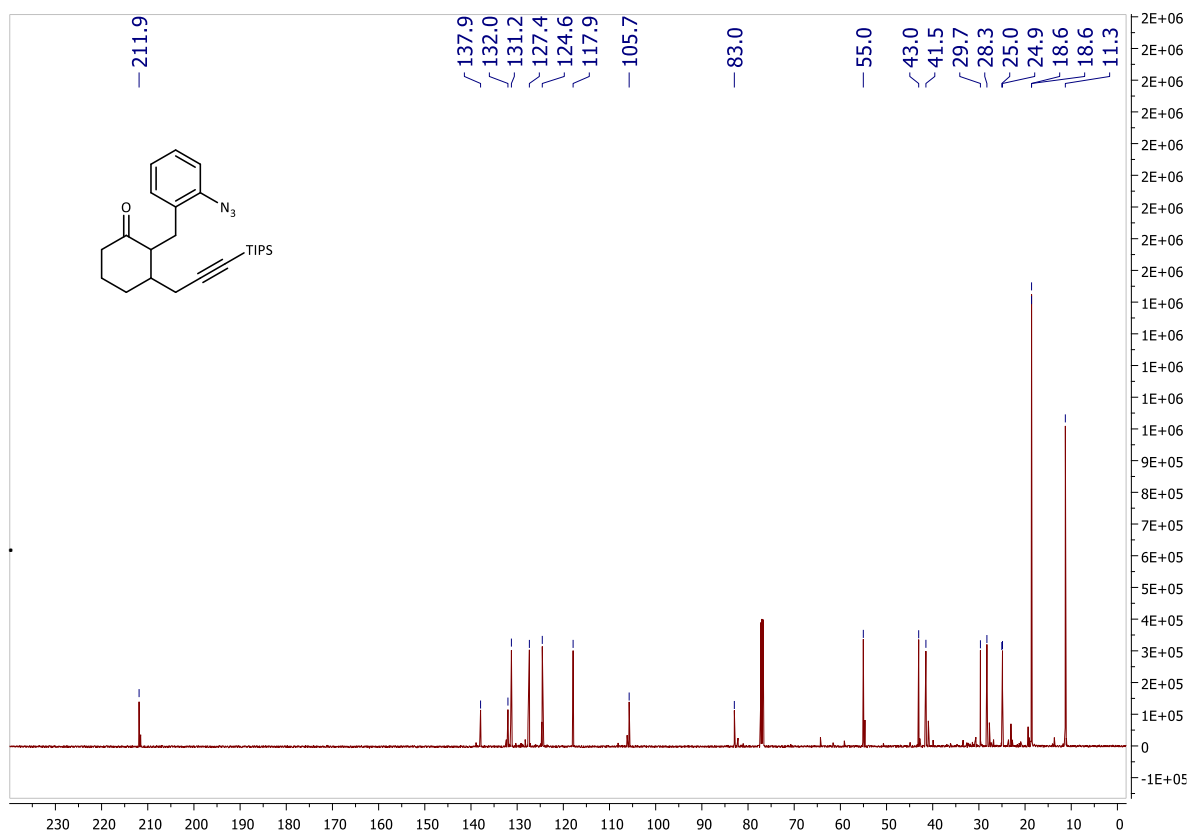
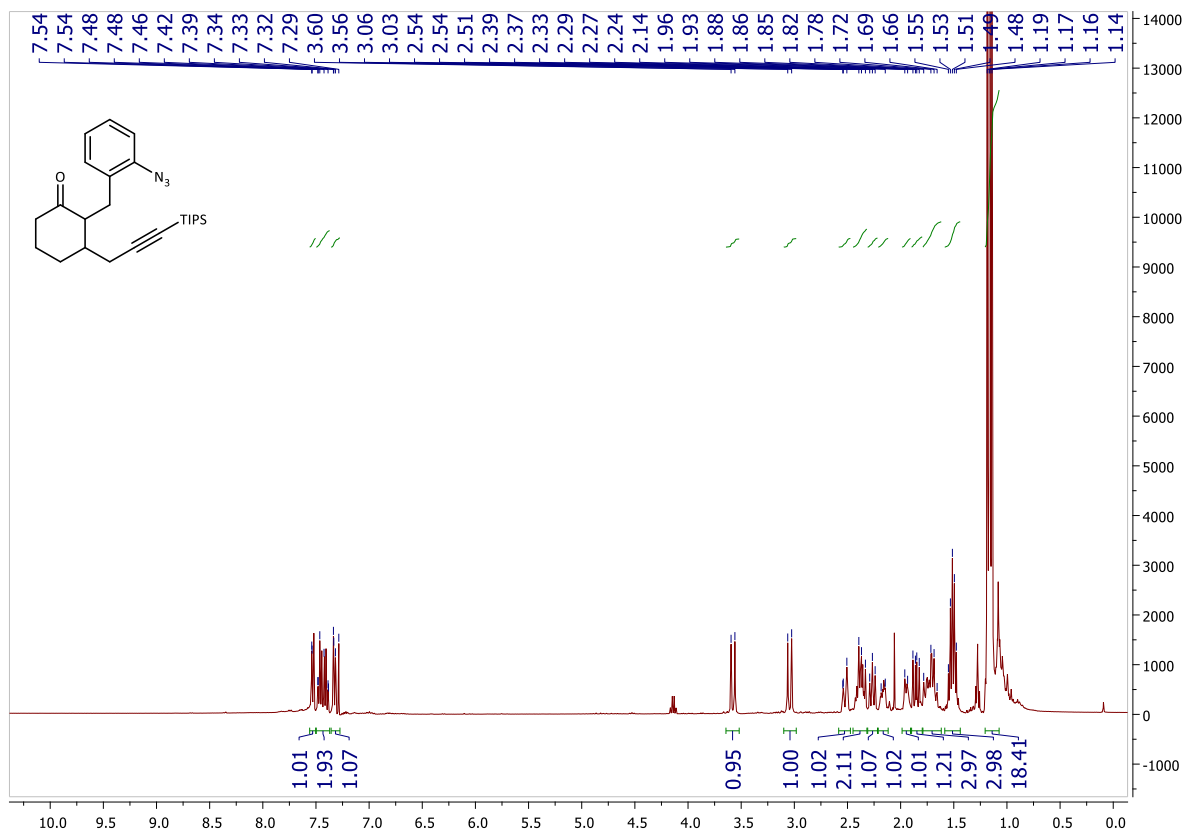
Supplementary Figure S62. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **1b**.



**Supplementary Figure S63.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.1.

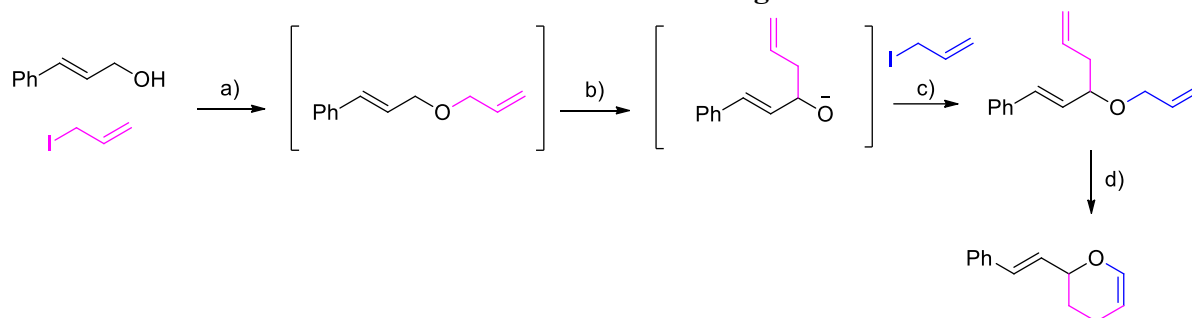


**Supplementary Figure S64.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.2.



**Supplementary Figure S65.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.3.

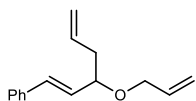
### Section S6.3 Reaction Mach3 described in main-text Figure 4a.



**Supplementary Figure S66. Reaction Mach3 described in main-text Figure 4a.** Reagents and conditions: (a) NaH, THF, 0 °C to rt, 1 h; (b) *n*BuLi, -78 °C, 0.5 h, 93%; (c) i. HMPA, rt, 0.5 h; ii. rt, 1 h, 66%; (d) i. Grubbs 1, PhMe, rt, 2.5 h; ii. NaOH, *i*PrOH, reflux, 14 h, 75%.

#### One-pot procedure

To a solution of cinnamyl, allyl or propargyl alcohol (1 mmol) in anhydrous THF (5 mL) at 0 °C was added NaH (60% in mineral oil, 60 mg, 1.5 mmol, 1.5 equiv) and the mixture was stirred for 30 min. Then allyl iodide (183  $\mu$ L, 2 mmol, 2 equiv) was added and the mixture was stirred at rt for 1 h. Next, reaction mixture was recooled to -78 °C and *n*BuLi (2.5 M in hexane, 0.8 mL, 2 mmol, 2 equiv) was added dropwise and the reaction mixture was stirred at -78 °C for 30 min, followed by the addition of HMPA (1.7 mL, 10 mmol, 10 equiv). Cooling bath was removed and the reaction mixture was stirred at rt for 0.5 h (till dark bloody red color). Next, allyl iodide (183  $\mu$ L, 2 mmol, 2 equiv) was added at once (becomes discolored) and the reaction mixture was stirred for 1 h. Upon completion, the reaction mixture was quenched sat. aq. solution of NH<sub>4</sub>Cl and warmed to rt. The aqueous layer was extracted with Et<sub>2</sub>O. The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated *in vacuo*. Crude product was purified by column chromatography.



#### (E)-3-(allyloxy)hexa-1,5-dien-1-ylbenzene **3a**

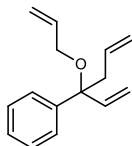
According to general procedure, the reaction was performed with cinnamyl alcohol as the starting material. After purification by column chromatography (pentane:Et<sub>2</sub>O 98:2) **3a** was obtained (142 mg, 66%) as a colorless oil.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  7.42 – 7.38 (m, 2H), 7.36 – 7.30 (m, 2H), 7.28 – 7.23 (m, 1H), 6.54 (d, *J* = 15.9 Hz, 1H), 6.10 (dd, *J* = 15.9, 7.9 Hz, 1H), 5.99 – 5.80 (m, 2H), 5.32 – 5.25 (m,

1H), 5.20 – 5.04 (m, 3H), 4.14 – 4.07 (m, 1H), 3.99 – 3.88 (m, 2H), 2.54 – 2.45 (m, 1H), 2.43 – 2.33 (m, 1H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 136.6, 135.1, 134.5, 132.3, 130.0, 128.6, 127.7, 126.5, 117.0, 116.7, 79.7, 69.3, 40.3;

The spectral data match those reported in the literature<sup>52</sup>.



### **(3-(allyloxy)hexa-1,5-dien-3-yl)benzene 3b**

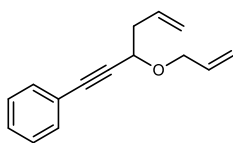
According to general procedure, the reaction was performed with  $\alpha$ -vinylbenzyl alcohol as the starting material. After purification by column chromatography (hexane:DCM 9:1) compound **3b** was obtained (176 mg, 82%) as a light yellow liquid.

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.45 – 7.41 (m, 2H), 7.37 – 7.31 (m, 2H), 7.28 – 7.24 (m, 1H), 6.01 (dd,  $J = 17.5, 10.9$  Hz, 1H), 5.94 (ddt,  $J = 17.2, 10.3, 5.2$  Hz, 1H), 5.75 – 5.66 (m, 1H), 5.37 – 5.30 (m, 3H), 5.16 – 5.12 (m, 1H), 5.05 – 5.02 (m, 1H), 5.01 – 4.99 (m, 1H), 3.88 – 3.80 (m, 2H), 2.85 – 2.74 (m, 2H);

<sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 142.9, 141.2, 135.4, 133.5, 128.0, 127.0, 126.8, 117.5, 115.5, 115.4, 81.1, 63.8, 42.1;

IR (film, CH<sub>2</sub>Cl<sub>2</sub>) 3079, 3015, 2980, 2917, 2859, 1953, 1838, 1683, 1641, 1600, 1492, 1446, 1409, 1378, 1360, 1312, 1289, 1259, 1197, 1181, 1123, 1064, 1032, 992, 917 cm<sup>-1</sup>;

**Anal.** Calcd for C<sub>15</sub>H<sub>18</sub>O: C, 84.07; H, 8.47. Found: C, 84.06; H, 8.50.



### **(3-(allyloxy)hex-5-en-1-yn-1-yl)benzene 3c**

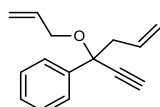
According to general procedure, the reaction was performed with 3-phenyl-2-propyn-1-ol as the starting material. After purification by column chromatography (hexane/acetone 99:1) compound **3c** was obtained (153 mg, 72%) as a yellow liquid.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.47 – 7.41 (m, 2H), 7.34 – 7.29 (m, 3H), 6.01 – 5.92 (m, 2H), 5.38 – 5.32 (m, 1H), 5.24 – 5.12 (m, 3H), 4.38 – 4.31 (m, 2H), 4.10 – 4.04 (m, 1H), 2.64 – 2.56 (m, 2H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 134.4, 133.7, 131.8, 128.3, 128.2, 122.7, 117.7, 117.4, 87.7, 86.2, 69.7, 68.9, 40.3;

**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3079, 3018, 2979, 2954, 2925, 2857, 2229, 1950, 1841, 1726, 1681, 1642, 1598, 1573, 1490, 1442, 1428, 1337, 1278, 1256, 1124, 1080, 995, 917, 756, 691 cm<sup>-1</sup>;

**Anal.** Calcd for C<sub>15</sub>H<sub>16</sub>O: C, 84.87; H, 7.60. Found: C, 84.79; H, 7.68.



### **(3-(allyloxy)hex-5-en-1-yn-3-yl)benzene 3d**

According to general procedure, the reaction was performed with 1-phenyl-2-propyn-1-ol as the starting material. After purification by column chromatography (hexane/EtOAc 99:1) compound **3d** was obtained (204 mg, 96%) as a yellow liquid.

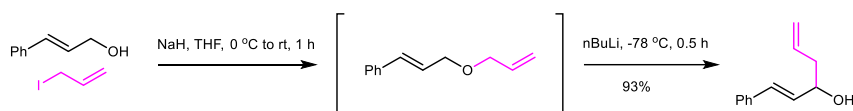
**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.61 – 7.55 (m, 2H), 7.39 – 7.33 (m, 2H), 7.33 – 7.28 (m, 1H), 5.97 - 5.88 (m, 1H), 5.82 - 5.72 (m, 1H), 5.32 - 5.26 (m, 1H), 5.17 – 5.11 (m, 1H), 5.05 – 4.98 (m, 2H), 4.17 - 4.11 (m, 1H), 3.72 - 3.66 (m, 1H), 2.84 – 2.78 (m, 1H), 2.76 (s, 1H), 2.66 – 2.59 (m, 1H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 140.9, 134.8, 132.8, 128.2, 127.9, 126.5, 118.3, 116.2, 82.6, 79.1, 76.8, 66.0, 49.4;

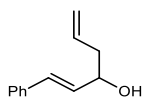
**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3303, 3079, 3017, 2955, 2925, 2856, 2242, 2110, 1954, 1837, 1728, 1679, 1643, 1601, 1489, 1447, 1424, 1378, 1361, 1314, 1286, 1261, 1202, 1178, 1124, 1057, 992, 918, 866, 763, 702, 660, 638 cm<sup>-1</sup>;

**Anal.** Calcd for C<sub>15</sub>H<sub>16</sub>O: C, 84.87; H, 7.60. Found: C, 84.81; H, 7.72.

### **Mechanistic validation by stepwise isolation**



**Supplementary Figure S67.** Reaction scheme of the stepwise process to confirm the second step intermediate.



**(E)-1-phenylhexa-1,5-dien-3-ol S6.3.1** <sup>53</sup>

To a solution of cinnamyl alcohol (129  $\mu$ L, 1 mmol) in anhydrous THF (5 mL) at 0  $^{\circ}$ C was added NaH (60% in mineral oil, 60 mg, 1.5 mmol, 1.5 equiv) and the mixture was stirred for 30 min. Then allyl iodide (183  $\mu$ L, 2 mmol, 2 equiv) was added and the mixture was stirred at rt for 1 h. Next, reaction mixture was recooled to -78  $^{\circ}$ C and *n*BuLi (2.5 M in hexane, 0.8 mL, 2 mmol, 2 equiv) was added dropwise and the reaction mixture was stirred at -78  $^{\circ}$ C for 30 min. Upon completion, the reaction mixture was quenched sat.  $\text{NH}_4\text{Cl}$  and warmed to rt. The aqueous layer was extracted with  $\text{Et}_2\text{O}$ . The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated *in vacuo*. Crude product was purified by column chromatography (hexane:EtOAc 9:1) to give **S6.3.1** (162 mg, 93%) as a colorless oil.

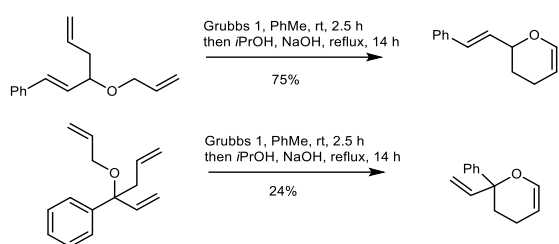
The spectral data match those reported in the literature<sup>53</sup>.

**$^1\text{H}$  NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.41 – 7.36 (m, 2H), 7.35 – 7.29 (m, 2H), 7.27 – 7.21 (m, 1H), 6.61 (dd,  $J$  = 15.9, 1.3 Hz, 1H), 6.25 (dd,  $J$  = 15.9, 6.3 Hz, 1H), 5.93 – 5.81 (m, 1H), 5.22 – 5.14 (m, 2H), 4.40 – 4.32 (m, 1H), 2.50 – 2.33 (m, 2H), 1.94 (d,  $J$  = 3.8 Hz, 1H);

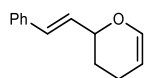
**$^{13}\text{C}$  NMR** (100 MHz,  $\text{CDCl}_3$ )  $\delta$  136.7, 134.1, 131.6, 130.4, 128.6, 127.7, 126.5, 118.4, 71.7, 42.0;



## Application examples



**Supplementary Figure S68.** Reaction scheme for the formation of dihydropyranes from the products of the Mach3 reaction.



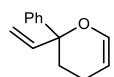
### (E)-2-styryl-3,4-dihydro-2H-pyran S6.3.2

To a solution of **3a** (42 mg, 0.2 mmol) in anhydrous PhMe (1.5 mL) at rt was added Grubbs 1 (8 mg, 0.01 mmol, 5 mol%) and the mixture was stirred for 2.5 h. Upon completion, *i*PrOH (0.38 mL) and NaOH (2 mg, 0.05 mmol, 25 mol%) were added and the mixture was refluxed for 14 h. Upon completion, the mixture was cooled to rt and then reaction mixture was directly transferred onto silica gel and purified by column chromatography (pentane/Et<sub>2</sub>O 98:2) to give **S6.3.2** (28 mg, 75%) as a colorless liquid.

The spectral data match those reported in the literature.<sup>53</sup>

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.40 (d, *J* = 7.5 Hz, 2H), 7.32 (t, *J* = 7.5 Hz, 2H), 7.26 - 7.23 (m, 1H), 6.64 (d, *J* = 16 Hz, 1H), 6.44 (d, *J* = 6.2 Hz, 1H), 6.27 (dd, *J* = 16, 6.2 Hz, 1H), 4.75 - 4.71 (m, 1H), 4.53 - 4.47 (m, 1H), 2.16 - 2.10 (m, 1H), 2.06 - 1.99 (m, 2H), 1.82 - 1.77 (m, 1H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 143.5, 136.7, 130.9, 129.0, 128.5, 127.7, 126.5, 100.6, 75.3, 28.3, 19.4;



### 2-phenyl-2-vinyl-3,4-dihydro-2H-pyran S6.3.3

To a solution of **3b** (139 mg, 0.65 mmol) in anhydrous PhMe (5 mL) at rt was added Grubbs 1 (26 mg, 0.033 mmol, 5 mol%) and the mixture was stirred for 2.5 h. Upon completion, *i*PrOH (0.8 mL) and NaOH (13 mg, 0.33 mmol, 50 mol%) were added and the mixture was refluxed

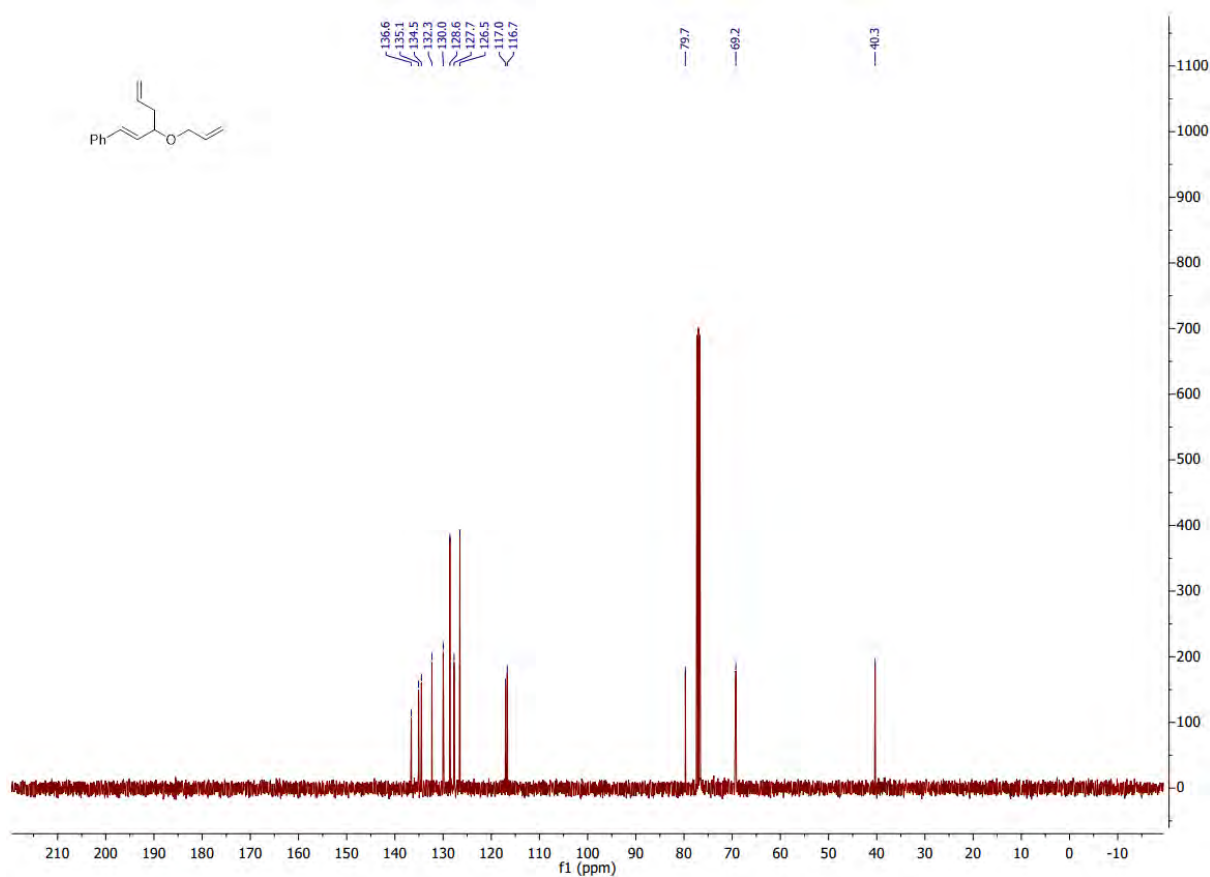
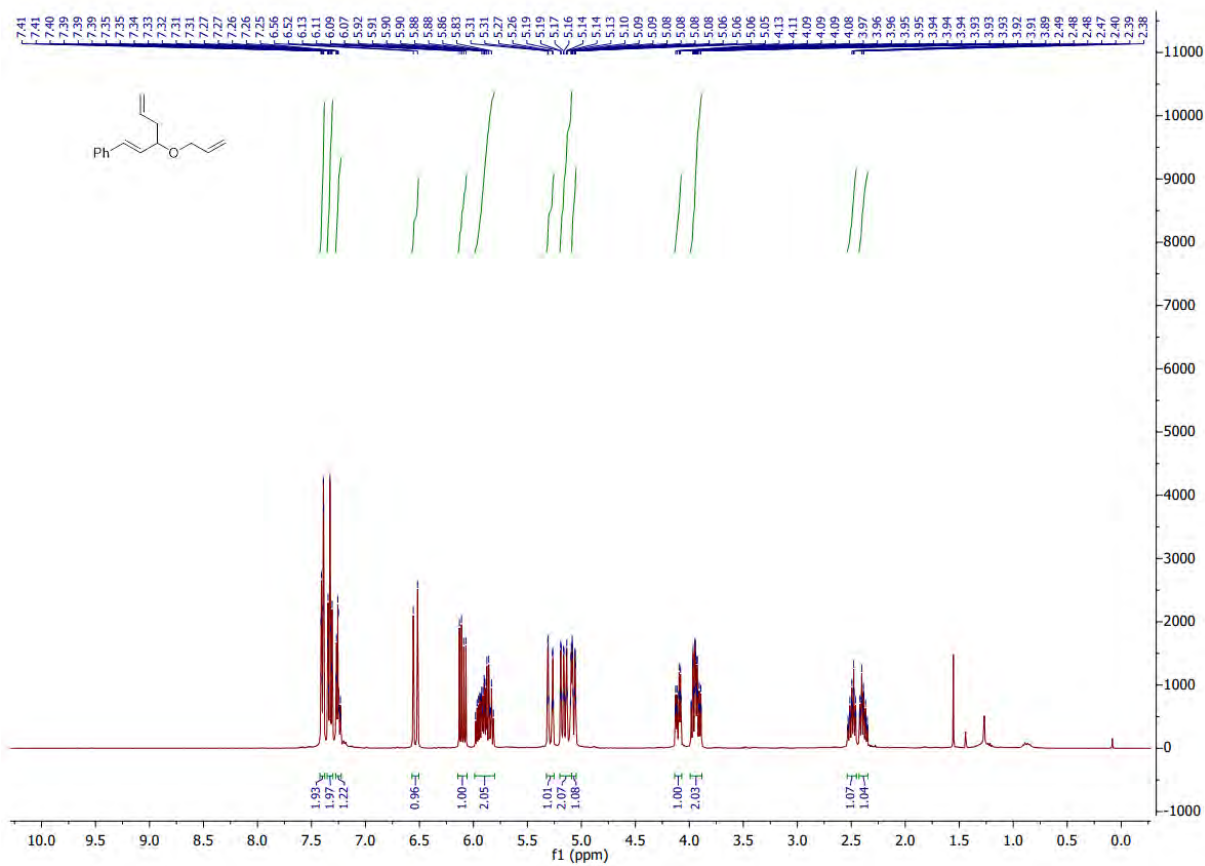
for 14 h. Upon completion, the mixture was cooled to rt and then reaction mixture was directly transferred onto silica gel and purified by column chromatography (hexane/EtOAc 99:1) to give **S6.3.3** (29 mg, 24%) as a colorless liquid.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.43 – 7.40 (m, 2H), 7.37 – 7.33 (m, 2H), 7.28 – 7.23 (m, 1H), 6.54 (m, 1H), 6.00 (dd, J = 17.3, 10.7 Hz, 1H), 5.24 – 5.13 (m, 2H), 4.68 (dt, J = 6.3, 3.8 Hz, 1H), 2.26 - 2.18 (m, 1H), 2.11 - 2.04 (m, 1H), 2.01 – 1.93 (m, 1H), 1.84 - 1.77 (m, 1H);

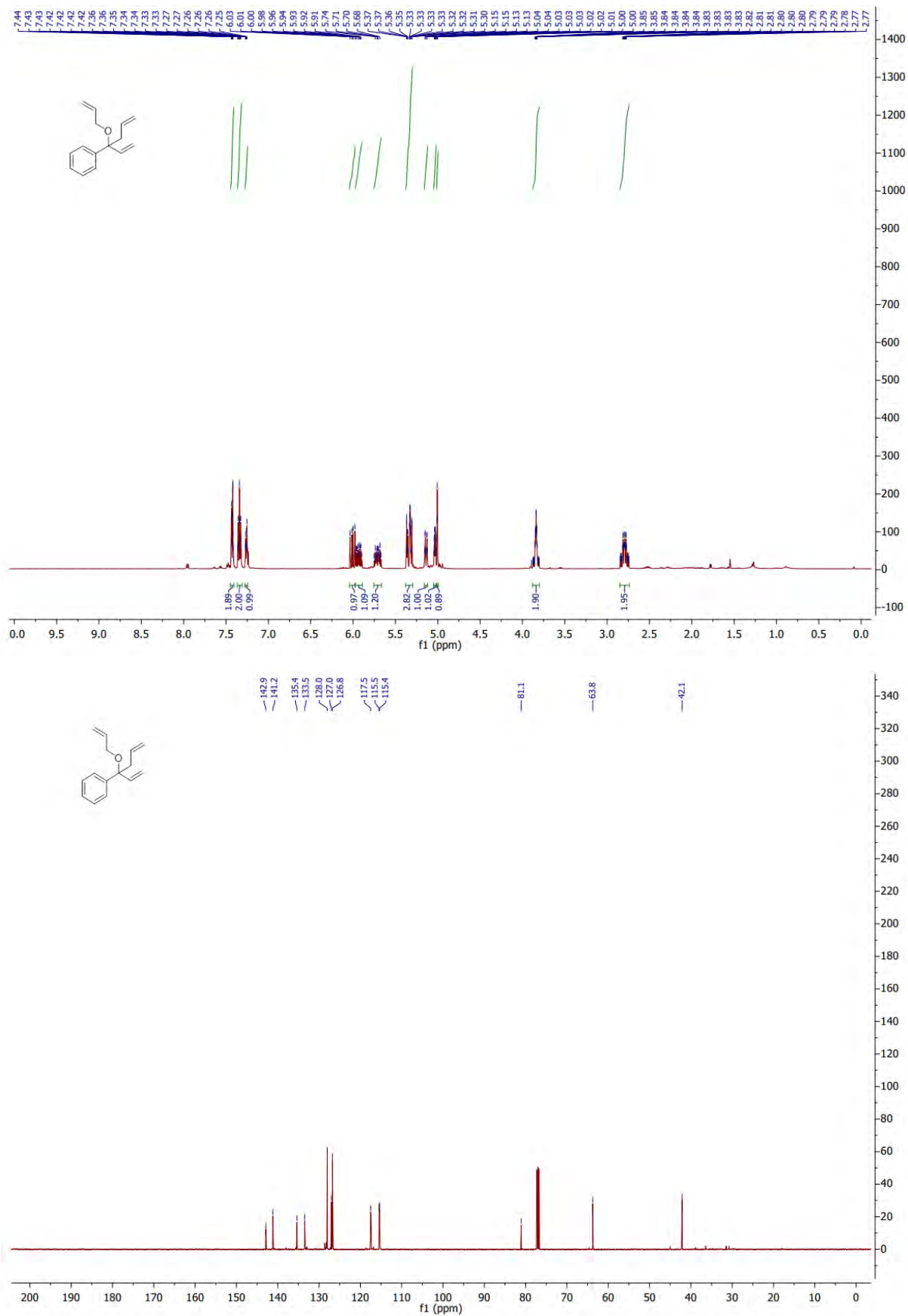
**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 143.5, 142.3, 141.3, 128.2, 127.0, 125.4, 113.7, 100.5, 79.6, 31.5, 17.7;

**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3432, 3086, 3060, 3031, 2925, 2854, 1958, 1885, 1724, 1681, 1599, 1493, 1448, 1408, 1382, 1252, 1179, 1085, 1014, 928, 873, 822, 762, 701 cm<sup>-1</sup>;

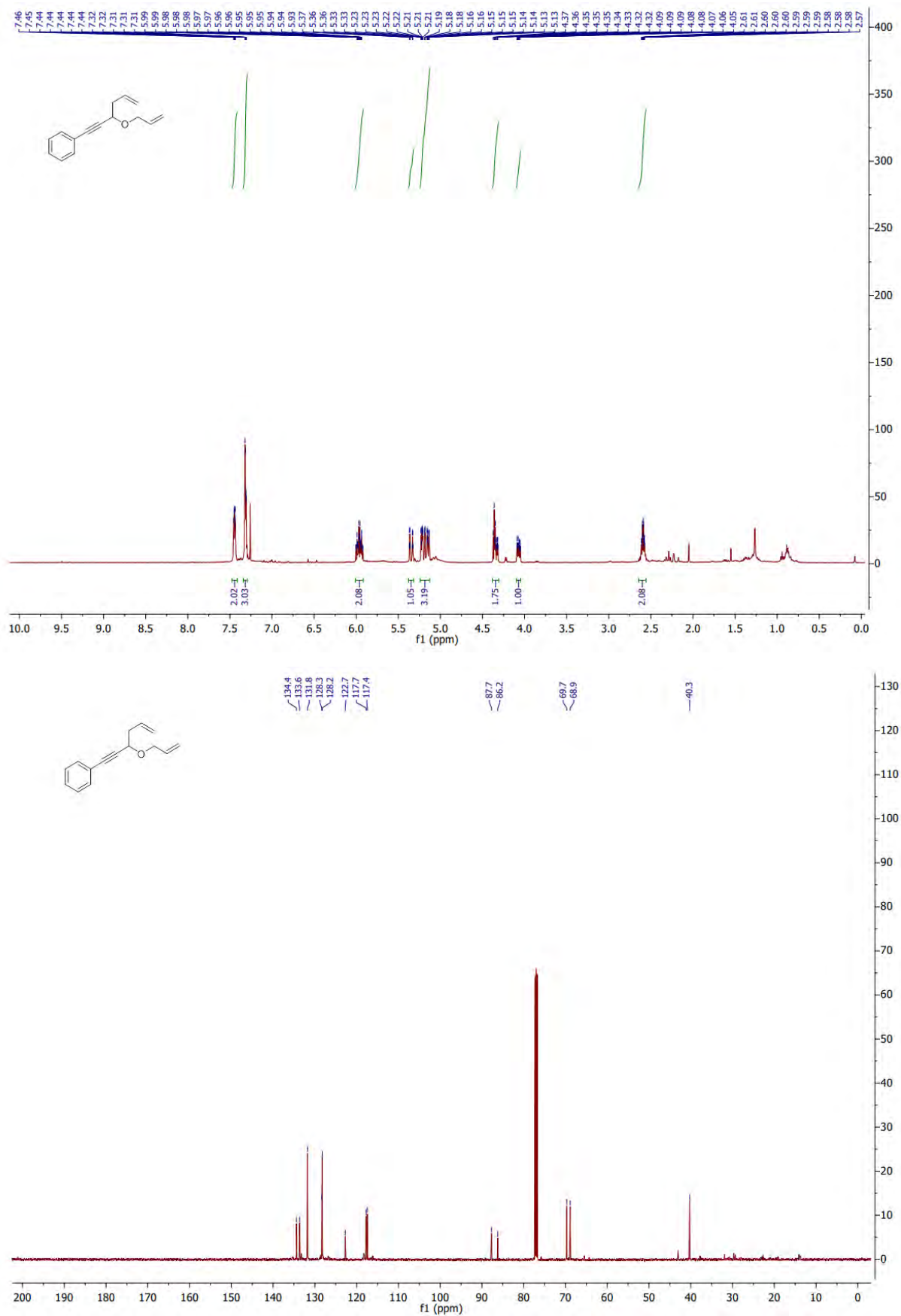
**Anal.** Calcd for C<sub>13</sub>H<sub>14</sub>O: C, 83.83; H, 7.58. Found: C, 83.82; H, 7.58.



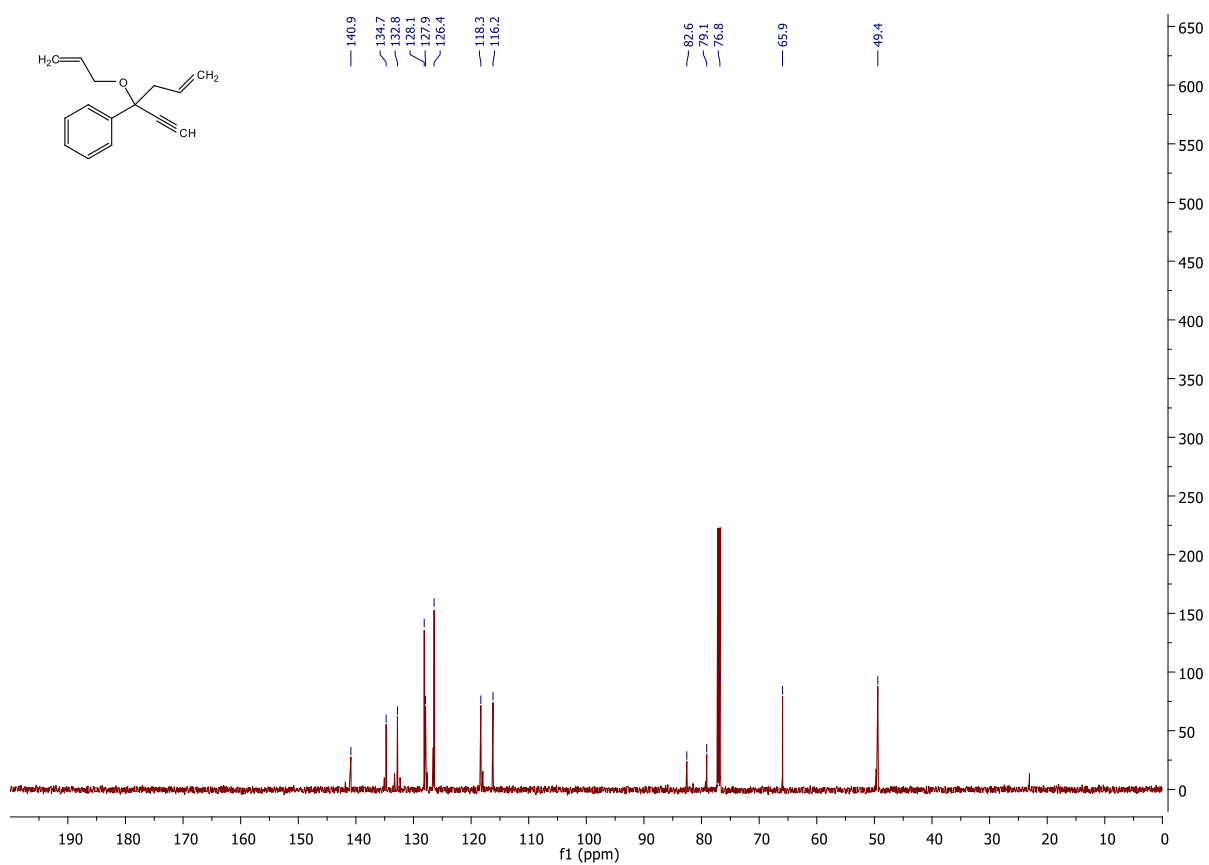
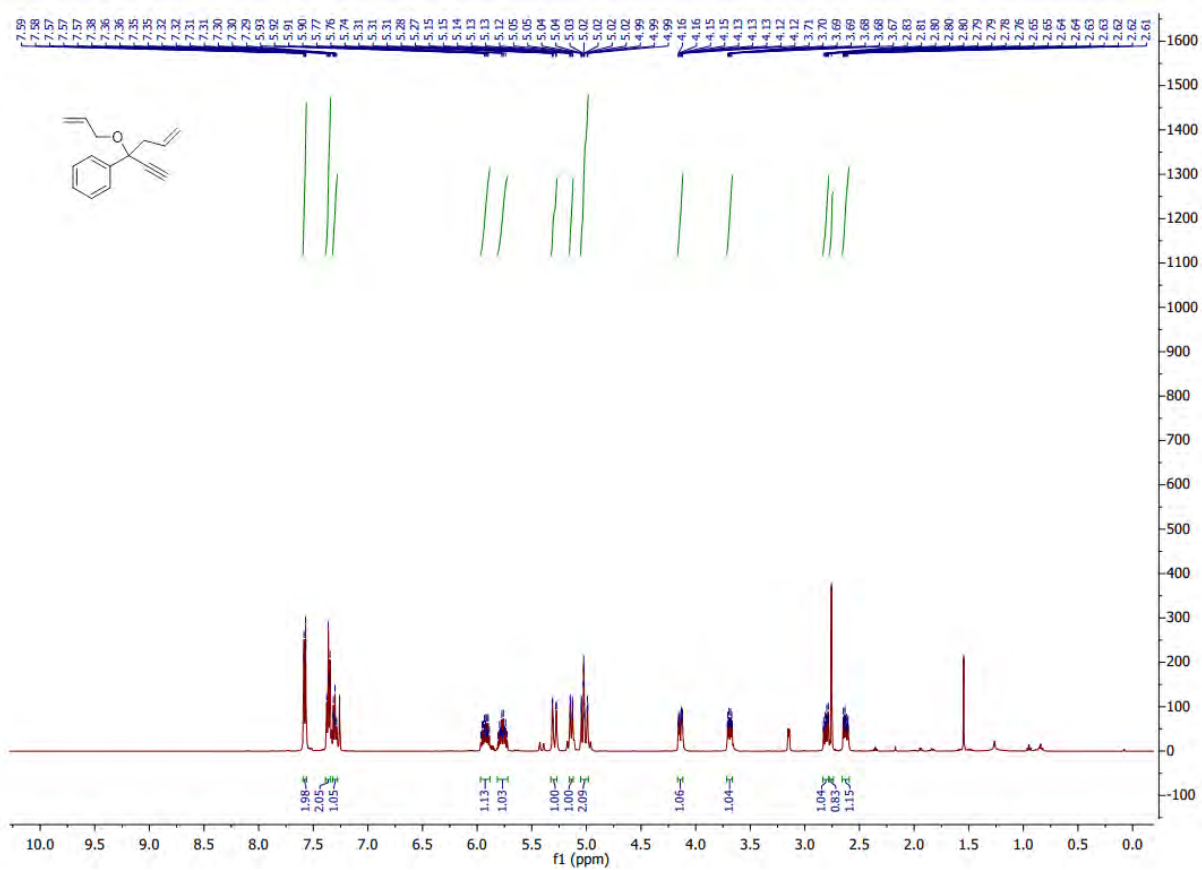
Supplementary Figure S69. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **3a**.



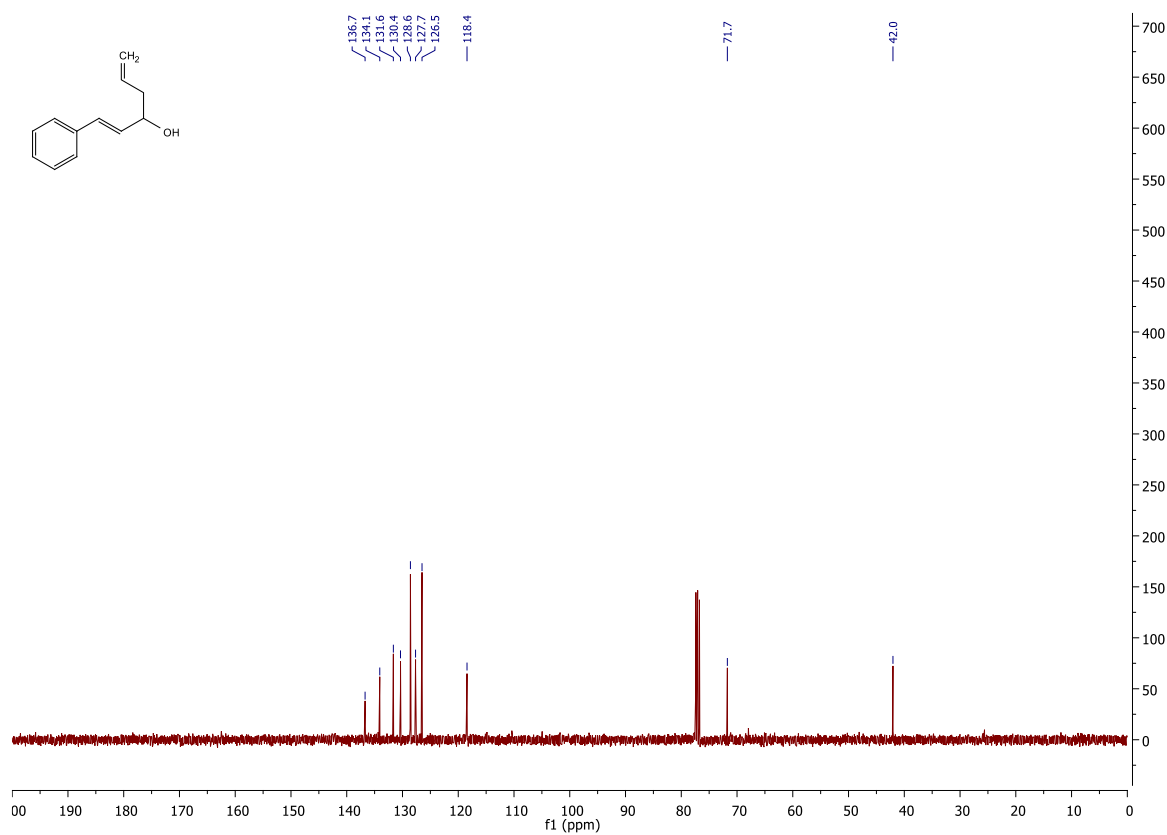
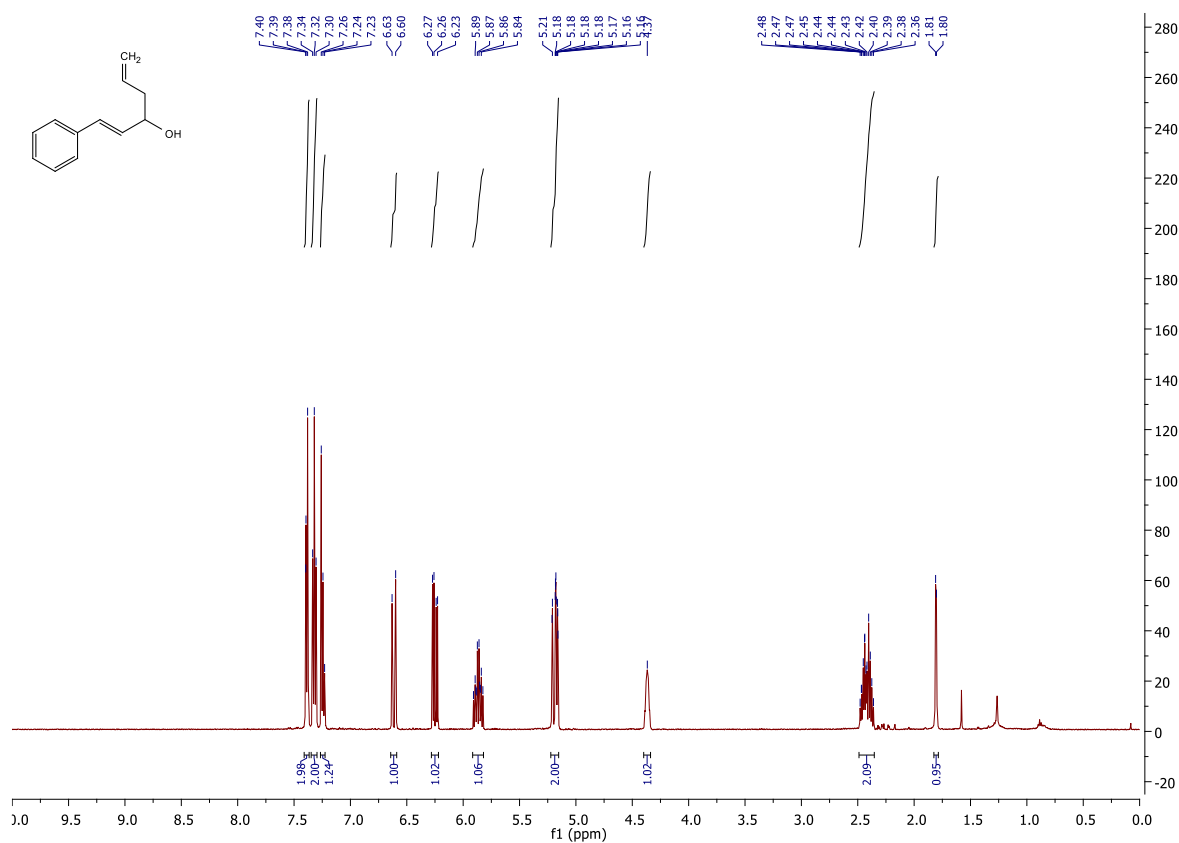
Supplementary Figure S70. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **3b**.



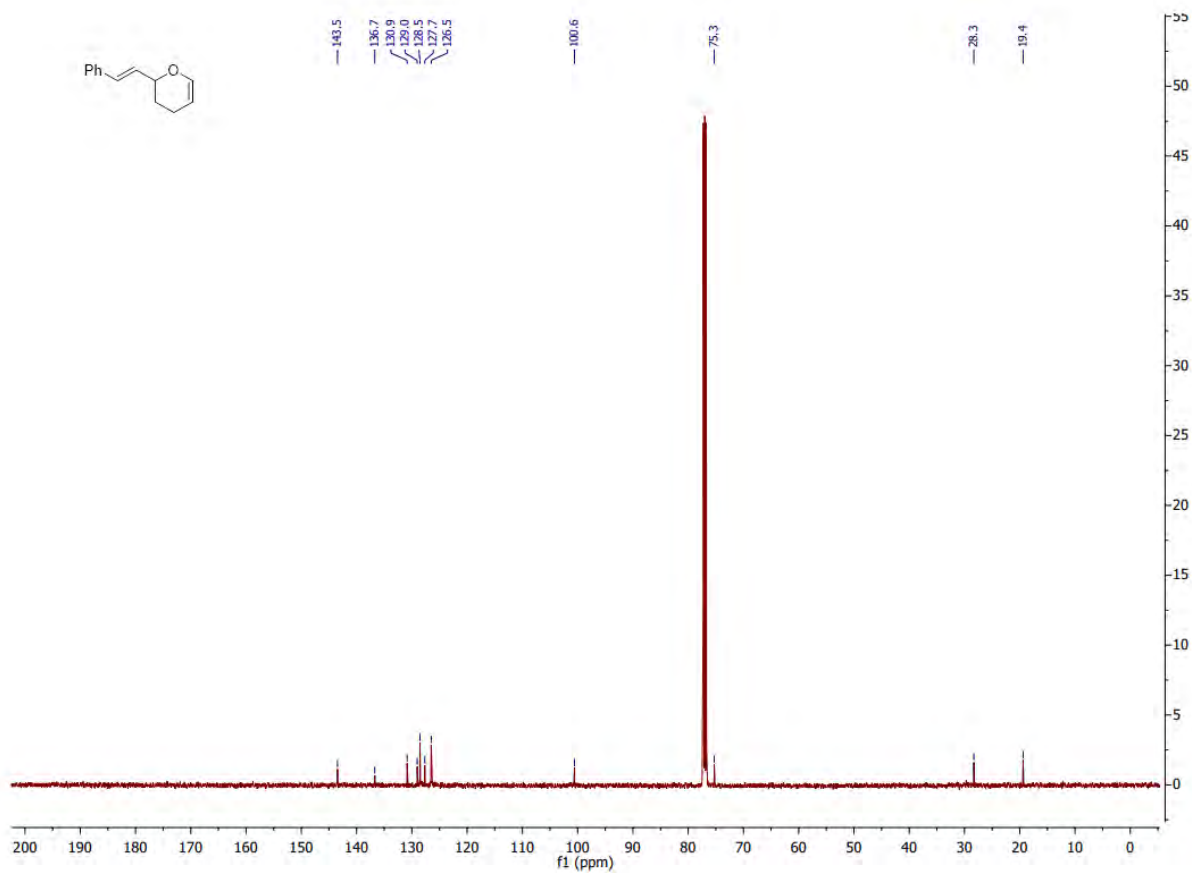
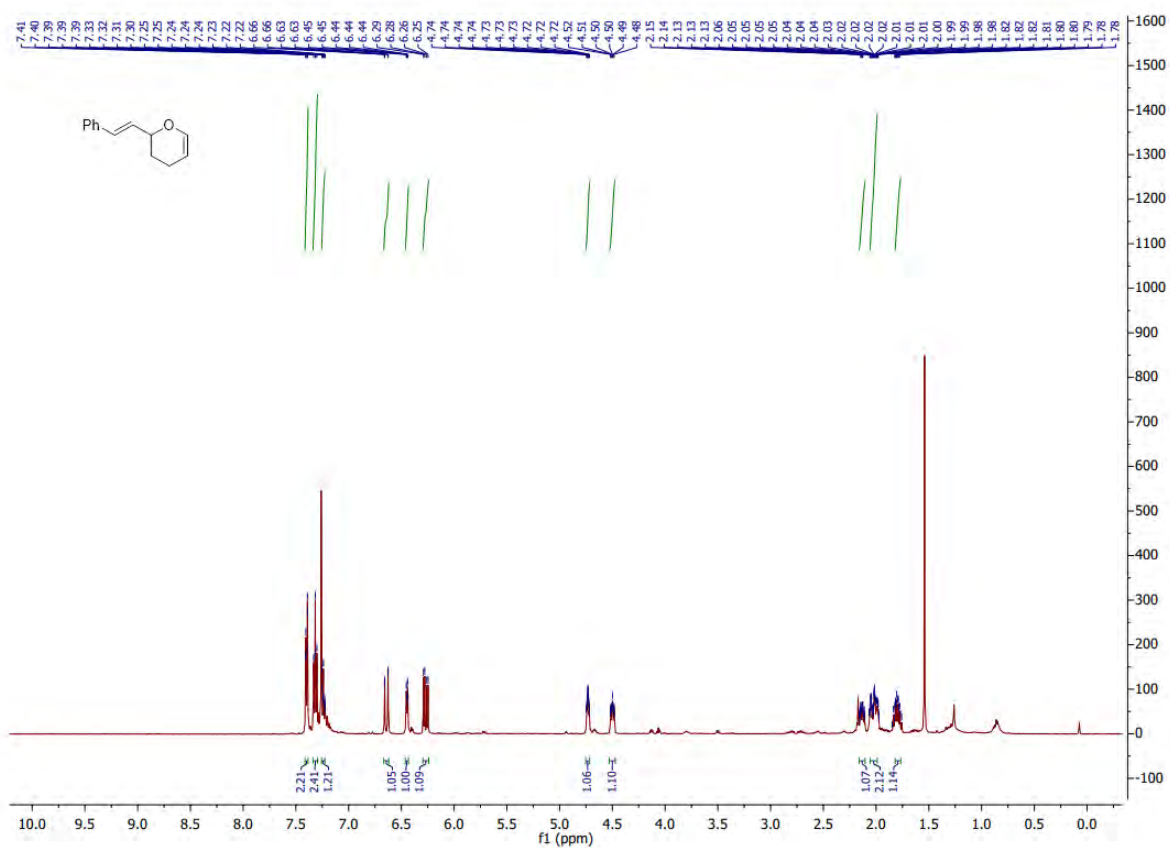
Supplementary Figure S71. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **3c**.



Supplementary Figure S72. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **3d**.

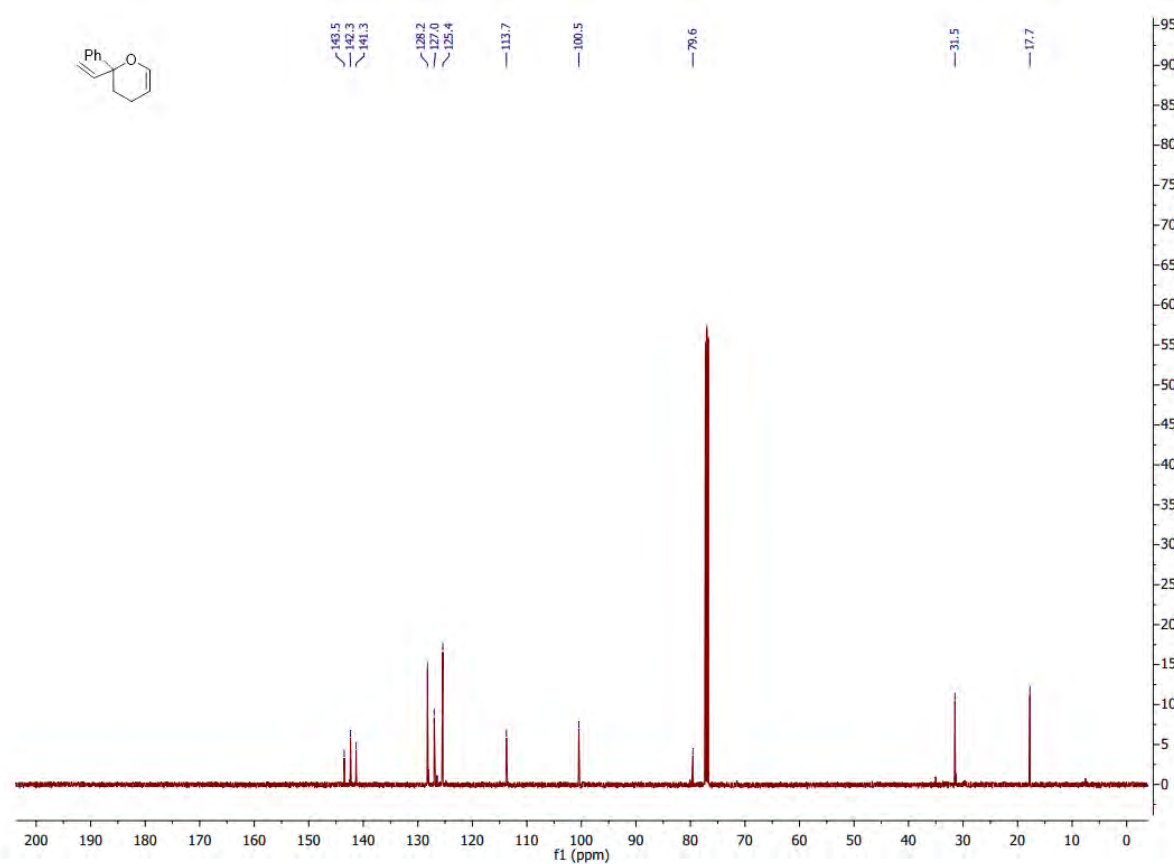
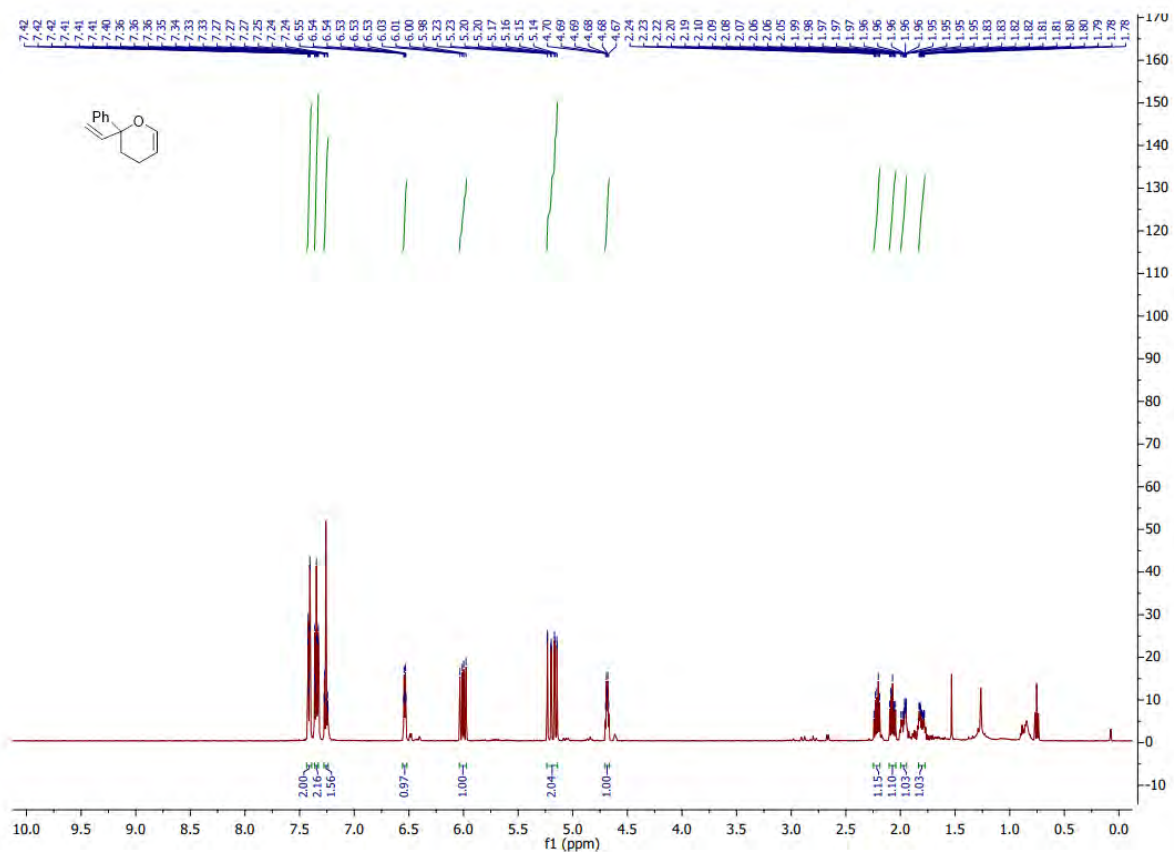


**Supplementary Figure S73.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.3.1.



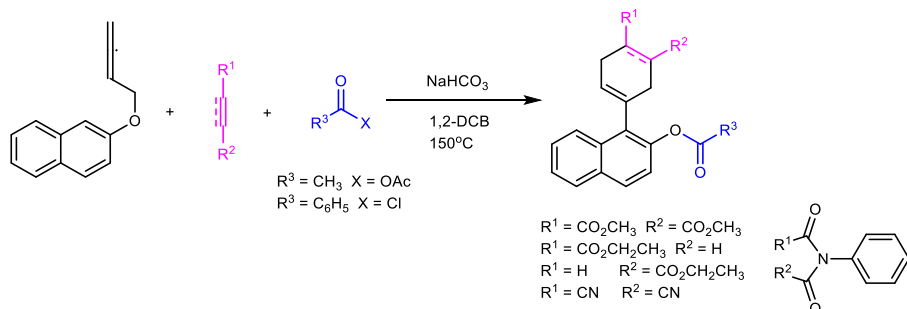
**Supplementary Figure S74.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.3.2.





**Supplementary Figure S75.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.3.3.

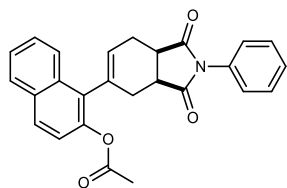
## Section S6.4. Reaction Mach4 described in main-text Figure 4b.



## Supplementary Figure S76. Reaction Mach4 described in main-text Figure 4b.

### One-pot procedure

To a Schlenk flask added 2-(buta-2,3-dien-1-yloxy)naphthalene (10 mg, 51  $\mu\text{mol}$ , 1 equiv), dienophile (102  $\mu\text{mol}$ , 2equiv), acylation/benzoylation reagent (102  $\mu\text{mol}$ , 2 equiv) and  $\text{NaHCO}_3$  (4.3 mg, 51  $\mu\text{mol}$ , 1 equiv) and then suspended in 1,2-dichlorobenzene (0.5 mL). Flask was flushed with argon gas, closed tightly and then placed in heating bath warmed to 150  $^\circ\text{C}$ . Reaction was stirred for a given time and then reaction mixture was directly transferred onto chromatography column and purified by column chromatography.



### 1-(1,3-dioxo-2-phenyl-2,3,3a,4,7,7a-hexahydro-1H-isoindol-5-yl)naphthalen-2-yl acetate **4a**

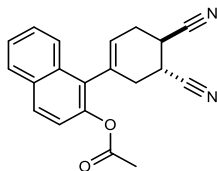
Dienophile: *N*-phenylmaleimide (17.7mg, 102  $\mu\text{mol}$ , 2 equiv); Acylation reagent: acetic anhydride (9.6 $\mu\text{L}$ , 10.4 mg, 102  $\mu\text{mol}$ , 2 equiv); Reaction time: 2.5h ; Chromatography eluent: (hexane/EtOAc 3:2); Yield: 20.2 mg, 96% a slightly yellow solid (mixture of isomers in proportion c.a. 1.7:1).

$^1\text{H NMR}$  (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.82 (*major*, m, 3H), 7.70 (*minor*, m, 3 H) 7.52 – 7.36 (*major + minor*, m, 6H), 7.25 (*major + minor*, m, 2H), 6.05 (*major + minor*, m, 1H), 3.49 – 3.31 (*major + minor*, m, 2H), 3.12 – 2.50 (*major + minor*, m, 4H), 2.31 (*major*, s, 3H) 1.79 (*minor*, s, 3H).

$^{13}\text{C NMR}$  (126 MHz,  $\text{CDCl}_3$ )  $\delta$  178.9 (*major*), 178.8 (*minor*), 178.4 (*major*), 178.3 (*minor*), 170.56 (*minor*), 169.27 (*major*), 145.2 (*major*), 144.78 (*minor*), 134.1 (*major*), 133.32 (*minor*), 132.45 (*minor*), 132.02 (*minor*), 131.97 (*major*), 131.6 (*major*), 130.47 (*major*), 130.44 (*minor*), 129.1 (*major*), 128.96 (*minor*), 128.85 (*major + minor*), 128.56 (*minor*), 128.43

(*major*), 128.33 (*major*), 128.04 (*minor*), 126.7 (*major*), 126.6 (*minor*), 126.23 (*minor*), 126.20 (*major*), 125.64 (*major*), 125.60 (*minor*), 125.35 (*minor*), 125.0 (*major*), 122.0 (*minor*), 121.3 (*major*), 39.61 (*major*), 39.47 (*minor*), 39.0 (*major*), 38.9 (*minor*), 29.44 (*minor*), 29.0 (*major*), 24.5 (*major*), 24.07 (*minor*), 21.0 (*major*), 20.75 (*minor*);

**HRMS** (EI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>26</sub>H<sub>21</sub>N<sub>1</sub>O<sub>4</sub>Na 434.1368; Found 434.1361.



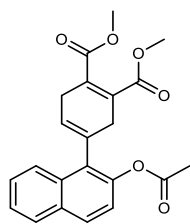
**1-(4,5-dicyanocyclohex-1-en-1-yl)naphthalen-2-yl acetate 4b.**

Dienophile: *trans*-1,2-dicyanoethylene (6.4μL, 8.0 mg, 102 μmol, 2 equiv); Acylation reagent: acetic anhydride (9.6μL, 10.4 mg, 102 μmol, 2 equiv); Reaction time: 3h; Chromatography eluent: (hexane/EtOAc 3:1); Yield: 7.6 mg, 47% a white solid.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.97 – 7.80 (m, 3H), 7.49 (m, 2H), 7.20 (m, 1H), 5.83 (m, 1H), 3.41 (m, 2H), 3.12 – 2.92 (m, 2H), 2.82 – 2.54 (m, 2H), 2.38 (m, 3H).

**<sup>13</sup>C NMR** (126 MHz, CDCl<sub>3</sub>) δ 170.2, 145.2, 132.1, 131.9, 130.1, 129.4, 128.2, 127.3, 125.9, 124.8, 124.5, 124.3, 121.5, 118.8, 118.6, 29.5, 27.1, 26.2, 25.8, 20.9;

**HRMS** (EI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>20</sub>H<sub>16</sub>N<sub>2</sub>O<sub>2</sub>Na 339.1109; Found 339.1110.



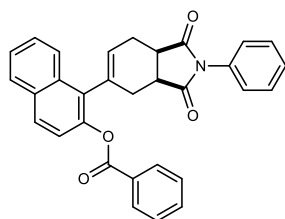
**Dimethyl 4-(2-acetoxynaphthalen-1-yl)cyclohexa-1,4-diene-1,2-dicarboxylate 4c**

Dienophile: dimethyl but-2-enedioate (12.5  $\mu$ L, 14.5 mg, 102  $\mu$ mol, 2equiv); Acylation reagent: acetic anhydride (9.6  $\mu$ L, 10.4 mg, 102  $\mu$ mol, 2 equiv); Reaction time: 2.5h ; Chromatography eluent: (hexane/EtOAc 3:1); Yield: 18.2 mg, 94%, slightly yellow liquid.

**$^1\text{H NMR}$**  (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.91-7.80 (m, 3H), 7.55-7.49 (m, 2H), 7.19 (m, 1H), 5.75 (s, 1H), 3.85 (s, 3H), 3.77 (s, 3H), 3.34-3.22 (m, 4H), 2.29 (s, 3H).

**$^{13}\text{C NMR}$**  (126 MHz,  $\text{CDCl}_3$ )  $\delta$  169.9, 168.5, 167.9, 145.2, 132.5, 132.5, 132.2, 131.9, 130.1, 129.0, 128.8, 128.3, 126.8, 125.7, 125.0, 123.7, 121.5, 52.4, 52.3, 31.5, 28.8, 20.9;

**HRMS** (EI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{22}\text{H}_{20}\text{O}_6\text{Na}$  403.1158; Found 403.1156.



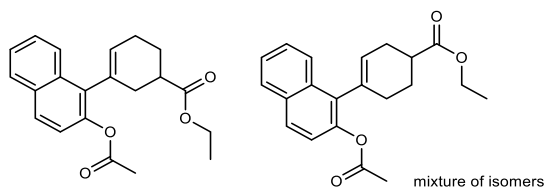
**1-(1,3-dioxo-2-phenyl-2,3,3a,4,7,7a-hexahydro-1H-isoindol-5-yl)naphthalen-2-yl benzoate 4d**

Dienophile: *N*-phenylmaleimide (17.7mg, 102  $\mu$ mol, 2 equiv); Acylation reagent: benzoyl chloride (11.8  $\mu$ L, 14.3 mg, 102  $\mu$ mol, 2 equiv); Reaction time: 3.5h; Chromatography eluent: (hexane/EtOAc 3:2); Yield: 15 mg, 62% a white solid.

**$^1\text{H NMR}$**  (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.37 (m, 1H), 7.92 (m, 2H), 7.87 – 7.83 (m, 2H), 7.69 (m, 1H), 7.55 (m, 1H), 7.50 – 7.45 (m, 2H), 7.41 (m, 4H), 7.37 – 7.34 (m, 1H), 7.20 (d,  $J = 7.5$  Hz, 2H), 5.77 (s, 1H), 5.58 (s, 1H), 4.39 (d,  $J = 8.9$  Hz, 1H), 4.32 (dd,  $J = 11.0, 8.0$  Hz, 1H), 4.18 (dd,  $J = 11.0, 7.7$  Hz, 1H), 3.95 (t,  $J = 7.9$  Hz, 1H), 3.67 (d,  $J = 8.9$  Hz, 1H).

**$^{13}\text{C NMR}$**  (126 MHz,  $\text{CDCl}_3$ )  $\delta$  177.12, 175.37, 166.09, 137.80, 133.86, 133.19, 131.89, 131.82, 130.24, 129.65, 129.10, 128.85, 128.73, 128.55, 128.36, 127.77, 127.00, 126.70, 126.47, 126.25, 125.03, 120.39, 111.99, 65.06, 44.92, 44.20, 42.96.

**HRMS** (EI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{31}\text{H}_{23}\text{N}_1\text{O}_4\text{Na}$  496.1525; Found 496.1528.



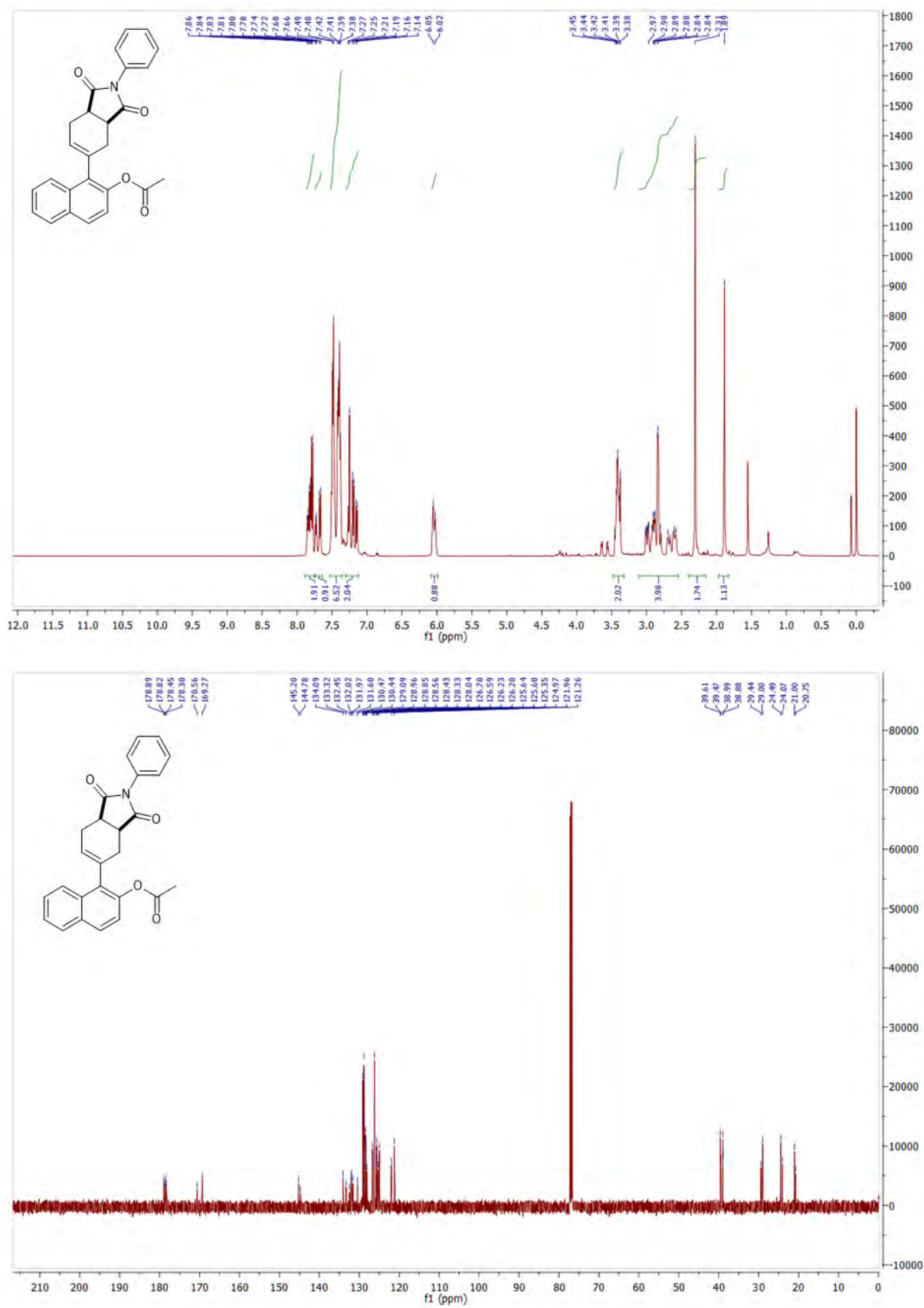
**Ethyl 3-(2-acetoxynaphthalen-1-yl)cyclohex-3-ene-1-carboxylate and ethyl 4-(2-acetoxynaphthalen-1-yl)cyclohex-3-ene-1-carboxylate (1:1 mixture) 4e**

Dienophile: ethyl acrylate (10.9 $\mu$ L, 10.2 mg, 102  $\mu$ mol, 2 equiv); Acylation reagent: acetic anhydride (9.6 $\mu$ L, 10.4 mg, 102  $\mu$ mol, 2 equiv); Reaction time: 22h; Chromatography eluent: (pentane/Et<sub>2</sub>O 9:1); Yield: 15.9 mg, 92% colorless liquid (mixture of isomers).

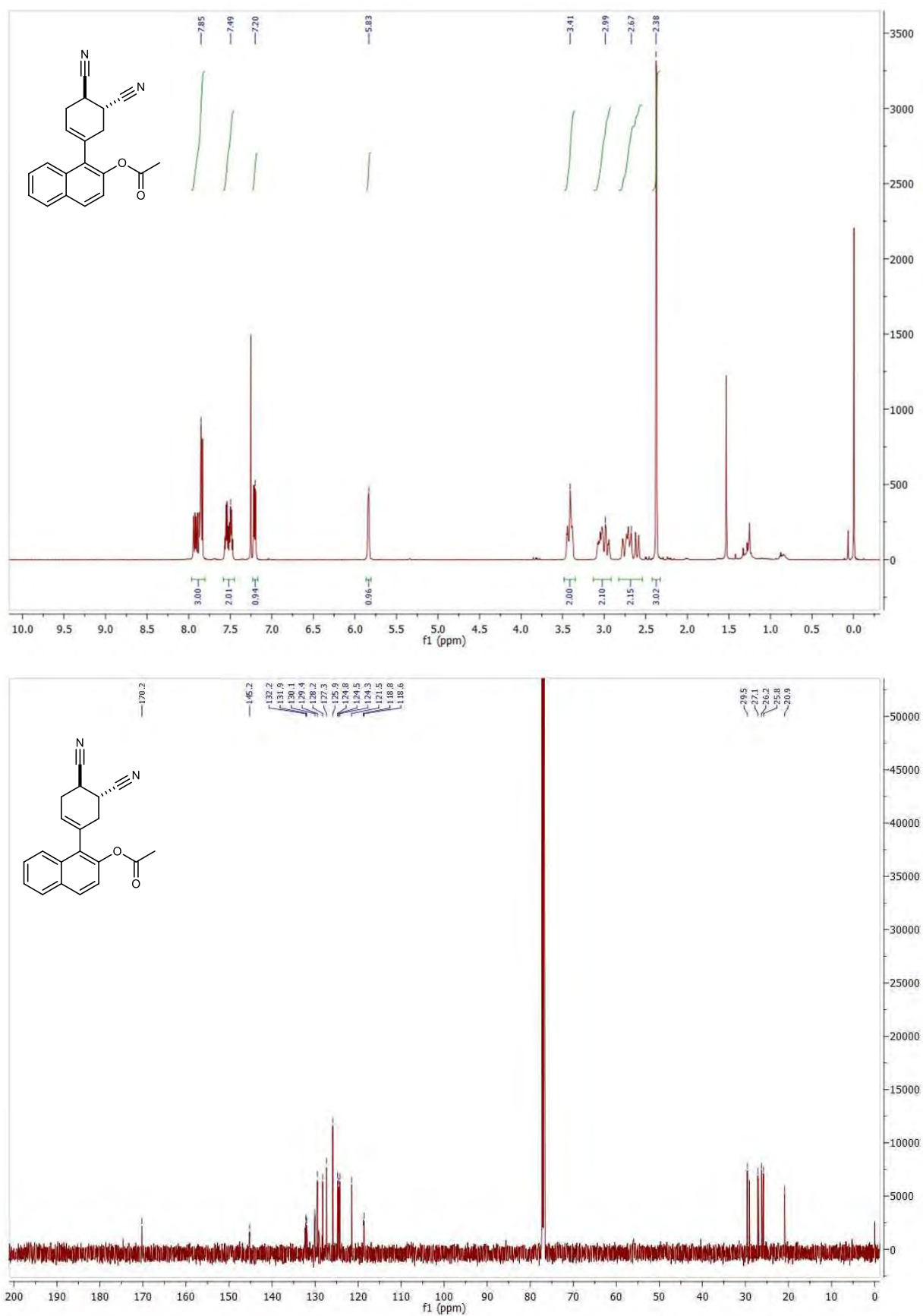
**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>)  $\delta$  8.04 – 7.72 (m, 3H), 7.52 – 7.40 (m, 2H), 7.22 – 7.15 (m, 1H), 5.71 (s, 1H), 4.28 – 4.06 (m, 2H), 2.88 – 2.65 (m, 1H), 2.62 – 2.43 (m, 2H), 2.34 (s, 2H), 2.30 (s + s, 3H), 2.17 (m, 1 H), 1.93 (m, 1H), 1.33 – 1.23 (m, 3H).

**<sup>13</sup>C NMR** (126 MHz, CDCl<sub>3</sub>)  $\delta$  175.62, 175.55, 169.95, 169.84, 144.98, 144.78, 132.54, 132.32, 132.15, 131.97, 131.88, 128.34, 128.28, 128.24, 128.13, 128.02, 126.68, 126.64, 126.44, 126.32, 125.71, 125.50, 125.47, 125.43, 125.40, 125.34, 121.56, 121.38, 60.49, 60.39, 39.13, 38.87, 31.67, 28.80, 28.63, 27.80, 25.73, 25.65, 20.86, 20.82, 14.28, 14.25.

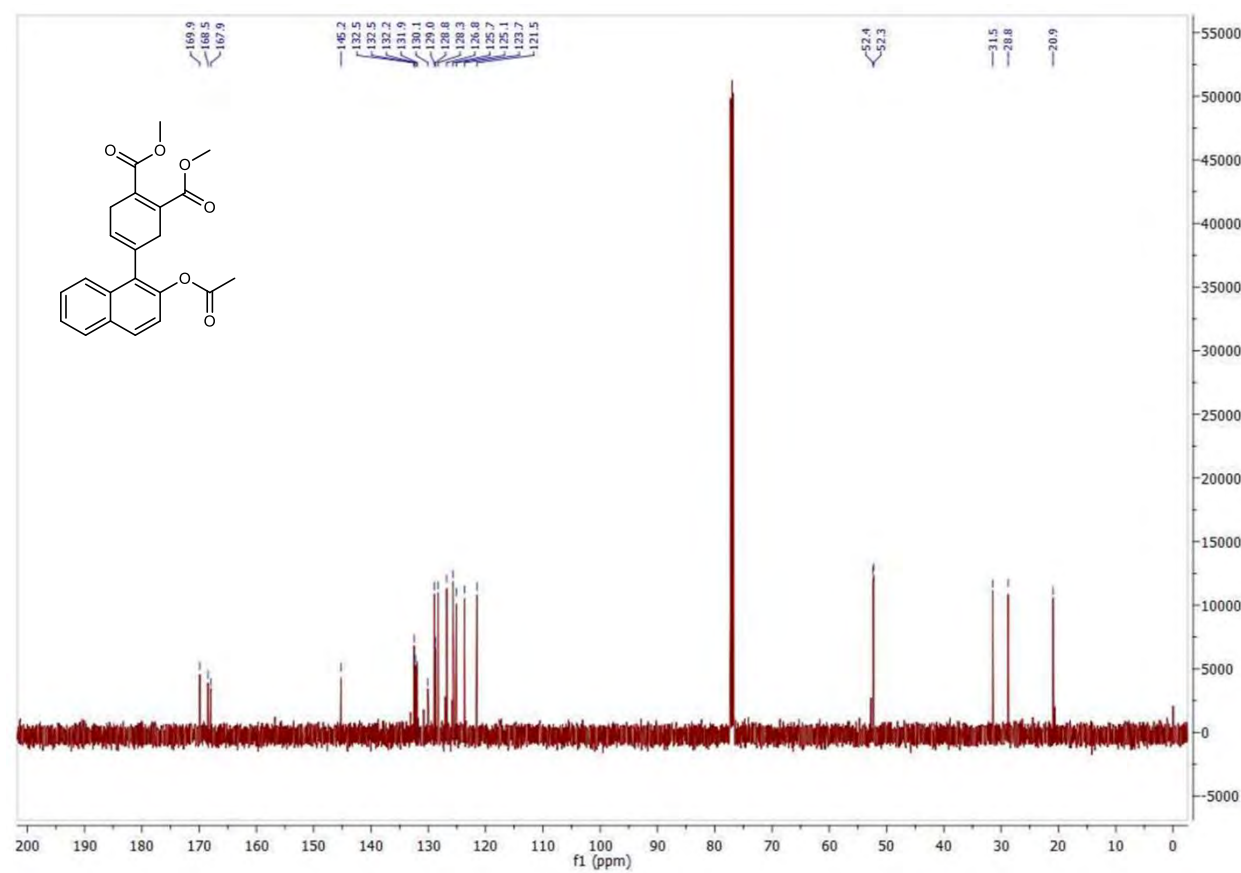
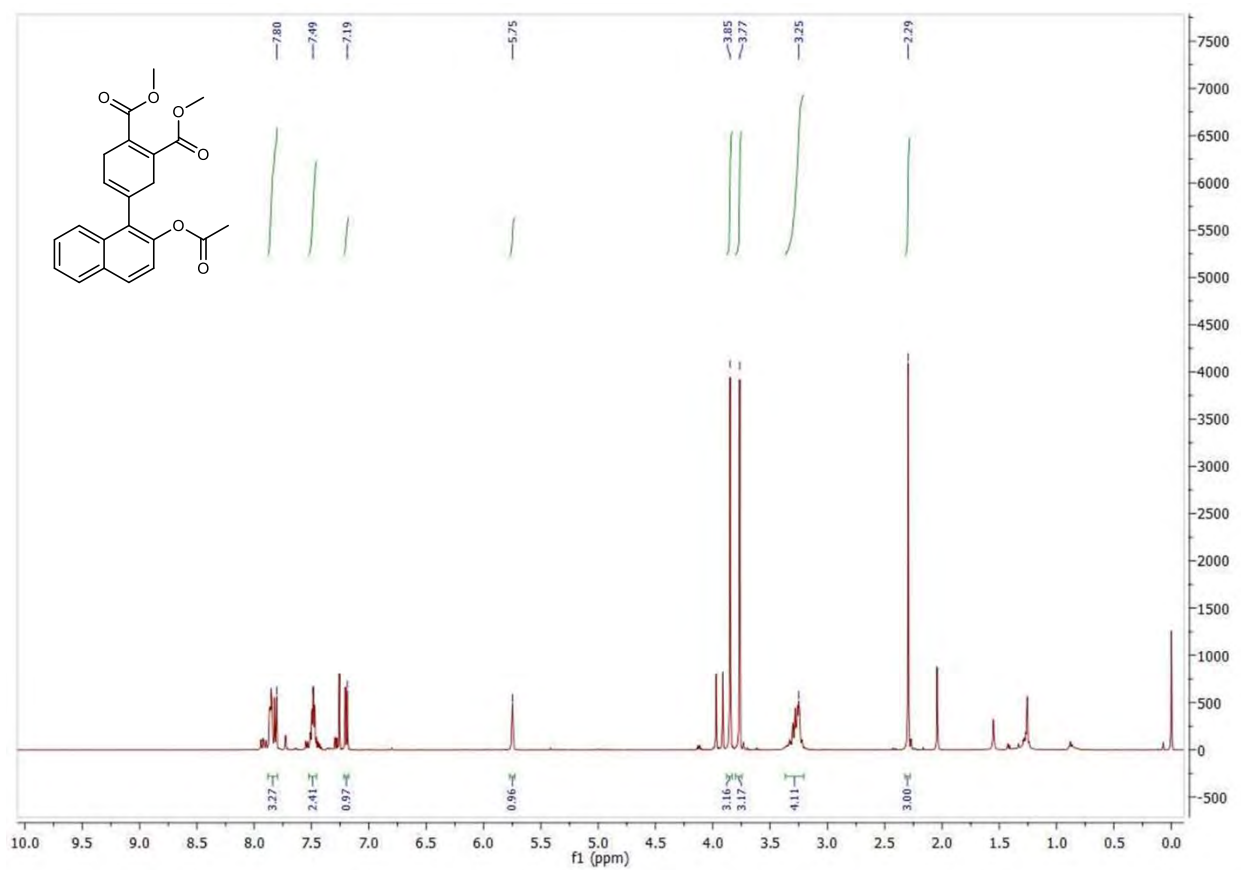
**HRMS** (EI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>21</sub>H<sub>22</sub>O<sub>4</sub>Na 361.1416; Found 361.1417.



Supplementary Figure S77.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound **4a**.

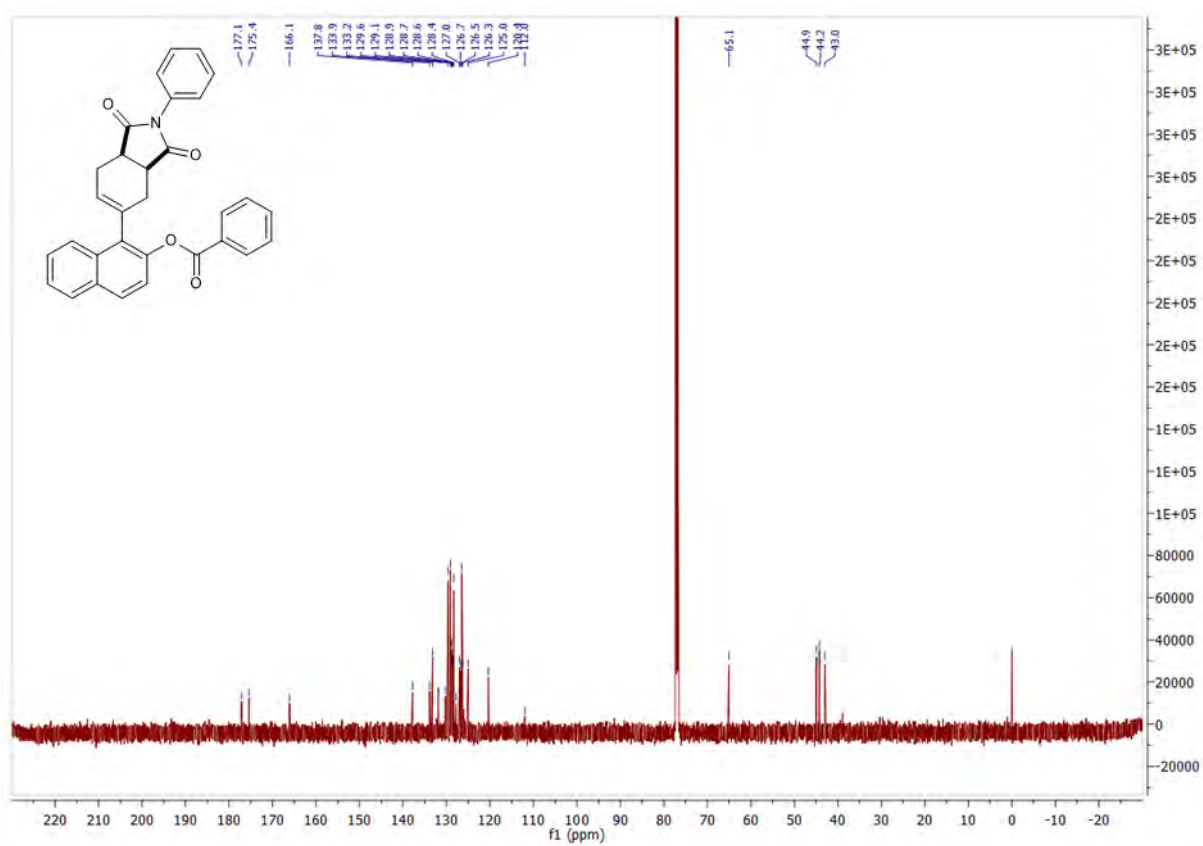
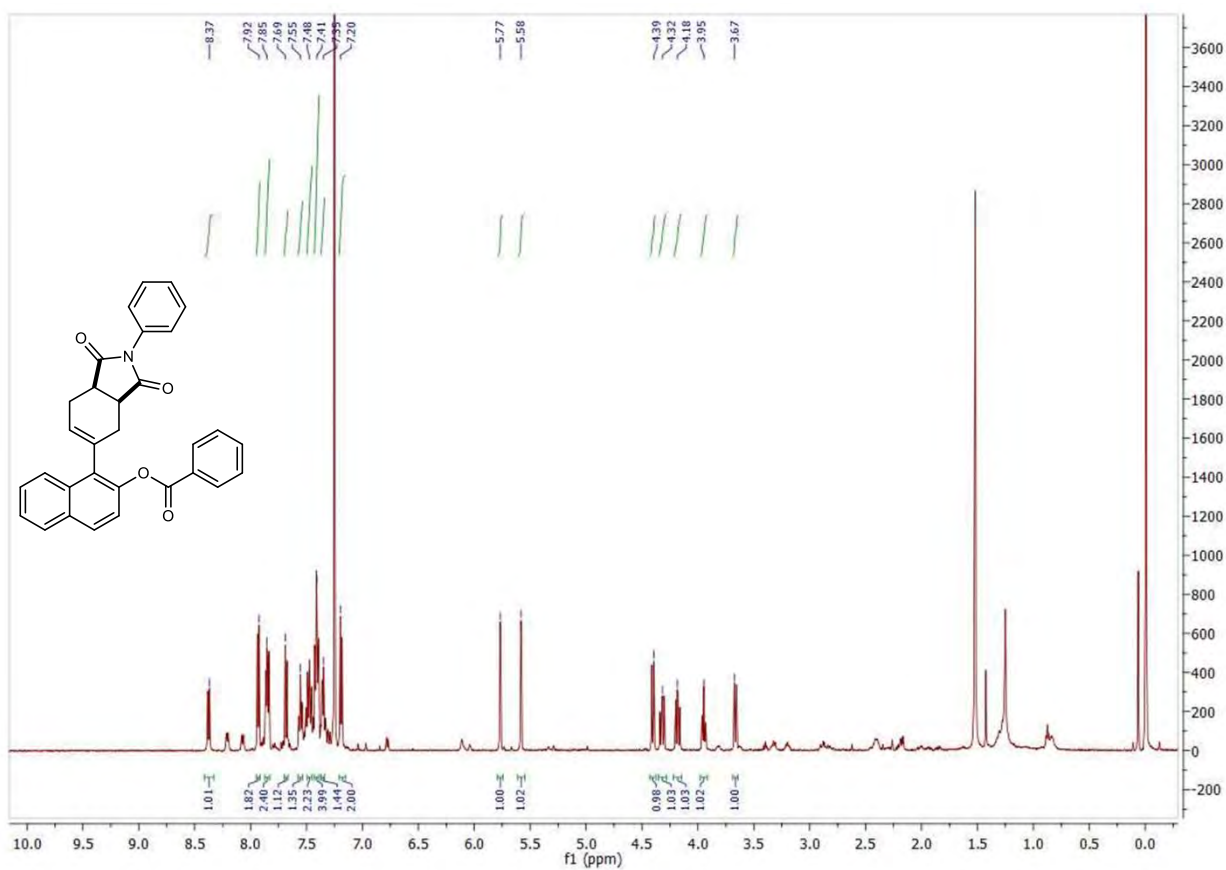


Supplementary Figure S78. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **4b**.

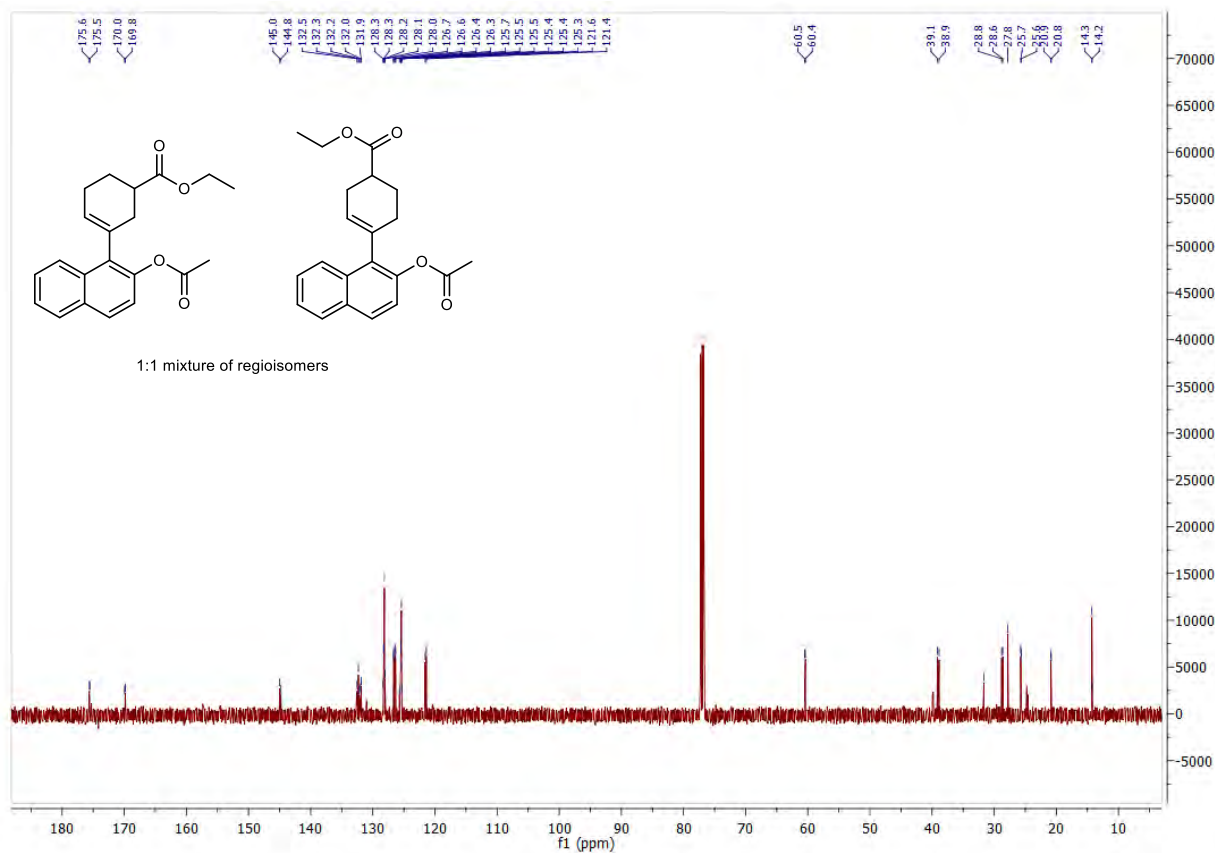
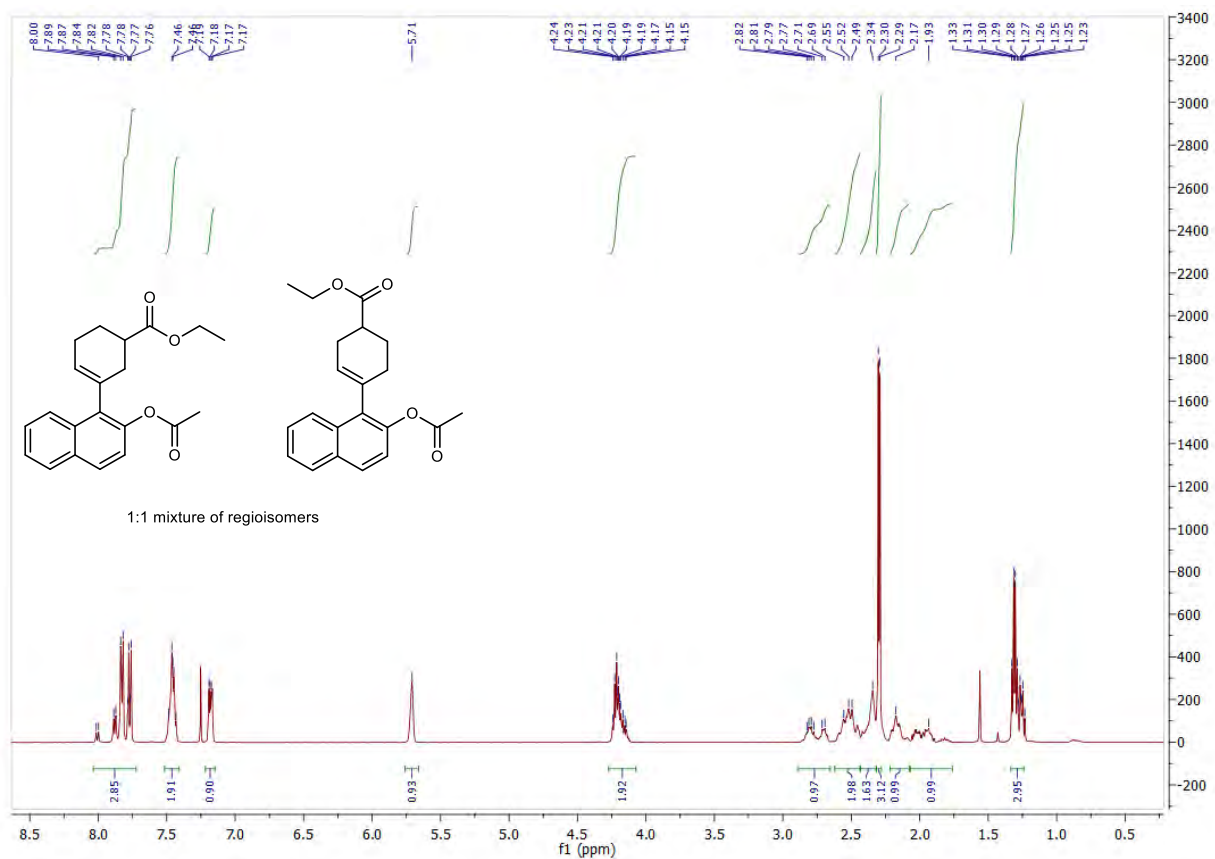


**Supplementary Figure S79.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **4c**.



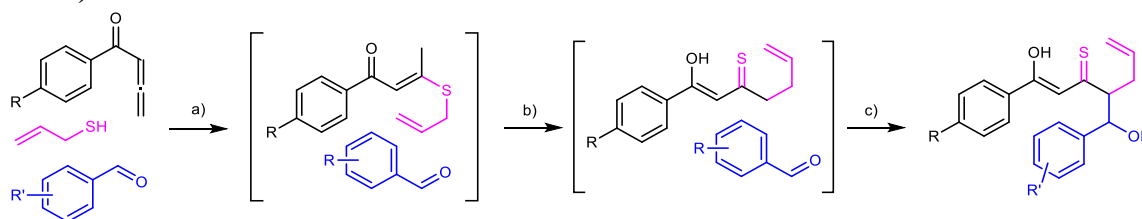


Supplementary Figure S80. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **4d**.



Supplementary Figure S81.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound **4e**.

**Section S6.5 Reaction Mach5 described in main-text Figure 4c (variant with the addition of DBU).**

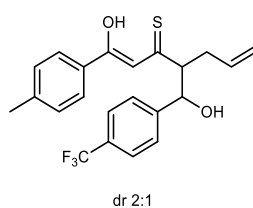


**Supplementary Figure S82. Mach5 described in main-text Figure 4c (variant with the addition of DBU).** Reagents and conditions: (a) *o*-DCB, 40 °C 1.5 h (b) 180 °C, 10 min. (c) DBU rt, 15 min to 1 h.

**One-pot procedure**

**General Procedure:**

To an argon-purged sealed flask containing anhydrous *o*-DCB (3.2 mL) was added terminal allene<sup>54</sup> (1 mmol), prop-2-ene-1-thiol (0.1 mL, 1.2 mmol, 1.2 equiv), and corresponding aldehyde (1 mmol, 1 equiv). The reaction flask was put into a pre-heated 40 °C bath and stirred for 1.5 h. After completion of the first step, monitored by TLC, the reaction mixture was put into a pre-heated 180°C oil bath, stirred for 10 min. then cooled to rt. Completion of the thio-Claisen rearrangement was monitored by TLC. Then, DBU (0.15 mL, 152.2 mg, 1 mmol, 1 equiv) was added to the mixture at once, stirred at rt until complete conversion, monitored by TLC. Upon completion, without evaporating the solvent, the crude mixture was directly loaded to the silica gel, purified by flash chromatography (Hexane/Et<sub>2</sub>O/AcOH) to yield the desired product.



**(Z)-1-hydroxy-4-(hydroxy(4-(trifluoromethyl)phenyl)methyl)-1-(*p*-tolyl)hepta-1,6-diene-3-thione **5a**.**

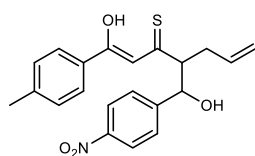
From *p*-trifluoromethylbenzaldehyde (0.14 mL, 174 mg, 1 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) **5a** was synthesized following the general procedure. After addition of DBU, the reaction was stirred for 15 min at rt and purified by column chromatography (Hexane/Et<sub>2</sub>O/AcOH 10:1:0.025) to give **5a** as two diastereomers (dr 1:2) (230 mg, 57%) as a thick red oil.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 16.01 (s, 1H Minor), 15.76 (d, *J* = 1.2 Hz, 1H Major), 7.74 – 7.69 (m, 2H Major + 2H Minor), 7.63 – 7.56 (m, 2H Major + 2H Minor), 7.54 – 7.49 (m, 2H Major + 2H Minor), 7.27 – 7.26 (m, 2H Minor), 7.25 – 7.24 (m, 2H Major), 6.83 (s, 1H Minor), 6.79 (s, 1H Major), 5.74 – 5.57 (m, 1H Major + 1H Minor), 5.10 (d, *J* = 7.3 Hz, 1H Minor), 5.08 (d, *J* = 5.2 Hz, 1H Major), 5.06 – 5.00 (m, 1H Major + 1H Minor), 4.98 – 4.93 (m, 1H Major + 1H Minor), 3.27 – 3.21 (m, 1H Minor), 3.23 – 3.17 (m, 1H Major), 2.83 – 2.74 (m, 1H Major), 2.58 – 2.50 (m, 1H Major + 1H Minor), 2.42 (s, 3H Major + 3H Minor), 2.26 – 2.19 (m, 1H Minor);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 216.0 (Major), 215.9 (Minor) 178.8 (Major), 178.5 (Minor), 146.1 (d, *J* = 1.3 Hz Minor), 146.0 (d, *J* = 1.4 Hz Major), 144.2 (Minor), 144.1 (Major), 135.5 (Major), 134.6 (Minor), 131.7 (Major), 131.6 (Minor), 130.1 (d, *J* = 29.3 Hz Minor), 130.08 (d, *J* = 29.2 Hz Major), 129.6 (Major + Minor), 129.2 (Major), 129.1 (Minor), 127.4 (Minor) 127.3 (Major), 127.1 (Minor), 126.8 (Major), 125.3 (q, *J* = 3.8 Hz Minor), 125.2 (q, *J* = 3.8 Hz Major), 121.3 (q, *J* = 279.0 Hz CF<sub>3</sub> Major), 121.3 (q, *J* = 278.4 Hz CF<sub>3</sub> Minor), 117.4 (Minor), 117.0 (Major), 114.1 (Minor), 112.9 (Major), 76.4 (Minor), 76.1 (Major), 65.1 (Major), 64.8 (Minor), 37.0 (Minor), 33.9 (Major), 21.73 (Minor), 21.72 (Major);

**IR** (film, DCM) 3424 (OH), 3076, 2979, 2919 (C-S), 1607, 1579, 1550, 1504, 1451, 1417, 1326 (CF<sub>3</sub>), 1254 (C=S), 1187, 1165, 1125, 1067, 1016, 918, 842, 809, 726, cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M-H]<sup>-</sup> Calcd for C<sub>22</sub>H<sub>20</sub>F<sub>3</sub>O<sub>2</sub>S<sup>-</sup> 405.1136; Found 405.1134.



dr 1.2:1

**(Z)-1-hydroxy-4-(hydroxy(4-nitrophenyl)methyl)-1-(p-tolyl)hepta-1,6-diene-3-thione 5c.**

From *p*-nitrobenzaldehyde (151.1 mg, 1 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) 5c was synthesized following the general procedure. After addition of DBU, the reaction was stirred for 15 min at rt and purified by column chromatography (Hexane/Et<sub>2</sub>O/AcOH 4:1:0.025) to give 5c as two diastereomers (1:1.2) (135 mg, 35%) as a red-brown thick oil.

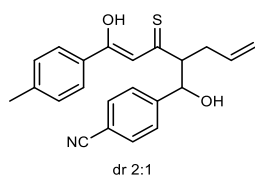
**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 16.08 – 16.04 (m, 1H Minor), 15.95 – 15.88 (m, 1H Major), 8.19 (d, *J* = 8.9 Hz, 2H Minor), 8.17 (d, *J* = 8.7 Hz, 2H Major), 7.75 (d, *J* = 8.3 Hz, 2H Minor), 7.74 (d, *J* = 8.3 Hz, 2H Major), 7.58 (d, *J* = 8.7 Hz, 2H Major), 7.55 (d, *J* = 8.6 Hz, 2H Minor),

7.27 – 7.24 (m, 2H Major + 2H Minor), 6.86 (s, 1H Minor), 6.84 (s, 1H Major), 5.70 – 5.60 (m, 1H Major + 1H Minor), 5.17 – 5.11 (m, 1H Major + 1H Minor), 5.05 – 5.00 (m, 1H Major + 1H Minor), 5.00 – 4.94 (m, 1H Major + 1H Minor), 3.27 – 3.22 (m, 1H Minor), 3.22 – 3.18 (m, 1H Major), 3.15 (br(s), OH Major + OH Minor), 2.82 – 2.74 (m, 1H Major), 2.63 – 2.56 (m, 1H Minor), 2.53 – 2.47 (m, 1H Major), 2.41 (s, 3H Major + 3H Minor), 2.30 – 2.23 (m, 1H Minor);

<sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 216.6 (Major), 216.1 (Minor), 178.7 (Major), 178.4 (Minor), 149.6 (Major), 149.4 (Minor), 147.4 (Minor), 147.3 (Major), 144.5 (Minor), 144.4 (Major), 135.2 (Major), 135.4 (Minor), 131.4 (Major), 131.2 (Minor), 129.7 (Major + Minor), 127.5 (Minor), 127.4 (Major), 127.35 (Major), 127.3 (Minor), 123.5 (Minor), 123.4 (Major), 117.7 (Minor), 117.2 (Major), 113.8 (Minor), 112.7 (Major), 76.0 (Minor), 75.8 (Major), 64.9 (Major), 64.5 (Minor), 37.1 (Minor), 33.9 (Major), 21.8 (Major + Minor);

IR (film, DCM) 3437 (OH), 3076, 2987, 2918, 2857 (C-S), 1606, 1578, 1550, 1520 (NO<sub>2</sub>), 1505, 1455, 1346 (NO<sub>2</sub>), 1257 (C=S), 1189, 1108, 1046, 1014, 918, 855, 807, 757, 725 cm<sup>-1</sup>;

HRMS (ESI) *m/z*: [M-H]<sup>-</sup> Calcd. for C<sub>21</sub>H<sub>20</sub>NO<sub>4</sub>S<sup>-</sup> 382.1123; Found 382.1111.



#### (Z)-4-(2-allyl-1,5-dihydroxy-3-thioxo-5-(*p*-tolyl)pent-4-en-1-yl)benzonitrile **5d**

From *p*-cyanobenzaldehyde (131.1 mg, 1 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) **5d** was synthesized following the general procedure. After addition of DBU the reaction was stirred for 1 h at rt and purified by column chromatography (Hexane/Et<sub>2</sub>O/AcOH 4:1:0.025) to give **5d** as two diastereomers (1:2) (89 mg, 24%) as a thick red oil.

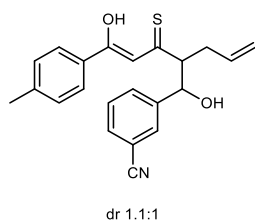
<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 16.06 (d, *J* = 0.6 Hz, 1H Minor), 15.89 (dd, *J* = 1.9, 1.0 Hz, 1H Major), 7.74 (d, *J* = 8.3 Hz, 2H Major + 2H Minor), 7.63 (d, *J* = 8.3 Hz, 2H Minor), 7.61 (d, *J* = 8.3 Hz, 2H Major), 7.52 (d, *J* = 8.1 Hz, 2H Major), 7.49 (d, *J* = 8.1 Hz, 2H Minor), 7.27 (d, *J* = 8.1 Hz, 2H Major + 2H Minor), 6.83 (s, 1H Minor), 6.81 (s, 1H Major), 5.68 – 5.60 (m, 1H Major + 1H Minor), 5.09 (d, *J* = 7.0 Hz, 1H Minor), 5.07 (d, *J* = 5.1 Hz, 1H Major), 5.04 – 5.00 (m, 1H Major + 1H Minor), 4.99 – 4.96 (m, 1H Minor), 4.96 – 4.93 (m, 1H Major), 3.21 (ddd, *J* = 9.5, 6.9, 5.2 Hz, 1H Minor), 3.17 (ddd, *J* = 10.4, 5.1, 4.0 Hz, 1H Major), 3.08 (br(s), OH

Major), 3.04 (br(s), OH Minor), 2.80 – 2.73 (m, 1H Major), 2.61 – 2.55 (m, 1H Minor), 2.52 – 2.47 (m, 1H Major), 2.42 (s, 3H Major + 3H Minor), 2.28 – 2.22 (m, 1H Minor);

<sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ 216.5 (Major), 216.1 (Minor), 178.7 (Major), 178.4 (Minor), 147.5 (Minor), 147.3 (Major), 144.35 (Minor), 144.4 (Major), 135.3 (Major), 134.4 (Minor), 132.1 (Minor), 132.0 (Major), 131.4 (Major), 131.3 (Minor), 129.7 (Major + Minor), 127.37 (Minor), 127.36 (Minor), 127.3 (Major), 127.2 (Major), 118.7 (Major), 118.6 (Minor), 117.6 (Minor), 117.2 (Major), 113.8 (Minor), 112.7 (Major), 111.6 (Minor), 111.3 (Major), 76.2 (Minor), 75.9 (Major), 64.9 (Major), 64.5 (Minor), 37.1 (Minor), 33.9 (Major), 21.7 (Major + Minor);

IR (film, DCM) 3459 (OH), 3073, 3033, 2979, 2920, 2859 (C-S), 2228 (CN), 1606, 1578, 1549, 1503, 1453, 1254 (C=S), 1189, 1120, 1017, 917, 835, 810, 733 cm<sup>-1</sup>.

HRMS (ESI) *m/z*: [M-H]<sup>-</sup> Calcd for C<sub>22</sub>H<sub>20</sub>NO<sub>2</sub>S<sup>-</sup> 322.1215; Found 322.1212.



### (Z)-3-(2-allyl-1,5-dihydroxy-3-thioxo-5-(p-tolyl)pent-4-en-1-yl) **5e**

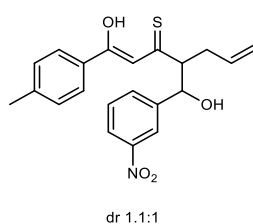
From *m*-cyanobenzaldehyde (131.1 mg, 1 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) **5e** was synthesized following the general procedure. After addition of DBU the reaction was stirred for 1 h at rt and purified by column chromatography (Hexane/Et<sub>2</sub>O/AcOH 4:1:0.025) to give **5e** as two diastereomers (1:1.1) (82 mg, 23%) as a thick red oil.

<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 16.08 (s, 1H Minor), 15.91 (s, 1H Major), 7.78 (d, *J* = 8.0 Hz, 2H Minor), 7.76 (d, *J* = 7.8 Hz, 2H Major), 7.72 (d, *J* = 15.7 Hz, 2H Minor), 7.63 (d, *J* = 7.8 Hz, 2H Major), 7.56 (d, *J* = 7.7 Hz, 1H Minor), 7.52 (d, *J* = 7.7 Hz, 1H Major), 7.45 (t, *J* = 7.8 Hz, 1H Minor), 7.41 (t, *J* = 7.7 Hz, 1H Major), 7.27 (d, *J* = 8.0 Hz, 2H Major + 2H Minor), 6.89 (s, 1H Minor), 6.84 (s, 1H Major), 5.69 – 5.59 (m, 1H Major + 1H Minor), 5.07 (d, *J* = 7.4 Hz, 1H Minor), 5.05 (d, *J* = 5.1 Hz, 1H Major), 5.04 – 4.97 (m, 1H Major + 1H Minor), 4.99 – 4.94 (m, 1H Major + 1H Minor), 3.24 – 3.20 (m, 1H Minor), 3.19 – 3.15 (m, 1H Major), 3.09 (br(s), OH Major), 2.98 (br(s), OH Minor), 2.80 – 2.74 (m, 1H Major), 2.60 – 2.52 (m, 1H Minor), 2.52 – 2.47 (m, 1H Major), 2.42 (s, 3H Major + 3H Minor), 2.60 – 2.52 (m, 1H Minor);

$^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  216.8 (Minor), 216.3 (Major), 178.6 (Minor), 178.4 (Major), 144.4 (Minor), 144.3 (Major), 143.7 (Minor), 143.6 (Major), 135.3 (Major), 134.4 (Minor), 131.5 (Minor), 131.2 (Minor), 131.19 (Major), 131.0 (Major), 130.3 (Minor), 130.1 (Major), 129.7 (Major + Minor), 129.1 (Minor), 128.9 (Major), 127.4 (Minor), 127.3 (Major), 118.8 (Major), 118.7 (Minor), 117.6 (Minor), 117.2 (Major), 113.9 (Major + Minor), 112.73 (Minor) 112.72 (Major), 112.5 (Minor), 112.3 (Major), 76.0 (Minor), 75.7 (Major), 64.9 (Major), 64.7 (Minor), 37.1 (Minor), 33.9 (Major), 21.7 (Major + Minor);

IR (film, DCM) 3460 (OH), 3075, 2923, 2856 (C-S), 2230 (CN), 1606, 1578, 1550, 1505, 1437, 1253 (C=S), 1187, 1121, 1100, 1050, 1018, 910, 810, 733  $\text{cm}^{-1}$ ;

HRMS (ESI)  $m/z$ :  $[\text{M}-\text{Na}]^+$  Calcd for  $\text{C}_{22}\text{H}_{21}\text{NNaO}_2\text{S}^+$  386.1191; Found 386.1190.



**(Z)-1-hydroxy-4-(hydroxy(3-nitrophenyl)methyl)-1-(p-tolyl)hepta-1,6-diene-3-thione 5f**

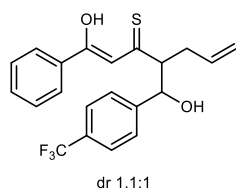
From *m*-nitrobenzaldehyde (151.1 mg, 1 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) 5f was synthesized following the general procedure. After addition of DBU the reaction was stirred for 15 min at rt and purified by column chromatography (Hexane/ $\text{Et}_2\text{O}$ / $\text{AcOH}$  4:1:0.025) to give 5f as two diastereomers (1:1.1) (80 mg, 21%) as a thick red oil.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  16.05 (s, 1H Minor), 15.89 (s, 1H Major), 8.31 – 8.29 (m, 1H Major), 8.29 – 8.28 (m, 1H Minor), 8.13 (ddd,  $J = 8.2, 2.3, 1.0$  Hz, 1H Minor), 8.09 (ddd,  $J = 8.3, 2.1, 1.0$  Hz, 1H Major), 7.77 (m, 3H Major), 7.76 – 7.71 (m, 3H Minor), 7.52 (t,  $J = 7.0$  Hz, 1H Major), 7.48 (t,  $J = 7.0$  Hz, 1H Minor), 7.27 (d,  $J = 7.6$  Hz, 2H Major + 2H Minor), 6.92 (s, 1H Minor), 6.87 (s, 1H Major), 5.72 – 5.59 (m, 1H Major + 1H Minor), 5.17 – 5.12 (m, 1H Major, 1H Minor), 5.06 – 4.99 (m, 1H Major + 1H Minor), 4.99 – 4.93 (m, 1H Major + 1H Minor), 3.31 – 3.26 (m, 1H Minor), 3.26 – 3.20 (m, 1H Major), 2.84 – 2.74 (m, 1H Major), 2.64 – 2.56 (m, 1H Minor), 2.54 – 2.47 (m, 1H Major), 2.42 (s, 3H Major + 3H Minor), 2.29 – 2.20 (m, 1H Minor);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 216.6 (Major), 216.1 (Minor), 178.7 (Major), 178.5 (Minor), 148.33 (Major), 148.27 (Minor), 144.40 (Minor), 144.37 (Major + Minor), 144.2 (Major), 135.2 (Major), 134.4 (Minor), 132.9 (Minor), 132.8 (Major), 131.4 (Major), 131.3 (Minor), 129.7 (Major + Minor), 129.3 (Minor), 129.1 (Major), 127.40 (Minor), 127.35 (Major), 122.8 (Minor), 122.6 (Major), 121.7 (Minor), 121.4 (Major), 117.7 (Minor), 117.2 (Major), 113.9 (Minor), 112.8 (Major), 76.0 (Minor), 75.7 (Major), 64.8 (Major), 64.6 (Minor), 37.1 (Minor), 34.0 (Major), 21.76 (Minor), 21.75 (Major);

IR (film, DCM) 3424 (OH), 3076, 2922, 285 (C-S), 2230 (CN), **1606, 1578, 1549**, 1530 (NO<sub>2</sub>), **1504, 1441**, 1349 (NO<sub>2</sub>), 1254 (C=S), 1188, **1197, 1063**, 916, 810, 738 cm<sup>-1</sup>;

HRMS (ESI) *m/z*: [M-Na]<sup>+</sup> Calcd for C<sub>21</sub>H<sub>21</sub>NNaO<sub>4</sub>S<sup>+</sup> 406.1089; Found 406.1085.



**(Z)-1-hydroxy-4-(hydroxy(4-(trifluoromethyl)phenyl)methyl)-1-phenylhepta-1,6-diene-3-thione 5g**

From *p*-trifluoromethylbenzaldehyde (0.1 mL, 122 mg, 0.7 mmol, 1 equiv) and 1-phenylbuta-2,3-dien-1-one (101 mg, 0.7 mmol, 1 equiv) **5g** was synthesized following the general procedure. After addition of DBU (0.1 mL, 0.7 mmol, 1 equiv.) the reaction was stirred for 15 min at rt and purified by column chromatography (Hexane/Et<sub>2</sub>O/AcOH 10:1:0.025) to give **5g** as two diastereomers (1:1.1) (97 mg, 35%) as a thick red oil.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 15.86 (s, 1H Minor), 15.61 (s, 1H Major), 7.84 – 7.77 (m, 2H Major + 2H Minor), 7.64 – 7.50 (m, 5H Major + 5H Minor), 7.48 – 7.43 (m, 2H Major + 2H Minor), 6.85 (s, 1H Minor), 6.78 (s, 1H Major), 5.75 – 5.57 (m, 1H Major + 1H Minor), 5.10 (d, *J* = 6.8 Hz, 1H Minor), 5.08 (d, *J* = 5.4 Hz, 1H Major), 5.06 – 4.95 (m, 2H Major + 2H Minor), 3.30 – 3.18 (m, 1H Major + 1H Minor), 2.86 (br(s), OH Minor), 2.81 – 2.75 (m, 1H Major), 2.78 (br(s), OH Major), 2.61 – 2.53 (m, 1H Major + 1H Minor), 2.28 – 2.17 (m, 1H Minor);

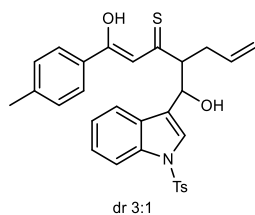
<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 216.3 (Major), 216.1 (Minor), 178.7 (Major), 178.4 (Minor), 146.0 (d, *J*<sub>4</sub> = 1.2 Hz Major), 146.08 (d, *J*<sub>4</sub> = 0.8 Hz Minor), 135.4 (Major), 134.7 (Major), 134.54 (Minor), 134.52 (Minor), 133.0 (Major), 132.9 (Minor), 129.9 (d, *J*<sub>2</sub> = 32.3 Hz Major), 129.8 (d, *J*<sub>2</sub> = 28.3 Hz Minor), 128.9 (Major + Minor), 127.3 (Minor), 127.2 (Major), 127.1



(Minor), 126.8 (Major), 125.4 (q,  $J_3 = 3.6$  Hz Minor), 125.3 (q,  $J_3 = 3.7$  Hz Major), 124.1 (d,  $J_1 = 271.9$  Hz Major + Minor), 117.5 (Major), 117.1 (Minor), 114.3 (Minor), 113.2 (Major), 76.5 (Major), 76.2 (Minor), 65.2 (Major), 64.9 (Minor), 37.0 (Minor), 34.0 (Major);

**IR** (film, DCM) 3431 (OH), 3074, 2979, 2915 (C-S), 1619, 1586, 1551, 1491, 1463, 1417, 1326 (CF<sub>3</sub>), 1252 (C=S), 1165, 1124, 1067, 1016, 919, 844, 775 cm<sup>-1</sup>;

**HRMS** (ESI)  $m/z$ : [M-H]<sup>-</sup> Calcd for C<sub>21</sub>H<sub>18</sub>F<sub>3</sub>O<sub>2</sub>S<sup>-</sup> 391.0980; Found 391.0976.



### **(Z)-1-hydroxy-4-(hydroxy(1-tosyl-1H-indol-3-yl)methyl)-1-(p-tolyl)hepta-1,6-diene-3-thione 5h**

From 1-tosyl-1H-indole-3-carbaldehyde<sup>55</sup> (300 mg, 1 mmol, 1 equiv) and 1-(p-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) **5h** was synthesized following the general procedure. After addition of DBU the reaction was stirred for 18 h at rt and purified by flash chromatography (Hexane/Et<sub>2</sub>O/AcOH 10:1:0.025) to give **5h** as two diastereomers (3:1) (60 mg, 11%) as a thick red oil.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$  15.82 (s, 1H Minor), 15.47 (s, 1H Major), 7.99 (d,  $J = 8.2$  Hz, 1H Minor), 7.94 (d,  $J = 7.9$  Hz, 1H Major), 7.77 (dd,  $J = 7.0, 1.0$  Hz, 1H Major), 7.70 (d,  $J = 7.6$  Hz, 1H Minor), 7.66 (d,  $J = 8.4$  Hz, 2H Minor), 7.59 (d,  $J = 8.4$  Hz, 2H Major), 7.55 (s, 1H Minor), 7.52 (s, 1H Major), 7.42 (d,  $J = 8.3$  Hz, 2H Minor), 7.34 – 7.28 (m, 4H Major + 2H Minor), 7.16 (d,  $J = 8.1$  Hz, 2H Minor), 7.13 (d,  $J = 8.1$  Hz, 2H Major), 7.07 (d,  $J = 8.1$  Hz, 2H Minor), 7.00 (d,  $J = 8.1$  Hz, 2H Major), 6.71 (s, 1H Minor), 6.61 (s, 1H Major), 5.76 – 5.61 (m, 1H Major + 1H Minor), 5.31 – 5.27 (m, 1H Minor), 5.18 (d,  $J = 6.7$  Hz, 1H Major), 5.07 – 5.00 (m, 1H Major + 1H Minor), 4.97 – 4.92 (m, 1H Major + 1H Minor), 3.54 – 3.49 (m, 1H Minor), 3.49 – 3.42 (m, 1H Major), 2.82 – 2.73 (m, 1H Major + 1H Minor), 2.77 (br(s), OH Minor), 2.67 (br(s), OH Major), 2.39 (s, 3H Minor), 2.38 (s, 3H Major), 2.36 – 2.33 (m, 1H Major), 2.23 (s, 3H Minor), 2.17 (s, 3H Major), 2.05 – 1.97 (m, 1H Minor);

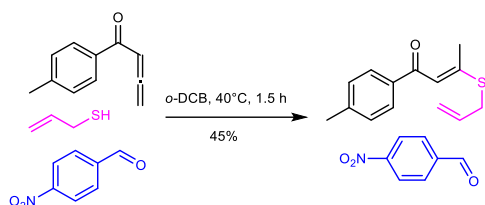
**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>)  $\delta$  215.2 (Major + Minor), 178.6 (Major), 178.3 (Minor), 144.8 (Minor), 144.75 (Major), 143.9 (Minor), 143.6 (Major), 135.7 (Major + Minor), 135.5 (Minor), 135.4 (Major), 134.92 (Major), 134.88 (Minor), 131.9 (Major), 131.7 (Minor), 129.78 (Major),

129.76 (Minor), 129.5 (Minor), 129.4 (Major), 128.8 (Minor), 128.7 (Major), 127.3 (Minor), 127.1 (Major), 126.6 (Minor), 126.5 (Major), 124.8 (Minor), 124.7 (Major), 124.4 (Major), 124.1 (Minor), 123.8 (Minor), 123.32 (Minor), 123.29 (Major), 122.9 (Major), 120.5 (Major), 120.3 (Minor), 117.3 (Minor), 116.8 (Major), 114.0 (Minor), 113.98 (Minor), 113.89 (Major), 113.2 (Major), 71.6 (Major), 70.8 (Minor), 62.8 (Major), 62.6 (Minor), 37.1 (Minor), 35.1 (Major), 21.7 (Minor), 21.6 (Major), 21.5 (Minor), 21.4 (Major);

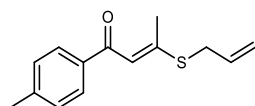
**IR** (film, DCM) 3430 (OH), 3066, 2953, 2922, 2856 (C-S), 1605, 1579, 1150, 1503, 1446, 1368, 1261, 1173, 1121, 1096, 1019, 916, 811, 747  $\text{cm}^{-1}$ ;

**HRMS** (ESI)  $m/z$ :  $[\text{M}-\text{H}]^-$  Calcd for  $\text{C}_{30}\text{H}_{29}\text{NNaO}_4\text{S}_2^+$  554.1436; Found 554.1444.

### Mechanistic validation by stepwise isolation



**Supplementary Figure S83.** Reaction scheme of the stepwise process to confirm the first step intermediate



### (E)-3-(allylthio)-1-(p-tolyl)but-2-en-1-one **S6.5.1**

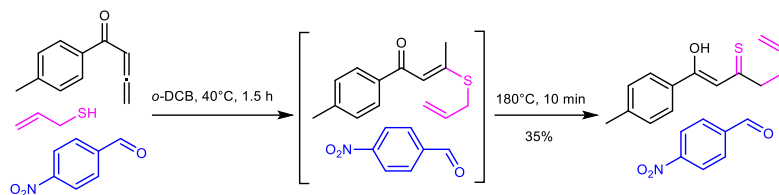
To an argon-purged sealed flask containing anhydrous *o*-DCB (3.2 mL) was added terminal allene (1 mmol), prop-2-ene-1-thiol (0.1 mL, 89 mg, 1.2 mmol, 1.2 equiv) and *p*-nitrobenzaldehyde (151 mg, 1 mmol, 1 equiv). The reaction flask was put into the bath preheated to 40 °C, and stirred for 1.5 h. After completion of the first step, monitored by TLC, without evaporating the solvent, the crude mixture was directly loaded to the silica gel, purified by flash chromatography (Hexane/Et<sub>2</sub>O 20:1) to yield **S6.5.1** as a yellow oil (105 mg, 45%).

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.79 (d,  $J$  = 8.1 Hz, 2H), 7.24 (d,  $J$  = 8.4 Hz, 2H), 6.67 (s, 1H), 5.92 (ddt,  $J$  = 16.8, 10.1, 6.5 Hz, 1H), 5.38 (dd,  $J$  = 16.8, 0.5 Hz, 1H), 5.28 (dd,  $J$  = 10.1, 0.4 Hz, 1H), 3.56 (d,  $J$  = 6.5 Hz, 2H), 2.49 (s, 3H), 2.40 (s, 3H);

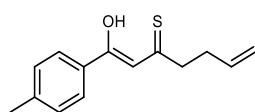
**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  187.6, 159.4, 142.7, 137.2, 131.7, 129.2, 128.1, 118.9, 114.3, 34.9, 21.7, 21.6;

**IR** (film, DCM) : 3084, 3028, 2952, 2919 (C-S), 1645 (C=O), 1606, 1557, 1424, 1375, 1234, 1181, 1110, 1051, 1018, 924, 858, 810  $\text{cm}^{-1}$ ;

**HRMS** (EI)  $m/z$ :  $[M]^+$  Calcd for  $\text{C}_{14}\text{H}_{16}\text{OS}^+$ : 232.0922; Found: 232.0926.



**Supplementary Figure S84.** Reaction scheme of the stepwise process to confirm the second step intermediate



### **(Z)-1-hydroxy-1-(p-tolyl)hepta-1,6-diene-3-thione S6.5.2**

To an argon-purged sealed flask containing anhydrous *o*-DCB (3.2 mL) was added terminal allene (1 mmol), prop-2-ene-1-thiol (0.1 mL, 89 mg, 1.2 mmol, 1.2 equiv) and *p*-nitrobenzaldehyde (151 mg, 1 mmol, 1 equiv). The reaction flask was put into a pre-heated 40°C bath and stirred for 1.5 h. After completion of the first step, monitored by TLC, the reaction mixture was put into pre-heated 180°C oil bath, stirred for 10 min. and cooled to rt. Completion of the thio-Claisen rearrangement was monitored by TLC. The crude mixture was directly loaded to the silica gel, purified by flash chromatography (Hexane/Et<sub>2</sub>O 5:1) to yield **S6.5.2** as red thick oil (82 mg, 35%).

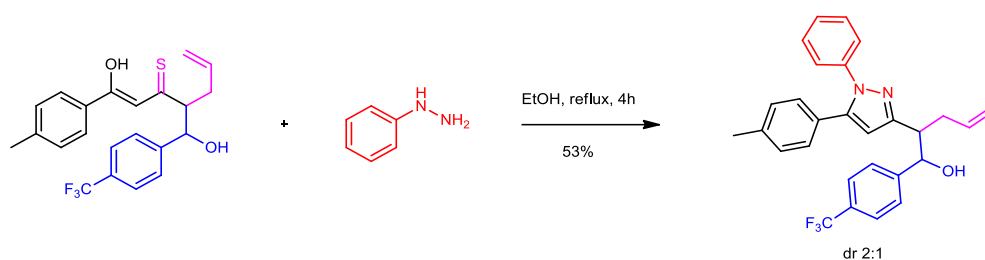
**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>)  $\delta$  14.30 (s, 1H), 7.84 (d,  $J = 8.2$  Hz, 2H), 7.27 (d,  $J = 8.2$  Hz, 2H), 7.01 (s, 1H), 6.02 – 5.72 (m, 1H), 5.10 (ddd,  $J = 17.0, 3.1, 1.5$  Hz, 1H), 5.03 (dd,  $J = 10.2, 1.4$  Hz, 1H), 2.87 – 2.79 (m, 2H), 2.54 – 2.49 (m, 2H), 2.42 (s, 3H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  207.1, 180.0, 143.4, 136.8, 132.9, 129.5, 127.3, 115.7, 111.7, 48.3, 33.9, 21.7;

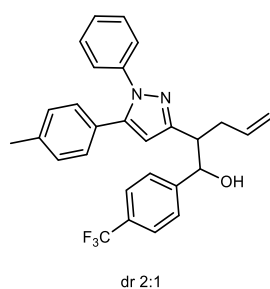
**IR** (film, DCM): 3077, 3031, 2978, 2919 (C-S), 1607, 1580, 1553, 1504, 1444, 1252 (C=S), 1187, 1119, 1050, 992, 915, 806  $\text{cm}^{-1}$ ;

**HRMS** (EI)  $m/z$ :  $[M]^+$  Calcd for  $\text{C}_{14}\text{H}_{16}\text{OS}^+$ : 232.0922; Found: 232.0926.

## Application example



## Supplementary Figure S85. Reaction scheme of pyrazole formation from the Mach5 product



### 2-(1-phenyl-5-(*p*-tolyl)-1H-pyrazol-3-yl)-1-(4-(trifluoromethyl)phenyl)pent-4-en-1-ol S6.5.3<sup>56</sup>

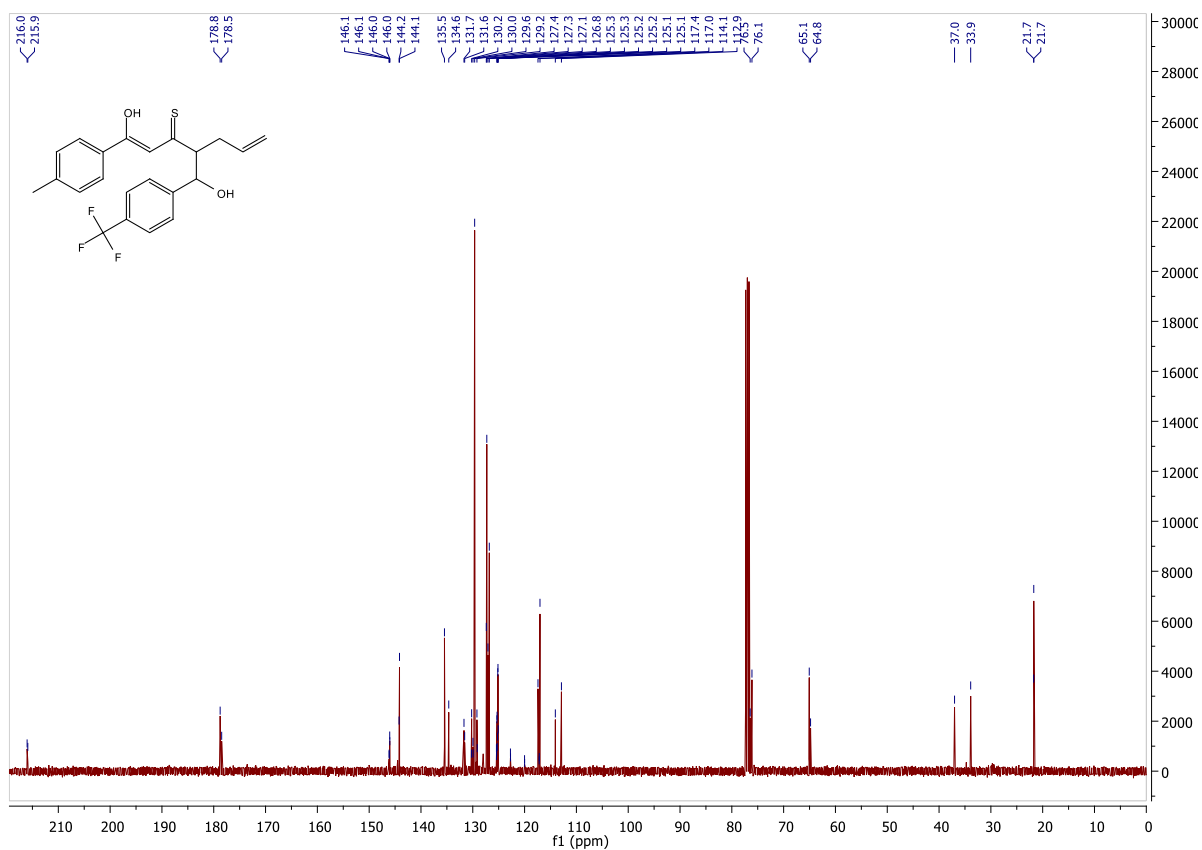
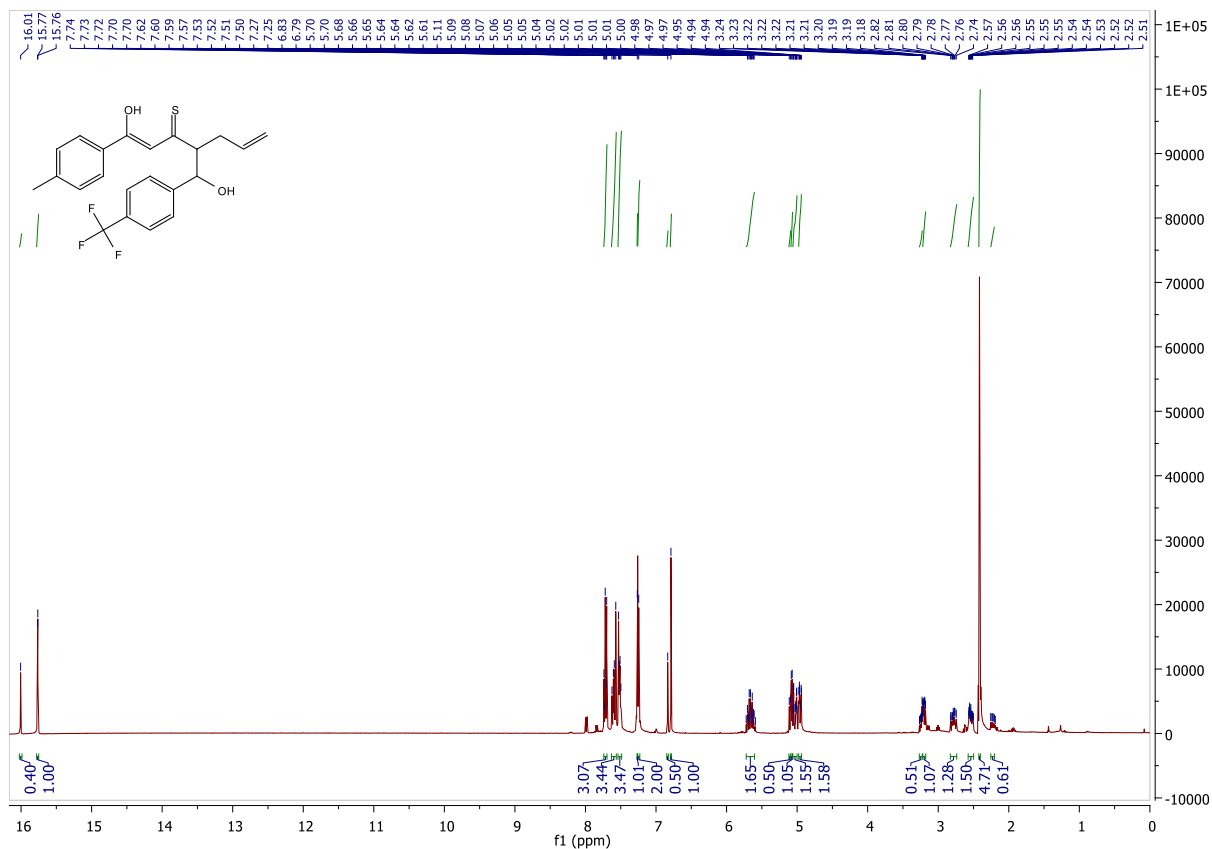
Phenyl hydrazine (0.04 mL, 42 mg, 0.39 mmol, 1.1 equiv.) was added to the solution of **5a** (as two diastereoisomers (2:1)) (144 mg, 0.35 mmol, 1 equiv) in EtOH (2.8 mL) and refluxed for 4 h. After the completion of the reaction (monitored by TLC), the reaction mixture was concentrated, dissolved in DCM, poured into water and extracted with DCM (x3), washed with water (x2) and brine, dried over MgSO<sub>4</sub>, filtered and concentrated. The crude mixture was purified by flash chromatography (Hexane/EtOAc 3:1) to give **S6.5.3** as two diastereomers (2:1) (86 mg, 53%) as thick orange oil.

<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.59 (d, *J* = 8.1 Hz, 2H Major), 7.56 (d, *J* = 8.1 Hz, 2H Minor), 7.47 (d, *J* = 8.0 Hz, 2H Major), 7.43 (d, *J* = 8.0 Hz, 2H Minor), 7.37 – 7.24 (m, 5H Major + 5H Minor), 7.13 – 7.03 (m, 4H Major + 4H Minor), 6.23 (s, 1H Major), 6.17 (s, 1H Minor), 5.84 (ddt, *J* = 17.0, 10.1, 6.9 Hz, 1H Minor), 5.75 (ddt, *J* = 17.0, 10.1, 6.9 Hz, 1H Major), 5.21 (d, *J* = 3.0 Hz, 1H Major), 5.09 (d, *J* = 17.1 Hz, 1H Minor), 5.05 (d, *J* = 10.2 Hz, 1H Minor), 5.01 – 4.94 (m, 2H Major + 1H Minor), 3.29 – 3.24 (m, 1H Major), 3.23 – 3.19 (m, 1H Minor), 2.57 – 2.53 (m, 1H Major), 2.54 – 2.50 (m, 1H Minor), 2.34 (s, 3H Major), 2.33 (s, 3H Minor), 2.32 – 2.28 (m, 1H Major + 1H Minor);

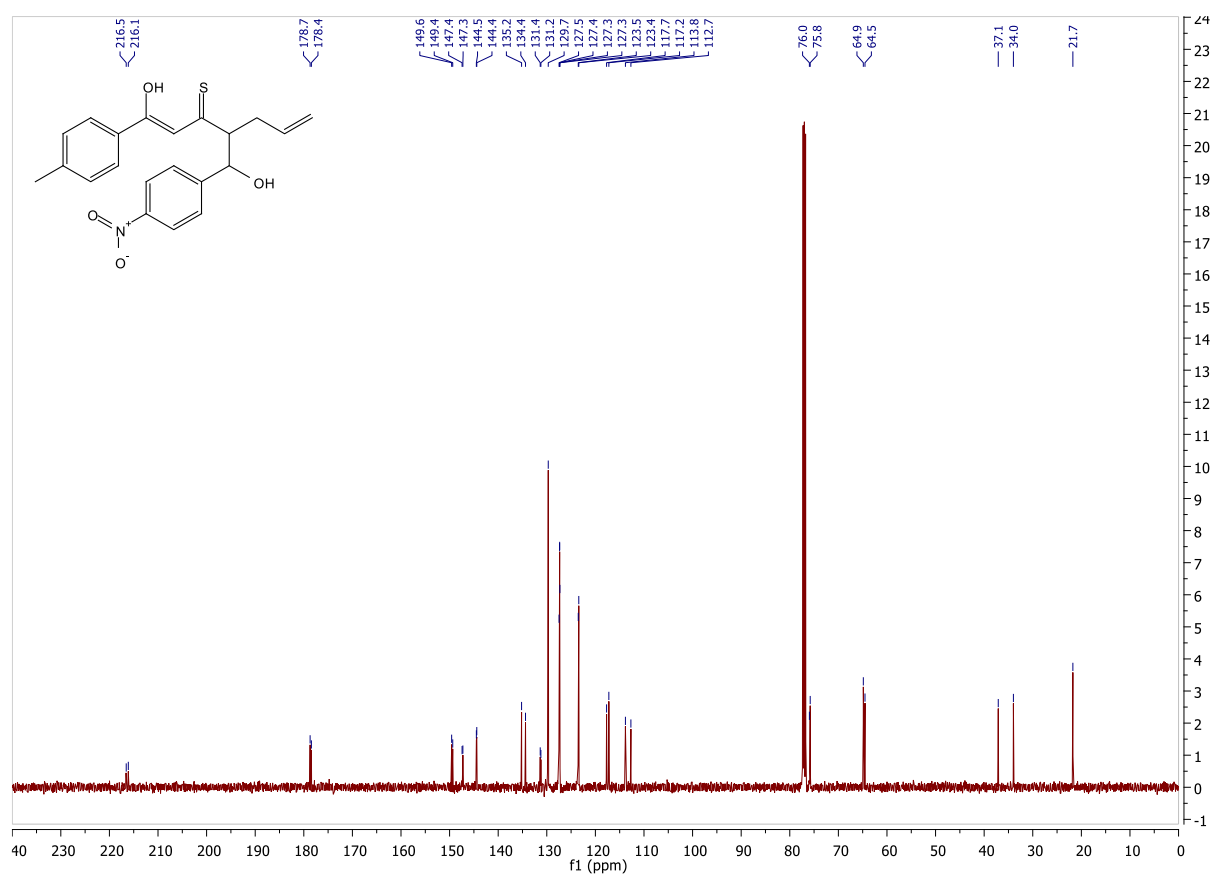
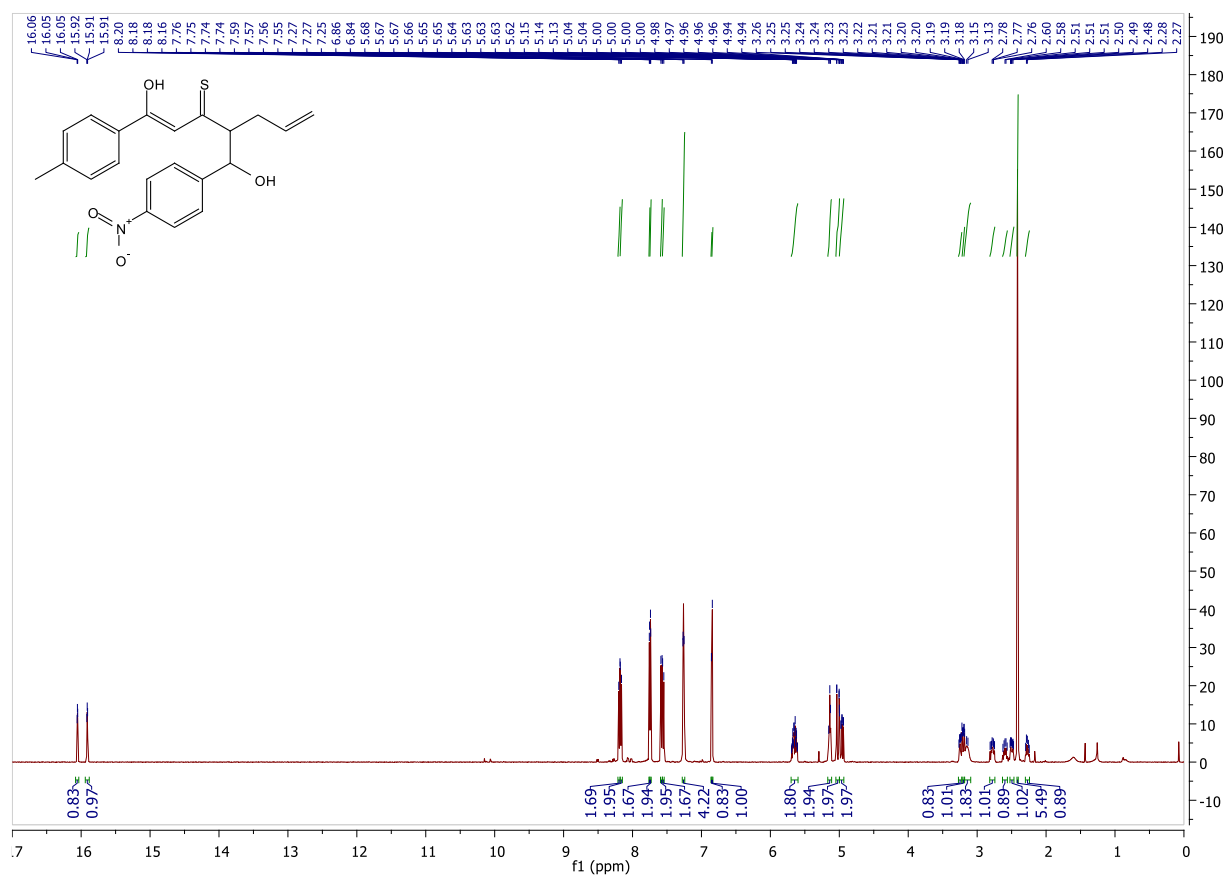
**<sup>13</sup>C NMR** (151 MHz, CDCl<sub>3</sub>) δ 153.8 (Major), 152.9 (Minor), 147.5 (Minor), 146.3 (Major), 143.8 (Minor), 143.6 (Major), 139.8 (Major), 139.7 (Minor), 138.4 (2C Major + 2C Minor) 136.6 (Major), 135.9 (Minor), 129.18 (Major), 129.15 (Minor), 128.89 (Minor), 128.88 (Major), 128.54 (Major), 128.52 (Minor), 127.5 (Minor), 127.4 (Major), 127.3 (Major), 127.2 (Minor), 126.8 (Minor), 126.7 (Major), 125.1 (Minor), 125.09 (Major), 124.99 (q,  $J_3 = 3.7$  Hz Minor), 124.85 (q,  $J_3 = 3.7$  Hz Major), 124.29 (q,  $J_1 = 271.9$  Hz Major), 124.25 (q,  $J_1 = 272.0$  Hz Minor), 116.9 (Minor), 116.4 (Major), 107.40 (Minor), 107.35 (Major), 75.4 (Minor), 75.2 (Major), 46.4 (Minor), 45.9 (Major), 36.4 (Minor), 32.7 (Major), 21.23 (Major), 21.22 (Minor);

**IR** (film, DCM) 3378 (OH), 3073, 2979, 2922, 2864, 1619, 1598, 1505, 1442, 1326, 1164, 1124, 1067, 1017, 913, 824, 465, 694 cm<sup>-1</sup>;

**HRMS** (EI)  $m/z$ : [M-H]<sup>-</sup> Calcd for C<sub>28</sub>H<sub>25</sub>F<sub>3</sub>N<sub>2</sub>O 462.1919 Found 462.1918.



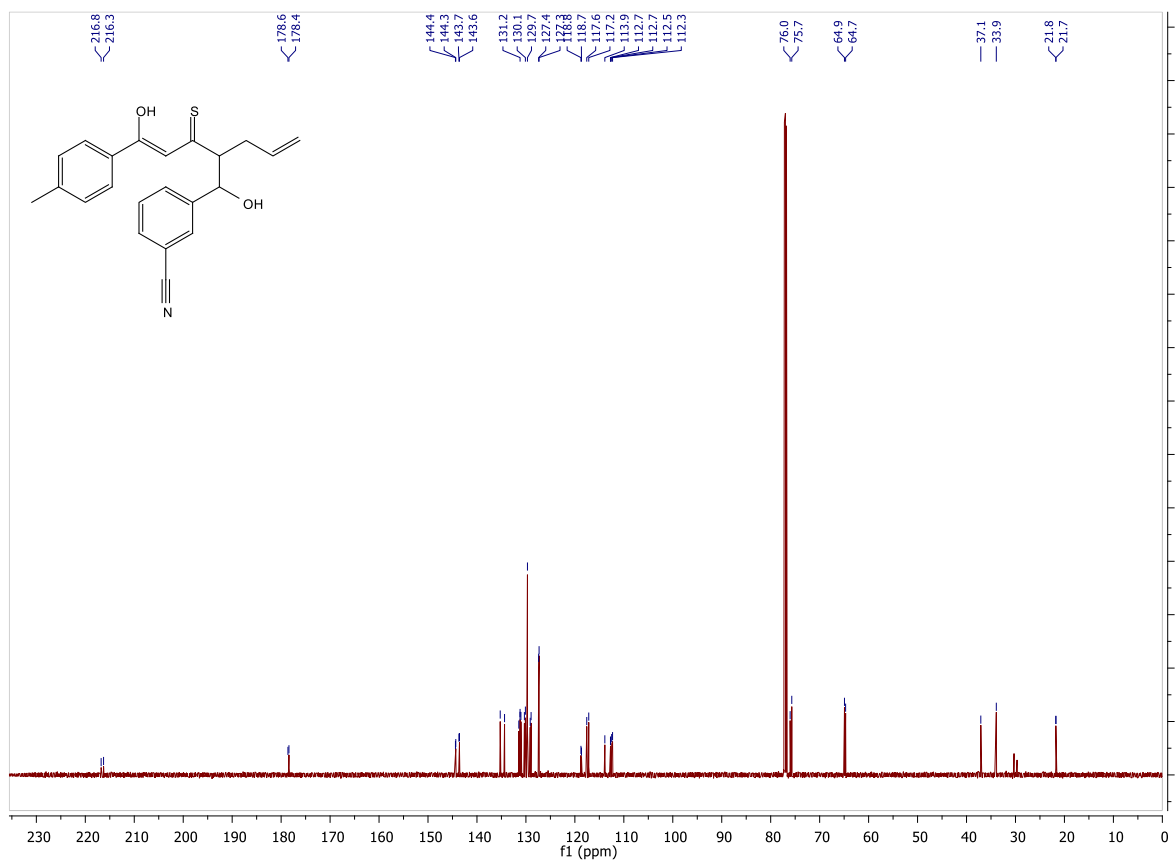
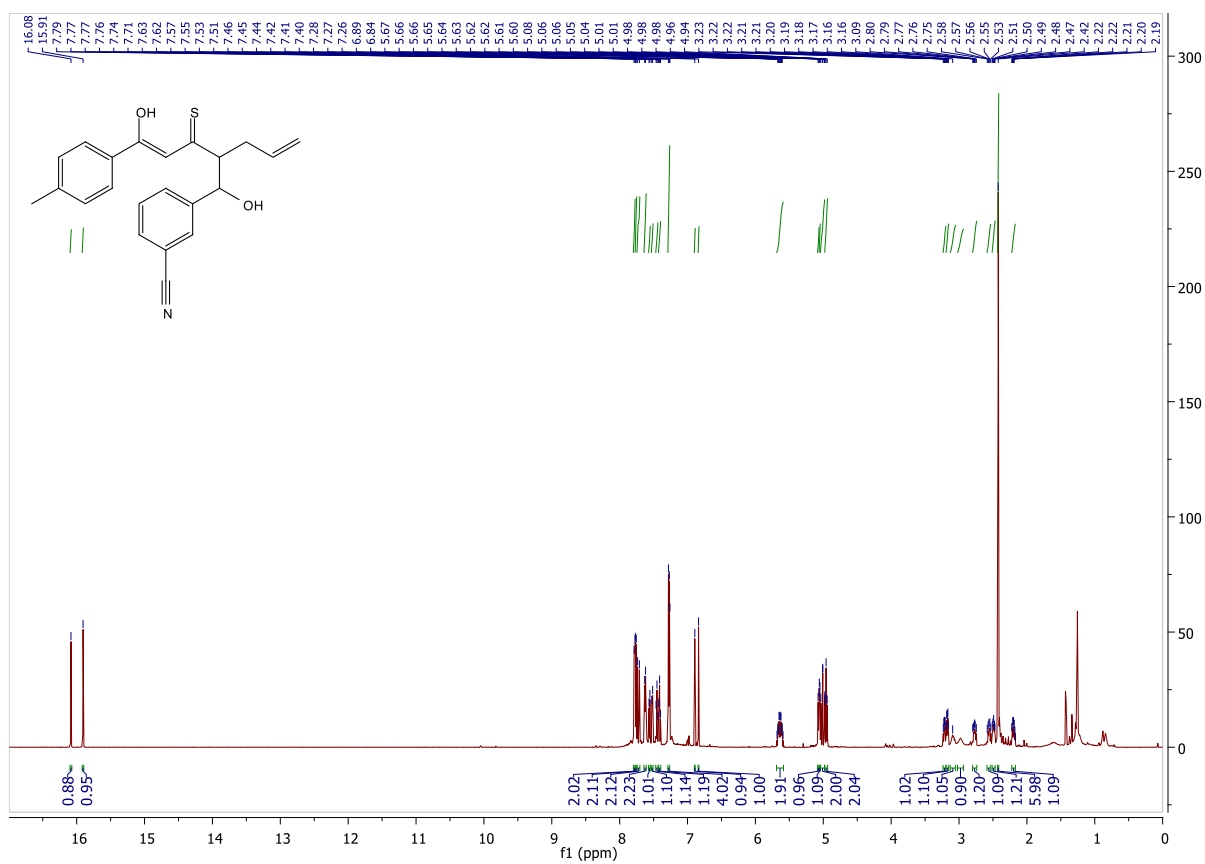
Supplementary Figure S86. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5a**.



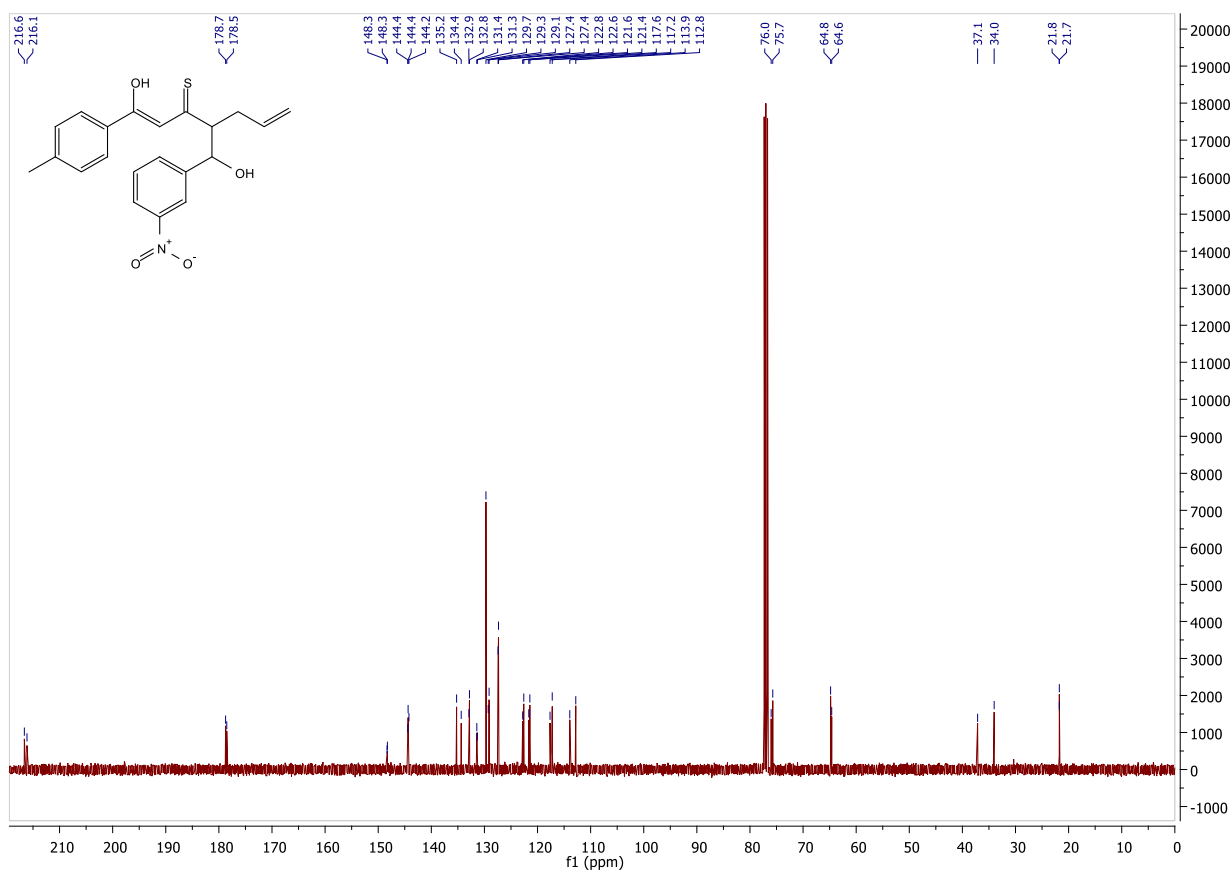
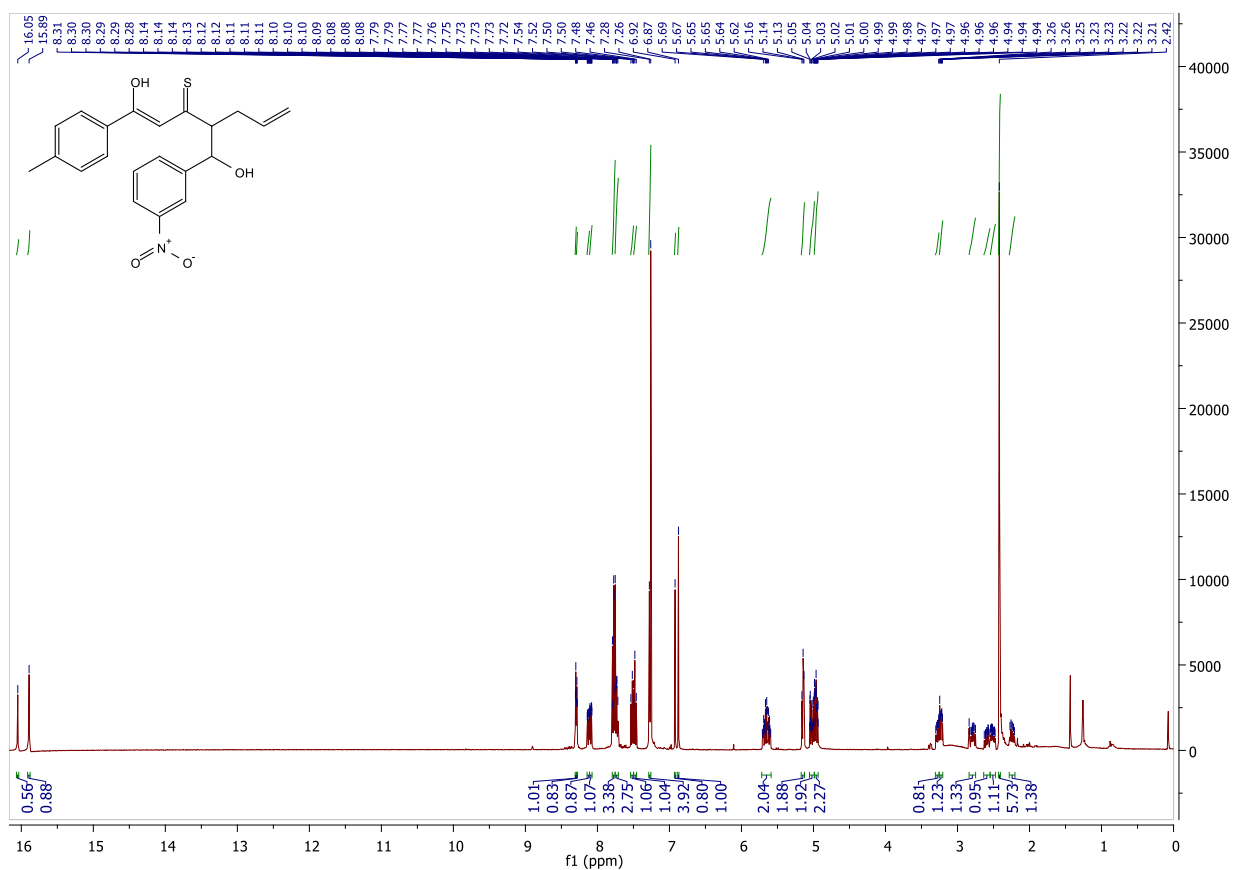
**Supplementary Figure S87.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5c**.



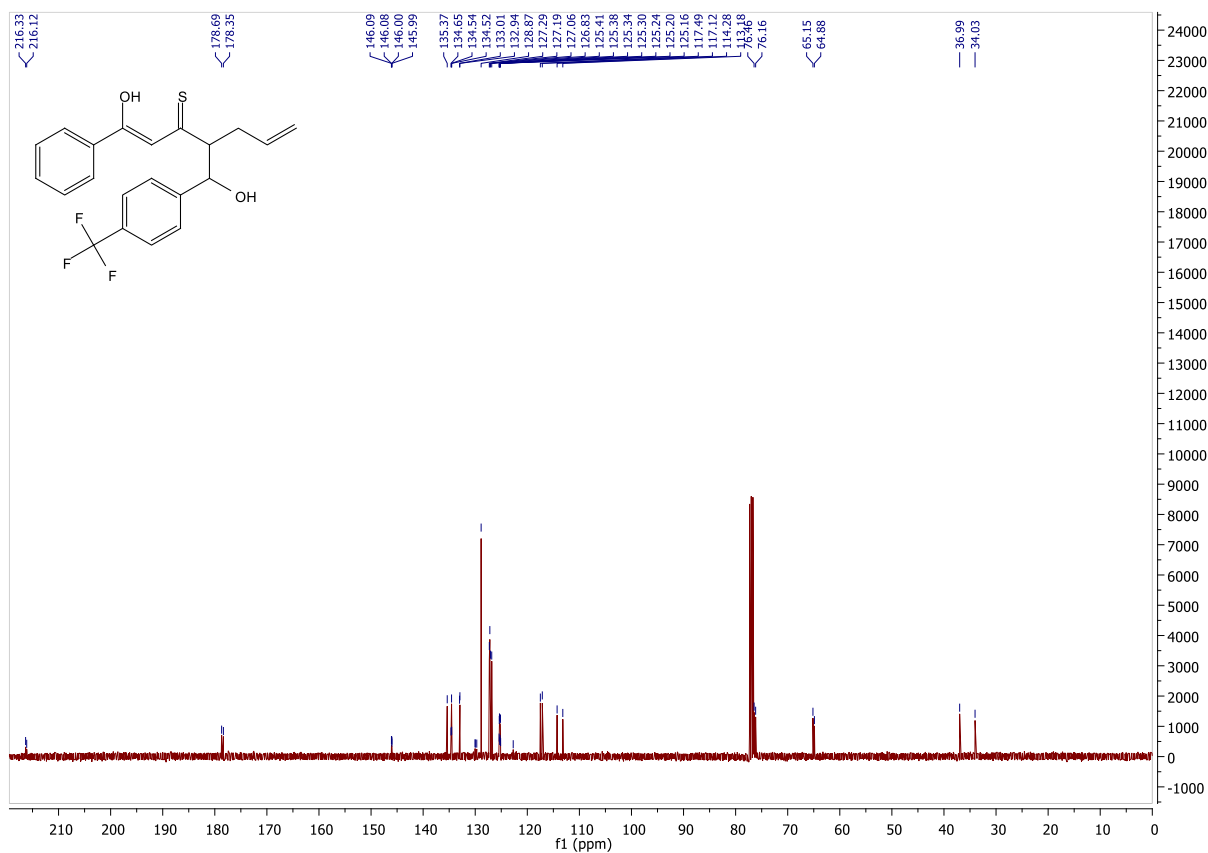
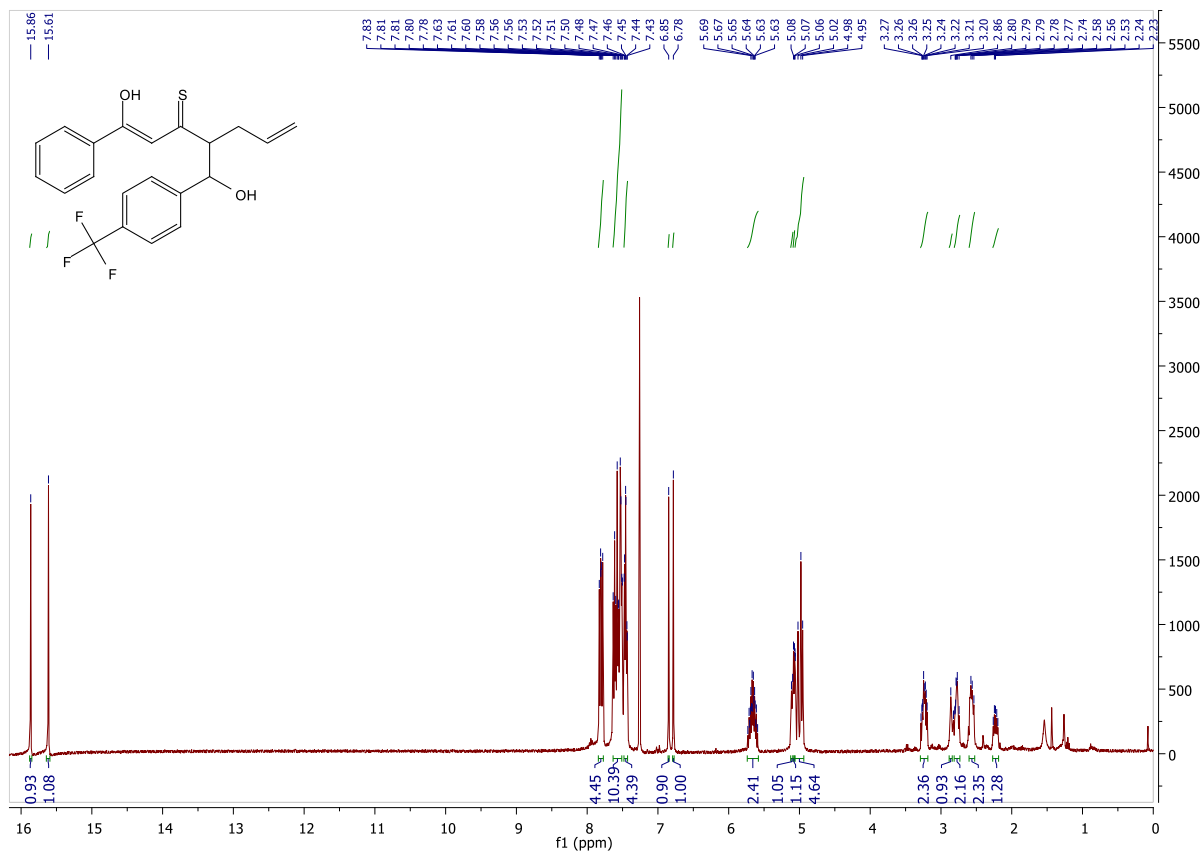




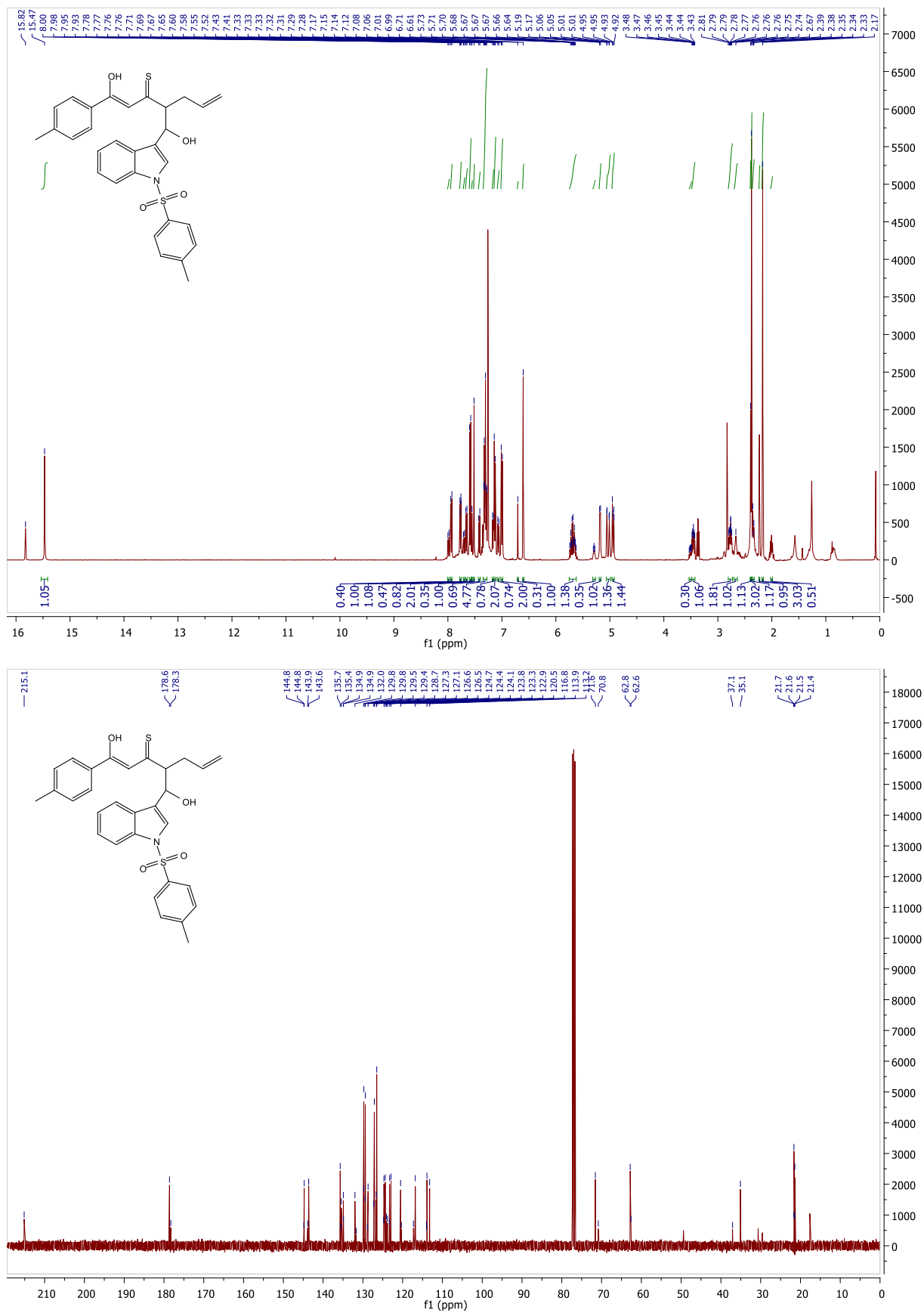
Supplementary Figure S89. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5e**.



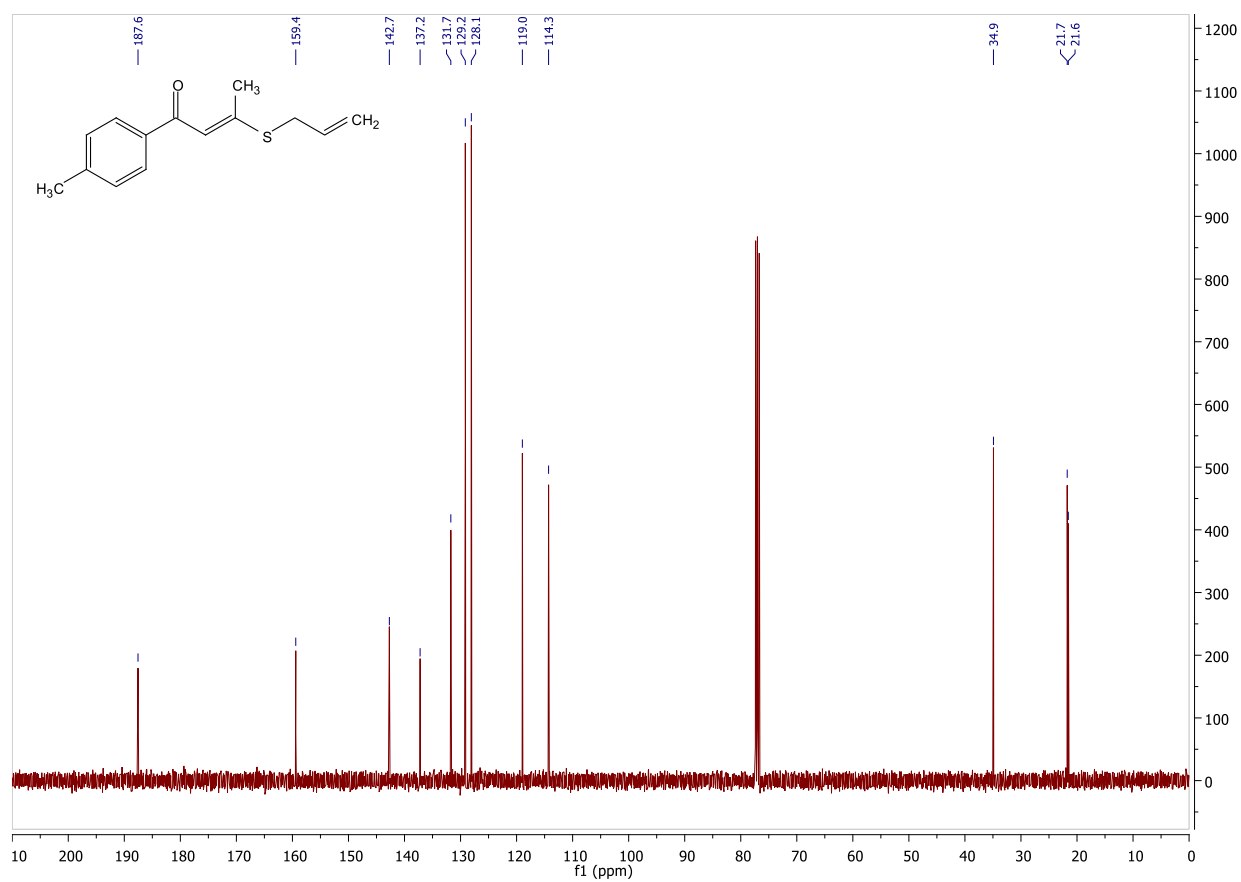
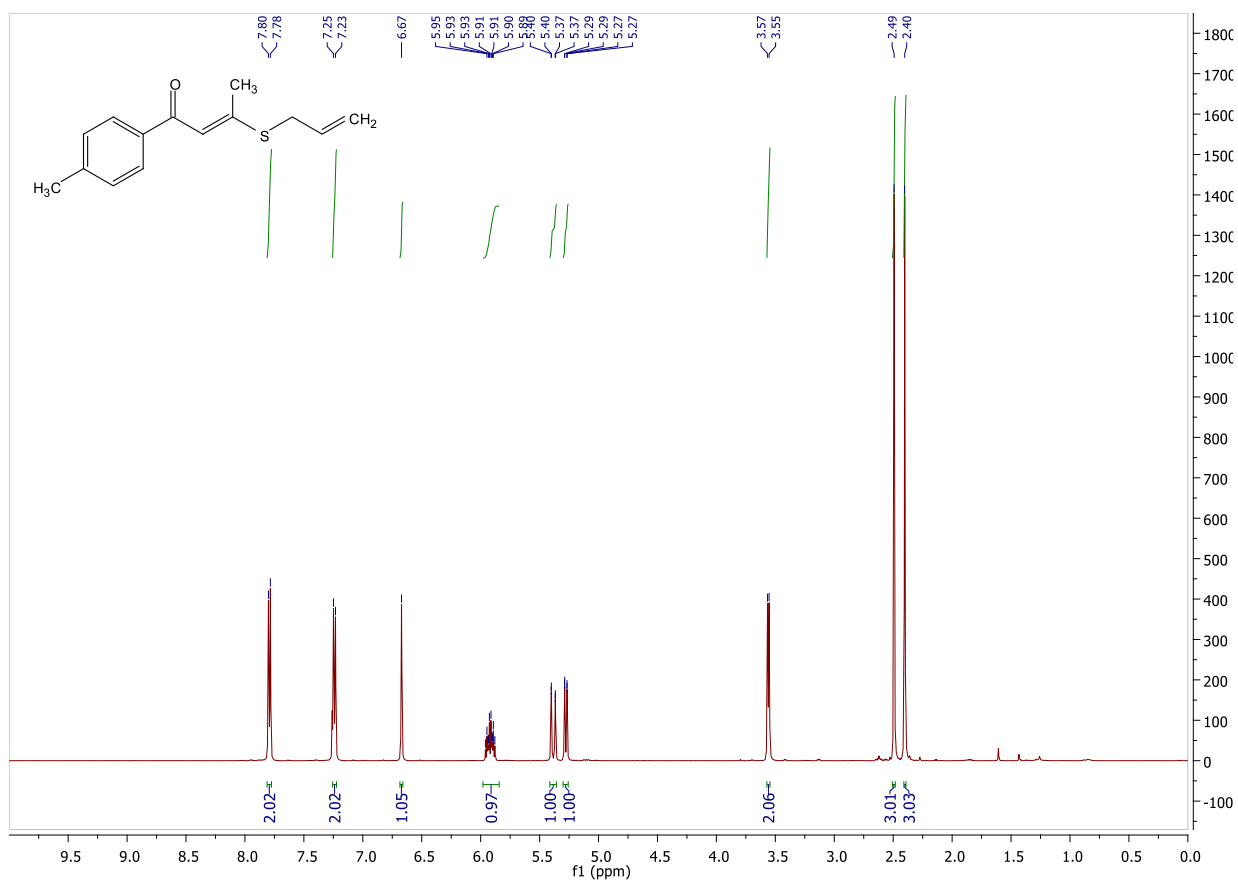
Supplementary Figure S90. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5f**.



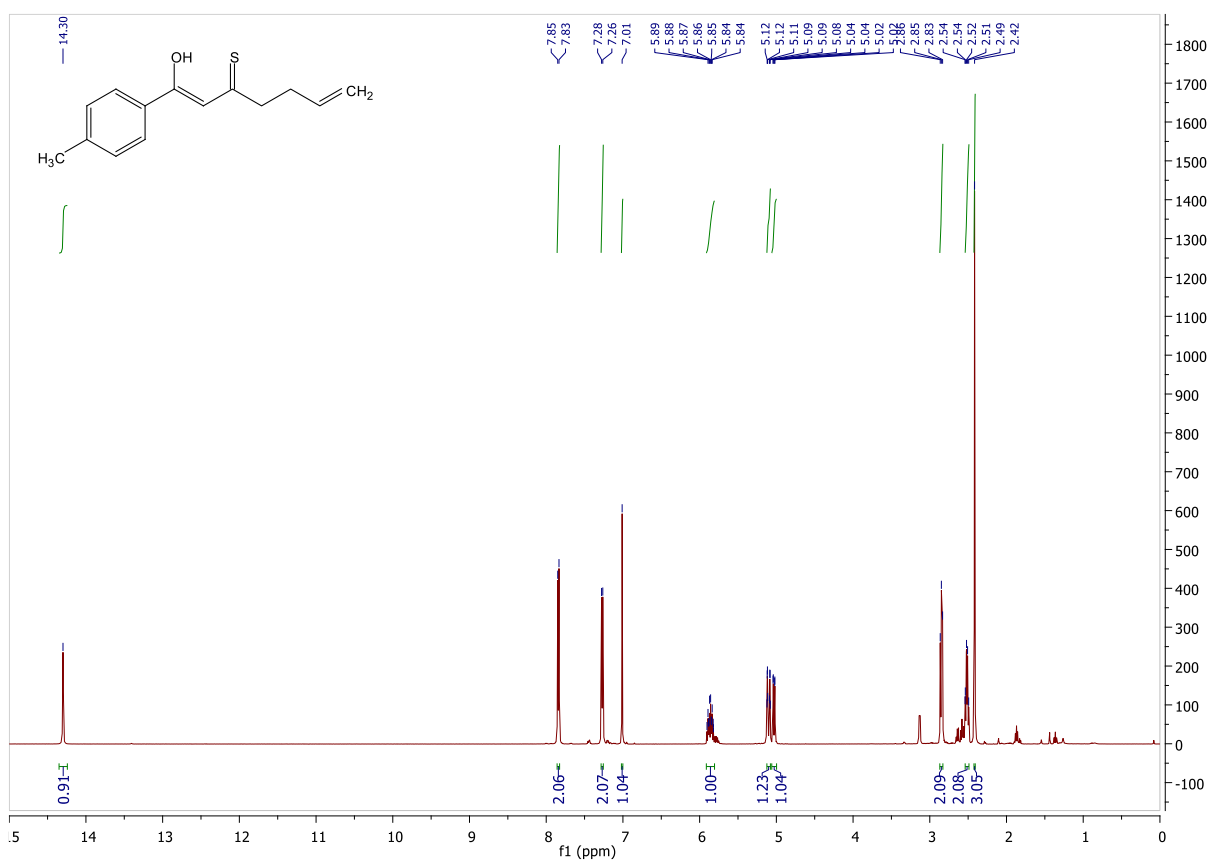
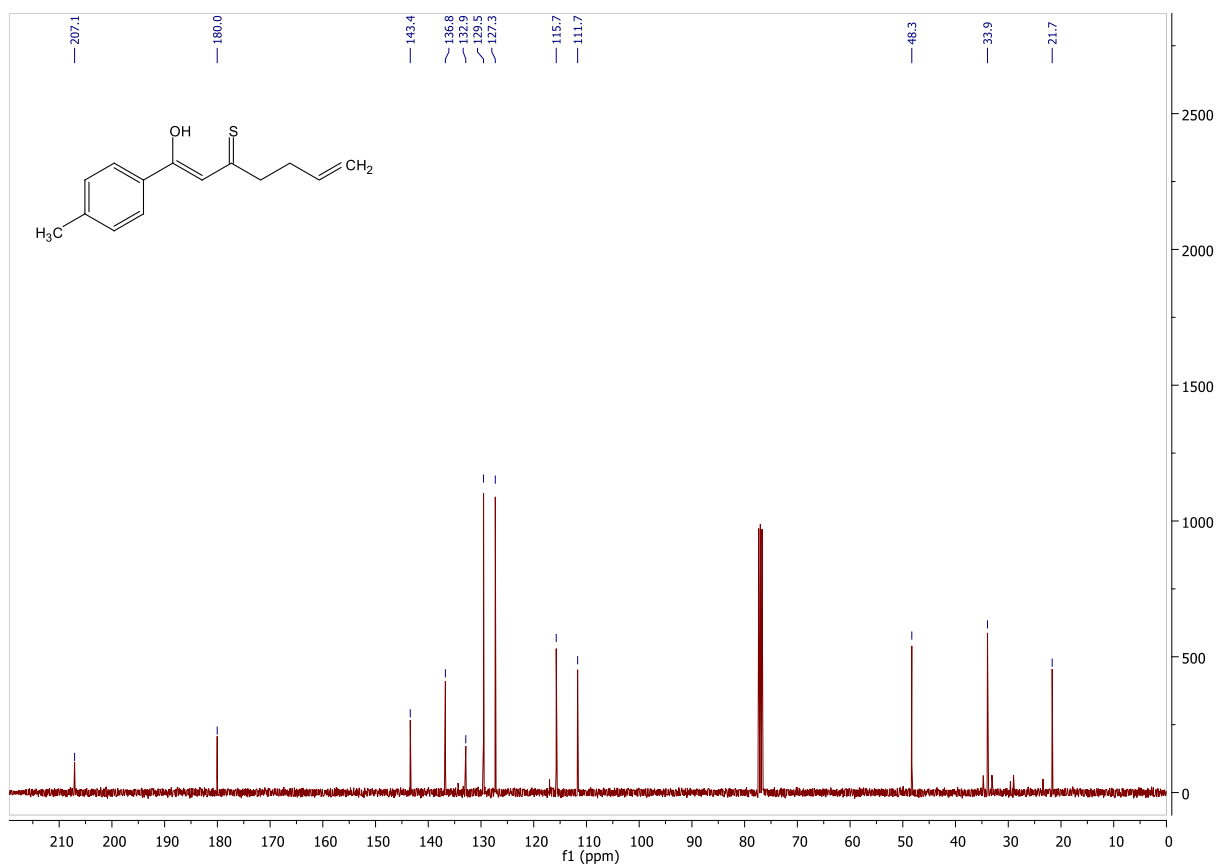
Supplementary Figure S91. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5g**.



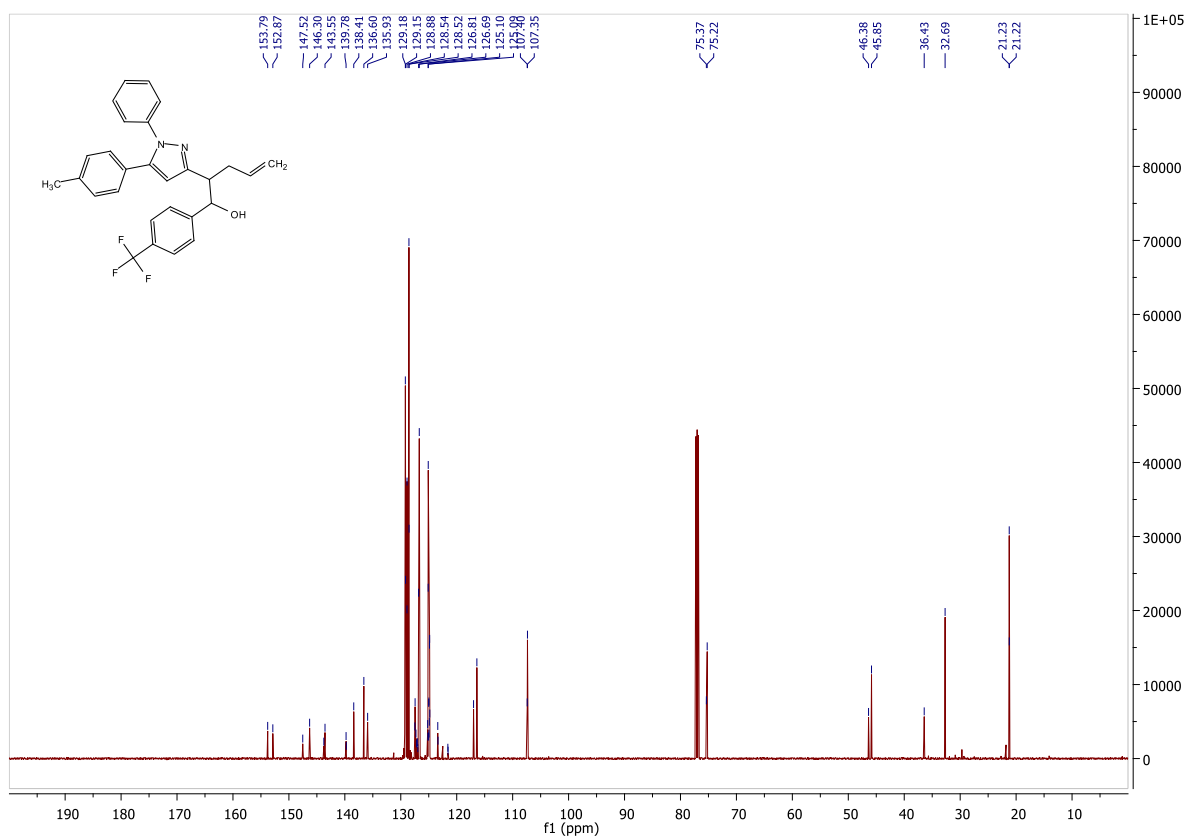
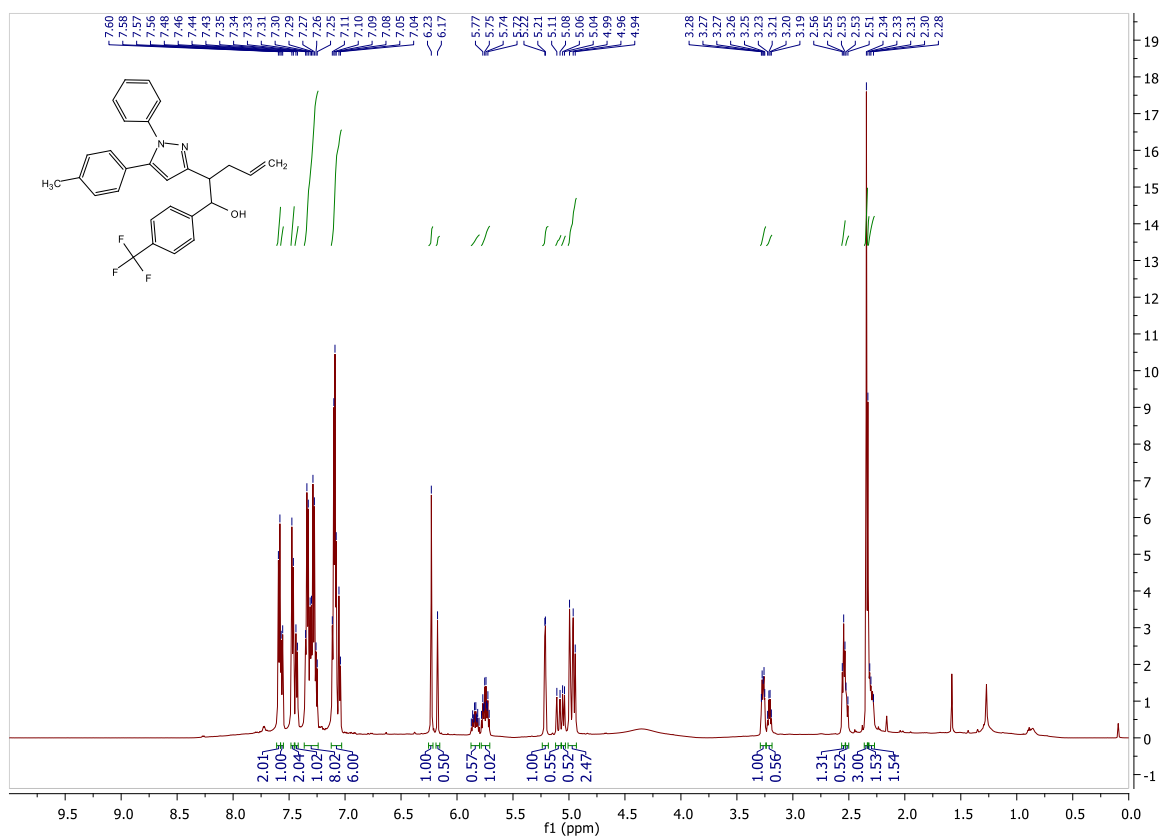
Supplementary Figure S92. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5h**.



**Supplementary Figure S93.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.5.1.

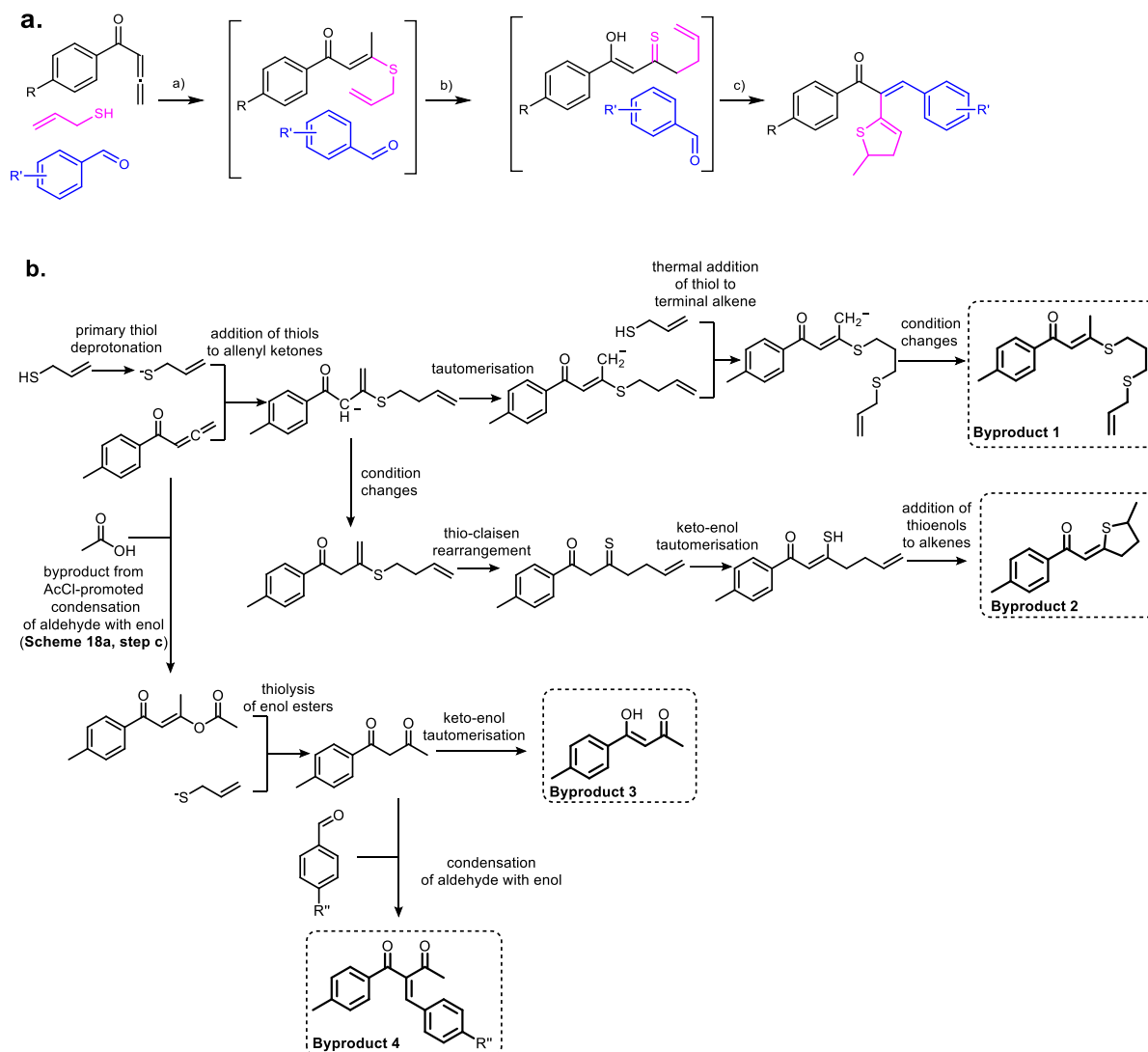


**Supplementary Figure S94.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.5.2.



**Supplementary Figure S95.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.5.3.

**Section S6.6 Reaction Mach6 described in main-text Figure 4c (variant with the addition of acetyl chloride).**



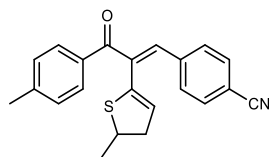
**Supplementary Figure S96. Reaction Mach6. a,** Reagents and conditions: (a) *o*-DCB, 40 °C, 1.5 h (b) 180 °C, 10 min. (c) acetyl chloride, rt, 18 h; **b,** identified byproducts formed during the reaction; all mechanistic steps are consistent with Allchemy's predictions.

**One-pot procedure**

To a sealed-flask containing anhydrous *o*-DCB (3.2 mL) was added terminal allene (1 mmol), prop-2-ene-1-thiol (0.1 mL, 89 mg, 1.2 mmol, 1.2 equiv), and corresponding aldehyde (1 mmol, 1 equiv) respectively. The reaction flask was put in pre-heated oil bath and stirred at 40 °C for 1.5 h. After completion of the first step, monitored by TLC, the reaction mixture was put into pre-heated oil bath, stirred at 180 °C for 10 min. and cooled to rt. Completion of the thio-Claisen rearrangement was monitored by TLC. Then, acetyl chloride (0.71 mL, 78.5 mg, 1 mmol, 1 equiv) was added to the mixture, stirred for 18 h at rt. Upon completion, without evaporating



the solvent, the crude mixture was directly loaded to the silica gel and purified by flash chromatography (Hexane/Et<sub>2</sub>O) to yield the desired product.



**(Z)-4-(2-(5-methyl-4,5-dihydrothiophen-2-yl)-3-oxo-3-(*p*-tolyl)prop-1-en-1-yl)benzotrile 5b**

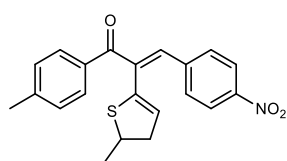
From *p*-cyanobenzaldehyde (262.2 mg, 2 mmol, 1 equiv), and 1-(*p*-tolyl)buta-2,3-dien-1-one (316.4 mg, 2 mmol, 1 equiv) 5b was synthesized following the general procedure. Purification by flash chromatography (Hexane/Et<sub>2</sub>O 4:1) gave 5b (100 mg, 14%) as a yellow oil.

<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.94 (d, *J* = 8.3 Hz, 2H), 7.70 (d, *J* = 8.3 Hz, 2H), 7.50 (s, 1H), 7.49 (d, *J* = 7.9 Hz, 2H), 7.30 (s, 1H), 7.28 (d, *J* = 7.9 Hz, 2H), 3.62 (h, *J* = 6.7 Hz, 1H), 3.24 (ddd, *J* = 14.6, 6.6, 1.9 Hz, 1H), 2.71 (ddd, *J* = 14.6, 7.0, 2.4 Hz, 1H), 2.42 (s, 3H), 1.42 (d, *J* = 6.7 Hz, 3H);

<sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ 187.9, 164.4, 145.2, 142.9, 140.9, 136.1, 132.2, 129.9, 129.3, 127.9, 123.5, 118.6, 111.4, 107.5, 42.8, 41.3, 21.6, 21.3;

IR (film, DCM) 3056, 3035, 2959, 2922, 2865 (C-S), 2226 (CN), 1625, 1605 (C=O), 1568, 1522, 1450, 1412, 1339, 1241, 1181, 1117, 1056, 970, 910, 884, 801 cm<sup>-1</sup>;

HRMS (ESI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>22</sub>H<sub>19</sub>NNaOS<sup>+</sup> 368.1085; Found 368.1088.



**(Z)-2-(5-methyl-4,5-dihydrothiophen-2-yl)-3-(4-nitrophenyl)-1-(*p*-tolyl)prop-2-en-1-one S6.6.1**

From *p*-nitrobenzaldehyde (151.1 mg, 1 mmol, 1 equiv) and 1-(*p*-tolyl)buta-2,3-dien-1-one (158.2 mg, 1 mmol, 1 equiv) S6.6.1 was synthesized following the general procedure. Purification by column chromatography (Hexane/Et<sub>2</sub>O 10:1) gave S6.6.1 (30 mg, 8%) as a yellow liquid oil.

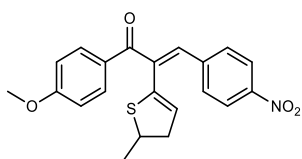
<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 8.26 (d, *J* = 8.7 Hz, 2H), 7.93 (d, *J* = 8.1 Hz, 2H), 7.54 (d, *J* = 8.7 Hz, 2H), 7.51 (s, 1H), 7.34 (s, 1H), 7.27 (d, *J* = 8.1 Hz, 2H), 3.67 – 3.57 (m, 1H), 3.25 (ddd,

$J = 14.7, 6.7, 1.7$  Hz, 1H), 2.72 (ddd,  $J = 14.7, 6.7, 2.3$  Hz, 1H), 2.41 (s, 3H), 1.41 (d,  $J = 6.7$  Hz, 3H);

$^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  187.9, 164.2, 146.9, 145.9, 142.9 (2C), 136.1, 130.1, 129.3, 127.9, 123.8, 123.1, 107.7, 42.8, 41.3, 21.6, 21.4;

IR (film, DCM) 3075, 2959, 2922, 2864 (C-S), 1625, 1605 (C=O), 1569, 1519 ( $\text{NO}_2$ ), 1342 ( $\text{NO}_2$ ), 1240, 1182, 1110, 1055, 970, 864, 799, 733  $\text{cm}^{-1}$ ;

HRMS (ESI)  $m/z$ :  $[\text{M}+\text{H}]^+$  Calcd for  $\text{C}_{21}\text{H}_{20}\text{NO}_3\text{S}^+$  366.1164; Found 366.1171.



**(Z)-1-(4-methoxyphenyl)-2-(5-methyl-4,5-dihydrothiophen-2-yl)-3-(4-nitrophenyl)prop-2-en-1-one S6.6.2**

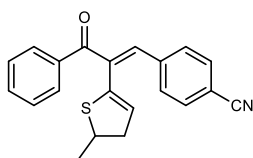
From *p*-nitrobenzaldehyde (131.1 mg, 1 mmol, 1 equiv), and 1-(4-methoxyphenyl)buta-2,3-dien-1-one (174.2 mg, 1 mmol, 1 equiv) **S6.6.2** was synthesized following the general procedure. Purification by flash chromatography (Hexane/ $\text{Et}_2\text{O}$  10:1) gave **S6.6.2** (32 mg, 8%) as a yellow oil.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.27 (d,  $J = 8.8$  Hz, 2H), 8.02 (d,  $J = 8.9$  Hz, 2H), 7.55 (d,  $J = 8.8$  Hz, 2H), 7.50 (s, 1H), 7.34 (s, 1H), 6.96 (d,  $J = 8.9$  Hz, 2H), 3.87 (s, 3H), 3.67 – 3.55 (m, 1H), 3.25 (ddd,  $J = 14.8, 6.7, 1.8$  Hz, 1H), 2.72 (ddd,  $J = 14.7, 6.7, 2.4$  Hz, 1H), 1.41 (d,  $J = 6.7$  Hz, 3H);

$^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  187.0, 163.7, 162.9, 146.9, 145.9, 142.9, 131.6, 130.1, 130.0, 123.8, 122.9, 113.8, 107.6, 55.4, 42.7, 41.3, 21.4;

IR (film, DCM) 3074, 2958, 2925, 2841 (C-S), 1623, 1597 (C=O), 1572, 1512 ( $\text{NO}_2$ ), 1459, 1341 ( $\text{NO}_2$ ), 1238, 1170, 1111, 1055, 969, 886, 805, 732  $\text{cm}^{-1}$ ;

HRMS (ESI)  $m/z$ :  $[\text{M}+\text{H}]^+$  Calcd for  $\text{C}_{21}\text{H}_{19}\text{NNaO}_4\text{S}^+$  404.0932; Found 404.0933.



**(Z)-4-(2-(5-methyl-4,5-dihydrothiophen-2-yl)-3-oxo-3-phenylprop-1-en-1-yl)benzonitrile  
S6.6.3**

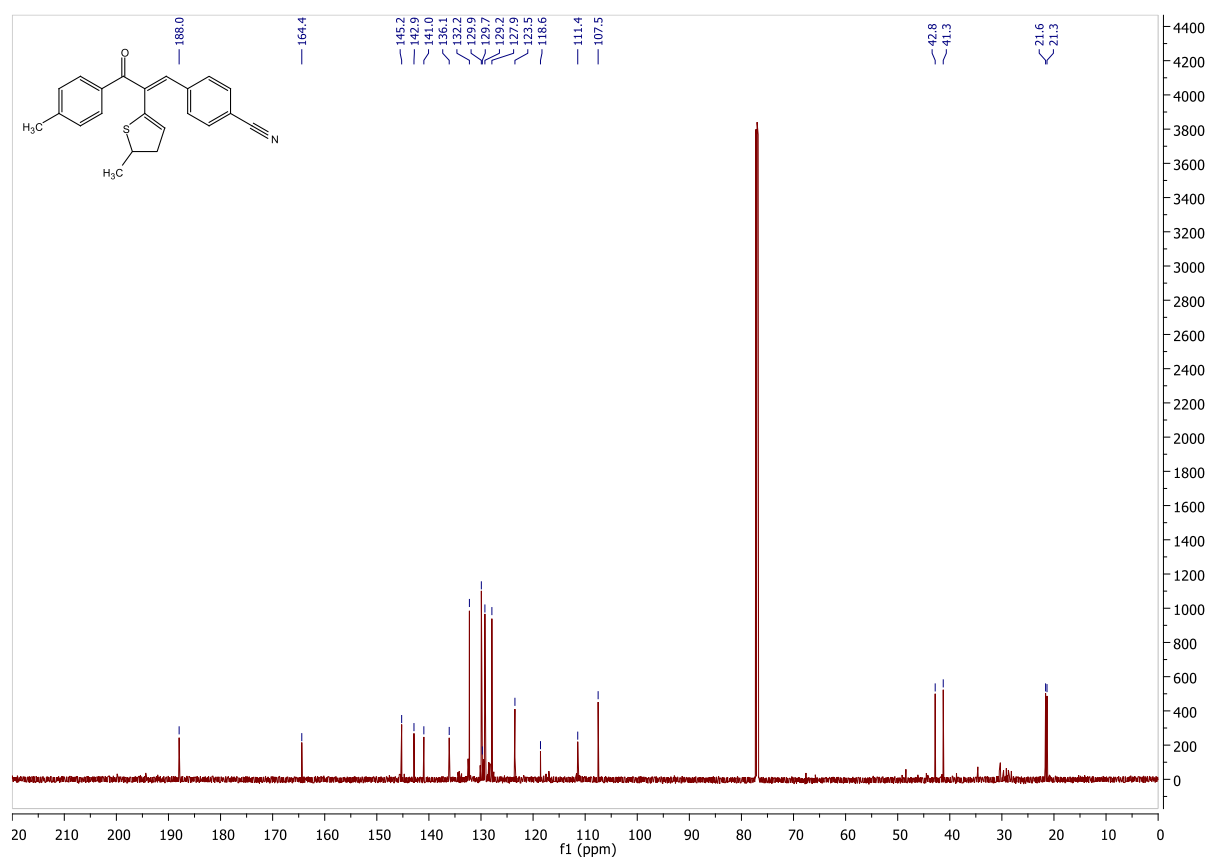
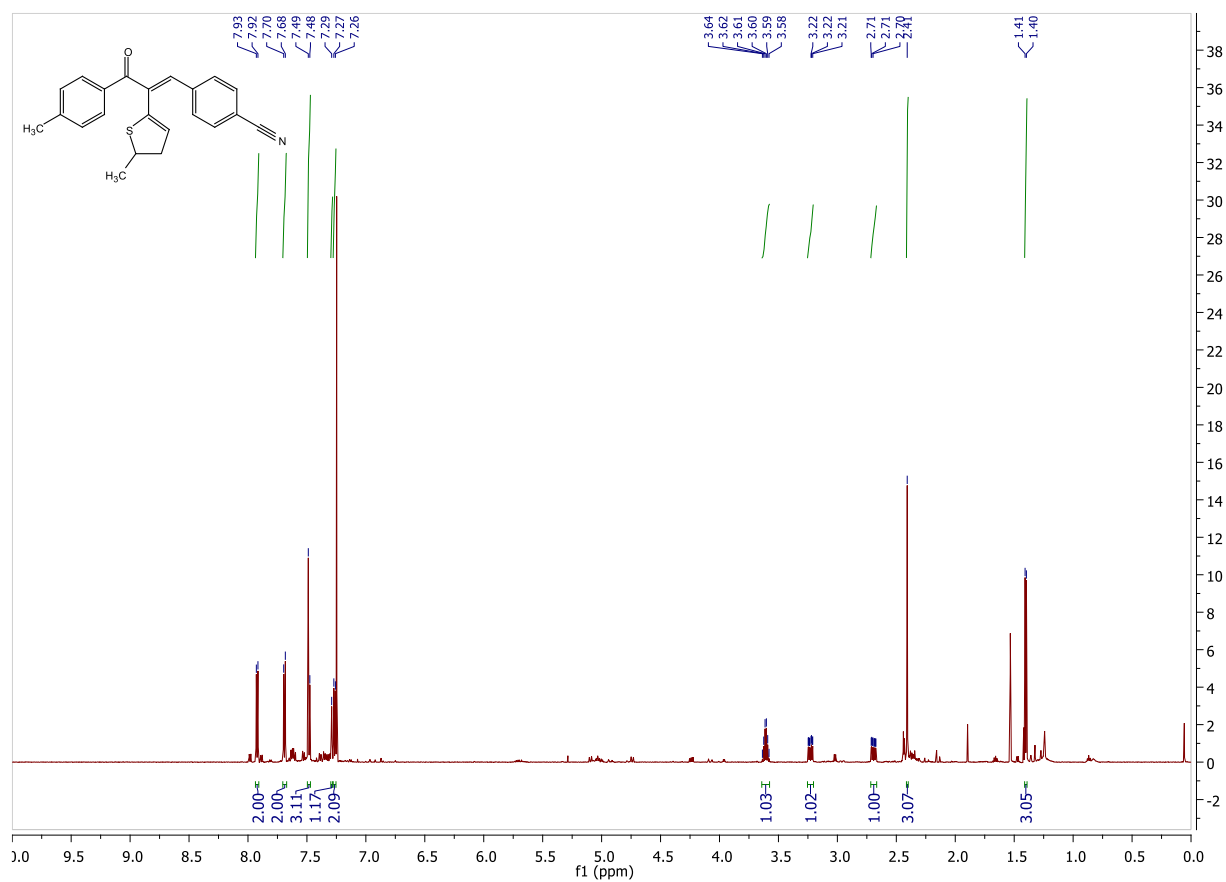
From *p*-cyanobenzaldehyde (131.1 mg, 1 mmol, 1 equiv), and 1-phenylbuta-2,3-dien-1-one (144.2 mg, 1 mmol, 1 equiv) **S6.6.3** was synthesized following the general procedure. Purification by flash chromatography (Hexane/Et<sub>2</sub>O 4:1) gave **S6.6.3** (25 mg, 8%) as a yellow oil.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 8.04 – 8.00 (m, 2H), 7.70 (d, *J* = 8.4 Hz, 2H), 7.52 – 7.51 (m, 1H), 7.51 (s, 1H), 7.51 – 7.47 (m, 4H), 7.31 (t, *J* = 1.8 Hz, 1H), 3.63 (h, *J* = 6.7 Hz, 1H), 3.25 (ddd, *J* = 14.7, 6.7, 1.8 Hz, 1H), 2.71 (ddd, *J* = 14.7, 6.7, 2.4 Hz, 1H), 1.42 (d, *J* = 6.7 Hz, 3H);

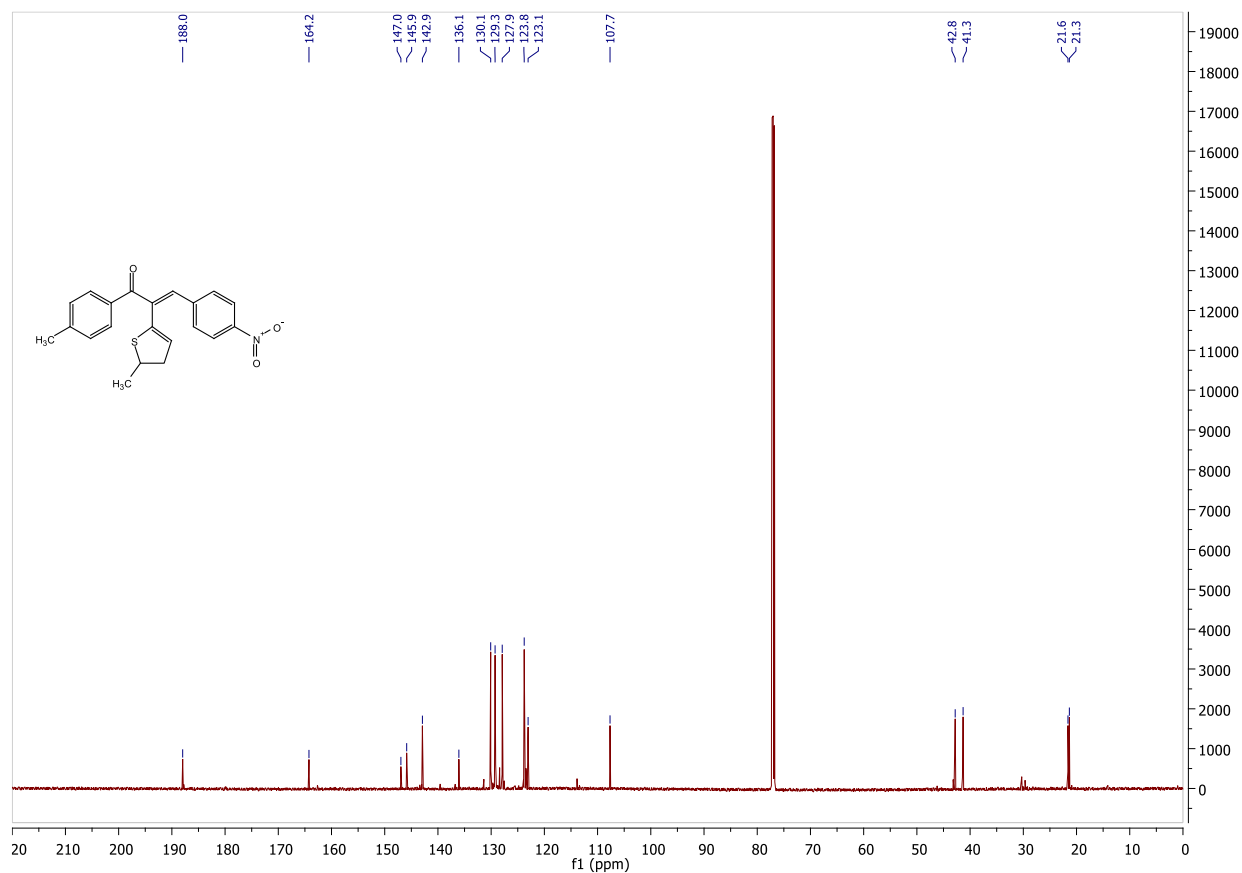
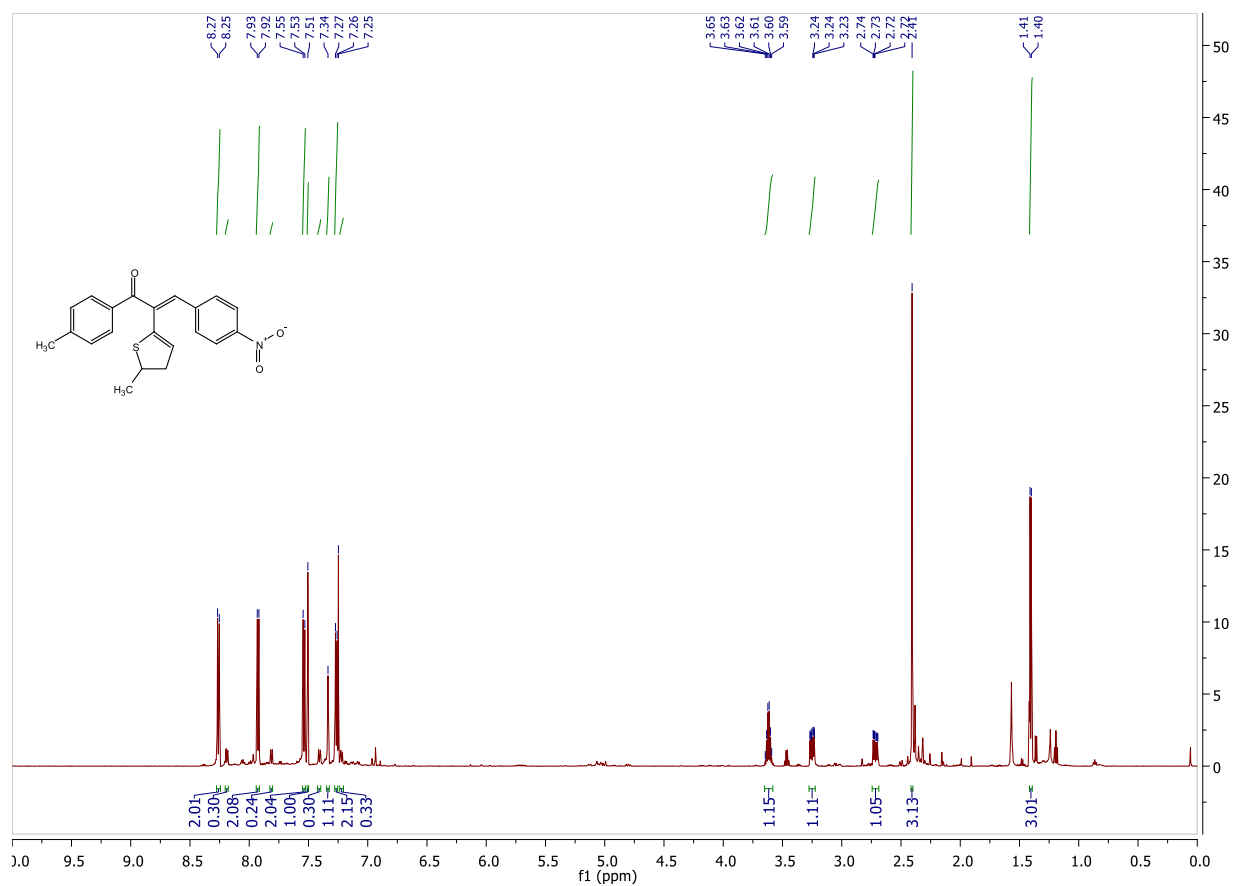
**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 188.3, 164.9, 145.2, 140.9, 138.8, 132.3, 132.2, 129.9, 128.6, 127.8, 123.6, 118.6, 111.6, 107.5, 42.9, 41.3, 21.3;

**IR** (film, DCM) 3059, 2956, 2923, 2853 (C-S), 2225 (CN), 1679, 1624, 1600, 1576, 1512 (NO<sub>2</sub>), 1340 (NO<sub>2</sub>), 1237, 1178, 1056, 970, 884, 830, 772 cm<sup>-1</sup>;

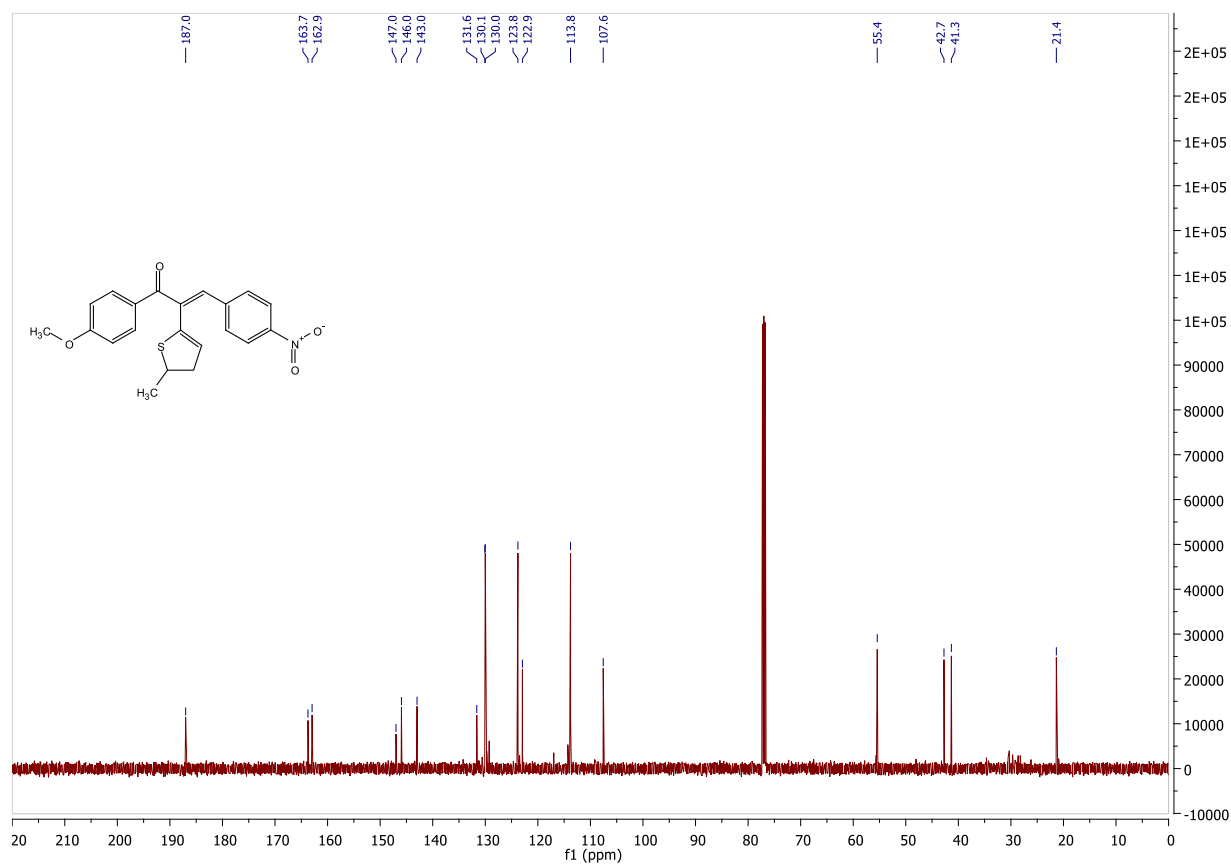
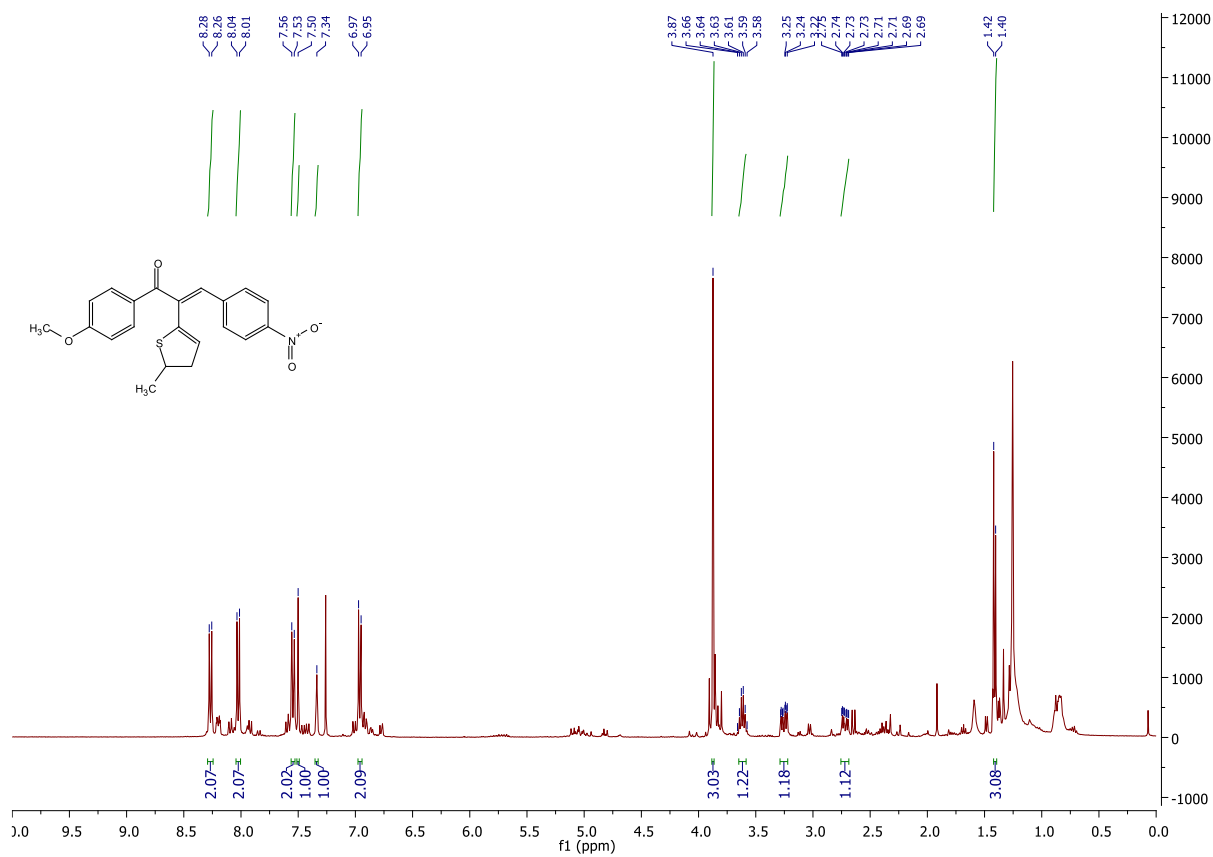
**HRMS** (ESI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>21</sub>H<sub>19</sub>NNaO<sub>4</sub>S<sup>+</sup> 332.1109; Found 332.1109.



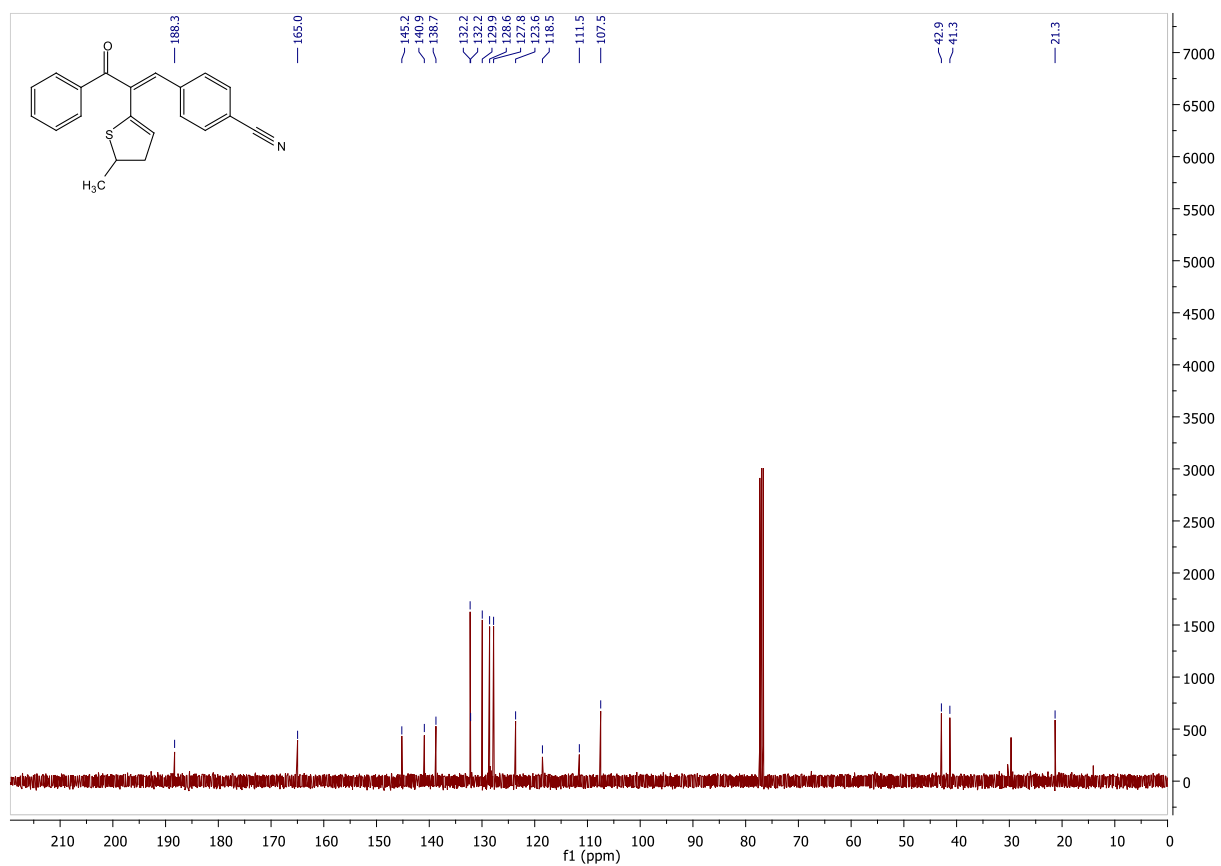
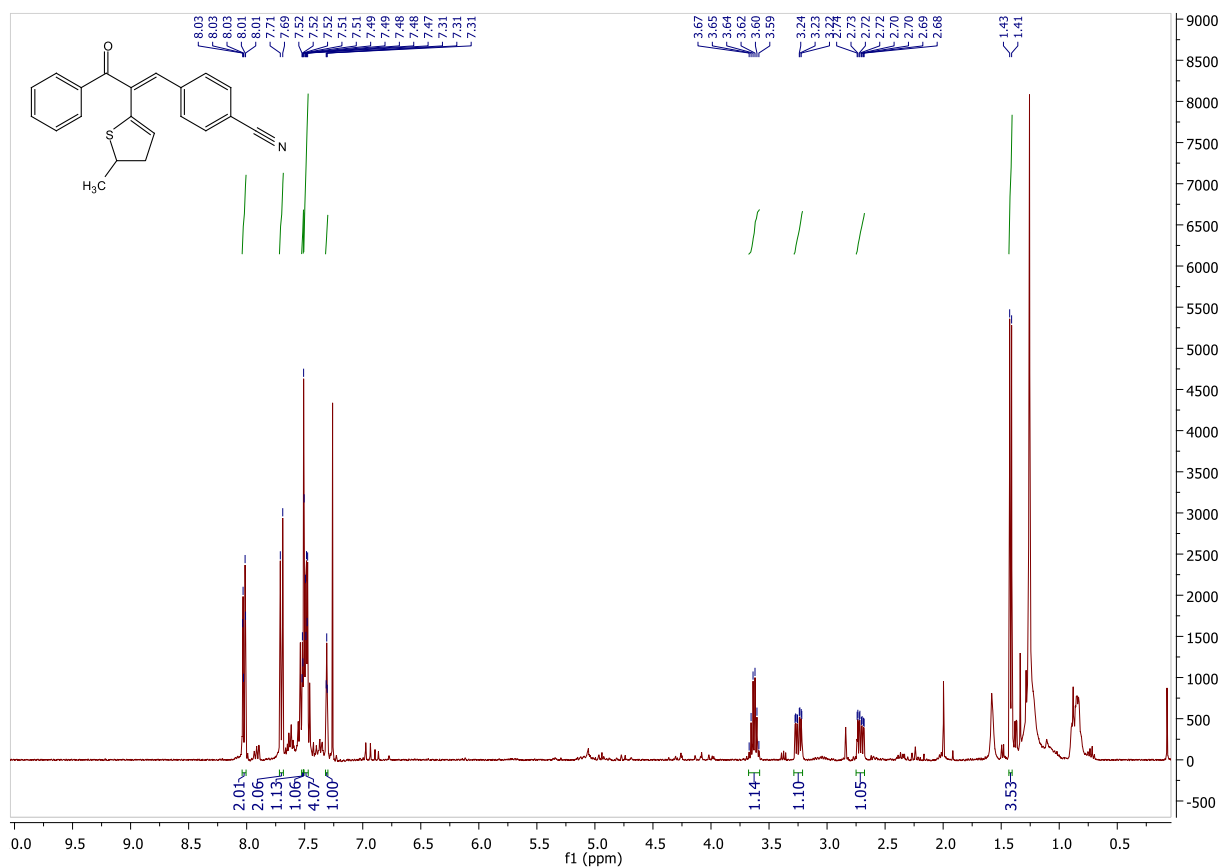
**Supplementary Figure S97.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **5b**.



**Supplementary Figure S98.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.6.1.

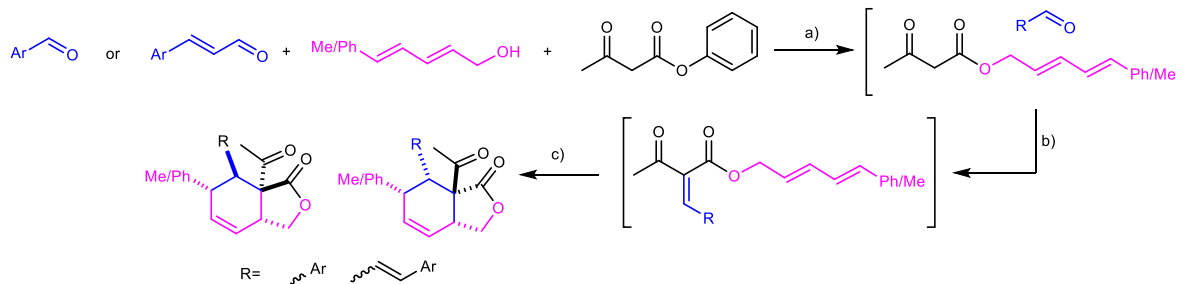


**Supplementary Figure S99.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.6.2.



**Supplementary Figure S100.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.6.3.

### Section S6.7 Reaction Mach7 described in main-text Figure 4d

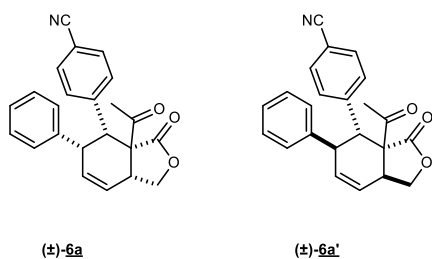


**Supplementary Figure S101. Reaction Mach7 described in main-text Figure 4d.** Reagents and conditions: (a) BHT, *o*-DCB, 100 °C, 2 h; (b) piperidinium acetate, MgSO<sub>4</sub>, rt, 2 to 5 days; (c) 210 °C, 0.75 to 2 h, 18% to 48%

### One-pot procedure

#### General procedure:

Anhydrous *o*-DCB (0.25 M), arylaldehyde (1.0 or 1.05 equiv), alcohol (1.1 equiv), butylhydroxytoluene (0.1 equiv), phenyl acetoacetate (1.0 equiv) were added in a seal tube under inert atmosphere. Reaction mixture was heated at 100 °C for 2 hours and then left to cool to rt. Sorbyl ester generation was confirmed by TLC. Piperidinium acetate<sup>57</sup> (0.2 equiv) and anhydrous MgSO<sub>4</sub> (3.9 equiv) were added and the reaction mixture was stirred vigorously at rt for a given time. Knoevenagel product generation was confirmed by TLC. The suspension was heated to 210°C for a given time. Purification depends on the structures and is described below.



4-((3a*S*,4*R*,5*S*,7a*R*)-3a-acetyl-3-oxo-5-phenyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile and 4-((3a*R*,4*S*,5*R*,7a*S*)-3a-acetyl-3-oxo-5-phenyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile **6a**

4-((3a*S*,4*R*,5*R*,7a*S*)-3a-acetyl-3-oxo-5-phenyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile and 4-((3a*R*,4*S*,5*S*,7a*R*)-3a-acetyl-3-oxo-5-phenyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile **6a'**,

According to general procedure, the reaction was performed with phenyl acetoacetate (200 mg, 1.12 mmol, 1.0 equiv), butylhydroxytoluene (24 mg, 0.109 mmol, 0.1 equiv), 4-



cyanobenzaldehyde (147 mg, 1.12 mmol, 1.0 equiv) and (2*E*,4*E*)-5-phenylpenta-2,4-dien-1-ol<sup>58</sup> (197 mg, 1.23 mmol, 1.1 equiv). Knoevenagel condensation was performed over 3 days. The suspension was heated to 210 °C for 2 h. The resulting mixture was diluted with EtOAc and water. The suspension was stirred until MgSO<sub>4</sub> was dissolved. Organic phase was separated, washed with water, dried over MgSO<sub>4</sub>, filtered and then evaporated at 75 °C, under 20 mbar. Crude product was purified by column chromatography (hexane/EtOAc 100:0 to 70:30) to afford **6a** (168 mg, 42%) as a yellowish thick oil and **6a'** (25 mg, 6%) as a yellowish thick oil.

### **6a**

<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.43 – 7.35 (m, 4H), 7.15 (br(t), *J* = 7.2 Hz, 2H), 7.10 (br(t), *J* = 7.3 Hz, 1H), 6.97 (br(d), *J* = 7.2 Hz, 2H), 6.43 (br(d), *J* = 10.1 Hz, 1H), 6.08 (dt, *J* = 10.1, 2.9 Hz, 1H), 4.49 – 4.44 (m, 1H), 4.05 (d, *J* = 5.6 Hz, 1H), 3.92 – 3.86 (m, 3H), 2.52 (s, 3H);

<sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ 202.0, 173.5, 142.1, 139.4, 131.9, 131.7, 130.8, 128.5, 128.4, 127.1, 125.4, 118.6, 111.8, 69.3, 64.3, 47.9, 42.1, 36.8, 26.4;

IR (film, CHCl<sub>3</sub>) 3028, 2921, 2227, 1761, 1711, 1184, 751 cm<sup>-1</sup>;

HRMS (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>23</sub>H<sub>19</sub>NO<sub>3</sub>Na 380.1263; Found 380.1262.

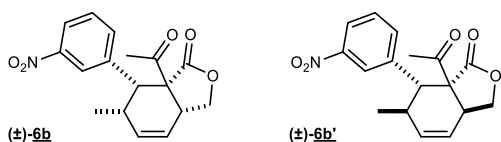
### **6a'**

<sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ 7.68 (d, *J* = 8.4 Hz, 2H), 7.56 (d, *J* = 8.4 Hz, 2H), 7.32 (t, *J* = 7.6 Hz, 2H), 7.24 (d, *J* = 7.3 Hz, 1H), 7.09 (d, *J* = 7.5 Hz, 2H), 6.44 (dt, *J* = 10.2, 2.1 Hz, 1H), 6.06 (dt, *J* = 10.2, 3.2 Hz, 1H), 4.46 (dd, *J* = 11.3, 7.6 Hz, 1H), 4.35 (t, *J* = 7.6 Hz, 1H), 4.13 (s, 1H), 3.94 (dd, *J* = 5.6, 3.0 Hz, 1H), 3.09 – 3.02 (m, 1H), 1.54 (s, 3H);

<sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ 199.5, 171.0, 145.6, 142.2, 132.5, 130.2, 129.0, 128.6, 127.4, 127.1, 126.2, 118.6, 112.0, 70.6, 64.3, 50.5, 47.0, 37.0, 25.8;

IR (film, CHCl<sub>3</sub>) 3026, 2925, 2862, 2228, 1770, 1714 cm<sup>-1</sup>;

HRMS (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>23</sub>H<sub>19</sub>NO<sub>3</sub>Na 380.1263; Found 380.1268.



**(3aR,6S,7R,7aS)-7a-acetyl-6-methyl-7-(3-nitrophenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** and **(3aS,6R,7S,7aR)-7a-acetyl-6-methyl-7-(3-nitrophenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** **6b**

**((3aS,6R,7R,7aS)-7a-acetyl-6-methyl-7-(3-nitrophenyl)-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one** and **(3aR,6S,7S,7aR)-7a-acetyl-6-methyl-7-(3-nitrophenyl)-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one** **6b'**

According to general procedure, the reaction was performed with phenyl acetoacetate (95 mg, 0.53 mmol, 1.0 equiv), butylhydroxytoluene (12 mg, 0.054 mmol, 0.1 equiv) 3-nitrobenzaldehyde (84 mg, 0.56 mmol, 1.05 equiv) and (2*E*,4*E*)-hexa-2,4-dien-1-ol (66  $\mu$ L, 57 mg, 0.59 mmol, 1.1 equiv). Knoevenagel condensation was performed over 3 days. The suspension was heated to 210  $^{\circ}$ C for 45 min. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10 to 80:20) to afford **6b** (47 mg, 28%) as a pale yellow solid and **6b'** (27 mg, 16%) as a pale yellow solid. The product **6b'** was reprecipitated from dichloromethane/hexane mixture to afford colorless crystals for X-ray analysis.

### **6b**

**mp:** 105-112  $^{\circ}$ C;

**$^1$ H NMR** (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.44 (s, 1H), 8.13 (ddd,  $J = 8.2, 2.1, 0.9$  Hz, 1H), 7.83 (d,  $J = 7.6$  Hz, 1H), 7.48 (t,  $J = 8.0$  Hz, 1H), 5.99 (dt,  $J = 10.1, 2.0$  Hz, 1H), 5.88 (dt,  $J = 10.1, 3.2$  Hz, 1H), 4.35 (t,  $J = 9.0$  Hz, 1H), 3.80 – 3.74 (m, 2H), 3.69 (t,  $J = 9$  Hz, 1H), 2.68 – 2.61 (m, 1H), 2.50 (s, 3H), 0.97 (d,  $J = 7.5$  Hz, 3H);

**$^{13}$ C NMR** (125 MHz,  $\text{CDCl}_3$ )  $\delta$  202.2, 174.0, 148.0, 138.9, 133.8, 129.6, 129.2, 128.4, 123.5, 123.0, 69.2, 64.6, 46.2, 36.6, 30.7, 26.2, 17.8;

**IR** (film,  $\text{CHCl}_3$ ) 3025, 2971, 2928, 2878, 1760, 1711, 1528, 1350  $\text{cm}^{-1}$ ;

**HRMS** (ESI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{17}\text{H}_{17}\text{NO}_5\text{Na}$  338.1004; Found 338.1006.

## **6b'**

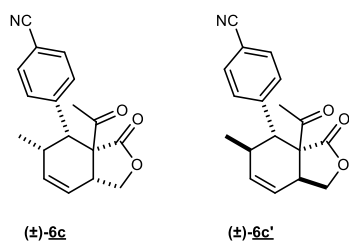
**mp:** 175-178 °C;

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 8.14 (dd, *J* = 7.1, 1.0 Hz, 2H), 7.74 (d, *J* = 7.8 Hz, 1H), 7.51 (td, *J* = 7.8, 1.6 Hz, 1H), 6.14 (dt, *J* = 10.1, 2.2 Hz, 1H), 5.72 (dt, *J* = 10.1, 3.1 Hz, 1H), 4.39 – 4.31 (m, 2H), 3.83 (s, 1H), 3.07 – 3.00 (m, 1H), 2.77 – 2.71 (m, *J* = 7.5, 2.8 Hz, 1H), 2.46 (s, 3H), 1.15 (d, *J* = 7.6 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 199.5, 171.2, 148.4, 142.4, 136.0, 130.8, 129.1, 123.8, 122.8, 122.7, 70.4, 64.5, 47.7, 37.7, 36.8, 26.8, 22.1;

**IR** (film, CHCl<sub>3</sub>) 3031, 2976, 2931, 2876, 1779, 1766, 1714, 1529, 1349 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>17</sub>H<sub>17</sub>NO<sub>5</sub>Na 338.1004; Found 338.1010.



**4-((3aS,4R,5S,7aR)-3a-acetyl-3-oxo-5-methyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile and 4-((3aR,4S,5R,7aS)-3a-acetyl-3-oxo-5-methyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile 6c**

**4-((3aS,4R,5R,7aS)-3a-acetyl-5-methyl-3-oxo-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile and 4-((3aR,4S,5S,7aR)-3a-acetyl-3-oxo-5-methyl-1,3,3a,4,5,7a-hexahydroisobenzofuran-4-yl)benzonitrile 6c',**

According to general procedure, the reaction was performed with phenyl acetoacetate (200 mg, 1.12 mmol, 1.0 equiv), butylhydroxytoluene (24 mg, 0.109 mmol, 0.1 equiv), 4-cyanobenzaldehyde (147 mg, 1.12 mmol, 1.0 equiv) and (2*E*,4*E*)-hexa-2,4-dien-1-ol (132 μL, 115 mg, 1.17 mmol, 1.1 equiv). Knoevenagel condensation was performed over 3 days. The suspension was heated to 210 °C for 2 h. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 100:10 to 80:20 to afford 6c (99 mg, 30%) as a yellowish thick oil and 6c' with a minor non separable isomer (50 mg, 15%) as a yellowish thick oil.

### 6c

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.65 (d, *J* = 8.5 Hz, 2H), 7.57 (d, *J* = 8.5 Hz, 2H), 5.95 (dt, *J* = 10.2, 2.1 Hz, 1H), 5.81 (dt, *J* = 10.2, 3.2 Hz, 1H), 4.31 (t, *J* = 8.8 Hz, 1H), 3.75 – 3.69 (m, 2H), 3.65 (t, *J* = 8.8 Hz, 1H), 2.64 – 2.58 (m, 1H), 2.47 (s, 3H), 0.96 (d, *J* = 7.5 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 202.3, 173.9, 142.3, 134.1, 132.1, 131.8, 123.3, 118.6, 112.0, 69.2, 64.6, 46.6, 36.9, 30.9, 26.2, 17.6;

**IR** (film, CHCl<sub>3</sub>) 3023, 2971, 2932, 2879, 2227, 1761, 1711 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>17</sub>NO<sub>3</sub>Na 318.1106; Found 318.1104.

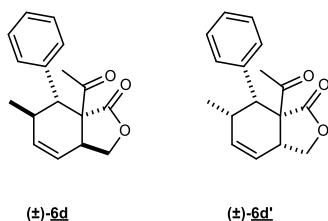
### 6c'

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.61 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.2 Hz, 2H), 6.10 (dt, *J* = 10.0, 2.2 Hz, 1H), 5.68 (dt, *J* = 10.0, 3.2 Hz, 1H), 4.37 – 4.29 (m, 2H), 3.76 (s, 1H), 3.05 – 2.97 (m, 1H), 2.74 – 2.66 (m, 1H), 2.44 (s, 3H), 1.13 (d, *J* = 7.7 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 199.7, 171.4, 145.9, 132.2, 130.9, 130.3, 122.9, 118.7, 111.8, 70.5, 64.6, 48.3, 37.7, 37.0, 27.0, 22.3;

**IR** (film, CHCl<sub>3</sub>) 3029, 2978, 2930, 2878, 2227, 1779, 1766, 1714 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>17</sub>NO<sub>3</sub>Na 318.1106; Found 318.1105.



**(3a*S*,6*R*,7*R*,7a*S*)-7a-acetyl-6-methyl-7-phenyl-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one and (3a*R*,6*S*,7*S*,7a*R*)-7a-acetyl-6-methyl-7-phenyl-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one 6d,**

**(3a*R*,6*S*,7*R*,7a*S*)-7a-acetyl-6-methyl-7-phenyl-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one and (3a*S*,6*R*,7*S*,7a*R*)-7a-acetyl-6-methyl-7-phenyl-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one 6d'.**

According to general procedure, the reaction was performed with phenyl acetoacetate (200 mg, 1.12 mmol, 1.0 equiv), butylhydroxytoluene (24 mg, 0.109 mmol, 0.1 equiv), benzaldehyde (114 μL, 119 mg, 1.12 mmol, 1.0 equiv) and (2*E*,4*E*)-hexa-2,4-dien-1-ol (132 μL, 115 mg, 1.17

mmol, 1.05 equiv). Knoevenagel condensation was performed over 5 days. The suspension was heated to 210 °C for 2h. The resulting mixture was diluted with EtOAc and water. The suspension was stirred until MgSO<sub>4</sub> was dissolved. Organic phase was separated, washed with water, dried over MgSO<sub>4</sub>, filtered and then evaporated at 75 °C, under 20 mbar. Crude product was purified by column chromatography (hexane/EtOAc 90:10) to afford **6d** (30 mg, 10%) as a pale yellow solid and **6d'** (25 mg, 8%) as a yellowish thick oil.

### **6d**

**mp:** 103-105°C;

**<sup>1</sup>H NMR** (500 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.35 (d, *J* = 7.4 Hz, 2H), 7.19 – 7.14 (m, 2H), 7.07 (t, *J* = 7.4 Hz, 1H), 5.64 (br(d), *J* = 10.0 Hz, 1H), 5.28 (dt, *J* = 10.0, 3.3 Hz, 1H), 4.08 (dd, *J* = 11.4, 7.4 Hz, 1H), 3.54 (s, 1H), 3.51 (t, *J* = 7.5 Hz, 1H), 2.94 – 2.87 (m, 1H), 2.42 – 2.35 (m, 1H), 2.10 (s, 3H), 0.71 (d, *J* = 7.7 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 200.6, 171.5, 140.6, 131.1, 129.3, 128.3, 127.5, 122.5, 70.2, 65.0, 48.2, 37.8, 36.7, 26.9, 22.2;

**IR** (film, CHCl<sub>3</sub>) 3030, 2976, 2926, 2876, 1781, 1713 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>17</sub>H<sub>18</sub>O<sub>3</sub>Na 293.1154; Found 293.1159.

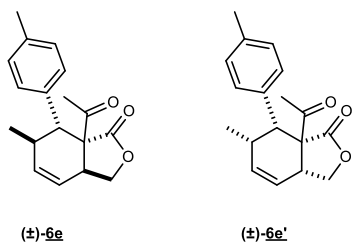
### **6d'**

**<sup>1</sup>H NMR** (600 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.45 (d, *J* = 7.5 Hz, 2H), 7.04 (t, *J* = 7.4 Hz, 2H), 7.01 – 6.98 (m, 1H), 5.53 (dt, *J* = 10.1, 2.1 Hz, 1H), 5.18 (dt, *J* = 10.1, 3.3 Hz, 1H), 3.53 (d, *J* = 5.5 Hz, 1H), 3.46 (dd, *J* = 9.1, 8.3 Hz, 1H), 3.35 – 3.29 (m, 1H), 3.23 (dd, *J* = 9.1, 8.3 Hz, 1H), 2.31 (ttt, *J* = 10.1, 5.2, 2.6 Hz, 1H), 2.19 (s, 3H), 0.76 (d, *J* = 7.5 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 203.2, 174.2, 136.1, 134.2, 131.2, 128.3, 128.0, 123.0, 68.9, 65.0, 46.7, 36.4, 30.8, 26.0, 17.8;

**IR** (film, CHCl<sub>3</sub>) 3027, 2971, 2927, 2878, 1761, 1709 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>17</sub>H<sub>18</sub>O<sub>3</sub>Na 293.1154; Found 293.1153.



**(3aS,6R,7R,7aS)-7a-acetyl-6-methyl-7-(*p*-tolyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one** and **(3aR,6S,7S,7aR)-7a-acetyl-6-methyl-7-(*p*-tolyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one** **6e**,

**(3aR,6S,7R,7aS)-7a-acetyl-6-methyl-7-(*p*-tolyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one** and **(3aS,6R,7S,7aR)-7a-acetyl-6-methyl-7-(*p*-tolyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6*H*)-one** **6e'**.

According to general procedure, the reaction was performed with phenyl acetoacetate (95 mg, 0.53 mmol, 1.0 equiv), butylhydroxytoluene (12 mg, 0.054 mmol, 0.1 equiv), 4-methylbenzaldehyde (66  $\mu$ L, 67 mg, 0.56 mmol, 1.05 equiv) and (2*E*,4*E*)-hexa-2,4-dien-1-ol (66  $\mu$ L, 57 mg, 0.59 mmol, 1.1 equiv). Knoevenagel condensation was performed over 2 days. The suspension was heated to 210 °C for 45 min. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10 to 80:20) to afford **6e** (29 mg, 19%) as a yellowish thick oil and **6e'** (22 mg, 15%) as a yellowish thick oil.

### **6e**

**<sup>1</sup>H NMR** (600 MHz, CDCl<sub>3</sub>)  $\delta$  7.21 (d, *J* = 8.0 Hz, 2H), 7.12 (d, *J* = 8.0 Hz, 2H), 6.06 (dt, *J* = 10.1, 2.3 Hz, 1H), 5.65 (dt, *J* = 10.1, 3.3 Hz, 1H), 4.30 (ddd, *J* = 20.6, 13.3, 7.6 Hz, 2H), 3.69 (s, 1H), 3.16 (ddt, *J* = 11.1, 5.4, 3.0 Hz, 1H), 2.75 – 2.69 (m, 1H), 2.44 (s, 3H), 2.32 (s, 3H), 1.10 (d, *J* = 7.6 Hz, 3H);

**<sup>13</sup>C NMR** (150 MHz, CDCl<sub>3</sub>)  $\delta$  200.8, 171.7, 137.7, 137.3, 131.2, 129.3, 129.1, 122.6, 70.3, 65.3, 48.0, 38.1, 36.8, 27.0, 22.4, 21.1;

**IR** (film, CHCl<sub>3</sub>) 3027, 2974, 2925, 2876, 1780, 1712 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>20</sub>O<sub>3</sub>Na 307.1310; Found 307.1312.

### **6e'**

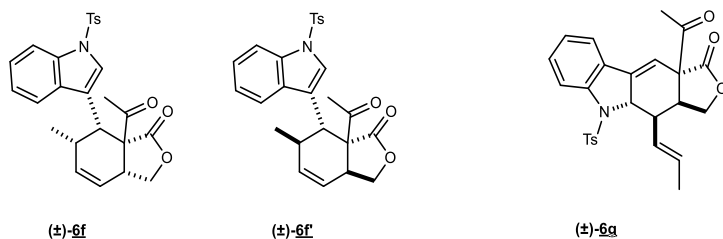
**<sup>1</sup>H NMR** (600 MHz, CDCl<sub>3</sub>)  $\delta$  7.37 (d, *J* = 8.2 Hz, 2H), 7.06 (d, *J* = 8.2 Hz, 2H), 5.92 (dt, *J* = 10.1, 1.9 Hz, 1H), 5.78 (dt, *J* = 10.1, 3.3 Hz, 1H), 4.21 (dd, *J* = 9.4, 8.2 Hz, 1H), 3.74 (tdd, *J* =

9.4, 6.2, 2.7 Hz, 1H), 3.65 (dd,  $J = 9.9, 8.2$  Hz, 1H), 3.62 (d,  $J = 5.4$  Hz, 1H), 2.54 (s, 3H), 2.53 – 2.47 (m, 1H), 2.30 (s, 3H), 0.99 (d,  $J = 7.4$  Hz, 3H);

$^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  203.3, 174.2, 137.6, 134.1, 132.8, 131.0, 129.0, 122.9, 68.9, 65.1, 46.4, 36.2, 30.7, 26.0, 21.2, 17.9;

IR (film,  $\text{CHCl}_3$ ) 3023, 2972, 2924, 2877, 1760, 1709, 1186  $\text{cm}^{-1}$ ;

HRMS (ESI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{18}\text{H}_{20}\text{O}_3\text{Na}$  307.1310; Found 307.1311.



**(3aR,6S,7R,7aS)-7a-acetyl-6-methyl-7-(1-tosyl-1H-indol-3-yl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one and (3aS,6R,7S,7aR)-7a-acetyl-6-methyl-7-(1-tosyl-1H-indol-3-yl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one 6f.**

**(3aS,6R,7R,7aS)-7a-acetyl-6-methyl-7-(1-tosyl-1H-indol-3-yl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one and (3aR,6S,7S,7aR)-7a-acetyl-6-methyl-7-(1-tosyl-1H-indol-3-yl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one 6f',**

**(3aS,4R,4aS,10aS)-10a-acetyl-4-((E)-prop-1-en-1-yl)-5-tosyl-3,3a,4,4a,5,10a-hexahydro-1H-furo[3,4-b]carbazol-1-one and (3aR,4S,4aR,10aR)-10a-acetyl-4-((E)-prop-1-en-1-yl)-5-tosyl-3,3a,4,4a,5,10a-hexahydro-1H-furo[3,4-b]carbazol-1-one 6g.**

According to general procedure, the reaction was performed with phenyl acetoacetate (200 mg, 1.12 mmol, 1.0 equiv), butylhydroxytoluene (24 mg, 0.109 mmol, 0.1 equiv), 1-tosyl-1H-indole-3-carbaldehyde<sup>55</sup> (335 mg, 1.12 mmol, 1.0 equiv) and (2E,4E)-hexa-2,4-dien-1-ol (132  $\mu\text{L}$ , 1.17 mmol, 1.05 equiv). Knoevenagel condensation was performed over 3 days. The suspension was heated to 210 °C for 2 h. The resulting mixture was diluted with EtOAc and water. The suspension was stirred until  $\text{MgSO}_4$  was dissolved. Organic phase was separated, washed with water, dried over  $\text{MgSO}_4$ , filtered and then evaporated at 75 °C, under 20 mbar. Crude product was purified by column chromatography (hexane/EtOAc 100:0 to 70:30) to afford 6f (83 mg, 16%) as a yellowish thick oil, 6f' with a minor non separable isomer (49 mg, 9%) as a yellowish thick oil, and the product of inverse demand Diels-Alder 6g (125 mg, 24%) as a brownish thick oil.

**6f**

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 8.00 (d, *J* = 7.9 Hz, 1H), 7.88 (s, 1H), 7.70 (d, *J* = 8.4 Hz, 2H), 7.61 (d, *J* = 7.5 Hz, 1H), 7.37 – 7.27 (m, 2H), 7.22 (d, *J* = 8.1 Hz, 2H), 5.97 (dt, *J* = 10.1, 2.5 Hz, 1H), 5.84 (dt, *J* = 10.1, 2.9 Hz, 1H), 4.31 (t, *J* = 8.8 Hz, 1H), 4.01 (d, *J* = 5.2 Hz, 1H), 3.87 – 3.80 (m, 1H), 3.74 – 3.66 (m, 1H), 2.61 – 2.51 (m, 1H), 2.40 (s, 3H), 2.35 (s, 3H), 0.89 (d, *J* = 7.4 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 202.8, 174.0, 145.0, 135.2, 134.6, 134.3, 131.5, 129.9, 126.6, 125.8, 125.2, 123.9, 123.4, 119.4, 117.9, 113.7, 69.4, 63.7, 36.9, 31.1 (2C), 26.1, 21.5, 17.1;

**IR** (film, CHCl<sub>3</sub>) 3024, 2968, 2926, 2877, 1776, 1701, 1448, 1368, 1173cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>26</sub>H<sub>25</sub>O<sub>5</sub>NaS 486.1351; Found 486.1353.

**6f'**

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.94 (d, *J* = 8.4 Hz, 1H), 7.83 – 7.79 (m, 1H), 7.73 (d, *J* = 8.4 Hz, 2H), 7.47 (s, 1H), 7.33 – 7.27 (m, 2H), 7.22 (d, *J* = 8.1 Hz, 2H), 6.13 (dt, *J* = 10.0, 2.3 Hz, 1H), 5.65 (dt, *J* = 10.0, 3.2 Hz, 1H), 4.36 – 4.26 (m, 2H), 3.90 (s, 1H), 3.29 – 3.21 (m, 1H), 2.80 – 2.70 (m, 1H), 2.46 (s, 3H), 2.34 (s, 3H), 1.14 (d, *J* = 7.7 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 200.3, 171.3, 145.2, 135.2, 134.9, 131.2, 130.5, 130.0, 127.0, 125.6, 125.1, 123.4, 123.0, 120.6, 113.7, 70.2, 65.1, 40.4, 38.8, 38.0, 27.2, 22.6, 21.7;

**IR** (film, CHCl<sub>3</sub>) 3031, 2966, 2925, 2876, 1780, 1715, 1447, 1366, 1174cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>26</sub>H<sub>25</sub>O<sub>5</sub>NaS 486.1351; Found 486.1355.

**6g**

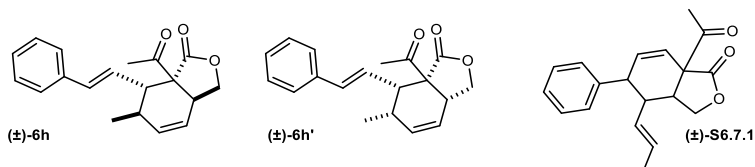
**<sup>1</sup>H NMR** (600 MHz, CDCl<sub>3</sub>) δ 7.77 (d, *J* = 8.3 Hz, 1H), 7.63 (d, *J* = 8.3 Hz, 2H), 7.39 (ddd, *J* = 7.7, 1.2, 0.6 Hz, 1H), 7.32 (ddd, *J* = 8.5, 7.5, 1.3 Hz, 1H), 7.23 (dd, *J* = 8.6, 0.6 Hz, 2H), 7.07 (td, *J* = 7.6, 0.9 Hz, 1H), 6.27 (d, *J* = 3.4 Hz, 1H), 5.69 (dq, *J* = 15.2, 6.4, 0.7 Hz, 1H), 5.03 (ddd, *J* = 15.4, 8.1, 1.7 Hz, 1H), 4.71 (t, *J* = 9.7 Hz, 1H), 4.30 (dd, *J* = 4.7, 3.4 Hz, 1H), 4.18 (dd, *J* = 9.7, 7.2 Hz, 1H), 3.54 – 3.49 (m, 2H), 2.43 (s, 3H), 2.36 (s, 3H), 1.55 (dd, *J* = 6.4, 1.5 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 201.2, 173.2, 145.12, 145.07, 137.9, 133.1, 131.1, 130.9, 130.1, 127.5, 127.3, 125.5, 124.5, 121.6, 116.0, 110.7, 69.9, 62.4, 60.7, 42.5, 39.9, 26.6, 21.7, 18.2;

**IR** (film, CHCl<sub>3</sub>) 3022, 2921, 2855, 1763, 1716, 1461, 1357, 1169cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>26</sub>H<sub>25</sub>O<sub>5</sub>NaS 486.1351; Found 486.1355.





**(3aS,6R,7S,7aS)-7a-acetyl-6-methyl-7-((E)-styryl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** and **(3aR,6S,7R,7aR)-7a-acetyl-6-methyl-7-((E)-styryl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** **6h**

**(3aR,6S,7S,7aS)-7a-acetyl-6-methyl-7-((E)-styryl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** and **(3aS,6R,7R,7aR)-7a-acetyl-6-methyl-7-((E)-styryl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** **6h'**

**7a-acetyl-5-phenyl-4-(prop-1-en-1-yl)-3,3a,4,5-tetrahydroisobenzofuran-1(7aH)-one**  
**S6.7.1**

According to general procedure, the reaction was performed with phenyl acetoacetate (95 mg, 0.53 mmol, 1.0 equiv), butylhydroxytoluene (12 mg, 0.054 mmol, 0.1 equiv), cinnamaldehyde (70  $\mu$ L, 73 mg, 0.56 mmol, 1.05 equiv) and (2E,4E)-hexa-2,4-dien-1-ol (66  $\mu$ L, 0.59 mmol, 1.1 equiv). Knoevenagel condensation was performed over 3 days. The suspension was heated to 210 °C for 45 min. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10 to 80:20) to afford **6h** (29 mg, 19%) as a brown oil, **6h'** (19 mg, 12%) as a yellow thick oil and the and the product of inverse demand Diels-Alder **S6.7.1** (15 mg, 9%) as a brown oil.

### **6h**

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$  7.38 – 7.35 (m,  $J$  = 5.4, 3.3 Hz, 2H), 7.32 – 7.28 (m,  $J$  = 8.2, 6.6 Hz, 2H), 7.26 – 7.21 (m,  $J$  = 7.1 Hz, 1H), 6.53 (d,  $J$  = 15.9 Hz, 1H), 6.27 (dd,  $J$  = 16.0, 7.2 Hz, 1H), 6.00 (dt,  $J$  = 10.0, 2.4 Hz, 1H), 5.53 (dt,  $J$  = 10.0, 3.2 Hz, 1H), 4.40 (s, 1H), 4.38 (d,  $J$  = 1.3 Hz, 1H), 3.27 (d,  $J$  = 7.3 Hz, 1H), 3.26 – 3.18 (m, 1H), 2.65 – 2.57 (m, 1H), 2.36 (s, 3H), 1.09 (d,  $J$  = 7.7 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>)  $\delta$  200.8, 172.1, 136.7, 133.3, 130.8, 128.7, 128.1, 127.9, 126.6, 122.3, 70.7, 64.9, 45.3, 37.7, 35.9, 27.1, 21.8;

**IR** (film, CHCl<sub>3</sub>) 3028, 2967, 2925, 2874, 1778, 1714 cm<sup>-1</sup>;

**HRMS** (ESI)  $m/z$ : [M+Na]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>3</sub>Na 319.1310; Found 319.1315.

## **6h'**

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.34 – 7.28 (m, 4H), 7.26 – 7.22 (m, 1H), 6.67 (d, *J* = 15.9 Hz, 1H), 6.09 (dd, *J* = 15.9, 9.8 Hz, 1H), 5.73 (ddd, *J* = 10.1, 3.4, 2.8 Hz, 1H), 5.64 (dtd, *J* = 10.1, 2.0, 0.8 Hz, 1H), 4.43 (dd, *J* = 9.1, 8.6 Hz, 1H), 3.84 (dd, *J* = 9.7, 8.5 Hz, 1H), 3.78 – 3.70 (m, 1H), 3.24 (dd, *J* = 9.8, 4.5 Hz, 1H), 2.53 (s, 3H), 2.32 – 2.25 (m, 1H), 1.04 (d, *J* = 7.4 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 202.9, 174.4, 136.8, 136.5, 132.6, 128.7, 128.0, 126.6, 123.8, 122.9, 69.6, 64.5, 46.4, 35.6, 30.7, 26.2, 18.4;

**IR** (film, CHCl<sub>3</sub>) 3026, 2965, 2925, 2874, 1763, 1710 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>3</sub>Na 319.1310; Found 319.1311.

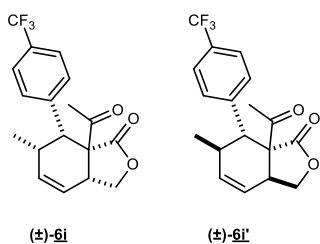
### **S6.7.1**

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.31 – 7.27 (m, 2H), 7.26 – 7.22 (m, 1H), 7.00 – 6.96 (m, 2H), 6.31 (dd, *J* = 10.0, 0.9 Hz, 1H), 6.18 (dd, *J* = 10.0, 6.1 Hz, 1H), 5.58 (dq, *J* = 15.0, 6.4 Hz, 1H), 4.76 (ddq, *J* = 15.0, 9.4, 1.6 Hz, 1H), 4.19 – 4.12 (m, 2H), 3.46 (t, *J* = 5.7 Hz, 1H), 2.94 (ddd, *J* = 12.9, 4.7, 2.4 Hz, 1H), 2.51 (s, 3H), 2.51 – 2.43 (m, 1H), 1.64 (dd, *J* = 6.4, 1.7 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 200.6, 173.7, 137.2, 134.1, 131.0, 129.9, 129.7, 128.3, 127.2, 122.3, 69.7, 64.3, 45.3, 43.5, 36.2, 26.6, 18.0;

**IR** (film, CHCl<sub>3</sub>) 3028, 2973, 2916, 2886, 1767, 1714 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>3</sub>Na 319.1310; Found 319.1315.



**(3aR,6S,7R,7aS)-7a-acetyl-6-methyl-7-(4-(trifluoromethyl)phenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** and **(3aS,6R,7S,7aR)-7a-acetyl-6-methyl-7-(4-(trifluoromethyl)phenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** **6i**

**(3aS,6R,7R,7aS)-7a-acetyl-6-methyl-7-(4-(trifluoromethyl)phenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** and **(3aR,6S,7S,7aR)-7a-acetyl-6-methyl-7-(4-(trifluoromethyl)phenyl)-3,3a,7,7a-tetrahydroisobenzofuran-1(6H)-one** **6i'**

According to general procedure, the reaction was performed with phenyl acetoacetate (200 mg, 1.12 mmol, 1.0 equiv), butylhydroxytoluene (24 mg, 0.109 mmol, 0.1 equiv), 4-trifluoromethylbenzaldehyde (153  $\mu$ L, 195 mg, 1.12 mmol, 1.0 equiv) and (2*E*,4*E*)-hexa-2,4-dien-1-ol (132  $\mu$ L, 115 mg, 1.17 mmol, 1.05 equiv). Knoevenagel condensation was performed over 2 days. The suspension was heated to 210°C for 45 min. The resulting mixture was diluted with EtOAc and water. The suspension was stirred until MgSO<sub>4</sub> was dissolved. Organic phase was separated, washed with water, dried over MgSO<sub>4</sub>, filtered and then evaporated at 75 °C, under 20 mbar. Crude product was purified by column chromatography (hexane/Et<sub>2</sub>O 95/5 to 70:30) to afford **6i** (67 mg, 18%) as a yellowish thick oil and **6i'** (53 mg, 14%) as a yellowish thick oil.

### **6i**

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.65 (d, *J* = 8.4 Hz, 2H), 7.53 (d, *J* = 8.4 Hz, 2H), 5.96 (dt, *J* = 10.1, 2.1 Hz, 1H), 5.81 (dt, *J* = 10.1, 3.2 Hz, 1H), 4.29 (t, *J* = 8.8 Hz, 1H), 3.77 – 3.71 (m, 2H), 3.65 (t, *J* = 8.8 Hz, 1H), 2.63 – 2.55 (m, 1H), 2.50 (s, 3H), 0.98 (d, *J* = 7.5 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>)  $\delta$  202.6, 174.0, 140.7, 134.1, 131.5, 130.1 (q, *J* = 32.4 Hz), 125.3 (q, *J* = 3.2 Hz), 124.1 (q, *J* = 272.1 Hz), 123.2 (s), 69.1, 64.7, 46.4, 36.7, 30.9, 26.2, 17.7;

**IR** (film, CHCl<sub>3</sub>) 3026, 2973, 2932, 2881, 1762, 1712, 1327 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>17</sub>O<sub>3</sub>F<sub>3</sub>Na 361.1027; Found 361.1024.

### **6i'**

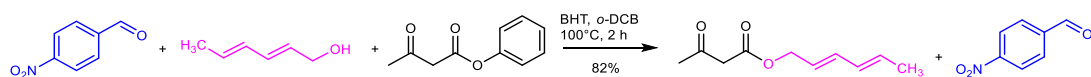
**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.57 (d, *J* = 8.2 Hz, 2H), 7.46 (d, *J* = 8.2 Hz, 2H), 6.10 (dt, *J* = 10.1, 2.1 Hz, 1H), 5.68 (dt, *J* = 10.1, 3.3 Hz, 1H), 4.33 (dq, *J* = 15.1, 7.7 Hz, 2H), 3.78 (s, 1H), 3.10 – 3.04 (m, 1H), 2.76 – 2.69 (m, 1H), 2.45 (s, 3H), 1.13 (d, *J* = 7.7 Hz, 3H);

**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>)  $\delta$  200.1, 171.5, 144.6, 131.0, 129.9 (q, *J* = 32.6 Hz), 129.9, 125.4 (q, *J* = 3.8 Hz), 124.2 (q, *J* = 272.0 Hz), 122.8, 70.5, 65.3, 48.1, 37.8, 36.9, 27.0, 22.3;

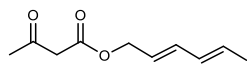
**IR** (film, CHCl<sub>3</sub>) 3033, 2974, 2929, 2880, 1780, 1715, 1326 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M+Na]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>17</sub>O<sub>3</sub>F<sub>3</sub>Na 361.1027; Found 361.1023.

### Mechanistic validation by stepwise isolation



**Supplementary Figure S102.** Reaction scheme of the stepwise process to confirm the first step intermediate



### (2E,4E)-hexa-2,4-dien-1-yl 3-oxobutanoate S6.7.2

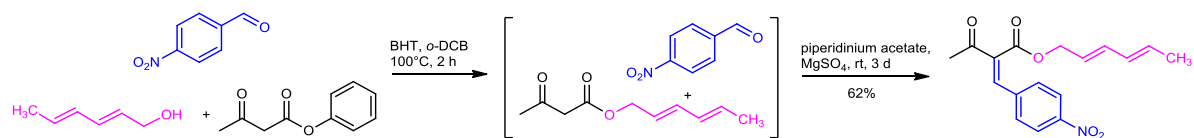
Anhydrous *o*-DCB (9.0 mL, 0.25 M), (2E,4E)-hexa-2,4-dien-1-ol (0.28 mL, 244 mg, 2.49 mmol, 1.1 equiv), butylhydroxytoluene (48 mg, 0.22 mmol, 0.1 equiv), 3-nitrobenzaldehyde (354 mg, 2.34 mmol, 1.05 equiv) and phenyl acetoacetate (400 mg, 2.24 mmol, 1.0 equiv) were added in a seal tube under inert atmosphere. Reaction mixture was heated at 100 °C for 2 h and then left to cool to rt. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10) to afford **S6.7.2** as a keto-enol mixture (ratio ketone/enol 9:1, 334 mg, 82%) as a colorless oil.

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 6.26 (dd, *J* = 15.2, 10.5 Hz, 1H), 6.08 – 6.00 (m, 1H), 5.76 (dq, *J* = 14.0, 6.8 Hz, 1H), 5.61 (dt, *J* = 14.6, 6.9 Hz, 1H), 4.63 (d, *J* = 6.7 Hz, 2H), 3.45 (s, 2H), 2.26 (s, 3H), 1.76 (dd, *J* = 6.7, 1.6 Hz, 3H);

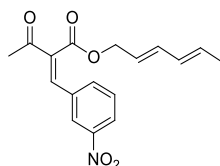
**<sup>13</sup>C NMR** (125 MHz, CDCl<sub>3</sub>) δ 200.6, 167.0, 135.8, 131.9, 130.4, 123.0, 66.0, 50.2, 30.3, 18.3;

**IR** (film, CHCl<sub>3</sub>) 3023, 2936, 2881, 1743, 1719, 1150 cm<sup>-1</sup>;

**HRMS** (ESI) *m/z*: [M-H]<sup>-</sup> Calcd for C<sub>10</sub>H<sub>13</sub>NO<sub>3</sub> 181.0865; Found 181.0867.



**Supplementary Figure S103.** Reaction scheme of the stepwise process to confirm the second step intermediate



### **(Z)-(2E,4E)-hexa-2,4-dien-1-yl 2-(3-nitrobenzylidene)-3-oxobutanoate S6.7.3**

Anhydrous *o*-DCB (9.0 mL, 0.25 M), (*2E,4E*)-hexa-2,4-dien-1-ol (0.28 mL, 244 mg, 2.49 mmol, 1.1 equiv), butylhydroxytoluene (48 mg, 0.22 mmol, 0.1 equiv), 3-nitrobenzaldehyde (354 mg, 2.34 mmol, 1.05 equiv) and phenyl acetoacetate (400 mg, 2.24 mmol, 1.0 equiv) were added in a seal tube under inert atmosphere. Reaction mixture was heated at 100 °C for 2 h and then left to cool to rt. Sorbyl ester generation was confirmed by TLC. Piperidinium acetate (64 mg, 0.44 mmol, 0.2 equiv) and anhydrous MgSO<sub>4</sub> (1.02 g, 8.47 mmol, 3.8 equiv) were added and the reaction mixture was stirred vigorously at rt for 3 days. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10 to 80:20) to afford **S6.7.3** (437 mg, 62 %) as a yellowish oil.

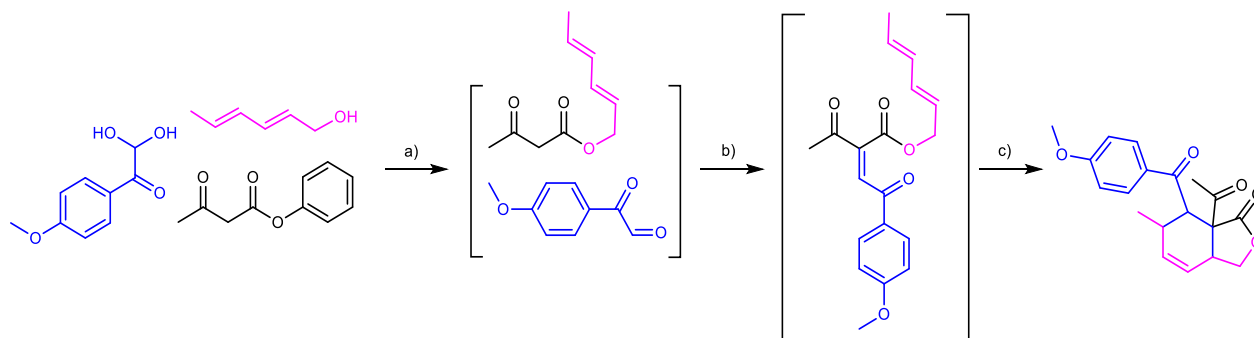
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.30 (t, *J* = 2.0 Hz, 1H), 8.22 (d, *J* = 8.2 Hz, 1H), 7.73 (d, *J* = 7.7 Hz, 1H), 7.59 (s, 1H), 7.55 (t, *J* = 8.0 Hz, 1H), 6.24 (dd, *J* = 15.2, 10.4 Hz, 1H), 6.00 (dd, *J* = 14.6, 10.9 Hz, 1H), 5.73 (dq, *J* = 13.8, 6.7 Hz, 1H), 5.58 (dt, *J* = 14.6, 7.0 Hz, 1H), 4.77 (d, *J* = 7.0 Hz, 2H), 2.43 (s, 3H), 1.75 (d, *J* = 6.7 Hz, 3H);

<sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 194.0, 166.7, 148.5, 138.3, 137.0, 136.7, 135.2, 134.7, 132.5, 130.2, 130.0, 124.9, 124.0, 121.9, 66.7, 27.0, 18.2;

IR (film, CHCl<sub>3</sub>) 3023, 2959, 2934, 2877, 1730, 1699, 1532, 1353, 1201 cm<sup>-1</sup>;

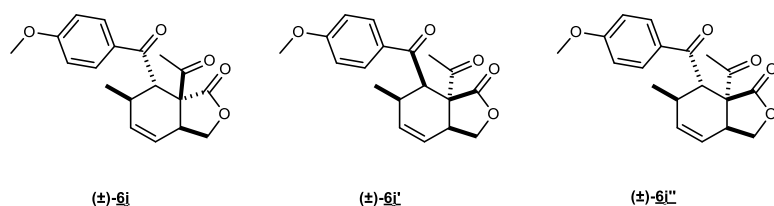
HRMS (EI) *m/z*: [M]<sup>+</sup> Calcd for C<sub>17</sub>H<sub>17</sub>NO<sub>5</sub> 315.1107; Found 315.1118.

**Reaction Mach7 described in main-text Figure 4d (variant with 2,2-dihydroxy-1-(4-methoxyphenyl)ethanone).**



**Supplementary Figure S104. Reaction Mach7 described in main-text Figure 4g-h (variant with 2,2-dihydroxy-1-(4-methoxyphenyl)ethanone).** Reagents and conditions: (a) BHT, MgSO<sub>4</sub>, *o*-DCB, 100°C, 2 h; (b) piperidinium acetate, rt, 3 days; (c) 210°C, 1.5 h, 57%.

#### One-pot, gram-scale procedure



**(3aS,6R,7S,7aR)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one and (3aR,6S,7R,7aS)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one 6i,**

**(3aS,6R,7R,7aS)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one and (3aR,6S,7S,7aR)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one 6i',**

**(3aS,6R,7S,7aS)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one and (3aR,6S,7R,7aR)-7a-acetyl-7-(4-methoxybenzoyl)-6-methyl-3a,6,7,7a-tetrahydroisobenzofuran-1(3H)-one 6i''**

Anhydrous *o*-DCB (49 mL, 0.25 M), 2,2-dihydroxy-1-(4-methoxyphenyl)ethanone<sup>59</sup> (2.14 g, 11.75 mmol, 1.05 equiv), (2*E*,4*E*)-hexa-2,4-dien-1-ol (1.38 mL, 1.20 g, 12.25 mmol, 1.1 equiv), butylhydroxytoluene (240 mg, 1.09 mmol, 0.1 equiv), and anhydrous MgSO<sub>4</sub> (5.10 g, 42.37 mmol, 3.8 equiv) were added in a seal tube under inert atmosphere. The suspension was stirred at rt for 10 min and then phenyl acetoacetate (2.0 g, 11.18 mmol, 1.0 equiv) was added. Reaction mixture was heated at 100 °C for 2 h and then left to cool to rt. Sorbyl ester generation was

confirmed by TLC. Piperidinium acetate (320 mg, 2.20 mmol, 0.2 equiv), were added and the reaction mixture was stirred vigorously at rt for 3 days. The suspension was heated to 210 °C for a 1.5 h. The resulting mixture was directly purified by column chromatography (hexane/EtOAc 90:10 to 80:20) to afford 2 fractions containing **6i'** and a mixture of **6i** and **6i''**. The fraction containing **6i** and **6i''** was repurified by chromatography (toluene/EtOAc 95/5) to afford **6i** (950 mg, 26% and **6i''** (328 mg, 9%) as a brown thick oil. The fraction containing **6i'** was repurified by chromatography (toluene/acetone 90:10) to afford **6i'** (872 mg, 24%) as a brown thick oil. **6i'** and **6i** were triturated in pentane/diethyl ether 8:2 affording respectively white and yellow solid.

### **6i**

**mp:** 128-131°C

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.96 (d, *J* = 8.9 Hz, 2H), 6.98 (d, *J* = 8.9 Hz, 2H), 6.08 (dt, *J* = 9.7, 2.9 Hz, 1H), 5.37 (dt, *J* = 9.7, 3.0 Hz, 1H), 4.59 (t, *J* = 7.8 Hz, 1H), 4.45 (dd, *J* = 11.7, 7.8 Hz, 1H), 4.08 (d, *J* = 2.8 Hz, 1H), 4.08 – 4.02 (m, 1H), 3.89 (s, 3H), 2.58 (dtt, *J* = 10.2, 5.1, 2.6 Hz, 1H), 2.34 (s, 3H), 1.23 (d, *J* = 7.5 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 200.3, 198.7, 172.5, 164.2, 131.0, 130.4, 130.0, 124.2, 114.2, 70.9, 64.9, 55.7, 47.9, 39.3, 34.8, 27.4, 22.3;

**IR** (film, CHCl<sub>3</sub>) 3032, 2969, 2933, 2842, 1770, 1713, 1669, 1599, 1254, 1173 cm<sup>-1</sup>;

**HRMS** (EI) *m/z*: [M]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>5</sub>Na 328.1311; Found 328.1306.

### **6i'**

**mp:** 151-153°C

**<sup>1</sup>H NMR** (500 MHz, CDCl<sub>3</sub>) δ 7.97 (d, *J* = 8.9 Hz, 2H), 6.99 (d, *J* = 8.9 Hz, 2H), 5.70 (ddd, *J* = 10.2, 3.4, 1.4 Hz, 1H), 5.64 (ddd, *J* = 10.2, 5.2, 1.8 Hz, 1H), 4.66 (dd, *J* = 8.4, 7.1 Hz, 1H), 4.60 (s, 1H), 4.15 (dd, *J* = 8.4, 1.4 Hz, 1H), 4.01 (ddd, *J* = 7.0, 3.4, 1.7 Hz, 1H), 3.89 (s, 3H), 2.64 (tt, *J* = 8.1, 5.9 Hz, 1H), 2.12 (s, 3H), 1.30 (d, *J* = 7.4 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 200.3, 197.9, 174.8, 164.2, 131.0, 129.3, 127.8, 126.1, 114.5, 73.61, 60.8, 55.7, 49.7, 37.1, 32.7, 25.4, 22.1;

**IR** (film, CHCl<sub>3</sub>) 3023, 2978, 2931, 2842, 1764, 1713, 1670, 1599, 1257, 1171 cm<sup>-1</sup>;

**HRMS** (EI) *m/z*: [M]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>5</sub>Na 328.1311; Found 328.1316.

**6j''**

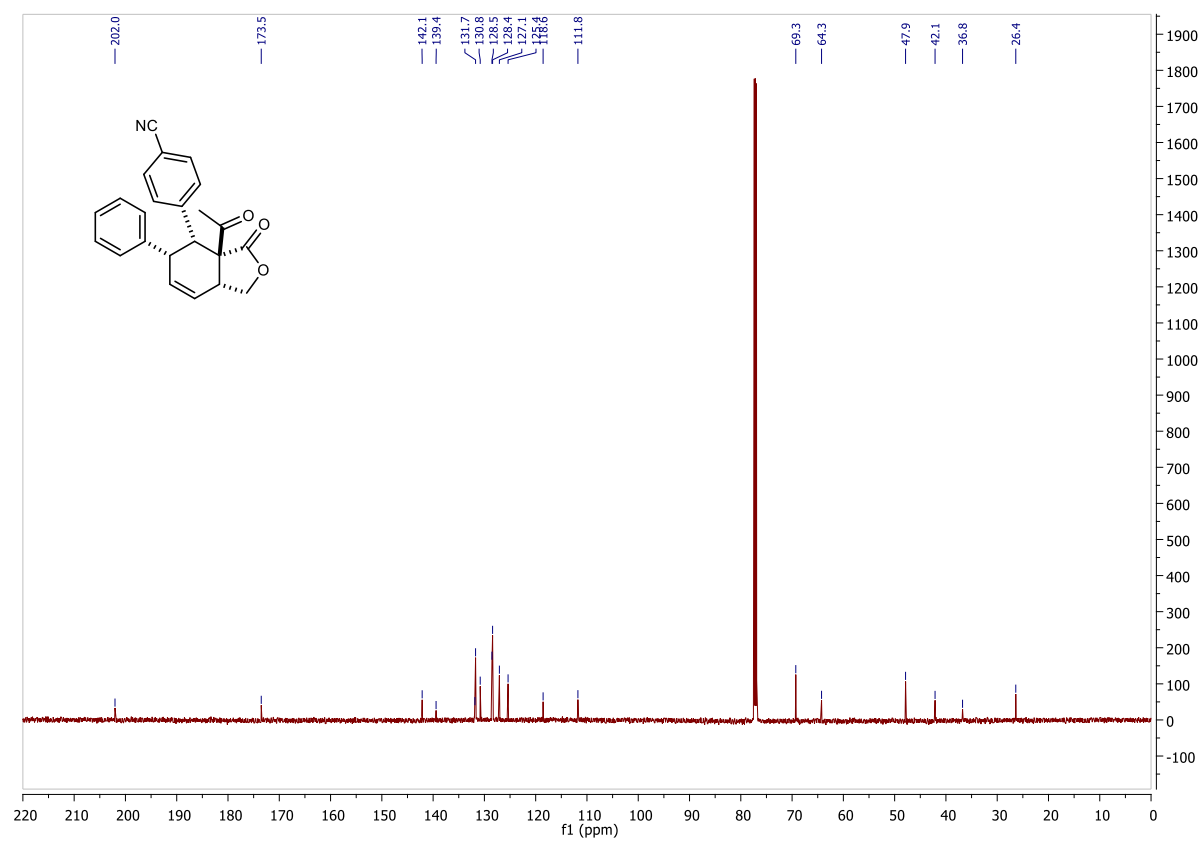
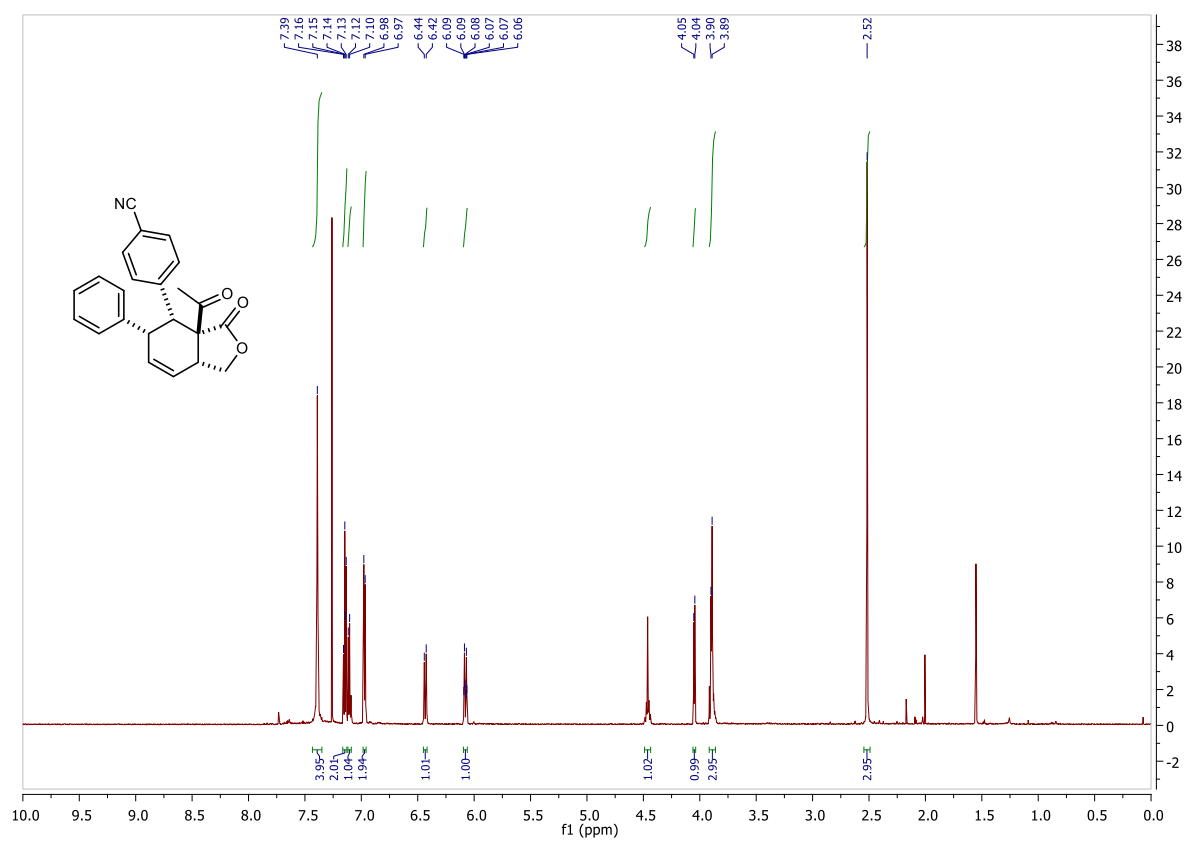
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.90 (d, *J* = 8.9 Hz, 2H), 6.92 (d, *J* = 8.9 Hz, 2H), 5.73 (dt, *J* = 10.0, 3.0 Hz, 1H), 5.51 (dt, *J* = 10.0, 2.5 Hz, 1H), 4.62 – 4.56 (m, 2H), 4.50 (t, *J* = 8.1 Hz, 1H), 3.86 (s, 3H), 3.53 (ddq, *J* = 9.5, 8.3, 2.8 Hz, 1H), 2.72 (dddd, *J* = 10.2, 5.3, 3.9, 1.7 Hz, 1H), 2.46 (s, 3H), 0.91 (d, *J* = 7.6 Hz, 3H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 202.7, 199.5, 175.3, 163.9, 132.1, 130.9, 130.3, 124.1, 114.0, 70.7, 61.8, 55.6, 46.3, 36.8, 30.6, 26.4, 19.0;

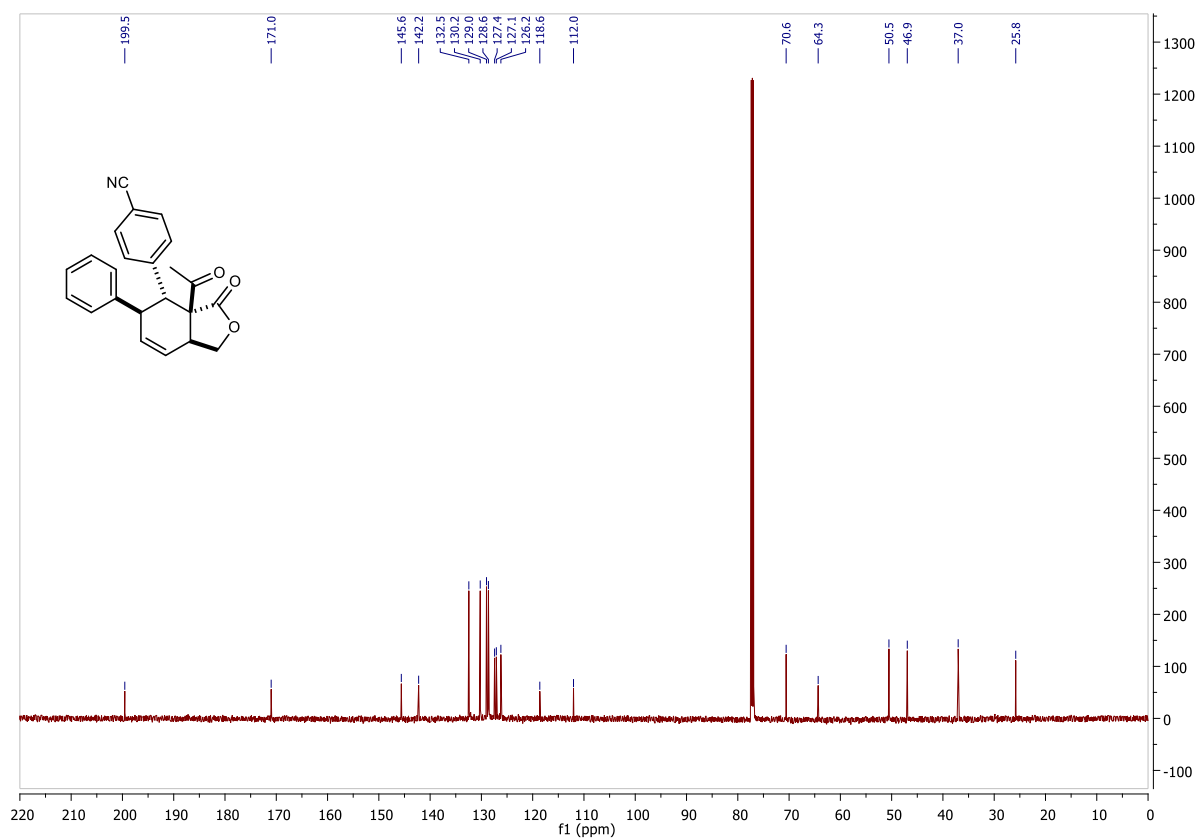
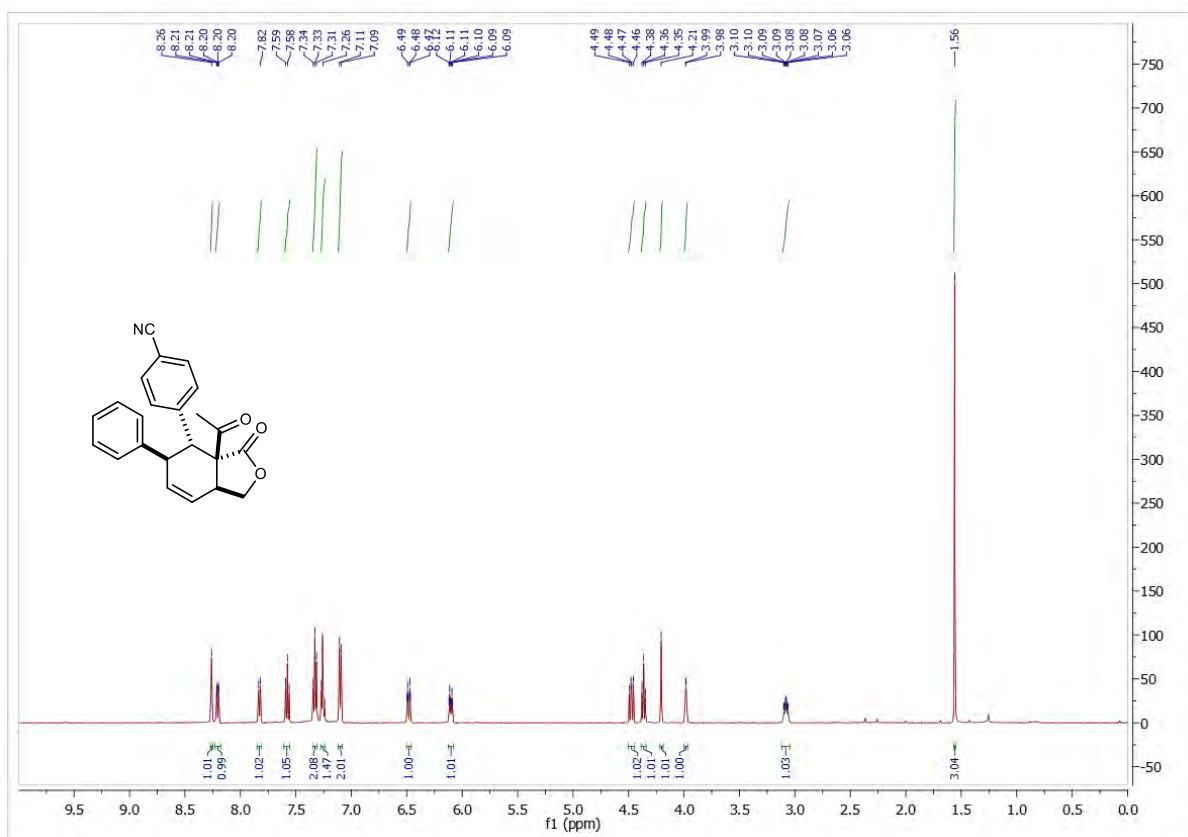
**IR** (film, CHCl<sub>3</sub>) 3023, 2969, 2935, 2842, 1756, 1711, 1668, 1599, 1240, 1173 cm<sup>-1</sup>;

**HRMS** (EI) *m/z*: [M]<sup>+</sup> Calcd for C<sub>19</sub>H<sub>20</sub>O<sub>5</sub>Na 328.1311; Found 328.1322.

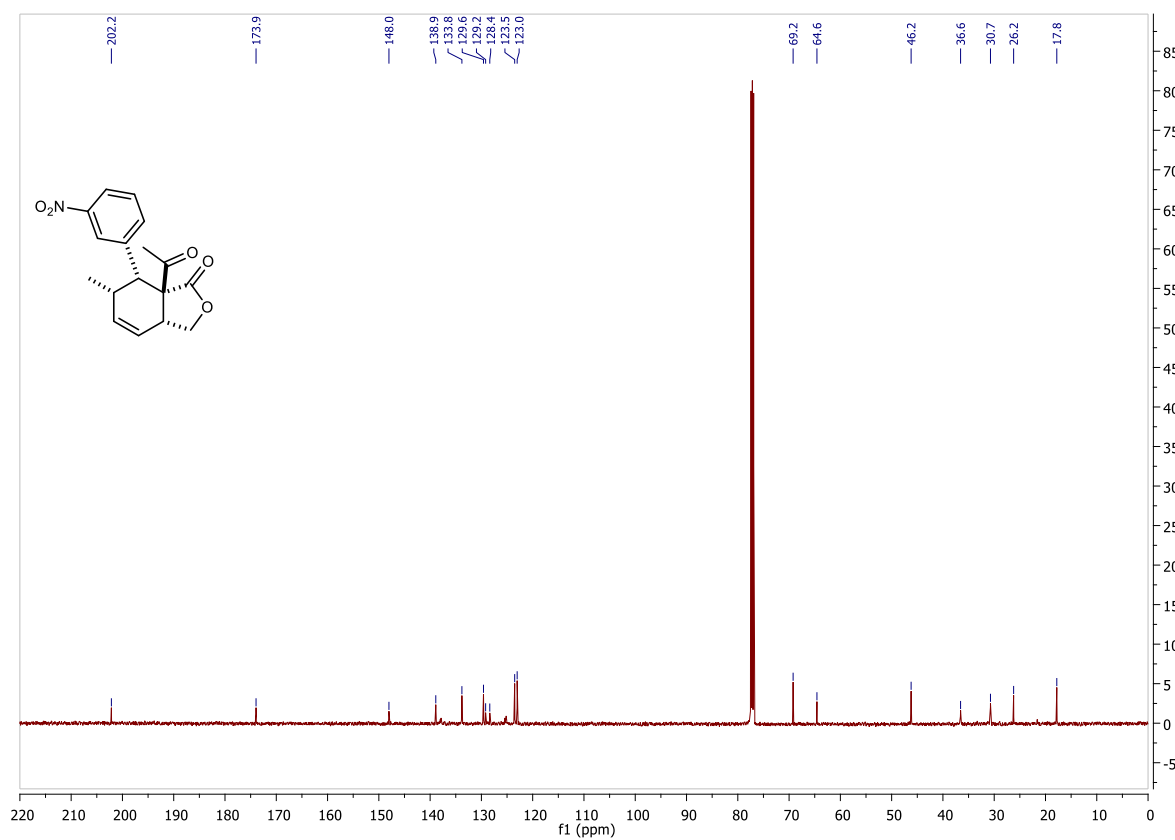
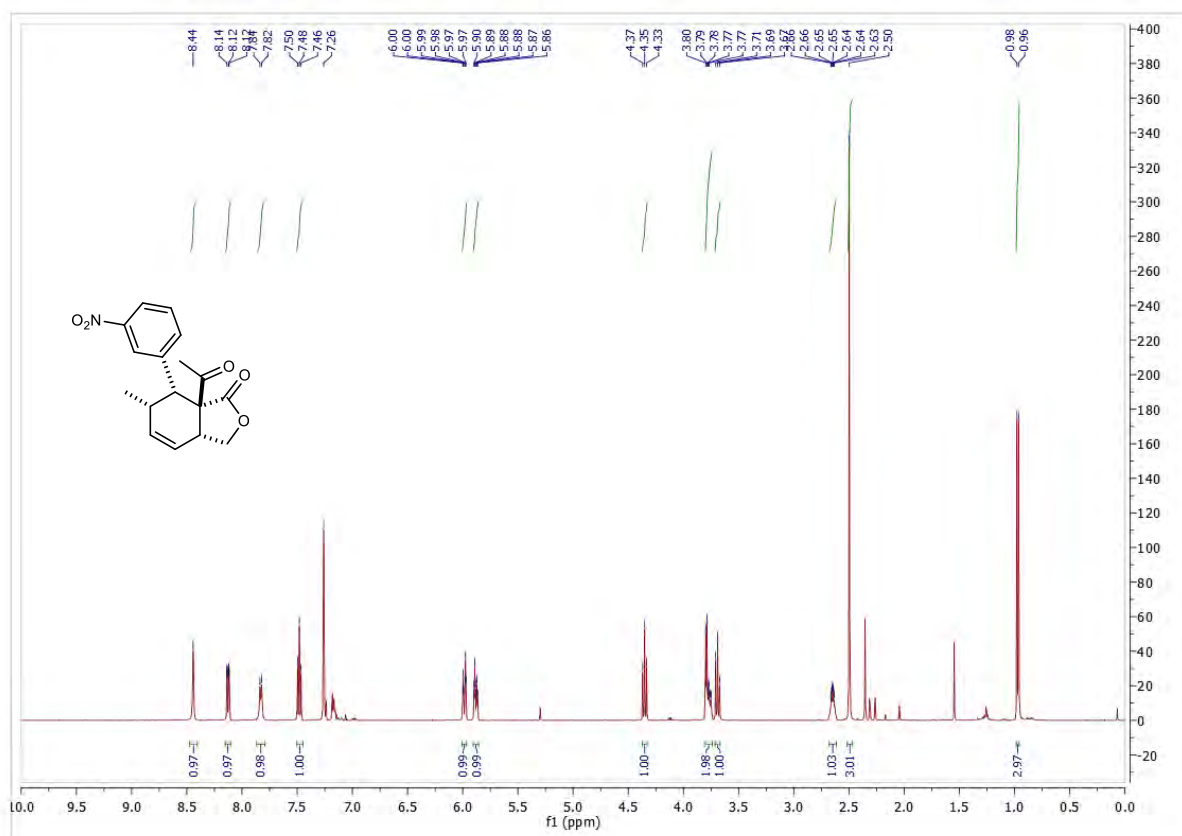




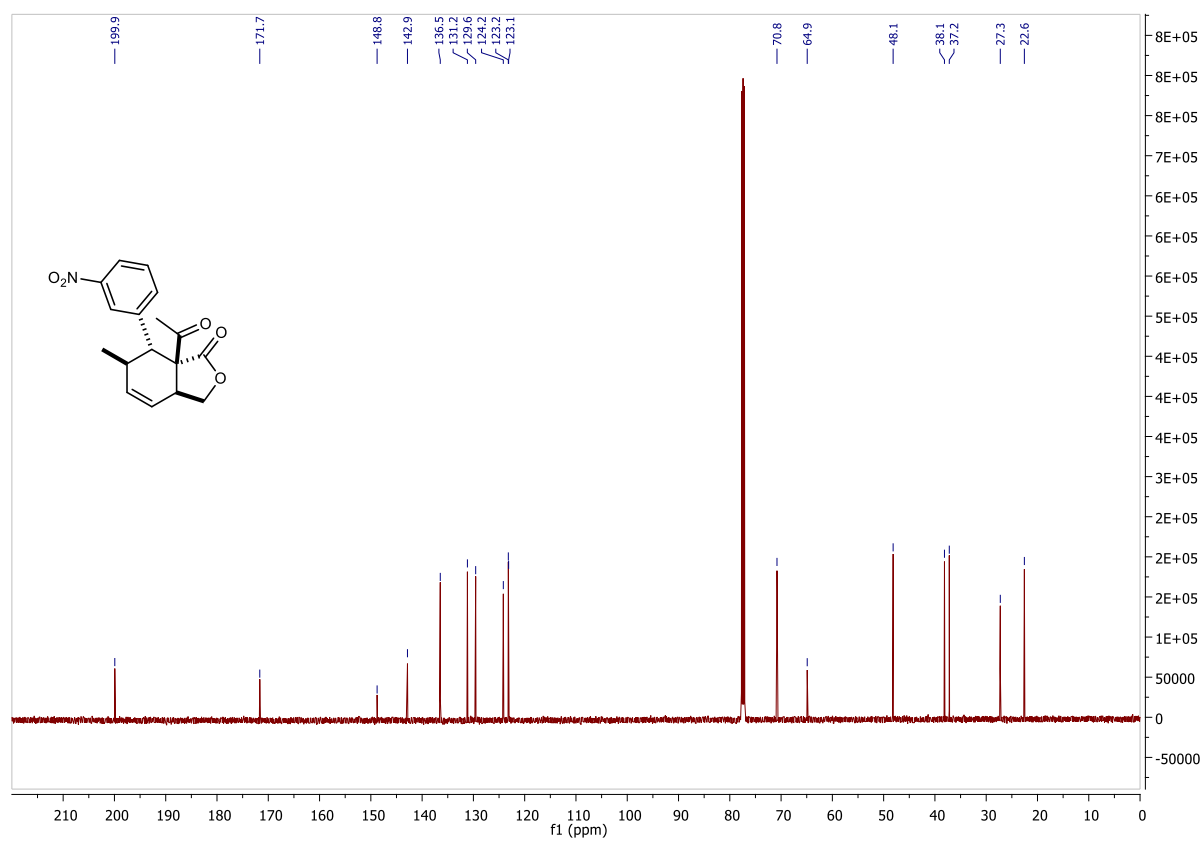
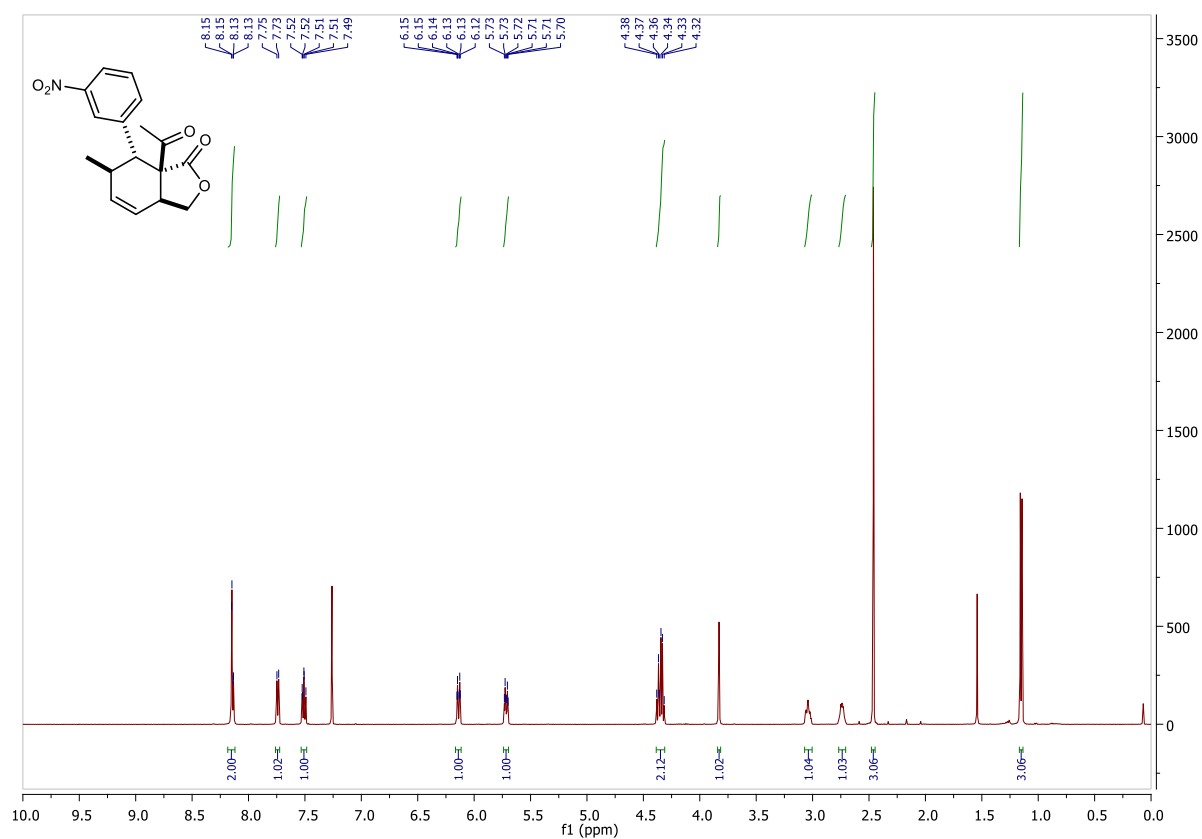
**Supplementary Figure S105.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6a**.



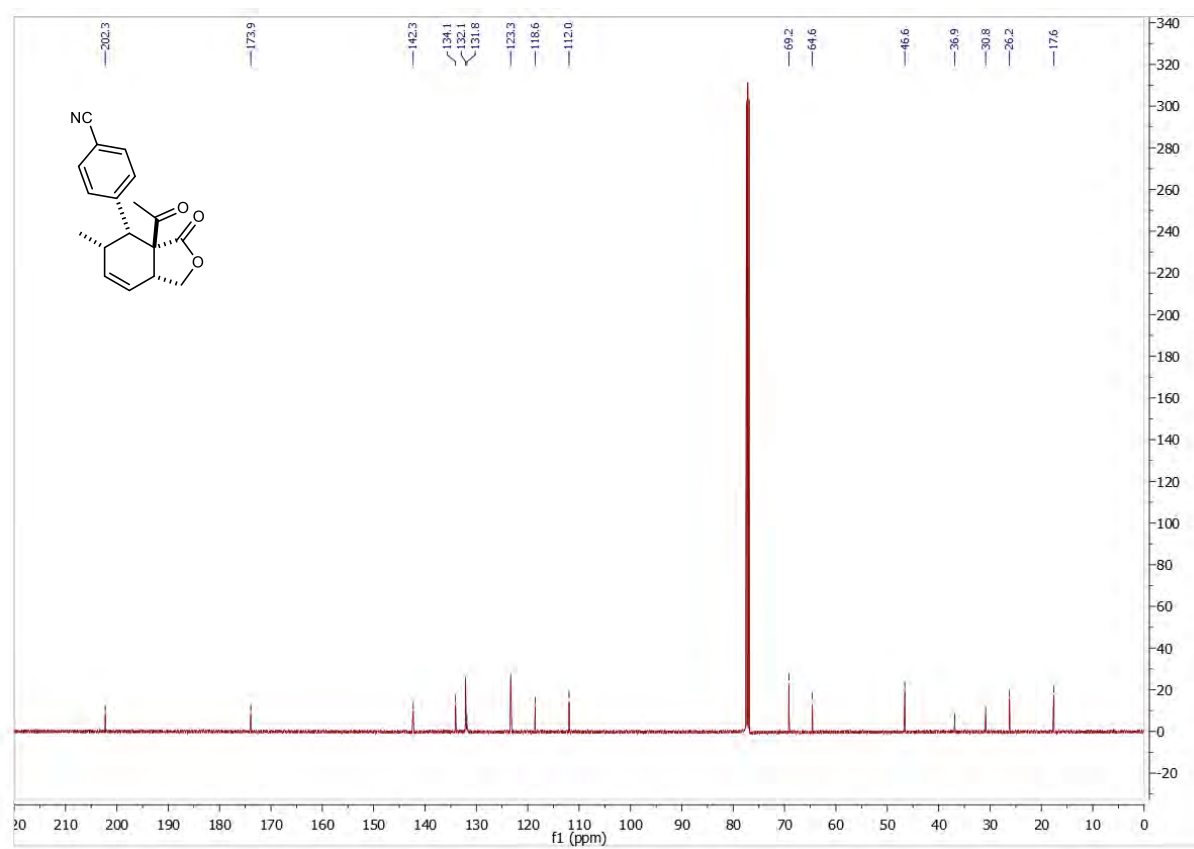
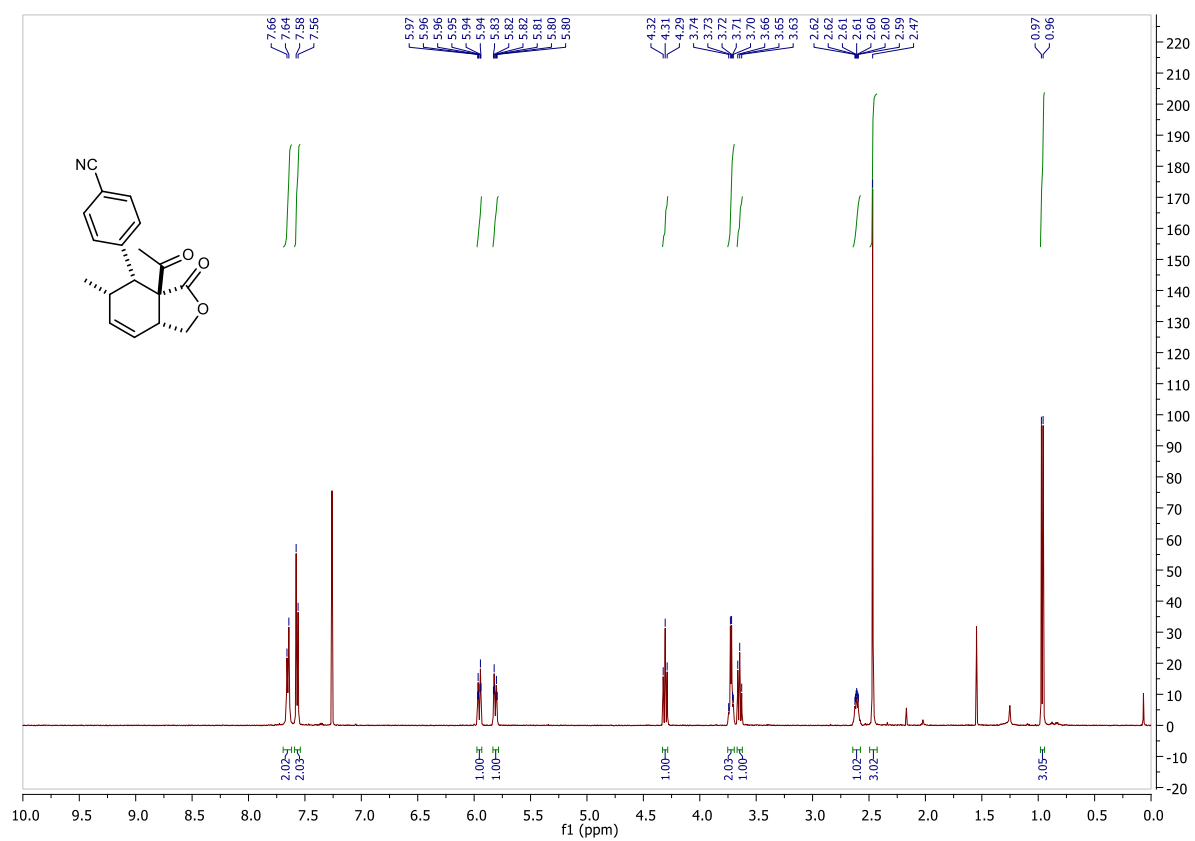
**Supplementary Figure S106.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6a'**.



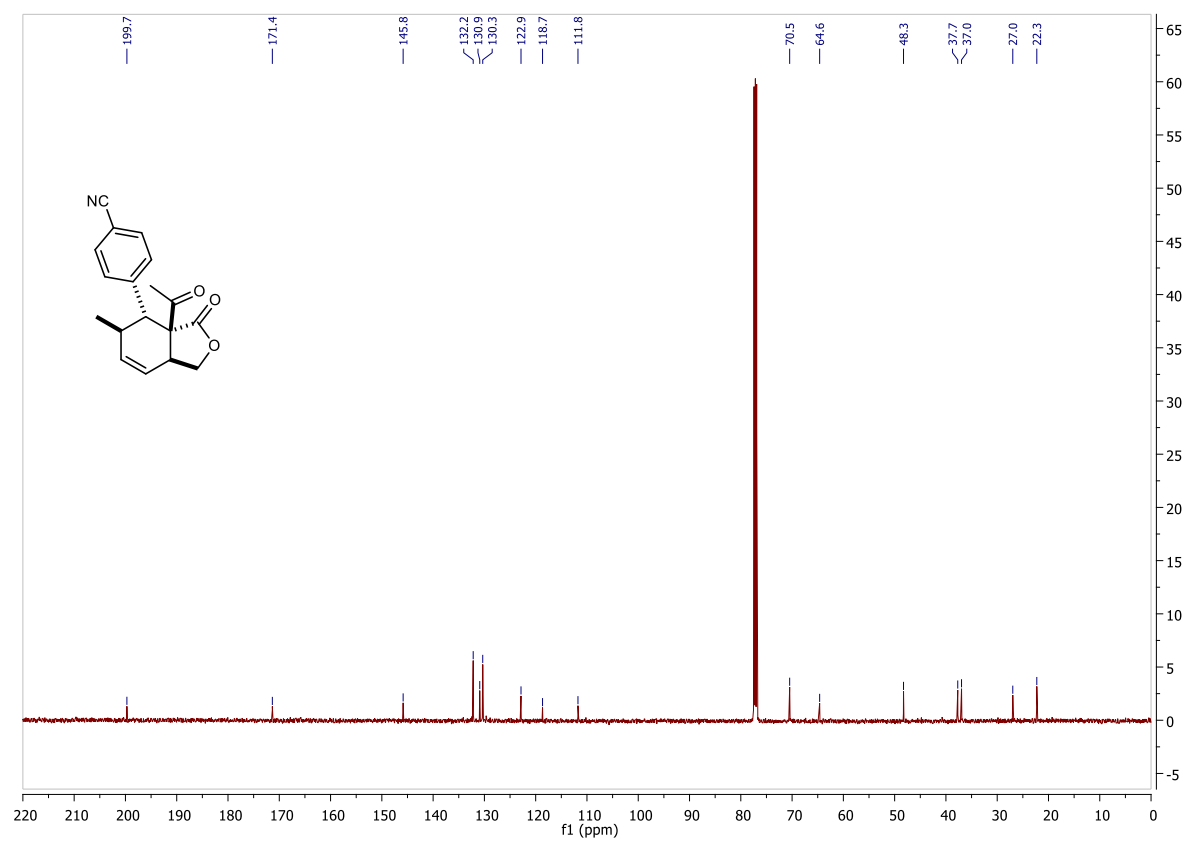
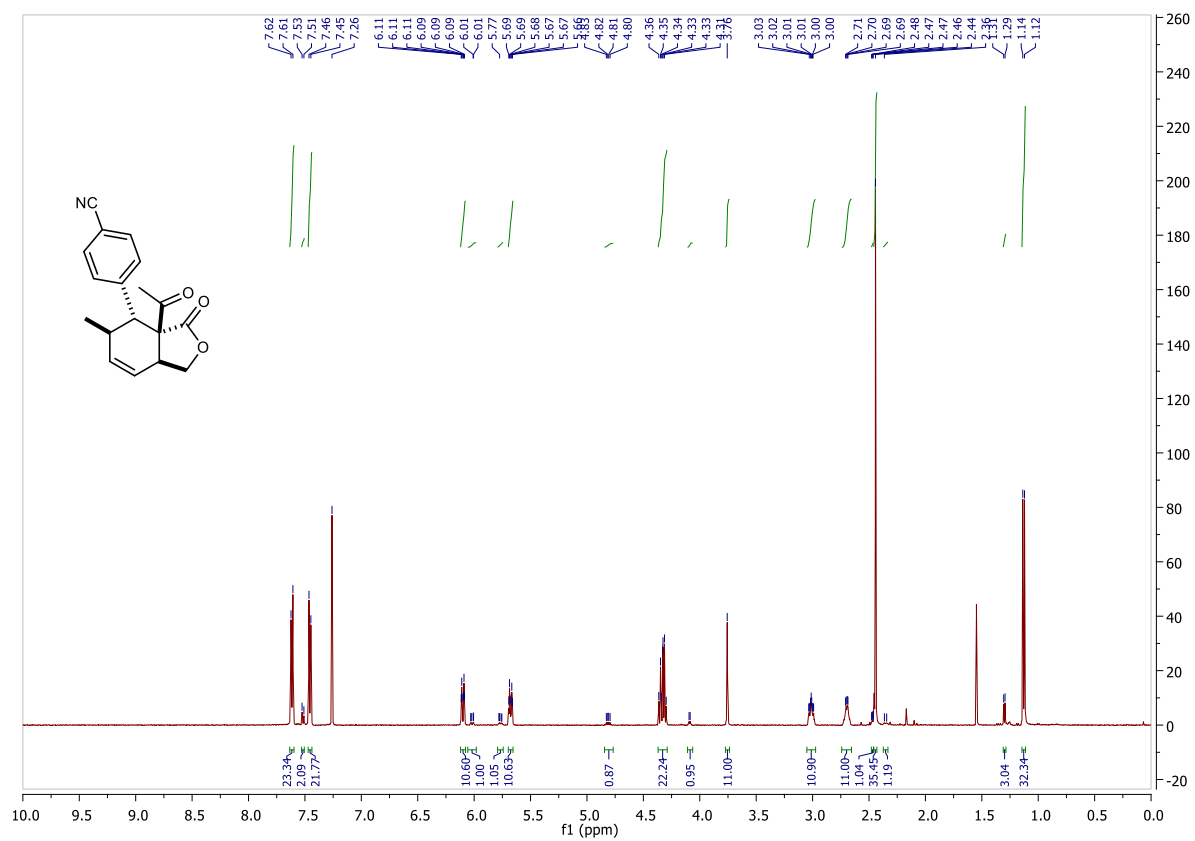
Supplementary Figure S107. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6b**.



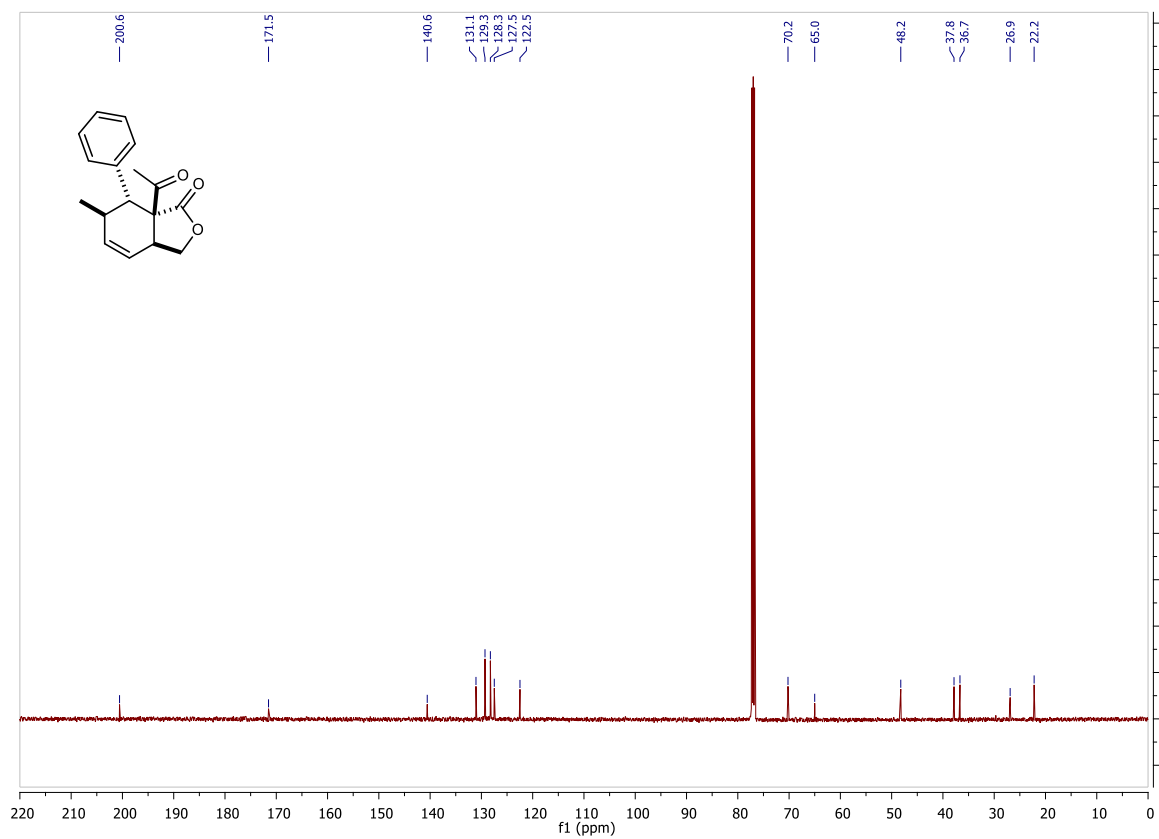
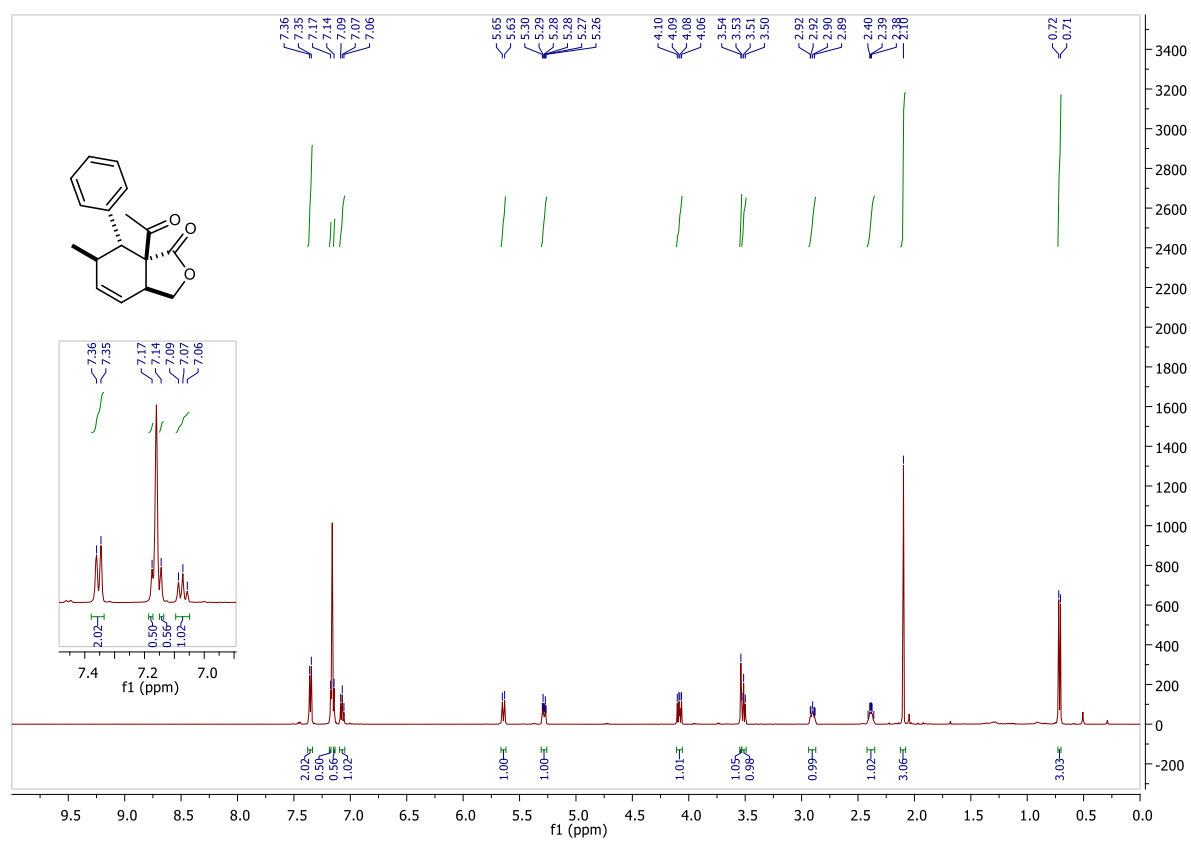
Supplementary Figure S108. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6b'**.



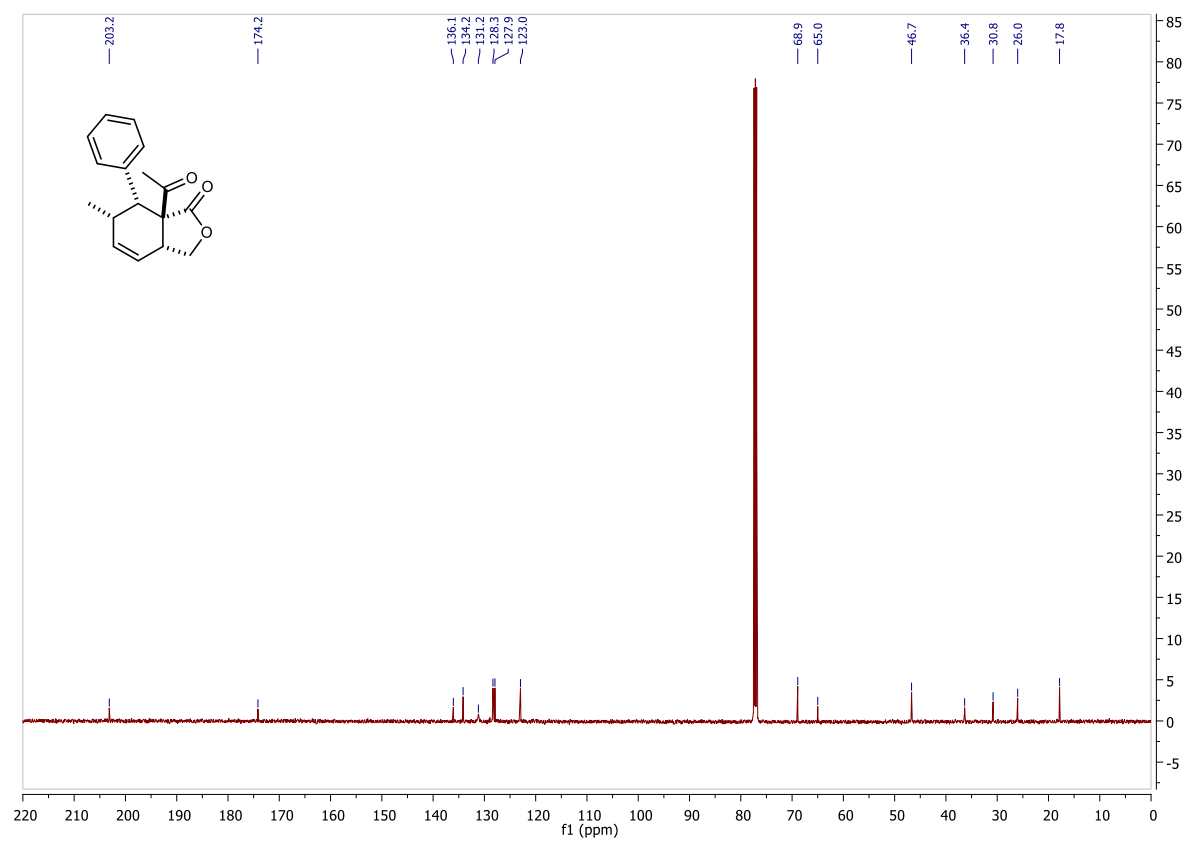
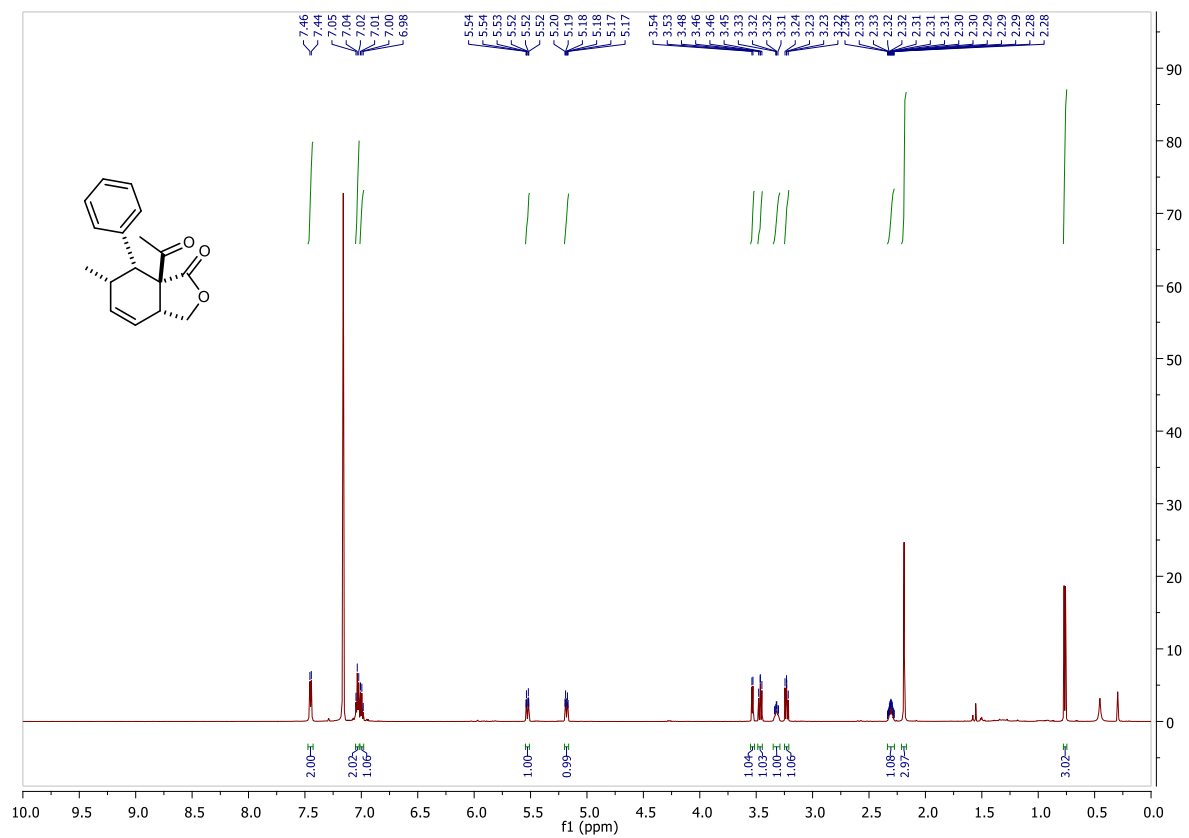
Supplementary Figure S109. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6c**.



Supplementary Figure S110. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6c**.

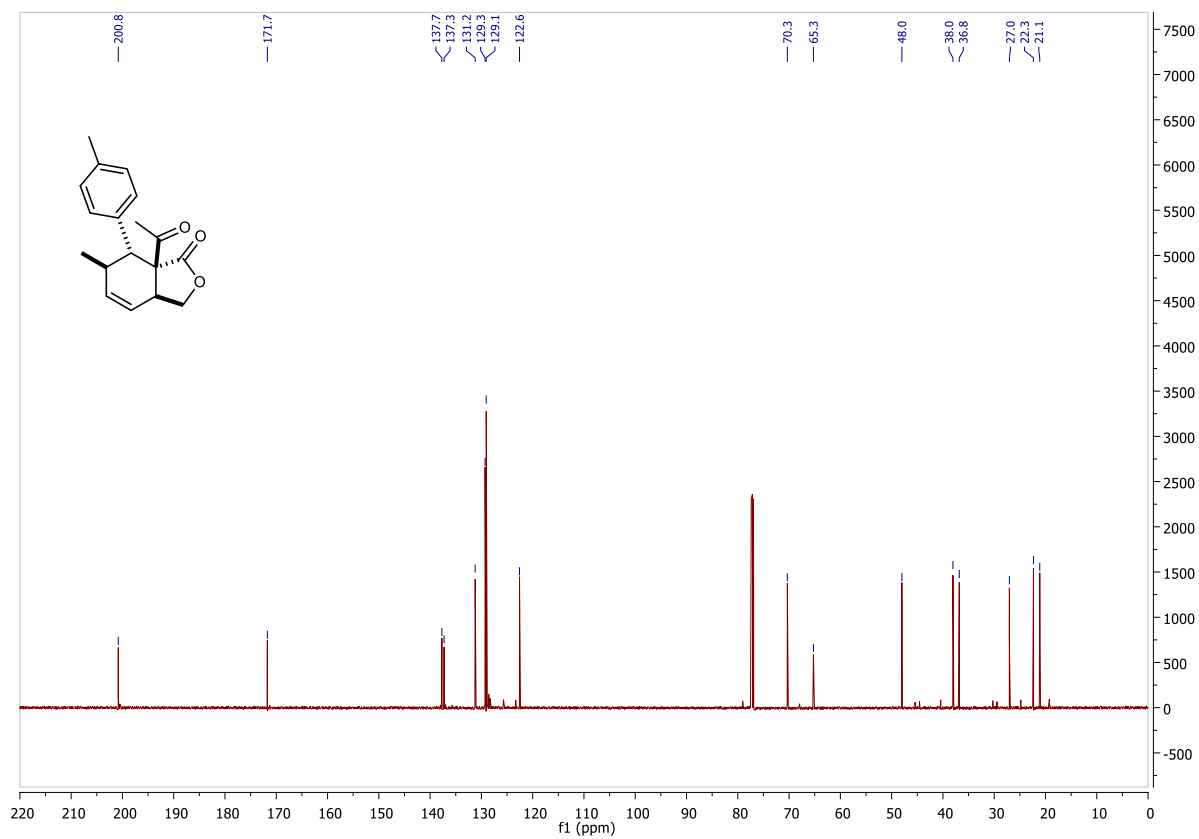
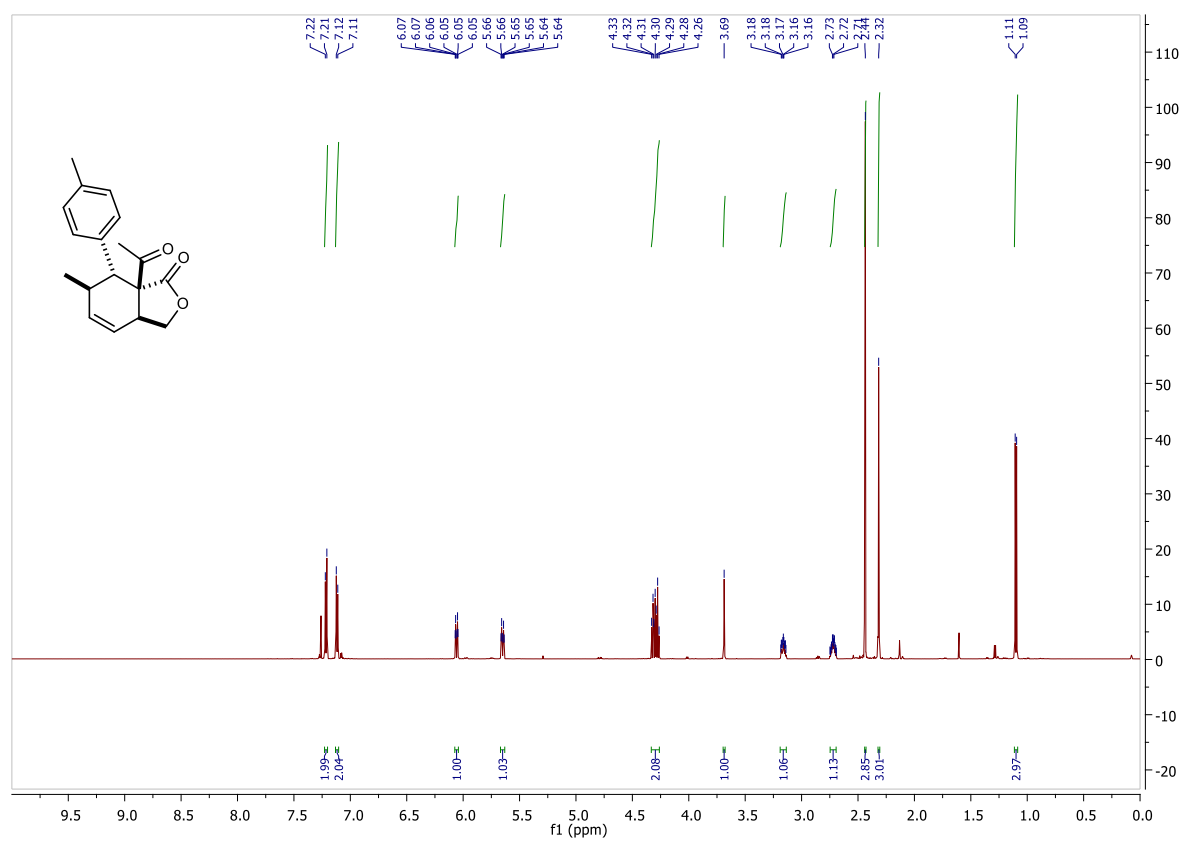


**Supplementary Figure S111.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6d**.

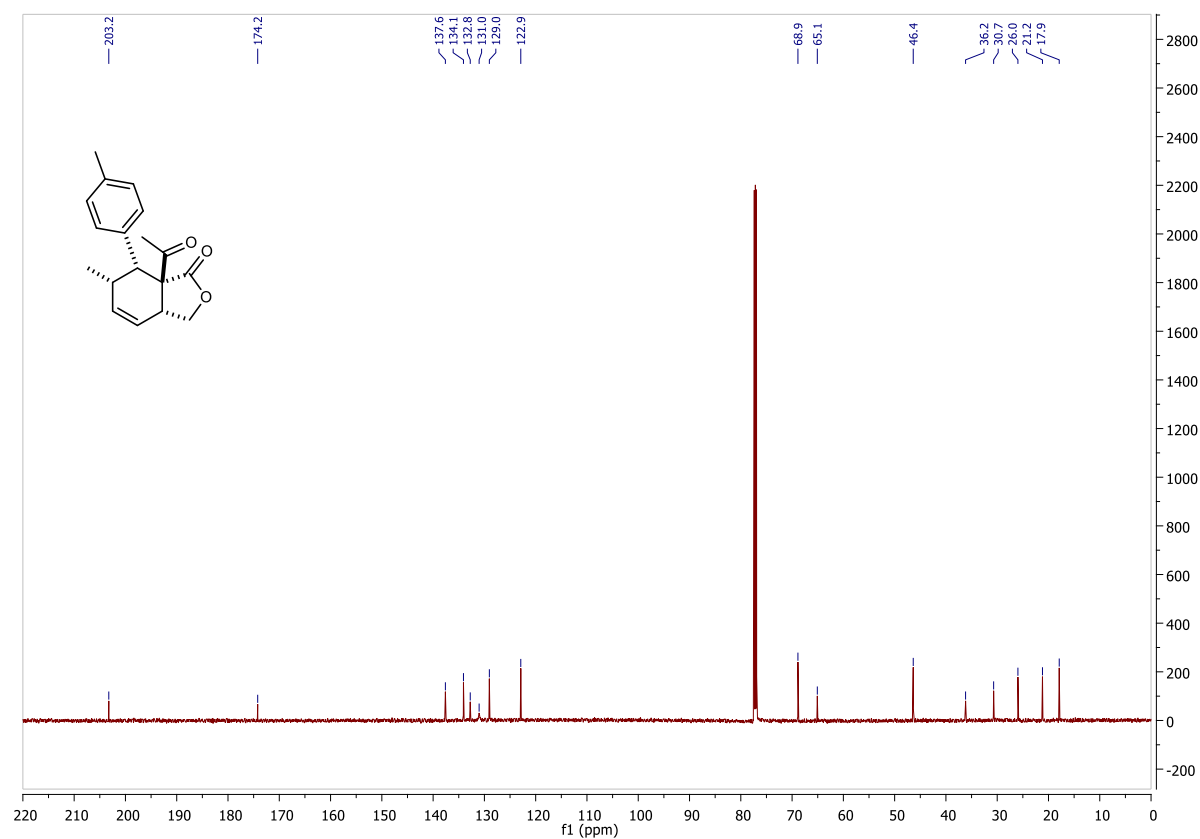
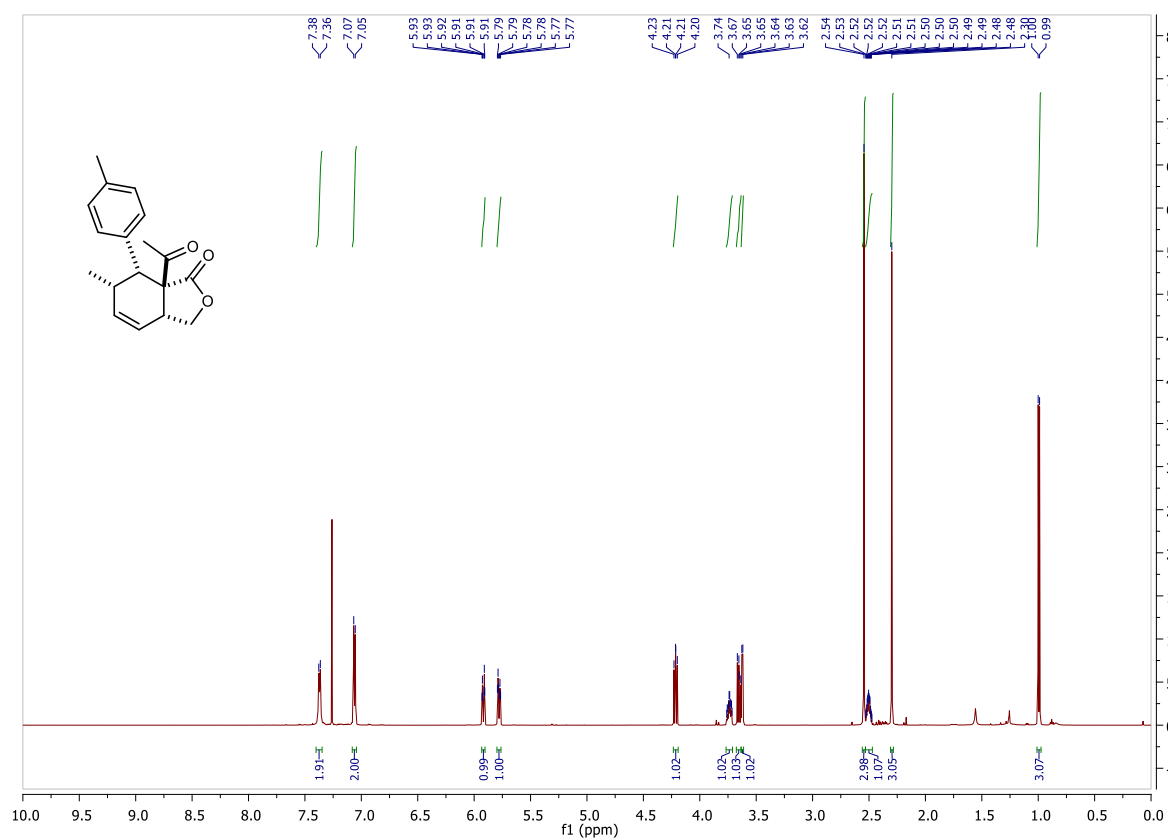


**Supplementary Figure S112.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6d'**.

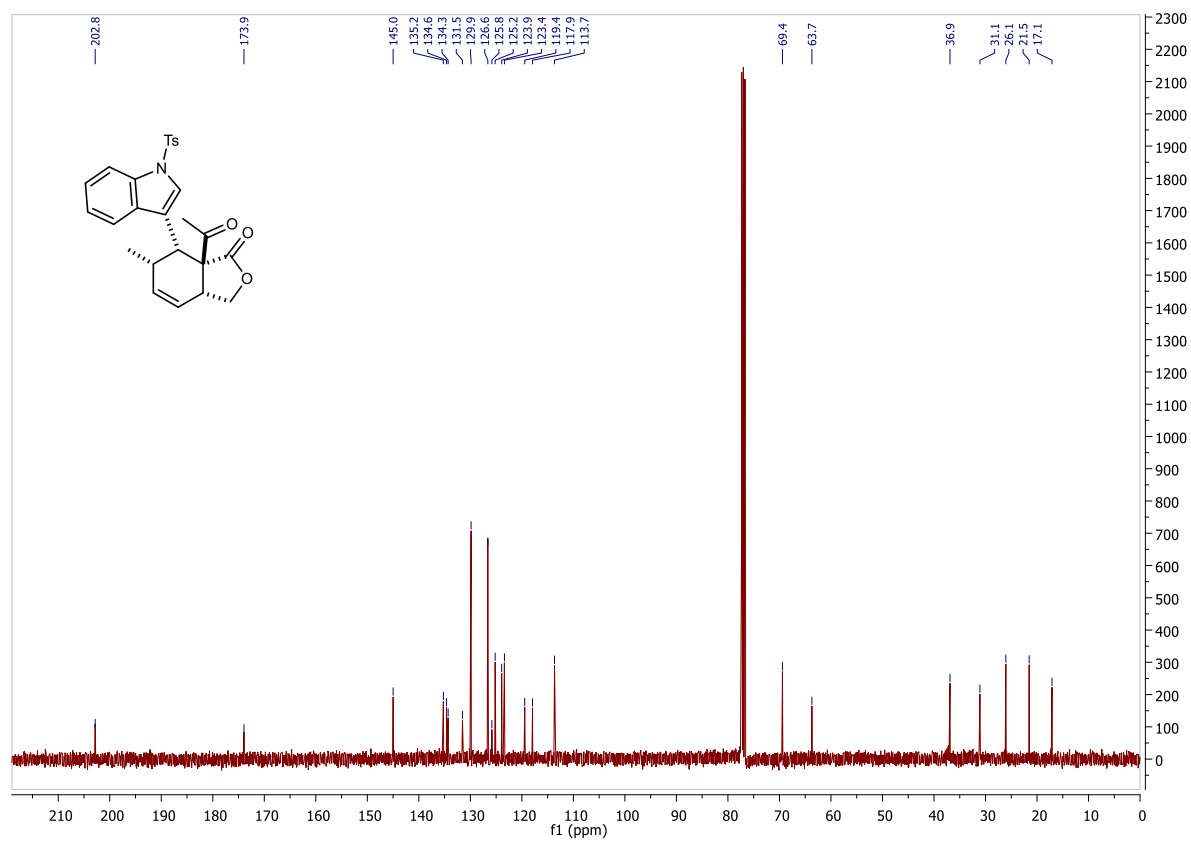
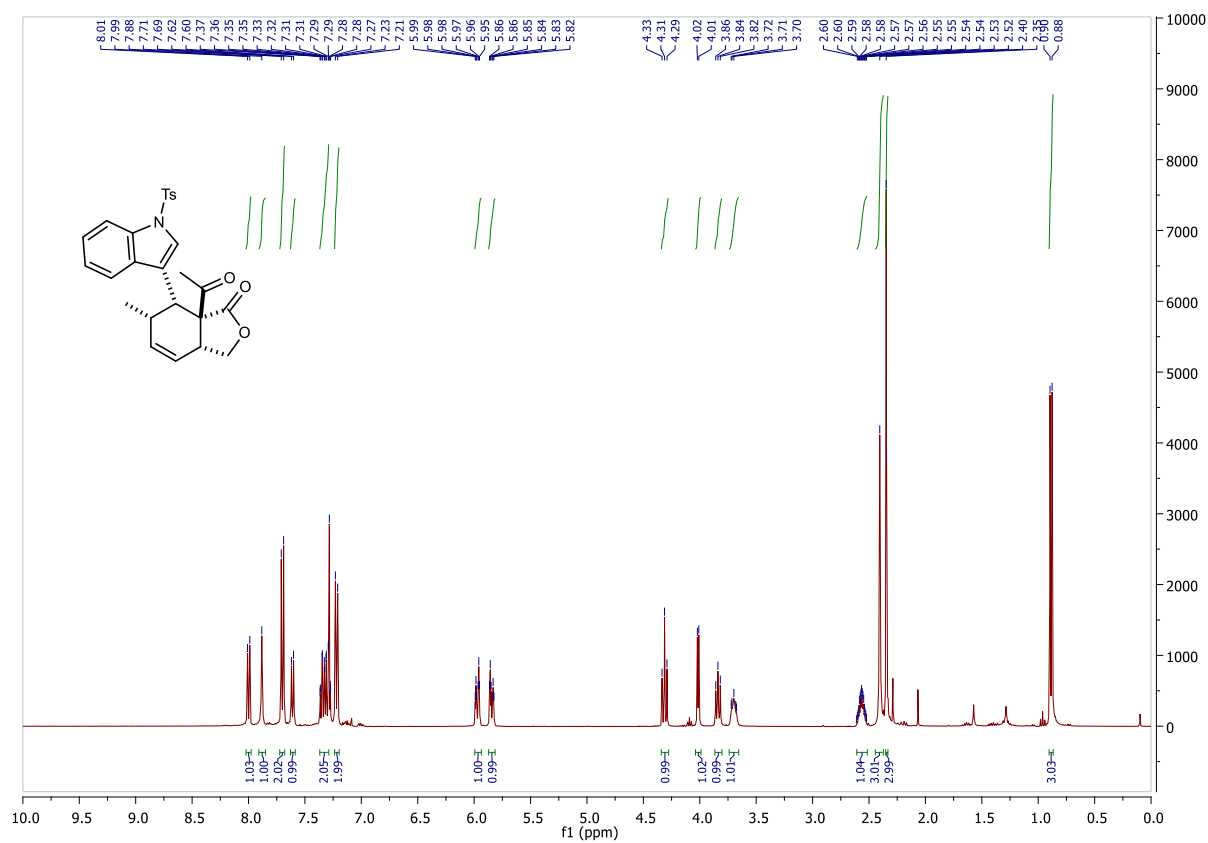




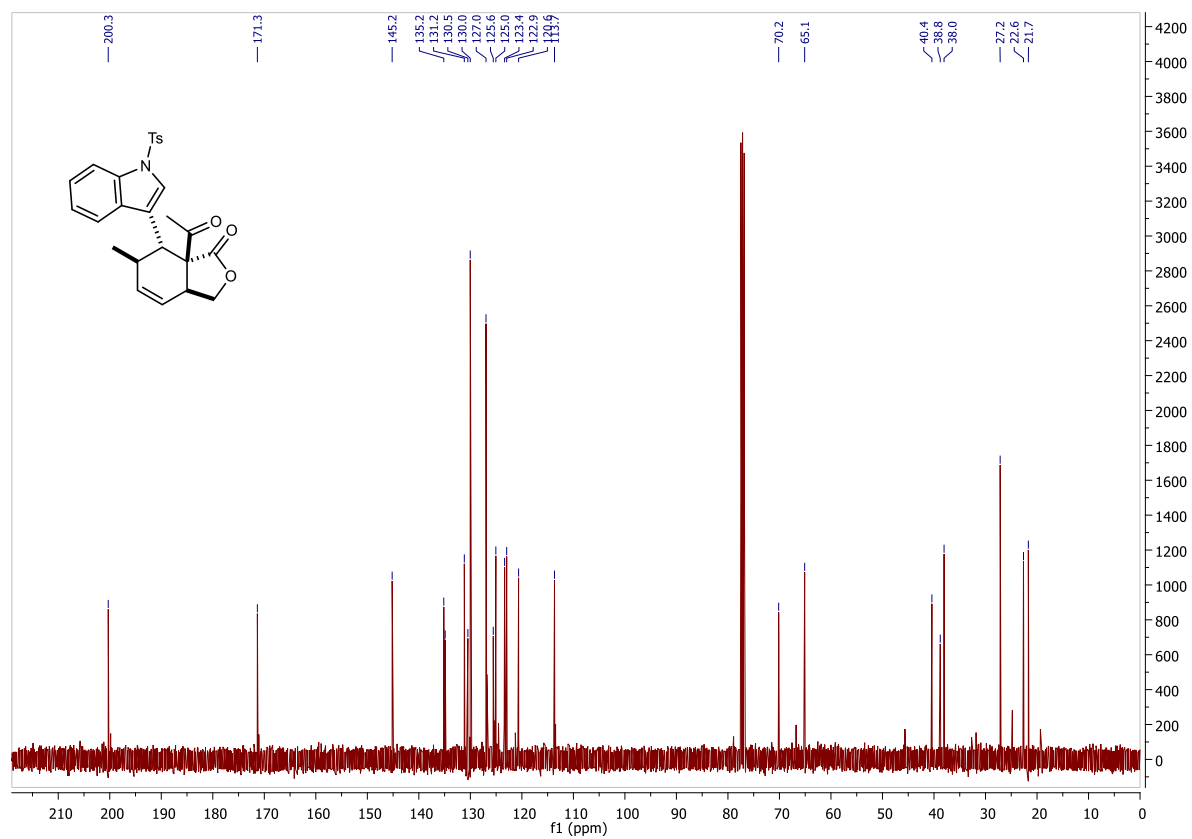
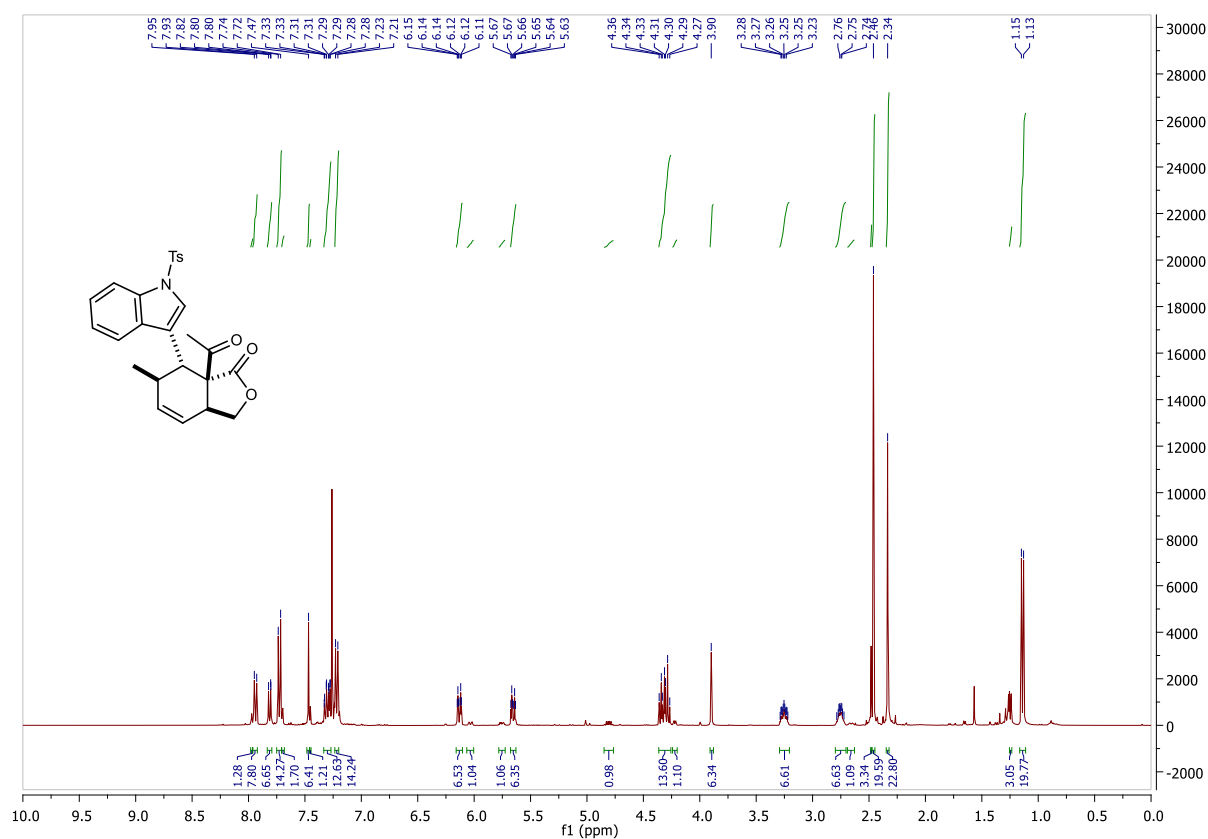
**Supplementary Figure S113.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6e**.



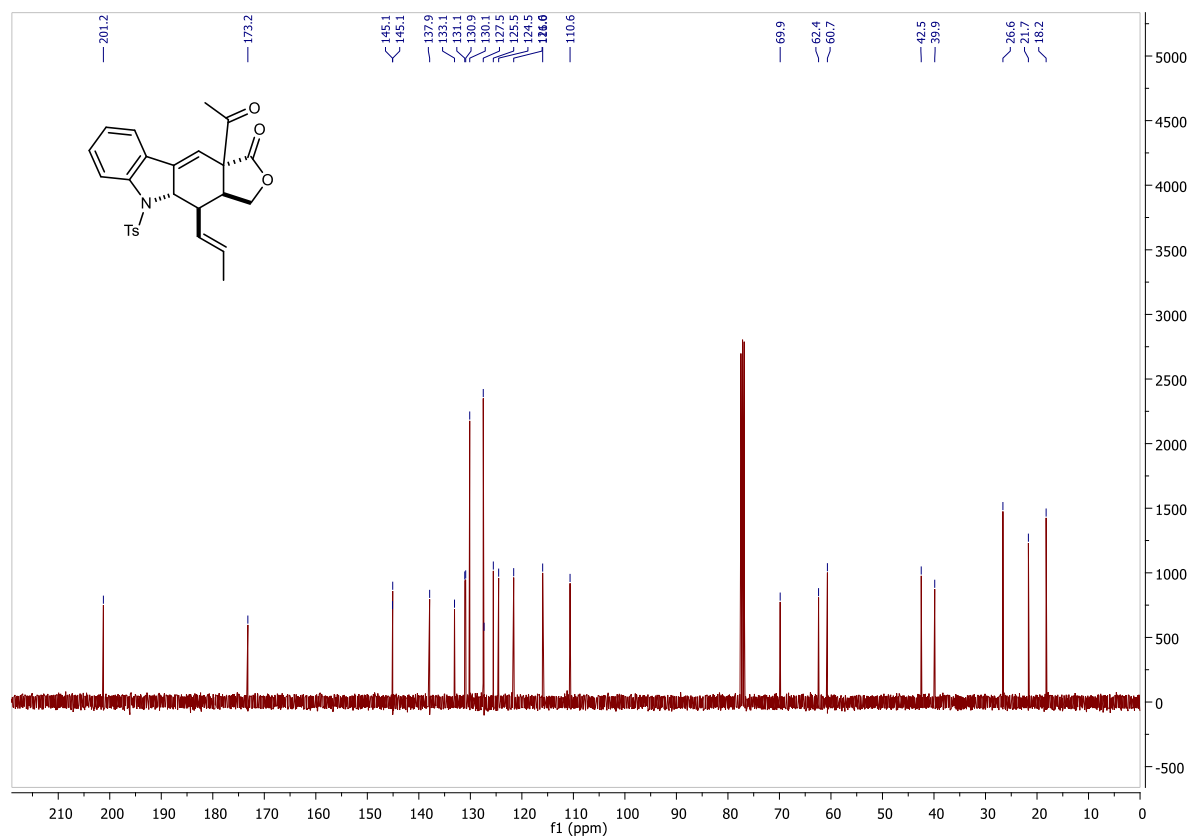
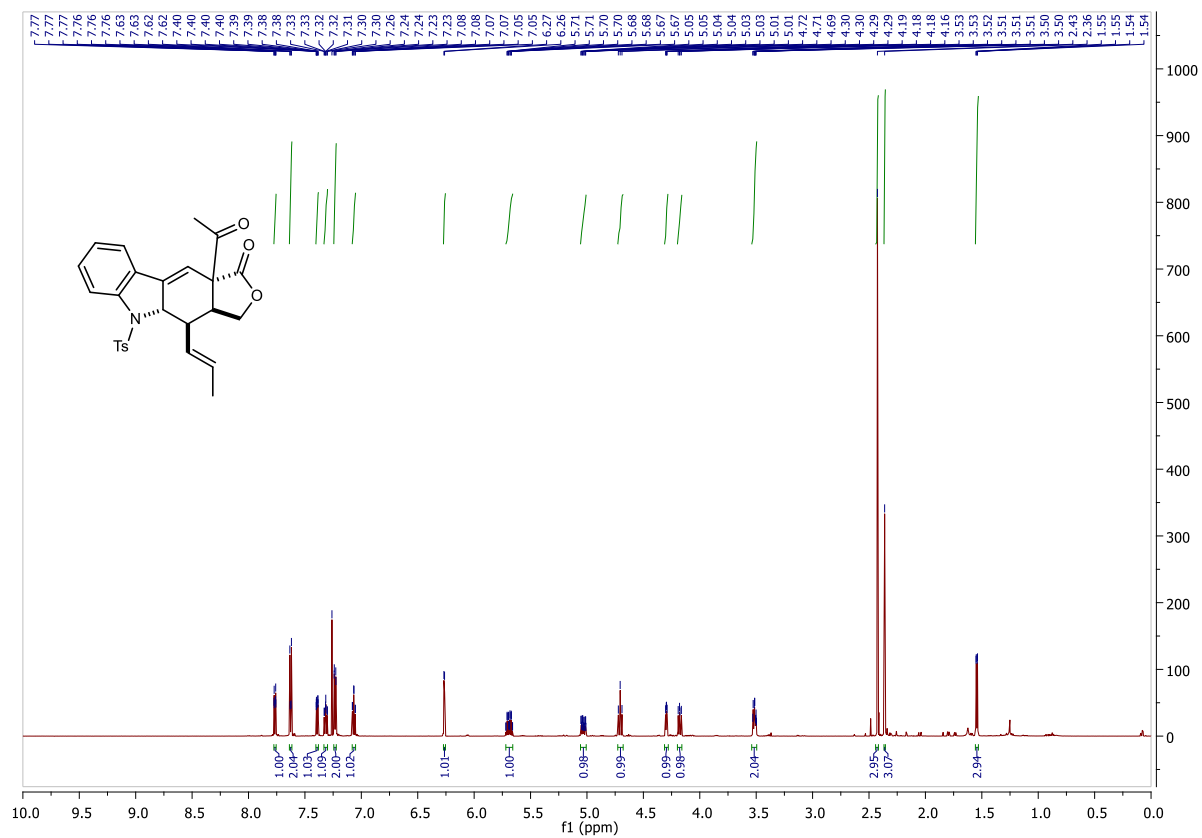
**Supplementary Figure S114.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6e**.



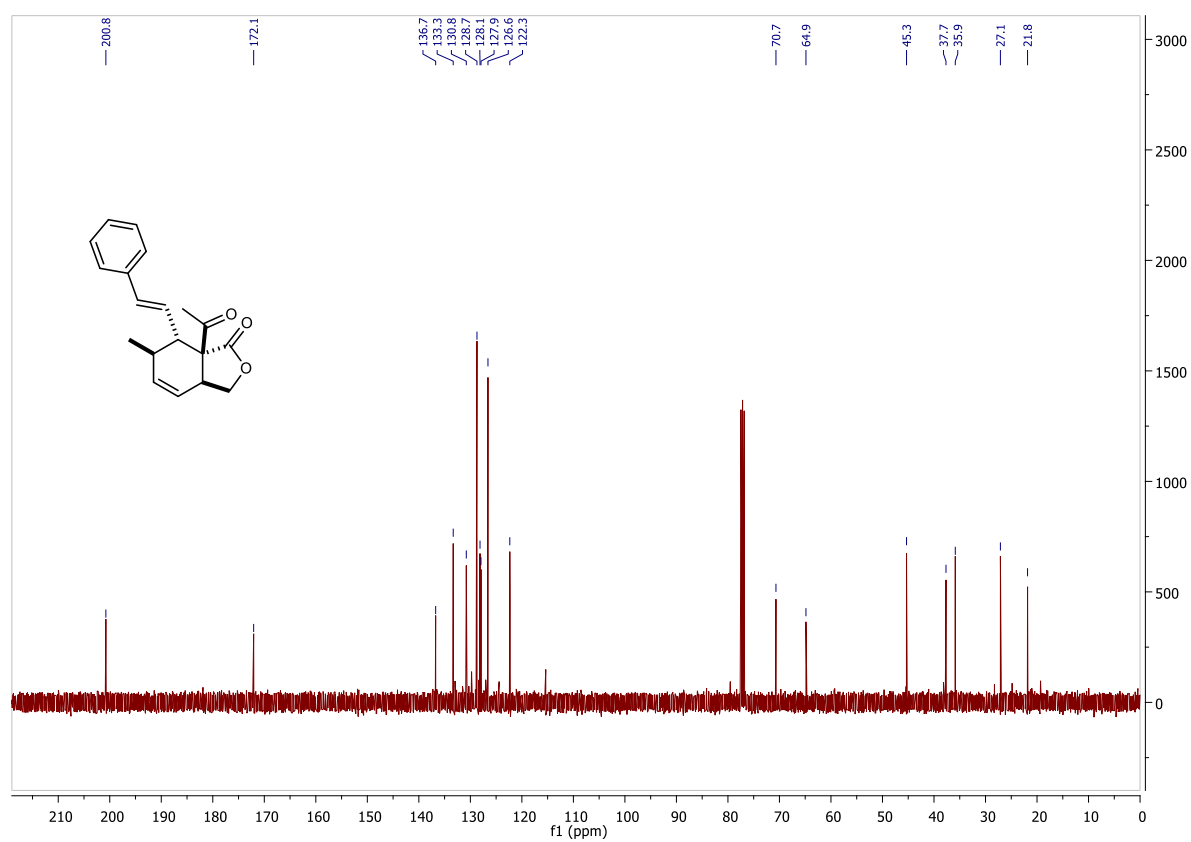
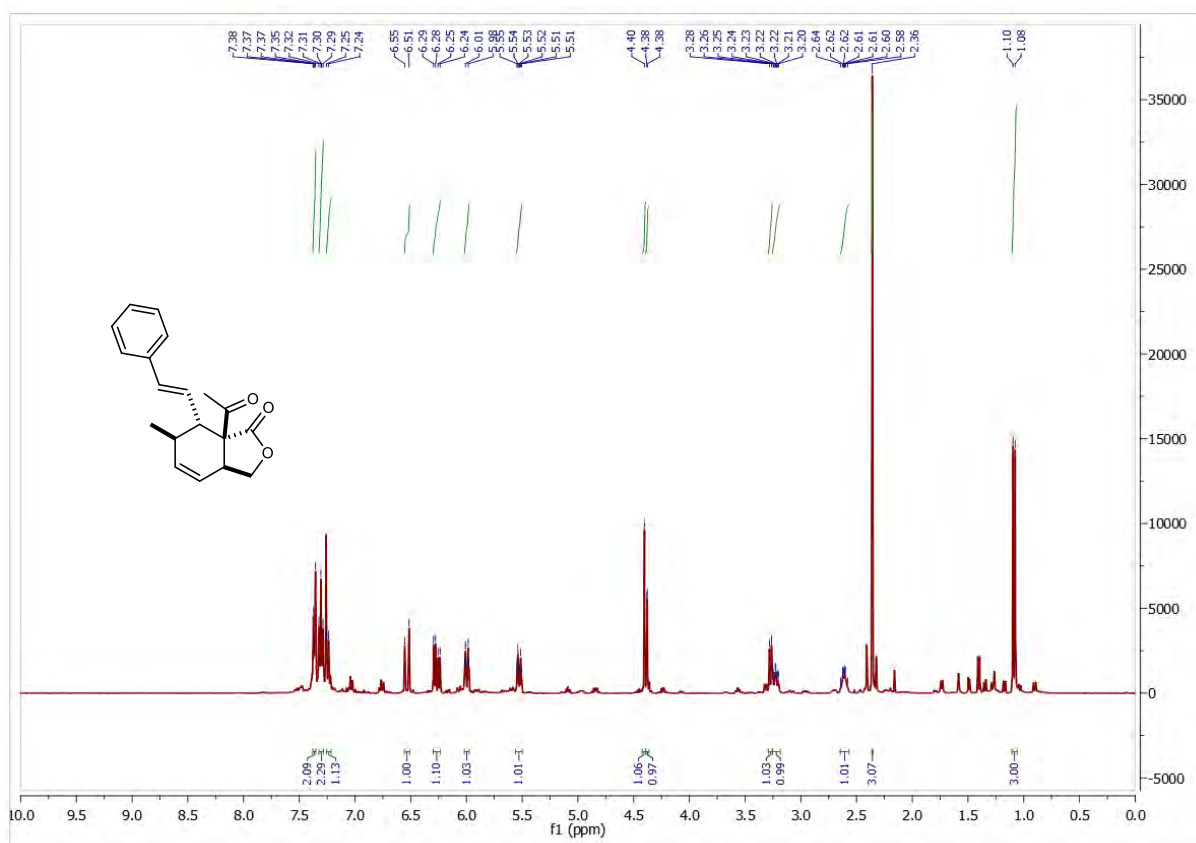
Supplementary Figure S115.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound **6f**



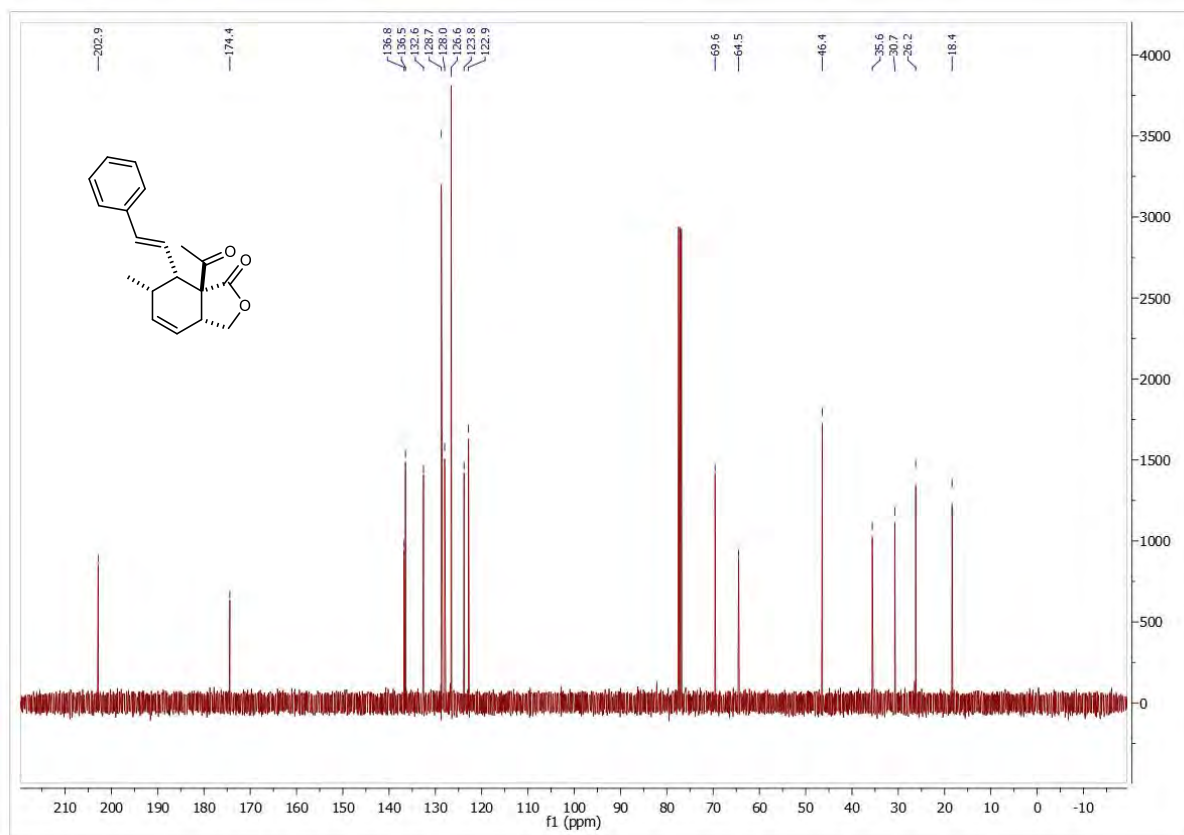
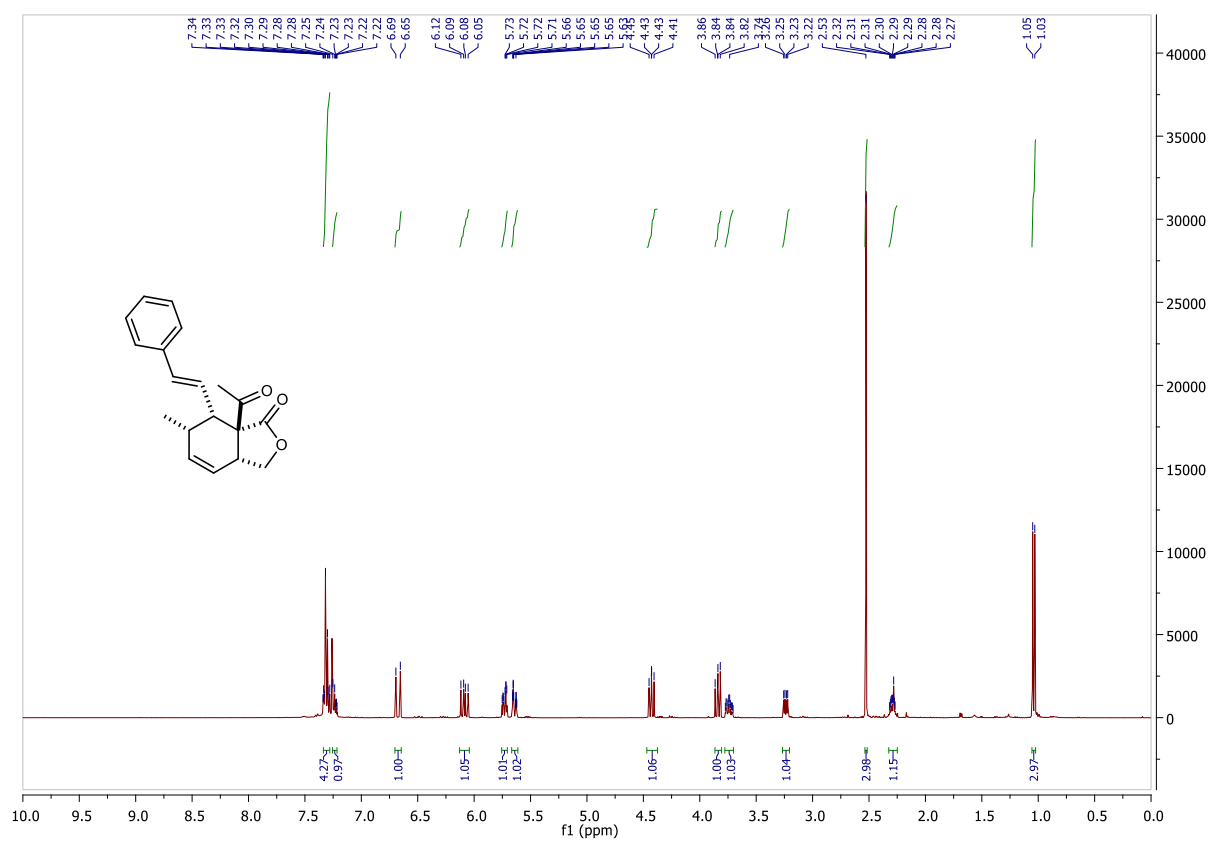
Supplementary Figure S116. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6f**.



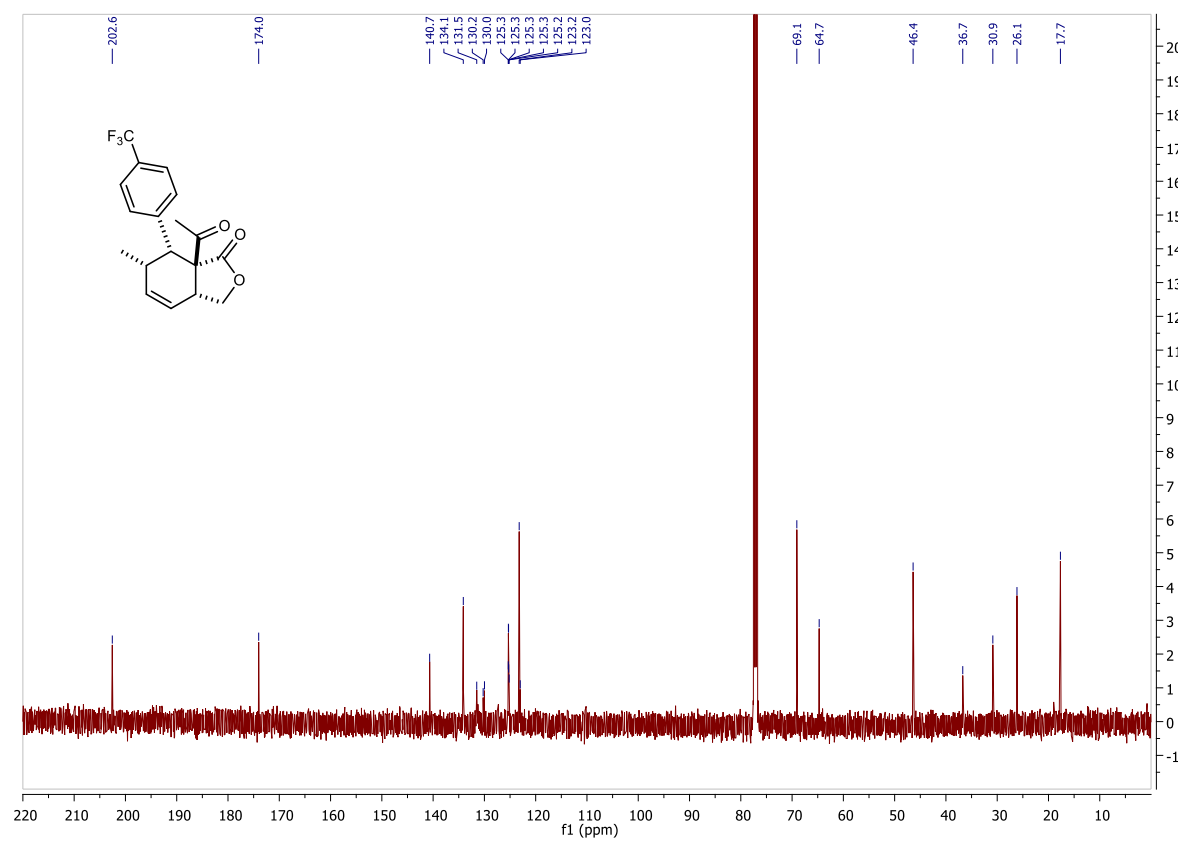
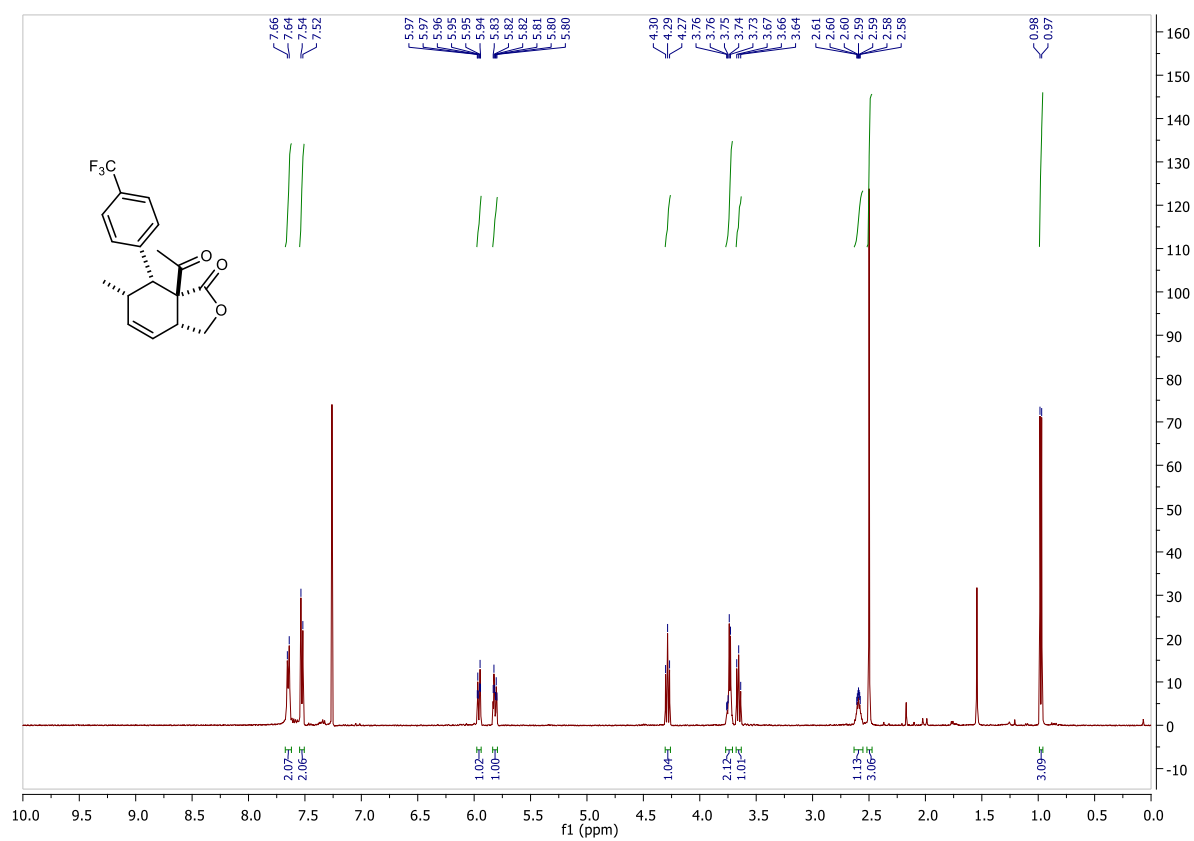
Supplementary Figure S117. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6g**.



Supplementary Figure S118. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6h**.

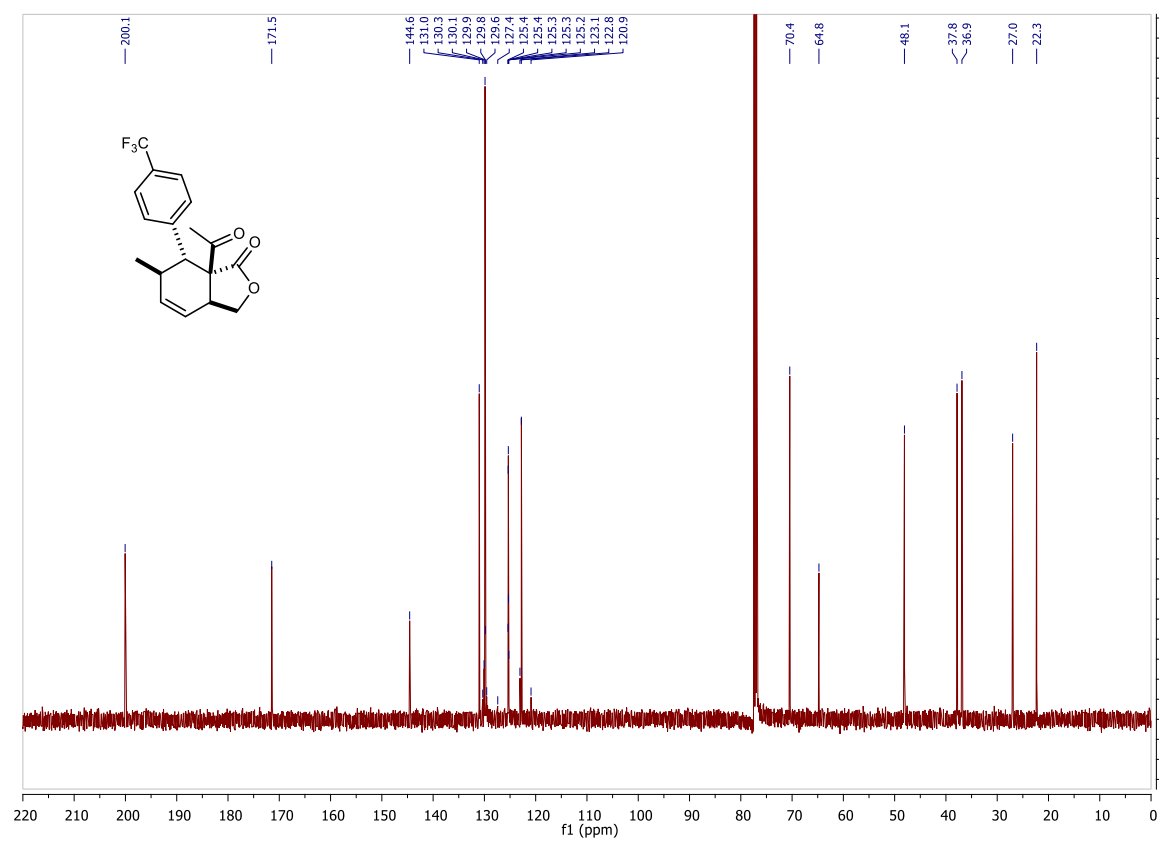
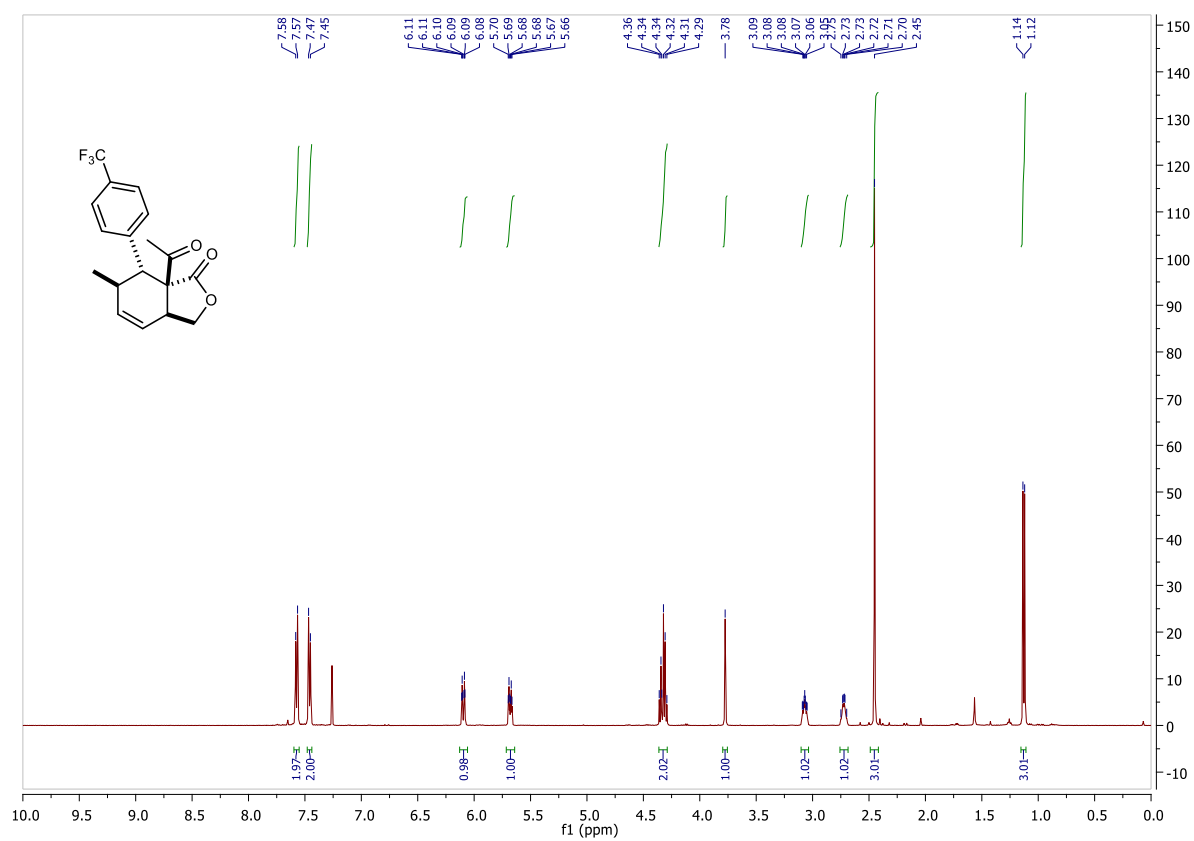


Supplementary Figure S119. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6h'**.

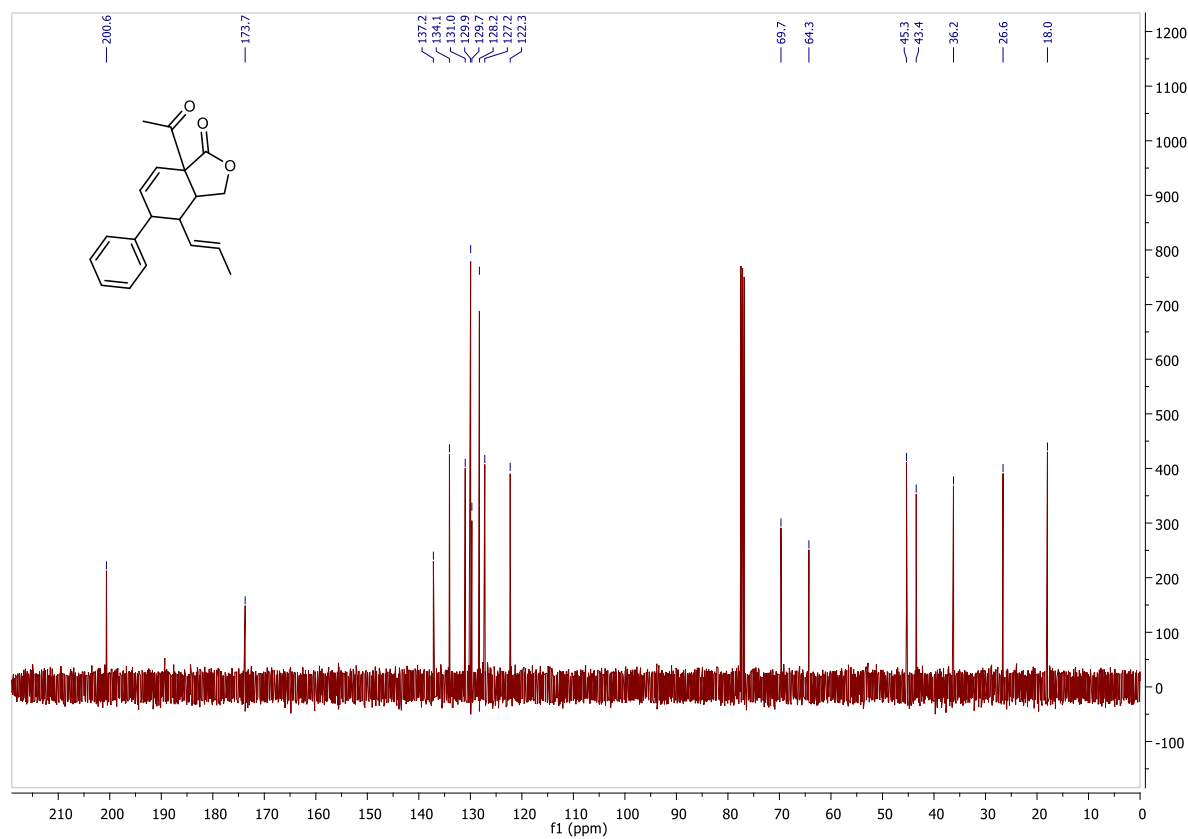
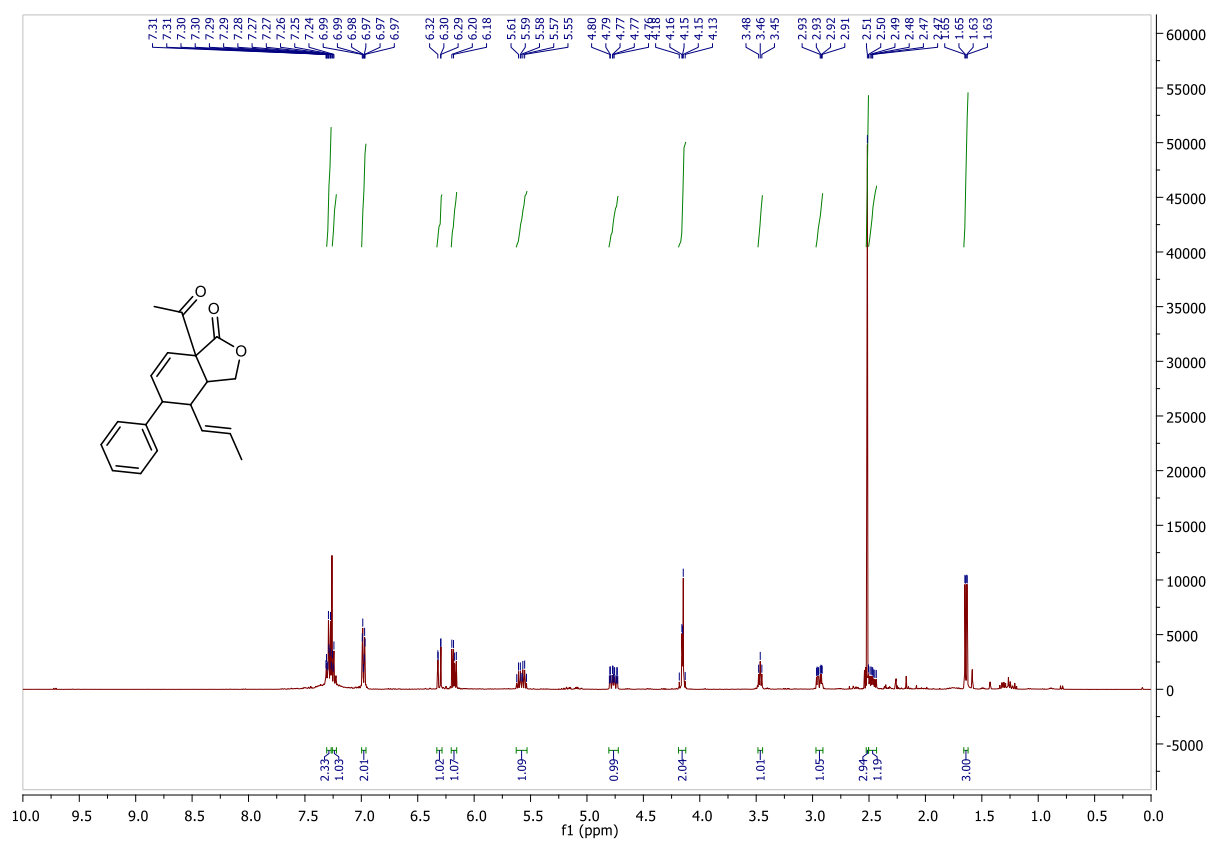


**Supplementary Figure S120.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6i**.

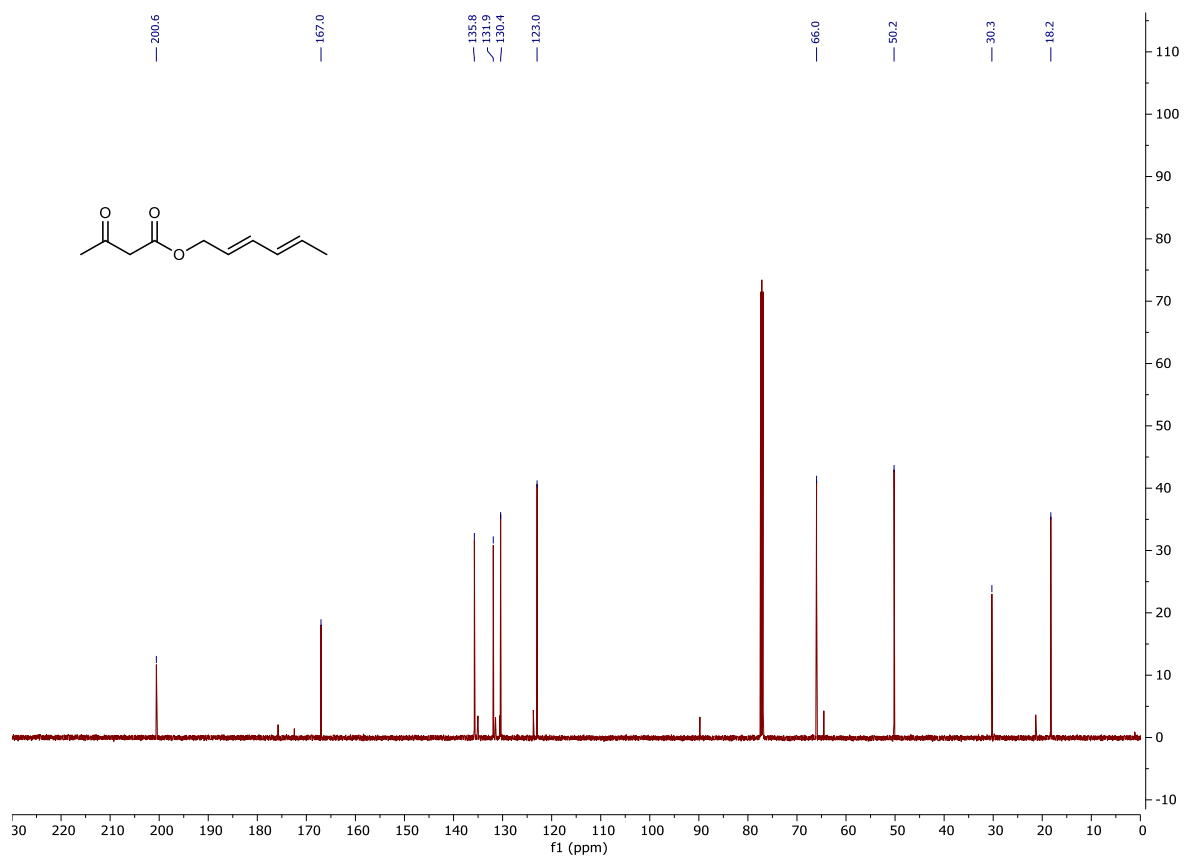
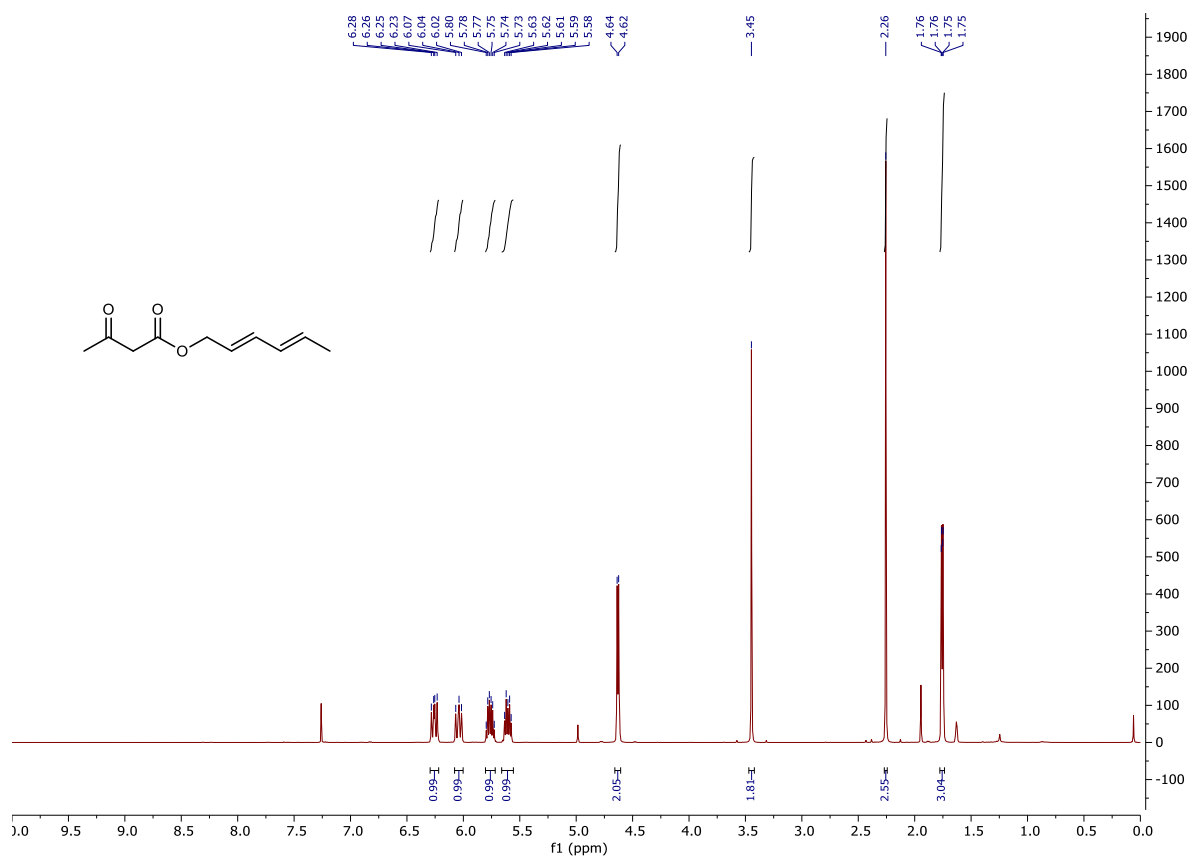




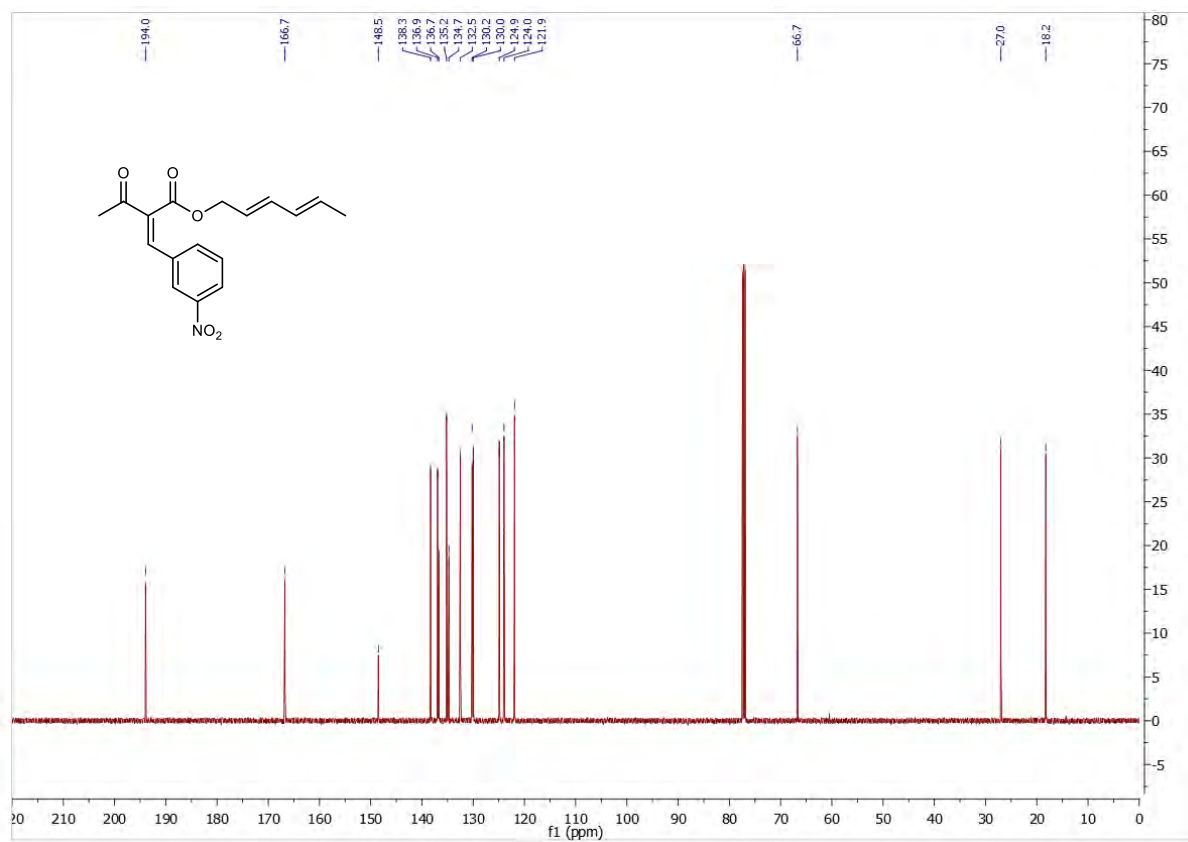
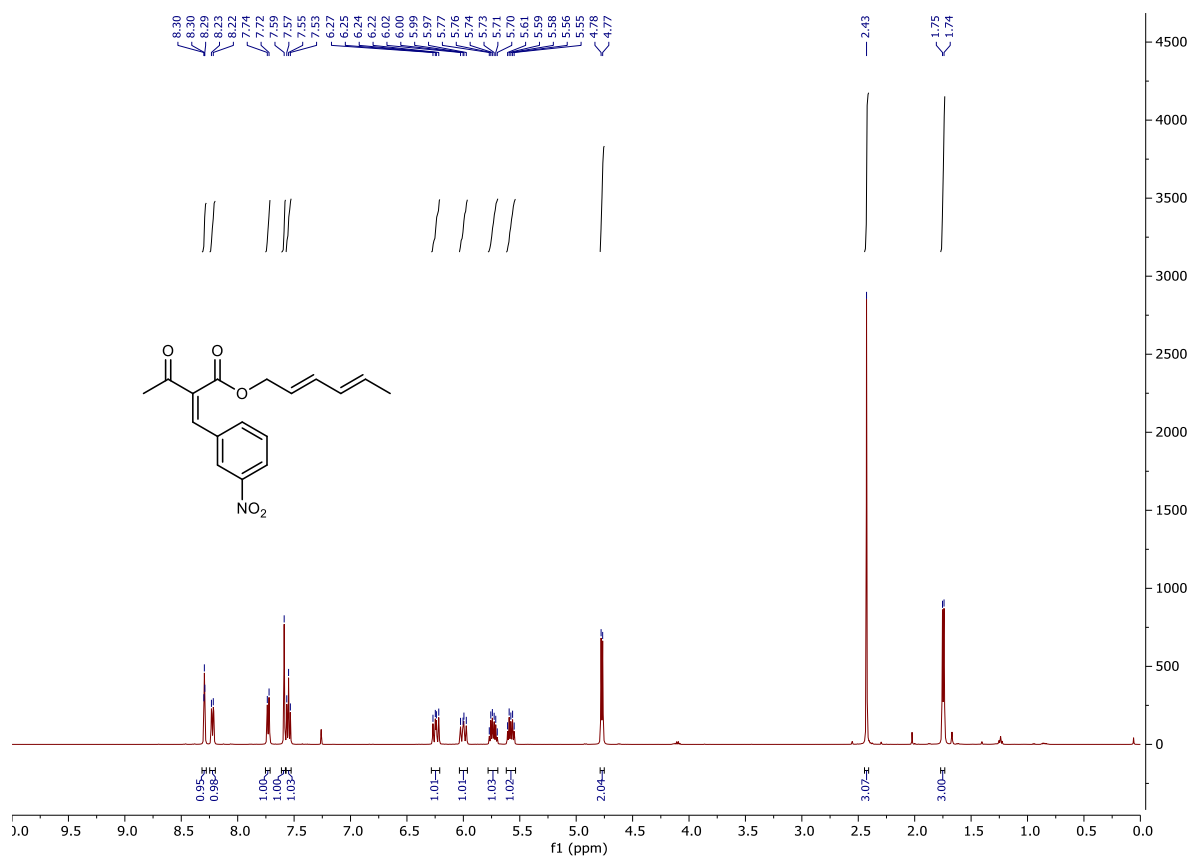
**Supplementary Figure S121.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6i'**



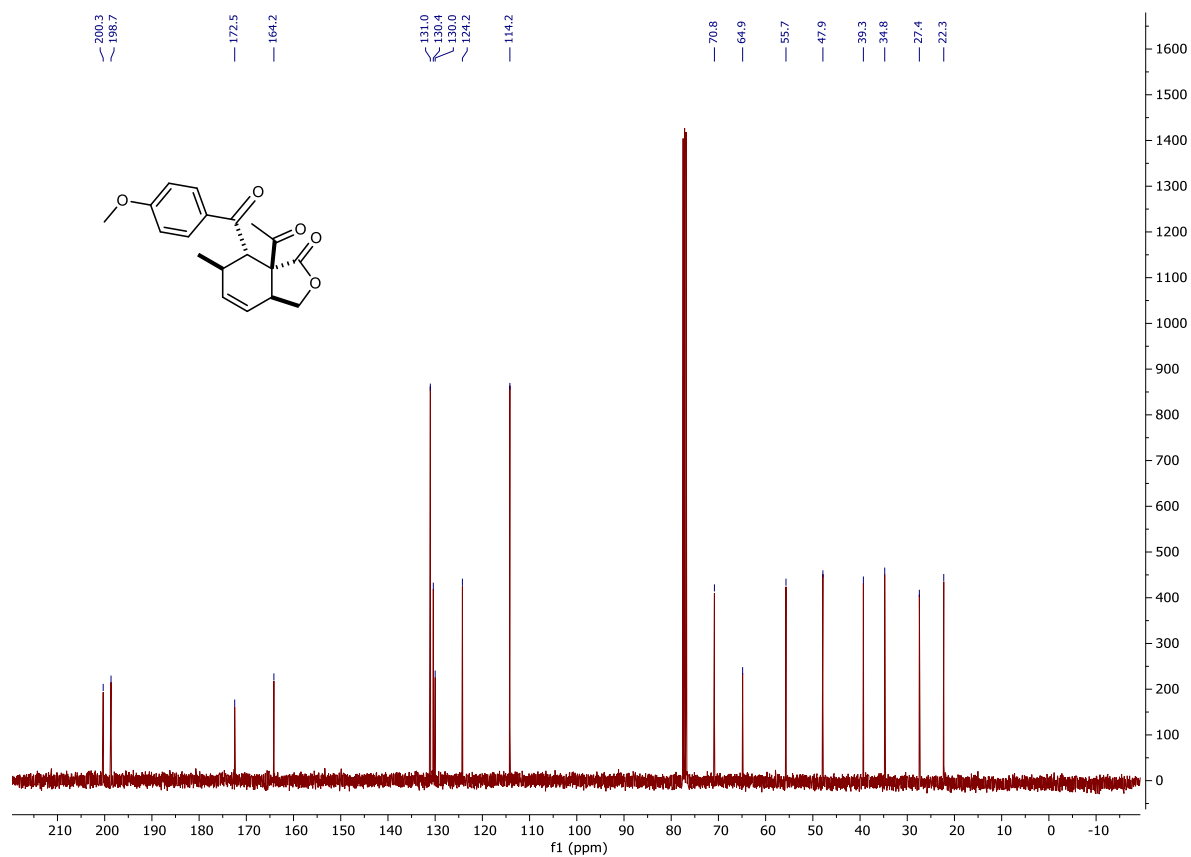
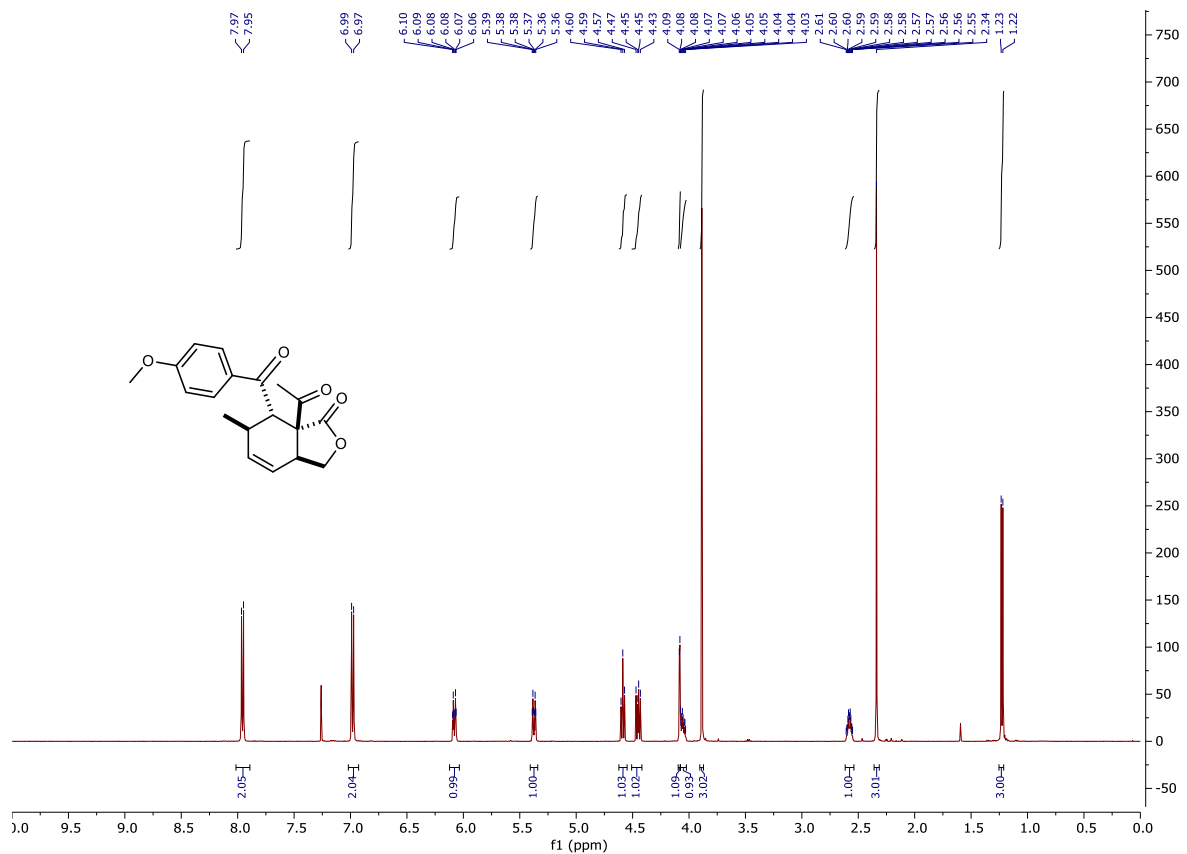
**Supplementary Figure S122.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **S6.7.1**.



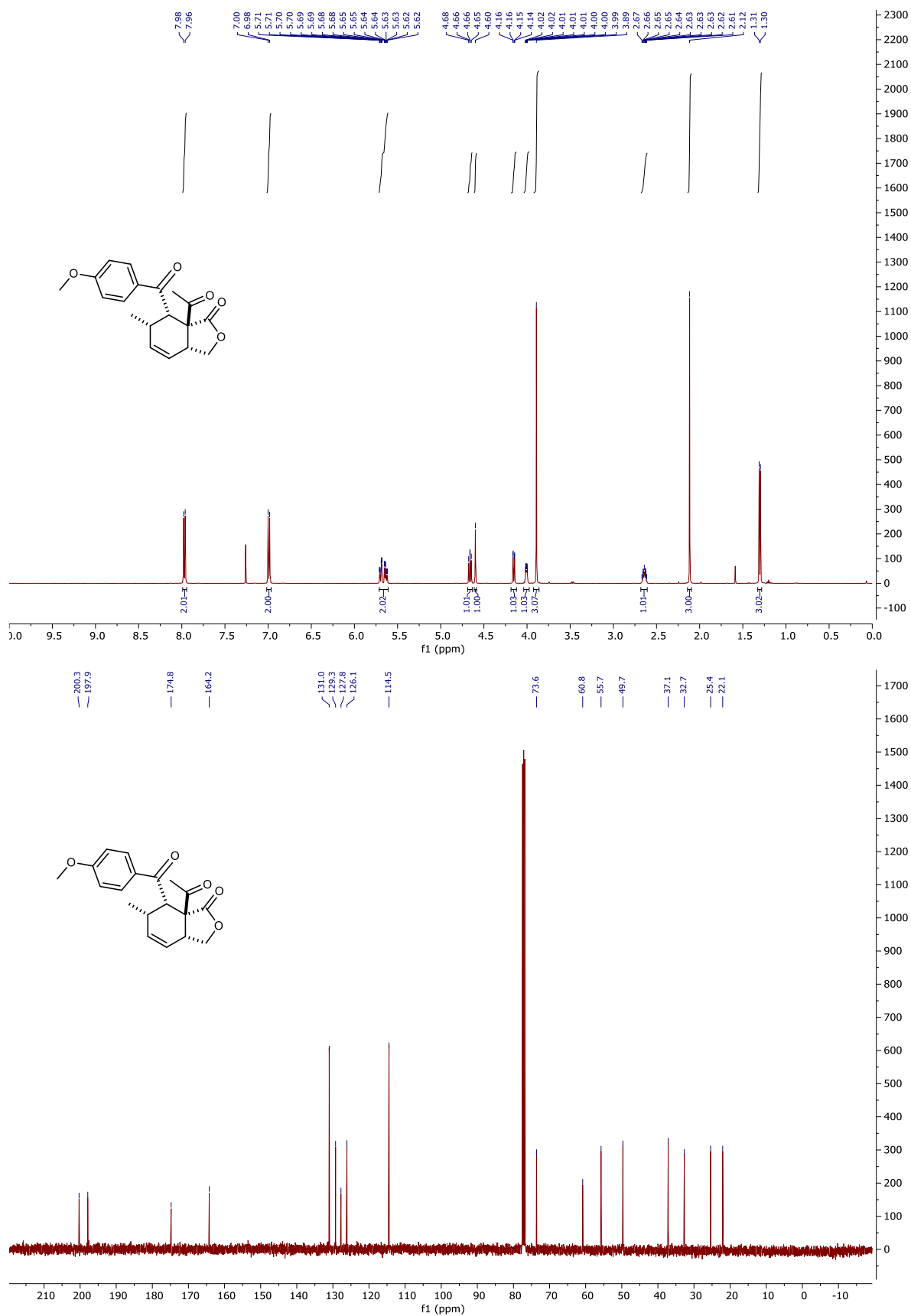
**Supplementary Figure S123.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.7.2.



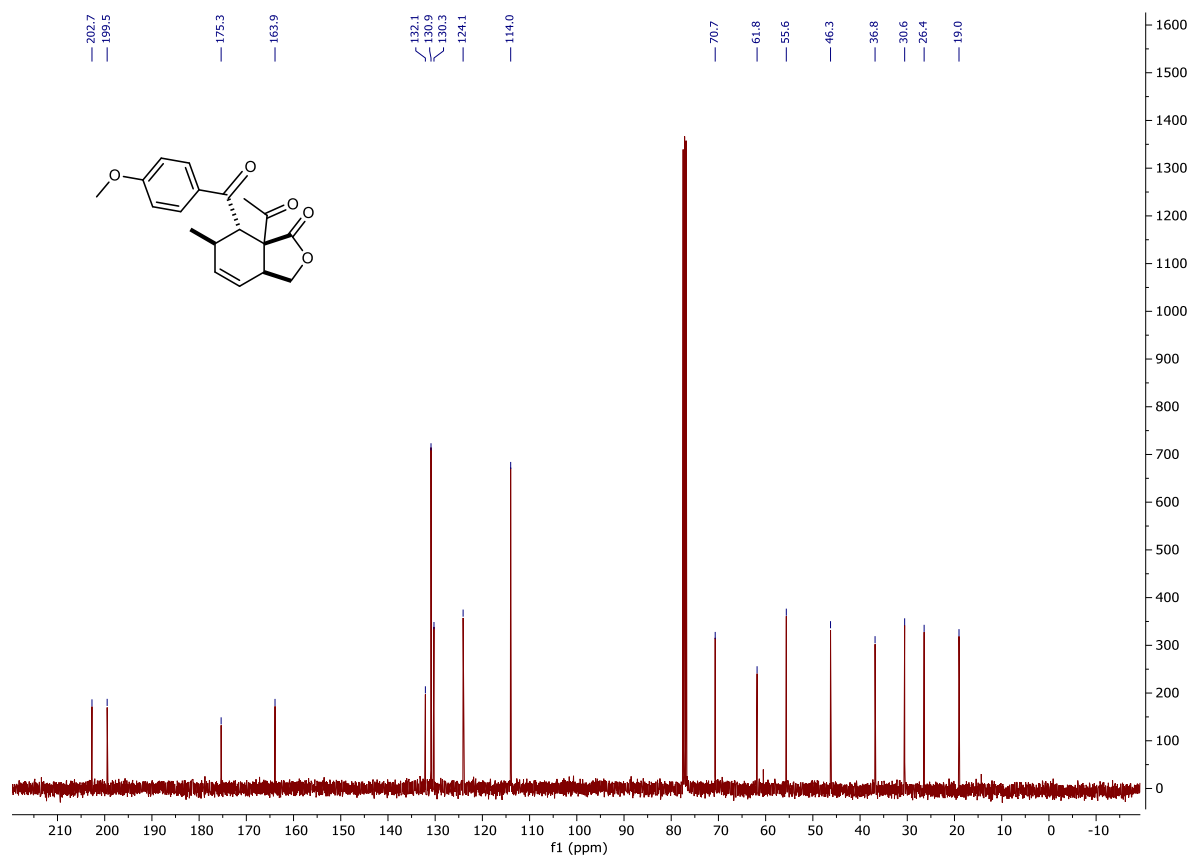
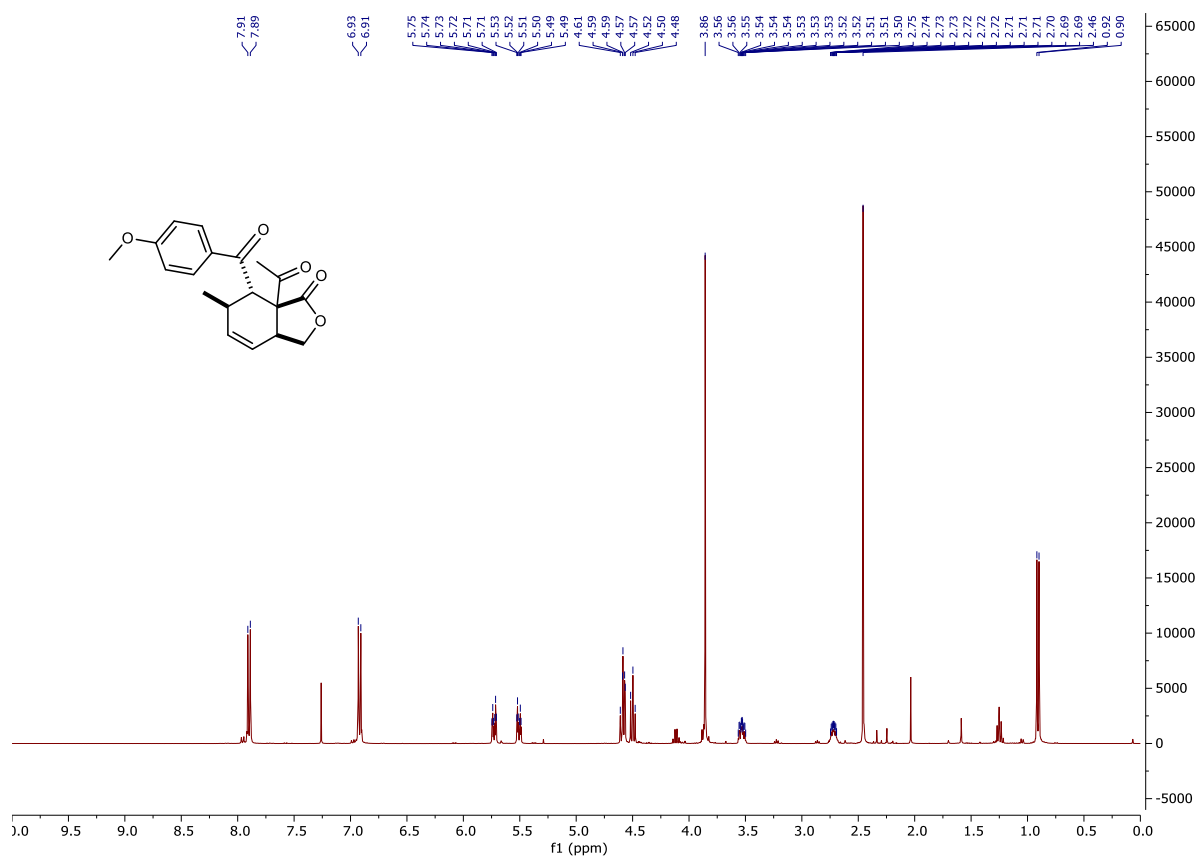
**Supplementary Figure S124.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **S6.7.3**.



**Supplementary Figure S125.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6j**.

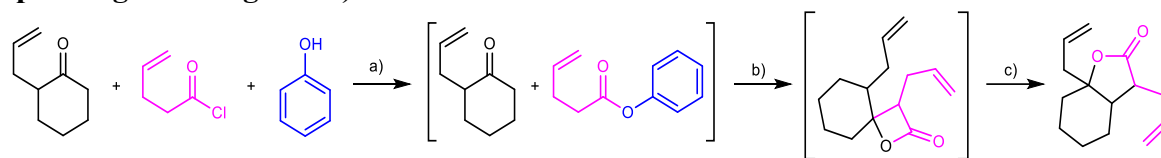


Supplementary Figure S126. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6j**.



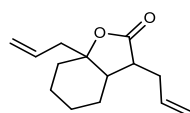
Supplementary Figure S127. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **6j2**.

**Section S6.8 Reaction Mach8 described in main-text Figure 5a-b (variant with the ring-expanding rearrangement).**



**Supplementary Figure S128. Reaction Mach8 described in main-text Figure 5a-b (variant with the ring-expanding rearrangement).** Reagents and conditions: (a) Pyridine, Et<sub>2</sub>O, 0 °C to rt, 1 h, then rt, 22 h, 98%; (b) NaHMDS, -78 °C, 30 min, then from -78 °C to -30 °C, 1.5 h, 60%; (c) MgBr<sub>2</sub>·Et<sub>2</sub>O, from -30 °C to rt, 3 h, then rt, 18 h, 31%

**One-pot procedure**



**3,7a-diallylhexahydrobenzofuran-2(3H)-one 7a.**

To a flask containing 2-allylcyclohexanone (0.146 mL, 1 mmol), phenol (0.104 g, 1.1 mmol, 1.1 equiv) and 4-pentenyl chloride (0.134 mL, 1.2 mmol, 1.2 equiv) was added anhydrous Et<sub>2</sub>O (5 mL) and the mixture was cooled to 0 °C. Anhydrous pyridine (0.162 mL, 2 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion of the ester **S6.8.1** formation the mixture was re-cooled to -78 °C and sodium bis(trimethylsilyl)amide solution (1.0 M in THF) (1.5 mL, 1.5 mmol, 1.5 equiv) was added quickly in one portion. The reaction mixture was stirred at -78 °C for 30 min and then was allowed to slowly warm to -30 °C over 1.5 h. Upon completion of the spiro-compound **S6.8.2** formation magnesium bromide ethyl etherate (1.291 g, 5 mmol, 5 equiv) was added in one portion at -30 °C. The mixture was allowed to warm to rt over 3 h and stirred at rt for 18 h. Upon completion reaction was diluted with Et<sub>2</sub>O (15 mL) and quenched with water (5 mL). The layers were separated and the aqueous phase was washed with Et<sub>2</sub>O (3 x 20 mL) and combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Crude product was purified by flash column chromatography (hexane/EtOAc 15:1) to give **7a** (68 mg, 31%) as a colorless oil.

**<sup>1</sup>H NMR** (400 MHz, C<sub>6</sub>D<sub>6</sub>) δ 5.73 – 5.58 (m, 1H), 5.54 (dddd, *J* = 16.7, 10.1, 8.3, 6.2 Hz, 1H), 5.00 – 4.81 (m, 4H), 2.68 – 2.50 (m, 2H), 2.17 – 1.95 (m, 2H), 1.88 – 1.78 (m, 2H), 1.72 (dt, *J* = 11.4, 5.8 Hz, 1H), 1.33 – 1.12 (m, 4H), 1.07 (ddd, *J* = 14.5, 11.7, 5.4 Hz, 1H), 0.80 – 0.62 (m, 2H);

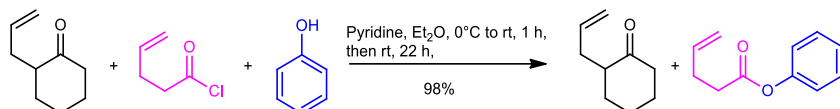


$^{13}\text{C}$  NMR (100 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  176.7, 136.3, 133.3, 118.7, 116.0, 83.0, 45.0, 43.9, 39.6, 32.5, 29.4, 24.2, 23.2, 20.9;

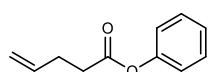
IR (film,  $\text{CH}_2\text{Cl}_2$ ) 3077, 3006, 2978, 2938, 2859, 1770, 1641, 1449, 1439  $\text{cm}^{-1}$ ;

HRMS (ESI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{14}\text{H}_{20}\text{O}_2\text{Na}$  243.1361; Found 243.1365.

### Mechanistic validation by stepwise isolation



**Supplementary Figure S129.** Reaction scheme of the stepwise process to confirm first step intermediate



### Phenyl pent-4-enoate **S6.8.1**<sup>60</sup>

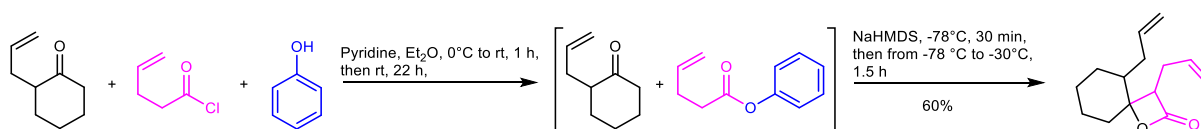
To a flask containing 2-allylcyclohexanone (0.073 mL, 0.5 mmol), phenol (0.052 g, 0.55 mmol, 1.1 equiv) and 4-pentenoyl chloride (0.067 mL, 0.6 mmol, 1.2 equiv) was added anhydrous  $\text{Et}_2\text{O}$  (2.5 mL) and the mixture was cooled to  $0^\circ\text{C}$ . Anhydrous pyridine (0.081 mL, 1 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion reaction was diluted with DCM (25 mL) and quenched with sat. aqueous solution of  $\text{NaHCO}_3$  (10 mL). The layers were separated and the organic phase was dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Crude product was purified by flash column chromatography (hexane/ $\text{EtOAc}$  20:1) to give **S6.8.1** (85 mg, 98%) as a colorless liquid.

$^1\text{H}$  NMR (500 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  7.10 – 6.99 (m, 4H), 6.94 – 6.87 (m, 1H), 5.74 – 5.63 (m, 1H), 5.03 – 4.88 (m, 2H), 2.27 – 2.23 (m, 4H);

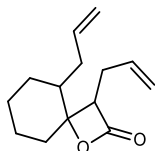
$^{13}\text{C}$  NMR (125 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  170.7, 151.6, 136.7, 129.5, 125.7, 122.0, 115.8, 33.7, 29.1;

IR (film,  $\text{CH}_2\text{Cl}_2$ ) 3077, 2980, 2922, 1761, 1641, 1594 1493  $\text{cm}^{-1}$ ;

HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{11}\text{H}_{12}\text{O}_2$  176.0837; Found 176.0829.



**Supplementary Figure S130.** Reaction scheme of the stepwise process to confirm the second step intermediate



### 3,5-diallyl-1-oxaspiro[3.5]nonan-2-one S6.8.2.

To a flask containing 2-allylcyclohexanone (0.073 mL, 0.5 mmol), phenol (0.052 g, 0.55 mmol, 1.1 equiv) and 4-pentenoyl chloride (0.067 mL, 0.6 mmol, 1.2 equiv) was added anhydrous Et<sub>2</sub>O (2.5 mL) and the mixture was cooled to 0 °C. Anhydrous pyridine (0.081 mL, 1 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion of the ester **S6.8.1** formation the mixture was re-cooled to -78 °C and sodium bis(trimethylsilyl)amide solution (1.0 M in THF) (0.75 mL, 0.75 mmol, 1.5 equiv) was added in one portion. The reaction mixture was stirred at -78 °C for 30 min and then was allowed to slowly warm to -30 °C over 1.5 h. Upon completion reaction was quenched at -30 °C by addition of NaOH (1 M in H<sub>2</sub>O) (1 mL). The mixture was diluted with Et<sub>2</sub>O (15 mL) and water (5 mL) and the layers were separated. Organic phase was washed with NaOH (1 M in H<sub>2</sub>O) (5 mL) and brine (5 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Crude product was purified by flash column chromatography (hexane/EtOAc 25:1) to give **S6.8.2** (65 mg, 60%) as a colorless oil.

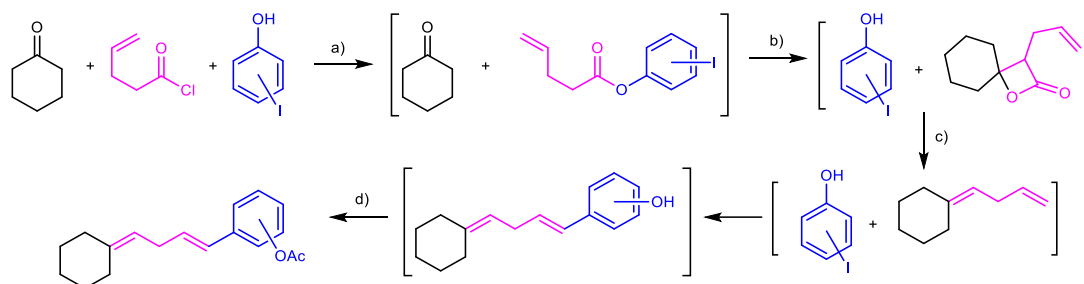
**<sup>1</sup>H NMR** (500 MHz, C<sub>6</sub>D<sub>6</sub>) δ 5.63 (dddd, *J* = 17.4, 10.3, 7.4, 5.5 Hz, 1H), 5.52 (dddd, *J* = 17.0, 10.2, 8.1, 5.9 Hz, 1H), 4.98 – 4.85 (m, 4H), 3.24 (dd, *J* = 8.7, 7.4 Hz, 1H), 2.28 – 2.18 (m, 2H), 2.07 – 1.98 (m, 1H), 1.90 – 1.78 (m, 1H), 1.64 – 1.49 (m, 1H), 1.42 – 1.34 (m, 3H), 1.28 – 1.16 (m, 4H), 0.96 – 0.87 (m, 1H);

**<sup>13</sup>C NMR** (125 MHz, C<sub>6</sub>D<sub>6</sub>) δ 170.1, 137.0, 135.1, 116.7, 116.6, 82.7, 54.1, 42.9, 32.8, 31.0, 28.5, 27.8, 23.6, 22.4;

**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3077, 2935, 2860, 1814, 1641, 1444 cm<sup>-1</sup>;

**HRMS** (APCI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>14</sub>H<sub>21</sub>O<sub>2</sub> 221.1542; Found 221.1543.

**Section S6.9 Reaction Mach9 described in main-text Figure 5a-b (variant with the application of Heck reaction).**



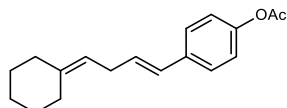
**Supplementary Figure S131. Reaction Mach9 described in main-text Figure 5a-b (variant with the application of Heck reaction).** Reagents and conditions: (a) Pyridine, THF, 0 °C to rt, 1 h, then rt, 22 h, 82% (4-iodophenol), 84% (3-iodophenol), 90% (2-iodophenol); (b) NaHMDS, -78 °C, 30 min, then from -78 °C to -30 °C, 1.5 h, 50% (4-iodophenol), 42% (3-iodophenol), 15% (2-iodophenol); (c) Pd(OAc)<sub>2</sub>, TBAC, Na<sub>2</sub>CO<sub>3</sub>, 100 °C, 24 h; (d) Ac<sub>2</sub>O, DMAP, Pyridine, rt, 18 h, 35% (4-iodophenol), 27% (3-iodophenol), 8% (2-iodophenol).

**One-pot procedure**

**General procedure:**

To a glass Schlenk vacuum ampule containing cyclohexanone (0.104 mL, 1 mmol), iodophenol (0.242 g, 1.1 mmol, 1.1 equiv) and 4-pentenoyl chloride (0.134 mL, 1.2 mmol, 1.2 equiv) was added anhydrous THF (5 mL) and the mixture was cooled to 0 °C. Anhydrous pyridine (0.162 mL, 2 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion of the ester **S6.9.1**, **S6.9.2** or **S6.9.3** formation the mixture was re-cooled to -78 °C and sodium bis(trimethylsilyl)amide solution (1.0 M in THF) (1.5 mL, 1.5 mmol, 1.5 equiv) was added in one portion. The reaction mixture was stirred at -78 °C for 30 min and then was allowed to slowly warm to -30 °C over 1.5 h. Upon completion of the spiro-compound **S6.9.4** formation, palladium(II) acetate (11 mg, 0.05 mmol, 5 mol%), tetrabutylammonium chloride (0.278 g, 1 mmol, 1 equiv) and sodium carbonate (0.530 g, 5 mmol, 5 equiv) were added in at -30 °C and the ampule was tightly closed with a Teflon plug valve. The reaction vessel was moved from the cooling bath at -30 °C directly to a heating block heated to 100 °C and the reaction mixture was stirred at 100 °C for 24 h. Upon completion reaction mixture was allowed to cool down to rt and was diluted with Et<sub>2</sub>O (25 mL) and was washed with sat. aqueous solution of NH<sub>4</sub>Cl (10 mL) followed by water (10 mL). Organic phase was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. **Due to immense problems with separation of the desired product from impurities during previous trials, the mixture was directly submitted**

**for acetylation to improve purification.** Crude product was dissolved in anhydrous pyridine (1.2 mL) and acetic anhydride (0.473 mL, 5 mmol, 5 equiv) and DMAP (12 mg, 0.1 mmol, 10 mol%) were added. The reaction mixture was stirred at rt for 18 h. Upon completion reaction was quenched with water (5 mL) and extracted with EtOAc (3 x 20 mL). Combined organic phases were washed with 1 M HCl (10 mL), sat. aqueous solution of NaHCO<sub>3</sub> (10 mL), water (10 mL) and brine (10 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Crude product was purified by flash column chromatography.



#### 4-(4-cyclohexylidenebut-1-en-1-yl)phenyl acetate **7b 4-OAc**.

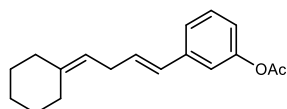
According to general procedure, the reaction was performed with 4-iodophenol. After purification by flash column chromatography (hexane/EtOAc 95:5) compound **7b 4-OAc** was obtained (94 mg, 35%) as a yellow oil.

**<sup>1</sup>H NMR** (400 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.15 (d, *J* = 6.4 Hz, 2H), 7.01 – 6.96 (m, 2H), 6.32 (dt, *J* = 15.8, 1.8 Hz, 1H), 6.04 (dt, *J* = 15.8, 6.4 Hz, 1H), 5.24 (tt, *J* = 7.3, 1.3 Hz, 1H), 2.85 (d, *J* = 5.8 Hz, 1H), 2.83 (d, *J* = 5.8 Hz, 1H), 2.14 – 2.07 (m, 4H), 1.76 (s, 3H), 1.52 – 1.42 (m, 6H);

**<sup>13</sup>C NMR** (100 MHz, C<sub>6</sub>D<sub>6</sub>) δ 168.4, 150.4, 141.1, 135.9, 130.1, 129.3, 127.2, 122.0, 118.8, 37.6, 31.0, 29.0, 29.0, 28.2, 27.2, 20.6;

**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3030, 2926, 2852, 1762, 1505 cm<sup>-1</sup>;

**HRMS** (EI) *m/z*: [M]<sup>+</sup> Calcd for C<sub>18</sub>H<sub>22</sub>O<sub>2</sub> 270.1620; Found 270.1631.



#### 3-(4-cyclohexylidenebut-1-en-1-yl)phenyl acetate **7b 3-OAc**.

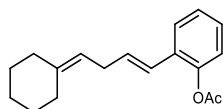
According to general procedure, the reaction was performed with 3-iodophenol. After purification by flash column chromatography (hexane/EtOAc 95:5) compound **7b 3-OAc** was obtained (72 mg, 27%) as a slightly yellow oil.

**<sup>1</sup>H NMR** (500 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.18 – 7.16 (m, 1H), 7.04 – 7.00 (m, 2H), 6.95 – 6.90 (m, 1H), 6.32 (dt, *J* = 15.8, 1.7 Hz, 1H), 6.08 (dt, *J* = 15.8, 6.4 Hz, 1H), 5.20 (tt, *J* = 7.3, 1.3 Hz, 1H), 2.81 (d, *J* = 6.9 Hz, 1H), 2.79 (d, *J* = 6.9 Hz, 1H), 2.13 – 2.02 (m, 4H), 1.77 (s, 3H), 1.55 – 1.37 (m, 6H);

$^{13}\text{C}$  NMR (125 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  168.0, 151.5, 140.8, 139.6, 130.7, 129.1, 128.9, 123.3, 120.0, 119.1, 118.2, 37.1, 30.6, 28.6, 28.6, 27.8, 26.8, 20.2;

IR (film,  $\text{CH}_2\text{Cl}_2$ ) 2926, 2852, 1766, 1605, 1580, 1486, 1444  $\text{cm}^{-1}$ ;

HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{18}\text{H}_{22}\text{O}_2$  270.1620; Found 270.1626.



### 2-(4-cyclohexylidenebut-1-en-1-yl)phenyl acetate **7b** 2-OAc.

According to general procedure, the reaction was performed with 2-iodophenol. After purification by flash column chromatography (hexane/EtOAc 25:1) compound **7b** 2-OAc was obtained (23 mg, 9%) as a colorless oil.

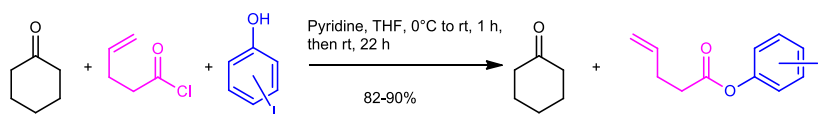
$^1\text{H}$  NMR (600 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  7.37 (dd,  $J = 7.7, 1.8$  Hz, 1H), 7.02 – 7.00 (m, 1H), 6.97 (td,  $J = 7.7, 1.8$  Hz, 1H), 6.95 – 6.91 (m, 1H), 6.62 (dt,  $J = 15.8, 1.8$  Hz, 1H), 6.12 (dt,  $J = 15.8, 6.4$  Hz, 1H), 5.19 (tt,  $J = 7.3, 1.2$  Hz, 1H), 2.86 – 2.81 (m, 2H), 2.10 – 2.04 (m, 4H), 1.84 (s, 3H), 1.52 – 1.47 (m, 2H), 1.45 – 1.41 (m, 4H);

$^{13}\text{C}$  NMR (150 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  168.3, 148.6, 141.4, 132.4, 131.1, 128.3, 126.9, 126.1, 123.7, 123.1, 118.5, 37.5, 31.3, 29.0, 29.0, 28.1, 27.2, 20.4;

IR (film,  $\text{CH}_2\text{Cl}_2$ ) 2926, 2852, 1766, 1484, 1447  $\text{cm}^{-1}$ ;

HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{18}\text{H}_{22}\text{O}_2$  270.1620; Found 270.1618.

### Mechanistic validation by stepwise isolation

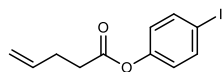


**Supplementary Figure S132.** Reaction scheme of the stepwise process to confirm the first step intermediate

### General procedure:

To a flask containing cyclohexanone (0.052 mL, 0.5 mmol), iodophenol (0.121 g, 0.55 mmol, 1.1 equiv) and 4-pentenoyl chloride (0.067 mL, 0.6 mmol, 1.2 equiv) was added anhydrous THF (2.5 mL) and the mixture was cooled to  $0^\circ\text{C}$ . Anhydrous pyridine (0.081 mL, 1 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion reaction was diluted with DCM (25 mL) and quenched with  $\text{NaHCO}_3$

(10 mL). The layers were separated and the organic phase was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Crude product was purified by flash column chromatography.



#### 4-iodophenyl pent-4-enoate **S6.9.1**.

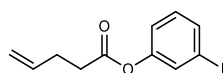
According to general procedure, the reaction was performed with 4-iodophenol. After purification by flash column chromatography (hexane/EtOAc 30:1) compound **S6.9.1** was obtained (125 mg, 82%) as a slightly yellow liquid.

<sup>1</sup>H NMR (400 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.35 – 7.30 (m, 2H), 6.59 – 6.52 (m, 2H), 5.73 – 5.56 (m, 1H), 5.01 – 4.89 (m, 2H), 2.21 – 2.18 (m, 4H);

<sup>13</sup>C NMR (100 MHz, C<sub>6</sub>D<sub>6</sub>) δ 170.3, 151.1, 138.7, 136.5, 124.0, 115.9, 89.7, 33.5, 29.0;

IR (film, CH<sub>2</sub>Cl<sub>2</sub>) 3504, 3079, 2979, 2919, 2852, 1761, 1641, 1580, 1482 cm<sup>-1</sup>;

HRMS (APCI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>11</sub>H<sub>12</sub>O<sub>2</sub>I 302.9882; Found 302.9880.



#### 3-iodophenyl pent-4-enoate **S6.9.2**.

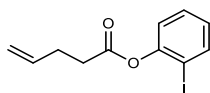
According to general procedure, the reaction was performed with 3-iodophenol. After purification by flash column chromatography (hexane/EtOAc 30:1) compound **S6.9.2** was obtained (128 mg, 84%) as a slightly yellow liquid.

<sup>1</sup>H NMR (400 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.45 (dd, *J* = 2.2, 1.6 Hz, 1H), 7.24 (ddd, *J* = 8.1, 1.6, 0.9 Hz, 1H), 6.85 (ddd, *J* = 8.1, 2.2, 0.9 Hz, 1H), 6.53 (t, *J* = 8.1 Hz, 1H), 5.71 – 5.53 (m, 1H), 5.00 – 4.85 (m, 2H), 2.19 – 2.17 (m, 4H);

<sup>13</sup>C NMR (100 MHz, C<sub>6</sub>D<sub>6</sub>) δ 170.3, 151.7, 136.5, 134.9, 131.1, 130.8, 121.4, 115.9, 93.8, 33.5, 29.0;

IR (film, CH<sub>2</sub>Cl<sub>2</sub>) 3076, 2978, 2920, 2853, 1761, 1641, 1580, 1466, 1420 cm<sup>-1</sup>;

HRMS (APCI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>11</sub>H<sub>12</sub>O<sub>2</sub>I 302.9882; Found 302.9883.



### 2-iodophenyl pent-4-enoate **S6.9.3**

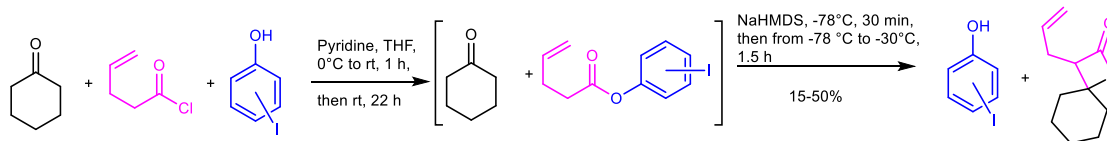
According to general procedure, the reaction was performed with 2-iodophenol. After purification by flash column chromatography (hexane/EtOAc 20:1) compound **S6.9.3** was obtained (136 mg, 90%) as a slightly yellow liquid.

**<sup>1</sup>H NMR** (400 MHz, C<sub>6</sub>D<sub>6</sub>) δ 7.51 (dd, *J* = 7.9, 1.5 Hz, 1H), 6.91 – 6.81 (m, 2H), 6.41 (ddd, *J* = 7.9, 7.1, 1.8 Hz, 1H), 5.73 (ddt, *J* = 16.7, 10.2, 6.4 Hz, 1H), 5.03 – 4.89 (m, 2H), 2.45 – 2.40 (m, 2H), 2.36 – 2.30 (m, 2H);

**<sup>13</sup>C NMR** (100 MHz, C<sub>6</sub>D<sub>6</sub>) δ 170.0, 151.9, 139.6, 136.7, 129.4, 127.5, 123.5, 115.9, 91.0, 33.8, 29.0;

**IR** (film, CH<sub>2</sub>Cl<sub>2</sub>) 3511, 3077, 2979, 2921, 2854, 1768, 1641, 1577, 1466, 1440, 1416 cm<sup>-1</sup>;

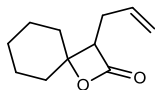
**HRMS** (APCI) *m/z*: [M+H]<sup>+</sup> Calcd for C<sub>11</sub>H<sub>12</sub>O<sub>2</sub>I 302.9882; Found 302.9885.



**Supplementary Figure S133.** Reaction scheme of the stepwise process to confirm the second step intermediate

#### General procedure:

To a flask containing cyclohexanone (0.052 mL, 0.5 mmol), iodophenol (0.121 g, 0.55 mmol, 1.1 equiv) and 4-pentenoyl chloride (0.067 mL, 0.6 mmol, 1.2 equiv) was added anhydrous THF (2.5 mL) and the mixture was cooled to 0 °C. Anhydrous pyridine (0.081 mL, 1 mmol, 2 equiv) was added dropwise and the mixture was allowed to warm to rt over 1 h and stirred at rt for 22 h. Upon completion of the ester **S6.9.1**, **S6.9.2** or **S6.9.3** formation the mixture was re-cooled to -78°C and sodium bis(trimethylsilyl)amide solution (1.0 M in THF) (0.75 mL, 0.75 mmol, 1.5 equiv) was added in one portion. The reaction mixture was stirred at -78°C for 30 min and then was allowed to slowly warm to -30 °C over 1.5 h. Upon completion reaction was quenched at -30°C by addition of NaOH (1 M in H<sub>2</sub>O) (1 mL). The mixture was diluted with Et<sub>2</sub>O (15 mL) and water (5 mL) and the layers were separated. Organic phase was washed with NaOH (1 M in H<sub>2</sub>O) (5 mL) and brine (5 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Crude product was purified by flash column chromatography.



### 3-allyl-1-oxaspiro[3.5]nonan-2-one S6.9.4

According to general procedure, the reaction was performed with *para*-, *meta*- or *ortho*-iodophenol. After purification by flash column chromatography (hexane/EtOAc 19:1) compound **S6.9.4** was obtained (45 mg, 50% (4-iodophenol)), (38 mg, 42% (3-iodophenol)) and (14 mg, 15% (2-iodophenol)) as a colorless oil.

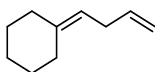
$^1\text{H NMR}$  (400 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  5.58 (dddd,  $J = 17.4, 10.3, 7.2, 5.5$  Hz, 1H), 4.97 – 4.81 (m, 2H), 2.71 (dd,  $J = 9.2, 7.2$  Hz, 1H), 2.27 – 2.16 (m, 1H), 2.07 – 1.91 (m, 1H), 1.48 – 1.38 (m, 4H), 1.32 – 1.06 (m, 5H), 0.96 – 0.85 (m, 1H);

$^{13}\text{C NMR}$  (100 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  170.2, 135.0, 116.6, 81.0, 57.5, 37.3, 31.1, 28.2, 25.1, 23.0, 22.3;

**IR** (film,  $\text{CH}_2\text{Cl}_2$ ) 3081, 2936, 2861, 1816, 1643, 1449  $\text{cm}^{-1}$ ;

**HRMS** (APCI)  $m/z$ :  $[\text{M}+\text{H}]^+$  Calcd for  $\text{C}_{11}\text{H}_{17}\text{O}_2$  181.1229; Found 181.1230.

### Additional reactivity investigations



### but-3-en-1-ylidenecyclohexane S6.9.5

Reaction below was performed to provide additional evidence that decarboxylation of **S6.9.4** reaction occurs in higher temperature in THF. Due to high volatility of the product, the reaction was performed in THF- $d_8$  to enable direct recording of NMR spectra of **S6.9.5**.

To a glass Schlenk vacuum ampule containing spiro-ester **S6.9.4** (0.045 g, 0.25 mmol) was added THF- $d_8$  and the ampule was tightly closed with a Teflon plug valve. The reaction vessel was placed in a heating block and the mixture was stirred at 100 °C for 18 h. Upon completion reaction mixture was allowed to cool down to rt and analytical data were recorded.

$^1\text{H NMR}$  (400 MHz, THF- $d_8$ )  $\delta$  5.82 – 5.70 (m, 1H), 5.12 – 5.04 (m, 1H), 4.98 (dq,  $J = 17.1, 1.8$  Hz, 1H), 4.89 (dq,  $J = 10.1, 1.8$  Hz, 1H), 2.73 (t,  $J = 6.9$  Hz, 2H), 2.18 – 2.04 (m, 4H), 1.59 – 1.44 (m, 6H);

$^{13}\text{C NMR}$  (100 MHz, THF- $d_8$ )  $\delta$  141.5, 138.7, 119.3, 114.4, 38.1, 32.3, 29.8, 29.5, 28.8, 28.0;

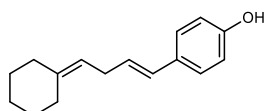
**IR** (film,  $\text{CH}_2\text{Cl}_2$ ) 3079, 2928, 2854, 1764, 1638, 1446  $\text{cm}^{-1}$ ;

**HRMS** (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{10}\text{H}_{16}$  136.1252; Found 136.1246.



### General procedure:

To a flask containing acetylated compounds **7b 4-OAc**, **7b 3-OAc**, **7b 2-OAc** were added THF/MeOH (1:1) ( $c = 0.05$ ) and potassium carbonate (4 equiv). The mixture was stirred at rt for 2 h. Upon completion reaction was diluted with Et<sub>2</sub>O (15 mL) and washed with 1 M HCl (5 mL), sat. aqueous solution of NaHCO<sub>3</sub> (5 mL), water (5 mL) and brine (5 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated *in vacuo*. Crude product was purified by flash column chromatography.



### 4-(4-cyclohexylidenebut-1-en-1-yl)phenol **S6.9.6**.

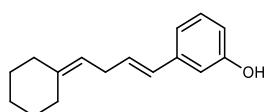
According to general procedure, the reaction was performed with **7b 4-OAc** (18 mg, 0.07 mmol). After purification by flash column chromatography (hexane/EtOAc 9:1) compound **S6.9.6** was obtained (13 mg, 86%) as a slightly yellow oil.

<sup>1</sup>H NMR (500 MHz, C<sub>6</sub>D<sub>6</sub>)  $\delta$  7.15 – 7.10 (m, 2H), 6.52 – 6.47 (m, 2H), 6.37 (dt,  $J = 15.8, 1.7$  Hz, 1H), 6.02 (dt,  $J = 15.8, 6.5$  Hz, 1H), 5.26 (tt,  $J = 7.3, 1.3$  Hz, 1H), 4.18 (brs, 1H), 2.89 – 2.84 (m, 2H), 2.14 – 2.08 (m, 4H), 1.53 – 1.42 (m, 6H);

<sup>13</sup>C NMR (125 MHz, C<sub>6</sub>D<sub>6</sub>)  $\delta$  155.5, 140.7, 131.1, 129.7, 127.7, 127.5, 119.3, 115.7, 37.6, 31.1, 29.0, 29.0, 28.2, 27.2;

IR (film, CH<sub>2</sub>Cl<sub>2</sub>) 3338, 3022, 2925, 2851, 1609, 1594, 1511, 1444 cm<sup>-1</sup>;

HRMS (EI)  $m/z$ : [M]<sup>+</sup> Calcd for C<sub>16</sub>H<sub>20</sub>O 228.1514; Found 228.1517.



### 3-(4-cyclohexylidenebut-1-en-1-yl)phenol **S6.9.7**.

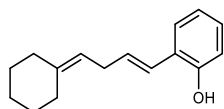
According to general procedure, the reaction was performed with **7b 3-OAc** (12 mg, 0.04 mmol). After purification by flash column chromatography (hexane/EtOAc 9:1) compound **S6.9.7** was obtained (9 mg, 89%) as a slightly yellow oil.

<sup>1</sup>H NMR (500 MHz, C<sub>6</sub>D<sub>6</sub>)  $\delta$  6.99 (t,  $J = 7.7$  Hz, 1H), 6.88 – 6.84 (m, 1H), 6.64 – 6.60 (m, 1H), 6.45 – 6.41 (m, 1H), 6.35 (dd,  $J = 15.7, 1.7$  Hz, 1H), 6.11 (dt,  $J = 15.7, 6.5$  Hz, 1H), 5.25 – 5.21 (m, 1H), 3.96 (brs, 1H), 2.86 – 2.81 (m, 2H), 2.13 – 2.08 (m, 4H), 1.52 – 1.43 (m, 6H);

$^{13}\text{C}$  NMR (125 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  156.6, 141.1, 140.1, 130.2, 130.0, 129.8, 119.0, 118.9, 114.2, 113.1, 37.5, 31.0, 29.0, 29.0, 28.2, 27.2;

IR (film,  $\text{CH}_2\text{Cl}_2$ ) 3342, 3025, 2925, 2852, 1606, 1583, 1492, 1447  $\text{cm}^{-1}$ ;

HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{16}\text{H}_{20}\text{O}$  228.1514; Found 228.1511.



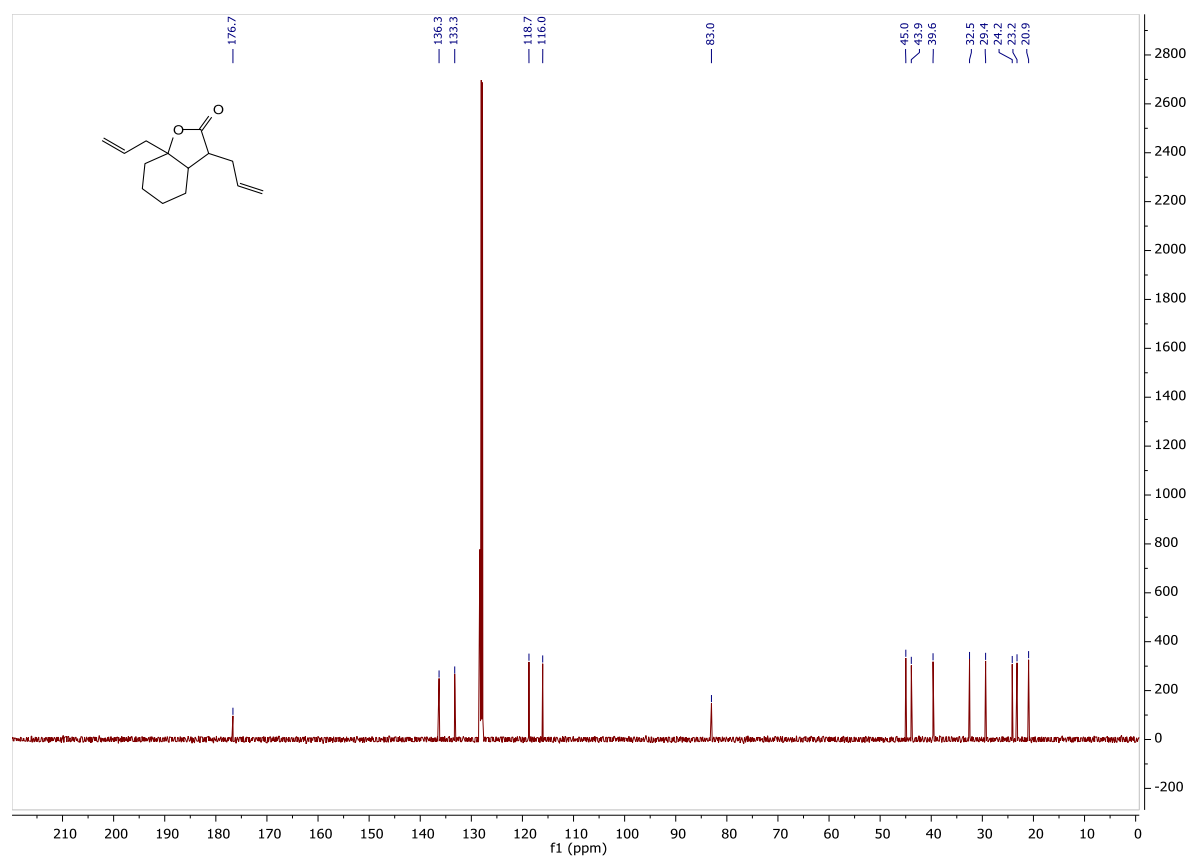
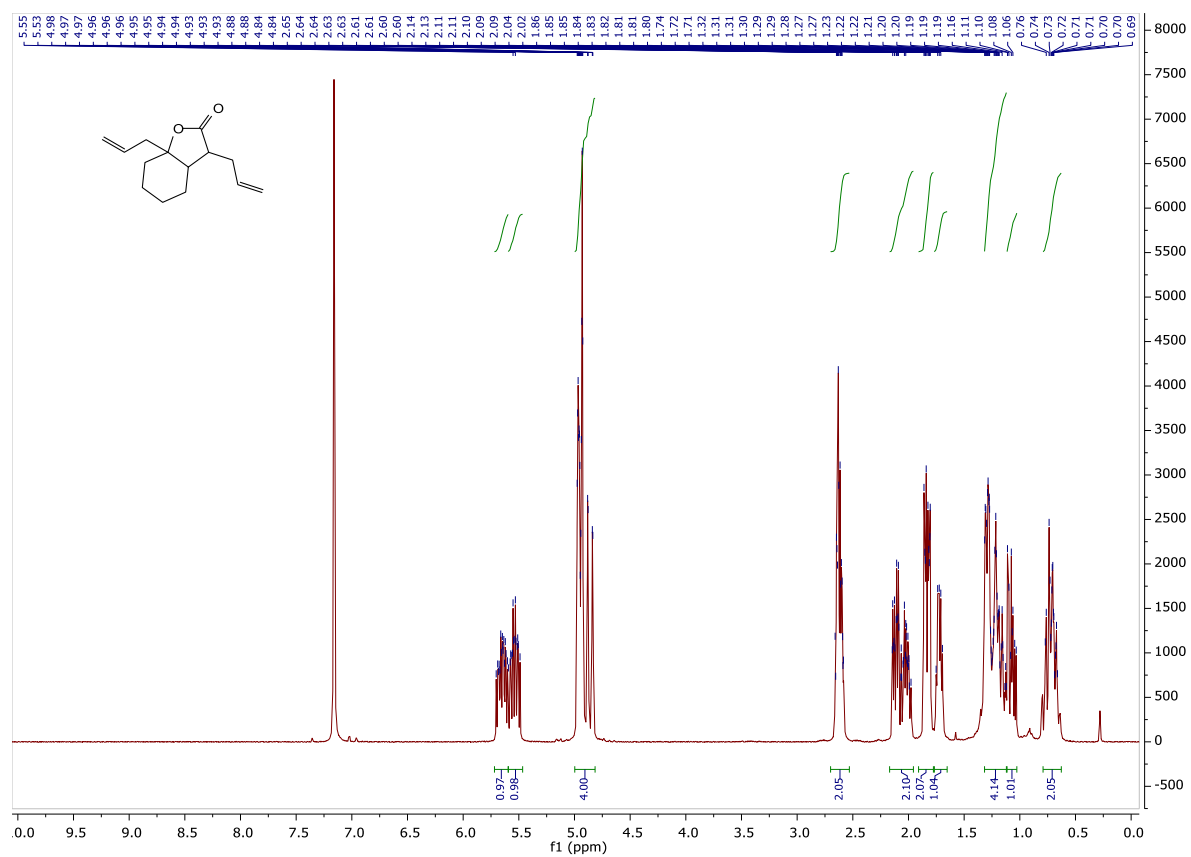
### 2-(4-cyclohexylidenebut-1-en-1-yl)phenol S6.9.8

According to general procedure, the reaction was performed with **7b 2-OAc** (20 mg, 0.07 mmol). After purification by flash column chromatography (hexane/EtOAc 95:5) compound **S6.9.8** was obtained (15 mg, 89%) as a colorless oil.

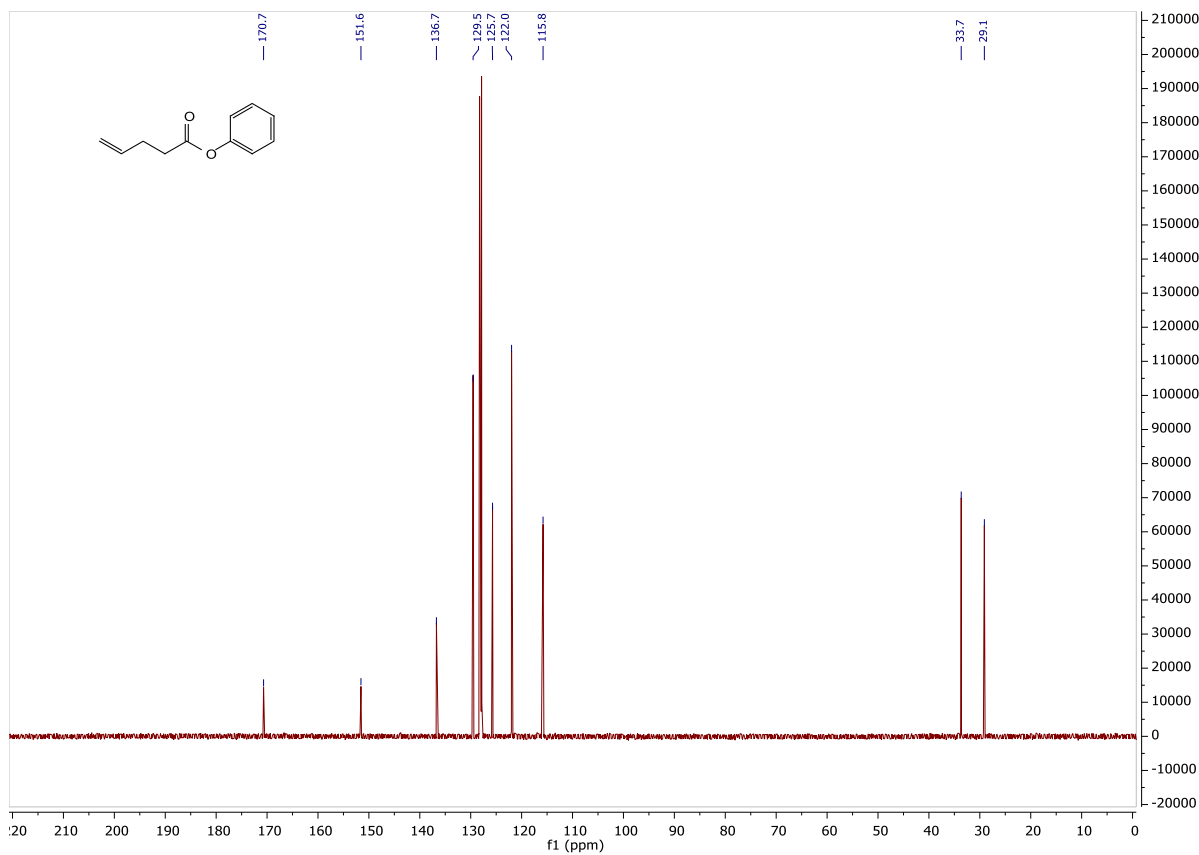
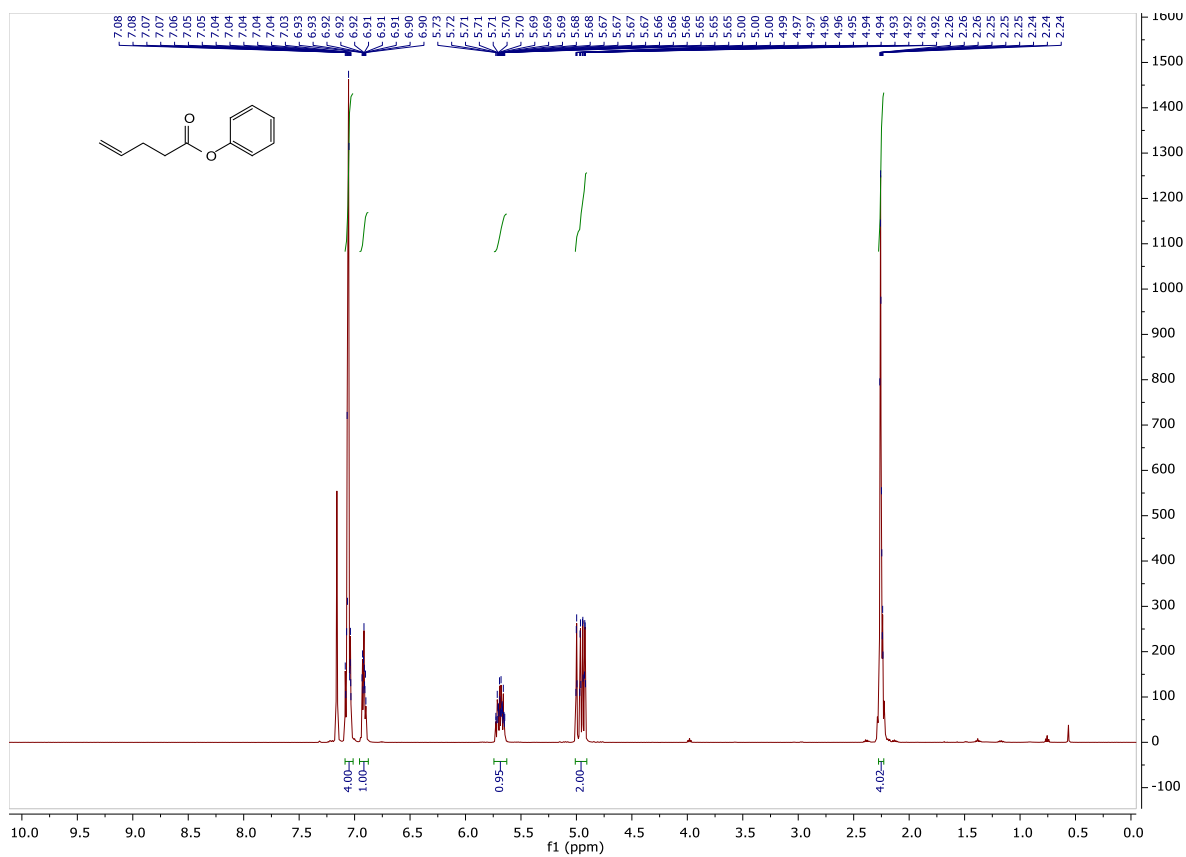
$^1\text{H}$  NMR (500 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  7.32 (dd,  $J = 7.7, 1.7$  Hz, 1H), 6.92 (td,  $J = 7.7, 1.7$  Hz, 1H), 6.80 – 6.74 (m, 2H), 6.33 – 6.29 (m, 1H), 6.17 (dt,  $J = 15.9, 6.5$  Hz, 1H), 5.22 (t,  $J = 7.3$  Hz, 1H), 4.22 (s, 1H), 2.88 – 2.83 (m, 2H), 2.12 – 2.04 (m, 4H), 1.53 – 1.40 (m, 6H);

$^{13}\text{C}$  NMR (125 MHz,  $\text{C}_6\text{D}_6$ )  $\delta$  153.2, 140.9, 131.4, 128.4, 127.6, 125.6, 124.7, 121.0, 119.0, 115.9, 37.5, 31.5, 29.0, 29.0, 28.1, 27.2;

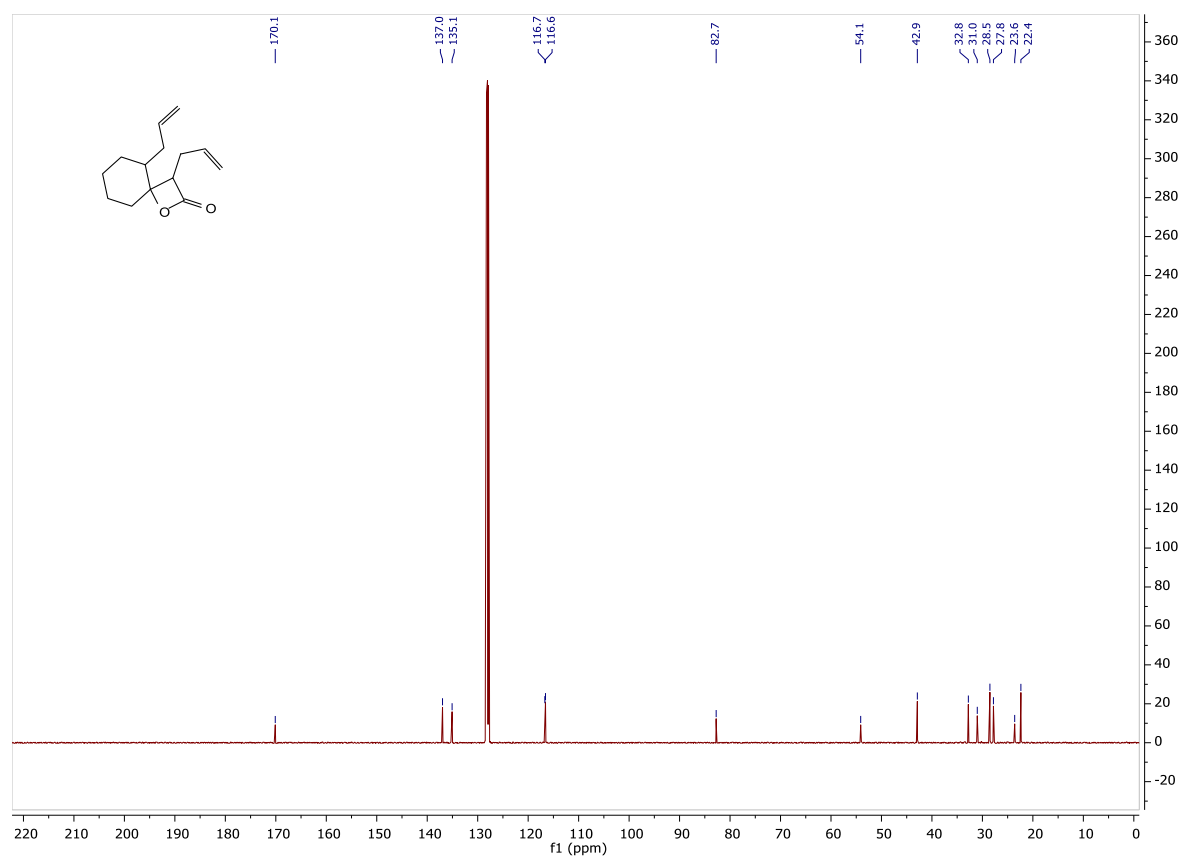
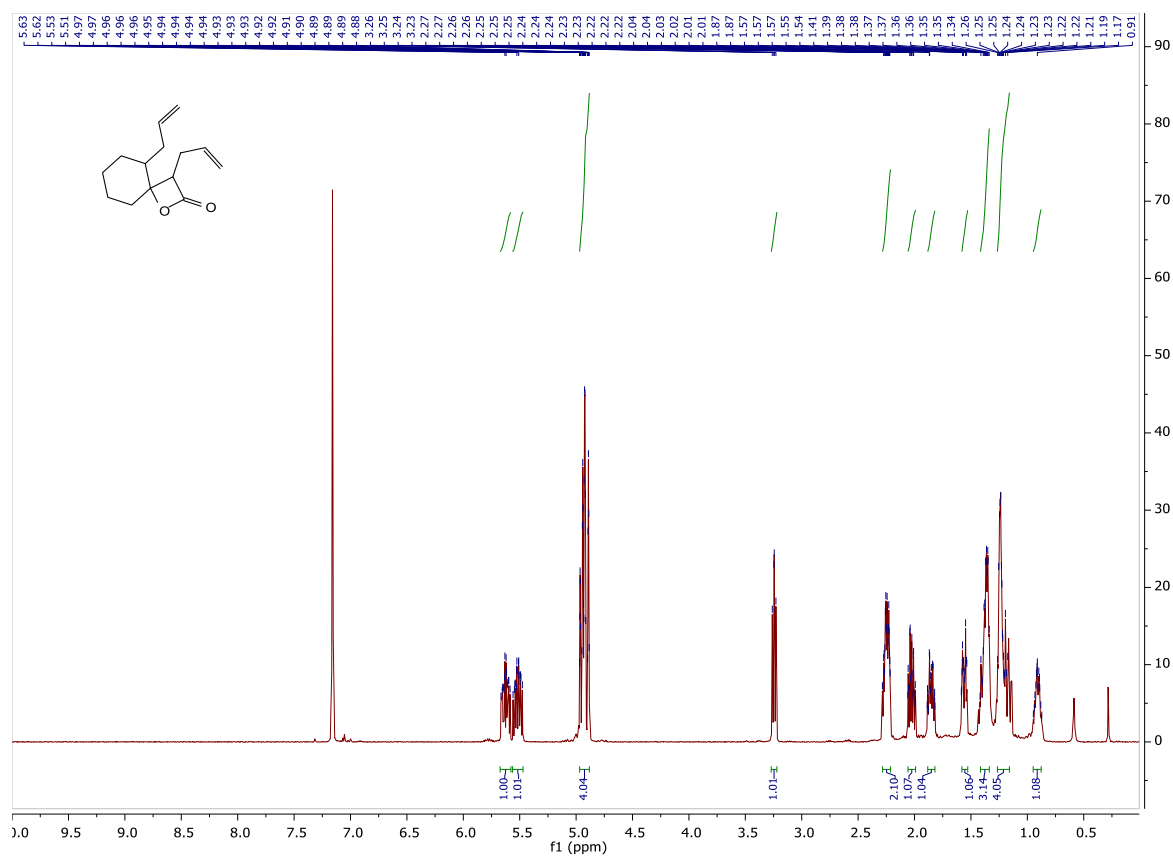
IR (film,  $\text{CH}_2\text{Cl}_2$ ) 3425, 2925, 2852, 1604, 1584, 1497, 1485, 1453  $\text{cm}^{-1}$ ; HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_{16}\text{H}_{20}\text{O}$  228.1514; Found 228.1519.



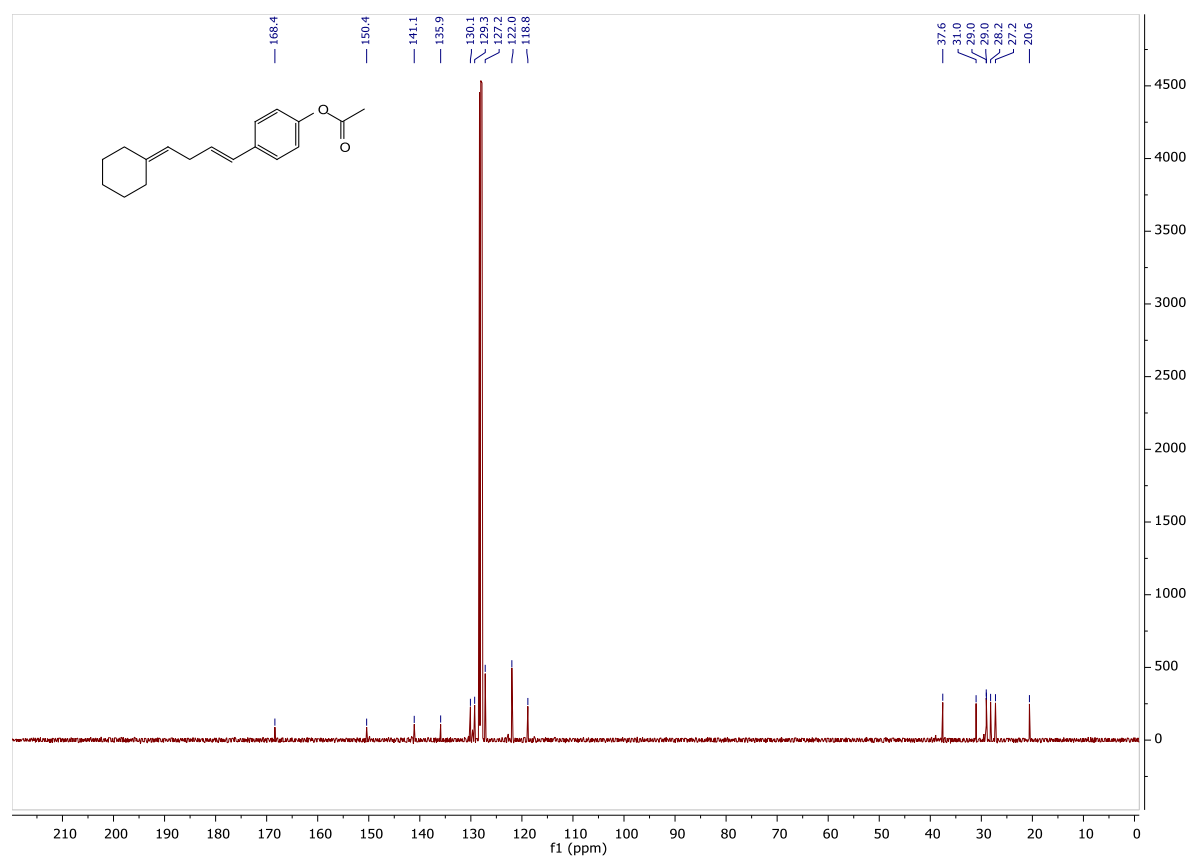
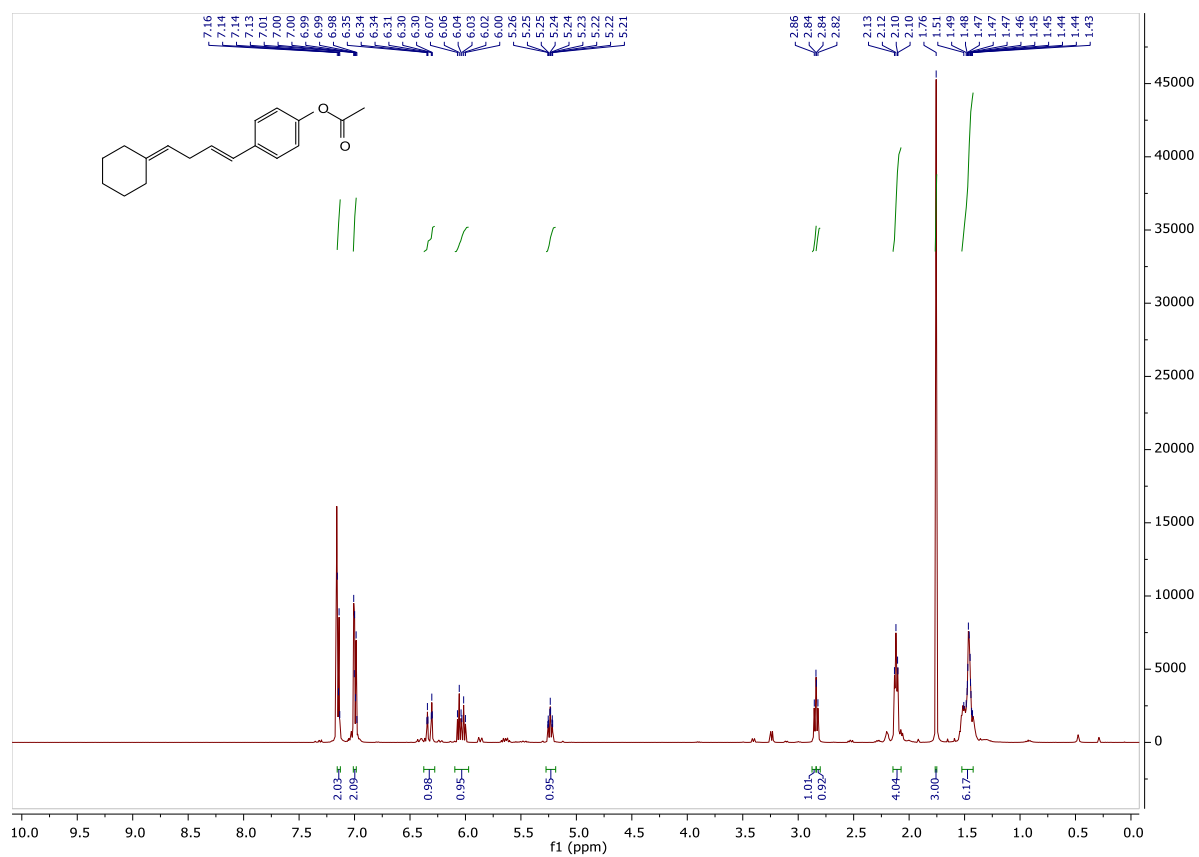
**Supplementary Figure S134.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **7a**.



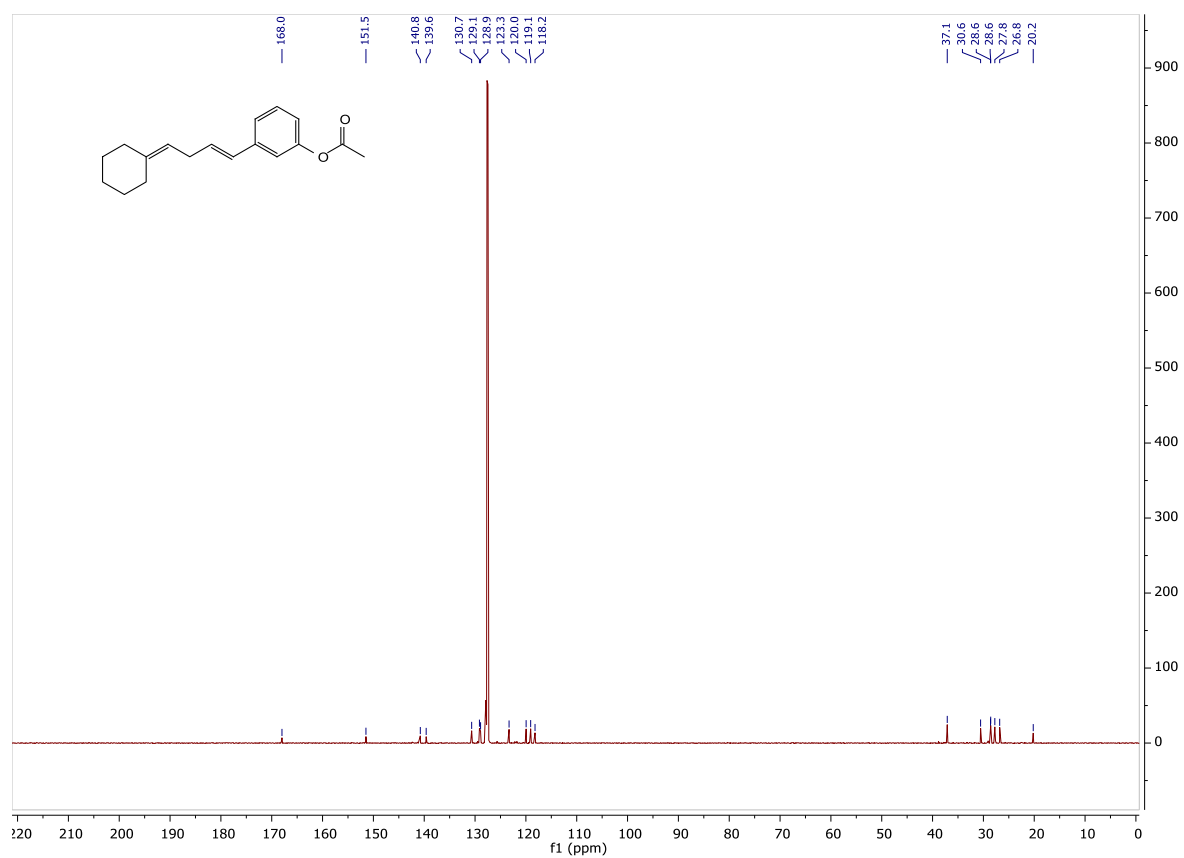
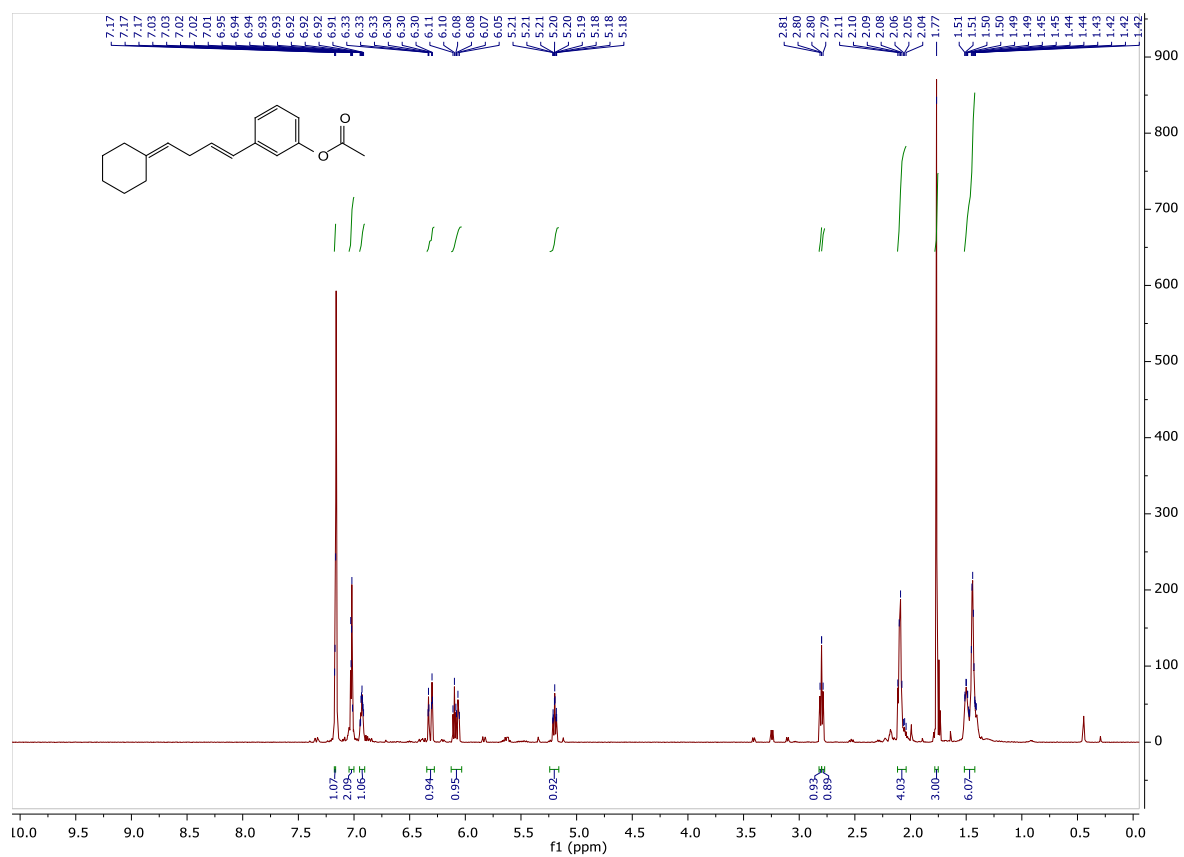
**Supplementary Figure S135.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.8.1.



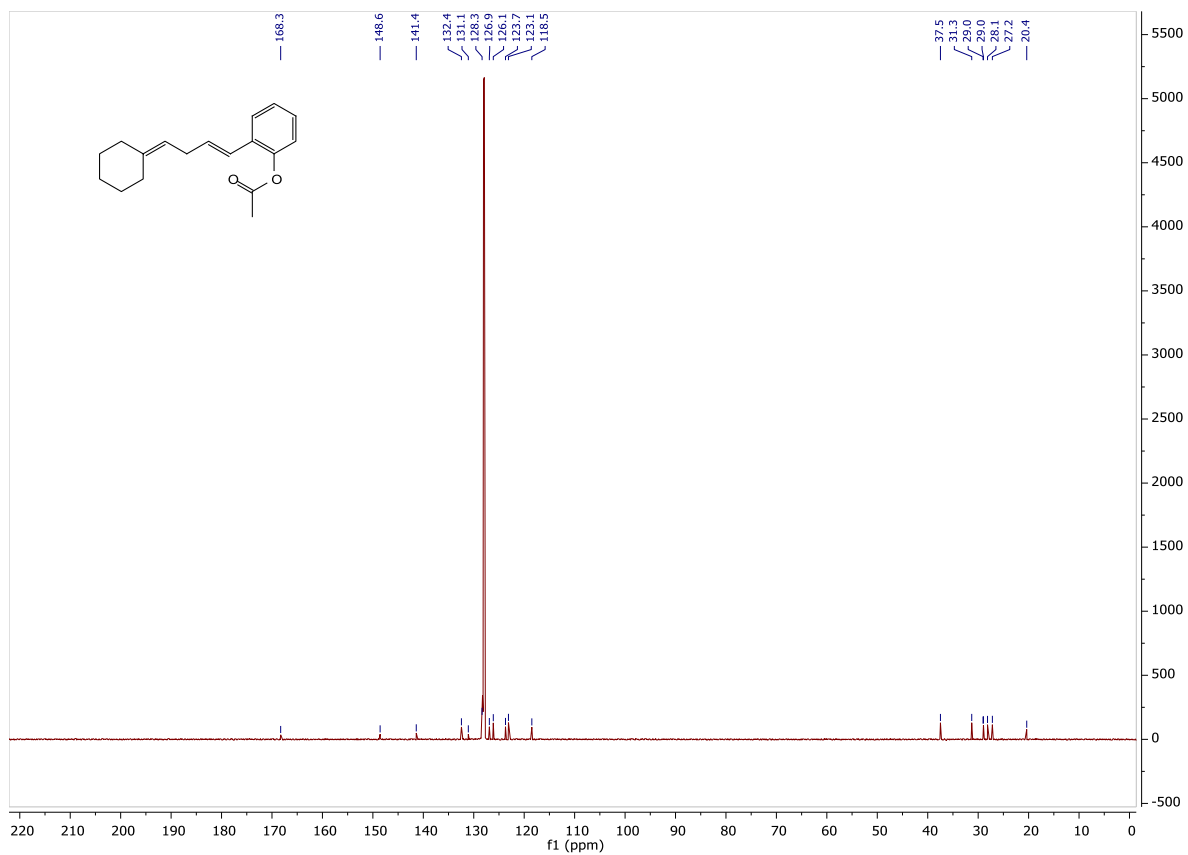
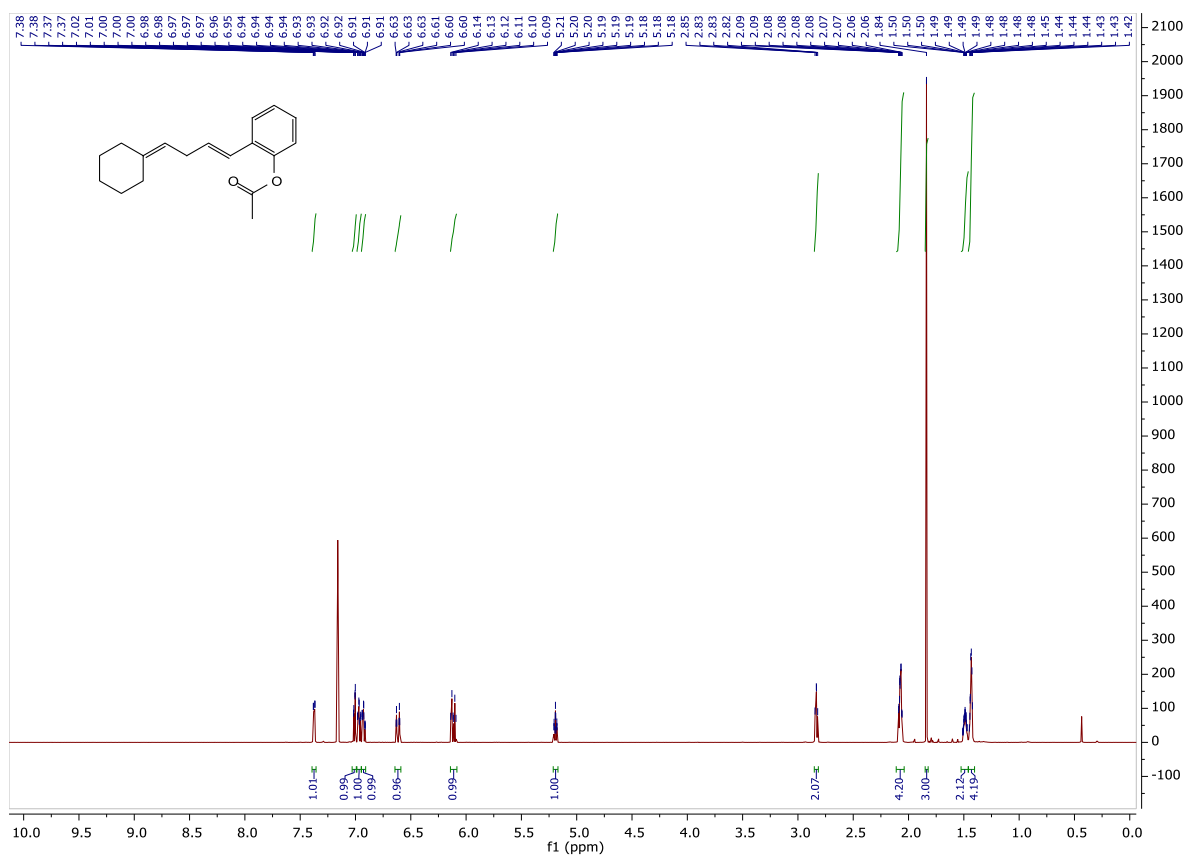
**Supplementary Figure S136.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **S6.8.2**.



**Supplementary Figure S137.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **7b** 4-OAc.

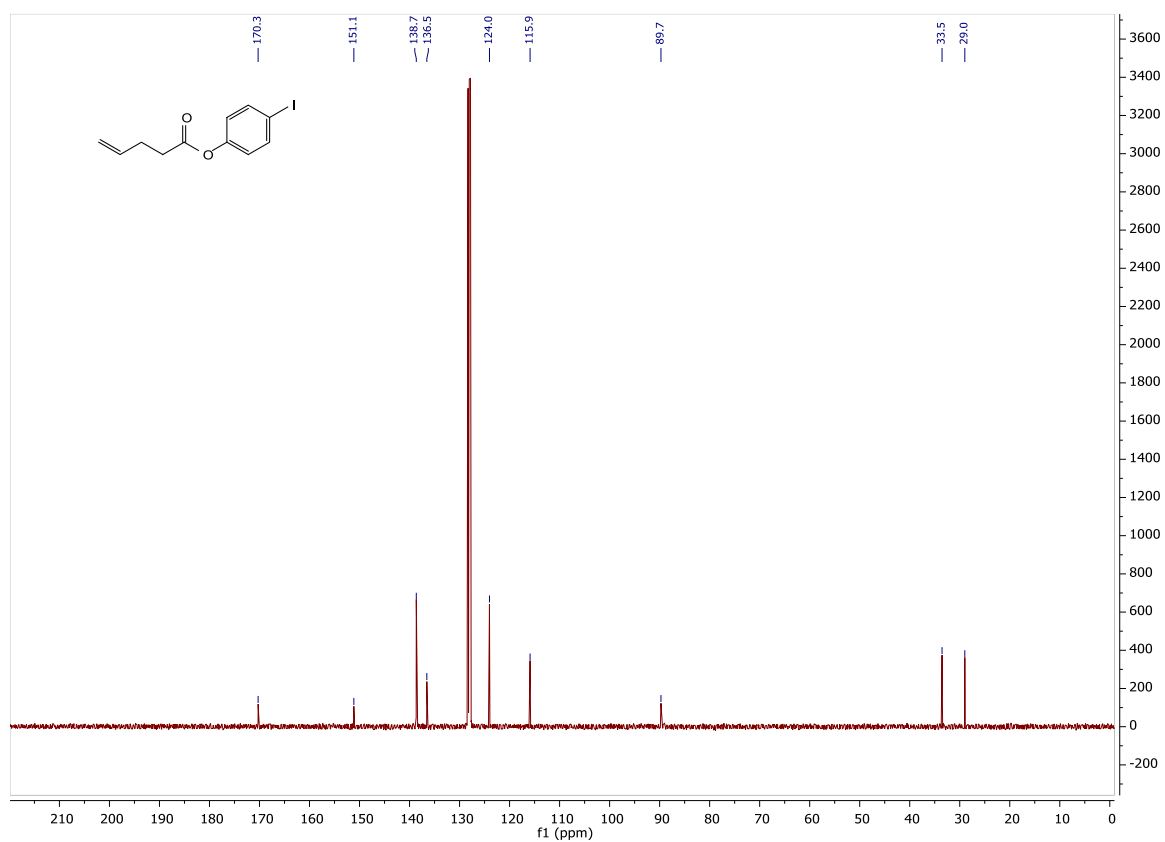
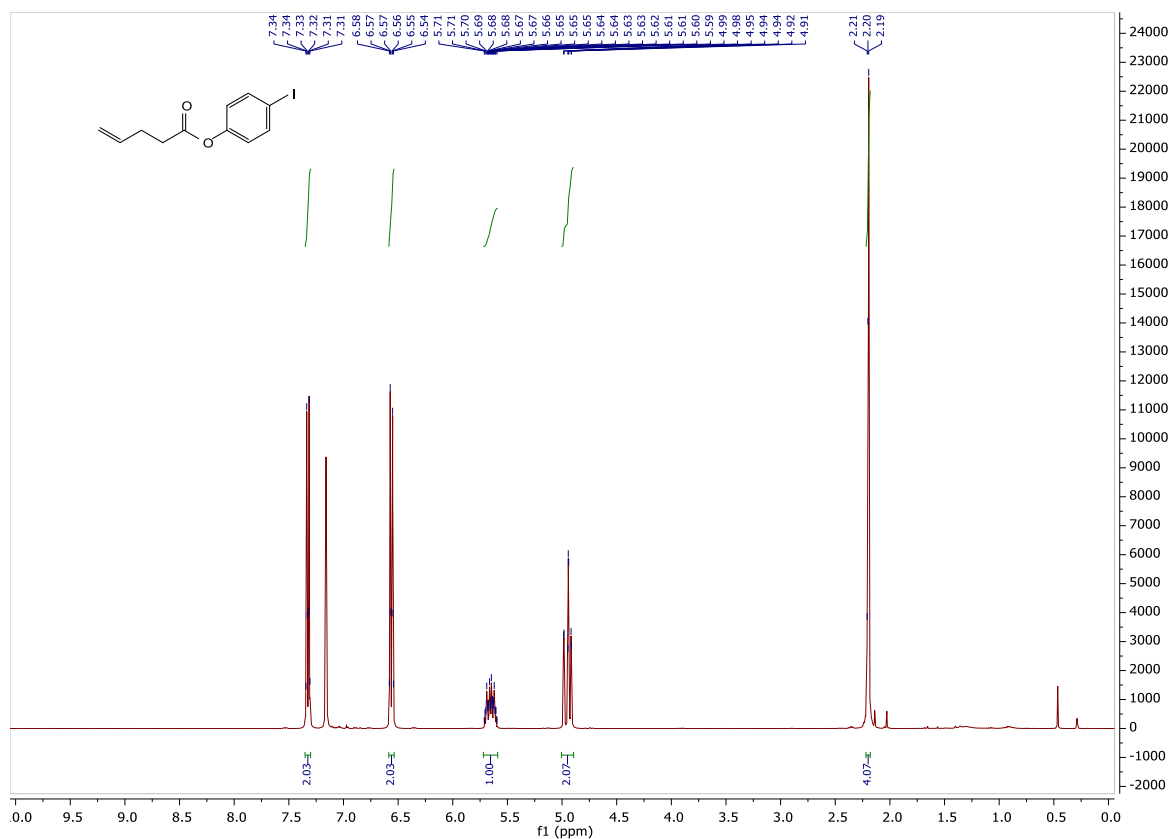


**Supplementary Figure S138.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **7b** 3-OAc.

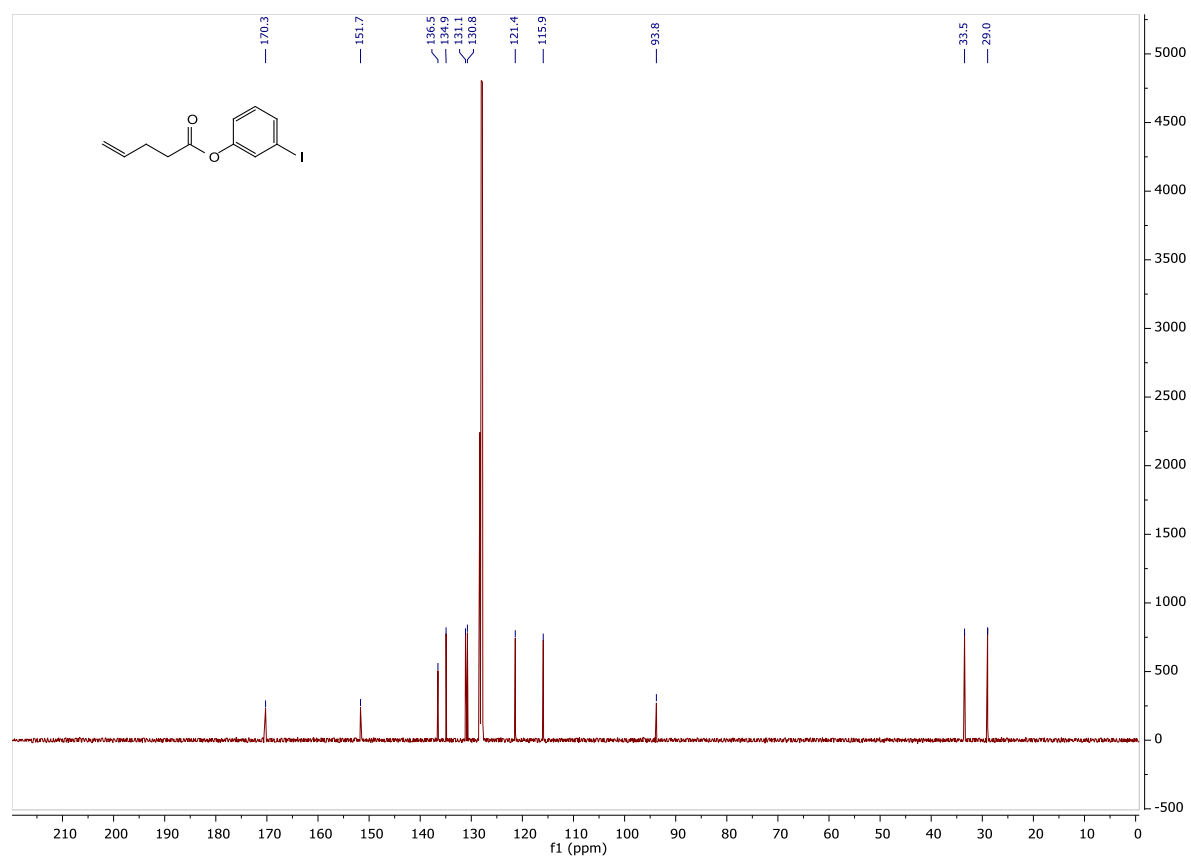
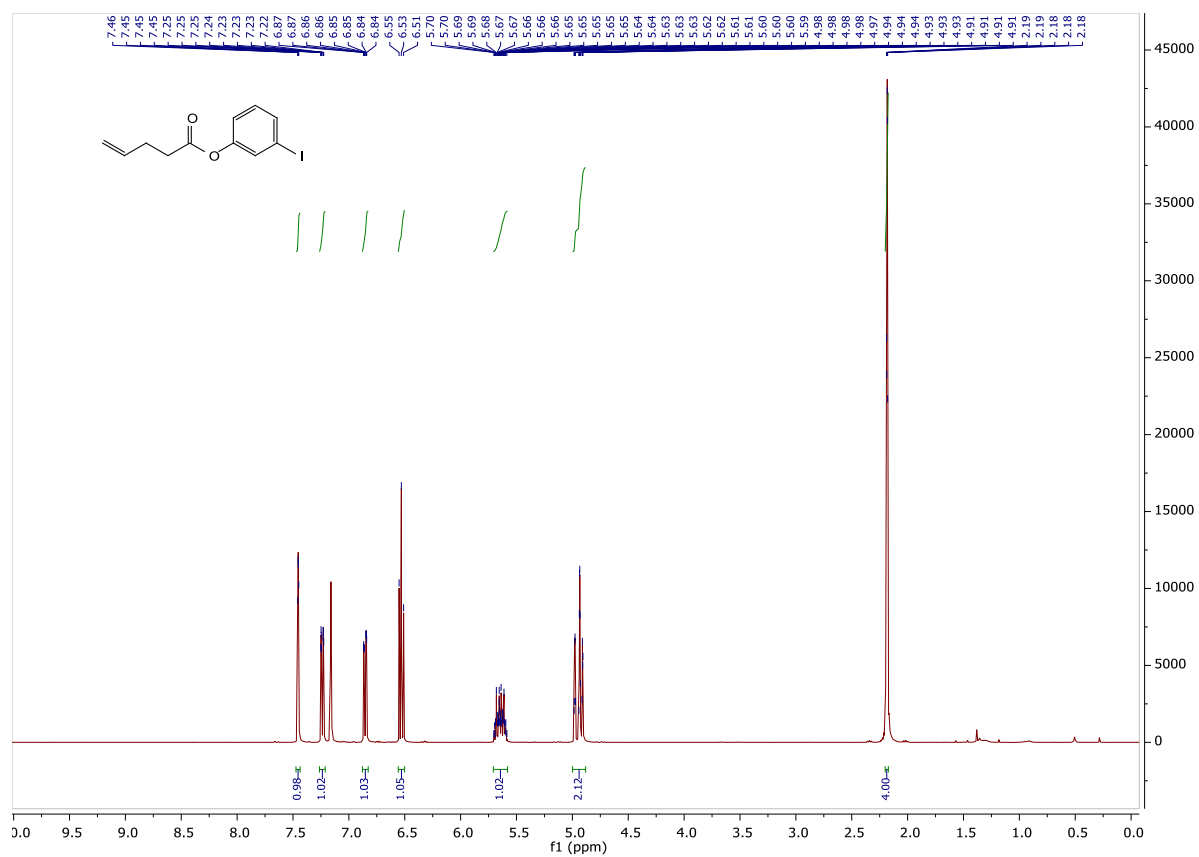


**Supplementary Figure S139.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **7b** 2-OAc.

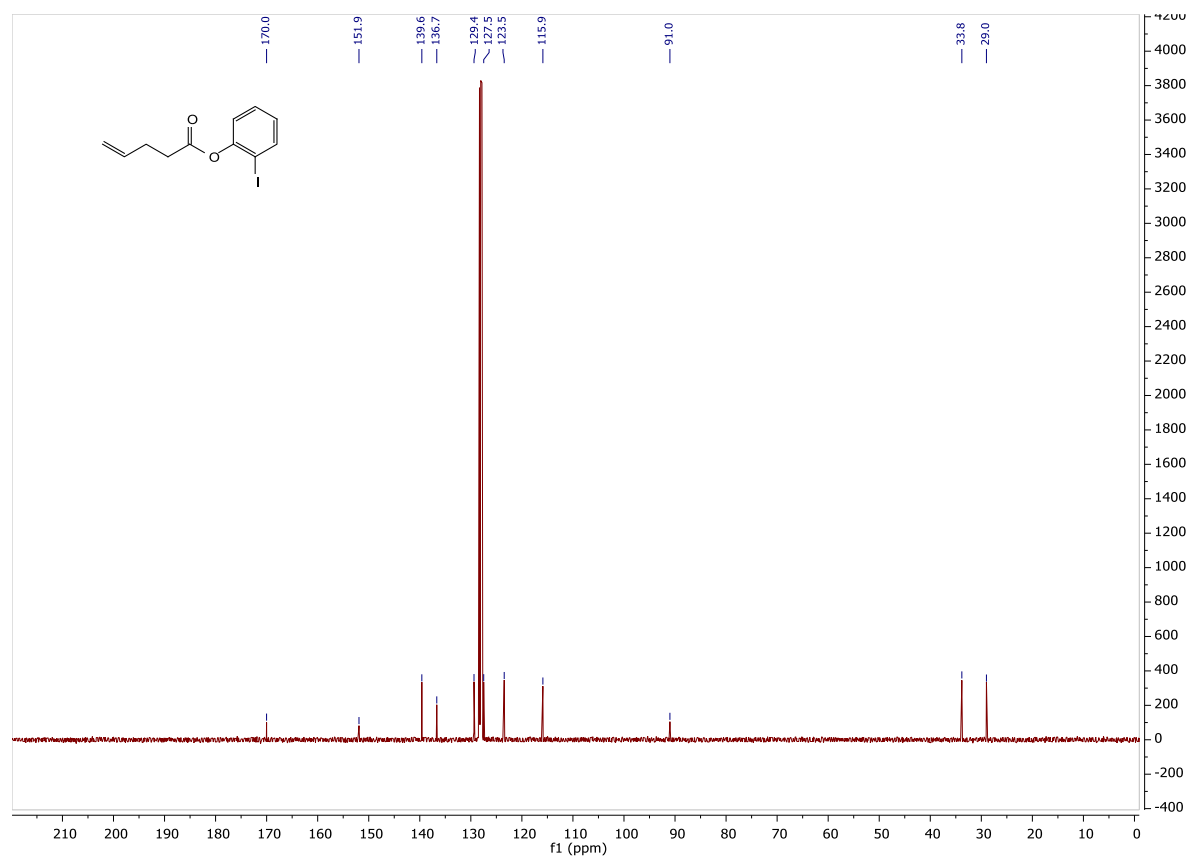
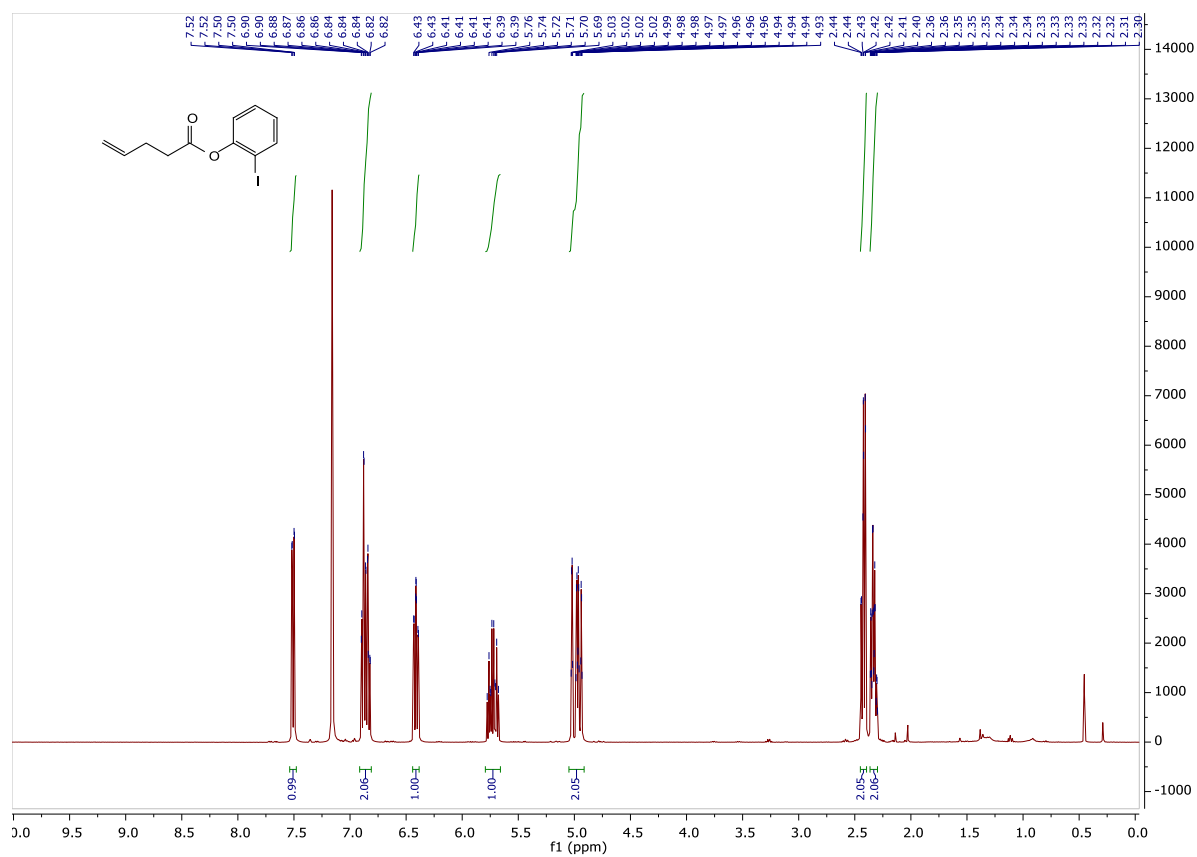




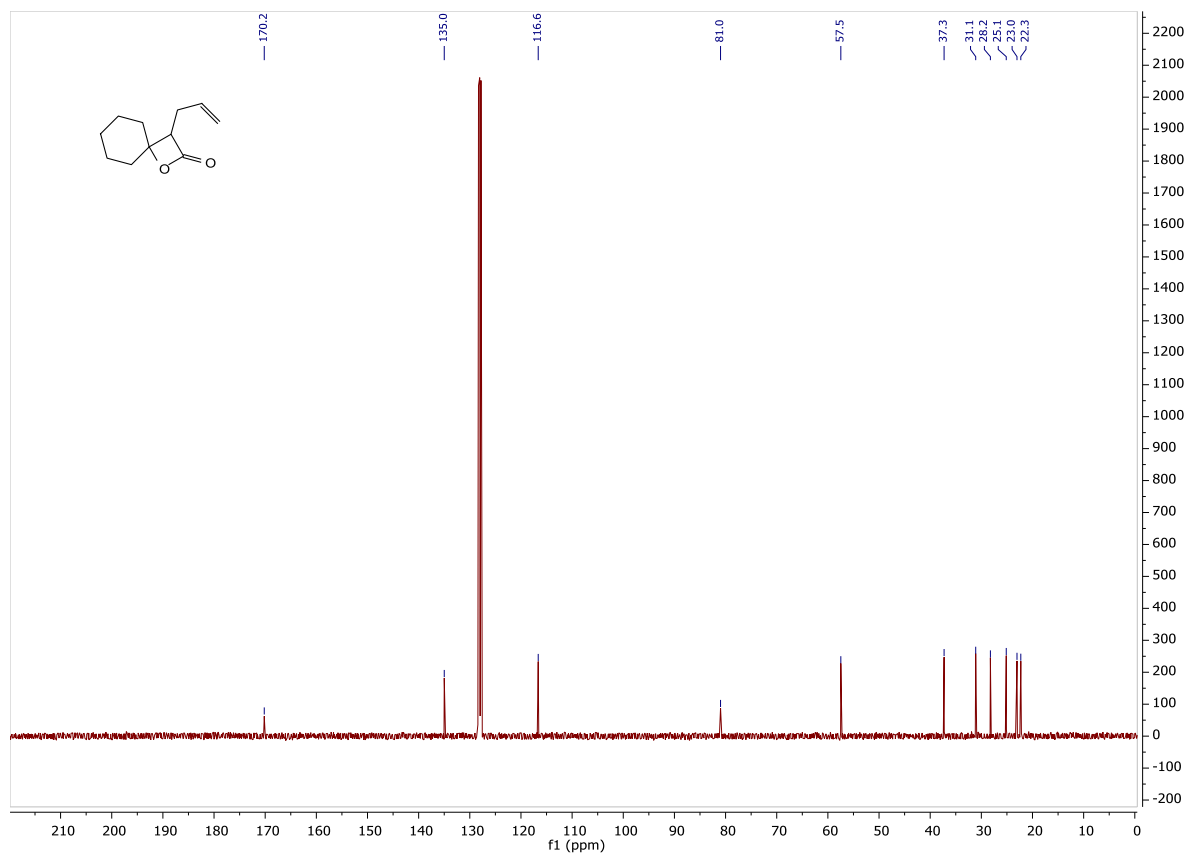
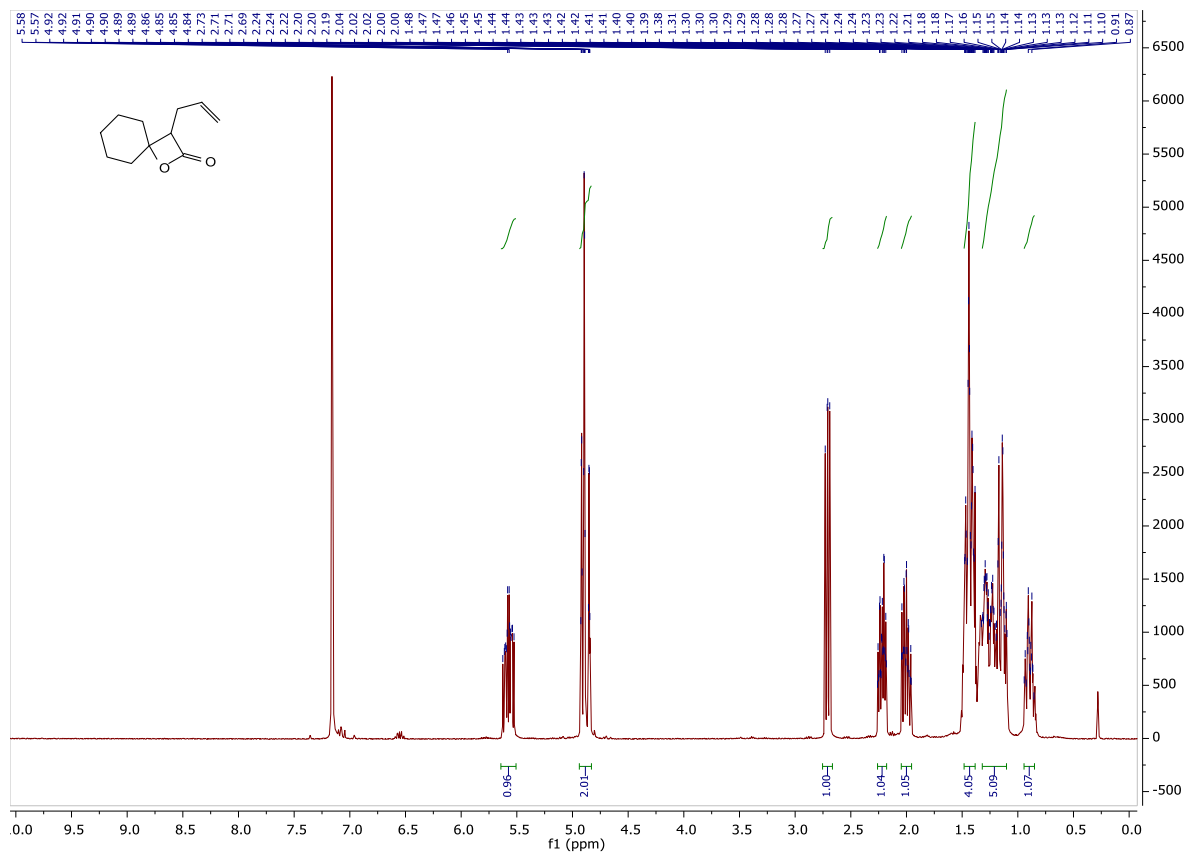
**Supplementary Figure S140.**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.9.1.



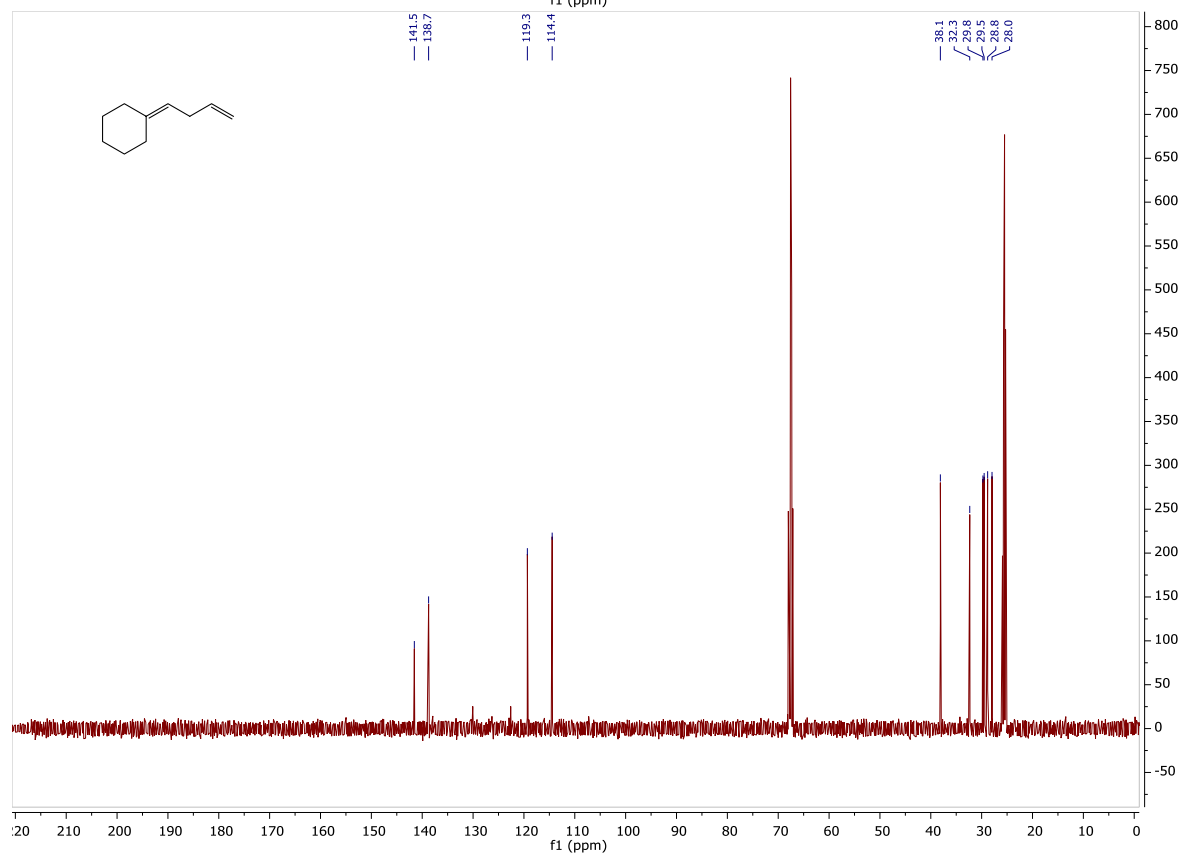
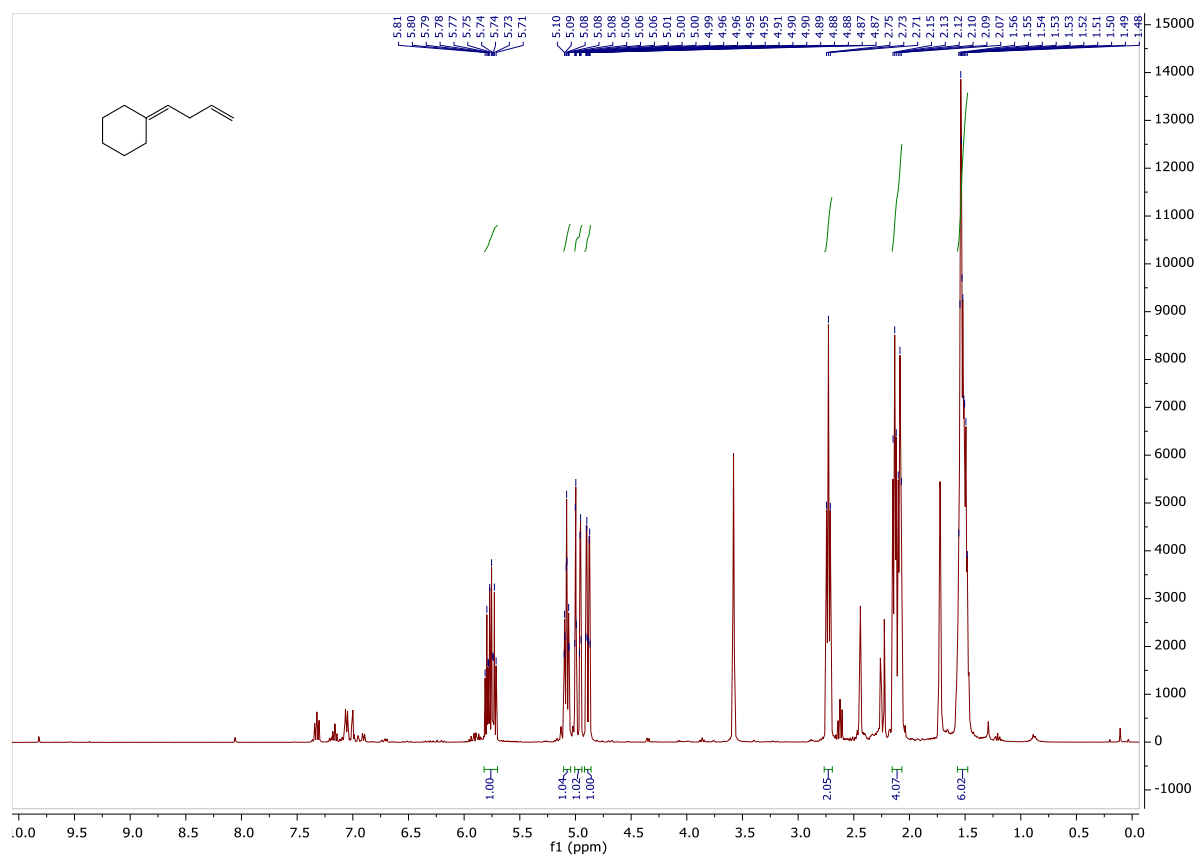
**Supplementary Figure S141.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.9.2.



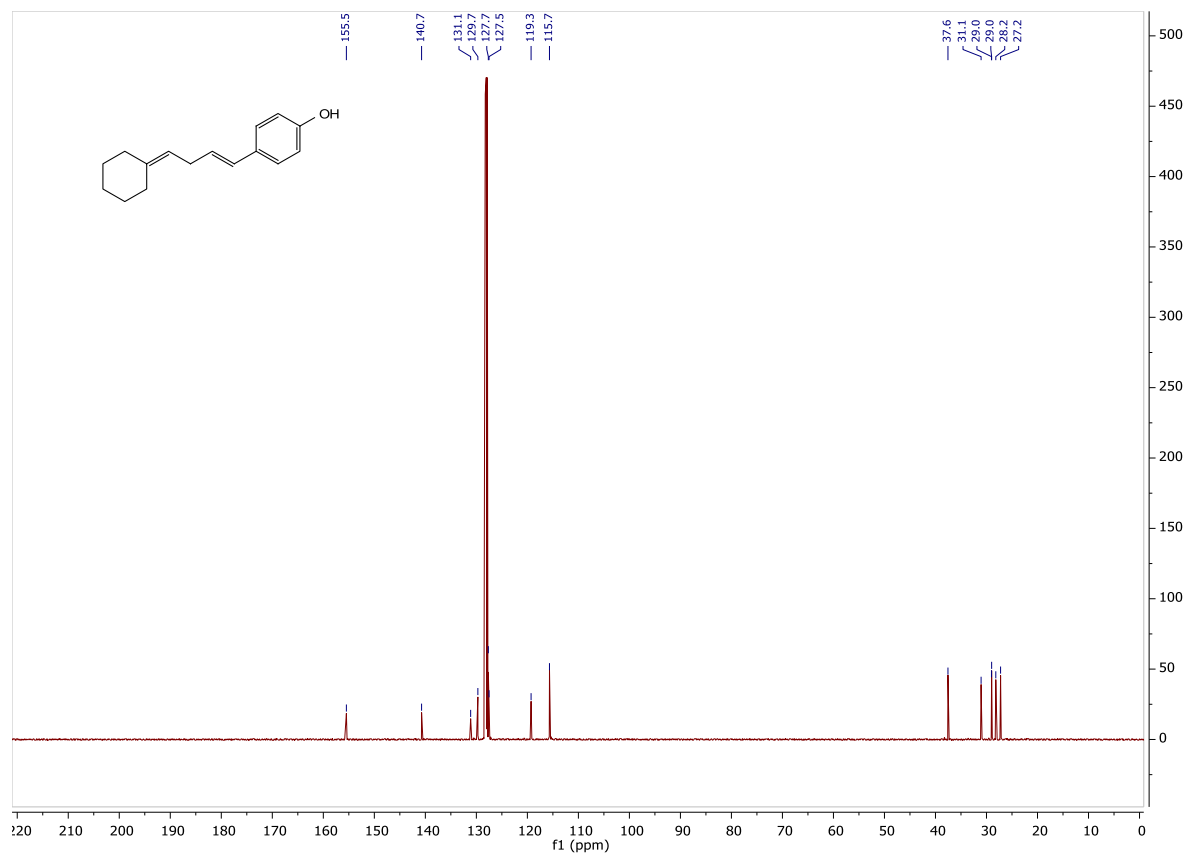
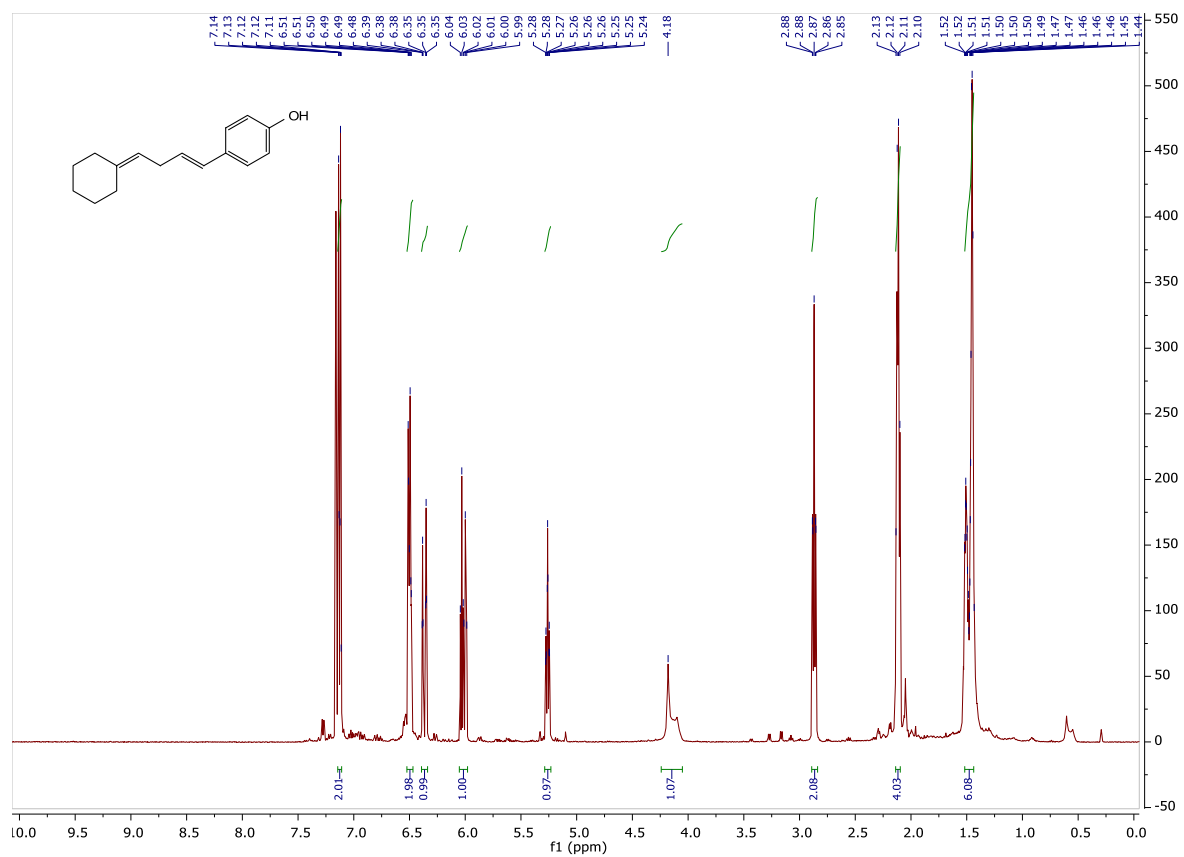
**Supplementary Figure S142.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **S6.9.3**.



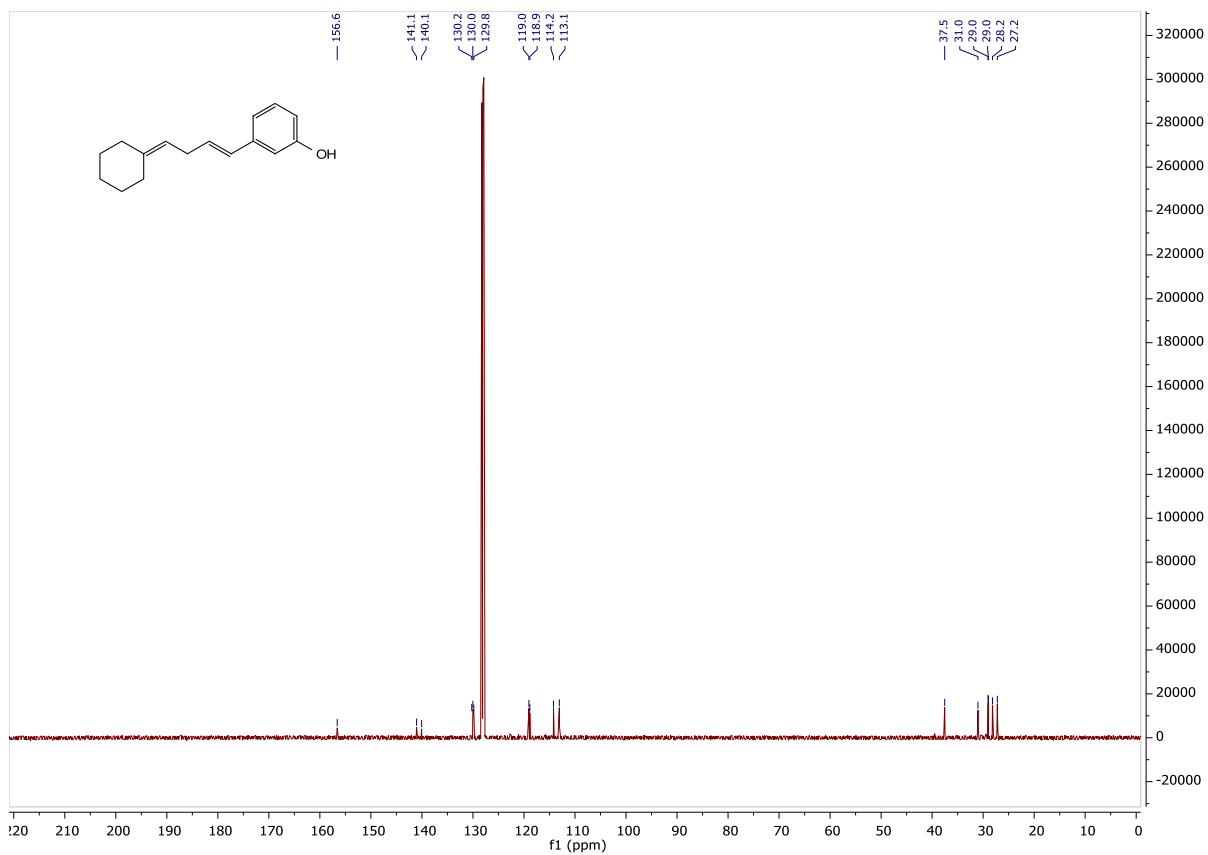
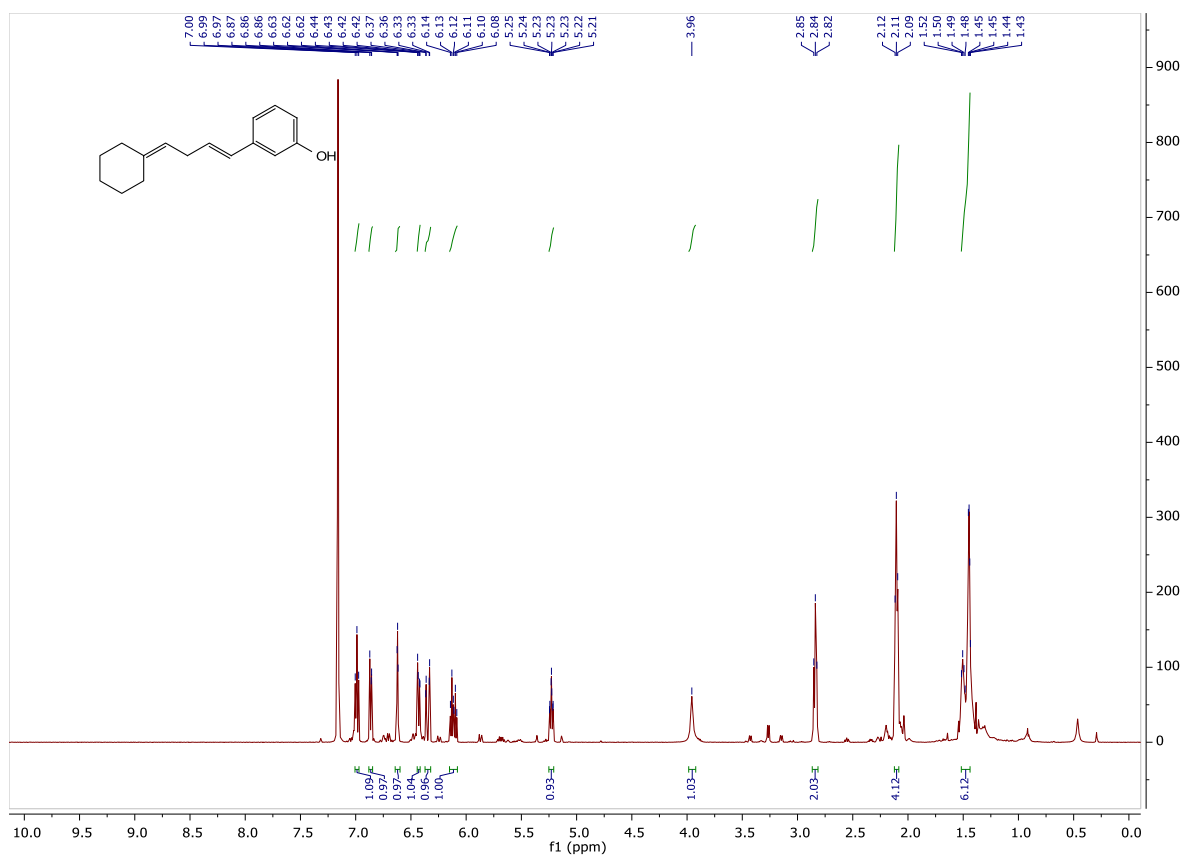
**Supplementary Figure S143.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.9.4.



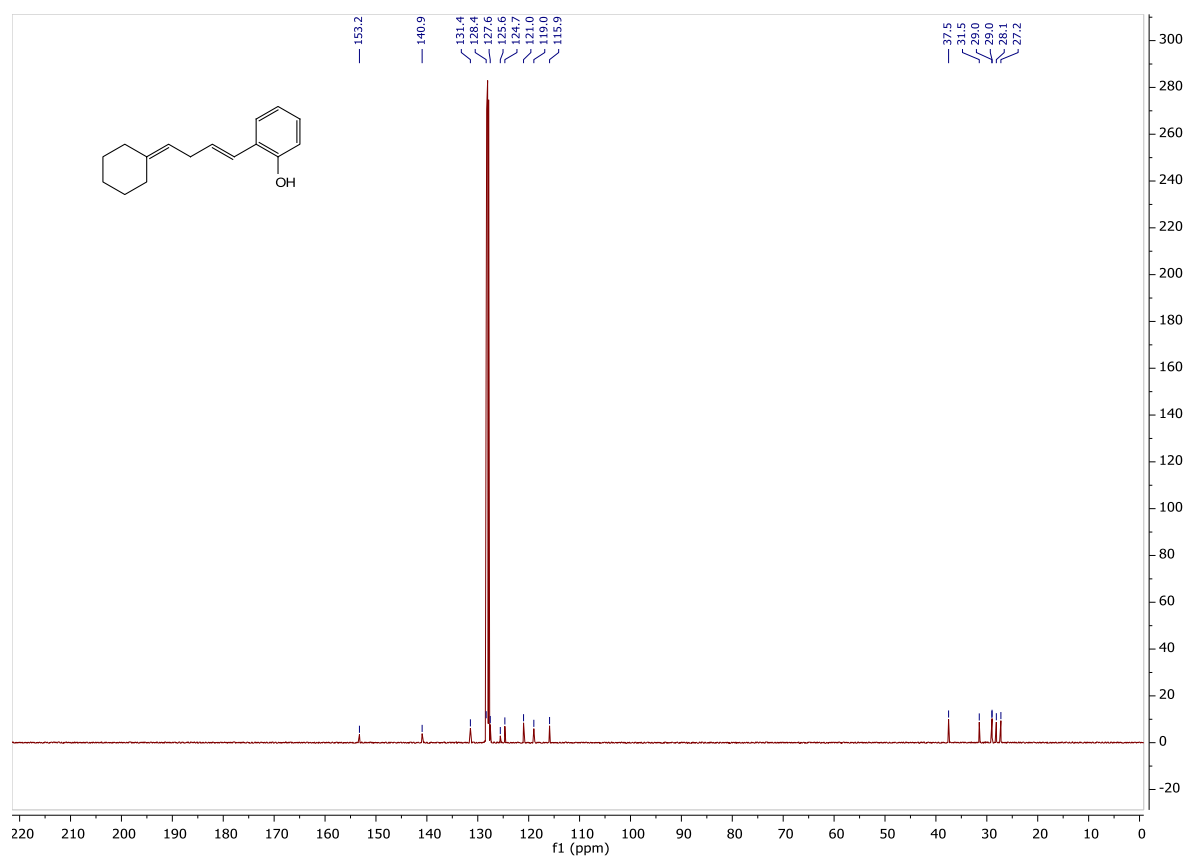
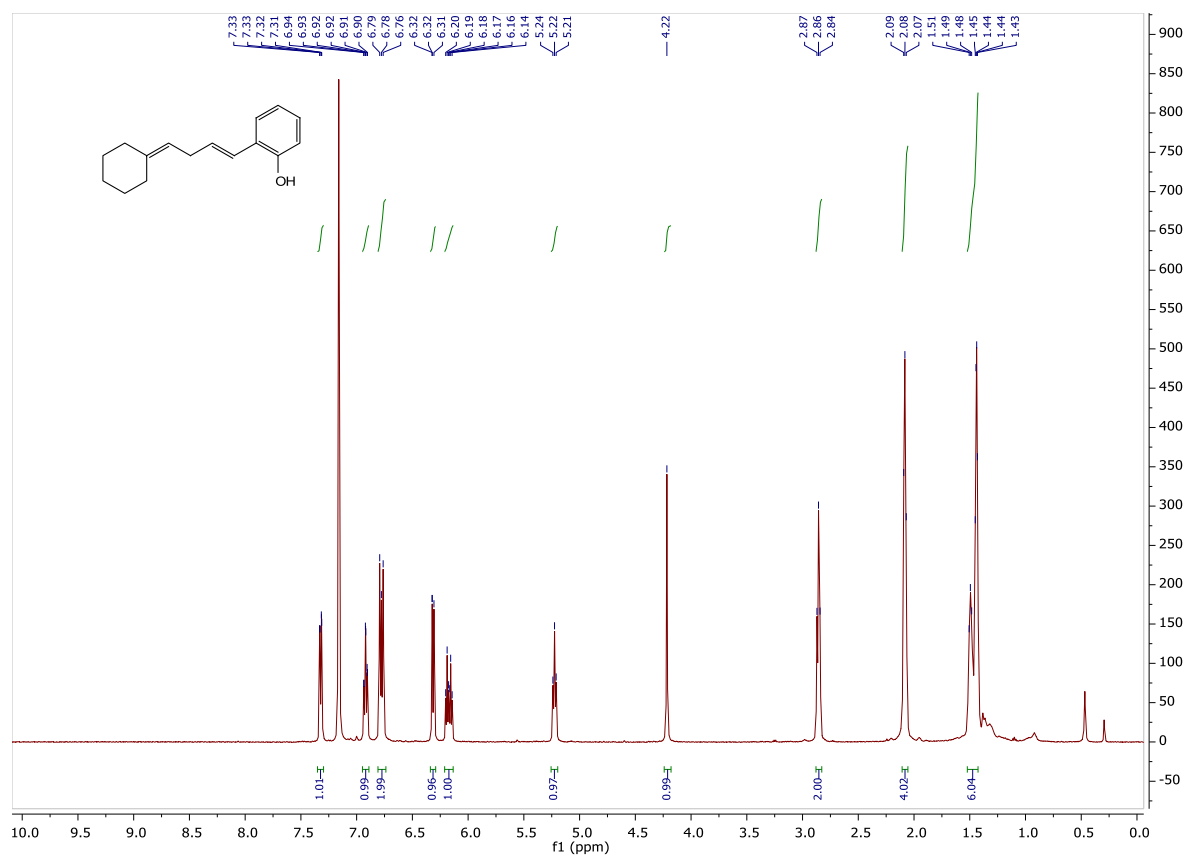
**Supplementary Figure S144.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.9.5.



Supplementary Figure S145.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.9.6.



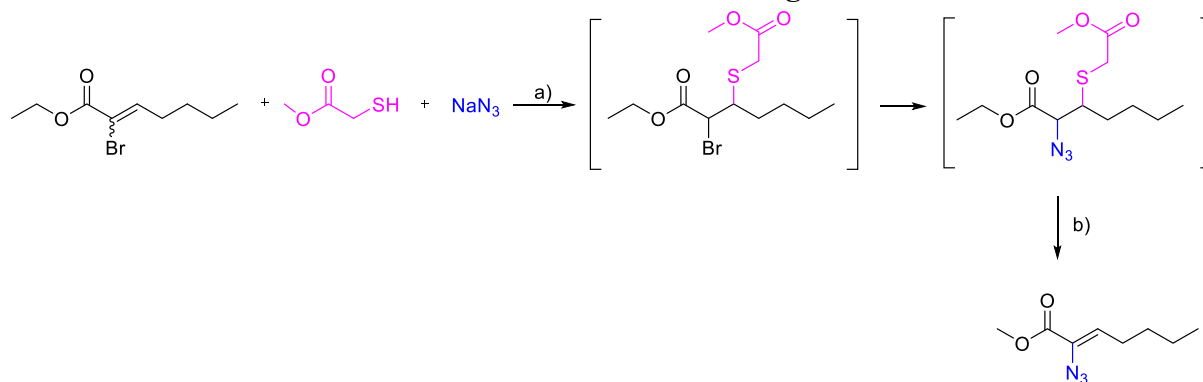
**Supplementary Figure S146.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **S6.9.7**.



**Supplementary Figure S147.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.9.8.

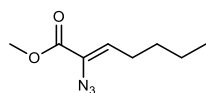


**Section S6.10 Reaction Mach10 described in main-text Figure 5c-e.**



**Supplementary Figure S148.** Non-organocatalytic variant of reaction **Mach10**. Reagents and conditions: (a) MeOH, rt, 5 h, (b) MeONa, rt, 10 min, 63%.

**One-pot procedure**



Methyl (Z)-2-azidohept-2-enoate **8b**

Ethyl 2-bromohept-2-enoate<sup>61</sup> (94 mg, 0.4 mmol), sodium azide (52 mg, 0.8 mmol, 2 equiv) and methyl thioglycolate (35.8  $\mu\text{L}$ , 0.4 mmol, 1 equiv) were stirred in methanol (2 mL) at rt for 5 h. Upon completion of the compound **S6.10.1** formation MeONa (43 mg, 0.8 mmol, 2 equiv) was added in one portion. The reaction mixture was stirred at rt for 10 min and then was quenched by addition of sat. aqueous solution of  $\text{NH}_4\text{Cl}$  (3 drops). Then reaction mixture was directly transferred onto silica gel and purified by column chromatography (pentane/DCM 9:1) to give **8b** (49 mg, 67%) as a colorless liquid.

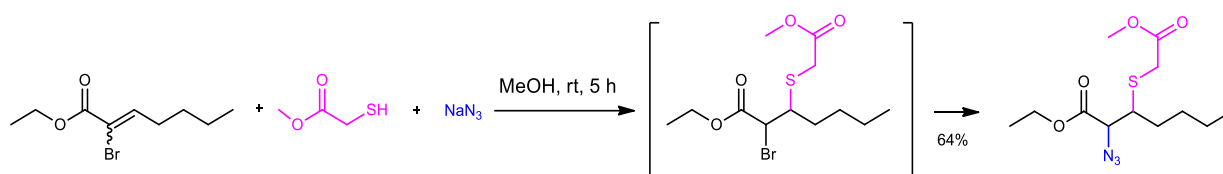
**$^1\text{H}$  NMR** (500 MHz,  $\text{CDCl}_3$ )  $\delta$  6.19 (t,  $J = 7.6$  Hz, 1H), 3.82 (s, 3H), 2.22 (q,  $J = 7.4$  Hz, 2H), 1.44 – 1.30 (m, 4H), 0.91 (t,  $J = 7.2$  Hz, 3H);

**$^{13}\text{C}$  NMR** (125 MHz,  $\text{CDCl}_3$ )  $\delta$  163.3, 132.1, 127.6, 52.5, 30.4, 27.0, 22.3, 13.8.

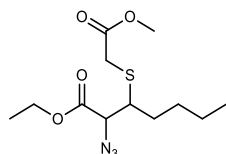
**IR** (film,  $\text{CH}_2\text{Cl}_2$ ) 2956, 2925, 2855, 2189, 2123, 1725, 1634, 1459, 1438, 1371, 1321, 1275, 1256, 1209, 1146, 1069, 932, 825, 758  $\text{cm}^{-1}$ ;

**Anal.** Calcd for  $\text{C}_8\text{H}_{13}\text{N}_3\text{O}_2$ : C, 52.45; H, 7.15. Found: C, 52.59; H, 7.37.

### Mechanistic validation by stepwise isolation



**Supplementary Figure S149.** Reaction scheme of the stepwise process to confirm second step intermediate



### Ethyl 2-azido-3-((2-methoxy-2-oxoethyl)thio)heptanoate **S6.10.1**

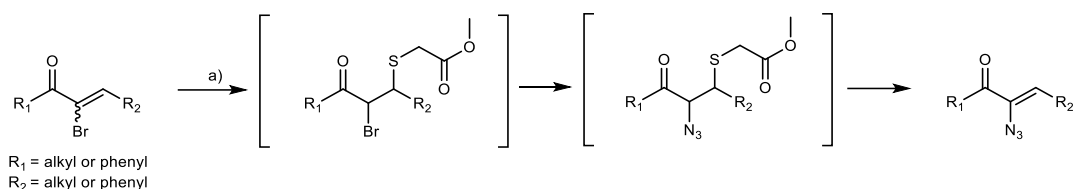
Ethyl 2-bromohept-2-enoate<sup>61</sup> (94 mg, 0.4 mmol), sodium azide (52 mg, 0.8 mmol, 2 equiv) and methyl thioglycolate (35.8  $\mu\text{L}$ , 0.4 mmol, 1 equiv) were dissolved in methanol (2 mL) at  $\text{rt}$ . Reaction was stirred for 5 h and then reaction mixture was directly transferred onto silica gel and purified by column chromatography (pentane/diethyl ether 9:1) to give **S6.10.1** (83 mg, 64%) as a colorless liquid.

**$^1\text{H NMR}$**  (600 MHz,  $\text{CDCl}_3$ )  $\delta$  4.32 – 4.23 (m, 3H), 4.23 – 4.17 (m, 1H), 3.74 (s, 1H), 3.72 (s, 3H), 3.66 (s, 1H), 3.48 (s, 1H), 3.40 – 3.14 (m, 4H), 2.47 (q,  $J = 7.4$  Hz, 1H), 1.70 – 1.57 (m, 3H), 1.54 – 1.36 (m, 4H), 1.35 – 1.31 (m, 7H), 0.91 (t,  $J = 7.0$  Hz, 5H).

**$^{13}\text{C NMR}$**  (150 MHz,  $\text{CDCl}_3$ )  $\delta$  170.7, 170.6, 170.0, 168.7, 168.7, 164.7, 152.3, 125.9, 66.4, 66.3, 62.2, 62.1, 61.5, 52.5, 52.4, 52.2, 47.9, 35.0, 33.4, 33.3, 32.5, 30.5, 30.4, 30.3, 29.1, 28.8, 22.4, 22.3, 14.2, 14.1, 13.9, 13.8.

**IR** (film,  $\text{CH}_2\text{Cl}_2$ ) 2956, 2932, 2872, 2861, 2113, 1741, 1606, 1465, 1437, 1369, 1277, 1200, 1159, 1134, 1025, 859, 754, 708  $\text{cm}^{-1}$ ;

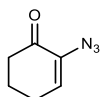
**HRMS** (ESI)  $m/z$ :  $[\text{M}+\text{Na}]^+$  Calcd for  $\text{C}_{12}\text{H}_{21}\text{N}_3\text{O}_4\text{SNa}$  326.1145; Found 326.1150.



**Supplementary Figure S150.** Reaction scheme of the *organocatalytic* reaction **Mach10** described in main-text **Figure 5c-e**. Reagents and conditions: (a)  $\text{NaN}_3$  (2 equiv), methyl thioglycolate (10 or 35 mol%), MeOH, rt, 17 h, 33-57%.

### One-pot, organocatalytic variant of Mach 10

$\alpha$ -bromo- $\alpha,\beta$ -unsaturated ketone (0.4 mmol), sodium azide (2 equiv) and methyl thioglycolate (10 mol% for **8d** and **8c** or 35 mol% for **8e** and **8f**) were stirred in methanol (2 mL). Reaction was stirred for 17 h and then reaction mixture was directly transferred onto silica gel and purified by column chromatography.



### 2-azidocyclohex-2-en-1-one **8d**<sup>62</sup>

According to general procedure, the reaction was performed with 2-bromocyclohex-2-en-1-one<sup>63</sup> as the starting material. After purification by column chromatography (pentane/Et<sub>2</sub>O 95:5) compound **8d** was obtained (25 mg, 46%) as a colorless liquid (**product is highly volatile!**).

The spectral data match those reported in the literature<sup>62</sup>.

**<sup>1</sup>H NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.43 (t,  $J = 4.6$  Hz, 1H), 2.56 – 2.51 (m, 2H), 2.43 (td,  $J = 6.1$ , 4.6 Hz, 2H), 2.05 – 1.97 (m, 2H);

**<sup>13</sup>C NMR** (100 MHz,  $\text{CDCl}_3$ )  $\delta$  194.0, 135.3, 132.7, 38.3, 25.1, 22.6;



### 2-azidocyclopent-2-en-1-one **8c**

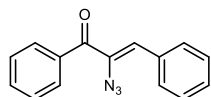
According to general procedure, the reaction was performed with 2-bromocyclopent-2-en-1-one<sup>64</sup> as the starting material. After purification by column chromatography (pentane/Et<sub>2</sub>O 95:5) compound **8c** was obtained (28 mg, 57%) as a colorless liquid (**product is highly volatile!**).

**<sup>1</sup>H NMR** (400 MHz,  $\text{CDCl}_3$ )  $\delta$  6.86 (t,  $J = 3.1$  Hz, 1H), 2.62 – 2.55 (m, 2H), 2.52 – 2.48 (m, 2H);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  202.7, 142.1, 139.9, 33.7, 23.8;

IR (film,  $\text{CH}_2\text{Cl}_2$ ) 2925, 2855, 2195, 2118, 2063, 1710, 1620, 1441, 1404, 1343, 1290, 1230, 1164, 1063, 1032, 1013, 994, 925, 868, 784, 768, 723  $\text{cm}^{-1}$ ;

HRMS (EI)  $m/z$ :  $[\text{M}]^+$  Calcd for  $\text{C}_5\text{H}_5\text{N}_3\text{O}$  123.0433; Found 123.0427.



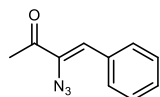
**(Z)-2-azido-1,3-diphenylprop-2-en-1-one 8e**<sup>65</sup>

According to general procedure, the reaction was performed with 2-bromo-1,3-diphenylprop-2-en-1-one<sup>66</sup> as the starting material. After purification by column chromatography (hexane/EtOAc 95:5) compound 8e was obtained (38 mg, 38%) as a yellowish solid.

The spectral data match those reported in the literature.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.84 – 7.78 (m, 4H), 7.62 (t,  $J$  = 8.0 Hz, 1H), 7.50 (t,  $J$  = 8.0 Hz, 2H), 7.42 – 7.33 (m, 3H), 6.47 (s, 1H);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  192.2, 136.8, 133.6, 133.1, 132.7, 130.6, 129.8, 129.7, 129.6, 128.6, 128.5;



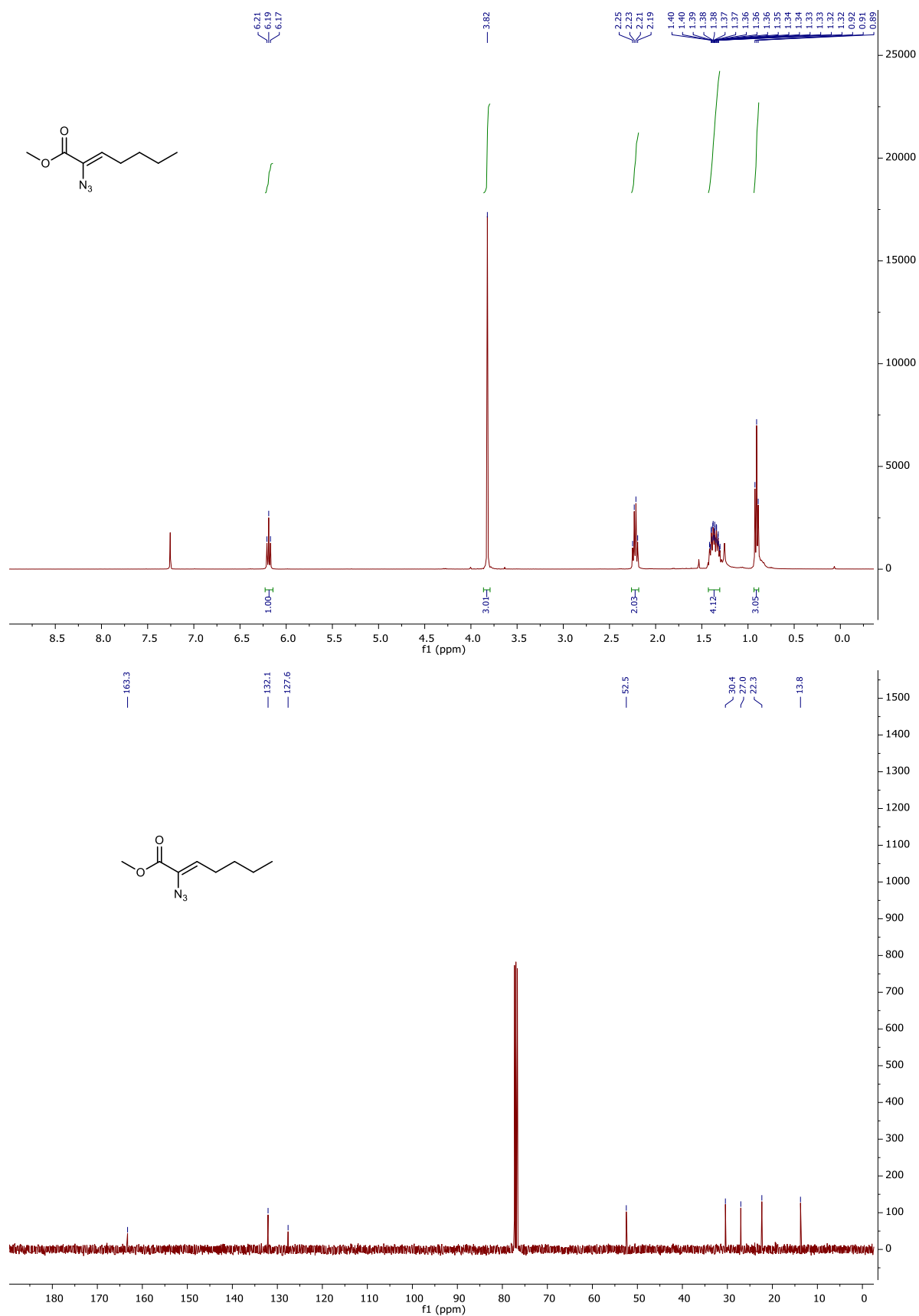
**(Z)-3-azido-4-phenylbut-3-en-2-one 8f**<sup>67</sup>

According to general procedure, the reaction was performed with (Z)-3-bromo-4-phenylbut-3-en-2-one<sup>68</sup> as the starting material. After purification by column chromatography (hexane/EtOAc 9:1) compound 8f was obtained (36 mg, 48%) as a colorless solid.

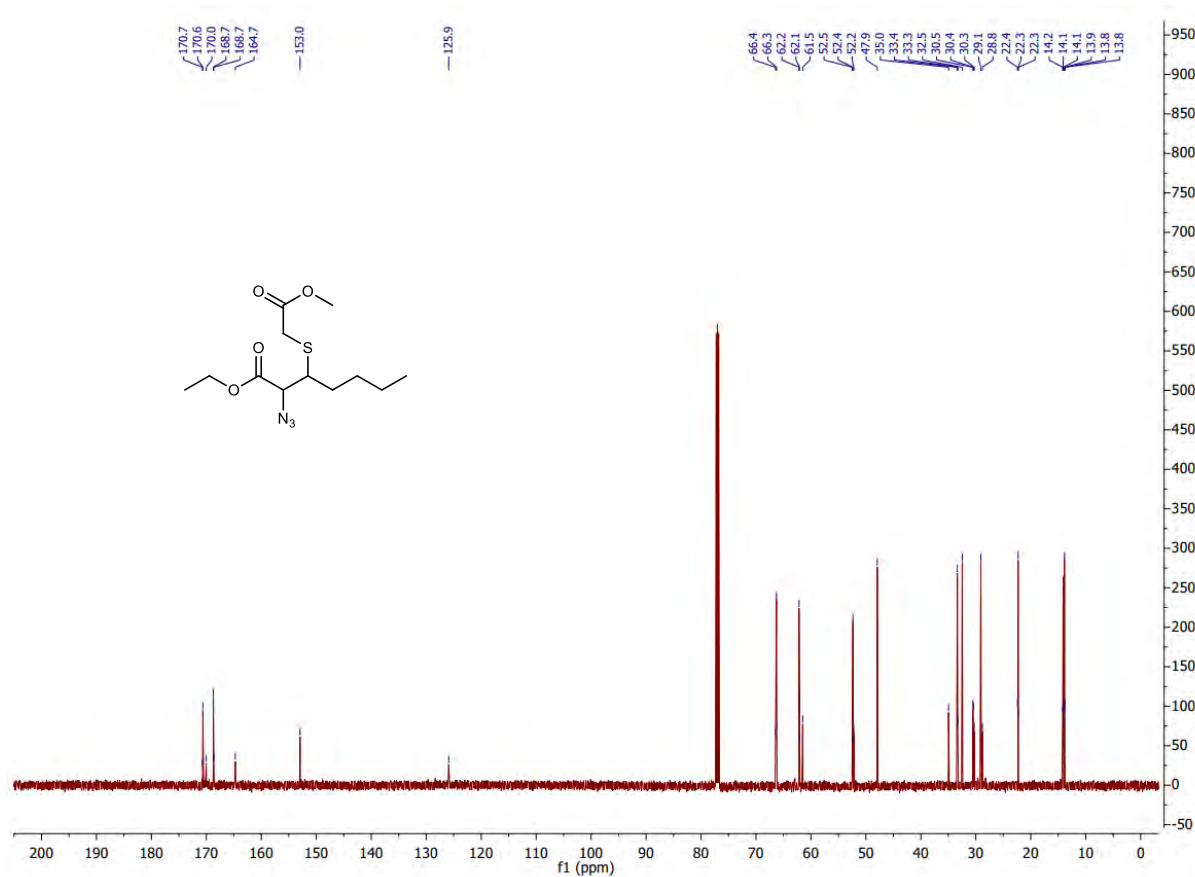
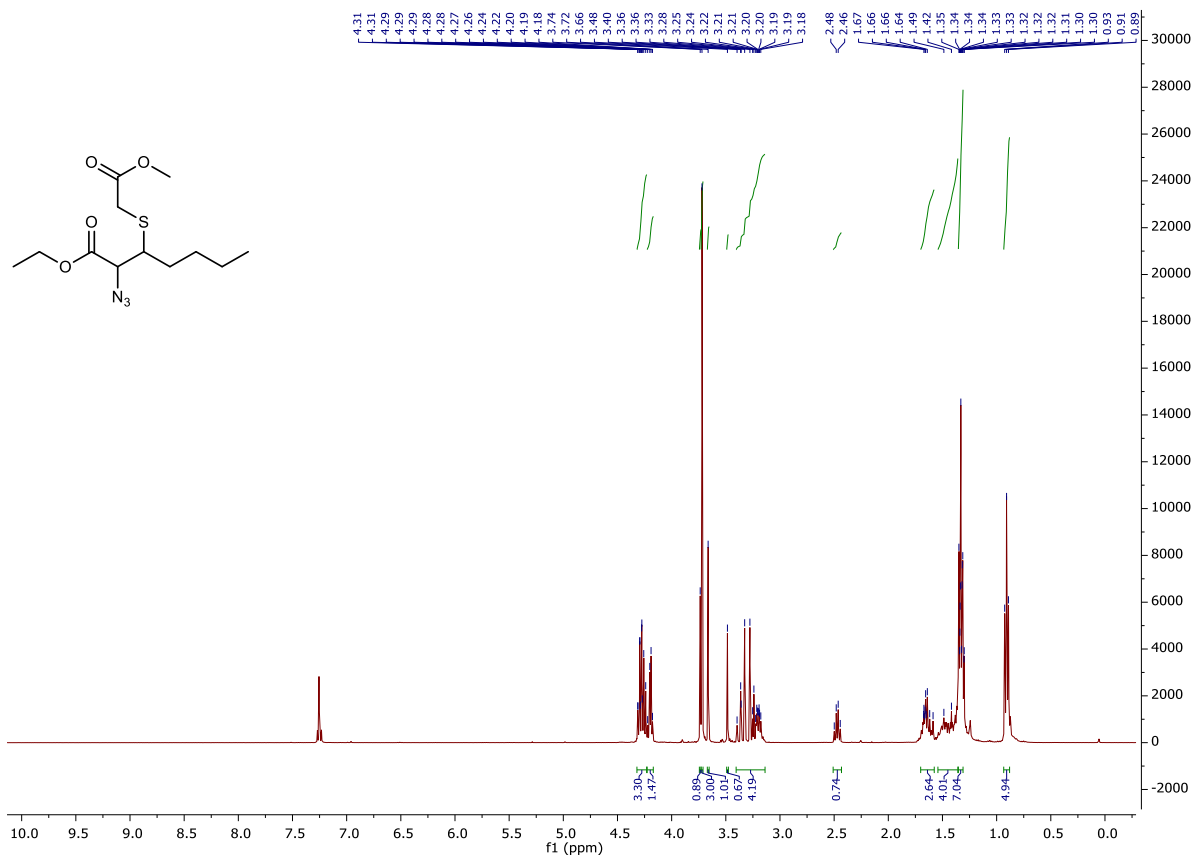
The spectral data match those reported in the literature.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.87 – 7.83 (m, 2H), 7.44 – 7.33 (m, 3H), 6.71 (s, 1H), 2.51 (s, 3H);

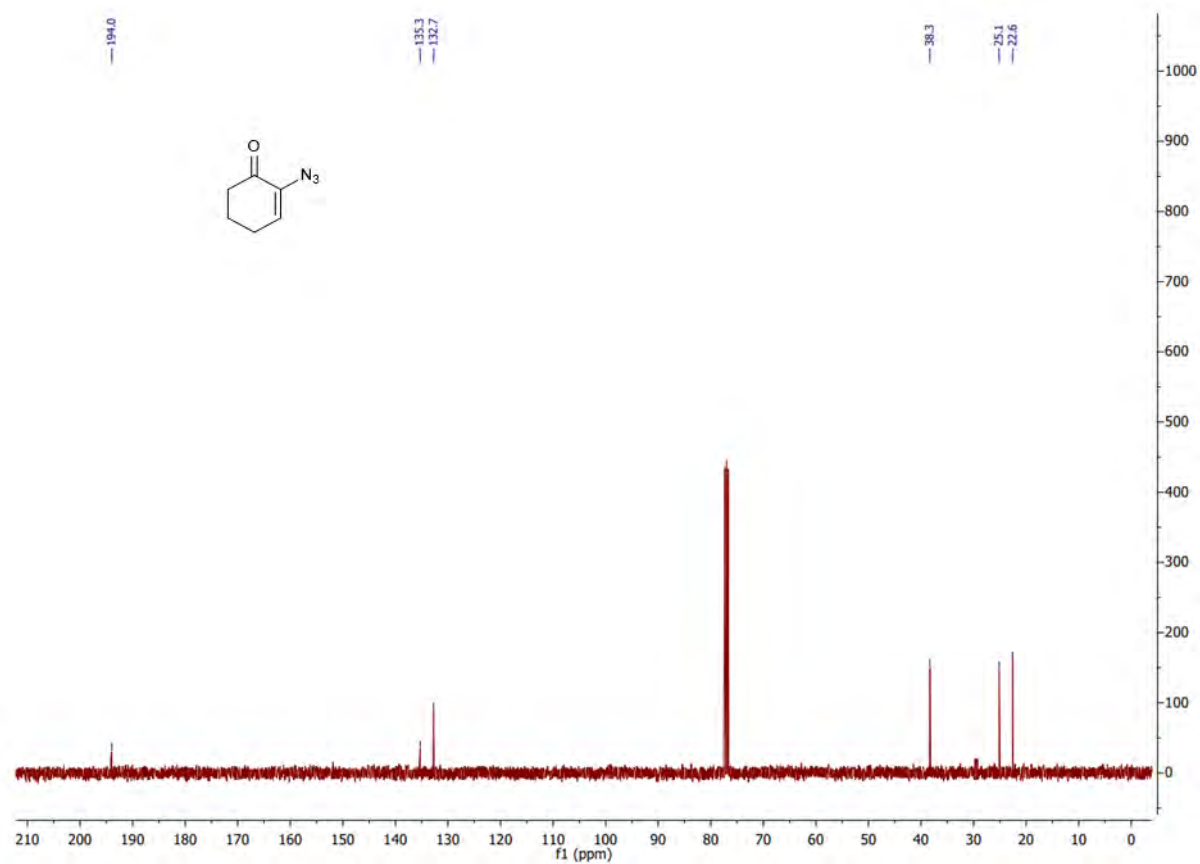
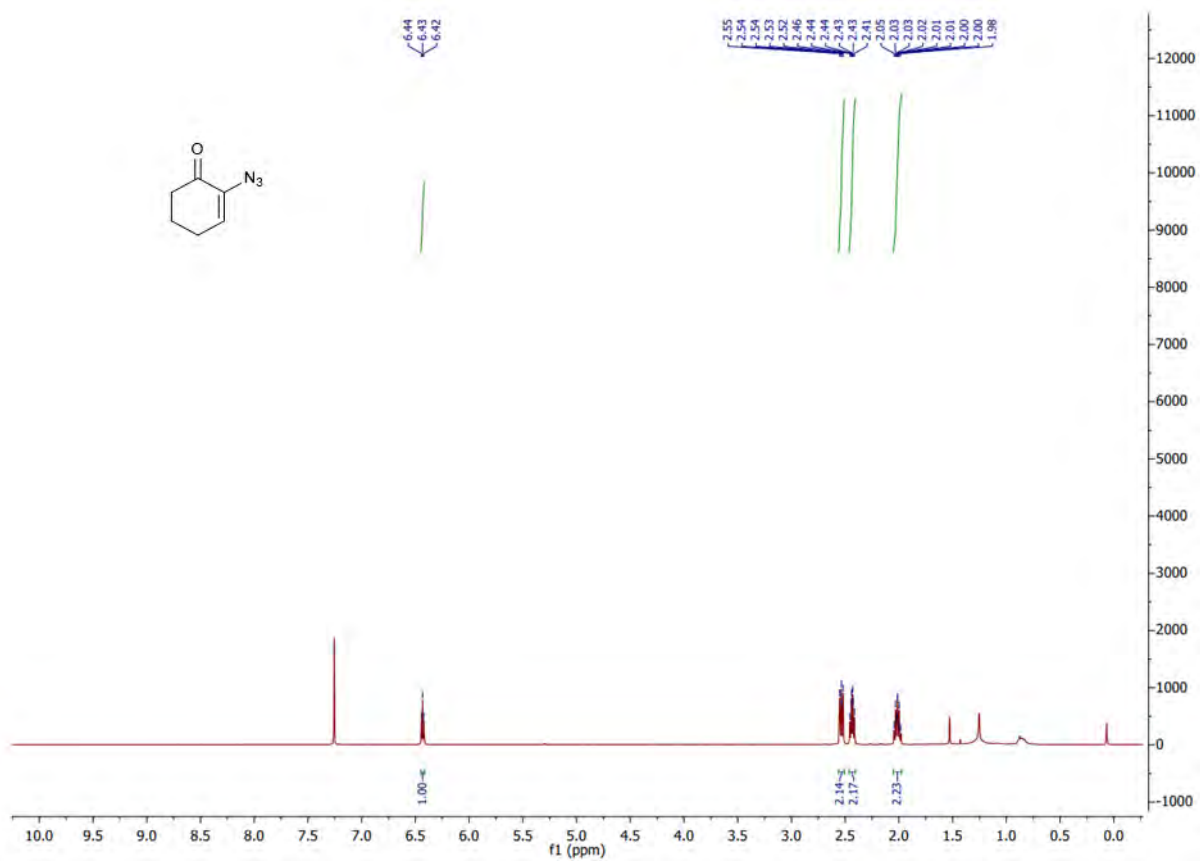
$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  194.3, 134.0, 133.2, 130.7, 129.8, 128.6, 126.7, 25.8;



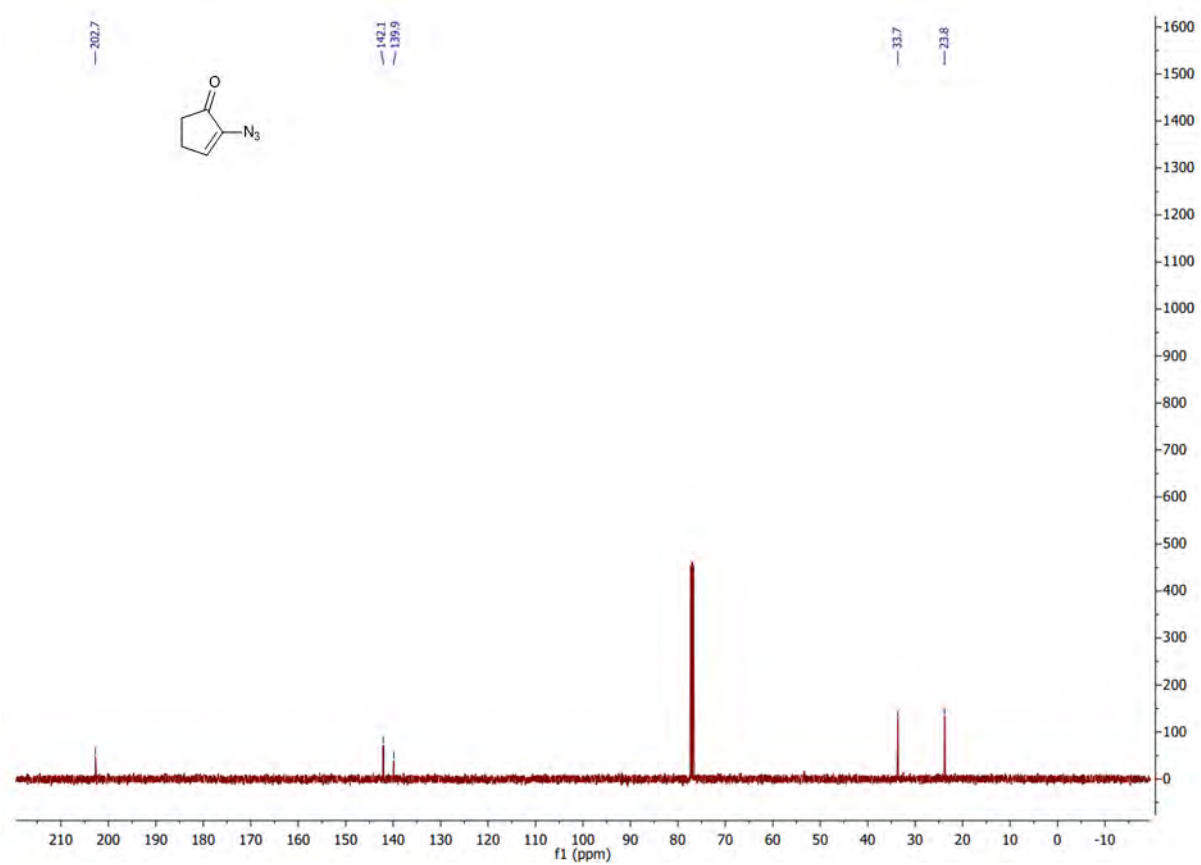
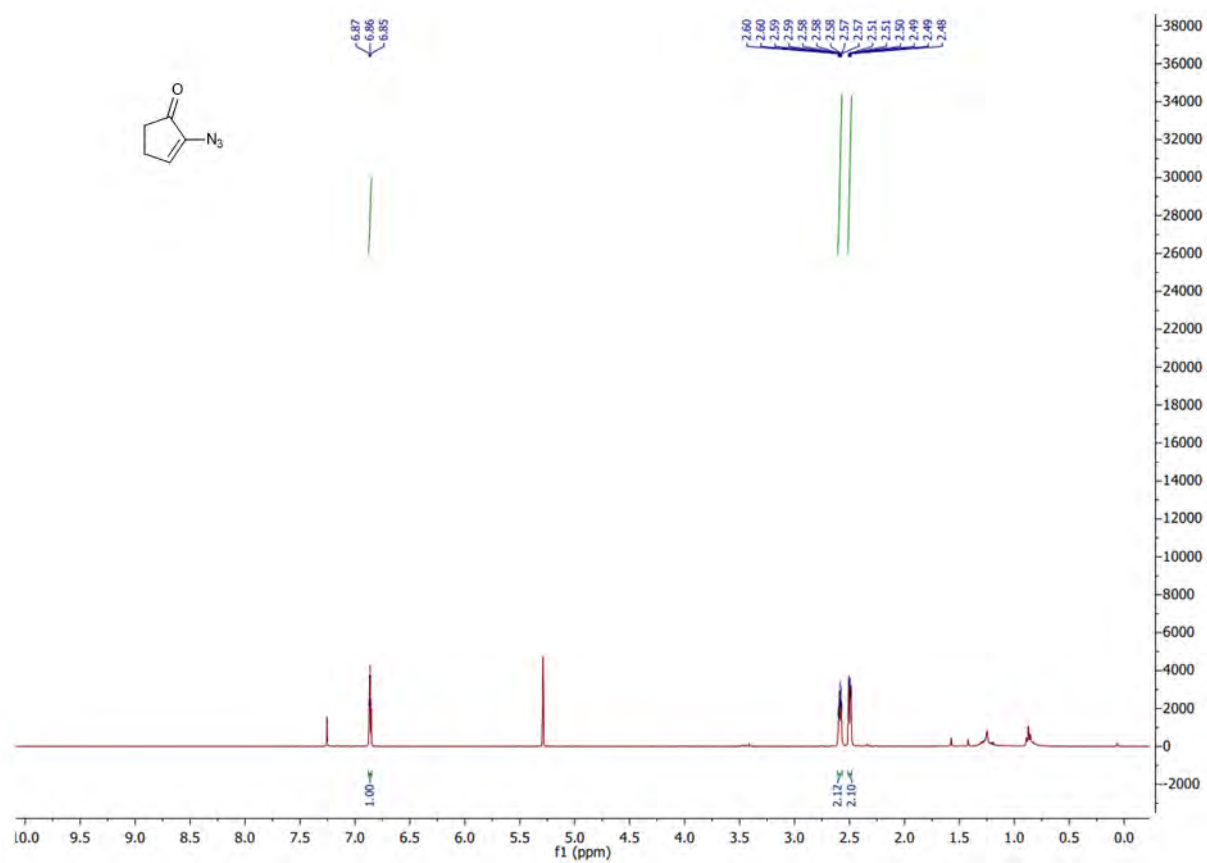
Supplementary Figure S151. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **8b**.



Supplementary Figure S152. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.10.1.

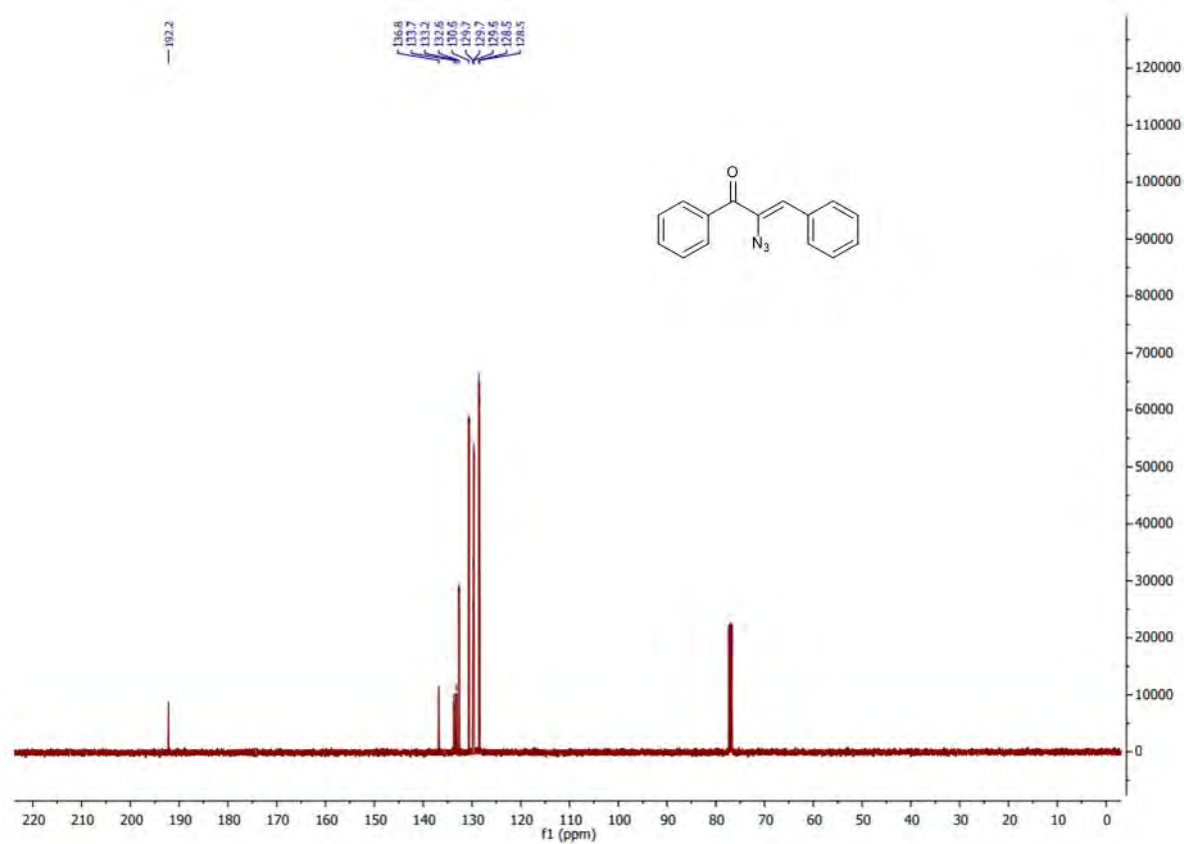
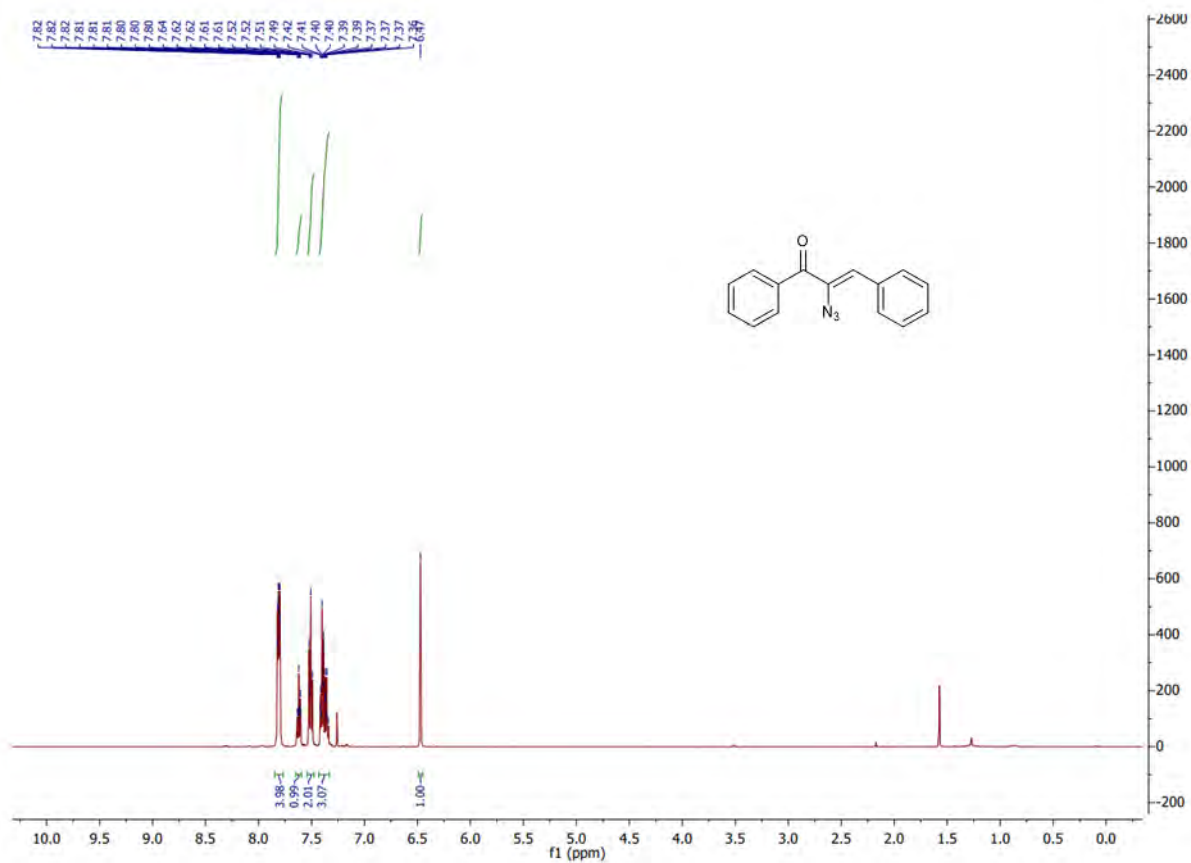


Supplementary Figure S153. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **8d**.

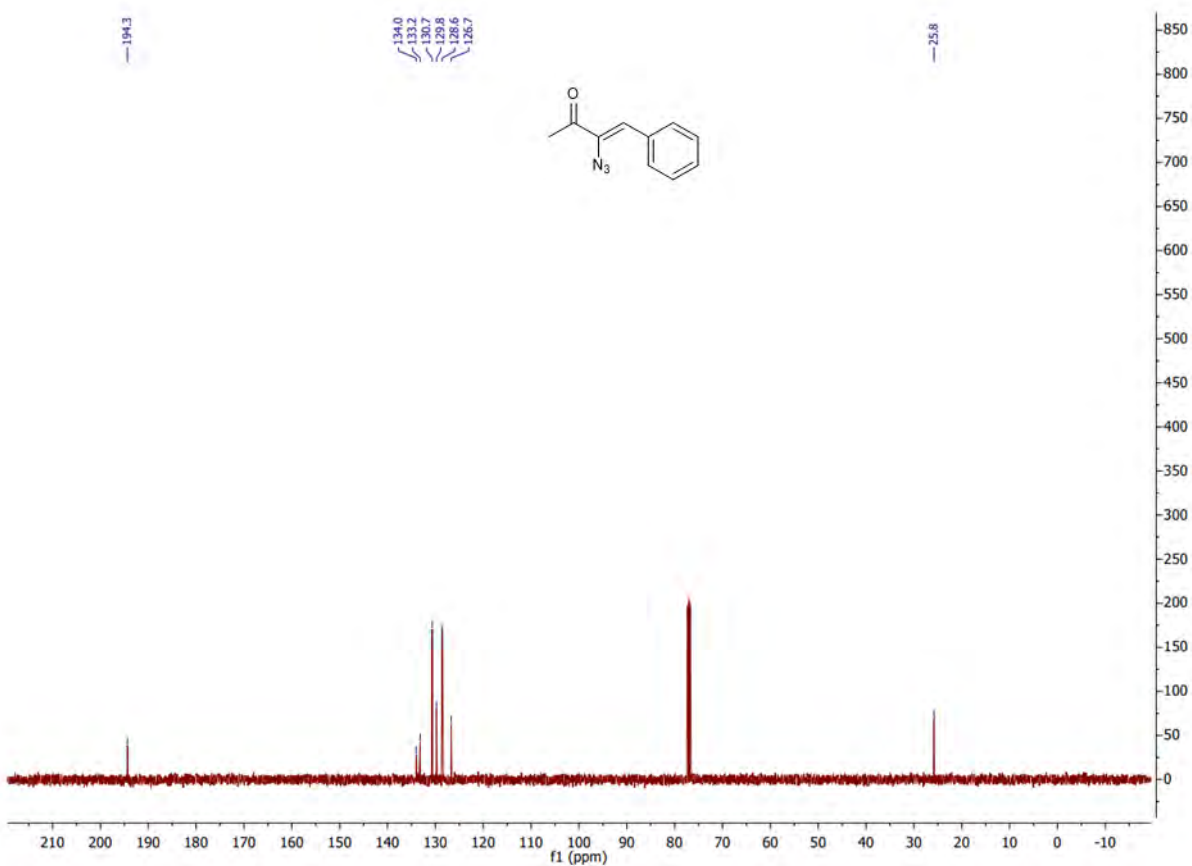
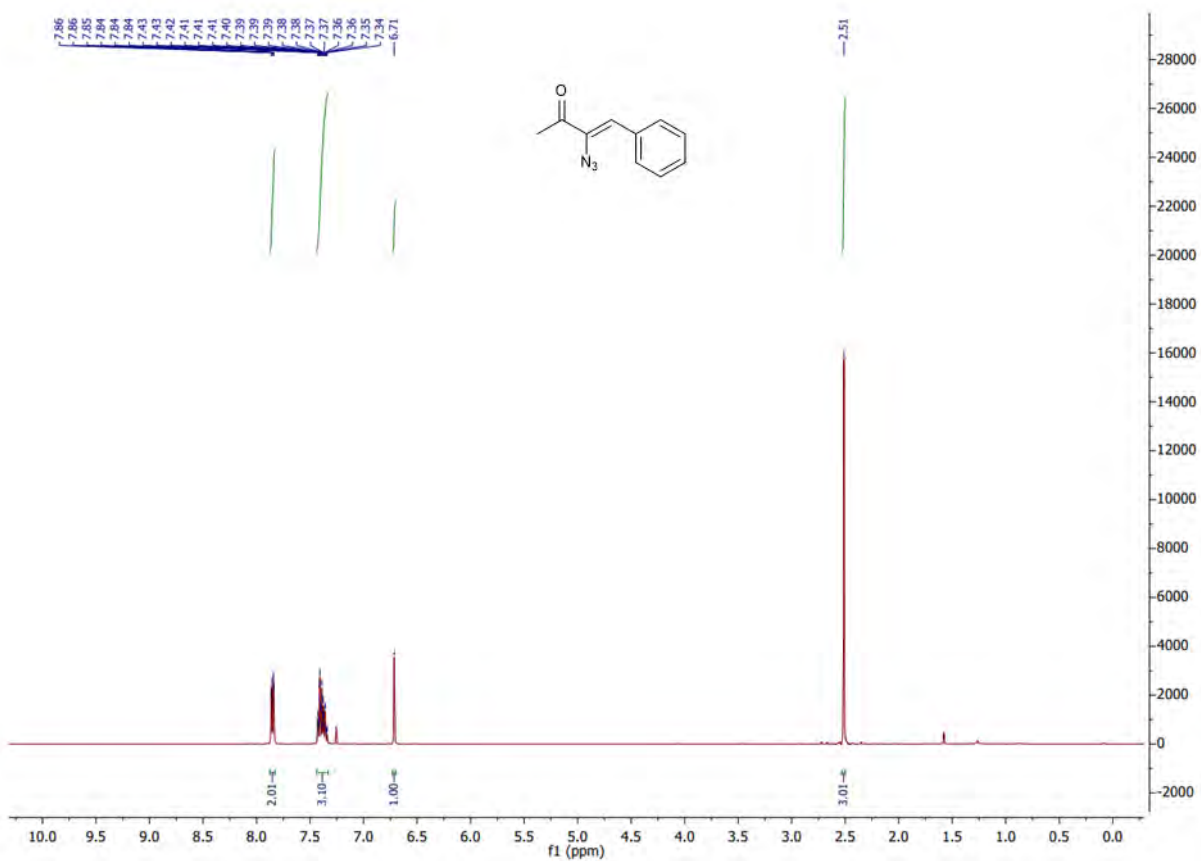


Supplementary Figure S154. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **8c**.



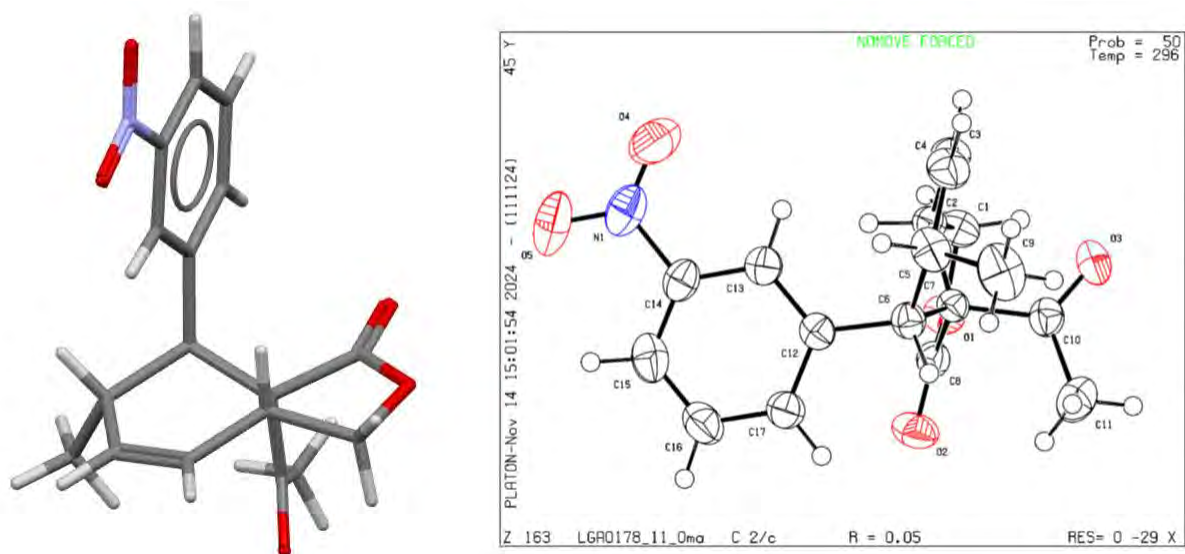


Supplementary Figure S155. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **8e**.



Supplementary Figure S156. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound **8f**.

## Section S6.11 X-ray crystallographic data for compound **6b'**



**Supplementary Figure S157.** X-ray structure of **6b'** (left) and probability ellipsoids of **6b'** (right)

### Experimental Summary

The measurement equipment was a Bruker X8 APEXII diffractometer using Cu-K $\alpha$  radiation ( $\lambda = 1.54178 \text{ \AA}$ ). A colorless needle-like specimen of C<sub>17</sub>H<sub>17</sub>NO<sub>5</sub>, approximate dimensions 0.078 mm x 0.147 mm x 0.645 mm, was used for the X-ray crystallographic analysis. The X-ray intensity data were measured.

A total of 4470 frames were collected. The total exposure time was 42.22 hours. The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using a monoclinic unit cell yielded a total of 43868 reflections to a maximum  $\theta$  angle of 68.50° (0.83 Å resolution), of which 2840 were independent (average redundancy 15.446, completeness = 99.4%,  $R_{\text{int}} = 8.10\%$ ,  $R_{\text{sig}} = 3.77\%$ ) and 1977 (69.61%) were greater than  $2\sigma(F^2)$ . The final cell constants of  $a = 27.1449(10) \text{ \AA}$ ,  $b = 7.8049(3) \text{ \AA}$ ,  $c = 18.5305(8) \text{ \AA}$ ,  $\beta = 127.874(4)^\circ$ , volume = 3099.0(3) Å<sup>3</sup>, are based upon the refinement of the XYZ-centroids of 9972 reflections above  $20 \sigma(I)$  with  $8.252^\circ < 2\theta < 125.5^\circ$ . Data were corrected for absorption effects using the numerical method (SADABS). The ratio of minimum to maximum apparent transmission was 0.742. The calculated minimum and maximum transmission coefficients (based on crystal size) are 0.6150 and 0.9380.

The structure was solved and refined using the Bruker SHELXTL Software Package, using the space group  $C 1 2/c 1$ , with  $Z = 8$  for the formula unit,  $C_{17}H_{17}NO_5$ . The final anisotropic full-matrix least-squares refinement on  $F^2$  with 265 variables converged at  $R1 = 4.76\%$ , for the observed data and  $wR2 = 12.51\%$  for all data. The goodness-of-fit was 1.056. The largest peak in the final difference electron density synthesis was  $0.182 \text{ e}^-/\text{\AA}^3$  and the largest hole was  $-0.275 \text{ e}^-/\text{\AA}^3$  with an RMS deviation of  $0.050 \text{ e}^-/\text{\AA}^3$ . On the basis of the final model, the calculated density was  $1.352 \text{ g/cm}^3$  and  $F(000)$ , 1328  $e^-$ .

**Supplementary Table S4. Sample and crystal data for LGA0178\_11.**

<b>Identification code</b>	LGA0178_11	
<b>Chemical formula</b>	$C_{17}H_{17}NO_5$	
<b>Formula weight</b>	315.31 g/mol	
<b>Temperature</b>	296(2) K	
<b>Wavelength</b>	1.54178 $\text{\AA}$	
<b>Crystal size</b>	0.078 x 0.147 x 0.645 mm	
<b>Crystal habit</b>	colorless needle	
<b>Crystal system</b>	monoclinic	
<b>Space group</b>	$C 1 2/c 1$	
<b>Unit cell dimensions</b>	$a = 27.1449(10) \text{\AA}$	$\alpha = 90^\circ$
	$b = 7.8049(3) \text{\AA}$	$\beta = 127.874(4)^\circ$
	$c = 18.5305(8) \text{\AA}$	$\gamma = 90^\circ$
<b>Volume</b>	$3099.0(3) \text{\AA}^3$	
<b>Z</b>	8	
<b>Density (calculated)</b>	$1.352 \text{ g/cm}^3$	
<b>Absorption coefficient</b>	$0.834 \text{ mm}^{-1}$	
<b>F(000)</b>	1328	
<b>Theta range for data collection</b>	4.13 to $68.50^\circ$	
<b>Index ranges</b>	$-31 \leq h \leq 32$ , $-9 \leq k \leq 9$ , $-22 \leq l \leq 22$	
<b>Reflections collected</b>	43868	
<b>Independent reflections</b>	2840 [ $R(\text{int}) = 0.0810$ ]	
<b>Coverage of independent reflections</b>	99.4%	
<b>Absorption correction</b>	numerical	
<b>Max. and min. transmission</b>	0.9380 and 0.6150	
<b>Structure solution technique</b>	direct methods	
<b>Structure solution program</b>	SHELXL-2014 (Sheldrick, 2014)	
<b>Refinement method</b>	Full-matrix least-squares on $F^2$	

<b>Refinement program</b>	SHELXL-2014 (Sheldrick, 2014)
<b>Function minimized</b>	$\Sigma w(F_o^2 - F_c^2)^2$
<b>Data / restraints / parameters</b>	2840 / 0 / 265
<b>Goodness-of-fit on F<sup>2</sup></b>	1.056
	1977
<b>Final R indices</b>	data; R1 = 0.0476, wR2 = 0.1101 I>2 $\sigma$ (I)
	all data R1 = 0.0757, wR2 = 0.1251
<b>Weighting scheme</b>	w=1/[ $\sigma^2(F_o^2)+(0.0552P)^2+1.9503P$ ] where P=(F <sub>o</sub> <sup>2</sup> +2F <sub>c</sub> <sup>2</sup> )/3
<b>Largest diff. peak and hole</b>	0.182 and -0.275 eÅ <sup>-3</sup>
<b>R.M.S. deviation from mean</b>	0.050 eÅ <sup>-3</sup>

**Supplementary Table S5. Atomic coordinates and equivalent isotropic atomic displacement parameters (Å<sup>2</sup>) for LGA0178 11.**

U(eq) is defined as one third of the trace of the orthogonalized U<sub>ij</sub> tensor.

	x/a	y/b	z/c	U(eq)
O1	0.21357(6)	0.33003(18)	0.30271(9)	0.0417(4)
O2	0.24571(7)	0.5207(2)	0.25041(10)	0.0508(4)
O3	0.29123(8)	0.5276(2)	0.50973(11)	0.0543(5)
O4	0.43724(9)	0.8448(2)	0.41155(14)	0.0752(6)
O5	0.43193(10)	0.8372(2)	0.29060(15)	0.0829(6)
N1	0.42751(9)	0.9130(3)	0.34441(16)	0.0562(6)
C1	0.23992(10)	0.2291(3)	0.38538(15)	0.0417(5)
C2	0.30952(9)	0.2457(2)	0.43499(13)	0.0327(5)
C3	0.35810(11)	0.2067(3)	0.53351(15)	0.0437(6)
C4	0.41319(11)	0.2806(3)	0.57735(16)	0.0489(6)
C5	0.43148(10)	0.4142(3)	0.53875(15)	0.0447(6)
C6	0.38015(9)	0.4543(3)	0.43507(14)	0.0346(5)
C7	0.31529(9)	0.4274(2)	0.41169(12)	0.0298(4)
C8	0.25680(9)	0.4344(3)	0.31228(14)	0.0347(5)
C9	0.45512(15)	0.5748(4)	0.5983(2)	0.0656(8)
C10	0.30276(9)	0.5667(3)	0.45836(14)	0.0379(5)
C11	0.30161(13)	0.7498(3)	0.43328(18)	0.0595(7)
C12	0.38716(9)	0.3514(3)	0.37258(13)	0.0336(5)
C13	0.40579(9)	0.1811(3)	0.38882(15)	0.0377(5)
C14	0.40894(10)	0.0936(3)	0.32704(15)	0.0411(5)
C15	0.39501(11)	0.1698(3)	0.24940(17)	0.0512(6)
C16	0.37776(11)	0.3398(4)	0.23399(16)	0.0521(6)
C17	0.37413(10)	0.4294(3)	0.29515(15)	0.0420(5)

**Supplementary Table S6. Bond lengths (Å) for LGA0178 11.**

O1-C8	1.348(2)	O1-C1	1.461(2)
O2-C8	1.198(2)	O3-C10	1.209(2)
O4-N1	1.224(3)	O5-N1	1.227(2)
N1-C14	1.464(3)	C1-C2	1.516(3)
C1-H1A	1.01(2)	C1-H1B	0.98(2)
C2-C3	1.486(3)	C2-C7	1.518(3)
C2-H2	0.999(19)	C3-C4	1.318(3)
C3-H3	1.01(2)	C4-C5	1.510(3)
C4-H4	1.00(3)	C5-C9	1.527(4)
C5-C6	1.563(3)	C5-H5	0.98(2)
C6-C12	1.514(3)	C6-C7	1.548(3)
C6-H6	1.01(2)	C7-C8	1.525(3)
C7-C10	1.552(3)	C9-H9A	0.99(3)
C9-H9B	1.00(3)	C9-H9C	0.98(3)
C10-C11	1.497(3)	C11-H11A	0.96
C11-H11B	0.96	C11-H11C	0.96
C12-C13	1.388(3)	C12-C17	1.390(3)
C13-C14	1.380(3)	C13-H13	0.97(2)
C14-C15	1.378(3)	C15-C16	1.377(4)
C15-H15	0.98(3)	C16-C17	1.386(3)
C16-H16	0.96(2)	C17-H17	0.96(2)

**Supplementary Table S7. Bond angles (°) for LGA0178 11.**

C8-O1-C1	111.20(16)	O4-N1-O5	123.3(2)
O4-N1-C14	118.2(2)	O5-N1-C14	118.5(2)
O1-C1-C2	102.44(16)	O1-C1-H1A	106.4(12)
C2-C1-H1A	113.4(12)	O1-C1-H1B	107.9(12)
C2-C1-H1B	111.5(13)	H1A-C1-H1B	114.1(17)
C3-C2-C1	124.27(18)	C3-C2-C7	111.83(17)
C1-C2-C7	102.43(16)	C3-C2-H2	107.4(11)
C1-C2-H2	103.1(11)	C7-C2-H2	106.5(11)
C4-C3-C2	118.9(2)	C4-C3-H3	123.4(12)
C2-C3-H3	117.5(12)	C3-C4-C5	126.1(2)
C3-C4-H4	119.5(15)	C5-C4-H4	114.3(15)
C4-C5-C9	109.8(2)	C4-C5-C6	114.60(18)
C9-C5-C6	112.7(2)	C4-C5-H5	107.7(12)
C9-C5-H5	107.4(12)	C6-C5-H5	104.1(12)
C12-C6-C7	111.68(16)	C12-C6-C5	113.45(17)

C7-C6-C5	108.61(16)	C12-C6-H6	106.4(11)
C7-C6-H6	108.6(11)	C5-C6-H6	107.9(11)
C2-C7-C8	99.45(15)	C2-C7-C6	110.21(16)
C8-C7-C6	119.66(15)	C2-C7-C10	113.58(15)
C8-C7-C10	103.09(15)	C6-C7-C10	110.48(15)
O2-C8-O1	121.81(19)	O2-C8-C7	129.76(19)
O1-C8-C7	108.30(16)	C5-C9-H9A	111.7(18)
C5-C9-H9B	111.5(15)	H9A-C9-H9B	105.(2)
C5-C9-H9C	113.4(16)	H9A-C9-H9C	106.(2)
H9B-C9-H9C	108.(2)	O3-C10-C11	121.28(19)
O3-C10-C7	120.85(18)	C11-C10-C7	117.76(18)
C10-C11-H11A	109.5	C10-C11-H11B	109.5
H11A-C11-H11B	109.5	C10-C11-H11C	109.5
H11A-C11-H11C	109.5	H11B-C11-H11C	109.5
C13-C12-C17	118.12(19)	C13-C12-C6	122.85(18)
C17-C12-C6	119.03(19)	C14-C13-C12	119.6(2)
C14-C13-H13	118.7(13)	C12-C13-H13	121.7(13)
C15-C14-C13	122.3(2)	C15-C14-N1	118.8(2)
C13-C14-N1	118.8(2)	C16-C15-C14	118.2(2)
C16-C15-H15	125.2(15)	C14-C15-H15	116.5(15)
C15-C16-C17	120.1(2)	C15-C16-H16	122.2(14)
C17-C16-H16	117.6(15)	C16-C17-C12	121.5(2)
C16-C17-H17	119.5(12)	C12-C17-H17	119.0(12)

**Supplementary Table S8. Torsion angles (°) for LGA0178 11.**

C8-O1-C1-C2	-17.2(2)	O1-C1-C2-C3	162.89(18)
O1-C1-C2-C7	35.2(2)	C1-C2-C3-C4	-155.6(2)
C7-C2-C3-C4	-32.0(3)	C2-C3-C4-C5	4.7(4)
C3-C4-C5-C9	123.5(3)	C3-C4-C5-C6	-4.6(3)
C4-C5-C6-C12	-94.3(2)	C9-C5-C6-C12	139.2(2)
C4-C5-C6-C7	30.5(2)	C9-C5-C6-C7	-96.0(2)
C3-C2-C7-C8	-174.03(16)	C1-C2-C7-C8	-38.85(18)
C3-C2-C7-C6	59.4(2)	C1-C2-C7-C6	-165.42(16)
C3-C2-C7-C10	-65.2(2)	C1-C2-C7-C10	70.0(2)
C12-C6-C7-C2	68.3(2)	C5-C6-C7-C2	-57.6(2)
C12-C6-C7-C8	-46.0(2)	C5-C6-C7-C8	-171.83(17)
C12-C6-C7-C10	-165.36(16)	C5-C6-C7-C10	68.8(2)
C1-O1-C8-O2	175.55(19)	C1-O1-C8-C7	-8.3(2)
C2-C7-C8-O2	-154.3(2)	C6-C7-C8-O2	-34.5(3)
C10-C7-C8-O2	88.6(2)	C2-C7-C8-O1	29.88(19)

C6-C7-C8-O1	149.74(16)	C10-C7-C8-O1	-87.18(18)
C2-C7-C10-O3	1.8(3)	C8-C7-C10-O3	108.3(2)
C6-C7-C10-O3	-122.7(2)	C2-C7-C10-C11	-174.61(19)
C8-C7-C10-C11	-68.0(2)	C6-C7-C10-C11	61.0(2)
C7-C6-C12-C13	-84.9(2)	C5-C6-C12-C13	38.2(3)
C7-C6-C12-C17	94.7(2)	C5-C6-C12-C17	-142.2(2)
C17-C12-C13-C14	-2.0(3)	C6-C12-C13-C14	177.63(18)
C12-C13-C14-C15	0.9(3)	C12-C13-C14-N1	-178.82(18)
O4-N1-C14-C15	-177.0(2)	O5-N1-C14-C15	2.3(3)
O4-N1-C14-C13	2.8(3)	O5-N1-C14-C13	-178.0(2)
C13-C14-C15-C16	0.4(3)	N1-C14-C15-C16	-179.8(2)
C14-C15-C16-C17	-0.7(4)	C15-C16-C17-C12	-0.4(4)
C13-C12-C17-C16	1.8(3)	C6-C12-C17-C16	-177.85(19)

**Supplementary Table S9. Anisotropic atomic displacement parameters ( $\text{\AA}^2$ ) for LGA0178 11.**

The anisotropic atomic displacement factor exponent takes the form:  $-2\pi^2 [ h^2 a^{*2} U_{11} + \dots + 2 h k a^* b^* U_{12} ]$

	U <sub>11</sub>	U <sub>22</sub>	U <sub>33</sub>	U <sub>23</sub>	U <sub>13</sub>	U <sub>12</sub>
O1	0.0306(8)	0.0491(9)	0.0379(8)	0.0020(7)	0.0173(7)	-0.0019(7)
O2	0.0477(10)	0.0628(10)	0.0378(8)	0.0160(8)	0.0242(8)	0.0080(8)
O3	0.0729(12)	0.0567(10)	0.0539(10)	-0.0042(8)	0.0494(10)	0.0030(8)
O4	0.0855(15)	0.0509(11)	0.0891(15)	0.0124(10)	0.0536(13)	0.0144(10)
O5	0.0885(15)	0.0666(13)	0.1139(17)	-0.0234(12)	0.0725(14)	0.0075(11)
N1	0.0451(13)	0.0474(12)	0.0755(16)	-0.0090(12)	0.0367(12)	0.0020(10)
C1	0.0414(14)	0.0449(13)	0.0413(13)	0.0024(11)	0.0266(11)	-0.0037(11)
C2	0.0337(12)	0.0333(11)	0.0325(11)	0.0018(9)	0.0210(10)	0.0004(9)
C3	0.0467(15)	0.0457(13)	0.0408(13)	0.0110(10)	0.0279(12)	0.0098(11)
C4	0.0417(15)	0.0630(16)	0.0324(13)	0.0066(11)	0.0178(12)	0.0092(12)
C5	0.0301(13)	0.0626(15)	0.0352(12)	-0.0038(11)	0.0170(11)	-0.0011(11)
C6	0.0338(12)	0.0363(12)	0.0367(11)	-0.0018(9)	0.0231(10)	-0.0020(9)
C7	0.0287(11)	0.0333(11)	0.0275(10)	0.0006(8)	0.0173(9)	0.0010(8)
C8	0.0333(12)	0.0370(11)	0.0353(12)	0.0009(9)	0.0219(10)	0.0039(9)
C9	0.0507(18)	0.085(2)	0.0510(17)	-0.0242(16)	0.0260(15)	-0.0211(16)
C10	0.0358(12)	0.0410(12)	0.0351(12)	-0.0028(9)	0.0208(10)	0.0016(10)
C11	0.0758(19)	0.0397(13)	0.0735(18)	-0.0030(12)	0.0512(16)	0.0040(12)
C12	0.0259(11)	0.0404(11)	0.0349(11)	0.0004(9)	0.0188(10)	-0.0005(9)
C13	0.0291(12)	0.0429(13)	0.0409(13)	0.0011(10)	0.0215(10)	0.0008(9)
C14	0.0324(12)	0.0398(12)	0.0497(14)	-0.0036(10)	0.0246(11)	0.0014(9)
C15	0.0427(14)	0.0659(17)	0.0471(14)	-0.0089(12)	0.0286(12)	0.0024(12)
C16	0.0512(16)	0.0700(17)	0.0380(14)	0.0063(12)	0.0289(12)	0.0094(13)



	U <sub>11</sub>	U <sub>22</sub>	U <sub>33</sub>	U <sub>23</sub>	U <sub>13</sub>	U <sub>12</sub>
C17	0.0402(13)	0.0469(14)	0.0427(13)	0.0069(11)	0.0273(11)	0.0067(11)

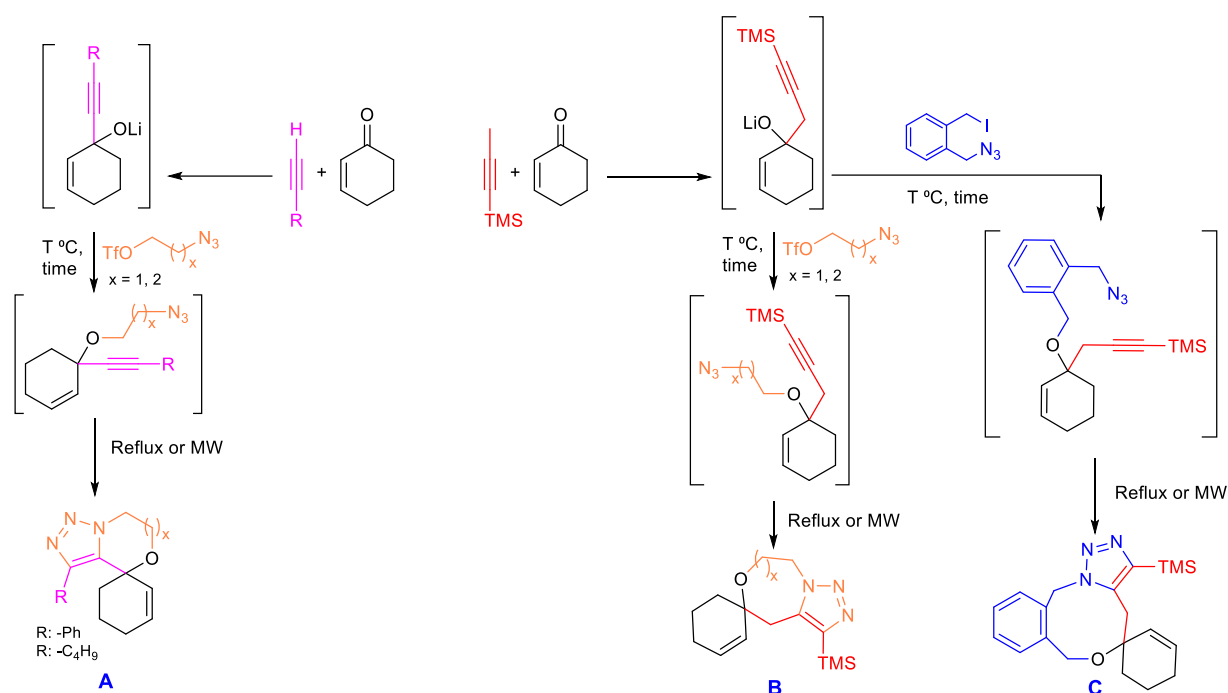
**Supplementary Table S10. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å<sup>2</sup>) for LGA0178 11.**

	x/a	y/b	z/c	U(eq)
H11A	0.2783	0.8168	0.4463	0.089
H11B	0.2823	0.7573	0.3693	0.089
H11C	0.3435	0.7926	0.4682	0.089
H2	0.3168(9)	0.168(2)	0.3996(13)	0.035(5)
H5	0.4663(10)	0.368(3)	0.5417(13)	0.040(6)
H6	0.3843(9)	0.578(3)	0.4247(13)	0.038(5)
H17	0.3638(10)	0.549(3)	0.2851(14)	0.042(6)
H3	0.3480(10)	0.117(3)	0.5619(14)	0.051(6)
H4	0.4465(12)	0.254(3)	0.6438(18)	0.068(7)
H1A	0.2255(10)	0.285(3)	0.4188(14)	0.047(6)
H1B	0.2257(10)	0.110(3)	0.3668(14)	0.050(6)
H13	0.4166(10)	0.121(3)	0.4428(16)	0.055(7)
H16	0.3693(11)	0.401(3)	0.1827(16)	0.059(7)
H15	0.4009(12)	0.099(3)	0.2115(17)	0.072(8)
H9A	0.4907(15)	0.549(4)	0.663(2)	0.091(10)
H9B	0.4716(13)	0.661(4)	0.5781(18)	0.074(9)
H9C	0.4231(13)	0.631(4)	0.5993(18)	0.079(9)

**Supplementary Table S11. Hydrogen bond distances (Å) and angles (°) for LGA0178 11.**

	Donor-H	Acceptor-H	Donor-Acceptor	Angle
C1-H1A...O3	1.01(2)	2.43(2)	2.956(3)	111.5(14)
C2-H2...O2	0.999(19)	2.476(19)	3.288(2)	138.0(15)
C11-H11B...O2	0.96	2.56	3.266(3)	130.4
C1-H1A...O3	1.01(2)	2.43(2)	2.956(3)	111.5(14)
C2-H2...O2	0.999(19)	2.476(19)	3.288(2)	138.0(15)
C11-H11B...O2	0.96	2.56	3.266(3)	130.4
C11-H11B...O2	0.96	2.56	3.266(3)	130.4
C2-H2...O2	0.999(19)	2.476(19)	3.288(2)	138.0(15)
C1-H1A...O3	1.01(2)	2.43(2)	2.956(3)	111.5(14)

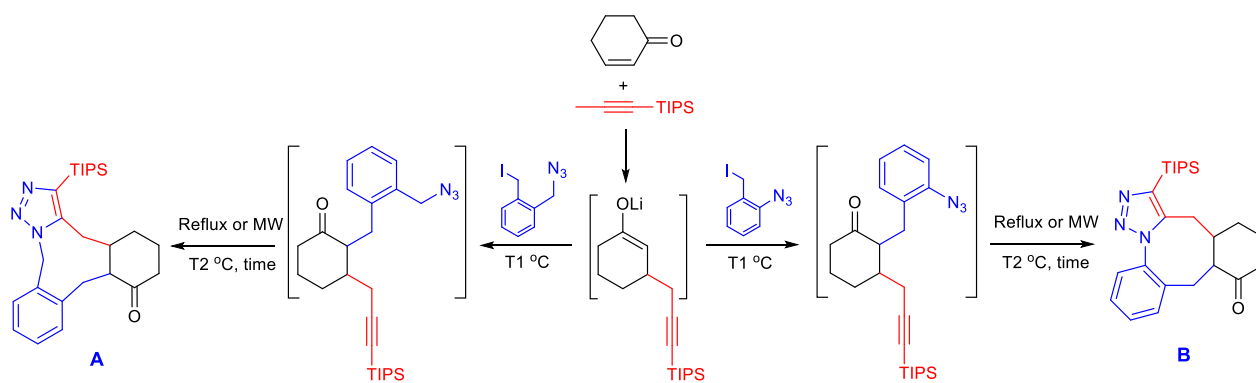
## S6.12. Reaction optimization tables.



Supplementary Table S12. Optimization table of reaction Mach1

Entry	Desired Product	Solvent(s)	T °C, time	Cyclization conditions	Yield %
1	A/B	THF	-20 °C to RT, up to 24 h	---	0%*
2	A/B/C	THF, Dioxane**	-20 °C to 0 °C, up to 24 h	Reflux	≤10**
3	A/B	THF, Dioxane**	-20 °C, 24 h	MW	≤10
4	C	THF, Dioxane**	50 °C, 12 h	MW	≤10
5	A	THF, Dioxane**	-20 °C, 24 h	Reflux	R = Ph, 33-38 R = C <sub>4</sub> H <sub>9</sub> , 34-44
6	B	THF, Dioxane**	-20 °C, 24 h	Reflux	X = 1, 12 X = 2, 34
7	C	THF, Dioxane**	50 °C, 12 h	Reflux	23

\* Polymerization of THF after the interaction with triflate at RT. \*\* Additional solvent was added to the reaction mixture prior to cyclizing. \*\*\* The lithium enolate in the first intermediate is quenched at 0 °C.

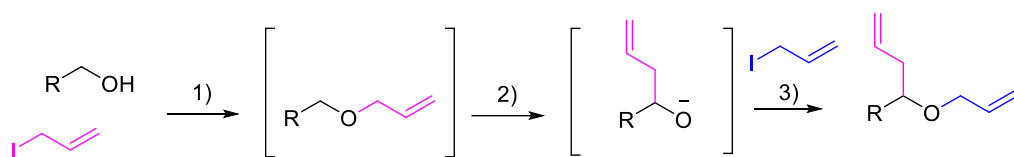


**Supplementary Table S13. Optimization table of reaction Mach2**

Entry	Desired Product	Solvent	T1 °C*	Cyclization conditions	T2 °C, time	Yield %
1	A	THF	-78	Reflux	120, 48 h	<10
2	A	THF, Dioxane**	-78	MW	145, 1 h	22
3	A	THF, Toluene***	-78	MW	180, 4 h	35
4	B	THF	-78	Reflux	120, 48 h	<10
5	B	THF, Dioxane**	-78	MW	185, 4 h	14

\*Lower T1 will result in a lower stepwise conversion and then a lower yield of the target compound.

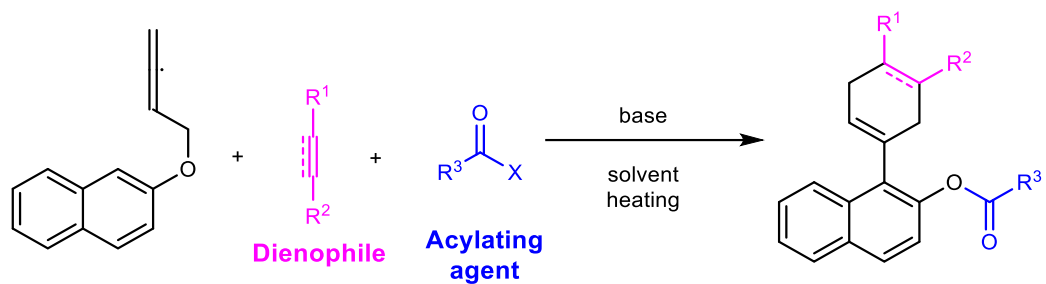
\*\* Additional solvent was added to the reaction mixture prior to cyclizing. \*\*\* The reaction mixture was evaporated and then redissolved in toluene prior to cyclizing. Although the yield was slightly higher, this modification of the condition does not fit in the formal "one-pot reaction" definition and consequently, is not reported as the optimal, one-pot yield.



**Supplementary Table S14. Optimization table of reaction Mach3**

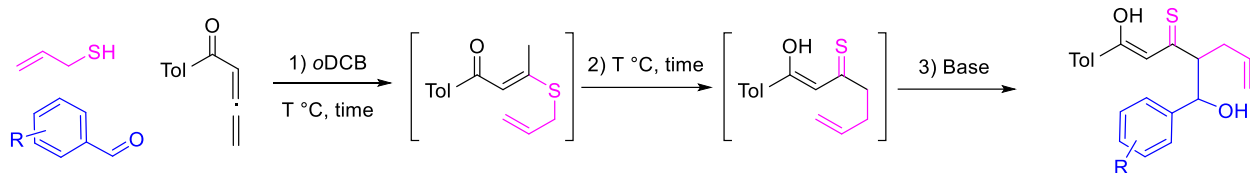
Entry	Base (equiv.)	Solvent	Condition	R	X (equiv.)	Yield
1	KOtBu (3)	THF (1M)	1) 0 °C to rt, 2 days 2) <i>n</i> -BuLi, -78°C to RT, ON	Ph	1) Br (2.2)	14%*
2	NaH (4)	THF (1M)	1) 0 °C to rt, 30 min 2) <i>n</i> -BuLi, -78°C to RT, ON 3) NaH (2), 0°C to RT, ON	Ph	1) Br (2.2)	40%*
3	NaH (4)	THF (1M)	1) 0 °C to rt, ON 2) LDA (2), -78°C to RT, ON	Ph	1) Br (4)	0%
4	NaH (4)	DMF (1M)	1) 0 °C to rt, 1 h 2) <i>n</i> BuLi, -78°C to RT, ON 3) NaH (2), 0°C to RT, ON	Ph	1) Br (4)	0%
5	NaH (1.5)	THF (1M)	1) 0 °C to rt, 1 h 2) <i>n</i> -BuLi, -78 °C, 0.5 h 3) HMPA (10), -78°C to RT, 0.5 h	Ph	1) I (2)** 2) - 3) I (2), 1 h	DP***
6	NaH (1.5)	THF (1M)	1) 0 °C to rt, 1 h 2) <i>n</i> -BuLi, -78 °C, 0.5 h 3) HMPA (10), -78°C to RT, 0.5 h	Cinnamyl	1) I (2) 3) I (2), 1 h	66%

DP: Desired Product. \* Only Wittig rearrangement product observed. \*\*Cl, Br, and OTs were also screened under the same conditions but were unsuccessful. \*\*\* DP could not be separated from impurities.



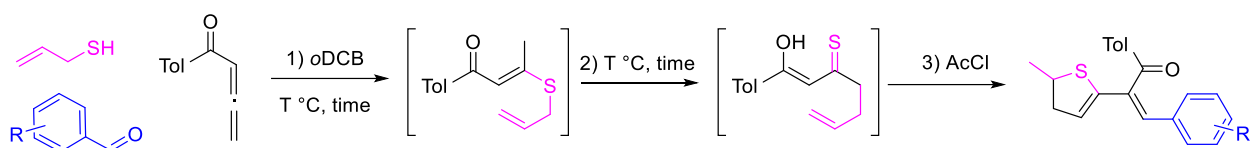
Supplementary Table S15. Optimization table of reaction Mach4

Entry	Dienophile	Acylating reagent	Base (equiv.)	Solvent	Condition	Yield
1	Dimethyl acetylene-dicarboxylate	Ac <sub>2</sub> O (2eq)	AcONa (1eq)	DCB	150°C, 3 h	84%
2	Dimethyl acetylene-dicarboxylate	Ac <sub>2</sub> O (2eq)	NaHCO <sub>3</sub> (1eq)	DCB	150°C, 2.5 h	94%
3	Diethyl acrylate	Ac <sub>2</sub> O (2eq)	AcONa (1eq)	DCB	150°C, 3 h	92%
4	<i>N</i> -phenyl-maleimide	BzCl (2eq)	NaHCO <sub>3</sub> (1eq)	DCB	150°C, 3.5 h	60%
5	<i>N</i> -phenyl-maleimide	Ac <sub>2</sub> O (2eq)	NaHCO <sub>3</sub> (1eq)	1,3,5-Trimethyl benzene	150°C, 5 h	77%
6	<i>N</i> -phenyl-maleimide	Ac <sub>2</sub> O (2eq)	NaHCO <sub>3</sub> (1eq)	DCB	150°C, 2.5 h	96%



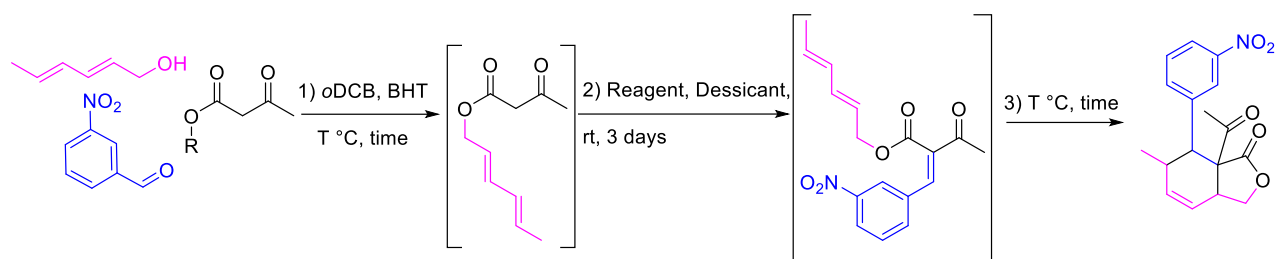
**Supplementary Table S16. Optimization table of reaction Mach5**

Entry	Base (equiv.)	Solvent (C)	T °C	R	Stepwise Conversion	Yield
1	DBU (1)	<i>o</i> -DCB (0.3M)	1) RT, ON	<i>p</i> -NO <sub>2</sub>	Complete	35%
			2) 180 °C, 10 min			
			3) RT, 15min			
2	Et <sub>3</sub> N (1)	<i>o</i> -DCB (0.3M)	1) RT, ON	<i>p</i> -NO <sub>2</sub>	Not Complete	Trace
			2) 180 °C, 10 min			
			3) RT, 15min			
3	DBU (1)	<i>o</i> -DCB (0.15M)	1) RT, ON	<i>p</i> -NO <sub>2</sub>	Complete	23%
			2) 180 °C, 10 min			
			3) RT, 15min			
4	DBU (1)	Toluene (0.3M)	1) RT, ON	<i>p</i> -NO <sub>2</sub>	Not Complete	0%
			2) 120 °C, 10 min			
			3) RT, 15min			
5	DBU (1)	<i>o</i> -DCB (0.3M)	1) RT, ON	<i>p</i> -CF <sub>3</sub>	Complete	36%
			2) 180°C, 10 min			
			3) RT, 15 min			
6	DBU (1)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -CF <sub>3</sub>	Complete	57%
			2) 180 °C, 10 min			
			3) RT, 15min			



**Supplementary Table S17. Optimization table of reaction Mach6**

Entry	Reagent (equiv.)	Solvent (C)	T °C	R	Conversion	Yield
1	AcCl (1)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -NO <sub>2</sub>	Complete	8%
			2) 180 °C, 10 min			
			3) RT, 18 h			
2	AcCl (1.5)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -NO <sub>2</sub>	Complete	5%
			2) 180 °C, 10 min			
			3) RT, 18 h			
3	AcCl (1) + MgSO <sub>4</sub> (or MS 4Å)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -NO <sub>2</sub>	Complete	0%
			2) 180 °C, 10 min			
			3) RT, 18 h			
4	PivCl (1)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -NO <sub>2</sub>	Not Complete	Trace
			2) 180 °C, 10 min			
			3) RT, ON			
5	AcCl (1)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -NO <sub>2</sub>	Complete	~11%
			2) 180 °C, 10 min			
			3) 90 °C, 4 h			
6	AcCl (1)	<i>o</i> -DCB (0.3M)	1) 40 °C, 1.5 h	<i>p</i> -CN	Complete	14%
			2) 180 °C, 10 min			
			3) RT, 18 h			

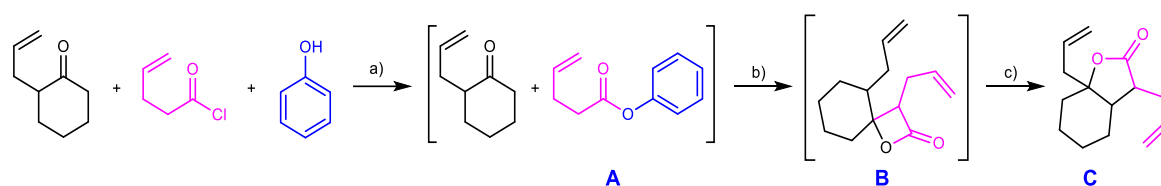


**Supplementary Table S18. Optimization table of reaction Mach7**

Entry	Solvent	T °C	Desiccant	Reagent	R	Yield
1	<i>o</i> -DCB (0.3M)	1) 150 °C, 2 h	Molecular sieves 4Å	1) Dowex (10 eq) 2) Piperidine (0.2 eq)	Me	Trace DP*
		2) 185 °C, ON				
		3) -				
2	<i>o</i> -DCB (0.3M)	1) 150 °C, 2h	Molecular sieves 4Å	1) Dowex (10 eq) 2) Piperidine (0.2 eq)	Me	13%
		2) RT, 2 days				
		3) 185 °C, 2.5 h				
3	<i>o</i> -DCB (0.3M)	1) 150 °C, 1 h	Molecular sieves 4Å	1) Dowex (10 eq) 2) Piperidinium acetate (0.2 eq)	Me	15%
		2) RT, 2 days				
		3) 185 °C, 4 h				
4	Toluene (0.3M)	1) 150 °C, 1 h	Molecular sieves 4Å	1) Dowex (10 eq) 2) Piperidinium acetate (0.2 eq)	Me	8%
		2) RT, 3 days				
		3) 185 °C, 4 h				
5	<i>o</i> -DCB (0.3M)	1) RT, ON	Molecular sieves 4Å	1) DMAP (0.2 eq.) 2) Piperidinium acetate (0.2 eq)	Ph	0% **
		2) RT, 3 days				
		3) 185 °C, 4 h				
6	<i>o</i> -DCB (0.3M)	1) 100 °C, 1 h	MgSO <sub>4</sub> (2 eq)	1) None 2) Piperidinium acetate (0.2 eq)	Ph	30%
		2) RT, 3 days				
		3) 210 °C, 30 min				
7	<i>o</i> -DCB (0.3M)	1) 100 °C, 1h	MgSO <sub>4</sub> (2 eq)	1) None 2) Piperidinium acetate (0.2 eq)	Ph	43%
		2) RT, 3 days				
		3) 185 °C, 4 h				
8	<i>o</i> -DCB (0.3M)	1) 100 °C, 1h	MgSO <sub>4</sub> (3.9 eq)	1) None 2) Piperidinium acetate (0.2 eq)	Ph	44%
		2) RT, 3 days				
		3) 210 °C, 45 min				

\*DP: Desired Product. \*\*Only decarboxylation

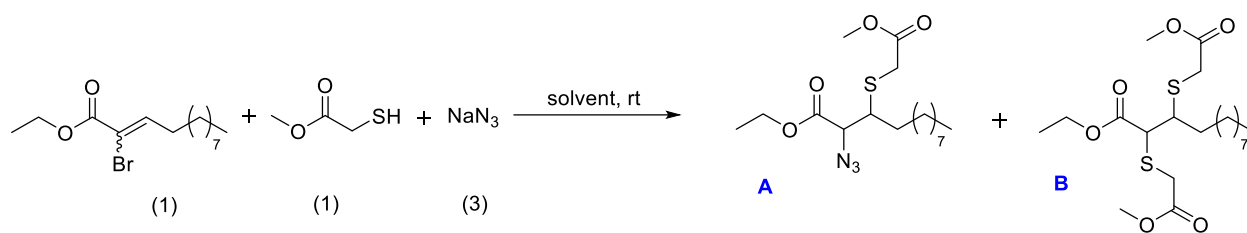




**Supplementary Table S19. Optimization table of reaction Mach8/9**

Entry	Solvent	Reagents (equiv.)	T °C, time	Yield %
1	Et <sub>2</sub> O (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	A: 98%
2	Et <sub>2</sub> O (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	B: 21%
		2) <i>n</i> -BuLi (1.5), DIPA (1.65)	2) -78°C, 0.5h then to -30°C in 1.5h	
3	Et <sub>2</sub> O (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	B: 60%
		2) NaHMDS (1.5)	2) -78°C, 0.5h then to -30°C in 1.5h	
4	Et <sub>2</sub> O (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	C: 21%
		2) NaHMDS (1.5)	2) -78°C, 0.5h then to -30°C in 1.5h	
		3) MgBr <sub>2</sub> •Et <sub>2</sub> O (1.5)	3) -30°C to rt in 3h, then rt, 18h	
5	Et <sub>2</sub> O (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	C: 31%
		2) NaHMDS (1.5)	2) -78°C, 0.5h then to -30°C in 1.5h	
		3) MgBr <sub>2</sub> •Et <sub>2</sub> O (5)	3) -30°C to rt in 3h, then rt, 18h	
6	THF (0.2 M)	1) Pyridine (2)	1) 0°C to rt in 1h then 22h rt	C: 17%
		2) NaHMDS (1.5)	2) -78°C, 0.5h then to -30°C in 1.5h	
		3) MgBr <sub>2</sub> •Et <sub>2</sub> O (5)	3) -30°C to rt in 3h, then rt, 18h	

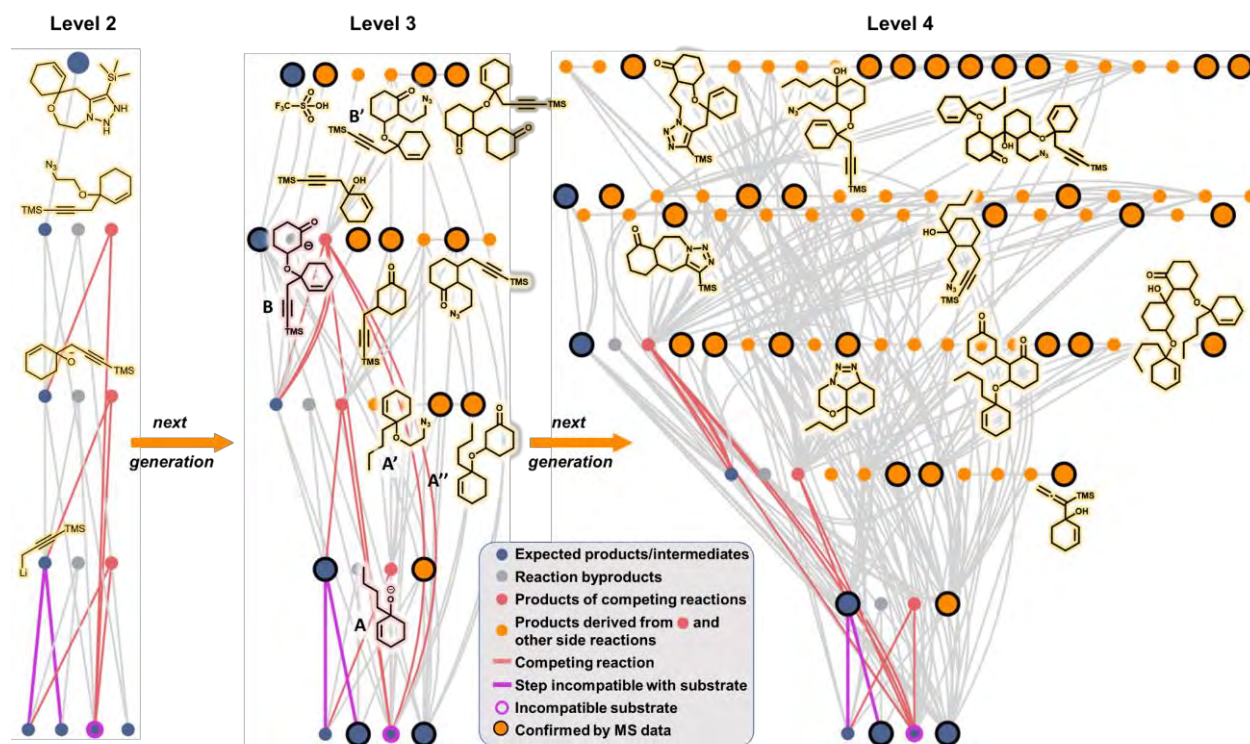
\*Mach9 continues the thus optimized sequence with straightforward Heck coupling under efficient conditions reported in the main-text Figure 5 (no additional optimization).



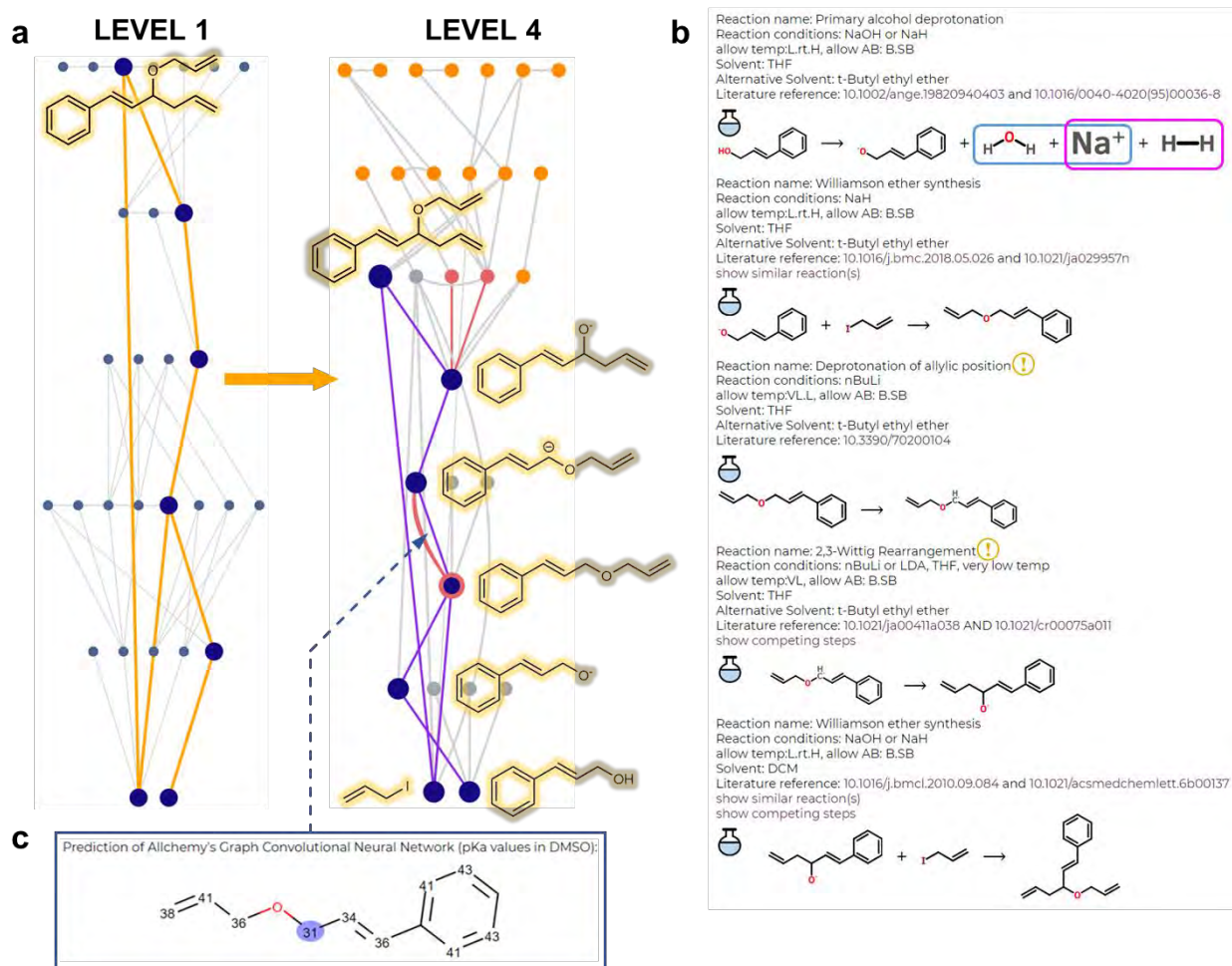
**Supplementary Table S20. Optimization table of reaction Mach10**

Entry	Solvent	Reagents (equiv.)	Time	A	B
1	DMF	-	3 d	37%	23%
2		-	3 h at $-35\text{ }^\circ\text{C}$ then 17 h at rt	45%	18%
3	DMSO	Extra $\text{NaN}_3$ (1)	3 d	40%	20%
4		-	4 d	32%	21%
5	MeCN	-	3 d	-	-
6		$\text{H}_2\text{O}$ (1)	4 d	17%	14%
7	Acetone	-	3 d	-	-
8		$\text{H}_2\text{O}$ (1)	15 h	28%	15%
9	1,4-dioxane	-	3 d	-	-
10		$\text{H}_2\text{O}$ (1)	4 d	16%	25%
11	Benzene	TBAB (0.1)	3 d	23%	21%
12	MeOH	-	3 d	<b>64%</b>	13%
13	EtOH	-	3 d	59%	20%

### S6.13. Allchemy screenshots of mechanistic networks and reaction pathways.

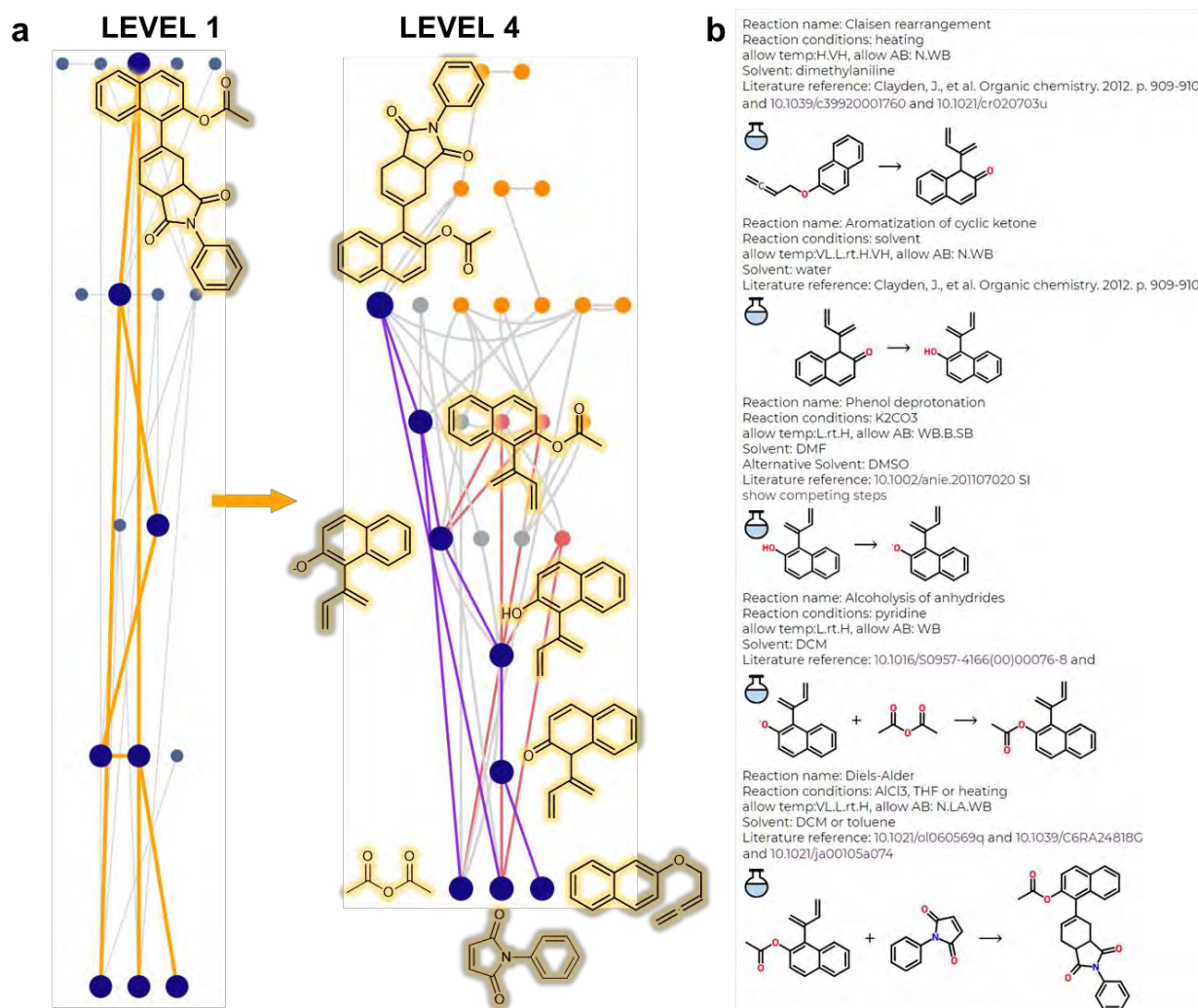


**Supplementary Figure S158. Different levels of analysis and additional ESI-MS assignments for the spirocyclic triazole path from Figure 3.** Allchemy MECH module screenshots of networks at Levels 2,3,4 with larger orange nodes denoting likely structural assignments (overlaid as highlighted structures, in addition to structures already shown in the main-text **Figure 3b**) of peaks observed in the ESI-MS of the crude-reaction mixture.



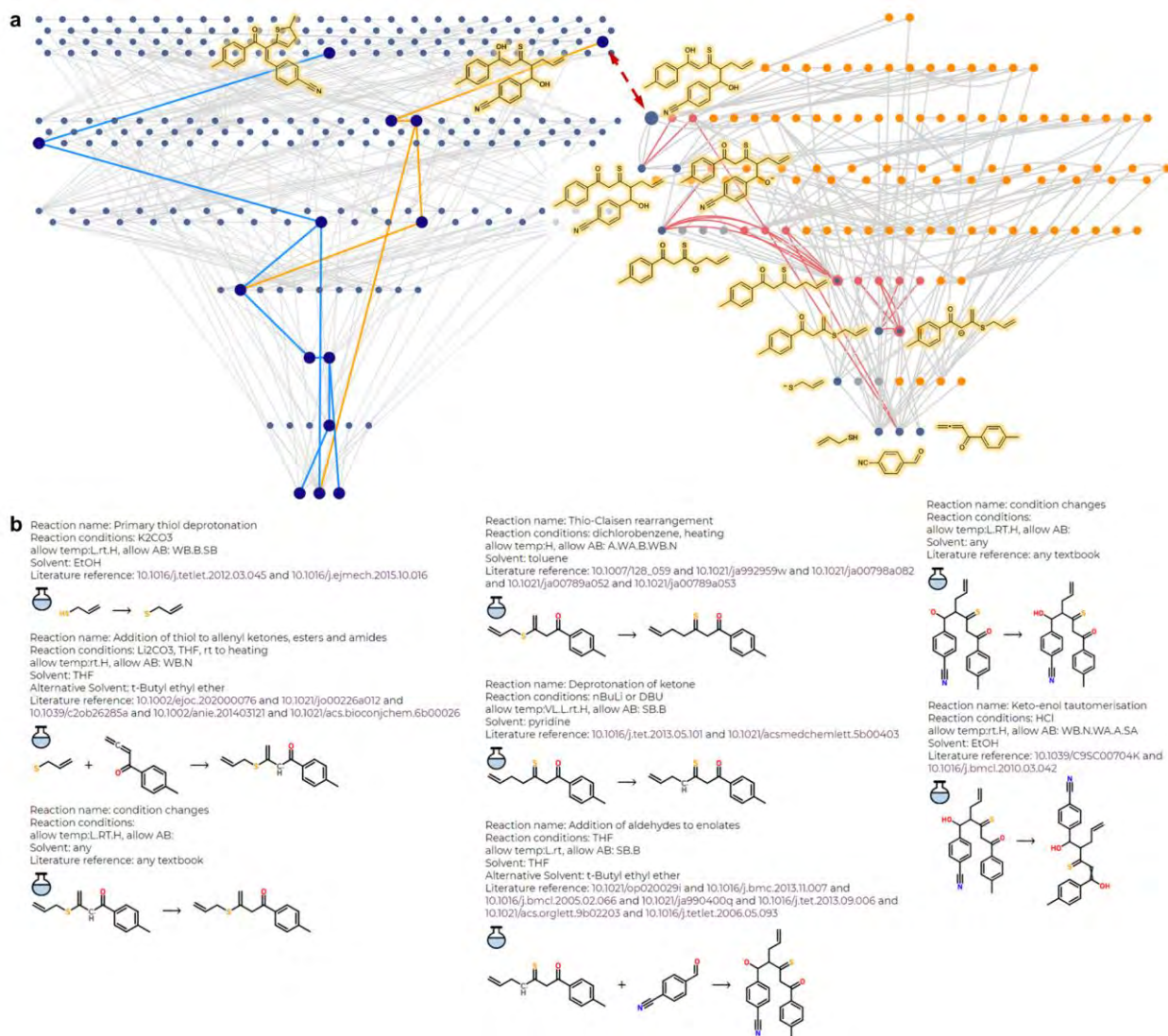
**Supplementary Figure S159. One-pot sequence from main-text Figure 4a, starting from allyl iodide and cinnamyl alcohol.** The sequence is presented as **a**, Level 1 and Level 4 networks and, **b**, list of mechanistic steps (both networks and the list are screenshots from Allchemy's MECH module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative literature references (i.e., publications(s) in which this type of chemistry was used, albeit not for the specific substrates, as Allchemy does not use literature precedents but generalized mechanistic steps; see Methods). Another hyperlink allows visualizing competing reactions whereas clicking the “flask” icon expands a given step to show its byproducts (here, for deprotonation of primary alcohols step, indicated by blue and pink frames, corresponding to alternative reagents for this reaction). The exclamation mark points to the key cross-reactivity conflict which renders this sequence one-pot rather than MCR

(excessive allyl iodide would react with *n*-butyllithium, hampering deprotonation and subsequent Wittig rearrangement. **c**, Allchemy's pKa filter predicting correctly the most acidic CH position in the deprotonation step.



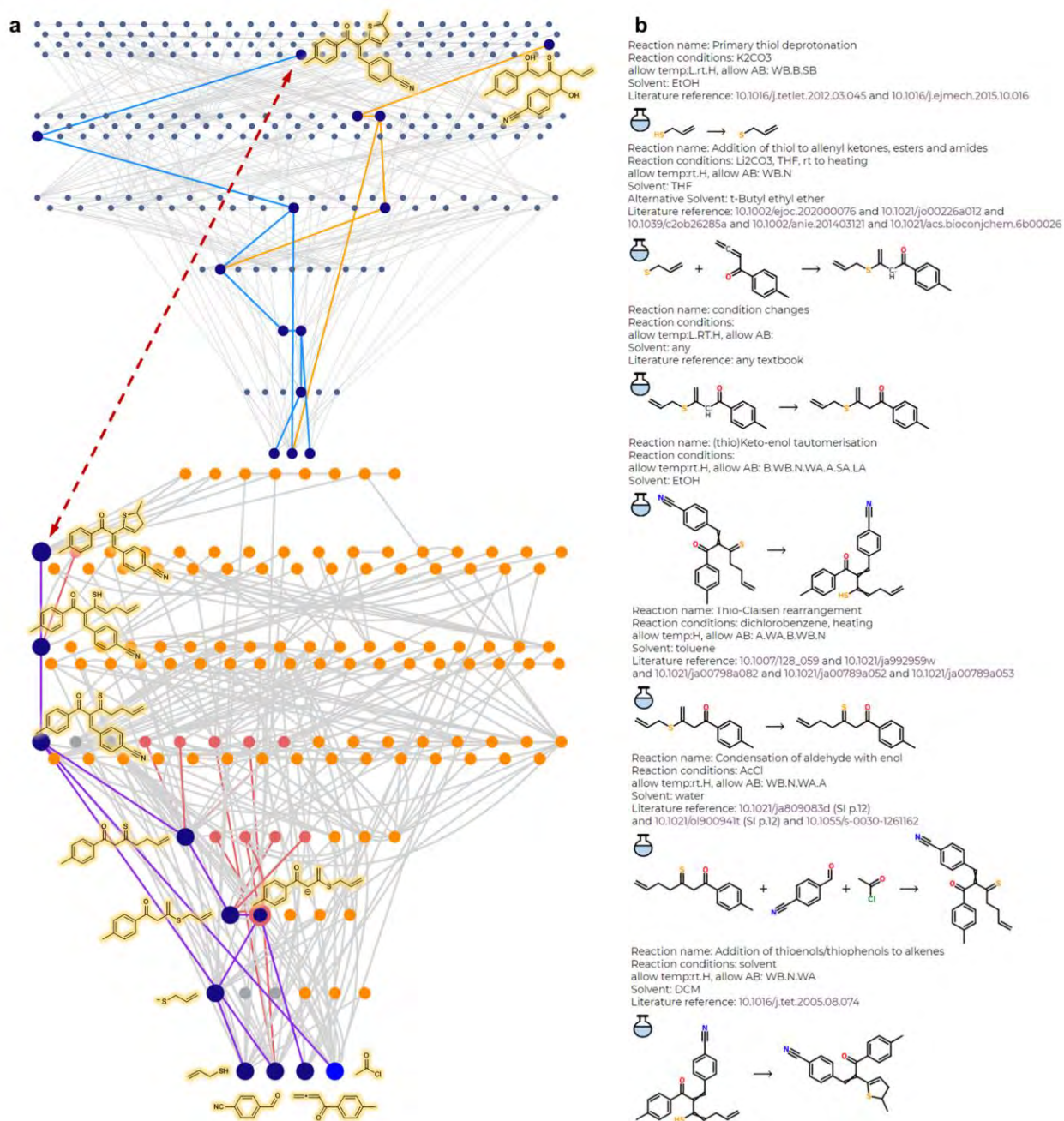
**Supplementary Figure S160. Multicomponent reaction from Figure 4c, starting from 2-(buta-2,3-dien-1-yloxy)naphthalene, *N*-phenylmaleimide and acetic anhydride. This MCR is presented here as **a**, Level 1 and Level 4 networks and **b**, list of mechanistic steps (both networks and the list are screenshots from Allchemy's MECH module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative**

literature references. As in Extended Figure 2, by-products of each reaction can be visualized by clicking on the “flask” icon.



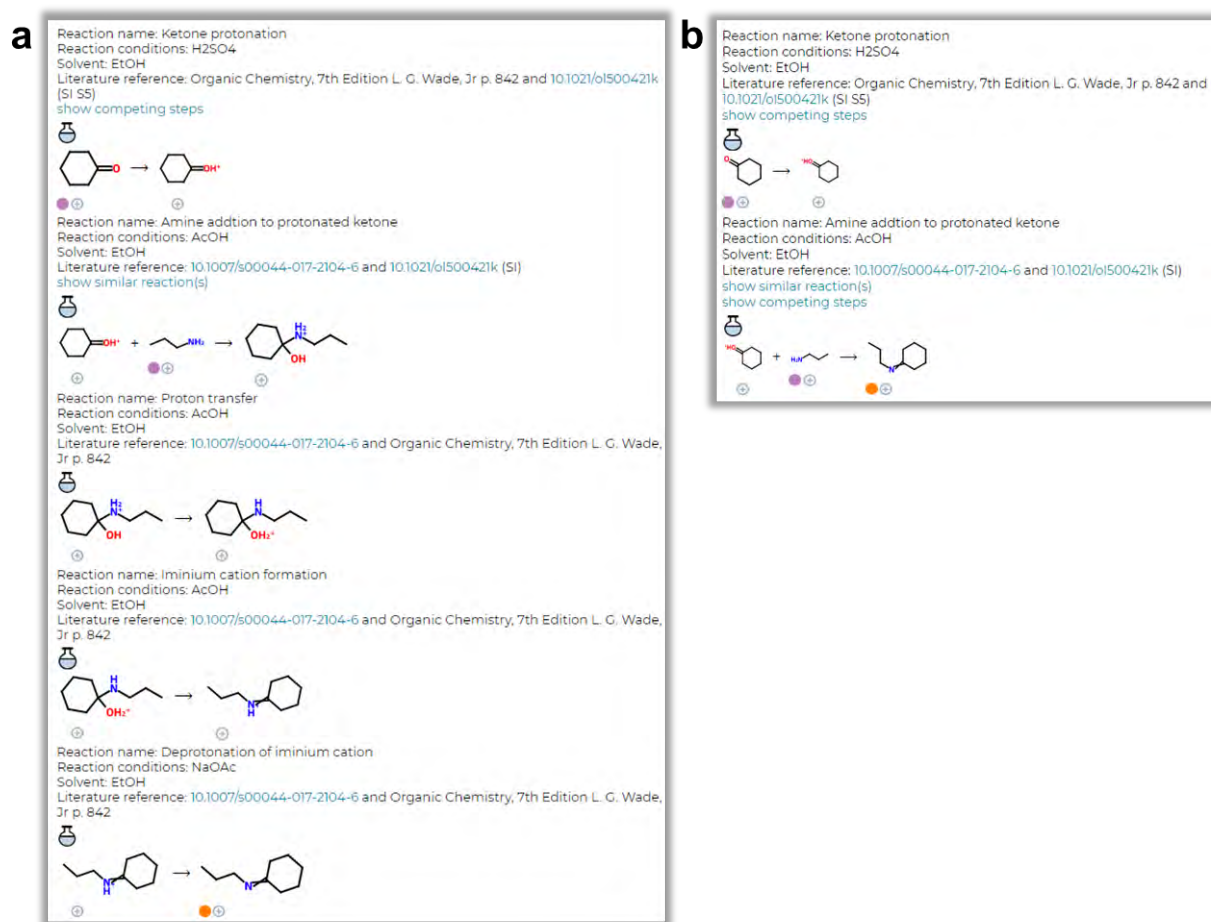
**Supplementary Figure S161. The first of the two MCRs from Figure 4e, starting from allyl thiol, 4-cyanobenzaldehyde and 1-(p-tolyl)buta-2,3-dien-1-one, catalyzed by DBU.** This MCR is presented here as **a**, Level 1 and Level 4 networks and, **b**, list of mechanistic steps (both networks and the list are screenshots from Allchemy’s MECH module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative literature references. As in Extended Figure 2, by-products of each reaction can be visualized

by clicking on the “flask” icon. In Level 4 network, terminal nodes corresponding to unstable molecules (charged, enols) that were not involved in any further reactions are not shown for clarity.



**Supplementary Figure S162. The second MCR from Figure 4e, starting from allyl thiol, 4-cyanobenzaldehyde and 1-(p-tolyl)buta-2,3-dien-1-one, catalyzed by acyl chloride. This MCR is presented here as a, Level 1 and Level 4 networks and, b, list of mechanistic steps**

(both networks and the list are screenshots from Allchemy's MECH module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative literature references. As in Extended Figure 2, by-products of each reaction can be visualized by clicking on the “flask” icon. In Level 4 network, terminal nodes corresponding to unstable molecules (charged, enols) that were not involved in any further reactions are not shown for clarity.



**Supplementary Figure S163. Individual mechanistic steps vs. “supersteps”.** Allchemy screenshots illustrating **a**, Imine formation divided into all mechanistic steps; and **b**, Imine formation with “short-cut” combining *Amine addition to protonated ketone*, *Proton transfer*, *Iminium cation formation* and *Deprotonation of iminium cation* into one “superstep”.



## Section S7. Supplementary References

1. Roszak R., Beker W., Molga K. & Grzybowski B. A. *J. Am. Chem. Soc.* **141**, 17142–17149 (2019).
2. Banfi, L., Basso, A., Lambruschini, C., Moni, L. & Riva, R. The 100 facets of the Passerini reaction. *Chem. Sci.* **12**, 15445-15472 (2021).
3. Reddy, D. S., Judd, W. R. & Aubé, J. Lewis Acid-Mediated Reactions of Alkyl Azides with  $\alpha,\beta$ -Unsaturated Ketones. *Org. Lett.* **5**, 3899-3902 (2003).
4. Kolarovič, A., Schnürch, M., & Mihovilovic, M. D. Tandem Catalysis: From Alkynoic Acids and Aryl Iodides to 1,2,3-Triazoles in One Pot. *J. Org. Chem.* **76**, 2613-2618 (2011).
5. Xu, M., Kuang, C., Wang, Z., Yang, Q., & Jiang, Y. A Novel Approach to 1-Monosubstituted 1,2,3-Triazoles by a Click Cycloaddition/Decarboxylation Process. *Synthesis* **2**, 223-228 (2010).
6. Spiteri, C. & Moses, J. E. Copper-Catalyzed Azide–Alkyne Cycloaddition: Regioselective Synthesis of 1,4,5-Trisubstituted 1,2,3-Triazoles. *Angew. Chem. Int. Ed.* **49**, 31-33 (2010).
7. Worell, B. T., Malik, J. A., & Fokin, V. V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science* **340**, 457-460 (2013)
8. Kee, J.-M., Villani, B., Carpenter, L. R., & Muir, T. W. Development of Stable Phosphohistidine Analogues. *J. Am. Chem. Soc.* **132**, 14327-14329 (2010)
9. Majireck, M. M. & Weinreb, S. M. A Study of the Scope and Regioselectivity of the Ruthenium-Catalyzed [3+2]-Cycloaddition of Azides with Internal Alkynes. *J. Org. Chem.* **71**, 8680-8683 (2006).
10. Johansson, J. R., Beke-Somfai, T., Said Stålsmeden, A., & Kann N. Ruthenium-Catalyzed Azide Alkyne Cycloaddition Reaction: Scope, Mechanism, and Applications. *Chem. Rev.* **116**, 14726-14768 (2016).
11. Singleton, E. and Hester E. Oosthuizen. "Metal isocyanide complexes." *Advances in organometallic chemistry*. Vol. 22. Academic Press, 1983, 209-310.
12. Klucznik, T., Syntrivanis, LD., Baś, S. *et al.* Computational prediction of complex cationic rearrangement outcomes. *Nature* **625**, 508–515 (2024).

13. Hare, S. R., Pemberton, R. P. & Tantillo, D. J. Navigating past a fork in the road: carbocation- $\pi$  interactions can manipulate dynamic behavior of reactions facing post-transition-state bifurcations. *J. Am. Chem. Soc.* **139**, 7485-7493 (2017).
14. Gutta, P. & Tantillo, D. J. Proton sandwiches: nonclassical carbocations with tetracoordinate protons. *Angew. Chem. Int. Ed.* **44**, 2719-2723 (2005).
15. Mita, T. et al. Prediction of high-yielding single-step or cascade pericyclic reactions for the synthesis of complex synthetic targets. *J. Am. Chem. Soc.* **144**, 22985-23000 (2022).
16. Hayashi, H. et al. In silico reaction screening with difluorocarbene for N-difluoroalkylative dearomatization of pyridines. *Nat. Synth.* **1**, 804-814 (2022).
17. Kobylanskii, I. J., Novikov, M. S. & Khlebnikov, A. F. Formation and reactivity of gem-difluoro-substituted pyridinium ylides: Experimental and DFT investigation. *J. Fluor. Chem.* **132**, 175-180 (2011).
18. Dong, S., Fu, X. & Xu, X. [3+ 2]-cycloaddition of catalytically generated pyridinium ylide: A general access to indolizine derivatives. *Asian J. Org. Chem.* **9**, 1133-1143 (2020).
19. Kayala, M. A. & Baldi, P. ReactionPredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning. *J. Chem. Inf. Model.* **52**, 2526-2540 (2012).
20. Wang, L. P. et al. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044-1048 (2014).
21. Blackmond, D. G. The origin of biological homochirality. *Cold Spring Harb. Perspect. Biol.* **11**, a032540 (2019).
22. Joung, J. F. et al. Reproducing reaction mechanisms with machine learning models trained on a large-scale mechanistic dataset. *Angew. Chem. Int. Ed.* e202411296 (2024).
23. Tavakoli, M., Chiu, Y. T. T., Baldi, P., Carlton, A. M. & Van Vranken, D. RMechDB: A public database of elementary radical reaction steps. *J. Chem. Inf. Model.* **63**, 1114-1123 (2023).
24. Wołos, A. et al. Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. *Science*, **369**, eaaw1955 (2020).
25. Baisch, U., Pagano, S., Zeuner, M. & Schnick, W. Carbon dioxide fixation by organolanthanides and thermal degradation into amorphous and higher condensed Ln/O/C/N Solids. *Eur. J. Inorg. Chem.* **2006**, 3517-3524 (2006).
26. Bierbaum, V. M., Grabowski, J. J., & DePuy, C. H. Gas-phase synthesis and reactions of nitrogen- and sulfur-containing anions. *J. Phys. Chem.* **88**, 1389-1393 (1984).
27. Park, H., Jung, Y. M., You, J. K., Hong, W. H., & Kim, J. N. Analysis of the CO<sub>2</sub> and NH<sub>3</sub> reaction in an aqueous solution by 2D IR COSY: Formation of bicarbonate and carbamate. *J. Phys. Chem. A* **112**, 6558-6562 (2008).

28. Kojima, Y. Method and apparatus for synthesizing urea. U.S. Patent 2006/0270872 A1 (2006).
29. Saravanakumar, D., Song, J., Lee, S., Hur, N. H., & Shin, W. Electrocatalytic conversion of carbon dioxide and nitrate ions to urea by a titania-nafion composite electrode. *ChemSusChem* **10**, 3999–4003 (2017).
30. Perret, A. & Krawczynski, A. M. Recherches sur les cyanamides métalliques. *Helv. Chim. Acta* **15**, 1009–1022 (1932).
31. Kimura, M., Ishiguro, H. & Tsukahara, K. Kinetics of reduction of iodine by oxalate and formate ions in an aqueous solution: Difference in mechanisms of the reactions of oxalate and formate ions with iodine. *J. Phys. Chem.* **94**, 4106–4110 (1990).
32. Iyyamperumal, R., Zhang, L., Henkelman, G. & Crooks, R. M. Efficient electrocatalytic oxidation of formic acid using Au@Pt dendrimer-encapsulated nanoparticles. *J. Am. Chem. Soc.* **135**, 5521–5524 (2013).
33. Smithies, D. & Hart, E. J. Radiation chemistry of aqueous formic acid solutions. Effect of concentration. *J. Am. Chem. Soc.* **82**, 4775–4779 (1960).
34. Huber, C., Eisenreich, W. & Wächtershäuser, G. Synthesis of  $\alpha$ -amino and  $\alpha$ -hydroxy acids under volcanic conditions: Implications for the origin of life. *Tetrahedron Lett.* **51**, 1069–1071 (2010).
35. Yu, C., Liu, B. & Hu, L. Efficient Baylis–Hillman reaction using stoichiometric base catalyst and an aqueous medium. *J. Org. Chem.* **66**, 5413–5418 (2001).
36. SMARTS tutorial. [http://www.daylight.com/dayhtml\\_tutorials/languages/smarts/](http://www.daylight.com/dayhtml_tutorials/languages/smarts/) (accessed 13 August 2024).
37. Zidan, A. *et al.*  $\beta$ -Lactam synthesis through diodomethane addition to amide dianions. *Angew. Chem. Int. Ed.* **56**, 12179–12183 (2017).
38. Medeiros, G. A. *et al.* Probing the mechanism of the Ugi four-component reaction with charge-tagged reagents by ESI-MS(/MS). *Chem. Commun.* **50**, 338–340 (2014).
39. Barreto, A. de F. S., Dos Santos, V. A. & Andrade, C. K. Z. Synthesis of acylhydrazino-peptomers, a new class of peptidomimetics, by consecutive Ugi and hydrazino-Ugi reactions. *Beilstein J. Org. Chem.* **12**, 2865–2872 (2016)
40. Marcaccini, S., Pepino, R., Polo, C. & Cruz Pozo, M. Studies on Isocyanides and Related Compounds; A Facile Synthesis of 4-Phenyl-1-(2H)phthalazinone-2-alkanoic Acid Amides. *Synthesis* **2001**, 0085–0088 (2001)


41. Christian, G., Stolzenberg, H. & Fehlhammer, W. P. Metal stabilized cyanoisocyanide; X-ray structure of  $[\text{Cr}(\text{CO})_5\text{CNCN}]$ . *J. Chem. Soc. Chem. Commun.* 184–185 (1982)
42. Burger, H. *et al.* Syntheses and  $^1\text{H}$ -,  $^{13}\text{C}$ - and  $^{15}\text{N}$ -NMR spectra of ethynyl isocyanide, H-C triple bond C-N triple bond C, D-C triple bond C-N triple bond C and prop-1-ynyl isocyanide, H $^3$ C-C triple bond C-N triple bond C, D $^3$ C-C triple bond C-N triple bond C: high resolution infrared spectrum of prop-1-ynyl isocyanide. *Chem. Eur. J.* **6**, 3377–3385 (2000)
43. Landrum, G. RDKit. <https://www.rdkit.org/> (accessed 13 August 2024).
44. Vlaar, T., Ruijter, E., Maes, B. U. W. & Orru, R. V. A. Palladium-catalyzed migratory insertion of isocyanides: an emerging platform in cross-coupling chemistry. *Angew. Chem. Int. Ed.* **52**, 7084–7097 (2013)
45. Suryawanshi, M. D. *et al.* Design, Synthesis of 3-(5-Substituted Phenyl-[1,3,4]oxadiazol-2-yl)-1H-indole and Its Microbial Activity. *Chem. Biodivers.* **20**, e202201017 (2023)
46. Robinson, B. The Fischer Indole Synthesis. *Chem. Rev.* **63**, 373–401 (1963)
47. Chaskar, A., Deokar, H., Padalkar, V., Phatangare, K. & Patil, S. K. Highly Efficient and Facile Green Approach for One-Pot Fischer Indole Synthesis. *J. Korean Chem. Soc.* **54**, 411–413 (2010)
48. Jeanty, M., Blu, J., Suzenet, F. & Guillaumet, G. Synthesis of 4- and 6-azaindoles via the Fischer reaction. *Org. Lett.* **11**, 5142–5145 (2009)
49. Thomae, D., Jeanty, M., Coste, J., Guillaumet, G. & Suzenet, F. Extending the scope of the Aza-Fischer synthesis of 4- and 6-azaindoles. *European J. Org. Chem.* **2013**, 3328–3336 (2013)
50. Corey, E. J. and Rucker, C. Useful synthetic reagents derived from 1-triisopropylsilylpropyne and 1,3-[triisopropylsilyl]propyne, direct, stereoselective synthesis of either or enynes *Tetrahedron Lett.* **23**, 719 (1982).
51. Yanagisawa, A., Okitsu, S. & Arai, T. Regioselective Propargylation of Carbonyl Compounds with (3-Bromoprop-1-ynyl)trimethylsilane Promoted by Reactive Barium. *Synlett* **11**, 1679-1682 (2005).
52. Schmidt, B. Ruthenium-Catalyzed Olefin Metathesis Double-Bond Isomerization Sequence. *J. Org. Chem.* **69**, 7672-7687 (2004).

53. Kobayashi, S. & Nagayama, S. A Polymer-Supported Scandium Catalyst. *J. Org. Chem.* **61**, 2256-2257 (1996).
54. Li, Y., Brand, J. P. & Waser, J. Gold-Catalyzed Regioselective Synthesis of 2- and 3-Alkynyl Furans. *Angew. Chem. Int. Ed.* **52**, 6743-6747 (2013).
55. Jones, B. T. *et al.* Complex Polyheterocycles and the Stereochemical Reassignment of Pileamartine A via Aza-Heck Triggered Aryl C–H Functionalization Cascades *J. Am. Chem. Soc.* **143**, 15593–15598 (2021).
56. Kumar, S. V. *et al.* Cyclocondensation of arylhydrazines with 1, 3-bis (het) arylmonothio-1, 3-diketones and 1, 3-bis (het) aryl-3-(methylthio)-2-propenones: Synthesis of 1-aryl-3, 5-bis (het) arylpyrazoles with complementary regioselectivity. *J. Org. Chem.* **78**, 4960-4973 (2013).
57. Anouti, M., Caillon-Caravanier, M., Dridi, Y., Galiano, H. & Lemordant, D. Synthesis and characterization of new pyrrolidinium based protic ionic liquids. Good and superionic liquids. *J. Phys. Chem. B.* **112**, 13335–13343 (2008).
58. Schultz, E. E., Lindsay, V. N. G. & Sarpong, R. Expedient Synthesis of Fused Azepine Derivatives using a Sequential Rhodium(II)-Catalyzed Cyclopropanation/1-Aza-Cope Rearrangement of Dienyltriazoles. *Angew. Chem.* **126**, 10062 –10066 (2014).
59. Reid, C. M. *et al.* Synthesis and anti-protozoal activity of C2-substituted polyazamacrocycles. *Bioorg. Med. Chem. Lett.* **18**, 2455–2458 (2008).
60. Zhang, Z.; Zhu, L. & Li, C. Copper-Catalyzed Carbotrifluoromethylation of Unactivated Alkenes Driven by Trifluoromethylation of Alkyl Radicals *Chin. J. Chem.* **2019**, *37*, 452—456.
61. Tago, K. & Kogen, H. Bis(2,2,2-trifluoroethyl)bromophosphonoacetate, a Novel HWE Reagent for the Preparation of (*E*)-alpha-Bromoacrylates: A General and Stereoselective Method for the Synthesis of Trisubstituted Alkenes. *Org. Lett.* **2**, 1975–1978 (2000).
62. Molina, P., Pastor, A. & Vilaplana, M. J. Vinyliminophosphorane-mediated preparation of 2-arylquinoline and 4-aryl-1-azaanthraquinone derivatives. X-Ray crystal structure of 1,2-dihydro-3H-indazolo[2,3-a]quinolin-4-one. *Tetrahedron*, **51**, 1265–1276 (1995).
63. Kim, K.-M. & Park, I.-H. A Convenient Halogenation of  $\alpha,\beta$ -Unsaturated Carbonyl Compounds with OXONE® and Hydrohalic Acid (HBr, HCl) *Synthesis* **16**, 2641-2644 (2004).

64. Murphy, J. A. *et al.* A radical based addition–elimination route for the preparation of indoles. *J. Chem. Soc., Perkin Trans. 1*, 2395–2408 (2000).
65. Suresh, R., Muthusubramanian, S., Nagaraj, M. & Manickam, G. Indium trichloride catalyzed regioselective synthesis of substituted pyrroles in water. *Tetrahedron Lett.* **54**, 1779–1784 (2013).
66. Huang, Z., Wang, L. & Huang, X. Stereoselective Synthesis of  $\alpha$ -Bromo- $\alpha,\beta$ -unsaturated Ketones via Wittig Reaction. *Synth. Commun.* **33**, 757–762 (2003).
67. Chao, A., Lujan-Montelongo, J. A. & Fleming, F. F. Isocyano Enones: Addition–Cyclization Cascade to Oxazoles. *Org. Lett.*, **18**, 3062–3065 (2016).
68. Moser, R., Boskovic, Z. V., Crowe, C. S. & Lipshutz, B. H. CuH-Catalyzed Enantioselective 1,2-Reductions of  $\alpha,\beta$ -Unsaturated Ketones. *J. Am. Chem. Soc.* **132**, 7852–7853 (2010).



Cite this: DOI: 10.1039/d5sc00070j

 All publication charges for this article have been paid for by the Royal Society of Chemistry

# Retro-forward synthesis design and experimental validation of potent structural analogs of known drugs†

Ahmad Makkawi,<sup>‡a</sup> Wiktor Beker,<sup>‡b</sup> Agnieszka Wotos,<sup>‡b</sup> Sabyasachi Manna,<sup>a</sup> Rafał Roszak,<sup>b</sup> Sara Szymkuć,<sup>b</sup> Martyna Moskal,<sup>b</sup> Aleksei Koshevarnikov,<sup>ab</sup> Karol Molga,<sup>b</sup> Anna Żądło-Dobrowolska<sup>\*a</sup> and Bartosz A. Grzybowski<sup>ID\*acd</sup>

Generation of structural analogs to “parent” molecule(s) of interest remains one of the important elements of drug development. Ideally, such analogs should be synthesizable by concise and robust synthetic routes. The current work illustrates how this process can be facilitated by a computational pipeline spanning (i) diversification of the parent *via* substructure replacements aimed at enhancing biological activity, (ii) retrosynthesis of the thus generated “replicas” to identify substrates, (iii) forward syntheses originating from these substrates (and synthetically versatile “auxiliaries”) and guided “towards” the parent, and (iv) evaluation of the candidates for target binding and other medicinal–chemical properties. This pipeline proposes syntheses of thousands of readily makeable analogs in a matter of minutes, and is deployed here to validate by experiment seven structural analogs of Ketoprofen and six analogs of Donepezil. The concise, computer-designed syntheses are confirmed in 12 out of 13 cases, offering access to several potent inhibitors. While the synthesis-design component is robust, binding affinities are predicted less accurately although still to the order-of-magnitude, which may be valuable in discerning promising from inadequate binders.

Received 5th January 2025  
Accepted 16th March 2025

DOI: 10.1039/d5sc00070j

rsc.li/chemical-science

## 1. Introduction

Recent years have brought revolutionary advances<sup>1–10</sup> in the use of computers to autonomously plan chemical syntheses of arbitrary targets, all the way up to complex natural products.<sup>8–10</sup> One of the prominent areas of application of these algorithms has been in drug discovery where the synthesis design is part of algorithmic pipelines<sup>11–15</sup> intended to predict target and off target binding as well as ADME-Tox properties. While the premium is, without doubt, on discovering potent candidates featuring unprecedented scaffolds<sup>16</sup> and binding modes, many drugs are derivatives of the old ones and “the best way to discover a new drug is to start with an old one”.<sup>17</sup> Accordingly, analogs structurally similar to some desired “parent” molecules continue to be sought in drug screening and lead-optimization campaigns<sup>18</sup> and efforts to accelerate this process with the help

of computers date back to at least the 1990s with many ingenious approaches undertaken since then.

There are two general schools of thought about the analog design problem. In the “classic” approach, a computer program performs *in silico* reactions to produce virtual molecules. The reactions are typically applied in the “forward” direction (starting from some static collections of starting materials<sup>19–25</sup>) and generate molecular spaces which, nowadays, can be quite enormous<sup>25</sup> (*e.g.*, a few billion virtual molecules in Enamine’s REAL space and up to 10 (ref. 20) in Merck KGaA’s MASSIV collection). These spaces are subsequently pruned for target similarity or other properties of interest. There are also solutions that use retrosynthetic pathways of the parent as input and generate analogs by replacing the original starting materials by structurally or functionally similar blocks.<sup>26</sup> The second family of approaches relies on the burgeoning generative AI models.<sup>27–34</sup> These methods often combine target-similarity with concurrent predictions of other properties.<sup>35</sup> While the very synthesizability of the generated structures has historically been a challenge,<sup>36–39</sup> there has been significant recent progress, as detailed in an excellent recent ref. 40. We observe, however, that irrespective of the approach taken, studies in this area are rarely accompanied by experimental validations of computer-designed syntheses (see ref. 23, 24 and 41) and/or of the predicted potency of the proposed analogs. Such validations are urgently needed as, ultimately, they will decide wider adoption

<sup>a</sup>Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland. E-mail: anna.zadlo@icho.edu.pl<sup>b</sup>Allchemy, Inc., Highland, IN, USA<sup>c</sup>Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan, 44919, Republic of Korea. E-mail: nanogryzbowski@gmail.com<sup>d</sup>Department of Chemistry, Ulsan Institute of Science and Technology, UNIST, Ulsan, 44919, Republic of Korea† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5sc00070j>

‡ Authors contributed equally.



of these methods (which, one may argue, is not guaranteed given the widely publicized, recent setbacks of AI in industrial drug discovery<sup>42</sup>).

With this in mind, the overriding objective of our current work is not to argue for the advantages or disadvantages of any particular approach (they all have some) but to test by experiment the particular analog-design pipeline we have been developing for several years now. This pipeline is along the abovementioned “classical” lines of *in silico* synthesis and encompasses substructure replacements within the parent (to diversify the parent scaffold and, hopefully, enhance biological activity), retrosynthetic generation of substrates retaining mutual reactivity, “guided” forward synthesis to produce large numbers of easily synthesizable structural analogs of the parent, and estimation of these candidates’ binding affinity to the desired target. Within this scheme, the two key questions we ask relate to the correctness of the computer-designed analog syntheses and to the accuracy of predicting their binding affinities.

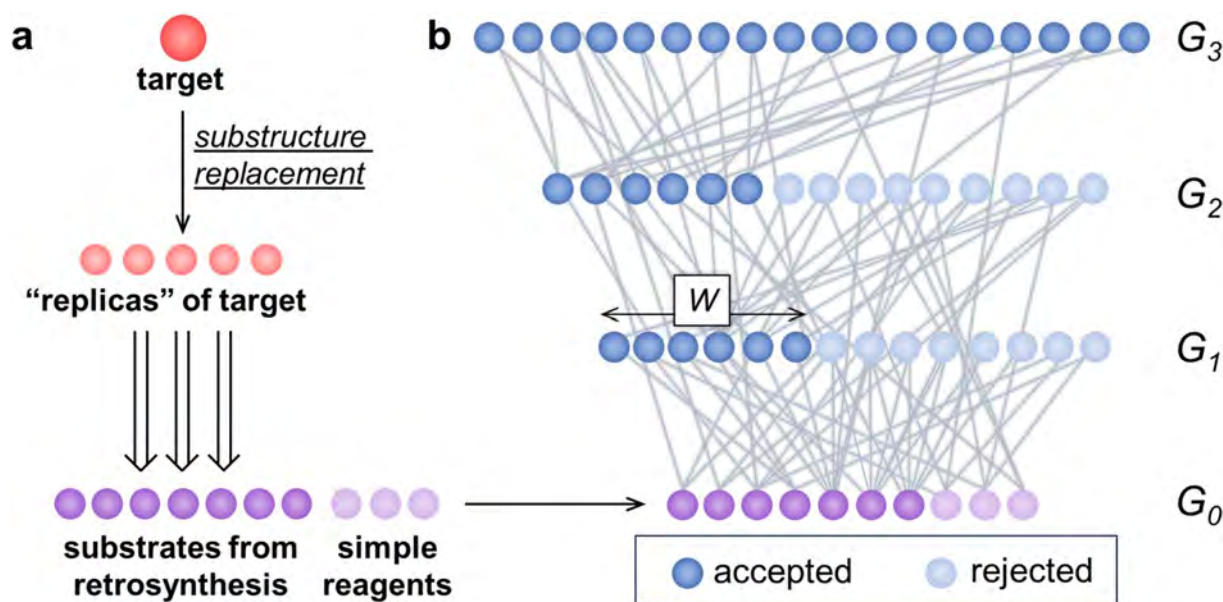
Considering two parent molecules (Ketoprofen and Donepezil), the outcomes of our studies are nuanced. On the positive side, experiments validate concise, computer-designed syntheses of seven analogs of Ketoprofen and five analogs of Donepezil (against one failed route). Six Ketoprofen analogs are  $\mu\text{M}$  binders to human cyclooxygenase-2 (COX-2), with one

offering slightly better binding than the parent drug ( $0.61 \mu\text{M}$  vs.  $0.69 \mu\text{M}$ ). For Donepezil, all five analogs show sub-micromolar binding to acetylcholinesterase, AChE, with one having nanomolar affinity close to that of the parent ( $36 \text{ nM}$  vs.  $21 \text{ nM}$ ). At the same time, binding predictions – which had guided selection of analogs for synthesis validation – by three different docking programs and a neural-network, match the experimental values only to within an order-of-magnitude. These results make us conclude that (i) the synthesis-planning aspects of computerized analog design are nowadays robust, and (ii) common affinity-prediction tools may help select promising binders but cannot discriminate between moderate ( $\mu\text{M}$ ) vs. high-affinity (nM) ones.

## 2. Results

### 2.1 Components of the computational pipeline (Fig. 1)

**2.1.1 Guided reaction networks.** Our analog-design algorithm rests on the application of the so-called guided reaction networks, described in several of our prior works.<sup>43–45</sup> Briefly, assuming a given collection of starting materials (“zero-th synthetic generation,”  $G_0$ ) – whose choice we will discuss in point 1.2 – the algorithm iteratively applies its knowledge-base of reaction transforms,  $\{R_i\}$ , in the “forward” direction (these reaction transforms are encoded as described in detail in ref. 46



**Fig. 1** Scheme of the algorithmic pipeline for substrate selection and subsequent analog generation. (a) For a given target/“parent” of interest (node colored in red), the algorithm first identifies its substructures that can be altered by replacements likely to result in activity enhancement.<sup>49</sup> This creates several (for a typical drug-like parent,  $\sim 10$ – $100$ ) parent replicas (light red). The algorithm then expands retrosynthetic networks to find viable synthetic routes to all of these molecules and retains their commercially available starting materials. Since easily makeable analogs are ultimately sought, retrosynthesis is limited to the depth of five steps and using only 180 reaction classes popular in medicinal chemistry. Retrosynthetic searches stop when reaching commercially available substrates, here, those from Mcule’s catalog, <https://mcule.com/database/>, of  $\sim 2.5$  million chemicals. The set of retrosynthetically-derived substrates (nodes colored violet) is further augmented by 23 simple yet synthetically useful chemicals (here, additional nodes colored in light violet). (b) All of the said substrates serve as the zero-th generation,  $G_0$ , for the guided forward search in which, after each round of reactions, only some  $W$  molecules most similar to the parent are retained (dark-blue nodes, here  $W = 150$ ). In this way, the forward synthesis is gradually “focused” towards parent’s structural analogs. We note that because of this truncation, the sizes of the networks do not explode. This also allows the use of larger collections of reaction transforms (here,  $\sim 25\,000$  rules from Allchemy) to achieve more synthetically diverse outcomes, while limiting the times of network propagation, typically to several minutes.





and 47). Application of  $\{R_i\}$  to  $G_0$  gives products in generation  $G_1$ . Then, molecules in  $G_0$  and  $G_1$  are combined and become available for yet another round of reactions. If this process is simply repeated without any additional restrictions to create  $G_2$ ,  $G_3$ , etc., the numbers of molecules produced in each generation and the overall size of the network increase very rapidly (as we showed in ref. 43, stronger than exponentially), and the calculation times become impractically long. Instead, to allow for efficient (minutes) exploration of the structural space yet to “guide” network expansion towards the desired “parent”, the numbers of products retained after each generation is restricted to some predetermined number  $W$  (a.k.a. “beam width”<sup>48</sup>) of those most similar to the said parent molecule. For instance, starting from 100 substrates in  $G_0$ , one typically produces thousands of molecules already in  $G_1$  – however, only a few hundred most parent-similar ones are retained as  $G_1$ , and allowed to further react with each other and with the  $G_0$  substrates to generate  $G_2$ . Regarding this scheme, we note that after a few initial generations, the produced molecules may become already as large as the parent itself. At this point, it makes little sense to allow them to further react with each other. Accordingly, after 1–2 initial generations, we impose an additional constraint that the molecules retained in a given  $G_i$  can react only with species of “earlier” generations (*i.e.*, up to  $G_{i-1}$ ) but no longer between themselves.

**2.1.2 Retrosynthesis and the choice of substrates.** The second pillar of our approach is the very choice of the  $G_0$  commercially available starting materials, which we wish to select judiciously (to offer the best chance of generating parent’s structural analogs) yet automatically (to avoid subjective and tedious selection by software’s human operator). This substrate set cannot be too large (as this might explode the forward networks) but should also be (i) diverse, capturing not only the key structural motifs of the parent (*e.g.*, rings or ring systems) but also motifs similar to them; and (ii) should be synthetically flexible, in the sense that the blocks should be able to engage in mutual reactions as much as possible and also in other reactions that functionalize these blocks.

Given these requirements, the straightforward method of disconnecting the parent into the starting materials by retrosynthesis may be overly simplistic. Retrosynthesis of only the parent molecule generates substrates that, in the forward direction, can be reassembled into the parent itself, into some intermediates *en route* to this parent, and usually some molecules in which alternative reactions of the starting materials are performed. However, this approach does not generate diverse analogs. The set of starting materials can, of course, be diversified by adding molecules similar to those found by retrosynthesis – we tested this approach early on but many of the “similar” to the starting materials featured functional group patterns that were unsuitable or problematic for the subsequent forward synthesis (see ESI, Section S2†). Mindful of this, we augmented the retrosynthetic protocol in two ways:

First, by performing substructure replacements (intended to enhance biological activity and digitized, in large part, according to Novartis’ tables from ref. 49) within the parent and only then performing retrosyntheses of these “replica” molecules.

The replacements are performed at either peripheral groups or internal motifs such as 1,3-disubstituted benzene rings or piperazines (see ESI, Fig. S2†). They increase structural diversity of the starting materials while, simultaneously, retaining functional groups necessary for the “mix-and-match” reactivity between blocks derived from different replicas. Here, in all retrosyntheses, we used the Allchemy algorithm<sup>50</sup> although other reliable retrosynthetic engines<sup>2–4,51</sup> can be used.

Second, by adding to  $G_0$  some simple yet synthetically useful chemicals with which to further modify and/or activate molecules within the forward networks. While various choices are possible, we used a static, “minimal” set of 23 popular chemicals (see ESI, Fig. S4†) chosen for synthetic versatility. For example, *N*-chlorosuccinimide, *N*-bromosuccinimide and nitric acid enable electrophilic aromatic substitutions, bis(pinacolato)diboron opens the way to Suzuki couplings, DAST can functionalize molecules with fluoride, mesyl chloride activates alcohols for  $S_N2$  reaction, ethyl magnesium bromide can engage in Kulinkovich cyclopropanation while hydrazine, thiourea, azide and trimethyl orthoformate enable formation of various heterocycles, etc. An example of a search with and without these additional chemicals is provided in the ESI, Section Section S3.†

**2.1.3 Property evaluation.** After the retro-forward searches produce the structural analogs, these candidates can be inspected by various substructure filters and ranked by property estimation algorithms, some of which will be discussed in specific examples below.

Regarding the synthesis components 1.1 and 1.2, we note that retrosynthesis may be significantly slower than guided network expansion. With Allchemy’s full reaction database, retrosynthetic analysis of a typical drug-like molecule takes only 1–5 min, but for the tens of replicas created in the first stage of the pipeline, these times are already in hours – that is, much longer than for the forward synthesis up to 3–4 generations (4–6 min). Consequently, these two elements of the pipeline use different sets of reaction rules: for retrosynthesis, to reduce the size of retrosynthetic search networks, only 180 reaction classes most popular in medicinal chemistry, and for the forward synthesis, all 25 307 reaction rules available in Allchemy. In this way, both retrosynthetic and forward searches complete within minutes which, based on the feedback on the pipeline’s external users appears to be a practically acceptable limit for this type of application. The entire pipeline, integrated in a form of a user-friendly WebApp is available for academic testing at <https://analog.allchemy.net/> with the user manual provided in the ESI, Section S1.† The source-code for guided network expansion is deposited at <https://zenodo.org/records/7371247>.

## 2.3 Application of the pipeline to specific targets

We applied the above approach (1)–(3) to identify readily synthesizable analogs to anti-inflammatory Ketoprofen and to Donepezil, a medication used to treat Alzheimer-type dementia.

**2.3.1 Ketoprofen analogs.** In this example, the algorithm commenced by generating 61 Ketoprofen’s substructure-replacement replicas (Fig. 2a). The retrosynthetic searches to all these targets were allowed to terminate at relatively large,



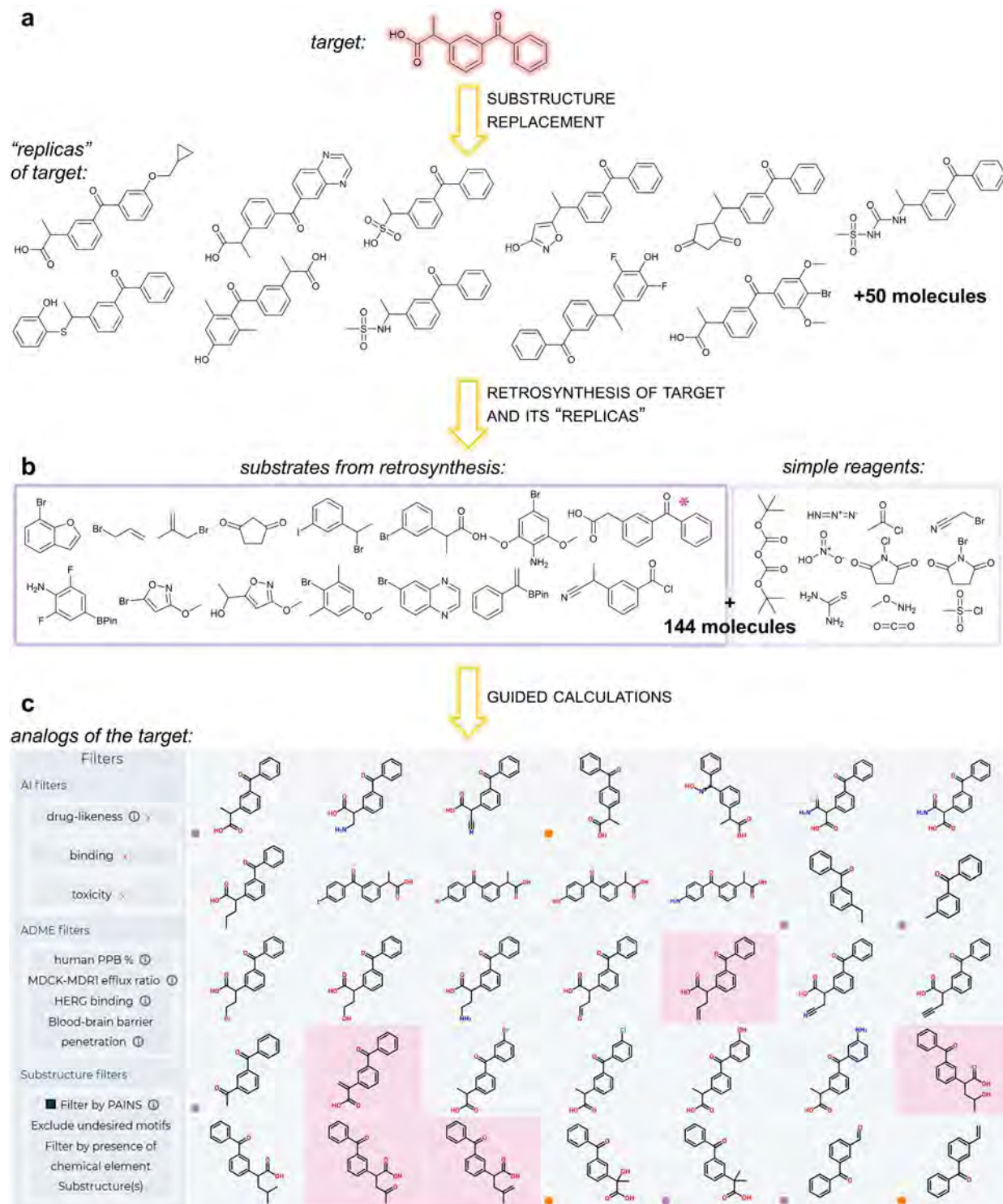
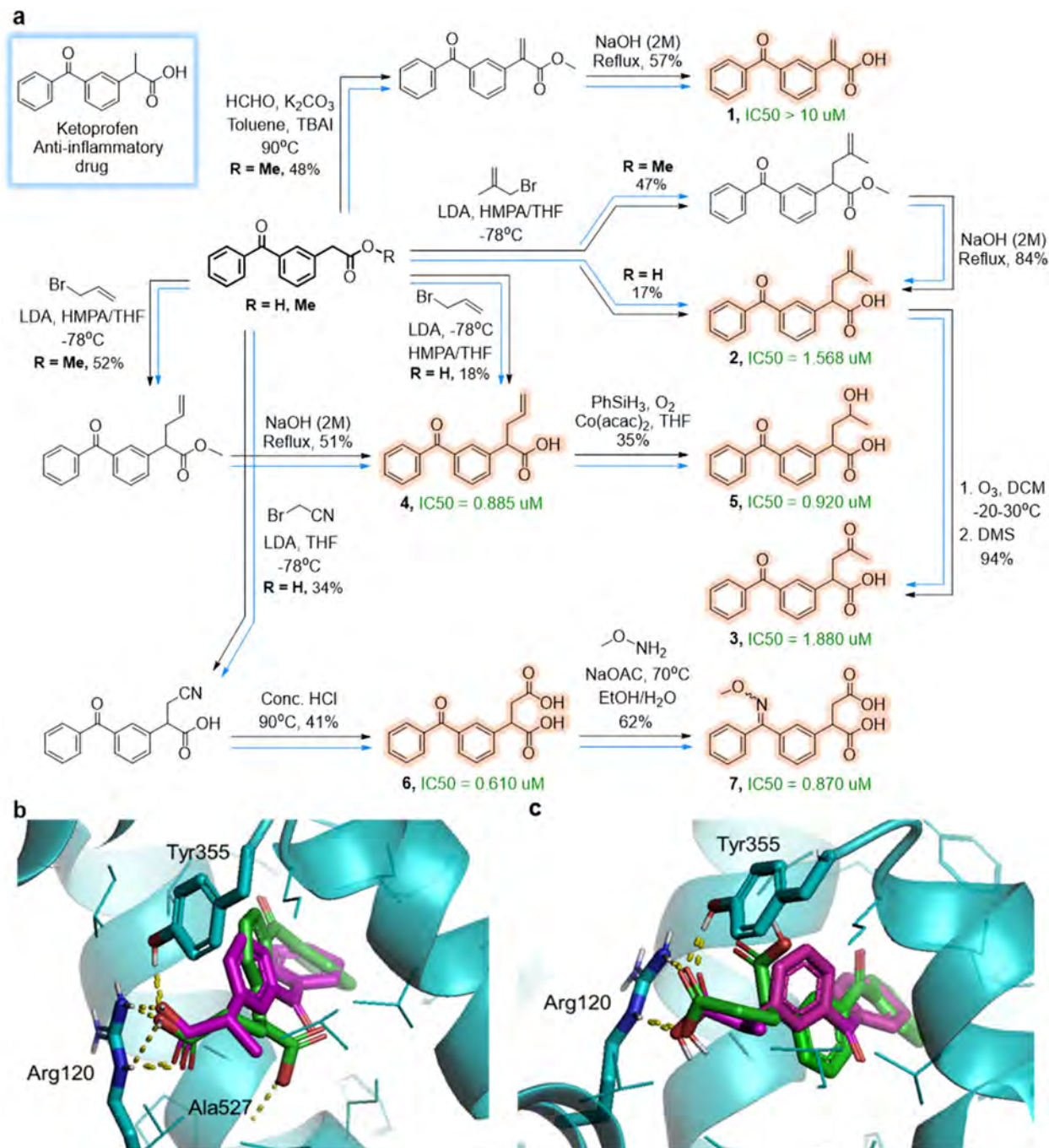


Fig. 2 Search for Ketoprofen's analogs. The panels illustrate the key steps along the analog design pipeline: (a) Ketoprofen and some of its replicas featuring substructure replacements aimed at enhancing biological activity;<sup>49</sup> (b) some of the substrates derived by retrosynthesis of molecules from (a) as well as, on the right, examples of the simple reagents with which to functionalize or activate the substrates (for all 23, see ESI, Fig. S4†); (c) Allchemy screenshot of some of the top-ranking analogs generated by guided expansion of a reaction network starting from molecules in (b). The analogs shown here are sorted by similarity to the Donepezil parent and are centered around the 2-(3-benzoylphenyl)acetic acid with modifications introduced mostly in the  $\alpha$ -position of carboxylic acid and as substituents on the aromatic ring. Clicking on any of the "tiles" provides synthesis details (these and other options to visualize the synthetic networks are detailed in the ESI, Section S1†). Some of the analogs ultimately committed to synthesis are shown on pink backgrounds. Visible on the left is the panel with medicinal-chemical functionalities by which the analogs can be filtered (e.g., general drug-likeness,<sup>79</sup> binding to specific proteins, various ADME-Tox models, PAINS and substructure-based filters, see user manual in the ESI†).





**Fig. 3** Ketoprofen's analogs committed to synthesis. (a) A small network of algorithm-planned syntheses of analogs **1** through **7** starting from 2-(3-benzoylphenyl)acetic acid and/or ester derivatives (R = H or R = Me). Experimental pathways (tracked by black arrows) follow the algorithm's original suggestions (blue arrows). Conditions and isolated yields are given next to reaction arrows. IC<sub>50</sub> values are given next to the analogs. (b) and (c) All analogs were docked into COX-2 protein. The alignment of docking poses from AutoDock 4 for (b) (S)-Ketoprofen and its analog (S)-6, and (c) (R)-Ketoprofen and its analog (R)-6. Key, protein-ligand hydrogen bonds are depicted as yellow dotted lines. The carboxylic acid moiety of (S)-Ketoprofen (colored in magenta) forms two hydrogen bonds with Arg120 and one with Tyr355. Its analog (S)-6 (b, colored in green) is predicted to be similarly positioned inside the active site and to form three hydrogen bonds: with Arg120, with Tyr355 and, for the second carboxylic acid moiety, with Ala527. For (R)-Ketoprofen (colored in magenta in panel (c)) and its analog (R)-6 (colored in green), the carboxylic acid moieties in both compounds form two hydrogen bonds (with Arg120 and with Tyr355). Raw files with these and all other docking experiments using AutoDock 4, AutoDock Vina, and Dock 6 (ref. 54–57) are deposited at <https://zenodo.org/records/14571461>.

commercially available starting materials such as (3-benzoylphenyl)acetic acid marked by pink asterisk in Fig. 2b; other notable substrates included allyl bromide derived by retrosynthesis of replicas containing cyclopropylmethoxy substituent.

The 62 retrosynthetic searches took 4 min 20 s (on a server with Intel Xeon Gold 5412U processor) and identified 151 starting materials. Then, the forward search from these substrates and additional set of 23 useful chemicals (four molecules



overlapped between these two sets) was propagated to  $G_3$  and produced, within 4 min 3 s, a guided reaction network encompassing 3692 products of which 781 (21%) had similarity to parent above 0.7 (ECFP6 (ref. 52) – based on Tversky's similarity with parameters (0.2, 0.8); see examples in Fig. 2c and results deposited under the “saved results” tab of the WebApp at <https://analogs.allchemy.net/>).

Within this set, we focused on molecules 1–7 that form a small synthetic cluster summarized in Fig. 3a. These analogs are all derivatives of the commercially available 2-(3-benzoylphenyl)acetic acid and were not reported or tested for cyclooxygenase-2, COX-2, binding before (only compound 1 was synthesized in ref. 53 but in the context of hydrocarboxylation methodology with no biological studies). Moreover, the proposed syntheses all appear concise and straightforward, and were successfully carried out in yields indicated next to the arrows (under conditions identical or very similar to those suggested by the software and without thorough optimization).

In selecting these particular structures, we were also guided by predictions of their binding affinities to COX-2. To this end, we used three popular docking programs, AutoDock 4, AutoDock Vina, Dock 6.<sup>54–57</sup> These programs used the 5IKR PDB structure and predicted the binding scores of the analogs 1–7 (averaged over possible stereoisomers of a given racemic analog) to be comparable or more favorable than those of Ketoprofen, which we also docked for reference (top portion of Table 1). Inspection of the docking poses (see examples in

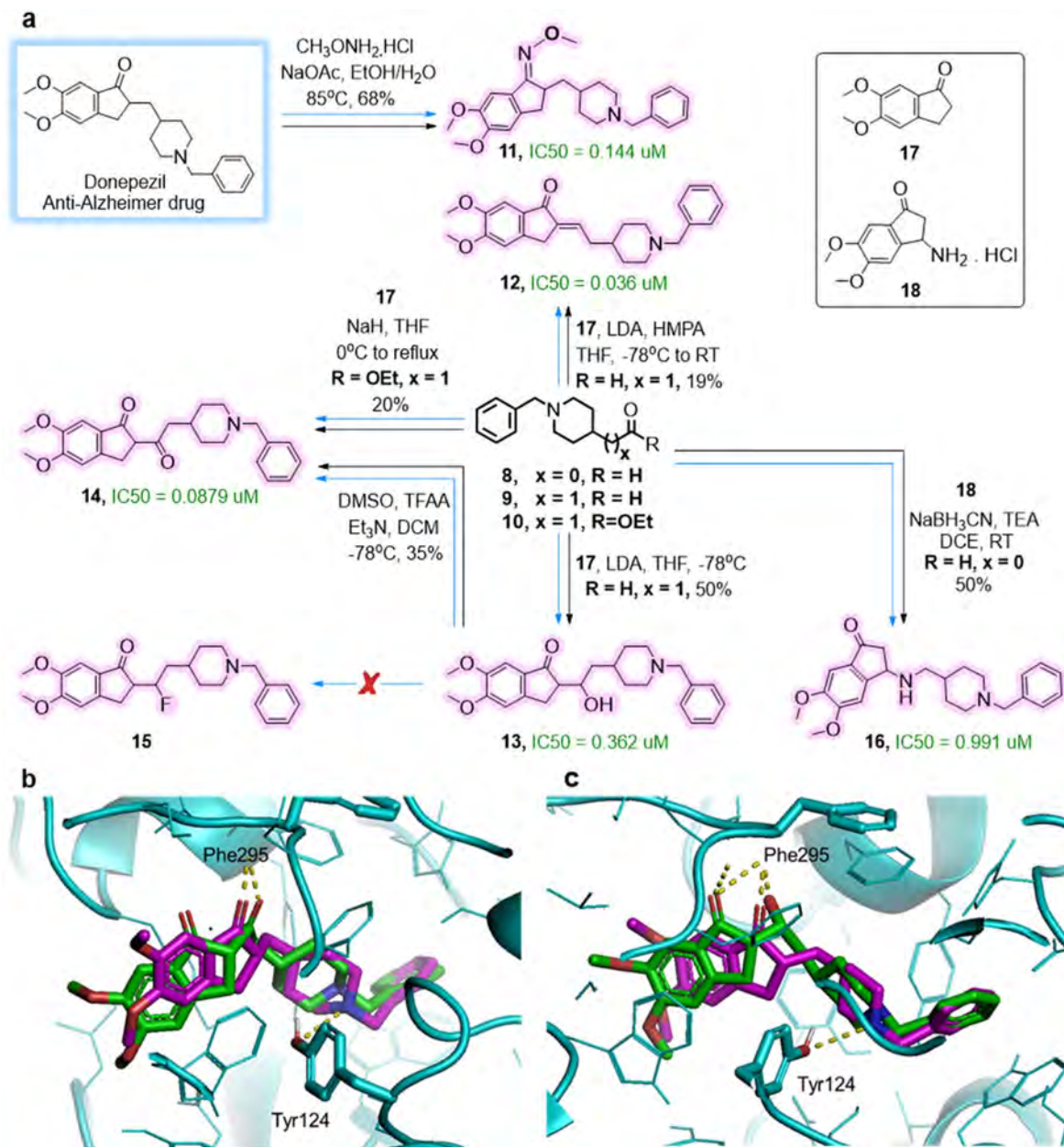
Fig. 3b and c and remaining analogs deposited at <https://zenodo.org/records/14571461>) revealed that these molecules should, indeed, be able to engage in numerous favorable interactions with the COX-2 active site. In parallel, we sought binding strength estimates using Allchemy's neural network, NN, trained on 1 752 921 protein assays from the ChEMBL 29 database<sup>58</sup> and spanning activity values for 3843 one-hot-encoded targets and 863 471 ligands represented as concatenation of Morgan fingerprints with radius 3 (using AllChem.-GetMorganFingerprint function in RDKit<sup>59</sup>), Xfp pharmacophore fingerprints<sup>60</sup> and selection of MACCS keys.<sup>61</sup> The architecture of this feedforward multitasking NN is similar to the one used in ref. 62 and 63, where it was shown to provide the best performance for ChEMBL activity prediction. The main difference is that we used batch normalization, and assigned different activity thresholds to different protein families (according to the values defined in ref. 64). These alterations slightly improved performance to ROC AUC = 0.87. Here, this network predicted that the affinities of the analogs we synthesized should be on-the-order-of micromolar.

To verify these predictions, we ran spectrofluorometric COX-2 human inhibition assay<sup>65,66</sup> and quantified the  $IC_{50}$  values. These values are marked in green next to the specific analogs in Fig. 3a. As seen, one analog, 1, is binding poorly, >10  $\mu$ M, but the remaining six are micromolar binders with one, 6, exhibiting slightly better affinity than the parent Ketoprofen, 0.61  $\mu$ M vs. 0.69  $\mu$ M.

**Table 1** Predicted binding affinities (or docking scores) and selected medicinal–chemical properties for the parent compounds and their analogs. Top part of the Table is for Ketoprofen and its analogs 1–7; bottom part is for Donepezil and its analogs 11, 14, 16. The first two columns specify, respectively,  $K_i$  values predicted by AutoDock Vina and by AutoDock 4. The third column has the docking score from Dock 6. The values shown are averages over all stereoisomers (files with the docked individual stereoisomers are deposited at <https://zenodo.org/records/14571461>). The remaining four columns give the Allchemy-predicted values of  $\log P$  calculated by group contribution method<sup>60</sup> as well as predictions of Allchemy's machine-learning models for hERG cardiotoxicity (on a 0–1 scale), degree of human plasma protein binding, hPPB, related to drug absorption and efficacy (0–100%), and degree of blood–brain barrier, BBB, penetration (0–1). Although none of the models used Ketoprofen or Donepezil molecules for training/testing, the key predictions match experimental data. For instance, the hERG model correctly predicts Donepezil to be cardiotoxic (values close to 1), which is in line with experimental studies evidencing that Donepezil binds to hERG<sup>61</sup> with  $IC_{50} = 1.3 \mu$ M and can cause QTc prolongation.<sup>62</sup> By contrast, the model predicts low values of hERG inhibition for Ketoprofen (and its analogs), which agrees with experiments confirming its low cardiotoxicity.<sup>63</sup> In turn, the hPPB model predicts high, >90% bound fractions for both Ketoprofen and Donepezil and their analogs – while lower hPPB values are, in principle, desired in drug design, these predicted values agree with experimental measurements for Ketoprofen (99%<sup>64</sup>) and Donepezil (95.6%<sup>65</sup>). In turn, the blood–brain barrier, BBB, penetration model predicts high values for Donepezil, which is expected for a CNS drug and experimentally confirmed.<sup>66</sup> Ketoprofen also crosses the BBB<sup>67</sup> but we note that some of the analogs, more polar ones, are predicted to have significantly lower values

Compound	$K_i$ AutoDock Vina ( $\mu$ M)	$K_i$ AutoDock ( $\mu$ M)	Dock 6 (score)	$\log P$	hERG (0–1)	hPPB (0–100%)	BBB (0–1)
Ketoprofen	0.613	0.284	–33.03	3.106	0.04	98.40 $\pm$ 1.20	0.91
1	0.413	0.202	–34.85	3.015	0.03	98.50 $\pm$ 1.00	0.84
2	0.814	0.268	–34.62	4.052	0.1	98.80 $\pm$ 1.30	0.64
3	0.579	0.161	–33.40	3.065	0.01	97.00 $\pm$ 1.30	0.42
4	1.138	0.224	–34.82	3.662	0.09	98.60 $\pm$ 1.20	0.69
5	0.538	0.348	–36.74	2.857	0.01	94.30 $\pm$ 1.20	0.46
6	0.451	0.154	–33.19	2.561	0	96.50 $\pm$ 2.70	0.37
7	0.534	0.264	–39.15	2.728	0	95.80 $\pm$ 3.60	0.29
Donepezil	0.046	0.037	–43.78	4.361	0.96	91.70 $\pm$ 1.30	0.92
11	0.086	0.061	–41.76	4.357	0.98	90.80 $\pm$ 1.10	0.9
12	0.064	0.069	–42.54	4.671	0.95	95.80 $\pm$ 1.20	0.75
13	0.204	0.097	–43.06	3.722	0.92	88.90 $\pm$ 1.10	0.66
14	0.037	0.053	–42.55	3.930	0.92	93.90 $\pm$ 4.70	0.77
16	0.027	0.011	–46.98	3.833	0.97	89.60 $\pm$ 1.20	0.75





**Fig. 4** Donepezil's analogs committed to synthesis. (a) A small network of algorithm-planned syntheses of analogs **11** through **16**. Blue arrows highlight Allchemy's suggestions, black arrows highlight the experimentally executed pathways. Conditions and isolated yields are given above reaction arrows.  $\text{IC}_{50}$  values are given next to the analogs. Note: As synthesis of **9** and **10** from simpler starting materials was previously described, we followed the literature procedure (**10** was made in two steps and aldehyde **9** from ester **10** in another 2 steps with Allchemy-proposed alcohol as an intermediate; ester **10** and alcohol are commercially available but relatively expensive). All analogs were docked into acetylcholinesterase from *Electrophorus electricus*. Docking poses from AutoDock Vina are aligned for (b) (*R*)-donepezil and its analog (*S*)-**14**, and (c) (*R*)-donepezil and its analog (*1R,2S*)-**13**. Key, protein-ligand hydrogen bond interactions are traced by yellow dotted lines. The carbonyl oxygen of (*R*)-Donepezil's (colored in magenta) indanone ring forms a hydrogen bond with Phe295 –NH whereas nitrogen from piperidine ring hydrogen-bonds with Tyr124 –OH. Its analog (*S*)-**14** (b, colored in green) forms hydrogen bonds with Phe295 –NH, but instead of carbonyl oxygen from indanone ring, carbonyl oxygen from 2-(1-benzylpiperidin-4-yl)acetyl side chain is involved. Analog (*1R,2S*)-**13** (c, colored in green) forms two hydrogen bonds between Phe295 –NH and carbonyl oxygen of indanone ring and hydroxyl oxygen of analog's side chain as well as and one hydrogen bond between carbonyl oxygen of indanone ring and Arg296 –NH<sub>2</sub>. Raw files with these and all other docking experiments using (AutoDock 4, AutoDock Vina, Dock 6)<sup>54–57</sup> are deposited at <https://zenodo.org/records/14571461>.

**2.3.2 Donepezil analogs.** We followed a similar protocol as for Ketoprofen. The algorithm generated 18 substructure-replacement replicas of the Donepezil parent. To reach commercially available substrates, the algorithm used

disconnections more central than in Ketoprofen example and reached, e.g., 5,6-dimethoxy-2,3-dihydro-1*H*-inden-1-one, **17**, as well as, 2-(1-benzylpiperidin-4-yl)ethanol and (1-benzylpiperidin-4-yl)acetic acid, that can be easily transformed



into aldehyde **9** – the core substrate, used in most syntheses – and ester **10**, respectively.

Retrosyntheses of the parent and all 18 replicas took only 3 min and generated 44 substrates which were combined with 23 synthetically useful, simple chemicals. Three molecules overlapped between these two sets so that the number of unique starting materials in  $G_0$  was 64. The forward search up to  $G_3$  generated, within 5 min, a network of 3619 products of which 2222 (~61%) had similarity to parent above 0.7.

Fig. 4a shows six molecules that were committed to synthesis: the replica **11** of the parent drug and analogs **12**–**15** forming a small cluster entailing only six reactions based on the aforementioned aldehyde **9**, ester **10** as well as ketone **17**. Additionally, we included the synthesis of analog **16** that had been found by an earlier version of the algorithm, and used **8** and **18** as starting materials. All these reactions were executed under computer-suggested conditions, save for the Swern rather than Dess–Martin oxidation of **13** to **14**. In five out of six cases, the reactions gave the expected products although in poor to moderate yields (no condition optimization was attempted). In one case, the fluorinated derivative **15** was not obtained because of elimination to **12**. While there are literature examples of fluorination proceeding on  $\beta$ -hydroxy ketones, Allchemy's  $pK_a$  model<sup>67</sup> predicts that the cH position within our cyclic aryl ketone is more acidic, hence promoting elimination vs. substitution.

As in the case of Ketoprofen, these analogs were chosen because most were not reported before (compound **11** is known as an intermediate in the synthesis of acetylcholinesterase, AChE, inhibitors but its activity was not evaluated<sup>68</sup>) and because their predicted binding scores (by AutoDock 4, AutoDock Vina, Dock 6)<sup>54–57</sup> were comparable or better than those of the Donepezil parent, which we also docked for reference (see bottom portion of Table 1). Allchemy's neural network also predicted micromolar-level binding.

These predictions were verified in spectrophotometric AChE inhibition assay,<sup>65,69</sup> which quantified the  $IC_{50}$  values given next to the specific analogs in Fig. 4a. The replica **11** had  $IC_{50} = 144$  nM and two analogs were submicromolar, 991 nM for **16** and 362 nM for **13**. However, the two remaining analogs were significantly more potent, 88 nM for **14** and 36 nM for **12**. This last value rivals the potency measured for Donepezil itself, 21 nM.

### 3. Discussion and conclusions

The above results substantiate three major conclusions. First, the quality of synthetic predictions appears satisfactory with 12 out of 13 analogs made according to the computer-designed routes. This is perhaps not unexpected given that predictions of programs such as Allchemy or Chematica/Synthia had previously been validated on targets more challenging than our analogs.<sup>3,8,10,44</sup> In this light, one can argue that all these simple syntheses could have been designed without much effort by a human expert – while this is true, the computer may be viewed as a useful “calculator” accelerating the straightforward but otherwise tedious steps of the design process, from the creation of replicas *via* substructure replacements, through the selection of commercially available substrates, to forward synthesis.

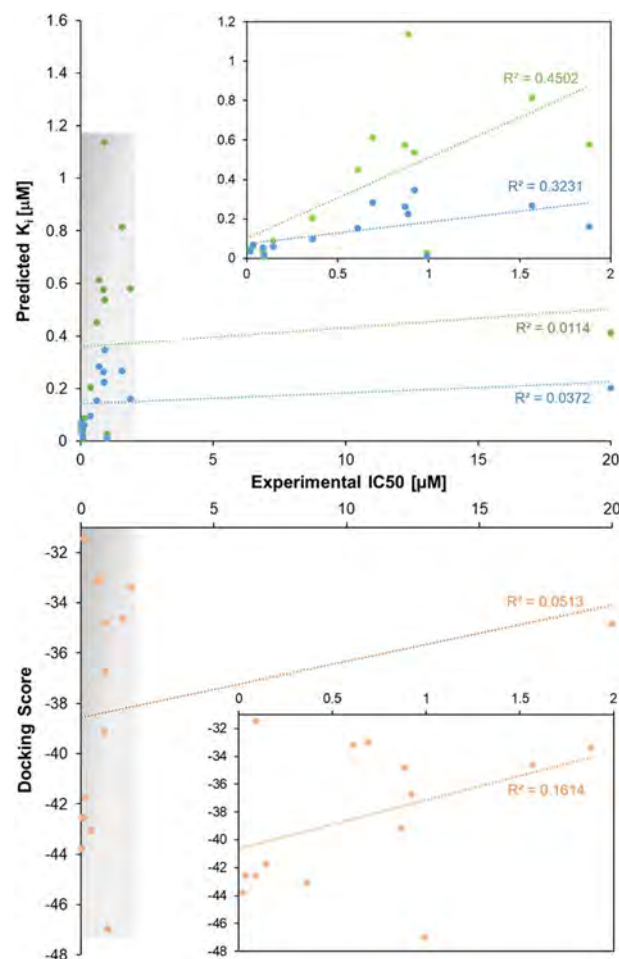


Fig. 5 The relationship between the experimental  $IC_{50}$  ( $\mu\text{M}$ ) and  $K_i$  ( $\mu\text{M}$ ) values (from AutoDock 4, blue, and AutoDock Vina, green) or the docking scores (from Dock 6, orange). Insets zoom on the regions which, in the full plots, are shaded in gray and do not include the poorly binding analog.

Second, whereas the synthesis part is robust, the prediction of analog's properties remains challenging. This is illustrated in Fig. 5 which plots the experimentally measured  $IC_{50}$ 's against the predictions of the docking programs – as seen, the correlations are quite poor and even discounting the outlier **1**, are limited to ~0.45 for AutoDock Vina (green markers in the inset to the upper plot). We observe that such values are in line with comparative studies of docking methods<sup>70</sup> where similar, low correlations were reported. In parallel, Allchemy's internal neural network trained on 1.75 million of protein assays from ChEMBL estimated the potency of our analogs to be micromolar, which is true for most but not all of them (*e.g.*, not for **1** which is  $>10$   $\mu\text{M}$  and not for **12** and **14** for which  $IC_{50}$  values are in tens of nM). Taken together, these results reinforce the view that neither the docking nor NN models can currently predict the affinity accurately – though they can perhaps be accurate to the order-of-magnitude, which can still be useful in distinguishing very poor from decent binders.

Third and last, it should be remembered that in developing potential drugs, binding affinity is but one of the important



metrics and one should also evaluate ADME-Tox properties. There are nowadays multiple machine-learning models to evaluate these properties (e.g., ref. 71–73 for hERG models, ref. 74 for plasma protein binding, PPB, or ref. 75–78 for blood-brain barrier, BBB, penetration) and some have also been implemented in Allchemy's pipeline (see user manual in ESI, Section S1†), offering respectable accuracies ~0.8–0.9 and realistic predictions (see Table 1 and its caption). Thus, one could also think of using these metrics as part of a multi-objective scoring function<sup>35</sup> to guide the synthesis of molecules that, ultimately, meet several desirable criteria at once. The multiobjective approach is, arguably, more conceptually elegant than the synthesize-and-then-evaluate pipeline we pursued here. On the other hand, it should be noted that ADME-Tox models suffer from the scarcity of publically available data (typically, few thousand molecules per model) and are largely untested (and likely less reliable) in out-of-box predictions. Using such metrics to make decisions about which molecules to synthesize can eliminate some interesting candidates from consideration. Our thinking when developing the “synthesis-first” pipeline was that synthesis planning – being the most robust component – should be unhindered by additional constraints, especially that it yields large collections of candidate molecules in very short times. These molecules can then be evaluated by other models/filters, with a human expert making choices which metrics to prioritize (or trust). This said, we recognize that if much larger synthetic spaces are to be explored (with much bigger substrate sets or higher beam width, *W*, see ref. 44 and 45), then multiobjective scoring should be considered to narrow down and accelerate this exploration.

## Data availability

All docking poses are deposited at <https://zenodo.org/records/14571461>. The source-code for guided network expansion is deposited at <https://zenodo.org/records/7371247>. Sample of the database of reaction templates is deposited at <https://zenodo.org/records/15001486>. All synthesis results are deposited under the “saved results” tab of the WebApp at <https://analogs.allchemy.net/> (see user manual in the ESI†). To test the WebApp under two-week, free academic access, please send an email from an academic address to [admin@allchemy.net](mailto:admin@allchemy.net) for access credentials.

## Author contributions

A. M. and S. M. performed syntheses. A. M. and A. Ż.-D. performed the binding assays. W. B., A. W., R. R., S. S., M. M., A. K. and K. M. developed the algorithms and the WebApp. A. Ż.-D. performed docking studies and supervised the syntheses. B. A. G. conceived and supervised the project.

## Conflicts of interest

W. B., A. W., R. R., S. S., M. M., A. K., K. M. and B. A. G. are consultants and/or stakeholders of Allchemy, Inc. Allchemy software is property of Allchemy Inc.

## Acknowledgements

Development of the Analogs module within the Allchemy platform (by W. B., A. W., R. R., S. S., M. M., A. K. and K. M.) was supported by internal funds of Allchemy, Inc. Syntheses performed by A. M. and analyses performed by A. K. were supported by the National Science Center, Poland (grant Maestro, # 2018/30/A/ST5/00529 to B. A. G.). Synthesis, binding assays and docking studies by A. Ż.-D. and S. M. were supported by the National Science Centre, Poland (grant SONATA 2020/39/D/ST4/01890 to A. Ż.-D.). Analysis of results and writing of the paper by B. A. G. was supported by the Institute for Basic Science, Korea (project code IBS-R020-D1). The authors thank the Mcule team for providing access to their catalog and standardizing the prices information therein to per-gram.

## References

- S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk and B. A. Grzybowski, *Angew. Chem., Int. Ed.*, 2016, **55**, 5904–5937.
- C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, *ACS Cent. Sci.*, 2017, **3**, 1237–1245.
- T. Klucznik, B. Mikulak-Klucznik, M. P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E. P. Gajewska, A. Touchkine, P. Dittwald, M. P. Startek, G. J. Kirkovits, R. Roszak, A. Adamski, B. Sieredzińska, M. Mrksich, S. L. J. Trice and B. A. Grzybowski, *Chem*, 2018, **4**, 522–532.
- M. H. S. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.
- C. W. Coley, D. A. Thomas, J. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison and K. F. Jensen, *Science*, 2019, **365**, eaax1566.
- A. F. de Almeida, R. Moreira and T. Rodrigues, *Nat. Rev. Chem.*, 2019, **3**, 589–604.
- P. Schwaller, A. C. Vaucher, R. Laplaza, C. Bunne, A. Krause, C. Corminboeuf and T. Laino, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1604.
- B. Mikulak-Klucznik, P. Gołębiowska, A. A. Bayly, O. Popik, T. Klucznik, S. Szymkuć, E. P. Gajewska, P. Dittwald, O. Staszewska-Krajewska, W. Beker, T. Badowski, K. A. Scheidt, K. Molga, J. Mlynarski, M. Mrksich and B. A. Grzybowski, *Nature*, 2020, **588**, 83–88.
- Y. Lin, R. Zhang, D. Wang and T. Cernak, *Science*, 2023, **379**, 453–457.
- T. Klucznik, L. D. Syntrivanis, S. Baś, B. Mikulak-Klucznik, M. Moskal, S. Szymkuć, J. Mlynarski, L. Gadina, W. Beker, M. D. Burke, K. Tiefenbacher and B. A. Grzybowski, *Nature*, 2024, **625**, 508–515.
- P. Jia, J. Pei, G. Wang, X. Pan, Y. Zhu, Y. Wu and L. Ouyang, *Green Synth. Catal.*, 2022, **3**, 11–24.
- A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev,



- Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo and A. Aspuru-Guzik, *Nat. Biotechnol.*, 2019, **37**, 1038–1040.
- 13 D. E. Graff, E. I. Shakhnovich and C. W. Coley, *Chem. Sci.*, 2021, **12**, 7866–7881.
- 14 T. J. Struble, J. C. Alvarez, S. P. Brown, M. Chytil, J. Cisar, R. L. Desjarlais, O. Engkvist, S. A. Frank, D. R. Greve, D. J. Griffin, X. Hou, J. W. Johannes, C. Kretsoulas, B. Lahue, M. Mathea, G. Mogk, C. A. Nicolaou, A. D. Palmer, D. J. Price, R. I. Robinson, S. Salentin, L. Xing, T. Jaakkola, W. H. Green, R. Barzilay, C. W. Coley and K. F. Jensen, *J. Med. Chem.*, 2020, **63**, 8667–8682.
- 15 A. V. Sadybekov and V. Katritch, *Nature*, 2023, **616**, 673–685.
- 16 R. Roszak, L. Gadina, A. Wołos, A. Makkawi, B. Mikulak-Klucznik, Y. Bilgi, K. Molga, P. Gołębiowska, O. Popik, T. Klucznik, S. Szymkuć, M. Moskal, S. Baś, R. Frydrych, J. Mlynarski, O. Vakuliuk, D. T. Gryko and B. A. Grzybowski, *Nat. Commun.*, 2024, **15**, 10285.
- 17 T. N. K. Raju, *Lancet*, 2000, **356**, 436.
- 18 J. Fischer and C. R. Ganellin, *Analogue-Based Drug Discovery II*, Wiley-VCH, 2010, DOI: [10.1002/9783527630035](https://doi.org/10.1002/9783527630035).
- 19 M. Hartenfeller and G. Schneider, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 742–759.
- 20 G. Schneider and U. Fechner, *Nat. Rev. Drug Discovery*, 2005, **4**, 649–663.
- 21 G. Schneider, M. L. Lee, M. Stahl and P. Schneider, *J. Comput. Aided Mol. Des.*, 2000, **14**, 487–494.
- 22 X. Q. Lewell, D. B. Judd, S. P. Watson and M. M. Hann, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 511–522.
- 23 H. M. Vinkers, M. R. de Jonge, F. F. D. Daeyaert, J. Heeres, L. M. H. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. Van Aken and P. A. J. Janssen, *J. Med. Chem.*, 2003, **46**, 2765–2773.
- 24 M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark and G. Schneider, *PLoS Comput. Biol.*, 2012, **8**, e1002380.
- 25 T. Hoffmann and M. Gastreich, *Drug Discovery Today*, 2019, **24**, 1148–1156.
- 26 U. Dolfus, H. Briem and M. Rarey, *J. Chem. Inf. Model.*, 2022, **62**, 3565–3576.
- 27 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.
- 28 M. H. S. Segler, T. Kogej, C. Tyrchan and M. P. Waller, *ACS Cent. Sci.*, 2018, **4**, 120–131.
- 29 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and B. Yoshua, in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2014, pp. 2672–2680.
- 30 D. C. Elton, Z. Boukouvalas, M. D. Fuge and P. W. Chung, *Mol. Syst. Des. Eng.*, 2019, **4**, 828–849.
- 31 J. Bradshaw, B. Paige, M. J. Kusner, M. H. S. Segler and J. M. Hernández-Lobato, *arXiv*, 2019, preprint, arXiv:1906.05221, DOI: [10.48550/arXiv.1906.05221](https://doi.org/10.48550/arXiv.1906.05221).
- 32 T. Blaschke, J. Arús-Pous, H. Chen, C. Margreitter, C. Tyrchan, O. Engkvist, K. Papadopoulos and A. Patronov, *J. Chem. Inf. Model.*, 2020, **60**, 5918–5922.
- 33 J. Bradshaw, B. Paige, M. J. Kusner, M. H. S. Segler and J. M. Hernández-Lobato, *arXiv*, 2020, preprint, arXiv:2012.11522, DOI: [10.48550/arXiv.2012.11522](https://doi.org/10.48550/arXiv.2012.11522).
- 34 S. K. Gottipati, B. Sattarov, S. Niu, Y. Pathak, H. Wei, S. Liu, K. M. J. Thomas, S. Blackburn, C. W. Coley, J. Tang, S. Chandar and Y. Bengio, *arXiv*, 2020, preprint, arXiv:2004.12485, DOI: [10.48550/arXiv.2004.12485](https://doi.org/10.48550/arXiv.2004.12485).
- 35 J. C. Fromer, D. E. Graff and C. W. Coley, *Digital Discovery*, 2024, **3**, 467–481.
- 36 W. Gao and C. W. Coley, *J. Chem. Inf. Model.*, 2020, **60**, 5714–5723.
- 37 P. Renz, D. Van Rompaey, J. K. Wegner, S. Hochreiter and G. Klambauer, *Drug Discovery Today: Technol.*, 2019, **32**, 55–63.
- 38 W. P. Walters, R. Barzilay and E. Opín, *Drug. Discovery*, 2021, **16**, 937–947.
- 39 M. Stanley and M. Segler, *Curr. Opin. Struct. Biol.*, 2023, **82**, 102658.
- 40 W. Gao, S. Luo and C. W. Coley, *arXiv*, 2024, preprint, arXiv:2410.03494, DOI: [10.48550/arXiv.2410.03494](https://doi.org/10.48550/arXiv.2410.03494).
- 41 S. K. Singh, K. King, C. Gannett, C. Chuong, S. Y. Joshi, C. Plate, P. Farzeen, E. M. Webb, L. Kumar-Kunche, J. Weger-Lucarelli, A. N. Lowell, A. M. Brown and S. A. Deshmukh, *J. Phys. Chem. Lett.*, 2023, **14**, 9490–9499.
- 42 D. Lowe, *AI Does Not Make It Easy (In the pipeline)*, *Science*, 2024, <https://www.science.org/content/blog-post/ai-does-not-make-it-easy>.
- 43 A. Wołos, R. Roszak, A. Żądło-Dobrowolska, W. Beker, B. Mikulak-Klucznik, G. Spólnik, M. Dygas, S. Szymkuć and B. A. Grzybowski, *Science*, 2020, **369**, eaaw1955.
- 44 A. Wołos, D. Koszelewski, R. Roszak, S. Szymkuć, M. Moskal, R. Ostaszewski, B. T. Herrera, J. M. Maier, G. Brezicki, J. Samuel, J. A. M. Lummiss, D. T. McQuade, L. Rogers and B. A. Grzybowski, *Nature*, 2022, **604**, 668–676.
- 45 R. Roszak, A. Wołos, M. Benke, Ł. Gleń, J. Konka, P. Jensen, P. Burghardt, A. Żądło-Dobrowolska, P. Janiuk, S. Szymkuć and B. A. Grzybowski, *Chem*, 2024, **10**, 952–970.
- 46 K. Molga, E. P. Gajewska, S. Szymkuć and B. A. Grzybowski, *React. Chem. Eng.*, 2019, **4**, 1506–1521.
- 47 F. Strieth-Kalthoff, S. Szymkuć, K. Molga, A. Aspuru-Guzik, F. Glorius and B. A. Grzybowski, *J. Am. Chem. Soc.*, 2024, **146**, 11005–11017.
- 48 B. A. Grzybowski, T. Badowski, K. Molga and S. Szymkuć, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2023, **13**, e1630.
- 49 P. Ertl, E. Altmann, S. Racine and R. Lewis, *Eur. J. Med. Chem.*, 2022, **238**, 114483.
- 50 B. Mikulak-Klucznik, T. Klucznik, W. Beker, M. Moskal and B. A. Grzybowski, *Chem*, 2024, **10**, 1319–1326.
- 51 S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist and E. Bjerrum, *J. Cheminform.*, 2020, **12**, 70.
- 52 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.
- 53 D. Liu, T. Ru, Z. Deng, L. Zhang, Y. Ning and F. E. Chen, *ACS Catal.*, 2023, **13**, 12868–12876.
- 54 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **16**, 2785–2791.





- 55 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
- 56 J. Eberhardt, D. Santos-Martins, A. F. Tillack and S. Forli, *J. Chem. Inf. Model.*, 2021, **61**, 3891–3898.
- 57 W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz and R. C. Rizzo, *J. Comput. Chem.*, 2015, **36**, 1132–1156.
- 58 ChEMBLdb, [http://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl\\_29/](http://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_29/), accessed February 28, 2025.
- 59 G. Landrum, *RDKit*, <http://www.rdkit.org>, accessed February 28, 2025.
- 60 M. Awale and J.-L. Reymond, *J. Chem. Inf. Model.*, 2014, **54**, 1892–1907.
- 61 J. L. Durant, B. A. Leland, D. R. Henry and J. G. Nourse, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1273–1280.
- 62 A. Mayr, G. Klambauer, T. Unterthiner, M. Steijaert, J. K. Wegner, H. Ceulemans, D.-A. Clevert and S. Hochreiter, *Chem. Sci.*, 2018, **9**, 5441–5451.
- 63 N. Bosc, F. Atkinson, E. Félix, A. Gaulton, A. Hersey and A. R. Leach, *J. Cheminform.*, 2019, **11**, 64.
- 64 K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen and R. Barzilay, *J. Chem. Inf. Model.*, 2019, **59**, 3370–3388.
- 65 S. Asghar, N. Mushtaq, A. Ahmed, L. Anwar, R. Munawar and S. Akhtar, *Molecules*, 2024, **29**, 490.
- 66 SigmaAldrich, Cyclooxygenase 2 human  $\geq 70\%$  (SDS-PAGE), 0858UN-UN1000UN, <https://www.sigmaaldrich.com/PL/en/product/sigma/c0858>, accessed 8 April 2025.
- 67 R. Roszak, W. Beker, K. Molga and B. A. Grzybowski, *J. Am. Chem. Soc.*, 2019, **141**, 17142–17149.
- 68 H. Sugimoto, Y. Iimura, Y. Yamanishi and K. Yamatsu, *J. Med. Chem.*, 1995, **38**, 4821–4829.
- 69 SigmaAldrich, *Acetylcholinesterase Inhibitor Screening Kit, 324KT-KT1KT*, <https://www.sigmaaldrich.com/PL/en/product/sigma/mak324>, accessed 8 April 2025.
- 70 A. Pecina, J. Fanfrlík, M. Lepšík and J. Řezáč, *Nat. Commun.*, 2024, **15**, 1127.
- 71 K. Ogura, T. Sato, H. Yuki and T. Honma, *Sci. Rep.*, 2019, **9**, 12220.
- 72 H. Kim and H. Nam, *Comput. Biol. Chem.*, 2020, **87**, 107286.
- 73 X. Zhang, J. Mao, M. Wei, Y. Qi and J. Z. H. Zhang, *J. Chem. Inf. Model.*, 2022, **62**, 1830–1839.
- 74 Z. Han, Z. Xia, J. Xia, I. V. Tetko and S. Wu, *Eur. J. Pharm. Sci.*, 2025, **204**, 106946.
- 75 F. Meng, Y. Xi, J. Huang and P. W. Ayers, *Sci. Data*, 2021, **8**, 289.
- 76 V. Kumar, A. Banerjee and K. Roy, *J. Chem. Inf. Model.*, 2024, **64**, 4298–4309.
- 77 S. Cherian Parakkal, R. Datta and D. Das, *Mol. Inf.*, 2022, **41**, 2100315.
- 78 B. Mazumdar, P. K. Deva Sarma, H. J. Mahanta and G. N. Sastry, *Comput. Biol. Med.*, 2023, **160**, 106984.
- 79 W. Beker, A. Wołos, S. Szymkuć and B. A. Grzybowski, *Nat. Mach. Intell.*, 2020, **2**, 457–465.
- 80 S. A. Wildman and G. M. Crippen, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 868–873.
- 81 Y. J. Chae, H. J. Lee, J. H. Jeon, I.-B. Kim, J.-S. Choi, K.-W. Sung and S. J. Hahn, *Brain Res.*, 2015, **1597**, 77–85.
- 82 S. M. Vogel, L. M. Mican and T. L. Smith, *Ment. Health Clin.*, 2019, **9**, 128–132.
- 83 L. Brandolini, A. Antonosante, C. Giorgio, M. Begnasco, M. d'Angelo, V. Castelli, E. Benedetti, A. Cimini and M. Allegretti, *Sci. Rep.*, 2020, **10**, 18337.
- 84 P. J. Hayball, R. L. Nation, F. Bochner, J. L. Newton, R. A. Massy-Westropp and D. P. G. Hamon, *Chirality*, 1991, **3**, 460–466.
- 85 S. L. Rogers, N. M. Cooper, R. Sukovaty, J. E. Pederson, J. N. Lee and L. T. Friedhoff, *Br. J. Clin. Pharmacol.*, 1998, **46**, 7–12.
- 86 W. A. Banks, *Adv. Drug. Del. Rev.*, 2012, **64**, 629–639.
- 87 H. Kokki, *Pediatric Drugs*, 2010, **12**, 313–329.



## SUPPLEMENTARY INFORMATION

### **Retro-forward synthesis design and experimental validation of potent structural analogs of known drugs.**

Ahmad Makkawi<sup>1+</sup>, Wiktor Beker<sup>2+</sup>, Agnieszka Wołos<sup>2+</sup>, Sabyasachi Manna<sup>1</sup>, Rafał Roszak<sup>2</sup>, Sara Szymkuć<sup>2</sup>, Martyna Moskał<sup>2</sup>, Aleksei Koshevarnikov<sup>1,2</sup>, Karol Molga<sup>2</sup>, Anna Żądło-Do-browolska<sup>1\*</sup> and Bartosz A. Grzybowski<sup>1,3,4\*</sup>

<sup>1</sup> Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland.

<sup>2</sup> Allchemy, Inc., Highland, IN, USA.

<sup>3</sup> Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea.

<sup>4</sup> Department of Chemistry, Ulsan Institute of Science and Technology, UNIST, Ulsan 44919, Republic of Korea.

+ Authors contributed equally

\*Correspondence to: [anna.zadlo@icho.edu.pl](mailto:anna.zadlo@icho.edu.pl) or [nanogrzybowski@gmail.com](mailto:nanogrzybowski@gmail.com)

## **TABLE OF CONTENTS**

**Section S1.** Allchemy User Guide. (p.S-3-)

**S1.1.** General information.

**S1.2.** Setting up new calculations.

**S1.3.** Browsing results.

**S1.4.** Analysis of results.

**S1.5.** Graph view.

**S1.6.** Saved results.

**Section S2.** A comment on the selection of starting materials based on the similarity to parent (p.S-16-)

**Section S3.** Addition of simple chemicals to the substrate set. (p.S-18-)

**Section S4.** Molecular docking Studies. (p.S-20-)

**S4.1.** AutoDock 4.2.6.

**S4.2.** Dock 6.12.

**S4.3.** AutoDock Vina 1.2.5.

**Section S5.** Procedures of the biological analysis. (p.S-29-)

**S5.1.** Cyclooxygenase-2 (COX-2) inhibition assay.

**S5.2.** Acetylcholinesterase (AChE) inhibition assay.

**Section S6.** Synthetic details. (p.S-32-)

**S6.1.** Ketoprofen's analogs described in main-text **Figure 3.a.**

**S6.2.** Donepezil's analogs described in main-text **Figure 4.a.**

**Section S7.** Tutorial for reaction rule coding. (p.S-76-)

**Section S8.** Supplementary references. (p.S-84-)

## Section S1. User Guide for Allchemy's WebApp associated with the article

### S1.1. General information

A WebApp to perform new calculations associated with this article is available for testing by academic users at <https://analogs.allchemy.net/> (upon sending a registration email from an academic address, and on a rolling basis in two-week slots). Because of server limitations, some constraints for calculations were added, namely: maximum number of generations - 5, maximum time of single calculation – 30 minutes. Because the graph view of results is based on SVG2 standard, it is recommended to use Google Chrome or other browsers using the same engine, e.g., Chromium or Opera.

### S1.2. Setting up new calculations

After logging in, user will be redirected to the “*New search*” tab (**Figure S1**), whereby calculations can be initiated. To set up the search, follow the instructions below:

- 1) Specify (optionally) allowed reaction conditions (**I**). It is possible to narrow down the search only to reactions occurring at particular temperatures, conditions and solvent classes.
- 2) Introduce the lead compound (**II**) using one or more input options: drawing editor, entering SMILES (“from text” option) or uploading .txt file (with SMILES) or .mol file. Only one target per calculation can be used.
- 3) Use “*propose*” functionality to automatically generate substrate collection or/and enter your own collection of substrates (**III**, from file, from editor, from text). When the “*propose*” functionality is selected, pop-up window with adjustable parameters appears (**Figure S2**).
- 4) Select motifs for substructure replacements. Allchemy automatically recognizes which types of motifs are present in the structure of the lead and presents the user only with possible options. All checkboxes are selected in default settings, but mind that more options selected will result in a larger set of substrates. Substructure replacements selected in calculations described in this publication were “*Non-fused 6-membered aromatic rings*” and “*Carboxylic acids*” for Ketoprofen and “*Ethers, Secondary Amines, Thioethers*” and “*Ketones*” for Donepezil. After clicking on the “*propose*” button, retrosynthetic calculations commence (it can take up to ~5 minutes) and the user will be redirected to the window with automatically generated starting materials and a set of auxiliary reagents (**Figures S3,S4**).
- 5) Review generated set of substrates (**Figure S3**). Generated building blocks, auxiliary reagents and compounds introduced by the user can be browsed separately, by clicking on the corresponding tabs (a). Selection of compounds via clicking on the structures (background of selected molecule(s) will change to blue) allows to remove unwanted starting materials (remove selected, b) or to narrow down the set only to selected

compounds (keep selected, c). Click on the “*next page*”/“*prev page*” buttons (d) to browse all substrates. When your set of starting materials is finalized, click on the “x” button in the top-right corner (e).

- 6) Displaying the target and/or substrates as well as further modifications are possible in the “*Show selected*” section (**Figure S1, IV**). Clicking on “*Show molecules*” in the “*Reactants*” section will re-open the window shown in **Figure S3**.
- 7) Set the number of synthetic generations (max. 5; **V**) and limit for maximum molecular weight for all generated molecules (typically set to 100-150 g/mol more than the molecular mass of the parent compound).
- 8) Adjust the beam-width (parameter  $W$ , **VI**) that prunes the molecules generated in each forward generation. Pruning slider set in the middle (default) corresponds to  $W = 150$ . Lower pruning translates into a higher  $W$  value whereas higher pruning, to a lower  $W$  value. Searches with higher  $W$  generate larger networks. They can sometimes be beneficial as they can incorporate intermediates that are, initially, not the most similar to target but, after additional modification, become so. Searches with lower  $W$  are more focused and more “greedy” in the sense that they accept intermediates whose similarity to target keeps increasing with synthetic generations,  $G$ .
- 9) Additional limits for intermediates can be set (optionally, **VII**): maximum number of heavy atoms, chiral centers and halogen atoms.
- 10) Set a search name for a calculation (optionally, **VIII**). They will be available in the “*Saved results*” tab under this name when calculations are finished.
- 11) Click on the “*Search*” button (**IX**) to start calculation. Depending on search settings and number of substrates, calculation should take between 5 and 30 minutes. Upper limit of calculation time is set to 30 minutes and after this time, calculations will be stopped and results, up to the last completed synthetic generation, will be displayed.

(I) **Select conditions**

Temperature range:

very low  low  rt  high  very high

Reaction conditions:

lewis acid  strongly acidic  acidic  mildly acidic  neutral  mildly basic  basic  strongly basic

Solvent(s):

polar  nonpolar  protic  aprotic

**Specify starting materials**

Specify molecules

(II) **Lead:** from file, from editor, from text

(III) **Reactants:** from file, from editor, from text, propose

(IV) **Show selected**

**Lead:** No molecules yet, Show molecules, Copy as smiles

**Reactants:** No molecules yet, Show molecules, Copy as smiles

**Search parameters**

(V) Number of synthetic generations:   
 Max molecular weight for intermediates:

**Advanced options**

(VI) Pruning low  high

(VII) Set limits for intermediates:

Max number of heavy atoms:   
 Max number of chiral centers:   
 Max number of halogens:

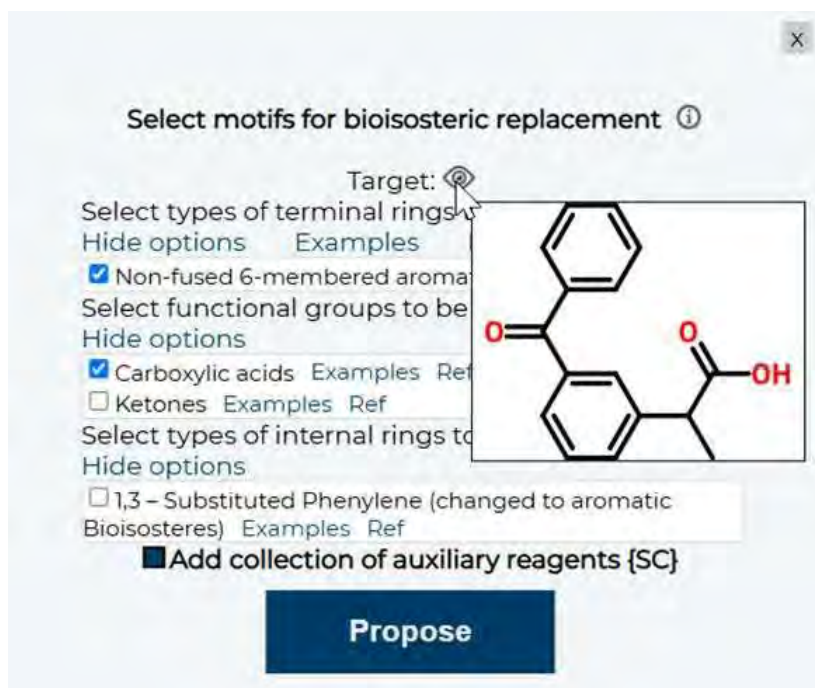
(VIII) Search name:

(IX) **Search**

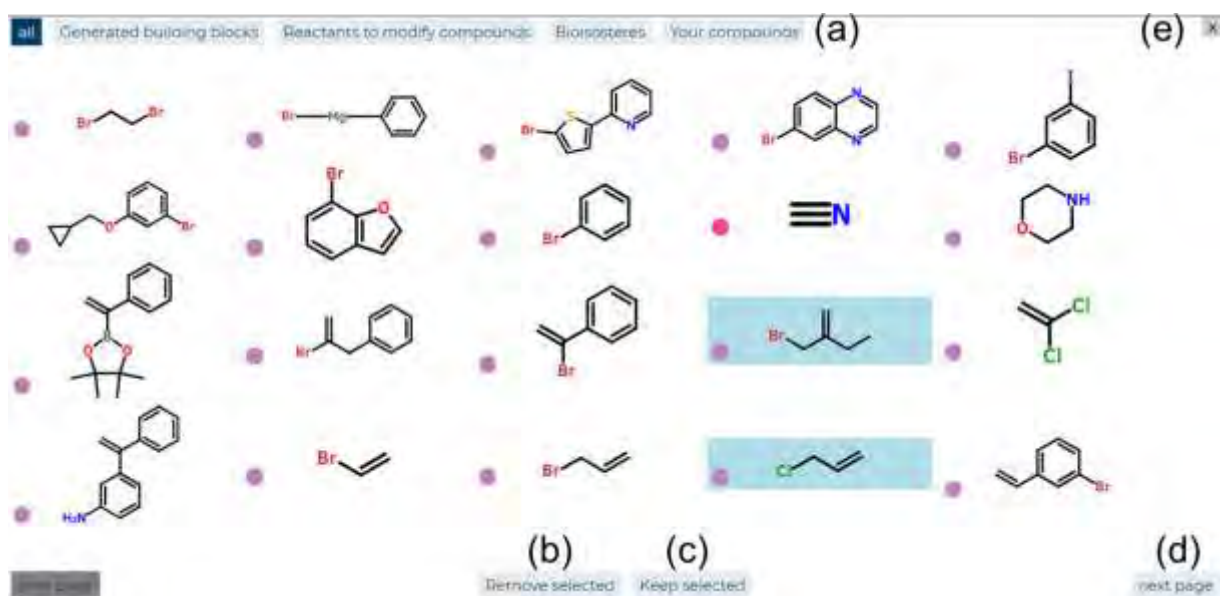
Clear form

**Figure S1. New search tab.** Setting up a new search requires: I) Constraining (optionally) allowed reaction conditions (temperature range, reaction conditions, solvents). II) Specification of a lead compound, III) Selection of starting materials via automatic “*propose*” functionality or/and addition of one’s own collection of substrates, IV) Reviewing of target and reactants (optional), V) Setting number of synthetic generations and mass limits and, optionally, VI) Adjustment of the pruning slider (beam-width parameter,  $W$ ), VII) Setting additional limits for

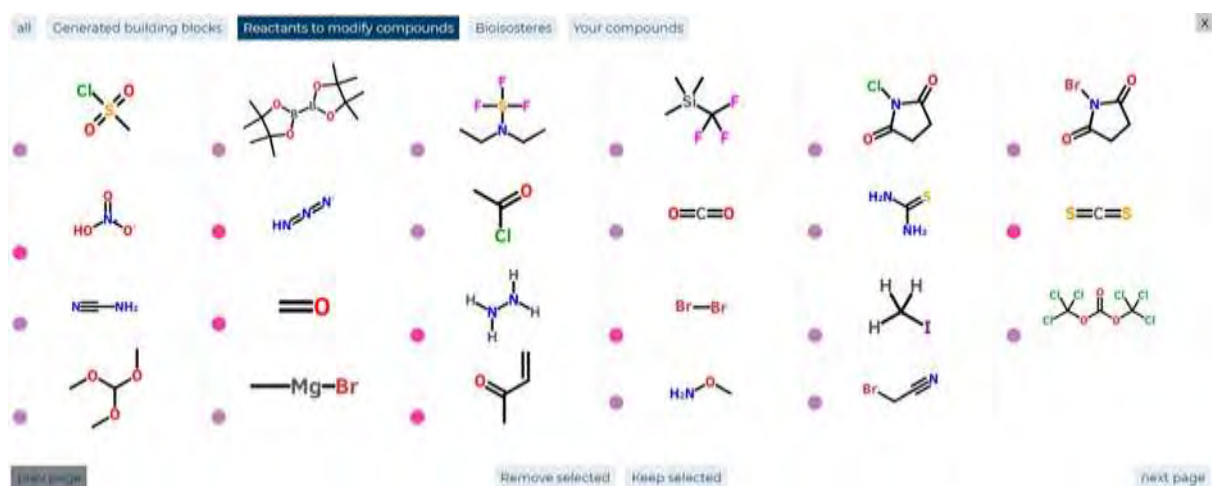
intermediates, and VIII) Naming calculations. Clicking on the “Search” button (IX) will start the calculations.



**Figure S2. Allchemy’s panel for the selection of substructure replacements.** Allchemy automatically recognizes which types of motifs are present in the structure of the lead and presents the user with possible replacement options. All checkboxes are selected in default settings, but please mind that more options selected will result in bigger set of substrates. Hovering over an “eye” icon will display structure of the parent for which analogs are sought. The replacements are divided into three categories: terminal rings, internal rings and functional groups, with each category further split into subcategories. e.g., *Non-fused 6-membered aromatic rings*, *Fused aromatic rings*, *Aliphatic rings*. Replacement of internal rings is performed based on the substitution pattern, for example, 2,5-disubstituted thiophene and 2,5-disubstituted pyridine are proposed as replacements for 1,4-substituted benzene whereas 2,4-disubstituted furan and 2,4-disubstituted pyrimidine are replacements for 1,3-substituted benzene (based on Tables 34 and 53 from *J. Med. Chem.* 2021, 64, 14046–14128).



**Figure S3. Browsing the set of proposed starting materials.** a) Building blocks generated from retrosynthesis of the parent/target molecules and its “replica” molecules derived from substructure replacements, set of simple auxiliary reagents and substrates/reagents introduced by the user can be displayed separately, by clicking on the corresponding tab. Selection of compounds (left-click on the structure) allows to b) remove or c) keep selected molecules. d) Click on the “next page” and “prev page” buttons to browse all the substrates. e) Close the window using the “x” button located in the top-right corner.



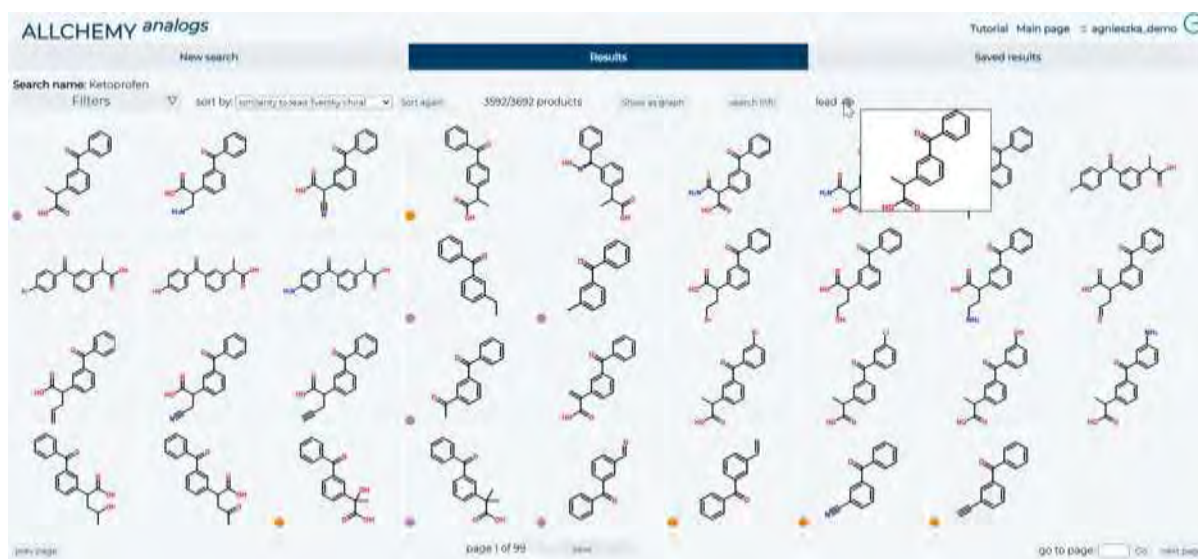
**Figure S4. Set of 23 auxiliary reagents – popular chemicals chosen for synthetic versatility.**

### S1.3. Browsing results

After calculation finishes, the user will be redirected to the results page. Generated compounds are displayed as panels of molecular structures (**Figure S5**). To browse the structures, user can use “next page” and “prev page” buttons located under the structures or use “go to page”

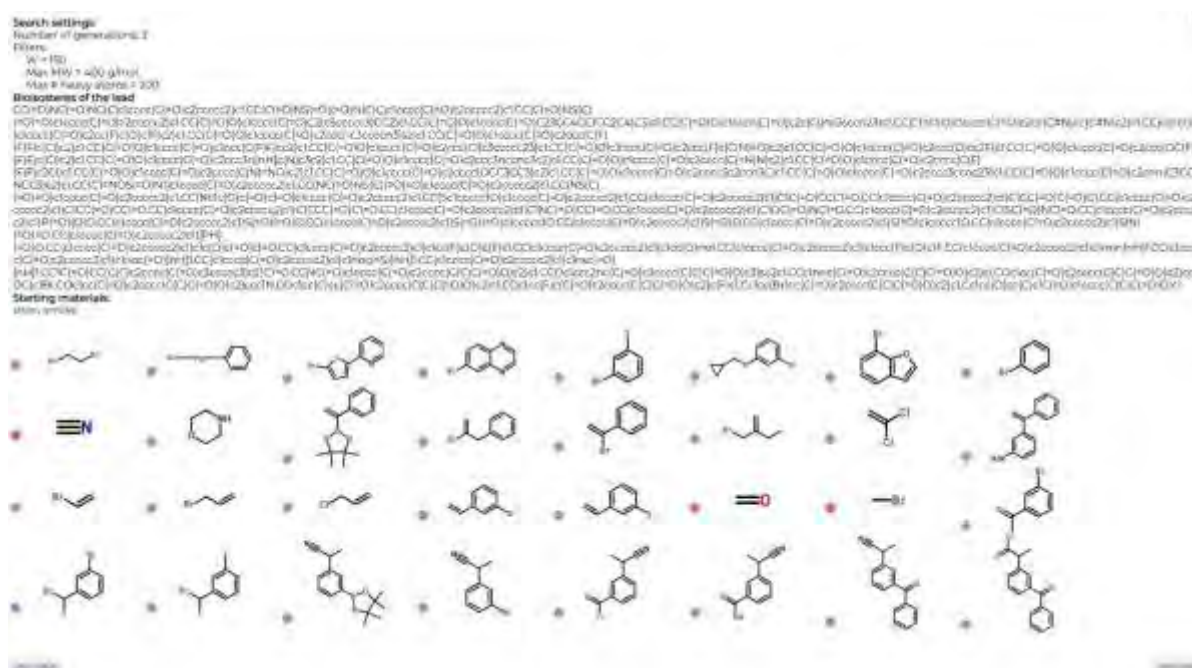


function (by typing number of the page of interest and clicking “go”). To visualize the parent/target, user can hover over the “eye” icon.



**Figure S5.** First page of the results of a calculation (here, analogs of Ketoprofen target) displayed as a default panel view.

Button “*search info*” located in the top panel allows the user to display information about search parameters (Figure S6) and lists of “replica” molecules of the lead (as list of SMILES) generated by substructure replacement as well as starting materials used in this calculation (both as structures and a list of SMILES).



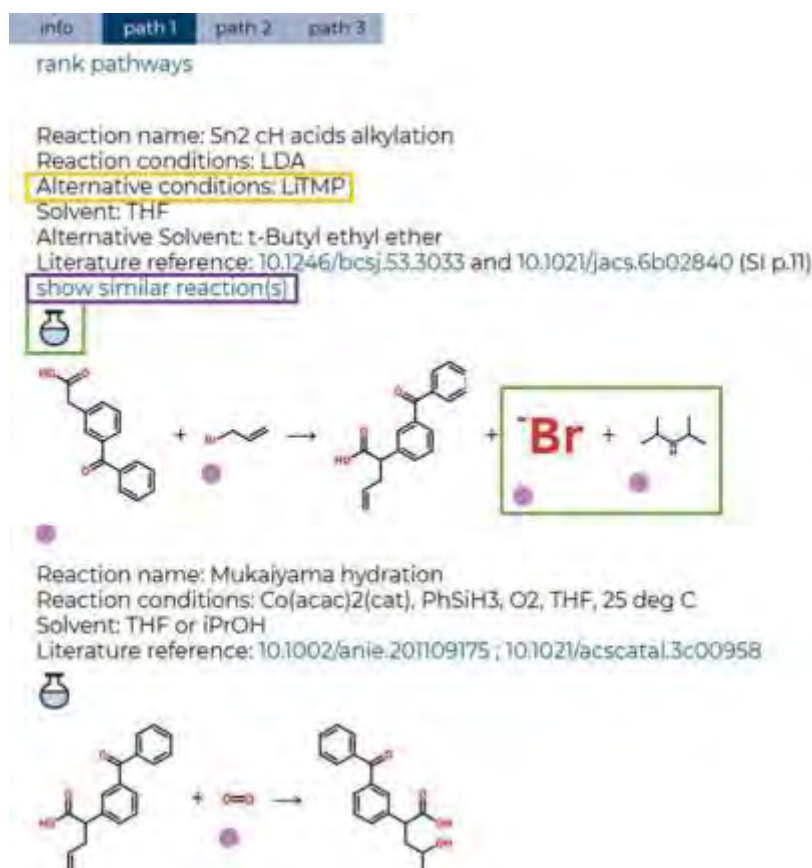
**Figure S6.** Information about search available after clicking “*search info*” button. Search parameters are listed at the top of the window. List of the “replica” molecules of the parent as

SMILES is available in the middle whereas structures of substrates are displayed along with corresponding list to SMILES and placed at the bottom of the pop-up window.

Clicking on any of the structures from **Figure S5** displays details of its synthesis – and, usually, multiple synthetic pathways found. As illustrated in **Figure S7**, each step lists not only the reaction scheme, but also reaction name, typical reaction conditions, typical solvent and illustrative literature reference(s) for this type of chemistry (mostly as DOI hyperlinks). Syntheses designed by the program also take into consideration environmental issues and practical aspects. In reactions involving harmful reagents or non-green solvents, greener alternatives are proposed, if a reasonable replacement is possible. For example, LiTMP is suggested as an alternative for LDA – highly flammable base (yellow frame in **Figure S7**). A practically important feature is inclusion of reaction by-products. After clicking on the flask icon, structures of by-products are displayed after the main product (green frame in **Figure S7**). In the highlighted example, diisopropylamine is generated from LDA during deprotonation of a carboxylic acid substrate. Additionally, if a proposed reaction has similar literature precedents in patent literature, “*show similar reaction(s)*” hyperlink will be available (purple frame in **Figure S7**). Clicking on the hyperlink opens a pop-up window with examples of similar reactions from patent literature. **Figure S8** shows four such reactions, similar to the specific instance of  $\alpha$ -alkylation of active methylene compound from **Figure S8**.

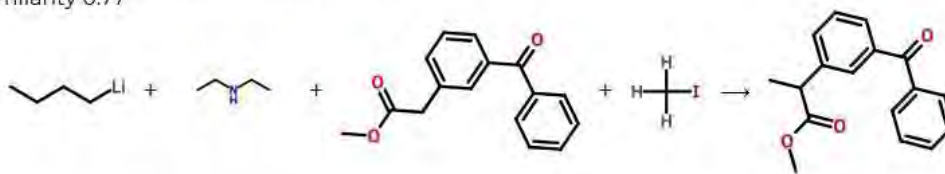
Additional pathways (if available) and molecular/pKa information about the target compound can be displayed by clicking on the appropriate grey tabs at the top of the window (**Figure S9**). Predictions of machine-learning models can also be found in the “*info*” tab for every molecule, namely:

- hERG cardiotoxicity (on the 0-1 scale with values <0.5 being desirable)
- degree of human plasma protein binding, hPPB, related to drug absorption and efficacy (0-100%, with very high values less desirable)
- MDCK-MDR1 efflux ratios estimating the degree to which the molecules are pumped out of the cells (values above ~2 suggest undesirable, active efflux)
- blood-brain barrier penetration (0–1, with higher values reflecting higher probability of BBB penetration)
- drug-likeness
- list of 20 protein targets for which a given molecule is predicted to have the highest affinity

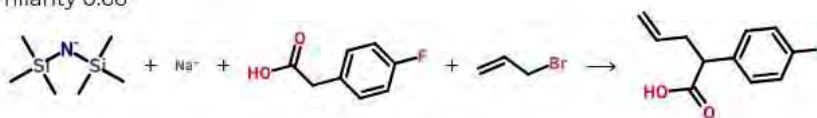


**Figure S7. Example of synthetic details available upon clicking on the molecule.** Additionally, other pathways (if available) and information about the molecule can be found on this page (by clicking appropriate grey buttons at the top). Each reaction scheme is accompanied by reaction name, typical reaction conditions, typical solvent and illustrative literature reference(s) for this type of chemistry. Alternative solvents/reagents are proposed for reactions involving harmful/non-green conditions, when reasonable replacement is possible (yellow frame). Clicking upon the flask icon displays by-products of a given reaction (green frame). “*Show similar reactions*” hyperlink, available for some of the reactions, enables the user to review instances of similar reactions from the patents (purple frame and **Figure S8**).

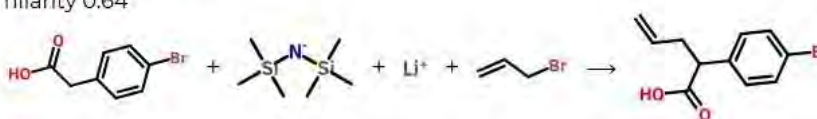
reference: US03995056 AND US03931302  
similarity 0.77



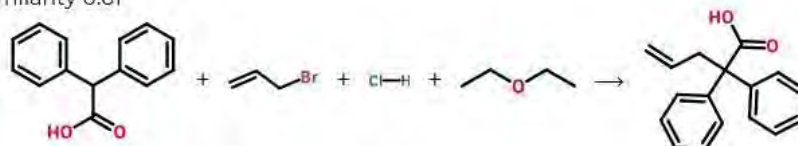
reference: US08349880B2  
yield: 43%  
similarity 0.66



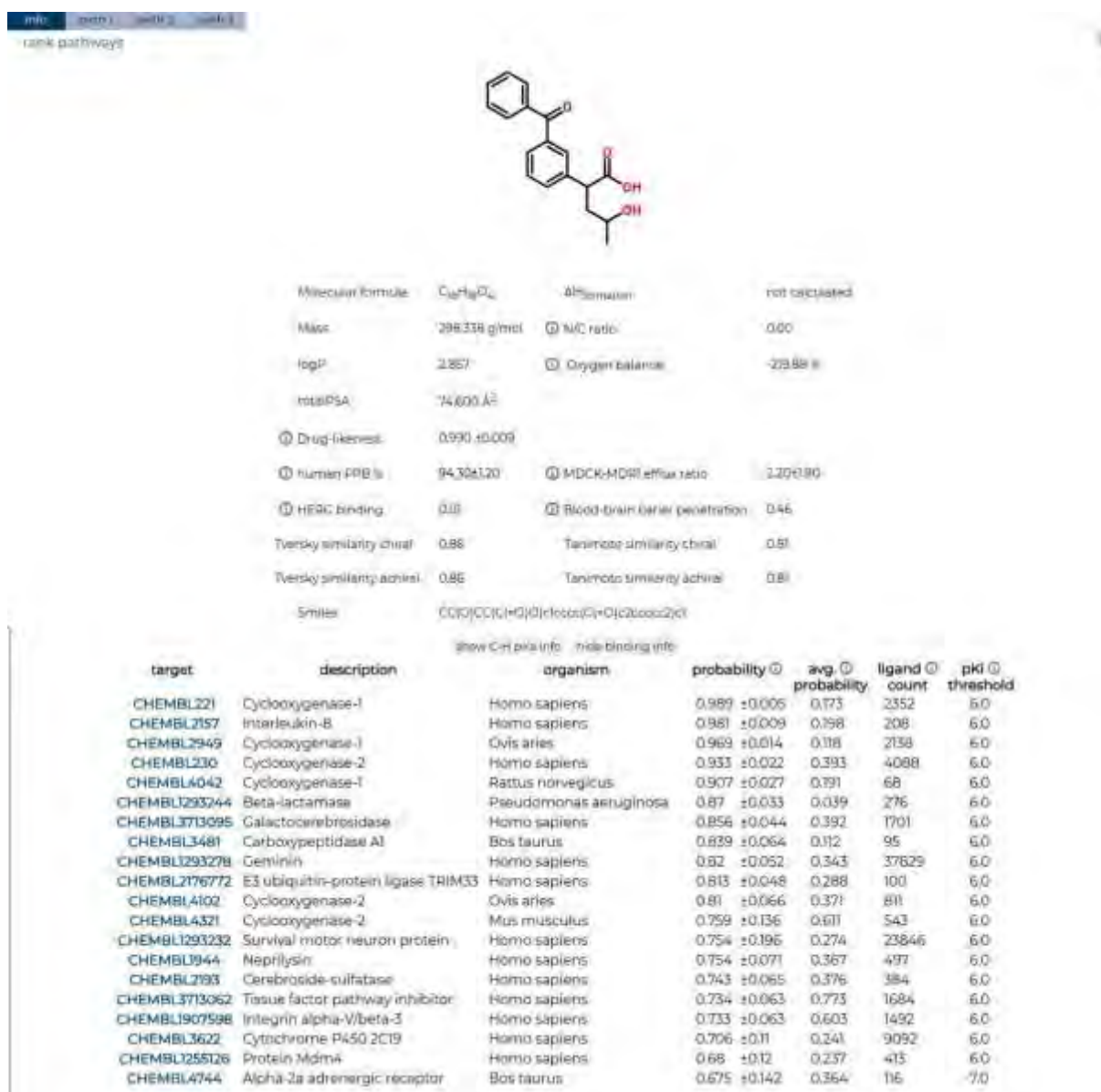
reference: US05708169  
similarity 0.64



reference: US05039691  
similarity 0.61



**Figure S8. Examples of similar reactions.** This window appears upon clicking on the “*show similar reactions*” hyperlink (purple frame in **Figure S8**). Examples are sorted by Tanimoto-based similarity to original reactions. Each reaction is accompanied by patent number(s) and similarity to original transformation. Some records contain also information about reaction yield (2<sup>nd</sup> transformation from the top). Reagents and sometimes solvents are often included in reaction visualization.



**Figure S9. Molecular information and predictions of machine-learning models can be displayed for any molecule upon clicking on the “*info*” tab.** List of 20 molecular targets with the highest predicted probability of binding contains: ChEMBLID as a hyperlink to the ChEMBL database, name of the molecular target, organism, predicted probability of binding, average activity probability for all ChEMBL ligands, number of ligands for this target in the training set and pKi threshold for the target, which is used to determine if a ligand is active.

#### S1.4. Analysis of results

Analysis of the results can be performed using “sort” and “filter” functionalities. In addition to sorting by similarity to target (default setting), user can choose from the list of nine parameters, i.e., mass, number of rings and oxygen balance (right part of **Figure S10**).

The image shows a software interface for chemical analysis. On the left, there is a 'Filters' panel with several sections:

- AI filters:** drug-likeness, binding, toxicity.
- ADME Filters:** human PPB %, MDCK-MDR1 efflux ratio, hERG binding, Blood-brain barrier penetration.
- Substructure filters:** Filter by PAINS, Exclude undesired motifs (Exclude all, custom selection), Filter by presence of chemical element, Substructure(s).
- Analog-Specific Filters:** Exclude intermediates, Show only bioisosteres similarity to lead.
- Structural filters:** Expand all filters, Undo changes, mass, logP, # H-bond acceptors, # H-bond donors, PSA (polar surface area), # of aromatic rings, # of aliphatic rings, # of rings formed, Fraction of heteroatoms, # of acidic groups, # of basic groups, # of acidic and basic groups, # of chiral centers, Fraction of sp<sup>3</sup> carbons, # of acyclic amides, # of rotatable bonds, # of halogen atoms, # of heavy atoms, # of all atoms, oxygen balance, Reset filters, Apply filters.

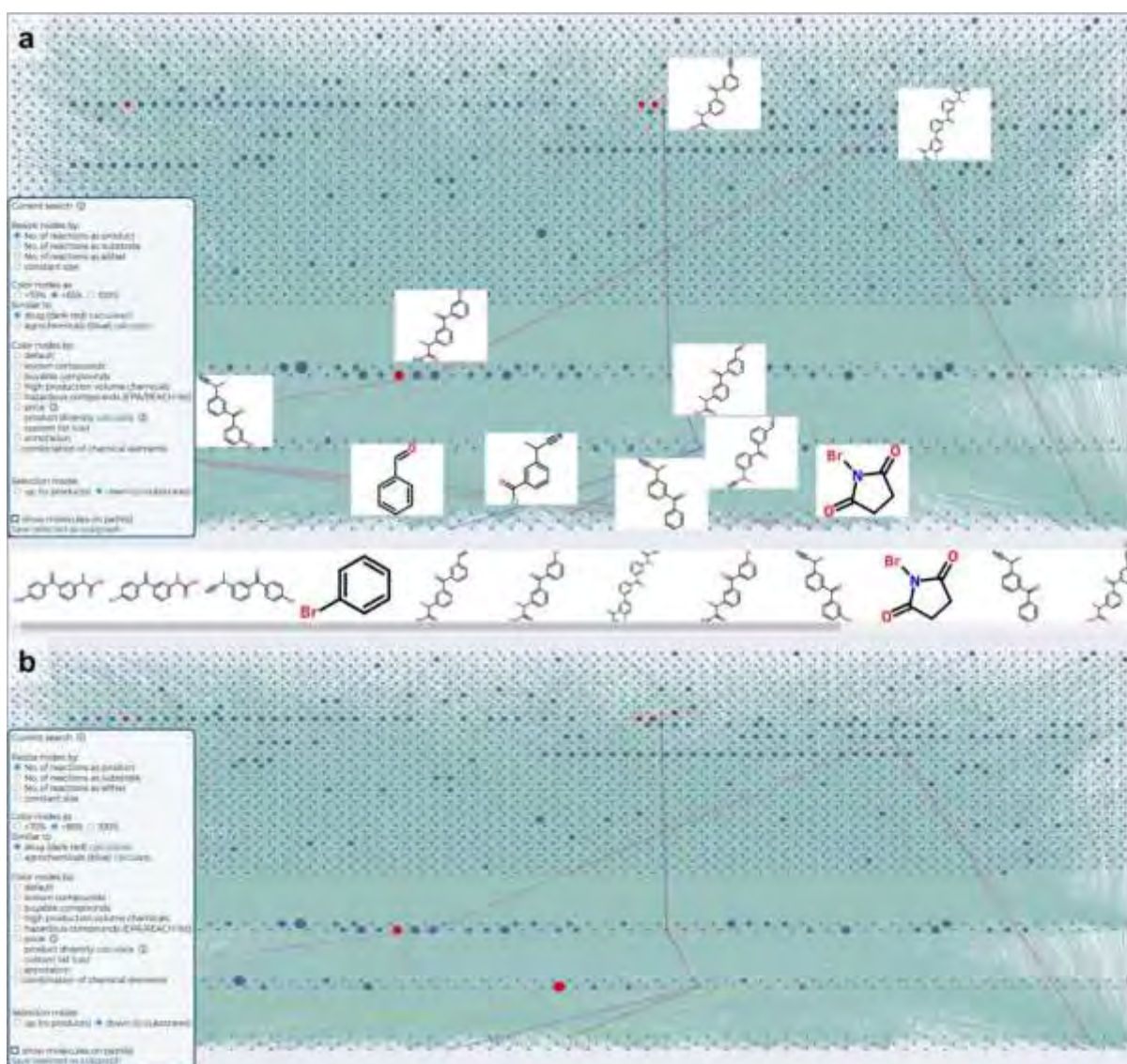
On the right, there is a 'sort by' dropdown menu with the following options: mass, # rings, # stereo centers, Balaban index, Bertz index, N/C ratio, oxygen balance, best binding, similarity to lead Tversky chiral (highlighted), similarity to lead Tversky achiral, similarity to lead Tanimoto chiral, similarity to lead Tanimoto achiral. Below the menu, there are several chemical structures of molecules, some with red and blue highlights.

List of the filters on the left comprises 5 general classes: AI filters for prediction of drug-likeness, binding to protein targets and toxicity; ADME filters - machine-learning models for prediction of human plasma protein binding, MCDK-MDR1 efflux ratio, hERG binding, blood-brain barrier penetration; Substructure filters to exclude undesired motifs, PAINS motifs or filter by presence of chemical elements or user-defined substructures; Analog-specific filters enable the user to exclude intermediates (molecules with functional groups absent in any drugs, selected by default), display only bioisosteres of the lead or molecules above user-defined similarity threshold; Structural filters with twenty physico-chemical and structural properties, i.e., mass, logP, number of aromatic or aliphatic rings (left part of **Figure S10**).

**Figure S10. Analysis of results can be performed via sorting (right) and/or filtering (left).** Filters comprise 5 general classes: AI filters, ADME filters, Substructure filters, Analog-specific filters and Structural filters. Multiple filters can be used simultaneously.

## S1.5. Graph view

In addition to the default panel view, results of the calculation can be displayed as a reaction graph/network (**Figure S11**). This modality is useful to track synthetic pathways leading to selected molecule(s) as well as for further analysis of results. Menu in the bottom-left corner allows to color or resize nodes of the graph, corresponding to particular molecules, by multiple parameters, i.e., number of incoming or outgoing connections, similarity to drugs or agrochemicals (calculated using Tanimoto similarity between Morgan fingerprints) or combination of chemical elements the molecules contain. Additionally, checkbox “show molecules on paths” may facilitate navigating the graph by displaying all intermediates next to the nodes (when number of structures is small enough, **Figure S11.a**) or as a panel over the graph (when number of structures would be too big to assure comfortable work with the graph, **Figure S11.b**).



**Figure S11. Results of calculations** (here, for the Ketoprofen parent) presented as a reaction graph/network. All molecules resulting from particular calculations are stacked into “layers” according to the synthetic generation,  $G_i$ , in which they were produced; nodes representing

substrates are in bottom layer. Hovering over any node displays the structure of the corresponding molecule, left-clicking on the node opens a window with synthetic pathway(s), while right-clicking traces on the graph the pathway leading to this molecule (colored lines connecting product with substrates via intermediates). Expandable panel located in the bottom-left corner allows for further graph analysis. The nodes can be colored by, i.a., similarity of a given molecule to existing drugs, known or hazardous compounds. Additionally, it is possible to resize nodes by number of reactions in which a molecule serves as product ( $k_{in}$ ), substrate ( $k_{out}$ ) or either. Here, nodes are resized by number of incoming connections (reactions leading to the molecule) and molecules with similarity to drugs  $>0.85$  are colored red. “*Selection mode*” section of the expandable panel allows to change the type of action resulting from right-clicking on the compound. In the default selection mode (“*down*”, illustrated in this Figure), right-clicking on the node traces the pathway generated in the smallest number of generations leading to this molecule, while in mode “*up*”, it traces all reactions in which the molecule serves as a substrate. Additionally, to facilitate the analysis of results, it is possible to use “*show molecules on path(s)*” checkbox. When selecting pathways with this functionality turned on, miniatures of all molecules participating in a given pathway will be displayed **a)** next to their corresponding nodes or **b)** as a panel over the graph.

## S1.6. Saved results

All calculations performed by the user will be automatically saved in “*Saved results*” tab under user-specified name. If the *search name* was not provided during search set-up, calculations will be saved under software-generated name. The result name can be changed anytime, using “pencil” icon located next to the current name. Additionally, all users will be provided with pre-calculated results described in this publication under the names “Ketoprofen” and “Donepezil” (**Figure S12**). Saved results can be filtered and sorted using different key-parameters, with “Recent” option (newest calculations on the top) being default.

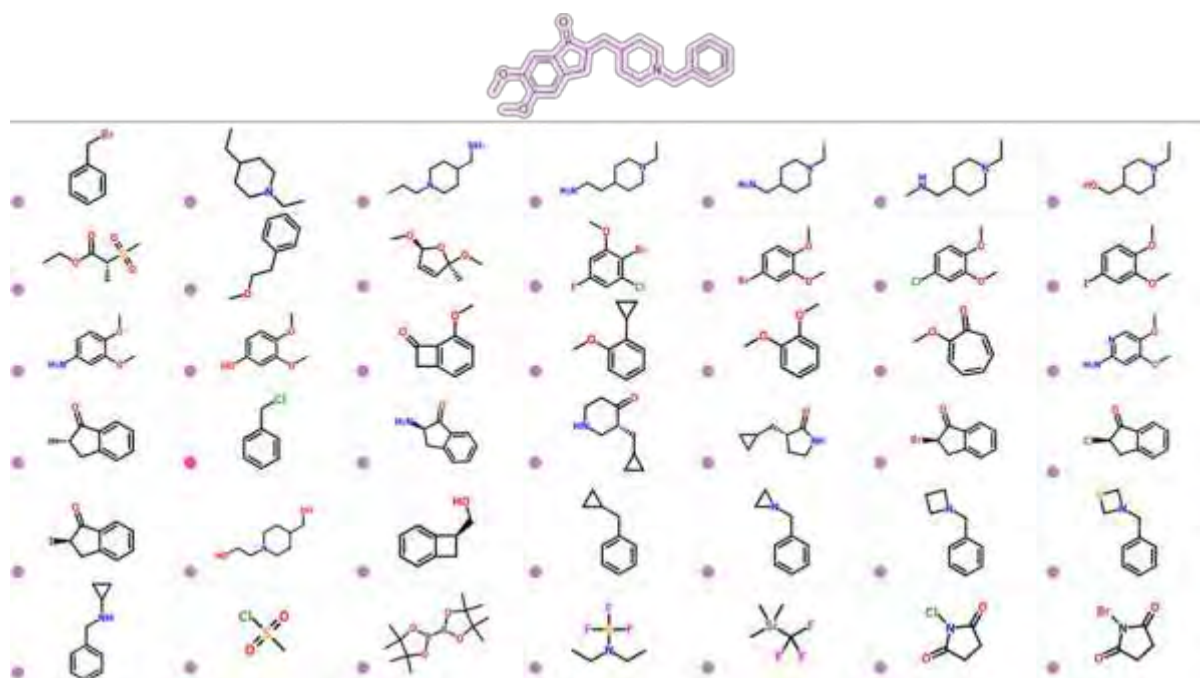
Name	Date	Num. results
Donepezil	2024-12-18 13:10	3619
Ketoprofen	2024-12-18 13:10	3692

**Figure S12. Saved results tab.** All finished calculations are automatically saved and available for future analysis in the “*Saved results*” tab. Additionally, all users will be provided with pre-calculated results described in this publication under the names “Ketoprofen” and “Donepezil”.



## Section S2. A comment on the selection of starting materials based on the similarity to parent

One of the possible approaches to analog generation is to select starting materials based on the similarity to the parent's various substructures. This can be done either by simple substructure matching of commercially available molecules (up to some size threshold) against the parent itself or by first performing retrosynthesis of the parent and then identifying commercial chemicals that are the most similar to these building blocks. One of the problems with such approaches is that metrics of structural similarity are not necessarily indicative of functional similarity (see discussion in ref<sup>S1-4</sup>), and hence our "replicas"-based approach. A more mundane problem is that by simply searching for similar fragments, one may lose the reactive handles by which these fragments could later be connected (see example in **Figure S13** below). In approaches incorporating initial retrosynthesis (main-text ref. 26), one can overcome this problem – at least in part – by stipulating that the similarity search be confined to molecules retaining the necessary reactive groups. This however, does not take into account situations in which a very similar fragment may be available but with a chemically equivalent handle – say, Cl vs Br, which are formally different but, in many reactions, exhibit similar reactivity. This problem, is again, potentially surmountable by introducing tables of groups that are "exchangeable". We tried a similarity approach early on but were not satisfied with the chemical flexibility of the generated blocks whereas construction of comprehensive lists of exchangeable groups (to accommodate some 25,307 reaction types in Allchemy) proved a prohibitive undertaking.



**Figure S13. Substrates chosen by similarity to the target molecule (Donepezil, on the top).** Some substrates lack any reactive functional group that could be used as attachment point in forward synthesis, while others, despite containing reactive functionality and structural pattern

from the target molecule, do not have a right partner to react with. For example, benzyl bromide (left-most molecule on the top) would be an excellent substrate for alkylation of 4-piperidyl fragment, but all six piperidine-containing molecules in the top-row are already substituted with ethyl/propyl groups which, as a consequence, precludes such alkylation.

### Section S3. Addition of simple chemicals to the substrate set.

In this section, we compare the forward syntheses of analogs with and without the use of the set of 23 “auxiliary” reagents from **Figure S4** (i.e., for a given target, searches starting from the same set of retrosynthesis-generated substrates but with vs. without 23 auxiliaries also included). In this comparison, all searches were up to 3 synthetic generations, and the limit for molecular weight of the generated molecules was set to 550 g/mol for Donepezil and 400 g/mol for Ketoprofen.

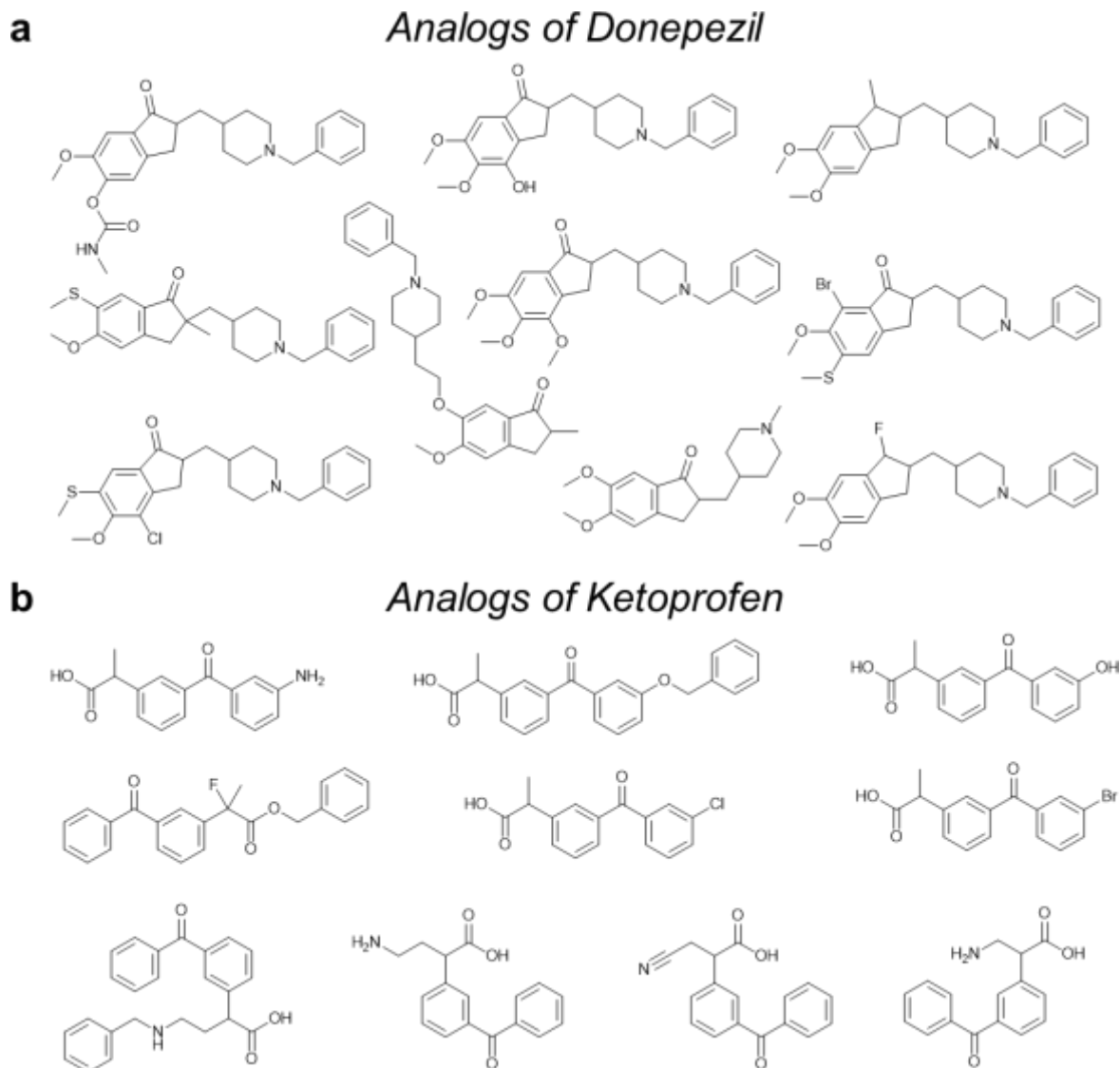
The results are summarized in **Table S1**. For both targets, inclusion of the 23 auxiliary reagents increased the number of generated products and analogs. For the total number of products up to G3, the increase was moderate (15% for Donepezil and 7% for Ketoprofen) but became more significant for analogs with target similarity  $\geq 0.7$  (27% and 28%) and for close analogs with target similarity  $\geq 0.85$  (33% and 35%). **Figure S14** shows examples of 10 close analogs for each target that are found only when the the auxiliary reagents are used.

**Table S1.** Comparison of forward calculations performed with and without the set of auxiliary reagents for Donepezil and Ketoprofen.

	<i>Donepezil</i>		<i>Ketoprofen</i>	
	Calculations <b>without</b> auxiliary reagents	Calculations <b>with</b> auxiliary reagents	Calculations <b>without</b> auxiliary reagents	Calculations <b>with</b> auxiliary reagents
<i>Number all of products</i>	3138	3619 (+15%)*	3459	3692 (+7%)*
<i>Number of analogs (similarity <math>\geq 0.7</math>)**</i>	1752	2222 (+27%)*	609	781 (+28%)*
<i>Number of close analogs (similarity <math>\geq 0.85</math>)**</i>	141	188 (+33%)*	69	93 (+35%)*

\* - numbers in the parentheses denotes percentage increase compared to calculations without auxiliary reagents

\*\* - similarity to target or one of target's “replicas”



**Figure S14.** Close analogs (similarity to target or one of target's "replicas"  $\geq 0.85$ ) of Donepezil and Ketoprofen found only in calculations starting from the substrate set augmented with the collection of 23 auxiliary reagents.

## Section S4. Molecular docking Studies

The crystal structures of COX-2 and AChE were extracted from the Protein Data Bank (<http://www.rcsb.org>, for COX-2 pdb code: 5IKR; the structure of mefenamic acid bound to human cyclooxygenase-2, resolution 2.34 Å, for AChE pdb code: *Electrophorus Electricus* acetylcholinesterase, resolution 4.20 Å). Each structure was cleaned of all water molecules and inhibitors as well as all non-interacting ions with Chimera software before being used in the docking studies. For COX-2, one of the two subunits was taken as the target structure. For AChE, one of the four subunits was taken as the target structure. Energy minimization of 3-dimensional structures of Ketoprofen and donepezil analogues were performed using Chem3D. Docking of 7 Ketoprofen analogues and Ketoprofen drug with COX-2 and 5 Donepezil analogues and donepezil drug with AChE was performed using three independent docking programs AutoDock 4.2.6, AutoDock Vina 1.2.5 and Dock 6 for comparison. The output files obtained from the docking study were visualized and analyzed by PyMol.

### S4.1. AutoDock 4.2.6

Polar hydrogens were added to the protein structures and merged with non-polar hydrogens for the accurate calculation of partial Gasteiger charges. AD4 atom type was assigned. For docking simulations employing the free energy function, hydrogens were added to the ligands, partial atomic charges were assigned, and torsions were defined. The grid box size was set at 55, 55, 70 Å (x, y, z) with center x = 43.334, y = 0.932, z = 59.875 and spacing 0.375 Å for COX-2 and 60, 60, 60 Å (x, y, z) with center x = 42.095, y = 66.809, z = -81.47 and spacing 0.375 Å for AChE. Docking was performed using the Lamarckian genetic algorithm. 50 independent runs per ligand using an initial population of 300 individuals with a mutation rate of 0.02 were evolved for 27000 generations. A maximum of 2.5 million energy evaluations was applied for each experiment. The results were clustered using a tolerance of 2.0 Å. The best ranked complexes of COX-2 with Ketoprofen analogs and AChE with Donepezil analogs were selected on the basis of binding free energy value.

### S4.2. Dock 6.12

First, protein and ligand for docking with Dock 6.12 were prepared using Chimera software DockPrep tool. Then, the molecular surface of the receptor, excluding hydrogen atoms and the ligand, was calculated and saved as DMS file. The Sphgen program was employed to generate spheres within the ligand binding site, which were then selected manually. Subsequently, the showbox program was used to create a box around the spheres, extending by an 8.0 Å margin in all directions. Program grid was used to precompute energy interactions between a dummy probe atom and all receptor atoms on a 0.3 Å resolution grid within the box. Grid creates the grid files necessary for rapid score evaluation. Finally, the docking was performed using the anchor-and-grow algorithm.

During docking, it was estimated that Donepezil analogs should form hydrogen bonds with Phe295, similar to crystallized Donepezil inside human acetylcholinesterase<sup>S5</sup>. Ketoprofen analogs were predicted to form hydrogen bonds with either Arg120 or Ser530, as is typical for most NSAID drugs.

### S4.3. AutoDock Vina 1.2.5

Docking studies were also performed using AutoDock Vina 1.2.5. The input files included receptor and ligand in pdbqt formats and docking box in txt format. Output was a list of poses ranked by predicted binding energy ( $-\Delta G$  in kcal/mol). Partial charges and polar hydrogens were added using AutoDock Tools and then exported as pdbqt files. The same size of the docking box was defined, 55, 55, 70 Å (x, y, z) with center x = 43.334, y = 0.932, z = 59.875 for COX-2 and 60, 60, 60 Å (x, y, z) with center x = 42.095, y = 66.809, z = -81.47 for AChE.

To obtain the maximum number of poses, we set the number of modes to 20 and the energy range to 4, exhaustiveness parameter to 100. AutoDock Vina samples the pose space based on a random seed value. Consequently, two runs with the same settings and structure files will typically produce different output poses.

**Table S2. Molecular modelling** of Ketoprofen, Donepezil and their analogs performed with AutoDock 4 software. Donepezil and its analogs were docked inside the active site of acetylcholinesterase from *Electrophorus electricus*, PDB code 1C2O. Ketoprofen and its analogs were docked inside the active site of human COX-2, PDB code 5IKR.

Compound	Binding affinity [kcal/mol]	Ki [μM]	Hydrogen bonds with aminoacids	Comment
( <i>S</i> )-Donepezil	-9.88	0.057	Tyr133, His447, Ser203	conformation 1 has inverted position inside the active site
( <i>R</i> )-Donepezil	-10.62	0.016	Phe295	-
(1 <i>R</i> ,2 <i>R</i> )- <b>10</b>	-9.11	0.209	Gly122	conformation 1 has inverted position inside the active site, conformation 20 binds to Phe295
(1 <i>S</i> ,2 <i>R</i> )- <b>10</b>	-9.86	0.059	Glu202, Ser125	-
(1 <i>R</i> ,2 <i>S</i> )- <b>10</b>	-9.76	0.07	Ser203, His447	-
(1 <i>S</i> ,2 <i>S</i> )- <b>10</b>	-9.98	0.048	Ser125, Glu202	-
( <i>S</i> )- <b>9</b>	-9.79	0.067	-	conformation 1 has inverted position inside the active site, conformation 16 binds to Phe295 and Arg296
( <i>R</i> )- <b>9</b>	-10.11	0.039	Ser125	conformation 6 binds to Phe295 and Tyr124
<b>11</b>	-9.77	0.069	Tyr337	-
( <i>R</i> )- <b>12</b>	-10.73	0.014	Tyr337	conformation 1 has inverted position inside the active site,

				conformation 2 binds to Phe295 and Tyr337
(S)-12	-11.12	0.007	Tyr133, Ser125	-
(1E,2R)-8	-10.02	0.045	Gly120, Gly120, Ser125, Tyr133	Conformation 7 binds to Phe295
(1E,2S)-8	-9.34	0.142	Tyr337, Tyr133, Tyr124	conformation 1 has inverted position inside the active site
(1Z,2R)-8	-10.75	0.013	-	-
(1Z,2S)-8	-10.03	0.044	Tyr133, Tyr124, Ser125	conformation 1 has inverted position inside the active site, conformation 15 binds to Phe295
(S)-Ketoprofen	-8.97	0.264	Arg120, Arg120, Tyr355	-
(R)-Ketoprofen	-8.89	0.304	Arg120, Arg120, Tyr355	-
<b>1</b>	-9.13	0.202	Arg120, Arg120, Tyr355	-
(S)-3	-9.15	0.197	Arg120, Arg120, Tyr355	-
(R)-3	-9.42	0.125	Ser530, Arg120, Arg120, Tyr355	-
(S)-4	-9.04	0.237	Arg120, Arg120, Arg120, Tyr355	-
(R)-4	-9.11	0.21	Arg120, Arg120, Arg120, Tyr355	-
(S)-2	-9.13	0.203	Ser530, Arg120, Arg120, Tyr355	-
(R)-2	-8.84	0.333	Arg120, Arg120, Tyr355	-
(2R,4R)-5	-8.95	0.277	Arg120, Arg120, Ser530, Tyr355	-
(2R,4S)-5	-8.61	0.492	Arg120, Arg120, Tyr355	-
(2S,4R)-5	-8.76	0.379	Arg120, Arg120, Tyr355	-
(2S,4S)-5	-9.02	0.244	Arg120, Arg120, Arg120, Ser530, Tyr355	-
(R)-6	-9.34	0.143	Arg120, Arg120, Tyr355	-
(S)-6	-9.26	0.164	Arg120, Arg120, Arg120, Tyr355, Ala527	-
(1E,2R)-7	-8.91	0.293	Arg120, Arg120, Tyr355, Ser530, Val523	-
(1E,2S)-7	-9.02	0.243	Arg120, Arg120, Tyr355, Ser530, Tyr385	-
(1Z,2R)-7	-9.8	0.066	Arg120, Arg120, Tyr355, Ser530	-
(1Z,2S)-7	-8.66	0.452	Arg120, Tyr355, Ser530	-

**Table S3. Molecular modelling** of Ketoprofen, Donepezil and their analogs performed with Dock 6 software. Donepezil and its analogs were docked inside the active site of acetylcholinesterase from *Electrophorus electricus*, PDB code 1C2O. Ketoprofen and its analogs were docked inside the active site of human COX-2, PDB code 5IKR.

Compound	Scoring	Hydrogen bonds with aminoacids	Comment
( <i>S</i> )-Donepezil	-43.4326	-	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
( <i>R</i> )-Donepezil	-44.1372	Ser293	conformation 57 binds to Phe295, Tyr124;
(1 <i>R</i> ,2 <i>R</i> )- <b>10</b>	-41.2404	-	conformation 13 binds to Phe295, Tyr337, Arg296;
(1 <i>S</i> ,2 <i>R</i> )- <b>10</b>	-42.304	Tyr124	-
(1 <i>R</i> ,2 <i>S</i> )- <b>10</b>	-42.8801	-	conformation 1 has inverted position inside the active site, conformation 45 binds to Phe295;
(1 <i>S</i> ,2 <i>S</i> )- <b>10</b>	-45.7982	Tyr337	no conformation binds to Phe295;
( <i>S</i> )- <b>9</b>	-44.3883	Ser125, Tyr124	conformation 1 has inverted position inside the active site, conformation 8 binds to Phe295;
( <i>R</i> )- <b>9</b>	-40.7053	-	no conformation binds to Phe295;
<b>11</b>	-42.5352	-	conformation 56 binds to Phe295;
( <i>R</i> )- <b>12</b>	-44.8583	Tyr337	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
( <i>S</i> )- <b>12</b>	-49.0946	Tyr124, Ser203	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
(1 <i>E</i> ,2 <i>R</i> )- <b>8</b>	-41.0654	Gln291	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
(1 <i>E</i> ,2 <i>S</i> )- <b>8</b>	-42.0311	Ser293, Ser293	no conformation with proper positioning binds to Phe295;
(1 <i>Z</i> ,2 <i>R</i> )- <b>8</b>	-41.9395	-	no conformation binds to Phe295;



(1 <i>Z</i> ,2 <i>S</i> )- <b>8</b>	-41.9907	Tyr124, Tyr341	conformation 1 has inverted position inside the active site, conformation 97 binds to Phe295;
( <i>S</i> )-Ketoprofen	-33.7019	-	conformation 4 binds to Arg120, Arg120, Ser530;
( <i>R</i> )-Ketoprofen	-32.3618	Arg120, Glu524	conformation 1 is docked outside the active site, conformation 8 binds to Arg120, Arg120, Tyr355, Ser530;
<b>1</b>	-34.8531	Ser530	-
( <i>S</i> )- <b>3</b>	-33.5769	-	conformation 1 is docked outside the active site, conformation 40 binds to Ser530;
( <i>R</i> )- <b>3</b>	-33.2132	-	conformation 1 is docked outside the active site, conformation 48 binds to Ser530, Tyr385;
( <i>S</i> )- <b>4</b>	-34.4899	-	-
( <i>R</i> )- <b>4</b>	-35.1559	Arg120, Phe470	conformation 1 is docked outside the active site, conformation 66 is docked inside but doesn't form interactions;
( <i>S</i> )- <b>2</b>	-35.2979	Arg120	conformation 1 is docked outside the active site, conformation 73 binds to Val523;
( <i>R</i> )- <b>2</b>	-33.9441	-	conformation 1 is docked outside the active site, conformation 34 is docked inside but doesn't form interactions;
(2 <i>R</i> ,4 <i>R</i> )- <b>5</b>	-34.621	Arg120, Arg120, Arg120	conformation 1 is docked outside the active site, conformation 52 binds to Ser530;
(2 <i>R</i> ,4 <i>S</i> )- <b>5</b>	-37.3397	Lys83	conformation 1 is docked outside the active site, conformation 60 binds to Ser530;
(2 <i>S</i> ,4 <i>R</i> )- <b>5</b>	-38.2374	Pro84, Glu524	conformation 1 is docked outside the active site, conformation 56 is docked inside but doesn't form interactions;
(2 <i>S</i> ,4 <i>S</i> )- <b>5</b>	-33.4627	Arg120, Arg120, Arg120, Glu524	conformation 1 is docked outside the active site, conformation 67 binds to Arg120, Arg120, Tyr355;

(R)-6	-29.1311	Arg120, Tyr355, Tyr115	conformation 1 is docked outside the active site, conformation 24 binds to Ser530, Tyr385;
(S)-6	-37.2414	-	Conformation 1 is docked outside the active site, conformation 62 binds to Arg120, Tyr355;
(1E,2R)-7	-41.2565	-	conformation 1 is docked outside the active site, conformation 62 binds to Arg120, Arg120, Tyr355;
(1E,2S)-7	-40.2117	Arg120, Arg120	conformation 1 is docked outside the active site, conformation 58 binds to Arg120, Arg120, Tyr355, Ser530, Tyr385;
(1Z,2R)-7	-35.3165	Arg120, Arg120, Tyr115	conformation 1 is docked outside the active site, conformation 58 binds to Ser530, Tyr385;
(1Z,2S)-7	-39.8196	Arg120, Arg120, Lys83	conformation 1 is docked outside the active site, conformation 60 binds to Arg120, Try355, Ser530, Tyr385;

**Table S4. Molecular modelling** of Ketoprofen, Donepezil and their analogs performed with AutoDock Vina software. Donepezil and its analogs were docked inside the active site of acetylcholinesterase from *Electrophorus electricus*, PDB code 1C2O. Ketoprofen and its analogs were docked inside the active site of human COX-2, PDB code 5IKR.

Compound	Binding affinity [kcal/mol]	Ki [ $\mu$ M]	Hydrogen bonds with aminoacids	Comment
(S)-Donepezil	-10	0.046	Ser203	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
(R)-Donepezil	-10	0.046	Ser203	conformation 1 has inverted position inside the active site, conformation 4 binds to Phe295, Tyr124;

<b>(1R,2R)-10</b>	-8.6	0.488	-	conformation 1 doesn't form hydrogen bonds with protein, no conformation with correct positioning binds to Phe295;
<b>(1S,2R)-10</b>	-9	0.249	Ser203, Gly122	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
<b>(1R,2S)-10</b>	-10.4	0.023	Phe295, Phe295, Arg296	-
<b>(1S,2S)-10</b>	-9.9	0.054	Ser203	conformation 1 has inverted position inside the active site, no conformation with correct positioning binds to Phe295;
<b>(S)-9</b>	-9.8	0.064	Phe295	-
<b>(R)-9</b>	-10.9	0.01	-	conformation 1 doesn't form hydrogen bonds with protein, conformation 2 binds to Phe295, Arg296, Tyr124;
<b>11</b>	-9.8	0.064	-	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
<b>(R)-12</b>	-10.7	0.014	Phe295	-
<b>(S)-12</b>	-10.1	0.039	Tyr124, Ser125, Ser203	conformation 1 has inverted position inside the active site, no conformation with correct positioning binds to Phe295;
<b>(1E,2R)-8</b>	-9.9	0.054	Arg296, Ser293, Ser293	no conformation binds to Phe295;
<b>(1E,2S)-8</b>	-9	0.249	Tyr124	no conformation binds to Phe295;
<b>(1Z,2R)-8</b>	-10.6	0.017	Tyr124	no conformation with correct positioning binds to Phe295;

(1Z,2S)-8	-10.4	0.023	Ser203, Tyr341, Tyr124 inv	conformation 1 has inverted position inside the active site, no conformation binds to Phe295;
(S)-Ketoprofen	-8.7	0.413	Ser530	-
(R)-Ketoprofen	-8.3	0.812	Arg44, Lys137, Tyr130	conformation 2 binds to Ser530;
<b>1</b>	-8.7	0.413	Arg120, Tyr355	conformation 4 binds to Ser530;
(S)-3	-8.5	0.579	Arg44, His39, Gln461	conformation 1 is docked outside the active site, conformation 10 is docked inside the active site and binds to Met522;
(R)-3	-8.5	0.579	His39, Gln461, Phe381, Arg44	conformation 1 is docked outside the active site, no conformation binds to Arg120 or Ser530;
(S)-4	-8.1	1.138	His39, Cys41, Cys37	conformation 1 is docked outside the active site, conformation 9 binds to Ser530;
(R)-4	-8.1	1.138	Gln461, Arg44, His39	conformation 1 is docked outside the active site, conformation 11 is docked inside but doesn't form hydrogen bonds;
(S)-2	-8.6	0.489	His39, Cys41	conformation 1 is docked outside the active site, conformation 2 binds to Arg120, Arg120, Tyr355;
(R)-2	-8.1	1.138	Arg44, Cys41, Cys41	conformation 1 is docked outside the active site, no conformation is docked inside;
(2R,4R)-5	-8.5	0.579	His39, Cys41, Phe367	conformation 1 is docked outside the active site, conformation 10 binds to Arg120, Arg120;
(2R,4S)-5	-8.5	0.579	His39	conformation 1 is docked outside the active site,

(2 <i>S</i> ,4 <i>R</i> )-5	-8.5	0.579	Arg44, Arg44, Cys47, Gly45	conformation 7 binds to Arg120, Arg120, Tyr355; conformation 1 is docked outside the active site; conformation 11 binds to Arg120, Tyr355;
(2 <i>S</i> ,4 <i>S</i> )-5	-8.7	0.413	Arg120, Arg120, Ser530, Tyr355	-
( <i>R</i> )-6	-8.6	0.489	Arg469, Arg44, Arg44, Tyr130, Gly45	conformation 1 is docked outside the active site, conformation 12 binds to Ser530;
( <i>S</i> )-6	-8.7	0.413	Asn43, His39, Arg44	conformation 1 is docked outside the active site, conformation 16 binds to Tyr385, Tyr385;
(1 <i>E</i> .2 <i>R</i> )-7	-8.2	0.961	His39, Gln461	conformation 1 is docked outside the active site, conformation 8 binds to Arg120, Arg120, Tyr355, Ser530, Ser530;
(1 <i>E</i> ,2 <i>S</i> )-7	-8.5	0.579	Arg120, Tyr355, Ser530, Tyr385	-
(1 <i>Z</i> ,2 <i>R</i> )-7	-8.7	0.413	Cys47, His39, Cys36	conformation 1 is docked outside the active site, no conformation binds inside;
(1 <i>Z</i> ,2 <i>S</i> )-7	-8.8	0.349	Gln42, Lys468, Glu465, Arg44, Cys47	conformation 1 is docked outside the active site, no conformation binds inside;

---

## Section S5. Procedures of the biological analysis

### S5.1. Cyclooxygenase-2 (COX-2) inhibition assay

The synthesized compounds were assessed for their COX-2 inhibitory potential using an adapted literature procedures<sup>S6,7</sup>. This procedure was carried out with a fluorescence-based assay by measuring the effects of the compounds on the peroxidase activity of COX-2 enzyme using 10-acetyl-3,7-dihydroxyphenoxazine (Amplex Red = Ampliflu) as a fluorescent probe. Under the action of the COX-2 enzyme, arachidonic acid is converted to prostaglandin G2 (PG2). The fluorescence produced by the probe is directly proportional to the PG2 formation.

Recombinant human COX-2 enzyme stock solution was prepared by diluting 5  $\mu\text{L}$  of  $\geq 8000$  units/mg protein solution supplied from Sigma–Aldrich with 35  $\mu\text{L}$  of buffer solution of pH 8.0. It was then divided into 4 batches of 10  $\mu\text{L}$  each, and kept at  $-78\text{ }^\circ\text{C}$ . At the time of measurement, one batch at a time was used and stored between  $-50\text{ }^\circ\text{C}$  and  $-20\text{ }^\circ\text{C}$ . 1 mM stock solution of Hemin was prepared by dissolving Hemin in anhydrous DMSO. 2 mM solution of test compounds in DMSO was diluted in buffer solution of pH 8.0, to prepare appropriate working solutions. Arachidonic acid (AA) working solution was prepared by mixing 100  $\mu\text{L}$  of 10 mM AA solution in ethanol, 100  $\mu\text{L}$  solution of 0.1M KOH and 800  $\mu\text{L}$  of ddH<sub>2</sub>O.

COX-2 enzyme (1  $\mu\text{L}$  from stock solution, 2.5U per cuvette), hemin cofactor (1  $\mu\text{L}$  from 1 mM solution in DMSO, final concentration: 10  $\mu\text{M}$ ), test compounds or the standard drugs (10  $\mu\text{L}$ , final concentrations ranged from 0.1  $\mu\text{M}$  to 100  $\mu\text{M}$ ) were incubated in reaction buffer (100 mM Tris-HCl-EDTA buffer, 77  $\mu\text{L}$ , pH 8.0) in a thermoshaker for a period of 15 min at room temperature and at 300 rpm. The reaction was initiated by addition of Amplex red reagent (Sigma-Aldrich, 1  $\mu\text{L}$  from 10 mM solution in DMSO, final concentration: 100  $\mu\text{M}$ ), followed by arachidonic acid/NaOH solution (10  $\mu\text{L}$  from 1 mM AA/NaOH stock solution). The assay was performed in a final volume of 100  $\mu\text{L}$ . After 2 min of incubation in the thermoshaker at room temperature and 300 rpm, the assay was immediately quantified (F6000, Shimadzu) based on the generated fluorescence of resorufin ( $\lambda_{\text{ex}} = 535\text{ nm}$ ,  $\lambda_{\text{em}} = 601\text{ nm}$ ). Any delay in the measurement would lead to inaccurate and asynchronous results. The measured fluorescence for the final results was calculated after subtraction of background activity. The mean fluorescence values were calculated to determine the percentage of residual activity achieved by treatment of the enzyme with each compound. Enzyme control (10  $\mu\text{L}$  of buffer was added instead of an inhibitor) serves as 100% activity. IC<sub>50</sub> values were determined from dose-response curves. Celecoxib and Ketoprofen served as standard drugs for comparison. Each experiment was conducted in triplicate.

Volume	Solution
1 $\mu$ L	COX-2 from stock solution, final conc. 2.5U/100
1 $\mu$ L	Hemin from 1 mM stock solution, final conc. 10 $\mu$ M
77 $\mu$ L	100 mM Tris-HCl-EDTA, pH 8.0
10 $\mu$ L	Inhibitor or drug from working solutions, final conc. ranged from 0.1 $\mu$ M to 100 $\mu$ M
1 $\mu$ L	Ampliflu from 10 mM solution in DMSO, final conc. 100 $\mu$ M
10 $\mu$ L	arachidonic acid/NaOH solution from 1 mM stock solution

## S5.2. Acetylcholinesterase (AChE) inhibition assay

The synthesized compounds were assessed for their AChE inhibitory potential using a modified Ellman's method<sup>S6</sup>. This protocol involves a two-step reaction, starting with the hydrolysis of acetylthiocholine iodide, catalyzed by the AChE enzyme, to produce thiocholine and acetic acid. In the second step, resultant thiocholine reacts with Ellman's reagent (5,5-dithiobis(2-nitro) benzoic acid, DTNB) generating TNB (5-thio-2-nitrobenzoic acid). TNB can be quantitatively measured at  $\lambda = 412$  nm.

100 U/mL enzyme stock solution was prepared by dissolving 292.5 U/mg protein supplied from Sigma-Aldrich (AChE, E.C. 3.1.1.7, as lyophilized powder) with ddH<sub>2</sub>O. 5 U/mL of AChE working solution was prepared directly before performing experiments by diluting stock solution with ddH<sub>2</sub>O. 2 mM solution of test compounds in DMSO was diluted in ddH<sub>2</sub>O to prepare appropriate working solutions. 2 mM stock solutions of potential inhibitors were prepared in dimethyl sulfoxide (DMSO). DdH<sub>2</sub>O was used to prepare different dilutions of inhibitors to obtain less than 2% (v/v) DMSO to avoid false positive results.

Each reaction was initiated by mixing sodium phosphate buffer (384  $\mu$ L, 50 mM, pH 7.4), the test compounds or the standard drugs (6  $\mu$ L final concentrations ranged from 0.1  $\mu$ M to 100  $\mu$ M), acetylcholinesterase enzyme from electric eel origin (10  $\mu$ L from 5U/mL to get final concentration of 0.05U per cuvette). The working solution was then incubated at room temperature for 15 min, after which 100  $\mu$ L of 2 mM acetylthiocholine iodide in sodium phosphate buffer and 100  $\mu$ L of 0.66 mM DTNB (5,5-dithiobis(2-nitro) benzoic acid) in sodium phosphate buffer were added. Reaction was incubated for 5 min at rt. Absorption corresponding to the formed chromophore was detected at 412 nm using a Shimadzu UV-1900 spectrophotometer. The assay was performed in a final volume of 600  $\mu$ L. A background reaction was also performed under the same conditions but without the inclusion of enzyme. In this assay, the blank consisted of sodium phosphate buffer. The mean absorbance values were calculated to determine the percentage of residual activity achieved by treatment of the enzyme with each compound. Enzyme control (6  $\mu$ L of buffer was added instead of an inhibitor) serves as 100% activity. IC<sub>50</sub> values were determined from dose-response curves. Donepezil served as standard drug for comparison. Each experiment was conducted in triplicate.

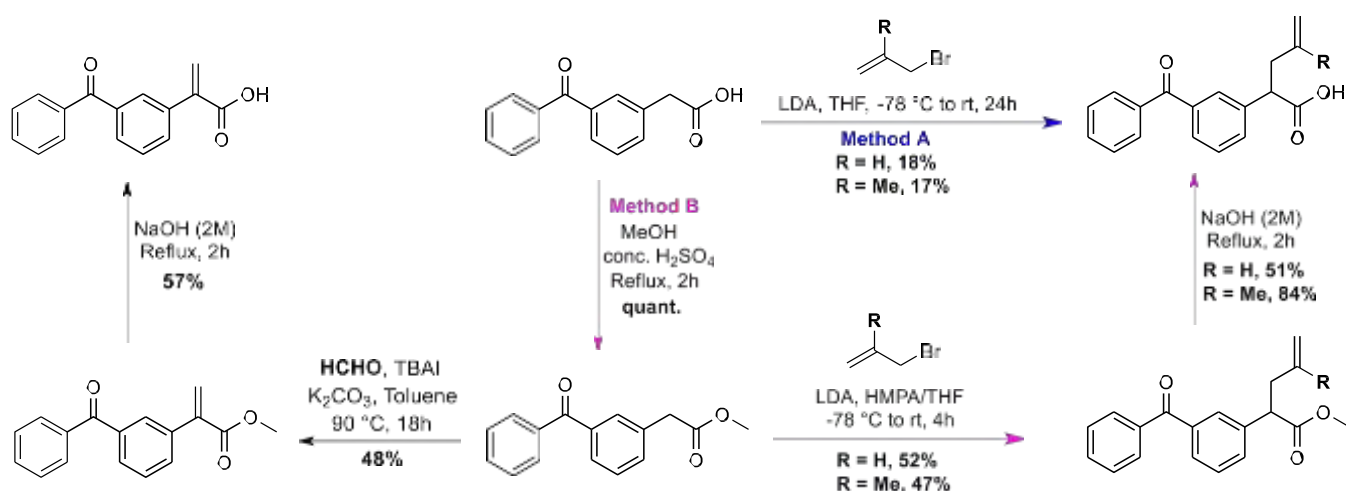
Volume	Solution
384 $\mu\text{L}$	sodium phosphate buffer (50 mM, pH 7.4)
6 $\mu\text{L}$	Inhibitor or drug from working solutions, final conc. ranged from 0.1 $\mu\text{M}$ to 100 $\mu\text{M}$
10 $\mu\text{L}$	AChE from 5U/mL working solution to get final concentration of 0.05U per cuvette
100 $\mu\text{L}$	acetylthiocholine iodide from 2 mM stock solution in sodium phosphate buffer
100 $\mu\text{L}$	DTNB from 0.66 mM stock solution in sodium phosphate buffer



## Section S6. Synthetic details

**General information.** All starting materials and reagents were purchased from Aldrich, Fisher, Alfa Aesar, TCI, or Ambeed and used without purification. All solvents used were freshly distilled prior to use.  $^1\text{H}$  NMR spectra were recorded at 400, 500 or 600 MHz, and  $^{13}\text{C}$  NMR spectra were recorded at 100, or 150 MHz with complete proton decoupling. Chemical shifts are given in  $\delta$  relative to the residual signals of the deuterated solvents. High-resolution mass spectra were acquired using electron ionization (EI) or electrospray ionization (ESI) mode with a time-of-flight detector. Infrared (IR) spectra were recorded on a Fourier transform infrared (FT-IR) spectrometer as a thin film on a NaCl plate (film). Elemental analysis was performed in the automatic analyzer UNiCube by Elementar, at a temperature of 1150 °C, and by the analysis of the percentage of C, H, N, S atoms. TLC was performed with aluminium plates coated with 60 F254 silica gel. Plates were visualized with UV light (254 nm) and by treatment with ethanolic *p*-anisaldehyde with sulfuric and glacial acetic acid followed by heating, aqueous cerium(IV) sulfate solution with molybdic and sulfuric acid followed by heating, or aqueous potassium permanganate with sodium hydroxide and potassium carbonate solution followed by heating or ethanolic vanillin with sulfuric acid followed by heating. Reaction products were purified by flash chromatography using silica gel 60 (230-400 mesh).

### Section S6.1. Ketoprofen's analogs described in main-text Figure 3.a



**Figure S15.** Reactions of compounds S6.1.1 to S6.1.7 described in main-text Figure 3.a.

#### General procedure S6.I for Method A

LDA (2M in THF, 2.5 equiv.) was added to 1 mL of anhydrous THF at  $-78\text{ }^\circ\text{C}$  under Ar atmosphere. A solution of 2-(3-benzoylphenyl)acetic acid (1 equiv.) was then added to the mixture. The reaction mixture was then slowly allowed to warm up to  $0\text{ }^\circ\text{C}$ , and stirred at that temperature for 1 h. Afterwards, the reaction mixture was again cooled down to  $-78\text{ }^\circ\text{C}$ , and the corresponding allyl bromide (4 equiv.) was added. The reaction mixture was allowed to warm up to room temperature and stirred for 24 h. It was then quenched with water (5 mL) followed by 1N HCl (5 mL), then extracted with ethyl acetate (3x10 mL). The combined organic layers were collected, washed with brine, dried over  $\text{MgSO}_4$ , and concentrated under

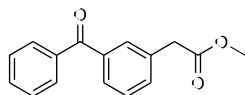
reduced pressure to obtain the crude product, which was then purified by column chromatography with hexane/EtOAc/AcOH (84/15/1) mixture, to afford the final compound.

### General procedure S6.II for Method B

To a solution of 2-(3-benzoylphenyl)acetic acid (1 equiv.) in MeOH (0.5 mL), was added 2 drops of concentrated H<sub>2</sub>SO<sub>4</sub> at room temperature, and then was refluxed for 2 h. Upon completion of the reaction (monitored by TLC), the volatiles were removed under reduced pressure. The resultant residue was diluted with EtOAc (10 mL) and then neutralized with aqueous NaHCO<sub>3</sub> solution. The mixture was extracted with EtOAc (3x10 mL). The combined organic phase was collected, washed with brine, dried over MgSO<sub>4</sub>, and concentrated under reduced pressure to afford methyl 2-(3-benzoylphenyl)acetate as yellow liquid (60 mg, quant.), which was used for the next step without further purification.

LDA (2M in THF, 1.2 equiv.) was added to 2 mL of anhydrous THF at -78 °C under argon atmosphere. A solution of methyl 2-(3-benzoylphenyl)acetate (1 equiv.) in dry THF (2 mL) was added to it dropwise, and the reaction mixture was then slowly allowed to warm up to 0 °C and stirred at that temperature for 1 h. Afterwards, the reaction mixture was again cooled back to -78 °C before HMPA (2 equiv.) and the corresponding allyl bromide (2 equiv.) were added. The reaction mixture was then allowed to warm up to room temperature and stirred for extra 4 h. It was then quenched with saturated NH<sub>4</sub>Cl (50 mL) and extracted with EtOAc (3x50 mL). The combined organic phases were dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated. The residue was purified by column chromatography on silica gel, using hexane/EtOAc (90/10), to afford the desired product.

The latter obtained ester (1 equiv.) was added to an aqueous solution of 2N NaOH (0.5 mL) and the reaction mixture was refluxed for 2 h. After cooling down to room temperature, water (10 mL) was added and the resulting mixture was extracted with diethyl ether (10 mL). The aqueous layer was then acidified with 3N aqueous HCl solution (pH<1, checked by litmus paper), and extracted with ethyl acetate (3x20 mL). The combined extracts were dried over MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure to give the crude mixture. It was then purified by column chromatography, with hexane/EtOAc/AcOH (84/15/1), to afford the desired final product.



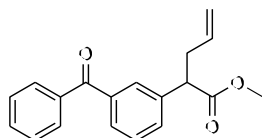
### Methyl 2-(3-benzoylphenyl)acetate S6.1.1

Following the general procedure S6.II for method B, using 2-(3-benzoylphenyl)acetic acid (60.06 mg, 0.25 mmol, 1 equiv.) in MeOH (0.5 mL), to afford the desired product methyl 2-(3-benzoylphenyl)acetate S6.1.1, as yellow liquid (62 mg, quant.).

The spectral data match those reported in the literature<sup>S8</sup>;

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.82-7.79 (m, 2 H), 7.73-7.68 (m, 2 H), 7.61-7.57 (m, 1 H), 7.53-7.42 (m, 4 H), 3.71 (s, 3 H), 3.70 (s, 3 H);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  196.3, 171.4, 137.8, 137.4, 134.2, 133.2, 132.4, 130.8, 129.9, 128.8, 128.4, 128.2, 52.0, 40.8.



### Methyl 2-(3-benzoylphenyl)pent-4-enoate S6.1.2

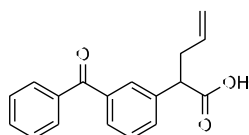
Following the general procedure S6.II for method B, using LDA (2M in THF, 1.2 mL, 2.359 mmol, 1.2 equiv.) in THF (2 mL), methyl 2-(3-benzoylphenyl)acetate (500 mg, 1.966 mmol, 1 equiv.) in THF (2 mL), HMPA (0.68 mL, 3.932 mmol, 2 equiv.) and neat allyl bromide (0.34 mL, 3.932 mmol, 2 equiv.), to afford desired product, methyl 2-(3-benzoylphenyl)pent-4-enoate S6.1.2, as colorless liquid (301 mg, 52%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.80-7.78 (m, 2 H), 7.74-7.73 (m, 1 H), 7.70-7.67 (m, 1 H), 7.61-7.54 (m, 2 H), 7.50-7.42 (m, 3 H), 5.76-5.66 (m, 1 H), 5.09-5.00 (m, 2 H), 3.73 (t,  $J$  = 7.76 Hz, 1 H), 3.67 (s, 3 H), 2.88-2.81 (m, 1 H), 2.58-2.51 (m, 1 H);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  196.4, 173.4, 138.8, 137.9, 137.5, 134.8, 132.5, 131.9, 130.1, 129.7, 129.2, 128.6, 128.3, 117.4, 52.1, 51.2, 37.4;

HRMS (ESI)  $m/z$ :  $[\text{M}]^+$  Calc. for  $\text{C}_{19}\text{H}_{18}\text{O}_3\text{Na}$ : 317.1154, found: 317.1159;

IR (film,  $\text{CDCl}_3$ ): 3063, 2951, 1737, 1669, 1598, 1438, 1317, 1280, 1229, 1198, 1164, 919, 718  $\text{cm}^{-1}$ .



### Methyl 2-(3-benzoylphenyl)pent-4-enoate S6.1.3

Following the general procedure S6.I for method A, using LDA (2M in THF, 0.78 mL, 1.560 mmol, 2.5 equiv.) in THF (1 mL), 2-(3-benzoylphenyl)acetic acid (150 mg, 0.624 mmol, 1 equiv.) in THF (1 mL) and neat allyl bromide (0.22 mL, 2.497 mmol, 4 equiv.), to afford 2-(3-benzoylphenyl)pent-4-enoic acid S6.1.3, as colorless oil (31.5 mg, 18% yield).

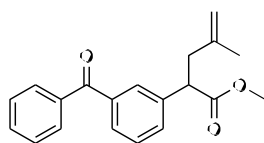
Following the general procedure S6.II for method B, using methyl 2-(3-benzoylphenyl)pent-4-enoate (25 mg, 0.0849 mmol, 1 equiv.) and 2N NaOH (0.5 mL), to afford 2-(3-benzoylphenyl)pent-4-enoic acid S6.1.3, as colorless oil (12.1 mg, 51% yield).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.79-7.76 (m, 3 H), 7.71-7.69 (m, 1 H), 7.60-7.55 (m, 2 H), 7.49-7.43 (m, 3 H), 5.77-5.67 (m, 1 H), 5.11-5.02 (m, 2 H), 3.74 (t,  $J$  = 7.69 Hz, 1 H), 2.87-2.82 (m, 1 H), 2.60-2.53 (m, 1 H);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  196.3, 178.0, 138.2, 138.0, 137.4, 134.4, 132.5, 132.0, 130.1, 129.8, 129.4, 128.7, 128.3, 117.7, 51.0, 37.1;

HRMS (ESI)  $m/z$ :  $[\text{M}]^+$  Calc. for  $\text{C}_{18}\text{H}_{16}\text{O}_3\text{Na}$ : 303.0997, found: 303.0996;

IR (film, CDCl<sub>3</sub>): 3067, 2921, 1734, 1708, 1658, 1597, 1318, 1281, 1178, 919, 720 cm<sup>-1</sup>.



#### Methyl 2-(3-benzoylphenyl)-4-methylpent-4-enoate S6.1.4

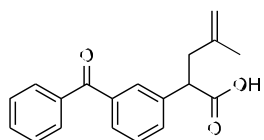
Following the general procedure S6.II for method B, using LDA (2M in THF, 0.3 mL, 0.594 mmol, 1.2 equiv.) in THF (2 mL), methyl 2-(3-benzoylphenyl)acetate (126 mg, 0.495 mmol, 1 equiv.) in THF (2 mL), HMPA (0.17 mL, 0.99 mmol, 2 equiv.), and neat 3-bromo-2-methylprop-1-ene (0.1 mL, 0.99 mmol, 2 equiv.) to afford methyl 2-(3-benzoylphenyl)-4-methylpent-4-enoate S6.1.4, as colorless liquid (72 mg, 47%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.80-7.75 (m, 3 H), 7.69-7.67 (m, 1 H), 7.61-7.56 (m, 2 H), 7.50-7.41 (m, 3 H), 4.76 (s, 1 H), 4.68 (s, 1 H), 3.91-3.87 (m, 1 H), 3.66 (s, 3 H), 2.85 (dd, *J* = 14.6, 8.7 Hz, 1 H), 2.48 (dd, *J* = 14.6, 6.9 Hz, 1 H), 1.72 (s, 3 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.4, 173.6, 142.1, 139.0, 137.9, 137.5, 132.5, 131.8, 130.1, 129.6, 129.1, 128.5, 128.3, 112.6, 52.1, 49.8, 41.2, 22.5;

HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>20</sub>H<sub>20</sub>O<sub>3</sub>Na: 331.1310, found: 331.1316;

IR (film, CDCl<sub>3</sub>): 3073, 2951, 1737, 1659, 1598, 1579, 1445, 1316, 1281, 1194, 1159, 718 cm<sup>-1</sup>.



#### 2-(3-benzoylphenyl)-4-methylpent-4-enoic acid S6.1.5

Following the general procedure S6.I for method A, using LDA (2M in THF, 0.52 mL, 1.04 mmol, 2.5 equiv.) in THF (1 mL), 2-(3-benzoylphenyl)acetic acid (100 mg, 0.416 mmol, 1 equiv.) in THF (1 mL) and 3-bromo-2-methylprop-1-ene (0.17 mL, 1.664 mmol, 4 equiv.), to afford 2-(3-benzoylphenyl)-4-methylpent-4-enoic acid S6.1.5, as colorless oil (21 mg, 17% yield).

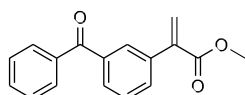
Following the general procedure S6.II for method B, using methyl 2-(3-benzoylphenyl)-4-methylpent-4-enoate (25 mg, 0.081 mmol, 1 equiv.) and 2N NaOH (0.5 mL), to afford 2-(3-benzoylphenyl)-4-methylpent-4-enoic acid S6.1.5, as white sticky solid (20 mg, 84% yield).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.79-7.77 (m, 3 H), 7.70-7.68 (m, 1 H), 7.60-7.56 (m, 2 H), 7.48-7.42 (m, 3 H), 4.77 (s, 1 H), 4.70 (s, 1 H), 3.88 (t, *J* = 7.79 Hz, 1 H), 2.87-2.81 (m, 1 H), 2.52-2.47 (m, 1 H), 1.71 (s, 3 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.4, 178.6, 141.7, 138.3, 137.9, 137.4, 132.5, 132.0, 130.1, 129.8, 129.4, 128.6, 128.3, 112.9, 49.6, 40.8, 22.5;

HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>19</sub>H<sub>18</sub>O<sub>3</sub>Na: 317.1154, found: 317.1155;

IR (film, CDCl<sub>3</sub>): 3074, 2967, 2926, 1735, 1708, 1658, 1597, 1446, 1282, 1178, 719 cm<sup>-1</sup>.



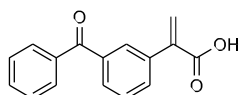
### Methyl 2-(3-benzoylphenyl)acrylate S6.1.6

To a solution of methyl 2-(3-benzoylphenyl)acetate (60 mg, 0.236 mmol, 1 equiv.) in toluene (1 mL), were added paraformaldehyde (21.26 mg, 0.708 mmol, 3 equiv.), tetrabutylammonium iodide (3.50 mg, 0.0094 mmol, 4 mol%) and K<sub>2</sub>CO<sub>3</sub> (97.8 mg, 0.708 mmol, 3 equiv.) at room temperature. The resulting mixture was stirred at 90 °C for 18 h. After completion of the reaction, water (20 mL) was added and the aqueous phase was extracted with EtOAc (2x20 mL). The combine organic phases were washed with brine, dried over MgSO<sub>4</sub>, and concentrated under reduced pressure. The obtained crude mixture was purified by column chromatography, using hexane/EtOAc (90/10) to afford methyl 2-(3-benzoylphenyl)acrylate **S6.1.6**, as yellow liquid (30 mg, 48% yield).

The spectral data match those reported in the literature<sup>S9</sup>;

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.85-7.81 (m, 3 H), 7.78-7.74 (m, 1 H), 7.65-7.63 (m, 1 H), 7.61-7.57 (m, 1 H), 7.50-7.45 (m, 3 H), 6.43 (d, *J* = 0.9 Hz, 1 H), 5.95 (d, *J* = 0.9 Hz, 1 H), 3.82 (s, 3 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.3, 166.7, 140.4, 137.5, 137.4, 136.9, 132.5, 132.3, 130.1, 129.9, 129.8, 128.3, 128.1, 128.0, 52.3.



### 2-(3-benzoylphenyl)acrylic acid S6.1.7

Following the hydrolysis step in the general procedure S6.II for method B, using Methyl 2-(3-benzoylphenyl)acrylate (26.6 mg, 0.0998 mmol, 1 equiv.) and 2N NaOH (0.5 mL), to afford 2-(3-benzoylphenyl)acrylic acid **S6.1.7** as light yellow sticky oil (12 mg, 47% yield).

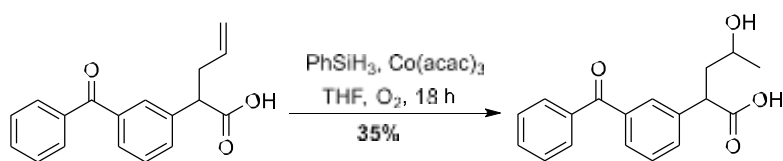
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.89-7.88 (m, 1 H), 7.84-7.81 (m, 2 H), 7.80-7.77 (m, 1 H), 7.69-7.66 (m, 1 H), 7.61-7.57 (m, 1 H), 7.51-7.47 (m, 3 H), 6.61 (s, 1 H), 6.10 (s, 1 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.4, 170.3, 139.6, 137.6, 137.4, 136.3, 132.6, 132.4, 130.4, 130.13, 130.10, 130.0, 128.3, 128.1.;

HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>16</sub>H<sub>12</sub>O<sub>3</sub>H: 253.0865, found: 253.0869;

IR (film, CDCl<sub>3</sub>): 2923, 2852, 1698, 1659, 1597, 1447, 1318, 1277, 1216, 707 cm<sup>-1</sup>.

### Synthesis of 2-(3-benzoylphenyl)-4-hydroxypentanoic acid S6.1.8



#### Scheme S1. Reaction scheme of compound S6.1.8 described in main-text Figure 3.a.

Under inert atmosphere, and to a solution of 2-(3-benzoylphenyl)pent-4-enoic acid S6.1.3 (43.0 mg, 0.153 mmol, 1 equiv.) in anhydrous THF (1 mL), was added phenylsilane (33.2 mg, 0.306 mmol, 2 equiv.), and  $\text{Co}(\text{acac})_3$  (2.71 mg, 0.0076 mmol, 5 mol%). An oxygen balloon was connected to the reaction system, and it was stirred for 24 h at room temperature. EtOAc (10 mL) and water (10 mL) were then added to the reaction mixture. The aqueous layer was extracted with EtOAc (3x20 mL), and the combined organic layers were collected, washed with brine, dried over anhydrous  $\text{MgSO}_4$ , and concentrated under reduced pressure to obtain the crude product, which was then purified by column chromatography, with hexane/EtOAc/AcOH (87/12/1) mixture, to afford 2-(3-benzoylphenyl)-4-hydroxypentanoic acid S6.1.8, as colorless liquid (16 mg, 35% yield).

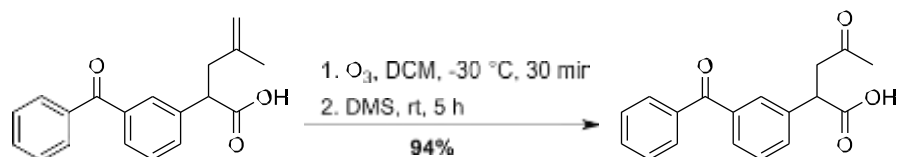
$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.82-7.78 (m, 4 H), 7.74-7.69 (m, 4 H), 7.62-7.53 (m, 4 H), 7.51-7.46 (m, 6 H), 4.87-4.79 (m, 1 H), 4.69-4.60 (m, m, 1 H), 4.04-3.95 (m, 2 H), 2.85-2.78 (m, 1 H), 2.63-2.55 (m, 1 H), 2.32-2.36 (m, 1 H), 2.11-2.02 (m, 1 H), 1.51 (d,  $J = 6.1$  Hz, 3 H), 1.47 (d,  $J = 6.3$  Hz, 3 H);

$^{13}\text{C NMR}$  (100 MHz,  $\text{CDCl}_3$ )  $\delta$  196.3, 196.2, 176.6, 176.3, 138.3, 138.1, 137.4, 137.3, 136.9, 132.6, 132.6, 132.1, 131.7, 130.1, 130.1, 129.6, 129.5, 129.4, 129.1, 128.9, 128.8, 128.4, 128.4, 75.1, 75.1, 47.5, 45.3, 39.5, 37.6, 21.0, 20.7;

**Elemental analysis:** Calc. for  $\text{C}_{18}\text{H}_{18}\text{O}_4$ : C, 72.47; H, 6.08; O, 21.45. Found: C, 73.38; H, 5.79; O, 20.83;

**IR** (film,  $\text{CDCl}_3$ ): 3060, 2977, 2927, 1769, 1658, 1283, 1716, 1135, 1056, 718  $\text{cm}^{-1}$ .

### Synthesis of 2-(3-benzoylphenyl)-4-oxopentanoic acid S6.1.9



#### Scheme S2. Reaction scheme of compound S6.1.9 described in main-text Figure 3.a.

Under inert atmosphere, the compound 2-(3-benzoylphenyl)-4-methylpent-4-enoic acid S6.1.5 (15 mg, 0.051 mmol, 1 equiv.) was dissolved in 7 mL of anhydrous DCM, and the mixture was purged with stream of ozone at  $-30$  °C, until the color of the reaction mixture turned into dark blue. The reaction mixture was then brought to  $0$  °C, and dimethyl sulfide (0.038 mL, 0.510 mmol, 10 equiv.) was added. The reaction mixture was allowed to warm up to room temperature and stirred for 5 h. After that, the reaction mixture was concentrated and

partitioned between diethyl ether (10 mL) and water (10 mL), the aqueous layer was extracted with diethyl ether (2x10 mL). the combined organic layers were washed with brine (20 mL), dried over anhydrous MgSO<sub>4</sub> and concentrated. The crude mixture was then purified by column chromatography, with hexane/EtOAc/AcOH (84/15/1) mixture, to obtain 2-(3-benzoylphenyl)-4-oxopentanoic acid **S6.1.9**, as colorless oil (14.2 mg, 94% yield).

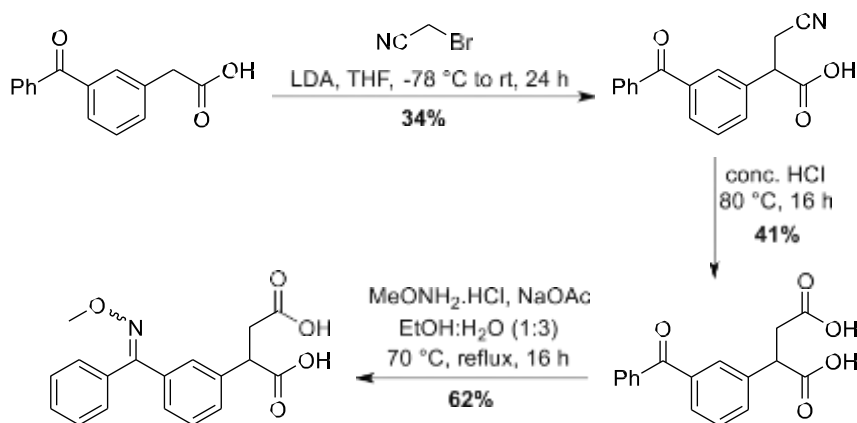
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.78-7.74 (m, 3 H), 7.69-7.67 (m, 1 H), 7.60-7.57 (m, 1 H), 7.54-7.52 (m, 1 H), 7.49-7.42 (m, 3 H), 4.20 (dd, *J* = 9.92, 4.52 Hz, 1 H), 3.38 (dd, *J* = 17.9, = 9.84 Hz, 1 H), 2.78 (dd, *J* = 18.0, 4.6 Hz, 1 H), 2.17 (s, 3 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 205.7, 196.3, 177.5, 138.2, 137.9, 137.2, 132.6, 132.0, 130.1, 129.6, 129.4, 128.8, 128.3, 46.3, 45.8, 29.9;

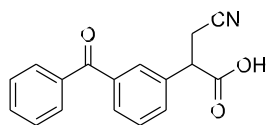
HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>18</sub>H<sub>16</sub>O<sub>4</sub>Na: 319.0946, found: 319.0948;

IR (film, CDCl<sub>3</sub>): 3061, 3027, 2956, 2924, 1714, 1658, 1597, 1318, 1283, 1163, 720 cm<sup>-1</sup>.

### Illustration of compounds **S6.1.10** to **S6.1.12** described in main-text Figure 3.a



**Figure S16.** Synthesis of compounds **S6.1.10** to **S6.1.12** described in main-text Figure 3.a.



### 2-(3-benzoylphenyl)-3-cyanopropanoic acid **S6.1.10**

LDA (2M in THF, 0.94 mL, 1.873 mmol, 2.5 equiv.) was added to 1 mL of anhydrous THF at -78 °C under argon atmosphere. A solution of 2-(3-benzoylphenyl)acetic acid (180 mg, 0.750 mmol, 1 equiv.) in THF (1 mL) was then added, and the reaction mixture was slowly allowed to warm up to 0 °C, and stirred at that temperature for 1 h. After that, the reaction mixture was again cooled down to -78 °C, and neat 2-bromoacetonitrile (0.078 mL, 1.125 mmol, 1.5 equiv.) was added at -78 °C. The reaction mixture was then allowed to warm up to room temperature and stirred for 24 h. Then, quenched with water (5 mL), followed by 1N HCl (5 mL), and it was extracted with ethyl acetate (3x10 mL). The combined organic layers were collected, washed with brine, dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure to

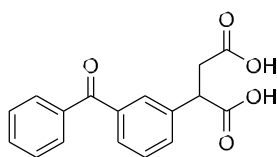
obtain the crude product, which was then purified by column chromatography, with hexane/EtOAc/AcOH(79/20/1) mixture, to afford 2-(3-benzoylphenyl)-3-cyanopropanoic acid **S6.1.10**, as yellow sticky oil (71.2 mg, 34% yield).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.80-7.75 (m, 4 H), 7.62-7.54 (m, 2 H), 7.52-7.46 (m, 3 H), 4.05 (t, *J* = 7.5 Hz, 1 H), 3.05 (dd, *J* = 16.9, 7.0 Hz, 1 H), 2.88 (dd, *J* = 16.9, 7.9 Hz, 1 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.2, 175.1, 138.4, 136.9, 135.6, 132.9, 131.7, 130.5, 130.1, 129.3, 128.4, 117.1, 47.2, 21.2;

HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>17</sub>H<sub>13</sub>NO<sub>3</sub>Na: 302.0793, found: 302.0797;

IR (film, CDCl<sub>3</sub>): 2958, 2925, 2853, 2252, 1712, 1658, 1598, 1494, 1447, 1318, 1281, 1185, 1081, 718 cm<sup>-1</sup>.



### 2-(3-benzoylphenyl)succinic acid **S6.1.11**

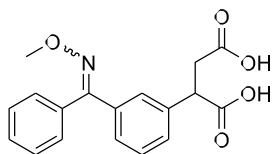
The obtained compound **S6.1.10** (71 mg, 0.254 mmol, 1 equiv.) was refluxed with 1 mL of conc. HCl for 16 h. After this, water (10 mL) and EtOAc (20 mL) were added to the reaction mixture, and was extracted with EtOAc (3x20 mL). The combined organic layers were collected, washed with brine, dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure to get the crude product, which was then purified by column chromatography, with hexane/EtOAc/AcOH (74/25/1) mixture, to afford 2-(3-benzoylphenyl)succinic acid **S6.1.11**, as colorless oil (31 mg, 41% yield).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.77-7.75 (m, 3 H), 7.70 (d, *J* = 7.6 Hz, 1 H), 7.60-7.54 (m, 2 H), 7.48-7.43 (m, 3 H), 4.18 (dd, *J* = 11.3, 3.96 Hz, 1 H), 3.28 (dd, *J* = 17.4, 11.4 Hz, 1 H), 2.73 (dd, *J* = 17.4, 4.04 Hz, 1 H);

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 196.2, 178.3, 177.3, 138.3, 137.2, 137.0, 132.7, 131.7, 130.1, 129.9, 129.4, 129.0, 128.4, 46.9, 37.3;

HRMS (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>17</sub>H<sub>14</sub>O<sub>5</sub>Na: 321.0739, found: 321.0741;

IR (film, CDCl<sub>3</sub>): 3027, 2926, 1712, 1658, 1597, 1416, 1319, 1284, 1718, 1001, 720 cm<sup>-1</sup>.



### 2-(3-((methoxyimino)(phenyl)methyl)phenyl)succinic acid **S6.1.12**

To a solution of compound **S6.1.11** (33.7 mg, 0.113 mmol, 1 equiv.) in EtOH:H<sub>2</sub>O (1:3, 1.33 mL), was added methoxamine hydrochloride (25.48 mg, 0.305 mmol, 2.7 equiv.) followed by anhydrous sodium acetate (40.78 mg, 0.497, 4.4 equiv.). The mixture was then refluxed for 16



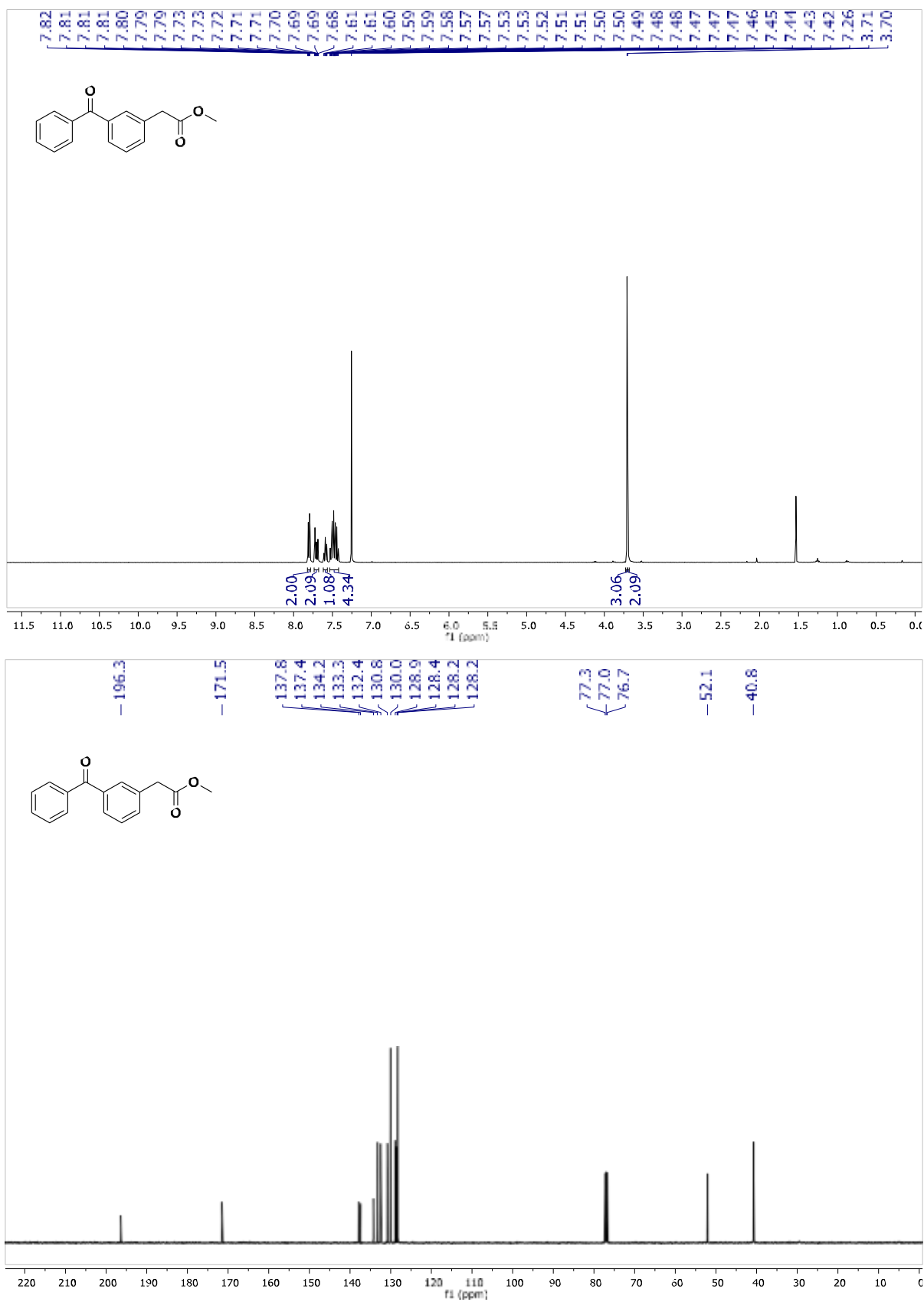
h. EtOAc (10 mL) and water (10 mL) were then added to the reaction mixture. The organic layer was extracted with EtOAc (3x20 mL). The combined organic layer was collected, washed with brine dried over anhydrous MgSO<sub>4</sub>, and concentrated under reduced pressure to obtain the crude product, which was then purified by column chromatography, with hexane/EtOAc/AcOH (84/15/1) mixture, to afford the target 2-(3-((methoxyimino)(phenyl)methyl)phenyl)succinic acid **S6.1.12**, as light yellow sticky oil (23 mg, 62% yield).

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.53 (s, 1 H), 7.45-7.39 (m, 6 H), 7.37-7.28 (m, 11 H), 4.14-4.08 (m, 2 H), 3.98 (s, 3H), 3.95 (s, 3 H), 3.29-3.20 (m, 2 H), 2.74-2.65 (m, 2 H);

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>) δ 178.9, 178.9, 177.8, 177.8, 156.1, 156.0, 137.2, 136.6, 136.3, 136.1, 134.0, 132.9, 129.4, 129.1, 129.1, 129.0, 128.9, 128.8, 128.4, 128.3, 128.2, 128.1, 127.9, 127.8, 127.1, 62.5, 62.4, 47.0, 46.9, 37.4;

**HRMS** (ESI) m/z: [M]<sup>+</sup> Calc. for C<sub>18</sub>H<sub>17</sub>NO<sub>5</sub>Na: 350.1004, found: 350.1008;

**IR** (film, CDCl<sub>3</sub>): 3026, 2936, 1712, 1442, 1422, 1289, 1252, 1216, 1184, 1054, 758 cm<sup>-1</sup>.



**Figure S17.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.1.

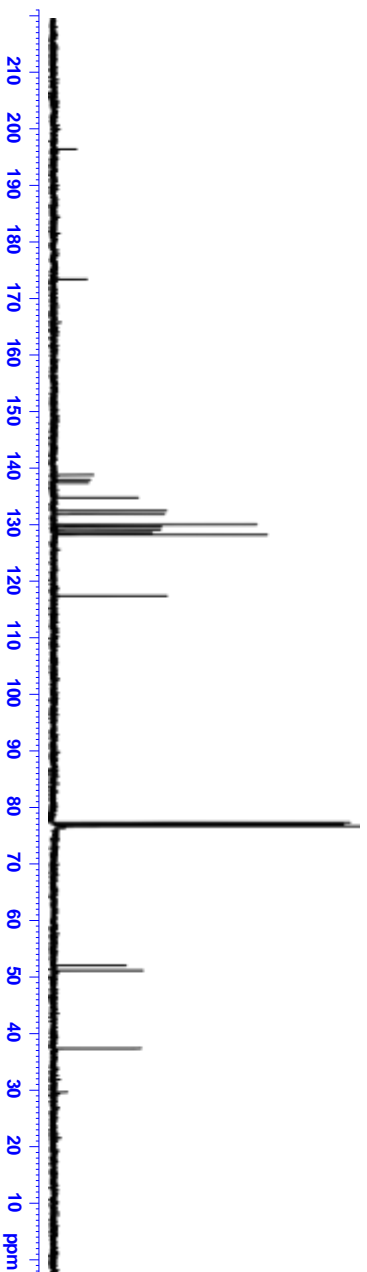
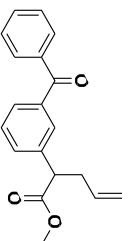
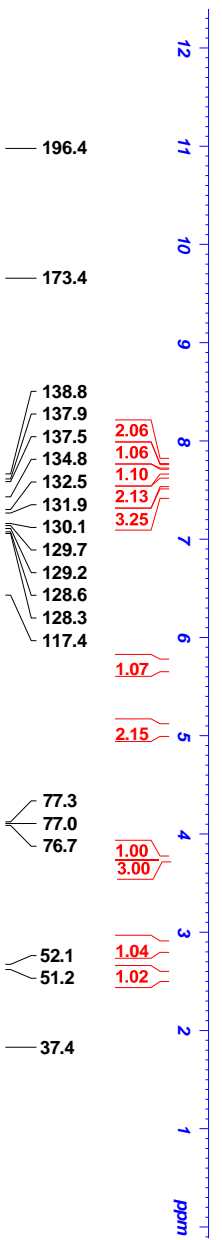
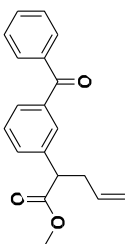
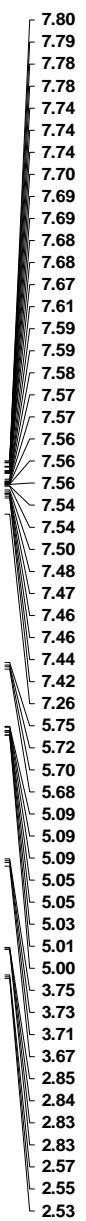


Figure S18. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.2.

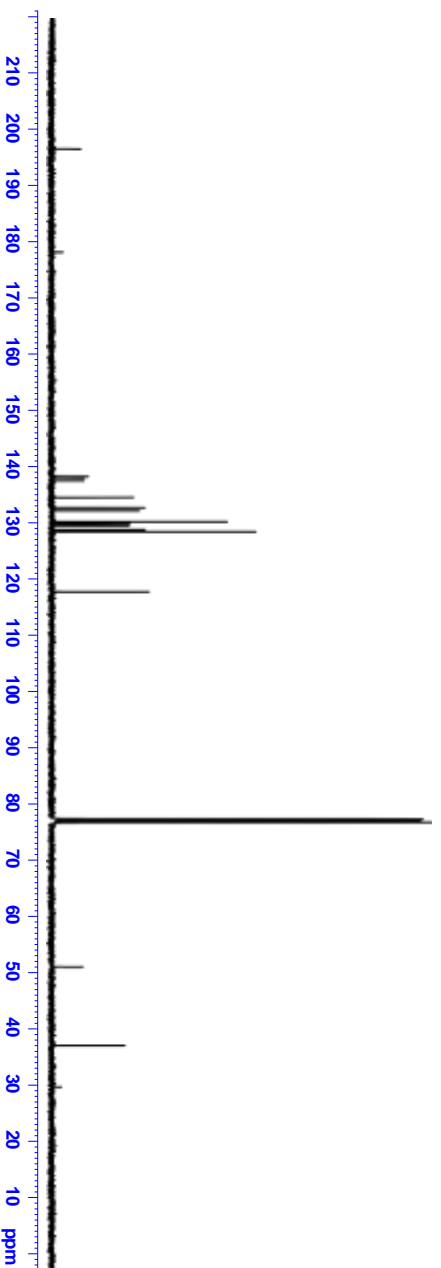
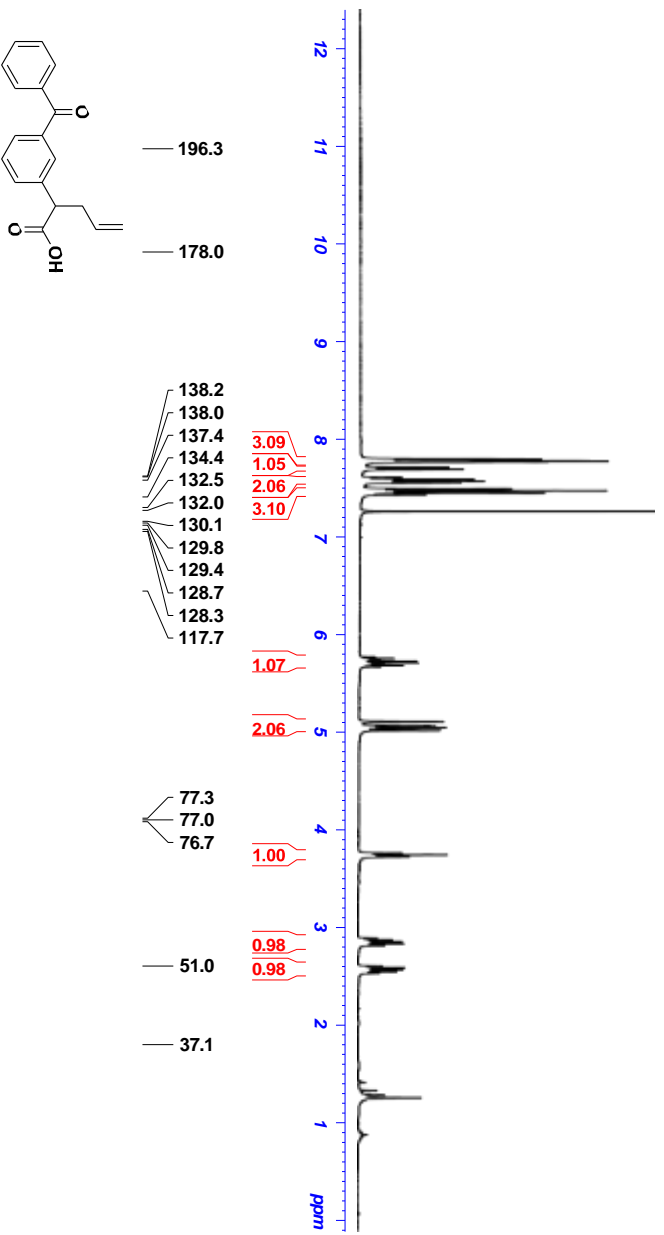
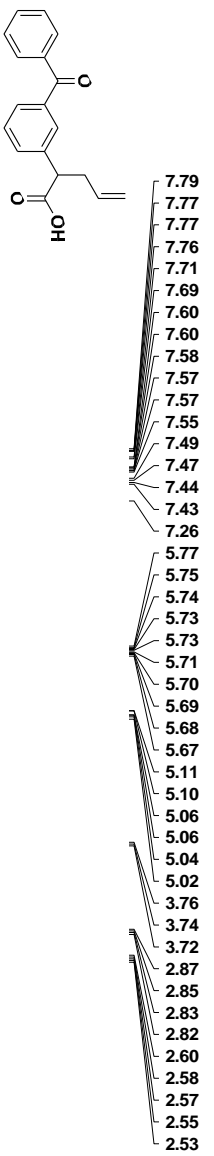
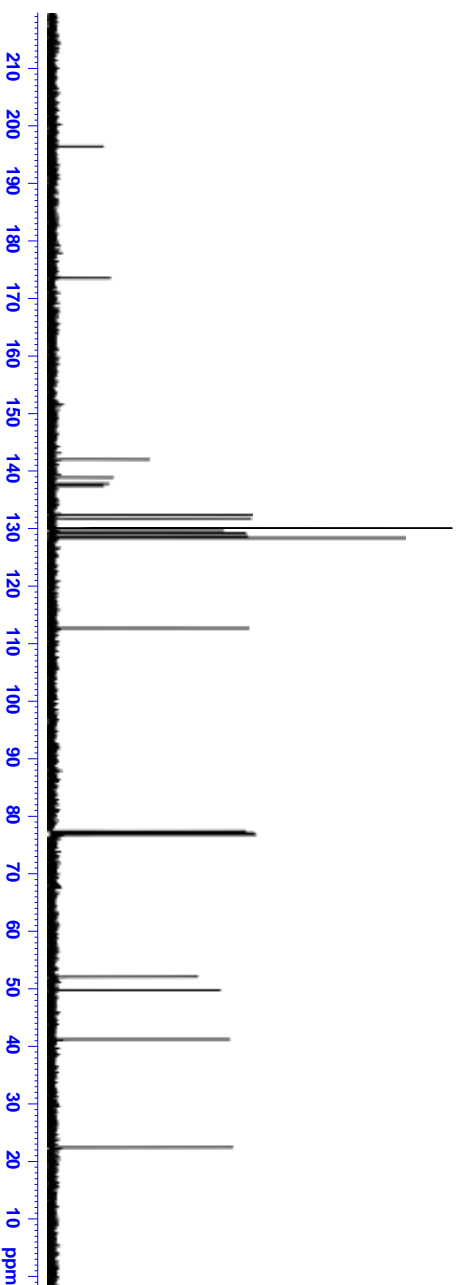
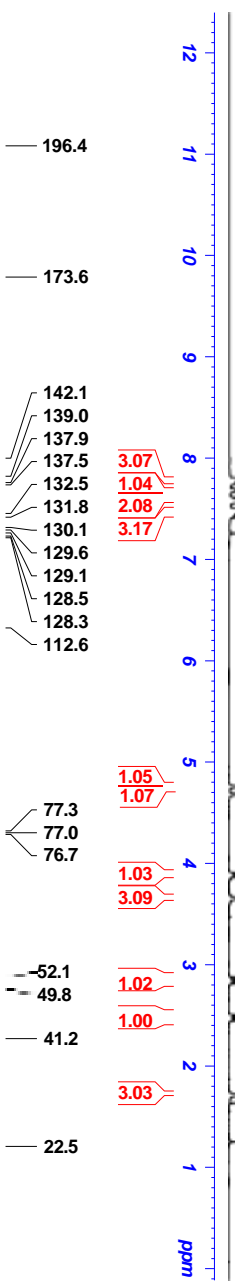
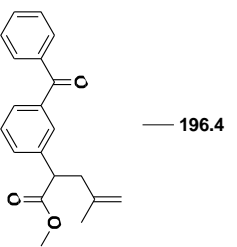
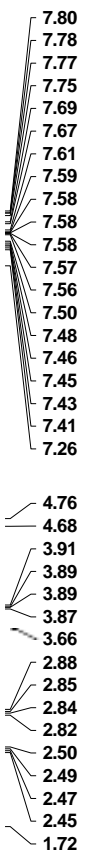
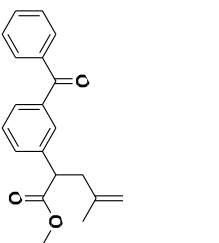
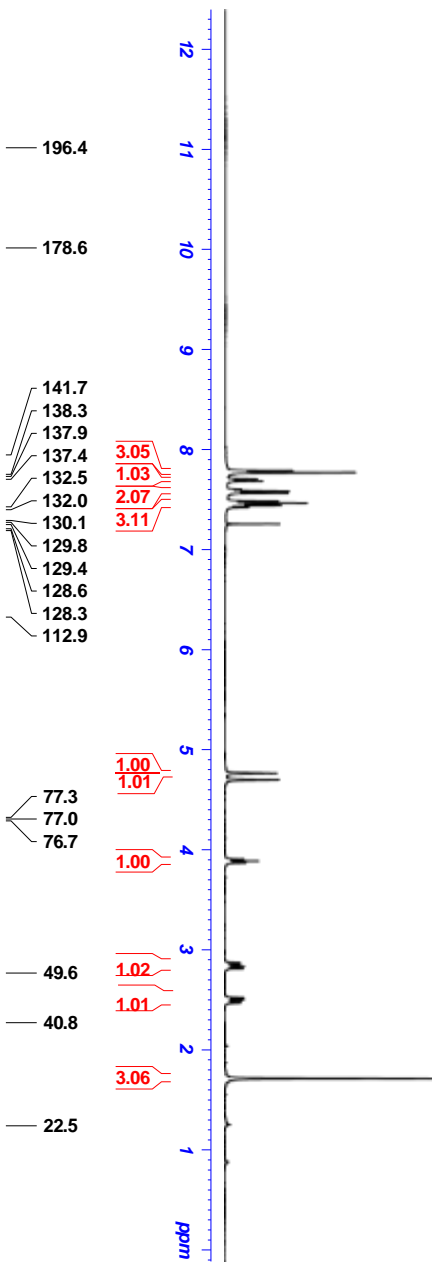
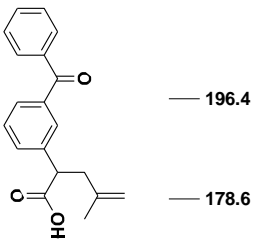
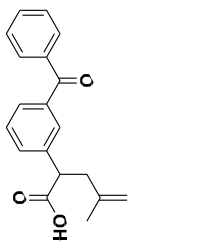


Figure S19. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.3.



**Figure S20.**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound **S6.1.4**.



196.4

178.6

141.7

138.3

137.9

137.4

132.5

132.0

130.1

129.8

129.4

128.6

128.3

112.9

77.3

77.0

76.7

49.6

40.8

22.5

3.05

1.03

2.07

3.11

1.00

1.01

1.00

1.02

1.01

3.06

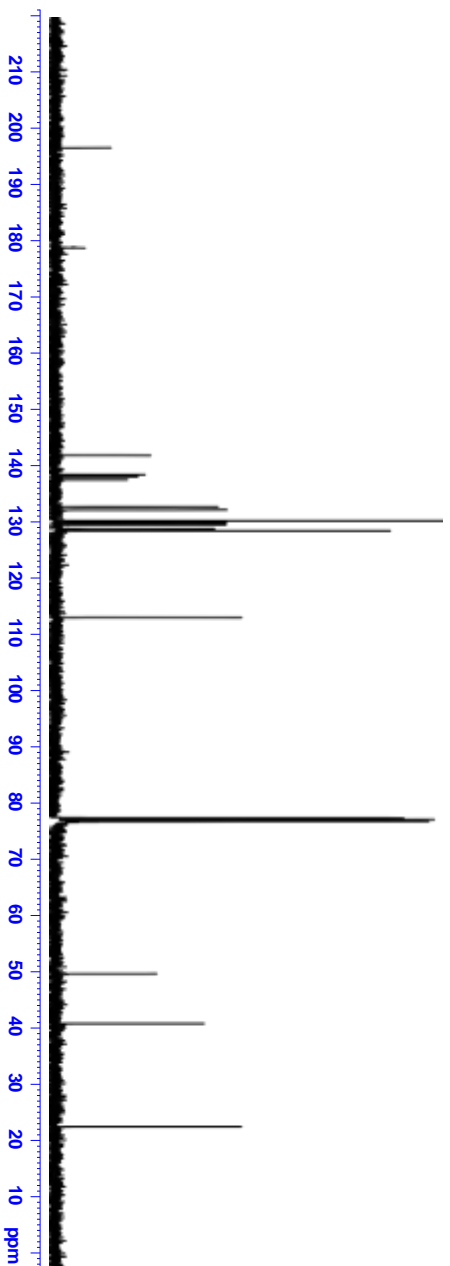
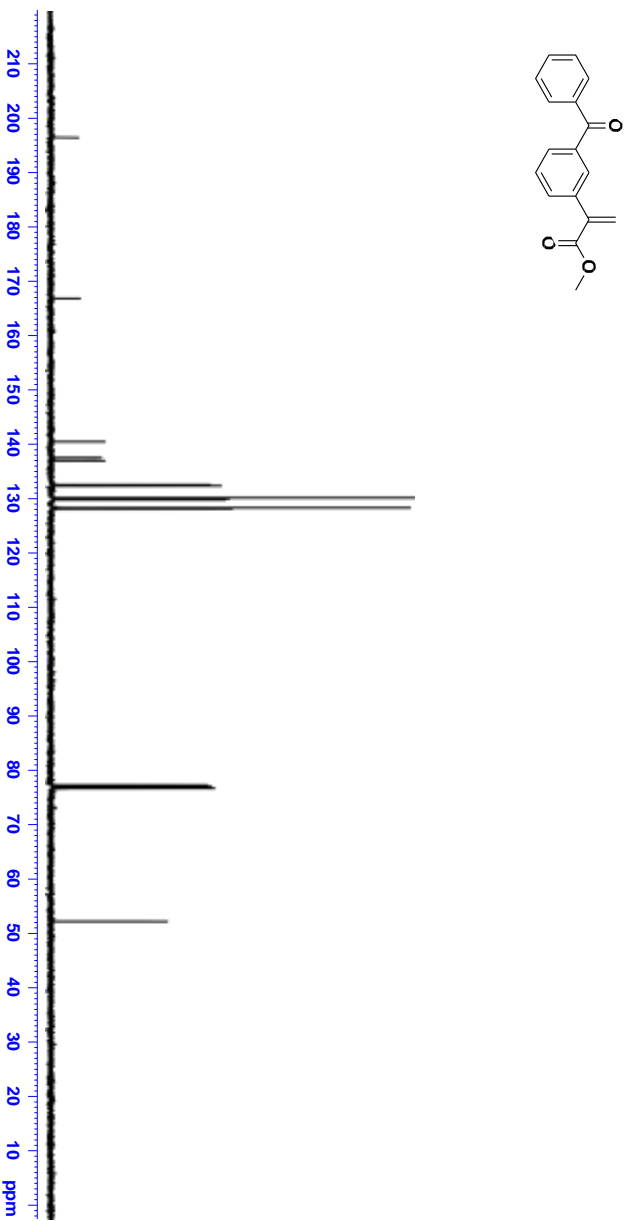
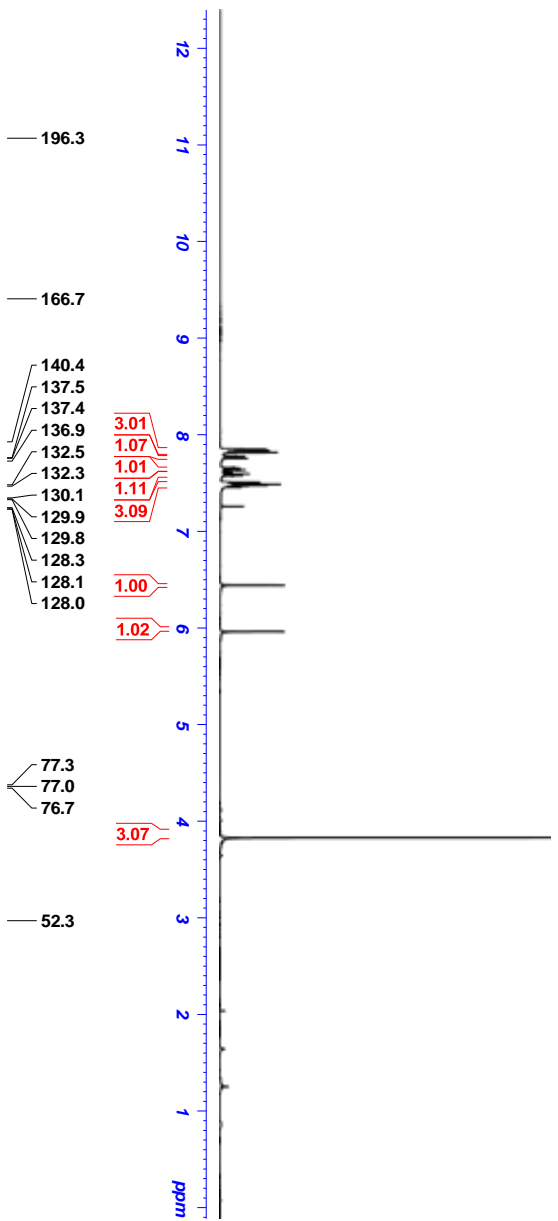
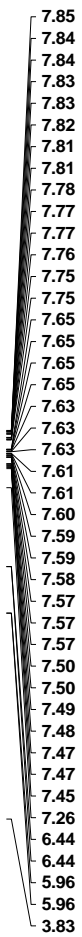


Figure S21. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.5.



**Figure S22.**  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.1.6.

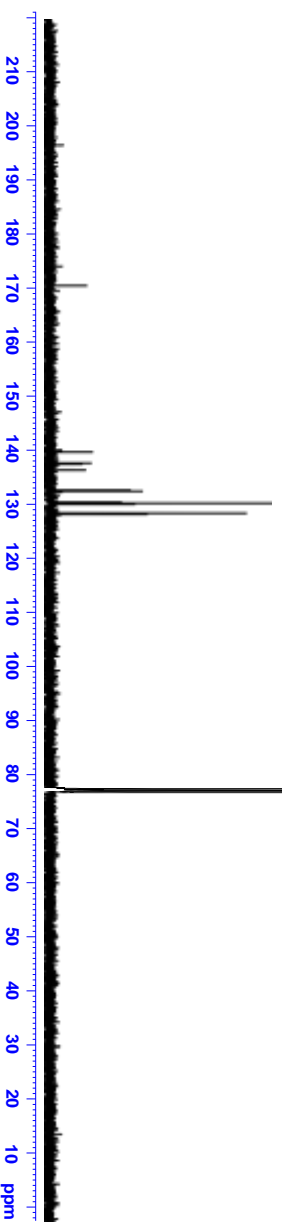
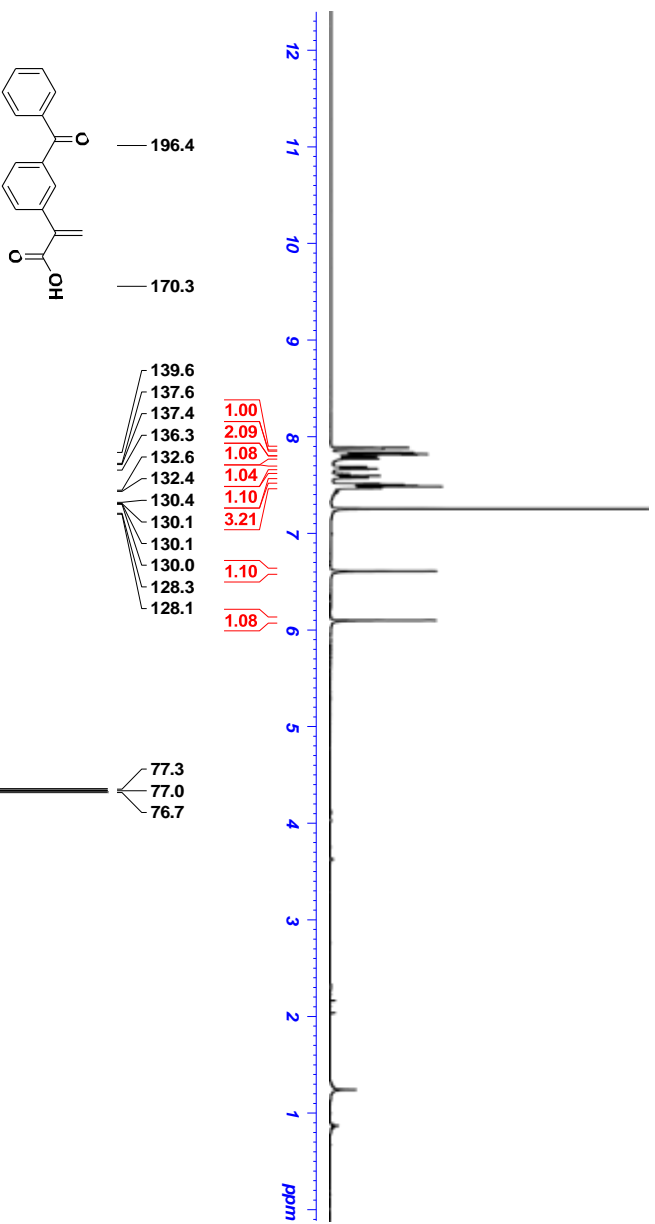
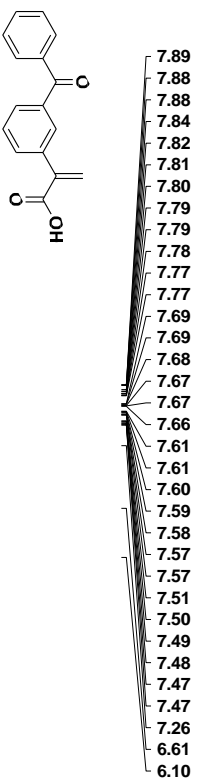


Figure S23. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.7.





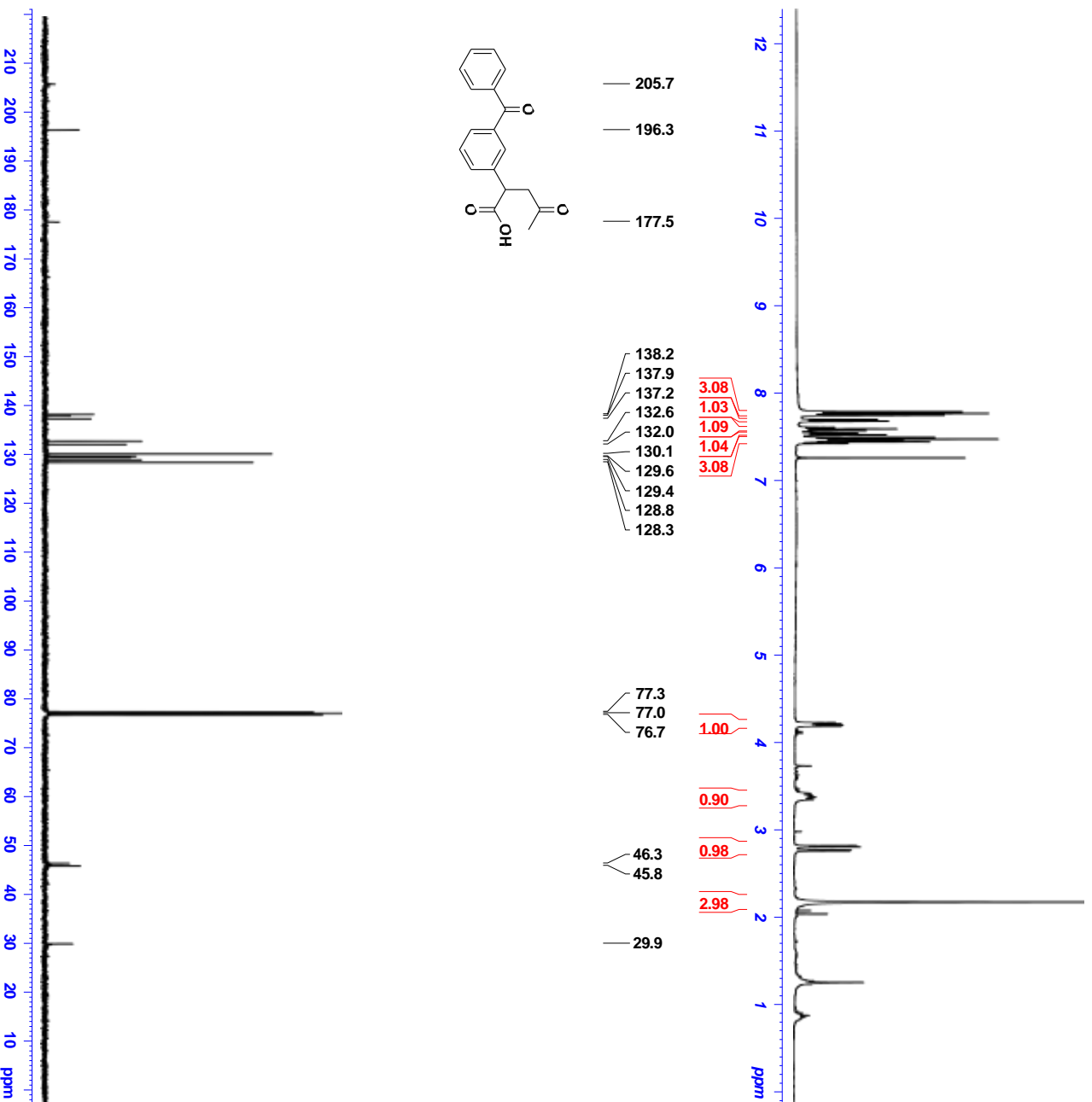
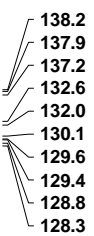
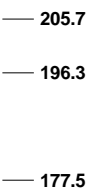
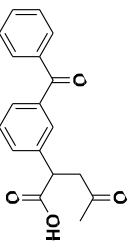
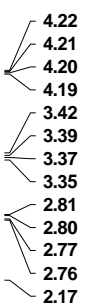
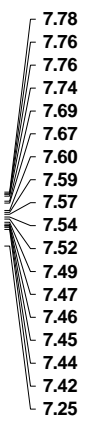
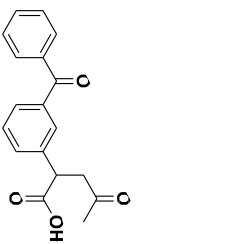


Figure S25. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.9.

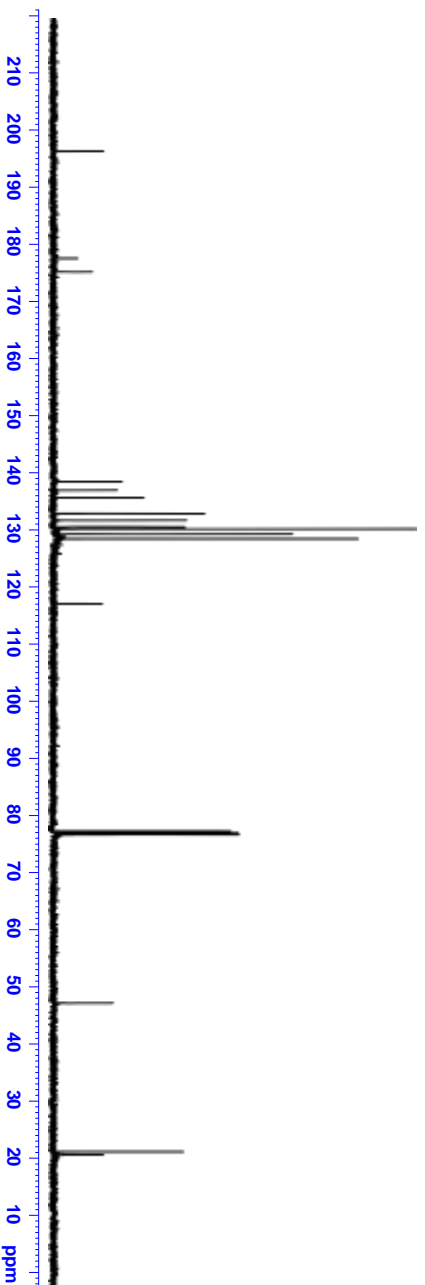
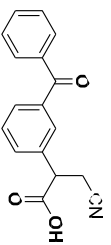
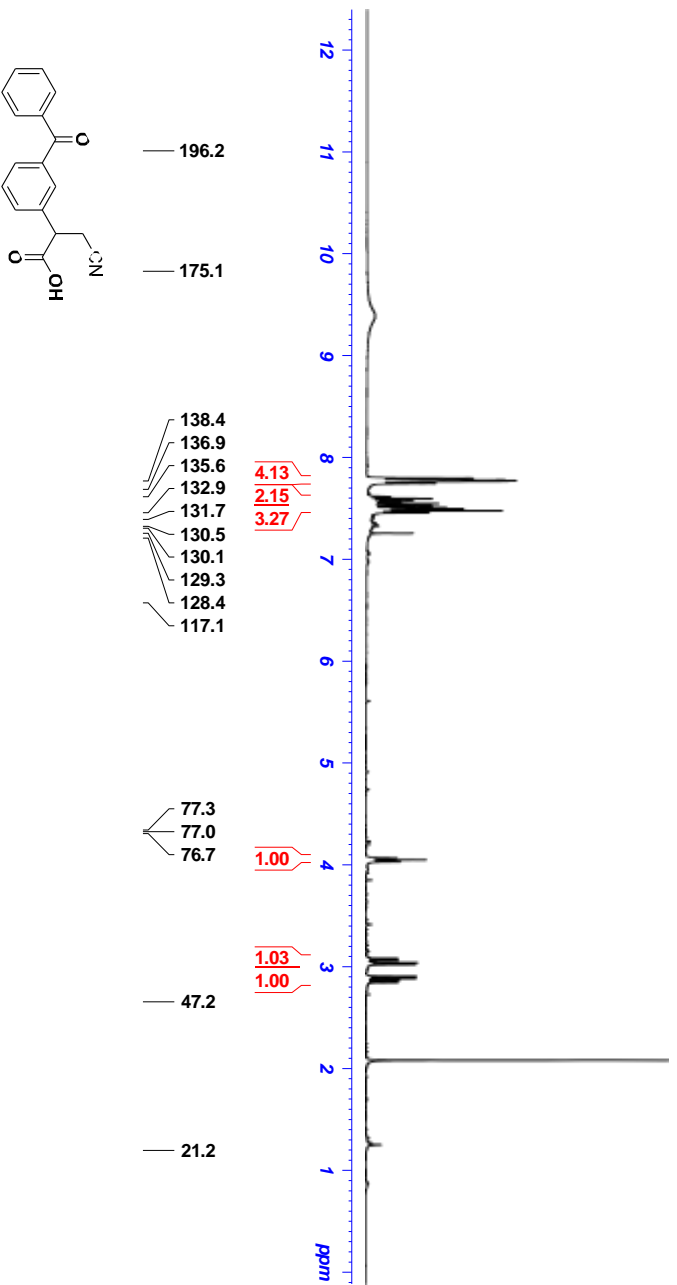
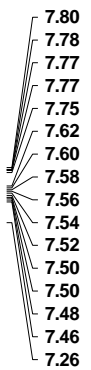
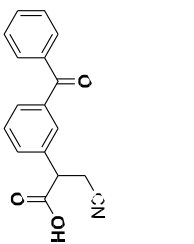


Figure S26. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.10.

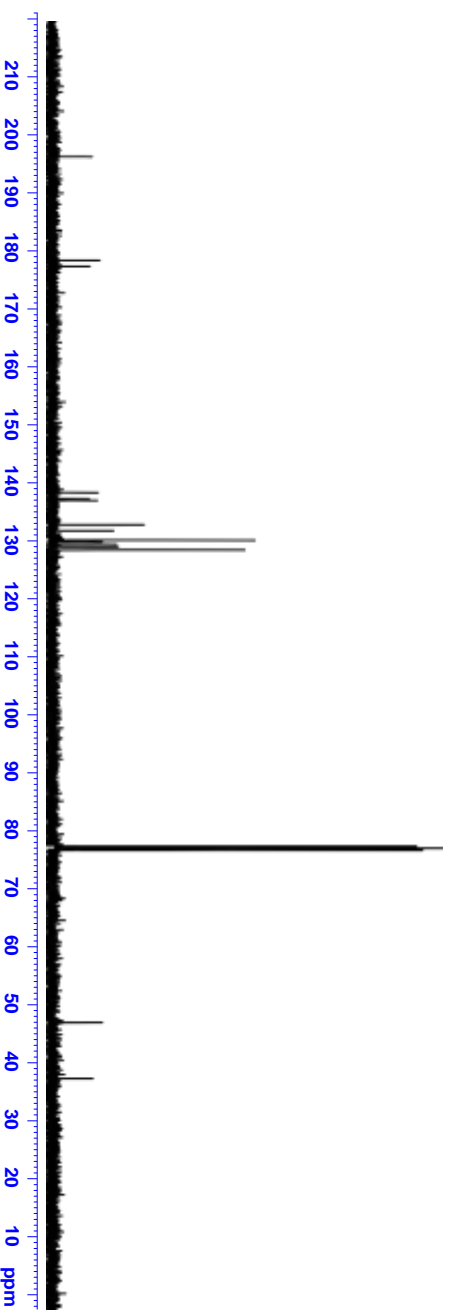
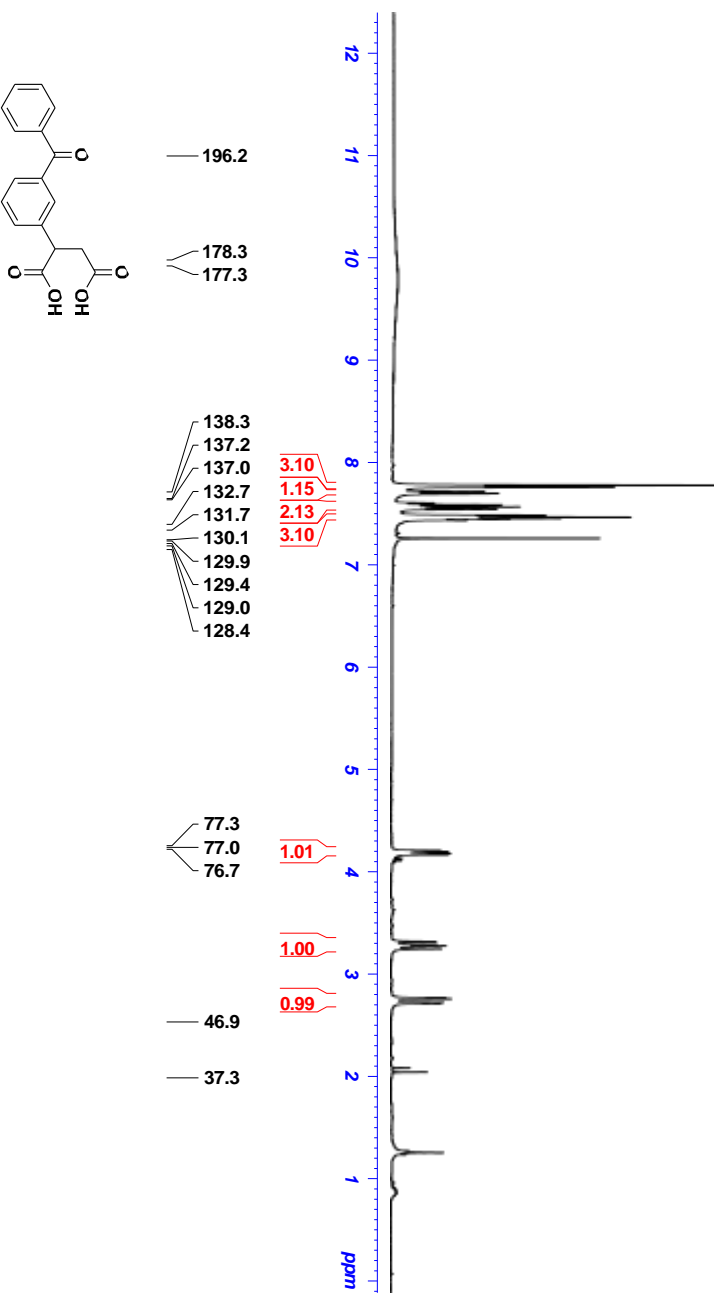
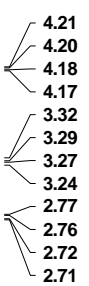
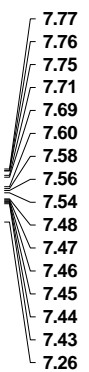
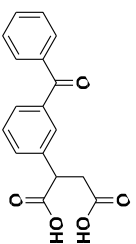


Figure S27. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.11.

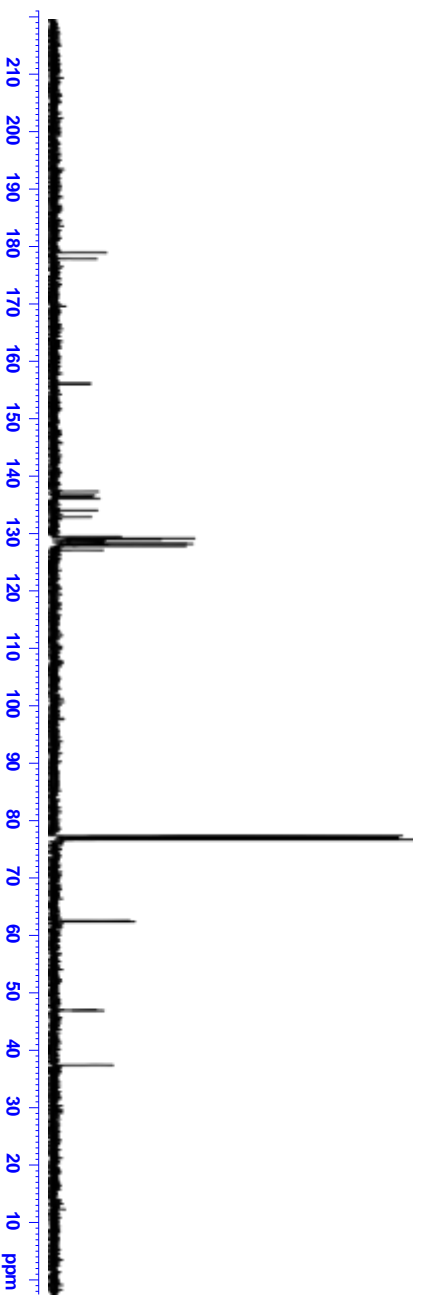
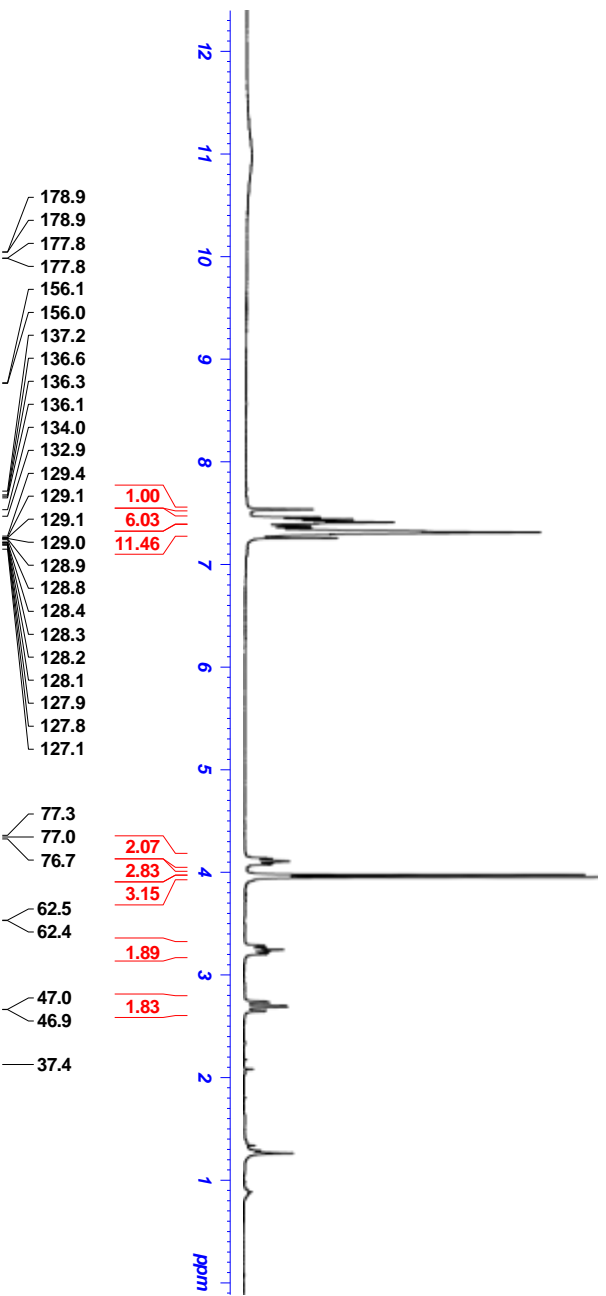
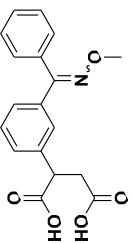
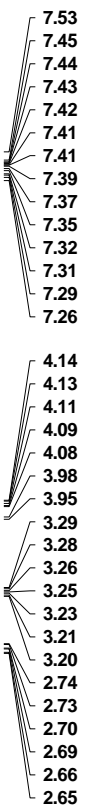
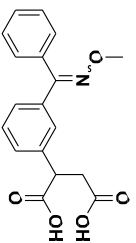
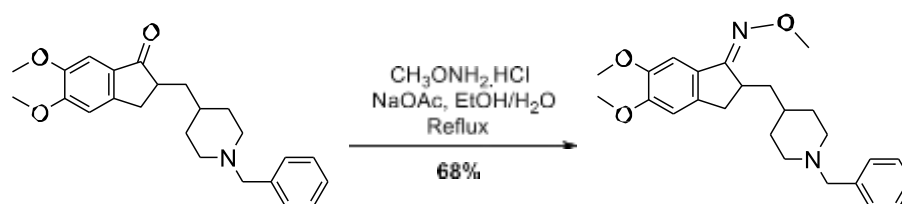


Figure S28. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.1.12.

## Section S6.2. Donepezil's analogs described in main-text Figure 4.a

### Synthesis of (E)-2-((1-benzylpiperidin-4-yl)methyl)-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one O-methyl oxime S6.2.1



### Scheme S3. Reaction scheme of compound S6.2.1 described in main-text Figure 4.a.

To a solution of donepezil hydrochloride (100 mg, 0.240 mmol, 1 equiv.) in EtOH/H<sub>2</sub>O (1.0/2.5 mL), was added O-methyl hydroxylamine hydrochloride (54.3 mg, 0.650 mmol, 2.7 equiv.) and sodium acetate (87.0 mg, 1.06 mmol, 4.4 equiv.) at room temperature. It was then refluxed for 20 h. The solvents mixture was directly evaporated, and the crude was purified on column chromatography with EtOAc/MeOH/acetic acid (90/10/2, v/v/v), to afford the desired product **S6.2.1** (67 mg, 68%), as light-yellow sticky oil.

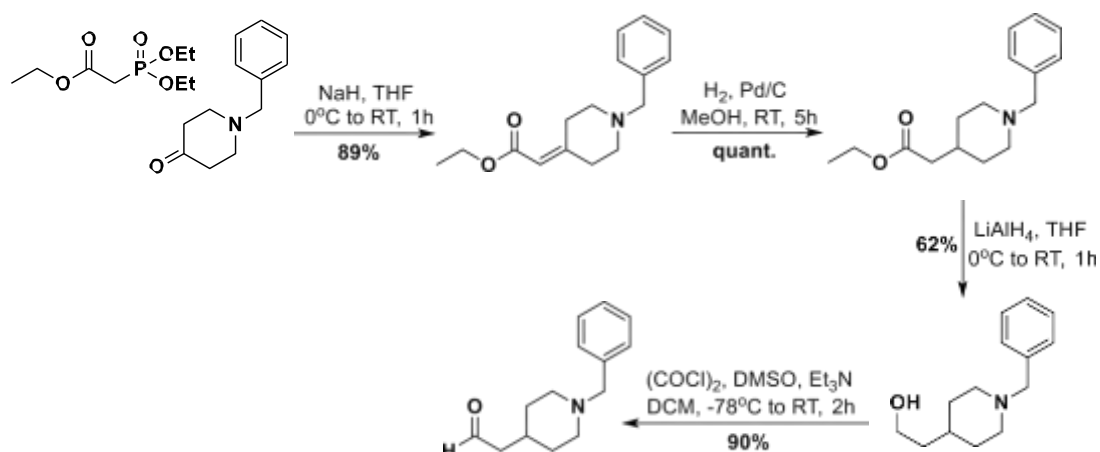
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.54-7.53 (m, 2H), 7.42-7.41 (m, 3H), 7.08 (s, 1H), 6.73 (s, 1H), 4.02 (s, 2H), 3.91-3.89 (m, 9H), 3.42-3.37 (m, 1H), 3.31-3.28 (m, 2H), 3.13-3.07 (m, 1H), 2.58-2.54 (m, 3H), 1.87-1.78 (m, 5H), 1.42-1.37 (m, 1H), 1.29 (m, 1H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 164.99, 151.84, 149.05 (overlapped), 139.55, 130.98 (overlapped), 129.38, 129.02 (overlapped), 127.47, 107.47, 103.27, 61.85, 61.15, 56.05, 55.96, 52.30, 52.13, 38.63, 37.80, 36.02, 32.91, 30.13, 29.69;

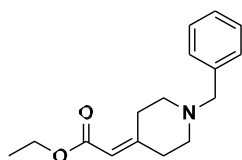
**HRMS** (ESI) m/z: [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>33</sub>N<sub>2</sub>O<sub>3</sub> 409.2491; Found 409.2497;

**IR** (film, CDCl<sub>3</sub>): 2952, 2926, 1717, 1602, 1500, 1460, 1332, 1043, 733 cm<sup>-1</sup>.

**General synthesis of the substrate 2-(1-benzylpiperidin-4-yl)acetaldehyde S6.2.5, described in main-text Figure 4.a**



**Figure S29. Synthesis of compounds S6.2.2 to S6.2.5.**



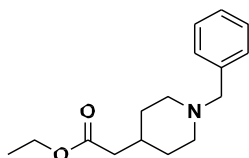
**Ethyl 2-(1-benzylpiperidin-4-ylidene)acetate S6.2.2**

Sodium hydride (60% in mineral oil, 0.253 g, 6.34 mmol, 1.2 equiv.) was added to anhydrous THF (8 mL) and the resulting mixture was stirred for 10 min in an ice bath. Then, triethyl phosphonoacetate (1.42 g, 6.34 mmol, 1.2 equiv.) was slowly added at the same temperature. The resulting solution was stirred for 30 min at RT, and cooled in an ice bath, followed by adding a solution of 1-benzyl-4-piperidone (1 g, 5.28 mmol, 1 equiv.) in anhydrous THF (3 mL). The resulting solution was stirred for no longer than 30 min at room temperature under inert atmosphere. The reaction mixture was then cooled back to 0 °C and a saturated aqueous solution of NH<sub>4</sub>Cl was added. The aqueous layer was extracted twice with EtOAc. The organic layer was separated, dried over MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. The concentrate was purified by Flash chromatography (Hex./EtOAc, 7/3, v/v) to afford the desired product **S6.2.2** (1.23 g, 89%), as a colorless liquid.

The spectral data match those reported in the literature<sup>S10</sup>;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.35-7.32 (m, 4H), 7.30-7.26 (m, 1H), 5.66 (s, 1H), 4.19-4.14 (q, *J* = 7.13 Hz, 2H), 3.55 (s, 2H), 3.03-3.00 (t, *J* = 5.59 Hz, 2H), 2.56-2.53 (m, 4H), 2.36-2.33 (t, *J* = 5.58 Hz, 2H), 1.31-1.28 (t, *J* = 7.14 Hz, 3H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 166.56, 159.56, 138.32, 129.05 (overlapped), 128.24 (overlapped), 127.07, 114.03, 62.59, 59.58, 54.54, 54.10, 36.82, 29.45, 14.32.



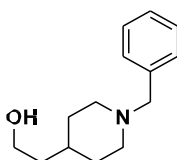
### Ethyl 2-(1-benzylpiperidin-4-yl)acetate **S6.2.3**

Following the procedure for the synthesis of compound **S6.2.3**<sup>S10</sup>. **S6.2.2** (1.22 g, 4.70 mmol) and Pd/C (0.035 g) in MeOH (8 mL) were used at RT. A colorless liquid of **S6.2.3** (quant.) was obtained, no further purification was required.

The spectral data match those reported in the literature<sup>S10</sup>;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>).  $\delta$  7.33-7.28 (m, 4H), 7.26-7.23 (m, 1H), 4.16-4.13 (m, 2H), 3.50 (s, 2H), 3.09-3.06 (d,  $J = 11.93$  Hz, 2H), 2.88-2.86 (d,  $J = 11.76$  Hz, 2H), 2.67-2.61 (dd,  $J = 13.01, 11.16$  Hz, 2H), 2.24-2.23 (m, 3H), 2.03-1.96 (m, 2H), 1.28-1.26 (m, 3H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>).  $\delta$  172.79, 138.57, 129.14 (overlapped), 128.12 (overlapped), 126.87, 63.40, 60.16, 53.52 (overlapped), 41.29, 32.97, 32.11 (overlapped), 14.27.



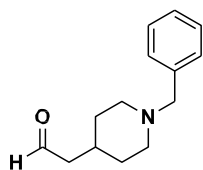
### 2-(1-benzylpiperidin-4-yl)ethan-1-ol **S6.2.4**

Following a modified procedure for the synthesis of compound **S6.2.4**<sup>S10</sup>. Compound **S6.2.3** (0.930 g, 3.55 mmol, 1 equiv.) and LiAlH<sub>4</sub> (1M in THF, 9.9 mmol, 2.5 equiv.) in THF (16 mL) were used. The temperature was allowed to warm up from 0°C to RT, and the reaction was kept stirring for 1h at RT. A colorless liquid of **S6.2.4** (0.480 g, 62%) was obtained, no further purification was required.

The spectral data match those reported in the literature<sup>S10</sup>;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>).  $\delta$  7.33-7.30 (m, 4H), 7.28-7.24 (m, 1H), 3.70-3.66 (t,  $J = 6.64$  Hz, 2H), 3.50 (s, 2H), 2.90-2.87 (d,  $J = 11.68$  Hz, 2H), 2.27 (m, 1H), 2.00-1.93 (m, 2H), 1.69-1.66 (m, 2H), 1.55-1.50 (m, 2H), 1.45-1.41 (m, 1H), 1.35-1.28 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>).  $\delta$  138.40, 129.27 (overlapped), 128.12 (overlapped), 126.90, 63.52, 60.41, 53.81 (overlapped), 39.46, 32.33, 32.31 (overlapped).



### 2-(1-benzylpiperidin-4-yl)acetaldehyde **S6.2.5**



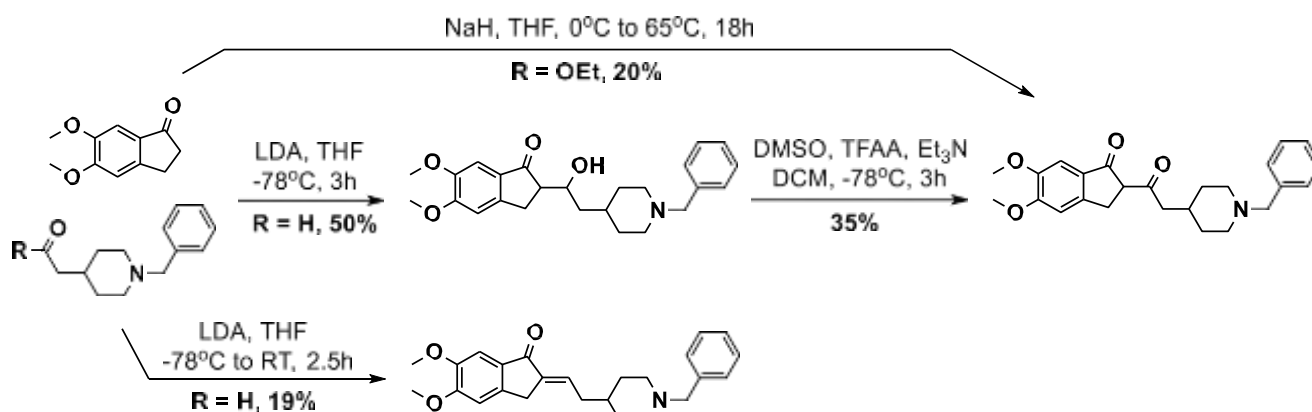
Following the same procedure for **S6.2.5** synthesis<sup>S10</sup>. Compound **S6.2.4** (0.475 g, 2.16 mmol, 1 equiv.) in anhydrous DCM (1 mL), oxalyl chloride (0.548 g, 4.32 mmol, 2 equiv.) in anhydrous DCM (11 mL), Et<sub>3</sub>N (0.931 g, 9.2 mmol, 4.25 equiv.), and DMSO (0.37 mL) were used at -78°C. Purification was done by using silica gel column chromatography (Hex./EtOAc, 4/6 to 2/8, v/v) to afford the desired product **S6.2.5** (0.420 g, 90%) as a colorless liquid.

The spectral data match those reported in the literature<sup>S10</sup>;

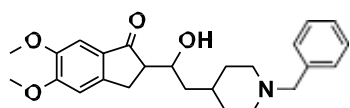
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 9.79 (m, 1H), 7.32-7.30 (m, 3H), 7.28-7.24 (m, 2H), 3.51 (s, 2H), 2.90-2.87 (d, *J* = 11.82 Hz, 2H), 2.39-2.37 (m, 2H), 2.06-1.99 (m, 2H), 1.95-1.88 (m, 1H), 1.72-1.69 (m, 2H), 1.42-1.31 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 202.18, 129.15 (overlapped), 128.16 (overlapped), 126.95, 63.39, 53.49 (overlapped), 50.49, 32.21 (overlapped), 30.66.

**General figure of compounds S6.2.6 to S6.2.8, described in main-text Figure 4.a**



**Figure S30. Reactions of compounds S6.2.6 to S6.2.8.**



**2-(2-(1-benzylpiperidin-4-yl)-1-hydroxyethyl)-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one S6.2.6.**

Under inert atmosphere, diisopropylamine (0.088 g, 0.87 mmol, 1.2 equiv.) was added to a flask containing anhydrous THF (2 mL), n-butyllithium (1.6 M in hexanes, 0.80 mmol, 1.1 equiv.) was added dropwise at -78 °C, and the mixture was stirred for 1 h at that temperature. To the reaction mixture was added 5,6-dimethoxy-1-indanone (0.140 g, 0.728 mmol, 1 equiv.) dissolved in THF (2 mL), and it was stirred at 0 °C for 1 h. The temperature was cooled back to -78 °C, and a solution of **S6.2.5** (0.190 g, 0.87 mmol, 1.2 equiv.) in THF (2 mL) was added. The reaction was stirred for 3 h between -78 °C and -50 °C before it was quenched with 2M HCl solution, and then extracted with EtOAc (2 x 30 mL). The organic extracts were washed with sat. NaHCO<sub>3</sub> and brine, then dried with anhydrous MgSO<sub>4</sub>. The solvent was removed in

vacuo, and the residue was purified using silica gel column chromatography (EtOAc/MeOH, 9/1, v/v) to afford the desired product **S6.2.6** (0.150 g, 50%), as white solids.

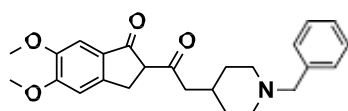
**m.p.** = 261-263 °C;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.33-7.30 (m, 4H), 7.27-7.23 (m, 1H), 7.18 (s, 1H), 6.90 (s, 1H), 4.54 (bs, 1H), 3.99 (s, 3H), 3.93 (s, 3H), 3.91-3.86 (m, 1H), 3.50 (s, 2H), 3.22-3.16 (m, 1H), 2.90-2.87 (d, *J* = 11.35 Hz, 2H), 2.71-2.63 (m, 2H), 2.06-1.96 (m, 2H), 1.84-1.81 (m, 1H), 1.74-1.67 (m, 2H), 1.64-1.57 (m, 1H), 1.39-1.30 (m, 2H), 1.28-1.15 (m, 1H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 208.4, 156.1, 149.7, 149.3, 138.7, 129.2 (overlapped), 129.1, 128.1 (overlapped), 126.8, 107.3, 104.3, 70.3, 63.5, 56.3, 56.1, 53.9, 53.7, 52.5, 43.0, 33.6, 31.6, 31.5, 29.7;

**HRMS** (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>32</sub>NO<sub>4</sub> 410.2331; Found 410.2332;

**IR** (film, CDCl<sub>3</sub>) 3443, 2923, 1683, 1501, 1316, 1265, 1121, 1040, 747 cm<sup>-1</sup>.



## 2-(2-(1-benzylpiperidin-4-yl)acetyl)-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one **S6.2.7**

### **Method A:**

To a solution of DMSO (47 μL) in DCM (0.5 mL) was added TFAA (0.067 g, 0.319 mmol, 2.9 equiv.) dropwise at -78 °C under argon. The resulting mixture was stirred at that temperature for 45 min. Then, a precooled solution of **S6.2.6** (0.045 g, 0.11 mmol, 1 equiv.) in DCM (0.4 mL) was added. The reaction mixture was stirred at -78 °C for 30 min, then at -15 °C for 15 min, and cooled back to -78°C. Et<sub>3</sub>N (0.100 g, 0.99 mmol, 9 equiv.) was then added, and the mixture was stirred at -78 °C for 45 min. The reaction was quenched by addition of sat. aqueous solution of NH<sub>4</sub>Cl and the mixture warmed to room temperature. The two phases were separated and the aqueous phase was extracted with DCM (3x) and then EtOAc (1x). The combined organic phase was washed with brine, dried with anhydrous MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (EtOAc/MeOH, 9/1, v/v) to afford the desired product **S6.2.7** (0.016 g, 35%) as white solids.

### **Method B:**

To a suspension of NaH (60% in mineral oil, 37.44 mg, 1.2 equiv.) in THF (0.5 mL) was added 5,6-dimethoxy-1-indanone (0.150 g, 0.78 mmol, 1 equiv.) in THF (1 mL) at 0 °C. After stirring for 10 min, **S6.2.3** (0.490 g, 1.87 mmol, 2.4 equiv.) was added dropwise. This mixture was refluxed for 18 h, cooled back to r.t. and poured onto ice-cold 1 M HCl (4 mL). Extraction was performed with EtOAc (15x3 mL), the organic phase was washed with brine, dried with anhydrous MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (EtOAc/MeOH, 9/1, v/v) to afford the desired product **S6.2.7** (0.063 g, 20%), as white solids.

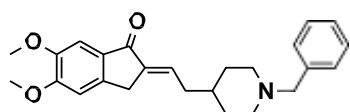
**m.p.** = 265-267 °C;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.33-7.32 (m, 5H), 7.27-7.24 (m, 1H), 6.96 (s, 1H), 3.99 (s, 3H), 3.95 (s, 3H), 3.93-3.91 (m, 1H), 3.51 (s, 2H), 2.91-2.88 (m, 3H), 2.32-2.30 (d, *J* = 7.18 Hz, 2H), 2.07-1.97 (m, 3H), 1.78-1.72 (m, 3H), 1.41-1.35 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 203.3 (194.5), 198.1 (172.9), 154.5 (154.4), 149.7 (149.5), 143.2 (144.5), 138.5 (138.7), 131.3 (134.6), 129.1 (overlapped), 128.1 (overlapped), 126.9 (126.8), 110.8 (overlapped), 107.4 (107.3), 63.4 (63.4), 61.9 (104.4), 56.2 (56.2-56.1), 53.8 (53.6) (overlapped), 40.6 (40.6), 34.0 (34.0), 32.5 (32.4-32.0), 29.6 (31.4);

**HRMS** (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>30</sub>NO<sub>4</sub> 408.2175; Found 408.2176;

**IR** (film, CDCl<sub>3</sub>) 2933, 1689, 1655, 1500, 1291, 1222, 1128, 750 cm<sup>-1</sup>.



**(E)-2-(2-(1-benzylpiperidin-4-yl)ethylidene)-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one**  
**S6.2.8**

Lithium diisopropylamide, LDA (2M in THF, 2.71 mmol, 2.0 equiv.) was added to 2 mL of anhydrous THF, the mixture was cooled to -78 °C before 5,6-dimethoxy-1-indanone (0.438 g, 2.28 mmol, 1.65 equiv.) in THF (5 mL) and (0.5 mL) of HMPA were added. The reaction mixture was stirred at that temperature for 15 min. A solution of **S6.2.5** (0.300 g, 1.38 mmol, 1 equiv.) in THF (5 mL) was then added, and the temperature was allowed to gradually increase to room temperature, followed by stirring for 2 h. An aqueous 1% ammonium chloride solution was added thereto, and the reaction mixture was extracted with EtOAc, The organic layers were collected, dried over MgSO<sub>4</sub>, filtered and concentrated in vacuo. The residue was purified by flash column chromatography (DCM/MeOH, 95/5, v/v) to afford the desired product **S6.2.8** (100 mg, 19%) as white solid.

**m.p.** = 243-245 °C;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.34-7.32 (m, 4H), 7.31 (s, 1H), 7.29-7.28 (m, 1H), 6.93 (s, 1H), 6.82-6.78 (t, *J* = 7.83 Hz, 1H), 4.00 (s, 3H), 3.95 (s, 3H), 3.57-3.55 (m, 4H), 2.98-2.95 (m, 2H), 2.28-2.24 (t, *J* = 7.30 Hz, 2H), 2.06-2.01 (m, 2H), 1.77-1.73 (m, 2H), 1.60-1.53 (m, 1H), 1.46-1.40 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 192.12, 155.33, 149.53, 144.56, 137.90, 134.38, 131.85, 129.88, 129.53, 129.49, 128.37, 128.25, 127.26, 107.31, 105.07, 63.01, 56.25, 56.14, 53.42, 53.23, 36.68, 35.73, 31.98, 29.84, 29.69;

**HRMS** (ESI) *m/z*: [M]<sup>+</sup> Calc. for C<sub>25</sub>H<sub>30</sub>NO<sub>3</sub> 392.2226; Found 392.2228;

**IR** (film, CDCl<sub>3</sub>) 2924, 1694, 1650, 1501, 1307, 1254, 1129 cm<sup>-1</sup>.

General synthesis of the substrate 1-benzylpiperidine-4-carbaldehyde S6.2.12, described in main-text Figure 4.a

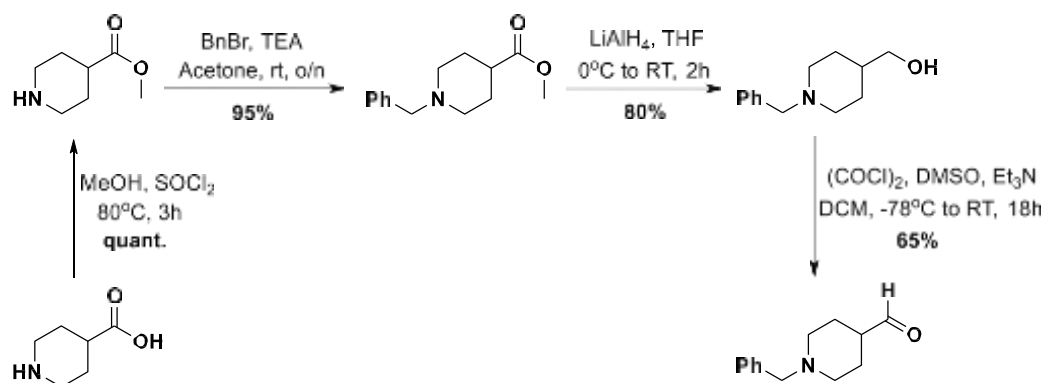
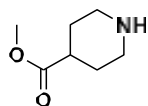


Figure S31. Synthesis of compounds S6.2.9 to S6.2.12.

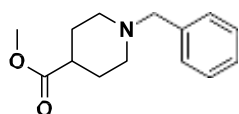


#### Methyl piperidine-4-carboxylate S6.2.9

To a solution of 4-piperidinecarboxylic acid (2 g, 15.48 mmol, 1 equiv.) in methanol (12 mL) was added thionyl chloride (1.84 g, 15.48 mmol, 1 equiv.) drop-wisely, the reaction is exothermic, so the temperature was gradually increased during the dropwise addition. Then, the mixture was set at 80 °C, and kept stirring for 3 h. After that, it was concentrated under reduced pressure to give the desired product **S6.2.9** (2.2 g, quant.), as white solids.

<sup>1</sup>H NMR (400 MHz, DMSO). δ 3.63 (s, 3H), 3.20-3.17 (m, 2H), 2.92-2.86 (m, 2H), 2.51 (m, 1H), 1.98-1.95 (m, 2H), 1.84-1.74 (m, 2H);

<sup>13</sup>C NMR (101 MHz, DMSO). δ 174.1, 52.2, 42.5 (overlapped), 38.1, 24.8 (overlapped).



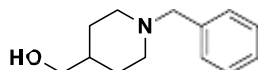
#### Methyl 1-benzylpiperidine-4-carboxylate S6.2.10

To a solution of methyl piperidine-4-carboxylate **S6.2.9** (2.2 g, 15.36 mmol, 1 equiv.) in acetone (18 mL), triethylamine (4.3 mL, 30.72 mmol, 2 equiv.) was added at r.t., followed by benzyl bromide (1.82 mL, 15.36 mmol, 1 equiv.). The mixture was stirred at r.t. for 16 h, and then it was quenched with water, and extracted with DCM. The organic layers were collected, dried over MgSO<sub>4</sub>, filtered and concentrated in vacuo. The crude yellow liquid was purified by flash chromatography (Hexane/EtOAc, 8/2, v/v) to afford the desired product **S6.2.10** (3.43 g, 95%), as light-yellow liquid.

The spectral data match those reported in the literature<sup>S11</sup>;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.34-7.30 (m, 4H), 7.29-7.25 (m, 1H), 3.69 (s, 3H), 3.53 (s, 2H), 2.89-2.87 (m, 2H), 2.37-2.29 (m, 1H), 2.10-2.05 (m, 2H), 1.93-1.90 (m, 2H), 1.85-1.75 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 175.7, 138.4, 129.1 (overlapped), 128.2 (overlapped), 127.0, 63.2, 52.9 (overlapped), 51.6, 41.1, 28.3 (overlapped).



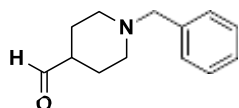
### (1-benzylpiperidin-4-yl)methanol **S6.2.11**

To a suspension of LiAlH<sub>4</sub> (1.67 g, 44.10 mmol, 3 equiv.) in 30 mL of anhydrous THF was added, dropwise, a solution of methyl ester **S6.2.10** (3.43 g, 14.70 mmol, 1 equiv.) in 30 mL of anhydrous THF, while maintaining the temperature at 0 °C. After the addition was completed, the reaction mixture was allowed to stir at r.t. for 2 h. The reaction was then cooled back to 0 °C, and 40 mL of cold water was added with vigorous stirring. Insoluble material was filtered off, and the filtrate was concentrated in vacuo to remove THF. The resulting aqueous suspension was extracted with EtOAc and dried over MgSO<sub>4</sub>, filtered and concentrated in vacuo to give us an oily liquid (2.4 g, 80%), as the desired product **S6.2.11**.

The spectral data match those reported in the literature<sup>S11</sup>;

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.34-7.30 (m, 4H), 7.28-7.26 (m, 1H), 3.52-3.50 (m, 4H), 2.95-2.92 (m, 2H), 2.02-1.96 (m, 2H), 1.75-1.71 (m, 2H), 1.57-1.47 (m, 1H), 1.36-1.26 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 138.5, 129.2 (overlapped), 128.1 (overlapped), 126.9, 68.0, 63.5, 53.4 (overlapped), 38.6, 28.8 (overlapped).



### 1-benzylpiperidine-4-carbaldehyde **S6.2.12**

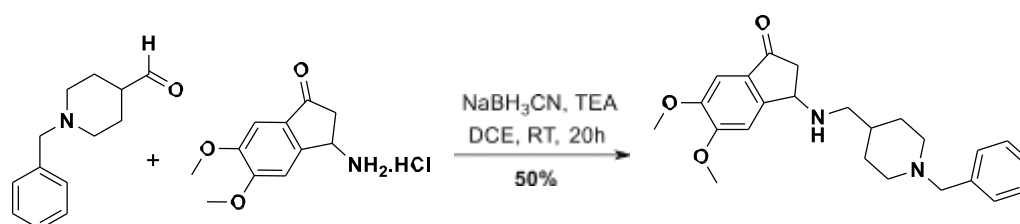
To a solution of oxalyl chloride (0.387 mL, 4.57 mmol, 2 equiv.) in 11 mL of anhydrous DCM was added dropwise dimethyl sulfoxide (0.4 mL, 5.63 mmol, 2.45 equiv.) at -60 °C. The reaction mixture was stirred at the same temperature for 10 min before adding the solution of the alcohol **S6.2.11** (0.470 g, 2.29 mmol, 1 equiv.) in 1 mL of anhydrous DCM. After 15 min, trimethylamine (1.36 mL, 9.728 mmol, 4.25 equiv.) was added dropwise. Then, reaction mixture was warmed to ambient temperature over 3 h, and then poured into water. The organic phase was separated, and the aqueous phase was extracted twice with DCM. The combined organic phase was dried over MgSO<sub>4</sub> and concentrated in vacuo. The residual oil was purified by flash chromatography (Hexane/EtOAc, 7/3, v/v) to afford the desired product **S6.2.12** as oily liquid (300 mg, 65%).

The spectral data match those reported in the literature<sup>S11</sup>;

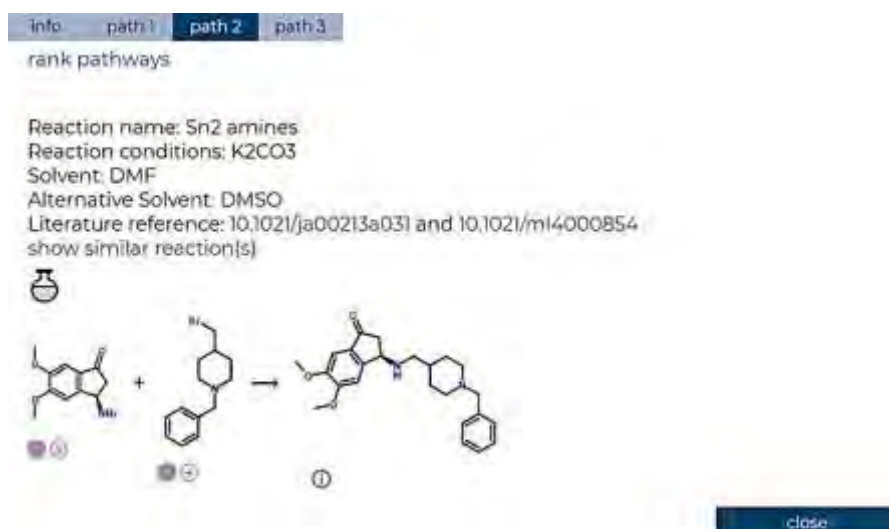
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 9.67 (s, 1H), 7.34-7.33 (m, 4H), 7.30-7.26 (m, 1H), 3.53 (s, 2H), 2.86-2.83 (m, 2H), 2.29-2.23 (m, 1H), 2.16-2.11 (m, 2H), 1.92-1.89 (m, 2H), 1.77-1.67 (m, 2H);

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>). δ 204.0, 129.1 (overlapped), 128.2 (overlapped), 127.1, 63.2, 52.5 (overlapped), 48.0, 25.4 (overlapped).

### Synthesis of 3-(((1-benzylpiperidin-4-yl)methyl)amino)-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one S6.2.13



**Scheme S4.** Reaction scheme of compound S6.2.13 described in main-text Figure 4.a.



**Figure S32.** A screenshot from Allchemy showing the route of synthesis for compound S6.2.13.

Under inert atmosphere, to a solution of 3-amino-5,6-dimethoxy-2,3-dihydro-1H-inden-1-one hydrochloride<sup>S12</sup> (0.180 g, 0.738 mmol, 1.3 equiv.) in DCE (2 mL) was added trimethylamine (0.134 mL, 0.965 mmol, 1.7 equiv.) dropwise at room temperature. 1-benzylpiperidine-4-carbaldehyde S6.2.12 (0.116 g, 0.568 mmol, 1 equiv.) and NaBH<sub>3</sub>CN (0.064 g, 1.02 mmol, 1.8 equiv.) were then added portion-wise, and the mixture was kept stirring at room temperature for 20 h. The reaction was monitored by TLC. The reaction was then quenched with aq. NaHCO<sub>3</sub>, extracted with EtOAc, the combined organic phase was dried over MgSO<sub>4</sub> and concentrated in vacuo, and the crude mixture was purified by flash chromatography (EtOAc/MeOH, 5/5, v/v) to give the desired product S6.2.13 (113 mg, 50%), as colorless sticky oil.

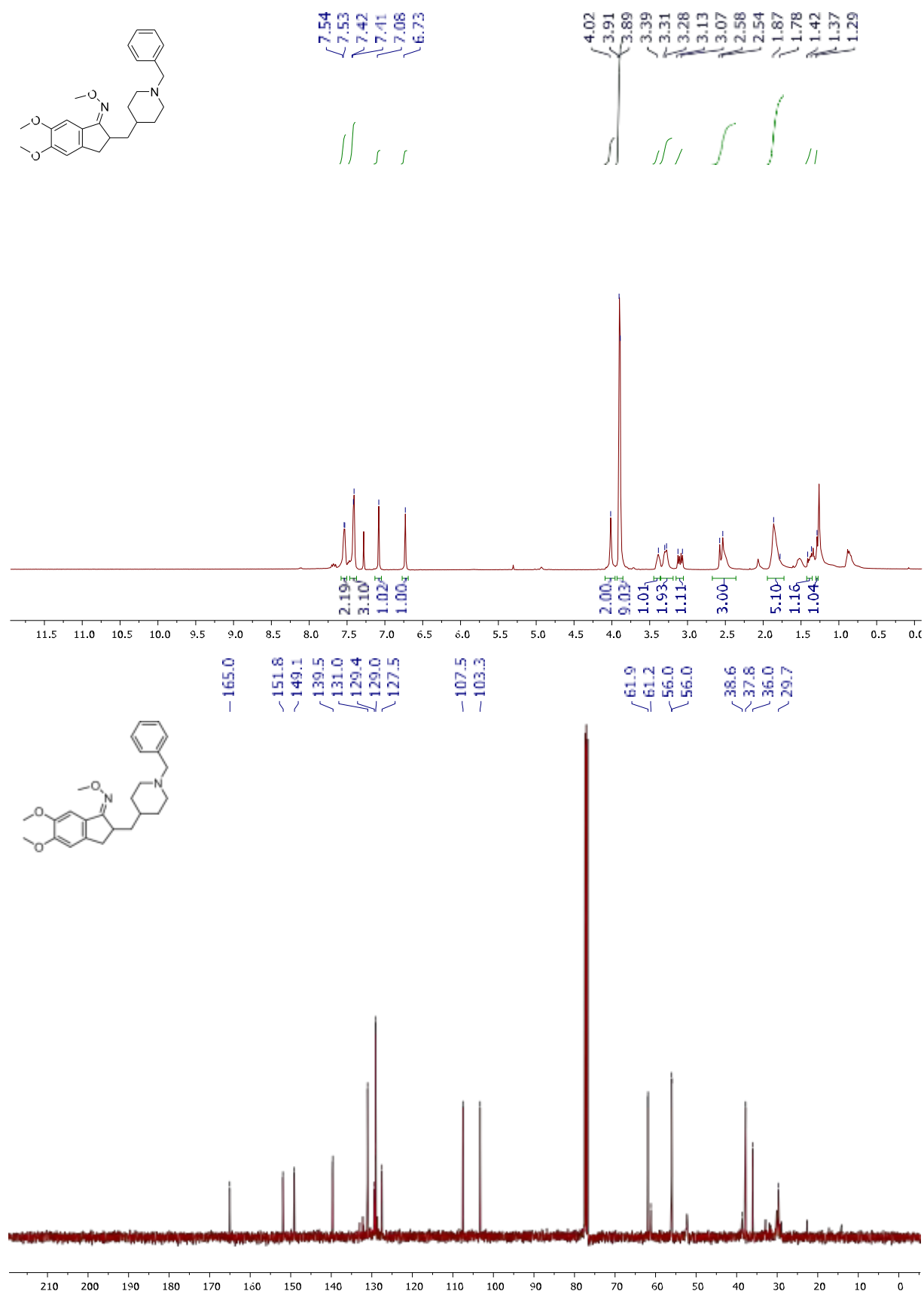
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>). δ 7.41-7.38 (m, 5H), 7.17 (s, 1H), 7.11 (s, 1H), 4.43-4.41 (dd, *J* = 2.48, 6.46 Hz, 1H), 4.01 (s, 3H), 3.93 (s, 3H), 3.87 (s, 2H), 3.26-3.23 (m, 2H), 2.98-2.91 (dd,

$J = 6.60, 18.57$  Hz, 1H), 2.65-2.55 (m, 2H), 2.52-2.47 (dd,  $J = 2.72, 18.57$  Hz, 1H), 2.38-2.32 (m, 2H), 1.89-1.86 (m, 2H), 1.35-1.27 (m, 3H);

**$^{13}\text{C}$  NMR** (101 MHz,  $\text{CDCl}_3$ ).  $\delta$  202.8, 155.7, 150.5, 130.2 (overlapped), 129.9, 128.7 (overlapped), 128.4, 106.8, 103.7, 62.1, 56.4, 56.2, 56.0, 52.8, 52.7, 52.1, 44.3, 35.8, 29.7, 28.9;

**HRMS** (ESI)  $m/z$ :  $[\text{M}]^+$  Calc. for  $\text{C}_{24}\text{H}_{31}\text{N}_2\text{O}_3$  395.2335; Found 395.2337;

**IR** (film,  $\text{CDCl}_3$ ) 3408, 2923, 2850, 1698, 1500, 1390, 1311, 1272  $\text{cm}^{-1}$ .



**Figure S33.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.1.



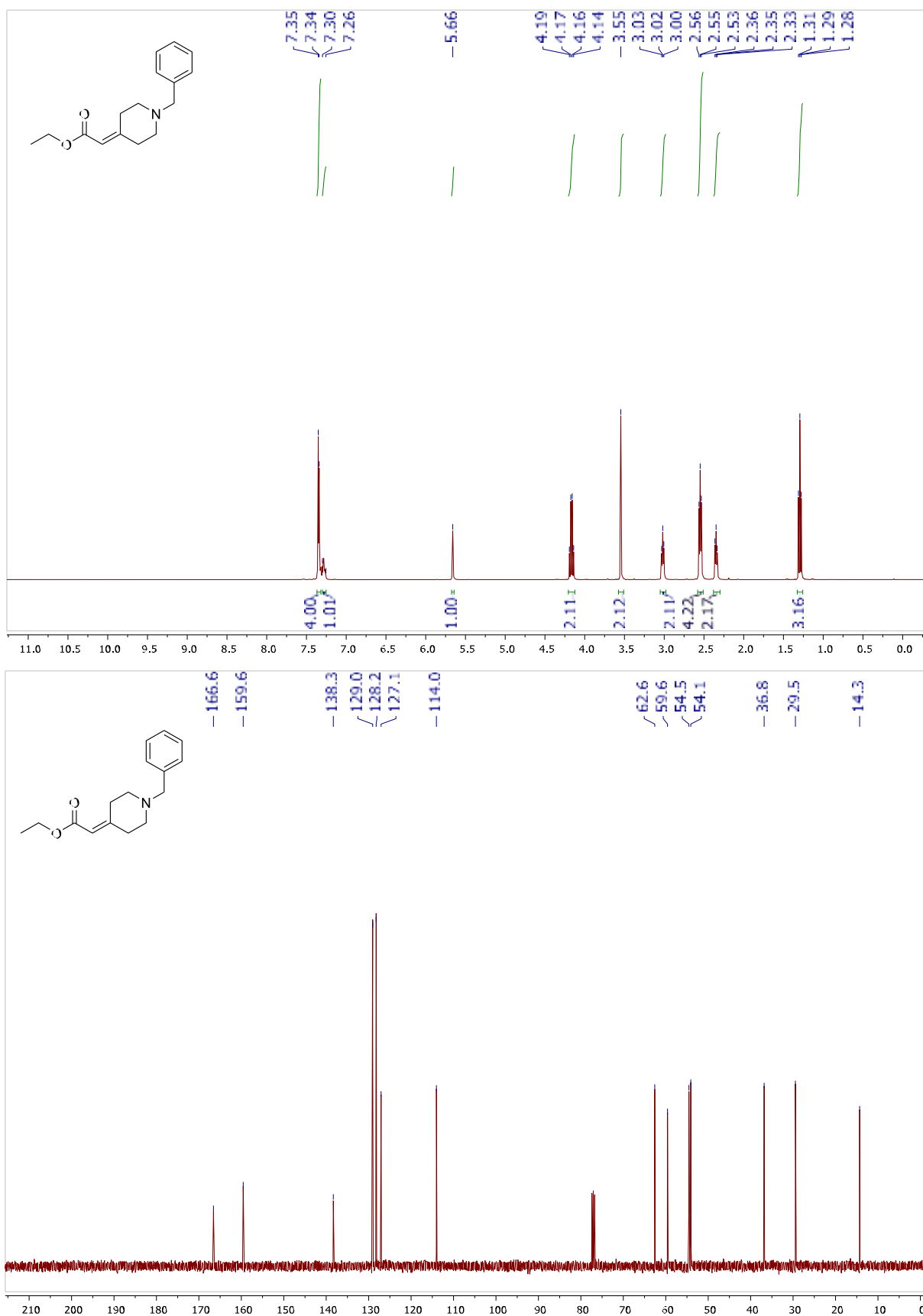


Figure S34. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.2.

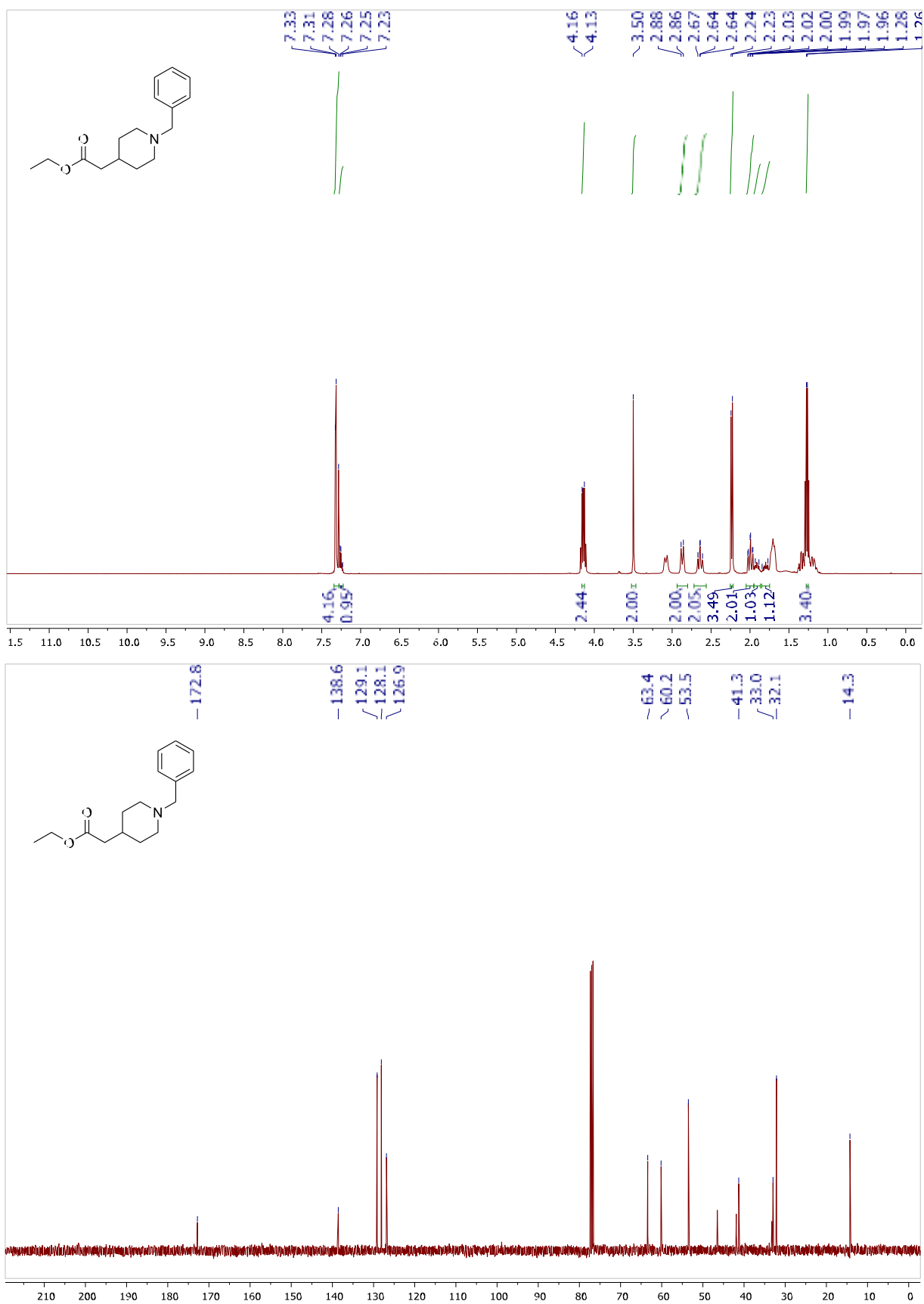


Figure S35. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.3.

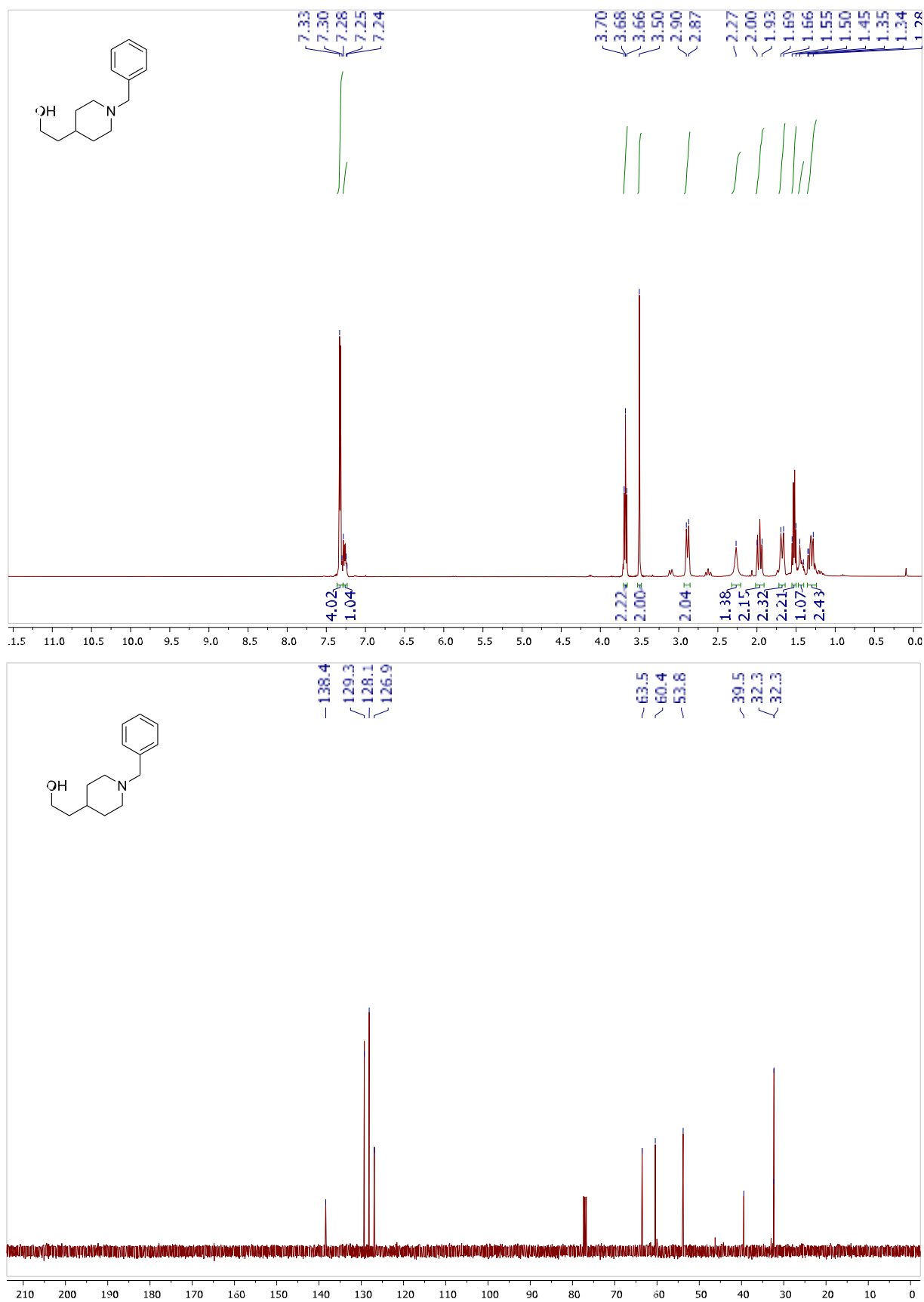
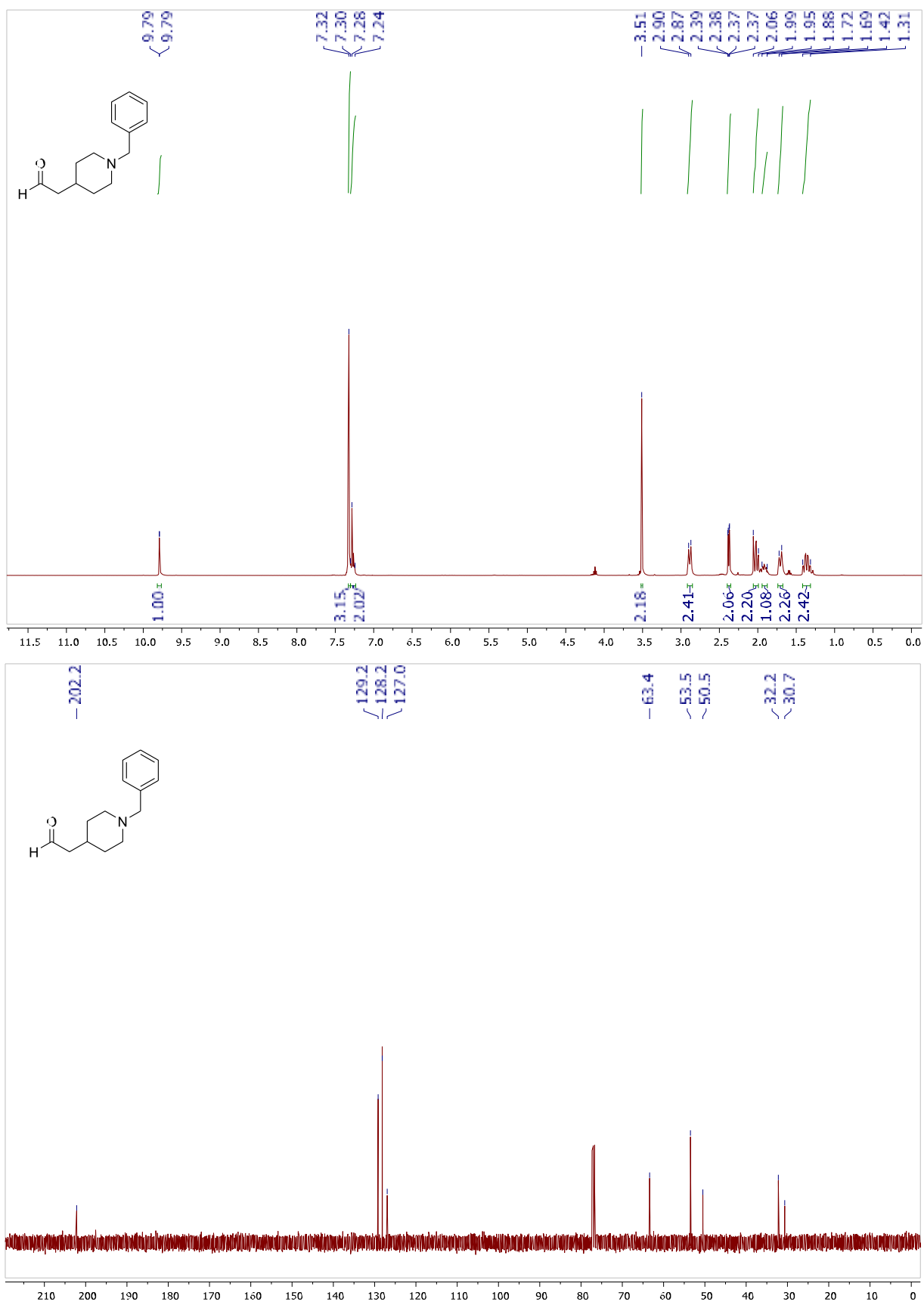


Figure S36. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.4.



**Figure S37.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.5.

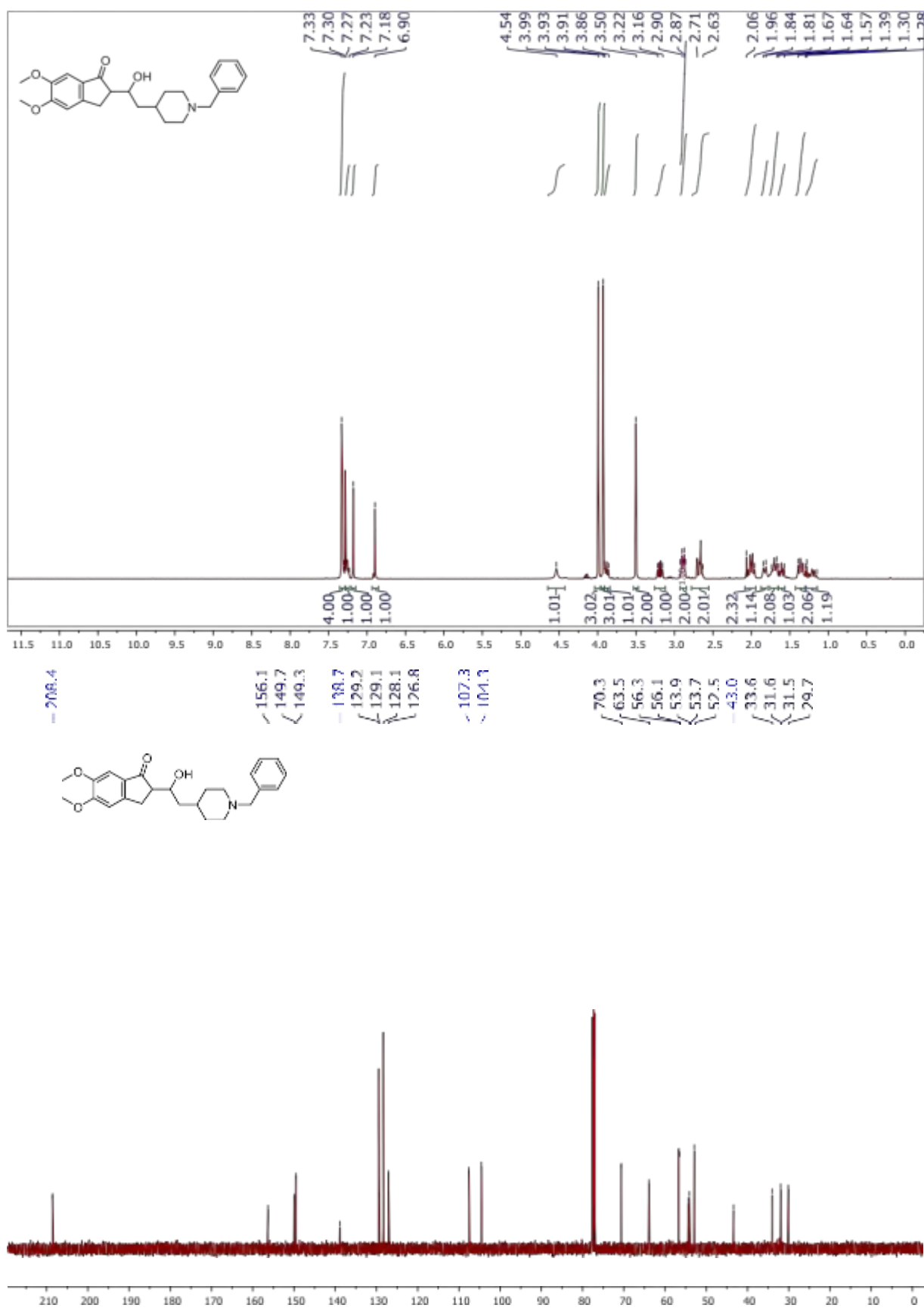


Figure S38.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.2.6.

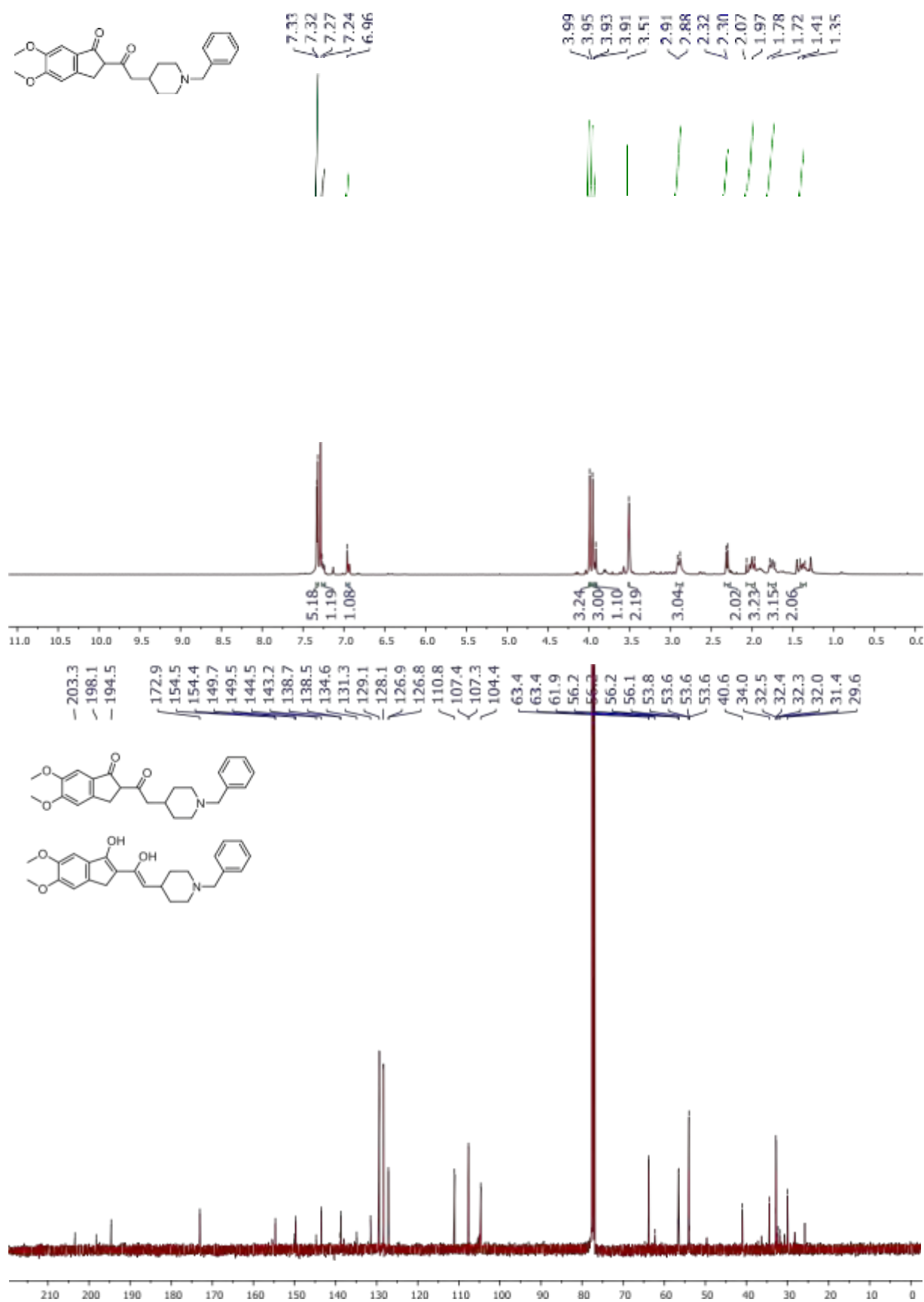
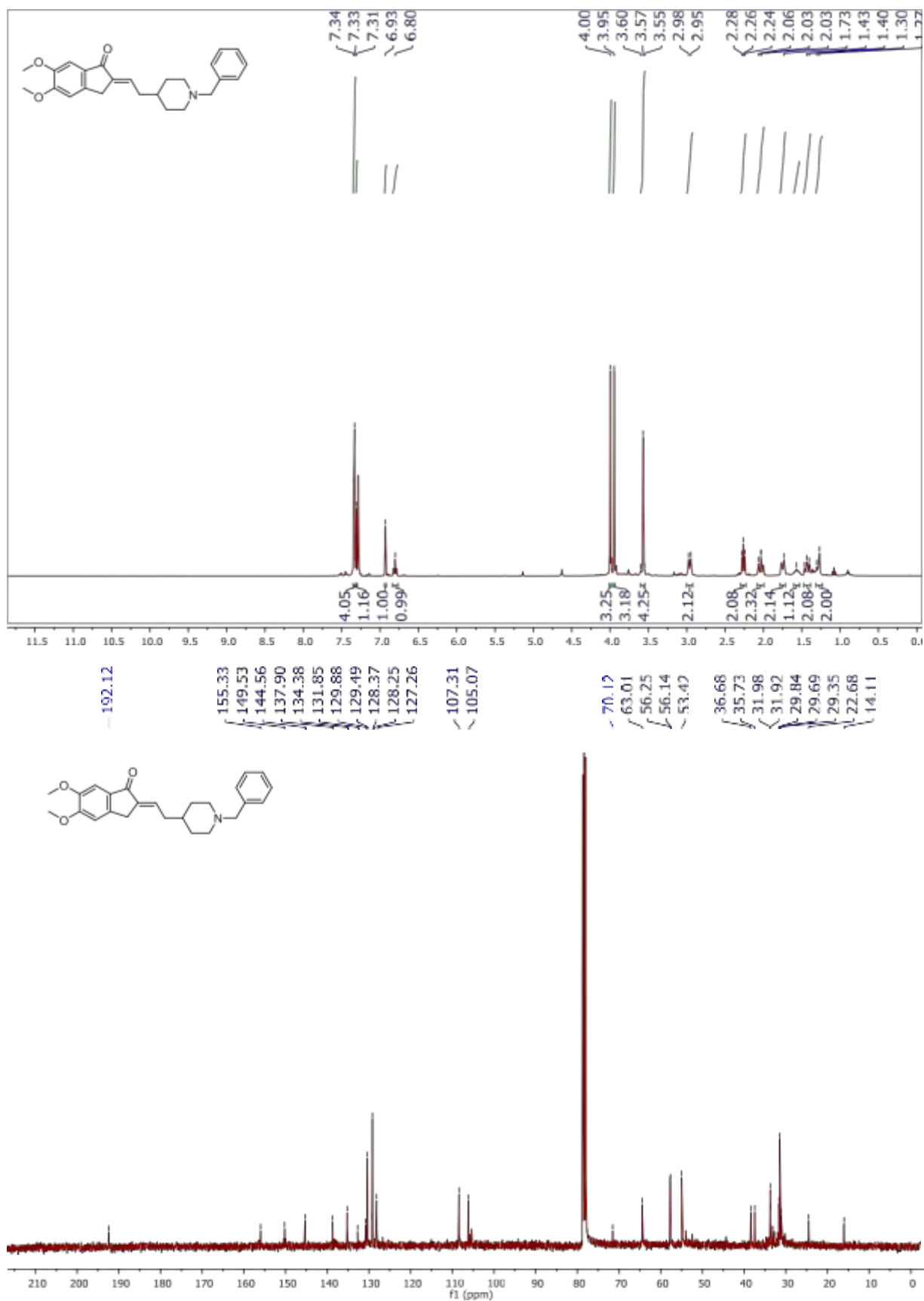


Figure S39.  $^1\text{H}$  NMR (top) and  $^{13}\text{C}$  NMR (bottom) spectra of compound S6.2.7.



**Figure S40.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.8.

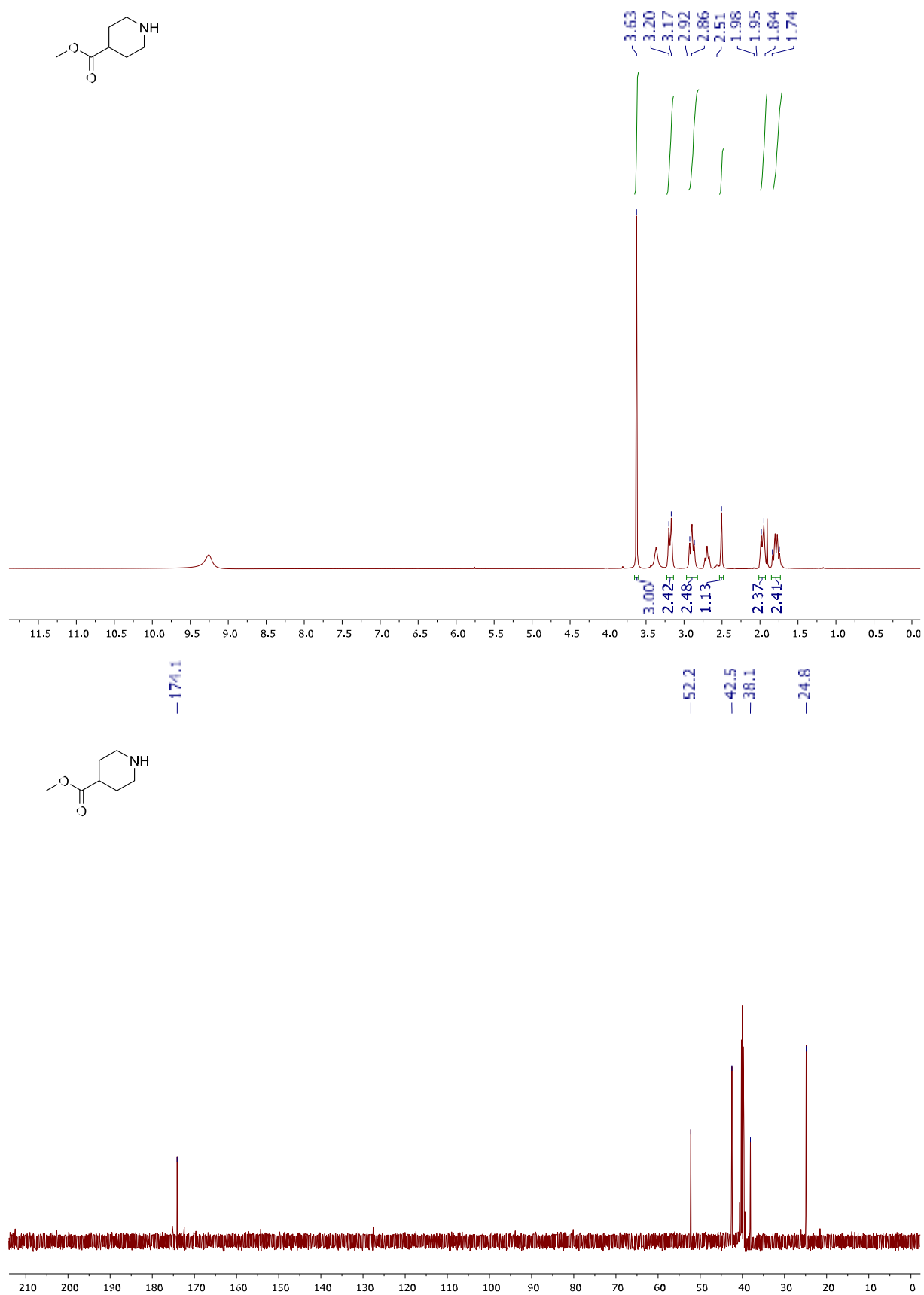
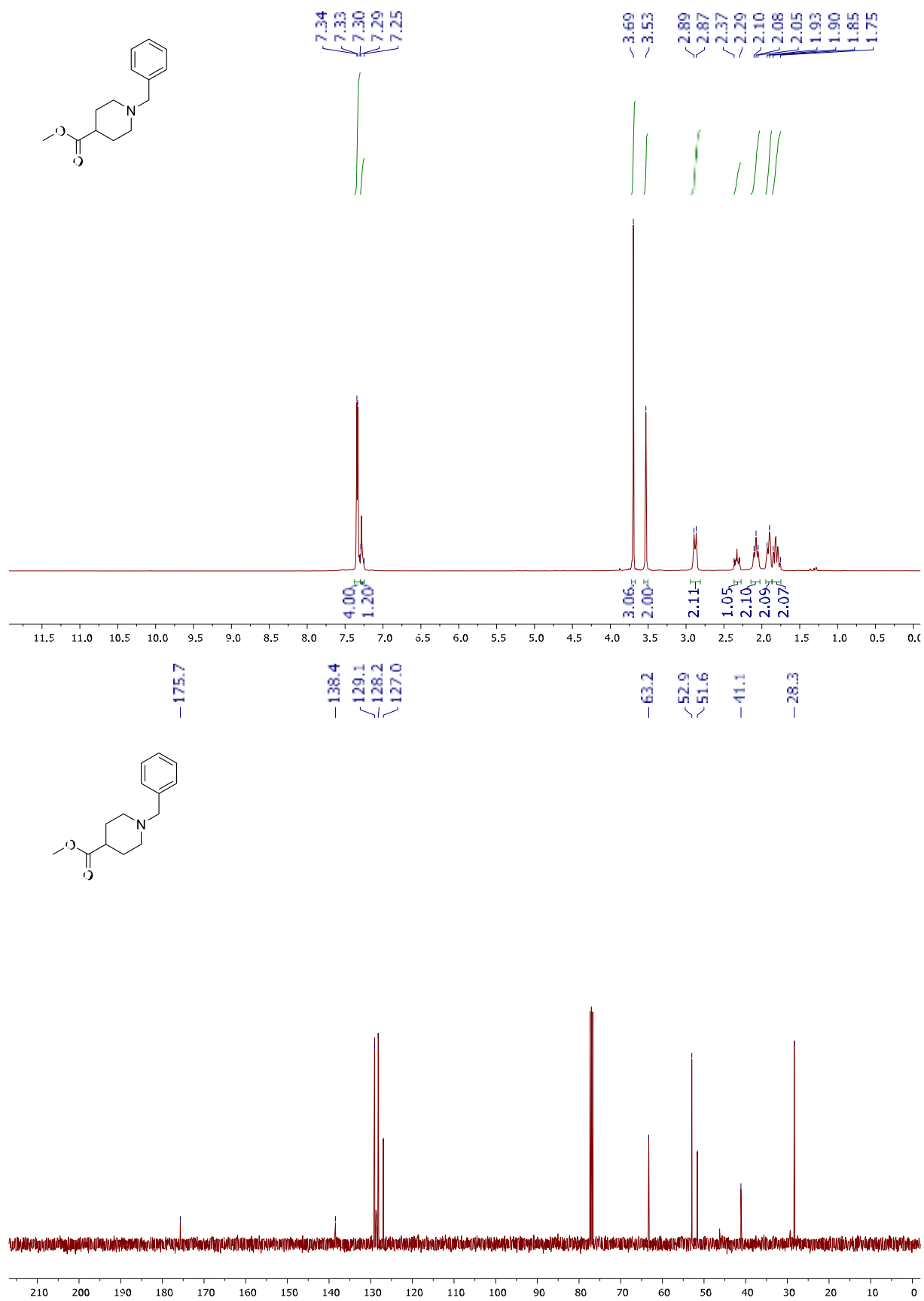
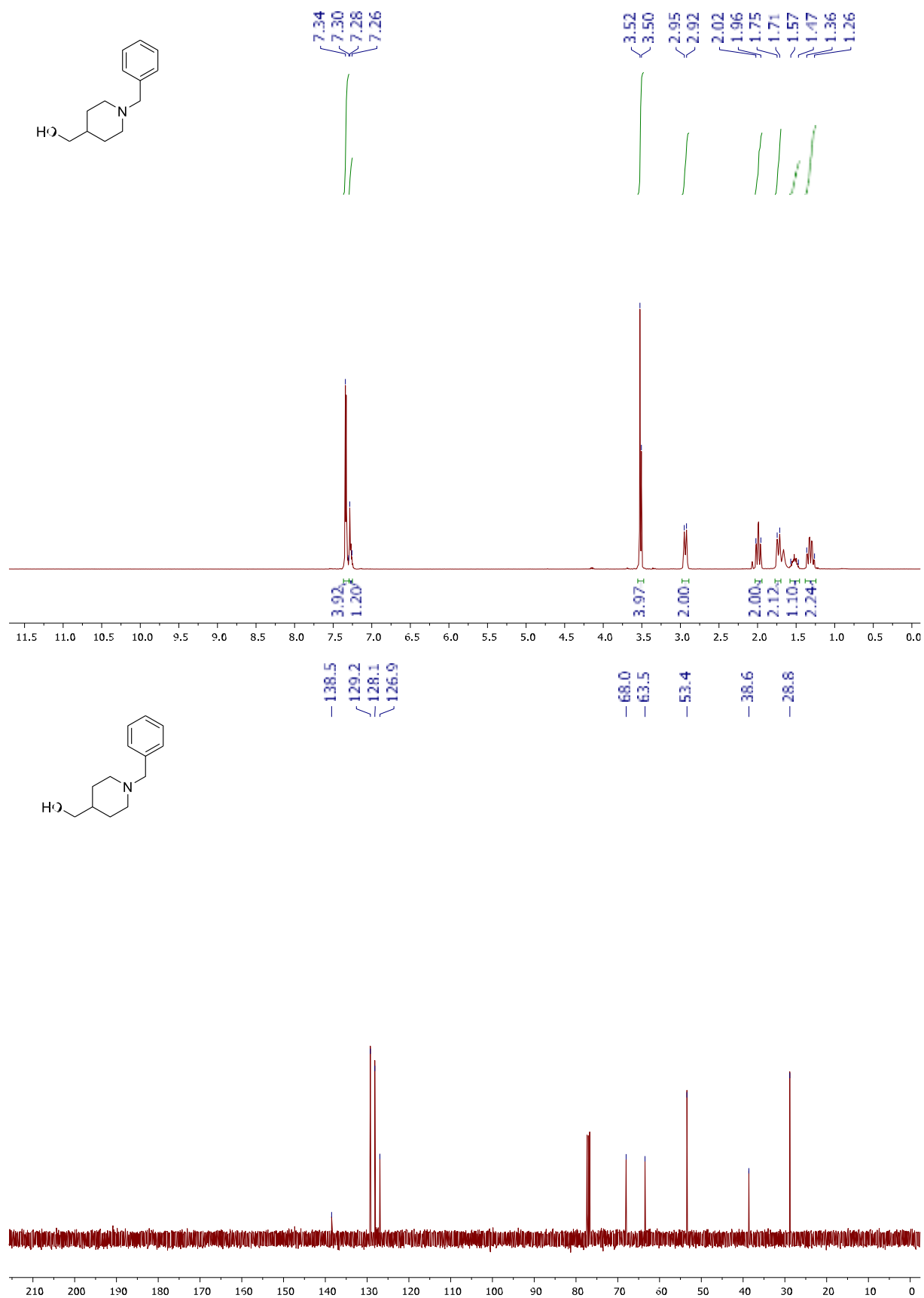


Figure S41. <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.9.

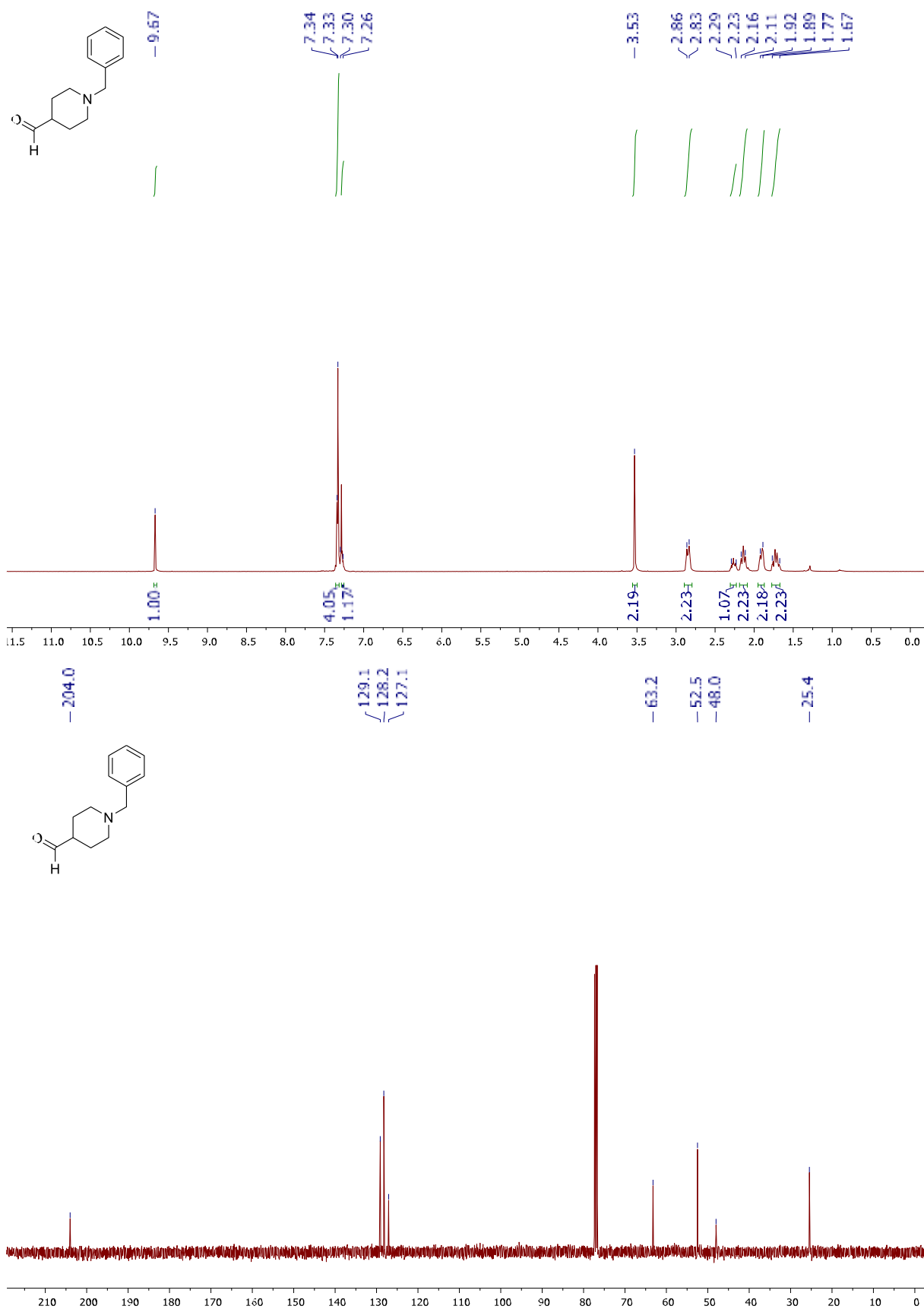




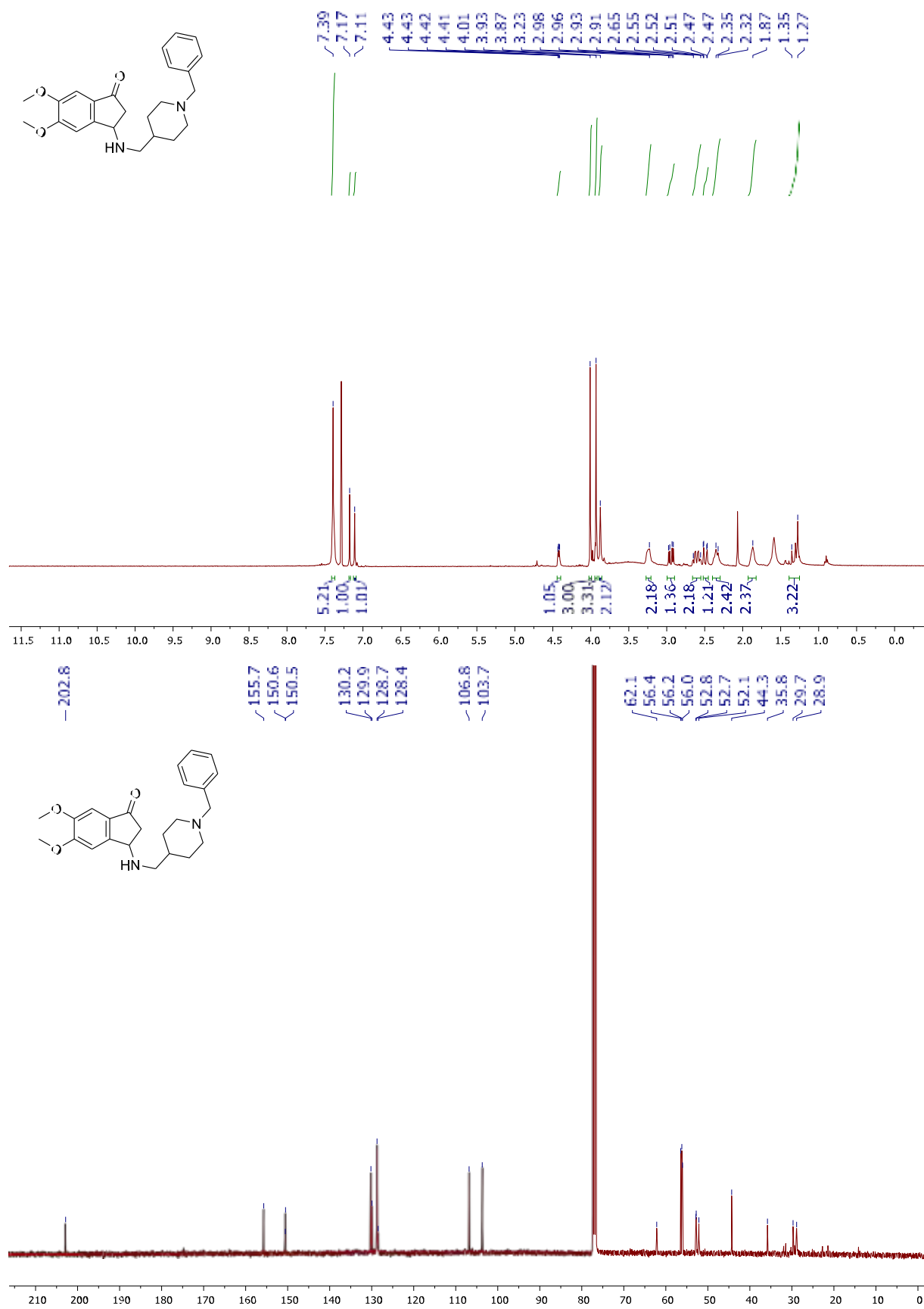
**Figure S42.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.10.



**Figure S43.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.11.



**Figure S44.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.12.



**Figure S45.** <sup>1</sup>H NMR (top) and <sup>13</sup>C NMR (bottom) spectra of compound S6.2.13.

## Section S7. Tutorial for reaction rule coding.

All synthetic analyses presented in the Manuscript were performed using *Analogs* module of the Allchemy software. This module uses two distinct components:

- i. retrosynthetic analysis of target replicas to obtain appropriate building blocks;
- ii. forward-synthesis analysis commencing from the obtained building blocks (along with added simple chemicals) to generate the structural analogs of the parent molecule.

The first component is limited to 180 reaction rules, covering chemistries most commonly used in medicinal chemistry. In contrast, the second component utilizes the entire collection of reaction rules available in Allchemy, totaling 25,307 rules. As in all previous works involving the Allchemy software, these rules are expert-coded (rather than machine extracted) to minimize the incorrect predictions in both retrosynthetic and forward-synthesis analyses. In this Section of Supporting Information, we provide general guidelines related to the coding of such reaction rules, and illustrate these guidelines with examples of reaction rules used for generation of Ketoprofen's and Donepezil's analogues.

The coding of each reaction rule begins with a careful examination of the underlying reaction mechanism, aiming to generalize the rule beyond published precedent(s). At this stage, the strict reaction "core" (the atoms that change their environments) and the admissible substituents flanking this core are determined. The scope of the admissible substituents must account for appropriate electronic (e.g., the presence of electron-withdrawing groups in CH-alkylations or aldol-type condensations) and steric (e.g., the lack of steric hindrance in  $\text{S}_{\text{N}}2$ -type substitutions) environments to minimize incorrect predictions. The reaction template is then written in the SMARTS notation (<https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>) which is a machine-readable notation representing molecules and reactions as alphanumeric strings. This notation allows for defining the lists of admissible substituents and incorporation of full atom mapping across the reaction. The reaction template is written in the forward direction, with set of substrates on the left side of the reaction arrow and main product and byproducts on the right side of the reaction arrow (*Substrate1.Substrate2.Substrate3*>>*MainProd.Byprod1.Byprod2*). Additionally, the same reaction template is written in the retrosynthetic direction (with main product and byproducts on the left side of reaction arrow and set of substrates on the right side of reaction arrow) to be available for both (retro and forward-synthesis) modalities available in Allchemy.

In the next step, the list of functional groups beyond that are outside of the reaction core and can interfere with expected reaction outcome needs to be defined. At this stage, the reaction conditions (e.g., the presence of strong acids or bases, reducing or oxidizing reagents, or the presence of water or other nucleophilic solvents) and reacting partners (nucleophilic, electrophilic or radical species) are carefully evaluated. Without this list of incompatible structural fragments, the software would predict implausible reaction outcomes which would fail in the experimental validation due to cross-reactivity, non-selectivity or degradation of unstable substrates (e.g., due to hydrolysis) under the conditions necessary to perform the given reaction.

In the third step, the additional reaction details are provided. Here, the reaction conditions are classified with respect to:

- i. general conditions (strongly acidic, acidic, neutral, basic, strongly basic, Lewis Acid);
- ii. temperature (very low, low, room temperature (rt), high, very high);
- iii. solvent classification (polar/nonpolar and protic/aprotic).

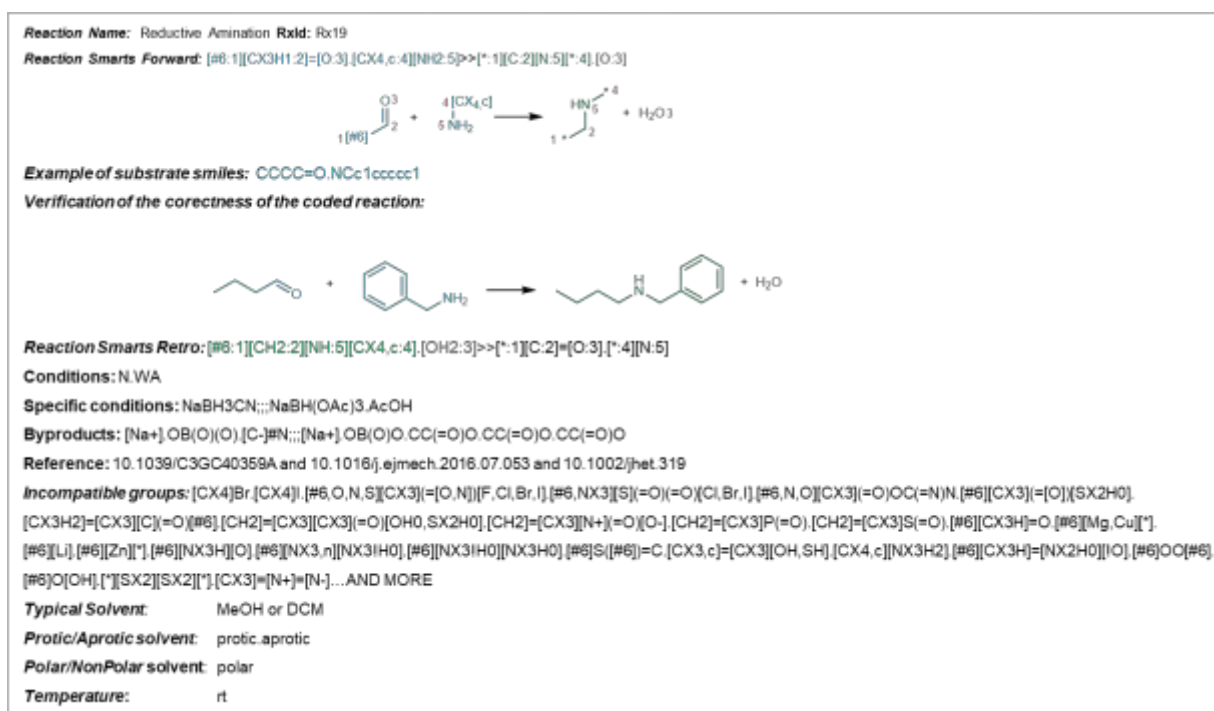
These classifications enable running the analyses avoiding unwanted reaction conditions, e.g., requiring extreme (cryogenic or very high) temperatures. These considerations can be included in the search settings as shown in the topmost part of Figure S1. Finally, the additional reaction metadata such as typical reaction conditions, literature reference illustrating a given type of chemistry, reaction name, or reaction byproducts (derived from reagents used) are added. All these data are available to the user and are displayed upon entering synthetic details (**Figure S7**).

With these general guidelines in mind, we will discuss in detail two examples of reaction rules used for disconnecting the parent Ketoprofen and Donepezil targets (and their structural replicas) and reassembling their structural analogues.

### **Example 1: Reductive amination.**

This type of chemistry allows for the preparation of secondary amines from primary amines and aldehydes. From a mechanistic point of view, this one-pot, two-stage reaction proceeds via

- i. formation of an imine via addition of an amine to an aldehyde and elimination of water and
- ii. reduction of the imine via addition of the hydride delivered by a borohydride reducing agent.



**Figure S46.** Coding of reaction rules used in the *Analogs* module. Example of a reaction rule for reductive amination used for preparation of Donepezil analogue **16**.

With reference to **Figure S46**, we begin coding the reaction template in forward direction by defining the set of substrates participating in this reaction. First, we specify the aldehyde starting material with atoms #1-#3. Here, the reaction core (atoms which change their environments) is defined with atoms #2 and #3 and represents the monosubstituted carbonyl group of an aldehyde which is written as [CX3H1:2]=[O:3] in SMARTS notation. The flanking substituent (#1) can be either aromatic or aliphatic sp<sup>3</sup>, sp<sup>2</sup> or sp hybridized carbon atom, which is written as [#6:1] in SMARTS notation with #6 denoting carbon's atomic number. Next, we define the primary amine participating in the reductive amination reaction with atoms #4 and #5. The amino group of primary amine with two hydrogen atoms attached is encoded as [NH2:5] in SMARTS notation, while the flanking substituent (#4) can be either an aliphatic or an aromatic carbon atom which is written as [CX4,c:4] in SMARTS. No additional constraints are added to aliphatic carbons at position #4 because this type of chemistry is validated even for very hindered amines with nitrogen atom attached to a quaternary carbon. With all substrates thus coded in SMARTS notation, we move to the product side of the reaction and start coding the reaction product and byproduct templates. As the software uses the substrates defined on the left side of reaction arrow to check applicability of a given reaction transform, the right side do not require the precise specification of substituents. Here, we code the template of the secondary amine with atoms #1,#2,#4,#5 as [\*:1][C:2][N:5][\*:4] and water byproduct with atom #3. Importantly,

all atoms which were lists or pseudolists such as [#6:1] defining various types of atoms (here, both aliphatic and aromatic) are written as “dummy” atoms denoted with stars. In this example, we use this “dummy” atom notation to code substituents flanking the carbonyl group of the aldehyde ([\*:1]) and the amine ([\*:4]). Finally, we rewrite the reaction template in the retro direction – now, with main reaction product (*green*) and byproducts (*grey*) on the left side of reaction arrow and substrate set (*blue*) on the right side of reaction arrow. In this example, the secondary amine main product is SMARTS-encoded as [#6:1][CH2:2][NH:5][CX4,c:4] while the water byproduct is written as [OH2:3].

In the next step, we provide an example of substrates for which a given reaction template can be applied and verify the correctness of coded reaction templates by i) running the forward template for these substrates and ii) running the retro template with products generated by the forward template – if both templates are correct we “regenerate” the original set of substrates. This tests are performed within the freely available RDKit (<https://www.rdkit.org/>) cheminformatics package.

After validation of the coded reaction templates, we move to the specification of functional groups present outside the reaction core and interfering with the expected reaction outcome. In this particular case, the reaction requires borohydride reducing agent and proceeds between an aldehyde electrophile and amine nucleophile. Accordingly, the list of competing functional groups includes:

1. strongly electrophilic groups which may react with amine nucleophile, e.g., alkyl bromides ('[CX4]Br'), alkyl iodides ('[CX4]I'), acyl halides ('[#6,O,N,S][CX3](=[O,N])[F,Cl,Br,I)'), sulfonyl halides ('[#6,NX3][S](=O)(=O)[Cl,Br,I)'), activated carboxylic acids ('[#6,N,O][CX3](=O)OC(=N)N'), thioesters ('[#6][CX3](=[O])[SX2H0]'),  $\beta$ -unsubstituted Michael acceptors ('[CX3H2]=[CX3][C](=O)[#6]', '[CH2]=[CX3][CX3](=O)[OH0,SX2H0]', '[CH2]=[CX3][N+](=O)[O-]', '[CH2]=[CX3]P(=O)', '[CH2]=[CX3]S(=O)'), and competitive aldehydes ('[#6][CX3H]=O')
2. strongly nucleophilic groups which may react with aldehyde substrate and/or imine intermediate, e.g., organomagnesium, organolithium, organozinc and organocuprates compounds ('[#6][Mg,Cu][\*]', '[#6][Li]', '[#6][Zn][\*]'), NH-hydroxylamines ('[#6][NX3H][O]'), NH-hydrazines ('[#6][NX3,n][NX3!H0]', '[#6][NX3!H0][NX3H0]'), ylides ('[#6]S([#6])=C'), enols and thioenols ('[CX3,c]=[CX3][OH,SH]'), and competitive primary amines ('[CX4,c][NX3H2]')



3. functional groups prone to reduction, e.g., aldimines ([\*][CX3H]=[NX2H0][!O]), peroxides ([\*]OO[\*]), hydroperoxides ([\*]O[OH]), disulfides ([\*][SX2][SX2][\*]), diazo ([CX3]=[N+]=[N-])

In the next step, we classify the reaction conditions (NaBH3CN or NaBH(OAc)3.AcOH) – here, the general reaction conditions are classified as ‘neutral or weakly acidic’ (*N.WA*), the temperature is classified as room temperature (*rt*) while the typical solvent (MeOH or DCM) is classified as *polar, protic* or *aprotic*.

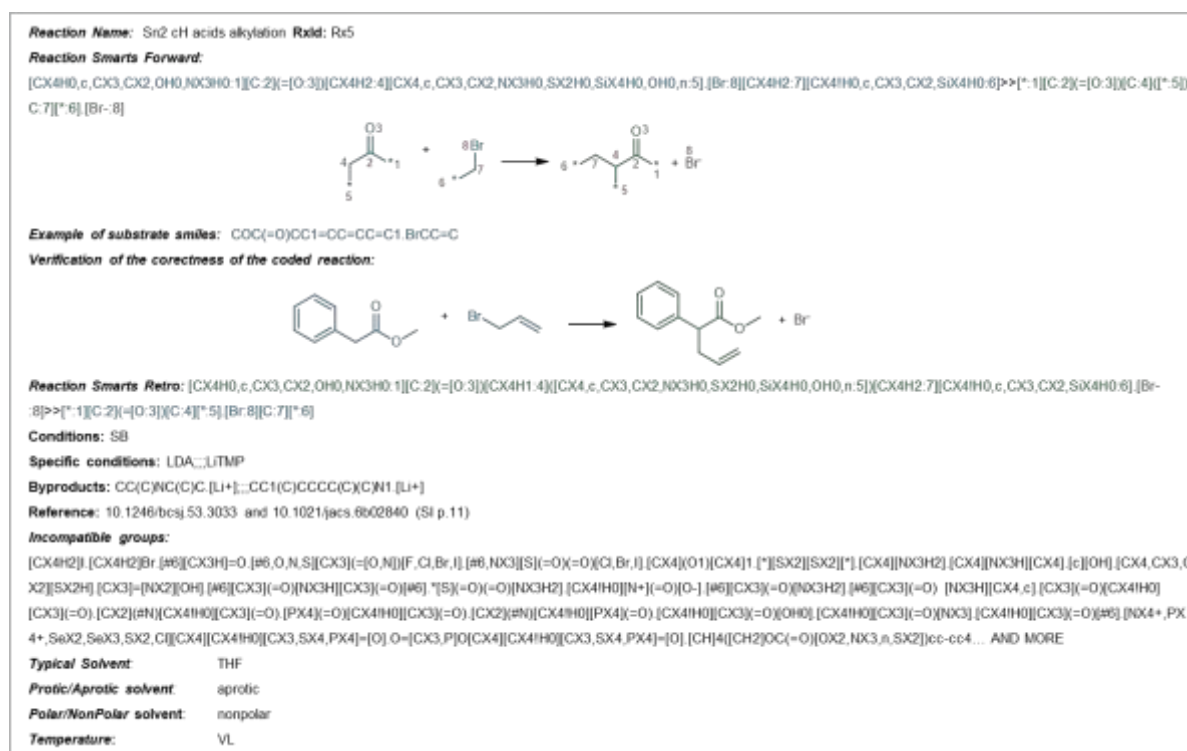
Finally, the “Reaction name” (“*Reductive Amination*”), reaction identifier (RxId, “*Rx19*”) reaction byproducts generated from reagents ([Na+].OB(O)(O).[C-]#N from NaBH3CN and [Na+].OB(O)O.CC(=O)O.CC(=O)O.CC(=O)O from NaBH(OAc)3) and references to illustrative examples of application of such reductive aminations are added.

### Example 2: C-H alkylation

In this example, we illustrate coding of a reaction rule covering the alkylation of enolates with primary alkyl bromides. This reaction transform allows for the preparation of functionalized ketones, esters or amides and was the key step in reassembling Ketoprofen’s analogues. From the mechanistic point of view, this one-pot two-stage reaction is carried out by i) deprotonation and enolisation of C-H acids and ii)  $S_N2$ -type substitution of the obtained enolate nucleophile with alkyl bromide electrophile. Accordingly, the coded reaction template must account for the presence of appropriate electron withdrawing groups (enabling the deprotonation and formation of the enolate) and define the steric environment surrounding the electrophilic center of primary alkyl bromide (excluding unreactive and sterically hindered neopentyl-type electrophiles).

With reference to **Figure S47**, we start coding the reaction template by defining the participating nucleophile with atoms #1-#5. First, we define the strict reaction core – here, atom #4 changes its environment and is limited to sterically unhindered aliphatic carbon atom with two hydrogens attached, which is written as [CX4H2:4] in SMARTS notation. Next, we move to the specification of electron withdrawing groups acidifying the #4 positions. In this particular reaction template, the allowed groups are limited to carboxylic acid esters, amides or ketones. The carbonyl group of each of these fragments is coded on atoms #2 and #3 and written as [C:2](=[O:3]) in SMARTS notation while atom #1 defines the flanking alkoxy-, amino- or carbon fragment behind the carbonyl group and is coded as [CX4H0,c,CX3,CX2,OH0,NX3H0:1]. The alkoxy- fragment of carboxylic acid esters is coded as “OH0” – oxygen atom with no hydrogens attached. The amino fragment of amides is limited

to amides with no hydrogens at nitrogen atom because N-H amides are preferentially deprotonated and alkylated at nitrogen rather than at  $\alpha$ -CH position. This requirement is written as 'NX3H0' is SMARTS notation representing sp<sup>3</sup>-hybridized nitrogen atom with no hydrogens attached. Finally, the ketones are allowed to have acidic hydrogens at only a single  $\alpha$ -position #4 to avoid non-selective reaction outcomes. Thus, position #1 of ketones is limited to alkyl carbons with no attached hydrogens ('CX4H0'), aromatic ('c') or aliphatic sp<sup>2</sup> ('CX3') or sp ('CX2') hybridized carbons. Finally, we define the list of substituents at position #5 flanking the reaction center. Here, we allow for the aliphatic ('CX4', 'CX3', 'CX2') or aromatic ('c') carbons, n-bound heterocycles ('n') and nitrogen, sulfur, oxygen or silicon heteroatoms with no attached hydrogens ('NX3H0', 'SX2H0', 'OX2H0', 'SiX4H0'). We note the alkylation of C-H acids with two electron withdrawing groups (e.g., ketoesters, ketoamides or diketones) is coded in the separate reaction rule because such alkylations can be performed under much milder conditions (with no strong base involved) and their scope of incompatible groups is quite different.



**Figure S47.** Coding of reaction rules used in the *Analogs* module. Example of a reaction rule for the alkylation of CH acids used here in the preparation of Ketoprofen's analogues.

Next, we continue the coding of reaction template by defining the alkyl bromide electrophile with atoms #6-#8. First, we define the strict reaction core: bromide leaving group ([Br:8]) and

the carbon atom of primary alkyl bromide involved in S<sub>N</sub>2-type displacement (#7), the latter limited to aliphatic carbon atom with two hydrogens attached ([CX4H2:7]). The neighboring position #6 ([CX4!H0,c,CX3,CX2,SiX4H0:6]) allows for the presence of aromatic ('c') or aliphatic sp<sup>2</sup> ('CX3') or sp ('CX2') hybridized carbons (defining highly reactive benzyl, allyl and propargyl bromides as electrophiles) or silicon atom. Aliphatic, sp<sup>3</sup>-hybridized carbons are also allowed but are required to have at least one hydrogen attached ([CX4!H0]) excluding sterically hindered and unreactive neopentyl-type electrophiles. With all substrates thus defined, we move to the coding of the main reaction product (*green* in **Figure S47**) with atoms #1-#7 and the bromide leaving group ([Br-:8], *grey* in **Figure S47**). As in the previous example, we replace all lists of substituents (here, at #1, #5 and #6) with 'dummy' atoms ([\*:1], [\*:5] and [\*:6]). Finally, we provide the example of substrates for which this reaction rule can be applied, rewrite the forward reaction template in retro direction with the main reaction product coded as [CX4H0,c,CX3,CX2,OH0,NX3H0:1][C:2](=[O:3])[CX4H1:4]([CX4,c,CX3,CX2,NX3H0,SX2H0,SiX4H0,OH0,n:5])[CX4H2:7][CX4!H0,c,CX3,CX2,SiX4H0:6], and validate both reaction templates in the *rdkit* package.

In the next step, we define the list of incompatible functional groups. The reaction proceeds between the alkyl bromide electrophile and enolate nucleophile, the latter generated with strong, non-nucleophilic base such as LDA or LiTMP. Accordingly, the list of functional groups interfering with the expected reaction outcome contains:

- 1) strongly electrophilic groups which may react with the enolate nucleophile, e.g., alkyl iodides ('[CX4H2]I') and bromides ('[CX4H2]Br'), aldehydes ('[#6][CX3H]=O'), acyl halides ('[#6,O,N,S][CX3](=[O,N])[F,Cl,Br,I]'), sulfonyl halides ('[#6,NX3][S](=O)(=O)[Cl,Br,I]'), epoxides ('[CX4](O1)[CX4]1'), disulfides ('[\*][SX2][SX2][\*]')
- 2) nucleophilic groups which may react (especially, after deprotonation) with alkyl bromide electrophile, e.g., aliphatic amines ('[CX4][NX3H2]' and '[CX4][NX3H][CX4]'), phenols ('[c][OH]'), thiols ('[CX4,CX3,CX2][SX2H]'), oximes ('[CX3]=[NX2][OH]'), imides ('[#6][CX3](=O)[NX3H][CX3](=O)[#6]'), sulfonamides ('\*[S](=O)(=O)[NX3H2]')
- 3) acidic groups which may interfere with the deprotonation and formation of enolate, e.g., nitroalkanes ([CX4!H0][N+](=O)[O-]), amides ('[#6][CX3](=O)[NX3H2]' and '[#6][CX3](=O)[NX3H][CX4,c]'), active methylene compounds ('[CX3](=O)[CX4!H0][CX3](=O)', '[CX2](#N)[CX4!H0][CX3](=O)', '[PX4](=O)[CX4!H0][CX3](=O)', '[CX2](#N)[CX4!H0][PX4](=O)') and competing enolizable esters ([CX4!H0]

[CX3](=O)[OH0]), amides ([CX4!H0][CX3](=O)[NX3]) and ketones ([CX4!H0][CX3](=O)[#6])

- 4) groups unstable under basic conditions, e.g., fragments prone to  $\beta$ -elimination ('[NX4+,PX4+,SeX2,SeX3,SX2,Cl][CX4][CX4!H0][CX3,SX4,PX4]=[O]' and 'O=[CX3,P]O[CX4][CX4!H0][CX3,SX4,PX4]=[O]') or Fmoc protecting group ('[CH]4([CH2]OC(=O)[OX2,NX3,n,SX2])cc-cc4').

In the next step, we classify the reaction conditions (LDA or LiTMP) – here, the general reaction conditions are classified as 'strongly basic' (*SB*) due to presence of LDA or LiTMP bases used for deprotonation (with pKa's ~36), the temperature at which such alkylations are usually performed (-78°C) is classified as very low (*VL*) while the typical solvent (THF) is classified as *nonpolar, aprotic*.

Finally, the "Reaction name" ("*Sn2 cH acids alkylation*"), reaction identifier (RxId, "Rx5"), reaction byproducts generated from reagents ('CC(C)NC(C)C.[Li+]' from LDA and 'CC1(C)CCCC(C)(C)N1.[Li+]' from LiTMP) and references to illustrative examples of application of such alkylations are added.

## Section S8. Supplementary References

- S1. G. Maggiora, M. Vogt, D. Stumpfe and J. Bajorath, *J. Med. Chem.*, 2013, **57**, 3186-3204.
- S2. D. Stumpfe, H. Hu and J. Bajorath, *J. Comp. Aid. Mol. Des.*, 2020, **34**, 929-942.
- S3. L. M. Lima and E. J. Barreiro, *Current Medicinal Chemistry*, 2005, **12**, 23-49.
- S4. Y. C. Martin, J. L. Kofron and L. M. Traphagen, *J. Med. Chem.*, 2002, **45**, 4350-4358.
- S5. K. V. Dileep, K. Ihara, C. Mishima-Tsumagari, M. Kukimoto-Niino, M. Yonemochi, K. Hanada, M. Shirouzu and K. Zhang, *Int. J. Biol. Macromol.*, 2022, **210**, 172-181.
- S6. S. Asghar, N. Mushtaq, A. Ahmed, L. Anwar, R. Munawar and S. Akhtar, *Molecules*, 2024, **29**, 490.
- S7. K. T. Kumar, B. Gorain, D. K. Roy, Zothanpuia, S. K. Samanta, M. Pal, P. Biswas, A. Roy, D. Adhikari, S. Karmakar and T. Sen, *J. Ethnopharmacol.*, 2008, **120**, 7-12.
- S8. C. Yu, R. Huang and F. W. Patureau, *Angew. Chem. Int. Ed.*, 2022, **61**, e202201142.
- S9. D. Liu, L. Zhang, J. Cheng, Q. Wei, Z. Jia and F. E. Chen, *Green Chem.*, 2024, **26**, 9690-9696.
- S10. J. Yan, J. Hu, A. Liu, L. He, X. Li and H. Wei, *Bioorg. Med. Chem.*, 2017, **25**, 2946-2955.
- S11. J. Alfaro-Lopez, T. Okayama, K. Hosohata, P. Davis, F. Porreca, H. I. Yamamura and V. J. Hruby, *J. Med. Chem.*, 1999, **42**, 5359-5368.
- S12. B. Guieu, C. Lecoutey, R. Legay, A. Davis, J. S. Oliveira-Santos, C. D. Altomare, M. Catto, C. Rochais and P. Dallemagne, *Molecules*, 2021, **26**, 80

**9. STATEMENTS OF CONTRIBUTIONS**

Warsaw – May 19, 2025

## Statement of Contribution

I, Ahmad Makkawi, declare that my contribution in the following articles:

Roszak, R.; Gadina, L.; Wolos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Performing the syntheses described Fig. 3 in the main text, including the identification and characterization of the products and side products
- Contributing in writing parts of the SI and the main text

Makkawi, A.; Wiktor, B.; Wolos, A.; Manna, S.; Roszak, R.; Szymkuć, S.; Moskal, M.; Koshevnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Performing the syntheses described in the main text, including the identification and characterization of the products
- Performing the binding assays for all compounds
- Performing the docking studies using only AutoDock
- Contributing in writing and drawing parts of the SI and the main text
- Combining and formatting the SI and the main text

Ahmad Makkawi

Signature:



Bartosz A. Grzybowski  
Distinguished Professor of Chemistry and  
Director  
Center for Algorithmic and Roboticized Synthesis, CARE  
Institute for Basic Science, UNIST Campus  
Ulsan, Republic of Korea



## Statement of Contribution

I, Prof. B. A. Grzybowski, declare that my contribution in the following articles:

Rozzak, R.; Gadina, L.; Wolos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Designing and developing Allchemy platform
- Supervising the project
- Writing the paper with help from other authors

Makkawi, A.; Wiktor, B.; Wolos, A.; Manna, S.; Rozzak, R.; Szymkuc, S.; Moskał, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, **2025**, DOI: 10.1039/D5SC00070J.

Consisted of:

- Designing and developing Allchemy platform
- Supervising the project
- Writing the paper with help from other authors

I hereby confirm the abovementioned contributions.

A handwritten signature in black ink, appearing to be 'B. Grzybowski', written over a light blue horizontal line.

Prof. Dr. Bartosz A. Grzybowski, June 9, 2025



Warsaw – May 14, 2025

## Statement of Contribution

I, Prof. D. T. Gryko, declare that my contribution in the following article:

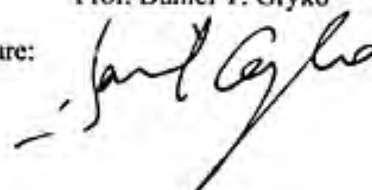
Rozzak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Helping with the evaluation of the kinetic networks.

Prof. Daniel T. Gryko

Signature:



Warsaw – May 14, 2025

## Statement of Contribution

I, Prof. J. Mlynarski, declare that my contribution in the following article:

Rozzak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Supervising the syntheses described in Supplementary Section S2, performed by S. Baś.

Prof. Jacek Mlynarski

Signature:



Warsaw – May 15, 2025

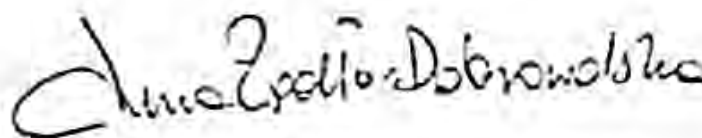
## Statement of Contribution

I, Anna Zadło-Dobrowolska, declare that my contribution in the following article:

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuc, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- the docking studies
- Supervising the syntheses
- Performing the computational calculations
- Contributing in writing of the SI and the main text



Anna Zadło-Dobrowolska

Signature:

Warsaw – May 14, 2025

## Statement of Contribution

I, Sara Szymkuć, declare that my contribution in the following articles:

Rozzak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Developing the MECH module within the Alchemy platform
- Performing analyses and calculations described in the paper
- Contributing in writing some parts in the main article

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Rozzak, R.; Szymkuć, S.; Moskal, M.; Koshevnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms of the ANALOGS module

Sara Szymkuć

Signature:



Warsaw – May 14, 2025

## Statement of Contribution

I, Martyna Moskal, declare that my contribution in the following articles:

Rozsak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołbiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Młynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

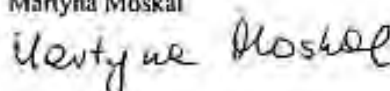
Consisted of:

- Developing the MECH module within the Allchemy platform
- Performing analyses and calculations described in the paper

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Rozsak, R.; Szymkuć, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOGS module

Martyna Moskal  
Signature: 

Warsaw – May 14, 2025

## Statement of Contribution

I, Tomasz Klucznik, declare that my contribution in the following article:

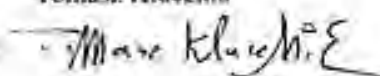
Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Developing the MECH module within the Alchemy platform
- Performing syntheses described in Fig. 4b in the main article

Tomasz Klucznik

Signature:



Warsaw – May 14, 2025

## Statement of Contribution

I, Barbara Mikulak-Klucznik, declare that my contribution in the following article:

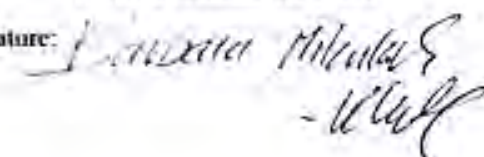
Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Developing the MECH module within the Alchemy platform
- Performing syntheses described in Fig. 4b in the main article

Barbara Mikulak-Klucznik

Signature:



The image shows a handwritten signature in black ink. The signature is written in a cursive style and appears to read 'Barbara Mikulak-Klucznik'. Below the main signature, there is a smaller, more stylized signature that looks like '- Klucznik'.

## Statement

I, Louis Gadina, hereby declare that my contribution in the following article:

Rozzak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Performing syntheses described in the main article, Fig. 4d as well as identifying and characterizing the resulting products
- Contributing to the creation of the dataset used for Figure 1C
- Performing the analysis presented in Figure 1C
- Contributing in writing parts of the SI and the main text

Louis Gadina

Signature:



14 05 2025



## Statement

I, Yasemin Bilgi, hereby declare that my contribution in the following article:

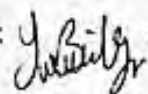
Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B., Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Performing syntheses described in the main article, Fig. 4a, along with P.G, as well as identifying and characterizing the resulting products
- Performing syntheses described in the main article, Fig. 4c as well as identifying and characterizing the resulting products
- Contributing to the creation of the dataset used for Figure 1C
- Contributing in writing parts of the SI and the main text

Yasemin Bilgi

Signature:



14.05.2025

Warsaw – May 15, 2025

## Statement of Contribution

I, Sabyasachi Manna, declare that my contribution in the following article:

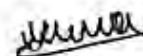
Makkawi, A.; Wiktor, B.; Wolos, A.; Manna, S.; Rószak, R.; Szymkuc, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Performing syntheses for some target compounds
- Contributing in writing some parts of the SI

Sabyasachi Manna

Signature:



Warsaw – May 15, 2025

## Statement of Contribution

I, Aleksei Koshevnikov, declare that my contribution in the following article:

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuc, S.; Moskał, M.; Koshevnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOG module

Aleksei Koshevnikov

Signature:



Warsaw - May 14, 2025

## Statement of Contribution

I, Karol Molga, declare that my contribution in the following articles:

Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Haś, S.; Frydrych, R.; Młynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Developing the MECII module within the Allchemy platform
- Performing analyses and calculations described in the paper
- Contributing in writing some parts of the SI and the main text

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuć, S.; Moskał, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, 2025, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOGS module

Karol Molga

Signature:



## Statement of Contribution

I, Agnieszka Wołos, declare that my contribution in the following articles:

Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Developing the MECH module within the Allchemy platform
- Performing analyses and calculations described in the paper
- Contributing in writing some parts of the SI and the main text

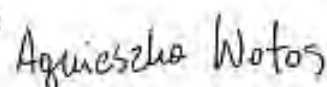
Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuć, S.; Moskał, M.; Koshevnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, **2025**, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOG module
- Contributing in writing some parts of the SI and the main text

Agnieszka Wołos

Signature:



Warsaw – May 14, 2025

## Statement of Contribution

I, Wiktor Beker, declare that my contribution in the following article:

Makkawi, A.; Wiktor, B.; Wolos, A.; Manna, S.; Roszak, R.; Szymkuc, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, **2025**, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOGS module

Wiktor Beker

Signature:



Warsaw – May 14, 2025

## Statement of Contribution

I, Rafal Roszak, declare that my contribution in the following articles:

Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Developing the MECH module within the Allchemy platform
- Performing analyses and calculations described in the paper

Makkawi, A.; Wiktor, B.; Wołos, A.; Manna, S.; Roszak, R.; Szymkuc, S.; Moskal, M.; Koshevarnikov, A.; Molga, K.; Zadło-Dobrowolska, A.; Grzybowski, B. A. Retro-forward synthesis design & experimental validation of potent structural analogs of known drugs. *Chem. Sci.*, **2025**, DOI: 10.1039/D5SC00070J.

Consisted of:

- Developing the algorithms and the WebApp of the ANALOG module



Rafal Roszak

Ulsan, South Korea – May 14, 2025

## Statement of Contribution

I, Rafał Frydrych, declare that my contribution in the following article:

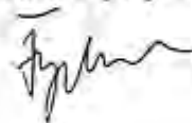
Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Młynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Helping in the syntheses described in Supplementary Section S2 with supervision from Prof. B. A. Grzybowski.

Rafał Frydrych

Signature:





May 14, 2025

## Statement of Contribution

I, Oskar Popik, declare that my contribution to the following article:

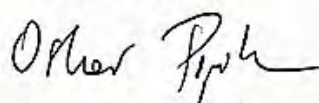
Rozzak, R.; Gadina, L.; Wolos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiewska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Performing syntheses described in Fig. 4a in the article, under the supervision of Prof. J. Mlynarski
- Contributing to the writing of the Supplementary Information section related to my synthesis

Oskar Popik

Signature:



## Statement of Contribution

I, Patrycja Gołębiowska, declare that my contribution in the following article:

Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Młynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Performing syntheses described in Figs. 4a and 4c in the main article
- Contributing in writing the parts in the SI related to my syntheses
- Helping in combining the SI

Patrycja Gołębiowska

Warsaw – May 14, 2025

Warsaw – May 14, 2025

## Statement of Contribution

I, Olena Vakuliuk, declare that my contribution in the following article:

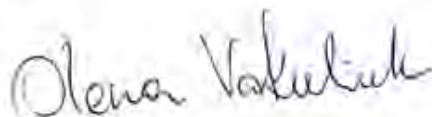
Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskał, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, 2024, 15, 10285.

Consisted of:

- Helping with the evaluation of the kinetic networks, along with Prof. D. T. Gryko.

Olena Vakuliuk

Signature:



Krakow – May 14, 2025

## Statement of Contribution

I, Sebastian Baś, declare that my contribution in the following article:

Roszak, R.; Gadina, L.; Wołos, A.; Makkawi, A.; Mikulak-Klucznik, B.; Bilgi, Y.; Molga, K.; Gołębiowska, P.; Popik, O.; Klucznik, T.; Szymkuć, S.; Moskal, M.; Baś, S.; Frydrych, R.; Mlynarski, J.; Vakuliuk, O.; Gryko, D. T.; and Grzybowski, B. A. Systematic, computational discovery of multicomponent and one-pot reactions. *Nature Communications*, **2024**, *15*, 10285.

Consisted of:

- Performing syntheses described in Supplementary Section S2 with supervision from Prof. J. Mlynarski.

Sebastian Baś

Signature:



Sebastian Baś  
18 05 2025  
0 14 08 10285