

Institute of Organic Chemistry
Polish Academy of Sciences

DOCTORAL DISSERTATION

Computational and Automated Approaches to Reaction Discovery and Chemical-Space Exploration

M.Sc. Yasemin Bilgi-Gadina

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

Supervisor: Prof. Bartosz A. Grzybowski

Warsaw, 2026

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)

and



“The Institute for Basic Science (IBS-R020-D1)”

Acknowledgements

I would like to express my deepest gratitude to my supervisor, **Prof. Bartosz A. Grzybowski**, for teaching me how a scientist should think. I am especially grateful for his scientific vision and for guiding me in developing meaningful ideas and seeking solutions to fundamental unanswered questions. Thanks to him, whenever I pursue a new idea, I have learned to ask myself, "*How many people in the world could this affect*?" His mentorship revealed to me the true beauty, purpose, and significance of science. Last but not least, I am deeply thankful to him for opening the door to a new world and giving me the opportunity to work in a highly advanced research laboratory in Ulsan, South Korea.

I would also like to thank all my colleagues from **Team XI** for their valuable collaboration.

I warmly thank my friends at ICHO after leaving my home country, for their warmth and kindness: **Ahmad Makkawi, Kitti Franciska Szabó, Jaqueline Stella Araujo Badarò, Raveena Jajoria, Parul Sahrawat, Sakshi Sareen.**

I would also like to acknowledge my colleagues in the **Grzybowski laboratory in South Korea** for creating the most inspiring and supportive working environment I have ever experienced. Their kindness, warmth, and scientific excellence made my time in Ulsan both professionally rewarding and personally unforgettable.

I would like to thank **Natalia Onishchenko** and **Diana Kolygina** from the bottom of my heart, who became my closest friends in South Korea and filled my time here with warmth, joy, and unforgettable memories. By taking me to so many places across the country, they helped me discover its beauty in a way I never could have experienced without them.

To my lifelong friends—my sisters and brothers not by blood, but by choice—who have always made my life better—**Tuğçe Alumur, Fulya Sarıefe, Müge Tezcan, Fikrican Dilek, İsmail Tahmaz, Müge Bilgen, Seda Erzurum, Mert Koç, Deniz Karataş-Cevizci, Murat Delman, Meltem Kutluer, Ece Yazıcı**—Thank you for your friendship, encouragement, and laughter, and for standing by my side through both the good times and the difficult ones. I owe special thanks to **Fikrican**, who helped me find this PhD position and directly influenced the course of my academic career. A special expression of gratitude goes to **Fikriye Sarıefe and Gülcan Soğancoğlu**, the mothers of two of my friends, for embracing me as their own daughter and for always caring for and supporting me.

To my family—**Gülfer Bilgi, Ferudun Bilgi, Christine Mouroux, Pascal Ramillon, Andrea Gadina**—thank you for your endless love, unwavering support, and belief in me. I am deeply grateful for everything you have done and continue to do for me.

Most importantly, I offer my most heartfelt gratitude to my husband, **Louis P.J-L. Gadina,** my greatest source of love and support. He entered my life as a colleague and became my family and the most important person in my life. Throughout this demanding journey, he stood beside me, believed in me during moments of doubt, celebrated every success, and walked every step of the way with me. His love, as strong and enduring as a carbon–carbon bond, together with his patience, encouragement, and understanding made this achievement possible. I am deeply grateful to have shared this journey with him. *Mon cœur, merci d'être mon chez-moi, ma force et mon plus beau voyage. Je t'aime, aujourd'hui et pour toujours.*

Abstract in English

My doctoral research addressed four connected challenges in modern synthetic chemistry: discovering new multicomponent and one-pot reactions using *Allchemy* software's *MECH* module; developing *AbSynth*, an open-access total-synthesis database, as a reliable, machine-readable collection of published synthetic knowledge for future computational research; robotically exploring multidimensional reaction spaces; and quantifying via automation large numbers of complex crude reaction mixtures.

In the first stage, as part of a broader team effort, I contributed to the experimental evaluation of *Allchemy-MECH*, which generates mechanistically plausible reaction pathways from selected starting materials. I first validated a computationally proposed one-pot sequence involving two Williamson ether syntheses connected by a [2,3]-Wittig rearrangement to afford branched diallylic ethers. One product was further converted by ring-closing metathesis into a synthetically useful cyclic enol ether. I then used *Allchemy-MECH* to identify conditionally divergent multicomponent pathways from allyl thiol, an aromatic aldehyde, and an allenyl ketone. These afforded hydroxylated monothio-β-diketones under DBU-mediated conditions and 2,3-dihydrothiophenes after addition of acetyl chloride; the MCR product was further converted into a substituted pyrazole, a common structural motif in many drugs. I developed the corresponding experimental procedures. These results – published in *Nature Communications*[1] – showed that mechanistic network analysis can propose experimentally applicable MCR and one-pot sequences while identifying reagent-compatibility constraints, the required order of addition, and condition-dependent divergence into different product classes.

In turn, the *AbSynth* project addressed the need for reliable, machine-readable data on complete total syntheses by assembling more than 3,000 published pathways as an open-access resource. I served as a co-corresponding author on this work (now published in *J. Am. Chem. Soc.*[2]) and my contributions spanned data curation, its analysis, structural and stereochemical quality control, correction of pathway errors, and coordination of the curation team. This work was essential because errors in structures, stereochemistry, step order, or pathway connections would distort subsequent analyses of synthetic strategy and route design. The resulting collection provides a reliable foundation for large-scale theoretical studies of synthetic-route organization, complexity progression, retrosynthetic logic, synthesis planning, cheminformatics, and chemical artificial intelligence.

The next stage of my doctoral research focused on robotic chemistry and multidimensional reaction-space exploration in the laboratory of the Grzybowski group in South Korea. In the work published in *Nature*[3], I studied the E1 and S_N1 reactions of fluorenols as model systems to establish how reactions could be executed robotically, quantified their yields by algorithmic UV–Vis spectral unmixing, and mapped the diverse outcomes across multidimensional condition "hyperspaces". These studies

provided the methodological foundation for applying reaction-hyperspace mapping to more complex chemical systems. The large number of concentration-dependent yield measurements also enabled kinetic and equilibrium models to be fitted to the E1 and S_N1 hyperspaces, providing estimates of kinetic and thermodynamic parameters despite each hyperspace being measured at a single reaction time.

I subsequently designed and carried out a hyperspace study of textbook olefin bromination reactions, now published in *Angewandte Chemie Int. Ed.*[4]. I contributed to defining the reaction hyperspaces, orchestrated on the Robowski platform and then analyzed approximately 1000 reactions, and characterized more than twelve compounds formed across these systems. Automated 80 MHz benchtop NMR acquisition and characteristic-signal fitting enabled hyperspace-wide quantification of the crude reaction mixtures. Unexpected complexity emerged from these textbook brominations, including previously unrecognized products, condition-dependent pathway switchovers, and broader reaction networks than their conventional descriptions suggest.

As an additional extension of this work – not included in the publication cycle constituting the dissertation – I designed and conducted hyperspace mapping of the Chan–Lam coupling, an important C–N bond-forming reaction widely used in synthetic and medicinal chemistry. Because the crude spectra showed extensive overlap and condition-dependent resonance changes, quantitative analysis used NMRanger, a reference-guided deconvolution method that fits selected spectral regions to determine the amounts of individual reaction components. I established the reaction-specific workflow by optimizing the fitting parameters, operating the method independently, and validating the derived yields through ground-truth comparisons, including isolations. This enabled quantitative mapping of coupling products, regioisomers, by-products, and unconsumed reagents across approximately 1,600 conditions. The NMRanger framework was developed by Louis P. J.-L. Gadina, while my contribution focused on its reaction-specific implementation, validation, and application to a complex catalytic system. The corresponding manuscript is currently under revision at *J. Am. Chem. Soc.* and is provided separately as additional material.

The studies constituting this dissertation trace a progression from computationally assisted reaction validation and chemical-data curation to independent expertise in robotic experimentation, reaction-hyperspace design, compound characterization, and automated quantitative analysis. The additional Chan–Lam/NMRanger study further extended these approaches to a complex catalytic system. Together, this work contributed both new chemical knowledge and practical methods for investigating complex reaction systems at a scale inaccessible through conventional experimentation alone.

Streszczenie w Języku Polskim

Moje badania doktorskie dotyczyły czterech powiązanych ze sobą wyzwań współczesnej chemii syntetycznej: odkrywania nowych reakcji wieloskładnikowych i reakcji typu one-pot z wykorzystaniem modułu MECH oprogramowania *Allchemy*; rozwoju *AbSynth*, ogólnodostępnej bazy danych syntez totalnych, jako wiarygodnego, maszynowo odczytywalnego zbioru informacji o opublikowanych syntezach przeznaczonego do przyszłych badań obliczeniowych; robotycznej eksploracji wielowymiarowych przestrzeni reakcyjnych; oraz zautomatyzowanej analizy ilościowej dużej liczby złożonych, surowych mieszanin poreakcyjnych.

W pierwszym etapie, w ramach szerszych prac zespołowych, uczestniczyłam w eksperymentalnej ocenie modułu *Allchemy-MECH*, który generuje mechanistycznie prawdopodobne ścieżki reakcji na podstawie wybranych materiałów wyjściowych. Najpierw zweryfikowałam doświadczalnie zaproponowaną obliczeniowo sekwencję typu one-pot obejmującą dwie syntezy eterów metodą Williamsona połączone przegrupowaniem [2,3]-Wittiga, prowadzącą do rozgałęzionych eterów diallilowych. Jeden z produktów został następnie przekształcony poprzez metatezę z zamknięciem pierścienia w syntetycznie użyteczny cykliczny eter enolowy. Następnie wykorzystałam *Allchemy-MECH* do identyfikacji wieloskładnikowych ścieżek reakcyjnych prowadzących do różnych produktów w zależności od warunków, z udziałem tiolu allilowego, aldehydu aromatycznego i ketonu allenylowego. W warunkach z udziałem DBU prowadziły one do hydroksylowanych monotio-β-diketonów, natomiast po dodaniu chlorku acetylu — do 2,3-dihydrotiofenów; produkt reakcji wieloskładnikowej został ponadto przekształcony w podstawiony pirazol, którego szkielet jest częstym motywem strukturalnym wielu leków. Opracowałam odpowiednie procedury eksperymentalne. Wyniki te — opublikowane w *Nature Communications*[1] — wykazały, że analiza mechanistycznych sieci reakcji może prowadzić do eksperymentalnie realizowalnych reakcji wieloskładnikowych i sekwencji typu one-pot, a jednocześnie pozwala identyfikować ograniczenia związane z kompatybilnością odczynników, wymaganą kolejność ich dodawania oraz zależne od warunków rozgałęzienie ścieżek prowadzących do różnych klas produktów.

Projekt *AbSynth* odpowiadał z kolei na potrzebę dostępu do wiarygodnych, maszynowo odczytywalnych danych dotyczących kompletnych syntez totalnych poprzez zgromadzenie ponad 3000 opublikowanych ścieżek syntetycznych w ogólnodostępnym zasobie. Byłam współautorką korespondencyjną tej pracy, obecnie opublikowanej w **J. Am. Chem. Soc.***[2], a mój wkład obejmował opracowywanie i analizę danych, kontrolę poprawności struktur chemicznych i stereochemii, korygowanie błędów w ścieżkach syntetycznych oraz koordynację pracy zespołu odpowiedzialnego za ich kurację. Prace te miały kluczowe znaczenie, ponieważ błędy w strukturach, stereochemii, kolejności etapów lub połączeniach pomiędzy etapami syntezy prowadziłyby do zniekształcenia późniejszych analiz strategii syntetycznych i projektowania dróg syntezy. Powstały zbiór stanowi wiarygodną podstawę do prowadzonych na dużą skalę badań teoretycznych nad organizacją ścieżek syntetycznych, zmianami złożoności w toku syntezy, logiką retrosyntetyczną, planowaniem syntez, chemoinformatyką oraz sztuczną inteligencją w chemii.

Kolejny etap moich badań doktorskich koncentrował się na chemii robotycznej i eksploracji wielowymiarowych przestrzeni reakcyjnych w laboratorium grupy Grzybowskiego w Korei Południowej. W pracy opublikowanej w *Nature*[3] badałam reakcje E1 i S_N1 fluorenoli jako układy modelowe służące opracowaniu sposobu prowadzenia reakcji za pomocą robota, wyznaczałam ich wydajności metodą algorytmicznego rozdzielania składowych widm UV–Vis oraz mapowałam

zróżnicowane wyniki reakcji w wielowymiarowych „hiperprzestrzeniach” warunków. Badania te stworzyły podstawę metodyczną do zastosowania mapowania hiperprzestrzeni reakcji w bardziej złożonych układach chemicznych. Duża liczba pomiarów wydajności uzyskanych dla różnych stężeń początkowych umożliwiła również dopasowanie modeli kinetycznych i równowagowych do hiperprzestrzeni reakcji E1 i S_N1 oraz oszacowanie parametrów kinetycznych i termodynamicznych, mimo że każdą hiperprzestrzeń badano tylko dla jednego czasu reakcji.

Następnie zaprojektowałam i przeprowadziłam badania hiperprzestrzeni podręcznikowych reakcji bromowania olefin, których wyniki zostały opublikowane w **Angewandte Chemie Int. Ed.**[4]. Uczestniczyłam w definiowaniu hiperprzestrzeni reakcji, koordynowałam wykonanie około 1000 reakcji na platformie Robowski, a następnie analizowałam ich wyniki i scharakteryzowałam ponad dwanaście związków powstających w badanych układach. Zautomatyzowana akwizycja widm na stołowym spektrometrze NMR o częstotliwości 80 MHz oraz dopasowywanie charakterystycznych sygnałów umożliwiły analizę ilościową surowych mieszanin poreakcyjnych w całych badanych hiperprzestrzeniach. Badania ujawniły nieoczekiwaną złożoność tych podręcznikowych reakcji bromowania, obejmującą wcześniej nierozpoznane produkty, zależne od warunków przełączanie dominujących ścieżek reakcji oraz sieci reakcyjne znacznie bardziej rozbudowane, niż wynikałoby to z ich konwencjonalnych opisów.

W ramach dodatkowego rozszerzenia tych badań — niewłączonego do cyklu publikacji stanowiącego podstawę rozprawy — zaprojektowałam i przeprowadziłam mapowanie hiperprzestrzeni sprzęgania Chan–Lam, ważnej reakcji tworzenia wiązań C–N, szeroko stosowanej w chemii syntetycznej i medycznej. Ze względu na znaczne nakładanie się sygnałów w widmach surowych mieszanin poreakcyjnych oraz zależne od warunków zmiany przesunięć chemicznych, w analizie ilościowej zastosowano NMRanger, metodę dekonwolucji opartą na widmach referencyjnych, która wykorzystuje dopasowywanie wybranych regionów widma do określania ilości poszczególnych składników mieszaniny reakcyjnej. Opracowałam procedurę analityczną dostosowaną do badanego układu reakcyjnego poprzez optymalizację parametrów dopasowania, samodzielne stosowanie metody oraz walidację uzyskanych wydajności przez porównanie z niezależnymi wynikami referencyjnymi, w tym z wydajnościami wyznaczonymi po izolacji produktów. Umożliwiło to ilościowe mapowanie produktów sprzęgania, regioizomerów, produktów ubocznych i nieprzereagowanych odczynników w około 1600 zestawach warunków reakcji. Platforma NMRanger została opracowana przez Louisa P. J.-L. Gadinę, natomiast mój wkład koncentrował się na jej wdrożeniu dostosowanym do badanego układu, walidacji i zastosowaniu w złożonym układzie katalitycznym. Odpowiadający tym badaniom manuskrypt znajduje się obecnie na etapie wprowadzania poprawek w ramach procesu recenzyjnego w J. Am. Chem. Soc. i został dołączony osobno jako materiał dodatkowy.

Badania stanowiące podstawę niniejszej rozprawy odzwierciedlają rozwój od wspomaganej obliczeniowo walidacji reakcji i kuracji danych chemicznych do samodzielnego prowadzenia eksperymentów robotycznych, projektowania hiperprzestrzeni reakcji, charakteryzowania związków oraz zautomatyzowanej analizy ilościowej. Dodatkowe badania sprzęgania Chan–Lam z wykorzystaniem NMRanger rozszerzyły zastosowanie tych podejść na złożony układ katalityczny. Łącznie prace te dostarczyły zarówno nowej wiedzy chemicznej, jak i praktycznych metod badania złożonych układów reakcyjnych w skali nieosiągalnej przy wykorzystaniu wyłącznie konwencjonalnych metod eksperymentalnych.

Table of Contents

List of Publications **9**

List of Publications Included in the Doctoral Dissertation 9

List of Publications Not Included in the Doctoral Dissertation 10

1. Research Aims and Objectives **11**

2. Introduction **13**

2.1. *Allchemy-MECH*: Computational reaction discovery 15

2.2. *AbSynth*: Machine-readable representation of total-synthesis pathways 18

2.3. Chemical automation, multidimensional reaction spaces, and automated analysis 20

3. Results and Discussion **26**

3.1. Validation of computationally predicted MCR and one-pot sequences 26

3.2. Construction, quality control, and analysis of the *AbSynth* pathway dataset 33

3.3. Robotic mapping of E1 and S_N1 reaction hyperspaces in fluorenol systems 37

3.4. Hidden reaction networks in textbook olefin brominations 41

4. Summary and Conclusions **46**

5. References **48**

6. Original Publications **55**

7. Author Declarations **111**

8. Additional Materials **165**

8.1. Chan–Lam Coupling Reaction Hyperspace and Automated NMR Quantification 165

8.2. Supporting Information for Chan–Lam Coupling Reaction Hyperspace and Automated NMR Quantification 183

List of Publications

List of Publications Included in the Doctoral Dissertation

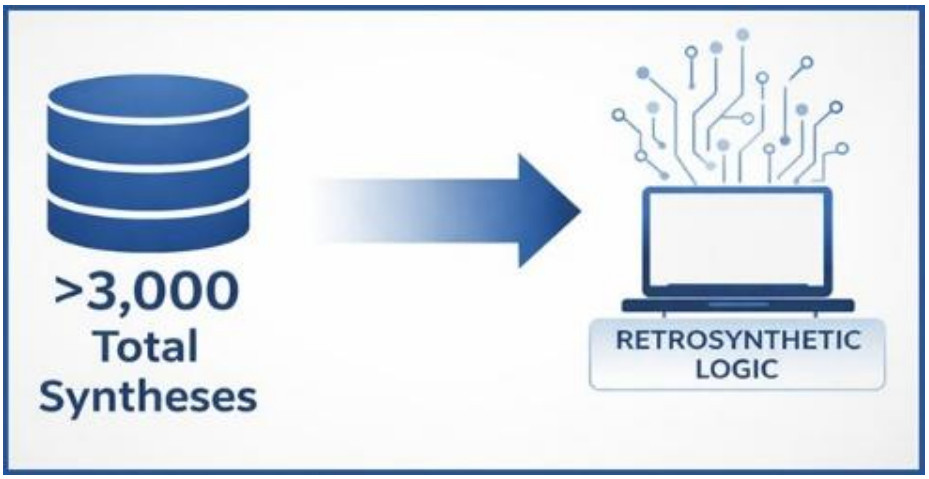
1. Systematic, computational discovery of multicomponent and one-pot reactions.

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. *Nature Communications*, **2024**, 15, 10285. https://doi.org/10.1038/s41467-024-54611-5.

2. *AbSynth*: A Century of Total Syntheses as a Machine-Readable Resource for Analyzing Strategic Logic in Organic Chemistry

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W. D.; Beker, W.; Saillard, E.; Łęgowski, K.; Roszak, R.; Kowalczyk, P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.*; Gryko, D. T.*; **Bilgi-Gadina, Y.*;** Grzybowski, B. A.* *J. Am. Chem. Soc.* **2026**. https://doi.org/10.1021/jacs.6c08434.

***AbSynth* portal:** https://absynth.allchemy.net/



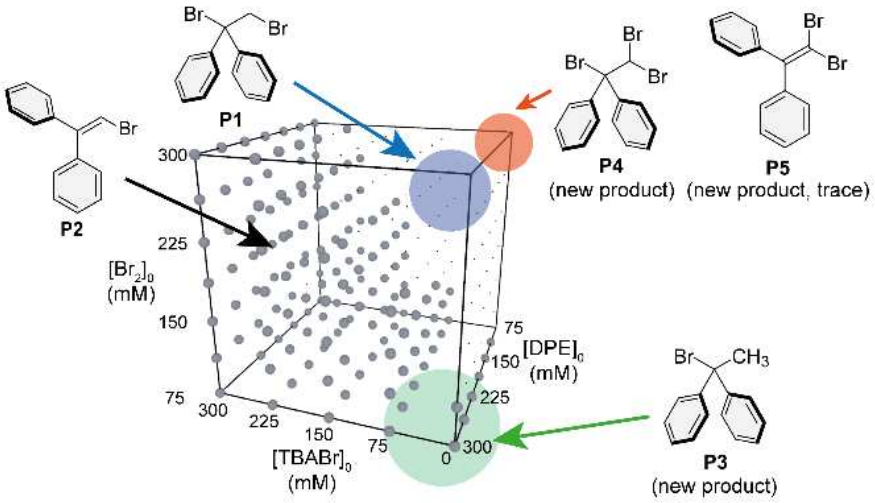
3. Robot-assisted mapping of chemical reaction hyperspaces and networks.

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; **Bilgi, Y.;** Gadina, L.; et al. *Nature* **645**, 922–931 (2025). https://doi.org/10.1038/s41586-025-09490-1

4. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations.

Jia, Y.$^{+}$; Wong, W.-S.$^{+}$; **Bilgi-Gadina, Y.$^{+}$;** Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. *Angew. Chem. Int. Ed.* **2026**, e9132891. https://doi.org/10.1002/anie.9132891.

+ These authors contributed equally.

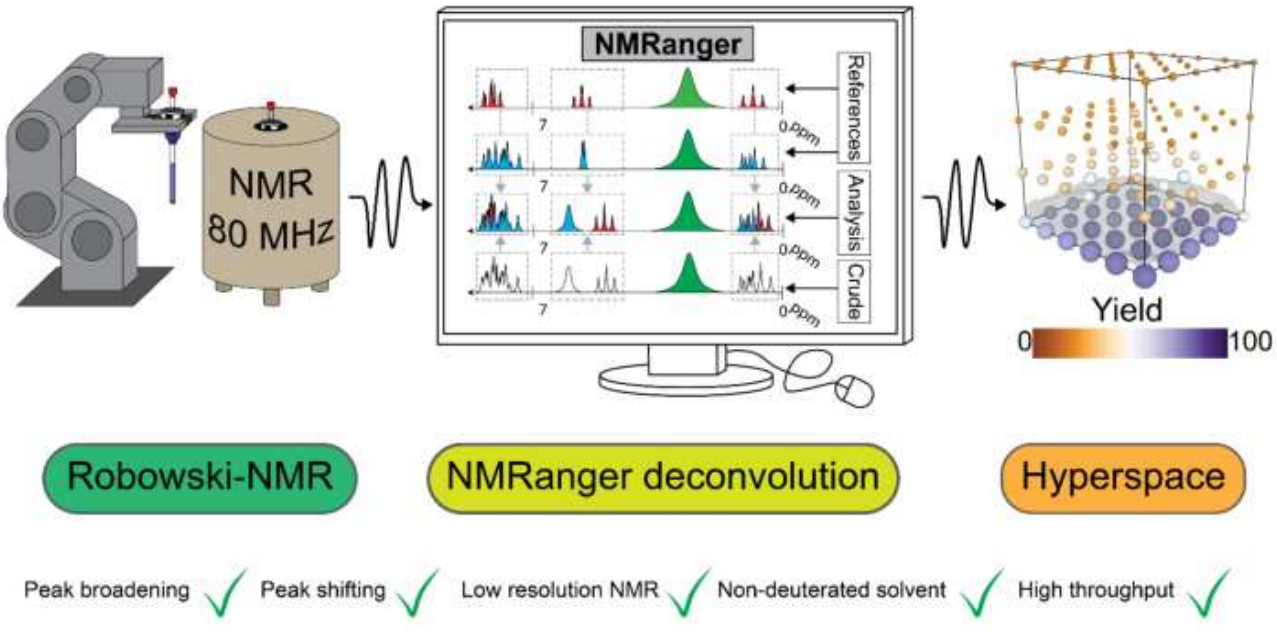


List of Publications Not Included in the Doctoral Dissertation

1. Automated NMRanger Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures

Yasemin Bilgi-Gadina, Louis P. J.-L. Gadina, Rafal Frydrych, Bartosz A. Grzybowski* *J. Am. Chem. Soc.* ***Submitted.*** **2026**. ***In revision.***

This manuscript is currently under revision and therefore does not form part of the publication cycle constituting the doctoral dissertation. It is included only as additional material and is not formally subject to evaluation as part of that publication cycle.



2. Studying the Outcomes and Mechanisms of Carbocationic Rearrangements Using Algorithm-Augmented Experimentation.

Gadina, L. P. J. -L.; Baś, S.; Makkawi, A.; **Bilgi-Gadina, Y**.; Klucznik, T.; Beker, W.; Grzybowski, B. A. *ChemistryEurope* **2026**, *4* (2), e202500189. https://doi.org/10.1002/ceur.202500189.

1. Research Aims and Objectives

This dissertation investigated complementary approaches to reaction discovery and the analysis of chemical reactivity, including computational prediction of new reaction pathways, curation and analysis of complete total-synthesis pathways, and robotic experimentation. The work combined experimental validation of computationally predicted reactions, construction and analysis of the machine-readable *AbSynth* dataset, and robotic exploration of reaction networks. Four objectives, detailed later in this section, defined the studies constituting the dissertation.[1–4]

Recent advances in computational chemistry and chemical AI, machine-readable chemical data, laboratory automation, and high-throughput analysis have created new opportunities for reaction discovery and systematic studies of chemical reactivity.[5–10] In this dissertation, these developments were investigated through two connected but distinct research directions. The first focused on computationally assisted reaction discovery and the organization of synthetic knowledge through *Allchemy-MECH*[1] and *AbSynth*.[2] The second focused on robotic exploration of multidimensional reaction "hyperspaces" and on analytical methods capable of quantifying the resulting complex reaction mixtures.[3,4,11]

The first aim of this dissertation was to determine whether reaction networks generated from mechanistic steps by *Allchemy-MECH* algorithms could provide meaningful starting points for the discovery of multicomponent and one-pot transformations. This is important because many synthesis-planning methods remain dependent on known reactions[12–14], whereas mechanistic exploration may identify transformations without direct precedent and anticipate practical constraints before experimental testing.

The second aim was to improve the quality and accessibility of route-level chemical data through the development of *AbSynth*. Complete total syntheses contain information on step order, branching, convergence, stereochemistry, and strategic decisions that is lost when reactions are stored only as isolated transformations. A reliable, open-access, machine-readable representation of such pathways is therefore needed for large-scale analysis, synthesis planning, benchmarking, and future chemical-AI applications.[2,8,15]

My doctoral work continued with an experimental direction centered on robotic chemistry. The principal aim of this experimental work was to establish reaction-hyperspace mapping as a method for understanding how chemical outcomes emerge from interacting variables.[3,16–18] Conventional optimization generally seeks to improve a predefined reaction outcome[19–21], whereas broader exploration of reaction spaces can reveal alternative products and previously unrecognized reactivity.[22,23] Robotic experimentation was therefore used to examine multidimensional condition

spaces reproducibly and to assess whether both productive and low-yielding regions could provide mechanistic and synthetic information.

Because the value of hyperspace mapping depends on analytical throughput, a further aim was to broaden the range of reactions accessible to automated analysis. The workflow started with UV–Vis spectral unmixing[3] and extended to automated benchtop NMR: a textbook olefin bromination reaction was used to establish quantification based on peak picking and fitting of characteristic signals[4], while the more demanding Chan–Lam coupling was used to test NMRanger[11] for strongly overlapped, shifting, and broadened crude spectra across a catalytic reaction space (see **Section 8** for details).

Accordingly, this dissertation aimed to:

1. Experimentally evaluate computationally predicted MCR and one-pot reactions;
2. Develop reliable, open-access, machine-readable total-synthesis pathways;
3. Establish robotic hyperspace mapping of condition-dependent reactivity;
4. Extend hyperspace analysis from UV–Vis to automated benchtop NMR.

2. Introduction

Chemical artificial intelligence (AI) and robotic chemistry connect computational decision-making with automated experimentation. Computational methods can propose synthetic routes or prioritize reaction conditions, while robotic platforms execute protocols and collect analytical measurements.[24–27] Their integration ranges from computer-generated synthetic routes executed on automated platforms[24] to closed-loop workflows in which experimental yields update a computational model that selects conditions and substrate combinations for the next round.[25]

Within chemical AI, a longstanding challenge is synthesis planning: given a target molecule, how can a viable synthetic route from accessible starting materials be identified? Retrosynthetic analysis works backwards from the target by disconnecting it into simpler precursors. Because a target can often be disconnected in several ways, multiple routes can be generated and evaluated according to chemical feasibility and selectivity.[28,29] The formalization of retrosynthetic analysis emerged from Corey's work in the 1960s. In 1967, he described general principles for planning complex-molecule synthesis,[30] and in 1969, Corey and Wipke introduced computer-assisted synthesis design, expressing elements of retrosynthetic reasoning in a form that could be processed algorithmically.[31]

At a larger scale, synthetic knowledge can be represented as reaction networks connecting compounds through reported transformations. The Grzybowski group were the pioneers of large-scale computational analysis of such networks. In 2005, the group represented millions of Beilstein compounds and reactions as the Network of Organic Chemistry, revealing a structured, evolving architecture containing highly connected molecules linking many parts of the network.[32,33] In 2012, this network was used to identify and experimentally validate one-pot sequences comprising several consecutive transformations[34] and, in a Grzybowski–Bishop collaboration, to optimize synthetic routes to single or multiple targets.[35] This network-based view was subsequently extended by other groups; for example, Lapkin and co-workers built Reaxys-based reaction networks for automated route generation and multi-criteria process evaluation.[36]

By 2018, computer-assisted planning could propose complete synthetic routes. Chematica, a synthesis-planning platform developed by the Grzybowski group and later developed into SYNTHIA, now marketed globally by Merck, used expert-encoded reaction rules and related network-search algorithms to drive synthesis planning; eight computer-designed syntheses were subsequently carried out experimentally.[28] In parallel, Segler, Preuss, and Waller developed a data-driven approach combining reaction templates, neural networks, and Monte Carlo tree search. The resulting routes were evaluated computationally and by chemists but were not carried out experimentally in that study.[29] In 2020, the Grzybowski group extended computer-assisted planning to complex natural products, validating three computer-designed syntheses.[6] In 2023, Cernak's group analyzed 50

SYNTHIA-generated routes by graph-edit analysis to identify high-impact synthetic steps, guiding the development of a three-step enantioselective synthesis of (−)-stemoamide.[37]

The Grzybowski group also addressed the complementary forward problem: starting from a defined set of substrates, what reaction pathways and products become accessible under specified conditions? In 2020, a forward-synthesis algorithm generated a large network of prebiotic reactions from a small set of primordial substrates, enabling analysis of synthetic connectivity, chemical emergence, and self-regeneration.[38] Two years later, a related study introduced an algorithm for identifying iterative reaction sequences in which the functionality required for continued molecular growth is regenerated; selected predictions were validated experimentally.[39] A further development of forward reaction exploration – this time, however, at the level of mechanistic transformations – was HopCat, which in 2024 combined mechanistic-step networks with physical-organic rules, quantum-chemical calculations, and kinetic modelling to predict complex cationic rearrangements.[40] In 2026, our group applied HopCat in algorithm-augmented experiments to investigate the outcomes and mechanisms of carbocationic rearrangements.[41] In 2024, our group also disclosed *Allchemy-MECH*—a part of this dissertation—extending mechanism-level network generation to the systematic discovery of multicomponent and one-pot reactions, several of which were experimentally validated.[1]

Alongside these advances, increasing attention was given to standardized, machine-readable representations of chemical information. The Open Reaction Database, introduced in 2021, established a framework for recording and sharing reaction data.[8] At the level of complete synthetic routes, our own *AbSynth* collection provides a machine-readable resource of curated total syntheses that preserves structures, stereochemistry, step order, and convergent branches for systematic analysis of multistep synthetic strategy.[2]

As computational methods became increasingly capable of proposing routes, conditions, and experiments, attention also turned to their efficient experimental execution. Burke's group demonstrated automated building-block-based small-molecule synthesis in 2015.[42] Cronin's group coupled robotic experimentation with reaction analysis and machine learning in 2018[26] and, in 2019, encoded synthetic procedures as programmable operations for modular robotic execution.[27] Also in 2019, Coley and co-workers connected AI-assisted synthesis planning to a reconfigurable robotic flow platform, extending computer-assisted planning from *in silico* design to automated laboratory execution.[24] Automation also enabled experimental results to guide subsequent experiments: Cooper's group combined a mobile robotic chemist with Bayesian experiment selection in 2020,[43] Doyle's group demonstrated Bayesian reaction optimization in 2021,[44] and in 2022 a Burke–Aspuru-Guzik–Schroeder–Grzybowski collaboration used closed-loop experimentation to identify general conditions for heteroaryl Suzuki–Miyaura coupling.[25]

By 2025, automated platforms were being used for high-throughput reaction discovery, yield optimization, and generation of standardized reaction data. The Grzybowski group pursued and pioneered a different objective with the Robowski hyperspace-mapping approach: determining systematically how reaction outcomes change across multidimensional parameter spaces and how these variations can help reconstruct the connectivity and kinetics of reaction's mechanistic networks.[3] Across thousands of conditions, the study quantified major products and identifiable by-products, revealed regions of unexpected reactivity and switchovers between major products, and used systematic variation of substrate proportions to reconstruct reaction networks and identify hidden intermediates and products.[3] In 2026, this approach was extended to the Pechmann and Biginelli reactions, catalyst-controlled radical cascades, and bromination reactions analysed by automated benchtop NMR.[4,16–18]

Together, these developments illustrate the convergence of computational synthesis planning, reaction-network analysis, machine-readable chemical data, and robotic experimentation as complementary approaches to exploring chemical reactivity. The Grzybowski group made foundational contributions to this progression, from large-scale organic reaction-network analysis and computer-assisted synthesis planning to forward and mechanism-based reaction discovery and, more recently, robotic hyperspace mapping. The following sections provide the specific background for the studies in this dissertation: computational reaction discovery by *Allchemy-MECH* (**Section 2.1**), machine-readable representation and analysis of total-synthesis pathways in *AbSynth*, (**Section 2.2**), and chemical automation, reaction-hyperspace mapping, and large-scale automated analysis and quantification of crude reaction mixtures (**Section 2.3**).

2.1. *Allchemy-MECH*: Computational reaction discovery

The discovery of new reaction classes expands the range of accessible molecular structures and can provide shorter and more economical routes to known compounds. Multicomponent reactions (MCRs) are especially valuable because they combine three or more starting materials into a single product, often forming several bonds and generating considerable molecular complexity in one operation.[45–47] Common substrate classes include aldehydes or ketones, amines, isocyanides, carboxylic acids. Well-established examples include the Ugi, Passerini, Biginelli, and Strecker reactions.[46,48,49] One-pot reactions follow a related strategy by carrying out consecutive transformations in the same vessel without isolating intermediates between individual steps, often using staged reagent addition or changes in reaction conditions.[50,51] By reducing the number of synthetic operations, purification procedures, solvent-intensive workups, and material losses associated with intermediate handling, both strategies contribute to step and pot economy and may improve the overall resource efficiency of a synthesis.[52,53] Their ability to generate complex and

structurally diverse products from relatively simple starting materials also makes them valuable for heterocycle synthesis, medicinal chemistry, and the rapid preparation of molecular libraries.[54]

The potential of multicomponent chemistry is illustrated by Robinson's synthesis of tropinone, in which a complex bicyclic framework was assembled directly from simple components in a single operation (**Figure 1a**).[55] However, designing new MCRs and one-pot sequences rationally remains challenging. The participating substrates must display complementary reactivity, the required intermediates must form in appropriate order, and the productive sequence must compete successfully against side reactions. In a true MCR, the components must coexist from the beginning of the reaction, whereas a one-pot sequence may require staged reagent addition or deliberate changes in the reaction conditions to prevent premature cross-reactivity.[50,51]

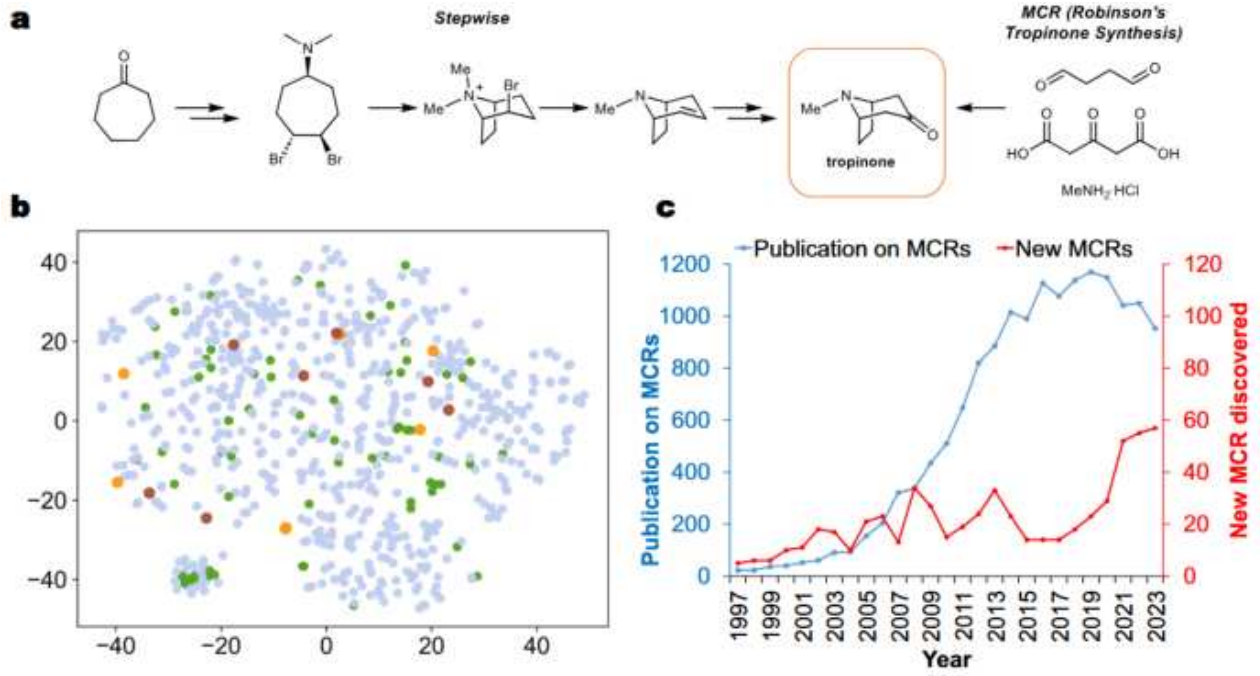


Figure 1. Significance and current discovery rate of multicomponent reactions.[1] **a)** A classic example illustrating the elegance and efficiency of Robinson's one-step, MCR synthesis of tropinone vs. prior, fifteen-step synthesis[55]. In the latter, only two key steps are shown. **b)** t-SNE projection "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.[1] 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[57] **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**. *The figure was adapted from Publication 1.*

Despite more than a century of research, the number of recognized MCR classes remains limited to several hundred.[1] The known transformations span a broad chemical space, but many regions remain underexplored (**Figure 1b**). Many established families rely on a relatively small set of extensively explored reactivity patterns, particularly those based on isocyanides, imines, and β-dicarbonyl compounds.[45,47,56] Consequently, a substantial part of the literature has focused on expanding the scope of established MCRs rather than introducing new mechanistic classes.[1] Despite

numerous publications on MCRs, the discovery of genuinely new MCR classes remains relatively rare (**Figure 1c**).[1] The rational design of new MCRs remains challenging because it requires the simultaneous evaluation of numerous elementary steps involving multiple substrates, transient intermediates, by-products, and competing pathways that may divert the reaction from the desired sequence.[1]

Computer-assisted reaction discovery has long aimed not only to retrieve known chemistry but also to generate reactions without direct precedent. Early studies by Ugi and co-workers described a mathematical model of constitutional chemistry in which ensembles of molecules were represented by atomic vectors and bond and electron (BE) matrices, and reactions by reaction (R) matrices. This representation enabled computer programs to systematically generate certain types of possible reactions.[58,59] Many contemporary synthesis-planning systems, however, rely mainly on recorded reactions or curated templates. Although effective for retrieving precedented chemistry and proposing routes to defined targets, their predictions are constrained by the knowledge encoded in their datasets.[6,15,60,61] Mechanistically informed approaches instead represent reactions as sequences of elementary or arrow-pushing steps that can be assembled into networks containing productive routes, competing pathways, by-products, and subsequent transformations.[1]

Allchemy-MECH is a pioneering effort in this area and applies this concept by iteratively using generalized mechanistic transforms (encoded in the SMARTS notation) to selected starting materials. Candidate sequences of mechanistic steps are assessed using physical-organic chemistry rules, condition compatibility, approximate kinetics, and possible cross-reactivity. Sequences whose components can coexist from the beginning may be classified as MCRs, whereas sequences requiring controlled reagent addition to avoid cross-reactivity are proposed as one-pot procedures. Competing reactions, by-products, and their subsequent transformations can also be included in the analysis.[1]

This strategy extends computer-assisted synthesis beyond the recombination of known, complete reactions and can identify sequences not previously recognized as distinct reaction classes. Computational plausibility alone, however, does not ensure experimental feasibility. Selectivity, substrate stability, reaction rates, and the effects of solvent, concentration, temperature, and mixing must still be evaluated experimentally to establish whether the proposed sequences proceed reproducibly under compatible conditions.[1] In the work forming the first part of this dissertation, I experimentally validated selected *Allchemy-MECH* predictions by developing and optimizing suitable experimental conditions. Validation of ***Mach3*** showed that the algorithm correctly identified a cross-reactivity constraint requiring a specific order of reagent addition, while ***Mach5*** and ***Mach6*** confirmed that the predicted mechanistic pathways could lead to distinct multicomponent reactions under different conditions.[1] These validations established that *Allchemy-MECH* could propose

experimentally accessible one-pot and multicomponent reactions while correctly identifying reagent-compatibility constraints, required order of addition, and condition-dependent pathway divergence.

2.2. *AbSynth*: Machine-readable representation of total-synthesis pathways

Published total syntheses provide a record of the multistep strategies used to construct complex molecules and therefore offer a valuable basis for identifying broader patterns in synthesis planning. However, many established complexity metrics and retrosynthetic heuristics were developed from relatively limited sets of total syntheses, typically numbering from tens to hundreds, or from larger reaction datasets dominated by examples outside the context of total synthesis.[2] To evaluate such concepts across thousands of complete routes, existing resources present different limitations. Total-synthesis repositories such as Chemistry by Design[62,63], SynArchive[64] and Hans Reich's Collection[65] primarily provide graphical schemes rather than machine-readable molecular data. Commercial databases such as Reaxys and SciFinder provide access to reactions reported in published syntheses. However, many total syntheses published before 1990 had not been digitized in these databases.[2] Reaxys also contains an appreciable entry-error rate of approximately 15.3%[2] while both Reaxys and SciFinder restrict public redistribution of their content.[2] In contrast, *AbSynth* provides a curated, machine-readable collection of complete total-synthesis routes that is freely available to academic users for computational analysis.[2]

To construct the *AbSynth* collection, complete synthetic routes were reconstructed from the primary literature and converted into a machine-readable form suitable for computational analysis. A representative example is shown in **Figure 2**.[2,61,66] Molecular structures from the entire collection are now freely available to the academic community as SMILES[2], while the *AbSynth* portal[67] preserves the progression of individual syntheses and their branches, allowing users to navigate both the main route and contributing sub-branches.[67] Because manual reconstruction and drawing of complex structures are susceptible to errors, the dataset underwent more than ten rounds of corrections and cross-checks. The remaining step-level error rate was estimated at approximately 3–4%, and identified errors continue to be corrected.[2]

In its reported form, *AbSynth* contains 3,077 total syntheses and 59,712 individual steps, covering almost a century of total-synthesis research up to 2024.[2] The collection was developed not only to provide complete total-synthesis routes in a machine-readable form but also to enable systematic analysis of the architecture and logic of total syntheses. The analyses examined network connectivity among routes, historical trends in pathway structure and convergence, the behavior of structural and synthetic complexity metrics along synthetic pathways, and the applicability of Corey's retrosynthetic heuristics.[2,68]

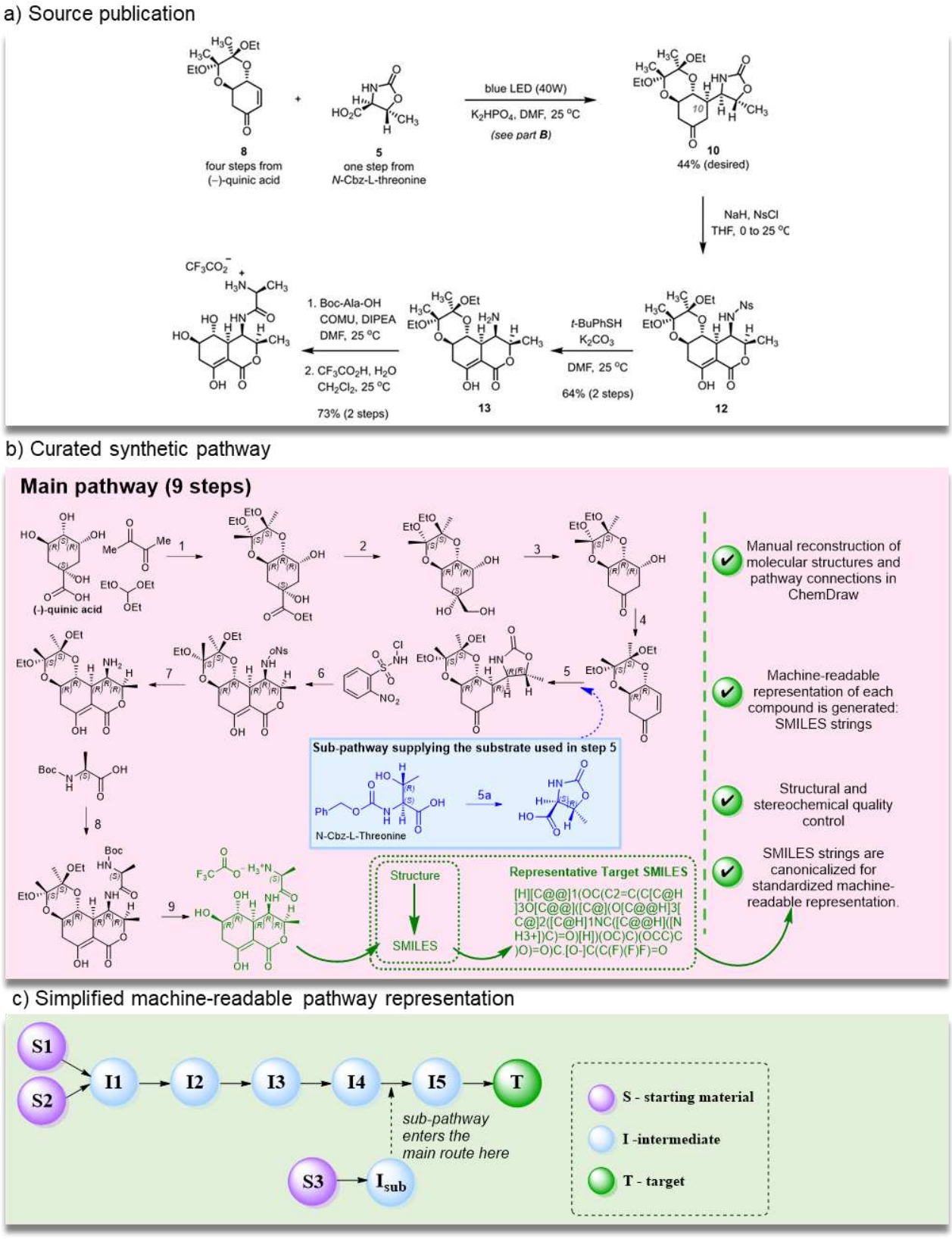


Figure 2. Conversion of a published synthesis into a machine-readable pathway. a) Representative literature scheme for the synthesis of (+)-actinobolin.[66] **b)** Manual reconstruction of the main route in ChemDraw and its contributing sub-pathway as connected molecular structures. Each compound was encoded as a SMILES string and recorded with its step assignment and pathway connections. The green dashed frame indicates the representative target SMILES. The reconstructed structures and stereochemistry were subsequently quality-controlled, followed by SMILES canonicalization. **c)** Simplified pathway representation showing the relationships among starting materials, intermediates, the sub-pathway, and the final target. Panel a was adapted from the original publication[66]; panels **b** and **c** illustrate the ***AbSynth curation workflow*** described in *Publication 2*.

These analyses showed that complex syntheses cannot be planned by considering only one reaction at a time. The tested complexity measures did not simply decrease after every step. Many steps instead prepared the molecule for a later reaction, for example by protecting a functional group or changing one functional group into another. Such steps may not simplify the molecular skeleton themselves, but they can be necessary before an important bond can be formed or disconnected.[2] The analysis of Corey's strategic bond-disconnection rules showed a similar limitation.[68] Many of these rules had low accuracy, meaning that even when a rule was applicable, the corresponding disconnection was often not used experimentally. In addition, some rules pointed to several possible bonds rather than identifying one clear disconnection site.[2] These findings highlight the need to consider several consecutive reactions as parts of longer synthetic sequences rather than evaluating each step in isolation. *AbSynth* was therefore used to identify recurring patterns in such multistep sequences, including how long sequences without changes to the molecular skeleton typically extend and how long different functional groups can persist unchanged along a synthetic route. By representing complete synthetic pathways in a machine-readable form, *AbSynth* enables such sequence-level features to be analyzed across thousands of syntheses, allows established concepts of synthesis planning to be tested across a much larger collection of complete synthetic routes, and reveals recurring patterns that can inform practical multistep synthesis planning.[2]

2.3. Chemical automation, multidimensional reaction spaces, and automated analysis

The increasing digitization of synthesis has established automated experimentation as an important component of modern synthetic chemistry.[69] Robotic platforms provide consistent liquid handling, timing, temperature control, and data recording, enabling reaction studies at a scale and reproducibility difficult to achieve manually. Miniaturization to micromole-scale reactions, with many reactions performed simultaneously in multi-vial plates, further allows numerous combinations of conditions to be examined while reducing reagent consumption, solvent use, and chemical waste. Integration with computational planning and data-analysis methods can additionally accelerate reaction development.[21,70–75]

Automated experimentation can serve several scientific purposes. In automated multistep synthesis, the principal aim is the reproducible execution of a predefined sequence.[71] In closed-loop optimization, an algorithm uses previous experimental results to select subsequent conditions and maximize an objective such as yield or selectivity.[21,72–75] Reaction-hyperspace mapping pioneered by our group uses automation for a different purpose: to investigate chemical reactions as networks of competing and consecutive transformations embedded within multidimensional spaces of experimental conditions.[3] Conventional reaction schemes commonly emphasize a principal, essentially linear pathway from reactants to a selected product, while competing pathways and other products receive less attention. The same reactants can, however, participate in broader reaction

networks containing alternative intermediates and products. Changes in concentration, stoichiometry, temperature, and other conditions can alter reaction rates and equilibria within these networks, thereby favoring or suppressing different reaction pathways.[3] Broad hyperspace scans can therefore reveal previously unrecognized products, switchovers between dominant outcomes, and condition-dependent relationships among reaction components that help reconstruct the reaction network.[3,4,16–18]

Figure 3 illustrates this progression from a conventional linear representation to a reaction network and, finally, to a network embedded within a reaction hyperspace. In the linear representation, attention is focused on the pathway leading from the reactants to a selected product, whereas the network representation additionally includes competing intermediates and pathways leading to alternative products. Embedding this network within a multidimensional hyperspace links these pathways to the experimental conditions under which they are favored. Each point in the hyperspace therefore represents a particular set of conditions and the corresponding reaction composition, while different regions can favor different products and pathways. Systematic mapping of these condition-dependent distributions can consequently provide information about the organization of the reaction network and how the relative importance of its pathways changes across the hyperspace.[3,4,16–18]

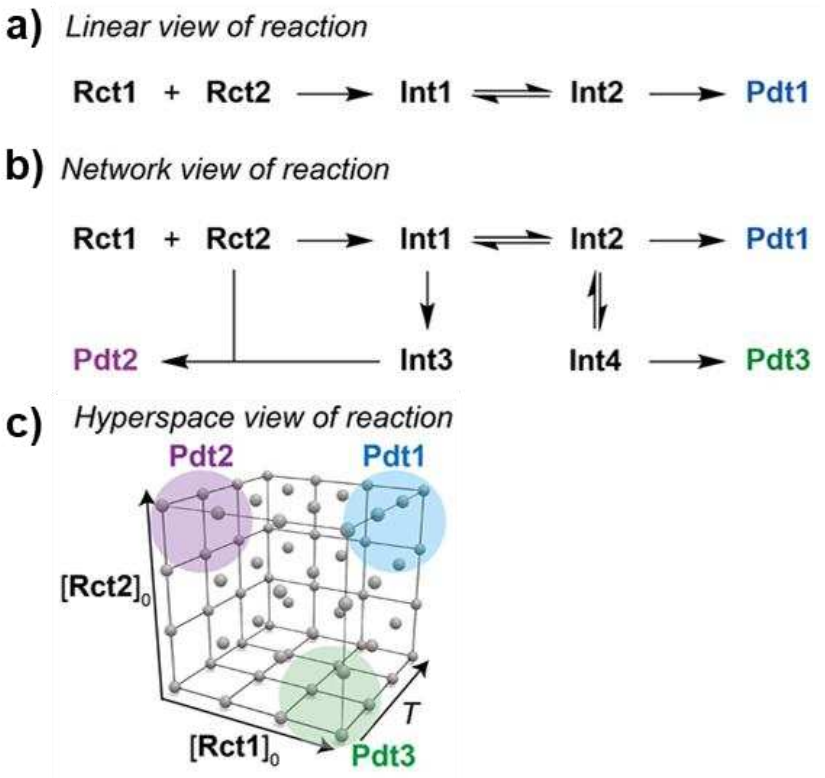


Figure 3. Linear, network, and hyperspace representations of a chemical reaction. a) In a conventional linear representation, reactants Rct1 and Rct2 are converted into product Pdt1 through hypothetical intermediates Int1 and Int2, while competing pathways and additional products are omitted. **b)** A network representation includes alternative intermediates and competing pathways leading to additional products, here Pdt2 and Pdt3. Changes in reaction conditions can alter the relative importance of these pathways, causing different products to become dominant. **c)** The reaction network can be viewed as embedded within a multidimensional hyperspace of experimental conditions. A three-dimensional space defined by the initial concentrations of Rct1 and Rct2 and temperature, *T*, is shown. The colored regions indicate conditions under which Pdt1 (*blue*), Pdt2 (*purple*), or Pdt3 (*green*) is the dominant product. *Reproduced from Publication 4.*

Automation is essential for this purpose because meaningful hyperspace reconstruction requires large, internally consistent datasets. Robotic execution allows the same dispensing, heating, timing, sampling, and analytical procedures to be applied across hundreds or thousands of reactions. Automation alone, however, is insufficient. Increasing reaction throughput has scientific value only when the analytical workflow can process the same number of samples as the reaction platform produces. If each sample requires manual processing, chromatographic separation, or product isolation, the effective throughput of the complete workflow remains low regardless of how rapidly the reactions are performed.

Several analytical approaches can be used for high-throughput reaction analysis, each providing different types of chemical information. Chromatographic methods, including HPLC and LC–MS, are widely used for the analysis of large reaction arrays, providing chemical separation and quantitative information; LC–MS additionally provides mass-to-charge information that can support product identification.[76,77] Direct mass-spectrometric screening can further increase throughput, although matrix components such as salts, solvents, reagents, catalysts, and by-products can influence analyte response.[77] Miniaturized photocapture methods provide very high throughput but require compatibility with both the capture chemistry and subsequent MS detection.[78] Optical spectral unmixing offers another rapid approach and has been used to reconstruct reaction hyperspaces by fitting crude UV–Vis spectra as combinations of reference spectra. Its applicability, however, decreases when components absorb weakly or have strongly overlapping absorption spectra.[3,16–18] The choice of analytical method therefore depends on how many samples must be analyzed, the complexity and concentration range of the reaction mixtures, whether individual components can be clearly distinguished, the required sensitivity, the amount of sample preparation needed, compatibility with the reaction matrix and automated workflow, and, for methods involving spectral fitting or deconvolution, the computational requirements of data analysis.

NMR spectroscopy provides a complementary and more general-scope approach because it provides structurally informative and compound-specific signals directly from crude reaction mixtures. Several automated NMR quantification methods have been developed, mainly for metabolomic applications. These include targeted approaches such as Chenomx and MagMet[79,80], probabilistic deconvolution methods such as Bayesil and BATMAN[81,82], reporter-signal methods such as AQuA[83], and the peak-picking-free Magnetstein algorithm.[84] Among these methods, BATMAN and Magnetstein were benchmarked on representative crude reaction spectra from the NMRanger study (see **Section 8.2**). Magnetstein showed poor quantitative performance and long calculation times, including failure to converge for one test case. BATMAN gave a reasonable result for the simpler mixture but lost substantial accuracy for the more complex spectrum and required considerably longer calculation times than NMRanger. In these benchmark tests, neither method

provided the combination of quantitative reliability, robustness, and computational efficiency required for hyperspace-wide analysis (**Table 1**).[11,81,84]

Table 1. Benchmark comparison of NMRanger, Magnetstein, and BATMAN on representative crude reaction spectra.

Algorithm	PhMorph/ $PhB(OH)_2$ (Calculation time)	TriOMePh/PhAza (Calculation time)
NMRanger	1.656 (30 sec)	0.770 (1min 34 sec)
Magnetstein without solvent mask	No convergence after 18h.	0.262 (2h53min)
Magnetstein with solvent mask	13.275 (3h5min)	0 (2h42min)
BATMAN	1.354 (4min 29s)	4.843 (28min 35s)
Ground Truth	1.775	0.757

Benchtop NMR is particularly compatible with robotic reaction studies because compact instruments can be integrated directly with automated sample transfer, spectrum acquisition, and data processing. In the olefin-bromination study constituting part of this dissertation, automated 80 MHz benchtop NMR was used to analyze nearly 1,000 crude reaction mixtures.[4] When each relevant component possessed a sufficiently resolved characteristic resonance, automated quantification could be performed by fitting these signals and converting the fitted areas into concentrations using calibration data (**Figure 4a**)[4]. This enabled consistent hyperspace-wide quantification without manual integration of every spectrum and extended the robotic workflow beyond systems suitable for UV–Vis spectral unmixing.

The applicability of characteristic-signal fitting is nevertheless limited by the requirement that diagnostic resonances remain sufficiently resolved throughout the investigated condition space. In more complex mixtures, several components may contribute to the same spectral region, while changes in concentration, acid–base interactions, sample composition, and metal coordination can alter chemical shifts, peak widths, intensities, and line shapes. These effects make quantification based on individual diagnostic peaks increasingly difficult.

As a subsequent extension of this analytical workflow, the NMRanger algorithm was developed to address strongly overlapping and condition-dependent spectra by fitting selected spectral regions using models derived from reference spectra (**Figure 4b**). Adjustable peak positions, widths, amplitudes, and Gaussian/Lorentzian contributions (**Figure 4c-e**) allow the reference-derived models to accommodate moderate spectral changes while simultaneously estimating the contributions of several reaction components. This approach therefore extends automated NMR analysis from characteristic-signal fitting to broader spectral deconvolution (see **Sections 8.1 and 8.2** for details).[11]

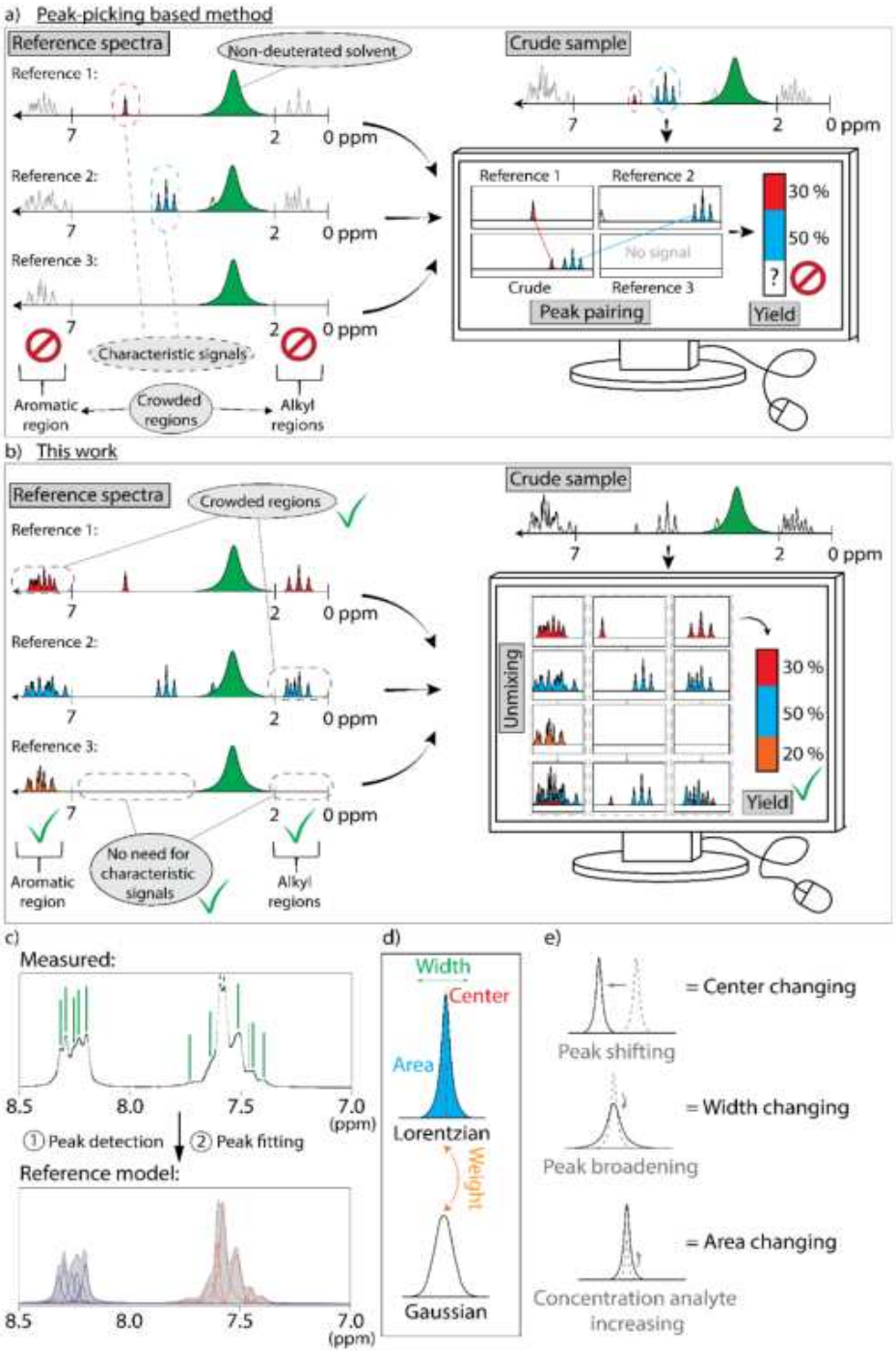


Figure 4. Conventional peak-picking vs. NMRanger algorithms. a) In peak-picking approaches (e.g., the algorithm we described in ref.[4]), reference spectra are stored in a library, and characteristic peaks from each reference are paired with corresponding peaks in the crude sample. The crude composition is then calculated from these matched characteristic signals. This approach works well as long as characteristic signals are cleanly resolved. **b)** NMRanger selected-region fitting approach. Over user-specified regions, reference spectra are converted into editable models based on the approximation of the peaks by pseudo-Voigt functions (see **c**). The crude spectra are fitted and quantified by unmixing into individual component contributions, using the reference models as spectral templates. **c)** Scheme illustrating conversion of a "crowded" region of the experimental ^{1}H NMR spectrum of phenylboronic acid into a functional model described as a sum of fitted peaks. Detected peaks, indicated by vertical green lines in the top panel, are fitted with pseudo-Voigt functions to generate the continuous model shown in the bottom panel. Individual peak contributions are colored red or blue according to the multiplet to which they belong. The resulting substrate model is shown as the grey-filled profile. **d)** Illustration of the parameters defining each pseudo-Voigt peak model, including amplitude, center, width, and Gaussian/Lorentzian weighting. **e)** Representative effects of changing pseudo-Voigt parameters during fitting, including peak shifting through center changes, peak broadening through width and Gaussian/Lorentzian weighting changes, and concentration changes through amplitude scaling. *Reproduced from Section 8.*

These capabilities are particularly important in reaction-hyperspace studies because the analytical method must tolerate systematic spectral variation across widely different conditions while quantifying desired products, competing products, by-products, and unconsumed reagents. As an additional extension of the doctoral research, NMRanger was applied to the Chan–Lam coupling, a demanding catalytic system in which the crude spectra are affected by extensive signal overlap, condition-dependent resonance shifts, broadening, and copper-induced spectral perturbations. This work extended the automated benchtop NMR workflow beyond characteristic-signal fitting to region-selective spectral deconvolution across a multidimensional catalytic reaction space. Because the corresponding manuscript is currently under revision, this study does not form part of the publication cycle constituting the dissertation and is discussed here only to provide methodological and scientific continuity. The full manuscript and its Supporting Information are provided separately as additional material in **Sections 8.1 and 8.2** [11]

3. Results and Discussion

My PhD studies began with the experimental validation of computationally proposed multicomponent and one-pot reaction sequences, followed by work on machine-readable synthesis data, robotic reaction-hyperspace exploration, and automated quantitative analysis of complex reaction mixtures. The results are discussed in this order to reflect the progression of the research.

3.1. Validation of computationally predicted MCR and one-pot sequences

The experimental validation began with the selection of substrate sets for computational analysis. The main criteria were mechanistic distinction from established MCR classes, a substantial increase in structural complexity, the use of commercially available or readily prepared substrates, and formation of potentially useful scaffolds. Computationally proposed sequences in which all substrates could coexist from the beginning were treated as multicomponent reactions. Reagents used only to promote or redirect subsequent transformations, without contributing atoms to the final product, could be introduced later without changing this classification. By contrast, when an atom-contributing substrate had to be added sequentially to avoid cross-reactivity, the sequence was classified as a one-pot procedure (**Figure 5**).[1,85]

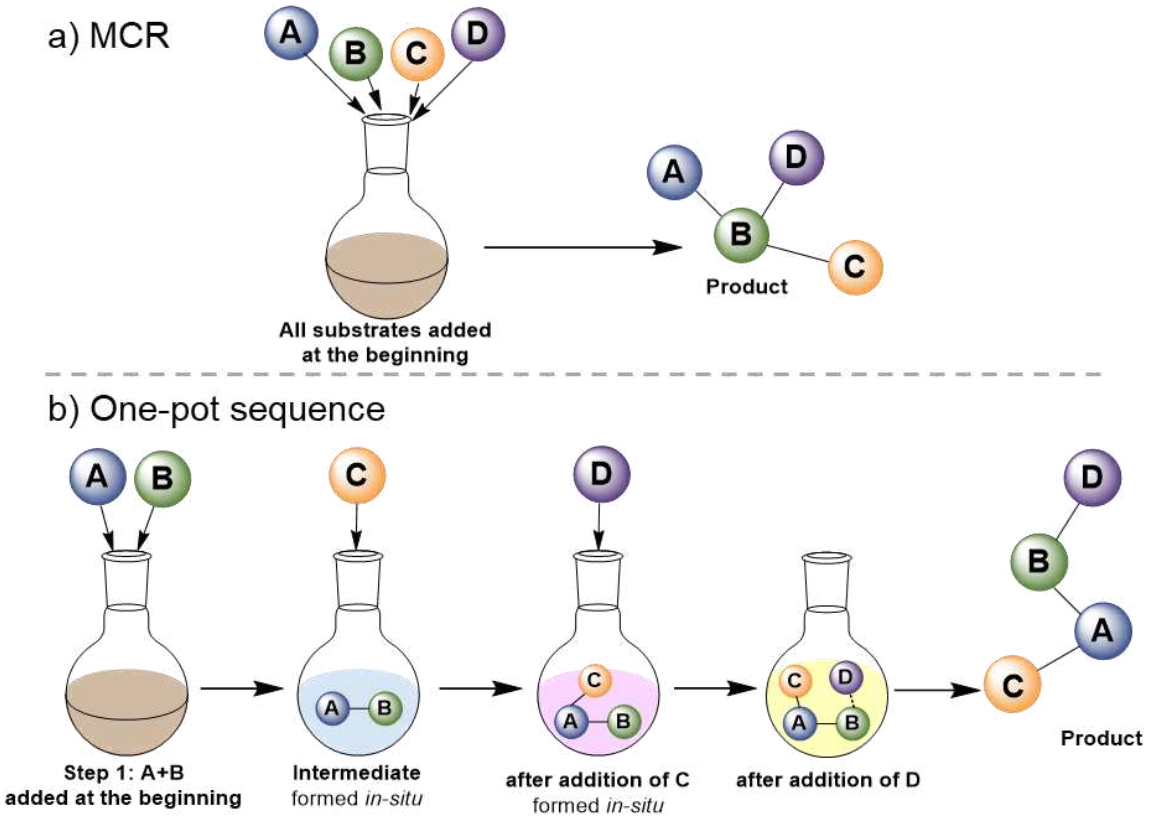


Figure 5. Schematic distinction between multicomponent and one-pot reaction sequences. a) In a MCR, all substrates, represented here as A, B, C, and D, are added at the beginning of the reaction and combine in a single operation to form the final product**.** Catalysts, bases, or other reagents that do not contribute atoms to the final product may be introduced later to change the reaction conditions. **b)** In a one-pot sequence, the substrates are introduced successively into the same reaction vessel. In the illustrated example, A and B first react to form an intermediate in situ, after which C and D are added sequentially to preserve the required order of transformations without isolating intermediates.

The first sequence I investigated experimentally was the ***Mach3*** ("Mach" notation was used to emphasize the reaction was discovered by a machine!) one-pot sequence, reported in Publication 1, yielding branched diallylic ethers through five consecutive mechanistic transformations: deprotonation of the β,γ-unsaturated alcohol, a first Williamson ether synthesis with allyl iodide, deprotonation at the allylic carbon, a [2,3]-Wittig rearrangement, and a second Williamson ether synthesis in which the resulting alkoxide was alkylated with allyl iodide.[1] The algorithm correctly classified this sequence as one-pot rather than multicomponent because excess allyl iodide can react with n-butyllithium, interfering with formation of the carbanion required for the [2,3]-Wittig rearrangement. The computational analysis therefore indicated that the first allylation had to be completed before addition of n-butyllithium. After completion of the rearrangement, a second portion of allyl iodide was introduced to alkylate the resulting alkoxide and furnish the final branched diallylic ether. *Allchemy's* C–H pKa filter also correctly identified the position expected to undergo deprotonation (**Figure 6**).[1,86]

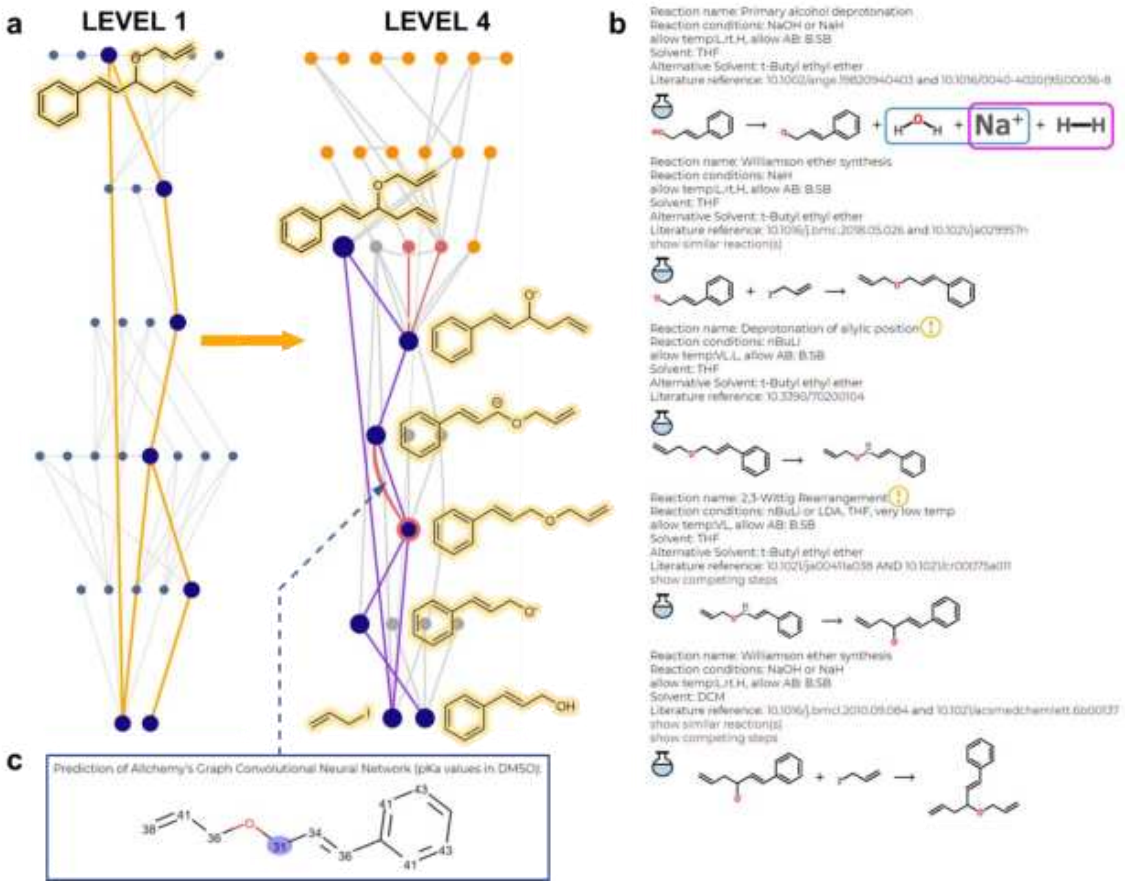


Figure 6. One-pot sequence, starting from allyl iodide and cinnamyl alcohol.[1] The sequence is presented as **a)** Level 1 and Level 4 (i.e., expanding side-products) networks and, **b)** list of mechanistic steps (both networks and the list are screenshots from *Allchemy*'s *MECH* module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative literature references. 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 p*K*a filter[86] predicting correctly the most acidic CH position in the deprotonation step. *The figure was reproduced from Publication 1*.[1]

Experimentally, cinnamyl alcohol was first treated with sodium hydride and allyl iodide in THF. The mixture was then cooled to −78 °C and treated with *n*-butyllithium, followed by HMPA and a second portion of allyl iodide. Under the optimized one-pot conditions, product **3a** was isolated in 66% yield, close to the approximately 68% yield predicted by the software. The reaction was also performed with related allylic and propargylic alcohols, furnishing products **3b–3d** in yields of 72–96% (**Figure 7**).[1]

The predicted product was considered synthetically attractive because its two terminal alkenes enabled subsequent ring-closing metathesis (RCM). This additional transformation was not compatible with the strongly basic one-pot sequence and was therefore performed separately. Treatment of **3a** with the first-generation Grubbs catalyst, followed by hydrolysis, afforded the corresponding cyclic enol ether in 75% yield (**Figure 7**, green frame). Cyclic enol ethers of this type have been used as scaffolds and intermediates in medicinal synthesis.[1,87,88] The ***Mach3*** reaction therefore provided both a test of the software's ability to recognize an experimentally necessary order of addition and a concise route to synthetically useful products.

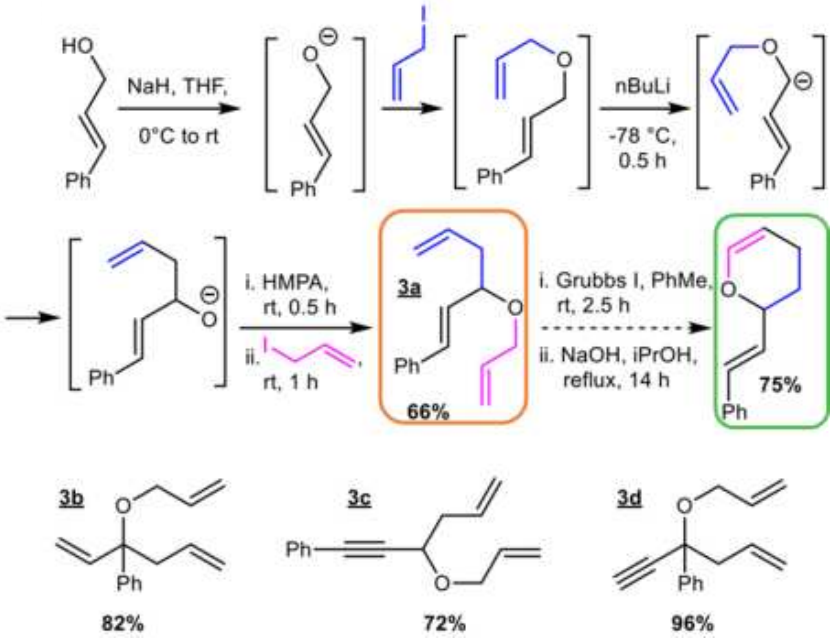


Figure 7. Computer-discovered one-pot sequence.[1] **a**) Scheme of a one-pot sequence designed by the MECH algorithm for the synthesis of branched allyl ethers. The sequence is detected as one-pot rather than the MCR because excessive allyl iodide would react with *n*-butyllithium, hampering deprotonation and subsequent Wittig rearrangement (see **Figure 6**). 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, framed in *green*. Additional derivatives **3b–3d** were prepared from allyl iodide and other commercially available β,γ-unsaturated alcohols.[1] *The figure was reproduced from Publication 1*.[1]

After this initial validation, I used the *Allchemy-MECH* module to search for new MCR sequences from a selected set of substrates. The aim was to identify new multicomponent sequences from simple reactants capable of participating in several consecutive mechanistic steps. In the search that led to the ***Mach5*** and ***Mach6*** reactions, reported in *Publication 1*, the selected substrates were allyl thiol, an aromatic aldehyde, and 1-(aryl)buta-2,3-dien-1-one. *Allchemy-MECH* identified two pathways that

shared the initial mechanistic sequence but diverged in the final stage: addition of DBU led to hydroxylated monothio-β-diketones, whereas addition of acetyl chloride (AcCl) led to 2,3-dihydrothiophenes. The predicted pathways, intermediates, by-products, and competing transformations are shown in **Figure 8** and **Figure 9** respectively.

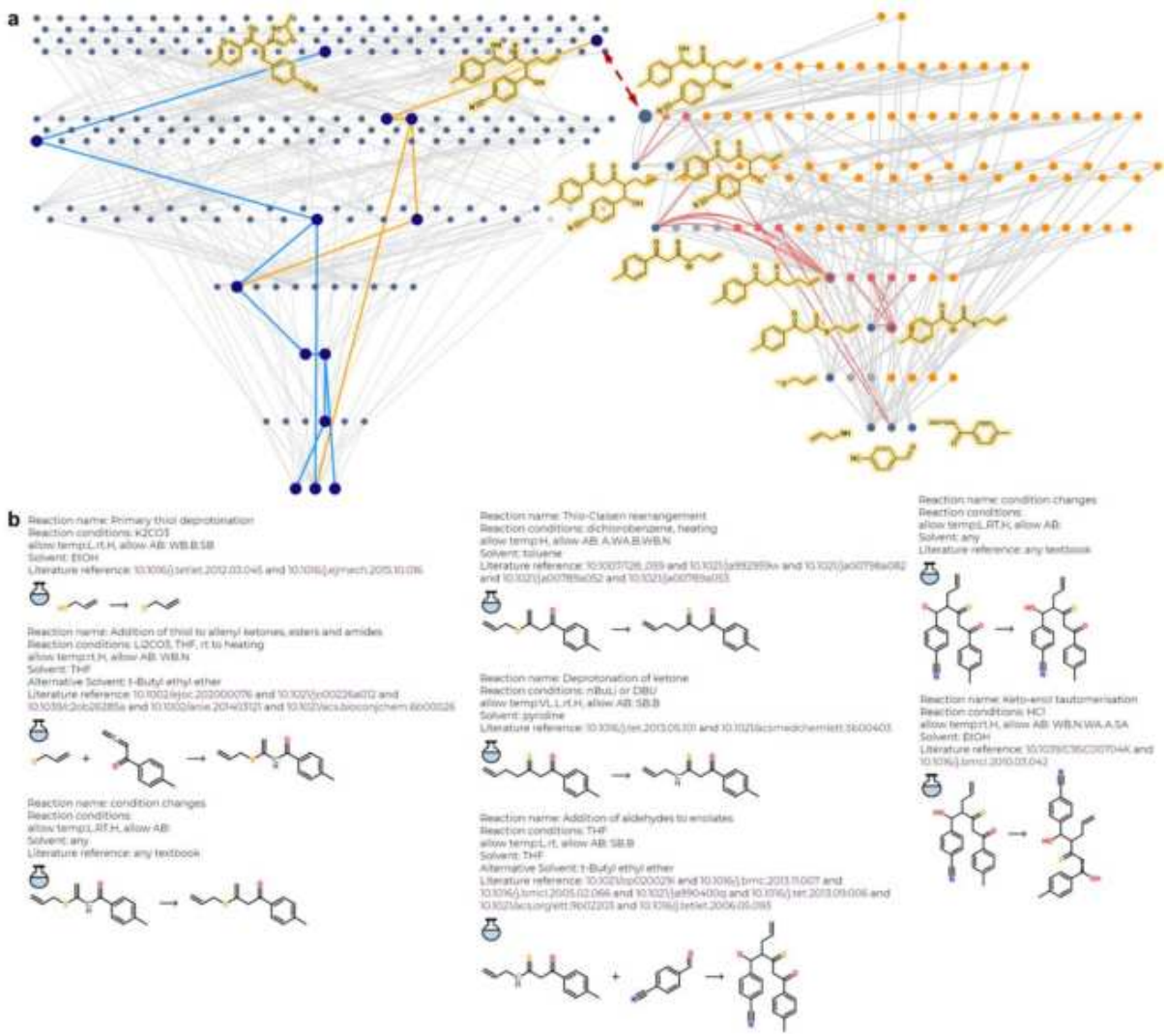


Figure 8. The first computationally-designed MCR, starting from allyl thiol, 4-cyanobenzaldehyde and 1-(p-tolyl)buta-2,3-dien-1-one, catalyzed by DBU. This MCR is presented here as **a)** Level 1 and Level 4 (i.e., expanding side-products) networks and **b)** list of mechanistic steps (both networks 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 **Figure 6**, 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. *The figure was reproduced from Publication 1.*

The calculated sequence began with deprotonation of allyl thiol and nucleophilic addition to the allenyl ketone. Isomerization of the initial adduct was followed by a thio-Claisen rearrangement to form a thioketone intermediate. From this common intermediate, the network could proceed along two different pathways depending on the subsequent reaction conditions. Importantly, all three substrates contributing atoms to the final products — allyl thiol, the aromatic aldehyde, and the allenyl ketone— were present from the beginning in both ***Mach5*** and ***Mach6***. DBU and AcCl were introduced later as non-product-incorporating reagents that directed the subsequent transformation; no atoms from either

reagent are retained in the isolated products. Upon addition of DBU, condensation of the thioketone with the aromatic aldehyde was predicted to occur at the less acidic methylene group, affording a hydroxylated monothio-β-diketone (**Figure 8**).[1] In the alternative ***Mach6*** pathway, addition of AcCl was predicted to trigger acetylation of the alcohol, followed by elimination of acetic acid to form a Knoevenagel-type adduct, thioketo–enol tautomerization, and cyclization to form a 2,3-dihydrothiophene (**Figure 9**).[1]

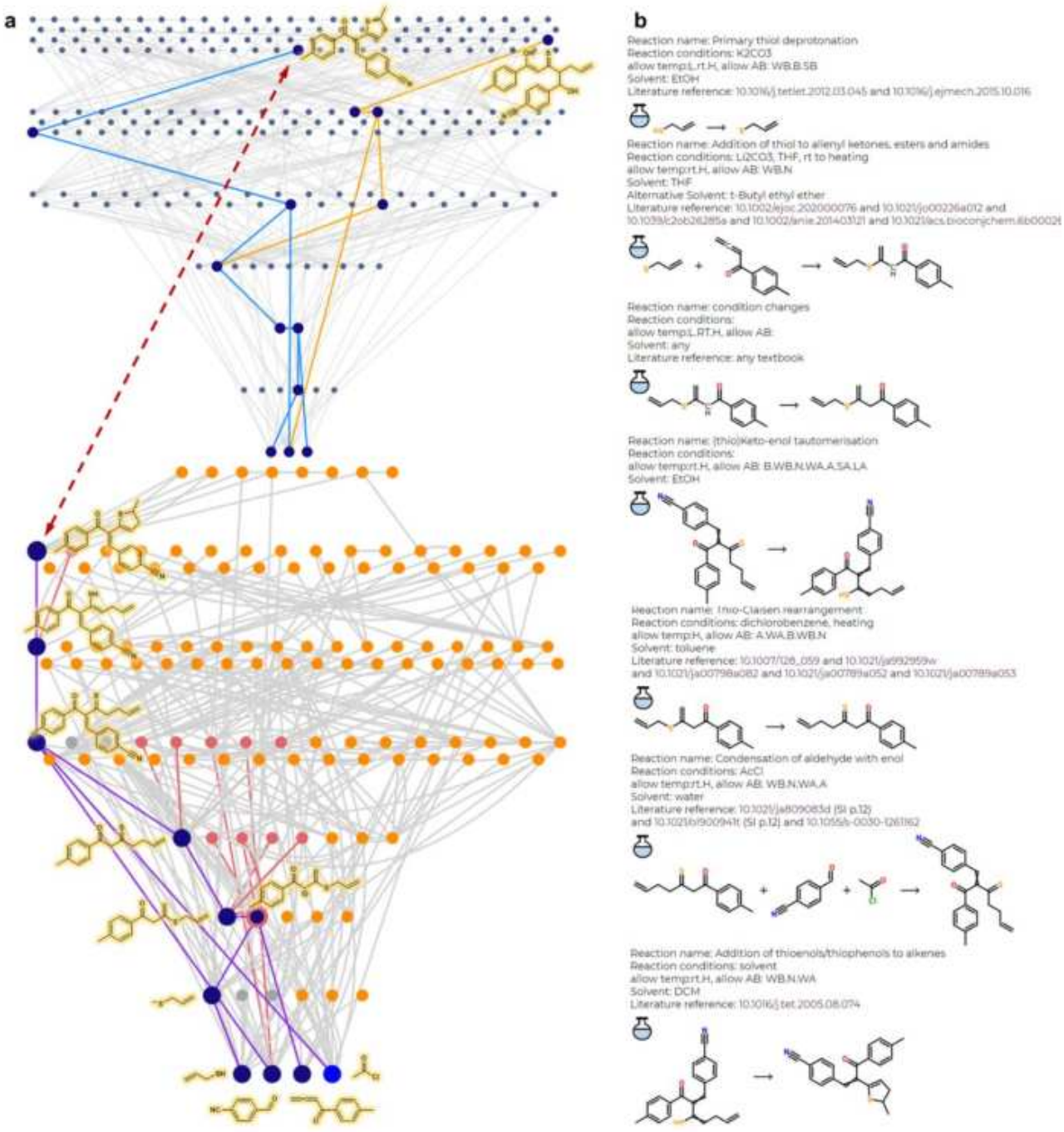


Figure 9. The second computationally-designed MCR, starting from allyl thiol, 4-cyanobenzaldehyde and 1-(p-tolyl)buta-2,3-dien-1-one, catalyzed by AcCl. This MCR is presented here as **a)** Level 1 and Level 4 networks and **b)** list of mechanistic steps (both networks and the list are screenshots from *Allchemy*'s MECH module). Every step is accompanied by name, typical reaction conditions, solvent, and hyperlinks to illustrative literature references. As in **Figure 6**, 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. *The figure was reproduced from Publication 1.*

For the base-mediated ***Mach5*** sequence, the base, solvent, concentration, temperature, reaction time, and aldehyde substituent were optimized. The best conditions used o-dichlorobenzene, whose high boiling point enabled the temperature required for the rearrangement: allyl thiol, the terminal allene, and the aldehyde were heated at 40 °C for 1.5 h and then at 180 °C for 10 min to promote the thio-Claisen rearrangement, after which DBU was added at room temperature to complete the condensation. DBU was essential, as triethylamine gave only trace product; lower concentration reduced the yield, and toluene gave incomplete conversion. Under these conditions, p-trifluoromethyl-substituted product **5a** was isolated in 57% yield, compared with the computationally estimated 43%. The reaction was extended to aromatic aldehydes bearing nitro, cyano, trifluoromethyl, and heteroaryl substituents, affording products **5c–5h** in 11–35% yield, and **5a** was further converted into the corresponding substituted pyrazole in 53% yield (**Figure 10**).[1]

The alternative ***Mach6*** pathway was tested by replacing DBU with AcCl after completion of the addition and thio-Claisen steps. The optimized procedure again used *o*-dichlorobenzene, with the initial reaction performed at 40 °C for 1.5 h and the rearrangement at 180 °C for 10 min. AcCl was then added, and the mixture was stirred at room temperature for 18 h. The *p*-cyano-substituted 2,3-dihydrothiophene **5b** was isolated in 14% yield, close to the computer-predicted yield of 12% (**Figure 10**). Increasing the AcCl loading, adding drying agents, using pivaloyl chloride, or raising the temperature of the final stage did not improve the yield.[1]

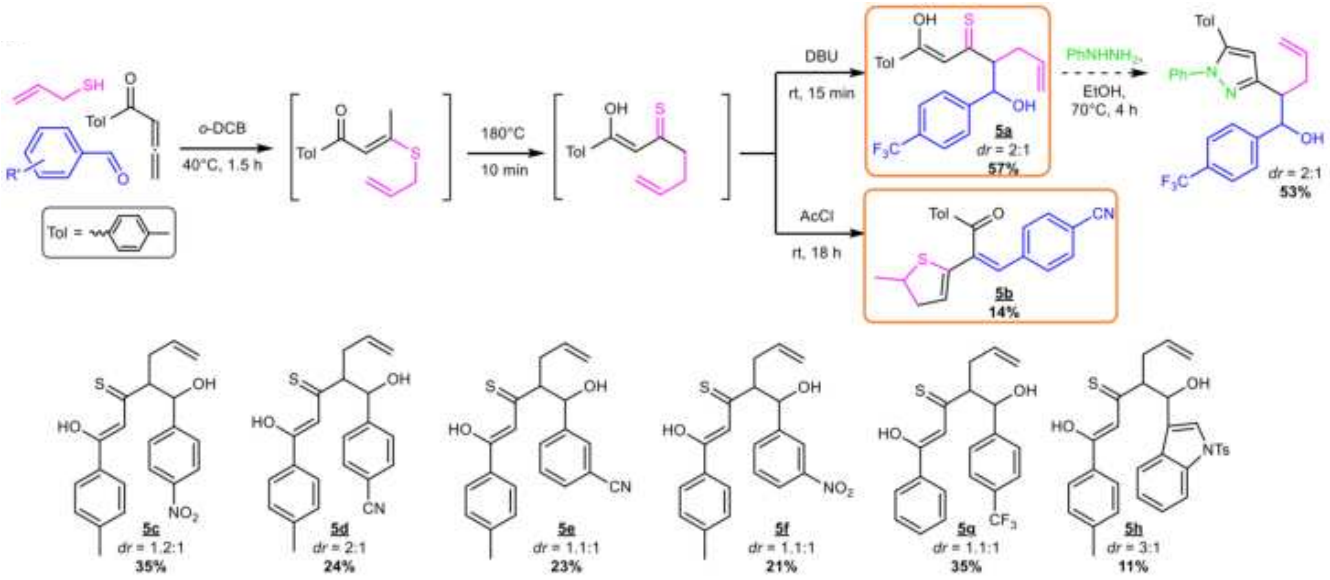


Figure 10. Conditionally divergent MCRs producing hydroxylated monothio-β-diketones under basic conditions and 2,3-dihydrothiophenes after addition of AcCl. The upper pathway gives unsaturated hydroxylated monothio-β-diketones, which exist predominantly in the thioenol tautomeric form, whereas the lower pathway gives 2,3-dihydrothiophenes after addition of AcCl in the final stage. Non-isolated intermediates are shown in brackets. The isolated products 5a, originally predicted for the upper pathway, and 5b, the highest-yielding product from the lower pathway, are framed in orange. The monothio-β-diketone 5a was subsequently reacted with phenylhydrazine, shown in green, to afford a substituted pyrazole. Additional products 5c–5h were prepared using the upper pathway. *Reproduced from Publication 1.*[1]

The supporting experiments also identified several ***Mach6*** by-products arising from competing thiol addition, thiolysis, tautomerization, and condensation pathways represented in the *Allchemy* network (**Figure 11**). These observations showed that the computational network captured not only the elementary transformations but also competing transformations contributing to the lower preparative yield.

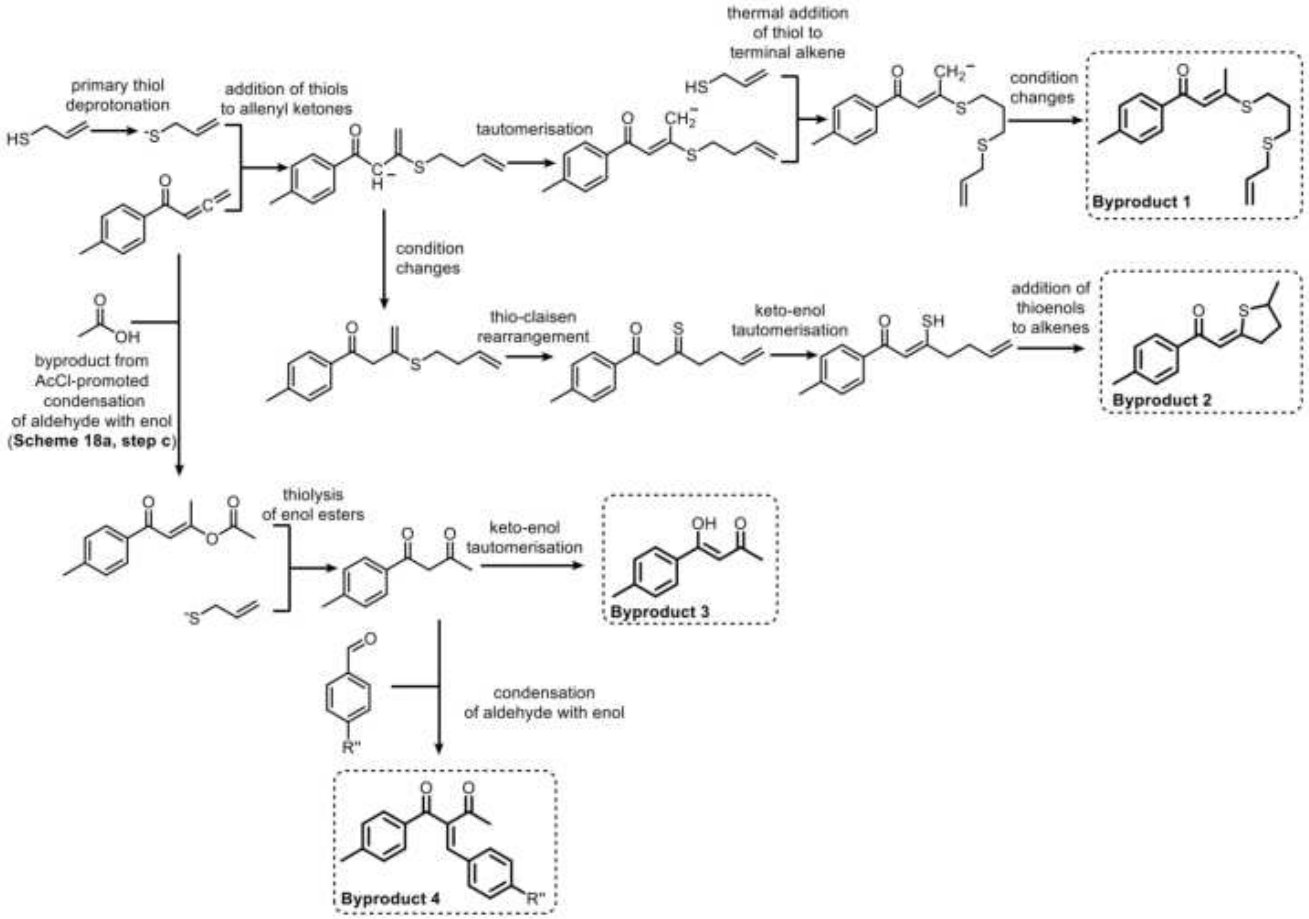


Figure 11. Identified byproducts formed during the reaction; all mechanistic steps are consistent with *Allchemy*'s predictions. *The figure was reproduced from Publication 1*.[1]

Together, the ***Mach3***, ***Mach5***, and ***Mach6*** studies demonstrated complementary applications of *Allchemy-MECH*. For ***Mach3***, the analysis identified a cross-reactivity constraint that required sequential reagent addition. For ***Mach5*** and ***Mach6***, it predicted condition-dependent divergence from the same starting materials into two product classes. Although several ***Mach5*** products were obtained in moderate to low yields, structurally complex monothio-β-diketones were assembled from three simple components in a single operation. The particularly low ***Mach6*** yield was attributed partly to competing pathways that generated several identified by-products. Overall, the predicted transformations agreed with experiment, showing that mechanistic network analysis can guide reaction discovery while accounting for reagent compatibility, pathway divergence, and competing reactions. Such computational approaches depend on chemical information being represented accurately in a structured, machine-readable form, a requirement that also motivated the development of the *AbSynth* pathway dataset.

3.2. Construction, quality control, and analysis of the *AbSynth* pathway dataset

Whereas *Allchemy-MECH* used structured mechanistic rules to generate new reaction hypotheses[1], *AbSynth* addressed a complementary challenge: the lack of openly available, reliable and machine-readable data describing complete synthetic routes at the highest level of complexity. Recognizing this limitation, we embarked on a multiyear effort to assemble the first openly available, comprehensive collection of total syntheses. An additional and fundamental motivation was to use this collection to learn the rules that govern the logic of total synthesis, beyond the Corey heuristics.[68]

Published total syntheses were reconstructed as connected pathways in which each molecular structure was encoded as a SMILES string and linked to its corresponding reaction step and pathway branch. By preserving reaction order, branching, convergence, stereochemistry, and the relationships among intermediates, *AbSynth* converted established synthetic strategies into a format suitable for large-scale computational analysis.[2]

The final dataset comprised 3,077 total syntheses and 59,712 individual steps, including 50,425 unique steps. The routes covered almost a century of synthetic chemistry and were released both as machine-readable records for computational analysis and through the *AbSynth* portal for interactive inspection of individual pathways (**Figure 12** and **Figure 13**).[67]

Target	DOI	Rx_Index	Substrates SMILES	Product SMILES	Reaction SMILES
Isomatrine	10.1021/jacs.2c09804	8001	COC(=O)CCCC=C	COC(=O)CCCC=O	COC(=O)CCCC=C>>COC
		8002	COC(=O)CCCC=O	COC(=O)CCCC=C(Br)Br	COC(=O)CCCC=O>>COC
		8003	COC(CCC/C=C(Br)\Br)=	COC(=O)CCC\C=C(/Br)C	COC(CCC/C=C(Br)\Br)=C
		8004	COC(=O)CCC\C=C(/Br)(	NC(=O)CCC\C=C(/Br)C=(	COC(=O)CCC\C=C(/Br)C
		8005	COCCl.NC(=O)CCC\C=C(	COCNC(=O)CCC\C=C(/B	COCCl.NC(=O)CCC\C=C
		8006	COCNC(CCC/C=C(C=C)	COCNC(=O)CCC\C=C(/C	COCNC(CCC/C=C(C=C)\
		8006d	O=C1N(CCCC#C)C(=O)(	NCCCC#C	O=C1N(CCCC#C)C(=O)C
		8006c	NCCCC#C	CC(C)(C)OC(=O)NCCCC#	NCCCC#C>>CC(C)(C)OC
		8006b	CC(C)(C)OC(=O)NCCCC	B\C=C\CCCNC(=O)OC(C	CC(C)(C)OC(=O)NCCCC
		8006a	CC(C)(O)C(C)(C)O.B\C=	CC(C)(C)OC(=O)NCCC\C	CC(C)(O)C(C)(C)O.B\C=C
		8007	COCNC(=O)CCC\C=C(/I	[H][C@@]12CCCC(=O)N	COCNC(=O)CCC\C=C(/C
		8008	[H][C@@]12CCCC(=O)I	[H][C@@]12CCCC(N1CC	[H][C@@]12CCCC(=O)N
		8009	[H][C@@]12CCCC(N1CI	[H][C@]12CCCN3CCC=C	[H][C@@]12CCCC(N1CC
		8010	[H][C@]12CCCN3CCC=(	[H][C@]12CCCN3CCC[C(	[H][C@]12CCCN3CCC=C

Rx_Name	Conditions	Author Name	Year
Ozonolysis of alkenes	O3.EDCM.-60°C to rt	Sherburn	2022
Olefination	CBr4.PPh3.DCM. 5°C.66%		
Negishi coupling	Pd(PPh3)4.THF.rt.80%		
Amidation	NH4Cl.AlMe3.toluene.60°C.95%		
N-Alkylation	NaHMDS.toluene.rt.34%		
Suzuki coupling	PdCl2(dppf).Ba(OH)2.8H2O.THF.H2O.rt.82%		
Phthalimide Removal	N2H4-H2O.EtOH.reflux to rt to reflux		
N-Boc Protection	Boc2O.rt.93% (2 steps)		
Borylation	BH3.THF.cyclohexene.0°C		
Boronic ester Formation	neat.rt.92% (2 steps)		
hetero-intramolecular Diels Alder	CF3CO2H.CHCl3.rt.50%		
BOC Removal	CF3COOH.DCM.rt		
hetero-intramolecular Diels Alder	HCHO.H2O.100°C.74% (2 steps)		
Hydrogenation	H2.PtO2.AcOH.25°C.25%		

Figure 12. Machine-readable representation of the *AbSynth* pathway data. Representative extract showing a total-synthesis pathway encoded as reaction records. Substrates and products are represented as SMILES strings, while each transformation is encoded as reaction SMILES and accompanied by its reaction name, conditions, pathway index, and bibliographic information.[2]

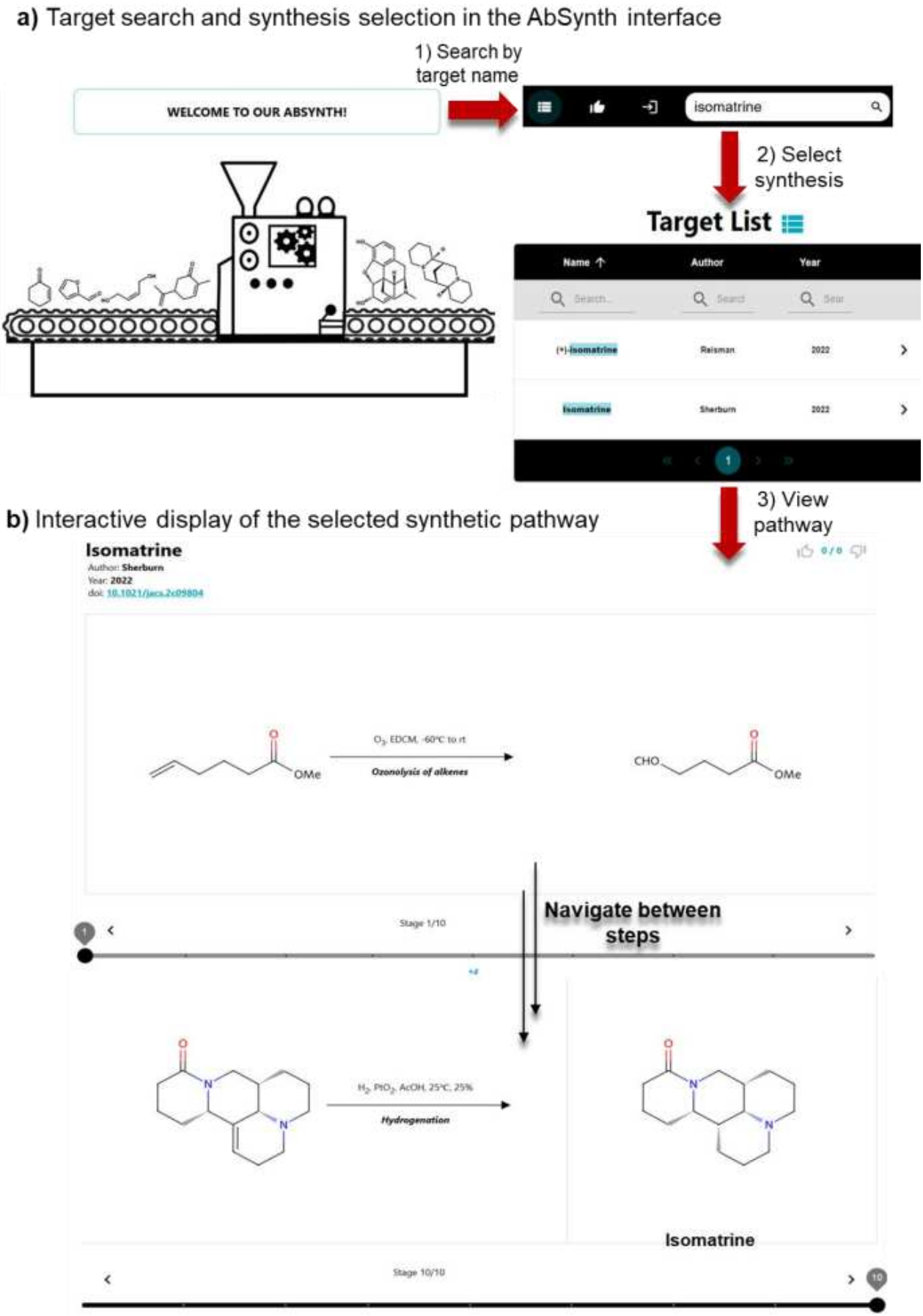


Figure 13. Interactive access to the *AbSynth* pathway collection. a) *AbSynth* WebApp used to search for a target from the available published syntheses. *AbSynth* is based on the curated machine-readable pathway data illustrated in **Figure 12. b)** Interactive display of the selected isomatrine synthesis, illustrated by its first and final reaction steps. Users can navigate through the complete pathway and inspect the corresponding structures, transformations, and reaction information.[67]

As part of my doctoral research, I curated and quality-controlled the synthetic pathways, corrected structural and stereochemical inconsistencies, coordinated the work of the curation team, and participated in the analyses of statistical trends ultimately described in our *J. Am. Chem. Soc.* publication on *AbSynth*.[2,67] Because over 30 contributors participated in compiling the dataset, shared representation rules were required to ensure that the pathways were encoded consistently. I trained the contributors to apply these rules during pathway curation. The rules covered the representation of salts, protecting groups, stereochemistry, multistep operations, branched and convergent pathways, intermediates reported in the source publications but not isolated experimentally, and generation of their SMILES strings using ChemDraw.[2]

Quality control was carried out iteratively. Curated pathways were checked against the source publications, and those containing inconsistencies were returned for correction (**Figure 14**). Over successive correction rounds, the number of pathways with detected errors decreased, while the number accepted as structurally consistent increased. The review focused particularly on errors that could propagate through a route, including incorrect stereochemistry, misplaced double bonds, missing reaction steps, missing atoms, and intermediates assigned to the wrong branch. Such errors can compromise the structural and stereochemical continuity of the complete pathway.[2]

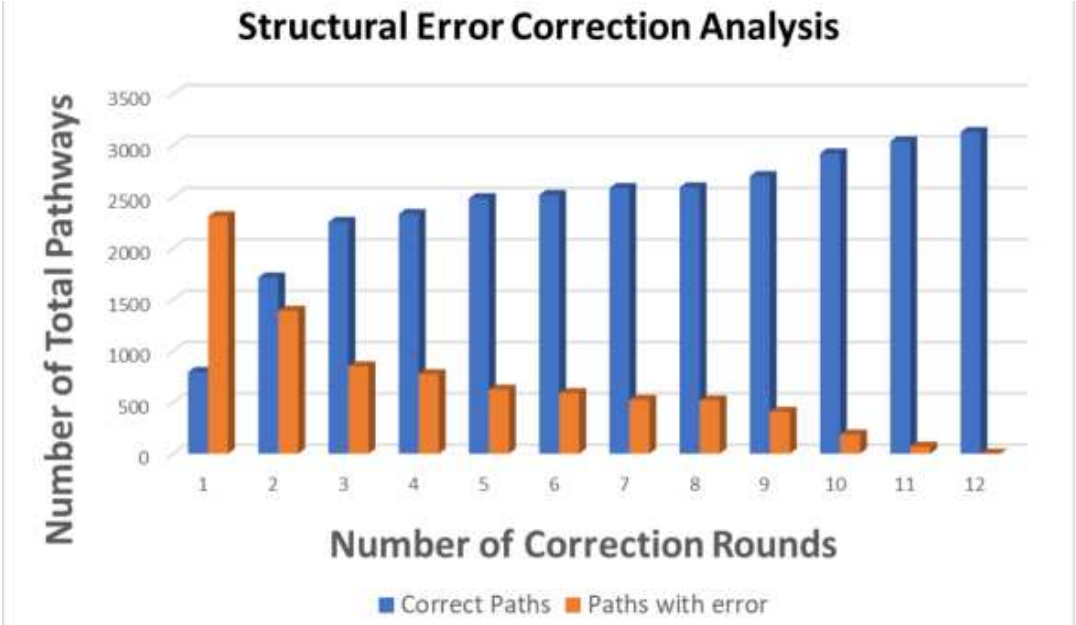


Figure 14. Error analysis of *AbSynth*'s synthesis pathways.[2] The x-axis shows the number of correction rounds, and the y-axis shows the number of pathways. Blue bars indicate pathways classified as correct (i.e., all steps correct), while orange bars indicate pathways with detected errors. Across 12 rounds of iterative manual correction and re-evaluation, the number of validated pathways progressively increased, while the number of error-containing pathways decreased steadily. *Reproduced from Publication 2*.

Given the scale of the dataset and the involvement of multiple curators, an independent expert quality assessment was essential both to estimate the remaining error rate in *AbSynth* and to benchmark its accuracy against an established commercial database. Ten randomly selected *AbSynth* pathways were therefore compared with the corresponding Reaxys records and source publications.[2,89] Among

170 manually inspected steps, six *AbSynth* entries contained errors, corresponding to 3.5%, whereas 26 Reaxys entries contained errors, corresponding to 15.3%. The identified *AbSynth* errors were corrected, and the remaining error rate in the full dataset was estimated at approximately 3–4%. This comparison provided a quantitative estimate of the uncertainty remaining after curation and showed that the curated *AbSynth* pathways had a comparatively low observed error rate after direct reconstruction from source publications and review (**Table 2**).

Table 2. Step-level error-rate comparison of Reaxys and *AbSynth* across ten randomly selected total-synthesis pathways. The table summarizes the number of evaluated steps, error-containing steps, error rates, and major error types identified in each dataset. *Reproduced from Publication 2*.[2]

Dataset	Total evaluated steps	Steps with errors	Error rate
Reaxys	170	26	15.3%
AbSynth	170	6	3.5%

Path 1	Target Compound	Total Steps	Reaxys errors	Reaxys error type	AbSynth errors	AbSynth error type
1	Piperarborenine B	7	0	N/A	0	N/A
2	(−)-Himandrine	31	13	Wrong stereo isomer	1	Wrong position of one functional group
3	Pleocarpenone	19	7	Missing steps: metal-complex structures	0	N/A
4	Tabersonine	15	0	N/A	3	Missing steps
5	Vineomycinone B2 Methyl Ester	21	1	Missing step / wrong stereoisomer	0	N/A
6	(+)-Salimabromide	18	0	N/A	1	Wrong stereoisomer
7	Plocamenone	7	2	Missing structure / wrong geometric isomer	1	Wrong position of one functional group
8	Xishacorene B	10	1	Wrong stereoisomer	0	N/A
9	11-Gorgiacerol	14	0	N/A	0	N/A
10	Brevisamide	28	2	Wrong stereoisomer	0	N/A
Total		**170**	**26**		**6**	

With the quality-controlled *AbSynth* collection established, I and other members of the collaboration used it to investigate general patterns in total-synthesis design. These analyses showed that molecular complexity does not increase monotonically throughout a total synthesis. Many steps leave the carbon skeleton largely unchanged but remain necessary because they introduce, remove, or modify functional groups required for later bond-forming or skeletal transformations. The study also examined how many consecutive steps can occur without changing the carbon skeleton and how long particular functional groups are retained before being modified or removed. *AbSynth* also allowed, for

the first time, to validate the applicability of the celebrated Corey heuristics – surprisingly, while the rules for the forbidden disconnections were reflected in the actual total syntheses, those suggesting the loci of preferred disconnections had very low predictive power. In other words, the ways in which Corey's rules were supposed to guide synthetic design are not reflected in the actual total syntheses performed over the past decades.[2,68]

Quantitative analysis of the non-skeletal modification sequences showed that they are usually short: in 90–95% of cases, they contain no more than four consecutive steps. The analysis of functional-group persistence further showed that some groups tend to be transformed or protected soon after their introduction, whereas more robust functionalities can remain unchanged over several subsequent steps. Together, these findings emphasized the importance of multistep reasoning in total-synthesis planning.[2]

Overall, the main outcome of my contribution is a quality-controlled, machine-readable dataset of complete total-synthesis pathways whose reliability supports route-level analysis and challenges the accepted views on synthesis design. By preserving accurate structures, stereochemistry, step order, and pathway connections, *AbSynth* enables systematic investigation of retrosynthetic strategy, functional-group management, bond-forming logic, and the development and evaluation of future synthesis-planning methods. Analysis of these pathways further showed the importance of multistep reasoning in total-synthesis design and revealed that several established heuristics for preferred retrosynthetic disconnections have limited predictive power when tested across thousands of complete syntheses.

3.3. Robotic mapping of E1 and S_N1 reaction hyperspaces in fluorenol systems

The next experimental stage of my doctoral research focused on robotic reaction-hyperspace mapping. Chemical reactions are usually examined under only a limited set of conditions, even though their outcomes are governed by several interacting variables. This raises broader questions that conventional optimization studies rarely answer: Can changing the conditions switch the reaction between different major products? Can previously unrecognized compounds appear only within narrow regions of the reaction space? Can correlations among product and by-product distributions reveal hidden intermediates or competing pathways? Can reaction outcomes be mapped in sufficient detail to reconstruct the underlying reaction network rather than merely identify one optimal condition? Reaction-hyperspace mapping addresses these questions by systematically measuring reaction composition across multidimensional combinations of concentration, stoichiometry, temperature, and other experimental variables. It therefore extends reaction study beyond optimization, providing a more complete view of chemical reactivity and revealing synthetic possibilities that may remain hidden in traditional bench chemistry.

A robotic workflow developed by our group enabled systematic reaction-hyperspace mapping. The first version of the house-built, low-cost Robowski platform combined automated reagent dispensing, sealed 54-vial plates, controlled heating, and automated UV–Vis analysis, allowing up to approximately 1,000 reactions per day. Reference spectra of identified substrates, products, and by-products were recorded at known concentrations for calibration. The robot then dispensed reagents across different concentrations and stoichiometries (**Figure 15a**). After heating in a preheated oven or an Eppendorf ThermoMixer® with shaking for the selected reaction time, the crude mixtures were analyzed automatically using a NanoDrop UV–Vis spectrophotometer. Each crude spectrum was fitted as a combination of the reference spectra to estimate component concentrations and yields by spectral unmixing (**Figure 15b,c**).[3]

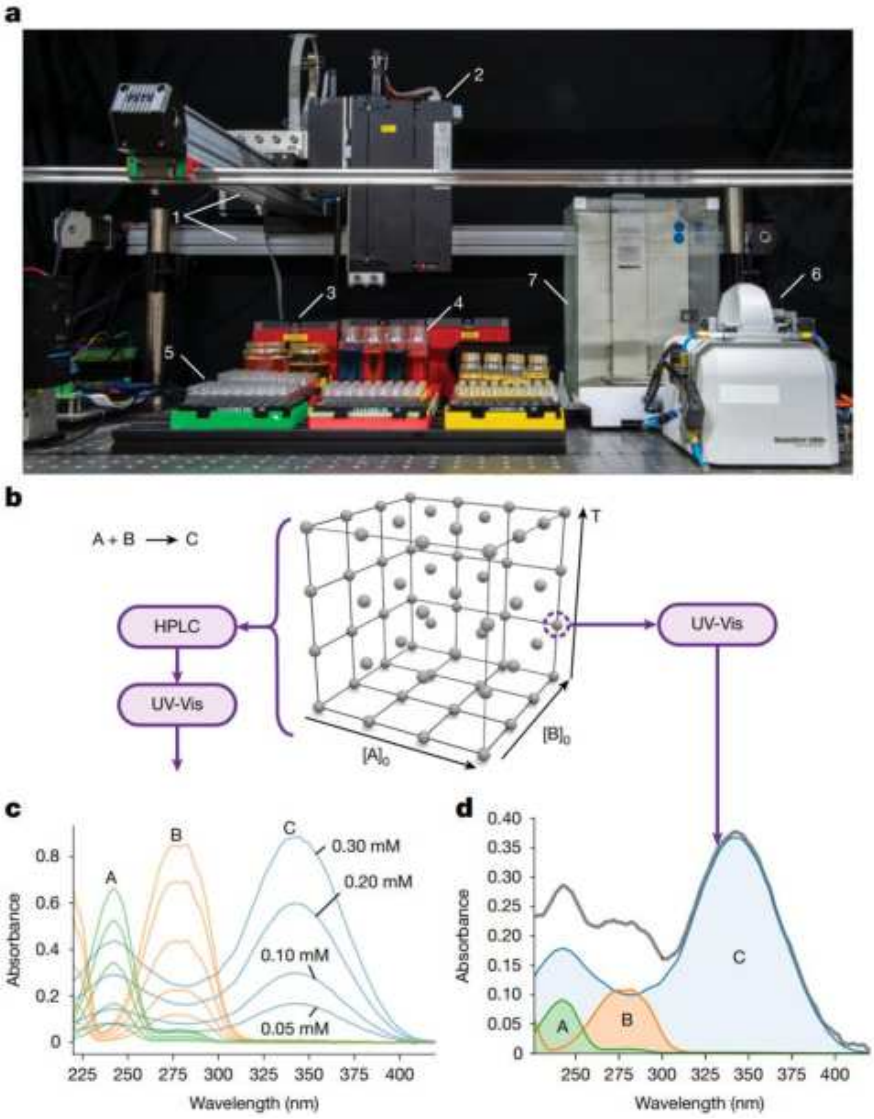


Figure 15. Robotic platform quantification workflow for reaction-hyperspace mapping.[3] **a**) Parts of the house-built, inexpensive (about $25K) system: **(1)** horizontal gantry; **(2)** liquid-handling module; **(3)** racks of pipette tips; **(4)** stock solutions; **(5)** 54-vial plates; **(6)** UV-Vis spectrophotometer; **(7)** balance for weighing of solutions (see Supplementary Videos 1 and 2 of Publication 3). **b)** Hypothetical A + B → C reaction space defined by initial concentrations $[A]_0$ and $[B]_0$ and temperature T. **c)** Reference UV–Vis spectra of isolated reaction components recorded at different concentrations. **d)** Quantification of an individual crude mixture by fitting the measured spectrum as a combination of the reference spectra. *Reproduced from Publication 3.*[3]

My work in robotic chemistry began with two mechanistically established carbocationic transformations: the E1 dehydration of 9-butyl-9H-fluoren-9-ol and the S_N1 substitution of 9-phenyl-9H-fluoren-9-ol. These reactions were selected as relatively simple model systems because their mechanisms and expected products were well defined, allowing the behavior of basic reaction hyperspaces to be examined before the workflow was applied to more complex systems.

For the E1 reaction, 9-butyl-9H-fluoren-9-ol (**13a**) was treated with HBr in acetonitrile. Protonation of the alcohol was followed by loss of water to form a fluorenyl carbocation, which then underwent deprotonation to give the alkene product (**Figure 16a**). The hyperspace comprised 775 conditions defined by the initial concentrations of the alcohol and HBr and by temperature, with all reactions analyzed after 4 h. The product yield changed smoothly across the investigated conditions and increased toward a single global maximum (**Figure 16c**). No significant S_N1 product or unexpected major by-product was detected; the mixtures contained mainly unreacted alcohol and the alkene product.

For the S_N1 reaction, 9-phenyl-9H-fluoren-9-ol was treated with HBr in 1,4-dioxane containing acetic acid. The phenyl-substituted substrate (**14a**) was chosen because alkene formation by elimination was prevented by its structure, allowing the substitution pathway to be studied without competing E1 reactivity. Protonation and loss of water generated a stabilized carbocation, which was trapped by bromide to form the substitution product (**Figure 16b**). The hyperspace contained 930 conditions defined by the initial concentrations of the alcohol and HBr and by temperature, with the reactions analyzed after 48 h. As in the E1 system, the product yield varied smoothly across the hyperspace and generally increased toward the global maximum, although a shoulder maximum was observed at low concentrations of the limiting alcohol. No additional major products were detected (**Figure 16d**).

Importantly, the E1 and S_N1 hyperspace experiments provided information beyond the distributions of product yields. The experimental yields were analyzed using mathematical models based on the proposed reaction mechanisms. For the E1 system, the corresponding kinetic equations were solved numerically, and the model parameters were adjusted so that the calculated product yields matched the experimental data across the hyperspace. This fitting provided estimates of parameters including reaction enthalpy (ΔH), reaction entropy (ΔS), and enthalpy of activation ($\Delta H^{\ddagger}$). For the S_N1 system, which had reached thermodynamic equilibrium above 26 °C, an equilibrium model was fitted to the experimental data at each temperature, and the temperature dependence of the fitted equilibrium constants was used to estimate ΔH and ΔS. Even though each hyperspace was measured at only one reaction time, the large number of yield measurements obtained at different starting concentrations helped restrict the range of model parameters consistent with the experimental data. The numerical

modelling and fitting were performed by co-author Yaroslav I. Sobolev, while I contributed to adapting the E1 and S_N1 reactions to the robotic platform (**Figure 16e,f**).[3]

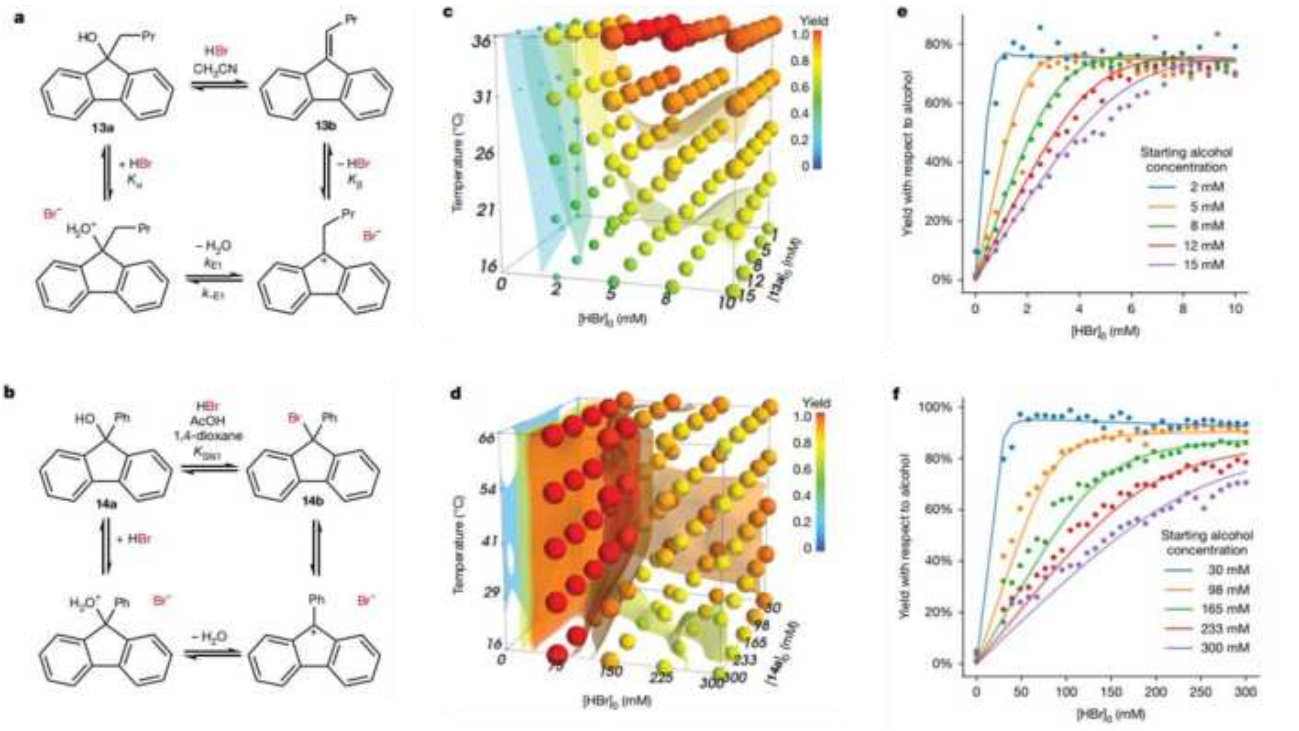


Figure 16. Mechanisms and yield distributions across the reaction spaces of E1 and S_N1 reactions.[3] **a,b)** Schemes of E1 **(a)** and S_N1 **(b)** reactions and their mechanisms. For the E1, 9-butyl-9H-fluoren-9-ol substrate **13a** and for the S_N1, to prevent elimination, 9-phenyl-9H-fluoren-9-ol **14a** were used. **c,d)** Corresponding yield distributions of products **13b** and **14b**. For clarity, only 125 conditions per cube are shown. Marker colours and sizes are proportional to yields. Yield color scales are the same for both cubes. Surfaces of equal yield are at 20%, 40%, 70% and 85%. **e,f)** Kinetics model (curves) fitted to experimental yields with respect to alcohol (circles) for E1 and S_N1 reactions. To better visualize the quality of fits, only two-dimensional snapshots of the entire dataset are shown, here T = 21 °C for E1 and T = 36 °C for S_N1 (for similar plots at all temperatures, see Supplementary Information **Sections 5.1** and **5.2** of ref.[3]). Horizontal axis has the initial concentrations of HBr, with the colour corresponding to different initial concentrations of the alcohol indicated in the keys. *Reproduced from Publication 3.*[3]

Together, these studies established that reaction-hyperspace mapping could provide quantitative kinetic and thermodynamic information in addition to mapping product yields. They also provided my first practical experience in designing and interpreting robotic reaction-hyperspace experiments and formed the basis for my subsequent work on more complex systems.

The E1 and S_N1 studies relied on UV–Vis detection because the substrates and products had sufficiently distinct absorption spectra for reliable spectral unmixing. This restricted the approach to compounds with suitable optical signals. To extend robotic hyperspace mapping to a broader range of organic reactions, the next stage incorporated automated benchtop NMR analysis, which provides compound-specific structural information without requiring distinct absorption spectra. Hidden reaction networks in textbook olefin brominations.

3.4. Hidden reaction networks in textbook olefin brominations

Olefin bromination was selected as the first reaction class investigated using automated benchtop NMR because it is one of the foundational transformations in organic chemistry (dating back to Berzelius' work from 1846) and is usually presented as a straightforward addition of bromine across a carbon–carbon double bond. Vicinal dibromides are also useful intermediates for elimination, substitution, cyclization, and further functional-group transformations.[90] Their apparently simple textbook representation made these reactions an informative test case for determining whether systematic variation of concentration, reagent loading, temperature, and reaction time could reveal additional products and competing pathways. The aim was therefore not only to identify favorable conditions, but also to determine whether these established transformations behaved as broader, condition-dependent reaction networks.[4]

I contributed to the design of the bromination hyperspaces, carried out the robotic experiments, and analyzed the resulting NMR spectra and product-distribution maps. The DPE hyperspace revealed five distinct products, while in the BDA system I isolated and characterized approximately thirteen (sic!) products formed under different conditions. On the basis of these experimental results, the proposed mechanisms were refined and the underlying reaction networks were reconstructed.

To analyze the large number of crude mixtures generated in these experiments, the study used a house-built automated benchtop NMR platform developed by other members of the group. Reaction mixtures were transferred automatically from the 54-vial plates into NMR tubes and measured using an 80 MHz benchtop spectrometer. After each measurement, the tube was removed, washed, dried, and returned for reuse, while a second tube was prepared and measured in parallel. This alternating workflow allowed one 54-vial plate to be analyzed in approximately 6 h (**Figure 17**).[4]

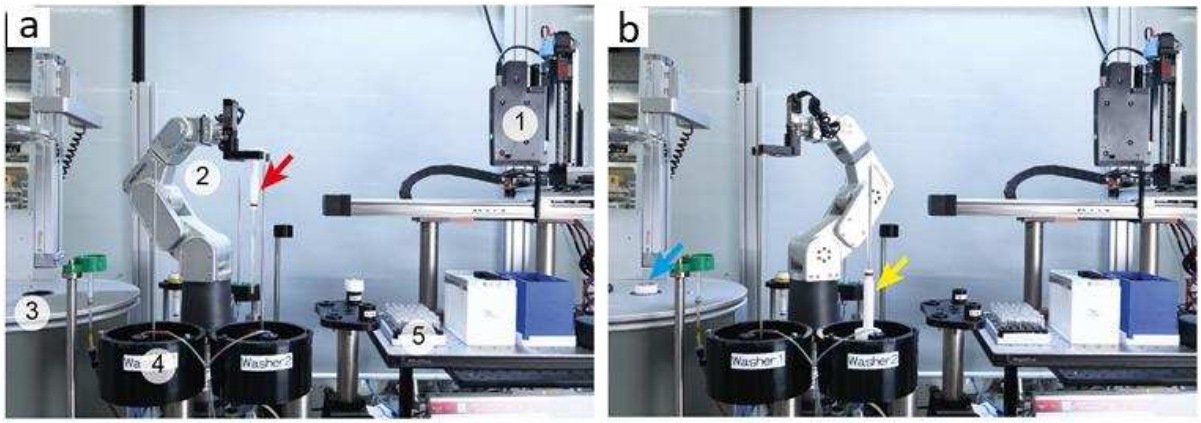


Figure 17. Automated benchtop NMR setup.[4] **a)** Photo of the automated-NMR system illustrating how the robotic arm is transferring an NMR tube (red arrow) containing reaction mixture to the spectrometer: **(1)** liquid handler used to transfer crude mixtures from reaction vials into NMR tubes; **(2)** robotic arm for transferring the tubes among different locations; **(3)** bench-top 80 MHz NMR spectrometer; **(4)** one of two washing stations for cleaning the tubes after each measurement; **(5)** 54-vial plate containing reaction crudes. **b)** The same NMR setup where one tube/sample is being measured (blue arrow) while the other tube (yellow arrow) is placed upside down onto the nozzle supplying the washing solvent. *Reproduced from Publication 4.*[4]

The acquired spectra were processed automatically using characteristic resonances selected for each component. Each selected resonance was fitted with a pseudo-Voigt line shape, allowing its position, width, and intensity to be adjusted to match the measured NMR signal. The fitted peak area was then converted into the amount of the corresponding compound using calibration data obtained from reference solutions of known concentration. This enabled consistent quantification across hundreds of crude reaction spectra without manual integration of every sample. The fitting workflow was developed and implemented by another member of the collaboration.[4]

Using the automated NMR workflow, I first investigated the bromination of 1,1-diphenylethylene (DPE). A total of 833 conditions were examined by varying the initial concentrations of DPE, Br_2, and ionic additives. The resulting hyperspace showed that the known vicinal dibromide **P1** and vinyl bromide **P2** dominated different regions, while three additional products appeared under more specific conditions (**Figure 18**). Under DPE-rich and bromine-poor conditions, the hydrobromination product **P3** reached 74% yield. In contrast, excess bromine favored the previously unrecognized tribromide **P4**, which became the major product with an NMR-determined yield of 91%, while the vinylidene dibromide product **P5** was detected only in trace amounts. These results showed that changes in the initial composition could redirect the same substrate toward different major products. Because these products differed substantially in stability, their structures were established using complementary approaches. **P3** could not be isolated because of rapid dehydrobromination and was confirmed *in situ* by ^{1}H NMR titration; P4 was characterized through its hydrolysis product **P4′**, whereas trace **P5** was identified directly by ^{13}C NMR.[4]

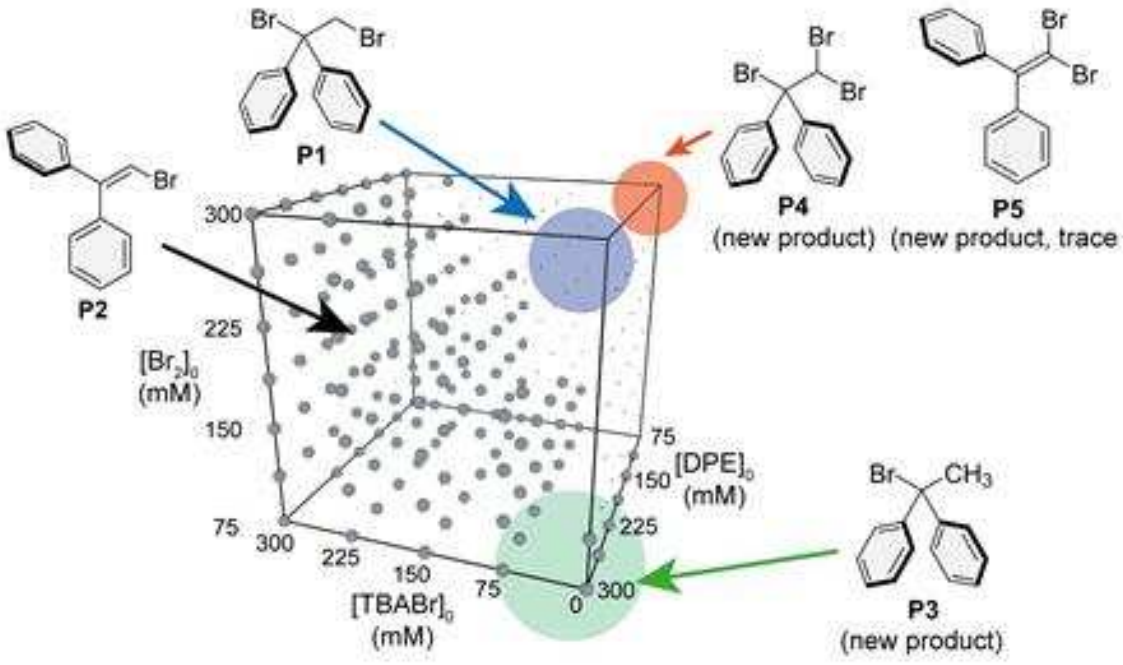


Figure 18. Product distributions for the bromination of DPE.[4] The hyperspace was defined by the initial concentrations of DPE, Br_2, and TBABr. The highlighted regions indicate conditions favoring the vicinal dibromide **P1**, vinyl bromide **P2**, hydrobromination product **P3**, tribromide **P4**, and trace dehydrobromination product **P5**. *Reproduced from Publication 4.*[4]

The product distributions, together with targeted NMR experiments, were used to reconstruct the proposed reaction network (**Figure 19**). Bromination of DPE first forms the π-complex DPE·2Br₂, followed by ionization to the β-bromocarbenium–tribromide ion pair **C1⁺·Br₃⁻.** Collapse of this ion pair gives the vicinal dibromide P1, whereas dissociation of the ion pair followed by β-deprotonation of **C1**⁺ by Br⁻ gives the vinyl bromide **P2** together with HBr. The HBr formed in this pathway can add to DPE to produce **P3**. Further bromination of **P2** proceeds through the corresponding π-complex and β,β-dibromocarbenium–tribromide ion pair **C3⁺·Br₃⁻**, whose collapse gives **P4**. **P5** can arise either by deprotonation of **C3**⁺ or by elimination of HBr from **P4**.[4]

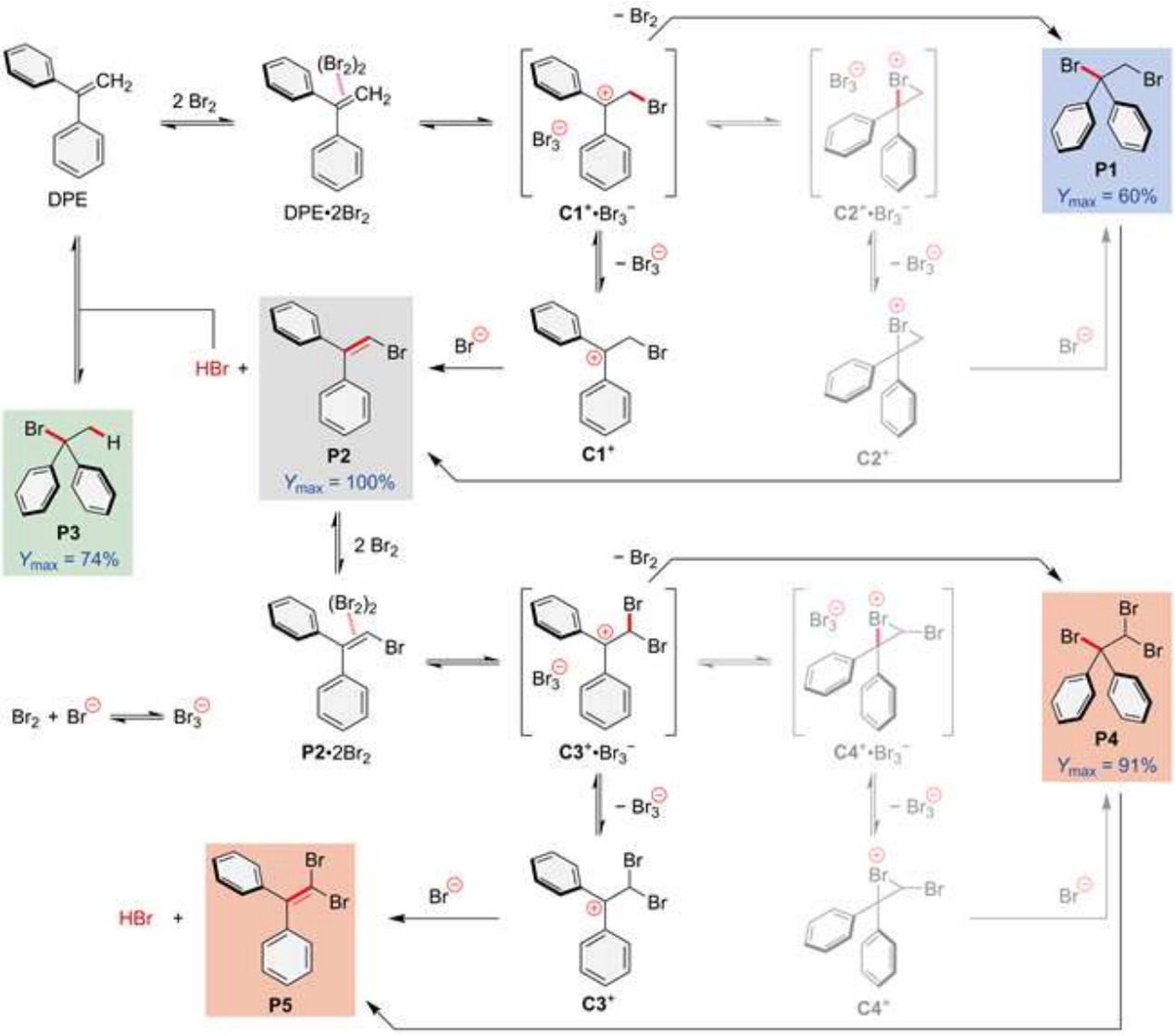


Figure 19. Proposed reaction network for the bromination of DPE.[4] DPE first forms the π-complex DPE·2Br₂, followed by ionization to the β-bromocarbenium–tribromide ion pair C1⁺·Br₃⁻. Collapse of this ion pair gives the vicinal dibromide **P1** (blue; Y_max = 60%), whereas dissociation and deprotonation afford the vinyl bromide **P2** (gray; Y_max = 100%) together with HBr. The released HBr can add to DPE to form the hydrobromination product **P3** (green; Y_max = 74%). Further bromination of P2 proceeds through the corresponding π-complex and β,β-dibromocarbenium–tribromide ion pair C3⁺·Br₃⁻ to give the tribromide **P4** (middle right, orange; Y_max = 91%), while elimination of HBr affords trace **P5** (bottom left, orange). The corresponding cyclic bromonium structures C2⁺ and C4⁺ are shown in gray because DFT optimization converged to the open carbocations C1⁺ and C3⁺. *Reproduced from Publication 4.*[4]

The second system I investigated was the bromination of benzylideneacetone (BDA), which displayed an even greater diversity of outcomes. The four-dimensional hyperspace comprised 48 combinations of BDA, Br_2, and TBABr concentrations, each analyzed after 24 and 48 h. Scaling up selected conditions and separating the crude mixtures by HPLC established >12 products in the BDA reaction network (**Figure 20b**). Previous literature described only the vicinal dibromide **Q1** and the α-brominated product **Q2**, obtained under different solvent conditions—**Q1** in chloroform and **Q2** by precipitation in acetic acid.[91,92] In the present DCE system, **Q1** appeared as anti and syn diastereomers, and both **Q1** and **Q2** were formed under different hyperspace coordinates (**Figure 20a**).[4]

Three compounds, **Q4**, **Q5**, and **Q7**, had not previously been reported. The chlorinated product **Q2′** was known only from separate chlorination methods; its formation under bromination conditions was therefore unexpected.[93,94] Control experiments indicated that TBABr reacted with DCE to release chloride, which displaced bromide in **Q2**[4]. The additional products **Q3**, **Q3′**, **Q6**, **Q7′**, and **Q8** arose through further halogenation, substitution, and elimination pathways. Thus, the two conventionally expected product types expanded into >12 observed species, including three previously unknown compounds and one known compound formed by an unreported pathway (**Figure 20b**).

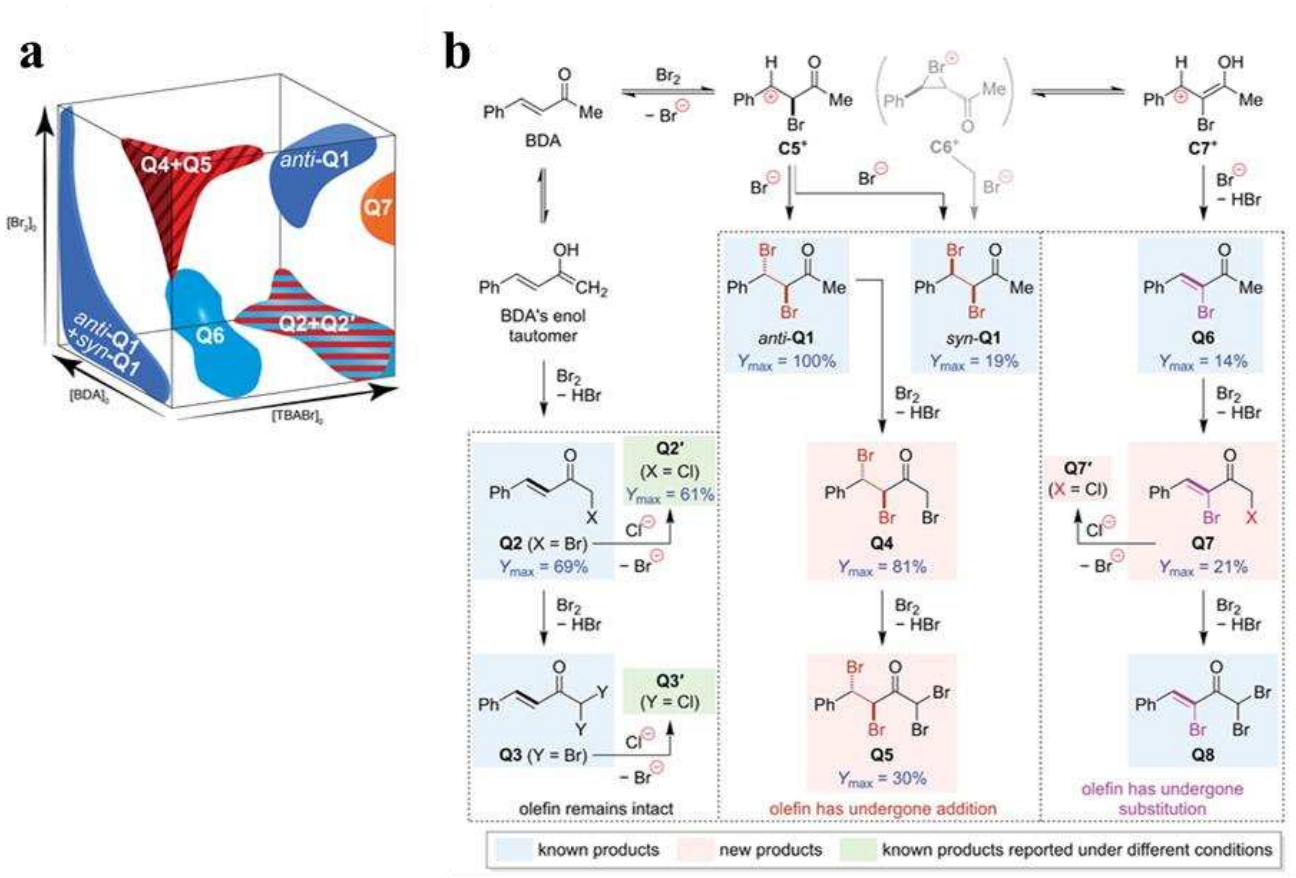


Figure 20. Product-forming regions and proposed reaction network for the bromination of BDA.[4] a) Schematic summary of the regions of the BDA hyperspace in which the principal products dominate as functions of the initial concentrations of BDA, Br_2, and TBABr. **b)** Proposed reaction network connecting the observed products through methyl-group bromination, bromine addition across the alkene, substitution, and further halogenation pathways. Maximum observed yields are indicated for the principal products. *Reproduced from Publication 4.*[4]

Anti-Q1, **Q2**, and the previously unreported **Q4** were major products in different hyperspace regions, reaching maximum yields of 96%, 69%, and 81%, respectively. **Syn-Q1**, **Q2′**, **Q5**, **Q6**, and **Q7** reached 10–50%, whereas **Q3**, **Q3′**, **Q7′**, and **Q8** were detected only in small quantities or could not be quantified reliably because of spectral overlap. These distributions revealed condition-dependent switching among addition, α-bromination, and further-bromination pathways (**Figure 20a**). Together with control experiments and structural characterization, they supported a network comprising three connected classes: methyl-group halogenation with the alkene retained, bromine addition followed by further bromination, and alkene substitution with additional halogenation **(Figure 20b)**. Comparison of the product concentrations after 24 and 48 h provided additional support for some connections in the proposed network. Under selected conditions, the decrease in **syn-Q1** was accompanied by a corresponding increase in **Q6**, consistent with a possible conversion of **syn-Q1** to **Q6**. A similar relationship supported the proposed conversion of **Q2** to **Q2′**.[4]

Together, the DPE and BDA studies demonstrated that automated NMR analysis could support hyperspace-scale investigation of complex reaction mixtures, reveal switches between major products, identify previously unrecognized compounds, and help reconstruct connected reaction pathways.

4. Summary and Conclusions

This dissertation demonstrates that chemical discovery benefits from integrating mechanistic computation, reliable machine-readable data, robotic experimentation, and automated quantitative analysis. Together, these approaches expand the range of reactions that can be proposed, explored, and interpreted beyond the limits of conventional small-scale experimentation.

The *Allchemy-MECH* studies showed that mechanistic reaction networks can guide the discovery of multicomponent and one-pot transformations beyond the retrieval of known reactions. The calculations identified promising products as well as practical requirements related to reagent compatibility, order of addition, competing pathways, and condition-dependent divergence. Experimental validation converted these predictions into workable synthetic procedures and established their scope and preparative utility.

The *AbSynth* contribution addressed a complementary requirement for computational chemistry: the reliability and accessibility of the information on which analysis and prediction depend. By reconstructing complete total syntheses as quality-controlled, machine-readable pathways, *AbSynth* preserves route-level information such as step order, branching, convergence, stereochemistry, and intermediate relationships. Importantly, its release as an open-access resource enables broad use in synthesis analysis, benchmarking, method development, and chemical-AI applications, while emphasizing the continuing importance of expert chemical curation. Analysis of these pathways further showed the importance of multistep reasoning in total-synthesis design and revealed the limited predictive power of several established heuristics for preferred retrosynthetic disconnections when tested across thousands of complete syntheses.

The robotic studies established reaction-hyperspace mapping as a means of determining how reactivity changes across multidimensional condition spaces, rather than treating yield optimization as the sole objective. The fluorenol systems demonstrated that robotic experimentation combined with UV–Vis spectral unmixing could generate internally consistent maps across hundreds of conditions. The large yield datasets also enabled kinetic and equilibrium models to be fitted to the E1 and S_N1 hyperspaces, providing estimates of kinetic and thermodynamic parameters even though each hyperspace was measured at a single reaction time. Extending the workflow to automated benchtop NMR broadened its applicability to reactions whose components could not be distinguished reliably by optical spectroscopy. In the olefin-bromination studies, automated sample transfer and spectrum acquisition, together with characteristic-signal fitting and calibration, enabled consistent quantitative analysis of nearly 1,000 crude reaction mixtures. The resulting product-distribution maps revealed switches between major pathways, solvent participation, previously unrecognized products, and reactivity confined to specific regions of condition space. These results show that both high- and low-

yielding conditions can provide mechanistic and synthetic information when the complete mixture composition is quantified reliably.

The broader implication of this dissertation is that the value of large-scale chemical experimentation lies not simply in generating more data, but in converting experiments into chemically interpretable information. For large robotic datasets, analytical throughput and quantitative reliability are therefore as important as experimental throughput. Automated benchtop NMR is especially valuable in this context because it provides compound-specific structural information directly from crude mixtures and, when coupled with reproducible signal fitting and calibration, makes hyperspace-wide quantification practical without manual analysis of every sample. By integrating mechanistic reasoning, curated chemical knowledge, robotic experimentation, and scalable quantitative analysis, this dissertation provides a framework for constructing more complete descriptions of reactivity that can support mechanistic understanding, reaction prediction, synthesis planning, and future autonomous experimentation.

5. References

(1) 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.; Grzybowski, B. A. Systematic, Computational Discovery of Multicomponent and One-Pot Reactions. *Nat. Commun.* **2024**, *15* (1), 10285. https://doi.org/10.1038/s41467-024-54611-5

(2) Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W. D.; Beker, W.; Saillard, E.; Łęgowski, K.; Roszak, R.; Kowalczyk, P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Centuryof Total Syntheses as a Machine-ReadableResource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026**. https://doi.org/10.1021/jacs.6c08434

(3) Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1

(4) Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**, e9132891. https://doi.org/10.1002/anie.9132891

(5) Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. Organic Reactivity from Mechanism to Machine Learning. *Nat. Rev. Chem.* **2021**, *5* (4), 240–255. https://doi.org/10.1038/s41570-021-00260-x

(6) Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska, O.; Beker, W.; Badowski, T.; Scheidt, K. A.; Molga, K.; Mlynarski, J.; Mrksich, M.; Grzybowski, B. A. Computational Planning of the Synthesis of Complex Natural Products. *Nature* **2020**, *588* (7836), 83–88. https://doi.org/10.1038/s41586-020-2855-y

(7) Collins, K. D.; Gensch, T.; Glorius, F. Contemporary Screening Approaches to Reaction Discovery and Development. *Nat. Chem.* **2014**, *6* (10), 859–871.

(8) Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143* (45), 18820–18826. https://doi.org/10.1021/jacs.1c09820

(9) Nsouli, R.; Galiyan, G.; Ackerman-Biegasiewicz, L. K. G. Advancing Organic Chemistry Using High-Throughput Experimentation. *Angew. Chem. Int. Ed.* **2025**, *64* (40), e202506588. https://doi.org/10.1002/anie.202506588

(10) King-Smith, E.; Berritt, S.; Bernier, L.; Hou, X.; Klug-McLeod, J. L.; Mustakis, J.; Sach, N. W.; Tucker, J. W.; Yang, Q.; Howard, R. M. Probing the Chemical 'Reactome'with High-Throughput Experimentation Data. *Nat. Chem.* **2024**, *16* (4), 633–643. https://doi.org/10.1038/s41557-023-01393-w

(11) Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Frydrych, R.; Grzybowski, B. A. Automated NMRanger Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures. *J. Am. Chem. Soc. Submitt.* **2026**.

(12) Strieth-Kalthoff, F.; Szymkuć, S.; Molga, K.; Aspuru-Guzik, A.; Glorius, F.; Grzybowski, B. A. Artificial Intelligence for Retrosynthetic Planning Needs Both Data and Expert Knowledge. *J. Am. Chem. Soc.* **2024**, jacs.4c00338. https://doi.org/10.1021/jacs.4c00338

(13) Segler, M. H.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555* (7698), 604–610. https://doi.org/10.1038/nature25978

(14) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51* (5), 1281–1289. https://doi.org/10.1021/acs.accounts.8b00087

(15) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **2016**, *55* (20), 5904–5937. https://doi.org/10.1002/anie.201506101

(16) Prajapati, B.; Wong, W.; Frydrych, R.; Kolygina, D. V.; Jia, Y.; Marazzi, B.; Vagkidis, N.; Schreiner, P. R.; Grzybowski, B. A. Robot-Assisted Reconstruction and Control of the Pechmann Reaction Network. *Angew. Chem.* **2026**, *138* (6), e18394. https://doi.org/10.1002/ange.202518394

(17) Jiang, Y.; Gadina, L.; Ahumada, J. C.; Jia, Y.; Wong, W.-S.; Kim, J.; Grzybowski, B. A. Hyperspace Mapping and Iterative Dimensionality Reduction Reveal Radical Cascade Networks Controlled by Perovskite-Type Catalysts. *Chem* **2026**.

(18) Matuszczyk, D.; Frydrych, R.; Wong, W.-S.; Jia, Y.; Sobolev, Y. I.; Hanus, J.; Lee, J.; Grzybowski, B. A. Hyperspace Exploration Using Robotics for the Discovery of Mechanistically Distinct Transformations and Complex Functional Products. *Nat. Synth.* **2026**, 1–10.

(19) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123* (6), 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798

(20) Tom, G.; Schmid, S. P.; Baird, S. G.; Cao, Y.; Darvish, K.; Hao, H.; Lo, S.; Pablo-García, S.; Rajaonson, E. M.; Skreta, M.; Yoshikawa, N.; Corapi, S.; Akkoc, G. D.; Strieth-Kalthoff, F.; Seifrid, M.; Aspuru-Guzik, A. Self-Driving Laboratories for Chemistry and Materials Science. *Chem. Rev.* **2024**, *124* (16), 9633–9732. https://doi.org/10.1021/acs.chemrev.4c00055

(21) Angello, N. H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E. R.; Roszak, R.; Wu, T. C.; Schroeder, C. M.; Aspuru-Guzik, A.; Grzybowski, B. A.; Burke, M. D. Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling. *Science* **2022**, *378* (6618), 399–405. https://doi.org/10.1126/science.adc8743

(22) Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to Explore Chemical Space Using Algorithms and Automation. *Nat. Rev. Chem.* **2019**, *3* (2), 119–128.

(23) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559* (7714), 377–381.

(24) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365* (6453), eaax1566. https://doi.org/10.1126/science.aax1566

(25) Angello, N. H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E. R.; Roszak, R.; Wu, T. C.; Schroeder, C. M.; Aspuru-Guzik, A.; Grzybowski, B. A.; Burke, M. D. Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling. *Science* **2022**, *378* (6618), 399–405. https://doi.org/10.1126/science.adc8743

(26) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559* (7714), 377–381.

(27) Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science* **2019**, *363* (6423), eaav2211. https://doi.org/10.1126/science.aav2211

(28) Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; Toutchkine, A.; Dittwald, P.; Startek, M. P.; Kirkovits, G. J.; Roszak, R.; Adamski, A.; Sieredzińska, B.; Mrksich, M.; Trice, S. L. J.; Grzybowski, B. A. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **2018**, *4* (3), 522–532. https://doi.org/10.1016/j.chempr.2018.02.002

(29) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555* (7698), 604–610. https://doi.org/10.1038/nature25978

(30) Corey, E. J. General Methods for the Construction of Complex Molecules. *Pure Appl. Chem.* **1967**, *14* (1), 19–38. https://doi.org/10.1351/pac196714010019

(31) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses: Pathways for Molecular Synthesis Can Be Devised with a Computer and Equipment for Graphical Communication. *Science* **1969**, *166* (3902), 178–192. https://doi.org/10.1126/science.166.3902.178

(32) Fialkowski, M.; Bishop, K. J. M.; Chubukov, V. A.; Campbell, C. J.; Grzybowski, B. A. Architecture and Evolution of Organic Chemistry. *Angew. Chem. Int. Ed.* **2005**, *44* (44), 7263–7269. https://doi.org/10.1002/anie.200502272

(33) Bishop, K. J. M.; Klajn, R.; Grzybowski, B. A. The Core and Most Useful Molecules in Organic Chemistry. *Angew. Chem.* **2006**, *118* (32), 5474–5480. https://doi.org/10.1002/ange.200600881

(34) Gothard, C. M.; Soh, S.; Gothard, N. A.; Kowalczyk, B.; Wei, Y.; Baytekin, B.; Grzybowski, B. A. Rewiring Chemistry: Algorithmic Discovery and Experimental Validation of One-Pot Reactions in the Network of Organic Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51* (32), 7922–7927. https://doi.org/10.1002/anie.201202155

(35) Kowalik, M.; Gothard, C. M.; Drews, A. M.; Gothard, N. A.; Weckiewicz, A.; Fuller, P. E.; Grzybowski, B. A.; Bishop, K. J. M. Parallel Optimization of Synthetic Pathways within the Network of Organic Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51* (32), 7928–7932. https://doi.org/10.1002/anie.201202209

(36) Jacob, P.-M.; Yamin, P.; Perez-Storey, C.; Hopgood, M.; Lapkin, A. A. Towards Automation of Chemical Process Route Selection Based on Data Mining. *Green Chem.* **2017**, *19* (1), 140–152. https://doi.org/10.1039/C6GC02482C

(37) Lin, Y.; Zhang, R.; Wang, D.; Cernak, T. Computer-Aided Key Step Generation in Alkaloid Total Synthesis. *Science* **2023**, *379* (6631), 453–457. https://doi.org/10.1126/science.ade8459

(38) Wołos, A.; Roszak, R.; Żądło-Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; Spólnik, G.; Dygas, M.; Szymkuć, S.; Grzybowski, B. A. Synthetic Connectivity, Emergence, and Self-Regeneration in the Network of Prebiotic Chemistry. *Science* **2020**, *369* (6511), eaaw1955. https://doi.org/10.1126/science.aaw1955

(39) Molga, K.; Szymkuć, S.; Gołębiowska, P.; Popik, O.; Dittwald, P.; Moskal, M.; Roszak, R.; Mlynarski, J.; Grzybowski, B. A. A Computer Algorithm to Discover Iterative Sequences of Organic Reactions. *Nat. Synth.* **2022**, *1* (1), 49–58. https://doi.org/10.1038/s44160-021-00010-3

(40) Klucznik, T.; Syntrivanis, L.-D.; Baś, S.; Mikulak-Klucznik, B.; Moskal, M.; Szymkuć, S.; Mlynarski, J.; Gadina, L.; Beker, W.; Burke, M. D.; Tiefenbacher, K.; Grzybowski, B. A. Computational Prediction of Complex Cationic Rearrangement Outcomes. *Nature* **2023**. https://doi.org/10.1038/s41586-023-06854-3

(41) Gadina, L. P. J. -L.; Baś, S.; Makkawi, A.; Bilgi-Gadina, Y.; Klucznik, T.; Beker, W.; Grzybowski, B. A. Studying the Outcomes and Mechanisms of Carbocationic Rearrangements Using Algorithm-Augmented Experimentation. *ChemistryEurope* **2026**, *4* (2), e202500189. https://doi.org/10.1002/ceur.202500189

(42) Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M. E.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. Synthesis of Many Different Types of Organic Small Molecules Using One Automated Process. *Science* **2015**, *347* (6227), 1221–1226. https://doi.org/10.1126/science.aaa5414

(43) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2

(44) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y

(45) Dömling, A.; Wang, W.; Wang, K. Chemistry and Biology Of Multicomponent Reactions. *Chem. Rev.* **2012**, *112* (6), 3083–3135. https://doi.org/10.1021/cr100233r

(46) Dömling, A.; Ugi, I. Multicomponent Reactions with Isocyanides. *Angew. Chem.* **2000**, *39* (18), 3168–3210. https://doi.org/10.1002/1521-3773(20000915)39:18%3C3168::AID-ANIE3168%3E3.0.CO;2-U

(47) Ganem, B. Strategies for Innovation in Multicomponent Reaction Design. *Acc. Chem. Res.* **2009**, *42* (3), 463–472. https://doi.org/10.1021/ar800214s

(48) Oliver Kappe, C. 100 Years of the Biginelli Dihydropyrimidine Synthesis. *Tetrahedron* **1993**, *49* (32), 6937–6963. https://doi.org/10.1016/S0040-4020(01)87971-0

(49) Strecker, A. Ueber Die Künstliche Bildung Der Milchsäure Und Einen Neuen, Dem Glycocoll Homologen Körper; *Justus Liebigs Ann. Chem.* **1850**, *75* (1), 27–45. https://doi.org/10.1002/jlac.18500750103

(50) Zhao, W.; Chen, F.-E. One-Pot Synthesis and Its Practical Application in Pharmaceutical Industry. *Curr. Org. Synth.* **2012**, *9* (6), 873–897. https://doi.org/10.2174/157017912803901619

(51) Broadwater, S. J.; Roth, S. L.; Price, K. E.; Kobašlija, M.; McQuade, D. T. One-Pot Multi-Step Synthesis: A Challenge Spawning Innovation. *Org. Biomol. Chem.* **2005**, *3* (16), 2899–2906. https://doi.org/10.1039/b506621m

(52) Hayashi, Y. Time and Pot Economy in Total Synthesis. *Acc. Chem. Res.* **2021**, *54* (6), 1385–1398. https://doi.org/10.1021/acs.accounts.0c00803

(53) Cioc, R. C.; Ruijter, E.; Orru, R. V. Multicomponent Reactions: Advanced Tools for Sustainable Organic Synthesis. *Green Chem.* **2014**, *16* (6), 2958–2975. https://doi.org/10.1039/c4gc00013g

(54) D'Souza, D. M.; Mueller, T. J. Multi-Component Syntheses of Heterocycles by Transition-Metal Catalysis. *Chem. Soc. Rev.* **2007**, *36* (7), 1095–1108. https://doi.org/10.1039/b608235c

(55) Medley, J. W.; Movassaghi, M. Robinson's Landmark Synthesis of Tropinone. *Chem. Commun.* **2013**, *49* (92), 10775–10777. https://doi.org/10.1039/c3cc44461a

(56) Ruijter, E.; Scheffelaar, R.; Orru, R. V. A. Multicomponent Reaction Design in the Quest for Molecular Complexity and Diversity. *Angew. Chem. Int. Ed.* **2011**, *50* (28), 6234–6246. https://doi.org/10.1002/anie.201006515

(57) 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.* **2015**, *55* (1), 39–53. https://doi.org/10.1021/ci5006614

(58) Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. New Applications of Computers in Chemistry. *Angew. Chem. Int. Ed. Engl.* **1979**, *18* (2), 111–123. https://doi.org/10.1002/anie.197901111

(59) Bauer, J.; Ugi, I. Chemical Reactions and Structures without Precedent Generated by Computer Program. *J. Chem. Res. Synop.* **1982**.

(60) Nicolaou, K. C. The Emergence of the Structure of the Molecule and the Art of Its Synthesis. *Angew. Chem. Int. Ed.* **2013**, *52* (1), 131–146. https://doi.org/10.1002/anie.201207081

(61) 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.* **2019**, *4* (9), 1506–1521. https://doi.org/10.1039/C9RE00076C

(62) Carstensen, G. Chemistry by Design. https://chemistrybydesign.oia.arizona.edu/ (accessed 2026-03-16)

(63) Draghici, C.; Njardarson, J. T. Chemistry By Design: A Web-Based Educational Flashcard for Exploring Synthetic Organic Chemistry. *J. Chem. Educ.* **2012**, *89* (8), 1080–1082. https://doi.org/10.1021/ed2006423

(64) SynArchive—The Organic Synthesis Archive. https://synarchive.com/ (accessed 2026-03-28)

(65) Hans Reich's Collection. Total Syntheses. https://organicchemistrydata.org/hansreich/resources/syntheses (accessed 2026-07-05)

(66) Tharra, P. R.; Mikhaylov, A. A.; Švejkar, J.; Gysin, M.; Hobbie, S. N.; Švenda, J. Short Synthesis of (+)-Actinobolin: Simple Entry to Complex Small-Molecule Inhibitors of Protein Synthesis. *Angew. Chem. Int. Ed.* **2022**, *61* (17), e202116520. https://doi.org/10.1002/anie.202116520

(67) AbSynth. https://absynth.allchemy.net/target_list

(68) Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis*; Wiley Interscience: New York, 1989.

(69) Davies, I. W. The Digitization of Organic Synthesis. *Nature* **2019**, *570* (7760), 175–181. https://doi.org/10.1038/s41586-019-1288-y

(70) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559* (7714), 377–381. https://doi.org/10.1038/s41586-018-0307-8

(71) Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning. *Science* **2019**, *365* (6453), eaax1566. https://doi.org/10.1126/science.aax1566

(72) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y

(73) Wang, J. Y.; Stevens, J. M.; Kariofillis, S. K.; Tom, M.-J.; Golden, D. L.; Li, J.; Tabora, J. E.; Parasram, M.; Shields, B. J.; Primer, D. N.; Hao, B.; Del Valle, D.; DiSomma, S.; Furman, A.; Zipp, G. G.; Melnikov, S.; Paulson, J.; Doyle, A. G. Identifying General Reaction Conditions by Bandit Optimization. *Nature* **2024**, *626* (8001), 1025–1033. https://doi.org/10.1038/s41586-024-07021-y

(74) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends Chem.* **2019**, *1* (3), 282–291. https://doi.org/10.1016/j.trechm.2019.02.007

(75) Reizman, B. J.; Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **2016**, *49* (9), 1786–1796. https://doi.org/10.1021/acs.accounts.6b00261

(76) Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett.* **2017**, *8* (6), 601–607. https://doi.org/10.1021/acsmedchemlett.7b00165

(77) Troshin, K.; Hartwig, J. F. Snap Deconvolution: An Informatics Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **2017**, *357* (6347), 175–181. https://doi.org/10.1126/science.aan1568

(78) Bayly, A. A.; McDonald, B. R.; Mrksich, M.; Scheidt, K. A. High-Throughput Photocapture Approach for Reaction Discovery. *Proc. Natl. Acad. Sci.* **2020**, *117* (24), 13261–13266. https://doi.org/10.1073/pnas.2003347117

(79) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of^{1} H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442. https://doi.org/10.1021/ac060209g

(80) Rout, M.; Lipfert, M.; Lee, B. L.; Berjanskii, M.; Assempour, N.; Fresno, R. V.; Cayuela, A. S.; Dong, Y.; Johnson, M.; Shahin, H.; Gautam, V.; Sajed, T.; Oler, E.; Peters, H.; Mandal, R.; Wishart, D. S. MagMet: A Fully Automated Web Server for Targeted Nuclear Magnetic Resonance Metabolomics of Plasma and Serum. *Magn. Reson. Chem.* **2023**, *61* (12), 681–704. https://doi.org/10.1002/mrc.5371

(81) Hao, J.; Astle, W.; De Iorio, M.; Ebbels, T. M. D. BATMAN—an R Package for the Automated Quantification of Metabolites from Nuclear Magnetic Resonance Spectra Using a Bayesian Model. *Bioinformatics* **2012**, *28* (15), 2088–2090. https://doi.org/10.1093/bioinformatics/bts308

(82) Astle, W.; De Iorio, M.; Richardson, S.; Stephens, D.; Ebbels, T. A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures. *J. Am. Stat. Assoc.* **2012**, *107* (500), 1259–1271. https://doi.org/10.1080/01621459.2012.695661

(83) Röhnisch, H. E.; Eriksson, J.; Müllner, E.; Agback, P.; Sandström, C.; Moazzami, A. A. AQuA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma. *Anal. Chem.* **2018**, *90* (3), 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324

(84) Domżał, B.; Nawrocka, E. K.; Gołowicz, D.; Ciach, M. A.; Miasojedow, B.; Kazimierczuk, K.; Gambin, A. Magnetstein: An Open-Source Tool for Quantitative NMR Mixture Analysis Robust to Low Resolution, Distorted Lineshapes, and Peak Shifts. *Anal. Chem.* **2024**, *96* (1), 188–196. https://doi.org/10.1021/acs.analchem.3c03594

(85) Zhu, J.; Wang, Q.; Wang, M. *Multicomponent Reactions in Organic Synthesis*; John Wiley & Sons, 2015.

(86) Roszak, R.; Beker, W.; Molga, K.; Grzybowski, B. A. Rapid and Accurate Prediction of p K_a Values of C–H Acids Using Graph Convolutional Neural Networks. *J. Am. Chem. Soc.* **2019**, *141* (43), 17142–17149. https://doi.org/10.1021/jacs.9b05895

(87) Ireland, R. E.; Armstrong, J. D.; Lebreton, 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.* **1993**, *115* (16), 7152–7165. https://doi.org/10.1021/ja00069a013

(88) 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.* **1987**, *109* (7), 2082–2089. https://doi.org/10.1021/ja00241a028

(89) Elsevier. Reaxys Predictive Retrosynthesis. https://www.elsevier.com/products/reaxys/predictive-retrosynthesis (accessed 2026-05-13)

(90) Kamal, R.; Kumar, R.; Kumar, V.; Bhardwaj, V. Synthetic Utilization of α,β-Chalcone Dibromide In Heterocyclic Chemistry and Stereoselective Debromination. *ChemistrySelect* **2019**, *4* (39), 11578–11603. https://doi.org/10.1002/slct.201902262

(91) Claisen, L.; Claparede, A. Condensationen von Ketonen Mit Aldehyden. *Berichte Dtsch. Chem. Ges.* **1881**, *14* (2), 2460–2468.

(92) Ghodsi, S.; Fouladi, S.; Safari, H. A.; Taghi-Ganji, M. Synthesis and Characterization of New Fluoroquinolones Containing an N Substituted Piperazine Moiety. *Asian J. Chem.* **2009**, *21* (4), 3037.

(93) Fraser, R. R.; Kong, F. The Chlorination of Ketones Using Trimeihylchlorosilane and Dimetwylsulfoxide With Bromide Ion Catalysis. *Synth. Commun.* **1988**, *18* (10), 1071–1077. https://doi.org/10.1080/00397918808060892

(94) Zhou, J.; Li, J.; Li, Y.; Wu, C.; He, G.; Yang, Q.; Zhou, Y.; Liu, H. Direct Synthesis of 3-Acylindoles through Rhodium(III)-Catalyzed Annulation of *N* -Phenylamidines with α-Cl Ketones. *Org. Lett.* **2018**, *20* (23), 7645–7649. https://doi.org/10.1021/acs.orglett.8b03383

6. Original Publications

nature communications

Article

https://doi.org/10.1038/s41467-024-54611-5

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

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[1–4] dating back to 1970s. In this context, reactions that build complex scaffolds from multiple simple components in one step (i.e., multicomponent reactions, MCRs[5–11]; Fig. 1a) and/or proceed sequentially in one pot[12–14] are of particular interest as they minimize separation and purification operations, and increase the overall step- and atom-economy[15] as well as "greenness"[16,17] 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, β-dicarbonyl, or imine-based MCRs) and their straightforward combinations[18] and extensions[19–21] 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[22–28] 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

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

https://doi.org/10.1038/s41467-024-54611-5

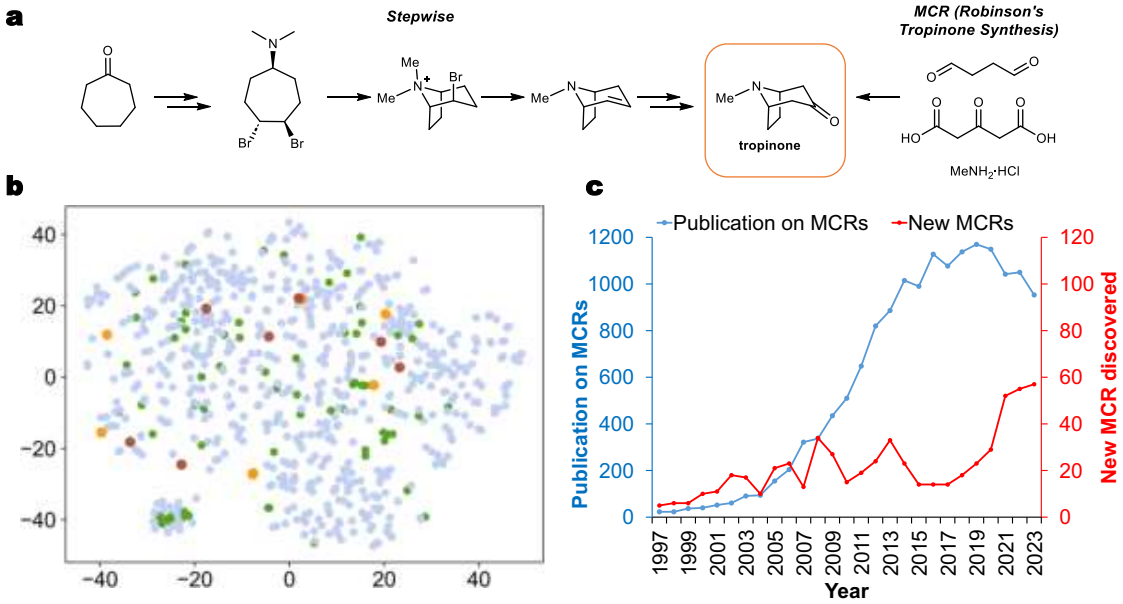


Fig. 1 | Significance and current discovery rate of multicomponent reactions. **a** A classic example illustrating the elegance and efficiency of Robinson's one-step, MCR synthesis of tropinone vs. prior, fifteen-step synthesis[10]. In the latter, only two key steps are shown. **b** t-SNE projection "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[83] (i.e., difference across a full reaction, not its mechanistic steps). The interactive t-SNE map is deposited at https://mcrmap.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[29], which has been used in computational chemistry for decades[30–35] (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[36], 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 unpreported reactions[37,38]. 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[39,40]. 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

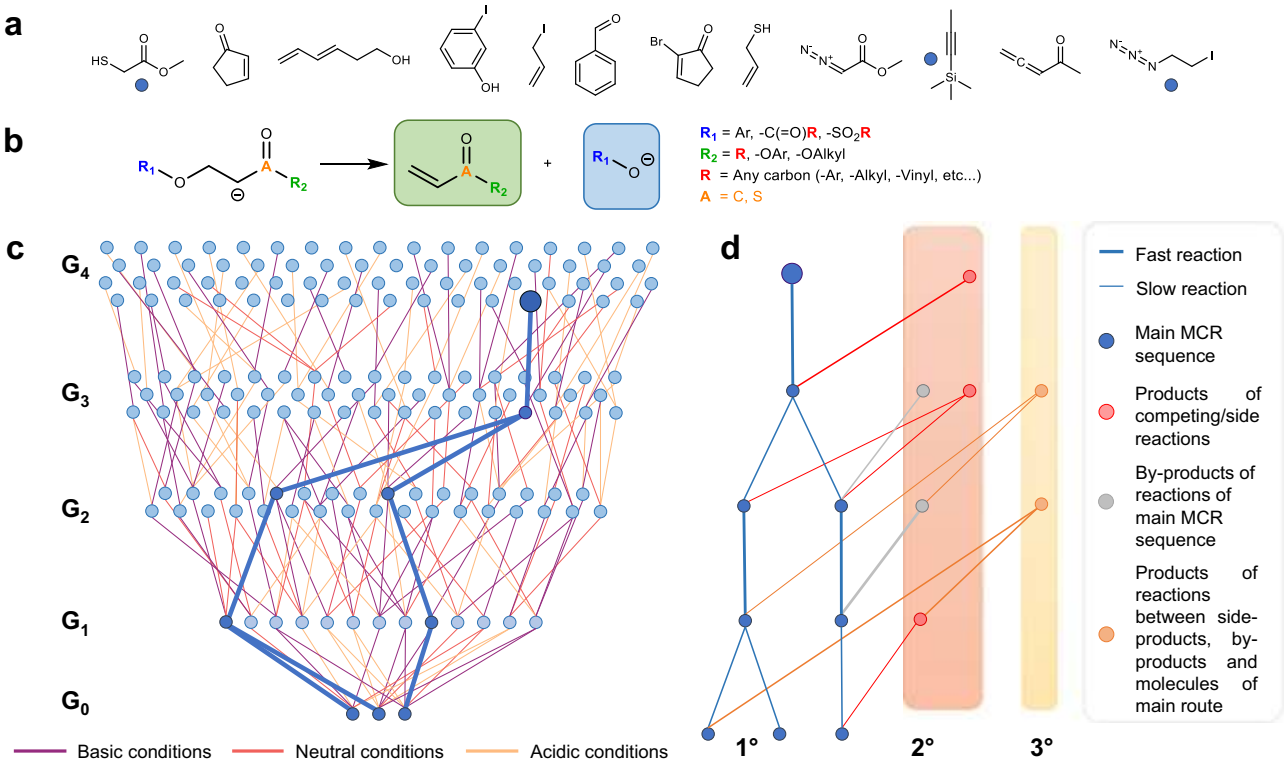


Fig. 2 | Key elements of the MECH algorithm to discover MCRs. a Examples of simple starting materials from the collection of ca. 2400 (see main text and Methods). **b** Abbreviated example of one of ~8000 mechanistic transforms, here E1cB elimination. Different positions can accommodate various substituents, 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 expended 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 that the main-path ones.

first-generation, G_1, of products and by-products, which are then iteratively reacted[23,25,36] to give generations G_2, G_3 (up to some user-specified generation n), resulting in rapidly expanding[26] 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[41] 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:

(i) Cannot combine steps requiring oxidative and reductive conditions, and cannot use water-sensitive steps after water-requiring ones;

(ii) Should use solvents of the same class, although protic solvents are allowed to be added to aprotic ones (but not vice versa);

(iii) Cannot change multiple times between non-overlapping high and low temperature ranges (which would be experimentally impractical);

(iv) 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

https://doi.org/10.1038/s41467-024-54611-5

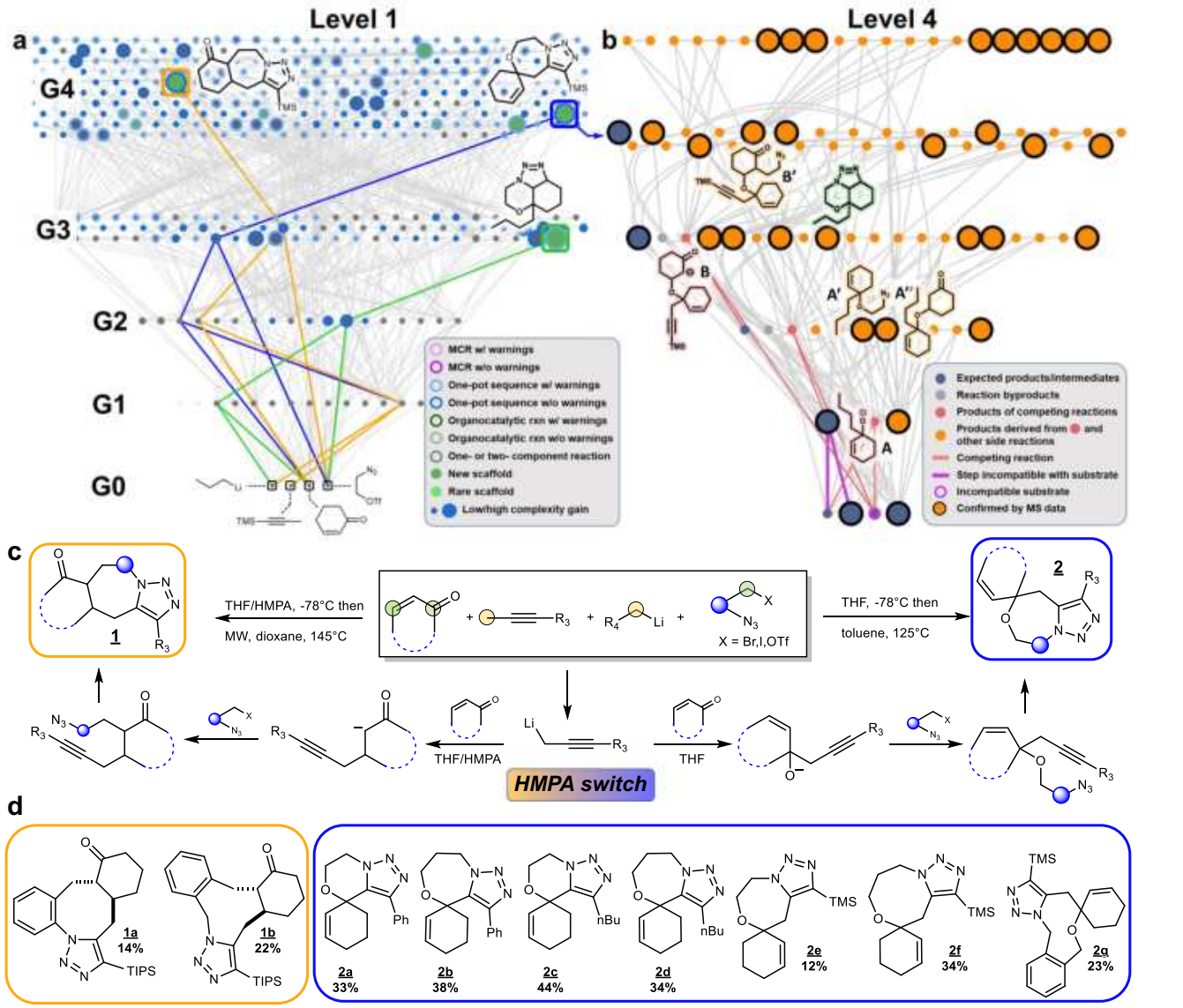


Fig. 3 | Example of algorithmically-discovered one-pot sequences and the corresponding mechanistic network expanded to Level 4. a Screenshot of Level 1 network propagated from cyclohexenone, trimethylsilylpropyne, *n*-butyllithium and azidotriflate substrates to $n = 4$ generations, G_4. The network 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 nBuLi/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.allchemy.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[22,24]) 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, azidohalide and alkyne, the mechanistic network propagated to G_4 (Fig. 3a) contains conditions-matched sequences leading to 391 products with MW < 500. Two compounds in G_4 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[44–47]. 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[48]; 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.

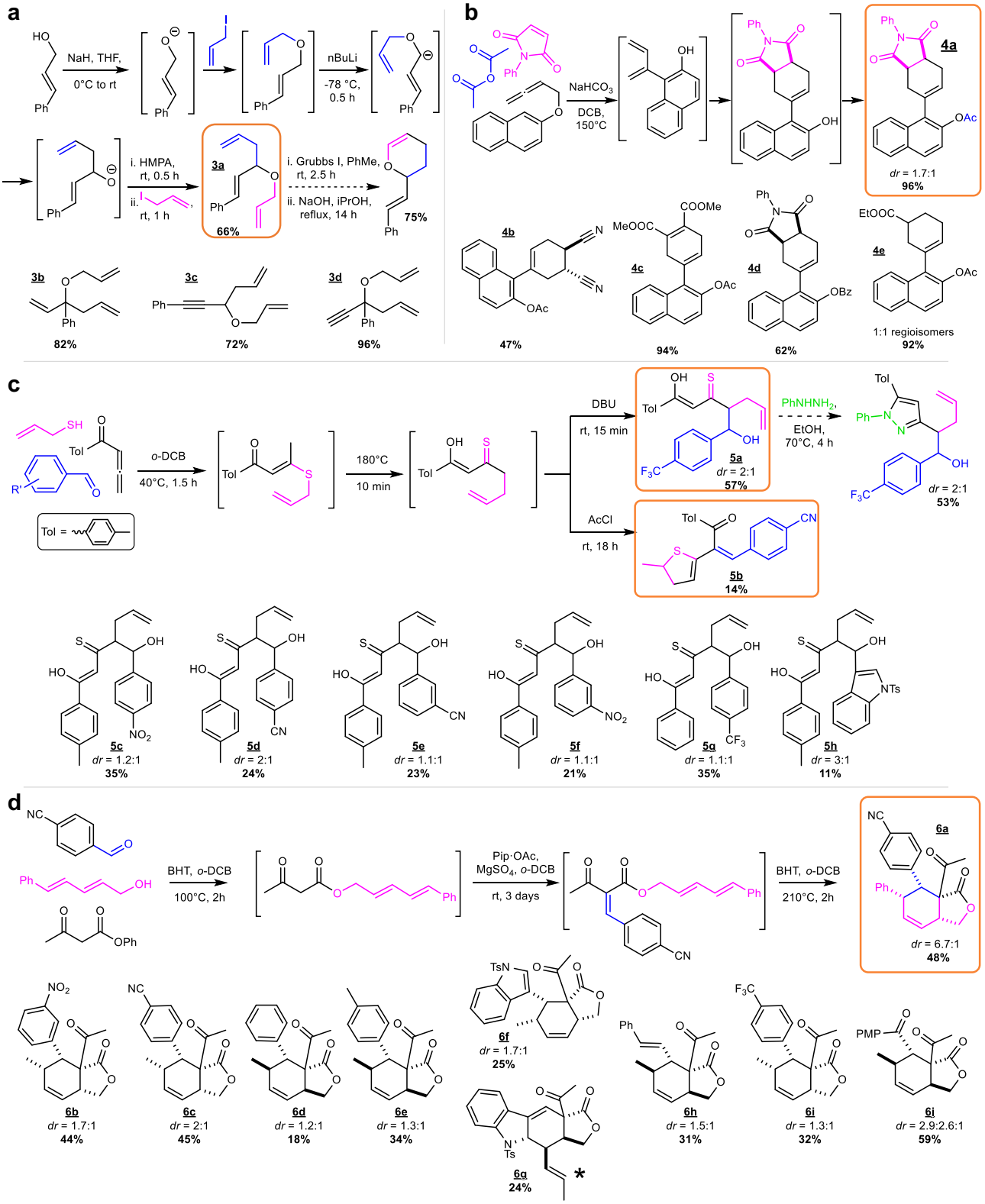
Another prediction for a one-pot, Mach3 sequence relying on a 2,3-Wittig rearrangement and leading to branched diallylic ethers **3a**–**3d**, is illustrated in Fig. 4a and Supplementary Fig. S159. This sequence was committed to experiment because, after metathesis (not compatible with one-pot conditions 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[49–52]. 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[53,54]. Scaffolds of this type are typically prepared via various multistep protocols[55–59]. 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 thioketone 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[60] to give substituted pyrazoles which are popular motifs of many drugs. By contrast, in Mach6, addition of acetyl chloride triggers a relatively rare[61] sequence of acetylation of alcohol, acidic elimination of acetic acid catalyzed by the in-situ generated HCl to give the Knoevenagel-type adduct, thioketo-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

https://doi.org/10.1038/s41467-024-54611-5



by partial decomposition of the product during purification), and their applications are less conspicuous[62].

The sequence underlying Mach7 MCR shown in Fig. 4d – leading to a scaffold akin to oblongolide natural products considered as potential algicide, herbicide[63] and antiviral[64] 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 β,γ-unsaturated alcohols. **b** Scheme of a MCR producing unsaturated β-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-β-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-β-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}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 β-lactone which, upon addition of $MgBr_2$, 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[67] 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, iodophenol 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 (α-bromo-α,β-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[68]. However, the program also indicated that that the C-H pKa of the α-azidoester be higher than that of the α-thioester – that is, the deprotonation (either by azide anion[69] 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 α-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 β-aryl ketones, 35 mol% thiol load is necessary due to the trapping of the thiol catalyst in the S_N2 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[70]) 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, neural, acidic), solvent class (protic/aprotic and polar/non-polar), temperature range (very low = <−20 °C, low = −20 to 20 °C, r.t., high = 40 to 150 °C, and very high = >150 °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,

https://doi.org/10.1038/s41467-024-54611-5

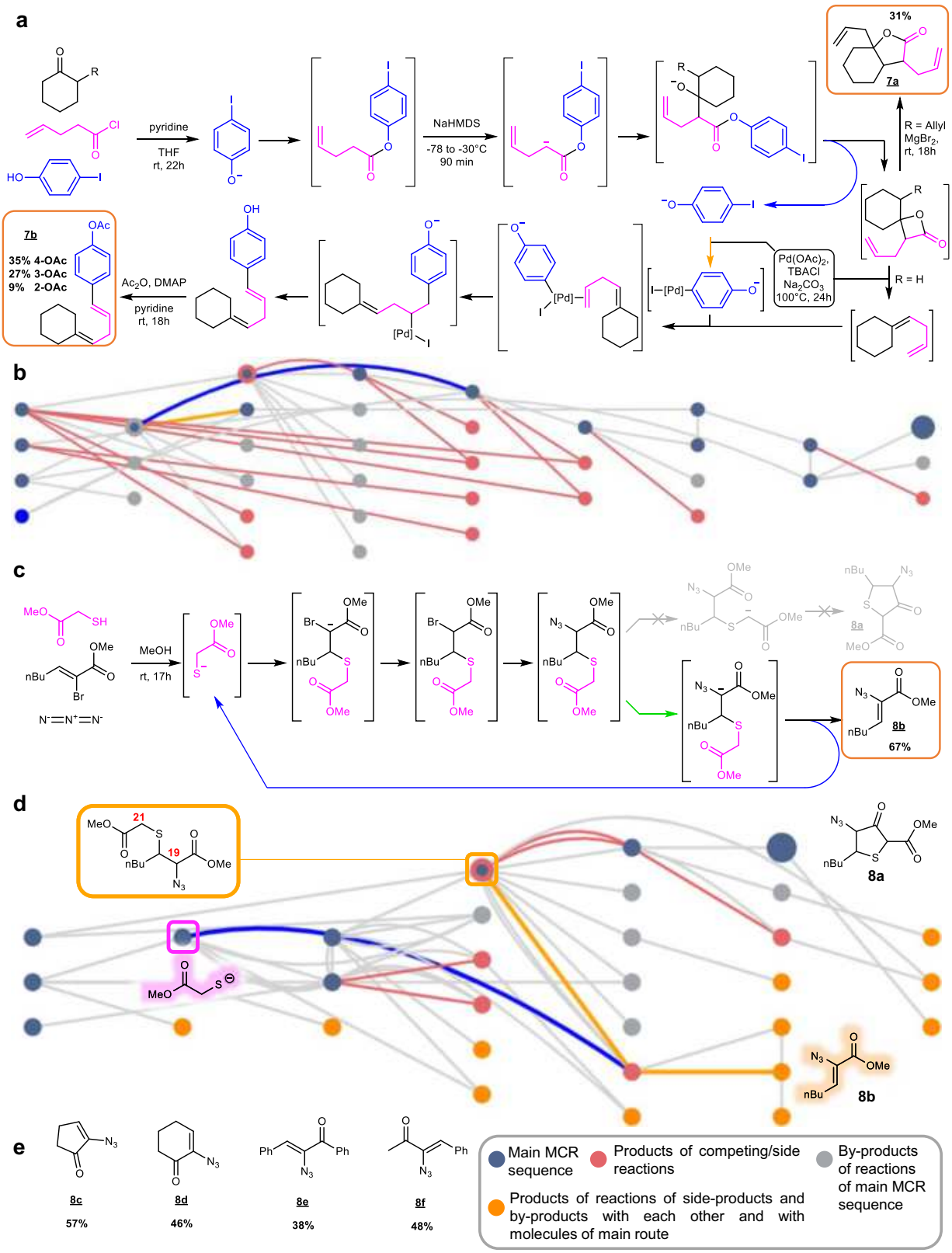


Fig. 5 | Computer-discovered substrate-reusing MCRs and an organocatalytic reaction. a Scheme of a MCR for the synthesis of arylated skipped dienes. Non-isolated intermediates are shown in brackets and the isolated products are framed in orange. The obtained dienes were separately acetylated for the purpose of purification. The bicyclic lactone **7a** (upper right) was obtained from substituted cyclohexanone (R = allyl) and phenol substrates when $MgBr{\cdot}Et_2O$ 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^2-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 α-bromoenones. 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., diethylaluminium 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[72]; – 7.5 h[73]; – 3 h[74]; – 5 min). Another example is S_N2 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[75], while reaction of a similar molecule, methyl 2-(1-bromo-2-methoxy-2-oxoethyl) benzoate, with potassium cyanide in DMF takes only 1 h[76]).

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[36]). 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[36]). Some rare types of steps involving π-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 ΔC is calculated along an atom-mapped path as (a·#Rearrangements + a·#RingsFormed + #BondsCreated + #BondsDisconnected), where $a = 5$ 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[77]. 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[78], the rate constants of mechanistic steps are approximated by using Mayr's nucleophilicity, N, and electrophilicity, E, indices[42,43] as $\log k_{20deg} \propto (N + E)$, which are further fine-tuned by corrections capturing relative reactivities, stoichiometries and amounts of various species in the mechanistic networks, $\ln k_i = \ln k_i^{Mayr} + \sum \text{corrections}(r_i)$. 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 (ρ^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[79–82].

Pre-curated collection of substrates available through Allchemy'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[77] 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 uniqe 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. User manuals are available in Supplementary Section S1. 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 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 α-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 π-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., Kobašlija, 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-arylglycoside 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. Chongquing, 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., Sarkar, 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., Srichomthong, 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 γ-butyrolactones. *Int. J. Mol. Sci.* **22**, 2769 (2021).
68. Le, H. V. et al. Design and mechanism of tetrahydrothiophene-based γ-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 hydroxylnaphthalene-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 Pigm* **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-arylisoquinolines from δ-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. Skoraczyński, 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 saraszymkuc@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

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/JACS
Article

AbSynth: A Century of Total Syntheses as a Machine-Readable Resource for Analyzing Strategic Logic in Organic Chemistry

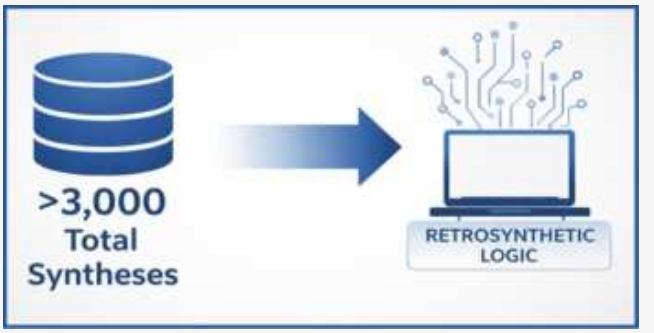
Maciej Czuba, Konrad Barnowski, Karol Molga, Zofia Pławecka, Agnieszka Wołos, Szymon Stodółkiewicz, Sebastian Baś, Wojciech Petrykowski, Wiktor Beker, Evan Saillard, Konrad Łęgowski, Rafał Roszak, Paweł Kowalczyk, Barbara Mikulak-Klucznik, Piotr Janasz, Emilia Mielke, Patrycja Pietruszewska, Maja Łazuchiewicz, Sylwia Wójcik, Jakub Zabielski, Dariusz Nostkiewicz, Gabriela Smaga, Ahmad Makkawi, Louis P. J.-L. Gadina, Jakub Skoczek, Tomasz Klucznik, Maxence Deschamps, Anna Żądło-Dobrowolska, Michał Michalak, Sara Szymkuć,* Daniel T. Gryko,* Yasemin Bilgi-Gadina,* and Bartosz A. Grzybowski*

Cite This: J. Am. Chem. Soc. 2026, 148, 38740−38753

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The design of synthetic routes to complex, stereochemically rich targets remains one of the most demanding challenges in organic chemistry. To interrogate the strategic logic underlying such efforts, we curated and digitized over 3,000 classical total syntheses comprising close to 60,000 individual steps, now freely available as the AbSynth collection. Analyses across this data set reveal that synthesis planning cannot be reduced to purely local, one-step-at-a-time decision-making: known complexity metrics and scoring functions fail to vary monotonically along synthetic trajectories, and conventional bond-disconnection heuristics apply only sporadically and lack the specificity required for automation. Instead, our study shows that nearly half of all steps correspond to structurally unproductive − yet strategically essential − plateaus of skeletal complexity. Navigating these plateaus requires retrosynthetic algorithms to embrace multistep reasoning rather than incremental scoring. We identify several such multistep heuristics, including tolerable lengths of nonskeletal disconnections and the "longevity" of functional groups across routes, and anticipate that AbSynth will serve as a resource for uncovering additional trends in synthesis design. By providing a century-spanning, machine-readable data set of complete routes, this work offers a foundation for advancing chemical AI toward the complexity of human expert planning.



1. INTRODUCTION

The design of synthetic routes to complex, stereochemically rich targets remains one of the most intellectually demanding challenges in organic chemistry, often likened to an art.[1] The difficulties span several interrelated levels. First, the search space is enormous: complex synthons can be matched by hundreds[2,3] or even thousands[4] of possible reactions, generating networks of retrosynthetic options that expand with synthetic depth, D, roughly exponentially, on the order of $\sim 100^{D}$ (see ref 2). At the same time, this combinatorial explosion is not accompanied by a proportional increase in the number of viable pathways connecting the target to commercial starting materials. For most complex targets, only one or a few total syntheses have ever been reported. This scarcity is not necessarily due to lack of interest or effort, but often reflects the fact that most candidate routes lead into dead ends − that is, molecules that cannot be constructed in a selective manner. Selectivity considerations introduce another dimension of complexity: planned disconnections must be compatible with all other parts of the molecule. Some incompatibilities can be managed with protecting groups, but these must be carefully chosen to remain in place as long as needed and yet be removable at the appropriate stage. In many ways, total synthesis design resembles a multidimensional version of Sudoku, where each move must satisfy constraints imposed by the elements already present. In terms of the branching factors of viable options a chemist may legitimately consider at each step, synthesis planning is far more complex than chess − and far more frustrating, since unlike chess where

Received: April 23, 2026
Revised: August 12, 2026
Accepted: August 13, 2026
Published: September 1, 2026



ACS Publications
© 2026 The Authors. Published by American Chemical Society

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

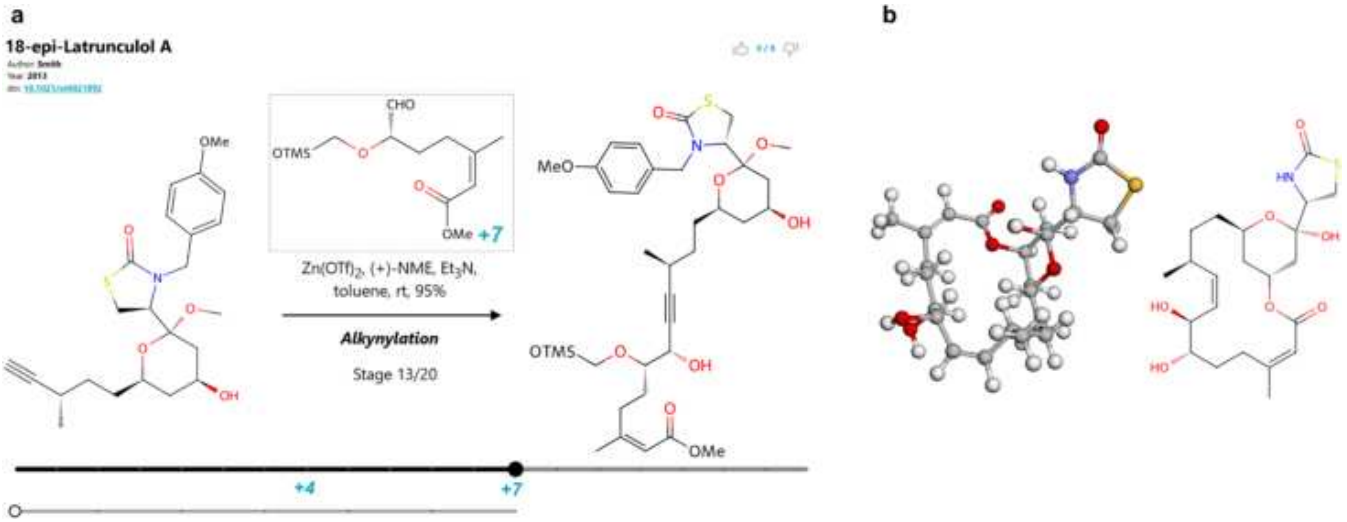


Figure 1. Screenshot from the AbSynth portal for total synthesis. AbSynth curates all 3,077 synthetic routes discussed in this study and is available at https://absynth.allchemy.net or in digitized form through a link at the end of this paper. (a) One of the steps in the synthesis of 18-epi-Latrunculol A. The slider at the bottom of the page can be used to navigate along the pathway. Steps where the pathway branches are indicated by the numbers below the slider (in blue), representing the length of each sub-branch. The smaller slider at the bottom can be used to navigate along the selected sub-branch. (b) Each molecule along the pathway can be zoomed in and displayed as a 3D model; shown here is the pathway target, 18-epi-Latrunculol A. The 3D view is useful given that the automatically generated 2D RDKit display – despite many in-house written patches to improve quality – is sometimes generating contorted molecule views (especially for caged molecules). Other options in AbSynth include searching by name, author, or year. For users who register and sign in, there are two additional options (1) Representative pathways can be downloaded as ChemDraw files (which can be useful for teaching purposes), and (2) The users can "vote" if they like or dislike specific steps (this is intended to gather, in due course, the rankings of synthetic elegance).

all games have some conclusion (win, loss, pat), many "games of synthesis" yield no viable solutions at all.

Faced with this incredible complexity, chemists have long sought principles or at least heuristics that would streamline the navigation of retrosynthetic networks. In his pioneering works, E.J. Corey[5] formulated heuristics specifying loci at which molecules should be preferentially disconnected (to, hopefully, guide searches into productive network branches), popular disconnections that should be combined into longer "strategic" sequences (to help navigate complexity "plateaus"), and motifs that should not be disconnected at all (to eliminate some unpromising branches). Others have worked on deriving metrics that would monotonously increase (or decrease) along the synthesis coordinate[6−15] – if such metrics existed, they would greatly simplify the design task to a greedy search, where one could simply take the best-scoring move (or perhaps a few best-scoring ones) for each synthon and smoothly construct the entire route without exploring the network's decoy branches.

However, majority of these heuristics and metrics have been derived either from relatively limited sets of total syntheses – e.g., only a few hundred available to Corey in the 1970s and 1980s, or the tens to hundreds used to define many complexity metrics[6−12,15] – or from larger reaction data sets but dominated by nontotal-synthesis examples.[13,14] Ideally, one would like to evaluate these measures – and also test hypotheses for additional "rules" – on a more comprehensive set of total syntheses, hopefully numbering in the thousands. Yet this entails its own hurdles. Repositories such as Chemistry by Design (CbD),[16,17] SynArchive,[18] or Hans Reich's Collection[19] would be excellent for this purpose, but they provide only images and no machine-readable information. In turn, in databases such as Reaxys or SciFinder, many total syntheses from before 1990 have not been digitized, the entry error rates (ca. 15−20% in Reaxys as estimated in refs [20,21] and SI, Section 3) are too high for accurate analysis, and Reaxys and SciFinder policies explicitly prohibit any public dissemination of their content. Confronted with these limitations, our first objective was to curate from scratch a large data set of total syntheses. This effort relied heavily – and perhaps uniquely – on the broad engagement of undergraduate students (from the 1904 Undergraduate Chemistry Club at the Jagiellonian University) working under the guidance of the more experienced colleagues from this and other institutions involved (Allchemy, JU, IOC PAS, IBS). Over several years, this group curated a set of over 3,000 total syntheses spanning close to 60,000 individual steps – this collection can be viewed in the AbSynth portal at https://absynth.allchemy.net/target_list (see Figure 1) and is made freely available to academic users in digital form from https://allchemy.net/#absynth.

With this data set in hand, we began to interrogate the architecture and logic of total syntheses at multiple levels: from the network connectivity of different routes, to trends in pathway structure, to the metrics related to the "depth of synthetic planning," and down to molecular-level reactivity. The trends observed at this last level are perhaps most interesting – documenting limitations of the existing heuristics and metrics, suggesting several new ones, and underscoring the importance of "strategic" planning over sequences of multiple steps, including those that may appear unproductive yet constitute roughly half of the data set. While analyses of this type cannot capture tacit knowledge (such as never-published failed routes, problematic substrate combinations, subtle reactivity patterns that are rarely reported, etc.), they are essential for further development of AI algorithms, which continue to rely heavily on a few publically available data sets (with the USPTO patent collection now being a standard) and are largely built on one-step-at-the-time scoring/"thinking" schemes. In this regard, the data set we disclose not only complements other public

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816172/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

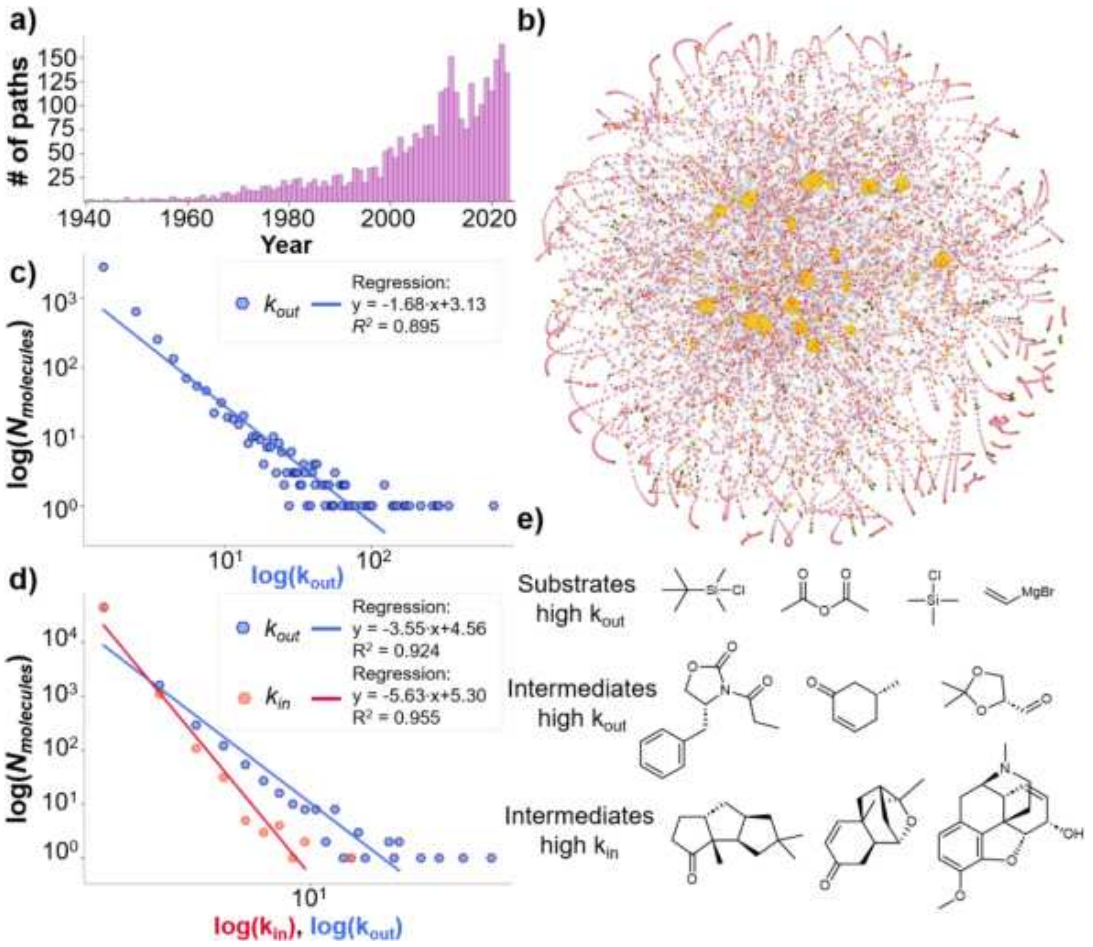


Figure 2. The network of total synthesis. (a) 3,077 total syntheses studied here and categorized according to the publication year. Notice rapid growth from 1970s onward but also a relative slowdown over the past decade. (b) A network of 300 randomly chosen syntheses in the so-called bipartite representation with*yellow* (substrate), *violet* (intermediates), and *green* (targets) nodes signifying molecules and *red* nodes denoting reaction operations. The radii of molecule nodes are proportional to their total connectivity, $k_{in} + k_{out}$. Two things to notice here are (i) the poor interconnectivity of the synthetic routes (many nonintersecting "lamella") yet (ii) the high connectivity of some of the substrates. Distributions in (c, d) are based on the full 3,077 data set. (c) A doubly logarithmic plot quantifying the distribution of the "outgoing" connectivities of the substrates (i.e., their node degrees k_{out}^{sub}) follows a power law. (d) The distributions of "incoming" connectivities, k_{in} (*red* markers and line fit) and "outgoing" connectivities, k_{out} (*blue* markers and line fit) of the intermediates are also power laws, indicating network's scale-free topology. (e) Examples of molecules with high "incoming" and "outgoing" connectivities. The most popular substrates, shown in the *top* row, include*tert*-butyldimethylsilyl chloride (k_{out} = 721), acetic acid anhydride (k_{out} = 680), and trimethylsilyl chloride (k_{out} = 335), all used primarily for protection of functional groups. Among substrates not employed for protection, the most popular one − vinylmagnesium bromide − has been used in 170 distinct reactions, mainly 1,2-additions to carbonyl moiety, i.a., in syntheses of Sphingofungin A, Resiniferatoxin and Mollanol A. In turn, many intermediates with a high number of "outgoing" connections (shown in the *middle* row) are fairly simple molecules with a defined stereocenter. For the example, (3R)-3-propionyl-4-benzyloxazolidin-2-one has been used in stereoselective aldol reactions, e.g., in syntheses of Dictyostatin, 6-deoxyerythronolide B and Aflastatin A. (R)-5-methyl-cyclohex-2-enone has served as a substrate in stereoselective additions of nucleophiles to Michael acceptors as well as in α-alkylations, e.g., in syntheses of Lycoflexine, Magellaninone and (+)-Quassin. (R)-Glyceraldehyde acetonide has been popular as a stereochemistry-inducing substrate in Darzens condensations, hetero-Diels-Alder cycloadditions and 1,2-additions of nucleophiles, e.g., in syntheses of (+)-Cavicularin, Phorboxazole A and Amphidinolide C2. Finally, intermediates with the highest "incoming" connectivities (*bottom* row) are typically very advanced molecules, used as key intermediates in multiple syntheses of popular natural products. For example, hirsutene norketone, a key intermediate of the Hirsutene syntheses, has been made in 14 different ways, i.a., by stereoselective alkylation, double bond reduction, oxidation of alcohol, and tributylstannane-mediated deoxygenation of oxalate. Another intermediate with high "incoming" connectivity, (1S,6S,7S,9S,10S)-9-methyl-8-oxatetracyclo[7.2.1.1^{7},^{10}.0^{1},^{6}]-tridec-2-en-4-one is a key intermediate in multiple syntheses of Platencimycin. It has been prepared, i.a., via stereoselective intramolecular Michael addition, stereoselective addition of alcohol to alkene, acetal hydrolysis, or double bond reduction. The third intermediate with the highest "incoming" connectivity is also a natural product − codeine − and the last intermediate in the synthesis of morphine. It has been produced, i.a., by stereoselective ketone reduction, stereoselective hydroamination, debromination, or carbamate reduction.

initiatives (notably the Open Reaction Database, ORD[22]) but also provides a unique foundation for the community to explore new paradigms in synthesis planning − a resource that can inspire more sophisticated AI models, catalyze creative human strategies, and ultimately help redefine how complex molecules are conceived and constructed.

2. RESULTS

2.1. Data Set Curation

Given the aforementioned limitations in the availability of computer-readable total synthesis data, we undertook a multiyear campaign to digitize over 3,000 total syntheses,

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

combining selection of "classic" targets from CbD and SynArchive with additional literature examples we found interesting (e.g., Cihunamide B,[23] Sesquiterpene (+)-Zizaene,[24] (±)-Isocyclocelabenzine,[25] etc.). In drawing the routes, we worked with source publications — this allowed us not only to avoid some of the mistakes in the existing repositories but also situations in which syntheses ended in complex, noncommercial starting materials, which would skew the analyses of path length and structure. In such cases, we consulted the background literature to find the pathways' "endings" tracing back to commercial substrates. Because the drawing of highly complex structures is inherently prone to human error, the project required extensive validation — ultimately more than ten rounds of corrections and cross-checks among participants. Still, occasional errors remain (estimated at 3–4% of all steps; see SI, Section 3), but these continue to be identified and corrected. In its final form, the AbSynth collection comprises 3,077 total syntheses and 59,712 individual steps, spanning almost a century of work up to 2024.

2.2. A Century of Routes Woven into a Sparse Network

We begin AbSynth analyses with a "global" look at all pathways. First, we observe that their numbers increased roughly exponentially with time (Figure 2a): there are only 39 routes from before World War II, ca. 100 for every five years between ∼1970–1990, and about 500 for every five of recent years, 2010–2024. While our set certainly does not capture*all* total syntheses, these statistics seem to confirm that key theoretical advances in the logic of synthesis planning (1970s and 1980s) were made when only several hundred paths were available — substantiating the need for a refreshed look with >3,000 routes now available for analysis.

Next, we consider the connectivity of these routes, representing the entire set as a large network. Figure 2b plots the network for a subset of 300 syntheses — what is immediately visible is its "lamellar" nature, with individual pathways sparsely interconnected. In fact, statistical analysis over the entire 3,077 set reveals that the average connectivity of the nonsubstrate nodes is small and around unity: the average number of "incoming" reaction connections to any node is k_{in} = 1.04, whereas the average number of "outgoing" connections from any intermediate is k_{out} = 1.08. For the substrates, k_{in}^{sub} = 0 by definition, whereas k_{out}^{sub} = 3.82, meaning that substrates are typically shared by several syntheses. Interestingly, distributions of both k_{in} and k_{out} (Figure 2c) as well as $k_{out}^{substrate}$ (Figure 2d) follow a power law, signifying the so-called scale-free architecture of the network and the presence of sparse but highly connected hubs (Figure 2e and corresponding caption). We observe that the scale-free architecture also characterizes the network of all organic reactions we had analyzed previously[26] — however, that Network of Organic Chemistry, NOC, has average connectivity around 2, whereas the connectivity of our network of total syntheses is only ∼1, again underscoring the fact that most natural products have (so far) been made along highly unique and nonintersecting routes.

2.3. Steady Averages, Daring Outliers, and the Rise of Convergence

Next, we turn to the nature of targets themselves. Figures 3a, b characterize the targets pursued over the years and plot, respectively, the numbers of stereocenters and rings (for other metrics yet similar conclusions, see SI Figure S1). On average, these values have increased only modestly but there is a clear

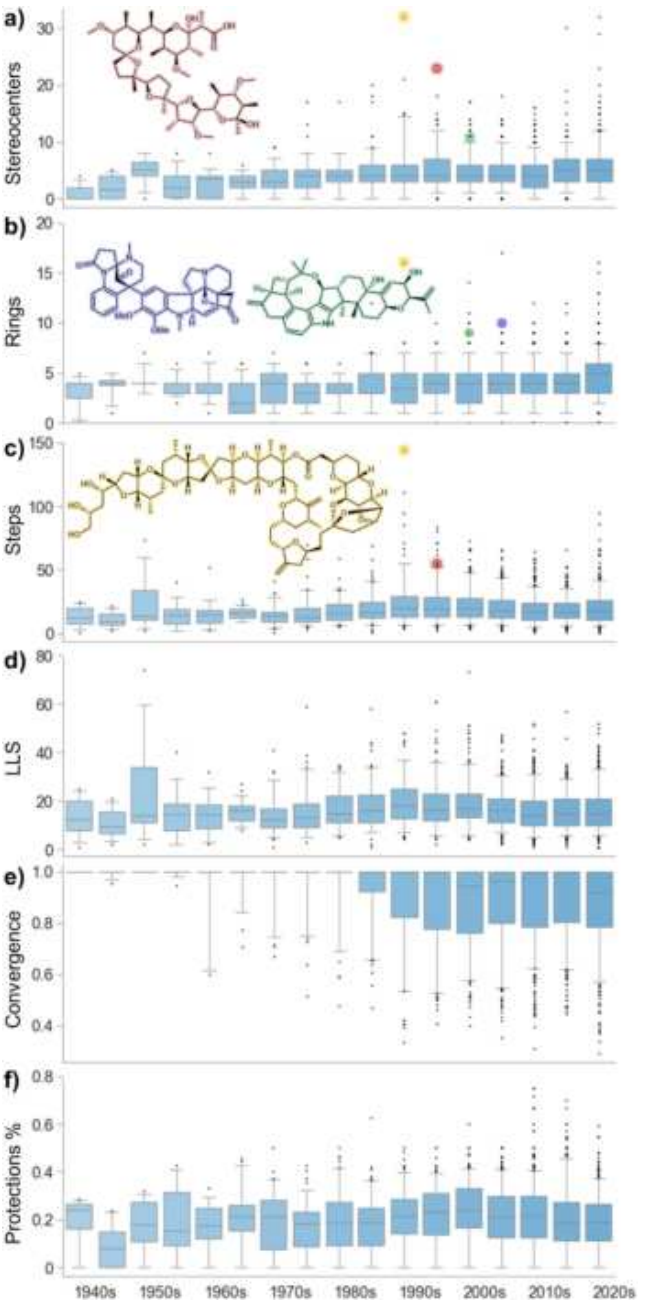


Figure 3. Time evolution of targets and path characteristics. (a, b) The plots quantify how the numbers of stereocenters and rings have evolved in the targets synthesized from 1940 to the present. Since WWII, the averages have not increased strongly although there is a growing number of "outlier" total syntheses of very complex targets denoted by individual black diamond markers. Some of these outliers are highlighted, e.g., Halichondrin B = structure and markers in *yellow*, Haplophytine = *blue*, Penitrem D = *green*, Lonomycin A = *red*. Graphs in (c, d) focus on pathway structure. The median total numbers of steps, *N*, and median lengths of the longest linear sequence, *LLS*, do not show marked increase (however, there is again a growing number of very-long-route "outliers"). (e) At the same time, starting in the 1980s, total syntheses became significantly more convergent. Convergence here is defined as *LLS*/*N* — i.e., this ratio is equal to 1 for fully linear pathways and <1 for convergent routes. (f) The ratio of the sum of protection and deprotection steps to *N* has also remained roughly constant. All plots are based on the analysis of n = 3,077 total syntheses. In all plots, boxplot elements indicate the first quartile, median, and third quartile; whiskers show the fifth and 95th percentiles, with outliers as black dots.

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740–38753

trend in the outliers — it can then be said that while the "average" complexity of natural products pursued has remained roughly constant, there are more and more successful attempts at extremely complex targets (e.g., those with >30 stereocenters and/or >10 rings), which is likely enabled by the advances in requisite methodologies. Figures 3c, d show that the total number of steps, *N*, per synthesis has also remained roughly constant, and so did the length of the longest linear sequence, *LLS*. By contrast, route convergence — here, defined as *LLS*/*N*, equal to 1 for fully linear paths and <1 for convergent ones — has improved perceptibly over the past ∼30 years (Figure 3e). Interestingly, this trend coincides with the birth of green chemistry and reflects one of its postulates[27,28] in that convergence usually translates into higher yields and less waste production. At the same time, it should be noted that total synthesis has not delivered on another green-chemistry principle to minimize the use of protecting groups — despite considerable attention from synthesis experts,[29] the ratio of protection steps to all steps has remained steady at ∼0.2 (Figure 3f). It may be that the complexity of natural products is so high that nonenzymatic syntheses with smaller numbers of protections/deprotections are simply not possible, at least given the currently available synthetic methods.

2.4. Structural Persistence and Planning Depth

Having examined the historical evolution of routes and the complexity of their targets, we now focus on how structural features persist along the pathways themselves — specifically, how successive intermediates remain related to their precursors, and what this persistence reveals about the effective depth of planning in total synthesis. To probe this, we defined an autocorrelation function $A(d) = \frac{1}{|P(d)|}\sum_{(i,j)\in P(d)} T(i, j)$, quantifying the average similarity of molecules separated along a synthetic route by a "synthetic distance" of *d* steps (Figure 4a). In this expression, $P(d)$ denotes all such pairs (i, j) in the 3,077 set for a given *d*, and $T(i, j)$ denotes Tanimoto similarity of molecules *i* and *j*. The linearity of the doubly logarithmic plot of $A(d)$ vs *d* (Figure 4b) evidence a power law, signaling the presence of "memory" or long-range persistence.[30] In turn, Figure 4c plots the Tanimoto similarity between any two molecules separated by *d* synthetic steps (across all 3,077 pathways in our collection) as a function of this synthetic distance, *d*. Interestingly, at about *d* = 10−20 steps, the average similarity (blue line) decreases to the level of the dotted red line, which corresponds to the Tanimoto similarity between any two molecules taken at random from the entire data set ZINC database[31] (similar value of approximately ∼10 steps can be obtained with Graph Edit Distance, see SI, Figure S7). Assuming that some residual similarity must be present to deem any sequence "purposefully designed," one can then say that synthesis experts plan, on average, up to ∼10−20 steps ahead — which, curiously, is akin to the depth of planning of chess masters[32] (but much lower than for the Go masters who, purportedly, can envision a position 40 steps ahead[33]).

This finding illustrates one of the key challenges for synthesis-design algorithms: to match purposeful human design that extends to such depths, these algorithms cannot be just meandering over the network of possible synthetic moves (as we saw, ca. 100^{10} across 10 steps) but must be guided by some "oracle" metrics mentioned in the Introduction and prompting our next question.

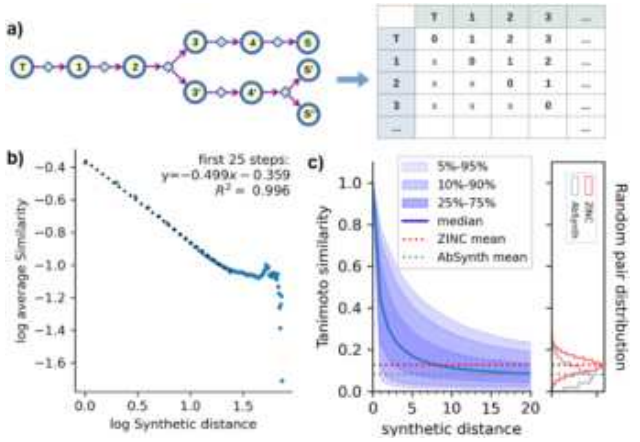


Figure 4. The depth of synthetic planning. (a) Scheme illustrating a hypothetical retrosynthetic pathway (shown as a bipartite graph) and, to the right, part of the distance matrix between different species. Note that since the graph is directed, only distances between a molecule and its precursors are defined. (b) For each possible distance along a given synthetic pathway, the Tanimoto similarity between the molecules at this distance is computed. This analysis is then repeated for all 3,077 pathways, and the average Tanimoto similarity, $A(d)$, for a given distance is calculated. The plot has the doubly logarithmic dependence of the average Tanimoto similarity on the average synthetic distance — as seen, up to ∼25 steps, this dependence is linear, signaling the presence of "memory" or long-range persistence.[30] (c) Distribution of Tanimoto similarity vs synthetic distance (for all molecules in all 3,077 syntheses) is plotted on a linear scale and compared to the average similarity between randomly drawn pairs of molecules from the AbSynth data set or ZINC database.[31] This dependence indicates that molecules separated by, on average, ∼10−20 synthetic steps are no more similar to each other than any two molecules drawn out from ZINC at random. This, in turn, can be interpreted as a complete loss of information, meaning that, akin to chess masters, synthetic experts plan total syntheses to the "depth" of ∼10−20 steps (of course, the distribution shows that there are also less frequent syntheses in which this depth is larger; lighter blue hues).

2.5. Do Synthesis-Planning "Oracles" Exist?

To answer this question, we examined how various metrics of structural and/or synthetic complexity change along the pathways. For decades, such measures have been sought[6−15] with the hope that a single metric might vary monotonically with synthesis progress, thereby allowing automated planners to advance steadily while avoiding decoy branches of the network. Yet examples of profiles for 10 total syntheses (Figure 5) evidence nonmonotonic behavior across all metrics considered, from the vintage Bertz index,[6] through further refined metrics of Whitlock,[7] Barone[8] and others,[9−12,15] to the modern ML-derived SCScore[13] and RAscore.[14] When such profiles were evaluated over all 3,077 routes, the distributions of correlation coefficients between a given metric and synthetic distance from the target were skewed toward high-correlation values of +1 or −1 (depending on the metric definition; SI, Figure S2) but the profiles also exhibited, on average, four local maxima (SI, Figure S3). The presence of such maxima indicates why none of these metrics is an "oracle" to guide synthesis planning as a simple greedy search along some steadily decreasing coordinate. This, of course, does not preclude the possibility that more effective metrics could be developed in the future. However, it is important to emphasize that, in realistic *de novo* design problems, such metrics must do more than behave monotonically along a given route (or set of routes). They must also

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

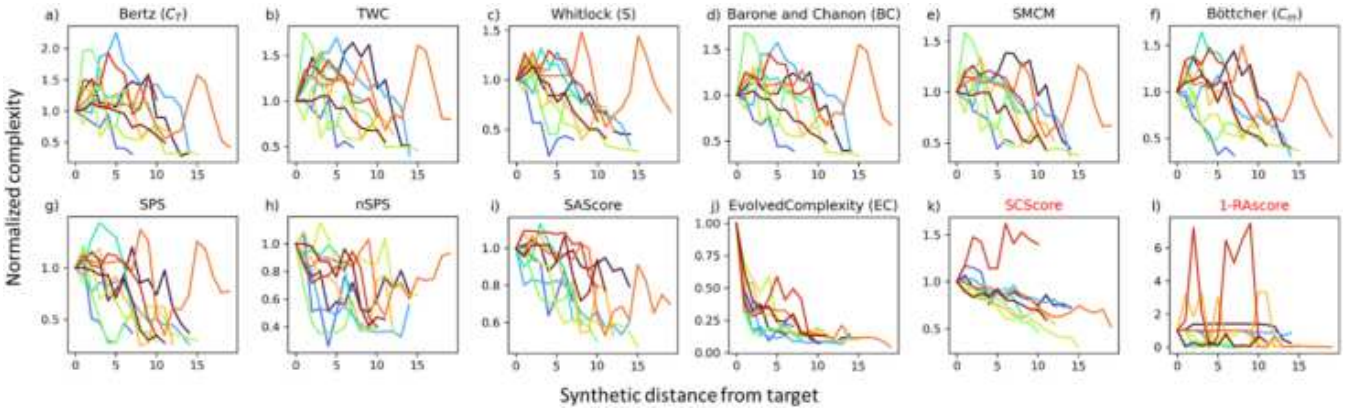


Figure 5. Examples of "complexity" profiles along synthetic routes. The horizontal axes of all plots count the numbers of synthetic steps from the target (i.e., zero corresponds to the target itself). The vertical axes have the relative complexity measures (normalized by the target's value) for the intermediates at a given step. For the clarity of presentation, we look here at 10 randomly chosen linear pathways (for the sake of visualization, we avoided branched syntheses, particularly given that linear syntheses comprise about 55% of the whole data set — similar plots could be plotted for branched routes, albeit at the expense of the said clarity). The panels are arranged roughly with the age of the metrics, with the oldest ones in the top row and the newest ones in the bottom row. Machine-Learning-based metrics are highlighted in red font. (a) Bertz score,[6] (b) Total Walk Count (TWC),[34] (c) Whitlock index,[7] (d) Barone and Chanon score,[8] (e) Synthetic Method Complexity Metric,[9] (f) Böttcher index,[10] (g) Spacial Score,[11] (h) normalized Spacial Score,[11] (i) Synthetic Accessibility Score (SAScore),[14] (j) EvolvedComplexity,[15] (k) Synthetic Complexity Score (SCScore),[13] (l) Retrosynthetic Accessibility score (RAscore; since this metric works in reverse, with 1 denoting accessible/simple targets and 0 otherwise, we took 1-RAscore for presentation).[14] The gross outlier on SCScore and RAscore is the synthesis of Epiandrosterone by Johnson in 1953.[35] For the distributions of correlation coefficients and local maxima over all 3,077 pathways, see SI Figures S2 and S3. For the correlations between individual metrics, see SI Figure S4.

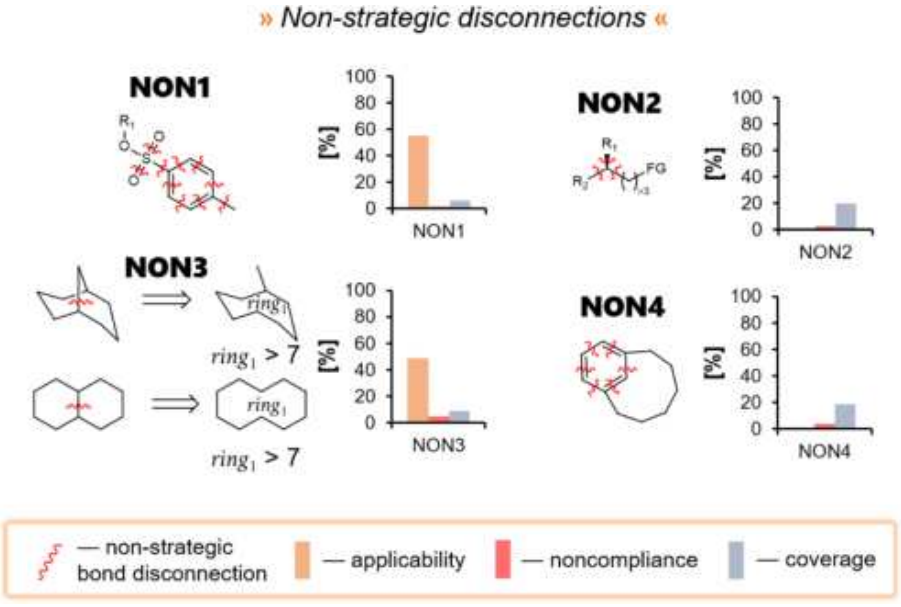


Figure 6. Schemes and statistics of structure-based and topological strategies for nonstrategic bond disconnections. All rules come from Chapter 3 of E.J. Corey's book.[5] This class of rules defines which disconnections should be disfavored, including: NON1) Building-*block*-type groups should not be internally disconnected; NON2) Bonds directly connected to remote stereocenters (far from any functional group or other stereocenter) should be preserved; NON3) disconnections of fused or bridged ring-systems, generating a new ring larger than 7-membered are disfavored; NON4) bonds within aromatic and heteroaromatic rings, which are part of bridged ring systems are not strategic. Red waves represent nonstrategic disconnections. If more than one red wave is placed over the structure, none of the indicated bonds should be disconnected according to the rule. Each rule is accompanied by the chart visually representing its applicability — orange bars (in what % of reaction data set, the given rule could be applied), noncompliance — red bars (in what % of applicable reactions, the indicated nonstrategic bond was actually disconnected), and coverage (what % of molecule's bond are covered by the rule).

discriminate productive solutions from the ∼100 other chemically plausible options that can match a given retron at each step — a challenge often overlooked when AI scoring functions are trained solely on published reaction collections, while disregarding alternative transformations that would also be chemically valid.

2.6. Can Corey's Heuristics Help the Algorithms?

If oracles are unavailable, perhaps classical heuristics can assist. Retrosynthetic reasoning has long been associated with Corey's celebrated rules,[5] particularly those governing bond disconnection preferences. These heuristics have undeniably helped

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816172/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

generations of chemists navigate complex synthetic landscapes. Yet, it remains an open question whether they embody a level of strategic logic sufficiently robust for *algorithmic* deployment – capable of guiding synthesis planning reliably whenever, or at least in most cases when, a retron presents a matching structural motif.

To explore this, we encoded four nonstrategic disconnection rules (i.e., cases where bond cleavage is typically discouraged) and a set of 22 strategic (i.e., preferred) disconnection rules. These rules were implemented RDKit's structural analysis functions and SMARTS-based pattern recognition, with the complete list provided in the SI Section 2.3. These "templates" were systematically applied to all 59,712 reaction steps across 3,077 total syntheses, enabling large-scale assessment of their prevalence and predictive utility.

For nonstrategic disconnections, we considered three metrics: (1) Applicability, counting the percentage of steps, in which a given rule template matched the retron and *could*, in principle, be considered/applied; (2) Noncompliance, the frequency with which a rule was violated but the indicated bond was nonetheless disconnected; and (3) Coverage, the proportion of all bonds in a molecule that are flagged as nonstrategic. As illustrated in Figure 6, noncompliance is consistently low (≤5%), suggesting strong adherence to these constraints. Rules NON1 and NON3 are the most frequently applied, though they cover only a modest fraction of bonds (6% and 9%, respectively). Nonetheless, these nonstrategic-disconnection heuristics do offer practical value in algorithmic retrosynthesis by helping eliminate unproductive disconnection sites.

For strategic disconnections, we focused on bonds not flagged by the strong NON heuristics. We retained the definition of applicability but introduced two additional metrics: (1) Accuracy, defined as the percentage of steps in which the retron-matching rule *was*, indeed, executed in experiment; and (2) Specificity, defined as the fraction of bonds/sites actually disconnected to all bonds/sites marked by a given rule (see Figure 7 for illustrative examples). For the rules to have high predictive power, they should be (i) highly accurate – meaning that whenever a given rule matches the retron, it should be experimentally actionable (i.e., really executed in our experimental 3,077 data set), and (ii) highly specific – ideally, the rule should match a given molecule at only one locus.

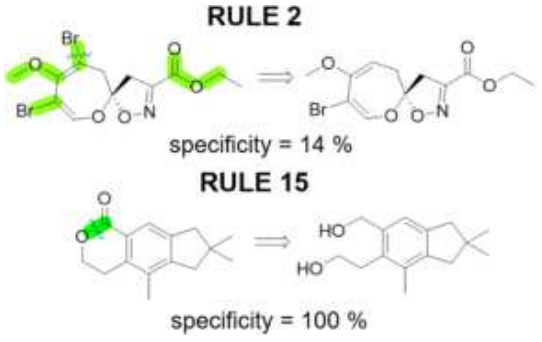


Figure 7. Illustration of rule specificity, defined as the fraction of bonds actually disconnected to those indicated by the rule. In the first example on the top, from the syntheses of 7-Deoxypsammaplysins K and O,[36] rule #2 (describing strategic disconnection of acyclic carbon-heteroatom bond) indicates seven bonds (highlighted in green) as possibly strategic. In the actual synthesis, only one of these bonds (marked by a blue wave) was disconnected, translating into 1/7 or 14% specificity. In the example on the bottom – taken from the total synthesis of Alcyopterosin O[37] – rule #15 (describing strategic disconnection of easily formed rings such as lactones, hemiacetals, hemiketals in fused ring systems) indicated only one bond as strategic, and this exact bond was actually disconnected (indicated by blue wave), resulting in 100% specificity.

Surprisingly, the results summarized in Figure 8 indicate that the specificity of only ca. half of the rules is high, while accuracy is quite low across the entire set, with only seven rules exceeding 25% accuracy (rules #2, #3, #4, #11, #13, #18, #19). Rule #2, which promotes disconnection of acyclic carbon-heteroatom bonds, is the most frequently applied and achieves the highest accuracy (46%), yet its specificity is poor (28%), i.e., typically flagging ca. three bonds of which only one is disconnected. Rule #13, which targets cocyclic bond pairs – specifically, a fusion bond and an adjacent off-exendo bond – stands out as the only rule combining moderate accuracy with high specificity, though its applicability is limited (6%).

2.7. The Importance of Multistep Planning

The conclusion – perhaps not entirely unexpected – from the above analyses is that strategic-disconnection rules, while valuable for human retrosynthetic reasoning, are not actionable *algorithmically* due to limited specificity and low accuracy. Crucially, they do not offer guidance on how to reach the molecular contexts in which they become applicable – for example, how to manipulate or protect functional groups to resolve reactivity conflicts, or how to achieve substitution patterns that activate specific bonds for disconnection. Human experts navigate these challenges using deep intuition and accumulated experience.

Algorithms, by contrast, derive no structural simplification from these manipulations, yet face the aforementioned,[2,3] 100^D, exponential growth in retrosynthetic graph size – expanding by a factor of ∼100 with each additional synthetic step explored. We previously encountered this issue in computational planning of natural product syntheses,[3] where incorporating strategic-disconnection rules into the scoring function failed to improve route identification. However, in the same work,[3] we observed that synthesis-planning algorithms markedly improved their performance upon inclusion of hard-wired multistep sequences. These included popular functional group interconversions (FGIs), some disconnections traditionally considered strategic, and a larger set derived algorithmically[38] – albeit without detailed knowledge of their applicability to total syntheses. AbSynth now provides a unique opportunity to analyze the actual multistep sequences used in total synthesis, offering a foundation for identifying which patterns may serve as productive heuristics for future algorithmic planning.

2.8. Multistep Heuristics from AbSynth Analysis

Building on this idea of multistep "thinking," we applied a statistical approach to search for regularities in step sequences. While most trends we considered were weak, some proved strong enough to provide practical guidance for synthesis design. The value of these heuristics – now implemented in our own retrosynthetic algorithms[4,39] – lies in pruning the unpromising branches of retrosynthetic search graphs, thereby helping to avoid "combinatorial explosion"[40] of options that have plagued computational synthesis design for decades.

2.8.1. Trimming Excursions over Synthetic Plateaus.

The first heuristics concerns the tolerable length of synthetic sequences that do not simplify the target skeleton directly but instead prepare it for subsequent disconnections. For simple molecules, such as most small-molecule drugs, successive

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740–38753

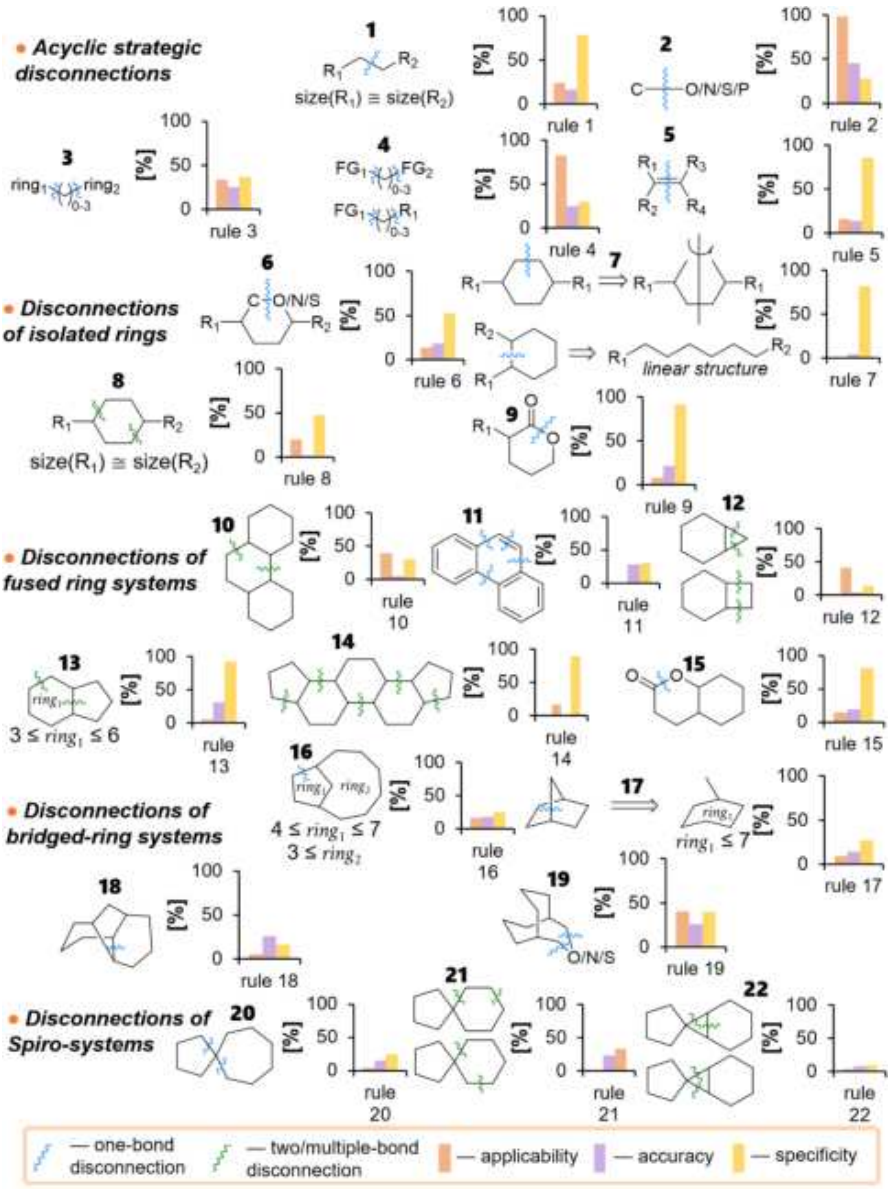


Figure 8. Schemes and statistics of structure-based and topological strategies for strategic bond disconnections. All rules come from Chapter 3 of E.J. Corey's book.[5] Rules described too vaguely and imprecisely have been omitted. Blue waves represent one-bond disconnections. If more than one blue wave is placed on the structure, any of the indicated bonds can be disconnected according to the rule. Green waves represent two-bond or multibond disconnections, which means that all indicated bonds have to be disconnected in the reaction. Each rule has been accompanied by the chart visually representing its applicability — orange bars (in what % of the experimental reaction data set, a given rule could be applied); accuracy — violet bars (in what % of applicable reactions, the rule correctly predicted a disconnection that was actually carried out experimentally); and specificity (what % of bonds predicted by the given rule have been disconnected). Strategic bond disconnections have been divided into five categories, based on the topology of the structure. The first class is for strategies promoting disconnections of acyclic bonds: 1) resulting in two molecules of approximately the same size, 2) disconnecting bonds between carbon and heteroatom, 3) applying to bonds attached directly to the ring or within 1−3 bond distance, along the pathway connecting two rings/ring systems, 4) applying to bonds along the path of 1−3 carbon atoms connecting two functional groups or within 1−3 carbon atoms distance from a functional group, 5) applying to internal alkenes. The second class groups rules promoting disconnection of isolated rings. These heuristics apply to 6) bonds between carbon and heteroatom in centrally located rings, 7) bonds in centrally located rings leading to symmetrical or linear molecules, 8) bond pairs resulting in two molecules of approximately the same size, 9) easily formed rings such as lactones, hemiacetals, hemiketals, with no restriction to central location. The third class describes heuristics promoting disconnection of fused rings and applies to 10) cocyclic pair of bonds which are exendo to a fusion bond in centrally located rings, 11) bonds of the central benzenoid rings in the polycyclic systems, 12) [2 + 1] and [2 + 2] of 3- and 4-membered rings, 13) cocyclic pair of bonds involving fusion bond and one bond away (offexendo), 14) multiple contiguous exendo bonds as candidates for polyannulation transforms, 15) easily formed rings, analogously to rule 9 for isolated rings. The fourth class is for rules promoting disconnections of bridged rings and applies to 16) exendo bonds within a 4−7 membered ring and exo to a ring larger than 3-membered, 17) generation of a new ring smaller than 8-membered, 18) bonds within a ring of maximum bridging (containing the highest number of bridgehead atoms), 19) heterobonds involving oxygen, nitrogen or sulfur. The fifth class describes suggested disconnections of spiro rings and applies to 20) exendo bonds, 21) bond pairs − an exendo bond and a cocyclic bond beta to it, 22) bond pairs − an exendo bond and a cocyclic bond next to it in a 3-membered ring.

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

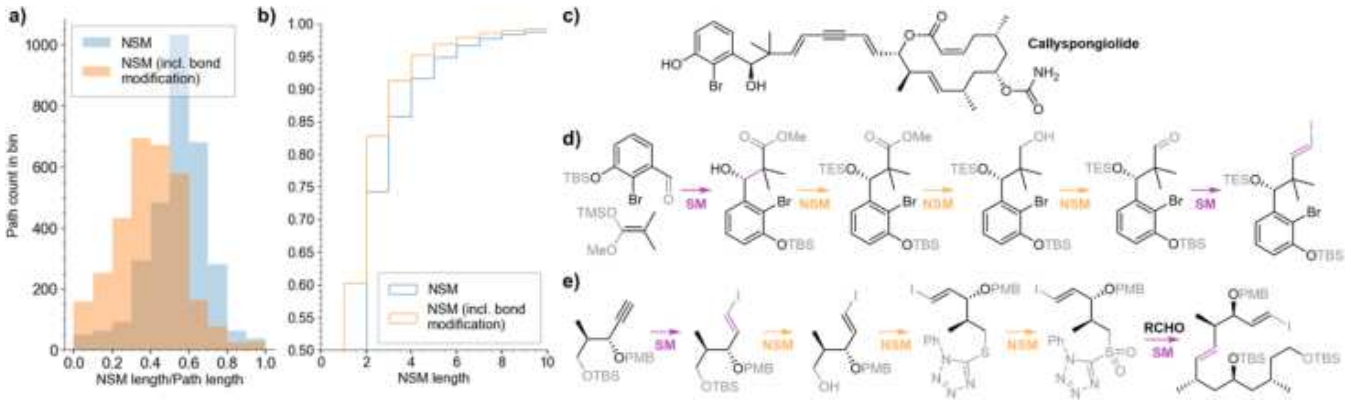


Figure 9. Nonskeletal modification (NSM) steps in synthetic pathways. (a) Distributions of the fractions of steps identified as nonskeletal modifications in all 3,077 total syntheses. The blue plot has the distribution when only skeletal bond creation is counted as skeletal modification (SM), orange plot has the distribution for which bond order change is also counted as SM. (b) Cumulative distribution of NSM sequences lengths. Color coding as in panel (a). (c) Structure of Callyspongiolide and (d, e) Examples of NSM sequences used in its synthesis.[41] Color-coding: Nonskeletal bonds and atoms are in *light-gray*; newly formed skeletal bonds are highlighted in *purple,* and so are the corresponding step arrows. (d) NSM sequence (protection of an alcohol, reduction of an ester, and oxidation of an alcohol) applied after the Kiyooka aldol SM enables synthesis of homologated aldehyde for subsequent vinyl iodide formation. (e) Three step NSM sequence (deprotection, Mitsunobu reaction, and oxidation) used for the preparation of sulfone required for Julia-Kocienski olefination, bridging prior and subsequent SM steps.

skeletal disconnections (i.e., disconnections of bonds present in the target molecule and thus simplifying the structure) are often sufficient to design a route — and algorithms can now perform such tasks with relative ease. In contrast, the total synthesis of complex natural products requires more intricate planning, with key disconnections interspersed by sequences of transformations that leave the molecular connectivity unchanged (e.g., functional group interconversions, protection or deprotection steps). Statistics in Figure 9a show that such nonskeletal modifications (NSMs, transformations that do not create any bonds present in the final target molecule) account for as much as 40% of all synthetic steps — or, if one includes cases where a skeletal bond is modified but not disconnected (e.g., converted from double to single), even more than 50%. Notwithstanding, it should be remembered that the sole role of NSMs is to resolve reactivity problems hampering the execution of structure-simplifying steps — as such, these NSM sequences are unproductive (but necessary!) and should be kept as short as possible. How short? One answer to this question is provided by the cumulative distributions plotted in Figure 9b, which indicate that in 90−95% of cases, human experts relied on NSM sequences no longer than four steps (irrespective of the stage of synthesis, see SI Figure S5, and nearly constant over time, Figure S8). This observation motivates heuristics stipulating that when expanding the retrosynthesis-planning graph, NSM branches longer than four steps should be truncated or at least heavily penalized.

2.8.2. Limiting Drag-Along of Problematic Groups. The second heuristics considers the "longevity" of functional groups along synthetic routes. To quantify this, we collected statistics of the length of sequences in which specific groups survived a certain number of steps. In Figure 10, these lengths are quantified as cumulative distributions and are color-coded such that the longer the *red* streak in a given row, the fewer synthetic examples exist in which the specific group survives any steps. For instance, thiols were found in the total of 29 sequences, but in 90% of them, they were consumed immediately after introduction (in the figure, *red*-colored sequence length = 0 extends to the fourth column, corresponding to 90% cumulative probability). This reflects a common knowledge that it is difficult to drag along a synthetic route highly reactive groups like the said thiols, but also alkyl iodides, triflates, or enolizable aldehydes. An interesting extension is that some *red* rows correspond to groups that are not extremely labile but rather polar (carboxylic acids, boronic acids, primary amines) — their minimal "longevity" may likely reflect isolation difficulties. By contrast, *green* rows correspond to robust functionalities (e.g., terminal alkenes, internal alkynes, or aryl bromides) that can easily persist over multiple steps. From the perspective of synthesis-planning algorithms, the rule we advocate is straightforward: synthetic branches in which red-row groups are not consumed (i.e., converted into other groups or protected) immediately after introduction should be removed.

2.8.3. Avoiding Combinations of Highly Reactive Groups. The third heuristic formalizes an intuitive expectation: synthons containing multiple reactive groups of different types (for example, COOH, NH_2, SH, and I all within one molecule) are not only difficult to synthesize but, even if prepared, are prone to undesired side reactions in subsequent steps. To quantify this expectation, we analyzed 30 reactive groups (including alkyl and vinyl halides, Michael acceptors, epoxides, thiols, imines, as listed in Figure 11a) and counted their occurrence across intermediates in 3,077 syntheses. The statistics in Figure 11b reveal that most molecules contained none of these reactive groups; approximately 14,000 intermediates contained one, ∼1,600 contained two groups of different types (e.g., one azide and one carboxylic acid), ∼120 contained three, and only 15 contained four or more. Moreover, decade-stratified analysis (see SI, Figure S6) showed that this pattern held consistently over time, with only a modest rise in multigroup intermediates from ∼1.5% to ∼3%. Based on these distributions, we propose a practical rule: intermediates containing more than three different reactive groups from the defined set should be removed from consideration — or, in a more stringent variant, even those with more than two.

3. CONCLUSIONS

Beyond satisfying curiosity about how human experts design synthesis routes to complex targets, the most consequential

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

	tot-seq	longest	25%	50%	75%	90%	95%	99%
Alcohol	1373	14	0	1	2	3	4	6
Carboxylic Acid	1039	9	0	0	1	1	2	3
Phenol	288	8	0	1	2	3	5	6
Thiol	31	2	0	0	0	0	1	2
Primary Amine	498	5	0	0	1	1	2	3
Secondary Amine	472	8	0	0	1	2	2	4
Primary Alkyl-Cl (act)	79	6	0	0	1	2	3	4
Primary Alkyl-Cl	91	5	0	1	2	4	5	5
Primary Alkyl-Br (act)	222	4	0	0	0	0	1	3
Primary Alkyl-Br	174	11	0	0	1	3	5	10
Primary Alkyl-I (act)	52	2	0	0	0	0	0	1
Primary Alkyl-I	292	6	0	0	0	1	2	4
Nitro	324	25	1	2	4	8	13	22
Azide	170	19	0	1	3	6	9	16
Terminal Alkene	1406	27	1	3	6	9	12	18
Terminal Alkyne	463	12	0	1	2	4	5	8
Internal Alkyne	601	25	0	1	3	6	7	11
Allene	76	11	0	1	1	6	7	11
Aryl-I	169	14	1	1	2	6	8	12
Aryl-OTf	79	4	0	0	0	1	2	3
Aryl-Br	384	23	1	2	4	7	10	17
Vinyl-Br	260	23	0	1	2	5	6	9
Vinyl-I	340	15	0	1	3	5	7	11
Vinyl-OTf	228	16	0	0	0	3	5	12
Boronic Acid	39	1	0	0	0	0	0	1
Boronic Ester	164	11	0	0	1	2	4	8
Enolizable Aldehyde	849	2	0	0	0	0	1	2
Terminal Epoxide	202	13	0	0	1	1	2	4
EWG-Alkene (β-CH2)	274	5	0	0	1	2	2	4
EWG-Alkene (β-sub)	1468	15	0	1	2	4	5	9

Figure 10. Functional group "longevity" along synthetic paths. For each group, the "tot-seq" column lists the total number of sequences (from all 3,077 syntheses) in which a given group was found. The "longest" column gives the longest sequence (measured in synthetic steps) over which the group survived intact (e.g., there exists a total synthesis of (−)-Flueggenine D[42] in which an unprotected alcohol was dragged along for 14 steps). The subsequent "#%" columns quantify the cumulative distribution of a specific sequence length. For instance, for alcohols, 25% of sequences had zero length − meaning the alcohol was immediately converted to another group. In 50% of sequences, the alcohol survived for, at most, one step (i.e., either zero or one step); in 75% of sequences, it survived for, at most, 2 steps, etc. The intuitive way to read these distributions is using the color-coding, where *red* highlights low sequence lengths − in other words, the longer the *red*/*reddish* streak in a given column, the fewer synthetic examples in which a group would survive any or one step. For discussion of specific examples, see the main text.

outcome of our analysis lies in its implications for computational synthesis design. It demonstrates that advanced synthesis planning cannot be reduced to a series of purely "local" decision-making tasks, in which the algorithm "thinks" only one-step-at-a-time. This reductionist view has long motivated the search for metrics that vary monotonically along a synthetic trajectory − had such "scoring functions" existed, they would allow algorithms to traverse synthesis space greedily (or nearly so), selecting the best option at each step without exploring vast graphs of possibilities. Yet our statistical study of more than 3,000 classical total syntheses reveals that no such universally applicable metrics have been identified. Likewise, if E. J. Corey's heuristics − historically so powerful in guiding human design − were consistently accurate and specific, they could direct bond disconnections at each step, narrowing the search space and mitigating the combinatorial explosion. As shown here, however, these heuristics apply only sporadically and lack the specificity required to automate most of the design process (although it would be interesting to see if a machine-learned function based on the *combination* of these rules would be more accurate).

In parallel, we observed (Section 2.8.1) that approximately half of the synthetic steps do not reduce skeletal complexity, though they are still essential to prepare the skeleton for key disconnections. Navigating these "plateaus" of structural complexity is precisely what makes synthesis planning so difficult for algorithms. Corey made an inspired contribution to this problem by introducing the concepts of *strategic sequences* and *tactical combinations*. At the time, however, such sequences were few and not embedded within algorithmic frameworks for full-route planning (the pioneering LHASA program operated one move at a time, with human choices guiding the expansion of the retrosynthetic graph). Today, by contrast, they can be generated algorithmically and at scale.[38] In our own work on computational synthesis design targeting natural products,[3,4] we found that incorporating such "hard-wired" sequences − some forming higher-level combinations as long as six steps − was a

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

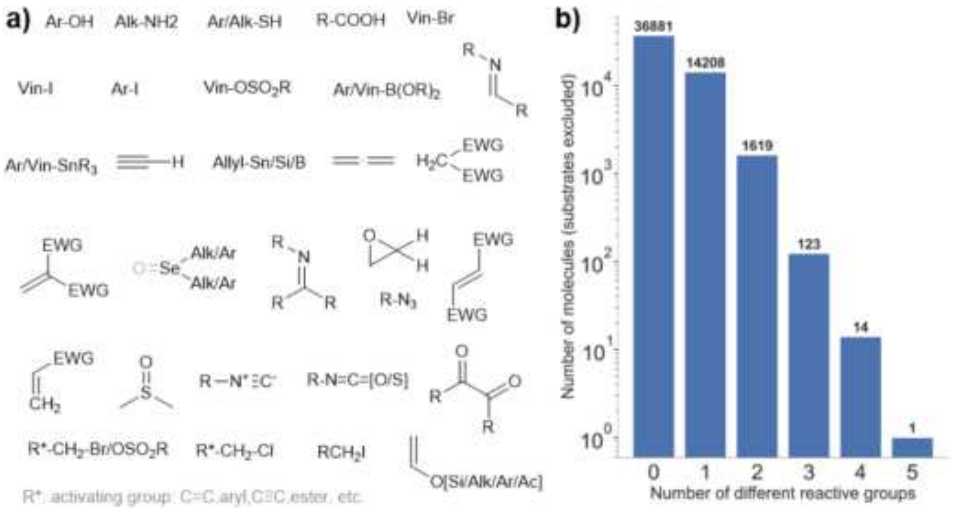


Figure 11. Statistics of reactive functional groups. (a) List of 30 functional groups considered reactive. (b) Numbers of compounds found in the set of 3,077 total syntheses, in which a given number of different reactive groups was found. Notice logarithmic vertical axis. While tens of thousands of molecules contain no or just one reactive group, there are only 15 that contain more than three different reactive groups.

key factor in improving algorithmic performance. This enabled access to targets such as Callyspongiolide, Scabrolide A, Conidiogenone B, or Taxuyunnanine D, although not yet to the next tier of complexity exemplified by molecules like Pentamycin, Taxol, or Palytoxin. In that study, we selected sequences based on subjective expert knowledge. However, collections like the 3,077-set examined here offer a more objective basis for identifying sequences that are, in fact, commonly used in total synthesis.

These considerations have implications not only for our own algorithms[2−4,38,39] but for the broader landscape of chemical AI and retrosynthetic tools. Despite nearly a decade of effort by many groups and architectural innovations in neural networks, virtually all these algorithms (e.g., ASKCOS,[43] AiZynth,[44,45] Reaxys,[46] SciFinder,[47] and many others) still operate in a one-step-at-a-time fashion, evaluating synthesis progress incrementally. This conceptual limitation may explain why such systems remain confined to relatively simple targets, where skeletal disconnections dominate,[48,49] (for additional evaluation of ASKCOS on 50 targets from AbSynth, see SI, Section S4). Of course, algorithmic approaches may also be constrained by factors such as the reliance on predefined templates, heuristics, and incomplete molecular representations. In this regard, reasoning agents promise to incorporate richer topological and contextual descriptions of molecules, enabling more adaptive and human-like decision making. Recently emerging techniques incorporating LLM-based reasoning algorithms[51] promise to do exactly that but are still limited by, e.g., the insufficient consideration of side reactions,[51] limited understanding of more advanced (and less popular) chemistry,[52] or prone to hallucination and errors in the predicted reactions.[53]

To tackle more complex molecules and gain traction among expert synthetic chemists, future methods (algorithmic or agentic) aiming at the construction of complete routes must embrace multistep reasoning − and this should be done not by bypassing strategic sequences and merely "jumping" between key disconnections, but by taking on the crucial task of matching and executing those strategic sequences themselves, rather than leaving it to the human user.[50]

In this study, we have illustrated several such heuristics, but importantly, the AbSynth collection itself provides a foundation for uncovering additional multistep trends. We hope that readers will explore these data further, extending the scope of retrosynthetic analysis and enriching the design strategies available to both human chemists and computational tools.

The final comment underscores the need to further expand and continuously curate the AbSynth data set − it is definitely work in progress. So far, the repository has been assembled by young enthusiasts from a distant country and a small, self-supporting startup. Looking ahead, however, we see no compelling reason why large organizations such as CAS (part of the nonprofit ACS) should not share at least portions of their aggregated SciFinder content with the broader community − content that is, in large measure, derived from research supported by public funding worldwide. Such a gesture would align squarely with the FAIR Data Principles[54] that ACS professes to uphold,[55] and would represent a meaningful step toward a more open and equitable scientific ecosystem. In parallel, we hope that the synthesis community will help augment AbSynth with their own contributions − not only successful total syntheses but also those that failed. Indeed, the latter would be especially valuable, as robust synthesis-planning algorithms cannot be trained without understanding the factors that cause reactions to fail (for a discussion of the importance of this largely unpublished "dark knowledge," see our recent ref 49). By embracing both successes and failures, AbSynth can ultimately evolve into a truly comprehensive resource − one that captures the full spectrum of synthetic practice and accelerates the development of more insightful, resilient, and creative approaches to computational synthesis design.

■ ASSOCIATED CONTENT

Data Availability Statement

All synthesis pathways are available for inspection through the AbSynth portal available at https://absynth.allchemy.net/target_list. Collection of SMILES from this dataset are available from https://allchemy.net/#absynth Codes for computer-encoded Corey heuristics are deposited at https://zenodo.org/records/21838008.

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.6c08434.

Additional details of data curation and statistical trends (PDF)

AUTHOR INFORMATION

Corresponding Authors

Bartosz A. Grzybowski − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland; Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea;* orcid.org/0000-0001-6613-4261; Email: nanogrzybowski@gmail.com

Yasemin Bilgi-Gadina − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland;* Email: bilgiyaseminn@gmail.com

Daniel T. Gryko − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland;* Email: danieltgryko@gmail.com

Sara Szymkuć − *Allchemy, Inc., Highland, Indiana 46322, United States*; Email: sara.szymkuc@gmail.com

Authors

Maciej Czuba − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Konrad Barnowski − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Karol Molga − *Allchemy, Inc., Highland, Indiana 46322, United States*

Zofia Pławecka − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Agnieszka Wołos − *Allchemy, Inc., Highland, Indiana 46322, United States*

Szymon Stodółkiewicz − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Sebastian Baś − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland;* orcid.org/0000-0003-4692-5757

Wojciech Petrykowski − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland;* orcid.org/0009-0007-5490-8371

Wiktor Beker − *Allchemy, Inc., Highland, Indiana 46322, United States*

Evan Saillard − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Konrad Łęgowski − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Rafał Roszak − *Allchemy, Inc., Highland, Indiana 46322, United States*

Paweł Kowalczyk − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland*

Barbara Mikulak-Klucznik − *Allchemy, Inc., Highland, Indiana 46322, United States*

Piotr Janasz − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Emilia Mielke − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland;* orcid.org/0000-0002-8196-0104

Patrycja Pietruszewska − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland;* orcid.org/0009-0001-2915-8886

Maja Łazuchiewicz − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Sylwia Wójcik − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Jakub Zabielski − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Dariusz Nostkiewicz − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Gabriela Smaga − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland;* orcid.org/0009-0008-7820-8107

Ahmad Makkawi − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland*

Louis P. J.-L. Gadina − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland*

Jakub Skoczek − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Tomasz Klucznik − *Allchemy, Inc., Highland, Indiana 46322, United States*

Maxence Deschamps − *1904 Undergraduate Chemistry Club at the Faculty of Chemistry, Jagiellonian University, Kraków 30-387, Poland*

Anna Żądło-Dobrowolska − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland;* orcid.org/0000-0003-4240-9851

Michał Michalak − *Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland;* orcid.org/0000-0002-1193-1551

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.6c08434

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Vice-Rector of JU, Prof. Wojciech Macyk, for supporting this work within the 1904 Undergraduate Chemistry Club at the Jagiellonian University. We also thank I. Stoń, A. Seltenreich, and A. Pawlik for help in the early stages of the project. This work received support from the internal funds of Allchemy, Inc. This project has also received funding from the European Research Council (ARCHIMEDES, 101097337). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. S.B. acknowledges support from the Polish

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816172/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

National Science Centre, Poland (SONATA17; UMO2021/43/D/ST4/02948). Conceptualization and manuscript writing by B.A.G. were generously supported by the taxpayers of South Korea through the Institute for Basic Science, project code IBS-R020-D1.

REFERENCES

(1) Nicolaou, K. C. The Emergence of the Structure of the Molecule and the Art of Its Synthesis. *Angew. Chem., Int. Ed.* **2013**, *52*, 131−146.
(2) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem., Int. Ed.* **2016**, *55*, 5904−5937.
(3) Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska, O.; Beker, W.; et al. Computational Planning of the Synthesis of Complex Natural Products. *Nature* **2020**, *588*, 83−88.
(4) Molga, K.; Beker, W.; Roszak, R.; Czerwiński, A.; Grzybowski, B. A. Hierarchical Reaction Logic Enables Computational Design of Complex Peptide Syntheses. *J. Am. Chem. Soc.* **2025**, *147*, 7644−7662.
(5) Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis*; Wiley-Interscience: New York, 1989.
(6) Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599−3601.
(7) Whitlock, H. W. On the Structure of Total Synthesis of Complex Natural Products. *J. Org. Chem.* **1998**, *63*, 7982−7989.
(8) Barone, R.; Chanon, M. A New and Simple Approach to Chemical Complexity. Application to the Synthesis of Natural Products. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 269−272.
(9) Allu, T. K.; Oprea, T. I. Rapid Evaluation of Synthetic and Molecular Complexity for *In Silico* Chemistry. *J. Chem. Inf. Model.* **2005**, *45*, 1237−1243.
(10) Böttcher, T. An Additive Definition of Molecular Complexity. *J. Chem. Inf. Model.* **2016**, *56*, 462−470.
(11) Krzyzanowski, A.; Pahl, A.; Grigalunas, M.; Waldmann, H. Spacial Score—A Comprehensive Topological Indicator for Small-Molecule Complexity. *J. Med. Chem.* **2023**, *66*, 12739−12750.
(12) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminf.* **2009**, *1*, 8.
(13) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. SCScore: Synthetic Complexity Learned from a Reaction Corpus. *J. Chem. Inf. Model.* **2018**, *58*, 252−261.
(14) Thakkar, A.; Chadimová, V.; Bjerrum, E. J.; Engkvist, O.; Reymond, J.-L. Retrosynthetic Accessibility Score (RAscore)—Rapid Machine-Learned Synthesizability Classification from AI-Driven Retrosynthetic Planning. *Chem. Sci.* **2021**, *12*, 3339−3349.
(15) Cuomo, A.; et al. EvolvedComplexity as a Total Synthesis Assessment Metric: Strychnine as a Case Study of Scoring Functions. *ChemRxiv* **2025**.
(16) Draghici, C.; Njardarson, J. T. Chemistry by Design: A Web-Based Educational Flashcard for Exploring Synthetic Organic Chemistry. *J. Chem. Educ* **2012**, *89*, 1080−1082.
(17) *Chemistry By Design*. https://chemistrybydesign.oia.arizona.edu/. (accessed March, 2026).
(18) *SynArchive—The Organic Synthesis Archive*. https://synarchive.com/. (accessed March, 2026).
(19) Hans Reich's Collection *Total Syntheses*. https://organicchemistrydata.org/hansreich/resources/syntheses. (accessed March 2026).
(20) Szymkuć, S.; Badowski, T.; Grzybowski, B. A. Is Organic Chemistry Really Growing Exponentially? *Angew. Chem., Int. Ed.* **2021**, *60*, 26226−26232.
(21) Chen, L.-Y.; Li, Y.-P. AutoTemplate: Enhancing Chemical Reaction Datasets for Machine Learning Applications in Organic Chemistry. *J. Cheminf.* **2024**, *16*, 74.
(22) Coley, C. W.; Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143*, 18820−18826.
(23) Yu, L.; Nagata, Y.; Nakamura, H. Atroposelective Total Synthesis of Cihunamide B. *J. Am. Chem. Soc.* **2024**, *146*, 2549−2555.
(24) Zhu, Y.; Zheng, J.; Evans, P. A. Intramolecular Rhodium-Catalyzed [(3 + 2+2)] Carbocyclization Reactions with Dienylidene-cyclopropanes: A Concise and Stereoselective Total Synthesis of the Sesquiterpene (+)-Zizaene. *J. Am. Chem. Soc.* **2023**, *145*, 3833−3838.
(25) Saikia, A. K.; Devi, N. R.; Shit, S.; Behera, B. K. Synthesis of 4-Vinyl-1,2,3,4-Tetrahydroisoquinoline from N-Tethered Benzyl-Alkenol Catalyzed by Indium(III) Chloride: Formal Synthesis of (±)-Isocyclocelabenzine. *Synthesis* **2020**, *52*, 1425−1434.
(26) Fialkowski, M.; Bishop, K. J. M.; Chubukov, V. A.; Campbell, C. J.; Grzybowski, B. A. Architecture and Evolution of Organic Chemistry. *Angew. Chem., Int. Ed.* **2005**, *44*, 7263−7269.
(27) Poliakoff, M.; Fitzpatrick, J. M.; Farren, T. R.; Anastas, P. T. Green Chemistry: Science and Politics of Change. *Science* **2002**, *297*, 807−810.
(28) Anastas, P. T.; Warner, J. C. *Green Chemistry: theory and Practice*; Oxford University Press: New York, 1998.
(29) Young, I. S.; Baran, P. S. Protecting-Group-Free Synthesis as an Opportunity for Invention. *Nat. Chem.* **2009**, *1*, 193−205.
(30) Witt, A.; Malamud, B. D. Quantification of Long-Range Persistence in Geophysical Time Series: Conventional and Benchmark-Based Improvement Techniques. *Surv. Geophys.* **2013**, *34*, 541−651.
(31) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(32) Harrell, E. *Magnus Carlsen: The 19-Year-Old King of Chess*. https://time.com/archive/6948922/magnus-carlsen-the-19-year-old-king-of-chess/ (accessed March 2026).
(33) Nakayama, N.*Memories of Kitani*Treasure Chest Enigma; Slate Shell198416−19
(34) Ruecker, G.; Ruecker, C. Counts of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683−695.
(35) Johnson, W. S.; et al. Total Synthesis of Epiandrosterone. *J. Am. Chem. Soc.* **1953**, *75*, 2275−2276.
(36) Zhang, L.; Wang, R.; Wang, C.; Liu, B.; Yang, Y.; Zhang, Z.; Huang, J.; Yang, Z. Concise Synthesis of 7-Deoxypsammaplysins K and O and 7-Deoxyceratinamide A by 1,3-Dipole Cycloaddition. *Org. Lett.* **2022**, *24*, 3786−3791.
(37) Park, K. H. K.; Frank, N.; Duarte, F.; Anderson, E. A. Collective synthesis of illudalane sesquiterpenes via cascade inverse electron demand (4 + 2) cycloadditions of thiophene S,S-dioxides. *J. Am. Chem. Soc.* **2022**, *144*, 10017−10024.
(38) Gajewska, E. P.; Szymkuć, S.; Dittwald, P.; Startek, M.; Popik, O.; Mlynarski, J.; Grzybowski, B. A. Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses. *Chem* **2020**, *6*, 280−293.
(39) Mikulak-Klucznik, B.; Klucznik, T.; Beker, W.; Moskal, M.; Grzybowski, B. A. Catalyst: Curtailing the Scalable Supply of Fentanyl by Using Chemical AI. *Chem* **2024**, *10*, 1319−1326.
(40) Corey, E. J.; Long, A. K.; Rubenstein, S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408−418.
(41) Zhou, J.; Gao, B.; Xu, Z.; Ye, T. Total Synthesis and Stereochemical Assignment of Callyspongiolide. *J. Am. Chem. Soc.* **2016**, *138*, 6948−6951.
(42) Jeon, S.; Lee, J.; Park, S.; Han, S. Total Synthesis of Dimeric *Securinega* Alkaloids (−)-Flueggenines D and I. *Chem. Sci.* **2020**, *11*, 10934−10938.
(43) Coley, C. W.; Thomas, D.; Lummiss, A.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **2019**, *365*, No. eaax1566.
(44) Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J. AiZynthFinder: a fast, robust and flexible

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

open-source software for retrosynthetic planning. *J. Cheminf.* **2020**, *12*, 70.
(45) Shields, J. D.; Howells, R.; Lamont, G.; Leilei, Y.; Madin, A.; Reimann, C. E.; Rezaei, H.; Reuillon, T.; Smith, B.; Thomson, C.; Zheng, Y.; Ziegler, R. E. AiZynth impact on medicinal chemistry practice at AstraZeneca. *RSC Med. Chem.* **2024**, *15*, 1085−1095.
(46) *Reaxys Predictive Retrosynthesis*. https://www.elsevier.com/products/reaxys/predictive-retrosynthesis.
(47) CAS *SciFinder Synthesis Planning*. https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder/synthesis-planning.
(48) Molga, K.; Dittwald, P.; Grzybowski, B. A. Navigating around Patented Routes by Preserving Specific Motifs along Computer-Planned Retrosynthetic Pathways. *Chem* **2019**, *5* (2), 460−473.
(49) Strieth-Kalthoff, F.; Szymkuć, S.; Molga, K.; Aspuru-Guzik, A.; Glorius, F.; Grzybowski, B. A. Artificial Intelligence for Retrosynthetic Planning Needs Both Data and Expert Knowledge. *J. Am. Chem. Soc.* **2024**, *146*, 11005−11017.
(50) Roh, J.; Joung, J. F.; Yu, K.; Tu, Z.; Bartholomew, G. L.; Santiago-Reyes, O. A.; Fong, M. H.; Sarpong, R.; Reisman, S. E.; Coley, C. W. *ACS Cent. Sci.* **2026**, *12*, 345−357.
(51) Zhao, P.-C.; Zhou, H.-Z.; Wang, Q.; Wu, Z.-Y.; Yu, H.; Shi, J.-Y. A Comprehensive Survey of AI-Based Retrosynthesis Planning: Datasets, Models, and Tools. *Innovation Informatics* **2026**, *2*, 100026.
(52) Zhang, C.; Lin, Q.; Zhu, B.; Yang, H.; Lian, X.; Deng, H.; Zheng, J.; Liao, K. SynAsk: Unleashing the Power of Large Language Models in Organic Synthesis. *Chem. Sci.* **2024**, *16*, 43−56.
(53) Xuan-Vu, N.; Armstrong, D.; Wehrbach, M.; Bran, A. M.; Jončev, Z.; Schwaller, P. Synthelite: Chemist-Aligned and Feasibility-Aware Synthesis Planning with LLMs. *arXiv [cs.AI]* **2025**.
(54) *The FAIR Data Principles - FORCE11*. https://force11.org/info/the-fair-data-principles (accessed March 2026).
(55) *ACS research data policy*. https://researcher-resources.acs.org/publish/data_policy (accessed March 2026).

Downloaded from pubs.acs.org/jacsat/article-pdf/148/36/38740/6816217/2/jacs.6c08434.pdf by ULSAN NATL INST SCIENCE AND TECHLGY user on 28 September 2026

https://doi.org/10.1021/jacs.6c08434
J. Am. Chem. Soc. 2026, 148, 38740−38753

Article

Robot-assisted mapping of chemical reaction hyperspaces and networks

https://doi.org/10.1038/s41586-025-09490-1

Received: 19 March 2025

Accepted: 4 August 2025

Published online: 24 September 2025

Open access

Check for updates

Yankai Jia[1,2,3], Rafał Frydrych[1,3], Yaroslav I. Sobolev[1], Wai-Shing Wong[1], Bibek Prajapati[1], Daniel Matuszczyk[1], Yasemin Bilgi[1], Louis Gadina[1], Juan Carlos Ahumada[1], Galymzhan Moldagulov[1,2], Namhun Kim[1], Eric S. Larsen[1], Maxence Deschamps[1], Yanqiu Jiang[1]✉ & Bartosz A. Grzybowski[1,2]✉

Despite decades of investigation, it remains unclear (and hard to predict[1–4]) how the outcomes of chemical reactions change over multidimensional 'hyperspaces' defined by reaction conditions[5]. Whereas human chemists can explore only a limited subset of these manifolds, automated platforms[6–12] can generate thousands of reactions in parallel. Yet, purification and yield quantification remain bottlenecks, constrained by time-consuming and resource-intensive analytical techniques. As a result, our understanding of reaction hyperspaces remains fragmentary[7,9,13–16]. Are yield distributions smooth or corrugated? Do they conceal mechanistically new reactions? Can major products vary across different regions? Here, to address these questions, we developed a low-cost robotic platform using primarily optical detection to quantify yields of products and by-products at unprecedented throughput and minimal cost per condition. Scanning hyperspaces across thousands of conditions, we find and prove mathematically that, for continuous variables (concentrations, temperatures), individual yield distributions are generally slow-varying. At the same time, we uncover hyperspace regions of unexpected reactivity as well as switchovers between major products. Moreover, by systematically surveying substrate proportions, we reconstruct underlying reaction networks and expose hidden intermediates and products—even in reactions studied for well over a century. This hyperspace-scanning approach provides a versatile and scalable framework for reaction optimization and discovery. Crucially, it can help identify conditions under which complex mixtures can be driven cleanly towards different major products, thereby expanding synthetic diversity while reducing chemical input requirements.

Owing to rapid advances in chemical automation, there has been a growing number of high-throughput campaigns to discover new modes of reactivity[6,7,12,16,17], maximize reaction yields (often under the guidance of various optimization algorithms[9,15,18]) and produce standardized data sought urgently[19] by chemical artificial intelligence. Our goals here were different, as we wished to reconstruct complete portraits of chemical reactions over multidimensional parameter spaces (concentrations, temperatures), with the quantification of yields not only for the major products but also for as many by-products as could be identified. Anticipating thousands of crude mixtures to analyse (here >9,000 in total), we wished to minimize the use of techniques such as nuclear magnetic resonance (NMR) or liquid chromatography–mass spectrometry (LC-MS), which, at their maximum throughput of only a few samples per hour, become costly and require prolonged or even dedicated access to equipment, which may be prohibitive for most academic groups. Accordingly, we relied heavily on inexpensive (cents per sample) and rapid (about 100 samples per hour) ultraviolet–visible (UV-Vis) detection augmented with algorithms to decompose complex spectra (that is, not just those featuring distinct and easy-to-interpret spectral features[20,21]) and with autocorrelation metrics to quantify missing information and detect anomalous outcomes.

Robotic set-up and analysis of reaction outcomes

Unless otherwise stated, experiments were performed on a robotic platform (Fig. 1a and Supplementary Video 1) house-built to support various organic solvents and harsh reagents and allowing us to execute and characterize up to roughly 1,000 reactions per day (for design blueprints and computer codes, see Methods and Supplementary Information Sections 1 and 2).

For a given reaction under study, this robot examines the hyperspace of conditions at points of some N-dimensional grid (for example, uniform in Fig. 1b), setting up reactions and acquiring UV-Vis absorption spectra at each point and at desired time(s). The acquisition of an individual spectrum takes, on average, roughly 8 s, whereas the accompanying pipetting and spectrometer washing–drying take about 50 s in total

[1]Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan, Republic of Korea. [2]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea. [3]These authors contributed equally: Yankai Jia, Rafał Frydrych. ✉e-mail: jyqinception@gmail.com; nanogrzybowski@gmail.com

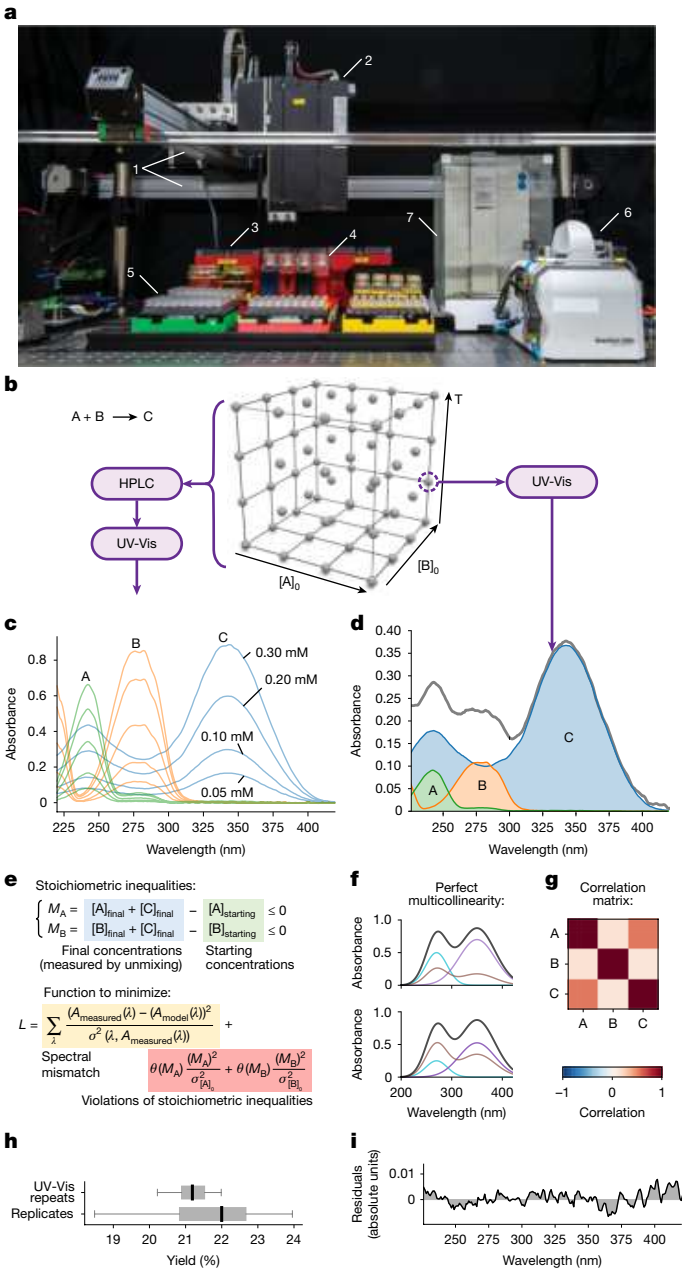


Fig. 1 | Automated reaction platform and optical yield determination. **a**, Parts of the house-built, inexpensive (about $25K) system: (1) horizontal gantry; (2) liquid-handling module; (3) racks of pipette tips; (4) stock solutions; (5) 54-vial plates; (6) UV-Vis spectrophotometer; (7) balance for weighing of solutions (see Supplementary Videos 1 and 2). **b**, Hypothetical A + B → C reaction space defined by initial concentrations $[A]_0$ and $[B]_0$ and temperature T. **c**, UV-Vis spectra of the HPLC-isolated species are measured at different concentrations (here for Claisen–Schmidt condensation from Fig. 2a). **d**, At each point of the conditions' space, the UV-Vis spectrum (grey) is decomposed—here nearly perfectly—into the component spectra from **c**, quantifying the concentrations/yields of these components in the crude mixture. **e**, Stoichiometry of reaction places constraints on possible concentrations (blue) of components in the crude mixture (see Methods). **f**, Example of a perfect linear dependence ('multicollinearity'), in which the brown spectrum in the top subplot happens to be a linear sum of the cyan and purple ones (scaled by, respectively, 0.50 and 0.25). Consequently, unmixing is an ill-conditioned problem, as the spectrum of the entire mixture (black) can be decomposed into infinitely many combinations of the brown, cyan and purple components (1:1:1 in the top and 0.5:0.75:2 in the bottom subplots). **g**, Correlation matrix for components A, B and C of the Claisen–Schmidt condensation from **d**. Low off-diagonal elements indicate no strong multicollinearity. **h**, Statistical distributions of yields with respect to the limiting reagent. Relative standard deviations are 2% between repeated optical detection performed on the same crude mixture ($n = 54$) and 5% between experimental repeats ($n = 27$) of the entire workflow for the same conditions (data from Fig. 5). **i**, Differences between experimental and modelled spectra (from **d**) as a function of wavelength. The fit is adequate, as this trace closely resembles uncorrelated ('white') noise.

(Supplementary Video 2). In a subsequent (and perhaps counterintuitive) step, the crudes from all hyperspace points are combined (Fig. 1b, left). The resultant mixture is separated by chromatography and the isolated fractions are identified by traditional spectroscopic (NMR, MS) analyses. This 'bulk' analysis is to identify the 'basis set' of reaction products that form in any appreciable quantities anywhere in the hyperspace. The UV-Vis absorption spectra of these purified products (as well as all substrates, solvents and reagents used) at different concentrations are taken to construct concentration–absorbance calibration curves (Fig. 1c). Then the crude and often complex UV-Vis spectra acquired at each point of the hyperspace (Fig. 1b, right) are fitted by linear combinations of the reference spectra of the basis-set components (Fig. 1d) using the so-called vector decomposition techniques (also known as spectral unmixing[22]). Overall, this protocol requires conventional high-performance liquid chromatography (HPLC)/NMR/MS analysis of only one complex mixture combining all hyperspace samples (Fig. 1c). It returns on this time investment manifolds by rapidly estimating the yields and mixture compositions at thousands of individual data points within the hyperspace of the reaction.

This general scheme merits two comments. First, it is important to reject potential unmixing solutions violating reaction stoichiometry (Fig. 1e) and to ensure stability of the fit. The latter benefits from the absorption spectra of components being of similar magnitudes and not linearly dependent (Fig. 1f), which is diagnosed by looking at off-diagonal elements of the correlation matrix between the concentrations of the components (Fig. 1g). Satisfying these conditions (1) is increasingly challenging as the number of components increases but (2) is helped by extending the spectral range as much as possible into the UV, at which the absorption bands of organic species are usually more numerous and narrower (see the example in Fig. 5b). With these precautions, yield estimates are within 5% (for example, a 20% yield would have a spread of 19–21%), with optical measurement and spectral unmixing contributing 2% and the remaining 3% caused by uncertainty of pipetting, fluctuations of temperature and residual evaporation (Fig. 1h and Supplementary Information Sections 2.10 and 2.11).

Second, the fitting procedure is accompanied by an algorithm to detect anomalous outcomes in some regions of the hyperspace. This is done by tracking differences between experimental and fitted spectra, calculating the variance of residuals (Fig. 1i) and evaluating the autocorrelation (across wavelengths, not time) of these residuals by the Durbin–Watson statistic. The mismatch is considered notable if the root mean square residual exceeds 0.01 absorbance units or if the Durbin–Watson statistic at 30 nm 'lag' deviates from the value of 2 (which corresponds to the absence of autocorrelation) more strongly than the respective statistic for the baseline of a given spectrophotometer. If systematic deviation is detected, it then signifies formation of a product unexpected based on the initial scan of the hyperspace (see Fig. 3d,e). Further mathematical details of spectral unmixing and anomaly detection are discussed in Supplementary Information Section 3.

Reaction scope

We confirmed that this method is applicable to a range of reactions used widely in both academic groups and industry, including various couplings, condensations, cycloadditions and rearrangements (reactions **1**–**8** on the green background in Extended Data Fig. 1a), as

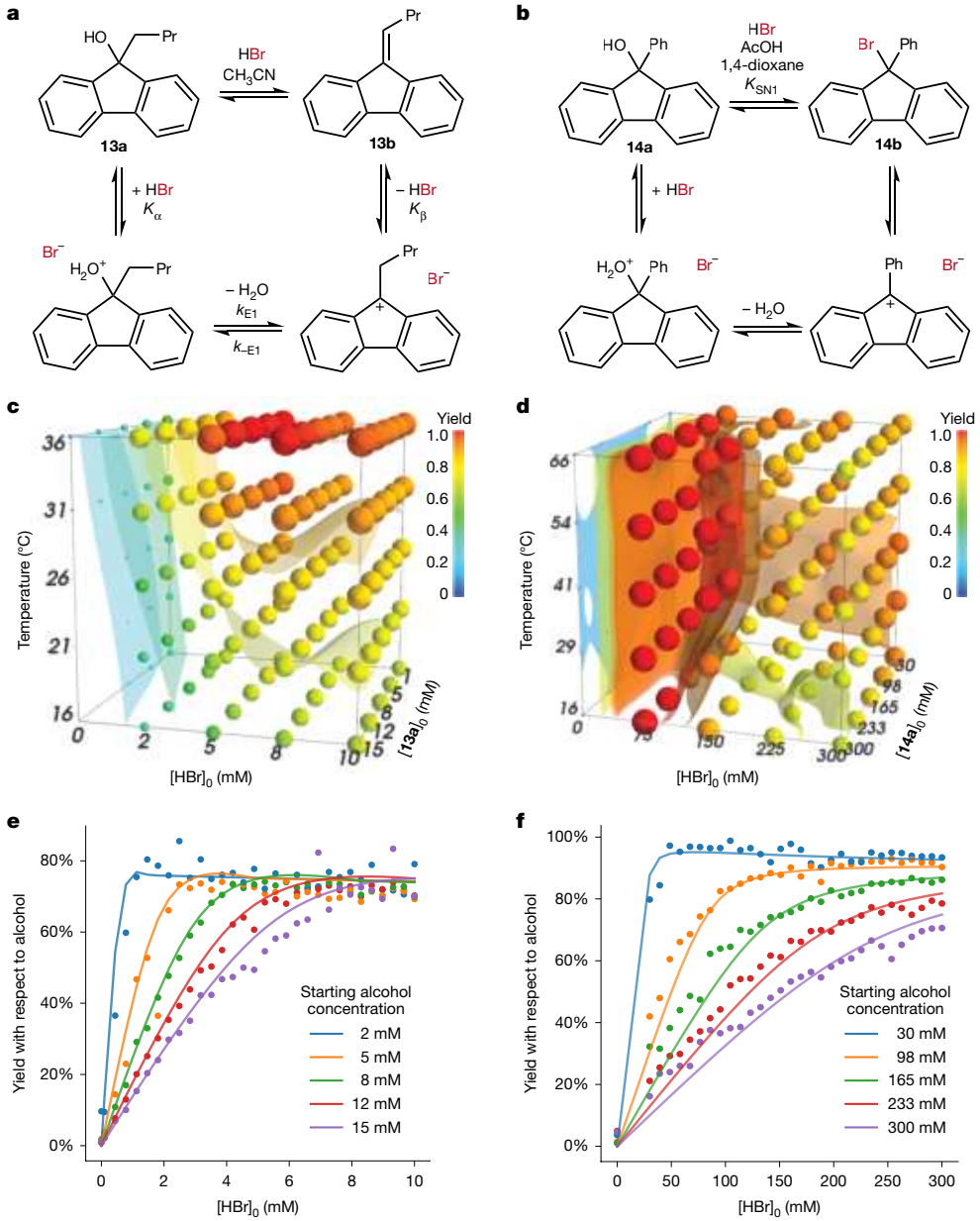


Fig. 2 | Yield distributions over the reaction spaces of E1 and S_N1 reactions. **a**,**b**, Schemes of E1 (**a**) and S_N1 (**b**) reactions and their mechanisms. For the former, we use 9-butyl-9*H*-fluoren-9-ol substrate **13a** and for the latter, to prevent elimination, 9-phenyl-9*H*-fluoren-9-ol **14a**. For E1, the parameters of the mechanism are the temperature dependences of equilibrium constant K_α for dynamic equilibrium between **13a** and its protonated form, of equilibrium constant K_β for dynamic equilibrium between **13b** and its protonated form, of pK_a of HBr in acetonitrile and the Eyring–Polanyi equation parameters for kinetic rate constants k_{E1} and k_{-E1} for water molecule elimination from the protonated **13a** (Supplementary Information Section 5.2). For S_N1 at a given temperature, the only parameter is the effective equilibrium constant K_{SN1} = [**14b**][H_2O]/([**14a**][HBr]) (Supplementary Information Section 5.1). E1 used acetonitrile as solvent. For S_N1, the solvent was dioxane with admixture of AcOH (to homogenize the biphasic dioxane–water mixture formed in situ and complicating optical detection). **c**,**d**, Corresponding yield distributions of products **13b** and **14b**. For clarity, only 125 conditions per cube are shown. Marker colours and sizes are proportional to yields. Yield colour scales are the same for both cubes. Surfaces of equal yield are at 20%, 40%, 70% and 85%. **e**,**f**, Kinetics model (curves) fitted to experimental yields with respect to alcohol (circles) for E1 and S_N1 reactions. To better visualize the quality of fits, only two-dimensional snapshots of the entire dataset are shown, here T = 21 °C for E1 and T = 36 °C for S_N1 (for similar plots at all temperatures, see Supplementary Information Sections 5.1 and 5.2). Horizontal axis has the initial concentrations of HBr, with the colour corresponding to different initial concentrations of the alcohol indicated in the keys.

well as substitutions, eliminations and multicomponent reactions (MCRs), discussed in detail in Figs. 2–6. Expectedly, the approach is not suitable for, for example, reactions of aliphatic scaffolds whose products give no signal in the UV-Vis range above 220 nm and some reactions are borderline in terms of fit quality, owing to the obstruction of the product signal by reactant/solvent peaks (for example, **9**–**12** in the yellow and red portions of Extended Data Fig. 1a). Notably, yields quantified by the robotized approach correlate strongly, R^2 = 0.96, with the yields of the same reactions performed, purified and analysed ex roboto (Extended Data Fig. 1b).

Reaction hyperspaces

With these capabilities, we began to explore and analyse hyperspaces of several classic reaction types differing in mechanistic complexity. In doing so, we focused on yield distributions at reaction time of several

hours (that is, the situation relevant to synthetic organic practice) and on hyperspace regions featuring maximal degree of variability (as opposed to regions in which, for example, concentrations are so high that yields are all 'saturated'; see Supplementary Fig. 121). For clarity, in Figs. 2–6, we plot only the subsets of data points, with all raw data available at Zenodo (https://doi.org/10.5281/zenodo.14880579).

Hyperspaces of basic reactions with no anomalies

Starting simple, we considered the spaces of the E1 elimination and S_N1 substitution[23,24] using, respectively, substrates **13a** and **14a**. The space of E1 was examined at $t = 4$ h for 775 conditions and that of S_N1 at $t = 48$ h for 930 conditions. The three-dimensional yield distributions shown in Fig. 2c,d for reaction products **13b** and **14b** are approximately concave, steadily increasing towards the global maximum (in bimolecular S_N1, there is shoulder maximum in the region in which the concentration of the limiting reagent **14a** is low; see also Extended Data Fig. 2). Mathematically, the 'steepness' of these surfaces can be quantified by the slopes of the product concentration with respect to the initial concentration of substrate(s), $D_{ij} = \partial C_i/\partial C_{0,j}$—as shown in Supplementary Fig. 166, the absolute values of these slopes, $|D_{ij}|$ are small, at or below unity. Analysis of residuals detects no anomalies and, under all conditions, only substrates and the product are present in various proportions (for example, E1 is not accompanied by S_N1 side reaction). Of note, yield data contained in the 'cubes' allow for fitting approximate kinetic models (by time-integrating the underlying kinetic equations) and for deriving reasonable values of some kinetic and thermodynamic parameters, such as ΔH, $\Delta H^{\ddagger}$ or ΔS (see caption to Fig. 2, Methods and Supplementary Information Sections 5.1 and 5.2 for derivations). Even though the data correspond to only one time point, the abundance of yield values for various substrate concentrations imposes hundreds of constraints against which a candidate model must simultaneously fit, thus limiting the acceptable parameters of the model. In the 'A four-dimensional hyperspace' section, we will use this approach to interrogate more complex mechanisms.

A hyperspace with anomalous outcomes

Next, we revisited the S_N1 reaction but with **15a** used as a substrate (Fig. 3a). The main product **15b** forms through the rearrangement of the carbocation **15c**, and its yield distribution over the 1,085 conditions fits closely to the expected first-order kinetics (see caption to Fig. 3b). The algorithm also detects a minor anthraquinone by-product **15d** congruent with the Diels–Alder reaction with singlet-state oxygen[25,26] and distributed as illustrated in Fig. 3c. Most notably, analysis of residuals (Fig. 1i) provides the first example of a systematic anomaly and an unexpected outcome in a narrow region of low HBr concentrations marked in pink in Fig. 3d. This species exhibits an intensely pink colour that gradually wanes over two days (and disappears rapidly during purification attempts). Its UV-Vis spectrum (Fig. 3e) does not agree with absorption spectra of **15c** or structurally similar carbocations[27–29], but extensive analyses based on MS as well as time-dependent density functional theory (TD-DFT) calculations (Supplementary Information Sections 4.13 and 6.2) support assignment as a carbocation dimer **15e** resulting from a reaction between **15c** and **15a** (with enough HBr to create **15c**, but with the amount of water introduced by this HBr insufficient to quench the intermediate into **15b**). To our best knowledge, dimerization of a substrate with its own derived carbocationic intermediate has only been observed under superacidic conditions at temperatures down to −50 °C (ref. 30) but never in the presence of quenching nucleophiles (here H_2O and Br^-) and under ambient conditions. We emphasize, however, that even with the presence of this anomaly, the hyperspace remains simple—that is, each of its constituent species features only one yield maximum—and the yield distributions of all species are slowly varying, $|D_{ij}| \approx 1$ (Supplementary Information Fig. 166), including **15e**, which forms over a narrow range but in only approximately 1 nM concentration.

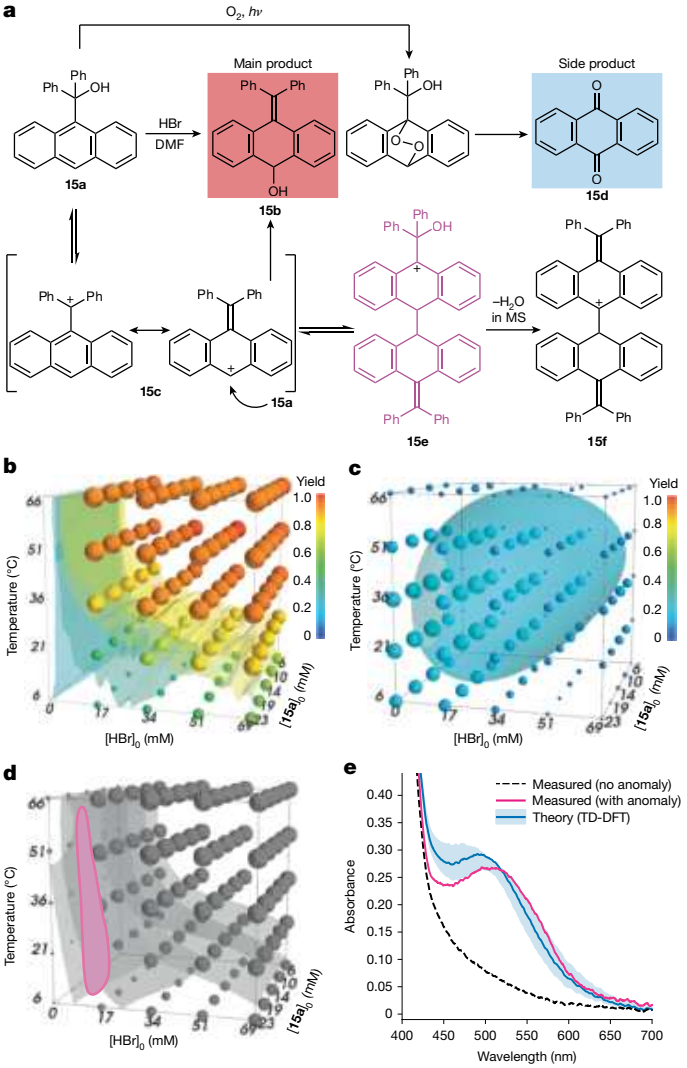


Fig. 3 | S_N1 space with an anomalous outcome. a, Scheme of a reaction between alcohol **15a** and HBr leading to major product **15b** through anthranyl carbocation **15c**. The minor anthraquinone product **15d** originates from the cycloaddition of substrate **15a** with singlet oxygen to give the endoperoxide, followed by subsequent fragmentations (for details, see Supplementary Fig. 32). The 'unexpected' carbocation dimer structure is assigned to **15e** (for details of structural assignment, see Supplementary Information Sections 4.13 and 6.2). **b**, Space of reaction yields of the main product **15b** under different conditions (reaction time: 24 h; for clarity, only 125 out of 1,085 points are shown; for all data, see Supplementary Information Section 4.15). Fitting the kinetic model to these experimental data reveals that yields of product **15b** scale as $1 - \exp\left(-tk_1[\mathrm{HBr}]\exp\left(-\frac{E_a}{k_BT}\right)\right)$, in which t is reaction time, $k_1\exp\left(-\frac{E_a}{k_BT}\right)$ is the kinetic rate constant at a given temperature (prefactor k_1 is a free parameter), E_a is the activation energy free parameter and [HBr] is the concentration of HBr—that is, these conversions increase with [HBr] and with temperature, do not depend on the initial concentration of the alcohol substrate and the rate-limiting step is formation of a carbocation **15c**, all as expected for S_N1 reaction. The isosurfaces correspond to 20%, 40%, 70% and 85% yield. **c**, Distribution of yields for the side product **15d**. The isosurface corresponds to 20% yield. **d**, Pink cloud delineates a narrow region of conditions in which an anomalous outcome—that is, carbocation dimer **15e**—is observed. **e**, Pink line is a typical UV-Vis spectrum from the region of conditions in which an anomalous outcome is observed. The dashed line is a typical experimental spectrum for which this anomaly is absent. Blue line is the ab initio theoretical absorption spectrum of 1.23 nM of **15e** after summation with the dashed 'baseline', blue band is the 1σ confidence interval (see Supplementary Information Section 6.2).

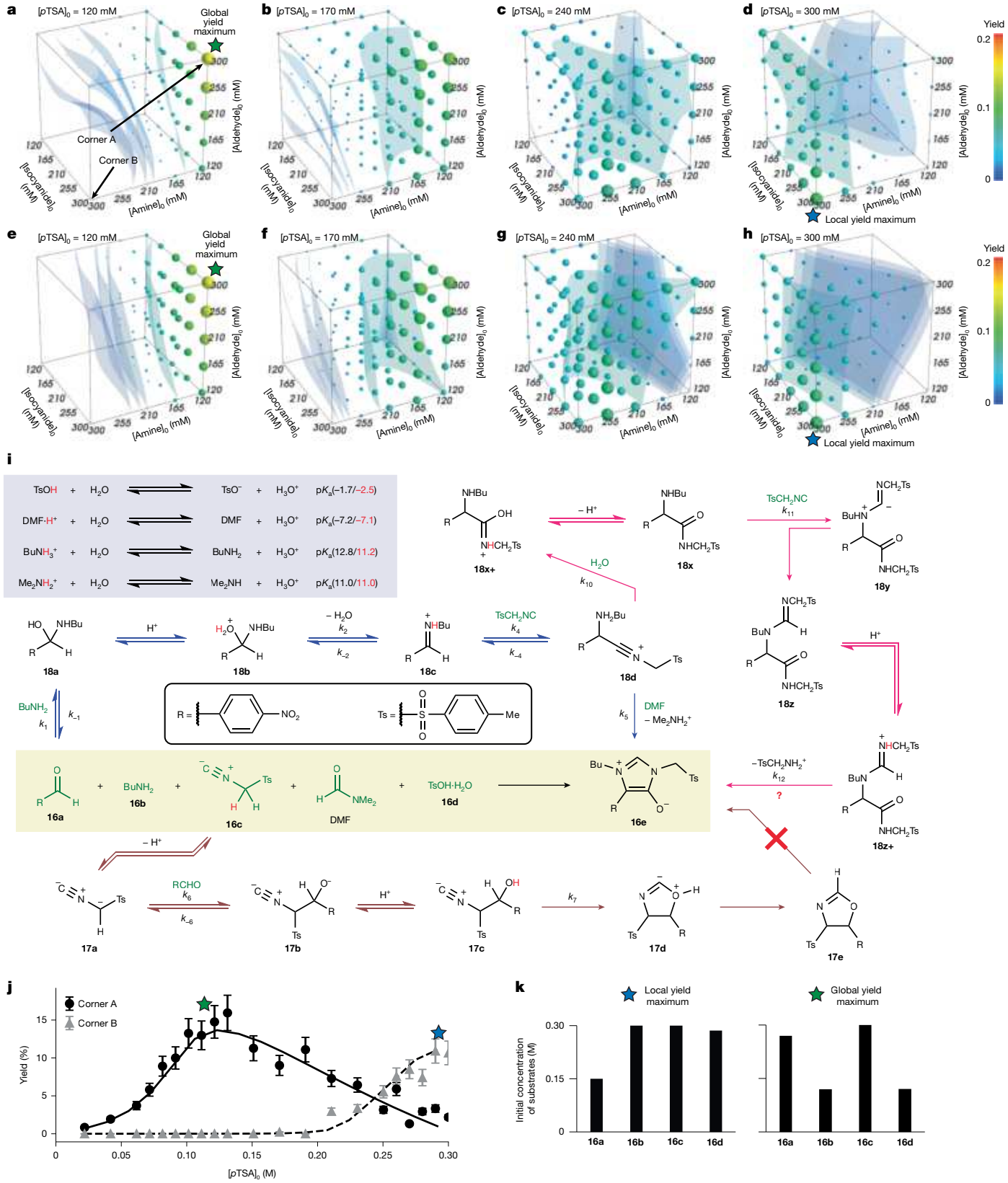


Fig. 4 | See next page for caption.

A four-dimensional hyperspace

In search for more topologically complex hyperspaces, we turned to reactions that involve more than two substrates and are based on much more complex mechanisms. As the first case, we interrogated a four-component Ugi-type reaction from ref. 31, illustrated in Fig. 4 and chosen because of distinct UV-Vis signal of the cyclized product **16e**. For this cyclization reaction, 3,234 conditions were investigated

Fig. 4 | Experimental and modelled hyperspace of a Ugi-type, four-component reaction. a–**d**, Snapshots of the reaction's four-dimensional hyperspace. The axes of each plot are initial amine, aldehyde and isocyanide concentrations for initial concentrations of *p*TSA indicated above each plot. Experimental yields (after 16 h at 26 °C) of product **16e** are colour-coded according to the legend in **d**. Only 125 interpolated conditions per cube are shown for clarity (see Supplementary Information Section 4.15 and Supplementary Video 3). Green and blue stars mark two distinct yield maxima. **e**–**h**, Corresponding yield distributions of **16e** predicted by a kinetic model fitted simultaneously to all experimental data (yield isosurfaces are at 0.5%, 1.0%, 1.5% and 5.0%). This model is based on a kinetic network outlined in **i** and further side reactions (see Supplementary Information Scheme 5). The original reaction of **16a**, **16b**, **16c** and DMF using initiator **16d** and leading to product **16e** is on the yellow background. The classic Ugi mechanism is summarized by blue arrows (for details, see Supplementary Information Scheme 6). A competing oxazoline-based path (brown) was experimentally excluded (red cross). A third mechanism (magenta), potentially active at high $[p\text{TSA}]_0$ and involving aldehyde, amine and isocyanide in a 1:1:2 ratio, did not improve model fit. The theoretical model (Supplementary Tables 19, 20, 22 and 23) comprised this entire mechanistic network of 12 reactions and 15 proton transfer steps. Best-fitted model followed the data closely while retaining pK_a values (in red font) of most species within 1–2 units from literature values (black font) used as initial guess (see Supplementary Information Section 5.3). **j**, Experimental and modelled yield profiles at the two maxima (corners A and B in **a**) along $[p\text{TSA}]_0$ confirm that yields at global/local maxima are higher than at any four-dimensional paths connecting them (see Supplementary Video 3 and Supplementary Fig. 161). Error bars correspond to 14.6% relative standard error. **k**, Initial concentrations at global and local maxima.

within a four-dimensional space defined by different initial concentrations of 4-nitrobenzaldehyde **16a**, *n*-butylamine **16b** and *p*-tosylmethyl isocyanide **16c** substrates (DMF substrate was also a solvent of constant concentration), as well as *p*-toluenesulfonic acid monohydrate, *p*TSA **16d**, acting as a reaction initiator.

The hyperspace illustrated in Fig. 4a–d and Supplementary Video 3 now reveals the presence of two distinct yield maxima for the heterocyclic product **16e**—a global maximum marked by a green star and a local maximum marked by a blue star. Further sampling of the conditions' grid in the region between the maxima confirms that they are separated in four-dimensional space by a region of low yield (Fig. 4j and Supplementary Figs. 160c and 161).

Notably, MS analyses (Supplementary Information Section 4.8) evidence distinct signals at these two maxima: the local maximum gives a peak *m*/*z* ratio corresponding to the iminium ion **18c** expected for the classical Ugi mechanism[32,33], whereas the global maximum also features a peak attributable to the oxazoline **17e**. This, in turn, may suggest an extra mechanism (brown arrows) through the oxazoline **17e** intermediate—alternative mechanisms for the Ugi reaction have, indeed, been postulated for decades but never proved[34,35]. However, such a possibility is directly disqualified by separate, ex roboto experiments in which isolated **17e** fails to give product **16e** on further reaction with *n*-butylamine in the presence of *p*TSA (Supplementary Information Section 4.8.4). Although other mechanisms can also be considered (see magenta arrows in Fig. 4 and corresponding caption), fitting the experimental hyperspace data to the kinetic network of 12 reactions and 15 proton transfer steps (Fig. 4i and theoretical details in Supplementary Information Section 5.3) suggests that an excellent agreement can be achieved (Fig. 4e–h) based only on the Ugi mechanism—that is, inclusion of extra mechanisms offers no perceptible improvement. Instead, the emergence of the two maxima can be attributed to the shifts in the equilibria underlying the network (Supplementary Information Sections 5.3.4 and 5.3.5). We also note that evolution of these maxima following changes in *p*TSA concentration is very gradual (Supplementary Video 3) and the yield distributions are characterized by very small 'slopes', $|D_{ij}| \lesssim 1$ (Supplementary Fig. 166).

A hyperspace supporting a switchable reaction network

Next, we considered the classic Hantzsch pyridine synthesis (Fig. 5), studied for almost 150 years (refs. 36–38) and interesting for several known intermediates and competing pathways (coloured in blue in Fig. 5c). Notably, examining the four-dimensional hyperspace (concentrations of the three substrates, 26 °C and 80 °C temperatures; a total of 2,582 conditions) revealed the presence of many more components than previously thought. Their isolation required several HPLC repurification cycles performed in a closed-loop manner. Specifically, after each cycle, the newly identified substances were added to the spectral unmixing algorithm and the global fits to experimental spectra from the entire hypercube gradually improved (Fig. 5a,b); meanwhile, more fractions were selected for the next round of purification if they featured yet-unassigned signals (Supplementary Information Sections 4.9 and 4.10). After eight such fitting–purification cycles, the mismatch between the fitted and experimental spectra reached instrumental noise (dashed green line in Fig. 5a), at which point our knowledge of the hyperspace composition can, for all practical reasons, be deemed nearly complete. This knowledge spans not only the seven known species[36–38] but also nine new ones (coloured red in Fig. 5c) that have not been reported in the classical Hantzsch reaction but can be of interest in the context of biological activity[39].

By considering the causal relationships between these and other species discovered within the hyperspace, it is possible to establish the synthetic connectivity of the isolated species (Supplementary Information Section 4.10). This analysis, ultimately, reconstructs the complete network of the Hantzsch reaction detailed in Fig. 5c with key steps confirmed by further ex roboto reactions (marked by green arrows). Notably, analysis of yield distributions (with the help of HPLC, as spectral unmixing of all 16 components is no longer unique) revealed that, by adjusting substrate concentrations at 80 °C, the network can be switched between three different major products (**19d**, **19e**, **19k**), each forming in >60% yield and with maxima located at different corners of the conditions' cube (Fig. 5d–f). One of these switchovers is between the Hantzsch ester (**19d**) and the Petrenko-Kritschenko product (**19e**)[40], meaning that even the so-called named reactions can, in reality, be part of the same hyperspace. Another observation is that, despite the switchovers, the individual yield distributions for individual species remain smooth, as visualized by the isosurfaces in Fig. 5d–f.

Five-dimensional compositional space

Finally, we consider a hyperspace in which dimensions correspond to systematic variations in composition. Here these compositions are the contents of different metals in the Prussian blue analogues (PBAs)[41] (Fig. 6), which are perovskite-type materials described by a general formula $KM_B[M_A(CN)_6]$ and widely studied in the context of catalysis and energy storage[42]. The robot surveys a five-dimensional space defined by the contents of metals at sites M_A (two types, Fe and Co) and M_B (five types, Mn, Fe, Co, Ni and Cu) over a uniform grid with the granularity of 0.2 (for example, $Mn_{0.2}Ni_{0.4}Cu_{0.4}$–$Fe_{0.6}Co_{0.4}$ PBA versus $Mn_{0.2}Ni_{0.2}Cu_{0.6}$–$Fe_{0.4}Co_{0.6}$ PBA, with molar fractions at A and B sites each summing to unity). Each of the 756 PBAs is prepared in situ by straightforward co-precipitation method (Fig. 6a,b) and is used to catalyse reaction of styrene **20a** and *t*-BuOOH to give styrene oxide **20c** (given that styrene oxide is the key intermediate to synthesize many fine chemicals and pharmaceuticals[43], this reaction has been the subject of several studies using various catalysts, including PBAs[44,45]). Figure 6c shows the distribution of reaction yields and Fig. 6d quantifies the selectivities with respect to benzaldehyde **20b**, the main by-product of the reaction. As seen, the yield hyperspace is much more corrugated than concentration–temperature hyperspaces from Figs. 2–5 and features

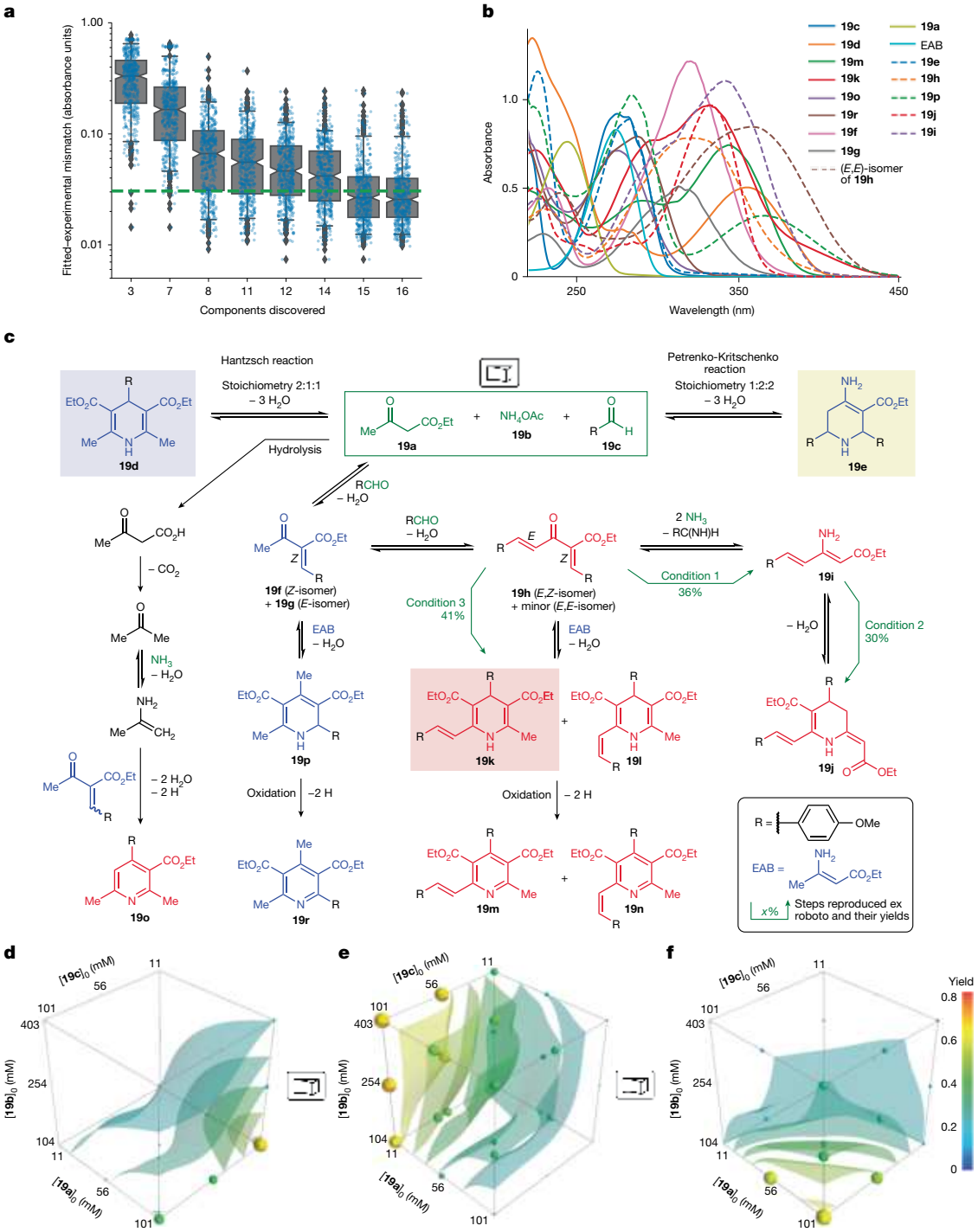


Fig. 5 | 'Switchable' hyperspace of the Hantzsch reaction network. **a**, Identification of reaction products by closed-loop repurification and refitting against four-dimensional hypercube UV-Vis spectra (see Supplementary Information Section 4.15). Within eight repurification rounds, 16 components (14 products plus unreacted **19a** and **19c**) were identified. The mismatch between fitted and experimental spectra (vertical axis) decreased to the median value of instrumental uncertainty (dashed green line; see Supplementary Information Section 4.9). The vertical axis shows the highest spectral mismatch per condition. Box plot elements indicate the first quartile, median and third quartile, whiskers show the 5th and 95th percentiles, with outliers as black diamonds. Blue scatter plot shows individual data points. **b**, UV-Vis 'basis set' of the 16 components. Their discovery follows the cycles in **a**: 3 components (**19a**, **19c**, **19d**), 7 components (+**19k**, **19r**, **19o**, **19m**), 8 components (+EAB), 11 components (+**19j**, **19f**, **19g**), 12 components (+**19e**), 14 components (+**19h**, **19p**), 15 components (+**19i**), 16 components (+**19h**). **c**, Reaction network: green, substrates; blue, isolated known products/intermediates; red, new products/intermediates. Network connectivity was verified by separate reactions (green arrows): (1) NH_4OAc, AcOH, 80 °C, 18 h; (2) ethyl acetoacetate, EtOH, 80 °C, 18 h; (3) EAB, EtOH, 80 °C, 24 h. Note the presence of products of two separate 'named reactions', Hantzsch (**19d**) and Petrenko-Kritschenko (**19e**) (mechanisms are shown in Supplementary Information Schemes 2 and 3. **d**–**f**, Within the hyperspace, the reaction network at 80 °C can be switched between three main products observed in different corners of the space: Hantzsch ester **19d** (maximum yield 63.5%, <3% **19p**) (**d**); Petrenko-Kritschenko product **19e** (maximum yield about 66%) (**e**); benzylidene-extended Hantzsch derivative **19k** (maximum yield 61.7%) (**f**). These yields (after 48 h) were quantified by HPLC and are colour-coded according to the legend. Isosurfaces correspond to 20%, 30%, 40%, 50% and 60% yield. For full experimental details, see Supplementary Information Sections 4.7, 4.9 and 4.10.

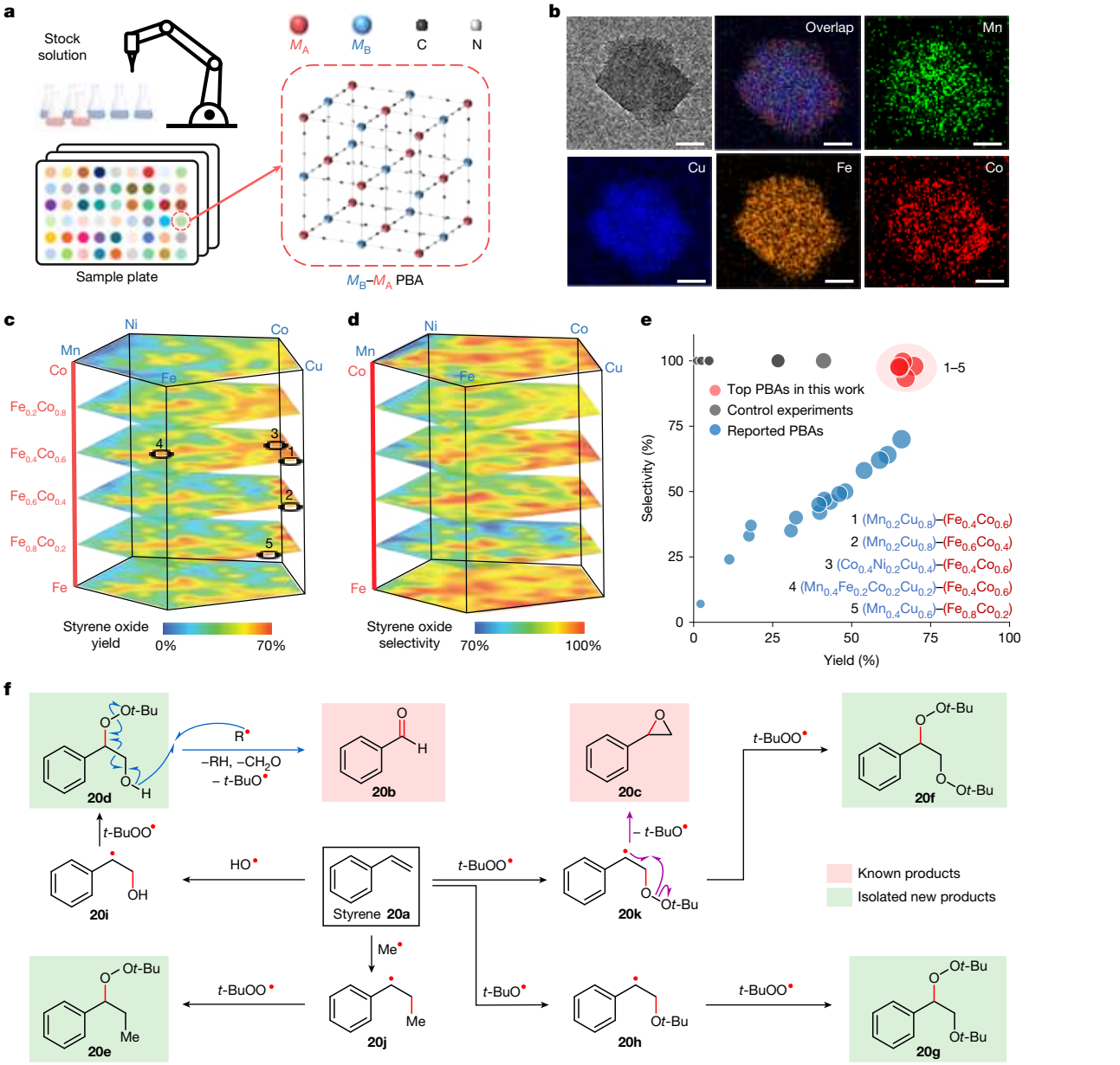


Fig. 6 | Five-dimensional space of catalyst compositions. a, 756 PBAs, catalysts with a general formula $KM_B[M_A(CN)_6]$, were prepared by the robot in separate vials using the co-precipitation method. Up to two aqueous solutions of M_A and up to five of M_B were pipetted automatically with specified volumes to prepare the PBA catalyst suspensions, each with a final volume of 200 µl in individual vials, followed by 24 h of ageing at room temperature. **b**, Irrespective of the composition, these nanomaterials were particles approximately 50 nm in size. Uniformity of metal distributions was confirmed by energy-dispersive spectroscopy. For large-area scanning electron microscopy images and all experimental details, see Supplementary Information Section 8. Scale bars, 20 nm. **c**, Five-dimensional hyperspace of yields of styrene oxide reconstructed by spectral unmixing. **d**, Five-dimensional hyperspace of selectivity of styrene oxide reconstructed by spectral unmixing. **e**, At least five PBAs (red markers 1–5 here and black circles in the corresponding hyperspace in **c**) showed yield–selectivity properties better than previously studied PBAs (blue markers), as well as some other controls (grey markers: Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, $Fe(CN)_6^{3+}$, $Co(CN)_6^{3-}$ and the last one without any catalyst; see raw data in Supplementary Table 34). **f**, The robotically reconstructed hyperspace allowed for the identification of four products unreported in previous PBA studies of this reaction (**20d**, **20e**, **20f** and **20g** shown on green backgrounds) and for the reconstruction of the reaction's mechanistic network involving four intermediates **20h**, **20i**, **20j** and **20k**. The blue fishhook arrows on species **20d** indicate the electron flow to furnish aldehyde **20b** (R = *t*-BuOO, *t*-BuO or HO), whereas the purple fishhook arrows on species **20k** indicate the electron flow to furnish epoxide **20c** (ref. 56). For the generation of radicals from *tert*-butyl hydroperoxide, see Supplementary Information Section 8.7.

several local yield maxima. Notably, it contains several PBA compositions that offer better yield–selectivity characteristics than previously reported PBAs[44,45] or other controls (Fig. 6e). As in other examples that we considered, the hyperspace reconstruction reveals the presence of unreported intermediates and by-products and helps reconstruct a mechanistic network (Fig. 6f) that had previously remained elusive[44–46].

Discussion

Hyperspace structure

One of the insights from these studies is the relatively simple structure of the concentration–temperature spaces, with yield distributions of any single species featuring, at most, two yield maxima and with the slope values $|D_{ij}|$ at or below unity (Supplementary Fig. 166).

As described in Supplementary Information Sections 7.1 and 7.2, it can be rigorously proved that, for strongly connected, directed hypergraphs[47,48] (that is, those that cannot be split into non-interacting subsets of reactions) and for first-order and pseudo-first-order reactions, $|D_{ij}| \leq 1$. For higher-order kinetics of individual reactions, the theoretical upper bound on $|D_{ij}|$ is too high to be practically relevant, but numerical studies of networks comprising first-order and second-order steps and no cycles confirm that $|D_{ij}|$ is also typically on the order of unity. A corollary to this result is that, along concentration coordinates $C_{0,j}$, $j = 1,\ldots,k$, the hypersurfaces can be examined faithfully at sparse intervals (see Supplementary Information Section 7.6). This, in turn, can help reduce the number of experiments during conditions' screening and optimization campaigns. For more discussion of hyperspace properties (in terms of topology, differential yields and Shannon information theory), see Supplementary Figs. 161 and 162 and Supplementary Information Section 4.12.

Naturally, these considerations do not apply to hyperspaces in which compositions are varied. For instance, different catalysts (as in Fig. 6) can lower the activation barriers substantially and to different extents. As a result, they can alter the network-wide kinetics more than changes over a limited range of concentrations or temperatures, resulting in 'steeper' concentration derivatives and in the hyperspace featuring many local yield maxima.

Network reconstruction

The method by which we reconstructed reaction networks is loosely analogous to the approaches used in electronics to reverse-engineer a circuit inside a 'black box' by applying inputs and analysing corresponding outputs. Here the inputs are the various substrate concentrations and temperatures and the outputs are the identities of the products found by the global analysis of the hyperspace. With the set of products identified, the 'wiring' of the network is prescribed by the general rules of chemical reactivity. In the cases studied here, the wirings could be reconstructed relatively readily by humans but, as illustrated in Supplementary Information Section 4.11, this process can also be aided by network algorithms operating either at the level of mechanistic steps[4,49] or full reactions[50], thereby linking the hyperspace of conditions to the 'space of reaction grammars'[5], a link argued to be key to chemical innovation[5]. In this effort, the broad investigation of the hyperspace is essential, as different inputs can serve to 'activate' or 'deactivate' different branches of the network (by modifying reaction rates and shifting equilibria), enabling the formation of as many products as possible. This also means investigating the reaction space at stoichiometries that a human chemist might not necessarily find intuitive or relevant to a particular reaction. Here this approach more than doubled the knowledge of the Hantzsch hyperspace and we have seen similar discovery enhancements in other reactions that we have studied since then (for example, Pechmann, Biginelli). From a technical point of view, our ability to interrogate entire spaces of conditions capitalizes on the use of inexpensive (cents per sample) optical detection, which reduces the need for costly (about $45–300 per sample) HPLC/quantitative NMR analyses by a factor of several hundred (for example, one HPLC/NMR cycle per 1,085 UV-Vis experiments to reconstruct the S_N1 network in Fig. 3; eight closed-loop cycles to ensure complete knowledge of the Hantzsch space examined by UV-Vis at 2,582 conditions). As long as the equilibrium is not fully reached (Supplementary Figs. 154 and 160), sweeping the reaction rates through control of starting concentrations accesses the type of data similar to conventional kinetic experiments (that 'sweep' the time to observation instead) and estimation of kinetic parameters becomes possible.

Network control

A related aspect is the practical ability to direct these reaction networks towards desired outcomes by simply adjusting concentrations (that is, without using different reagents, as done in some recent excellent papers, such as ref. 51). For simple reactions, such stoichiometric control is, of course, well known (for example, we can influence the ratio of monosubstituted to disubstituted product of alkyl dibromide by using one versus two equivalents of the nucleophile). However, for multicomponent reaction mixtures, it is less obvious that they can be pushed cleanly towards different major products. We have seen this capability realized for the Hantzsch reaction, in which the yields of major products **19d**, **19e** and **19k** were maximized at concentration ratios very different from those we may expect based on product stoichiometry alone. For instance, product **19e** incorporates one copy of substrate **19a**, two copies of **19b** and two copies of **19c**, but in the hyperspace in Fig. 5e, the regions of appreciable yield, >40%, start from about 1:4.5:2 ratio (and continue for even higher molar excess of **19b** at the yield maximum, which, however, becomes wasteful in terms of substrate use; see Extended Data Fig. 2 and Supplementary Information Section 4.17).

Conclusions

In conclusion, this work is an initial effort—made possible by modern reaction automation—to understand the structure of reaction hyperspaces, one of the five foundational spaces of mathematical chemistry[5]. It reinforces the view of chemical reactions as networks[4,51,52] embedded in multidimensional spaces of conditions and, at least in some cases, switchable between different major products, which is reminiscent of some biochemical networks[53,54] and promising in terms of diversity-oriented synthesis[55]. The rapid and cost-effective approach to hyperspace reconstruction can (1) systematize and accelerate reaction discovery and optimization and (2) foster fundamental research on reaction networks as dynamic systems. This effort can be aided by hyperspace visualization and analysis tools such as those we developed and make available at Zenodo (https://doi.org/10.5281/zenodo.14880579; see also Supplementary Video 4). In a broader context, inclusion of our experimental yield maps to the benchmarks used for testing the general-purpose yield optimization algorithms[9,11,15] would appreciably widen the diversity of these benchmarks, given the current scarcity of such datasets[7,9,13,15]. In future work, we aim to extend our robotic platform to reactions that require solid dispensing and/or strictly oxygen-free conditions (this will broaden the scope of hyperspaces we can analyse), to accelerate dispensing and measuring operations (to analyse fast reactions) and also to monitor hyperspace evolution over time, which will be necessary to reconstruct highly nonlinear mechanisms (for example, oscillating reactions).

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09490-1.

1. Skoraczyński, G. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **7**, 3582 (2017).
2. Saebi, M. et al. On the use of real-world datasets for reaction yield prediction. *Chem. Sci.* **14**, 4997–5005 (2023).
3. 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).
4. Szymkuć, S., Wołos, A., Roszak, R. & Grzybowski, B. A. Estimation of multicomponent reactions' yields from networks of mechanistic steps. *Nat. Commun.* **15**, 10286 (2024).
5. Restrepo, G. Spaces of mathematical chemistry. *Theory Biosci.* **143**, 237–251 (2024).
6. Granda, J. M., Donina, L., Dragone, V., Long, D. L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
7. Lin, S. et al. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236 (2018).
8. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).

9. Angello, N. H. et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399–405 (2022).
10. Rohrbach, S. et al. Digitization and validation of a chemical synthesis literature database in the ChemPU. *Science* **377**, 172–180 (2022).
11. Slattery, A. et al. Automated self-optimization, intensification, and scale-up of photocatalysis in flow. *Science* **383**, eadj1817 (2024).
12. Dai, T. et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **635**, 890–897 (2024).
13. Buitrago Santanilla, A. et al. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015).
14. Davies, I. W. The digitization of organic synthesis. *Nature* **570**, 175–181 (2019).
15. Shields, B. J. et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
16. Wilbraham, L., Mehr, S. H. M. & Cronin, L. Digitizing chemistry using the chemical processing unit: from synthesis to discovery. *Acc. Chem. Res.* **54**, 253–262 (2021).
17. Mahjour, B. et al. Rapid planning and analysis of high-throughput experiment arrays for reaction discovery. *Nat. Commun.* **14**, 3924 (2023).
18. Wang, J. Y. et al. Identifying general reaction conditions by bandit optimization. *Nature* **626**, 1025–1033 (2024).
19. Strieth-Kalthoff, F. et al. Artificial intelligence for retrosynthetic planning needs both data and expert knowledge. *J. Am. Chem. Soc.* **146**, 11005–11017 (2024).
20. Stadler, E. et al. A versatile method for the determination of photochemical quantum yields via online UV-Vis spectroscopy. *Photochem. Photobiol. Sci.* **17**, 660–669 (2018).
21. Lu, J.-M. et al. Roboticized AI-assisted microfluidic photocatalytic synthesis and screening up to 10,000 reactions per day. *Nat. Commun.* **15**, 8826 (2024).
22. Bioucas-Dias, J. M. et al. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**, 354–379 (2012).
23. Banert, K. & Kurnianto, A. Nucleophile substitution bei 4,4-dimethyl-2-adamantyl-substraten: rückseitenangriff bei 2-adamantan-derivaten. *Chem. Ber.* **119**, 3826–3841 (1986).
24. Thibblin, A. & Sidhu, H. Mechanisms of competing solvolytic elimination and substitution reactions. The role of ion-pair intermediates in aqueous solvents. *J. Chem. Soc., Perkin Trans. 2* **2**, 1423–1428 (1994).
25. Clennan, E. L. Aromatic endoperoxides. *Photochem. Photobiol.* **99**, 204–220 (2023).
26. Klaper, M., Wessig, P. & Linker, T. Base catalysed decomposition of anthracene endoperoxide. *Chem. Commun.* **52**, 1210–1213 (2016).
27. Ammer, J., Sailer, C. F., Riedle, E. & Mayr, H. Photolytic generation of benzhydryl cations and radicals from quaternary phosphonium salts: how highly reactive carbocations survive their first nanoseconds. *J. Am. Chem. Soc.* **134**, 11481–11494 (2012).
28. Nishimae, Y., Kurata, H. & Oda, M. Arylbis(9-anthryl)methyl cations: highly crowded, near infrared light absorbing hydrocarbon cations. *Angew. Chem. Int. Ed.* **43**, 4947–4950 (2004).
29. Nojima, M., Takagi, M., Morinaga, M., Nagao, G. & Tokura, N. Reaction of some triarylmethyl radicals, polyarylalkenes, and 9,10-dihydro-9,10-epidioxyanthracenes with sulphur dioxide; detection of radicals and/or cations derived from the corresponding cation radicals. *J. Chem. Soc. Perkin Trans. 1* **5**, 488–495 (1978).
30. Hollenstein, S. & Laali, K. K. Efficient conversion of 9-isopropenylphenanthrene to 4,6,6-trimethyl-6*H*-benz[*de*]anthracene in FSO3H; 5,6-dihydro-4*H*-benzanthracen-4-ium ion and its charge delocalization mode. *Chem. Commun.* 2145–2146 (1997).
31. Cankařová, N., Nemec, I. & Krchňák, V. *p*-TSA-mediated four-component reaction: one-step access to mesoionic 1*H*-imidazol-3-ium-4-olates, direct NHC precursors. *Adv. Synth. Catal.* **364**, 2996–3003 (2022).
32. 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).
33. Rocha, R. O., Rodrigues, M. O. & Neto, B. A. D. Review on the Ugi multicomponent reaction mechanism and the use of fluorescent derivatives as functional chromophores. *ACS Omega* **5**, 972–979 (2020).
34. Alvim, H. G. O., da Silva Júnior, E. N. & Neto, B. A. D. What do we know about multicomponent reactions? Mechanisms and trends for the Biginelli, Hantzsch, Mannich, Passerini and Ugi MCRs. *RSC Adv.* **4**, 54282–54299 (2014).
35. Chéron, N., Ramozzi, R., Kaïm, L. E., Grimaud, L. & Fleurat-Lessard, P. Challenging 50 years of established views on Ugi reaction: a theoretical approach. *J. Org. Chem.* **77**, 1361–1366 (2012).
36. Hantzsch, A. Condensationsprodukte aus Aldehydammoniak und ketonartigen Verbindungen. *Ber. Dtsch. Chem. Ges.* **14**, 1637–1638 (1881).
37. Shen, L. et al. A revisit to the Hantzsch reaction: unexpected products beyond 1,4-dihydropyridines. *Green Chem.* **11**, 1414–1420 (2009).
38. Santos, V. G. et al. The multicomponent Hantzsch reaction: comprehensive mass spectrometry monitoring using charge-tagged reagents. *Chem. Eur. J.* **20**, 12808–12816 (2014).
39. Chang, C.-C. et al. Antagonism of 4-substituted 1,4-dihydropyridine-3,5-dicarboxylates toward voltage-dependent L-type Ca^{2+} channels $Ca_V1.3$ and $Ca_V1.2$. *Bioorg. Med. Chem.* **18**, 3147–3158 (2010).
40. Petrenko-Kritschenko, P. Über die kondensation des acetondicarbonsäureesters mit aldehyden, ammoniak und aminen. *J. Prakt. Chem.* **85**, 1–37 (1912).
41. Singh, B. & Indra, A. Prussian blue- and Prussian blue analogue-derived materials: progress and prospects for electrochemical energy conversion. *Mater. Today Energy* **16**, 100404 (2020).
42. Li, W. et al. Chemical properties, structural properties, and energy storage applications of Prussian blue analogues. *Small* **15**, 1900470 (2019).
43. Choo, J. P. S. & Li, Z. Styrene oxide isomerase catalyzed Meinwald rearrangement reaction: discovery and application in single-step and one-pot cascade reactions. *Org. Process Res. Dev.* **26**, 1960–1970 (2022).
44. Guo, S. et al. Synthesis of trimetallic Prussian blue analogues and catalytic application for the epoxidation of styrene. *Ind. Eng. Chem. Res.* **59**, 13831–13840 (2020).
45. Liang, Y. et al. Prussian blue analogues as heterogeneous catalysts for epoxidation of styrene. *RSC Adv.* **5**, 17993–17999 (2015).
46. Zhang, L., Zhang, Z., He, X., Zhang, F. & Zhang, Z. Regulation of the products of styrene oxidation. *Chem. Eng. Res. Des.* **120**, 171–178 (2017).
47. Pal, A. et al. Finding thermodynamically favorable pathways in chemical reaction networks using flows in hypergraphs and mixed-integer linear programming. *J. Chem. Inf. Model.* **65**, 6772–6787 (2025).
48. Grzybowski, B. A., Bishop, K. J. M., Kowalczyk, B. & Wilmer, C. E. The 'wired' universe of organic chemistry. *Nat. Chem.* **1**, 31–36 (2009).
49. Krzeszewski, M. et al. Computer-generated, mechanistic networks assist in assigning the outcomes of complex multicomponent reactions. *J. Am. Chem. Soc.* **147**, 15636–15644 (2025).
50. Mikulak-Klucznik, B., Klucznik, T., Beker, W., Moskal, M. & Grzybowski, B. A. Catalyst: curtailing the scalable supply of fentanyl by using chemical AI. *Chem* **10**, 1319–1326 (2024).
51. Mahjour, B., Shen, Y., Liu, W. & Cernak, T. A map of the amine–carboxylic acid coupling system. *Nature* **580**, 71–75 (2020).
52. Baltussen, M. G., de Jong, T. J., Duez, Q., Robinson, W. E. & Huck, W. T. S. Chemical reservoir computation in a self-organizing reaction network. *Nature* **631**, 549–555 (2024).
53. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588 (2006).
54. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **497**, 619–623 (2013).
55. Wołos, A. et al. Computer-designed repurposing of chemical wastes into drugs. *Nature* **604**, 668–676 (2022).
56. Halder, J. et al. Insight of solvent effect on CeO_2 catalyzed oxidation of styrene with tert-butyl hydroperoxide: a combined experimental and theoretical approach. *Catal. Commun.* **164**, 106413 (2022).

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) 2025

Methods

Automation platform

Reagent addition was carried out using a commercial pipetting module (ZEUS, Hamilton), optimized for pipetting a range of organic solvents in aliquots ranging from 10 µl to 1 ml, with volumetric errors of less than 1% for volumes greater than 50 µl and less than 2% for volumes between 10 and 50 µl. Liquids were dispensed into 2-ml glass vials arranged in custom-designed 54-well plates, with each reaction typically set to a final volume of 500 µl. Following reagent addition, vials were hermetically sealed using a flat lid, with a rubber sheet and perfluoroalkoxy film placed between the lid and the vial. Initial mixing was performed on an orbital shaker at 250 rpm for 5 min. Subsequent stirring is omitted, as separate studies (see Supplementary Information Sections 2.9 and 4.5) have shown that, for vials of these dimensions, passive mixing over the course of the reaction (hours) is as efficient as mechanical stirring, thereby eliminating the need for cumbersome stir bars.

After reactions were run for desired times, the crude mixture from each vial was, if necessary, automatically diluted to align with the detection range of the UV-Vis spectrophotometer (NanoDrop, Thermo Fisher Scientific). The operation of the spectrophotometer was also automated, including lid opening and closing, along with the flushing and drying of the measuring pedestal (see Supplementary Video 2). All of these operations were orchestrated by house-written software. The entire assembly was placed inside a hood and, if needed, constantly purged with nitrogen. The manual operations in the entire workflow were changing the 54-well plates, covering them with rubber/perfluoroalkoxy sheets and placing on the orbital shaker. For some hyperspaces, two such systems were used. Further technical details and the blueprints for system replication are provided in Supplementary Information Section 2 and at Zenodo (https://doi.org/10.5281/zenodo.14880579)[57]. The software design is illustrated in Supplementary Fig. 2 and the source code is deposited at https://github.com/yaroslavsobolev/robowski-maps.

The full cubes of 775 E1 conditions (five temperatures, 16–36 °C; five initial substrate concentrations, 1.5–15.0 mM; 31 HBr concentrations, 0.1–10.0 mM; reaction time: 4 h) and 930 S_N1 conditions (six temperatures; 16–66 °C; five 9-butyl-9*H*-fluoren-9-ol substrate concentrations, 0.03–0.30 M; 31 HBr concentrations; reaction time: 48 h) are provided in Supplementary Information Section 4.15, with all raw data in the Zenodo repository.

Stoichiometry constraints

Stoichiometry constraints, as shown in Fig. 1e, were included into the spectral unmixing algorithm by modifying the log-likelihood function L to be minimized: L is a sum of squared spectral mismatch (yellow) of experimental $A_{\mathrm{measured}}(\lambda)$ and modelled $A_{\mathrm{model}}(\lambda)$ absorbances, divided by the instrumental variance of the spectrophotometer $\sigma^2(\lambda, A_{\mathrm{measured}}(\lambda))$, and the term representing the violation of stoichiometric inequalities (red) containing the Heaviside function $\theta(x)$ and the experimental uncertainties $\sigma_{[A]_0}$ and $\sigma_{[B]_0}$ of preparing the reaction mixture with given starting concentrations. See Supplementary Information Section 3 for more details about the spectral unmixing algorithm

Isosurface extraction

The isosurfaces in Figs. 2–5 were extracted by the marching cubes algorithm[58] operating on a 50 × 50 × 50 regular grid obtained through the radial basis functions interpolator applied to the raw data. This method was implemented in the HyperspaceViewer software, which is provided in the Supplementary Information and is open-sourced at a GitLab repository (https://gitlab.com/az-steak/hyperspace_viewer).

Kinetic fitting

The kinetic fitting in Fig. 2e,f was done by numerically integrating the equations of the mechanism and optimizing the parameters to achieve the best-fit of the model's predicted product yields simultaneously to all experimental data. This procedure gave reasonable estimates of the reactions' thermodynamic parameters: $\Delta H = -30.7 \pm 1.4$ kJ mol^{-1} for S_N1 and $\Delta H = 81.98^{+3.52}_{-1.64}$ kJ mol^{-1}, $\Delta S = 273.8^{+10.8}_{-4.9}$ J mol^{-1} K^{-1} and $\Delta H^{\ddagger} = 21.10^{+0.33}_{-0.49}$ kJ mol^{-1} (transition state enthalpy) for E1. For derivations of these values and comparisons against related literature examples, see Supplementary Information Sections 5.1 and 5.2.

Initial guess of kinetic parameters (Fig. 4e–h) was provided by the Markov chain Monte Carlo algorithm[59], kinetic equations at each condition were integrated numerically and the discrepancy against the experimental yield distribution (over the hyperspace) was minimized using the trust region reflective algorithm[60]. See Supplementary Information Section 5.3 for further details of the kinetic model.

PBA-catalysed styrene epoxidation

PBAs were synthesized following a reported co-precipitation method[61]. Two stock solutions containing metal M_A (Fe, Co) were prepared by dissolving $K_3[Fe(CN)_6]$ or $K_3[Co(CN)_6]$ in deionized water to form clear solutions of type A (0.10 M). Five stock solutions containing metal M_B (Mn, Fe, Co, Ni, Cu) were prepared separately by dissolving corresponding metal nitrates/chlorides ($Mn(NO_3)_2$, $FeCl_2$, $Co(NO_3)_2$, $Ni(NO_3)_2$, $Cu(NO_3)_2$, each at 0.10 M) in the presence of trisodium citrate dihydrate ($Na_3C_6H_5O_7{\cdot}2H_2O$, 0.1125 M) in deionized water to form clear solutions of type B. Afterwards, mixing of solutions of type A and type B were performed in specified volume ratios using the automated liquid-handling system according to the target composition. For instance, for a PBA with composition $Mn_{0.2}Ni_{0.4}Cu_{0.4}$–$Fe_{0.6}Co_{0.4}$, the sequence and volume of stock solution addition were as follows: 20 µl $Mn(NO_3)_2$, 40 µl $Ni(NO_3)_2$, 40 µl $Cu(NO_3)_2$, 60 µl $K_3[Fe(CN)_6]$ and 40 µl $K_3[Co(CN)_6]$. The total volume of solutions of type A was 100 µl and the total volume of solutions of type B was also 100 µl—therefore, the total volume of each PBA solution was always 200 µl. After sealing, each 54-vial well plate housing different mixed solutions was shaken for 15 min (250 rpm) and aged at room temperature for 24 h to form PBAs.

For a typical set of styrene epoxidation reactions, the reaction stock solution was prepared as follows: first, styrene (0.50 M) and *tert*-butyl hydrogen peroxide (0.75 M) were dissolved in 100 ml acetonitrile. Then, 30 ml cetyltrimethylammonium bromide aqueous solution (0.25 M) was added to the mixture. Subsequently, the obtained mixture was sonicated for 5 min to obtain a homogeneous stock solution. Afterwards, the robotic system pipetted 1.3-ml aliquots of the thus prepared stock solution into the vials containing previously prepared PBAs (200 µl aqueous suspension). The reaction vials were then sealed and every reaction plate was placed on a thermal shaker (72 °C, 700 rpm) for 6 h. After reaction, when the PBAs sedimented, the robot acquired 40 µl of the supernatant from every vial, diluted it 1,000 times and then measured the UV-Vis spectrum with the NanoDrop spectrophotometer. The yield of styrene oxide (y_{so}) and benzaldehyde (y_{alde}) for each sample can be obtained by spectral unmixing and the selectivity of styrene, calculated by $y_{so}/(y_{so} + y_{alde}) \times 100\%$, was then obtained.

Data availability

All data in support of the findings of this study are included in the Supplementary Information; furthermore, raw spectra and their fitted results for all automated reactions (9,362 data points in total) are deposited at Zenodo (https://doi.org/10.5281/zenodo.14880579)[57]. Deposition numbers 2418839 (**15b**) and 2418838 (**15b**·C_3H_7NO) containing the crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre at www.ccdc.cam.ac.uk/data_request/cif.

Code availability

All codes and pipelines used for this study are provided at GitHub (https://github.com/yaroslavsobolev/robowski-maps). HyperspaceViewer

software for the visualization and analysis of hyperspaces is deposited at GitLab (https://gitlab.com/az-steak/hyperspace_viewer) and Zenodo (https://doi.org/10.5281/zenodo.14880579)[57], along with the source code and user manual, and as the HyperspaceViewer.zip Supplementary File. Furthermore, We release a Google Colab notebook (https://colab.research.google.com/drive/1uXaoWjkHYaPoCt8N6q4GzYY3e92T_nC-?usp=sharing) with the step-by-step tutorial of the core functionality of calibration and spectral unmixing for the Claisen–Schmidt condensation case shown in Fig. 1c,d.

57. Jia, Y. et al. Code and raw data for 'Robot-assisted mapping of chemical reaction hyperspaces and networks'. *Zenodo* https://doi.org/10.5281/zenodo.14880579 (2025).
58. Schroeder, W., Martin, K. & Lorensen, B. *The Visualization Toolkit: An Object-oriented Approach to 3D Graphics* 4th edn (Kitware, 2006).
59. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **5**, 65–80 (2010).
60. Branch, M. A., Coleman, T. F. & Li, Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM J. Sci. Comput.* **21**, 1–23 (1999).
61. Du, M. et al. High-entropy Prussian blue analogues and their oxide family as sulfur hosts for lithium-sulfur batteries. *Angew. Chem. Int. Ed.* **61**, e202209350 (2022).

Acknowledgements We thank A. Shevlyakov for helpful discussions. We thank Fan Guo for assistance with the EDS analysis. This work was generously supported by the taxpayers of South Korea through the Institute for Basic Science, project code IBS-R020-D1.

Author contributions Y. Jia and Y.I.S. constructed the automation hardware, designed its software, conducted experiments with R.F. and B.P., performed data analysis and wrote the initial draft with R.F. R.F. contributed to automation design, led the chemical part of the project, selected, performed and interpreted all reactions in roboto and performed MS and MS/MS mechanistic studies. Y.I.S. designed software for spectral unmixing and numerical models of reaction kinetics (with input from R.F. and W.-S.W.), fitted kinetics to the experimental data, performed TD-DFT calculations of absorption spectra, quantum thermochemistry calculations and theoretical study of the smoothness of yield maps. B.P. performed all Hantzsch reactions with the robotic platform and partial HPLC purifications of the Hantzsch reaction and performed one reaction for versatility test. L.G. proposed the reaction, made the proof-of-concept test and determined products structure and reaction mechanism for PBA catalysts. J.C.A. worked on synthesis, products isolation, structure determination and mechanism analysis for PBA-catalysed reactions. J.C.A. also performed and analysed the Suzuki–Miyaura coupling experiments. Y. Jiang worked on PBA synthesis, reactions on robotic platform, data analysis and wrote the initial draft for the PBA part. M.D. developed the HyperspaceViewer software. L.G. wrote the manual for the HyperspaceViewer software. W.-S.W. worked on four reactions for the versatility test, assigned structures to unexpected Hantzsch reaction products and proposed the corresponding mechanisms. Y.B. optimized S_N1 and E1 reactions and their adaptation to the robotic platform. G.M. and others conducted a reaction versatility test. G.M. and Y.B. conducted the system accuracy test. D.M. contributed to robotic synthesis (Hantzsch reaction), separation of crude mixtures, purification, NMR analysis of isolated compounds and performed one reaction for the versatility test. N.K. analysed single-crystal X-ray diffraction data. E.S.L. conducted synthesis and aided with the analysis and determination for the rearrangement reaction of anthracen-9-yldiphenylmethanol and analysed several sets of single-crystal X-ray diffraction data to determine their three-dimensional structures. B.A.G. conceived and supervised the project.

Competing interests The authors declare no competing interests.

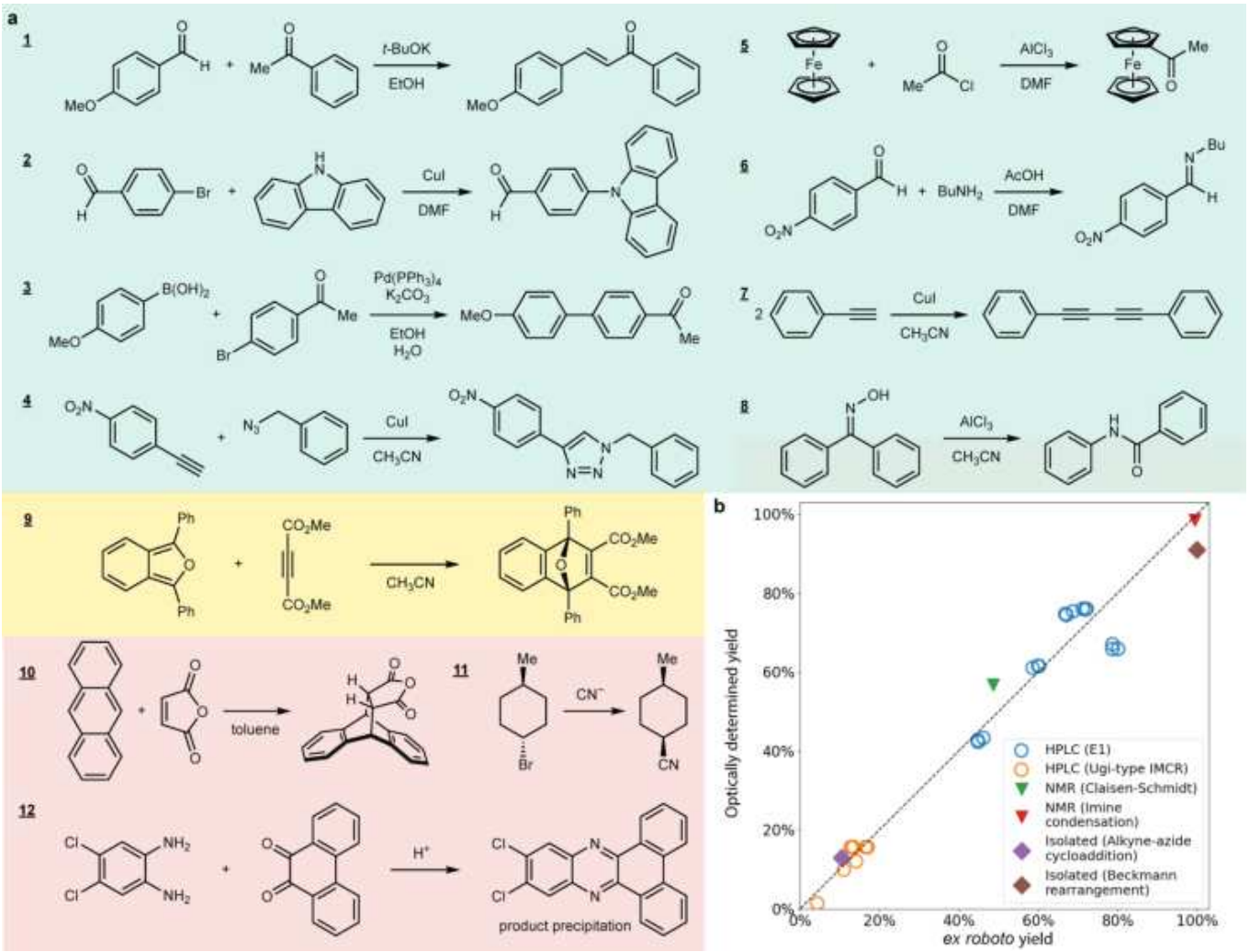
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09490-1.

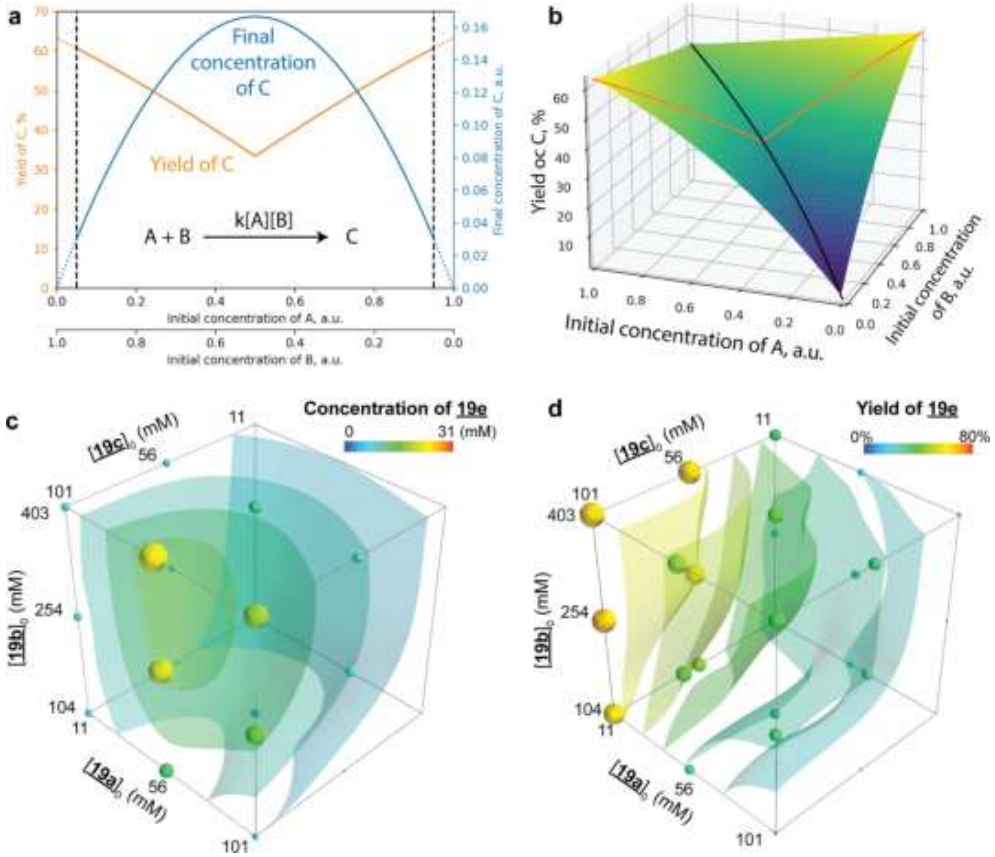
Correspondence and requests for materials should be addressed to Yanqiu Jiang or Bartosz A. Grzybowski.

Peer review information *Nature* thanks Guillermo Restrepo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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



Extended Data Fig. 1 | Partial scope and quality of optical yield estimates. **a**, Examples of reactions compatible (green background), borderline (yellow) and incompatible (pink) with yield determination using UV-Vis and spectral unmixing. The compatible reactions include but are not limited to Claisen–Schmidt condensation **1**, Ullmann-type coupling **2**, Suzuki–Miyaura cross-coupling **3**, Cu(I)-catalysed alkyne-azide cycloaddition **4**, Friedel–Crafts acylation **5**, imine condensation **6**, Glaser-type coupling **7** or Beckmann rearrangement **8**. The Diels–Alder reaction **9** shown on yellow background is borderline in terms of the quality of unmixing (the overlap of the UV-Vis signals from the product and from dimethyl acetylenedicarboxylate prevents accurate fitting, leading to an overestimation of the yield). For the Diels–Alder reaction **10** shown on red background, the band of the product (λ_{max} < 286 nm) is obscured by the toluene solvent (λ_{max} = 286 nm). Other reactions unsuitable for the method are, for example, S_N2 reaction of alkyl bromide **11** (not absorbing above 220 nm) or condensation of dichlorodiaminobenzene with phenanthrenequinone **12** (poor solubility of the product). **b**, R^2 = 0.96 correlation between the optically determined yields versus the yields of the same reactions determined by conventional means (HPLC for E1 from Fig. 3 and Ugi-type MCR from Fig. 4; ^{1}H NMR for Claisen–Schmidt condensation **1** and imine condensation **6**; the weights of isolated compounds for Cu(I)-catalysed alkyne-azide cycloaddition **4** and Beckmann rearrangement **8**). For numerical values of each data point, see Supplementary Tables 5–7 in Supplementary Information Sections 4.3 and 4.4.



Extended Data Fig. 2 | Concentrations versus yields near domain boundaries and the need to sample domain interiors. a, Taking as an example a basic bimolecular reaction A + B → C (with kinetic rate k[A][B], reaction time t = 1 and kinetic constant k = 1), the plot has the yield and concentration of C for various combinations of A and B substrates. **b**, The plot shows the yield surface explicitly (black curve is the boundary at which the limiting reactant switches from A to B). As seen, yield—defined with respect to the concentration of the limiting reagent—tends to maximize at the boundaries of the region (orange curve in **a**) but the concentration of the product tends to maximize in its interior (blue curve in **a**). There are two implications of this 'inverse' dependence: (1) yields at very low concentrations may be artificially high, whereas—in reality—they involve very high (wasteful) excesses of some reagents; (2) detecting and quantifying the product reliably may be compromised in such a low-concentration regime. Accordingly, in the current work—especially for the multicomponent mixtures—we have worked over concentration ranges that do not extend to very small values, thereby assuring reliable detection while avoiding highly wasteful excess ratios. Such a 'practical' range is shown schematically in **a** by the dashed vertical lines. Another corollary is that discovery of unexpected products may be severely compromised when using a strategy of investigating only the edges and boundaries of an n-dimensional region while leaving the interior unexplored—a strategy that may otherwise seem intuitive to a chemist and has some merit in maximization of the yield of known products. **c**,**d**, Comparison of final concentration map (**c**) to yield map (**d**) for product **19e** in Hantzsch reaction. Note the difference between the conditions that maximize concentration versus ones that maximize yield: yield is maximized at the vertices of the cube, whereas concentration is maximized in the middle of the front face and top edge, with higher concentrations protruding deeper into the interior of the cube (see blue curve in **a**). The isosurfaces correspond to 5 mM, 10 mM, 15 mM, 20 mM, 25 mM and 30 mM concentrations of **19e** in **c**, 20%, 30%, 40%, 50% and 60% yields of **19e** in **d** and are calculated as explained in the caption of Fig. 2c,d.

Check for updates

Angewandte Chemie International Edition

Angewandte International Edition Chemie

GDCh German Chemical Society

How to cite: *Angew. Chem. Int. Ed.* **2026**, e9132891
doi.org/10.1002/anie.9132891

RESEARCH ARTICLE OPEN ACCESS

From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations

Yankai Jia[1] | Wai-Shing Wong[1] | Yasemin Bilgi-Gadina[1,2] | Louis P. J.-L. Gadina[1] | King-Lam Kwong[1] | Yanqiu Jiang[1] | Nikolaos Vagkidis[1,3] | Benito Marazzi[1,3] | Peter R. Schreiner[3] | Bartosz A. Grzybowski[1,2,4]

[1]IBS Center for Algorithmic and Robotized Synthesis (CARS), Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea | [2]Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland | [3]Institute of Organic Chemistry, Justus Liebig University, Giessen, Germany | [4]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

Correspondence: Peter R. Schreiner (prs@uni-giessen.de) | Bartosz A. Grzybowski (nanogrzybowski@gmail.com)

Received: 8 March 2026 | **Revised:** 21 May 2026 | **Accepted:** 1 June 2026

ABSTRACT

Chemical reactions are traditionally depicted as linear equations optimized toward a single product, with other outcomes treated as minor side products. However, advances in reaction automation now allow systematic exploration of reaction "hyperspaces"—that is, multidimensional landscapes of conditions—showing that even archetypal transformations can yield previously unrecognized dominant products. Using the Robowski robotic platform with low-cost NMR detection, we scanned nearly 1000 conditions for the bromination of 1,1-diphenylethylene (DPE) and benzylideneacetone (BDA). These maps revealed new major products, condition-dependent switchovers between addition and substitution pathways, and novel substitution patterns with potential for derivatization into biologically active scaffolds. In turn, correlations and anticorrelations in product distributions across the hyperspace facilitated reconstruction of mechanistic networks connecting different outcomes. Together, our findings show that even foundational reactions, long considered mechanistically complete, harbor hidden complexity and behave as dynamic, switchable networks embedded in conditions' space.

1 | Introduction

Recent advances in reaction automation, coupled with algorithms capable of deconvoluting complex reaction mixtures [1], have made it possible to map the outcomes of organic reactions exhaustively over multidimensional "hyperspaces" of reaction conditions. In this hyperspace perspective, reactions are no longer viewed as simple, "linear" conversions optimized toward a single desired product but as intricate, dynamic networks embedded in the hyperspace and potentially switchable between different major products (Figure 1a,b). In our recent studies [1–4], hyperspace scanning revealed unexpected complexity even in classical transformations long presumed to be mechanistically settled. For example, in the Pechmann coumarin synthesis (1884), the scan uncovered 11 previously unreported products, including biologically active benzofurans formed through non-classical pathways [2]. In the Biginelli reaction (1891), hyperspace exploration revealed a previously unknown seven-component transformation yielding scaffolds capable of chirality-specific dimerization and selective metal binding [3]. Likewise, in multicomponent reactions involving canonical 1,3-dicarbonyl substrates, the scans exposed new cyclization modes inaccessible through traditional one-variable-at-a-time experimentation [1, 4]. However, while such multicomponent systems are inherently rich in reactivity, it remains unclear whether simpler, foundational transformations harbor comparable hyperspace diversity. To probe this question, we turned to olefin bromination—a cornerstone of organic chemistry first reported by Berzelius in 1846 [5] and mechanistically dissected for more than a century [6–11]. On one hand, we did not expect the classic olefins (e.g., aliphatic alkenes, stilbenes,

Yankai Jia, Wai-Shing Wong, and Yasemin Bilgi-Gadina contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2026 The Author(s). *Angewandte Chemie International Edition* published by Wiley-VCH GmbH

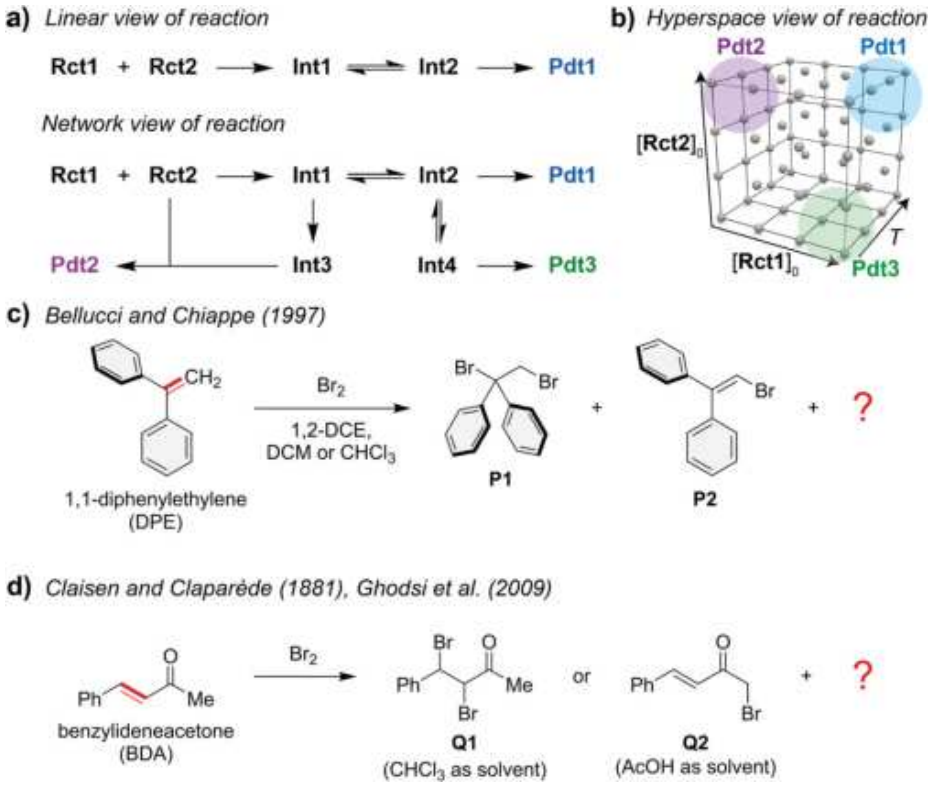


FIGURE 1 | Reactions as networks embedded in the hyperspaces of conditions. (a) Linear versus network views of a chemical reaction. In the linear view, reactants **Rct1** and **Rct2** give the desired product **Pdt1** via some hypothetical intermediates **Int1** and **Int2**; by-products and side-reactions are not considered productive and are typically omitted. By contrast, in the network view, it is essential to account for the possible competing reactions, here those involving additional intermediates **Int3** and **Int4** and additional end-species **Pdt2** and **Pdt3**, which are classically considered by-products but can become dominant products when the conditions are changed and the network is "switched." (b) Such switchable reaction networks are embedded in multidimensional hyperspaces of conditions. In our current work, the space is four-dimensional, but in this scheme we visualize a three-dimensional example of a space defined by the initial concentrations of **Rct1** and **Rct2** and temperature, *T*. The color coding indicates regions in which products **Pdt1** (*blue*), **Pdt2** (*purple*), and **Pdt3** (*green*) from panel (a) are dominant. (c) Bromination of DPE is known to give two different products, **P1** and **P2**, for different conditions. (d) Bromination of BDA is known to give the olefin bromination product **Q1**, as well as the methyl-brominated product **Q2**, with the olefin remaining unreacted. The central question of this work is whether additional products exist in the hyperspaces of these reactions.

or styrenes) to support reactivity channels alternative to classic electrophilic addition yielding vicinal dibromide products. On the other hand, we hypothesized that olefins prone to the addition-elimination process may support more intricate mechanistic networks leading to condition-dependent switchability between different major products. For instance, it is well-documented that 1,1-diphenylethylene (DPE; Figure 1c) can react with Br_2 to afford predominantly either the vicinal dibromide **P1** (the addition product) or the vinyl bromide **P2** (the substitution product), depending on conditions [12]. Whether this system supports *additional* major products elsewhere in hyperspace has never been examined. Similarly, the bifunctional substrate benzylideneacetone (BDA; Figure 1d) yields the addition product **Q1** in chloroform [13], whereas in acetic acid it precipitates as the methyl-substitution product **Q2** [14]. Whether other major products exist—and whether they can arise within the same solvent under different hyperspace coordinates—remains unknown. Here, we address these questions using the Robowski platform. In contrast to our previous implementations relying on UV–Vis detection [1–4], we integrate Robowski with a custom, low-cost NMR module operating in a fully automated closed loop of sample preparation, transfer, spectrum acquisition, and tube washing/drying. Using this system, we scan the hyperspaces of DPE and BDA bromination across nearly 1000 conditions. In both cases, we uncover regions of previously unrecognized reactivity, including major products not described in the literature, and substitution patterns that are synthetically useful for derivatization into biologically active compounds [15–19]. We then use the correlations (or anticorrelations) of product distributions across the hyperspace to help reconstruct the mechanistic networks leading to different products. These findings highlight the power of hyperspace-guided discovery and demonstrate that even archetypal reactions—long considered mechanistically complete—can harbor rich, condition-dependent reactivity. More broadly, this study serves as a proof of principle: hyperspace scanning, enabled by accessible robotic platforms, can redefine the boundaries of reaction-space exploration and establish reactions as dynamic, switchable networks embedded in conditions' space.

2 | Results and Discussion

2.1 | Construction of the Robotic Platform

For the preparation of all reaction mixtures, our studies relied on the Robowski system described in detail, along with replication

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

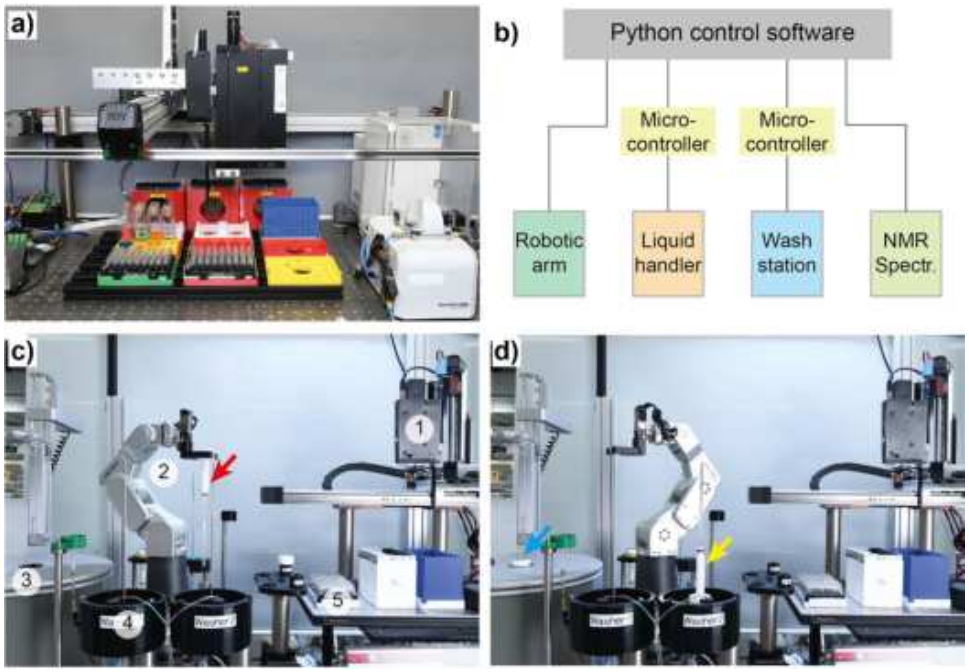


FIGURE 2 | Robowski and Robowski-NMR setups. (a) Photo of the basic, house-built Robowski setup for reaction preparation. For details, see ref. [1]. (b) Scheme for the automated NMR setup. (c) Photo of the Robowski-NMR system illustrating how the robotic arm is transferring an NMR tube (red arrow) containing a reaction mixture to the spectrometer: 1 = liquid handler for transferring liquids from reaction vials (prepared as in panel [a]) to NMR tubes; 2 = robotic arm for transferring the tubes among different locations; 3 = bench-top NMR spectrometer; 4 = one of two washing stations for cleaning the tubes after each measurement; 5 = 54-vial plate containing reaction crudes. (d) The same NMR setup where one tube/sample is being measured (blue arrow) while the other tube (yellow arrow) is placed upside down onto the nozzle supplying the washing solvent. Operation of this system is further illustrated in Movie S1.

blueprints, in ref. [1]. This in-house-built, inexpensive (~$25K) platform supports various organic solvents and harsh reagents and can be programmed to execute desired sequences of reagent addition, mixing, and temperature control over 54-vial plates of 2 mL vials (Figure 2a). In previous works, we used Robowski in conjunction with optical detection to approximate reaction yields based on the so-called unmixing of UV–Vis spectra of reaction crudes. While powerful in scope, we also noted that the unmixing approach is not without limitations and is ill-suited to quantify reaction outcomes, in which UV–Vis spectra of multiple products are highly correlated or weak (for a detailed discussion, see Supporting Information of ref. [1]). This is the case in olefin bromination reactions, calling for an alternative detection modality. Accordingly, we first set out to construct a Robowski-NMR system combining several components illustrated in Figure 2b, Section 2 of the Supporting Information, and Movie S1. First, a pipetting module was used to transfer the reaction crudes from the 54-vial plates (prepared by traditional Robowski from Figure 2a) to the NMR tubes. Once the tube was filled, a robotic arm (Meca500, Mecademic) transferred it to an 80 MHz benchtop NMR (Spinsolve 80, Magritek). After the spectrum was acquired (~7 min per sample), the robot removed the tube, turned it upside down, and threaded it onto a nozzle supplying a high-velocity stream of ethanol as a rinsing solvent for 30 s. After this first washing step, the tube was rinsed for the second time for 30 s at another washing station (Figure 2b) and was finally carried to a drying station under a stream of hot air (2 min). As these washing/drying operations took place, the robot was already filling in another NMR tube and acquiring its NMR spectrum. The control hardware (utilizing an Arduino microcontroller, see Supporting Information, Section 2) and house-written software orchestrated the entire process such that only these two tubes were in constant, alternating use—as one was being cleaned/dried, the other was being filled in/measured. At specified intervals (every 5 samples), the robot also inserted into the NMR a standard sample (95% D_2O with 5% H_2O) for shimming to ensure magnetic field homogeneity. In this manner, the system's throughput was limited only by the supply of samples to measure. The closed-loop design allowed us to process one 54-vial plate within ~6 h—this was significantly shorter than for using a carousel system, in which cleaning of the NMR tubes after each run added considerable time (~2 h per plate) and required additional manual operations of transferring reaction mixtures from vials into NMR tubes and reinserting tubes into the carousel. The longest continuous run we performed was 72 h. From an engineering point of view, we found this setup easier to operate and more adaptable to different types of samples/solvents than the flow NMR technique, which we had initially considered but which involved multiple operational challenges (e.g., time-consuming optimization of flow injection, material carryover for concentrated samples, and careful flow control) [20, 21]. In parallel, all NMR spectra acquired were processed by house-written scripts that perform spectral phasing, baseline correction, and peak fitting with Lorentzian/pseudo-Voigt functions. All codes for running the robot and for spectra processing as well as blueprints for system replication are deposited at Zenodo (https://doi.org/10.5281/zenodo.17080099).

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | Background Experiments With Simple Olefins

As a benchmark, we first performed brominations of three classic olefins: *cis*-stilbene [8], styrene [22], and 2-vinylpyridine. For each olefin, we tested four conditions with $[\text{olefin}]_0$ and $[Br_2]_0$ each set at either 75 or 300 mM. The NMR spectra (see Supporting Information, Section 3.1) of all these reaction mixtures indicated the formation of only the expected electrophilic addition product (i.e., the vicinal dibromide). Given no other reactivity channels and additional products, we did not investigate in more detail the hyperspaces for such simple olefins.

2.3 | Hyperspace of DPE Bromination

Instead, to probe the potential for expanded reactivity, we considered a four-dimensional space of reaction conditions defined by various initial concentrations of 1,1-diphenylethylene (DPE), Br_2 and an ionic additive, tetra-*n*-butylammonium bromide (TBABr), tetra-*n*-butylammonium tetrafluoroborate ($TBABF_4$), tetra-*n*-butylphosphonium bromide (TBPBr), or tetra-*n*-butylammonium tribromide ($TBABr_3$). Compared to TBABr, $TBABF_4$ represents the use of a non-nucleophilic counteranion (BF_4^-) instead of Br^-; TBPBr represents the use of a countercation (TBP^+) larger than TBA^+; and for $TBABr_3$, the Br_3^- ion is less basic than Br^- and can perturb the complexation equilibria of Br_2 differently than Br^- [23, 24]. The concentration boundaries for each three-dimensional DPE–Br2–additive subspace were chosen based on several criteria. For the upper boundary, we considered the maximal solubility of these reactants as well as the problems associated with distortion of NMR signals in the regime of excessively high concentrations (which made quantification of yields problematic). For the lower boundary, we wished the concentrations to remain high enough to allow for reliable quantification of even minor products. We also wished to keep the concentration ranges synthetically relevant—wide enough to produce variations in reaction outcomes but narrow enough to avoid unrealistic and wasteful stoichiometric ratios. For TBABr, the three-dimensional subspace was probed at seven concentrations of $[DPE]_0$ (75 and 300 mM in even intervals), seven concentrations of $[Br_2]_0$ (70–300 mM), and five concentrations of an ionic additive (0, 75, 150, 225, 300 mM), translating into $7 \times 7 \times 5 = 245$ conditions. For other salts, the zero-salt condition was not repeated, and each space spanned $7 \times 7 \times 4 = 196$ conditions. Altogether, the four-dimensional space was probed at $245 + 3 \times 196 = 833$ conditions. Robowski prepared all reaction mixtures by pipetting into each 2 mL vial the desired volume of the 1,2-dichloroethane (1,2-DCE) solvent followed by the desired volumes of stock solutions of salt (1.20 M for TBABr, $TBABF_4$, and TBPBr; 939 mM for $TBABr_3$), Br_2 (876 mM), and DPE (876 mM), such that the total volume of each mixture was 800 μL. The plates were then sealed and shaken to homogenize all reaction mixtures, incubated in an oven at 26 °C in darkness for 20 h, and then opened and transferred to the Robowski-NMR station to acquire ^{1}H NMR spectra of all reaction crudes.

2.3.1 | Mapping of Product Yield Distributions

Our exploration of product yield distributions started with TBABr. For this particular system, Bellucci and Chiappe [12] have previously reported 11 experiments evidencing formation of the addition product **P1** and the substitution product **P2**, with the latter favored for decreasing reactant concentrations. We also see these products in our spectra with diagnostic, cleanly resolved singlets at 4.51, 5.44, and 6.80 ppm assigned to **P1**, DPE, and **P2**, respectively (see Supporting Information, Section 3.2). These signals were fitted with Lorentzian/pseudo-Voigt functions, integrated, and converted to absolute concentrations using external calibration curves for DPE (423, 211, 106, 53, and 26 mM) and TBABr (0, 75, 150, 225, and 300 mM; see Supporting Information, Section 5.4). The yields for **P1** and **P2** were then quantified with respect to the limiting reactant, $Y_i = [\mathbf{P}_i]_{\text{final}}/\min([DPE]_0, [Br_2]_0 + [TBABr_3]_0)$, where we account for the fact that Br_3^- anion from $TBABr_3$ is also a source of Br_2 equivalent. Yield distributions were plotted using our HyperspaceViewer software [1] (https://gitlab.com/az-steak/hyperspace_viewer) and are shown in Figure 3a,b. As seen, for the addition product **P1**, the yields are maximal (60%) for $\{[DPE]_0, [Br_2]_0, [TBABr]_0\} = 262.5, 262.5, 0$ mM and gradually decrease away from this corner along all three axes (i.e., with decreasing $[DPE]_0$, decreasing $[Br_2]_0$ and increasing $[TBABr]_0$). By contrast, the substitution product **P2** forms in > 90% yields at high values of $[TBABr]_0 = 225$ and 300 mM regardless of $[DPE]_0$ and $[Br_2]_0$.

These outcomes are in line with the prior work [12], but they do not account for all possible products. To see this, we quantified the fraction of "unaccounted" DPE consumption—that is, the fraction of DPE that was converted but could not be explained by the formation of the already-identified products. For stoichiometric 1-to-1 conversions of DPE into products, defined the Residual Unaccounted Fraction of DPE as $\text{RUF} = 1 - \sum_{\text{known},\, i} [\mathbf{P}_i]_{\text{final}}/([DPE]_0 - [DPE]_{\text{final}})$. As illustrated in Figure 3c, the RUF map (based on the knowledge of only **P1** and **P2**) features maxima near two corners, namely (**A**), corresponding to excess of DPE and low $[TBABr]_0$, and (**B**), corresponding to excess Br_2 and low $[TBABr]_0$. The RUF values at corners **A** and **B** are 0.4 and 1.0, respectively. Hence, we proceeded to inspect the spectra collected at these two corners.

The ^{1}H NMR spectra of the crudes near corner **A** feature a singlet at $\delta = 2.46$ ppm. This product **P3** could not be isolated in pure form because of its lability upon aqueous workup, during which it eliminates to regenerate DPE [25, 26] (Figure 3g). However, the singlet's chemical shift and lack of spin–spin coupling partners are consistent with the structure of 1-bromo-1,1-diphenylethane, which we indeed confirmed in situ by ^{1}H NMR titration of DPE with HBr (Figure 3g and details in the Supporting Information, Section 3.5) The highest NMR-based yield of **P3** was estimated at 74% under two conditions of $\{[DPE]_0, [Br_2]_0, [TBABr]_0\} = 300, 75, 0$ mM and $262.5, 75, 0$ mM, as illustrated in the yield map in Figure 3d. We observe that the yield for **P3** decreases moving away from corner **A**—gradually as $[DPE]_0$ decreases and $[Br_2]_0$ increases, but rapidly as $[TBABr]_0$ increases.

In turn, crude ^{1}H NMR spectra near corner **B** feature a singlet at $\delta = 6.73$ ppm, which is slightly higher field than that for **P2** at $\delta = 6.80$ ppm. Upon aqueous workup and preparative thin-layer chromatography, we isolated 2,2-dibromo-1,1-diphenylethan-1-ol (**P4′**) and confirmed this assignment by 2D NMR spectroscopy (see Supporting Information, Section 3.4). However, this compound featured a singlet at $\delta = 6.77$ ppm, suggesting the possibility

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

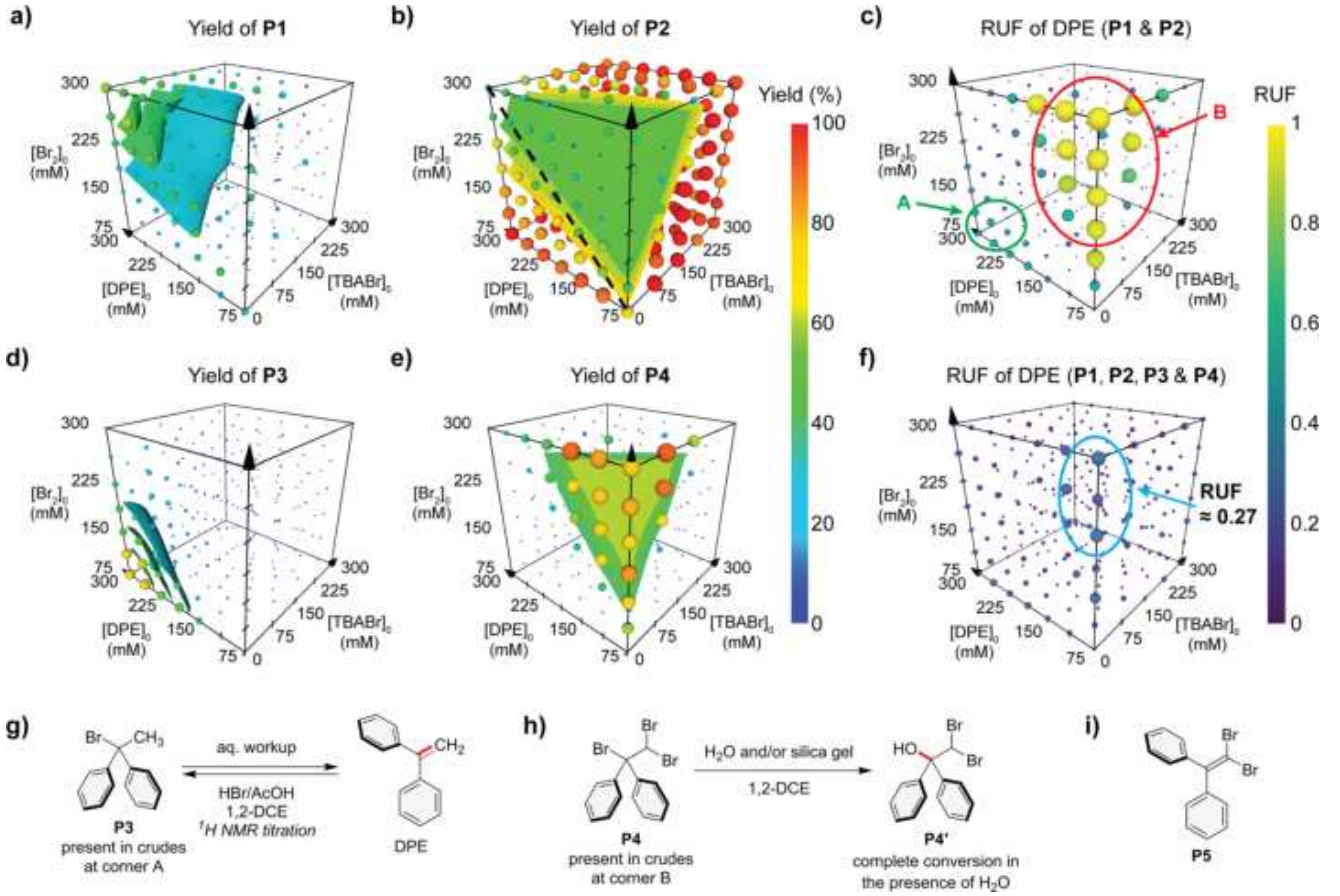


FIGURE 3 | Hyperspace distribution of products of DBE bromination. (a) Yield map of the known vicinal dibromide **P1** (isosurfaces displayed at 30%, 40%, and 50%). (b) Yield map of the known vinyl bromide **P2** (isosurfaces displayed at 50%, 60%, and 70%). The equimolar diagonal ($[DPE]_0 = [Br_2]_0$) when $[TBABr]_0 = 0$ is indicated as a black dashed line. (c) Map of the fraction of consumed DPE that could not be accounted for by the formation of **P1** and **P2**, generated using the equation for "residual unaccounted fraction": $RUF = 1 - \sum_{known,\,i}[\mathbf{P}_i]_{final}/([DPE]_0 - [DPE]_{final})$. The value of RUF varies from 0 (all reacted DPE accounted for) to 1 (all reacted DPE unaccounted for). The RUFs at the local maxima at corners **A** and **B** are, respectively, 0.4 and 1.0, indicating limited understanding of these regions. (d) Yield map of previously unreported hydrobromination product **P3** (isosurfaces displayed at 40%, 50%, and 60%). (e) Yield map of previously unreported tribromo derivative **P4** (isosurfaces displayed at 40%, 50%, and 60%). (f) Map of RUF of DPE based on the knowledge of **P1**, **P2**, **P3** and **P4**. Within the region with excess Br_2 and $[TBABr]_0 = 0$ (*cyan circle*), RUF ~ 0.27, and the vinylidene dibromide **P5** was identified therein under one condition by ^{13}C NMR spectroscopy. Colorbars to the right of the yield maps (a, b, d, e) and the RUF maps (c, f) indicate the yields and RUF values, respectively. (g) The previously unrecognized product **P3** is found in the crude near corner **A**. This product cannot be isolated in pure form because of rapid dehydrobromination to give DPE. The identity of **P3** was indirectly confirmed by ^1H NMR titration of DPE against HBr/AcOH. (h) In the presence of water, the unreported product **P4** hydrolyzed to alcohol **P4′** which was isolated for characterization. (i) The molecular structure of unreported **P5** identified by ^{13}C NMR spectroscopy under one condition at corner **B**.

of hydrolysis (during purification) of the parent 1,1,2-tribromo-2,2-diphenylethane (**P4**) formed in situ (Figure 3h). To verify this hypothesis, we performed a separate set of experiments in which robot-prepared samples featuring the highest intensity of the **P4**-indicative signal were subject to hydrolysis—this diagnostic peak gradually disappeared while the **P4′** signal at $\delta = 6.77$ grew in intensity (for details, see Supporting Information, Section 3.6). Across the space of conditions studied, the highest NMR-determined yield of **P4** was 91% for $\{[DPE]_0, [Br_2]_0, [TBABr]_0\} =$ 75, 300, 75 mM (see yield map in Figure 3e).

Having quantified the distributions of products **P1** through **P4**, we repeated the RUF analysis of the DPE substrate that reacted but was unaccounted for by the knowledge of these four products. As illustrated in Figure 3f, the values of RUF were now only a few percent, with the exception of a region near corner **B**, at which RUF was up to 0.27. Since this region overlaps with the region over which **P4** forms, we hypothesized it may be due to the presence of 1,1-dibromo-2,2-diphenylethylene **P5** (Figure 3i), which could result from the dehydrobromination of **P4**. To test this hypothesis, we repeated 16 conditions of high $[Br_2]_0$ around corner **B** and acquired their crude ^{13}C NMR spectra (100 MHz) after 24 h (^{13}C NMR spectra were preferable since **P5** lacks any characteristic olefinic or aliphatic protons and can be easily overlooked by ^1H NMR). These spectra were compared against the authentic sample synthesized separately by the treatment of benzophenone with carbon tetrabromide and triphenylphosphine [27]. Six out of 16 spectra showed all six ^{13}C resonances due to **P5**, including the three diagnostic peaks of the non-protonated carbon atoms (see Supporting Information, Table S2 for the 16 conditions and Figure S20 for the spectra). However, this product was formed only in trace amounts, and

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

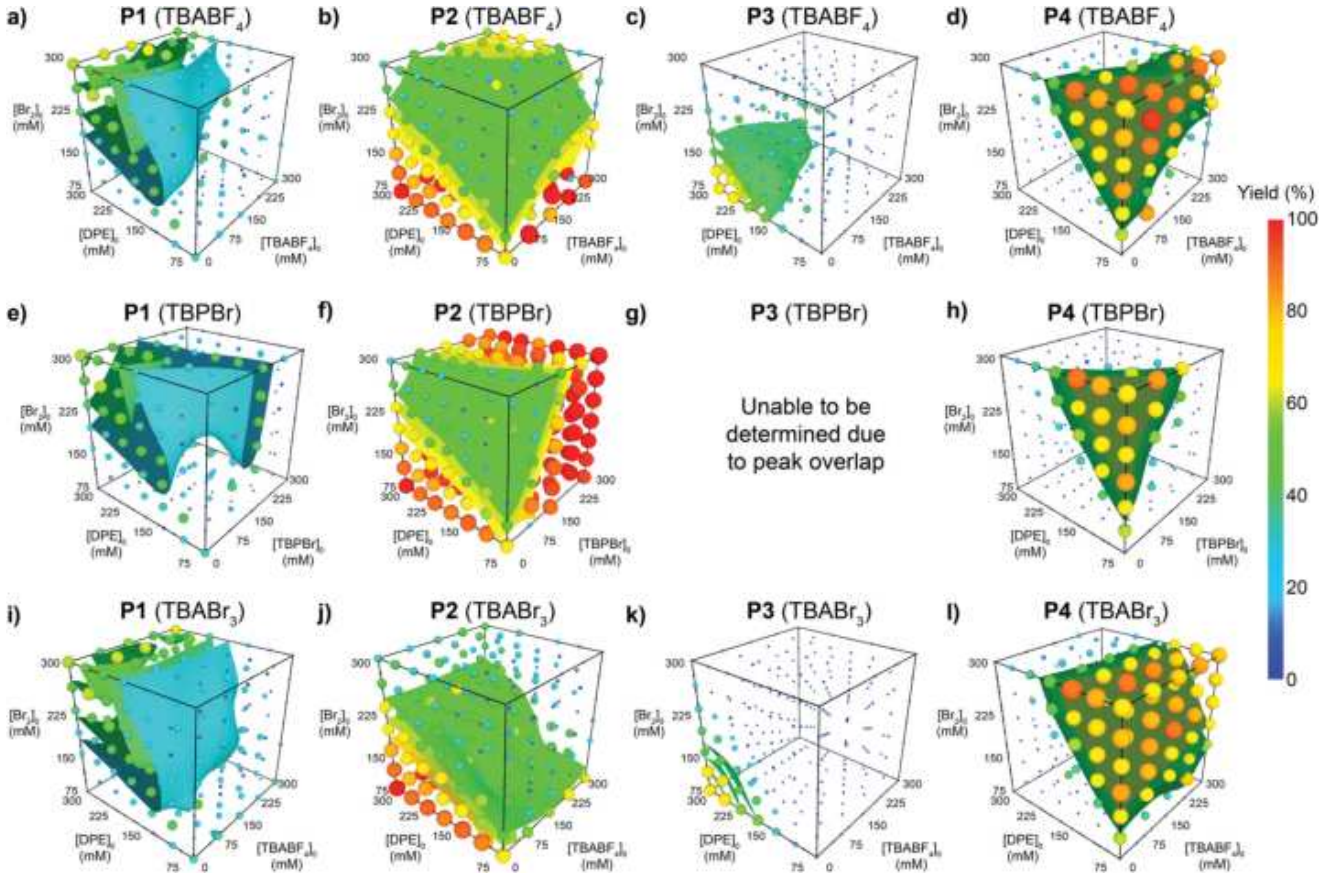


FIGURE 4 | Yield maps of DPE bromination using alternative ionic additives. The isosurfaces represent {30%, 40%, 50%}, {50%, 60%, 70%}, {40%, 50%, 60%}, and {40%, 50%, 60%} for **P1**, **P2**, **P3** and **P4**, respectively.

yield quantification by quantitative ^{13}C NMR spectroscopy in the presence of a paramagnetic relaxation agent [28, 29] was not pursued.

We note that hyperspace maps for the three other salts ($TBABF_4$, TBPBr, and $TBABr_3$) generally resembled those obtained with TBABr, though with subtle differences attributable to the unique properties of their cations. For $TBABF_4$, the maximal, 63% yield of **P1** is similar to that with TBABr (63% vs. 60%), but the region of appreciable (30%–50%) yields is broader (Figure 4a). At the same time, yields of **P2** are systematically lower (40%–85% vs. 85%–100% for TBABr) (Figure 4b). This increase in the **P1**/**P2** ratio upon changing the salt's counteranion from Br^- to BF_4^- can be reasonably ascribed to much slower elimination of **P1** to **P2** by the much weaker base BF_4^-. The maximal yields of **P3** and **P4** are similar to those with TBABr but the regions of appreciable yield are broader, for **P4** extending from the **B** corner toward high values of $[TBABF_4]_0$ (Figure 4c,d). Based on the mechanistic network from Figure 5b, this can be attributed to the inability of $TBABF_4$ to form the Br_2–Br^- complex, increasing steady-state concentration of Br_2 and, in effect, promoting conversion of **P2** to the π-complex **P2**•$2Br_2$ and the ion pair $\mathbf{C3}^+ \bullet Br_3^-$, eventually leading to more **P4**. For TBPBr, the maximum yield of **P1** is also ~60%, but the distribution of **P1** has an additional shoulder along the $[TBPBr]_0$ = 150 mM plane where a local, 47% yield maximum is located (Figure 4e). The yields around this local maximum remain fair, 25%–40%, compared to 0%–30% over the same regions for TBABr and $TBABF_4$. At the same time, yields of **P2** are slightly but systematically lower than for TBABr (Figure 4f). This trend may be due to a higher extent of Br_2–Br^- complexation to give Br_3^-, which may originate from a tighter ion pairing of the larger TBP^+ cation and Br_3^- anion. Consequently, a higher steady-state concentration of Br_3^- suppresses dissociation of the ion pair $\mathbf{C1}^+ \bullet Br_3^-$, limiting the formation of **P2**. The stronger Br_2–Br^- complexation also implies a lower steady-state concentration of Br_2, which is manifest in the slightly decreased yields of tribromide **P4** in the same region (Figure 4h). For **P3**, the ^1H resonance ($\delta \sim 2.68$ ppm) is not observed because of the overlap with the broad multiplet due to the α-CH_2 group of TBP^+ (Figure 4g). For $TBABr_3$, the most striking feature is that yields of **P2** are low, < 50%, over more than half of the space, while the regions of good yield for both **P1** and **P4** are broader than for other salts (Figure 4i–l). This can be explained by the Br_3^- inhibiting the dissociation of the ion pair $\mathbf{C1}^+ \bullet Br_3^-$ (thus favoring the formation of **P1**). In addition, Br_3^- serves as an additional source of molecular Br_2 such that a greater portion of **P2** formed can be further brominated to give **P4** via the intermediates **P2**•$2Br_2$ and $\mathbf{C3}^+ \bullet Br_3^-$.

2.3.2 | Mechanistic Network of DPE Bromination

Analysis of yield maps for the major products **P1–P4** (Figure 3a–e) and minor product **P5**, together with chemical reasoning, facilitated reconstruction of a mechanistic network for DPE bromination (Figure 5c). The elements of this network leading

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

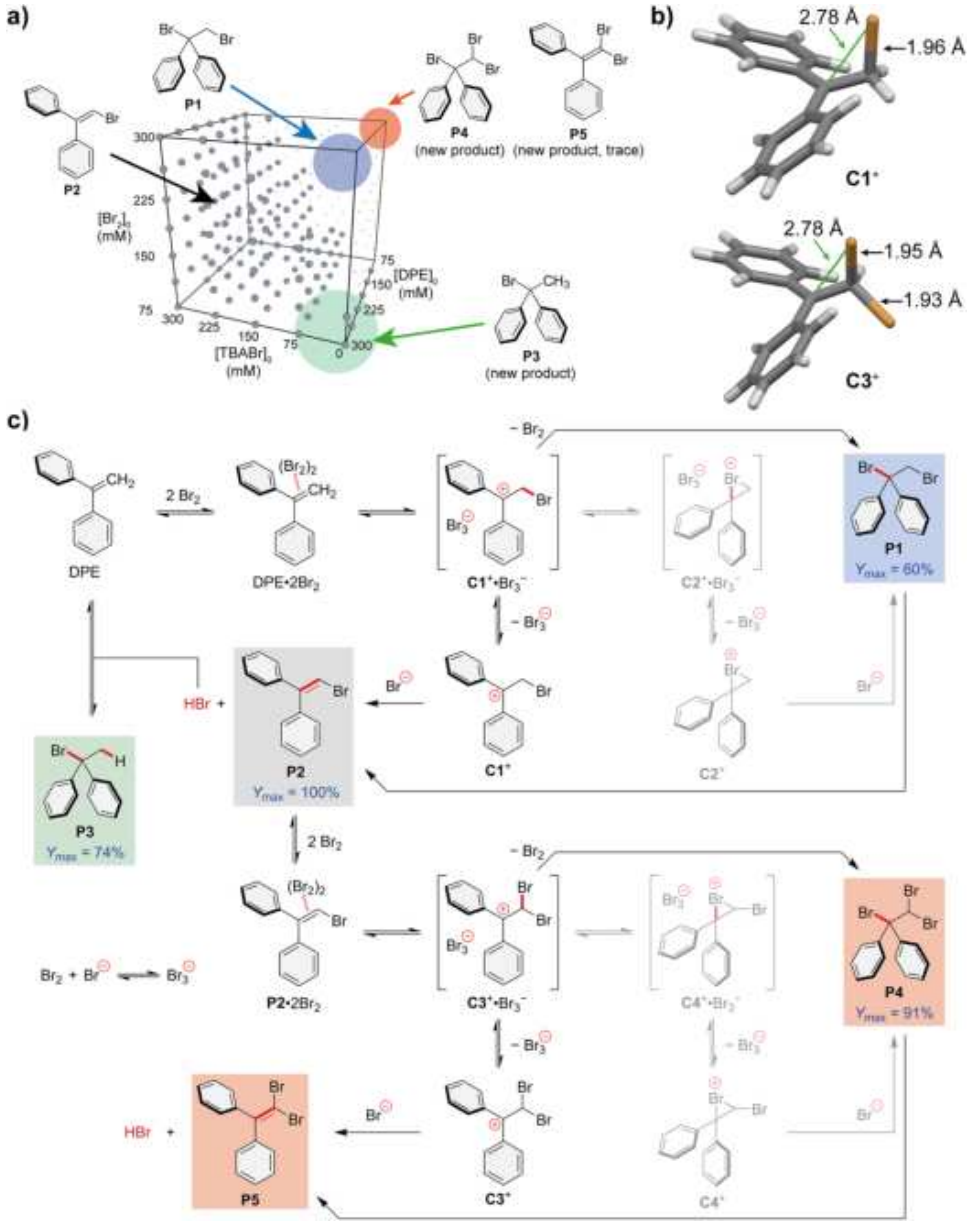


FIGURE 5 | Summary of the hyperspace product distributions and reconstruction of the plausible reaction network. (a) Hyperspace for the bromination of DPE showing the regions where the two known products (**P1** and **P2**) and three new products (**P3**, **P4,** and **P5**) are high yielding. Note that axes are rotated compared to Figures 3 and 4 to better visualize the distribution of gray points corresponding to the broad region of > 85% yield for **P2**. The vicinal dibromide **P1** is localized where $[DPE]_0$ and $[Br_2]_0$ are high and $[TBABr]_0 = 0$ (*blue shade*). The yield of the hydrobromination product **P3** is maximized with excess DPE and $[TBABr]_0 = 0$ (*green shade*). The tribromo derivative **P4** has maximum yield when Br_2 is in excess and $[TBABr]_0 = 0$ (*red shade*); vinylidene dibromide **P5** is also found therein as a trace product. (b) DFT-optimized geometries of the β-bromocarbenium ion $\mathbf{C1}^+$ (top) and β,β-dibromocarbenium ion $\mathbf{C3}^+$ (bottom) on the level of ωB97X-D/def2-TZVP/CPCM(1,2-DCE). In both structures, the distance between the carbocationic center and the Br atom is 2.78 Å (*green arrows*), which is significantly longer than the full C–Br bonds (1.93–1.96 Å, *black arrows*). Geometry optimization of the plausible cyclic bromonium ion $\mathbf{C2}^+$ always results in the open carbocation structure $\mathbf{C1}^+$, suggesting that $\mathbf{C2}^+$ does not exist as a local minimum. For further theoretical details, see Supporting Information, Sections 6.1 and 6.2. (c) Yield distributions and correlations of yield gradients allow for reconstructing the reaction's plausible mechanistic network. Maximum NMR yields (Y_{max}) are indicated below the structures of **P1**–**P4**. In accordance with the generally accepted mechanism [30], a rapid pre-equilibrium between DPE and Br_2 in 1:2 ratio yields the π-complex DPE•$2Br_2$, which subsequently undergoes a rate-limiting ionization to give, in solvents of moderate polarity, a tight carbenium–tribromide ion pair $\mathbf{C1}^+ \bullet Br_3^-$. These initial steps are supported by the second-order dependence of Br_2 on the reaction rate for various olefins [31–33] and DPE [12] in chlorinated solvents even at $[Br_2]_0$ as low as 10^{-4} M. For the ion pair $\mathbf{C1}^+ \bullet Br_3^-$, the constituent β-bromocarbenium ion $\mathbf{C1}^+$ is proposed based on Hammett plots constructed from kinetic data of substituted 1,1-diarylethylenes and the fitted reaction rate constant (ρ) [34, 35]. The cyclic bromonium ion $\mathbf{C2}^+$ (i.e., the bridged analogue of $\mathbf{C1}^+$) [10, 22] and the ion pair ($\mathbf{C2}^+ \bullet Br_3^-$) are in *light gray* because DFT calculations show that $\mathbf{C2}^+$ always converges to the open carbenium ion $\mathbf{C1}^+$ (see caption to panel b). In the accepted mechanism, the vicinal dibromide **P1** can be furnished from the ion pair $\mathbf{C1}^+ \bullet Br_3^-$ when a bond is formed between the benzhydrylic carbocationic center and the terminal Br atom of Br_3^- with the elimination of Br_2 [33, 36]. This pathway is consistent with the occurrence of **P1**'s maximum yield in our hyperspace at the corner with high $[DPE]_0$ and high $[Br_2]_0$ because the extent of $\mathbf{C1}^+ \bullet Br_3^-$ ion pair

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to the known products **P1** and **P2** are well established [31–36] and are narrated in the caption to Figure 5b. Regarding the unprecedented products **P3–P5**, we first recall that the maximal yield of **P3** is observed at corner **A** of the hyperspace (excess DPE with low $[TBABr]_0$) and drops gradually when $[DPE]_0$ decreases and $[Br]_0$ increases. In fact, a similar local yield trend around corner **A** is noticeable for the vinyl bromide **P2**. In other words, the yield gradients of **P2** and **P3** roughly correlate at the vicinity of corner **A** on the zero-salt plane. This observation corroborates the proposal that the Markovnikov hydrobromination of DPE to afford **P3** consumes the HBr by-product accompanying the formation of **P2**. As for the new tribromo derivative **P4**, the high-yielding region is confined near corner **B** (excess Br_2 with low $[TBABr]_0$) of the hyperspace, and the yield generally decreases as $[DPE]_0$ increases, $[Br_2]_0$ decreases and $[TBABr]_0$ increases. Interestingly, we observe a rough anti-correlation of such a yield gradient between **P4** and the vinyl bromide **P2**, where the latter's yield is zero at corner **B** and gradually increases along the three directions away from the corner. Thus, it is reasonable to assume that the formation of the tribromo derivative **P4** takes place at the expense of vinyl bromide **P2**. Mechanistically, **P2** may react further with excess Br_2, leading to the formation of the π-complex **P2**•$2Br_2$ and the β,β-dibromocarbenium–tribromide ion pair $\mathbf{C3}^+ \bullet Br_3^-$ [30]. The spontaneous collapse of $\mathbf{C3}^+ \bullet Br_3^-$ can yield the new tribromo derivative **P4**. Finally, the trace vinylidene dibromide **P5** can be afforded either from the dissociated β,β-dibromocarbenium ion $\mathbf{C3}^+$ via proton loss to a Br^- anion, or from the elimination of HBr from the tribromo derivative **P4**. We note that DFT calculations (Figure 5b and Supporting Information, Section 6.2) are consistent with the proposed open-carbocationic structures of $\mathbf{C1}^+$ (shown in the upper part of Figure 5c) and $\mathbf{C3}^+$. Overall, while the majority of the connections in the network are inferred, they are congruent with the hyperspace product distributions and accepted mechanistic knowledge; two transformations, DPE + HBr → **P3** and **P4** → **P5** + HBr, are experimentally validated by ^{1}H NMR titration results (Figure S13) and ^{13}C NMR spectroscopy (Figure S21).

2.3.3 | Hyperspace of Benzylideneacetone, BDA, Bromination

As the second case study, we explored the four-dimensional hyperspace of the bromination of an electron-poor BDA. This space was probed at four initial concentrations of BDA (75, 150, 225, and 300 mM), four initial concentrations of Br_2 (75, 150, 225, and 300 mM), and three initial concentrations of TBABr (0, 75, and 150 mM) at two reaction times (24 and 48 h). All reaction mixtures were prepared on the Robowski platform by sequentially pipetting 1,2-DCE solvent and three stock solutions of TBABr, Br_2, and BDA (each with a concentration of 876 mM) and were then subjected to homogenization and incubation at 26°C in darkness.

This time, the hyperspace was much more complex than for DPE and housed no less than 12 products shown in Figure 6 and isolated by scaling up specifically selected conditions on Robowski and then purifying the different crudes by HPLC (for details, see Supporting Information, Section 4.1). Using their reference ^{1}H NMR spectra to define diagnostic peaks, all crude spectra acquired at each point of the hyperspace were processed with our in-house, fully automated pipeline, which performed spectral phasing and baseline correction, fitted peaks using pseudo-Voigt functions, and assigned these fitted peaks to product-characteristic signals (see examples in the Supporting Information, Section 5.3). Product yields across the hyperspace were then computed from these assigned signals using external calibration. Repetitions under four representative conditions were performed to establish the reproducibility of our yield quantification protocol (see Supporting Information, Section 4.4).

Three species (*anti*-**Q1, Q2**, and **Q4**) were major products in different regions of the hyperspace (maximum yield of each > 60%), five (*syn*-**Q1**, **Q2′**, **Q5, Q6,** and **Q7**) had maximum yields between 10% and 50%, and the remaining four (**Q3**, **Q3′**, **Q7′,** and **Q8**) were not quantified in terms of yield distributions either because of severe spectral overlap or because of their minute quantities. All of the products were stable in the same solvent, including the **Q1** and **Q2** pair, which, as we narrated in the introduction, was previously obtained only in two different solvents ($CHCl_3$, AcOH) with **Q2** precipitating in AcOH. Three products (**Q4**, **Q5,** and **Q7**) were never reported before, whereas the chlorinated one (**Q2′**) has been prepared from BDA under different conditions, that is, using TMSCl/DMSO [37] or $BnMe_3N^+ \bullet ICl_2^-$ [38] as the chlorinating agent.

Figure 6a–h show individual yield distributions of the quantified products at 24 h (for *anti*-**Q1**, *syn*-**Q1**, **Q4**, **Q2**, **Q6,** and **Q7**) or 48 h (**Q5** and **Q2′**, for which appreciable yields became available at longer times). For the known addition product *anti*-**Q1** in Figure 6a, the global maximum (96%) is located at $\{[BDA]_0, [Br_2]_0, [TBABr]_0\}$ = 75, 225, 75 mM. Notably, a local maximum (82%) is located at $\{[BDA]_0, [Br_2]_0, [TBABr]_0\}$ = 225, 75, 0 mM and is surrounded by nine conditions with yields > 60% for $[BDA]_0 \geq [Br_2]_0$ on the zero-salt plane—these conditions correspond to the typical, no-salt literature conditions [13, 39] for the preparation of **Q1**. Within this same region, the diastereomer *syn*-**Q1** is found as a minor product with a maximum yield of 19% (Figure 6b). Next, the major product **Q4** is formed in up to 81% yield at $\{[BDA]_0, [Br_2]_0, [TBABr]_0\}$ = 75, 225, 0 mM at 24 h (Figure 6c). **Q5** is also produced within this region albeit in much smaller yields, up to 30% at $\{[BDA]_0, [Br_2]_0, [TBABr]_0\}$ = 75, 300, 0 mM at 48 h

dissociation is minimal and favors such a spontaneous collapse. In order to form the vinyl bromide **P2**, the ion pair $\mathbf{C1}^+ \bullet Br_3^-$ must undergo dissociation to give free ions of $\mathbf{C1}^+$. The β-deprotonation of $\mathbf{C1}^+$ by a Br^- anion leads to **P2** and the by-product HBr. As concentration decreases from 300 to 75 mM along the equimolar diagonal ($[DPE]_0 = [Br_2]_0$) on the zero-salt plane ($[TBABr]_0 = 0$), the yield of **P2** increases from 32% to 70%. This is in accord with the necessity of the dissociation of $\mathbf{C1}^+ \bullet Br_3^-$ which happens to a larger extent upon dilution. For the discussion of the lower portion of the network, leading to unprecedented products **P3–P5**, see main text. *Note:* As in the upper portion for the network, the cyclic bromonium ion $\mathbf{C4}^+$ (i.e., the bridge analogue of $\mathbf{C3}^+$) and the ion pair ($\mathbf{C4}^+ \bullet Br_3^-$) are in *light gray* because DFT computations show that $\mathbf{C4}^+$ always converges to the open carbenium ion $\mathbf{C3}^+$ (for further theoretical details, see Supporting Information, Section 6.2).

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

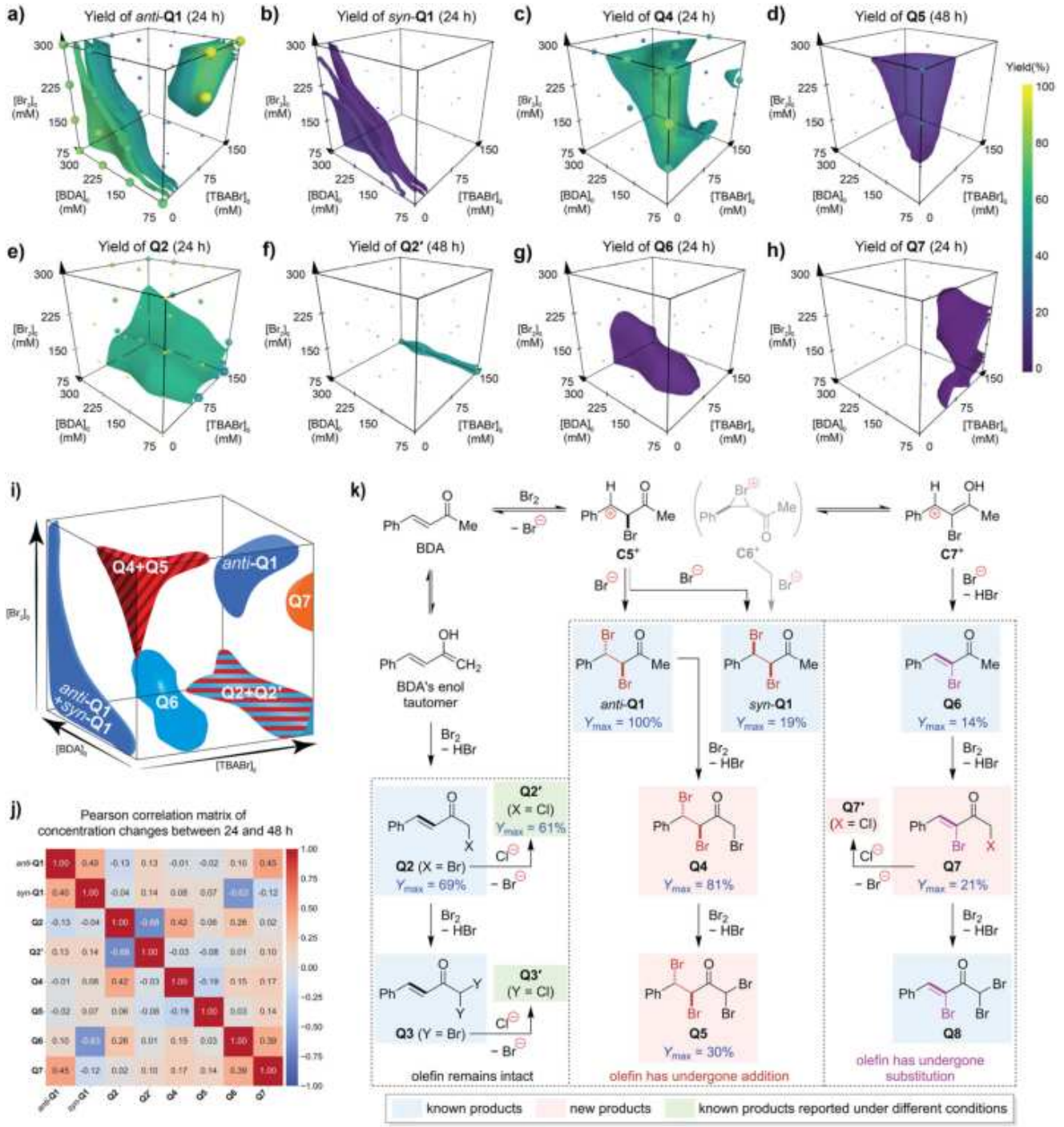


FIGURE 6 | Exploration of the hyperspace of benzylideneacetone bromination. (a–h) Yield maps for *anti*-**Q1**, *syn*-**Q1**, **Q4**, **Q5**, **Q2**, **Q2′**, **Q6,** and **Q7**. The reaction time is 24 h for these products except for **Q5** and **Q2′** (reaction time = 48 h). The isosurfaces are at 50%, 60%, and 70% (for a and c); 10%, 12%, and 15% (for b and d); 40 and 50% (for e); 50% (for f); 10% (for g); 8%, 10%, and 12% (for h). The set of eight yield maps at the other reaction time (48 or 24 h) is provided in the SI, Section 4.3. (i) Schematic illustration of the distributions of products from (a–h). (j) Correlation matrix of the concentration changes for *anti*-**Q1**, *syn*-**Q1**, **Q2**, **Q2′**, **Q4**, **Q5**, **Q6,** and **Q7**. The Pearson correlation between compounds X and Y is given by $r_{XY} = \sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y}) / \sqrt{\sum_{k=1}^{n}(x_k - \bar{x})^2 \sum_{k=1}^{n}(y_k - \bar{y})^2}$ where x_k and y_k are the concentration changes between 24 and 48 h under the k-th experimental condition. (k) Reconstructed mechanistic network. The maximum yields (Y_{max}) for the eight quantifiable products at either 24 or 48 h reaction times are shown below the compound labels. The putative cyclic bromonium ion **C6**$^+$ is shown in light gray since it always converges to the open-carbocationic structure **C5**$^+$ upon geometry optimization (for further theoretical details, see Supporting Information, Section 6.2).

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Figure 6d). In turn, the known product **Q2** shows appreciable yields when $[Br_2]_0$ = 75 mM and when $[TBABr]_0$ is non-zero, with its yield maximized (69%) at {$[BDA]_0$, $[Br_2]_0$, $[TBABr]_0$} = 150, 75, 75 mM (Figure 6e). Along the edge of {$[Br_2]_0$, $[TBABr]_0$} = 75, 150 mM, three out of four conditions give yields of 67%–69%. Interestingly, the related chloro-substituted product **Q2′** is colocalized with **Q2** and at 48 h reaches 57%–60% yields at four conditions along the said edge (Figure 6f). The minor product **Q6** has the yield maximum (14%) at {$[BDA]_0$, $[Br_2]_0$, $[TBABr]_0$} = 150, 75, 75 mM (Figure 6g), whereas the 20% yield maximum of **Q7** is located at {$[BDA]_0$, $[Br_2]_0$, $[TBABr]_0$} = 75, 150, 150 mM (Figure 6h). The various regions occupied by the eight quantified products in the 3D subspace are schematically summarized in Figure 6i—this map emphasizes multiple switchovers between different products, and the striped regions denote the presence of two non-isomeric products.

2.3.4 | Mechanistic Network of BDA Bromination

These distributions and the hyperspace colocalization of different species allowed for inferring connections between the species involved and supported the reconstruction of a plausible mechanistic network for BDA bromination. In Figure 6k, the products identified are grouped into three categories: (1) the olefin of BDA undergoes electrophilic addition, (2) the olefin remains intact, and (3) the olefin undergoes substitution. In each category, the methyl group of BDA is halogenated to different extents. Regarding group (1), BDA reacts with Br_2 to generate the open β-bromocarbenium ion **C5**$^+$ (in DFT computations, the plausible cyclic bromonium ion **C6**$^+$, shown in light gray in Figure 6k, always converges to **C5**$^+$ upon geometry optimization). DFT computations further show that the *anti*-attack of the **C5**$^+$ by Br^- is barrierless to furnish *anti*-**Q1**. However, the *syn*-attack of **C5**$^+$ by Br^- always converges to regenerate BDA and Br_2 instead of leading to *syn*-**Q1** (for details, see Supporting Information, Section 6.3). These theoretical results are in agreement with the preference for *anti*-**Q1** over *syn*-**Q1** in this corner of the hyperspace. As mentioned, *syn*-**Q1** is absent in the vicinity of *anti*-**Q1**'s global maximum where Br_2 is in excess and $[TBABr]_0$ = 75 or 150 mM. The exclusive formation of *anti*-**Q1** in this region is attributed to the *anti*-attack of **C5**$^+$ by the large excess of Br_3^- which, according to DFT computations, is also barrierless. Lastly, when Br_2 is the excess reactant and TBABr is absent, the methyl group of *anti*-**Q1** can be further brominated to give **Q4** as the major product and **Q5** as the minor product.

For product category (2), the methyl group of BDA is brominated via the enol tautomer to afford the monobromide **Q2** (major product) and dibromide **Q3** (trace product). However, **Q2** was found to also undergo substitution by Cl^- to yield the corresponding monochloride **Q2′**. In control experiments, the solvent 1,2-DCE was found to react with TBABr to give 1-bromo-2-chloroethane. This suggests that Cl^- released from the solvent molecules underlies the unexpected bromine-to-chlorine substitution observed for **Q2** (see Supporting Information, Section 4.5).

For product category (3), the keto form **C6**$^+$ must tautomerize to the enol form **C7**$^+$ which is then deprotonated by Br^- to give the known, minor product **Q6**. The methyl group of **Q6** can be further mono- and dibrominated to give **Q7** (minor product) and **Q8** (trace product), respectively. Again, **Q7** was found to be substituted by free Cl^- to give the trace product **Q7′**.

Finally, since we performed our analyses at two time points (24 and 48 h), we were also interested in whether and how the disappearance of some products may be related to the formation of others—such correlations could further corroborate the reaction mechanism. Accordingly, we performed Pearson correlation analysis for the changes in concentration of the eight measurable products between 24 and 48 h. In particular, if one product is converted into another in this time period, their concentration changes should be strongly anti-correlated. As shown in the correlation matrix (Figure 6j), only two product pairs give such anti-correlations, namely the **Q2**/**Q2′** pair (r = −0.68) and the *syn*-**Q1**/**Q6** pair (r = −0.63). Both anti-correlations have p values < 0.01 and are thus statistically significant (for details, see Supporting Information, Section 4.6 and Figure S65). First, r = −0.68 between **Q2** and **Q2′**, which is consistent with the proposed mechanism whereby **Q2** is slowly displaced by Cl^- (which, in turn, is slowly liberated from 1,2-DCE by TBABr) to produce the chloro-substituted derivative **Q2′**. Second, *syn*-**Q1** and **Q6** have r = −0.63, pinpointing the possibility that *syn*-**Q1** is slowly eliminated to **Q6** even in the absence of added Br^-. In contrast, the major diastereomer *anti*-**Q1** lacks correlation with **Q6** (r = 0.10), suggesting a difference in reactivity between the two diastereomers toward dehydrobromination. Strictly speaking, these two transformations are merely reasonable hypotheses based on the correlation analysis. In principle, anti-correlations can also arise when two independent product formation pathways respond oppositely to condition variables. Mindful of this, we plotted the maps of concentration changes of *syn*-**Q1** and **Q6** from 24 to 48 h, focusing on a region along the salt-free plane where six conditions exhibit negative values of Δ[*syn*-**Q1**] and also positive values of Δ[**Q6**] (Figure S66). For all these six conditions, the decrease in [*syn*-**Q1**] is roughly compensated by the increase in [**Q6**] (Table S8). This material balance further supports the hypothesis of an overall *syn*-**Q1** → **Q6** conversion. Analogous analysis of concentration changes and material balance was performed for the **Q2**/**Q2′** pair, also corroborating the proposed conversion from **Q2** to **Q2′** (Figure S67 and Table S9) under specific conditions of the hyperspace. In the future, it would be desirable to validate these mechanistically inferred transformations experimentally by kinetic measurements. All in all, this type of "global" analysis of yield data enabled the uncovering of not only the differential reactivity of diastereomers, but also the participation of solvent molecules in the reaction network. In the broad sense, this implies chemists should pay attention to potential indirect involvement of solvents (even halogenated solvents that are generally considered inert), particularly within "extreme" regions of the conditions' hyperspace.

3 | Conclusions

Our exploration of hyperspaces in olefin bromination demonstrates that even "textbook" reactions may conceal unexpected diversity. Simple olefins behaved predictably, but substrates such as DPE and BDA produced multiple condition-dependent outcomes—including previously unreported major products and mechanistic pathways—highlighting how comprehensively our

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

workflow can elucidate complex reaction systems. Here, this comprehensiveness was in part due to the systematic grid search of the hyperspace. Naturally, this is a resource-intensive strategy, and one may legitimately envision the use of gradient-based and adaptive sampling methods, which would be more effective in optimizing yields of products detected across broad regions of the hyperspace—it should be remembered, however, that such methods could also miss products that form in low yields only within small hyperspace regions (e.g., **Q2′** here). Looking forward, two aspects are critical. First, automation of hyperspace exploration is indispensable for fundamental discovery: without robotic preparation and high-throughput detection, the multidimensional hyperspaces cannot be reconstructed, and the hidden mechanistic networks remain inaccessible. Automation here is not about accelerating known chemistry, but about opening new windows onto reactivity. Second, the synergy between hyperspace mapping and correlation analysis of yield distributions, interpreted alongside widespread computational approaches such as DFT, transforms descriptive product maps into mechanistic reconstructions. With these components in place, hyperspace scanning can be generalized to other transformations, redefining reactions as dynamic, condition-embedded networks and establishing a method for mechanistic discovery at scale.

Author Contributions

Yankai Jia: investigation, methodology, validation, software, data curation, formal analysis, writing – original draft. **Wai-Shing Wong:** investigation, methodology, validation, software, data curation, formal analysis, writing – original draft. **Yasemin Bilgi-Gadina:** investigation, methodology, software, data curation, formal analysis, writing – original draft. **Louis P. J.-L. Gadina:** investigation, methodology, validation, software, data curation. **King-Lam Kwong:** software, data curation, investigation. **Yanqiu Jiang:** software, data curation, investigation. **Nikolaos Vagkidis:** data curation, investigation. **Benito Marazzi:** data curation, investigation. **Peter R. Schreiner:** funding acquisition, writing - review and editing, project administration, resources, supervision. **Bartosz A. Grzybowski:** conceptualization, funding acquisition, writing – original draft, methodology, supervision, resources, project administration, writing – review and editing.

Acknowledgments

The authors thank the taxpayers of Korea for generous funding under Award IBS-R020-D1 (to Bartosz A. Grzybowski). This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Advanced Grant Number. 101054751, "COLDOC," to Peter R. Schreiner). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors are grateful to Dr. Yaroslav I. Sobolev for the mass spectral fitting of compound **P4** and the theoretical calculations. Computations were performed on the Olaf supercomputer, operated by the Research Solution Center at the Institute for Basic Science. The authors also appreciate the discussion about mass spectral data with Prof. Dietmar Kuck from Bielefeld University.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.17080099.

References

1. Y. Jia, R. Frydrych, Y. I. Sobolev, et al., "Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks," *Nature* 645 (2025): 922–931, https://doi.org/10.1038/s41586-025-09490-1.

2. B. Prajapati, W.-S. Wong, R. Frydrych, et al., "Robot-Assisted Reconstruction and Control of the Pechmann Reaction Network," *Angewandte Chemie International Edition* 65 (2026): e18394, https://doi.org/10.1002/anie.202518394.

3. D. Matuszczyk, R. Frydrych, W.-S. Wong, et al., "Hyperspace Exploration Using Robotics for the Discovery of Mechanistically Distinct Transformations and Complex Functional Products," *Nature Synthesis* (2026), https://doi.org/10.1038/s44160-026-01096-3.

4. W.-S. Wong, Y. Jia, R. Frydrych, et al., Systems-level decoding of reaction mechanisms: From hyperspace yield distributions to kinetic networks.

5. J. Berzelius, Jahres-Bericht über die Fortschritte der Chemie und Mineralogie (Laupp'sche Buchhandlung, 1846), 616–621.

6. J. Meisenheimer, "Über Substitution und Addition," *Justus Liebigs Annalen der Chemie* 456 (1927): 126–151, https://doi.org/10.1002/jlac.19274560108.

7. J. B. Conn, G. B. Kistiakowsky, and E. A. Smith, "Heats of Organic Reactions. VII. Addition of Halogens to Olefins," *Journal of the American Chemical Society* 60 (1938): 2764–2771, https://doi.org/10.1021/ja01278a063.

8. R. E. Buckles, J. M. Bader, and R. J. Thurmaier, "Stereospecificity of the Addition of Bromine to *cis*- and *trans*-Stilbene," *Journal of Organic Chemistry* 27 (1962): 4523–4527, https://doi.org/10.1021/jo01059a097.

9. M. F. Ruasse and J. E. Dubois, "Reactivity of Ethylenic Compounds. Bromination. XXX. Electrophilic Bromination of Aromatic Conjugated Olefins. I. Evaluation of a Competitive Path Mechanism in Bromination of Trans Monosubstituted Stilbenes," *Journal of Organic Chemistry* 37 (1972): 1770–1778, https://doi.org/10.1021/jo00976a023.

10. J.-E. Dubois, M.-F. Ruasse, and A. Argile, "Styrene Bromination. The Competition Between Carbonium and Bromonium Ion Pathways Evidenced by Increasing the Electronic Demand on the Bromine Atom," *Tetrahedron Letters* 19 (1978): 177–180, https://doi.org/10.1016/S0040-4039(01)85077-2.

11. H. Mayr, E. Will, U. W. Heigl, and C. Schade, "Bromination of Octmethylcyclopentene—The Irregular Reactivity of a Sterically Hindered Cycloalkene," *Tetrahedron* 42 (1986): 2519–2522, https://doi.org/10.1016/0040-4020(86)80016-3.

12. G. Bellucci and C. Chiappe, "A Kinetic, Product and Kinetic Isotope Effect Investigation of the Bromination of 1,1-Diphenylethylenes and of Their 2,2-Dideuterio Derivatives," *Journal of the Chemical Society, Perkin Transactions* 2 (1997): 581–584, https://doi.org/10.1039/a605390d.

13. L. Claisen and A. Claparède, "Condensationen von Ketonen mit Aldehyden," *Berichte der Deutschen Chemischen Gesellschaft* 14 (1881): 2460–2468, https://doi.org/10.1002/cber.188101402192.

14. S. Ghodsi, S. Fouladi, H. A. Safari, and M. Taghi-Ganji, "Synthesis and Characterization of New Fluoroquinolones Containing an N Substituted Piperazine Moiety," *Asian Journal of Chemistry* 21 (2009): 3037.

15. X.-C. Li, D. Ferreira, M. R. Jacob, et al., "Antifungal Cyclopentenediones From Pipercoruscans," *Journal of the American Chemical Society* 126 (2004): 6872–6873, https://doi.org/10.1021/ja048081c.

16. H. Kusama, Y. Onizawa, and N. Iwasawa, "W(CO) 5 (L)-Catalyzed Tandem Intramolecular Cyclopropanation/Cope Rearrangement for the Stereoselective Construction of Bicyclo[5.3.0]doecane Framework," *Journal of the American Chemical Society* 128 (2006): 16500, https://doi.org/10.1021/ja0671924.

17. R. Horák, L. Kvapil, K. Motyka, et al., "Synthesis of 2-Alkenyl-3-hydroxyquinolin-4(1*H*)-ones as Promising Antimicrobial and Fluorescent Agents," *Tetrahedron* 74 (2018): 366–374, https://doi.org/10.1016/j.tet.2017.12.010.

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

18. R. Kamal, R. Kumar, V. Kumar, and V. Bhardwaj, "Synthetic Utilization of α,β-Chalcone Dibromide in Heterocyclic Chemistry and Stereoselective Debromination," *ChemistrySelect* 4 (2019): 11578–11603, https://doi.org/10.1002/slct.201902262.

19. Á. Ramírez-Trinidad, K. Carrillo-Jaimes, J. A. Rivera-Chávez, and E. Hernández-Vázquez, "Synthesis and Cytotoxic/Antimicrobial Screening of 2-Alkenylimidazo[1,2-*a*]pyrimidines," *Medicinal Chemistry Research* 32 (2023): 144–157, https://doi.org/10.1007/s00044-022-02997-6.

20. G. C. Leo, A. Krikava, and G. W. Caldwell, "Application of Flow NMR to an Open-Access Pharmaceutical Environment," *Analytical Chemistry* 75 (2003): 1954–1957, https://doi.org/10.1021/ac026389l.

21. M. V. Gomez and A. de la Hoz, "NMR Reaction Monitoring in Flow Synthesis," *Beilstein Journal of Organic Chemistry* 13 (2017): 285–300, https://doi.org/10.3762/bjoc.13.31.

22. M. F. Ruasse, A. Argile, and J. E. Dubois, "Reactivity of Ethylenic Compounds. 42. The Multipathway Mechanism in Styrene Bromination: Competition Between Bromonium and Carbonium Ion Intermediates," *Journal of the American Chemical Society* 100 (1978): 7645–7652, https://doi.org/10.1021/ja00492a035.

23. R. O. Griffith, A. McKeown, and A. G. Winn, "The Bromine-Bromide-Tribromide Equilibrium," *Transactions of the Faraday Society* 28 (1932): 101–107, https://doi.org/10.1039/tf9322800101.

24. G. Bellucci, R. Bianchini, C. Chiappe, and R. Ambrosetti, "The Formation of Pentabromide Ions From Bromine and Bromide in Moderate Polarity Aprotic Solvents and Their Possible Involvement in the Product Determining Step of Olefin," *Journal of the American Chemical Society* 111 (1989): 199–202, https://doi.org/10.1021/ja00183a031.

25. J. P. Moerdyk and C. W. Bielawski, "Dihaloimidazolidinediones as Versatile Halodehydrating Agents," *Chemistry – A European Journal* 20 (2014): 13487–13490, https://doi.org/10.1002/chem.201403407.

26. X.-M. Cui, Y.-H. Guan, N. Li, et al., "A Mild and Efficient Method for Bromination of Alcohols Using α,α-Dibromo-β-dicarbonyl Compounds as Halogen Sources," *Tetrahedron Letters* 55 (2014): 90–93, https://doi.org/10.1016/j.tetlet.2013.10.120.

27. G.-F. Zhang, Z.-Q. Chen, M. P. Aldred, et al., "Direct Validation of the Restriction of Intramolecular Rotation Hypothesis via the Synthesis of Novel *ortho*-Methyl Substituted Tetraphenylethenes and Their Application in Cell Imaging," *Chemical Communications* 50 (2014): 12058–12060, https://doi.org/10.1039/C4CC04241G.

28. H. H. Fischer, M. Seiler, T. S. Ertl, et al., "Quantification Studies in Continuous-Flow ^{13}C Nuclear Magnetic Resonance Spectroscopy by Use of Immobilized Paramagnetic Relaxation Agents," *Journal of Physical Chemistry B* 107 (2003): 4879–4886, https://doi.org/10.1021/jp021631d.

29. A. Bara-Estaún, M. C. Harder, C. L. Lyall, J. P. Lowe, E. Suturina, and U. Hintermair, "Paramagnetic Relaxation Agents for Enhancing Temporal Resolution and Sensitivity in Multinuclear FlowNMR Spectroscopy," *Chemistry – A European Journal* 29 (2023): e202300215, https://doi.org/10.1002/chem.202300215.

30. G. Bellucci, C. Chiappe, and F. Marioni, "Evidence for a Reversible Electrophilic Step in Olefin Bromination. The Case of Stilbenes," *Journal of the American Chemical Society* 109 (1987): 515–522, https://doi.org/10.1021/ja00236a032.

31. A. Modro, G. H. Schmid, and K. Yates, "Electrophilic Additions to Multiple Bonds. 2. Medium Effect on Bromine Additions to alkenes," *Journal of Organic Chemistry* 42 (1977): 3673–3676, https://doi.org/10.1021/jo00443a008.

32. P. B. D. de la Mare and R. D. Wilson, "The Kinetics and Mechanisms of Additions to Olefinic Systems. Part 24. Reactions of 3-substituted Cholest-5-enes with Sources of Electrophilic Bromine," *Journal of the Chemical Society, Perkin Transactions 2* 1977 (1977): 2048–2054.

33. G. Bellucci, R. Bianchini, R. Ambrosetti, and G. Ingrosso, "Comparison of Molecular Bromine and Tribromide Ion as Brominating Reagents. 1. Kinetic Evidence for Different Mechanisms of Addition to Cyclohexene," *Journal of Organic Chemistry* 50 (1985): 3313–3318, https://doi.org/10.1021/jo00218a013.

34. A. F. Hegarty, J. E. Dubois, and E. D. Bergmann, "Bromination of 1,1-Diphenylethylenes. I. Kinetic Study of Monosubstituted Derivatives in Methanol," *Journal of Organic Chemistry* 37 (1972): 2218–2222, https://doi.org/10.1021/jo00979a002.

35. A. F. Hegarty, J. E. Dubois, J. S. Lomas, W. V. Wright, and E. D. Bergmann, "Bromination of 1,1-Diphenylethylenes. II. Resonance Saturation and Geometrical Effects on the Reactivity Multiply Substituted Derivatives," *Journal of Organic Chemistry* 37 (1972): 2222–2228, https://doi.org/10.1021/jo00979a003.

36. G. Bellucci, R. Bianchini, and R. Ambrosetti, "Direct Evidence for Bromine-Olefin Charge-Transfer Complexes as Essential Intermediates of the Fast Ionic Addition of Bromine to Cyclohexene," *Journal of the American Chemical Society* 107 (1985): 2464–2471, https://doi.org/10.1021/ja00294a044.

37. R. R. Fraser and F. Kong, "The Chlorination of Ketones Using Trimeihylchlorosilane and Dimetwylsulfoxide with Bromide Ion Catalysis," *Synthetic Communications* 18 (1988): 1071, https://doi.org/10.1080/00397918808060892.

38. J. Zhou, J. Li, Y. Li, et al., "Direct Synthesis of 3-Acylindoles Through Rhodium(III)-Catalyzed Annulation of N -Phenylamidines with α-Cl Ketones," *Organic Letters* 20 (2018): 7645–7649, https://doi.org/10.1021/acs.orglett.8b03383.

39. N. H. Cromwell, "α,β-Unsaturated Aminoketones. III. α- and β-Morpholinobenzalacetones," *Journal of the American Chemical Society* 62 (1940): 3470–3473, https://doi.org/10.1021/ja01869a051.

40. R. Neumann, F. de la Vega, and A. Bar-On, "α-Bromo Carbonyl Compounds as Promoters for the Synthesis of (2-Bromoethyl)benzene by the Anti-Markovnikov Addition of Hydrogen Bromide to Styrene," *Journal of Organic Chemistry* 60 (1995): 1315–1318, https://doi.org/10.1021/jo00110a040.

41. J. Panda, J. Sahoo, J. Dutta, H. S. Biswal, and G. Sahoo, "Spectroscopic and Computational Study of the Organocatalytic Umpolung of Bromocations: An Accelerated Stereoselective Dibromination Protocol," *Chemistry – A European Journal* 29 (2023): e202300675, https://doi.org/10.1002/chem.202300675.

42. V. Pace, L. Castoldi, and W. Holzer, "Synthesis of α,β-Unsaturated α′-Haloketones Through the Chemoselective Addition of Halomethyllithiums to Weinreb Amides," *Journal of Organic Chemistry* 78 (2013): 7764–7770, https://doi.org/10.1021/jo401236t.

43. T. M. Lipon, I. D. Marpna, K. Wanniang, et al., "Selenium Dioxide-Mediated Bromination of α,β-Unsaturated Ketones Using *N*-Bromosuccinimide in the Presence of *p*-Toluenesulfonic Acid: A Versatile Route for the Synthesis of α′-Bromo-4-arylbut-3-en-2-one and α′,α′-Dibromo-4-arylbut-3-en-2-one," *ACS Omega* 6 (2021): 27466–27477, https://doi.org/10.1021/acsomega.1c04352.

44. K. A. Leets, A. Pacherille, and J. D. Chisholm, "Tandem Oxidation-Bromination of Allylic Alcohols With a TEMPO-Oxone-Et_4NBr Reactant System," *Tetrahedron Letters* 103 (2022): 153994, https://doi.org/10.1016/j.tetlet.2022.153994.

45. J.-C. Xiang, J.-W. Wang, P. Yuan, J.-T. Ma, A.-X. Wu, and Z.-X. Liao, "Switching Over of the Chemoselectivity: I_2-DMSO-Enabled α,α-Dichlorination of Functionalized Methyl Ketones," *Journal of Organic Chemistry* 87 (2022): 15101–15113, https://doi.org/10.1021/acs.joc.2c01591.

46. G. Cavallo, P. Metrangolo, R. Milani, et al., "The Halogen Bond," *Chemical Reviews* 116 (2016): 2478–2601, https://doi.org/10.1021/acs.chemrev.5b00484.

47. M. A. Branch, T. F. Coleman, and Y. Li, "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems," *SIAM Journal of Scientific Computing* 21 (1999): 1, https://doi.org/10.1137/S1064827595289108.

48. J.-D. Chai and M. Head-Gordon, "Long-Range Corrected Hybrid Density Functionals With Damped Atom–atom Dispersion Corrections,"

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Physical Chemistry Chemical Physics 10 (2008): 6615–6620, https://doi.org/10.1039/b810189b.

49. F. Weigend and R. Ahlrichs, "Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy," *Physical Chemistry Chemical Physics* 7 (2005): 3297–3305, https://doi.org/10.1039/b508541a.

50. F. Weigend, "Accurate Coulomb-fitting Basis Sets for H to Rn," *Physical Chemistry Chemical Physics* 8 (2006): 1057–1065, https://doi.org/10.1039/b515623h.

51. A. Hellweg, C. Hättig, S. Höfener, and W. Klopper, "Optimized Accurate Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations for the Atoms Rb to Rn," *Theoretical Chemistry Accounts* 117 (2007): 587–597, https://doi.org/10.1007/s00214-007-0250-5.

52. A. Hellweg and D. Rappoport, "Development of New Auxiliary Basis Functions of the Karlsruhe Segmented Contracted Basis Sets Including Diffuse Basis Functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC Calculations," *Physical Chemistry Chemical Physics* 17 (2015): 1010–1017, https://doi.org/10.1039/C4CP04286G.

53. V. Barone and M. Cossi, "Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model," *Journal of Physical Chemistry A* 102 (1998): 1995.

54. S. Grimme, "Exploration of Chemical Compound, Conformer, and Reaction Space With Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations," *Journal of Chemical Theory and Computation* 15 (2019): 2847–2862, https://doi.org/10.1021/acs.jctc.9b00143.

55. P. Pracht, F. Bohle, and S. Grimme, "Automated Exploration of the Low-Energy Chemical Space With Fast Quantum Chemical Methods," *Physical Chemistry Chemical Physics* 22 (2020): 7169–7192, https://doi.org/10.1039/C9CP06869D.

56. P. Pracht, S. Grimme, C. Bannwarth, et al., "CREST—A Program for the Exploration of Low-Energy Molecular Chemical Space," *Journal of Chemical Physics* 160 (2024): 114110, https://doi.org/10.1063/5.0197592.

57. C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method With Multipole Electrostatics and Density-Dependent Dispersion Contributions," *Journal of Chemical Theory and Computation* 15 (2019): 1652–1671, https://doi.org/10.1021/acs.jctc.8b01176.

58. S. Ehlert, M. Stahn, S. Spicher, and S. Grimme, "Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods," *Journal of Chemical Theory and Computation* 17 (2021): 4250–4261, https://doi.org/10.1021/acs.jctc.1c00471.

59. J. Baker, "An Algorithm for the Location of Transition States," *Journal of Computational Chemistry* 7 (1986): 385–395, https://doi.org/10.1002/jcc.540070402.

60. H. Horn, H. Weiß, M. Háser, M. Ehrig, and R. Ahlrichs, "Prescreening of Two-Electron Integral Derivatives in SCF Gradient and Hessian Calculations," *Journal of Computational Chemistry* 12 (1991): 1058–1064, https://doi.org/10.1002/jcc.540120903.

61. H. B. Schlegel, "Geometry Optimization on Potential Energy Surfaces," in *Modern Electronic Structure Theory*, ed. D. R. Yarkony (World Scientific, 1995): 459–500, https://doi.org/10.1142/9789812832108_0008.

62. F. Eckert, P. Pulay, and H.-J. Werner, "Ab Initio Geometry Optimization for Large Molecules," *Journal of Computational Chemistry* 18 (1997): 1473–1483, https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1473::AID-JCC5%3e3.0.CO;2-G.

63. T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch, "Optimization of the r^{2}SCAN-3c Composite Electronic-Structure Method for Use With Slater-Type Orbital Basis Sets," *Journal of Physical Chemistry A* 126 (2022): 3826–3838, https://doi.org/10.1021/acs.jpca.2c02951.

64. M. Müller, A. Hansen, and S. Grimme, "ωB97X-3c: A Composite Range-Separated Hybrid DFT Method With a Molecule-Optimized Polarized Valence Double-ζ Basis Set," *Journal of Chemical Physics* 158 (2023): 014103..

65. G. Santra, N. Sylvetsky, and J. M. L. Martin, "Minimally Empirical Double-Hybrid Functionals Trained Against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4," *Journal of Physical Chemistry A* 123 (2019): 5129–5143, https://doi.org/10.1021/acs.jpca.9b03157.

66. S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu," *Journal of Chemical Physics* 132 (2010): 154104, https://doi.org/10.1063/1.3382344.

67. F. Neese, "An Improvement of the Resolution of the Identity Approximation for the Formation of the Coulomb Matrix," *Journal of Computational Chemistry* 24 (2003): 1740–1747, https://doi.org/10.1002/jcc.10318.

68. S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the Damping Function in Dispersion Corrected Density Functional Theory," *Journal of Computational Chemistry* 32 (2011): 1456–1465, https://doi.org/10.1002/jcc.21759.

69. D. Bykov, T. Petrenko, R. Izsák, et al., "Efficient Implementation of the Analytic Second Derivatives of Hartree–Fock and Hybrid DFT Energies: A Detailed Analysis of Different Approximations," *Molecular Physics* 113 (2015): 1961, https://doi.org/10.1080/00268976.2015.1025114.

70. F. Neese, "Software Update: The ORCA Program System—Version 5.0," *WIREs Computational Molecular Science* 12 (2022): e1606, https://doi.org/10.1002/wcms.1606.

71. F. Neese, "The SHARK Integral Generation and Digestion System," *Journal of Computational Chemistry* 44 (2023): 381–396, https://doi.org/10.1002/jcc.26942.

72. D. S. Tikhonov, I. Gordiy, D. A. Iakovlev, et al., "Harmonic Scale Factors of Fundamental Transitions for Dispersion-Corrected Quantum Chemical Methods," *ChemPhysChem* 25 (2024): e202400547, https://doi.org/10.1002/cphc.202400547.

73. T. Gerlach, S. Müller, A. G. De Castilla, and I. Smirnova, "An Open Source COSMO-RS Implementation and Parameterization Supporting the Efficient Implementation of Multiple Segment Descriptors," *Fluid Phase Equilibria* 560 (2022): 113472, https://doi.org/10.1016/j.fluid.2022.113472.

74. S. Müller, T. Nevolianis, M. Garcia-Ratés, C. Riplinger, K. Leonhard, and I. Smirnova, "Predicting Solvation Free Energies for Neutral Molecules in Any Solvent With Open COSMO-RS," *Fluid Phase Equilibria* 589 (2025): 114250.

75. A. Klamt, "Conductor-Like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena," *Journal of Physical Chemistry* 99 (1995): 2224–2235, https://doi.org/10.1021/j100007a062.

76. A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz, "Refinement and Parametrization of COSMO-RS," *Journal of Physical Chemistry A* 102 (1998): 5074–5085, https://doi.org/10.1021/jp980017s.

77. K. K. Irikura and D. J. Frurip, eds., *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics* (American Chemical Society, 1998).

78. P. Pracht and S. Grimme, "Calculation of Absolute Molecular Entropies and Heat Capacities Made Simple," *Chemical Science* 12 (2021): 6551–6568, https://doi.org/10.1039/D1SC00621E.

79. G. Mills, H. Jónsson, and G. K. Schenter, "Reversible Work Transition State Theory: Application to Dissociative Adsorption of Hydrogen," *Surface Science* 324 (1995): 305–337, https://doi.org/10.1016/0039-6028(94)00731-4.

80. H. Jónsson, G. Mills, and K. W. Jacobsen, "Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions," in *Classical and Quantum Dynamics in Condensed Phase Simulations*, eds. B. J. Berne,

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

G. Ciccotti, and D. F. Coker (World Scientific Publishing Co Pte Ltd, 1998), 385–404, https://doi.org/10.1142/9789812839664_0016.

81. G. Henkelman and H. Jónsson, "Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points," *Journal of Chemical Physics* 113 (2000): 9978–9985, https://doi.org/10.1063/1.1323224.

82. V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, et al., "Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined With Eigenvector Following," *Journal of Chemical Theory and Computation* 17 (2021): 4929–4945, https://doi.org/10.1021/acs.jctc.1c00462.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

The robotic setup, experimental procedures, analytical and spectral data, methods for NMR data processing, and theoretical calculations of this work are available in the Supporting Information of this article. Codes, datasets, and output of this work are deposited at Zenodo (https://doi.org/10.5281/zenodo.17080099). The authors have cited additional references within the Supporting Information [40–82].

Supporting File 1: anie73351-sup-0001-SuppMat.pdf.

Supporting File 2: anie73351-sup-0002-MovieS1.mp4.

15213773, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.9132891 by Ulsan National Institute Of, Wiley Online Library on [19/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7. Author Declarations

Ulsan- 11/09/2026

Statement of Contribution

I, Yasemin Bilgi-Gadina, 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.; Bass, 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. https://doi.org/10.1038/s41467-024-54611-5.

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

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, 645 (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1

Consisted of:

- Optimization and adaptation to the robotic platform of fluorenol E1, S_N1 reactions

Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**, e9132891. https://doi.org/10.1002/anie.9132891

Consisted of:

- Investigation, methodology, experimental validation, software, data curation, formal analysis, writing the original draft.

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W. D.; Beker, W.; Saillard, E.; Łęgowski, K.; Roszak, R.; Kowalczyk, P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Centuryof Total Syntheses as a Machine-ReadableResource for Analyzing Strategic Logic in Organic Chemistry. J. Am. Chem. Soc. 2026. https://doi.org/10.1021/jacs.6c08434

Consisted of:

- Data curation, quality control, team management, writing the SI.

Bilgi-Gadina Y., Louis P. J.-L. Gadina, Rafal Frydrych, Bartosz A. Grzybowski* Automated NMRcraft Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures *J. Am. Chem. Soc.* ***Submitted. 2026. In revision.***

Consisted of:

- Designing and performing the experiments, characterizations, writing the paper and SI.

Yasemin Bilgi-Gadina



Ulsan- 11/09/2026

Statement of Contribution

I, Prof. Bartosz A. Grzybowski, 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.; Bass, 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. https://doi.org/10.1038/s41467-024-54611-5.

Consisted of:

- Designing and developing Allchemy platform
- Supervising the project
- Writing the paper with help from other authors

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, 645 (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1

Consisted of:

- Supervising the project
- Writing the paper with help from other authors

Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. Angew. Chem. Int. Ed. **2026**, e9132891. https://doi.org/10.1002/anie.9132891

Consisted of:

- Supervising the project
- Writing the paper with help from other authors

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W. D.; Beker, W.; Saillard, E.; Łęgowski, K.; Roszak, R.; Kowalczyk, P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Centuryof Total Syntheses as a Machine-ReadableResource for Analyzing Strategic Logic in Organic Chemistry. J. Am. Chem. Soc. **2026**. https://doi.org/10.1021/jacs.6c08434

Consisted of:

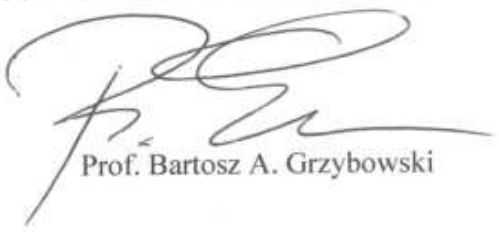
- Supervising the project
- Writing the paper with help from other authors

Bilgi-Gadina Y., Louis P. J.-L. Gadina, Rafal Frydrych, Bartosz A. Grzybowski* Automated NMRcraft Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures *J. Am. Chem. Soc.* ***Submitted***. **2026**. ***In revision.***

Consisted of:

- Supervising the project
- Writing the paper with help from other authors

I hereby confirm the abovementioned contributions



Prof. Bartosz A. Grzybowski

Institute of Organic Chemistry
Polish Academy of Sciences

Professor Daniel Gryko
Director

+48 22 343 23 21
daniel.gryko@icho.edu.pl

Warsaw 4th August 2026.

Statement of Contribution

I, Prof. Daniel T. Gryko, 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:

- Helping with the evaluation of the kinetic networks.

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributed to student recruitment and funding.



Yours sincerely

Daniel Gryko

Warszawa- 30/07/2026

Statement of Contribution

I, Prof. Jacek Młynarski, 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:

- Supervising the syntheses described in Supplementary Section S2, performed by S. Baś.



Prof. Jacek Młynarski

Giessen- 03/08/2026

Statement of Contribution

I, Prof. Peter R. Schreiner, declare that my contribution in the following articles:

Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**.

Consisted of:

- Funding acquisition, writing - review and editing, project administration, resources, supervision.

Prof. Peter R. Schreiner

Signature:

Warsaw- 31/07/2026

Statement of Contribution

I, Ahmad Makkawi, hereby 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:

- Performing the syntheses described in 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

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributed to data curation and analysis.

Ahmad Makkawi



Warszawa- 03/08/2026

Statement of Contribution

I, Anna Żądło-Dobrowolska, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Curated the synthesis data.

Anna Żądło-Dobrowolska

Anna Żądło-Dobrowolska

Signature:

Warsaw- 30/07/2026

Statement of Contribution

I, Barbara Mikulak-Klucznik, 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 syntheses described in Fig. 4b in the main article

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributing to data curation and analysis.

Barbara Mikulak-Klucznik

Signature: Barbara Mikulak-Klucznik

Giessen- 01/08/2026

Statement of Contribution

I, Benito Marazzi, declare that my contribution in the following articles:

Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**.

Consisted of:

- Data curation, investigation.

Benito Marazzi



Signature:

Ulsan- 08/09/2026

Statement of Contribution

I, Bibek Prajapati, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Conducted experiments.
- Performed all Hantzsch reactions using the robotic platform.
- Carried out partial HPLC purification of the Hantzsch reaction products.
- Performed one reaction for the versatility test.

Bibek Prajapati

Signature:



Ulsan- 08/09/2026

Statement of Contribution

I, Daniel Matuszczyk, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Contributed to the robotic synthesis of the Hantzsch reaction.
- Separated crude reaction mixtures.
- Purified the products.
- Performed NMR analysis of the isolated compounds.
- Performed one reaction for the versatility test.

Daniel Matuszczyk

Signature:



Kraków- 30/07/2026

Statement of Contribution

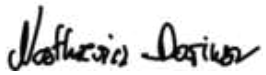
I, Dariusz Nostkiewicz, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Developed the AbSynth WebApp

Dariusz Nostkiewicz



Warszawa- 21/09/2026

Statement of Contribution

I, Michał Michalak, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026**, *148*, 38740–38753.

Consisted of:

- Curated the synthesis data.

Michał Michalak

Signature:

Ulsan- 08/09/2026

Statement of Contribution

I, Eric S. Larsen, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Conducted synthesis for the rearrangement reaction of anthracen-9-yldiphenylmethanol.
- Assisted with the analysis and structural determination of the rearrangement products.
- Analysed several sets of single-crystal X-ray diffraction data to determine their three-dimensional structures.

Eric S. Larsen

Signature:

Kraków- 30/07/2026

Statement of Contribution

I, Gabriela Smaga hereby declare that my contribution in the following article:

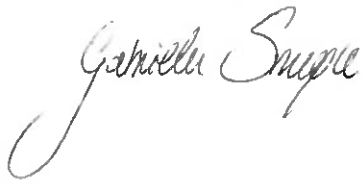
Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Gabriela Smaga

Signature:



Ulsan- 08/09/2026

Statement of Contribution

I, Galymzhan Moldagulov, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Conducted the reaction versatility and the system accuracy test.

Galymzhan Moldagulov

Signature:

PODPIS ZAUFANY
JAKUB
ZABIELSKI
30.07.2026 21:36:16 GMT+0200
Dokument podpisany elektronicznie
podpisem zaufanym

Kraków- 00/00/2026

Statement of Contribution

I, Jakub Zabielski, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Developed the AbSynth WebApp

Jakub Zabielski

Signature:

Kraków- 30/07/2026

Statement of Contribution

I, Jakub Skoczek, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Jakub Skoczek

Signature:



Barcelona- 30/07/2026

Statement of Contribution

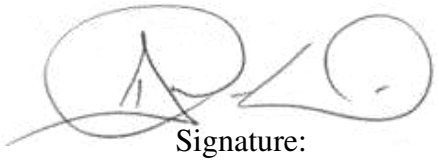
I, Juan Carlos Ahumada, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Performed synthesis for the PBA-catalyzed reactions and isolated the products.
- Determined product structures for the PBA-catalyzed reactions.
- Contributed to mechanistic analysis of the PBA-catalyzed reactions.
- Performed and analyzed the Suzuki–Miyaura coupling experiments.

Juan Carlos Ahumada



Signature:

Warsaw- 08/08/2026

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.; 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
- Contributing in writing some parts of the SI and the main text

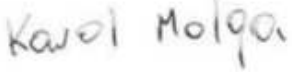
Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Performed statistical analyses
- Wrote computer codes
- Contributed to data curation.

Karol Molga

Signature:



Ulsan- 00/00/2026

Statement of Contribution

I, King-Lam Kwong, declare that my contribution in the following articles:

Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**.

Consisted of:

- Software, data curation, investigation.

King-Lam Kwong



Signature:

PODPIS ZAUFANY
KONRAD
BARNOWSKI
03.08.2026 09:54:34 GMT+0200
Dokument podpisany elektronicznie
podpisem zaufanym

Kraków- 08/03/2026

Statement of Contribution

I, Konrad Barnowski, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Konrad Barnowski

The document was e-signed.

Kraków- 30/07/2026

Statement of Contribution

I, Konrad Łęgowski, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Konrad Łęgowski

Signature:

Ulsan- 11/09/2026

Statement of Contribution

I, Louis P.J-L. Gadina, 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.; Bass, 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. https://doi.org/10.1038/s41467-024-54611-5.

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 and performing the analysis
- Contributing in writing parts of the SI and the main text

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, 645 (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1

Consisted of:

- Proposed the reaction for the PBA-catalysed reactions.
- Performed the proof-of-concept test for the PBA-catalysed reactions.
- Determined the product structures and reaction mechanism for the PBA-catalysed reactions.
- Wrote the manual for the HyperspaceViewer software.

Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**, e9132891. https://doi.org/10.1002/anie.9132891

Consisted of:

- Designed and implemented the NMR spectrum analysis code and contributed to data analysis.

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W. D.; Beker, W.; Saillard, E.; Łęgowski, K.; Roszak, R.; Kowalczyk, P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Centuryof Total Syntheses as a Machine-ReadableResource for Analyzing Strategic Logic in Organic Chemistry. J. Am. Chem. Soc. 2026. https://doi.org/10.1021/jacs.6c08434

Consisted of:

- Contributed to data curation and analysis.

Bilgi-Gadina Y., Louis P. J.-L. Gadina, Rafal Frydrych, Bartosz A. Grzybowski* Automated NMRcraft Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures *J. Am. Chem. Soc.* ***Submitted. 2026. In revision.***

Consisted of:

- Developed automated NMR quantification framework.

Louis P.J-L. Gadina



Kraków- 30/07/2026

Statement of Contribution

I, Maciej Czuba, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Maciej Czuba

Signature: 

Kraków- 01/08/2026

Statement of Contribution

I, Maja Łazuchiewicz, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Maja Łazuchiewicz

Signature:

Maja Łazuchiewicz

Scanned with CamScanner

Warsawa- 10/08/2026

Statement of Contribution

I, Martyna Moskal, 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 Allchemy platform
- Performing analyses and calculations described in the paper

Martyna Moskal

Signature:

Martyna Moskal

Châteauroux- 31/07/2026

Statement of Contribution

I, Maxence Deschamps, declare that my contribution in the following articles:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Developed the HyperspaceViewer software.

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Maxence Deschamps

Signature:



Warsaw, 30/07/2026

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.; 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:

- Helping with the evaluation of the kinetic networks, along with Prof. D. T. Gryko.

Olena Vakuliuk

Signature:



Kraków- 31/07/2026

Statement of Contribution

I, Patrycja Pietruszewska, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Patrycja Pietruszewska

Pietruszewska

Kraków- 31/07/2026

Statement of Contribution

I, Piotr Janasz, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Piotr Janasz

Piotr Janasz

Ulsan- 15/09/2026

Statement of Contribution

I. Rafał Frydrych. 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:

- Helping in the syntheses described in Supplementary Section S2 with supervision from Prof. B. A. Grzybowski.

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Contributed to the automation design.
- Led the chemical component of the project.
- Selected, performed, and interpreted all reactions conducted using the robotic platform.
- Performed MS and MS/MS mechanistic studies.
- Conducted experiments together with Y. Jia, Y.I.S., and B.P.
- Contributed to data analysis.
- Co-wrote the initial draft with Y. Jia

Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Frydrych, R.; Grzybowski, B. A. Automated NMRcraft Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures. *J. Am. Chem. Soc. Submitt.* **2026**.

Consisted of:

- Developed the HPLC method and assisted with the acquisition of ground truths by HPLC.

Rafał Frydrych

Signature:

Kraków- 01/08/2026

Statement of Contribution

I, Saillard Evan, hereby declare that my contribution in the following article:

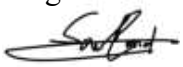
Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.

Saillard Evan

Signature:



Warsaw- 21/09/2026

Statement of Contribution

I, Sara Szymkuć, 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
- Contributing in writing some parts in the main article

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributed to student recruitment and funding.

Sara Szymkuć

Signature:

SARA ANNA SZYMKUĆ
2026.09.25 15:06:33 +02:00

Kraków- 30/07/2026

Statement of Contribution

I, Sebastian Baś, hereby declare that my contribution to 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:

- Performing synthesis described in Supplementary Section S2 with supervision from Prof. J. Mlynarski.

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributed to data curation and analysis.
- Contributed to student recruitment and coordinated work among student groups, as well as to the management and verification of the web app preparation

Your sincerely,

Sebastian Baś



Signature:

Kraków- 31/07/2026

Statement of Contribution

I, Szymon Stodółkiewicz, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Szymon Stodółkiewicz



Ulsan- 08/09/2026

Statement of Contribution

I, Wai-Shing Wong, declare that my contribution in the following articles:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Provided input for the numerical models of reaction kinetics.
- Performed four reactions for the versatility test.
- Assigned structures to unexpected Hantzsch reaction products.
- Proposed the corresponding reaction mechanisms.

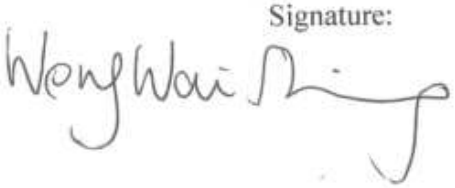
Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. Angew. Chem. Int. Ed. 2026, e9132891. https://doi.org/10.1002/anie.9132891

Consisted of:

- Investigation, methodology, validation, software, data curation, formal analysis, writing – original draft.

Wai-Shing Wong

Signature:



Warsaw- 07/30/2026

Statement of Contribution

I, Wiktor Beker, declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Performed statistical analyses
- Wrote computer codes
- Contributed to data curation.

Wiktor Beker

Signature:



Warszawa- 30/07/2026

Statement of Contribution

I, Wojciech Petrykowski, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Wojciech Petrykowski

Wojciech Petrykowski

Nanjing, China - 31/07/2026

Statement of Contribution

I, Yankai Jia, declare that my contribution in the following articles:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Constructed the automation hardware and designed the automation software.
- Conducted experiments.
- Performed data analysis.
- Wrote the initial draft.

Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**.

Consisted of:

- Developed Automated benchtop-NMR and the automation software.
- Investigation, methodology, validation, data curation, formal analysis, writing – original draft.

Yankai Jia

Signature: 

Ulsan- 08/09/2026

Statement of Contribution

I, Yanqiu Jiang, declare that my contribution in the following articles:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Synthesized the PBA catalysts.
- Performed reactions using the robotic platform.
- Conducted data analysis.
- Wrote the initial draft of the PBA-related section.

Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. Angew. Chem. Int. Ed. 2026, e9132891. https://doi.org/10.1002/anie.9132891

Consisted of:

- Software, data curation, investigation.

Yanqiu Jiang

Signature:

Ulsan- 08/09/2026

Statement of Contribution

I, Yaroslav I. Sobolev, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Constructed the automation hardware and designed the automation software.
- Conducted experiments.
- Designed the software for spectral unmixing.
- Developed numerical models of reaction kinetics.
- Fitted the kinetic models to the experimental data.
- Performed TD-DFT calculations of absorption spectra.
- Performed quantum thermochemistry calculations.
- Conducted the theoretical study of the smoothness of yield maps.

Yaroslav I. Sobolev



Signature:

Kraków- 30/07/2026

Statement of Contribution

I, Zofia Pławecka, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Zofia Pławecka

Signature:

Zofia Pławecka

Kraków- 30/07/2026

Statement of Contribution

I, Emilia Mielke, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Emilia Mielke

Emilia Mielke

Ulsan- 08/08/2026

Statement of Contribution

I, Namhun Kim, declare that my contribution in the following article:

Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G.; Kim, N.; Larsen, E. S.; Deschamps, M.; Jiang, Y.; Grzybowski, B. A. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

Consisted of:

- Analysed single-crystal X-ray diffraction data.

Namhun Kim

Signature:

Giessen- 03/08/2026

Statement of Contribution

I, Nikolaos Vagkidis, declare that my contribution in the following articles:

Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**.

Consisted of:

- Data curation, investigation.

Nikolaos Vagkidis

Signature:

Huskvarna - 02/08/2026

Statement of Contribution

I, Oskar Popik, 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 Fig. 4a in the article, under the supervision of Prof. J. Młynarski
- Contributing to the writing of the Supplementary Information section related to my synthesis

Oskar Popik

Signature:



Warszawa, 11.08.2026 r.

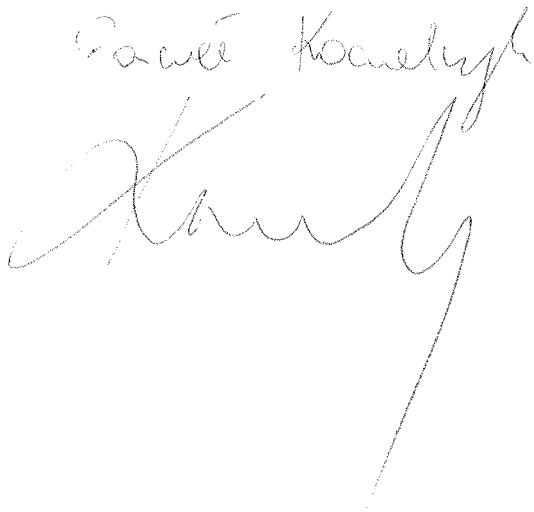
Statement of Contribution

I, Paweł Kowalczyk, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Mołga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.D.; Beker W.; Saillard E.; Łęgowski K.; Roszak R.; Kowalczyk P.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* ***2026.***

Consisted of:

- Curated the synthesis data.



Warsaw, 2nd September 2026

Statement of Contribution

I, Rafał 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

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Performed statistical analyses
- Wrote computer codes
- Contributed to data curation.

Rafał Roszak

Signature:

Kraków- 03/08/2026

Statement of Contribution

I, Sylwia Wójcik, hereby declare that my contribution in the following article:

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.*

Consisted of:

- Curated the synthesis data.

Sylwia Wójcik

Signature:

Sylwia Wójcik

Warszawa- 11/08/2026

Statement of Contribution

I, Tomasz Klucznik, 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 syntheses described in Fig. 4b in the main article

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Contributed to data curation and analysis.

Tomasz Klucznik

Signature:

Warszawa- 30/07/2026

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.; 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
- Contributing in writing some parts of the SI and the main text

Czuba, M.; Barnowski, K.; Molga, K.; Pławecka, Z.; Wołos, A.; Stodółkiewicz, S.; Baś, S.; Petrykowski, W.; Mikulak-Klucznik, B.; Janasz, P.; Mielke, E.; Pietruszewska, P.; Łazuchiewicz, M.; Wójcik, S.; Zabielski, J.; Nostkiewicz, D.; Smaga, G.; Makkawi, A.; Gadina, L. P. J.-L.; Skoczek, J.; Klucznik, T.; Deschamps, M.; Żądło-Dobrowolska, A.; Michalak, M.; Szymkuć, S.; Gryko, D. T.; Bilgi-Gadina, Y.; Grzybowski, B. A. AbSynth: A Century of Total Syntheses as a Machine Readable Resource for Analyzing Strategic Logic in Organic Chemistry. *J. Am. Chem. Soc.* **2026.**

Consisted of:

- Performed statistical analyses
- Wrote computer codes
- Contributed to data curation.

Agnieszka Wołos

Signature:

Agnieszka Wołos

Ulsan- 29/09/2026

Statement of Contribution

I, Yasemin Bilgi-Gadina, declare the contribution of Patrycja Gołębiowska to the following article on her behalf. Patrycja Gołębiowska left the group several years ago, and I have since lost contact with her. I attempted to reach her using the email address available to me but did not receive a response. It is possible that her contact details have changed, and I am not aware of any other means of contacting her.

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 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



Yasemin Bilgi-Gadina

8. Additional Materials

8.1.Chan–Lam Coupling Reaction Hyperspace and Automated NMR Quantification

Automated NMRanger Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures

Yasemin Bilgi-Gadina[1,2], Louis P. J.-L. Gadina[1,2], Rafal Frydrych[1], Bartosz A. Grzybowski[1,3]*

[1]Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan, 44919, Republic of Korea. [2]Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, 01-224, Poland. [3]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea.

KEYWORDS: NMRanger, automated NMR analysis, Chan-Lam coupling, Reaction hyperspaces, Robotics

ABSTRACT: Quantitative analysis of crude reaction mixtures remains a major bottleneck in high throughput and hyperspace-level chemical discovery, where overlapping, shifting, and broadened NMR signals hinder reliable product and by-product quantification. Here, we introduce NMRanger, a region selective NMR fitting framework that deconvolves complex ^{1}H NMR spectra into contributions from a basis set of reference compounds while tolerating peak broadening, chemical shift drift, and multiplet distortion. Integrated with the Robowski-NMR automation platform, NMRanger enables continuous acquisition and automated analysis of crude mixtures across large reaction arrays. We apply this workflow to map 1591 conditions of the Chan–Lam coupling, reconstructing hyperspace-wide distributions of coupling products, by-products, and unconsumed reagents. These maps reveal how atmosphere, temperature, base loading, substrate concentration, and copper stoichiometry jointly shape the balance between C–N coupling and side reactions. Spot check validation against GPC isolation, HPLC analysis or manual integration based on internal standard confirms that NMRanger yields track experimental trends and provide quantitatively useful estimates. In doing so, NMRanger makes quantitative crude-mixture NMR analysis feasible at hyperspace scale, enabling reaction-landscape and condition-space reconstructions that were previously inaccessible with conventional NMR workflows. To support broad adoption, we also disclose the full NMRanger source code.

INTRODUCTION

Modern chemical automation systems now allow the execution of hundreds to thousands of reactions per day, supporting tasks that range from multistep syntheses,[1–9] to large-scale reaction optimization,[10–15] to systematic exploration of multidimensions "hyperspaces" of conditions[16–20]. However, as preparation and execution of reactions become faster and more reproducible, a new bottleneck has emerged: the *quantitative* analysis of complex reaction mixtures, particularly crucial in discovery-oriented tasks.[16–27] Traditional approaches such as product isolation are too slow, sample consuming, and costly to be practical for routine analysis of large reaction arrays. Likewise, manual interpretation of spectroscopic data becomes altogether impractical at the scales enabled by automation. We have encountered this bottleneck acutely in hyperspace scans aimed at discovering conditions that support new patterns of reactivity[18–20] or at reconstructing hyperspace wide distributions of both products and by-products[16,17,20] from which mechanistic networks can be inferred.[16–20] To date, we have worked around this problem using unmixing of UV-vis spectra[16–19] (limited to UV-vis active species) or semi-automated NMR analyses of characteristic signals.[20] This characteristic-peak approach (**Figure 1a**) works well when signals from each substance are clearly resolved but becomes inadequate when multiplets not only overlap but also shift and broaden from sample to sample due to concentration effects, acid-base interactions, sample heterogeneity, or metal-containing species. Reference-guided and peak-picking-free methods for NMR quantification offer partial solutions, but they are not designed for reaction-specific crude mixture quantification (see ref [28] and later in the text). Accordingly, to mitigate these limitations, we introduce here NMRanger, a region-selective NMR-fitting framework for quantitative reaction analysis that unmixes complex NMR spectra into the "basis set" of components while tolerating peak broadening and chemical-shift variation (**Figure 1b**). We use this algorithm in conjunction with the Robowski-NMR platform to explore the hyperspace of the Chan-Lam coupling, broadly used in natural-product, medicinal, pharmaceutical, and late-stage functionalization chemistry.[29–35] Mapping this reaction across a multidimensional condition space constitutes the first application of the hyperspace concept to organometallic catalysis**,** in a setting where crude-mixture complexity and strong condition dependence place high demands on quantitative NMR analysis. These demands arise because the distributions of products and by-products – and therefore the crude NMR spectra – are highly sensitive to numerous reaction parameters, including the seven "dimensions" we probe (atmosphere, temperature, reaction time, phenylboronic acid equivalents, base equivalents, substrate concentration, and copper loading) across four substrates (morpholine, 7-azaindole, 5-bromo-7-azaindole, and 1H-indazole). Deconvolution of these spectra is particularly non-trivial because aromatic substrates, products, and by-products require NMR unmixing in multiplet-rich regions while concentration effects, acid–base interactions, sample heterogeneity, non-deuterated solvent, and copper species further broaden resonances and shift peaks through coordination and paramagnetic effects.

Despite these difficulties, however, our analyses across the hyperspace of 1591 distinct reaction conditions not only confirm previously established regions of maximal yield for the desired coupling products[36–39] but also reveal how boronic-acid-derived by-products (e.g., benzene, phenol, and diaryl ether) and regioisomeric distributions are organized throughout the hyperspace. Together, these results show that NMRanger enables rapid, automated quantification of complex crude mixtures – far faster than manual analysis and beyond the reach of existing NMR software – with clear potential for broader application to diverse reaction classes.

RESULTS AND DISCUSSION

1. Algorithm details. The key algorithmic objective of this work is to automate interpretation of ^{1}H NMR spectra of crude mixtures obtained when scanning a given reaction under a large number of conditions, be it during a reaction optimization campaign[21,22,25] or when mapping out this reaction's complete hyperspace[16–20]. Comprehensive studies of a broad range of transformations – from foundational brominations[20] to complex multicomponent reactions[16,17,19]– indicate that a hyperspace of a given reaction houses from a few[16] to teens [17–20] of products forming in appreciable yields. Upon isolation, these products form a "basis set" of spectral signatures, $S_i(\delta_{\mathrm{ppm}})$, with which we aim to fit as closely as possible the crude spectra $S_{\mathrm{crude}}(\delta_{\mathrm{ppm}})$ of all hyperspace conditions and thus reconstruct the hyperspace-wide distributions of all products and by-products so that $S_{\mathrm{crude}}(\delta_{\mathrm{ppm}}) \approx \sum_i A_i S_i$, where A_i is the abundance of an analyte i. As mentioned in the introduction, the difficulty in this kind of fitting is that signals from S_i components may not only overlap but also shift and broaden. In principle, there exist several automated and semi-automated NMR quantification methods, developed mainly in metabolomics, to compare an experimental ^{1}H NMR spectrum with reference information from known metabolites and to estimate compound amounts from fitted or matched reference contributions. Targeted profiling methods such as Chenomx and MagMet use libraries of peak positions, intensities, linewidths, and J-couplings.[40,41] Bayesil and BATMAN use probabilistic deconvolution: Bayesil uses sequential Monte Carlo inference, whereas BATMAN uses Bayesian Lorentzian multiplet templates, MCMC sampling, and wavelets for unassigned signals.[42–44] AQuA uses one reporter signal per metabolite and a calibration matrix to correct signal interference.[45] The recent Magnetstein algorithm provides a related peak-picking-free strategy by treating ^{1}H NMR spectra as signal-intensity distributions and fitting mixtures as weighted combinations of library spectra using the Wasserstein distance, which potentially makes its analysis less dependent on predefined positions of individual signals.[28] However, when these methods were tested on the crude-mixture spectra from our reactions, their performance was poor, as described in Supplementary Information, SI, **Section 2.** Faced with the limitations of these general-purpose models, we embarked on an effort to develop an algorithm that would offer a practical compromise between peak-based and peak-picking-free methods while modeling possible distortions of components' spectra with as little supervision as possible. This NMRanger framework natively tolerates moderate peak broadening and chemical-shift variation. Corner cases, including contaminated references, unusually pronounced peak shifts, or weak signals, can be addressed by targeted edits to the basis set, allowing specific challenges associated with crude reaction mixture analysis to be handled directly.

In brief, NMRanger uses region-based modeling to enable knowledge transfer from narrow to broader spectral regions. The user first defines one or more spectral regions to be processed in order (see User Manual in the SI, **Section 3**), each modeled as a weighted sum of analyte spectra while allowing possible spectral distortions. Narrow regions (0.5–2 ppm wide), containing well-resolved informative signals, are typically processed first because they involve fewer overlapping components and provide rapid (seconds) concentration estimates for the analytes visible in the selected window. These estimates are then transferred to broader, more congested regions (3–5 ppm wide) as concentration constraints, with limited allowed variation, for analytes that also contribute to the broader-region fit. Because broader regions generally involve more components and stronger overlap, the corresponding inverse problem is more computationally demanding (2–20 min per region on a single CPU core). By transferring concentration information from narrow to broader regions, this knowledge-transfer strategy restricts the space of possible compositions and thereby alleviates underdetermination in congested spectral domains.

Once regions are defined, we first build a model for a spectrum $S_{n,i}(\delta_{\mathrm{ppm}})$ of i-th pure reference compound in the n-th region, see **Figure 1c**.

$$S_{n,i}(\delta_{\mathrm{ppm}}) = \sum_{c_{i,j} \in r_n} V(\delta_{\mathrm{ppm}}, a_{i,j}, c_{i,j}, w_{i,j}, p_{i,j}) \qquad (1)$$

where $S_{n,i}$ is the spectral intensity, δ_{ppm} is the chemical shift, $V(\delta_{\mathrm{ppm}}, a_{i,j}, c_{i,j}, w_{i,j}, p_{i,j})$ is the parametrized profile of a single NMR peak, and summation is performed only over j-th peaks whose center $c_{i,j}$ belongs to the n-th region r_n. Spectral lineshapes in NMR are defined by both the exponential relaxation, which yields a purely Lorentzian lineshape, and sources of additional broadening that include residual spatial inhomogeneity of magnetic field (imperfect shimming), variations of bulk magnetic susceptibility, and unresolved multiplets – all of which result in a convolution of a basic Lorentzian profile with a Gaussian profile. Result of this convolution is, by definition, a Voigt function. A standard approach to alleviating the computation cost of Voigt function with negligible loss of accuracy is to approximate it with pseudo-Voigt profile[46–48]:

$$V(\delta_{\mathrm{ppm}}, a, c, w, p) = p\frac{a}{w\sqrt{2\pi}}e^{-\frac{(\delta_{\mathrm{ppm}}-c)^2}{2w^2}} + (1-p)\frac{a}{\pi}\frac{w\sqrt{2\ln 2}}{(\delta_{\mathrm{ppm}}-c)^2 + w^2 2\ln 2} \quad (2)$$

where a is the area of the peak function, w the width, c the center, and p is the weight of Gaussian peak function (**Figure 1d,e**). Note that $\int_{-\infty}^{\infty} V(\delta_{\mathrm{ppm}}, a, c, w, p)\, d\delta_{\mathrm{ppm}} = a$ irrespective of values of c, w, and p. The consequence is that fitting $V(\delta_{\mathrm{ppm}}, a, c, w, p)$ to experimental data in a finite region yields correct analytical peak area a with infinite integration bounds, irrespective of the width of the region in which $V(\delta_{\mathrm{ppm}}, a, c, w, p)$ was fitted. The number of peaks and initial guesses of their centers c_j are determined by SciPy's find_peaks method[49], which identifies local maxima whose height exceeds five standard deviations of the instrumental noise of the NMR spectrum (determined from peak-free regions at both ends of the spectrum, > 36 ppm and < −26 ppm in our case).

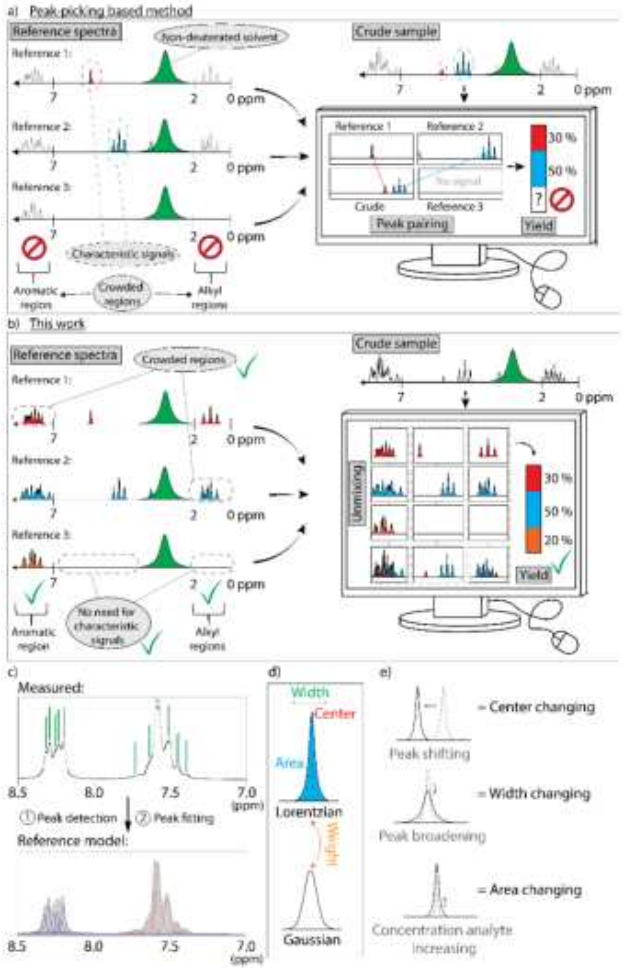


Figure 1. Conventional peak-picking vs. NMRanger algorithms. a) In peak-picking approaches (e.g., the algorithm we described in ref [20]), reference spectra are stored in a library, and isolated characteristic peaks from each reference are paired with corresponding peaks in the crude sample. The crude composition is then calculated from these matched characteristic signals. This approach works well as long as characteristic signals are cleanly resolved (e.g., in reference spectra #1 and #2 here) but fails when there is no characteristic signal without significant overlap (e.g., ref. spectrum #3 has only signals in the crowded, aromatic region). **b)** NMRanger selected-region fitting approach. Over user-specified regions, reference spectra are converted into editable models based on the approximation of the peaks by pseudo-Voigt functions (see **c**). The crude spectra are fitted and quantified by unmixing into individual component contributions, using the reference models as spectral templates. **c)** Scheme illustrating conversion of a "crowded" region of the experimental 1H NMR spectrum of Phenylboronic acid into a functional model described as a sum of fitted peaks. Detected peaks, indicated by vertical green lines in the top panel, are fitted with pseudo-Voigt functions to generate the continuous model shown in the bottom panel. Individual peak contributions are colored red or blue according to the multiplet to which they belong. The resulting substrate model is shown as the grey-filled profile. **d)** Illustration of the parameters defining each pseudo-Voigt peak model, including amplitude, center, width, and Gaussian/Lorentzian weighting. **e)** Representative effects of changing pseudo-Voigt parameters during fitting, including peak shifting through center changes, peak broadening through width and Gaussian/Lorentzian weighting changes, and concentration changes through amplitude scaling.

Parameters $a_{i,j}$, $c_{i,j}$, $w_{i,j}$, $p_{i,j}$ for $c_{i,j} \in r_n$, are optimized by minimizing the squared deviation between the modelled spectrum $S_{n,i}(\delta_{\mathrm{ppm}})$ and the experimental $S_{i,ref}(\delta_{\mathrm{ppm}})$ within the spectral range r_n:

$$\int_{\delta_{\mathrm{ppm}} \in r_n} \left(S_{i,ref}(\delta_{\mathrm{ppm}}) - \sum_{c_{i,j} \in r_n} V(\delta_{\mathrm{ppm}}, a_{i,j}, c_{i,j}, w_{i,j}, p_{i,j}) \right)^2 d\delta_{\mathrm{ppm}} \quad (3)$$

using the trust-region reflective algorithm.[50]

Next, multiplets are identified. A pair of peaks j_1, j_2 are assumed to belong to the same multiplet if $|c_{i,j_1} - c_{i,j_2}| < \tau$ where the threshold τ is set to a ppm equivalent of 20 Hz (0.25 ppm for 80 MHz ^{1}H resonance). We set the threshold to 20 Hz because commonly observed coupling constants typically fall within 2–16 Hz, although *trans*-olefin coupling constants can occasionally exceed this range.

Thus, established spectral models for individual reference compounds are used for analysis of crude mixture spectra as follows. The spectrum of the crude is modeled as,

$$S_{n,crude}(\delta_{\mathrm{ppm}}) = \sum_i \tilde{A}_{i,n} \sum_{c_{i,j} \in r_n} V(\delta_{\mathrm{ppm}}, a_{i,j}, c_{i,j} + \tilde{\Delta}_{i,m}, \tilde{v}_{i,m} w_{i,j}, \tilde{p}_{i,j}) \quad (4)$$

where tilde diacritic sign ˜ marks free parameters, $\tilde{A}_{i,n}$ is the scaling factor proportional to the concentration of the i-th analyte, $\tilde{v}_{i,m}$ is the broadening factor for the multiplet m, $\tilde{p}_{i,j}$ is the Gaussian weight, $\tilde{\Delta}_{i,m}$ is the drift of m-th multiplet of the i-th analyte. Here it is implied that m is the index of multiplet to which j-th peak belongs. The broadening factor $\tilde{v}_{i,m}$ uses as initial guess 1 and bounds are [0.5, 2] by default to accommodate for possible broadening of the peaks. $\tilde{p}_{i,j}$ uses $p_{i,j}$ as initial guess but can vary from 0 to 1 as the weight of the Gaussian component may vary across different mixtures. Bounds for $\tilde{A}_{i,n}$ are described below. The best-fit parameters are obtained by minimizing the sum of squares of the residuals between the model and the experimental spectrum of the crude across the range r_n, similar to expression (3).

The scaling factor $\tilde{A}_{i,n}$ in principle should be identical for all ranges n, but here we allow it to take a region-specific value so that estimates obtained in one region can be propagated as constraints in subsequent regions. Let's assume that a fitting was performed for the range r_{n-1} and it has yielded the best-fit value $\tilde{A}_{i,n-1}$. When fitting the range r_n, we should expect $\tilde{A}_{i,n-1} = \tilde{A}_{i,n}$. However, in practice, during fitting of the n-th region, we allow a constrained variation $\tilde{A}_{i,n-1} = \tilde{A}_{i,n}(1 \pm \varepsilon)$, where ε is a small value (in our study 0.05) to accommodate for experimental noise and baseline variation. Noteworthy, the results of this procedure depend on the order of processing with uncertainty propagating along this order. Accordingly, the good practice is to begin by processing regions containing better-resolved peaks first as their uncertainty is low.

Finally, for each i-th analyte, concentration is determined by

$$C_{i,n} = C_0 \tilde{A}_{i,n} \frac{N_{0,r_n}}{N_{i,r_n}} \frac{\sum_{c_{i,j} \in r_n} a_{i,j}}{\sum_{c_{0,j} \in r_n} a_{0,j}} \quad (5)$$

where N_{i,r_n} is the number of protons observable in the analyzed range r_n, C_0 is the concentration of the external standard (i=0) measured separately. This assumes comparable NMR response factors across analytes, which is reasonable here because the quantified signals correspond to the same class of nuclei, namely aromatic protons. These concentrations are ultimately used to reconstruct the analyte distribution across the hyperspace of conditions explored.

While moderate peak broadening, meaning a width increase of less than 30%, is accommodated automatically in the model via adjustment of $\tilde{v}_{i,m}$ during the fit, severe broadening requires additional care. Strong coordination or paramagnetic effects can attenuate resonances to the point that they merge with the baseline or become partially undetectable, leading to underestimated integrals. In these cases, we construct the respective reference model using spectra of coordination-prone substrates measured in conditions recreating the broadening, for instance in presence of metallic salt as detailed in the SI, **Section 1.13**. NMRanger source code is available for download via at = via GitHub at https://github.com/GadinaL/NMRanger.

2. Integration with the Robowski-NMR platform. With the NMRanger algorithm thus developed and deployed, we used it to map the hyperspace of the Chan-Lam coupling over 1591 distinct conditions. All reactions were performed on the Robowski-NMR automation platform described in our recent work [20] and whose replication blueprints were disclosed in ref [16,20]. In brief, Robowski is an in-house-built, inexpensive (~$25K) platform designed to execute user-specified sequences of reagent addition, mixing, and temperature control over 54-vial plates of 2 mL vials (**Figure 2a**, images 1 and 2, SI **Video 1**). Robowski-NMR couples this reaction preparation module with the system to handle NMR tubes and acquire spectra on an 80 MHz benchtop NMR (Spinsolve 80, Magritek). First, a pipetting module transfers the reaction crudes from the 54-vial plates (prepared by traditional Robowski from **Figure 2a**, image 3/5) to an NMR tube which is then inserted into the NMR machine by a Meca500 (Mecademic) robotic arm. After the spectrum is acquired, the same arm removes the tube from the NMR, turns it upside down, threads onto two consecutive washing stations supplying high-velocity stream of ethanol as rinsing solvent (30 s each) and then transfers to a drying station under a stream of hot air (2 min). As these washing/drying operations take place, the robot is already filling in another test-tube and acquiring its NMR spectrum. In this way, the system can operate in a continuous (for up to 72 hrs), closed-loop manner using only two NMR tubes (for further details and all computer codes to control the system, see ref [20]).

We used this system to map our condition spaces of four nitrogen coupling partners: morpholine, 7-azaindole, 5-bromo-7-azaindole, and 1H-indazole, with phenylboronic acid as the aryl source and $Cu(OAc)_2$ as the copper catalyst. Depending on the substrate, the reaction hyperspace included up to seven dimensions: temperature, atmosphere, reaction time, phenylboronic acid equivalents, base equivalents, substrate concentration, and copper loading. Phenylboronic acid equivalents were varied because excess arylboronic acid is often required due to side reactions. Base equivalents were varied because bases can influence N–H deprotonation, catalyst activation, transmetalation, and

boronic-acid-derived oxidation or protodeboronation.[32,38] Substrate concentration and copper loading were also varied, with copper screened from catalytic to stoichiometric amounts because catalytic Chan–Lam variants require efficient Cu(I)-to-Cu(II) reoxidation by air or O_2.[38] All reactions were performed in non-deuterated 1,2-dichloroethane (DCE) at 26 °C, a common temperature for Chan–Lam coupling, and also at elevated temperatures (40 and 75 °C for morpholine, 50 and 100 °C for azaindoles, and 50 °C for indazole) to test heating effects. Because $Cu(OAc)_2$ is not soluble in DCE, copper source was first added to each vial as an aqueous solution. The plate was then dried in a vacuum oven at 40 °C to remove water, and the remaining reagents were subsequently added in DCE. The atmosphere was varied because oxygen participates in the copper redox cycle, and oxygen availability differs between open-air and oxygen-saturated sealed vials – accordingly, we performed reactions under air as well as in vials sparged with O_2 (though a 3D-printed multi-nozzle device applied over the array of 54 vials, see SI, **Figure S1** and **Video 2**) for 60 seconds before sealing the plates with a hermetic lid.[51,52] We note that standard Chan–Lam procedures at larger scales often use either a continuous oxygen supply or open-air conditions. In our miniaturized workflow, however, sealed vials were required to prevent non-negligible concentration changes from even modest solvent evaporation, and they also enabled us to probe how limited oxygen availability influences catalytic turnover and –by-product formation.

3. Hyperspace mapping. The crux of the hyperspace approach – detailed in refs.[16–20]– is that while we acquired the ^{1}H NMR spectra of all 1591 crude reaction mixtures (which was expeditious on Robowski-NMR), we did not isolate/purify the components of each of the 1570 reactions, which would be extremely time consuming and costly. Instead, for each sub-space defined by a particular substrate, we combined the individual reaction crudes and then purified them on Gel Permeation Chromatography (GPC) or Medium-Pressure Liquid Chromatography (MPLC) system to isolate the main components (i.e., species present in appreciable amounts anywhere in the conditions space). The ^{1}H NMR spectra of these "basis set"[16-20] components were then used as references with which to unmix all 1591 individual reaction crudes using NMRanger. All spectra across the condition spaces for all substrates could be adequately explained with, depending on the substrate, four to seven components isolated over at most two re-purification rounds of the combined crudes (for mass-balance deficit due to minor unassigned or poorly resolved signals, see the SI, **Section 1,** additional hyperspaces). Overall, rather than performing 1591 separate HPLC analyses, we reconstructed the hyperspaces (at least in detail sufficient to interpret reaction's mechanism) and confirmed the identity of some unexpected products (from the scale-up of the alkylation/cyclization reaction between bromo azaindole and solvent, see **Figure 2b**) with only ten HPLC or MPLC runs. This ca. hundred fifty-fold reduction in effort is the hallmark of the hyperspace approach, enabling systematic mapping of reactivity landscapes that would otherwise be inaccessible.

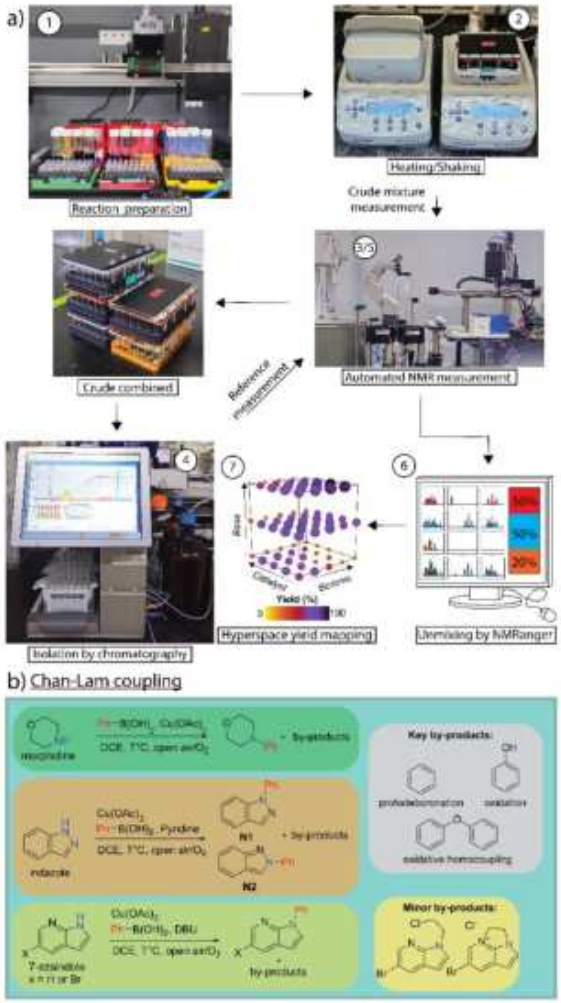


Figure 2. Experimental and computational workflow from reaction setup to hyperspace quantification. a) Overview of the automated workflow: **(1)** reaction setup of reactions in Robowski's 54-vial custom plates; **(2)** heating and shaking the plates transferred to the Eppendorf ThermoMixer® C for the desired reaction time; **(3 & 5)** Robowski-NMR performs cycles of automated transfer of the crudes into NMR tubes, acquisition of NMR spectra, and tube washing/drying for reusal (see SI **Video 1**); **(4)** isolation of reference compounds from combined crude mixtures; **(5)** acquisition of reference NMR spectra; **(6)** NMRanger fitting and unmixing of crude reaction spectra; and **(7)** conversion of fitted concentrations or yields into reaction hyperspace maps (here, for clarity, only a three-dimensional sub-space for one of the substrates is shown). **b)** Our current study used the workflow from **(1)-(7)** to map out the distribution of products and by-products across the hyperspace of Chan–Lam N-arylation reactions. The scheme shows the substrates used (morpholine, 7-azaindole/5-bromo-7-azaindole, and 1H-indazole with phenylboronic acid) as well as the key by-products isolated in the hyperspace.

4. Product and byproduct distributions over specific hyperspaces.

4.1. Hyperspace of morpholine. We began with morpholine, a simple nitrogen coupling partner whose *N*-phenyl derivatives are medicinally relevant.[53,54] Because morpholine itself is basic, no additional base was used.[55] In total, 748 conditions were screened across initial morpholine concentrations (20-110 mM), $Cu(OAc)_2$ loading (0.1–1.1 equiv), $PhB(OH)_2$ equivalents (1.0–2.0 equiv), temperature (26, 40, or 75 °C), and atmosphere (open air or O_2-saturated; for details see caption to **Figure 3**). The 20 mM lower limit was chosen for 80 MHz NMR sensitivity, and the upper concentration was constrained by $PhB(OH)_2$ solubility in DCE. The NMR spectra contained overlapping aromatic signals from the coupling product, $PhB(OH)_2$, and boronic-acid-derived by-products, but this overlap could be resolved using the default NMRanger fitting workflow over a single aromatic region from 6.5 to 8.5 ppm, which covered the relevant reference spectra while avoiding the broad non-deuterated solvent signal (obscuring aliphatic peaks of *N*-phenylmorpholine). A representative fitting example is shown in **Figure 3a**.

Figures 3b-f illustrate yield distributions for the main product of the coupling, *N*-phenylmorpholine. Focusing on the vertical axis, we see that formation of this product in appreciable yields requires $Cu(OAc)_2$ loading above 0.1 equiv. regardless of morpholine concentration or $PhB(OH)_2$ equivalents. Across the horizontal planes of the cubes, $PhB(OH)_2$ equivalents improve product formation moderately, mainly when sufficient copper is present (0.3 equiv. and above), consistent with the need for excess arylboronic acid when competing boronic-acid-derived pathways consume the aryl source. In most cases, yields become relatively stable at approximately 1.6 equiv and above. In the air, 26 °C, 24 h space (**Figure 3b**) an inverse trend is observed, with lower $PhB(OH)_2$ loading giving higher yields – this, however, likely reflects incomplete reaction at the shorter time point. In the corresponding 72 h space (**Figure 3c**), product formation increases with higher $PhB(OH)_2$ loading and then stabilizes at higher equivalents. The O_2, 40 °C, 24 h and O_2, 75 °C, 24 h spaces (**Figures 3e,f**) are the least sensitive to $PhB(OH)_2$ loading, with similar yields observed across the map.

By contrast, the concentration of morpholine has a much more pronounced and interesting effect in that the highest yields are observed for 20 mM (especially with high copper loading and sufficient $PhB(OH)_2$) and decrease for higher concentrations. In sealed 54-vial plates, higher substrate loading may create relative oxygen limitation, affecting Cu(I)-to-Cu(II) reoxidation and the balance between coupling and boronic-acid-derived side reactions. In this regard, examination of atmosphere and temperature variables showed that heating alone is insufficient to boost product formation. Under open air at 26 °C, increasing the reaction time from 24 to 72 h increases the maximum yield (from 65 to 81.4%), indicating continued reaction progress. Heating to 40 °C under air increases the maximum yield slightly at 24 h to 68.9%, whereas 40 °C under O_2 gives the highest 24 h yields (84.4%) and broadens the set of conditions with appreciable product formation, especially at lower-to-intermediate morpholine concentrations. At higher morpholine concentration, especially 50 mM, O_2 still helps (Y_{max} = 60%) but does not fully overcome the concentration-dependent limitation. Increasing to 75 °C under O_2, tested only at 20 mM based on the lower-temperature results, does not improve the reaction relative to 40 °C/O_2 and gives a maximum yield of 69.7% (**Figure 3b-f**).

In turn, **Figures 3g-j** illustrate the distributions of byproducts and unconsumed boronic acid for the 26 °C, air, 72 h subspace. The major byproduct is benzene (**Figure 3g**), indicating that protodeboronation competes with C–N coupling – in fact, its distribution is a "mirror image" of that of the *N*-phenylmorpholine main product, with yields increasing with increasing concentration of morpholine (only 10% for 20 mM yet 78% for 110 mM). In contrast, phenol and diphenyl ether remain below 10% under all morpholine conditions (**Figure 3h,i**). Unconsumed $PhB(OH)_2$ appears in lower-yielding regions (**Figure 3j**). Additional spaces for other conditions are provided in the SI, **Section 1.4**. Jointly, these results indicate that protodeboronation is the dominant boronic-acid-derived side pathway, whereas boronic acid oxidation and oxidative etherification are minor under the screened conditions. Thus, low *N*-phenylmorpholine yields can arise from either incomplete arylboronic acid consumption or diversion of $PhB(OH)_2$ to benzene (SI, **Figures S10-S13**).

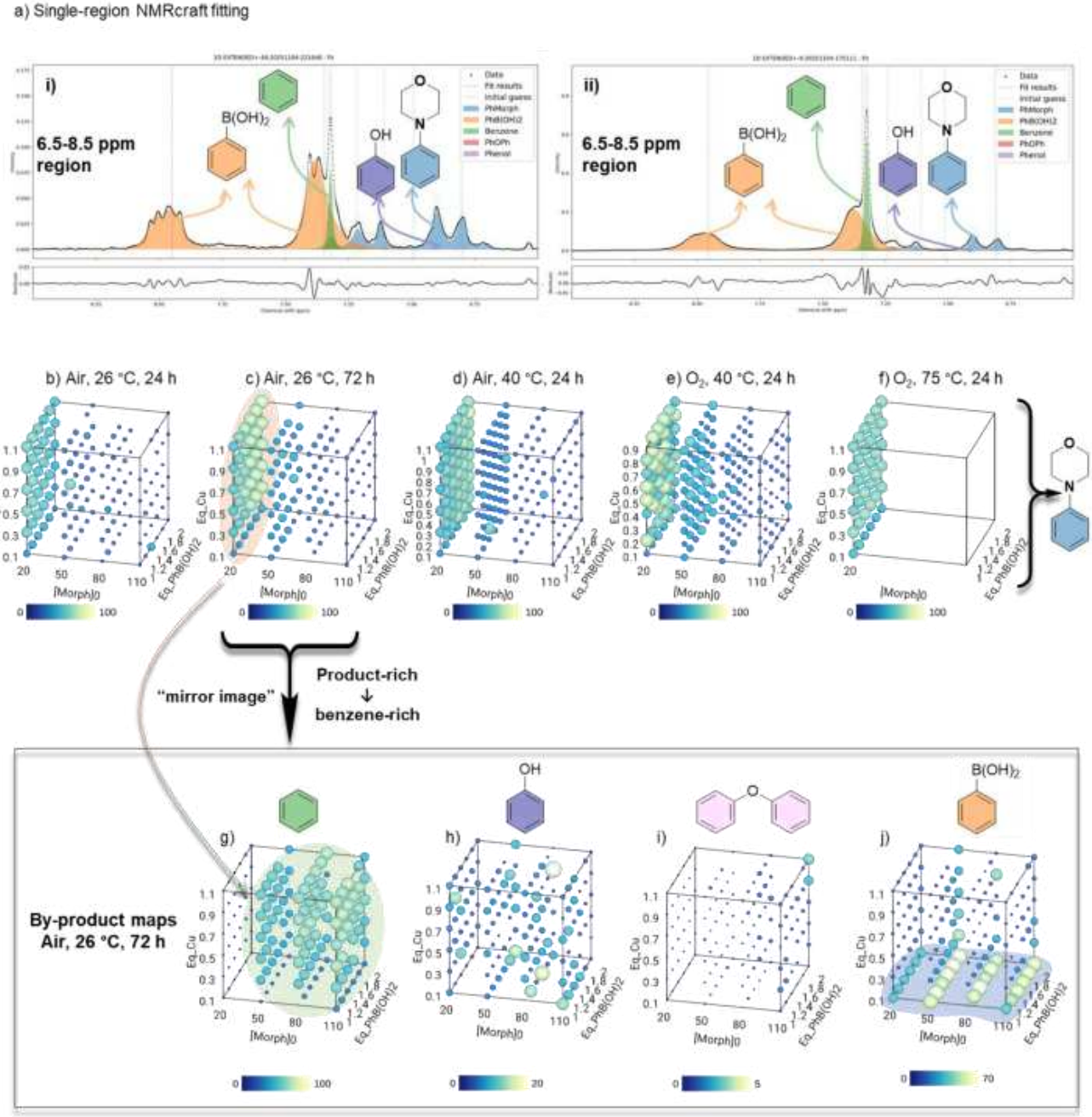


Figure 3. NMRanger fitting and 3D snapshots of morpholine reaction hyperspace. a) Representative single-region NMRanger fit of a morpholine crude reaction mixture under the conditions of 20 mM morpholine concentration, O_2 atmosphere, 40 °C, in the presence of 0.1 $Cu(OAc)_2$ and 1.8 equiv $PhB(OH)_2$. Panel **(i)** shows the example of fit with moderate broadening and panel **(ii)** shows the example of fit with stronger broadening of $PhB(OH)_2$. The overall fit (*black line)* is performed in the 6.5–8.5 ppm aromatic region and shows the fitted contributions (*colored*) from $PhB(OH)_2$, benzene, N-phenylmorpholine, and phenol. The traces below spectra give the difference between the fitted and experimental spectra; these differences do not exceed 0.025 intensity unit. **b-f)** N-phenylmorpholine yield maps under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively. **g)** Benzene by-product map for the air, 26 °C, 72 h condition. The orange shade in panel **c** marks the product-rich region at 20 mM concentration, while the green shade in panel **g** marks the benzene-rich region observed at higher concentrations, illustrating the "switchover" between N-phenylmorpholine and benzene formation. **h,i)** Maps of phenol and diphenyl ether for the air, 26 °C, 72 h space. **j)** Map of unconsumed $PhB(OH)_2$ for air, 26 °C, 72 h. The blue shade highlights the low-copper-loading region, where more unconsumed $PhB(OH)_2$ remained, especially at higher concentrations. Note: The entire morpholine campaign spanned 748 conditions. Dimensions were not uniformly explored. Moreover, experimental outliers (i.e., full evaporation in a vial or copper splash in the vacuum oven) were ignored. The set of experiments is detailed as: air, 26 °C, 72 h: 4×6×6 = 144 conditions –7 outliers; air, 26 °C, 24 h: 4×6×6 = 144 – 8 outliers; air, 40 °C, 24 h: 4 x 6 x 6 = 144, no outliers; air, 40 °C, 24 h: additional boronic acid loadings: 2×4×5 = 40 –1 outlier; air, 40 °C, 24 h: additional $Cu(OAc)_2$ loadings, 2×5×5 = 50, no outliers; O_2, 40 °C, 24 h: 4×9×6 = 216 – 10 outliers; O2, 75 °C, 24 h: 1×6×6 = 36, no outliers, for a total of 748 conditions. For details, see SI, **Section 1.3.**

4.2 Hyperspace of 7-azaindole. We next examined the 7-azaindole substrate, which required the addition of a base; DBU equivalents were therefore included as an additional dimension, and the reaction was screened across 356 conditions (see SI, **Table S2** for details of reaction-space "dimensions").[56] Because the spectra showed greater aromatic overlap, quantification used multiple NMR regions for the fitting: 5.7–7.0 ppm region containing characteristic signals from 1-phenyl-7-azaindole, $PhB(OH)_2$, and phenol was fitted first, and the algorithm used the resulting abundance estimates to solve the wider 5.7–9.5 ppm region. This broader window encompassed all signals present in the reference spectra of the fully aromatic components, so it retained the complete peak information needed for quantification while excluding the non-deuterated solvent region, where no azaindole-derived peaks are present. A representative multiple-NMR-region fitting example is shown in **Figure 4a** and the hyperspace yield distributions of the coupling product are quantified in **Figures 4b-j**.

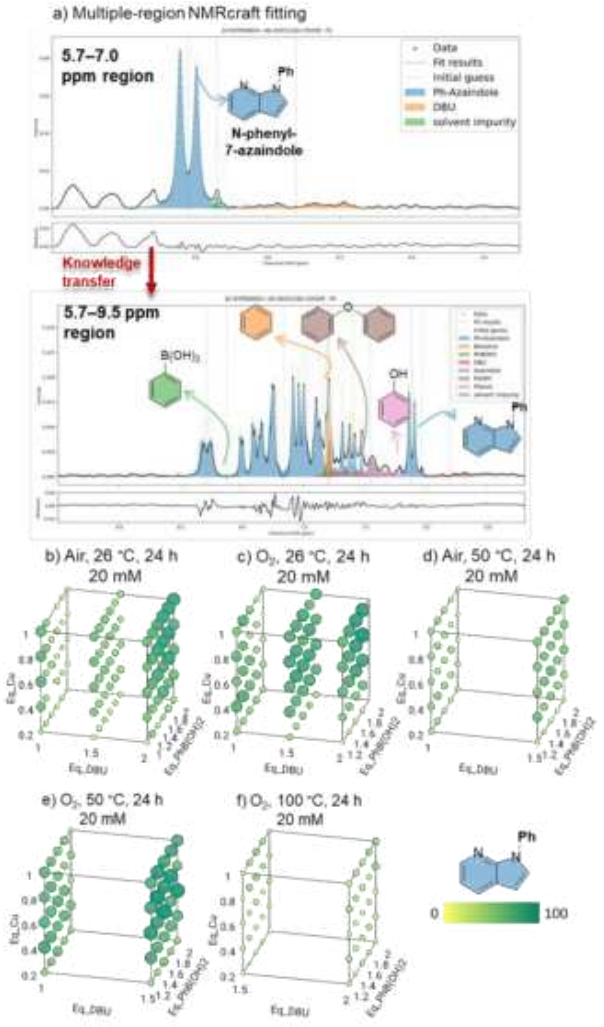


Figure 4. NMRanger fitting and 3D snapshots of 7-azaindole reaction hyperspace. a) Representative multiple-region NMRanger fitting of a 7-azaindole crude reaction mixture under the conditions of 20 mM initial 7-azaindole concentration, O_2 atmosphere, 50 °C, in the presence of 0.6 equiv $Cu(OAc)_2$, 1.2 equiv $PhB(OH)_2$, and 1.5 equiv DBU. A lower-overlap diagnostic region from 5.7–7.0 ppm is fitted first, and the resulting abundance information is transferred to a broader 5.7–9.5 ppm region for final fitting of contributions from N-phenyl-7-azaindole, $PhB(OH)_2$, diphenyl ether, benzene, and phenol (*colored*). The traces below the spectra give the difference between the fitted and experimental spectra; these differences do not exceed 0.025 intensity unit **b–f)** N-phenyl-7-azaindole yield spaces obtained at 20 mM 7-azaindole under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

As seen in these maps, oxygen availability and DBU loading strongly affect product formation. At 26 °C under open air, N-phenyl-7-azaindole forms across a broad hyperspace region, with higher yields generally associated with higher $Cu(OAc)_2$ (above 0.2 equiv) and stoichiometric-to-high DBU (1 to 2 equiv). O_2 saturation at 20 mM initial concentration of 7-azaindole, 26 °C, and 2 equiv base, increases the maximum yield from 83.4% to 91.9%, confirming once again that oxygen availability improves coupling (**Figure 4b,c**). Temperature has a non-monotonic effect. At 50 °C under air, in the presence of 1.5 equiv DBU, *N*-phenylation (max yield 60.3%) is less efficient than under O_2 (max yield 94.9%). However, further heating to 100 °C under O_2 decreases product formation across much of the hyperspace, with a maximum yield of only 47.3% even in the presence of high base loading (2.0 equiv of DBU, **Figure 4d-f**). In turn, increasing substrate concentration reduces *N*-phenyl-7-azaindole formation – at 20 mM, high yields up to 92% are observed across a broad range of $Cu(OAc)_2$ and DBU loadings, but decrease to, at most, 51.6% at 50 mM and 39.7% at100 mM and 2.0 equiv DBU (**Figures 5a,b**). This concentration dependence parallels the morpholine case under oxygen-saturated conditions but is more pronounced, showing that higher substrate loading can limit coupling even when reagent equivalents are scaled accordingly.

The by-product profiles match the morpholine case, with benzene as the main boronic-acid-derived by-product in regions where C–N coupling is low. At 50 °C, under air, benzene's maximal yield reaches 51.8% (**Figure 5c**). In contrast, phenol and diphenyl ether remain low, < 5.5% for both (see **Figures S30, S31**). At 50 °C under air with 1.5 equiv DBU, ~ 40% of unconsumed $PhB(OH)_2$ remains in the low-copper region, especially at 0.2 equiv $Cu(OAc)_2$, (**Figure 5d**). Under O_2 at 100 °C, with 1.5 equiv DBU, the amount of unconsumed $PhB(OH)_2$ is higher, ranging from 47% to 71% (**Figure S27**).

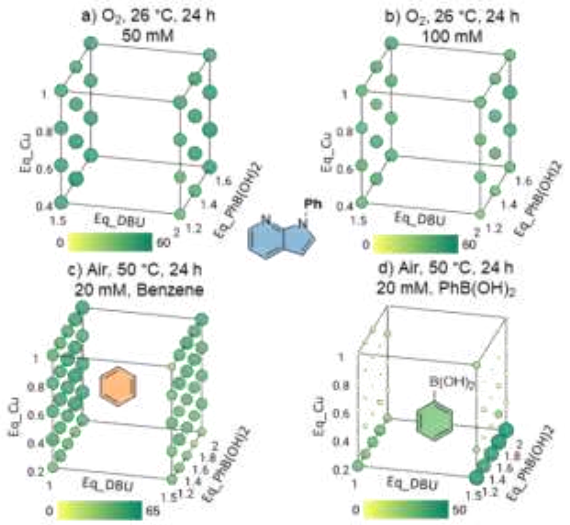


Figure 5. Azaindole reaction spaces quantifying concentration-dependent product formation, benzene by-product, and unconsumed phenylboronic acid. a,b) Concentration-dependent O_2-saturated spaces at 26 °C, 24 h, with initial 7-azaindole concentrations of 50 and 100 mM, respectively. **c)** Benzene by-product space for the air, 50 °C, 24 h, 20 mM initial substrate concentration. **d)** Unconsumed $PhB(OH)_2$ hyperspace under air, 50 °C, 24 h, 20 mM initial substrate concentration.

4.3. Hyperspace of 5-bromo-7-azaindole. To evaluate effects of the substitution on the azaindole scaffold, we examined 5-bromo-7-azaindole substrate under O_2-saturated conditions as they produced higher yield for the analogue 7-azaindole. In total, 148 conditions were analyzed. Because the spectra also contained overlapping aromatic signals, quantification mostly used multiple NMR regions for the fitting. The 7.85–8.70 ppm region containing signals from *N*-phenyl-5-bromo-7-azaindole, unreacted 5-bromo-7-azaindole, and $PhB(OH)_2$ was fitted first, and the abundance estimates were transferred to the wider region (5.8 to 9.5 ppm) for final unmixing (**Figure 6a**).

The 5-bromo product follows the same broad trends as 7-azaindole. At 26 °C, it forms across the hyperspace, reaching a maximum yield of 85.5%. At the lowest copper loading of 0.2 equiv $Cu(OAc)_2$, up to 64% of unconsumed 5-bromo-7-azaindole (**Figure S48**) and up to 60.3% unconsumed $PhB(OH)_2$ (at the highest $PhB(OH)_2$ loading, **Figure 7d**) are observed. These results indicate that low copper loading mainly leads to incomplete substrate conversion rather than increased protodeboronation. Heating to 50 °C gives the most productive hyperspace, increasing product formation to a maximum yield of 95.8%, while unconsumed $PhB(OH)_2$ is bounded at 39.3% at the lowest copper (0.2 equiv) and highest boronic acid (2 equiv) loading. At 100 °C, product formation decreases, limiting the yield of *N*-phenylation to 67.1% (see **Figures 6b-d** for the spaces of *N*-phenylation and **Figures 7d-f** for unconsumed $PhB(OH)_2$).

The boronic-acid-derived by-product profile also resembles the morpholine and 7-azaindole cases: benzene is the main by-product, reaching a maximum yield of 30% at 100 °C. Diphenyl ether is detected in up to 5.8% yield, while the yields of phenol are 16.6%, 23.9%, and 20.8% at 26, 50, and 100 °C, respectively. Benzene by-product maps are given in **Figure 7a-c**, and those for other by-products are provided in the SI, **Section 1.8**.

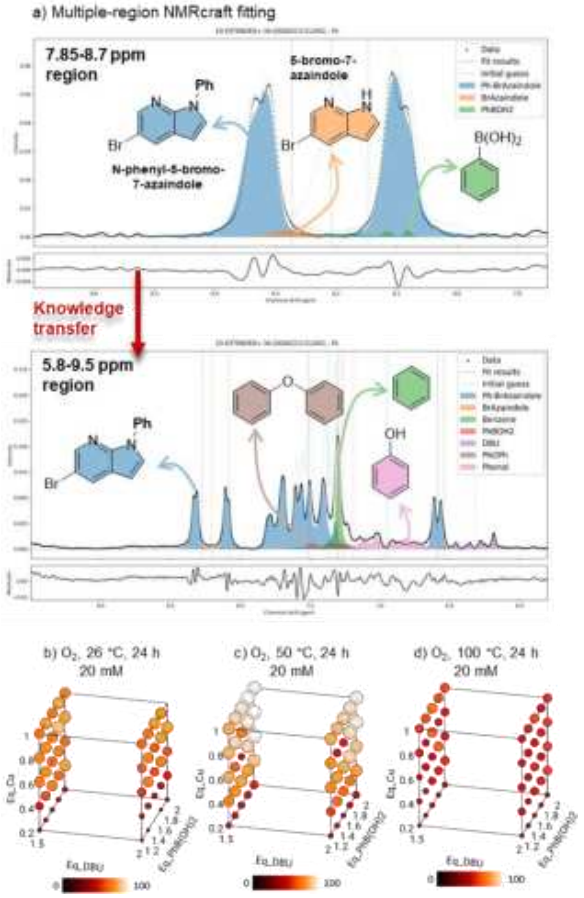


Figure 6. NMRanger fitting and 5-bromo-7-azaindole reaction hyperspaces. a) Representative multiple-region NMRanger fitting of a 5-bromo-7-azaindole crude reaction mixture under the conditions of 20 mM initial concentration of 5-bromo-7-azaindole, O_2 atmosphere, 26 °C, in the presence of 0.8 equiv $Cu(OAc)_2$, 1.8 equiv $PhB(OH)_2$, and 2 equiv DBU. The 7.85–8.70 ppm characteristic region is fitted first and transferred to the broader 5.8–9.5 ppm region for final fitting of contributions from N-phenyl-5-bromo-7-azaindole, diphenyl ether, benzene, phenol, and unconsumed $PhB(OH)_2$ (*colored*). The traces below the spectra give the difference between the fitted and experimental spectra; these differences do not exceed 0.025 intensity unit **b–d)** N-phenyl-5-bromo-7-azaindole yield hyperspaces at 20 mM concentration of 5-bromo-7-azaindole under O_2-saturated conditions at 26, 50, and 100 °C for 24 h, respectively.

This said, an effect distinct to 5-bromo-7-azaindole appears at 100 °C, revealing unexpected solvent-derived reactivity. DCE is not fully inert, and a solvent-derived *N*-alkylation product of 5-bromo-7-azaindole is isolated along with a second, more polar species tentatively assigned as a cyclized pyridinium chloride product (**Figure 7g**). However, this cyclized species forms only in minute amounts and are unstable after isolation, preventing further characterization.

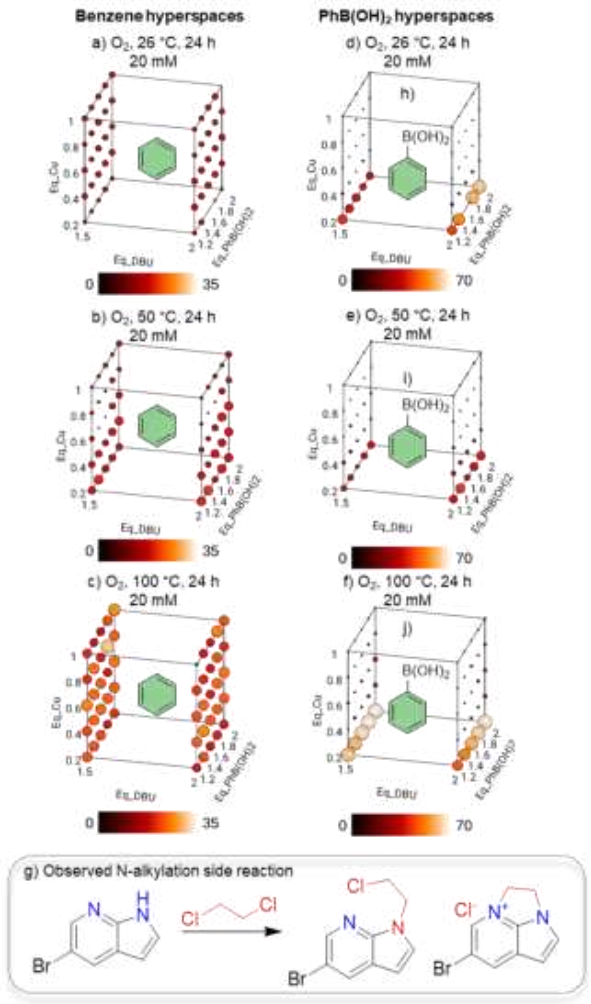


Figure 7. Maps for benzene by-product and unconsumed phenylboronic acid as well as solvent-derived N-alkylation side reaction. a-c) Benzene by-product hyperspace for 20 mM initial substrate concentration under O_2, for 24 h, and at 26, 50, 100 °C respectively. **d-f)** Unconsumed $PhB(OH)_2$ hyperspaces for 20 mM initial substrate concentration under O_2, for 24 h, and at 26, 50, 100 °C respectively. **g)** Observed N-alkylation side reaction of 5-bromo-7-azaindole in DCE, giving chloroethylated azaindole products.

4.4. Hyperspace of 1H-Indazole. Finally, we examined 1H-indazole as the last substrate, for which *N*-arylation can give 1-phenyl-1H-indazole (N1) and 2-phenyl-2H-indazole (N2) regioisomers. The reaction parameters that we studied for this substrate were pyridine equivalents, $Cu(OAc)_2$ loading, and $PhB(OH)_2$ equivalents.[36] In total, 339 conditions were analyzed (see SI, **Table S3** details of reaction-space "dimensions"). Most spectra were quantified by fitting across multiple NMR regions. The 8.0–8.8 ppm region containing characteristic signals from both regioisomers and $PhB(OH)_2$ was fitted first, followed by final unmixing over the broader, 6.6–9.0 ppm range (**Figure 8a**). As for the azaindole cases, this aromatic window contained all indazole-derived peaks. Very low-conversion conditions, including some low-copper/high-base/excess-$PhB(OH)_2$, higher-concentration, and different copper and base sources screens, were quantified using only the wider 6.6–9.0 ppm region because product signals were only trace (for details, see the SI, **Sections 1.9, 1.10**).

At 20 mM 1H-indazole, 1-phenyl-1H-indazole (N1) is consistently the dominant regioisomer across the screened conditions, whereas 2-phenyl-2H-indazole (N2) forms only as a minor product. Higher overall coupling yields are generally observed with high copper loading and pyridine equivalents. Unlike the other substrates, oxygen saturation produces only a modest increase in the maximum N1 yield, from 66.5% under air to 70.5% under O_2 at 26 °C (**Figures 8b,c**). The 50 °C hyperspace trends and maximum yields are similar to those obtained at 26 °C (**Figures 8d,e**). N2 formation remains lower across all conditions, with maximum yields ranging from 22.2 to 31.4% under air and O_2 (**Figures 9a–d**). Oxygen and temperature cause only small changes in the N1/N2 ratio; their main effect is on total conversion rather than regioselectivity. Representative N1/N2 ratio maps are shown in **Figure 9e** and **Figure S65**.

Excess pyridine and $PhB(OH)_2$ loadings (3 and 4 equiv) were screened in the presence of low copper loading (0.2 equiv). These experiments show that excess base or arylboronic acid does not overcome low copper loading (Y_{max} = 48%). When $Cu(OAc)_2$ loading is already sufficient (1 equiv), increasing pyridine and $PhB(OH)_2$ decreases the maximum yield (Y_{max} = 57%). The hyperspace maps for these studies are provided in the SI**, Section 1.10**.

Similar to the other substrates, increasing the initial concentration of 1H-indazole noticeably reduces the formation of both regioisomers. At 65 mM substrate concentration, the maximum N1-phenylation yields are only 19.3% at 26 °C and 22.3% at 50 °C (**Figure S67**).

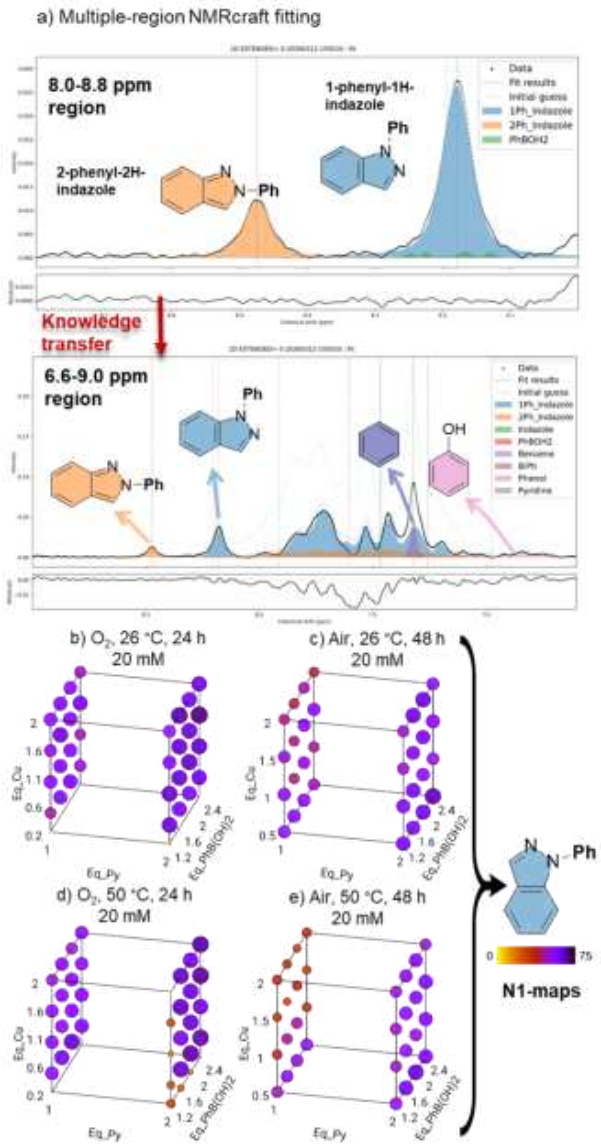


Figure 8. NMRanger fitting and indazole reaction major isomer hyperspaces. a) Representative multiple-region NMRanger fitting of an indazole crude reaction mixture under the conditions of 20 mM initial 1H-indazole concentration, O_2 atmosphere, 26 °C, 1.0 equiv $Cu(OAc)_2$, 1.2 equiv $PhB(OH)_2$, and 1.0 equiv pyridine. A lower-overlap characteristic region from 8.0–8.8 ppm is fitted first, and the resulting abundance information is transferred to a broader 6.6–9.0 ppm region for fitting of contributions from the N1-phenylated product, the N2-phenylated regioisomer, benzene, and phenol (*colored*). The traces below the spectra give the difference between the fitted and experimental spectrum; these differences do not exceed 0.025 intensity unit **b–e)** 1-Phenyl-1H-indazole yield hyperspaces at 20 mM initial 1H-indazole concentration under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively

The indazole by-product maps also echo trends for other substrates, with benzene as the dominant boronic-acid-derived side product. At 20 mM initial indazole concentration, benzene reaches maximum yields up to ~72% (see SI, **Figure S62**) while phenol and diphenyl ether form only in trace amounts under all conditions (see SI, **Figures S63, S64**).

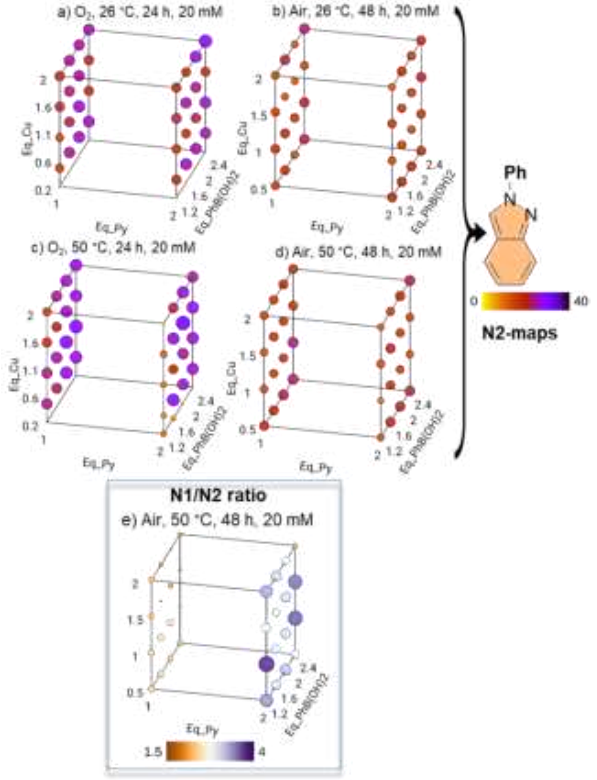


Figure 9. NMRanger fitting and snapshot of the hyperspace for the minor indazole isomer. a–d) 2-Phenyl-2H-indazole yield hyperspaces under the same conditions shown in **Figure 8, panels b–e,** respectively. **e)** Map of the N1/N2 regioisomer ratio for the air, 50 °C, 48 h condition at 20 mM initial 1H-indazole concentration.

The role of base was further examined through pyridine-free and alternative-base screens. In the absence of pyridine, phenyl-1H-indazole does not form or forms only in trace amounts – instead, high benzene concentrations are observed (SI, **Figure S73**) using DBU, DIPEA, DMAP, or TEA under O_2-saturated conditions at 26 °C gives yields systematically worse than with pyridine (50% from DMAP and trace for other bases). The N1/N2 ratio changes only modestly, indicating that base identity does not meaningfully affect regioselectivity (for hyperspaces, see SI, **Figure 73**).

We also tested other copper sources (copper(II) 2-ethylhexanoate, $CuBr_2$, $Cu(acac)_2$,and CuI) at selected eight points (copper 0.2, 0.5, 1.0 equiv, $PhB(OH)_2$ 1.2, 2.0, 3.0 equiv, and pyridine 2.0 equiv) Among the alternative Cu(II) sources, copper(II) 2-ethylhexanoate gives the highest N1-phenylation yield of 57%, $CuBr_2$ screen gives

34.8%, while $Cu(acac)_2$ is largely ineffective. CuI, tested as a Cu(I) source, yields only 13.9% product under O_2 – because Chan-Lam coupling is typically performed with Cu(II) salts, this low but detectable CuI reactivity is perhaps notable (**Figure S74**). In the commonly proposed mechanism[38,57,58], transmetalation from the boronic acid forms an organocopper intermediate, followed by generation of a higher-valent Cu(III) species that undergoes C–N bond-forming reductive elimination. Effective catalytic turnover therefore requires access to Cu(II)/Cu(III) redox states, and Cu(I) must first be oxidized under O_2 before entering the catalytic cycle. The limited product formation with CuI may therefore reflect slow or incomplete Cu(I) to Cu(II) oxidation under these conditions[38,59,60]. Overall, neither the base nor the copper-source screen improves indazole product formation or regioselectivity relative to the pyridine/$Cu(OAc)_2$ system.

5. Mechanistic Interpretation of the Hyperspace Maps. Taken together, the hyperspace trends agree with the established Chan–Lam mechanism[38] (**Figure 10**), and also provide mechanistic insight by showing which reaction variables control different parts of the catalytic cycle: oxygen availability governs oxidative turnover, copper loading controls boronic acid consumption, and base loading supports catalyst activation and N–H substrate deprotonation.

In the catalytic cycle, dimeric copper species (I) undergoes amine-assisted denucleation to form mononuclear Cu(II) complex (II). Complex (II) reacts with $PhB(OH)_2$ acid to give Cu(II)–organoboron complex (III), followed by transmetalation to aryl–Cu(II)–amine complex (IV). Oxidation/ disproportionation gives Cu(III) intermediate (V), and reductive elimination forms the C–N coupled product with release of Cu(I)OAc (VI). Reoxidation of Cu(I) to Cu(II) by O_2/air regenerates the active copper species and continues the cycle.

The hyperspace maps connect these mechanistic steps to the reaction outcomes and show that oxygen availability affects the reaction in two main ways. Taking morpholine as the example substrate, we note that reactions performed under air at 26 °C for 24 h show slower substrate conversion than oxygen-saturated reactions. This likely reflects slower oxygen transfer from the vial headspace into the solution and, consequently, less efficient reoxidation of Cu(I) to Cu(II). The improved yield observed after 72 h under air supports this interpretation, indicating that the reaction proceeds under oxygen-limited conditions but requires longer time to reach higher conversion (**Figure 10b**, **panel iii**). At the same time, at higher morpholine concentrations of 50, 80, and 110 mM, the reaction requires more oxygen, but the amount of oxygen available in each vial remains limited. A similar concentration-dependent decrease in product formation was also observed for the other N-coupling partners, as discussed above. Under these conditions, Cu(I) accumulation in the presence of proton sources may promote protodeboronation, leading to increased benzene formation (**Figure 10b**, panel iii; see also SI, **Section 1** for the other hyperspaces).[38,59]

The azaindole and indazole maps show similar oxygen- and copper-dependent trends, but the role of base differs from that in the morpholine reaction. Morpholine acts as both the N-coupling partner and a basic amine additive, whereas azaindoles and indazoles require separate bases, such as DBU or pyridine, to support copper catalyst activation (denucleation) and acid trapping.[38,59] Since these bases are not coupling substrates, increasing their loading can more effectively promote deprotonation of the N–H substrate and neutralize acid generated during the reaction, explaining why higher DBU or pyridine loading generally improved azaindole and indazole *N*-arylation, whereas higher morpholine concentration reduced N-phenylmorpholine formation (**Figure 10b**, **panel i**).

Unconsumed $PhB(OH)_2$ hyperspaces, particularly for the azaindole series, further highlight the copper dependence of the catalytic cycle. In **Figure 10b**, **panel ii**, unconsumed $PhB(OH)_2$ remains high in the low-$Cu(OAc)_2$ region and decreases as copper loading increases, showing that insufficient copper limits boronic acid consumption. This behavior is consistent with the requirement for available Cu(II) to support oxidation/disproportionation and generate the higher-valent Cu(III) intermediates required for C–N bond formation. Together, the hyperspace maps link oxygen availability, copper loading, base loading, residual $PhB(OH)_2$, and benzene formation to key steps of the Chan–Lam catalytic cycle.

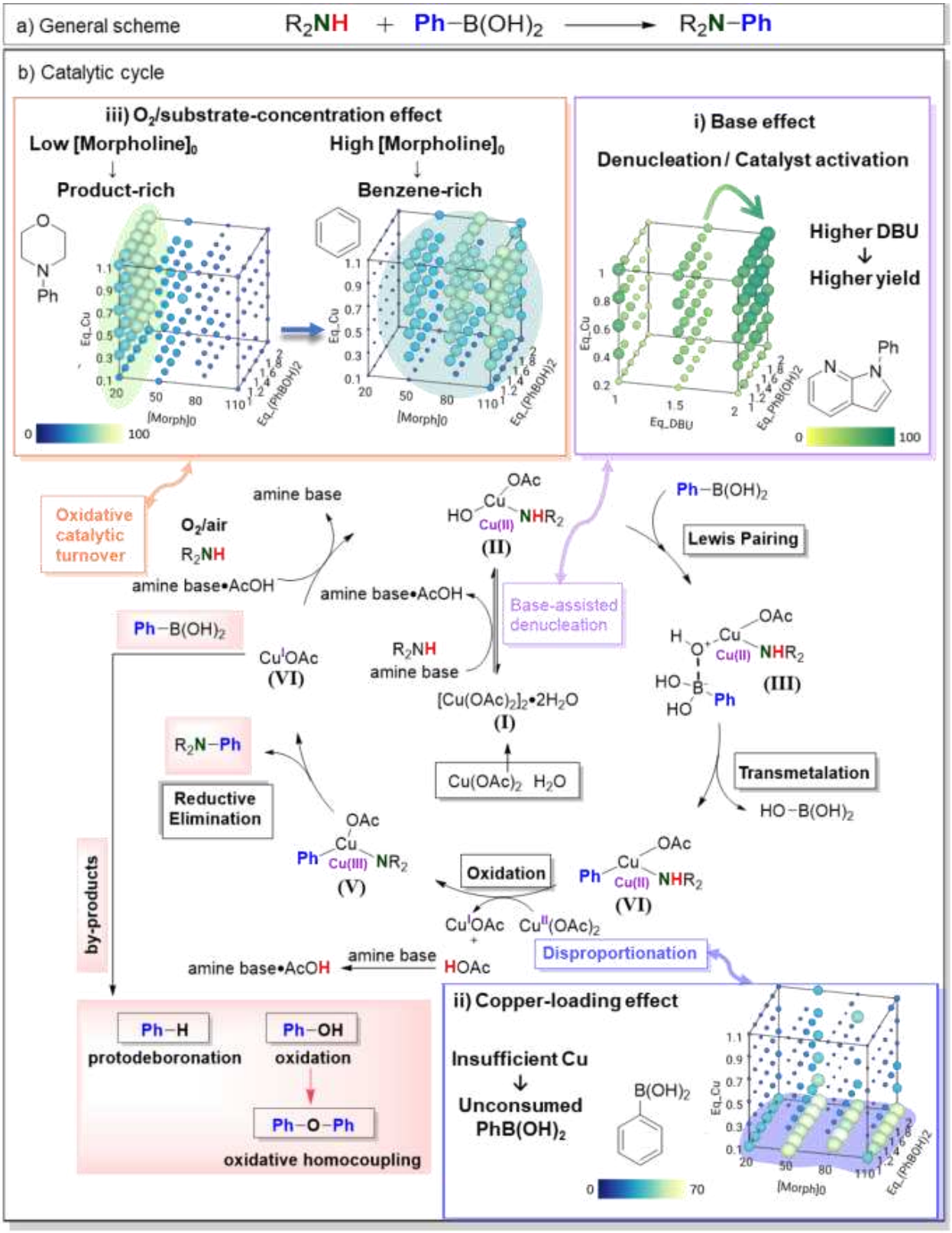


Figure 10. Established mechanism of Chan–Lam coupling and its connection to hyperspace trends. a) General Chan–Lam *N*-arylation scheme between phenylboronic acid and an N–H coupling partner to form the corresponding N-phenylated product. **b)** Chan–Lam catalytic cycle. The framed hyperspaces connect key mechanistic steps to reaction outcomes. The 7-azaindole hyperspace **(i)**, obtained under air at 26 °C for 24 h with 20 mM initial 7-azaindole concentration, is shown in the purple frame and illustrates the role of amine base loading, where higher DBU equivalents improve N-arylation by supporting denucleation and catalyst activation.[38] The residual $PhB(OH)_2$ hyperspace **(ii)**, obtained under air at 26 °C for 72 h, is shown in the blue frame and highlights the low-copper region, where insufficient copper loading limits boronic acid consumption and reduces access to the higher-valent copper intermediates required for productive C–N coupling. The morpholine hyperspace **(iii)**, obtained under air at 26 °C for 72 h and shown in the orange frame, illustrates the effects of oxygen availability and substrate concentration on oxidative catalytic turnover: lower morpholine concentration favors *N*-phenylmorpholine formation (left map), whereas higher morpholine concentration shifts the reaction toward benzene formation through protodeboronation (right).

6. Accuracy of the NMRanger hyperspace analyses. Our final goal was to assess by independent methods the accuracy of the yields determined by NMRanger – such a comparison is essential to verify that the trends observed across the hyperspace are meaningful. To this end, we performed a series of representative spot checks across the spaces described in Sections 4.1-4.4. In these studies, we compared the NMRanger results to the yields estimated based on chromatographic analyses (GPC or HPLC) and/or to NMR yields estimated using internal standard via MNova software. To obtain sufficient amounts of material for the GPC/HPLC analyses, we pooled multiple samples from 27-54 vials (executed under the same condition). For the MNova analyses, a known amount of internal standard was added, selected peaks were integrated, and the product yield was calculated from the relative integrations. The results plotted in **Figure 11** show overall good agreement between reference and NMRanger-derived yields, supporting the quantitative reliability of the NMRanger approach. Representative examples are summarized in **Table 1**, with the complete dataset provided in the SI, **Table S10**. Across all 65 validation measurements, regardless of the ground-truth method used, NMRanger gave an R^2 of 0.801 and a mean absolute error of 6.38%.

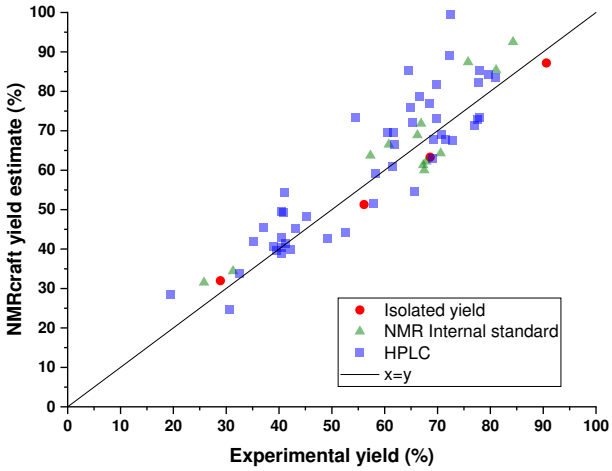


Figure 11. Validation of NMRanger yield estimates. NMRanger-estimated yields are plotted against reference experimental yields determined by GPC isolation (*red circles*), HPLC (*blue squares*), or internal-standard NMR (*green triangles*). The black line indicates perfect agreement between estimated and experimental yields. The mean absolute errors were 4.2, 6.01, and 6.7 yield %, with corresponding (R^2) values of 0.96, 0.86, and 0.766 for isolated yield, internal-standard NMR, and HPLC, respectively.

Table 1. Examples of head-to-head comparisons of yields estimated by NMRanger vs. other methods (GPC, HPLC, internal standard NMR). 1,3,5-Trimethoxybenzene is used as internal standard. All 65 comparisons are provided in the SI, **Table S10**.

Substrate / Concentration / Condition	Product	NMRanger Yield%	Comparative Yield%	Additional comments
Morpholine; 20 mM; 40°C/O_2; 0.9 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$	N-phenylmorpholine	56.10%	51.30%	Isolated yield on GPC, 54 vials pooled
	Unconsumed $PhB(OH)_2$	28.90%	32.00%	Isolated yield on GPC, 54 vials pooled
7-azaindole; 20 mM; 26°C/O_2; 1.0 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	68.60%	63.30%	Isolated yield on GPC, 27 vials pooled
7-azaindole; 20 mM; 26°C/O_2; 0.6 equiv $Cu(OAc)_2$; 1.4 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	81.10%	85.40%	NMR yield determined by using an internal standard.
7-azaindole; 20 mM; 26°C/O_2; 0.2 equiv $Cu(OAc)_2$; 1.2 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	31.10%	34.30%	NMR yield determined by using an internal standard.
7-azaindole; 20 mM;26°C/O_2; 0.2 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	40.56%	40.55%	Yield determined by HPLC
7-azaindole; 20 mM; 26°C/O_2; 0.6 equiv $Cu(OAc)_2$; 1.4 equiv $PhB(OH)_2$; 1.0 equiv DBU	N-Phenyl-7-azaindole	61.88%	66.75%	Yield determined by HPLC
7-azaindole; 20 mM; 26°C/O_2; 1.0 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	77.92%	73.70%	Yield determined by HPLC
7-azaindole; 20 mM; 26°C/O_2; 0.2 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 1.0 equiv DBU	N-Phenyl-7-azaindole	41.00%	54.50%	Yield determined by HPLC
7-azaindole; 20 mM; 26°C/O_2; 0.8 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	66.60%	78.60%	Yield determined by HPLC
Executed under the same condition: 5-bromo-7-azaindole; 20 mM; 26°C/O_2; 0.8 equiv $Cu(OAc)_2$; 1.8 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-5-bromo-7-azaindole	67.30%	61.30%	NMR yield is determined by using an internal standard.
	N-Phenyl-5-bromo-7-azaindole	67.90%	62.20%	NMR yield is determined by using an internal standard.
	N-Phenyl-5-bromo-7-azaindole	66.20%	68.90%	NMR yield is determined by using an internal standard.
1H-indazole; 20 mM; 26°C/O_2; 2.0 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 2.0 equiv pyridine	1-phenyl-1H-indazole:2-phenyl-2H-indazole (N1/N2)	90.60%	87.20%	Isolated yield on GPC, 27 vials pooled (4:1 ratio vs 3.6:1)

CONCLUSIONS

In conclusion, we have disclosed NMRanger, an algorithm for quantitative deconvolution of ^{1}H NMR spectra of complex mixtures into contributions from their constituent species. Within our own research[16–20], NMRanger has become the analytical backbone for reconstructing reaction hyperspaces – that is, multidimensional condition landscapes in which mechanistic pathways, flux distributions, and pathway switchovers can be inferred directly from crude mixtures. These reconstructions consistently reveal[16–20] that many reactions are not single, static mechanisms but switchable networks whose dominant pathways shift as atmosphere, stoichiometry, concentration, or temperature are varied. The ability to quantify crude mixtures reliably across thousands of conditions is what makes this network-level view operational rather than conceptual.

At the same time, NMRanger is equally applicable to more conventional reaction-optimization campaigns in which the major components are known but their proportions vary across conditions. Because the algorithm tolerates peak overlap, drift, and broadening – and can incorporate new components when they appear – it provides a practical route to quantitative analysis in workflows where isolation is impractical or where rapid iteration is essential.

Looking forward, we anticipate that NMRanger will support broader efforts to treat chemical reactivity as a navigable, high-dimensional space rather than a collection of isolated experiments. As automated platforms continue to scale, the ability to extract quantitative information directly from crude mixtures will be central to mapping reaction landscapes, identifying mechanistic switchovers, and discovering new reactivity patterns. To facilitate these directions, we provide the full NMRanger source code for adaptation to diverse reaction classes and automated systems.

ASSOCIATED CONTENT

Supporting Information. Experimental procedures, additional experimental data, supplementary videos, NMR characterization data, description of the code used, additional hyperspaces, additional discussions referred to in the main text. This material is available free of charge via the Internet at http://pubs.acs.org.

NMRanger source code. NMRanger source code is available via GitHub at https://github.com/GadinaL/NMRanger and via Zenodo at https://zenodo.org/records/21373963.

Raw data, analysis results, HPLC chromatograms, 3D model for the sparging adaptor. Raw NMR Fid, NMR spectra as CSVs of reference analytes and crude reaction mixture, analyses results as CSVs, 3D model to reproduce the adaptor used for oxygen sparging and HPLC chromatograms are available free of charge via via Zenodo at https://zenodo.org/records/21373963.

Corresponding Author

* Bartosz. A. Grzybowski – Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea; Email: nanogrzybowski@gmail.com

Author Contributions

Y.B.G. designed and performed the experiments and analyses. L.P.J.-L.G. developed NMRanger. R.F. developed the HPLC method and assisted with the acquisition of ground truths by HPLC. B.A.G. conceived and supervised the project. Y.B.G., L.P.J.-L.G., and B.A.G. wrote the manuscript.

Notes

The authors declare no competing interest.

ACKNOWLEDGMENT

This work was supported by the Institute for Basic Science (IBS-R020-D1). We thank Dr. Yaroslav I. Sobolev and Dr. Wai Shing Wong for valuable discussions during this project.

ABBREVIATIONS

NMR, Nuclear Magnetic Resonance; DCE, 1,2-dichloroethane; $Cu(OAc)_2$, copper(II) acetate; $PhB(OH)_2$, phenyl boronic acid; PhH, benzene; PhOH, phenol; PhOPh, diphenyl ether; DBU, 1,8-Diazabicyclo[5.4.0]undec-7-ene; $CuBr_2$, copper(II) bromide; CuI, copper (I) iodide.

REFERENCES

1. Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M. E.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. Synthesis of Many Different Types of Organic Small Molecules Using One Automated Process. *Science* **2015**, *347* (6227), 1221–1226. https://doi.org/10.1126/science.aaa5414.

2. Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J. M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P. J.; Angelone, D.; Cronin, L. Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language. *Science* **2019**, *363* (6423), eaav2211. https://doi.org/10.1126/science.aav2211.

3. Chatterjee, S.; Guidi, M.; Seeberger, P. H.; Gilmore, K. Automated Radial Synthesis of Organic Molecules. *Nature* **2020**, *579* (7799), 379–384. https://doi.org/10.1038/s41586-020-2083-5.

4. Liu, C.; Xie, J.; Wu, W.; Wang, M.; Chen, W.; Idres, S. B.; Rong, J.; Deng, L.-W.; Khan, S. A.; Wu, J. Automated Synthesis of Prexasertib and Derivatives Enabled by Continuous-Flow Solid-Phase Synthesis. *Nat. Chem.* **2021**, *13* (5), 451–457. https://doi.org/10.1038/s41557-021-00662-w.

5. Bubliauskas, A.; Blair, D. J.; Powell-Davies, H.; Kitson, P. J.; Burke, M. D.; Cronin, L. Digitizing Chemical Synthesis in 3D Printed Reactionware. *Angew. Chem.* **2022**, *134* (24), e202116108. https://doi.org/10.1002/ange.202116108.

6. Blair, D. J.; Chitti, S.; Trobe, M.; Kostyra, D. M.; Haley, H. M.; Hansen, R. L.; Ballmer, S. G.; Woods, T. J.; Wang, W.; Mubayi, V. Automated Iterative C Sp 3–C Bond Formation. *Nature* **2022**, *604* (7904), 92–97. https://doi.org/10.1038/s41586-022-04491-w.

7. Wang, W.; Angello, N. H.; Blair, D. J.; Tyrikos-Ergas, T.; Krueger, W. H.; Medine, K. N.; LaPorte, A. J.; Berger, J. M.; Burke, M. D. Rapid Automated Iterative Small-Molecule Synthesis. *Nat. Synth.* **2024**, *3* (8), 1031–1038. https://doi.org/10.1038/s44160-024-00558-w.

8. Strieth-Kalthoff, F.; Hao, H.; Rathore, V.; Derasp, J.; Gaudin, T.; Angello, N. H.; Seifrid, M.; Trushina, E.; Guy, M.; Liu, J.; *et al* Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters. *Science* **2024**, *384* (6697), eadk9227. https://doi.org/10.1126/science.adk9227.

9. Rauschen, R.; Ayme, J.-F.; Matysiak, B. M.; Thomas, D.; Cronin, L. A Programmable Modular Robot for the Synthesis of Molecular Machines. *Chem* **2025**, *11* (8). DOI: 10.1016/j.chempr.2025.102504.

10. McMullen, J. P.; Jensen, K. F. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. *Org. Process Res. Dev.* **2010**, *14* (5), 1169–1176. https://doi.org/10.1021/op100123e.

11. Reizman, B. J.; Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **2016**, *49* (9), 1786–1796. https://doi.org/10.1021/acs.accounts.6b00261.

12. Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2.

13. Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y.

14. Angello, N. H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E. R.; Roszak, R.; Wu, T. C.; Schroeder, C. M.; Aspuru-Guzik, A.; Grzybowski, B. A.; Burke, M. D. Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling. *Science* **2022**, *378* (6618), 399–405. https://doi.org/10.1126/science.adc8743.

15. Wang, J. Y.; Stevens, J. M.; Kariofillis, S. K.; Tom, M.-J.; Golden, D. L.; Li, J.; Tabora, J. E.; Parasram, M.; Shields, B. J.; Primer, D. N. Identifying General Reaction Conditions by Bandit Optimization. *Nature* **2024**, *626* (8001), 1025–1033. https://doi.org/10.1038/s41586-024-07021-y.

16. Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1.

17. Prajapati, B.; Wong, W.; Frydrych, R.; Kolygina, D. V.; Jia, Y.; Marazzi, B.; Vagkidis, N.; Schreiner, P. R.; Grzybowski, B. A. Robot-Assisted Reconstruction and Control of the Pechmann Reaction Network *Angew. Chem. Int. Ed.* **2026**, 65, e18394. https://doi.org/10.1002/anie.202518394.

18. Jiang, Y.; Gadina, L.; Ahumada, J. C.; Jia, Y.; Wong, W.-S.; Kim, J.; Grzybowski, B. A. Hyperspace Mapping and Iterative Dimensionality Reduction Reveal Radical Cascade Networks Controlled by Perovskite-Type Catalysts. *Chem* **2026**, 102987. https://doi.org/10.1016/j.chempr.2026.102987

19. Matuszczyk, D.; Frydrych, R.; Wong, W.-S.; Jia, Y.; Sobolev, Y. I.; Hanus, J.; Lee, J.; Grzybowski, B. A. Hyperspace Exploration Using Robotics for the Discovery of Mechanistically Distinct Transformations and Complex Functional Products. *Nat. Synth.* **2026**. https://doi.org/10.1038/s44160-026-01096-3

20. Jia, Y.; Wong, W.; Bilgi-Gadina, Y.; Gadina, L. P. J. -L.; Kwong, K.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* **2026**, e9132891. https://doi.org/10.1002/anie.9132891

21. Katzenburg, F.; Boser, F.; Schäfer, F. R.; Pflüger, P. M.; Glorius, F. Calibration-Free Quantification and Automated Data Analysis for High-Throughput Reaction Screening. *Digit. Discov.* **2025**, *4* (2), 384–392. https://doi.org/10.1021/jacs.6c08584.

22. Hu, M.; Yang, L.; Twarog, N.; Ochoada, J.; Li, Y.; Vrettos, E. I.; Torres-Hernandez, A. X.; Martinez, J. B.; Bhatia, J.; Young, B. M. Continuous Collective Analysis of Chemical Reactions. *Nature* **2024**, *636* (8042), 374–379. https://doi.org/10.1038/s41586-024-08211-4.

23. Bayly, A. A.; McDonald, B. R.; Mrksich, M.; Scheidt, K. A. High-Throughput Photocapture Approach for Reaction Discovery. *Proc. Natl. Acad. Sci.* **2020**, *117* (24), 13261–13266. https://doi.org/10.1073/pnas.2003347117.

24. Cabrera-Pardo, J. R.; Chai, D. I.; Liu, S.; Mrksich, M.; Kozmin, S. A. Label-Assisted Mass Spectrometry for the Acceleration of Reaction Discovery and Optimization. *Nat. Chem.* **2013**, *5* (5), 423–427. https://doi.org/10.1038/nchem.1612.

25. Das, J.; Zhang, X.; Tan, Y.; Abdelalim, M.; Lu, T.; Mauro, C.; Regan, C. J.; Cernak, T. A 50,688-Reaction Data Set Reveals General Ligands and Mechanistic Diversity in C–N Couplings. *J. Am. Chem. Soc.* **2026**, *148* (25), 26401–26412. https://doi.org/10.1021/jacs.6c05959.

26. Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. *Nat. Commun.* **2023**, *14* (1), 3924. https://doi.org/10.1038/s41467-023-39531-0.

27. Troshin, K.; Hartwig, J. F. Snap Deconvolution: An Informatics Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **2017**, *357* (6347), 175–181. https://doi.org/10.1126/science.aan1568.

28. Domżał, B.; Nawrocka, E. K.; Gołowicz, D.; Ciach, M. A.; Miasojedow, B.; Kazimierczuk, K.; Gambin, A. Magnetstein: An Open-Source Tool for Quantitative NMR Mixture Analysis Robust to Low Resolution, Distorted Lineshapes, and Peak Shifts. *Anal. Chem.* **2024**, *96* (1), 188–196. https://doi.org/10.1021/acs.analchem.3c03594.

29. Ricci, A. *Amino Group Chemistry: From Synthesis to the Life Sciences*; John Wiley & Sons, 2008.

30. West, M. J.; Fyfe, J. W. B.; Vantourout, J. C.; Watson, A. J. B. Mechanistic Development and Recent Applications of the Chan–Lam Amination. *Chem. Rev.* **2019**, *119* (24), 12491–12523. https://doi.org/10.1021/acs.chemrev.9b00491.

31. Devi, P. S.; Saranya, S.; Anilkumar, G. Recent Advances in Chan–Lam Coupling Reaction. *Catal. Sci. Technol.* **2024**, *14* (9), 2320–2351. https://doi.org/10.1039/d4cy00034j.

32. Qiao, J. X.; Lam, P. Y. S. Recent Advances in Chan–Lam Coupling Reaction: Copper-Promoted C–Heteroatom Bond Cross-Coupling Reactions with Boronic Acids and Derivatives. In *Boronic Acids*; Hall, D. G., Ed.; Wiley, 2011; pp 315–361. https://doi.org/10.1002/9783527639328.ch6.

33. Munir, I.; Zahoor, A. F.; Rasool, N.; Naqvi, S. A. R.; Zia, K. M.; Ahmad, R. Synthetic Applications and Methodology Development of Chan–Lam Coupling: A Review. *Mol. Divers.* **2019**, *23* (1), 215–259. https://doi.org/10.1007/s11030-018-9870-z.

34. G.-Simonian, N.; Serafino, A.; Neuville, L.; Masson, G. Chan–Lam Cross-Coupling With Alkylboron Reagents: From Transmetallation to Radical Pathways. *Eur. J. Org. Chem.* **2026**, *29* (18), e70389. https://doi.org/10.1002/ejoc.70389.

35. Xiao, H.; Cao, P.; Xiong, W.; Yang, T. Overcoming Limitations in Chan-Lam Amination with Alkylboronic Esters via Aminyl Radical Substitution. *Nat. Commun.* **2025**, *16* (1), 9121. https://doi.org/10.1038/s41467-025-64155-x.

36. Lam, P. Y.; Clark, C. G.; Saubern, S.; Adams, J.; Winters, M. P.; Chan, D. M.; Combs, A. New Aryl/Heteroaryl C-N Bond Cross-Coupling Reactions via Arylboronic Acid/Cupric Acetate Arylation. *Tetrahedron Lett.* **1998**, *39* (19), 2941–2944. https://doi.org/10.1016/S0040-4039(98)00504-8

37. Chan, D. M.; Monaco, K. L.; Wang, R.-P.; Winters, M. P. New N-and O-Arylations with Phenylboronic Acids and Cupric Acetate. *Tetrahedron Lett.* **1998**, *39* (19), 2933–2936. https://doi.org/10.1016/S0040-4039(98)00503-6.

38. Vantourout, J. C.; Miras, H. N.; Isidro-Llobet, A.; Sproules, S.; Watson, A. J. B. Spectroscopic Studies of the Chan–Lam Amination: A Mechanism-Inspired Solution to Boronic Ester Reactivity. *J. Am. Chem. Soc.* **2017**, *139* (13), 4769–4779. https://doi.org/10.1021/jacs.6b12800.

39. Evans, D. A.; Katz, J. L.; West, T. R. Synthesis of Diaryl Ethers through the Copper-Promoted Arylation of Phenols with Arylboronic Acids. An Expedient Synthesis of Thyroxine. *Tetrahedron Lett.* **1998**, *39* (19), 2937–2940. https://doi.org/10.1016/S0040-4039(98)00502-4.

40. Rout, M.; Lipfert, M.; Lee, B. L.; Berjanskii, M.; Assempour, N.; Fresno, R. V.; Cayuela, A. S.; Dong, Y.; Johnson, M.; Shahin, H.; Gautam, V.; Sajed, T.; Oler, E.; Peters, H.; Mandal, R.; Wishart, D. S. MagMet: A Fully Automated Web Server for Targeted Nuclear Magnetic Resonance Metabolomics of Plasma and Serum. *Magn. Reson. Chem.* **2023**, *61* (12), 681–704. https://doi.org/10.1002/mrc.5371.

41. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of^1 H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442. https://doi.org/10.1021/ac060209g.

42. Hao, J.; Astle, W.; De Iorio, M.; Ebbels, T. M. BATMAN—an R Package for the Automated Quantification of Metabolites from Nuclear Magnetic Resonance Spectra Using a Bayesian Model. *Bioinformatics* **2012**, *28* (15), 2088–2090. https://doi.org/10.1093/bioinformatics/bts308.

43. Astle, W.; De Iorio, M.; Richardson, S.; Stephens, D.; Ebbels, T. A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures. *J. Am. Stat. Assoc.* **2012**, *107* (500), 1259–1271. https://doi.org/10.1080/01621459.2012.695661.

44. Ravanbakhsh, S.; Liu, P.; Bjordahl, T. C.; Mandal, R.; Grant, J. R.; Wilson, M.; Eisner, R.; Sinelnikov, I.; Hu, X.; Luchinat, C. Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PloS One* **2015**, *10* (5), e0124219. https://doi.org/10.1371/journal.pone.0132873.

45. Röhnisch, H. E.; Eriksson, J.; Müllner, E.; Agback, P.; Sandström, C.; Moazzami, A. A. AQuA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma. *Anal. Chem.* **2018**, *90* (3), 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324.

46. Bruce, S. D.; Higinbotham, J.; Marshall, I.; Beswick, P. H. An Analytical Derivation of a Popular Approximation of the Voigt Function for Quantification of NMR Spectra. *J. Magn. Reson.* **2000**, *142* (1), 57–63. https://doi.org/10.1006/jmre.1999.1911.

47. Hellwig, J.; Neymeyr, K. On the Uniqueness of Continuous and Discrete Hard Models of NMR-Spectra. *J. Math. Chem.* **2025**, *63* (1), 267–296. https://doi.org/10.1007/s10910-024-01673-x.

48. Maddams, W. F. The Scope and Limitations of Curve Fitting. *Appl. Spectrosc.* **1980**, *34* (3), 245–267. https://doi.org/10.1366/0003702804730312.

49. Virtanen, P.; Gommers, R.; Oliphant, T. E. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

50. Branch, M. A.; Coleman, T. F.; Li, Y. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM J. Sci. Comput.* **1999**, *21* (1), 1–23. https://doi.org/10.1137/S1064827595289108.

51. III. Experiments on the Quantity of Gases Absorbed by Water, at Different Temperatures, and under Different Pressures. *Philos. Trans. R. Soc. Lond.* **1803**, *93*, 29–274. https://doi.org/10.1098/rstl.1803.0004.

52. Hone, C. A.; Kappe, C. O. The Use of Molecular Oxygen for Liquid Phase Aerobic Oxidations in Continuous Flow. In *Accounts on Sustainable Flow Chemistry*; Noël, T., Luque, R., Eds.; Topics in Current Chemistry Collections; Springer International Publishing: Cham, 2020; pp 67–110. https://doi.org/10.1007/978-3-030-36572-1_3.

53. Al-Soliemy, A. M.; Farghaly, T. A.; Abbas, E. M. H.; Shaaban, M. R.; Zayed, M. E. M.; El-Naggar, T. B. A. Synthesis of Thiazolyl-N-Phenylmorpholine Derivatives and Their Biological Activities. *Med. Chem.* **2021**, *17* (7), 790–805. https://doi.org/10.2174/1573406416666200517103435.

54. Solis, E.; Suyama, J. A.; Lazenka, M. F.; DeFelice, L. J.; Negus, S. S.; Blough, B. E.; Banks, M. L. Dissociable Effects of the Prodrug Phendimetrazine and Its Metabolite Phenmetrazine at Dopamine Transporters. *Sci. Rep.* **2016**, *6* (1), 31385. https://doi.org/10.1038/srep31385.

55. Quach, T. D.; Batey, R. A. Ligand- and Base-Free Copper(II)-Catalyzed C–N Bond Formation: Cross-Coupling Reactions of Organoboron Compounds with Aliphatic Amines and Anilines. *Org. Lett.* **2003**, *5* (23), 4397–4400. https://doi.org/10.1021/ol035681s.

56. Mondal, K.; Mukhopadhyay, N.; Sengupta, A.; Roy, T.; Das, P. Exploiting Coordination Behavior of 7-Azaindole for Mechanistic Investigation of Chan-Lam Coupling and Application to 7-Azaindole Based Pharmacophores. *Chem. – Eur. J.* **2023**, *29* (16), e202203718. https://doi.org/10.1002/chem.202203718.

57. King, A. E.; Brunold, T. C.; Stahl, S. S. Mechanistic Study of Copper-Catalyzed Aerobic Oxidative Coupling of Arylboronic Esters and Methanol: Insights into an Organometallic Oxidase Reaction. *J. Am. Chem. Soc.* **2009**, *131* (14), 5044–5045. https://doi.org/10.1021/ja9006657.

58. King, A. E.; Ryland, B. L.; Brunold, T. C.; Stahl, S. S. Kinetic and Spectroscopic Studies of Aerobic Copper(II)-Catalyzed Methoxylation of Arylboronic Esters and Insights into Aryl Transmetalation to Copper(II). *Organometallics* **2012**, *31* (22), 7948–7957. https://doi.org/10.1021/om300586p.

59. Chu, L.; Qing, F.-L. Copper-Mediated Oxidative Trifluoromethylation of Boronic Acids. *Org. Lett.* **2010**, *12* (21), 5060–5063. https://doi.org/10.1021/ol1023135.

60. Walker, B. R.; Manabe, S.; Brusoe, A. T.; Sevov, C. S. Mediator-Enabled Electrocatalysis with Ligandless Copper for Anaerobic Chan–Lam Coupling Reactions. *J. Am. Chem. Soc.* **2021**, *143* (16), 6257–6265. https://doi.org/10.1021/jacs.1c02103.

8.2.Supporting Information for Chan–Lam Coupling Reaction Hyperspace and Automated NMR Quantification

Supporting Information for manuscript entitled: **"*Automated NMRanger Deconvolution Enables Hyperspace-Wide Quantification of Crude Reaction Mixtures*"**

Yasemin Bilgi-Gadina[1,2], Louis P. J.-L. Gadina[1,2], Rafal Frydrych[1], Bartosz A. Grzybowski[1,3]*

AUTHOR ADDRESSES. [1]Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan, 44919, Republic of Korea.

[2]Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, 01-224, Poland.

[3]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea.

* Correspondence to: E-mail: nanogrzybowski@gmail.com

Table of Contents

1. Experimental 187

1.1. General Information 187

1.2. Robowski Reaction Setup, NMR Quantification, and Product Characterization 188

1.2.1. General Methods: Robowski reaction setup, Robowski-NMR measurements, isolations, and reference spectra recording. 188

1.3. Morpholine Chan-Lam Coupling 190

1.3.1. Characterizations and Reference Spectra for Morpholine Quantification 192

1.3.2. Calibration and fitting of morpholine 198

1.4. Additional hyperspaces of morpholine 200

1.5. 7-azaindole 203

1.5.1. Characterizations and Reference Spectra for 7-azaindole Quantification 204

1.5.2. Calibration and fitting of 7-azaindole 213

1.6. Additional hyperspaces of 7-azaindole 217

1.7. 5-bromo-7-azaindole 222

1.7.1. Characterizations and Reference Spectra for 5-bromo-7-azaindole Quantification 223

1.7.2. Calibration and fitting of 5-bromo-7-azaindole 233

1.8. Additional hyperspaces of 5-bromo-7-azaindole 237

1.9. 1H-indazole 239

1.9.1. Characterizations and Reference Spectra for 1H-indazole Quantification 241

1.9.2. Calibration and fitting of 1H-indazole 247

1.10. Additional hyperspaces of 1H-Indazole 251

1.11. NMRanger accuracy validation experiments 261

1.11.1. Morpholine validations 261

1.11.2. 7-azaindole validations 261

1.11.3. 5-bromo7-azaindole validations 270

1.11.4. 1H-indazole validations 271

1.11.5. Accuracy analysis of whole validations 272

1.12. Confirmation of chemical shift and product assignment by NMR titration 275
1.13. Determination of observable protons after copper complexation 277
1.14. DBU reference for azaindoles 279
2. Other NMR software available 282
2.1. Benchmark against Magnetstein and BATMAN 283
3. NMRanger's Manual 285
3.1. Requirement 285
3.2. Parameters 285
3.2.1. General Parameters 285
3.2.2. Substrate-specific Parameters 289
3.2.3. Workflow 291
3.3. Illustrated tutorial 293
3.3.1. Preparation of the environment. 293
3.3.2. Interacting with NMRanger. 295
3.3.3. Demonstration of single-range analysis with NMRanger. 297
3.3.4. Demonstration of multiple-range analysis with NMRanger. 302
3.3.5. Preparing a set of parameters for a new substrate. 303
4. NMRanger Parameter Selection for Chan-Lam Coupling 314
4.1. General substrate-specific parameters selection 314
4.2. Additional substrate-specific treatments 317
5. Robowski-NMR Acquisition Parameters Selection 326
5.1. T_1/T_2 relaxation measurements 326
6. References 332

1. Experimental

1.1. General Information

All compounds and solvents used for synthesis were purchased from Sigma-Aldrich, TCI, or BLD pharma and were used as purchased without purification. HPLC grade 1,2-dichloroethane (1,2 DCE) was used as the reaction solvent. For substrates morpholine, 7-azaindole, 5-bromo-7-azaindole, and 1H-indazole, the Magritek Spinsolve 80 MHz benchtop NMR spectrometer was used to measure the ^{1}H NMR spectra of all reaction crudes using our Robowski-NMR platform. Chemical shift values are reported in parts per million (ppm) using the non-deuterated solvent signal of 1,2-Dichloroethane (δ = 3.73 ppm) as the internal reference. The ^{1}H, ^{13}C, and 2D NMR spectra of isolated products (in $CDCl_3$) were recorded on 400 MHz FT– NMR Bruker Avance III HD spectrometer. The solvent residual signals of $CDCl_3$ were: δ = 7.26 ppm in the ^{1}H NMR spectra and δ = 77.16 ppm in the ^{13}C NMR spectra. High-resolution mass spectra (HRMS) were obtained on a JEOL AccuTOF 4G+ DART mass spectrometer using ESI ionization. Chromatographic purifications were carried out using either recycling preparative HPLC or medium-pressure liquid chromatography (MPLC). Recycling preparative HPLC was performed on a JAI LaboACE LC-7080 Plus equipped with a RID (RI-700 LA) and UV–Vis (UV–VIS4ch 800LA) detectors and preparative Gel Permeation Chromatography (GPC) columns JAIGEL-2HR Plus and JAIGEL-2.5HR Plus (size 20.0 mm ID × 600 mm each). MPLC was performed using a Biotage® Selekt flash chromatography system with Biotage® Sfär Silica 60 µm prepacked flash chromatography columns. Solid deposition/dry loading was performed using Biotage® Sfär DLV Empty 10 or 25 g Columns with frits.

1.2. Robowski Reaction Setup, NMR Quantification, and Product Characterization

1.2.1. General Methods: Robowski reaction setup, Robowski-NMR measurements, isolations, and reference spectra recording.

A. Reaction set-up. Reaction mixtures were prepared by automated liquid handling of stock solutions into 2 mL vials using Robowski. First, a 75 or 100 mM $Cu(OAc)_2$ stock solution in water was added to the reaction vials and evaporated in a vacuum oven at 40 °C for 4-8 h, depending on the volume added. After evaporation, 1,2-dichloroethane (DCE) was added, followed by addition of the remaining stock solutions. The stock solutions comprised $PhB(OH)_2$ (250 mM), the N-coupling partner, and, where applicable, a base; the values in parentheses indicate the corresponding stock concentrations. The *N*-coupling partners were morpholine (1000 mM), 7-azaindole (1000 mM), 5-bromo-7-azaindole (100 mM), and 1H-indazole (150 mM). No additional base was used for morpholine reactions, whereas DBU (1000 mM) and pyridine (2000 mM) were used for azaindole and indazole reactions, respectively.

For reactions under air, the vials were sealed directly after preparation. For O_2-saturated conditions, the entire 54-vial reaction plate was sparged simultaneously for 60 s using a 3D-printed multi-nozzle plate fitted over all reaction vials, **Figure S1**, followed immediately by sealing. The `.SLDPRT` file is provided for design modification in SolidWorks. The `.3MF` file can be imported into compatible slicing software, *PrusaSlicer,* where the print parameters are defined, and the printer-ready file is generated. Both files are available from the Zenodo repository: https://zenodo.org/records/21373963. The sealed plates were then placed in the preheated Eppendorf ThermoMixer C at the corresponding temperature and shaken at 400 rpm for the selected reaction time (24, 48, and 72 h, depending on the substrate).

B. Measurements. After completion of the selected reaction time, the plates containing the crude reaction mixtures were transferred directly to Robowski-NMR for automated measurements without further workup (see **Supplementary Video 1**). ^{1}H NMR spectra were acquired using a Magritek Spinsolve 80 MHz benchtop NMR spectrometer under the acquisition parameters described in the **Section 5.1**, which were selected based on the results of T_1/T_2 experiments.

C. Isolations and reference preparation. After the measurements, the crude reaction mixtures from the plates were combined, and the solvent was evaporated. The remaining copper salts were removed by filtration through a short silica pad, which was washed with DCE. The filtrate was then evaporated, and when sufficient material was available, the crude mixture was purified by MPLC or GPC to isolate the coupling product and other compounds.

The isolated products were then used to prepare solutions of known concentration in DCE. These solutions were measured on the 80 MHz benchtop NMR spectrometer to record reference spectra, which were added to the NMRanger reference library and used for automated spectral fitting and quantification.

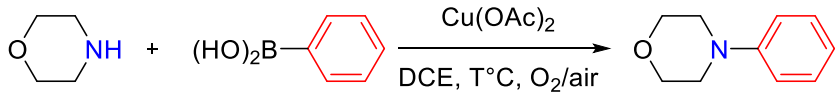


Figure S1. 3D-printed multi-nozzle plate fitted over the 54-vial reaction plate for simultaneous oxygen sparging before sealing. The nozzles are aligned with individual vials, allowing the needles to enter the reaction solutions and deliver oxygen directly into each vial by bubbling.

D. NMRanger quantification. The crude reaction spectra were quantified using NMRanger. For morpholine reactions, fitting was performed over the aromatic regions (≥5.7 and ≤9.5 ppm). This region contained the aromatic signals of coupling products, starting materials (except morpholine), $PhB(OH)_2$, phenol, and diphenyl ether (PhOPh). Representative fitting examples are shown in **Sections 1.3.2, 1.5.2, 1.7.2, 1.9.2.** After processing a given dataset, NMRanger outputs a CSV file containing the estimated integrations and concentrations of all fitted components (see **Section 3** for the **User Manual**). Product yields are calculated in the same file from the estimated product concentration, $C_{i,n}$ of an analyte i in the analyzed range n by:

$Yield\ (\%) = C_{i,n} \ / \ C_{lim,t_0} \times 100$, where C_{lim,t_0} is the initial concentration of the limiting reagent. This CSV file was then submitted to the Hyperspace Viewer[1] to visualize and analyze the reaction outcomes as hyperspace maps.

1.3. Morpholine Chan-Lam Coupling



Reaction design. For morpholine, a six-dimensional reaction space was explored. The reaction variables are $[Morpholine]_0$ (20, 50, 80, and 110 mM)), $Cu(OAc)_2$ loading (0.1–1.1 equiv in 0.2 intervals), $PhB(OH)_2$ loading (1–2 equiv in 0.2 intervals), temperature (26, 40, and 75 °C), atmosphere (air or O_2), and reaction time. For the 40 °C, O_2-saturated conditions, an additional copper loading range of 0.2–0.8 equiv in 0.2 intervals was screened across all morpholine concentrations. For the 40 °C/air conditions, the same additional $Cu(OAc)_2$ loading range was screened, together with expanded $PhB(OH)_2$ loadings of 1.5, 2.2, 2.8, and 3.5 equiv at 20 and 50 mM initial morpholine concentrations.

Following General Method **A**, reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, with a final reaction volume of 1.1 mL. First, a 100 mM $Cu(OAc)_2$ stock solution in water was added to the reaction vials and evaporated in a vacuum oven at 40 °C for 4–8 h. After evaporation, DCE was added, followed by the remaining DCE stock solutions: $PhB(OH)_2$ (250 mM) and morpholine (1000 mM). When required, the reaction mixtures were sparged with oxygen, sealed, and placed in a preheated ThermoMixer C for the required reaction

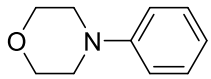
time. Preparation of one 54-vial reaction plate took approximately 30 min. In total, 748 independent reaction conditions were prepared (**Table S1**).

Table S1. Summary of the screening dimensions and included reaction conditions for the morpholine dataset.

Atmosphere	T°C	Time	Screening set	Planned conditions	Excluded outliers	Included conditions
Air	26 °C	72 h	Main screening	4×6×6 = 144	7	137
Air	26 °C	24 h	Main screening	4×6×6 = 144	8	136
Air	40 °C	24 h	Main screening	4×6×6 = 144	0	144
Air	40 °C	24 h	Additional $PhB(OH)_2$ loadings	2×4×5 = 40	1	39
Air	40 °C	24 h	Additional $Cu(OAc)_2$ loadings	2×5×5 = 50	0	50
O_2	40 °C	24 h	Main screening	4×9×6 = 216	10	206
O_2	75 °C	24 h	Main screening	1×6×6 = 36	0	36
Total				**774**	**26**	**748**

After completion of the selected reaction time, the plates containing the crude reaction mixtures were measured by Robowski-NMR, General Method **B**. Then crude mixtures were combined for purification on MPLC as described in the General Method **C**. The isolated compounds were then measured in DCE in known concentrations (**Figure S3-Figure S7**) and added to the NMRanger basis reference set and used for the NMRanger quantification, General Method **D**.

1.3.1. Characterizations and Reference Spectra for Morpholine Quantification



N-phenylmorpholine[2]: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.34 – 7.24 (m, 2H), 6.98 – 6.86 (m, 3H), 3.90 – 3.86 (m, 4H), 3.33 – 2.97 (m, 4H).

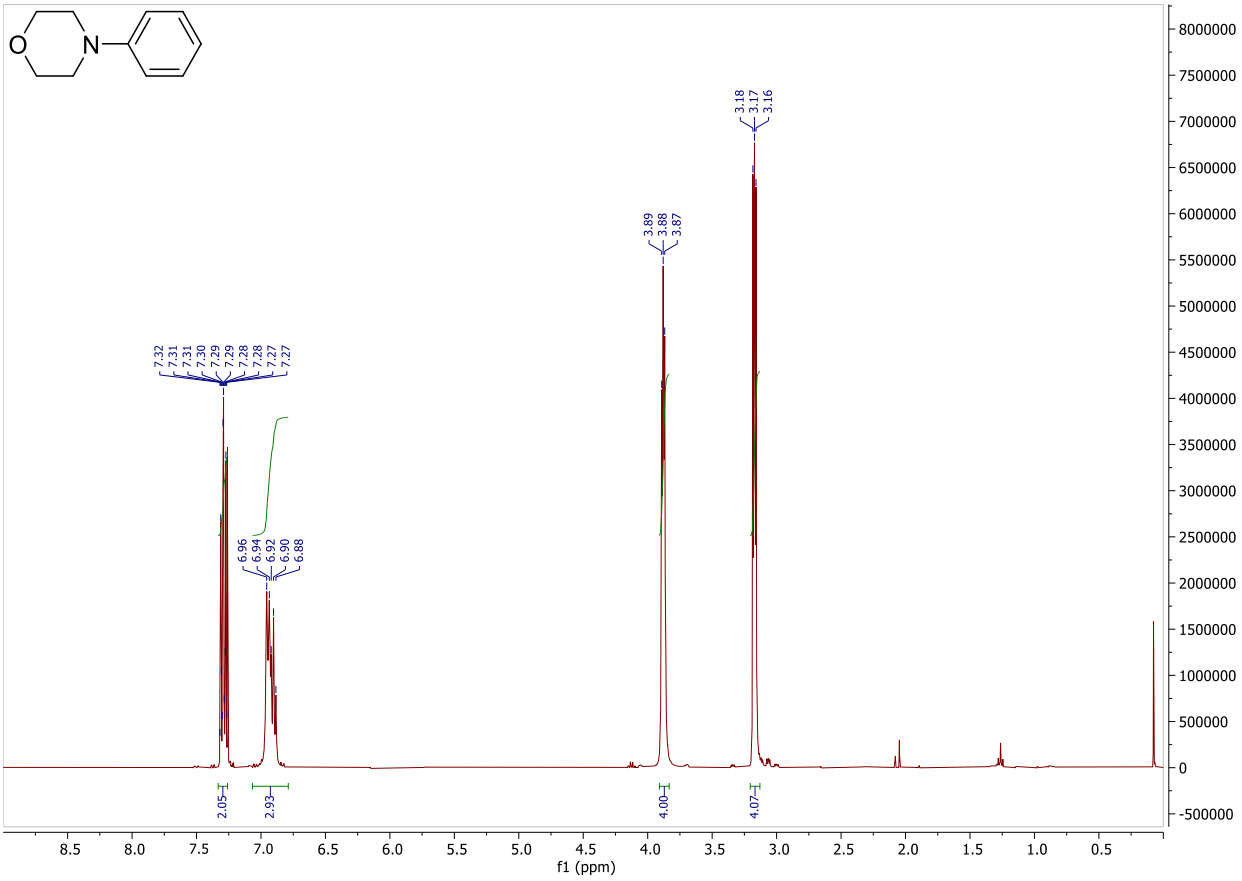


Figure S2. ^{1}H NMR spectrum of 4-phenylmorpholine in $CDCl_3$.

Reference Spectra preparation in DCE. Isolated N-phenylmorpholine was measured in DCE for the reference library. For phenylboronic acid, benzene, phenol, and diphenyl ether, commercially available compounds were used. All spectra were recorded on an 80 MHz NMR spectrometer with 32 scans, a 6.4 s acquisition time, a 10 s repetition time, and a 30° pulse angle (see **Section 5.1** for parameter selection).

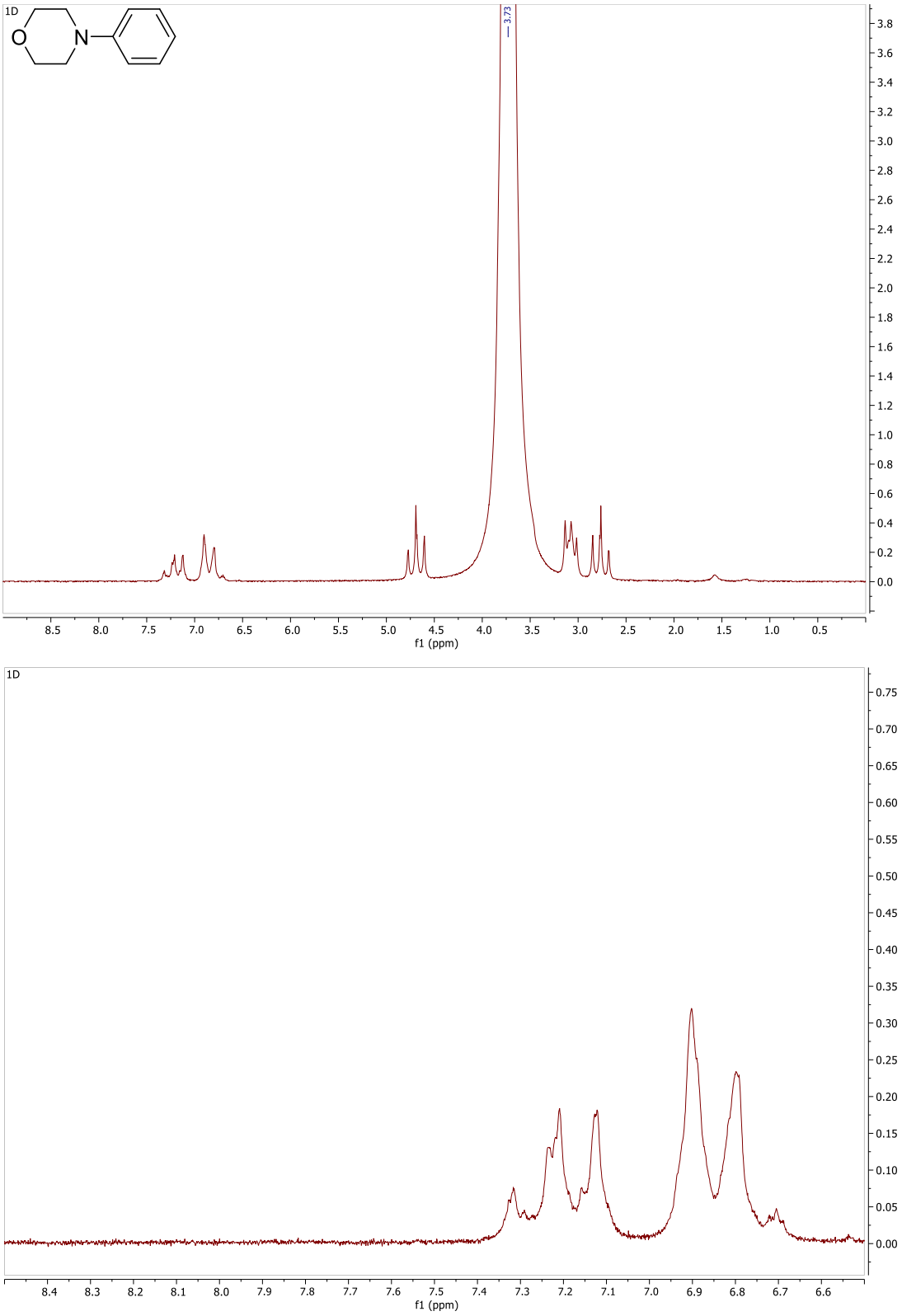


Figure S3. ^{1}H NMR spectrum of N-phenylmorpholine in DCE at 90 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.5–8.5 ppm, (*bottom*) used for NMRanger quantification.

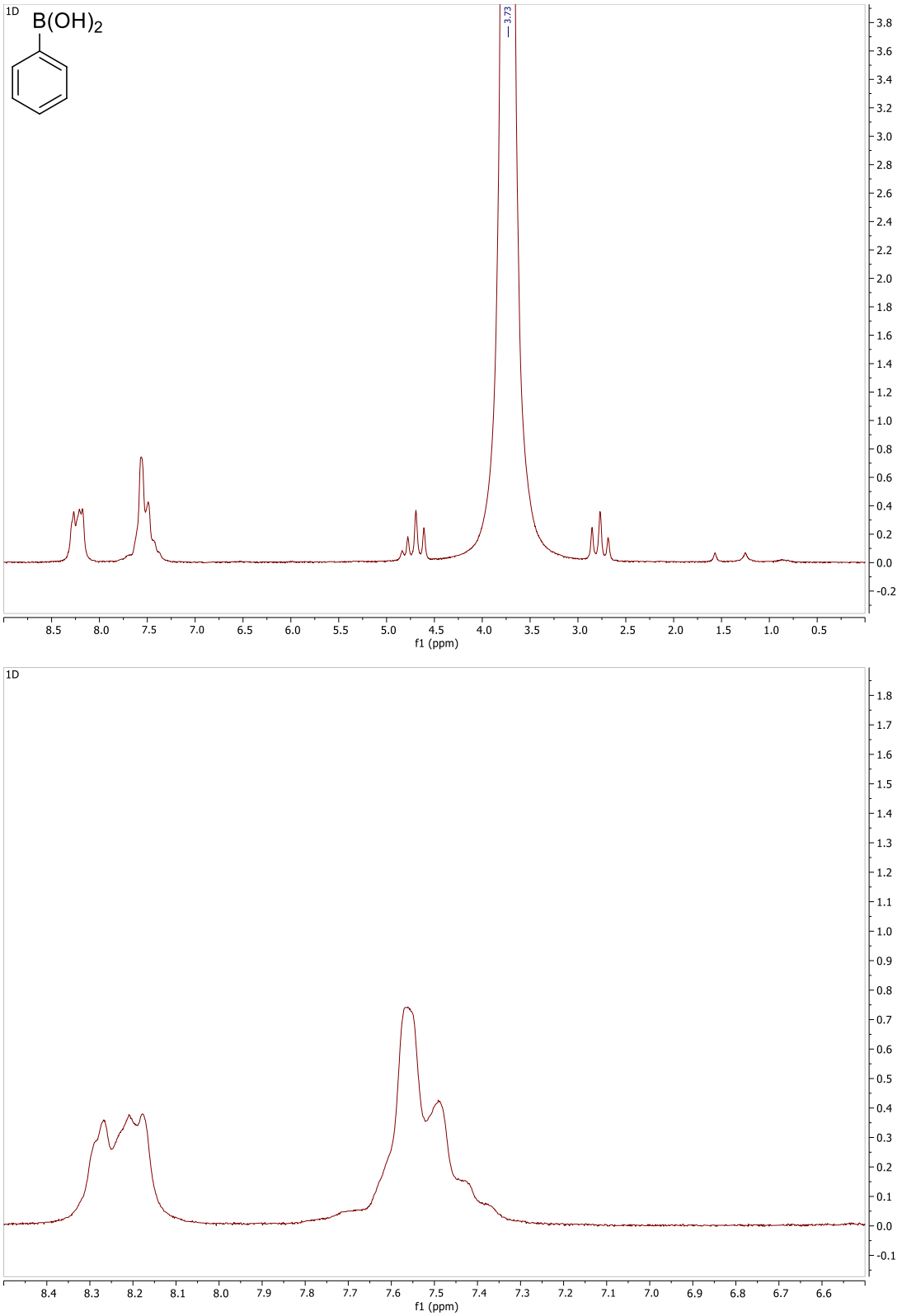


Figure S4. ^{1}H NMR spectrum of phenyl boronic acid in DCE at 250 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.5–8.5 ppm, (*bottom*) used for NMRanger quantification.

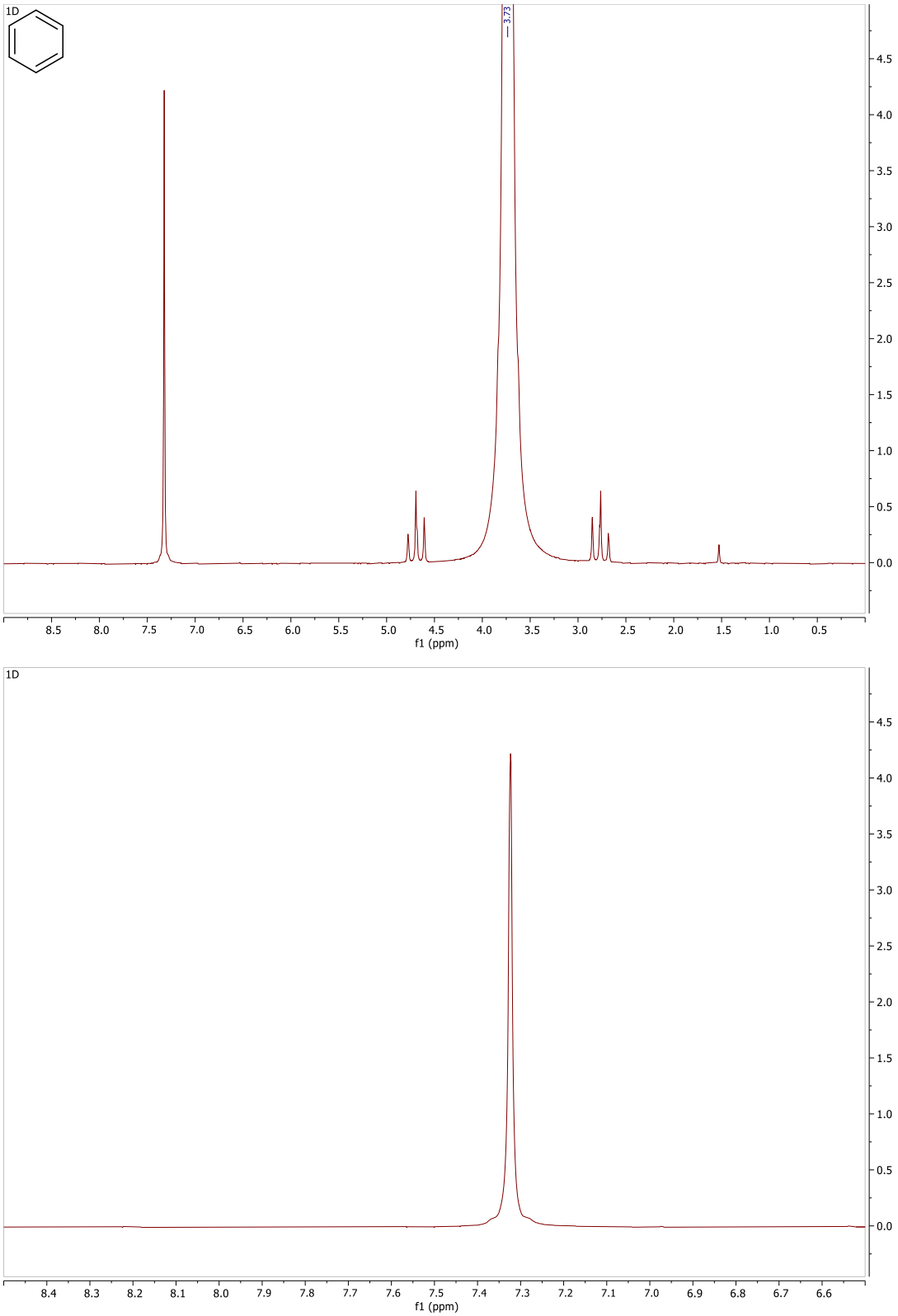


Figure S5. ^{1}H NMR spectrum of benzene in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.5–8.5 ppm, (*bottom*) used for NMRanger quantification.

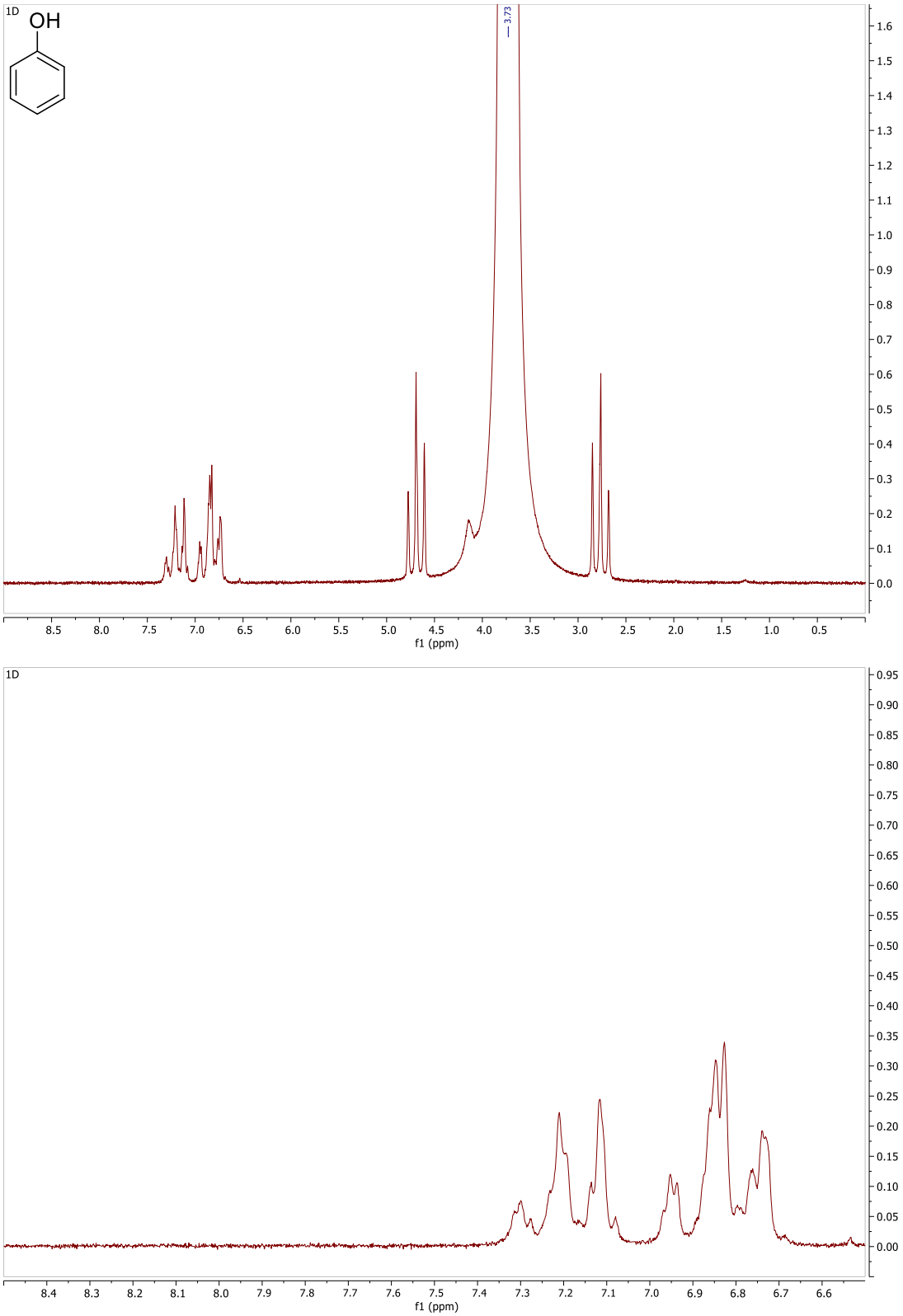


Figure S6. ^{1}H NMR spectrum of phenol in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.5–8.5 ppm, (*bottom*) used for NMRanger quantification.

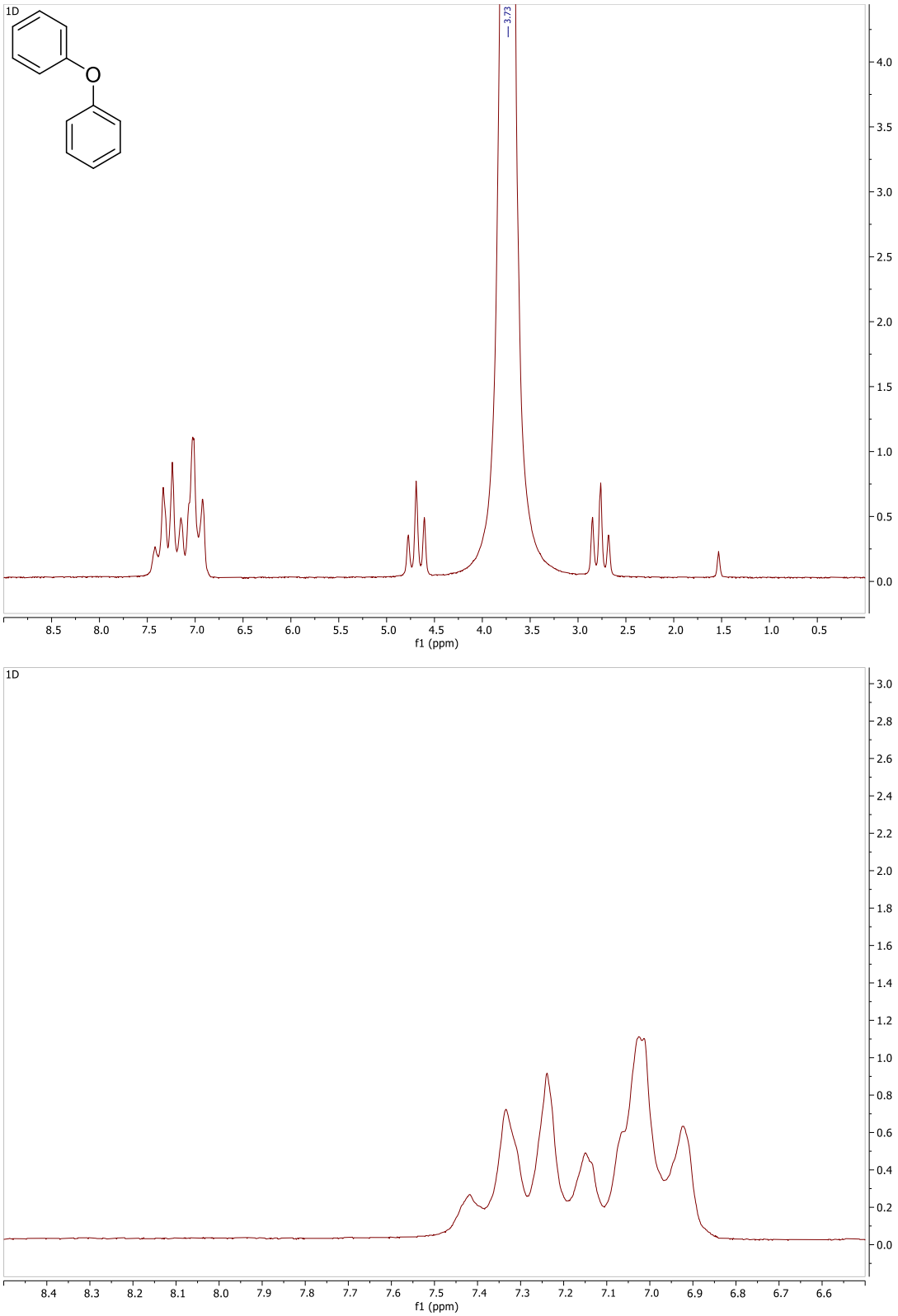


Figure S7. ^{1}H NMR spectrum of diphenyl ether in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.5–8.5 ppm, (*bottom*) used for NMRanger quantification.

1.3.2. Calibration and fitting of morpholine

A calibration curve of N-phenylmorpholine was generated automatically by NMRanger using calibrant solutions prepared in DCE at known concentrations and added in basis set (**Figure S8**). These calibrants were used as external standards for estimated yield calculations. These reference spectra were then used for single-region NMRanger fitting of crude morpholine reaction mixtures over the 6.5–8.5 ppm aromatic region, where N-phenylmorpholine, $PhB(OH)_2$, benzene, phenol, and diphenyl ether signals overlap. Representative fits are shown in **Figure S9**, illustrating reconstruction of spectra with different product/by-product compositions before conversion into the morpholine hyperspaces as discussed in **Section 4.1 of the main text**.

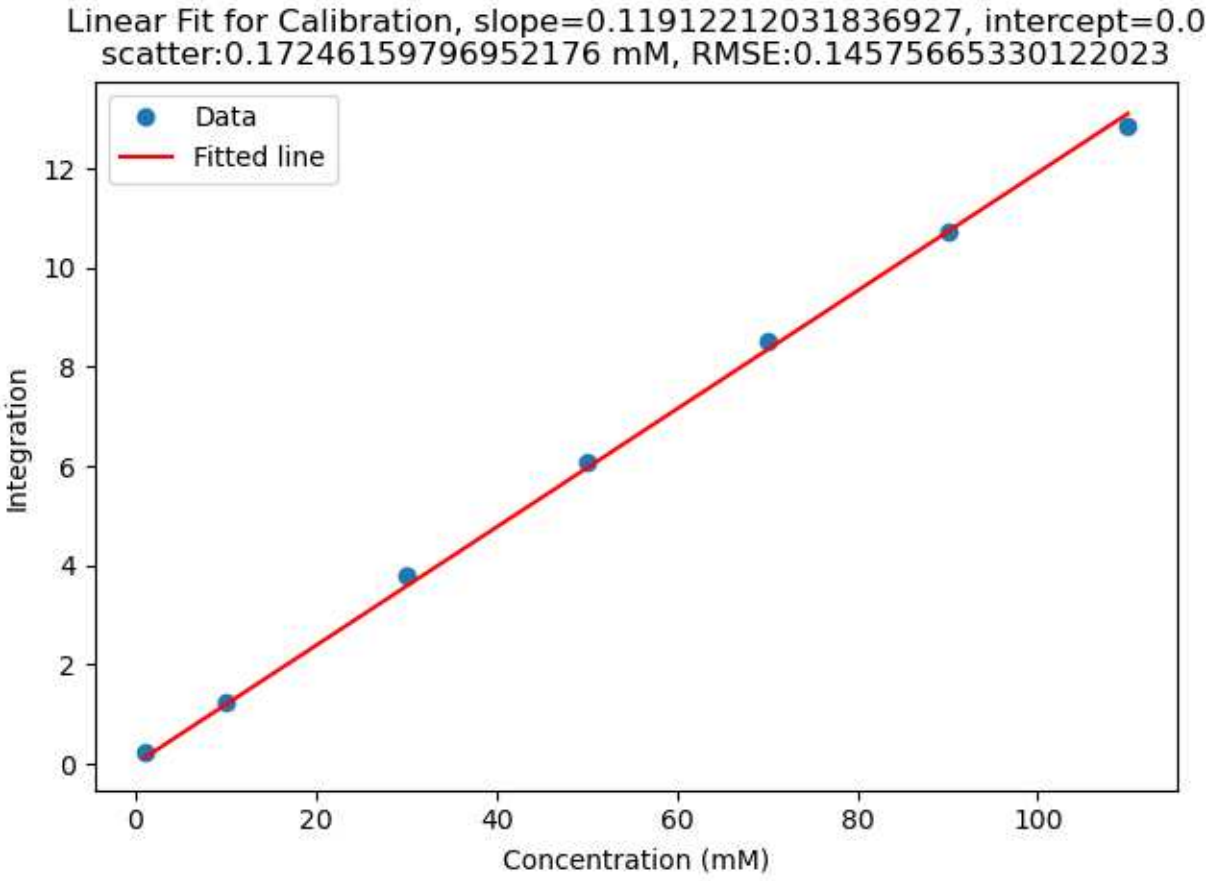


Figure S8. Calibration curve of N-phenylmorpholine generated automatically by NMRanger. Solutions containing known concentrations of N-phenylmorpholine were prepared in DCE at 1, 10, 30, 50, 70, 90, and 110 mM. The upper concentration was chosen based on the expected maximum product concentration in the morpholine coupling reactions.

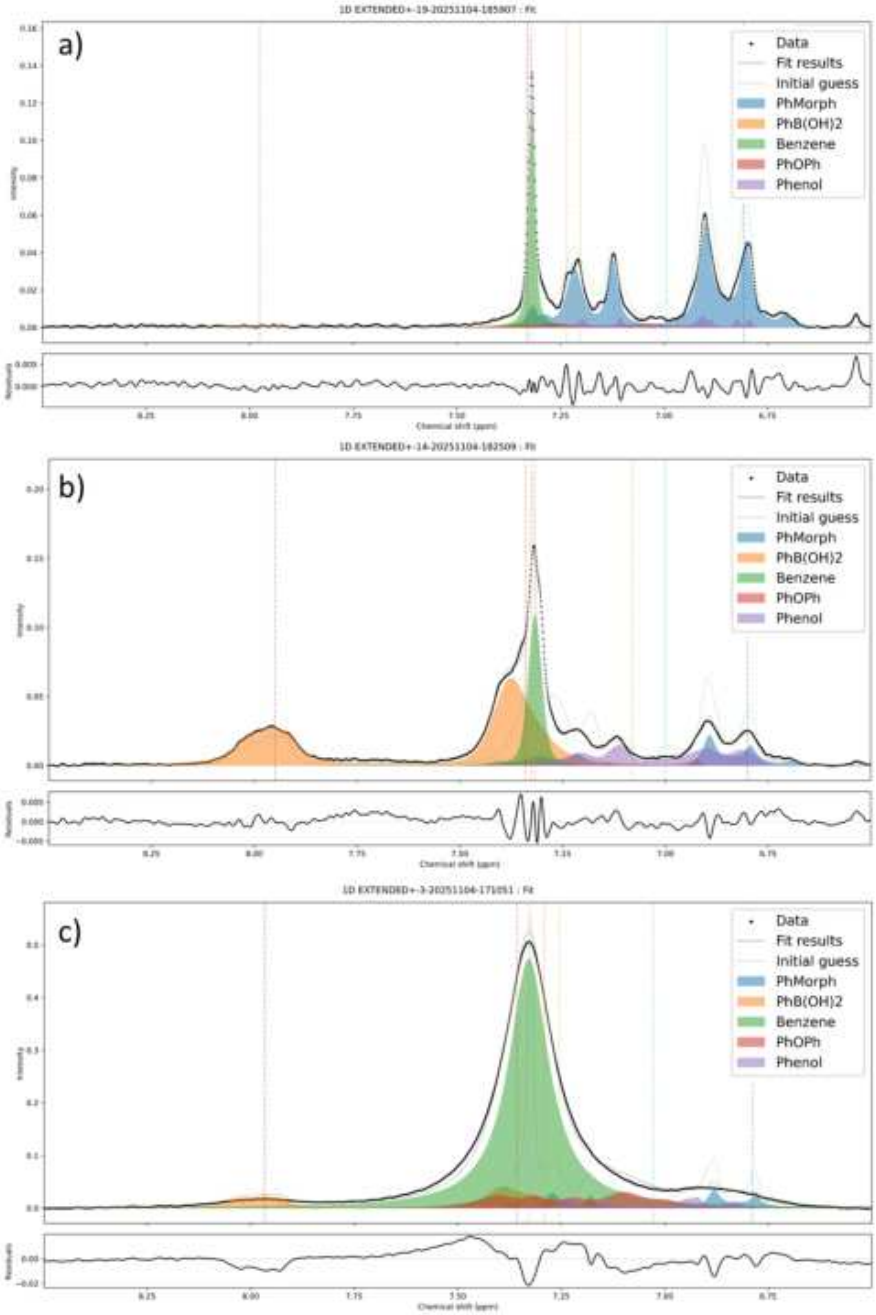


Figure S9. Representative single-region NMRanger fits for morpholine crude reaction mixtures. Fitting was performed over the 6.5–8.5 ppm aromatic region, containing overlapping signals from N-phenylmorpholine (*blue*), $PhB(OH)_2$ (*orange*), benzene (*green*), phenol (*lavender*), and diphenyl ether (*rose*). Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. a) Good fit: Product-rich mixture containing benzene, with close agreement between the fitted and experimental spectra; the small, mostly unstructured residuals indicate accurate reproduction of the peak positions and component contributions. b) More overlapped spectrum containing broadened $PhB(OH)_2$, benzene, lower product formation, and phenol; the main spectral features are reproduced, with only a small mismatch around the broadened $PhB(OH)_2$ resonance and larger residuals in the congested $PhB(OH)_2$/benzene region. c) Poor fit: Benzene-rich spectrum in which the dominant benzene contribution is captured, but mismatch remains around the broadened resonance and distorted baseline.

1.4. Additional hyperspaces of morpholine

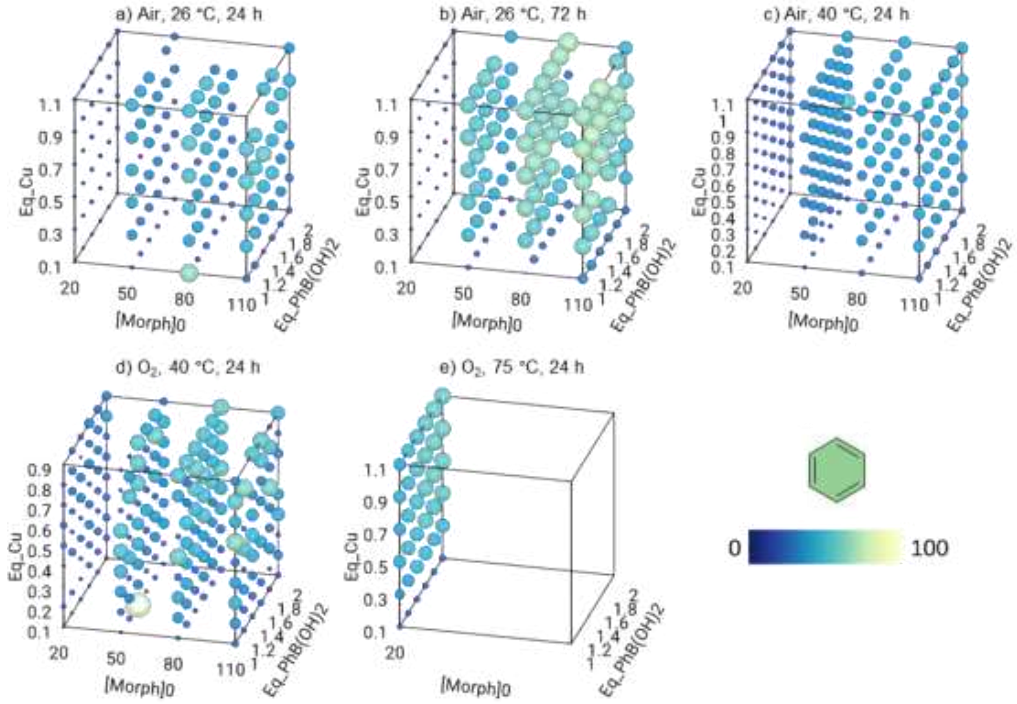


Figure S10. Benzene by-product formation hyperspaces for the morpholine Chan-Lam coupling. a–e) Yield maps under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively.

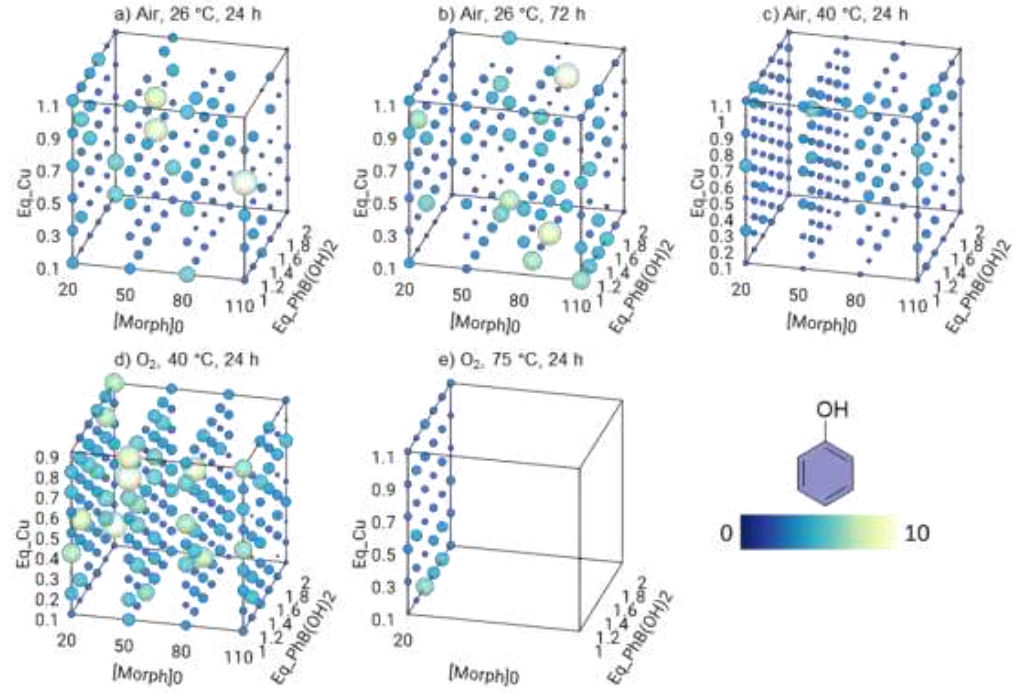


Figure S11. Phenol by-product formation hyperspaces for the morpholine Chan-Lam coupling a-e) Yield maps under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively.

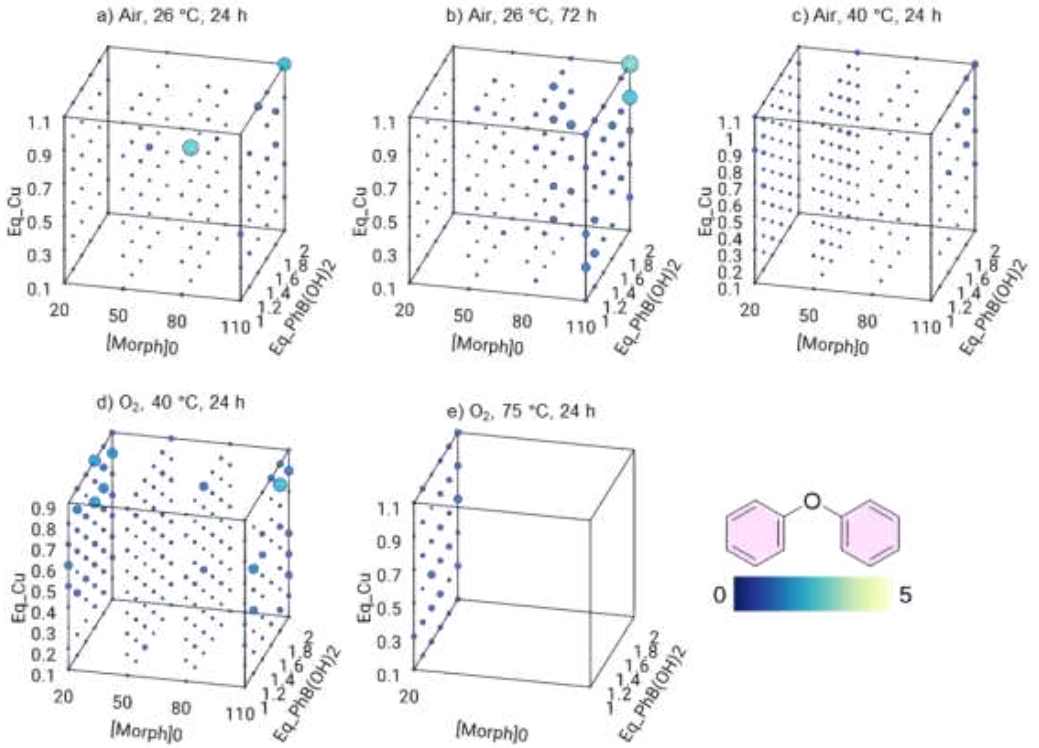


Figure S12. PhOPh by-product formation hyperspaces for the morpholine Chan-Lam coupling a-e) Yield maps under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively.

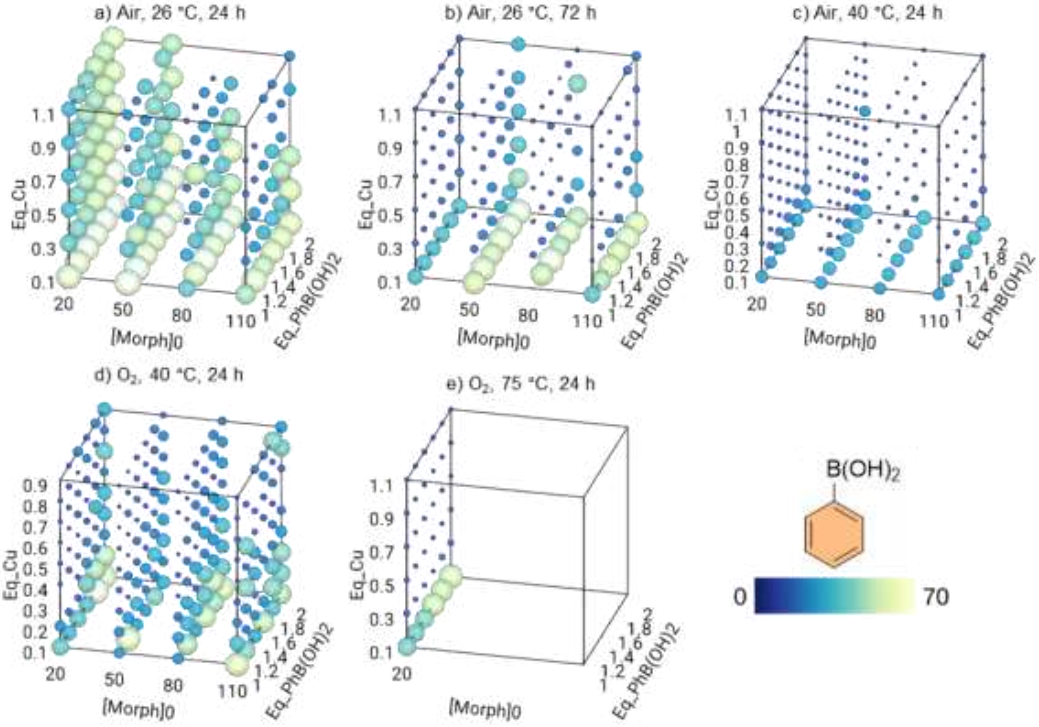


Figure S13. Unconsumed $PhB(OH)_2$ hyperspaces for the morpholine Chan-Lam coupling a-e) Yield maps under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively.

Complete mass balance was not expected in all cases. When evaluated relative to phenylboronic acid (see General Method **D** for calculation details), the aryl source for the coupling products and aryl-derived by-products, the average mass balances are 94, 88, 61, 81, and ~100% for conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O2, 40 °C, 24 h; and O2, 75 °C, 24 h, respectively. (**Figure S14**). The remaining deficit most likely arises, at least in part, from phenylboronic acid precipitation after the reaction, consistent with its limited solubility in DCE and the observed residue on the vial walls.

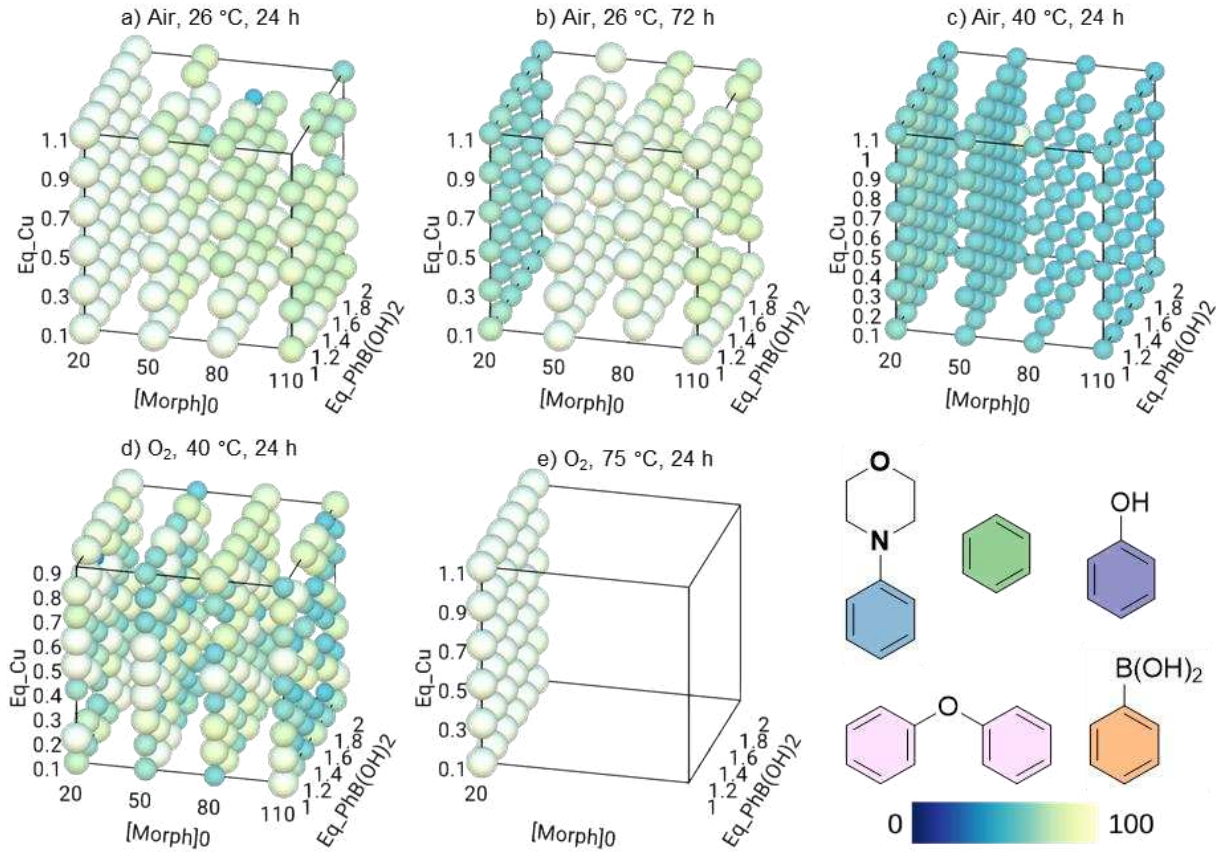
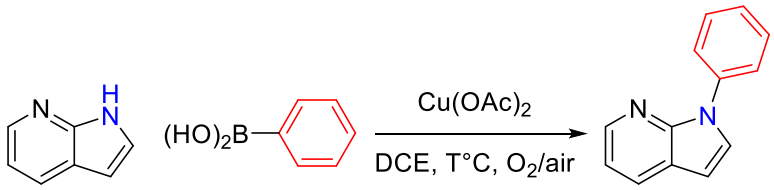


Figure S14. Mass-balance hyperspaces for the morpholine Chan-Lam coupling a-e) Mass balance was determined relative to the initial amount of $PhB(OH)_2$ by summing all quantified phenylboronic acid-derived species: N-phenylmorpholine, benzene, phenol, diphenyl ether, and unconsumed $PhB(OH)_2$. Maps are shown under different reaction conditions: air, 26 °C, 24 h; air, 26 °C, 72 h; air, 40 °C, 24 h; O_2, 40 °C, 24 h; and O_2, 75 °C, 24 h, respectively from (**a**) to (**e**).

1.5. 7-azaindole



Reaction design. For 7-azaindole, a six-dimensional reaction space was explored. The reaction variables are [7-azaindole]$_0$ (20, 50, and 100 mM)), $Cu(OAc)_2$ loading (0.2–1.0 equiv in 0.2 intervals), $PhB(OH)_2$ loading (1.2–2 equiv in 0.2 intervals), temperature (26, 50, and 100 °C), DBU loading (1, 1.5, and 2 equiv), and atmosphere (air or O_2). All reactions are performed for 24 h.

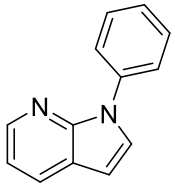
Following General Method **A**, reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, with a final reaction volume of 1.1 mL. First, a 100 mM $Cu(OAc)_2$ stock solution in water was added to the reaction vials and evaporated in a vacuum oven at 40 °C for 4–8 h. After evaporation, DCE was added, followed by the remaining DCE stock solutions: $PhB(OH)_2$ (250 mM), 7-azaindole (1000 mM), and DBU (1000 mM). When required, the reaction mixtures were sparged with oxygen, sealed, and placed in a preheated ThermoMixer for the required reaction time. Preparation of one 54-vial reaction plate took approximately 45 min. In total, 356 independent reaction conditions were prepared (**Table S2**).

Table S2. Summary of the screening dimensions and included reaction conditions for the 7-azaindole dataset.

Atmosphere	T°C	Initial concentration	Screening dimensions	Excluded/missing	Included conditions
Air	26 °C	20 mM	6×5×3 = 90	2 outliers	88
O_2	26 °C	20 mM	5×5×3 = 75	5 missing points	70
Air	50 °C	20 mM	5×5×2 = 50	0	50
O_2	50 °C	20 mM	5×5×2 = 50	0	50
O_2	100 °C	20 mM	5×5×2 = 50	0	50
O_2	26 °C	50 mM	4×3×2 = 24	0	24
O_2	26 °C	100 mM	4×3×2 = 24	0	24
Total			363	7	**356**

After completion of the selected reaction time, the plates containing the crude reaction mixtures were measured by Robowski-NMR, General Method **B**. Then crude mixtures were combined for purification on MPLC, see General Method **C**. The isolated compounds were then measured in DCE in a known concentrations (**Figure S16-Figure S22**) and added to the NMRanger basis reference set and used for the NMRanger quantification, General Method **D**.

1.5.1. Characterizations and Reference Spectra for 7-azaindole Quantification



N-phenyl-7-azaindole[3]**: ^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.39 (dd, *J* = 4.7, 1.6 Hz, 1H), 7.99 (dd, *J* = 7.8, 1.6 Hz, 1H), 7.80 – 7.71 (m, 2H), 7.58 – 7.49 (m, 3H), 7.39 – 7.30 (m, 1H), 7.15 (dd, *J* = 7.8, 4.7 Hz, 1H), 6.64 (d, *J* = 3.6 Hz, 1H); **^{13}C NMR** (101 MHz, CDCl3) δ 147.3, 143.5, 138.5, 129.4, 129.2, 127.9, 126.4, 124.0, 121.7, 116.7, 101.7.

Reference Spectra preparation in DCE. Isolated N-phenyl-7-azaindole was measured in DCE for the reference library. For phenylboronic acid, DBU, benzene, phenol, and diphenyl ether, commercially available compounds were used. All spectra were recorded on an 80 MHz NMR spectrometer with 32 scans, a 6.4 s acquisition time, a 15 s repetition time, and a 90° pulse angle (see **Section 5.1** for parameter selection).

DBU resonances were located mainly in the aliphatic region beyond the broad solvent signal. However, additional signals were observed between 5.0 and 6.5 ppm and assigned to vinyl chloride formed by DBU-induced elimination of DCE (see **Section 1.14** for characterizations). One proton resonance from vinyl chloride falls within the selected fitting window and was therefore included in the reference basis set.

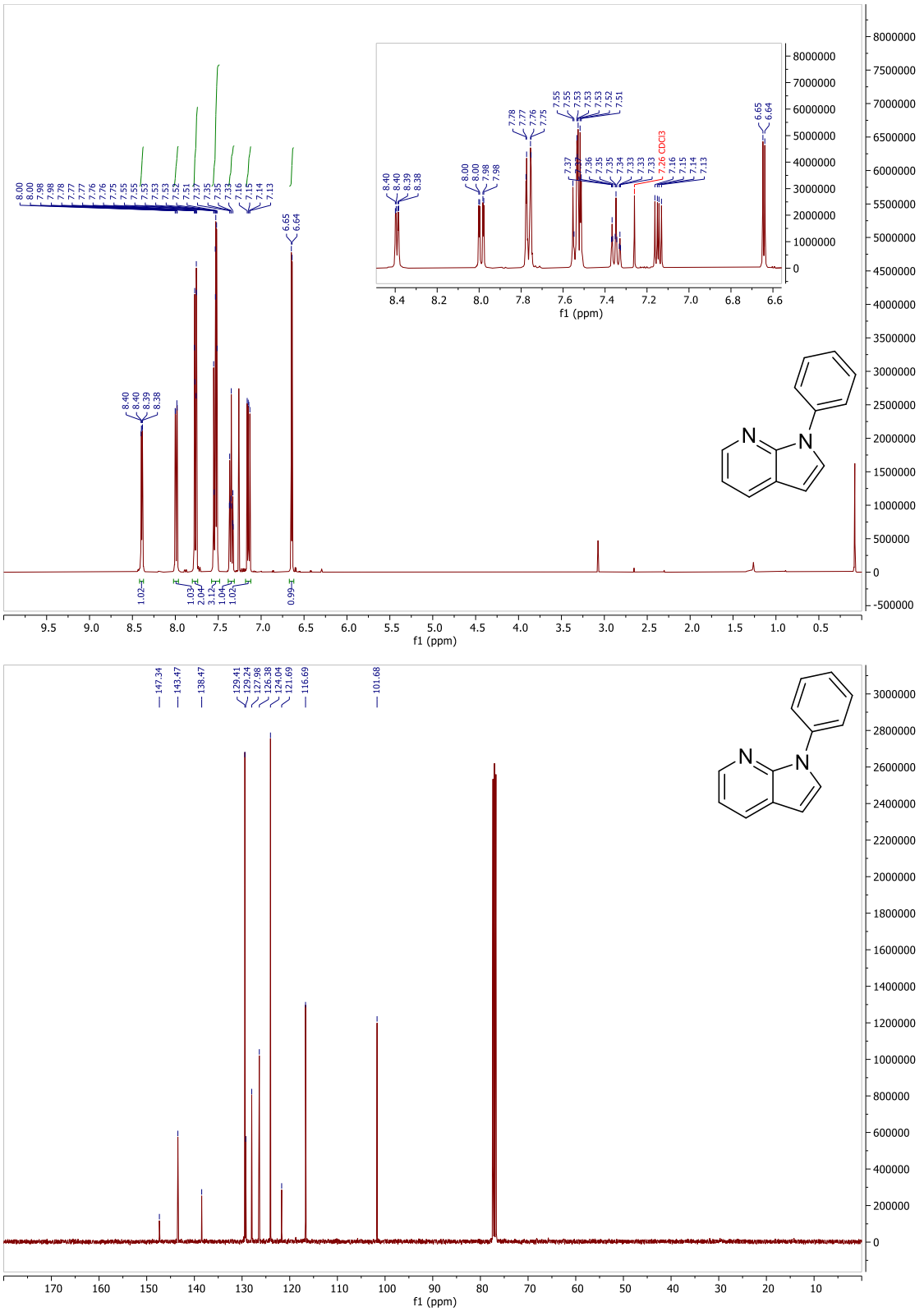


Figure S15. ^{1}H NMR (*top*) and ^{13}C NMR (*bottom*) spectra of N-phenyl-7-azaindole in $CDCl_3$.

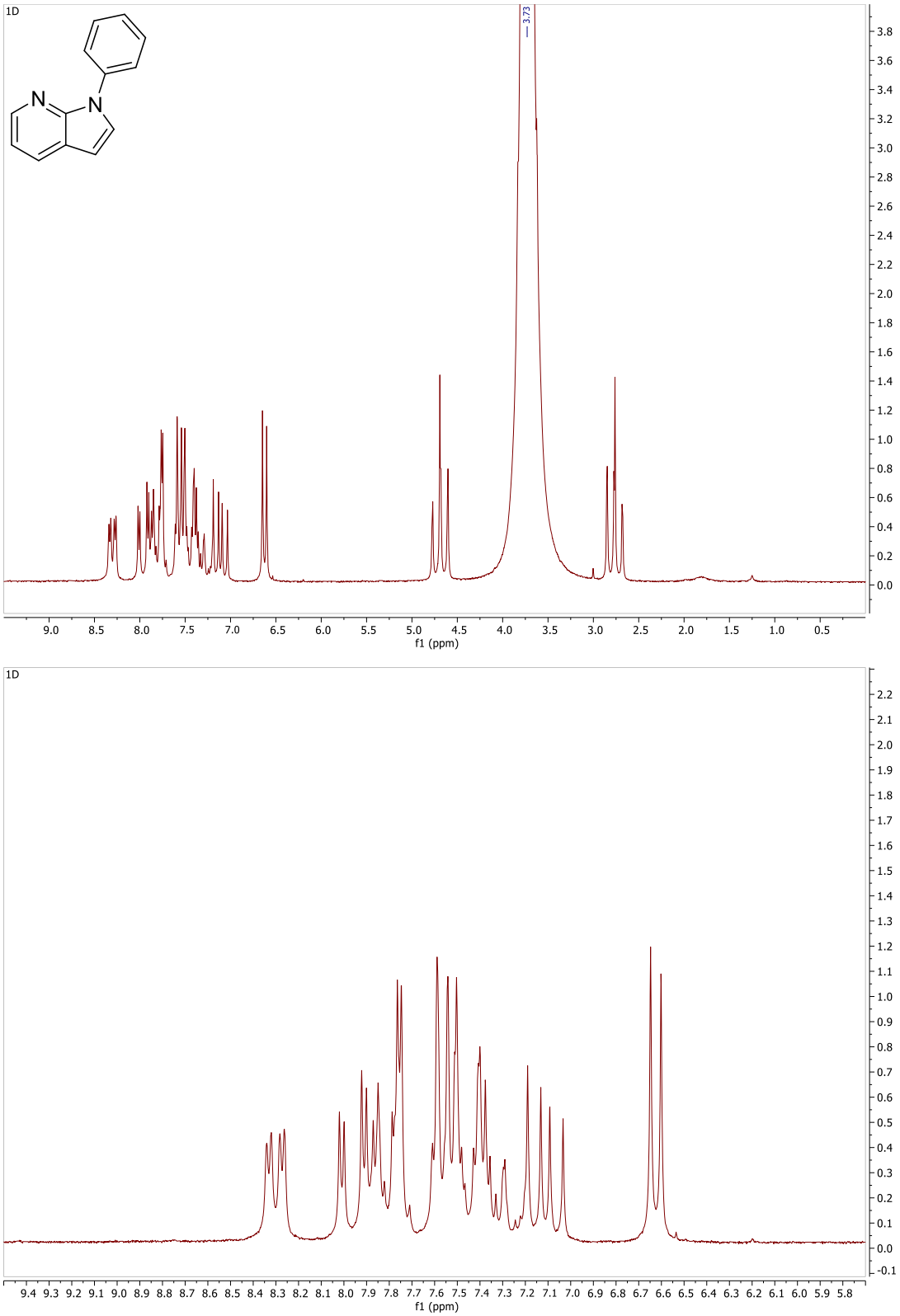


Figure S16. ^{1}H NMR spectrum of N-phenyl-7-azaindole in DCE at 150 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

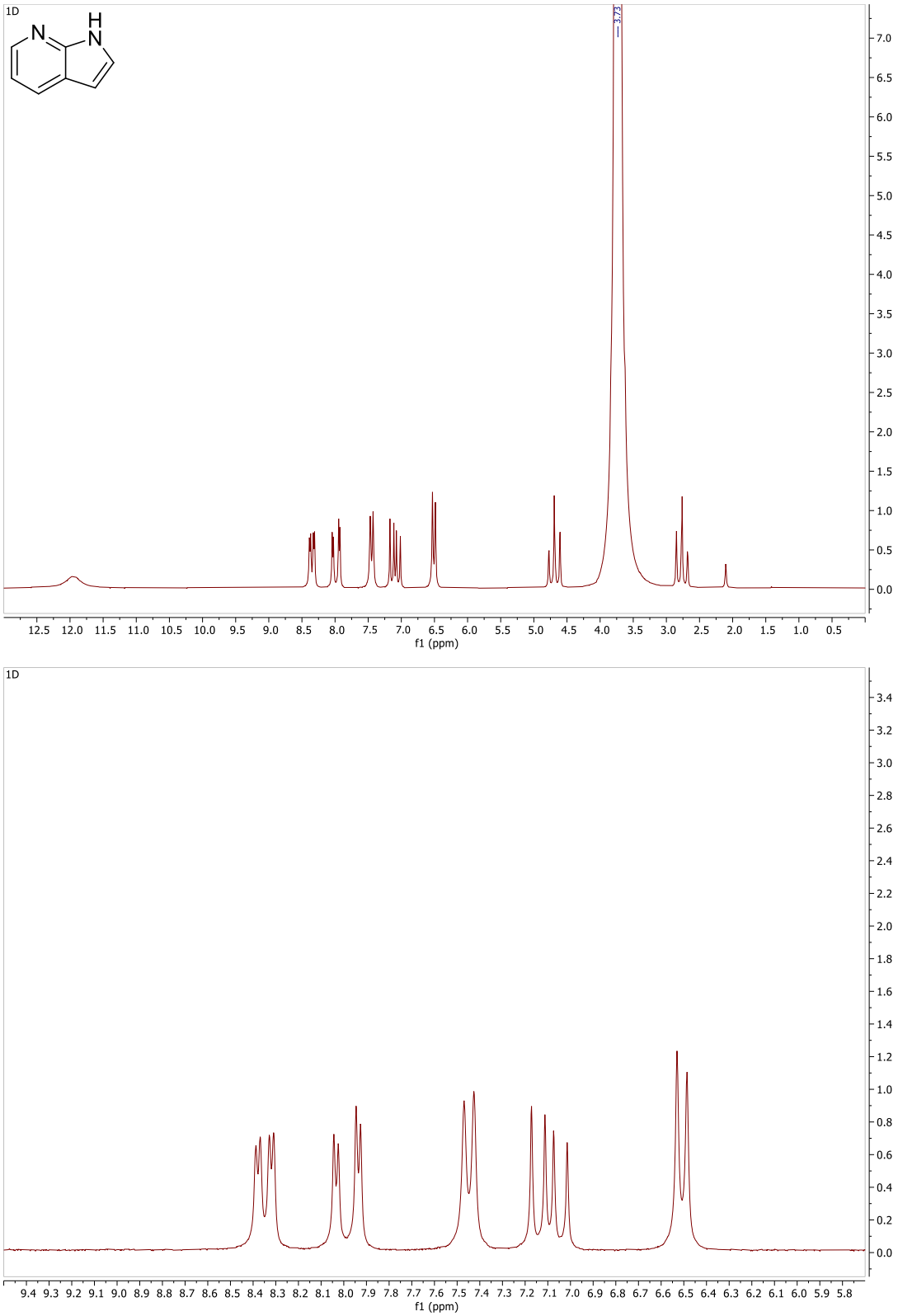


Figure S17. ^{1}H NMR spectrum of 7-azaindole in DCE at 150 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

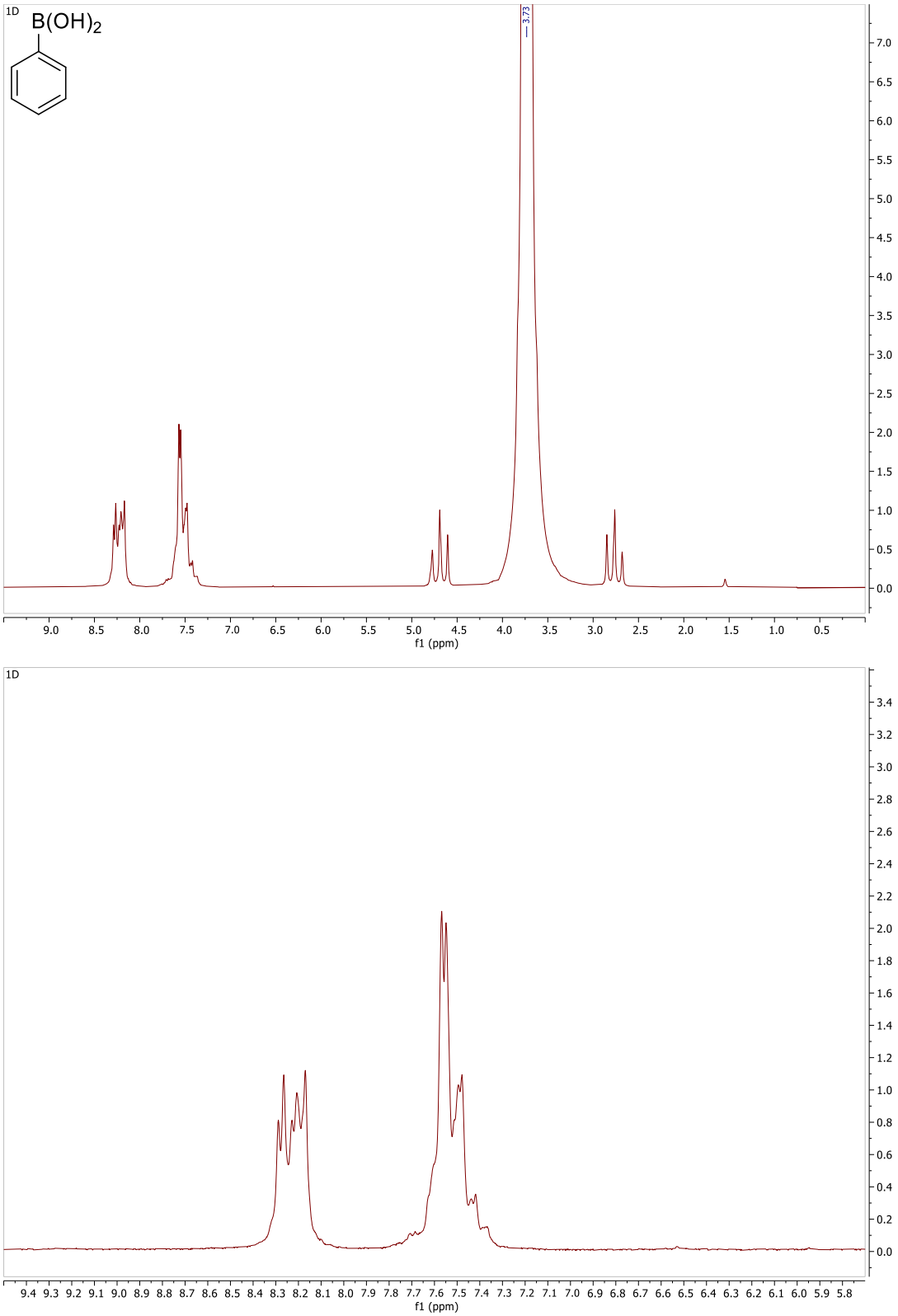


Figure S18. ^{1}H NMR spectrum of $PhB(OH)_2$ in DCE at 250 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

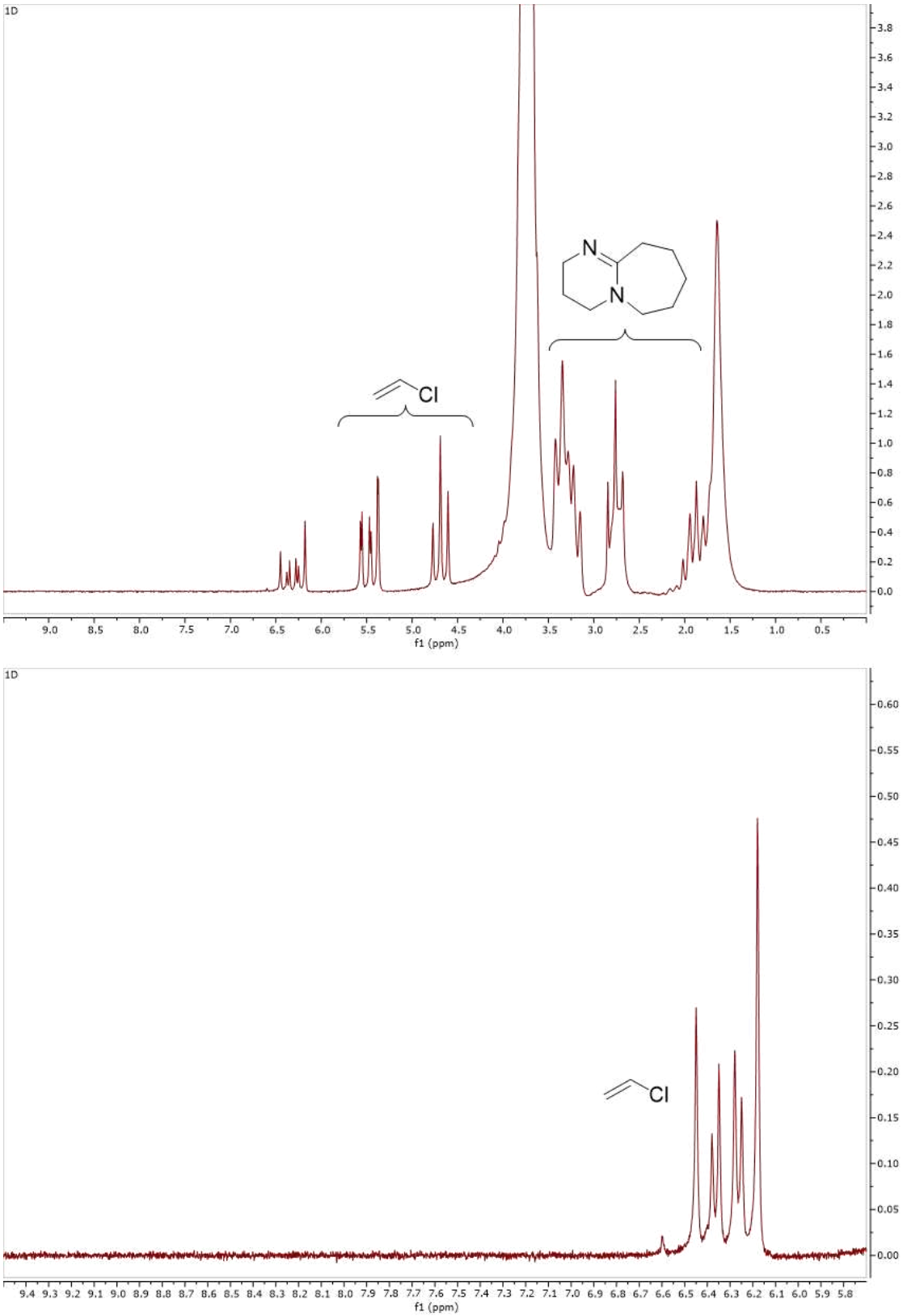


Figure S19. ^{1}H NMR spectrum of DBU including vinyl chloride impurity in DCE at 250 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) contains one proton that belongs to vinyl chloride, used for NMRanger quantification.

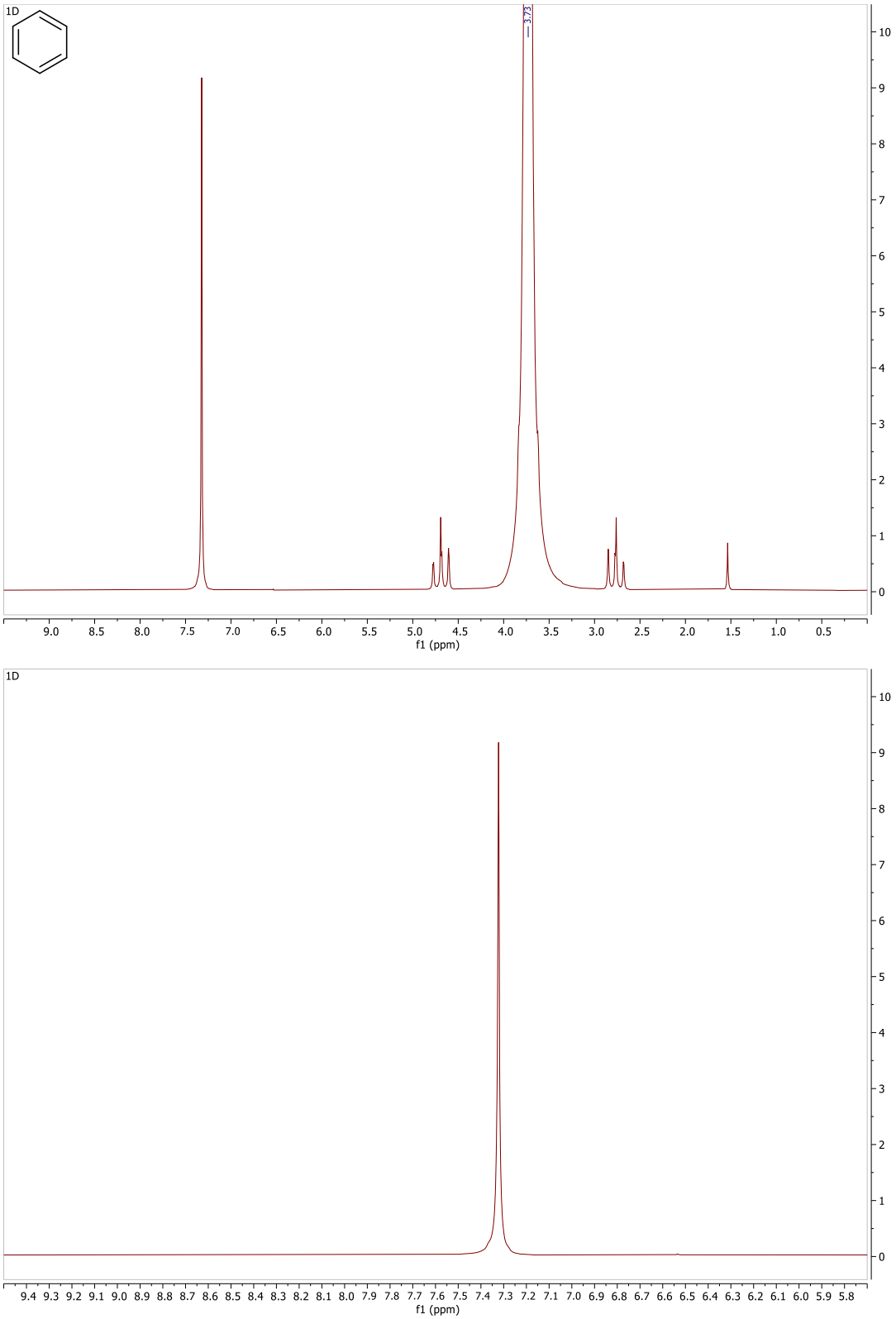


Figure S20. ^{1}H NMR spectrum of benzene in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

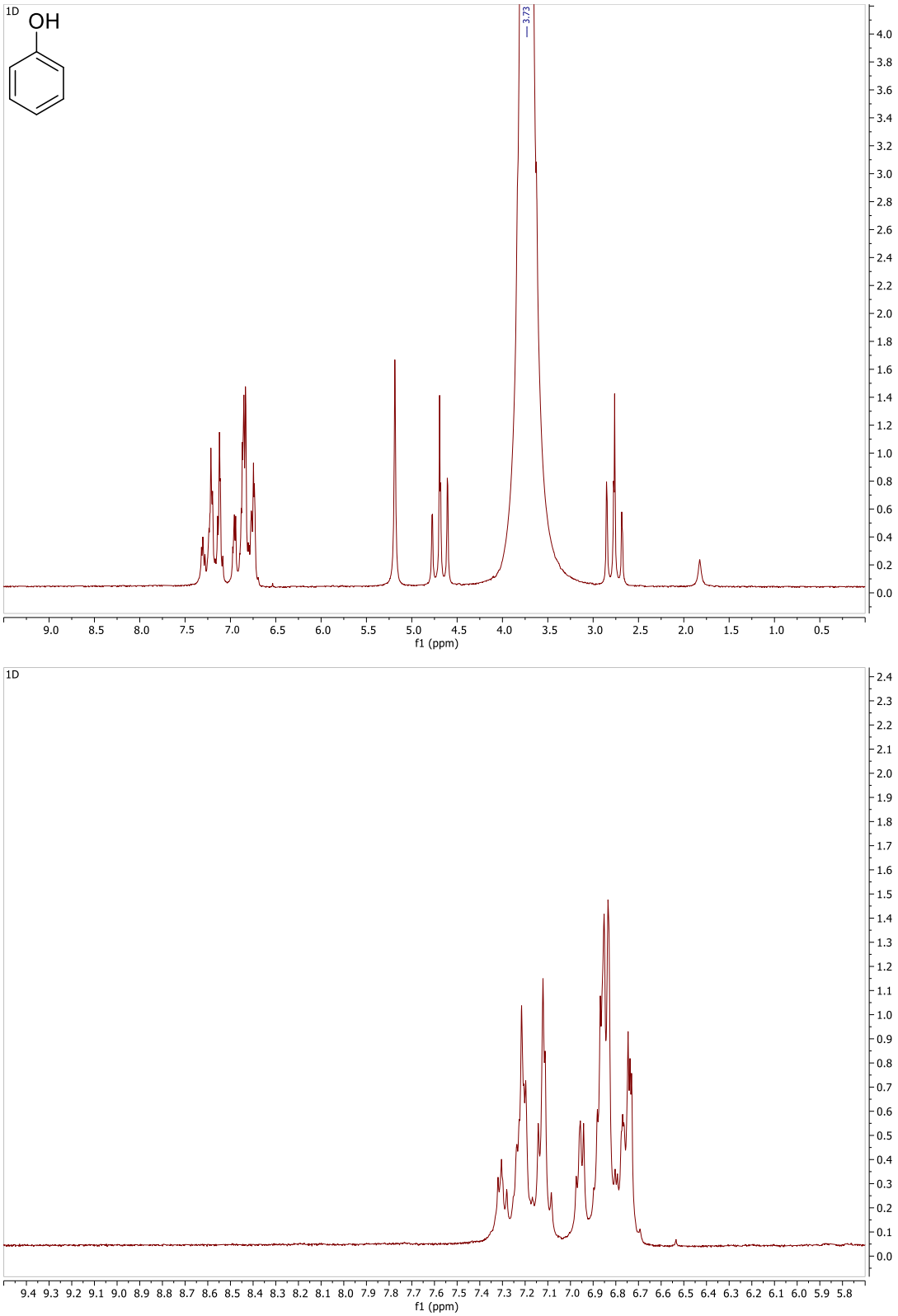


Figure S21. ^{1}H NMR spectrum of phenol in DCE at 200 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

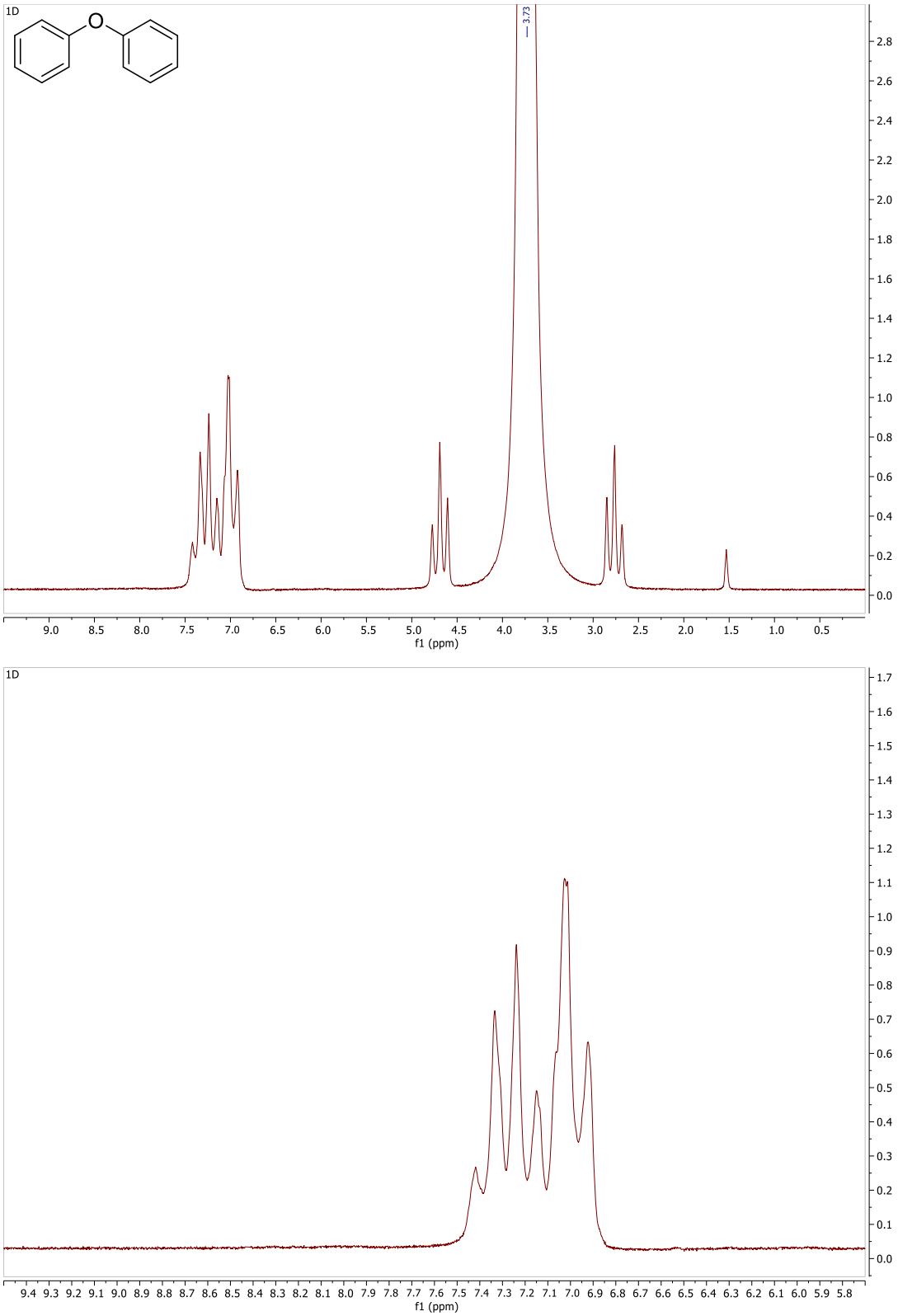


Figure S22. ^{1}H NMR spectrum of PhOPh in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.7–9.5 ppm, (*bottom*) used for NMRanger quantification.

1.5.2. Calibration and fitting of 7-azaindole

A calibration curve of N-phenyl-7-azaindole was generated automatically by NMRanger using calibrant solutions prepared in DCE at known concentrations and added in basis set (**Figure S23**). These calibrants served as external standards for yield estimation. Crude reaction mixtures of 7-azaindole were then quantified using multiple-region NMRanger fitting. The characteristic 5.7–7.0 ppm region, containing signals from N-phenyl-7-azaindole, $PhB(OH)_2$, and phenol was fitted first, and the resulting abundance estimates were transferred to guide fitting of the broader 5.7–9.5 ppm region. Because all compounds are fully aromatic, this broader range contained all reference resonances required for their quantification while excluding the non-deuterated solvent region, which contains no azaindole or phenyl boronic acid-derived compound signals. Representative (good/acceptable/poor) fits are shown in **Figure S24**-**Figure S26**, illustrating spectral reconstruction across samples with different product and by-product compositions before their incorporation into the reaction hyperspaces discussed in **Section 4.2 of the main text**.

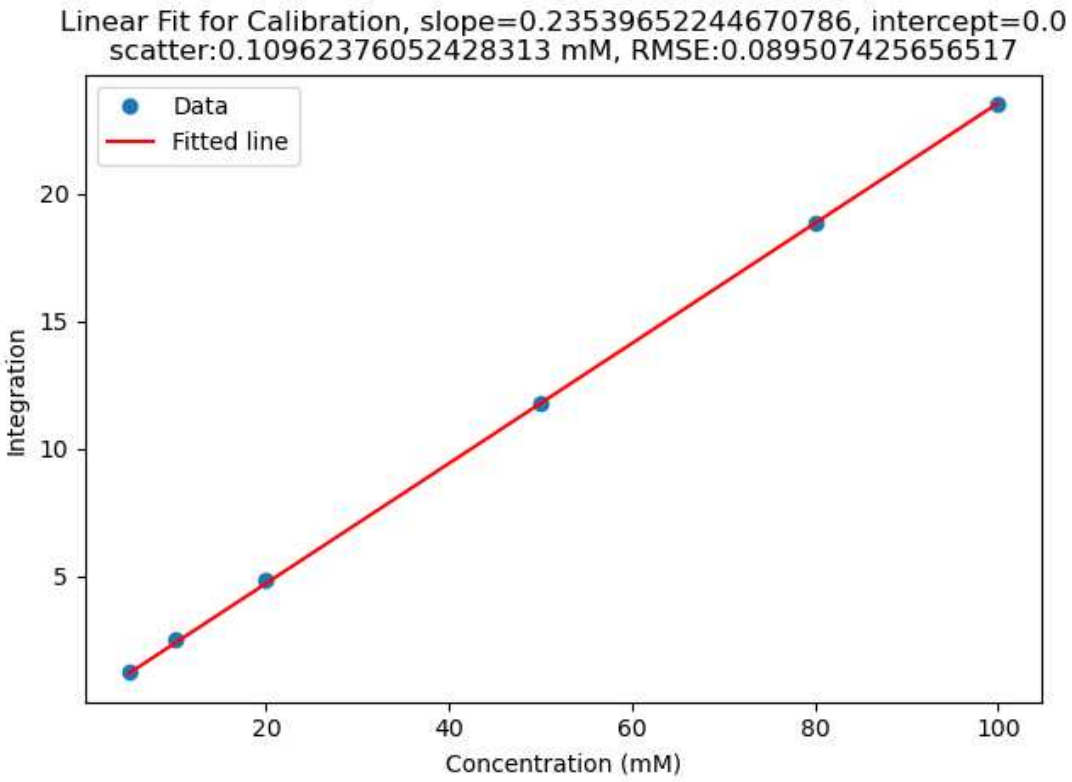


Figure S23. **Calibration curve of N-phenyl-7-azaindole generated automatically by NMRanger.** Solutions containing known concentrations of N-phenyl-7-azaindole were prepared in DCE at 5, 10, 20, 50, 80, 100 mM. The upper concentration was chosen based on the expected maximum product concentration in the morpholine coupling reactions.

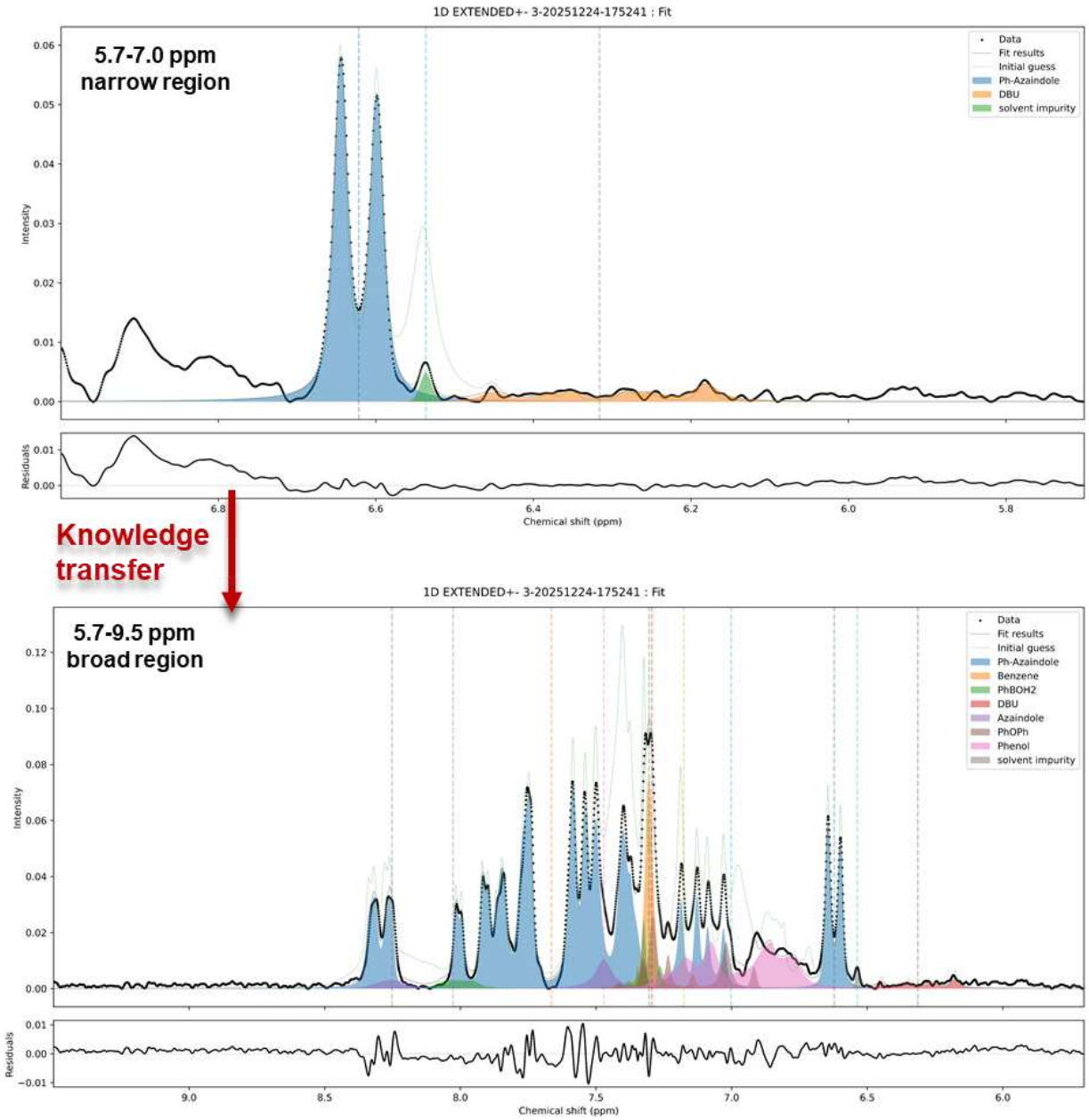


Figure S24. Representative multiple-region NMRanger good fit for 7-azaindole crude reaction mixtures. Fitting was performed first over 5.7–7.0 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 5.7–9.5 ppm region, containing overlapping signals from N-phenyl-7-azaindole (*blue*), 7-azaindole (*purple*), $PhB(OH)_2$ (*green*), benzene (*orange*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 7-azaindole, O_2 atmosphere, 26 °C, in the presence of 0.6 equiv $Cu(OAc)_2$, 1.4 equiv $PhB(OH)_2$, and 1.0 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Good fit for a highly overlapped 7-azaindole crude spectrum; the fitted line closely follows the main peak shapes and intensities across the aromatic region and accurately reproduces the peak positions, with only moderate residuals in the most congested regions.

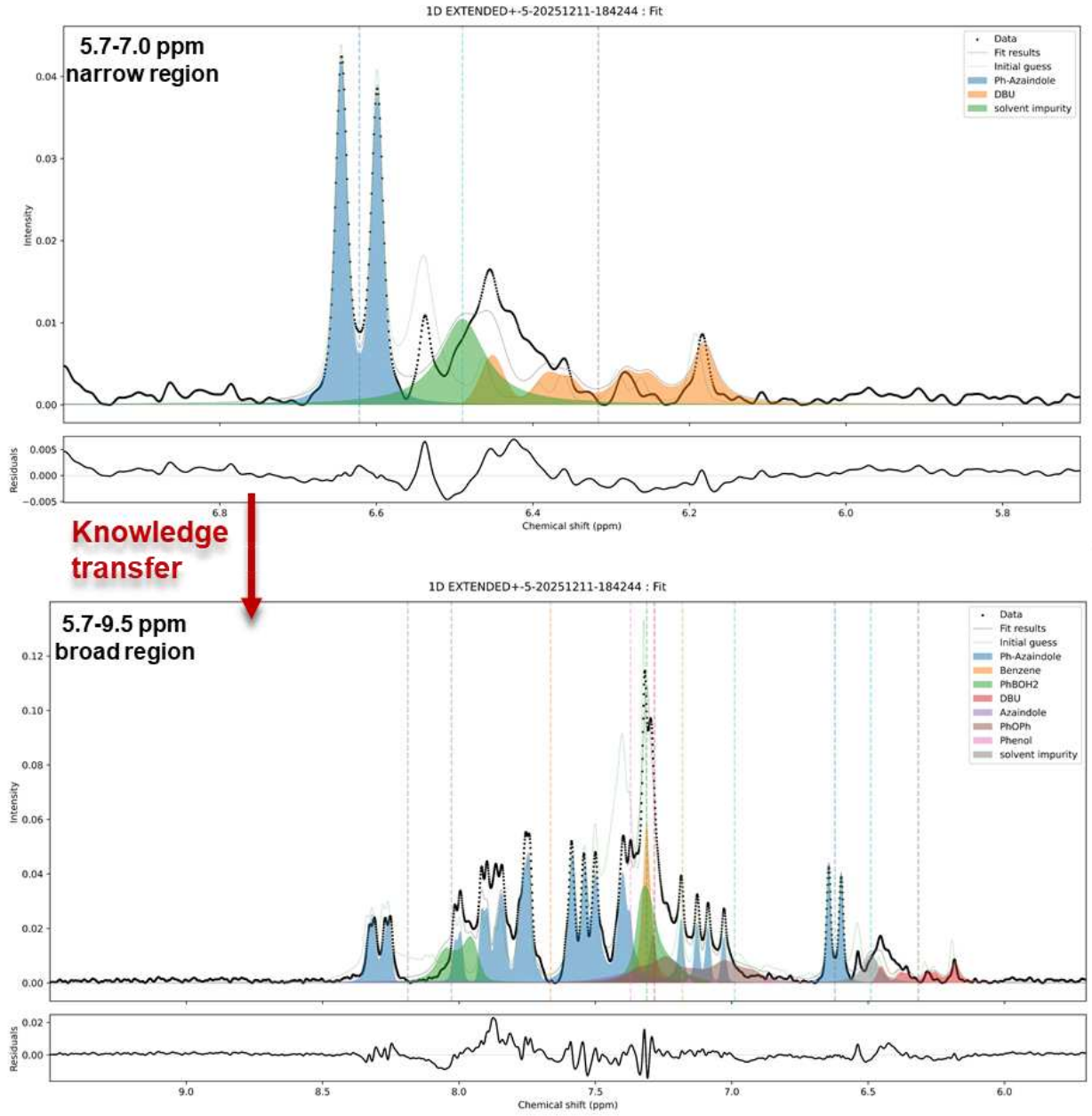


Figure S25. Representative multiple-region NMRanger acceptable fit for 7-azaindole crude reaction mixtures. Fitting was performed first over 5.7–7.0 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 5.7–9.5 ppm region, containing overlapping signals from N-phenyl-7-azaindole (*blue*), 7-azaindole (*purple*), $PhB(OH)_2$ (*green*), benzene (*orange*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 7-azaindole, O_2 atmosphere, 26 °C, in the presence of 0.2 equiv $Cu(OAc)_2$, 1.2 equiv $PhB(OH)_2$, and 1.0 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Acceptable fit: The fitted line shows good agreement with the experimental spectrum for the main N-phenyl-7-azaindole resonances around 6.6–6.8 ppm. The fit remains acceptable in the overlapping 6.4–6.6 ppm region and across the congested 7.2–8.0 ppm range, where only slight mismatches and structured residuals are observed.

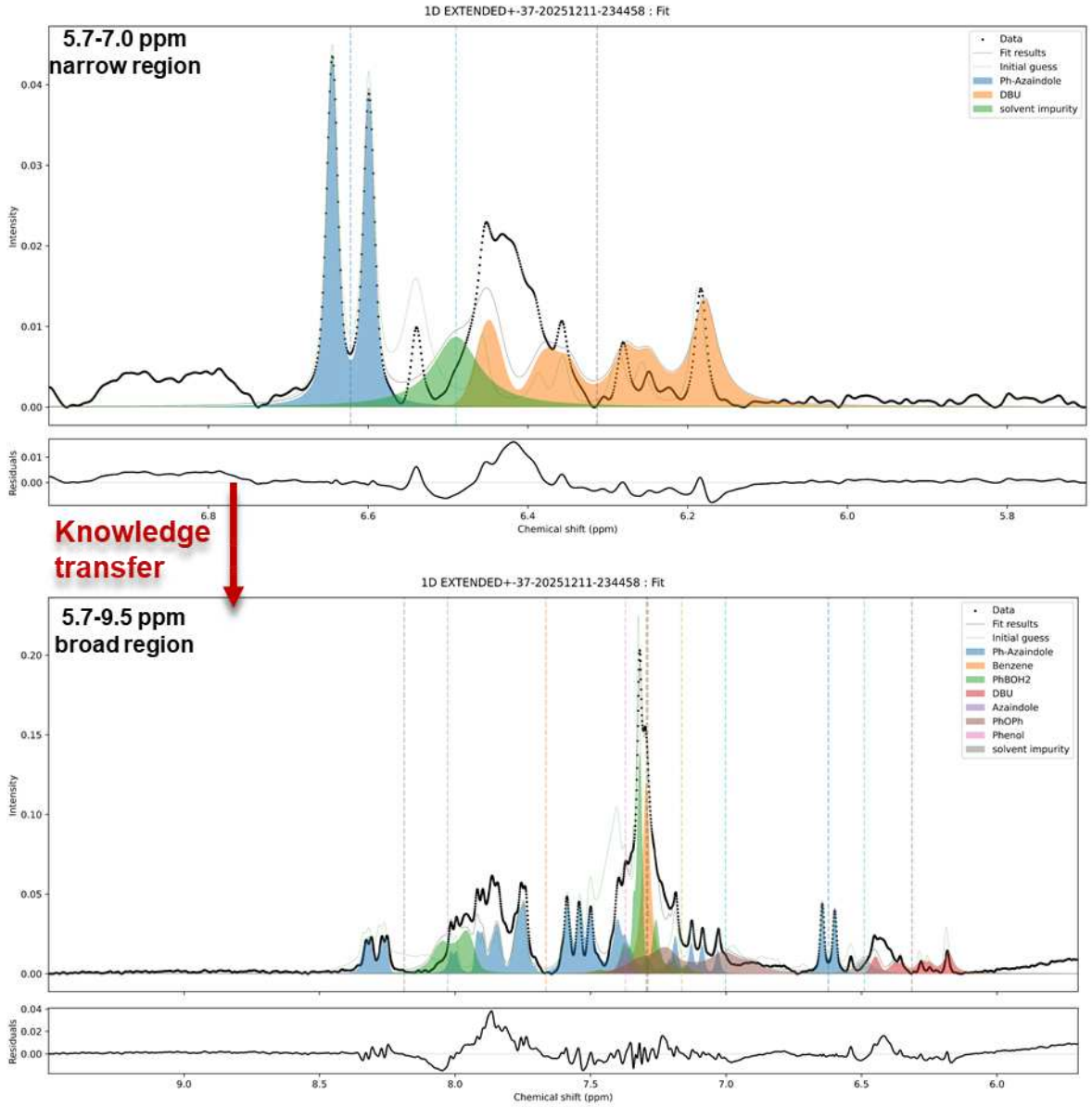


Figure S26. Representative multiple-region NMRanger poor fit for 7-azaindole crude reaction mixtures. Fitting was performed first over 5.7–7.0 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 5.7–9.5 ppm region, containing overlapping signals from N-phenyl-7-azaindole (*blue*), 7-azaindole (*purple*), $PhB(OH)_2$ (*green*), benzene (*orange*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 7-azaindole, O_2 atmosphere, 26 °C, in the presence of 0.2 equiv $Cu(OAc)_2$, 1.8 equiv PhB(OH)2, and 1.5 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Poor fit for a highly overlapped 7-azaindole crude spectrum; the fitted line reproduces the dominant component distribution, although visible mismatch remains in the crowded 8.0–7.8 ppm region and around minor baseline-distorted signals.

1.6. Additional hyperspaces of 7-azaindole

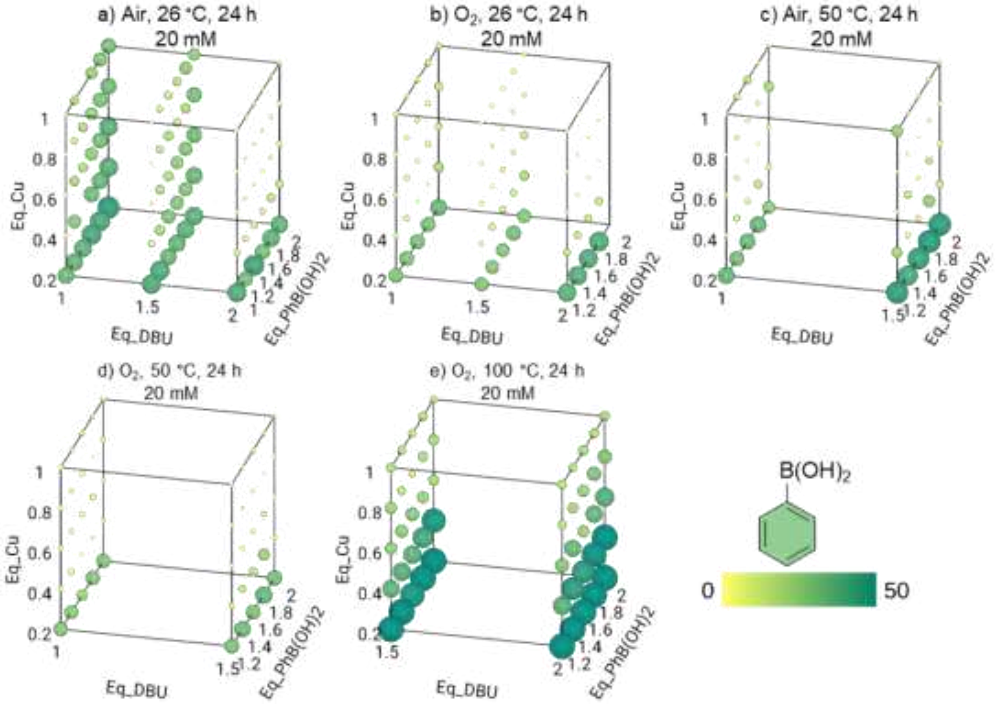


Figure S27. Unconsumed $PhB(OH)_2$ hyperspaces for the 7-azaindole Chan-Lam coupling a-e) Yield maps under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

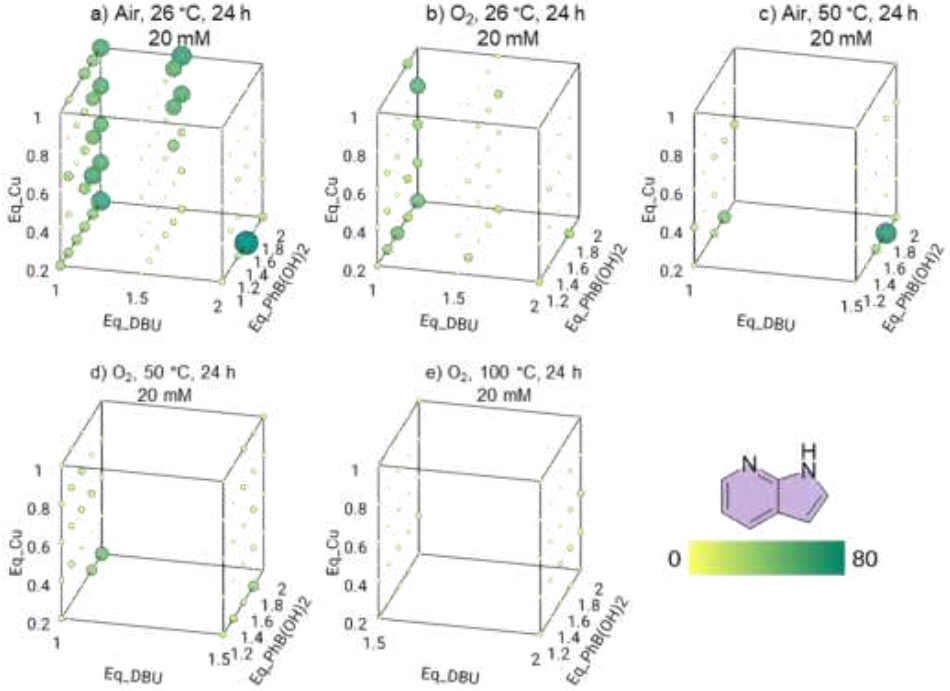


Figure S28. Unconsumed starting material hyperspaces for the 7-azaindole Chan-Lam coupling a-e) Yield maps under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

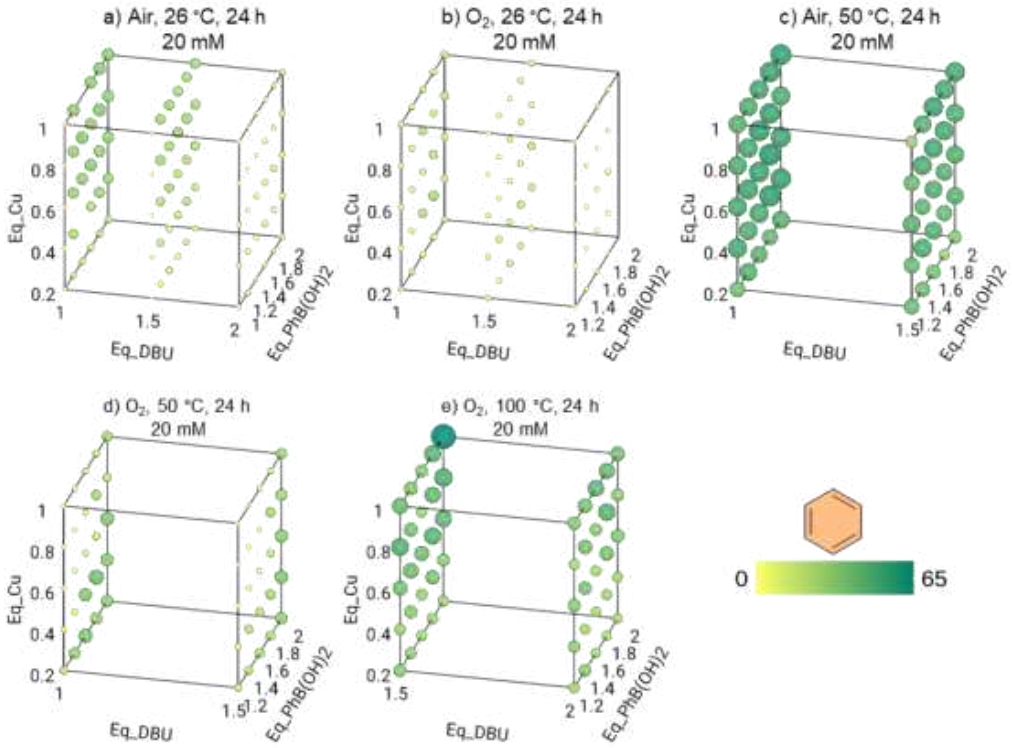


Figure S29. Benzene by-product formation hyperspaces for the 7-azaindole Chan-Lam coupling. a–e) Yield maps under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

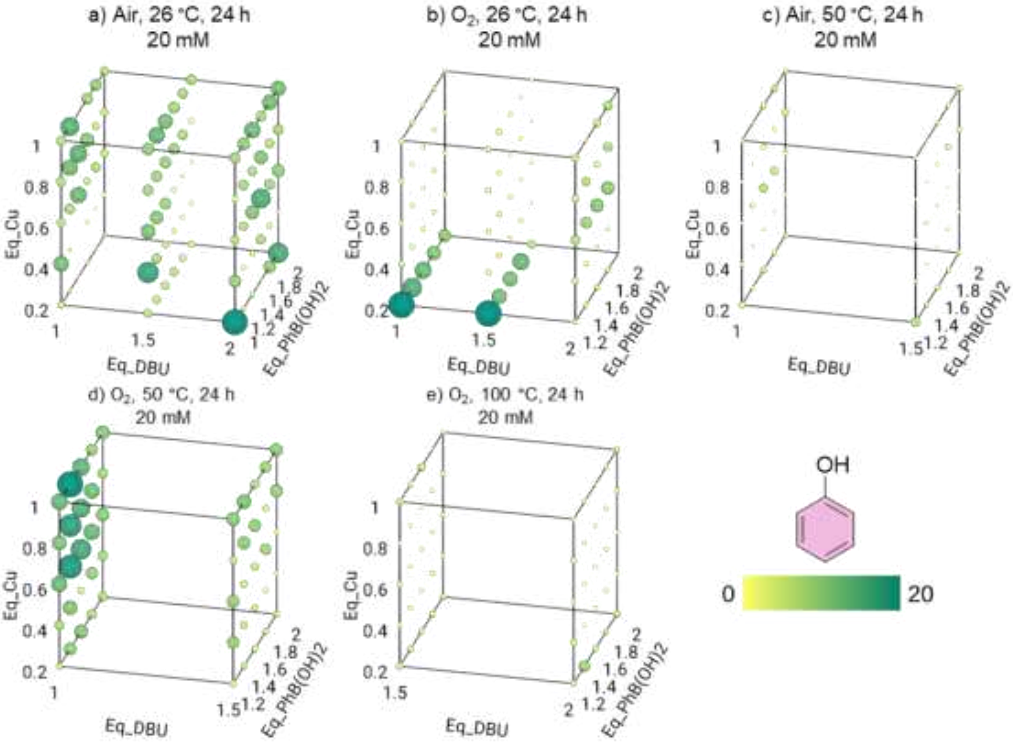


Figure S30. Phenol by-product formation hyperspaces for the 7-azaindole Chan-Lam coupling. a–e) Yield maps under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

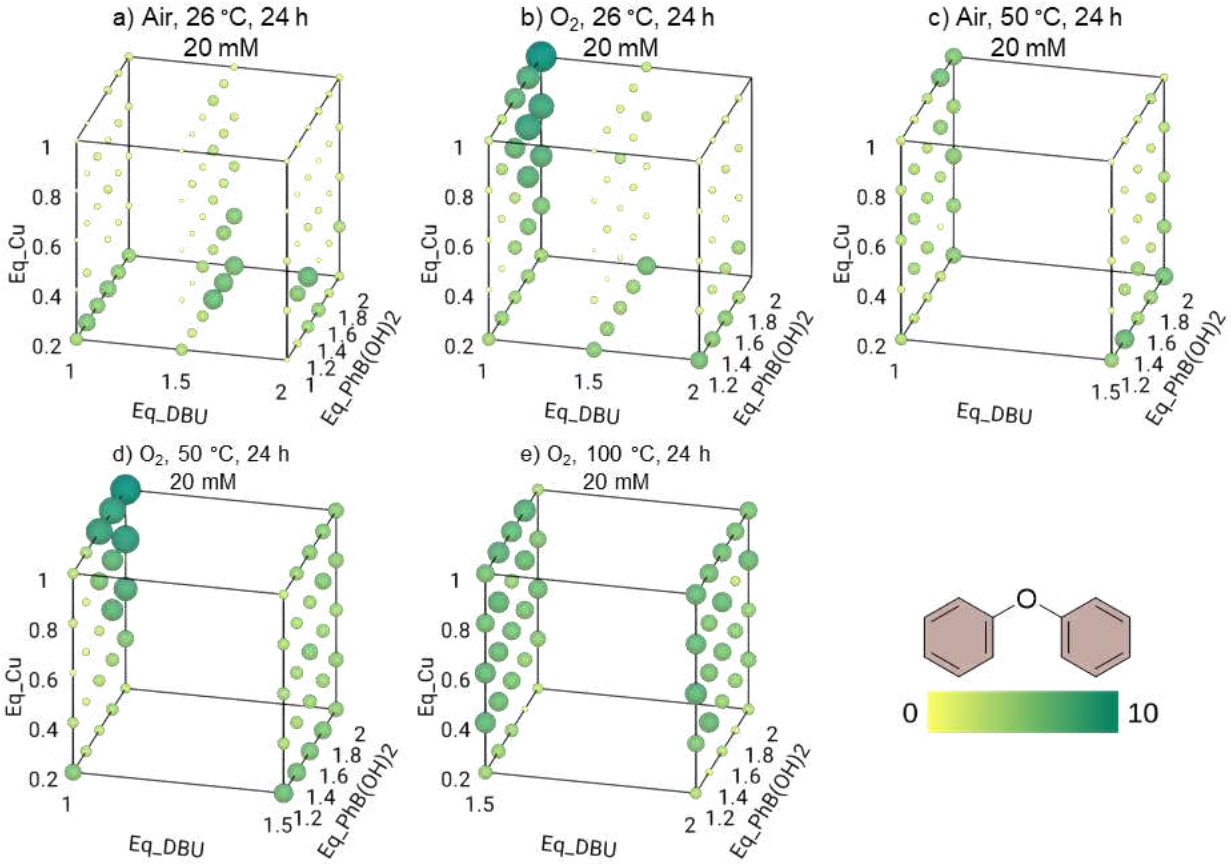


Figure S31. PhOPh by-product formation hyperspaces for the 7-azaindole Chan-Lam coupling. a–e) Yield maps under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

When evaluated relative to phenylboronic acid (see General Method **D** for calculation details), the aryl source for the coupling products and aryl-derived by-products, the average mass balances are 74, 68, 84, 79, and 86% for conditions: air, 26 °C, 24 h; O2, 26 °C, 24 h; air, 50 °C, 24 h; O2, 50 °C, 24 h; and O2, 100 °C, 24 h, respectively. (**Figure S32**). The remaining deficit most likely arises, at least in part, from phenylboronic acid precipitation after the reaction, consistent with its limited solubility in DCE and the observed residue on the vial walls.

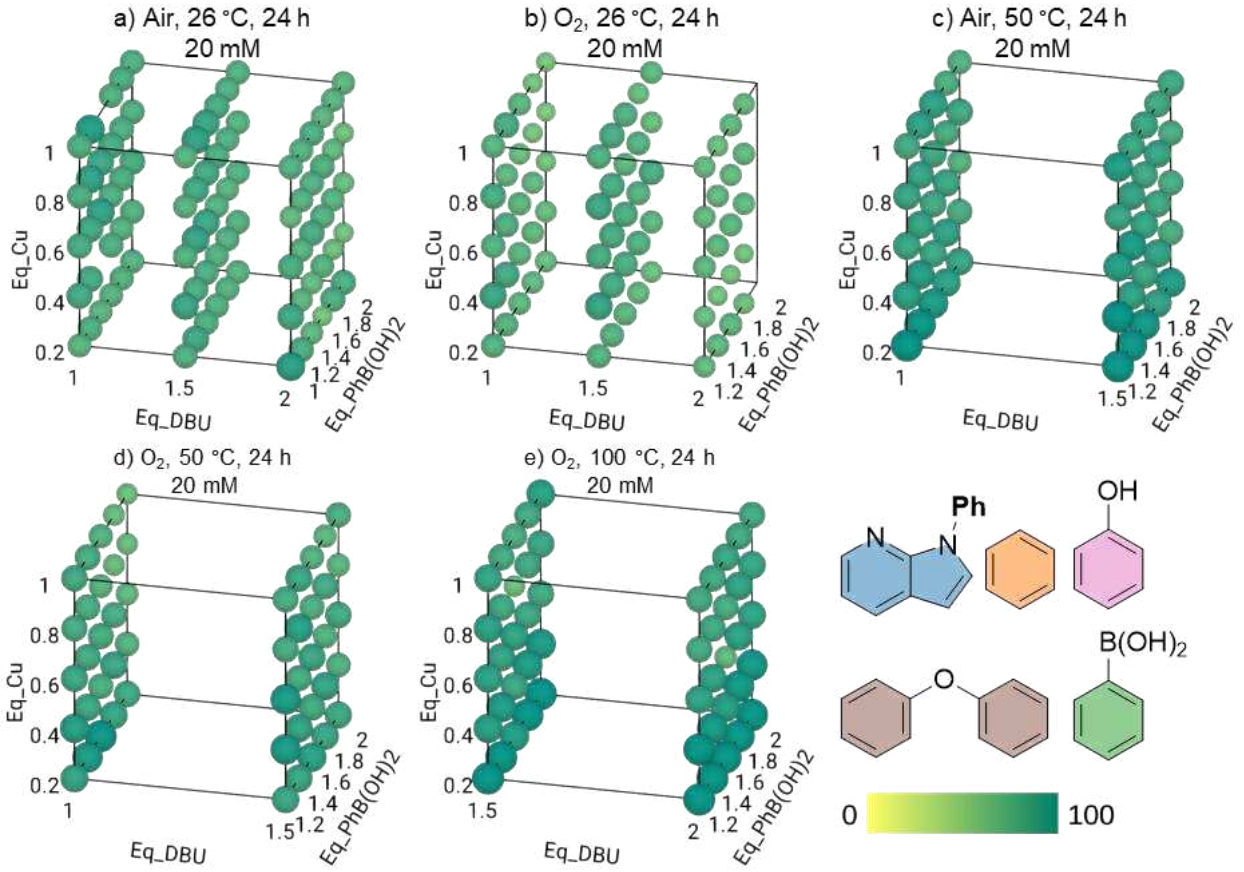


Figure S32. Mass-balance hyperspaces for the 7-azaindole Chan-Lam coupling a-e) Mass balance was determined relative to the initial amount of $PhB(OH)_2$ by summing all quantified phenylboronic acid-derived species: N-phenyl-,7-azaindole, benzene, phenol, diphenyl ether, and unconsumed $PhB(OH)_2$. Maps are shown under different reaction conditions: air, 26 °C, 24 h; O_2, 26 °C, 24 h; air, 50 °C, 24 h; O_2, 50 °C, 24 h; and O_2, 100 °C, 24 h, respectively.

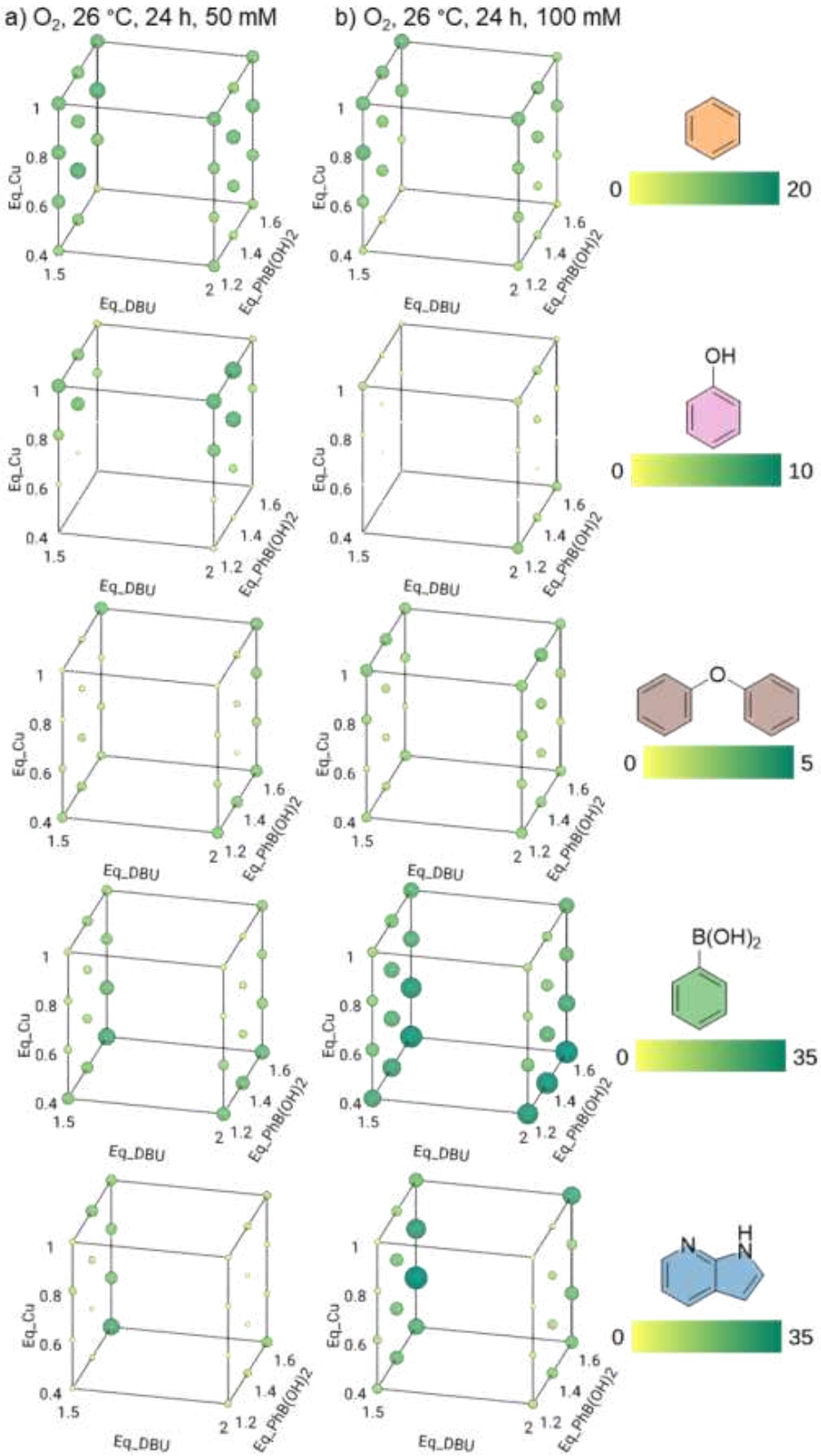
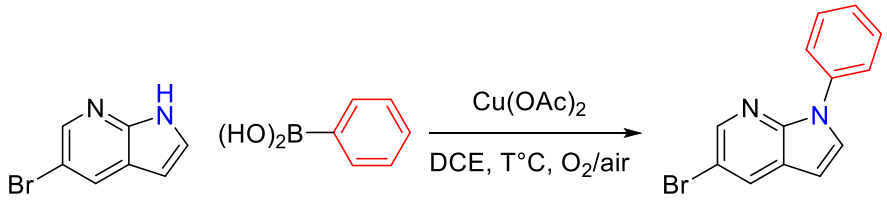


Figure S33. By-product and unconsumed reagent hyperspaces of 7-azaindole under high concentration. a) for 50 mM initial 7-azaindole concentration. **b)** for 100 mM initial concentration.

1.7. 5-bromo-7-azaindole

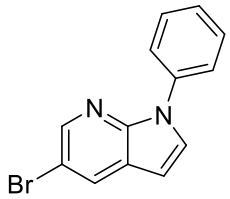


Reaction design. For 5-bromo7-azaindole, a four-dimensional reaction space was explored at 20 mM, under O_2. The reaction variables are $Cu(OAc)_2$ loading (0.2–1.0 equiv in 0.2 intervals), $PhB(OH)_2$ loading (1.2–2 equiv in 0.2 intervals), temperature (26, 50, and 100 °C), DBU loading (1.5, and 2 equiv).

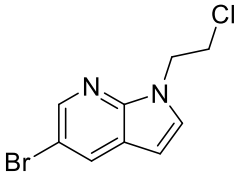
Following General Method **A**, reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, with a final reaction volume of 1.1 mL. First, a 75 mM $Cu(OAc)_2$ stock solution in water was added to the reaction vials and evaporated in a vacuum oven at 40 °C for 4-8 h. After evaporation, DCE was added, followed by the remaining DCE stock solutions: $PhB(OH)_2$ (250 mM), 5-bromo7-azaindole (100 mM), and DBU (2000 mM). When required, the reaction mixtures were sparged with oxygen, sealed, and placed in a preheated ThermoMixer C for the required reaction time. Preparation of one 54-vial reaction plate took approximately 45 min. Three sets of 5×5×2 conditions were prepared, corresponding to 150 experiments in total. After excluding two experimental outliers, the final dataset comprised 148 independent reaction conditions.

After completion of the selected reaction time, the plates containing the crude reaction mixtures were measured by Robowski-NMR, General Method **B**. Then crude mixtures are combined for purification on MPLC, see General Method **C**. The isolated compounds were then measured in DCE in a known concentrations (**Figure S42-Figure S43**) and added to the NMRanger basis reference set and used for the NMRanger quantification, General Method **D**.

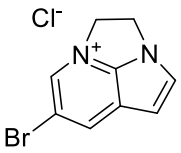
1.7.1. Characterizations and Reference Spectra for 5-bromo-7-azaindole Quantification



N-phenyl-5-bromo-7-azaindole[4]**: ^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.38 (d, *J* = 2.2 Hz, 1H), 8.09 (d, *J* = 2.2 Hz, 1H), 7.75 – 7.67 (m, 2H), 7.58 – 7.48 (m, 3H), 7.36 (t, *J* = 7.4 Hz, 1H), 6.58 (d, *J* = 3.6 Hz, 1H).; **^{13}C NMR** (101 MHz, $CDCl_3$) δ 145.9, 144.2, 138.2, 131.3, 129.6, 129.5, 126.9, 124.1, 123.3, 112.7, 101.2.



N-chloroethyl-5-bromo-7-azaindole:^{1}H NMR (400 MHz, $CDCl_3$) δ 8.32 (d, *J* = 2.2 Hz, 1H), 8.03 (d, *J* = 2.2 Hz, 1H), 7.30 (d, *J* = 3.5 Hz, 1H), 6.42 (d, *J* = 3.5 Hz, 1H), 4.58 (t, *J* = 6.0 Hz, 2H), 3.90 (t, *J* = 6.0 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 145.6, 143.4, 130.9, 130.3, 122.4, 112.0, 99.4, 46.8, 42.9; HRMS (ESI) m/z: [M + H]+ Calcd for $C_9H_9BrClN_2$: 258.9632, Found 258.9638.



6-bromo-1,2-dihydro-2a,7a-diazacyclopenta[cd]inden-7a-ium chloride (cyclized pyridinium chloride): ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.25 (d, *J* = 1.8 Hz, 1H), 7.87 (d, *J* = 2.7 Hz, 1H), 7.85 (d, *J* = 1.8 Hz, 1H), 6.66 (d, *J* = 2.7 Hz, 1H), 5.01 – 4.94 (m, 2H), 4.22 – 4.14 (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 146.4, 134.2, 131.6, 130.6, 128.6, 115.6, 101.9, 56.1, 41.8.

Reference Spectra preparation in DCE. Isolated N-phenyl-5-bromo-7-azaindole was measured in DCE for the reference library. For phenylboronic acid, DBU, benzene, phenol, and diphenyl ether, commercially available compounds were used. All spectra were recorded on an 80 MHz NMR spectrometer with 32 scans, a 6.4 s acquisition time, a 15 s repetition time, and a 90° pulse angle as 7-azaindole (see **Section 5.1** for parameter selection). See **Section 1.5.1** for phenylboronic acid, DBU, benzene, phenol, and diphenyl ether spectra.

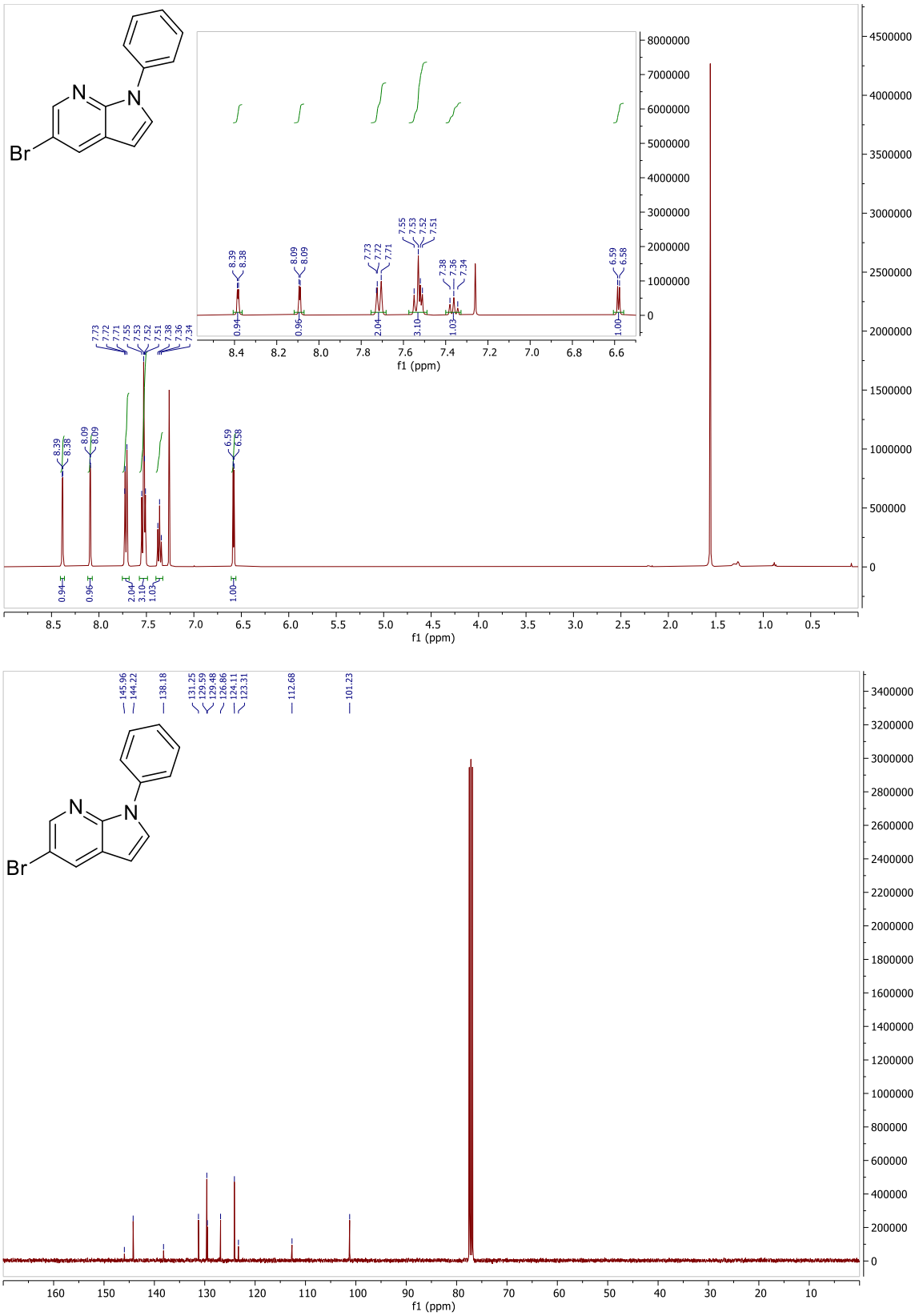


Figure S34. ^{1}H NMR (*top*) and ^{13}C NMR (*bottom*) spectra of N-phenyl-5-bromo7-azaindole in $CDCl_3$.

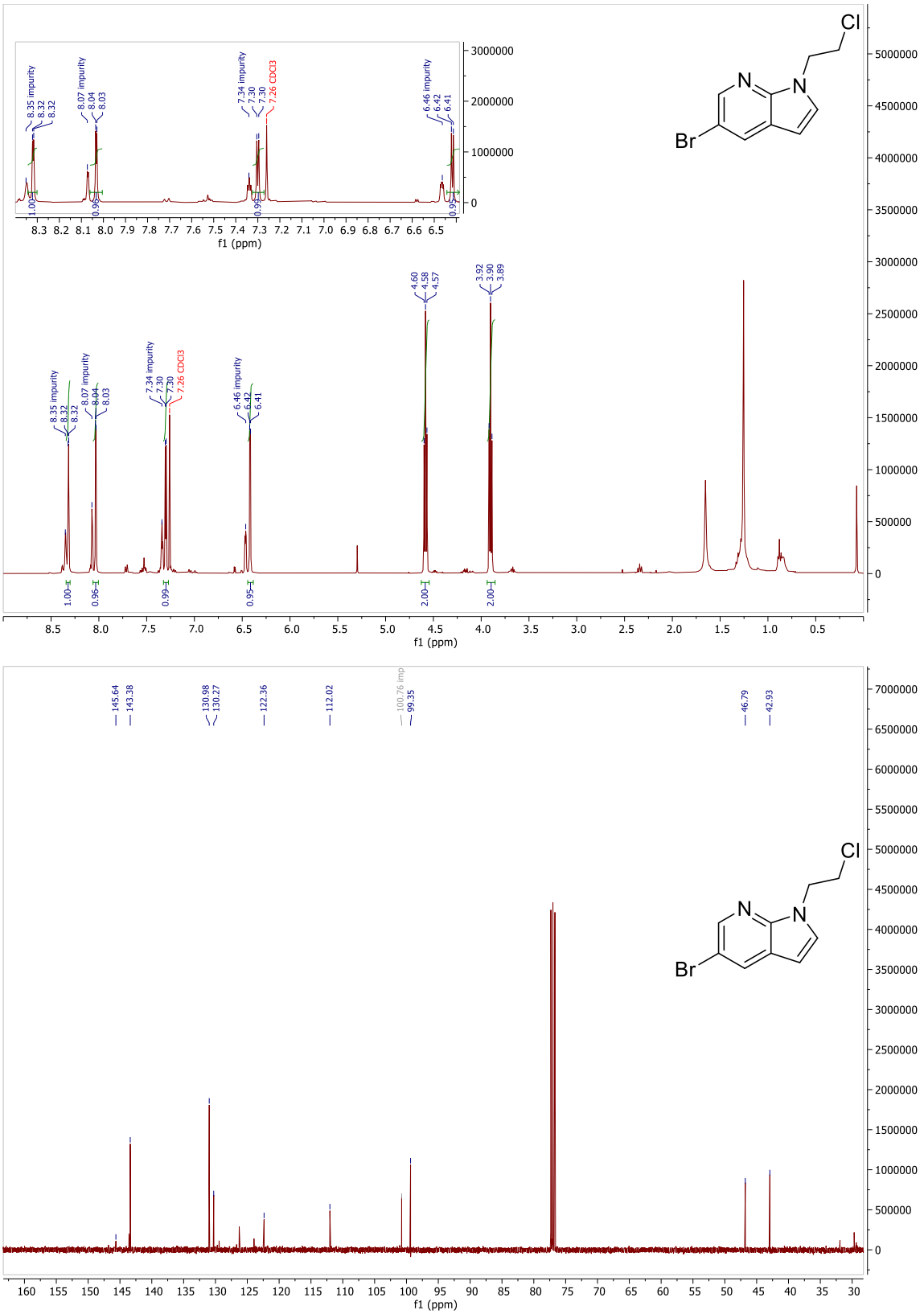


Figure S35. ^{1}H NMR (*top*) and ^{13}C NMR (*bottom*) spectra of N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

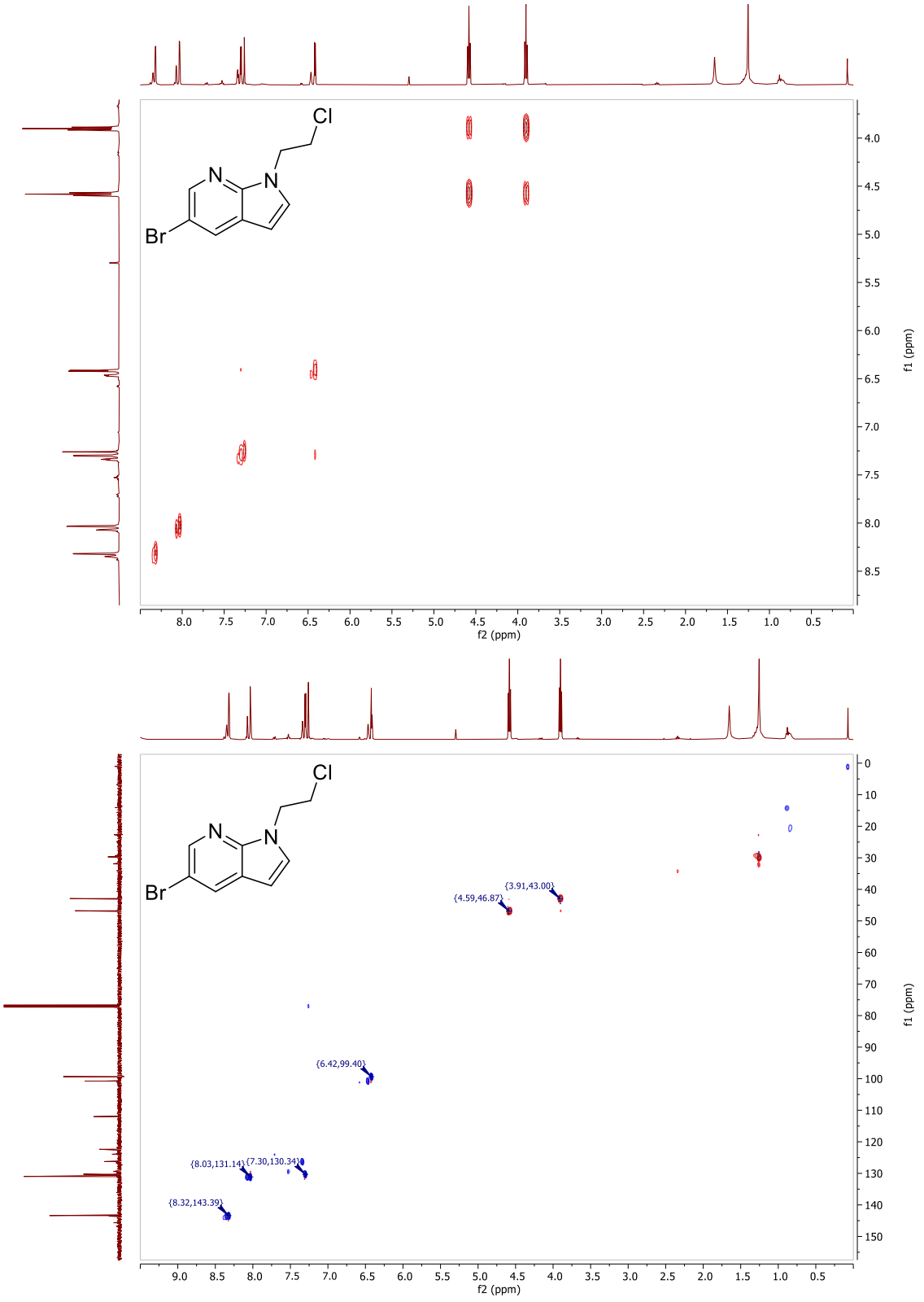


Figure S36. COSY (*top*) and multiplicity-edited HSQC (^{1}H/DEPT 135) (*bottom*) spectra of N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

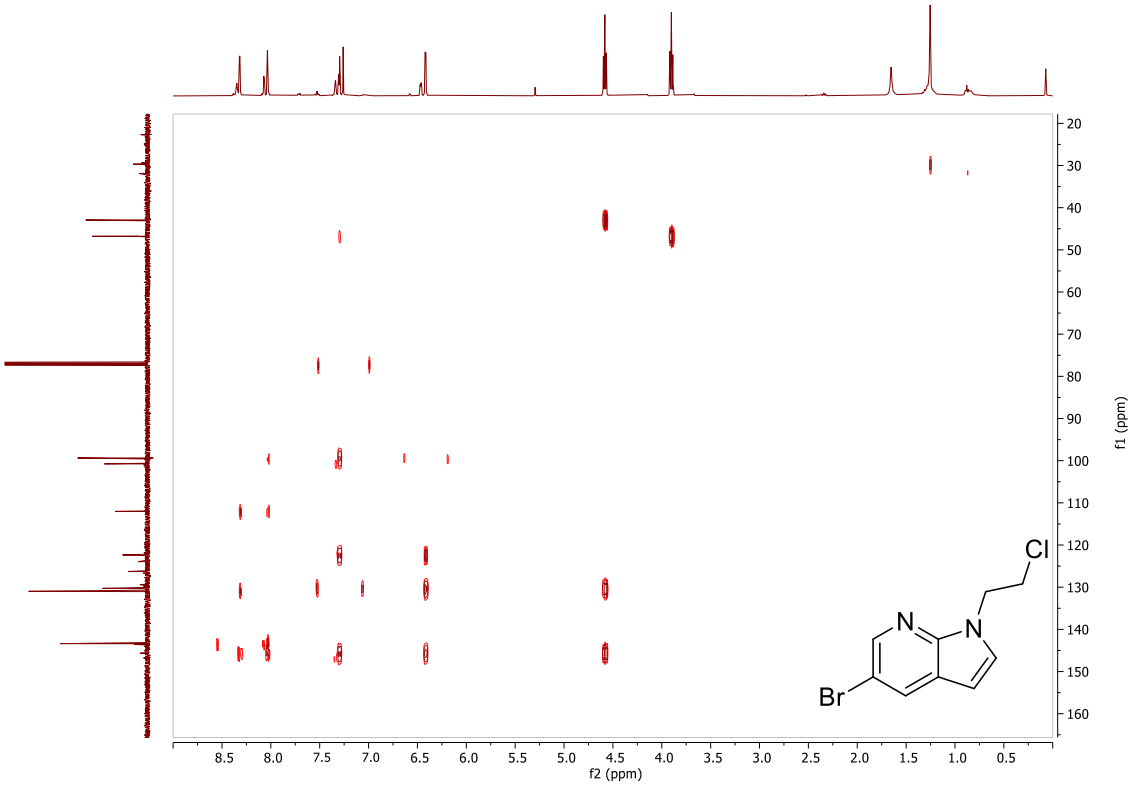


Figure S37. HMBC spectra of N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

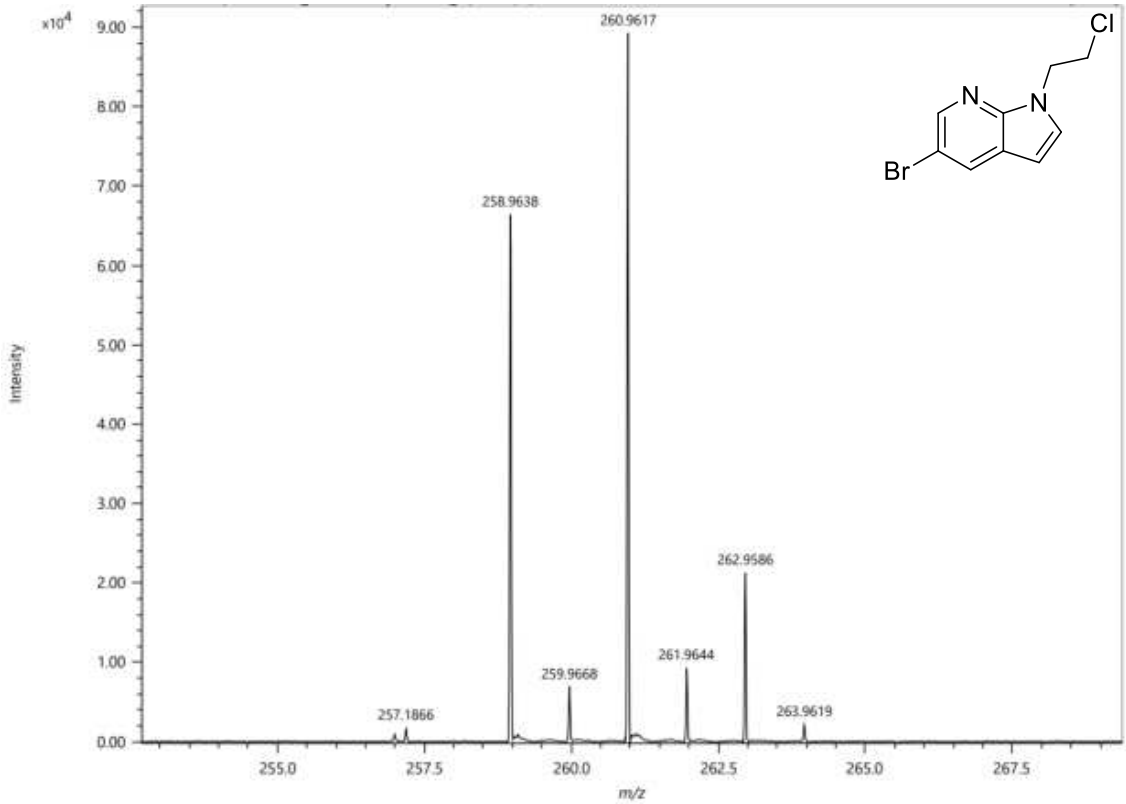


Figure S38. HRMS of N-chloroethyl-5-bromo-7-azaindole.

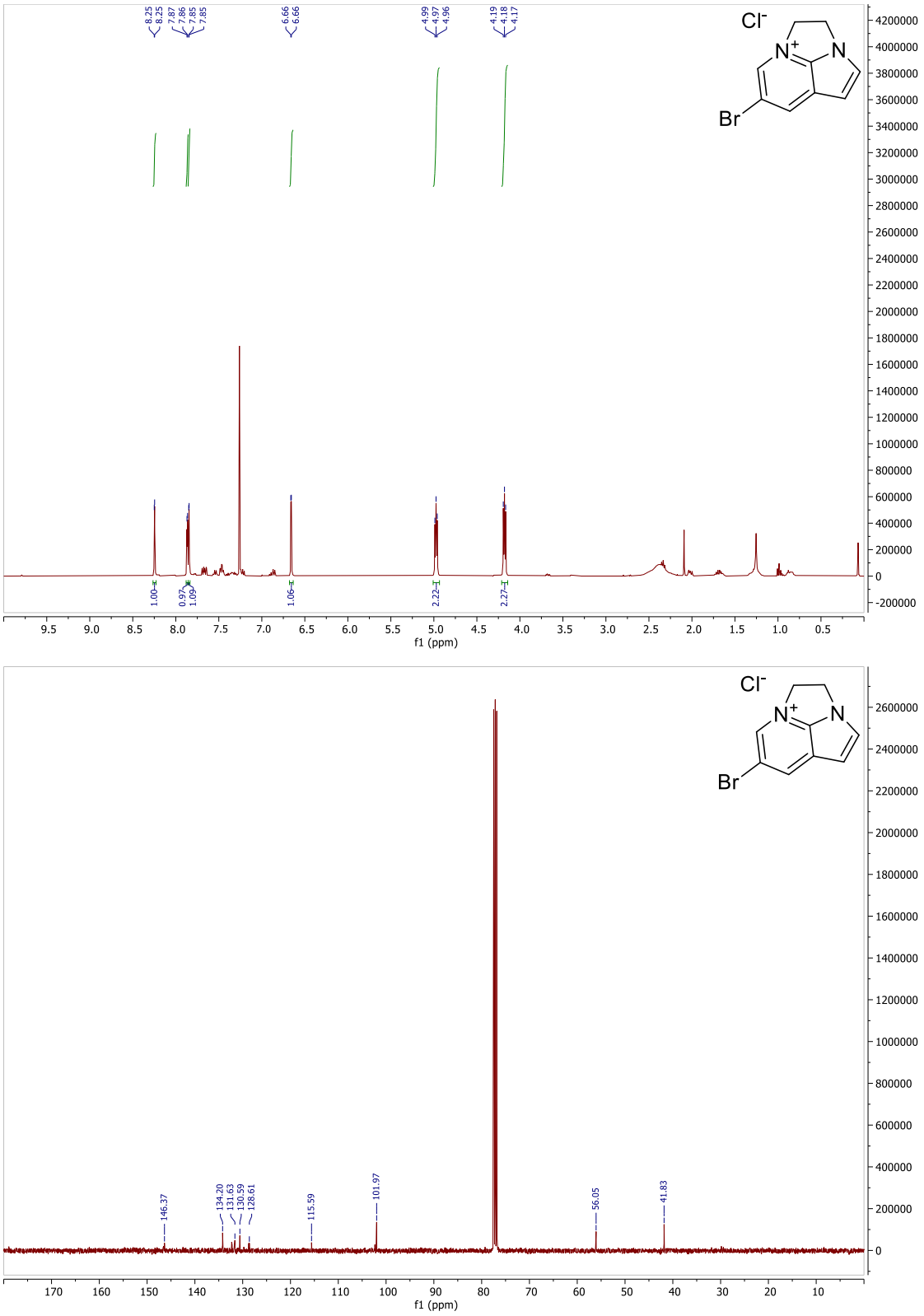


Figure S39. ^{1}H NMR (*top*) and ^{13}C NMR (*bottom*) spectra of cyclized N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

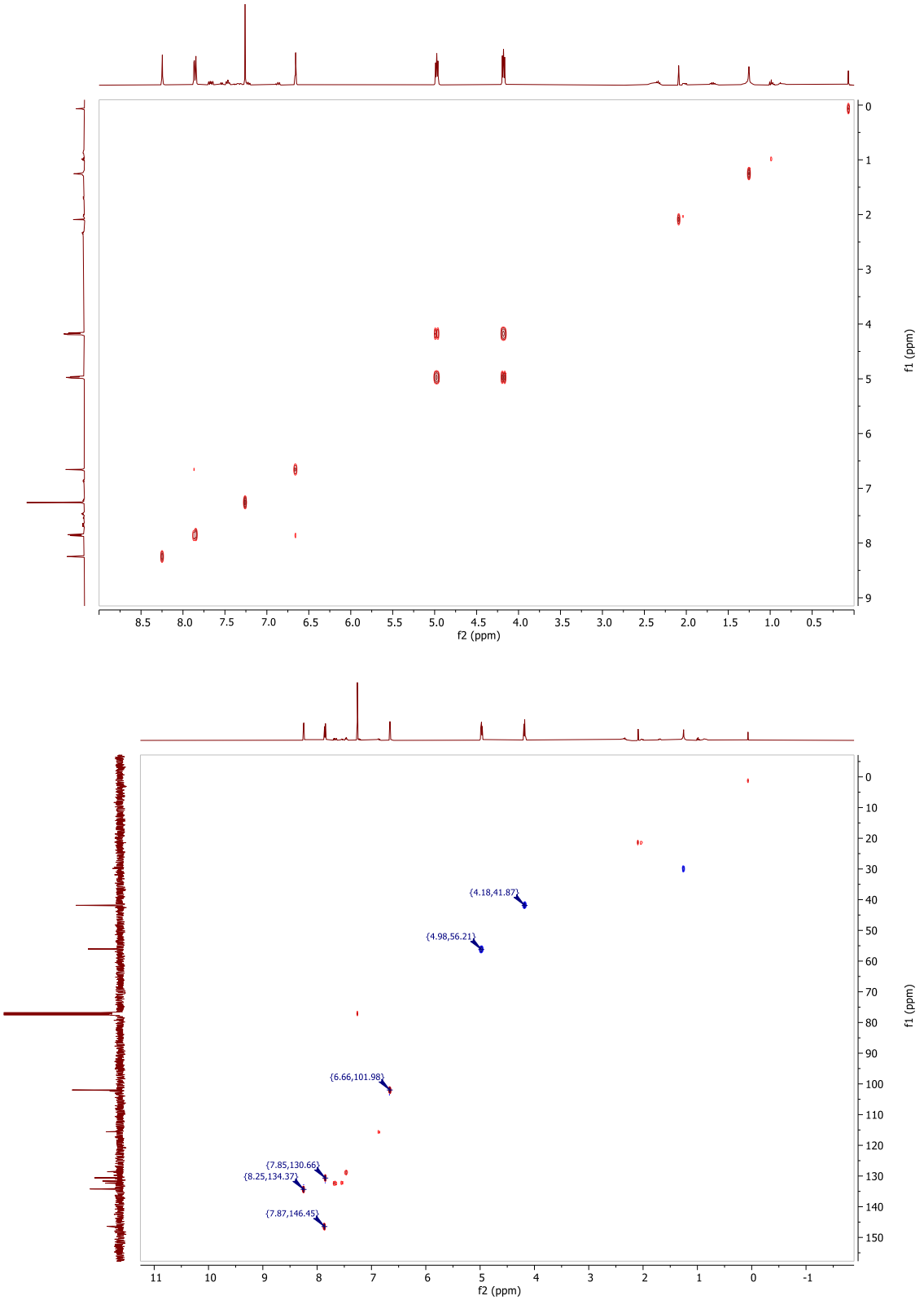


Figure S40. COSY (*top*) and multiplicity-edited HSQC (^{1}H/DEPT 135) (*bottom*) spectra of cyclized N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

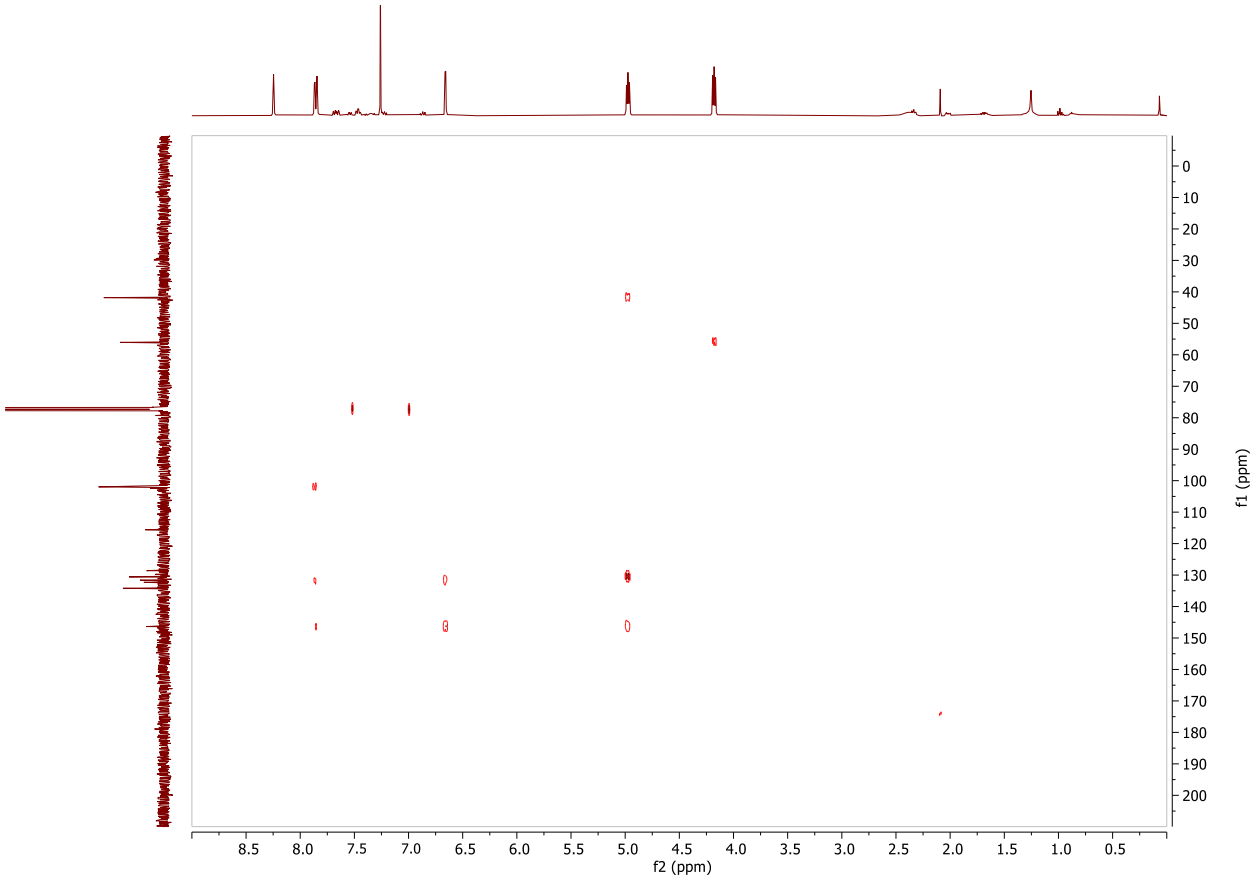


Figure S41. HMBC spectra of cyclized N-chloroethyl-5-bromo-7-azaindole in $CDCl_3$.

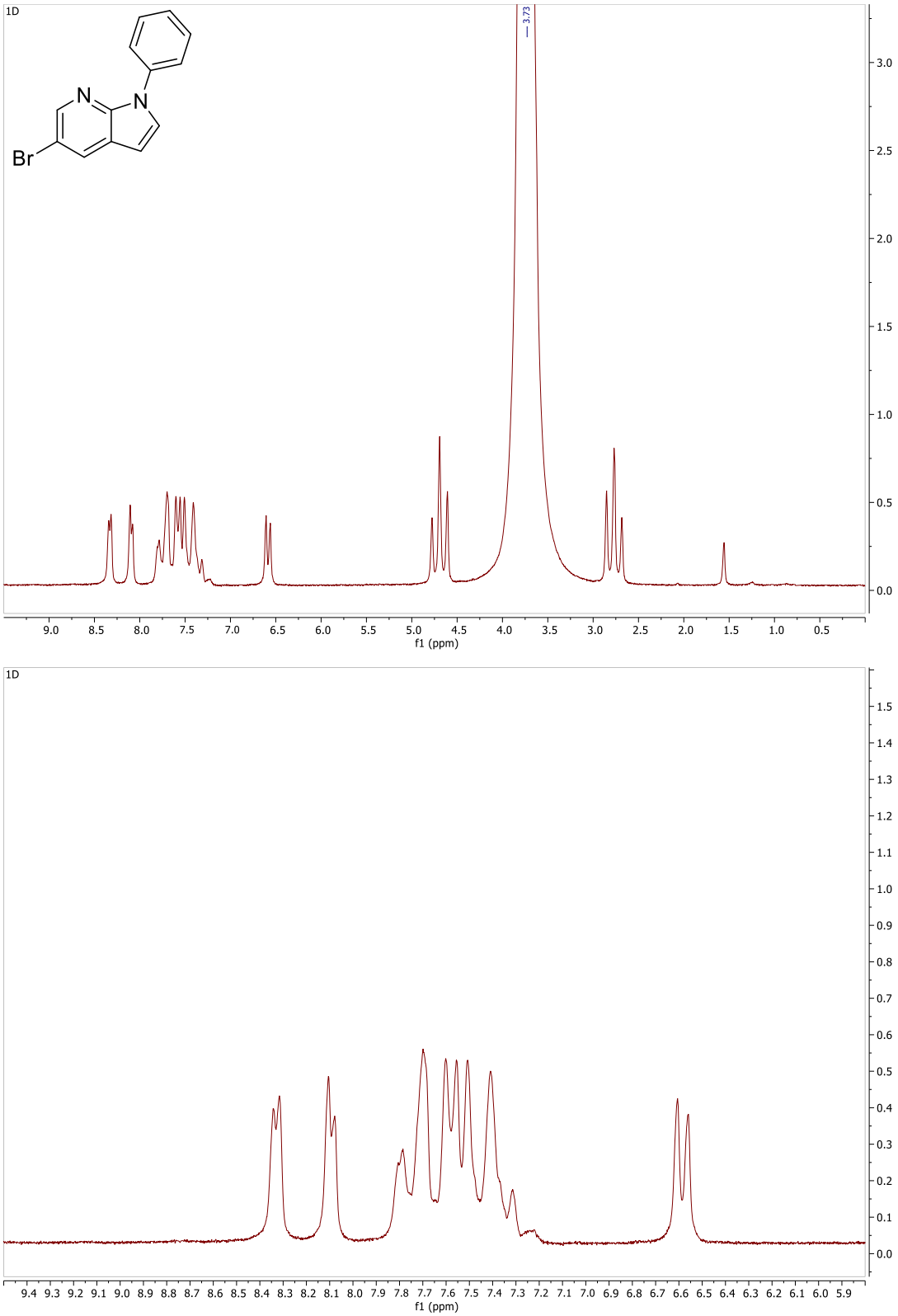


Figure S42. ^{1}H NMR spectrum of N-phenyl-5-bromo-7-azaindole in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.8–9.5 ppm, (*bottom*) used for NMRanger quantification.

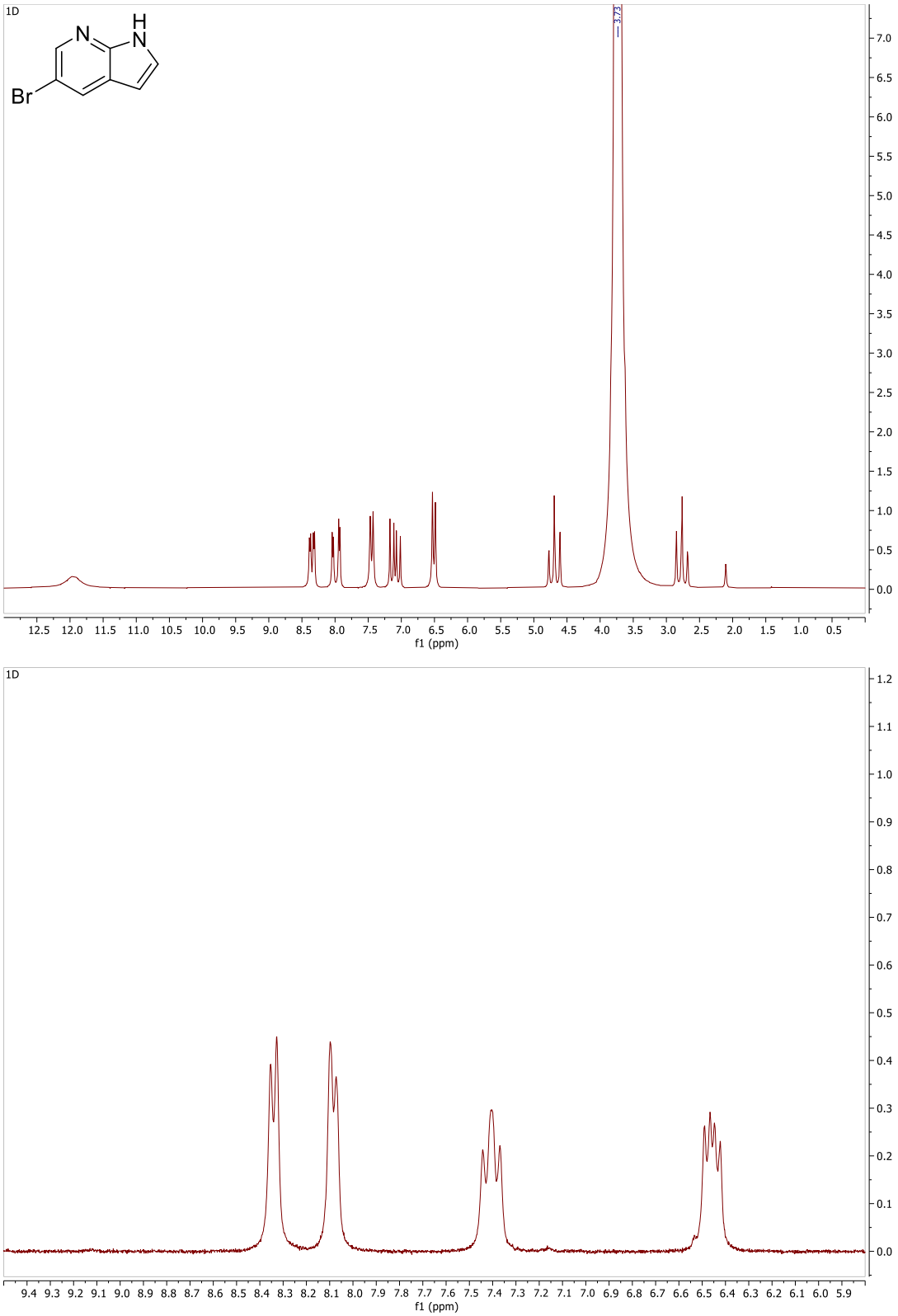


Figure S43. ^{1}H NMR spectrum of 5-bromo-7-azaindole in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 5.8–9.5 ppm, (*bottom*) used for NMRanger quantification.

1.7.2. Calibration and fitting of 5-bromo-7-azaindole

A calibration curve for N-phenyl-5-bromo-7-azaindole was generated automatically in NMRanger using DCE calibrant solutions of known concentration incorporated into the basis set (**Figure S44**). These calibrants served as external standards for yield estimation. Crude 5-bromo-7-azaindole reaction mixtures were quantified by multiple-region fitting: the characteristic 7.85–8.70 ppm region, containing product and starting-material signals, was fitted first, and the resulting abundance estimates were transferred to guide fitting of the broader 5.80–9.50 ppm region. Because all components are fully aromatic, this range included all reference resonances required for quantification while excluding the non-deuterated solvent region, which contains no bromoazaindole-derived signals. At 0.2 equiv copper loading, product formation was too low for reliable transfer from the characteristic region; these samples were therefore quantified by single-range fitting over 5.8–9.5 ppm. Representative good, acceptable, and poor fits are shown in **Figure S45**-**Figure S47** illustrating spectral reconstruction across different product and by-product compositions before incorporation into the reaction hyperspaces discussed in **Section 4.3 of the main text.**

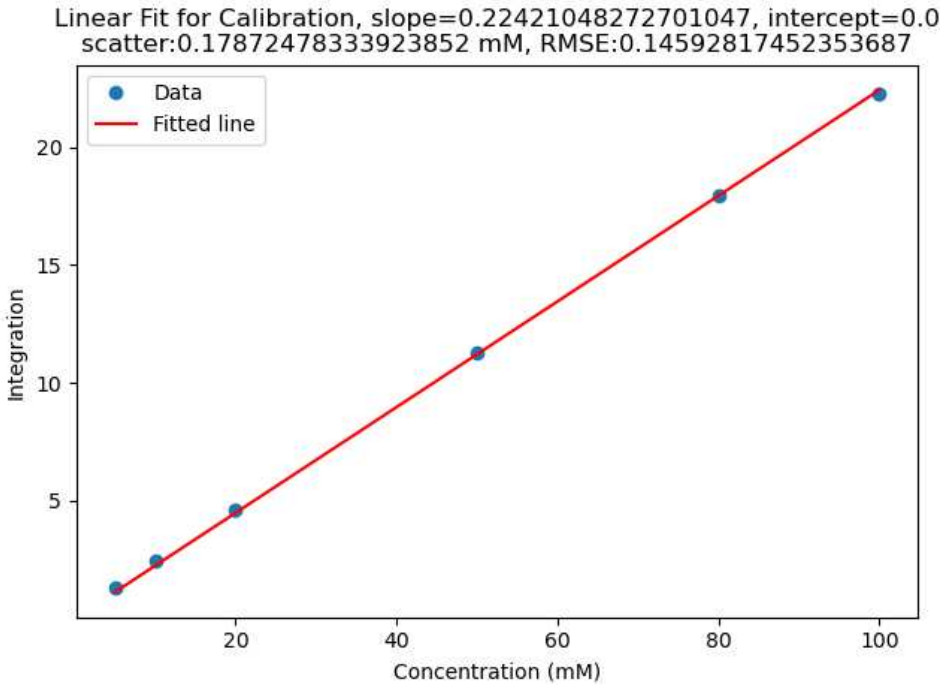


Figure S44. Calibration curve of N-phenyl-5-bromo-7-azaindole generated automatically by NMRanger. Solutions containing known concentrations of N-phenyl-5-bromo-7-azaindole were prepared in DCE at 5, 10, 20, 50, 80, 100 mM. The upper concentration was chosen based on the expected maximum product concentration in the morpholine coupling reactions.

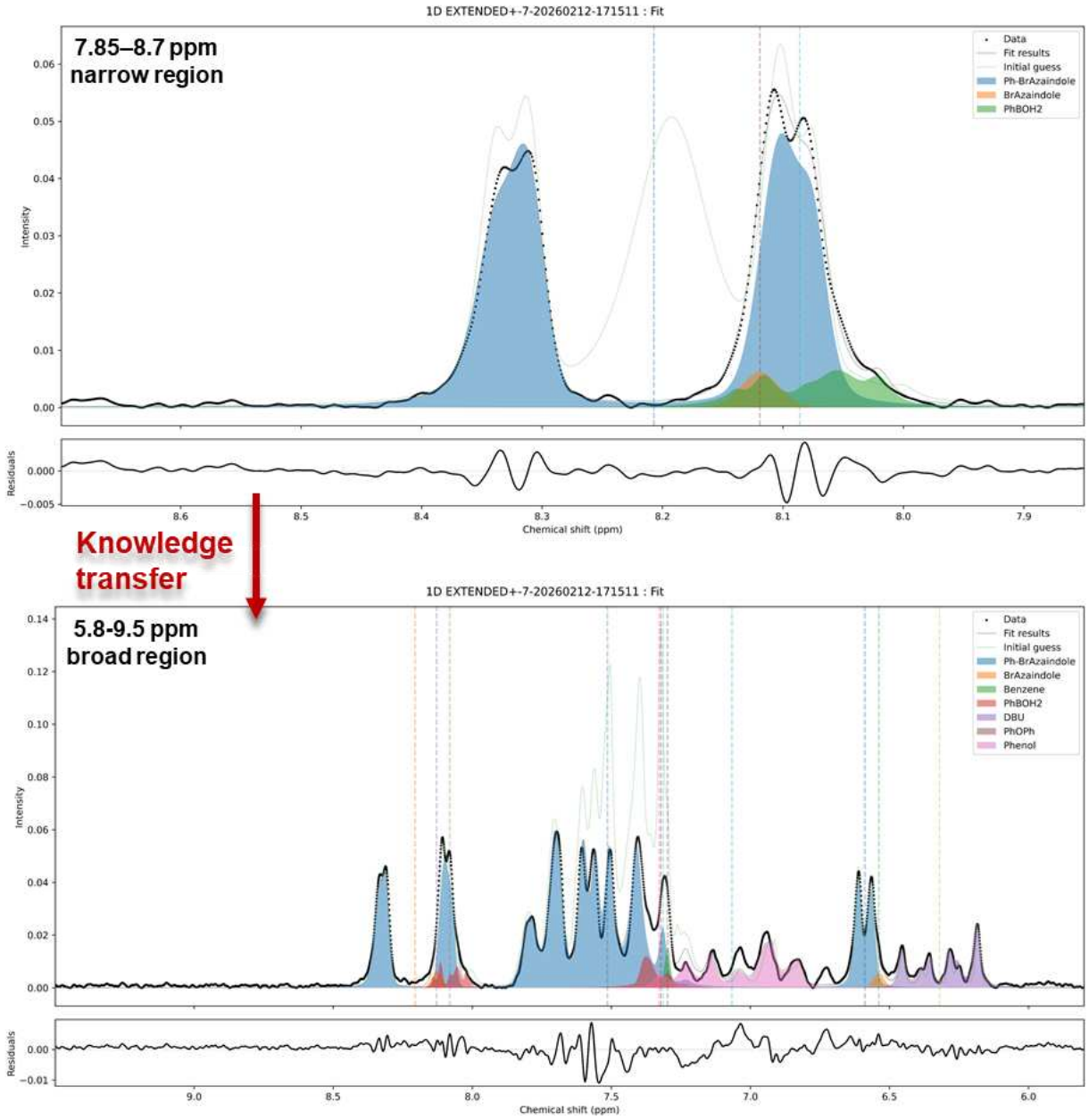


Figure S45. Representative multiple-region NMRanger good fit for 5-bromo-7-azaindole crude reaction mixtures. Fitting was performed first over the 7.85–8.7 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 5.8–9.5 ppm region, containing overlapping signals from N-phenyl-5-bromo-7-azaindole (*blue*), 5-bromo-7-azaindole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*green*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 5-bromo-7-azaindole, O_2 atmosphere, 50 °C, in the presence of 0.6 equiv $Cu(OAc)_2$, 1.4 equiv $PhB(OH)_2$, and 1.5 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Good fit for a highly overlapped 5-bromo-7-azaindole crude spectrum; the fitted line closely follows the main peak shapes and intensities across the aromatic region, with only moderate residuals in the most congested regions.

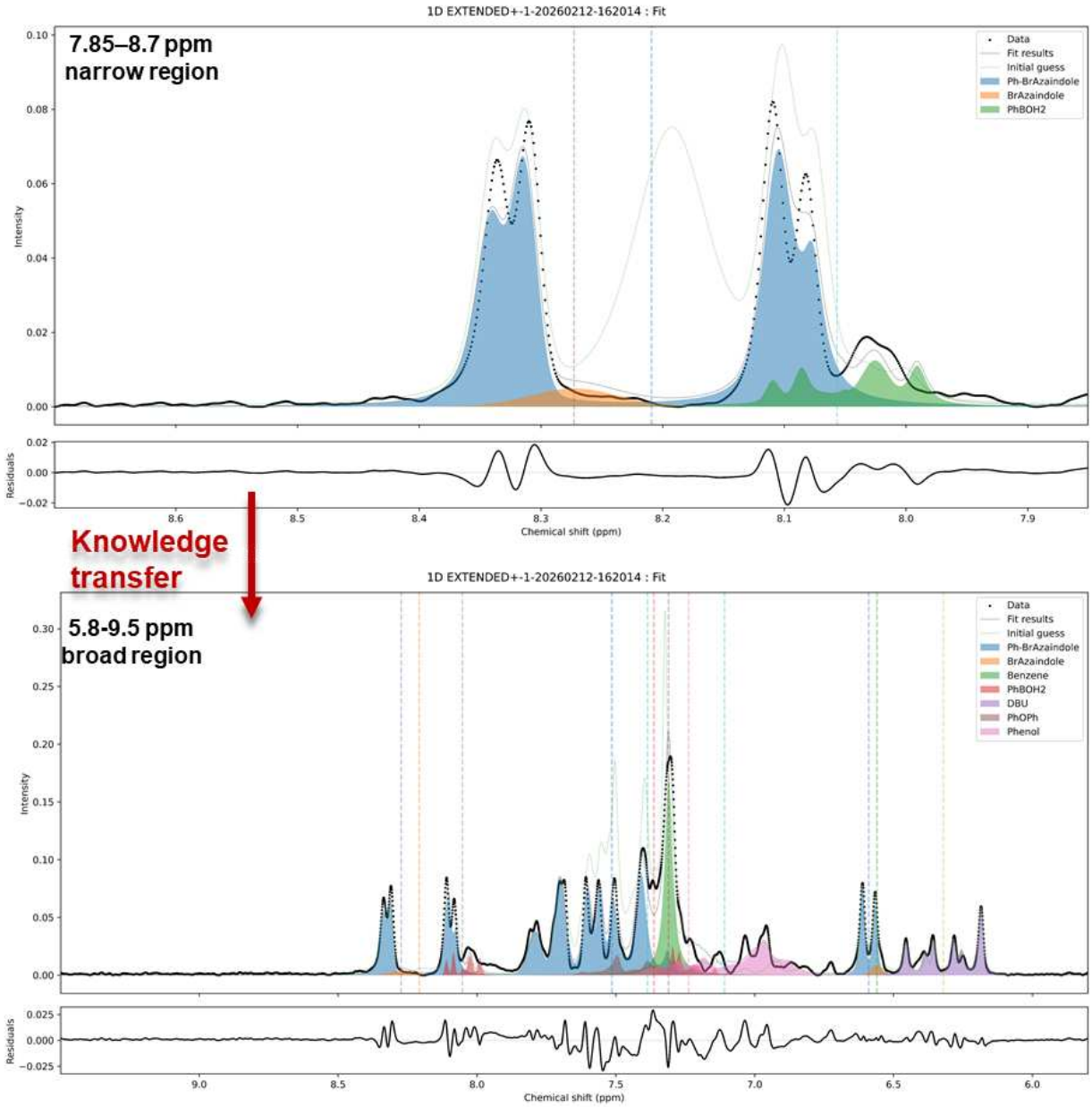


Figure S46. Representative multiple-region NMRanger acceptable fit for 5-bromo-7-azaindole crude reaction mixtures. Fitting was performed first over the 7.85–8.7 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 5.8–9.5 ppm region, containing overlapping signals from N-phenyl-5-bromo-7-azaindole (*blue*), 5-bromo-7-azaindole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*green*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 5-bromo-7-azaindole, O_2 atmosphere, 50 °C, in the presence of 0.4 equiv $Cu(OAc)_2$, 2.0 equiv $PhB(OH)_2$, and 2.0 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Acceptable fit for a highly overlapped 5-bromo-7-azaindole crude spectrum. The fitted line reproduces the main experimental features reasonably well, especially the dominant product/by-product signals around 8.4–8.0 ppm (product), 7.6–7.3 ppm (product and benzene), and the sharper signals near 6.6 ppm (product). However, there are clear structured residuals in the most congested aromatic region, especially around 7.5–7.2 ppm, where the fit does not fully match the peak shapes and intensities. There is also some mismatch in the lower-intensity region around 7.1–6.9 ppm.

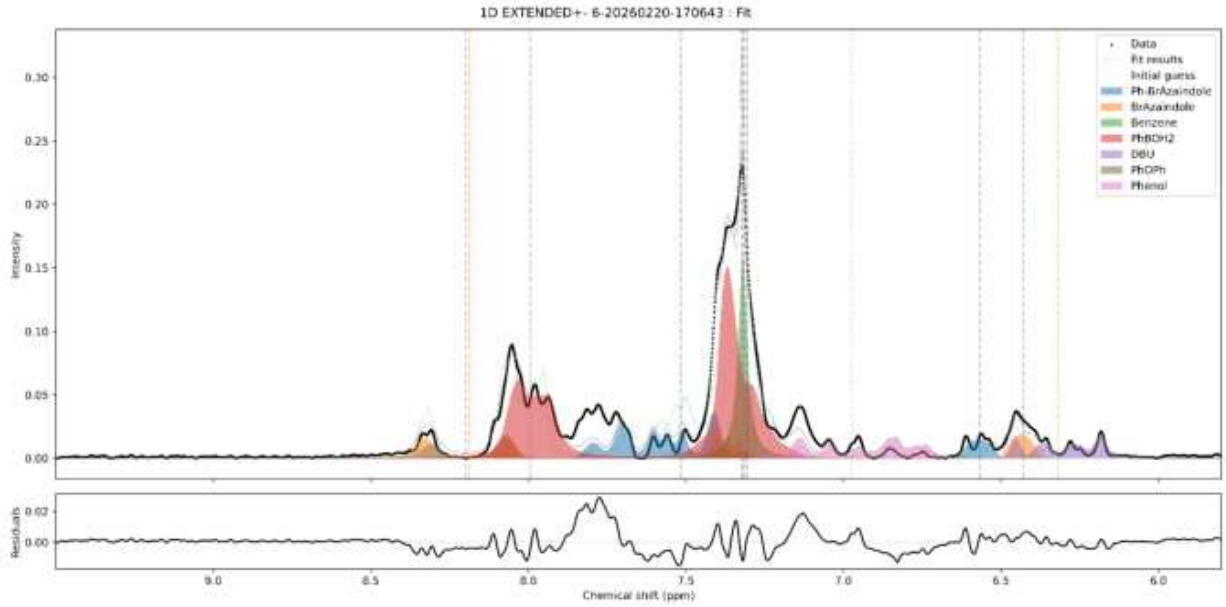


Figure S47. Representative multiple-region NMRanger poor fit for a 5-bromo-7-azaindole crude reaction mixture. Because of the low conversion and absence of characteristic product resonances, fitting was performed only over the wide 5.8–9.5 ppm aromatic region. This region contains overlapping signals from N-phenyl-5-bromo-7-azaindole (*blue*), 5-bromo-7-azaindole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*green*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 5-bromo-7-azaindole, O_2 atmosphere, 100 °C, in the presence of 0.8 equiv $Cu(OAc)_2$, 2.0 equiv $PhB(OH)_2$, and 1.5 equiv DBU. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Poor fit for a broadened 5-bromo-7-azaindole crude spectrum, where there is not much coupling product formation. The fitted line reproduces the dominant central intensity around 7.4–7.3 ppm (phenyl boronic acid and benzene), but the broader features around 8.1–7.9 ppm and 7.2–7.0 ppm are not fully matched, giving structured residuals in these regions.

1.8. Additional hyperspaces of 5-bromo-7-azaindole

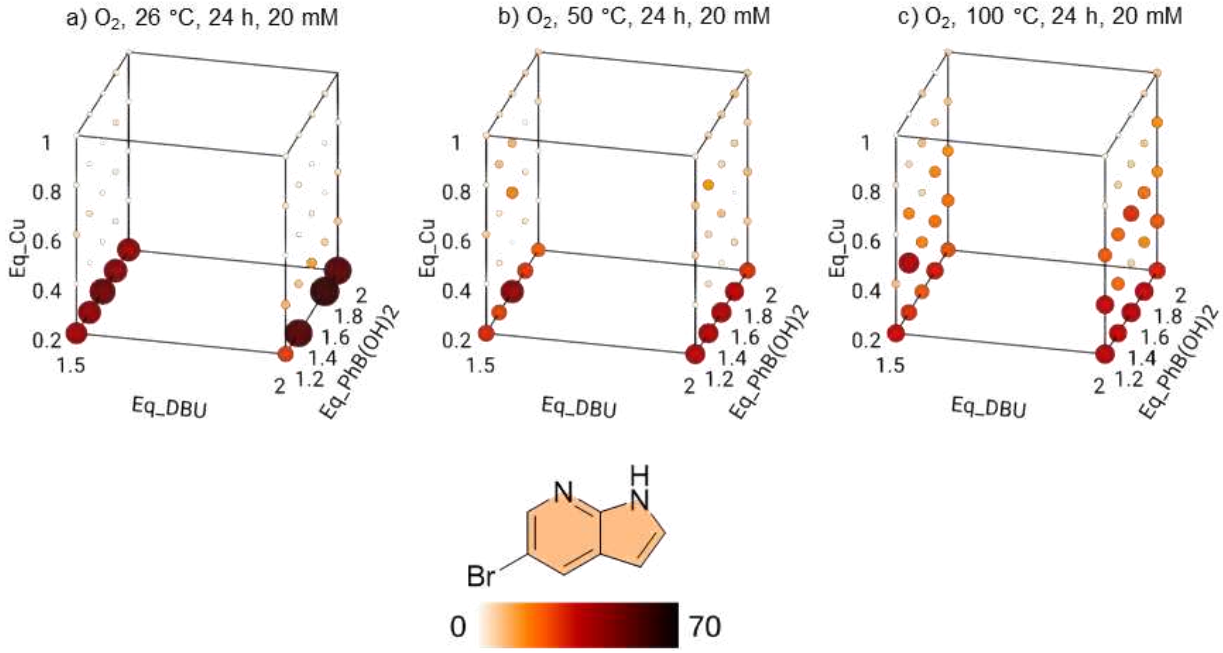


Figure S48. Unconsumed starting material hyperspaces for the 5-bromo-7-azaindole Chan-Lam coupling a-c) Yield maps under different reaction conditions: 20 mM, O_2, for 24 h for all, and at 26, 50, 100 °C respectively.

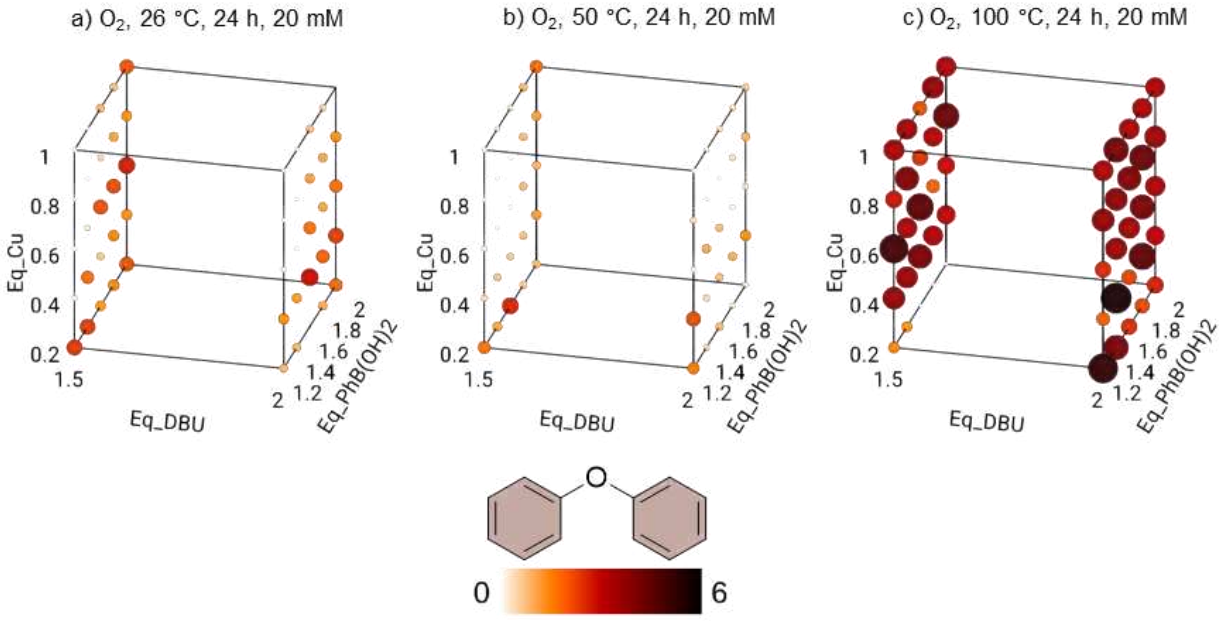


Figure S49. PhOPh by-product formation hyperspaces for the 5-bromo-7-azaindole Chan-Lam coupling. a–c) Yield maps under different reaction conditions: 20 mM, O_2, for 24 h for all, and at 26, 50, 100 °C respectively.

When we evaluated the mass balance relative to phenylboronic acid (see General Method **D** for calculation details), the aryl source for the coupling products and aryl-derived by-products, the average mass balances are 66.5, 80, and 80% for conditions: 20 mM, O_2, for 24 h for all, and at 26, 50, 100 °C respectively. (**Figure S50**). The remaining deficit most likely arises, at least in part, from phenylboronic acid precipitation after the reaction, consistent with its limited solubility in DCE and the observed residue on the vial walls.

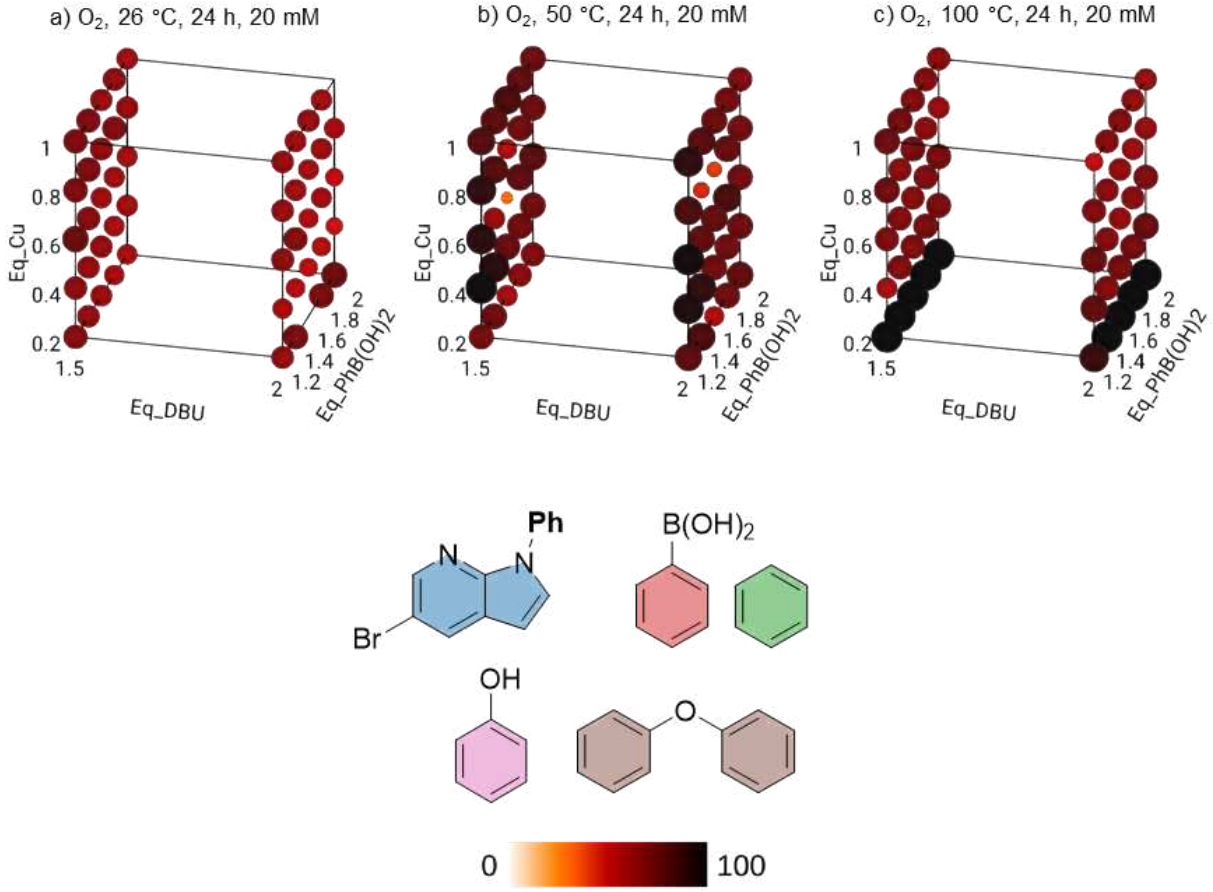
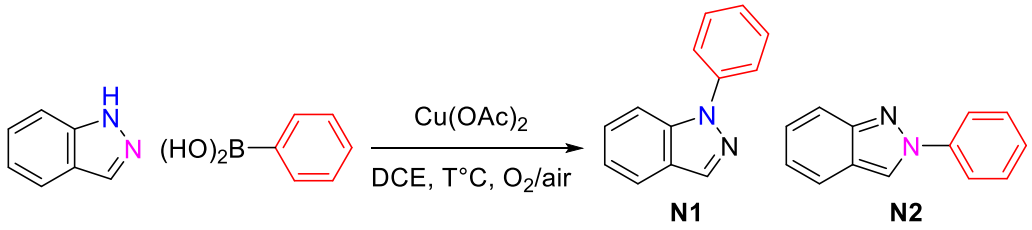


Figure S50. Mass-balance hyperspaces for the 5-bromo-7-azaindole Chan-Lam coupling a-e) Mass balance was determined relative to the initial amount of $PhB(OH)_2$ by summing all quantified phenylboronic acid-derived species: N-phenyl-5-bromo-7-azaindole, benzene, phenol, diphenyl ether, and unconsumed $PhB(OH)_2$. Maps are shown under different reaction conditions: 20 mM, O_2, for 24 h for all, and at 26, 50, 100 °C respectively.

1.9. 1H-indazole



Reaction design. For 1H-indazole, a seven-dimensional reaction space was explored. The reaction variables are [1H-indazole]$_0$ (20 and 65 mM), $Cu(OAc)_2$ loading (0.2 and 0.5-2 equiv in 0.5 intervals), $PhB(OH)_2$ loading (1.2–2.4 equiv in 0.4 intervals), pyridine loading (1 and 2 equiv), temperature (26, and 50 °C), atmosphere (air or O_2), and reaction time.

Following General Method **A**, reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, with a final reaction volume of 1.1 mL. First, a 75 mM $Cu(OAc)_2$ stock solution in water was added to the reaction vials and evaporated in a vacuum oven at 40 °C for 4–8 h. After evaporation, DCE was added, followed by the remaining DCE stock solutions: $PhB(OH)_2$ (250 mM), 1H-indazole(100 mM for 20mM initial indazole and 150 mM for 65 mM initial concentration), and pyridine (2000 mM for 20 mM initial reaction concentration and 4000 mM for 65 mM initial reaction concentration). When required, the reaction mixtures were sparged with oxygen, sealed, and placed in a preheated ThermoMixer C for the required reaction time. Preparation of one 54-vial reaction plate took approximately 45 min. In total, 339 independent reaction conditions were prepared (**Table S3**).

After completion of the selected reaction time, the plates containing the crude reaction mixtures were measured by Robowski-NMR, General Method **B**. Then crude mixtures were combined for purification on MPLC as described in the General Method **C**. The isolated compounds were then measured in DCE in a known concentrations (**Figure S53-Figure S55**) and added to the NMRanger basis reference set and used for the NMRanger quantification, General Method **D**.

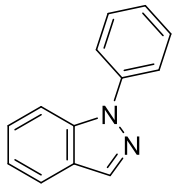
Reference Spectra preparation in DCE. Isolated 1-phenyl-1H-indazole and 2-phenyl-2H-indazole were measured in DCE for the reference library. For phenylboronic acid, benzene, phenol, and diphenyl ether, commercially available compounds were used. All spectra were recorded on an 80 MHz NMR spectrometer with 32 scans, a 6.4 s acquisition time, a 15 s repetition time, and

a 90° pulse angle (see **Section 5.1** for parameter selection). See **Section 1.5.1** for phenylboronic acid, benzene, phenol, and diphenyl ether spectra.

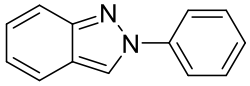
Table S3. Summary of the screening dimensions and included reaction conditions for the 1H-indazole dataset.

Initial concentration/ Atmosphere	T°C	Time	Screening set	Additional conditions	Screening dimensions	Included conditions
20 mM, O_2	26 °C	24 h	Main screening	6 conditions at 0.2 equiv $Cu(OAc)_2$	2×4×4	38
20 mM, O_2	26 °C	24 h	Excess $PhB(OH)_2$ and base	—	2×3×3	18
20 mM, O_2	26 °C	24 h	No pyridine	—	2×4×4	16
20 mM, air	26 °C	48 h	Main screening	—	2×4×4	32
20 mM, O_2	50 °C	24 h	Main screening	3 conditions at 0.2 equiv $Cu(OAc)_2$	2×4×4	35
20 mM, O_2	50 °C	24 h	No pyridine	—	4×4	16
20 mM, air	50 °C	48 h	Main screening	—	2×4×4	32
65 mM, O_2	26 °C	24 h	Concentration screening	—	4×3×3	36
65 mM, O_2	50 °C	24 h	Concentration screening	—	4×3×3	36
20 mM, O_2	26 °C	24 h	Alternative copper sources	4 copper sources	12 conditions per copper source	48
20 mM, O_2	26 °C	24 h	Alternative bases	4 bases	8 conditions per base	32
					Total	**339**

1.9.1. Characterizations and Reference Spectra for 1H-indazole Quantification



1-phenyl-1H-indazole (N1)[5]**: ^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.21 (d, *J* = 1.0 Hz, 1H), 7.84 – 7.72 (m, 4H), 7.59 – 7.51 (m, 2H), 7.44 (ddd, *J* = 8.3, 6.9, 1.1 Hz, 1H), 7.40 – 7.34 (m, 1H), 7.24 (ddd, *J* = 7.9, 6.9, 0.9 Hz, 1H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 140.4, 138.9, 135.4, 129.6, 127.3, 126.8, 125.5, 122.9, 121.7, 121.5, 110.6.



2-phenyl-2H-indazole (N2)[5]**: ^{1}H NMR** ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.42 (d, *J* = 1.1 Hz, 1H), 7.95 – 7.87 (m, 2H), 7.80 (dd, *J* = 8.8, 1.1 Hz, 1H), 7.72 (dt, *J* = 8.5, 1.1 Hz, 1H), 7.60 – 7.49 (m, 2H), 7.45 – 7.36 (m, 1H), 7.33 (ddd, *J* = 8.8, 6.6, 1.1 Hz, 1H), 7.12 (ddd, *J* = 8.4, 6.6, 0.9 Hz, 1H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 149.9, 140.7, 129.7, 128.1, 126.9, 122.9, 122.6, 121.2, 120.6, 120.5, 118.1.

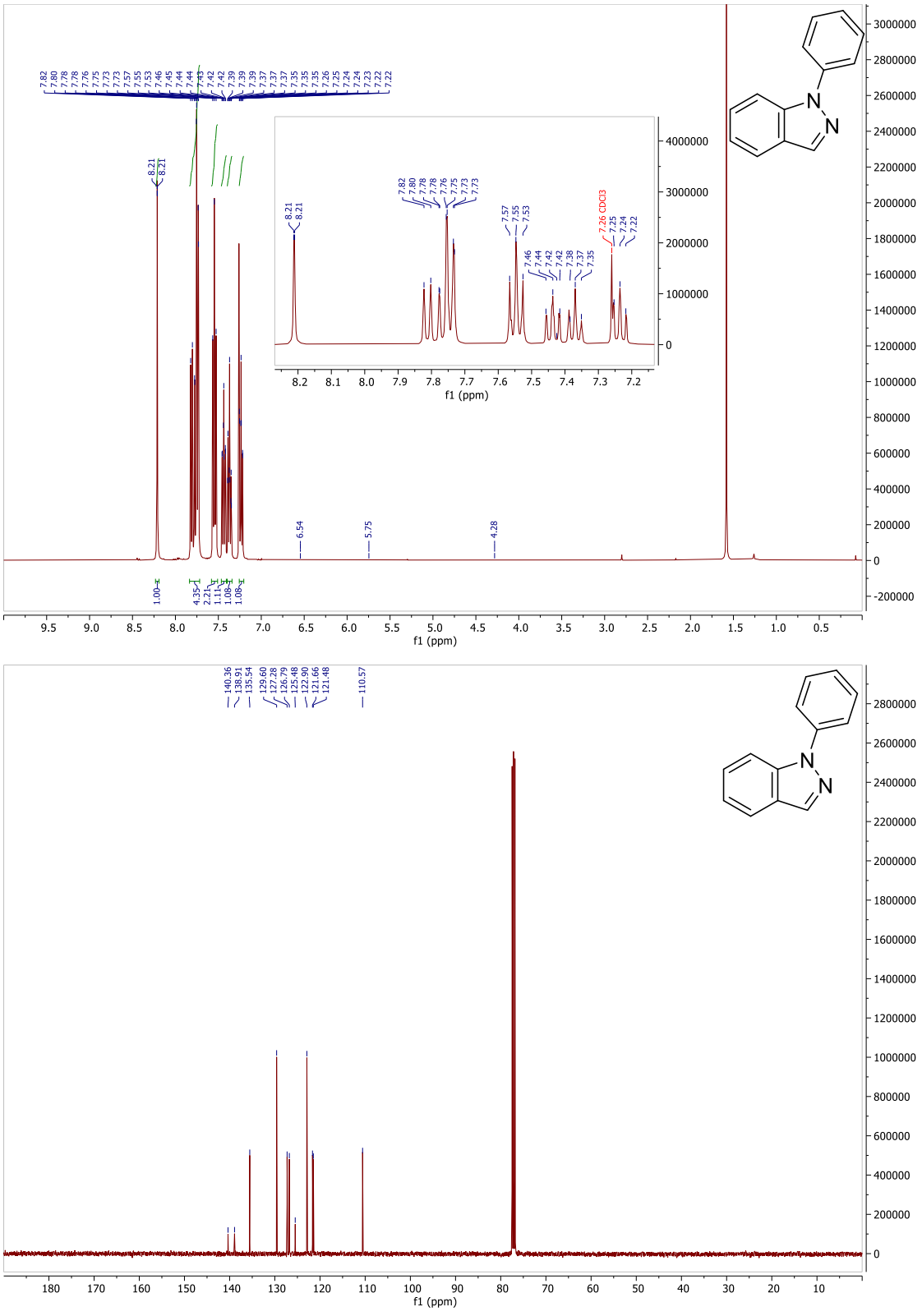


Figure S51. ^{1}H NMR (*top*) and ^{13}C NMR (*bottom*) spectra of 1-phenyl-1H-indazole in $CDCl_3$.

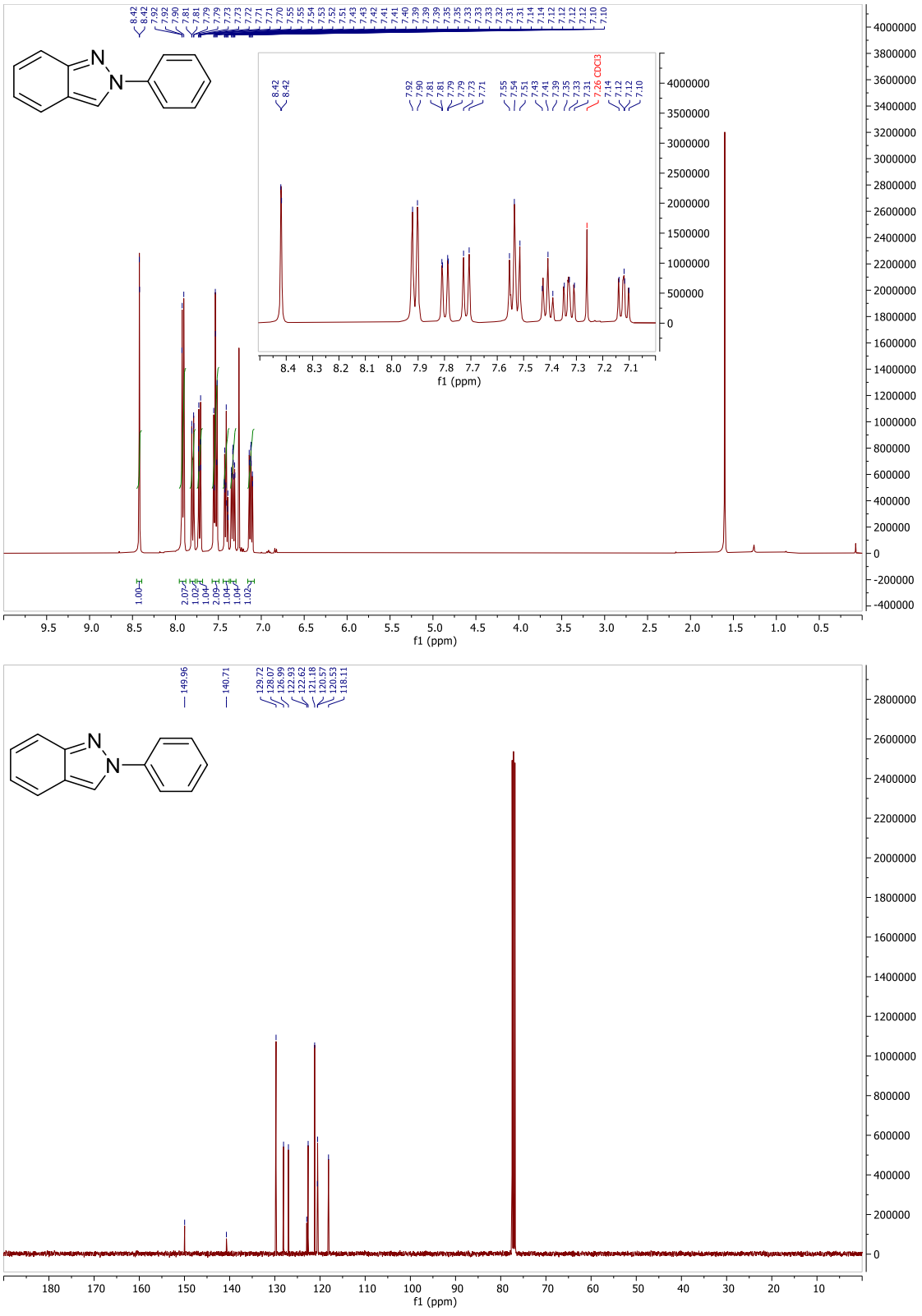


Figure S52. ^{1}H NMR (top) and ^{13}C NMR (*bottom*) spectra of 2-phenyl-2H-indazole in $CDCl_3$.

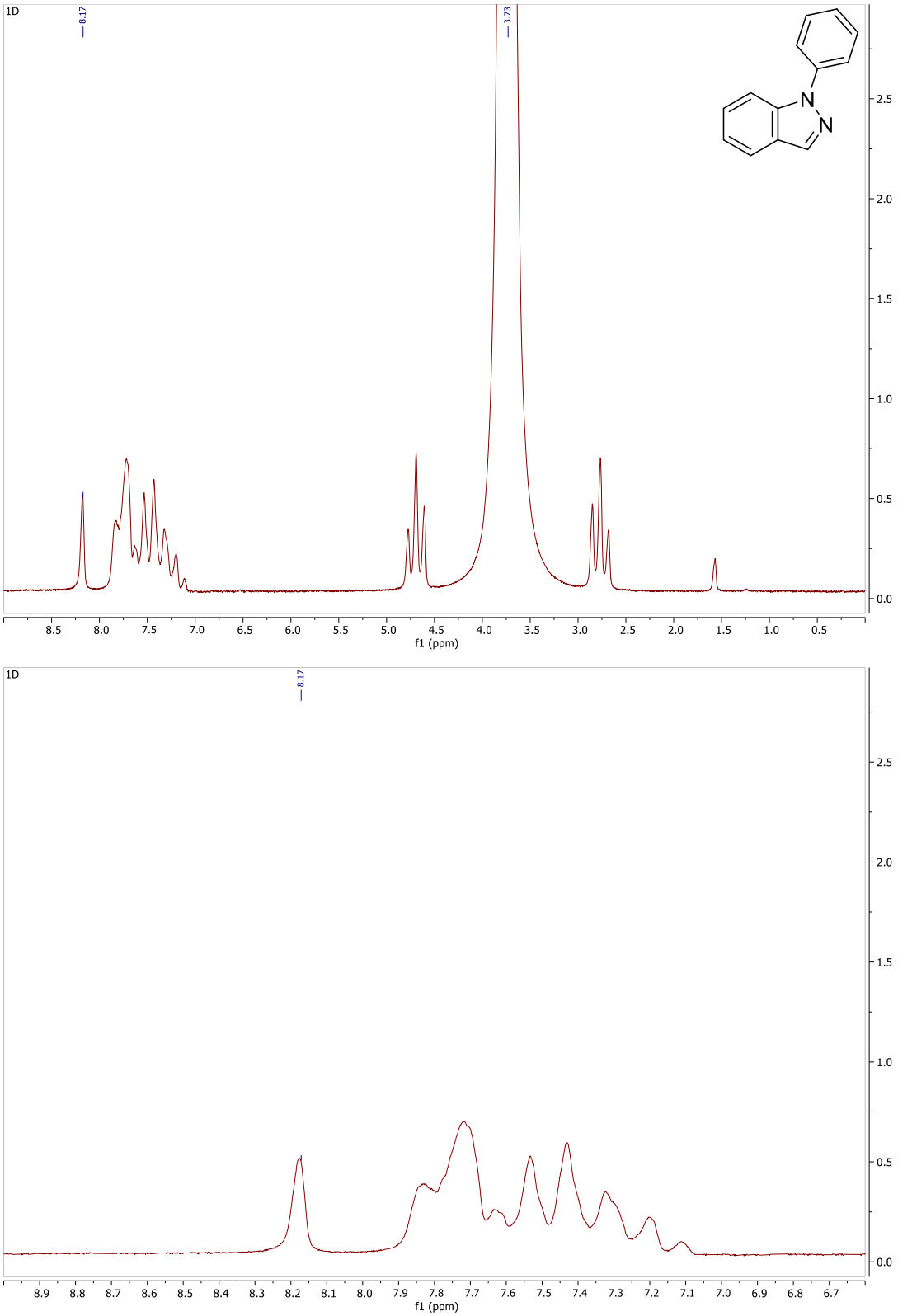


Figure S53. ^{1}H NMR spectrum of 1-phenyl-1H-indazole in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.6–9.0 ppm, (*bottom*) used for NMRanger quantification.

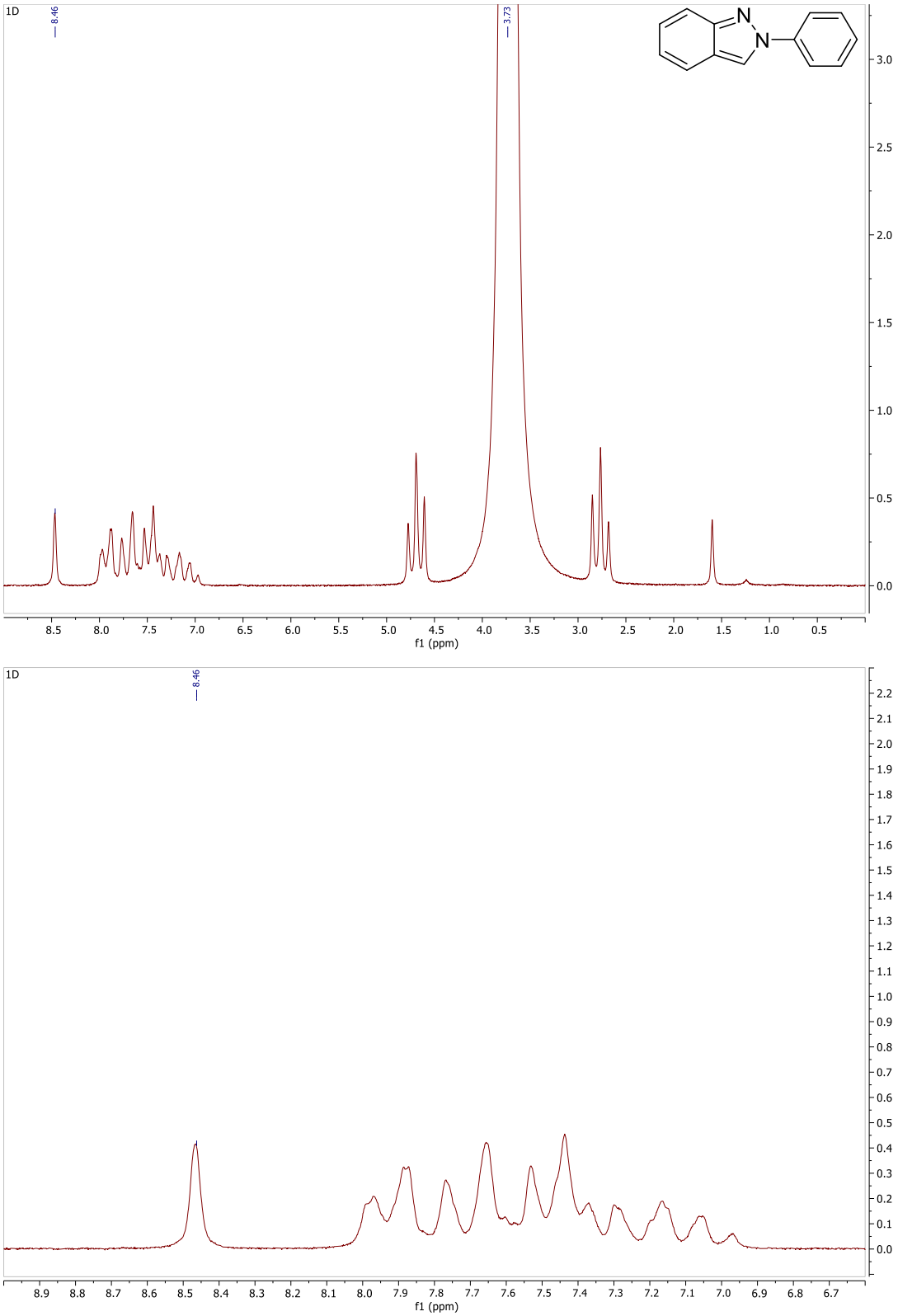


Figure S54. ^{1}H NMR spectrum of 2-phenyl-2H-indazole in DCE at 60 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.6–9.0 ppm, (*bottom*) used for NMRanger quantification.

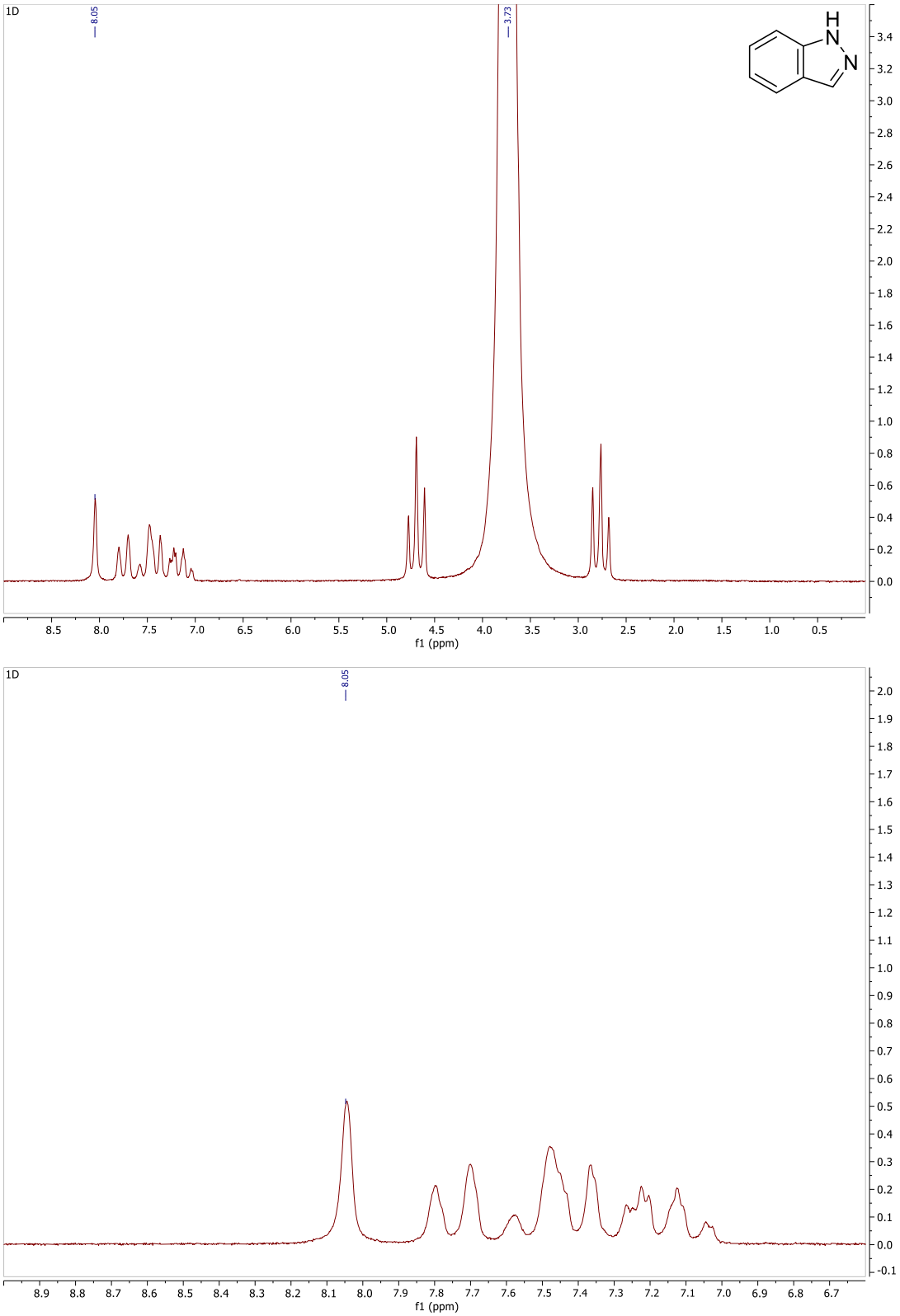


Figure S55. ^{1}H NMR spectrum of 1H-indazole in DCE at 100 mM concentration. Full spectrum (*top*), zoomed aromatic region, 6.6–9.0 ppm, (*bottom*) used for NMRanger quantification.

1.9.2. Calibration and fitting of 1H-indazole

A calibration curve for 1-phenyl-1H-indazole was generated automatically In NMRanger using DCE calibrant solutions of known concentration incorporated into the basis set (**Figure S56**). These calibrants served as external standards for yield estimation. Crude 1H-indazole reaction mixtures were quantified by multiple-region fitting: the characteristic 8.0–8.8 ppm region, containing product, was fitted first, and the resulting abundance estimates were transferred to guide fitting of the broader 6.6–9.0 ppm region. Because all components are fully aromatic, this range included all reference resonances required for quantification while excluding the non-deuterated solvent region, which contains no indazole or phenyl boronic acid-derived compound signals. Representative good, acceptable, and poor fits are shown in **Figure S57**-**Figure S59** illustrating spectral reconstruction across different product and by-product compositions before incorporation into the reaction hyperspaces discussed in **Section 4.3 of the main text.**

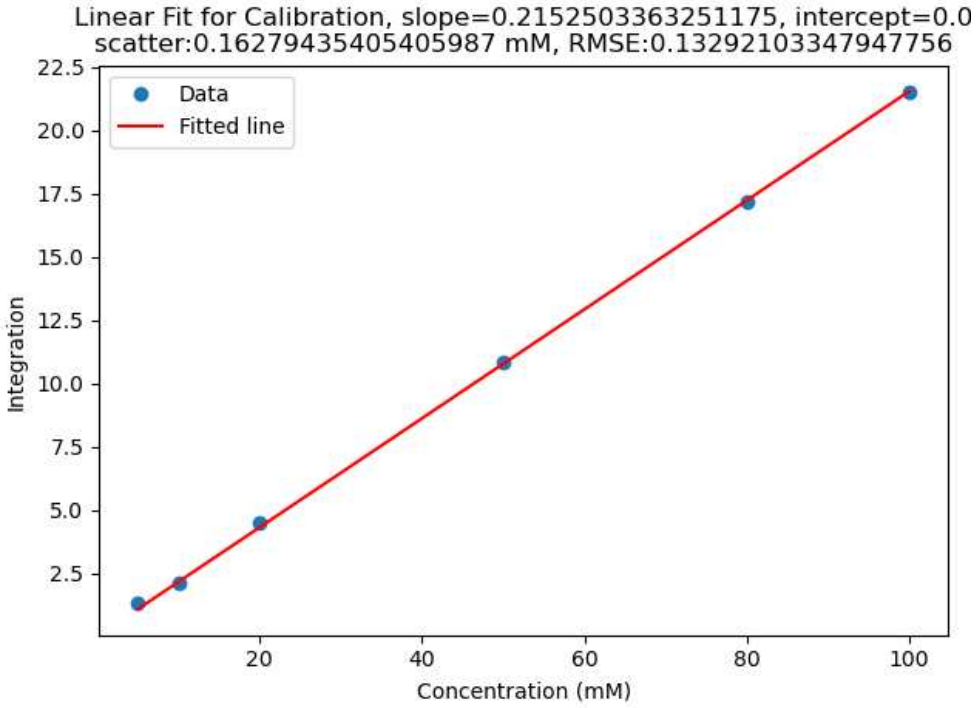


Figure S56. Calibration curve of 1-phenyl-1H-indazole generated automatically by NMRanger. Solutions containing known concentrations of 1-phenyl-1H-indazole were prepared in DCE at 5, 10, 20, 50, 80, 100 mM. The upper concentration was chosen based on the expected maximum product concentration in the morpholine coupling reactions.

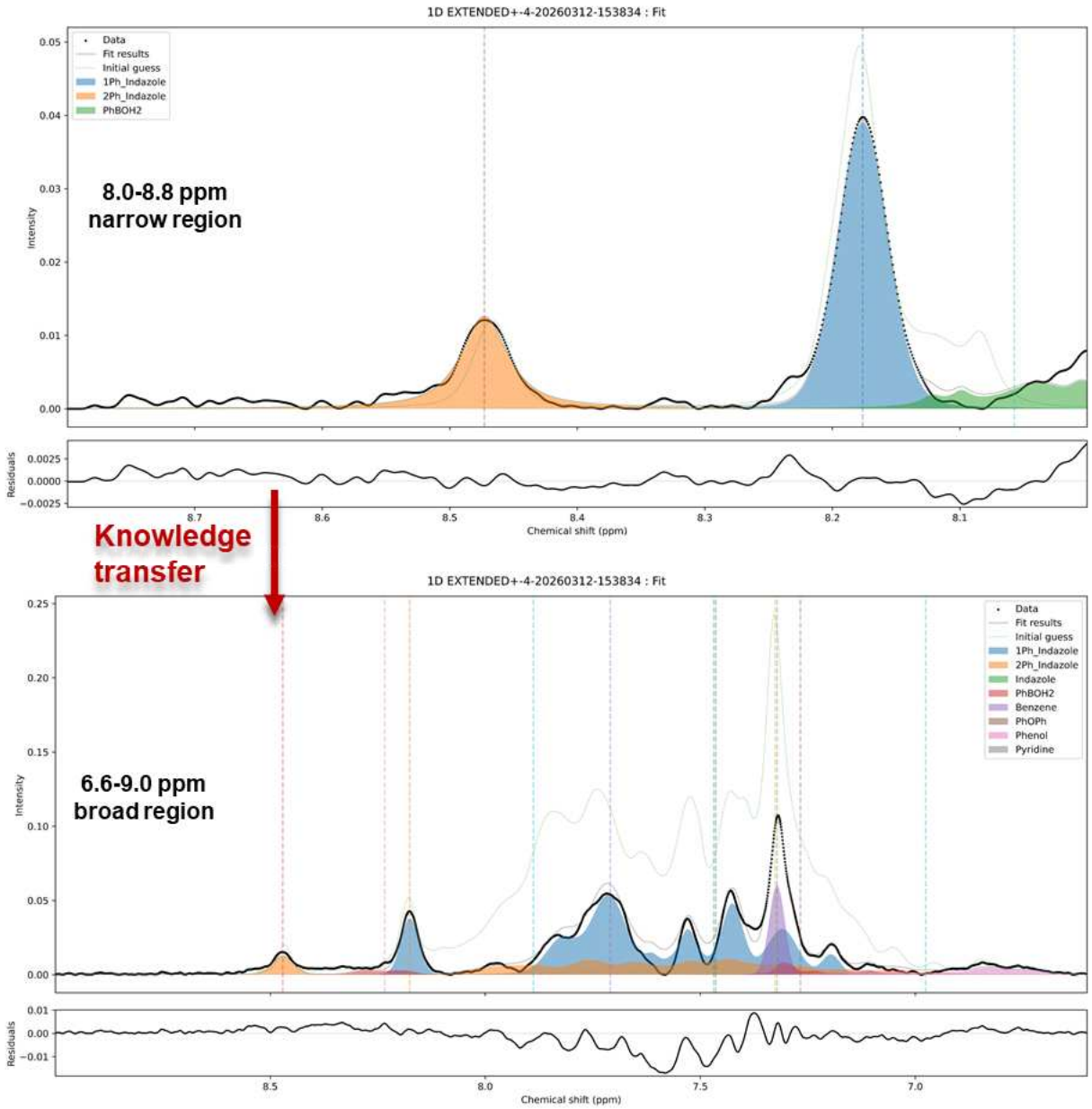


Figure S57. Representative multiple-region NMRanger good fit for 1H-indazole crude reaction mixtures. Fitting was performed first over the 8.0–8.8 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 6.6–9.0 ppm region, containing overlapping signals from 1-phenyl-1H-indazole (*blue*), 2-phenyl-2H-indazole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*lavender*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 1H-indazole, O_2 atmosphere, 26 °C, in the presence of 0.5 equiv $Cu(OAc)_2$, 1.2 equiv $PhB(OH)_2$, and 1.0 equiv pyridine. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. Good fit for a highly overlapped 1H-indazole crude spectrum; the fitted line closely reproduces the resonances.

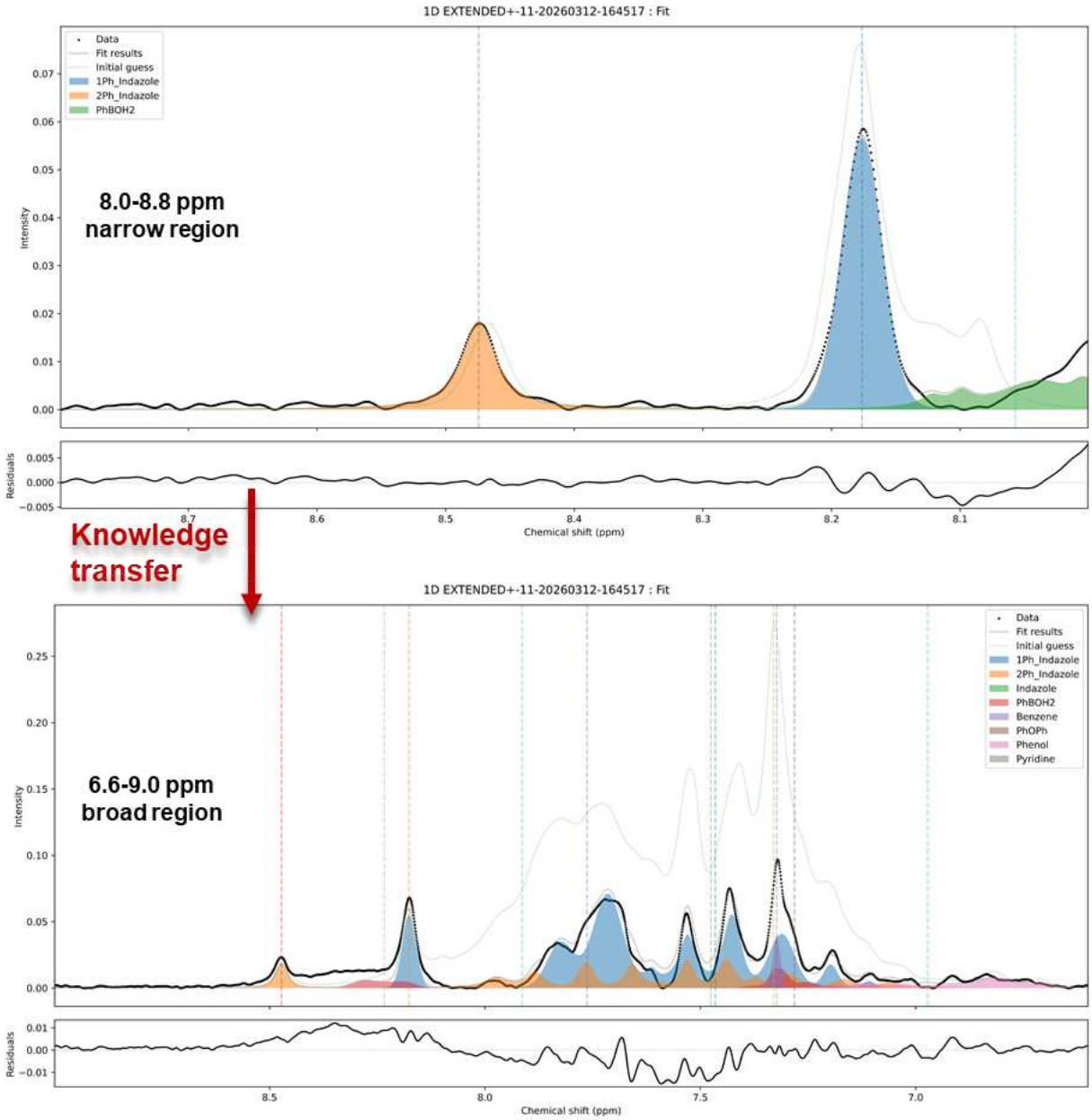


Figure S58. Representative multiple-region NMRanger acceptable fit for 1H-indazole crude reaction mixtures. Fitting was performed first over the 8.0–8.8 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 6.6–9.0 ppm region, containing overlapping signals from 1-phenyl-1H-indazole (*blue*), 2-phenyl-2H-indazole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*lavender*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 1H-indazole, O_2 atmosphere, 26 °C, in the presence of 2.0 equiv $Cu(OAc)_2$, 1.2 equiv $PhB(OH)_2$, and 2.0 equiv pyridine. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. The fitted line follows the main resonances and overall spectral envelope, but shows noticeable mismatches in the congested 7.4–7.8 ppm region.

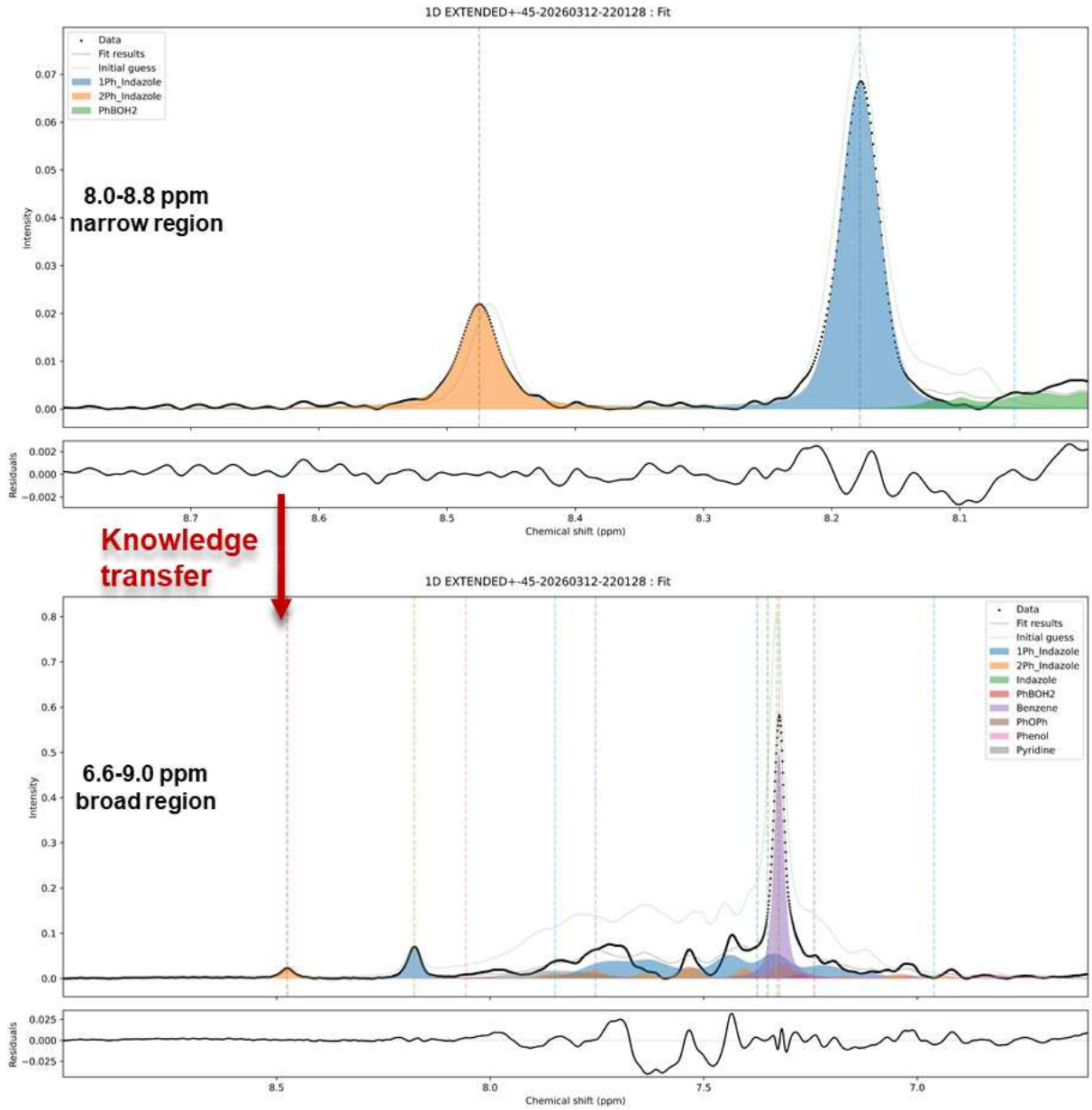


Figure S59. Representative poor multiple-region NMRanger fit for 1H-indazole crude reaction mixtures. Fitting was performed first over the 8.0–8.8 ppm aromatic region, and the resulting abundance estimates were used to solve the wider 6.6–9.0 ppm region, containing overlapping signals from 1-phenyl-1H-indazole (*blue*), 2-phenyl-2H-indazole (*orange*), $PhB(OH)_2$ (*rose*), benzene (*lavender*), pyridine (*grey*), phenol (*pink*), and diphenyl ether (*coral*) under the conditions of 20 mM initial concentration of 1H-indazole, O_2 atmosphere, 26 °C, in the presence of 1.0 equiv $Cu(OAc)_2$, 2.4 equiv $PhB(OH)_2$, and 2.0 equiv pyridine. Experimental spectra are shown as *black points*, the fitted spectra as *grey lines*, and individual fitted component contributions as *colored areas*; residuals are shown below each fit. In the 8.0–8.8 ppm narrow region, the fitted line reproduces the characteristic 1-phenyl-1H-indazole and 2-phenyl-2H-indazole resonances well. After knowledge transfer to the broader 6.6–9.0 ppm region, the dominant benzene peak near 7.3 ppm and the overall component distribution are captured; however, the low product yield results in weak product signals that are more difficult to fit, leading to noticeable mismatch across the heavily overlapped 7.4–7.8 ppm region.

1.10. Additional hyperspaces of 1H-Indazole

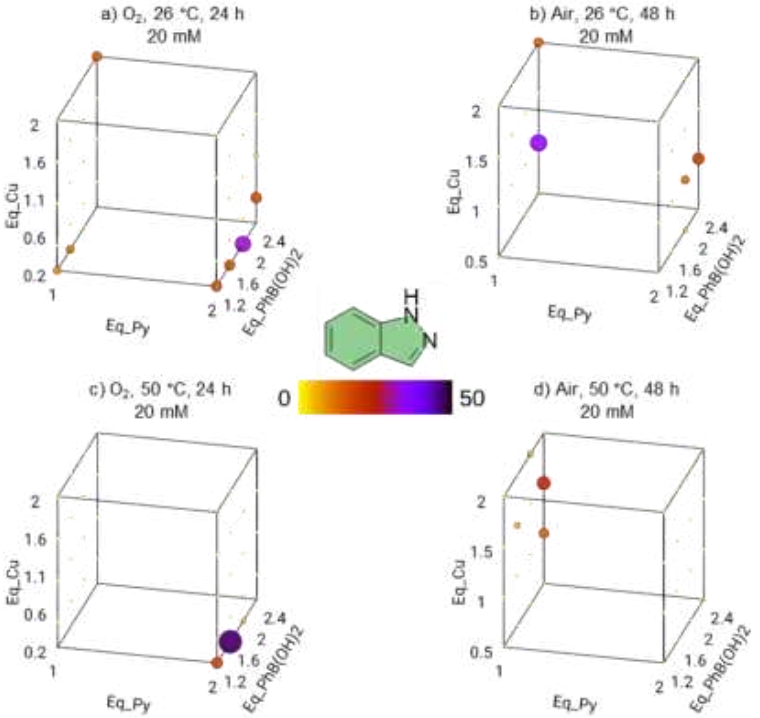


Figure S60. Unconsumed starting material hyperspaces for the 1H-indazole Chan-Lam coupling a-d) Yield maps under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

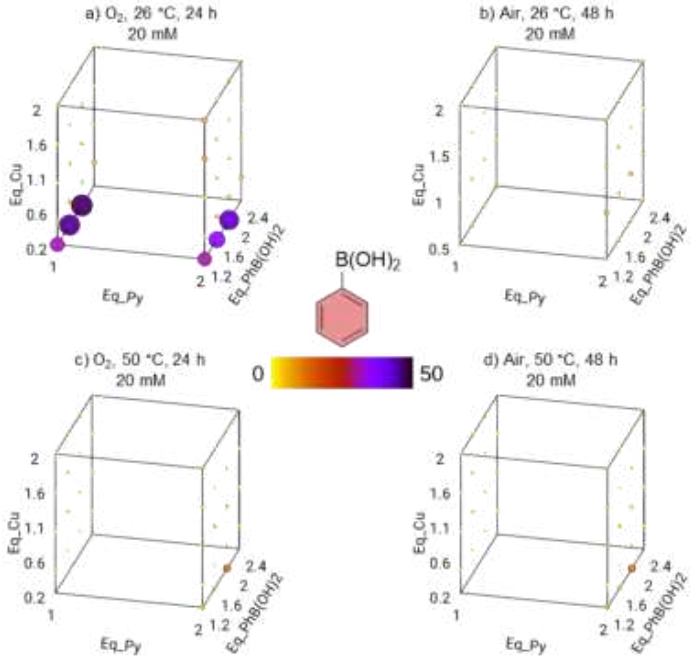


Figure S61. Unconsumed $PhB(OH)_2$ hyperspaces for the 1H-indazole Chan-Lam coupling a-d) Yield maps under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

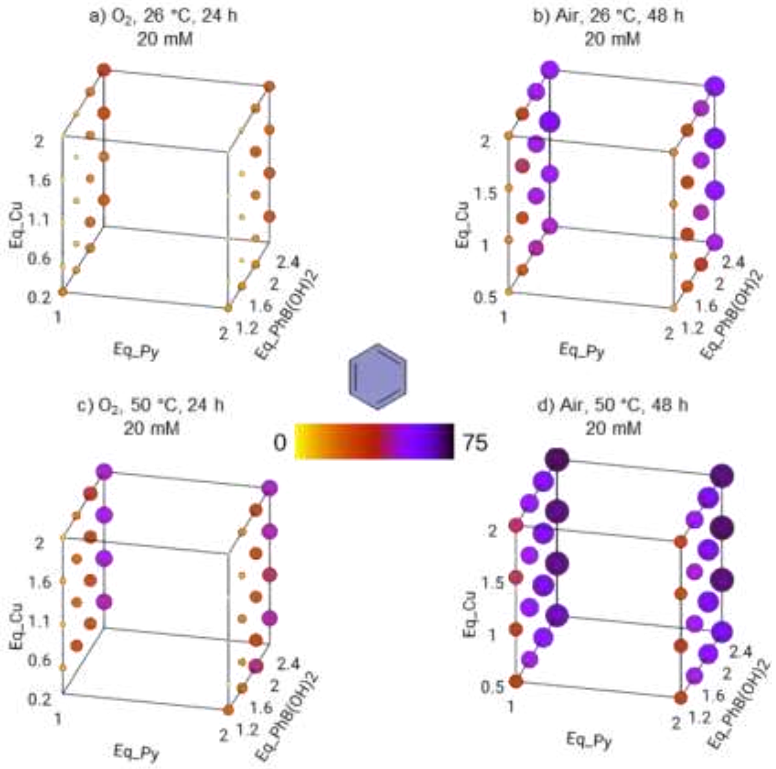


Figure S62. Benzene by-product formation hyperspaces for the 1H-indazole Chan-Lam coupling. a-d) Yield maps under different reaction conditions: O2, 26 °C, 24 h; air, 26 °C, 48 h; O2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

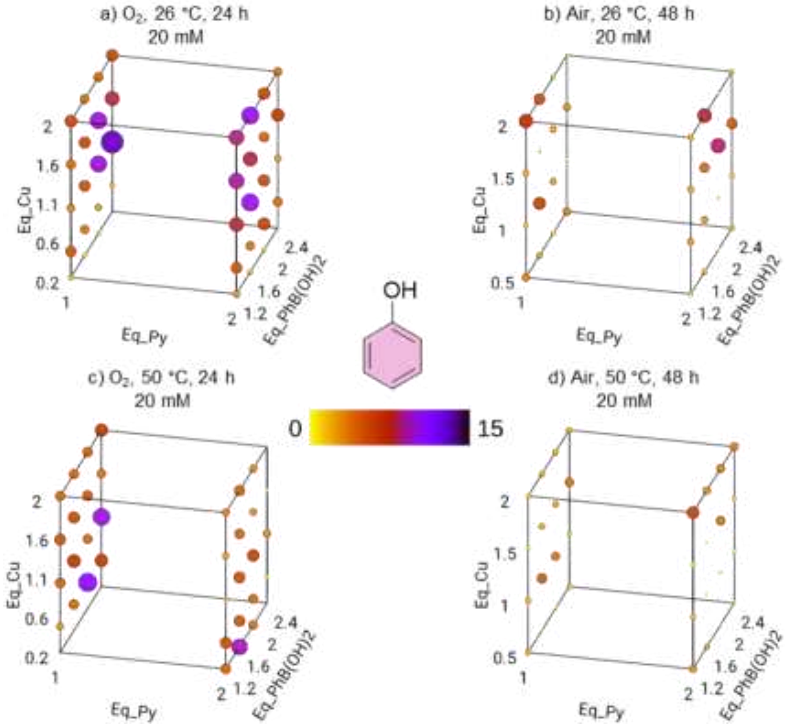


Figure S63. Phenol by-product formation hyperspaces for the 1H-indazole Chan-Lam coupling. a-d) Yield maps under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

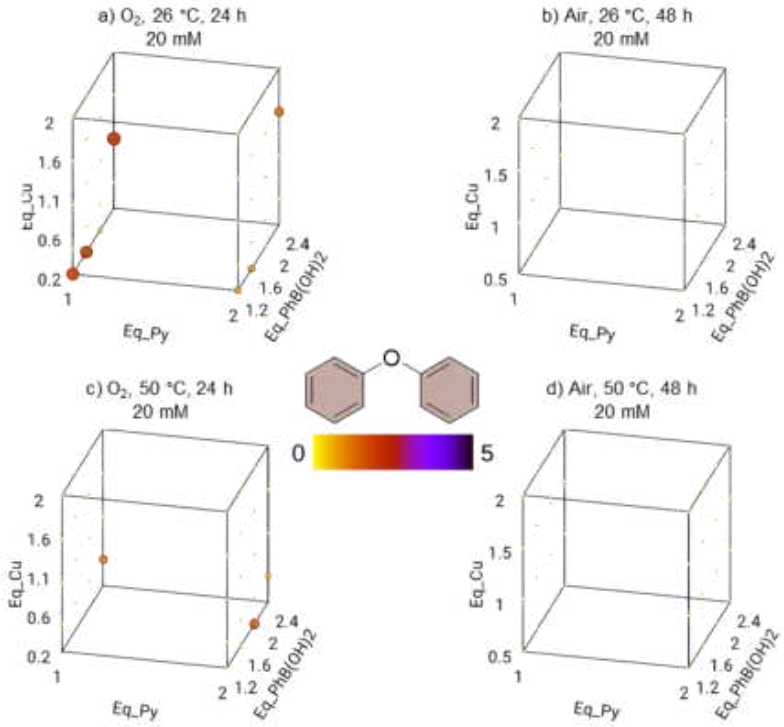


Figure S64. PhOPh by-product formation hyperspaces for the 1H-indazole Chan-Lam coupling. a-d) Yield maps under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

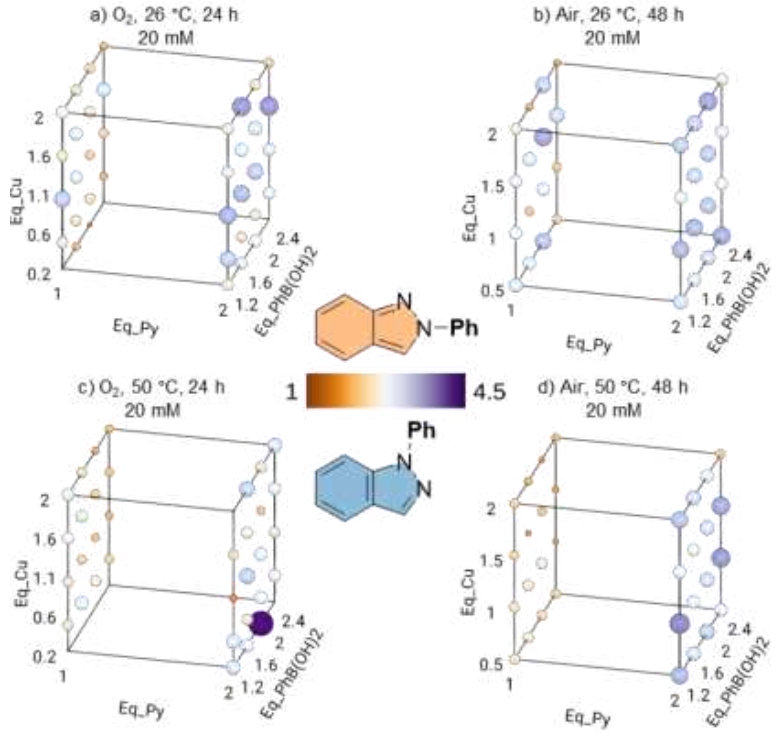


Figure S65. N1/N2 ratio hyperspaces for the 1H-indazole Chan-Lam coupling. a-d) Yield maps under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

When we evaluated the mass balance relative to phenylboronic acid (see General Method **D** for calculation details), the aryl source for the coupling products and aryl-derived by-products, the average mass balances are 73, 87, 76, and 97% for O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively. (**Figure S66**). The remaining deficit most likely arises, at least in part, from phenylboronic acid precipitation after the reaction, consistent with its limited solubility in DCE and the observed residue on the vial walls.

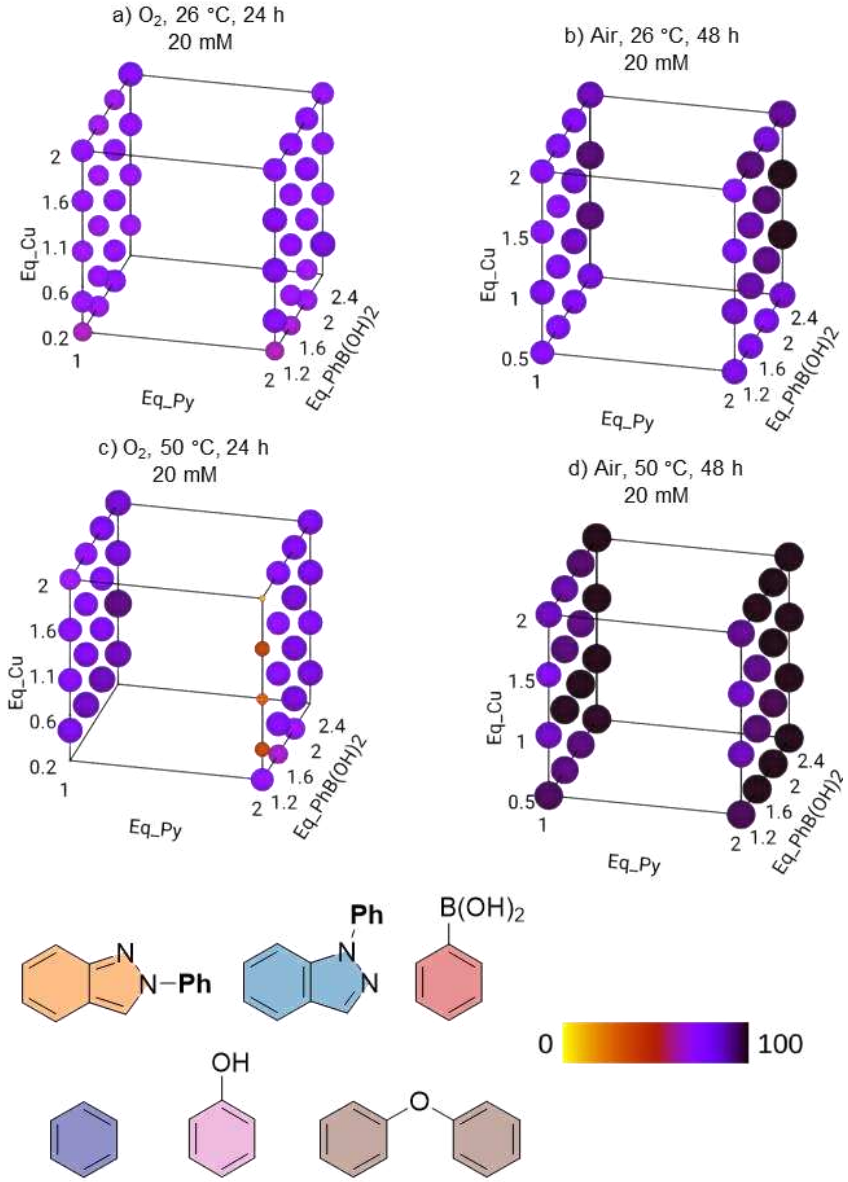


Figure S66. Mass-balance hyperspaces for the 1H-indazole Chan-Lam coupling a-e) Mass balance was determined relative to the initial amount of $PhB(OH)_2$ by summing all quantified phenylboronic acid-derived species: 1-phenyl-1H-indazole, 2-phenyl-2H-indazole, benzene, phenol, diphenyl ether, and unconsumed $PhB(OH)_2$. Maps are shown under different reaction conditions: O_2, 26 °C, 24 h; air, 26 °C, 48 h; O_2, 50 °C, 24 h; and air, 50 °C, 48 h, respectively.

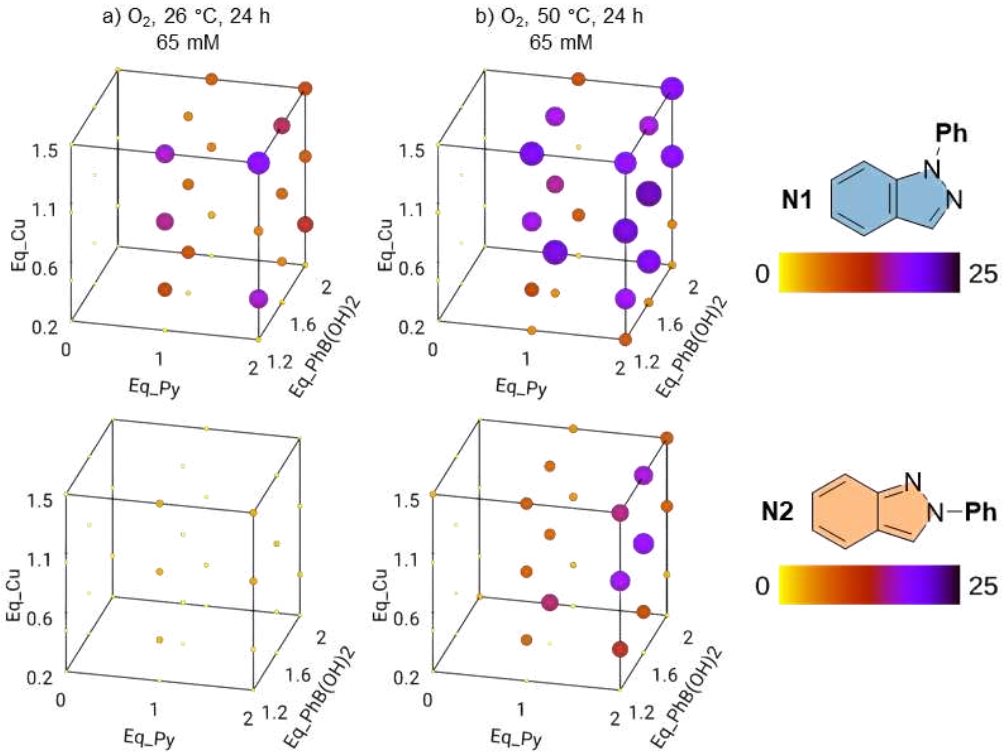


Figure S67. Coupling product yield hyperspaces at 65 mM initial 1H-indazole concentration. a) O_2, 24 h, at 26 °C. b) O_2, 24 h, at 50 °C.

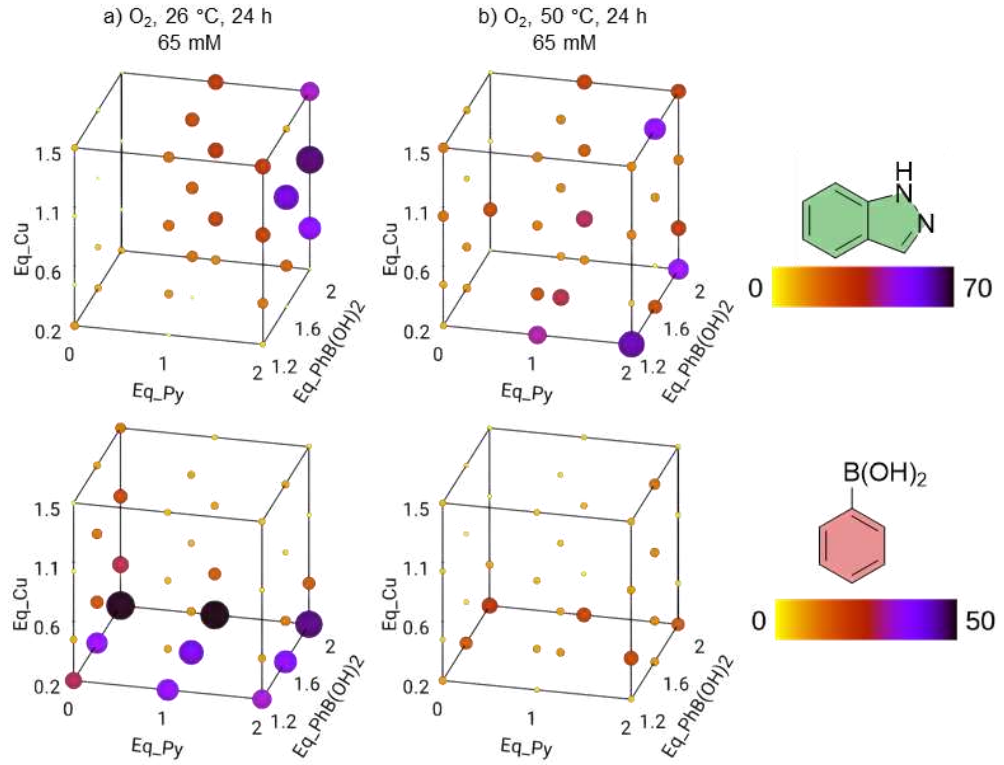


Figure S68. Unconsumed starting material yield hyperspaces at 65 mM initial 1H-indazole concentration. Unconsumed indazole (*top*), unconsumed $PhB(OH)_2$ (*bottom*). a) O_2, 24 h, at 26 °C. b) O_2, 24 h, at 50 °C.

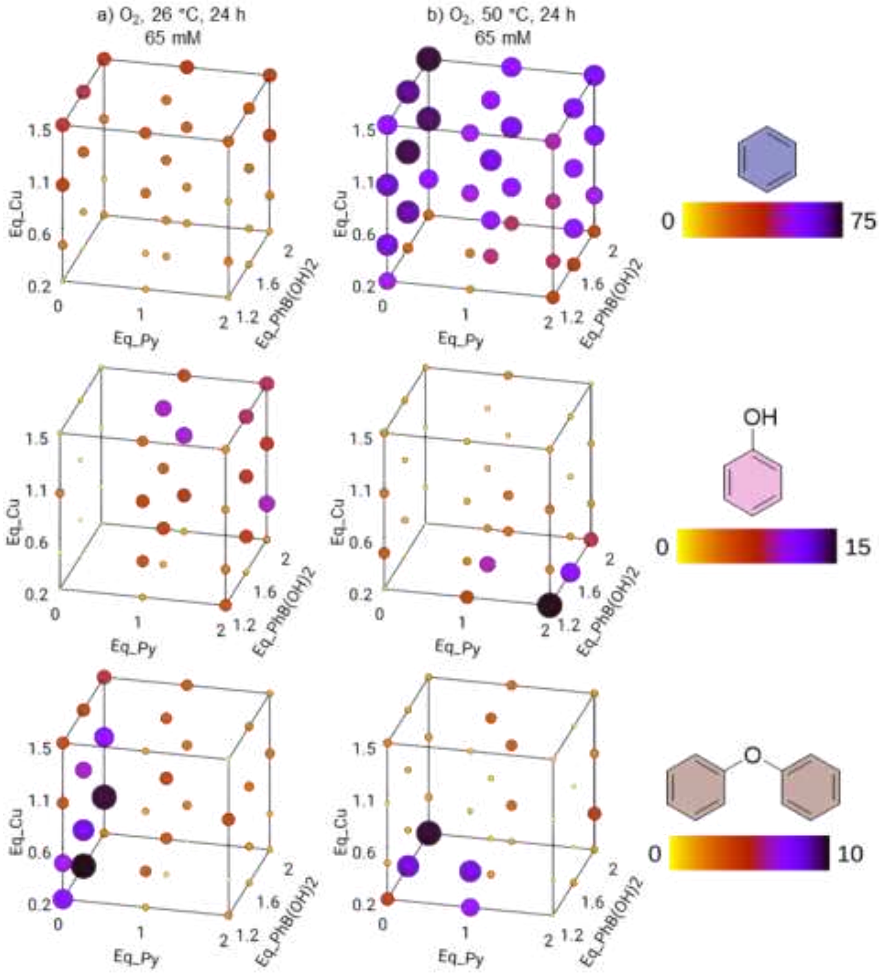


Figure S69. By-product yield hyperspaces at 65 mM initial 1H-indazole concentration. Benzene (*top*), phenol (*middle*), PhOPh (*bottom*). a) O_2, 24 h, at 26 °C. b) O_2, 24 h, at 50 °C.

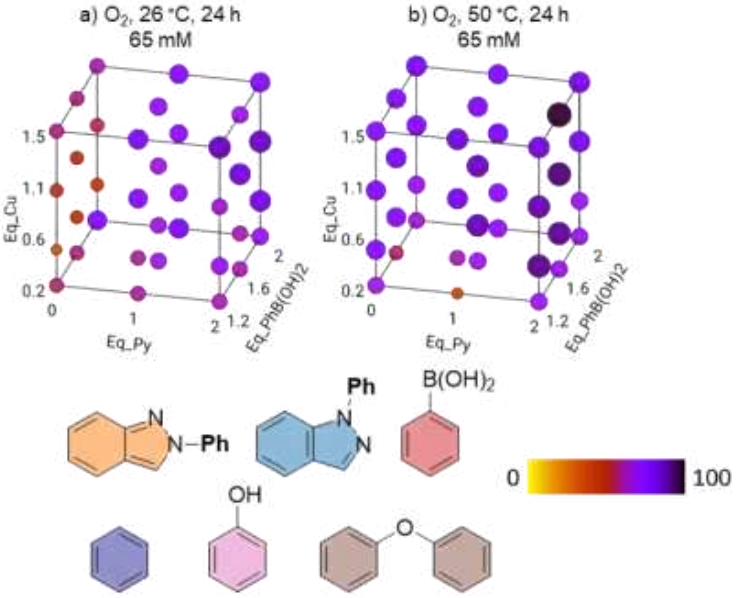


Figure S70. Mass balance hyperspaces at 65 mM initial 1H-indazole concentration. a) O_2, 24 h, at 26 °C. b) O_2, 24 h, at 50 °C.

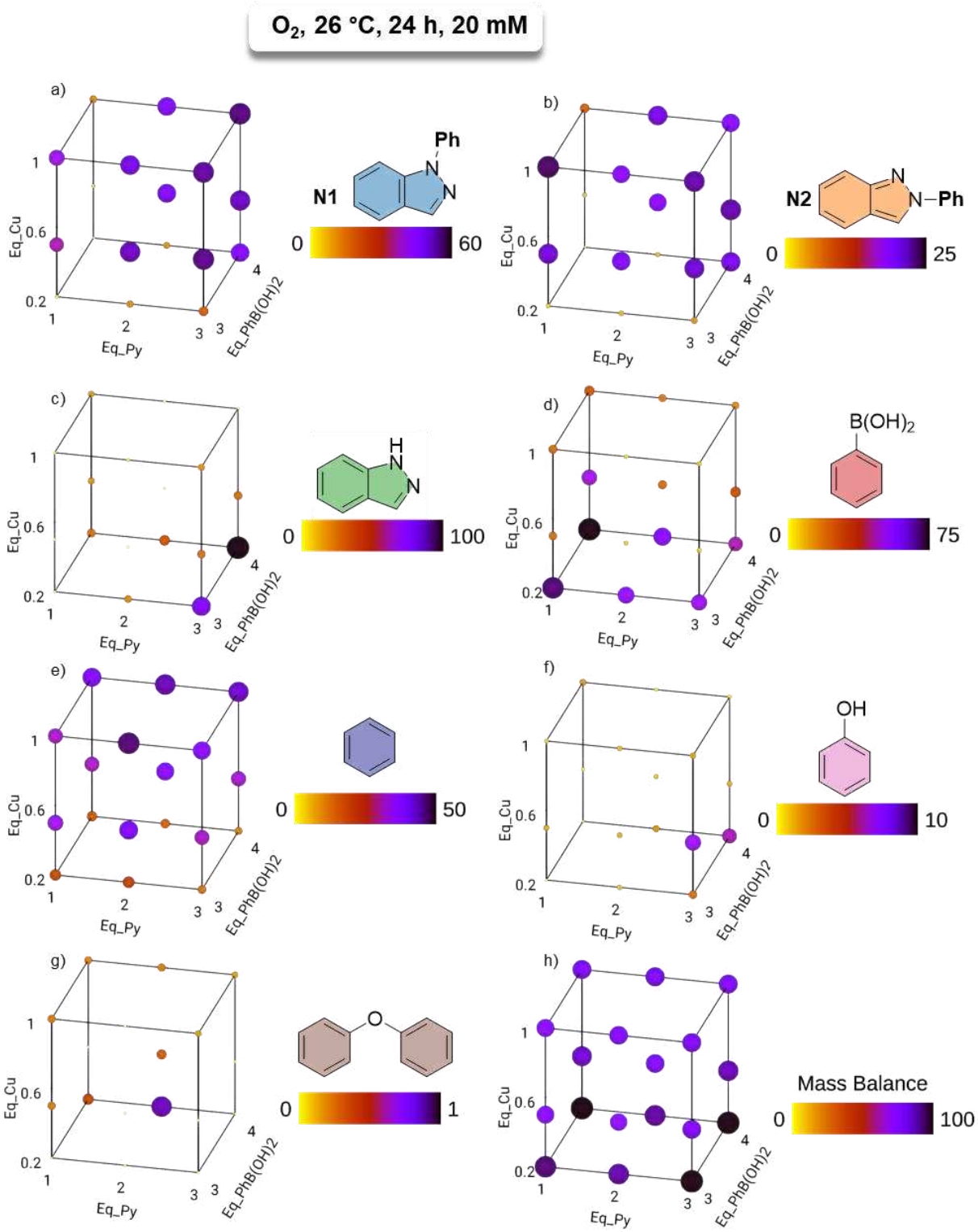


Figure S71. Hyperspaces for additional boronic acid and base equivalents. All reactions are under the same general condition; O_2, 26 °C, 24 h, 20 mM initial concentration. a) 1-phenyl-1H-indazole yield map. b) 2-phenyl-2H-indazole yield map. c) Unconsumed indazole yield map. d) Unconsumed $PhB(OH)_2$ yield map. e) Benzene by-product yield map. f) Phenol by-product yield map. g) PhOPh by-product yield map. h) Mass-balance map.

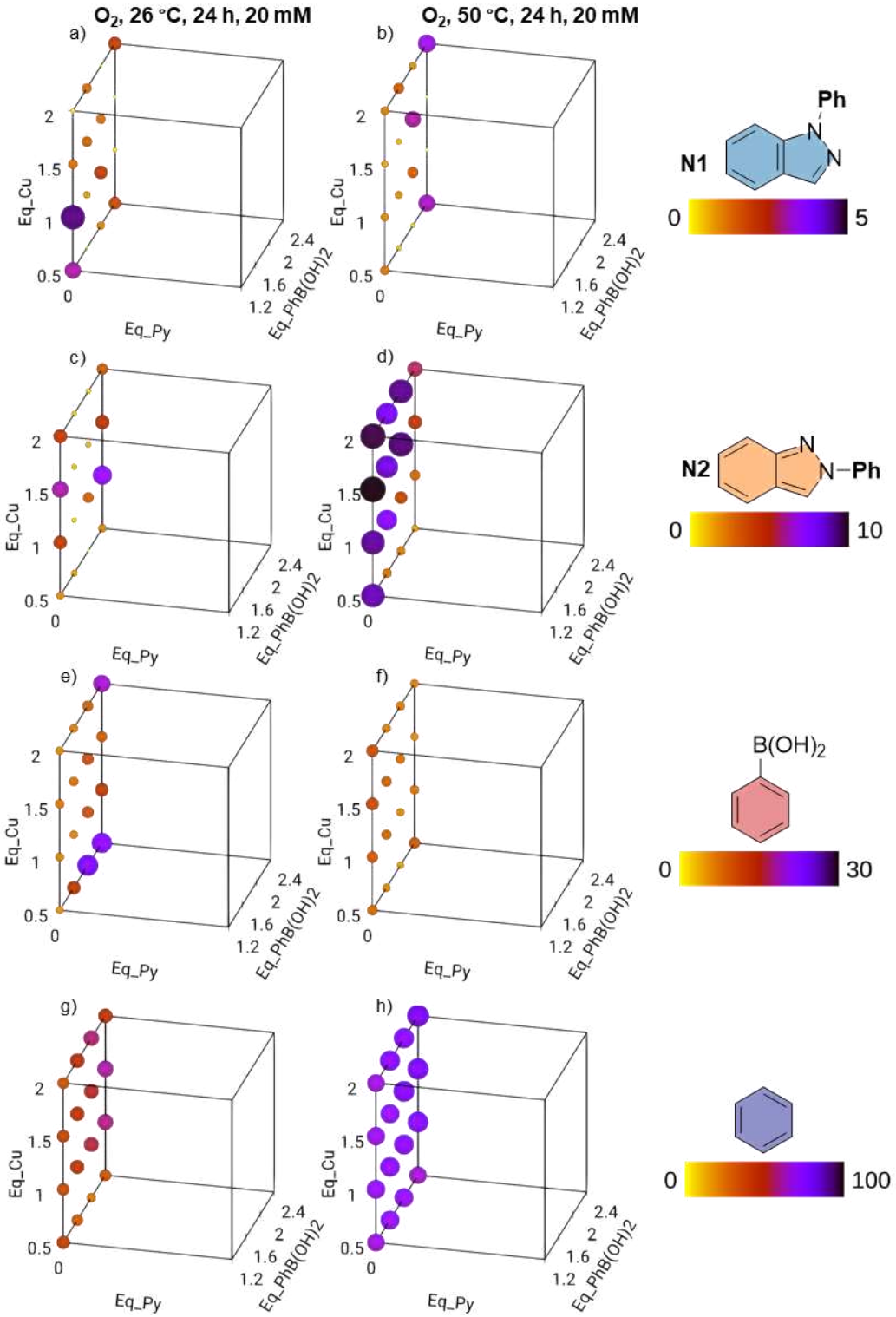


Figure S72. 2D Hyperspaces for 1H indazole reaction when no base used under O_2 at 26 and 50 °C for 24 h. a) 1-phenyl-1H-indazole yield map at 26 (*left*) and 50 °C (*right*). b) 2-phenyl-2H-indazole yield map at 26 (*left*) and 50 °C (*right*). c) Unconsumed $PhB(OH)_2$ yield map at 26 (*left*) and 50 °C (*right*). e) Benzene by-product yield map at 26 (*left*) and 50 °C (*right*).

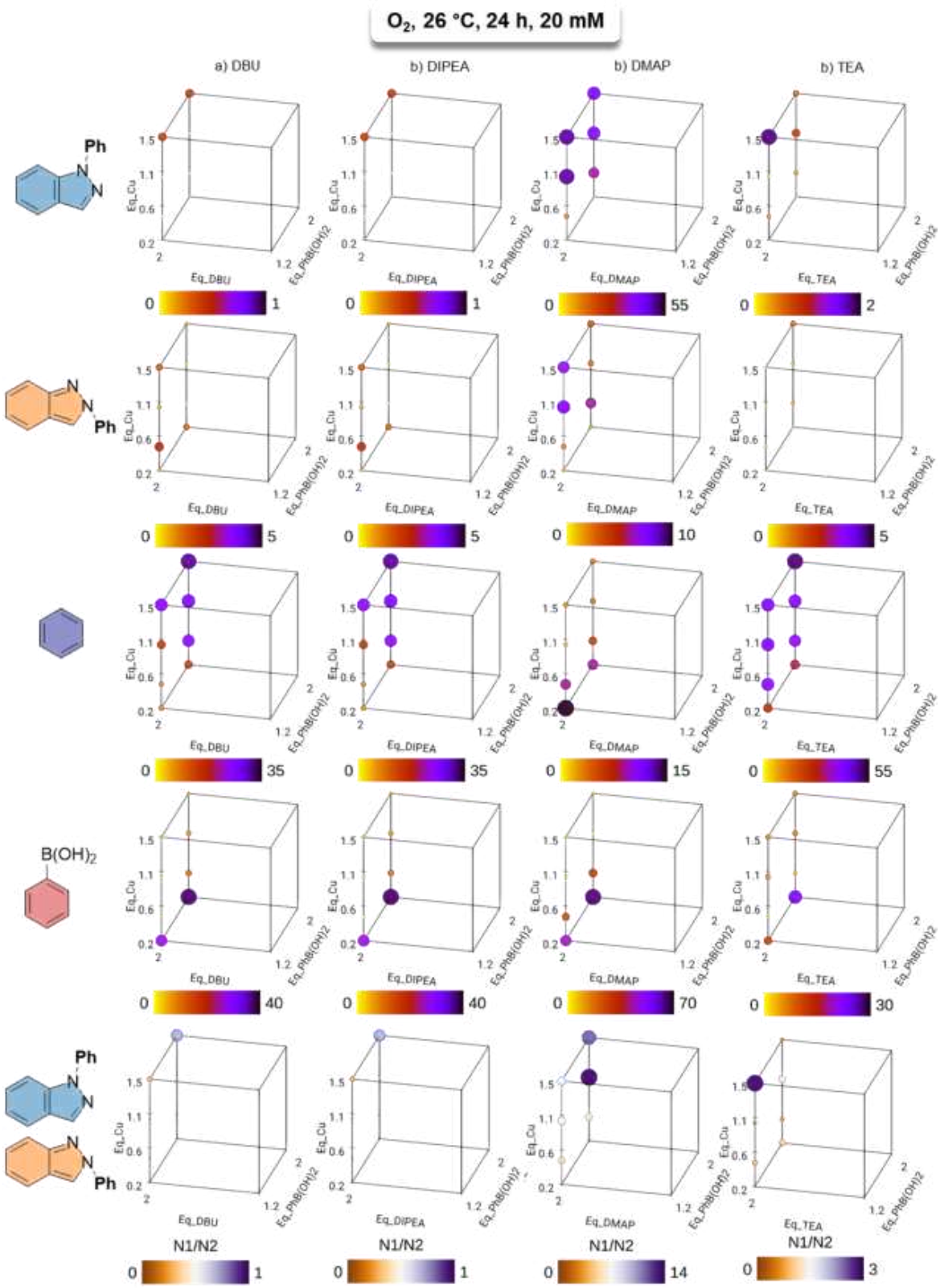


Figure S73. Reaction hyperspaces for 1H-indazole under O_2 at 26 °C for 24 h, showing the yields of N1, N2, benzene, unconsumed $PhB(OH)_2$, and N1/N2 ratio from top to bottom. Columns correspond to the base used: a) DBU hyperspaces. **b)** DIPEA hyperspaces. **c)** DMAP hyperspaces. **d)** TEA hyperspaces.

In the presence of DMAP, NMRanger estimated the N1/N2 ratio at up to 14%; however, this value is likely overestimated because the N2 signal is close to the noise level and therefore below the reliable sensitivity range of the fitting method. Manual integration of the corresponding resonances indicated an approximate maximum N2 contribution of 6% (**Figure S73**).

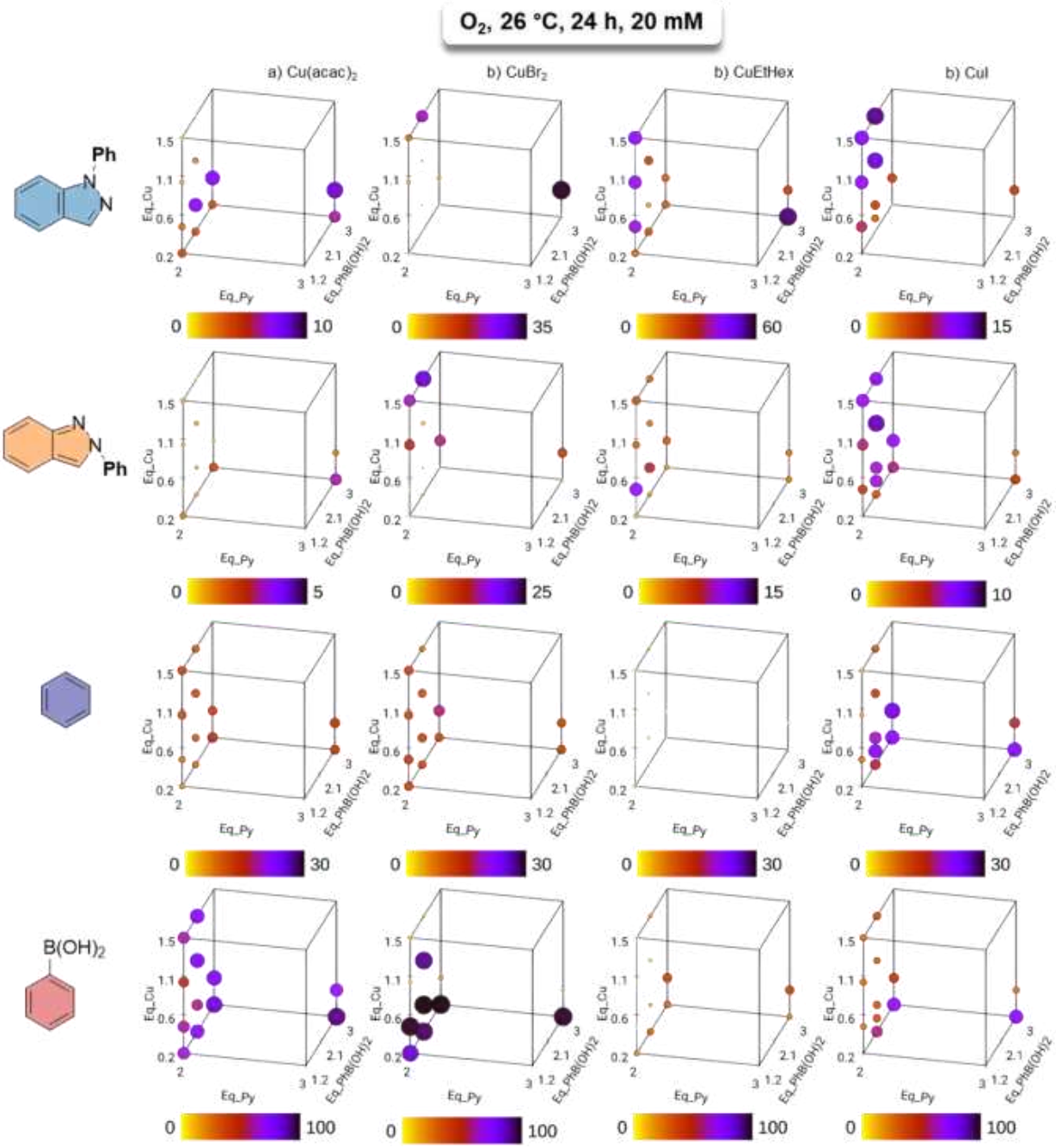


Figure S74. Reaction hyperspaces for 1H-indazole under O_2 at 26 °C for 24 h, showing the yields of N1, N2, benzene, and unconsumed $PhB(OH)_2$ from top to bottom. Columns correspond to the base used: **a)** $Cu(acac)_2$ hyperspaces. **b)** $CuBr_2$ hyperspaces. **c)** copper(II) 2-ethylhexanoate (CuEtHex) hyperspaces. **d)** CuI hyperspaces.

1.11. NMRanger accuracy validation experiments

1.11.1. Morpholine validations

Following General Method **A**, 54 reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, each with a final volume of 1000 µL. All reactions were performed under identical conditions: 20 mM initial morpholine concentration, 0.9 equiv $Cu(OAc)_2$, and 2.0 equiv $PhB(OH)_2$ under O_2 at 40 °C for 24 h. A 75 mM aqueous $Cu(OAc)_2$ stock solution was first dispensed into the vials and evaporated in a vacuum oven at 40 °C for 4–8 h. After drying, 820 µL of DCE was added to each vial, followed by 20 µL of a 1000 mM morpholine stock solution and 160 µL of a 250 mM $PhB(OH)_2$ stock solution, both prepared in DCE. The reaction mixtures were sparged with oxygen, the vial plate was sealed, and the reactions were incubated in a preheated ThermoMixer for 24 h.

After completion, the contents of all 54 vials were combined and concentrated to dryness. The residue was passed through a short silica pad to remove residual copper salts, washed with DCE, and evaporated. The resulting crude mixture was dissolved in HPLC-grade $CHCl_3$ and filtered through a 0.2 µm PTFE syringe filter (13 mm) before injection onto the GPC column.

Each vial contained 0.020 mmol of morpholine in a total reaction volume of 1000 µL, corresponding to 1.08 mmol across 54 reaction vials. N-Phenylmorpholine was isolated in 90.41 mg (0.55 mmol), corresponding to a 51.3% isolated yield. Under these conditions, NMRanger estimated a yield of 56.1%, in close agreement with the isolated yield.

1.11.2. 7-azaindole validations

Three independent validation methods are used to determine NMRanger accuracy, isolation, NMR yield determined by using an internal standard, and analytical HPLC analysis.

1) Isolation. Following General Method **A**, 27 reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, each with a final volume of 1150 µL. All reactions were performed under identical conditions: 20 mM initial 7-azaindole concentration, under O_2 at 26 °C for 24 h and 1.0 equiv $Cu(OAc)_2$, 2.0 equiv $PhB(OH)_2$, and 1.5 equiv DBU. A 75 mM aqueous $Cu(OAc)_2$ stock solution was first dispensed into the vials and evaporated in a vacuum oven at 40 °C for 4–8 h. After drying, 908.5 µL of DCE was added to each vial, followed by 23

µL of a 1000 mM 7-azaindole stock solution, 184 µL of a 250 mM $PhB(OH)_2$ stock solution, and 34.5 µL of a 1000 mM DBU stock solution, all prepared in DCE. The reaction mixtures were sparged with oxygen, the vial plate was sealed, and the reactions were incubated in a preheated ThermoMixer for 24 h.

After completion, the contents of all 27 vials were combined and concentrated to dryness. The residue was passed through a short silica pad to remove residual copper salts, washed with DCE, and evaporated. The resulting crude mixture was dissolved in HPLC-grade $CHCl_3$ and filtered through a 0.2 µm PTFE syringe filter (13 mm) before injection onto the GPC column.

Each vial contained 0.022 mmol of 7-azaindole in a total reaction volume of 1150 µL, corresponding to 0.594 mmol across 27 reaction vials. N-Phenyl-7-azaindole was isolated in 73.07 mg (0.376 mmol), corresponding to a 63.3% isolated yield. Under these conditions, NMRanger estimated a yield of 68.6%, in close agreement with the isolated yield.

Unconsumed $PhB(OH)_2$ was also isolated from the same pooled sample as 38.05 mg (0.312 mmol), corresponding to 28.9% of the initial amount. NMRanger estimated 32.0% unconsumed $PhB(OH)_2$, again in close agreement with the isolated value.

2) NMR yield determined by using an internal standard. Following General Method **A**, eight reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, each with a final volume of 1150 µL. Reactions were performed with 20 mM 7-azaindole under O_2 at 26 °C for 24 h. A 75 mM aqueous $Cu(OAc)_2$ stock solution was dispensed and evaporated at 40 °C for 4–8 h, after which DCE, 23 µL of 1000 mM 7-azaindole, and the required volumes of 250 mM $PhB(OH)_2$ and 1000 mM DBU stock solutions were added (**Table S4**). The mixtures were sparged with O_2, sealed, and incubated in a preheated ThermoMixer for 24 h.

After completion, a known amount of 1,3,5-trimethoxybenzene (TriOMePh) was added to each NMR tube. Yields estimated by NMRanger were compared with manual internal-standard integrations performed in Mnova, giving a mean absolute difference of 6.58 (**Table S5**). As an additional cross-check, the known TriOMePh concentrations were compared with the values estimated by NMRanger and showed good agreement, with a mean absolute difference of 1.71 mM (**Table S6**).

Table S4. Volumes of $Cu(OAc)_2$, DCE, 7-azaindole, DBU, and $PhB(OH)_2$ stock solutions added to each reaction vial.

local_index	vol_$Cu(OAc)_2$	vol_DCE	vol_Azaindole	vol_DBU	vol_$PhB(OH)_2$
1	184	963.7	23	34.5	128.8
3	306.67	970.6	23	46	110.4
12	306.67	982.1	23	34.5	110.4
17	61.33	982.1	23	34.5	110.4
25	245.33	963.7	23	34.5	128.8
26	61.33	970.6	23	46	110.4
27	122.67	933.8	23	46	147.2
36	122.67	915.4	23	46	165.6

Table S5. Reagent equivalents and comparison of N-phenyl-7-azaindole yields determined by NMRanger and by internal-standard NMR using TriOMePh.

Equivalent $Cu(OAc)_2$;$PhB(OH)_2$;DBU	TriOMePh (mg)	Product NMRanger Yield %	Product Cpmparative Yield %	Absolute Difference %
0.6;1.4;1.5	0.87	81.1	85.4	4.3
1.0;1.2;2.0	0.88	67.5	60.0	7.5
1.0;1.2;1.5	0.94	75.8	87.4	11.6
0.2;1.2;1.5	0.93	31.1	34.3	3.2
0.2;1.2;2.0	0.86	25.8	31.5	5.7
0.4;1.6;2.0	0.83	57.3	63.7	6.4
0.4;1.8;2.0	0.91	60.8	66.5	5.7
0.4;1.4;1.5	4.22	84.3	92.5	8.2
			Mean Absolute Error %	6.58

Table S6. Comparison of the ground-truth and NMRanger-determined concentrations of TriOMePh.

TriOMePh (mg)	DCE (mL)	$[Ground\ Truth]_0$	$[NMRanger]_0$	Absolute Difference (mM)
0.87	0.4	12.93	12.43	0.50
0.88	0.5	10.46	11.51	1.05
0.94	0.5	11.18	9.95	1.23
0.93	0.5	11.06	8.79	2.27
0.86	0.5	10.23	8.84	1.39
0.83	0.5	9.87	8.97	0.90
0.91	0.5	10.82	9.52	1.30
4.22	0.5	50.18	45.09	5.09
			Mean Absolute Error (in mM)	1.71

3) Analytical HPLC analysis. A set of 48 reactions was prepared at an initial 7-azaindole concentration of 20 mM under O_2 at 26 °C. Each reaction mixture was analyzed independently by both Robowski NMR and HPLC. HPLC analyses were performed using a photodiode array detector (PDA) with a mobile phase consisting of acetonitrile (solvent A) and water containing 0.05% TFA (solvent B), at a flow rate of 0.8 mL/min and a column-oven temperature of 30 °C. The gradient program was as follows: 30% B at 0.01 min, 20% B at 10.0 min, held at 20% B until 15.0 min, increased to 30% B at 20.0 min, and held at 30% B until 40.0 min. The total run time is 40 min per sample.

For HPLC quantification, a calibration curve for N-phenyl-7-azaindole was generated from standards of known concentration and applied analogously to the external-calibration procedure used in NMRanger **(Figure S75)**. Product yields determined by HPLC were compared directly with the corresponding NMRanger estimates, and separate reaction-hyperspace cubes were generated from the HPLC and NMRanger datasets **(Figure S76)**. Representative HPLC chromatograms from pure N-phenyl-7-azaindole used for calibrations in **Figure S77** and from the crude reaction mixtures are shown in **Figure S78**. The entire report of HPLC can be obtained from Zenodo repository: https://zenodo.org/records/21373963. The NMRanger-estimated yields showed good agreement with the HPLC results, with R^2=0.766 and a mean absolute error of 6.7% (**Table S7**).

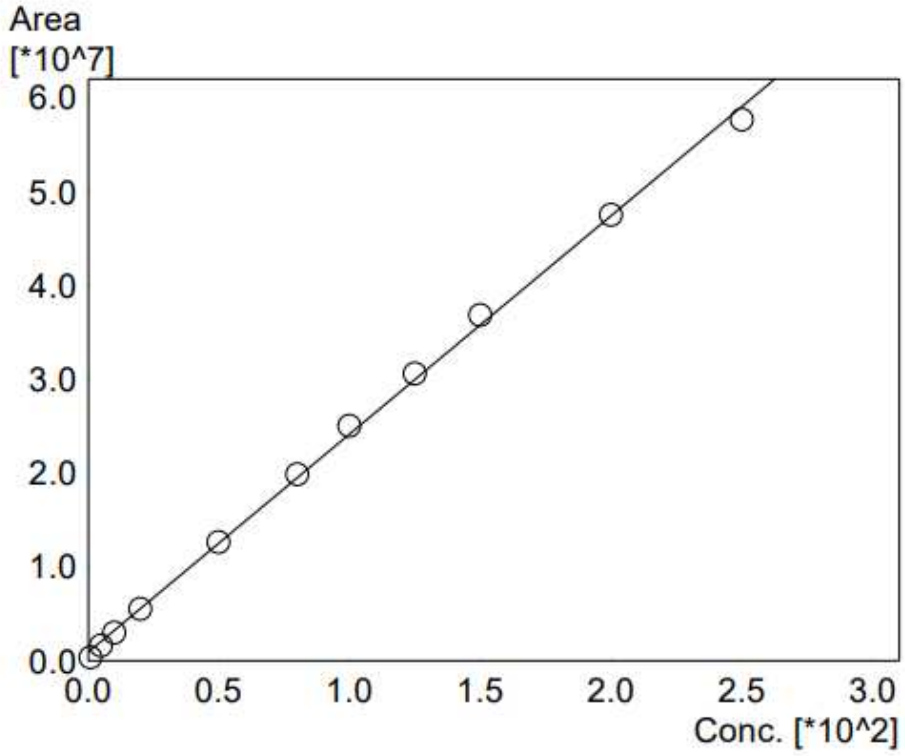


Figure S75. HPLC calibration curve of N-phenyl-7-azaindole. Solutions containing known concentrations of N-phenyl-7-azaindole were prepared in DCE at 1, 5, 10, 20, 50, 80, 100, 125, 150, 200, and 250 mM.

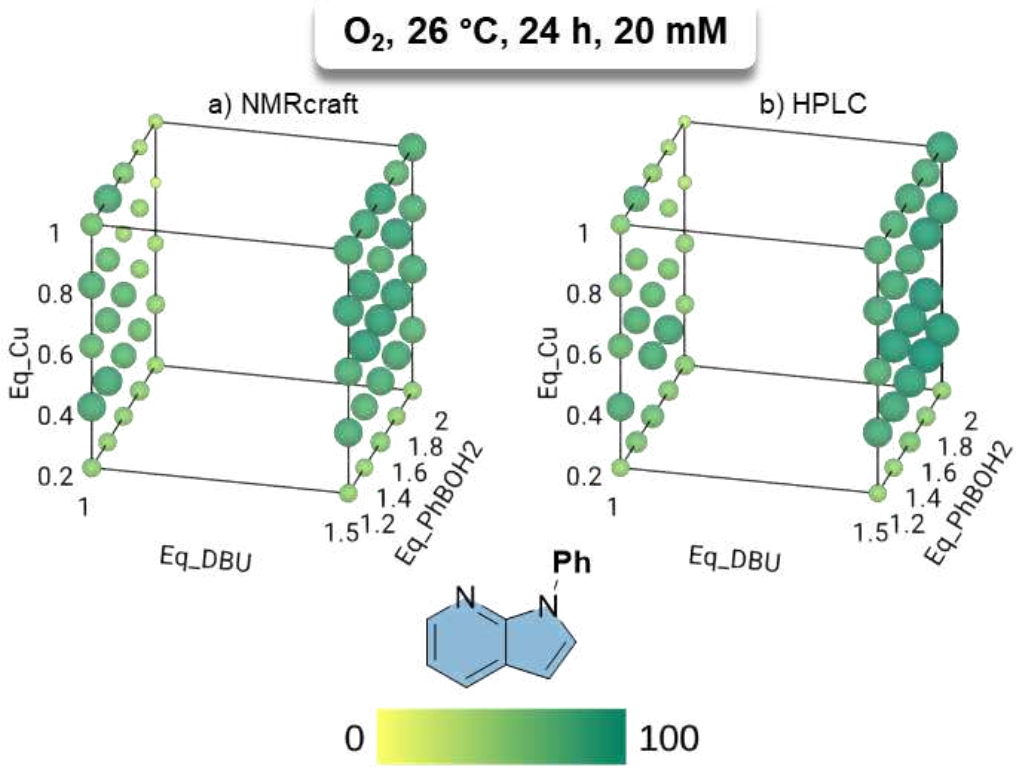


Figure S76. NMRanger vs HPLC hyperspaces. a) NMRanger yield map. **b)** HPLC yield map.

Table S7. All comparative yields determined by HPLC for N-Phenyl-7-azaindole, all performed under the conditions of 20 mM 7-azaindole; 26°C/O_2

Conditions: equiv_Cu(OAc)$_2$; equiv_PhB(OH)$_2$; equiv_DBU	**NMRanger Yield%**	**Comparative HPLC Yield %**	**Absolute difference %**	**Relative Difference %**
0.4; 1.4; 1.0	72.53	99.54	27.02	27.14
1.0; 2.0; 1.0	30.54	24.74	5.80	-23.46
0.8; 1.8; 1.0	39.55	39.67	0.12	0.30
0.6; 1.4; 1.0	61.88	66.45	4.57	6.87
1.0; 1.8; 1.0	32.54	33.70	1.15	3.41
0.2; 1.2; 1.0	42.10	40.02	2.08	-5.20
0.4; 1.6; 1.0	61.59	69.68	8.09	11.61
0.8; 2.0; 1.5	66.61	78.64	12.03	15.30
0.6; 2.0; 1.0	35.10	41.86	6.76	16.15
1.0; 1.8; 1.5	60.59	69.58	9.00	12.93
0.6; 1.5; 1.5	77.95	85.21	7.26	8.52
0.4; 2.0; 1.0	41.20	41.41	0.21	0.51
0.2; 2.0; 1.5	40.56	40.36	0.19	-0.48
0.8; 1.6; 1.5	65.21	72.12	6.91	9.58
0.4; 1.6; 1.5	64.90	75.80	10.90	14.38

1.0; 1.6; 1.5	77.92	73.36	4.56	-6.22
0.2; 1.4; 1.5	38.98	40.61	1.64	4.03
0.2; 2.0; 1.0	37.12	45.44	8.32	18.30
0.8; 2.0; 1.0	19.44	28.37	8.93	31.48
0.8; 1.4; 1.0	57.82	51.66	6.16	-11.92
0.6; 1.2; 1.0	58.31	59.28	0.97	1.63
0.6; 1.6; 1.0	61.55	61.02	0.53	-0.87
1.0; 1.2; 1.5	71.48	67.84	3.65	-5.37
0.2; 1.8; 1.0	43.09	45.29	2.20	4.86
0.4; 1.6; 1.5	69.73	81.73	11.99	14.67
0.6; 1.4; 1.5	81.00	83.62	2.61	3.12
0.4; 2.0; 1.5	64.49	85.21	20.72	24.32
0.2; 1.2; 1.5	40.38	38.77	1.61	-4.15
0.4; 1.2; 1.5	69.82	73.21	3.39	4.63
0.2; 1.4; 1.0	40.51	49.42	8.91	18.03
0.8; 1.2; 1.5	76.91	71.37	5.54	-7.76
1.0; 1.4; 1.0	69.14	62.86	6.28	-9.99
1.0; 2.0; 1.5	68.56	77.00	8.43	10.95
0.4; 1.8; 1.0	54.58	73.41	18.84	25.66
0.4; 1.8; 1.5	72.19	89.09	16.90	18.97
Table S7. Continued.				
0.6; 1.6; 1.5	79.60	84.36	4.76	5.65
0.2; 1.8; 1.5	40.40	42.95	2.55	5.93
1.0; 1.4; 1.5	72.93	67.49	5.44	-8.06
0.6; 1.8; 1.0	40.86	49.17	8.32	16.91
0.6; 1.2; 1.5	69.30	67.74	1.57	-2.31
1.0; 1.6; 1.0	49.26	42.55	6.70	-15.75
0.8; 1.8; 1.5	77.75	82.17	4.42	5.38
0.8; 1.4; 1.5	77.52	72.97	4.55	-6.24
0.8; 1.2; 1.0	65.62	54.60	11.02	-20.19
0.2; 1.6; 1.5	45.11	48.33	3.22	6.66
0.2; 1.6; 1.0	41.03	54.45	13.42	24.65
0.4; 1.2; 1.0	70.82	69.18	1.64	-2.37
1.0; 1.2; 1.0	52.56	44.20	8.36	-18.92
	Mean Absolute Error	**6.7**	**R^2**	**0.766**

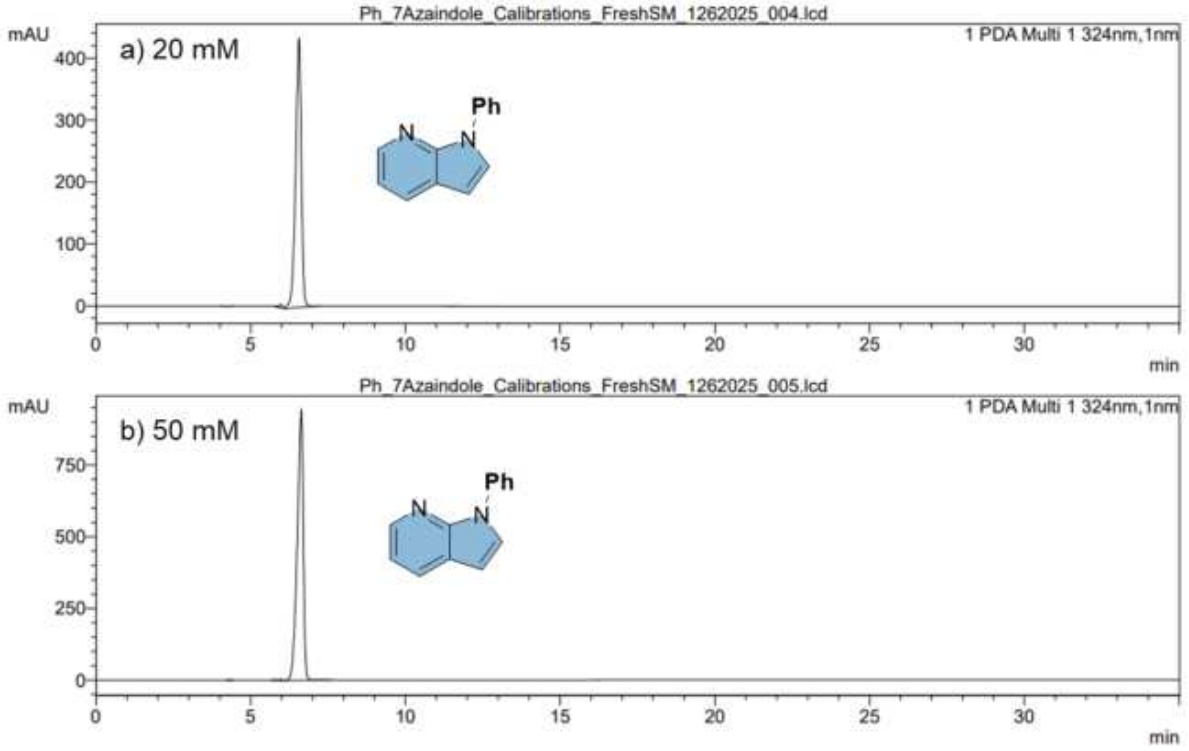


Figure S77. Representative pure N-Phenyl-7-azaindole HPLC chromatogram at 20 and 50 mM.

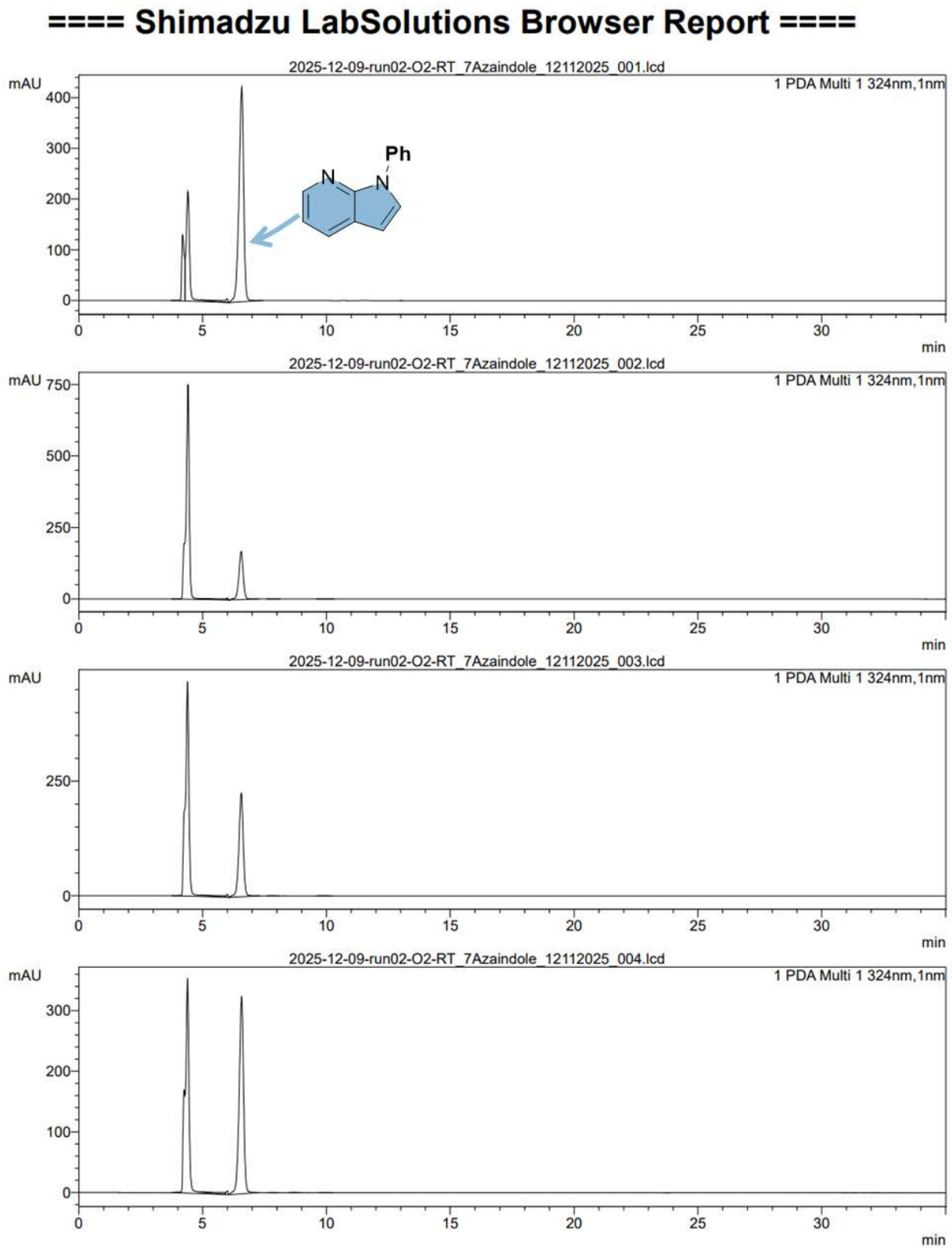


Figure S78. HPLC chromatograms.

1.11.3. 5-bromo7-azaindole validations

NMR yield determined by using an internal standard. Following General Method **A**, five reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, each with a final volume of 1100 µL. Reactions were performed with 20 mM 5-bromo-7-azaindole under O_2 at 26 °C for 24 h; and 0.8 equiv $Cu(OAc)_2$, 1.8 equiv $PhB(OH)_2$, and 1.5 equiv DBU. A 75 mM aqueous $Cu(OAc)_2$ stock solution was dispensed and evaporated at 40 °C for 4–8 h. After drying, 705.1 µL of DCE was added to each vial, followed by 220 µL of a 100 mM 5-bromo-7-azaindole stock solution, 158.4 µL of a 250 mM $PhB(OH)_2$ stock solution, and 16.5 µL of a 2000 mM DBU stock solution, all prepared in DCE. The mixtures were sparged with O_2, sealed, and incubated in a preheated ThermoMixer for 24 h.

After completion, a known amount of 1,3,5-trimethoxybenzene (TriOMePh) was added to each NMR tube. Yields estimated by NMRanger were compared with manual internal-standard integrations performed in Mnova, giving a mean absolute difference of 5.12 (**Table S8**). Some variation was observed between replicate vials, likely due to experimental variability rather than the yield-estimation method, since both quantification approaches gave similar average values. Small vial-to-vial differences in oxygen saturation may contribute to this variation, as oxygen mass transfer into the reaction solution can differ slightly between vials due to variations in gas flow. NMRanger gave a standard deviation of 1.7%, compared with 4.5% for manual internal-standard integration (**Table S8**) supporting reproducible yield quantification consistent with the internal-standard method.

Table S8. Comparison of N-phenyl-5-bromo-7-azaindole yields estimated by NMRanger and determined by manual internal-standard NMR integration using TriOMePh.

Condition	TriOMePh (mg)	NMRanger Yield %	Comparative Yield %	Absolute Difference %
Repeated five times under the same condition: 0.8 equiv $Cu(OAc)_2$; 1.8 equiv $PhB(OH)_2$; 1.5 equiv DBU	1.34	67.3	61.3	6.0
	1.11	66.9	71.8	4.9
	1.34	67.9	62.2	5.7
	1.33	70.6	64.3	6.3
	1.04	66.2	68.9	2.7
	Standard Deviation	**1.7**	**4.5**	
			Mean Absolute Error	**5.12**

1.11.4. 1H-indazole validations

Isolation. Following General Method A, 27 reaction mixtures were prepared in 2 mL vials by automated liquid handling using Robowski, each with a final volume of 1100 µL. Reactions were performed with 20 mM 1H-indazole, 2.0 equiv $Cu(OAc)_2$, 1.6 equiv $PhB(OH)_2$, and 2.0 equiv pyridine under O_2 at 26 °C for 24 h. A 75 mM aqueous $Cu(OAc)_2$ stock solution was dispensed and evaporated at 40 °C for 4–8 h. After drying, 717.2 µL DCE, 220 µL of 100 mM 1H-indazole, 140.8 µL of 250 mM $PhB(OH)_2$, and 22 µL of 2000 mM pyridine, all in DCE, were added. The mixtures were sparged with O_2, sealed, and incubated in a preheated ThermoMixer for 24 h.

After completion, all 27 reactions were combined, concentrated, passed through a short silica pad to remove residual copper salts, washed with 1,2-DCE, and evaporated. The crude mixture was dissolved in HPLC-grade $CHCl_3$ and filtered through a 0.2 µm PTFE syringe filter before GPC purification.

Each vial contained 0.022 mmol of 1H-indazole, corresponding to 0.594 mmol in total. A 1-phenyl-1H-indazole/2-phenyl-2H-indazole (N1/N2) mixture was isolated as 104.6 mg (0.538 mmol, 90.6% combined yield). NMRanger estimated a combined yield of 87.2%, in close agreement with the isolated value. Reproducibility was assessed using six randomly selected samples analyzed by Robowski NMR; the NMRanger-estimated N1 and N2 yields showed standard deviations of 3.0% and 1.7%, respectively. The mean absolute difference from the isolated yield was 3.9%, supporting reproducible quantification consistent with the ground-truth value (**Table S9**).

Table S9. Reproducibility of NMRanger quantification for 1-phenyl-1H-indazole (N1) and 2-phenyl-2H-indazole (N2) across six replicate reactions, compared with the combined isolated yield. Standard deviations for the N1 and N2 yields and the mean absolute error relative to the isolated yield are shown.

Condition	N1	N2	Total % (N1+N2)	Isolated Yield %	Absolute Difference %
Repeated five times under the same condition: 2.0 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 2.0 equiv pyridine	66.1	22.3	88.4	90.6	2.2
	68.9	20.1	89.0	90.6	1.6
	62.3	23.3	85.6	90.6	5.0
	61.8	23.1	84.8	90.6	5.8
	63.9	21.1	85.0	90.6	5.6
	68.1	19.1	87.2	90.6	3.4
Standard Deviation	**3.0**	**1.7**	**Mean Absolute Error**		**3.9**

1.11.5. Accuracy analysis of whole validations

Validation against independent ground-truth measurements. NMRanger yield estimates were evaluated against 65 independent measurements obtained using isolated yields, internal-standard NMR, HPLC quantification, and replicate experiments. Across the full dataset, NMRanger showed good agreement with the corresponding ground-truth values, with R^2=0.801 and a mean absolute error of 6.38% as discussed in the **main text, Section 6**. The complete comparison is provided in **Table S10**.

Table S10. Complete validation dataset comparing NMRanger-estimated yields with ground-truth values obtained by isolated yield, internal-standard NMR, HPLC, or replicate analysis across 65 measurements.

Conditions: equiv $Cu(OAc)_2$; equiv $PhB(OH)_2$; equiv DBU	Product	NMRanger Yield %	Comparative Yield %	Absolute difference %
Morpholine; 20 mM; 40°C/O_2; 0.9 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$	N-phenylmorpholine	56.10	51.30	4.80
	Unconsumed $PhB(OH)_2$	28.90	32.00	3.10
7-azaindole; 20 mM; 26°C/O_2; 1.0 equiv $Cu(OAc)_2$; 2.0 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	68.60	63.30	5.30
20 mM 1H-indazole; under 26°C/O_2; 2.0 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 2.0 equiv pyridine	1-phenyl-1H-indazole:2-phenyl-2H-indazole (N1/N2)	90.60	87.20	3.40
7-azaindole; 20 mM; 26°C/O_2; 0.6 equiv $Cu(OAc)_2$; 1.4 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	81.10	85.40	4.30

Table S10. Continued.

7-azaindole; 20 mM; under 26°C/O_2; 1.0 equiv $Cu(OAc)_2$; 1.2equiv $PhB(OH)_2$; 2 equiv DBU	N-Phenyl-7-azaindole	67.50	60.00	7.50
7-azaindole; 20 mM; under 26°C/O_2; 1.0equiv $Cu(OAc)_2$; 1.2 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	75.80	87.40	11.60
7-azaindole; 20 mM; 26°C/O_2; 0.2 equiv $Cu(OAc)_2$; 1.2 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	31.30	34.40	3.10
7-azaindole; 20 mM; under 26°C/O_2; 0.2 equiv $Cu(OAc)_2$; 1.2 equiv $PhB(OH)_2$; 2.0 equiv DBU	N-Phenyl-7-azaindole	25.80	31.50	5.70
7-azaindole; 20 mM; under 26°C/O_2; 0.4 equiv $Cu(OAc)_2$; 1.6 equiv $PhB(OH)_2$; 2.0 equiv DBU	N-Phenyl-7-azaindole	57.30	63.70	6.40
7-azaindole; 20 mM; under 26°C/O_2; 0.4 equiv $Cu(OAc)_2$; 1.8 equiv $PhB(OH)_2$; 2.0 equiv DBU	N-Phenyl-7-azaindole	60.80	66.50	5.70
20 mM 7-azaindole; under 26°C/O_2; 0.4 equiv $Cu(OAc)_2$; 1.4 equiv $PhB(OH)_2$; 1.5 equiv DBU	N-Phenyl-7-azaindole	84.30	92.50	8.20
Repeated five times under the same condition: 5-bromo-7-azaindole; 20 mM; 26°C/O2; 0.8 equiv Cu(OAc)2; 1.8 equiv PhB(OH)2; 1.5 equiv DBU	N-Phenyl-5-bromo-7-azaindole	67.30	61.30	6.00
		66.90	71.80	4.90
		67.90	62.20	5.70
		70.60	64.30	6.30
		66.20	68.90	2.70

Table S10. Continued. All reactions below here are performed under the same general of 20 mM 7-azaindole substrae; 26°C/O_2.

0.4; 1.4; 1.0	N-Phenyl-7-azaindole	72.53	99.54	27.02
1.0; 2.0; 1.0	N-Phenyl-7-azaindole	30.54	24.74	5.80
0.8; 1.8; 1.0	N-Phenyl-7-azaindole	39.55	39.67	0.12
0.6; 1.4; 1.0	N-Phenyl-7-azaindole	61.88	66.45	4.57
1.0; 1.8; 1.0	N-Phenyl-7-azaindole	32.54	33.70	1.15
0.2; 1.2; 1.0	N-Phenyl-7-azaindole	42.10	40.02	2.08
0.4; 1.6; 1.0	N-Phenyl-7-azaindole	61.59	69.68	8.09
0.8; 2.0; 1.5	N-Phenyl-7-azaindole	66.61	78.64	12.03
0.6; 2.0; 1.0	N-Phenyl-7-azaindole	35.10	41.86	6.76
1.0; 1.8; 1.5	N-Phenyl-7-azaindole	60.59	69.58	9.00
0.6; 1.5; 1.5	N-Phenyl-7-azaindole	77.95	85.21	7.26
0.4; 2.0; 1.0	N-Phenyl-7-azaindole	41.20	41.41	0.21
0.2; 2.0; 1.5	N-Phenyl-7-azaindole	40.56	40.36	0.19
0.8; 1.6; 1.5	N-Phenyl-7-azaindole	65.21	72.12	6.91
0.4; 1.6; 1.5	N-Phenyl-7-azaindole	64.90	75.80	10.90
1.0; 1.6; 1.5	N-Phenyl-7-azaindole	77.92	73.36	4.56
0.2; 1.4; 1.5	N-Phenyl-7-azaindole	38.98	40.61	1.64
0.2; 2.0; 1.0	N-Phenyl-7-azaindole	37.12	45.44	8.32
0.8; 2.0; 1.0	N-Phenyl-7-azaindole	19.44	28.37	8.93
0.8; 1.4; 1.0	N-Phenyl-7-azaindole	57.82	51.66	6.16
0.6; 1.2; 1.0	N-Phenyl-7-azaindole	58.31	59.28	0.97
0.6; 1.6; 1.0	N-Phenyl-7-azaindole	61.55	61.02	0.53
1.0; 1.2; 1.5	N-Phenyl-7-azaindole	71.48	67.84	3.65
0.2; 1.8; 1.0	N-Phenyl-7-azaindole	43.09	45.29	2.20
0.4; 1.6; 1.5	N-Phenyl-7-azaindole	69.73	81.73	11.99
0.6; 1.4; 1.5	N-Phenyl-7-azaindole	81.00	83.62	2.61
0.4; 2.0; 1.5	N-Phenyl-7-azaindole	64.49	85.21	20.72
0.2; 1.2; 1.5	N-Phenyl-7-azaindole	40.38	38.77	1.61
0.4; 1.2; 1.5	N-Phenyl-7-azaindole	69.82	73.21	3.39
0.2; 1.4; 1.0	N-Phenyl-7-azaindole	40.51	49.42	8.91
0.8; 1.2; 1.5	N-Phenyl-7-azaindole	76.91	71.37	5.54
1.0; 1.4; 1.0	N-Phenyl-7-azaindole	69.14	62.86	6.28
1.0; 2.0; 1.5	N-Phenyl-7-azaindole	68.56	77.00	8.43
0.4; 1.8; 1.0	N-Phenyl-7-azaindole	54.58	73.41	18.84
0.4; 1.8; 1.5	N-Phenyl-7-azaindole	72.19	89.09	16.90
0.6; 1.6; 1.5	N-Phenyl-7-azaindole	79.60	84.36	4.76

Table S10. Continued. All reactions below here are performed under the same general of 20 mM 7-azaindole; 26°C/O_2.

0.2; 1.8; 1.5	N-Phenyl-7-azaindole	40.40	42.95	2.55
1.0; 1.4; 1.5	N-Phenyl-7-azaindole	72.93	67.49	5.44
0.6; 1.8; 1.0	N-Phenyl-7-azaindole	40.86	49.17	8.32
0.6; 1.2; 1.5	N-Phenyl-7-azaindole	69.30	67.74	1.57
1.0; 1.6; 1.0	N-Phenyl-7-azaindole	49.26	42.55	6.70
0.8; 1.8; 1.5	N-Phenyl-7-azaindole	77.75	82.17	4.42
0.8; 1.4; 1.5	N-Phenyl-7-azaindole	77.52	72.97	4.55
0.8; 1.2; 1.0	N-Phenyl-7-azaindole	65.62	54.60	11.02
0.2; 1.6; 1.5	N-Phenyl-7-azaindole	45.11	48.33	3.22
0.2; 1.6; 1.0	N-Phenyl-7-azaindole	41.03	54.45	13.42
0.4; 1.2; 1.0	N-Phenyl-7-azaindole	70.82	69.18	1.64
1.0; 1.2; 1.0	N-Phenyl-7-azaindole	52.56	44.20	8.36
Results:	**Mean Absolute Error**	**6.38**	**R square**	**0.801**

1.12. Confirmation of chemical shift and product assignment by NMR titration

To confirm the identities of compounds whose resonances shifted in the crude reaction mixture, NMR titration experiments were performed for $PhB(OH)_2$ and phenol. As shown in **Figure S79**, increasing amounts of $PhB(OH)_2$ were added to a crude sample. The intensities of the shifted resonances increased proportionally without the appearance of new signals, confirming their assignment to $PhB(OH)_2$.

A similar response was observed for phenol. Upon addition of increasing amounts of phenol to the crude reaction mixture, the intensity of the assigned resonance increased without the appearance of new peaks, confirming its identity and the corresponding chemical-shift assignment (**Figure S80**).

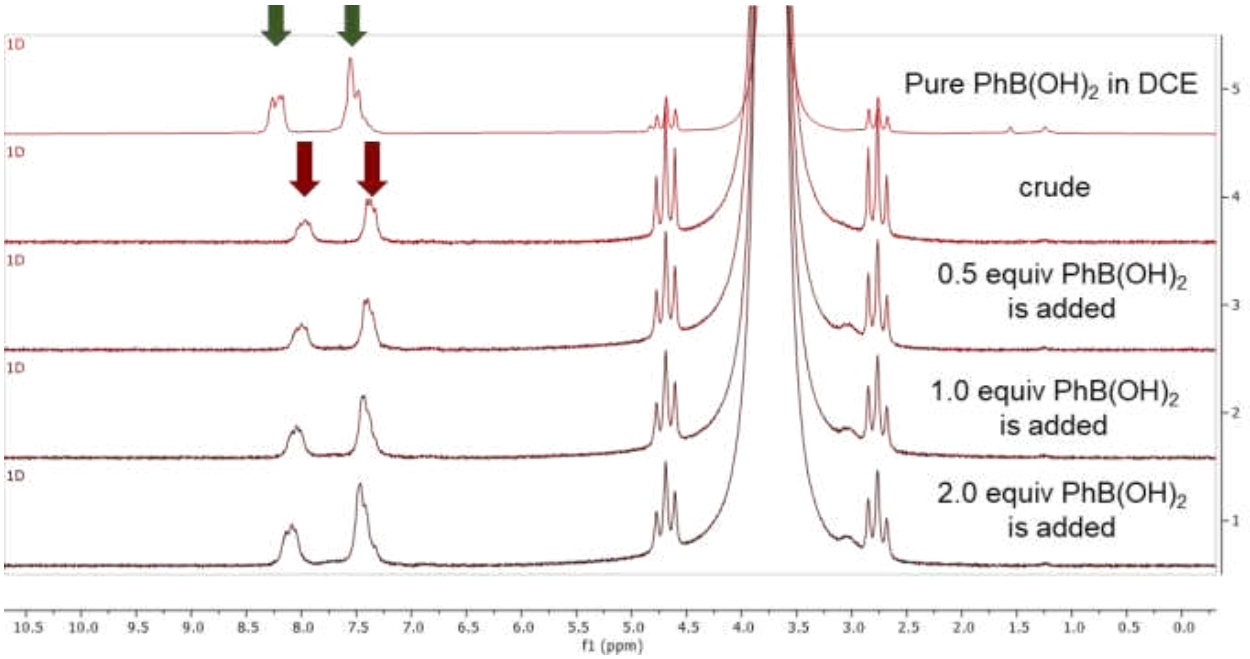


Figure S79. NMR titration of a crude reaction mixture with $PhB(OH)_2$. The spectrum of pure $PhB(OH)_2$ in DCE is shown at the top, with the corresponding aromatic resonances indicated by green arrows. In the crude mixture, these signals appear shifted, as marked by the red arrows. Successive addition of 0.5, 1.0, and 2.0 equiv $PhB(OH)_2$ increased the intensity of the shifted resonances without generating new peaks, confirming their assignment to $PhB(OH)_2$.

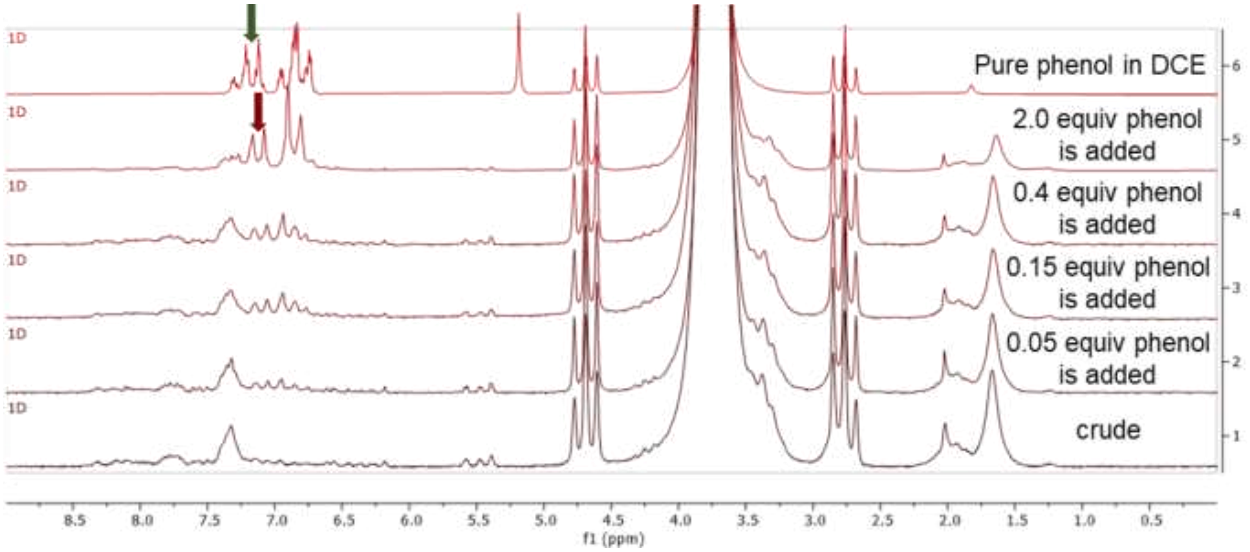


Figure S80. NMR titration of a crude reaction mixture with phenol. The spectrum of pure phenol in DCE is shown at the top, with the aromatic phenol multiplet indicated by the green arrow. In the crude mixture, this multiplet appears shifted, as marked by the red arrow. Successive addition of 0.05, 0.15, 0.4, and 2.0 equiv phenol progressively increased the intensity of the shifted multiplet without generating new peaks, confirming its assignment to phenol.

1.13. Determination of observable protons after copper complexation

In the particular case of Chan-Lam reaction, Phenyl morpholine showed only moderate distortion, consistent with negligible coordination effects. In contrast, the heterocyclic starting materials showed substantially greater signal perturbation upon exposure to copper. When the loss of an individual resonance could be identified directly, such as disappearance of a singlet in the 5-bromo-7-azaindole spectrum, the observable-proton count N_{i,r_n} was adjusted accordingly.

For more strongly distorted spectra, including those of 1H-indazole and 7-azaindole, empirical N_{i,r_n} values were estimated using **expression (5) in the main text** and an internal standard from spectra in which both C_0 and $C_{i,n}$ were known. Because these values are derived from broadened and partially suppressed signals, they are inherently less precise. The corresponding starting material–copper spectra were then used as model references, with proton counts assigned according to the experimentally observable resonances.

For the arylated indazole and azaindole products, no loss of detectable proton signals attributable to copper coordination was observed. Any moderate line broadening was accommodated by the fitted peak-width parameter, allowing product quantification using the default NMRanger settings. Related NMR spectra for the observable proton assessments are shown in **Figure S81**-**Figure S83**.

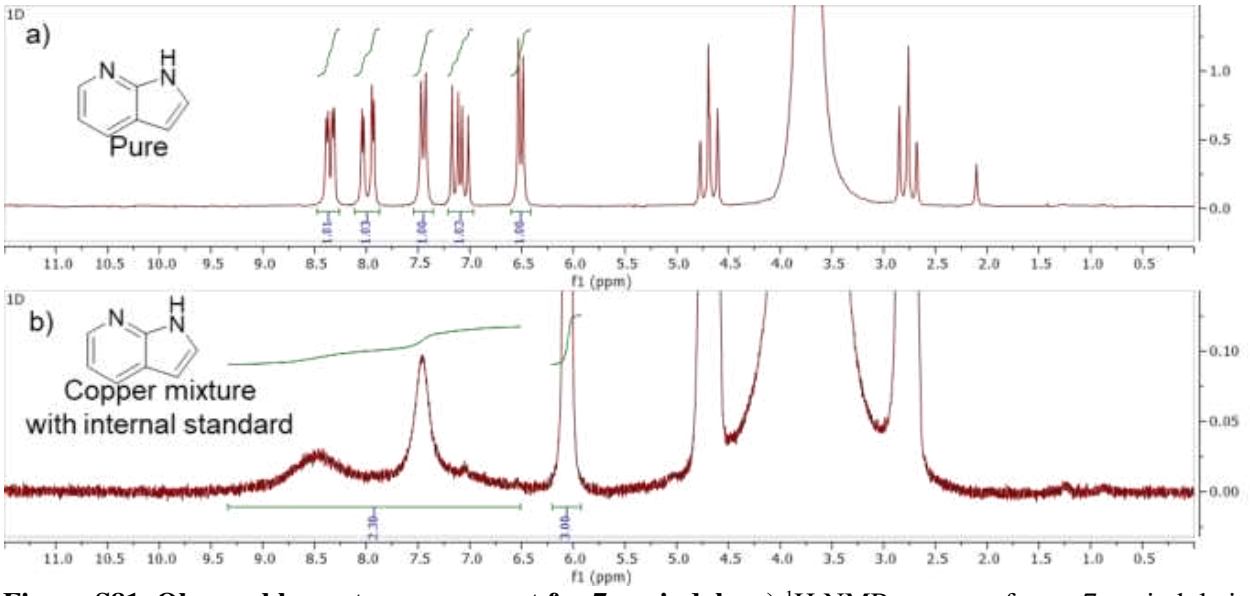


Figure S81. Observable-proton assessment for 7-azaindole. a) ^{1}H NMR spectra of pure 7-azaindole in DCE and the corresponding copper-containing sample measured with an internal standard. **b)** Integration of the remaining 7-azaindole resonances, indicating an effective observable proton count of 2.3 after copper complexation.

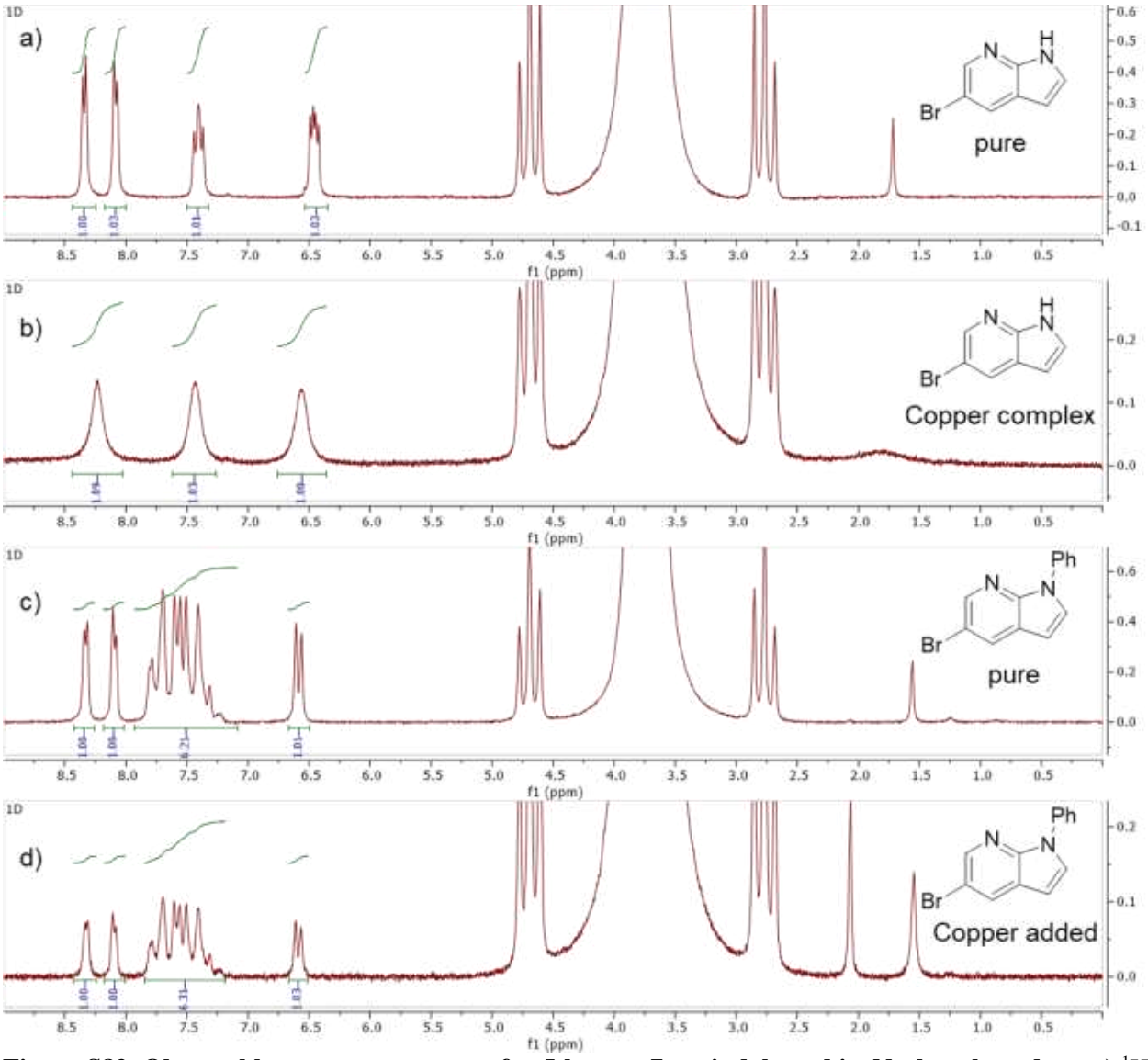


Figure S82. Observable-proton assessment for 5-bromo-7-azaindole and its N-phenyl product. a) ^{1}H NMR spectrum of pure 5-bromo-7-azaindole in DCE. **b)** After copper complexation, the singlet near 8.1 ppm disappears, corresponding to the loss of one observable proton, while the remaining resonances are broadened. **c)** ^{1}H NMR spectrum of pure N-phenyl-5-bromo-7-azaindole. **d)** After copper addition, the product retains all characteristic resonances, with only moderate broadening and no loss of observable protons.

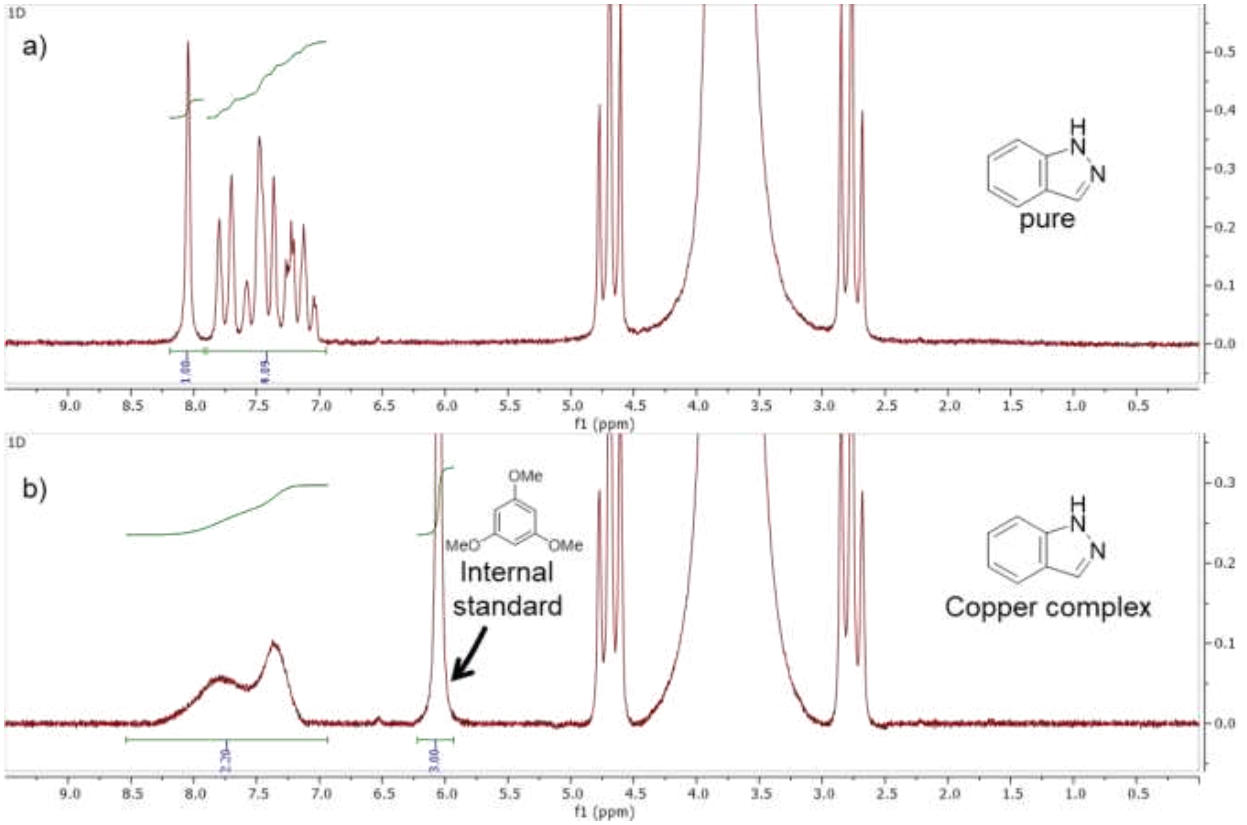
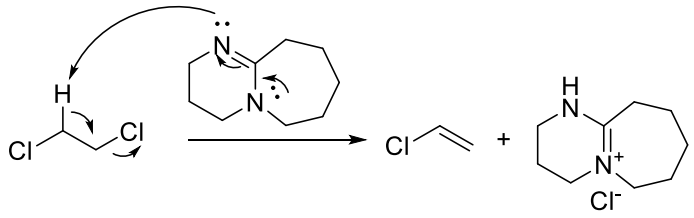


Figure S83. Observable-proton assessment for 1H-indazole. a) ^{1}H NMR spectrum of pure 1H-indazole in DCE. **b)** After copper complexation, substantial broadening and signal suppression are observed in the aromatic region. Using the internal-standard signal near 6.1 ppm, the remaining indazole resonances were quantified, giving an effective observable proton count of approximately 2.3.

1.14. DBU reference for azaindoles

DBU was used as the base for reactions involving 7-azaindole and 5-bromo-7-azaindole substrates. The proton signals of DBU are expected either near 10.0 ppm or below 4.0 ppm, outside the spectral regions used by NMRanger for quantitation, which are located between 5.7 and 9.5 ppm.

However, during acquisition of the DBU reference spectrum in DCE, an additional group of signals was observed between 5.0 and 6.5 ppm. This group included a doublet at 5.38 ppm with an apparent coupling constant of 0.9 Hz, a doublet of doublets at 5.51 ppm with apparent coupling constants of 8.01 and 1.49 Hz, and a more complex multiplet at 6.31 ppm (see **Figure S84a**). These signals were assigned to vinyl chloride, formed by DBU-induced elimination of DCE, as shown in **Scheme S1**.



Scheme S1. Base induced elimination of Dichloroethane

Vinyl chloride has previously been reported by ^{1}H NMR in DCE, with signals at 5.43, 5.50, and 6.30 ppm,[6] in good agreement with the signals observed here at 5.38, 5.51, and 6.31 ppm, respectively. In addition, the coupling constants of vinyl chloride are known to depend on both solvent polarity and concentration.[7,8] Therefore, small differences between the reported and observed coupling patterns are expected.

To further support this assignment, an ABC spin system was simulated using the NMRdb simulator (see **Figure S84b**), can be reproduced at https://app.nmrium.com/simulation)[9], based on reported coupling constants.[7] for vinyl chloride (see **Figure S84c**). The simulated 80 MHz spectrum reproduced the main features of the experimental multiplet pattern, matching closely the experimental observation.

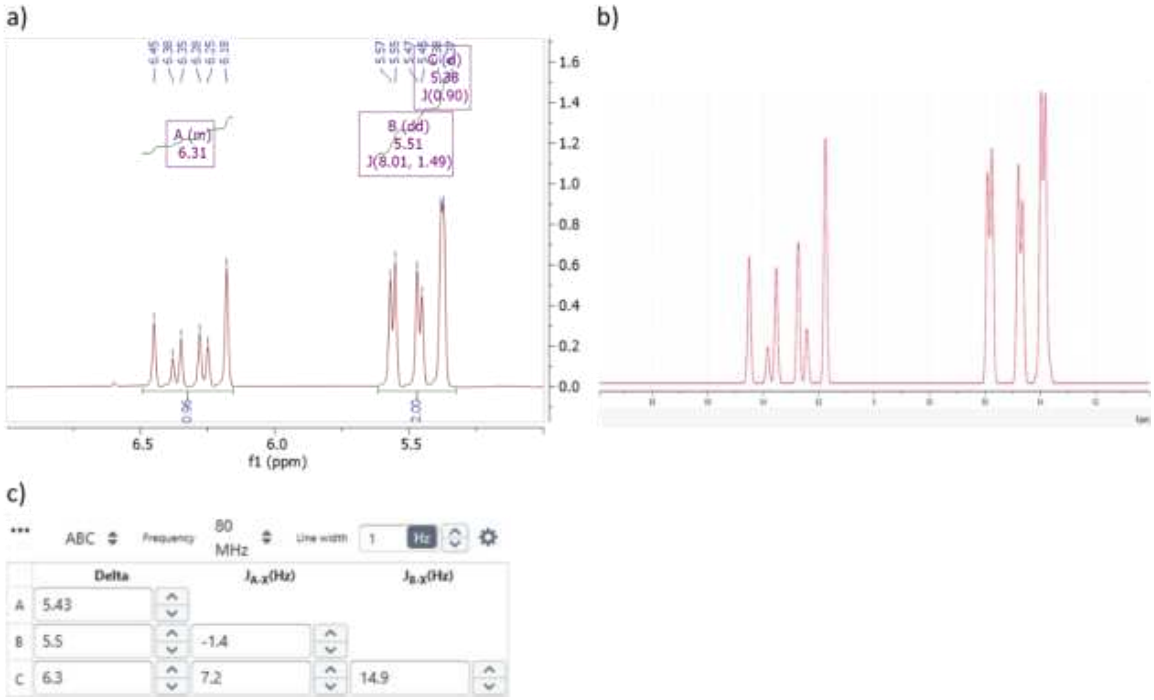


Figure S84. Experimental ^{1}H NMR signals attributed to vinyl chloride and comparison with simulated spectra. a) Experimental spectrum acquired after dissolving DBU in DCE, shown in the 5.0–7.0 ppm region. **b)** Simulated spectrum of vinyl chloride generated as an ABC spin system using literature-based chemical shifts and coupling constants. c) Chemical shifts and coupling constants used for the simulation.

To confirm that vinyl chloride was formed by reaction of DBU with DCE, and was not present as an initial contamination, DBU was dissolved in DCE and the mixture was monitored over time by ^{1}H NMR spectroscopy (**Figure S85**). The signals assigned to vinyl chloride gradually appeared in the spectrum, with multiplets visible around 6.3 and 5.5 ppm. After 60 min, even the weaker signal near 6.3 ppm was clearly observed. An additional peak was also observed to shift upfield over time; this signal was tentatively attributed to HCl generated during the elimination reaction.

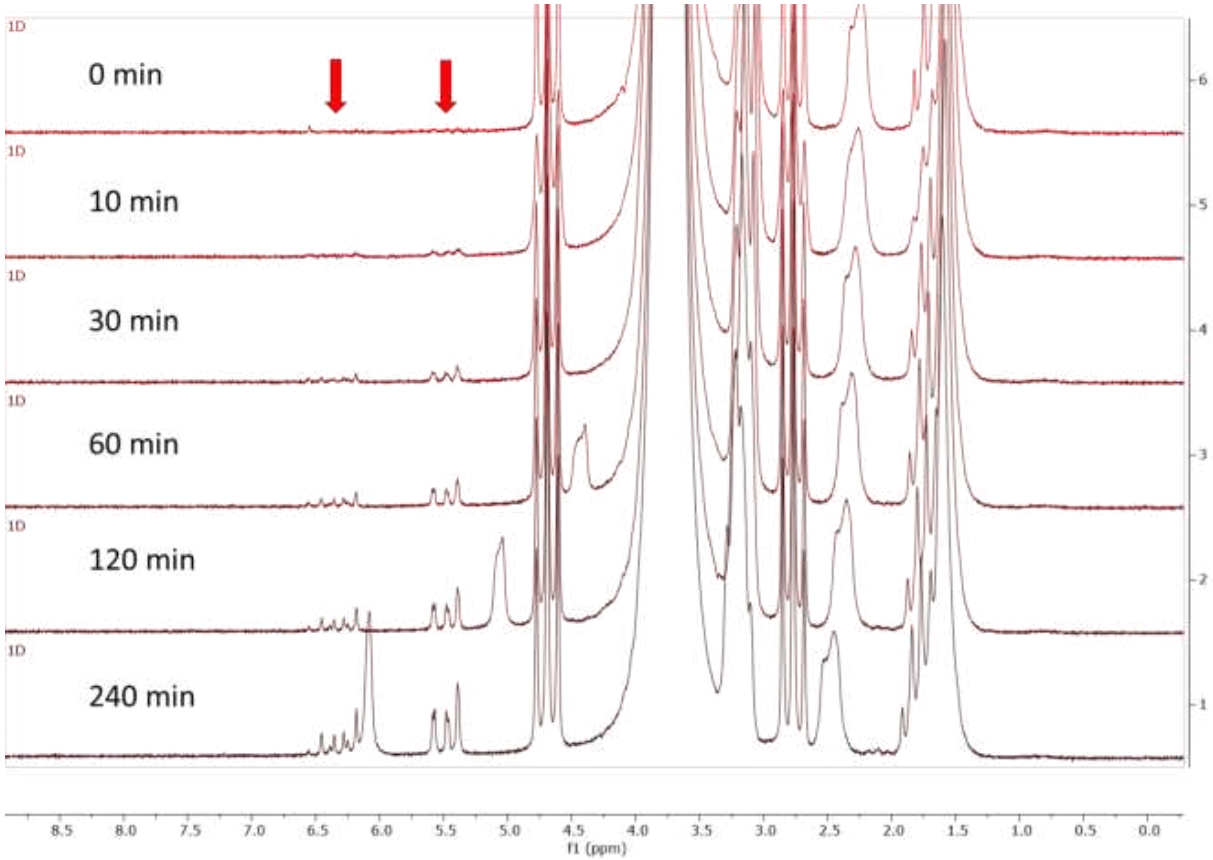


Figure S85. Stack of NMR spectrum of DBU in DCE over time. The red arrow points to the location of the vinyl chloride multiplets.

Because vinyl chloride has a boiling point of -13.4 °C, only the fraction dissolved in the NMR solvent is expected to be detected. The total amount generated during the reaction therefore cannot be quantified reliably, as vinyl chloride can partition between the solution and the gas phase. Nevertheless, the vinyl chloride signals observed in the DBU reference spectrum were included in the NMRanger model. This allowed the additional signals observed experimentally in the crude reaction mixtures to be modelled as a separate spectral component. As a result, the presence of this side reaction does not interfere with the quantitation of the target reaction components by NMRanger.

2. Other NMR software available

Although several software tools for NMR-based crude mixture analysis have been reported in the literature, benchmarking against most of them was not feasible, either because they were not readily accessible, could not be straightforwardly deployed, or were not directly applicable to our datasets.

Publication	Incompatibility causes
Rout, M.; Lipfert, M.; Lee, B. L.; Berjanskii, M.; Assempour, N.; Fresno, R. V.; Cayuela, A. S.; Dong, Y.; Johnson, M.; Shahin, H.; Gautam, V.; Sajed, T.; Oler, E.; Peters, H.; Mandal, R.; Wishart, D. S. MagMet: A Fully Automated Web Server for Targeted Nuclear Magnetic Resonance Metabolomics of Plasma and Serum. *Magn. Reson. Chem.* **2023**, *61* (12), 681–704. https://doi.org/10.1002/mrc.5371.	- Fixed reference metabolite list - Specific pulse sequence required - Partially deuterated samples only
Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of^{1} H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442. https://doi.org/10.1021/ac060209g.	- No testable code or platform available
Hao, J.; Astle, W.; De Iorio, M.; Ebbels, T. M. BATMAN—an R Package for the Automated Quantification of Metabolites from Nuclear Magnetic Resonance Spectra Using a Bayesian Model. *Bioinformatics* **2012**, *28* (15), 2088–2090.	None.
Astle, W.; De Iorio, M.; Richardson, S.; Stephens, D.; Ebbels, T. A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures. *J. Am. Stat. Assoc.* **2012**, *107* (500), 1259–1271. https://doi.org/10.1080/01621459.2012.695661.	- No testable code or platform available
Ravanbakhsh, S.; Liu, P.; Bjordahl, T. C.; Mandal, R.; Grant, J. R.; Wilson, M.; Eisner, R.; Sinelnikov, I.; Hu, X.; Luchinat, C. Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PloS One* **2015**, *10* (5), e0124219.	- Fixed reference metabolite list - Limited to 500/600 MHz - Specific pulse sequence required
Röhnisch, H. E.; Eriksson, J.; Müllner, E.; Agback, P.; Sandström, C.; Moazzami, A. A. AQuA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma. *Anal. Chem.* **2018**, *90* (3), 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324.	- Only pseudo-code available
Domżał, B.; Nawrocka, E. K.; Gołowicz, D.; Ciach, M. A.; Miasojedow, B.; Kazimierczuk, K.; Gambin, A. Magnetstein: An Open-Source Tool for Quantitative NMR Mixture Analysis Robust to Low Resolution, Distorted Lineshapes, and Peak Shifts. *Anal. Chem.* **2024**, *96* (1), 188–196. https://doi.org/10.1021/acs.analchem.3c03594.	None.

2.1. Benchmark against Magnetstein and BATMAN

We benchmarked NMRanger against Magnetstein and BATMAN using two crude reaction mixtures. Magnetstein analyses were performed through the web service (https://bioputer.mimuw.edu.pl). BATMAN analyses were performed using the same chemical-shift tolerance as NMRanger (0.08 ppm) and the same reference offset for phenylboronic acid. The input and output files, or screenshots where applicable, are available in a Zenodo repository (https://zenodo.org/records/21396796) as `Magnetstein_input_and_results.zip` and `BATMAN_input_and_results.zip`

(During the review process, the Reviewers are kindly asked to use the following link: https://zenodo.org/records/21396796?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjJjNTc0YzUxLWNhYTEtNGRiMi05YTY0LTU4OTRkMDg3YTU0ZCIsImRhdGEiOnt9LCJyYW5kb20iOiIzYTRjMTVkMzYyNmIwNGIyOWVkNzlhZWQzY2UwZGFhOCJ9.yHPZ_xSGzlCMNgyGDfqzCBDP5BF0OfTWsb8kmoxT-AGI-sL5ynB08Ew4bDVzm83a7k53bdff76cFrv8D1grTWQ).

BATMAN calculations were performed in triplicate using different random seeds to assess convergence of the fit. The reported BATMAN results correspond to the average of these three runs.

Ground-truth compositions were obtained independently for two substrates: *N*-phenyl morpholine and *N*-phenyl 7-azaindole.

For phenyl morpholine, NMRanger was evaluated over a single spectral window, from 6.5 to 8.5 ppm. For phenyl azaindole, two spectral windows were used: 5.7–7.0 ppm and 5.7–9.5 ppm. Magnetstein was run with its default parameters: $\kappa_{mixture} = 0.25$ and $\kappa_{components} = 0.22$.

Because NMRanger allows selected spectral regions to be excluded and can process spectra acquired in non-deuterated solvents, each raw dataset was analyzed twice with Magnetstein. In the first analysis, the spectra were used without modification. In the second, the solvent region was masked by setting it to zero, providing a fairer comparison between the two approaches.

Magnetstein reports the relative proportions of the reference compounds rather than direct quantities. In addition, some quantified products could not be isolated because of their volatility.

Therefore, the figure of merit was defined as the ratio between two analytes for which the ground truth was known.

The first case study involved the quantification of phenylboronic acid ($PhB(OH)_2$) and phenyl morpholine (PhMorph) in a crude reaction mixture. The ground truth for this mixture was determined by isolation after gel permeation chromatography (GPC). The second case study involved the quantification of phenyl azaindole (PhAza) using a known amount of 1,3,5-trimethoxyphenyl (TriOMePh) as an internal standard.

Algorithm	PhMorph/ $PhB(OH)_2$ (Calculation time)	TriOMePh/PhAza (Calculation time)
NMRanger	1.656 (30 sec)	0.770 (1min 34 sec)
Magnetstein without solvent mask	No convergence after 18h.	0.262 (2h53min)
Magnetstein with solvent mask	13.275 (3h5min)	0 (2h42min)
BATMAN	1.354 (4min 29s)	4.843 (28min 35s)
Ground Truth	1.775	0.757

In both case studies, Magnetstein showed the lowest performance. This may reflect the extent of chemical-shift variation and line broadening in crude reaction mixtures, which may be difficult to capture using a generalized parametrization. We cannot exclude the possibility that alternative parameter values, namely $\kappa_{mixture}$ and $\kappa_{components}$, could improve Magnetstein performance. However, the computational cost of each analysis made systematic tuning of these parameters prohibitively time-consuming. BATMAN performed only slightly worse than NMRanger for the simpler phenyl morpholine spectrum, although with slower convergence and less accurate quantification. In contrast, its performance deteriorated markedly for the more complex phenyl azaindole crude mixture, indicating reduced robustness as spectral complexity increased.

3. NMRanger's Manual

The following section is a user manual focused on the use of NMRanger's to quantify components within reaction mixtures. First, technical requirements are given to run the algorithm, then the available parameters to tune are detailed for the more proficient users. Finally, an illustrated "step-by-step" tutorial is provided so that even users with little to none prior experience with code can run the demonstration and then adapt it to their use cases.

3.1. Requirement

All data fitting and graphical displays were performed using Python 3.12.12. Required packages are: NumPy 1.26.4, SciPy 1.13.1, Scikit-learn 1.5.1, Pandas 2.2.2, tqdm 4.66.5, matplotlib 3.9.2, Pybaselines 1.2.1.

NMRanger uses CSV files as input for NMR analysis. In this work, raw FID files were converted to CSV format using the MestReNova script published in ref.[10] A copy of this script is provided with the Supplementary Information. Details of the preprocessing procedure are described in ref.[10]

3.2. Parameters

NMRanger parameters are divided into two categories: **general parameters** and **substrate-specific parameters**. General parameters are defined in the `Default parameters` section and are applied by default to all calculations. Substrate-specific parameters are defined in the `Substrate definition` section and apply only to a given crude mixture type, for example a specific substrate or substrate combination. Any default parameter can be copied into a substrate definition and edited to override the global value.

3.2.1. General Parameters

Parameter	Default value	Description
`debug`	False	Enables interactive display after each fit. This displays the fitting results for each reference and analyzed sample, as well as the calibration-curve generation. This option should only be used with

		MAX_WORKER = 1, as interactive display with parallel processing can lead to unexpected behavior.
use_model_multiplicity	True	Clusters signals into multiplets for fitting. Peaks belonging to the same multiplet are constrained to move together, reducing the number of fitted parameters. If set to False, each peak is treated independently, which may decrease fit stability and increase calculation time.
transfer_results_across_ranges	True	Uses the results from one fitted spectral window to constrain the estimated abundance of the same analytes in subsequent spectral regions. If set to False, each spectral region is analyzed independently.
tie_global_gl	False	Uses a common Gaussian/Lorentzian mixing weight for all peaks of a given reference. This can accelerate calculations for references with many peaks, but may reduce fit quality if different peaks have different lineshape characteristics.
width_bounds	None	Optional bounds for the pseudo-Voigt width parameter, either None or (min, max). The lower bound should be greater than zero. This parameter should not be interpreted directly as a linewidth in ppm because its effect depends on the pseudo-Voigt parametrization.
detection_threshold	None	Sensitivity threshold used for peak detection during model generation. If expected peaks are missing, a value of 0.5 can be used as a starting point and lowered until all relevant peaks are identified.
additional_model	None	Allows additional reference models to be generated directly from predefined parameter sets instead of being loaded from reference NMR spectra. See the substrate Example in the script for examples.
multiplet_distance_threshold	None	Maximum distance, in ppm, between two peaks for them to be assigned to the same multiplet. If set to None, this value

		is calculated as `coupling_constant_threshold / NMR_frequency`.
`coupling_constant_threshold`	20	Maximum distance, in Hz, between two peaks for them to be assigned to the same multiplet when multiplet_distance_threshold = None.
`special_treatment`	None	Allows manual correction of the model generated from reference spectra. For example, peaks can be moved, added, or removed. See the substrate Example in the script for examples.
`custom_baseline`	None	Specifies the baseline-correction method. Currently supported values are None, default, strong_spline, weak_spline, exponential, and linear. Additional baseline methods can be added in `make_baseline_correction)`.
`calibration_concentrations`	[]	Concentrations of the reference solutions used to generate the external calibration curve. This list can be left empty if calibration spectra are provided directly.
`calibration_integrations`	[]	Corresponding peak integrals for the external calibration curve, for example when these values were measured manually. If provided, NMRanger skips automatic calibration-curve generation and uses these values directly. When used together with `calibration_concentrations`, both lists must have matching order and length.
`filter_list`	[[]]	References to exclude from fitting. Each sublist corresponds to one analyzed spectral region. References listed in a given sublist are not used for fitting that region.
`sample_to_check`	[]	Restricts the analysis to selected samples. If sample indices are provided, only the corresponding samples are analyzed. Indexing is zero-based.
`ftol`	1e-12	Termination criterion passed to `curve_fit` for the fitting procedure. Higher values result in earlier

		termination and faster calculations, but may produce looser fits.
xtol	1e-12	Termination criterion passed to curve_fit for the fitting procedure. Higher values result in earlier termination and faster calculations, but may produce looser fits.
gtol	1e-12	Termination criterion passed to curve_fit for the fitting procedure. Higher values result in earlier termination and faster calculations, but may produce looser fits.
tolerance_across_ranges	0.05	Allowed deviation in analyte abundance between spectral regions when transfer_results_across_ranges = True. The default value corresponds to 5%. This parameter has no effect when transfer_results_across_ranges = False.
tolerance_ppm	0.08	Maximum tolerated chemical-shift deviation for fitted peaks, in ppm.
verbose	2	Controls the amount of information displayed during fitting. 0 displays no information, 1 displays only termination information, and 2 displays progress information during the analysis.
smoothing	5	Gaussian smoothing applied during preprocessing. This value is passed as the sigma parameter to gaussian_filter from scipy.ndimage.
solver_type	Local	Selects the fitting strategy. Local uses curve_fit with the trust-region reflective algorithm. Global uses differential evolution as the optimization method.
MAX_WORKER	1	Number of processors used for the analysis. Each sample is handled by one processor. When debug = True, this value should be kept at 1.
NMR_frequency	80	Spectrometer frequency, in MHz, for the observed nucleus.

3.2.2. Substrate-specific Parameters

Substrate-specific parameters define the input files, spectral regions, reference compounds, calibration data, and proton-normalization factors used for a given crude mixture. These parameters are defined in the `Substrate definition` section of the script and override the default parameters only for the corresponding substrate or substrate combination.

For each substrate definition, the user provides the paths to the reference spectra, calibration spectra, and samples to analyze. The user also specifies the spectral windows used for fitting, the labels assigned to each reference compound, and the number of observable protons represented by each reference in the selected regions. These proton counts are used to normalize the fitted peak areas before conversion into concentration.

Parameter	Example	value Description
`path_to_Analyte`	r"C\User\References\CSV_Analyte.csv"	Path to the CSV file containing the NMR spectrum of the reference analyte. The CSV file should contain chemical shift in the first column and intensity in the second column.
`path_calibration_intensity`	[r"C\User\References\CSV_Analyte_at_10 mM", r"C\User\References\CSV_Analyte_at_20 mM "]	List of paths to CSV files used to generate the external calibration curve. Each file corresponds to a reference analyte measured at a known concentration. This list can be left empty if calibration integrals are provided manually in `calibration_integrations`.
`path_to_test`	r"C\User\Crudes\one_crude_sample.csv"	Path to a single crude-mixture CSV file used for testing or checking the substrate definition.
`master_path`	r" C\User\Crudes "	Path to the folder containing the individual sample folders to be analyzed.
`peaks_info`	[[1, 5], [2,10]]	List of spectral regions to analyze, in ppm. Each region is defined as [lower_bound, upper_bound]. Regions are analyzed in the order provided. The first region should contain observable signals from the reference product used to generate the calibration curve.
`calibration_concentrations`	[1, 2]	List of concentrations, in mM, used to generate the external calibration curve.

		The order must match either the spectra listed in `path_calibration_intensity` or the manually provided values in `calibration_integrations`.
`calibration_integrations`	[5,10]	List of manually provided integrals used to generate the calibration curve. If this list is provided, `path_calibration_intensity` is not used for calibration-curve generation. The order must match `calibration_concentrations`.
`calibration_integrations_label`	"Analyte"	Label of the reference compound used for calibration. The proton count associated with this label in the analyzed NMR region is used to normalize the calibration curve.
`labels_model`	["Analyte", “Impurity A”]	List of labels assigned to the reference compounds used in the fit. The order must match references_spectra and H_coefficients.
`H_coefficients`	np.array([1,2],[2,2])	Number of observable protons represented by each reference compound in the analyzed spectral region or regions. These values are used to normalize fitted peak areas before conversion into concentration. The order must match `labels_model`. If a reference has no usable signal in a given region but must still be included in the data structure, its H coefficient can be set to 1 to avoid normalization issues.
`references_spectra`	[path_to_Analyte]	List of paths to the CSV files of the reference compounds used for fitting. The order must match labels_model and H_coefficients.
`filter_list`	[[“Impurity A”],[]]	List of references to exclude from fitting in each analyzed spectral region. Each sublist corresponds to one region in peaks_info. For example, [["Impurity A "], []] excludes Impurity A from the first region but uses all references in the second region.

3.2.3. Workflow

For each substrate, the user first defines the reference spectra, calibration data, sample folder, spectral regions to analyze, reference labels, proton-normalization coefficients, and optional substrate-specific corrections. These corrections can include baseline choices, exclusion of selected references in specific spectral regions, manual addition or removal of peaks, or chemical-shift offsets.

NMR spectra are first pre-processed using an MNova script that executes the required processing functions without use of the graphical interface. The script, available in the project repository at Zenodo repository: https://zenodo.org/records/21373963, automatically loads spectra from the `/Result` subfolders of a selected parent directory, performs spectral trimming, baseline correction, phase correction, peak picking, and chemical-shift referencing, and exports the processed spectra as .csv files, for the further details see the supporting information of ref [10]. For each spectral region, NMRanger loads the reference `CSV` files, extracts the selected ppm window, applies smoothing and baseline correction, and detects the peaks present in each reference spectrum. The detected peaks are then fitted with pseudo-Voigt functions to convert each reference spectrum into a functional model described by peak amplitude, center, width, and Gaussian/Lorentzian mixing weight. Optional manual corrections are applied at this stage when required, and peaks can be grouped into multiplets to preserve their relative spacing during subsequent fitting.

An external calibration curve is then generated. If manual calibration integrals are provided, they are used directly. Otherwise, NMRanger fits the calibration spectra using the selected reference model, sums the fitted peak amplitudes, normalizes them by the corresponding number of observable protons, and relates the resulting integrals to the known calibration concentrations.

The crude reaction spectra are then analyzed. For each sample, the selected spectral region is extracted, smoothed, baseline-corrected, and fitted as a linear combination of the reference models. During this fit, each analyte is represented by a scale factor applied to its reference model, while peak positions, widths, and lineshape parameters are allowed to vary within user-defined bounds. When multiplet-aware fitting is enabled, peaks belonging to the same multiplet are constrained to

shift together. When several spectral regions are analyzed sequentially, the concentration estimates from a previously fitted region can be used to constrain the fit of subsequent regions.

Finally, fitted amplitudes are summed for each analyte, normalized by the effective number of observable protons, and converted into concentrations using the calibration curve. For each analyzed sample, NMRanger saves the fitted concentrations together with diagnostic plots showing the fitted spectrum, individual model contributions, residuals, and covariance matrix.

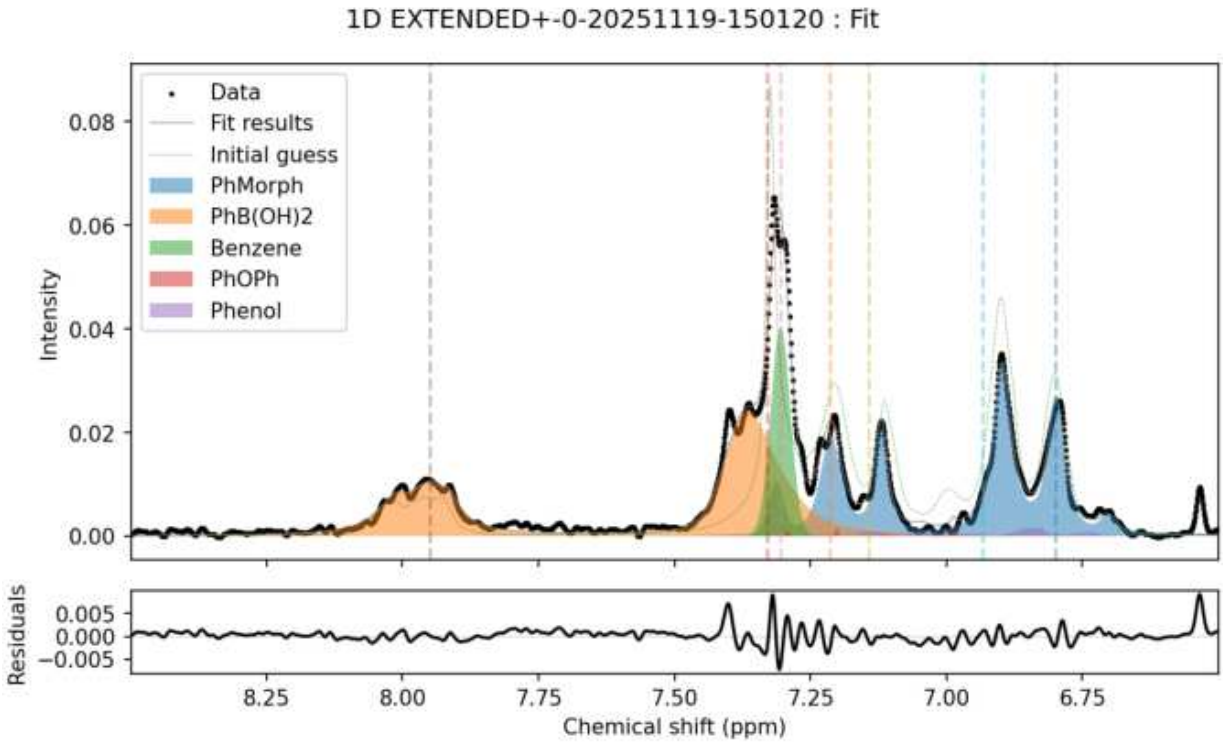


Figure S86. Representative fit from the spectral unmixing of a crude phenylboronic acid/morpholine coupling mixture. Top: experimental data are shown as black points, the optimized global fit as a grey line, and the initial guess as a dashed green line. Colored filled areas represent the fitted contributions of phenyl morpholine, phenylboronic acid, benzene, diphenyl ether, and phenol. Vertical dashed lines mark the fitted multiplet centers. Bottom: residuals after fitting, with the grey horizontal line indicating zero residual.

3.3. Illustrated tutorial

This tutorial explains how to prepare your computer to use **NMRanger** for NMR data analysis. It is intended for users with little or no coding experience, so the steps are described in detail and supported by illustrations where needed.

First, the tutorial will guide the user through the setup of a code editor, using **Visual Studio Code** — or **VS Code** — as an example. Both softwares are free to download. It will then explain how to create a suitable working environment with **Anaconda Navigator**, which will allow you to run NMRanger with the required Python version and libraries. This guide assumes that you just freshly installed both **VS Code** and **Anaconda Navigator**, and that you downloaded the file `NMRanger.zip` containing both the script `NMRanger.py` and the folder `Data`.

The tutorial will then show how to prepare the input parameters for an analysis using a demo dataset. This dataset will be available from both a GitHub repository and a Zenodo repository:

GitHub repository: https://github.com/GadinaL/NMRanger
Zenodo repository: https://zenodo.org/records/21373963

Once the parameters are prepared, the tutorial will explain how to run NMRanger and how to identify whether the analysis has started correctly.

VS Code and Anaconda are used here because they offer a simple and convenient setup. More experienced users may use other tools to edit the files or manage their Python environment if they prefer.

3.3.1. Preparation of the environment.

When using **Anaconda Navigator** for the first time, you will need to create an environment for NMRanger. To make this setup easier, we provide a file named `NMRanger-environment.yml`. This file can be used to automatically create an environment containing all the libraries required to run NMRanger.

1) Open Anaconda Navigator to reach the “Home” screen and click on “Environment” in the top-left of the screen:

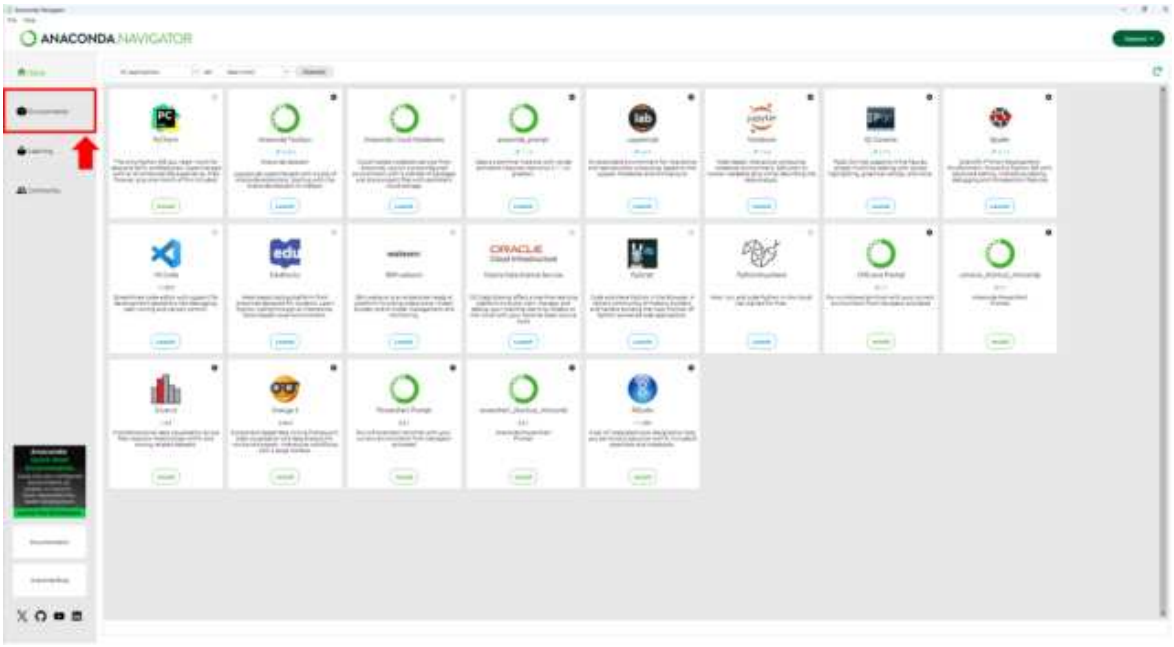


Figure S87. Anaconda Navigator “Home” screen. “Environment” panel is highlighted by a red frame and pointed by a red arrow.

2) In the “Environment” tab, click on the “Import” button at the bottom-left of the screen:



Figure S88. Anaconda Navigator “Environment” screen. “Import” button is highlighted by a red frame and pointed by a red arrow.

2) Click the black square next to the text field below **Local drive**. In the file selection window, open the folder where NMRanger was unzipped. Select the file named `NMRanger-environnement.yml`, then click **Open**. Back in **Anaconda Navigator**, click the **Import** button to create the environment.

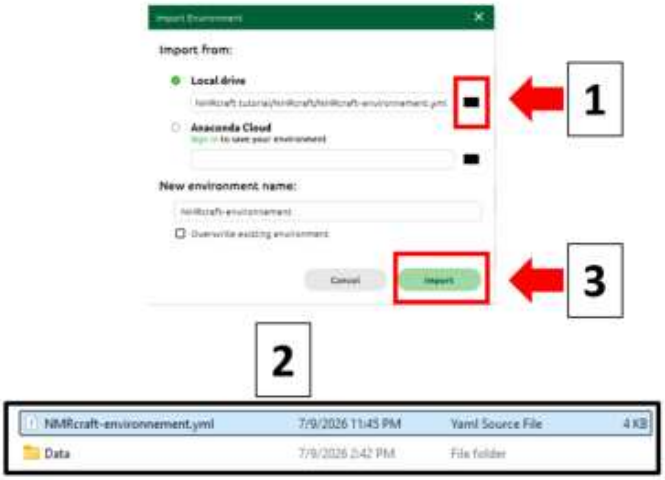


Figure S89. Anaconda navigator environment importer.

The environment is now ready. These steps need to be performed only once and do not have to be repeated between each use of **NMRanger**.

3.3.2. Interacting with NMRanger.

To run NMRanger, you must first activate the correct environment in **Anaconda Navigator**. Then, **VS Code** must be opened from the **Home** screen of Anaconda Navigator.

1. In Anaconda Navigator, go to the **Home** screen.
2. Check the environment selected in the central drop-down menu.
3. Make sure the NMRanger environment is selected.
4. If another environment is selected, click the drop-down menu and choose the NMRanger environment.
5. Once the correct environment is active, click the **Launch** button under **VS Code**.

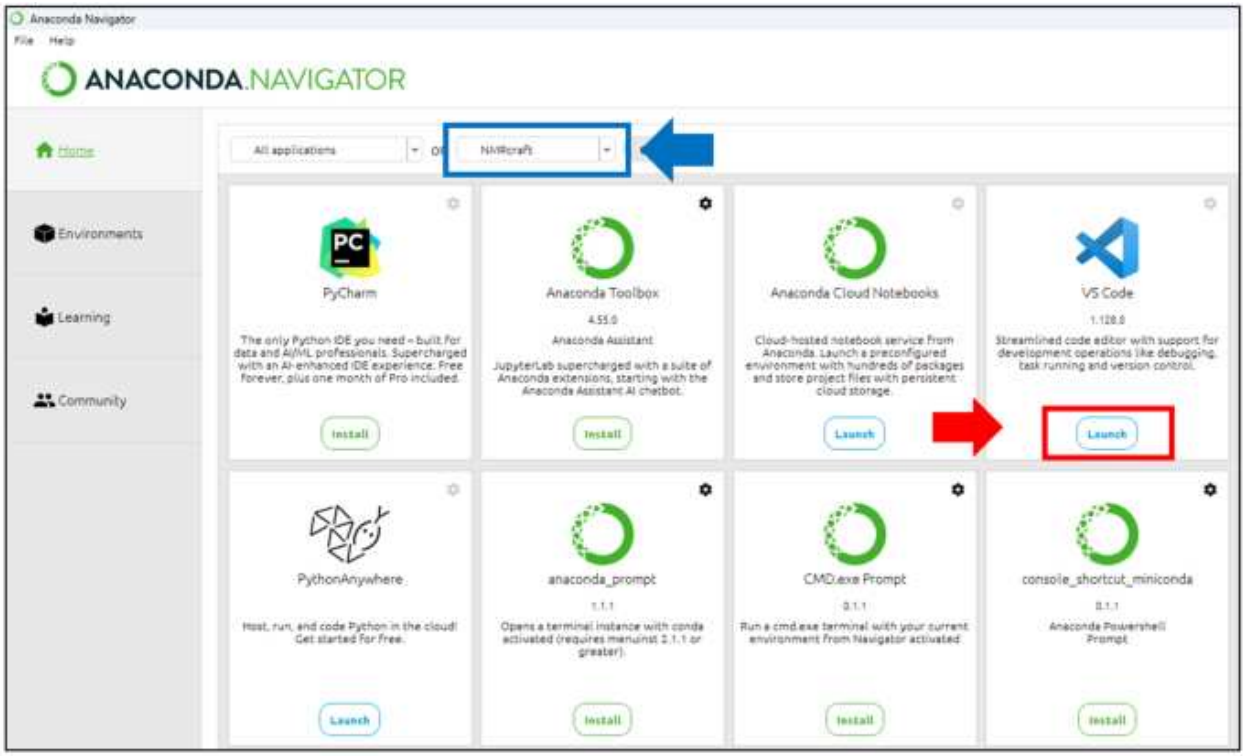


Figure S90. Anaconda Navigator "Home" screen. The environment drop-down panel is highlighted by a blue frame and pointed by a blue arrow. The VS Code launch button is highlighted by a red frame and pointed by a red arrow. Note that the position of VS Code on the screen may vary depending on your screen resolution and the application already installed in your session.

VS Code will open the NMRanger code. In this file, you will specify the location of your data files and choose the parameters for your analysis.

At the beginning of the code, four main sections are visible:

1. **Library imports** — highlighted by a red frame in Figure S91. This section loads the Python libraries required by NMRanger. Users who are not familiar with Python can ignore this section.
2. **Default parameters** — highlighted by a green frame in Figure S91. These parameters are applied to all analyses, regardless of the selected substrate.
3. **Substrate selection** — highlighted by a yellow frame in Figure S91. This section controls which substrate-specific parameters will be used for the analysis.
4. **Substrate-specific inputs** — highlighted by a blue frame in Figure S91. These parameters are only applied when the corresponding substrate is selected.

The role of each parameter is explained in more detail in the **Parameters** section of the Supporting Information.

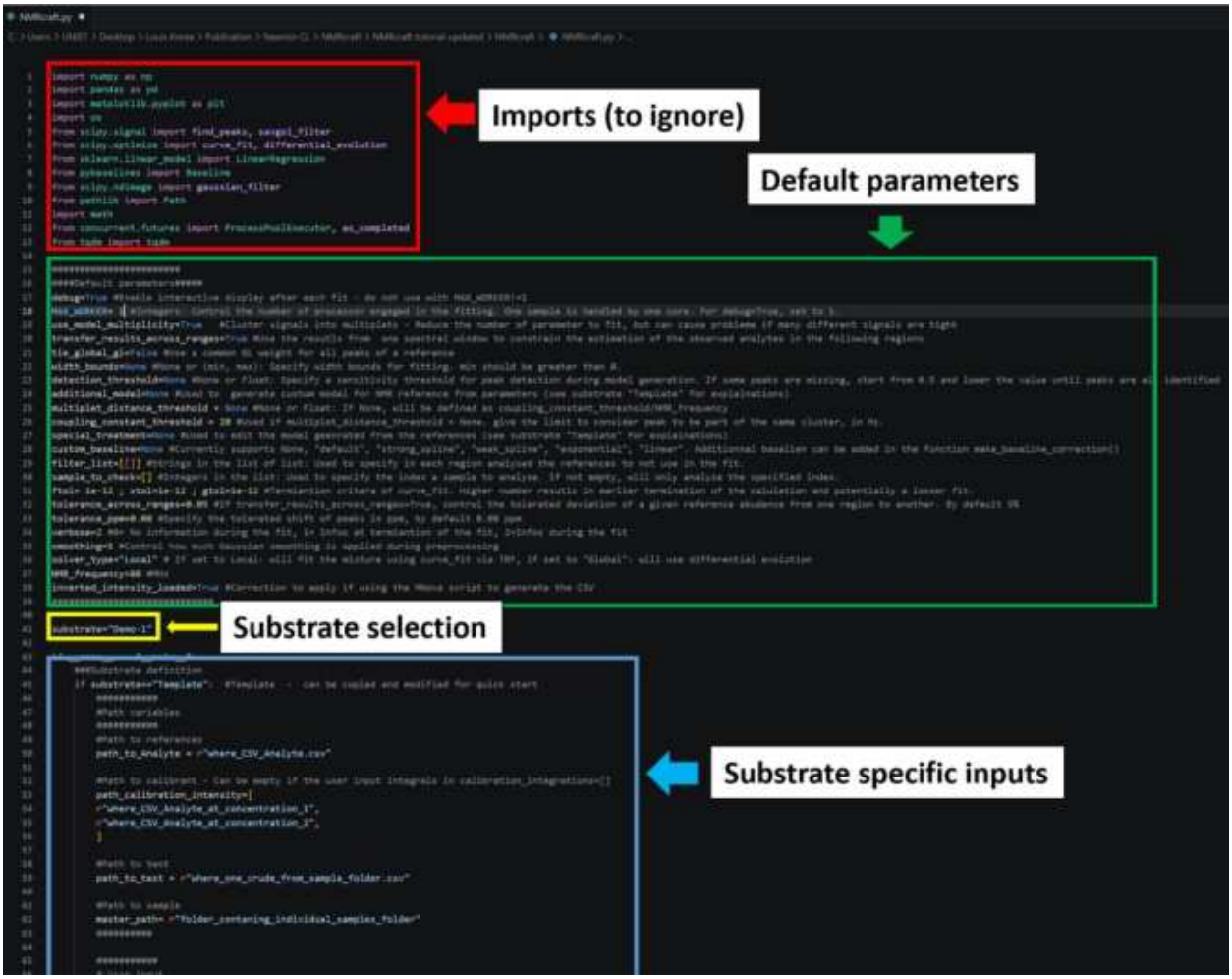


Figure S91. VS Code screen after loading NMRanger. The red frame shows the library imports, which can be ignored by users who are not familiar with Python. The green frame shows the default parameters applied to all analyses. The yellow frame shows the substrate selection, which controls which substrate-specific parameters are used. The blue frame shows the substrate-specific inputs, which are applied only when the corresponding substrate is selected.

3.3.3. Demonstration of single-range analysis with NMRanger.

For a quick demonstration of a single-range analysis, run the script by pressing **F5** or **Ctrl + F5**. You can also click **Run** at the top of the VS Code window and select **Run Without Debugging**.

This demonstration will perform the following steps:

1. generate the reference model;
2. run a test unmixing to check whether the selected parameters give reasonable results;
3. generate a calibration curve;
4. analyse 10 crude reaction mixtures obtained from the reaction of morpholine and phenylboronic acid with copper diacetate.

The ^{1}H NMR spectra of the crude mixtures were measured on an 80 MHz NMR spectrometer using a 30° pulse angle.

During model generation, each reference spectrum will be loaded one after another. For each reference, two pop-up windows will appear:

- a peak decomposition of the reference spectrum, including the fitting residuals, as shown in **Figure S92**;
- a preview of the model after optional corrections, as shown in **Figure S93**. These corrections are used to handle specific cases where the automatic model may need adjustment.

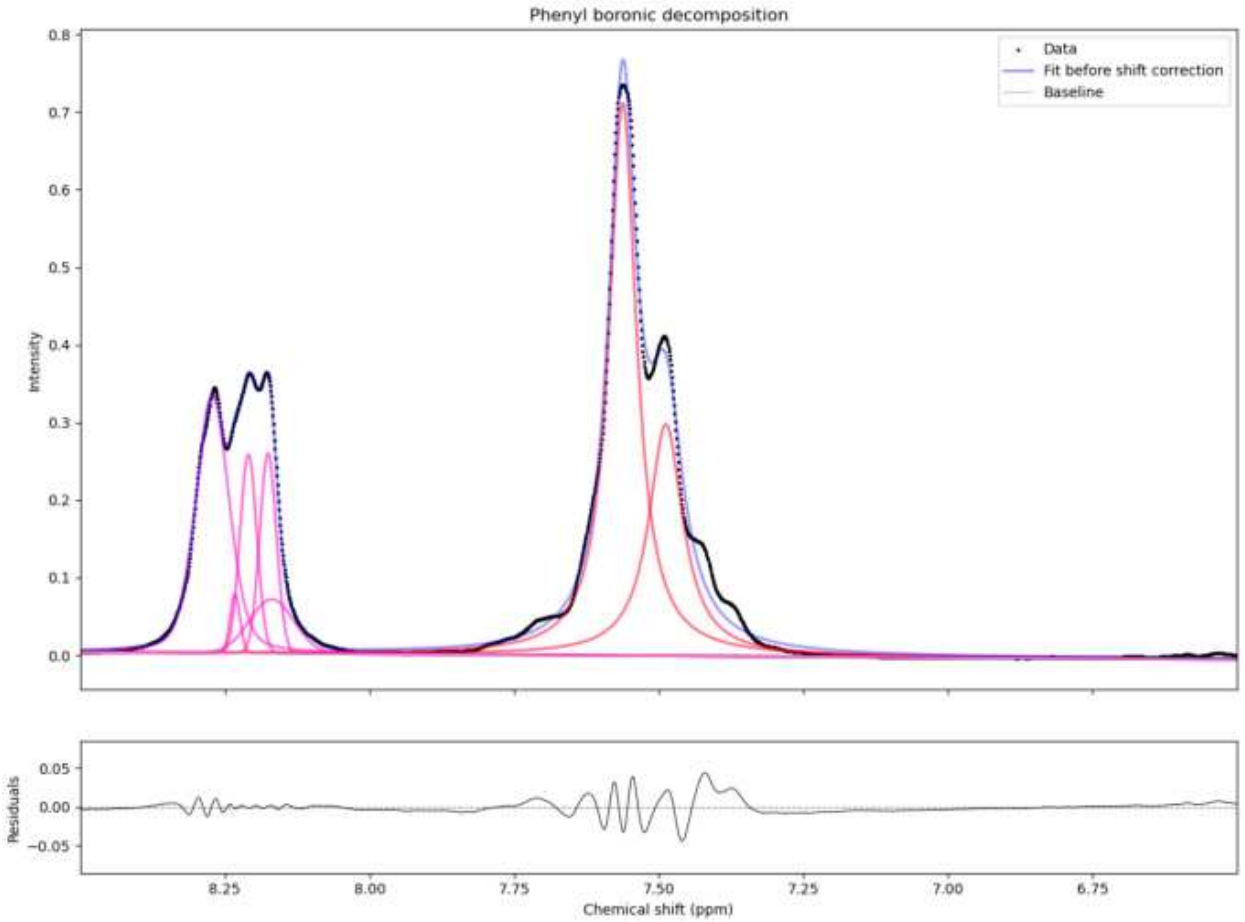


Figure S92. Example of decomposition of a reference spectrum into a sum of Voigt functions. The top panel shows the phenylboronic acid reference spectrum and its fitted model. The experimental data are shown as black dots, and the overall model is shown as a pale blue line, which overlaps closely with the data in most regions. The pink and red curves represent groups of Voigt peaks assigned to different multiplets. The bottom panel shows the fitting residuals as a black line, with the dashed grey line marking the zero level. The residuals do not exceed 0.05 intensity units.

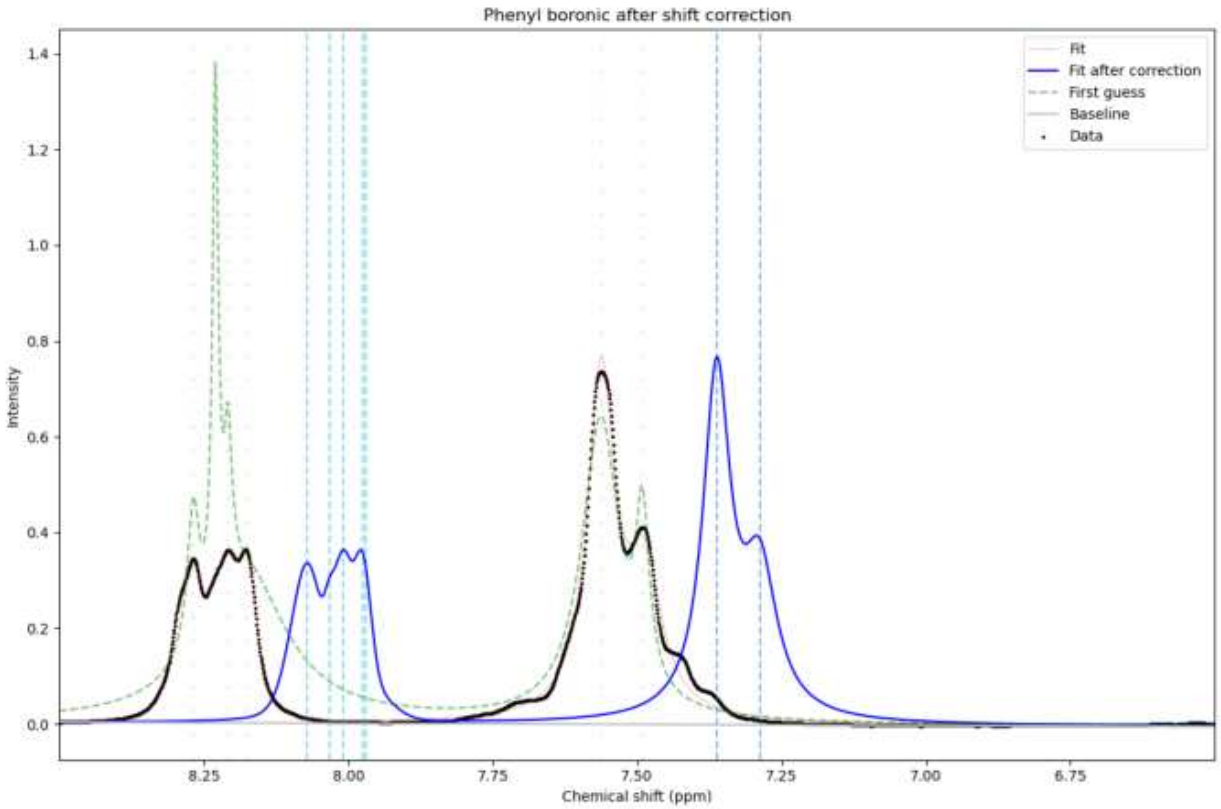


Figure S93. Correction preview after generation of the phenylboronic acid model. The vertical dashed lines indicate the detected peak positions and are color-coded by multiplet. The black dots show the experimental data, the green line shows the initial fit, and the blue line shows the corrected model used for mixture analysis. Here, the peak positions were shifted by −0.2 ppm to account for an unusually large shift observed in the crude mixtures. Such manual corrections are only needed in special cases, as NMRanger normally handles smaller peak shifts automatically. Increasing `tolerance_ppm` could also allow larger shifts, but would reduce confidence for the identification of other analytes. This correction should only be applied after a doping experiment confirms the assignment of the shifted signals. See **Doping experiments** in the Supporting Information. This illustrate how NMRanger can be adapted to handle corner cases.

After all reference models will have been generated, NMRanger will perform a test unmixing on one selected crude mixture. This step allows the user to check whether the chosen parameters give a good fit. Two windows will be displayed:

- the fitted spectrum and fit result, as shown in **Figure S94**;
- the covariance matrix of the fitted parameters, as shown in **Figure S95**.

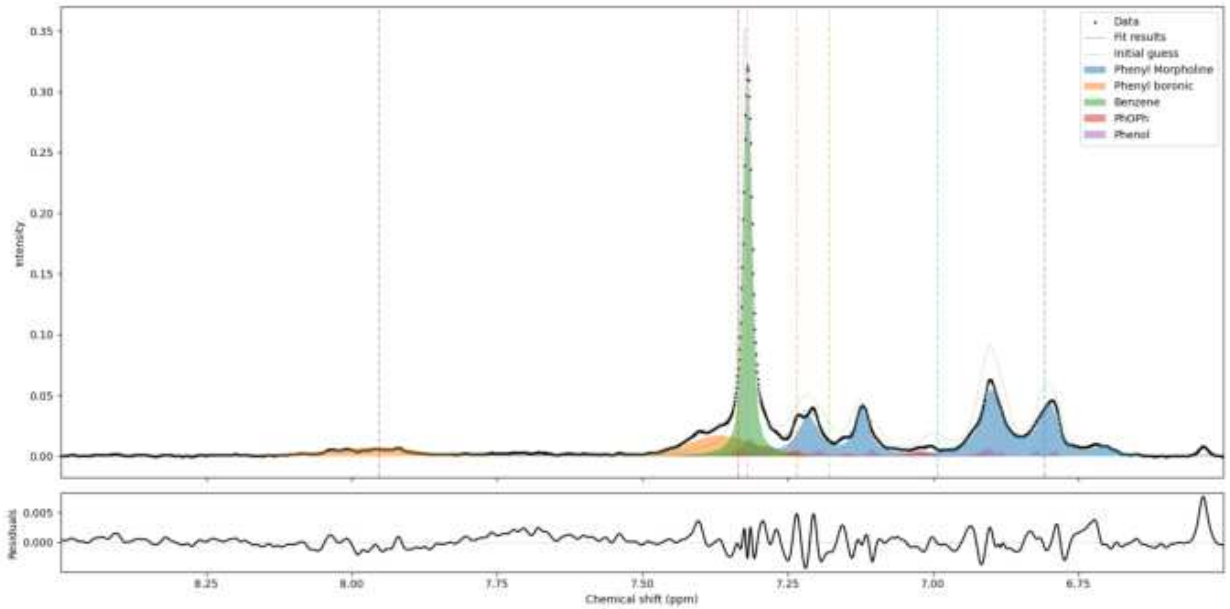


Figure S94. Example of crude mixture analysis. In the top panel, the black dots show the experimental data, the grey line shows the best fit, and the dashed green line shows the initial guess. The colored filled areas represent the fitted contributions of each analyte: phenyl morpholine in blue, phenylboronic acid in orange, benzene in green, diphenyl ether in red, and phenol in purple. The vertical dashed lines indicate the fitted centers of the multiplets (7 multiplets in this case). The bottom panel shows the fitting residuals as a black line, with the zero level indicated by a grey line. In this example, the residuals remain below 0.005 intensity units.

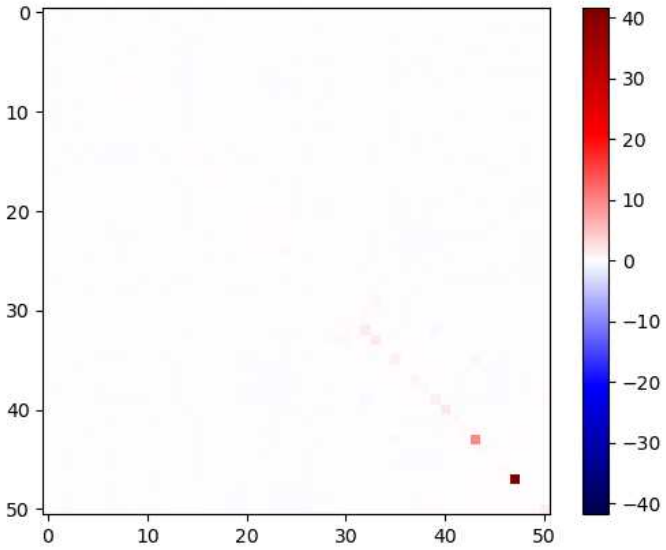


Figure S95. Example of covariance matrix generated after fitting. The diagonal elements indicate the uncertainty associated with each fitted parameter, while off-diagonal elements indicate covariance between pairs of parameters. In this example, 7 multiplets were fitted. Parameters 0–6 correspond to the scale factors, parameters 7–13 to the width scaling factors, parameters 14–20 to the shifts applied to the multiplet centers, and parameters 21–50 to the Gaussian weights of the individual peaks. The Gaussian weights can show higher uncertainties because the exact peak shape may be poorly defined, especially when peaks are not well resolved. Strong off-diagonal features indicate that two parameters influence each other strongly, which may suggest an ill-defined fit and should prompt closer inspection.

Next, a calibration curve will be generated using a series of phenyl morpholine spectra measured at different concentrations, as shown in **Figure S96**.

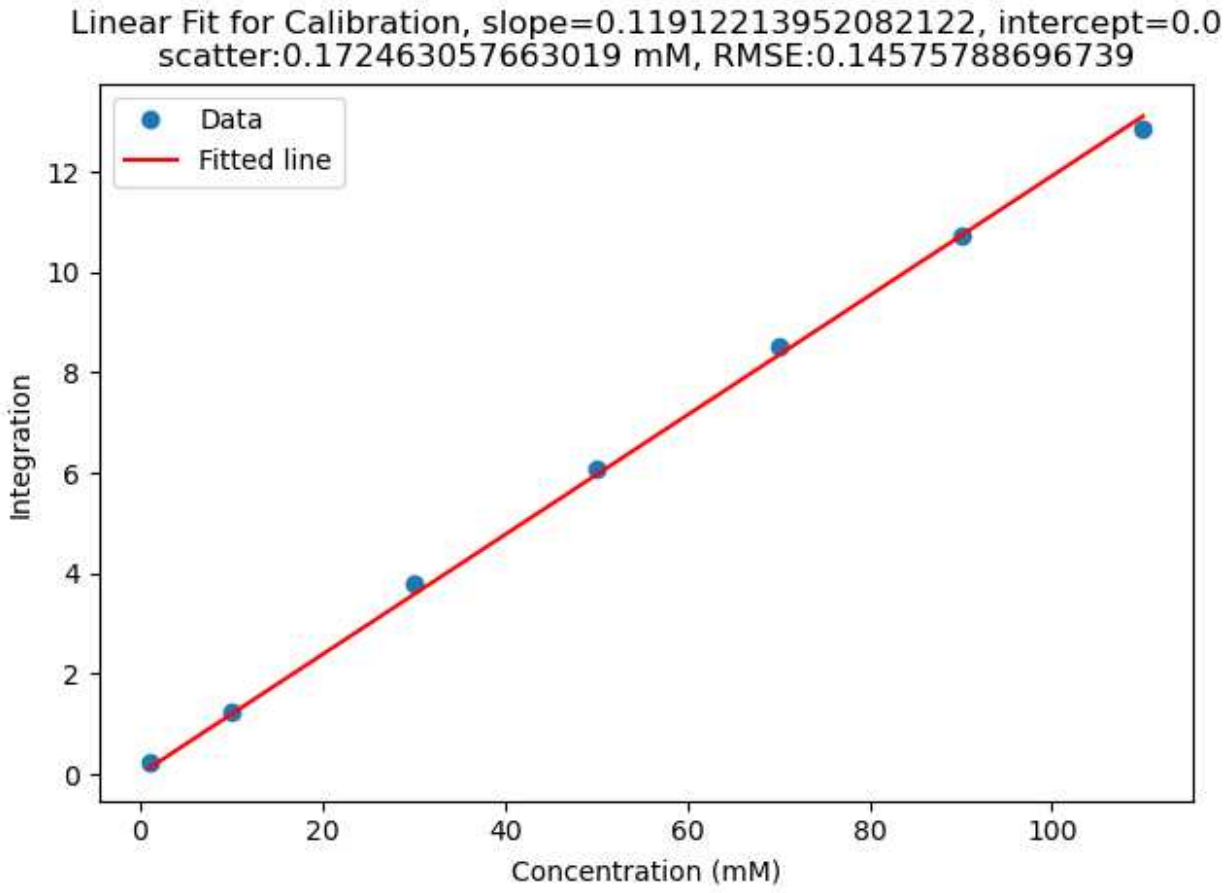


Figure S96. Example of calibration curve generated by NMRanger for phenyl boronic acid.

Finally, NMRanger will analyze each crude sample. After each calculation, two windows will be displayed: the fitted spectrum and fit result and the covariance matrix of the fitted parameters.

For a faster analysis, the calculations can also be run in parallel. To do this, go to the default parameters and set:

```
debug = False
MAX_WORKER = 5
```

or:

```
debug = False
MAX_WORKER = 10
```

In this mode, the full dataset will be analyzed without visual pop-up windows. Instead, the fit images will be saved directly in each experiment folder. Depending on the number of workers selected and the computer used, this step may take from a few seconds to about one minute.

In both modes, **NMRanger** will create a CSV file named `integrations_and_concentrations.csv` after the analysis. This file summarizes the detected composition of each mixture.

Important note: do not use: `debug = True` when: `MAX_WORKER` is different from `1`. This can cause unstable or unexpected behaviour.

3.3.4. Demonstration of multiple-range analysis with NMRanger.

For a quick demonstration of a multiple-range analysis, first set the substrate selection to:

```
substrate = "Demo-2"
```

Then run the script by pressing **F5** or **Ctrl + F5**. You can also click **Run** at the top of the VS Code window and select **Run Without Debugging**.

This demonstration will perform the following steps:

1. generate the reference models;
2. run a test unmixing to check whether the selected parameters give reasonable results;
3. generate a calibration curve;
4. analyze 10 crude reaction mixtures obtained from the reaction of azaindole and phenylboronic acid with copper diacetate in the presence of DBU.

The ^{1}H NMR spectra of the crude mixtures were measured on an 80 MHz NMR spectrometer using a 90° pulse angle.

The same steps described in the **Demonstration of single-range analysis** section will be performed. First, NMRanger will analyze the [5.7, 7.0] ppm region. After the 10 crude samples have been analyzed in this region, the full sequence will be repeated for the [5.7, 9.5] ppm region.

For a faster analysis, the calculations can also be run in parallel. To do this, go to the default parameters and set:

```
debug = False
MAX_WORKER = 5
```

or:

```
debug = False
MAX_WORKER = 10
```

In this mode, the full analysis will run in parallel without visual pop-up windows. Instead, the fit images will be saved directly in each experiment folder. Depending on the number of workers selected and the computer used, this operation may take approximately 15 to 30 minutes.

Important note: do not use: `debug = True` when: `MAX_WORKER` is different from `1`. This can cause unstable or unexpected behavior.

3.3.5. Preparing a set of parameters for a new substrate.

To analyze a new mixture, some parameters must be defined specifically for that mixture. These parameters are selected through the substrate selection line, highlighted by the yellow frame in **Figure S91**.

For each new substrate or mixture type, the following information must be provided:

1. the paths to the data files;
2. the NMR region to analyze;
3. the concentrations used to build the calibration curve;
4. the reference used for calibration, so that NMRanger knows how many protons are present in the selected NMR region;
5. the list of reference labels to include in the basis set;
6. the paths to the CSV files of the reference spectra, in the same order as the reference labels.

The calibration concentrations must be written as an ordered list. This order must match the order of the calibration CSV files. Similarly, the list of reference CSV files must follow the same order as the list of reference labels.

Optional corrections can also be added after these required parameters. These corrections are only needed in specific cases, for example when one reference requires a manual peak shift correction.

1) For a quick start, NMRanger includes a `Template` substrate. This template can be copied and modified to define the parameters for a new mixture. To use it, scroll down in the code until you find the section:

```
if substrate == "Template":
```

Then copy the full template section, from:

```
if substrate == "Template":
```

to:

```
filter_list = [[]]
```

as shown in **Figure S97**.

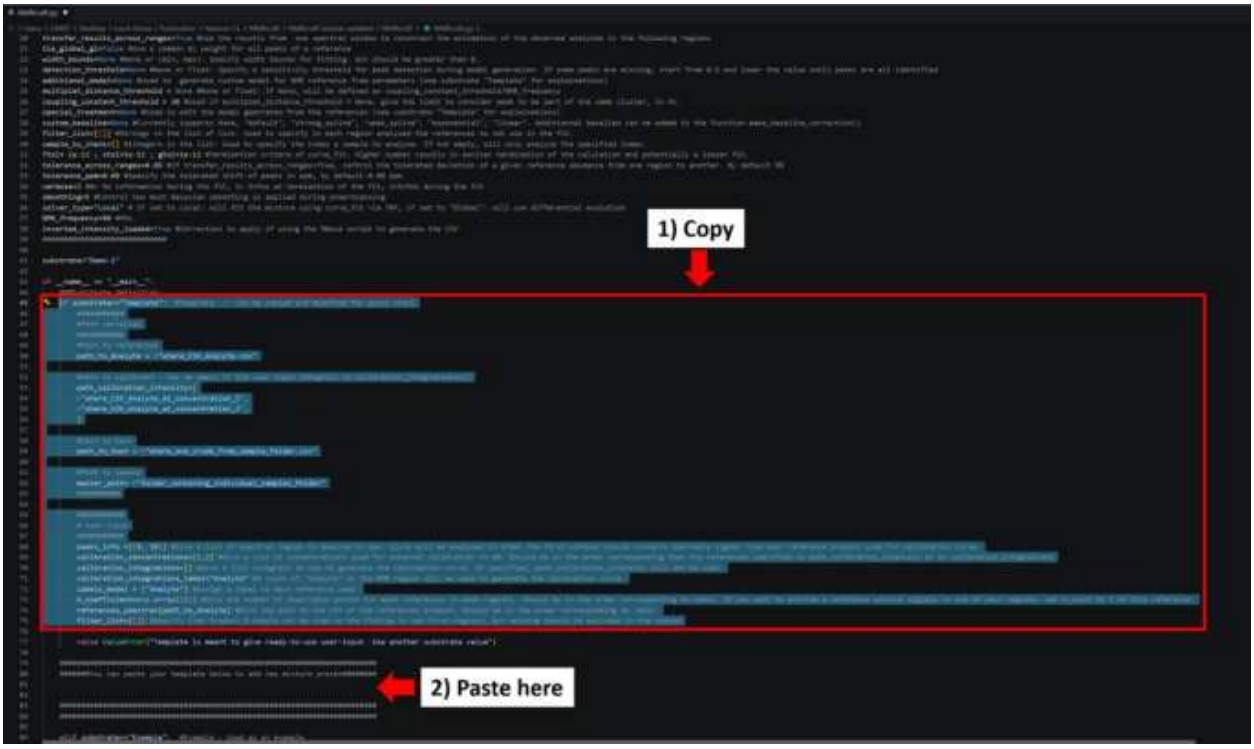


Figure S97. Code screenshot showing the location of the substrate input template and where to duplicate it.

After pasting the template, your screen should look like **Figure S98**:

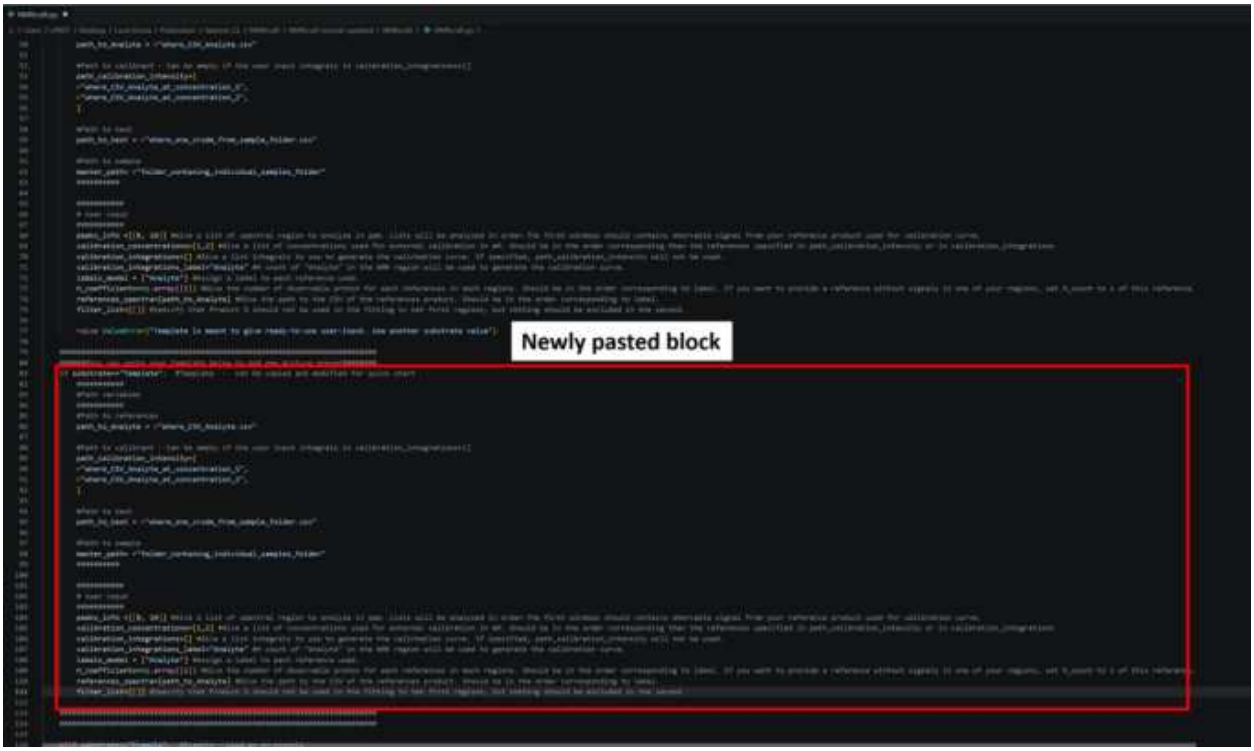


Figure S98. VS Code screenshot after pasting the substrate input template. The pasted block must be indented one level to the right, except for the first line defining the substrate condition.

Make sure that all pasted lines are indented one level to the right, except the first line:

```
if substrate == "Template":
```

Only this first line should stay aligned with the other substrate definitions, as shown **Figure S98**.

2) Next, modify the beginning of the copied block:

1. Replace `if` with `elif`.
2. Replace `"Template"` with the name of your new substrate.

For this example, we will use:

```
elif substrate == "My_substrate":
```

3) Then define the CSV files that will be used for the analysis.

In the folder where NMRanger.zip was extracted, demonstration datasets are provided. Open Windows Explorer and go to:

```
NMRanger\Data\References\Demo-1
```

Copy the path to each reference CSV file and paste it in the section below:

```
# Path to references
```

Each reference analyte is defined by one variable, for example:

```
path_to_analyte_1                                                    =
r"C:\Users\Downloads\NMRanger\Data\References\Demo-
1\Benzene\data.csv"
```

The path must start with `r"` and end with `"`. This tells Python to read the Windows path correctly.

To keep the same color code as in the first demonstration, use the following order:

```
path_to_analyte_1 → Phenyl morpholine reference
path_to_analyte_2 → Phenyl boronic acid reference
path_to_analyte_3 → Benzene reference
path_to_analyte_4 → Diphenyl ether reference
path_to_analyte_5 → Phenol reference
path_to_analyte_6 → Morpholine reference
```

At this stage, your screen should look like **Figure S99**.

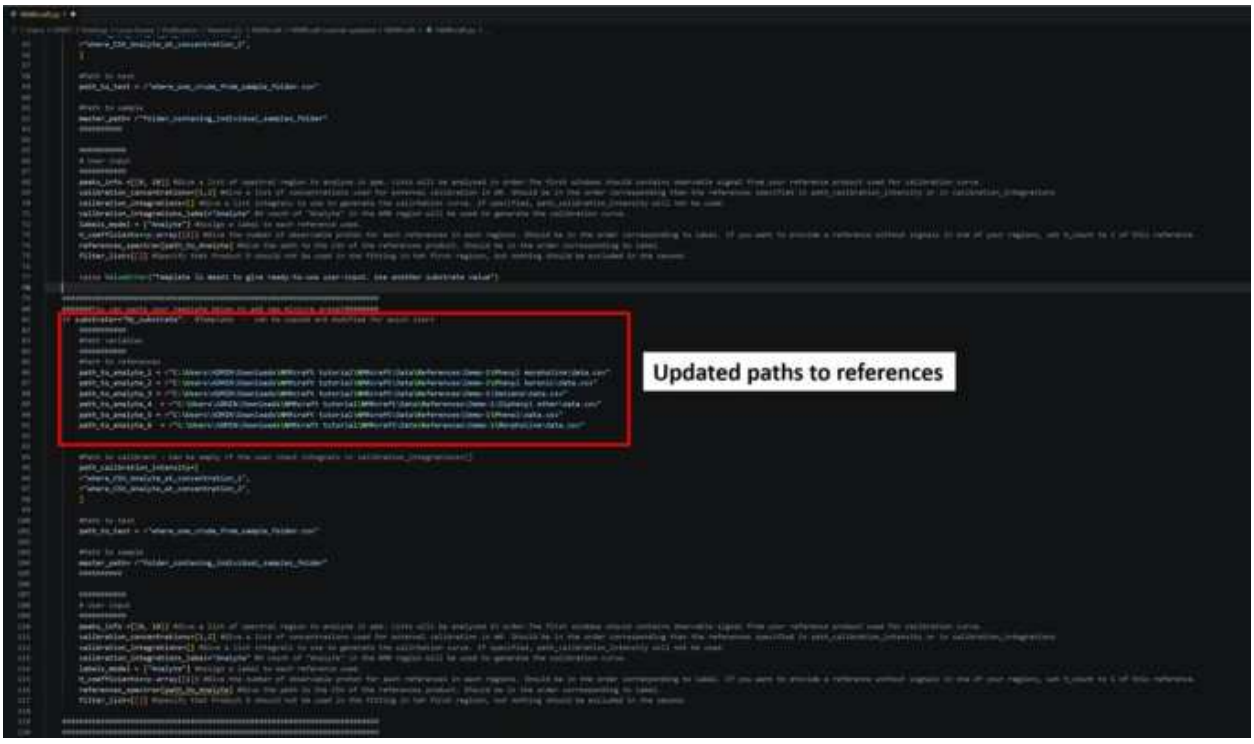


Figure S99. VS Code screenshot after defining the paths to the reference CSV files for the new substrate.

4) Update the paths to the CSV files used to generate the calibration curve.

In the folder where NMRanger.zip was extracted, demonstration datasets are provided. Open Windows Explorer and go to:

`NMRanger\Data\References\Demo-1\Calibration`

Copy the path to each calibration CSV file and paste it in the section below:

`# Path to calibrant`

The calibration paths must be written as an ordered list. Each path must start with `r"` and end with `"`, as in the example below:

```
r"C:\Users\Downloads\NMRanger\Data\References\Demo-1\Calibration\PhMorph_1mM\data.csv"
```

For this example, enter the calibration paths in increasing order of concentration, from the lowest concentration to the highest concentration. This order must match the order of the concentrations that will be entered later.

At this stage, your screen should look like **Figure S100**.



Figure S100. VS Code screenshot after defining the paths to the calibration CSV files. The paths are listed in increasing order of concentration.

5) Provide the path to one representative crude mixture and to the folder containing all crude samples.

The representative mixture will be used to test whether the selected parameters give a good fit. Its path must be entered below:

```
# Path to test
```

Open Windows Explorer and go to the following folder inside the extracted NMRanger folder:

```
NMRanger\Data\Datasets\Demo-1\Results\1D    EXTENDED+-0-20251104-
165215
```

Copy the path to the corresponding `data.csv` file and paste it after:

```
path_to_test =
```

The folder containing all crude samples must then be entered below:

```
# Path to sample
```

For this example, copy the path to the full Results folder:

```
NMRanger\Data\Datasets\Demo-1\Results
```

and paste it after:

```
master_path =
```

Each Windows path should start with `r"` and end with `"`, as shown below:

```
path_to_test  =  r"C:\Users\Downloads\NMRanger\Data\Datasets\Demo-
1\Results\1D EXTENDED+-0-20251104-165215\data.csv"

master_path  =  r"C:\Users\Downloads\NMRanger\Data\Datasets\Demo-
1\Results"
```

At this stage, your screen should look like **Figure S101**.

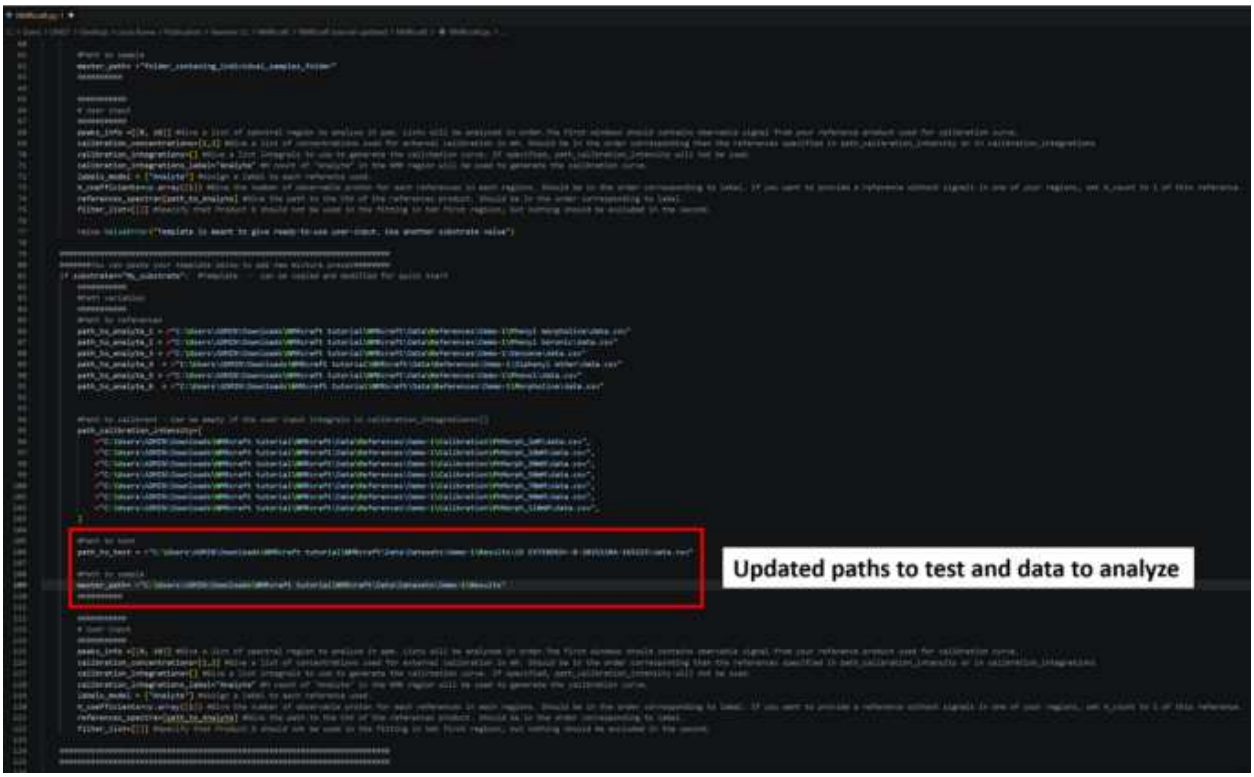


Figure S101. VS Code screenshot after defining the path to the representative test mixture and the path to the folder containing the crude samples to analyse.

6) Define the NMR region to analyze.

The NMR region is defined by its lower and upper ppm limits:

```
[lower_limit, upper_limit]
```

The regions must be written as a list of regions. This means that each region is placed inside its own pair of brackets, and all regions are placed inside a second pair of brackets.

For example, a single region must be written as:

```
peak_infos = [[6.5, 8.5]]
```

Do not write:

```
peak_infos = [6.5, 8.5]
```

because NMRanger expects a list of regions, even when only one region is analysed.

For this example, change:

```
peak_infos = [[0, 10]]
```

to:

```
peak_infos = [[6.5, 8.5]]
```

7) Provide the concentrations used for the calibration curve.

The concentrations must be written in the same order as the calibration CSV files listed below:

```
# Path to calibrant
```

For this example, use:

```
calibration_concentrations = [1, 10, 30, 50, 70, 90, 110]
```

8) Leave `calibration_integrations` empty.

For this tutorial, the calibration curve will be generated directly from the calibration spectra. Therefore, keep:

```
calibration_integrations = []
```

This parameter can also be used to provide integration values directly, instead of calibration spectra. In that case, NMRanger would use the values in `calibration_integrations` to generate the calibration curve and would not load calibration spectra.

9) Define the calibration reference label.

The parameter `calibration_integrations_label` tells NMRanger which reference compound is used for the calibration curve. This is needed so that NMRanger can use the correct number of protons in the selected NMR region.

For this example, use:

```
calibration_integrations_label = "PhMorph"
```

10) Define the reference labels, proton numbers, and reference paths.

First, provide the labels of the reference spectra used in the analysis. These labels must be written in the same order as the paths defined in the section:

```
# Path to references
```

For this example, use:

```
labels_model = ["PhMorph", "PhB(OH)2", "Benzene", "PhOPh",
"Phenol", "Morpholine"]
```

Next, provide the number of protons visible in the selected NMR region for each reference. The order must match the order of `labels_model`.

For this example, use:

```
nb_protons = [np.array([5, 5, 6, 10, 5, 1])]
```

The external brackets indicate the list of analyzed regions. Since only one region is analysed here, only one `np.array(...)` is needed.

If several NMR regions are analyzed, add one `np.array(...)` per region. For example, for three regions, the structure would be:

```
nb_protons = [
    np.array([...]),
    np.array([...]),
    np.array([...])
]
```

Finally, provide the list of paths to the reference spectra. The order must again match `labels_model`.

For this example, use:

```
model_path = [
    path_to_analyte_1,
    path_to_analyte_2,
    path_to_analyte_3,
    path_to_analyte_4,
    path_to_analyte_5,
    path_to_analyte_6
]
```

11) Leave `filter_list` unchanged.

For this single-region example, keep:

```
filter_list = [[]]
```

This parameter is used to exclude one or more analytes from a specific NMR region. For example, it can be useful when a peak is partly outside the selected region. In that case, the baseline close to the truncated peak may be biased, and the analyte should not be quantified in that region.

If several NMR regions are analysed, `filter_list` must contain one internal list per region. For example, for three NMR regions, use:

```
filter_list = [[], [], []]
```

At this stage, the user input section should look like **Figure S102**.

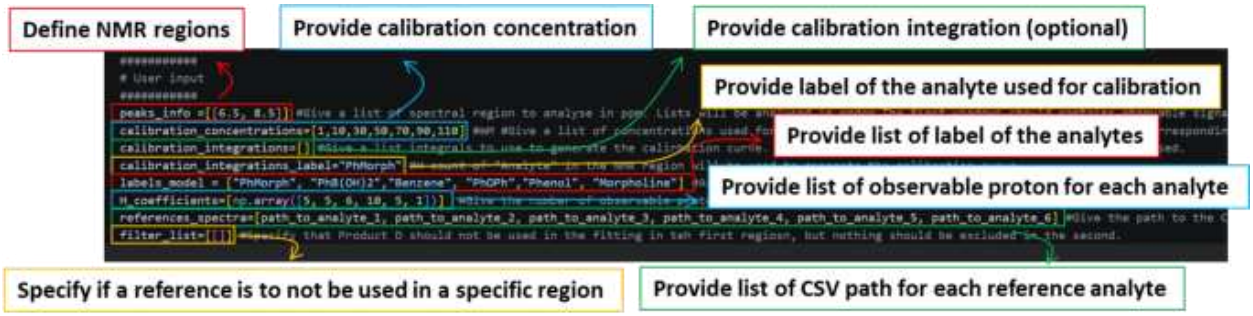


Figure S102. VS Code screenshot after defining the NMR region, calibration information, reference labels, proton numbers, reference paths, and filter list for the new substrate.

After following these steps, the new parameter set is defined and ready for a basic NMRanger analysis.

To use it, update the substrate selection line, highlighted by the yellow frame in Figure S7, to:

```
substrate = "My_substrate"
```

Then run the script by pressing **F5** or **Ctrl + F5**. You can also click **Run** at the top of the VS Code window and select **Run Without Debugging**.

For this tutorial, one additional correction will be added. This correction is needed because phenylboronic acid shows an unusually large peak shift in these crude mixtures.

1. Add the following block just after `filter_list`:

```
special_treatment = {
    "PhB(OH)2": {
        1: {
            "rule_type": "offset",
            "shift_offset": -0.20,
            "range": "> 6.0"
        },
    },
}
```

The variable `special_treatment` is used to define specific corrections for selected analytes.

In this example, the correction is applied to the analyte labelled `"PhB(OH)2"`. The rule tells NMRanger to identify peaks with centers above 6.0 ppm and shift their model position by −0.20 ppm. This compensates for the unusually large chemical shift observed for phenylboronic acid in the crude mixtures.

A detailed list of the available correction rules and their parameters is provided in the **Substrate-specific Parameters** section.

The code should now look like **Figure S103**.

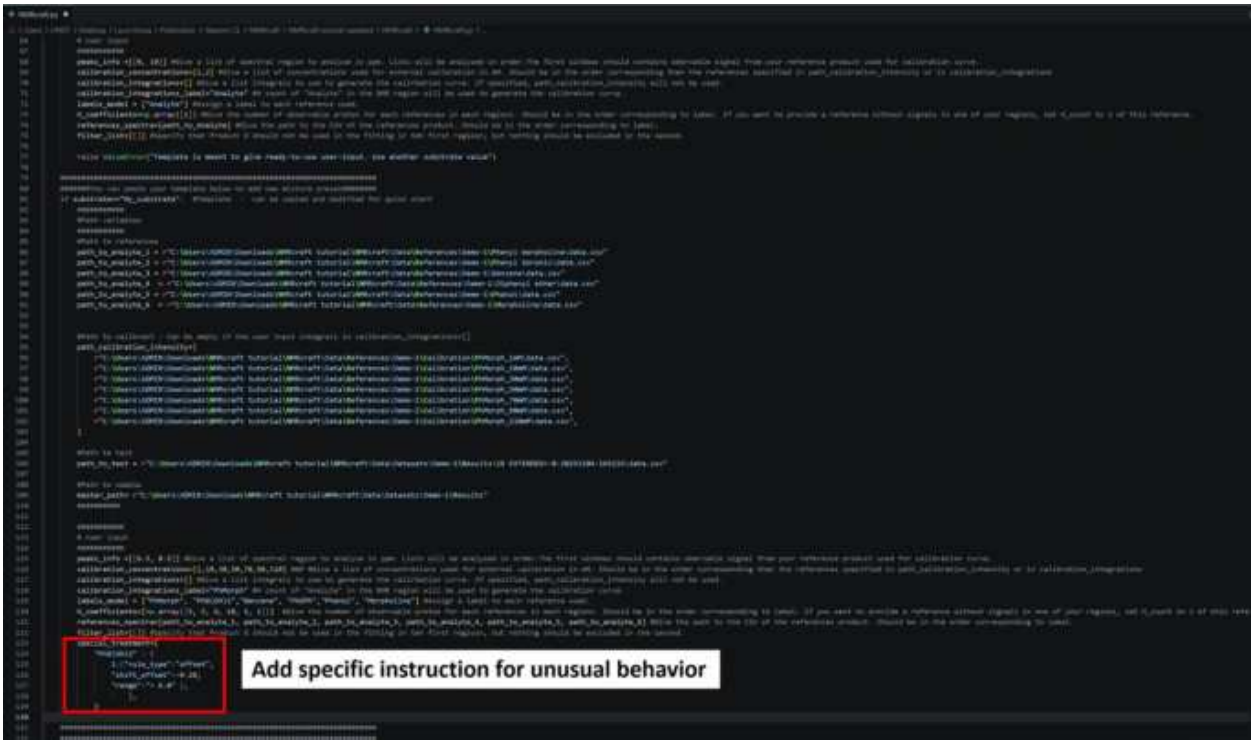


Figure S103. VS Code screenshot after adding the `special_treatment` **correction for phenylboronic acid.** This correction shifts selected peak positions by −0.20 ppm to account for the unusually large chemical shift observed in the crude mixtures.

NMRanger is now ready for analysis. To run the script, press **F5** or **Ctrl + F5**. You can also click **Run** at the top of the VS Code window and select **Run Without Debugging**.

4. NMRanger Parameter Selection for Chan-Lam Coupling

4.1. General substrate-specific parameters selection

Substrate-specific NMRanger parameters for Chan–Lam reaction analysis were selected to retain all resonances required for reliable quantification, exclude non-informative spectral regions or regions affected by overlapping solvent and impurity signals, and improve fitting stability in highly overlapped crude spectra. The 7-azaindole workflow is presented below as a representative example (**Figure S104**).

```
# User input
##########
peaks_info =[[5.7, 7.0],[5.7, 9.5]]
calibration_integrations=[]
calibration_concentrations=[5,10,20,50,80,100] #mM
calibration_integrations_label="Ph-Azaindole"
labels_model = ["Ph-Azaindole", "Benzene", "PhBOH2","DBU","Azaindole", "PhOPh","Phenol"]
H_coefficients=[np.array([1, 0, 0, 1, 1, 5, 3]),np.array([10, 6, 5, 1, 2.3, 10, 5])]
references_spectra=[path_to_Ph_Azaindole, path_to_Benzene_90, path_to_PhBOH2_90, path_to_DBU_90, path_to_AzaCu, path_to_PhOPh, path_to_phenol_90]
filter_list=[["PhOPh","Phenol","Azaindole"],[]]
sample_to_check=[]
```

Figure S104. Representative NMRanger substrate-specific parameters used for 7-azaindole reactions. Screenshot of the parameter settings in VS Code.

For 7-azaindole, the 5.7–9.5 ppm window was selected because all analytes relevant to quantification are aromatic and therefore have their diagnostic resonances within this region. The aliphatic region was deliberately excluded because it is dominated by the broad residual non-deuterated solvent signal and does not contain azaindole-derived resonances needed for quantification. Within the selected aromatic window, the less congested 5.7–7.0 ppm region was fitted first to obtain a reliable abundance estimate for N-phenyl-7-azaindole, which was then transferred to guide fitting of the full 5.7–9.5 ppm range. This region is therefore sufficient for complete quantification while avoiding solvent-induced baseline distortion and unnecessary spectral interference (**Figure S105**).

The `H_coefficients` values were assigned according to the number of observable protons contributed by each compound within the selected fitting region (**Figure S104**). Although free 7-azaindole contains five aromatic protons in the 5.7–9.5 ppm window, observable-proton experiments showed that copper complexation suppresses part of its NMR signal; therefore, an effective proton count of 2.3 was used for 7-azaindole. The procedure used to determine the observable proton count is described in **Section 1.13**.

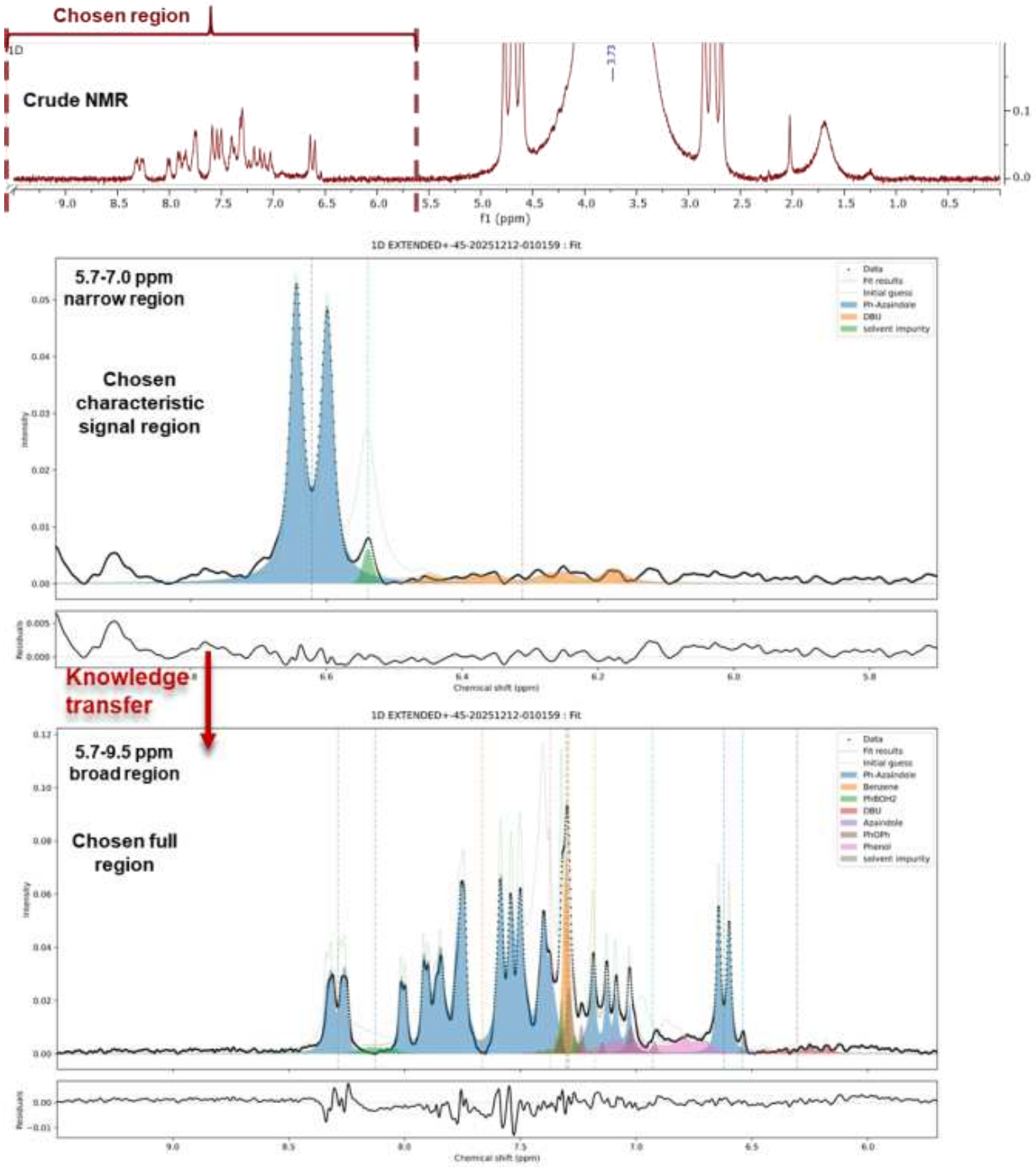


Figure S105. Selection of fitting regions and sequential knowledge transfer for NMRanger analysis of 7-azaindole reaction mixtures. The aromatic 5.7–9.5 ppm window was selected because it contains all resonances required for quantification, while the aliphatic region was excluded due to the broad residual non-deuterated solvent signal. The less congested 5.7–7.0 ppm region was fitted first to obtain a reliable estimate of N-phenyl-7-azaindole abundance, which was then transferred to guide fitting of the full 5.7–9.5 ppm region containing all modeled reaction components.

In the `filter_list`, 7-azaindole was excluded from the narrow fitting window because copper complexation suppresses its resonances in this region. Diphenyl ether was also excluded because only part of its multiplet is visible within the narrow window, making reliable quantification difficult. Phenol was excluded as well because its signals overlap with those of diphenyl ether and, in the absence of a reliable diphenyl ether contribution, including phenol could promote overfitting (**Figure S106**). These components were therefore fitted only in the broader 5.7–9.5 ppm region.

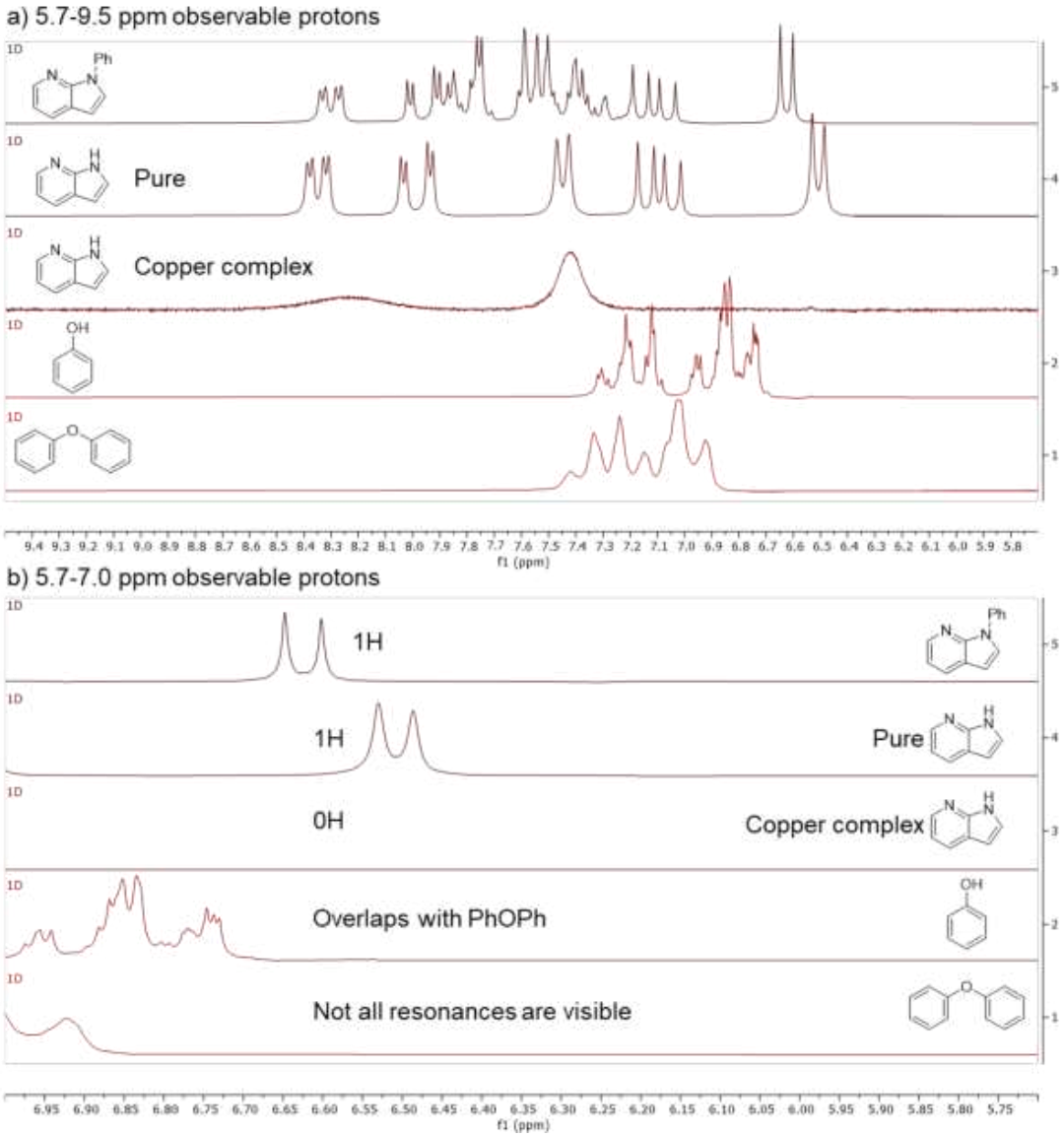


Figure S106. Observable resonances of the reference components in the broad 5.7–9.5 ppm region (a) and the narrow 5.7–7.0 ppm window (b). Within the narrow window, N-phenyl-7-azaindole and free 7-azaindole each contribute one observable proton, whereas the copper-complexed 7-azaindole (which we use as reference) signal is not observed. The phenol resonances overlap with those of diphenyl ether, and only part of the diphenyl ether multiplet is visible.

4.2. Additional substrate-specific treatments

NMRanger includes configurable compound-specific treatments to account for spectral features not adequately represented by the unmodified reference spectra. Additional signals from non-isolatable impurities can be incorporated using the `additional_model` by specifying their proton count and initial peak parameters, as illustrated in **Figure S107**, for a solvent impurity at 6.54 ppm. The initial guess is defined by four parameters: peak area (a), center (c), width (w), and weight of Gaussian peak function (p). Further details of the fitting algorithm are provided in **Section 1 of the main text.**

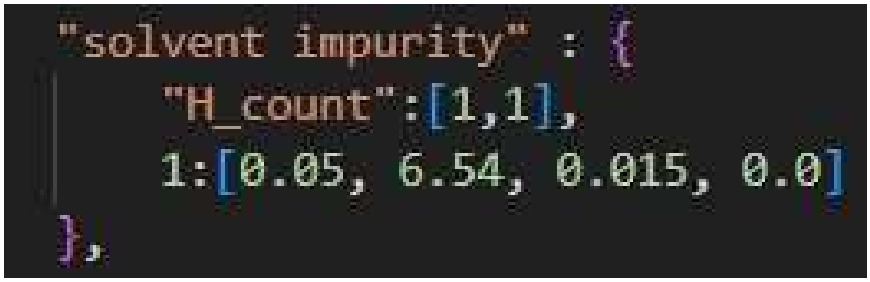


Figure S107. Additional model: solvent impurity. Definition of an additional model for an unidentified solvent impurity. When the proton count is unknown, `H_count` is set to 1; this value is not used for quantification but allows the additional signal to be represented during fitting. The parameter list [0.05, 6.54, 0.015, 0.0] defines the initial values of a, c, w, p. The peak center c is the most critical parameter because it specifies the peak position, here 6.54 ppm. Initial values for the amplitude a and width w are typically selected on the basis of the expected order of magnitude, commonly within 0.005–0.01, while $p = 0$ defines a purely Lorentzian line shape. After initialization, NMRanger optimizes these parameters during fitting.

The `special_treatment` instructions allow specific components to be adjusted before fitting to the reference analyte spectrum.

The `remove` option is used to exclude a resonance within a defined chemical-shift range when this signal is known not to originate from the assigned reference compound. For example, as shown in **Figure S108** , a solvent-impurity signal appears close to the DBU resonances and could otherwise be incorrectly fitted as part of the DBU contribution. Removing this narrow region prevents this misassignment.

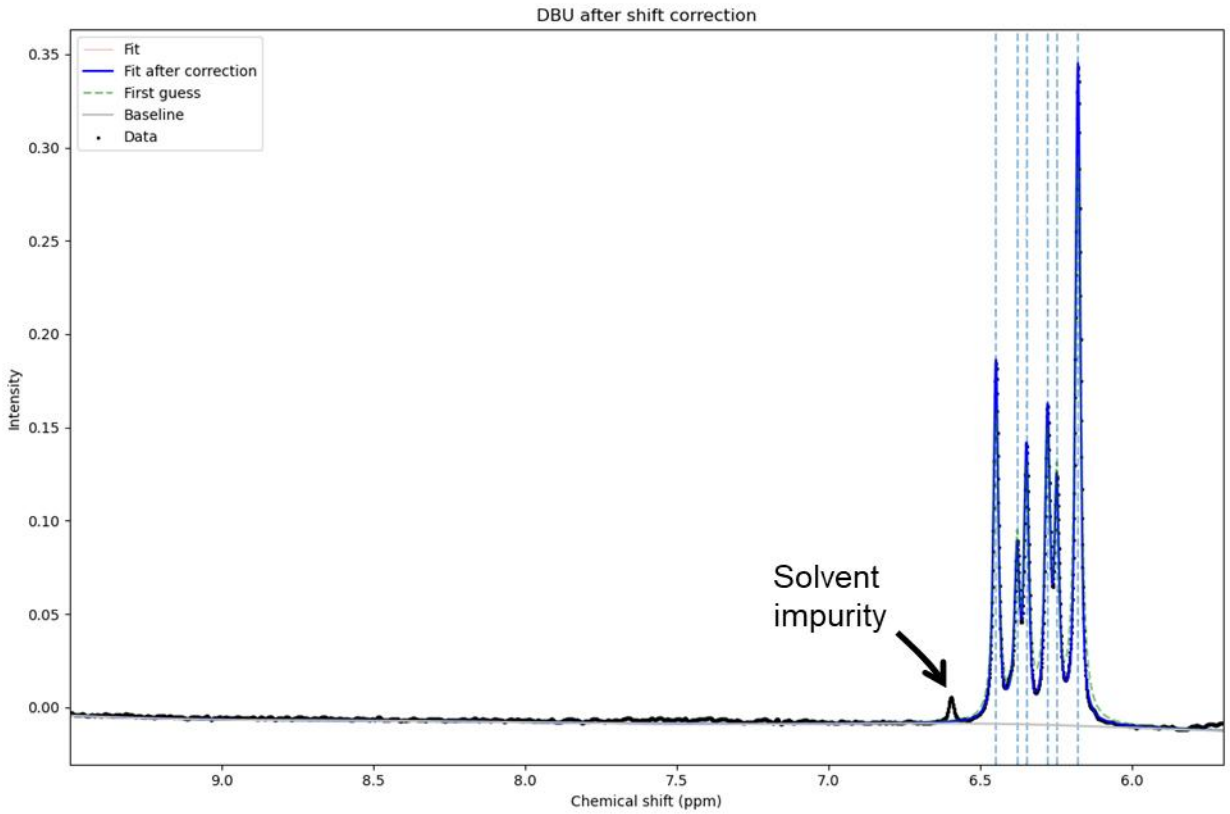


Figure S108. Application of the `remove` rule to the DBU reference. A solvent-impurity resonance near 6.54 ppm lies close to the DBU signals and could otherwise be misassigned as part of the DBU contribution. Excluding this narrow region prevents interference while preserving the fit of the remaining DBU resonances.

The `add` option allows additional peaks to be introduced manually by defining their position, width, and intensity. This is useful when a peak is not detected by the default `find_peaks` settings, for example because it has a low maximum intensity. In the 7-azaindole reference measured in the presence of copper, broad signals were observed, leading to low peak maxima. NMRanger detected a sharp peak around 7.45 ppm, but missed a broader signal around 8.3 ppm (**Figure S109**). Manually adding this peak during generation of the 7-azaindole model allowed the main features of the spectrum to be reproduced more faithfully before analysis of the crude mixtures. Note that the peak is added before the generation of the reference model so that its parameters (peak area, center, width and gaussian weight) are optimized at the same time than all the other peak already detected.

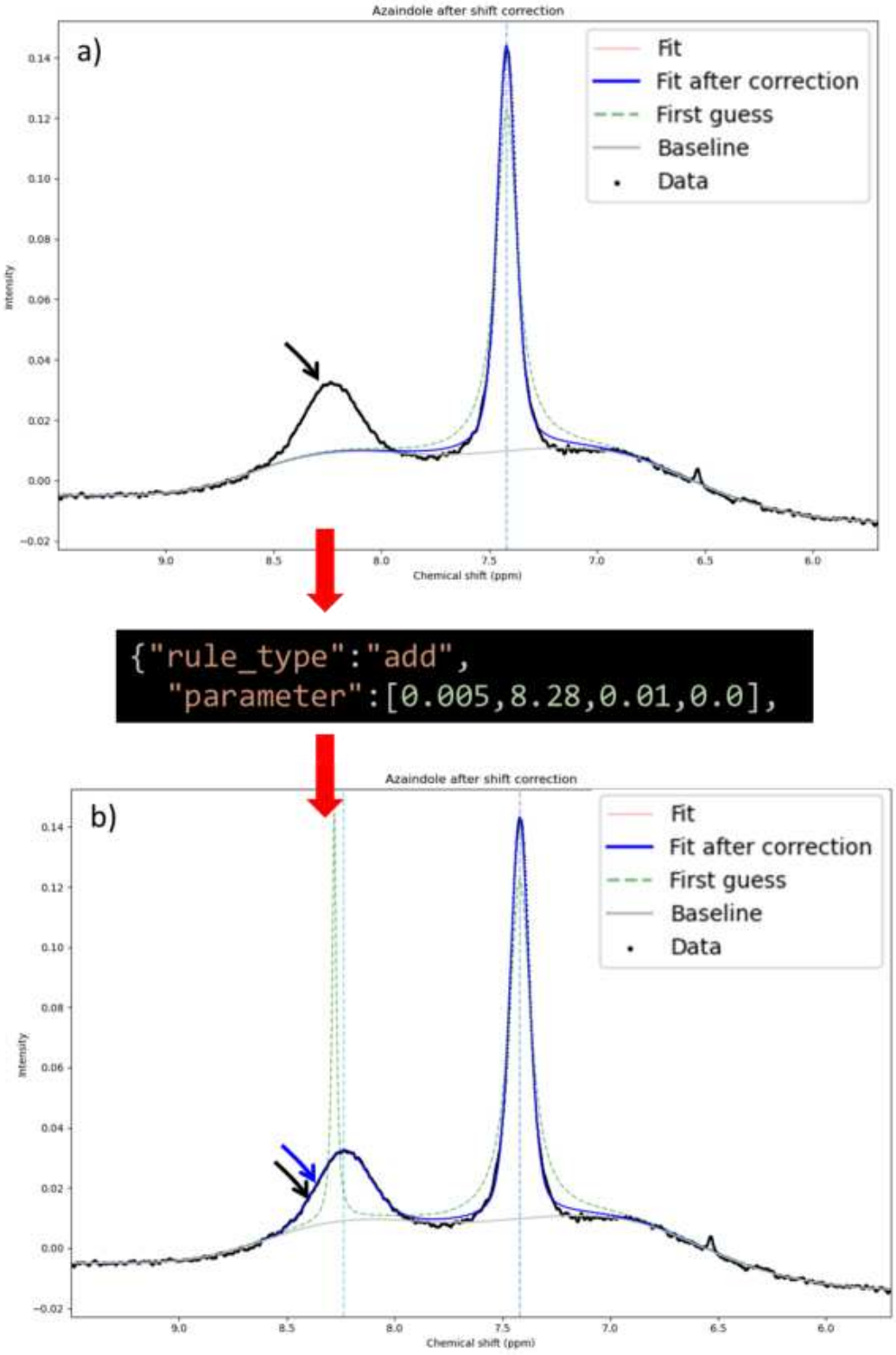


Figure S109. Application of the `add` and `weak_spline` rule to the 7-azaindole–copper reference. a) Before correction, the broad signal near 8.3 ppm is not represented by the initial reference model. **b)** Introduction of an additional peak at this position improves reconstruction of the missing resonance while retaining the fit of the main signal near 7.4 ppm.

The `baseline` option allows an alternative baseline correction to be applied to individual components. This provides additional local flexibility when the default baseline correction is not sufficient. For 7-azaindole, substantial line broadening produced baseline distortions that were not adequately described by the default settings. The `weak_spline` rule was therefore applied to fit a smooth, weakly constrained spline to better reproduce characteristic peaks of 7-azaindole in presence of copper (**Figure S109**)

The `offset` option is used to shift reference peaks within a defined ppm range. This correction is useful when a systematic chemical-shift difference is observed between a pure reference spectrum and the corresponding signals in crude reaction mixtures. In most cases, this correction is not required because NMRanger can natively account for moderate peak shifts during fitting. However, some special cases may show unusually large and reproducible shifts. One possible solution would be to increase the allowed peak-position tolerance, controlled by `tolerance_ppm`. However, increasing this tolerance for all components can make the identification of other analytes less reliable. The `offset` option provides a more targeted solution. If a specific analyte always shows a shift of similar magnitude, its expected peak positions can be corrected before fitting, without broadening the fitting boundaries for all other analytes. For example, during the analysis of crude mixtures containing $PhB(OH)_2$, its resonances were observed approximately **0.20 ppm downfield** compared with the pure reference spectrum. This systematic shift was therefore included as an `offset` correction in the NMRanger `special_treatment` parameters, as shown in **Figure S110**.

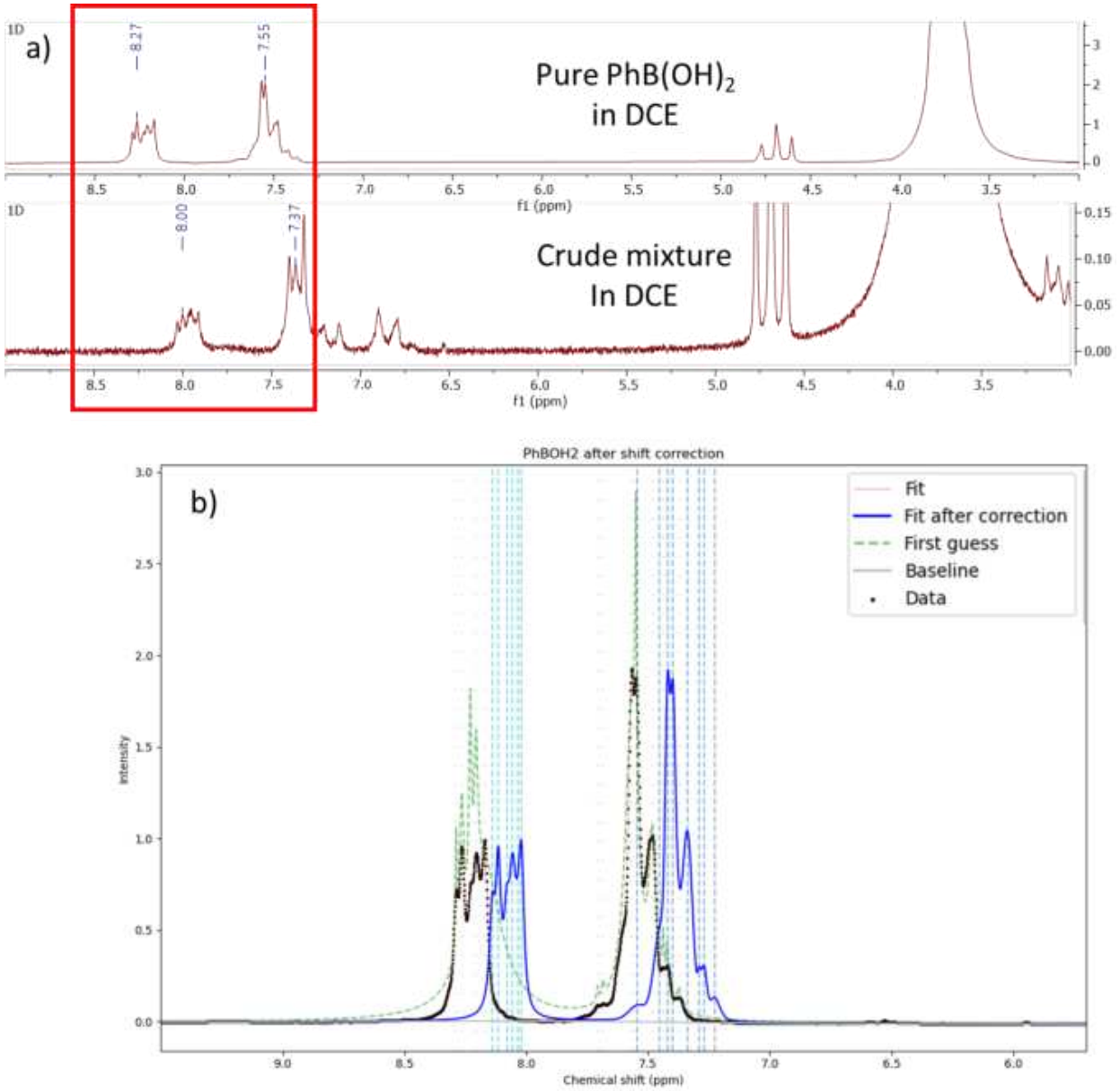


Figure S110. Application of the `offset` rule to the PhB(OH)$_2$. **a)** Comparison of pure PhB(OH)$_2$ in DCE with a crude reaction mixture, showing systematic shifts of the aromatic resonances from approximately 8.27 and 7.55 ppm in the pure reference to 8.00 and 7.37 ppm in the crude spectrum. **b)** Incorporation of the corresponding regional chemical-shift offset into the NMRanger special-treatment parameters improves alignment of the PhB(OH)$_2$ reference with the experimental spectrum and enables accurate fitting of the shifted resonances.

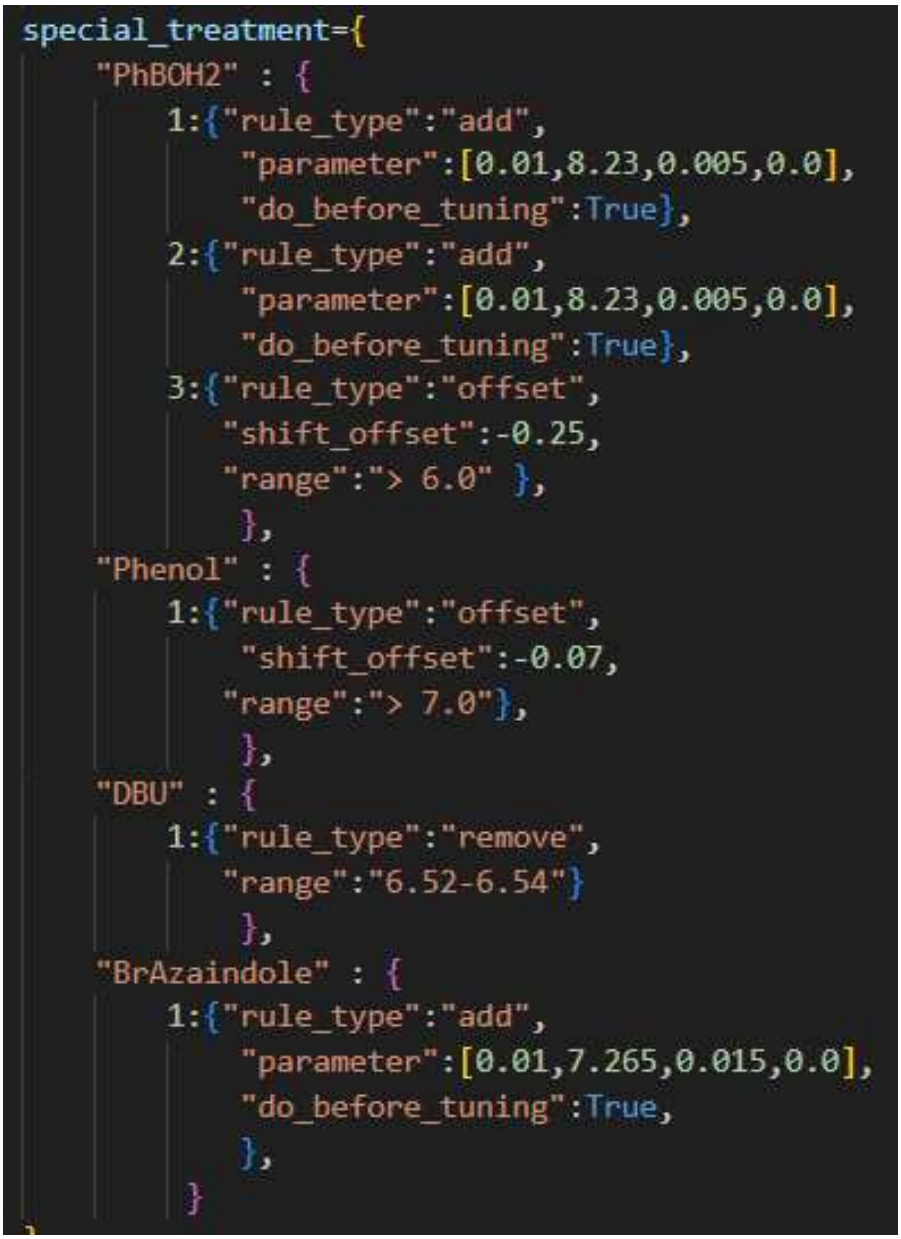
The complete set of substrate-specific corrections used for the 7-azaindole dataset is summarized in the parameter block shown below. These settings combine the `add`, `offset`, `remove`, and `baseline` rules described above and were applied before fitting the crude reaction spectra (**Figure S111**-**Figure S114**).

```
special_treatment={
    "PhB(OH)2" : {
        1:{"rule_type":"remove",
           "range":">8.5"
           },
        2:{"rule_type":"add",
           "parameter":[0.01,8.23,0.005,0.0],
           "do_before_tuning":True},
        3:{"rule_type":"add",
           "parameter":[0.01,8.23,0.005,0.0],
           "do_before_tuning":True},
        4:{"rule_type":"offset",
           "shift_offset":-0.20,
           "range":"> 6.0" },
           },
    "PhMorph" : {
        1:{"rule_type":"remove",
                "range":"6.5-6.6"
                },
        2:{"rule_type":"add",
           "parameter":[0.002,6.7,0.005,0.0],
           "do_before_tuning":True},
    },
    "Phenol" : {
        1:{"rule_type":"offset",
           "shift_offset":-0.07,
           "range":"> 7.0"
           },
```

Figure S111. NMRanger `special_treatment` settings used for morpholine reactions, including compound-specific `add`, `offset`, `remove`, and `baseline` rules. Screenshot of the parameter definition in VS Code.



Figure S112. NMRanger `special_treatment` settings used for 7-azaindole reactions, including compound-specific `add`, `offset`, `remove`, and `baseline` rules. Screenshot of the parameter definition in VS Code.



```
special_treatment={
    "PhBOH2" : {
        1:{"rule_type":"add",
           "parameter":[0.01,8.23,0.005,0.0],
           "do_before_tuning":True},
        2:{"rule_type":"add",
           "parameter":[0.01,8.23,0.005,0.0],
           "do_before_tuning":True},
        3:{"rule_type":"offset",
          "shift_offset":-0.25,
          "range":"> 6.0" },
           },
    "Phenol" : {
        1:{"rule_type":"offset",
           "shift_offset":-0.07,
          "range":"> 7.0"},
           },
    "DBU" : {
        1:{"rule_type":"remove",
          "range":"6.52-6.54"}
           },
    "BrAzaindole" : {
        1:{"rule_type":"add",
           "parameter":[0.01,7.265,0.015,0.0],
           "do_before_tuning":True,
           },
     }
}
```

Figure S113. NMRanger `special_treatment` settings used for 5-bromo-7-azaindole reactions, including compound-specific `add`, `offset`, `remove`, and `baseline` rules. Screenshot of the parameter definition in VS Code.



Figure S114. NMRanger `special_treatment` settings used for 1H-indazole reactions, including compound-specific `add`, `offset`, `remove`, and `baseline` rules. Screenshot of the parameter definition in VS Code.

5. Robowski-NMR Acquisition Parameters Selection

5.1. T_1/T_2 relaxation measurements

Preliminary T_1/T_2 relaxation experiments were performed on representative reference coupling product of 7-azaindole and the internal standard to optimize the acquisition parameters for automated ^1H NMR quantification. T_1 measurements were used to determine how long the magnetization required to recover between scans and to select an appropriate repetition time for quantitative signal comparison. The repetition time was chosen using the standard quantitative NMR (qNMR) criterion that the delay between scans should be approximately five times the longest measured T_1 ($5\times$ T_1), allowing near-complete magnetization recovery (>99%). T_2 measurements were used to evaluate signal decay, linewidth behavior, and the acquisition time required to capture the FID decay. The acquisition time was selected using the practical criterion that the acquisition window should be approximately $3 \times T_2$, allowing most of the FID decay to be captured while avoiding unnecessarily long acquisition of noise.[11–13]

For N-phenyl-7-azaindole, the measured T_1 values among the selected aromatic signals were 2.71–3.11 s; therefore, a 15 s repetition time was selected, corresponding to approximately $5 \times T_1$. Based on the standard qNMR recovery criterion, this repetition time allows ca. 99% recovery of the longitudinal magnetization before the next scan.[11] The measured T_2 values were 1.63–2.57 s, giving an approximate $3 \times T_2$ range of 4.9–7.7 s. The selected 6.4 s acquisition time falls within this range. The final Robowski-NMR acquisition parameters were set to 32 scans, a 6.4 s acquisition time, a 15 s repetition time, and a 90° pulse angle, corresponding to approximately 8 min per sample (**Figure S115**).

T_1/T_2 relaxation measurements were also performed for the internal standard, 1,3,5-trimethoxybenzene. The measured relaxation times were T_1 = 2.79 s and T_2 = 2.61 s, which are within the range measured for N-phenyl-7-azaindole. Therefore, the selected 15 s repetition time and 6.4 s acquisition time were also suitable for the internal standard under the same Robowski-NMR acquisition conditions (**Figure S116**).

Because both the representative 7-azaindole coupling product and the internal standard showed compatible relaxation behavior, the same acquisition parameters were applied to the 5-bromo-7-

azaindole and indazole substrate classes. This is appropriate because the quantified signals for these substrates were primarily aromatic protons, which are expected to have broadly comparable relaxation behavior across these structurally related compounds. In addition, greater signal overlap and the presence of multiple aromatic components in these substrates required more spectral information for reliable fitting, supporting the use of a 90° pulse angle.

The 30° pulse condition was retained for morpholine because the spectra were relatively simple: N-phenylmorpholine is symmetric and contains fewer distinct signals. Although a 90° pulse angle could also have been used, the 30° pulse angle allowed shorter measurements while still providing sufficient signal quality for reliable quantification. Details of the morpholine-specific acquisition settings are discussed in the next section.

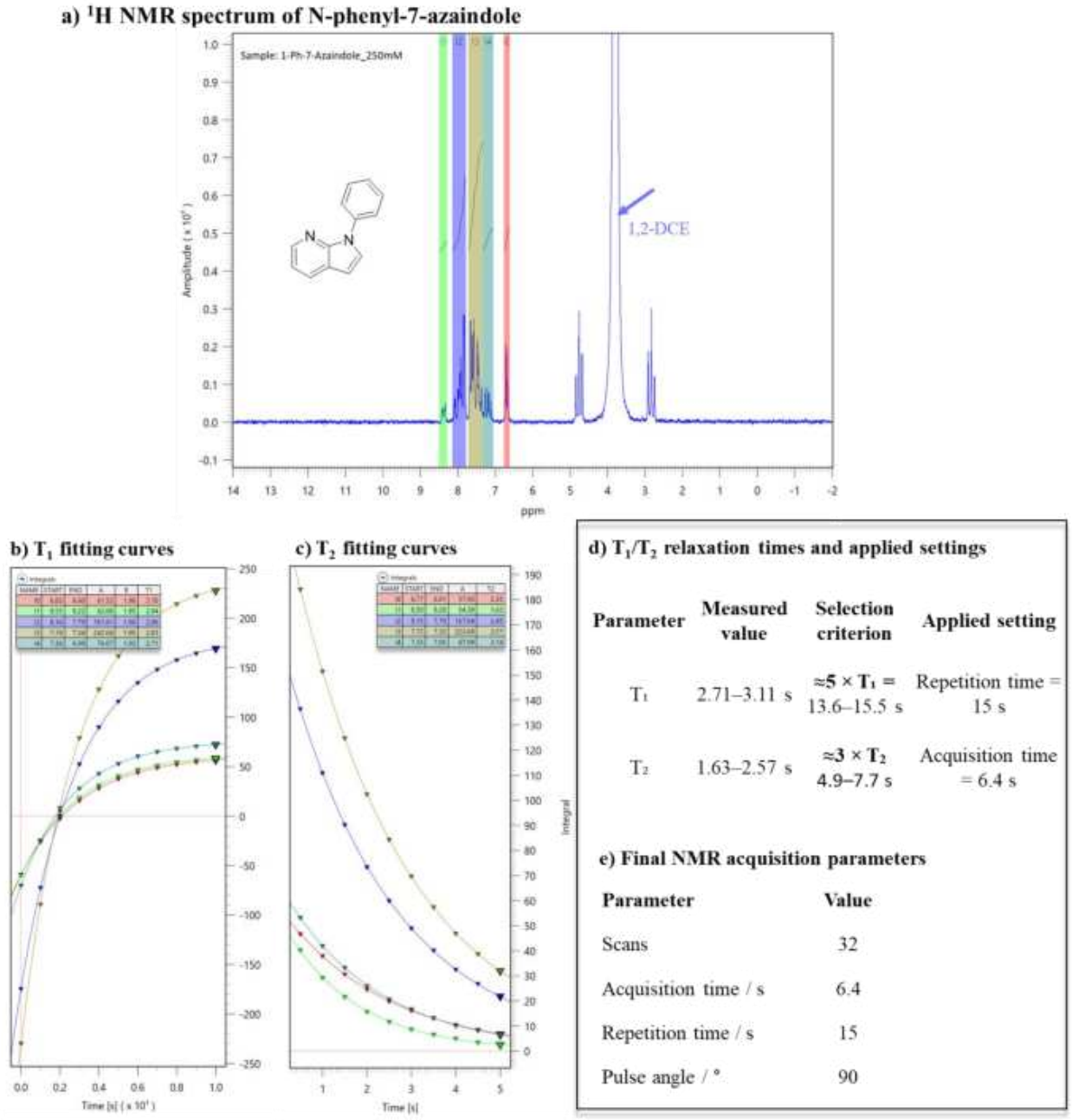
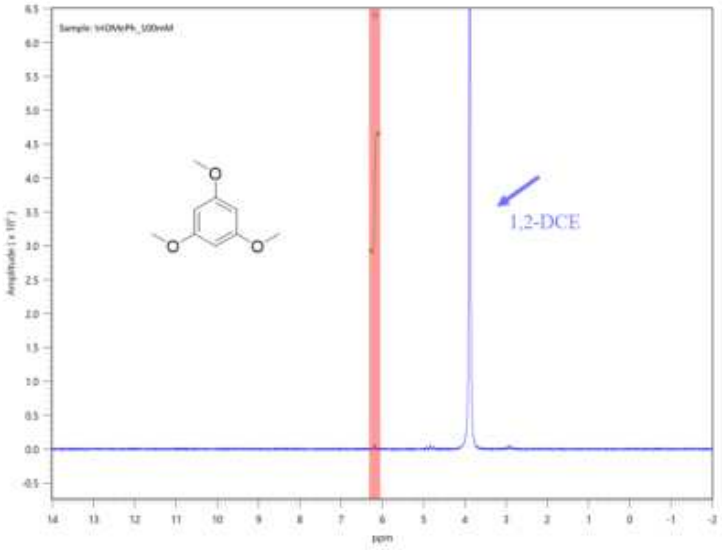
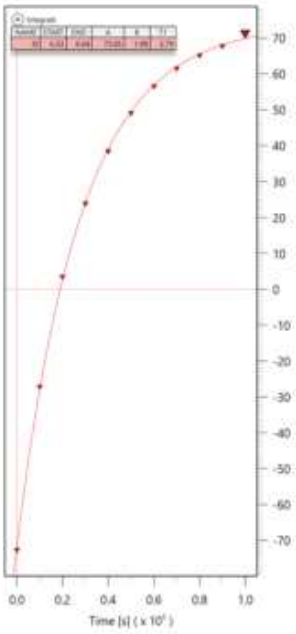


Figure S115. T_1/T_2 relaxation measurements for N-phenyl-7-azaindole used to select Robowski-NMR acquisition parameters. a) Full ^{1}H NMR spectrum of N-phenyl-7-azaindole, with the aromatic signals used for relaxation analysis highlighted. **b)** T_1 fitting curves obtained from the highlighted signal integrals listed in the inset table. **c)** fitting curves obtained from the highlighted signal integrals listed in the inset table. **d)** Summary of measured relaxation times, selection criteria, and applied acquisition settings. The measured relaxation times were T_1 = 2.71–3.11 s and T_2 = 1.63–2.57 s. The selected repetition time of 15 s corresponds to approximately 5 × T_1, while the selected acquisition time of 6.4 s is close to the practical ≈3 × T_2 criterion, supporting its use for automated Robowski-NMR measurements. **e)** Final Robowski-NMR acquisition parameters used for automated measurements.

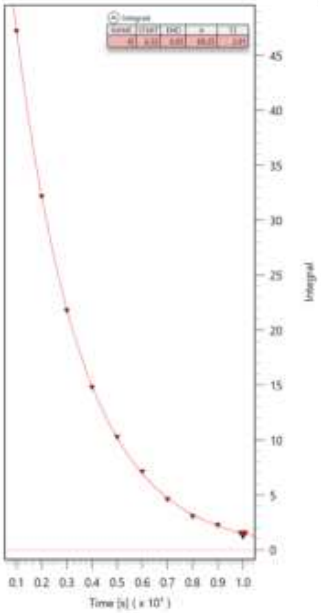
a) ^{1}H NMR spectrum of N-phenyl-7-azaindole



b) T_1 fitting curves



c) T_2 fitting curves



d) T_1/T_2 relaxation times and applied settings

Parameter	Measured value	Selection criterion	Applied setting
T_1	2.79 s	≈5 × T_1 = 14.0 s	Repetition time = 15 s
T_2	2.61 s	≈3 × T_2 7.8 s	Acquisition time = 6.4 s

e) Final NMR acquisition parameters

Parameter	Value
Scans	32
Acquisition time / s	6.4
Repetition time / s	15
Pulse angle / °	90

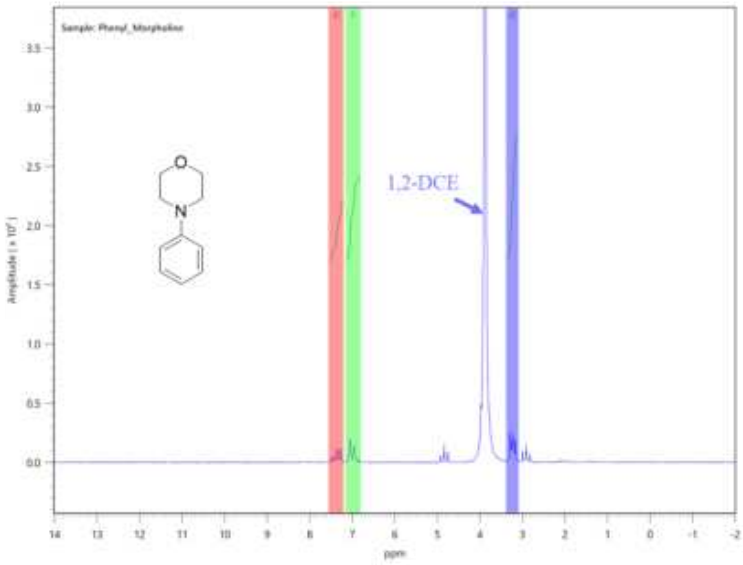
Figure S116. T_1/T_2 relaxation measurements for N-phenyl-7-azaindole used to select Robowski-NMR acquisition parameters. a) Full ^{1}H NMR spectrum of 1,3,5-trimethoxybenzene, with the signal used for relaxation analysis highlighted. **b)** T_1 fitting curves obtained from the highlighted signal integrals listed in the inset table. **c)** T_2 fitting curve obtained from the highlighted signal integral listed in the inset table. **d)** The measured relaxation times were T_1 = 2.79 s and T_2 = 2.61 s. The 15 s repetition time satisfies the approximate 5 × T_1 criterion, and the 6.4 s acquisition time is close to the practical ≈3 × T_2 criterion, supporting the use of these acquisition parameters for Robowski-NMR quantification with 1,3,5-trimethoxybenzene as the internal standard. **e)** Final Robowski-NMR acquisition parameters used for automated measurements.

For *N*-phenylmorpholine, the morpholine reaction plates were first measured using the initial Robowski-NMR acquisition settings of 32 scans, 6.4 s acquisition time, 10 s repetition time, and a 30° pulse angle. After these measurements, T_1/T_2 relaxation experiments were performed on the N-phenylmorpholine reference compound to evaluate whether these acquisition parameters were suitable for quantitative analysis.

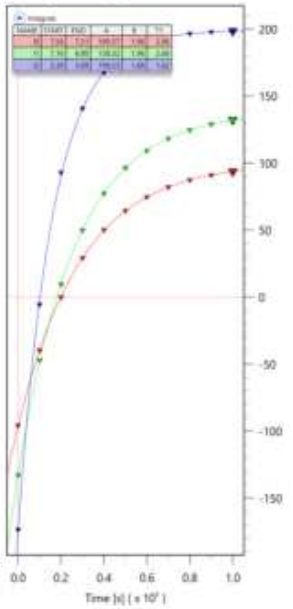
The measured T_1 values among the selected signals were 1.62–2.96 s, while the measured T_2 values were 1.62–2.66 s. The 10 s repetition time used for the morpholine reaction plates is shorter than 5 × the longest measured T_1; the use of a 30° pulse angle reduces saturation effects compared with a 90° pulse and allows faster acquisition.[11] In addition, the calibration/reference spectra and the reaction spectra were acquired using the same acquisition parameters. Therefore, the calibration accounted for the selected repetition time, acquisition time, and pulse angle under matched measurement conditions.

The measured T_2 values give an approximate 3 × T_2 range of 4.8–8.0 s, and the selected 6.4 s acquisition time falls within this range, supporting sufficient FID acquisition. Together, these results indicate that the initial Robowski-NMR settings used for the morpholine reactions were suitable for automated quantification while enabling shorter measurement times (~5 min per sample). This was further supported by the agreement between NMR-based quantification and isolated-yield validation for selected morpholine reaction conditions (**Figure S117**).

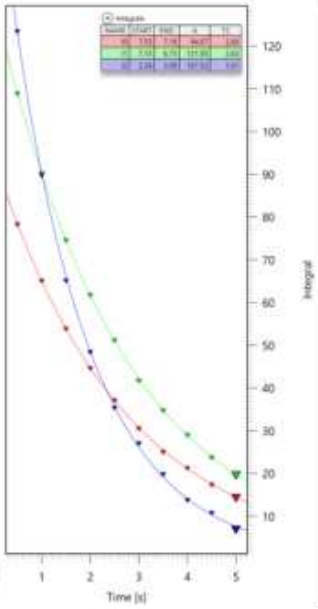
a) ^{1}H NMR spectrum of N-phenyl-morpholine



b) T_1 fitting curves



c) T_2 fitting curves



d) T_1/T_2 relaxation times and applied settings

Parameter	Measured value	Selection criterion	Applied setting
T_1	1.62–2.96 s	≈5 × T_1 = 8.1–14.8 s	Repetition time = 15 s
T_2	1.62–2.66 s	≈3 × T_2 4.8–8.0 s	Acquisition time = 6.4 s

e) Final NMR acquisition parameters

Parameter	Value
Scans	32
Acquisition time / s	6.4
Repetition time / s	10
Pulse angle / °	30

Figure S117. T_1/T_2 relaxation measurements for N-phenyl-7-azaindole used to select Robowski-NMR acquisition parameters. a) Full ^{1}H NMR spectrum of N-phenylmorpholine, with the signals used for relaxation analysis are highlighted. **b)** T_1 fitting curves obtained from the highlighted signal integrals listed in the inset table. **c)** The fitting curves obtained from the highlighted signal integrals listed in the inset table. **d)** Summary of measured relaxation times, selection criteria, and applied acquisition settings. The measured relaxation times were T_1 = 1.62–2.96 s and T_2 = 1.62–2.66 s. The selected acquisition time of 6.4 s falls within the practical ≈3 × T_2 range, supporting its use for automated Robowski-NMR measurements. **e)** Final Robowski-NMR acquisition parameters used for automated measurements.

6. References

1. Jia, Y.; Frydrych, R.; Sobolev, Y. I.; Wong, W.-S.; Prajapati, B.; Matuszczyk, D.; Bilgi, Y.; Gadina, L.; Ahumada, J. C.; Moldagulov, G. Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks. *Nature* **2025**, *645* (8082), 922–931. https://doi.org/10.1038/s41586-025-09490-1
2. Quach, T. D.; Batey, R. A. Ligand- and Base-Free Copper(II)-Catalyzed C−N Bond Formation: Cross-Coupling Reactions of Organoboron Compounds with Aliphatic Amines and Anilines. *Org. Lett.* **2003**, *5* (23), 4397–4400. https://doi.org/10.1021/ol035681s.
3. Mondal, K.; Mukhopadhyay, N.; Sengupta, A.; Roy, T.; Das, P. Exploiting Coordination Behavior of 7-Azaindole for Mechanistic Investigation of Chan-Lam Coupling and Application to 7-Azaindole Based Pharmacophores. *Chem. – Eur. J.* **2023**, *29* (16), e202203718. https://doi.org/10.1002/chem.202203718.
4. Banerjee, S.; Mishra, M.; Punniyamurthy, T. Copper-Catalyzed C7-Selective C–H/N–H Cross-Dehydrogenative Coupling of Indolines with Sulfoximines. *Org. Lett.* **2022**, *24* (43), 7997–8001. https://doi.org/10.1021/acs.orglett.2c03190.
5. Lam, P. Y. S.; Clark, C. G.; Saubern, S.; Adams, J.; Winters, M. P.; Chan, D. M. T.; Combs, A. New Aryl/Heteroaryl C□N Bond Cross-Coupling Reactions via Arylboronic Acid/Cupric Acetate Arylation. *Tetrahedron Lett.* **1998**, *39* (19), 2941–2944. https://doi.org/10.1016/S0040-4039(98)00504-8.
6. Koerner, T.; Brown, R. S.; Gainsforth, J. L.; Klobukowski, M. Electrophilic Bromination of Ethylene and Ethylene- d_4 . A Combined Experimental and Theoretical Study. *J. Am. Chem. Soc.* **1998**, *120* (23), 5628–5636. https://doi.org/10.1021/ja974151m.
7. Banwell, C. N.; Sheppard, N. High Resolution Nuclear-Magnetic-Resonance Spectra of Hydrocarbon Groupings: IV. ABC Spectra of Vinyl Compounds. *Mol. Phys.* **1960**, *3* (4), 351–369. https://doi.org/10.1080/00268976000100411.
8. Watts, V. S.; Goldstein, J. H. Dependence of Some Ethylenic *J* Gem Values on Solvent and Concentration. *J. Chem. Phys.* **1965**, *42* (1), 228–233. https://doi.org/10.1063/1.1695680.
9. Castillo, A. M.; Patiny, L.; Wist, J. Fast and Accurate Algorithm for the Simulation of NMR Spectra of Large Spin Systems. *J. Magn. Reson.* **2011**, *209* (2), 123–130. https://doi.org/10.1016/j.jmr.2010.12.008.
10. Jia, Y.; Wong, W.-S.; Bilgi-Gadina, Y.; Gadina, L. P. J.-L.; Kwong, K.-L.; Jiang, Y.; Vagkidis, N.; Marazzi, B.; Schreiner, P. R.; Grzybowski, B. A. From Berzelius to Hyperspace: Previously Unrecognized Network of Reactivity in Textbook Brominations. *Angew. Chem. Int. Ed.* ***2026***. *Accepted.*
11. Mestrelab Research. *qNMR: the best pulse length for acquisition Home » Resource » qNMR: the best pulse length for acquisition*. https://mestrelab.com/resources/qnmr-the-best-pulse-length-for-acquisition-2.html.
12. Claridge, T. D. *High-Resolution NMR Techniques in Organic Chemistry*; Elsevier, 2016; Vol. 27.
13. Amin, N.; Claridge, T. Quantitative NMR Spectroscopy. University of Oxford, 2017. https://nmr.chem.ox.ac.uk/files/quantitativenmrpdf.