DISSERTATION

Advancing Multiscale Modelling to Describe Chemical Reactions Under Experimental Conditions

by

Radu Alexandru Talmazan

Mat. Nr. 01018433

to obtain the degree

Doctor rerum naturalium (Dr. rer. nat.)

Under the supervision of

Univ. Prof. Dr. Georg Kent Hellerup Madsen

Assistant Prof. Dr. Maren Podewitz

Faculty of Technical Chemistry

Institute of Materials Chemistry

Chemical Selectivity and Reactivity

Wien, 2024

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass die vorliegende Arbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen von mir selbstständig erstellt wurde. Alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, sind in dieser Arbeit genannt und aufgelistet. Die aus den Quellen wörtlich entnommenen Stellen, sind als solche kenntlich gemacht. Das Thema dieser Arbeit wurde von mir bisher weder im In- noch Ausland einer Beurteilerin/einem Beurteiler zur Begutachtung in irgendeiner Form als Prüfungsarbeit vorgelegt. Diese Arbeit stimmt mit der von den Begutachterinnen/Begutachtern beurteilten Arbeit überein.

Acknowledgement

I find myself at the end of a journey years in the making and it only seems fit to thank my companions, friends, loved ones and all other individuals who have shaped my path. Their unwavering support, encouragement, occasional pressure, and wisdom have motivated me to pull through any challenges that arose.

To my advisor, Assistant Prof. Dr. Maren Podewitz, thank you for being more than a mentor. You have been my guiding light through the vast ocean that is computational chemistry. Your patience, brilliance, and belief in my potential were more than I could ever ask for. Thank you for accepting me under you wing and allowing me so much freedom in my work. I hope I could at least partially repay you through the ever-present coffee and cookies.

The lab/office was my second (maybe third) home. Martin Mähr, who is positively bussin' and never caps, thank you for opening my eye to the wonders of Cathedrals, wherever they may be. Lukas Magenheim, whose resilience and dedication never cease to impress—thank you for showing me the wonders of Steinrigl. Jan Pecak, the one who took refuge from the wet lab, thank you for being a great friend and welcoming me to the TU Wien. To all other members of the lab (Paul, Hannah, Jorge – yes, you are one of us now), it was a pleasure to have met you and thanks for making work fun. Science is better when shared.

To Associate Prof. Dr. Thomas Hofer and Prof. Dr. Christopher Stein, your scientific insight and ability never ceases to amaze me. Thank you for setting an example for me to strive for.

To Univ. Prof. Dr. Georg Kent Hellerup Madsen, thank you for welcoming me at the TU Wien and helping me get settled in, as well as always stepping up whenever I needed assistance.

To Univ.Prof. Dipl.-Ing. Dr.techn. Karl Kirchner and Prof. Dr. Ivan Castillo Pérez for their invaluable experimental insights and the desire to work together in solving chemical problems.

To my family, especially my parents, who helped me follow my passion and were with me every step of the way, including through my occasional my rants about p-values and frustrations. I could not have done this without you.

To my sister, who patiently texted me daily to e-mail people, thank you for your dedication in pushing me to achieve.

To Leo, thank you for your staying by my side and encouraging me to pursue my goals, and especially for the many corrections of various weird and unmotivated abstracts and texts.

To Hannes, you know what you did…

To Wang Yin, your ideas are a breath of fresh air, and the fun never ends when we are together.

In this acknowledgment, I weave together words of gratitude, but they cannot fully capture the depth of my appreciation. To everyone who played a part—thank you.

With heartfelt thanks,

Radu

Table of Contents

1. Preface

This thesis is written in a cumulative fashion, with a general introduction chapter, followed by a method section which covers the methodology used in all the publications contained within this thesis. The manuscripts of the publications are located in the results and discussion sections. A conclusion summarises the findings and sets an outlook for future research.

2. Abstract

The elucidation of reaction mechanisms is a key field of research in computational chemistry. To obtain an accurate representation of the behaviour of molecules there are two main considerations that need to be accounted for. Firstly, the computational model, namely the underlying theory used to describe the system, must be chosen to be as exact as possible. Secondly, the chemical model, which consists of the molecule under study and the environment surrounding it, must closely mimic the experimental conditions. This also means that the flexibility and dynamic nature needs to be considered in mechanistic investigations, particularly for flexible systems. While computational methods have been steadily improving, allowing for larger systems to be tackled, an improvement in the chemical model is essential to achieve an operando model of chemical reactions. While full first-principle models are not feasible yet, within this thesis, we developed tailored multiscale description to capture the reactivity and reaction mechanism of (transition-)metal containing systems: The first step towards a more precise model is the generation of solution structures as explicit solvent molecules may participate in reactions or aids in the stabilization of large molecules. Yet the set-up of such calculations required many manual steps and expert knowledge. To this end, we developed the PyConSolv package that streamlines the generation of parameters for MM force-fields for metal- and nonmetal-containing complexes. This tool allows the exploration of the potential energy surface using MM molecular dynamics (MD) and, subsequently, the generation of conformers in explicit solvent. It facilitates the description of catalytic transition-metal complexes which require considerable modelling effort, such as Cu-Calix[8]arene, and led to more realistic results/structures. The ability to evaluate reactivity in explicit solvent allowed us to investigate the different reactivity of two Cu-calixa[8]arene stereoisomers and to identify the cause of the poor catalytic activity of one of them. It also provided an opportunity to evaluate the effect of the macrocyclic calixarene on the C-N coupling reaction mechanism. In an effort to evaluate the dynamic effects on the C-N bond formation, we developed a QM/MM MD simulation protocol for evaluation in explicit solvent. Our study is one of the first to address dynamics in non-aqueous solvent and to achieve sufficient sampling - it therefore represents a significant step towards mimicking the exact chemical environment found in experiments. To extract chemically relevant information from the vast simulation data required a computer-assisted evaluation. To obtain results that can be understood by any chemist, we turned to interpretable machine learning techniques (such as Lasso, Random Forest, and Logistic Regression) within a consensus model. Combined with dimensionality reduction methods (PCA, LDA, and tICA) and a Granger Causality model, we determined and quantified the reaction coordinate for the C-N coupling reaction when described under experimental conditions, that is at finite temperature and pressure including dynamics. By demonstrating the advantages of setting up tailored multiscale approaches for computational chemical investigations, as well as the development of tools to lower the entry to these methods, we push the chemical model a step closer to a digital twin of the experiment.

3. Abstract

Die Aufklärung von Reaktionsmechanismen ist ein zentrales Forschungsgebiet der computergestützten Chemie. Um eine genaue Darstellung des Verhaltens von Molekülen zu erhalten, müssen zwei wesentliche Aspekte berücksichtigt werden. Erstens muss das Rechenmodell, d. h. die zugrundeliegende Theorie zur Beschreibung des Systems, so genau wie möglich gewählt werden. Zweitens muss das chemische Modell, bestehend aus dem zu untersuchenden Molekül und seiner Umgebung, die experimentellen Bedingungen genau nachbilden. Dies hat zur Folge, dass Flexibilität und Dynamik bei mechanistischen Untersuchungen, insbesondere bei nicht-rigiden Systemen mit einbezogen werden müssen. Während die stetige Weiterentwicklung der Berechnungsmethoden heutzutage die Modellierung großer Systeme erlaubt, so ist eine Verbesserung des chemischen Modells ebenso unerlässlich, um ein operando-Modell zur Beschreibung chemischer Reaktionen zu erhalten. Da umfassende First-Principle-Modelle derzeit noch nicht realisierbar sind, wurden im Rahmen dieser Doktorarbeit maßgeschneiderte Multiskalenbeschreibungen entwickelt, um Reaktivität und Reaktionsmechanismen (Übergangs-)metallhaltiger Systeme zu charakterisieren: Der erste Schritt zu einem genaueren Modell ist dabei die Generierung von Strukturen im Lösungsmittel, denn explizite Lösungsmittelmoleküle können an Reaktionen beteiligt sein oder zur Stabilisierung großer Moleküle beitragen. Die Durchführung solcher Berechnungen erforderte jedoch viele manuelle Schritte und Expertenwissen. Um dies zu vermeiden, wurde die Software PyConSolv entwickelt, die die Generierung von MM Kraftfeldparametern für metall- und nichtmetallhaltige Komplexe vereinfacht. Dieses Tool ermöglicht nicht nur die Erkundung von potenziellen Energieoberfläche mittels MM-Molekulardynamik (MD), sondern auch die Erzeugung von Konformeren im expliziten Lösungsmittel. Somit können nun auch für große katalytische Übergangsmetallkomplexe, wie z. B. Cu-Calix[8]arene, effizient realistische Lösungsmittelstrukturen erhalten werden. Diese Methodik ermöglicht ebenso die Untersuchung der unterschiedlichen Reaktivitäten zweier Cu-Calixa[8]aren-Stereoisomere und die Ermittlung der Faktoren, die zu einer schlechte katalytische Aktivität von einem der beiden Isomere führt. Darüber hinaus kann nun bewertet werden, inwiefern das makrozyklische Calixaren den Reaktionsmechanismus des Cu-Calix[8]aren Katalysators, hier der C-N Kopplung, beeinflusst. Um die Auswirkungen dynamischer Effekte auf die Bildung der C-N-Bindung zu evaluieren, wurde ein QM/MM-MD-Simulationsprotokoll zur Beschreibung des Systems im expliziten Lösungsmittel entwickelt. Diese Studie ist eine der ersten, die sich mit der Dynamik in nichtwässrigen Lösungsmitteln befasst und eine ausreichende Abtastrate der Potentialfläche erreicht. Dies ist ein wichtiger Schritt, um die exakte chemische Umgebung in Computerexperimenten nachzubilden, und ein Meilenstein für die Entwicklung von operando-Modellen. Um chemisch relevante Informationen aus den umfangreichen Simulationsdaten zu extrahieren, ist eine computergestützte Auswertung erforderlich. Zur Auswertung wurden Methoden des maschinellen Lernens (wie Lasso, Random Forest und logistische Regression) innerhalb eines Konsensmodells angewandt, die eine chemische Interpretierbarkeit der Resultate ermöglichen und somit für jede:n Chemiker:in verständlich sind. In Kombination mit Methoden zur Dimensionalitätsreduktion (PCA, LDA und tICA) und einem Granger-Kausalitätsmodell konnte so die Reaktionskoordinate des C-N-Kopplungsschrittes unter Berücksichtigung der experimentellen Bedingungen, d.h. bei Standardtemperatur und unter Einbeziehung dynamischer Effekte, bestimmt und quantifiziert werden. Durch die Demonstration der Vorteile maßgeschneiderter Multiskalenansätze für chemische Berechnungen und durch die Entwicklung von Werkzeugen zur Erleichterung des Zugangs zu diesen Methoden, bringen wir das chemische Modell einen Schritt näher an den digitalen Zwilling des Experiments.

4. Included Publications

Published:

*** Denotes shared first authorship**

1) **Talmazan, R. A.**; Podewitz, M. PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent. J. Chem. Inf. Model. 2023, 63 (17), 5400–5407. https://doi.org/10.1021/acs.jcim.3c00798.
2) **Talmazan, R. A***.; Refugio Monroy, J.*; del Río-Portilla, F.; Castillo, I.; Podewitz, M. Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]Arene Supramolecular Catalyst. ChemCatChem 2022, 14 (20). https://doi.org/10.1002/cctc.202200662.
3) Berlanga-Vázquez, A.*; **Talmazan, R. A.***; Reyes-Mata, C. A.; Percástegui, E. G.; Flores-Alamo, M.; Podewitz, M.; Castillo, I. Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]Arene C−S Coupling. Eur. J. Inorg. Chem. 2022. https://doi.org/10.1002/ejic.202200596.
4) Himmelbauer, D.*; **Talmazan, R., A.***; Weber, S.*; Pecak, J.*; Thun-Hohenstein, A.; Geissler, M.-S.; Pachmann, L.; Pignitter, M.; Podewitz, M.; Kirchner, K. No Transition Metals Required – Oxygen Promoted Synthesis of Imines from Primary Alcohols and Amines under Ambient Conditions. Chem. – Eur. J. 2023, 29 (29), e202300094. https://doi.org/10.1002/chem.202300094.

Submitted:

5) **Talmazan, R., A.**; Gamper J.; Castillo, I.; Hofer, T., S.; Podewitz, M. From Data to Chemistry: Revealing Causality and Reaction Coordinates through Interpretable Machine Learning in Supramolecular Transition Metal Catalysis, https://doi.org/10.26434/chemrxiv-2024-nd20j

Other Publications:

6) Pecak, J.; **Talmazan, R. A.**; Svatunek, D.; Kirchner, K.; Podewitz, M. Is Mn(I) More Promising Than Fe(II)–A Comparison of Mn vs Fe Complexes for Olefin Metathesis, Organometallics 2024 43 (4), 457-466 https://doi.org/10.1021/acs.organomet.3c00398
7) Goetzfried, S. K.; Gallati, C. M.; Cziferszky, M.; **Talmazan, R. A.**; Wurst, K.; Liedl, K. R.; Podewitz, M.; Gust, R. N-Heterocyclic Carbene Gold(I) Complexes: Mechanism of the Ligand Scrambling Reaction and Their Oxidation to Gold(III) in Aqueous Solutions. Inorg. Chem. 2020, 59 (20), 15312–15323. https://doi.org/10.1021/acs.inorgchem.0c02298.
8) **Talmazan, R. A.**; Liedl, K. R.; Kräutler, B.; Podewitz, M. The Intermolecular Anthracene-Transfer in a Regiospecific Antipodal C60 Difunctionalization. Org. Biomol. Chem. 2020, 18 (21), 4090–4103. https://doi.org/10.1039/D0OB00520G.
9) Spathopoulou, A.; Sauerwein, G. A.; Marteau, V.; Podlesnic, M.; Lindlbauer, T.; Kipura, T.; Hotze, M.; Gabassi, E.; Kruszewski, K.; Koskuvi, M.; Réthelyi, J. M.; Apáti, Á.; Conti, L.; Ku, M.; Koal, T.; Müller, U.; **Talmazan, R. A.**; Ojansuu, I.; Vaurio, O.; Lähteenvuo, M.; Lehtonen, Š.; Mertens, J.; Kwiatkowski, M.; Günther, K.; Tiihonen, J.; Koistinaho, J.; Trajanoski, Z.; Edenhofer, F. Integrative Metabolomics-Genomics Analysis Identifies Key Networks in a Stem Cell-Based Model of Schizophrenia. Mol. Psychiatry 2024. https://doi.org/10.1038/s41380-024-02568-8.

5. Statement of Contribution

1) Radu A. Talmazan (RAT) is the sole first author and his contributions include the whole computational work for the study and the programming of the PyConSolv python package. Maren Podewitz (MP) alongside RAT developed the concept and was involved in the writing of the manuscript.
2) RAT is a shared first author and his contributions include the entire computational work and analysis for the study. Refugio J. Monroy, F. del Río-Portilla and I. Castillo (IC) devised and performed the experimental studies. MP devised the computational concept. All authors contributed to the writing and editing of the manuscript.
3) RAT is a shared first author and his contributions include the whole computational work for the study. A. Berlanga-Vázquez, C. A. Reyes-Mata, E. G. Percástegui, M. Flores-Alamo and IC devised and performed the experimental studies. MP and RAT devised the computational concept and analysed the results. All authors contributed to the writing and editing of the manuscript.
4) RAT is a shared first author and his contributions include the whole computational work for the study. D. Himmelbauer and K. Kirchner provided the initial idea and devised the initial experiments. S. Weber and J. Pecak (JP) performed the in-depth experimental studies with help from A. Thun-Hohenstein, M.S. Geissler, L. Pachmann and M.Pignitter. MP, JP, and RAT devised the computational concept. All authors contributed to the writing and editing of the manuscript.
5) The project was conceived by RAT and MP, while IC provided chemical input for the system, required to devise the project. RAT performed the QM/MM/MD calculations together with Thomas S. Hofer (TSH), while Jakob Gamper implemented the rescaling barostat into the QMCFC package, specifically for this project. RAT devised the Jupyter Notebook to conduct the analyses. Analyses were edited by MP. RAT wrote the original draft, MP, TSH and IC edited the manuscript.

6. Introduction

Chemical reactions lie at the heart of countless natural and synthetic processes, from biological transformations to industrial catalysis. Understanding the underlying mechanisms and predicting their outcomes is essential for advancing science and technology. However, the intrinsic complexity of chemical reactions poses significant challenges, both from experimental and computational points of view.

In conjunction with experimental advancements, computational chemistry has played a pivotal role in designing and understanding catalytic systems by unravelling reaction mechanisms.[1–3] However, despite significant efforts, quantum chemistry still faces limitations in its predictive capabilities.[1] As the systems under investigation become increasingly complex, a need arises to enhance not only the underlying electronic structure theory employed for computations, but also the overall chemical model describing the system in its environment.[4] In fact, errors stemming from overly simplistic chemical models can often surpass those arising from approximate theoretical methodologies like density functional theory (DFT).[1,5] Consequently, the ultimate objective is to create a chemical model that serves as a 'digital twin' of the laboratory conditions—replicating realistic conditions, such as finite temperature and pressure, as well as including solvent molecules explicitly. Given the impracticality of full ab-initio quantum chemical calculations, tailored multiscale approaches must be devised, accounting for conformational flexibility, explicit solvation effects, and the dynamic nature of chemical reactions, while the individual solution depends on the system under study and research question to be answered.

A vast majority of computational studies are performed starting from singular structures to describe reactivity, in what is known as a static approach. While this technique is often suitable for small, rigid structures, conformational sampling is essential for flexible systems.[4,6,7] Recently, more accessible tools have allowed for a transition from single structures to structural ensembles, [8,9] providing a more detailed insight into the reactivity and selectivity of catalytic and other reaction mechanisms. [4,7] Yet this is far from being routine.

Another critical factor to address is solvation. The most commonly employed treatments of solvation are through a continuum description of the bulk solvent.[10–12] However, due to their implicit nature, such methods fall short in capturing explicit interactions between solute and solvent. Additionally, when combined with dispersion corrections and a small basis set, these models tend to favour overly compact structures with numerous intramolecular bonds—resulting in an inadequate representation of a solvated system. [9,13] Consequently, it becomes essential to incorporate explicit solvation effects into any realistic model. Ideally, this can be achieved through a comprehensive condensed phase model or a microsolvation approach. [14,15]

While representing structural ensembles within an explicit solvent represents a substantial enhancement of the chemical model, it remains crucial to acknowledge that molecules, and chemistry in general, inherently exhibit dynamic behaviour. By following the motion of the nuclei during a chemical reaction, such as in a molecular dynamics simulation, we may paint a more accurate picture of reaction mechanisms, as we go beyond the electronic energy profile, incorporating dynamic effects.

Bringing these effects into mechanistic studies pushes the chemical model closer to reality. However, the increase in complexity is accompanied by greater computational costs, particularly in the context of ab-initio molecular dynamics. To mitigate this limitation, multiscale methods—such as quantum mechanics/molecular mechanics (QM/MM)—have emerged. In QM/MM, the core the system under investigation , e.g. the catalytic centre and substrates, is treated at the quantum mechanical (QM) level, while the surrounding environment is described classically, using molecular mechanics (MM).[16–18] When combined with molecular dynamics (MD), this approach allows us to explore the dynamic behaviour of molecules. Transitioning to QM/MM MD simulations in explicit solvent extends our sampling timescales to the order of nanoseconds, surpassing the limitations of pure QM methods. Increased sampling is essential, as reactions can be considered as rare events. Thus, repeated sampling becomes essential for statistically meaningful insights into energy barriers and changes in structural features.

As conformational sampling and dynamics begin to play a greater role in mechanistic investigations, it became apparent that a gap needed to be filled, when it comes to setting up small systems for MD simulations in various non-aqueous solvents at an MM level, particularly those containing transition metals. We lowered the barrier to entry by creating an easy-to-use python-based package which simplifies the setup of an explicitly solvated system and its subsequent analysis.[9] Thus, parameters can be generated on-the-fly for the desired solute and the complete system can be solvated, including possible counterions with ease.

While setting up multiscale methods, combined with MD, presents its own challenges, these simulations generate vast amounts of data, which must be interpreted in order to gain chemical insight. As the resulting dataset exhibits very high dimensionality (3N coordinates, combined with information regarding energy, forces, QM properties etc.), an automated, computer driven analysis is required. To obtain chemical information from the dataset, statistical analysis and machine learning have shown to be key elements. Interpretable machine learning techniques—such as decision trees, random forests, or logistic regression—perform well in extracting relevant information from large datasets and presenting it in easily understandable formats. Additionally, dimensionality reduction techniques[19–21] and feature selection methods,[22–24] can help reduce the highly dimensional space of chemical reactions, rendering a more easily understandable overview of the data. These methods hold promise for delivering chemical insights into mechanisms and reactivity.

While a lot of focus has recently been placed on machine learning approaches, statistical analysis also offers new, until now unexplored, avenues for chemical data analysis. A notable example is causality, a widespread concept, applied across various scientific domains from economics to biology and climate research. [25–36] However, it has not been explored to study chemical reactions, let alone more complex transition-metal catalysis. Molecular dynamics trajectories represent an ideal case for applying this econometric analysis, as they bare a stark resemblance to the data resulting from world markets, with high dimensional variables, presented in the form of time-series.

In a pursuit of understanding chemical reactions at a fundamental level, we exemplify the aforementioned concepts by using various chemical systems and creating tailored multiscale models for their investigation. We demonstrate that implicit solvation and microsolvation, without the need for conformational sampling, are accurate enough to decipher the mechanism of a small rigid system, consisting of benzyl alcohol in the presence of K-tBu, environmental oxygen and water. Even though the synthesis of imines via the oxidation of alcohols to aldehydes is generally performed with metal complexes, an environmental oxygen catalysed reaction has shown high yields under normal laboratory conditions. The mechanistic investigation revealed a remarkably complex radical reaction network, showing how challenging even the investigation of a small system can be.[37]

Additionally, the reactivity and reaction mechanism of a Cu(I)-(2,9-dimethyl-1,10-phenanthroyl)-2,3,4,6,7,8-hexamethyl-*p*-tert-butylcalix[8]arene system, was very thoroughly investigated. This specific catalyst has demonstrated remarkable efficacy in facilitating C-N and C-S coupling reactions with earth abundant metals—an essential process with far-reaching implications across various scientific domains.[38–43] The investigated system poses particular challenges in its computational treatment, requiring a dynamic, ensemble based sampling in explicit solvent for accurate results, as implicit solvation results in a collapse of the highly flexible catalytic cage.[13,44,45] We showed through MM/MD simulations in explicit chloroform that the position of the phenanthroline bridge is key for maintaining an active catalytic centre.[44] The mechanism for this system was established, through QM, to be a sequence of oxidative addition / reductive elimination steps and finally, with the use of a tailored hybrid QM/MM/MD model and statistical methods, we were able to approach an operando description of the system and fully explain the dynamic effects involved in the crucial reaction steps. We developed and applied a high-throughput multiscale QM/MM/MD protocol and combined it with a statistical analysis to elucidate the reaction coordinate and infer causality. We exemplified how both supervised and unsupervised interpretable machine learning techniques can be used to extract chemically relevant information.[45] By avoiding common pitfalls in such approaches,[46] we showed how performance can be enhanced by employing a consensus approach that combines multiple interpretable machine learning methods to reduce data complexity.

7. Methodology

Computational chemistry relies on computers to solve complex chemical problems. When performing a computational investigation of a chemical system, there are two key aspects to consider. On the one hand, the computational model, represents the level of theory used to describe the system. This can be either through the use of electronic structure theory (EST), molecular mechanics (MM) or a combination of these. On the other hand, the chemical model refers to how the molecule under study is described, whether it is modelled by a smaller, cut-down system or not, whether solvent and associated counterions are included in the model and if so, how they are handled (implicitly or explicitly).

While theoretical methods are at the base of computational chemistry, the overall goal of the field remains the explanation or prediction of reactivity and chemical behaviour. Whether one relies on QM, MM or hybrid approaches, the resulting data and numbers must be analysed and transferred into chemical information, in order to understand phenomena happening at the microscopic level. With increasing system complexity and the introduction of dynamic effects, such as in QM/MM/MD, a large amount of data is generated, which cannot be interpreted observationally. To tackle this challenge, statistical approaches, particularly machine learning, can be used to extract chemical information, such as the reaction coordinate, from the dataset.

The following chapters will give a concise overview over the foundations of electronic structure theory, molecular mechanics and molecular dynamics and present more details concerning the specific methodology used and developed for the published studies present within this manuscript. It will also cover hybrid multiscale methods, as well as the aspects of solvation and chemical environment. Likewise, a short statistics chapter will delve into ways to infer chemically relevant information from big data using machine learning and causality.

7.1. Electronic Structure Theory

Electronic structure theory (EST) describes the behaviour of electrons in atoms or molecules. As the nuclei and electrons are very small, elemental particles, an accurate description of their behaviour can only be achieved by taking into account their quantum nature, that is by describing them with quantum mechanics. In quantum chemistry, typically only electrons are treated quantum mechanically, while nuclei are described as charged point particles. The Schrödinger equation precisely describes how the intricate electrons and nuclei interactions govern the atomic world. It takes the form of:

$$\hat{H}\psi = E\psi \qquad (1)$$

Wherein $\hat{H}$ represents the Hamiltonian, ψ represents the wavefunction of the system and E represents the ground state energy of the system.

The Hamiltonian can be further broken down as:

$$\hat{H} = -\overbrace{\sum_{A=1}^{M} \frac{\hbar^2}{2M_{nuc}} \nabla_A^2}^{Kinetic\ Energy\ of\ Nuclei} - \overbrace{\sum_{i=1}^{N} \frac{\hbar^2}{2m_{elec}} \nabla_i^2}^{Kinetic\ Energy\ of\ Electrons} + \overbrace{\sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{4\pi\varepsilon_0} \frac{e^2}{r_i - r_j}}^{Electron-Electron\ Interaction} - \overbrace{\sum_{i=1}^{N} \sum_{A=1}^{M} \frac{1}{4\pi\varepsilon_0} \frac{e \cdot Z_A}{r_i - R_A}}^{Electron-Nucleus\ Interaction} + \overbrace{\sum_{A=1}^{M} \sum_{B>A}^{M} \frac{e^2}{4\pi\varepsilon_0} \frac{Z_A Z_B}{R_A - R_B}}^{Nulceus-Nucleus\ Interaction} \qquad (2)$$

Where the variables and constants are defined as:

M – mass of the nucleus, au

$\hbar$ – reduced Planck's constant, $1.054571817 \times 10^{-34}$ J s

∇ –, differential operator in cartesian coordinates in 3-dimensional Euclidean space

m – mass of electron

ε_0 – permittivity of vacuum

r_i – coordinates of electron in 3-dimensional Euclidean space

R_A – coordinates of nucleus A in 3-dimensional Euclidean space

Z_A – atomic number of nucleus A

e – charge of electron in C

The Hamiltonian describes the total interaction between the electrons and nuclei in a molecule. The first two terms in Eq. (2) describe the kinetic energies of the nuclei and electrons. The electron-nucleus interactions represent are attractive in nature, while the electron-electron and nucleus-nucleus interactions are repulsive in nature.

To simplify the calculations, QM primarily operates within the framework of the Born-Oppenheimer Approximation, which posits that electrons are significantly lighter and therefore significantly faster than nuclei. Consequently, electrons rapidly adjust their distribution to optimize their arrangement for any given nuclear configuration. This results in the term for the kinetic energy of the nuclei to be reduced to 0, while the nucleus-nucleus repulsion remains constant, hence, a decoupling of electronic and nuclear coordinates is achieved.

As there is no analytical solution for the Schrödinger equation for systems with more than one electron, further approximations need to be introduced. Within the field of EST, two main approaches exist. Wavefunction methods, such as Hartree-Fock theory, perturbation theory or coupled cluster approaches, which solve the Schrödinger equations through approximating the electronic wavefunction. In fact, coupled cluster with full singles, doubles and perturbative triples treatment calculations are considered to be the current gold standard, as they achieve chemical accuracy for a large number of systems, provided that no multiconfigurational character is present. However, wavefunction methods are generally very costly in terms of calculation power. Density Functional Theory (DFT), in contrast, reformulates the many-body problem in terms of the electron density. In 1927, Thomas and Fermi proposed that the energy of the system is intrinsically linked to the electron density, the probability distribution of electrons in space, as a functional of the energy.[47,48] In 1964, Hohenberg and Kohn expanded on the idea, by proposing that the external potential of a system uniquely determines the electron density and that by figuring out the exact energy functional, the ground state energy of the system can be localized, when minimizing it with respect to the electron density.[49]

Another breakthrough came a year later in 1965, when Kohn and Sham introduced the concept of a surrogate system, consisting of quasi particles that can be thought of as non-interacting electrons, which mimics the behaviour of the real system.[50] These quasi-particles move in an effective potential, which includes the external potential and exchange-correlation term. The introduction of the surrogate system allowed an approximate calculation of the kinetic energy, the electron-electron interaction, and the electron-nuclei interaction, while only a small fraction of the energy functional remains unknown. For this remainder no analytical solution is known but it has to be approximated. The definition of this functional is still under active study, as locating the exact functional will result in the exact solution.[51–53]

This approach brings calculation speed improvements, as within KS-DFT the wave function can be described by a single determinant and if the exchange-correlation functional was known, this single determinant would yield the correct ground state energy and density. This results in remarkable accuracy, provided the correct density functional is chosen for the system at hand. Various functional classes exist for DFT, each with their own trade-off regarding speed and accuracy.[54]

Other QM approaches which have recently (re-)gained importance are semi-empirical methods, particularly GFN2-xTB.[55] These methods rely on an explicit description of the valence shell electrons with a minimal basis-set, while the core electron density remains frozen, in addition to approximating several integrals by parameters. This trick provides the benefit of extremely rapid calculations, with the downside of requiring thorough benchmarking, as the resulting structures may show gross deviations from those calculated with a higher level of theory such as DFT.[56]

7.2. Molecular Mechanics

Molecular mechanics represents a paradigm that depicts molecular structures as a simplified ball and stick system, relying on classical mechanics to compute the potential energy. Each atom represents a ball with a radius, mass, and charge, while every bond is considered as a spring, with a harmonic potential function.

To be able to evaluate the potential energy of a system, molecular mechanics makes use of a set of predetermined parameters – such as mass, charge, bond, angle, and torsion – and a functional form, collectively known as a force-field. The equations governing the force-field involve a bonded part and a non-bonded part, as shown in the equations of the AMBER force-fields[57] below:

$$E_{total} = E_{bonded} + E_{non-bonded} \qquad (3)$$

$$E_{bonded} = \sum_{bonds} k_b\,(r - r_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2 + \sum_{dihedrals} V_n\,[1 + \cos(n\phi - \gamma)] \qquad (4)$$

$$E_{non-bonded} = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \qquad (5)$$

Where the variables are defined as:

k_b – force constant for a bond in Kcal/mol/Å^2

r_0 – equilibrium bond distance in Å

k_θ – force constant for an angle in Kcal/mol/radian^2

r_θ – equilibrium bond angle in radians

V_n – torsion potential in kcal/mol

n – periodicity

γ – phase in degrees

A_{ij}– parameters for atom pair i-j, defined as εR_{min}^{12}, where ε is depth of the potential well and R$_{min}$ is the van der Waals radius of the atom pair

B_{ij} – parameters for atom pair i-j, defined as $2\varepsilon R_{min}^{6}$, where ε is depth of the potential well and R$_{min}$ is the van der Waals radius of the atom pair

R_{ij} – distance between atom i and atom j in Å

ε – depth of the potential energy well of the atom pair

The parameters for the creation of a force-field are generally derived from high level QM calculations, sometimes in conjunction with experimental data. The bonded part of the force-filed can be obtain via geometry scans or directly from the Hessian matrix of the systems under study (a time-consuming endeavour).[58] For the derivation of charges required for the nonbonded terms, popular force-fields, such as AMBER, use an electrostatic potential fitting on the reference molecular structure, where partial charges are optimized to best describe the QM electrostatic potential.[59,60]

By relying of a force-field approach, calculations can be sped up greatly, allowing for the calculation of systems containing millions of atoms. While accuracy takes a hit, force fields are still performant enough to allow for the replication of larger scale effects, conformational changes, and explicit solvent interactions.[60–62] However, due to the inherent nature of MM force-fields, it is not possible to describe bond-breaking or formation between atoms. Due to this limitation, chemical reactions cannot be described and investigated with standard force-field approaches.

7.3. Solvation

Including solvation effects in the description of a chemical system is a key component, as all homogeneous chemical reaction take place in condensed phase. There are three main approaches to treating solvation effects within computational chemistry, schematically depicted in Figure 1.

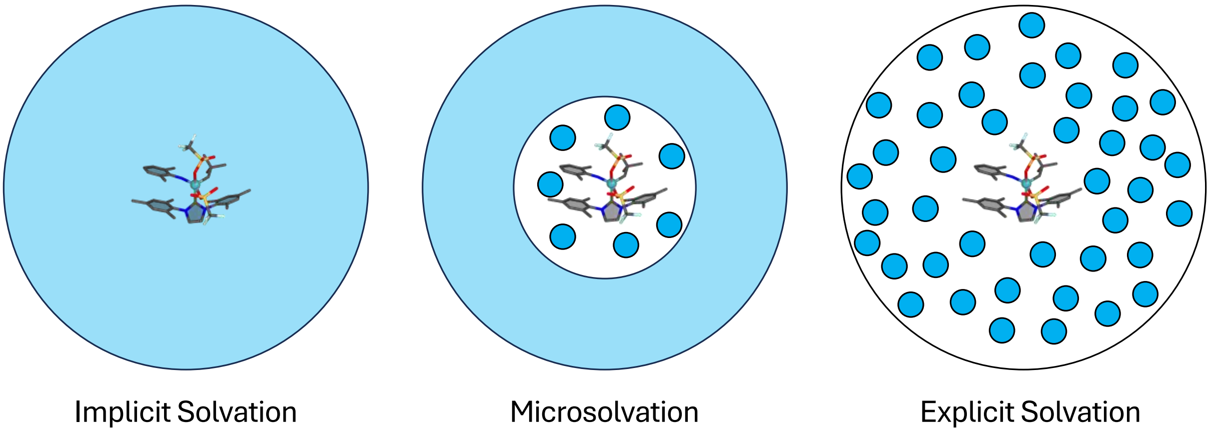


Figure 1. The three main solvation models. Left: Implicit solvation model also referred to as continuum solvation; Middle: Hybrid microsolvation model that describes few solvents explicitly and the rest with a continuum. Right: Explicit solvation model, which treats all solvent molecules explicitly.

The easiest way to consider the effects of solvents is to rely on an implicit solvent (Figure 1 (left)), which describes the surrounding environment using a continuum model, which mimics certain properties (permittivity and refraction index) of the chosen solvent.[12] This model comes with very low computational costs and performs remarkably well, when explicit solute-solvent interactions do not need to be considered.

A second way to treat solvent effects is to explicitly consider all solvent molecules surrounding the solute (Figure 1 (right)). While this approach is computationally much more demanding than implicit solvation, it comes with the advantage of being able to determine the precise solute-solvent interactions, which are essential for certain molecules, such as macrocycles, where a cavity collapse can happen when using just implicit solvent models.

The third way to deal with solvation is a hybrid explicit/implicit approach, known as microsolvation. Here, only a few essential solvent molecules are treated explicitly, with the entire system being surrounded by a dielectric continuum. This approach results in lower computational costs than a fully explicit solvation model, while capturing essential interactions which elude implicit solvation models. However, the great difficulty of this approach is determining the number and placement of the individual solvent molecules. This can be done either stochastically or through physics-based approaches, with the latter showing better performance, yet requiring more computational time.[15]

7.4. Hybrid Approaches

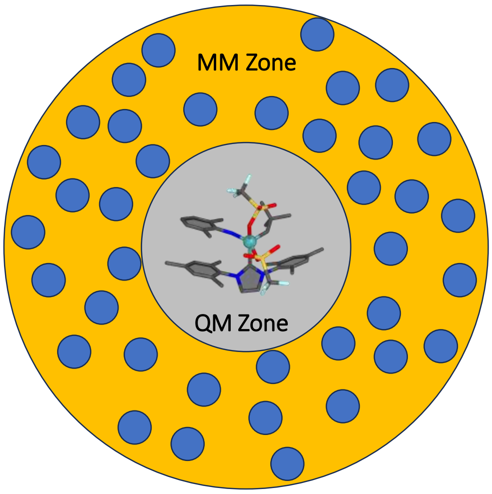


Figure 2. Schematic representation of a hybrid QM/MM system. The solute molecule is calculated at a QM level (gray), while the solvent (blue circles) is calculated at an MM level (yellow).

While QM delivers very accurate results, it is limited to small systems, up to a few hundred atoms, due to the high computational demands. On the other hand, MM offers the possibility to deal with systems going up to several million atoms, while sacrificing some accuracy and the ability to deal with bond breaking.

A way to bridge the gap between the two was introduced by Karplus, Warshel and Levitt, who proposed splitting the system into a core region, to be treated with QM, and an outside region, treated with MM.[16–18] This QM/MM hybrid approach represents an excellent compromise, allowing for vastly reduced calculation time, compared to fully ab-initio models, while also greatly increasing the accuracy in the description of the chemical model.

The main downside of QM/MM is the relatively complicated setup of the system (Figure 2). The QM zone must be large enough to capture all the important effects, while remaining small enough to allow for reduced calculation time. While several techniques exist for the automation of this decision process, it is still not a trivial matter. [63,64] Another complexity of QM/MM is the treatment of the interaction between the two distinct regions. The total Hamiltonian of the system is then divided into three separate terms, with one Hamiltonian for the QM part, one for the MM part and one for the QM/MM interaction:

$$\hat{H}_{Total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \tag{6}$$

The $\hat{H}_{QM/MM}$ term can be defined in several ways. The most straightforward approach, widely used in the ONIOM model,[65,66] is the subtractive scheme, where the Hamiltonian for the complete system at the MM level is calculated, added to that of the QM region and lastly, a MM Hamiltonian of the QM region at the MM level is subtracted:

$$\hat{H}_{QM/MM} = \hat{H}_{MM(A+B)} + \hat{H}_{QM(A)} - \hat{H}_{MM(A)} \tag{7}$$

This scheme has the benefit of being quite simple to implement, it does not require any code modifications and avoids any double counting of interactions. However, the QM zone is not polarized by the MM background.

An alternative scheme is the additive scheme, wherein the two regions are calculated independently at their respective levels of theory and an interaction term is added:

$$\hat{H}_{QM/MM} = \hat{H}_{MM(B)} + \hat{H}_{QM(A)} + \hat{H}_{QM/MM(B-A)} \tag{8}$$

This scheme has some advantages, such as overcoming the need for force-field parameters for the QM region and no double counting of interactions. These interactions can be calculated with various levels of sophistication. The simplest way, mechanical embedding, treats the electrostatic interaction between the two regions at the MM level.[65,66] This ansatz can result in significant error, as generally the changes in the QM region are not reflected in the MM force-field. Additionally, the QM region is not affected by the electrostatics of the MM region. A more complex approach, electrostatic embedding, expands on the mechanical embedding, by calculating the QM region in the presence of the MM point charges.[67,68] This results in an improved description of the QM zone, alongside increased computational costs. A major drawback is that the MM zone does not react to the changes in the QM system. This shortcoming is remedied by a polarized embedding approach, where the force-field parameters are updated as the QM system changes, in a self-consistent way. This method is significantly more costly, yet it provides the highest accuracy of all QM/MM approaches to date.[69,70]

Another significant element of setting up QM/MM systems is the treatment of bonds which cross the zone boundaries, which is generally the case for systems which surpass 50-100 atoms. Several schemes exist to deal with this issue.

In the boundary atom schemes, a dummy atom (link-atom), generally a hydrogen, is introduced which replaces the partner atom of the bond which crosses the boundary. It satisfies the valence of the QM system, capping the broken bond. This model can be improved, by adjusting the position of the link-atom and accompanying it with a potential function, thus mimicking the behaviour of the atom it replaces. [71]

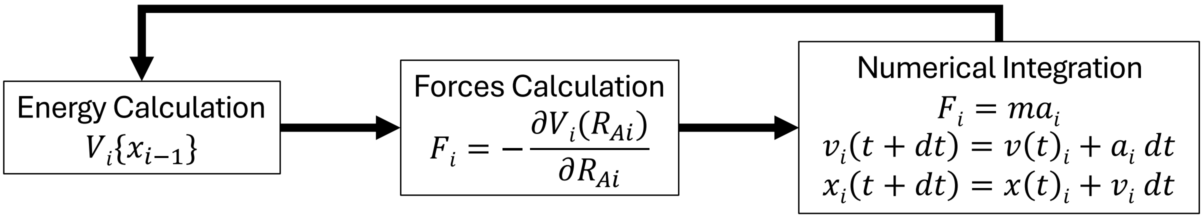
An alternative to boundary atoms is placing a frontier orbital to satisfy the valence of the atom with the broken covalent bond.[72,73] This approach, referred to as the localized orbital scheme, utilizes a doubly occupied orbital, attached to the QM atom, which is kept frozen during calculations.[74] However, this method is more costly, compared to link-atom approaches, as additional parametrization steps need to be performed for the generation of the orbitals. Furthermore, there is little advantage to using localized orbitals, rather than link-atoms.[75]

7.5. Dynamics

As chemical systems are dynamic in nature, a static picture calculated either with DFT, MM or hybrid QM/MM represents a simplification. One way of accounting for the dynamic nature of molecules are molecular dynamics (MD) simulations, which allow the study of the time-evolution of the system. MD relies on satisfying Newton's equations of motions, allowing us to observe the dynamic evolution of a molecular system.

$$F = -\frac{\partial V(R_A)}{\partial R_A} \quad (9)$$

The force (F) acting upon the atoms (A) can be derived as the partial derivative of the potential energy (V) with respect to the atomic coordinates (R_A). Thus, the system can be propagated by solving the equations of motion and updating the atomic coordinates by discretizing the time into finite steps, as shown in Scheme 1 below.



Scheme 1. Basic concept of MD system propagation, where x_i denotes the positions of the atoms, v_i the velocities, a_i the acceleration and t the timestep.

Typically, for MD, the systems are subjected to environmental temperature and pressure. This setup allows for the sampling of conformations that are generally accessible under environmental conditions. Depending on the chosen conditions and the properties that are to be recovered from the calculations, the simulation can take place under one of several different ensembles. The most basic ensemble, the microcanonical ensemble (NVE), keeps the number of particles (N), volume (V) and energy (E) constant for the system. This corresponds to a system which cannot exchange heat or particles with its environment. A more largely used approach is to perform the simulations in the canonical ensemble (NVT), where the temperature (T) and not the energy is kept constant in addition to number of particles (N) and volume (V). This is achieved via the use of a thermostat, which scales the kinetic energy of the system in such a way as to keep the kinetic energy, thus temperature, constant.[76,77] Another alternative is given by the isothermal-isobaric ensemble (NPT), where pressure, instead of volume is kept constant. By using a barostat, the size of the simulation box can be adjusted, to keep a constant pressure.[78,79]

Regardless of the ensemble used, if the starting structure is in a deep energy well, it might not be able to overcome the barriers given the time frame of the simulation, the system is stuck in a minimum and cannot explore more of the potential energy surface (PES). In these cases, a switch to enhanced sampling methods, such as metadynamics[80] or accelerated molecular dynamics (AMD)[81] can prove useful. Enhanced sampling methods allow for faster sampling of conformational space by global rescaling of the PES of the molecule. Metadynamics allows for a faster exploration of the structural space by biasing the already visited conformations. This comes as the expense of requiring a definition of a collective variable, which represents the structural parameter changes that are of interest, however, it is not always trivial to define and requires chemical knowledge about the system. Another approach is aMD, which adds a boost potential to the true potential energy of the system, when it is below a certain threshold. This way, deep energy wells are flattened, allowing for a considerably faster sampling of the PES.

MD can be performed with either QM, MM or QM/MM hybrid methods as the underlying base. While MM is used for large systems, QM methods are mostly used for small ab-initio investigations. As a middle-ground, semi-empirical methods can be used to bridge the gap between the two either in a QM or in a QM/MM approach, in particular semi-empirical GFN2-xTB in combination with metadynamics is used to explore the conformations of molecules in implicit solvent.[8] In all instances of this thesis, nuclear motion was treated classically, that is a Born-Oppenheimer MD simulations were performed.

7.6. Statistical Analysis and Causality

Dynamic investigations of chemical systems result in the creation of vast datasets, which need to be processed in an automated fashion, to obtain a deeper chemical interpretation. A common characteristic of these datasets is the high dimensionality, such as a large number of atomic coordinates and properties. This “curse of dimensionality” renders the raw data difficult to interpret for chemists, thus the use of statistical analysis, such as machine learning approaches, dimensionality reduction techniques and causality models can aid in the interpretation of results.

Machine learning methods can be divided into several categories, depending on the data they rely on or on how easy it is to interpret the algorithm behind the method. The techniques can be divided into supervised (require labelled data) or unsupervised methods (do not require labels), depending on the data fed to the algorithm. The labels represent a way to separate the points in the dataset, by assigning each one to a category, which can be based, for example, on chemical knowledge (educt, transition state, product categories). Similarly, the methods can be classified on whether the results can be easily traced back to the input data, dividing them into interpretable and non-interpretable approaches. As we are interested in chemical interpretability/explainability of the results, we use interpretable ML approaches only.

To tackle the high dimensionality of the data (e.g. coordinates, properties resulting from MD simulations, also known as features) dimensionality reduction methods, such as principal component analysis (PCA),[19] Linear Discriminant Analysis (LDA)[21] or time-lagged Independent Component Analysis (tICA)[20] are often used.[82–84] PCA aims to reduce the high dimensional space by projecting the features of the dataset into a lower dimensional one. This lower dimension is defined by a new orthogonal set of axes or directions, called principal components, which aim to capture the maximum variance within the data. tICA acts in a comparable way to PCA, reducing the dimensionality of the dataset by projecting it onto a new set of directions, with the difference that rather than maximising the variance in the dataset, it tries to maximise the autocorrelation in the data at a given lag-time. Thus, it is able to identify degrees of freedom which show significant correlation with their lagged value, which, in the context of MD, lead to the identification of slow motions in a system. While both PCA and tICA are unsupervised methods, LDA represents a supervised dimensionality reduction approach. LDA aims to find a new set of dimensions which maximises the variation and at the same time the separation between the various labelled classes within the dataset (e.g. products and educts). In all cases, the new dimensions are defined as a linear combination of the original features (system properties and/or coordinates in the case of MD), thus all methods are easily interpretable.

To pre-process the dataset and enhance the performance of dimensionality reduction methods, feature selection can also be performed. Generally, this involves training a classifier, a model constructed to separate between the various labelled classes in the dataset, such as random forest[85] or logistic regression[24]. By examining the contribution of each feature to the classification performance, an importance value can be assigned. By selecting only the features which aid in the separation of the various classes in the dataset, extraneous components can be discarded, thus reducing the overall dimensionality.

Classification models can also be used independently, to quantify the differences between the various classes they have been trained on, such as conformations or states. For example, a random forest or decision tree can be used to examine the distributions of the various features and compare them between states.

While the aforementioned methods can reduce the dataset to something more easily interpretable, e.g., a small number of internal coordinates to define a reaction coordinate, investigating the nature of the relationship between the various features can bring further insight. Most analysis methods focus or rely on the correlation relationships between features, while causal interactions are not considered. However, as MD simulations are time series, inferring causal relationships between variables is possible. One way to investigate such interactions is to use the Granger causality model.[86] This model attempts to predict the future values of one variable (Y) using its past values. If the prediction improves when also considering the past values of a second variable (X), then the model indicates that X causes Y. While the model is fairly simple, it has proven to be quite powerful in various scientific fields. This statistical model can be applied to any data which contains multiple variables in the form of time-series, such as MD trajectories. Naturally, Granger Causality requires some assumptions about the underlying dataset, such as stationarity and cointegration, which can be verified through statistical tests.[86–90]

Another key factor in statistics has to deal with repeated sampling of a single event. It is critical to perform repeated sampling, to ascertain that the effects witnessed are not just happenstance but can be generalized to the event itself. This is especially relevant when looking at chemical reactions, especially when attempting to infer causality or define general reaction coordinates. However, such an approach necessitates false detection rate corrections, another commonly used statistical technique, which automatically adjusts the p-value statistic with increased observations of events.[91]

8. Developed Methods and Investigated Systems

This chapter contains information about the systems that have been investigated. It also exemplifies the methods and workflow which were developed to advance the chemical model and obtain a more accurate understanding of chemical reactivity.

8.1. PyConSolv: Force-Field Parametrization and MD Analysis

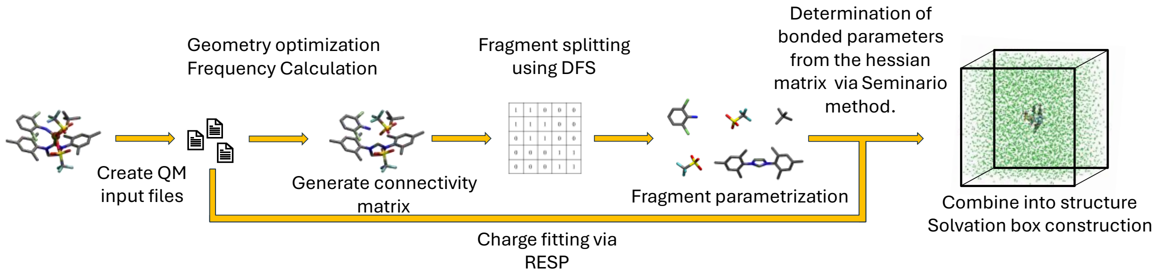


Figure 3. Automated parametrization procedure using PyConSolv, adapted from Ref. 9.

One significant downside of molecular mechanics is the lack of existing “off-the-shelf”-parameters for the treatment of metal-containing systems. Metals pose a particular challenge due to their complex binding behaviour. Thus, if a system containing one or several metals must be treated at a molecular mechanics level, it requires custom parameters. This issue can be overcome through automatic methods, which derive the appropriate parameters from quantum mechanical calculations,[58] however, the process is involved and requires significant user input. To simplify this task, we introduced the PyConSolv package, which can automatically generate parameters with minimal input, following the workflow shown in Figure 3**Error! Reference source not found.**, requiring minimal intervention from the user.[9]

8.2. Supramolecular transition-metal complexes - operando modelling and evaluation

As molecules in solution are dynamic in nature, it is only natural to perform a dynamic evaluation of the reaction mechanism, considering explicit solvent. To obtain a statistically relevant picture of the reaction, repeated sampling is required, in a high-throughput fashion.

To this end, we developed a QM/MM/MD model, relying of highly efficient semi-empirical methods as the QM theory, thus allowing for massive sampling. The 1.5 million resulting datapoints represent a very high-dimensional dataset, consisting of coordinates and energies. We showed that by relying on interpretable machine learning techniques, we can directly obtain chemically relevant information and define the reaction coordinate. Furthermore, by utilising a statistical model we were able to infer the Granger causality between the reaction coordinate components, delivering an unprecedented new way to study chemical reactions.[45]

8.3. Synthesis of Imines from Primary Alcohols and Amines

The use of alcohols as alkylation surrogates for aldehydes is a highly desirable strategy in organic chemistry, particularly as alcohols can be directly sourced from biomass. Over the past decades, researchers have developed efficient catalytic methods for in situ generation of carbonyl functionalities, relying on both noble and nonprecious transition metals.

Notably, benzyl alcohols can undergo rapid oxidation to form the corresponding carbonyl compounds when treated with strong bases under atmospheric conditions. Condensation reactions involving benzyl alcohols can occur in the presence of base (sometimes catalytic amounts) without the need for transition-metal complexes. These transformations are predominantly driven by alkali metal-catalysed Meerwein–Ponndorf–Verley reduction/Oppenauer oxidation (MPV-O) reactions.[92 92]

Another noteworthy transformation involves the synthesis of imines from benzyl alcohols and anilines—a reaction known since the 1950s but often overlooked in the organometallic community. Le Berre and colleagues investigated the oxidation of diphenylmethanol using KOtBu and oxygen, resulting in benzophenone and KOOH (originally reported as KO_2).[93] Unlike transition metal-catalysed reactions, the mechanistic details of this process remain relatively unexplored.

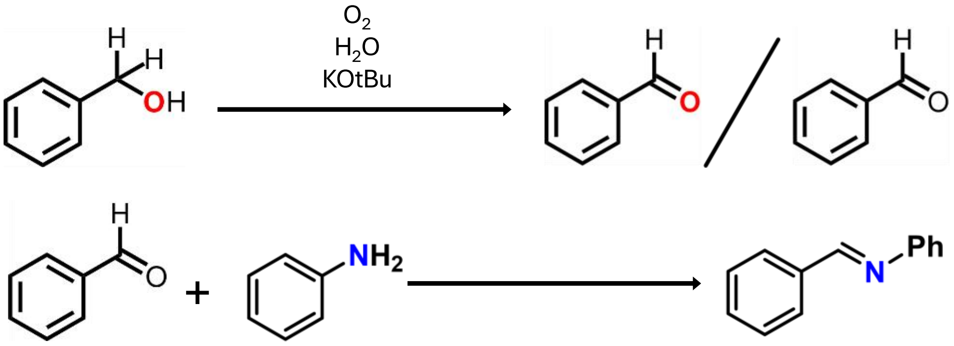


Figure 4. Formal reaction scheme for the oxidation of alcohols to aldehydes using environmental oxygen. This is followed by the formation of imine from the aldehyde and aniline.

While the experimental coupling reaction of alcohols and amines promoted by oxygen at room temperature was investigated by our collaboration partners (see Figure 4), a detailed mechanistic understanding was missing. We used density functional theory along with implicit and explicit solvation and incorporation of counter ions to propose a comprehensive reaction mechanism in agreement with experimental observations. We revealed a complex radical reaction network, leading to the formation of the aldehyde, via several distinct pathways. Our modelling required the evaluation of minimal energy crossing points between various spin surfaces to identify the most favourable pathways, while our chemical model included the counter ion and explicit solvent molecules. This augmented chemical model was necessary to achieve accessible barriers.[37]

8.4. Supramolecular Nanoreactor for C-X Coupling

The formation of C–X bonds (X = S, N) is a process in high demand due to the importance of amine and sulphur derivatives as fundamental building blocks for biologically active compounds and high-performance materials.[38–42,94] To address this issue, the development of novel, highly efficient catalysts based on Earth-abundant metals (such as copper or iron) becomes crucial for environmentally friendly and sustainable production processes.

However, Earth-abundant metal systems, particularly those involving copper, are susceptible to deactivation patterns, including dimerization and oxidation from Cu(I) to Cu(II). Recent approaches to incorporate copper as a catalytic centre have turned to bioinspired supramolecular catalysis. In this strategy, transition metals are encapsulated within confined spaces or cavities, creating a constrained chemical environment for the active species.[95] These cavities also impose steric restrictions during catalytic transformations. The result is efficient collision between substrates and the anchored metal site, leading to a high density of local interactions. This behaviour mimics the active sites of metalloenzymes and often results in beneficial effects, such as enhanced activity and selectivity.

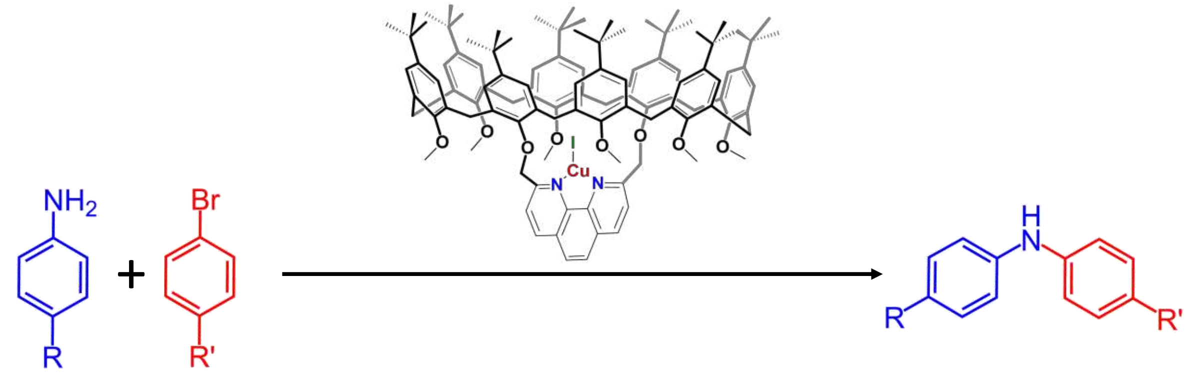


Figure 5. C-N coupling reaction with the [Cu(C_8PhenMe$_6$)I] catalyst.

A promising candidate for C-X coupling reactions is [Cu(C_8PhenMe$_6$)I], depicted in Figure 5. This catalytic system features a Cu(I) centre, surrounded by a macrocyclic calix[8]arene cage and exhibits remarkable catalytic activity in C−X cross-coupling reactions in experiments. Given these promising attributes, investigating C_8-based encapsulation of Cu-phenanthroline complexes for C−X coupling becomes a compelling avenue to develop highly efficient catalysts and broaden the scope of applicable reactions.[13,44] However, such highly flexible catalytic systems pose a challenge, as they require conformational sampling in explicit solvent, due to the tendency to favour intramolecular bonds when implicit solvation is used in combination with a small basis set.

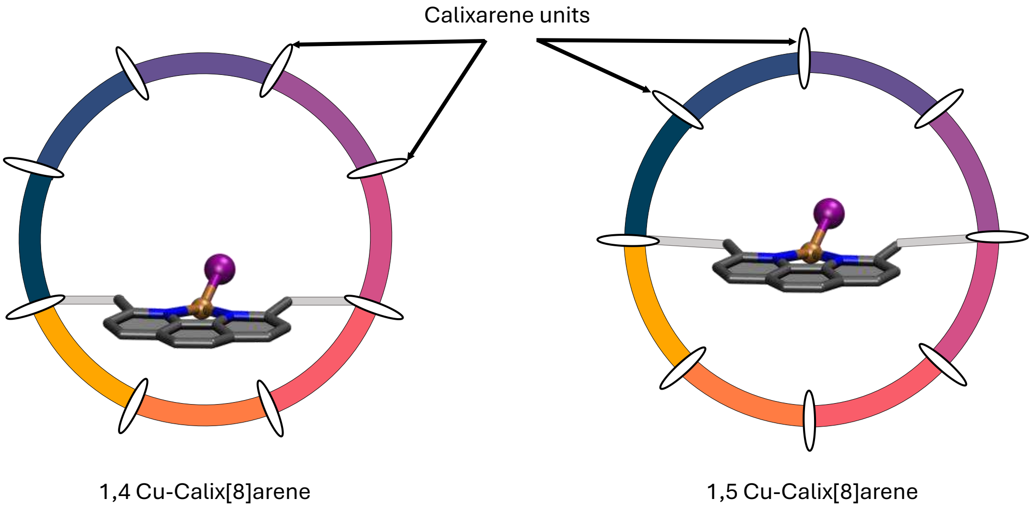


Figure 6. Two regio-isomers of the Cu-Calix[8]arene, defines by the calixarene units to which the phenanthroline is bound.

The computational investigations revealed that the substitution pattern of the calixarenes is critical for maintaining reactivity (see Figure 6). If the phenanthroline bridge is bound to the 1 and 4 calixarenes, then the catalyst becomes deactivated. MD simulations revealed how the catalytic centre becomes exposed due to the lopsided cavity.[44]

9. Results and Discussion

As this work is a cumulative thesis, this section contains a compilation of four original research articles that are published in peer-reviewed journals and one additional article that is currently under evaluation, but, as of the time of writing, being published only on a preprint server.

The included publications are listed in the included publications section and their text is included in the appendix.

10. Conclusion

By exploring the reactivity of various chemical systems, from small rigid molecules, to large, flexible transition metal catalysts, we were able to show the importance of constructing an appropriate chemical model. Thus, we developed tailored multiscale computational protocols to model experimental conditions for several chemical systems.

It is apparent that solvent plays a vital role in all the systems under investigation. The complex reaction network for the transition metal free oxidation of alcohols to aldehydes can only be explained via solvent oxygen exchange steps. Likewise, obtaining realistic structures with the large, flexible Cu-Calix[8]arene catalysts is only possible when modelling them under explicit solvation.

By developing tools, such as PyConSolv, to simplify the parametrization of molecules, particularly transition-metal catalysts, for use in MM studies, we open up the avenue for explicit solvation studies to become mainstream. Furthermore, having an adequate force-field description of systems allows for the transition to conformational ensembles or even dynamic studies.

Taking the chemical model a step further, investigating chemical reactions with high-throughput QM/MM MD allowed us to study the bond formation dynamics at experimental conditions. By developing a workflow to identify and quantify the reaction coordinate from a set of trajectories, we detected chemically intuitive and less intuitive structural changes that led to bond formation. Through a combination of interpretable machine learning methods, chemically relevant information can be easily extracted from the high dimensional dataset. By taking the analysis a step further and utilising a causality model, the relationship between the individual contributions to the reaction coordinate can be decomposed into a sequence of individual motions, identifying a cause for the reaction to happen.

The methodology and studies contained within this work serve as a template to study the reaction dynamics of any chemical system under realistic conditions. It drives the development operando chemical models in the quest for predictive computational chemistry.

11. References

(1) Fey, N.; Lynam, J. M. Computational Mechanistic Study in Organometallic Catalysis: Why Prediction Is Still a Challenge. *WIREs Comput. Mol. Sci.* **2022**, *12* (4), e1590. https://doi.org/10.1002/wcms.1590.

(2) Sciortino, G.; Maseras, F. Computational Study of Homogeneous Multimetallic Cooperative Catalysis. *Top. Catal.* **2022**, *65* (1), 105–117. https://doi.org/10.1007/s11244-021-01493-2.

(3) Artús Suàrez, L.; Balcells, D.; Nova, A. Computational Studies on the Mechanisms for Deaminative Amide Hydrogenation by Homogeneous Bifunctional Catalysts. *Top. Catal.* **2022**, *65* (1), 82–95. https://doi.org/10.1007/s11244-021-01542-w.

(4) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9* (8), 6803–6813. https://doi.org/10.1021/acscatal.9b01537.

(5) Besora, M.; Braga, A. A. C.; Ujaque, G.; Maseras, F.; Lledós, A. The Importance of Conformational Search: A Test Case on the Catalytic Cycle of the Suzuki–Miyaura Cross-Coupling. *Theor. Chem. Acc.* **2011**, *128* (4–6), 639–646. https://doi.org/10.1007/s00214-010-0823-6.

(6) Podewitz, M.; Sen, S.; Buchmeiser, M. R. On the Origin of E-Selectivity in the Ring-Opening Metathesis Polymerization with Molybdenum Imido Alkylidene N-Heterocyclic Carbene Complexes. *Organometallics* **2021**, *40* (15), 2478–2488. https://doi.org/10.1021/acs.organomet.1c00229.

(7) Eisenstein, O.; Ujaque, G.; Lledós, A. What Makes a Good (Computed) Energy Profile? In *New Directions in the Modeling of Organometallic Reactions*; Lledós, A., Ujaque, G., Eds.; Topics in Organometallic Chemistry; Springer International Publishing: Cham, 2020; pp 1–38. https://doi.org/10.1007/3418_2020_57.

(8) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST—A Program for the Exploration of Low-Energy Molecular Chemical Space. *J. Chem. Phys.* **2024**, *160* (11), 114110. https://doi.org/10.1063/5.0197592.

(9) Talmazan, R. A.; Podewitz, M. PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent. *J. Chem. Inf. Model.* **2023**, *63* (17), 5400–5407. https://doi.org/10.1021/acs.jcim.3c00798.

(10) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001. https://doi.org/10.1021/jp9716997.

(11) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc. Perkin Trans. 2* **1993**, No. 5, 799–805. https://doi.org/10.1039/P29930000799.

(12) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. https://doi.org/10.1021/jp810292n.

(13) Talmazan, R. A.; Refugio Monroy, J.; del Río-Portilla, F.; Castillo, I.; Podewitz, M. Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]Arene Supramolecular Catalyst. *ChemCatChem* **2022**, *14* (20). https://doi.org/10.1002/cctc.202200662.

(14) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic Microsolvation Approach with a Cluster-Continuum Scheme and Conformational Sampling. *J. Comput. Chem.* **2020**, *41* (12), 1144–1155. https://doi.org/10.1002/jcc.26161.

(15) Steiner, M.; Holzknecht, T.; Schauperl, M.; Podewitz, M. Quantum Chemical Microsolvation by Automated Water Placement. *Molecules* **2021**, *26* (6), 1793. https://doi.org/10.3390/molecules26061793.

(16) Warshel, A.; Karplus, M. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.* **1972**, *94* (16), 5612–5625. https://doi.org/10.1021/ja00771a014.

(17) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249. https://doi.org/10.1016/0022-2836(76)90311-9.

(18) Gao, J. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29* (6), 298–305. https://doi.org/10.1021/ar950140r.

(19) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

(20) Molgedey, L.; Schuster, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **1994**, *72* (23), 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634.

(21) Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Mullers, K. R. Fisher Discriminant Analysis with Kernels. In *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*; 1999; pp 41–48. https://doi.org/10.1109/NNSP.1999.788121.

(22) Chormunge, S.; Jena, S. Correlation Based Feature Selection with Clustering for High Dimensional Data. *J. Electr. Syst. Inf. Technol.* **2018**, *5* (3), 542–549. https://doi.org/10.1016/j.jesit.2017.06.004.

(23) Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58* (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

(24) Cox, D. R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. B Methodol.* **1958**, *20* (2), 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x.

(25) Imbens, G. W.; Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: Cambridge, 2015. https://doi.org/10.1017/CBO9781139025751.

(26) Xing, Q.; Wu, C.; Chen, F.; Liu, J.; Pradhan, P.; Bryan, B. A.; Schaubroeck, T.; Carrasco, L. R.; Gonsamo, A.; Li, Y.; Chen, X.; Deng, X.; Albanese, A.; Li, Y.; Xu, Z. Intranational Synergies and Trade-Offs Reveal Common and Differentiated Priorities of Sustainable Development Goals in China. *Nat. Commun.* **2024**, *15* (1), 2251. https://doi.org/10.1038/s41467-024-46491-6.

(27) Cárdenas-García, P. J.; Brida, J. G.; Segarra, V. Modeling the Link between Tourism and Economic Development: Evidence from Homogeneous Panels of Countries. *Humanit. Soc. Sci. Commun.* **2024**, *11* (1), 1–12. https://doi.org/10.1057/s41599-024-02826-8.

(28) Kaufmann, R. K.; Newberry, D.; Xin, C.; Gopal, S. Feedbacks among Electric Vehicle Adoption, Charging, and the Cost and Installation of Rooftop Solar Photovoltaics. *Nat. Energy* **2021**, *6* (2), 143–149. https://doi.org/10.1038/s41560-020-00746-w.

(29) Tanaka, T.; Guo, J. International Price Volatility Transmission and Structural Change: A Market Connectivity Analysis in the Beef Sector. *Humanit. Soc. Sci. Commun.* **2020**, *7* (1), 1–13. https://doi.org/10.1057/s41599-020-00657-x.

(30) Ouyang, T.; Liu, F.; Huang, B. Dynamic Econometric Analysis on Influencing Factors of Production Efficiency in Construction Industry of Guangxi Province in China. *Sci. Rep.* **2022**, *12* (1), 17509. https://doi.org/10.1038/s41598-022-22374-y.

(31) Le, T. Increased Impact of the El Niño–Southern Oscillation on Global Vegetation under Future Warming Environment. *Sci. Rep.* **2023**, *13* (1), 14459. https://doi.org/10.1038/s41598-023-41590-8.

(32) Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; Glymour, C.; Kretschmer, M.; Mahecha, M. D.; Muñoz-Marí, J.; van Nes, E. H.; Peters, J.; Quax, R.; Reichstein, M.; Scheffer, M.; Schölkopf, B.; Spirtes, P.; Sugihara, G.; Sun, J.; Zhang, K.; Zscheischler, J. Inferring Causation

from Time Series in Earth System Sciences. *Nat. Commun.* **2019**, *10* (1), 2553. https://doi.org/10.1038/s41467-019-10105-3.

(33) Hill, S. M.; Heiser, L. M.; Cokelaer, T.; Unger, M.; Nesser, N. K.; Carlin, D. E.; Zhang, Y.; Sokolov, A.; Paull, E. O.; Wong, C. K.; Graim, K.; Bivol, A.; Wang, H.; Zhu, F.; Afsari, B.; Danilova, L. V.; Favorov, A. V.; Lee, W. S.; Taylor, D.; Hu, C. W.; Long, B. L.; Noren, D. P.; Bisberg, A. J.; Mills, G. B.; Gray, J. W.; Kellen, M.; Norman, T.; Friend, S.; Qutub, A. A.; Fertig, E. J.; Guan, Y.; Song, M.; Stuart, J. M.; Spellman, P. T.; Koeppl, H.; Stolovitzky, G.; Saez-Rodriguez, J.; Mukherjee, S. Inferring Causal Molecular Networks: Empirical Assessment through a Community-Based Effort. *Nat. Methods* **2016**, *13* (4), 310–318. https://doi.org/10.1038/nmeth.3773.

(34) Walter, J. Establishing Microbiome Causality to Tackle Malnutrition. *Nat. Microbiol.* **2024**, *9* (4), 884–885. https://doi.org/10.1038/s41564-024-01653-6.

(35) Friston, K. J.; Harrison, L.; Penny, W. Dynamic Causal Modelling. *NeuroImage* **2003**, *19* (4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7.

(36) Zheng, J.; Anderson, K. L.; Leal, S. L.; Shestyuk, A.; Gulsen, G.; Mnatsakanyan, L.; Vadera, S.; Hsu, F. P. K.; Yassa, M. A.; Knight, R. T.; Lin, J. J. Amygdala-Hippocampal Dynamics during Salient Information Processing. *Nat. Commun.* **2017**, *8* (1), 14413. https://doi.org/10.1038/ncomms14413.

(37) Himmelbauer, D.; Talmazan, R.; Weber, S.; Pecak, J.; Thun-Hohenstein, A.; Geissler, M.-S.; Pachmann, L.; Pignitter, M.; Podewitz, M.; Kirchner, K. No Transition Metals Required – Oxygen Promoted Synthesis of Imines from Primary Alcohols and Amines under Ambient Conditions. *Chem. – Eur. J.* **2023**, *29* (29), e202300094. https://doi.org/10.1002/chem.202300094.

(38) Buskes, M. J.; Blanco, M.-J. Impact of Cross-Coupling Reactions in Drug Discovery and Development. *Molecules* **2020**, *25* (15), 3493. https://doi.org/10.3390/molecules25153493.

(39) Gay, M.; Carato, P.; Coevoet, M.; Renault, N.; Larchanché, P.-E.; Barczyk, A.; Yous, S.; Buée, L.; Sergeant, N.; Melnyk, P. New Phenylaniline Derivatives as Modulators of Amyloid Protein Precursor Metabolism. *Bioorg. Med. Chem.* **2018**, *26* (8), 2151–2164. https://doi.org/10.1016/j.bmc.2018.03.016.

(40) Jiang, R.; Li, L.; Sheng, T.; Hu, G.; Chen, Y.; Wang, L. Edge-Site Engineering of Atomically Dispersed Fe–N4 by Selective C–N Bond Cleavage for Enhanced Oxygen Reduction Reaction Activities. *J. Am. Chem. Soc.* **2018**, *140* (37), 11594–11598. https://doi.org/10.1021/jacs.8b07294.

(41) Hayakawa, S.; Kawasaki, A.; Hong, Y.; Uraguchi, D.; Ooi, T.; Kim, D.; Akutagawa, T.; Fukui, N.; Shinokubo, H. Inserting Nitrogen: An Effective Concept To Create Nonplanar and Stimuli-Responsive Perylene Bisimide Analogues. *J. Am. Chem. Soc.* **2019**, *141* (50), 19807–19816. https://doi.org/10.1021/jacs.9b09556.

(42) Izumi, S.; Higginbotham, H. F.; Nyga, A.; Stachelek, P.; Tohnai, N.; Silva, P. de; Data, P.; Takeda, Y.; Minakata, S. Thermally Activated Delayed Fluorescent Donor–Acceptor–Donor–Acceptor π-Conjugated Macrocycle for Organic Light-Emitting Diodes. *J. Am. Chem. Soc.* **2020**, *142* (3), 1482–1491. https://doi.org/10.1021/jacs.9b11578.

(43) Picini, F.; Schneider, S.; Gavat, O.; Vargas Jentzsch, A.; Tan, J.; Maaloum, M.; Strub, J.-M.; Tokunaga, S.; Lehn, J.-M.; Moulin, E.; Giuseppone, N. Supramolecular Polymerization of Triarylamine-Based Macrocycles into Electroactive Nanotubes. *J. Am. Chem. Soc.* **2021**, *143* (17), 6498–6504. https://doi.org/10.1021/jacs.1c00623.

(44) Berlanga-Vázquez, A.; Talmazan, R. A.; Reyes-Mata, C. A.; Percástegui, E. G.; Flores-Alamo, M.; Podewitz, M.; Castillo, I. Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]Arene C−S Coupling. *Eur. J. Inorg. Chem.* **2022**. https://doi.org/10.1002/ejic.202200596.

(45) Talmazan, R. A.; Gamper, J.; Castillo, I.; Hofer, T. S.; Podewitz, M. From Data to Chemistry: Revealing Causality and Reaction Coordinates through Interpretable Machine Learning in Supramolecular Transition Metal Catalysis. ChemRxiv June 25, 2024. https://doi.org/10.26434/chemrxiv-2024-nd20j.

(46) Chen, V.; Yang, M.; Cui, W.; Kim, J. S.; Talwalkar, A.; Ma, J. Applying Interpretable Machine Learning in Computational Biology—Pitfalls, Recommendations and Opportunities for New Developments. *Nat. Methods* **2024**, *21* (8), 1454–1461. https://doi.org/10.1038/s41592-024-02359-7.

(47) Thomas, L. H. The Calculation of Atomic Fields. *Math. Proc. Camb. Philos. Soc.* **1927**, *23* (5), 542–548. https://doi.org/10.1017/S0305004100011683.

(48) Fermi, E. Statistical Method to Determine Some Properties of Atoms. *Rendiconti Lincei Sci. Fis. E Nat.* **2011**, *22*, 283–314.

(49) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. https://doi.org/10.1103/PhysRev.136.B864.

(50) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133.

(51) Kaupp, M.; Wodyński, A.; Arbuznikov, A. V.; Fürst, S.; Schattenberg, C. J. Toward the Next Generation of Density Functionals: Escaping the Zero-Sum Game by Using the Exact-Exchange Energy Density. *Acc. Chem. Res.* **2024**, *57* (13), 1815–1826. https://doi.org/10.1021/acs.accounts.4c00209.

(52) Becke, A. D. Doubling down on Density-Functional Theory. *J. Chem. Phys.* **2023**, *159* (24), 241101. https://doi.org/10.1063/5.0178236.

(53) Teale, A. M.; Helgaker, T.; Savin, A.; Adamo, C.; Aradi, B.; Arbuznikov, A. V.; Ayers, P. W.; Baerends, E. J.; Barone, V.; Calaminici, P.; Cancès, E.; Carter, E. A.; Chattaraj, P. K.; Chermette, H.; Ciofini, I.; Crawford, T. D.; Proft, F. D.; Dobson, J. F.; Draxl, C.; Frauenheim, T.; Fromager, E.; Fuentealba, P.; Gagliardi, L.; Galli, G.; Gao, J.; Geerlings, P.; Gidopoulos, N.; Gill, P. M. W.; Gori-Giorgi, P.; Görling, A.; Gould, T.; Grimme, S.; Gritsenko, O.; Jensen, H. J. A.; Johnson, E. R.; Jones, R. O.; Kaupp, M.; Köster, A. M.; Kronik, L.; Krylov, A. I.; Kvaal, S.; Laestadius, A.; Levy, M.; Lewin, M.; Liu, S.; Loos, P.-F.; Maitra, N. T.; Neese, F.; Perdew, J. P.; Pernal, K.; Pernot, P.; Piecuch, P.; Rebolini, E.; Reining, L.; Romaniello, P.; Ruzsinszky, A.; Salahub, D. R.; Scheffler, M.; Schwerdtfeger, P.; Staroverov, V. N.; Sun, J.; Tellgren, E.; Tozer, D. J.; Trickey, S. B.; Ullrich, C. A.; Vela, A.; Vignale, G.; Wesolowski, T. A.; Xu, X.; Yang, W. DFT Exchange: Sharing Perspectives on the Workhorse of Quantum Chemistry and Materials Science. *Phys. Chem. Chem. Phys.* **2022**, *24* (47), 28700–28781. https://doi.org/10.1039/D2CP02827A.

(54) Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User*. *Aust. J. Chem.* **2019**, *72* (8), 563–573. https://doi.org/10.1071/CH19023.

(55) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

(56) Podewitz, M. Towards Predictive Computational Catalysis – a Case Study of Olefin Metathesis with Mo Imido Alkylidene N-Heterocyclic Carbene Catalysts. In *Chemical Modelling: Volume 17*; Bahmann, H., Tremblay, J. C., Eds.; The Royal Society of Chemistry, 2022; Vol. 17, p 0. https://doi.org/10.1039/9781839169342-00001.

(57) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K. AMBER20. *Univ. Calif. San Franc.* **2020**.

(58) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60* (7), 1271–1277. https://doi.org/10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W.

(59) Chiacchio, M. A.; Legnani, L.; Fassi, E. M. A.; Roda, G.; Grazioso, G. Development of AMBER Parameters for Molecular Simulations of Selected Boron-Based Covalent Ligands. *Molecules* **2023**, *28* (6), 2866. https://doi.org/10.3390/molecules28062866.

(60) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open

Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17* (10), 6262–6280. https://doi.org/10.1021/acs.jctc.1c00571.

(61) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236. https://doi.org/10.1021/ja9621760.

(62) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(63) Brandt, F.; Jacob, C. R. Efficient Automatic Construction of Atom-Economical QM Regions with Point-Charge Variation Analysis. *Phys. Chem. Chem. Phys.* **2023**, *25* (20), 14484–14495. https://doi.org/10.1039/D3CP01263H.

(64) Csizi, K.-S.; Reiher, M. Universal QM/MM Approaches for General Nanoscale Applications. *WIREs Comput. Mol. Sci.* **2023**, *13* (4), e1656. https://doi.org/10.1002/wcms.1656.

(65) Maseras, F.; Morokuma, K. IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16* (9), 1170–1179. https://doi.org/10.1002/jcc.540160911.

(66) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100* (50), 19357–19363. https://doi.org/10.1021/jp962071j.

(67) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH3Cl + Cl− Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7* (6), 718–730. https://doi.org/10.1002/jcc.540070604.

(68) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733. https://doi.org/10.1002/jcc.540110605.

(69) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6* (12), 3721–3734. https://doi.org/10.1021/ct1003803.

(70) Poulsen, T. D.; Kongsted, J.; Osted, A.; Ogilby, P. R.; Mikkelsen, K. V. The Combined Multiconfigurational Self-Consistent-Field/Molecular Mechanics Wave Function Approach. *J. Chem. Phys.* **2001**, *115* (6), 2393–2400. https://doi.org/10.1063/1.1374559.

(71) Rode, B. M.; Hofer, T. S.; Randolf, B. R.; Schwenk, C. F.; Xenides, D.; Vchirawongkwin, V. Ab Initio Quantum Mechanical Charge Field (QMCF) Molecular Dynamics: A QM/MM – MD Procedure for Accurate Simulations of Ions and Complexes. *Theor. Chem. Acc.* **2006**, *115* (2), 77–85. https://doi.org/10.1007/s00214-005-0049-1.

(72) Assfeld, X.; Rivail, J.-L. Quantum Chemical Computations on Parts of Large Molecules: The Ab Initio Local Self Consistent Field Method. *Chem. Phys. Lett.* **1996**, *263* (1), 100–106. https://doi.org/10.1016/S0009-2614(96)01165-7.

(73) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102* (24), 4714–4721. https://doi.org/10.1021/jp9809890.

(74) Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20* (14), 1468–1494. https://doi.org/10.1002/(SICI)1096-987X(19991115)20:14<1468::AID-JCC2>3.0.CO;2-0.

(75) Rodríguez, A.; Oliva, C.; González, M.; van der Kamp, M.; Mulholland, A. J. Comparison of Different Quantum Mechanical/Molecular Mechanics Boundary Treatments in the Reaction of the Hepatitis C Virus NS3 Protease with the NS5A/5B Substrate. *J. Phys. Chem. B* **2007**, *111* (44), 12909–12915. https://doi.org/10.1021/jp0743469.

(76) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(77) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(78) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72* (4), 2384–2393. https://doi.org/10.1063/1.439486.

(79) Hoover, W. G. Constant-Pressure Equations of Motion. *Phys. Rev. A* **1986**, *34* (3), 2499–2500. https://doi.org/10.1103/PhysRevA.34.2499.

(80) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(81) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929. https://doi.org/10.1063/1.1755656.

(82) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Bioinforma.* **1993**, *17* (4), 412–425. https://doi.org/10.1002/prot.340170408.

(83) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 015102. https://doi.org/10.1063/1.4811489.

(84) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9* (4), 2000–2009. https://doi.org/10.1021/ct300878a.

(85) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(86) Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* **1969**, *37* (3), 424–438. https://doi.org/10.2307/1912791.

(87) Toda, H. Y.; Yamamoto, T. Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *J. Econom.* **1995**, *66* (1), 225–250. https://doi.org/10.1016/0304-4076(94)01616-8.

(88) Dickey, D. A.; Fuller, W. A. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74* (366), 427–431. https://doi.org/10.2307/2286348.

(89) Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *J. Econom.* **1992**, *54* (1), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y.

(90) Johansen, S. Statistical Analysis of Cointegration Vectors. *J. Econ. Dyn. Control* **1988**, *12* (2), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3.

(91) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57* (1), 289–300.

(92) Polshettiwar, V.; Varma, R. S. Revisiting the Meerwein–Ponndorf–Verley Reduction: A Sustainable Protocol for Transfer Hydrogenation of Aldehydes and Ketones. *Green Chem.* **2009**, *11* (9), 1313–1316. https://doi.org/10.1039/B913079A.

(93) A. Leberre. *Bull. Soc. Chim. Fr.* **1961**, 1543–1549.

(94) Fitzner, M.; Wuitschik, G.; Koller, R. J.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us about Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11* (48), 13085–13093. https://doi.org/10.1039/D0SC04074F.

(95) Guzman-Percastegui, E.; Hernandez, D. J.; Castillo, I. Calix [8] Arene Nanoreactor for Cu (i)-Catalysed C–S Coupling. *Chem. Commun.* **2016**, *52* (15), 3111–3114.

12. Appendix

The appendix contains the publications associated with this manuscript and their respective supporting information.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent

R. A. Talmazan and M. Podewitz*

Cite This: J. Chem. Inf. Model. 2023, 63, 5400−5407

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We introduce PyConSolv, a freely available Python package that automates the generation of conformers of metal- and nonmetal-containing complexes in explicit solvent, through classical molecular dynamics simulations. Using a streamlined workflow and interfacing with widely used computational chemistry software, PyConSolv is an all-in-one tool for the generation of conformers in any solvent. Input requirements are minimal; only the geometry of the structure and the desired solvent in xyz (XMOL) format are needed. The package can also account for charged systems, by including arbitrary counterions in the simulation. A bonded model parametrization is performed automatically, utilizing AmberTools, ORCA, and Multiwfn software packages. PyConSolv provides a selection of preparametrized solvents and counterions for use in classical molecular dynamics simulations. We show the applicability of our package on a number of (transition-metal-containing) systems. The software is provided open source and free of charge.

■ INTRODUCTION

As quantum mechanical (QM) methods and computer hardware evolve and become more efficient, bigger systems,[1−3] that were previously far out of reach, can be tackled using quantum chemical wave function methods.[4,5] As the system size increases, it becomes necessary to consider greater flexibility, which in turn presents new challenges that need to be properly addressed.[6,7] One common approach is to generate conformational ensembles, which can be created in various ways.[8]

While the importance of conformers is well-known and a key feature in drug design pipelines,[9,10] it has gathered some attention in the field of homogeneous catalysis as early as the 1990s.[11−18] However, conformer searches are by no means widespread in standard mechanistic studies today. A systematic exploration of the conformational space, by generating rotamers, can be employed for simple systems, with few degrees of freedom, but quickly becomes unfeasible for larger molecules. While several more advanced methods for the generation of conformers exist, they are generally geared toward biological compounds.[19−22] Methods such as torsional Monte Carlo, augmented with Low Mode searching, offer a way to sample the conformational space, but are limited by computational costs when explicit solvent is necessary.[23,24] Alternatively, molecular dynamics (MD) methods can be employed for the generation of conformers, by running simulations of the system (in explicit solvent) and clustering the resulting trajectory.[25] For this purpose, ab initio MD (AIMD)[26] or classical MD (cMD), in various flavors, can be used, each with its own advantages and disadvantages, as outlined below.

For AIMD, the calculation of the electronic structure is generally considered to be the limiting factor for the overall sampling. To overcome this issue, the use of very efficient semiempirical methods, such as GFN2-xTB,[27] provides a good compromise. One tailor-made approach for conformer generation is realized in the Conformer Rotamer Ensemble

Received: May 25, 2023
Published: August 22, 2023

ACS Publications

https://doi.org/10.1021/acs.jcim.3c00798

Tool (CREST),[28] which is shown to perform very well for a large a number of systems,[29] yet it is not without its downsides. Due to the inherent limitations of semiempirical quantum chemical methods, GFN2-xTB and CREST do not perform well for typical transition-metal systems, and neither structures nor energies are reliable without any further processing.[30] Another caveat for CREST is the lack of explicit solvation. While implicit solvation models perform admirably for many complexes,[31,32] they fail to provide reasonable structures for systems containing a cavity,[15] as intramolecular bonds are heavily favored. This can be counteracted by using explicit solvation, either in microsolvation approaches,[33−35] that use a few solvent molecules, or in condensed phase calculations, but in any case, at the cost of massively increasing the computational requirements.

On the other hand, cMD simulations rely on a force-field approach, which leads to a great speedup in calculations. Consequently, it allows for explicit treatment of solvent molecules, of course, at some cost in accuracy. While many force fields exist, developed with either specific[36−40] or general use,[41−45] running a cMD simulation for metal-containing systems requires the generation of custom parameters for the system under study in order to obtain reasonable results. While there are tools available for the parametrization of organic molecules,[46] the parametrization of transition metal-containing complexes is a more involved process, requiring a special approach.[47−52] As metals can form very complex structures, with widely varying coordination numbers,[53] there are no predefined atom parameters available in the commonly used general force fields. A full parametrization of the metal center is therefore required. One method available in the AmberTools[54] suite, developed for parametrization of metal-containing biological systems, is the Metal Center Parameter Builder (MCPB).[55] MCPB utilized a bonded model[47] to describe the metal ion and its surrounding environment. This approach requires defining the bonded and nonbonded parameters of the force field, for the metal atom and its ligands. These parameters entail structural information, force constants, and potentials for bonds, angles, and dihedrals, as well as partial charges and van der Waals parameters. They are generated based on either experimental or QM optimized structures, by deriving their values from force constants and partial charges.[55] We would like to stress here that the force-field parameters are generated in such a way that the QM optimized structure is retained. While the use of a bonded model, as implemented in MCPB, provides accurate results for many different (transition-)metal-containing complexes,[14,15,47,55−57] the generation of parameters is a tedious and error prone process, due to the amount of user intervention that is required.

In this work, we present a user-friendly Python package, which builds upon AmberTools,[54] to provide an automated process for generating conformers of arbitrary, metal-containing, or metal-free complexes, in explicit solvent. The user only needs to provide an input structure, the desired QM method to be used for the geometry optimization and force constant calculations, the total system charge and multiplicity, and the solvent to be used to the simulation (if applicable). We provide full support for a large number of preparametrized solvents, as well as the ability to parametrize any solvent of choice. Additionally, as many metal-containing complexes are associated with a counterion, we provide seven preparametrized ions, as well as the ability to parametrize an ion of choice. By interfacing to ORCA 5,[58] the user has access to state-of-the-art QM methods for structure calculations. For system charge assignment, an interface to Multiwfn 3.8,[59] offers the user a plethora of charge calculations schemes, in addition to the Merz−Kollman RESP scheme[60,61] recommended by MCPB. After the parametrization is complete, the system is solvated and a suggested equilibration procedure is offered to the user, using the Amber MD package.[62,63] Once a simulation is performed, the resulting trajectory file can be analyzed via a script provided by PyConSolv. The clustering itself is based on the cpptraj package[64] and returns a set of conformers, ranked by their QM energy, based on single-point calculations.

■ METHODOLOGY

In this work, conformational sampling is performed for a number of structures in explicit solvent using PyConSolv to show the applicability as a proof-of-concept method. The systems of choice are a copper containing calix[8]arene catalyst for C−N coupling,[15] a molybdenum-based catalyst for olefin metathesis,[74,73] and a metal-free hydrogenobyric acid,[77] as well as the vitamin B_{12} metabolite methylcobalamin.[78]

Parametrization. The protocol implemented in the PyConSolv Python package is shown in Figure 1. The input required is a simple XMOL xyz-formatted molecular structure. Input files for ORCA 5 are generated automatically, and geometry optimization and subsequent frequency calculations are performed at the electronic structure theory level chosen by the user. The optimized geometry is taken as input for subsequent parametrization steps. As the metal center parametrization builds upon the MCPB.py package provided in AmberTools, the steps follow those recommended in the aforementioned package, deriving the force-field parameters using the well-established Seminario method.[65] The structure must be split into fragments which are to be parametrized separately, with each atom requiring a unique identifier in the pdb files. This step is automated by building a connectivity matrix based on atom radii and pair distances of the atoms. If two atoms are closer together than 60% of the sum of their atomic radii, they are considered to be bonded. If one of the atoms is a metal, the connectivity is checked, but it is not added to the matrix, as the metal needs to be parametrized separately from the rest of the fragments. Using a depth first search algorithm[66] to traverse the connectivity matrix, we are able to identify each individual fragment. The user is then presented with an interactive window, where a Lewis structure of each fragment is displayed. Here, we require the user to provide total fragment charges to perform the parametrization of each individual fragment using antechamber.[46] The ORCA 5 output files are analyzed and converted into inputs for Multiwfn 3.8 and MCPB, also accounting for the usage of effective core potentials[67] for heavy atoms. The RESP charges for the system are calculated using Multiwfn, with the recommended settings, as in the Multiwfn user manual (see SI). The MCPB input file is created, taking into account the carbon−metal bonds present in the system, which are not natively recognized. The metal center and all atoms directly bound to it have a new atom type assigned. The bonded parameters are derived from the Hessian matrix of the system using the Seminario method.[65] The previously calculated RESP charges are taken into account and enter the nonbonded part of the force-field parameters. For the ligands, the initial guesses for the parameters are based on modifications of the GAFF2

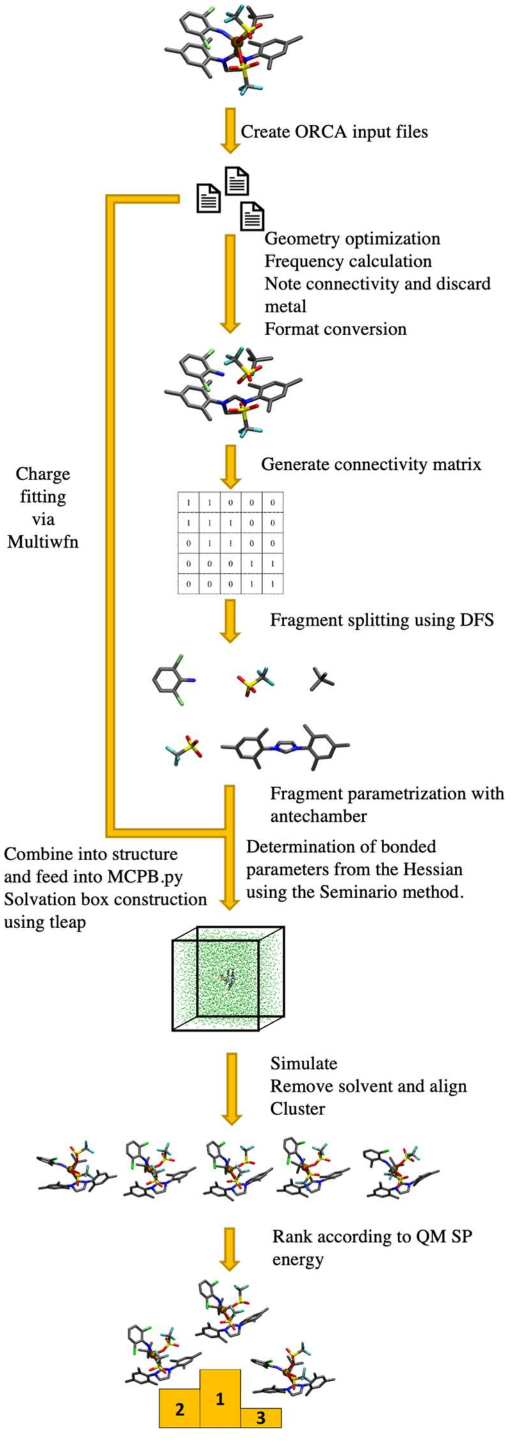


Figure 1. Automated PyConSolv workflow: Generation of the force-field parameters, setup of the simulation box, equilibration, production run, and analysis (for details, see manual in SI).

force-field parameters. Using the parameters generated by MCPB, a simulation box is then set up, with either one of the 18 preconfigured solvents or any other user-defined solvent. Subsequently, an equilibration script, which follows an extensive heating and cooling procedure as outlined by Wallnoefer et al.,[68] is provided, as well as an input file for a simulation of 100 ns in the NVT ensemble, with $T = 300$ K.

For a detailed guided workflow, please refer to the documentation in the SI or the online information available on GitHub (https://github.com/PodewitzLab/PyConSolv).

Supported Solvents and Ions. As part of the PyConSolv package, the solvents listed in Table S1 (SI) have been preparametrized and it has been verified that the equilibrated solvent box densities at 300 K are close to the experimentally determined densities. All parametrizations were performed on structures optimized with BP86/def2-SVP/D4,[69−72] with the exception of acetonitrile, for which no D4 corrections were used, as the resulting structure was incorrect. For parametrization, antechamber was used, with RESP charges.

In addition to solvents, the following common counterions have been parametrized: ScF_6^{3-}, BF_4^-, $B[Ar^F]_4^-$, $B(Ph)_4^-$, PF_6^-, OTf^-, and ClO_4^-. The structures were optimized with BP86/def2-SVP/D4[69−72] and all charges were computed with RESP. During parametrization for BF_4^- and ClO_4^-, the angle parameters could not be determined automatically with antechamber/MCPB.py. In these cases, a scan of the angle was performed with ORCA and the parameter were assigned manually.

■ ANALYSIS

After a cMD production run is completed, the resulting trajectory can be analyzed using PyConSolv. This can be done in two ways, either using one of the provided shell scripts or using the Python implementation present within PyConSolv. To obtain relevant conformers, the trajectory must be clustered. As we use a root-mean-square deviation (RMSD) of the distance of all heavy atoms as the clustering criteria, it is vital that the trajectory is properly aligned. This is resolved by aligning the complex based on a list of atom indices, provided by the user. The clustering can then be performed using one of the four methods provided by cpptraj: dbscan, hierarchical, k-means, or dpeaks.[64] On the resulting clusters, single-point calculations are performed, using the same electronic structure theory level that was used during the parametrization, including implicit solvation and dispersion corrections, as detailed in each test case. Finally, a list of clusters, ranked by the single-point energy, is presented to the user.

■ RESULTS

The PyConSolv workflow, as explained above, was applied to the following systems: a Cu(I)-calix[8]arene,[15] a Mo imido alkylidene N-heterocyclic carbene catalyst,[73,74] hydrogenobyric acid,[75−77] and methylcobalamin.[78] A suitable functional was chosen for each system, along with the appropriate solvent, as to mimic the experimental conditions. All charges were calculated using RESP. After parametrization, the system was equilibrated, and a 100 ns cMD production run was performed, using the simulation input file provided by PyConSolv. The clusters were generated using k-means, with the script provided by PyConSolv, using the RMSD of all nonhydrogen atoms as the distance metric. The simulations were aligned on the "rigid" parts of the structure, as described in each case below. A total of 10 clusters were chosen for each system and evaluated.

As a note about runtime, when comparing PyConSolv with CREST, it can be seen that with increase in system size, the PyConSolv workflow becomes noticeably faster, as the time-consuming step becomes the frequency calculation, rather than the cMD simulation (see SI Table S2 for timings).

Case 1: Cu(I)-Calix[8]arene. Cu(I)-phenanthroyl encapsulated by calix[8]arene is a catalyst that has proven to be very complex to model, with the macrocyclic cage being particularly mobile. This inherent flexibility needs to be taken into account, in order to obtain an accurate energy profile for the reaction[15] or to explain the difference in activity between various regioisomers.[14]

Using PyConSolv, with PBE0/def2-SVP/D3+CPCM-(Chloroform),[31,72,79,80] the system shown in Figure 2 was parametrized and solvated in chloroform. The trajectory was aligned on the copper and phenanthroline moieties. The clusters shown in Figure 2 represent those with the lowest energy, as computed via DFT single points, with the above specified density functional, basis set, and dispersion correction. For comparison, structures generated with CREST using GFN2-xTB and implicit solvation (chloroform) have been generated as well. From the representative cluster depicted in Figure 2, it can be seen that the calix[8]arene cage is collapsed around the copper center, when compared to the structures generated with PyConSolv. This is an artifact caused by the lack of explicit solvation in the CREST conformers and the tendency of dispersion corrections to favor compact structures with many intramolecular interactions. The collapsed structure is in stark contrast to those generated in explicit solvent via PyConSolv (see Figure 2, bottom left). As it is known from experiments that an intact cavity is vital for any catalysis to take place,[14,15] PyConSolv yields structures that are a better representation of the solution phase conformations.

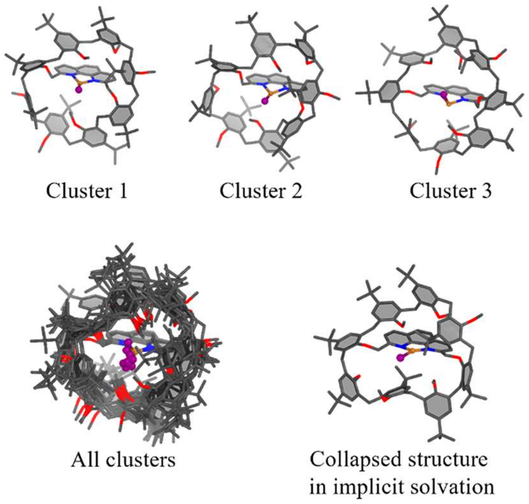


Figure 2. Cu(I)-calix[8]arene clusters. Upper panel: Top three clusters with lowest QM energy. Lower panel: All 10 clusters superimposed (left). The bottom right image depicts a collapsed cage structure in implicit solvation.

Case 2: Molybdenum Imido Alkylidene N-Heterocyclic Carbene (NHC) Catalyst. Molybdenum imido alkylidene N-heterocyclic carbene (NHC) catalysts represent one avenue of performing olefin metathesis.[73,74] It was shown during the investigation of the reaction mechanism, that conformer generation is crucial for explaining the reactivity.[16]

For the Mo-based catalyst, the structures generated with PyConSolv used BP86/def2-SVP/D3BJ+CPCM-(CH_2Cl_2)[31,69,70,72,81] for the force-field parametrization. The trajectory was aligned on the molybdenum, the carbon of the aryl group, and the NHC carbon bound to the molybdenum. It can be seen that the simulation captured the movement of the ligands, as shown in Figure 3. The triflate ligands show the most prominent flexibility. For comparison, structures were generated with CREST as well (see Figure S1, Supporting Information). The movement of the triflate groups is more pronounced in the CREST conformers, likely due to the enhanced sampling method implemented within its code. While similar enhanced sampling methods can easily be utilized in conjunction with PyConSolv, this is beyond the scope of this work. The NHC moiety, along with the aryl ligand, appear to be considerably more rigid. While the CREST conformers display a rotation in the NHC group, this is not present in the cMD simulation set up by PyConSolv. However, the underlying electronic structure method of CREST, GFN2-xTB, has shown to be less appropriate for this particular catalyst, and as such, the resulting structures require further refinement using DFT. In constrast, the Mo center coordination geometry in the force field is constrained to QM optimized one. Consequently, an adequate structure is conserved in the MD simulations at the cost that changes in the Mo coordination geometry cannot be captured. Yet, both sets of conformers look rather similar at first glance; hence, the solvent does not seem to have a great influence on the overall structure of the catalyst.

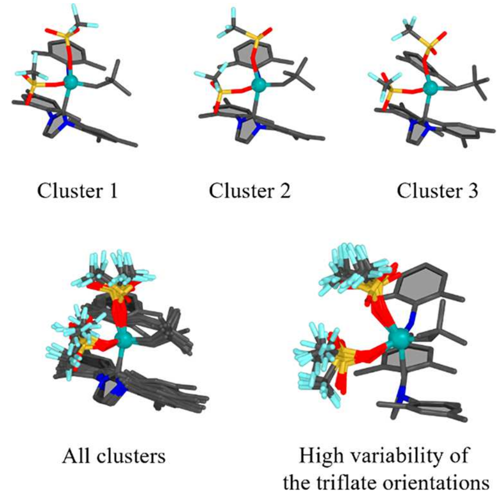


Figure 3. Mo imido alkylidene NHC catalyst clusters. Upper panel: Top three ranked clusters ranked by QM energy. Lower panel: All 10 clusters superimposed (left). The bottom right image highlights the high flexibility of the triflate groups.

Case 3: Hydrogenobyric Acid. Hydrogenobyric acid represents the metal-free precursor of vitamin B_{12} from which numerous native or artificial metal-cobalamins can be synthesized.[75−77] Its structural variability may prove to be of interest for the design of new drugs.[82,83] The electronic structure calculation for parametrization was performed with BP86/def2-SVP/D4+CPCM(Water)[31,69−72] and the trajectory was aligned on the nitrogen atoms of the corrin ring.

It can be seen from Figure 4 that the flexibility of the side chains is quite well captured, along with the flexibility of the corrin ring. The test case also illustrates that PyConSolv works equally well for metal-free systems.

Case 4: Methylcobalamin. Methylcobalamin is a vitamin B_{12} metabolite.[78] For the PyConSolv procedure, as in case 3, for parametrization, BP86/def2-SVP/D4+CPCM-(Water)[31,69−72] was the method of choice. The alignment was performed, as before, on the nitrogen atoms of the corrin ring.

It can be seen from the clusters in Figure 5 that the methylcobalamin structure is fairly rigid. The stabilization of the corrin ring is apparent when compared to that of the

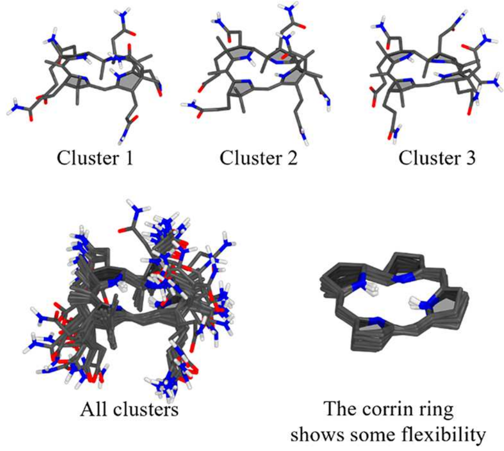


Figure 4. Hydrogenobyric acid clusters. Upper panel: Top three clusters with lowest QM energy. Lower panel: All 10 clusters superimposed (left) and a view of the corrin ring of the clusters (right).

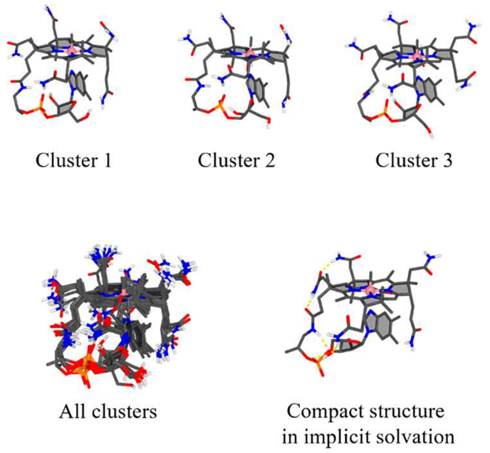


Figure 5. Methylcobalamin clusters. Upper panel: Top three clusters ranked by QM energy. Lower panel: All 10 clusters superimposed (left) and an example of a structure optimized in implicit solvent, dominated by intramolecular hydrogen bonding (right).

hydrogenobyric acid. It can be seen that all amide groups show some flexibility, while always pointing toward the solvent.

For sake of comparison, investigation of the complex in implicit solvent yielded a structure with many intramolecular hydrogen bonds of the amides and a loop conformation that is very different to the ones in explicit solvent (see Figure 5 bottom right; Figure S2, Supporting Information). These intramolecular hydrogen bonds may affect the reactivity of the complex, and it is unlikely that they are formed under experimental conditions. Consequently, PyConSolv provides more realistic liquid phase structures here.

Conformer Evaluation. The conformations generated by clustering of the trajectory are subjected to single-point energy calculations at the same level of theory as the parametrization to determine a ranking. Although this is a simplistic approach, it provides a good first guess as to which structures should be used for further refinement. While this is a fast method, it ignores the effects that explicit solvation might have on the overall stability of the system. To account for this, one would have to move to more complicated and costly methods such as QM/MM embedding schemes or resort to microsolvation models. An interface to the FEBISS[34] microsolvation package is planned for the near future.

■ CONCLUSION

In this work, we present a Python package to automate conformer generation of (metal-containing) complexes in explicit solvent. Our tool performs parametrization of metal-containing structures and sets up a workflow for simulation and analysis. The automatization makes conformer generation accessible to the nonexpert user. It allows for a straightforward implementation of conformational sampling (in explicit solvent) in standard computational chemistry workflows, e.g., for the determination of reaction mechanisms. With increasing system size, our Python package outperforms semiempirical methods, such as CREST. A subsequent reoptimization of the obtained clusters with DFT ensures that true energy minima are found.

While standard classical MD simulations can sometimes have difficulties in exhaustively exploring the potential energy surface, especially when the structure is trapped in a deep minimum, this obstacle can be overcome by resorting to enhanced sampling methods such as accelerated molecular dynamics[84] or metadynamics.[85] Alternatively, a more complex approach of Monte Carlo with Low Mode search, can be combined with MD to speed up the PES exploration. These methods add considerable complexity and are not currently implemented in PyConSolv. Another extension would be the implementation of MM methods, which would allow for the changes of the coordination around the metal center,[86,87] to switch between different coordination geometries to better reflect the nature of the metal atoms.

Nevertheless, the presented case studies show that conformer generation in explicit solvent leads to more realistic representations of the liquid phase structure. Our tool is a stepping stone toward more realistic modeling of reaction mechanisms and thus a prerequisite for calculations to become predictive.

■ ASSOCIATED CONTENT

Data Availability Statement

The PyConSolv package is available open source, free of charge, on GitHub: https://github.com/PodewitzLab/PyConSolv. Likewise, it can be installed from the popular PyPi software repository using pip install PyConSolv.

SI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c00798.

> Details of computational methodology including settings used for Multwfn, CREST, and solvent parametrization, as well as for the equilibration of MD simulations with AMBER; additional comparison between implicit and explicit solvent conformers; tutorial on how to use PyConSolv (PDF)

■ AUTHOR INFORMATION

Corresponding Author

M. Podewitz − *Institute of Materials Chemistry, TU Wien, A-1060 Wien, Austria;* orcid.org/0000-0001-7256-1219; Email: maren.podewitz@tuwien.ac.at

Author

R. A. Talmazan − *Institute of Materials Chemistry, TU Wien, A-1060 Wien, Austria*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.3c00798

Funding

Open Access is funded by the Austrian Science Fund (FWF).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the FWF for financial support (P33528), project granted to M.P.

■ REFERENCES

(1) Bowler, D. R.; Miyazaki, T. Calculations for Millions of Atoms with Density Functional Theory: Linear Scaling Shows Its Potential. *J. Phys.: Condens. Matter* **2010**, *22*, 074207.

(2) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chem., Int. Ed.* **2018**, *57* (16), 4170−4176.

(3) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50* (3), 539−543.

(4) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148* (1), No. 011101.

(5) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139* (13), 134101.

(6) Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Körner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M. A Real-World Perspective on Molecular Design: Miniperspective. *J. Med. Chem.* **2016**, *59* (9), 4087−4102.

(7) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem. Int. Ed.* **2022**, *61* (42), e202205735.

(8) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57* (8), 1747−1756.

(9) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035−4061.

(10) Shim, J.; MacKerell, A. D., Jr. Computational Ligand-Based Rational Design: Role of Conformational Sampling and Force Fields in Model Development. *Med. Chem. Commun.* **2011**, *2* (5), 356.

(11) Besora, M.; Braga, A. A. C.; Ujaque, G.; Maseras, F.; Lledós, A. The Importance of Conformational Search: A Test Case on the Catalytic Cycle of the Suzuki−Miyaura Cross-Coupling. *Theor. Chem. Acc.* **2011**, *128* (4−6), 639−646.

(12) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9* (8), 6803−6813.

(13) Eisenstein, O.; Ujaque, G.; Lledós, A. What Makes a Good (Computed) Energy Profile? In *New Directions in the Modeling of Organometallic Reactions*; Lledós, A., Ujaque, G., Eds.; Topics in Organometallic Chemistry; Springer International Publishing: Cham, 2020; pp 1−38. DOI: 10.1007/3418_2020_57.

(14) Berlanga-Vázquez, A.; Talmazan, R. A.; Reyes-Mata, C. A.; Percástegui, E. G.; Flores-Alamo, M.; Podewitz, M.; Castillo, I. Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]Arene C−S Coupling. *Eur. J. Inorg. Chem.* **2023**, *26*, e202200596.

(15) Talmazan, R. A.; Refugio Monroy, J.; del Río-Portilla, F.; Castillo, I.; Podewitz, M. Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]Arene Supramolecular Catalyst. *ChemCatChem.* **2022**, *14* (20), e202200662.

(16) Podewitz, M.; Sen, S.; Buchmeiser, M. R. On the Origin of E-Selectivity in the Ring-Opening Metathesis Polymerization with Molybdenum Imido Alkylidene N-Heterocyclic Carbene Complexes. *Organometallics* **2021**, *40* (15), 2478−2488.

(17) Norrby, P.-O.; Kolb, H. C.; Sharpless, K. B. Toward an Understanding of the High Enantioselectivity in the Osmium-Catalyzed Asymmetric Dihydroxylation. 2. A Qualitative Molecular Mechanics Approach. *J. Am. Chem. Soc.* **1994**, *116* (19), 8470−8478.

(18) Maseras, F.; Morokuma, K. IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16* (9), 1170−1179.

(19) Kolossváry, I.; Guida, W. C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *J. Am. Chem. Soc.* **1996**, *118* (21), 5011−5019.

(20) Cole, J. C.; Korb, O.; McCabe, P.; Read, M. G.; Taylor, R. Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *J. Chem. Inf. Model.* **2018**, *58* (3), 615−629.

(21) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019**, *59* (2), 731−742.

(22) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50* (4), 534−546.

(23) Parish, C.; Lombardi, R.; Sinclair, K.; Smith, E.; Goldberg, A.; Rappleye, M.; Dure, M. A Comparison of the Low Mode and Monte Carlo Conformational Search Methods. *J. Mol. Graph. Model.* **2002**, *21* (2), 129−150.

(24) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching. *J. Am. Chem. Soc.* **1990**, *112* (4), 1419−1427.

(25) Gaalswyk, K.; Rowley, C. N. An Explicit-Solvent Conformation Search Method Using Open Software. *PeerJ.* **2016**, *4*, e2088.

(26) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, 1st ed.; Cambridge University Press, 2009. DOI: 10.1017/CBO9780511609633.

(27) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652−1671.

(28) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22* (14), 7169−7192.

(29) Bursch, M.; Hansen, A.; Pracht, P.; Kohn, J. T.; Grimme, S. Theoretical Study on Conformational Energies of Transition Metal Complexes. *Phys. Chem. Chem. Phys.* **2021**, *23* (1), 287−299.

(30) Podewitz, M. Towards Predictive Computational Catalysis − a Case Study of Olefin Metathesis with Mo Imido Alkylidene N-Heterocyclic Carbene Catalysts. In *Chemical Modelling*; Bahmann, H., Tremblay, J. C., Eds.; The Royal Society of Chemistry, 2022; Vol. *17*, p 1−23. DOI: 10.1039/9781839169342-00001.

(31) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995−2001.

(32) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17* (7), 4250−4261.

(33) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic Microsolvation Approach with a Cluster-Continuum Scheme and

Conformational Sampling. *J. Comput. Chem.* **2020**, *41* (12), 1144−1155.

(34) Steiner, M.; Holzknecht, T.; Schauperl, M.; Podewitz, M. Quantum Chemical Microsolvation by Automated Water Placement. *Molecules* **2021**, *26* (6), 1793.

(35) Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S. Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods. *J. Chem. Theory Comput.* **2022**, *18* (5), 3174−3189.

(36) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528−552.

(37) Dickson, C. J.; Walker, R. C.; Gould, I. R. Lipid21: Complex Lipid Membrane Simulations with AMBER. *J. Chem. Theory Comput.* **2022**, *18* (3), 1726−1736.

(38) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114−4127.

(39) Madarász, Á.; Berta, D.; Paton, R. S. Development of a True Transition State Force Field from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2016**, *12* (4), 1833−1844.

(40) Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Commun.* **2018**, *54* (60), 8294−8311.

(41) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(42) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586−3616.

(43) van Gunsteren, W. F.; Daura, X.; Mark, A. E. GROMOS Force Field. In *Encyclopedia of Computational Chemistry*; von Rague Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., Schreiner, P. R., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2002. DOI: 10.1002/0470845015.cga011

(44) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.

(45) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.

(46) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247−260.

(47) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935−2947.

(48) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for + 2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733−2748.

(49) Li, P.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10* (1), 289−297.

(50) Aaqvist, J.; Warshel, A. Free Energy Relationships in Metalloenzyme-Catalyzed Reactions. Calculations of the Effects of Metal Ion Substitutions in Staphylococcal Nuclease. *J. Am. Chem. Soc.* **1990**, *112* (8), 2860−2868.

(51) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125* (50), 15671−15682.

(52) Quinn, T. R.; Patel, H. N.; Koh, K. H.; Haines, B. E.; Norrby, P.-O.; Helquist, P.; Wiest, O. Automated Fitting of Transition State Force Fields for Biomolecular Simulations. *PLoS One* **2022**, *17* (3), No. e0264960.

(53) Frenking, G.; Fröhlich, N. The Nature of the Bonding in Transition-Metal Compounds. *Chem. Rev.* **2000**, *100* (2), 717−774.

(54) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Panano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A.*AMBER 2020*; University of California: San Francisco, 2020.

(55) Li, P.; Merz, K. M. MCPB.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56* (4), 599−604.

(56) Kastner, D. W.; Nandy, A.; Mehmood, R.; Kulik, H. J. Mechanistic Insights into Substrate Positioning That Distinguish Non-Heme Fe(II)/α-Ketoglutarate-Dependent Halogenases and Hydroxylases. *ACS Catal.* **2023**, *13* (4), 2489−2501.

(57) Piskorz, T. K.; Martí-Centelles, V.; Young, T. A.; Lusby, P. J.; Duarte, F. Computational Modeling of Supramolecular Metallo-Organic Cages−Challenges and Opportunities. *ACS Catal.* **2022**, *12*, 5806−5826.

(58) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108.

(59) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580−592.

(60) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5* (2), 129−145.

(61) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269−10280.

(62) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542−1555.

(63) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878−3888.

(64) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084−3095.

(65) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60* (7), 1271−1277.

(66) Tarjan, R. Depth-First Search and Linear Graph Algorithms. In *12th Annual Symposium on Switching and Automata Theory (swat 1971)*; IEEE: East Lansing, MI, USA, 1971; pp 114−121. DOI: 10.1109/SWAT.1971.10

(67) Kahn, L. R.; Goddard, W. A., III Ab Initio Effective Potentials for Use in Molecular Calculations. *J. Chem. Phys.* **1972**, *56* (6), 2685−2701.

(68) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A Challenging System: Free Energy Prediction for Factor Xa. *J. Comput. Chem.* **2011**, *32* (8), 1743−1752.

(69) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098−3100.

(70) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33* (12), 8822−8824.

(71) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *J. Chem. Phys.* **2019**, *150* (15), 154122.

(72) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297−3305.

(73) Herz, K.; Podewitz, M.; Stöhr, L.; Wang, D.; Frey, W.; Liedl, K. R.; Sen, S.; Buchmeiser, M. R. Mechanism of Olefin Metathesis with Neutral and Cationic Molybdenum Imido Alkylidene N-Heterocyclic Carbene Complexes. *J. Am. Chem. Soc.* **2019**, *141* (20), 8264−8276.

(74) Buchmeiser, M. R.; Sen, S.; Unold, J.; Frey, W. N-Heterocyclic Carbene, High Oxidation State Molybdenum Alkylidene Complexes: Functional-Group-Tolerant Cationic Metathesis Catalysts. *Angew. Chem., Int. Ed.* **2014**, *53* (35), 9384−9388.

(75) Kieninger, C.; Wurst, K.; Podewitz, M.; Stanley, M.; Deery, E.; Lawrence, A. D.; Liedl, K. R.; Warren, M. J.; Kräutler, B. Replacement of the Cobalt Center of Vitamin B12 by Nickel: Nibalamin and Nibyric Acid Prepared from Metal-Free B_{12} Ligands Hydrogenobalamin and Hydrogenobyric Acid. *Angew. Chem. Int. Ed.* **2020**, *59* (45), 20129−20136.

(76) Kieninger, C.; Baker, J. A.; Podewitz, M.; Wurst, K.; Jockusch, S.; Lawrence, A. D.; Deery, E.; Gruber, K.; Liedl, K. R.; Warren, M. J.; Kräutler, B. Zinc Substitution of Cobalt in Vitamin B_{12}: Zincobyric Acid and Zincobalamin as Luminescent Structural B12-Mimics. *Angew. Chem. Int. Ed.* **2019**, *58* (41), 14568−14572.

(77) Kieninger, C.; Deery, E.; Lawrence, A. D.; Podewitz, M.; Wurst, K.; Nemoto-Smith, E.; Widner, F. J.; Baker, J. A.; Jockusch, S.; Kreutz, C. R.; Liedl, K. R.; Gruber, K.; Warren, M. J.; Kräutler, B. The Hydrogenobyric Acid Structure Reveals the Corrin Ligand as an Entatic State Module Empowering B_{12} Cofactors for Catalysis. *Angew. Chem. Int. Ed.* **2019**, *58* (31), 10756−10760.

(78) Kräutler, B. 2.11 - Cobalt Enzymes. In *Comprehensive Inorganic Chemistry III*, Third ed.; Reedijk, J., Poeppelmeier, K. R., Eds.; Elsevier: Oxford, 2023; pp 268−301. DOI: 10.1016/B978-0-12-823144-9.00146-1

(79) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0Model. *J. Chem. Phys.* **1999**, *110* (13), 6158−6170.

(80) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.

(81) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456−1465.

(82) Clardy, S. M.; Allis, D. G.; Fairchild, T. J.; Doyle, R. P. Vitamin B_{12} in Drug Delivery: Breaking through the Barriers to a B12 Bioconjugate Pharmaceutical. *Expert Opin. Drug Delivery* **2011**, *8* (1), 127−140.

(83) Shell, T. A.; Lawrence, D. S. Vitamin B_{12}: A Tunable, Long Wavelength, Light-Responsive Platform for Launching Therapeutic Agents. *Acc. Chem. Res.* **2015**, *48* (11), 2866−2874.

(84) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919−11929.

(85) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562−12566.

(86) Schmid, M. H.; Das, A. K.; Landis, C. R.; Meuwly, M. Multi-State VALBOND for Atomistic Simulations of Hypervalent Molecules, Metal Complexes, and Reactions. *J. Chem. Theory Comput.* **2018**, *14* (7), 3565−3578.

(87) Foscato, M.; Deeth, R. J.; Jensen, V. R. Integration of Ligand Field Molecular Mechanics in Tinker. *J. Chem. Inf. Model.* **2015**, *55* (6), 1282−1290.

Supporting Information:

PyConSolv:
A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent

R. A. Talmazan[1] and M. Podewitz[1]*

[1] Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Wien (Austria), maren.podewitz@tuwien.ac.at

Multiwfn charge settings

Multiwfn 3.8[1] is used to compute the charges for the parametrizations, as follows.

Nuclear + Electronic Merz-Kollman RESP[1] type atomic charge mode is used with 6 points/Å^2. All atoms are used for fitting 4 layers, with the scaling factors: 1.4,1.6,1.8,2.0. Automatic radii are utilized for fitting, missing radii are taken from UFF and scaled by 1/1.2. For fitting, the tightness parameter is 0.1, the restraint strengths for the two stages are set to 0.0005 and 0.0010, respectively. The maximum number of iterations is 50 and the convergence threshold is set to 10^{-6}. Charge equivalence constraints are enabled for CH_2 and CH_3 groups. No charge constraint settings are used and the connectivity is determined automatically.

CREST settings

CREST[2] was used for generating conformations as comparisons for the chosen test cases. For this, the GFN2-xTB[3] method was chosen, using ALPB[4] as the solvent model, with an appropriate solvent for each system. The energy threshold, RMSD and conformer pair energy differences were set to 6 kcal/mol, 0.125Å and 0.05 kcal/mol, as detailed in the CREST documentation. For the case of the Mo-catalyst, the energy threshold was raised to 50 kcal/mol, because a low number of conformers was obtained.

Solvent Parametrization

The solvents were parametrized using RESP charges and antechamber, after which they were equilibrated and a short, 10ns, classical molecular dynamics (cMD) run was performed to examine the density.

Counterions. The counterions shown in Figure S1 ($B[Ar^F]_4^-$, BF_4^-, OTf^-, ScF_6^{3-}, ClO_4^-, PF_6^-, and $B(Ph)_4^-$) have been parametrized and are available in PyConSolv.

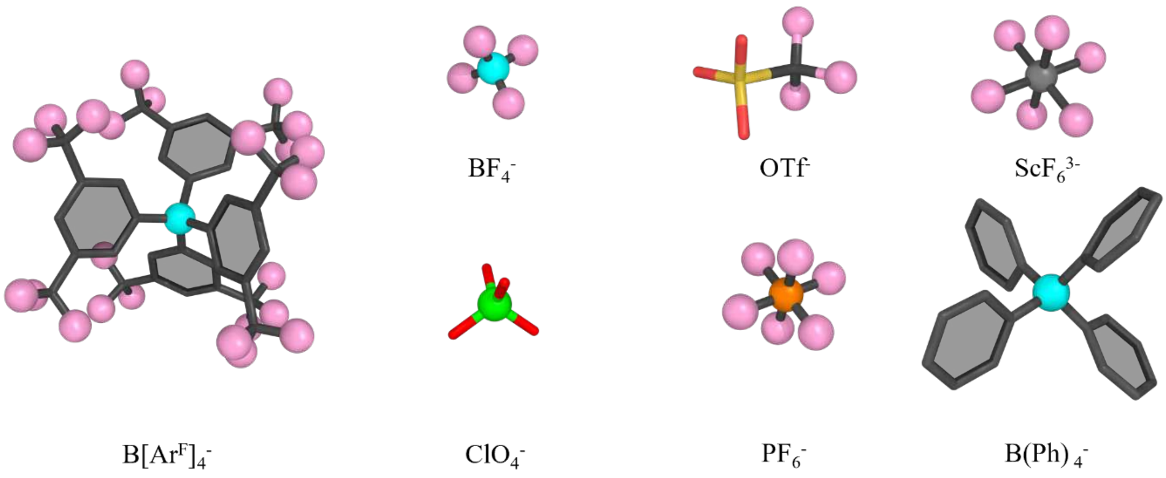


Figure S1 Structures of parametrized counterions; color coding: C atoms in grey, B atoms in cyan, F in pink, S in yellow, O in red, Sc in silver, P in orange.

Table S1. Density of boxes with parametrized solvent, after 10ns cMD. Densities taken from CRC Handbook of Chemistry and Physics, 95th Edition.[5] * denotes gas phase, while the calculated density is in the condensed phase.

Solvent	Exp. Density [g/cm^3]	Calc. Density [g/cm^3]	Density standard deviation [g/cm^3]
Acetone	0.7845	0.8160	0.0016
Acetonitrile	0.7857	0.8261	0.0013
Ammonia	0.696*	0.9523	0.0014
Benzene	0.8765	0.8519	0.0022
Tetrachloromethane	1.5940	1.5659	0.0036
Dichloromethane	1.3266	1.2617	0.0026
Chloroform	1.4788	1.4365	0.0029
Cyclohexane	0.7739	0.7496	0.0024
DMF	0.9445	0.9950	0.0020
DMSO	1.1010	1.1492	0.0022
Ethanol	0.7893	0.8142	0.0017
Hexane	0.6606	0.6339	0.0018
Methanol	0.7914	0.8171	0.0017
1-Octanol	0.8262	0.8180	0.0029
Pyridine	0.9819	0.9937	0.0022
THF	0.8833	0.8912	0.0019
Toluene	0.8623	0.8411	0.0025

Timings

The timing for the complete run of PyConSolv for each the testcase system is detailed in Table S2. The calculations were performed on a Ryzen R9 5900x CPU, using 12 threads and a total system memory of 64 GB. The simulations were performed using the GPU accelerated implementation of pmemd[6] on an RTX 3060 GPU.

Table S2. Timings for electronic structure calculations and simulation steps for PyConSolv and CREST.

	PyConSolv						CREST
System	SCF Cycle / h	Frequency Calculation / h	Antechamber / h	Equilibration / h	Simulation / h	Total / h	Total / h
Cu(I)-Calix[8]arene	0.11	17.84	3.08	0.67	9.01	30.71	483.32
Mo-NHC	0.05	1.07	0.08	0.33	8.45	9.98	10.68
Hydrogenobyric Acid	0.04	0.00	0.08	0.77	14.05	14.95	42.21
Methylcobalamin	0.14	5.91	0.50	1.75	22.53	30.83	423.52

Amber MD settings

To perform the classical molecular dynamics simulations, the Amber20[6] package was used, in combination with AmberTools23[6]. For the system equilibration, a multi-step approach with repeated heating and cooling was chosen, as follows:

1. The entire simulation box is restrained using cartesian restraints (restraint weight = 1000 kcal mol^{-1} Å^{-2}), with the exception of the hydrogen atoms, which undergo an energy minimization for 500 cycles.
2. The solute heavy atoms remain restrained and the whole system is minimized for 1000 cycles (restraint weight = 1000 kcal mol^{-1} Å^{-2}).
3. The system is equilibrated under constant volume to achieve a temperature of 300K, using Langevin dynamics[7] with the collision frequency of 2.0 ps^{-1}, a timestep of 0.001 ps and the SHAKE[8] algorithm for hydrogen bonds.
4. The solute heavy atoms remain restrained (restraint weight = 1000 kcal mol^{-1} Å^{-2}) and the whole system is equilibrated at constant pressure (1 atm), to achieve a reasonable system density and eliminate the voids created by the automated solvent placement in tleap. This is achieved using Langevin dynamics with the collision frequency of 2.0 ps^{-1}, a timestep of 0.001 ps and the SHAKE algorithm for hydrogen bonds.
5. The system is subsequently cooled under constant volume to achieve a temperature of 100K, using Langevin dynamics with the collision frequency of 2.0 ps^{-1}, a timestep of 0.001 ps and the SHAKE algorithm for hydrogen bonds.
6. A stepwise reduction of the restraints takes place, from 1000 to 0 kcal mol^{-1} Å^{-2}, minimizing at each step (1000, 500, 200, 100, 50, 20, 10, 5, 4, 3, 2, 1, 0.5 kcal mol^{-1} Å^{-2})
7. The system is heated up under constant volume to achieve a temperature of 300K, using Langevin dynamics with the collision frequency of 2.0 ps^{-1}, a timestep of 0.001 ps and the SHAKE algorithm for hydrogen bonds.

8. A final constant pressure (1 atm) equilibration takes place, using Langevin dynamics with the collision frequency of 2.0 ps^{-1}, a timestep of 0.001 ps and the SHAKE algorithm for hydrogen bonds.

The simulations then take place under constant pressure (1 atm, Berendsen barostat[9]) with isotropic position scaling, a pressure relaxation time of 2.0 ps with Langevin dynamics, SHAKE for the hydrogen atoms and a timestep of 0.001 ps.

The size of the simulation box is automatically detected based on the size of the system and a cubic shape is chosen in all instances. These settings resulted in the following number of particles for the simulations: Cu(I)-Calix[8]arene in chloroform: 966; molybdenum N-heterocyclic carbene (NHC) catalyst in dichloromethane: 844; hydrogenobyric acid in water: 5529; methylcobalamine in water: 6766.

Case 2: Molybdenum N-heterocyclic carbene (NHC) catalyst

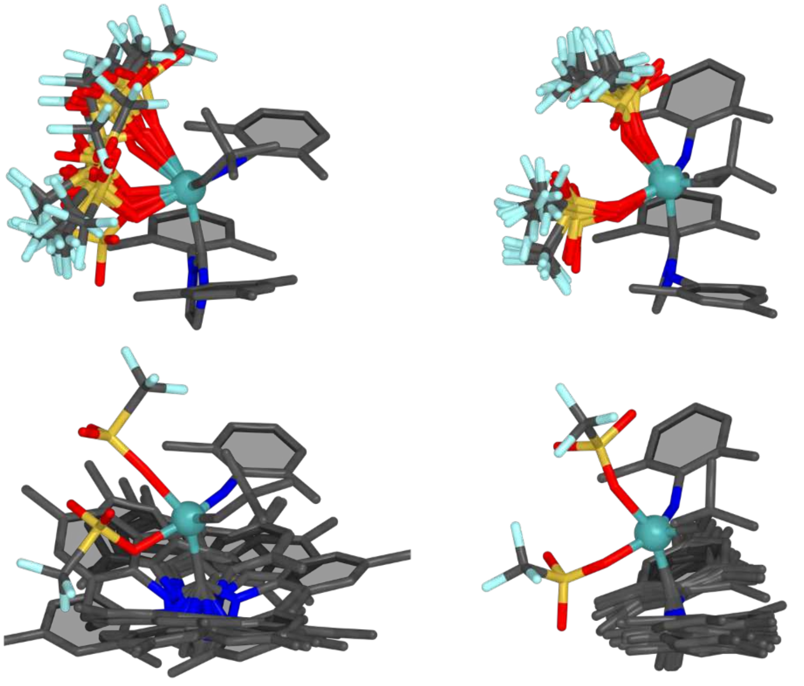

CREST Conformers PyConSolv Conformers

Figure S2 Comparison of conformers generated with CREST and PyConSolv.

In the case of the Molybdenum catalyst, it is apparent that CREST[2] achieves a greater sampling for the conformational space, thanks to its metadynamics based approach. PyConSolv can achieve further sampling through prolonging the classical molecular dynamics simulations or employing enhanced sampling techniques.

Case 4: Methylcobalamin

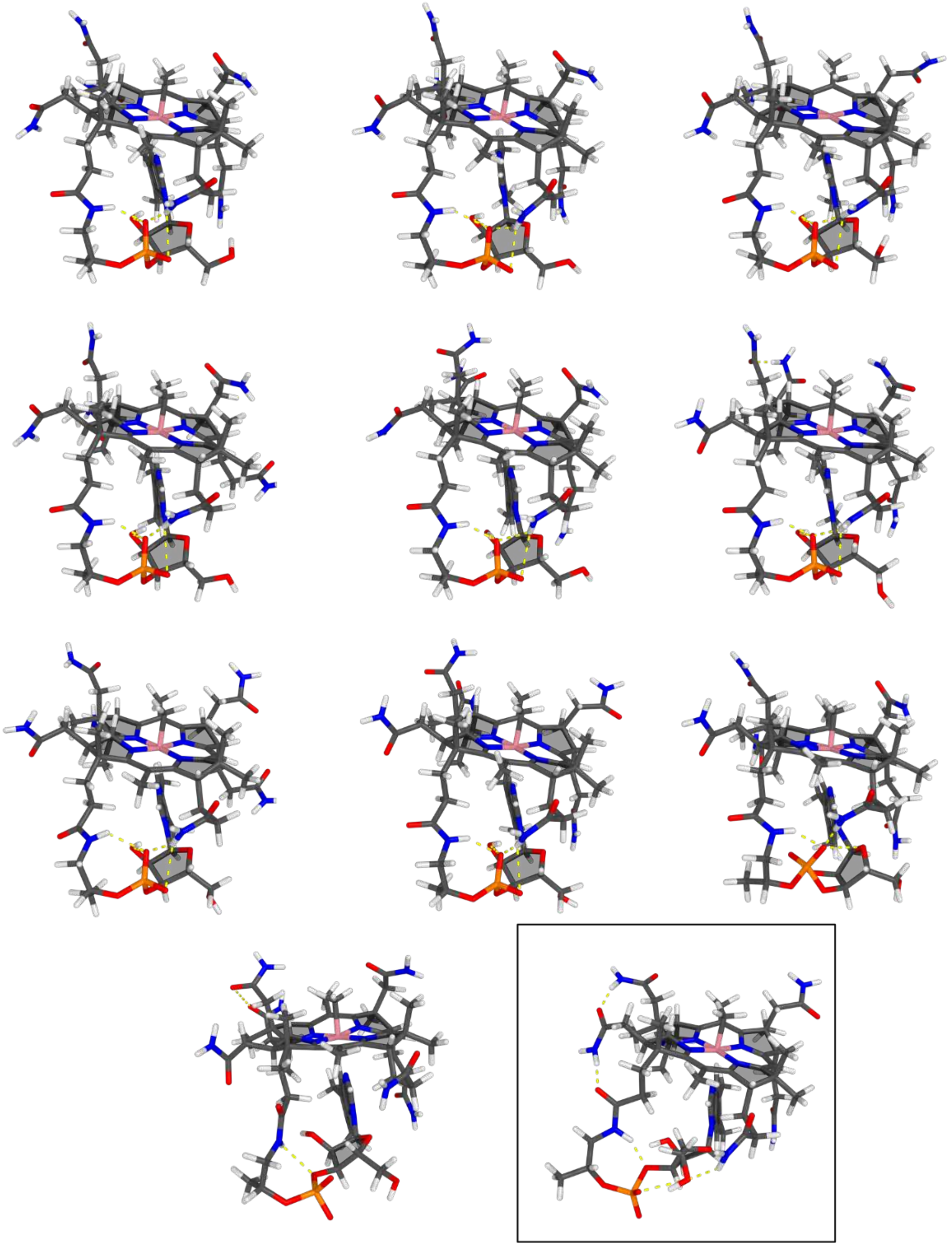


Figure S3. Conformers generated with PyConSolv, compared to a conformer with significant intramolecular hydrogen bonding from implicit solvation DFT optimizations (black rectangle).

Because of to the use of explicit solvation, unphysical intramolecular hydrogen bonding is minimized, when conformers are generated using PyConSolv, as shown in Figure S3.

Package Usage

The package was installed following the instructions on the github page(https://github.com/PodewitzLab/PyConSolv).

A new folder for the chosen system was created and the structure was saved as an XMOL file (input.xyz). This serves as the input for the pyconsolv command.

The parametrization was started using the following command: “pyconsolv input.xyz -m MMM -b BBB -s SSS -d DDD”, where MMM represents the DFT functional, BBB represents the basis set, SSS represents the solvent and DDD represents the dispersion correction model.

After the geometry optimization and frequency calculation are complete, a pop-up window appears, where the charges for each of the detected fragments are assigned (see Figure S4).

Following the charge assignment, the parameters are created, the system solvated and the equilibration is automatically started.

The simulation run is started in the newly created simulation folder, using the “run_simulation.sh” script, generated by PyConSolv.

Following the 100ns cMD production run, the analysis was performed using “pyconsolv -a sim-01”, choosing kmeans as the clustering method and default values (10 mininum points per cluster, 10 clusters total).

For a more detailed guide, please consult the user manual provided with PyConSolv, which can be found on the github repository.

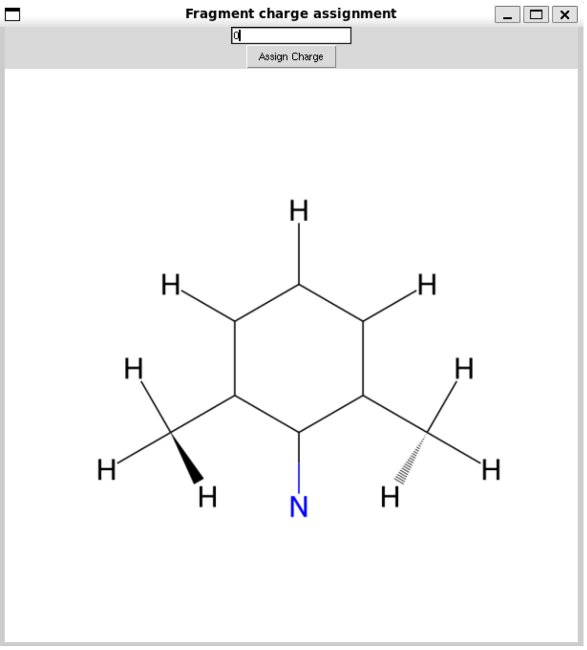


Figure S4 Fragment charge assignment dialogue. Here the user must provide the charges for each fragment of the initial molecule.

References

(1) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580–592. https://doi.org/10.1002/jcc.22885.

(2) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22* (14), 7169–7192. https://doi.org/10.1039/C9CP06869D.

(3) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

(4) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *J. Chem. Theory Comput.* **2021**, *17* (7), 4250–4261. https://doi.org/10.1021/acs.jctc.1c00471.

(5) Haynes, W. M. *CRC Handbook of Chemistry and Physics, 95th Edition*, 95th ed.; CRC Press: Hoboken, **2014**.

(6) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K. AMBER20. *University of California, San Francisco* **2020**.

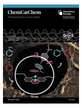
(7) Adelman, S. A.; Doll, J. D. Generalized Langevin Equation Approach for Atom/Solid-surface Scattering: General Formulation for Classical Scattering off Harmonic Solids. *J. Chem. Phys.* **2008**, *64* (6), 2375–2388. https://doi.org/10.1063/1.432526.

(8) Ciccotti, G.; Ryckaert, J. P. Molecular Dynamics Simulation of Rigid Molecules. *Computer Physics Reports* **1986**, *4* (6), 346–392. https://doi.org/10.1016/0167-7977(86)90022-5.

(9) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

ChemCatChem
Research Article
doi.org/10.1002/cctc.202200662
Chemistry Europe
European Chemical Societies Publishing

www.chemcatchem.org



Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]arene Supramolecular Catalyst

Radu A. Talmazan[+],[a, b] J. Refugio Monroy[+],[c, d] Federico del Río-Portilla,[c] Ivan Castillo,*[c] and Maren Podewitz*[a]

Development of C–N coupling methodologies based on Earth-abundant metals is a promising strategy in homogeneous catalysis for sustainable processes. However, such systems suffer from deactivation and low catalytic activity. We here report that encapsulation of Cu(I) within the phenanthroyl-containing calix[8]arene derivative 1,5-(2,9-dimethyl-1,10-phenanthroyl)-2,3,4,6,7,8-hexamethyl-*p*-tert-butylcalix[8]arene (**C$_8$PhenMe$_6$**) significantly enhances C–N coupling activity up to 92% yield in the reaction of aryl halides and aryl amines, with low catalyst loadings (2.5% mol). A tailored multiscale computational protocol based on Molecular Dynamics simulations and DFT investigations revealed an oxidative addition/reductive elimination process of the supramolecular catalyst [**Cu(C$_8$PhenMe$_6$)I**]. The computational investigations uncovered the origins of the enhanced catalytic activity over its molecular analogues: Catalyst deactivation through dimerization is prevented, and product release facilitated. Capturing the dynamic profile of the macrocycle and the impact of non-covalent interactions on reactivity allows for the rationalization of the behavior of the flexible supramolecular catalysts employed.

Introduction

C–N bond formation is a reaction with a large demand, as amine derivatives are essential building blocks for multiple biologically active molecules,[1] as well as play a key role in the design of high-performance materials.[2] In this regard, the development of novel, highly efficient catalysts based on Earth-abundant metals, such as copper or iron, is an essential strategy to explore environmentally more benign and sustainable production processes.

In recent years, computational chemistry has played a fundamental role in the development of new catalysts by uncovering reaction mechanisms and rationalizing reactivity and selectivity.[3] In fact, quantum chemical investigations have successfully been used not only to outline a mechanism, but also to understand the steric and electronic factors that govern catalytic systems, and to establish structure-activity relationships.[4] Consequently, complementing experimental efforts, computational studies can delineate the underlying mechanisms of the homogenous transition-metal catalyzed C–N bond formation – a crucial step to evolve from the main strategies based on palladium,[5] to extensions that employ Earth-abundant metals like copper.[6] However, appropriate computational protocols, adapted to the nature of the specific system, to obtain reliable results are required.

As Earth-abundant systems such as those with copper are prone to dimerization and other deactivation patterns, including oxidation from Cu(I) to Cu(II), recent efforts to incorporate it as catalytic center have resorted to bioinspired supramolecular catalysis.[7] Encapsulation of transition metals is a general strategy where a confined space or cavity provides a constrained chemical environment for the active species, while also imposing steric restrictions during the catalytic transformation.[8] The cavity allows efficient collisions between

---

[a] R. A. Talmazan,[+] Prof. M. Podewitz
Institute of Materials Chemistry
TU Wien
Getreidemarkt 9
1060 Vienna (Austria)
E-mail: maren.podewitz@tuwien.ac.at
Homepage: www.podewitz-lab.org
[b] R. A. Talmazan[+]
Institute of General, Inorganic, and Theoretical Chemistry and Center of Molecular Biosciences
University of Innsbruck
Innrain 80/82
6020 Innsbruck (Austria)
[c] J. Refugio Monroy,[+] Prof. F. del Río-Portilla, Prof. I. Castillo
Instituto de Química
Universidad Nacional Autónoma de México
Circuito Exterior
CU, Ciudad de México 04510 (México)
E-mail: joseivan@unam.mx
Homepage: https://iquimica.unam.mx/dr-ivan-castillo-perez
[d] J. Refugio Monroy[+]
Present address:
Department of Chemistry
Humboldt Universität zu Berlin
Brook-Taylor-Strasse 2
12489 Berlin (Germany)

[+] *Authors contributed equally.*

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cctc.202200662

This publication is part of the Catalysis Talents and part of the Catalysis in Latin America Special Collection. Please check the ChemCatChem homepage for more articles in the collection.

substrates and the anchored metal site, creating a high density of local interactions that usually results in beneficial effects such as higher activity and selectivity, imitating the behavior of metalloenzymatic active sites.[9]

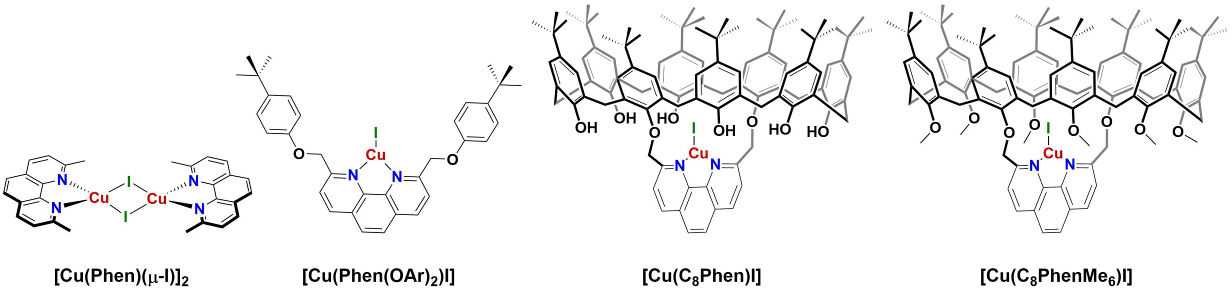
In this context, calix[n]arenes (n = 4–8) represent a group of cavitands that has been used as hosts for different metal centers including copper, aimed at emulating enzymatic catalysis.[10] However, the small calix[4,6]arenes tend to promote catalytic transformations that occur outside their cavities due to the limited confined space. Prior studies suggest that *p-tert*-butylcalix[8]arene (**C_8**) is an outstanding option for supramolecular catalysis.[11,12] Its larger and flexible cavity makes it amenable for metal ion encapsulation by using a coordinative moiety, while simultaneously allowing metal-substrate interactions inside its cavity. Despite these potential advantages, the number of catalytic systems based on **C_8** is significantly lower than those reported with the smaller calix[4,6]arenes.[11] One reason that hindered the rational development of supramolecular systems is the lack of an adequate cost-efficient computational methodology that sheds light on the intricate interplay of cavity conformations and active center accessibility. Such a methodology requires a description of the reaction mechanism at the catalytic center with electronic structure theory, while also incorporating the potentially vast conformational space of the macrocyclic ligand.

A promising candidate for macrocyclic encapsulation with calix[8]arene is [**Cu(Phen)(μ-I)**]$_2$ (Phen = 2,9-dimethyl-1,10-phenanthroline, Scheme 1). Its monomeric form has been shown to be active towards C–N coupling, even though with high ligand and metal loadings (10-20% mol), and only the most active substrates were converted with acceptable yields.[13] On the other hand, encapsulating Cu(I) with a bidentate phenanthroyl moiety in the recently reported **C_8**-based cavitand 1,5-(2,9-dimethyl-1,10-phenanthroyl)-*p-tert*-butylcalix[8]arene (**C_8Phen**)[14] allows for assembly of the catalytic system [**Cu(C_8Phen)I**] (Scheme 1). This supramolecular system exhibits remarkable catalytic C–S cross-coupling capability.[12] Consequently, exploring **C_8**-based encapsulation of Cu-phenanthroline complexes for C–N coupling reactions is an evident choice to develop highly efficient catalysts and to expand the range of reactions.

For a rational development of more efficient catalysts, insight into the underlying reaction mechanism is essential. C–N coupling reactions are known to follow different reaction pathways, involving either a single electron transfer (SET), an atom transfer mechanism with charged species, or an oxidative addition/reductive elimination sequence (see e.g., Ref 15a). These mechanistic possibilities have been investigated with computational methods, as well as experiments, with varying results.[3a,15,16] It appears that the nature of the catalyst, as well as the substrate, dictate the pathway,[16] with phenanthroline-based catalysts and aryl substrates generally following an oxidative addition/reductive elimination path.[16d] To identify the most viable mechanism, all alternatives need to be carefully evaluated in experimental and theoretical investigations.

In this context, the computational modelling of supramolecular systems including large calixarenes is not straightforward due to their intrinsic flexibility and hence, conformational diversity.[3e] This aspect must not be neglected to accurately assess the conformational space available to these types of catalysts. However, conformer sampling is not mature in transition-metal catalysis and even less established when flexible macrocyclic ligands are involved. The reason for it is the inherent difficulty to describe such compounds with standard conformer generators, and tailor-made approaches must be developed for these systems. To that end, enhanced sampling techniques such as accelerated molecular dynamics have proven successful to capture the conformational space of macrocycles, and this approach will be exploited here.[17]

Thus, we describe how encapsulation within a macrocyclic **C_8** cavitand significantly increases the C–N coupling activity of a Cu-phenanthroyl-derived catalyst, with CuI as an inexpensive and abundant metal source to assemble the supramolecular pre-catalyst [**Cu(C_8PhenMe$_6$)I**] (**C_8PhenMe$_6$** = 1,5-(2,9-dimethyl-1,10-phenanthroyl)-2,3,4,6,7,8-hexamethyl-*p*-tert-butyl-calix[8]arene, Scheme 1). The experimental results are corroborated with a tailored multiscale computational protocol to elucidate the reaction mechanism of C–N bond formation, mediated by [**Cu(C_8PhenMe$_6$)I**], and to determine the origin of its higher catalytic activity when compared to the molecular analogue [**Cu(Phen)(μ-I)**]$_2$. Our modelling approach is based on enhanced sampling techniques and quantum chemical investigations, while computational results are validated and complemented by extensive experimental characterization and activity studies.



[Cu(Phen)(μ-I)]$_2$ [Cu(Phen(OAr)$_2$)I] [Cu(C_8Phen)I] [Cu(C_8PhenMe$_6$)I]

Scheme 1. Molecular pre-catalysts [**Cu(Phen)(μ-I)**]$_2$ and [**Cu(Phen(OAr)$_2$)I**], as well as the parent supramolecular system **C_8**-based [**Cu(C_8Phen)I**], and the new methylated system [**Cu(C_8PhenMe$_6$)I**].

Results and Discussion

Synthesis and Characterization

In the new cavitand **C_8PhenMe$_6$**, the hydroxy groups from the parent **C_8Phen** were substituted by methoxy groups, which do not participate in acid-base equilibria under the basic conditions required for catalytic C–N coupling.[18] Complete methylation of the hydroxy groups of **C_8Phen** was achieved with a slight excess of MeI after deprotonation with NaH in anhydrous THF. The ^{1}H NMR spectrum of **C_8PhenMe$_6$** does not display signals beyond δ 8.50 ppm, which are associated with hydroxy groups in calix[8]arenes (Figure 1a).[18,19] Instead, the presence of two new singlets at δ 2.47 (6H, MeO$_{int}$) and 3.30 ppm (12H, MeO$_{ext}$) attest to the –OMe substitution. The methoxy groups are labelled as MeO$_{int}$, for those facing inside the calixarene, and MeO$_{ext}$, for those outside, based on the shielding or deshielding caused by the aromatic groups. Full NMR spectroscopic analysis of **C_8PhenMe$_6$** is presented in the Electronic Supporting Information (SI Figures S1–S4), including COSY NMR that allowed the assignment of signals corresponding to the methylene bridges of the cavitand CH$_{2C8}$ at δ 3.49 ppm (d, 4H, CH$_{2C8(d)}$), 3.59 (d, 4H, CH$_{2C8(b)}$), 3.98 (d, 4H, CH$_{2C8(a)}$), and 4.34 (d, 4H, CH$_{2C8(c)}$), while the methylene connector to the phenanthroyl moiety was detected at 4.27 ppm (s, 4H, CH$_{2Phen}$). Assignment of the signals was facilitated by Variable Temperature (VT) ^{1}H NMR spectroscopy (Figure 1b). The characteristic three-singlet pattern for the *tert*-butyl groups of a 1,5-substituted calix[8]arene was clearly discernible at δ 1.23, 1.14, and 0.90 ppm with a 1:1:2 ratio above 353 K; minor peaks likely correspond to different conformers (see computational section). The aromatic signals corresponding to the phenanthroyl moiety are well resolved at high temperature, being clearly identified at δ 8.06 and 7.55 ppm. The signals corresponding to the cavitand phenolic framework (Ar$_{C8}$) were observed from δ 7.55 to 6.65 ppm. The IR spectrum of the new cavitand in Figure S5 does not show bands above 3315 cm^{-1} that may be assigned to $\nu_{O\text{-}H}$, further supporting the complete methylation of the hydroxy groups. Positive-ion Fast Atom Bombardment Mass Spectrometry (FAB$^+$ MS) in Figure S6 shows an intense peak at $m/z = 1586$, assigned to the protonated molecular ion [**C_8PhenMe$_6$H**]$^+$. Finally, addition of CuI to an equimolar amount of **C_8PhenMe$_6$** in toluene suspension resulted in pre-catalyst [**Cu(C_8PhenMe$_6$)I**].

Characterization of [**Cu(C_8PhenMe$_6$)I**] was initially established by ^{1}H NMR spectroscopy and aided by VT ^{1}H NMR studies (Figures S7–S9 in the SI). The compound exhibits a symmetry plane as the free cavitand **C_8PhenMe$_6$**, based on the three *tert*-butyl signals in a 1:1:2 ratio, and the four doublets corresponding to the calix[8]arene methylene bridges CH$_{2C8}$ at δ 3.75 (m, 4H, CH$_{2C8(a)}$), 3.81 (m, 4H, CH$_{2C8(d)}$), 4.18 (m, 4H, CH$_{2C8(b)}$), and 4.21 (d, $J = 16$ Hz, 4H, CH$_{2C8(c)}$) observed at 393 K. The monomeric nature of [**Cu(C_8PhenMe$_6$)I**] was confirmed by DOSY ^{1}H NMR spectroscopy (Figure 2), where the free cavitand **C_8PhenMe$_6$** (Figure 2a) and the complex [**Cu(C_8PhenMe$_6$)I**] (Figure 2b) exhibit identical diffusion coefficients and hydrodynamic radii ($r_H = 24.90$ Å).[20] This observation suggests that the average size of **C_8PhenMe$_6$** is retained after CuI is anchored, indicating that **C_8PhenMe$_6$** may play a dual role as ligand: Confining the metal center (and potential substrates) within the cavity, while simultaneously preventing dimerization that deactivates other phenanthroline-based pre-catalysts (see also computational investigations, *vide infra*). Confirmation of the proposed identity of [**Cu(C_8PhenMe$_6$)I**] was established by FAB$^+$ MS, with a peak at $m/z = 1649$ assigned to [**Cu(C_8PhenMe$_6$)**]$^+$ (Figure 3a). Complex formation is facilitated by the lability of the Cu(I) ion,[21] with FAB$^+$ MS obtained after 30 min of stirring at 25 °C being identical to that displayed in Figure 3a. Furthermore, air stability of the pre-catalyst was confirmed through EPR spectroscopic studies at X-band frequency. [**Cu(C_8PhenMe$_6$)I**] is EPR-silent as expected for a diamagnetic Cu(I) d^{10} species (Figure 3b), and exposure of the isolated solid to air for 30 days resulted in

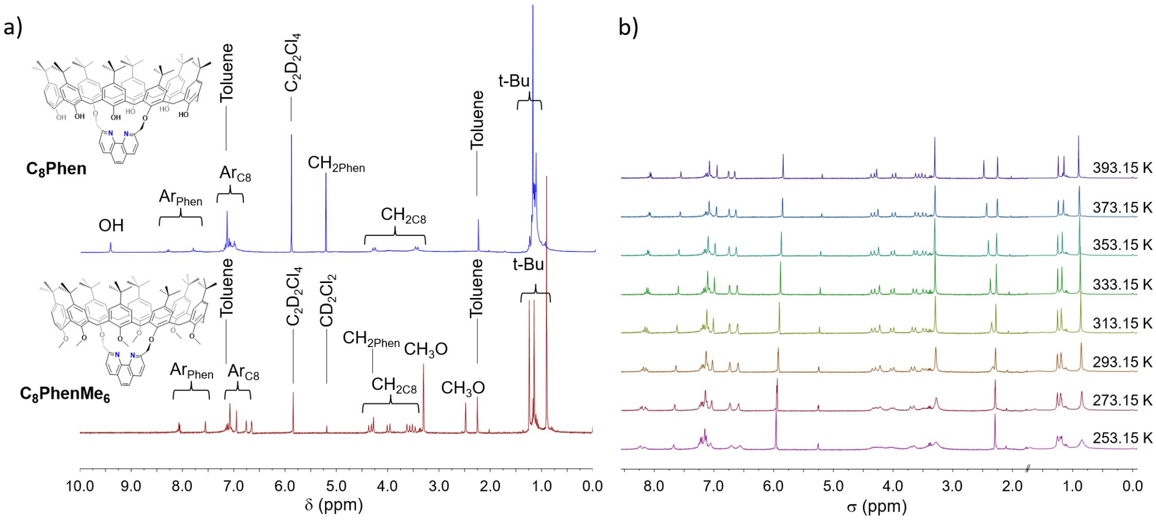


Figure 1. a) Comparison of ^{1}H NMR spectra of the parent cavitand **C_8Phen** (blue trace) and the new cavitand **C_8PhenMe$_6$** (red trace) in $C_2D_2Cl_4$. b) VT ^{1}H NMR spectra of **C_8PhenMe$_6$** in $C_2D_2Cl_4$ from 253.15 to 393.15 K.

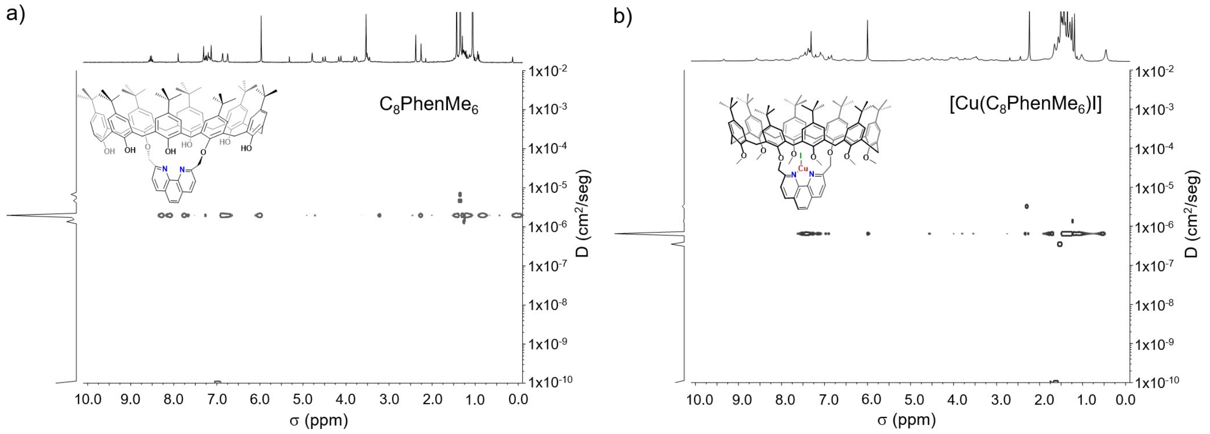


Figure 2. a) ^{1}H 500 MHz DOSY spectrum of 8 mM **C_8PhenMe$_6$** in 1,1,2,2-tetrachloroethane-d_2|toluene-d_8 mixture (500|50 μL) at 298 K, diffusion = 6.49×10^{-7} cm^2 s^{-1}, r_H = 24.90 Å; b) ^{1}H DOSY spectrum of 8 mM [**Cu(C_8PhenMe$_6$)I**] in 1,1,2,2-tetrachloroethane-d_2|toluene-d_8 at 298 K, diffusion = 6.49×10^{-7} cm^2 s^{-1}, r_H = 24.90 Å.

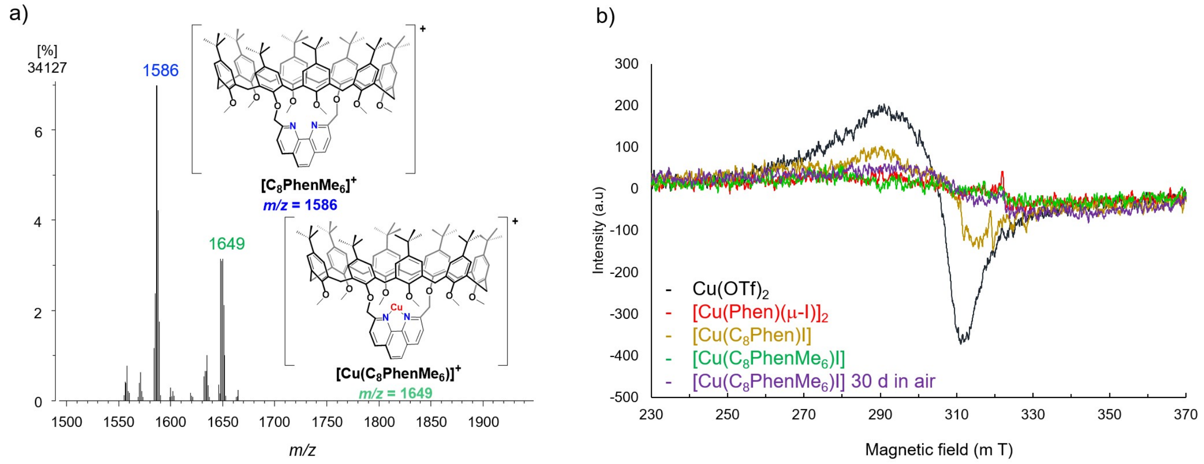


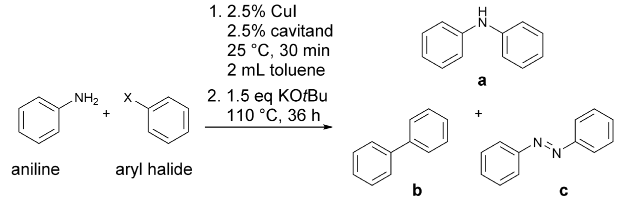
Figure 3. a) Experimental FAB$^+$ MS of [**Cu(C_8PhenMe$_6$)I**] showing peaks assigned to [**C_8PhenMe$_6$**]$^+$ and [**Cu(C_8PhenMe$_6$)**]$^+$; b) X-band EPR spectra of toluene suspensions of [**Cu(C_8PhenMe$_6$)I**] and reference compounds at 77 K.

virtually no change. In contrast, exposure to air for a few minutes resulted in significant signal intensity for [**Cu(C_8Phen)I**], consistent with oxidation to the EPR active Cu(II) species. [**Cu(Phen)(μ-I)**]$_2$ appears to be protected towards oxidation due to its dimeric nature;[22] Cu(OTf)$_2$ was employed as reference (OTf$^-$ = trifluoromethanesulfonate), see spectra in Figure 3b.

Catalytic Proof of Concept

As proof of concept to test the properties of the new cavitand **C_8PhenMe$_6$** in combination with Cu(I), the C–N coupling catalytic activity of the system was evaluated with 2.5 % loading of each. Aniline and bromobenzene in equimolar amounts were chosen as substrates as benchmark test; toluene was used as solvent, avoiding the commonly employed and high-boiling dimethylsulfoxide. Initial assessment of the reaction was carried out with several bases, KO*t*Bu affording the best results, as shown in Tables 1 and S1 in the SI. In a second set of evaluations, different sources of Cu(I) were tested to form the **C_8PhenMe$_6$**/Cu(I) catalytic system, with CuI performing better than CuBr or CuCl. Subsequently, the aryl halide substrate was varied, revealing that [**Cu(C_8PhenMe$_6$)I**] discriminates substrate by size: A significantly better performance was achieved for the production of diphenylamine (product **a**) when bromobenzene (PhBr) was used, rather than iodobenzene (PhI). Although iodide acts as a better leaving group that facilitates the potentially rate-limiting oxidative addition step (*vide infra*),[23] PhBr was determined to be a better substrate in related C–S couplings catalyzed by [**Cu(C_8Phen)X**] (X = Cl, I).[12] The isolated yield of product **a** with PhBr (73 %, entry 2 in Table 1) is significantly better than with PhI (56 %, entry 3 in Table 1). This result supports the notion that the substrates must access the confined space of the calixarene cavity to interact with the Cu(I) center. Thus, the rate of C–N coupling must be controlled by this steric restriction in combination with ease of oxidative addition of the aryl halide, since chlorobenzene (PhCl) did not react (Table 1, entry 1). Methylation of the phenolic moieties is

Table 1. Optimization of the reaction conditions for C–N cross-coupling.[a]



entry	cavitand	aryl halide	yield[b] [%] a	yield[b] [%] b	yield[b] [%] c
1	**C_8PhenMe$_6$**	chlorobenzene	–	–	–
2	**C_8PhenMe$_6$**	bromobenzene	73	8	–
3	**C_8PhenMe$_6$**	iodobenzene	56	10	–
4[c]	**C_8Phen**	bromobenzene	50	12	7
5[d]	**Phen**	bromobenzene	20	10	–
6[d]	**Phen(OAr)$_2$**	bromobenzene	26	8	10
7[e]	none	bromobenzene	–	9	–
8[f]	**C_8PhenMe$_6$**	bromobenzene	68	6	–

[a] The evaluations were performed in a Schlenk flask (25 mL) with 2 mL of toluene, under dinitrogen atmosphere, temperature was raised to 110 °C in an oil bath for 36 h. Reaction conditions: molar ratio of 1:1 of aryl halide and aniline, with 2.5% mol of cavitand and 2.5% mol of CuI. [b] Isolated yields after column chromatography. [c] 8% of triphenylamine (see Figures S20–21) isolated as double coupling product. [d] 10% mol of ligand and 10% mol of CuI. [e] without cavitand and 10% mol of CuI. [f] Mercury drop test was performed.

key to achieve high yields of C–N coupling products, since **[Cu(C_8Phen)I]** afforded only 50% of product **a** (entry 4 in Table 1). This may be attributed to the potential interference of deprotonated ArOH groups of **C_8Phen** in the presence of the strong base KO*t*Bu. Although a reasonable amount of **a** was obtained (50%), collateral formation of homocoupling products of aryl halide (biphenyl, product **b**) and aniline (azobenzene, product **c**) were observed in low (7–12%) yields. The formation of **b** is a commonly observed secondary catalytic path in Cu(I)-based systems,[24] particularly since KO*t*Bu can also promote biaryl coupling reactions.[25] Aniline oxidation to generate **c** has been associated with Cu(I) complexes in the presence of an oxidant.[26] In this evaluation, the cavitand **C_8Phen** does not seem to protect the Cu(I) center towards oxidation to Cu(II) by adventitious oxygen as its methylated counterpart **C_8PhenMe$_6$**. Thus, aniline oxidation to generate **c** may occur at the Cu(II) center, regenerating Cu(I). Acid-base equilibria also affect the catalytic performance, since deprotonation of the hydroxy groups of **C_8Phen** is likely favored over that of aniline (even when coordinated to Cu(I)) necessary for C–N coupling. The presence of Cu(II) in the **C_8Phen** system was confirmed by EPR analysis as mentioned above, with approximately 20% of **[Cu(C_8Phe)I]** initially present oxidized to Cu(II) based on integration and comparison to the external standard Cu(OTf)$_2$ (Figure 3b). In contrast, **C_8PhenMe$_6$** is more selective for C–N coupling, since no trace of azobenzene (**c**) formation was observed, and this may be due to the prevalence of Cu(I) as confirmed by EPR spectroscopic analysis of [**Cu(C_8PhenMe$_6$)I**]. Therefore, **C_8PhenMe$_6$** stabilizes the reduced species under strongly basic conditions, even upon exposure to air for 30 days, with a maximum of 5% Cu(II) present based on integration of the incipient EPR signal. These observations confirm that the chemical environment provided by the **C_8PhenMe$_6$** cavity stabilizes Cu(I) and prevents alternative acid-base equilibria that may take place within the OH-containing **C_8Phen**.

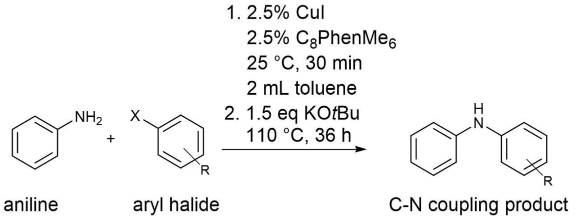
To evaluate the influence of the cavity on catalytic performance, we compared the **C_8PhenMe$_6$**/CuI system with the molecular analogues **Phen** and 2,9-dimethoxy(4-*tert*-butylphenyl)-1,10-phenanthroline (**Phen(OAr)$_2$**). Although substoichiometric quantities of ligand and CuI were employed (10% mol, entries 5 and 6, Table 1), no products were detected initially by thin layer chromatography (12-24 h of reaction time), and only low yields of **a** were consistently obtained (<26% at 36 h), further supporting the enhancement of C–N coupling activity due to macrocyclic encapsulation. The low conversion may be attributed to the known ligand redistribution tendency of **Phen** that results in complexes with 2:1 ligand/Cu(I) stoichiometry in solution.[22] Indeed [**Cu(Phen)$_2$**]$^+$ was detected by Direct Analysis in Real Time (DART) MS at $m/z=479$ (Figure S10). Halogen ligands tend to favor Cu–X–Cu (X = Cl, Br, I) bridge formation that limit the availability of the Cu(I) centers for catalytic reactions,[12] as evidenced by DOSY NMR spectroscopy of **Phen** in Figure S11 that affords a value of $r_H=2.38$ Å *vs* that of [**Cu(Phen)(μ-I)**]$_2$ in Figure S12 that is more than double at $r_H=7.25$ Å, consistent with a dimeric formulation. These experimental findings are in line with quantum chemical calculations, where dimerization yielding a Cu–X–Cu structural motif was energetically favored (see computational section of the results for more details).

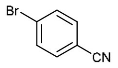
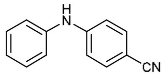
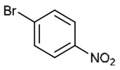
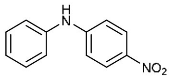
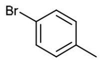
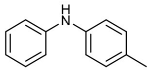
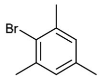
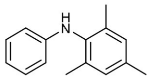
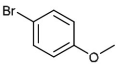
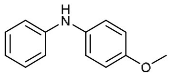
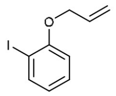
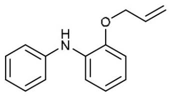
In the case of **Phen(OAr)$_2$** that also features the bidentate phenanthroline fragment, we hypothesized that the bulky substituents akin to those in **C_8PhenMe$_6$** may inhibit the dimerization observed with **Phen** (for DART MS of **Phen(OAr)$_2$** see Figure S13, for ^{1}H NMR characterization of **[Cu(Phen(OAr)$_2$)]** see Figure S14). Nevertheless, DOSY NMR spectroscopy (Figures S15–S16) shows that the calculated hydrodynamic radius of the complex formulated as **[Cu(Phen(OAr)$_2$)I]$_2$** ($r_H=10.56$ A) is about twice that of the free ligand **Phen(OAr)$_2$** ($r_H=5.65$ A). Furthermore FAB$^+$ MS attests to the presence of the redistribution product **[Cu(Phen(OAr)$_2$)$_2$]**$^+$ at $m/z=1071.5$ (Figure S17), along with monometallic **[Cu(Phen(OAr)$_2$)]**$^+$ at $m/z=567$. This implies that bulky substituents do not completely prevent ligand redistribution of the corresponding Cu(I) complex with **Phen(OAr)$_2$**. Thus, the limited catalytic performance of **[Cu(Phen(OAr)$_2$)I]** is likely due to equilibration with **[Cu(Phen(OAr)$_2$)$_2$]**$^+$, and the absence of the confined space provided by the cavity in **C_8PhenMe$_6$**. The combined observations indicate that such cavity is necessary to obtain high yields of the C–N coupling products, likely by increasing the local concentration of substrates during the catalytic process. This would allow more efficient interactions of the substrates with the Cu(I) center. We also tested C–N coupling with substoichiometric quantities of CuI (10% mol) in the absence of ligands (entry 7, Table 1), and no coupling product **a** was obtained. Finally, a mercury drop test did not alter the yield of **a** with **[Cu(C_8PhenMe$_6$)I]**;[27] although it does not provide definitive proof of a homogeneous catalytic process, we found no evidence of metallic copper.

Substrate Scope

The scope of C–N coupling with [**Cu(C$_8$PhenMe$_6$)I**] was explored by using different bromobenzene derivatives, as shown in Table 2. As usually observed in these types of coupling reactions, electron-deficient substrates result in higher yields of the desired products (entries 1–2 in Table 2). Moderate yields were obtained when electron-rich substrates were tested (entries 3–5), and also when *ortho* substituents relative to the halide are present (see also entry 6). The overall results evidence the good catalytic performance of [**Cu(C$_8$PhenMe$_6$)I**], particularly when the catalyst loading is compared to that of related molecular systems, which typically ranges from 5 to 10% mol.[14,24] The use of 1-allyloxy-2-iodobenzene as substrate afforded the expected coupling product, isolated in moderate yield (entry 6, 51%). Nonetheless, we did not find evidence of the formation of the cyclic product expected when a radical mechanism is involved in C–N bond formation.[24b] Thus, [**Cu(C$_8$PhenMe$_6$)I**] does not appear to promote the formation of aryl radical intermediates, see theoretical section for mechanistic details.

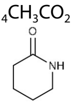
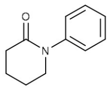
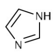
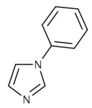
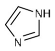
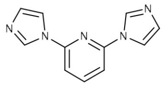
Table 2. Scope of the C–N cross-coupling catalyzed by in situ formed [Cu(C$_8$PhenMe$_6$)I].[a]



entry	aryl halide	C–N coupling product	yield[b] [%]
1			76
2			83
3			68
4			50
5			56
6			51

[a] Same conditions as in Table 1. [b] Isolated yields after column chromatography.

Table 3. C–N coupling with different nitrogen sources with arylbromide substrates catalyzed by in situ formed [Cu(C$_8$PhenMe$_6$)I].[a]

entry	*N*-source	C–N coupling product	yield[b] [%]
1	$NH_4CH_3CO_2$		–
2			76
3			82
4[c]			92

[a] Same conditions as in Table 1, using bromobenzene as substrate. [b] Isolated yields after column chromatography. [c] 2,6-dibromopyridine was used as substrate with imidazole in 1:2 ratio.

To gain further insight on the scope of the C–N reaction, different nitrogen-based substrates were tested under the standard catalytic conditions (Table 3). The results show that ammonium acetate is not a good substrate. Also, γ-valerolactam as a source of nitrogen reagent (entry 2, Table 3), offered no significant improvement in yield (76%). In contrast, when a more nucleophilic and less sterically hindered reagent like imidazole was employed (entry 3, Table 3), a higher yield was obtained (82%). Double substitution for the preparation of 2,6-disubstituted pyridine-bridged compound is also possible and resulted in an excellent yield (92%), avoiding high-boiling solvents or hazardous reaction conditions for this synthetic target.[28] These results indicate that smaller amines are preferred, supporting the notion that the use of [**Cu(C$_8$PhenMe$_6$)I**] may result in predominant steric, rather than electronic control. For characterization of coupling products, see Figures S18–S43 in the SI.

Computational Mechanistic Investigations

To substantiate the experimental findings and shed light on the mechanism underlying the C–N coupling, detailed theoretical studies were carried out. Due to the size and the inherent flexibility of [**Cu(C$_8$PhenMe$_6$)I**], initial mechanistic studies were carried out on the simpler [**Cu(Phen)I**] (Figure 4, for a molecular structure see Figure S45) as model system. Although experimental evidence shows that this species is a poor catalyst, its low performance may be attributed to deactivation upon dimerization to [**Cu(Phen)(μ-I)**]$_2$ (compare Figure S46). Indeed, dispersion-corrected density functional theory (DFT) investigations in implicit toluene (see Computational Methodology for details) predicted relative electronic and free energies to thermodynamically favor dimerization with $\Delta E = -71$ kJ mol^{-1} and $\Delta G^{298} = -39$ kJ mol^{-1} (Figure 4). Thus, dimerization appears to be one major factor for deactivation and investigation of [**Cu(Phen)I**] as model system is ideally suited to gain mechanistic insights prior to tackling the full-fledged catalyst [**Cu(C$_8$PhenMe$_6$)I**].

A schematic reaction mechanism for the C–N bond formation between aniline and bromobenzene mediated by [**Cu(Phen)I**] (denoted as **1**) is depicted in Figure 5. In the initial step, aniline coordinates to the neutral **1** to afford **2** in an exergonic reaction ($\Delta G^{298} = -28$ kJ mol^{-1}, $\Delta E = -51$ kJ mol^{-1}).

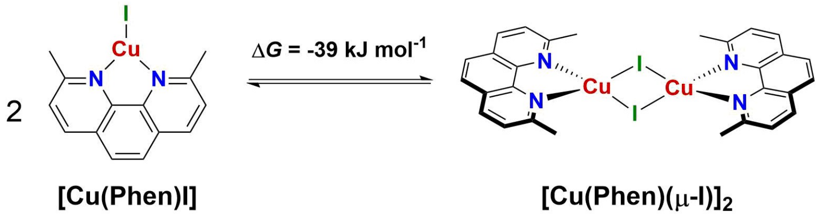


Figure 4. Dimerization of **[Cu(Phen)I]** to **[Cu(Phen)(μ-I)]$_2$**; reaction free energies were calculated with PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO (for details see Computational Methodology).

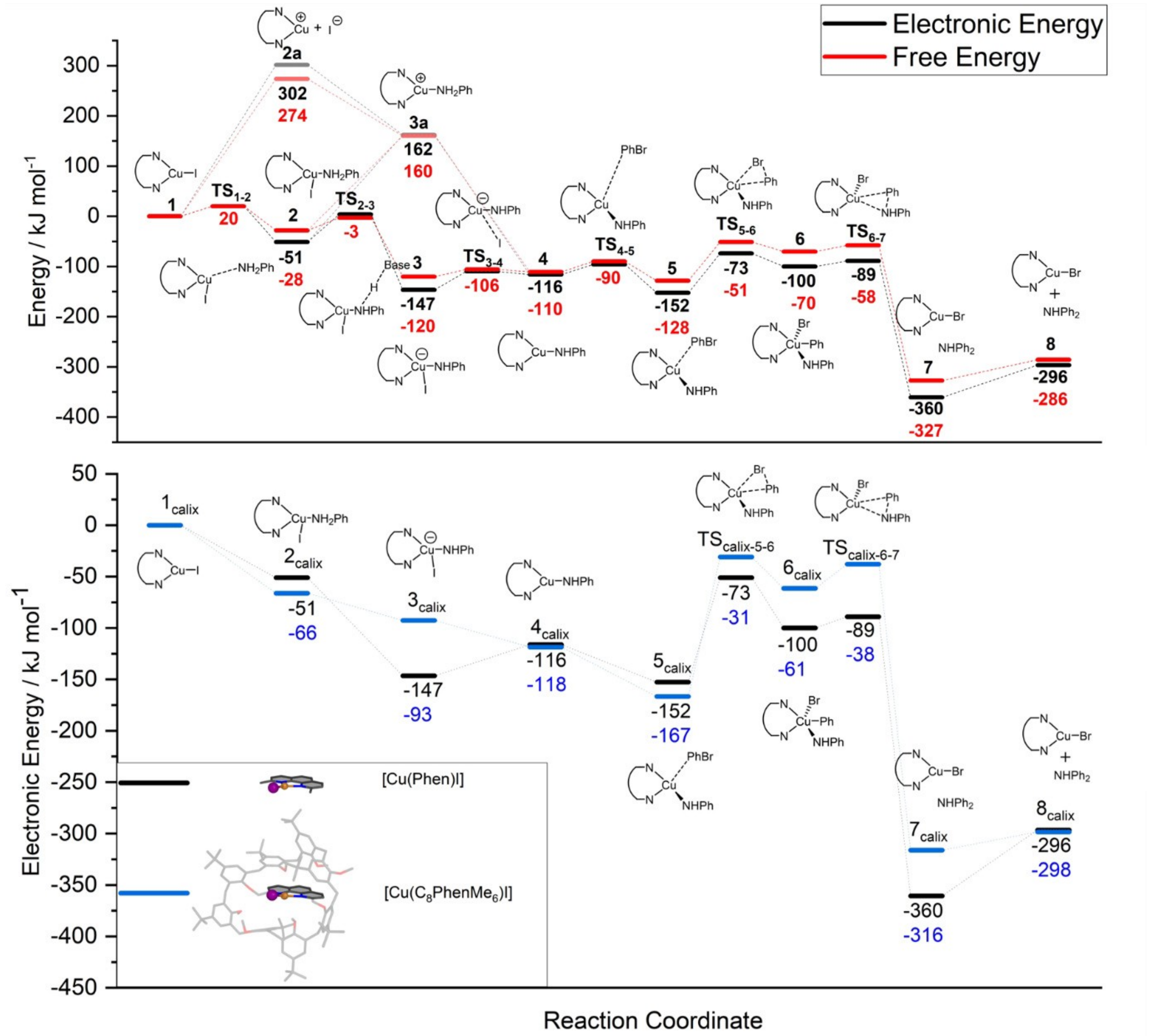


Figure 5. Top: Calculated energy diagram for the C–N coupling of aniline and PhBr with the model catalyst [**Cu(Phen)I**]. Relative electronic energies are given in black horizontal bars, whereas relative free energies are indicated by red bars. All energies are in kJ mol^{-1}, and were calculated with PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO. Bottom: Calculated energy diagram for the C–N coupling of aniline and PhBr with [**Cu(Phen)I**] (black bars) and [**Cu(C$_8$PhenMe$_6$)I**] (blue bars). All energies are relative electronic energies in kJ mol^{-1}, and were calculated with PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO; for details see Computational Methodology.

This adduct formation is monotonously downhill and hence diffusion controlled. For such processes, typical free energy barriers are about 20 kJ mol^{-1} – a value we assumed for **TS$_{1\text{-}2}$**.[29] The optimized structure of **2** shows π-π stacking of the phenyl ring of the aniline and the phenanthroyl moiety, at a distance of 3.5 Å (Figure S47). In the next step, deprotonation of coordinated aniline via **TS$_{2\text{-}3}$** (Figure S48) occurs, for which we estimated a free energy barrier of approximately 25 kJ mol^{-1} (see Computational Methodology for details). As a result, the negatively charged intermediate **3** is formed, with an additional stability of $\Delta G^{298} = -120$ kJ mol^{-1} and $\Delta E = -147$ kJ mol^{-1}. The aniline phenyl ring moves closer to the phenanthroyl fragment at 3.3 Å, while the rest of the structure remains largely unchanged (Figure S49). For the subsequent dissociation of iodide, we estimate a free energy barrier for **TS$_{3\text{-}4}$** of 14 kJ/mol (compare Figure S50; see Computational Methodology for details) to yield intermediate **4** with $\Delta G^{298} = -110$ kJ mol^{-1} and $\Delta E = -116$ kJ mol^{-1}. Here, Cu regains a trigonal planar geometry (Figure S51). Alternatively, **4** can be accessed by iodide dissociation from **2**, resulting in **2a** (Figure S52), which would be followed by coordination of aniline to give **3a** (Figure S53), and finally deprotonation. This alternative pathway can be ruled out based on the highly unfavorable energetic profile, which results in relative free energies of $\Delta G^{298} = 274$ and 160 kJ mol^{-1} for **2a** and **3a**, respectively. Approach of the phenyl bromide to **4** proceeds via **TS$_{4\text{-}5}$**. Once again, this step presents a monotonous decrease in electronic energy. We assumed diffusion control and – in the absence of the calixarene cavity – estimated a barrier of $\Delta G^{\neq} \approx 20$ kJ mol^{-1} (**TS$_{4\text{-}5}$**).[28] The formation is facilitated by π-π stacking interactions between PhBr and the phenanthroyl moiety, situated at 3.3 Å.

Structure **5** (Figure S54) represents the encounter complex from which the Br–C(Ph) oxidative addition step begins. **5** is stabilized by newly formed non-covalent interactions, with $\Delta G^{298} = -128$ kJ mol^{-1} and $\Delta E = -152$ kJ mol^{-1}. The oxidative addition to form **6** resulted in a free energy barrier for **TS$_{5\text{-}6}$** of $\Delta G^{\neq} =$ 77 kJ mol^{-1}. The structure of **TS$_{5\text{-}6}$** in Figure 6 (left) suggests concerted bond cleavage along with Cu–C(Ph) and Cu–Br bond formation. The Cu(III) intermediate **6** ($\Delta G^{298} = -70$ kJ mol^{-1} and $\Delta E = -100$ kJ mol^{-1}, Figure S55) then undergoes C–N coupling by reductive elimination to yield the observed product, associated with a low free energy barrier calculated at $\Delta G^{\neq} = 12$ kJ mol^{-1}. The transition state **TS$_{6\text{-}7}$** ($\Delta G^{298} = -58$ kJ mol^{-1} and $\Delta E = -89$ kJ mol^{-1}), depicted in Figure 6 (right), has the two aromatic rings of the products at a 70 °C ∠Cu–N angle as they approach, while increasing the distance to the metal center. Product formation is energetically favored, resulting in a reaction free energy of $\Delta G^{298} = -327$ kJ mol^{-1} ($\Delta E = -360$ kJ mol^{-1}) for the catalyst-product complex **7** (Figure S56). The final structure, **8**, represents the infinitely separated catalyst and product ($\Delta G^{298} = -286$ kJ mol^{-1} and $\Delta E = -296$ kJ mol^{-1}). Note that the extraordinary high stability of **7** compared to **8**, is a result of π-π stacking interactions of the phenanthroline and one phenyl ring of the aryl amine, as well as a stabilizing intermolecular Br–H interaction. For the subsequent second catalytic cycle, the initial structure is the bromo-equivalent of **1.** As no significant changes in the reaction profile were observed (see Figure S57), we refrain from further discussion here.

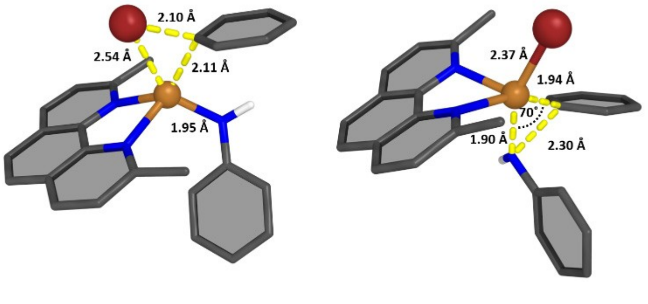


Figure 6. Left: Structure of the transition state for the oxidative addition **TS$_{5\text{-}6}$**, Right: Transition state for the reductive elimination **TS$_{6\text{-}7}$**. In both cases, important structural parameters are given. The two structures were optimized with PBE0/def2-SVP/D3 in implicit toluene.

The calculated rate determining step or state, according to Kozuch and Shaik's energy span model,[30] is the difference between the intermediate and the transition state that maximize the energy span. In our case, it is intermediate **5** and the transition state **TS$_{5\text{-}6}$**. For this oxidative addition step, the free energy barrier was calculated at $\Delta G^{\neq} = 77$ kJ mol^{-1}. This moderate barrier suggests that the reaction should readily take place at experimental reaction conditions starting with monomeric [**Cu(Phen)I**]. Apart from dimerization as a deactivation pathway, strong interactions between [**Cu(Phen)Br**] and the product may also result in loss of activity or complete deactivation of the catalyst.

Although other mechanisms, such as single electron transfer (SET) or atom transfer are discussed in the literature,[15] they all involve open-shell Cu(II) and radical species that can be ruled out here. No experimental evidence of such species was found. Hence, an oxidative addition/reductive elimination mechanism is the only viable reaction pathway that agrees with all experimental data. A recent computational study on a similar [**Cu(Phen)I**] catalyst predicted the reaction barriers of all alternative mechanisms to be higher.[16d]

Effect of Confinement

To estimate the effect of confinement on the reaction mechanism, we investigated the supramolecular catalysts and its intermediates by developing a multiscale approach workflow (Figure S58). Due to the flexibility of macrocyclic ligands, enhanced sampling methods were necessary to capture their conformational diversity. In addition, simulation in the full condensed phase, that is, in explicit solvation, was required to retain the cavity. Therefore, we performed accelerated Molecular Dynamics (MD) simulations of the ligand **C$_8$PhenMe$_6$** in chloroform to explore the potential energy surface. Although the experiments were conducted in toluene, deviations were found to be small (see Figure S59) and computational costs were significantly lower. To also investigate the corresponding Cu(I) complex and selected reaction intermediates as well as transition states, force field parameters had to be generated prior to simulation. After the simulation runs, obtained trajectories were clustered to yield a few representative and structurally diverse conformers. The initial structures thus generated were re-optimized with DFT (see Computational Methodology for details).

Overall, investigation of the supramolecular system as a whole revealed a similar energetic profile as [**Cu(Phen)I**], as shown in Figure 5 (bottom). For structures $\mathbf{2}_{\text{calix}}$, $\mathbf{4}_{\text{calix}}$, $\mathbf{5}_{\text{calix}}$ and $\mathbf{8}_{\text{calix}}$ the relative stabilities differ by less than 15 kJ mol^{-1}. Notable differences are found for $\mathbf{3}_{\text{calix}}$, $\mathbf{6}_{\text{calix}}$, and $\mathbf{7}_{\text{calix}}$. While the deviation of 54 kJ mol^{-1} between the small model and the supramolecular structure of $\mathbf{3}_{\text{calix}}$ arises potentially due to insufficient sampling, the higher relative energy of $\mathbf{6}_{\text{calix}}$ can be attributed to the very bulky nature of the intermediate, which occupies the calixarene cavity in its entirety, forcing the cage into unfavorable conformations. In addition to these intermediates, we also converged transition-states for the oxidative addition ($\mathbf{TS}_{\text{calix-5-6}}$) and reductive elimination ($\mathbf{TS}_{\text{calix-6-7}}$) steps, respectively. Although the predicted barriers are somewhat higher than in the molecular case, this should not be overemphasized because we did not run MD simulations on these structures. Consequently, we were unable to predict the lowest conformation properly and the reported numbers can be considered as upper bounds to the true TS barriers. Moving forward in the reaction, the release in energy from $\mathbf{6}_{\text{calix}}$ towards $\mathbf{7}_{\text{calix}}$, at −255 kJ mol^{-1} is similar to that of the transition from **6** to **7**, at −239 kJ mol^{-1}. A structural difference that can be remarked between **7** and $\mathbf{7}_{\text{calix}}$ is that the hydrogen bond between the amine and the bromine is no longer present. Instead, it has been replaced by a hydrogen bond to a calixarene oxygen atom, thus weakening the product-catalyst interaction. Consequently, the release of product from the calixarene cavity is facilitated, with $\mathbf{8}_{\text{calix}}$ being only 18 kJ mol^{-1} higher in energy than the previous intermediate, compared to 60 kJ mol^{-1} for **8**. As these are electronic energies only, thermal and entropic corrections will make this step exergonic.

Analysis of the dynamics of the full system revealed a response of the methoxyphenyl units surrounding the catalytic center upon C–N coupling, depicted in Figure 7. Close inspection of the structures of the optimized supramolecular variants, $\mathbf{1}_{\text{calix}}$ through $\mathbf{7}_{\text{calix}}$ (Figure S60) reveals that the methoxyphenyl units rotate to adapt the cavity around the reaction center. The rotation of the methoxyphenyl units makes room for the formation of the C–N coupling product, as depicted in Figure 7. To quantify differences in the calixarene flexibility, given by the ability of each methoxyphenyl moiety to rotate, we estimated the dihedral entropy of the species as the reaction progresses from $\mathbf{1}_{\text{calix}}$ to $\mathbf{6}_{\text{calix}}$, as shown in Figure S61. Clearly, a rigidification of the ring structure is observed along the reaction coordinate. The fivefold coordinated copper center in **6** and the bulky nature of the aniline and phenyl ligands are responsible for this effect. The large volume taken up by structures **5** and **6** also suggests that the binding position of the phenantroyl bridge is vital in achieving a working catalyst. In this context, the 1,5-substitution pattern is likely crucial for the C–N coupling steps, since this positioning allows for enough space for the bromoarene substrate to approach the Cu(I) center, while potentially rejecting larger substrates and thereby offering size selectivity. Thus, the cavity plays a critical role in limiting the size of the substrates approaching the reaction center. In addition, it prevents unwanted side reaction, such as dimerization, as found for the molecular analogue **1**. Furthermore, it potentially facilitates the reaction by activating intermediate $\mathbf{6}_{\text{calix}}$ and by preventing a strong catalyst-product interaction (see **7** and **8** vs. $\mathbf{7}_{\text{calix}}$ and $\mathbf{8}_{\text{calix}}$ in Figure 5).

Conclusions

In summary, we demonstrated how a macrocyclic encapsulation with a $\mathbf{C_8}$ based cavitand enhances C–N coupling activity of [$\mathbf{Cu(C_8PhenMe_6)I}$] when compared to its molecular analogue [**Cu(Phen)I**], and determined the origin of this increase in catalytic activity. The supramolecular catalyst achieves C–N coupling reactions in toluene with low loadings featuring Earth-abundant copper. The system offers a chemical environment that is inert to basic conditions, while also favoring interaction

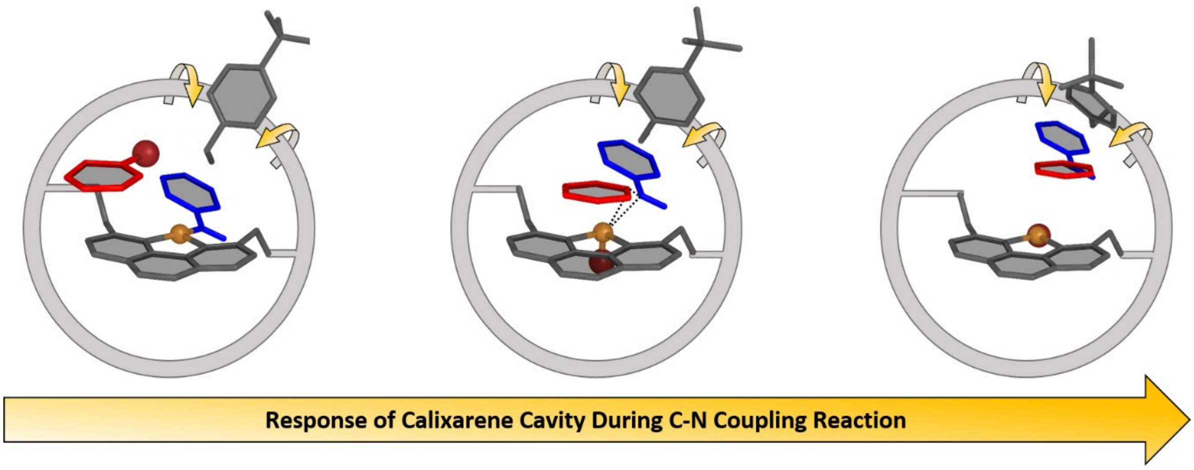


Figure 7. Schematic representation of the calixarene response during the C–N coupling reaction. The calixarene is depicted as a grey circle, surrounding the catalytic center, while one methoxyphenyl unit is shown in detail to highlight the rotation. As PhBr (red) enters the cavity (left), rotation of the methoxyphenyl unit begins, indicated by yellow arrows. C–N coupling starts taking place once PhBr approaches the reaction center (middle), along with continued rotation of the methoxyphenyl. Upon completion of the reaction (right), the product exits the cavity and the methoxyphenyl unit points in the opposite direction.

of the metal center with the substrates. Based on the experimental and theoretical results, the catalytic process is kinetically limited by the oxidative addition of the haloarene, and by the steric restrictions imposed by the macrocycle suggesting that a 1,5-phenanthroline bridge is vital for catalysis. Computational studies further revealed that prevention of catalyst aggregation, activation of intermediates as well as facilitation of product release fuels C–N coupling activity in comparison to the molecular analogue. **[Cu(C$_8$PhenMe$_6$)I]** exhibits better catalytic performance with small amines, in combination with bromoarene derivatives that fit within the cavity with more ease than their iodoarene counterparts. This supramolecular system evidences the advantages of the encapsulation of transition metals, such as higher activity and size selectivity. The detailed mechanistic insights should allow rational modulation of the reactivity by tailoring the properties of the calixarene, or other flexible macrocyclic scaffolds. The presented strategy may be extended to related homogeneous catalytic processes that are currently under investigation in our laboratory.

Experimental

Unless otherwise specified, the reagents were purchased from commercial suppliers and used without further treatment. Solvents were dried and degasified using standard techniques. Deuterated solvents were degassed prior to use. Catalytic evaluations were performed in oven dried Schlenk flasks under vacuum/N$_2$ (purity $>$99.998%, Praxair). Melting points were determined with an Electrothermal Mel-Temp apparatus and are uncorrected. ^{1}H and ^{13}C NMR spectra were recorded at 300 and 75 MHz with a JEOL Eclipse or Varian Inova 500 MHz spectrometer, the chemical shifts are reported relative to the residual solvent peaks. DOSY experiments were performed using Oneshot DOSY pulse sequence;[20] a duration of 1 ms diffusion-encoding gradient pulse were used in 0.1 s steps. 65,536 data points were acquired with 256 transients each. IR spectra were collected as KBr pellets on a Bruker Tensor 27 spectrometer from 4000 to 400 cm^{-1}. Elemental analyses were performed with a Vario-Micro V2.0.11 elemental analyzer. FAB$^+$ ion mass spectra were obtained on JEOL SX-102 A spectrometer operated at an accelerating voltage of 10 kV with *m*-nitrobenzyl alcohol as a solvent (matrix). DART MS were recorded on a JEOL JMS–T100LC spectrometer.

Synthesis of C$_8$PhenMe$_6$. Precursor **C$_8$Phen** (500 mg, 0.33 mol) and excess NaH (52 mg, 2.17 mmol) were suspended in anhydrous THF (10 mL) under dinitrogen, followed by addition of MeI (312 mg, 2.17 mmol), the mixture was held at 25 °C for 48 h to achieve complete methylation of the phenolic positions; the formation of a white suspension was noticed. Volatiles were evaporated under reduced pressure, and the crude solid was extracted with 50 mL mixture of dichloromethane/brine (1:1). The organic phase was dried over NaSO$_4$, and after slow solvent evaporation a colorless solid was obtained. The product was purified by column chromatography on silica gel with dichloromethane as eluant, and washed with hexanes, resulting in a colorless solid in 80% yield (418 mg, 0.26 mmol). M.p. $>$210 °C. δ_H (300 MHz, C$_2$D$_2$Cl$_4$, 293.15 K) 8.06 (m, 4H, Ar$_{Phen(ac)}$), 7.55 (s, 2H, Ar$_{Phen(b)}$), 7.07 (m, 4H, Ar$_{C8(c)}$), 6.94 (m, 4H, Ar$_{C8(a)}$), 6.75 (m, 4H, Ar$_{C8(d)}$), 6.65 (m, 4H, Ar$_{C8(d)}$), 4.34 (d, J=16.5 Hz, 4H, CH$_{2C8(c)}$), 4.27 (s, 4H, CH$_{2Phen}$), 3.98 (d, J=15.9 Hz, 4H, CH$_{2C8(a)}$), 3.59 (d, J=16.1 Hz, 4H, CH$_{2C8(b)}$), 3.49 (d, J=15.7 Hz, 4H, CH$_{2C8(d)}$), 3.30 (s, 12 H, MeO$_{ext}$), 2.47 (s, 6H, MeO$_{ext(a)}$), 2.25 (s, 6H, MeO$_{ext(b)}$), 1.23 (s, 18H, *t*-Bu$_{ext(b)}$), 1.14 (s, 18H, *t*-Bu$_{ext(a)}$), 0.90 (s, 36H, *t*-Bu$_{int}$). ν_{max} (KBr) 2955 (CH$_3$ and CH$_2$), 2866 (CH$_3$ and CH$_2$) and 1479 (Ar). *m/z* (FAB$^+$) 1587 **[C$_8$PhenMe$_6$H]**$^+$. Elemental analysis for C$_{108}$H$_{134}$N$_2$O$_8$·H$_2$O·CHCl$_3$, found: C, 75.87; H, 8.00; N, 1.62; requires: C, 75.88, H, 7.86, N, 1.43.

Synthesis of [Cu(C$_8$PhenMe$_6$)I]. The cavitand **C$_8$PhenMe$_6$** (30 mg, 0.018 mmol) and CuI (3.5 mg, 0.018 mmol) were placed in a 50 mL Schlenk tube with 2 mL of toluene under dinitrogen with constant stirring. Volatiles were evaporated after 2 h of stirring under reduced pressure, resulting in an orange crystalline solid in 92% yield (30 mg, 0.016 mmol), m.p. $>$230 °C. δ_H (300 MHz, toluene-d$_8$, 393.15 K) 8.34 (d, J=8.5 Hz, 1 H, Ar$_{Phen}$), 8.18 (d, 1 H, Ar$_{Phen}$), 7.68 (m, 1 H, Ar$_{Phen}$), 7.62 (m, 1 H, Ar$_{Phen}$), 7.16 (m, 8 H, Ar$_{C8}$), 7.09 (m, 3 H, Ar$_{toluene}$), 7.00 (m, 8 H, Ar$_{toluene}$), 6.98 (m, 2 H, Ar$_{toluene}$), 5.25 (m, 2 H, CH$_{2Phen}$), 4.87 (d, J=16 Hz, 2 H, CH$_{2exo}$), 4.73 (m, 2 H, CH$_{2exo}$), 4.52 (m, 5 H, CH$_{2exo}$ y CH$_{2Phen}$), 4.21 (d, J=16 Hz, 4 H, CH$_{2exo}$), 3.75 (m, 4H, CH$_{2exo}$), 4.18 (m, 4 H, CH$_{2endo}$), 3.81 (m, 4 H, CH$_{2endo}$), 3.59 (s, 9 H, CH$_{3int}$), 3.13 (s, 9 H, CH$_{3ext}$), 2.09 (s, 3 H, CH$_{3toluene}$), 1.30 (s, 36 H, *t*-Bu$_{ext}$), 1.18 (m, 36 H, *t*-Bu$_{int}$). *m/z* (FAB$^+$) 1649 **[Cu(C$_8$PhenMe$_6$)]**$^+$.

General procedure for coupling reactions. Typically, 4 mg (0.02 mmol, 2.5% mol) of CuI and 30 mg (0.02 mmol, 2.5% mol) of **C$_8$PhenMe$_6$** in 2.0 mL of toluene were stirred in a 25 mL Schlenk flask for 30 min under dinitrogen. Then, 73 μL of aniline (74 mg, 0.80 mmol), an equimolar amount of the aryl bromide, and KO*t*Bu (150 mg, 1.2 mmol) were added, and the temperature was raised to 110 °C for 36 h. After this time, the products were isolated by flash column chromatography on basic alumina as stationary phase; the mobile phase employed consisted of a hexanes/dichloromethane gradient, starting with 0 and up to 20–30% dichloromethane, depending on the polarity of the product obtained. Their identity was confirmed by comparison with literature spectroscopic data (see Figures S20-S45 in the SI).[24b,56] The same reaction conditions were employed for comparative studies with **C$_8$Phen**, and the molecular analogues **[Cu(Phen(OAr)$_2$)I]** and **[Cu(Phen)I]$_2$**, except that the catalyst loadings corresponded to 10% mol for the latter two complexes.

Computational Methodology

A multiscale modelling protocol was designed to describe the supramolecular catalyst **[Cu(C$_8$PhenMe$_6$)I]** (see Figure S58 for a schematic overview). We first studied the C–N coupling mechanism with monomeric **[Cu(Phen)I]** (**1**) as the catalyst, before assessing the flexibility of the macrocyclic ligand and studying the supramolecular system.

Quantum chemical studies of [Cu(Phen)I]. The initial geometry for **1** was constructed using Gaussview 6.0.16.[31a] Unless otherwise noted, quantum chemical calculations were performed with Turbomole 7.3.[31b,c] All structures were roughly preoptimized in *C*1 symmetry using the BP86[32] functional and def2-SVP[33] basis set with empirical dispersion corrections of the D3 type,[34] as well as implicit solvent corrections using the Conductor-Like Screening Model (COSMO)[35] as implemented in Turbomole. A dielectric constant of ε=2.4 was chosen to model toluene. The structures were further optimized with the PBE0 functional[36] which is known to provide good results when investigating structures involving transition metals[37,38] and the same basis set, solvation model and dispersion corrections as described above. Single point energy calculations were performed on the PBE0 optimized structures with the def2-TZVP basis set[33] using the resolution-of-identity technique for speed-up[39] and D3 corrections. Again, calculations were performed in toluene modelled as implicit solvent, to obtain more accurate electronic energy values. Harmonic frequency calculations were performed on the optimized structures to verify that they are

Chemistry Europe
European Chemical Societies Publishing

indeed energy minima. Reported energies are either relative electronic energies ΔE (PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/-COSMO) or relative free energies ΔG[40,41] (see SI for details on free energy calculations) at 298.15 K, all given in kJ mol^{-1}.

The transition states (TS) were obtained by using the eigenvector following method implemented in Turbomole.[31b,c] A successful transition state was confirmed by the presence of exactly one imaginary frequency matching the reaction coordinate.

In some cases, transition states could not be localized by the eigenvector following method. In such cases, tailored strategies were used to approximate the TSs. For the association reactions *via* **TS$_{1\text{-}2}$** and **TS$_{4\text{-}5}$**, the energy was found to monotonously decrease when the ligand approached the catalysts, which indicates a diffusion-controlled transition state. To verify this, the structures of the so obtained intermediates were modified in such a way that the two molecules are 6 Å apart and then left to freely optimize. The optimizations completed reaching the structure of the intermediate, providing further proof of a diffusion-controlled transition state. For these cases, as a suitable estimate for the transition state free energy, 20 kJ mol^{-1} was taken as suggested in literature.[29] Transition states for dissociation reactions could not be localized with eigenvector following methods either, because dissociations were found to be energetically uphill in the electronic energy. However, as the reactants separate, they obtain additional entropy. To approximate the transition state energies, a method put forward by Baik was applied[42] accounting for the onset of entropy upon dissociation, which has been successful for Au catalysts in the past[43] and proved useful in estimating **TS$_{4\text{-}5}$** (Figure S44), see SI for more details. Particularly challenging was the hydrogen abstraction (**TS$_{2\text{-}3}$**). As the deprotonation in experiment is facilitated by a base, we approximated the barrier for hydrogen abstraction by modelling the reaction path of hydrogen transfer from **2** to the base (see Figure S48 for structure). It is noteworthy here that all approximated reaction barriers are significantly lower than the rate determining state **TS$_{5\text{-}6}$**, for which a transition state was localized with eigenvector following.

Molecular Dynamics simulations to assess conformational flexibility. To obtain a representative structural ensemble of the macrocyclic ligand **C$_8$PhenMe$_6$** for subsequent DFT studies, the conformational spaces was explored in the absence and in the presence of Cu(I), using classical accelerated Molecular Dynamics (aMD) simulation – one approach shown to yield reliable conformational ensembles for macrocycles.[17] A sampling in explicit solvent proved to be essential to retain intact cavities, as initial sampling attempts in implicit solvent resulted in their collapse. In the absence of Cu(I), the initial structure of **C$_8$PhenMe$_6$** was adapted from the crystal structure of a calix[8]arene/Cs[14] and appropriate structural modification. Force field parameters for **C$_8$PhenMe$_6$** were generated using Amber's Generalized Force Field (GAFF),[44] partial charges were obtained with RESP fitting. MD simulations were carried in chloroform as well as benzene (as a model for toluene used in experiment). As results were largely similar (see Figure S59), we restricted ourselves to chloroform only to reduce computational costs.

To obtain adequate sampling of the catalyst, we performed twelve 1 μs aMD simulations, followed by a hierarchical clustering to obtain a structurally diverse ensemble (see details SI for more details on these calculations performed with Amber).[45–51]

In addition, we simulated the assembled supramolecular structures **1**$_{\text{calix}}$ and **6** $_{\text{calix}}$, to assess how the flexibility of the cavity changes upon Cu coordination. Therefore, a parametrization of the metal ion at the center of the molecule was required. MCPB.py[52] was selected as the tool of choice for the parametrization step. The initial structure was created in Gaussview by taking a cluster representative from the simulation without metal, after which the respective QM optimized structure was fitted in (see Assembly of supramolecular catalyst for DFT investigations). The resulting structures were then optimized at the PBE0/def2-SVP/D3/COSMO level. The next step involved separating the metal center and ligands into individual substructures and following the MCPB.py procedure. The structure optimizations and frequency calculations were performed at the PBE0/def2-SVP/D3 level. Using the resulting parameters, equilibration, generation of boosting parameters and production runs in explicit chloroform were performed according to the protocol outlined above.

To investigate the flexibility of the calixarene units, a dihedral analysis was used, to characterize the rotations between each calixarene unit. Dihedral angles were assigned between the planes of each of the six-membered rings of the calixarene units, as changes in these angles accurately describe the movements within the calixarene ring. Using the X-entropy script[53] to perform an entropic analysis on the angle distributions, the relative entropy values were obtained, thus, leading to an accurate estimation of the flexibility.

Assembly of supramolecular catalyst for DFT investigations. Once the [**Cu(Phen)I**] (**1**) structure pathway investigation was completed, the next step was to study the effects of the confinement by the calix[8]arene on the energetic landscape. To this end, a fitting of the reaction intermediates back into the **C$_8$** ring was necessary. Cluster representatives of the molecular dynamics simulations were selected using a hierarchical clustering algorithm. The cluster representative with the largest population of the aMD simulation of structure **1**$_{\text{calix}}$ was chosen as the starting point for a fitting of all QM optimized structures. The fitting process was automated using a python script via PyMOL,[54] which aligned the phenanthroyl backbones of the large and small models. The calixarene ring, along with the backbone from the large model were kept and combined with the remaining atoms of the small model. The results of the fitting were then pre-optimized with GFN2-xTB,[55] followed by an optimization with PBE0/def2-SV(P)/D3/COSMO. Final single point energies were again calculated with PBE0/def2-TZVP/D3/COSMO.

Structural parameter measurements and visualizations were done with PyMOL.[54]

Supporting Information

Detailed characterization of cavitand **C$_8$PhenMe$_6$** and the supramolecular system [**Cu(C$_8$PhenMe$_6$)I**], characterization of coupling products, details concerning the theoretical investigations.

Acknowledgements

We thank M. Orta for combustion analysis, V. Gómez-Vidales for EPR, M. A. Peña-González for NMR assistance, and A. S. Kamenik for help with the accelerated MD simulations. Computational results have been achieved using the HPC infrastructure of the University of Innsbruck (leo3e) and the Vienna Scientific Cluster VSC3. CONACYT (Beca 604082, Proyecto A1-S-8682), Instituto de Química-UNAM, and the Austrian Science Fund (FWF, M-2005 and P33528) are acknowledged for financial support.

Conflict of Interest

The authors declare no competing financial interest.

Data Availability Statement

Experimental characterization of the **C_8PhenMe$_6$** and the supramolecular system **[Cu(C_8PhenMe$_6$)I]** as well as all coupling products can be found in the Supporting Information. Coordinates of all investigated structures are supplied in an xyz file with the DOI: 10.5281/zenodo.7008048 that can be downloaded at https://zenodo.org/.

Keywords: C–N Coupling · Copper · Multiscale Modelling Supramolecular Catalysis

[1] a) M. J. Buskes, M. J. Blanco, *Molecules* **2020**, *25*, 3493–3515; b) M. Gay, P. Carato, M. Coevoet, N. Renault, P. Larchanché, A. Barczyk, S. Yous, L. Buée, N. Sergeant, P. Melnyk, *Bioorg. Med. Chem.* **2018**, *26*, 2151–2164; c) M. Fitzner, G. Wuitschik, R. Koller, J. M. Adam, T. Schindler, J. L. Reymond, *Chem. Sci.* **2020**, *11*, 13085–13093.

[2] a) R. Jiang, L. Li, T. Sheng, G. Hu, Y. Chen, L. Wang, *J. Am. Chem. Soc.* **2018**, *140*, 11594–11598; b) S. Hayakawa, A. Kawasaki, Y. Hong, D. Uraguchi, T. Ooi, D. Kim, T. Akutagawa, N. Fukui, H. Shinokubo, *J. Am. Chem. Soc.* **2019**, *141*, 19807–19816; c) S. Izumi, H. F. Higginbotham, A. Nyga, P. Stachelek, N. Tohnai, P. de Silva, P. Data, Y. Takeda, S. Minakata, *J. Am. Chem. Soc.* **2020**, *142*, 1482–1491; d) F. Picini, S. Schneider, O. Gavat, A. Vargas-Jentzsch, J. Tan, M. Maaloum, J. M. Strub, S. Tokunaga, J. M. Lehn, E. Moulin, N. Giuseppone, *J. Am. Chem. Soc.* **2021**, *143*, 6498–6504.

[3] a) G. O. Jones, P. Liu, K. N. Houk, S. L. Buchwald, *J. Am. Chem. Soc.* **2010**, *132*, 6205–6213; b) K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, E. A. Pidko, *Chem. Rev.* **2019**, *119*, 2453–252; c) S. Ahn, M. Hong, M. Sundararajan, D. H. Ess, M. H. Baik, *Chem. Rev.* **2019**, *119*, 6509–6560; d) J. N. Harvey, F. Himo, F. Maseras, L. Perrin, *ACS Catal.* **2019**, *9*, 6803–6813; e) T. K. Piskorz, V. Martí-Centelles, T. A. Young, P. J. Lusby, F. Duarte, *ACS Catal.* **2022**, *12*, 5806–5826.

[4] a) M. D. Wodrich, B. Sawatlon, E. Solel, S. Kozuch, C. Corminboeuf, *ACS Catal.* **2019**, *9*, 5716–5725; b) D. J. Durand, N. Fey, *Chem. Rev.* **2019**, *119*, 6561–6594; c) T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, G. Doyle, *Science* **2018**, *360*, 186–190.

[5] a) P. Ruiz-Castillo, S. L. Buchwald, *Chem. Rev.* **2016**, *116*, 12564–12649; b) J. M. Dennis, N. A. White, R. Y. Liu, S. L. Buchwald, *J. Am. Chem. Soc.* **2018**, *140*, 4721–4725; c) R. Dorel, C. P. Grugel, A. M. Haydl, *Angew. Chem. Int. Ed.* **2019**, *131*, 17276–17287.

[6] a) X. Guo, H. Rao, H. Fu, Y. Jiang, Y. Zhao, *Adv. Synth. Catal.* **2006**, *348*, 2197–2202; b) I. Güell, X. Ribas, *Eur. J. Org. Chem.* **2014**, *15*, 3188–3195; c) E. B. Corcoran, M. T. Pirnot, S. Lin, S. D. Dreher, D. A. Dirocco, I. W. Davies, S. L. Buchwald, D. W. Macmillan, *Science* **2016**, *353*, 279–283; d) A. Bartoszewicz, C. D. Matier, G. C. Fu, *J. Am. Chem. Soc.* **2019**, *141*, 14864–14869.

[7] a) Z. J. Wang, K. N. Clary, R. G. Bergman, K. N. Raymond, F. D. Toste, *Nat. Chem.* **2013**, *5*, 100–103; b) D. M. Kaphan, M. D. Levin, R. G. Bergman, K. N. Raymond, F. D. Toste, *Science* **2015**, *350*, 1235–1238; c) B. K. Yan, M. Fujita, *Science* **2015**, *350*,1165–1166.

[8] a) M. Raynal, P. Ballester, A. Vidal-Ferran, P. van Leeuwen, *Chem. Soc. Rev.* **2014**, *43*, 1660–1733; b) L. J. Jongkind, X. Caumes, A. P. T. Hartendorp, J. N. H. Reek, *Acc. Chem. Res.* **2018**, *51*, 2115–2128; c) V. Mouarrawis, R. Plessius, J. I. van der Vlugt, J. N. H. Reek, *Front. Chem.* **2018**, *6*, 1–20.

[9] a) V. Marcos, A. J. Stephens, J. Jaramillo-Garcia, A. L. Nussbaumer, S. L. Woltering, A. Valero, J. Lemonnier, I. J. Vitorica-Yrezabal, D. A. Leigh, *Science.* **2016**, *352*, 1555–1559; b) J. Okuda, *Nat. Catal.* **2018**, *1*, 639–640.

[10] a) P. Nandi, A. Solovyov, A. Okrut, A. Katz, *ACS Catal.* **2014**, *4*, 2492–2495; b) A. Gorbunov, D. Cheshkov, V. Kovalev, I. Vatsouro, *Chem. Eur. J.* **2015**, *21*, 9528–9534; c) L. Wang, X. Gao, X. Hang, X. Zhu, H. Han, W. Liao, W. Chen, *J. Am. Chem. Soc.* **2016**, *138*, 16236–16239; d) X. X. Yang, W. D. Yu, X. Y. Yi, L. J. Li, C. Li, *Chem. Commun.* **2020**, *56*, 14035–14038; e) O. Santoro, C. Redshaw, *Coord. Chem. Rev.* **2021**, *448*, 214173.

[11] a) I. Abdellah, P. Kasongo, A. Labattut, R. Guillot, E. Schulz, C. Martini, V. Huc, *Dalton Trans.* **2018**, *47*, 13843–13848; b) A. Labattut, A. Abdellah, J. Buendia, S. A. Fayssal, E. Adhel, D. Dragoe, C. Martini, E. Schulz, V. Huc, *Catalysts* **2020**, *10*, 1081–1094.

[12] E. Guzmán-Percástegui, D. J. Hernández, I. Castillo, *Chem. Commun.* **2016**, *52*, 3111–3114.

[13] R. A. Altman, S. L. Buchwald, *Org. Lett.* **2006**, *8*, 2779–2782.

[14] D. J. Hernández, H. Vázquez-Lima, P. Guadarrama, D. Martínez-Otero, I. Castillo, *Tetrahedron Lett.* **2013**, *54*, 4930–4933.

[15] a) S. Li, Y. Lan, *Chem. Commun.* **2020**, *56*, 6609–6619; b) M. Rovira, L. Jašíková, E. Andris, F. Acuña-Parés, M. Soler, I. Güell, M.-Z. Wang, L. Gómez, J. M. Luis, J. Roithová, X. Ribas, *Chem. Commun.* **2017**, *53*, 8786–8789.

[16] a) P. Larsson, C. Wallentin, P.-O. Norrby, *ChemCatChem* **2014**, *6*, 1277–1282; b) K. Gurjar, R. Sharma, *ChemCatChem* **2017**, *9*, 862–869; c) K. Gurjar, D. V. Morarji, *Organometallics* **2019**, *38*, 2502–2511; d) D. M. Andrada, S. M. Soria-Castro, D. A. Caminos, J. E. Argüello, A. B. Peñéñory, *Catalysts* **2017**, *7*, 1–17; e) G. Lefèvre, G. Franc, A. Tlili, C. Adamo, M. Taillefer, I. Ciofini, A. Jutand, *Organometallics* **2012**, *31*, 7694–7707.

[17] A. S. Kamenik, U. Lessel, J. E. Fuchs, T. Fox, K. R. Liedl, *J. Chem. Inf. Model.* **2018**, *58*, 982–992.

[18] a) F. Cunsolo, G. M. L. Consoli, M. Piattelli, P. Neri, *J. Org. Chem.* **1998**, *63*, 6852–6858; b) X. Zeng, A. S. Batsanov, M. R. Bryce, *J. Org. Chem.* **2006**, *71*, 9589–9594.

[19] a) M. Oguz, A. Gul, S. Karakurt, M. Yilmaz, *ChemistrySelect* **2020**, *5*, 12250–12254; b) A. N. Blokhin, A. B. Razina, A. E. Bursian, A. V. Ten'kovtsev, *Polym. Sci. Ser. B.* **2021**, *63*, 52–62.

[20] R. W. Adams, J. A. Aguilar, G. A. Morris, M. Nilsson, *Org. Biomol. Chem.* **2011**, *9*, 7062–7064.

[21] a) G. Evano, N. Blanchard, M. Toumi, *Chem. Rev.* **2008**, *108*, 3054–3131; b) D. S. Surry, S. L. Buchwald, *Chem. Sci.* **2010**, *1*, 13–31.

[22] R. K. Gujadhur, C. G. Bates, D. Venkataraman, *Org. Lett.* **2001**, *3*, 4315–4317.

[23] a) H. Xu, C. Wolf, *Chem. Commun.* **2009**, 1715–1717; b) W. Zhou, M. Fan, J. Yin, Y. Jiang, D. Ma, *J. Am. Chem. Soc.* **2015**, *137*, 11942–11945.

[24] a) S. Saito, *J. Inclusion Phenom. Macrocyclic Chem.* **2015**, *82*, 437–451; b) X. Ding, M. Huang, Z. Yi, D. Du, X. Zhu, Y. Wan, *J. Org. Chem.* **2017**, *82*, 5416–5423; c) B. Górski, A. L. Barthelemy, J. J. Douglas, F. Juliá, D. Leonori, *Nat. Catal.* **2021**, *4*, 623–630.

[25] S. Yanagisawa, K. Ueda, T. Taniguchi, K. Itami, *Org. Lett.* **2008**, *10*, 4673–4676.

[26] a) C. B. R. Reddy, S. R. Reddy, S. Naidu, *Catal. Commun.* **2014**, *56*, 50–54; b) S. Singh, P. Chauhan, M. Ravi, I. Taneja, P. Yadav, *RSC Adv.* **2015**, *5*, 61876–61880.

[27] V. M. Chernyshev, A. V. Astakhov, I. E. Chikunov, R. V. Tyurin, D. B. Eremin, G. S. Ranny, V. N. Khrustalev, V. P. Ananikov, *ACS Catal.* **2019**, *9*, 2984–2995.

[28] a) K. Inamoto, J. Kuroda, K. Hiroya, Y. Noda, M. Watanabe, T. Sakamoto, *Organometallics* **2006**, *25*, 3095–3098; b) T. Tu, X. Feng, Z. Wang, X. Liu, *Dalton Trans.* **2010**, *39*, 10598–10600; c) L. Wang, N. Liu, B. Dai, H. Hu, *Eur. J. Org. Chem.* **2014**, *4*, 6493–6500.

[29] C. McMullin, J. Jover, J. N. Harvery, N. Fey, *Dalton Trans.* **2010**, *39*, 10833–10836.

[30] S. Kozuch, S. Shaik, *Acc. Chem. Res.* **2011**, *44*, 101–110.

[31] a) *GaussView*, Version 6, R. Dennington, T. A. Keith, J. M. Millam. Semichem Inc., Shawnee Mission, KS, **2016**; b) F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100; c) TURBOMOLE V 7.3 2018, a development of the University of Karlsruhe and Forschungszentrum Karlsruhe GmnH., *TURBOMOLE.* Since 2007: TURBOMOLE GmbH, available from http://www.turbomole.com/.

[32] a) J. Perdew, *Phys. Rev. B.* **1986**, *33*, 8822–8824; b) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.

[33] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[34] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[35] a) A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin.* **1993**, *2*, 799–805; b) A. Schäfer, A. Klamt, D. Sattel, J. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.

[36] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[37] T. Weymuth, E. Couzijn, P. Chen, M. Reiher, *J. Chem. Theory Comput.* **2014**, *10*, 3092–3103.

[38] Y. Minenkov, E. Chermak, L. Cavallo, *J. Chem. Theory Comput.* **2016**, *12*, 1542–1560.

[39] a) F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143–152; b) K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283–289; c) K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* **1997**, *97*, 119–124; d) R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2004**, *6*, 5119–5121.

[40] G. Luchini, J. V. Alegre-Requena, I. Funes-Ardoiz, R. S. Paton, *F1000Research* **2020**, *291*.

[41] M. K. Kesharwani, B. Brauer, J. M. L. Martin, *J. Phys. Chem. A* **2015**, *119*, 1701–1714.

[42] H. Ryu, H. Kim, J. Park, S. Kim, M. H. Baik, *Organometallics* **2018**, *37*, 3228–3239.

[43] S. K. Goetzfried, C. M. Gallati, M. Cziferszky, R. A. Talmazan, K. Wurst, K. R. Liedl, M. Podewitz, R. Gust, *Inorg. Chem.* **2020**, *59*, 15312–15323.

[44] a) J. Wang, R. Wolf, J. Caldwell, P. Kollman, D. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174; b) J. Wang, W. Wang, P. Kollman, D. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

[45] H. Wallnoefer, S. Handschuh, K. R. Liedl, T. Fox, *J. Phys. Chem. B.* **2010**, *114*, 7405–7412.

[46] a) R. Salomon-Ferrer, D. Case, R. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, 3, 198–210; b) D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R .E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman. AMBER 2018, Reference Manual. University of California, San Francisco. **2018**.

[47] S. Adelman, J. Doll, *J. Chem. Phys.* **1976**, *64*, 2375–2388.

[48] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, J. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[49] G. Ciccotti, J. Ryckaert, *Comput. Phys. Commun.* **1986**, *4*, 345–392.

[50] a) D. Hamelberg, C. A. F. de Oliveira, J. A. McCammon, *J. Chem. Phys.* **2007**, *127*, 1–9; b) D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, *120*, 11919–11929.

[51] L. Pierce, R. Salomon-Ferrer, C. A. F. de Oliveira, J. A. McCammon, R. Walker, *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.

[52] P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599–604.

[53] J. Kraml, P. K. Quoika, A. S. Kamenik, K. R. Liedl, *J. Chem. Inf. Model.* **2021**, *61*, 1533–1538.

[54] *The PyMOL Molecular Graphics System*, Version 2.0, Schrödinger, LLC. **2017**.

[55] a) C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *11*, 1–49; b) C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

[56] a) F. Du, Q. Zhou, Y. Fu, Y. Chen, Y. Wu, G. Chen, *Synlett.* **2019**, *30*, 2161–2168; b) S. De, J. Yin, D. Ma, *Org. Lett.* **2017**, *19*, 4864–4867.

Manuscript received: May 23, 2022
Revised manuscript received: July 11, 2022
Accepted manuscript online: July 23, 2022
Version of record online: September 1, 2022

ChemCatChem

Supporting Information

Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]arene Supramolecular Catalyst

Radu A. Talmazan^{+}, J. Refugio Monroy^{+}, Federico del Río-Portilla, Ivan Castillo,* and Maren Podewitz*

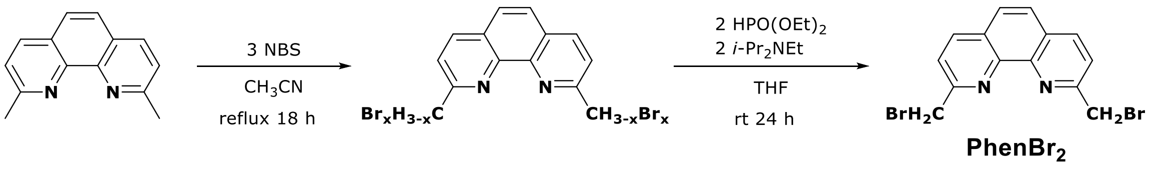
Table of Contents

S.1. General experimental procedures

Unless otherwise specified, the reagents were purchased from commercial suppliers and used without further treatment. Solvents were dried and degassed using standard techniques.[1] Deuterated solvents were degassed prior to use by 3 freeze-pump-thaw cycles. Catalysis evaluations were performed in oven-dried Schlenk flasks under vacuum/N_2 (purity > 99.998%, Praxair). ^{1}H and ^{13}C NMR spectra were recorded at RT with a 300 MHz JEOL Eclipse spectrometer, chemical shifts are reported according to the residual solvent peaks. IR spectra were recorded as KBr pellets on a Bruker Tensor 27 spectrometer from 4000 to 400 cm^{-1}. Elemental analyses were performed with a Vario-Micro V2.0.11 elemental analyser. Positive ion FAB$^+$ mass spectra were obtained on a JEOL SX-102A spectrometer operated at an accelerating voltage of 10 kV in a *m*-nitrobenzyl alcohol matrix. DART MS were recorded on a JEOL JMS-T100LC spectrometer. Melting points were determined with an Electrothermal Mel-Temp apparatus and are uncorrected.

S.2. Synthetic procedures

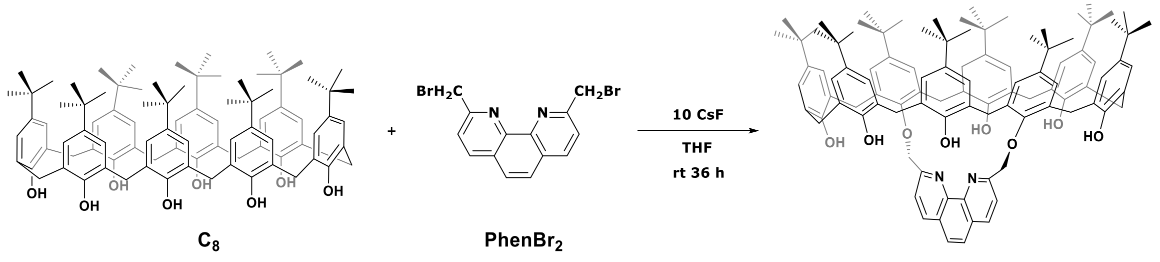
S.2.1. Synthesis of 2,9-bis(bromomethyl)-1,10-phenanthroline ($PhenBr_2$)



Scheme 1. Synthesis of 2,9-bis(bromomethyl)-1,10-phenanthroline ($PhenBr_2$).

Optimisation of the previously reported protocol[2] was adapted, with the reaction time modified. 2,9-dimethyl-1,10-phenanthroline (0.88 g, 4.17 mmol) was dissolved in 30 mL of acetonitrile in a 250 mL round-bottom Schlenk flask covered with black paper foil. Under constant stirring, recrystallised *N*-bromosuccinimide (NBS 2.23 g, 12.51 mmol) was added to the solution, then the temperature was raised to reflux with an oil bath and was maintained for 18 h. After that, volatiles were evaporated, the crude solid was extracted with diethyl ether, and washed with a saturated solution of $NaHCO_3$ (50 mL). Then, the solid was dissolved in 40 mL of THF, the brown solution was cooled to 0 °C. Thereupon, diethyl phosphite (1.1 mL, 8.34 mmol, 98%) and *N,N*-diisopropylethylamine (1.453 mL, 8.34 mmol, 99%) were added to the solution. After 24 h, volatiles were evaporated and the product was purified by column chromatography on silica gel with dichloromethane as eluant, resulting in an orange crystalline solid in 40% overall yield (0.61 g, 1.68 mmol), m.p. > 110 °C. δ_H (300 MHz, $CDCl_3$, 393.15 K) 8.26 (d, J = 8.34 Hz, 2 H, $Ar_{Phen(c)}$), 7.92 (d, J = 8.33 Hz, 2 H, $Ar_{Phen(a)}$), 7.82 (d, J = 7.80 , 2H, $Ar_{Phen(b)}$), 7.26 (s, 3 H, $CDCl_3$), 4.97 (s, 4 H, CH_{2Phen}). *m/z* (DART-MS) 367 [**$PhenBr_2H$**]$^+$.

S.2.2. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-*p*-tert-butylcalix[8]arene (C$_8$Phen)



C$_8$

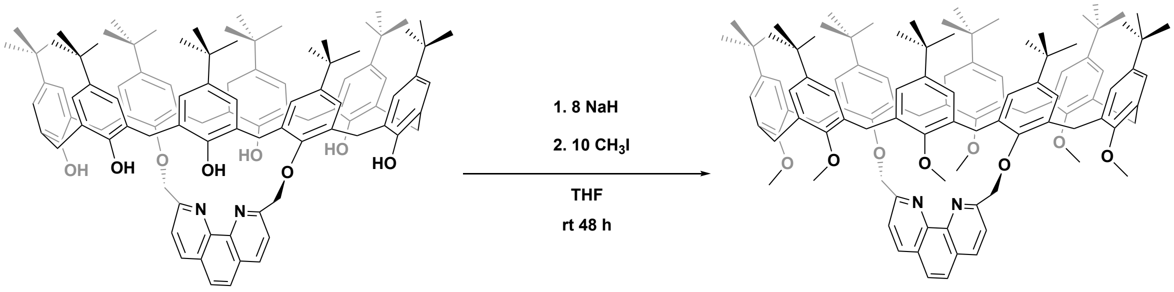
PhenBr$_2$

C$_8$Phen

Scheme 2. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-*p*-tert-butylcalix[8]arene (**C$_8$Phen**).

C$_8$Phen was obtained according to the previously reported protocol,[2] with modifications in the purification procedure. Previously synthesized *p*-tert-butylcalix[8]arene (obtained according to the literature procedure,[3] 0.80 g, 0.62 mmol) and CsF (0.94 g, 6.22 mmol) were dried in a round-bottom Schlenk flask at 120 °C for 2 h. After cooling to room temperature, the solids were dissolved in 15 mL of THF, the formation of a white suspension was observed. The temperature was increased to 50 °C for 12 h while stirring, resulting in a yellow solution. **PhenBr$_2$** (0.28 g, 0.73 mmol) was then added, and the mixture was stirred at room temperature for 36 h. Then, volatiles were evaporated under reduced pressure, and the crude solid was dissolved in 10 mL of chloroform/toluene (10:1). The organic phase was washed with 15 mL of 0.1 M HCl, followed by 50 mL of saturated NaHCO$_3$ solution, and extracted with 20 mL of chloroform. The organic phase was dried over CaCO$_3$, and after slow solvent evaporation an orange solid was obtained. The product was purified by column chromatography on silica gel with dichloromethane as eluant, and the product was washed with hexanes, resulting in a white crystalline solid in 86% yield (0.78 g, 0.52 mmol), m.p. > 230 °C. δ_H (300 MHz, CDCl$_3$, 393.15 K) 9.48 (s, 8 H, OH), 8.39 (d, J = 8.27 Hz, 2 H, Ar$_{\text{Phen(b)}}$), 7.78 (m, 2 H, Ar$_{\text{Phen(a)}}$), 7.24 (s, 2 H, Ar$_{\text{Phen(b)}}$), 7.18 (m, 16 H, Ar$_{\text{C8}}$), 5.26 (s, 4 H, CH$_{\text{2Phen}}$), 4.28 (d, J = 12.9, 8 H, CH$_{\text{2exo}}$), 3.50 (d, J = 12.9, 8 H, CH$_{\text{2endo}}$), 1.25 (m, 48 H, *t*-Bu), 1.17 (m, 24 H, *t*-Bu). *m/z* (FAB$^+$) 1502 [**C$_8$PhenH**]$^+$.

S.2.3. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-2,3,4,6,7,8-hexamethyl-*p*-*tert*-butylcalix[8]arene (C$_8$PhenMe$_6$)

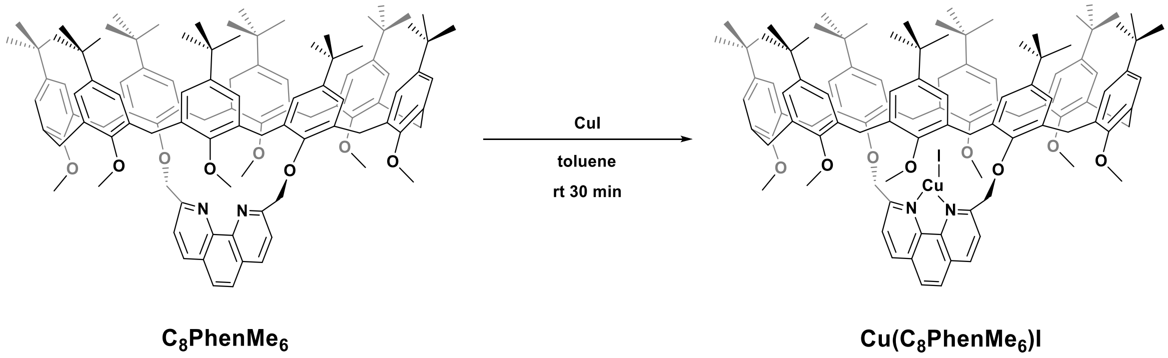


C$_8$Phen **C$_8$PhenMe$_6$**

Scheme 3. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-2,3,4,6,7,8-hexamethyl-*p*-*tert*-butylcalix[8]arene (**C$_8$PhenMe$_6$**).

Under inert atmosphere, the previously synthetized **C$_8$Phen** (500 mg, 0.33 mmol) was dissolved in 10 mL of anhydrous THF in a 50 mL round bottom Schlenk flask, NaH (52 mg, 2.17 mmol, 90%) was added, after one hour under constant stirring, the formation of a yellow precipitate was observed. Prior addition of CH_3I (312 mg, 2.17 mmol), the Schlenk flask was covered with black paper foil, then, the mixture was stirred at room temperature for 48 h, the formation of a white suspension was noticed. Volatiles were evaporated under reduced pressure, and the crude solid obtained was extracted with 50 mL mixture of dichloromethane/brine (1:1). The organic phase was dried over Na_2SO_4, and after slow evaporation of the solvent a colourless solid was obtained. The product was purified by column chromatography on silica gel with dichloromethane as eluant, and the product was washed with hexanes, resulting in a white crystalline solid in 80% yield (418 mg, 0.26 mmol). m.p. > 210 °C. δ_H (300 MHz, $C_2D_2Cl_4$, 293.15 K) 8.06 (m, 4H, $Ar_{Phen(ac)}$), 7.55 (s, 2H, $Ar_{Phen(b)}$), 7.07 (m, 4H, $Ar_{C8(c)}$), 6.94 (m, 4H, $Ar_{C8(a)}$), 6.75 (m, 4H, $Ar_{C8(d)}$), 6.65 (m, 4H, $Ar_{C8(d)}$), 4.34 (d, *J* = 16.5 Hz, 4H, $CH_{2C8(c)}$), 4.27 (s, 4H, CH_{2Phen}), 3.98 (d, *J* = 15.9 Hz, 4H, $CH_{2C8(a)}$), 3.59 (d, *J* = 16.1 Hz, 4H, $CH_{2C8(b)}$),), 3.49 (d, *J* = 15.7 Hz, 4H, $CH_{2C8(d)}$), 3.30 (s, 12 H, MeO_{ext}), 2.47 (s, 6H, $MeO_{ext(a)}$), 2.25 (s, 6H, $MeO_{ext(b)}$), 1.23 (s, 18H, *t*-$Bu_{ext(b)}$), 1.14 (s, 18H, *t*-$Bu_{ext(a)}$), 0.90 (s, 36H, *t*-Bu_{int}). ν ν_{max} (KBr) 2955 (CH_3 and CH_2), 2866 (CH_3 and CH_2) y 1479 (Ar). *m/z* (FAB$^+$) 1587 [**C$_8$PhenMe$_6$H**]$^+$. Elemental analysis for $C_{108}H_{134}N_2O_8 \cdot H_2O \cdot CHCl_3$, found: C, 75.87; H, 8.00; N, 1.62; requires: C, 75.88, H, 7.86, N, 1.43.

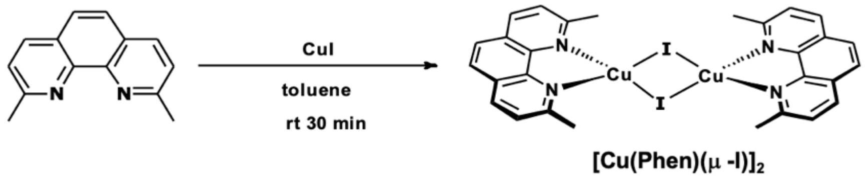
S.2.4. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-2,3,4,6,7,8-hexamethyl-*p*-tert-butylcalix[8]arene copper(I) complex [Cu(C_8PhenMe$_6$)I]



Scheme 4. Synthesis of 1,5-(2,9-dimethyl-1,10-phenanthroline)-2,3,4,6,7,8-hexamethyl-*p*-tert-butylcalix[8]arene copper(I) complex [**Cu(C_8PhenMe$_6$)I**].

C_8PhenMe$_6$ (30 mg, 0.018 mmol) and CuI (3.5 mg, 0.018 mmol) were placed in a 50 mL Schlenk tube with 2 mL of toluene under constant stirring for 30 min at room temperature under N_2. Volatiles were then evaporated under reduced pressure, resulting in an orange crystalline solid in 92% yield (30 mg, 0.016 mmol), m.p. > 230 °C. δ_H (300 MHz, toluene-d_8, 393.15 K) 8.34 (d, J = 8.5 Hz, 1 H, Ar_{Phen}), 8.18 (d, 1 H, Ar_{Phen}), 7.68 (m, 1 H, Ar_{Phen}), 7.62 (m, 1 H, Ar_{Phen}), 7.16 (m, 8 H, Ar_{C8}), 7.09 (m, 3 H, $Ar_{toluene}$), 7.00 (m, 8 H, $Ar_{toluene}$), 6.98 (m, 2 H, $Ar_{toluene}$), 5.25 (m, 2 H, CH_{2Phen}), 4.87 (d, J = 16 Hz, 2 H, CH_{2exo}), 4.73 (m, 2 H, CH_{2exo}), 4.52 (m, 5 H, CH_{2exo} y CH_{2Phen}), 4.21 (d, J = 16 Hz, 4 H, CH_{2exo}), 3.75 (m, 4H, CH_{2exo}), 4.18 (m, 4 H, CH_{2endo}), 3.81 (m, 4 H, CH_{2endo}), 3.59 (s, 9 H, CH_{3int}), 3.13 (s, 9 H, CH_{3ext}), 2.09 (s, 3 H, $CH_{3toluene}$), 1.30 (s, 36 H, *t*-Bu_{ext}), 1.18 (m, 36 H, *t*-Bu_{int}). *m/z* (FAB^+) 1649 [**Cu(C_8PhenMe$_6$)**]$^+$.

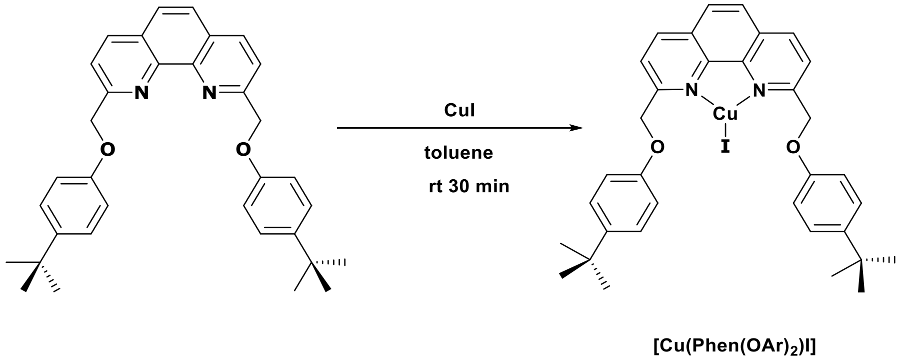
S.2.5. Synthesis of the bis(2,9-dimethyl-1,10-phenanthroline) copper(I) complex [Cu(Phen)(μ-I)]$_2$



Scheme 5. Synthesis of the bis(2,9-dimethyl-1,10-phenanthroline) copper(I) complex **[Cu(Phen)(μ-I)]**$_2$.

2,9-dimethyl-1,10-phenanthroline (32 mg, 0.15 mmol) and CuI (29 mg, 0.15 mmol) were dissolved in 2 mL toluene in a 50 mL Schlenk tube under N_2, the mixture was under constant stirring for 30 min at room temperature. Volatiles were evaporated under reduced pressure, resulting in a red solid in 37% yield (45 mg, 0.06 mmol). The low yield is due to the conditions were imposed to obtain the (1:1) Phen-CuI complex. m.p. > 230 °C. δ_H (300 MHz, toluene-d8, 393.15 K) 8.55 (d, *J* = 7.80 Hz, 2 H, $Ar_{Phen(c)}$), 8.06 (s, 2 H, $Ar_{Phen(a)}$), 7.82 (d, *J* = 7.9 Hz, 2 H, $Ar_{Phen(b)}$), 7.26 (s, 3 H, $CDCl_3$), 2.45 (s, 6 H, CH_{3Phen}). *m/z* (DART-MS) 479 **[Cu(Phen)₂]**$^+$.

S.2.6. Synthesis of the 2,9-bis((4-(tert-butyl)phenoxy)methyl)-1,10-phenanthroline copper(I) complex [Cu(Phen(OAr)₂)I]



Scheme 6. Synthesis of the 2,9-bis((4-(tert-butyl)phenoxy)methyl)-1,10-phenanthroline copper(I) complex **[Cu(Phen(OAr)₂)I]**.

2,9-bis((4-(*tert*-butyl)phenoxy)methyl)-1,10-phenanthroline (obtained according to the literature procedure,[4] 60 mg, 0.12 mol) and CuI (23 mg, 0.12 mol) were dissolved in 2 mL toluene in a 50 mL Schlenk tube under N_2, the mixture was stirred for 30 min at room temperature. Volatiles were evaporated under reduced pressure, resulting in a red solid in 93% yield (76 mg, 0.11 mmol), m.p. > 230 °C. δ_H (300 MHz, $CDCl_3$, 298.15 K) 8.31 (d, *J* = 8.3 Hz, 2H), 7.97 (d, *J* = 8.3 Hz, 2H), 7.81 (s, 2H), 7.34 (d, *J* = 8.9 Hz, 4 H), 7.26 (s, 3H, $CDCl_3$), 7.02 (m, 4H), 5.65 (s, 4H), 1.31 (s, 18H). *m/z* (ESI-MS) 567.3 **[Cu(Phen(OAr)₂)]**$^+$.

S.3. Characterization

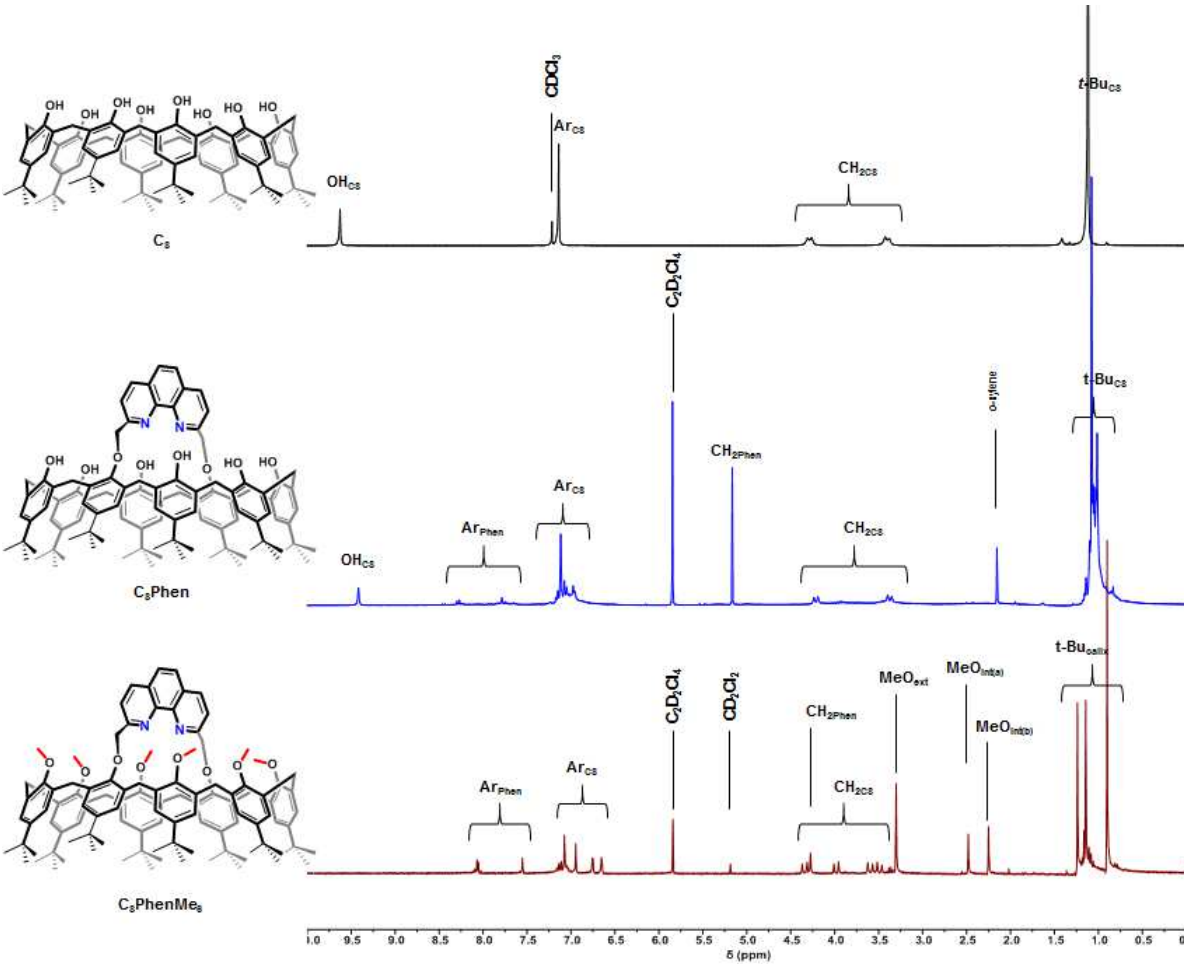
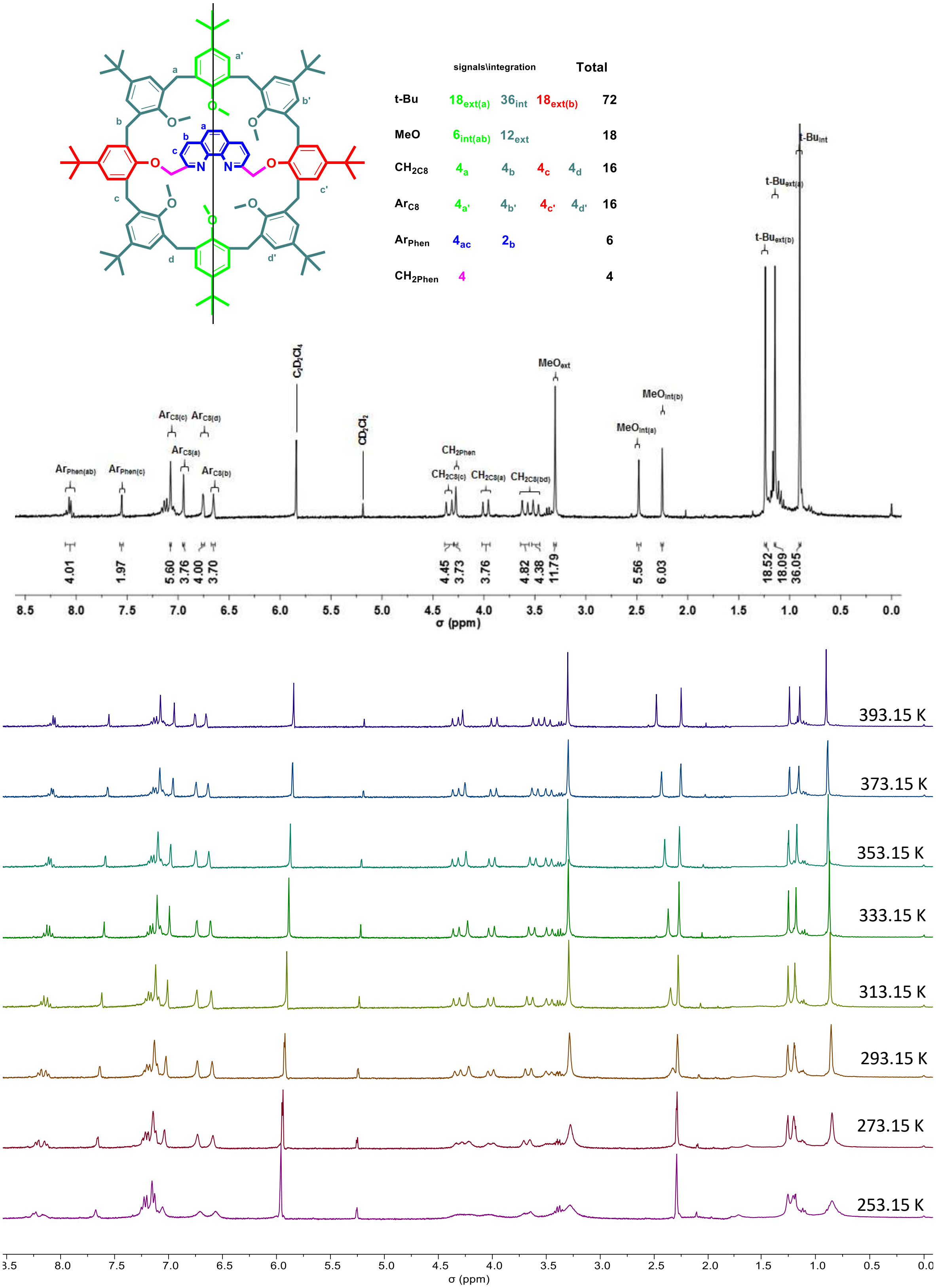


Figure S1. Comparison of the ^{1}H NMR spectra of the macrocycles **C$_8$** in CDCl$_3$ (black) at 293.15 K, **C$_8$Phen** in C$_2$D$_2$Cl$_4$ (blue) at 293.15 K, and **C$_8$PhenMe$_6$** in C$_2$D$_2$Cl$_4$ (red) at 393.15 K.



	signals\integration				Total
t-Bu	$18_{ext(a)}$	36_{int}	$18_{ext(b)}$		72
MeO	$6_{int(ab)}$	12_{ext}			18
CH_{2C8}	4_a	4_b	4_c	4_d	16
Ar_{C8}	$4_{a'}$	$4_{b'}$	$4_{c'}$	$4_{d'}$	16
Ar_{Phen}	4_{ac}	2_b			6
CH_{2Phen}	4				4

Figure S2. Top: analysis of the ^1H NMR spectrum of **$C_8PhenMe_6$** in $C_2D_2Cl_4$ at 393.15 K. Bottom: VT ^1H NMR spectra of **$C_8PhenMe_6$** in $C_2D_2Cl_4$ from 253.15 to 393.15 K.

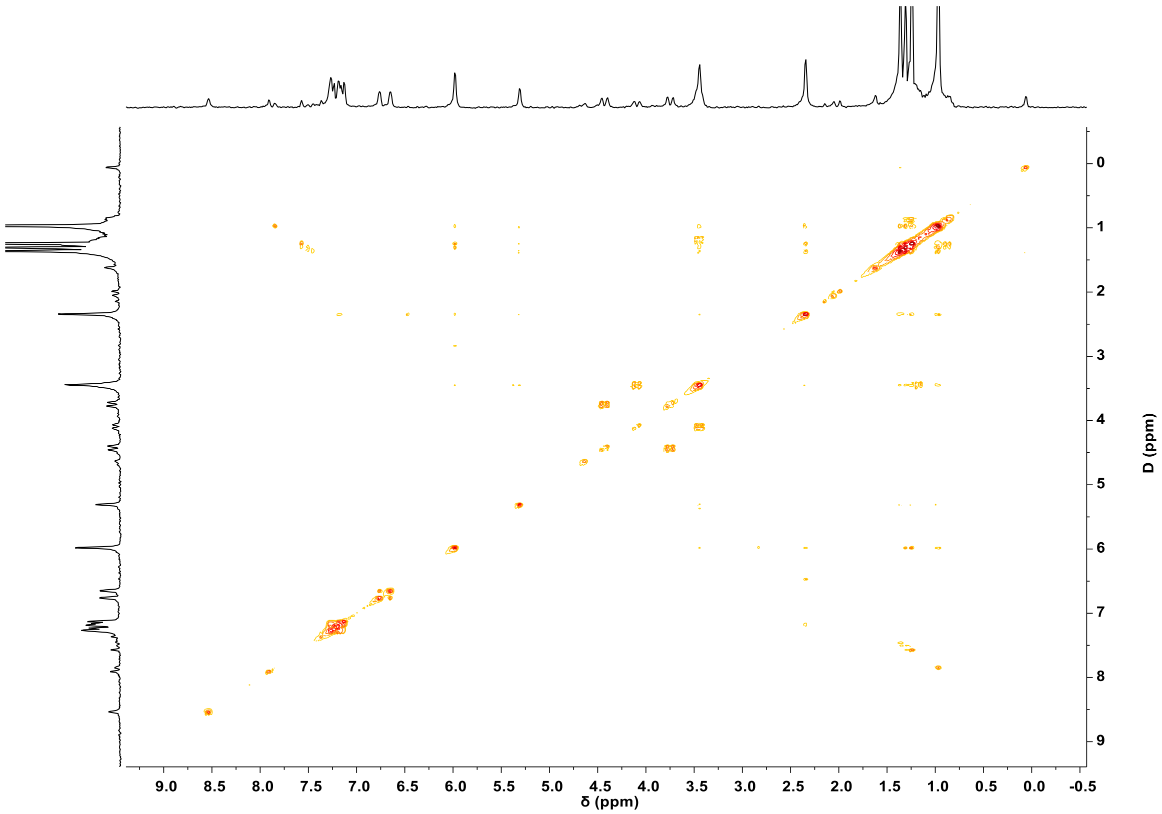


Figure S3. COSY NMR spectrum of **C$_8$PhenMe$_6$** in $C_2D_2Cl_4$ at 298.15 K.

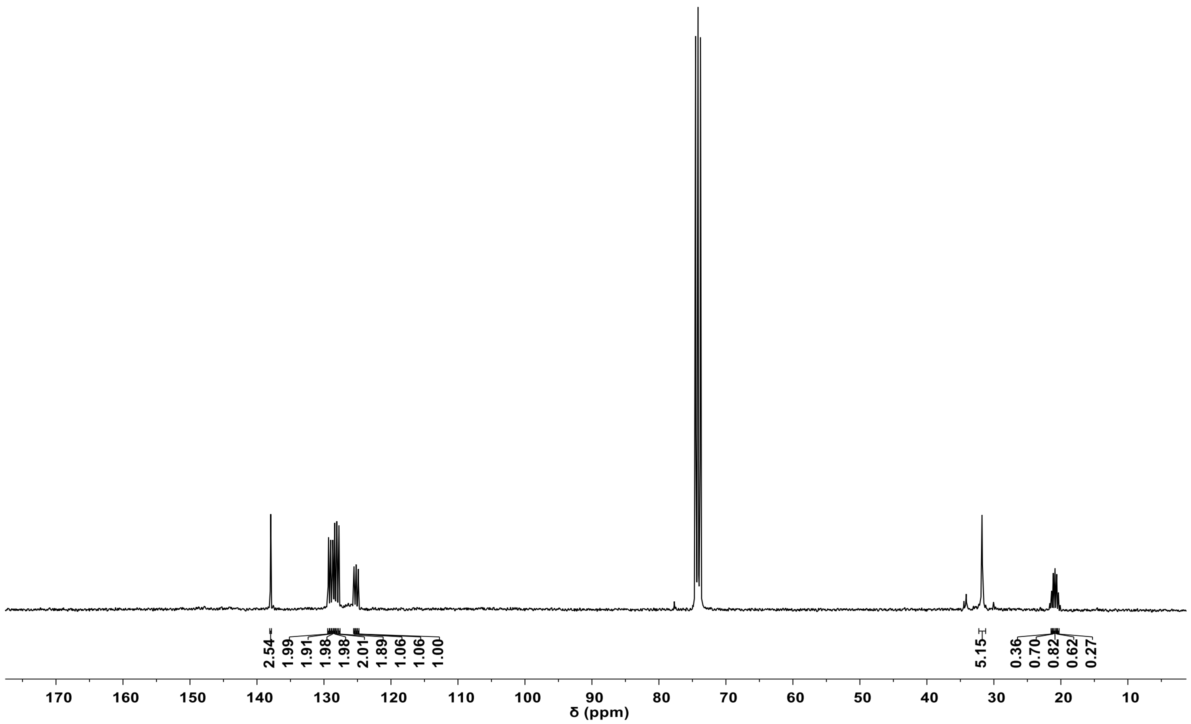


Figure S4. ^{13}C NMR spectrum of **C$_8$PhenMe$_6$** in $C_2D_2Cl_4$ at 298.15 K.

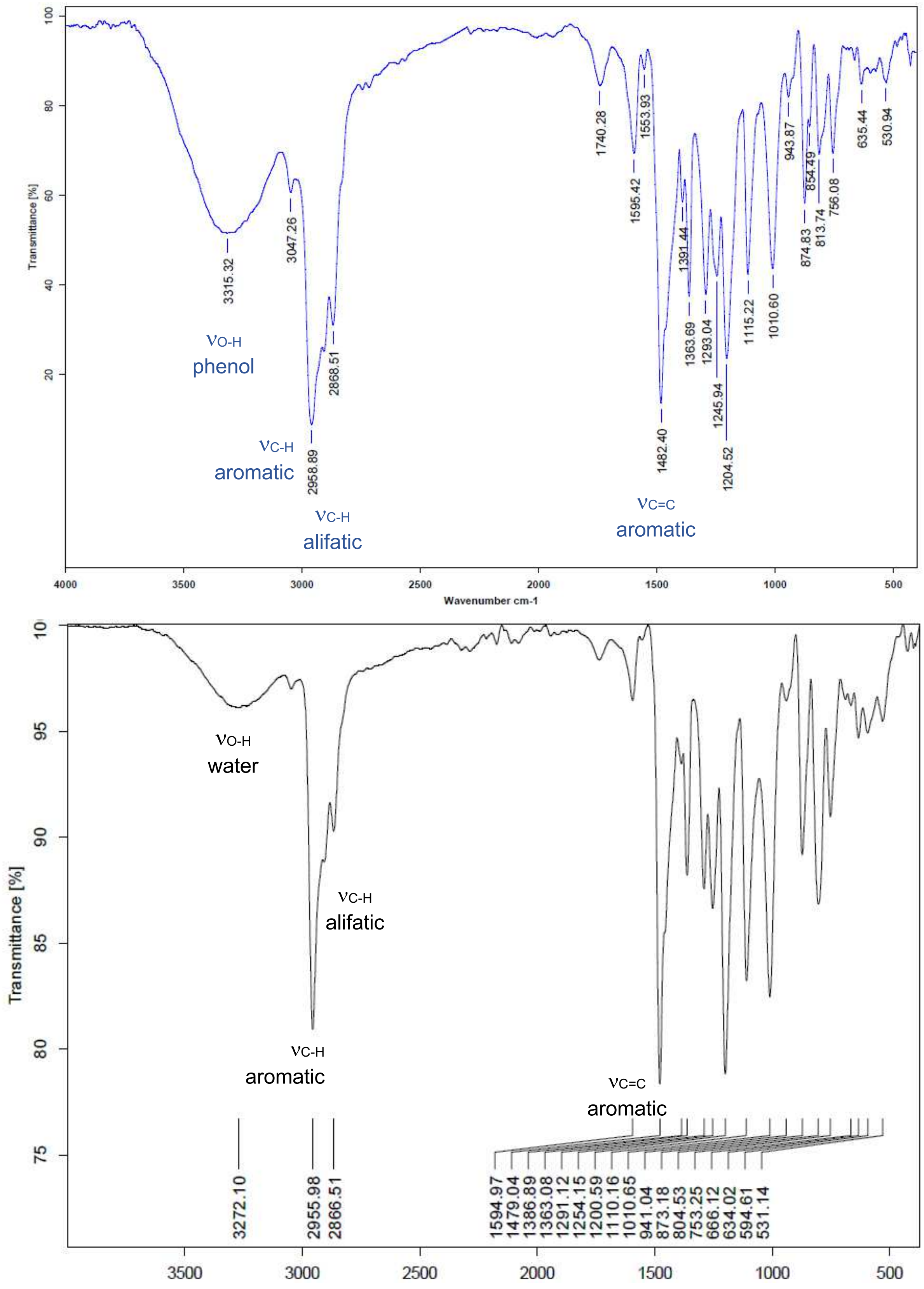


Figure S5. Comparison of IR spectra of **C$_8$Phen** (blue) and **C$_8$PhenMe$_6$** (black).

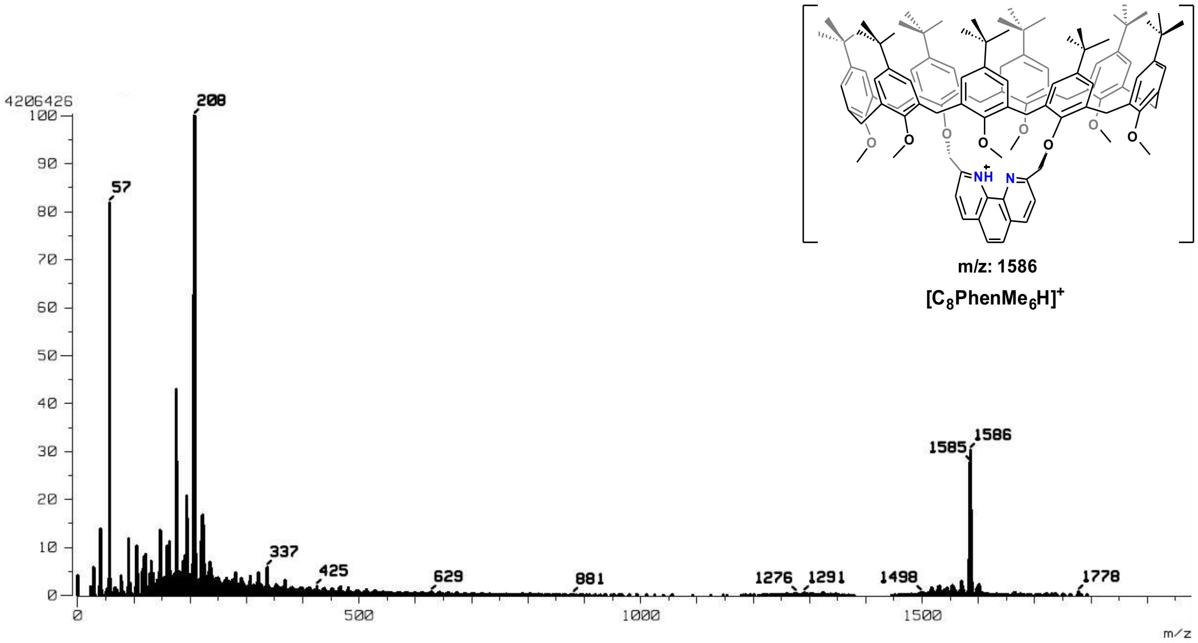


Figure S6. FAB$^+$ MS of **C$_8$PhenMe$_6$**.

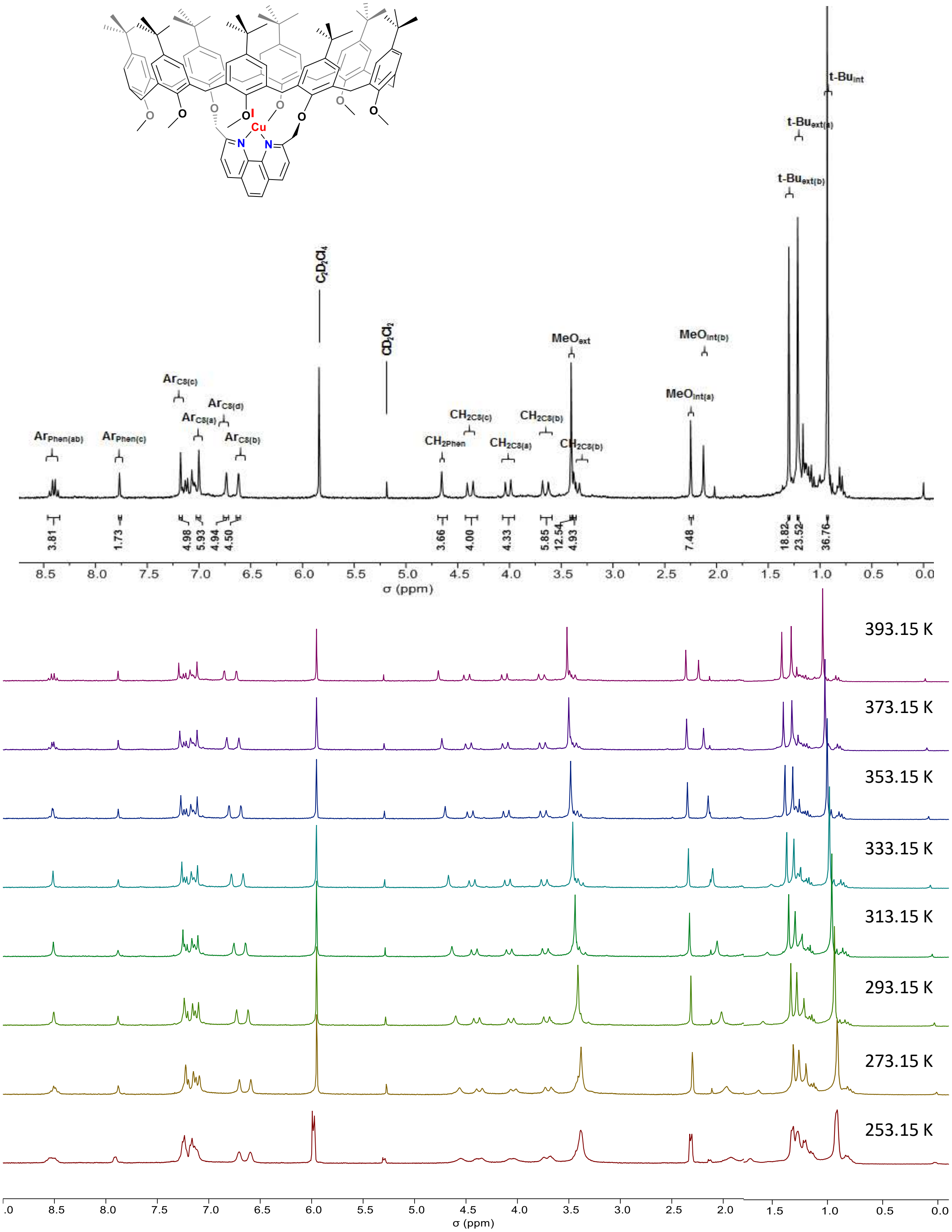


Figure S7. Top: analysis of the ^{1}H NMR spectrum of **[Cu(C$_8$PhenMe$_6$)I]** in $C_2D_2Cl_4$ at 393.15 K. Bottom: VT ^{1}H NMR spectra in $C_2D_2Cl_4$ from 253.15 to 393.15 K.

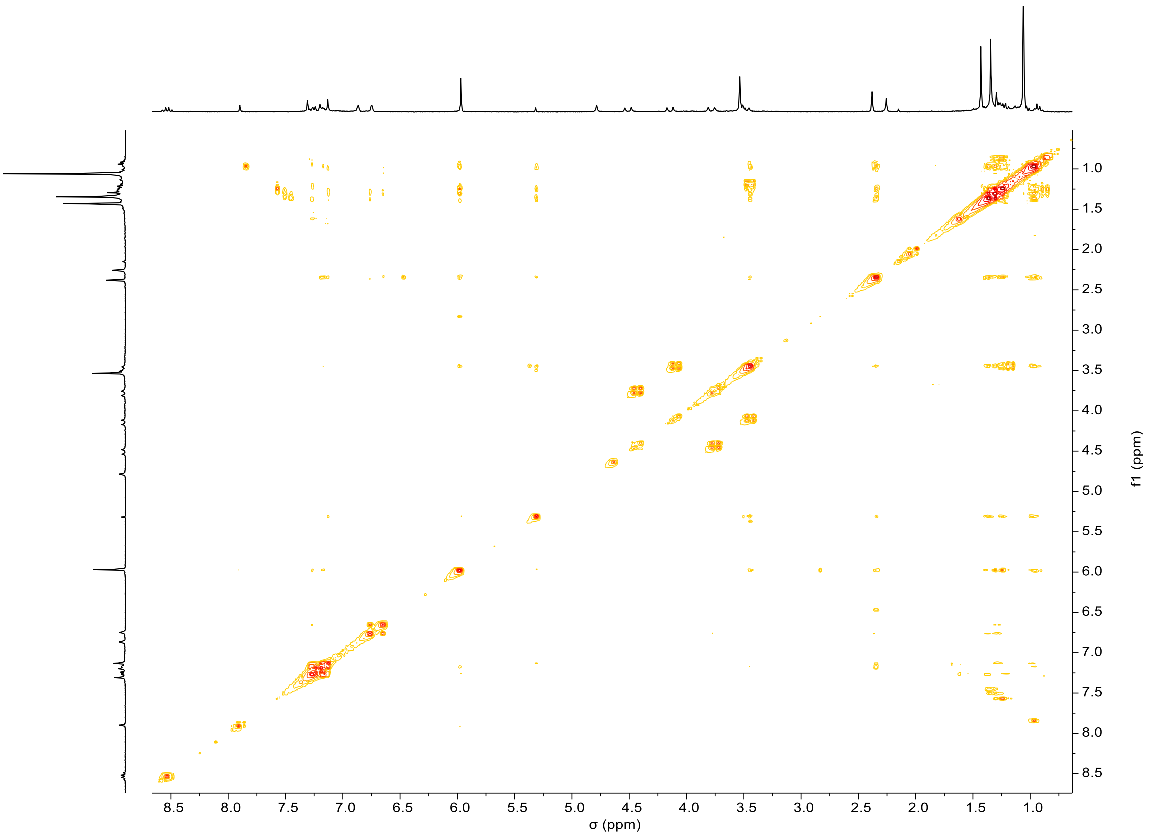


Figure S8. COSY NMR spectrum of **[Cu(C_8PhenMe$_6$)I]** in $C_2D_2Cl_4$ at 298.15 K.

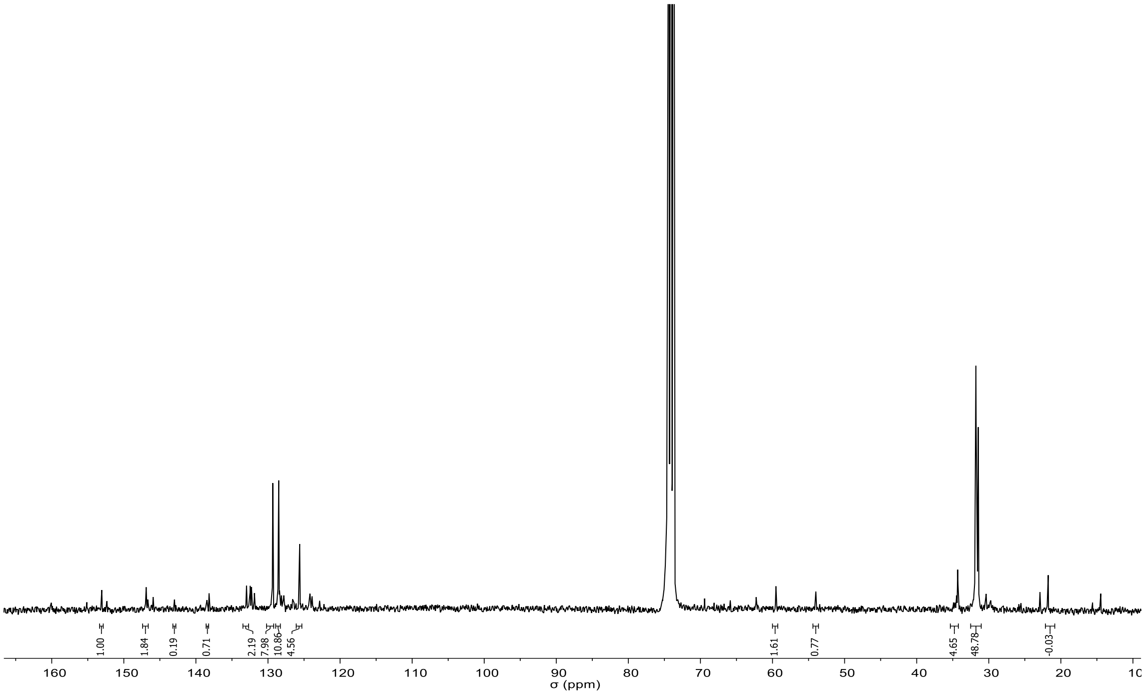


Figure S9. ^{13}C NMR spectrum of **[Cu(C_8PhenMe$_6$)I]** in $C_2D_2Cl_4$ at 293.15 K.

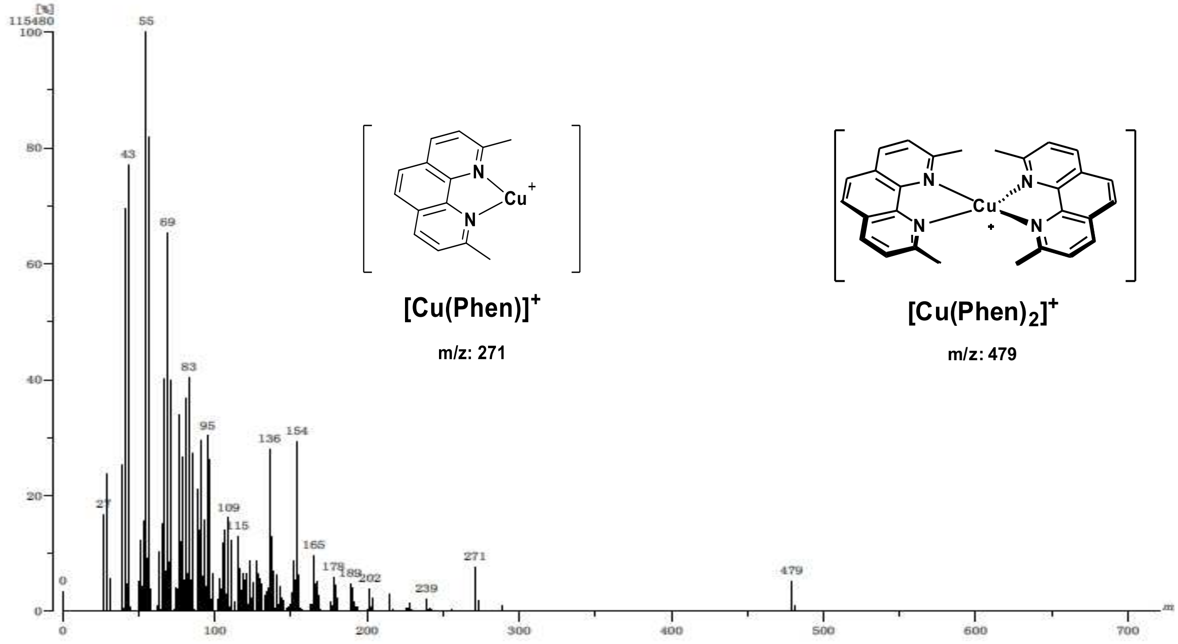


Figure S10. DART MS of **[Cu(Phen)I]$_2$** in CHCl$_3$.

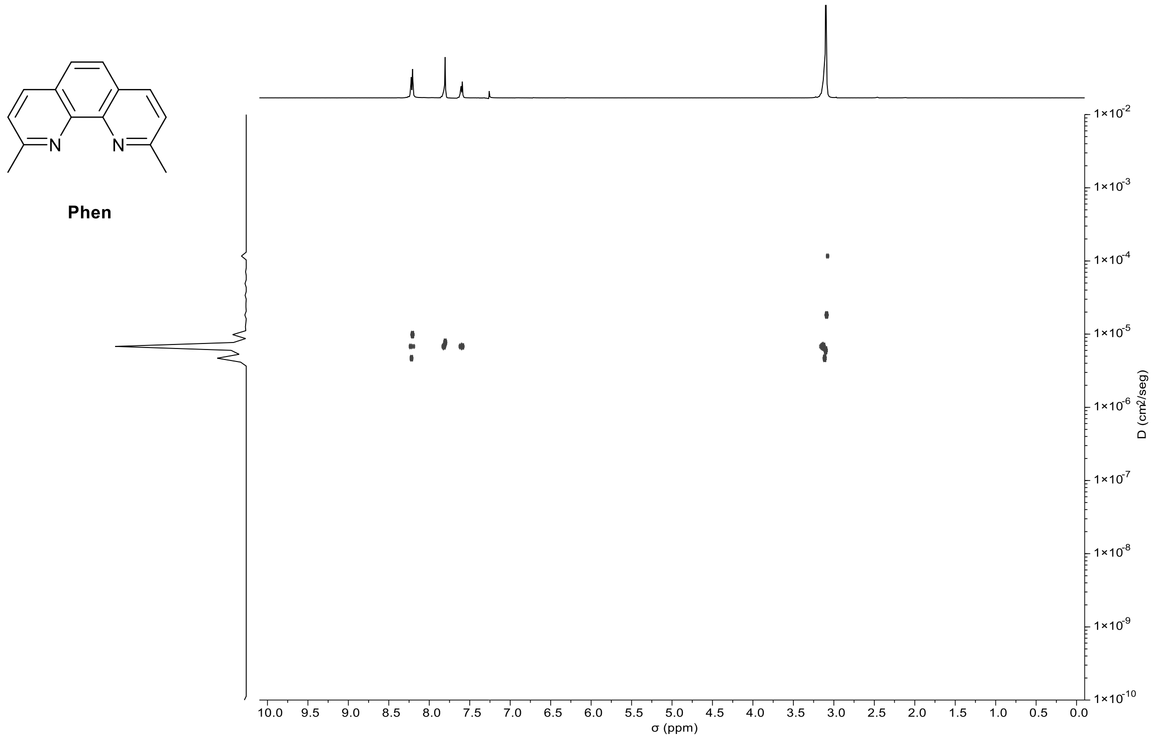


Figure S11. ^{1}H DOSY spectrum of the ligand **Phen** (8 mM in CDCl$_3$|toluene-d_8 500 µL|50 µL) at 298 K, diffusion = 6.80 x 10^{-6} cm^2/s, r$_H$ = 2.38 Å.

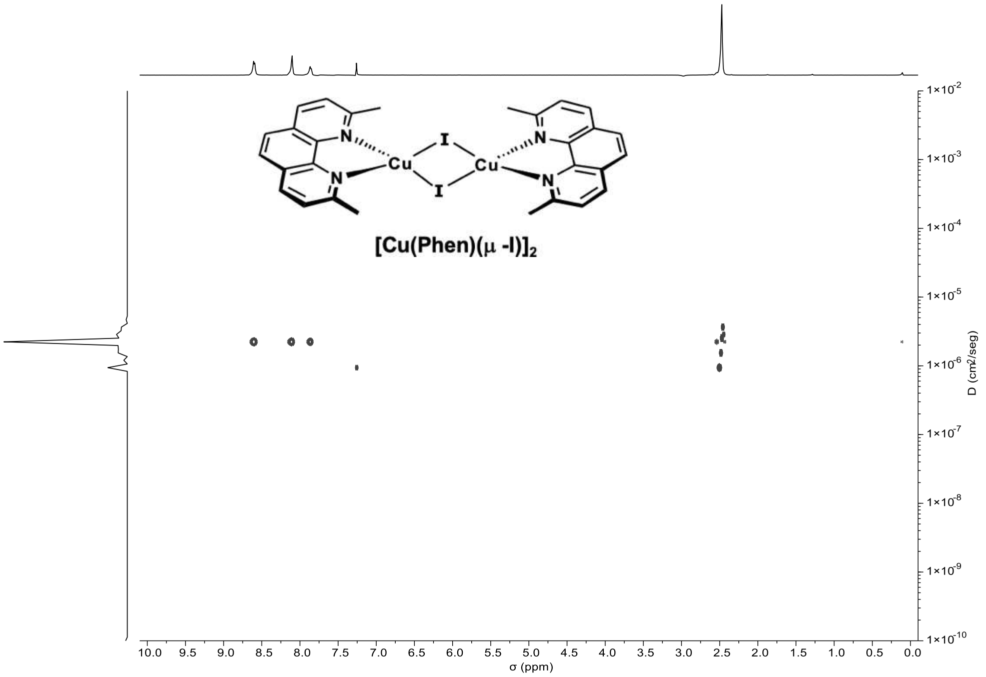


Figure S12. ^{1}H DOSY NMR spectrum of **[Cu(Phen)(μ-I)]$_2$** (8 mM in CDCl$_3$|toluene-d_8 500 μL|50 μL) at 298 K, diffusion = 2.23 x 10^{-6} cm^2/s, r$_H$ = 7.25 Å.

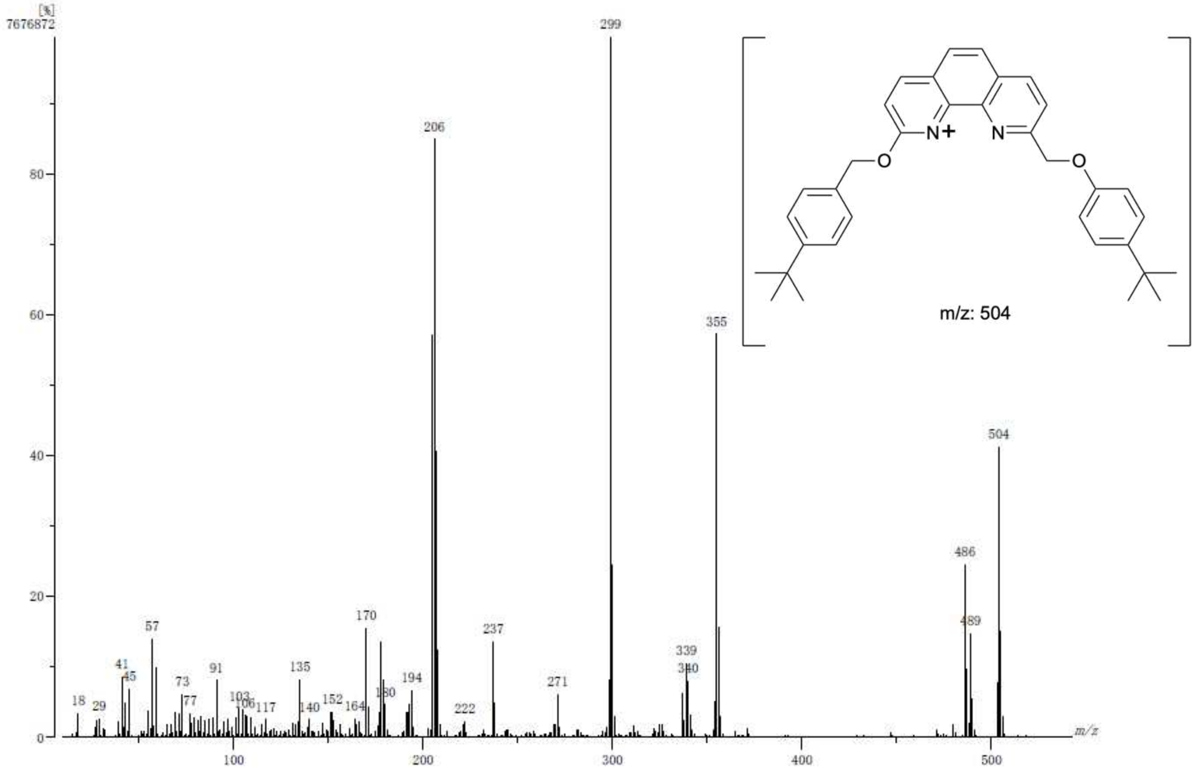


Figure S13. DART MS of **Phen(OAr)$_2$.**

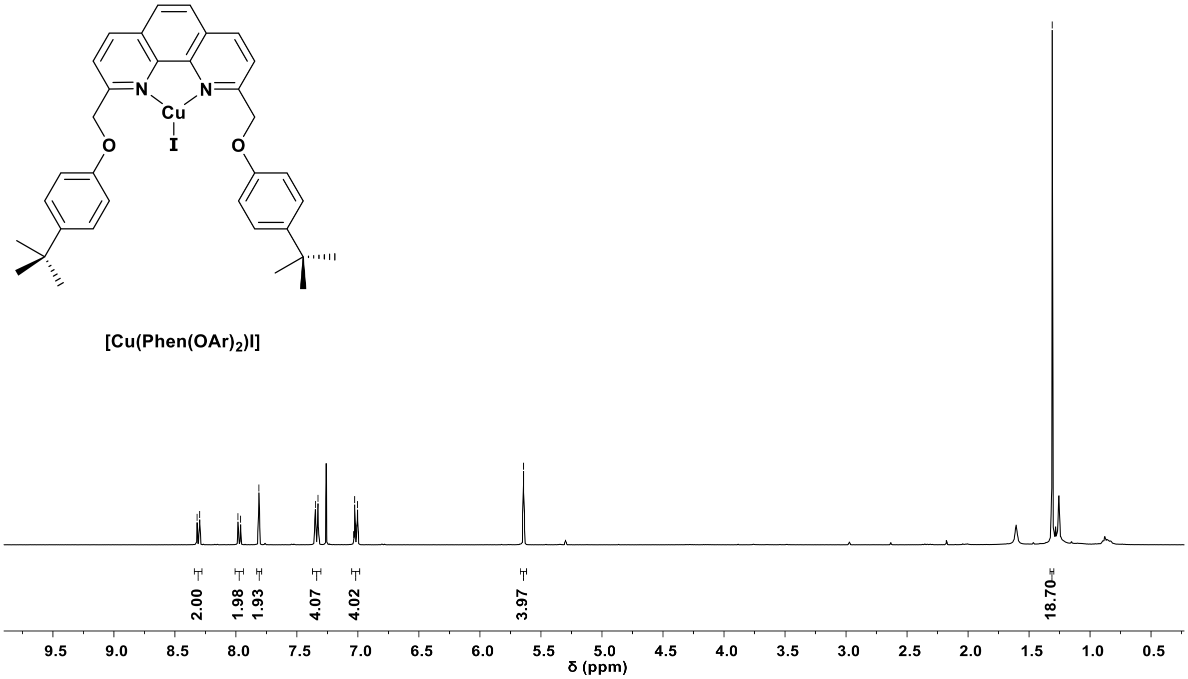


Figure S14. ^{1}H NMR spectrum of **[Cu(Phen(OAr)$_2$I]** in CDCl$_3$ at 293.15 K.

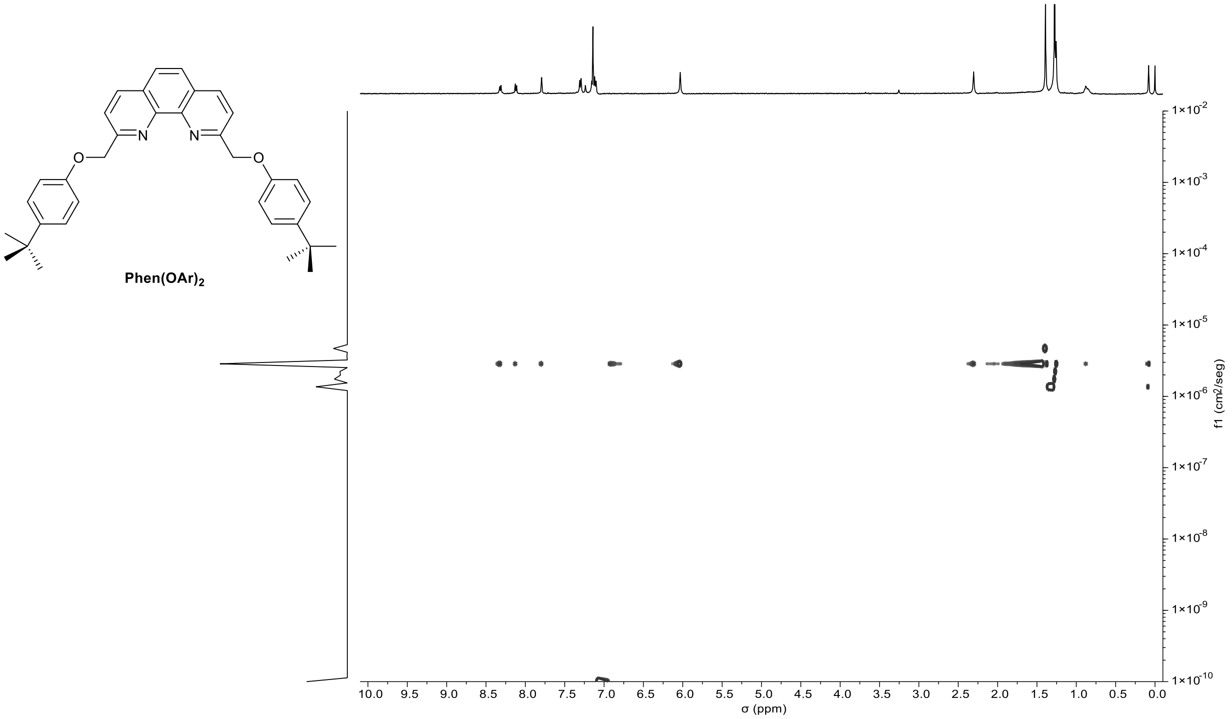


Figure S15. ^{1}H DOSY spectrum of **Phen(OAr)$_2$** (8 mM in CDCl$_3$|toluene-d_8 500 µL|50 µL) at 298 K, diffusion = 2.86 x 10^{-6} cm^2/s, r$_H$ = 5.65 Å.

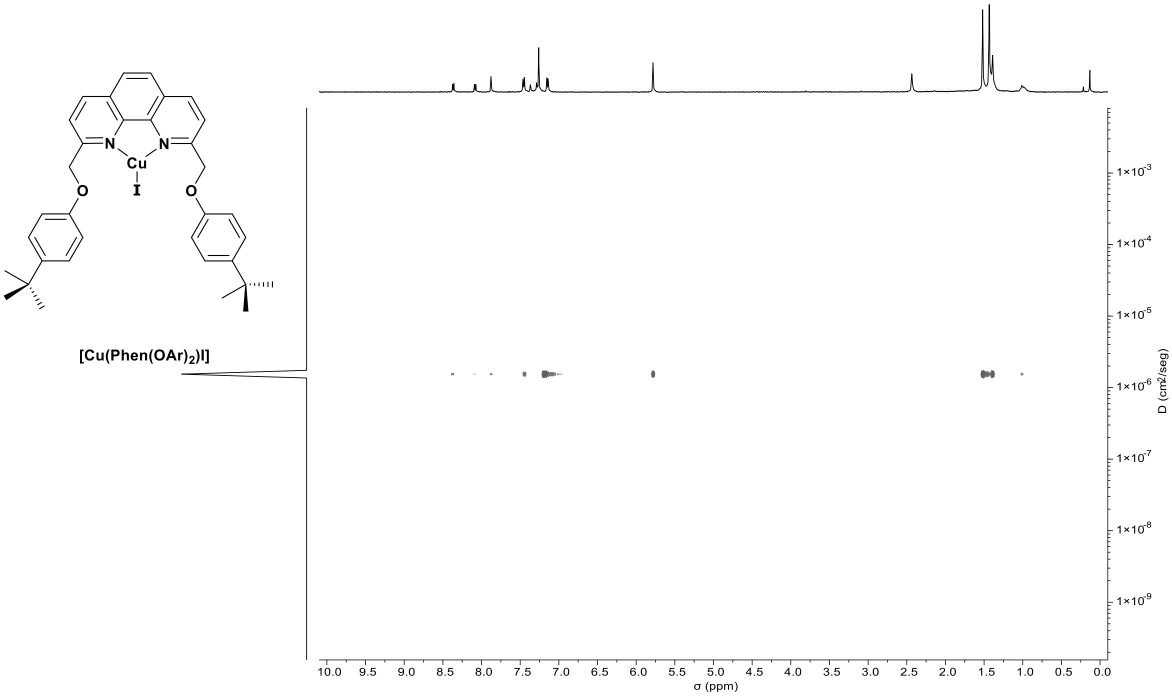


Figure S16. DOSY NMR spectrum of **[Cu(Phen(OAr)$_2$I]** (8 mM in CDCl$_3$|toluene-*d*$_8$ 500 µL|50 µL) at 298 K, diffusion = 1.53 x 10^{-6} cm^2/s, r$_H$ = 10.56 Å.

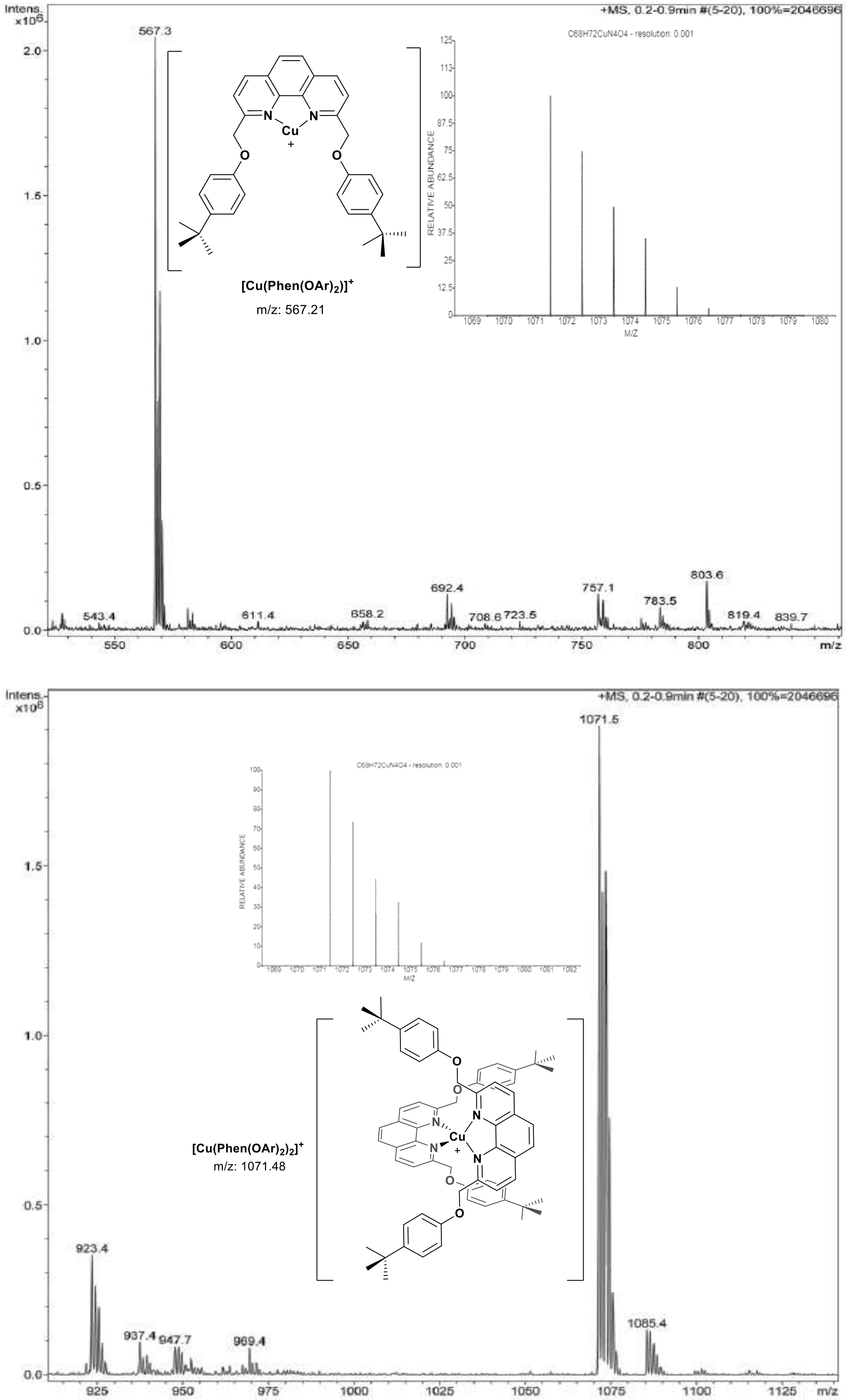
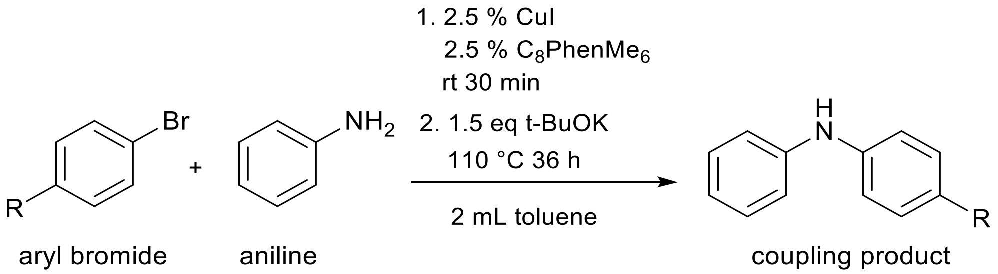


Figure S17. Top: ESI MS of **[Cu(Phen(OAr)$_2$)]$^+$** and its calculated isotopic pattern. Bottom: high *m/z* portion of ESI MS showing **[Cu(Phen(OAr)$_2$)$_2$]$^+$** and its calculated isotopic pattern.

S.4. General procedure for C-N cross-coupling reactions and characterization of the coupling products

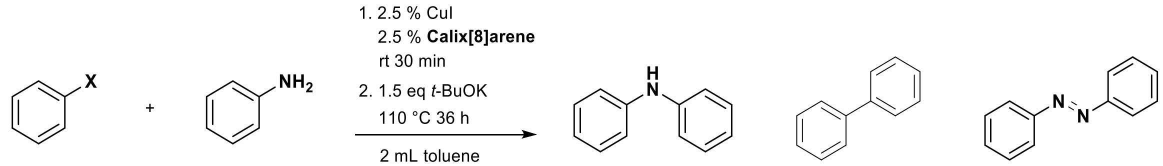


Scheme 7. General coupling procedure.

A 25 mL Schlenk flask was charged with 0.02 mmol (4 mg, 2.5 % mol) of CuI and 0.02 mmol (30 mg, 2.5 % mol) of **C$_8$PhenMe$_6$** and 2.0 mL of toluene, and the mixture was stirred for 30 min under dinitrogen. After addition of 0.80 mmol (73 μL, 74 mg) of aniline, 0.80 mmol of the corresponding aryl bromide, and 1.2 mmol (150 mg) of *t*-BuOK, the temperature was raised to 110 °C in an oil bath. Reactions were monitored by TLC carried out on 0.25 mm plates coated with silica gel, using UV light (254 nm) as visualising agent, until the consumption of the starting materials, or when a significant amount of the likely coupling products was evident, then the reaction was quenched by exposing it to air. The crude products were extracted with diethyl ether, separated and isolated by flash column chromatography performed on aluminium oxide 90 basic Macherey-Nagel as the stationary phase and hexane-dichloromethane at different ratios as the eluent. Yields refer to isolated compounds and their identity was confirmed by comparison with literature spectroscopic data. The same conditions were employed for comparison with cavitand **C$_8$Phen**, the catalyst blank **[Cu(Phen)I]$_2$**, and the molecular analogue **[Cu(Phen(OAr)$_2$)I]**; mercury drop tests were carried out under the same conditions.

S.4.1. Optimisation of the C-N cross-coupling reaction

Table S1. Optimization of the reaction conditions for C-N cross-coupling.[a]



aryl halide **aniline** **1** **2** **3**

Entry	Ligand	Aryl halide	Base	Yield[b] (%)		
				1	2	3
1	**C$_8$PhenMe$_6$**	chlorobenzene	*t*-BuOK	---	---	---
2	**C$_8$PhenMe$_6$**	bromobenzene	*t*-BuOK	73	8	---
3	**C$_8$PhenMe$_6$**	iodobenzene	*t*-BuOK	56	10	---
4	**C$_8$PhenMe$_6$**	bromobenzene	MeONa	20	10	---
5	**C$_8$PhenMe$_6$**	bromobenzene	NaOH	---	12	---
6	**C$_8$PhenMe$_6$**	bromobenzene	DBU	---	8	---
7	**C$_8$PhenMe$_6$**	bromobenzene	Cs_2CO_3	---	10	---
8[c]	**C$_8$Phen**	bromobenzene	*t*-BuOK	50	12	7
9[c]	**C$_8$Phen**	iodobenzene	*t*-BuOK	50	12	7
10[d]	**Phen**	bromobenzene	*t*-BuOK	20	10	---
11[d]	**Phen(OAr)$_2$**	bromobenzene	*t*-BuOK	26	8	10
12[e]	Without	bromobenzene	*t*-BuOK	---	9	---
13[f]	**C$_8$PhenMe$_6$**	bromobenzene	*t*-BuOK	68	6	---
14[g]	**C$_8$PhenMe$_6$**	bromobenzene	*t*-BuOK	52	12	---
15[h]	**C$_8$PhenMe$_6$**	bromobenzene	*t*-BuOK	---	---	---
16[i]	**C$_8$PhenMe$_6$**	bromobenzene	*t*-BuOK	---	---	---

[a] The evaluations were performed in a Schlenk flask (25 mL) with 2 mL of toluene, temperature was raised to 110 °C in an oil bath for 36 h under a nitrogen atmosphere. Reaction conditions: molar ratio of 1:1 of aryl halide and aniline, with 2.5 % mol of ligand and 2.5 % mol of CuI. [b] Isolated yields after column chromatography. [c] Also 15 % of triphenylamine was isolated. [d] 10 % mol of ligand and 10 % mol of CuI. [e] Without ligand and 10 % mol of CuI. [f] Mercury drop-test was performed. [g] THF as solvent. [h] Neat. [i] 2.5 % mol of CuCl.

S.4.2. Characterisation of the C-N coupling products

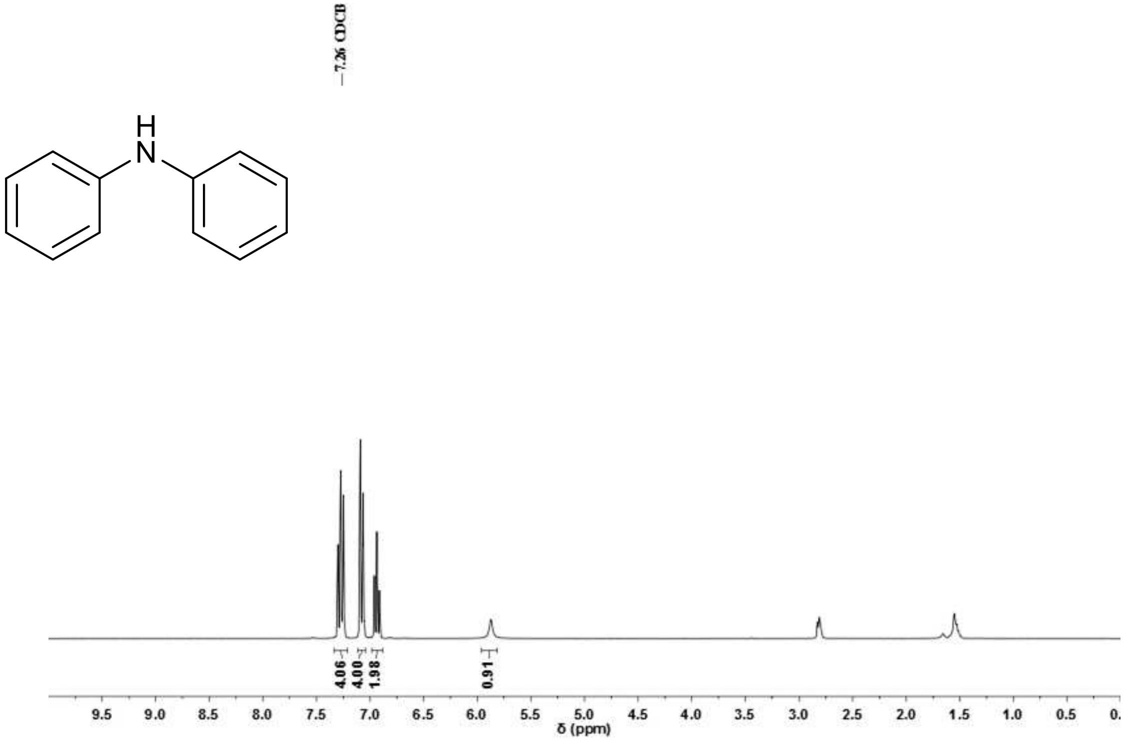


Figure S18. ^{1}H NMR spectrum of diphenylamine.

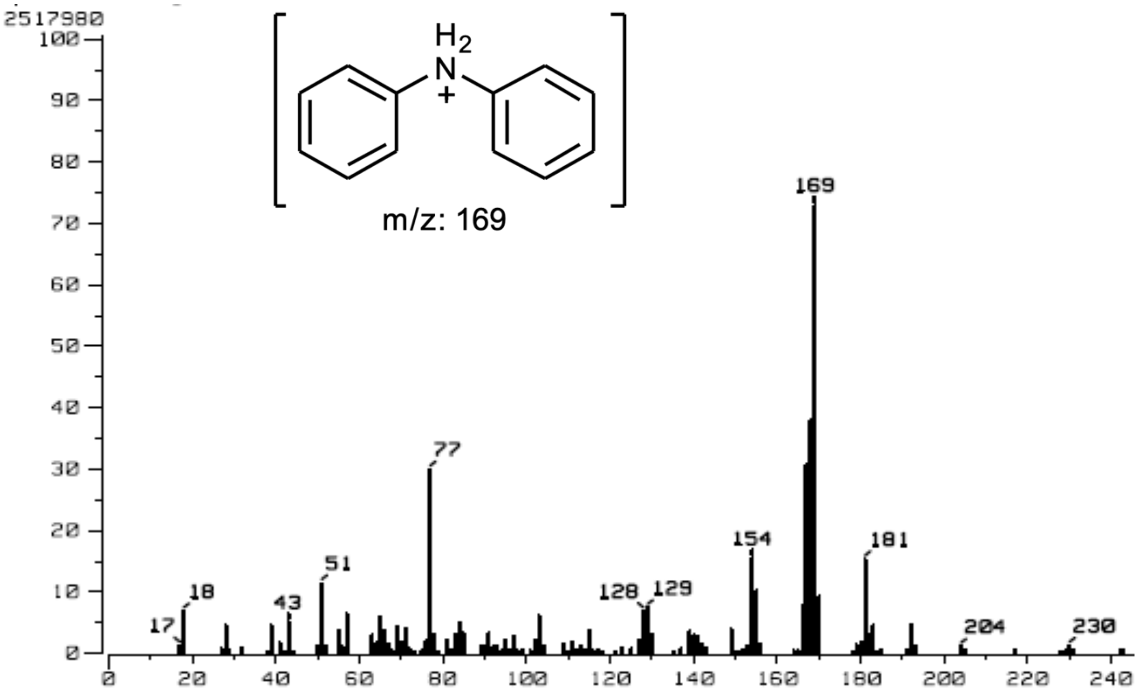


Figure S19. DART MS of diphenylamine.

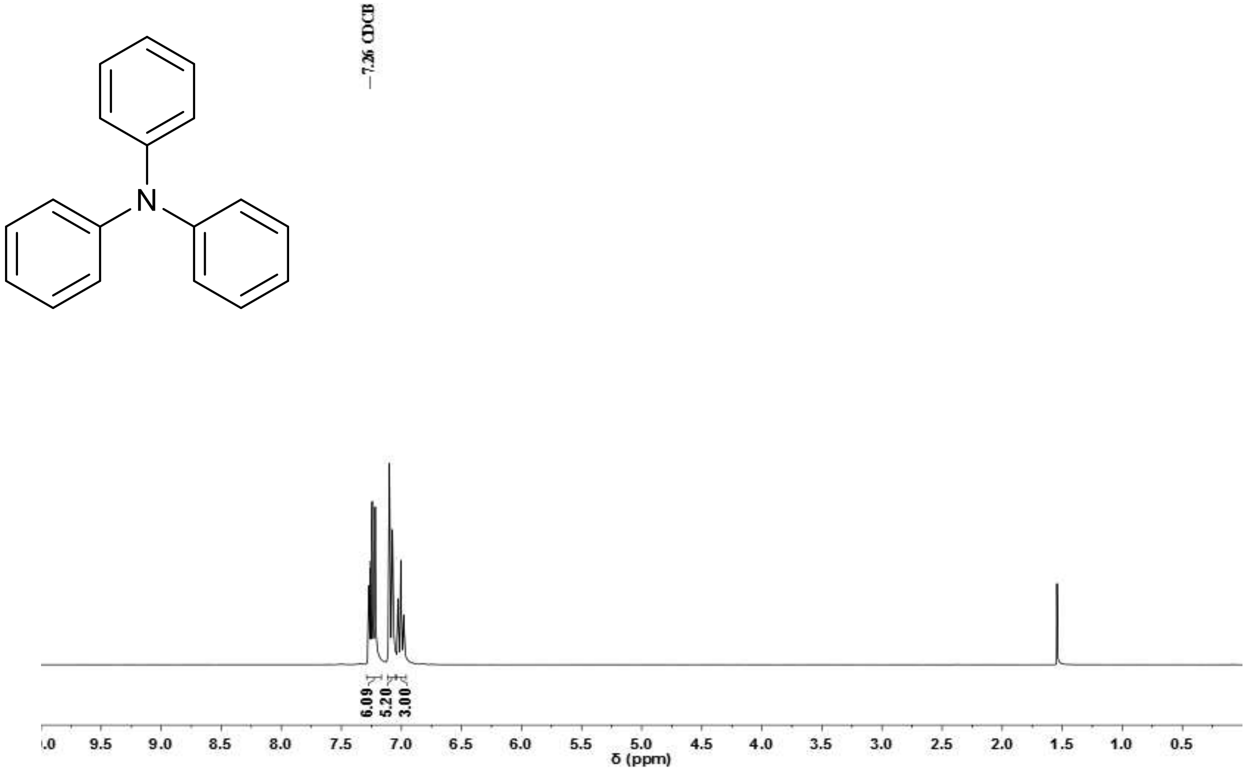


Figure S20. ^{1}H NMR spectrum of triphenylamine.

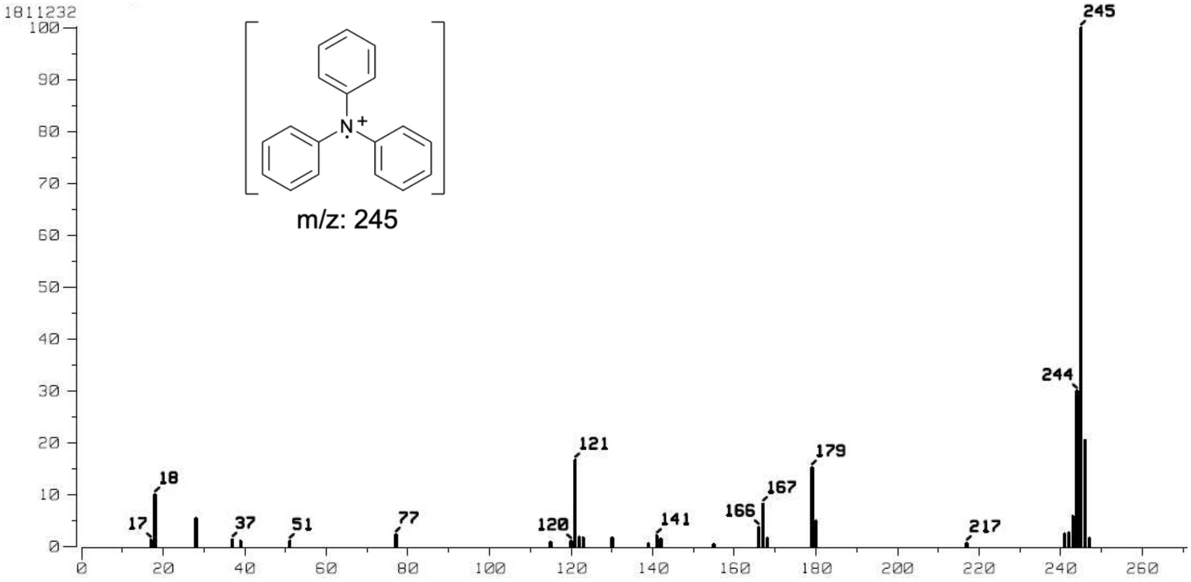


Figure S21. DART MS of triphenylamine.

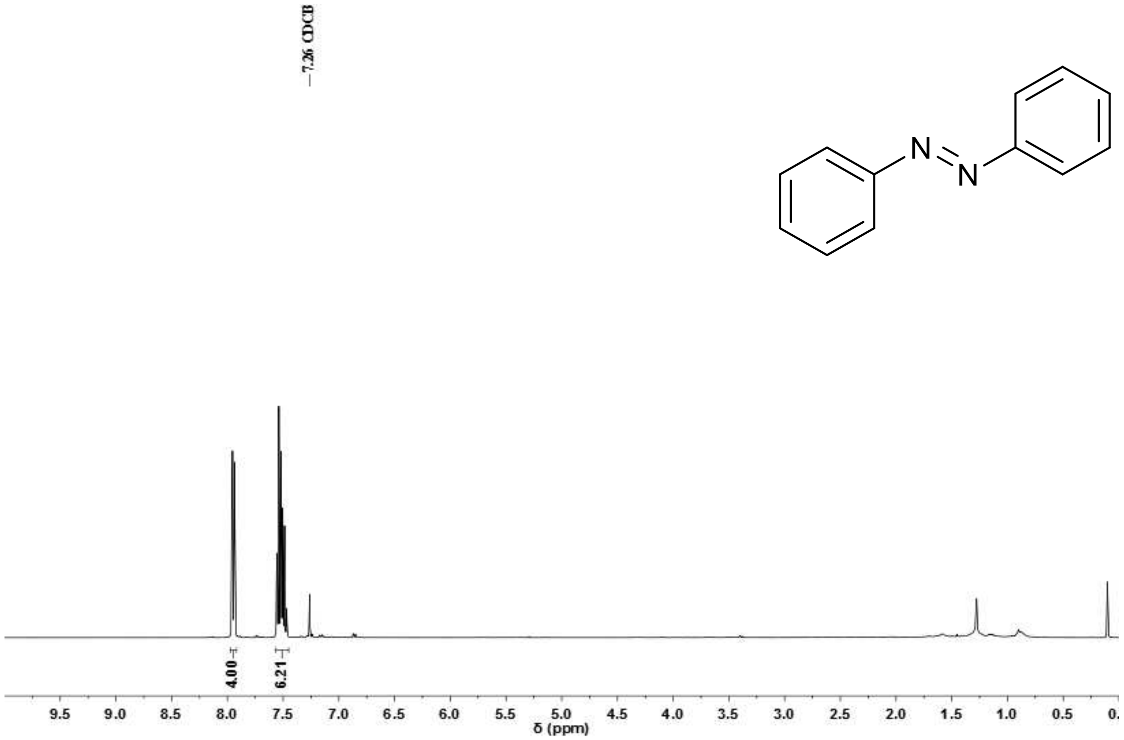


Figure S22. ^{1}H NMR spectrum of azobenzene.

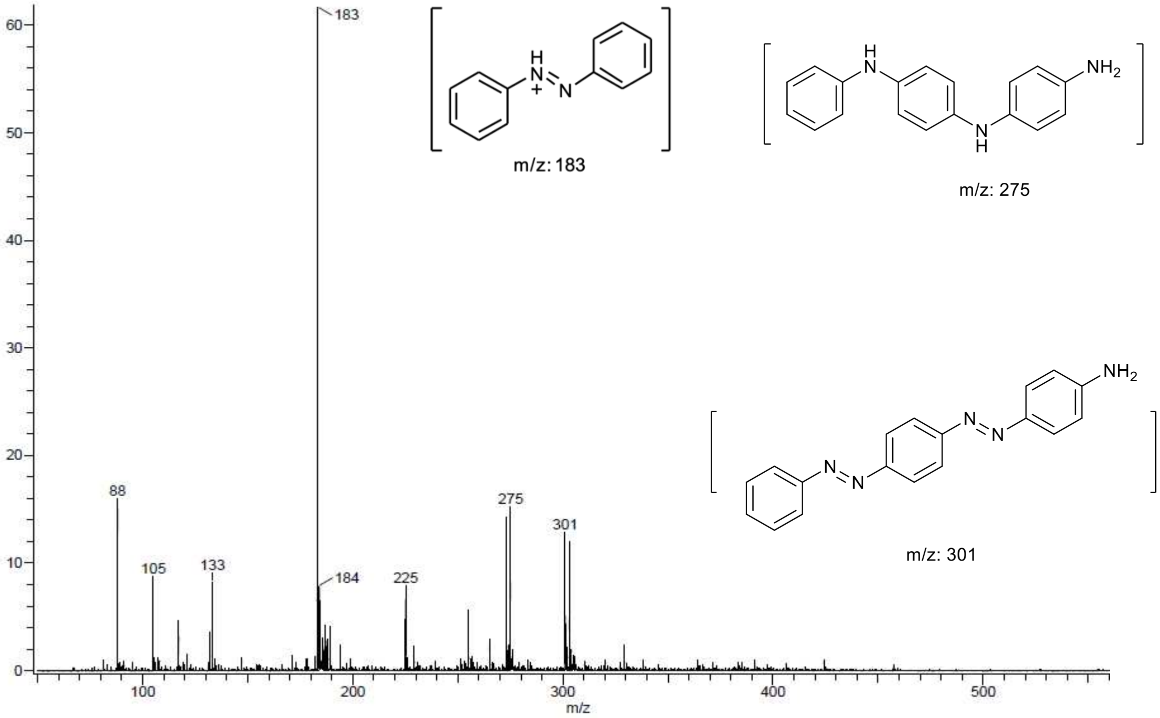


Figure S23. DART MS of azobenzene.

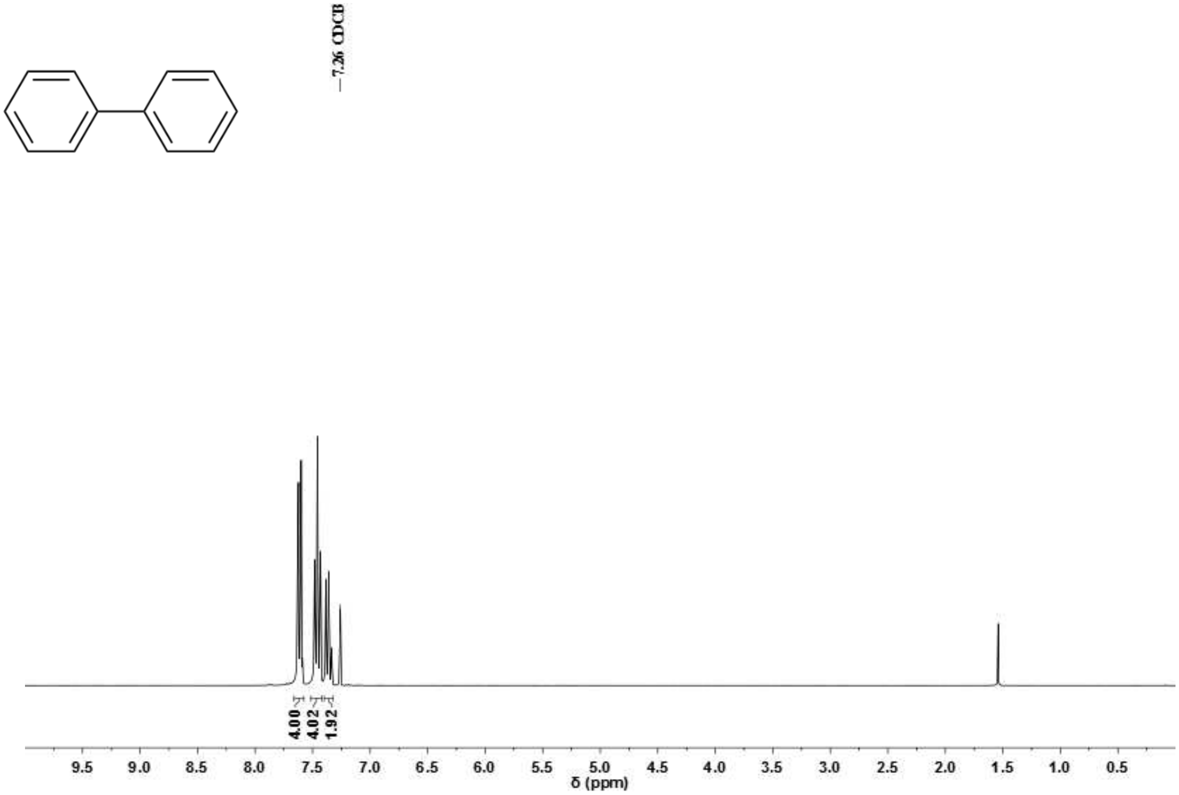


Figure S24. ^{1}H NMR spectrum of biphenyl.

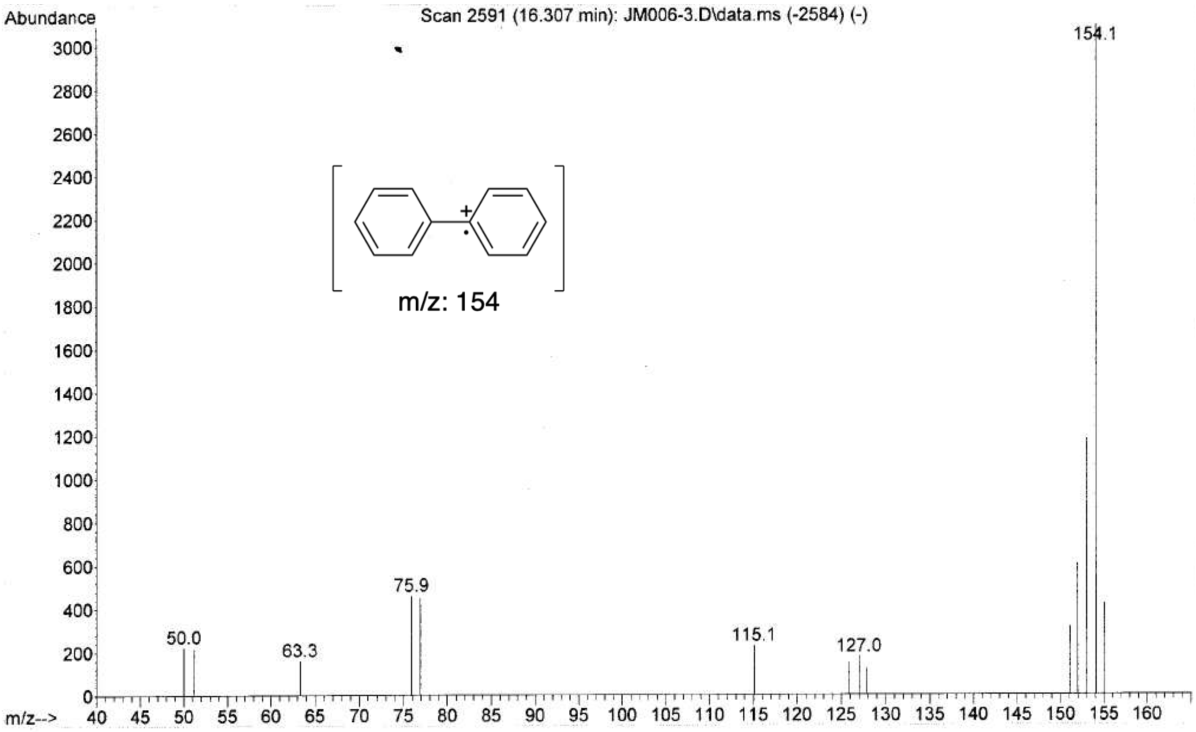


Figure S25. DART MS of biphenyl.

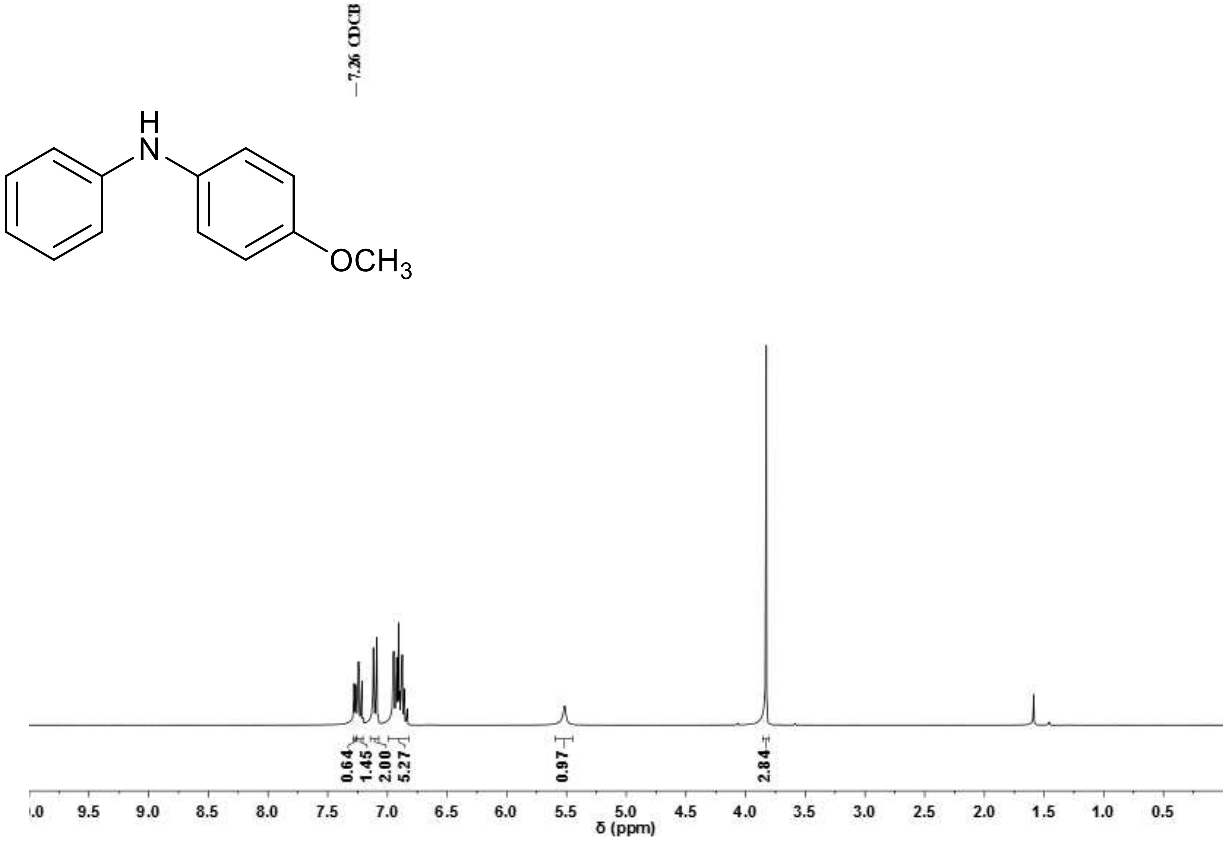


Figure S26. ^{1}H NMR spectrum of 4-methoxydiphenylamine.

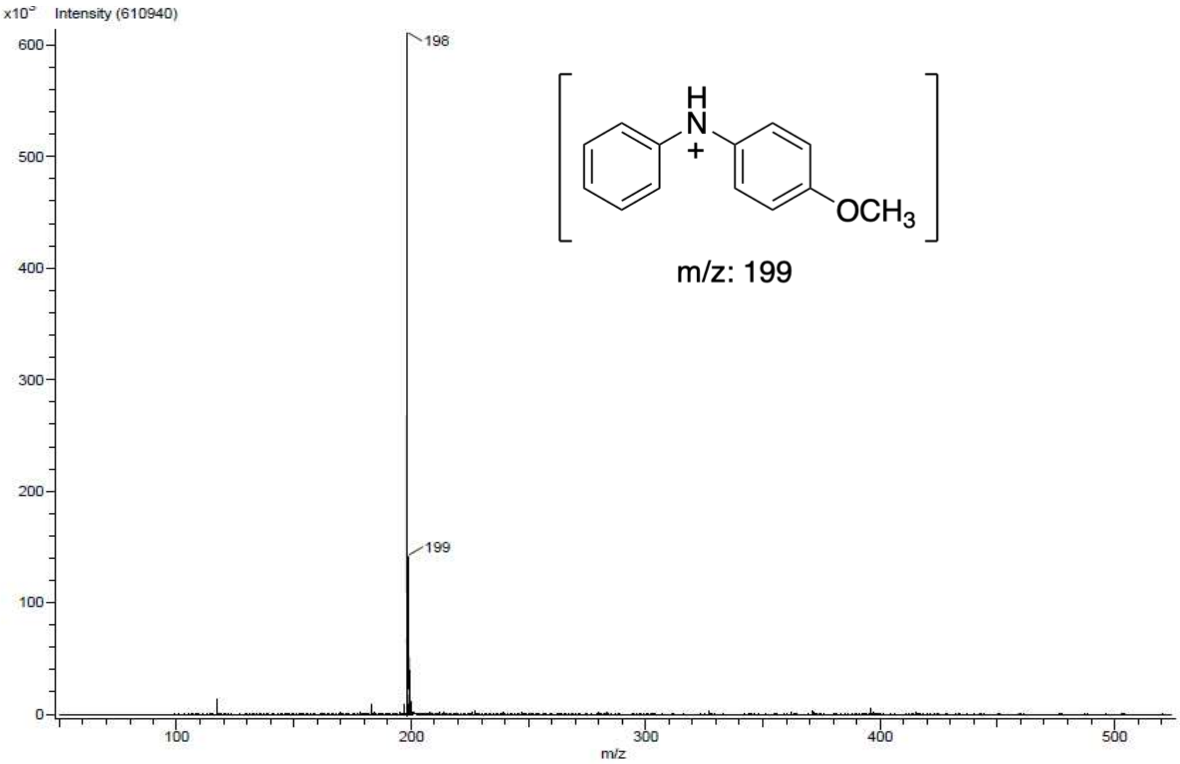


Figure S27. DART MS of 4-methoxydiphenylamine.

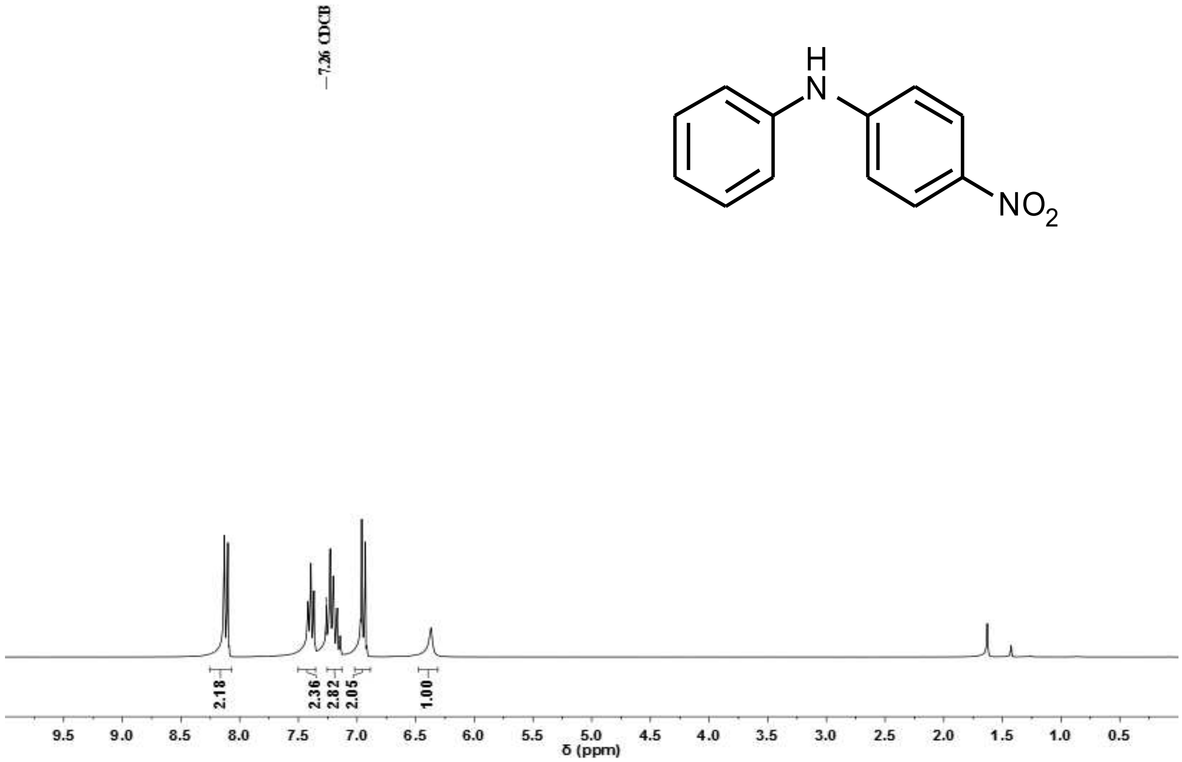


Figure S28. ^{1}H NMR spectrum of 4-nitrodiphenylamine.

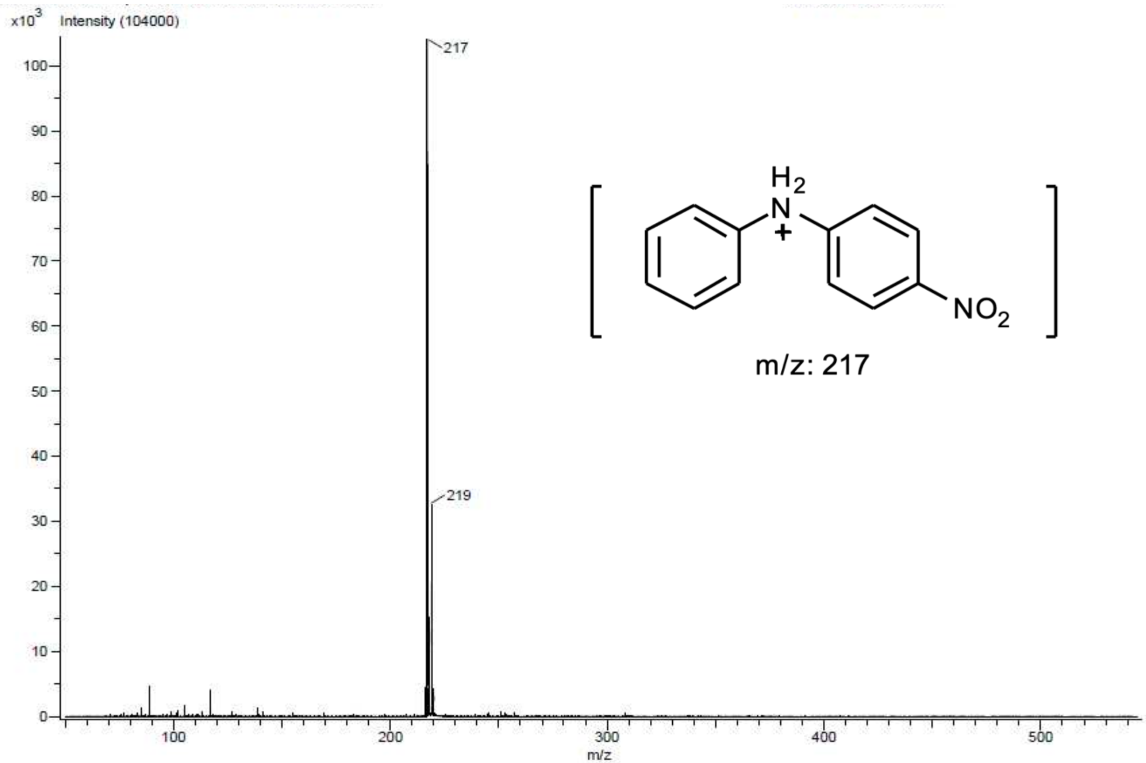


Figure S29. DART MS of 4-nitrodiphenylamine.

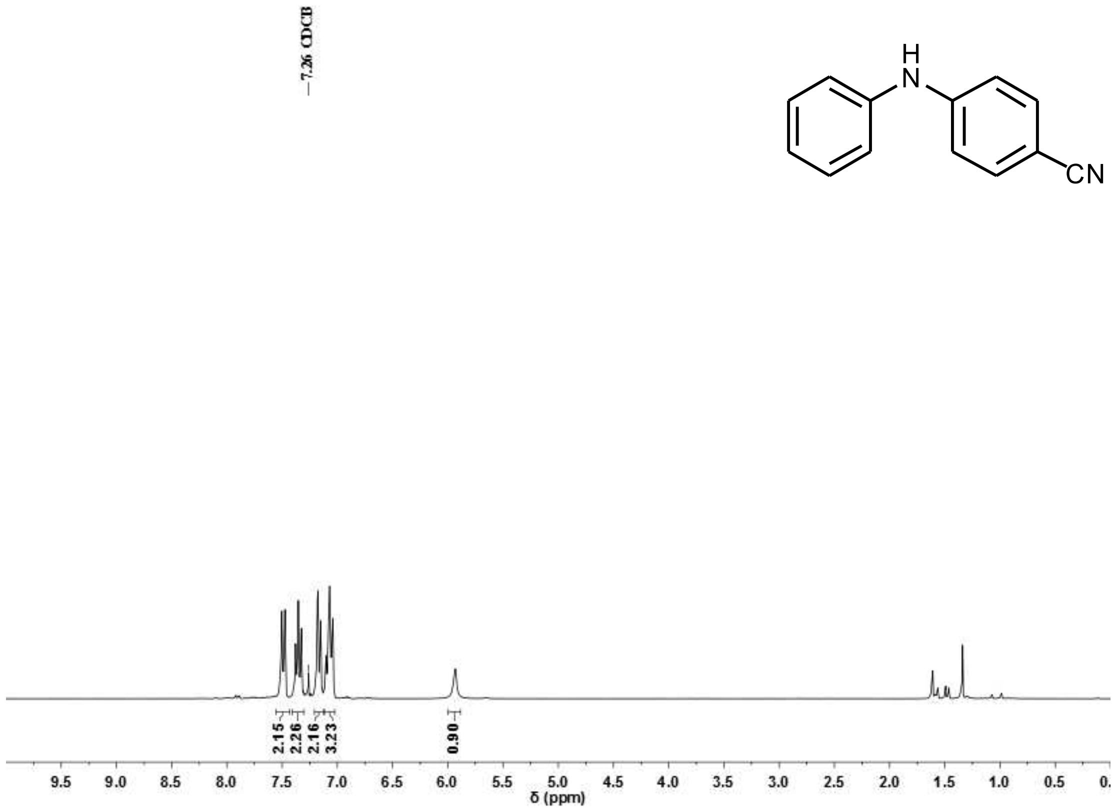


Figure S30. ^{1}H NMR spectrum of 4-cyanodiphenylamine.

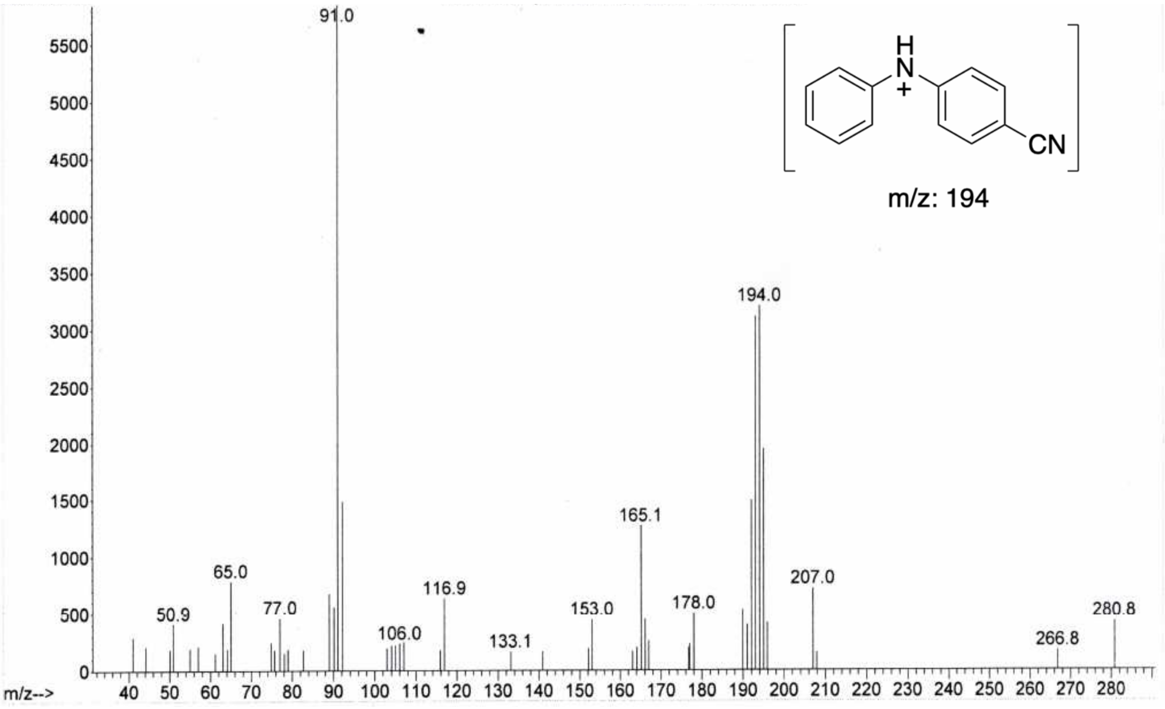


Figure S31. DART MS of 4-cyanodiphenylamine.

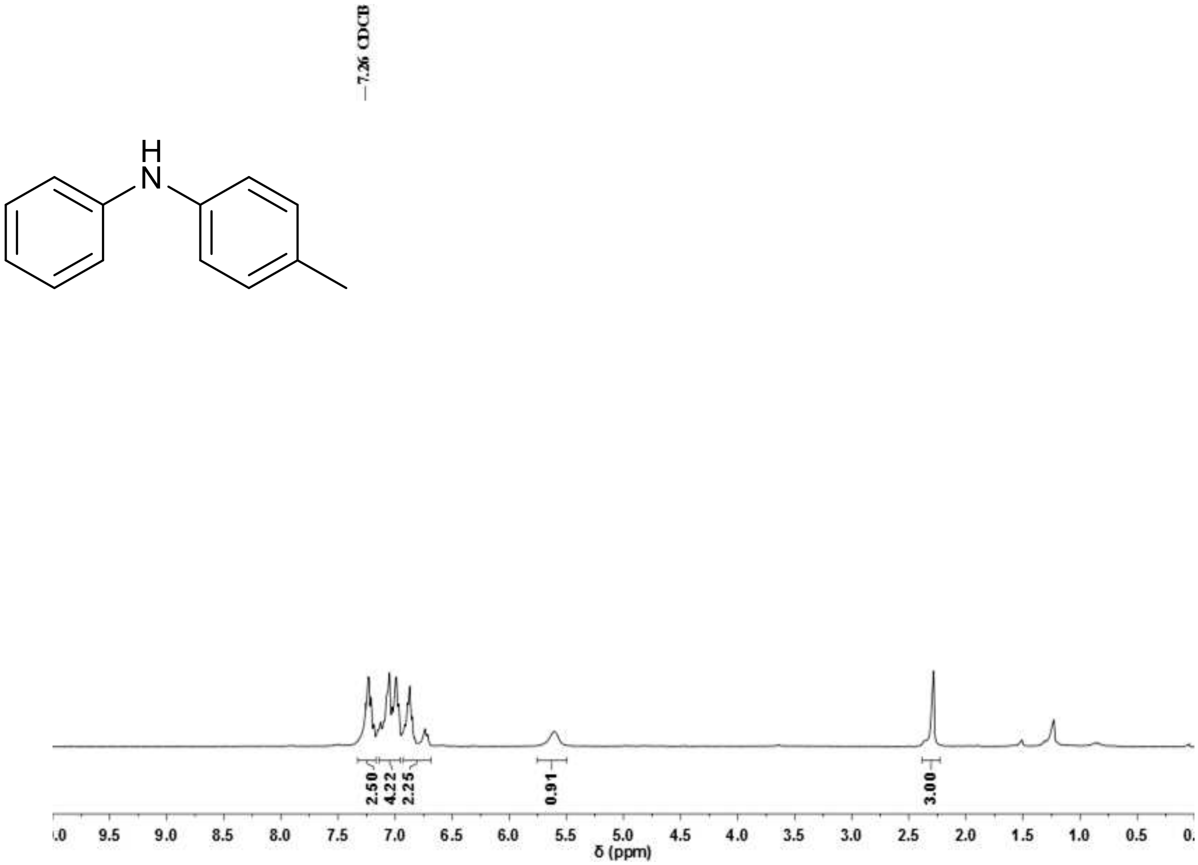


Figure S32. ^{1}H NMR spectrum of 4-methyldiphenylamine.

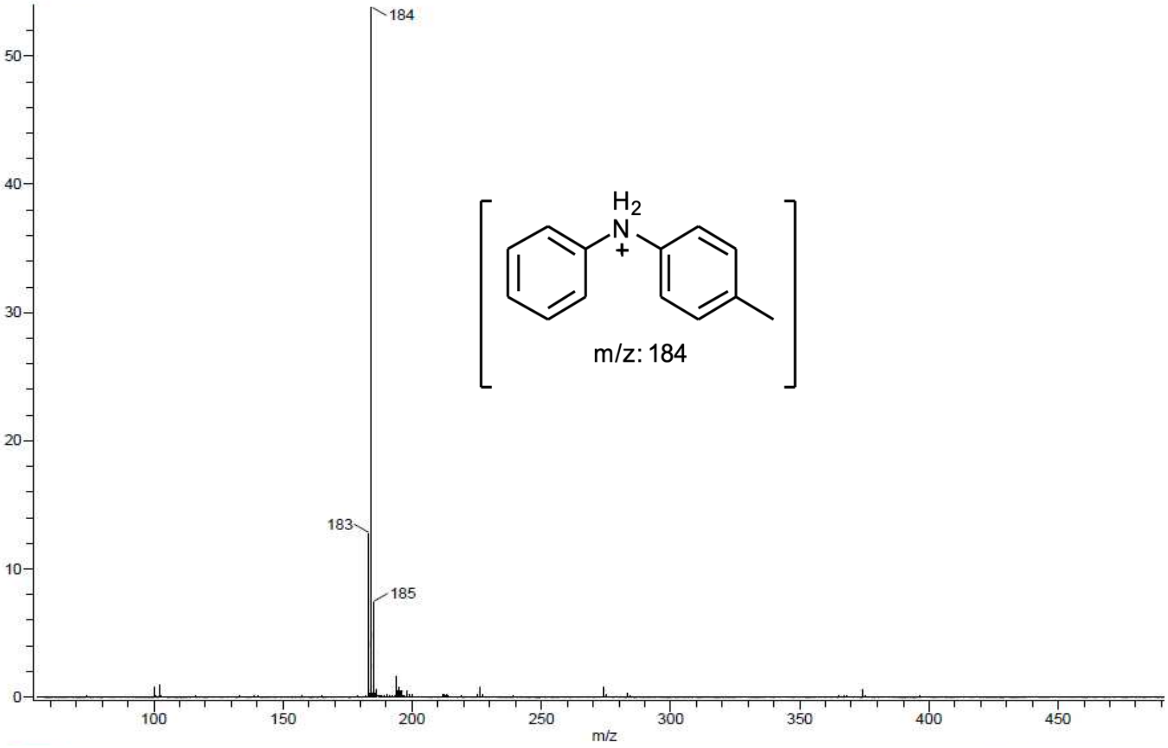


Figure S33. DART MS of 4-methyldiphenylamine.

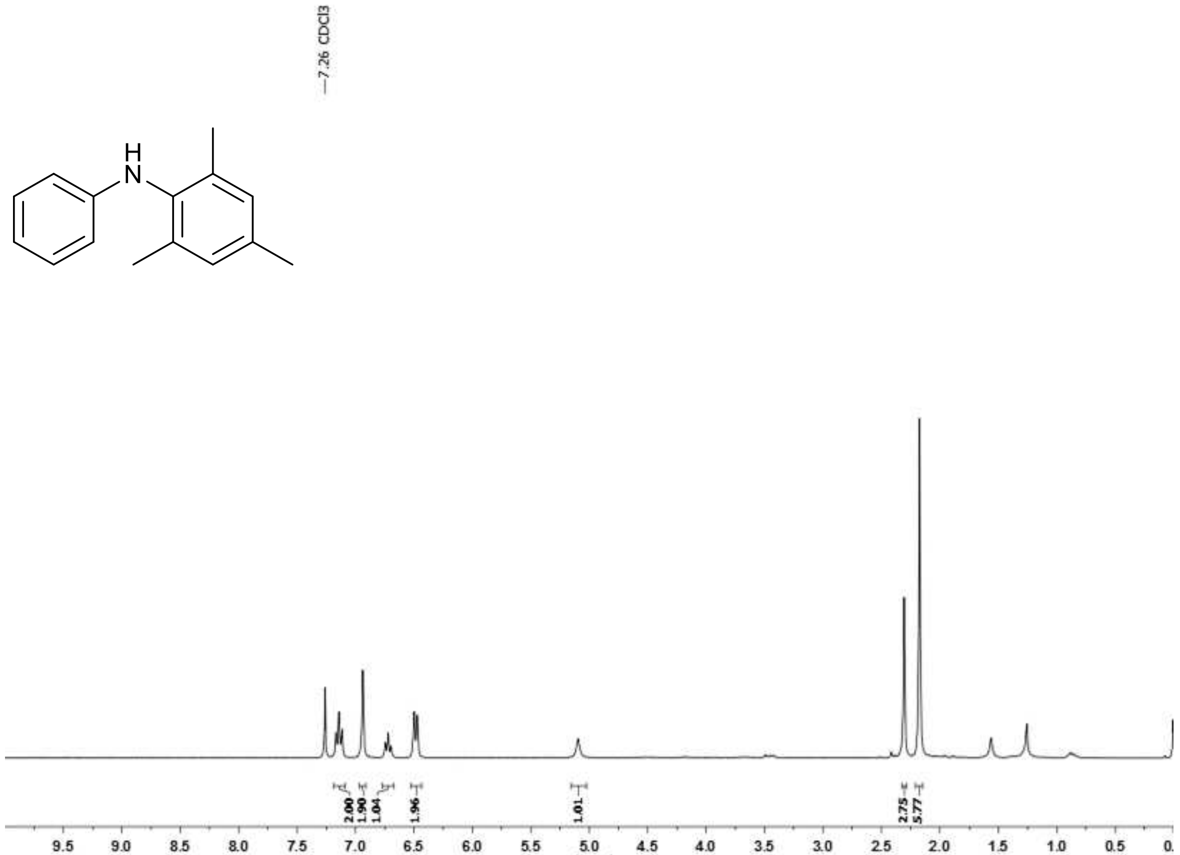


Figure S34. ^{1}H NMR spectrum of 2,4,6-trimethyldiphenylamine.

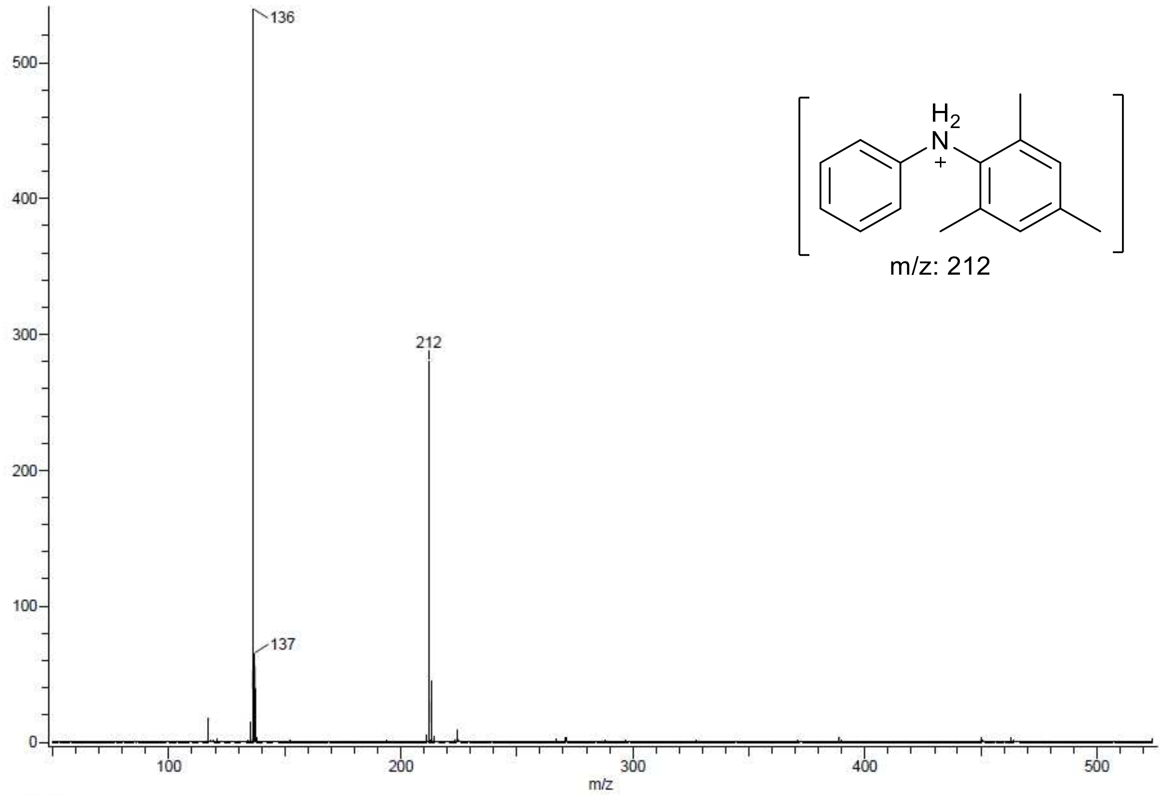


Figure S35. DART MS of 2,4,6-trimethyldiphenylamine.

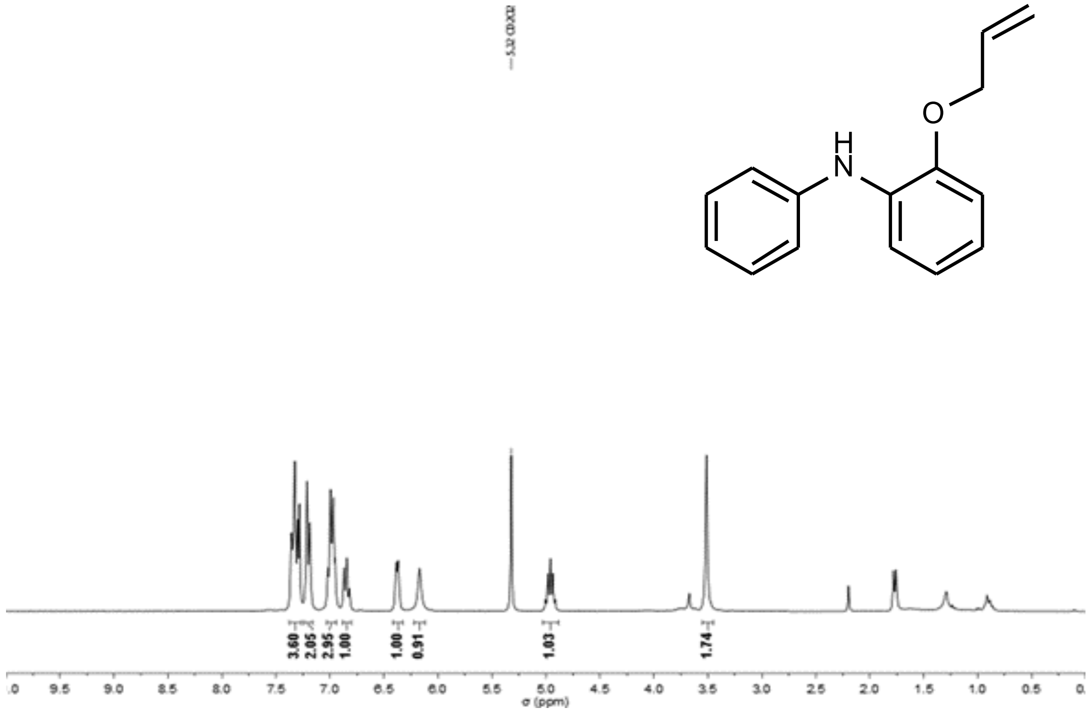


Figure S36. ^{1}H NMR spectrum of 2-(allyloxy)-N-phenylaniline.

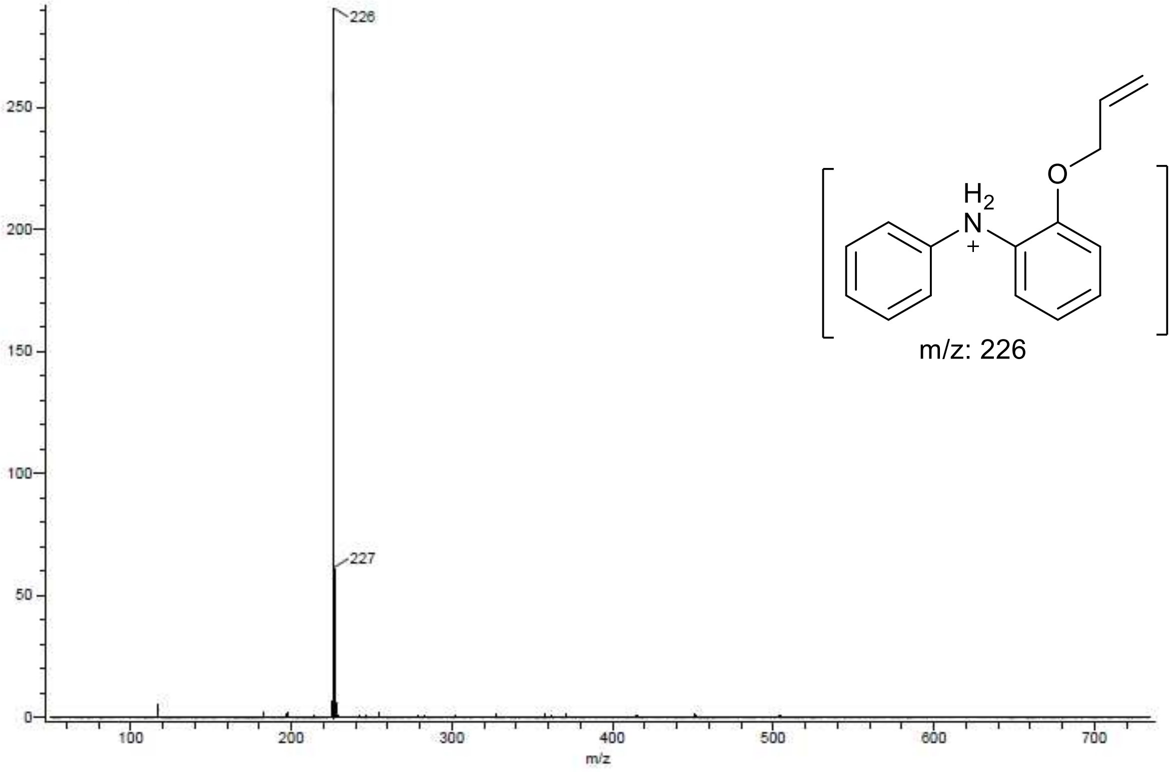


Figure S37. DART MS of 2-(allyloxy)-N-phenylaniline.

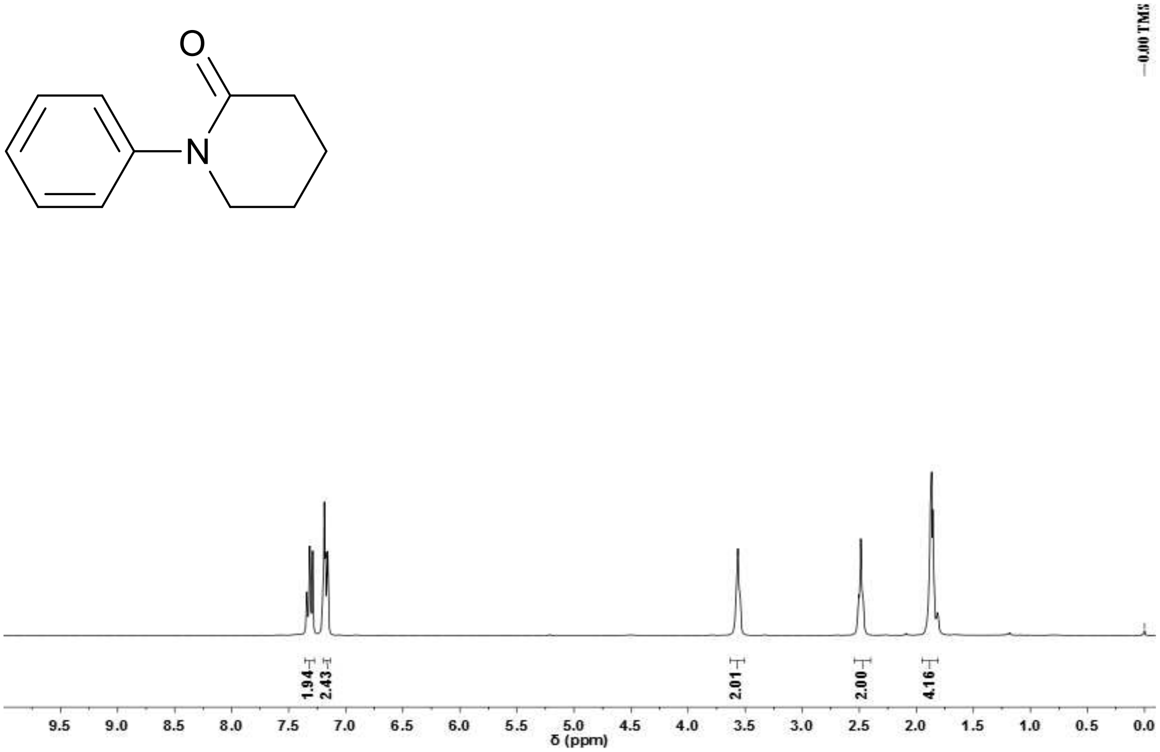


Figure S38. ^{1}H NMR spectrum of 1-phenyl-2-piperidinone.

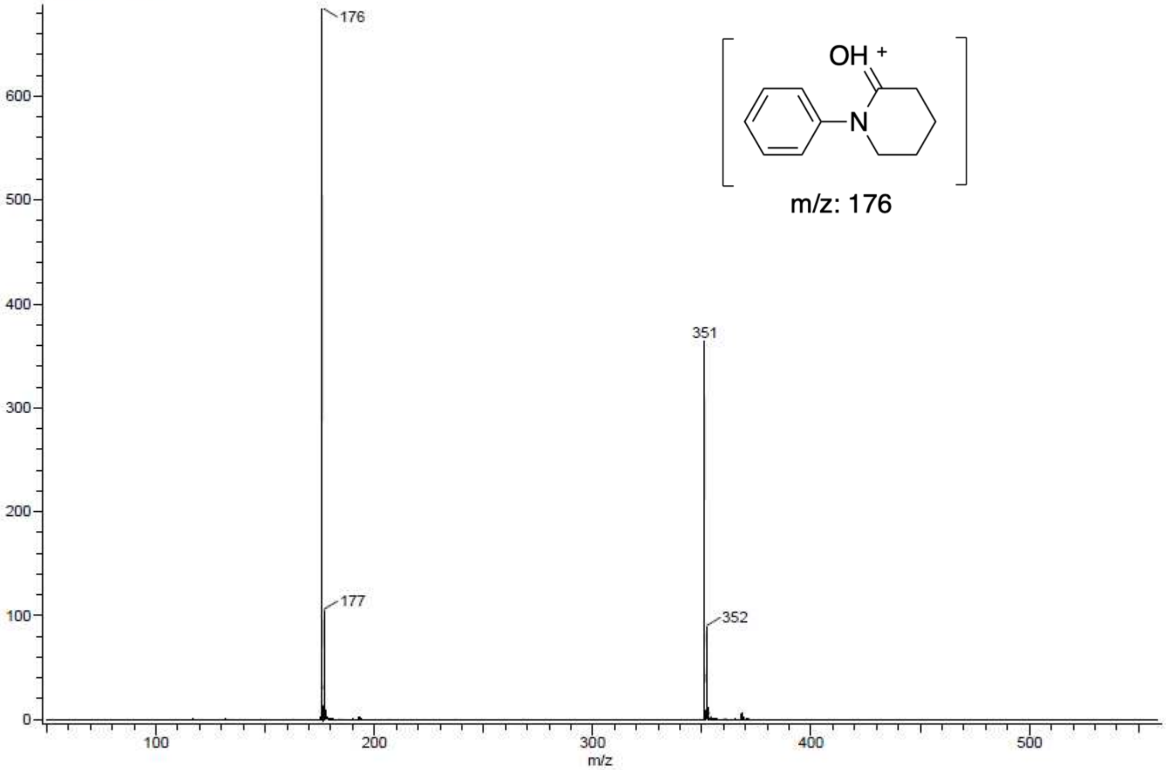


Figure S39. DART MS of 1-phenyl-2-piperidinone.

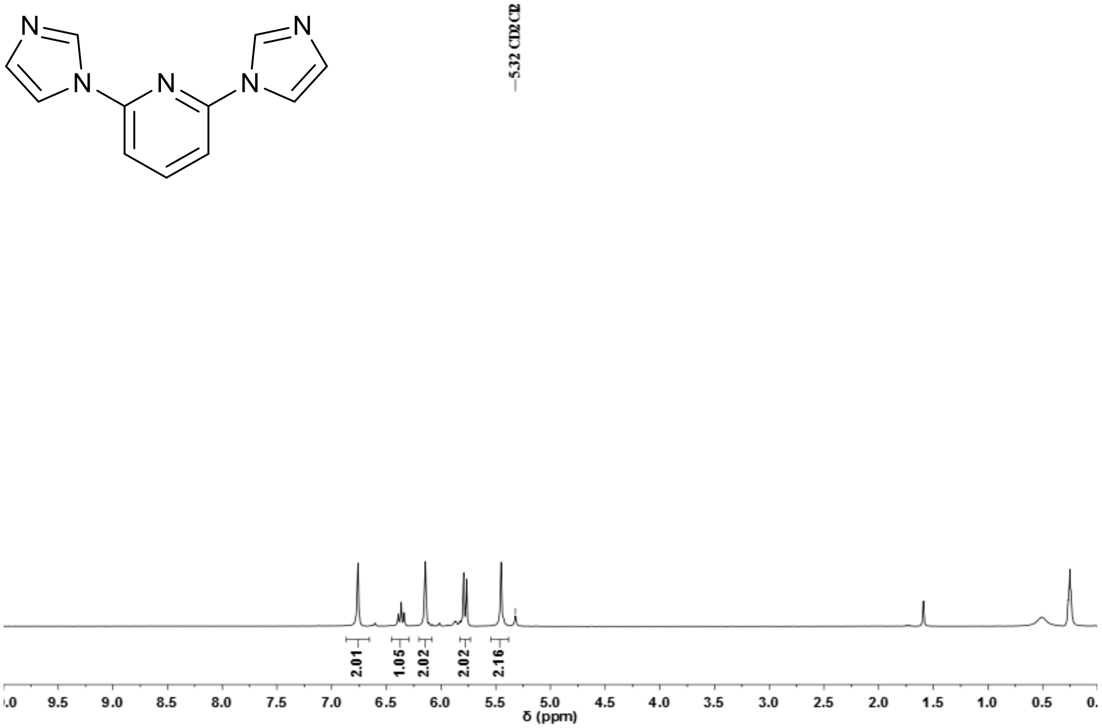


Figure S40. ^{1}H NMR spectrum of 2,6-di(1H-imidazol-1-yl)pyridine.

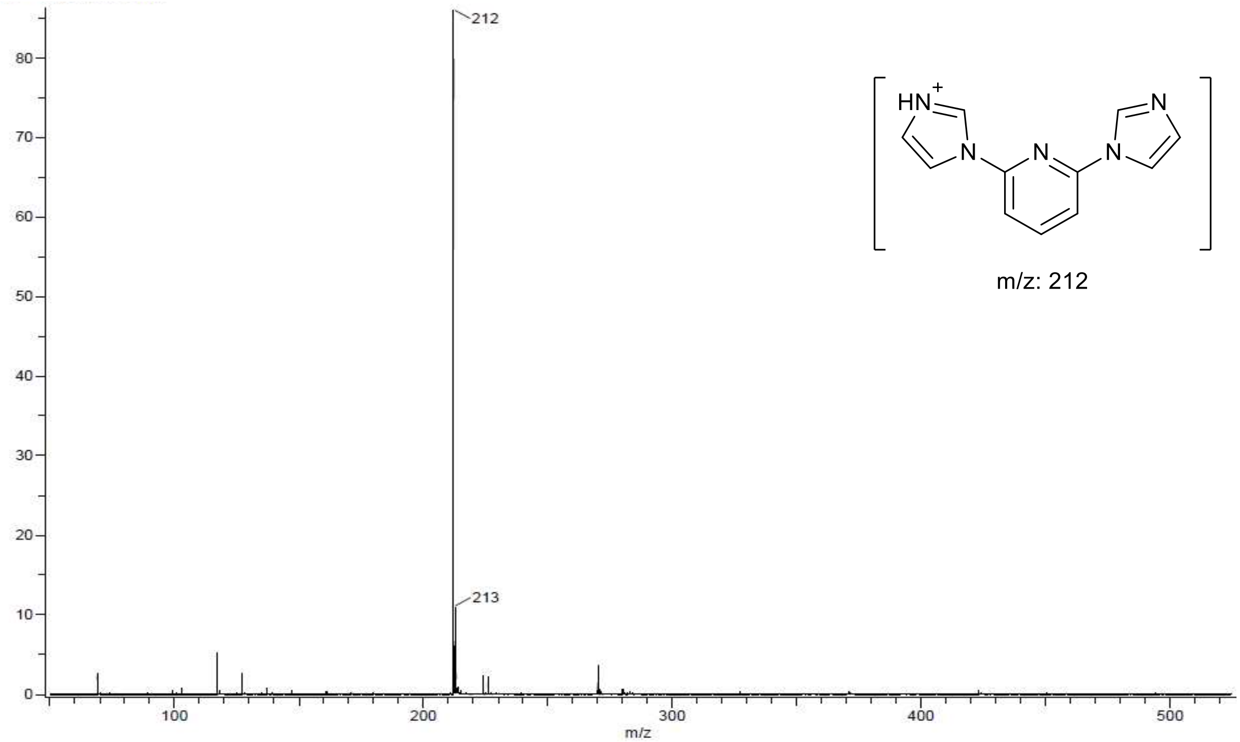


Figure S41. DART MS of 2,6-di(1H-imidazol-1-yl)pyridine.

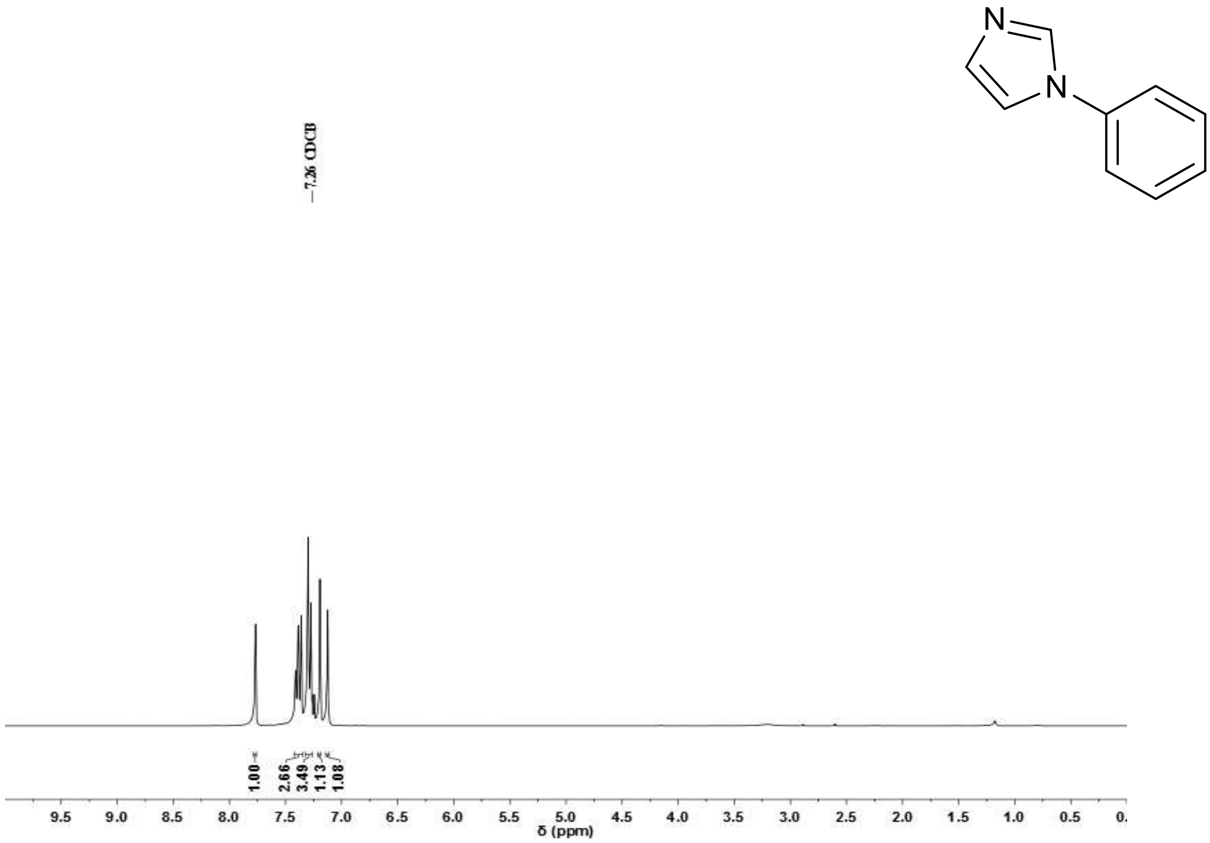


Figure S42. ^{1}H NMR spectrum of 1-phenylimidazole.

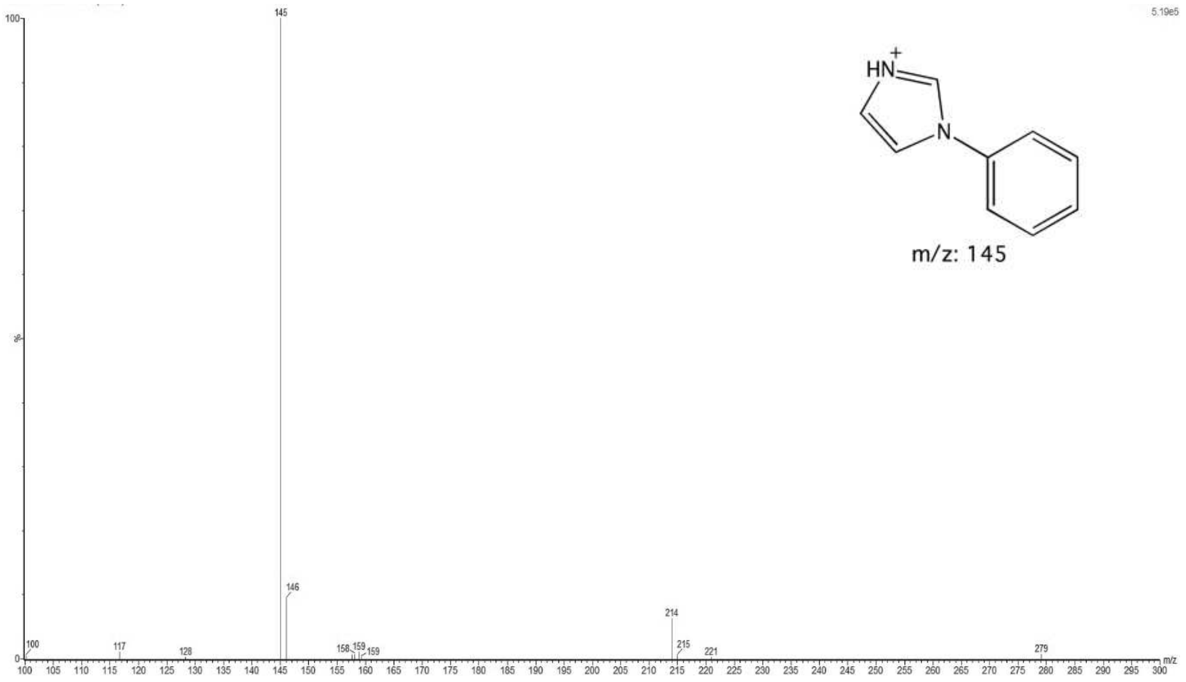


Figure S43. DART MS of 1-phenylimidazole.

S5. Computational Investigations

S5.1 Computational Methodology

S5.1.1 Free Energy Calculation

Harmonic frequency calculations were performed on the optimized structures to verify that they are indeed energy minima. The Goodvibes python script[5] was used to obtain the zero-point energy and thermal corrections for the selected structures, by calculating an approximate partition function using the rigid-rotator-quasiharmonic oscillator. To improve accuracy, the frequencies were scaled using the scaling factor 1.0340.[6] Frequencies below 100 cm^{-1} were set to 100 cm^{-1} and the gas phase values were corrected to a solution-phase standard state of 1 mol L^{-1}. Reported energies are either relative electronic energies ΔE (PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/-COSMO) or relative free energies ΔG at 298.15 K, all given in kJ mol^{-1}.

S5.1.2 Molecular Dynamics Simulations

Structures were equilibrated using a modified procedure by Wallnöfer *at al.*[7] that involves extensive heating and cooling to achieve equilibration. The production runs (in explicit chloroform) were carried out in NpT ensembles at 300 K, using Amber18.[8] Temperature was regulated with the Langevin thermostat,[9] whereas the pressure was kept at 1 bar using the Berendsen barostat.[10] The SHAKE algorithm[11] was used to restrain hydrogen bonds allowing for a time step of 2 fs; coordinates were saved every 10 ps, simulating a total of 1 µs. Accelerated MD (aMD) simulations[12] were carried out using the dual-boost algorithm implemented in Amber18,[8] where a bias was applied on the total potential and an additional boost on the dihedral term.[8] A total of 12 aMD simulations, each with 1 µs, were performed with various boosting parameters. These settings were derived as proposed by Pierce *et al.*[13] by performing 100 ns classical MD simulations. All trajectories of the production runs were combined, and the structures were aligned on the phenanthroyl bridge. A hierarchical clustering was applied on the combined trajectories to obtain a structurally divers ensemble. We did not re-weight the trajectories because (i) accurate re-weighting is difficult to achieve within aMD and energy minima remain such regardless of their relative energies, (ii) our goal was to obtain structures that are as diverse as possible, since relative energies of the conformers will change upon Cu(I) coordination.

S5.2 Computational protocol to determine the dissociation transition state $\mathbf{TS_{3\text{-}4}}$

The transition state $\mathbf{TS_{3\text{-}4}}$ presents a particular difficulty in obtaining a valid structure, as the electronic energy is rising in a monotonic fashion, as the dissociation coordinate is elongated. While this makes the determination of the exact structure difficult, the energy barrier of this TS can still be estimated. By manually creating structures along the reaction coordinate, 0.25 Å apart, and performing a restrained optimization as well as frequency calculations, the initial steps of the reaction can be characterized (Figure S44). When another datapoint is added, with the product structures infinitely separated, a jump in free energy can be observed. By fitting a sigmoid function on the values of $-T\Delta S$ (black dotted line), the free energy along the dissociation (red line) can be estimated from adding the calculated enthalpy (blue dotted line) to $-T\Delta S$ (black dotted line). The highest point is taken as the estimated for $\mathbf{TS_{3\text{-}4}}$. The approximated transition-state structure is depicted in Figure S50.

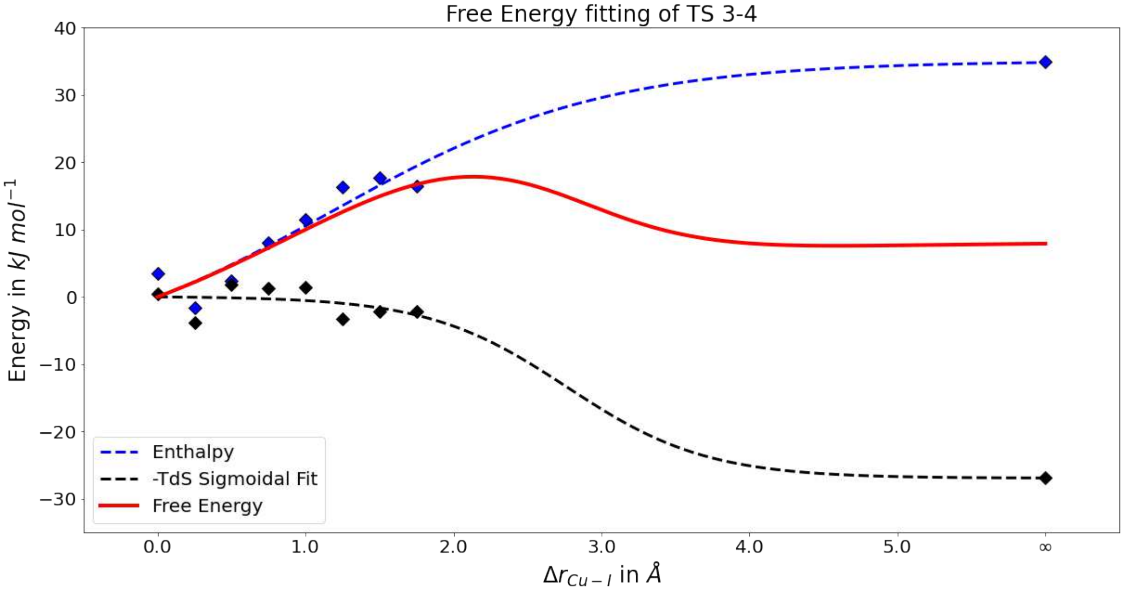


Figure S44. Free energy fitting of **TS$_{3\text{-}4}$**. Structures were generated along the reaction coordinate in steps of 0.25 Å (blue dots) and interpolated by a sigmoid blue dashed line. Optimizations and frequency calculations were carried out at the PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO level (black dots) and sigmoid functions were fitted for the enthalpy and -TΔS (black dashed line). The free energy was calculated by summing up the sigmoids and the maximum value was taken as the TS free energy estimate (red line).

S5.3 Structures of reactive species

The transition states **TS$_{2\text{-}3}$** (proton abstraction) and **TS$_{4\text{-}5}$** (iodide abstraction) are diffusion controlled, hence, a barrier of 20 kJ/mol is assumed and no structures are reported.

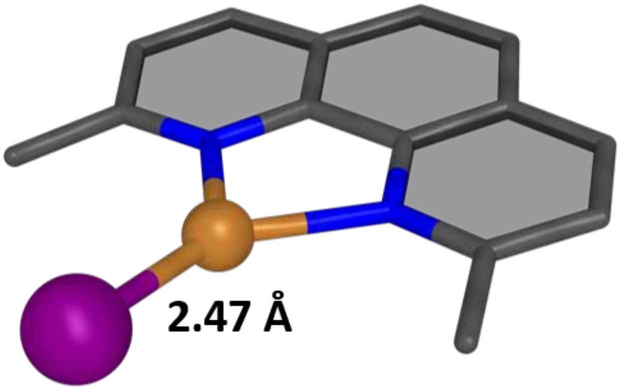


Figure S45. Structure **1**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

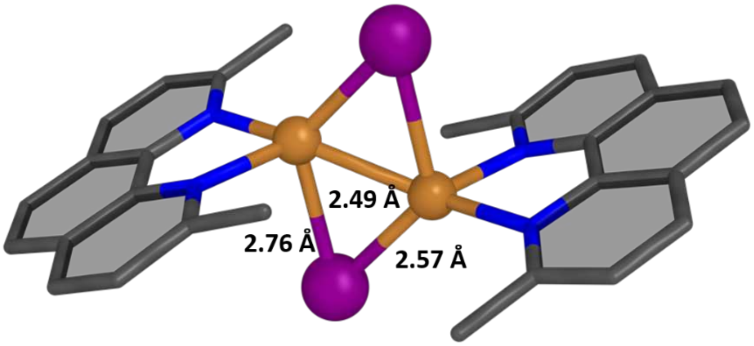


Figure S46. Structure [**Cu(Phen)(μ-I)**]$_2$ optimized with PBE0/def2-SVP/D3/COSMO(toluene).

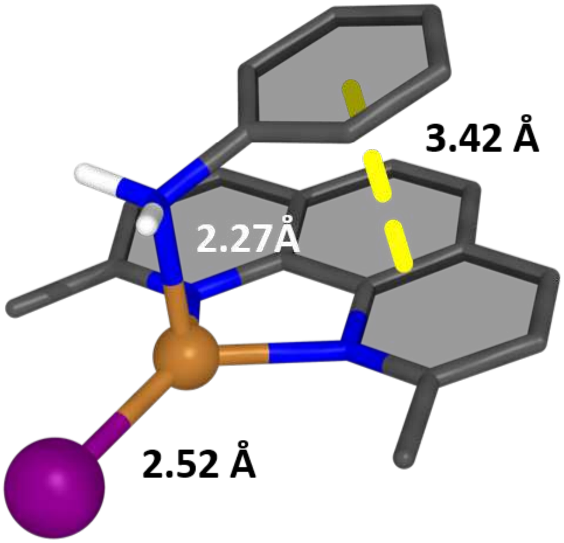


Figure S47. Structure **2**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

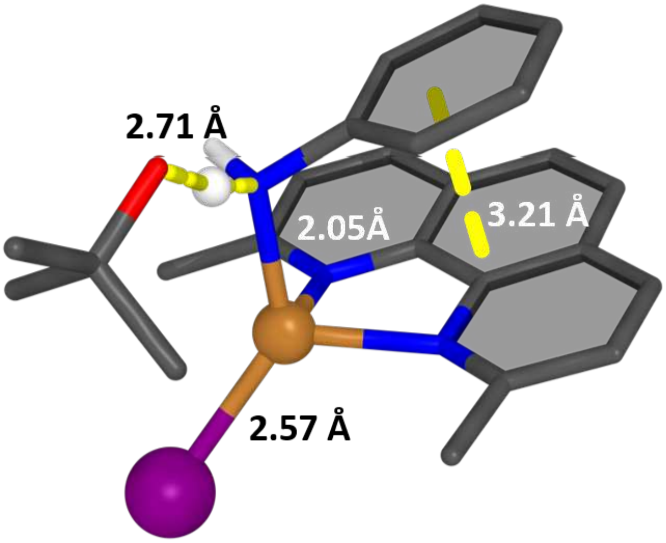


Figure S48. Structure of approximate $\textbf{TS}_{2\text{-}3}$, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

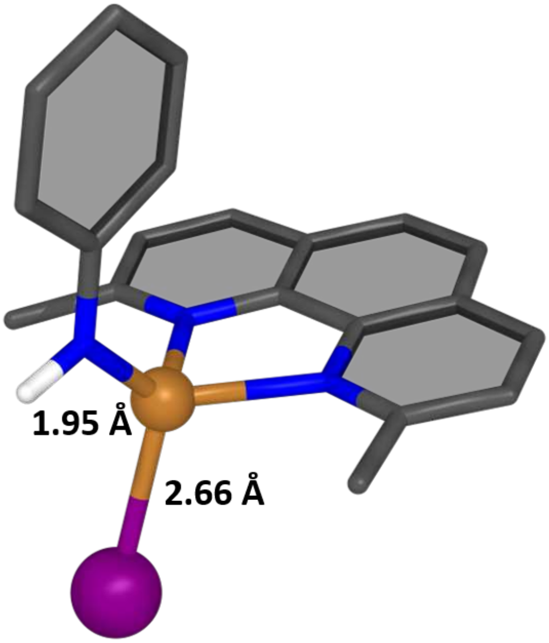


Figure S49. Structure **3**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

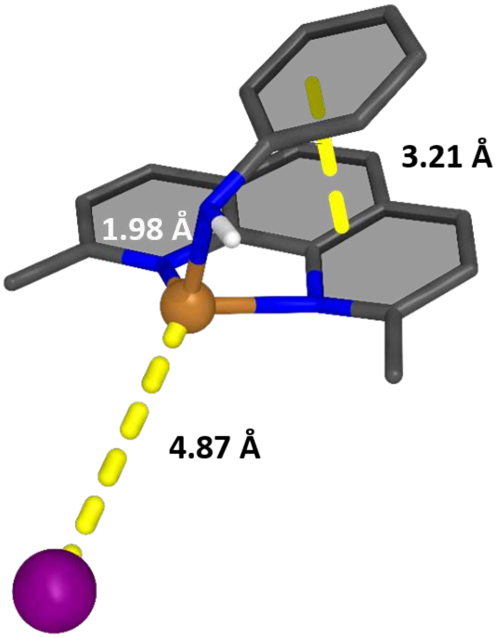


Figure S50. Structure **TS$_{3\text{-}4}$**, Estimate optimized with PBE0/def2-SVP/D3/COSMO(toluene).

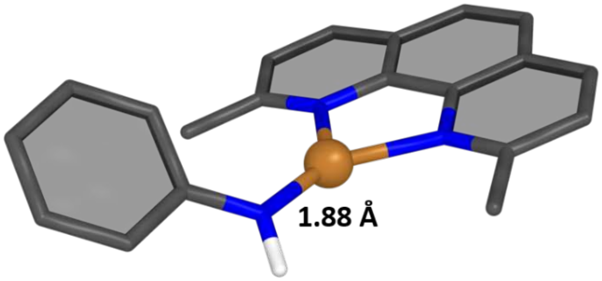


Figure S51. Structure **4**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

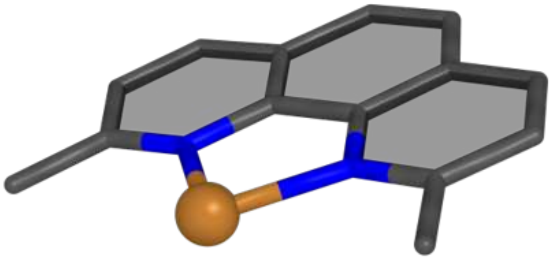


Figure S52. Structure **2a** optimized with PBE0/def2-SVP/D3/COSMO(toluene), counter ion is omitted.

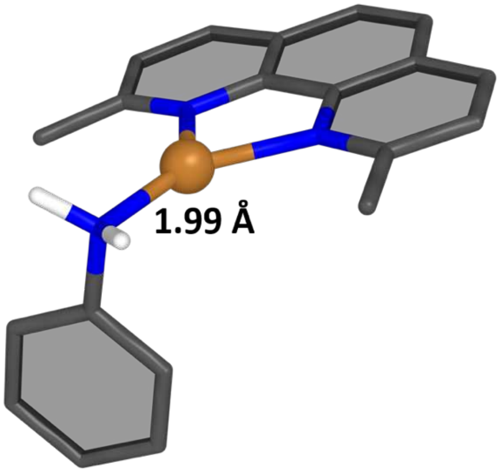


Figure S53. Structure **3a** optimized with PBE0/def2-SVP/D3/COSMO(toluene), counter ion is omitted

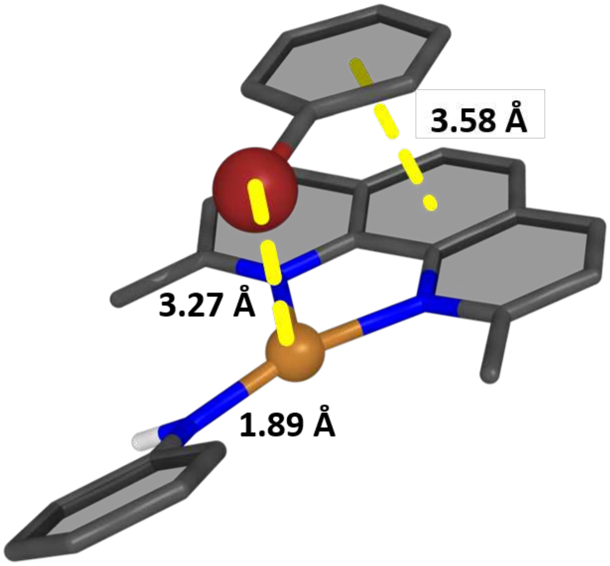


Figure S54. Structure **5**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

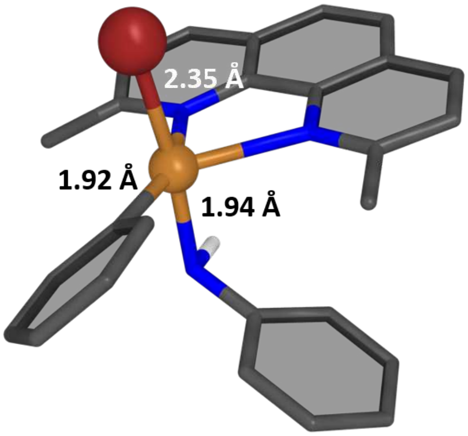


Figure S55. Structure **6**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

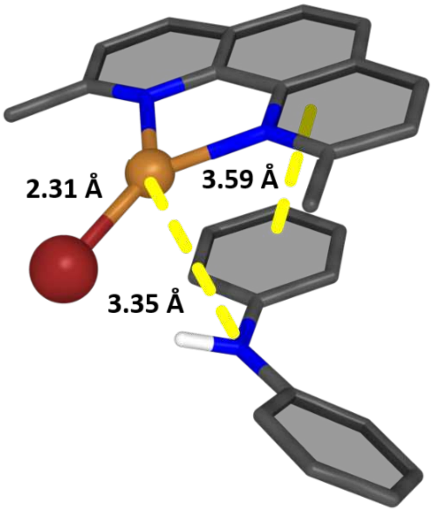


Figure S56. Structure **7**, optimized with PBE0/def2-SVP/D3/COSMO(toluene).

S5.4 Subsequent cycles of the reaction

The first cycle of the reaction begins with iodine and ends with bromine attached to the Cu atom. All further cycles of the reaction take place starting with the Br analogue of structure. An investigation of the electronic energy path of the Br equivalents of structures **1** to **4** is shown in Figure S57. The reaction energies are largely similar, with the main differences being in the case of **3** and **4** with the bromine intermediates being 7 and 20 kJ mol^{-1} higher in energy.

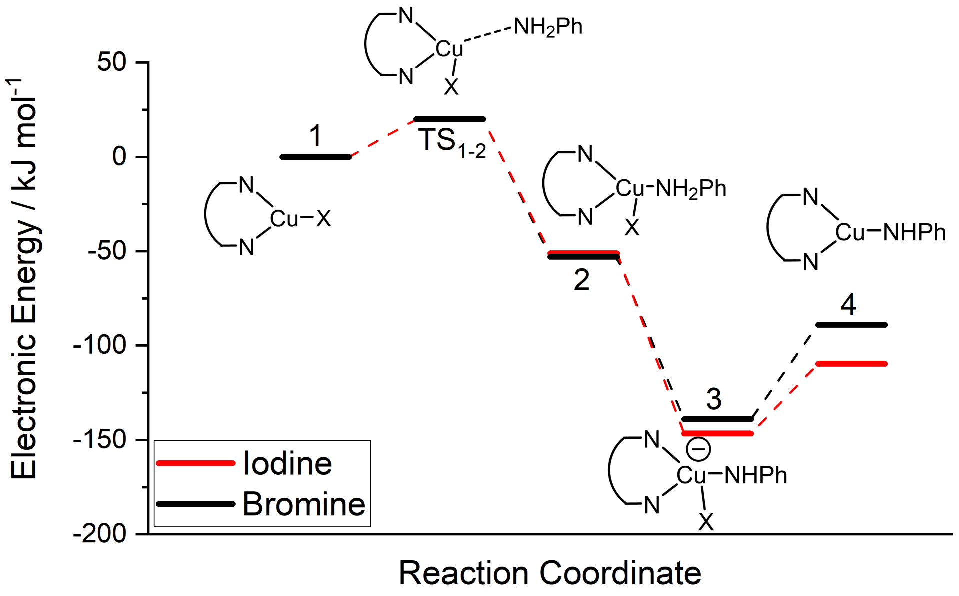


Figure S57. Energy diagram of structures **1**-**4** for the first (iodine) and subsequent (bromine) cycles of the reaction. X stands for the respective halogen atom. Energies were computed with PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO.

S5.5 Investigation of the supramolecular complexe

The supramolecular catalyst is a difficult structure to fully investigate, as QM calculations are computationally intensive and do not take into account the flexibility of the calixarene ring. MD simulations can not only describe the flexibility of the cage, but they allow for the inclusion of explicit solvent, which is essential for preventing the calixarene cavity from collapsing. This comes at the cost of an inability to take into account any bond formation or dissociation. To this end, we developed a protocol, as shown in Figure S58, where we combine MD simulations with QM optimizations, in such a way that we are able to obtain snapshots along the reaction pathway. By utilizing DFT for elucidating the reaction path on a model system, we can pinpoint the important intermediates involved in the reaction. Meanwhile, on the MD side, we can perform simulations of the catalyst to get an idea about the cage movements. By clustering the simulations, we are able to acquire a few representative structures, to which we can fit the various intermediates. These structures can then subsequently serve as starting points for more simulations of the various intermediates, as well as constructing a reaction profile for the whole supramolecular catalyst – being able to capture the effect on confinement on the reaction path. Specificities of each calculation step can be found in the respective subsections of the computational methodology.

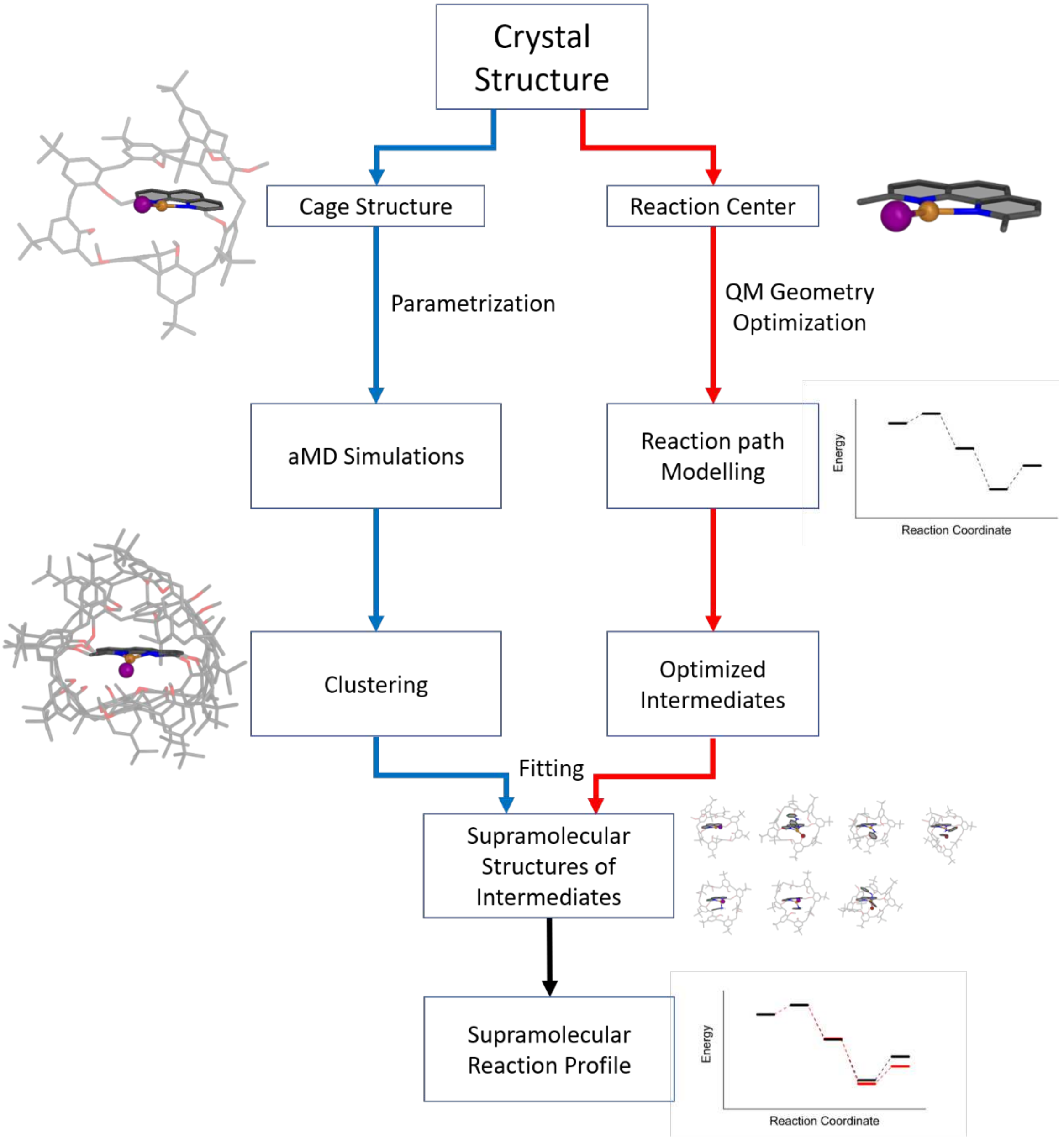


Figure S58. Workflow combining QM and MD simulations, developed to investigate the supramolecular catalyst.

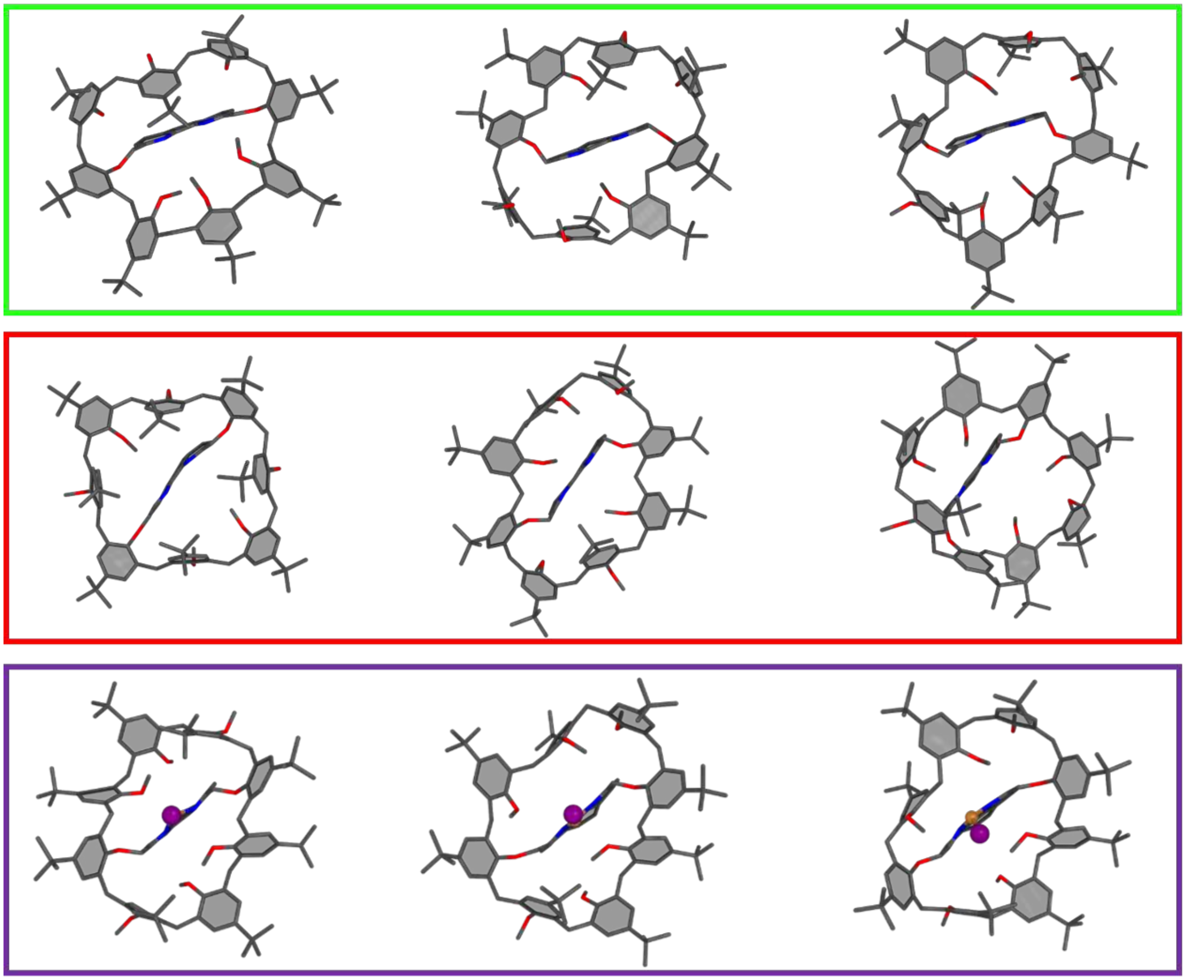
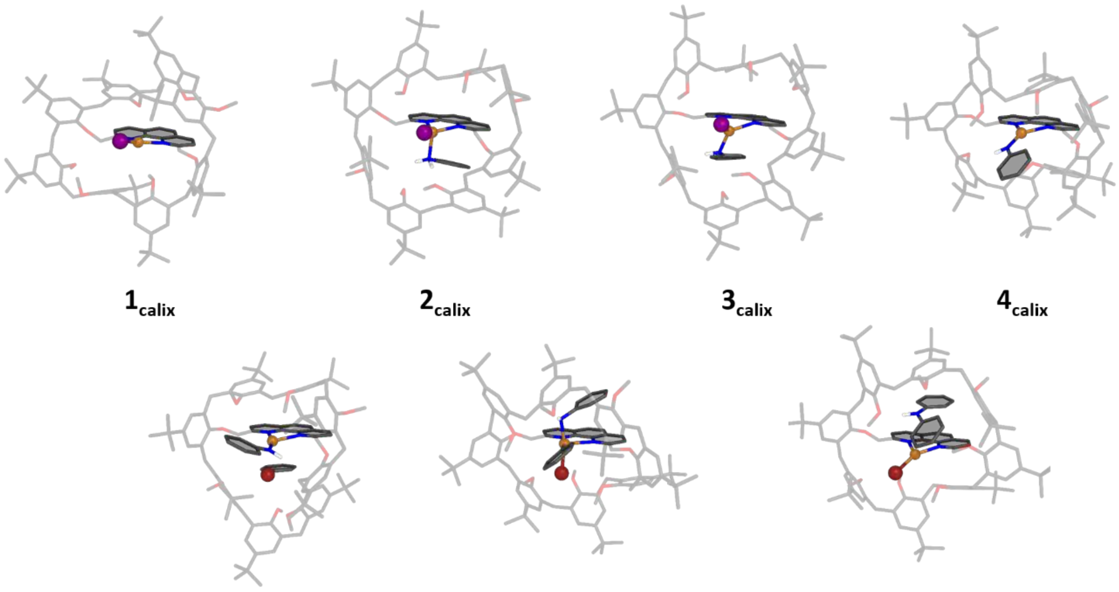


Figure S59. Top: Three most populated clusters of the aMD simulations of the calixarene ligand without Cu coordination (**C_8PhenMe$_6$**) in toluene. Middle: most populated clusters of **C_8PhenMe$_6$** in chloroform. Bottom: Most populated the clusters for the simulation of **1_{calix}** in explicit chloroform.

The fitting of the small phenanthroyl system was performed on multiple clusters from different simulations. Clustering the aMD simulations provided with an overview of the dynamics of the calixarene macrocycle. When comparing the main clusters in the simulations with implicit solvent (not shown) and chloroform, it was observed that explicit solvation was essential to retain the cavity. When comparing the top three most populated clusters of the aMD simulations of the uncoordinated ligand **C_8PhenMe$_6$** in toluene and in chloroform (top and middle panel in Figure S59), it is evident that the structures are very similar. Indeed, even CuI coordination yielding $\mathbf{1}_{\mathbf{calix}}$ (bottom panel in Figure S59) did not change much, but very similar clusters were obtained for the simulation of $\mathbf{1}_{\mathbf{calix}}$ in chloroform compared to simulation of the ligand alone with cavities of similar size. As results for chloroform and toluene were similar, all subsequent calculations were performed in chloroform only to reduce computational costs.

As the simulations yielded well-defined cavities, fitting of structures **2** to **7** deemed possible without too many clashes. To this end, structures **2** through **7** were fitted into the cavity, using the phenantroyl bridge as a base for alignment, yielding $\mathbf{2}_{\mathbf{calix}}$ through $\mathbf{7}_{\mathbf{calix}}$. Using the number of steric clashes as the criterion, each fitted structure was ranked. The most populated cluster of the simulation of $\mathbf{1}_{\mathbf{calix}}$ showed the least amount of clashing and thus was chosen as the template for subsequent optimizations.



$\mathbf{1}_{\mathrm{calix}}$ $\mathbf{2}_{\mathrm{calix}}$ $\mathbf{3}_{\mathrm{calix}}$ $\mathbf{4}_{\mathrm{calix}}$

$\mathbf{5}_{\mathrm{calix}}$ $\mathbf{6}_{\mathrm{calix}}$ $\mathbf{7}_{\mathrm{calix}}$

Figure S60. Structures of the supramolecular calixarene intermediates, optimized with PBE0/def2-TZVP/D3/COSMO//PBE0/def2-SVP/D3/COSMO.

The resulting structures were subsequently optimized with PBE0/def2-SVP/D3+COSMO, followed by single point calculations with the larger def2-TZVP basis set and the energetic landscape was evaluated. Final structures are depicted in Figure S60.

As simulations of structure $\mathbf{1}_{\mathrm{calix}}$ as well as $\mathbf{6}_{\mathrm{calix}}$ were performed, the flexibility of the calixarene ring was also investigated. It was observed that the calixarene units exhibits less movement during the simulation of $\mathbf{6}_{\mathrm{calix}}$. This was quantified as the relative entropy of the diherdrals between the individial units calculated with the X-entropy script.[14] In Figure S61, it can be seen that the distributions of the dihedrals in the simulated structure $\mathbf{6}_{\mathrm{calix}}$ are considerably narrower than those of $\mathbf{1}_{\mathrm{calix}}$.

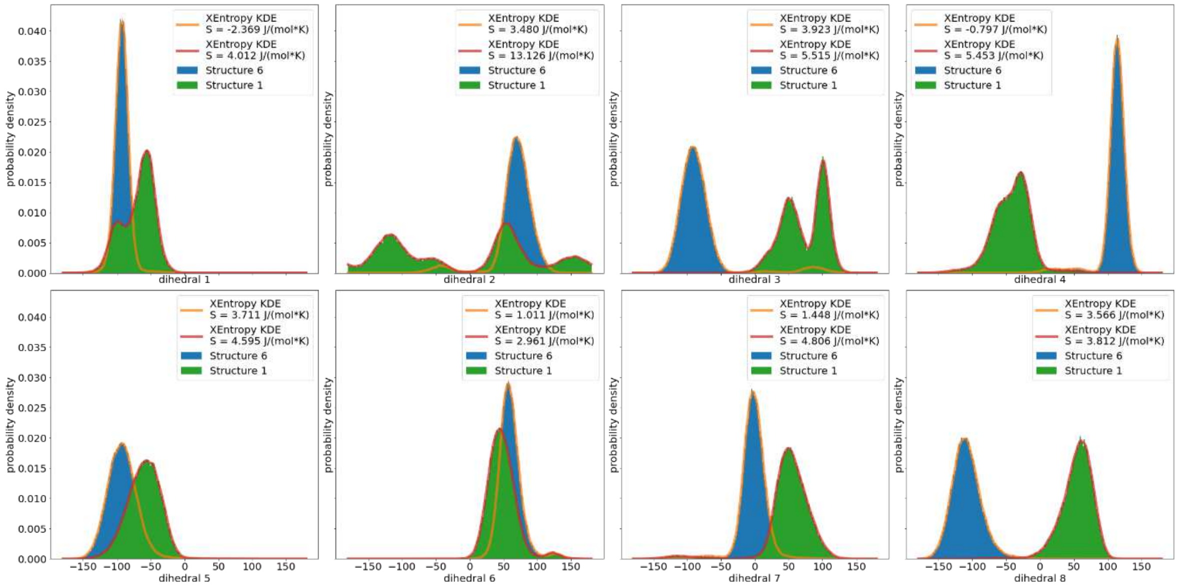


Figure S61. Entropy values of the dihedral distributions, as calculated with the X-entropy script. The green distribution, contoured by red, represents the values from the simulation of **1_{calix}**. The blue distribution, contoured by orange, represents the values obtained from the simulation of **6_{calix}**.

S6. References

(1) Armarego, W. L. F.; Chai, C. L. L. Purification of Organic Chemicals; **2009**.

(2) Hernández, D. J.; Vázquez-Lima, H.; Guadarrama, P.; Martínez-Otero, D.; Castillo, I. Tetrahedron Lett. **2013**, 54, 4930–4933.

(3) Munch, J. H.; C. D. Gutsche. p-tert-butylcalix[8]arene. Org. Synth. **1990**, 68, 243.

(4) Chandler, C. J.; Deady, L. W.; Reiss, J. A. J. Heterocyclic Chem. **1981**, 18, 599.

(5) Luchini G.; Alegre-Requena J. V.; Funes-Ardoiz I.; Paton R. S. GoodVibes: automated thermochemistry for heterogeneous computational chemistry data. F1000Research **2020**, 291.

(6) Kesharwani, M.; Brauer, B.; Martin, L. Frequency and Zero-Point Vibrational Energy Scale Factors for Double-Hybrid Density Functionals (and Other Selected Methods): Can Anharmonic Force Fields Be Avoided? J. Phys. Chem. A **2015**, 119, 1701–1714.

(7) Wallnoefer, H; Handschuh, S.; Liedl, K.; Fox, T. Stabilizing of a Globular Protein by a Highly Complex Water Network: A Molecular Dynamics Simulation Study on Factor Xa. J. Phys. Chem. B. **2010**, 114, 7405–7412.

(8) a) Salomon-Ferrer, R.; Case, D.; Walker, R. An overview of the Amber biomolecular simulation package. Wiley Interdiscip. Rev. Comput. Mol. Sci. **2013**, 3, 198–210. b) D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R .E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman. AMBER 2018, Reference Manual. University of California, San Francisco. **2018**.

(9) Adelman, S.; Doll, J. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering off Harmonic Solids. J. Chem. Phys. **1976**, 64, 2375–2388.

(10) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. Molecular-Dynamics With Coupling to an External Bath. J. Chem. Phys. **1984**, 81, 3684–3690.

(11) Ciccotti, G.; Ryckaert, J. Molecular-Dynamics Simulation of Rigid Molecules. Comput. Phys. Commun. **1986**, 4, 345-392.

(12) a) Hamelberg, D.; de Oliveira, C.; McCammon, J. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J. Chem. Phys. 2007, 127, 1-9. b) Hamelberg, D.; Mongan, J.; McCammon, J. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J. Chem. Phys. **2004**, 120, 11919–11929.

(13) Pierce, L.; Salomon-Ferrer, R.; F. de Oliveira, C.; McCammon, J.; Walker, R. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. J. Chem. Theory Comput. **2012**, 8, 2997–3002.

(14) Kraml, J.; Hofer, F.; Quoika, P. K.; Kamenik, A. S.; Liedl, K. R. *J. Chem. Inf. Model.* **2021**, *61*, 1533–1538.

EurJIC
European Journal of Inorganic Chemistry

Research Article
doi.org/10.1002/ejic.202200596

Chemistry Europe
European Chemical
Societies Publishing

www.eurjic.org

Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]arene C–S Coupling

Armando Berlanga-Vázquez,[a] Radu A. Talmazan,[b] Carlos A. Reyes-Mata,[a] Edmundo G. Percástegui,[c] Marcos Flores-Alamo,[d] Maren Podewitz,*[b] and Ivan Castillo*[a]

Functionalization of the phenolic rim of *p-tert*-butylcalix[8]arene with phenanthroline to create a cavity leads to formation of two regioisomers. Substitution of positions 1 and 5 produces the known C_{2v}-symmetric regioisomer 1,5-(2,9-dimethyl-1,10-phenanthroyl)-*p-tert*-butylcalix[8]arene (**L**^{1,5}), while substitution of positions 1 and 4 produces the C_s-symmetric regioisomer 1,4-(2,9-dimethyl-1,10-phenanthroyl)-*p-tert*-butylcalix[8]arene (**L**^{1,4}) described herein. [**Cu(L**^{1,4}**)I**] was synthesized from **L**^{1,4} and CuI in good yield and characterized spectroscopically. To evaluate the effect of its cavity on catalysis, Ullmann-type C–S coupling was chosen as proof-of-concept. Selected aryl halides were used, and the results compared with the previously reported Cu(I)/**L**^{1,5} system. Only highly activated aryl halides generate the C–S coupling product in moderate yields with the Cu(I)/**L**^{1,4} system. To shed light on these observations, detailed computational investigations were carried out, revealing the influence of the calix[8]arene macrocyclic morphology on the accessible conformations. The **L**^{1,4} regioisomer undergoes a deformation that does not occur with **L**^{1,5}, resulting in an exposed catalytic center, presumably the cause of the low activity of the former system. The 1,4-connectivity was confirmed in the solid-state structure of the byproduct [**Cu(L**^{1,4}**–H)**(CH_3CN)$_2$] that features Cu(I) coordinated inside a cleft defined by the macrocyclic framework.

Introduction

The application of molecules that provide a confined space as catalysts has been the focus of increasing attention in recent years.[1–4] These systems present a cavity as their main feature,[5] where they can host diverse reactants in a chemical environment that may differ significantly from that provided by the bulk solution. In this context, the macrocycles known as calix[n]arenes have been used as nanoreactors for a large range of processes, including as stabilizers for supported molecular noble-metal catalysts,[6] organic transformations,[7–10] small molecule activation,[11–13] as well as C–C[14,15] and C–N[16] cross-coupling reactions. Although the examples cited create a confined environment for the catalytic reaction to proceed, the metal involved is found outside the cavity of the calixarene (*exo*-coordinated). Very few examples of *endo*-coordinated metal complexes exist,[17,18] and even fewer that participate in catalytic transformations.[19] Calixarenes are synthetically available on a large scale and can be functionalized with relative ease.[20] In the specific case of calix[8]arenes that feature eight phenolic moieties, functionalization requires the use of a base, with CsF or Cs_2CO_3 commonly employed due to the additional templating effect of caesium, which allows control of the regioselectivity.[21,22]

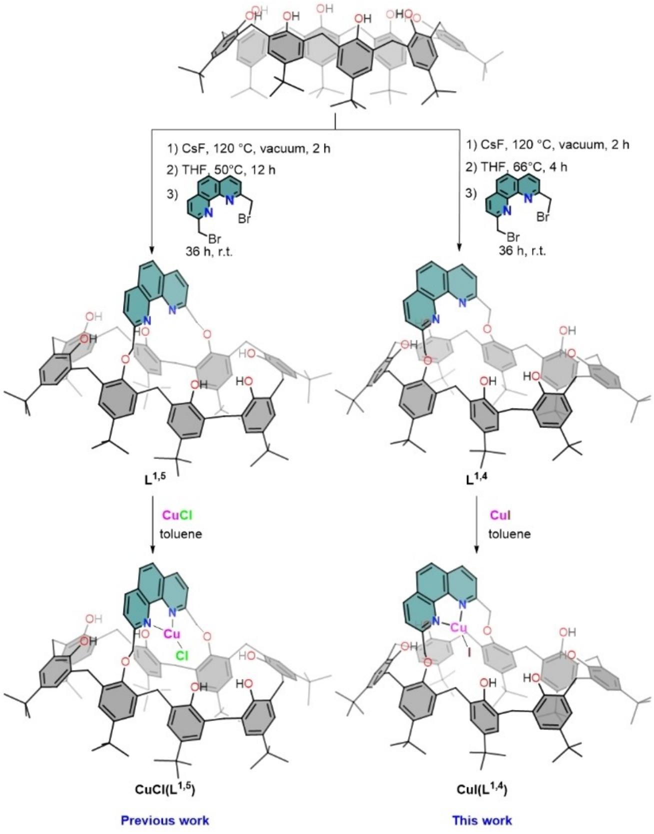
Our research group reported the catalytic activity of copper-based calix[8]arene derivatives for C–S cross coupling reactions, with a bridging phenanthroline motif introduced at the 1 and 5 positions of the phenolic rim (**L**^{1,5} in Scheme 1, left). The Cu(I) complex of **L**^{1,5} is very active for C–S coupling compared to molecular analogues, and displays a higher reactivity towards arylbromides, relative to aryliodides. [23] Under the conditions employed to obtain the 1,5-functionalized calix[8]arene, a side-product obtained in low yield from the reaction mixture was suspected to correspond to a regioisomer with different attachment points to the phenolic positions. Modification of the conditions resulted in improved yields of the latter compound that based on the isolated mass, corresponds to an almost 1:1 mixture of **L**^{1,5} and the new product (Scheme 1, right). To investigate the identity and properties of the previously unidentified regioisomer, we herein report the isolation and characterization of the new compound as the 1,4-phenanthroline substituted calix[8]arene (**L**^{1,4}) and the catalytic activity of its Cu(I) complexes in C–S coupling reactions for comparison

---

[a] A. Berlanga-Vázquez, C. A. Reyes-Mata, Prof. I. Castillo
Instituto de Química, Universidad Nacional Autónoma de México
Ciudad Universitaria, 04510, Ciudad de México (México)
E-mail: joseivan@unam.mx
iquimica.unam.mx/dr-ivan-castillo-perez
[b] R. A. Talmazan, Prof. M. Podewitz
Institute of Materials Chemistry, TU Wien
Getreidemarkt 9, 1060 Vienna (Austria)
E-mail: maren.podewitz@tuwien.ac.at
www.podewitz-lab.org
[c] Prof. E. G. Percástegui
Centro Conjunto de Investigación en Química Sustentable UAEM-UNAM
Carretera Toluca-Atlacomulco km 14.5
Toluca, 50200, Estado de México (México)
[d] Dr. M. Flores-Alamo
Facultad de Química, División de Estudios de Posgrado
Universidad Nacional Autónoma de México
Ciudad Universitaria, 04510, Ciudad de México (México)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/ejic.202200596

Part of the Special Collection on "Inorganic Reaction Mechanisms".



Scheme 1. Synthetic route for [**Cu(L^{1,5})Cl**] (references 23, 25) and [**Cu(L^{1,4})I**] (this work).

with the **L**^{1,5}-based system. The 1,4-substitution pattern may produce significant differences in the properties of the cavity, and concomitantly modify the catalytic activity of its Cu(I) complex. Molecular Dynamics (MD) simulations in explicit solvent, accompanied by Density Functional Theory (DFT) calculations capture the dynamic behavior of the flexible macrocycle and its copper complexes. They revealed that regioisomer-induced conformational changes of the calix[8]arene cavity are likely responsible for the dramatic difference in the catalytic performance of both systems, with the 1,4-regioisomer acting as a poor C–S coupling catalyst. Our studies provide guidance for the future development of better calixarene-based nanoreactors.[24]

Results and Discussion

Synthesis and Characterization

Calixarene derivatives. The functionalization of calix[8]arene with a phenanthroline bridge has previously been reported by our group.[25] In this O-alkylation process, one equivalent of calix[8]arene reacts with 1.2 equivs. of 2,9-bis(bromomethyl)-1,10-phenanthroline in THF in the presence of 10 equivs. of CsF as base and template. Deprotonation at 50 °C overnight leads to 82–86 % of the C_{2v}-symmetric regioisomer **L**^{1,5} in the original procedure,[25] which made identification of the small amount of side product unviable. In contrast, heating to reflux temperature (66 °C) for 4 h followed by addition of of 2,9-bis(bromomethyl)-1,10-phenanthroline gives rise to a nearly 1:1 mixture of **L**^{1,5} and **L**^{1,4} (Scheme 1). TLC analysis of the latter reaction mixture with 3:1 hexane/acetone as eluant shows the presence of only two products with very similar R_F (0.225 for **L**^{1,5} and 0.275 for **L**^{1,4}, determined after spectroscopic characterization, Figures S1–S7), making it challenging to separate via column chromatography. Separation of the two isomers involved crystallization from 8:1 $CHCl_3$/toluene mixture and cooling to −30 °C of the caesium complexes formed in the presence of excess CsF during alkylation with 2,9-bis(bromomethyl)-1,10-phenanthroline, resulting in the solid identified previously as [**Cs(L^{1,5}–H)**] as a white crystalline solid.[25] Separate treatment of [**Cs(L^{1,5}–H)**] dissolved in CH_2Cl_2, and the pale-yellow mother liquor containing [**Cs(L^{1,4}–H)**] with 0.1 N HCl each afforded the pure **L**^{1,5} and **L**^{1,4}, assessed initially by TLC analysis. The previously unidentified **L**^{1,4} was characterized by MALDI-TOF MS, with a peak at $m/z = 1501.7$ corresponding to the molecular ion (Figure S3). ^1H NMR spectroscopic analysis of **L**^{1,4} was undertaken in $C_2D_2Cl_4$ solution, since this solvent results in better resolved peaks for **L**^{1,5}.[23] Noticeable differences include the number of signals arising from the methylene protons of the calixarene framework: the less symmetric **L**^{1,4} should give rise to 5 doublets detected as broad peaks likely due to fluxional processes (Figure S4, see also theoretical section). For comparison, **L**^{1,5} displays only a pair of 1:1 signals due to the symmetry plane defined by the phenanthroline moiety that includes the O-atoms in 1 and 5 positions (Figure S5). The flexibility of macrocyclic **L**^{1,4} is also reflected in the poorly resolved ^{13}C NMR spectrum (Figure S6). The signals sharpen in ^1H NMR data acquired at higher temperature, but they do not coalesce completely up to 120 °C (Figures S7-S8).

Copper complexes. Once isolation and identification of **L**^{1,4} was achieved, the corresponding Cu(I) complex was obtained upon treatment of a toluene solution of the macrocycle with a stoichiometric amount of copper iodide dissolved in a minimum amount of acetonitrile under inert atmosphere. Cooling to −30 °C resulted in a brown microcrystalline solid in 82 % yield, formulated as [**Cu(L^{1,4})I**] (see below). The IR spectrum evidences a negligible displacement of the C–N band after Cu(I) complexation at 1591 cm^{-1}, relative to the free ligand at 1594 cm^{-1} (Figures S1 and S2).[33] Characterization by ^1H NMR spectroscopy confirms that the complex is diamagnetic, as expected for a Cu(I) d^{10} center (Figures S5 and S8). The apparent flexibility of the ligand in solution is restricted upon Cu(I) coordination; this can be discerned in the methylene group signals, as they get resolved into seven doublets with close to 1:1:1:1:1:1:2 ratio at room temperature in $C_2D_2Cl_4$ (Figures S8). MALDI-TOF MS of the complex obtained in acetonitrile solution shows a peak at $m/z = 1503.6$ corresponding to protonated $[\mathbf{L^{1,4}}+H]^+$, and in addition to the usually observed peak at $m/z = 1564.6$ $[Cu(\mathbf{L^{1,4}})]^+$ with low intensity, an iodide-containing species at $m/z = 1774.1$ assigned as $[Cu_2(\mathbf{L^{1,4}})(H_2O)I]^+$, based on its mass and isotopic distribution (Figure S9); formation of this dicopper ion in the ionization chamber is attributed to redistribution of labile Cu(I).

The presence of iodide as ligand is supported by addition of AgOTf to an acetonitrile solution of **[Cu($L^{1,4}$)I]**, which resulted in AgI as a water-insoluble, pale-yellow precipitate. **[Cu($L^{1,4}$)I]** is EPR silent at X-band frequency, as expected for a diamagnetic d^{10} ion (Figure S10). Finally, combustion analysis is consistent with the formulation as [Cu($\mathbf{L}^{1,4}$)(CH_3CN)(H_2O)I], likely due to the presence of acetonitrile from crystallization (Figure S8) and moisture absorbed during handling in air for analysis.

A separate batch of pale-yellow crystals were obtained by evaporation of a concentrated 9:1 acetonitrile/toluene solution in the triclinic space group *P*-1, and identified as [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$] based on X-ray diffraction. However, this species is only a minor product and the bulk of the isolated material corresponds to **[Cu($L^{1,4}$)I]**, as stated previously and confirmed by combustion analysis, MS, NMR spectroscopy, and reactivity with AgOTf. Hence, the low yield of [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$] did not allow us to test its catalytic activity extensively (see below). The crystals lose solvent rapidly once removed from the mother liquor at room temperature, and they diffract poorly despite immediate mounting on a glass capillary under a stream of cold N_2. Nonetheless, careful and rapid handling of the crystals allowed the collection of diffraction data for [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$], Figure 1. It consists of a calix[8]arene with the 2,9-dimethylphenanthroyl group bridging the 1 and 4 phenolic positions. One phenolic oxygen donor [O4] is bound to the Cu(I) center, which is chelated by the phenanthroline moiety; its coordination environment is complemented by two acetonitrile molecules as additional N-donors. Notably, the original iodide counterion is absent in these yellow crystals and a deprotonated phenol of the calixarene framework acts as charge compensator. The local geometry of the Cu(I) ion in [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$] can be described as a very distorted square pyramid $\tau^5 = 0.228$,[26] with the two phenanthroline N-atoms, O4 and one acetonitrile N-donor at the base of the pyramid, and the second acetonitrile molecule in the apex. The copper-nitrogen bond distances Cu1-N1 and Cu1-N2 (2.064 and 2.079 Å), and the N1-Cu1-N2 angle of 80.0° compare well with related Cu(I) phenanthroline complexes.[27,28] The Cu1-O4 length of 2.746 Å, is shorter than the sum of the van der Waals radii of copper and oxygen (2.9 Å).[29,30] Selected bond lengths (Å) and angles (°) are presented in Table 1. Crystallographic data and structure refinement details are presented in Table S1 in the Supporting Information.

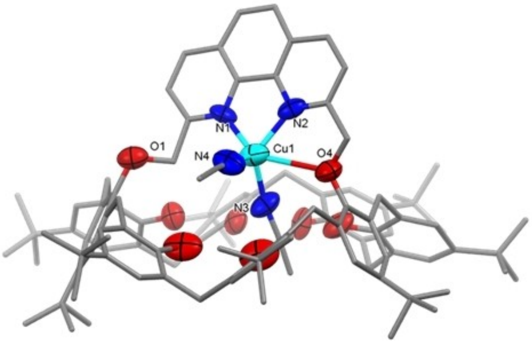


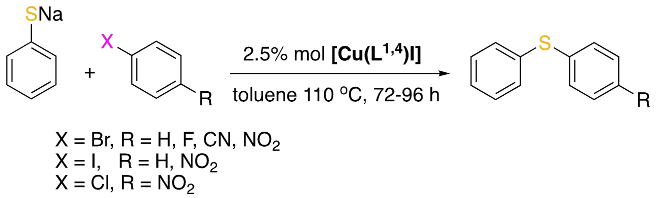
Figure 1. Mercury diagram of [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$] (H-atoms and solvent molecules omitted, C-atoms shown as wireframe for clarity). Ellipsoids are presented at the 50% probability level.

The conformation adopted by the substituted calix[8]arene in [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$] can be described as 1,2,3,4-alternate, where all phenolic groups are in *syn* orientation.[31] This conformation has been previously observed in calix[8]arene complexes of lanthanides, thorium, and molybdenum, although in all those cases the phenolic oxygen atoms are the only donors present.[32] The arrangement is stabilized by an intramolecular hydrogen-bonding network between the phenolic OH groups, which distributes the anionic charge among the phenol groups. The phenanthroyl bridge connecting the 1 and 4 positions defines a bimacrocyclic calixarene scaffold with an asymmetric cleft composed of a small and a large cycle, integrated by one 23- and one 31-membered rings, respectively. In contrast, the cavity of **L**^{1,5} consists of two symmetric 27-membered rings. For comparative purposes, the small cavity of **L**^{1,4} is only marginally larger than those of calix[4]arene and calix[5]arenes (16- and 20-membered rings), and slightly smaller than that of calix[6]arene (24-membered ring).

Catalytic evaluation. In previous work, **[Cu($L^{1,5}$)Cl]** had been tested as a catalyst precursor for C–S cross couplings, affording arylthioether products in 70–95% yields with a variety of aryl halides. The activity of this monometallic complex is significantly different from that observed when using small molecular complexes. For example, Bates et al. reported the dimeric complexes [(Me_2phen)Cu(μ-I)]$_2$ and [(Me_2phen)Cu(μ-SC_6H_5)]$_2$ that required 10% mol CuI and 2,9-dimethyl-1,10-phenanthroline (Me_2phen) loads for acceptable catalytic activity.[34] In contrast, the use of 2.5% mol loads of **[Cu($L^{1,5}$)Cl]** displayed high activity and unusual size-selectivity, demonstrating the influence of the cavity in the process that allows the less hindered substrates to access and interact with the copper center, thus favoring shorter reaction times. For direct comparison, we initially tested the catalytic activity of in situ formed **[Cu($L^{1,4}$)Cl]** in C–S couplings under analogous reaction conditions, with poor results. Although the difference is only the position of the phenanthroyl moiety, this modifies the size of the cavity, as well as the energetically accessible conformations in solution (see calculations below). Thus, Ullmann-type C–S couplings with the potentially more active iodide complex **[Cu($L^{1,4}$)I]** as catalyst precursor were carried out following the conditions shown in Scheme 2. We used bromobenzene in initial tests, based on the preference (size-selectivity) over

Table 1. Selected bond lengths (Å) and angles (°) for [**Cu**(**L**^{1,4}–**H**)(CH_3CN)$_2$].

Bond lengths		Bond angles	
Cu1-N4	1.944(10)	N4-Cu1-N3	111.6(3)
Cu1-N3	1.987(8)	N4-Cu1-N1	107.5(4)
Cu1-N1	2.064(8)	N3-Cu1-N1	118.2(3)
Cu1-N2	2.079(8)	N4-Cu1-N2	110.0(3)
Cu1-O4	2.746(9)	N1-Cu1-N2	80.0(3)
		O4-Cu1-N2	62.9(2)
		O4-Cu1-N3	75.3(3)
		O4-Cu1-N4	101.1(3)

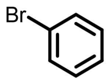
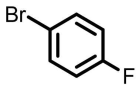
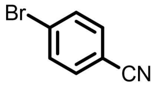
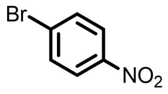
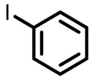
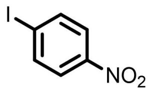
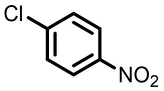


Scheme 2. C–S Ullmann-type couplings with **[Cu(L^{1,4})I]** as catalyst precursor.

iodobenzene exhibited previously by **[Cu(L^{1,5})Cl]**, and the activated 4-fluorobromobenzene, 4-bromobenzonitrile, and 4-nitrobromobenzene (entries 1–4 in Table 1). The catalyst and sodium thiophenolate were suspended in toluene and stirred for an hour, followed by addition of the corresponding arylbromide. The reactions were monitored by TLC analysis at 24 and 72 h intervals, and after such time the mixtures were cooled to room temperature for analysis. The products were separated by column chromatography and isolated or identified whenever necessary by ^{1}H NMR spectroscopy and DART-MS; the yields of isolated coupling product are shown in Table 2. Although the reactions were carried out under inert atmosphere, **[Cu(L^{1,4})I]** can be handled in air, based on EPR analysis of samples of the complex exposed to air for several days (Figure S11). This is attributed to the steric protection of Cu(I) provided by the calixarene scaffold.

Bromobenzene was the first substrate tested, and no cross-coupling product was detected by TLC analysis, even after 72 h of reaction at 110 °C; only reactants and catalyst were observed. This was also the case after chromatographic separation. Catalyst recovery was partially successful: although the complex appears to decompose in silica, the ligand was isolated with a maximum of 80 % yield. In a few specific cases of the coupling reactions, the copper complex was recovered as well in ca. 30 % yield. Unlike copper catalyst with **L^{1,5}**, for which the cross-coupling reaction generates diarylsulfides in 15 h with up to 95 % yield, **[Cu(L^{1,4})I]** is not a viable catalyst with the simplest substrate bromobenzene. Therefore, we decided to try substrates with activating (electron withdrawing) groups in the *para* position (F, CN, and NO$_2$). Even though 4-fluorobromobenzene may be considered as activated relative to bromobenzene, no reaction with sodium thiophenolate was observed. When 4-bromobenzonitrile was employed as substrate, the coupling product was spotted by TLC, ^{1}H NMR studies, and mass spectra analysis (Figures S12 and S13), but this was isolated only in trace amounts. Coupling was successful only with the most highly activated 4-nitrobromobenzene, affording a new compound that was isolated as an intense yellow oil after column chromatography. DART-MS shows a peak corresponding to the desired coupling product (4-nitrophenyl)phenylsulfide at $m/z = 232$ for $[M+H]^+$ (Figure S14). ^{1}H NMR spectroscopy in Figure S14 confirms the identity of the coupling product, corresponding to 35 % yield.

Table 2. Ullmann-type C–S cross-coupling of aryl halides using **[Cu(L^{1,5})Cl]**[a] and **[Cu(L^{1,4})I]**.

Entry	ArX	Product yield [%]	
		[Cu(L^{1,5})Cl][a]	**[Cu(L^{1,4})I]**
1		95	0
2		NA	0
3		94	Trace
4		95	35
5		88	0
6		NA	31
7		94	19

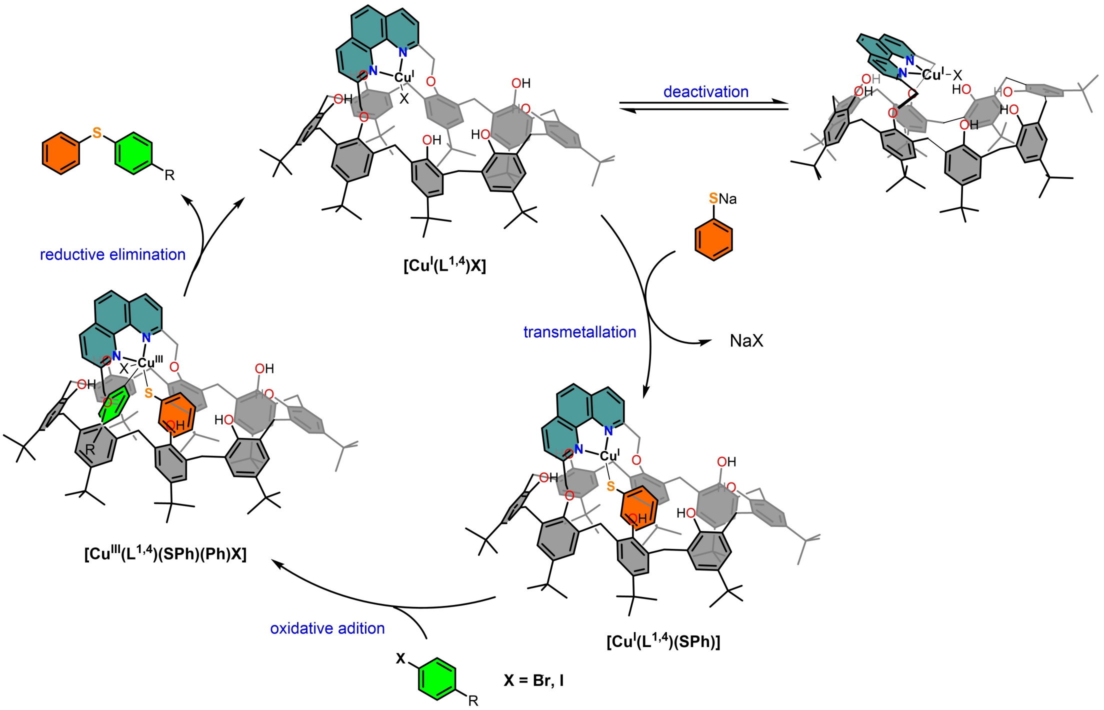
[a] Previous work from Ref. [23]. NA: Not available.

Considering that **[Cu(L^{1,4})I]** requires the most active substituted bromobenzene substrate to perform the coupling reaction, three additional tests were carried out using iodobenzene, 4-nitroiodobenzene and 4-nitrochlorobenzene. These substrates allowed us to evaluate the effect of the size of the halide in reactivity under the same conditions employed with arylbromides. The results of these experiments are summarized in Table 2 entries 5–7. When using iodobenzene, no product could be identified as diphenyl sulfide after column chromatography. Reactions with 4-nitroiodobenzene and 4-nitrochlorobenzene afforded the coupling product (4-nitrophenyl)phenylsulfide, isolated after column chromatography, corresponding to 31 % and 19 % yields respectively (see DART-MS in Figures S16 and S17). These observations evidence that coupling reactions proceed only when using *highly activated* aryl halides with **[Cu(L^{1,4})I]** as catalyst, albeit in moderate yields. This contrasts with the high activity established for the 1,5-substituted system with all types of substrates, although the size preference for aryl bromides over aryl iodides is retained. The low catalytic activity of **[Cu(L^{1,4})I]** can be ascribed to the difference of the cavity size and shape, and potentially to the conformations that may be accessible for both macrocycles. While in **[Cu(L^{1,5})Cl]** the Cu(I) ion is oriented towards the center of the cavity, allowing the substrates to enter and interact with the metal ion, in **[Cu(L^{1,4})I]** the position of the Cu(I) ion is shifted to one side, restricting the conformations that may lead to high conversions with any arylhalide.

A common phenomenon observed in all reactions tested was that biphenyl was not detected. This supports the notion that in **[Cu(L^{1,4})I]** the space to access the copper center is highly restricted by the conformations available to the calixarene framework. The extent of restriction is such that two aryl halides cannot be fitted simultaneously for homocoupling to take place, regardless of the identity of the halogen substituent on the aromatic substrates. When C–S coupling does occur, the mechanism must be similar to that postulated for these types of systems, where initial formation of Cu(I)-thiophenolate is followed by oxidative addition of the arylhalide, and product formation takes place by reductive elimination in the last step in Scheme 3. As an alternative, if **[Cu(L^{1,4}–H)**$(CH_3CN)_2$] were present, the calixarene phenolate moiety (**L^{1,4}–H)** may be protonated by the more acidic thiophenol PhSH, which would lead to **[Cu(L^{1,4})(SPh)]** in Scheme 3; the rest of the mechanism should be identical to the one shown. This catalytic reaction was indeed tested with the small amount of [**Cu**-**(L^{1,4}–H)**$(CH_3CN)_2$] available and PhSH/PhBr, with no products observed (see Experimental Section for details).

Computational conformational studies. To shed light on the reasons for the considerably lower activity of **[Cu(L^{1,4})X]** relative to **[Cu(L^{1,5})X]**, a theoretical investigation of the structures was undertaken. Initial structure optimizations with DFT did not show a big difference in energy for the two regioisomeric complexes (**[Cu(L^{1,4})I]** and **[Cu(L^{1,5})I]**, shown in Figure S18), with **[Cu(L^{1,4})I]** being slightly more stable by 10 kJ mol^{-1}. To explore the flexibility of the calixarene macrocycle in solution, we performed accelerated MD (aMD) simulations of **[Cu(L^{1,4})I]** and **[Cu(L^{1,5})I]** in explicit chloroform (see computational methodology for details) — a methodology established in a previous study on a similar system.[35] To quantify the change in conformational flexibility induced by the bridging phenanthroyl moiety at the 1,4 and 1,5 positions, an analysis of the dihedral angles between the individual calixarene phenolic units was performed. Based on their distribution, an estimate for the dihedral entropy as an indicator of inherent flexibility can be obtained via kernel density estimation (see computational details). For a given dihedral distribution, the entropy value can be obtained by analyzing the area under the curve (the larger the area, the higher the entropy). As seen in Figure 2, where the distributions of the individual dihedral angles are plotted for both complexes, the **[Cu(L^{1,4})I]** structures exhibit significantly more flexibility, especially in the larger loop, as evidenced by the higher entropy values.

The MD simulations also revealed an interesting difference in the behavior of the two structures, with **[Cu(L^{1,4})I]** displaying a deformation of the cavity. This is in stark contrast to the symmetric **[Cu(L^{1,5})I]** catalyst. As depicted in Figure 3 (full structures in Figure S19 and S20), the catalytic center becomes exposed to solvent as the framework deforms. Transition from the deformed to the intact cavity happens several times over the course of a single 1 μs aMD simulation, as well as during the 100 ns conventional MD simulation used to obtain the aMD parameters (see computational details). Upon optimization of the cluster representatives of the two distinct conformations with DFT, it is evident that the deformed structure is slightly more stable than the intact one ($\Delta E = -10$ kJ mol^{-1}). These findings provide evidence that the deformation of the cavity is likely to occur under experimental conditions and may explain the low catalytic activity of **[Cu(L^{1,4})I]** relative to **[Cu(L^{1,5})I]**. Although the deformed calculated structures suggest that dimerization might be feasible, no experimental evidence for deactivation through Cu–I–Cu bridge formation, i.e. dimerization, was found. If the reaction were to proceed in the



Scheme 3. Proposed mechanism for the (inefficient) C–S cross coupling of thiophenolate with aryl halides using **[Cu(L^{1,4})I]** as catalyst, including potential deactivation pathway.

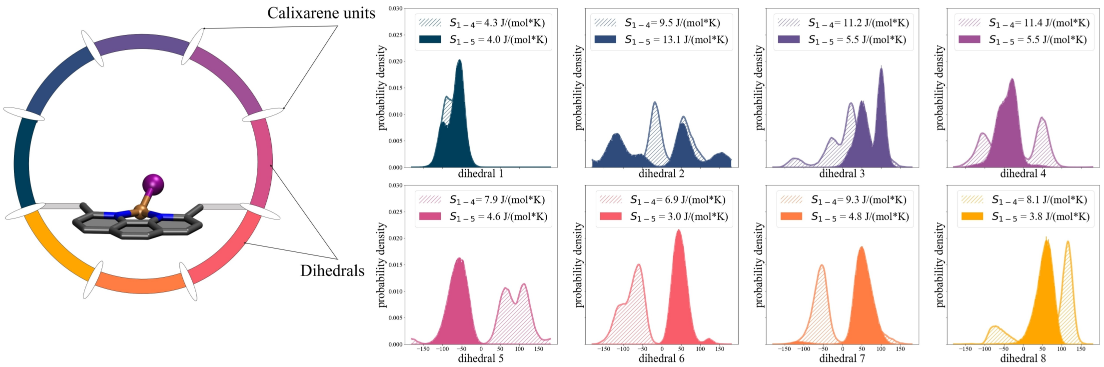


Figure 2. Left: schematic representation of **[Cu(L^{1,4})I]** depicting the position of each dihedral angle. Right: dihedral entropy estimated from the area under the curve of **[Cu(L^{1,4})I]** (patterned curves) and **[Cu(L^{1,5})I]** (filled curves) as observed during aMD simulations.

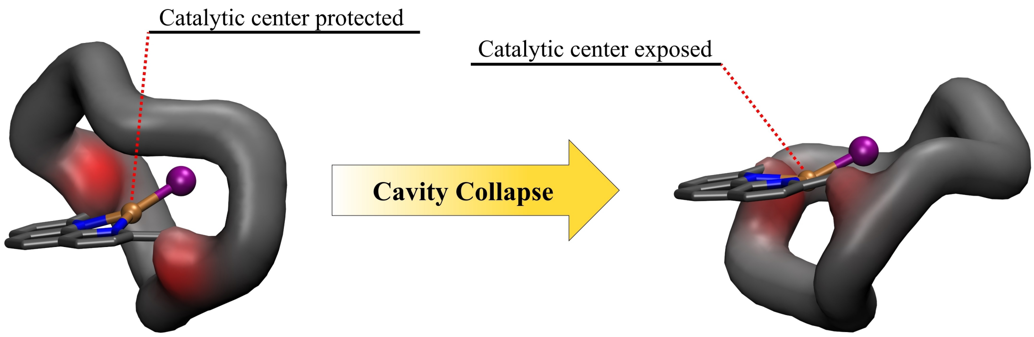


Figure 3. Schematic representation of the intact and deformed structures of **[Cu(L^{1,4})I]** observed during the aMD simulations in explicit solvent. The calixarene macrocycle is displayed as wiggle. Left: front and side views of the intact conformation; right: deformed structure with exposed Cu(I) center. Detailed structures shown in Figures S17 and S18 in the Supporting Information.

deformed conformation with an exposed catalytic center, replacement of iodide for thiophenolate would require more open space around the Cu center. The transmetallation is a two-step process, wherein a bulky, negatively charged intermediate, **[Cu(L^{1,4})I(SPh)]$^-$**, would be formed upon the addition of thiophenolate. This would be followed by the dissociation of iodide, resulting in **[Cu(L^{1,4})(SPh)]** (Scheme 3). To investigate the first step of the transmetallation, we modelled the intermediate as **[Cu(L^{1,4})I(SPh)]$^-$Na$^+$** to balance the charge in the simulation (see computational methodology for details). Given that thiophenolate is added as a sodium salt and that the solvent is non polar, it is reasonable to assume that the Na$^+$ remains in the vicinity of the negatively charged **[Cu(L^{1,4})I(SPh)]$^-$** complex. Our aMD simulations performed on the **[Cu(L^{1,4})I(SPh)]$^-$Na$^+$** intermediate, revealed a preference for the deformed structure. This finding supports the hypothesis that the large steric demand of the intermediate prevents the calixarene from flipping back to its original position and restoring the cavity. Indeed, the starting structure transitioned to a deformed one within the 100 ns conventional MD simulation. Transitions between conformations are rare in a 1 µs simulation, where only 2.2% of the structures show an intact cage. Moreover, a simple experimental transmetallation test between **[Cu(L^{1,4})I]** and NaSPh at room temperature did not afford evidence of the formation of [Cu(**L^{1,4}**)(**SPh**)] by MALDI-TOF MS. Re-optimization of the deformed **[Cu(L^{1,4})I(SPh)]$^-$Na$^+$** structures, obtained from MD simulations after clustering of the trajectory, with DFT revealed that these very compact conformations (see Figures S21–S22) are surprisingly stable. In fact, the deformed structure is 165 kJ mol^{-1} more stable than that with an intact cage. This finding is in line with results from MD simulations, where the intact cage conformation was hardly populated. Attempts to optimize the subsequent putative reactive intermediate **[Cu(L^{1,4})X(SPh)(Ph)]** (X=Br, I), proposed in Scheme 3 failed, with the calculations directly leading to the formation of the C–S coupling product (see Figure S23). This result suggests that sterically demanding species are not stable within the coordination environment of **[Cu(L^{1,4})I]**. This means either the oxidative addition cannot take place, or it is immediately followed by reductive elimination. In any case, the computational investigations highlight the importance of the 1,5-substitution pattern of **L^{1,5}** to preserve a reaction cavity that results in high catalytic activity, allowing the reaction to proceed via the known oxidative addition/reductive elimination pathway.

Conclusion

We have successfully isolated and characterized the 1,4-regioisomer of *p-tert*-butylcalix[8]arene substituted with a phenanthroyl moiety **L^{1,4}** by easy separation of the cesium complex of **L^{1,5}** via crystallization. **L^{1,4}** was employed as ligand to obtain the new copper(I) complex **[Cu(L^{1,4})I]**. Spectroscopic studies and structure optimization of **[Cu(L^{1,4})I]** by theoretical methods evidenced that the asymmetric cleft provided by **L^{1,4}** forces Cu(I) towards one side of the calixarene framework, relative to its symmetric [**Cu(L^{1,5})I]** counterpart. Catalytic C–S cross-coupling reactions of sodium thiophenolate with different aryl halides and **[Cu(L^{1,4})I]** as catalyst were tested. Product formation only occurs when highly activated 4-nitroaryl halides were used, although the yields obtained reached only a maximum value of 35% despite extended reaction times. The computational investigation of the two regioisomers delivers new insights into the conformational dynamics that can affect supramolecular catalytic activity. The high flexibility of the asymmetric cage resulted in a deformation of the calix[8]arene structure in **[Cu(L^{1,4})I]** that is likely the cause for the low catalytic activity. Our studies highlight the importance of the conformational landscape available for inherently flexible macrocycles, particularly when aimed at providing a reactive pocket for catalytic transformations. X-ray diffraction studies of the minor product [**Cu(L^{1,4}−H)**($CH_3CN)_2$] confirm the nature of the 1,4-isomer, as well as the rare *endo*-oriented coordination of the metal center.[17] Studies with different transition-metal ions hosted within **L^{1,4}** and **L^{1,5}** to further explore their properties are currently underway.

Experimental Section

Materials and methods: All preparations and manipulations of air sensitive compounds were carried out under dinitrogen atmosphere using an MBraun glovebox or standard Schlenk techniques. Copper iodide (CuI), sodium thiophenolate ($NaSC_6H_5$), bromobenzene (C_6H_5Br), 4-fluorobromobenzene (BrC_6H_4F), 4-bromobenzonitrile (BrC_6H_4CN), 4-nitrobromobenzene ($BrC_6H_4NO_2$), iodobenzene (IC_6H_5), 4-nitroiodobenzene ($IC_6H_4NO_2$) and 4-nitrochlorobenzene ($ClC_6H_4NO_2$) were purchased from Sigma-Aldrich and used without further purification. Toluene (Sigma-Aldrich, 99.5%) was dried over Na/benzophenone, distilled under N_2, and degassed via three freeze-pump-thaw cycles.[36] ^{1}H NMR spectra were recorded on a JEOL Eclipse spectrometer operating at 300 MHz and referenced to the residual solvent signal of $C_2D_2Cl_4$ (s, $\delta = 6.0$ ppm). MALDI-TOF MS were acquired with a Bruker Microflex MALDI-TOF mass spectrometer using a 2,5-dihydroxybenzoic acid matrix. ATR-IR spectra were recorded on a Frontier Perkin Elmer FTIR spectrometer. EPR spectra were acquired in quartz tubes with a JEOL JES TE300 spectrometer operating at X-band frequency (9.4 GHz) at 100 KHz field modulation, with a cylindrical cavity (TE011 mode). Thin layer chromatography (TLC) was conducted using aluminum-backed TLC Silica Gel 60 F_{254}. A mixture of hexane/acetone 3:1 was used as eluant. Visualization of developed plates was performed under UV-vis light (245 nm) or with an iodine chamber. Retardation factors (R_F) were calculated as recommended by IUPAC.[37] Column chromatography was performed using 70–230 mm silica gel as stationary phase and hexane as eluant.

Crystallographic details: Due to the high instability of single crystals of [**Cu(L^{1,4}−H)**($CH_3CN)_2$], several crystals had to be tested. A suitable single crystal was mounted on a glass fiber; crystallographic data were collected with an Oxford Diffraction Gemini "A" diffractometer with a CCD area detector, with $\lambda_{Mo K\alpha} = 0.71073$ Å at 170 K. Unit cell parameters were determined with a set of three runs of 15 frames (1° in ω). The double pass method of scanning was used to exclude any noise.[38] The collected frames were integrated by using an orientation matrix determined from the narrow frame scans. Final cell constants were determined by a global refinement; collected data were corrected for absorbance by using analytical numeric absorption correction using a multifaceted crystal model based on expressions upon the Laue symmetry with equivalent reflections. Structures solutions and refinement were carried out with the SHELXS-2014[39] and SHELXL-2014[40] packages. WinGX v2018.[41] software was used to prepare material for publication. Full-matrix least-squares refinement was carried out by minimizing $(Fo^2-Fc^2)^2$. All non-hydrogen atoms were refined anisotropically. H atoms attached to C atoms were placed in geometrically idealized positions and refined as riding on their parent atoms, with C–H = 0.95–0.99 Å and with $U_{iso}(H) = 1.2U_{eq}(C)$ for aromatic and methylene groups, and $1.5U_{eq}(C)$ for methyl groups. The *tert*-butyl groups located on C20, C31, C53, C75 and C97 are disordered. On the other hand, the solvent molecules were significantly disordered and could not be modelled properly, thus SQUEEZE,[42] a part of the PLATON package of crystallographic software, was used to calculate the solvents disorder areas and remove their contributions to the overall intensity data. The disordered solvent area is centered on the 0.084 0.411 0.428 position and showed an estimated total of 127 electrons and a void volume of 2271 Å^3. Crystallographic data for the complex is presented in Table S1 of the Supporting Information.

Deposition Number(s) 1441364 contain(s) the supplementary crystallographic data for this paper. These data are provided free of charge by the joint Cambridge Crystallographic Data Centre and Fachinformationszentrum Karlsruhe Access Structures service.

Synthesis of L^{1,4} and L^{1,5}: *p-tert*-butylcalix[8]arene (1.45 g, 0.91 mmol) and CsF (1.38 g, 9.1 mmol) were dried for two hours at 120 °C under vacuum. After allowing to cool to room temperature, anhydrous THF was added, and the mixture was heated to reflux for 4 h. Upon cooling to room temperature, 2,9-bis(bromomethyl)-1,10-phenanthroline (400 mg, 1.09 mmol) was added to the mixture and stirred for 36 h. The solvent was then evaporated under reduced pressure to afford a yellow solid that was dissolved in a $CHCl_3$/toluene 8:1 mixture and placed in a −30 °C freezer until a white precipitate appeared after 3–4 days. The solid was filtered to remove **[Cs(L^{1,5}−H)]** (41% yield of **L^{1,5}** after neutralization) and the solution containing **[Cs(L^{1,4}−H)]** was washed with 50 mL of HCl (0.1 M) to remove cesium as CsCl, and then washed with a saturated solution of $NaHCO_3$ for neutralization, affording **L^{1,4}** in 46% yield; m. p. = 200–202 °C (dec). IR (ATR) n (cm^{-1}) = 2959 (C–H), 1594 (C=N), 1481 (C=C). MALDI-TOF MS *m/z*: 1501 [**L^{1,4}** + H]$^+$. ^{1}H NMR ($C_2D_2Cl_4$, 300 MHz): δ 9.53 (s, 6H, OH), 8.42 (d, J = 8.03 Hz, 2H Ar_{phen}), 7.91 (m, 2H, Ar_{phen}), 7.81 (s, 2H, Ar_{phen}), 7.21 (m, 16H, Ar_{calix}), 5.53 (m, 2H, $-CH_2$-$_{phen}$), 5.17 (s, 2H, $-CH_2$-$_{phen}$), 4.43 (m, 4H, $-CH_2$-$_{calix}$), 4.10 (m, (8H, $-CH_2$-$_{calix}$), 3.63 (m, (4H, $-CH_2$-$_{calix}$), 1.40-1.30 (m, 72H, ^{t}Bu). Elemental analysis for $C_{103.5}H_{123.5}ClN_{2.5}O_{8.5}$ [**L^{1,4}**(0.5CH_3CN,0.5CH_2Cl_2, 0.5H_2O)] calcd.: C 78.98, H 7.91, N 2.22; found: C 78.70, H 7.62, N 2.11.

Synthesis of [Cu(L^{1,4})I]: To a solution of **L^{1,4}** (390 mg, 0.26 mmol) in 5 mL of anhydrous, and deoxygenated toluene inside a glovebox was added CuI (49 mg, 0.26 mmol). The brown solution was stirred at room temperature overnight. Toluene was then evaporated under reduced pressure to afford a brown solid that was dissolved in an acetonitrile/toluene 9:1 mixture. Slow evaporation afforded a small amount of yellow crystals (<5% [**Cu(L^{1,4}–H)**(CH$_3$CN)$_2$]) that were collected for X-ray diffraction; elemental analysis for $C_{104}H_{124}CuN_3O_9$ [Cu(**L^{1,4}–H**)(CH$_3$CN)(H$_2$O)] calcd.: C 76.93, H 7.70, N 2.59; found: C 76.90, H 7.46, N 2.73. The remaining solution was placed in a −30 °C freezer, resulting in a brown microcrystalline solid (360 mg, 82% yield); m. p. = 170–172 °C (dec). IR ATR ν (cm^{-1}) = 2954 (C–H), 1591 (C=N), 1481 (C=C). MALDI-TOF MS *m/z*: 1564 [Cu(**L^{1,4}**)]$^+$. ^{1}H NMR (C$_2$D$_2$Cl$_4$, 300 MHz): δ 9.17 (s, 1H, OH), 8.56 (d, J = 8.0 Hz, 2H Ar$_{phen}$), 8.31 (s, 2H, Ar$_{phen}$), 7.94 (m, 2H, Ar$_{phen}$), 6.98 (m, 16H, Ar$_{calix}$), 6.50 (d, J = 16.1 Hz, 2H, $-CH_{2\text{-}phen}$), 5.82 (d, J = 16.1 Hz, 2H, $-CH_{2\text{-}phen}$), 4.69 (d, J = 12.8 Hz, 2H, $-CH_{2\text{-}calix}$), 4.35 (d, J = 12.4 Hz, 2H, $-CH_{2\text{-}calix}$), 4.15 (d, J = 13.4 Hz, 2H, $-CH_{2\text{-}calix}$), 3.93 (d, J = 15.7 Hz, 2H, $-CH_{2\text{-}calix}$), 3.78 (d, J = 15.0 Hz, 2H, —CH2-$_{calix}$), 3.61 (d, J = 12.9 Hz, 2H, $-CH_{2\text{-}calix}$), 3.53 (d, J = 14.6 Hz, 4H, $-CH_{2\text{-}calix}$), 1.37 (s, 9H, $^{t-}$Bu), 1.30 (s, 18H, $^{t-}$Bu), 1.26 (s, 24H, $^{t-}$Bu), 1.23 (s, 24H, $^{t-}$Bu), 1.07 (s, 24H, $^{t-}$Bu). Elemental analysis for $C_{104}H_{127}Cl_2CuIN_3O_{10}$ [Cu(**L^{1,4}**)(CH$_3$CN)(2H$_2$O)$_2$I] calcd.: C 70.59, H 7.23, N 2.37; found: C 70.39, H 6.99, N 2.38.

Coupling reactions: In a typical procedure, sodium thiophenolate (93 mg, 0.71 mmol) was added to a solution of **[Cu(L^{1,4})I]** (30 mg, 2.5% mol) in 5 mL of anhydrous toluene in a 100 mL Schlenk flask under dinitrogen atmosphere, the mixture was stirred for 1 h at room temperature followed by the addition of 0.71 mmol of the corresponding aryl bromide. The mixture was heated to 110 °C for 24–72 h. An analogous test was run with 30 mg [**Cu(L^{1,4}–H)**(CH$_3$CN)$_2$], 0.71 mmol thiophenol and 0.71 mmol bromobenzene; this resulted in no detectable products, as judged in all cases by TLC analysis on silica gel; final products when present were purified and isolated after column chromatography with hexanes/dichloromethane gradient (starting with 100% hexanes) as eluant.

Quantum chemical studies: The initial geometries of **[Cu(L^{1,4})I]** and **[Cu(L^{1,5})I]** for the optimizations were constructed by modification of available crystal structures using Chemcraft 1.8.[43] Quantum chemical calculations were performed with Turbomole 7.5,[44] unless otherwise specified. All investigated structures were fully optimized using the PBE0[45] functional and def2-SVP[46] basis set with empirical dispersion corrections of the D3 type,[47] along with implicit solvent corrections using the Conductor-Like Screening Model (COSMO).[48] A dielectric constant of ε = 2.4 was chosen to model toluene. The PBE0 functional is known to provide good results when investigating structures involving transition metals.[49,50] To obtain more accurate electronic energies, the def2-TZVP basis set[46,51–54] was used, alongside the same settings as mentioned above. The minima were verified to be true by running numerical frequency calculations which were then checked for the absence of imaginary frequencies.

Molecular dynamics simulations: Obtaining a representative structural ensemble of the calixarene required that the conformational spaces of these two species are examined along the reaction pathway, for both the **L^{1,4}** and the **L^{1,5}** type of structures, by selecting several intermediates. The exploration of the potential energy surface was done using classical accelerated MD simulations, a method proven to deliver reliable conformational ensembles for macrocycles.[20] Sampling in explicit solvent proved to be essential to retain intact cavities, as attempts in implicit solvent resulted in the cavity collapsing upon itself. Test calculations on **[Cu(L^{1,5})I]** have shown that there is no significant difference whether chloroform or toluene was used.[35] To this end, chloroform was picked as the explicit solvent of choice, due to a faster equilibration. We chose to simulate the supramolecular structures **[Cu(L^{1,4})I]** and **[Cu(L^{1,5})I]**, to assess how the flexibility of the cavity changes depending on the position of the bridge. We also simulated the intermediate **[Cu(L^{1,4})I(SPh)]$^-$Na$^+$**, which is the first species to be formed in the transmetallation process.[35] Therefore, a parametrization of the metal at the center of the molecule was required. MCPB.py[55] was selected as the tool of choice for the parametrization step. The structure optimizations and frequency calculations for the metal center parametrization were performed at the PBE0/def2-SVP/D3 level. **[Cu(L^{1,4})I(SPh)]$^-$Na$^+$** was simulated using Na$^+$ as an explicit counterion to reduce the net charge of the system to 0. This is in line with experimental conditions, where thiophenolate was added as a sodium salt.

To perform the initial equilibrations prior to simulation production runs, a modified procedure by Wallnöfer et al.[56] that involves extensive heating and cooling was employed. The production runs (in explicit chloroform) were carried out in NpT ensembles at 300 K, using Amber20.[57] The temperature was regulated with the Langevin thermostat,[58] and the pressure was kept at 1 bar using the Berendsen barostat.[59] The SHAKE algorithm[60] was used to restrain hydrogen bonds allowing for a time step of 2 fs; coordinates were saved every 10 ps, simulating a total of 1 μs. Accelerated MD simulations were carried out using the dual-boost algorithm implemented in Amber20, where a bias was applied on the total potential and an additional boost on the dihedral term. Several aMD simulations, each with 1μs were performed with various boosting parameters. These settings were derived by performing 100 ns classical MD simulations, as proposed by Pierce et al.[61] The obtained structures were aligned on the phenanthroyl bridge. A hierarchical clustering was applied on the trajectories to obtain a structurally diverse ensemble. The reweighting of the aMD simulations was not necessary, as the energetics were not a concern, since we aimed at obtaining a large conformational ensemble, from which representative structures were selected and further optimized with DFT. To investigate the flexibility of the calixarene units, a dihedral analysis was used, to characterize the rotations between each moiety. Dihedral angles were assigned between the planes of each of the 6-membered rings of the calixarene units, as changes in these angles accurately describe the movements within the calixarene ring. The relative entropy values were obtained by using the X-entropy[62] script that performs an entropic analysis on the angle distributions by kernel density estimation. Thus, it leads to an accurate estimation of the flexibility, whereby the more flexible the system, the larger the area under the dihedral distribution curve, and the larger the entropy. DFT calculations on the cluster representatives or structure optimizations were performed with the protocol described in the previous section. Structural parameter measurements and visualizations were done with PyMol and VMD.[63,64]

Acknowledgements

The authors thank María del Carmen García-González for DART-MS, Lucero Ríos-Ruiz for MALDI-TOF, Virginia Gómez-Vidales for EPR measurements, María de los Angeles Peña for NMR spectroscopy, and María de la Paz Orta for combustion analysis. A. B.-V., C. R.-M. and I. C. thank Instituto de Química, UNAM, and Conacyt (Becas 736348, 308307), Proyecto A1-S-8682 for financial support. M. P. thanks the FWF for financial support (P 33528).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article. Calculated structures are deposited at https://zenodo.org/record/7444229.

Keywords: Calixarenes · Conformational sampling · Copper · DFT calculations · Macrocycles

[1] T. Chavagnan, D. Sémeril, D. Matt, L. Toupet, *Eur. J. Org. Chem.* **2017**, *2017*, 313–323.
[2] H. Zhou, J. Tan, X. Zhang, *Chem. Asian J.* **2019**, *14*, 3240–3250.
[3] V. Mouarrawis, R. Plessius, J. I. van der Vlugt, J. N. H. Reek, *Front. Chem.* **2018**, *6*, 623.
[4] L. Leclercq, G. Douyère, V. Nardello-Rataj, *Catalysts* **2019**, *9*, 163.
[5] C. Deraedt, D. Astruc, *Coord. Chem. Rev.* **2016**, *324*, 106–122.
[6] C. Schöttle, E. Guan, A. Okrut, N. A. Grosso-Giordano, A. Palermo, A. Solovyov, B. C. Gates, A. Katz, *J. Am. Chem. Soc.* **2019**, *141*, 4010–4015.
[7] P. Molenveld, J. F. J. Engbersen, D. N. Reinhoudt, *Chem. Soc. Rev.* **2000**, *29*, 75–86.
[8] C. Dieleman, S. Steyer, C. Jeunesse, D. Matt, *J. Chem. Soc. Dalton Trans.* **2001**, 2508–2517.
[9] E. Karakhanov, T. Buchneva, A. Maximov, M. Zavertyaeva, *J. Mol. Catal. A* **2002**, *184*, 11–17.
[10] Z. X. Xu, G. K. Li, C. F. Chen, Z. T. Huang, *Tetrahedron* **2008**, *64*, 8668–8675.
[11] I. Di Bari, G. Granata, G. M. L. Consoli, S. Sortino, *New J. Chem.* **2018**, *42*, 18096–18101.
[12] X. Hang, W. Yang, S. Wang, H. Han, W. Liao, J. Jia, *ACS Appl. Nano Mater.* **2019**, *2*, 4232–4237.
[13] N. Noll, F. Würthner, *Chem. Eur. J.* **2021**, *27*, 444–450.
[14] A. Labattut, S. Abi Fayssal, J. Buendia, I. Abdellah, V. Huc, C. Martini, E. Schulz, *React. Chem. Eng.* **2020**, *5*, 1509–1514.
[15] N. Narkhede, B. Uttam, C. P. Rao, *ACS Omega* **2019**, *4*, 4908–4917.
[16] S. Abi Fayssal, T. Naret, V. Huc, J. Buendia, C. Martini, E. Schulz, *Catal. Sci. Technol.* **2021**, *11*, 5223–5231.
[17] C.-G. Yan, J. Han, L. Li, D.-M. Liu, *J. Coord. Chem.* **2009**, *62*, 825–832.
[18] N. Le Poul, Y. Le Mest, I. Jabin, O. Reinaud, *Acc. Chem. Res.* **2015**, *48*,2097–2106.
[19] G. Izzet, J. Zeitouny, H. Akdas-Killig, Y. Frapart, S. Ménage, B. Douziech, I. Jabin, Y. Le Mest, O. Reinaud, *J. Am. Chem. Soc.* **2008**, *130*, 9514–9523.
[20] V. Böhmer, *Angew. Chem. Int. Ed.* **1995**, *34*, 713–745.
[21] G. M. L. Consoli, F. Cunsolo, C. Geraci, P. Neri, *Org. Lett.* **2001**, *3*, 1605–1608.
[22] C. Gaeta, L. Gregoli, M. Martino, P. Neri, *Tetrahedron Lett.* **2002**, *43*, 8875–8878.
[23] E. Guzmán-Percástegui, D. J. Hernández, I. Castillo, *Chem. Commun.* **2016**, *52*, 3111–3114.
[24] R. H. Crabtree, *Chem. Rev.* **2015**, *115*, 127–150.
[25] D. J. Hernández, H. Vázquez-Lima, P. Guadarrama, D. Martínez-Otero, I. Castillo, *Tetrahedron Lett.* **2013**, *54*, 4930–4933.
[26] A. W. Addison, T. N. Rao, J. Reedijk, J. Van Rijn, G. C. Verschoor, *J. Chem. Soc. Dalton Trans.* **1984**, 1349–1356.
[27] J. Cody, J. Dennisson, J. Gilmore, D. G. VanDerveer, M. M. Henary, A. Gabrielli, C. D. Sherrill, Y. Zhang, C. P. Pan, C. Burda, C. J. Fahrni, *Inorg. Chem.* **2003**, *42*, 4918–4929.
[28] C. Chen, Z. Weng, J. F. Hartwig, *Organometallics* **2012**, *31*, 8031–8037.
[29] S. S. Batsanov, *Inorg. Mater.* **2001**, *37*, 871–885.
[30] A. O. Legendre, F. C. Andrade, S. R. Ananias, A. E. Mauro, A. V. G. Netto, R. H. A. Santos, J. G. Ferreira, F. R. Martins, *J. Braz. Chem. Soc.* **2006**, *17*, 1683–1688.
[31] P. R. Martínez-Alanis, I. Castillo, *Tetrahedron Lett.* **2005**, *46*, 8845–8848.
[32] C. D. Gutsche, *Calixarenes Revisited*, Royal Society Of Chemistry, **1998**.
[33] K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2008**.
[34] C. G. Bates, R. K. Gujadhur, D. Venkataraman, *Org. Lett.* **2002**, *4*, 2803–2806.
[35] R. A. Talmazan, R. Monroy, F. del Río-Portilla, I. Castillo, M. Podewitz, *ChemCatChem* **2022**, e202200662.
[36] G. Luyckx, J. Ceulemans, *Bull. Soc. Chim. Belg.* **1987**, *96*, 151–163.
[37] L. S. Ettre, *Pure Appl. Chem.* **1993**, *65*, 819–872.
[38] CrysAlisPro, *version 1.171.36.32; Oxford Diffr. Ltd. Abingdon, U. K.* **2013**.
[39] G. M. Sheldrick, *Acta Crystallogr. Sect. A* **2015**, *71*, 3–8.
[40] G. M. Sheldrick, *Acta Crystallogr. Sect. C* **2015**, *71*, 3–8.
[41] L. J. Farrugia, *J. Appl. Crystallogr.* **2012**, *45*, 849–854.
[42] A. L. Spek, *Acta Crystallogr. Sect. C* **2015**, *71*, 9–18.
[43] G. A. Zhurko, D. A. Zhurko, Chemcraft. Version 1.8 (Build 610b). www.chemcraftprog.com.
[44] TURBOMOLE V7.5 **2021**, a) development of the University of Karlsruhe and Forschungszentrum Karlsruhe GmnH, *TURBOMOLE*.
[45] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.
[46] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, **3297–3305**.
[47] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[48] A. Schäfer, A. Klamt, D. Sattel, J. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.
[49] S. K. Goetzfried, C. M. Gallati, M. Cziferszky, R. A. Talmazan, K. Wurst, K. R. Liedl, M. Podewitz, R. Gust, *Inorg. Chem.* **2020**, *59*, 15312–15323.
[50] M. M. Islam, H. Tomiyasu, T. Matsumoto, J. Tanaka, P. E. Georghiou, C. Redshaw, T. Yamato, *Org. Biomol. Chem.* **2015**, *13*, 9055–9064.
[51] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143–152.
[52] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283–289.
[53] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* **1997**, *97*, 119–124.
[54] R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2004**, *6*, 5119–5121.
[55] P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
[56] H. Wallnöefer, S. Handschuh, K. R. Liedl, T. Fox, *J. Phys. Chem. B.* **2010**, *114*, 7405–7412.
[57] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, C. Jin, K. Kasavajhala, M. C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xue, D. M. York, S. Zhao, P. A. Kollman. Amber **2021**, University of California, San Francisco.
[58] S. A. Adelman, J. D. Doll, *J. Chem. Phys.* **1976**, *64*, 2375–2388.
[59] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
[60] G. Ciccotti, J. P. Ryckaert, *Comput. Phys. Rep.* **1986**, *4*, 346–392.
[61] L. C. T. Pierce, R. Salomon-Ferrer, C. A. F. de Oliveira, J. A. McCammon, R. C. Walker, *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
[62] J. Kraml, F. Hofer, P. K. Quoika, A. S. Kamenik, K. R. Liedl, *J. Chem. Inf. Model.* **2021**, *61*, 1533–1538.
[63] L. Schrodinger, The PyMOL Molecular Graphics System. **2015**.
[64] W. Humphrey, A. Dalke, K. Schulten, *J. Molec. Graphics* **1994**, *14*, 33–38.

Manuscript received: September 21, 2022
Revised manuscript received: November 14, 2022
Accepted manuscript online: December 2, 2022

Corrections added on December, 28 2022, after first online publication: addition of 1) VMD in the last sentence of the Experimental Part, together with the corresponding reference [64]; 2) the link to the repository in the Data Availability Statement.

European Journal of Inorganic Chemistry

Supporting Information

Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]arene C–S Coupling

Armando Berlanga-Vázquez, Radu A. Talmazan, Carlos A. Reyes-Mata,
Edmundo G. Percástegui, Marcos Flores-Alamo, Maren Podewitz,* and Ivan Castillo*

Contents

Synthesis of precursors

***p-tert*-butylcalix[8]arene**. This was synthesized according to the literature procedure,[1] from *p-tert*-butylphenol (4.00 g, 0.026 mol), paraformaldehyde (3.60 g, 0.120 mol), and sodium hydroxide (355 mg, 0.008 mol) in xylene, by heating to reflux for 24 h under inert atmosphere. After cooling to room temperature, the solid formed was filtered and washed with cold xylene. The product was obtained as a white crystalline solid in 80% yield (3.40 g); m.p. > 250°C (dec). ^{1}H NMR ($CDCl_3$, 300 MHz): δ 9.62 (s, 8H; OH), 7.17 (m, 16H; Ar), 4.36 (m, 8H; $–CH_2–$), 3.50 (m, 8H; $–CH_2–$), 1.25 (s, 72H; $^{t-}$Bu).

2,9-bis(bromomethyl)-1,10-phenanthroline ($NeoBr_2$). In a round bottom Schlenk flask, neocuproine (1.50 g, 0.007 mol) and three equivalents of recrystallized N-bromosuccinimide (4.00 g, 0.02 mol) were dissolved in CH_3CN and heated to reflux for 18 h under inert atmosphere. After cooling to room temperature and removing of the solvent under reduced pressure, the mixture was washed with diethylether. The solid was neutralized with a saturated solution of $NaHCO_3$. Then, the solid was dissolved in THF and cooled to 0 °C, two equivalents of both $HPO(OEt)_2$ and $^{i-}Pr_2NEt$ were added, the mixture was warmed to room temperature, and stirred for 24 h for debromination. Then the solvent was evaporated under reduce pressure, and the product was purified by column chromatography using CH_2Cl_2 as eluant to obtain the corresponding dibromide in 30% yield (800 mg); m.p. 110–115°C (dec). ^{1}H NMR ($CDCl_3$, 300 MHz): δ 8.26 (d, J = 8.34 Hz, 2H, Ar), 7.92 (d, J = 8.33 Hz, 2H, Ar), 7.81 (s, 2H, Ar), 4.97 (s, 4H, CH_2Br). DART-MS *m/z* = 367 $[NeoBr_2H]^+$.

Characterization of $L^{1,4}$ and [Cu($L^{1,4}$)I]:

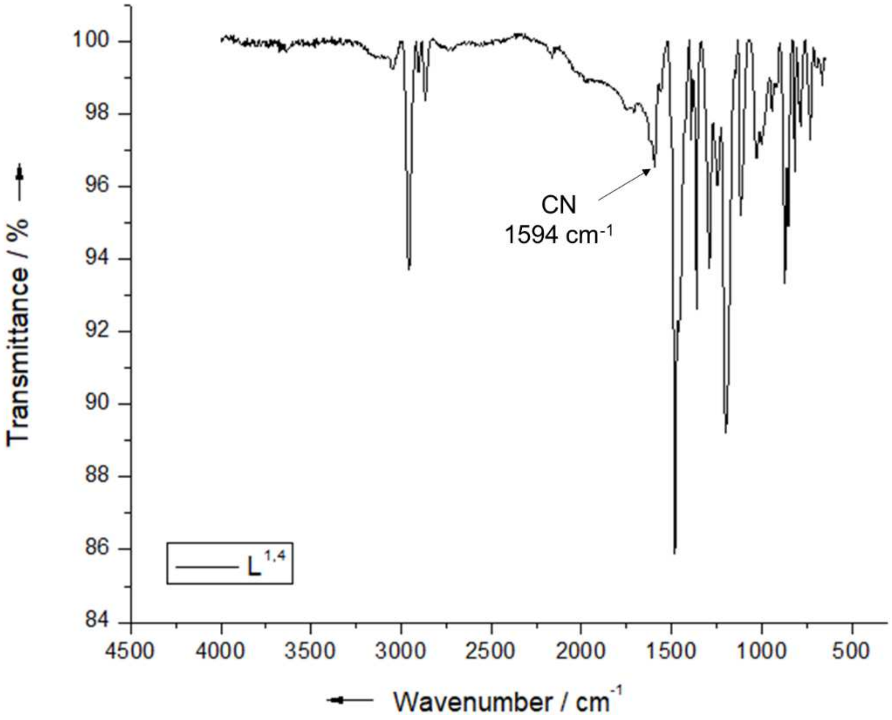


Figure S1. ATR-IR spectrum of $\mathbf{L^{1,4}}$.

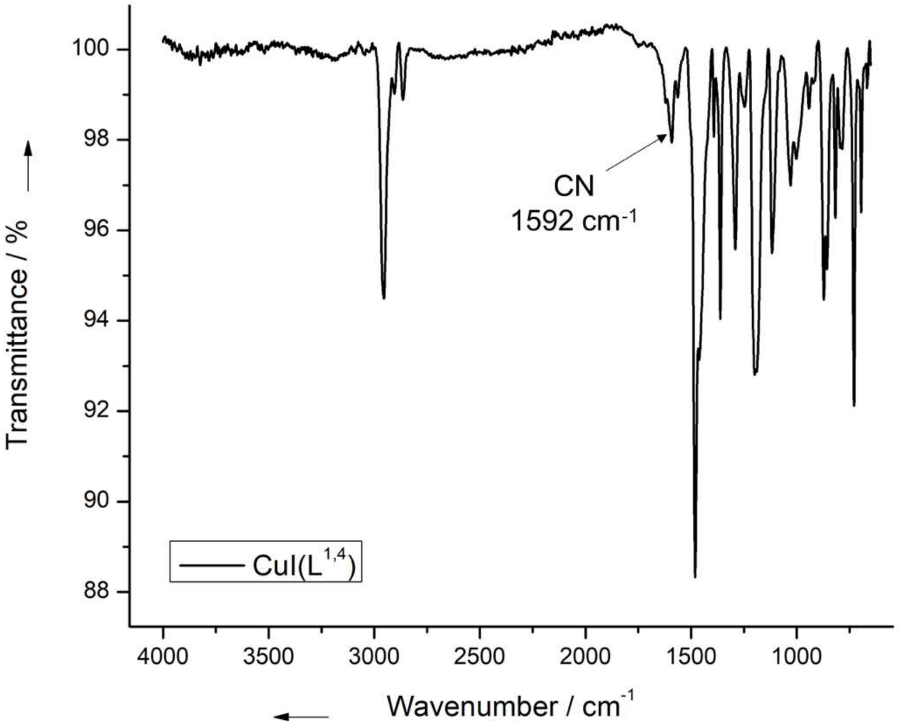


Figure S2. ATR-IR spectrum of **[Cu($\mathbf{L^{1,4}}$)I]**.

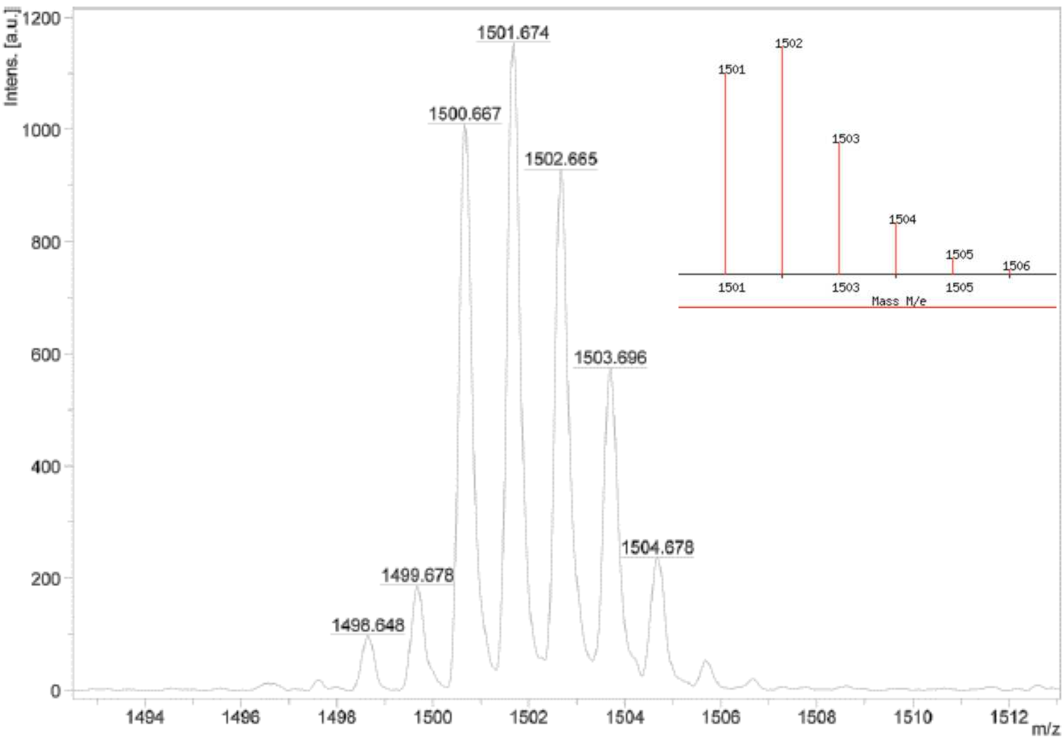


Figure S3. MALDI-TOF MS of $\mathbf{L}^{1,4}$; inset, simulated spectrum of $[\mathbf{L}^{1,4}]^+$.

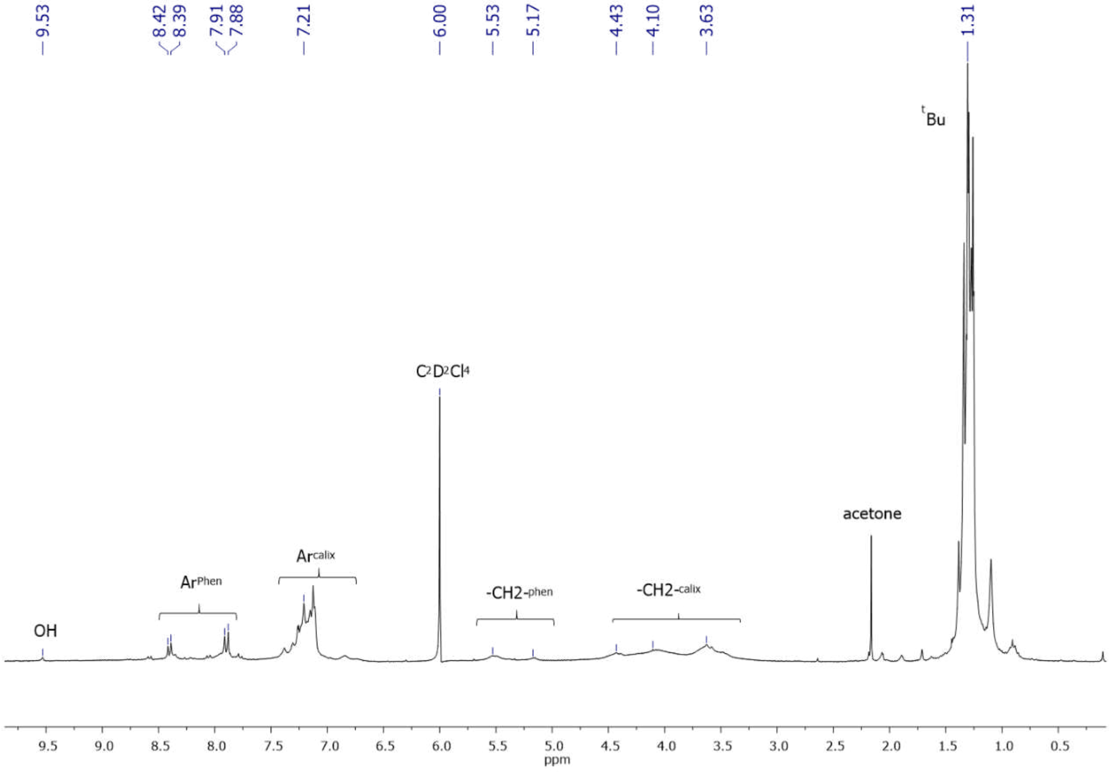


Figure S4. ^{1}H NMR spectrum of $\mathbf{L}^{1,4}$ in $C_2D_2Cl_4$ at room temperature.

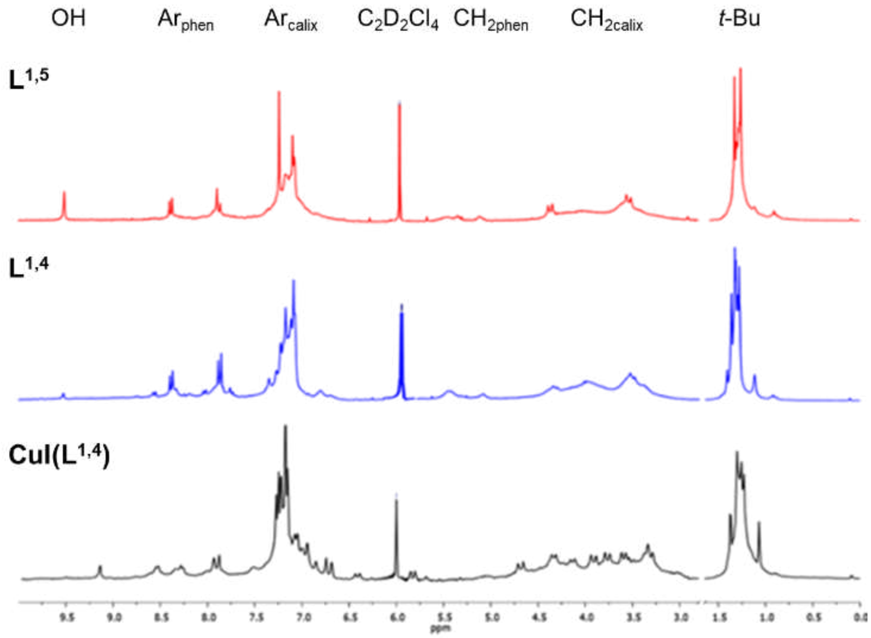


Figure S5. Comparative ^{1}H NMR spectra of **L^{1,5}** (red trace), **L^{1,4}** (blue trace) and **[Cu(L^{1,4})I]** (black trace) in $C_2D_2Cl_4$ at room temperature; assignments labelled on top.

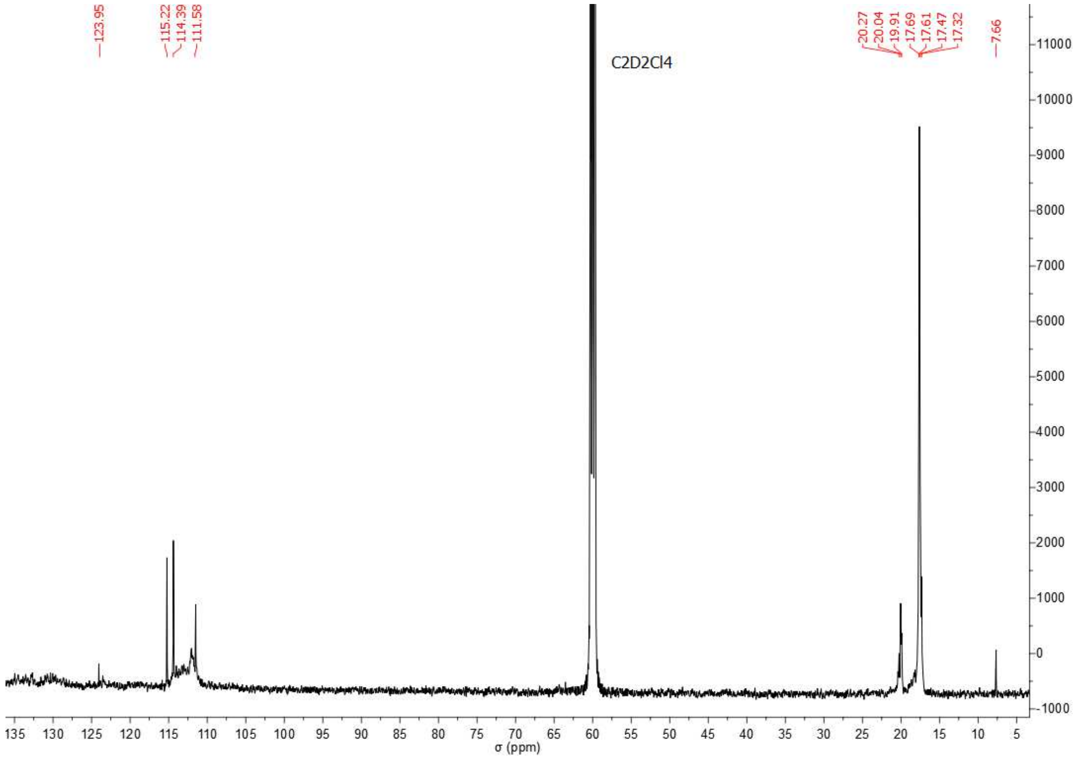


Figure S6. ^{13}C NMR spectrum of **L^{1,4}** in $C_2D_2Cl_4$ collected at room temperature for 12 h.

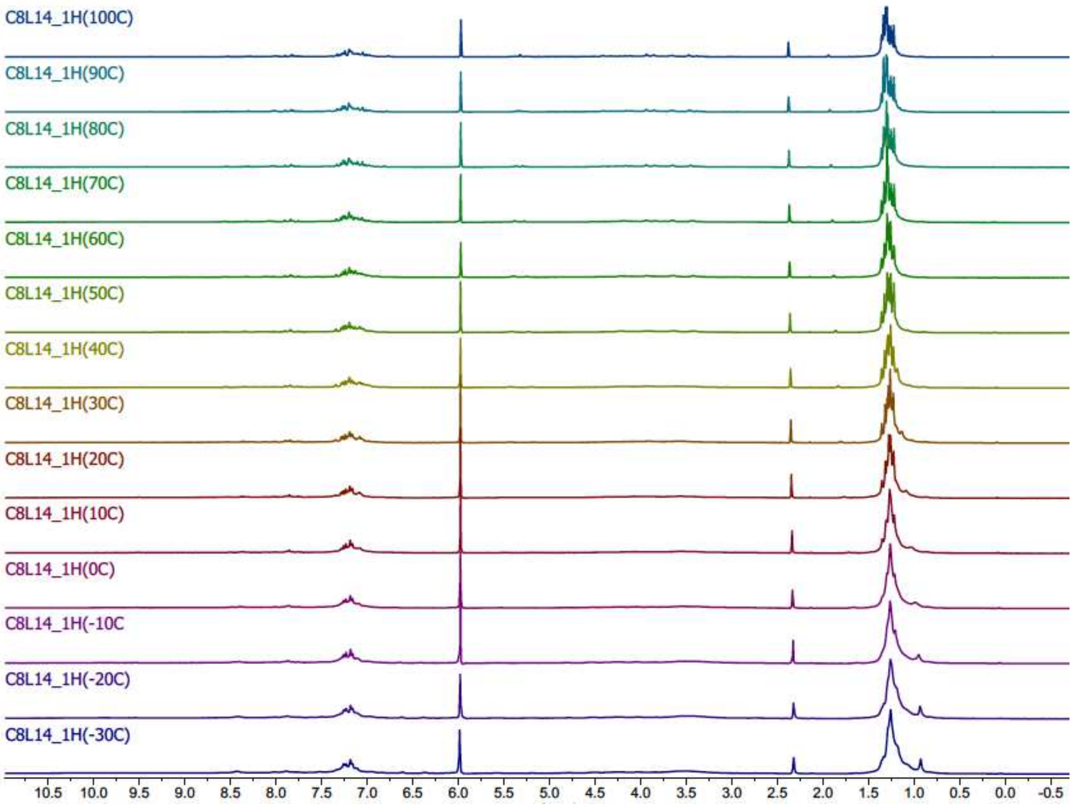


Figure S7. Variable-Temperature ^{1}H NMR spectra of $\mathbf{L}^{1,4}$ in $C_2D_2Cl_4$ (at 6.0 ppm) from -30 (bottom) to 100°C (top).

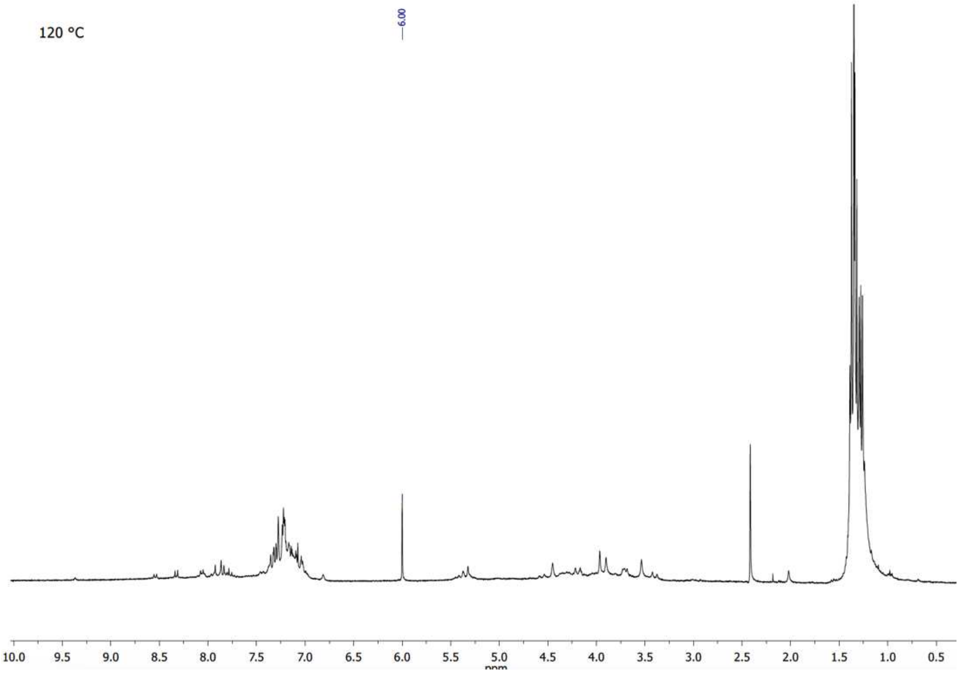


Figure S8. ^{1}H NMR spectrum of $\mathbf{L}^{1,4}$ in $C_2D_2Cl_4$ (at 6.0 ppm) at 120°C.

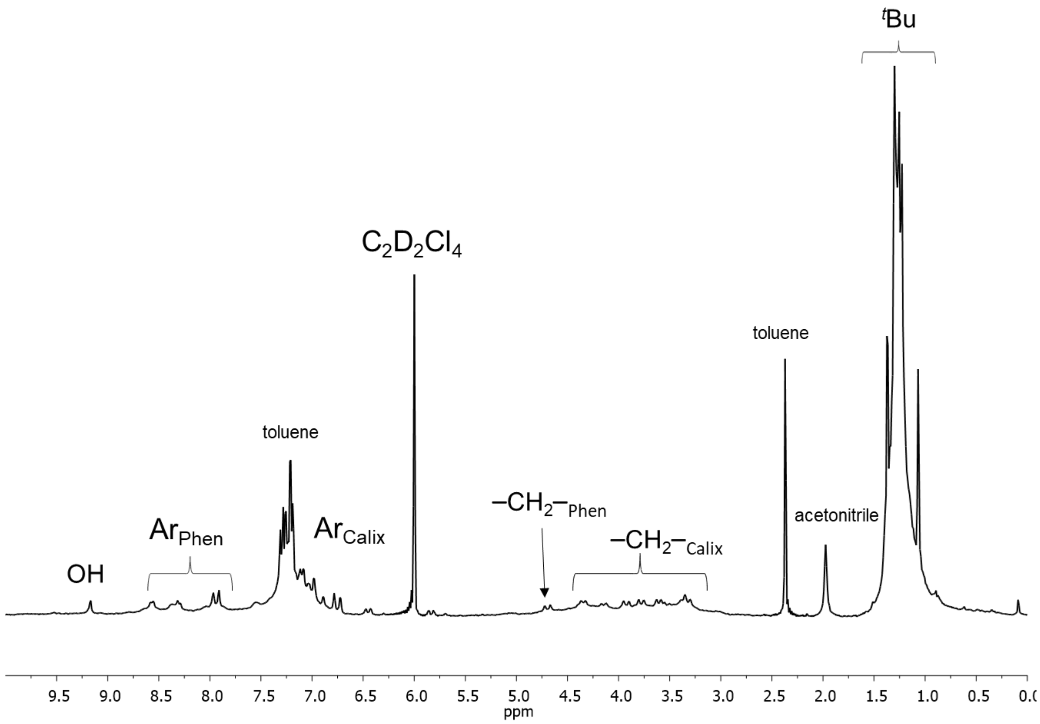


Figure S9. ^{1}H NMR spectrum of **[Cu(L^{1,4})I]** in $C_2D_2Cl_4$ at room temperature.

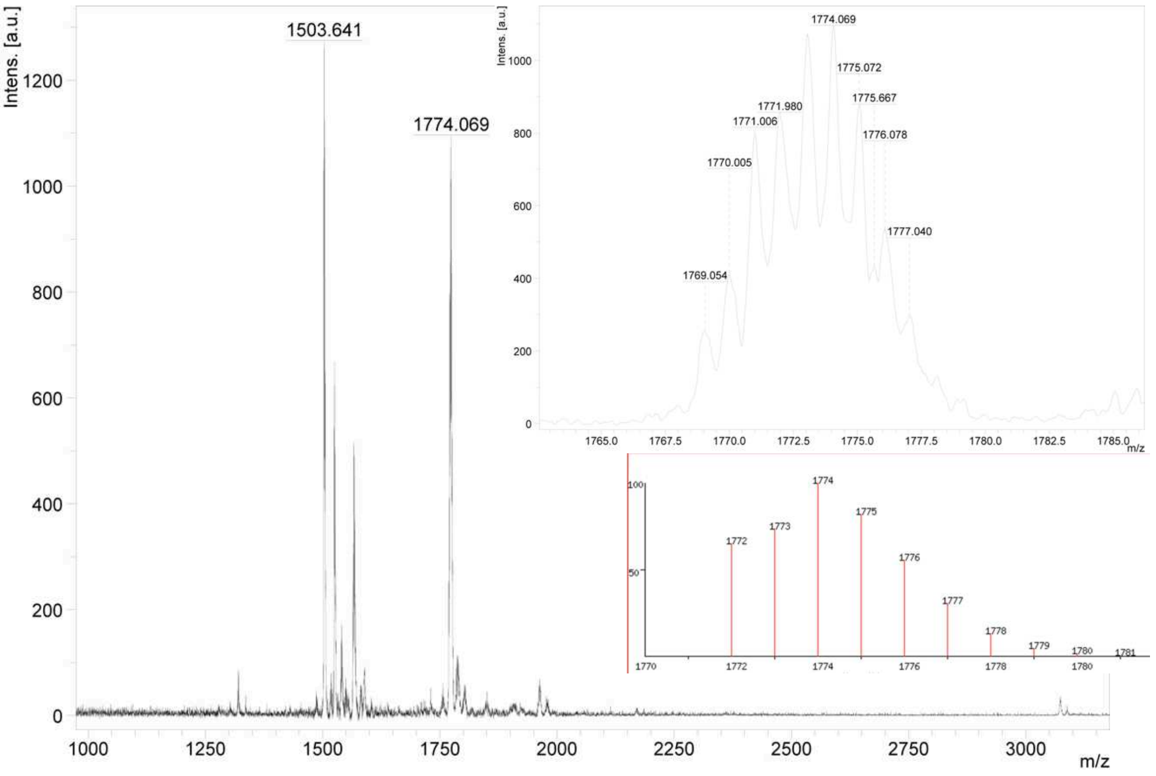


Figure S10. MALDI-TOF MS of [**Cu(L^{1,4})I**]; inset: experimental isotopic distribution of peak at 1774.1 (top) and simulation (bottom).

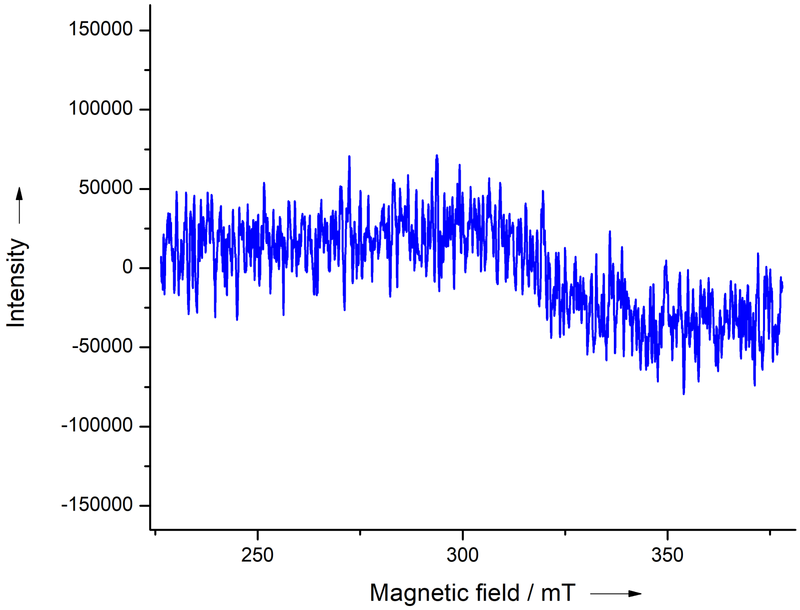


Figure S11. EPR spectrum of **[Cu(L^{1,4})I]**.

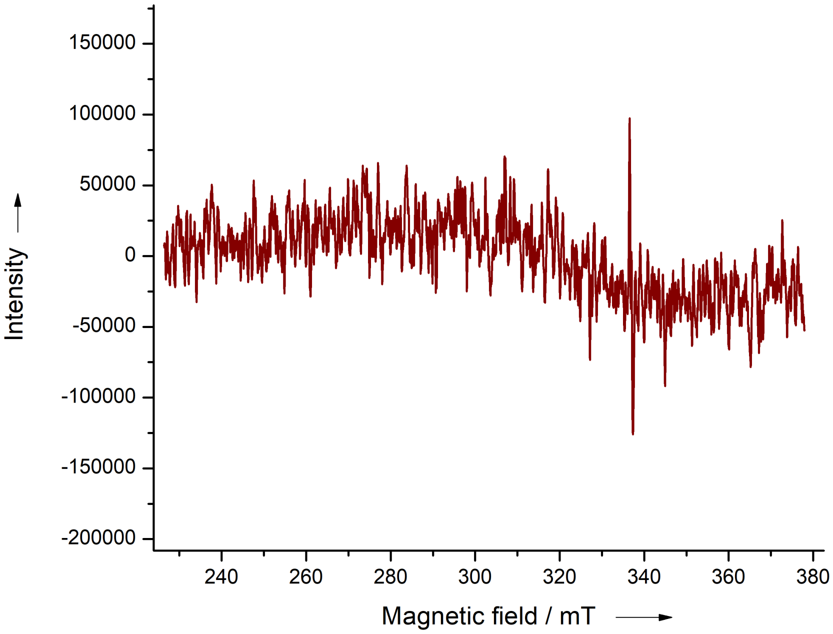


Figure S12. EPR spectrum of **[Cu(L^{1,4})I]** after exposure to air for several days.

Table S1. Crystal data and structure refinement for [$\mathbf{Cu(L^{1,4}\text{-}H)}$($CH_3CN)_2$]

Empirical formula	$C_{110}H_{131}CuN_6O_8$
Formula weight	1728.74
Temperature	170(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	$P\bar{1}$
Unit cell dimensions	a = 19.151(3) Å; α = 101.152(11)°
	b = 19.238(2) Å; β = 109.114(13)°
	c = 20.013(3) Å; γ = 97.228(11)°
Volume	6691.4(17) Å^3
Z	2
Density (calculated)	0.858 Mg/m^3
Absorption coefficient	0.206 mm^{-1}
F(000)	1852
Crystal size	0.290 x 0.230 x 0.120 mm^3
Theta range for data collection	3.450 to 25.350°
Index ranges	$-23 \le h \le 22$, $-23 \le k \le 23$, $-24 \le l \le 24$
Reflections collected	61714
Independent reflections	24351 [R(int) = 0.2146]
Completeness to θ = 25.242°	99.4 %
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	24351 / 901 / 906
Goodness-of-fit on F^2	0.875
Final R indices [I>2 σ(I)]	R1 = 0.1553, wR2 = 0.3266
R indices (all data)	R1 = 0.3139, wR2 = 0.3739
Largest diff. peak and hole	0.422 and -0.327 e.Å^{-3}

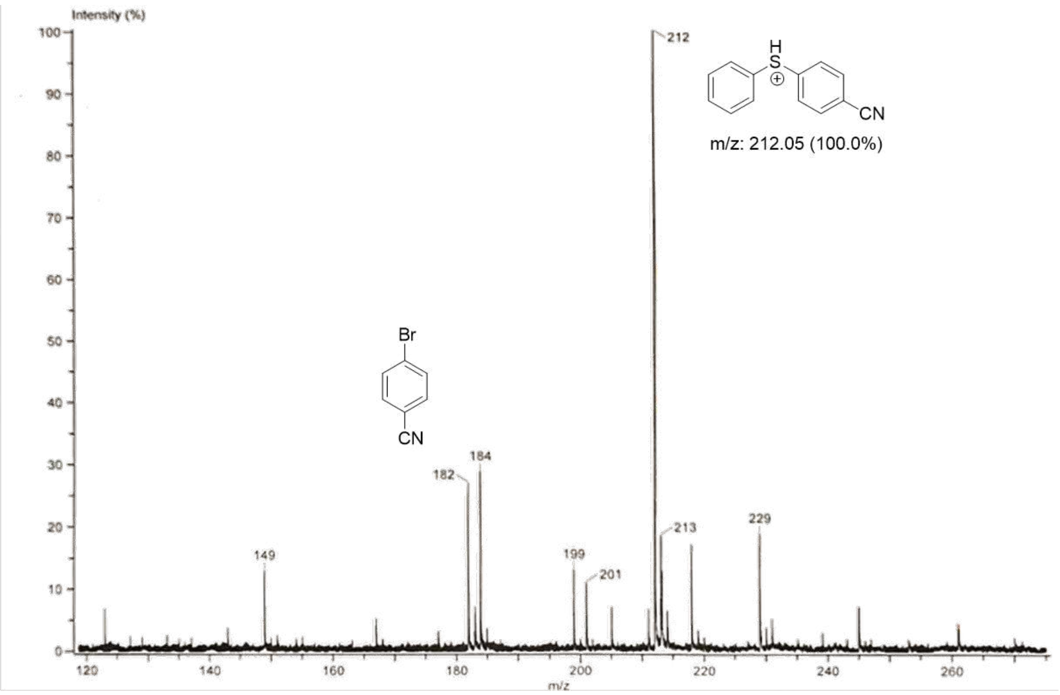


Figure S13. DART-MS of (4-nitrophenyl)phenylsulfide from reaction of thiophenolate with 4-bromobenzenonitrile and **[Cu(L^{1,4})I]** as catalyst.

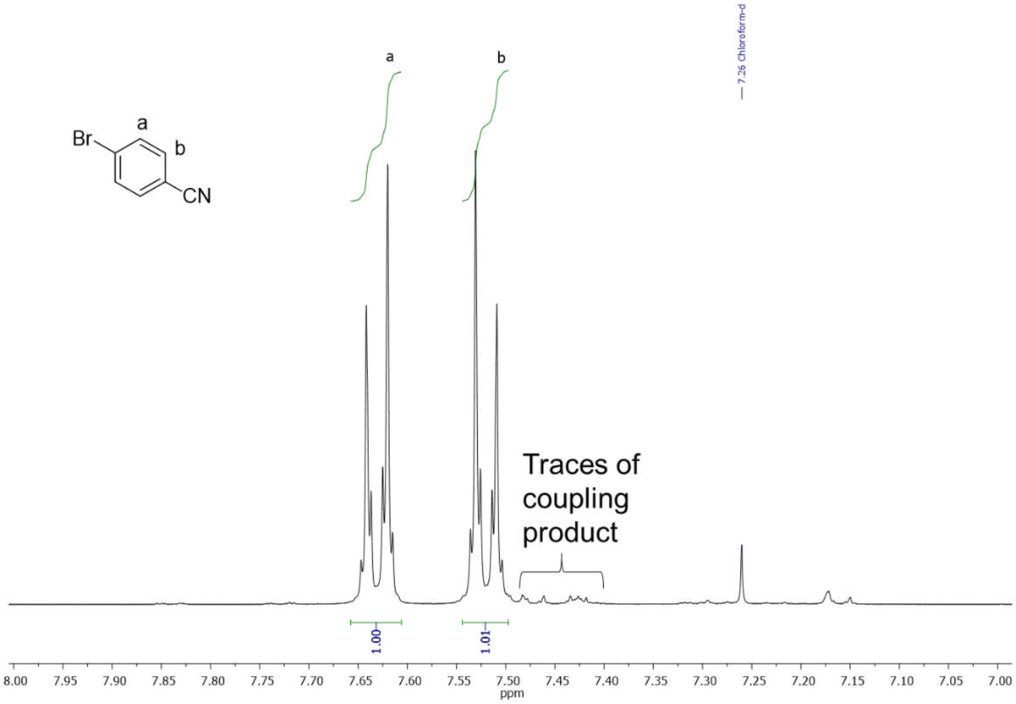


Figure S14. ^{1}H NMR spectrum of 4-bromobenzenonitrile in $CDCl_3$ at room temperature from reaction of thiophenolate with 4-bromobenzenonitrile and **[Cu(L^{1,4})I]** as catalyst.

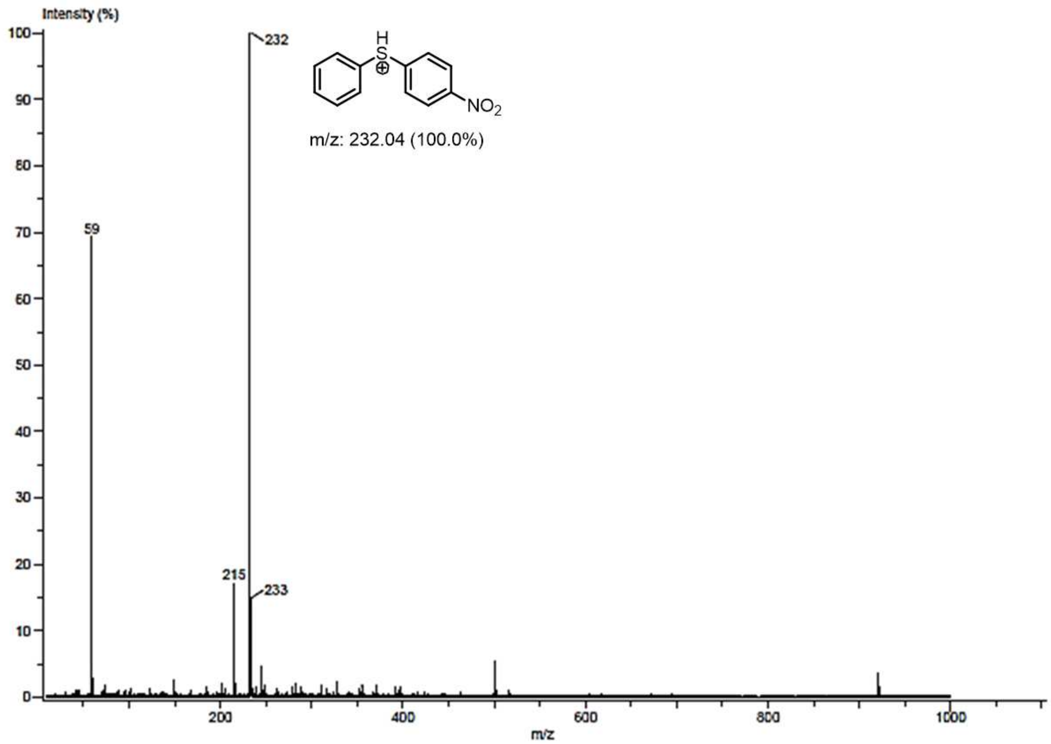


Figure S15. DART-MS of (4-nitrophenyl)phenylsulfide from reaction of thiophenolate with 4-nitrobromobenzene and **[Cu(L^{1,4})I]** as catalyst.

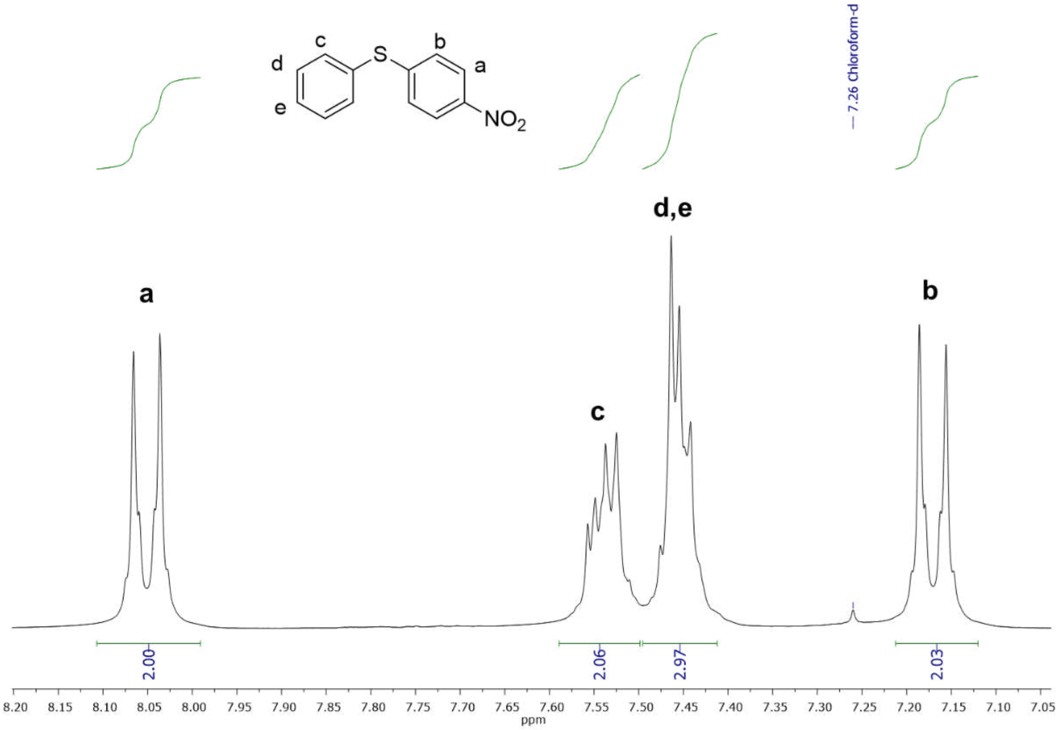


Figure S16. ^{1}H NMR spectrum of (4-nitrophenyl)phenylsulfide in $CDCl_3$ at room temperature from reaction of thiophenolate with 1-bromo-4-nitrobenzene and **[Cu(L^{1,4})I]** as catalyst.

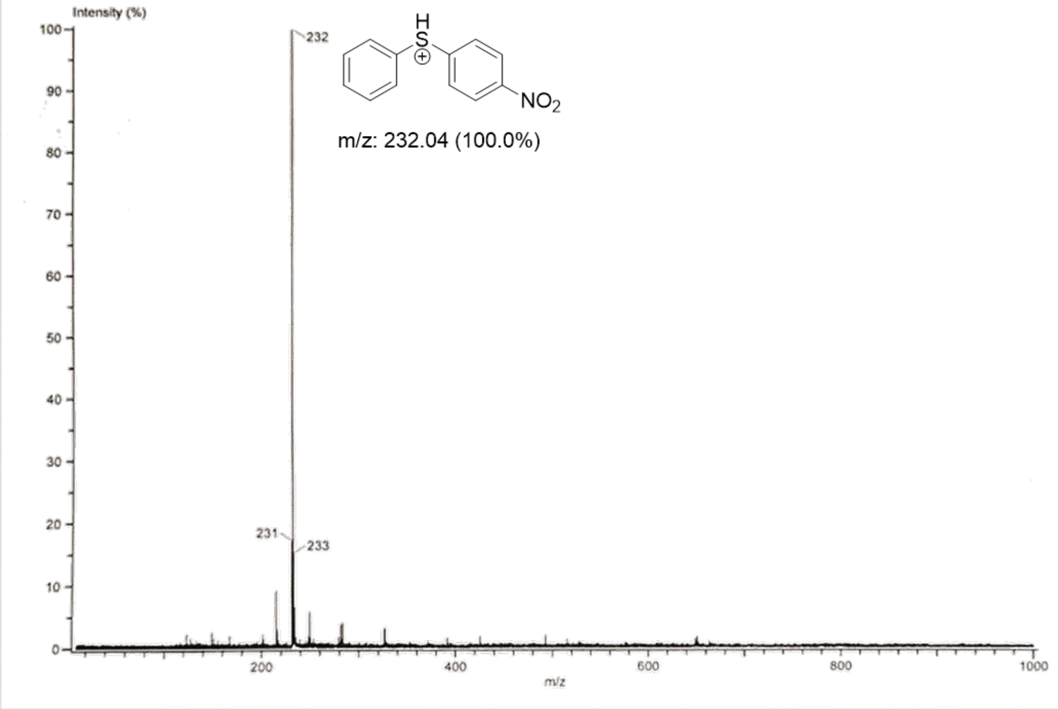


Figure S17. DART-MS of (4-nitrophenyl)phenylsulfide from reaction of thiophenolate with 4-nitroiodobenzene and **[Cu(L^{1,4})I]** as catalyst.

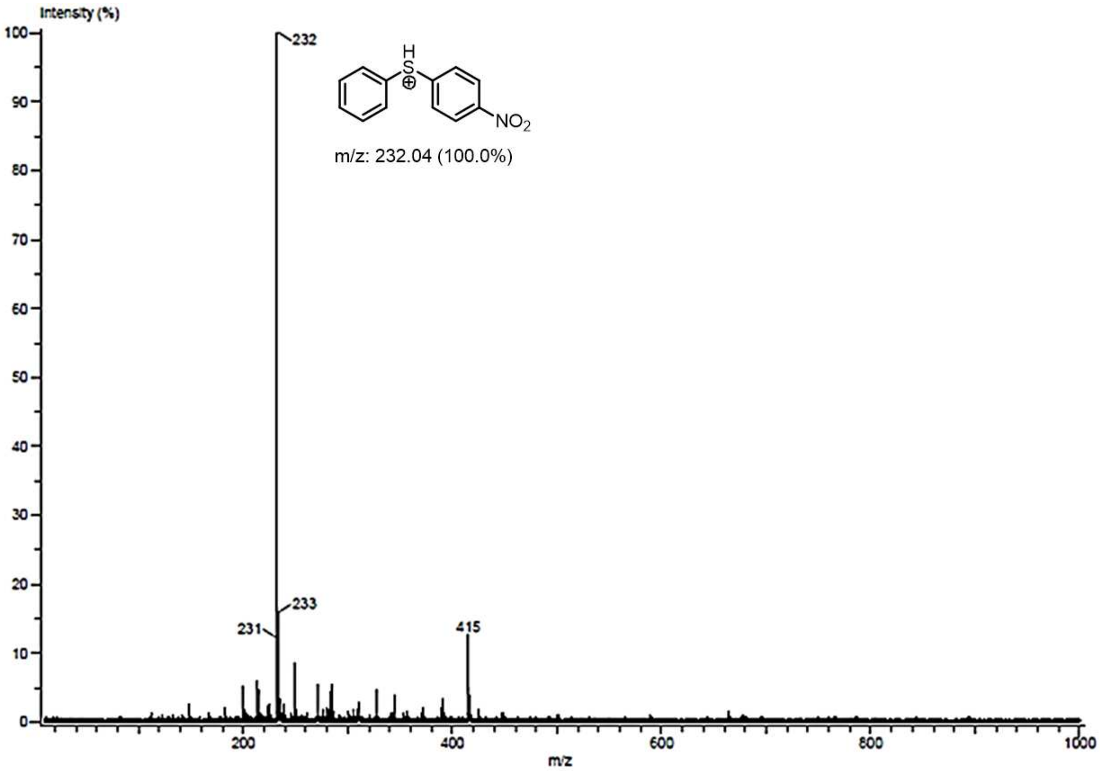


Figure S18. DART-MS of (4-nitrophenyl)phenylsulfide from reaction of thiophenolate with 4-nitrochlorobenzene and **[Cu(L^{1,4})I]** as catalyst.

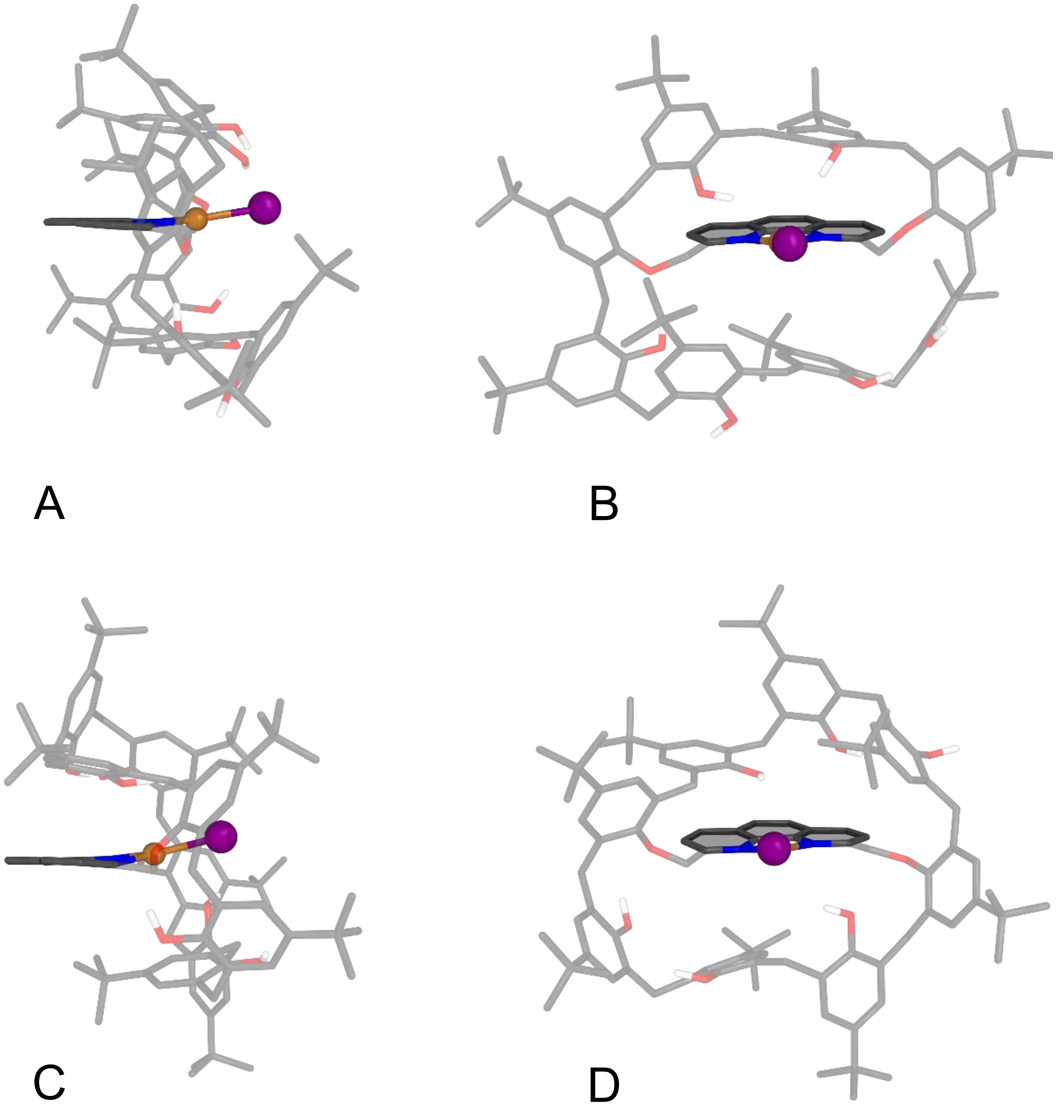


Figure S19. Top: Side (A) and top (B) view of the optimized initial **[Cu($L^{1,4}$)I]** structure with intact cage before sampling. Bottom: Side (C) and top (D) view of the of the optimized **[Cu($L^{1,5}$)I]** structure with intact cage. Optimizations performed with PBE0/def2-SVP/D3 in toluene modelled as implicit solvent with COSMO. Initial structures were taken from the available x-ray crystal structure and modified accordingly.

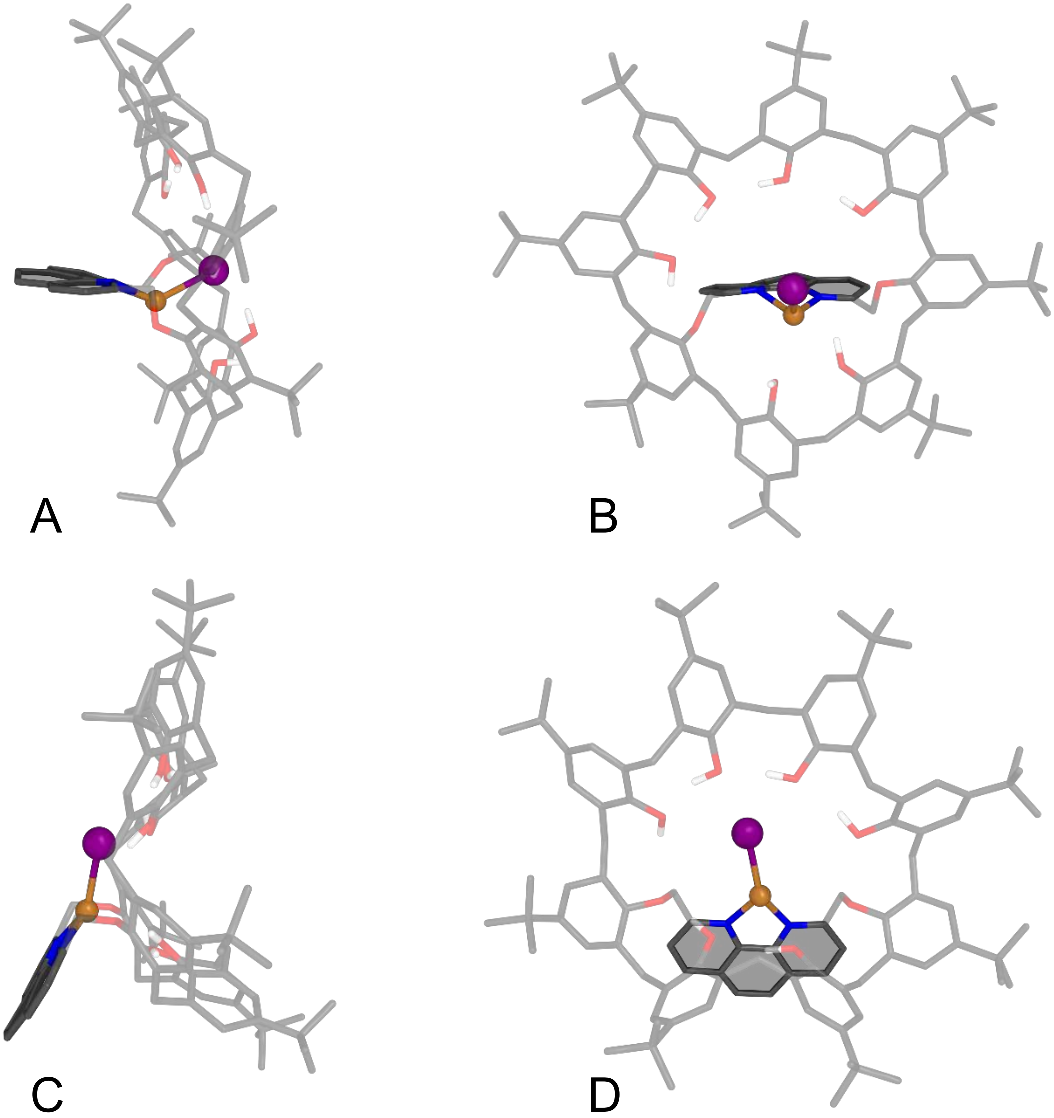


Figure S20. Top: Side (A) and top (B) view of the cluster representative of **[Cu(L^{1,4})I]** complex as obtained from the aMD simulations showing the *intact* cage. Bottom: Side (C) and top (D) view of the of the cluster representative of **[Cu(L^{1,4})I]** as obtained from the aMD simulations showing the *collapsed* cage. The above structures did *not* undergo DFT optimization.

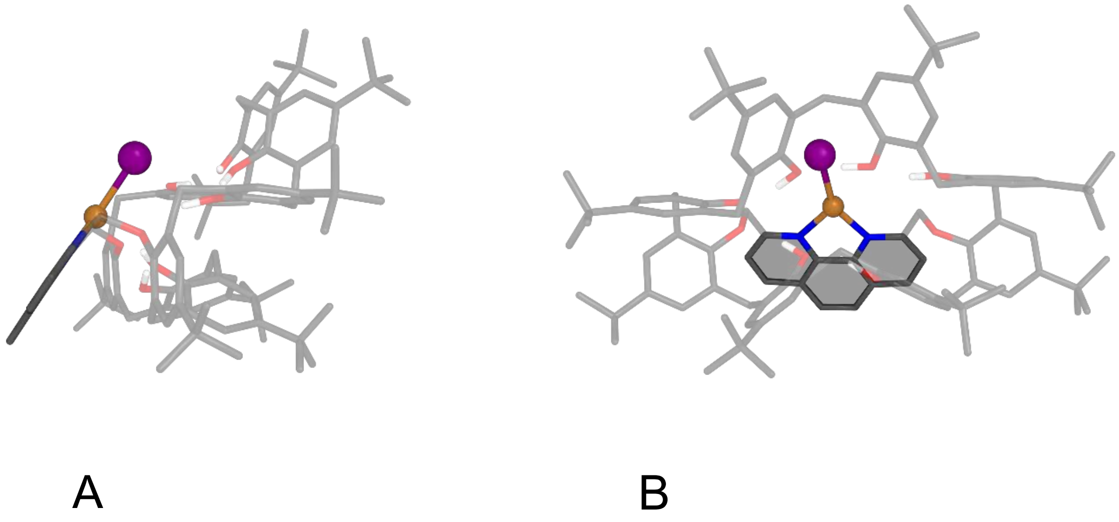


Figure S21. Side (A) and top (B) view of the DFT optimized cluster representative of **[Cu(L^{1,4})I]** structure showing the collapsed cage. Optimizations performed with PBE0/def2-SVP/D3 in toluene modelled as implicit solvent with COSMO.

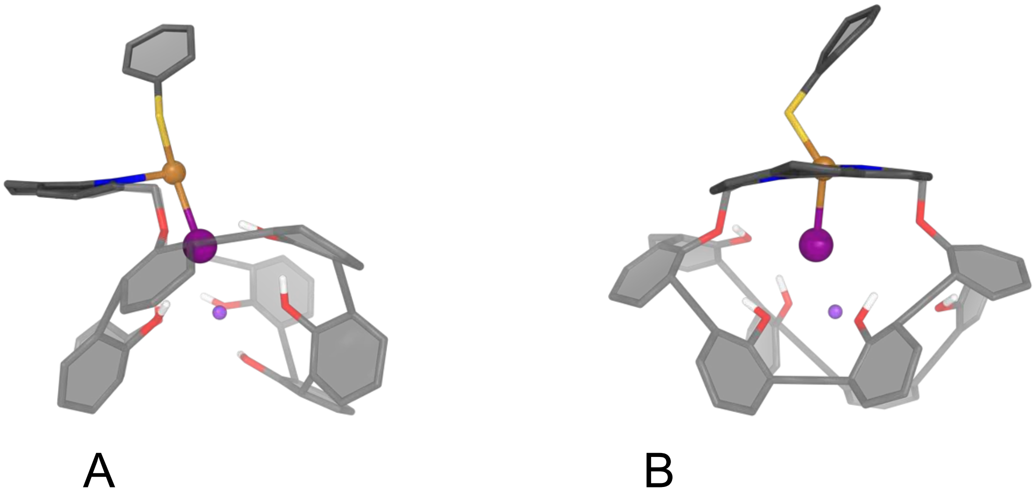


Figure S22. Top: Side (A) and top (B) view of the cluster representative of **[Cu(L^{1,4})I(SPh)]$^{-}$Na^{+}** structure showing the *collapsed* cage. For clarity of the picture, *t*-butyl groups and nonpolar hydrogen atoms were omitted for clarity, calixarene units were directly connected not displaying the connecting carbon atoms.

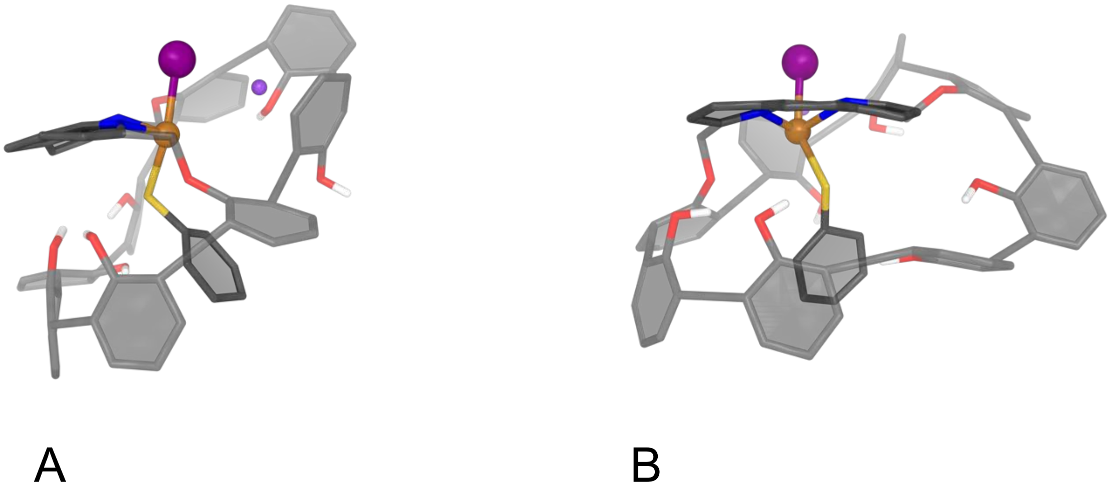


Figure S23. Top: Side (A) and front (B) view of the cluster representative of **$[Cu(L^{1,4})I(SPh)]^{-}Na^{+}$** structure showing the *intact* cage. For clarity of the picture, *t*-butyl groups and nonpolar hydrogen atoms were omitted for clarity, calixarene units were directly connected not displaying the connecting carbon atoms.

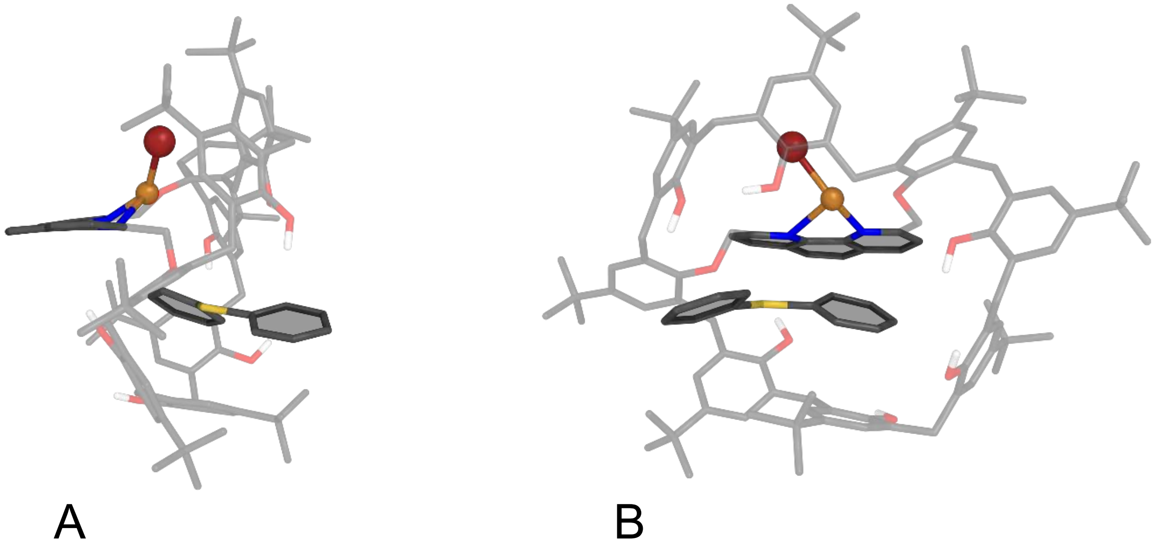


Figure S24. Top: Side (A) and top (B) view of an example of failed optimization of the reaction intermediate **$[Cu(L^{1,4})Br(SPh)(Ph)]$**. The reaction yielded the formation of the coupling product and regeneration of the catalyst. Optimizations performed with PBE0/Def2-SVP/D3 in toluene modelled as implicit solvent with COSMO.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202300094
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

No Transition Metals Required – Oxygen Promoted Synthesis of Imines from Primary Alcohols and Amines under Ambient Conditions

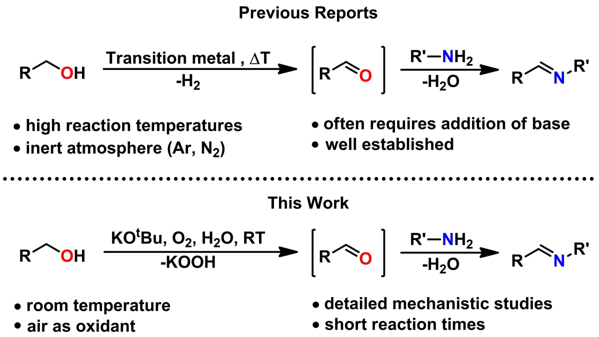
Daniel Himmelbauer+,[a] Radu Talmazan+,[b] Stefan Weber+,[a] Jan Pecak+,[a] Antonio Thun-Hohenstein,[a] Maxine-Sophie Geissler,[a] Lukas Pachmann,[a] Marc Pignitter,[c] Maren Podewitz,*[b] and Karl Kirchner*[a]

Abstract: The synthesis of imines denotes a cornerstone in organic chemistry. The use of alcohols as renewable substituents for carbonyl-functionality represents an attractive opportunity. Consequently, carbonyl moieties can be in situ generated from alcohols upon transition-metal catalysis under inert atmosphere. Alternatively, bases can be utilized under aerobic conditions. In this context, we report the synthesis of imines from benzyl alcohols and anilines, promoted by KO^{t}Bu under aerobic conditions at room temperature, in the absence of any transition-metal catalyst. A detailed investigation of the radical mechanism of the underlying reaction is presented. This reveals a complex reaction network fully supporting the experimental findings.

Introduction

The utilization of alcohols as alkylation surrogates for aldehydes represents a highly desirable reaction route in organic chemistry, since alcohols can be directly obtained from biomass.[1] Therefore, efficient catalytic procedures for in situ generation of carbonyl moieties, based on noble-[2] and non-precious transition metals[3] were described in the last decades by a manifold of researchers, including reports from our group.[4] However, it has been shown that benzyl alcohols can be swiftly oxidized to the corresponding carbonyls upon treatment with a strong base under atmospheric conditions.[5] In fact, various condensation reactions[6] can be carried out with benzyl alcohols in presence of base (sometimes in catalytic amounts) without transition-metal complexes. It was found that the majority of these transformations are driven by alkali metal catalyzed Meerwein–Ponndorf–Verley reduction/Oppenauer oxidation (MPV-O) reactions.[7] A similar transformation represents the synthesis of imines from benzyl alcohols and anilines[8] – a reaction that has been known since the 1950s but seems to be largely overseen in the organometallic's community (Scheme 1). Additionally, Le Berre and co-workers studied the oxidation of the related diphenylmethanol with KO^{t}Bu and oxygen to benzophenone and KOOH, originally reported as KO_2.[9]

In contrast to transition metal catalyzed reactions, only little mechanistic insights are provided so far for this process. We describe here an efficient coupling of alcohols and amines promoted by oxygen at room temperature and propose a



Scheme 1. Transition metal catalyzed synthesis of imines from alcohols and amines (top) and base/oxygen promoted imine synthesis in the absence of a transition-metal catalyst (bottom).

---

[a] Dr. D. Himmelbauer,+ Dr. S. Weber,+ Dr. J. Pecak,+ A. Thun-Hohenstein, M.-S. Geissler, L. Pachmann, Prof. K. Kirchner
Institute of Applied Synthetic Chemistry
TU Wien
Getreidemarkt 9/163-AC, A-1060 Wien (Austria)
E-mail: karl.kirchner@tuwien.ac.at

[b] R. Talmazan,+ Prof. M. Podewitz
Institute of Materials Chemistry
TU Wien
Getreidemarkt 9, A-1060 Wien (Austria)
E-mail: maren.podewitz@tuwien.ac.at

[c] Prof. M. Pignitter
Department of Physiological Chemistry
Faculty of Chemistry
University of Vienna
Althanstrasse 14, 1090 Wien (Austria)

[+] These authors contributed equally to this manuscript.

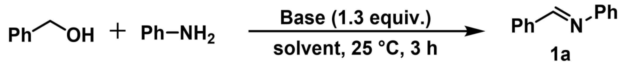
Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202300094

detailed reaction mechanism based on both experimental findings and extensive theoretical studies.

Results and Discussion

In order to examine the impact of various parameters in oxygen-promoted imine formation, optimization reactions were carried out, utilizing benzyl alcohol and aniline as model system. Selected results are depicted in Table 1. At first, different bases were investigated, showing good reactivity for strong bases. In this context, quantitative product formation could be observed for KO^{t}Bu (Table 1, entry 1) at room temperature in toluene. Notably, under the same reaction conditions, only traces of product were formed with KOH. Concerning the solvent, high conversions were detected in toluene and THF. It should be noted that poor reactivity was observed, if the reaction was carried out under argon atmosphere and under UV-irradiation (Table 1, entries 8 and 9). Interestingly, TEMPO can be used as efficient promotor under oxygen-free conditions (Table 1, entry 10). Having established the optimal parameters, scope and limitation of the introduced protocol was explored. In this context, anilines, containing electron-donating or -withdrawing groups, were reacted with benzyl alcohol to give excellent yields (Table 2 entries **1 a**–**1 r**). Furthermore, the condensation of substituted benzyl alcohols with anilines was examined (Table 2 entries **1 s**–**1 ah**), showing high reactivity for most substrates. The introduced protocol tolerates a wide range of functional groups including halides, (thio)ethers, alkenes, acetals, the nitro-group as well as electron-rich or electron-poor heterocycles. However, a massive drop in reactivity was observed for the reaction of aniline with aliphatic 1-octanol (see Table 2 entry **1 ai**).

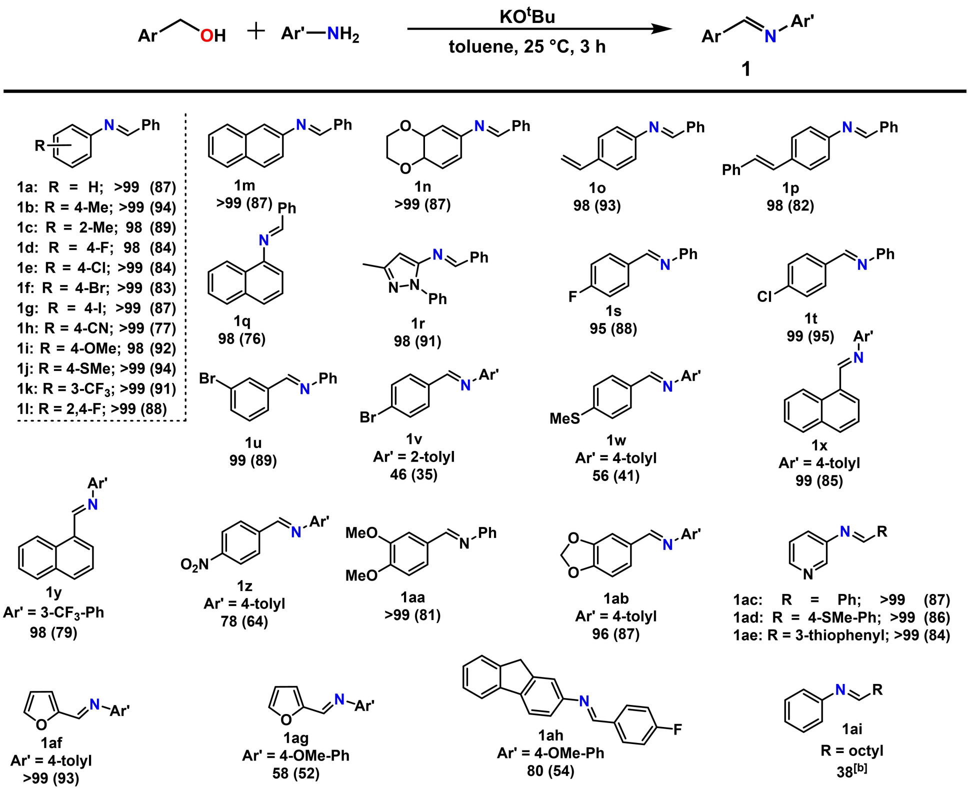
Table 1. Optimization reaction for oxygen-promoted imine formation.[a]



Entry	Base	Solvent	Conversion[b] [%]
1	**KO^{t}Bu**	**toluene**	**>99**
2	KOH	toluene	traces
3	KHMDS	toluene	66
4	NaO^{t}Bu	toluene	50
5	LiO^{t}Bu	toluene	60
6	KO^{t}Bu	THF	>99
7	KO^{t}Bu	MeCN	11
8[c]	KO^{t}Bu	toluene	5
9[c,d]	KO^{t}Bu	toluene	19
10[c,e]	KO^{t}Bu	toluene	82

[a] Conditions: benzyl alcohol (1.0 mmol), aniline (1.2 mmol), base (1.3 mmol), solvent (6 mL), 25 °C, open reaction vessel, 3 h. [b] determined by GC-MS. [c] under Ar. [d] under UV-irradiation. [e] 1 equiv. TEMPO added.

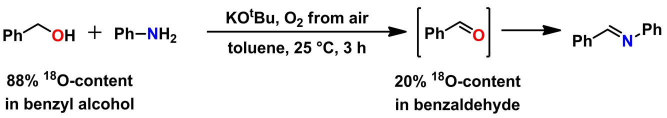
Table 2. Substrate scope of oxygen-promoted imine formation.[a]



[a] Conditions: benzyl alcohol (1.0 mmol), aniline (1.2 mmol), KO^{t}Bu (1.3 mmol), solvent (6 mL), 25 °C, open reaction vessel, 3 h, conversion determined by GC-MS, isolated yield given in parenthesis. [b] 24 h.

Encouraged by the high reactivity of the developed methodology we performed mechanistic investigations, employing benzyl alcohol and aniline as model substrates for all experimental and computational studies. At first, the role of KO^{t}Bu was explored. It should be noted that the purity of KO^{t}Bu and several other substrates was carefully analyzed to exclude traces of potentially (catalytically) active impurities using ICP-MS. We could reproduce all results with various batches of KO^{t}Bu from different suppliers with purity levels from 98% up to $>$99.99%. Furthermore, upon stochiometric addition of 18-crown-6 as potassium scavenger 88% conversion was detected: This suggests that the metal ion is not involved in the rate-determining step, ruling out an alkali metal MPV-O pathway (cf. computational investigation). Next, we investigated whether this transformation is driven by a radical mechanism. Although the addition of TEMPO did not lead to a drop in reactivity, the reaction pathway clearly involves radicals, since the addition of the typical radical trap 3,4-chromanediol completely suppressed all reactivity. KO^{t}Bu seems to simply act as a base, forming the benzylate salt, which then forms a radical with oxygen from the air (see below). In fact, potassium benzylate, PhCH$_2$OK, can be used as starting material, resulting in moderate yield due to high moisture sensitivity of the salt.

The impact of oxygen on the reaction was investigated and full conversion was observed for the model reaction within 15 min under a pure O$_2$ atmosphere. Next, deuterated benzyl alcohol-α,α-D$_2$ was used as substrate. No drop of deuterium content was detected in the product. Surprisingly, if ^{18}O-enriched benzyl alcohol was employed as starting material, a massive drop in ^{18}O-content was observed in the in situ generated benzaldehyde (Scheme 2). This could be attributed to an exchange with oxygen from air or to a scrambling reaction with water (cf. computational investigation).



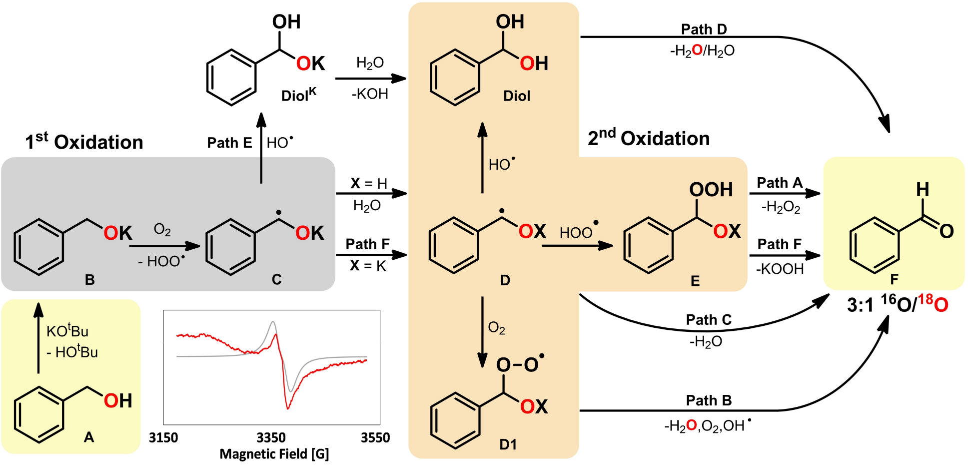
Scheme 2. ^{18}O-Labeling experiment for the base/oxygen promoted synthesis of imines.

Furthermore, we examined whether catalytic amounts of KO^{t}Bu are sufficient for this transformation as described by several researchers under more forcing conditions.[8] If 30 mol% of KO^{t}Bu are used in the model reaction under O$_2$-atmosphere, a turnover number of approx. 2.2 was observed, which proves that the reaction can in general be operated in a catalytic fashion.

To gain further insights into the reaction mechanism, we carried out extensive quantum chemical calculations (DLPNO-CCSD(T)/def2-QZVP//PBE0/def2-TZVP/D4 in toluene,[10] see Supporting Information for details), through which we were able to determine a complex radical pathway. Formation of the benzyl aldehyde can be achieved with and without loss in ^{18}O-labeling, shown in Scheme 3. The reaction mechanism can be separated into several reaction paths starting from path A, which can branch off at various reaction steps resulting in pathways (paths B to F) that depend on intermediates generated in preceding reaction steps. As some of the intermediates are rather short lived, we refrain from quantifying the likelihood of these pathways to occur. Overall, the reaction mechanism can be subdivided into two subsequent hydrogen atom abstraction (HAT) steps.

From the simplified reaction free energy diagram presented in Figure 1 showing relevant intermediates and transition states (for full diagram see Supporting Information Figures S7 and S8), it can be seen that benzyl alcohol **A** reacts with the tert-butyl base to form the potassium salt **B**. This step of the reaction is favored, with a $\Delta G = -5.3$ kcal/mol. Notably, this step can itself be skipped completely as the reaction was experimentally shown to proceed when starting directly from the potassium salt. This also indicates that the tert-butyl base does not play a significant role in reaction steps following **B**.

The salt proceeds to interact with molecular oxygen, which leads to oxidation and the formation of the radical **C** via hydrogen atom abstraction (HAT) depicted in **TS^{B-C}**. The free



Scheme 3. Simplified reaction mechanism for the formation of benzaldehyde from benzyl alcohol through various pathways. The red O-atoms represent the ^{18}O-labeling. The inserted EPR spectrum depicts a radical species (**C** or **D**) observed in experiment.

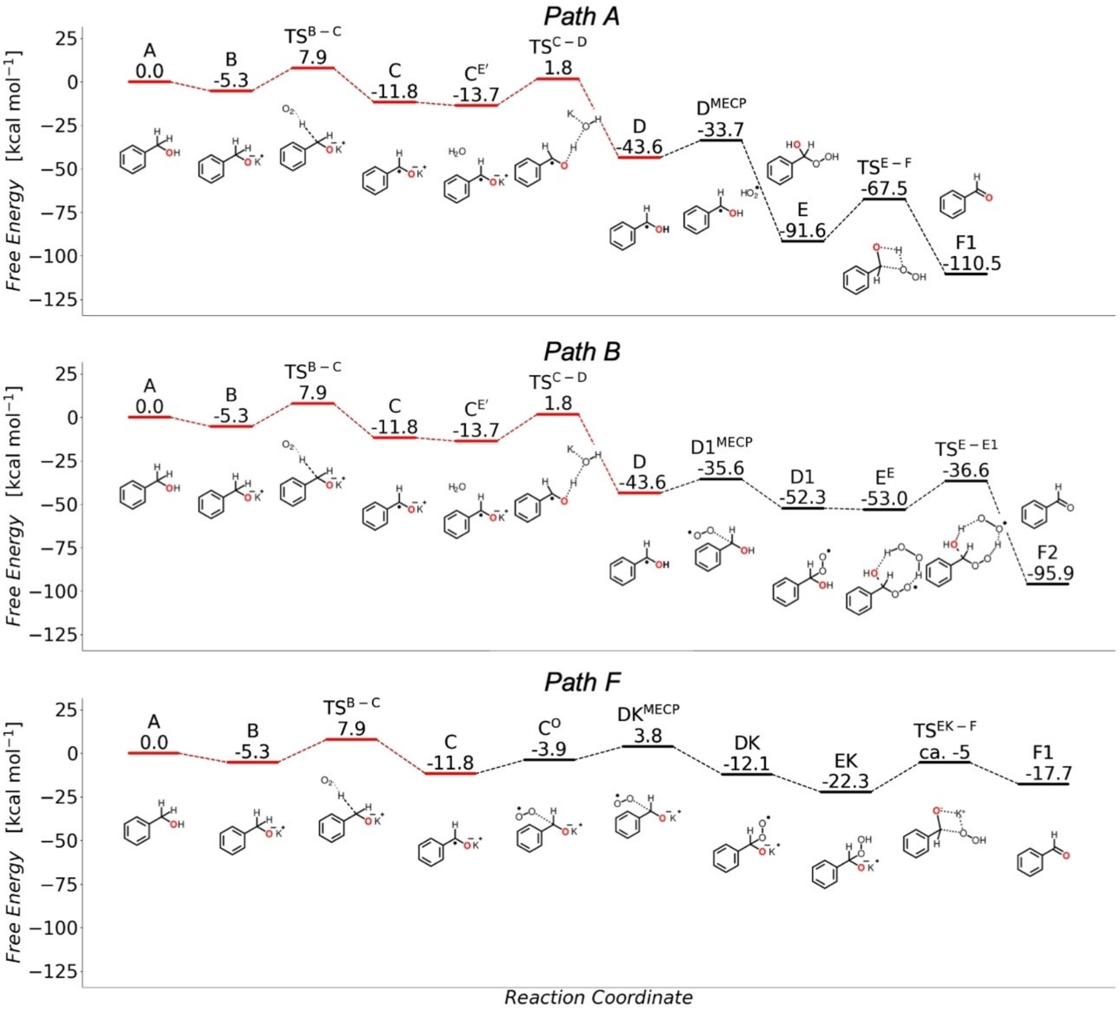


Figure 1. Simplified energy profiles of the various radical reaction pathways to form benzaldehyde. Paths A and B are water assisted, while path F denotes the water-free pathway. The value for the transition state **TS**$^{EK-F}$ is only an estimate, based on reaction path optimization. All values are free energies ΔG, in kcal/mol, calculated with DLPNO-CCSD(T)/def2-QZVP//PBE0/def2-TZVP/D4. The ^{18}O-label is depicted in red. Due to the nature of CCSD, accurate energies are not achievable for the MECP points. In there cases, the energy differences are obtained from B2PLYP/def2-TZVP/D4//PBE0/def2-TZVP/D4 (for more information see computational methodology section and SI).

energy barrier for this step is $\Delta G^{\neq} = 13.2$ kcal/mol, while the formation of the **C** radical itself is favored ($\Delta G = -11.8$ kcal/mol). With the formation of **C**, a rather stable peroxyl radical is liberated. In presence of residual water, the **C**$^{E'}$ complex ($\Delta G = -13.7$ kcal/mol) is formed. The potassium ion assists proton abstraction from water, forming the radical **D**, which is strongly favored, with $\Delta G = -43.6$ kcal/mol. The energy barrier for this step is given by **TS**$^{C-D}$ and has a free energy of $\Delta G^{\neq} = 15.5$ kcal/mol. For both **C** and **D** g_{iso} values were calculated,[11] with $g_{iso,C} = 2.003$ and $g_{iso,D} = 2.004$. A radical species was also detected in EPR spectroscopy ($g_{iso} = 2.001$, see Scheme 3) and our g-tensor calculations indicate that this structure must be either **C** or **D** or a mixture of both (see Figure S9 for details).

In path A, the initiation of the reaction, **D** can undergo a spin crossing event, in the proximity of the peroxyl radical generated previously, with an energy barrier of $\Delta G^{\neq} = 9.9$ kcal/mol for **D**^{MECP}, resulting in the formation of **E** ($\Delta G = -91.6$ kcal/mol). Such hydroperoxide species are found to be stable for similar compounds.[12] **E** then undergoes a scrambling reaction with an energy barrier of $\Delta G^{\neq} = 24.1$ kcal/mol, through **TS**$^{E-F}$, resulting in the final product, **F1**. In path A, the initial ^{18}O-labeled oxygen of the benzyl alcohol is retained.

From structure **D**, the reaction can also proceed in the presence of environmental O_2. Following path B (see Figure 1) the reaction with oxygen results in a spin crossover event, with a barrier of $\Delta G^{\neq} = 8.0$ kcal/mol for **D1**^{MECP}, culminating in the formation of **D1** (with $\Delta G = -52.3$ kcal/mol). **D1** can then re-enter path A, propagating the radical reaction by HAT from **B** under generation of **E** with a free energy of this reaction step of $\Delta G = -21.9$ kcal/mol (not shown in diagram). Alternatively, **D1** can proceed via **E**^{E} ($\Delta G = -53.0$ kcal/mol), provided that the H_2O_2 generated during the path A is in proximity. Continuing from **E**^{E}, the structure undergoes a series of rearrangements, with the highest energy point represented by **TS**$^{E-E1}$, at $\Delta G = -36.6$ kcal/mol. The rate determining barrier for these steps is

$\Delta G^{\neq} = 16.4$ kcal/mol, the second HAT. In the final product of the reaction, **F2**, the ^{18}O-labeled oxygen atom is substituted.

Moreover, in the presence of a hydroxyl radical (generated in path B), the additional path C and path D branching off from **D** are also possible (see Figure S8). Depending on the orientation of the two radicals in the encounter structure, path C can be followed, through the **F1**$^{\text{MECP}}$ spin crossover point (second HAT). This presents a barrier of $\Delta G^{\neq} = 10.8$ kcal/mol and results in the **F1** product, as it transitions into the singlet spin state. Alternatively (path D), a different orientation of the hydroxyl radical pushes the structure into forming a diol. The spin transition barrier for **Diol**$^{\text{MECP}}$ is $\Delta G^{\neq} = 11.5$ kcal/mol. Once on the singlet surface, **Diol** is created ($\Delta G = -36.7$ kcal/mol). This is subsequently followed by a scrambling reaction with a barrier of $\Delta G^{\neq} = 18.5$ kcal/mol, that has an equal probability of resulting in **F1** (labelled benzaldehyde) or **F2** (non-labelled benzaldehyde). The cumulative barrier for this path is $\Delta G^{\neq} = 25.4$ kcal/mol.

Another possibility for branching exists in **C** (path E), provided that an OH radical is present (see Figure S8). The spin crossover **C**$^{\text{MECP}}$ that takes place has a barrier of $\Delta G^{\neq} = 15.1$ kcal/mol and results in **Diol**$^{\text{K}}$ ($\Delta G = -36.1$ kcal/mol), the K^+ salt analogue of **Diol**. This structure can then react in the presence of water to result in the formation of **Diol** with a barrier for this reaction step of $\Delta G^{\neq} = 17.2$ kcal/mol (**TS**$^{\text{C-Diol}}$). Afterwards the **Diol** intermediate re-enters path D via **TS**$^{\text{Diol-F}}$ (scrambling) to yield **F1** and **F2** with equal probability.

Many alternate paths towards the product **F** starting from each intermediate were attempted, but the transition states were found to be either very high or calculations converged to previously known intermediates. In our proposed mechanism water plays a crucial role to form (neutral) low energy intermediates and transition states. In contrast, the initial investigation by Le Berre hinted at a water-free reaction but was missing mechanistic details.[9] Nevertheless, we modelled this pathway in analogy to our hitherto determined mechanism. In this reaction cascade (path F), **C** interacts with O_2 forming the potassium analogue of **D1**, denoted as **DK**, which is with $\Delta G = -12.1$ kcal/mol significantly less stable than the potassium free counterpart **D1** ($\Delta G = -52.3$ kcal/mol). Likewise, the barrier for this transformation was estimated to be around 16 kcal/mol compared to $\Delta G^{\neq} = 13.6$ kcal/mol for the water-assisted pathway. Radical propagation mediated by **DK** resulting in **EK** is also less favorable than in the water-assisted pathway ($\Delta G = -10.2$ kcal/mol vs. $\Delta G = -21.9$ kcal/mol), while the final formation of benzaldehyde is even exergonic with $\Delta G = +4.6$ kcal/mol. The relative stability of **F** in the water-free pathway is thereby only $\Delta G = -17.7$ kcal/mol as opposed to $\Delta G = -110.5$ kcal/mol in the water-assisted pathway. Although the water-free Le Berre mechanism is feasible, it is thermodynamically less favorable. In addition, we would like to note here that the consecutive reaction with amine leads to stochiometric amounts of water.

The termination of the reaction can happen via paths C, D or E, as well as through the decomposition of H_2O_2. This can be generated in path A, or by the combination of two OH radicals formed via path B.

A quantitative estimation for the ratio of **F1** to **F2** is very challenging, because several of the pathways depend on the presence and proximity of short-lived radical species. From a kinetic perspective, path B presents the lowest overall barriers, $\Delta G^{\neq}_{\text{max}} = 16.4$ kcal/mol, which suggest that given the right environment, **F2** would be produced in higher amounts – a finding in line with experimental observations. Still, this reaction pathway is kinetically competing with the water-free pathway, for which a rate-determining barrier of ca. 17 kcal/mol is estimated.

The fast condensation to the final imine, from the amine and the in situ generated benzaldehyde proceeds via a well-known exergonic C–N coupling pathway with a reaction barrier of $\Delta G^{\neq} = 16$ kcal/mol.[13] The released water promotes the main water-assisted reaction pathway – accelerating benzaldehyde formation. The experimentally determined drop in ^{18}O-content can therefore originate from exchange with oxygen (path B) or from scrambling with water (paths D and E). Alternatively, ^{18}O-labeled benzaldehyde can undergo base mediated exchange with water, as known from fundamental organic chemistry. However, this exchange has to compete with the rapid imine formation.

To determine the extent to which each reaction pathway is followed, an exhaustive microkinetic modelling would have to be performed, which is beyond the scope of this paper.

In conclusion, we report the first concise reaction mechanism of the transition metal free in situ formation of benzaldehyde under ambient conditions. Our combined experimental and computational studies unambiguously showed that dioxygen (and water) is essential for this complex radical process. We could show that KO^{t}Bu acts solely as a deprotonating agent, that the concentration of O_2 is critical for a rapid conversion and that the presence of a transition metal catalyst is not required.

The overall mechanism can be decomposed into two separate HAT reactions that eventually lead to oxidation of benzyl alcohol to benzaldehyde. EPR spectroscopy was able to confirm the presence of a radical intermediate and corroborate theoretical calculations. The forgoing of expensive transition-metal catalysts in the formation of aldehydes from alcohols is highly desirable. Thus, the detailed mechanism provides starting points for future investigations and optimizations of this and similar systems.

Experimental Section

General protocol for N-Alkylation: KO^{t}Bu (146 mg, 1.3 mmol) was put into a 10 mL flat-bottom vial. Then a toluene (6 mL) solution of the aniline derivative (110 μL, 1.2 mmol) and alcohol (104 μL, 1.0 mmol) was added. The reaction mixture was stirred for 3 h at room temperature, while exposed to ambient atmospheric conditions. A sample was taken for GC-MS analysis. Afterwards, the reaction mixture was filtered over a pad of Celite, and the solvent was removed under reduced pressure to yield the crude product which was purified via recrystallization or column chromatography.

Computational methodology: Structures were fully optimized using the PBE0 density functional,[10e] the def2-TZVP basis set,[10d] D4 corrections.[10f] Calculations were performed in toluene, modelled as implicit solvent.[10g] Structures were verified to energy minima or transition states by analysis of the Hessian, showing either no or exactly one imaginary frequency that matches the reaction coordinate, respectively. Final electronic energies were calculated with local pair-natural orbital (DLPNO) based singles- and doubles, and perturbative triples coupled cluster theory (CCSD(T)),[10a] while thermodynamic corrections to obtain free energies were calculated with DFT (see above) using the rigid-rotor quasi harmonic oscillator model. Minimum energy crossing points were carried out with a method proposed by Harvey.[14] EPR tensors were calculated with the double-hybrid density functional B2PLYP[15] and gauge-invariant independent atomic orbitals, while spectra were visualized with Easyspin.[16] All calculations were performed with ORCA 5.0.3.[17] Subsequent to the completion of this work, it was made known that ORCA 5.0.3 contained a bug in the implementation of D4, which could occasionally lead to wrong geometries and energies. As this was fixed in orca 5.0.4, we have re-optimized the structures and computed the corresponding DFT energies where applicable. There were no notable discrepancies between the two ORCA versions, in the case of our system.

Acknowledgements

MP would like to thank the Austrian Science Fund (FWF) for financial support (P 33528). KK gratefully acknowledges the financial support of FWF through projects No. P 33016-N and P 32570-N.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available. In the supplementary material of this article, the coordinates of all calculated structures are available in xmol format under https://doi.org/10.5281/zenodo.7626699.

Keywords: DFT · imines · oxidation reactions · radical mechanisms · transition metal free

[1] a) G. Guillena, D. J. Ramon, M. Yus, *Chem. Rev.* **2010**, *110*, 1611–1641; b) K. Barta, P. C. Ford, *Acc. Chem. Res.* **2014**, *47*, 1503–1512.

[2] a) B. Gnanaprakasam, J. Zhang, D. Milstein, *Angew. Chem. Int. Ed.* **2010**, *49*, 1468–1471; *Angew. Chem.* **2010**, *122*, 1510–1513; b) D. Srimani, Y. Ben-David, D. Milstein, *Angew. Chem. Int. Ed.* **2013**, *52*, 4012–4015; *Angew. Chem.* **2013**, *125*, 4104–4107; c) S. Imm, S. Bähn, L. Neubert, H. Neumann, M. Beller, *Angew. Chem. Int. Ed.* **2010**, *49*, 8126–8129; *Angew. Chem.* **2010**, *122*, 8303–8306; d) M. H. S. A. Hamid, J. M. J. Williams, *Chem. Commun.* **2007**, 725–727; e) T. Zweifel, J.-V. Naubron, H. Grützmacher, *Angew. Chem. Int. Ed.* **2009**, *48*, 559–563; *Angew. Chem.* **2009**, *121*, 567–571; f) R. Kawahara, K. Fujita, R. Yamaguchi, *J. Am. Chem. Soc.* **2010**, *132*, 15108–15111; g) B. Blank, S. Michlik, R. Kempe, *Adv. Synth. Catal.* **2009**, *351*, 2903–2911; h) M. A. Esteruelas, V. Lezáun, A. Martínez, M. Oliván, E. Oñate, *Organometallics* **2017**, *36*, 2996–3004.

[3] a) G. Zhang, S. K. Hanson, *Org. Lett.* **2013**, *15*, 650–653; b) K. Paudel, S. Xu, O. Hietsoi, B. Pandey, C. Onuh, K. Ding, *Organometallics* **2021**, *40*, 418–426; c) T. Yan, B. L. T. Feringa, K. Barta, *Nat. Commun.* **2014**, *5*, 5602–5609; d) A. J. Rawlings, L. J. Diorazio, M. Wills, *Org. Lett.* **2015**, *17*, 1086–1089; e) A. Mukherjee, A. Nerush, G. Leitus, L. J. W. Shimon, Y. Ben-David, N. A. Espinosa Jalapa, D. Milstein, *J. Am. Chem. Soc.* **2016**, *138*, 4298–4301; f) R. Fertig, T. Irrgang, F. Freitag, J. Zander, R. Kempe, *ACS Catal.* **2018**, *8*, 8525–8530.

[4] a) M. Mastalir, G. Tomsu, E. Pittenauer, G. Allmaier, K. Kirchner, *Org. Lett.* **2016**, *18*, 3462–3465; b) M. Mastalir, M. Glatz, N. Gorgas, B. Stöger, E. Pittenauer, G. Allmaier, L. F. Veiros, K. Kirchner, *Chem. Eur. J.* **2016**, *22*, 12316–12320; c) M. Mastalir, M. Glatz, E. Pittenauer, G. Allmaier, K. Kirchner, *J. Am. Chem. Soc.* **2016**, *138*, 15543–15546; d) M. Mastalir, M. Glatz, E. Pittenauer, G. Allmaier, K. Kirchner, *J. Am. Chem. Soc.* **2017**, *139*, 8812–8815.

[5] a) G. A. Russell, E. G. Janzen, *J. Am. Chem. Soc.* **1962**, *84*, 4153–4154; b) X. Wang, D. Z. Wang, *Tetrahedron* **2011**, *67*, 3406–341; c) L.-H. Zhou, X.-Q. Yu, L. Pu, *Tetrahedron Lett.* **2010**, *51*, 475–477; d) W. Zhang, M. Liu, H. Wu, J. Ding, J. Cheng, *Tetrahedron Lett.* **2008**, 5336–5338.

[6] a) L. J. Allen, R. H. Crabtree, *Green Chem.* **2010**, *12*, 1362–1364; b) Q. Xu, J. Chen, Q. Liu, *Adv. Synth. Catal.* **2013**, *355*, 697–704; c) Q. Xu, J. Chen, H. Tian, X. Yuan, S. Li, C. Zhou, J. Liu, *Angew. Chem. Int. Ed.* **2014**, *53*, 225–229; d) N. Shao, J. Rodriguez, A. Quintard, *Org. Lett.* **2020**, *22*, 7197–7201; e) C.-Z. Yao, Q.-Q. Li, M.-M. Wang, X.-S. Ning, Y.-B. Kang, *Chem. Commun.* **2015**, *51*, 7729–7732; f) B. C. Roy, I. A. Ansari, S. A. Samim, S. Kundu, *Chem. Asian J.* **2019**, *14*, 2215–2219; g) M. Xiao, X. Yue, R. Xu, W. Tang, D. Xue, C. Li, M. Lei, J. Xiao, C. Wang, *Angew. Chem. Int. Ed.* **2019**, *58*, 10528–10536; *Angew. Chem.* **2019**, *131*, 10638–10646; h) Q. Xu, Q. Li, X. Zhu, J. Chen, *Adv. Snth. Catal.* **2013**, *355*, 73–80; i) C. Wang, C. Chen, J. Han, J. Zhang, Y. Yao, Y. Zhao, *Eur. J. Org. Chem.* **2015**, 2972–2977; j) Q.-Q. Li, Z.-F. Xiao, C.-Z. Yao, H.-X. Zheng, Y.-B. Kang, *Org. Lett.* **2015**, *17*, 5328–5331; k) X.-H. Lu, Y.-W. Sun, X.-L. Wie, C. Peng, D. Zhou, Q.-H. Xia, *Catal. Commun.* **2014**, *55*, 78–82; l) X. Li, S. Li, Q. Li, X. Dong, Y. Li, X. Yu, Q. Xu, *Tetrahedron* **2016**, *72*, 264–272; m) S. Tivari, P. K. Singh, P. P. Singh, V. Srivastava, *RSC Adv.* **2022**, *12*, 35221–35226.

[7] a) V. Polshettiwar, R. S. Varma, *Green Chem.* **2009**, *11*, 1313–1316; b) A. Ouali, J.-P. Majoral, A.-M. Caminade, M. Taillefer, *ChemCatChem* **2009**, *1*, 504–509; c) J. Ballester, A.-M. Caminade, J.-P. Majoral, M. Taillefer, A. Ouali, *Catal. Commun.* **2014**, *47*, 58–62.

[8] a) Y. Sprinzak, *J. Am. Chem. Soc.* **1956**, *78*, 3207–3208; b) J. Xu, R. Zhuang, L. Bao, G. Tang, Y. Zhao, *Green Chem.* **2012**, *14*, 2384–2387; c) R. R. Donthiri, R. D. Patil, S. Adimurthy, *Eur. J. Org. Chem.* **2012**, 4457–4460.

[9] A. Leberre, *Bull. Soc. Chim. Fr.* **1961**, 1543–1549.

[10] a) C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106; b) F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305; c) J. Zheng, X. Xu, D. G. Truhlar, *Theor. Chem. Acc.* **2011**, *128*, 295–305; d) F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065; e) C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170; f) E. Caldeweyher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2017**, *147*, 034112; g) V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

[11] a) A. Hellweg, C. Hattig, S. Hofener, W. Klopper, *Theor. Chem. Acc.* **2007**, *117*, 587–597; b) G. L. Stoychev, A. A. Auer, F. Neese, *J. Chem. Theory Comput.* **2018**, *14*, 4756–4771; c) V. A. Tran, F. Neese, *J. Chem. Phys.* **2020**, *153*, 054105.

[12] J. M. Anglada, R. Crehuet, J. S. Francisco, *Eur. J. Chem.* **2016**, *22*, 18092–18100.

[13] Y. Q. Ding, Y. Z. Cui, T. D. Li, *J. Phys. Chem. A* **2015**, *119*, 4252–4260 .

[14] J. N. Harvey, M. Aschi, H. Schwarz, W. Koch Theor, *Chem. Acc.* **1998**, *99*, 95.

[15] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 34108.

[16] S. Stoll, A. Schweiger, *J. Magn. Reson.* **2006**, *178*, 42–55.

[17] a) F. Neese, *WIREs Comput. Mol. Sci.* **2022**,*12*, e1606; b) F. Neese, F. Wennemohs, U. Becker, C. Riplinger, *J. Chem. Phys.* **2020**, *152*, 224108.

Manuscript received: January 11, 2023
Accepted manuscript online: March 3, 2023
Version of record online: April 11, 2023

Chemistry–A European Journal

Supporting Information

No Transition Metals Required – Oxygen Promoted Synthesis of Imines from Primary Alcohols and Amines under Ambient Conditions

Daniel Himmelbauer, Radu Talmazan, Stefan Weber, Jan Pecak, Antonio Thun-Hohenstein, Maxine-Sophie Geissler, Lukas Pachmann, Marc Pignitter, Maren Podewitz,* and Karl Kirchner*

1. General Information

All reactions were performed under ambient conditions and solvents were used as purchased, if not stated otherwise. All aniline and benzyl alcohol substrates as well as all bases were purchased from commercially available sources like Sigma-Aldrich, Acros Organics or TCI and used without further purification. The deuterated solvents were purchased from Eurisotope and dried over 3 Å molecular sieves. ^{1}H and ^{13}C{^{1}H} spectra were recorded on Bruker AVANCE-250, AVANCE-400, and AVANCE-600 spectrometers. ^{1}H and ^{13}C{^{1}H} NMR spectra were referenced internally to residual protio-solvent, and solvent resonances, respectively, and are reported relative to tetramethylsilane ($\delta = 0$ ppm).

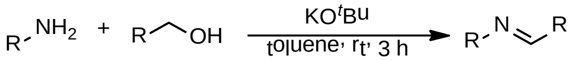
High resolution-accurate mass data mass spectra were recorded on a hybrid Maxis Qq-aoTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) fitted with an ESI-source. Measured accurate mass data of the [M]+ ions for confirming calculated elemental compositions were typically within ±5 ppm accuracy. The mass calibration was done with a commercial mixture of perfluorinated trialkyl-triazines (ES Tuning Mix, Agilent Technologies, Santa Clara, CA, USA).

GC–MS analysis was conducted on an ISQ LT Single quadrupole MS (Thermo Fisher) directly interfaced to a TRACE 1300 Gas Chromatographic systems (Thermo Fisher), using a Rxi-5Sil MS (30 m, 0.25mm ID) cross-bonded dimethyl polysiloxane capillary column.

CW-EPR spectroscopic measurements were performed on an Xband Bruker Elexsys-II E500 EPR spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany) in solution at 100 K. A highsensitivity cavity (SHQE1119) was used for measurements, setting the microwave frequency to 9.86 GHz, the modulation frequency to 100 kHz, the center field to 6000 G, the sweep width to 12000 G, the sweep time to 120 s, the modulation amplitude to 4 G, the microwave power to 15.9 mW, the conversion time to 7.33 ms, and the resolution to 4096 points. The spectra were analyzed using the Bruker Xepr software.

2. Experimental protocols

General protocol for N-Alkylation



KO^{t}Bu (146 mg, 1.3 mmol) was put into a 10 ml flat-bottom vial. Then a toluene (6 ml) solution of the aniline derivative (110 µl, 1.2 mmol) and alcohol (104 µl, 1.0 mmol) was added. The reaction mixture was stirred for 3 h at room temperature, while exposed to ambient atmospheric conditions. A sample was taken for GC-MS analysis with n-dodecane used as internal standard. Afterwards, the reaction mixture was filtered over a pad of Celite, and the solvent was removed under reduced pressure to yield the crude product which was purified and isolated as described below.

Further optimization reactions

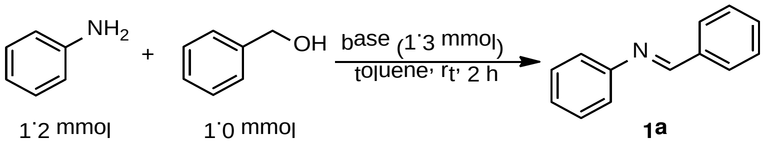


Table S1: Further optimization reactions of aniline with benzyl alcohol.^{a}

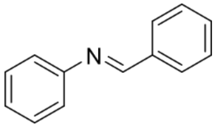
Entry	Base	Conversion [%]	Additive
1	KO^{t}Bu	>99	TEMPO
2^b	KO^{t}Bu	82	TEMPO
3^b	-	2	TEMPO
4	KO^{t}Bu	0	3,4-chromanediol
5^{b,c}	KO^{t}Bu	19	-

^{a} Reaction conditions: Aniline (55 µl, 0.6 mmol), benzyl alcohol (52 µl, 0.5 mmol), KO^{t}Bu (73 mg, 0.65 mmol), solvent (6 ml), additive (0.65 mmol), ambient conditions, room temperature. Conversion of **1a** was determined by GC/MS. ^{b} Exclusion of oxygen. ^{c} Reaction solution was radiated with UV light.

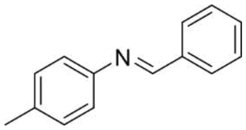
ICP-MS analysis

ICP-MS measurements were carried out on an iCAP Qc quadrupole inductively coupled plasma-mass spectrometer (ICP-MS, Thermo, Bremen), using typical instrumental settings for kinetic energy discrimination mode. Potassium *tert*-butoxide used in all performed and described experiments was carefully analyzed. Samples were diluted with 1% nitric acid solution, prepared from plasma-pure nitric acid and double distilled water. Additionally, a selection of organic substrates was analyzed by LA-ICP-MS. Only ppb traces of transition metals such as Pd or Pt were found.

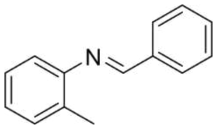
3. Spectroscopic Data for Imine Products



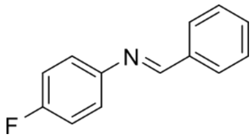
(E)-N-benzylideneaniline (**1a**)[1]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 157 mg (87 % isolated yield) as white solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.48 (s, 1H), 7.92 - 7.90 (m, 2H), 7.49 (dd, J = 5.1 Hz, J = 1.9 Hz 3H), 7.42 - 7.38 (m, 2H), 7.26 - 7.20 (m, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.6, 152.5, 136.8, 131.7, 129.5, 129.1, 129.1, 126.3, 121.2.



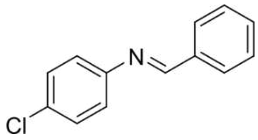
(E)-N-benzylidene-4-methylaniline (**1b**)[2]. The crude product was purified by Al_2O_3 120 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 183 mg (94 % isolated yield) as yellow liquid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.48 (s, 1H), 7.92 - 7.88 (m, 2H), 7.50 - 7.47 (m, 3H), 7.23 - 7.20 (m, 2H), 7.15 - 7.12 (m, 2H), 2.37 (s, 3H, CH_3). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.7, 149.8, 136.9, 136.3, 131.5, 130.1, 129.1, 129.0, 121.1, 21.1.



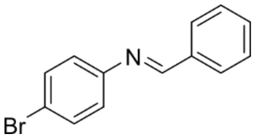
(E)-1-phenyl-N-(o-tolyl)methanimine (**1c**)[3]. The crude product was purified by Al_2O_3 120 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 173 mg (89 % isolated yield) as yellow liquid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.38 (s, 1H), 7.95 - 7.91 (m, 2H), 7.51 - 7.48 (m, 3H), 7.24 - 7.20 (m, 2H), 7.13 (td, J = 7.4 Hz, J = 1.3 Hz, 1H), 6.94 (dd, J = 7.6 Hz, J = 1.1 Hz, 1H), 2.35 (s, 3H, CH_3). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.8, 151.6, 137.0, 132.3, 131.6, 130.6, 129.1, 129.1, 127.1, 126.0, 117.9, 17.9.



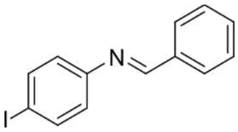
(E)-N-benzylidene-4-fluoroaniline (**1d**)[2]. The crude product was purified by Al_2O_3 120 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 2:1) yielding 167 mg (84 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.46 (s, 1H), 7.91 - 7.89 (m, 2H), 7.50 - 7.49 (m, 3H), 7.23 - 7.20 (m, 2H), 7.12 - 7.08 (m, 2H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 162.8, 160.5, 160.4, 148.6, 136.7, 131.8, 129.1 (d, J = 7.1 Hz), 122.7 (d, J = 8.3 Hz), 116.2 (d, J = 22.6 Hz).



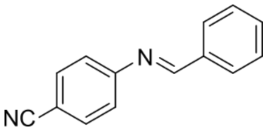
(E)-N-benzylidene-4-chloroaniline (**1e**)[4]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 181 mg (84 % isolated yield) as white solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.45 (s, 1H), 7.92 - 7.89 (m, 2H), 7.52 - 7.47 (m, 3H), 7.39 - 7.36 (m, 2H), 7.19 - 7.15 (m, 2H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 161.1, 151.0, 136.5, 132.0, 131.7, 129.6, 129.2, 129.2, 122.7.



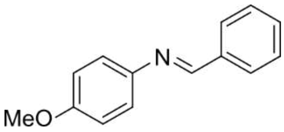
(E)-N-benzylidene-4-bromoaniline (**1f**)[4]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 216 mg (83 % isolated yield) as white solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.45 (s, 1H), 7.92 - 7.89 (m, 2H), 7.54 - 7.48 (m, 5H), 7.13 - 7.09 (m, 2H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 161.1, 151.5, 136.5, 132.6, 132.0, 129.2, 129.2, 123.0, 119.5.



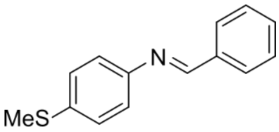
(E)-N-benzylidene-4-iodoaniline (**1g**)[4]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 268mg (87 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.44 (s, 1H), 7.91 - 7.89 (m, 2H), 7.74 - 7.70 (m, 2H), 7.52 - 7.47 (m, 3H). 7.00 - 6.96 (m, 2H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 161.2, 152.2, 138.6, 136.5, 132.0, 129.2, 129.2, 123.4, 90.4.



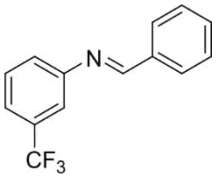
(E)-4-(benzylideneamino)benzonitrile (**1h**)[5].The crude product was purified by recrystallisation in *n*-heptane/Et_2O (1 ml, 1:1) yielding 185 mg (77 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.43 (s, 1H), 7.92 (m, 2H), 7.70 (m, 2H), 7.55 - 7.49 (m, 3H), 7.24 (m, 2H).**^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 162.8, 156.5, 136.1, 133.8, 132.5, 129.5, 129.3, 121.9, 119.3, 109.4.



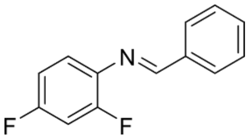
(E)-*N*-benzylidene-4-methoxyaniline (**1i**)[6]. The crude product was purified by Al_2O_3 60 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 2:1) yielding 194 mg (92 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.50 (s, 1H), 7.89 (m, 2H), 7.48 (m, 3H), 7.24 (m, 2H), 6.94 (m, 2H), 3.82 (s, 3H).**^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.8, 158.5, 145.2, 137.0, 131.3, 129.1, 128.9, 122.5, 114.7, 55.9.



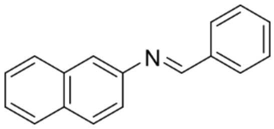
(E)-N-benzylidene-4-(methylthio)aniline (**1j**)[7]. The crude product was purified by recrystallisation in *n*-heptane/toluene (2 ml, 5:1) yielding 214 mg (94 % isolated yield) as brown solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.49 (s, 1H), 7.90 (m, 2H), 7.49 (m, 3H), 7.30 (m, 2H), 7.19 (m, 2H), 2.51 (s, 3H).**^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.9, 149.6, 136.8, 136.5, 131.7, 129.2, 129.1, 127.8, 121.9, 16.4.



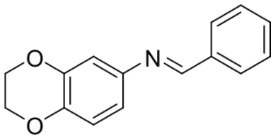
(E)-N-benzylidene-4-(trifluoromethyl)aniline (**1k**)[8]. The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 1:2) yielding 228 mg (91 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.49 (s, 1H), 7.93 (m, 2H), 7.56 – 7.48 (m, 5H), 7.46 (m, 1H), 7.40 (m, 1H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 162.2, 153.1, 136.3, 132.2, 131.7 (d, J = 32.1 Hz), 130.2, 129.4, 129.2, 124.7 124.6 (d, J = 272 Hz), 122.7 (q, J = 3.8 Hz), 118.2 (q, J = 3.8 Hz).



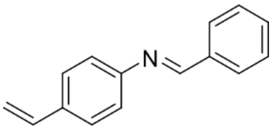
(E)-N-benzylidene-2,4-difluoroaniline (**1l**)[9].The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane) yielding 192 mg (88 % isolated yield) as yellow liquid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.52 (s, 1H), 7.93 - 7.91 (m, 2H), 7.55 - 7.47 (m, 3H), 7.20 - 7.18 (m, 1H), 6.97 - 6.90 (m, 2H).**^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 163.0 (t, J = 2.2 Hz), 162.2 (d, J = 11.2 Hz), 159.8 (d, J = 11.3 Hz), 157.0 (d, J = 12.0 Hz), 154.5 (d, J = 12.3 Hz), 136.8 (dd, J = 10.4 Hz, J = 3.7 Hz), 136.4, 132.2, 129.2 (d, J = 4.9 Hz), 122.5 (dd, J = 9.6 Hz, J = 3.2 Hz), 111.7 (dd, J = 22.1 Hz, J = 4.0 Hz), 104.9 (dd, J = 26.3 Hz, J = 24.3 Hz).



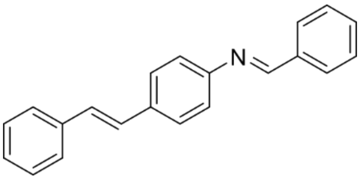
(E)-N-benzylidene-2-napthylamine (**1m**)[10]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 268mg (87 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.62 (s, 1H), 7.99 - 7.97 (m, 2H), 7.91 - 7.87 (m, 3H), 7.63 (m, 1H), 7.54 - 7.45 (m, 6H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.8, 150.0, 136.8, 134.6, 132.4, 131.8, 129.3, 129.2, 129.2, 128.3, 128.1, 126.8, 125.8, 121.5, 118.1.



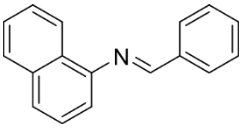
(E)-N-benzylidene-2,3-dihydrobenzo[b][1,4]dioxin-6-amine (**1n**)[11]. The crude product was purified by Al_2O_3 60 (6 g, basic activated) column chromatography (*n*-heptane/toluene/Et_2O 2:1:0.1) yielding 207 mg (87 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.46 (s, 1H), 7.87 (m, 2H), 7.47 (m, 3H), 6.87 (dd, J = 8.3 Hz, J = 0.5 Hz, 1H), 6.81 – 6.77 (m, 2H), 4.27 (s 4H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.9, 146.0, 144.3, 142.8, 136.9, 131.4, 129.1, 128.9, 117.8 114.9, 110.0, 65.0.



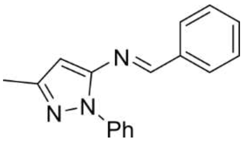
(E)-*N*-benzylidene-4-vinylaniline (**1o**)[12]. The crude product was purified by recrystallisation in *n*-heptane (4 ml) yielding 193 mg (93 % isolated yield) as yellow liquid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.50 (s, 1H), 7.91 (m, 2H), 7.48 (m, 5H), 7.20 (m, 2H), 6.75 (dd, *J* = 17.6 Hz, *J* = 10.9 Hz, 1H), 5.76 (dd, *J* = 17.6 Hz, *J* = 0.9 Hz, 1H), 5.24 (dd, *J* = 10.9 Hz, *J* = 0.9 Hz, 1H). **$^{13}C\{^{1}H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.3, 151.9, 136.8, 136.7, 135.9, 131.8, 129.2, 129.1, 127.4 121.5, 113.5.



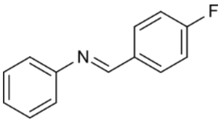
(E)-N-benzylidene-4-aminostilbene (**1p**)[4]. The crude product was purified by recrystallisation in *n*-heptane/toluene (4 ml, 1:1) yielding 232 mg (82 % isolated yield) as gold/brown solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.53 (s, 1H), 7.93 (m, 2H), 7.56 (m, 4H), 7.50 (m, 3H), 7.38 (m, 2H), 6.26 (s, 3H), 7.16 (d, *J* = 2.5 H, 2H). **$^{13}C\{^{1}H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.2, 151.7, 137.8, 136.8, 135.7, 131.8, 129.2, 129.1, 129.1, 128.5, 128.4, 128.0, 127.7, 126.8, 121.8, 53.8.



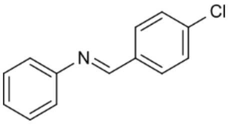
(E)-N-benzylidene-1-napthylamine (**1q**)[2]. The crude product was purified by Al_2O_3 60 (7 g, basic activated) column chromatography (*n*-heptane) yielding 175 mg (76 % isolated yield) as orange oil. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.58 (s, 1H), 8.36 - 8.33 (m, 1H), 8.05 - 8.03 (m, 2H), 7.89 - 7.86 (m, 1H), 7.75 - 7.73 (m, 1H), 7.55 - 7.47 (m, 6H), 7.10 - 7.08 (m, 1H).**$^{13}C\{^{1}H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.8, 149.7, 136.9, 134.4, 131.9, 129.3, 129.3, 129.2 128.0, 126.8, 126.5, 126.1, 126.1, 124.3, 113.0.



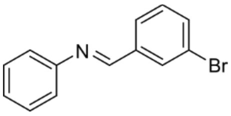
(E)-N-benzylidene-3-methyl-1-phenyl-1H-pyrazol-5-amine (**1r**). The crude product was purified by Al_2O_3 60 (6 g, basic activated) column chromatography (*n*-heptane/toluene/Et_2O 2:1:0.1) yielding 237 mg (91 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.66 (s, 1H), 7.87 (m, 2H), 7.75 (m, 2H), 7.47 (m, 5H), 7.32 (m, 1H), 6.22 (s, 1H), 2.33 (s, 3H). **$^{13}C\{^{1}H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.7, 150.8, 149.5, 136.4, 132.3, 129.4, 129.3, 128.9, 126.6, 124.2, 93.6, 14.3. HR-MS (ESI$^+$, CH_3CN/MeOH + 1 % H_2O): m/z calcd for $C_{17}H_{16}N_3$ $[M+H]^+$ 262.1339, found 262.1331.



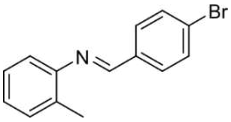
(E)-N-(4-fluorobenzylidene)aniline (**1s**)[2]. The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 185 mg (88 % isolated yield) as white solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.44 (s, 1H), 7.92 (m, 2H), 7.40 (m, 2H), 7.26 – 7.16 (m, 5H). **$^{13}C\{^{1}H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 165.1 (d, *J* = 251 Hz), 159.1, 152.3, 133.2 (d, *J* = 3.2 Hz), 131.1 (d, *J* = 8.8 Hz), 129.6, 126.4, 121.2, 116.2 (d. *J* = 21.9).



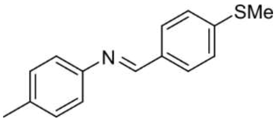
(E)-N-(4-chlorobenzylidene)aniline (**1t**)[4]. The crude product was purified by recrystallisation in *n*-heptane (2 ml) yielding 204 mg (95 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.45 (s, 1H), 7.87 (m, 2H), 7.47 (m, 2H), 7.40 (m, 2H), 7.27 – 7.20 (m, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.1, 152.1, 137.5, 135.4, 130.3, 129.6, 129.4, 126.5, 121.2.



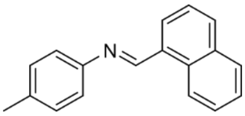
(E)-N-(3-bromobenzylidene)aniline (**1u**)[13]. The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 232 mg (89 % isolated yield) as yellow liquid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.42 (s, 1H), 8.11 (m, 1H), 7.82 (dt, J = 7.7 Hz, J = 1.2 Hz, 1H), 7.63 (ddd, J = 8.0 Hz, J = 2.1 Hz, J = 1.1 Hz, 1H), 7.43 – 7.36 (m, 3H), 7.28 – 7.21 (m, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.8, 151.9, 138.8, 134.5, 131.5, 130.8, 129.6, 128.0, 126.7, 123.3, 121.2.



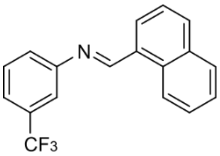
(E)-1-(4-bromophenyl)-N-(o-tolyl)methanimine (**1v**)[14]. The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 95 mg (46 % isolated yield) as yellow oil. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.34 (s, 1H), 7.81 (m, 2H), 7.63 (m, 2H), 7.21 (m, 2H), 7.13 (td, J = 7.4 Hz, J = 1.4 Hz, 1H), 6.94 (dd, J = 7.7 Hz, J = 1.4 Hz, 1H), 2.34 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.4, 151.1, 136.0, 132.5, 132.3, 130.6, 130.5, 127.1, 126.3, 125.9, 117.7, 17.9.



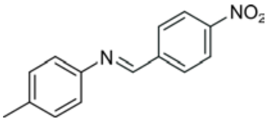
(E)-1-(4-(methylthio)phenyl)-N-(p-tolyl)methanimine (**1w**)[15]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 96 mg (41 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.42 (s, 1H), 7.80 (m, 2H), 7.31 (m, 2H), 7.31 (m, 2H), 7.20 (m, 2H), 7.12 (m, 2H), 2.54 (s, 3H), 2.36 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.0, 149.8, 143.5, 136.2, 133.5, 130.1, 129.3, 126.0, 121.1, 21.1, 15.3.



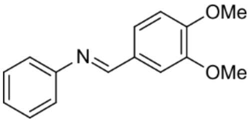
(E)-1-(naphthalen-1-yl)-N-(p-tolyl)methanimine (**1x**)[16]. The crude product was purified by Al_2O_3 150 (2 g, basic activated) column chromatography (*n*-heptane) yielding 208 mg (85 % isolated yield) as yellow oil. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 9.19 (d, J = 8.7 Hz, 1H), 9.14 (s, 1H), 8.13 (dd, J = 7.2 Hz, J = 1.1 Hz, 1H), 8.01 (d, J = 8.2 Hz, 1H), 7.97 (m, 1H), 7.68 (ddd, J = 8.5 Hz, J = 6.2 Hz, J = 1.5 Hz, 1H), 7.62 (m, 2H), 7.29 (m, 4H), 2.45 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 159.6, 150.4, 136.3, 134.4, 132.1, 132.0, 131.8, 130.3, 130.2, 129.1, 127.7, 126.6, 125.7, 124.8, 121.2, 21.2.



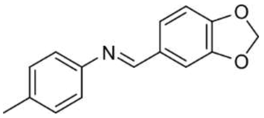
(E)-1-(naphthalen-1-yl)-*N*-(3-(trifluoromethyl)phenyl)methanimine (**1y**). The crude product was purified by Al_2O_3 150 (2 g, basic activated) column chromatography (*n*-heptane) yielding 236 mg (79 % isolated yield) as yellow oil. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 9.16 (d, *J* = 8.8 Hz, 1H), 9.09 (s, 1H), 8.11 (dd, *J* = 7.2 Hz, *J* = 1.1 Hz, 1H), 8.04 (d, *J* = 8.2 Hz, 1H), 7.97 (m, 1H), 7.69 - 7.47 (m, 6H), 7.49 (m, 1H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 162.2, 153.6, 134.4, 132.9, 131.9, 131.8, 131.6, 131.4, 131.3, 130.3, 129.2, 128.1, 126.8, 126.0, 125.7, 124.8, 124.7, 123.3, 122.7, 122.7, 122.7, 122.6, 118.3, 118.3, 118.2, 118.2. HR-MS (ESI$^+$, CH_3CN/MeOH + 1 % H_2O): m/z calcd for $C_{18}H_{12}F_3N$ $[M+H]^+$ 300.0995, found 300.0989.



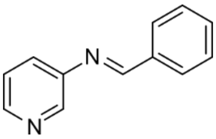
(E)-4-methyl-N-(4-nitrobenzylidene)aniline (**1z**)[16]. The crude product was purified by Al_2O_3 150 (4 g, basic activated) column chromatography (*n*-heptane/Et_2O 1:1) yielding 125 mg (64 % isolated yield) as yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.59 (s, 1H), 8.31 (m, 2H), 8.08 (m, 2H), 7.22 (m, 4H), 2.38 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 156.9, 149.6, 148.7, 142.3, 137.7, 130.3, 129.6, 124.3, 121.4, 21.2.



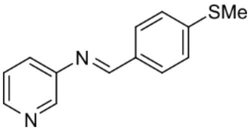
(E)-N-(3,4-Dimethoxybenzylidene)aniline (**1aa**)[17]. The crude product was purified by Al_2O_3 150 (2 g, basic activated) column chromatography (*n*-heptane/Et_2O 1:1) yielding 195 mg (81 % isolated yield) as yellow oil. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.37 (s, 1H), 7.60 (d, *J* = 7.6 Hz, 1H), 7.39 (m, 2H), 7.34 (dd, *J* = 8.2 Hz, *J* = 1.9 Hz, 1H), 7.24 - 7.18 (m, 3H), 6.95 (d, *J* = 8.2 Hz, 1H), 3.93 (s, 3H), 3.90 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 160.1, 152.7, 152.7, 150.0, 130.0, 129.5, 125.9, 124.5, 121.2, 111.1, 109.6, 56.3, 56.2.



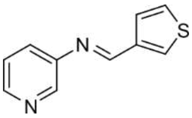
(E)-1-(benzo[d][1,3]dioxol-5-yl)-N-(p-tolyl)methanimine (**1ab**)[10]. The crude product was purified by recrystallisation in *n*-heptane (1 ml) yielding 208 mg (87 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.36 (s, 1H), 7.51 (d, *J* = 7.5 Hz, 1H), 7.27 (dd, *J* = 8.0 Hz, *J* = 1.6 Hz, 1H), 7.19 (m, 2H), 7.10 (m, 2H), 6.89 (d, *J* = 8.0 Hz, 1H), 6.04 (s, 2H), 2.36 (s, 3H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.8, 150.8, 149.8, 148.9, 136.0, 131.9, 130.1, 125.9, 121.1, 108.5, 106.9, 102.2, 21.1.



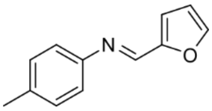
(E)-*N*-benzylidene-3-pyridinamine (**1ac**)[10]. The crude product was purified by Al_2O_3 150 (6 g, basic activated) column chromatography (*n*-heptane/toluene 1:1) yielding 116 mg (87 % isolated yield) as yellow liquid. **^1H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.49 (s, 1H), 8.48 – 8.45 (m, 2H), 7.93 (m, 2H), 7.54 – 7.50 (m, 4H), 7.33 (ddd, *J* = 8.1 Hz, *J* = 4.7 Hz, *J* = 0.8 Hz, 2H). **$^{13}C\{^1H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 162.3, 148.1, 147.5, 143.2, 136.3, 132.2, 129.3, 129.2, 127.8 124.0.



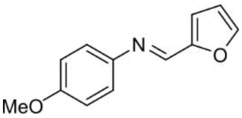
(E)-1-(4-(methylthio)phenyl)-N-(pyridin-3-yl)methanimine (**1ad**)[18]. The crude product was purified by Al_2O_3 150 (10 g, basic activated) column chromatography (Et_2O) yielding 196 mg (86 % isolated yield) as white solid. **^1H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.44 (m, 3H), 7.83 (m, 2H), 7.51 (ddd, *J* = 8.1 Hz, *J* = 2.6 Hz, *J* = 1.6 Hz, 1H), 7.51 (m, 3H), 2.54 (s, 3H). **$^{13}C\{^1H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 161.6, 148.2, 147.4, 144.6, 143.2, 132.9, 129.6, 127.8, 125.9, 124.0, 15.2.



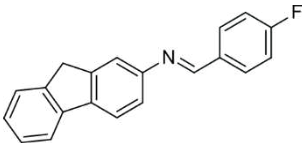
(E)-N-(pyridin-3-yl)-1-(thiophen-3-yl)methanimine (**1ae**). The crude product was purified by Al_2O_3 150 (2 g, basic activated) column chromatography (Et_2O) yielding 158 mg (84 % isolated yield) as yellow oil. **^1H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.49 (s, 1H), 8.44 (m, 2H), 7.88 (dd, *J* = 2.9 Hz, *J* = 1.1 Hz, 1H), 7.68 (ddd, *J* = 5.1 Hz, *J* = 1.2 Hz, *J* = 0.4 Hz, 1H), 7.50 (ddd, *J* = 8.1 Hz, *J* = 2.6 Hz, *J* = 1.6 Hz, 1H), 7.43 (ddd, *J* = 5.1 Hz, *J* = 2.9 Hz, *J* = 0.7 Hz, 1H), 7.31 (ddd, *J* = 8.1 Hz, *J* = 4.8 Hz, *J* = 0.7 Hz, 1H). **$^{13}C\{^1H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 156.3, 148.2, 147.4, 143.1, 141.0, 131.6, 127.8, 127.4, 126.1, 124.0. HR-MS (ESI^+, CH_3CN/MeOH + 1 % H_2O): m/z calcd for $C_{10}H_8N_2S$ $[M+H]^+$ 189.0481, found 189.0485.



(E)-1-(furan-2-yl)-N-(p-tolyl)methanimine (**1af**)[19]. The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/Et_2O 20:1) yielding 172 mg (93 % isolated yield) as yellow oil. **^1H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.30 (s, 1H), 7.62 (d, *J* = 1.7 Hz, 1H), 7.20 (m, 2H), 7.13 (m, 2H), 6.95 (dd, *J* = 3.5 Hz, *J* = 0.6 Hz, 1H), 6.57 (dd, *J* = 3.5 Hz, *J* = 1.8 Hz, 1H), 2.36 (s, 3H). **$^{13}C\{^1H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 153.0, 149.3, 147.3, 145.8, 136.6, 130.2, 121.2, 115.9, 112.5, 21.1.



(E)-1-(furan-2-yl)-N-(4-methoxyphenyl)methanimine (**1ag**)[20].The crude product was purified by Al_2O_3 150 (12 g, basic activated) column chromatography (*n*-heptane/toluene/Et_2O 1:1:0.1) yielding 105 mg (52 % isolated yield) as yellow oil. **^1H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.31 (s, 1H), 7.61 (d, *J* = 1.7 Hz, 1H), 7.22 (m, 2H), 6.92 (m, 3H), 6.57 (dd, *J* = 3.5 Hz, *J* = 1.8 Hz, 1H), 3.81 (s, 3H). **$^{13}C\{^1H\}$ NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 158.9, 153.1, 146.1, 145.7, 144.7, 122.6, 115.5, 114.8, 112.5, 55.8.



(E)-N-(4-fluorobenzylidene)-9H-fluoren-2-amine (**1ah**). The crude product was purified by Al_2O_3 150 (4 g, basic activated) column chromatography (*n*-heptane/acetone 1:1) yielding 155 mg (54 % isolated yield) as pale-yellow solid. **^{1}H NMR** (δ, 400 MHz, CD_2Cl_2, 25 °C): 8.54 (s, 1H), 7.95 (m, 2H), 7.80 (t, J = 8.1 Hz, 2H), 7.56 (m, 1H), 7.43 (m, 1H), 7.39 (m, 1H), 7.31 (dd, J = 7.4 Hz, J = 1.2 Hz, 1H), 7.27 (m, 1H), 7.20 (m, 2H), 3.95 (s, 2H). **^{13}C{^{1}H} NMR** (δ, 101 MHz, CD_2Cl_2, 25 °C): 166.3, 163.8, 158.3, 151.1, 145.0, 143.9, 141.7, 140.3, 133.4, 131.1, 131.1, 127.2, 126.9, 125.4, 120.7, 120.5, 120.1, 118.0, 116.3, 116.1, 37.3. HR-MS (ESI$^+$, CH_3CN/MeOH + 1 % H_2O): m/z calcd for $C_{20}H_{15}FN$ [M+H]$^+$ 288.1183, found 288.1187.

4. Experimental mechanistic studies

N-Alkylation of aniline under pure O_2 atmosphere

KO^{t}Bu (0.65 mmol) and molecular sieves (15 pellets, 4 Å) were put into a 10 mL screwcap-vial containing a magnetic stir bar. Then a toluene (3 mL) solution containing aniline (0.60 mmol) and benzyl alcohol (0.50 mmol) was added. The vial was closed and flushed with pure oxygen for 1 min and then stirred for 15 min. A GC-MS analysis showed full conversion.

N-Alkylation of aniline in the presence of 18-crown-6 ether

KO^{t}Bu (0.65 mmol) and molecular sieves (15 pellets, 4 Å) were put into a 10 mL flat-bottom vial containing a magnetic stir bar. Then a toluene (3 mL) solution containing aniline (0.60 mmol), benzyl alcohol (0.50 mmol) and 18-crown-6 ether (0.65 mmol) was added. The reaction mixture was stirred for 3 h at room temperature, while exposed to ambient atmospheric conditions. A GC-MS analysis showed 88% conversion.

Synthesis of benzyl alcohol [^{18}O]

Benzyl alcohol [^{18}O] was synthesized by slightly modified procedure from the literature.[21]

Toluene-4-sulforic acid (2mg) was added to methyl orthobenzoate (925.3 µl, 5.38 mmol, 1 equiv.) in a Schlenk tube under argon atmosphere. [^{18}O]Water (99% ^{18}O) (100 µl, 5.54 mmol, 1.03 equiv.) was added under rapid stirring. The mixture became homogenous after 2 minutes of stirring. The solution was stirred for further 30 minutes. Methanol was gently removed in vacuum (100 mbar, 30 minutes). 10 mL dry diethyl ether was added, and the mixture was cooled to 0 °C. Solid lithium aluminum hydride (217.8 mg, 5.65 mmol, 1.05 equiv.) was added in batch. The suspension was stirred for 5 minutes at 0 °C and 16 hours at room temperature. Water (2 mL) and saturated ammonium chloride solution (2 mL) were slowly added. The aqueous phase was extracted with *n*-pentane (3x 10 mL). The combined organic phases were dried over sodium sulfate and the solvent was gently removed, yielding benzyl alcohol [^{18}O] (290 mg, 49%) as colorless liquid.

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.37-7.35 (m, *4*H), 7.34 – 7.27 (m, 1H), 4.70 (d, *J* = 5.8 Hz, 2H), 1.63 (t, *J* = 5.9 Hz, 1H) ppm. **^{13}C{^{1}H} NMR** (100 MHz, $CDCl_3$, δ): 141.0, 128.7, 127.8, 127.1, 65.5 ppm.

The ^{18}O-content was assigned by mass spectrometry, using EI-MS from GC-MS data.[22]

Calculation: (Signal of m/z 110) / [(Signal of m/z 110) + (Signal of m/z 108)]:
70.02 / (70.02 + 9.93) = 88 % ^{18}O in benzyl alcohol

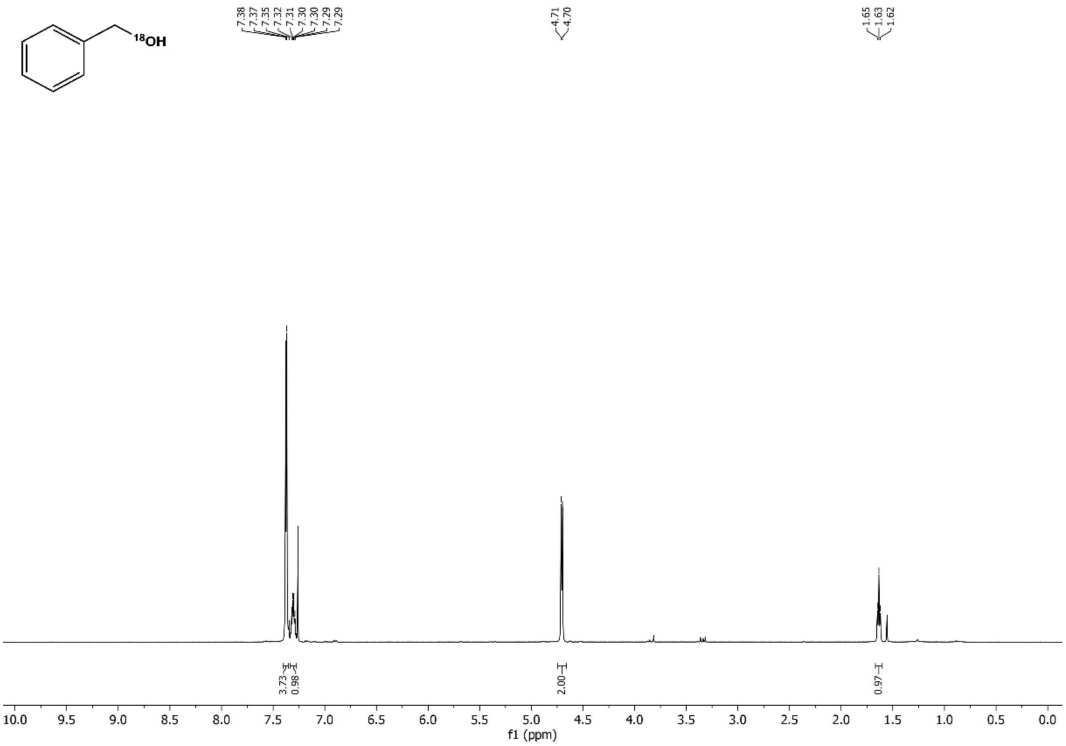


Figure S1 ^{1}H NMR spectrum of benzyl alcohol [^{18}O].

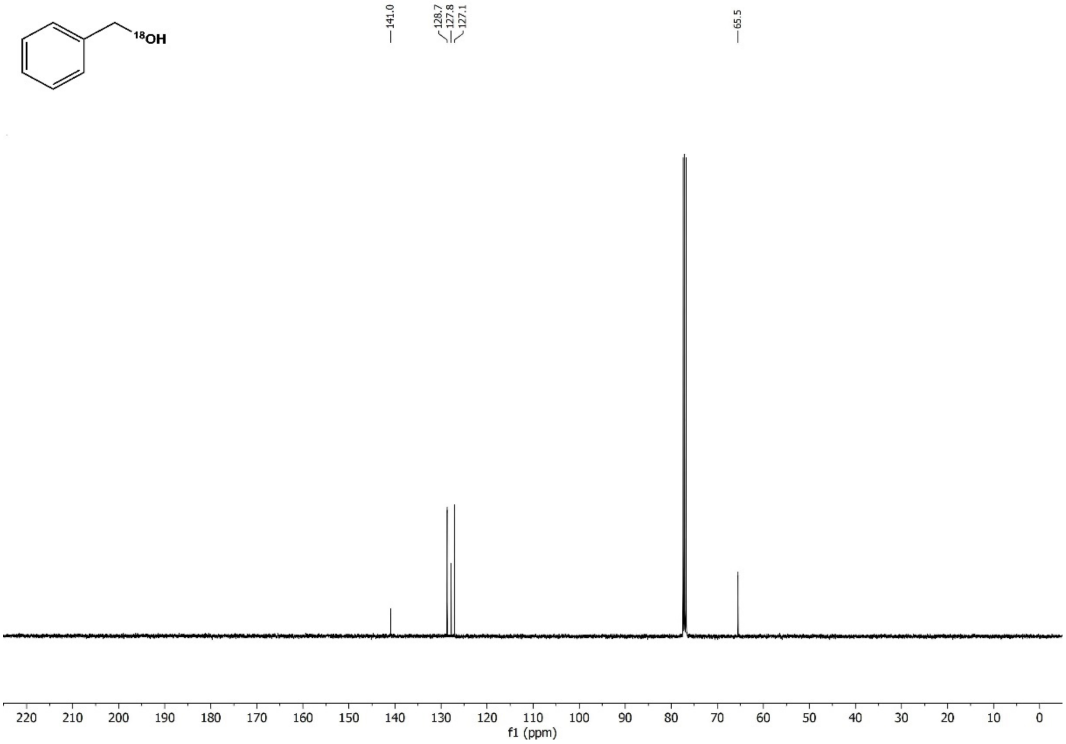


Figure S2 ^{13}C{^{1}H} NMR spectrum of benzyl alcohol [^{18}O].

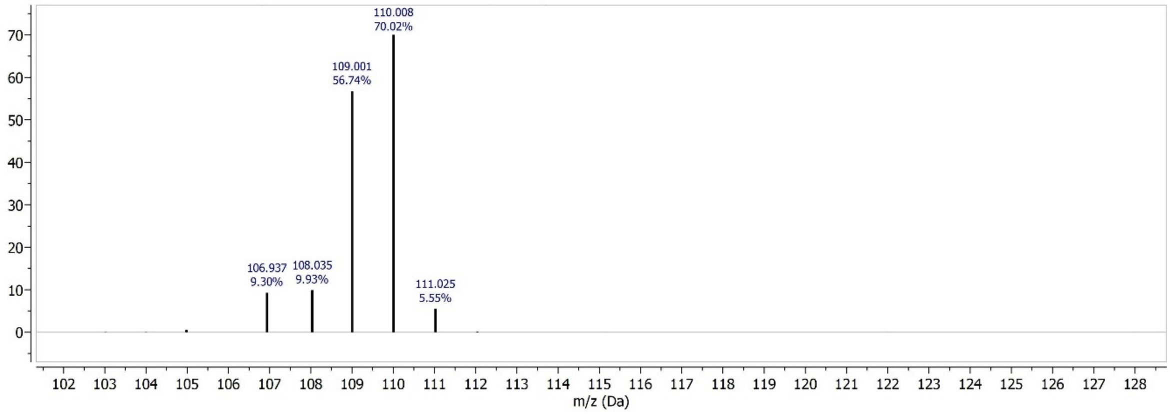


Figure S3 EI-MS spectrum of benzyl alcohol [^{18}O].

Synthesis of potassium benzylate (PhCH$_2$OK)

In an Argon filled glovebox benzyl alcohol (1.5 mL, 14.4 mmol) was reacted neat with metal K (0.5 g). The reaction mixture was stirred for 6 h at room temperature, and the solidified residue washed twice with pentane (3 mL). The off-white solid was dried in vacuum. Yield: 1.72 g (81 %)

^{1}H NMR (250 MHz, THF-d8): 7.37 (d, J = 7.2 Hz, 2H), 7.21 (t, J = 7.3 Hz, 2H), 7.09 (t, J = 7.2 Hz, 1H), 4.75 (s, 2H).

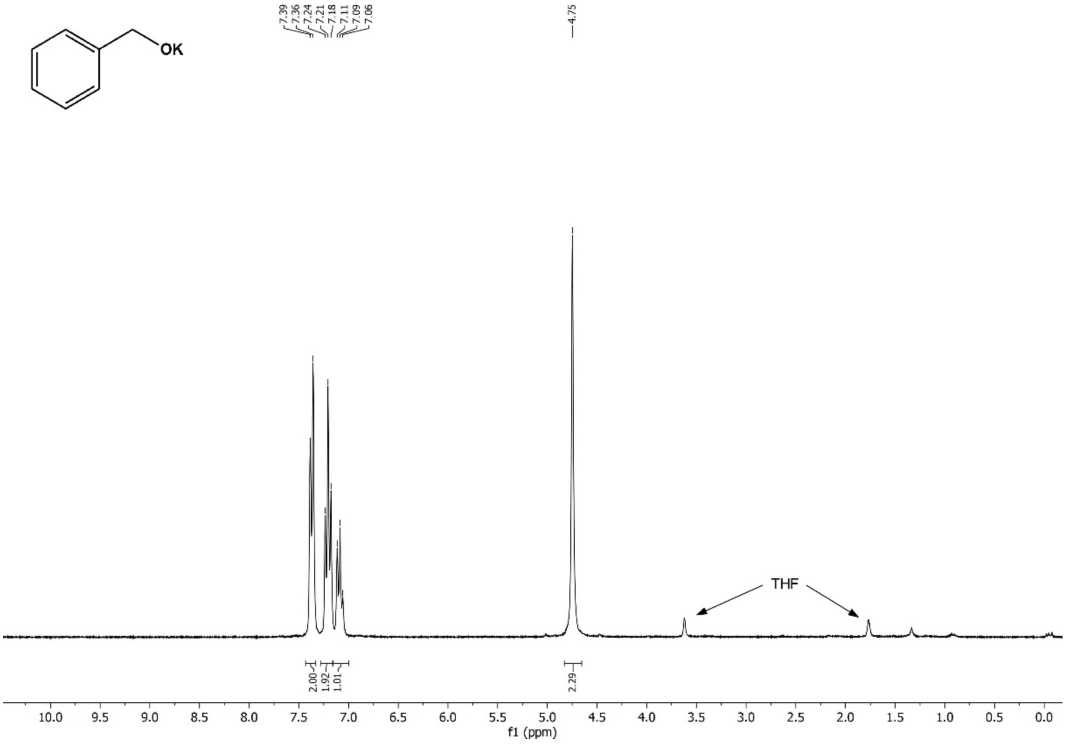
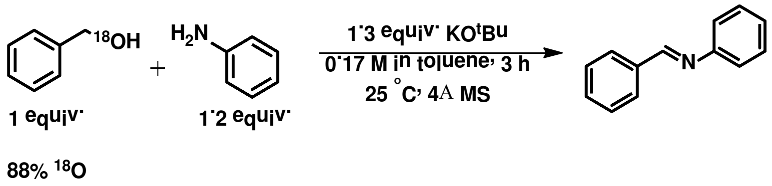


Figure S4 ^{1}H NMR of potassium benzylate.

N-Alkylation of aniline with benzyl alcohol [^{18}O]



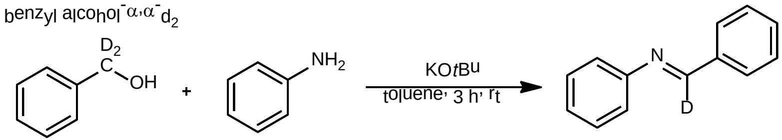
The reaction was carried out according to the procedure described above, employing benzyl alcohol [^{18}O] instead of standard benzyl alcohol. Samples for GC-MS analysis were taken every 15 min. The ^{18}O content was determined by EI-MS.

20±2% ^{18}O content was found in *in situ* formed benzaldehyde upon reaction progress.

N-Alkylation of aniline using PhCH2OK salt

The reaction was carried out according to the procedure described above, employing the potassium benzylate salt instead of standard benzyl alcohol and KO^{t}Bu. The manipulation of reactants was carried out in a glovebox, whereas the reaction itself was carried out under ambient atmospheric conditions. A GC-MS analysis showed 54% conversion.

N-Alkylation of aniline with benzyl alcohol-α,α-d2



1a-D

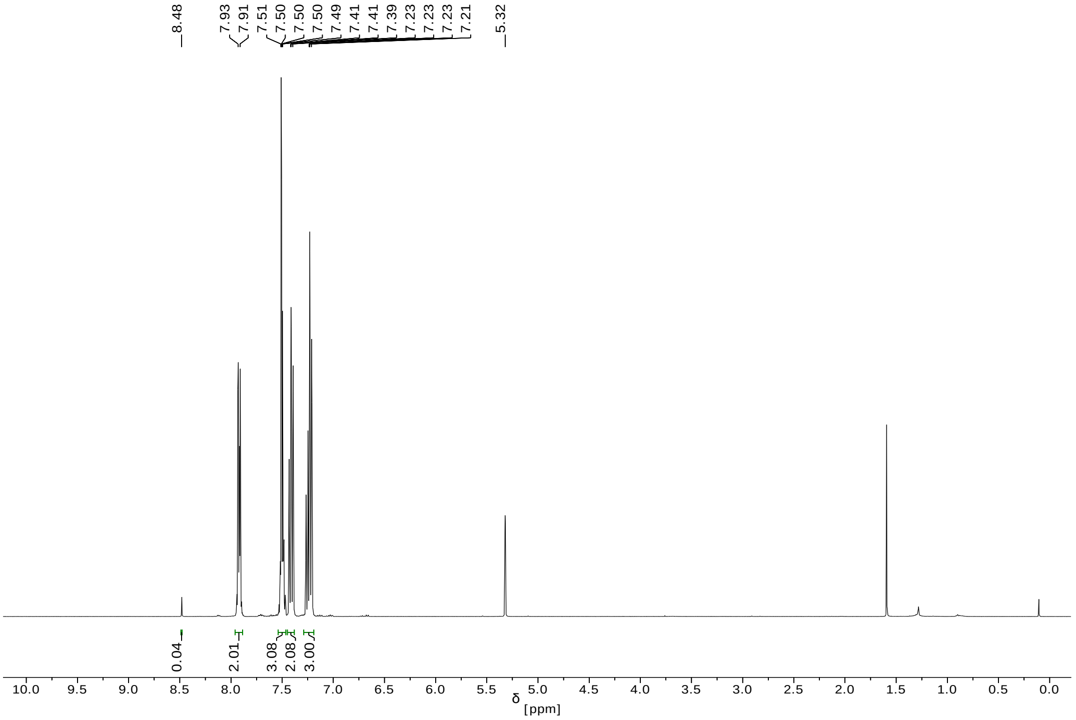


Figure S5. ^{1}H NMR of **1a-D.** Reaction conditions: aniline (55 µl, 0.6 mmol), benzyl alcohol-α,α-d2 (52 µl, 0.5 mmol, 96% D), KO^{t}Bu (73 mg, 0.65 mmol), molecular sieves (15-20 pellets, 4 Å) and toluene (3 ml), room temperature, 3 h. Work up same as **1a.**

EPR monitoring of N-Alkylation

In a glovebox, a flame-dried EPR tube was charged with a 0.5 mL of a solution of KO^{t}Bu, aniline and benzyl alcohol in toluene (prepared as described in the general procedure above) and sealed. A consecutive EPR experiment (100 K and room temperature) of this sample showed no presence of paramagnetic material. The measurement itself was repeated after 1 h with identical results.

Subsequently the air-tight sample tube was opened, and 2 mL of air was injected with a long steel canula into the solution. The sample was consecutively cooled down to 100 K and subjected to EPR measurement. A signal was detected with g_{iso} = 2.001 and hyperfine coupling A(^{1}H) = 18 G distinctive for an organic carbon-based radical.

A similar experiment was carried out with $PhCH_2OK$ salt and aniline to give identical results.

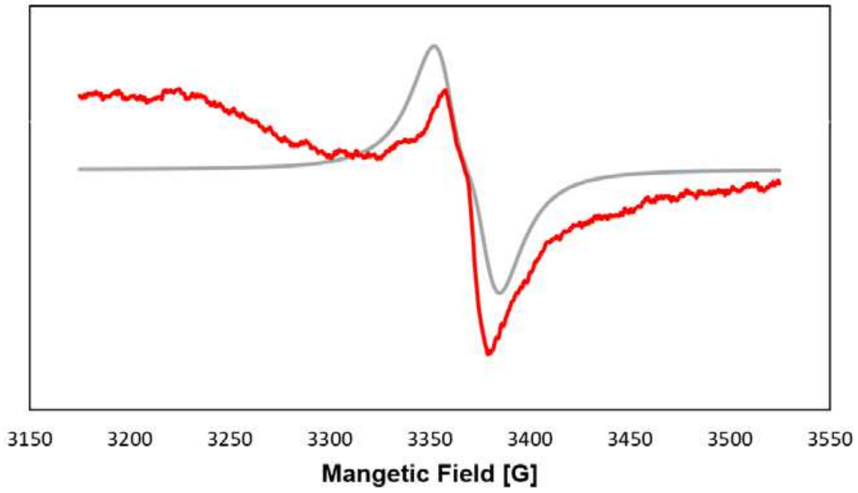


Figure S6. X-Band cw-EPR spectrum of the reaction mixture at 100 K in frozen toluene glass. The obtained signal is depicted in red and the computer simulation in grey color.

5. Computational studies

For the structure optimization, Density Functional Theory (DFT) was the method of choice, using the PBE0[23a] functional, def2-TZVP[23b] basis set with D4[23c] dispersion corrections as well as implicit solvation via the CPCM[23d] implementation (toluene was used as solvent). All calculations were performed using the ORCA 5.0.3[23e,f] software suite. To validate the accuracy of the DFT method, structures **A** and **D** were optimized using domain based local pair-natural orbital (DLPNO) based singles- and doubles, and perturbative triples coupled cluster (CCSD(T))[23g] with the def2-TZVP basis set an the CPCM implicit solvent model and minimal differences in the geometry were observed. The energetics confirmed this, as the energy change between single point calculations performed at the DLPNO-CCSD(T)/def2-QZVP[23b]/CPCM level on both DFT and CCSD optimized structure was less than 0.25 kcal/mol. To speed up calculation coulomb fitting was used.[23h,i]

By performing single point calculations using B2PLYLP[23i]/def2-TZVP/D4/CPCM on the structures and comparing them to CCSD values, we observed very small energy differences of up to 3 kcal/mol. As it is known that CCSD does not handle mixed spin states well, it was unfeasible to obtain accurate energy values for the minimum energy crossing points (MECP). As such, the DFT double hybrid values were taken for the energy barriers related to these spin-transition points along the reaction pathway. The MECP structures were calculated using a method by Harvey *et al as* implemented in ORCA.[24] Here, both the energy of spin surface 1 and the energy difference between the two spin surfaces are minimized. The structures were verified to be true minimum geometries by running surface crossing adapted frequency calculations, which showed no negative values.

The full proposed mechanism is depicted in Figure S7, with a multitude of conditional branching points, depending on the immediate environment of the molecule. While the mechanism itself is not exhaustive, it can provide greater insight into the possible species formed and their interactions.

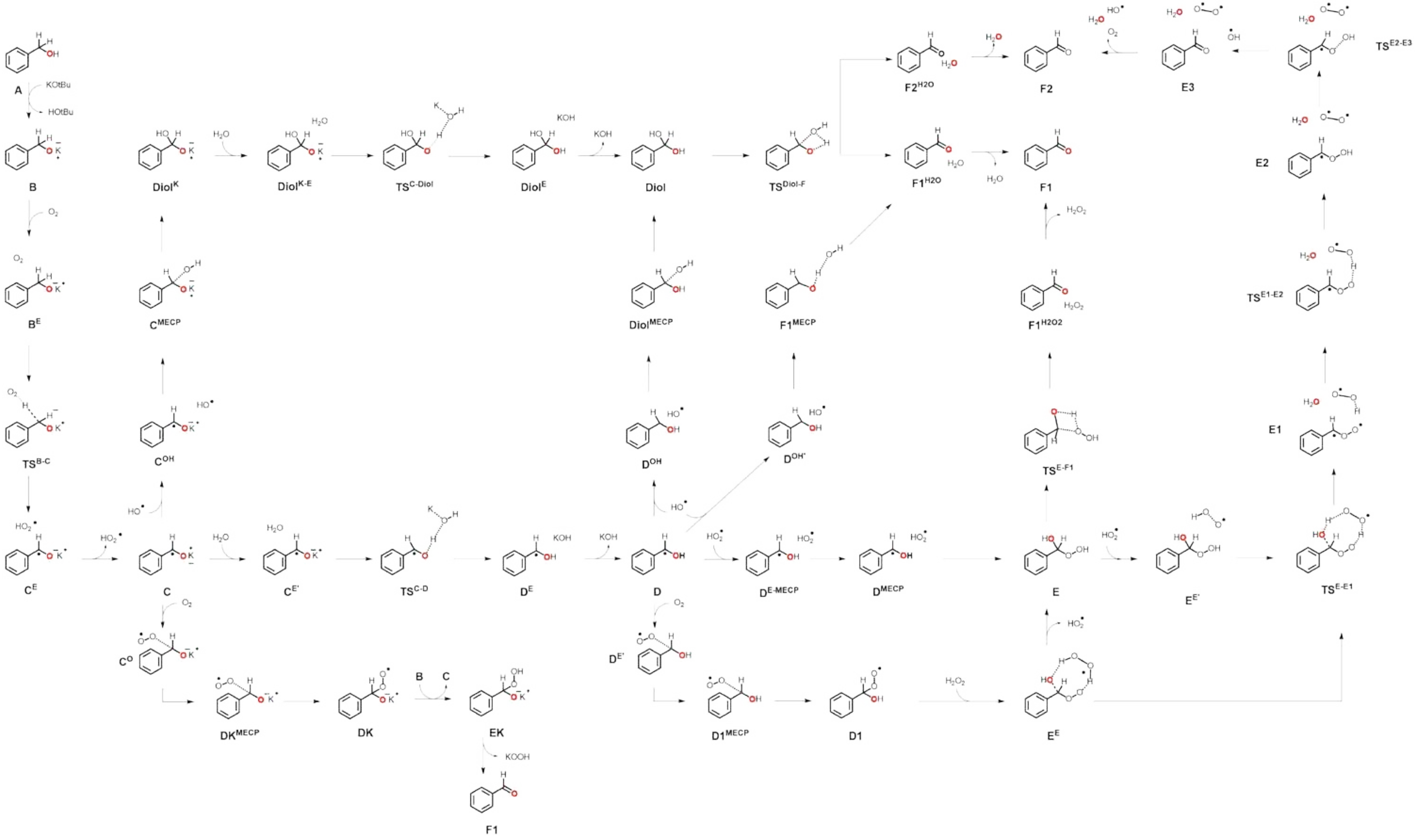


Figure S7. Reaction network of the oxidation of benzylalcohol to benzaldehyde.

Figure S8 depicts the reaction pathways C to E that depend on the existence of the short-lived hydroxyl radical and are therefore less likely to take place, even though the calculated thermodynamics and kinetics suggest that they are viable.

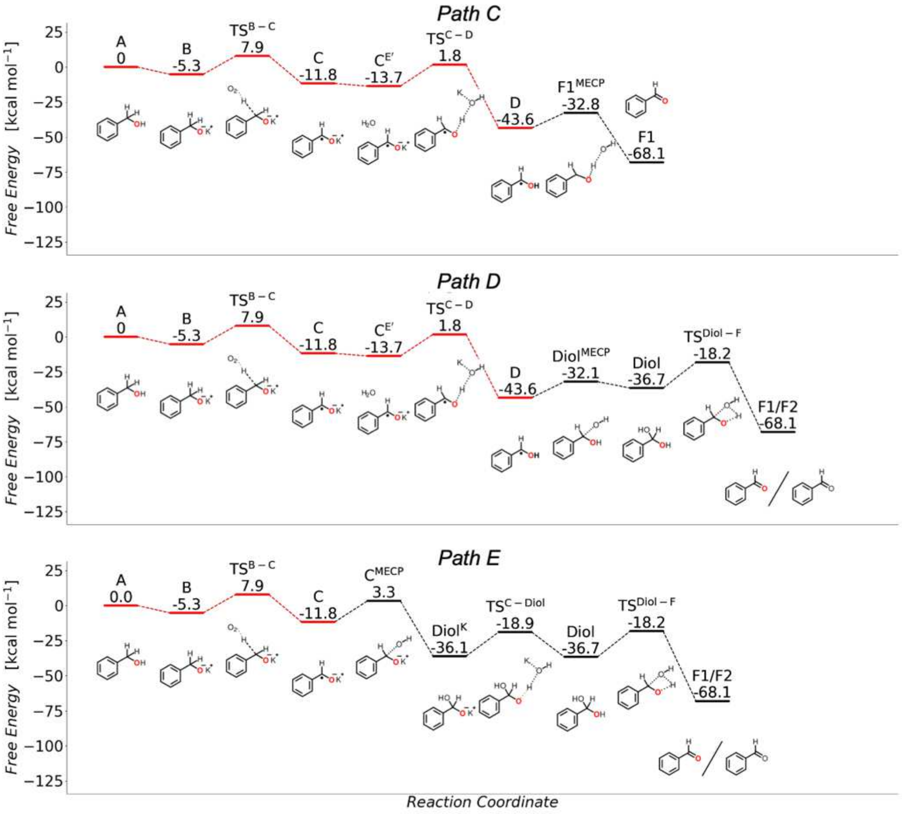


Figure S8. Pathways C-E of the water-assisted mechanism, branching out from C and D respectively. As these pathways depend on the presence of a short-lived hydroxyl radical, they are less likely to take place. All values are free energies Δ*G*, in kcal/mol, calculated with DLPNO-CCSD(T)/def2-QZVP//PBE0/def2-TZVP/D4. The ^{18}O-label is depicted in red.

For the EPR calculations, the method of choice was B2PLYP/def2-TZVP/D4/CPCM, using gauge independent atomic orbitals. We performed calculations on all intermediate radicals, and **C/D** have the g_{iso} values closest to the experimental results. Simulations of the EPR spectra were performed with Easyspin[25] and do not provide clear evidence whether **C** or **D** is the experimentally observed radical species or a combination of both as a small shoulder is visible in the experimental spectra, compare Figures S6 and S9.

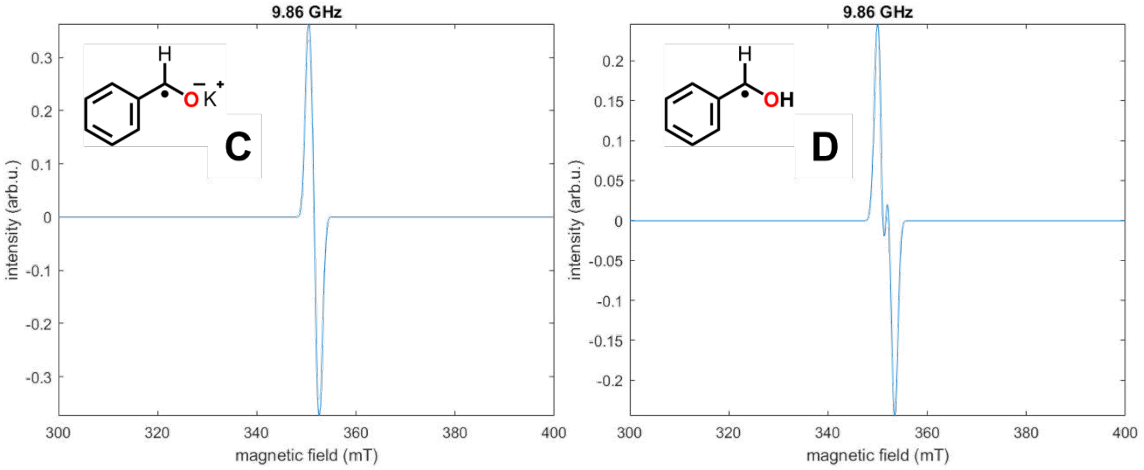
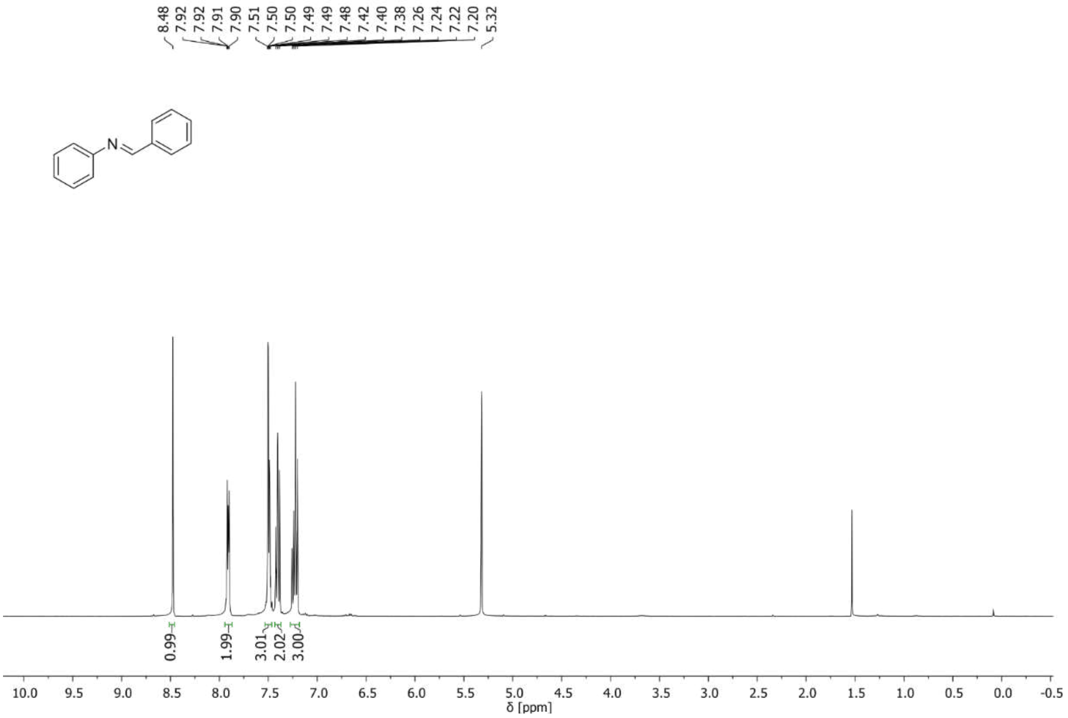
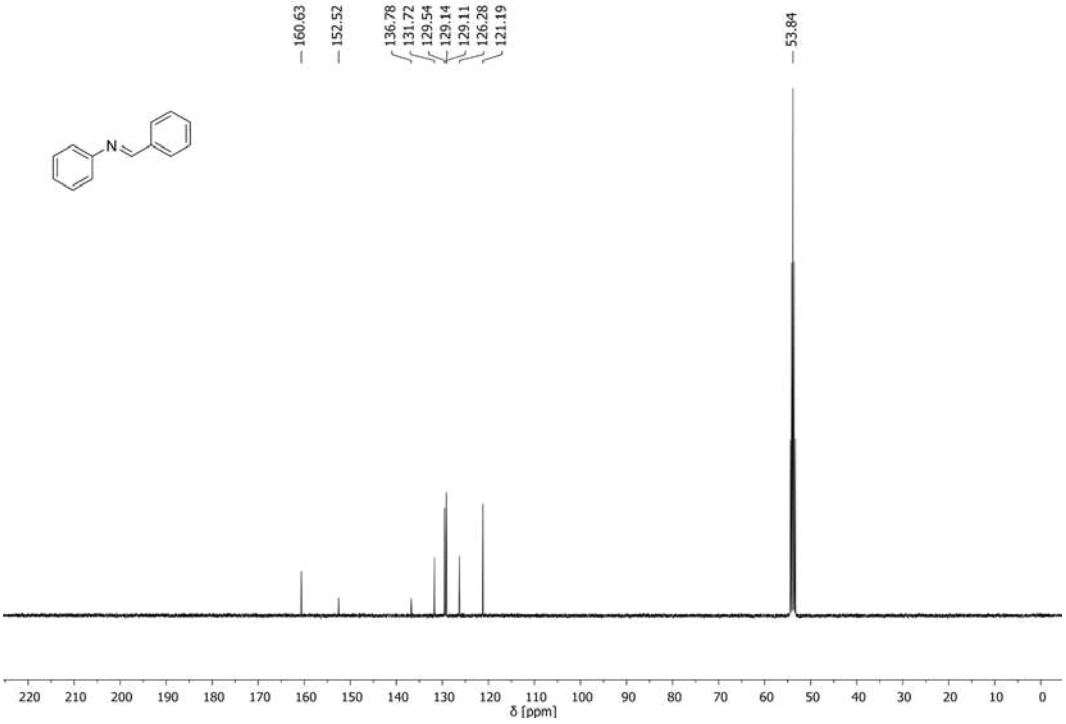


Figure S9. Simulated EPR spectrum at 100 K. Values obtained from B2PLYLP/def2-TZVP/D4/CPCM calculations using gauge independent atomic orbitals.

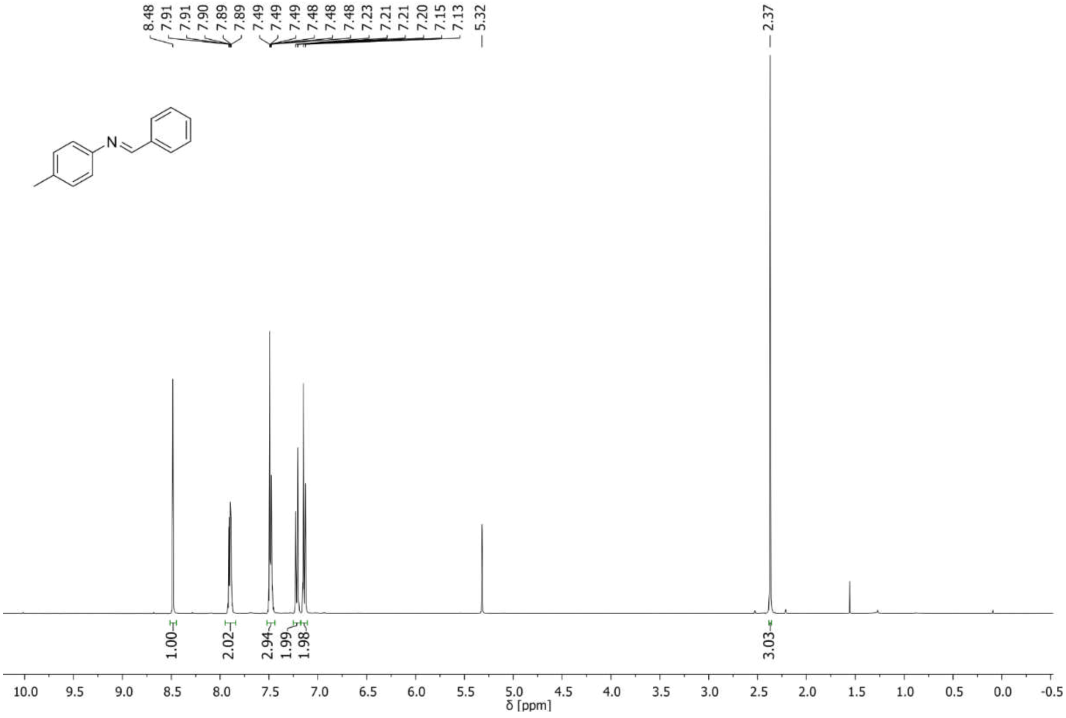
6. NMR Spectra of Imines Products



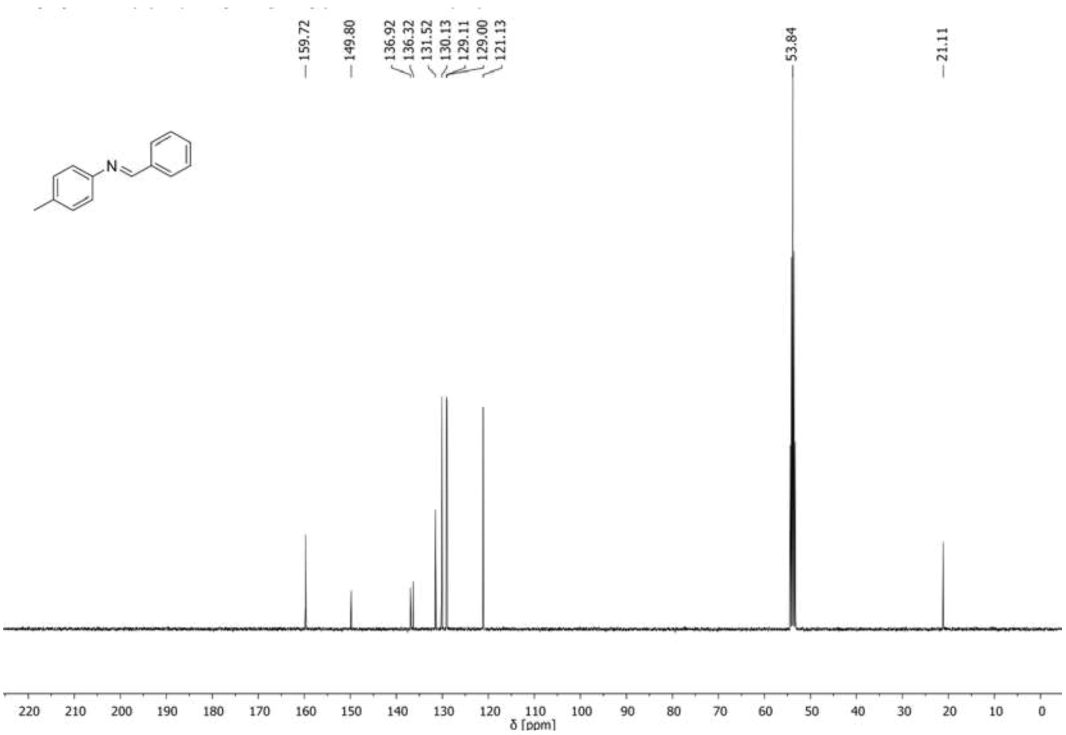
^{1}H NMR: (E)-N-benzylideneaniline (**1a**)



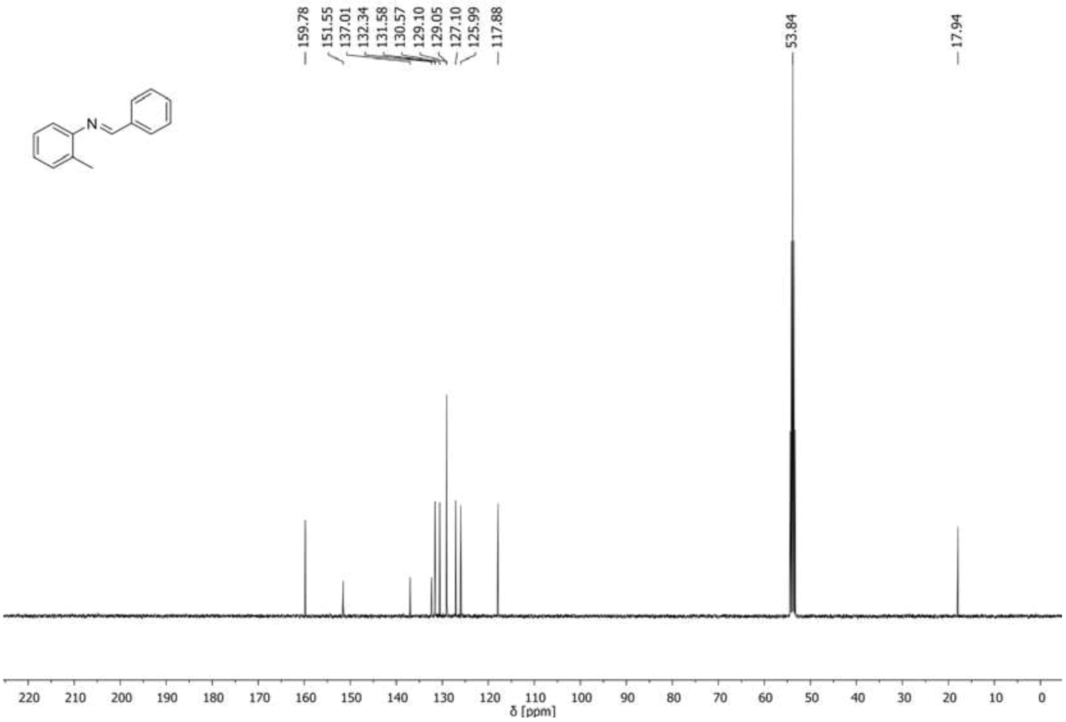
^{13}C{^{1}H} NMR: (E)-N-benzylideneaniline (**1a**)



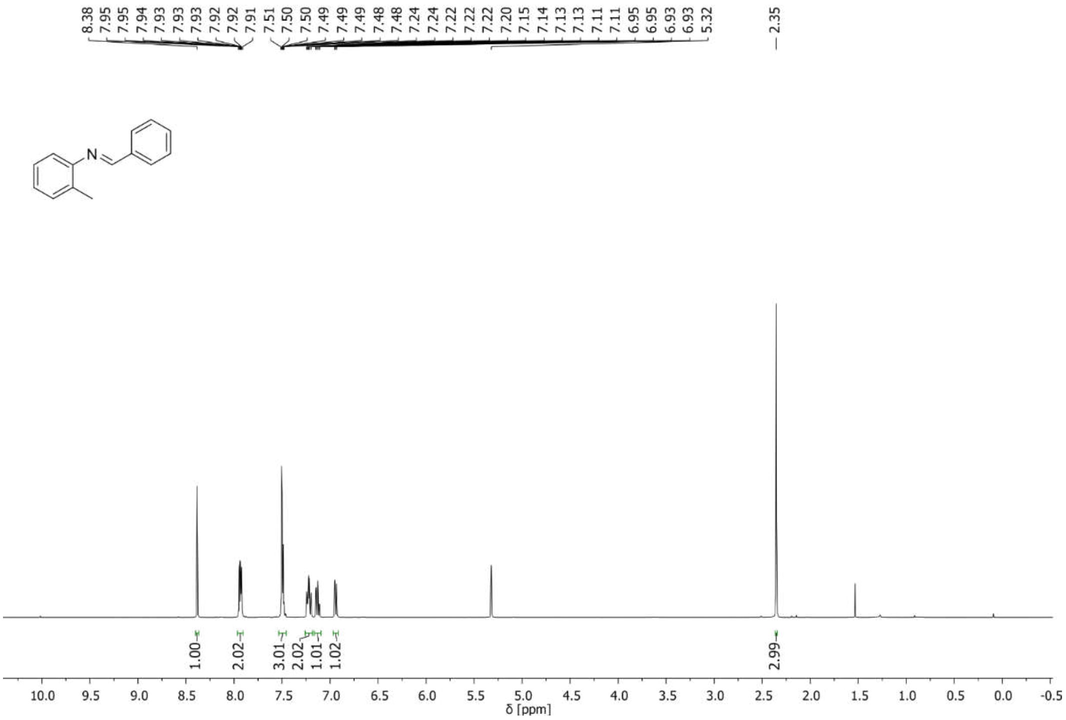
^{1}H NMR: (E)-1-phenyl-N-(p-tolyl)methanimine (**1b**)



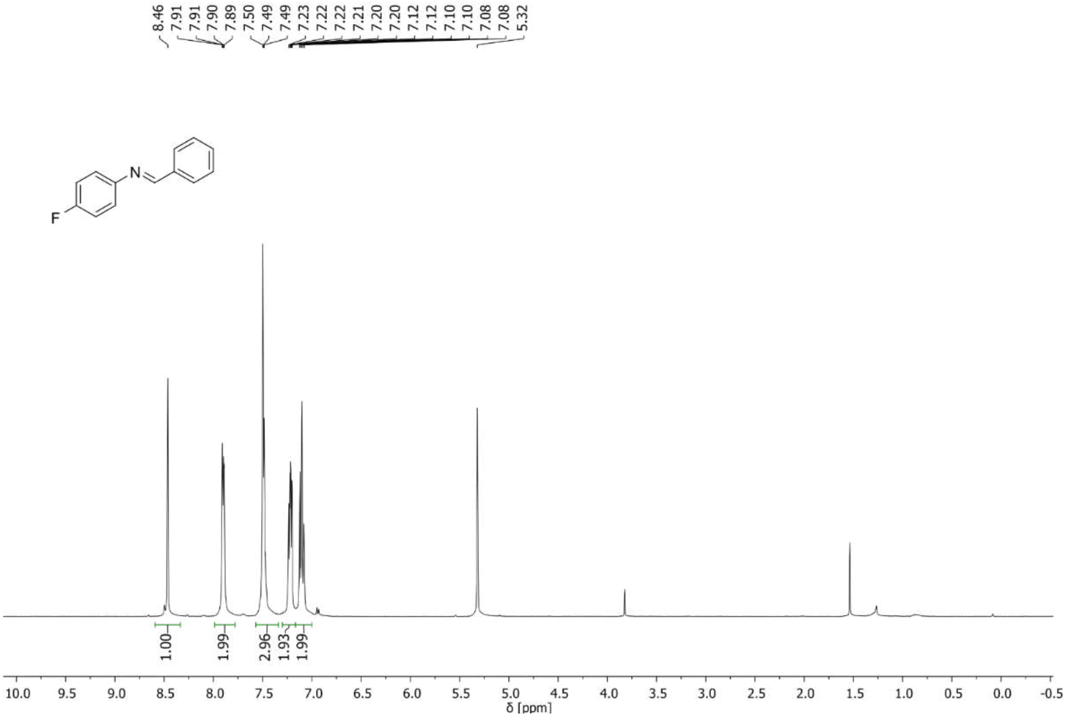
$^{13}C\{^{1}H\}$ NMR: (E)-1-phenyl-N-(p-tolyl)methanimine (**1b**)



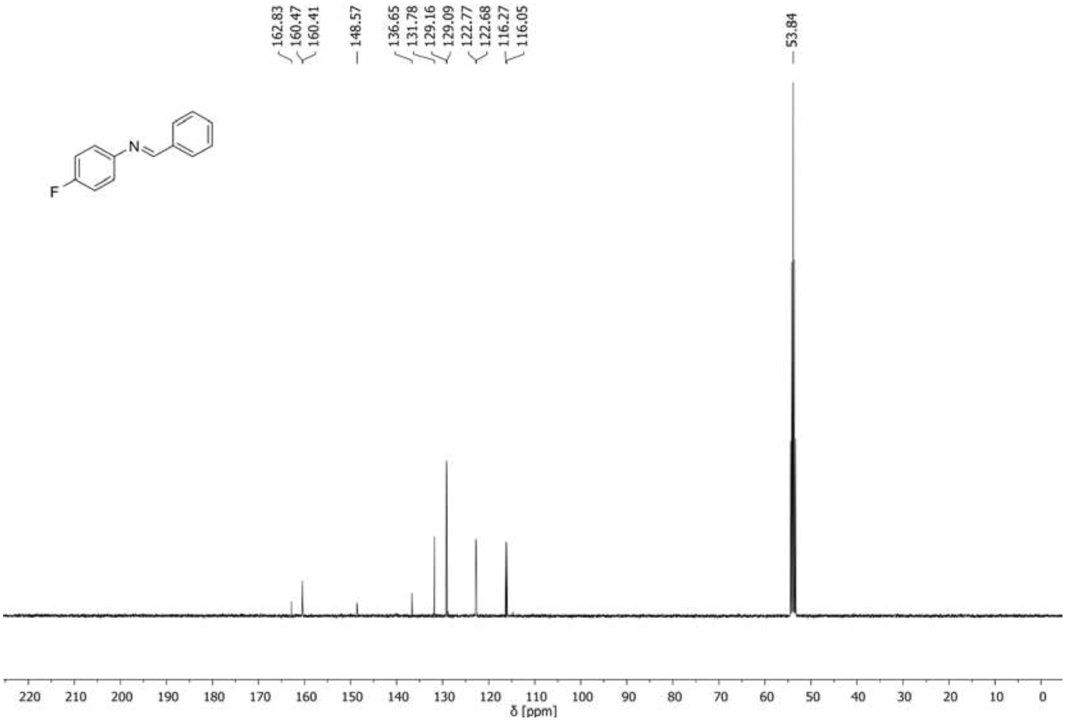
^{1}H NMR: (E)-1-phenyl-N-(o-tolyl)methanimine (**1c**)



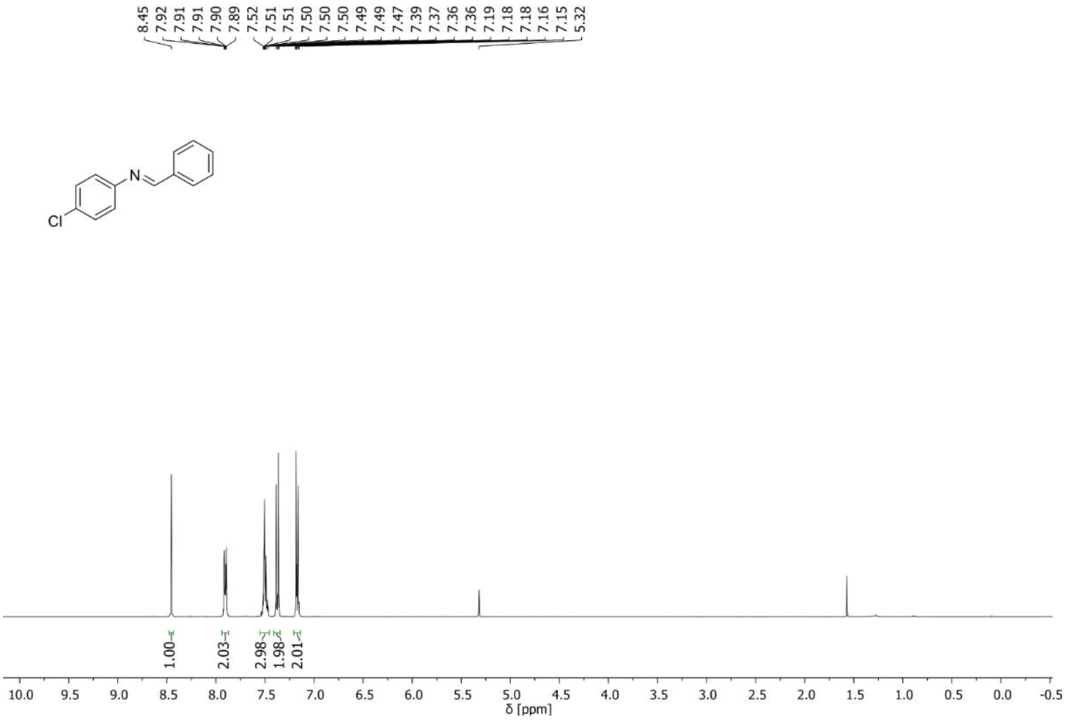
$^{13}C\{^{1}H\}$ NMR: (E)-1-phenyl-N-(o-tolyl)methanimine (**1c**)



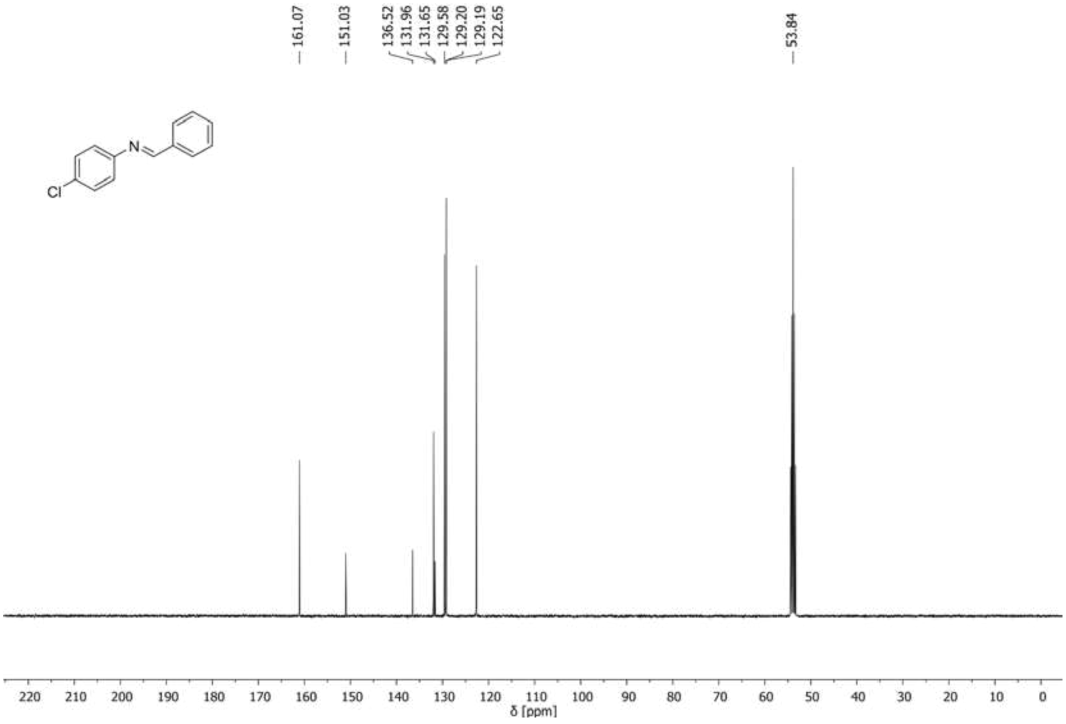
^{1}H NMR: (E)-N-benzylidene-4-fluoroaniline (**1d**)



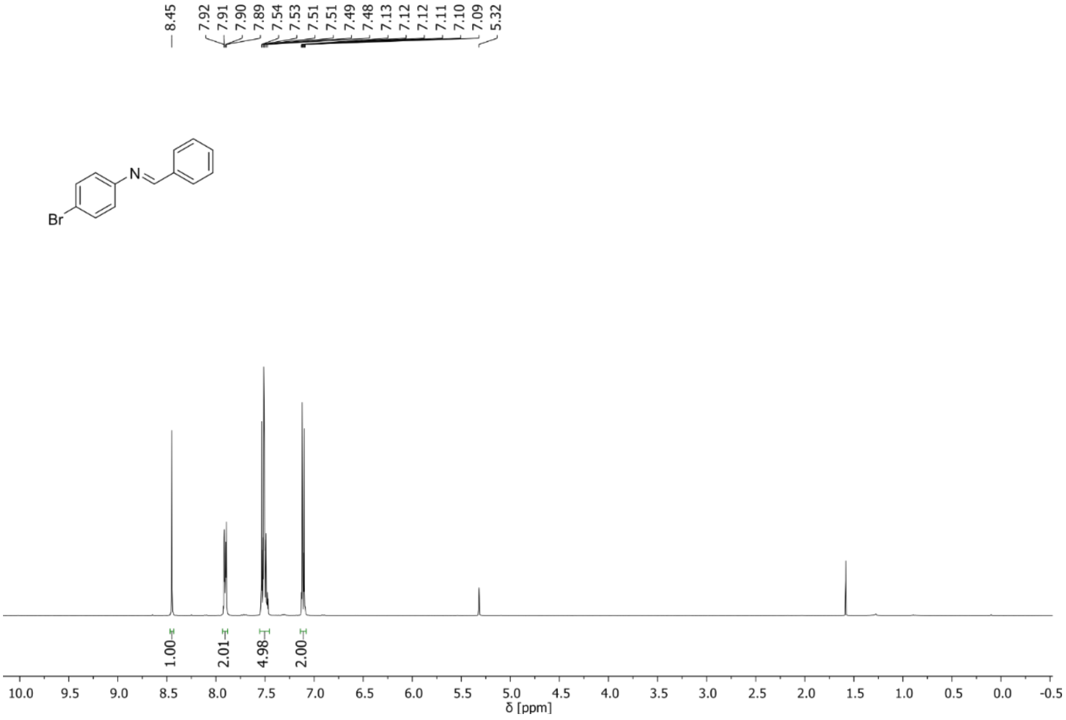
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-fluoroaniline (**1d**)



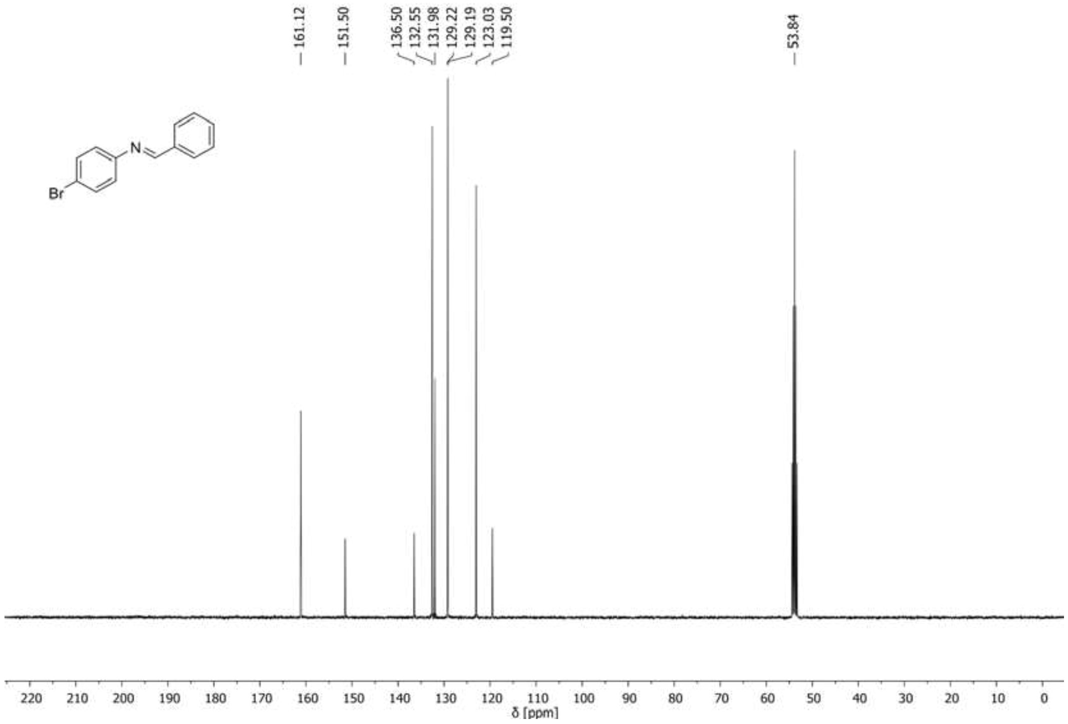
^{1}H NMR: (E)-N-benzylidene-4-chloroaniline (**1e**)



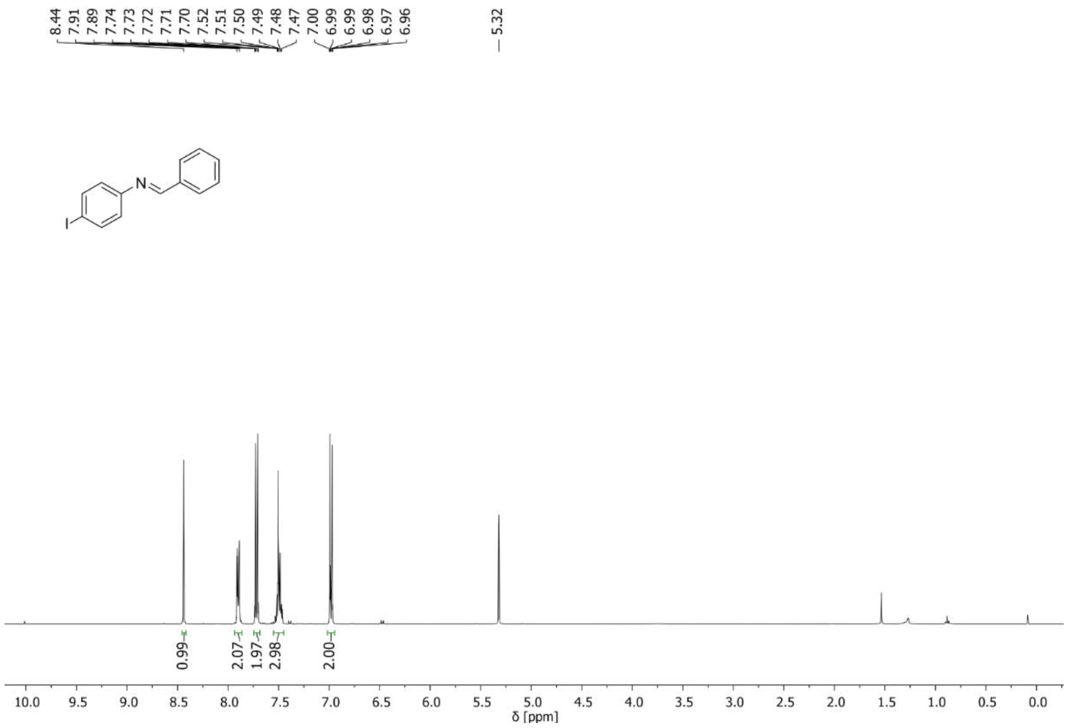
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-chloroaniline (**1e**)



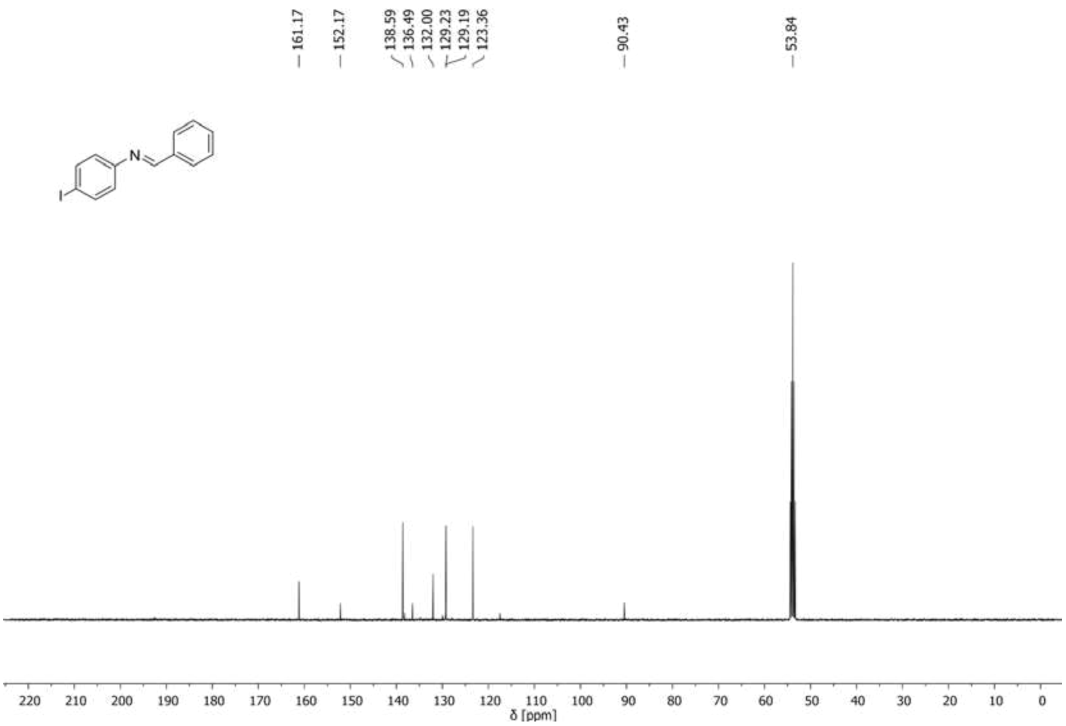
^{1}H NMR: (E)-N-benzylidene-4-bromoaniline (**1f**)



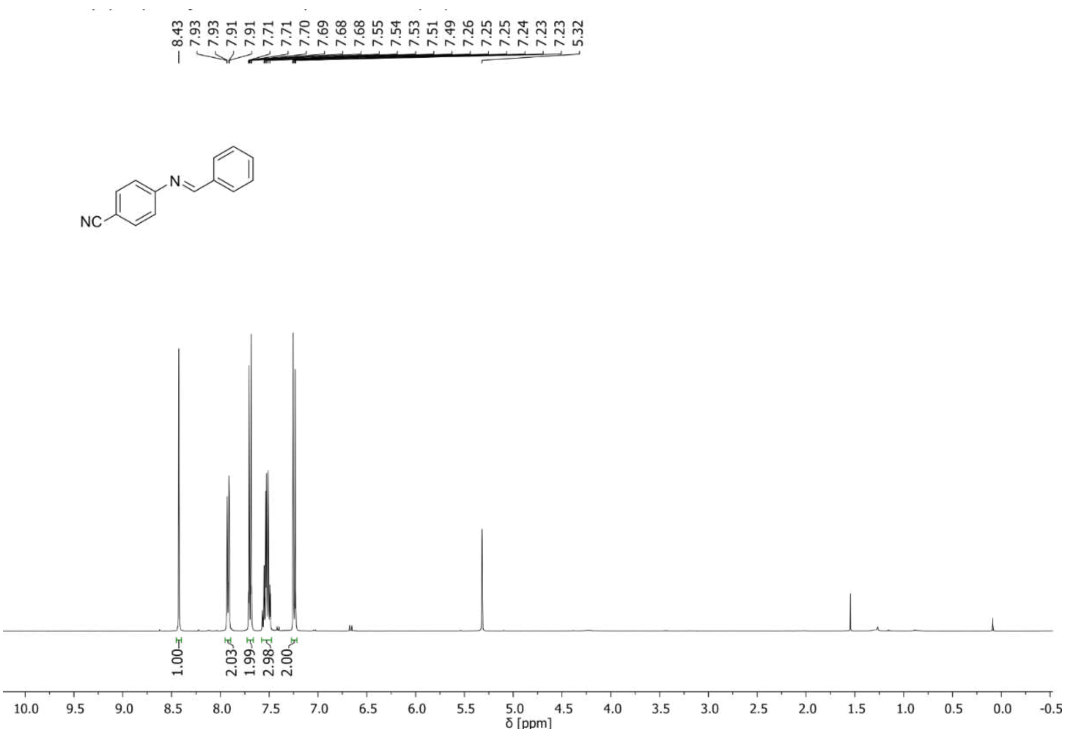
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-bromoaniline (**1f**)



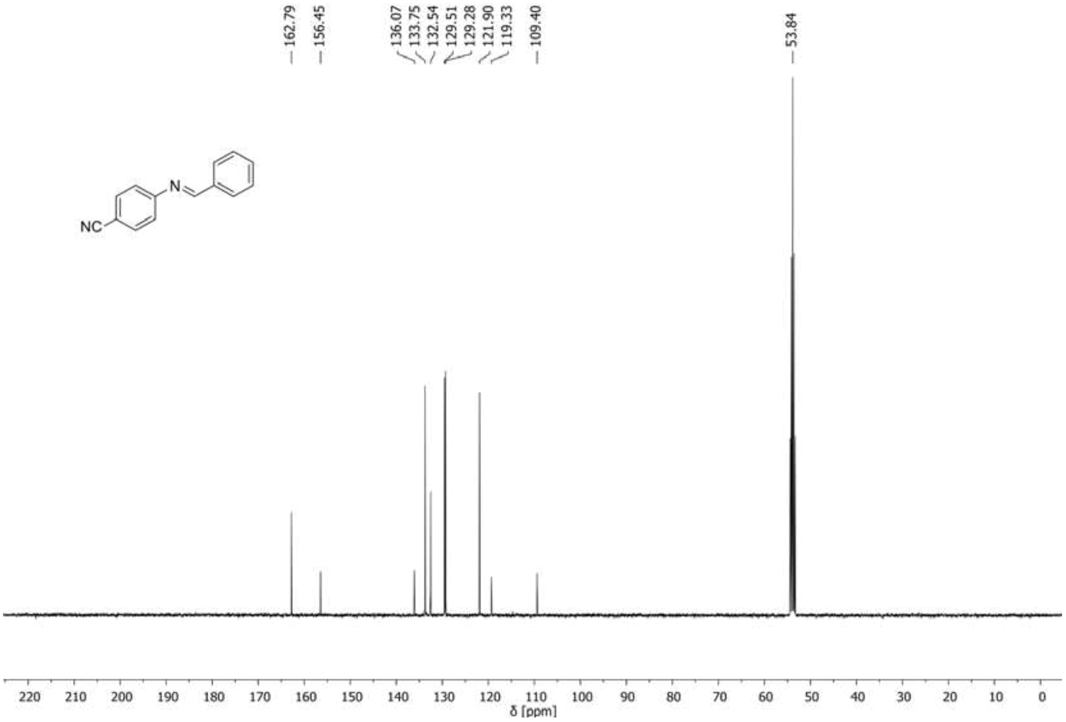
^{1}H NMR: (E)-N-benzylidene-4-iodoaniline (**1g**)



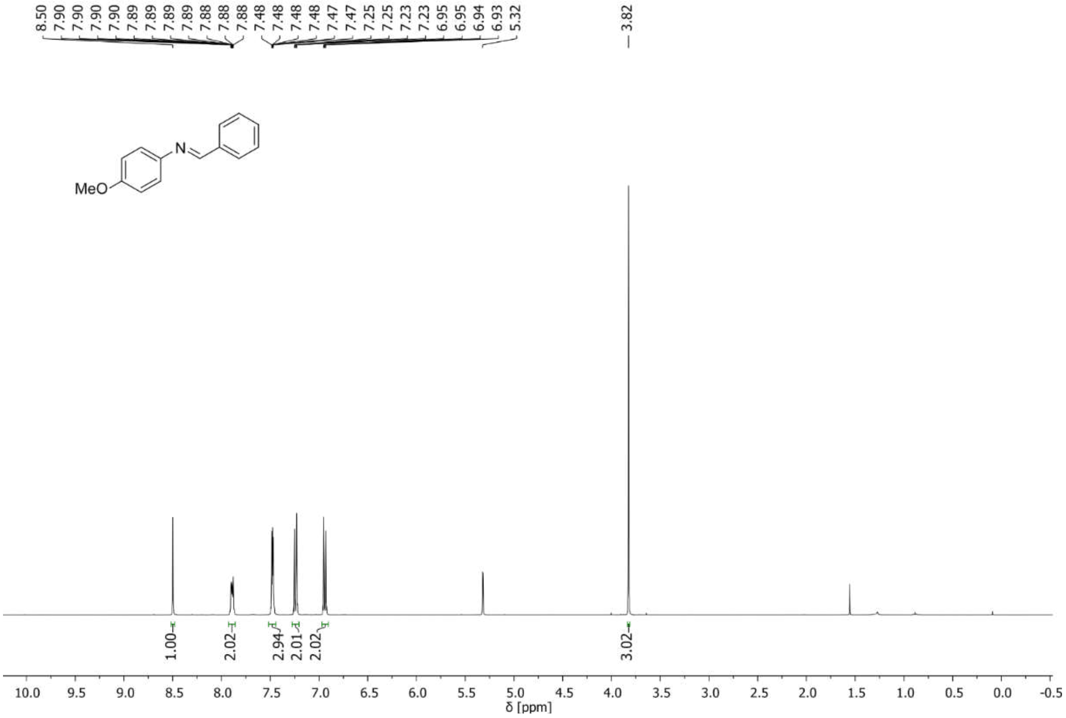
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-iodoaniline (**1g**)



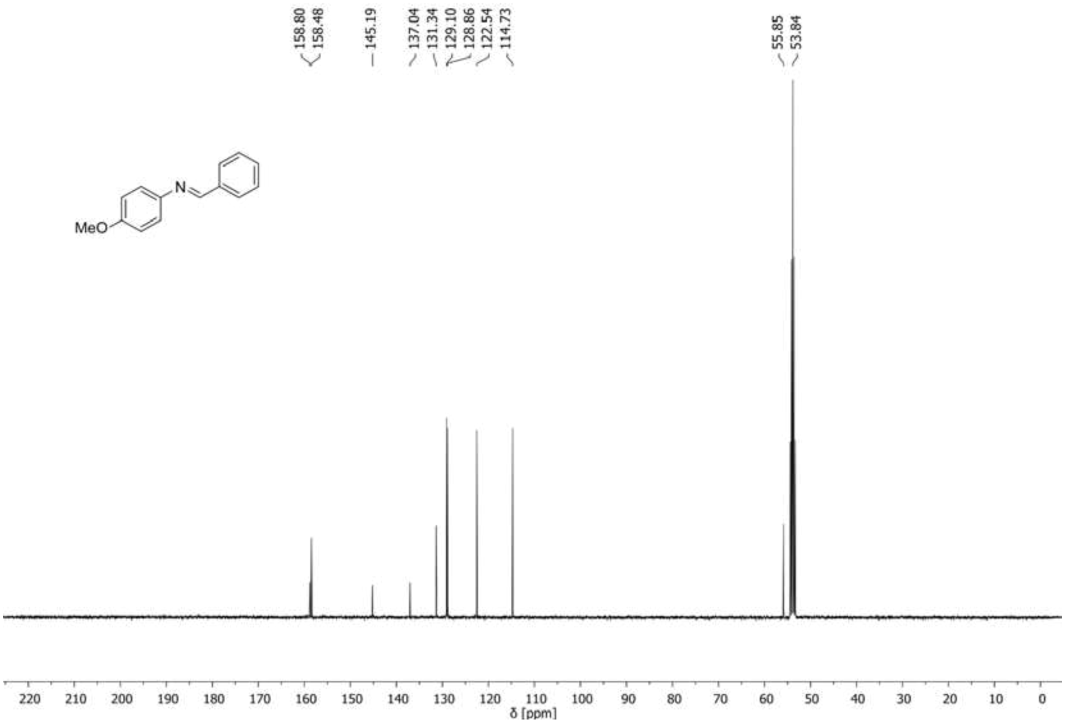
^{1}H NMR: (E)-4-(benzylideneamino)benzonitrile (**1h**)



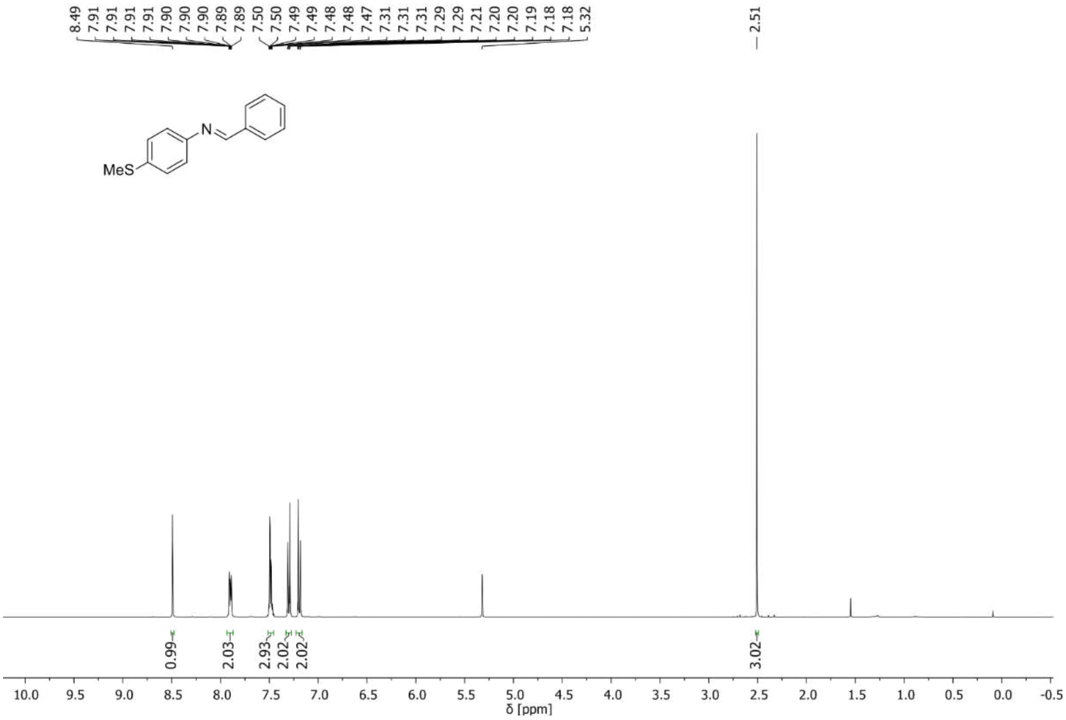
^{13}C{^{1}H} NMR: (E)-4-(benzylideneamino)benzonitrile (**1h**)



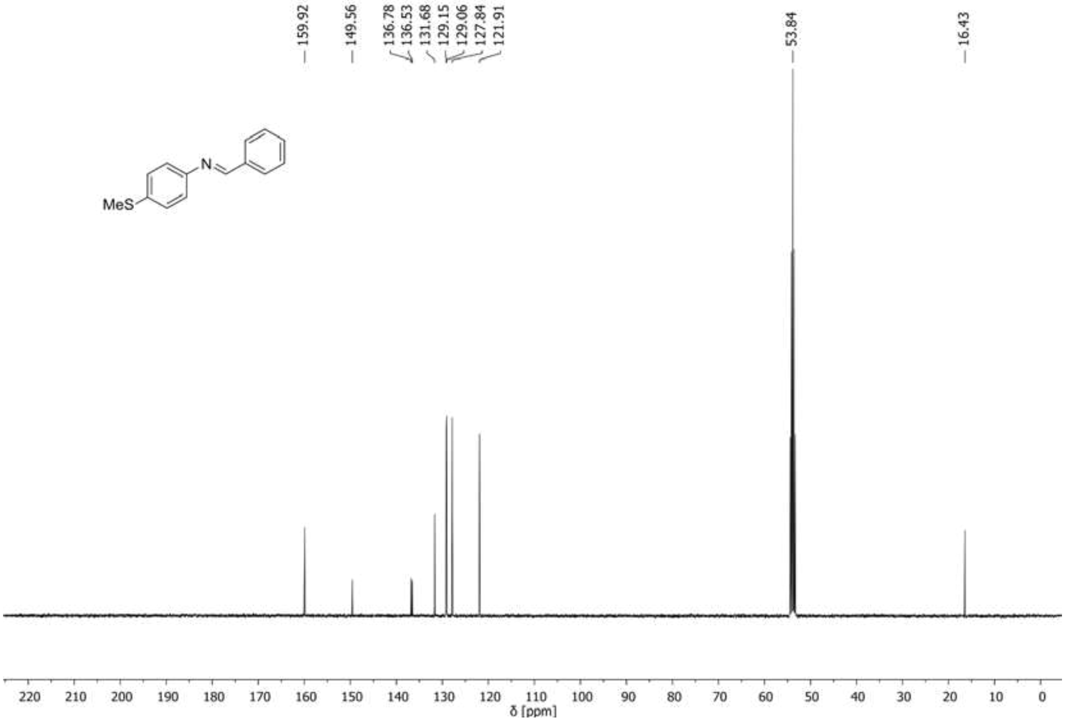
^{1}H NMR: (E)-N-benzylidene-4-methoxyaniline (**1i**)



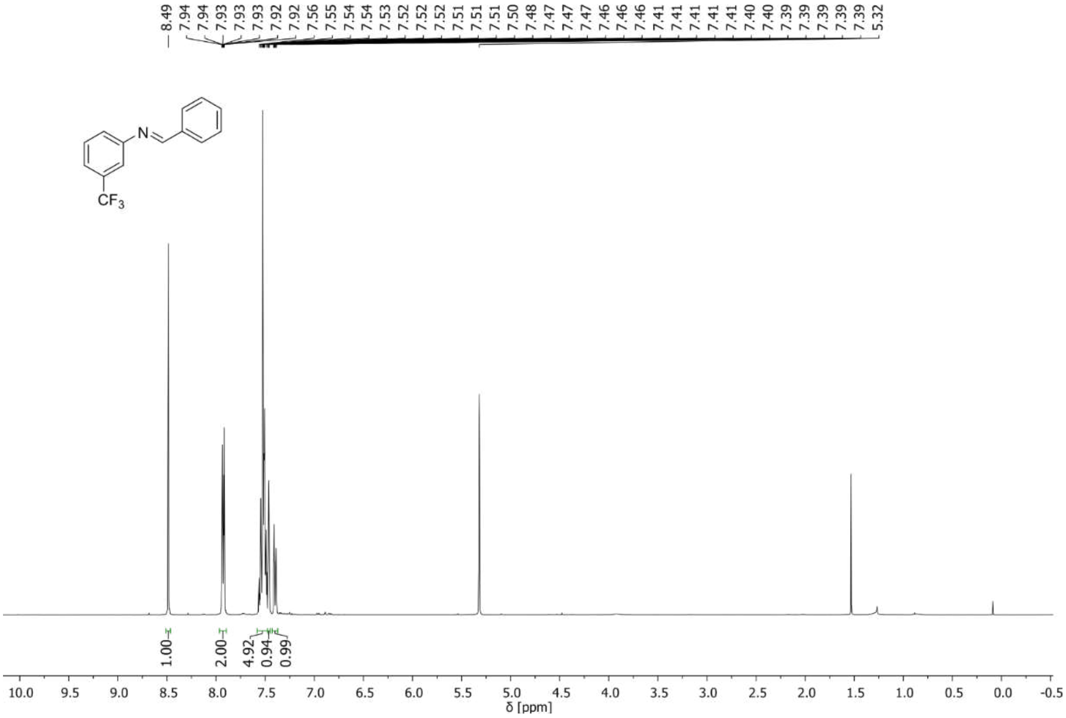
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-methoxyaniline (**1i**)



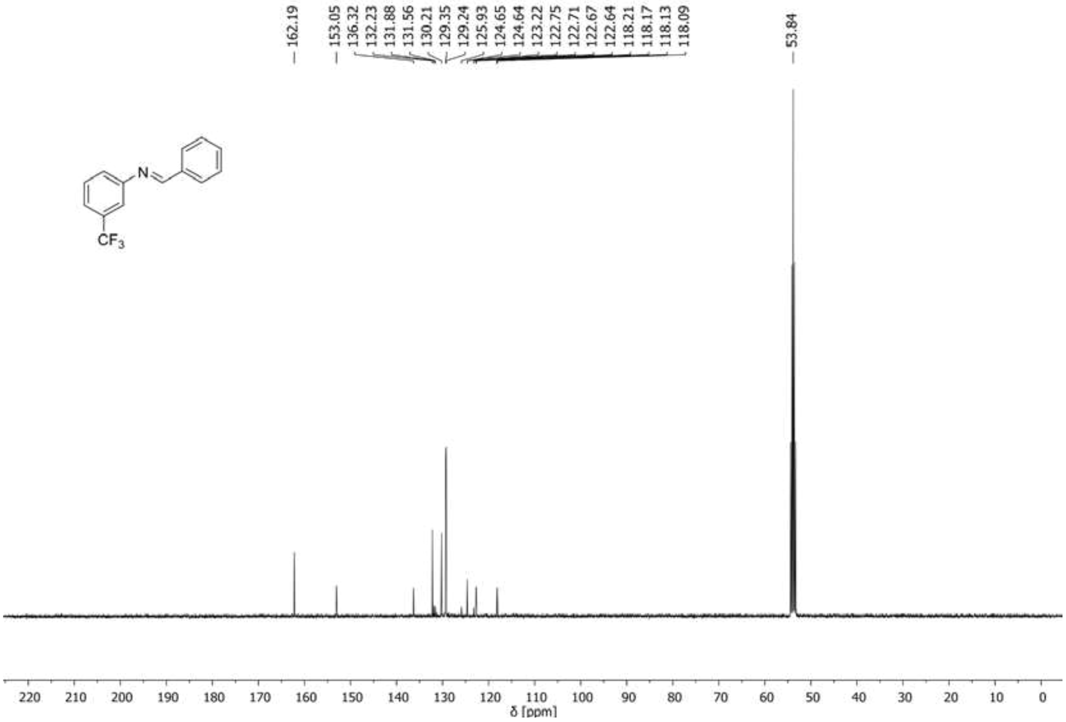
^{1}H NMR: (E)-N-benzylidene-4-(methylthio)aniline (**1j**)



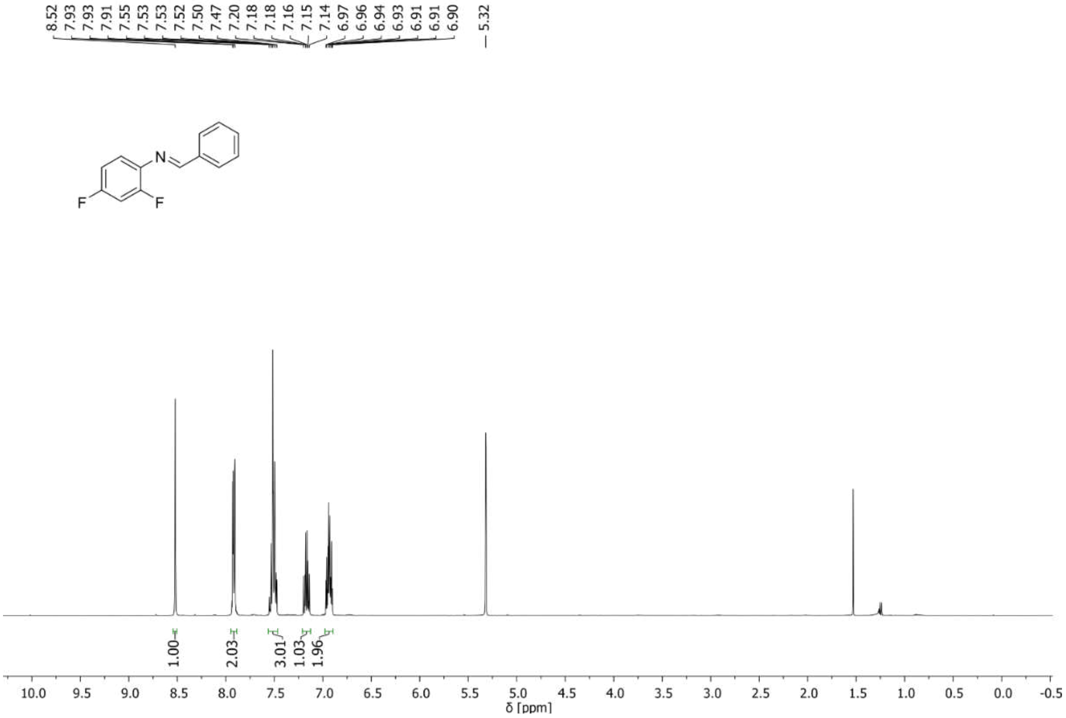
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-4-(methylthio)aniline (**1j**)



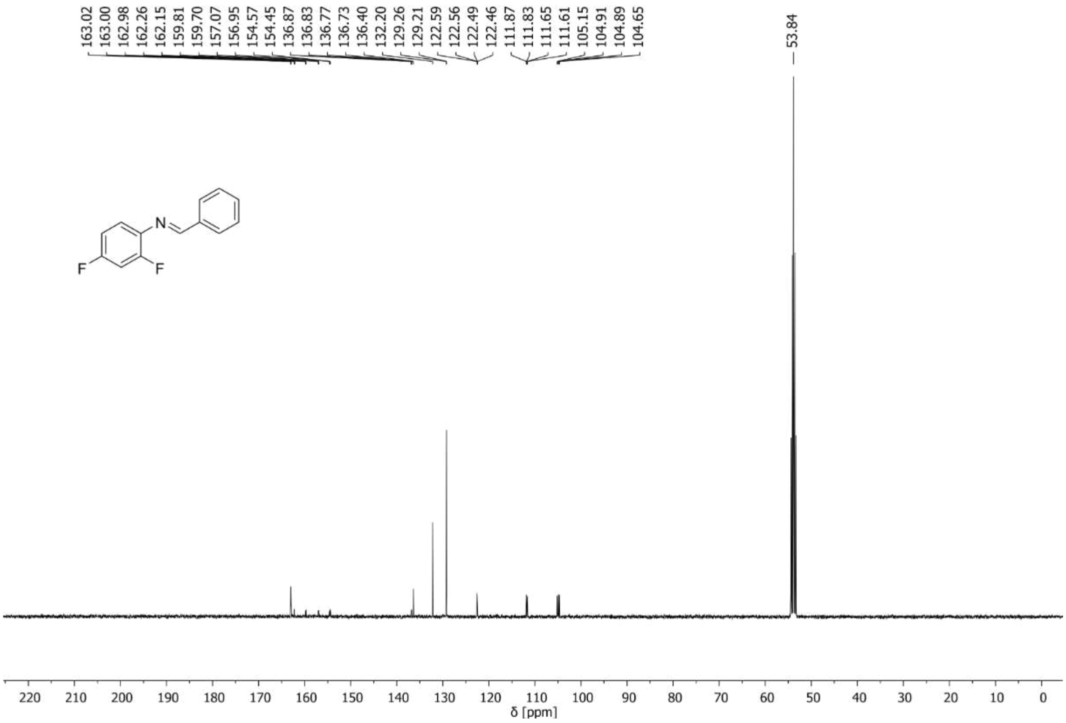
^{1}H NMR: (E)-N-benzylidene-3-(trifluoromethyl)aniline (**1k**)



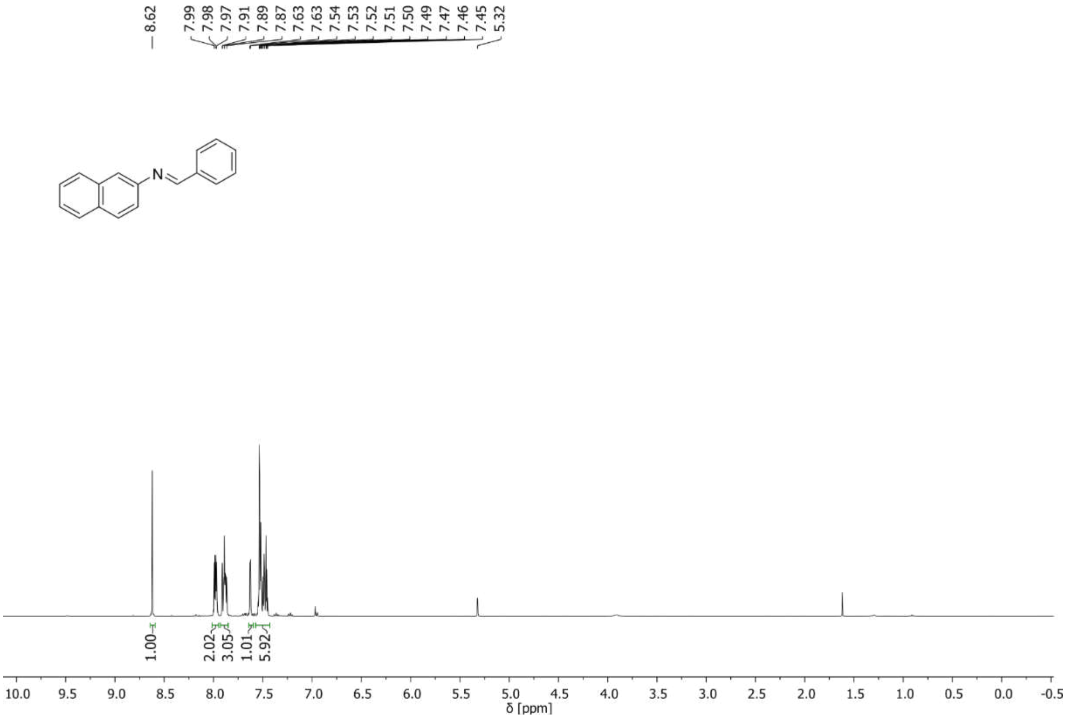
^{13}C{^{1}H} NMR: (E)-N-benzylidene-3-(trifluoromethyl)aniline (**1k**)



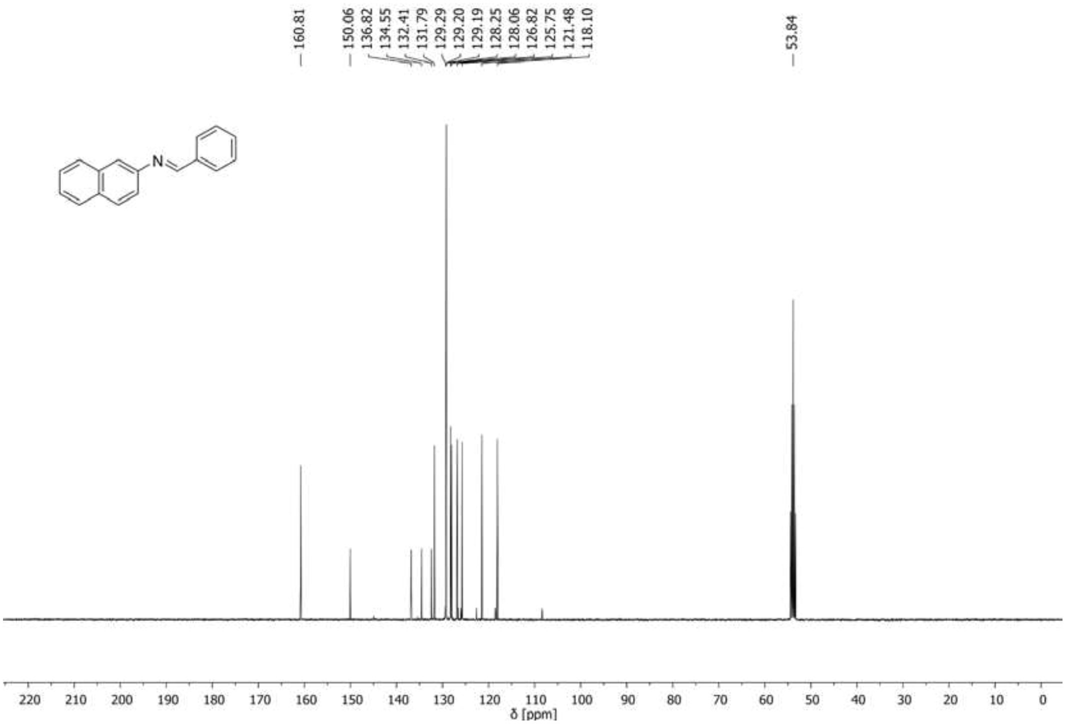
^{1}H NMR: (E)-N-benzylidene-2,4-difluoroaniline (**1l**)



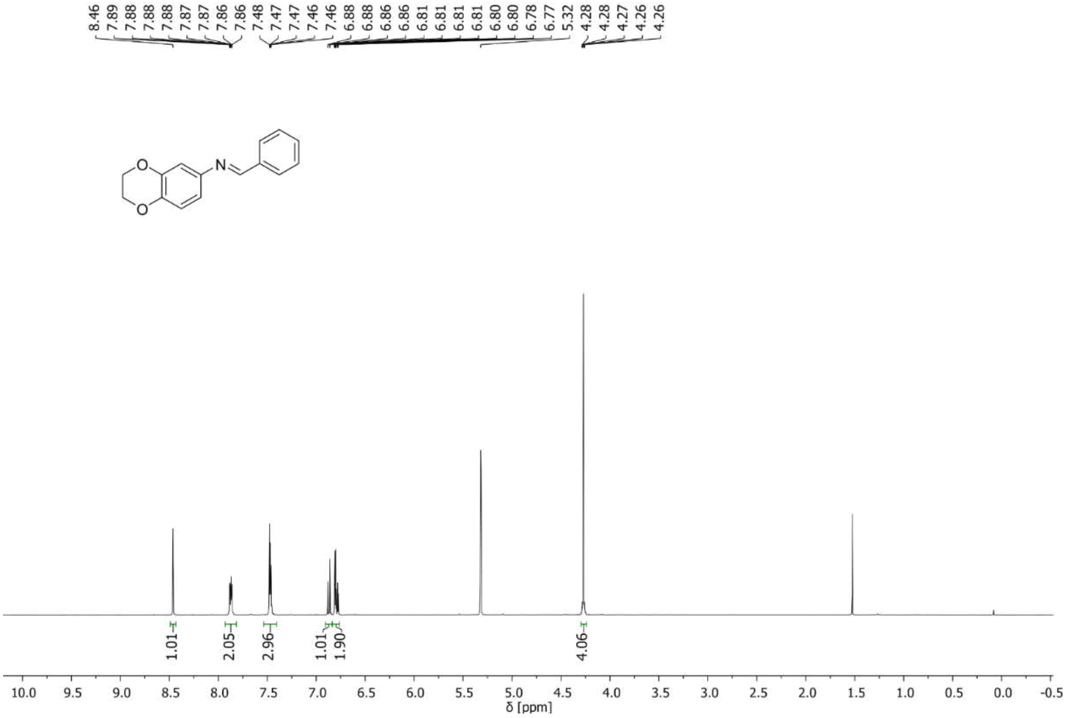
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-2,4-difluoroaniline (**1l**)



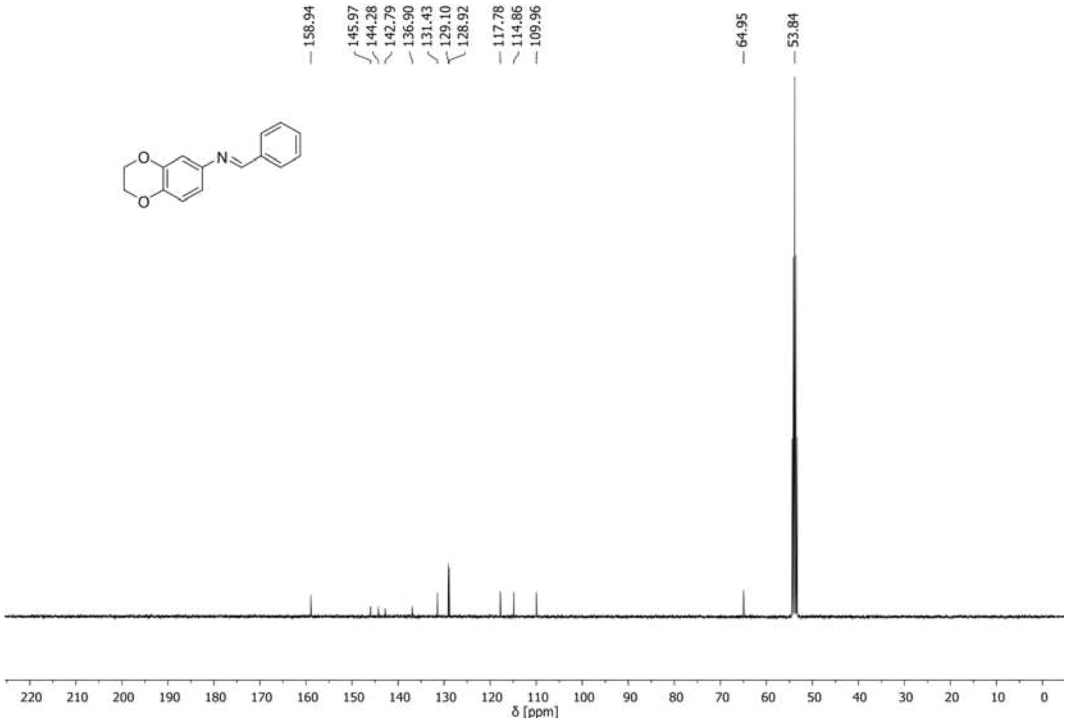
^{1}H NMR: (E)-N-benzylidene-2-naphthylamine (**1m**)



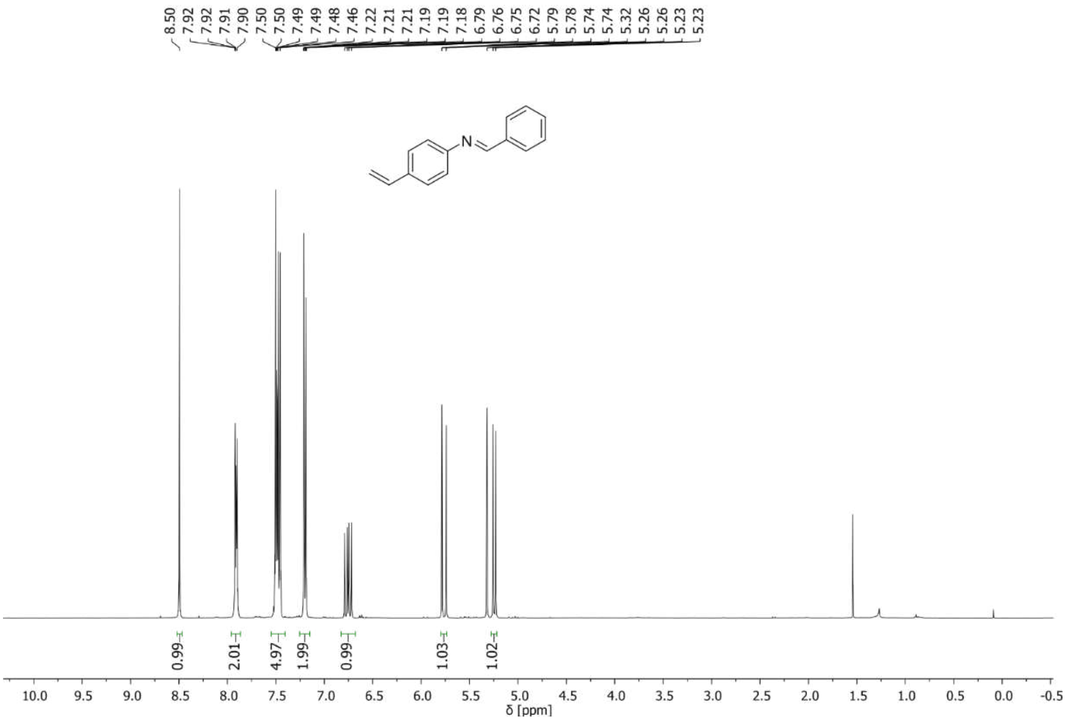
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-2-naphthylamine (**1m**)



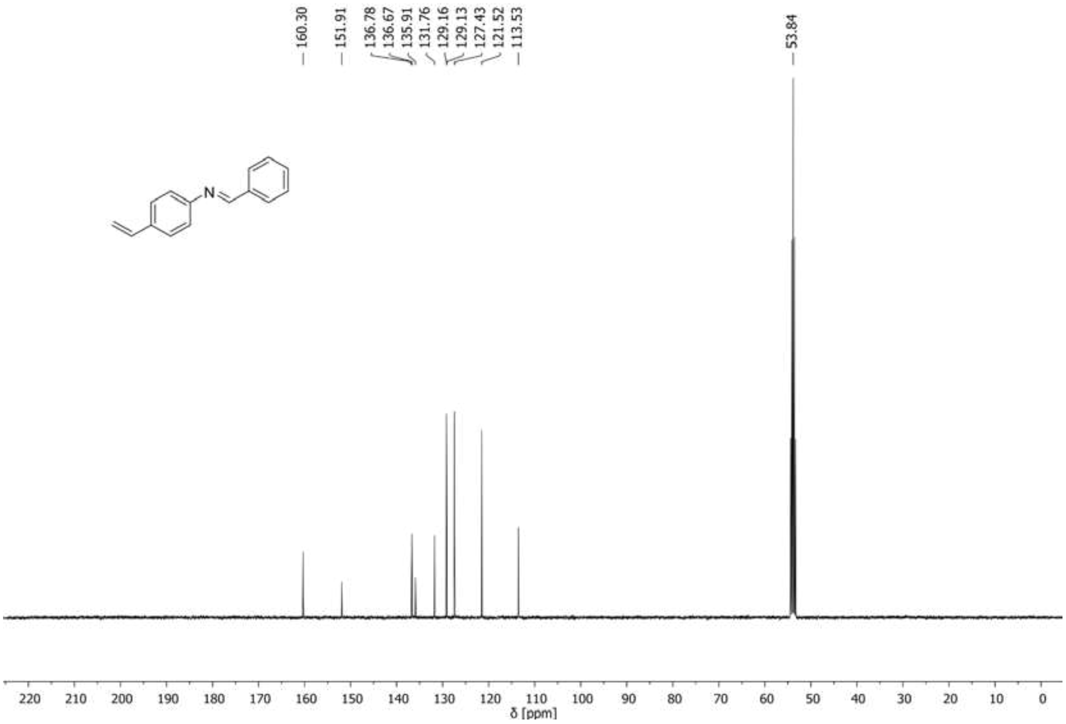
^{1}H NMR: (E)-N-benzylidene-2,3-dihydrobenzo[b][1,4]dioxin-6-amine (**1n**)



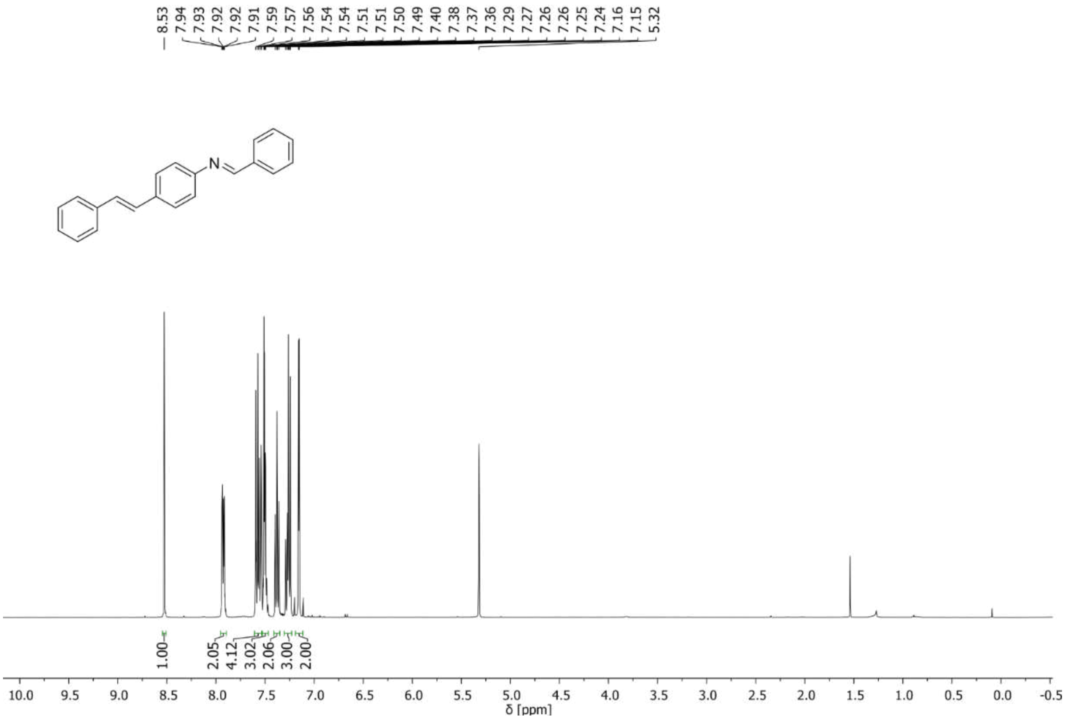
$^{13}C\{^{1}H\}$ NMR: (E)-N-benzylidene-2,3-dihydrobenzo[b][1,4]dioxin-6-amine (**1n**)



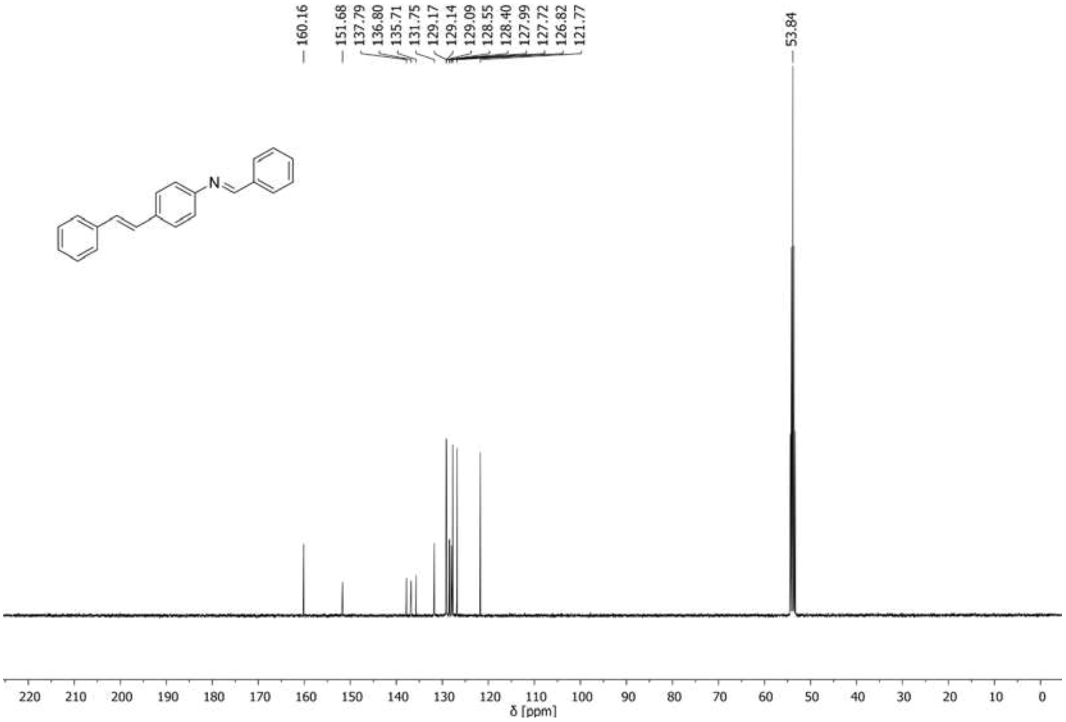
^{1}H NMR: (E)-N-benzylidene-4-vinylaniline (**1o**)



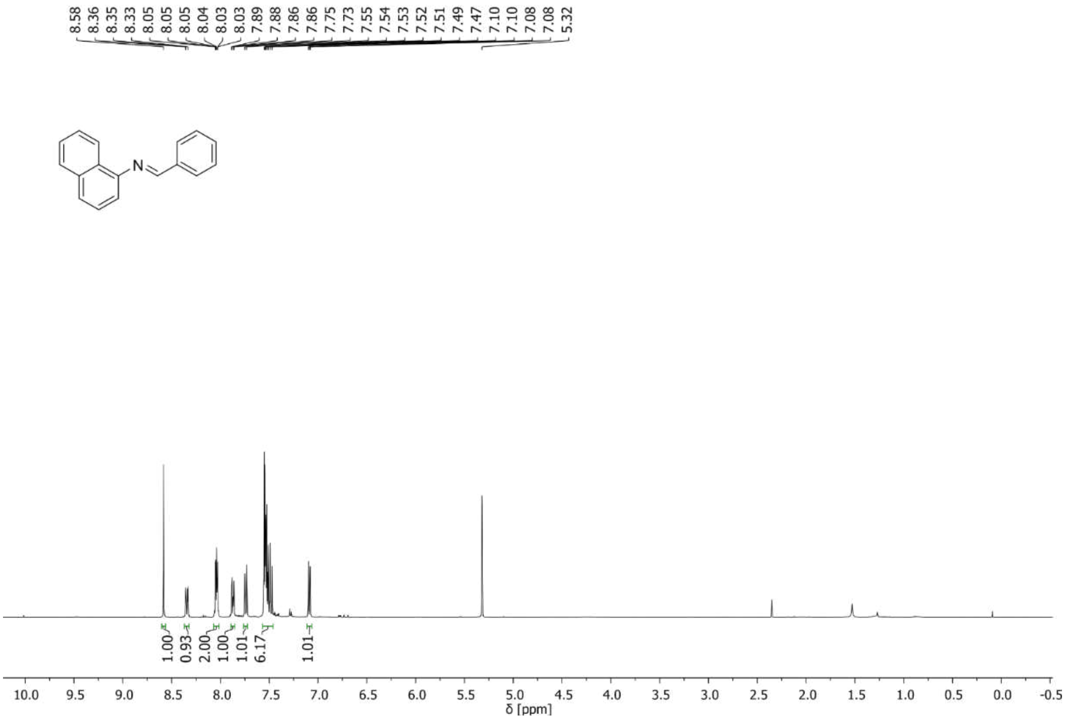
^{13}C{^{1}H} NMR: (E)-N-benzylidene-4-vinylaniline (**1o**)



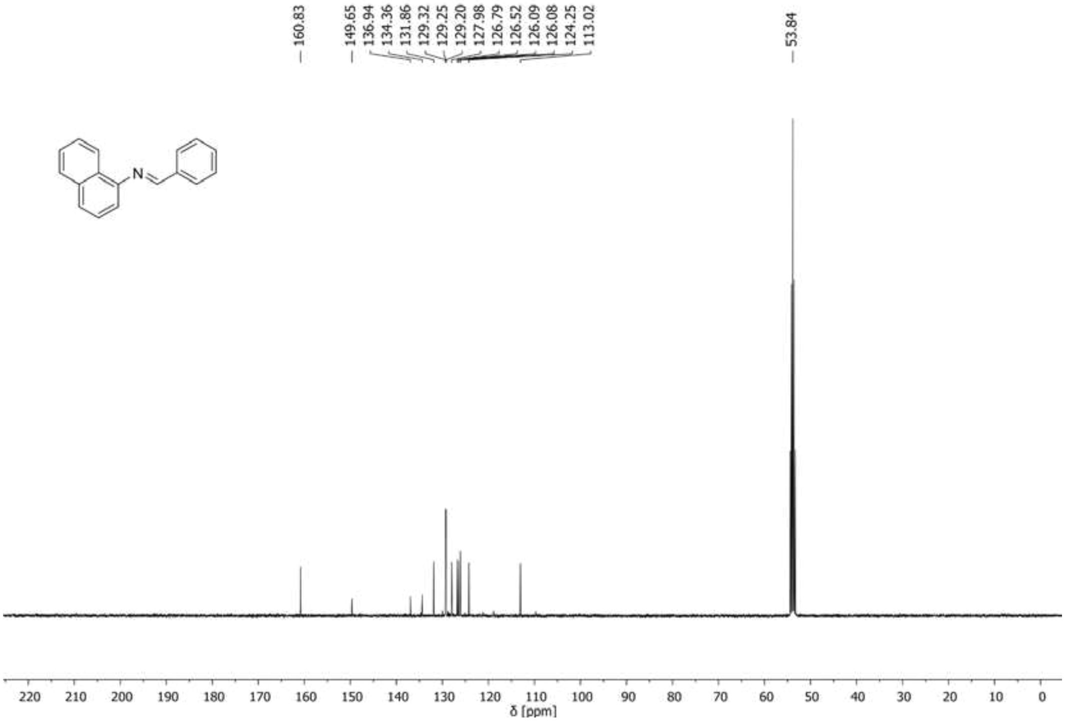
^{1}H NMR: (E)-N-benzylidene-4-aminostilbene (**1p**)



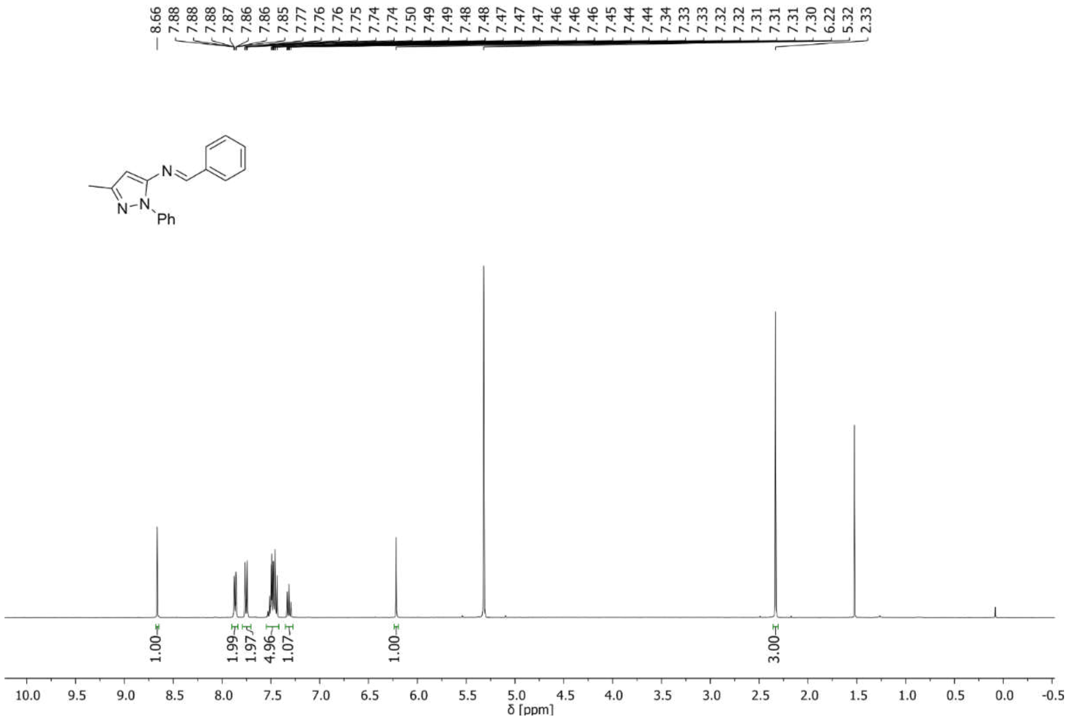
^{13}C{^{1}H} NMR: (E)-N-benzylidene-4-aminostilbene (**1p**)



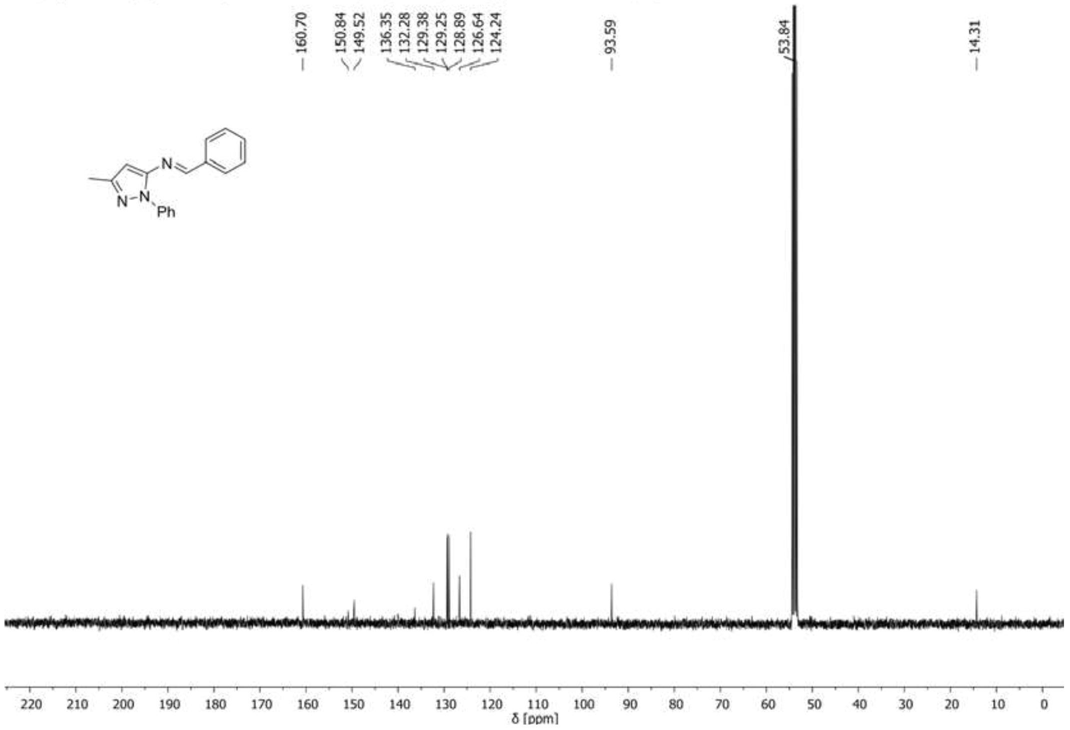
^{1}H NMR: (E)-N-benzylidene-4-naphthylamine (**1q**)



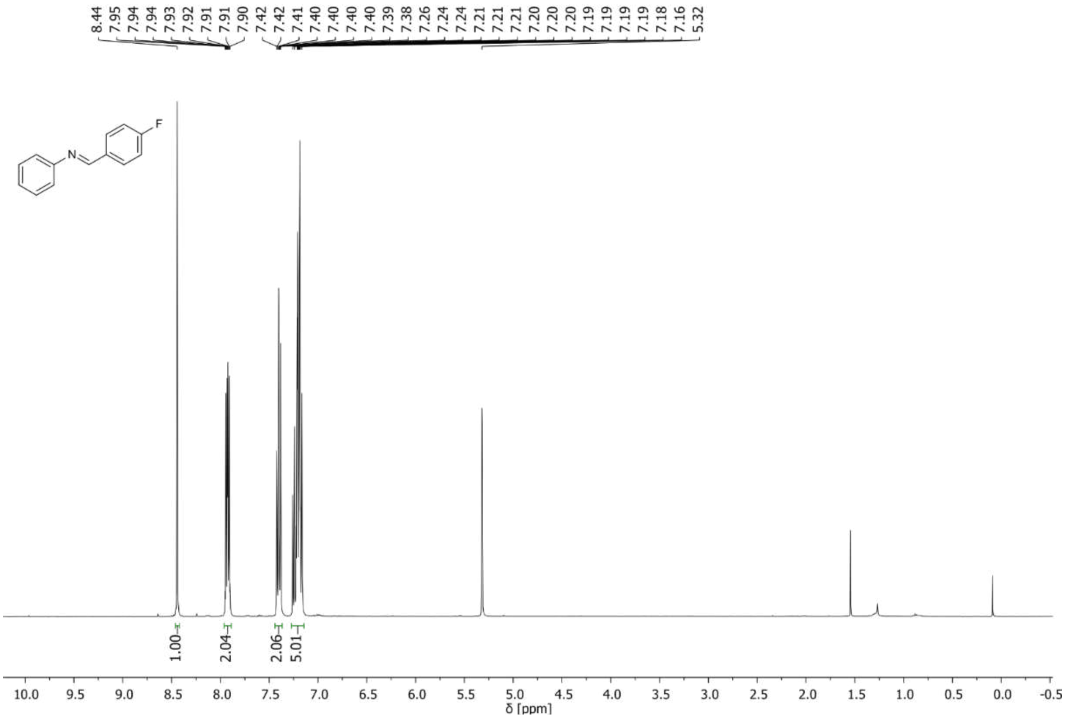
^{13}C{^{1}H} NMR: (E)-N-benzylidene-4-naphthylamine (**1q**)



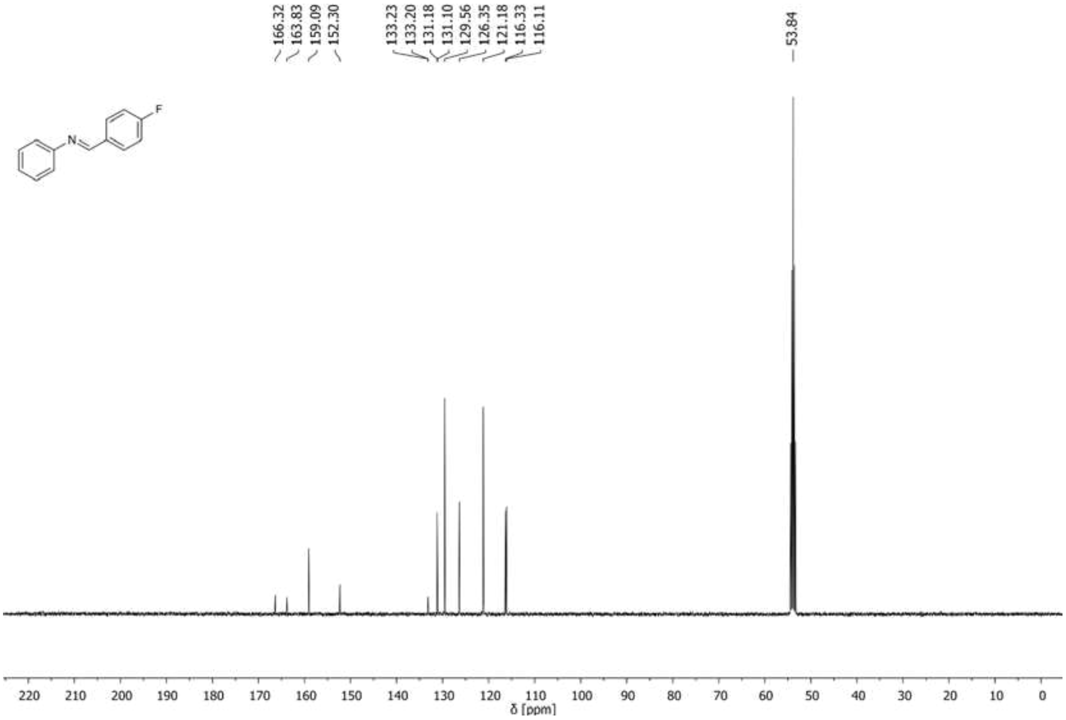
^{1}H NMR: (E)-N-benzylidene-3-methyl-1-phenyl-1H-pyrazol-5-amine (**1r**)



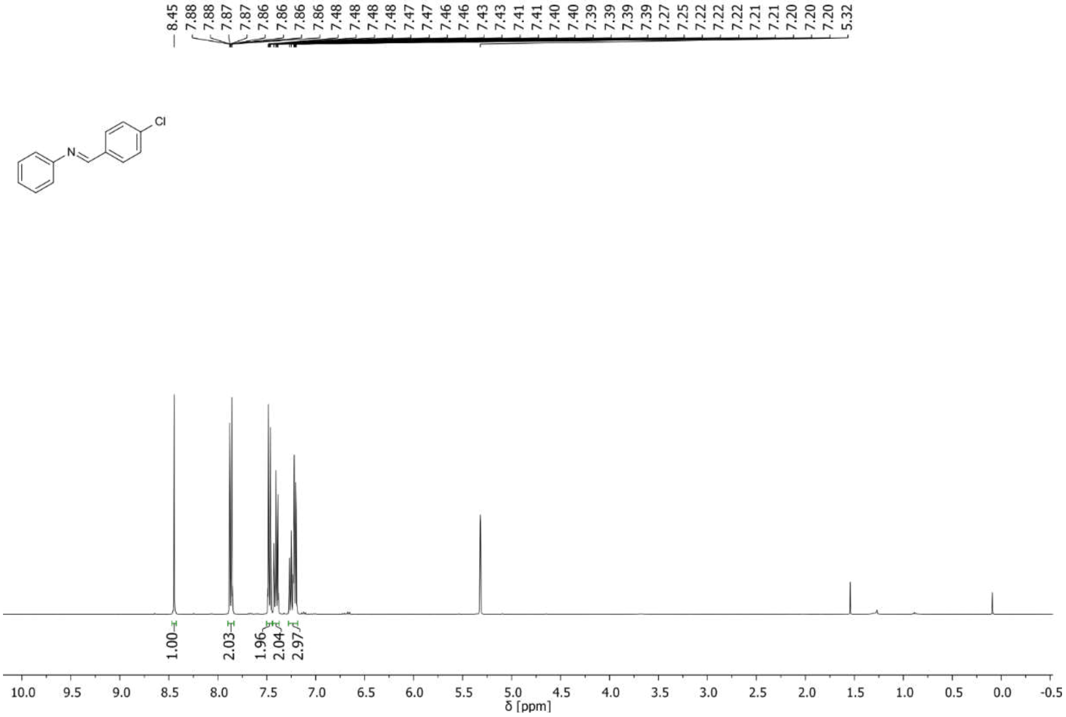
^{13}C{^{1}H} NMR: (E)-N-benzylidene-3-methyl-1-phenyl-1H-pyrazol-5-amine (**1r**)



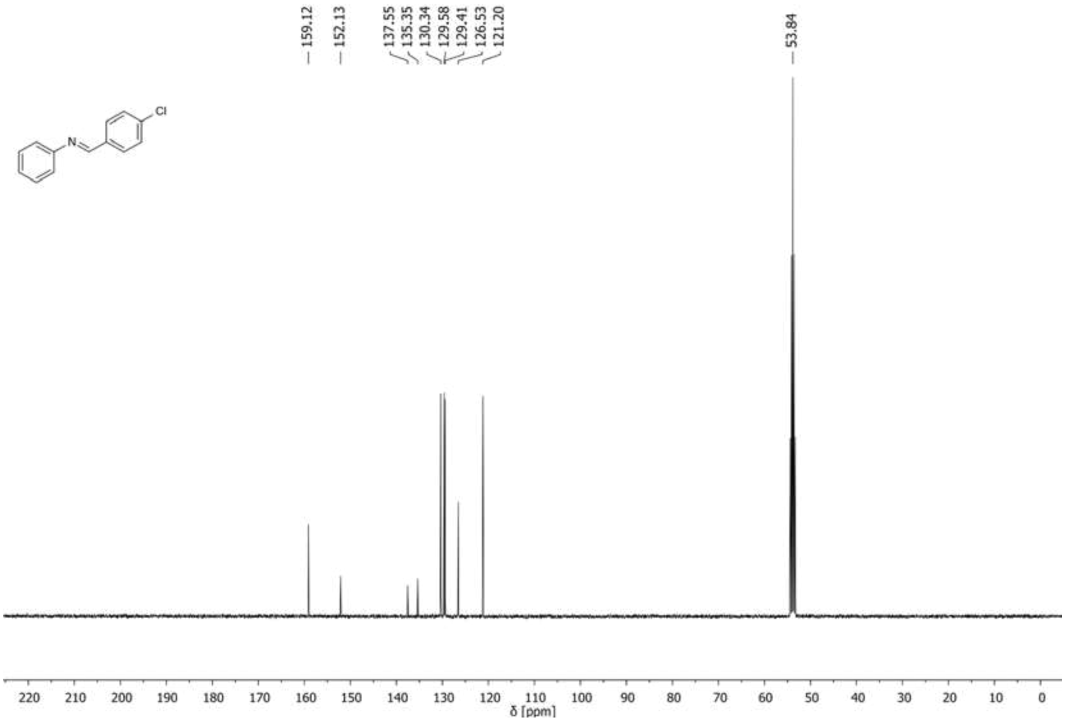
^{1}H NMR: (E)-N-(4-fluorobenzylidene)aniline (**1s**)



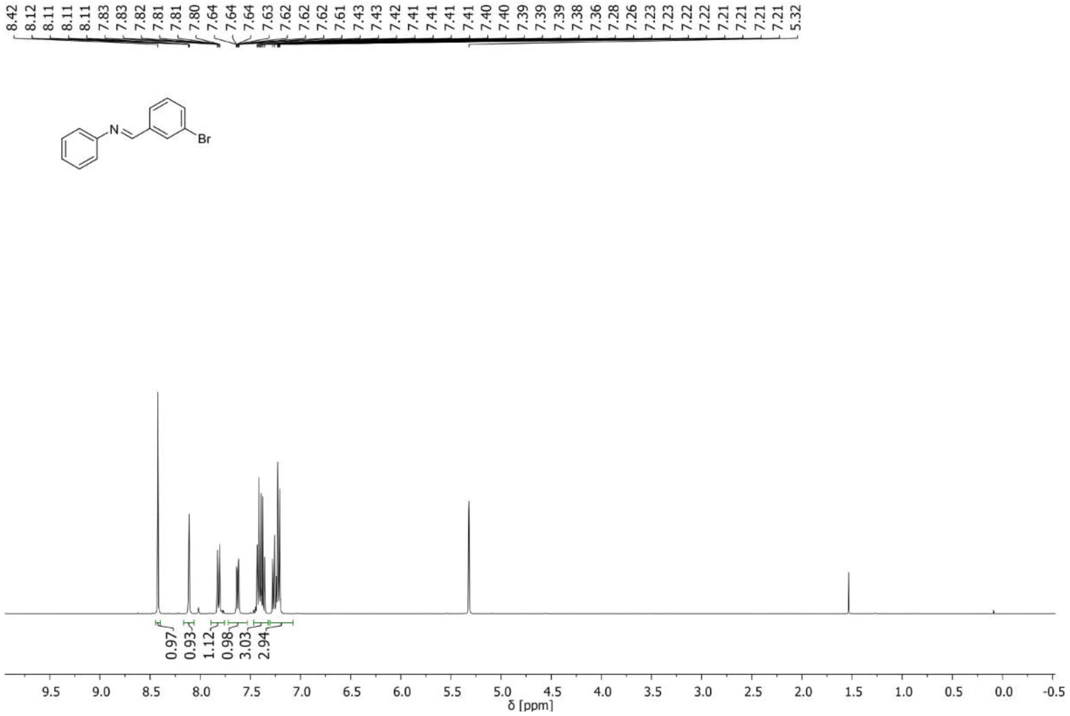
^{13}C{^{1}H} NMR: (E)-N-(4-fluorobenzylidene)aniline (**1s**)



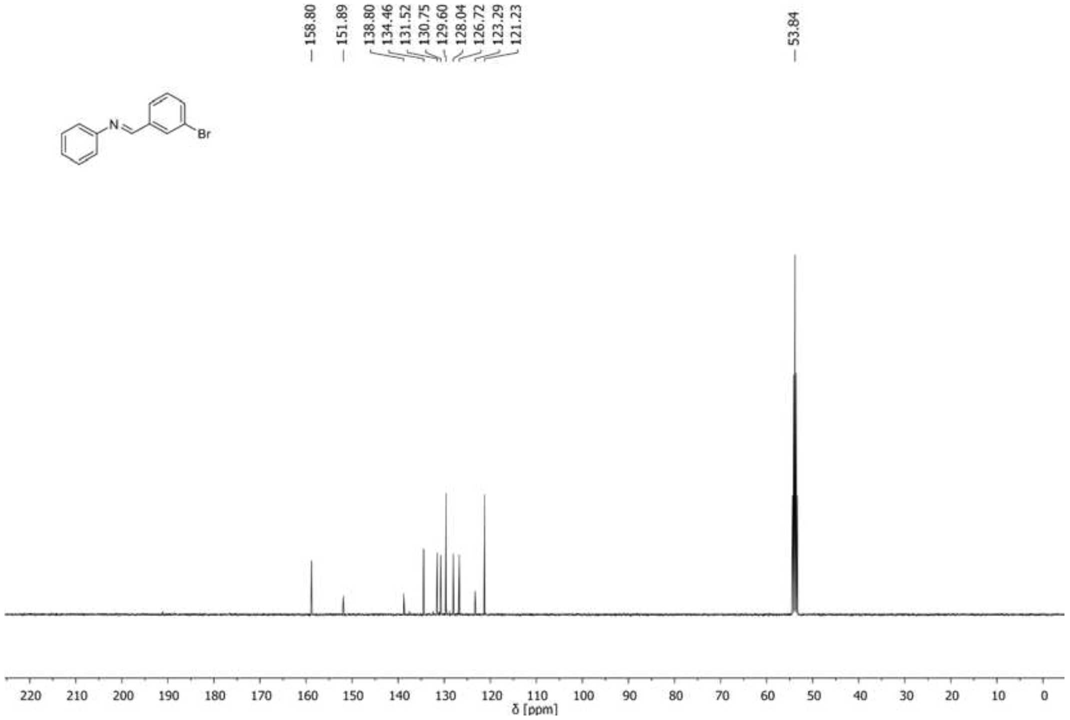
^{1}H NMR: (E)-N-(4-chlorobenzylidene)aniline (**1t**)



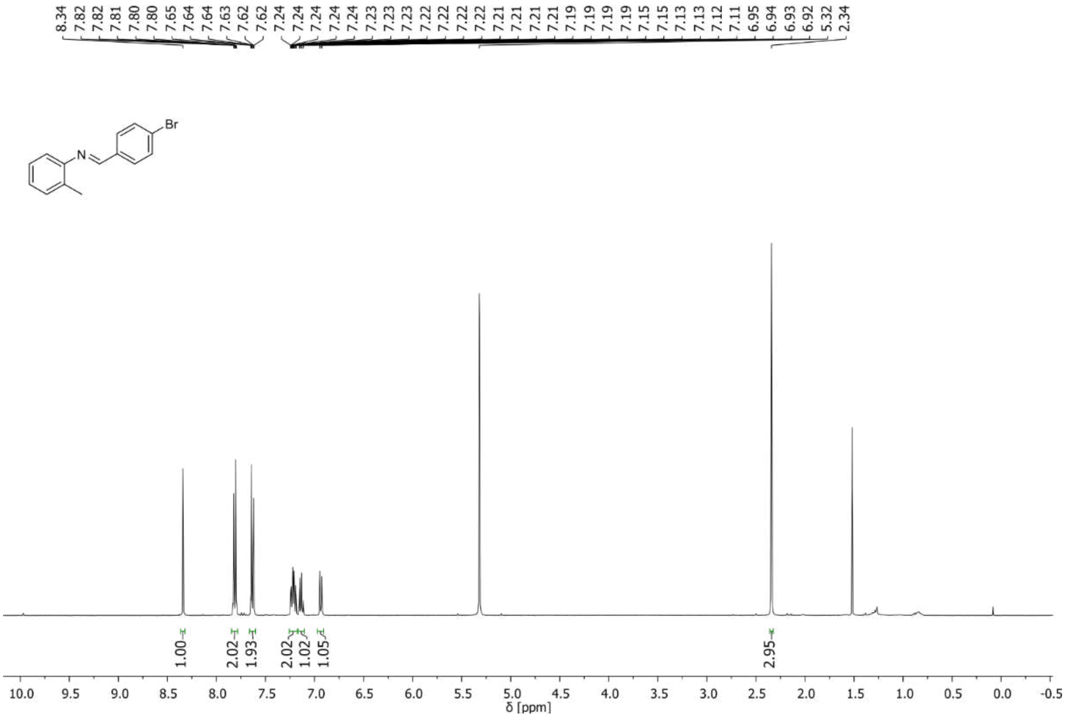
$^{13}C\{^{1}H\}$ NMR: (E)-N-(4-chlorobenzylidene)aniline (**1t**)



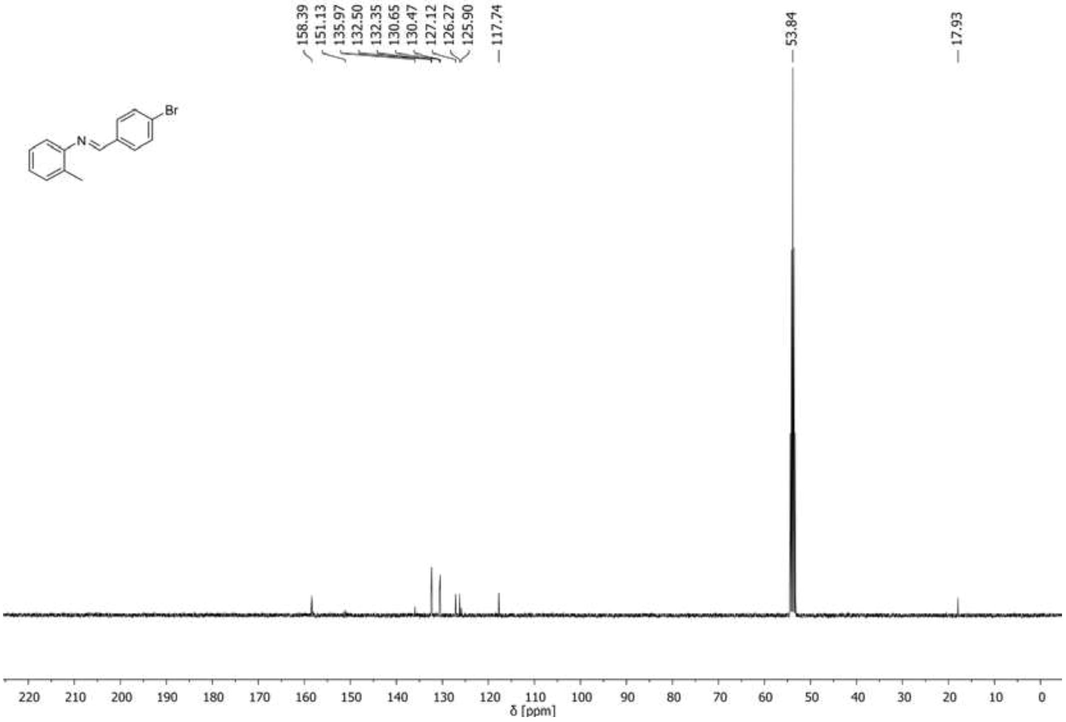
^{1}H NMR: (E)-N-(3-bromobenzylidene)aniline (**1u**)



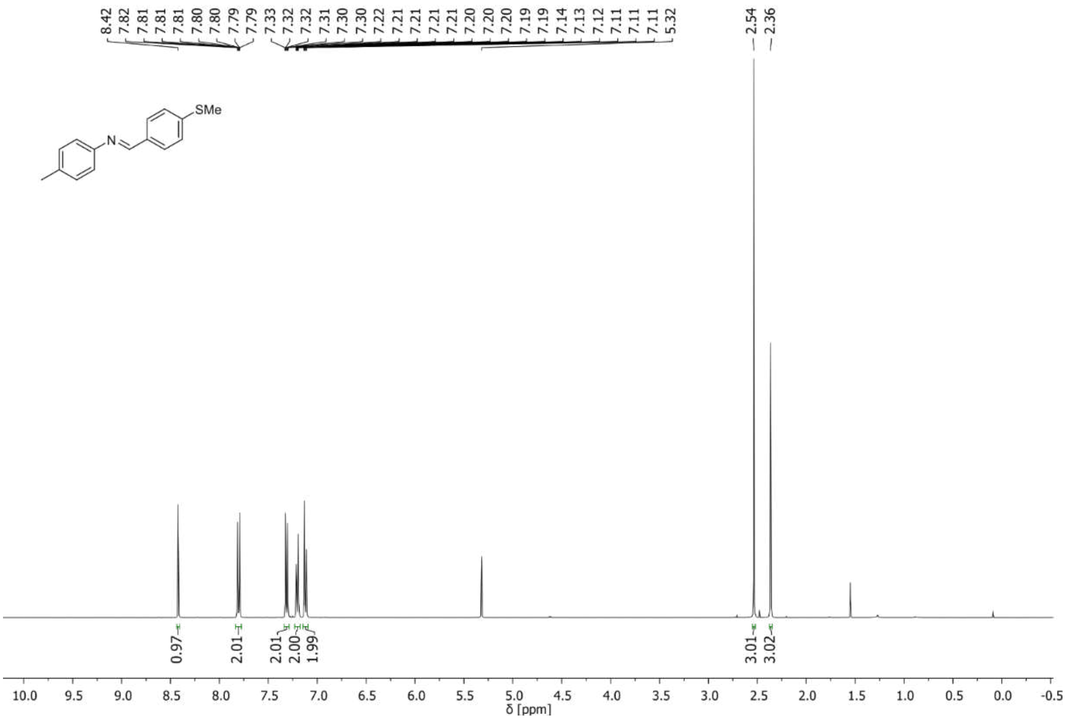
^{13}C{^{1}H} NMR: (E)-N-(3-bromobenzylidene)aniline (**1u**)



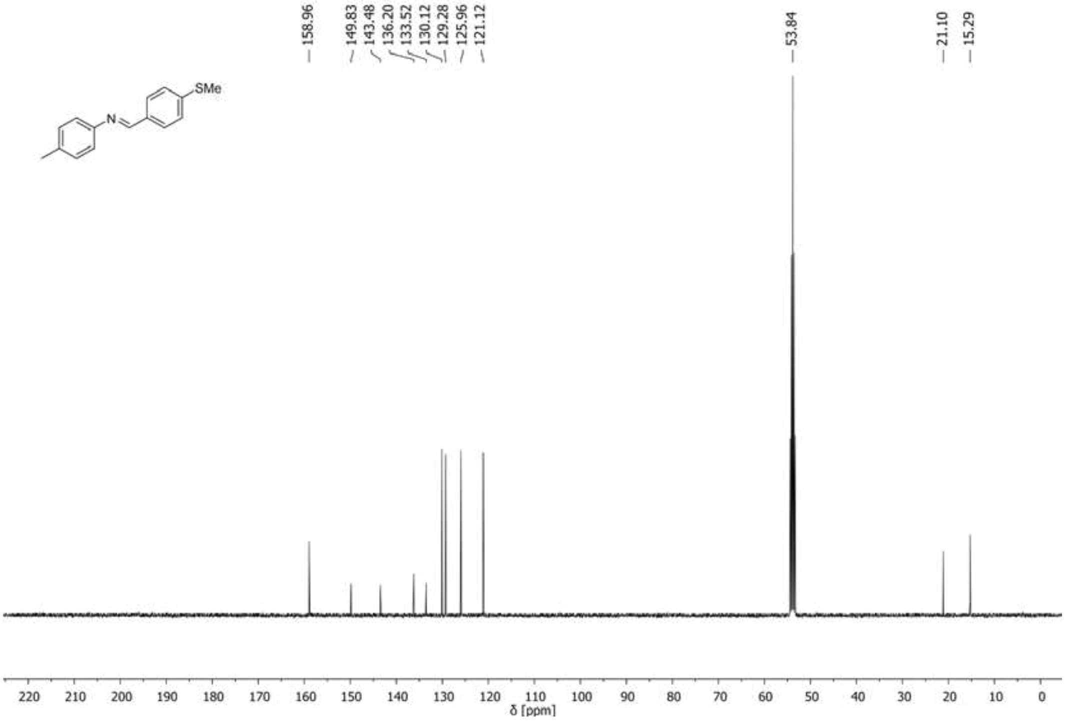
^{1}H NMR: (E)-1-(4-bromophenyl)-N-(o-tolyl)methanimine (**1v**)



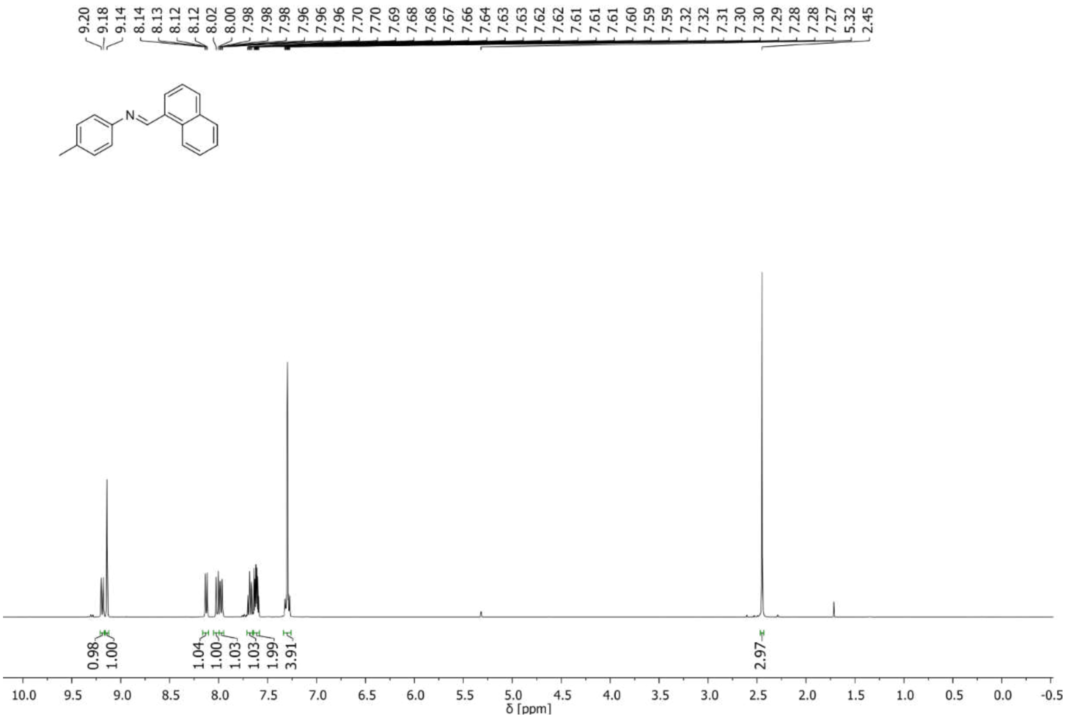
^{13}C{^{1}H} NMR: (E)-1-(4-bromophenyl)-N-(o-tolyl)methanimine (**1v**)



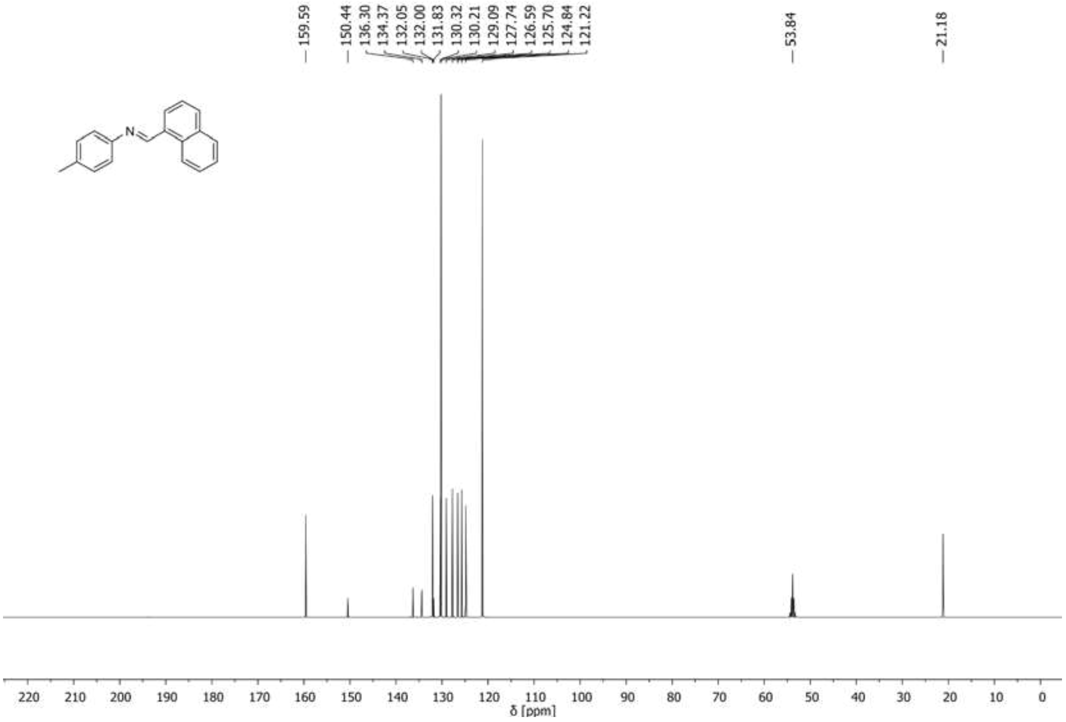
^{1}H NMR: (E)-1-(4-(methylthio)phenyl)-N-(p-tolyl)methanimine (**1w**)



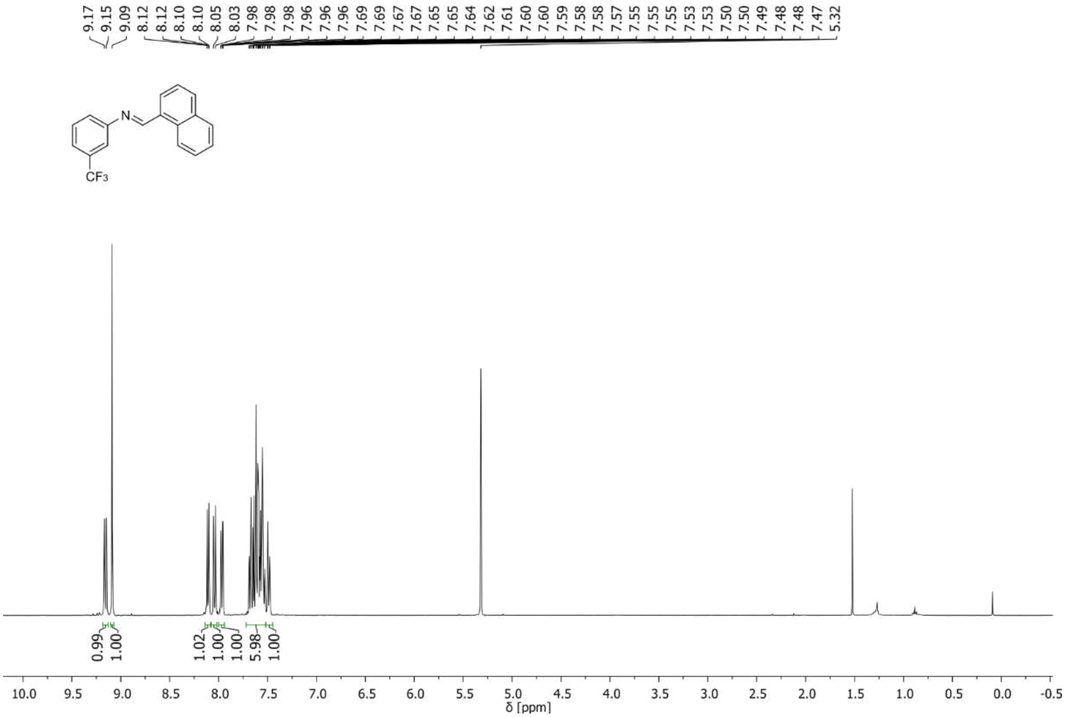
^{13}C{^{1}H} NMR: (E)-1-(4-(methylthio)phenyl)-N-(p-tolyl)methanimine (**1w**)



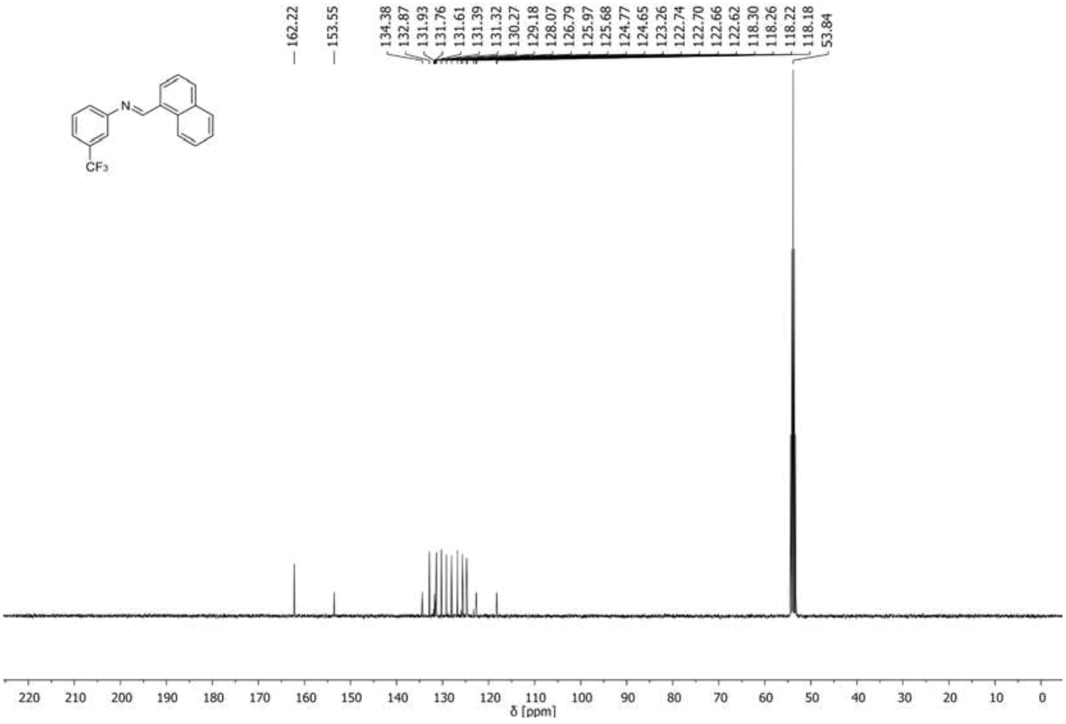
^{1}H NMR: (E)-1-(naphthalen-1-yl)-N-(p-tolyl)methanimine (**1x**)



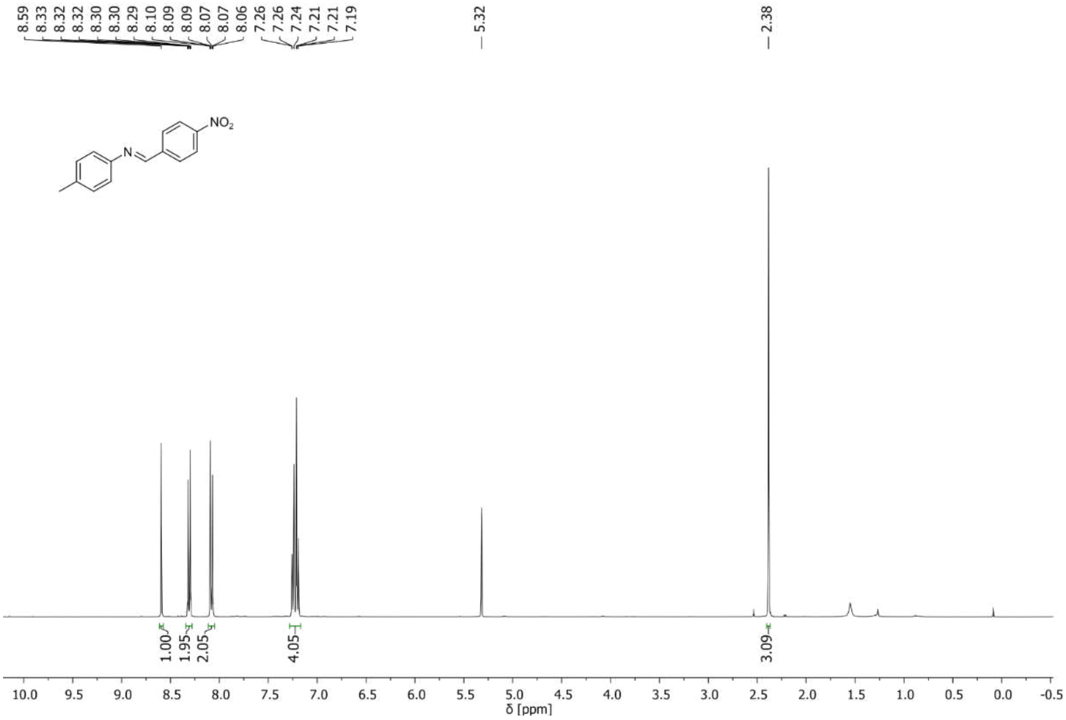
^{13}C{^{1}H} NMR: (E)-1-(naphthalen-1-yl)-N-(p-tolyl)methanimine (**1x**)



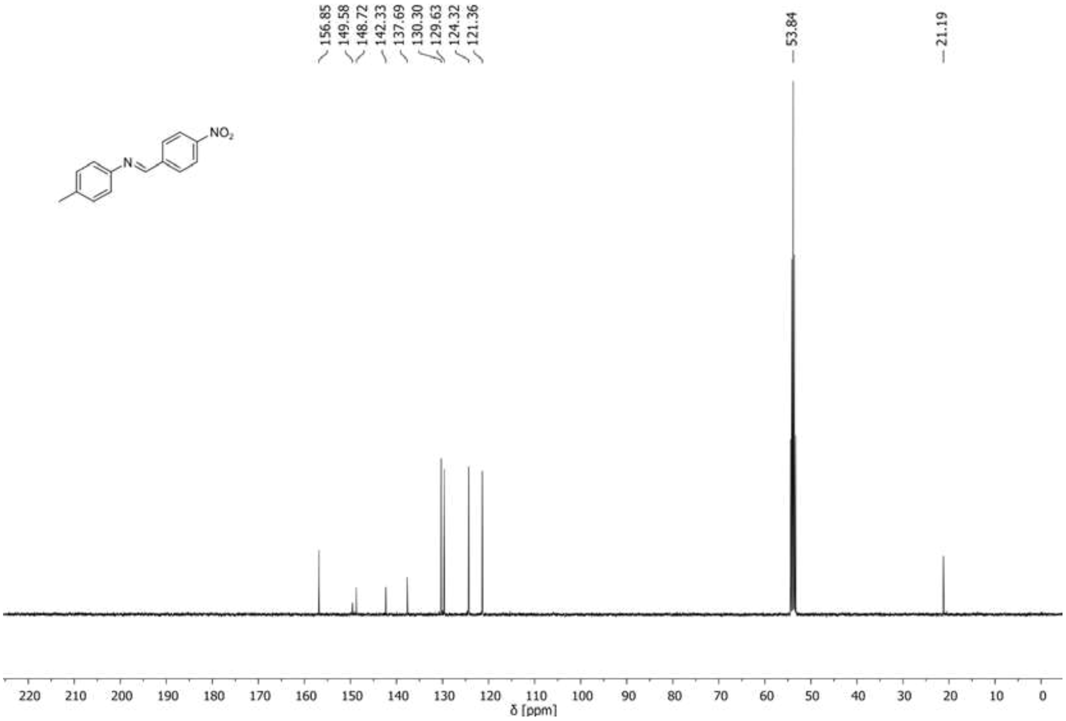
^{1}H NMR: (E)-1-(furan-2-yl)-N-(3-(trifluoromethyl)phenyl)methanimine (**1y**)



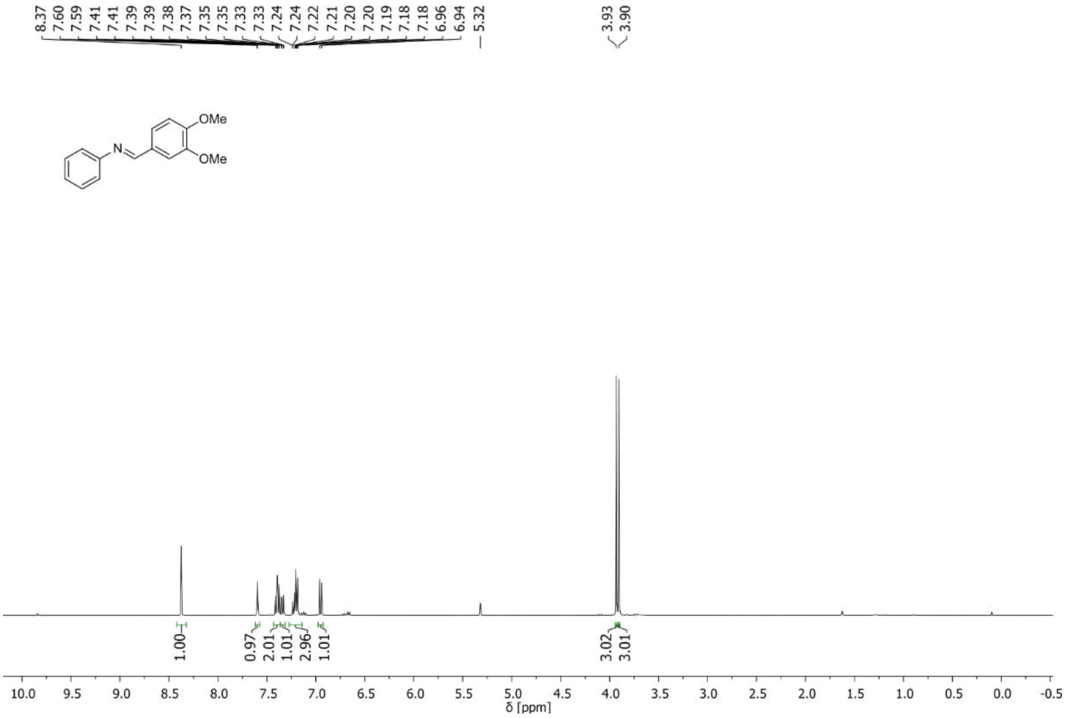
^{13}C{^{1}H} NMR: (E)-1-(furan-2-yl)-N-(3-(trifluoromethyl)phenyl)methanimine (**1y**)



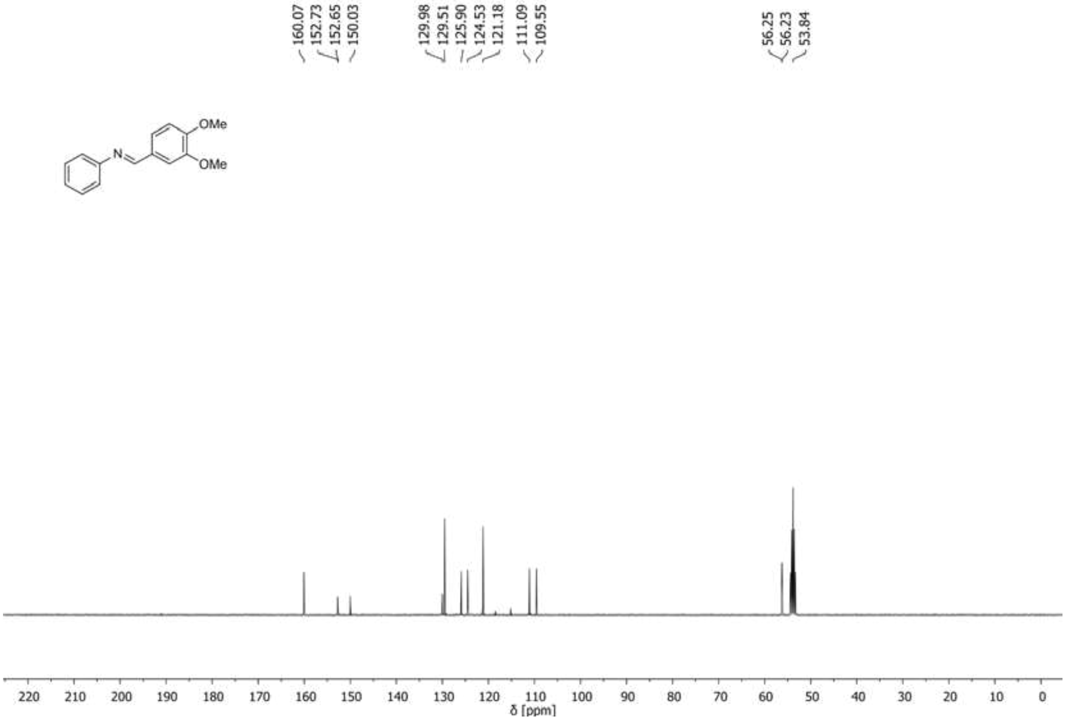
^{1}H NMR: (E)-4-methyl-N-(4-nitrobenzylidene)aniline (**1z**)



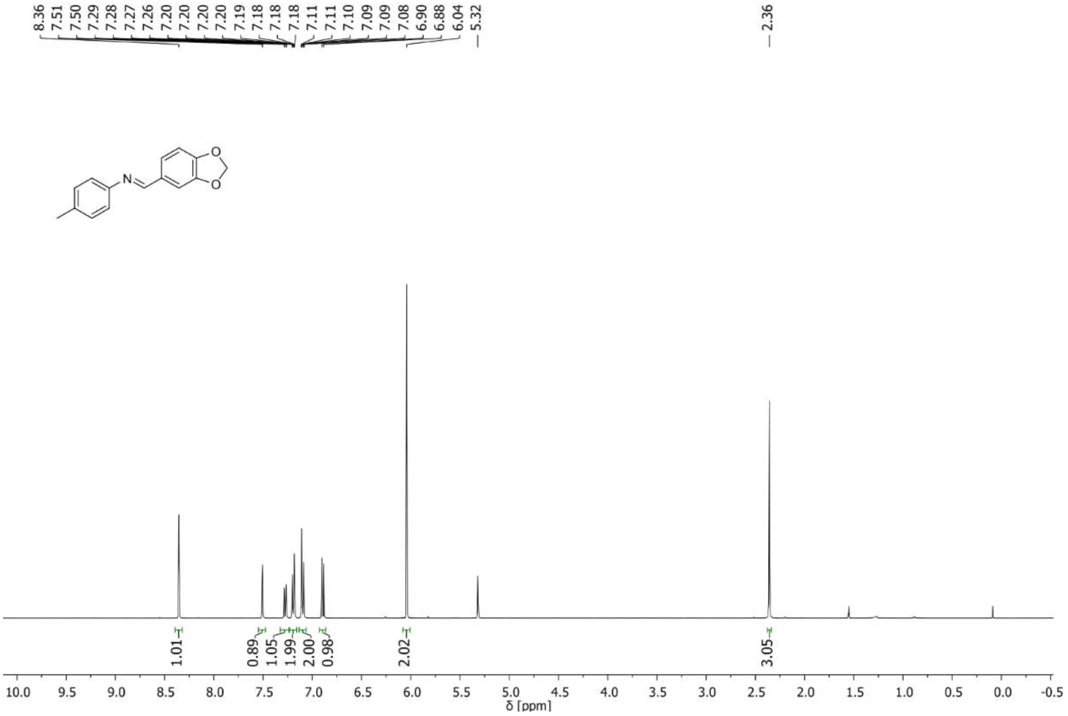
^{13}C{^{1}H} NMR: (E)-4-methyl-N-(4-nitrobenzylidene)aniline (**1z**)



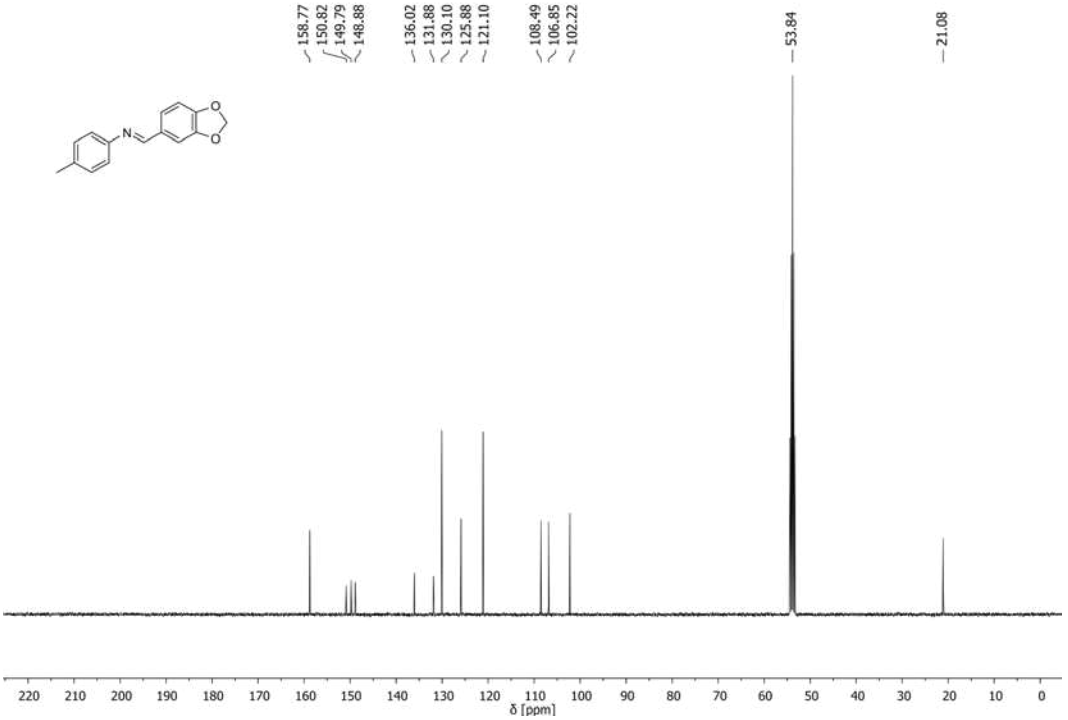
^{1}H NMR: (E)-N-(3,4-Dimethoxybenzylidene)aniline (**1aa**)



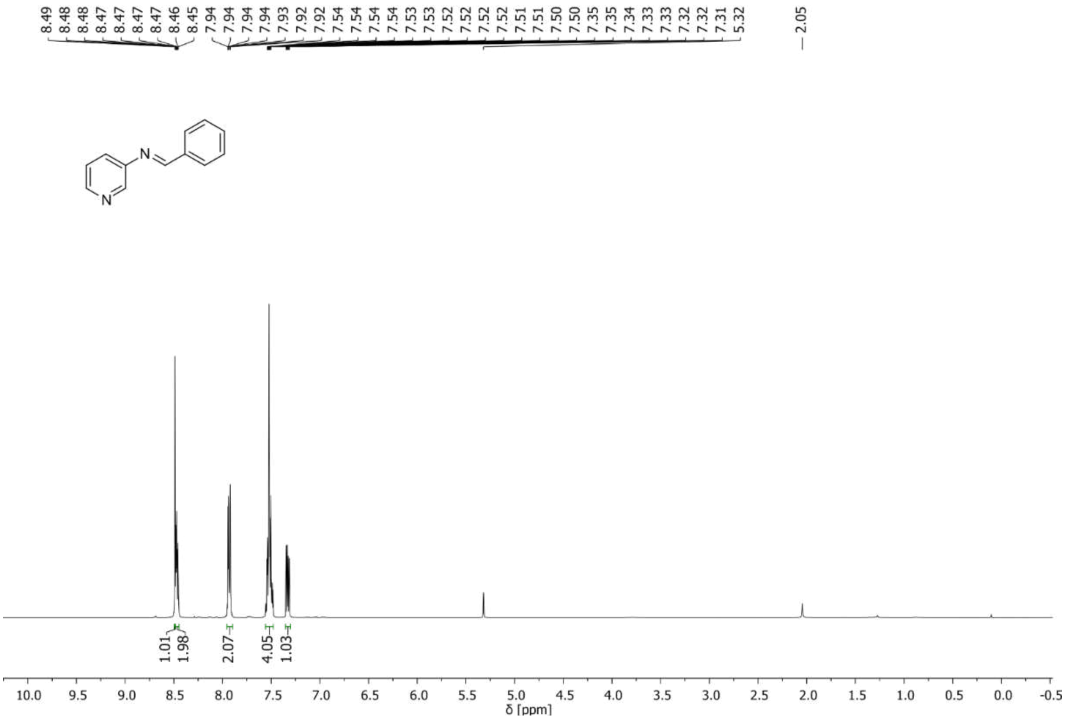
^{13}C{^{1}H} NMR: (E)-N-(3,4-Dimethoxybenzylidene)aniline (**1aa**)



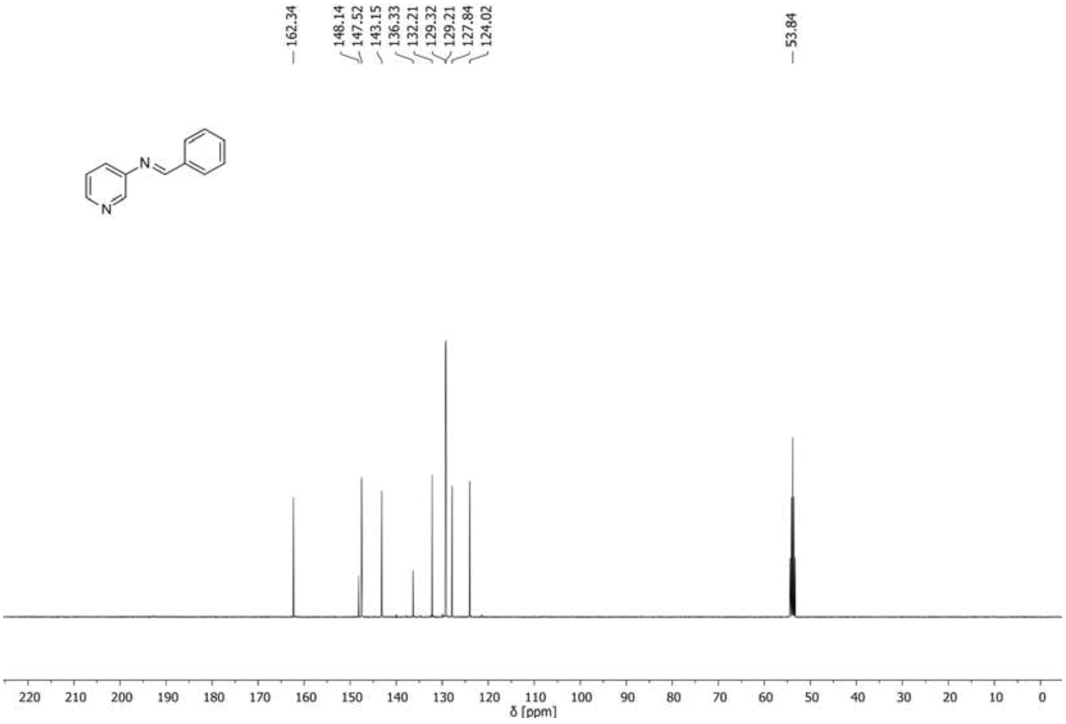
^{1}H NMR: (E)-1-(benzo[d][1,3]dioxol-5-yl)-N-(p-tolyl)methanimine (**1ab**)



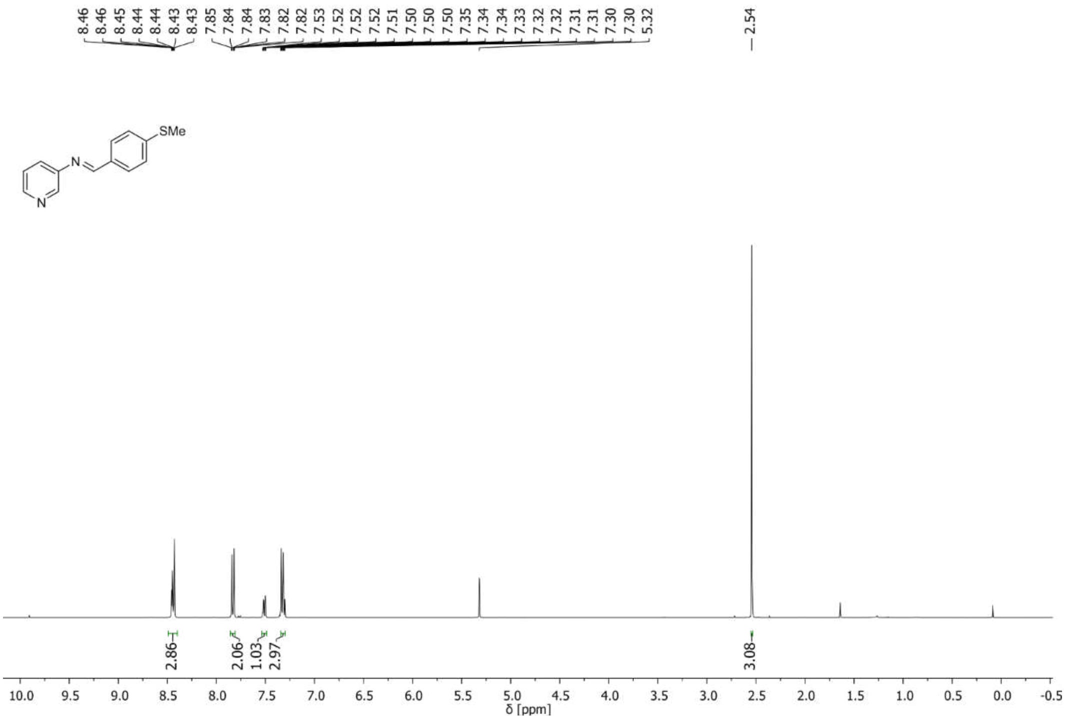
^{13}C{^{1}H} NMR: (E)-1-(benzo[d][1,3]dioxol-5-yl)-N-(p-tolyl)methanimine (**1ab**)



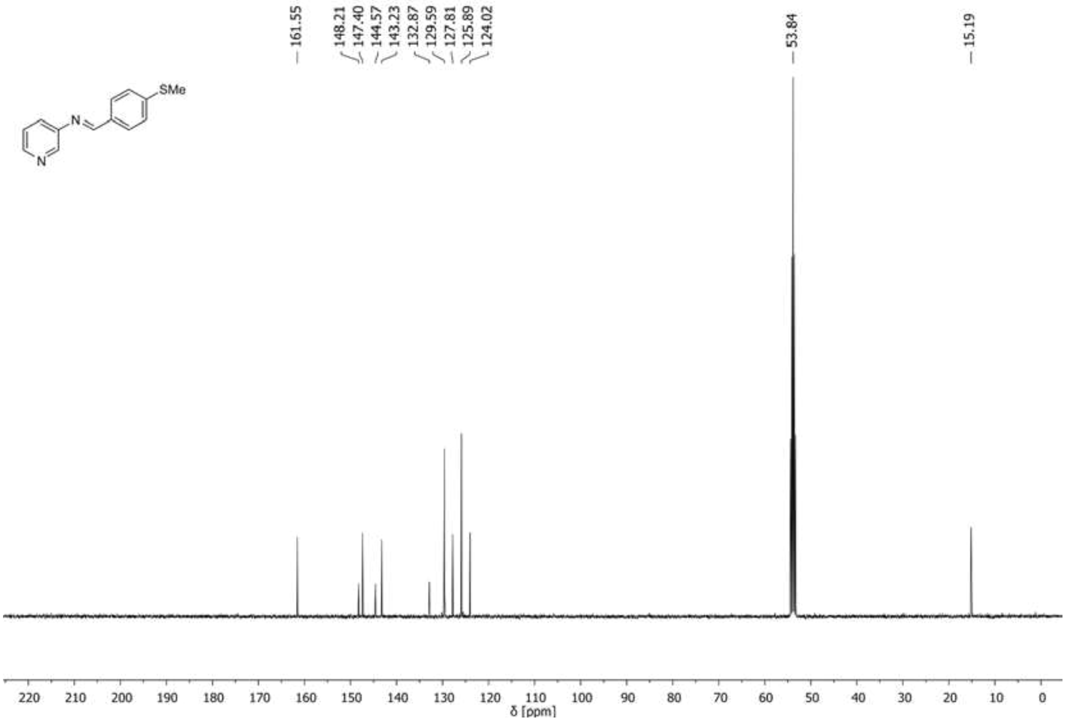
^{1}H NMR: (E)-N-benzylidene-3-pyridinamine (**1ac**)



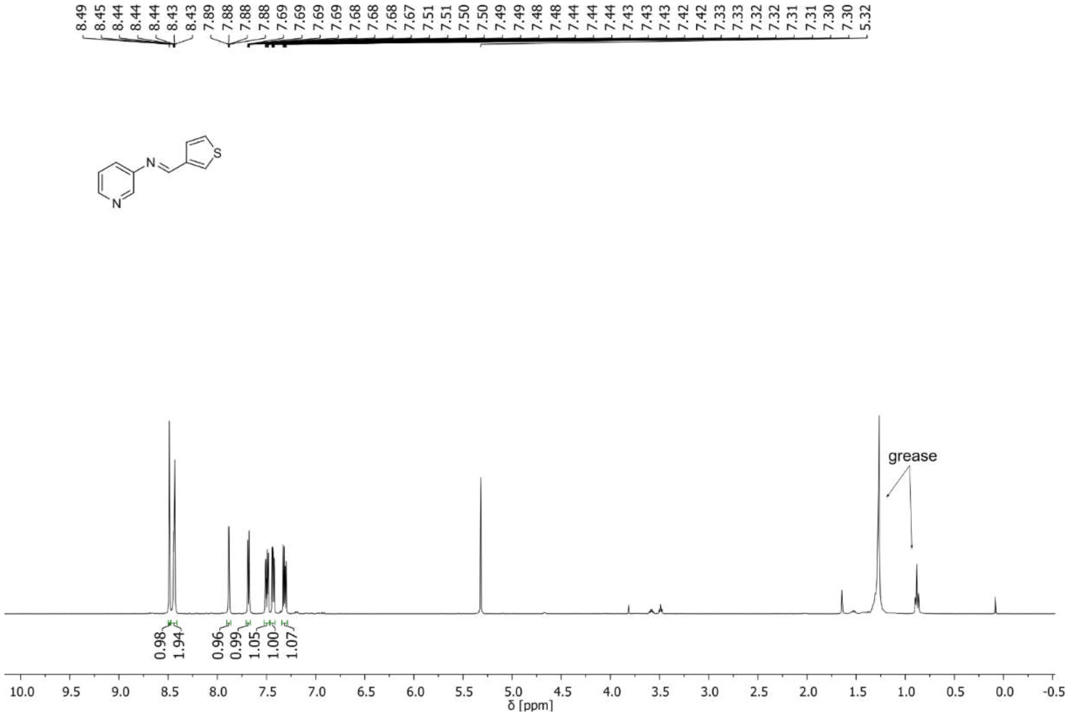
^{13}C{^{1}H} NMR: (E)-N-benzylidene-3-pyridinamine (**1ac**)



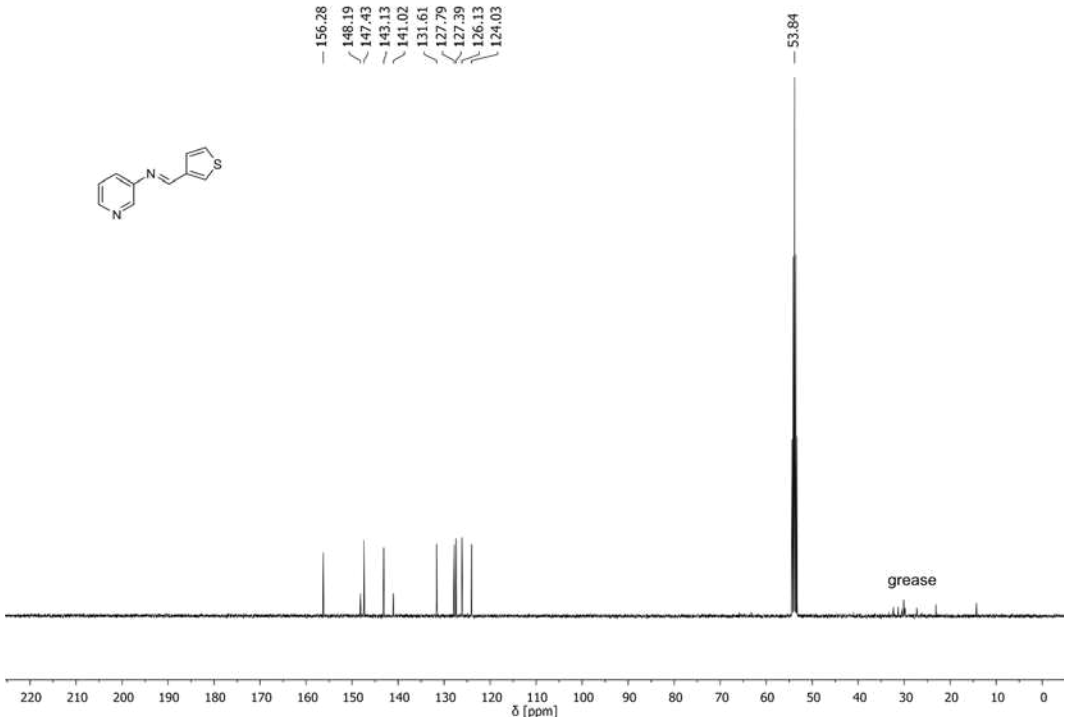
^{1}H NMR: (E)-1-(4-(methylthio)phenyl)-N-(pyridine-3-yl)methanimine (**1ad**)



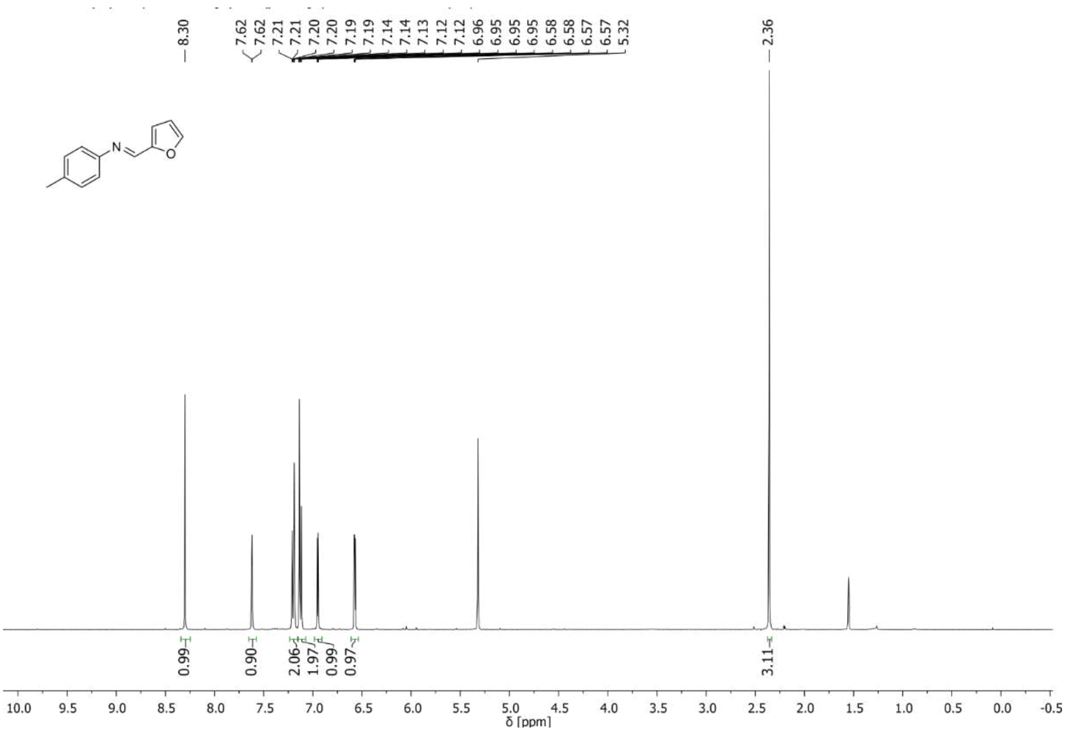
^{13}C{^{1}H} NMR: (E)-1-(4-(methylthio)phenyl)-N-(pyridine-3-yl)methanimine (**1ad**)



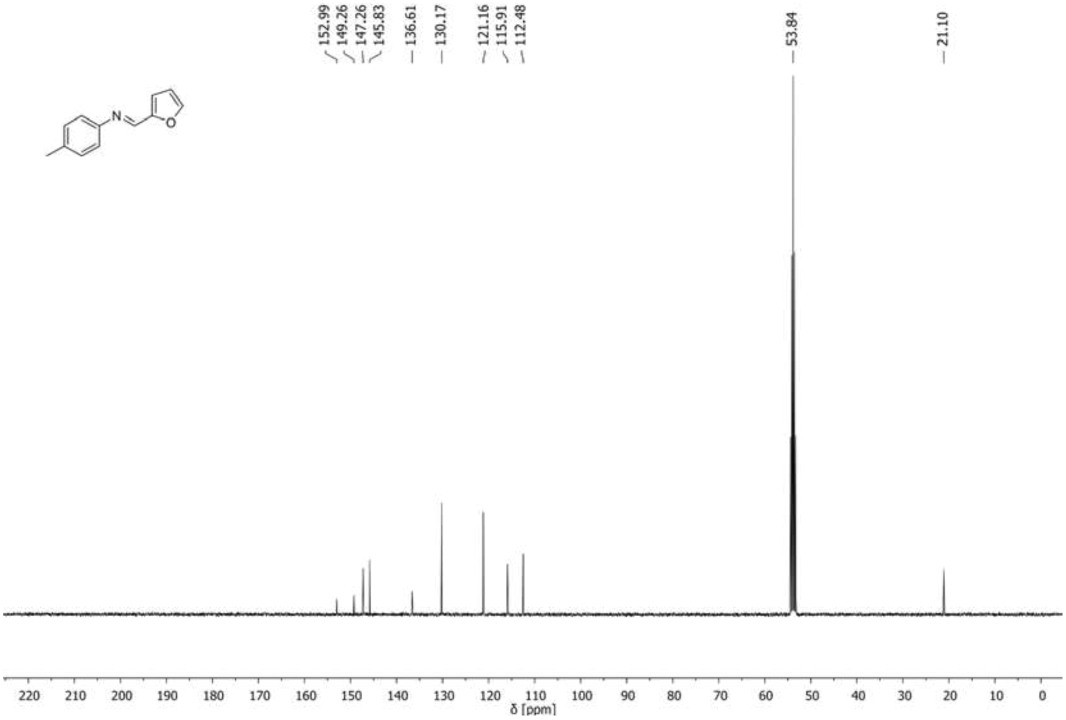
^{1}H NMR: (E)-N-(pyridin-3-yl)-1-(thiophen-3-yl)methanimine (**1ae**)



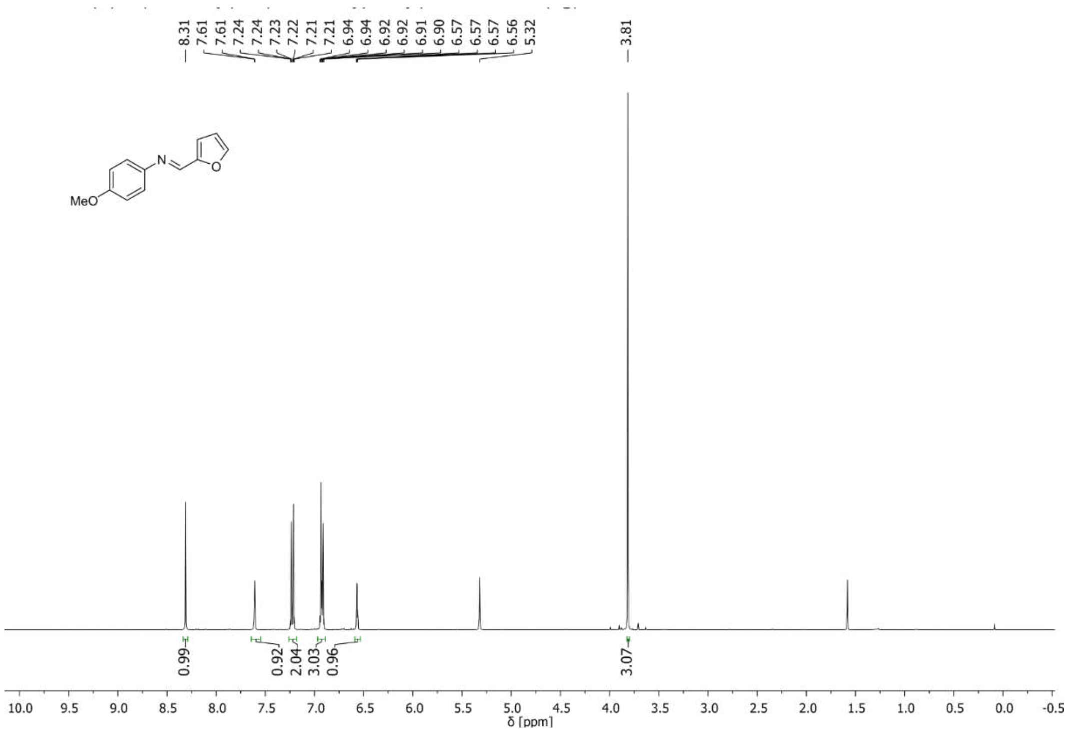
^{13}C{^{1}H} NMR: (E)-N-(pyridin-3-yl)-1-(thiophen-3-yl)methanimine (**1ae**)



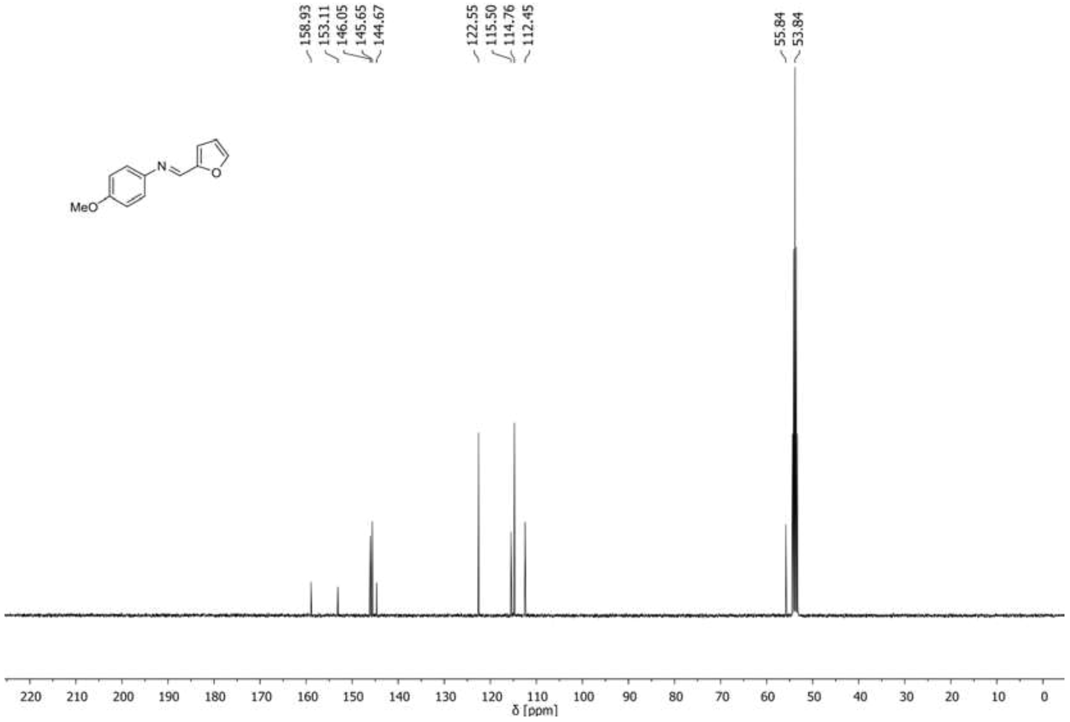
^{1}H NMR: (E)-1-(furan-2-yl)-N-(p-tolyl)methanimine (**1af**)



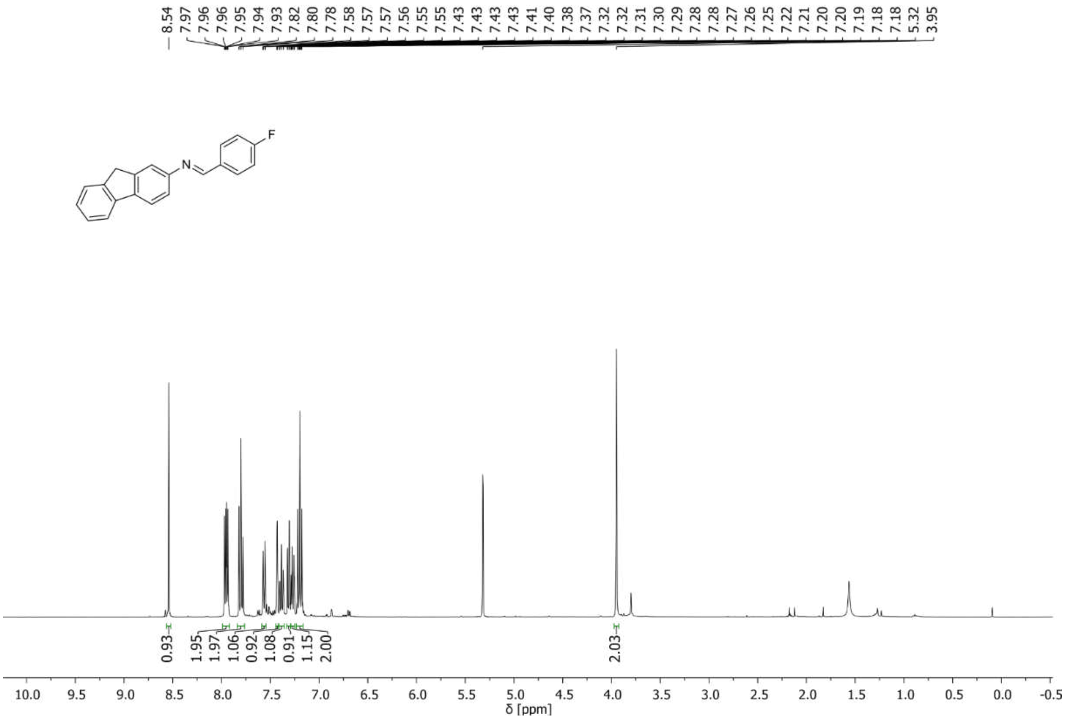
^{13}C{^{1}H} NMR: (E)-1-(furan-2-yl)-N-(p-tolyl)methanimine (**1af**)



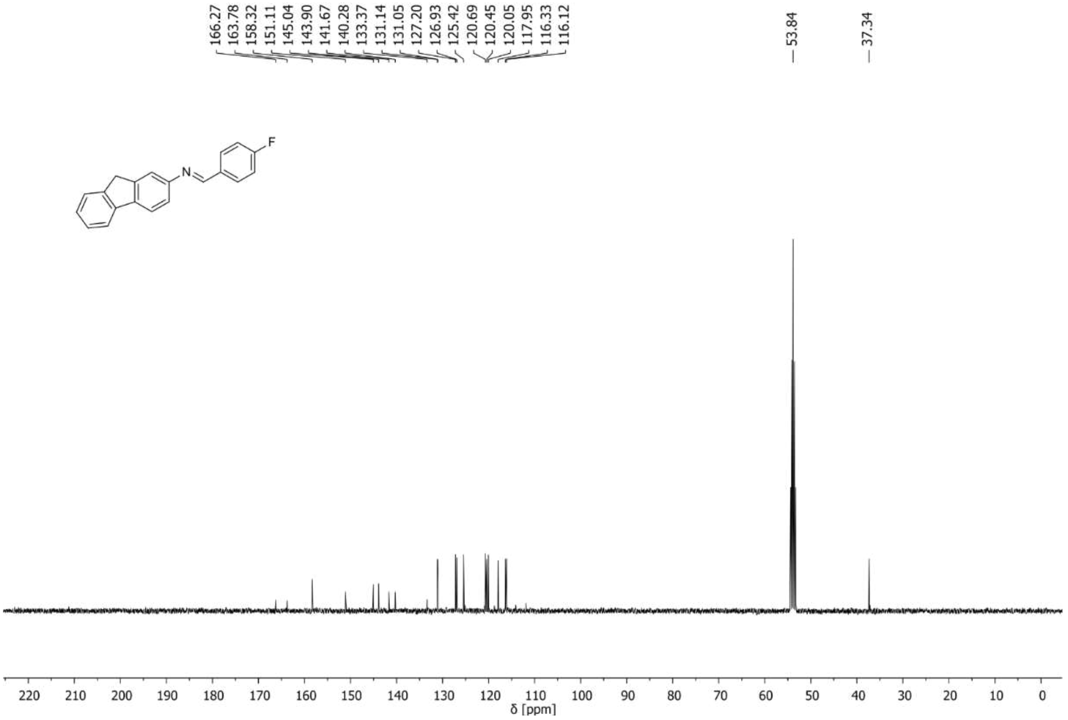
^{1}H NMR: (E)-1-(furan-2-yl)-N-(4-methoxyphenyl)methanimine (**1ag**)



^{13}C{^{1}H} NMR: (E)-1-(furan-2-yl)-N-(4-methoxyphenyl)methanimine (**1ag**)



^{1}H NMR: (E)-N-(4-fluorobenzylidene)-9H-fluoren-2-amine (**1ah**)



^{13}C{^{1}H} NMR: (E)-N-(4-fluorobenzylidene)-9H-fluoren-2-amine (**1ah**)

7. References

(1) H. Naka, D. Koseki, Y. Kondo, *Adv. Synth. Catal.* **2008**, *350*, 1901-1906.

(2) L. Tang, H. Sun, Y. Li, Z. Zha, Z. Wang, *Green Chem*. **2012**, *14*, 3423-3428.

(3) J. Wu, C. Darcel, *J. Org. Chem.* **2021**, *86*, 1023-1036.

(4) R. Fertig, T. Irrgang, F. Freitag, J. Zander, R. Kempe, *ACS Catal.* **2018,** *8*, 8525-8530.

(5) D. Riemer, W. Schilling, A.G. Götz, Y. Zhang, S. Gehrke, I. Tkach, O. Holloczki, *ACS Catal.* **2018**, *8*, 11679-11687.

(6) X. Hong, H. Wang, B. Liu, B. Xu, *Chem. Commun*. **2014**, *50*, 14129-14132.

(7) F.J. Goetz, *J. Heterocycl. Chem.* **1968**, *5*, 501-507.

(8) S.M. Landge, V. Atanassova, M. Thimmaiah, B. Torok, *Tetrahedron Lett.* **2007**, *48*, 5161-5164.

(9) A.L. Iglesias, J.J. Garcia, *J. Mol. Catal. A: Chem.* **2009**, *298*, 51-59.

(10) M. Mastalir, M. Glatz, N. Gorgas, B. Stöger, E. Pittenauer, G. Allmaier, L.F. Veiros, K. Kirchner, *Chem. Eur. J.* **2016**, *22*, 12316-12320.

(11) L.M. Kammer, M. Krumb, B. Spitzbarth, B. Lipp, J. Kuhlborn, J. Busold, O.M. Mulina, A.O. Terentev, T. Opatz, *Org. Lett.* **2020**, *22*, 3318-3322.

(12) B. Li, Y. Wang, Q. Chi, Z. Yuan, B. Liu, Z. Zhang, *New J. Chem*. **2021**, *45*, 4464-4471.

(13) N. Azizi, M. Edrisi, *Monatsh. Chem.* **2015**, *146***,** 1695-1698.

(14) G.J. Zhang, F.Y. Tang, X.Y. Wang, P. An, L.Q. Wang, Y.N. Liu, *ACS Sustainable Chem. Eng*. **2020**, *8*, 6118-6126.

(15) E. Manivannan, S.C. Chaturvedi, *Bioorg. Med. Chem.* **2012**, *20*, 7119-7127.

(16) F. Stanek, R. Pawlowski, P. Morawska, R. Bujok, M. Stodulski, *Org. Biomol. Chem.* **2020**, *18*, 2103-2112.

(17) L. Wang, B. Chen, L. Ren, H. Zhang, Y. Liu, S. Gao, *Chin. J. Catal.* **2015**, *36*, 19-23.

(18) A. Tejeria, Y. Perez-Pertejo, R.M. Reguera, R. Balaña-Fouce, C. Alonso, M. Fuertes, M. Gonzalez, G. Rubiales, F. Palacios. *Eur. J. Med. Chem.* **2016**, *124*, 740-749.

(19) W. Gong, M. Han, C. Chen, Y. Lin, G. Wang, H. Zhang, H. Zhao, *ChemCatChem* **2020**, *12*, 5948-5958.

(20) J.C. Anderson, G.P. Howell, R.M. Lawrence, C.S. Wilson, *J. Org. Chem.* **2005**, *70*, 5665-5670.

(21) D.J. Young, M.J.T. Robinson, *J. Labelled Cpd. Radiopham.* **2000**, *43*, 121-126.

(22) W. Nam, S.Y. Oh, M.H. Lim, M.H. Choi, S.Y. Han, G.J. Jhon, *Chem. Commun.* **2000**, 1787–1788.

(23) a) C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158-6170; b) F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305; c) E. Caldeweyher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2017**, *147*, 034112 d) V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995-2001; e) F. Neese. *WIREs ComputMol Sci.* **2022**,*12,* e1606; f) F. Neese, F. Wennemohs, U. Becker, C. Riplinger. *J. Chem. Phys.* **2020**, *152*, 224108; g). C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106; h) J. Zheng, X. Xu, D. G. Truhlar, *Theor. Chem. Acc.* **2011**, *128*, 295; i) F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065; j) S. Grimme. *J. Chem. Phys.* **2022**, *124*, 034108.

(24) J. N. Harvey, M. Aschi, H. Schwarz, W. Koch. *Theor. Chem. Acc*. **1998**, *99*, 95.

(25) S. Stoll, A. Schweiger, *J. Magn. Reson.* **2006**, *178*, 42-55.

From Data to Chemistry: Revealing Causality and Reaction Coordinates through Interpretable Machine Learning in Supramolecular Transition Metal Catalysis

R. A. Talmazan[1], J. Gamper[2], I. Castillo[3], T.S. Hofer*[2], M. Podewitz*[1]

[1] Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Wien (Austria), maren.podewitz@tuwien.ac.at

[2] Institute of General, Inorganic and Theoretical Chemistry, Leopold Franzens University of Innsbruck, Innrain 80/82, 6020, Innsbruck (Austria), t.hofer@uibk.ac.at

[3] Instituto de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Ciudad de México (México)

Abstract

Supramolecular transition metal catalysts with tailored reaction environments allow for the usage of abundant 3d metals as catalytic centres, leading to more sustainable chemical processes. However, such catalysts are large and flexible systems with intricate interactions, resulting in complex reaction coordinates. To capture their dynamic nature, we developed a broadly applicable, high-throughput workflow, leveraging quantum mechanics/molecular mechanics (QM/MM) molecular dynamics in explicit solvent, to investigate a Cu(I)-calix[8]arene catalysed C-N coupling reaction. The system complexity and high amount of data generated from sampling the reaction require automated analyses. To identify and quantify the reaction coordinate from noisy simulation trajectories, we applied interpretable machine learning techniques (Lasso, Random Forest, Logistic Regression) in a consensus model, alongside dimensionality reduction methods (PCA, LDA, tICA). Leveraging a Granger Causality model, we go beyond the traditional view of a reaction coordinate, by defining it as a sequence of molecular motions that led up to the reaction.

Introduction

Catalysis plays a pivotal role in chemistry. By lowering reaction barriers, catalysts facilitate chemical transformation under mild conditions, contributing to sustainable and economic processes. Nature has perfected this principle in enzymes, through precise control of the environment surrounding the catalytic centre and substrate.[1] In an effort to mimic the tight control over the environment, the field of supramolecular catalysis chemistry has emerged.[2–6] For example, the use of a macrocycle allows the substitution of rare earth metals with more abundant counterparts, while maintaining high catalytic performance.[7,8]

Complimentary to experimental advances, computational chemistry has played a key role in the design and understanding of catalytic systems, by elucidating reaction mechanisms.[9–12] Despite considerable efforts, quantum chemistry is limited in its predictive abilities.[13] As these systems grow in complexity, there is a need to improve not only on the underlying electronic structure theory used to study them, but also the chemical model that describes the catalytic system in its environment.[14] Often, the errors introduced by a too simplistic chemical model exceed those arising from the use of an approximate theoretical methodology such as DFT.[13,15] Consequently, the goal is to create a chemical model which is a "digital twin" of the reaction flask, where the in-silico procedure can fully replicate experimental conditions, meaning a catalyst in explicit solvent, at finite temperature and pressure. As a full ab-initio quantum chemical (QC) level is not feasible, a tailored multiscale strategy has to be developed, accounting for conformational flexibility, explicit solvation effects and the dynamic nature of chemical reactions.

While the majority of computational mechanistic studies are performed with a single structure, with conformer search recently gaining popularity[14,16,17] thanks to easily accessible tools,[18,19] a transition to structure ensembles provides a more complete picture.[14,17]

Another crucial aspect which needs to be considered is solvation. While widely used implicit solvation models provide remarkable performance,[20–22] they cannot describe explicit interactions between solute and solvent. In combination with dispersion corrections and a small basis set, they often favour very compact structures with many intramolecular bonds[7,19] – a poor representation of a solvated system. It is therefore crucial to include explicit solvation in any realistic model, ideally via a full condensed phase calculation or through microsolvation.[23,24]

Describing catalyst ensembles in explicit solvent is a significant improvement of the chemical model, yet it is important to recognize that catalysts are, by nature, dynamic entities. By tracking the motion of the nuclei during a reaction, we may observe a different reaction pathway,[25,26] The dynamic effects are well documented for organic molecules in implicit solvents,[27] yet rarely investigated in transition metal catalysis.[25,27,28]

While the inclusion of dynamics in mechanistic studies makes a model more realistic, it adds complexity and significantly increases computational cost. This limitation can be addressed by relying on multiscale methods such as quantum mechanics/molecular mechanics (QM/MM) which describe the catalytic centre and substrates at a QM level, while the surrounding environment is treated with MM.[29–31] This hybrid methodology can be combined with molecular dymanics (MD) to investigate the dynamic behaviour. A transition to QM/MM MD in explicit solvent allows for sampling timescales to the order of nanoseconds, magnitudes higher than in a pure QM approach. Due to the rare nature of the reactions, repeated sampling is a requirement for statistically relevant information regarding energy barriers and structural information.

While the setup of such multiscale methods is a challenge in itself, a vast amount of data is generated from the simulations. As it is not feasible to interpret them by observation only, an automatic way of processing reaction trajectories is needed. Machine learning (ML) approaches can be used to evaluate the data and extract condensed results which contain insight into the reaction coordinate.[28] While neural network approaches[32] have garnered a lot of attention recently, they are not very well suited for data analysis approaches where understanding of the process is required, due to their black box nature, which makes them very difficult to evaluate.[33] Instead, interpretable machine learning techniques, such as decision trees, random forests or logistic regression, offer good performance in extracting relevant information from large datasets and presenting them in easily understandable ways. In addition, dimensionality reduction techniques—Principal Component Analysis[34] (PCA) or time-lagged Independent Component Analysis[35] (tICA) effectively detect combined coordinates from the trajectories, revealing the key motions of a system. Yet these dimensionality reduction techniques have almost exclusively been applied to biomolecules[36–38] with few exceptions.[39,40] A combination of aforementioned methods offers great promise to detect a cumulative reaction coordinate from a multitude of independent trajectories, providing chemical insight into the mechanism and reactivity of the system. To the best of our knowledge these combined methods have not been applied to study reaction mechanisms in explicit solvent, let alone large supramolecular transition metal catalysts.

Another aspect that has until now been neglected in chemistry is causality. While the concept is widespread across various scientific domains[41]—ranging from economics[42–46] and climate research[47–51] to biology[52,53] and medical studies[54–56]—it remains surprisingly absent in the field of chemistry. Although a handful of precedents in biomolecular simulations exists,[57–59] it has not been explored to study chemical reactions, not to mention transition-metal catalysis. As trajectories are essentially discrete time series, containing the various degrees of freedom of the system, causality can be statistically inferred from the analysis of these trajectories. Consequently, the reaction coordinate can be decomposed into a sequence of motions leading up to the reaction, exposing the intricate interplay of functional groups of the system, offering an unprecedented view of reactivity.

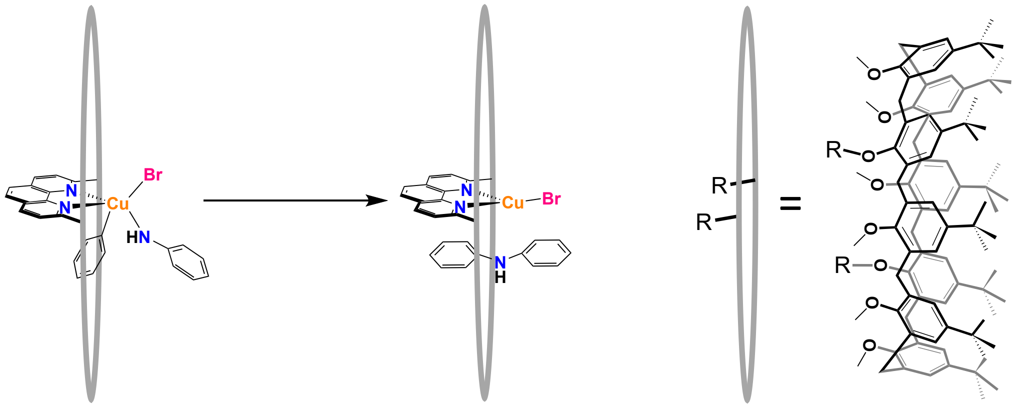


Figure 1. The supramolecular system calix[8]arene-based [Cu(C_8PhenMe$_6$)I] catalyses the C-N coupling reaction of phenyl bromide and aniline.

A supramolecular catalyst that has shown remarkable catalytic efficacy for C-N coupling is the the Cu(I)-1,5-(2,9-dimethyl-1,10-phenanthroyl)-2,3,4,6,7,8-hexamethyl-*p*-tert-butylcalix[8]arene system.[7] The macrocyclic ligand allows for usage of earth abundant metals, such as Cu, an essential step towards more sustainable chemical processes.[60–65] The investigated system necessitates explicit solvation for accurate results, as implicit solvation models lead to a collapse of the macrocyclic cage, which compromises catalytic activity.[7] To account for the conformational flexibility of the supramolecular cage[7,66] and the dynamic nature of the system as a whole, a dynamic ensemble-based approach is required to study the reaction. While the mechanism of this catalysts was established to be a sequence of oxidative addition / reductive elimination,[7] the dynamic effects of the system, in particular the contribution of the cage, and the influence of explicit solvent molecules are completely unknown. Consequently, we developed a multiscale QM/MM MD approach to understand the bond formation dynamics of the C-N coupling step with the Cu-calix[8]arene catalyst in explicit chloroform. By relying on the GFN2-xTB[67] method to describe the QM part, we could achieve massive sampling, resulting in 152 individual reaction trajectories. To extract chemically relevant information from these data, we employed supervised and unsupervised interpretable machine learning dimensionality reduction models, in order to identify the cumulative reaction coordinate and to detect critical movements in the structure. A consensus approach combining individual machine learning techniques improved the performance, while random forest models and decision rules allowed us to quantify the reaction coordinate. Finally, we employed the statistical Granger Causality analysis model[68,69] to decompose the reaction coordinate into a sequence of individual consecutive movements. This work serves as a widely applicable template for any mechanistic investigations, revealing and quantifying complex reaction coordinates, alongside causal effects derived from the individual movements leading up to the reaction.

Results

We obtained 152 QM/MM MD reaction trajectories for the C-N coupling step. Out of these, 142 reacted within 20 ps of simulation time. From these trajectories we evaluated the reaction energy and labelled the structural data accordingly as educt, transition state, or product – resulting in three ensembles. We then used this information to identify the reaction coordinate, quantify it and identify a sequence of movements leading to the reaction.

Reaction Energetics Analysis

We analysed 142 simulations where a reaction occurred to extract insights about the C-N coupling process. As the reaction happened spontaneously during the simulations, the energy profile can be obtained directly (SI Figure S2 A) and served as indicator to label the three distinct states, educt, transition state, and product. The reaction energy was obtained by averaging the ensembles of the educt and product states and amounted to -212 ± 25 kJ mol$^{-1.}$ A sigmoid fit through the smoothened energy profile of each simulation (see SI, Figure S2 B and C) allowed identification and the calculation of the energy barrier to be 13 ± 9 kJ mol^{-1}. These GFN2-xTB reaction energies are in excellent agreement with full DFT data, obtained with PBE0/def2-SVP/D3 (see SI, Table S2).

Extracting Chemical Information from Structural Data

To obtain information about the changes in chemical structure from the reaction trajectories with a total of over 1.5 million frames, we resorted to interpretable machine learning approaches.

Determination of a suitable coordinate system

A standard method to extract reaction coordinates from trajectories, either in biomolecular or reaction dynamics studies, is Principal Component Analysis (PCA) in cartesian coordinate space. However, this approach proved unsuccessful for the Cu-calix[8]arene catalyst due to the difficult to properly align this highly flexible system not showing any separation between the three states (Figure S3).

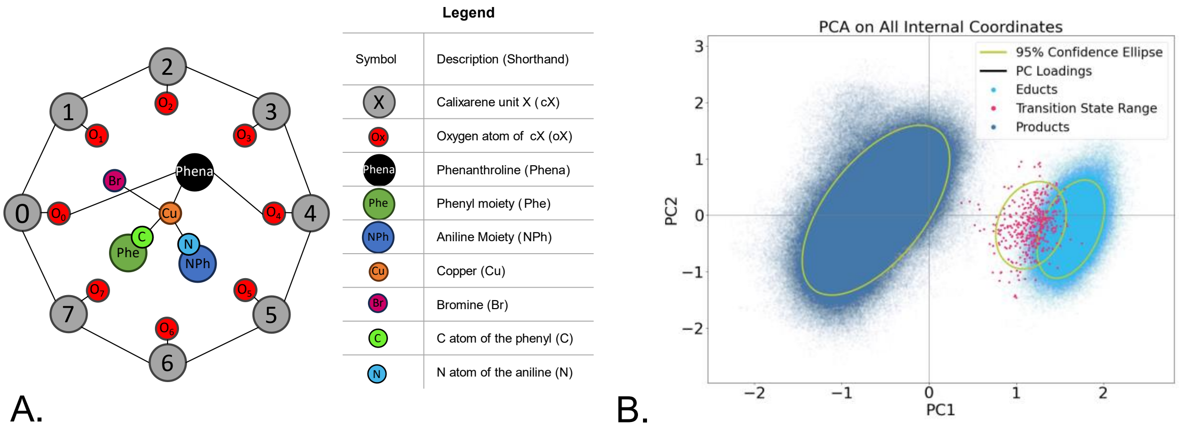


Figure 2. A. Schematic depiction of the calixarene, showing the centres of mass used for the calculation of the reduced set of internal coordinates; B. PCA performed on the reduced internal coordinates.

To achieve good separation between the three states, educt, TS, and product, we developed a reduced internal coordinate description of the system (see Figure 2A) to minimize the noise from highly correlated coordinates.[82] Hence, we described rigid fragments, such as the individual calixarene units (cX), the phenanthroline bridge (Phena), the phenyl (Phe) and aniline (NPh) moieties by their respective centres of mass (Figure 2A). An overview over the distribution of the internal coordinates which define the reduced model can be found in the supporting Information, Figure S10. This internal coordinate set nicely separates educts and products in the PCA space (Figure 2B), but still shows overlap between educts and transition states. Analysing the loadings of the principal components (see SI Figure S5), we can see that the main contributions belong to the C-N, NPh-Phe, Cu-N distances, as well as to the Phena-Cu-Br angle, describing changes in the coordination at the Cu centre as the product is formed.

A second less common method, to reveal the reaction coordinates in biomolecular studies, is the time-lagged Independent Component Analysis (tICA). While PCA focuses on the largest variance in the dataset, which generally corresponds to fast molecular motion, tICA can be used to separate and extract the internal coordinates which exhibit the strongest time-correlations for a chosen lag time, thus revealing slow movements in the system.

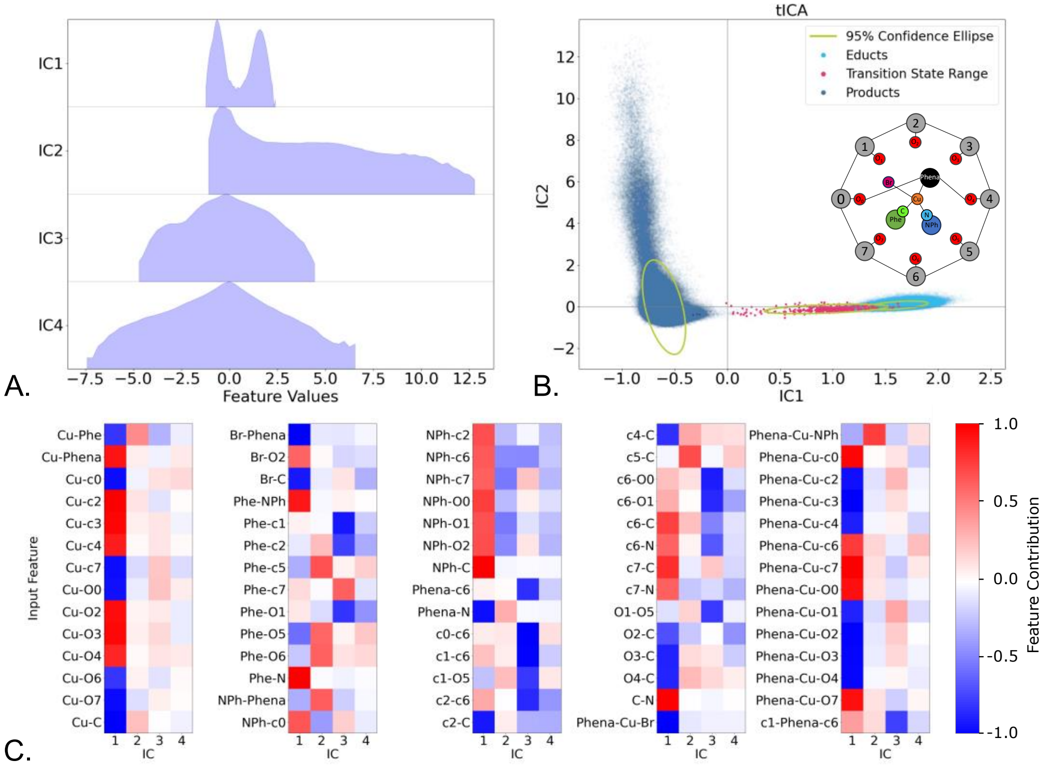


Figure 3. Time-Lagged Independent Component Analysis of the MD trajectories, using a lag time of 20 fs. A. Distribution of the structural ensembles (Feature Values) over each of the four independent components; B. Projection of the ensembles in the space of independent components 1 and 2, with the ensemble colouring performed a posteriori; C. Normalized internal coordinate (feature) contribution to each independent component.

The tICA performed on the reduced internal coordinates set, resulted in a good separation of the product and educt states, mainly across the first independent component (IC1), as shown in Figure 3A-B, yet the transition state ensemble cannot be fully separated from the educts. When taking into consideration further ICs (Figure 3A and 3B), we observe a broad distribution of the product ensemble, indicating significant conformational flexibility. The contributions to ICs can be traced down, by relating the contribution strength (Figure 3C) to the degree of freedom it corresponds to (Figure 3B insert). A positive contribution (coloured in red) means that the respective feature increases as the values of the IC increases, while a negative contribution (coloured in blue) means the feature increases as the IC values decrease. The absolute value of a contribution (colour intensity) represents the importance of the feature in defining the IC.

IC1 reveals the changes at the reaction centre (see 1st column, Figure 3C), related the Cu adopting a planar configuration upon product formation. Of particular interest are the strong contributions of the distance of NPh to the c0, c2, c6 and c7 (2nd column bottom/3rd column top) and their similarity to the C-N distance (4th column bottom), indicating the movement of the product deeper into the cavity. Additionally, a translation of the calixarene units can be inferred, when looking at the changes in the Phena-Cu-cX and Phena-Cu-oX angles (5th column). Their similar contribution values indicate a translation, rather than a rotation. This indicates that the distance between the phenyl part of the product and the calixarene units decrease as the reaction proceeds, which is indicative of π-π stacking between the product and calixarene cage. IC2 acts to separate various conformers within the product ensemble.

Improved Reaction Coordinate Detection Through Supervised Methods

We intended to further improve the separation of the three states in the PCA by utilizing the labelling of the data, indicating each structure as educt, product or transition state. Using this information, we trained a model that maximized the separation between the three ensembles and simultaneously reduced the number of internal coordinates (features) to those that contribute the most to the separation. The latter is referred to as feature elimination. There are various ways of achieving this goal and we tested a selection of them in PCA and Linear Discriminant Analysis (LDA) dimensionality reduction approaches (see Figure 4). While PCAs depend on the technique used to eliminate the features (Figure 4A-C and SI Figure S4), LDA is independent of the feature selection and consistently yields excellent separation (Figure 4D-F and SI Figure S4). While LDA results in compact distinct ensembles, PCA also achieves separation within the ensembles depending on the feature elimination approach that was used.

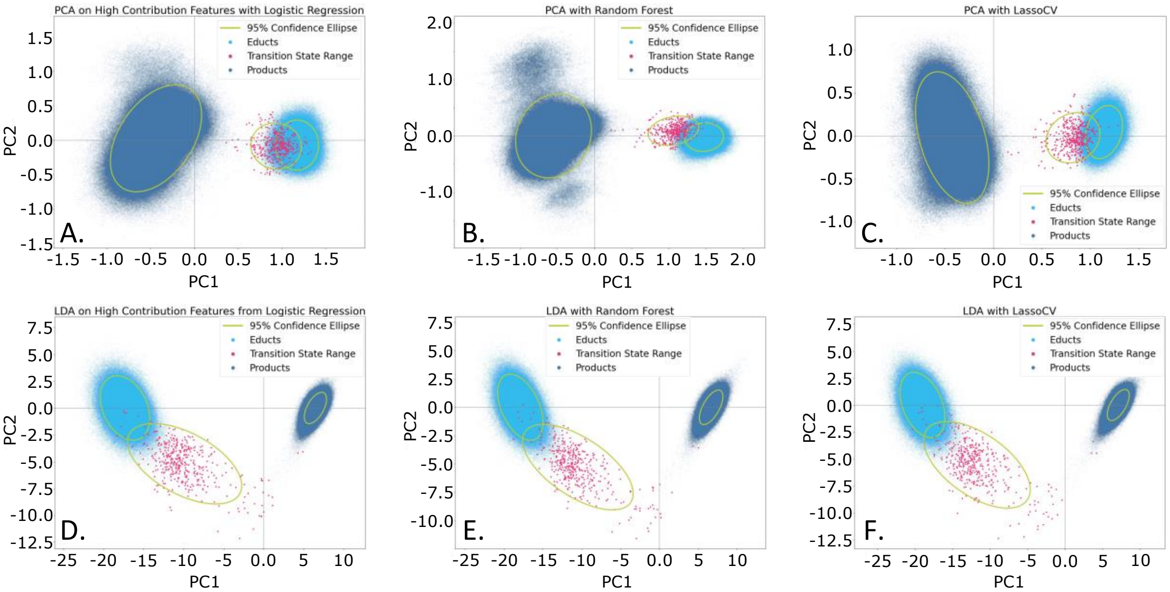


Figure 4. Dimensionality reduction performance with various feature reduction methods: A. PCA with Logistic Regression with mean-based cutoff; B. PCA with Random Forest with mean-based cutoff; C. PCA with LassoCV; D. LDA with Recursive feature elimination with cross validation using a logistic regression classifier; E. LDA with Recursive feature elimination with cross validation using a random forest classifier; F. LDA with LassoCV.

Amongst the top contributors are the change in the Phena-Cu-Br angle, as well as the C-N distance. Secondary features such as the distance between the product and phenanthroline bridge also play an important role for the LR classifier. The RF identifies the distances between the Cu and several calixarene units as being important, as well as the distance between the aniline and phenanthroline. The separation with the LassoCV approach is superior to that obtained with the logistic regression, as the educt and transition state ensembles are better separated. However, the there is only one ensemble visible for the product state. The principal contributors are the Cu-N, Cu-Phe, Cu-C, Phe-O3 and Phe-O2 distances.

It can be observed that the various feature selection methods show variability and differences in the selected internal coordinates and separation performance.

Revealing Causality in the Reaction Coordinate – Information from Consensus Model

In order to combat the variability associated with the various feature selection methods, we switched to a consensus model, which integrates the features highlighted by all supervised ML methods. The model identifies 49 features (internal coordinates) (Figure 5A) of high importance. We performed an additional PCA (Figure 5B) on the consensus features. Additional insight can be obtained by performing hierarchical clustering on the consensus features (see SI) and calculating the correlations between them, resulting in a clustermap (Figure 5C), which reorders the internal coordinates according to their correlation to each other.

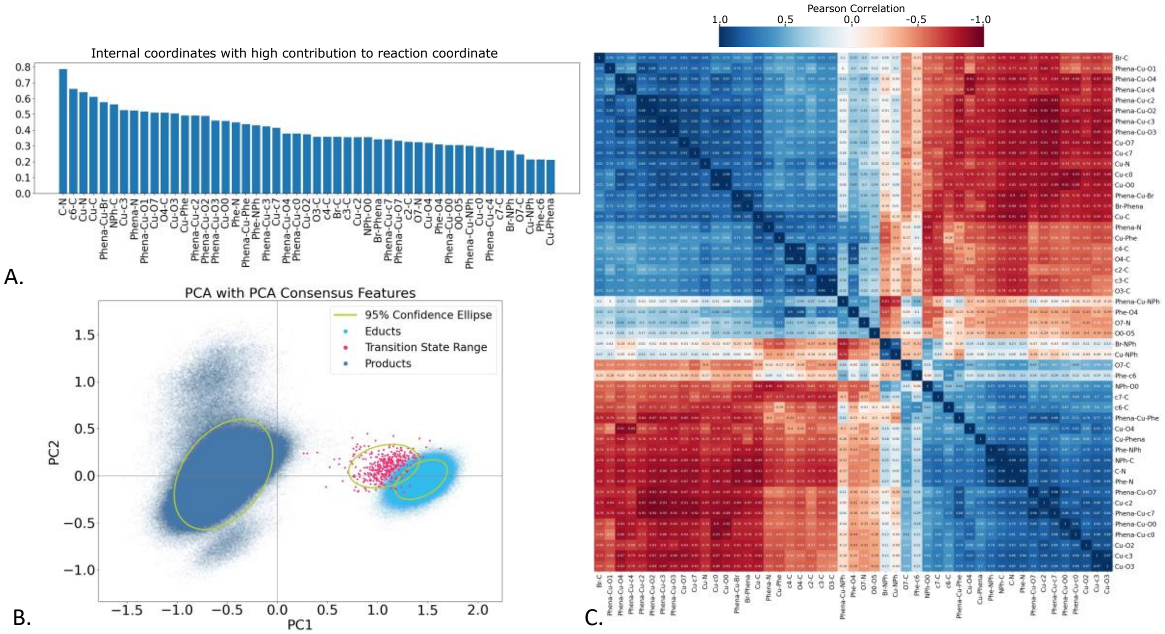


Figure 5. Analysis of internal coordinates deemed as high importance by the consensus model. A. Internal coordinates and their contribution; B. PCA of the consensus features; C. Clustermap of the internal coordinates Pearson correlations.

Judging the importance of the consensus model features (Figure 5A), we see that the C-N bond distance is most important (in agreement with chemical intuition), alongside several distances and angles corresponding to the reaction centre. Notably, the c6-C distance is also deemed highly important, which indicates that the cage indeed plays a role in defining the reaction coordinate. This small set of internal coordinates yields almost perfect separation of the three states as well as distinguishing the product conformations (Figure 5B), thereby outperforming any of the individual feature selection methods.

When looking at the clustermap plot (Figure 5C), we see two distinct regions of the complex with opposite correlation. Upon closer inspection of the individual features and their clustering, we can see the internal coordinates corresponding to the each calixarene unit group together. Coordinates that show positive correlation with the C-N bond (depicted in blue) decrease in value as the reaction proceeds, while those showing a negative correlation (depicted in red) increase during the reaction. Hence, the distances between Phe and NPh (comprising the product) and calixarene units 6 and 7 decrease, indicating π-π stacking interactions (depicted in blue), whereas C moves away from calixarene units 2-4 (coloured in red).

While the correlation analysis shows which movements take place in a correlated fashion, it is also interesting to evaluate the causality of these movements and how they propagate through the system. As tICA, a time-lag based method, revealed new information regarding the cage movement and product interaction, we decided to apply the Granger causality model, also time-lag based, to evaluate the reaction trajectories.

Using the coordinates resulting from the consensus model, we can construct a propagation cluster map of our system (Figure 6), which allows us to observe how the internal coordinates influence the C-N bond formation. This combines the hierarchical clustering of the consensus features with the results of the Granger causality model. Thus, we can infer the causality between coordinates and groups of coordinates.

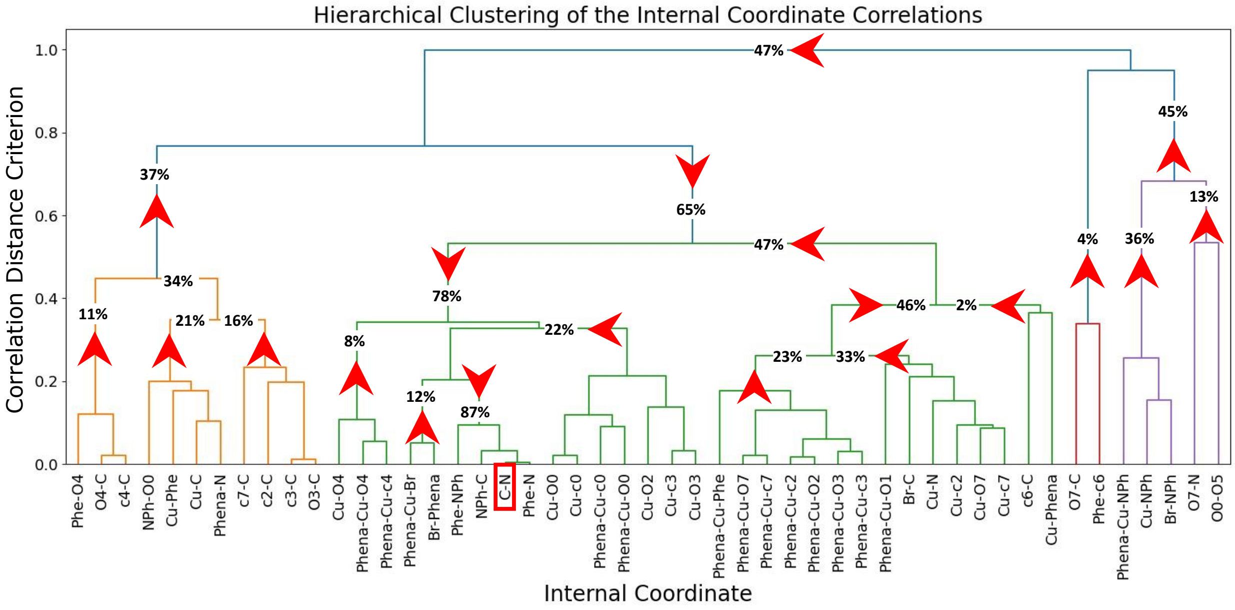


Figure 6. Propagation cluster map of the system, leading to the C-N bond formation. The numbers represent the total causality effects uncovered by the respective coordinate(s) and the red arrows indicate the direction of causality.

It allows us to examine the effects leading to the change in the C-N distance, identified as the most important feature in the consensus model. Causality was evaluated between each of the 49 consensus features and then examined with regards to the C-N bond distance. Thus, Figure 6 can be read as a map of movements leading to the coupling reaction, by choosing a starting point and following the arrows towards the highlighted C-N feature. The numbers next to the arrows indicate the number of trajectories where the C-N bond formation can be attributed to the respective features. In 52 trajectories out of 139, the coupling is triggered by a move in the orange cluster, consisting of the calixarene 2,3,4 and 7 to C distances as well as the proximity of Phe to the Cu. When we include information about the Phena-Cu-Br angle, alongside Cu and Br distances to NPh, from the purple cluster, as well as information from the red cluster, we can infer causality in 90 of the trajectories. The green cluster consists mostly of information related to the position of the calixarenes around the reaction centre. When we combine all information together, we can infer that the C-N coupling is caused by movements in the calixarene cage, alongside the Cu coordination change in 121 trajectories. Naturally the causality analysis can be expanded to include the sources of changes in each feature, however this quickly becomes a very complex multidimensional problem.

Quantification of Reaction Coordinates

While the previous approaches identified the relevant internal degrees of freedom which determine the reaction coordinate, as a next step, we sought to quantify it, that is, to define ranges which separate the data into the three ensembles. A method to achieve this, is the decision tree, which splits ensembles by setting cut-offs based on the coordinates that show the largest distribution differences between classes.

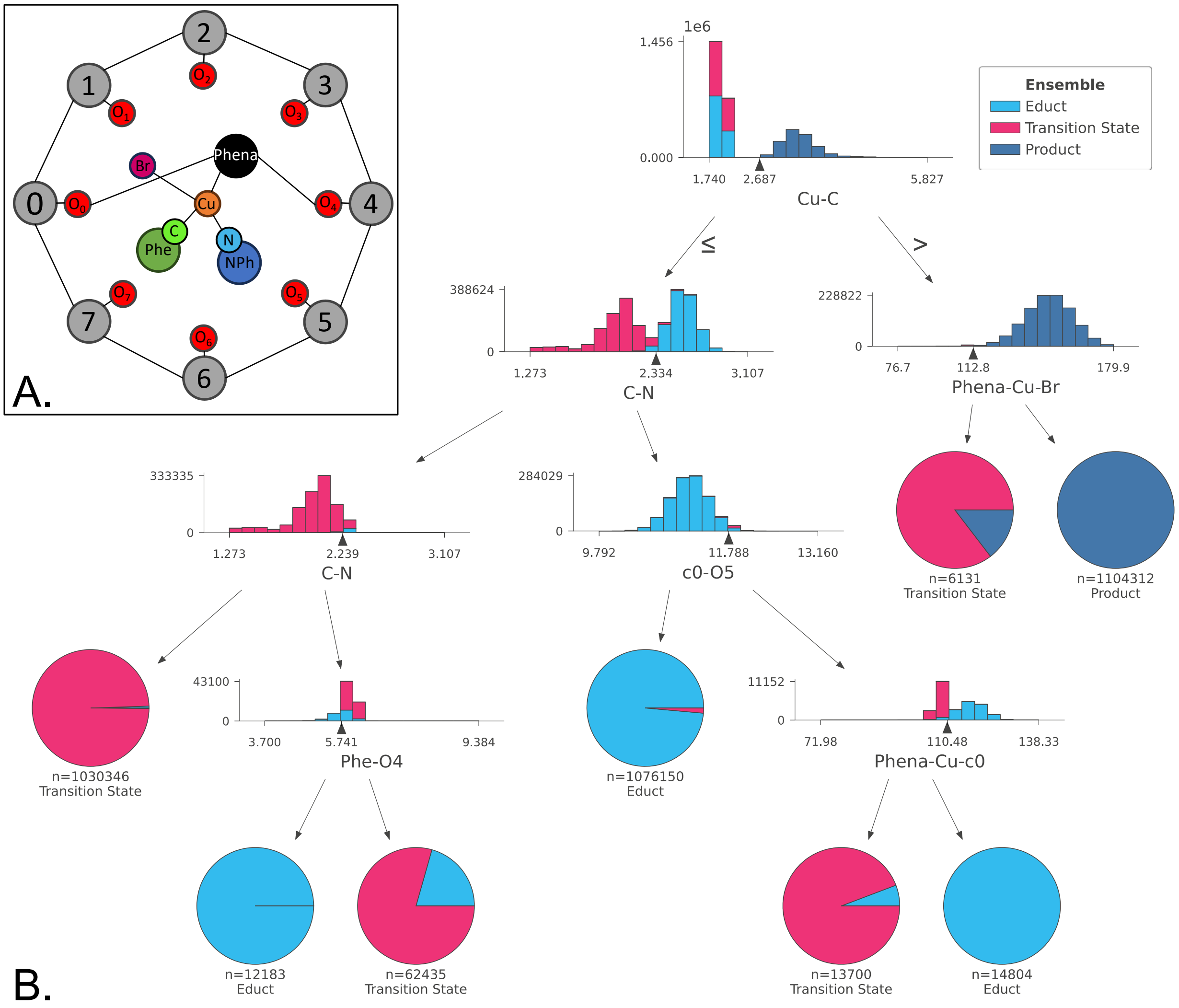


Figure 7. A. Model of the calixarene with the reduced centres of mass; B. Decision Tree classifier used to interpret the differences between the 3 structure groups,trained on the balanced classes.

The decision tree in Figure 7 has been trained on the whole dataset, with balanced weighting given to the groups via oversampling, in order to remove biasing against the transition state ensemble, which contains significantly fewer structures. The results of an unbalanced tree can be found in Figure S7.

Analysing the tree, we can see that the Cu-C distance plays a key role in the splitting of the educts and products, with the majority of transition states being grouped with the educt class. The remaining transition state contamination of the product ensemble can be separated by taking the angle determined by the Phenanthroline bridge, Cu and Br atoms into account (Phena-Cu-Br), where values below 112.8° are indicative of a transition state.

To differentiate between the transition states combined with educts, the C-N bond represents an effective metric, where values higher than 2.33 Å indicate an educt, while distances below 2.24 Å indicate a transition state. The educt states that do exhibit a C-N bond distance similar to that of the TS can be identified by a smaller phenyl to calixarene unit 4 oxygen atom distance (Phe-O4). When transition states exhibit a C-N bond distance over 2.33 Å, we can perform a selection based on the calixarene c0 and calixarene O5 distance (which should be greater than 11.79 Å) and the angle defined by the Phenanthroline bridge, Cu and c0, where values below 110.5 indicate a TS.

The reliability of a single decision tree-based approach can often be improved by utilizing a random forest classifier and averaging the results. In general, this improves accuracy, but reduces the interpretability. To overcome this limitation decision rules can be used, providing a semantic understanding of the RF classifier. We used 30 decision trees, each trained on a subset of the data, to yield the random forest. When applied to our dataset, this method provides rules for each of the three classes, as seen in Figure 8, below. A complete diagram of the decision rules can be found in the SI, Figure S8.

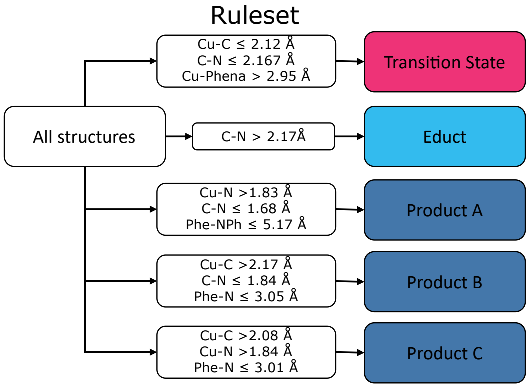


Figure 8. Graphical depiction of the decision rules derived from the random forest approach, Product A, B, C refers to different conformer ensembles within the product category.

Notably, the decision rules approach identifies three distinct rule sets for defining a product. This correlates with the findings from the consensus PCA and indicates the existence of various conformers in the product ensemble.

Discussion

Taking advantage of semi-empirical quantum chemistry methods allowed us to obtain a large sample set of the reductive elimination step of the C-N coupling reaction with a Cu-calix[8]arene catalyst, using a hybrid QM/MM molecular dynamics approach. Since the amount of generated data, alongside the very high dimensionality, makes the reaction difficult to interpret by visual analysis, statistical methods and machine learning techniques were used to extract chemically relevant information from the dataset.

To investigate the C-N coupling step we generated 152 reaction trajectories, amounting to more than 3 ns of total sampling time. We observed a convergence of the PCA with increased sampling (see SI, Figure S9), we assumed the total simulation time to be sufficient. As we observed spontaneous C-N coupling in 142 of the 152 trajectories, we could directly analyse these unbiased simulations.

By analysing the reaction energy profiles, we were able to quantify the reaction energy, as well as the reaction barrier, including uncertainty values. While the reaction barrier computed with a static modelling approach is within 10 kJ mol^{-1} (13 ± 9 kJ mol^{-1} vs 23 kJ mol^{-1}), the reaction energy exhibits some differences, -212 ± 25 kJ mol^{-1} vs -255 kJ mol^{-1}.[7] The remaining differences in the reaction energy can be attributed to the different chemical models that were used: in the static model, the investigated structure was obtained at 0 K, representing the bottom of the potential energy surface, while the dynamic study not only averages over all conformations but also considers thermal energy, hence, the structures are *not* 0 K structures and hence no minima on the potential energy surface. Alongside energetics, the reaction profile allowed for the categorization of the structures into three ensembles, namely educt, transition state and product. This was a key step in improving the reaction coordinate detection, as it allowed for the use of supervised learning methods to reduce the coordinate space.

To yield any separation between the three states in the PCA, we had to transform cartesian coordinates to a reduced set of internal coordinates, to minimize the number of highly correlated coordinates in the data set, thus reducing noise. While product states could be separated from educts and transition states, the latter two still showed overlap. In contrast, standard PCA on the cartesian coordinates resulted in no separation of the three states. We suspect the poor performance stems from failure to fully eliminate rotational and translational degrees of freedom from the system. Complementary to PCA tICA was used to identify slow movements of the system. While the analysis revealed some structural insights into the product conformer ensemble – revealing π-π stacking interactions between the coupling product, phenanthroline, and calixarene – it was not able to separate educts from transition states. Hence, tICA could not fully identify the reaction coordinate.

Utilizing the labelled data in PCA and LDA combined with supervised ML approaches resulted in a much better separation of the three ensembles. While the performance of PCA was highly dependent on the internal coordinate set, LDA showed remarkable separation between the three ensembles, highlighting the robustness of the method. A consensus model developed to combine the performance of the various dimensionality and feature reduction methods, identified 49 internal coordinates to be relevant, with the C-N distance being the most prominent one, which is in agreement with chemical intuition.

By applying the Granger causality model, pioneered in econometrics and widely applied in other scientific fields, [41–56] we were able to decompose the reaction coordinate into a sequence of molecular motions. The four main components of the reaction coordinate, separated in time and happening as a domino effect, are depicted schematically in Figure 9: i) the calixarene cage tilts perpendicularly to the phenanthroline; ii) the Cu changes coordination to become planar; iii) the C-N bond distance shortens; iv) π-π stacking effects drive the movement of the product inside the cage and below the phenanthroline. It appears that in the majority of the sampled data, the C-N coupling distance is largely influenced by a movement in the upper cage of the calixarene, namely units 2,3 and 4, which can be seen as the trigger for the reaction. This works presents a causal analysis, based on the robust Granger model for the C-N bond formation as a proof of principle. However, the model relies on the assumptions of stationary time-series and provides limited insight into the scale of the relationship between cause and effect. A transition to more advanced causal discovery models, such as PC Momentary Conditional Independence (PCMCI)[112] based methods would allow for analysis of non-stationary time series and better false positive control but is beyond the scope of this work.

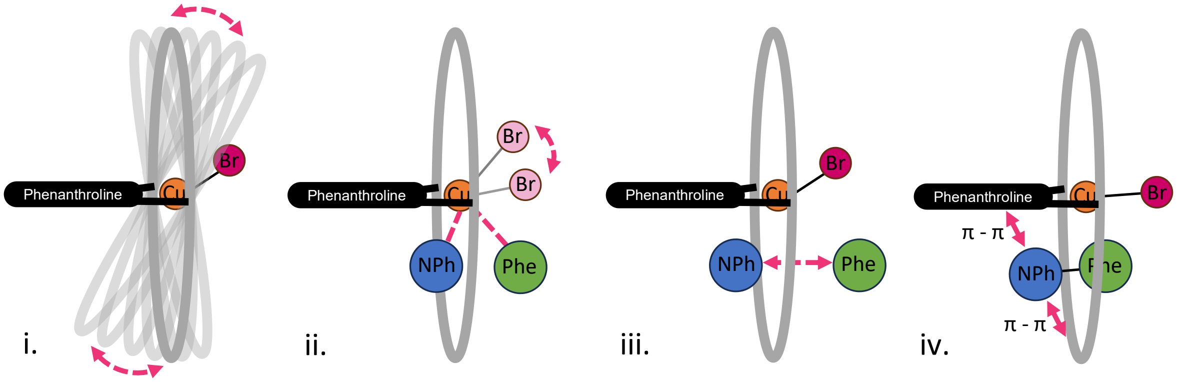


Figure 9. Schematic representation of the movements corresponding to the reaction coordinate of the C-N coupling reaction. i) Tilting movement of the calixarene cage; ii) Change in coordination of the copper centre; iii) Shortening of the C-N bond; iv) π-π stacking effects stabilise the product in the cavity.

To quantify the changes in internal coordinates, we employed a decision tree trained on the three ensembles, which highlighted the Cu-C and C-N distances as being the main feature for separating products from educts and transition states, respectively. However, decision trees are dependent on the initialisation condition and must be limited in their size to maintain interpretability. A transition to the decision rule approach overcomes some of the limitations by assigning a semantic interpretation to a random forest classifier, thus, clearly separating the three ensembles. In addition, three distinct product conformers are identified, each with a different set of rules defining them. These conformations will likely converge as the product diffuses out of the cavity.

To the best of our knowledge there is only a single study of the reaction dynamics of a transition-metal complex with explicit solvent with repeated sampling of the reaction step.[28] This study on Fe-oxo-mediated C–H functionalization reactions by Joy et. al used kinetic energies, quantum numbers and velocities to distinguish between two different dynamic reaction pathways.[28] While they also used ML for feature selection, their focus is on physical chemical factors that impact reactivity. Our focus lies on the investigation of structural changes to ease interpretation and to facilitate causality analysis, which has never been attempted for chemical reactions but opens a completely new angle on how to understand reactivity.

From all statistical analyses, we can see that the transition state structures tend to be close to the educt or even overlays with the educt space (see e.g. unsupervised PCA Figure 2, tICA Figure 3 or supervised PCA Figure 5), we surmise that this is an indication that we have an early transition-state here. Hence, this information can be leveraged for future catalyst design.

Conclusion

We developed a workflow to identify and quantify the reaction coordinate from a set of trajectories, detecting chemically intuitive and less intuitive contributions.

By devising a high throughput QM/MM MD workflow, we were able to study the C-N coupling reaction dynamics of a supramolecular Cu-calix[8]arene catalyst under experimental conditions. This development is a crucial step towards a predictive operando model for complex catalytic reactions. It allowed us to extract not only reaction energies and barriers with uncertainties but also provide insights into the intricate dynamic nature of the macrocyclic transition metal catalyst in explicit solvent.

To process the vast amount of data, interpretable machine learning techniques proved to be invaluable, thanks to the ability to map results back to structural changes. However, a consensus model is needed to eliminate the inherent variability/instability of the individual ML approaches.

By performing a causal analysis of the internal coordinates of a system that contribute to the categorisation of the ensembles, an extra temporal dimension can be added to the reaction coordinate, allowing us to explain the chemical reaction as a sequence of movements leading up to C-N bond formation. One can then utilize this information to pinpoint the exact source (group of atoms) that triggers the reaction and by suitable chemical mutation tweak the system in such way that the reactivity is enhanced. By checking the outcome of decision trees and rules run on this modified system, one can gauge the impact of a specific change on the reaction coordinate. Through the implementation of the methodology in a straight forward workflow and the restriction to analysis of structural parameters, this technique is accessible to the non-expert user, while the obtained results, that is changes in coordinates during the reaction, can easily be understood by a general chemist.

Our methodology was demonstrated on a highly flexible Cu-calix[8]arene catalyst, but serves as a blueprint to identify and quantify the reaction coordinate in any dynamic chemical system, from small (in)organic complexes to large (bio-)molecules.

Methods

Workflow for Determination of Complex Reaction Coordinates

The multistep protocol developed to investigate the C-N coupling dynamics with a Cu-calix[8]arene catalyst is highlighted in Figure 10. It involves high throughput explicit solvent MD sampling of the reaction step, followed by machine learning analysis, where consensus features are extracted. These are utilized for qualitative and quantitative analysis of the reaction coordinate. As a last step, the time evolution of the system is considered by applying a causality model that allows to redefine the reaction coordinate as a sequence of individual movements of groups of atoms.

Simulation Protocol

For the QM/MM MD simulation, we chose the in-house developed ab-initio quantum mechanical charge field (QMCF) molecular dynamics approach[70] using the link-bond method to describe the bonds crossing between QM and MM.[71,72] The simulation parameters were set up using the GAFF[73] force-field[74], using the PyConSolv[19] 1.0.0 tool, with default settings. For the geometry optimization, the PBE0 functional[75] was used with the def2-SVP basis set[76] and D3 dispersion corrections[77] in implicit chloroform, using CPCM.[20] The system was solvated in a cubic periodic box with 1708 chloroform molecules. For detailed information regarding the simulation parameters, see the supporting information. The 54 atoms at the centre of the calixarene cage were included in the QM zone (see SI Figure S1). The quantum mechanical calculations were performed at two distinct levels. The semi-empirical method GFN2-xTB was utilized, providing a vast speedup of the QM calculation. As semi-empirical methods require benchmarking,[78–80] a full DFT reference was used, with PBE0/def2-SVP/D3 using Turbomole.[81]

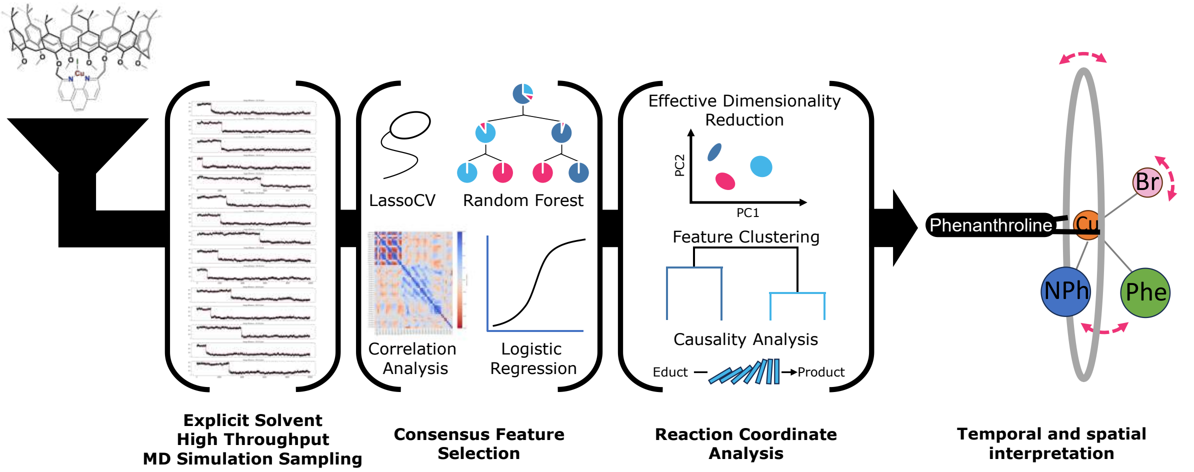


Figure 10. General workflow for extracting highly detailed information about the reaction coordinate from both a correlational and causal point of view.

To generate an appropriate starting structure the system was equilibrated and a 100ns MM/MD simulation was carried out, using the multistep protocol implemented in PyConSolv.[19] We conducted 152 simulation runs using the QM/MM MD protocol of the reductive elimination step, employing GFN2-xTB as QM method. Among these runs, 142 simulations captured the reductive elimination step and were utilized for energy and structural analysis.

Ensemble Labelling and Reaction Energy

From the MD simulation trajectories, cartesian coordinates and QM energies of the catalyst were extracted. The QM energy was used to label the individual frames as educt, transition state and product, while employing a filter function to minimize random fluctuations. The transition states were identified as the frames that define the last energy maximum before the large drop in energy associated with the formation of the product (See SI for details). The reaction barrier was defined as the difference between the transition state structure energy and the maximum value of a sigmoid function fitted through the reaction profile. The reaction energy, with uncertainty, was calculated based on the energy difference between the average energy of the product and educt ensembles, respectively. As DFT sampling on a similar scale as GFN2-xTB is not achievable, we cannot compare the reaction energies using the DFT educts and products. Thus, we used the fitted sigmoids to calculate the energy difference between the educts and product, by subtracting the highest sigmoid value from the lowest.

Identification of the Reaction Coordinate

We resorted to three coordinate systems to describe the reaction coordinate and significant secondary movements. Firstly, a cartesian coordinate system, obtained by combining all simulations and aligning them on the rigid phenanthroline bridge. This helps mitigate the impact of rotational and translational movements of the system during the simulations.[82] Secondly, we generated a set of internal coordinates for all trajectories, using the MDanalysis package.[83] This has the advantage of removing the issue of noise due to alignment artefacts, yet introduces more correlational effects.[82] Thirdly, we generated a reduced set of internal coordinates, describing highly rigid chemical moieties by their centre of mass (see SI for details)

To determine the importance of each feature in distinguishing between the three ensembles (educt, TS, product), and therefore its contribution to the reaction coordinate, we performed dimensionality reduction on the coordinate sets, aiming to increase separation between ensembles.

To this end, we utilized Principal Component Analysis (PCA)[34] and Time-lagged Independent Component Analysis (tICA),[35] which complement each other in regards to addressing the variances present in the dataset,[84,85] as unsupervised methods. For supervised dimensionality reduction, we opt for Linear Discriminant Analysis (LDA)[86] due to its efficacy in separating distinct classes within a given dataset. To further enhance the separation capability of PCA and LDA, several methods of automated feature selection were chosen and implemented, namely Recursive Feature Elimination with Cross Validation[87] (RFECV) using both Random Forest [88] (RF) and Logistic Regression[89] as classifier models, and Lasso[90] with Cross Validation (LassoCV), all using fivefold stratified cross-validation.[91] The RF and Logistic Regression classifiers were also evaluated manually for performance, eliminating features under a certain threshold (see SI for details). Additionally, a simple Pearson correlation evaluation between the energy and each feature was performed, removing any coordinate which exhibited low correlation (higher than -0.75 and lower than 0.75). For the final feature selection, we employed a consensus model that extracted high-contributing features. These features were identified by performing a PCA and LDA based on the Recursive Feature Elimination with Cross-Validation (RFECV), LassoCV, and manual assessments approaches. Subsequently, we computed the average contributions to the first three principal components, as well as the two LDA components. The selected features needed to appear in at least 70% of the elimination models to be deemed significant. This process yielded a set of 49 internal coordinates for the PCA and 46 for the LDA. These coordinates were further subjected to causal inference (as detailed in the Statistical Methods section) and correlation interpretation. To enhance comprehensibility, we applied a hierarchical clustering algorithm to group together similar movements within the consensus features.[92]

In tandem with automated feature reduction methodologies, we leveraged decision trees to highlight features from our dataset. These decision trees underwent training under two distinct conditions: one with class balancing and another without. Furthermore, we adopted a decision rule framework, employing the skoperules[93] library. Within this framework, a random forest bagging classifier, consisting of 30 decision trees, was used to provide a semantically quantitative characterization of the three ensembles.

Statistical Model for Causality Inference

We aimed to infer the cause of the onset of the chemical reaction. To this end, we focused our investigation on the educt and TS ensembles of each trajectory. We eliminated trajectories with few educt structures (less than 100), thus having a total of 139 reaction sampling events. To infer Granger Causality[68] (GC) we followed the protocol outlined by Toda and Yamamoto.[69] This involved performing the augmented Dickey-Fuller[94] and Kwiatkowski–Phillips–Schmidt–Shin[95] tests to ascertain the stationarity of the various time series. Most of the time series were deemed stationary, with only a couple of the features presenting non-stationarity, which were rendered stationary through differencing (see SI for the critical and test statistics for each test and trajectory).[96] A multivariate vector autoregression model[97,98] was constructed and fitted with lag times varying from 0 to 50, for each time series. The appropriate lag time was chosen for each trajectory, based on the Akaike information criterion[97,99] (see SI for additional information). The correlated time series were checked for cointegration using the Johansen test.[100] Finally, we calculated a GC matrix for each feature, for every simulation, resulting in a total of 139 matrices, using Vector Autoregressive (VAR) models trained at the appropriate lag time. To account for the occurrence of false positives with repeated sampling of the reaction, we applied the false discovery rate (FDR) correction proposed by Benjamini and Hochberg,[101] with a threshold alpha of 0.1. The threshold for the GC test was set to $p < 0.05$. The full p-values for the causality matrices, non-FDR corrected, can be found in the SI, along with the results of all statistical tests.

Technical Note

Our implementation leveraged the following Python packages for data processing, statistical analysis and machine learning: pandas 2.2.1,[102] numpy 1.26.4,[103] scikit-learn 1.4.1.post1,[104] statsmodels 0.14.1,[105] scipy 1.12.0,[106] mdanalysis 2.7.0,[83,107] skoperules 1.0.1,[93] and pyemma 2.5.12.[108] For visualizations, we utilized dtreeviz 2.2.217,[109] seaborn 0.13.2,[110] and matplotlib 3.8.318.[111] The code used for analysis is made available as a jupyter notebook on Github (https://github.com/PodewitzLab/MLReactCoord) and will be integrated in the analysis suite of a future version of PyConSolv (https://github.com/PodewitzLab/PyConSolv).

Data availability

The p-values for the statistical analysis are available on Github (https://github.com/PodewitzLab/MLReactCoord), alongside two example trajectories. Due to the size of the simulation data, it can be made available upon request from the corresponding authors.

Code availability

The python code used for analysis is made available on Github (https://github.com/PodewitzLab/MLReactCoord) and the respective functions will be implemented in a future version of PyConSolv (https://github.com/PodewitzLab/PyConSolv) to facilitate a broad applicability.

Author contributions

The project was conceived by R.A.T. and M.P., while I.C provided chemical input for the system, required to devise the project. R.A.T. performed the QM/MM/MD calculations together with T.S.H, while J.G. implemented the rescaling barostat into the QMCFC package, specifically for this project. R.A.T. devised the Jupyter Notebook to conduct the analyses. Analyses were edited by M.P. R.A.T wrote the original draft, M.P. T.S.H and I.C. edited the draft. All authors agree with the final version of the draft. M. P. acquired the funding and supervised this project. The computational resources were provided by T.S.H and M.P.

Acknowledgment

The authors could like to thank Jonny Proppe for fruitful discussion. M. P. would like to thank the Austrian Science Fund (FWF) for financial support (P-33528).

Author Information

Corresponding Authors

T.S. Hofer - Institute of Inorganic and Theoretical Chemistry, Leopold Franzens University of Innsbruck, Innrain 80/82, 6020, Innsbruck (Austria); https://orcid.org/0000-0002-6559-1513
Email: t.hofer@uibk.ac.at

M. Podewitz- Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Wien (Austria); https://orcid.org/0000-0001-7256-1219
Email: maren.podewitz@tuwien.ac.at

Authors

R. A. Talmazan - Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Wien (Austria); https://orcid.org/0000-0001-6678-7801

J. Gamper - Institute of Inorganic and Theoretical Chemistry, Leopold Franzens University of Innsbruck, Innrain 80/82, 6020, Innsbruck (Austria); https://orcid.org/0000-0003-1136-2536

I. Castillo - Instituto de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Ciudad de México (México); https://orcid.org/0000-0002-4876-4339

References

(1) Ringe, D.; Petsko, G. A. How Enzymes Work. *Science* **2008**, *320* (5882), 1428–1429. https://doi.org/10.1126/science.1159747.
(2) Koblenz, T. S.; Wassenaar, J.; Reek, J. N. H. Reactivity within a Confined Self-Assembled Nanospace. *Chem. Soc. Rev.* **2008**, *37* (2), 247–262. https://doi.org/10.1039/B614961H.
(3) Raynal, M.; Ballester, P.; Vidal-Ferran, A.; Leeuwen, P. W. N. M. van. Supramolecular Catalysis. Part 1: Non-Covalent Interactions as a Tool for Building and Modifying Homogeneous Catalysts. *Chem. Soc. Rev.* **2014**, *43* (5), 1660–1733. https://doi.org/10.1039/C3CS60027K.
(4) Raynal, M.; Ballester, P.; Vidal-Ferran, A.; Leeuwen, P. W. N. M. van. Supramolecular Catalysis. Part 2: Artificial Enzyme Mimics. *Chem. Soc. Rev.* **2014**, *43* (5), 1734–1787. https://doi.org/10.1039/C3CS60037H.
(5) Pachisia, S.; Gupta, R. Supramolecular Catalysis: The Role of H-Bonding Interactions in Substrate Orientation and Activation. *Dalton Trans.* **2021**, *50* (42), 14951–14966. https://doi.org/10.1039/D1DT02131A.
(6) Olivo, G.; Capocasa, G.; Giudice, D. D.; Lanzalunga, O.; Stefano, S. D. New Horizons for Catalysis Disclosed by Supramolecular Chemistry. *Chem. Soc. Rev.* **2021**, *50* (13), 7681–7724. https://doi.org/10.1039/D1CS00175B.

(7) Talmazan, R. A.; Refugio Monroy, J.; del Río-Portilla, F.; Castillo, I.; Podewitz, M. Encapsulation Enhances the Catalytic Activity of C-N Coupling: Reaction Mechanism of a Cu(I)/Calix[8]Arene Supramolecular Catalyst. *ChemCatChem* **2022**, *14* (20). https://doi.org/10.1002/cctc.202200662.
(8) Guzmán-Percástegui, E.; J. Hernández, D.; Castillo, I. Calix[8]Arene Nanoreactor for Cu(i)-Catalysed C–S Coupling. *Chemical Communications* **2016**, *52* (15), 3111–3114. https://doi.org/10.1039/C5CC09232A.
(9) Sciortino, G.; Maseras, F. Computational Study of Homogeneous Multimetallic Cooperative Catalysis. *Top Catal* **2022**, *65* (1), 105–117. https://doi.org/10.1007/s11244-021-01493-2.
(10) Artús Suàrez, L.; Balcells, D.; Nova, A. Computational Studies on the Mechanisms for Deaminative Amide Hydrogenation by Homogeneous Bifunctional Catalysts. *Top Catal* **2022**, *65* (1), 82–95. https://doi.org/10.1007/s11244-021-01542-w.
(11) Jones, G. O.; Liu, P.; Houk, K. N.; Buchwald, S. L. Computational Explorations of Mechanisms and Ligand-Directed Selectivities of Copper-Catalyzed Ullmann-Type Reactions. *J. Am. Chem. Soc.* **2010**, *132* (17), 6205–6213. https://doi.org/10.1021/ja100739h.
(12) Larsson, P.-F.; Wallentin, C.-J.; Norrby, P.-O. Mechanistic Aspects of Submol % Copper-Catalyzed C⍰N Cross-Coupling. *ChemCatChem* **2014**, *6* (5), 1277–1282. https://doi.org/10.1002/cctc.201301088.
(13) Fey, N.; Lynam, J. M. Computational Mechanistic Study in Organometallic Catalysis: Why Prediction Is Still a Challenge. *WIREs Computational Molecular Science* **2022**, *12* (4), e1590. https://doi.org/10.1002/wcms.1590.
(14) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9* (8), 6803–6813. https://doi.org/10.1021/acscatal.9b01537.
(15) Besora, M.; Braga, A. A. C.; Ujaque, G.; Maseras, F.; Lledós, A. The Importance of Conformational Search: A Test Case on the Catalytic Cycle of the Suzuki–Miyaura Cross-Coupling. *Theor Chem Acc* **2011**, *128* (4–6), 639–646. https://doi.org/10.1007/s00214-010-0823-6.
(16) Podewitz, M.; Sen, S.; Buchmeiser, M. R. On the Origin of E-Selectivity in the Ring-Opening Metathesis Polymerization with Molybdenum Imido Alkylidene N-Heterocyclic Carbene Complexes. *Organometallics* **2021**, *40* (15), 2478–2488. https://doi.org/10.1021/acs.organomet.1c00229.
(17) Eisenstein, O.; Ujaque, G.; Lledós, A. What Makes a Good (Computed) Energy Profile? In *New Directions in the Modeling of Organometallic Reactions*; Lledós, A., Ujaque, G., Eds.; Topics in Organometallic Chemistry; Springer International Publishing: Cham, 2020; pp 1–38. https://doi.org/10.1007/3418_2020_57.
(18) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST—A Program for the Exploration of Low-Energy Molecular Chemical Space. *The Journal of Chemical Physics* **2024**, *160* (11), 114110. https://doi.org/10.1063/5.0197592.
(19) Talmazan, R. A.; Podewitz, M. PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent. *J. Chem. Inf. Model.* **2023**, *63* (17), 5400–5407. https://doi.org/10.1021/acs.jcim.3c00798.
(20) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001. https://doi.org/10.1021/jp9716997.
(21) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, No. 5, 799–805. https://doi.org/10.1039/P29930000799.
(22) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and

Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. https://doi.org/10.1021/jp810292n.

(23) Simm, G. N.; Türtscher, P. L.; Reiher, M. Systematic Microsolvation Approach with a Cluster-Continuum Scheme and Conformational Sampling. *Journal of Computational Chemistry* **2020**, *41* (12), 1144–1155. https://doi.org/10.1002/jcc.26161.

(24) Steiner, M.; Holzknecht, T.; Schauperl, M.; Podewitz, M. Quantum Chemical Microsolvation by Automated Water Placement. *Molecules* **2021**, *26* (6), 1793. https://doi.org/10.3390/molecules26061793.

(25) Joy, J.; Ess, D. H. Direct Dynamics Trajectories Demonstrate Dynamic Matching and Nonstatistical Radical Pair Intermediates during Fe-Oxo-Mediated C–H Functionalization Reactions. *J. Am. Chem. Soc.* **2023**, *145* (13), 7628–7637. https://doi.org/10.1021/jacs.3c01196.

(26) Yang, Z.; Jamieson, C. S.; Xue, X.-S.; Garcia-Borràs, M.; Benton, T.; Dong, X.; Liu, F.; Houk, K. N. Mechanisms and Dynamics of Reactions Involving Entropic Intermediates. *TRECHEM* **2019**, *1* (1), 22–34. https://doi.org/10.1016/j.trechm.2019.01.009.

(27) Ess, D. H. Quasiclassical Direct Dynamics Trajectory Simulations of Organometallic Reactions. *Acc. Chem. Res.* **2021**, *54* (23), 4410–4422. https://doi.org/10.1021/acs.accounts.1c00575.

(28) Joy, J.; Schaefer, A. J.; Teynor, M. S.; Ess, D. H. Dynamical Origin of Rebound versus Dissociation Selectivity during Fe-Oxo-Mediated C–H Functionalization Reactions. *J. Am. Chem. Soc.* **2024**, *146* (4), 2452–2464. https://doi.org/10.1021/jacs.3c09891.

(29) Warshel, A.; Karplus, M. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.* **1972**, *94* (16), 5612–5625. https://doi.org/10.1021/ja00771a014.

(30) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *Journal of Molecular Biology* **1976**, *103* (2), 227–249. https://doi.org/10.1016/0022-2836(76)90311-9.

(31) Gao, J. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29* (6), 298–305. https://doi.org/10.1021/ar950140r.

(32) Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J. Computational Discovery of Transition-Metal Complexes: From High-Throughput Screening to Machine Learning. *Chem. Rev.* **2021**, *121* (16), 9927–10000. https://doi.org/10.1021/acs.chemrev.1c00347.

(33) Montavon, G.; Samek, W.; Müller, K.-R. Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing* **2018**, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011.

(34) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* **1987**, *2* (1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

(35) Molgedey, L.; Schuster, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **1994**, *72* (23), 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634.

(36) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Structure, Function, and Bioinformatics* **1993**, *17* (4), 412–425. https://doi.org/10.1002/prot.340170408.

(37) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *The Journal of Chemical Physics* **2013**, *139* (1), 015102. https://doi.org/10.1063/1.4811489.

(38) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J Chem Theory Comput* **2013**, *9* (4), 2000–2009. https://doi.org/10.1021/ct300878a.

(39) Hare, S. R.; Bratholm, L. A.; Glowacki, D. R.; Carpenter, B. K. Low Dimensional Representations along Intrinsic Reaction Coordinates and Molecular Dynamics Trajectories Using Interatomic Distance Matrices. *Chem. Sci.* **2019**, *10* (43), 9954–9968. https://doi.org/10.1039/C9SC02742D.

(40) *Reaction Space Projector (ReSPer) for Visualizing Dynamic Reaction Routes Based on Reduced-Dimension Space | Topics in Current Chemistry*. https://link.springer.com/article/10.1007/s41061-022-00377-7 (accessed 2024-06-04).

(41) Imbens, G. W.; Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: Cambridge, 2015. https://doi.org/10.1017/CBO9781139025751.

(42) Xing, Q.; Wu, C.; Chen, F.; Liu, J.; Pradhan, P.; Bryan, B. A.; Schaubroeck, T.; Carrasco, L. R.; Gonsamo, A.; Li, Y.; Chen, X.; Deng, X.; Albanese, A.; Li, Y.; Xu, Z. Intranational Synergies and Trade-Offs Reveal Common and Differentiated Priorities of Sustainable Development Goals in China. *Nat Commun* **2024**, *15* (1), 2251. https://doi.org/10.1038/s41467-024-46491-6.

(43) Cárdenas-García, P. J.; Brida, J. G.; Segarra, V. Modeling the Link between Tourism and Economic Development: Evidence from Homogeneous Panels of Countries. *Humanit Soc Sci Commun* **2024**, *11* (1), 1–12. https://doi.org/10.1057/s41599-024-02826-8.

(44) Kaufmann, R. K.; Newberry, D.; Xin, C.; Gopal, S. Feedbacks among Electric Vehicle Adoption, Charging, and the Cost and Installation of Rooftop Solar Photovoltaics. *Nat Energy* **2021**, *6* (2), 143–149. https://doi.org/10.1038/s41560-020-00746-w.

(45) Tanaka, T.; Guo, J. International Price Volatility Transmission and Structural Change: A Market Connectivity Analysis in the Beef Sector. *Humanit Soc Sci Commun* **2020**, *7* (1), 1–13. https://doi.org/10.1057/s41599-020-00657-x.

(46) Ouyang, T.; Liu, F.; Huang, B. Dynamic Econometric Analysis on Influencing Factors of Production Efficiency in Construction Industry of Guangxi Province in China. *Sci Rep* **2022**, *12* (1), 17509. https://doi.org/10.1038/s41598-022-22374-y.

(47) Le, T. Increased Impact of the El Niño–Southern Oscillation on Global Vegetation under Future Warming Environment. *Sci Rep* **2023**, *13* (1), 14459. https://doi.org/10.1038/s41598-023-41590-8.

(48) Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; Glymour, C.; Kretschmer, M.; Mahecha, M. D.; Muñoz-Marí, J.; van Nes, E. H.; Peters, J.; Quax, R.; Reichstein, M.; Scheffer, M.; Schölkopf, B.; Spirtes, P.; Sugihara, G.; Sun, J.; Zhang, K.; Zscheischler, J. Inferring Causation from Time Series in Earth System Sciences. *Nat Commun* **2019**, *10* (1), 2553. https://doi.org/10.1038/s41467-019-10105-3.

(49) Kretschmer, M.; Coumou, D.; Donges, J. F.; Runge, J. Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate* **2016**, *29* (11), 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1.

(50) Triacca, U. Is Granger Causality Analysis Appropriate to Investigate the Relationship between Atmospheric Concentration of Carbon Dioxide and Global Surface Air Temperature? *Theor. Appl. Climatol.* **2005**, *81* (3), 133–135. https://doi.org/10.1007/s00704-004-0112-1.

(51) McGraw, M. C.; Barnes, E. A. Memory Matters: A Case for Granger Causality in Climate Variability Studies. *Journal of Climate* **2018**, *31* (8), 3289–3300. https://doi.org/10.1175/JCLI-D-17-0334.1.

(52) Hill, S. M.; Heiser, L. M.; Cokelaer, T.; Unger, M.; Nesser, N. K.; Carlin, D. E.; Zhang, Y.; Sokolov, A.; Paull, E. O.; Wong, C. K.; Graim, K.; Bivol, A.; Wang, H.; Zhu, F.; Afsari, B.; Danilova, L. V.; Favorov, A. V.; Lee, W. S.; Taylor, D.; Hu, C. W.; Long, B. L.; Noren, D. P.; Bisberg, A. J.; Mills, G. B.; Gray, J. W.; Kellen, M.; Norman, T.; Friend, S.; Qutub, A. A.; Fertig, E. J.; Guan, Y.; Song, M.; Stuart, J. M.; Spellman, P. T.; Koeppl, H.; Stolovitzky, G.; Saez-Rodriguez, J.; Mukherjee, S. Inferring Causal Molecular Networks: Empirical Assessment through a Community-Based Effort. *Nat Methods* **2016**, *13* (4), 310–318. https://doi.org/10.1038/nmeth.3773.

(53) Walter, J. Establishing Microbiome Causality to Tackle Malnutrition. *Nat Microbiol* **2024**, *9* (4), 884–885. https://doi.org/10.1038/s41564-024-01653-6.

(54) Friston, K. J.; Harrison, L.; Penny, W. Dynamic Causal Modelling. *NeuroImage* **2003**, *19* (4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7.

(55) Duggento, A.; Passamonti, L.; Valenza, G.; Barbieri, R.; Guerrisi, M.; Toschi, N. Multivariate Granger Causality Unveils Directed Parietal to Prefrontal Cortex Connectivity during Task-Free MRI. *Sci Rep* **2018**, *8* (1), 5571. https://doi.org/10.1038/s41598-018-23996-x.

(56) Zheng, J.; Anderson, K. L.; Leal, S. L.; Shestyuk, A.; Gulsen, G.; Mnatsakanyan, L.; Vadera, S.; Hsu, F. P. K.; Yassa, M. A.; Knight, R. T.; Lin, J. J. Amygdala-Hippocampal Dynamics during Salient Information Processing. *Nat Commun* **2017**, *8* (1), 14413. https://doi.org/10.1038/ncomms14413.

(57) Kamberaj, H.; van der Vaart, A. Correlated Motions and Interactions at the Onset of the DNA-Induced Partial Unfolding of Ets-1. *Biophys J* **2009**, *96* (4), 1307–1317. https://doi.org/10.1016/j.bpj.2008.11.019.

(58) Kamberaj, H.; van der Vaart, A. Extracting the Causality of Correlated Motions from Molecular Dynamics Simulations. *Biophys J* **2009**, *97* (6), 1747–1755. https://doi.org/10.1016/j.bpj.2009.07.019.

(59) Sobieraj, M.; Setny, P. Granger Causality Analysis of Chignolin Folding. *J. Chem. Theory Comput.* **2022**, *18* (3), 1936–1944. https://doi.org/10.1021/acs.jctc.1c00945.

(60) Buskes, M. J.; Blanco, M.-J. Impact of Cross-Coupling Reactions in Drug Discovery and Development. *Molecules* **2020**, *25* (15), 3493. https://doi.org/10.3390/molecules25153493.

(61) Gay, M.; Carato, P.; Coevoet, M.; Renault, N.; Larchanché, P.-E.; Barczyk, A.; Yous, S.; Buée, L.; Sergeant, N.; Melnyk, P. New Phenylaniline Derivatives as Modulators of Amyloid Protein Precursor Metabolism. *Bioorganic & Medicinal Chemistry* **2018**, *26* (8), 2151–2164. https://doi.org/10.1016/j.bmc.2018.03.016.

(62) Jiang, R.; Li, L.; Sheng, T.; Hu, G.; Chen, Y.; Wang, L. Edge-Site Engineering of Atomically Dispersed Fe–N4 by Selective C–N Bond Cleavage for Enhanced Oxygen Reduction Reaction Activities. *J. Am. Chem. Soc.* **2018**, *140* (37), 11594–11598. https://doi.org/10.1021/jacs.8b07294.

(63) Hayakawa, S.; Kawasaki, A.; Hong, Y.; Uraguchi, D.; Ooi, T.; Kim, D.; Akutagawa, T.; Fukui, N.; Shinokubo, H. Inserting Nitrogen: An Effective Concept To Create Nonplanar and Stimuli-Responsive Perylene Bisimide Analogues. *J. Am. Chem. Soc.* **2019**, *141* (50), 19807–19816. https://doi.org/10.1021/jacs.9b09556.

(64) Izumi, S.; Higginbotham, H. F.; Nyga, A.; Stachelek, P.; Tohnai, N.; Silva, P. de; Data, P.; Takeda, Y.; Minakata, S. Thermally Activated Delayed Fluorescent Donor–Acceptor–Donor–Acceptor π-Conjugated Macrocycle for Organic Light-Emitting Diodes. *J. Am. Chem. Soc.* **2020**, *142* (3), 1482–1491. https://doi.org/10.1021/jacs.9b11578.

(65) Picini, F.; Schneider, S.; Gavat, O.; Vargas Jentzsch, A.; Tan, J.; Maaloum, M.; Strub, J.-M.; Tokunaga, S.; Lehn, J.-M.; Moulin, E.; Giuseppone, N. Supramolecular Polymerization of Triarylamine-Based Macrocycles into Electroactive Nanotubes. *J. Am. Chem. Soc.* **2021**, *143* (17), 6498–6504. https://doi.org/10.1021/jacs.1c00623.

(66) Berlanga-Vázquez, A.; Talmazan, R. A.; Reyes-Mata, C. A.; Percástegui, E. G.; Flores-Alamo, M.; Podewitz, M.; Castillo, I. Conformational Effects of Regioisomeric Substitution on the Catalytic Activity of Copper/Calix[8]Arene C−S Coupling. *Eur J Inorg Chem* **2022**. https://doi.org/10.1002/ejic.202200596.

(67) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

(68) Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* **1969**, *37* (3), 424–438. https://doi.org/10.2307/1912791.
(69) Toda, H. Y.; Yamamoto, T. Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics* **1995**, *66* (1), 225–250. https://doi.org/10.1016/0304-4076(94)01616-8.
(70) Rode, B. M.; Hofer, T. S.; Randolf, B. R.; Schwenk, C. F.; Xenides, D.; Vchirawongkwin, V. Ab Initio Quantum Mechanical Charge Field (QMCF) Molecular Dynamics: A QM/MM – MD Procedure for Accurate Simulations of Ions and Complexes. *Theor Chem Acc* **2006**, *115* (2), 77–85. https://doi.org/10.1007/s00214-005-0049-1.
(71) Amara, P.; Field, M. J. Evaluation of an Ab Initio Quantum Mechanical/Molecular Mechanical Hybrid-Potential Link-Atom Method. *Theor Chem Acc* **2003**, *109* (1), 43–52. https://doi.org/10.1007/s00214-002-0413-3.
(72) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH3Cl + Cl− Exchange Reaction and Gas Phase Protonation of Polyethers. *Journal of Computational Chemistry* **1986**, *7* (6), 718–730. https://doi.org/10.1002/jcc.540070604.
(73) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K. AMBER20. *University of California, San Francisco* **2020**.
(74) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.
(75) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *The Journal of Chemical Physics* **1999**, *110* (13), 6158–6170. https://doi.org/10.1063/1.478522.
(76) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. https://doi.org/10.1039/B508541A.
(77) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J Chem Phys* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.
(78) Bursch, M.; Hansen, A.; Pracht, P.; Kohn, J. T.; Grimme, S. Theoretical Study on Conformational Energies of Transition Metal Complexes. *Phys. Chem. Chem. Phys.* **2021**, *23* (1), 287–299. https://doi.org/10.1039/D0CP04696E.
(79) Husch, T.; Vaucher, A. C.; Reiher, M. Semiempirical Molecular Orbital Models Based on the Neglect of Diatomic Differential Overlap Approximation. *International Journal of Quantum Chemistry* **2018**, *118* (24), e25799. https://doi.org/10.1002/qua.25799.
(80) Podewitz, M. Towards Predictive Computational Catalysis – a Case Study of Olefin Metathesis with Mo Imido Alkylidene N-Heterocyclic Carbene Catalysts. In *Chemical Modelling: Volume 17*; Bahmann, H., Tremblay, J. C., Eds.; The Royal Society of Chemistry, 2022; Vol. 17, p 0. https://doi.org/10.1039/9781839169342-00001.
(81) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; Hellweg, A.; Helmich-Paris, B.; Holzer, C.; Huniar, U.; Kaupp, M.; Marefat Khah, A.; Karbalaei Khani, S.; Müller, T.; Mack, F.; Nguyen, B. D.; Parker, S. M.; Perlt, E.; Rappoport, D.; Reiter, K.; Roy, S.; Rückert, M.; Schmitz, G.; Sierka, M.; Tapavicza, E.; Tew, D. P.; van Wüllen, C.; Voora, V. K.; Weigend, F.; Wodyński, A.; Yu, J. M. TURBOMOLE: Modular Program Suite for Ab Initio Quantum-Chemical and Condensed-Matter Simulations. *The Journal of Chemical Physics* **2020**, *152* (18), 184107. https://doi.org/10.1063/5.0004635.

(82) Sittel, F.; Jain, A.; Stock, G. Principal Component Analysis of Molecular Dynamics: On the Use of Cartesian vs. Internal Coordinates. *The Journal of Chemical Physics* **2014**, *141* (1), 014111. https://doi.org/10.1063/1.4885338.

(83) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105. https://doi.org/10.25080/Majora-629e541a-00e.

(84) Schultze, S.; Grubmüller, H. Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *J. Chem. Theory Comput.* **2021**, *17* (9), 5766–5776. https://doi.org/10.1021/acs.jctc.1c00273.

(85) Wu, H.; Nüske, F.; Paul, F.; Klus, S.; Koltai, P.; Noé, F. Variational Koopman Models: Slow Collective Variables and Molecular Kinetics from Short off-Equilibrium Simulations. *The Journal of Chemical Physics* **2017**, *146* (15), 154104. https://doi.org/10.1063/1.4979344.

(86) Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Mullers, K. R. Fisher Discriminant Analysis with Kernels. In *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*; 1999; pp 41–48. https://doi.org/10.1109/NNSP.1999.788121.

(87) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning* **2002**, *46* (1), 389–422. https://doi.org/10.1023/A:1012487302797.

(88) Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(89) Cox, D. R. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society: Series B (Methodological)* **1958**, *20* (2), 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x.

(90) Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58* (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

(91) Zeng, X.; Martinez, T. R. Distribution-Balanced Stratified Cross-Validation for Accuracy Estimation. *Journal of Experimental & Theoretical Artificial Intelligence* **2000**, *12* (1), 1–12. https://doi.org/10.1080/095281300146272.

(92) Chormunge, S.; Jena, S. Correlation Based Feature Selection with Clustering for High Dimensional Data. *Journal of Electrical Systems and Information Technology* **2018**, *5* (3), 542–549. https://doi.org/10.1016/j.jesit.2017.06.004.

(93) Scikit-Learn-Contrib/Skope-Rules, 2024. https://github.com/scikit-learn-contrib/skope-rules (accessed 2024-04-15).

(94) Dickey, D. A.; Fuller, W. A. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* **1979**, *74* (366), 427–431. https://doi.org/10.2307/2286348.

(95) Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics* **1992**, *54* (1), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y.

(96) Hyndman, R. J.; Athanasopoulos, G. Forecasting: Principles and Practice. **2018**.

(97) Lütkepohl, H. *Introduction to Multiple Time Series Analysis*; Springer: Berlin, Heidelberg, 1991. https://doi.org/10.1007/978-3-662-02691-5.

(98) Sims, C. A. Macroeconomics and Reality. *Econometrica* **1980**, *48* (1), 1–48. https://doi.org/10.2307/1912017.

(99) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **1974**, *19* (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

(100) Johansen, S. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* **1988**, *12* (2), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3.

(101) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995**, *57* (1), 289–300.

(102) The pandas development team. Pandas-Dev/Pandas: Pandas, 2024. https://doi.org/10.5281/zenodo.10957263.

(103) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(104) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12* (85), 2825–2830.

(105) Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference*; Walt, S. van der, Millman, J., Eds.; 2010; pp 92–96. https://doi.org/10.25080/Majora-92bf1922-011.

(106) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. 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.

(107) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(108) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

(109) Parr, T. Parrt/Dtreeviz, 2024. https://github.com/parrt/dtreeviz (accessed 2024-04-15).

(110) Waskom, M. L. Seaborn: Statistical Data Visualization. *Journal of Open Source Software* **2021**, *6* (60), 3021. https://doi.org/10.21105/joss.03021.

(111) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(112) Runge, J.; Nowack, P.; Kretschmer, M.; Flaxman, S.; Sejdinovic, D. Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets. *Science Advances* **2019**, *5* (11), eaau4996. https://doi.org/10.1126/sciadv.aau4996.

Supporting Information for

From Data to Chemistry: Revealing Causality and Reaction Coordinates through Interpretable Machine Learning in Supramolecular Transition Metal Catalysis

R. A. Talmazan[1], J. Gamper[2], I. Castillo[3], T.S. Hofer*[2], M. Podewitz*[1]

[1] Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Wien (Austria), maren.podewitz@tuwien.ac.at

[2] Institute of Inorganic and Theoretical Chemistry, Leopold Franzens University of Innsbruck, Innrain 80/82, 6020, Innsbruck(Austria)

[3] Instituto de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Ciudad de México (México)

Table of Contents

1. Simulation protocol

The simulation temperature was maintained using the Bussi-Donadio-Parrinello thermostat with a was set to keep the temperature at 298.15K.[1] The velocity verlet[2]/RATTLE[3] was employed to integrate the equations of motion with a timestep of 2fs, with all hydrogen atom bonds being constrained. The long-range electrostatic interactions were considered via the reaction field method[4] with a Coulombic cutoff of 12.5 Å.

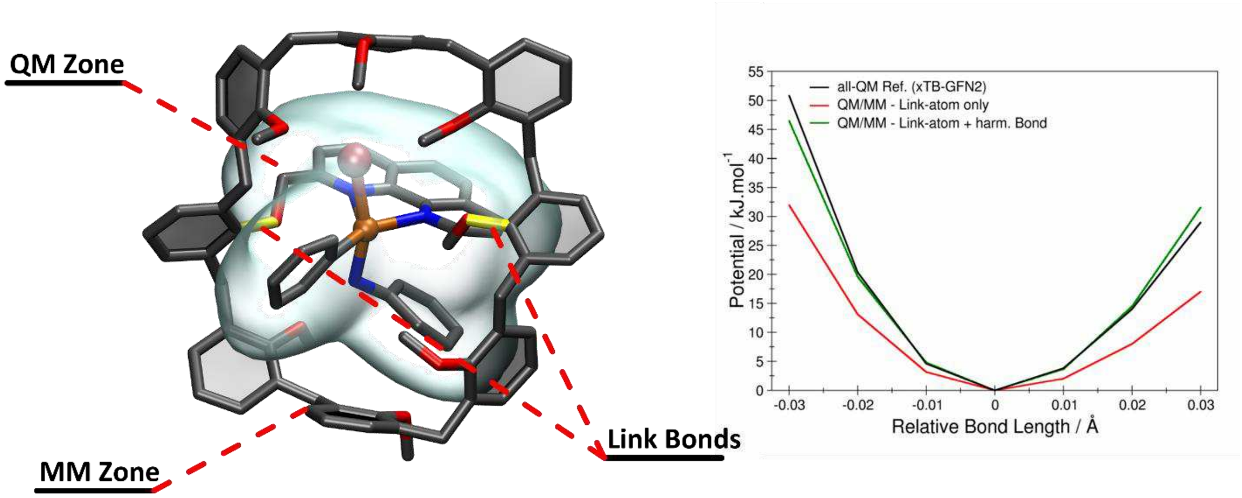


Figure S1 The centre of the figure shows the QM and MM areas of the model system. The 3 potential curves belong to the bond stretching energy of the C-O link bond. The black curve represents the energy calculated at a full QM level using xTB-GFN2. The red curve represents the energy calculated for the QM/MM setup. The green curve is the result of applying a properly adjusted MM potential along the link-bond. It can be seen that the green and black curves are in very good agreement in the minimum region.

The link bonds were defined between the oxygen atoms bonded to the phenanthroline bridge and the calixarene units 1 and 5 (see Figure S1). The parameters for the link bonds were calculated for both DFT and GFN2-xTB approaches (Table S1).

Through the use of a smoothing function applied to all particles located in the 0.2 Å smoothing region, the transition between the QM and MM regions was smoothed. The QM-MM transition was facilitated by using the link-atom approach. The bond linking the oxygen atom of the phenanthroline bridge with the carbon of the calixarene unit was partitioned and the unsaturated valence was satisfied by placing a hydrogen atom at a position determined by:

$$r_H = \rho(r_O - r_C) \qquad (1)$$

To obtain the correct parameters for the link-bond, full QM and MM calculations of a model system was performed (for full details on the model system, consult the SI). In Figure 1, the energy of the bond stretching can be seen to be very similar for both the model system in the case of a full QM approach, as well as in the link-bond, when the appropriate harmonic potential is applied.

In Table 1 the resulting parameters of the link-bond can be seen for both methods.

Table 1 Link bond parameters for the two methods.

	r_{eq} (Å)	ρ_{link}	K_{link}
PBE0/def2-SVP/D3	1.3687	0.7124	335.786
xTB-GFN2	1.3866	0.7085	154.4160

2. Reaction Energy and Labelling

The data that is extracted from the MD simulations consists of QM energies, as well as coordinates for the entire catalyst. The energy data contains the potential energy of the QM zone, embedded in the electrostatic point charges of the solvent molecules. As a consequence, random fluctuations, typically below 4 kJ/mol and categorised as noise, can be observed. To enhance the clarity of reaction energy trends, we employed the Savitzky–Golay filter[5] with a bandwidth of 10 and a second-degree polynomial fit to smooth the reaction profile. Following this, data was then categorized into three distinct groups: educts, transition states, and products. The educts and products are clearly distinguished, based on a large energy difference between the two ensembles (See Figure S2). The transition state corresponds to the energy maximum reached just before the significant energy drop associated with C-N bond formation (See Figure S2). To account for structural noise, we selected structures from frames immediately preceding and following the energy peak, resulting in a total of three structures per simulation run. These labels (Educt, TS and Product) were used for training the supervised machine learning models.

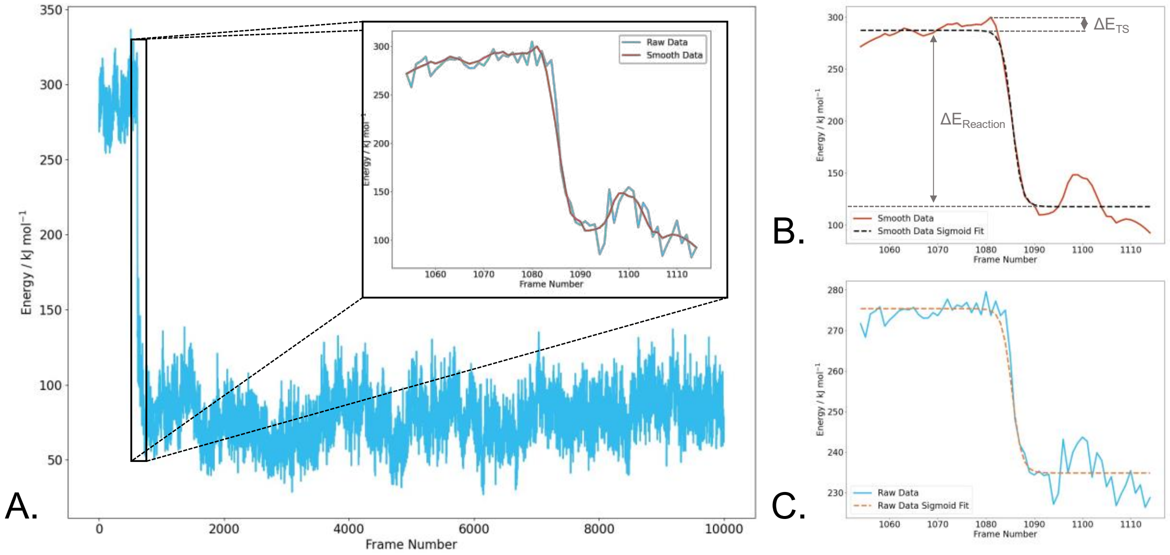


Figure S2 Energy Diagram example of a QM/MM MD simulation of the C-N coupling step of phenylbromide and aniline using a Cu(I)-calix[8]arene catalyst. A: Energy fluctuations for the whole duration of a simulation. Zoom focuses on the reductive elimination range in the simulation. The blue curve represents the raw energy data, the red curve represents the smoothed energy data; B: Energy curve and the sigmoid functions fitted to the smoothed reaction profile used for the reaction barrier calculations; C: Energy curve and the sigmoid functions fitted to the raw reaction profile used for the reaction energy calculations. Energies are for the QM zone, calculated with GFN2-xTB.

The reaction energy was calculated by analysing 30 frames before and after the transition state and fitting a sigmoid function to the energy profile. The reaction energy is defined as the difference between the minimum and maximum values obtained from this sigmoid fit, while the reaction barrier corresponds to the difference between the transition state structure and the maximum value of the sigmoid function (See Figure S2). Both raw data and smoothed data are used to compute these energy values to estimate the effect of the smoothing function.

We find that the GFN2-xTB reaction energies are in very good agreement with those computed with PBE0/def2-SVP/D3, as shown in Table S2.

Table S2 Reaction energies and barriers computed from the GFN2-xTB and DFT trajectories. All values are given in kJ mol^{-1}.

		Mean	Median	Standard Deviation
Smooth GFN2-xTB	Reaction	-165.0	-166.2	12.2
	Barrier	13.1	13.2	8.9
Raw GFN2-xTB	Reaction	-163.3	-164.2	12.3
	Barrier	11.8	12.5	13.4
Raw DFT	Reaction	-168.7		12.7
	Barrier	17.7		5.1

3. Coordinate Systems

We performed a PCA on the cartesian coordinates of the simulation trajectories of the Cu(I)-calix[8]arene (Figure S3A).This requires the simulation trajectories to be aligned, so that the PCA is able to separate the data according to the largest variance present within, thus capturing changes in system configuration.

As the calixarene system is highly flexible, there is considerable difficulty associated with performing a correct alignment of the simulation trajectories. This results in certain system-wide translational and rotational degrees of freedom remaining embedded in the cartesian coordinates. The PCA performance was lacklustre (Figure S3), where the 3 labelled groups cannot be properly separated. When investigating the Pearson correlation of the system coordinates, we observe signs of rotational motion (see Figure S3), indicating excessive noise in the trajectory coordinates.

Another strongly correlated area in the matrix is represented by the reaction centre, which corresponds to the formation of the C-N bond and rearrangement of the atoms surrounding the copper.

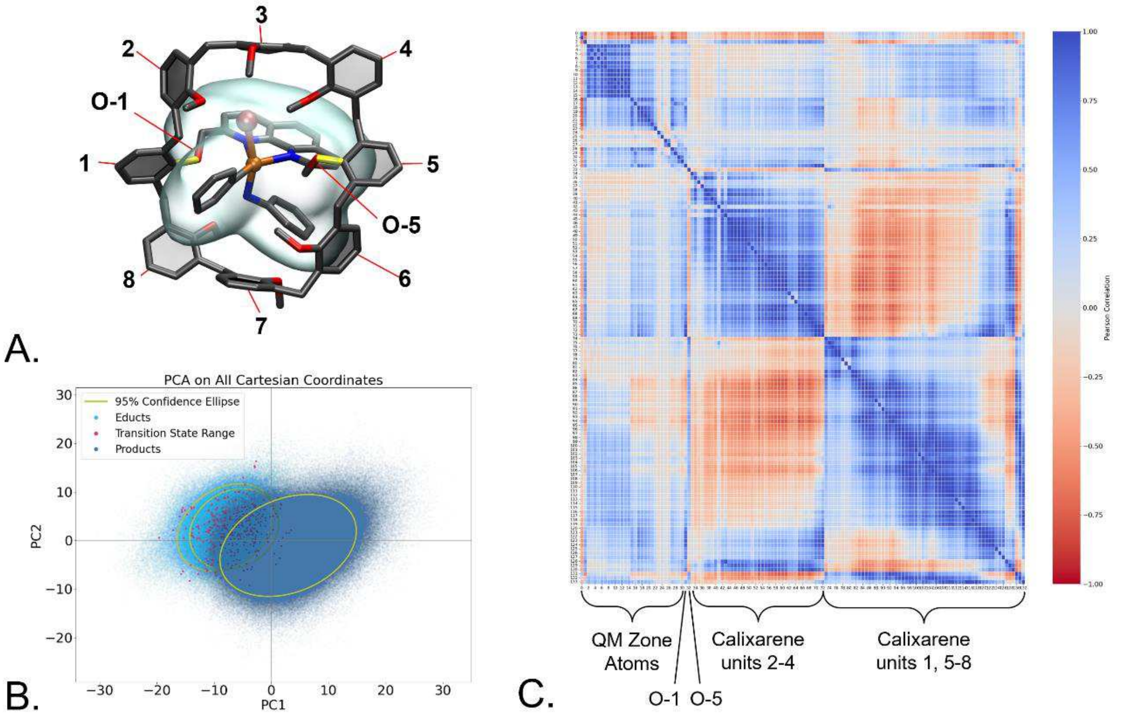


Figure S3 A. Calixarene structure with the numbered calixarene units and oxygen atoms. The bubble surrounds the atoms in the QM zone of the QM/MM MD. B. PCA Analysis performed on the XYZ coordinates of all heavy atoms of the calixarene, on all data points across simulations. Colouring performed according to the ensembles: light blue: educt; magenta: transition state; dark blue: product. C. Pearson correlation matrix of the cartesian coordinates.

To improve performance, we resorted to a highly reduced representation of the system. Specifically, we reduced each calixarene unit, the phenanthroline unit, and the phenyl and aniline moieties to their respective centres of mass. Furthermore, we augmented the internal coordinates by incorporating critical individual atoms. This included oxygen atoms associated with the calixarene, as well as copper (Cu) and bromine (Br) atoms, along with carbon (C) and nitrogen (N) atoms participating in the coupling bond. The reduction in dimensionality comes with little loss of information, as the atoms in the highly rigid calixarene units always move together. All coordinate systems were normalised, using min-max normalization, to improve the performance of the classification models. No normalisation was used for the decision tree and decision rules approach, as we aimed to quantify the features which distinguish the three ensembles.

4. Feature elimination

Using the ensemble labels derived from the reaction energetics, we were able to utilize supervised methods for feature reduction, in order to achieve a better separation of the reaction states. By feature reduction, we refer to a systematic elimination of internal coordinates, also known as features, which show no significant differences between the three ensembles. On the resulting reduced feature sets, we performed PCA (unsupervised) and Linear Discriminant Analysis (LDA, supervised) dimensionality reduction and evaluated the separation of the three ensembles, as well as the feature contributions to each principal component.

The feature elimination was performed using a fully manual approach, based on the correlation coefficient between the energy and features, a semi-automatic approach that involved training a classifier and using a cutoff for feature importance based on the mean of all feature importances, as well as a fully automated method. The latter involved automated recursive feature elimination with cross validation (RFECV) technique, with the random forest (RF) and logistic regression (LR) classifiers, and Lasso with cross validation (LassoCV). The feature importances for the models were extracted from the weights (Logistic regression) or mean impurity decrease indices (random forest).

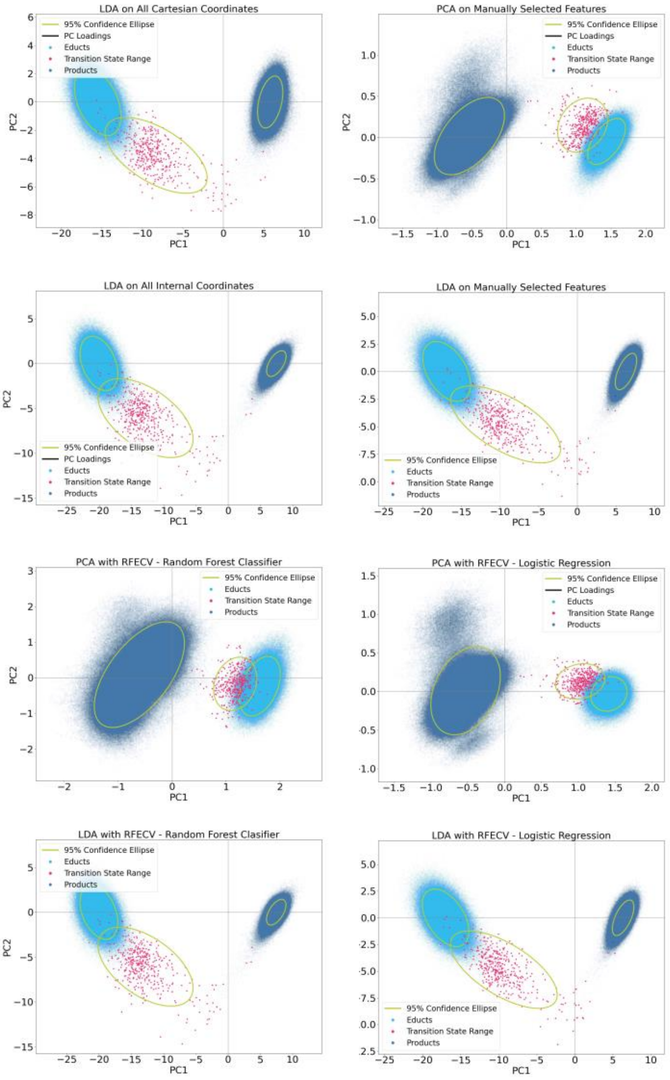


Figure S4. PCA and LDA on the features selected by the various dimensionality methods.

We performed a PCA and an LDA for each of the aforementioned dimensionality reduction and observed greatly varying performance between the approaches in the PCA approach, while the LDA was largely invariant (Figure S4).

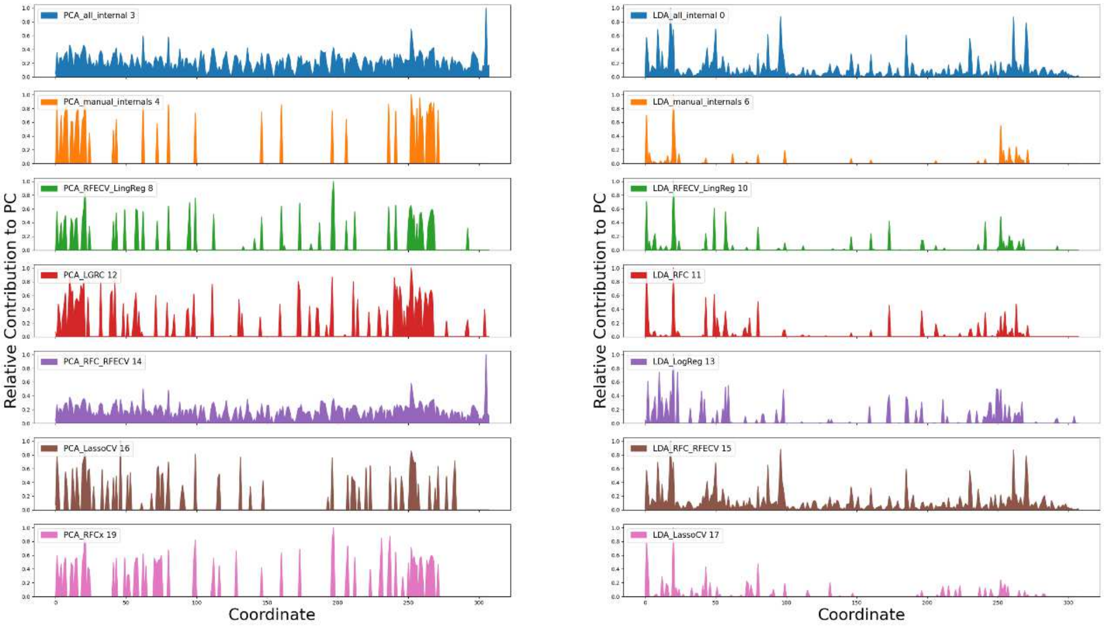


Figure S5. Feature contributions from the various methods. The contributions are derived as the mean contribution to the principal components.

When examining the feature contributions of both PCA and LDA principal components (Figure S5), the results differ slightly, depending on the classifier which was chosen.

The consensus model evaluation of the internal coordinates revealed a large overlap between the PCA and LDA features (Figure S6). We decided to choose the final consensus model based on the feature contributions from both LDA and PCA, in order to maximise the performance in dimensionality reduction, using a best of both worlds type approach.

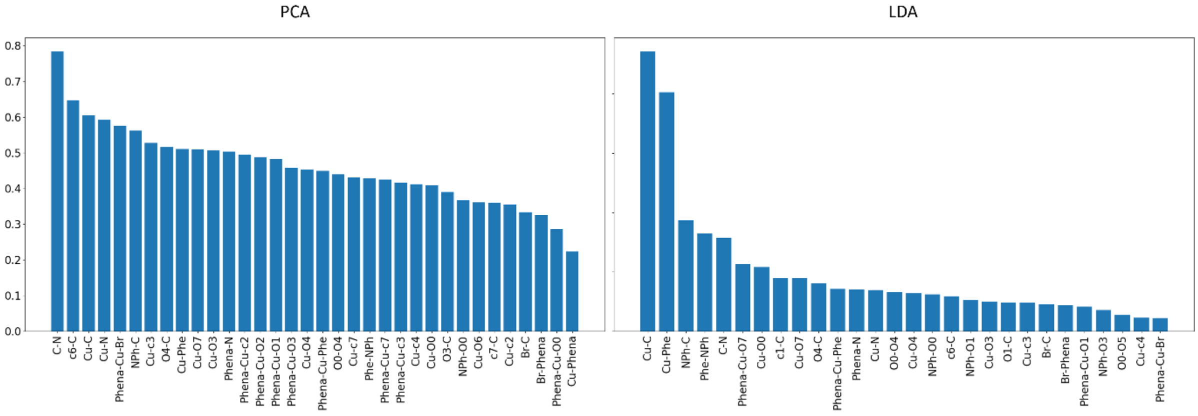


Figure S6. High importance features from LDA and PCA consensus models.

5. Decision Trees

This classifier separates the data based on precise cutoff values of the underlying features. The larger the tree structure, the more prone to overfitting it is. At the same time, the larger tree offers more detailed insight into what separates the structures. As we are not interested in performing predictions, an over-fitted tree is not of concern, but it’s structure must be pruned, by limiting the size of the tree, in order to keep it interpretable.

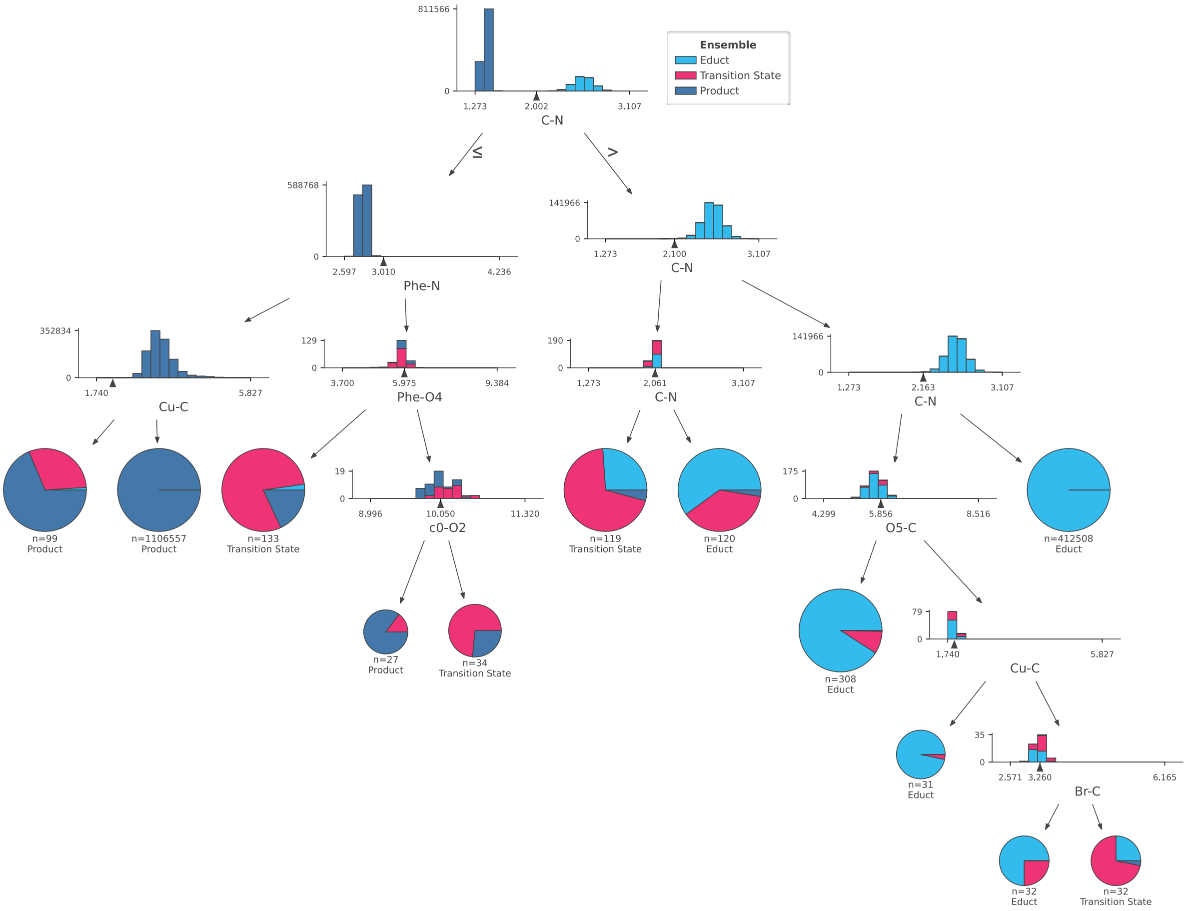


Figure S7. Decision tree trained on the unbalanced dataset.

When training the decision tree, it’s important to ensure that the classes are balanced, as otherwise the class with lower representations might be disfavoured compared to the others. This was the case for our dataset, as the transition state ensemble contains only 426 structures, compared to the over 100000 for the other ensembles (Figure S6). To combat this, we used oversampling. By repeatedly drawing from each ensemble, we were able to equalize the number of data points between all ensembles. This way, each class used in the decision tree becomes equally important.

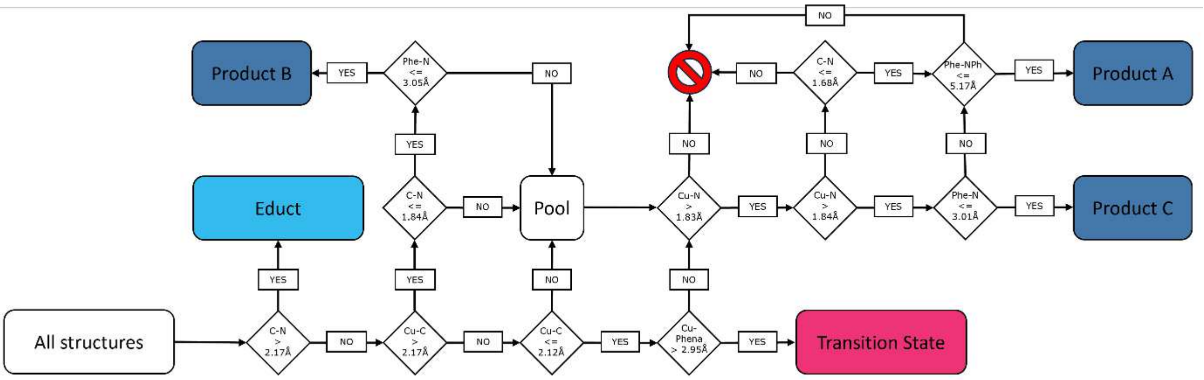


Figure S8. Decision rules depicted in diagram form.

The decision approach provides a semantic output of the decision tree, which can be transposed in a decision workflow as shown in Figure S8. The vast majority of structures can then be assigned to the appropriate ensemble, with only a few (less than 0.001%) not finding an assignment.

6. Robustness and convergence of sampling

We can evaluate the robustness of the sampling by performing dimensionality reduction on the consensus features using a reduced number of simulations and comparing it to the full dataset. Here we can see (Figure S9) that taking 10 random simulations from the dataset results in a PCA which resembles that of the full dataset, yet fails to distinguish between the product conformations.

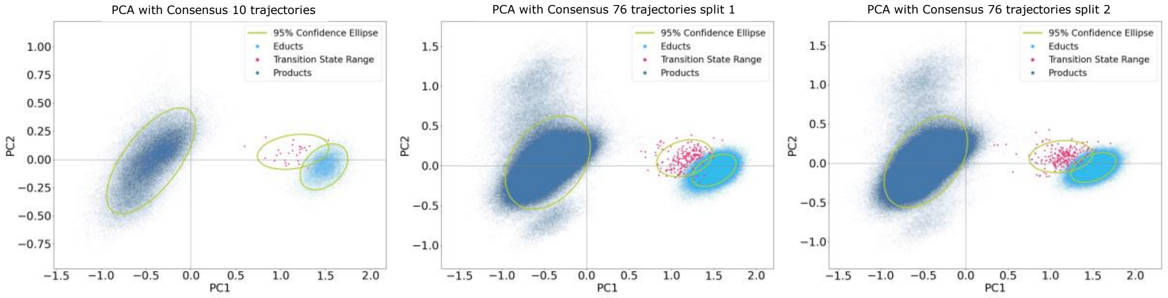


Figure S9. PCA performed on the consensus features using 10 trajectories, 76 trajectories and the remaining 76 trajectories.

When we randomly split the trajectories into two datasets with 76 runs each and perform the PCA, we see that they are very similar to each other in the separation of classes and product conformations. They strongly resemble the results of the PCA performed on the complete dataset. This indicates a convergence of the results and sufficient sampling of the reaction step.

7. Internal coordinates overview

In the histogram plots below (Figure S10), the distribution of all internal coordinates is shown, with calculated mean and standard deviation values for each ensemble.

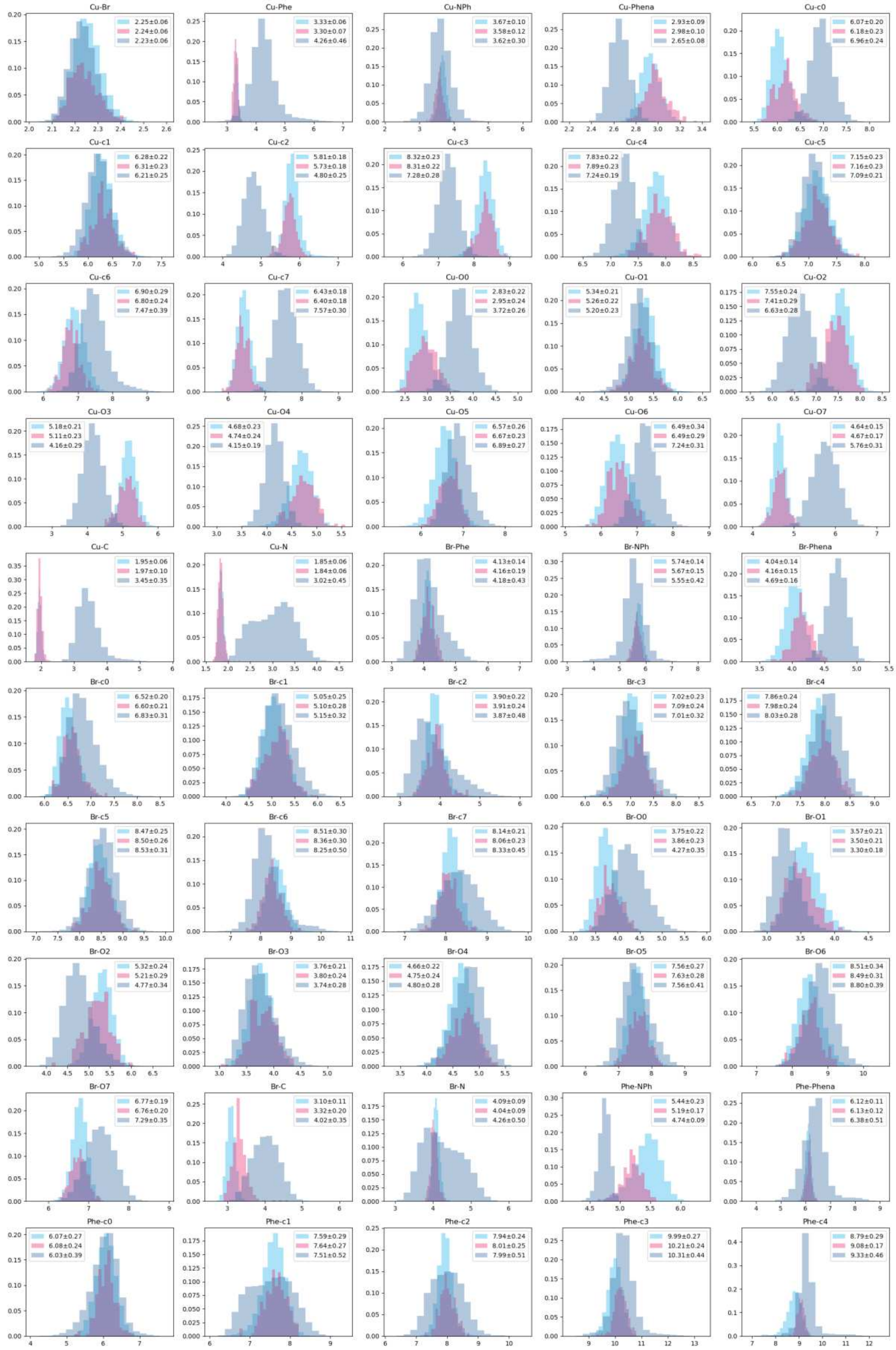


Figure S10 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

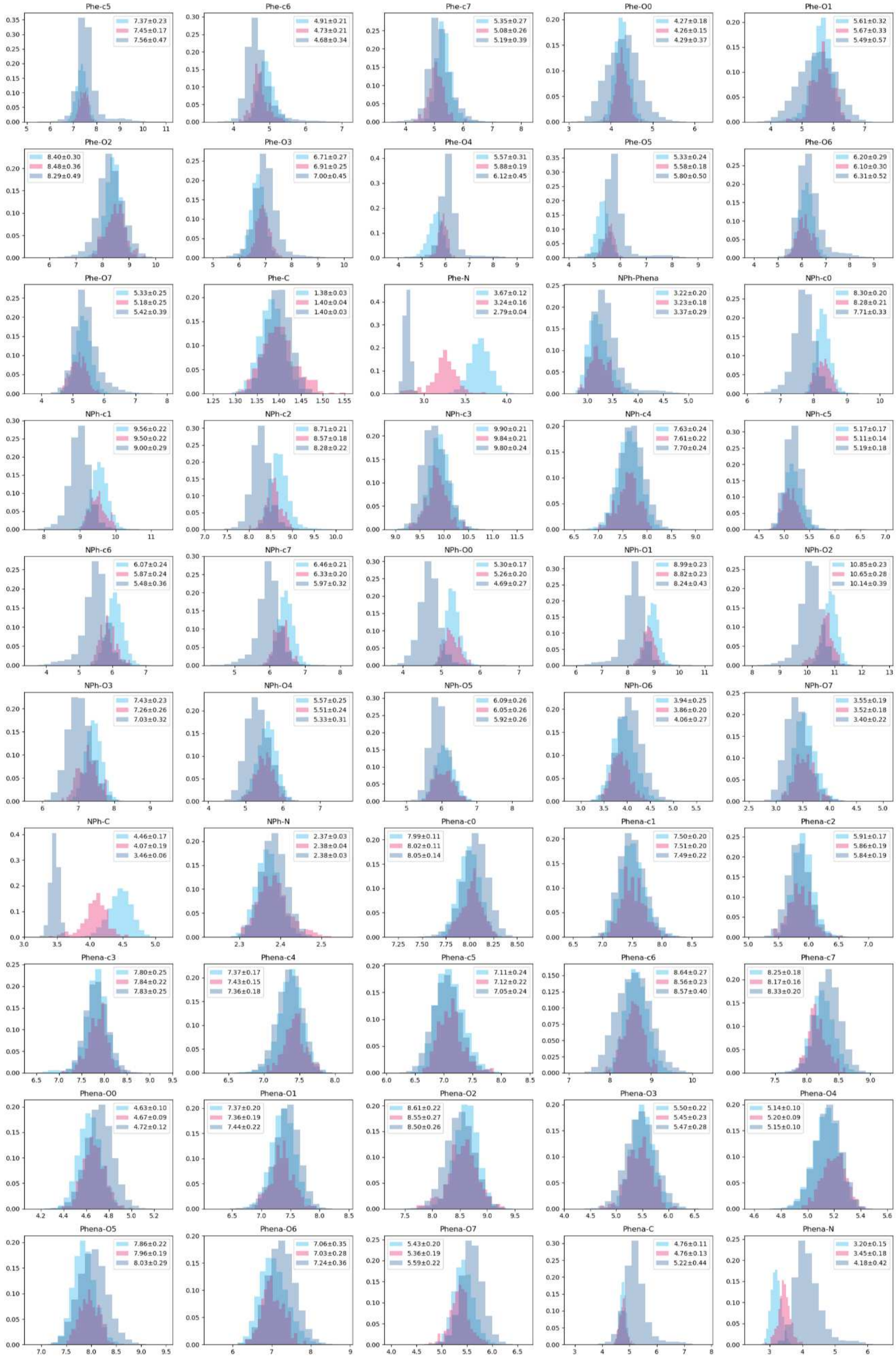


Figure S110 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

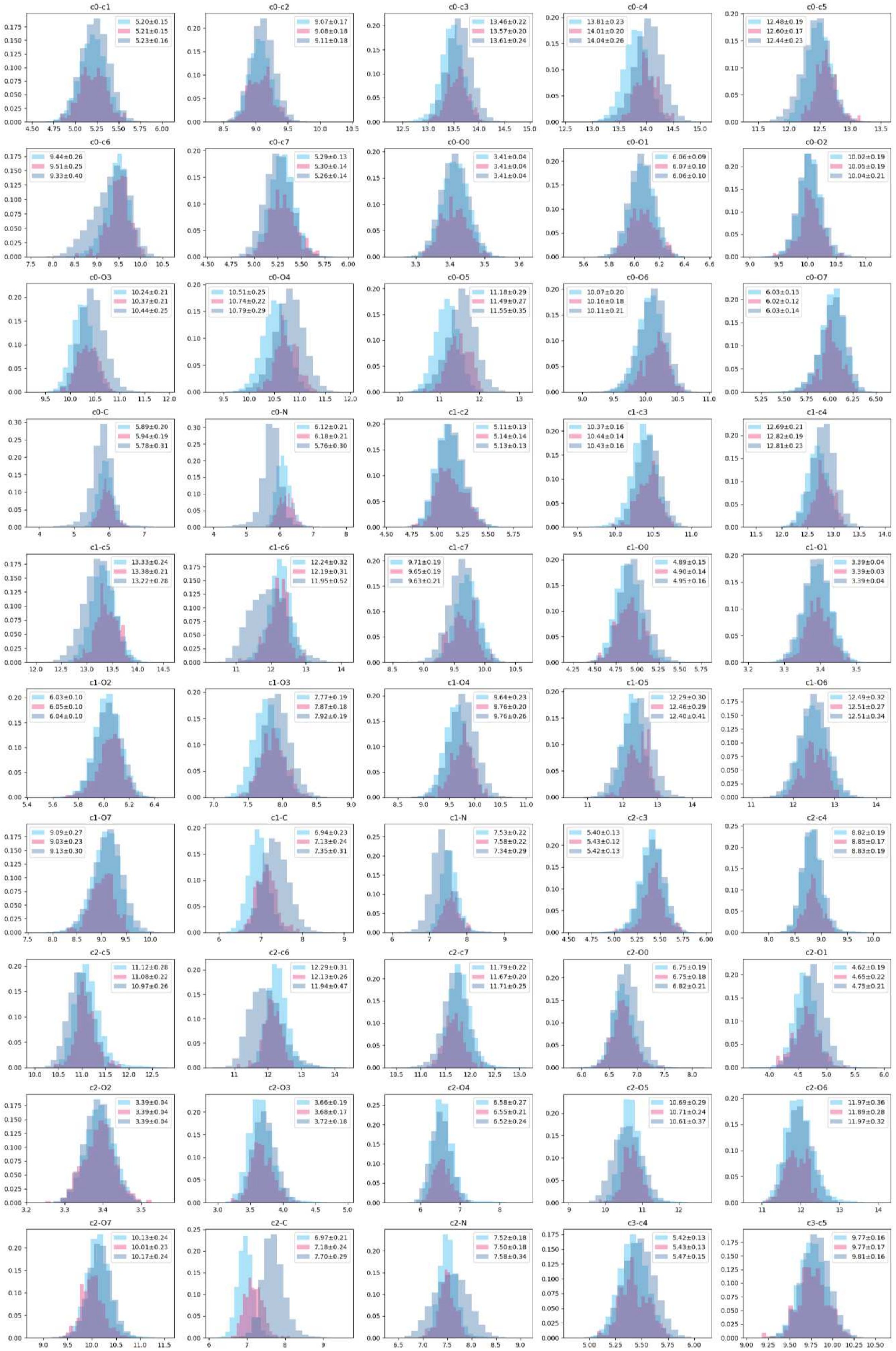


Figure S120 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

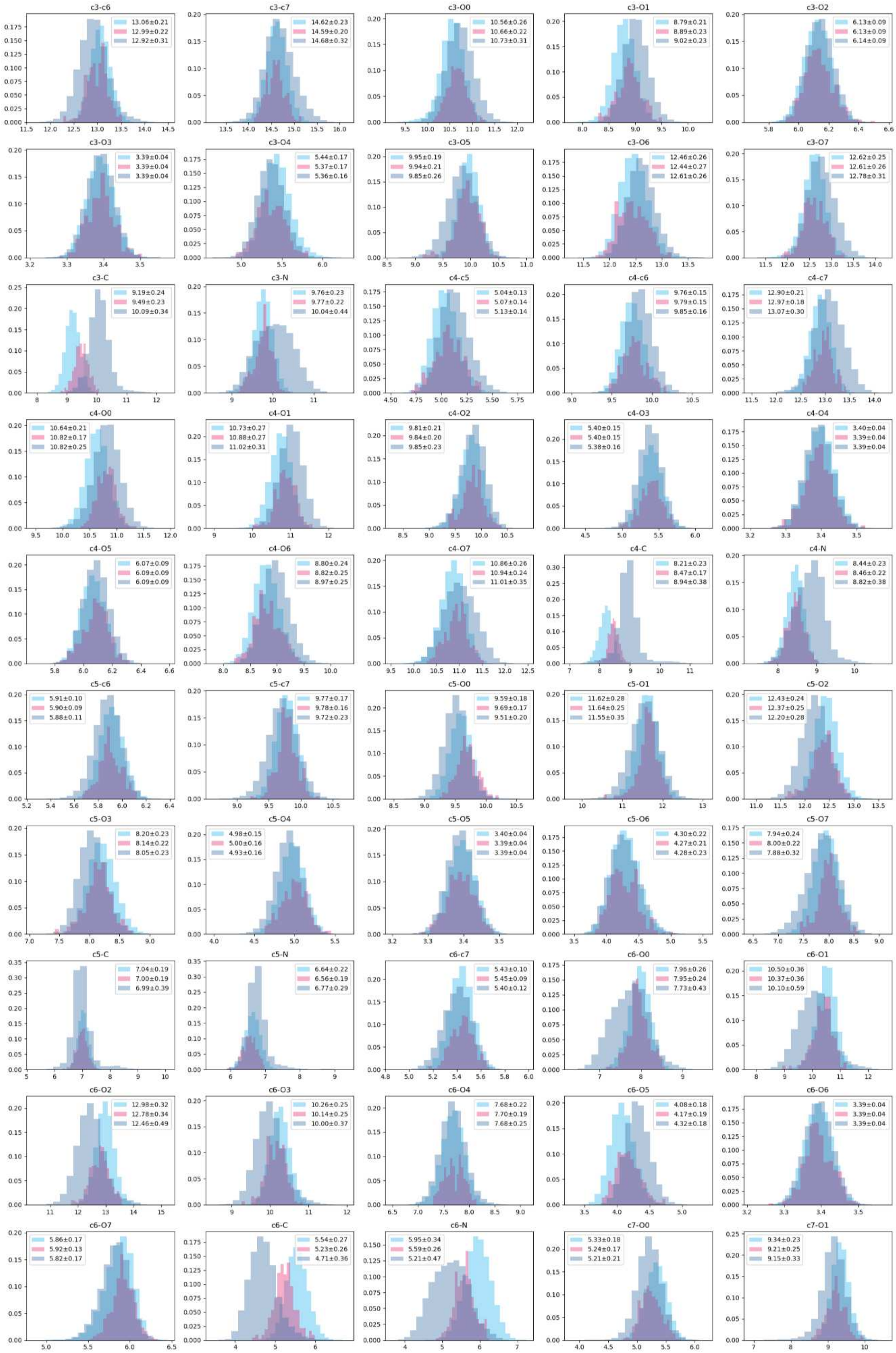


Figure S130 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

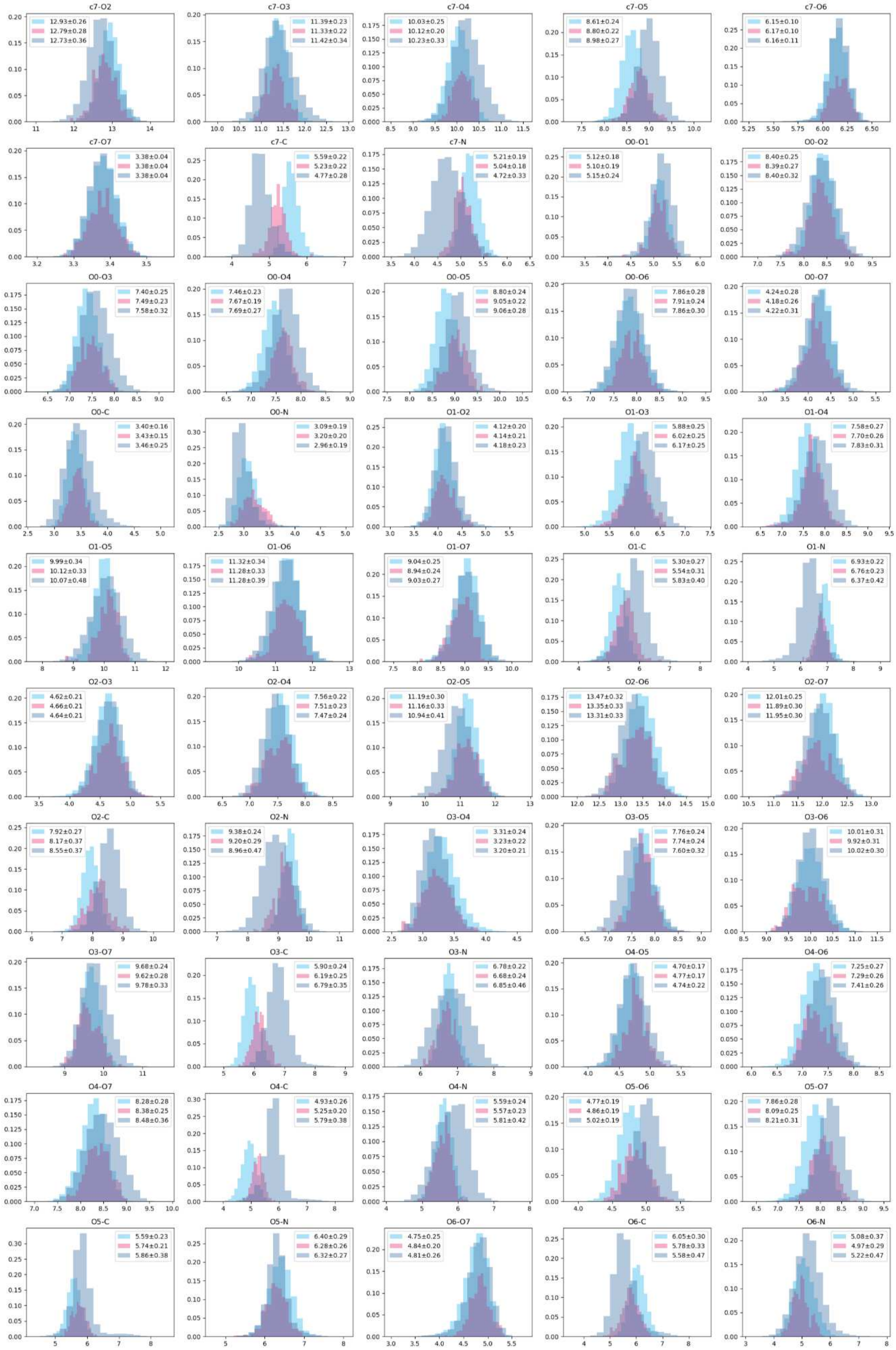


Figure S140 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

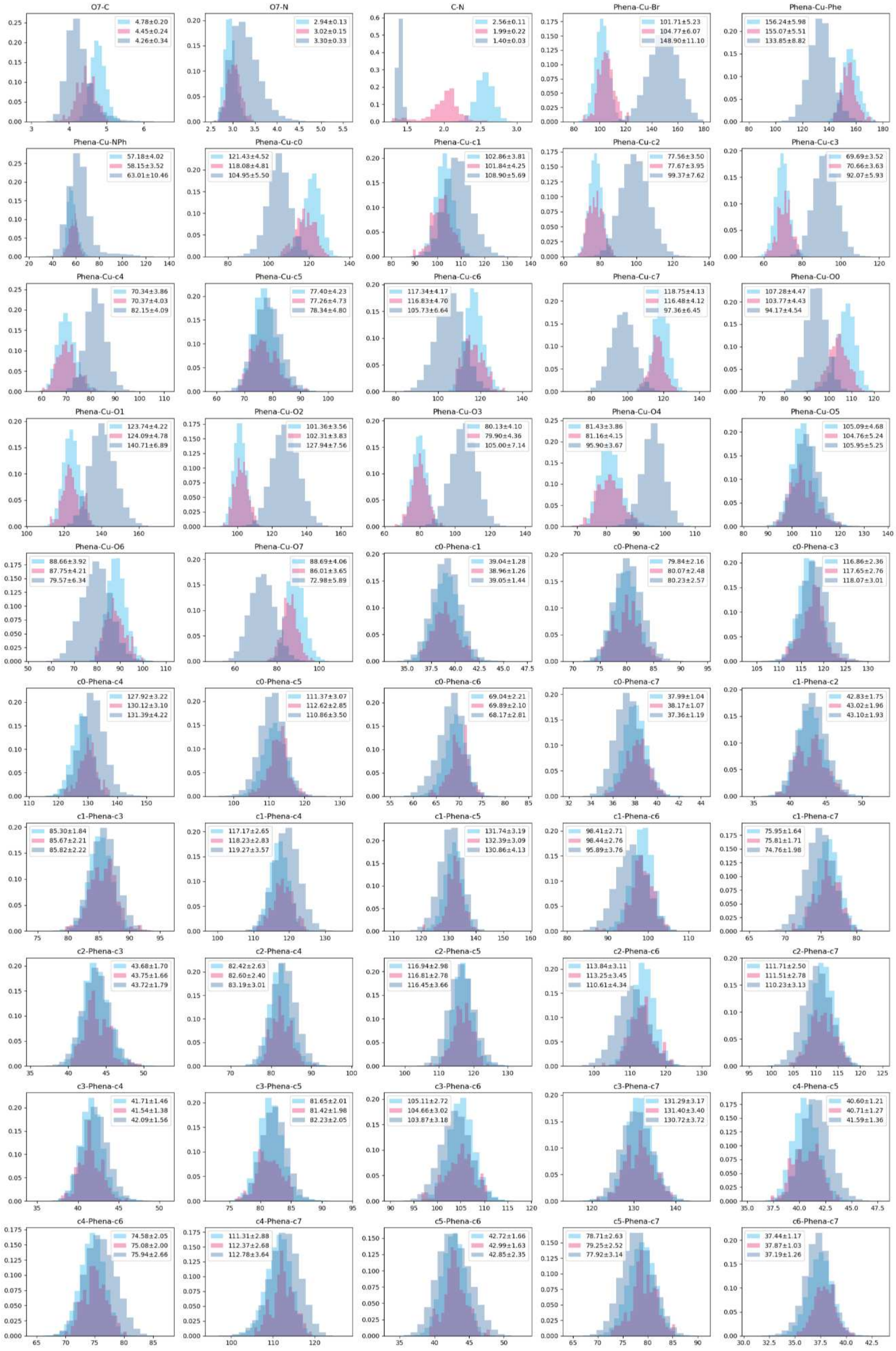


Figure S150 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation. Continues on next page.

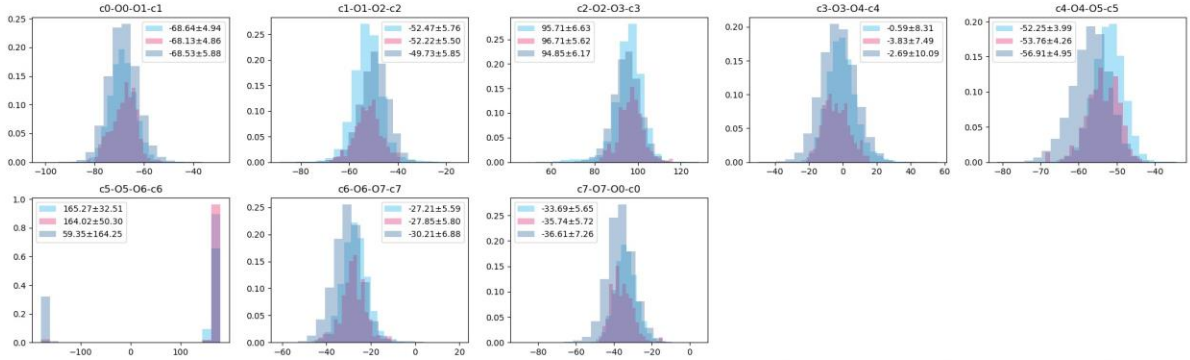


Figure S16 Histogram plot of the distribution of internal coordinates for each ensemble with mean value and standard deviation.

8. References

(1) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(2) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76* (1), 637–649. https://doi.org/10.1063/1.442716.

(3) Andersen, H. C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *Journal of Computational Physics* **1983**, *52* (1), 24–34. https://doi.org/10.1016/0021-9991(83)90014-1.

(4) Barker, J. A.; Watts, R. O. Monte Carlo Studies of the Dielectric Properties of Water-like Models. *Molecular Physics* **1973**, *26* (3), 789–792. https://doi.org/10.1080/00268977300102101.

(5) Savitzky, Abraham.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36* (8), 1627–1639. https://doi.org/10.1021/ac60214a047.