UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

General, Data-Driven Representations and Algorithms for Catalyst Design and Discovery

Permalink

https://escholarship.org/uc/item/1gx560s9

ISBN

9798189279041

Author

Smith, Andrew Laban

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

General, Data-Driven Representations and Algorithms for Catalyst Design and Discovery

By

Andrew Laban Smith

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor F. Dean Toste, Chair

Professor John F. Hartwig

Professor Gerard Marriott

Spring 2026

Abstract

General, Data-Driven Representations and Algorithms for Catalyst Design and Discovery

by

Andrew Laban Smith

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor F. Dean Toste, Chair

Catalysts lower overall reaction barriers without themselves being consumed; however, the precise properties enabling the desired activity and selectivity can be challenging to identify and exploit. In this dissertation we leverage synthesis coupled with data science-based toolkits to discover low-dimensional structure-activity relationships (SARs), linking simple catalyst properties to its performance. From the realized SARs, we can deploy machine learning algorithms to automate design, accelerating access to the most promising candidates.

Chapter 1 – we provide a perspective on the physical organic basis of catalysis and molecular representations that seek to unite catalyst *structure* to its *function*. We reflect on strategies that unite simple representations with high chemical fidelity. Together, we expect these goals to improve predictive and rational design under complex settings.

Chapter 2 – we present a data-driven approach to interrogate the activity and selectivity of a supramolecular cage-mediated cascade reduction. Data science-based inquiry identifies univariate regressions and classifications that precisely quantify the emergence of size-exclusion across the higher-order assembly within the host cavity. Moreover, simple and experimentally validated surrogates for competing, attractive non-covalent interactions are identified.

Chapter 3 – we deploy statistical inquiry to identify a single descriptor – active site-based buried volume – is uniquely relevant for the construction of models for stereoselectivity by diverse sets Brønsted acid organocatalysts. We show the representation transcends steric inference and likely accommodates a diverse array of stereoelectronic interactions responsible for stereoinduction.

Chapter 4 – we generalize the relationship between active site-based buried volume and stereoselectivity by Brønsted acid organocatalysts. We present a holistic interpretation based on *complementarity* where a chiral catalyst preferentially matches one chiral transition state. From this perspective, we construct predictive Gaussian regression models that generalize stereoselectivity beyond empirical examples onto new, disparate and unfamiliar solutions.

Chapter 5 – we conclude our extensive investigation on generalizing stereoselectivity of Brønsted acid organocatalysts by presenting an algorithm for the design of stereoselective solutions. The leverages active learning-based machine learning to efficiently navigate chemical space, affording the *inverse design* of highly stereoselective Brønsted acid organocatalysts.

Chapter 6 – we expand the reaction manifold of Brønsted acid organocatalysts to a suite of open-shell reactions, including: group transfer, alkene additions, and excited state quenching paradigms. Dithiophosphoric acids are uniquely capable of accessing these reaction modes and we conduct rigorous mechanistic inquiry to rationalize this. From these foundations, we expect our findings will enable data-driven investigations into new, stereoselective open-shell reactions.

Dedication

To perseverance.

Table of Contents

Dedication ..... **i**
Acknowledgements ..... **v**
Chapter I – Molecular representations for chemical phenomena **1**
I.1 – Preface 2
I.2 – Catalysis as a chemical phenomenon 2
I.3 – Molecular representations 9
I.4 – Outstanding challenges of molecular representations 23
I.5 – References 25
Chapter II – Models for catalytic activity and stereoselectivity within confined spaces **31**
II.1 – Preface 32
II.2 – Introduction 32
II.3 – Results 33
II.4 – Discussion and Outlook 38
II.5 – References 45
Chapter III – Simple generalizations of stereoselectivity over diverse sets of acid organocatalysts **47**
III.1 – Preface 49
III.2 – Introduction 49
III.3 – Results 50
III.4 – Discussion and Outlook 55
III.5 – References 61
Chapter IV – Abstracting complementarity for the design of stereoselective acid organocatalysts. **65**
IV.1 – Preface 67
IV.2 – Introduction 67
IV.3 – Results 68
IV.4 – Discussion and Outlook 73
IV.5 – References 78
Chapter V – Algorithmic design of stereoselective acid organocatalysts **81**
V.1 – Preface 83
V.2 – Introduction 83
V.3 – Results 84
V.4 – Discussion and Outlook 89
V.5 – References 95
Chapter VI – Recasting acid organocatalysts for a suite of radical reactions **97**
VI.1 – Preface 99
VI.2 – Introduction 99
VI.3 – Results 100
VI.4 – Discussion and Outlook 105
VI.5 – References 112
Appendix I – General Experimental and Computational Methods **115**
I.1 – Materials and Methods 115

I.2 – General Calculation and Statistical Considerations 116
I.3 – Molecular descriptor calculation workflow 117
I.4 – Non-covalent interaction calculation workflow 122
I.5 – References 127
Appendix II – Supplementary Information for Chapter II 129
II.1 – Synthesis of Supramolecular Assemblies 129
II.2 – Standard Synthetic Procedures 131
II.3 – Kinetics Experiments 132
II.4 – Experiment Replication and Expanded Controls 135
II.5 – Example NMR Spectra for Catalytic Reactivity and Kinetics 138
II.6 – Extended Discussion on Figure II-2C 140
II.7 – Extended Discussion on Figure II-3B 143
II.8 – Extended Discussion on Figure II-4A 146
II.9 – Extended Discussion on Figure II-4B 149
II.10 – Extended Discussion on Figure II-4C 152
II.11 – Extended Discussion on Figure II-5A 155
II.12 – Extended Discussion on Figure II-5B 158
II.13 – Extended Discussion on Figure II-5C 160
II.14 – Extended Discussion on Figure II-5D 166
II.15 – Substrate and Product Standards Synthesis 172
II.16 – Characterization Data of Novel Compounds 174
II.17 – NMR Spectra 177
II.18 – XYZ Coordinates 196
II.19 – References 231
Appendix III – Supplementary Information for Chapter III 235
III.1 – Specific Comments on Computational Methods 235
III.2 – Detailed Statistical Analyses 250
III.3 – Regression Performance over Structurally Diverse Datasets 254
III.4 – References 288
Appendix IV – Supplementary Information for Chapter IV 293
IV.1 – Specific Comments on Materials and Methods 293
IV.2 – Detailed Statistical Analyses 298
IV.3 – Experimental Benchmarks and Datasets 326
IV.4 – Empirical Bias Powers Discovery of Stereoselective Solutions 340
IV.5 – Synthetic Protocols 353
IV.6 – References 398
IV.7 – NMR Spectra 403
Appendix V – Supplementary Information for Chapter V 453
V.1 – Specific Comments on Materials and Methods 453
V.2 – Extended Discussion on Fig. V-2 455
V.3 – Extended Discussion on Fig. V-3 459
V.4 – Extended Discussion on Fig. V-4 468
V.5 – Extended Discussion on Fig. V-5 476

V.6 – Synthetic Methods 482
V.7 – References 499
V.8 – NMR Spectra 501
Appendix VI – Supplementary Information for Chapter VI 519
VI.1 – Specific Comments on Materials and Methods 519
VI.2 – Reaction optimization 520
VI.3 – Decarboxylation 525
VI.4 – Representative reaction NMR data 539
VI.5 – Time course data 543
VI.6 – Mechanistic studies 545
VI.7 – Synthetic Protocols 571
VI.8 – Products 582
VI.9 – Computational Data 584
VI.10 – NMR Spectra 590
VI.11 – References 600

Acknowledgements

I would like to thank my advisor, Dean, for his guidance and support throughout my program. I would also like to thank the members of my committee – John, Richmond and Tom – as well as the staff across the department, especially Jessie, Hasan, and Kathy, for the time and care they have devoted to supporting my progress. I am grateful to my closest colleague, Anna, and my closest friends at Berkeley, Gabe, Fadi and Andrew, for their encouragement at each step along this journey. Finally, I am grateful for my mom – Lynn, my dad – Richard, and my sister – Emily, for their unconditional and unwavering commitment to my wellbeing. I look forward to the opportunity to return each of these favors in kind.

Chapter I – **Molecular representations for chemical phenomena**

I.1 – Preface

Synthetic chemistry seeks to understand and deploy the fundamental laws governing the interactions between atoms and molecules. Although centuries of human knowledge inform expectations of chemical reactivity, abstracting observations into simplified frameworks remains a persistent challenge to realize the underlying basis of these phenomena. For *structure to inform function*, chemists must deploy a *chemical representation* – a learned manifestation of structure that reflects chemically meaningful properties. With the advent of data chemistry, such *representations* have evolved beyond hand-drawn structures towards progressively more complex informatics. These data-rich frameworks subsequently seek to model similarly complex phenomena, often at the expense of interpretability. Herein, we reflect on the foundation of *structure-activity* relationships – *molecular representations*. We first survey fundamentals characterizing the general, complex phenomenon of catalysis. After, we track increasingly high-dimensional *molecular representations* that aim to enhance our understanding and decision-making for catalysis. This discussion aims to highlight strategies that exploit *emergent simplicity*, where even complex chemical phenomena are described using parsimonious representations and models. We conclude our discussion with a perspective on common failures of molecular representations, seeking to inspire readers to prioritize scalable, low-dimensional frameworks to rationalize their own observations.

I.2 – Catalysis as a chemical phenomenon

Reactions between two molecules are characterized by their *spontaneity* (thermodynamics) and *practicality* (kinetics). Accordingly, a *favorable* reaction suggests two distinct (and often unspecified) meanings: (1) that a reaction is exergonic (either enthalpically and/or entropically favorable as defined by Gibb's free energy, $\Delta G < 0$), or (2) that a reaction occurs at a practical rate (ex. proceeds at a time-scale suitable for observation, subject to analytical technique).

Towards the initial implication, decades of research have been directed to establishing mechanisms that even drive *contra*-thermodynamic processes, coupling endergonic chemical transformations with other exergonic steps (ex. photonic energy, chemical fuel). Accordingly, modern synthetic chemistry is primarily focused on the latter of the two implications – aiming to *accelerate* the desired reaction. To accomplish this, chemists will seek to identify additives that transform the mechanistic landscape uniting the starting materials and the products. By navigating new reactive intermediates, these additives can be engineered to afford crucial stabilization to the emergent transition states, lowering the overall barrier of the process while itself not being consumed.

These additives are known as *catalysts*. As the mechanism of activity is distinct for any given transformation, rational and predictive design is an outstanding objective for synthetic chemists. We first introduce fundamental theories underlying catalytic phenomena. Then, we explore chemical properties and interactions generally exploited to realize catalysis.

- ***Fundamentals of catalysis***

All catalytic phenomena navigate interactions between a catalyst and substrate across (1) an initial association, (2) a transition state, and (3) the subsequent release of product. Successful catalysts *accelerate* some desired reaction. This requires that the catalyst reduces the overall barrier *en route* to the product. Such a requirement has profound (and sometimes counterintuitive) consequences on what constitutes *catalysis*. As *written in stone* by Eric Anslyn and Denis Dougherty (alongside countless chemists):[1]

Catalysis requires binding to the transition state stronger than the ground state

For the *overall* barrier of a reaction to be reduced by a catalyst, the largest transition state barrier must be reduced by a larger magnitude relative to all other interactions. At a glance, this implication seems trivial, but a variety of scenarios characterizing interactions between a catalyst and substrate highlight how simple interactions between two species is insufficient to accelerate reactivity. Typically, association is considered a favorable phenomenon, implicated by the spontaneous formation of the catalyst-substrate complex. This resultant ground-state *stabilization* necessarily slows the rate of reaction by the substrate. Unless the transition state navigated by the catalyst-substrate complex is stabilized by a larger magnitude than the ground state, the barrier to the reaction of the substrate is larger in the presence of the catalyst.

Beyond *transition state stabilization*, another coupled (and somewhat misleading) paradigm is *ground state destabilization*. The concept semantically begs the idea that the catalyst-substrate complex raises the ground state energy of the substrate, reducing the subsequent transition state barrier. In practice, this *destabilization* is a function of repulsive non-covalent interactions (discussed later), distorting the substrate energy landscape to favor configurations more akin to the transition state (ex. conformation, reaction coordinate). Taken literally, ground state destabilization violates the fundamental principle of catalysis – *a catalyst binds to the transition state stronger than the ground state*. After all, if a catalyst-substrate complex is higher in energy than the substrate *alone*, then the largest barrier to the desired reaction remains the same. The substrate and catalyst complex proceed through the *same*, unperturbed transition state!

Instead, when ground state *destabilization* is invoked as a catalytic phenomenon, it is better characterized by considering the nature of the proposed rate acceleration. *Ground state destabilization* typically describes binding of a substrate in an *activated* conformation. Here, *activated* refers to a conformation that more closely resembles the reactive transition state. By binding an *activated* conformation, a catalyst *preorganizes* the substrate, reducing entropic penalties at the transition state. Altogether, ground state destabilization rationalizes catalyst-substrate interactions that are intuitively related to the unbound substrate. Still, the paradigm requires that the transition state is stabilized to a larger extent than the ground state.

- ***General mechanisms of selective catalysis***

Between one substrate and one product, a successfully catalyzed process must reduce only a single reaction barrier. However, many methods navigate bifurcated reaction pathways where reactive intermediates can proceed to many in a set of products. In these scenarios, the formation of just one product is defined as a *selective* process. In practice, *selectivity* emerges in a variety of scenarios, including: chemo-, regio-, diastereo-, and stereoselectivity.

Although *selectivity* boasts a broad definition, *selective* catalysis relies on the same set of fundamental principles. Under thermodynamic constraints, selective outcomes exclusively relate to the relative ground state energies between the substrate and products, independent of the catalyst identity (ΔG). But, under kinetic constraints, selective outcomes are directly related to the largest transition state barrier *en route* to the set of products accessible ($\Delta G^{\ddagger}$). Accordingly, principles describing selectivity in reactions under *kinetic control* seek to rationalize relationships between competing transition states. Below, we discuss two powerful paradigms that help qualify these relative differences in transition state barriers in catalytic reactions.

◊ *Curtin-Hammett principle*

The Curtin-Hammett principle describes the selective formation of two products in a reaction operating under kinetic control. The paradigm asserts that the distribution of products corresponds to the energy difference in the product-determining transition state barriers ($\Delta\Delta G^{\ddagger}$). By inspection, the principle is a pairwise implementation of the familiar dogma of catalysis. It critically implies that the relative stability of reactive *intermediates* is irrelevant to the overall product distribution. This counterintuitive expectation is frequently invoked in stereoselective catalysis, where a variety of diastereomeric catalyst-substrate conformations can be sampled prior to the enantiodetermining step. Importantly, this principle only holds true given all other states (ex. conformations, catalyst-intermediate complexes) are under rapid, prior equilibrium.

One notable demonstration of the Curtin-Hammett principle in stereoselective catalysis is by Halpern in his report on the mechanism of the Noyori hydrogenation, an asymmetric process mediated by chiral bisphosphine-[Rh] complexes.[2] Halpern showed the major diastereomer of the catalyst-substrate complex (the only one detected in solution) surprisingly corresponds to the opposite enantiomer of the product formed. Instead, the minor diastereomer of the catalyst-substrate complex leads to the observed stereochemical outcome. By substantially reducing the temperature of the reaction, Halpern was able to *freeze* out conformational exchange, excluding any rapid prior equilibria prior to the enantiodetermining reduction. Subsequent treatment with hydrogen (H_2) led to complete consumption of the minor diastereomer, while the major diastereomer remained unreacted. Coupled with kinetics-derived insight into absolute reaction barriers from each diastereomeric complex, Halpern clearly demonstrated Noyori's stereoselective reduction operates under Curtin-Hammett control.

◊ *Hammond postulate*

The Hammond postulate qualifies the relationship between transition state barriers between analogous reactions. Specifically, the postulate characterizes a transition state as a fleeting, high-energy intermediate. The structure of this state is expected to resembles the structure of the nearest ground state. In exergonic reactions ($\Delta G < 0$), the Hammond postulate characterizes the transition state as *starting material-like*. In endergonic reaction ($\Delta G > 0$), the transition state is instead characterized as *product-like*. This principle is often deployed when comparing analogous reactions. For example, given two endergonic reactions where one forms a much higher energy product, the Hammond postulate would qualify this pathway as *slower* because the transition state is expected to be *similarly* higher in energy.

The Hammond postulate can be invoked to rationalize electronic effects across selectivity-determining reactions. Notably, Jacobsen and coworkers rigorously assigned an emergent electronic relationship in the stereoselective [Mn]-salen catalyzed epoxidation of alkenes.[3,4] In their report, the authors observed electron-poor salens (ex. bearing electron-withdrawing groups like -NO_2) furnished the epoxidation with drastically reduced enantioselectivity compared to electron-rich salens (ex. bearing electron-donating groups like -OMe, 22% v 96% *e.e.*). Robust, linear relationships correlating the electron density (represented by Hammett value, σ_p – see below) to the observed stereoselectivity provide quantitative support for this claim. Along with other mechanistic insight provided by isotopic labeling and Eyring analyses, the authors conclude that the initial oxo-addition by the olefin is enantiodetermining. With electron-donating salens, this step is rendered more exergonic by virtue of electronic stabilization of the electrophilic intermediate. Deploying the Hammond postulate, the authors suggest that transition states of the electron-rich [Mn]-salen complexes are thus more *product-like*. Because chirality is set at this intermediate, it is reasoned that the enantiodetermining transition state can better discriminate between developing stereocenters.

- ***Intermolecular interactions of binding***

The fundamental basis of catalysis, regardless of the complexity of the reaction pathway, relies on *binding to the transition state stronger than the ground state*. In this statement, *binding* refers to *non-covalent interactions* – a broad category of attractive and repulsive intermolecular forces. Terms like *steric* or *electronic* are invoked to characterize the interaction underlying the observed force. Yet, in all cases, electron density (organized over molecular orbitals) is the theoretical foundation. Below, we survey various nuances of non-covalent interactions and highlight cases where each interaction enables stereoselective catalysis.

◊ *Ion-ion, ion-dipole, and dipole-dipole interactions*

One of the most familiar intermolecular interactions to most scientists (let alone chemists) is the *ionic* interaction. The interaction (E_s) is defined by *Coulomb's law* (**Eqn. I-1**), where q refers to

charge (in Coulombs), ϵ refers to permittivity of the medium (ex. solvent), and r refers to distance between the two charges:

$$E_s = \frac{1}{4\pi\epsilon} * \frac{q_1 q_2}{r^2} \qquad (\mathbf{Eqn. I-1})$$

Unlike most intermolecular forces, ionic interactions are exceptionally *long-range*. Accordingly, strategies to modulate ionic interactions aim to influence attributes besides the magnitude of the interacting charges. For instance, the relationship relies on ϵ where less polar solvents generally have lower permittivity. Accordingly, hydrophobic and/or desolvated catalyst-substrate complexes can display unexpectedly powerful ionic interactions.

By extension, coordinating and/or polar solvents can attenuate ionic interactions by virtue of larger permittivity (ϵ) and by charge separation (lit. solvation, increases r). These new interactions compete with ion pairing. They are classified as *ion-dipole* interactions (**Eqn. I-2**), where μ refers to dipole moment (in Debye), α refers to the alignment between the dipole (directional) and interacting charge, q, with a distance dependence increased by a factor of r^{-1}:

$$E \propto \frac{q\mu}{r^3}\cos(\alpha) \qquad (\mathbf{Eqn. I-2})$$

Dipole moments, when aligned, are also capable of productive interactions with other dipole moments (**Eqn. I-3**), now decaying by yet another factor of r^{-1}:

$$E \propto \frac{\mu_1 \mu_2}{r^4} \qquad (\mathbf{Eqn. I-3})$$

Many catalytic systems rely on these charge and dipole-based interactions. For example, *hydrogen bonding* is one case of *dipole-dipole* interactions where the strong directionality of a highly polarized H-atom (bound to an electronegative atom) helps to enforce organized, dipolar interactions with other basic functional groups (ex. lone pairs of acceptors like N and O atoms).

Still, directionality is not a prerequisite to enforce selectivity across spatially sensitive reactions like those encountered in stereoselectivity. In one prominent example, Phipps and coworkers reported a generalizable approach for stereoselective transition metal catalysis based on the presence of chiral cations.[5] By simply appending a proximal anionic sulfonate group on an achiral ligand backbone, the authors demonstrate the addition of chiral cations (ex. naturally derived quinine, <$5/mmol) helps promote the remote, stereoselective borylation of arenes by simple iridium precatalysts. While ionic interactions are the basis of association between the achiral active catalyst and the chiral additive, mechanistic interrogation reveals a critical H-bonding network is required for successful enantiodiscrimination. In substrates without the requisite, chelating H-bond donor (*N*-Me amide rather than *N*-H amide), the authors show the desymmetrization only proceeds with attenuated selectivity (8% v. 73% *e.e.*), even in the presence of the coordinating chiral cation.

Besides Phipps' demonstration, countless reports of stereoselective catalysis leveraging ion/dipole non-covalent interactions are known (ex. Jacobsen's H-bonding small molecule catalysts,[6] Terada's Brønsted acids,[7] etc.).

◊ *Ion-π and π-stacking interactions*

Beyond point charges and dipole moments, molecules can also display *quadrupolar moments*. These electronic features are hallmarks of aromatic rings. A holistic outcome of constituent *p*-atomic orbitals, aromatic systems are *quadrupolar* with a distinct 2D-symmetry characterized by charge density localized on either face of the aromatic system. Accordingly, the relative electrostatic field over an aromatic system expects facial anionic character (π-bonds) and peripheral cationic character (polarized $C(sp^2)$–H bonds are electrophilic at H). The unique electrostatic configuration of these *quadrupolar* moments endows π-rich systems with attractive interactions that are collectively characterized *π-stacking*.

Traditionally, electron-rich π-systems favorably interact with cations, furnishing *cation- π* interactions.[8] Alkali metal ions like sodium can afford interactions up to 30 kcal/mol, drastically exceeding other *strong* non-covalent interactions like H-bonding interactions (typically, up to 5 kcal/mol). Along with other higher-order bonding interactions (ex. δ-bonding), these intermolecular forces can help to justify the stability of the familiar sandwich complexes where multivalent transition metals chelate to two π-systems (ex. ferrocene, cobaltocene, uranocene). In catalysis, electron-rich π-systems found in amino acids (ex. phenol from tyrosine, indole from tryptophan) are natively concentrated in certain domains of enzymes responsible for cation recognition, aptly characterized as the *cation-π box* or *aromatic box*.[9–11]

On the other hand, electron-poor π-systems (ex. perfluorobenzene) display an *inversion* in quadrupolar moment relative to electron-rich π-systems. The electrostatic configuration is opposite – faces of the π-system are now electropositive, while peripheral atoms are electronegative. This inversion endows these π-systems with complementary properties, such as anionic association. Notably, the Matile group has explored the utility of electron-poor naphthalene-diimides (NDI) as *anion-π* acceptors for broad applications in catalysis.[12] While their research program highlights the utility of these model systems to accelerate reactivity across otherwise destabilized anionic intermediates, the strategy has not observed widespread adoption for catalyst design beyond demonstration.

Finally, *π-π interactions* can also emerge between quadrupolar moment. Subject to continuous debate, their basis, strength, and structure observe frequent corrections and adjustments. Currently, the prevailing theory characterizing π-π interactions surprisingly assigns π-face interactions as repulsive. Instead, through-space π-σ dipoles are aligned to maximize the observed interactions.[13–15] Beyond electrostatics, associating π-systems also benefit entropically from desolvation. These enthalpic and entropic considerations collectively contribute to a complex profile underlying any π-π interaction. In catalysis, these features are best realized through rigorous mechanistic analysis, including substituent-effects (ex. Hammett features, see below) and crystal structures, clarifying the nuanced energy contributors in each case.[16] However, despite the persistence of methodological flaws in theory, computational calculations are routinely presented as primary

evidence for these multifaceted interactions (ex. energy decomposition analysis with SAPT,[17] Multifwn[18]).

◊ *Pauli repulsion and London dispersion forces*

While permanent charges (and charge distributions) are the basis of many non-covalent interactions, even nonpolar species, lacking any charge, dipole, quadrupole, etc. associate through fleeting interactions known as *London dispersion forces* (or *van der Waals forces*). In any given molecular orbital, electrons can be randomly and fleetingly distributed in a configuration that resembles a dipole moment – defined as an *instantaneous dipole.* As molecular size increases, the chance of *instantaneous dipoles* occurring also increases.

Critically, the *attractive* nature of dispersion forces appears contrary to familiar *repulsive* interactions between filled orbitals – called *Pauli repulsion*. But, these two competing ideas are *not* mutually exclusive. Instead, they are simultaneously captured with the Lennard-Jones potential (E_{LJ}, **Eqn. I-4**), where *attractive* interactions scale against r^{-6} while *repulsive* interactions scale against r^{-12} for some experimentally fitted parameters a and b.

$$E_{LJ} = E_{disp} + E_{Pauli} = \frac{-a}{r^6} + \frac{b}{r^{12}} \qquad (\mathbf{Eqn.\, I-4})$$

In general, electron clouds (ex. larger atomic and molecular orbitals) featuring a high propensity for this fleeting electronic reorganization are defined as *polarizable*. As dispersion interactions are intrinsic to all atoms and molecules, *polarizable* groups are a hallmark of many catalysts. The collective energy accumulated from many dispersive interactions can furnish surprising outcomes, completely overriding outcomes expected by traditional steric repulsion-based models.[19] To accurately quantify such interactions, dispersion-corrected theoretical methods are required (ex. Grimme-dispersion correction, GD3).[20]

For example, organocatalyzed transfer hydrogenations are well-known to rely on dispersion to enforce stereoselectivity.[21] Recently, List and coworkers have reported the rational design of a dispersion-enhanced Brønsted acid organocatalyst for transfer hydrogenation of unsaturated carbonyls.[22] The authors concede their previous solutions for the method rely on a bespoke Hantzch ester derivative. Their previous reagent-based strategy enforced stereoselective dispersion interactions with the chiral organocatalyst at the stereodetermining transition state. By empirically exploring new chiral acid/amine combinations with greater quantities of polarizable groups (ex. isopropyl groups, saturated spirocycles), the authors were able to realize a stereoselective solution from commercially available Hantzch ester (up to 99% *e.e.*, 22 ex.). Transition state calculations between catalytic systems with and without the key polarizable groups further demonstrate the pronounced effects of the engineered, attractive dispersion forces.

- ***Seeking rational design of catalysts***

The above discussion highlights the transferable foundations governing any catalytic phenomenon. The deceptively simple principles reflect complex (and often specific) non-covalent interactions across catalysts, substrates, and reactive intermediates. Yet, despite such simplifying generalizations, catalyst solutions realized are almost never rationally inferred. Instead, the mechanisms assigned to their success are only prescribed after the fact. Of course, a strong understanding of successful strategies can help to navigate catalyst chemical space, but the challenge remains – what are the true relationships between these interactions and catalytic activity? Accordingly, we assert:

Rational design of catalysts remains an outstanding objective in synthetic chemistry

In the next section, we introduce and explore *molecular representations*. We will show how chemically meaningful properties can be inferred from progressively complex data structures. Ultimately, an underlying objective will be to illustrate how these *representations* improve our capacity to deploy the well-formed, but abstract foundations of catalytic paradigms.

I.3 – Molecular representations

Structure informs function is an organizing principle to transform observations of natural phenomena into transferable principles. In catalysis, chemists will seek to characterize which structural features are responsible for the observed activity in some reaction. However, molecular structures are not intrinsically meaningful. Of course, molecular connectivity necessarily captures all the necessary information to justify a specific molecule's function, but connectivity itself is not exactly *intuitive*. Instead, chemists will abstract *connectivity*-based patterns into *properties*. By linking structure to *property* and thus *property* to function, chemists can realize profound generalizations of reactivity over disparate sets of molecules, regardless of their underlying connectivity (ex. isolobal analogy for organometallic bonding[23]).

To realize the rational design of catalysts, chemists seek to deploy *properties* that capture chemically meaningful criteria related to the observed function. Previously, we surveyed specific mechanisms through which catalysts can bind transition states. Yet, such catalyst-substrate interactions are altogether too complex to predict *a priori*. Instead, data-driven approaches help to track the emergent patterns between catalyst *property* and *function*. These properties are quantified as *molecular representations* (or *descriptors*). While initial strategies by chemists leverage empirical data as the foundation for such properties, the development of computational chemistry (and the expansion of computational capabilities in the age of information) has permitted the application of substantially more complex representations called *embeddings*. Over the next few sections, we introduce increasingly more abstract representations, beginning with properties defined from small, empirical observations, and evolving to unsupervised *embeddings* that leverage large data to natively abstract molecular connectivity to its function. Throughout, we will

provide commentary qualifying the utility of the *simplified* intuition drawn from such molecular representations.

- ***Empirical representations***

Initial molecular representations sought to capture relationships between transition states as dictated by principles like the Hammond postulate. Quantitative relationships known as *linear free energy relationships* (LFER) correlate thermodynamic criteria of reactions to the transition state, furnishing powerful insight into relationships across reactions. In each case, a thermodynamic relationship is defined by some ratio with a baseline reference. Then, these empirical ratios are correlated to new experimental observations, furnishing analogous mechanistic insights relating back to the original empirical foundations. We review two of the most popular LFERs for electronic and steric-controlled phenomena – *Hammett* and *Taft* values.

◊ *Hammett values*

Chemists often assign observed reactivity differences to latent electronic effects. To elucidate such interactions, they deploy *Hammett values* – a descriptor defined by a substituent's relative modulation to the acidity of benzoic acid. Here, Hammett constructed ratios (σ_x) between the degree of ionization between substituted (K_X) and unsubstituted benzoic acids (K_H, **Eqn. I-5**).[24]

$$\sigma_x = \log\left(\frac{K_X}{K_H}\right) \qquad (\mathbf{Eqn.\,I-5})$$

From this scale, Hammett values harness the intuitive relationship to thermodynamic acidity. Familiar electron-donating groups like *p*-NMe_2 are represented with $\sigma_x < 0$ as benzoic acid ionization is disfavored, while familiar electron-withdrawing groups like *p*-CF_3 are represented with $\sigma_x > 0$ as benzoic acid ionization is more favorable. Accordingly, researchers seek to relate these empirically quantified *electronic effects* to draw insights into the dependence of reactive intermediates on charge stabilization. Linear relationships are drawn for some alternative thermodynamic (*K*) or kinetic (k) observation (**Eqn. I-6**), now scaling the relationship to Hammett's σ_x by a parameter, ρ.

$$\log\left(\frac{\mathrm{k}_X}{\mathrm{k}_H}\right) = \rho\sigma_x \qquad (\mathbf{Eqn.\,I-6})$$

The ρ parameter is the subject for most conclusions drawn from Hammett plots. When $\rho > 0$, the analysis indicates a positive correlation between electron-withdrawing character ($\sigma_x > 0$) and reaction rate, suggesting the buildup of *more* negative charge in the transition state. On the other hand, when $\rho < 0$, the analysis reflects a positive correlation between electron-donating character ($\sigma_x < 0$) and reaction rate, instead suggesting the buildup of *more* positive charge in the transition state. In either case, the transition state need not have positive or negative charge, but rather only the *buildup*. For example, consider a cationic substrate that loses positive charge in a rate-

determining transition state. The corresponding Hammett analysis would show $\rho > 0$, reflecting the comparatively *more negatively charged* transition state, albeit still cationic.

While further advances have since been reported to closely characterize charge buildup across various model reactions, the most popular Hammett-variants aim to extract the relative contributions of resonance and induction. For instance, 4-methoxybenzoic acid bears a strong, resonance-donating methoxy group ($\sigma = -0.28$), while 3-methoxybenzoic acid observes a weak, inductively-withdrawing methoxy group ($\sigma = +0.13$). These contrasting resonance donation and inductive withdrawing effects have lead to the tabulation of both σ_p and σ_m values, corresponding to *para*- and *meta*-substituted benzoic acids respectively. Large tables of common Hammett values ($\sigma_p,\ \sigma_m, \sigma^+, \sigma^-$, etc.) have since been reported for quick-and-easy lookup.

◊ *Taft values*

Besides electronic effects, chemists routinely invoke steric interactions (ex. *repulsion*) when justifying differences in reaction outcomes. To quantify such a principle, Taft correlated the hydrolysis of methyl esters to the steric bulk of the neighboring substituent.[25] In this scheme, the addition of water forming the characteristic tetrahedral intermediate is assigned as the rate-determining step. As this step is formally charge-neutral (both the starting material and the tetrahedral intermediate are cationic in mechanism of acid-catalyzed hydrolysis), Taft assigned any change in rate to the steric hindrance of the neighboring group. Thus, Taft values (E_s) are empirical surrogates representing steric bulk, defined by the ratio of the rate of acid-catalyzed hydrolysis by substituted methyl esters (k_s) relative to methyl acetate (Me, **Eqn. I-7**):

$$E_s = \log\left(\frac{\mathrm{k}_s}{\mathrm{k}_{\mathrm{Me}}}\right) \qquad (\mathbf{Eqn.\,I-7})$$

Unlike Hammett values, Taft values are not nearly as commonplace. Moreover, their formulation is somewhat flawed as it remains partially contaminated by electronic contributions to the rate of hydrolysis. Instead, chemists routinely quantify steric bulk directly using computational approaches, bypassing analogies derived from empirical results. Still, for those eager to deploy the experimentally derived representations, Taft analyses follow the same form as Hammett's, now using the *steric sensitivity* factor δ instead (**Eqn. I-8**).

$$\log\left(\frac{\mathrm{k}_s}{\mathrm{k}_{\mathrm{Me}}}\right) = \delta E_s \qquad (\mathbf{Eqn.\,I-8})$$

- ***Mechanism-informed, in silico representations***

Although the empirically derived Hammett and Taft values promise a holistic and *grounded* view of the stereoelectronic effects furnished by substituents, they are primarily confounded by violated assumptions made by less *simple* substituents. For instance, the ionization of *o*-substituted benzoic acids can be influenced by through-space interactions (ex. H-bonding, Pauli repulsion) to the adjacent substituent, contaminating the impact by the substituent's resonance or induction effects.

Besides these empirical flaws, many systems studied by chemists are not comprised exclusively of *model* substituents. Familiar, but comparatively complex substituents (ex. 2,3,4,5,6-pentafluorophenyl) escape the scope of these simple, empirical representations. Accordingly, researchers turn to other strategies to impute stereoelectronic properties of broader, bespoke chemical space. These strategies involve an increasing reliance on computational resources. However, they are often evoke clear, mechanistic rationale, motivating their perception as well-formed chemical representations. We highlight three cases below, each illustrating a nuanced picture of substituent (and broadly, substrate) steric and electronic effects.

- ***Calculated NMR shifts***

Nuclear magnetic resonance (NMR) is a routine experimental technique used by organic chemists to elucidate chemical structures. For NMR-active nuclei (ex. ^{1}H, ^{13}C, ^{19}F, ^{31}P, etc.), NMR spectra display collections of signals corresponding to each chemically (and magnetically) inequivalent nucleus. Each signal has two components: (1) *intensity* and (2) *chemical shift* (δ). The latter is mathematically related to *magnetic shielding effects* (**Eqn. I-9**), where σ_d and σ_p are diamagnetic and paramagnetic contributions respectively:

$$\delta \propto \sigma = \sigma_d + \sigma_p \qquad (\mathbf{Eqn. I-9})$$

For example, ^{1}H nuclei possess relatively constant paramagnetic contributions (related to asymmetric electron distributions around the atom) to *shielding*. Instead, diamagnetic contributions (related to neighboring group effects) are the primary contributors. For nuclei in *analogous* states (ex. same functional group), this tensor can then operate as a quantitative surrogate for electronic effects. As NMR shielding tensors can be readily calculated (ex. nmr=giao keyword with Gaussian16), *chemical shifts* can serve as mechanistically adjacent representations for holistic, electronic effects.

In one powerful demonstration, Rovis and coworkers leveraged calculated ^{13}C NMR shifts to classify chemoselective outcomes in a Curtin-Hammett controlled Giese-addition.[26] In their report, the authors leveraged a rapid, reversible HAT agent (Ph_3SiSH) to sample many different nucleophilic, *C*-centered radicals. It was quickly realized that the observed chemoselectivity in Giese addition product did not correspond to BDE of the activated C–H bonds, but rather the nucleophilicity of the radical intermediate. Such an electronic component directly correlates to NMR shifts. The authors showed that chemoselectivity closely follows trends in calculated ^{13}C shift. Even in substrates with more than five reactive sites, positions with the *largest* calculated ^{13}C shifts were always observed to react *faster*, highlighting the predictive opportunity afforded by the representation.

◊ *Vibrational frequency*

Vibrational modes correspond to molecule-scale compressions of bonds abstracted as spring-like harmonic oscillators. This analogy enables chemists to derive mathematical relationships from

Hook's law (**Eqn. I-10**), where v is frequency (correlating to energy), k is an empirical factor related to the bond, and μ is the *reduced mass* across the two bound atoms (**Eqn. I-11**).

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad (\mathbf{Eqn.\,I-10})$$

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \qquad (\mathbf{Eqn.\,I-11})$$

In practice, the energies of each molecular oscillation can be experimentally detected using infrared spectroscopy (reported in wavenumber, cm^{-1}). For molecules with similar connectivity, analogous vibrational mode can reveal electronic perturbations, by virtue of the empirical force constant, k. Although this component lacks a clear mathematical definition, chemists can interpret *any* changes to holistically capture a range of stereoelectronic effects.

Historically, vibrational modes were invoked to quantify electronic effects influencing the degree of bonding between two atoms. Tolman characterized many of substituted $Ni(CO)_3$ complexes.[27] Here, the identity of the substituent influences the degree of metal-ligand back-bonding between Ni and the bound carbonyl groups. As these groups display a distinct, characteristic wavenumber (around 2050 cm^{-1}), Tolman could empirically quantify the electron-donating character of various ligands in such organometallic complexes. Electron-donating ligands promote metal-ligand back-bonding and thus lower the observed carbonyl stretching frequency. This specific carbonyl-based wavenumber has since been defined as the *Tolman electronic parameter* (TEP).

Much later, Milo and Sigman deployed *calculated* vibrational modes to holistically capture complex stereoelectronic effects that escape other traditional LFER approaches.[28–30] With multivariate linear regression models, selected frequencies were correlated to experimental objectives, such as stereoselectivity. Since their original report, the approach has been largely superseded by other simpler representations that do not require expensive frequency calculations. Although calculated vibrational modes are not commonplace representations, their intuitive and historical utility remains noteworthy of mention. In many modern analyses, chemists will report models with representations that lack clear relationships to mechanism-informed phenomena. Consequentially, it becomes challenging to recognize *how* such a representation meaningfully enhances chemical intuition. As calculated vibrational frequencies are fundamentally *grounded* descriptors, their application as holistic representations of general stereoelectronic substituent effects remains justified.

◊ *Transition-state parameterization*

As catalysis seeks to stabilize certain transition states *en route* to products of interest, it seems trivial to expect that transition states calculations would enable rational design. However, transition state characterization is *expensive*. Searching for these proposed conformations can be computationally arduous, let alone vague. Still, once robustly validated with empirical data,

computed transition states can be worth their trouble. In some reports, representations extracted from these proposed complexes have been instrumental in furnishing valuable LFERs to complex objectives like stereoselectivity. Below, we highlight two ideas stemming from transition state parameterization.

Recently, Sigman and Doyle have shown models trained on sparse data (limited in scope relative to the chemical space of interest) can furnish transferable predictions of stereoselectivity, even to out-of-set substrates and catalysts absent from their training data.[31] Critical to their success was the parameterization of general transition states corresponding to key steps excepted across many of the reactions of interest. The parameters extracted include various steric, geometric, or electronic *scalar representations* (discussed below). Representations extracted from ground state complexes, ligands alone, or other precatalyst structures were altogether unsuitable (larger error, larger variance) relative to the transition state-based parameterization deployed.

Although transition states themselves provide ample information into key interactions responsible for the observed barriers, representations that do not require such expensive calculations are much more accessible to the average chemist. Work by Toste and Sigman sought to address this challenge by abstracting transition state parameterization into smaller, interaction-based objectives. One of the most complex non-covalent interactions realized through transition parameterization is π-π interactions. Toste and Sigman leveraged simplified symmetry considerations to draw analogies between small molecule calculations and the corresponding interactions in enantiodetermining transition states.[32] Such small molecule analogies are routinely deployed in correlations for more complex π-π interactions, often larger, intractable complexes that cannot be reasonably computed.[33,34]

- ***Scalar representations (0-D)***

Mechanism-informed representations promise clear insight into the basis of an emergent relationship between *structure* and *function*. But, the precise mechanism by which any catalyst operates is hardly accessible *a priori.* Opportunities related to mechanism-adjacent generalizations (ex. transition-state parameterization) seem to address this challenge. Still, these generalizations remain reliant on *expert-level* insight to anticipate plausible relationships.

A promising alternative instead seeks to abstract chemical *structure* into *any property* of interest. Here, the representation is not experimentally derived, but rather intuitively reflects some molecular characteristic that is known to be important for activity. For instance, while π- π interactions are often invoked to justify transition-state stabilization, an agnostic property might simply relate to the *size* of one of the π-systems. Although circumstantial in correlation, the abstract representation quantifies critical chemical events at a fraction of the cost. By leveraging such generalizations, chemists sacrifice *interpretability* to maximize *utility*. Because these representations are highly accessible and do not require otherwise complex computations, chemists routinely deploy them in data-driven predictive chemistry.

Here, we highlight the simplest of such descriptors – *scalar* representations. Their minimal dimensionality (lit. *one* value) renders them incredibly lightweight and interpretable (ex. easy to calculate, store, visualize, read, etc.). Moreover, these representations can be computed for a wide range of chemically familiar properties, including geometric, steric, and electronic features, for application in LFER-like relationships.

◊ *Geometric scalar representations*

As chemical *structure* is abstracted further from clear relationships to function, one of the clearest abstractions corresponds to chemical bonding. Features like *bond lengths* are closely related to formal attributes like *bond order*, but also functional properties like *bond dissociation energy* (BDE). Empirically, geometric features can be directly measured from X-ray crystal structure data. To accelerate access to such representations, chemists can deploy computational methods. Of note, cheap, open-source semi-empirical methods like as Grimme's semi-empirical tight binding (xTB) furnish quality structural data in fractions of second for reasonably sized molecules relevant to organic chemists.[35]

Beyond one binding pair of atoms, chemists can also characterize *molecule*-scale geometric features. These analogs to *bond length* are known as *sterimol* values. First reported by Verloop, the values are quite literally *length* (L, the longest distance along the primary axis), *width* (B_1, the shortest distance orthogonal to the primary axis), and *height* (B_5, the longest distance orthogonal to the primary axis) of a molecule.[36] To compute such dimensions, Verloop leveraged *van der Waals radii* (also, *bondi* values) to quantify the size of atoms across the molecule.[37,38]

Because molecules are not static conformations, *sterimol* values are often reported as *ensemble*-averaged representations. The most common of which include *ground state* (lowest-energy conformer), *Boltzmann averaged*, and *minimum* (minimum-valued conformer). Paton and coworkers leveraged this particular nuance to improve the utility of predictive models relating *sterimol* to experimental outcomes like stereoselectivity.[39] Static representations of geometry furnished weaker models (lower R^2) relative to conformationally-aware *sterimol* features. To assist in the complex calculations required to complete these analyses, the group has reported a collection of Python packages (wSterimol,[40] DBSTEP[41]).

◊ *Steric scalar representations*

Lightweight *steric* features promise rapid insight into the nuances of attractive dispersion and Pauli repulsion. Historically, chemists have sought to characterize the *size* of certain atoms. Features like *van der Waals radii* were designed to rationalize atomic size. Within this radius, interacting atoms are expected to experience severe *Pauli repulsion* (recall, r^{-12} scaling).

To generalize this behavior to substituents, a variant of Taft values were designed called *Charton* values.[42] These steric representations critically merge physical representations of *van der Waals* radii with more common, group-based intuition in synthetic chemistry. Owing to their

interpretability and broader domain, *Charton* values have observed substantial applications in LFER-like analyses by synthetic chemists.[43–46] Yet, these empirically adjacent descriptors have (again) been largely superseded by more general and accessible features.

Chemists also quantify steric features by computing *molecular volumes*. Historically, these volumes could be imputed from crystal structure data. Most notably, Tolman developed a steric representation for phosphine ligands known as the *Tolman cone angle.*[47,48] Empirically, Tolman quantified the space occupied by ligands in $Ni(CO)_3$ by measuring how *wide* a ligand occupies a metal's primary coordination sphere, approximated with a *cone* and its aperture *angle*.

To extend this concept to substituents besides phosphine ligands, Cavallo reported a new feature – *buried volume*.[49,50] Here, a sphere is drawn around a point of interest (typically 3.5Å radius). Then, the *percent steric occupancy* of other atoms in the sphere (approximated with *bondi* radii) is measured. The occupancy is reported as the *buried volume*, specified by radius and position.

This feature is by far the most impactful low-dimensional representation of *steric* phenomena.[50] Countless reports leveraging the feature have been made across many different fields of catalysis. Valuable abstractions corresponding to critical *functions* such as absolute relationships for *catalytic activity* have been realized, highlighting the ability of the feature to capture complex interactions beyond standard *steric hindrance*.[51,52] In this thesis, we report extensively on the advantageous utility of buried volume to generalize over stereoselective outcomes by diverse sets of Brønsted acid organocatalysts (see **Chp. 3-5**). Accordingly, further discussions of the feature's utility are reserved for later.

◊ *Electronic scalar representations*

Chemists approximate *electronic*-like features with computational methods that quantify mechanism-adjacent properties. Atom-level descriptors not only include *calculated NMR shifts* (discussed above), but also abstract ideas like *partial charge* descriptors (ex. Hirshfeld,[53] Mulliken,[54] ChelpG,[55] and more). These features evoke expectations surrounding *electronegativity* of specific atoms, an intuitive concept for many organic chemists. In this analogy, electronegative atoms are expected to concentrate electron density, furnishing *negative partial charges*. Electron-poor domains are then characterized by *positive partial charges*. These representations are derived from computed, *molecule*-level perturbations to baseline expectations of atomic charge density. While the specific theoretical basis for any given *partial charge* representation is distinct, they correlate well to familiar, empirically derived Hammett values, motivating their adoption as representations for electronic substituent effects.[56]

Molecule-level electronic descriptors seek to capture holistic effects of each substituent contribution. For instance, electrophilic functional groups like carbonyls undergo addition reactions at the *C*-atom. Theoretically, we rationalize this phenomenon with molecular orbitals, where the unfilled, π^* orbital (LUMO) possess a larger coefficient at the site undergoing nucleophilic attack. To then evaluate the propensity of analogous carbonyl groups to undergo

addition, we then anticipate *lower lying* LUMOs are more accessible in energy, and thus more reactive (ex. LUMO lowering activation). Computationally, this idea can be quantified by computing an absolute value for orbital energies. While exact values are sensitive to basis set and altogether lack any physical relevance, *relative* values are highly meaningful. Thus, following from the above consideration, LUMO values can be computed for a range of molecules and quantitatively related to the observed rates of addition. Besides LUMO, other relevant frontier orbitals can also be calculated (ex. HOMO, SOMO, etc.). The energy gaps of these computed orbitals further correspond to physically meaningful electron transitions, a property with substantial relevance to photochemistry.

- ***Vector representations (1-D)***

A natural extension to individual *scalar* representations is to concatenate collections of *stereoelectronic* properties – denoted here as *vectors*. These extended representations are deployed with only circumstantial relevance to some chemical function. Typically, the representations are the output of *descriptor calculators* where one molecule is transformed into 1,000's of potential *scalar* representations, spanning geometric, steric, and electronic properties (ex. Mordred,[57] AutoQChem,[58] AutoDFT[59]). Thus, strategies that leverage these large *vector*-based representations seek to generalize broader relationships between chemical *structure* and *function* with a variety of statistical tools.

◊ *Composite, reduced representations*

In most analytical models, the relationships between domain and range are drawn with a variety of linear and nonlinear combinations of fitted parameters and selected representations. But, when dimensionality is excessive, these combinations can *overfit* – furnishing models that simply interpolate each point between domain and range by virtue of the large number of parameters. Thus, statistical tools are deployed to maximize the information captured by each feature, while minimizing the total number of dimensions.

The most popular technique is called *principal component analysis*. Here, linear algebra techniques solve for *principal components* corresponding to linear combinations across *all* dimensions, now parameterized with the largest portion of total *variance*. Statistically, models fundamentally seek to correlate *variance* (ex. distribution, spread) between variables. By reconstructing data that maximizes the total *variance* within a domain, *principal component analysis* essentially concentrates the total information associated with datasets of interest along singular dimensions. Subsequent relationships drawn to *principal components* then use the narrower set of information-rich *vectors* rather than the diffuse, repetitive *scalar* constituents (ex. up to 90% of the original variance across the domain).

Because *principal components* are linear combinations of potentially 1,000's of features, chemists rarely deploy these tools directly in models. Instead, chemists will leverage *principal component*-reduced vectors to visualize *chemical space*. This insight helps to group molecules by *chemical*

similarity, given the chemical space is defined by chemically meaningful properties. The interpretability of this analysis relates back to the linear combinations of features comprising each principal component. The fitted coefficients that scale each constituent feature can be interpreted as the *relative impact* of the feature on the *principal component*. As many features reflect similar properties, they are strongly correlated. So, *principal components* can reduce into *property*-based components (ex. geometric, steric, electronic).[60,61] The numerical values of the principal components are not physically meaningful. But relative differences are often interpreted as abstract (and not specific) differences across chemical properties.

◊ *Categorical representations*

To an extreme, dimensionality reduction can be applied to simply *categorize* a molecule by its identity. Without any expert insight, chemists frequently deploy such categorizations – known as *one-hot encoding* (OHE). For instance, solvent is a surprisingly challenging component to represent in chemical reactions. Sometimes, solvent directly participates in reaction mechanisms (ex. specific acid/base catalysis), while at other times it furnishes complex electric fields interacting with polarized reactive intermediates. Empirical parameters such as *dielectric constant* (recall, ϵ) incompletely generalize the complex electronic interactions that drastically influence both transition state enthalpy and entropy.

To account for the impact of solvent in *structure-activity* relationships, chemists deploy OHEs. Each value is considered as an identity, asking – *is it this specific category?* Compiled as a *vector*, OHE representations are comprised of $(n-1)$ 0's and a single 1 (for n total choices). Consequentially, each OHE is orthogonal, with no information regarding relative relationships between similar solvents. While this seems like a severe shortcoming, the representation requires *no* prior calculations *nor* empirical relevance. It merely represents a discrete empirical choice.

Accordingly, this representation has observed significant utility in *active learning* campaigns that seek to train models through iterations of empirical data. These approaches natively learn chemical relevance associated with each OHE through clever statistical tricks, bypassing any need to introduce the information from the outset. Some analyses have even shown that active learning models do not advantageously converge to ideal outcomes when provided chemically meaningful descriptors, further motivating the application of the *categorical* representations.[62]

◊ *Natural language representations*

Quantitative representations of chemical properties seem intuitive owing to their relationship to empirical observations (ex. spectroscopy). But, chemists routinely communicate molecules using chemical structures. A *vector* representation of structure corresponds to *language-based* representations. Chemists can leverage complex nomenclature standardized by IUPAC to *name* molecules, or they can leverage other simplified notations (ex. molecular formula). The most useful is the *SMILES* representation (*Simplified Molecular Input Line Entry System*) which seeks to capture molecular connectivity in a lightweight syntax suitable for diverse chemoinformatic

applications.[63] Many chemistry softwares are equipped with functions that directly operate on or extract *SMILES* strings from chemical structures (ex. RDKit,[64] ChemDraw).

Since its original report, the syntax has observed many iterations and evolutions, expanding its utility for more specific functions. One of the most common chemoinformatic tasks is *substructure analysis*. To enable rapid search over *ultra*-large databases, *SMARTS* syntax was developed (*SMILES Arbitrary Target Specification*).

As many of these notations have strict *grammar* rules, *de novo generation* of *SMILES/SMARTS* strings typically produces invalid molecules, that lack a translation to a real chemical structure. To address this challenge, Aspuru-Guzik and coworkers created a new chemical syntax that always decodes into tangible representations, the *SELFIES* representation (*SELF-referencing Embedded Strings*).[65,66] By obviating the requirement to learn implicit grammar rules, the guaranteed validity of the syntax is thought to promote the training quality of datasets in machine learning tasks leveraging the representations in high-dimensional, unsupervised tasks.[67]

- ***Matrix representations (2-D and beyond)***

It is apparent that informatics strongly influences the design of high-dimensional molecular representations. Instead of intercepting chemical intuition with mechanistic insight, these abstractions strongly prioritize *accessibility* and *generality*. In previous cases, datasets under consideration might consist of the complete set of experiments completed in a manuscript ($<10^2$ molecules). As chemical space expands ($>10^6$ molecules), two organizing ideas emerge. First, higher dimensional representations are fundamentally capable of representing *more* molecules. By leveraging progressively less expert insight, the representations boast broader applicability. Second, these representations that span chemical space (ex. *SMILES*) necessarily encode the same mechanism-adjacent properties. After all, these representations reflect the same molecule. Thus, intuitive properties chemists familiarly associate are simply embedded into hidden dimensions and relationships, inaccessible to inspection but accessible with model training.

Beyond 1-D concatenations of properties and/or categories, cheminformatics seeks to characterize molecules akin to the way chemists natively interpret structure – via *networks of chemical bonds*. This approach can be rendered algorithmic and lightweight, embedding atom and molecule-level representations by their *connectivity* or *occupancy*. The resulting *matrix representations* are much higher in dimensionality, typically assuming the form $m\ x\ n$ for some m atoms and n attributes considered. Significant efforts are then deployed to curate datasets suitable for the training of machine learning models. The central objective is the trained model meaningfully infers the hidden chemical laws necessarily embedded into *chemical networks*.

◊ *Fingerprint representations*

Atom-level representations are a convenient beginning for approaching molecule-scale questions. *Fingerprint* representations are lightweight features that capture which atoms a given atom is

bound to. Because the primary coordination sphere of an atom reflects too little information, *fingerprints* are typically measured up to 4-6 atoms in depth. The most common *fingerprint embeddings* are known as the *Morgan fingerprint* (also, *extended-connectivity fingerprint-4, ECFP-4*).[68] These embeddings can be readily generated using any number of cheminformatics packages.[64] Once produced, familiar dimensionality reduction techniques can help to visualize *chemical space* defined by the *fingerprint embedding*, devoid of other expert-informed chemical properties.

Although these representations necessarily capture the emergence of familiar concepts in synthetic chemistry (ex. functional groups, substituents), through-space interactions – a hallmark of noncovalent interactions – are wholly lost. Thus, atom-level *fingerprints* are typically exchanged for higher-level, *molecule*-level embeddings that advantageously capture a complete picture of molecule structure.

◊ *Grid-based representations*

Molecule-level embeddings intuitively extend from *grid-based* representations. Intuitively, molecular connectivity can be completely defined from pairwise distances, angles, and dihedrals derived from *xyz*-based coordinates. In fact, *xyz*-coordinates are often the primary representation reported by chemists, providing a handle for computational reproducibility when generating more precise chemical insight (ex. DFT calculations).

Equipped with *xyz*-coordinates, chemists have devised other grid-based analyses to transform the information into other composite features. One notable instance is *average steric occupancy* (ASO), first reported by the Denmark group.[69,70] Here, *xyz*-embeddings are overlaid with thousands of equally distributed grid points. At each position, the algorithm observes whether the molecule (defined atomically by *bondi* radii) occupies the space at the grid point. Averaging these values across many conformers, ASO presents a high-dimensional projection of *xyz*-based ensembles. The authors first deployed this representation to characterize confinement of chiral acid organocatalysts in a stereoselectivity prediction task. As ASO is comprised of many thousands of grid points, the representation requires *neural network* architectures to learn the complex chemical information necessarily embedded into the ensemble-derived feature. The Denmark group has since deployed the representation for machine learning applications in other conformer-dependent analyses like oligopeptide catalysis.[71,72]

By inspection, it is clear *xyz*-based representations are relatively expensive. Although routinely deployed on small scale, the embeddings require a suite of computational tools to generate (ex. molecular dynamics, semi-empirical tight binding).

◊ *Graph-based representations*

Graph-based representations merge the lightweight utility of *fingerprints* with the molecule-level insight of *grid*-based descriptors. Here, molecular connectivity is captured with an $m\ x\ m$ graph

for m total atoms. Primarily, such graph-based representations are deployed in *reaction prediction* tasks. Given sufficient data, the trained models can realize powerful insight, enhancing human intuition for *retrosynthesis*[73,74] and *forward synthesis.*[75]

Although *reaction prediction* requires a review to fully capture the nuances of dataset curation, processing, and application, we highlight here one notable development by Coley and Jensen. In their report, the authors sought to train a model capable of *regioselectivity* predictions for some forward reaction (ex. EAS site-selectivity, etc.).[76] While *graph*-based representations natively convey a complete picture of connectivity, Coley enhanced each atom-level feature with additional chemically meaningful properties. As these substantially increased the computational rigor of the representation (1,000's of seconds per evaluation), the authors simultaneously trained a neural network to predict the corresponding atom-level *scalar* representations (ex. partial charges, Fukui indices, reduced evaluation to 0.1's of seconds). Together, this information-dense representation promoted the quality of the *regioselectivity prediction* task relative to models that solely learned from native connectivity. As the training set increased in size, the model architecture lacking atom-level embeddings also approached the quality of the enhanced model. This relative improvement is one example whereby simply increasing training data (and dimensions) arbitrarily improves model quality. More broadly, Coley's work suggests mechanism-adjacent insight can substantially improve learning, especially in cases where training data is sparse and expensive to curate/collect.

- ***Machine-learned representations***

Finally, we introduce *machine-learned representations* for chemistry. At this point, we have tracked the evolution of molecular descriptors from empirical foundations all the way to abstracted *graph*-based approaches. However, the remaining idea we explore relates to the constitution of the representation itself. We have asserted syntactic representations (ex. *SMILES*, *SELFIES*) necessarily convey *all* chemical information. Therefore, the most lightweight approach to *any* modeling task would be to directly learn *all* chemical properties from these natural language representations.

Over the past decade, transformers and graph neural networks (GNN) have revolutionized the field of machine learning.[77] Notably, each architecture leverages chemical inputs simply defined by *identity*, without any additional, mechanism-adjacent information. Below, we discuss two implementations of chemical language transformers and GNN. We critically address the surprising *methods* by which these models *learn* chemical properties, contrasting it to chemically intuitive approaches natively deployed by human chemists.

  - ◊ *General pretrained transformers (GPT)*

Pretrained transformers boast *broad* utility for general chemical knowledge. The expectation is that given enough data, the pretrained transformer will learn fundamental relationships within the domain relevant to all downstream tasks (ex. chemical bonding, polarity, etc.). From these footholds, users can look to append new model layers to *specialize* the pretrained transformer using

small, curated subsets for specific tasks (ex. property prediction). This approach sidesteps the need for a *universal* chemical model that accurately performs *each* task by providing a modular baseline model that is generally *alright*.

One recent, notable implementation of the transformer framework for chemistry is ChemBERTa.[78] Here, the initial report sought to learn relationships between 52 thousand tokens (ex. letters) from 77 million *SMILES* representations (supplied from the public PubChem repository). During training, the authors masked (lit. hidden) up to 15% of the input tokens to the model. Then, the model was prompted to reconstruct the complete *SMILES* string from the incomplete input. By repeating this process over many millions of *SMILES* representations, the model was expected to learn fundamental relationships to chemical *connectivity* and, by extension, chemical *property*.

Once this *pretraining* phase is completed, *fine-tuning* can be conducted with only 1,000's of datapoints for specific tasks. In their report, the authors show that by simply *increasing* the size of the *pretraining* dataset (lit. reconstruction of more, masked *SMILES*), the performance of the subsequently fine-tuned models also increased, despite not receiving *more* specific training data information (*scaling law*). This highlights how powerful *molecular connectivity* can be to construct powerful *machine-learned* representations, regardless of their perceived distance from tangible chemical intuition.

◊ *Contrastive learning*

Besides transformer architectures, graph neural networks are also useful frameworks to build powerful embeddings. We specifically highlight that application of *contrastive learning* by Coley and coworkers to learn molecular properties solely from reported substate scope (present in the proprietary CAS Content Collection™).[79] The authors transform published datasets into molecular graphs using a variety of open-source tools that leverage computer vision to transform printed structures into graph-based representations. Then, a *graph isomorphism network* (GIN) model is trained against a simple objective known as *contrastive learning – reduce* the distance between similar-performing molecules within the same substrate scope, and *increase* the distance between dissimilar molecules and any other molecules *not* present in the scope.

This seemingly naïve objective takes for granted *publication bias* where chemists only report *positive* data, skewing literature-derived datasets towards *positive* outcomes. Still, after training on nearly 7,000 reaction scopes (of aryl halide transformations) spanning over 64,000 reactions, the authors showed the learned embeddings do capture familiar chemical intuition. Subsequent *fine-tuning* of the model's embeddings for property prediction tasks (ex. buried volume, partial charge) shows their *contrastive learning* approach outperforms other *fingerprint* and *GPT*-based embeddings. Altogether, their approach to learning *molecular embeddings* couples extremely well with the existing structure of literature-derived datasets. Automated approaches merging computer vision and natural language processing are likely capable of building larger models based on their *contrastive learning* strategy, extending the utility of the embedding beyond aryl halides, to other familiar reactive intermediates.

I.4 – Outstanding challenges of molecular representations

Our extensive discussions above showcase how mechanistic foundations in catalysis can be captured with precise (or general) molecular representations. Empirically derived features (ex. Hammett, Taft) directly correspond to stereoelectronic effects on reaction outcomes, as expected by the Hammond postulate. Accordingly, generalizations of these traditional LFERs have led to the widespread adoption of surrogate representations for a range of molecular properties. Cheminformatics introduces a new idea based on new objectives of molecular representations – *if all chemical information is stored in structure, then models can learn the associated features*. Higher dimensional representations (ex. *fingerprints, matrices*, *embeddings*) provide readily accessible inputs for broad chemical space. Subsequent model training on tasks with little relevance to reactivity still furnishes powerful representations because the models instead learn underlying, hidden patterns intrinsically stored within *connectivity.*

Still, countless molecular representations are designed, tested, and reported annually. No single representation has yet been realized to fully link chemical *structure* to *function*. Moreover, while high-dimensional embeddings empower models to learn from unstructured chemical information, they are completely intractable to humans, destroying their capacity to analytically enhance *human* chemical intuition. To conclude this chapter, we briefly provide a perspective on two outstanding challenges associated with molecular representations – their *accessibility* and *utility*.

- ***Accessibility of the representation***

The principle of parsimony (*Occam's razor*) suggests that the *simplest answer* is often the *best*. Given a representation that reasonably captures the relevant chemical properties of interest, it seems trivial to suggest that the most useful feature is then the *simplest*. At face value, simple representations evoke expectations about low dimensionality or minimal complexity. Instead, we assert the simplest representation should one that captures chemistry as we natively interpret it – through *molecular connectivity*. Fingerprints and molecular graphs are incredibly lightweight representations that can serve as robust inputs for models. The challenge is that these representations require substantial training to link the encoded *structure* to *function*.

At the end of our above discussion, we describe how *abstract* the learning objectives are for *machine-learned embeddings*. Most chemists would have expected the models to be trained on a large corpus of highly curated computational or empirical datasets. Instead, these models are instead trained on reconstructing arbitrary patterns in *molecular connectivity*. By simply increasing the volume of this *pretraining* data, the utility of the learned embeddings can be drastically improved in downstream fine-tuned tasks for more relevant objectives.

We expect familiarity with these approaches will improve as chemists become more literate in machine learning-based toolkits. Accordingly, we classify these high-dimensional, abstracted embeddings as *highly accessible*. Their valuable information is a single *pretrained* step from the simplest molecular representations – *connectivity*-based *SMILES* or molecular graphs. Their utility

then requires ML literacy to execute supervised fine-tuning (note, Coley's smallest training split utilized only 15 ex.!). We predict this literacy will improve over the next few years, and we encourage readers to recognize LLM-based agents (ex. ChatGPT, Gemini) can readily generate custom scripts to accomplish such tasks, obviating the new for code literacy to execute.

- ***Utility of the representation***

Occam's razor also expects a plausible *answer*. Many chemical representations are not intrinsically related to chemical function. Traditionally, lack of correlations between empirical descriptors like Hammett values and observations suggest lack of *mechanistic relevance*. However, complex molecular representations (ex. *machine learned embeddings*) are far from providing mechanistic insight. Accordingly, simple molecular representations must also be fundamentally *useful* for a given task.

First, we present *domain of applicability*. Representations can themselves be considered functions of *structure*. As a result, some structures may exist beyond the *domain* of the function. Accordingly, these *molecular representations* cannot capture these structural features. These challenges are especially destructive in modern structure-activity relationships. Often, chemists vaguely prioritize *analytical interpretability*, and then deploy molecular representations derived from simplified structures tangentially related to chemicals under analysis. Accordingly, these surrogate abstractions lack transferability, let alone *utility*. As chemistry seeks to learn from previous observations, patterns drawn using features that cannot be extended to other molecules lack relevance beyond their singular application.

Second, we reflect on the objective of *mechanistically informative* representations. Synthetic chemists are exceptional problem solvers. Centuries of progress have been completed without the need for *machine-learned* embeddings or cheminformatics, furnishing valuable rules and theories for chemical reactivity. These insights are principally the result of hypotheses supported by transferrable *chemical mechanisms*. As such, representations that clearly relate to *mechanism* remain the most popular (ex. Hammett). Even *machine-learned embeddings* are fine-tuned on tasks like Hammett value prediction, showcasing the persistent interest chemists have in these established descriptors. Merging *mechanism* with more accessible (and applicable) representations is an outstanding objective of the field.

—

In the remainder of this thesis, we discuss our contributions in identifying transferable molecular representations in catalysis. We leverage both direct and coincidental evidence to motivate mechanistic interpretations, seeking to power chemical intuition with quantitative insight.

I.5 – References

(1) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(2) Halpern, J. Mechanism and Stereoselectivity of Asymmetric Hydrogenation. *Science* **1982**, *217* (4558), 401–407.

(3) Jacobsen, E. N.; Zhang, W.; Guler, M. L. Electronic Tuning of Asymmetric Catalysts. *J. Am. Chem. Soc.* **1991**, *113* (17), 6703–6704.

(4) Palucki, M.; Finney, N. S.; Pospisil, P. J.; Güler, M. L.; Ishida, T.; Jacobsen, E. N. The Mechanistic Basis for Electronic Effects on Enantioselectivity in the (Salen)Mn(III)-Catalyzed Epoxidation Reaction. *J. Am. Chem. Soc.* **1998**, *120* (5), 948–954.

(5) Genov, G. R.; Douthwaite, J. L.; Lahdenperä, A. S. K.; Gibson, D. C.; Phipps, R. J. Enantioselective Remote C–H Activation Directed by a Chiral Cation. *Science* **2020**, *367* (6483), 1246–1251.

(6) Doyle, A. G.; Jacobsen, E. N. Small-Molecule H-Bond Donors in Asymmetric Catalysis. *Chem. Rev.* **2007**, *107* (12), 5713–5743.

(7) Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M. Complete Field Guide to Asymmetric BINOL-Phosphate Derived Brønsted Acid and Metal Catalysis: History and Classification by Mode of Activation; Brønsted Acidity, Hydrogen Bonding, Ion Pairing, and Metal Phosphates. *Chem. Rev.* **2014**, *114* (18), 9047–9153.

(8) Wheeler, S. E.; Houk, K. N. Substituent Effects in Cation/π Interactions and Electrostatic Potentials above the Centers of Substituted Benzenes Are Due Primarily to Through-Space Effects of the Substituents. *J. Am. Chem. Soc.* **2009**, *131* (9), 3126–3127.

(9) Dougherty, D. A. The Cation−π Interaction. *Acc. Chem. Res.* **2013**, *46* (4), 885–893.

(10) Dougherty, D. A. The Cation−π Interaction in Chemistry and Biology. *Chem. Rev.* **2025**, *125* (5), 2793–2808.

(11) Cheng, J.; Goldstein, R.; Gershenson, A.; Stec, B.; Roberts, M. F. The Cation-π Box Is a Specific Phosphatidylcholine Membrane Targeting Motif. *J Biol Chem* **2013**, *288* (21), 14863–14873.

(12) Zhao, Y.; Cotelle, Y.; Liu, L.; López-Andarias, J.; Bornhof, A.-B.; Akamatsu, M.; Sakai, N.; Matile, S. The Emergence of Anion−π Catalysis. *Acc. Chem. Res.* **2018**, *51* (9), 2255–2263.

(13) Xiao, Q.; Levine, M. S.; Iverson, B. L. Rethinking the Terms "π-Stacking" and "π–π Stacking" Again: A Proposal to Clarify the Language of Aromatic Interactions. *J. Am. Chem. Soc.* **2026**, *148* (15), 15331–15340.

(14) Wheeler, S. E. Local Nature of Substituent Effects in Stacking Interactions. *J. Am. Chem. Soc.* **2011**, *133* (26), 10262–10274.

(15) Wheeler, S. E.; Houk, K. N. Through-Space Effects of Substituents Dominate Molecular Electrostatic Potentials of Substituted Arenes. *J. Chem. Theory Comput.* **2009**, *5* (9), 2301–2312.

(16) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543* (7647), 637–646.

(17) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van Der Waals Complexes. *Chem. Rev.* **1994**, *94* (7), 1887–1930.

(18) Lu, T. A Comprehensive Electron Wavefunction Analysis Toolbox for Chemists, Multiwfn | The Journal of Chemical Physics | AIP Publishing. *J. Chem. Phys.* **2024**, *161*, 082503.

(19) Eschmann, C.; Song, L.; Schreiner, P. R. London Dispersion Interactions Rather than Steric Hindrance Determine the Enantioselectivity of the Corey–Bakshi–Shibata Reduction. *Angewandte Chemie International Edition* **2021**, *60* (9), 4823–4832.
(20) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116* (9), 5105–5154.
(21) Domanski, M. H. J.; Fuhrmann, M.; Schreiner, P. R. London Dispersion as a Design Element in Molecular Catalysis. *J. Am. Chem. Soc.* **2025**, *147* (33), 29611–29623.
(22) Leinung, W.; Mitschke, B.; Leutzsch, M.; Wakchaure, V. N.; Maji, R.; List, B. From Reagent to Catalyst: Dispersion-Driven Design of a General Asymmetric Transfer Hydrogenation Catalyst. *J. Am. Chem. Soc.* **2025**.
(23) Hoffmann, R. Building Bridges Between Inorganic and Organic Chemistry (Nobel Lecture). *Angewandte Chemie International Edition in English* **1982**, *21* (10), 711–724.
(24) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, *17* (1), 125–136.
(25) Taft, R. W. Jr. Linear Steric Energy Relationships. *J. Am. Chem. Soc.* **1953**, *75* (18), 4538–4539.
(26) Shen, Y.; Funez-Ardoiz, I.; Schoenebeck, F.; Rovis, T. Site-Selective α-C–H Functionalization of Trialkylamines via Reversible Hydrogen Atom Transfer Catalysis. *J. Am. Chem. Soc.* **2021**, *143* (45), 18952–18959.
(27) Tolman, C. A. Electron Donor-Acceptor Properties of Phosphorus Ligands. Substituent Additivity. *J. Am. Chem. Soc.* **1970**, *92* (10), 2953–2956.
(28) Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* **2014**, *507* (7491), 210–214.
(29) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, *347* (6223), 737–743.
(30) Santiago, C. B.; Milo, A.; Sigman, M. S. Developing a Modern Approach To Account for Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **2016**, *138* (40), 13424–13430.
(31) Gallarati, S.; Bucci, E. M.; Doyle, A. G.; Sigman, M. S. Transferable Enantioselectivity Models from Sparse Data. *Nature* **2026**, *651* (8106), 637–646.
(32) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
(33) Tsai, C.-C.; Sandford, C.; Wu, T.; Chen, B.; Sigman, M. S.; Toste, F. D. Enantioselective Intramolecular Allylic Substitution via Synergistic Palladium/Chiral Phosphoric Acid Catalysis: Insight into Stereoinduction through Statistical Modeling.
(34) Craescu, C. V.; David, C. D.; Heafner, E. D.; Raymond, K. N.; Bergman, R. G.; Toste, F. D. Attractive Noncovalent Interactions versus Steric Confinement in Asymmetric Supramolecular Catalysis. *J. Am. Chem. Soc.* **2026**, *148* (1), 1306–1315.
(35) 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.
(36) Verloop, A.; Hoogenstraaten, W.; Tipker, J. Chapter 4 - Development and Application of New Steric Substituent Parameters in Drug Design. In *Drug Design*; Ariëns, E. J., Ed.;

Medicinal Chemistry: A Series of Monographs; Academic Press: Amsterdam, 1976; Vol. 11, pp 165–207.
(37) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441–451.
(38) Mantina, M.; Chamberlin, A. C.; Valero, R.; Cramer, C. J.; Truhlar, D. G. Consistent van Der Waals Radii for the Whole Main Group. *J. Phys. Chem. A* **2009**, *113* (19), 5806–5812.
(39) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9* (3), 2313–2323.
(40) Patonlab/wSterimol, 2026.
(41) Patonlab/DBSTEP, 2026.
(42) Charton, M. Steric Effects. I. Esterification and Acid-Catalyzed Hydrolysis of Esters. *J. Am. Chem. Soc.* **1975**, *97* (6), 1552–1556.
(43) Gustafson, J. L.; Sigman, M. S.; Miller, S. J. Linear Free-Energy Relationship Analysis of a Catalytic Desymmetrization Reaction of a Diarylmethane-Bis(Phenol). *Org. Lett.* **2010**, *12* (12), 2794–2797.
(44) Harper, K. C.; Bess, E. N.; Sigman, M. S. Multidimensional Steric Parameters in the Analysis of Asymmetric Catalytic Reactions. *Nature Chem* **2012**, *4* (5), 366–374.
(45) Harper, K. C.; Sigman, M. S. Three-Dimensional Correlation of Steric and Electronic Free Energy Relationships Guides Asymmetric Propargylation. *Science* **2011**, *333* (6051), 1875–1878.
(46) Förster, H.; Vögtle, F. Steric Interactions in Organic Chemistry: Spatial Requirements of Substituents. *Angewandte Chemie International Edition in English* **1977**, *16* (7), 429–441.
(47) Hartwig, J. F. *Organotransition Metal Chemistry: From Bonding to Catalysis*, 1st ed.; University Science Books, 2009.
(48) Tolman, C. A. Steric Effects of Phosphorus Ligands in Organometallic Chemistry and Homogeneous Catalysis. *Chem. Rev.* **1977**, *77* (3), 313–348.
(49) Hillier, A. C.; Sommer, W. J.; Yong, B. S.; Petersen, J. L.; Cavallo, L.; Nolan, S. P. A Combined Experimental and Theoretical Study Examining the Binding of N-Heterocyclic Carbenes (NHC) to the Cp*RuCl (Cp* = H5-C5Me5) Moiety: Insight into Stereoelectronic Differences between Unsaturated and Saturated NHC Ligands. *Organometallics* **2003**, *22* (21), 4322–4326.
(50) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11* (10), 872–879.
(51) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308.
(52) Escayola, S.; Bahri-Laleh, N.; Poater, A. % V Bur Index and Steric Maps: From Predictive Catalysis to Machine Learning. *Chemical Society Reviews* **2024**, *53* (2), 853–882.
(53) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoret. Chim. Acta* **1977**, *44* (2), 129–138.
(54) Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23* (10), 1833–1840.
(55) Breneman, C.; Wiberg, K. B. Determining Atom-centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comp. Chem.* **1990**, *11* (3), 361–373.

(56) Luchini, G.; Paton, R. S. Bottom-Up Atomistic Descriptions of Top-Down Macroscopic Measurements: Computational Benchmarks for Hammett Electronic Parameters. *ACS Phys. Chem Au* **2024**, *4* (3), 259–267.
(57) Mordred-Descriptor/Mordred, 2026.
(58) Doyle-Lab-Ucla/Auto-Qchem, 2026.
(59) shenvilab. Shenvilab/Autodft, 2026.
(60) Dotson, J. J.; Van Dijk, L.; Timmerman, J. C.; Grosslight, S.; Walroth, R. C.; Gosselin, F.; Püntener, K.; Mack, K. A.; Sigman, M. S. Data-Driven Multi-Objective Optimization Tactics for Catalytic Asymmetric Reactions Using Bisphosphine Ligands. *J. Am. Chem. Soc.* **2023**, *145* (1), 110–121.
(61) Van Dijk, L.; Haas, B. C.; Lim, N.-K.; Clagg, K.; Dotson, J. J.; Treacy, S. M.; Piechowicz, K. A.; Roytman, V. A.; Zhang, H.; Toste, F. D.; Miller, S. J.; Gosselin, F.; Sigman, M. S. Data Science-Enabled Palladium-Catalyzed Enantioselective Aryl-Carbonylation of Sulfonimidamides. *J. Am. Chem. Soc.* **2023**, *145* (38), 20959–20967.
(62) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96.
(63) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36.
(64) *RDKit*.
(65) Aspuru-Guzik-Group/Selfies, 2026.
(66) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-Referencing Embedded Strings (SELFIES): A 100% Robust Molecular String Representation. *Mach. Learn.: Sci. Technol.* **2020**, *1* (4), 045024.
(67) Krenn, M.; Ai, Q.; Barthel, S.; Carson, N.; Frei, A.; Frey, N. C.; Friederich, P.; Gaudin, T.; Gayle, A. A.; Jablonka, K. M.; Lameiro, R. F.; Lemm, D.; Lo, A.; Moosavi, S. M.; Nápoles-Duarte, J. M.; Nigam, A.; Pollice, R.; Rajan, K.; Schatzschneider, U.; Schwaller, P.; Skreta, M.; Smit, B.; Strieth-Kalthoff, F.; Sun, C.; Tom, G.; Falk von Rudorff, G.; Wang, A.; White, A. D.; Young, A.; Yu, R.; Aspuru-Guzik, A. SELFIES and the Future of Molecular String Representations. *Patterns* **2022**, *3* (10), 100588.
(68) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.
(69) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631.
(70) Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020**, *120* (3), 1620–1689.
(71) Schnitzer, T.; Schnurr, M.; Zahrt, A. F.; Sakhaee, N.; Denmark, S. E.; Wennemers, H. Machine Learning to Develop Peptide Catalysts─Successes, Limitations, and Opportunities. *ACS Cent. Sci.* **2024**, *10* (2), 367–373.
(72) Olen, C. L.; Zahrt, A. F.; Reilly, S. W.; Schultz, D.; Emerson, K.; Candito, D.; Wang, X.; Strotman, N. A.; Denmark, S. E. Chemoinformatic Catalyst Selection Methods for the Optimization of Copper–Bis(Oxazoline)-Mediated, Asymmetric, Vinylogous Mukaiyama Aldol Reactions. *ACS Catal.* **2024**, *14* (4), 2642–2655.

(73) Lin, Y.; Zhang, R.; Wang, D.; Cernak, T. Computer-Aided Key Step Generation in Alkaloid Total Synthesis. *Science* **2023**, *379* (6631), 453–457.
(74) Roh, J.; Joung, J. F.; Yu, K.; Tu, Z.; Bartholomew, G. L.; Santiago-Reyes, O. A.; Fong, M. H.; Sarpong, R.; Reisman, S. E.; Coley, C. W. Higher-Level Strategies for Computer-Aided Retrosynthesis. *ACS Cent. Sci.* **2026**, *12* (3), 345–357.
(75) Tu, Z.; Choure, S. J.; Fong, M. H.; Roh, J.; Levin, I.; Yu, K.; Joung, J. F.; Morgan, N.; Li, S.-C.; Sun, X.; Lin, H.; Murnin, M.; Liles, J. P.; Struble, T. J.; Fortunato, M. E.; Liu, M.; Green, W. H.; Jensen, K. F.; Coley, C. W. ASKCOS: Open-Source, Data-Driven Synthesis Planning. *Acc. Chem. Res.* **2025**, *58* (11), 1764–1775.
(76) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-Selectivity Prediction with a Machine-Learned Reaction Representation and on-the-Fly Quantum Mechanical Descriptors. *Chem. Sci.* **2021**, *12* (6), 2198–2208.
(77) Luong, K.-D.; Singh, A. Application of Transformers in Cheminformatics. *J. Chem. Inf. Model.* **2024**, *64* (11), 4392–4409.
(78) Chithrananda, S.; Grand, G.; Ramsundar, B. ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction. arXiv October 23, 2020.
(79) Gao, W.; Raghavan, P.; Shprints, R.; Coley, C. W. Revealing the Relationship between Publication Bias and Chemical Reactivity with Contrastive Learning. *J. Am. Chem. Soc.* **2025**, *147* (10), 8959–8968.

Chapter II – **Models for catalytic activity and stereoselectivity within confined spaces**

Portions of this chapter were adapted from:

Treacy, S. M.†; Smith, A. L.†; et al. *J. Am. Chem. Soc.* **2024,** *146*, 29792–29800. (DOI)

Heafner, L. Smith, A. L.†; Craescu-Valentina, C.†, et al. *Chem* **2025,** *11*, 102368. (DOI)

II.1 – Preface

Catalysis of multicomponent transformations requires controlled assembly of reactants within the active site. Supramolecular scaffolds possess synthetic microenvironments that enable precise modulation over noncovalent interactions (NCIs) engaged by reactive, encapsulated species. While molecular properties that describe the behavior of single guests in host cavities have been studied extensively, multicomponent transformations remain challenging to design and deploy. Here, simple univariate regression and threshold analyses are employed to model reactivity in a cascade reduction of azaarenes catalyzed by water-soluble metal organic cages. Yield and stereoselectivity models help deduce unknown mechanisms of reactivity by the multicomponent, host–guest complexes. Furthermore, a comprehensive model is established for NCIs driving stereoselectivity in the reported host–guest adducts.

II.2 – Introduction

In nature, enzymes leverage an array of carefully tuned non-covalent interactions (NCIs) to decrease activation barriers to reactions along complex biosynthetic pathways[1,2]. To mimic these enzyme binding pockets, chemists have engineered supramolecular assemblies capable of imposing NCIs on encapsulated molecules to enable reactivity not observed in bulk solution [3–5]. Our group has long studied the host-guest complexation within the distinct microenvironments of anionic metal catecholate covalent organic cages and demonstrated their ability to catalyze various transformations. [6,7] (**Fig. II-1A**) More recently, our group has reported numerous synthetic methodologies leveraging reactivity across multiple reaction components, simultaneously assembled within the host cavity, furnishing intermediate *ternary complexes*.[8–10] Our current mechanistic understanding suggests our reported, host-mediated transformations are not only limited by access into these *ternary complexes*, but also influenced by NCIs along the dynamic, reaction coordinate. As these reactive complexes are highly transient, general chemical features that govern these transformations have remained elusive.[11–14]

To analyze complex reactivity, data science has emerged as an invaluable tool to deconstruct key features that correlate to desired reaction outcomes (**Fig. II-1B**). While machine learning-based predictions have rapidly grown in popularity, analysis by simple linear regression models and univariate thresholds is essential to abstract complex reactivity onto a handful of intuitive, chemically tangible principles. [15–17] We posited that data science tools would elucidate the principles that govern access to coencapsulated intermediates and their interactions within active site cavities. Modeling the yield of transformations proceeding through ternary complexes necessarily illuminates chemical features that mediate access into these adducts. Additionally, models conveying any host-mediated chemo-, regio-, or stereoselectivity should provide crucial information regarding the stereoelectronic perturbations of encapsulated intermediates imposed by the cavity microenvironment.

Because previous work from our group has disclosed a host-mediated reduction facilitated by the coencapsulation of both substrate and pyridine-borane, it was anticipated critical insight could be furnished by the discovery of additional, host-mediated, multicomponent reactivity.[9] Specifically, we hypothesized encapsulated azaarenes are poised for a protonation-reduction cascade with pyridine-boranes to furnish valuable semi-hydrogenated products. [18,19] (**Fig. II-1C**)

Herein, we employ a data science-based analysis, coupled with careful experimental design, to provide quantitative insight into the substrate and host features responsible for the observed, supramolecular catalytic activity. Specifically, a yield prediction model based on size exclusion by the host cavity was successfully applied across systematically modified hosts. This model generalizes substrate binding principles of ternary host-guest adducts that undergo the reported reaction. Additionally, transient NCIs responsible for stereoselectivity in the reported reduction cascade could be rationalized through substrate parameterization and calculated NCIs within the binding pocket. Overall, we showcase how a data science-powered workflow precisely quantifies the features of substrates that affect reactivity and stereoselectivity in supramolecular catalysis.

II.3 – Results

- ***Establishing the azaarene reduction cascade and mechanism***

We first considered whether the reduction cascade was possible with host **I**, a Ga_4L_6 coordination cage (L = *N*,*N*′-bis(2,3-dihydroxybenzoyl)-1,5-diaminonaphthalene). The reduction of model azaarene 2-methylquinoline **1** with pyridine-borane mediated by host **I** furnished the semi-hydrogenated, 1,2,3,4-tetrahydro-2-methylquinoline in 97% yield (Entry 1, **Fig. II-2A**). The absence of catalyst or the addition of a strongly encapsulating cationic guest results in low yields of the desired product, which supports catalytic activation of the quinoline and reductant by host **I** occurs within the cavity (Entries 2-4, **Fig. II-2A**). Conducting this protocol at lower temperatures affords the reduced product at attenuated yields and the reductant is required for the desired reactivity (Entries 5-6, **Fig. II-2A**).

As the reaction necessarily proceeds through a coencapsulated complex, we first systematically evaluated how substrate size influences the reduction. Since host **I** permits encapsulation of guests through aperture dilation [20], we hypothesized that volume of the most compact conformer for a given substrate may correlate with the observed reactivity. To interrogate this possibility, we calculated the minimum molecular volumes of ten additional mono- and disubstituted quinolines. This was accomplished by first generating conformational ensembles within a 1.8 kcal/mol window using Grimme's CREST package (see **Appendix I**). Then, the molecular volume was computed using a descriptor calculator from the open-source RDKit package and extracted the minimum value from across the ensemble. The molecular volumes of the quinolines considered range over 140-210 $Å^3$ (**Fig. II-2B**). The reduction of each of the substrates was then evaluated under the conditions established above and their yield was recorded. A linear model between yield and minimum molecular volume was anticipated to characterize how host-dependent size

exclusion of the reactive complex emerges as a response to guest size. An analogous analysis can be completed using univariate threshold detection, where substrates are binned into active or inactive categories based on their yield and the molecular volume of these substrates is used to detect where size exclusion formally emerges. Consistent with the hypothesis that guest size relates to reaction yield, a linear correlation was observed where the minimum molecular volume of the substrate captured experimental yield with an R^2 of 0.77 (**Fig. II-2C**). The corresponding threshold analysis was similarly robust with an active volume of 172 $Å^3$ at a reactivity threshold of 30% (MCC = 0.83).

The data science-based analysis supports the hypothesis that quinoline molecular volume has a profound influence on access into the reactive, ternary complex according to molecular volume. Despite the observation of minimal to no reactivity with substrates with minimum molecular volumes >200 $Å^3$, encapsulation studies of 2-substituted quinolines **1-4** with volumes ranging from 140-205 $Å^3$ show that even the largest members quantitatively encapsulate within the cavity of host **I** under reaction concentrations (see **Appendix II – NMR Spectra**). Thus, in accordance with the above correlations between quinoline volume and reaction yield, the rate of subsequent coencapsulation of the reductant was anticipated to become slower as quinoline volume increases. This was reinforced with an initial rates kinetic analysis of the reduction that is consistent with first-order rate dependence on catalyst and pyridine-borane concentrations, but zero-order dependence on quinoline concentration which we ascribe to saturation in this substrate (see **Supplementary Fig. II-1-4**). In agreement with the former quinoline encapsulation studies, these data indicated that the formation of a resting-state complex between the host and the quinoline guest occurs prior to pyridine-borane coencapsulation. To interrogate hydride delivery as the potential rate-determining step in this reaction, kinetic isotope effect experiments were conducted via one-pot competition between $PyBH_3$ and $PyBD_3$ which yielded no significant difference in D versus H incorporation at either position on the product (see **Supplementary Fig. II-5-6**). The lack of a kinetic isotope effect in combination with the first-order rate dependence on reductant is consistent with rate-limiting coencapsulation of pyridine-borane to access the reactive complex.

- ***Reactivity principles conserved across hosts***

Because pyridine-borane coencapsulation is implicated as the turnover-limiting step, the yield of the host-mediated reduction formally relates to how one host-guest complex can accommodate strain along a second encapsulation event. Accelerating this second encapsulation event rationally implies greater yield in the transformation; however, guest exchange rates for these bimolecular events are unknown. Instead, only their unimolecular congeners are precedented [20]. Over the past decade, a modest array of hosts with variable guest-exchange rates has been achieved by systematic modifications to the metal vertices of the host and the ligand linker of the metal organic cage [10] (**Fig. II-3A**). Tuning the metal vertices of the cage modulates guest-exchange rates but does not change the absolute size of the host cavity [21]. The prototypical **Ga (host I)** vertex can be exchanged to **Al (host II)**, affording hosts with slower guest exchange, or alternatively **In (host III)**, affording hosts with more rapid guest exchange, according to the strength of the metal-oxygen

bonds of the scaffold. The ligand linker of the assembly can also be substituted, as the naphthalene spacer of host **I** can be replaced with pyrene furnishing host **IV** [22]. As a result, host **IV** not only exhibits a larger aperture, lowering barriers to guest encapsulation, but also a larger host cavity altogether, enabling an absolute increase to the size of guests tolerated.

With consideration of these principles discovered in unimolecular contexts, reaction yield was leveraged as a platform to measure access into the reactive ternary complexes. Thus, additional datasets for each host assembly **II-IV** were generated from the azaarene scope described in Figure 2B and subjected each to linear regression and univariate threshold analysis. Across all hosts, good-to-excellent correlations between yield and minimum molecular volume were observed (up to R^2 = 0.84, **Fig. II-3B**). Additionally, excellent univariate threshold detection over minimum molecular volume was observed (up to MCC = 1.0). The In host **III** was expected to provide a more flexible assembly, suitable for more rapid guest exchange. Indeed, the In host **III** shows an upward slant of the yield correlation as well as an increased threshold for reactivity (184 $Å^3$ versus 172 $Å^3$, MCC = 0.83) relative to Ga host **I**, indicating larger substrates more easily permit coencapsulation with pyridine borane, thus accessing higher yields, such as product **3k** now formed in **66**% yield (vs. 35% with host **I**). In contrast, Al host **II** is expected to have lower exchange rates. The observed correlation and threshold analysis reflect this expectation showing a pronounced downward slant and depressed threshold for reactivity relative to host **I**, (163 $Å^3$ versus 172 $Å^3$, MCC = 1.00) indicating a severe kinetic barrier to coencapsulation of pyridine borane with many quinolines screened. Finally, the Ga host **IV** with a larger pyrene spacer is expected to not only have faster exchange rates, but also a larger absolute limit to guest encapsulation due to its larger cavity. The observed correlation reflects this by a rightward shift of the yield correlation, now corresponding to higher yields across the 130-210 $Å^3$ minimum molecular volume series as well as a best-in-class threshold for reactivity (189 $Å^3$ versus 172 $Å^3$, MCC = 1.00). Overall, our protocol quantitatively substantiates relationships between guest size and flexibility of the host catalyst for guest coencapsulation, transforming previously established qualitative principles for more complex applications

- ***Elucidating stereoelectronic effects on host-catalyzed, cascade reduction***

To elaborate which features of the intermediate host-guest complex influence the second encapsulation event, the scope of quinolines considered was expanded for the reduction. As pyrene-based host **IV** broadly displayed improved yields in the previous dataset, we assessed the generality of the yield prediction model based on minimum molecular volume of the quinoline substrate. The expanded scope considered a wider range of substituted quinolines with varying steric bulk and additional functional groups (**Fig. II-4A**). After performing the reactions with host **IV** and comparing the outcomes to the predicted yields, the observed minimum molecular volume exhibits little correlation to the experimental yield (R^2 = 0.20, **Fig. II-4A**).

To investigate the origin of this result, we hypothesized the minimum molecular volume of the quinoline guest erroneously assumes each quinoline maintains a shape and corresponding

stereoelectronic character. As such, the expanded substrate scope would be better represented by separate groups of quinolines, where each group shares similar stereoelectronic properties such that their host binding behavior is internally consistent. The substrate scope was categorized by clustering all quinolines considered over a principal component-reduced chemical space built from the *minimum molecular volume* dataset. The *minimum molecular volume* dataset refers to the full array of features extracted from the conformer of each quinoline that has the minimum molecular volume across the ensemble computed (see **Appendix I**). This analysis revealed two clusters, wherein the first cluster was dominated by 2- and 3-substituted quinolines while the second cluster was best represented by all other substitution patterns (**Fig. II-4B**). Within each cluster, yield correlations built from minimum molecular volume remained poor, and univariate threshold analyses lost coherence. Therefore, we screened suitable bivariate correlations and noticed minimum molecular volume and molecular dipole now rescued yield predictions (up to $R^2 = 0.80$, **Fig. II-4C**). Interrogation of each cluster's yield correlation revealed molecular dipole contributed to each cluster's prediction in the opposite sense. As molecular dipole is related to the conformation of the molecule, we interpret the improved bivariate linear correlations to indicate that molecular dipole differentially corrects each yield prediction based solely on molecular volume to account for apparent shape differences between each cluster (see **Appendix II – Extended Discussion on Fig. II-4C**). Whereas the previous models explicitly modeled steric effects on tertiary complex inhibition, an expanded chemical space demonstrated that adequate representation of molecular shape with confounding electronic features of the substrates is essential to understand global reactivity trends.

- ***Illuminating transient active-site interactions***

While yield prediction tasks quantified host and guest features influencing access to coencapsulated intermediates, the behavior of the coencapsulated complex itself remained elusive. Here, the diastereoselectivity of the reduction can provide valuable information to understand the behavior of transient intermediates within the host cavity. Akin to data science analyses of homogenous reactions as well as enzymatic transformations [23–30], we envisioned we could adapt the data science workflow to agnostically probe for features that correlate to the measured selectivity, where conclusions generated by this data science-driven inquiry can inspire unprecedented models describing outcomes from host-mediated catalysis.

Indeed, various 2,3-disubstituted azaarenes underwent the reduction with high diastereoselectivity for the *anti*-isomer. In this expanded scope, modest-to-good selectivity from 2:1 to 11:1 (*anti/syn*) was observed for the reduction (**Fig. II-5A**). Solution state reduction of **1** also favors the *anti* diastereomer but only in an attenuated 3:1 selectivity (compared to 10:1 by host **IV**). When correlations were evaluated between substrate features and the experimental stereoselectivity, a model based on the molecule's width (quantified by sterimol BMin) and electronegativity emerged among the best performing ($R^2 = 0.78$, LOO-$R^2 = 0.73$, L2O-$R^2 = 0.65$; see **Supplementary Fig. II-25**). Inspection of the correlation's coefficients indicated that molecules with narrower widths and higher electronegativity are reduced with higher stereoselectivity. The combination of

properties in the correlation could describe how accessible the p-system of the quinoline is both sterically and electronically, but it remains unclear what specific interactions are relevant for the observed reactivity.

To further evaluate how each reactive species contributes to the observed stereoselectivity, a structurally diverse series of different pyridine-boranes were then prepared and screened with quinoline **3j**. Now, a more pronounced range of diastereoselectivity was observed (11:1 to 1:3, *anti/syn*), where some reductants even favored the formation of the *syn* diastereomer (**Fig. II-5B**). Surveyed correlations revealed the minimum 4 Å buried volume at *C*-4 of the reductant excellently correlates to the observed selectivity (R^2 = 0.95). Here, stereoselectivity tends to invert as this buried volume increases. In line with the interpretation proposed for the quinoline series, access to favorable NCIs promotes the diastereoselectivity for the *anti*-isomer.

From these data science-based footholds, we then looked to deconvolute the specific NCIs relevant to the observed stereochemical outcomes. In this sense, the diastereoselectivity measured for the quinolines from Figure 5A can be recapitulated with competing non-covalent interactions between pro-*anti* and pro-*syn* complexes for the intermediate iminium species with the pyrene wall of the host (**Fig. II-5C**). These interactions display a strong correlation to the measured diastereoselectivity (R^2 = 0.87). We also revisited the effect of changing the pyridine-borane reductant on NCIs within the coencapsulated complex. Here, we observed the calculated non-covalent interaction between the reductant and the iminium intermediate in the pro-*anti* complex provided a robust correlation (R^2 = 0.96, **Fig. II-5D**). This correlation indicates when the interaction between the cationic intermediate and the reductant strengthens, the selectivity in the reduction gradually shifts from the *anti* to the *syn* diastereomer. This suggests a delicate balance between host-quinoline and quinoline-reductant interactions contributes to diastereoselectivity.

Altogether, we inferred the following model for diastereoinduction of the transient coencapsulated intermediates (**Fig. II-6**). Beginning microscopically from the singly reduced dihydroquinoline, protonation leads to one of two iminium species. Here, we propose competing NCIs serve as the selectivity-determining criteria. In the pro-*anti* case, the host wall, approximated with a single pyrene molecule, achieves an energetically favorable, uninhibited interaction, while the pro-*syn* case is disrupted by steric interactions with the pyrene wall. However, as the reductant achieves stronger interactions with the proposed intermediate, the selectivity gradually inverts to favor the *syn* diastereomer. In this pathway, the reductant can engage in more energetically favorable, sterically unhindered interactions with the reactive species. Through a data science-powered workflow we have gained valuable mechanistic understanding of the key NCIs that engender stereoselectivity in supramolecular host-catalyzed reactions of coencapsulated intermediates.

II.4 – Discussion and Outlook

Herein, we showcase how data science tools reveal complex dynamics in a catalyst structure-agnostic workflow. Coupling these tools with data generated from a supramolecular host-catalyzed cascade reduction of azaarenes transforms over a decade of mechanistic investigation of metal organic cages into quantifiable, rational synthetic choices. Our work highlights how supramolecular catalysis uniquely facilitates rational adaptations to host size and flexibility, enabling the design of non-covalent interaction rich active sites that productively manipulate reactive, ternary complexes. Critically, our workflow showcases how selection of chemically meaningful features provides interpretable models that synergize and enhance our expert understanding of how to exploit active sites in catalysis.

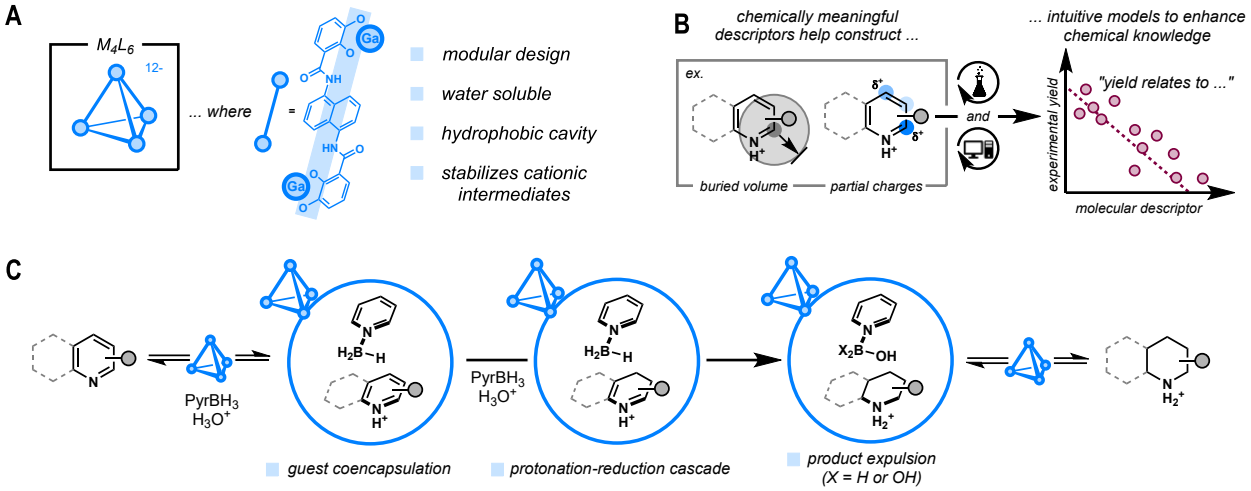


Fig. II-1 | Investigating reactive, ternary complexes requires chemically coherent systems and models. (**a**) Metal-catecholate cages possess synthetically tractable, biomimetic binding pockets. (**b**) Data science helps generate chemical knowledge based on correlations between chemically meaningful descriptors and experimental observables. (**c**) Proposed model azaarene reduction cascade to interrogate ternary complex formation and reactivity.

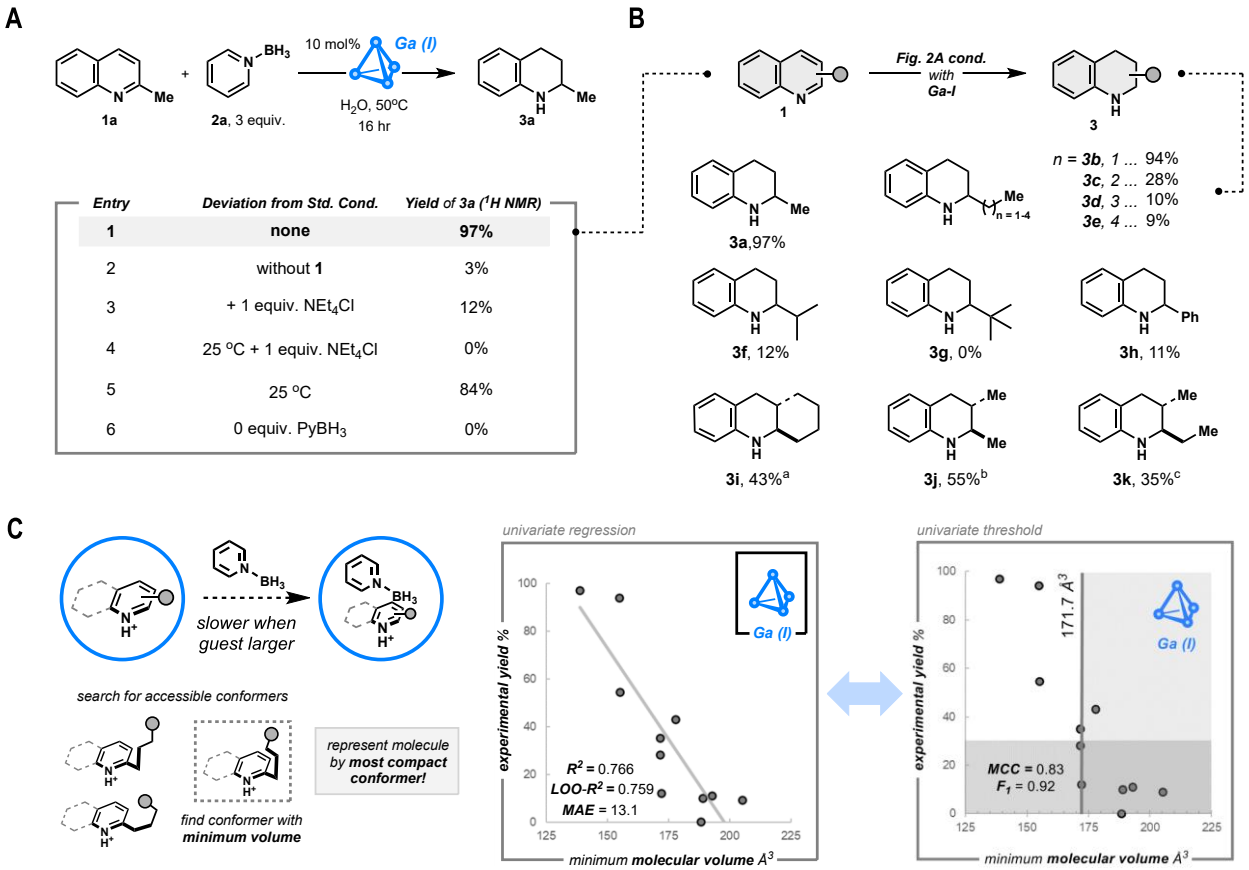


Fig. II-2 | Establishing the host-mediated azaarene reduction cascade reactivity, scope, and mechanism. (a) Reaction discovery and controls. **(b)** Systematic series of quinolines, see SI Table S6, [a]4:1 *dr*, [b]9:1 *dr*, [c]9:1 *dr*. **(c)** Representation of molecular volume in its most compact conformer linearly correlates ($R^2 = 0.77$) with yields of supramolecular catalyzed transformation and are well represented by a univariate threshold analysis (yield threshold = 30%, detected molecular volume threshold = 171.7 Å^3, MCC = 0.83).

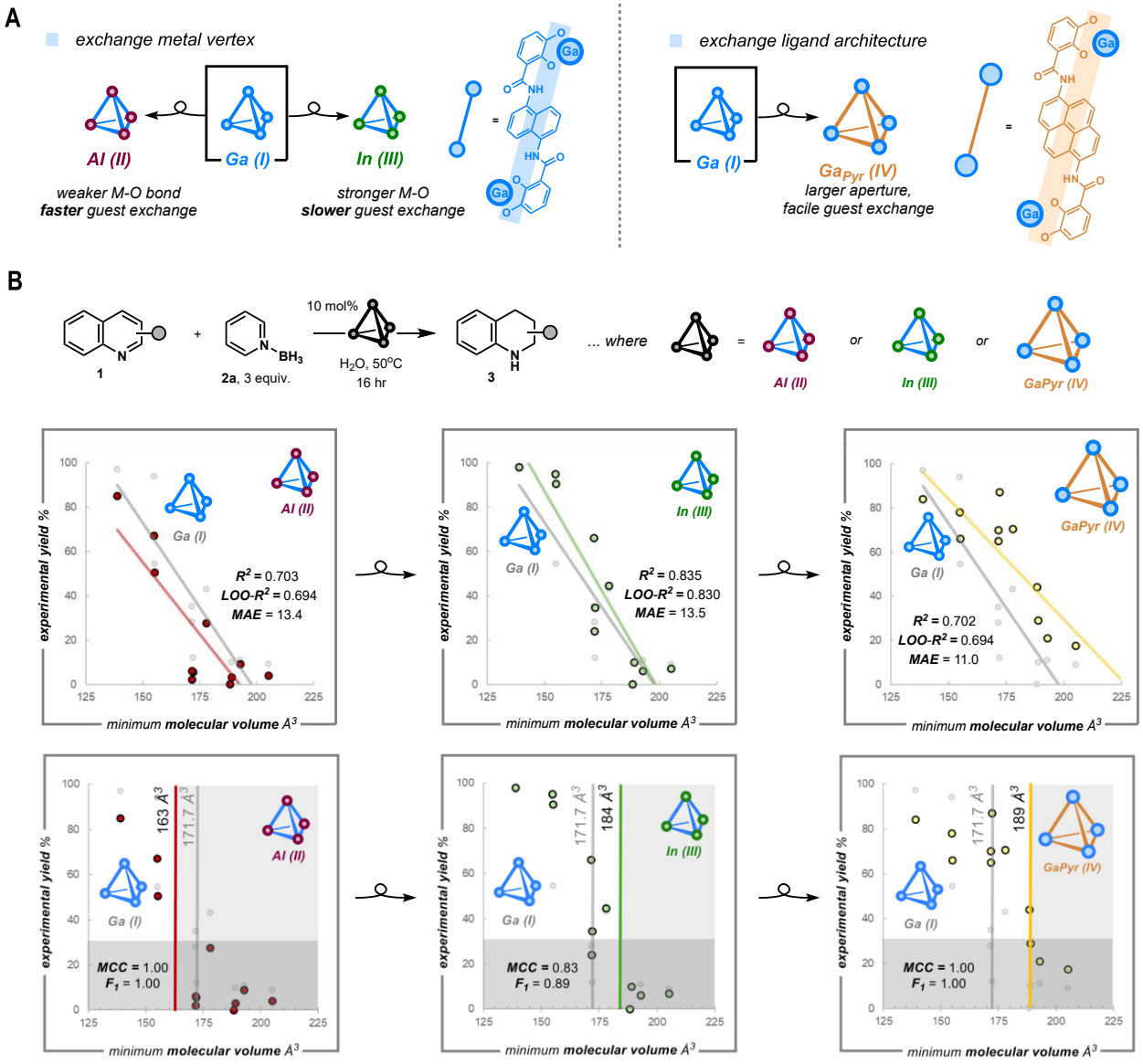


Fig. II-3 | Chemical principles describing reactivity of ternary complexes transfers across distinct hosts assemblies. (**a**) Exchange of the metal cation of the supramolecular cage increases the flexibility of the cavity, leading to differences in guest exchange rates. Using a pyrene ligand architecture increases the size of the cavity and enables more facile guest exchange. (**b**) Minimum molecular volume maintains robust linear correlations across the catalyst series (up to $R^2 = 0.84$), with corresponding excellent univariate threshold analyses (yield threshold = 30%, up to MCC = 1.0), see SI Table S6-S9.

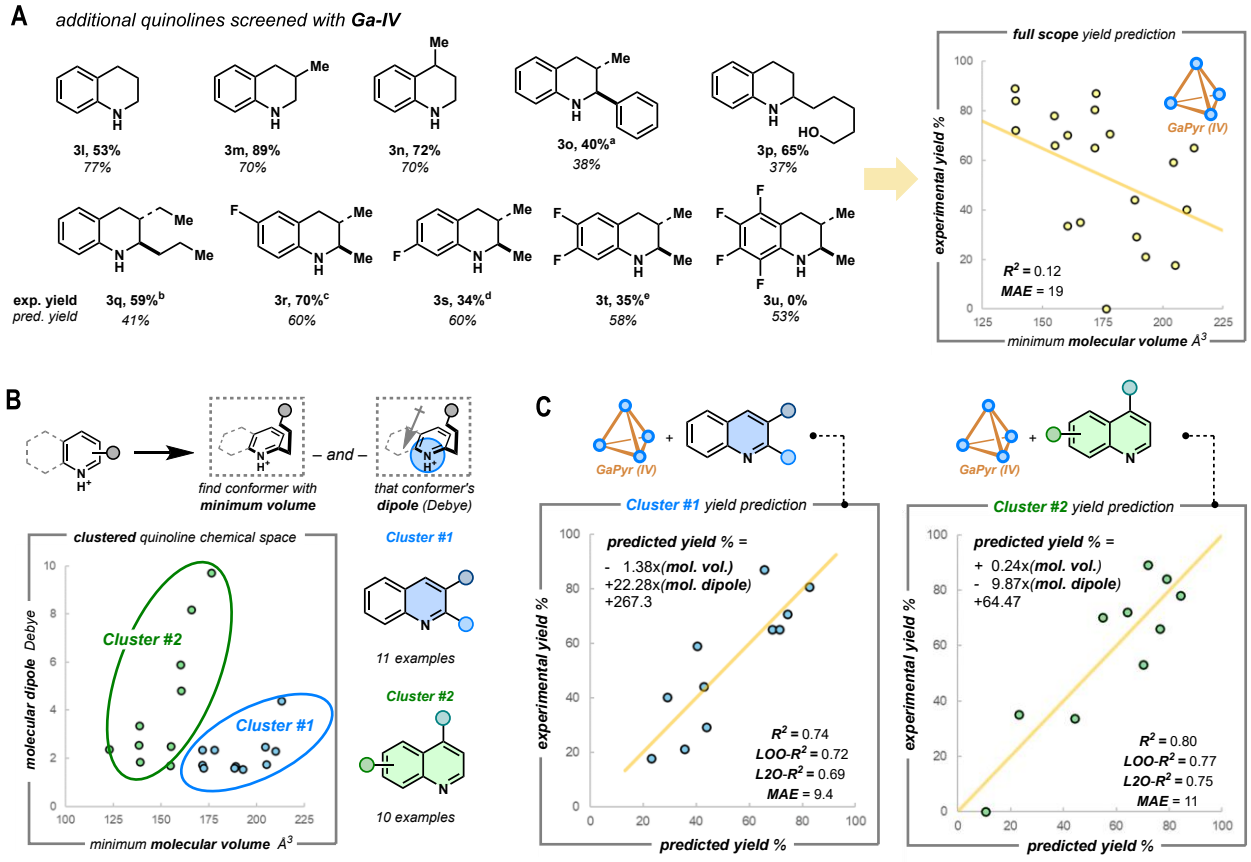


Fig. II-4 | Reactivity of ternary complex over an expanded substrate scope divergently correlates to substrate stereoelectronic features. (a) Additional quinolines scope and limited prediction capacity of minimum molecular volume on reaction yield, see SI Table S10, [a]1.7:1 *dr* [b]11.2:1 *dr* [c]10.5:1 *dr* [d]4.6:1 *dr* [e]4.5:1 *dr*. (**b**) Unsupervised clustering reveals two classes of quinoline substrates. (**c**) Cluster specific bivariate linear regression using molecular volume and molecular dipole differentially correlates to reaction yield, up to $R^2 = 0.80$, see SI Table S11.

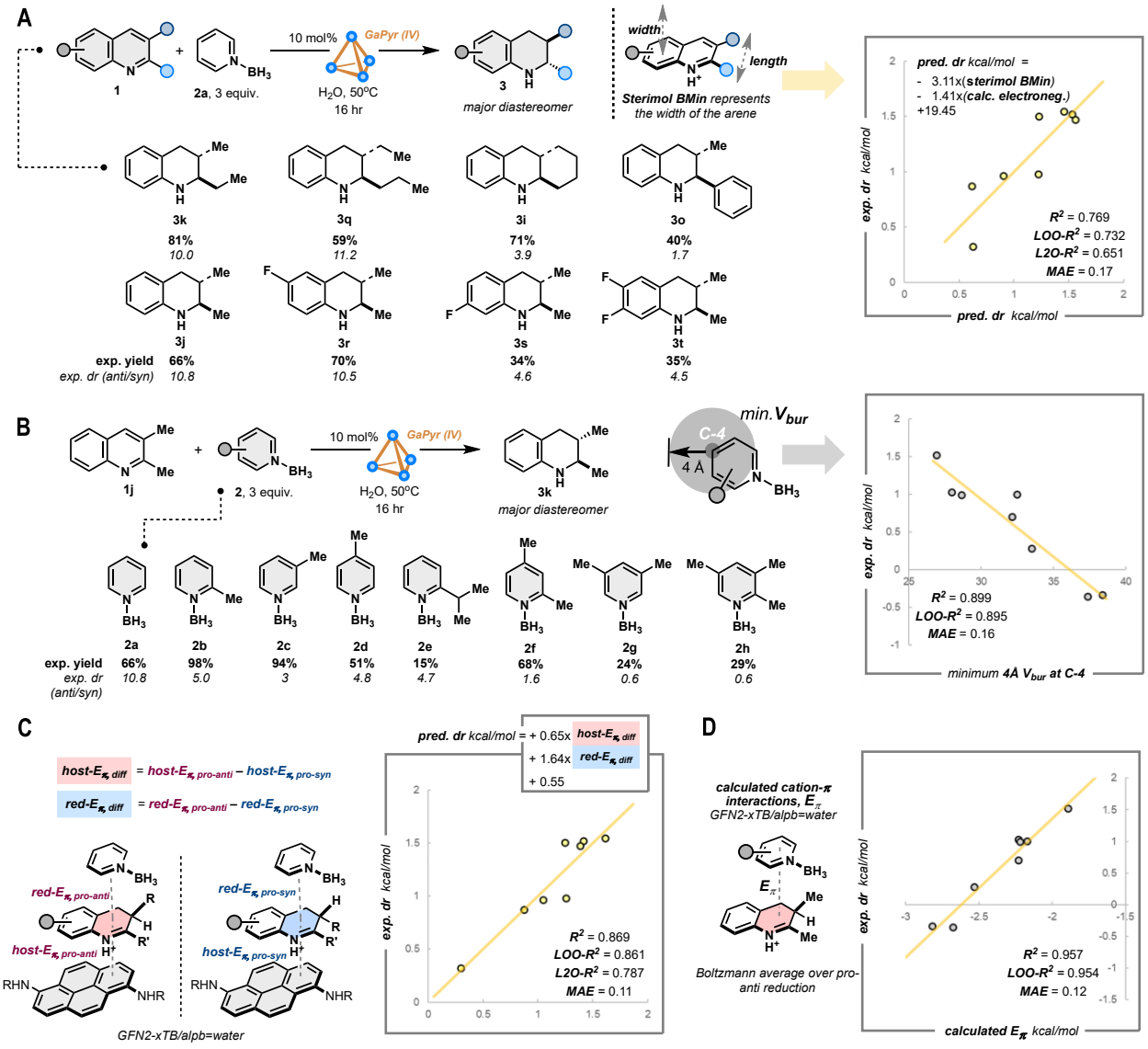


Fig. II-5 | Evaluation of transient, stereoselective interactions across the reactive, ternary complex within the host-mediated azaarene reduction cascade. (a) Sterimol BMin and electronegativity of the quinoline substrate correlate to the diastereoselectivity of the azaarene reduction, R2 = 0.77, see SI Table S14. (**b**) Buried volume at C-4 of pyridine-borane correlates to differences in diastereoselectivity of the 1j reduction, R2 = 0.90, see SI Table S15. (**c**) NCIs across proposed selectivity determining ternary complex recapitulates observed diastereoselectivity for a series of quinolines with 2a, R2 = 0.87, see SI Table S17 (**d**) NCIs across proposed selectivity determining ternary complex recapitulates observed diastereoselectivity for a series of pyridine boranes with 1j, R2 = 0.96, see SI Table 18.

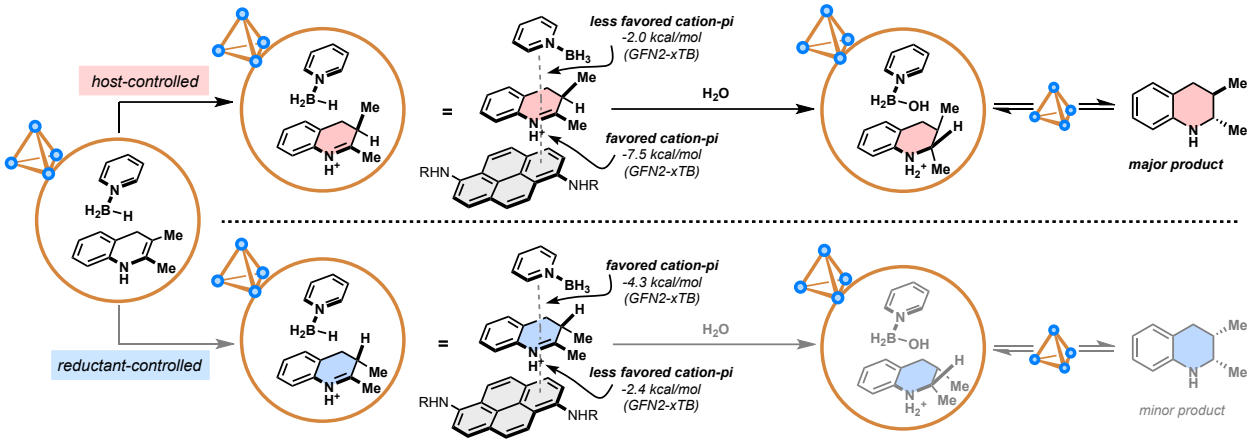


Fig. II-6 | Comprehensive model for the NCIs driving stereoselectivity in the reported reactive, ternary host-guest adducts.

II.5 – References

(1) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew Chem Int Ed* **2004**, *43* (48), 6596–6616.

(2) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53* (14), 5061–5084.

(3) Yoshizawa, M.; Tamura, M.; Fujita, M. Diels-Alder in Aqueous Molecular Hosts: Unusual Regioselectivity and Efficient Catalysis. *Science* **2006**, *312* (5771), 251–254.

(4) Kang, J.; Rebek, J. Acceleration of a Diels–Alder Reaction by a Self-Assembled Molecular Capsule. *Nature* **1997**, *385* (6611), 50–52.

(5) Zhang, Q.; Tiefenbacher, K. Terpene Cyclization Catalysed inside a Self-Assembled Cavity. *Nature Chem* **2015**, *7* (3), 197–202.

(6) Kaphan, D. M.; Levin, M. D.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. A Supramolecular Microenvironment Strategy for Transition Metal Catalysis. *Science* **2015**, *350* (6265), 1235–1238.

(7) Hong, C. M.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Self-Assembled Tetrahedral Hosts as Supramolecular Catalysts. *Acc. Chem. Res.* **2018**, *51* (10), 2447–2455.

(8) Bierschenk, S. M.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. A Nanovessel-Catalyzed Three-Component Aza-Darzens Reaction. *J. Am. Chem. Soc.* **2020**, *142* (2), 733–737.

(9) Morimoto, M.; Cao, W.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Chemoselective and Site-Selective Reductions Catalyzed by a Supramolecular Host and a Pyridine–Borane Cofactor. *J. Am. Chem. Soc.* **2021**, *143* (4), 2108–2114.

(10) Bierschenk, S. M.; Pan, J. Y.; Settineri, N. S.; Warzok, U.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Impact of Host Flexibility on Selectivity in a Supramolecular Host-Catalyzed Enantioselective Aza-Darzens Reaction. *J. Am. Chem. Soc.* **2022**, *144* (25), 11425–11433.

(11) 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* (10), 5806–5826.

(12) Rizzuto, F. J.; Von Krbek, L. K. S.; Nitschke, J. R. Strategies for Binding Multiple Guests in Metal–Organic Cages. *Nat Rev Chem* **2019**, *3* (4), 204–222.

(13) Parrilla-Gutiérrez, J. M.; Granda, J. M.; Ayme, J.-F.; Bajczyk, M. D.; Wilbraham, L.; Cronin, L. Electron Density-Based GPT for Optimization and Suggestion of Host–Guest Binders. *Nat Comput Sci* **2024**, *4* (3), 200–209.

(14) Norjmaa, G.; Vidossich, P.; Maréchal, J.-D.; Ujaque, G. Modeling Kinetics and Thermodynamics of Guest Encapsulation into the $[M_4L_6]^{12-}$ Supramolecular Organometallic Cage. *J. Chem. Inf. Model.* **2021**, *61* (9), 4370–4381.

(15) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8* (8), 7403–7429.

(16) Kitchin, J. R. Machine Learning in Catalysis. *Nat Catal* **2018**, *1* (4), 230–232.

(17) Suvarna, M.; Pérez-Ramírez, J. Embracing Data Science in Catalysis Research. *Nat Catal* **2024**, *7* (6), 624–635.

(18) Ma, J.; Chen, S.; Bellotti, P.; Guo, R.; Schäfer, F.; Heusler, A.; Zhang, X.; Daniliuc, C.; Brown, M. K.; Houk, K. N.; Glorius, F. Photochemical Intermolecular Dearomative Cycloaddition of Bicyclic Azaarenes with Alkenes. *Science* **2021**, *371* (6536), 1338–1345.

(19) Lückemeier, L.; De Vos, T.; Schlichter, L.; Gutheil, C.; Daniliuc, C. G.; Glorius, F. Chemoselective Heterogeneous Hydrogenation of Sulfur Containing Quinolines under Mild Conditions. *J. Am. Chem. Soc.* **2024**, *146* (9), 5864–5871.
(20) Davis, A. V.; Raymond, K. N. The Big Squeeze: Guest Exchange in an M_4L_6 Supramolecular Host. *J. Am. Chem. Soc.* **2005**, *127* (21), 7912–7919.
(21) Hong, C. M.; Morimoto, M.; Kapustin, E. A.; Alzakhem, N.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Deconvoluting the Role of Charge in a Supramolecular Catalyst. *J. Am. Chem. Soc.* **2018**, *140* (21), 6591–6595.
(22) Hart-Cooper, W. M.; Zhao, C.; Triano, R. M.; Yaghoubi, P.; Ozores, H. L.; Burford, K. N.; Toste, F. D.; Bergman, R. G.; Raymond, K. N. The Effect of Host Structure on the Selectivity and Mechanism of Supramolecular Catalysis of Prins Cyclizations. *Chem. Sci.* **2015**, *6* (2), 1383–1393.
(23) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7* (10), 1622–1637.
(24) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, *347* (6223), 737–743.
(25) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543* (7647), 637–646.
(26) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
(27) Toste, F. D.; Sigman, M. S.; Miller, S. J. Pursuit of Noncovalent Interactions for Strategic Site-Selective Catalysis. *Acc. Chem. Res.* **2017**, *50* (3), 609–615.
(28) Crawford, J. M.; Kingston, C.; Toste, F. D.; Sigman, M. S. Data Science Meets Physical Organic Chemistry. *Acc. Chem. Res.* **2021**, *54* (16), 3136–3148.
(29) Kelly, S. P.; Shende, V. V.; Flynn, A. R.; Dan, Q.; Ye, Y.; Smith, J. L.; Tsukamoto, S.; Sigman, M. S.; Sherman, D. H. Data Science-Driven Analysis of Substrate-Permissive Diketopiperazine Reverse Prenyltransferase NotF: Applications in Protein Engineering and Cascade Biocatalytic Synthesis of (−)-Eurotiumin A. *J. Am. Chem. Soc.* **2022**, *144* (42), 19326–19336.
(30) Clements, H. D.; Flynn, A. R.; Nicholls, B. T.; Grosheva, D.; Lefave, S. J.; Merriman, M. T.; Hyster, T. K.; Sigman, M. S. Using Data Science for Mechanistic Insights and Selectivity Predictions in a Non-Natural Biocatalytic Reaction. *J. Am. Chem. Soc.* **2023**, *145* (32), 17656–17664.

Chapter III – **Simple generalizations of stereoselectivity over diverse sets of acid organocatalysts**

Portions of this chapter were adapted from:

Smith, A. L.; Toste, F. D. *J. Am. Chem. Soc.* **2026**, *148*, 2792-2800. (DOI)

III.1 – Preface

Stereoselective catalysis is a paradigm of complexity, navigating arrays of specific, yet weak non-covalent interactions to achieve asymmetric induction. Accordingly, structure-activity relationships modeling stereoselectivity have required bespoke representations, high dimensionality, and/or large datasets to furnish productive trends, hampering their subsequent interpretability and utility. Here, we report active site-based buried volume is a uniquely advantageous descriptor for the construction of models to stereoselectivity by Brønsted acid organocatalysts. Through statistical analyses of 50+ datasets across nearly 200 distinct catalysts, we realize active site-based buried volume is an exceptional representation, accommodating both extensive structural diversity (within datasets) and functional diversity (across datasets). We show the descriptor's value likely relates to its surprising capacity to generalize beyond simple steric interactions and account for diverse stereoelectronic effects consequential to reported stereoselectivity. As such, we anticipate future modeling of stereoselectivity by Brønsted acid are greatly simplified via the concise representation – active site-based buried volume.

III.2 – Introduction

In chemistry, *structure informs function* is a paradigm invoked to justify the emergence of structure-activity relationships (SARs). In theory, chemical properties are paired with structural changes made, and inferences are drawn to rationalize the emergent patterns with reaction outcomes. To build quantitative relationships, *molecular representations* are used as surrogates for chemical properties (**Fig. III-1A**).[1] While many traditional SARs are constructed with *representations* rooted in specific experimental phenomena (ex. Hammett,[2] Taft,[3] Charton[4]), such *representations* display narrow *domains of applicability*[5] – culminating in critical *discontinuity failures* where certain molecules of interest cannot be described by the model altogether. To address this, modern SARs are typically constructed from calculated representations (ex. sterimol,[6] NBO partial charge,[7] etc.) that are intrinsically more generalizable.[8–10] But, as *molecular representations* become more abstract, their subsequent interpretations become increasingly unintuitive, limiting the production of useful and *general* chemical rationale.[11,12]

The construction of *intuitive* SARs is essential for understanding complex chemical phenomena. In synthetic chemistry, stereoselective catalysis is a paradigm of complexity. Here, a chiral catalyst specifically engages in an array of weak diastereoselective interactions to kinetically steer reactive intermediates towards one stereoisomer over another (**Fig. III-1B**).[1,13,14] Notably, Brønsted acid organocatalysts (BAOCs) are one versatile tool popularized over the past two decades, enabling a wide array of applications, including: Brønsted acid/base catalysis,[15,16] cooperative transition metal catalysis,[17] and counter-anion directed catalysis,[18] among others (**Fig. III-1C**). Their functional diversity arises from the pronounced structural modularity of BAOC scaffolds (**Fig. III-1D**).[19,20] To navigate this structural and functional dimensionality, synthetic chemists have made considerable progress engineering stereoselective BAOCs from two criteria – acidity[15,21] and

confinement.[22,23] These properties are routinely *qualitatively* invoked, but their quantitative *representations* have lacked coherence and interpretability.

Seminal reports by Cavallo,[24,25] Denmark,[26] List,[27] Reid,[28] and Sigman[29] have demonstrated calculated *representations* of confinement by BAOCs can help to construct SARs for stereoselectivity. However, these models have relied on unintuitive, bespoke representations for confinement, are complicated by excessive dimensionality (ex. too many other descriptors in the model), or are evaluated on only a very narrow set(s) of reactions. Although confinement reflects a powerful *qualitative* criterion for the construction of stereoselective BAOCs, a precise *molecular representation*, relevant across the breadth of their utility, has yet to be realized.

We herein report a powerful representation for the stereoselectivity by BAOCs – *active site-based* buried volume – emerging from statistical analyses of over 50+ datasets (**Fig. III-1E**). Its superior performance was benchmarked against that of many familiar and meaningful molecular descriptors, highlighting its unique ability to correlate to stereoselectivity across many datasets, regardless of the underlying structural and functional diversity of the BAOCs considered. We show that the representation, while traditionally reflecting *confinement* – a steric phenomenon – likely generalizes across an array of stereoelectronic non-covalent interactions.

III.3 – Results

- ***Active site-based buried volume emerges as an advantageous representation***

Stereoselective catalysis relies on an array of weak, but specific, non-covalent interactions (NCIs) between a chiral catalyst and a substrate. Accordingly, a *robust* representation for stereoselective catalysts necessarily (1) accommodates structural diversity of the deployed chiral catalysts, (2) generalizes across the accessible functional manifolds, and (3) mediates mechanistic differences observed from each transformation.

In the case of BAOCs, a range of molecular descriptors have emerged as promising candidates, as reported in SARs by Cavallo,[25] Denmark, Reid,[28] Sigman,[9] and our own group.[29] Qualitatively, these descriptors aim to reflect *confinement* of the BAOC's active site (*vida supra*) through various geometric and steric representations, including: bond length, angle, sterimol,[30] and buried volume[24] (**Fig. III-2A**). Beyond representations of *confinement*, many reported models also invoke stereoelectronic contributions to stereoselectivity from adjacent, modular substituents. We account for these expectations by considering an array of *substituent-based* features calculated from the analogous *aryl bromide* as a surrogate (**Fig. III-2A**).[31] Finally, as the *acidity* of BAOCs is often invoked to contextualize the differences across a series of chiral acids, we include calculated pK_a to reflect this expectation (see **Appendix I** and **Appendix III – Molecular Descriptor Calculation Workflow** for full details; see published **dataset_active-site.csv** and **dataset_substituent.csv** for full list of computed descriptors used in this work).[32]

Equipped with an array of candidates for general representations for BAOCs, we then collated a set of 50+ reaction datasets from the literature, culminating in nearly 200 unique BAOCs modeled (see reported **dataset_rxn.xlsx** for full list of curated datasets). To ensure the impact of any representation considered is *clearly* detected, univariate linear regression models trained on each reaction dataset serve as the basis for analysis, avoiding data-intensive, high-dimensional models (ex. neural networks, random forest, multivariate linear regression, etc.) that readily overfit on smaller datasets and become increasingly challenging to interpret (avg. dataset size in *this work* – 11).[12,33] To compare the overall performance of each representation, we evaluated regression *accuracy* (R^2, MAE – mean absolute error) and *robustness* (*p*-value, LOO R^2 and LOO MAE – leave-one-out cross-validation) across each dataset.[34] For representations spanning multiple atoms, dimensions, and ensembles (ex. partial charge at different atoms, buried volume across different radii), only the most *robust* regression model (largest LOO R^2 across the set) for a given dataset was considered in further analyses.

Across the 50+ reaction datasets analyzed, *active site-based* buried volume uniquely emerged as the most *accurate* and *robust* representation (67% of all models are *statistically meaningful* with >95% confidence, 48% with 99% confidence, avg. $R^2 = 0.57$, avg. MAE = 0.21; **Fig. III-2B** and **Fig. III-2C**). At the strictest confidence shown, the average descriptor only constructed meaningful univariate models for one fifth of the datasets (avg. 22% with >99% confidence). Surprisingly, other active site-based descriptors like sterimol L and BMin, which were expected to convey *confinement* like *active site-based* buried volume, are among the last-ranked descriptors with very sparse utility (avg. 9% with >99% confidence). Instead, the next highest ranked descriptor is *substituent-based* buried volume (39% with >99% confidence; avg. $R^2 = 0.56$, avg. MAE = 0.22). The descriptor has been deployed to represent *confinement* akin to its *active site-based* analog but apparently cannot correlate to enantioselectivity as generally across the datasets (**Fig. III-2B**; see **Appendix III – Detailed Statistical Analyses** and reported **dataset_scores.xlsx** for complete statistical analyses; see **Supplementary Fig. III-21-74** for a complete list of univariate regression models by *active site-based* buried volume).

- ***Active site-based buried volume accommodates structural heterogeneity***

Buried volume captures the largest proportion of *meaningful* univariate models relative to other descriptors, but its position – either *active site-based* or *substituent-based* –influences its utility. The juxtaposition between the two suggests the better-performing *active site-based* representation conveys salient structural changes in a more general manner. It was anticipated that *active site-based* buried volume will advantageously correlate to stereoselectivity when the reaction datasets are constructed from structurally diverse sets of BAOCs, while *substituent-based* buried volume should generate robust correlations when the datasets are dominated by changes to the modular substituent. So, we organized the datasets by their corresponding structural diversity (**Fig. III-3A**; see **Appendix III – Detailed Statistical Analyses** for classification criteria). 27 of the 54 reactions considered can be classified as *limited diversity* – essentially, the dataset has one or fewer examples

of catalysts differing beyond the modular substituent (ex. different chiral backbones and/or active sites). The remaining 27 reactions are classified as *significant diversity*.

Over the *limited diversity* datasets, *active site-based* and *substituent-based* buried volumes both perform well. As these datasets are populated with BAOCs varying primarily at the modular substituent, the *substituent-based* representation more precisely reflects the chemical changes (avg. $R^2 = 0.65$, avg. MAE = 0.19 kcal/mol; **Fig. III-3B**), while the *active site-based* representation performs at a similar caliber (avg. $R^2 = 0.64$, avg. MAE = 0.19 kcal/mol). On the other hand, across *significant diversity* datasts, *active site-based* buried volume outperforms its *substituent-based* analog. The heterogeneity of the datasets requires the descriptors to quantitatively reflect structural changes across the acids. As such, the utility of *substituent-based* buried volume substantially decreases (avg. $R^2 = 0.47$, avg. MAE = 0.25 kcal/mol; **Fig. III-3C**), while *active site-based* buried volume's is relatively sustained (avg. $R^2 = 0.51$, avg. MAE = 0.23 kcal/mol).

To highlight the impact structural heterogeneity, we showcase the utility of *active site-based* and *substituent-based* buried volume in two reaction datasets. R-**X**, a seminal dataset reported by Zahrt, et al.[26] describes the stereoselective addition of a thiol to an imine (**Fig. III-3D**). In this work, we model the average performance of a given BAOC across the 16 combinations of substrates reported by Zahrt, et al. Across the BAOCs considered in the dataset, only two examples are constructed from distinct chiral backbones (13% of the dataset). Accordingly, as most structural changes are concentrated at the modular substituent, *substituent-based* buried volume perfectly recapitulates the reported stereoselectivity ($R^2 = 0.98$, MAE = 0.04 kcal/mol). Even though *active site-based* buried volume cannot reflect the chemical changes with such precision, a useful univariate regression is still observed ($R^2 = 0.70$, MAE = 0.20 kcal/mol). Notably, other representations of *confinement*, such as sterimol, which have been included in multivariate SARs modeling this dataset fail to produce univariate relationships.

Unlike the homogeneous dataset of R-**X**, the dataset R-**XXXVII**[35] describes the application of a small, but diverse array of BAOCs in a multicomponent photochemical transformation (**Fig. III-3E**). Critically, it was observed in the original report that choice of chiral backbone significantly altered stereoselectivity (ref. TRIP-**CPA**, TRIP-**H_8CPA**, TRIP-**SPA** in **Fig. III-3E**). The trend drawn by *active site-based* buried volume seamlessly captures the enantioinduction cooperatively afforded by the chiral backbone and the modular substituents ($R^2 = 0.74$, MAE = 0.13 kcal/mol). Of course, *substituent-based* buried volume cannot reflect such structural changes over the dataset, so it fails to correlate ($R^2 = 0.15$, MAE = 0.24 kcal/mol; **Supplementary Fig. III-7-9**).

- ***Active site-based buried volume retains utility across catalytic manifolds***

The success of *active site-based* buried volume across *significant diversity* datasets supports the expectation that the descriptor will convey the collective consequences of each individual structural feature on stereoinduction. However, it is unclear whether such performance is also evidence of the descriptor's sustained utility across all BAOC reactivity modes. Traditionally, BAOCs were deployed in acid-catalyzed additions to π-bonds (ex. imine additions).[16] Yet, modern

applications of BAOCs have expanded their functional arsenal beyond additions to include stereoselective eliminations and substitutions under a range of conditions (ex. thermal, phase-transfer,[36] and photochemical). If *active site-based* buried volume is truly general, then its utility must be demonstrated across the array of reaction manifolds accessible to BAOCs.

To investigate this, we partitioned the 50+ reaction datasets into three distinct classes corresponding to catalytic activity: (1) acid-mediated additions, (2) acid-mediated substitutions and eliminations, (3) and complex, multicomponent paradigms (**Fig. III-4A**). Indeed, across each of these categories, *active site-based* buried volume furnishes *accurate* and *robust* regressions (*addition* – avg. $R^2 = 0.56$, *elimination and substitution* – avg. $R^2 = 0.63$, and *complex* – avg. $R^2 = 0.52$; **Fig. III-4B**). Other alternative representations of confinement, like sterimol, which have previously reported utility for specific acid-mediated *additions*, fail to accommodate the remaining functional diversity available to the BAOCs (ex. with sterimol_L *addition* – avg. $R^2 = 0.23$, *elimination and substitution* – avg. $R^2 = 0.28$, and *complex* – avg. $R^2 = 0.21$; **Supplementary Fig. III-10-13** for full analyses to other representations of *confinement*).

To highlight the consistent utility of *active site-based* buried volume across reaction manifolds, we showcase one reaction from each category enumerated above. While many acid-mediated *additions* rely on basic functional groups (ex. imines) to strongly bind to the chiral acid active site, stereoselective solutions have emerged for substrates lacking such motifs. One early example of this is R-**XV**, a stereoselective aldol-like addition between an enamide and a glyoxal reported by Terada and coworkers (**Fig. III-4C**).[37] The report realized that traditional stereoselective intuition – *increasing* confinement – does not furnish the desired adduct with selectivity (ref. CPA-**2** with 8% *e.e.*). Instead, fine-tuning confinement at the active site was required to achieve the desired alcohol with high selectivity (ref. CPA-**3** with 98% *e.e.*). *Active site-based* buried volume reflects Terada's counterintuitive findings, furnishing a useful univariate regression model ($R^2 = 0.81$, MAE = 0.35 kcal/mol).

Beyond acid-mediated *additions*, many stereoselective eliminations (ex. atroposelective condensations) and substitutions mediated by BAOCs have been reported. Here, we highlight R-**XLVII**, an intermolecular substitution of azetidine with thiol nucleophiles (**Fig. III-4D**).[38] The authors screen a small (7 ex.) but diverse (3 distinct scaffolds) set of BAOCs to realize a solution (up to 91% *e.e.*). *Active site-based* buried volume not only accommodates the structural heterogeneity of the dataset, but also the BAOC-mediated substitution, yielding an excellent linear model ($R^2 = 0.91$, MAE = 0.20 kcal/mol).

Finally, though BAOCs are typically employed as *chiral acids*, their applications have evolved beyond the scope of simple acid/base chemistry to include applications as ligands in asymmetric transition metal catalysis and chiral counterions (ex. asymmetric counterion directed catalysis, ACDC). For years, our group has investigated stereoselective phase-transfer oxidations mediated by chiral phosphates. One notable example from our group describes a phase-transfer catalyzed Pummerer-type cyclization that had required remote structural changes to enhance the observed

stereoselectivity (R-**XIV**, **Fig. III-4E**)[39]. In our original report, a multivariate regression from a collection of sterimol descriptors was required to reflect the impact of the distant structural changes to stereoselectivity (three terms, $R^2 = 0.96$). Here, we show *active site-based* buried volume alone readily recapitulates the relationship (one term, $R^2 = 0.85$, MAE = 0.06, **Fig. III-4E**). In comparison, *active site-based* sterimol features are unsuitable for univariate relationships (ex. with sterimol L alone, $R^2 = 0.17$, MAE = 0.13, **Fig. III-4E**)

- ***Active site-based buried volume holistically generalizes stereoelectronic effects***

Across the 50+ reaction datasets considered herein, *active site-based* buried volume uniquely reflects *confinement* as a quantitative representation for stereoselectivity by BAOCs, regardless of the underlying structure and function. Yet, *confinement* is often considered a *steric* phenomenon, where models for stereoselectivity invoke steric repulsion as the primary destabilizing criterion. More recently, chemists prefer to invoke attractive non-covalent interactions as a primary design criterion for stereoselectivity. From this perspective, *confinement* qualitatively reflects a propensity for attractive dispersion forces.[22,40,41] Still, a preponderance of other stereoelectronic non-covalent interactions (NCIs) have proven relevant for stereoselective catalysis (ex. π- π, cation- π, etc., **Fig. III-5A**).[13,14,42,43] Is *active site-based* buried volume capable of reflecting such diverse stereoelectronic NCIs?

In theory, electronic contributions to NCIs (ex. cation-π or π-π interactions) are defined from the interaction of charges across a distance (ex. ion-dipole scales with $1/r^2$ where r represents distance between charges, van der Waals attraction scales with $1/r^6$, **Fig. III-5B**).[1,44] It follows that active site-based buried volume, which intimately relates percent steric occupancy across a specific distance (lit. volume), should formally capture such spatial interactions through this relationship. Additionally, it is generally observed that larger π-systems (readily reflected by buried volume) access stronger electrostatic NCIs (ex. cation-π) as a consequence of their greater polarizability.[44–47] In fact, our group has previously shown robust, univariate relationships exist between buried volume and the strength of corresponding, stereoselective cation-π NCIs in some putative, stereo-determining reaction complexes.[48,49] As highlighted in R-**XV** above, *active site-based* buried volume reflects stereoselectivity for a phase-transfer catalyst where stereoinduction was attributed to strengthening C–H---π interactions in the major transition state.[39] Thus, across the datasets considered, *active site-based* buried volume likely reflects an array of stereoselective interactions by BAOCs, including those that are principally electronic in nature.

To ensure that such examples are not merely instances of *cooperative* steric and electronic control, we turned to the literature to identify BAOC-mediated transformations where stereoselectivity was explicitly realized via *electronic* NCIs. Indeed, elucidated with extensive computational work, Wheeler and coworkers have realized such electronic control was observed in dataset R-**XXVII**, a stereoselective condensation furnishing cyclic aminals, whereby biased coulombic interactions across the highly polarized active site drive stereoselectivity (**Fig. III-5C**).[50,51] Notably, this condensation is analogous to two other condensations analyzed herein, R-**XI** and R-**XXII**. As

expected, *substituent-based* electronic representations like ChelpG charge furnish meaningful univariate trends across the collection of transformation operating via electronic control (up to R^2 = 0.79, MAE = 0.18 kcal/mol; **Fig. III-5D**, see **Supplementary Fig. III-15-20** for remaining electronic representations). *Active site-based* buried volume similarly recapitulates these trends (R^2 = 0.75, MAE = 0.13 kcal/mol; **Fig. III-5E**), supporting the expectation that it can reflect principally electronic, stereoselective NCIs.

III.4 – Discussion and Outlook

We have demonstrated *active site-based* buried volume is a uniquely advantageous descriptor for the construction of simple and holistic relationships to stereoselectivity by BAOCs. These catalysts alter stereoinduction through an array of diverse NCIs collectively conferred by many distinct structural elements. As such, there is not one universally shared mechanistic attribute specifically exploited by the most selective catalysts. Still, *active site-based* buried volume navigates this stereoselective heterogeneity of each reaction dataset by standardizing spatial interactions, even across structurally diverse BAOCs. More broadly, intentional investigation of a general chemical representation for *complex functions* (ex. stereoselectivity) across any chemistry may reveal surprisingly simple and intuitive relationships.

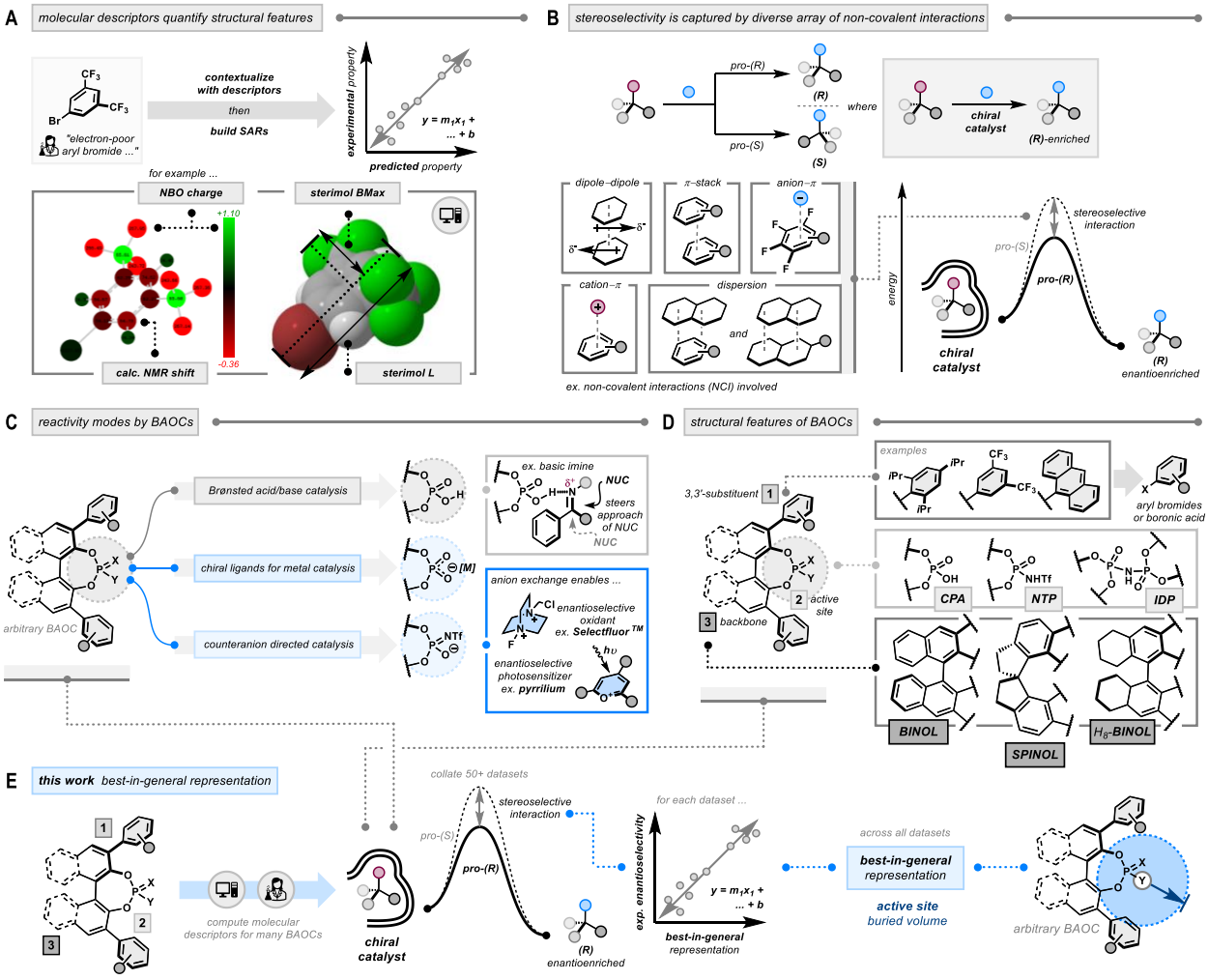


Fig. III-1 | Bronsted acid organocatalysts (BAOCs) navigate an array of non-covalent interactions (NCIs) to furnish stereoselective outcomes across manifold reaction paradigms. **(a)** Molecular descriptors are quantitative representations of chemical structures and are readily deployed to construct structure-activity relationships (SARs). (**b**) Stereoselective catalysts favor the formation of one stereoisomer over another via the collective impact of many diastereoselective NCIs. (**c**) BAOCs access a wide range of reactivity modes. (**d**) BAOCs boast enormous structural tunability over the chiral scaffold surrounding the Brønsted acid active site. (**e**) *This work – active site-based* buried volume emerges as a powerful representation for stereoselectivity by BAOCs from statistical analyses of univariate SARs across many datasets.

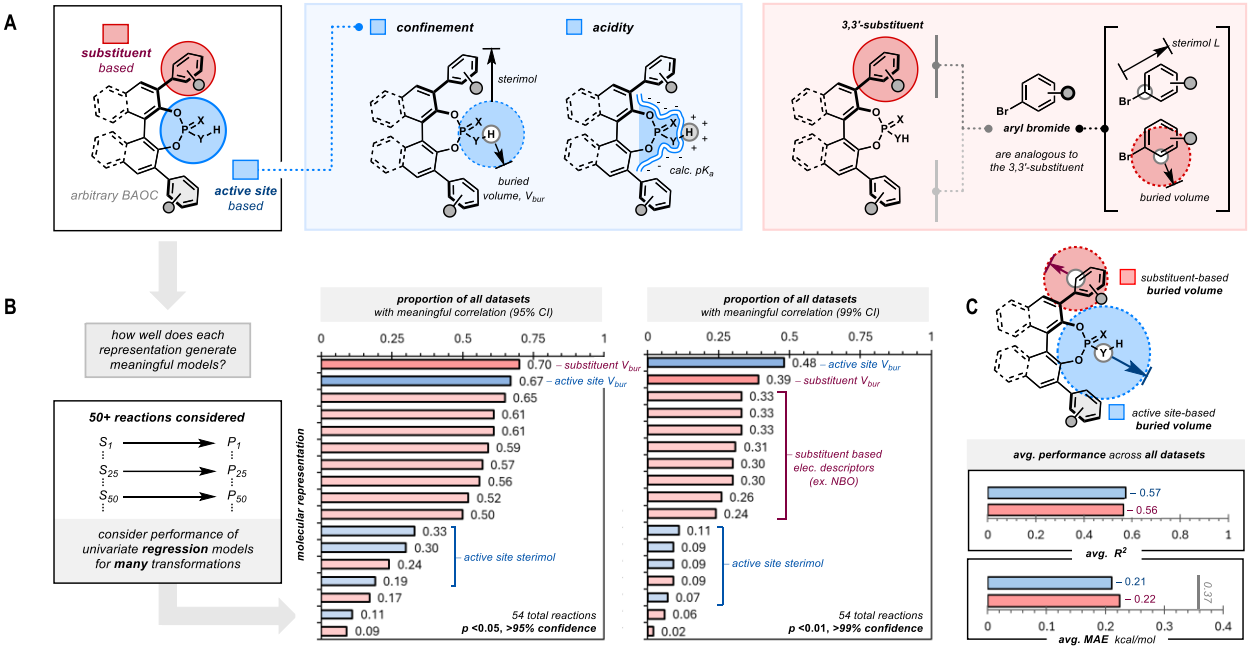


Fig. III-2 | Collation of candidate molecular descriptors and detection of a general representation for stereoselectivity. (a) *Active site-based* representations reflect existing synthetic design criteria (ex. acidity, confinement) without specific attention to underlying scaffold, while *substituent-based* representations are derived from simplified analogs (aryl bromides). (**b**) Proportion of meaningful univariate regression models constructed by each descriptor at increasingly strict confidence intervals (up to 99% confidence) across the set of 50+ reaction datasets considered in this work. (**c**) On average, *active site-based* and *substituent-based* buried volumes construct the most *accurate* and *robust* univariate regression models (in comparison, dummy predictor of *mean* for each dataset furnishes an avg. MAE = 0.37 kcal/mol).

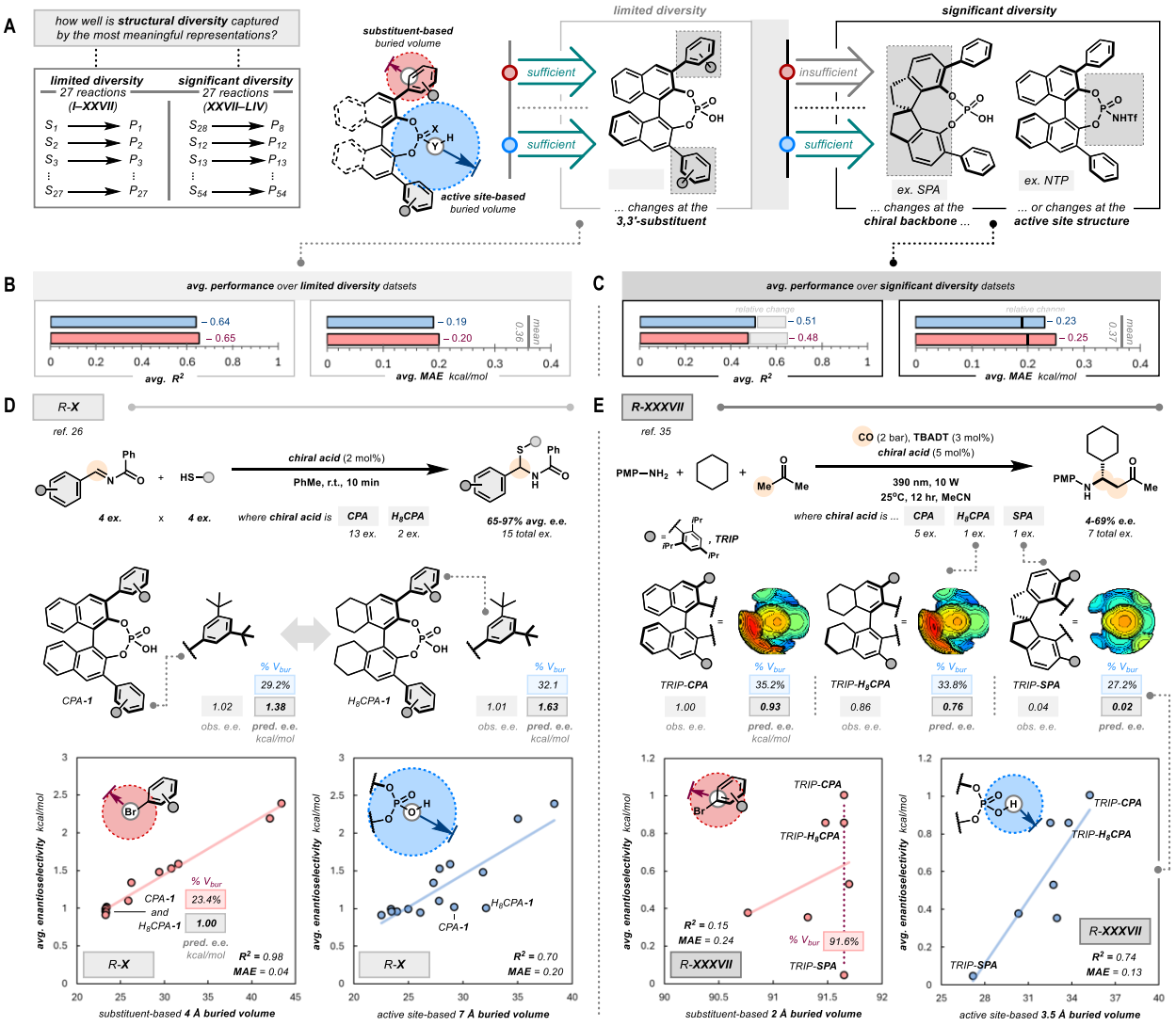


Fig. III-3 | *Active site-based buried volume accommodates substantial structural diversity within datasets*. (a) Organizing the reaction datasets with respect to the structural diversity of the BAOCs will help to qualify the generality of candidate representations. (**b**) *Limited diversity* datasets showcase both *active site-based* and *substituent-based* buried volumes are comparable, while (**c**) *significant diversity* datasets showcase *active site-based* buried volume is comparatively more robust. (**d**) R-**X** from Zahrt, et al.[26] highlights how both *active site-based* and *substituent-based* buried volumes furnish strong trends in a *limited diversity* dataset. (**e**) R-**XXXVII** from Ding, et al.[35] highlights how *active site-based* buried volume furnishes trends across diverse datasets.

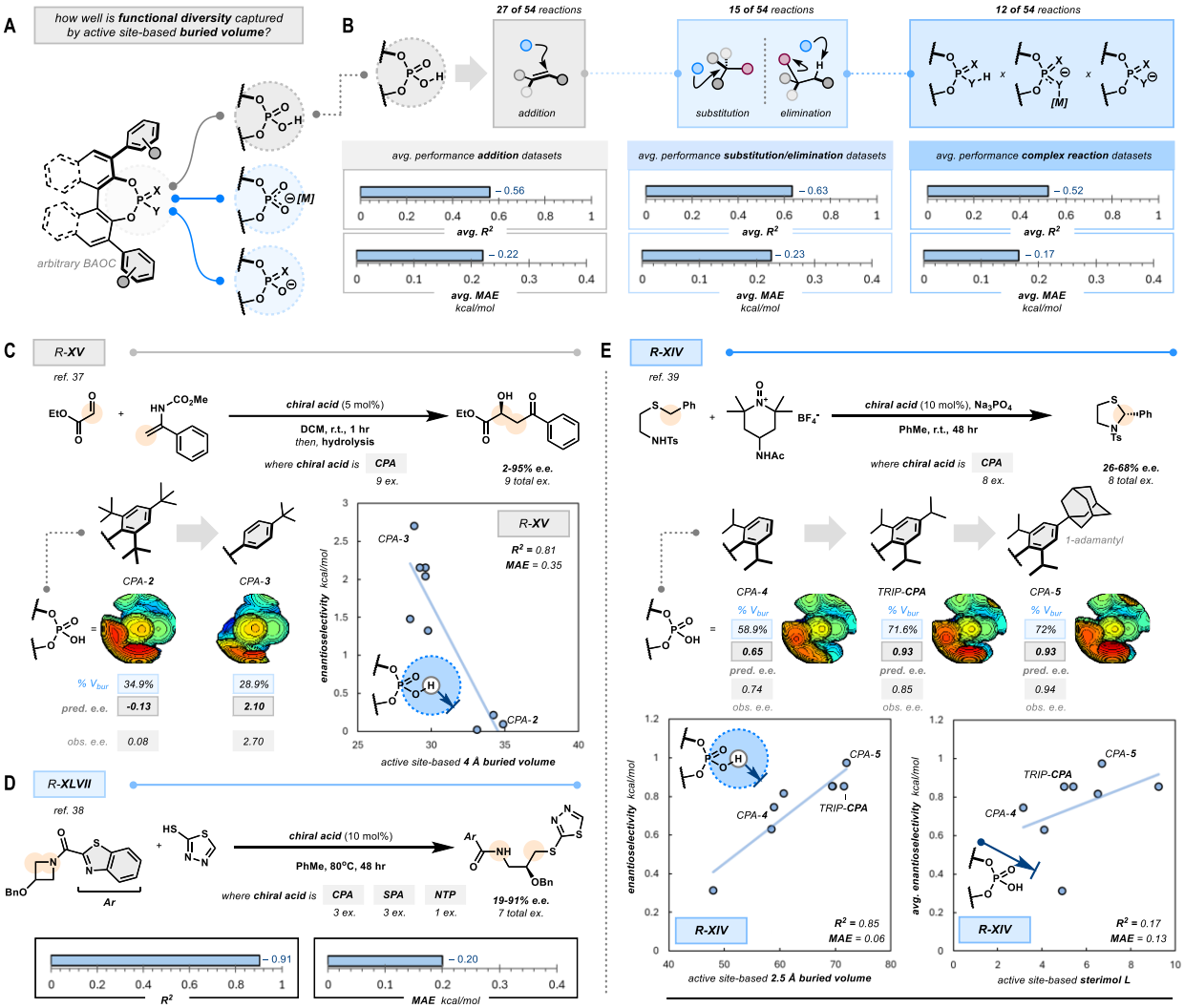


Fig. III-4 | ***Active site-based*** **buried volume retains utility across BAOC reaction manifolds.** **(a)** BAOC reactivity can be organized into three categories reflecting the role of the acid, where (**b**) *active site-based* buried volume continues to furnish *accurate* trends across each category. (**c**) R-**XV** by Terada, et al.[37] highlights how *active site-based* buried volume can reflect counterintuitive relationships for stereoselective additions under confinement. (**d**) R-**XLVII** by Wang, et al.[38] highlights how *active site-based* buried volume can accommodate substantial structural diversity, even within a stereoselective substitution. (**e**) R-**XIV** by Toste, et al.[39] highlights how *active site-based* buried volume serves as a privileged representation for stereoselectivity by BAOCs, even within complex, multicomponent transformations like phase-transfer catalysis.

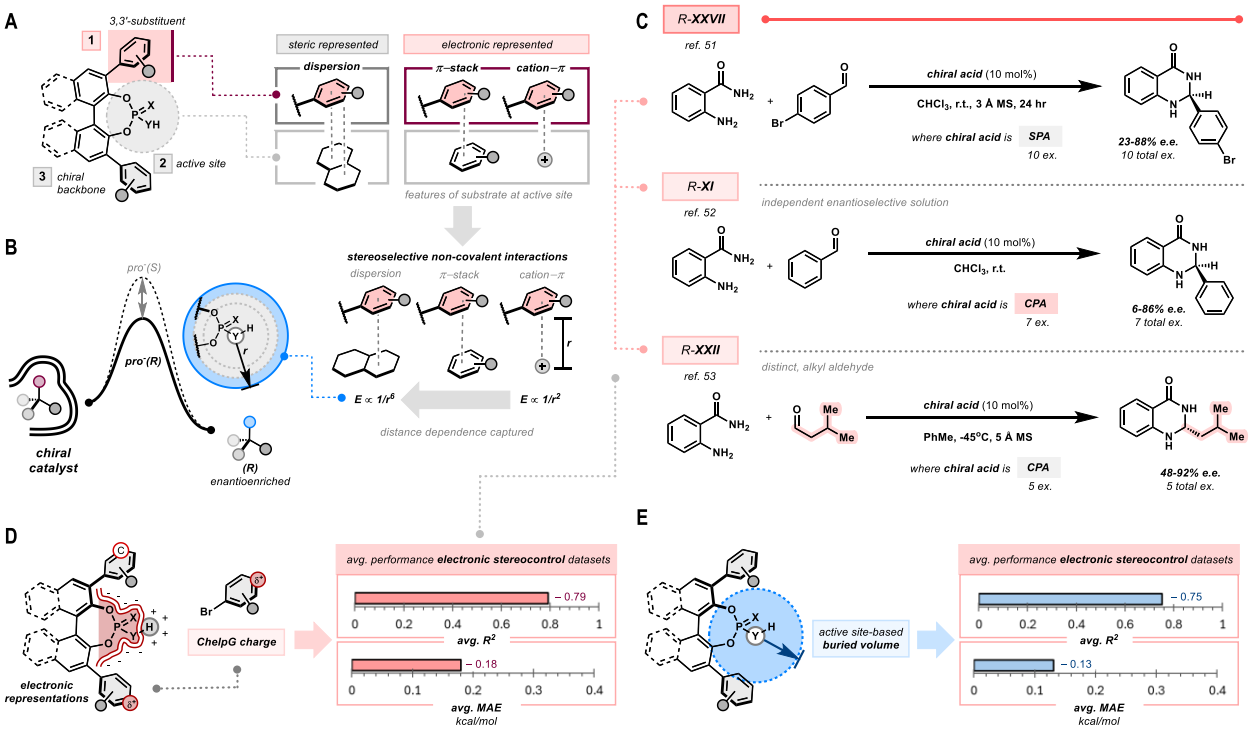


Fig. III-5 | Active site-based buried volume holistically generalizes stereoelectronic effects across reaction datasets considered. (**a**) Stereoselective NCIs can be categorized as predominantly steric or electronic-mediated. (**b**) Interaction energies of various stereoselective NCIs relate to distance (denoted *r*) which is reflected by *active site-based* buried volume. (**c**) R-**XXVII** is a dataset reported by Wheeler, et al.[51] to operate via electronic stereocontrol. R-XI[52] and R-**XXII**[53] transformations likely operate similarly by analogy to R-**XXVII**. (**d**) Electronic representations like *substituent-based* ChelpG charge furnish excellent trends (averaged across R-**XXVII**, R-**XI**, and R-**XXII**). (**e**) *Active site-based* buried volume also furnishes excellent univariate trends (averaged across R-**XXVII**, R-**XI**, and R-**XXII**).

III.5 – References

(1) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(2) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, *17* (1), 125–136.

(3) Taft, R. W. Jr. Linear Steric Energy Relationships. *J. Am. Chem. Soc.* **1953**, *75* (18).

(4) Charton, M. Steric Effects. I. Esterification and Acid-Catalyzed Hydrolysis of Esters. *J. Am. Chem. Soc.* **1975**, *97* (6), 1552–1556.

(5) Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* **2014**, *507* (7491), 210–214.

(6) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9* (3), 2313–2323.

(7) Weinhold, F. Natural Bond Orbital Analysis: A Critical Overview of Relationships to Alternative Bonding Perspectives. *Journal of Computational Chemistry* **2012**, *33* (30), 2363–2379.

(8) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7* (10), 1622–1637.

(9) B. Santiago, C.; Guo, J.-Y.; S. Sigman, M. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chemical Science* **2018**, *9* (9), 2398–2412.

(10) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54* (4), 827–836.

(11) Shalit Peleg, H.; Milo, A. Small Data Can Play a Big Role in Chemical Discovery. *Angewandte Chemie International Edition* **2023**, *62* (26).

(12) Haas, B. C.; Kalyani, D.; Sigman, M. S. Applying Statistical Modeling Strategies to Sparse Datasets in Synthetic Chemistry. *Science Advances* **2025**, *11* (1), eadt3013.

(13) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543* (7647), 637–646.

(14) Toste, F. D.; Sigman, M. S.; Miller, S. J. Pursuit of Noncovalent Interactions for Strategic Site-Selective Catalysis. *Acc. Chem. Res.* **2017**, *50* (3), 609–615.

(15) Akiyama, T. Stronger Brønsted Acids. *Chem. Rev.* **2007**, *107* (12), 5744–5758.

(16) Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M. Complete Field Guide to Asymmetric BINOL-Phosphate Derived Brønsted Acid and Metal Catalysis: History and Classification by Mode of Activation; Brønsted Acidity, Hydrogen Bonding, Ion Pairing, and Metal Phosphates. *Chem. Rev.* **2014**, *114* (18), 9047–9153.

(17) Chen, D.-F.; Han, Z.-Y.; Zhou, X.-L.; Gong, L.-Z. Asymmetric Organocatalysis Combined with Metal Catalysis: Concept, Proof of Concept, and Beyond. *Acc. Chem. Res.* **2014**, *47* (8), 2365–2377.

(18) Mahlau, M.; List, B. Asymmetric Counteranion-Directed Catalysis: Concept, Definition, and Applications. *Angewandte Chemie International Edition* **2013**, *52* (2), 518–533.

(19) Handjaya, J. P.; Patankar, N.; Reid, J. P. The Diversity and Evolution of Chiral Brønsted Acid Structures. *Chemistry – A European Journal* **2024**, *30* (46).

(20) Betinol, I. O.; Kuang, Y.; Mulley, B. P.; Reid, J. P. Controlling Stereoselectivity with Noncovalent Interactions in Chiral Phosphoric Acid Organocatalysis. *Chem. Rev.* **2025**, *125* (8), 4184–4286.

(21) Akiyama, T.; Mori, K. Stronger Brønsted Acids: Recent Progress. *Chem. Rev.* **2015**, *115* (17), 9277–9306.
(22) Mitschke, B.; Turberg, M.; List, B. Confinement as a Unifying Element in Selective Catalysis. *Chem* **2020**, *6* (10), 2515–2532.
(23) Schreyer, L.; Properzi, R.; List, B. IDPi Catalysis. *Angew. Chem. Int. Ed.* **2019**, *58* (37), 12761–12777.
(24) Hillier, A. C.; Sommer, W. J.; Yong, B. S.; Petersen, J. L.; Cavallo, L.; Nolan, S. P. A Combined Experimental and Theoretical Study Examining the Binding of N-Heterocyclic Carbenes (NHC) to the Cp*RuCl (Cp* = H5-C5Me5) Moiety: Insight into Stereoelectronic Differences between Unsaturated and Saturated NHC Ligands. *Organometallics* **2003**, *22* (21), 4322–4326.
(25) Antinucci, G.; Dereli, B.; Vittoria, A.; Budzelaar, P. H. M.; Cipullo, R.; Goryunov, G. P.; Kulyabin, P. S.; Uborsky, D. V.; Cavallo, L.; Ehm, C.; Voskoboynikov, A. Z.; Busico, V. Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction. *ACS Catal.* **2022**, *12* (12), 6934–6945.
(26) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424).
(27) Wakchaure, V. N.; DeSnoo, W.; Laconsay, C. J.; Leutzsch, M.; Tsuji, N.; Tantillo, D. J.; List, B. Catalytic Asymmetric Cationic Shifts of Aliphatic Hydrocarbons. *Nature* **2024**, *625* (7994), 287–292.
(28) Reid, J. P.; Sigman, M. S. Holistic Prediction of Enantioselectivity in Asymmetric Catalysis. *Nature* **2019**, *571* (7765), 343–348.
(29) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, *347* (6223), 737–743.
(30) Verloop, A.; Hoogenstraaten, W.; Tipker, J. Chapter 4 - Development and Application of New Steric Substituent Parameters in Drug Design. In *Drug Design*; Ariëns, E. J., Ed.; Medicinal Chemistry: A Series of Monographs; Academic Press: Amsterdam, 1976; Vol. 11, pp 165–207.
(31) Kariofillis, S. K.; Jiang, S.; Żurański, A. M.; Gandhi, S. S.; Martinez Alvarado, J. I.; Doyle, A. G. Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources. *J. Am. Chem. Soc.* **2022**, *144* (2), 1045–1055.
(32) Peng, B.; Ma, J.; Guo, J.; Gong, Y.; Wang, R.; Zhang, Y.; Zeng, J.; Chen, W.-W.; Ding, K.; Zhao, B. A Powerful Chiral Super Brønsted C–H Acid for Asymmetric Catalysis. *J. Am. Chem. Soc.* **2022**, *144* (7), 2853–2860.
(33) Dalmau, D.; Sigman, M. S.; Alegre-Requena, J. V. Machine Learning Workflows beyond Linear Models in Low-Data Regimes. *Chem. Sci.* **2025**.
(34) *Model Selection and Multimodel Inference*; Burnham, K. P., Anderson, D. R., Eds.; Springer: New York, NY, 2004.
(35) Ding, W.-W.; He, Z.-Y.; Sayed, M.; Zhou, Y.; Han, Z.-Y.; Gong, L.-Z. Enantioselective Synthesis of β- and α-Amino Ketones through Reversible Alkane Carbonylation. *Nat. Synth* **2024**, *3* (4), 507–516.
(36) Rauniyar, V.; Lackner, A. D.; Hamilton, G. L.; Toste, F. D. Asymmetric Electrophilic Fluorination Using an Anionic Chiral Phase-Transfer Catalyst. *Science* **2011**, *334* (6063), 1681–1684.

(37) Terada, M.; Soga, K.; Momiyama, N. Enantioselective Activation of Aldehydes by Chiral Phosphoric Acid Catalysts in an Aza-Ene-Type Reaction between Glyoxylate and Enecarbamate. *Angewandte Chemie International Edition* **2008**, *47* (22), 4122–4125.
(38) Wang, Z.; Sheong, F. K.; Sung, H. H. Y.; Williams, I. D.; Lin, Z.; Sun, J. Catalytic Enantioselective Intermolecular Desymmetrization of Azetidines. *J. Am. Chem. Soc.* **2015**, *137* (18), 5895–5898.
(39) Biswas, S.; Kubota, K.; Orlandi, M.; Turberg, M.; Miles, D. H.; Sigman, M. S.; Toste, F. D. Enantioselective Synthesis of N,S-Acetals by an Oxidative Pummerer-Type Transformation Using Phase-Transfer Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (2), 589–593.
(40) Eschmann, C.; Song, L.; Schreiner, P. R. London Dispersion Interactions Rather than Steric Hindrance Determine the Enantioselectivity of the Corey–Bakshi–Shibata Reduction. *Angewandte Chemie International Edition* **2021**, *60* (9), 4823–4832.
(41) Leinung, W.; Mitschke, B.; Leutzsch, M.; Wakchaure, V. N.; Maji, R.; List, B. From Reagent to Catalyst: Dispersion-Driven Design of a General Asymmetric Transfer Hydrogenation Catalyst. *J. Am. Chem. Soc.* **2025**.
(42) Knowles, R. R.; Lin, S.; Jacobsen, E. N. Enantioselective Thiourea-Catalyzed Cationic Polycyclizations. *J. Am. Chem. Soc.* **2010**, *132* (14), 5030–5032.
(43) Wheeler, S. E. Local Nature of Substituent Effects in Stacking Interactions. *J. Am. Chem. Soc.* **2011**, *133* (26), 10262–10274.
(44) Dougherty, D. A. The Cation−π Interaction in Chemistry and Biology. *Chem. Rev.* **2025**, *125* (5), 2793–2808.
(45) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
(46) Vijay, D.; Sastry, G. N. Exploring the Size Dependence of Cyclic and Acyclic π-Systems on Cation–π Binding. *Phys. Chem. Chem. Phys.* **2008**, *10* (4), 582–590.
(47) Chen, F.; Chen, Y.; Chang, X.-Y.; He, D.; Yang, Q.; Wang, D. Z.; Xu, C.; Yu, P.; Xing, X. Polarizability Matters in Enantio-Selection. *Nat Commun* **2024**, *15* (1), 3394.
(48) Treacy, S. M.; Smith, A. L.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science. *J. Am. Chem. Soc.* **2024**, *146* (43), 29792–29800.
(49) Heafner, E. D.; Smith, A. L.; Craescu, C. V.; Raymond, K. N.; Bergman, R. G.; Toste, F. D. Probing Enantioinduction in Confined Chiral Spaces through Asymmetric Oxime Reductions. *Chem* **2025**, *11* (4), 102368.
(50) Maji, R.; Mallojjala, S. C.; Wheeler, S. E. Electrostatic Interactions in Asymmetric Organocatalysis. *Acc. Chem. Res.* **2023**, *56* (14), 1990–2000.
(51) Laconsay, C. J.; Seguin, T. J.; Wheeler, S. E. Modulating Stereoselectivity through Electrostatic Interactions in a SPINOL-Phosphoric Acid-Catalyzed Synthesis of 2,3-Dihydroquinazolinones. *ACS Catal.* **2020**, *10* (20), 12292–12299.
(52) Rueping, M.; Antonchick, A. P. Catalytic Asymmetric Aminoallylation of Aldehydes: A Catalytic Enantioselective Aza-Cope Rearrangement. *Angewandte Chemie International Edition* **2008**, *47* (52), 10090–10093.
(53) Cheng, X.; Vellalath, S.; Goddard, R.; List, B. Direct Catalytic Asymmetric Synthesis of Cyclic Aminals from Aldehydes. *J. Am. Chem. Soc.* **2008**, *130* (47), 15786–15787.

Chapter IV – **Abstracting complementarity for the design of stereoselective acid organocatalysts**

Portions of this chapter were adapted from:

Smith, A. L.; Le, C.; Toste, F. D. *Manuscript in revision.*

IV.1 – Preface

Complementarity is invoked to holistically characterize interactions between a catalyst and its substrate. Although the paradigm maintains utility across many fields of chemistry, it is challenging to proactively deploy, requiring *post hoc* analyses to demonstrate. Here, we report active site-based buried volume furnishes a quantitative and generalizable representation for complementarity by Brønsted acid organocatalysts. The low-dimensional, geometric abstraction helps construct analytical Gaussian models encoding *complementarity.* From these footholds, experimental validation realizes new stereoselective solutions across diverse structural motifs and functional modes. Unlike other heuristics for stereoinduction, active site-based buried volume guides the design of stereoselective solutions without requiring detailed attention to the mechanisms of stereoinduction. We anticipate the successful abstraction of high-dimensional, stereoselective interactions by *complementarity* into a single, low-dimensional descriptor will offer a systematic framework for future catalyst design.

IV.2 – Introduction

Selectivity in catalysis has long been attributed to *complementarity* between catalyst active site and substrate. As the name implies, *complementarity* reflects an *ideal* fit between the two species. Historically, the phenomenon characterized the specificity of enzymes where their activity relates to *binding to the transition state stronger than the ground state*.[1,2] Yet this idea is not exclusive to homogeneous systems. Even heterogeneous catalysts (ex. zeolites) leverage *complementarity*-like intuition to justify selective outcomes from a collection of possible products.[3,4] In general, *complementarity* is principally attributed to the selective *stabilization* of matched intermediate(s) and/or *destabilization* of others via an array of non-covalent interactions (NCIs, **Fig. IV-1A**).[5,6] Each NCI displays varying dependence on distance, where close interactions are disfavored due to Pauli repulsion while far interactions lose electronic coupling. Interactions at *just the right* distance maximize stabilization (**Fig. IV-1B**).[7–9] The collective action of many interactions culminates in the observed *complementarity*.[2] As the field has evolved to consider increasingly complex chemical phenomena, the paradigm has remained a unifying criterion for the design of catalysts achieving chemo-, regio-, and stereoselectivity.[4,10–13]

Although *complementarity* promises an intuitive platform for catalyst design, molecular representations for the paradigm are not well-formed. Accordingly, construction of *complementary* catalysts is rarely predictive. Instead, the principle is often only inferred by elegant, *post hoc* rationalizations drawn from rigorous mechanistic and computational investigations.[14–18] These challenges are especially persistent in the context of stereoselectivity. Within a single transformation, structurally analogous catalysts can display disparate stereoinduction. Without clear avenues to reflect *complementarity* for stereoselectivity, subsequent structure-activity relationships (SARs) drawn can require high-dimensional models (ex. multivariate regression, machine learning-based models) comprised of case-specific molecular features with only circumstantial relevance.[6,19–24] Simple, qualitative heuristics like quadrant diagrams or

stereochemical models can emerge, but they lack quantitative insight. As a result, their persistent utility is quite rare whereby only a few instances are routinely engaged by chemists (ex. Sharpless' dihydroxylation,[25] Evans' auxiliary,[26] Ellman's sulfinamide[27]).

Identifying a molecular representation that clearly reflects *complementarity* would empower chemists to proactively leverage the principle for catalyst design. Such a representation must generalize across structural features and functional applications of the catalyst, escaping any bespoke relationships (**Fig. IV-1C**). In the context of stereoselectivity, our group has recently reported active site-based buried volume is a robust and holistic representation for diverse sets of Brønsted acid organocatalysts (BAOCs).[28] These catalysts are constructed from myriad structural motifs and consequentially access a wide range of functional modes (ex. acid/base, cooperative transition metal catalysis, etc.) enabling chemists with asymmetric access to ever-expanding regions of chemical space.[4,29–31] As active site-based buried volume is a low-dimensional geometric abstraction of the catalyst active site, we wondered whether the representation can reflect *complementarity*.[32,33] If so, models constructed with *active site-based* buried volume alone should anticipate stereoselective BAOCs.

In this work, we show active site-based buried volume quantitatively reflects *complementarity,* transforming the fundamental principle into a practical design tool (**Fig. IV-1D**). We experimentally demonstrate models constructed from active site-based buried volume can analytically represent catalysts with optimal *complementarity*. Furthermore, we demonstrate the representation can transfer *complementarity* across disparate scaffolds, transcending traditional biases surrounding stereoinduction. Altogether, this simple geometric abstraction readily steers chemists to ideal regions of catalyst chemical space, furnishing a framework for the design of selective catalysts.

IV.3 – Results

- ***Modeling complementarity with active site-based buried volume***

Complementarity by a stereoselective catalyst suggests the catalyst active site selectively *stabilizes* one diastereomeric transition state and/or *destabilizes* the other.[2] Such a phenomenon implies a reaction-specific optimum, pairing each substrate with a potentially unique catalyst. Catalysts approaching *complementarity* are expected to furnish increasingly selective outcomes, while those diverging should perform relatively worse (**Fig. IV-2A**).

Previously, we have reported active site-based buried volume generally furnishes robust univariate linear regressions to stereoselectivity by BAOCs (50+ reaction datasets). The original report described the representation as a reflection of BAOC *confinement* – a quality describing the relative constriction of an active site and typically attributed to stereoselectivity. Based on the theoretical implications of *complementarity* discussed above, one single *confinement* cannot be uniquely stereoselective across *all* reactions. To validate this, the distribution of the most stereoselective BAOCs was plotted against all BAOCs calculated in the 50+ datasets considered

herein (see **Appendix III** and **Appendix IV** for dataset details). Immediately, it can be recognized that no specific BAOC confinement is uniquely stereoselective (**Fig. IV-2B**, max. Kolmogorov–Smirnov statistic = 0.23, $p = 0.08$).

This analysis implies there are reaction-specific regions of chemical space occupied by *complementary* stereoselective catalysts. As we approach such regions, BAOCs should be increasingly stereoselective. In our original report, we detected these gradients with linear regression. However, if BAOCs deviate from the *complementary* solution, both more and less confined catalysts should perform worse. This behavior critically implies a nonlinear and non-monotonic modality. Although many of the datasets previously considered could be meaningfully modeled by simple linear regression, physically meaningful fits accommodating this *complementary* behavior would be necessarily advantageous (**Fig. IV-2C** through **Fig. 3-2F**). [34,35] In fact, based on R^2, mean absolute error (MAE), leave-one-out cross validation (LOO-CV), and Akaike information criteria (AIC), Gaussian regression furnishes the most meaningful fits across all datasets relative to other linear and non-linear regression-based analyses (Gaussian, avg. R^2 = 0.72, avg. MAE = 0.16 kcal/mol; see **Appendix IV** – **Detailed Statistical Analyses**, **Supplementary Fig. IV-2-56** and **Supplementary Table IV-3-7**).

Well-populated datasets exhibiting the hallmarks of *complementarity* showcase the unique utility of Gaussian regression. We highlight two such datasets here – one reflecting a stereoselective condensation and one reporting access to chiral-at-*N* molecules. R-**I** describes a stereoselective condensation furnishing a chiral spirocycle (n = 11, **Fig. IV-2G**).[36] After screening a diverse array of BAOCs, the authors realized only a moderate degree of confinement was required to confer stereoselectivity in the condensation (ref. **CPA-1**). Further increases to confinement only eroded the observed selectivity (ref. **Anth-NTP**). This sensitive evolution of *complementarity* across the wide net of BAOCs screened is excellently captured by Gaussian regression with active site-based buried volume (Gaussian, $R^2 = 0.94$, MAE = 0.15 kcal/mol).

R-**II** describes access to chiral-at-*N* stereocenters from BAOC-mediated oxidation and substitution (n = 12, **Fig. IV-2H**).[37] The authors realized the stereoselectivity from two distinct BAOC scaffolds – chiral phosphoric acids (CPA) and spirophosphoric acids (SPA). The former is a traditional catalyst, but familiar analogs only furnished the *N*-heterocycle in low enantioenrichment (ex. **TCYP-CPA**, obs. 7% *e.e.*). Instead, fine increases to confinement at the active site of SPAs were key to realize stereoselectivity. While sterically bulky CPA-based substituents harmed selectivity, sterically bulky SPA-based substituents were essential (ex. TCYP-**SPA**, obs. 78% *e.e.*). Gaussian regression with active site-based buried volume readily accommodates this emergent *complementarity* (Gaussian, $R^2 = 0.80$, MAE = 0.11 kcal/mol).

- ***Transferring stereoselective solutions across disparate, unreported scaffolds***

The success of *complementarity* to inform stereoselective solutions above supports the expectation that this criterion conveys the collective consequence of each structural element on stereoinduction. However, each of the above datasets considered only explores stereoselective

solutions within analogous scaffolds, where stereoselective NCIs are primarily modulated by altering the NCIs from 3,3'-substituents decorating analogous backbones (**Fig. IV-3A**).[6] This design is so common that empirical discovery often exclusively relies on evaluation of BAOCs solely modulated at this position. After all, simple heuristics like quadrant diagrams are frequently deployed for qualitative stereochemical predictions (**Fig. IV-3B**). Yet, these patterns require specific catalyst architectures to maintain utility. Does the low-dimensional abstraction of *complementarity* continue to reflect selectivity even when disparate catalysts (ex. without 3,3'-substituents) are under consideration?

To investigate whether the representation of *complementarity* transcends structural biases of canonical BAOCs, we considered an alternative architecture critically lacking 3,3'-substituents along the chiral backbone. This new BAOC is constructed from a *S*-aryl sulfonamide (SIA) with two flanking phosphoryl groups (denoted as P2-SIA). Now, modular substitution of the aryl group at *S* furnishes a diverse array of P2-SIAs. We report divergent access to 13 P2-SIAs prepared in three steps from trityl NSO (**Fig. IV-3C**).

Equipped with the previously unreported P2-SIAs, we first validated whether relationships to active site-based buried volume and stereoselectivity could be observed with datasets including traditional BAOCs like CPAs and NTPs (R-**III** through R-**VII**, see **Appendix IV – Experimental Benchmarks and Datasets**). In total, five benchmark transformations were evaluated with P2-SIAs, reflecting the reaction diversity present across all datasets considered, including: thermal or photochemical reactions, additions or eliminations, and inter- or intramolecular reactions (avg. benchmark dataset size = 18 ex. with up to 13 ex. of P2-SIA).[38–42] Indeed, across these benchmarks, we observe the P2-SIA-containing datasets are well represented by Gaussian regression constructed with active site-based buried volume (**Fig. IV-3D**; Gaussian, avg. $R^2 = 0.72$, avg. MAE = 0.17 kcal/mol).

To assess if *complementarity* truly generalizes across each scaffold, we revisited whether active site-based buried volume encodes relationships between P2-SIA and more familiar BAOCs like CPA and NTP. On inspection, P2-SIAs occupy a complementary region of chemical space unpopulated relative to traditional scaffolds (**Fig. IV-3E**). As such, we expect in some reactions that P2-SIAs are more *complementary* compared to one scaffold or another. By extension, these P2-SIAs would be advantageously stereoselective.

This is critically the case in R-**III**. Here, Yamamoto, et al. observed the privileged analog **TRIP-NTP** furnishes an enantioselective protonation of silyl enol ether in modest stereoselectivity (rep. 54% *e.e.*, obs. 60% *e.e.*), while the privileged **TRIP-CPA** is unsuitable for the transformation (reported n.d., obs. 8% *e.e.*).[38] First, we probed whether Yamamoto's **TRIP-NTP** was *complementary* by screening an additional five NTPs populating the surrounding regions. It was observed both more and less confined BAOCs were less stereoselective compared to Yamamoto's reported solution (Gaussian, *train* $R^2 = 0.77$, *train* MAE = 0.10 kcal/mol). Over this same domain, P2-SIAs populate the chemical space occupied by less confined NTPs but still modestly approach

TRIP-NTP. This analysis anticipates that the closest P2-SIA (P2-**5**) is the *next* most stereoselective (pred. 36% *e.e.*), exceeding all other BAOCs. Consistent with the model, P2-**5** affords the stereoselective protonation in superior stereoselectivity to the remaining dataset (obs. 36% *e.e.*, next most stereoselective obs. 30% *e.e.*). Moreover, the performance of the remaining seven P2-SIAs is very well-aligned with predictions made by the NTP-and-CPA-trained model (*test* MAE = 0.05 kcal/mol), highlighting the transferability of the representation for *complementarity*, even across disparate catalytic scaffolds lacking familiar structural motifs.

- ***Complementarity anticipates new stereoselective solutions***

The application of Gaussian regression models constructed with *active site-based* buried volume promises the discovery of *complementary* BAOCs. If *complementarity* is to be a practical tool, then it must be proactively leveraged to predict previously unknown stereoselective solutions (**Fig. IV-4A**). Yet, at the outset of catalyst discovery, chemists often turn to *privileged* solutions. In the context of BAOCs, *privileged* solutions drastically minimize the number of 3,3'-substituents under consideration. In fact, across the 50+ datasets considered herein, 2,4,6-triisopropylphenyl (TRIP), 9-anthracenyl, and 3,5-bistrifluoromethyl are the most popular, deployed in nearly 80% of all the reactions (note – the average substituent is present in less than 10%, **Fig. IV-4B**). We wondered whether Gaussian regression trained using only BAOCs from these *empirical* choices accurately anticipates if other, less popular BAOCs are more stereoselective. To this end, we deployed Gaussian regression from empirical data to nominate *complementary* regions of chemical space. Then, we leveraged this insight to identify new stereoselective BAOCs.

As proof-of-concept, we revisited R-**I** from **Fig. IV-2F**. In this dataset, a unique anthracene-derived substituent was designed to furnish the condensation in high stereoselectivity (**CPA-1**, rep. 94% *e.e.*), exceeding other pouplar, anthracenyl substituted BAOCs (**Anth-CPA, Anth-NTP**, and **Anth-SPA**, rep. up to 78% *e.e.*, 1.51 kcal/mol). We trained a Gaussian regression model solely on the *empirically popular* BAOCs – those constructed with the above, high-frequency 3,3'-substituents (n = 5, *train* R^2 = 0.97, *train* MAE = 0.06 kcal/mol, **Fig. IV-4C**). Still, among the other 6 BAOCs, the model predicts **CPA-1** to be the most stereoselective (pred. 89% *e.e.*, 2.05 kcal/mol), expecting to exceed the training set's best instance by +0.54 kcal/mol. This prediction is in strong agreement with the reported outcome (rep. 94% *e.e.*, 2.54 kcal/mol). Altogether, the test set is well-represented by the *empirical* model (*test* R^2 = 0.64, *test* MAE = 0.48 kcal/mol).

Motivated by this initial success, we sought to apply this approach to identify stereoselective candidates in a catalyst optimization campaign. Building on our prior report of an enantioselective allenoate–Claisen rearrangement,[43] we hypothesized that propargylic esters could function as conjugate acceptors with allyl amines to generate analogous reactive intermediates (**Fig. IV-4D**, see **Appendix IV – Empirical Bias Powers Disocvery of Stereoselective Solutions** for experimental details). Initial evaluation revealed that the privileged **TRIP-CPA** catalyzed the desired transformation with high diastereoselectivity and moderate enantioselectivity (obs. >20:1 *d.r.* and 53% *e.e.*). However, screening of other empirically successful catalysts, such as **TCYP-**

CPA, resulted in only marginal improvements in selectivity. To guide further optimization, we trained a Gaussian regression model on the compiled dataset (*train* R^2 = 0.85, *train* MAE = 0.093 kcal/mol). The model predicted a narrow region of opportunity between **TRIP-CPA** and **TCYP-CPA**. Systematic evaluation of candidate 3,3′-substituents, informed by computed buried volume parameters, identified a single promising CPA substituent – 2,4,6-tricyclopentylphenyl (pred. 61% *e.e.*). Notably, this motif appeared in only two examples within the 50+ datasets considered herein. Synthesis of the BAOC (**CPA-6**) and experimental evaluation confirmed the prediction, delivering an optimal result (obs. 67% *e.e.*, +0.12 kcal/mol over training set). In contrast, the analogous 2,4,6-tricycloheptylphenyl substituted (**CPA-7**) only displayed attenuated selectivity as predicted by the model (pred. 38% *e.e.,* obs. 41% *e.e.*). Additional derivatives resembling **TRIP-CPA** were designed and prepared, but none were expected nor observed to improve over **CPA-6**. Across the BAOCs prepared, good agreement with the trained Gaussian model was observed (*test* R^2 = 0.71, *test* MAE = 0.103 kcal/mol).

Finally, we aimed to leverage *complementarity* predicted from training data lacking extraordinarily popular BAOCs. Here, we investigated a BAOC-promoted photochemical [2+2] originally reported by Yoon, et al. (**Fig. IV-4E**, see **Appendix IV – Experimental Benchmarks and Datasets** for experimental details).[41,44] The authors quickly realized a suitable solution based on the acidic *N*-triflylphosphoramide (NTP) scaffold with **NTP-8** (obs. 96% *e.e.*). Accordingly, the authors attributed the success to stronger NCIs afforded by the stronger acid (rel. to **CPA-8**, obs. 90% *e.e.*). Critically, this logic does not distinguish the role of the auxiliary 3,3'-substituents which also modulates the observed selectivity. So, we looked to transform the selectivity-determining logic with the principle of *complementarity*. From the narrow foothold of Yoon's reported examples, a Gaussian curve can be traced, revealing the chemical space with opportunities for stereoselectivity. We then prepared an additional six BAOCs spanning the domains around this region and evaluated their performance in Yoon's photochemical [2+2]. Notably, pyrene-substituted NTP (**NTP-9**) occupies the region anticipated to be advantageously *complementary*. In fact, **NTP-9** displayed superior selectivity (obs. 98% *e.e.*, 1.83 kcal/mol), surpassing Yoon's previous solution (obs. 96% *e.e.,* 1.54 kcal/mol – improved by +0.29 kcal/mol). Likewise, although the initial dataset was small, good agreement with predictions was observed across all six examples (*test* MAE = 0.60 kcal/mol).

Gaussian regression with active site buried volume accurately anticipates stereoselective solutions, highlighting experimental opportunities to reveal optimized solutions over training data. Whereas three examples showcasing this are highlighted above, an additional seven analyses are detailed in **Appendix IV** (**Supplementary Fig. IV-64-70**). To summarize, we observed the Gaussian regression models solely trained on empirically popular catalysts readily identified more stereoselective solutions from substantially less popular choices (avg. improvement over training data = +0.45 kcal/mol). Beyond predicting *best-in-class* stereoselective choices, the trained models also generalized well across the less popular, *test set* catalysts (avg. *test* R2 = 0.73, avg. *test* MAE = 0.25 kcal/mol).

IV.4 – Discussion and Outlook

Complementarity has promised a fundamental paradigm for the design of stereoselective catalysts, yet its realization has long been restricted to retrospective analyses. Here, we have shown that a low-dimensional geometric abstraction of catalyst active sites can reflect *complementarity* and render it operational. Across a wide array of acid-catalyzed reactions, the representation was deployed to anticipate stereoselective solutions from both familiar and previously unreported BAOC scaffolds (ex. P2-SIA). By reframing *complementarity* from a qualitative principle into a quantifiable descriptor, the discovery of new stereoselective catalysts can be approached systematically.

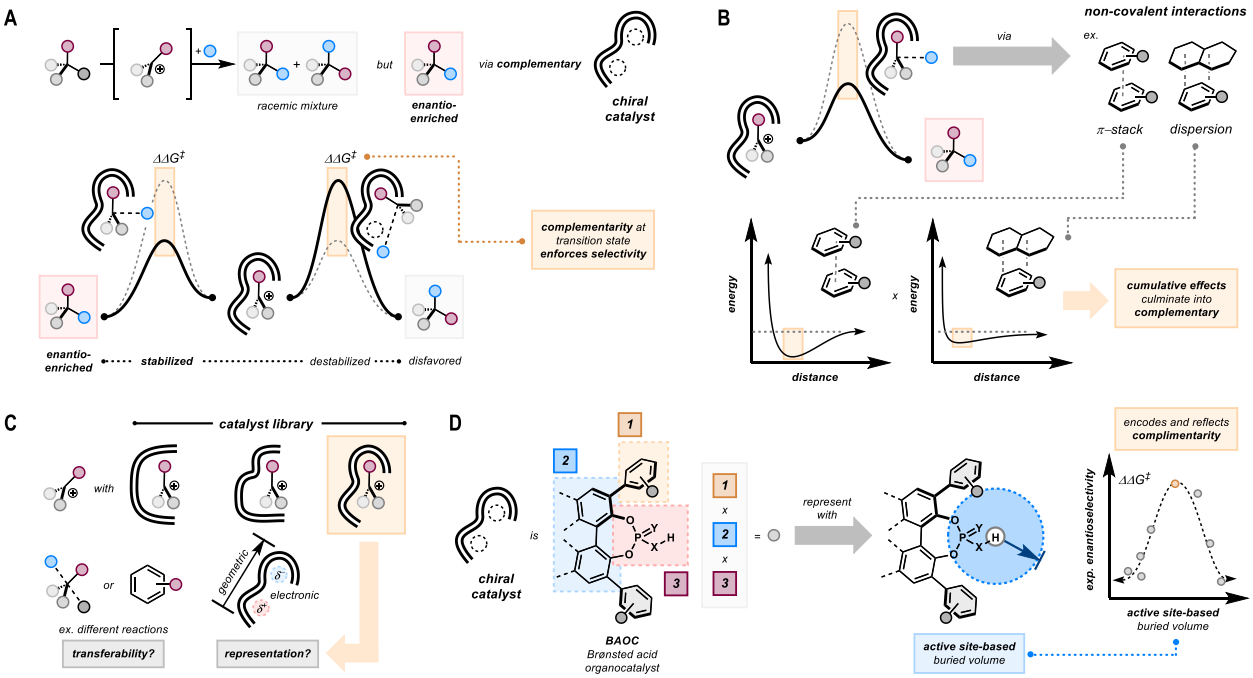


Fig. IV-1 | Complementarity between catalyst and substrate drives selectivity. (a) Catalysts can exploit complementarity at the transition state to drive the selective formation of one product over another (ex. stereoselectivity). (**b**) Complementarity is the cumulative effect of many weak non-covalent interactions (NCIs). (**c**) Complementarity is often recognized *post hoc*. Rendering the principle operational requires a molecular descriptor that readily characterizes each catalyst (ex. representation) and is relevant across all relevant reactivity (ex. transferrable). (**d**) *This work* – active site-based buried volume, in conjugation with Gaussian regression, intuitively encodes and reflects complementarity of these catalysts in stereoselective reactions.

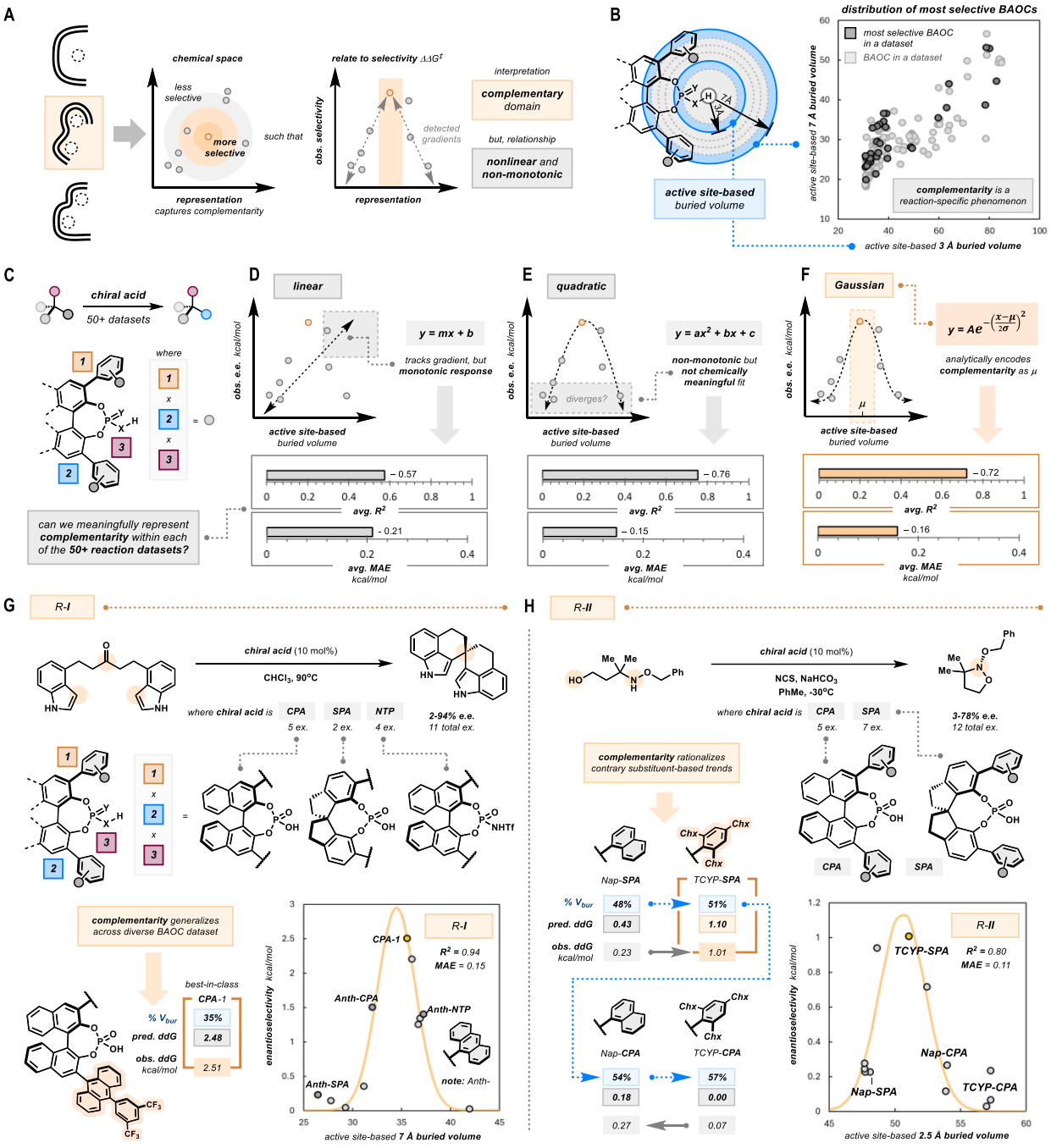


Fig. IV-2 | Complementarity is analytically represented by Gaussian regression. (a) Molecular representations for complementarity must capture its nonlinear, nonmonotonic nature. (**b**) No single BAOC *confinement* universally confers stereoselectivity, aligning with the reaction-specific basis of *complementarity*. (**c**) The average performance (avg. R^2, MAE) by active site-based buried volume of myriad BAOCs in 50+ reaction datasets shows that (**d**) linear regression cannot reflect complementarity as well as (**e**) non-physical, quadratic regression or (**f**) chemically meaningful, Gaussian regression. (**g-h**) Datasets R-**I**[36] and R-**II**[37] exhibit hallmarks of complementarity and are well-represented by Gaussian regression to active site buried volume.

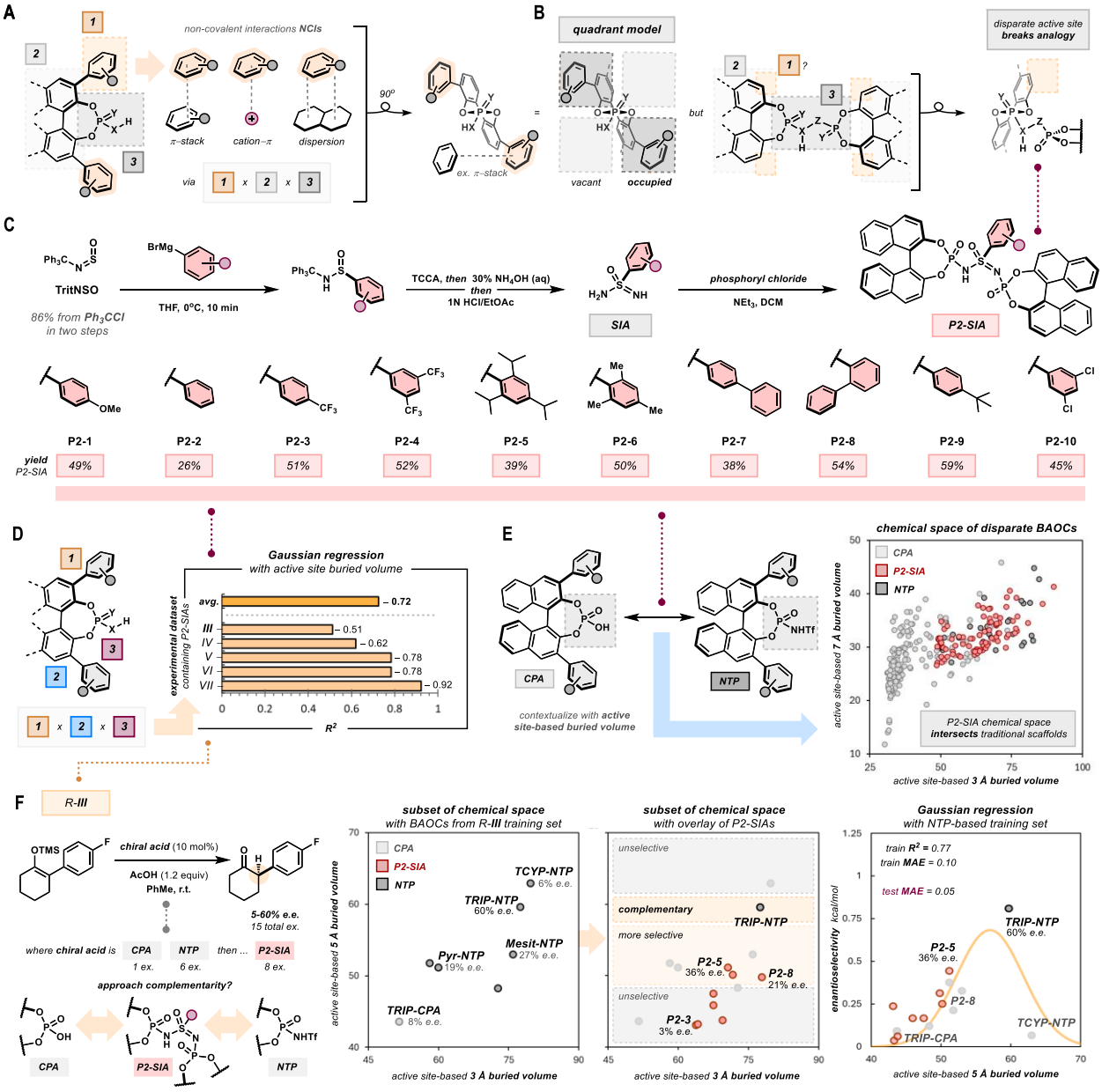


Fig. IV-3 | Active site-based buried volume coherently reflects complementarity across disparate catalyst scaffolds. (**a**) Stereoselective NCIs are traditionally conferred by modular 3,3'-substituents decorating chiral backbones. (**b**) Simple stereochemical models reflecting complementarity require analogous sense of enantioinduction. (**c**) Synthetic access into an unreported, disparate catalyst scaffold – P2-SIA (see **Appendix IV – Experimental Benchmarks and Datasets** for full synthetic details). (**d**) Stereoselectivity by BAOCs in five experimental benchmarks (R-**III** through R-**VII**) including P2-SIAs are well-represented by Gaussian regression with active site-based buried volume (see Supplementary Information for benchmark reaction details and datasets). (**e**) P2-SIAs intersect the region of chemical space between traditional BAOCs like CPAs and NTPs. (**f**) Dataset R-**III**[38] showcases how *near*-complementary P2-SIAs (e.g. P2-5) are advantageously stereoselective relative to traditional BAOCs like CPAs and NTPs.

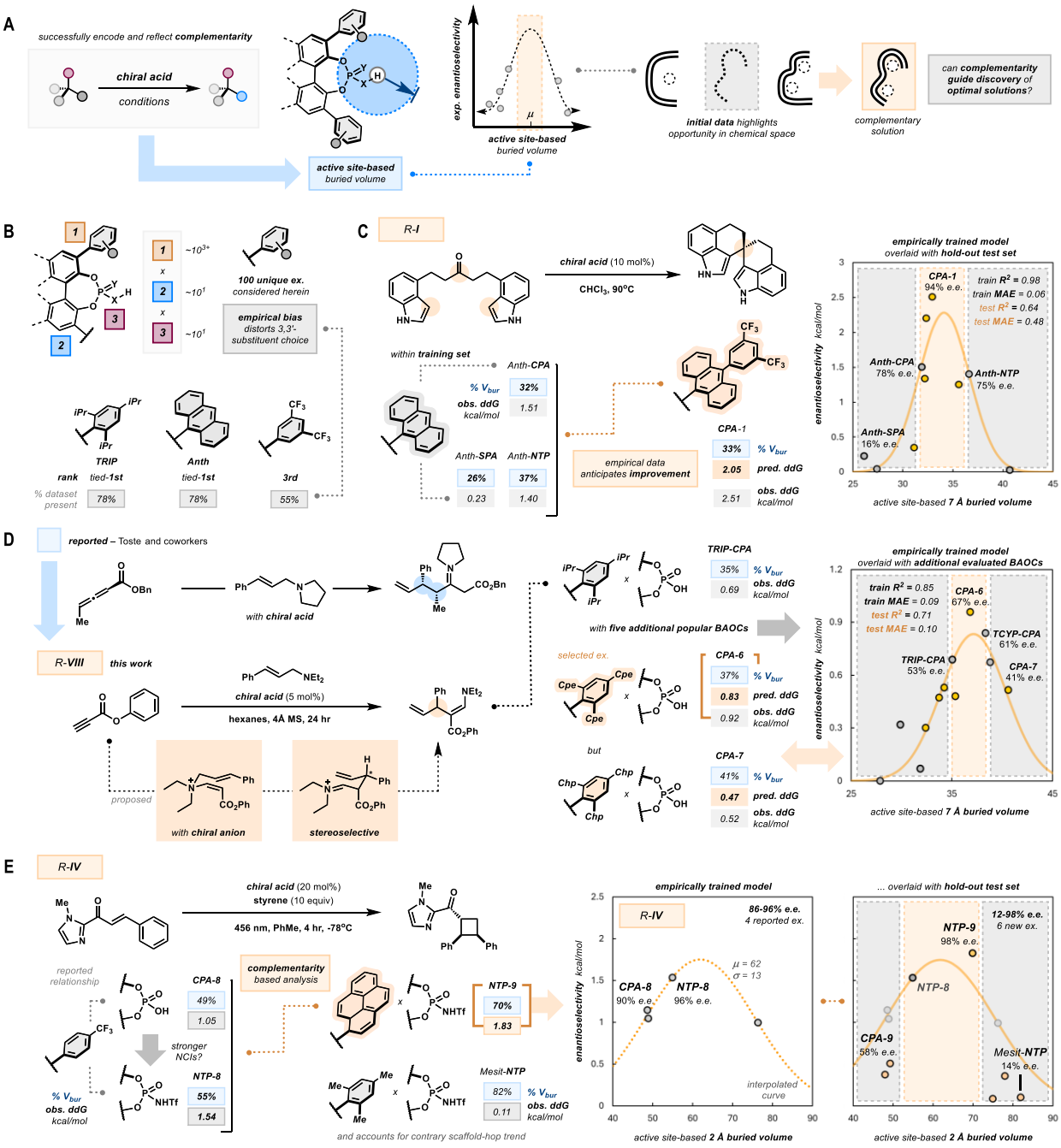


Fig. IV-4 | Complementarity guides data-driven optimization of catalyst stereoselectivity. (**a**) Proactive application of *complementarity* promises optimization for greater stereoselectivity. (**b**) Many datasets are constructed with empirical preference for only a few specific, *privileged* BAOCs. (**c**) Empirical training set (dark gray) from reaction R-I furnishes a Gaussian regression model that anticipates and predicts optimal stereoselective solutions from less popular options (gold). (**d**) R-**VIII** showcases how empirical data (dark gray) can be leveraged to anticipate the utility of new candidates (gold) during reaction discovery. (**e**) Reaction R-**IV** inspired by Yoon, et al.[41] highlights the data-driven application of complementarity to inform catalyst optimization.

IV.5 – References

(1) Nelson, D. L.; Cox, M., M. *Lehninger Principles of Biochemistry*, 4th ed.; W.H. Freeman & Co.

(2) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(3) Kaeding, W. W.; Chu, C.; Young, L. B.; Weinstein, B.; Butter, S. A. Selective Alkylation of Toluene with Methanol to Produce Para-Xylene. *Journal of Catalysis* **1981**, *67* (1), 159–174.

(4) Mitschke, B.; Turberg, M.; List, B. Confinement as a Unifying Element in Selective Catalysis. *Chem* **2020**, *6* (10), 2515–2532.

(5) Sak, M. H.; Jacobsen, E. N. Selective Noncovalent Catalysis with Small Molecules. *Chem. Rev.* **2025**, *125* (17), 8319–8369.

(6) Betinol, I. O.; Kuang, Y.; Mulley, B. P.; Reid, J. P. Controlling Stereoselectivity with Noncovalent Interactions in Chiral Phosphoric Acid Organocatalysis. *Chem. Rev.* **2025**, *125* (8), 4184–4286.

(7) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.

(8) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543* (7647), 637–646.

(9) Treacy, S. M.; Smith, A. L.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science. *J. Am. Chem. Soc.* **2024**, *146* (43), 29792–29800.

(10) Mouarrawis, V.; Plessius, R.; van der Vlugt, J. I.; Reek, J. N. H. Confinement Effects in Catalysis Using Well-Defined Materials and Cages. *Front. Chem.* **2018**, *6*.

(11) Morimoto, M.; Bierschenk, S. M.; Xia, K. T.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Advances in Supramolecular Host-Mediated Reactivity. *Nat Catal* **2020**, *3* (12), 969–984.

(12) Grimm, J. A. A.; Zhou, H.; Properzi, R.; Leutzsch, M.; Bistoni, G.; Nienhaus, J.; List, B. Catalytic Asymmetric Synthesis of Cannabinoids and Menthol from Neral. *Nature* **2023**, *615* (7953), 634–639.

(13) Luo, N.; Turberg, M.; Leutzsch, M.; Mitschke, B.; Brunen, S.; Wakchaure, V. N.; Nöthling, N.; Schelwies, M.; Pelzer, R.; List, B. The Catalytic Asymmetric Polyene Cyclization of Homofarnesol to Ambrox. *Nature* **2024**, *632* (8026), 795–801.

(14) Reid, J. P.; Goodman, J. M. Goldilocks Catalysts: Computational Insights into the Role of the 3,3′ Substituents on the Selectivity of BINOL-Derived Phosphoric Acid Catalysts. *J. Am. Chem. Soc.* **2016**, *138* (25), 7910–7917.

(15) Champagne, P. A.; Houk, K. N. Origins of Selectivity and General Model for Chiral Phosphoric Acid-Catalyzed Oxetane Desymmetrizations. *J. Am. Chem. Soc.* **2016**, *138* (38), 12356–12359.

(16) Alexandre Champagne, P. Identifying the True Origins of Selectivity in Chiral Phosphoric Acid Catalyzed N -Acyl-Azetidine Desymmetrizations. *Chemical Science* **2021**, *12* (47), 15662–15672.

(17) Duarte, F.; Paton, R. S. Molecular Recognition in Asymmetric Counteranion Catalysis: Understanding Chiral Phosphate-Mediated Desymmetrization. *J. Am. Chem. Soc.* **2017**, *139* (26), 8886–8896.

(18) Reid, J. P.; Goodman, J. M. Selecting Chiral BINOL-Derived Phosphoric Acid Catalysts: General Model To Identify Steric Features Essential for Enantioselectivity. *Chemistry – A European Journal* **2017**, *23* (57), 14248–14260.
(19) B. Santiago, C.; Guo, J.-Y.; S. Sigman, M. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chemical Science* **2018**, *9* (9), 2398–2412.
(20) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54* (4), 827–836.
(21) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, *347* (6223), 737–743.
(22) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631.
(23) Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020**, *120* (3), 1620–1689.
(24) Tsuji, N.; Sidorov, P.; Zhu, C.; Nagata, Y.; Gimadiev, T.; Varnek, A.; List, B. Predicting Highly Enantioselective Catalysts Using Tunable Fragment Descriptors. *Angewandte Chemie International Edition* **2023**, *62* (11), e202218659.
(25) Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B. Catalytic Asymmetric Dihydroxylation. *Chem. Rev.* **1994**, *94* (8), 2483–2547.
(26) Evans, D. A.; Bartroli, J.; Shih, T. L. Enantioselective Aldol Condensations. 2. Erythro-Selective Chiral Aldol Condensations via Boron Enolates. *J. Am. Chem. Soc.* **1981**, *103* (8), 2127–2129.
(27) Liu, G.; Cogan, D. A.; Ellman, J. A. Catalytic Asymmetric Synthesis of Tert-Butanesulfinamide. Application to the Asymmetric Synthesis of Amines. *J. Am. Chem. Soc.* **1997**, *119* (41), 9913–9914.
(28) Smith, A. L.; Toste, F. D. Stereoselective Generalizations over Diverse Sets of Chiral Acids Enabled by Buried Volume. *J. Am. Chem. Soc.* **2026**, jacs.5c20342.
(29) Akiyama, T. Stronger Brønsted Acids. *Chem. Rev.* **2007**, *107* (12), 5744–5758.
(30) Handjaya, J. P.; Patankar, N.; Reid, J. P. The Diversity and Evolution of Chiral Brønsted Acid Structures. *Chemistry – A European Journal* **2024**, *30* (46), e202400921.
(31) Hamilton, G. L.; Kang, E. J.; Mba, M.; Toste, F. D. A Powerful Chiral Counterion Strategy for Asymmetric Transition Metal Catalysis. *Science* **2007**, *317*.
(32) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308.
(33) Antinucci, G.; Dereli, B.; Vittoria, A.; Budzelaar, P. H. M.; Cipullo, R.; Goryunov, G. P.; Kulyabin, P. S.; Uborsky, D. V.; Cavallo, L.; Ehm, C.; Voskoboynikov, A. Z.; Busico, V. Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction. *ACS Catal.* **2022**, *12* (12), 6934–6945.
(34) Reid, J. P.; Sigman, M. S. Comparing Quantitative Prediction Methods for the Discovery of Small-Molecule Chiral Catalysts. *Nat Rev Chem* **2018**, *2* (10), 290–305.
(35) Ardkhean, R.; Roth, P. M. C.; Maksymowicz, R. M.; Curran, A.; Peng, Q.; Paton, R. S.; Fletcher, S. P. Enantioselective Conjugate Addition Catalyzed by a Copper Phosphoramidite

Complex: Computational and Experimental Exploration of Asymmetric Induction. *ACS Catal.* **2017**, *7* (10), 6729–6737.
(36) Zhao, H.-W.; Jiang, F.; Chen, S.; Hu, J.; Xiang, S.-H.; Ding, W.-Y.; Lu, W.; Tan, B. Organocatalytic Asymmetric Construction and Application of Axially Chiral Spiro-Bisindoles. *Angewandte Chemie International Edition* **2025**, *64* (12), e202422951.
(37) Wu, S.; Chen, P.; Duan, M.; Jiang, P.-Y.; Zhou, Q.; Xiang, S.-H.; Houk, K. N.; Tan, B. Controlling Pyramidal Nitrogen Chirality by Asymmetric Organocatalysis. *Nature* **2025**, *647* (8091), 897–905.
(38) Cheon, C. H.; Yamamoto, H. A Brønsted Acid Catalyst for the Enantioselective Protonation Reaction. *J. Am. Chem. Soc.* **2008**, *130* (29), 9246–9247.
(39) Saito, K.; Moriya, Y.; Akiyama, T. Chiral Phosphoric Acid Catalyzed Asymmetric Synthesis of 2-Substituted 2,3-Dihydro-4-Quinolones by a Protecting-Group-Free Approach. *Org. Lett.* **2015**, *17* (13), 3202–3205.
(40) Liao, S.; Čorić, I.; Wang, Q.; List, B. Activation of H2O2 by Chiral Confined Brønsted Acids: A Highly Enantioselective Catalytic Sulfoxidation. *J. Am. Chem. Soc.* **2012**, *134* (26), 10765–10768.
(41) Sherbrook, E. M.; Genzink, M. J.; Park, B.; Guzei, I. A.; Baik, M.-H.; Yoon, T. P. Chiral Brønsted Acid-Controlled Intermolecular Asymmetric [2 + 2] Photocycloadditions. *Nat Commun* **2021**, *12* (1), 5735.
(42) Kwon, Y.; Li, J.; Reid, J. P.; Crawford, J. M.; Jacob, R.; Sigman, M. S.; Toste, F. D.; Miller, S. J. Disparate Catalytic Scaffolds for Atroposelective Cyclodehydration. *J. Am. Chem. Soc.* **2019**, *141* (16), 6698–6705.
(43) Miró, J.; Gensch, T.; Ellwart, M.; Han, S.-J.; Lin, H.-H.; Sigman, M. S.; Toste, F. D. Enantioselective Allenoate-Claisen Rearrangement Using Chiral Phosphate Catalysts. *J. Am. Chem. Soc.* **2020**, *142* (13), 6390–6399.
(44) Sherbrook, E. M.; Jung, H.; Cho, D.; Baik, M.-H.; Yoon, T. P. Brønsted Acid Catalysis of Photosensitized Cycloadditions. *Chem. Sci.* **2020**, *11* (3), 856–861.

Chapter V – **Algorithmic design of stereoselective acid organocatalysts**

Portions of this chapter were adapted from:

Smith, A. L.; Wei, N.; Toste, F. D. *Manuscript in preparation.*

V.1 – Preface

Privileged catalysts are chemical solutions for an ever-expanding range of synthetic applications. However, failure of these tools devolves chemical discovery into an aimless, empirical campaign. For complex phenomena like stereoselective catalysis, experimental progress is inhibited by access into and exploration of chemical space, requiring multistep routes to synthesize just a single, likely unsuccessful, candidate. Here, we present a Bayesian framework for the algorithmic, formal inverse design of stereoselective Brønsted acid organocatalysts (BAOCs). The platform leverages active site-based buried volume as an *in silico* representation for BAOCs, translating catalyst *structure* to stereoselective *function* in a reaction-agnostic manner. We leverage the platform for the algorithmic, *in silico* nomination of new BAOCs based on the properties of privileged alternatives. Experimental validation demonstrates that nominated candidates furnish highly stereoselective solutions, even in cases where little prior data was available to guide catalyst design. Collectively, the platform inverts structure-property relationships to precisely nominate high-value candidates in complex, stereoselective tasks.

V.2 – Introduction

Synthetic chemistry relies on the discovery of conditions that afford a target transformation. Traditionally, chemists have resorted to empirical evaluation across dimensions like temperature, solvent, catalyst, and more.[1] More recently, approaches like high throughput experimentation and data-driven algorithms have improved the efficiency of discovery by transforming reaction outcomes into navigable objective landscapes (**Fig. V-1A**).[2–7] Still, design-make-test-analyze cycles (DMTA) require that experimental trials can be readily evaluated. Unless chemists can engage commercial or readily synthesized candidates, synthesis fundamentally constrains evaluation and thus significantly hinders discovery.

Accordingly, synthetic constraints lead chemists to prioritize the evaluation of privileged choices, minimizing the need for *de novo* exploration by offering empirically reliable solutions. When successful, these privileged solutions are especially efficient, limiting search to a narrow region of chemical space.[8,9] However, failure of these choices devolves exploration to a lengthy empirical campaign with little insight into requisite features for the desired transformation. This limitation is especially pronounced in stereoselective catalysis. For any given reaction, effective solutions are often specific, relying on precise non-covalent interactions to guide reactive intermediates to the desired stereoisomer.[10–12] Even small structural changes in the catalyst can drastically alter the observed selectivity. As stereoselective catalysts often require multistep syntheses to prepare, data-driven strategies that rely on rapid navigation of chemical space are compromised.

Computational tools that help link catalyst structure to function promise an *in silico*, property-based approach for discovery (**Fig. V-1B**). While machine learning-based models like neural networks or random forests can learn complex, unstructured relationships, they require substantial training data, an impractical expectation under severe synthetic burden.[4,13] Conversely, analytical

models like linear regression are more suitable for small datasets, but they can lack generality and transferability.[14,15] In the context of stereoselective catalysis, our group has reported that active site-based buried volume is a uniquely advantageous molecular representation for Brønsted acid organocatalysts (BAOCs).[16] Although the representation simply describes the relative steric occupancy around the active site of the catalyst, the models constructed reveal a chemical space organized against stereoselectivity.[17–19] It follows that synthesis of candidates only in the most promising regions should minimize the burden required by experimental evaluation.

This perspective immediately raises a fundamental challenge – *which structure corresponds to the desired property?* Without sufficient *in silico* data, neither analytical nor machine learned models can *inversely design* a catalyst structure given some property (**Fig. V-1C**).[20] Moreover, stereoselective catalysts are constructed from a collection of diverse structural features, culminating in a combinatorial explosion that cannot be exhaustively computed. To circumvent this, we envisioned the efficient exploration of the property landscape with Bayesian optimization, rapidly solving for catalysts structures only proximal to target regions and thereby reducing the synthetic burden associated with discovery.[21,22]

We herein disclose a Bayesian-powered framework for the inverse design of stereoselective BAOCs. We first demonstrate active site buried volume accelerates the empirical discovery of stereoselective BAOCs in simulated active learning campaigns across 50+ literature-derived datasets. From this success, we show latent opportunities in chemical space can remain, even after finding the *best* reported opportunity. Yet, these *possibilities* are only theoretical – they lack specific structures corresponding to their *in silico* representations. To algorithmically unite structure and property, we leverage a Bayesian-powered framework to efficiently discover catalyst bearing targeted properties. These candidates are then deployed in both known and previously unreported BAOC-catalyzed reactions, showcasing the utility of the platform to discover stereoselective solutions solely from *in silico* property space. Moreover, excellent stereoselective solutions are predictively designed from catalyst scaffolds with no known prior utility (up to 88% *e.e.*), even surpassing the performance of traditional inspiration. We anticipate this work establishes an algorithmic alternative to empirical discovery of stereoselective solutions.

V.3 – Results

- ***Buried volume accelerates optimization campaigns over preselected options***

Robust structure-activity relationships (SARs) promise rational catalyst design by quantitatively linking molecular structure to function through chemically meaningful representations. However, as discussed, *accurate* and *quantitative* generalizations are prohibitively expensive to *completely* evaluate (total no. conditions $>10^6$).[23] Instead, chemists typically prioritize the realization of only specific conditions that improve certain metrics. Accordingly, active learning strategies like Bayesian optimization leverage policies to efficiently navigate complex landscapes under these

limited budgets. Thus, we sought to evaluate whether the empirical discovery of stereoselective BAOCs could be accelerated through judicious choice of molecular representation.

With 50+ compiled literature datasets spanning broad structural *and* functional diversity of reported BAOCs,[16] simulated active learning campaigns illustrate the utility of a representation to guide optimization. From each reported dataset, one datapoint was randomly selected to initialize the simulation. The Bayesian surrogate model was fit, and the acquisition function evaluated the likelihood of improvement over the existing observations (see **Appendix V** for full model details, acquisition function is expected improvement, EI). The most promising choice was then made sequentially from the remaining set of reported options. Once the *most* stereoselective BAOC was chosen, the simulation was terminated and restarted from a different initialization (100 runs/dataset). Finally, we quantified how often simulations successfully converged to the best choice available, how fast it converged, and whether the model expected further evaluations would be promising, even though the *best* choice had been made.

We initially compared the utility of different molecular representations for BAOCs provided restrictive experimental budgets. As mentioned above, our group has recently reported that BAOCs can be advantageously represented by active site-based buried volume when modeling their stereoselectivity. Indeed, over the active learning campaigns conducted, buried volume accelerates empirical discovery of the most stereoselective option, especially under limited budgets. Although the average dataset contains over 12 possible BAOCs to choose, the representation helped realize the best choice nearly 70% of the time when provided a budget of only *four* additional experiments (**Fig. V-2A**, see **Supplementary Table V-3** for tabulated metrics). Critically, this utility is distinct from other competing molecular representations for confinement and acidity – two oft cited qualities by synthetic chemists when evaluating BAOCs. Sterimol, which captures geometric properties of the BAOC active site, only provides a small improvement over baseline, realizing the best choice about 60% of the time. On the other hand, calculated pK_a, a benchmarked surrogate for catalyst acidity, is altogether ineffective and statistically indistinguishable from random choice (**Fig. V-2A**). Only when the experimental budget is substantially increased do other molecular representations converge to familiar utility.

Although rapid convergence is promising, we were also interested to see if the active learning model can recognize if the *most* stereoselective choice was chosen. Approaching the *best* possible choice, the model should reflect strong prioritization (ex. *large* expected improvement, EI). Once made, the model should expect the remaining choices to offer substantially reduced opportunity (ex. *low* EI). From the campaign conducted, we plotted the average difference between EI *at* and *after* making the best possible choice in a dataset to visualize this insight (avg. ΔEI, **Fig. V-2B**, see **Supplementary Table V-4** for tabulated metrics). Under all experimental budgets, only models trained with active site-based buried volume representations accurately realize the *best* choice is unique (avg. $\Delta EI < 0$). Each other representation erroneously expects *better* opportunity among the remaining, less-stereoselective options (avg. $\Delta EI > 0$), suggesting the realized convergence

was not intentional. Only with larger experimental budgets is this invalid expectation attenuated (sterimol avg. ΔEI = 0.0; calculated pK_a avg. ΔEI = +0.06; with budget = 8, **Fig. V-2B**).

- ***In silico, formal inverse design of BAOCs.***

Above, we have shown that models trained on buried volume accurately recognize when the best BAOC has been selected. Still, these campaigns were restricted only to reported choices. We wondered whether these models might expect additional stereoselective opportunity beyond the known catalysts. To evaluate this, we extracted the maximum expected improvement (EI) across the buried volume space accessible to BAOCs (max. EI). Although the latent opportunity was typically small (avg. max. EI = 0.01), some datasets exhibited substantially larger chances (up to max. EI = 0.05). In these cases, the model predicts unrealized opportunity in sparsely populated regions near known solutions (**Supplementary Fig. V-5**). However, these *latent* prospects are defined only by buried volume, disjoint from tangible structures for evaluation.

Accordingly, we seek catalyst discovery from the inverse design formulation – *given a target region of buried volume, identify structures that realize it* (**Fig. V-3A**). Because active site-based buried volume empirically organizes catalyst structure with respect to stereoselective function, proximity to target buried volume implies target stereoselectivity. Yet, the combinatorial nature of BAOC structure precludes direct inversion of this mapping. So, we again deploy Bayesian optimization as an efficient strategy to navigate chemical space under data scarcity where exhaustive exploration is intractable. This closed-loop exploration cycle iteratively proposes BAOC structures expected to progressively approach a target buried volume, and thus target stereoselectivity (**Fig. V-3B**).

We first evaluated whether active site-based buried volume alone encodes sufficient information to guide structural discovery by attempting to rediscover the privileged catalyst TRIP-CPA (TRIP = 2,4,6-triisopropylphenyl, CPA = chiral phosphoric acid). The Bayesian framework was initialized from a random set of eight catalysts and tasked with minimizing the distance in buried volume to that of TRIP-CPA using aryl bromide-derived descriptors. Despite the absence of explicit structural information encoded in the buried volume-based objective, the workflow identified TRIP-CPA within six optimization cycles (<2% of the total space sampled, **Fig. V-3C**). This successful rediscovery serves as a controlled test of whether buried volume is sufficient to guide structural convergence. However, this result alone does not establish whether convergence reflects unique chemical insight encoded by the descriptor or more general behavior of the Bayesian-powered framework.

To benchmark this behavior, we repeated the rediscovery campaign across multiple random initializations and compared performance to optimization guided by Tanimoto similarity, a structure-based metric (**Fig. V-3D**, **Supplementary Table V-5**). While Tanimoto coefficient predictably drives rapid structural convergence, active site-based buried volume also furnishes comparable convergence, despite lacking any explicit structural encoding (**Fig. V-3E-F**, **Supplementary Table V-6-7**). As expected, random search, a baseline metric describing

uninformed selection, fails to identify the target at all. We observe similar trends for the rediscovery of other privileged BAOCs (see **Supplementary Fig. V-8-10**, **Supplementary Table V-8-10**), suggesting that active site-based buried volume holistically encodes the features necessary for formal inverse design from catalyst chemical space.

- ***Algorithmic generalization of a privileged solution with buried volume***

Stereoselective transformations enabled by Brønsted acid organocatalysts (BAOCs) are often guided by privileged catalyst scaffolds, where empirical precedent narrows the scope of candidates evaluated. A canonical example is transfer hydrogenation using Hantzsch ester, where stereocontrol is typically achieved through cooperative activation—most commonly via iminium intermediates that enhance non-covalent interactions with the chiral phosphoric acid (**Fig. V-4A**).[24–26] However, in substrate classes where electrophilic activation is intrinsically high, such intermediates may be unnecessary, altering the mode of stereoinduction. We sought to evaluate this scenario in a conjugate reduction of benzylidenes, where aromatic stabilization facilitates direct reduction without exogenous activation (**Fig. V-4B**).[27] In this regime, the absence of enforced catalyst–substrate organization presents a challenge for stereocontrol.[9,28]

To probe this behavior, we first evaluated the widely deployed TRIP-**CPA** catalyst under standard conditions. While complete conversion of the benzylidene substrate was observed, only modest stereoselectivity was achieved (25% e.e., **Fig. V-4C**), consistent with diminished control in the absence of cooperative activation. Notably, reaction outcome proved highly sensitive to solvent, with stereoselectivity varying from 0–62% e.e. across a small solvent screen. This pronounced dependence on solvent polarity suggests that external factors, rather than catalyst structure alone, play a dominant role in governing stereoinduction, motivating further mechanistic investigation.

Mechanistic interrogation indicates that stereoselectivity is governed by an array of subtle electronic effects rather than simple steric control (**Fig. V-4D**). A strong dependence of stereoselectivity on solvent polarity ($\Delta\Delta G^{\ddagger} \propto 1/\epsilon$) is consistent with charge-separated transition states, implicating ion-pairing interactions in the enantiodetermining step (**Fig. V-4E**). Consistent with this interpretation, counterion variation reveals a systematic dependence of stereoselectivity on ionic radius ($Li^+ > Na^+ > K^+ > Cs^+$), supporting a role for electrostatic interactions in governing stereoinduction (**Fig. V-4F**). Separately, ^{31}P NMR analysis indicates the formation of dynamic catalyst–additive complexes in solution, suggesting that multiple catalyst-associated species may be present under reaction conditions (see **Supplementary Fig. V-14**). However, attempts to rationalize such electronic stereoinduction through traditional substituent-based Hammett descriptors revealed no substantial relationship (avg. *e.e.* = 65%, obs. $\rho = 0.05$, **Supplementary Fig. V-17**). Collectively, these observations suggest that traditional mechanistic insights into stereoinduction do not translate into clear strategies for catalyst selection in this transformation.

Given the absence of reliable empirical guidance, we turned to our algorithmic framework to identify catalysts occupying regions of buried volume proximal to TRIP-**CPA**. Bayesian optimization rapidly populated this region of chemical space, nominating both familiar and

underexplored catalyst architectures (**Supplementary Fig. V-18**). Among these, CPA-**1**, a familiar analog to TRIP-**CPA**, was identified and found to furnish comparable stereoselectivity (61% e.e.). More notably, the model also identified a rare substituent (CPA-**2**) that occupies a similar region of buried volume. Despite its low empirical prevalence, CPA-**2** also exhibited stereoselectivity on par with the privileged alternatives – 52% e.e., consistent with property-driven expectations.

To validate the stereoselective success is not universal to all CPAs, we then synthesized and tested the next most empirically prevalent CPAs (CPA-**3** to CPA-**8**). Although these catalysts are frequently prepared and screened in reaction discovery campaigns, none stereoselectively furnish the product (6 ex., avg. 11% *e.e.*, **Fig. V-5G**). In fact, only CPA-**3** (9-anthracenyl substituted, closest in proximity to TRIP-**CPA** from the empirical set) furnished the product in any meaningful stereoselectivity (obs. 29% *e.e.*). These results demonstrate how property-powered insight can be leveraged to identify catalyst structures that are consistently relevant for stereoinduction, whereas other empirically prioritized strategies fail to furnish any meaningful selectivity.

- ***Algorithmic transfer of stereoselectivity onto previously unreported scaffolds***

Above, the algorithmic platform was applied to identify catalysts within the vicinity of a privileged solution (TRIP-CPA), furnishing unexpected solutions that are proximal to known design motifs. To challenge whether this framework can extend beyond empirical precedent, we next sought to design stereoselective catalysts from structurally distinct scaffolds defined solely by targeted active site-based buried volume.

To this end, we targeted a class of Brønsted acid organocatalysts lacking the canonical 3,3'-substituents that are typically invoked to modulate stereoselective non-covalent interactions (**Fig. V-5A**).[29] Instead, we designed CPA-like derivatives in which these substituents are replaced by a modular S-substituted sulfonamide positioned at the active site (denoted P1-SOAs). Although such scaffolds deviate from conventional BAOC architectures, prior analyses have demonstrated that active site-based buried volume remains a valid representation for stereoselectivity across catalysts with and without 3,3'-substitution. Accordingly, P1-SOAs provide a stringent test of whether stereoselective solutions can be accessed through property-driven design independent of established structural motifs.

Because the introduction of a sulfonamide attenuates the bifunctional Brønsted acid/base character typical of BAOCs, P1-SOAs are expected to operate primarily as chiral Brønsted acids rather than through cooperative activation. Accordingly, we evaluated these catalysts in an asymmetric Friedel–Crafts-type arylation,[30] a transformation known to be governed by chiral Brønsted acid activation. In this transformation, TRIP-CPA furnishes the desired product in high enantioselectivity (83% e.e.), providing a well-defined reference point for stereocontrol. Using this transformation as a target, the algorithmic platform was deployed to identify P1-SOAs whose active site-based buried volume matches that of TRIP-CPA. After six optimization cycles, candidate structures were selected based on synthetic accessibility (**Fig. V-5B-C**, **Supplementary Fig. V-21-22**). Notably, the model nominated P1-SOAs near the target region of buried volume

bearing aryl substituents not yet encountered in BAOC design (P1-**1** and P1-**2**). Evaluation of these candidates revealed excellent stereoselectivity (up to 88% e.e., avg. 78% e.e.), despite the absence of prior precedent for either their reactivity or stereochemical performance.

To assess whether stereoselectivity is intrinsic to the P1-SOA scaffold or specific to the algorithmically identified structures, we synthesized an additional set of P1-SOAs bearing substituents commonly encountered in BAOC literature. We observed these catalysts exhibited moderate activity (avg. 39% e.e., **Fig. V-5D**), but none approached the performance of the model-nominated P1-**1**. Closely related analogs, like biphenyl-substituted P1-**9** furnished the arylation in above average but attenuated selectivity, consistent with the property-based insight (P1-**9**, 62% *e.e.* vs. related P1-**1**, 88% *e.e.*). These results highlight how stereoselectivity is not necessarily unique to catalyst scaffold but may be generalized to specific regions of catalyst space as defined by buried volume. By leveraging the active learning framework, even valuable synthetic candidates from previously unreported catalyst scaffolds can be algorithmically realized from these representations.

V.4 – Discussion and Outlook

Discovery of stereoselective catalysts has long relied on empirical intuition, often biased to emergent, privileged solutions. We have shown that active site-based buried volume can serve as a valuable representation powering the inverse design of BAOCs from targeted properties. We have experimentally demonstrated that this representation can not only generalize candidates beyond privileged solutions, but also even identify highly stereoselective candidates without any prior information regarding their utility. Although we have limited our investigations to BAOCs in this work, we anticipate the inverse design paradigm from any chemically meaningful representations can rapidly identify valuable candidates from extreme data scarcity – in both experimental and computational domains.

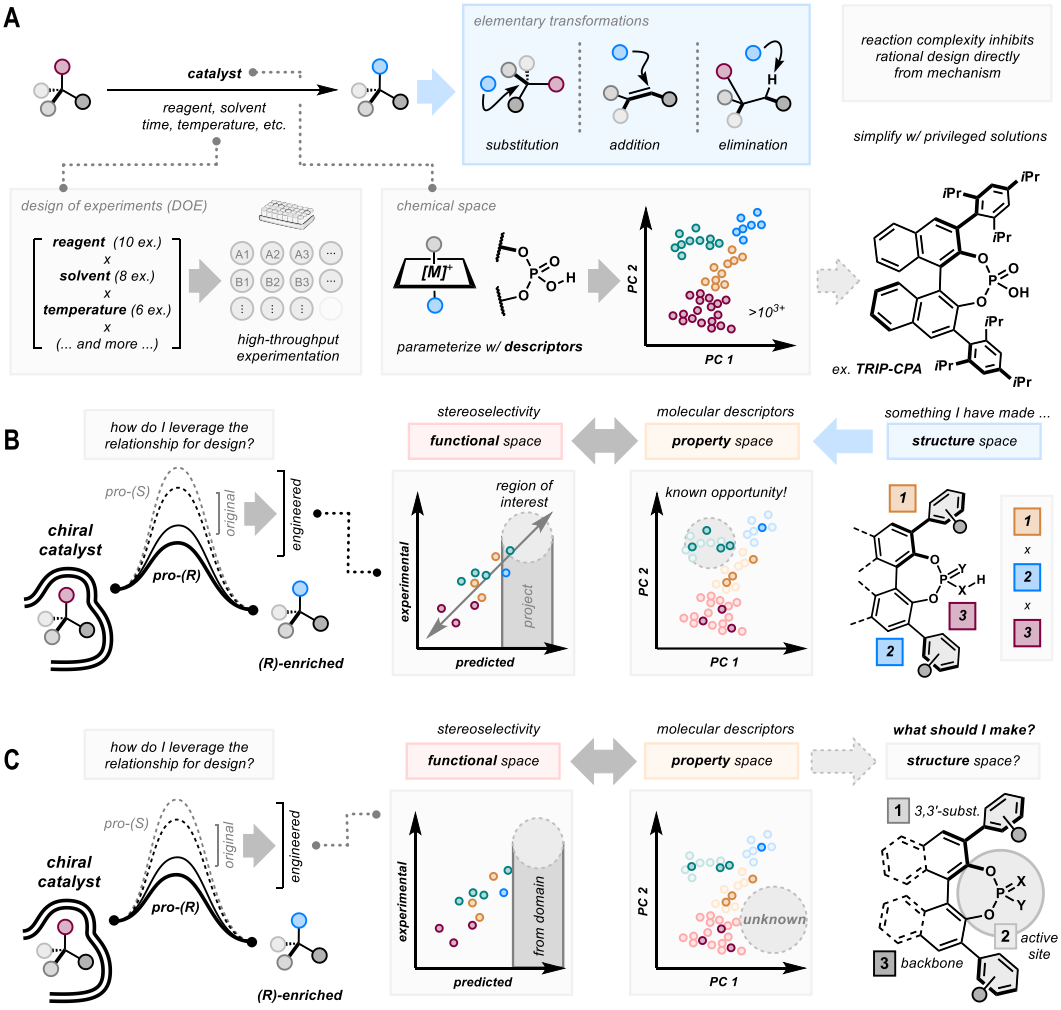


Fig. V-1 | Complex structure-property relationships confound the discovery of molecules with intended function. (**a**) Discovery leverages rational insight coupled with data-driven experimentation to realize solutions. *Privileged* approaches drastically simplify this complexity by presenting a promising opportunity *a priori*. (**b**) Emergent opportunities from data-driven insight require tangible structural solutions from targeted properties. (**c**) Inverse design is required to identify catalyst structures from promising, but *unpopulated* property space.

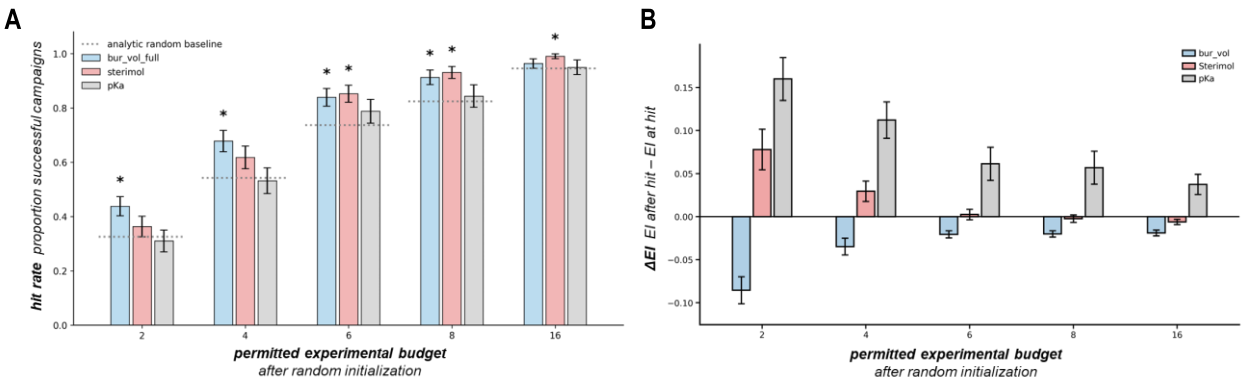


Fig. V-2 | Buried volume-based representation accelerates Bayesian-powered optimization of stereoselective acid catalysts from *preselected* sets. (**a**) Across 50+ literature-derived datasets, Bayesian active learning-based optimization campaigns rapidly identify the *best* chiral acid catalyst from available options when leveraging active site-based buried volume representations. ('*' refers to statistically significant different from the *random* baseline, $p < 0.05$) (**b)** Active learning models trained on active site-based buried volume representations accurately realize the *best* catalyst was observed (avg. ΔEI < 0), while other representations of confinement and acidity are erroneously optimistic (avg. ΔEI > 0). See **Appendix V** for full tabulated metrics.

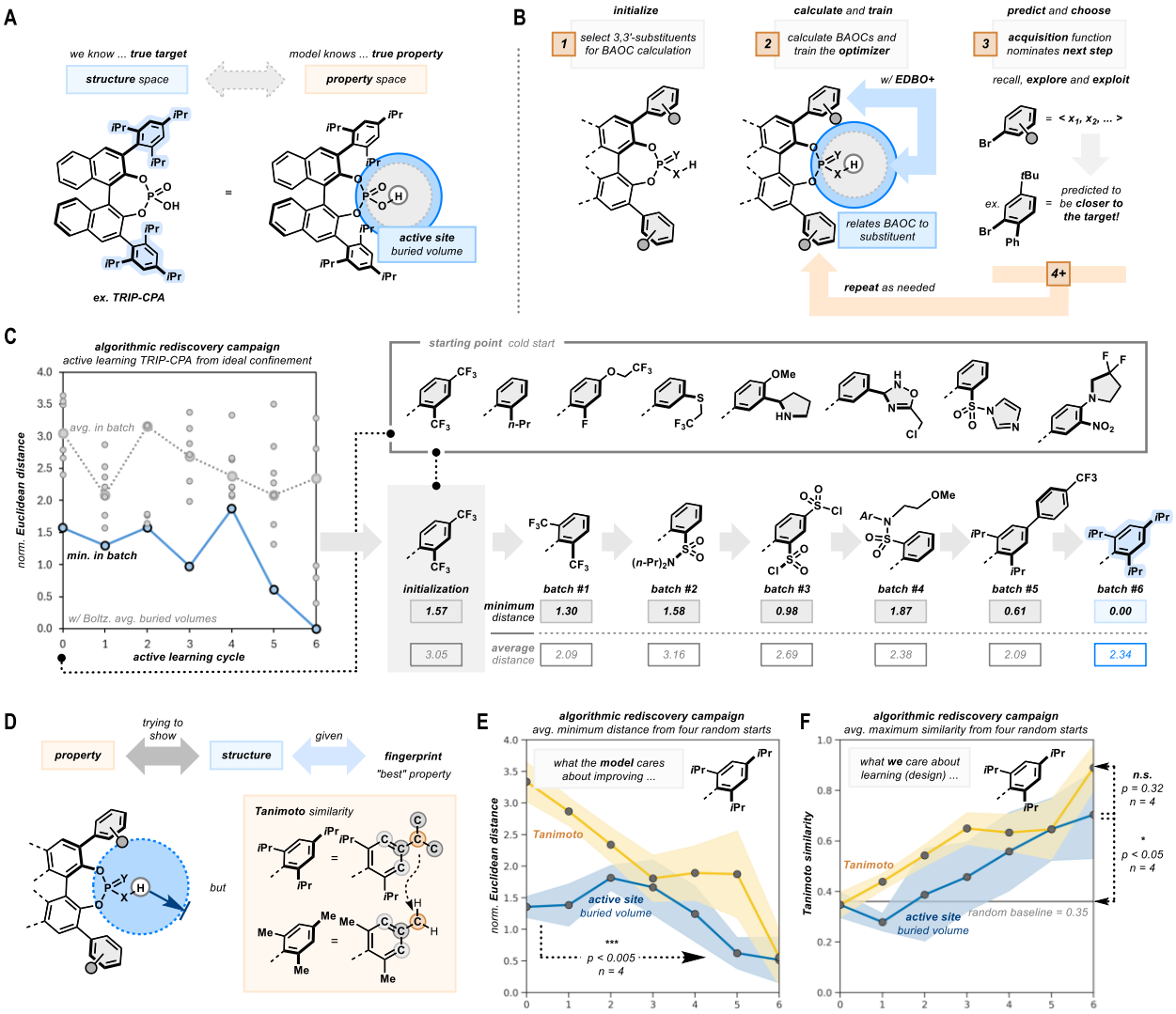


Fig. V-3 | Bayesian optimization-powered workflow accelerates the *in silico* discovery of chiral acid catalysts from active site-based buried volume representation. (a) Inverse design seeks to invert the relationship between structure (chiral acid catalyst) and property (active site-based buried volume). **(b)** Our formal inverse design workflow iteratively learns the relationship between structure and property through an active-learning loop where structures are selected based on their predicted proximity to the target property. Repeated data-generation efficiently improves the quality of these predictions without requiring complete computational characterization. **(c)** Convergent rediscovery of stereoselective chiral acids showcases active site-based buried volume sufficiently guides active learning-powered inverse design. **(d)** Active site-based buried volume is comparable to Tanimoto index, a traditional structural similarity metric. **(e-f)** Repeated rediscovery campaigns highlight the comparable utility of active site-based buried volume for the design of precise structural solutions.

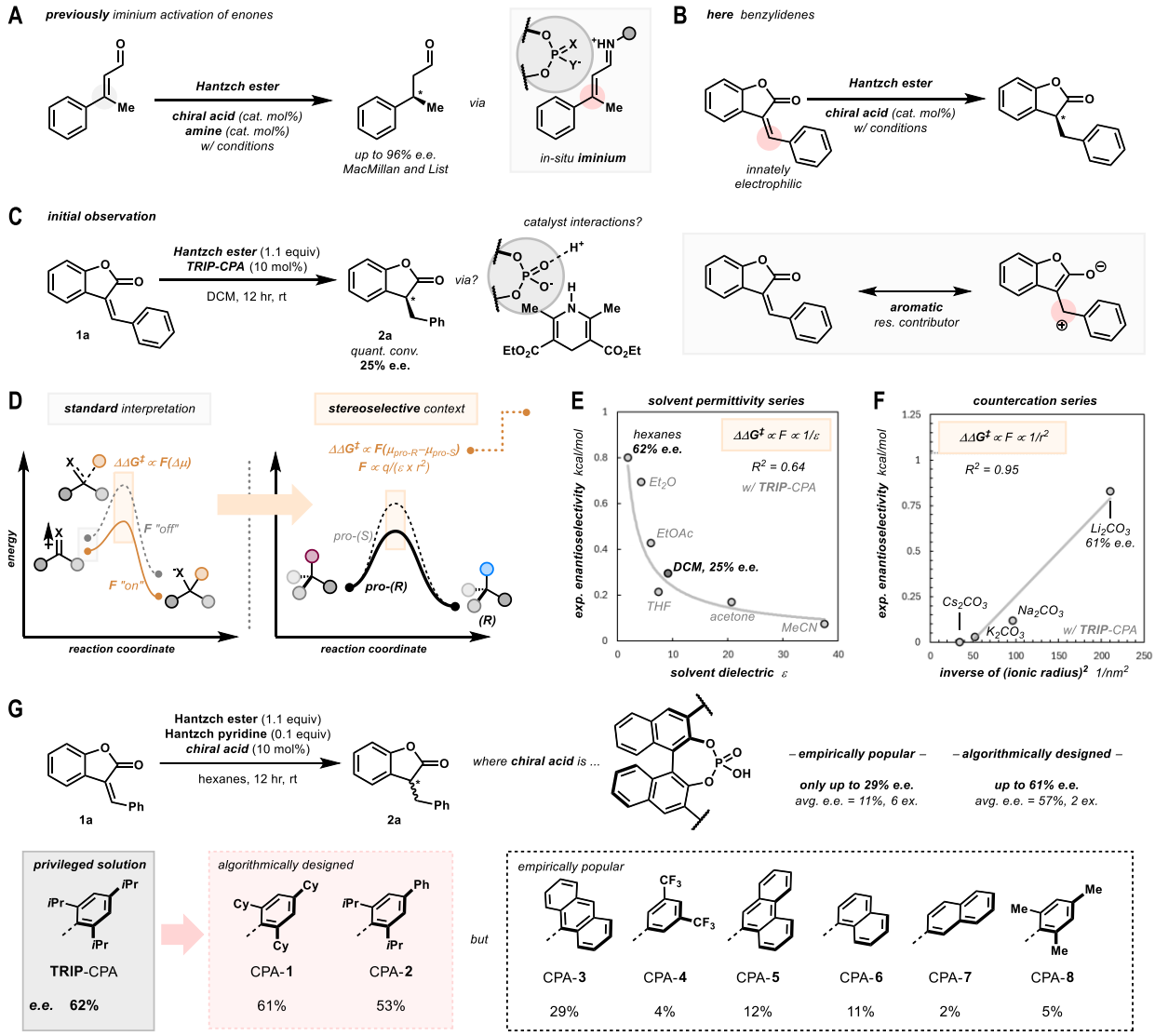


Fig. V-4 | Data-driven and algorithmic discovery of stereoselective solutions for previously unreported reactivity. (a-b) Asymmetric conjugate reduction of benzylidenes leverages innate substrate electrophilicity unlike familiar organocatalytic strategies. **(c)** *Privileged* BAOC, TRIP-CPA, affords stereoselective opportunity in conjugate reduction (obs. 25% *e.e.*). **(d-f)** Mechanistic insight suggests a complex interplay of catalytic activation and electronic stereocontrol to realize the observed stereoselectivity but does not obviously beg catalyst design insight. **(g)** Algorithmically designed BAOCs familiar to TRIP-CPA (similar active site-based buried volume) exhibit advantageous stereoselective performance in the asymmetric conjugate reduction of (up to 62% *e.e.*), whereas empirically popular BAOCs are altogether ineffective (up to 29% *e.e.*, 6 ex.).

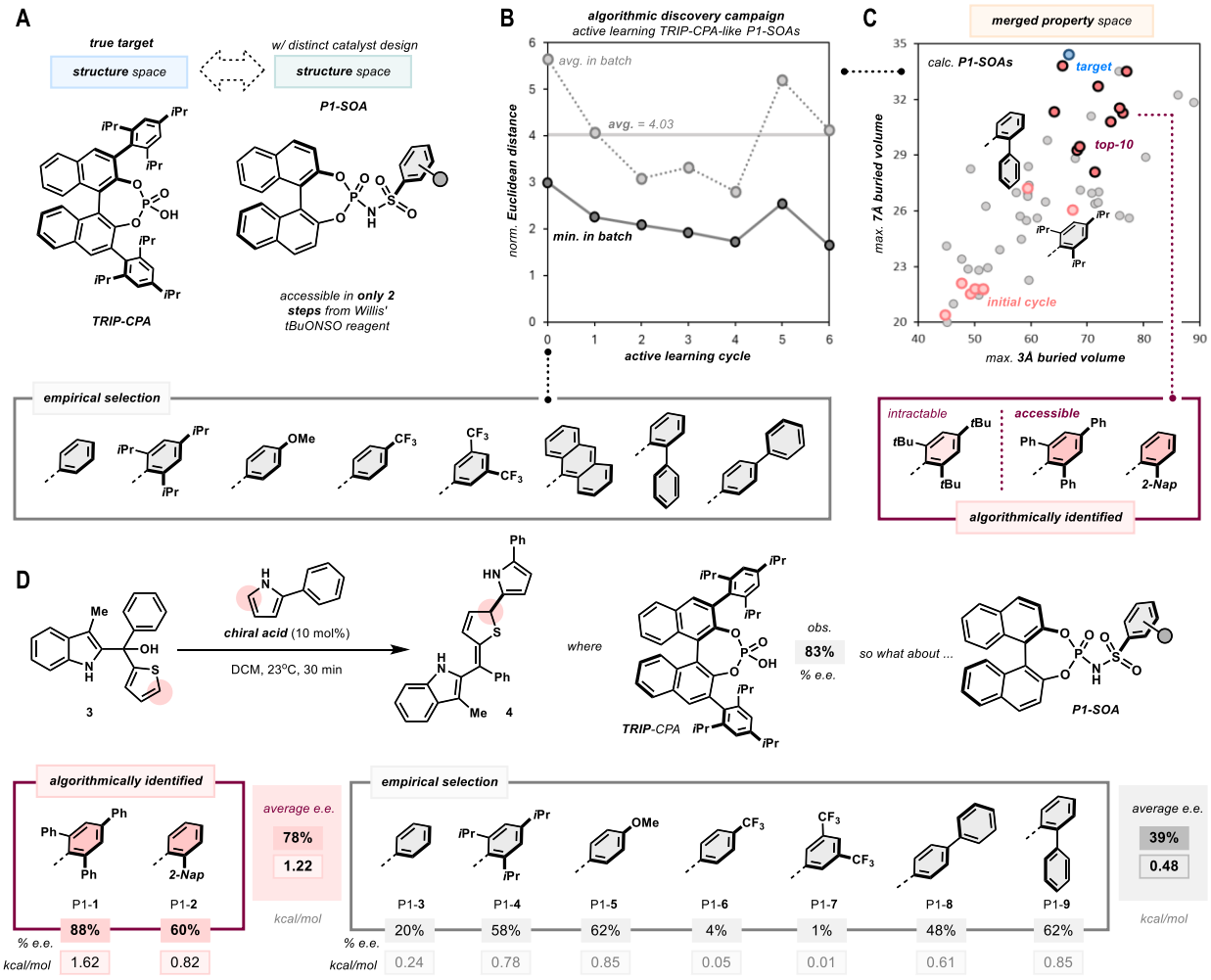


Fig. V-5 | Algorithmic translation of privileged stereoselective solutions onto new catalyst frameworks. (a) 3,3'-substituents along chiral Brønsted acids are privileged structural elements to modulate stereoinduction. The proposed P1-SOA framework lacks these substituents and instead display modular *S*-substitution. **(b-c)** Algorithmic, *in silico* discovery of P1-SOAs resembling the privileged **TRIP**-CPA. Initialization occurred from empirically familiar choices for substituents. The algorithm identified new promising candidates near the target after six learning cycles (<2% total search space). **(d)** Algorithmically designed P1-SOAs are advantageously stereoselective in a Friedel-Crafts-type arylation.

V.5 – References

(1) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123* (6), 3089–3126.

(2) Kulik, H. J.; Sigman, M. S. Advancing Discovery in Chemistry with Artificial Intelligence: From Reaction Outcomes to New Materials and Catalysts. *Acc. Chem. Res.* **2021**, *54* (10), 2335–2336.

(3) B. Santiago, C.; Guo, J.-Y.; S. Sigman, M. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chemical Science* **2018**, *9* (9), 2398–2412.

(4) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190.

(5) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96.

(6) Koscher, B. A.; Canty, R. B.; McDonald, M. A.; Greenman, K. P.; McGill, C. J.; Bilodeau, C. L.; Jin, W.; Wu, H.; Vermeire, F. H.; Jin, B.; Hart, T.; Kulesza, T.; Li, S.-C.; Jaakkola, T. S.; Barzilay, R.; Gómez-Bombarelli, R.; Green, W. H.; Jensen, K. F. Autonomous, Multiproperty-Driven Molecular Discovery: From Predictions to Measurements and Back. *Science* **2023**, *382* (6677), eadi1407.

(7) Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous Chemical Research with Large Language Models. *Nature* **2023**, *624* (7992), 570–578.

(8) Yoon, T. P.; Jacobsen, E. N. Privileged Chiral Catalysts. *Science* **2003**, *299* (5613), 1691–1693.

(9) Betinol, I. O.; Kuang, Y.; Mulley, B. P.; Reid, J. P. Controlling Stereoselectivity with Noncovalent Interactions in Chiral Phosphoric Acid Organocatalysis. *Chem. Rev.* **2025**, *125* (8), 4184–4286.

(10) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(11) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543* (7647), 637–646.

(12) Sak, M. H.; Jacobsen, E. N. Selective Noncovalent Catalysis with Small Molecules. *Chem. Rev.* **2025**, *125* (17), 8319–8369.

(13) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631.

(14) Dalmau, D.; Sigman, M. S.; Alegre-Requena, J. V. Machine Learning Workflows beyond Linear Models in Low-Data Regimes. *Chem. Sci.* **2025**.

(15) Gallarati, S.; Bucci, E. M.; Doyle, A. G.; Sigman, M. S. Transferable Enantioselectivity Models from Sparse Data. *Nature* **2026**, *651* (8106), 637–646.

(16) Smith, A. L.; Toste, F. D. Stereoselective Generalizations over Diverse Sets of Chiral Acids Enabled by Buried Volume. *J. Am. Chem. Soc.* **2026**, jacs.5c20342.

(17) Escayola, S.; Bahri-Laleh, N.; Poater, A. % V Bur Index and Steric Maps: From Predictive Catalysis to Machine Learning. *Chemical Society Reviews* **2024**, *53* (2), 853–882.

(18) Hillier, A. C.; Sommer, W. J.; Yong, B. S.; Petersen, J. L.; Cavallo, L.; Nolan, S. P. A Combined Experimental and Theoretical Study Examining the Binding of N-Heterocyclic

Carbenes (NHC) to the Cp*RuCl (Cp* = H5-C5Me5) Moiety: Insight into Stereoelectronic Differences between Unsaturated and Saturated NHC Ligands. *Organometallics* **2003**, *22* (21), 4322–4326.

(19) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11* (10), 872–879.

(20) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361* (6400), 360–365.

(21) Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022**, *144* (43), 19999–20007.

(22) Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **2018**, *4* (9), 1134–1145.

(23) Coley, C. W. Defining and Exploring Chemical Spaces. *Trends in Chemistry* **2021**, *3* (2), 133–145.

(24) Ouellet, S. G.; Walji, A. M.; Macmillan, D. W. C. Enantioselective Organocatalytic Transfer Hydrogenation Reactions Using Hantzsch Esters. *Acc. Chem. Res.* **2007**, *40* (12), 1327–1339.

(25) Martin, N. J. A.; List, B. Highly Enantioselective Transfer Hydrogenation of α,β-Unsaturated Ketones. *J. Am. Chem. Soc.* **2006**, *128* (41), 13368–13369.

(26) Hoffmann, S.; Seayad, A. M.; List, B. A Powerful Brønsted Acid Catalyst for the Organocatalytic Asymmetric Transfer Hydrogenation of Imines. *Angewandte Chemie International Edition* **2005**, *44* (45), 7424–7427.

(27) Allgäuer, D. S.; Jangra, H.; Asahara, H.; Li, Z.; Chen, Q.; Zipse, H.; Ofial, A. R.; Mayr, H. Quantification and Theoretical Analysis of the Electrophilicities of Michael Acceptors. *J. Am. Chem. Soc.* **2017**, *139* (38), 13318–13329.

(28) Schreyer, L.; Properzi, R.; List, B. IDPi Catalysis. *Angewandte Chemie International Edition* **2019**, *58* (37), 12761–12777.

(29) Handjaya, J. P.; Patankar, N.; Reid, J. P. The Diversity and Evolution of Chiral Brønsted Acid Structures. *Chemistry – A European Journal* **2024**, *30* (46), e202400921.

(30) Li, X.; Duan, M.; Yu, P.; Houk, K. N.; Sun, J. Organocatalytic Enantioselective Dearomatization of Thiophenes by 1,10-Conjugate Addition of Indole Imine Methides. *Nat Commun* **2021**, *12* (1), 4881.

Chapter VI – **Recasting acid organocatalysts for a suite of radical reactions**

Portions of this chapter were adapted from:

Smith, A. L.†; Moeckel, A. C.†; Tandean, E. M.; Toste, F. D. *Manuscript in revision.*

VI.1 – Preface

Photoredox catalysis requires the simultaneous action of many reactive, open-shell species. General strategies that impose control over each step *en route* to the desired product are essential to manage these competing pathways. We report dithiophosphoric acids, highly acidic Brønsted acid organocatalysts, are versatile tools in a suite of open-shell reactions, including group transfers, additions, and fragmentations. The catalysts have a strong preference for ionization to the dithiophosphate anion which competitively associates and quenches cationic photocatalysts, generating reactive dithiophosphoryl radicals. While these intermediates are known H-atom abstraction agents, we demonstrate that dithiophosphoric acids are also competent H-atom donation agents. Moreover, we observe dithiophosphoryl radicals undergo reversible π-additions into alkenes, enabling homolytic substitutions. Due to the competitive quenching of cationic photocatalysts by dithiophosphates, kinetic control across reactive species can emerge under photochemical conditions. We realize this paradigm to selectively discriminate substrates and products in a voltage-independent manner. Collectively, these results introduce dithiophosphoric acids as multifunctional catalysts with broad opportunity across many open-shell reactions

VI.2 – Introduction

Photoredox catalysis requires the careful orchestration of many reactive, open-shell species to achieve the desired transformation. Substrates, reagents, catalysts, products and more can all interact with the deployed photocatalyst.[1–3] Therefore, successful catalysis depends on precise photochemical events in which one species among many is productively activated (**Fig. VI-1A**). Following photon absorption, the photocatalyst generates an excited state that induces energy or electron transfer. Downstream elementary reactions can include group transfer, additions, and fragmentations, each mediated by auxiliary catalysts. Failure to follow the requisite sequence leads to unproductive quenching and deleterious side reactions, owing to the high-energy intermediates generated during photochemical events.

Strategies that impose control over each of these elementary steps promise new opportunities for photochemical reactions. In the field of organocatalysis, synthetic chemists routinely construct catalysts with predictable structure-activity relationships. Brønsted acid organocatalysts (BAOCs) are one versatile platform in which function is closely tied to acidity.[4–6] Systematic installation of electron-withdrawing groups at the catalyst active site precisely modulates acidity, enabling diverse applications. In photochemical contexts, BAOCs have been widely deployed as polar, *closed-shell* catalysts functioning as bases, hydrogen-bond donors, acids, and counter-anions, as demonstrated by List,[7,8] Nicewicz,[9] Knowles,[10–12] Phipps,[13,14] and more (**Fig. VI-1B**).

It is widely recognized that the choice of photocatalyst counterion can have a profound impact on photochemical activity.[15–17] Yet, these species are typically spectators to reactivity beyond modulating the photocatalyst itself. To this end, we wondered whether a *functional* counterion could influence *both* photocatalyst-mediated events and downstream reactivity. Notably, BAOCs

bearing polarizable atoms like dithiophosphoric acids can access radical, *open-shell* intermediates following photochemical oxidation of the dithiophosphate anion. The dithiophosphoryl radicals generated are highly electrophilic and have only observed exclusive application as downstream H-atom abstraction (HAA) catalysts.[18–22] Whereas traditional cocatalysts (ex. amines, thiols,[23] phosphines[24]) drive equilibria towards neutral and cationic intermediates, dithiophosphoric acids strongly favor the formation of the conjugate base, dithiophosphate anion (**Fig. VI-1C**). We hypothesized dithiophosphoric acids could engage complementary, ionic interactions with common cationic photocatalysts, providing an intuitive handle to modulate both photochemical events and subsequent transformations *en route* to the product. To this end, investigating the activity of dithiophosphoric acids, anions, and open-shell analogs across a suite of processes could elucidate their broader applications across photoredox catalysis.

We herein report the application of dithiophosphate catalysts in a suite of radical reactions (**Fig. VI-1D**). Despite the strong thermodynamic driving force for ionization of dithiophosphoric acids, we demonstrate these species indeed function as competent HAA and HAD catalysts. We realize their fate as HAD agents relies on steering their charge-driven association with cationic photocatalyst. This ionic association was easily modulated through solvent choice, affording precise control over photochemical activation of the dithiophosphate or substrate. Moreover, exploiting this modular association between photocatalyst and dithiophosphates unlocks new chemoselective photochemical paradigms governed by competitive photocatalyst quenching among the reactive species. Beyond group transfer reactivity, we also observed dithiophosphoryl radicals can undergo *anti*-Markovnikov addition into π-bonds, a complementary outcome to familiar Markovnikov reactivity by dithiophosphoric acids in polar paradigms. We deploy this covalent phenomenon to enable catalytic conjugate substitutions that afford olefin difunctionalizations. Collectively, these findings recast acid organocatalysts as versatile open-shell tools for photochemical reactions.

VI.3 – Results

- ***HAD is implicated in reversible group transfers***

To explore the utility of dithiophosphoric acids beyond HAA,[20] we first inspected whether it was theoretically practical for the catalysts to also mediate HAD, the microscopic reverse (**Fig. VI-2A**). Typically, synthetic strategies for H-atom transfer (HAT) involve distinct catalysts for HAA and HAD (**Fig. VI-2B**).[25,26] Such systems will rely on highly exergonic HAA and HAD events to drive the reactions forward. In contrast, reversible HAT catalysts operate over flattened thermodynamic landscapes in which polarity-matched interactions accelerate reaction kinetics. Rovis,[27] Wendlandt,[28,29] and others have shown that thiol catalysts like triphenylsilane thiol (S–H BDE = 88 kcal/mol) engage in reversible HAT reactions. Although dithiophosphoryl radicals readily abstract C–H bonds exceeding 90 kcal/mol, the calculated S–H bond strength of dithiophosphoric acids is surprisingly only 82 kcal/mol (see **Supplementary Table VI-19**). These

values imply that the observed HAA reactivity may be substantially endothermic (>10 kcal/mol). Accordingly, HAD from dithiophosphoric acids, should be thermodynamically favorable.

To probe whether dithiophosphoric acids can operate as reversible HAT agents, we investigated the C–H epimerization of vicinal diols as a benchmark transformation. Wendlandt and coworkers have reported that triphenylsilane thiol (Ph_3SiSH) catalyzes the *cis*-to-*trans* isomerization of 1,2-cyclohexane diol ***cis*-S1** via reversible HAT, ultimately siphoning the substrate to the thermodynamically preferred ***trans*-S1** isomer (obs. 74%, **Fig. VI-2C**).[29] Under analogous conditions, we found that tetrabutylammonium dithiophosphate **P-1** also promotes the *cis*-to-*trans* isomerization of **S1** in the presence of iridium photocatalyst **Ir-1** (obs. 64% in 48 hr, **Fig. VI-2C**, **Supplementary Tables VI-1-4**). Time-course analyses beginning from either pure ***cis*-S1** or ***trans*-S1** revealed convergence to a photostationary state comparable to previously reported equilibrium and thus consistent with reversible HAT reactivity (**Fig. VI-2D**).

We sought to gain further insight into the HAD event implicated. Two deuterium labeling experiments were designed to probe the intermolecular HAT process. In one crossover experiment between ***trans*-S1** and *D*-enriched ***trans*-S1**, no isotopic exchange between the two substrates was detected (**Fig. VI-2E**, **Supplementary Fig. VI-12**). Likewise, when *D*-enriched ***trans*-S1** was subjected to conditions permitting proton transfer with putative dithiophosphoric acid intermediates, no loss of isotopic enrichment was observed (**Fig. VI-2F**, **Supplementary Fig. VI-13**). These results contrast with Wendlandt's report with triphenylsilane thiol (Ph_3SiSH) in which significant isotope exchange is observed. As the isomerization is observed with **P-1**, these observations suggest that HAD by **P-1** occurs through rapid, in-cage rebound between the dithiophosphoric acid and the *C*-centered radical intermediate.

We further demonstrate rapid HAD by a radical clock experiment with **S2**. Under reversible HAT conditions, the *C*-centered radical from **S2** can undergo a ring-opening, cyclopropyl-carbinyl rearrangement, furnishing cyclic ketone **S3** (**Fig. VI-2G**). In the presence of dithiophosphate catalyst, C–H epimerization furnishes nearly half the ring-opened product than was observed with Ph_3SiSH after 24 hr. The diminished ring-opening by the dithiophosphate catalyst is consistent with a more rapid HAD rebound, outcompeting the kinetically favorable cyclopropyl carbinyl rearrangement ($k_{open} \sim 10^7\ s^{-1}$).[30] These experiments collectively demonstrate the emergent HAD capability of dithiophosphoric acids and their complementarity to other familiar thiol cocatalysts (also see **Supplementary Fig. VI-14** for TEMPO trap experiment).

- ***Engaging productive HAD by dithiophosphoric acids***

Although dithiophosphoric acids are capable of HAD within reversible HAT cascades, we challenged whether its rapid HAD could be productively engaged under reaction conditions that favor the ionization of the acid to the dithiophosphate anion. To investigate this possibility, we examined their application in a photochemical decarboxylation, converting a carboxylic acid **S4** to the corresponding alkane **S5** (**Fig. VI-3A**).[31] Typically, a base catalyst is added to promote the formation of the carboxylate anion. Then, photochemical oxidation drives decarboxylation to a *C*-

centered radical that is quenched by HAD from a thiol catalyst. The resulting thiyl radical is reduced by the photocatalyst to a basic thiolate. Finally, this species can accept a proton from the conjugate acid of the base catalyst, closing the catalytic cycle.

While most steps in this sequence are reasonable with dithiophosphoric acids, sluggish turnover was expected because of unfavorable proton transfer to the dithiophosphate anion. Surprisingly, evaluation of the decarboxylation reaction under a variety of conditions (**Supplementary Table VI-5-6**) revealed dithiophosphoric acid **P-2** can afford the transformation to **S5** in moderate yields (up to 59%, **Fig. VI-3A**). Solvent choice exerted a dramatic influence on reactivity whereby changing from MeCN to DCM completely suppressed reactivity (obs. 0%). Initially, we suspected the unfavorable thermodynamic equilibrium forming the requisite dithiophosphoric acid attenuates HAD and halts conversion. However, ^{31}P NMR analysis revealed full conversion of the acid precatalyst **P-2** to the conjugate base in both MeCN and DCM under the reaction conditions, indicating that differences in protonation equilibria are unlikely to account for the solvent-dependent reactivity (**Supplementary Fig. VI-48-49**).

Because the decarboxylation reaction sequence relies on oxidation of the carboxylate anion by the cationic photocatalyst, we next hypothesized competitive photocatalyst quenching by the anionic dithiophosphate may account for the suppressed reactivity in DCM. Stern-Volmer analysis indicates nearly 10-fold increase in K_{SV} in DCM by dithiophosphate compared to the carboxylic acid under reaction conditions (**Fig. VI-3B**, **Supplementary Table VI-17**). Only a 2-fold increase in K_{SV} was observed in MeCN. UV/Vis Job plot analysis indicates a clean 1:1 association between dithiophosphate anion and the cationic photocatalyst (**Fig. VI-3C, Supplementary Fig. VI-46**). Moreover, titration analyses show dithiophosphate associates to the photocatalyst nearly an order-of-magnitude stronger in DCM than in MeCN (dithiophosphate K_a = 5500 M^{-1} in DCM rel. to K_a = 650 M^{-1} in MeCN; **Fig. VI-3D**). In contrast, association between the substrate and the photocatalyst under reaction conditions is weaker and only comparable in MeCN (K_a = 100 M^{-1} in DCM and K_a = 270 M^{-1} in MeCN; **Fig. VI-3D**, **Supplementary Table VI-20**). After all, the higher dielectric constant of MeCN (ϵ = 38) should support the solvation of charged species more effectively than DCM (ϵ = 9). These observations indicate any emergent failures of dithiophosphoric acids behaving as HAD agents arise from competitive photocatalyst quenching by the corresponding dithiophosphate anion. Simple solvent selection can mitigate this ionic interaction to restore productive electron transfer steps.

To validate this mechanistic insight, we examined a second HAD-dependent transformation, photochemical hydroamination (**Fig. VI-3E**). Here, oxidation of alkene **S4** by the photocatalyst generates a radical cation that undergoes nucleophilic attack with an amine to form the C–N bond and an adjacent *C*-centered radical. Subsequent HAD, photocatalyst turnover, and proton transfer complete the catalytic cycle furnishing **S5**. Because the substrate is neutral prior to oxidation, the reaction does not benefit from ionic interactions promoting productive photocatalyst association. Mitigating competitive association to the photocatalyst by deploying dithiophosphoric acid rather than dithiophosphate anion should permit the desired reactivity. In fact, evaluation of a benchmark

intramolecular photochemical hydroamination[32] revealed dithiophosphoric acid is quite competent, affording hydroamination yields up to 77% in MeCN (**Supplementary Table VI-7-9**). In contrast, the corresponding dithiophosphate salt strongly binds to the photocatalyst and is significantly less productive (only up to 21%, **Supplementary Table VI-17**, **VI-21**, **Supplementary Fig. VI-46**). These results reinforce that modulating the association of dithiophosphate species with the cationic photocatalyst can greatly influence electron-transfer kinetics, unlocking HAD-dependent reactivity even under unfavorable thermodynamic landscapes.

- ***Catalytic olefin functionalization mediated by dithiophosphoryl radicals***

In the photochemical hydroamination above, we not only observed attenuated yields of **S7**, but also plateaued conversion with high loadings of dithiophosphate anion **P-1**. Under these conditions, competitive quenching by the anion should generate increased concentrations of dithiophosphoryl radical. At such high radical concentrations, dimerization to a disulfide-like species might be expected. However, ^{31}P NMR analysis of the photochemical hydroamination reaction mixture shows the formation of a distinct species appearing as a pair of signals near +110 ppm (note – disulfide expected at +103 ppm, anion expected at +133 ppm, **Fig. VI-4A**, **Supplementary Fig. VI-50**).

To account for the loss of catalytic activity in the reaction, we hypothesized the electrophilic dithiophosphoryl radicals may be forming adducts with the olefin substrate. After all, these reactions involve electron-rich styrenes which are polarity-matched olefins for π-addition.[33] In the photochemical hydroamination, radical addition of the catalyst to the substrate **S6** would furnish a diastereomeric adduct, producing a pair of ^{31}P signals. To validate this hypothesis, α-methylstyrene was subjected to the reaction conditions of the photochemical hydroamination (**Fig. VI-4B**). If the dithiophosphoryl radical adds to the olefin, the expected *anti*-Markovnikov product would furnish an analogous pair of ^{31}P signals. Consistent with this hypothesis, the reaction of dithiophosphate with α-methylstyrene furnishes the analogous adduct, indicated by the pair of ^{31}P signals and further confirmed by ^{31}P-^{1}H coupling that resolves into two triplets (**Supplementary Fig. VI-51** for complete spectra).

This covalent intermediate reflects a distinct opportunity for catalysis. Our group has previously reported *polar* olefin addition by dithiophosphoric acid can catalyze olefin hydroamination through a conjugate substitution pathway **(Fig. VI-4C**).[34] Now, in this report, the dithiophosphoryl radical adds into an olefin furnishing an *anti*-Markovnikov adduct. Conjugate substitution by another radical would release the covalently bound dithiophosphate, affording a complementary 1,2-difunctionalization compared to the 1,1'-difunctionalization of the polar pathway.

We demonstrate the catalytic utility of this homolytic conjugate substitution using the annulation of vinyl cyclopropane **S8** with electron-rich olefins, a canonical transformation previously investigated by Maruoka,[35] Miller,[36–38] and others[39] (**Fig. VI-4D**). Under our conditions, dithiophosphoryl radicals are generated from dithiophosphate precatalysts in the presence of an

initiator (ex. dibenzoyl peroxide). The dithiophosphoryl species then rapidly promotes formation of the cyclopentane adduct **S9** in quantitative yields (**Supplementary Table VI-10-13)**. Time-course analysis reveals rapid activity by the dithiophosphoryl radical under irradiation, completely converting the vinyl cyclopropane to the cyclopentane product in under 20 minutes at room temperature (**Fig. VI-4E**). Even with a 10-fold reduction in dithiophosphate loading, cyclopentane **S9** is obtained in 73% yield after 75 minutes (0.6 mol% **P-1**). Control reactions lacking either radical initiator or dithiophosphate furnish no cyclopentane product, supporting a mechanism involving radical addition followed by conjugate substitution. These results highlight the rapid reactivity of dithiophosphoryl radicals and demonstrate their catalytic utility beyond group transfer, downstream processes.

- ***Competitive photocatalyst association enables new paradigm***

We have thus far identified the capabilities of dithiophosphoric acids and radicals across a suite of downstream, photochemical reactions. Distinct from canonical conditions, the dithiophosphate anion competitively associates and quenches the excited state of cationic photocatalyst employed. We questioned whether competitive photocatalyst quenching alone could be leveraged as a design principle, independent of any downstream activity.

Certain photochemical transformations rely exclusively on electron transfer to and from the photocatalyst to afford the desired product (**Fig. VI-5A**). In such systems, both the substrate and product are redox-active and can each interact with the photocatalyst. A canonical example is the photochemical [2+2] cycloaddition of electron-rich styrenes.[40] In this transformation, oxidation of the olefin generates a radical cation that undergoes rapid dimerization to generate a cyclobutanyl radical cation. Subsequent reduction furnishes the desired cyclobutane product. However, the product is not inert – it can also be oxidized, returning to the prior reactive intermediate. Here, the cyclobutanyl radical cation can undergo cycloreversion, siphoning product back into starting material and enabling deleterious decomposition pathways.

Given the longstanding interest in this reaction class, significant effort has been devoted to mitigating product oxidation and cycloreversion (**Fig. VI-5B**). A prominent strategy, reported by Nicewicz and coworkers relies on electron-relay catalysts that differentiate between the substrate and product based on oxidation potentials (+1.1 V and +1.3 V vs. SCE, respectively).[41] A similar voltage-gated approach by Yoon and coworkers instead carefully designs the photocatalyst to operate within the narrow redox window (ex. $[Ru(bpm)_3]^{2+}$ with excited state potential, +1.2 V vs. SCE).[42] While these strategies exploit voltage as a thermodynamic criterion to gate redox events, we instead considered whether this selectivity could be recast into a kinetic paradigm defined by photocatalyst quenching propensity (**Fig. VI-5C)**. Through judicious addition of dithiophosphate anion, we could modulate the relative rate of excited-state quenching, permitting the oxidation of the substrate early in the reaction while inhibiting oxidation of the product later.

Using the photochemical [2+2] of *trans*-anethole **S10** as a benchmark, we observed the starting material only moderately quenches the pyrylium photocatalyst (K_{SV} = 130 M^{-1}), while

dithiophosphate drastically quenches (K_{SV} = 6,500 M^{-1}, **Fig. VI-5C**, **Supplementary Table VI-17**). These relative quenching efficiencies suggest that competitive behavior can be achieved within a narrow concentration regime. At **P-1** loadings below 2 mol%, photocatalyst quenching should be dominated by the styrene substrate, enabling productive cycloaddition. As the reaction progresses and substrate is consumed, the relative contribution of dithiophosphate to excited-state quenching increases, attenuating further oxidation events. Consistent with this model, systematic variation of **P-1** loading revealed a maximum in product **S11** formation at intermediate loadings, with diminished conversion of **S10** only observed at higher concentrations of **P-1** (**Fig. VI-5D**, **Supplementary Table VI-14-15**). In the absence of dithiophosphate, cyclobutane **S11** forms as expected, but with diminished mass balance and yield, indicating product decomposition. Across the conditions screened, competitive quenching by the dithiophosphate anion helped furnish cyclobutane **S11** with a 2.5-fold improvement over the photocatalyst alone.

We anticipated that this effect would not be observed in a photochemical [2+2] where the product lacks an electron-rich arene as a handle to initiate oxidative cycloreversion. To this end, we selected electron-rich enamine **S14** which forms the corresponding cyclobutane **S15**. Under the reaction conditions, similar yields and mass balance were detected at both 0% and 0.3 mol% **P-1** loading (the lowest screened), while higher **P-1** loadings only lead to attenuated yields (**Fig. VI-5E, Supplementary Table VI-17**). This data is consistent with the dithiophosphate only acting as a competitive quencher, mitigating the efficiency of substrate activation.

We then sought to validate the utility of the paradigm with the electron-rich styrene **S16**, an analog to *trans*-anethole bearing an additional methoxy group. We observed **S16** quenches the pyrylium photocatalyst with similar efficiency to the model substrate **S10** (K_{SV} = 120 M^{-1}). As expected, **S16** also displays an advantageous response with dithiophosphate, where intermediate **P-1** loadings promote access to the desired cyclobutane **S17** (1.5-fold improvement, **Fig. VI-5F**, **Supplementary Table VI-18**). Collectively, these observations introduce the utility of dithiophosphate anions beyond traditional photochemical and downstream reaction paradigms, showcasing how their anionic character can unlock a kinetic phenomenon for chemoselective photochemical reactions.

VI.4 – Discussion and Outlook

Photochemical reactions rely on the careful orchestration of multiple reactive open-shell species *en route* to the desired product. Failures often arise from unproductive electron transfer and deleterious side reactivity. In this work, we demonstrate dithiophosphoric acids are also operational organocatalysts across a suite of radical reactions. Their pronounced acidity endows the catalysts with predominantly anionic character, enabling competitive association to cationic photocatalysts. By mitigating this ionic interaction through solvent and catalyst choice, conventional photochemical transformations are accessible, while deliberate exploitation of this association unlocks new paradigms governed by competitive photocatalyst quenching. In addition, we show that dithiophosphoryl radicals can generate covalent adducts, and this reactivity can be

leveraged in catalytic, conjugate substitutions to furnish outcomes complementary to analogous polar processes. Altogether, our findings recast acid organocatalysts as intuitive tools for photochemical reactions. Because dithiophosphoric acids are proximal to chiral Brønsted acid organocatalysts, these catalysts also provide a promising entry point into stereoselective, open-shell transformations. We are actively investigating their application across the suite of radical reactions described and plan to report our findings in due course.

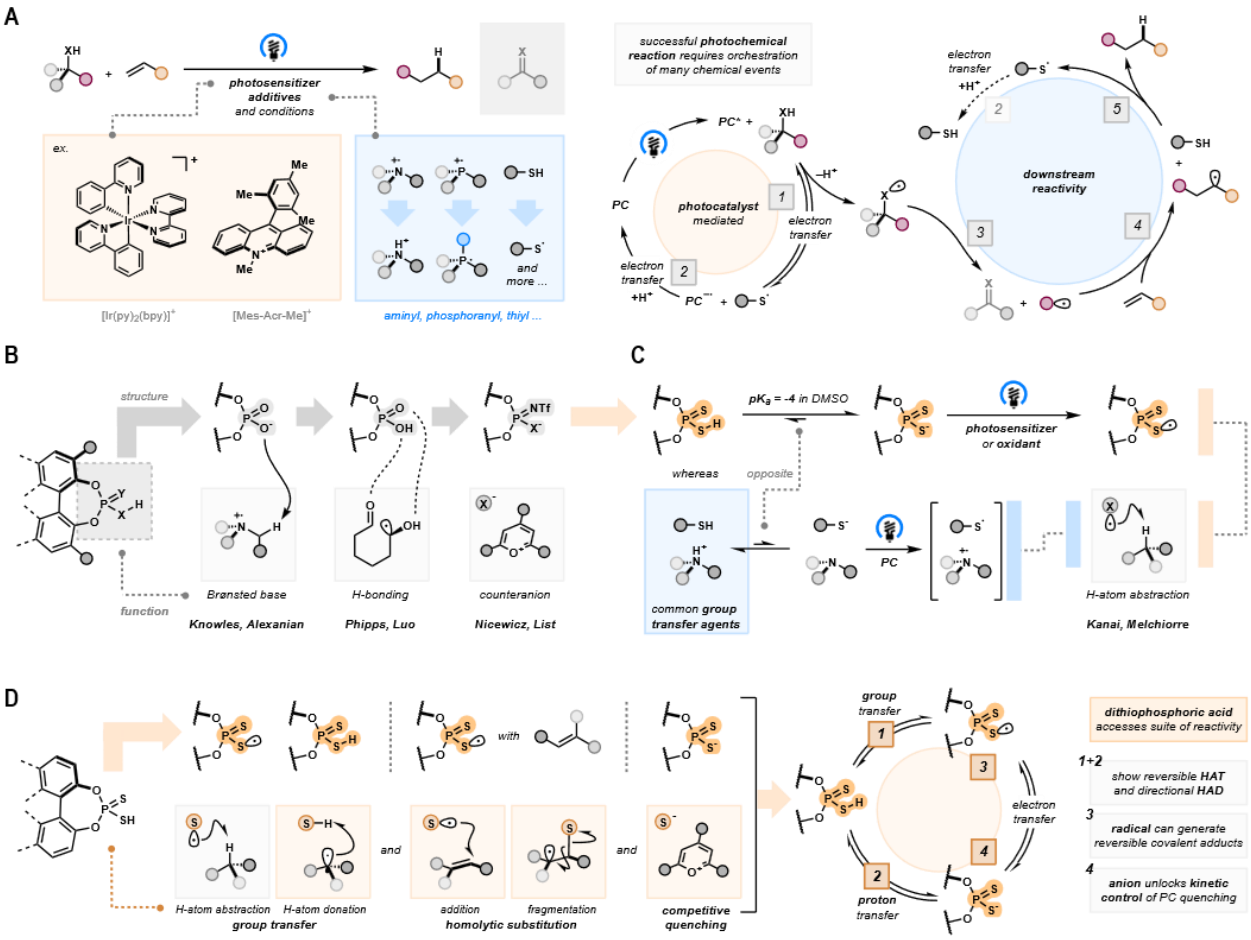


Fig. VI-1 | Photochemical paradigms rely on the collective operation of many open-shell steps. (**a**) Elementary steps in photochemical reactions include photoinduced energy/electron transfer, group transfers, additions, fragmentations, and more. (**b**) Brønsted acid organocatalysts (BAOCs) have activity closely tied to their acidity. (**c**) Common photoredox cocatalysts generate neutral or cationic intermediates, while dithiophosphoric acids thermodynamically prefer the formation of anionic conjugate base. (**d**) *This work* – dithiophosphoric acids access complementary, charge-driven intuition for the design of many photochemical paradigms involving redox, group transfer, addition, and fragmentation.

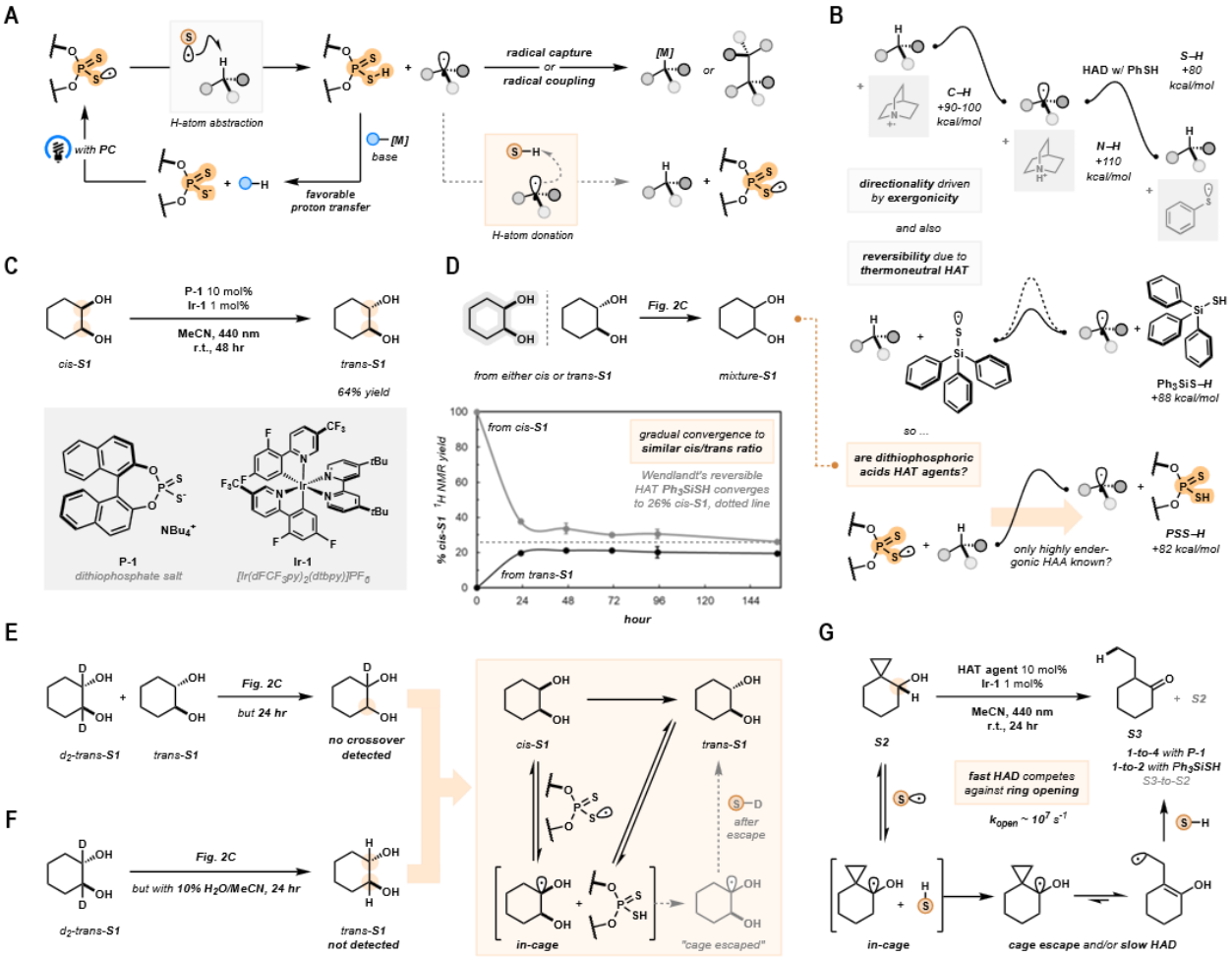


Fig. VI-2 | C–H epimerization implicates HAD by dithiophosphoric acids. (a) Dithiophosphoric acids are typically neutralized prior to activity beyond HAA. **(b)** Reversible HAT paradigms require neutral thermodynamic landscapes between HAA and HAD reactions. **(c)** Dithiophosphates mediate C–H epimerization of vicinal diols akin to known, reversible HAT agents like Ph_3SiSH. **(d)** Time course analyses reveal the convergent C–H epimerization to a photostationary state comparable to the thermodynamic equilibrium between *cis* and *trans* isomers of **S1**. Isotope labeling experiments like **(e)** crossover and **(f)** enrichment suggest HAD proceeds through rapid, in-cage rebound between dithiophosphoric acid and the putative *C*-centered radical intermediate. **(g)** Radical clock experiment supports expectation for rapid HAD by dithiophosphoric acids relative to familiar HAT agents like Ph_3SiSH.

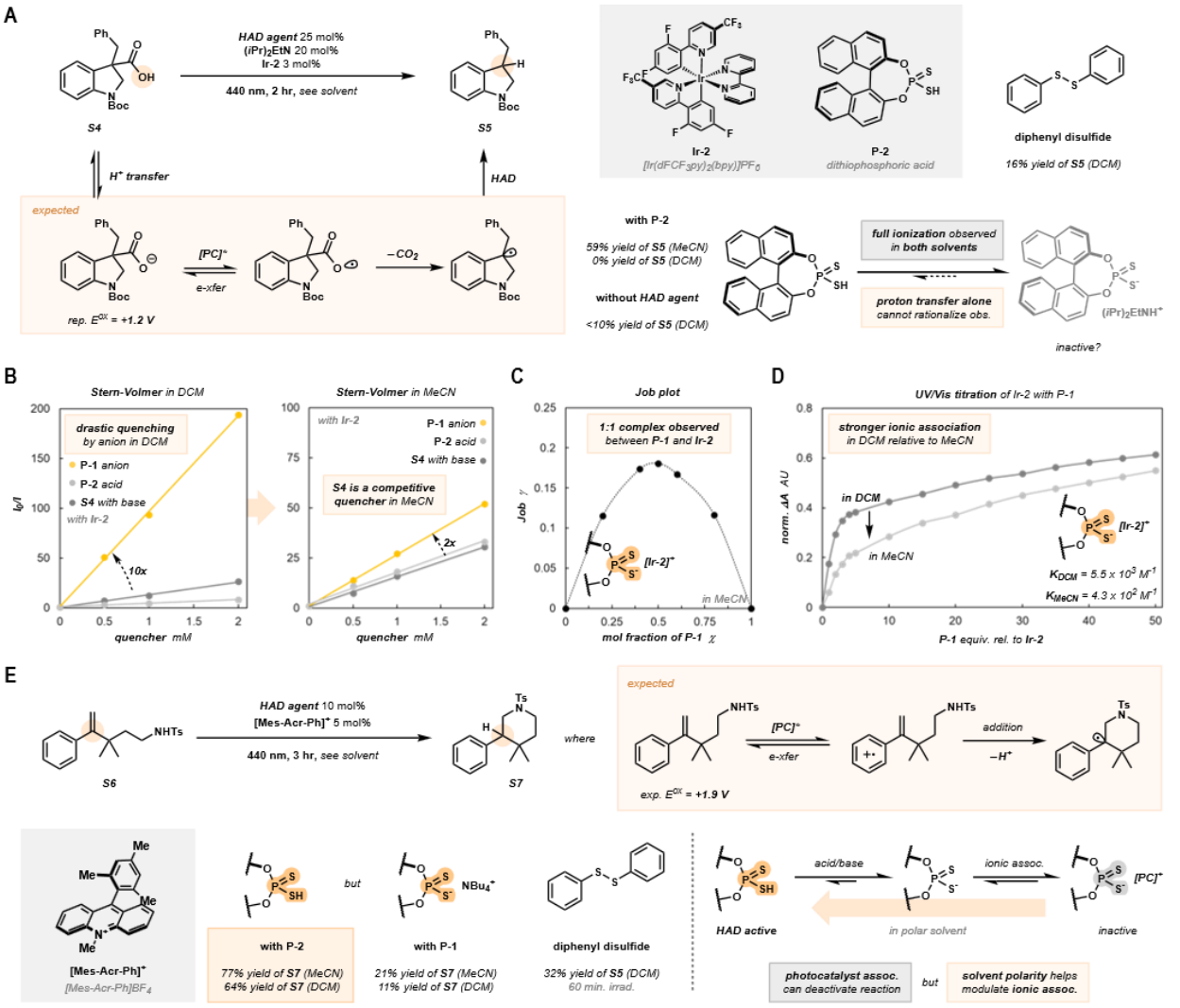


Fig. VI-3 | Productive HAD reactivity by dithiophosphoric acids requires mitigation of photocatalyst association. (a) Photochemical decarboxylation leverages three catalysts to orchestrate the conversion of a carboxylic acid to an alkane. Dithiophosphates are only suitable for HAD in a photochemical decarboxylation in specific solvents. **(b)** Stern-Volmer quenching data implicates dithiophosphates are exceptional quenchers relative to the substrate in DCM relative to MeCN. UV/Vis titration of dithiophosphate **P-1** with **Ir-2** indicates **(c)** 1:1 binding from a Job plot analysis, and **(d)** strong binding by the dithiophosphate in DCM, but attenuated binding in MeCN. **(e)** Neutral intermediates in a photochemical hydroamination must outcompete with anionic species for cationic photocatalyst quenching. Still, reactivity can be rescued through judicious choice of dithiophosphoric acid precatalyst and solvent.

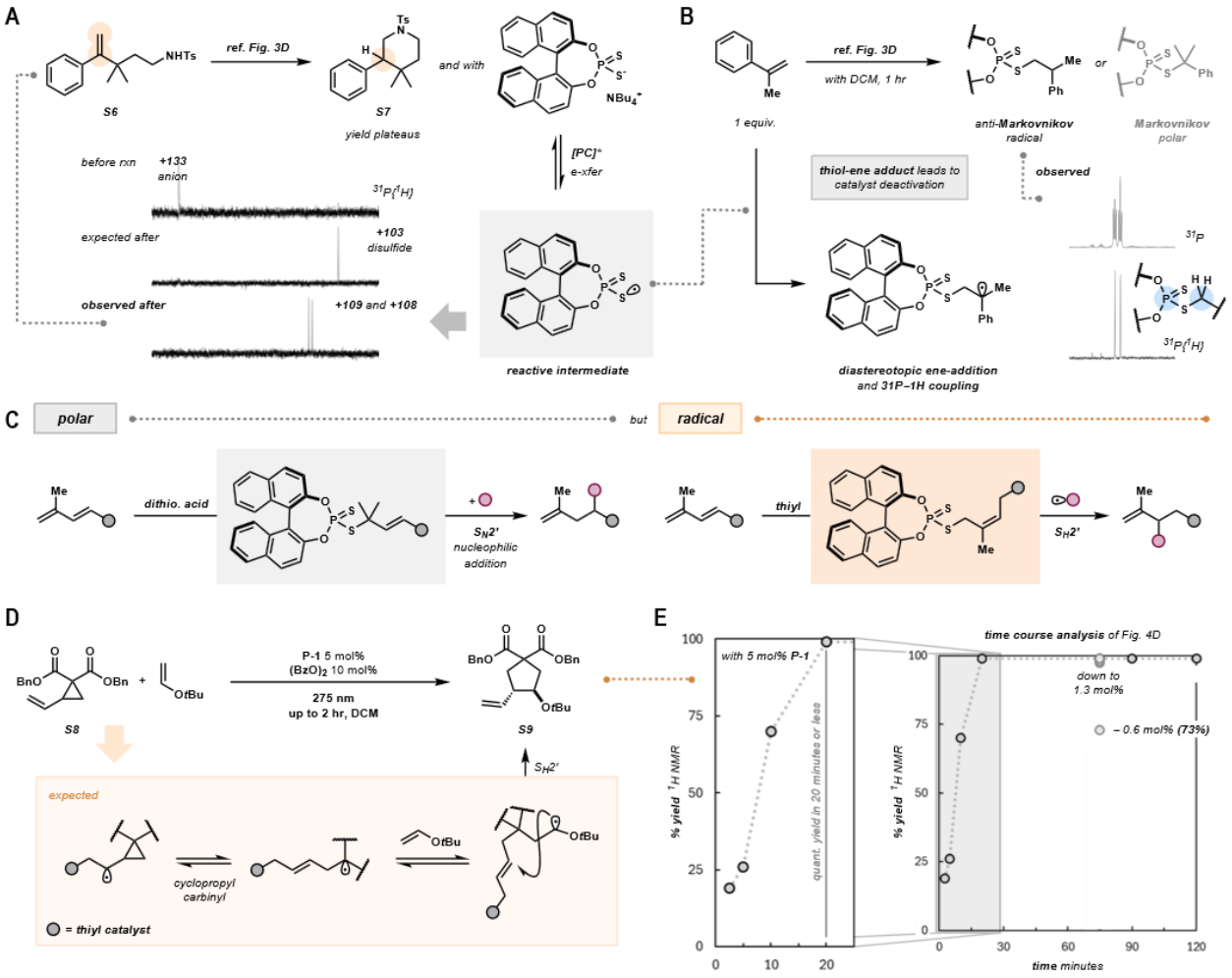


Fig. VI-4 | Covalent adducts from radical additions can be leveraged for organocatalyzed homolytic substitutions of olefins. (a) Loss of catalytic activity in the benchmark photochemical hydroamination of **S6** is attributed to dithiophosphoryl radical adduct to **S7**. **(b)** Model substrate α-methylstyrene validates radical addition by thiyl-like catalytic intermediate under reaction conditions. **(c)** Covalent adducts are strategic for conjugate substitutions in *polar* paradigms and could be leveraged for analogous conjugate *homolytic* substitutions in *radical* paradigms. **(d)** Open-shell catalysis leverages covalent adducts for formal [3+2] annulations of vinyl cyclopropanes. **(e)** Dithiophosphoric acids are highly active catalysts in conjugate homolytic substitution reactions.

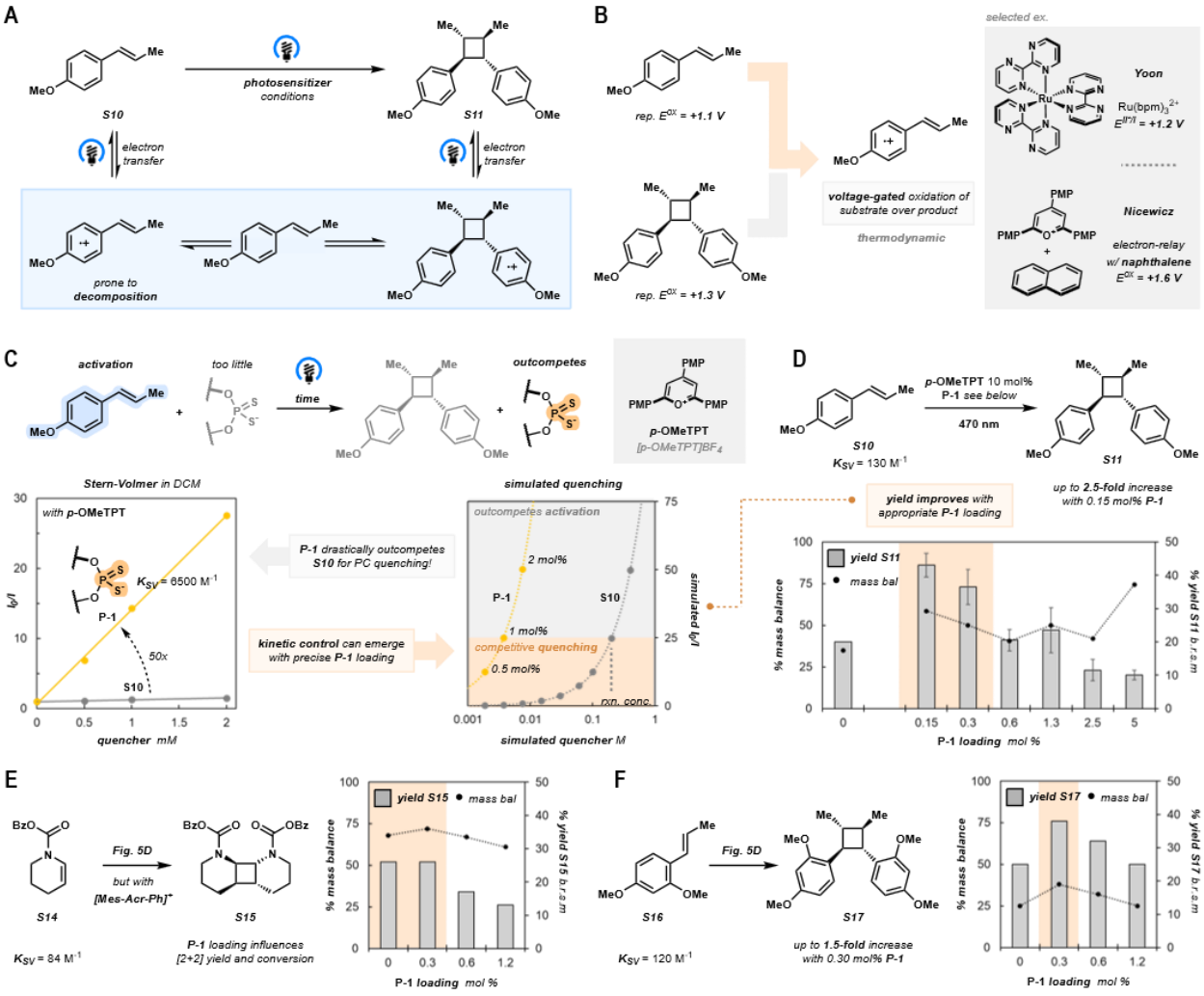


Fig. VI-5 | Competitive photocatalyst quenching by dithiophosphate anion unlock emergent chemoselectivity paradigms. (a) Photochemical [2+2] cycloadditions of electron-rich styrenes navigate redox activation of both substrate and product. **(b)** Voltage-gated phenomena are a prominent thermodynamic strategy to discriminate substrate and product activation. **(c)** Competitive photocatalyst quenching between the substrate, product, and an additive could unlock a kinetic chemoselectivity paradigm. Stern-Volmer quenching of pyrylium by substrate **S10** is substantially less efficient than dithiophosphate **P-1**, revealing a concentration regime where kinetic control can emerge. **(d-f)** Systematic variation of dithiophosphate loading demonstrates competitive quenching-based control where substrate activation is permitted, while product activation is attenuated, leading to advantageous yield and mass balance (see **Supplementary Table 14-18** for complete details).

VI.5 – References

(1) Shaw, M. H.; Twilton, J.; MacMillan, D. W. C. Photoredox Catalysis in Organic Chemistry. *J. Org. Chem.* **2016**, *81* (16), 6898–6926.

(2) Romero, N. A.; Nicewicz, D. A. Organic Photoredox Catalysis. *Chem. Rev.* **2016**, *116* (17), 10075–10166.

(3) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.

(4) Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M. Complete Field Guide to Asymmetric BINOL-Phosphate Derived Brønsted Acid and Metal Catalysis: History and Classification by Mode of Activation; Brønsted Acidity, Hydrogen Bonding, Ion Pairing, and Metal Phosphates. *Chem. Rev.* **2014**, *114* (18), 9047–9153.

(5) Akiyama, T.; Mori, K. Stronger Brønsted Acids: Recent Progress. *Chem. Rev.* **2015**, *115* (17), 9277–9306.

(6) Mahlau, M.; List, B. Asymmetric Counteranion-Directed Catalysis: Concept, Definition, and Applications. *Angewandte Chemie International Edition* **2013**, *52* (2), 518–533.

(7) Das, S.; Zhu, C.; Demirbas, D.; Bill, E.; Kanta De, C.; List, B. Asymmetric Counteranion-Directed Photoredox Catalysis. *Science* **2023**, *379*, 494–499.

(8) Zhu, C.; Das, S.; Guin, A.; De, C. K.; List, B. Organocatalytic Regio- and Stereoselective Cyclopropanation of Olefins. *Nat Catal* **2025**, *8* (5), 487–494.

(9) Morse, P. D.; Nguyen, T. M.; Cruz, C. L.; Nicewicz, D. A. Enantioselective Counter-Anions in Photoredox Catalysis: The Asymmetric Cation Radical Diels-Alder Reaction. *Tetrahedron* **2018**, *74* (26), 3266–3272.

(10) Shin, N. Y.; Ryss, J. M.; Zhang, X.; Miller, S. J.; Knowles, R. R. Light-Driven Deracemization Enabled by Excited-State Electron Transfer. *Science* **2019**, *366* (6463), 364–369.

(11) Roos, C. B.; Demaerel, J.; Graff, D. E.; Knowles, R. R. Enantioselective Hydroamination of Alkenes with Sulfonamides Enabled by Proton-Coupled Electron Transfer. *J. Am. Chem. Soc.* **2020**, *142* (13), 5974–5979.

(12) Xu, E. Y.; Werth, J.; Roos, C. B.; Bendelsmith, A. J.; Sigman, M. S.; Knowles, R. R. Noncovalent Stabilization of Radical Intermediates in the Enantioselective Hydroamination of Alkenes with Sulfonamides. *J. Am. Chem. Soc.* **2022**, *144* (41), 18948–18958.

(13) Gillespie, J. E.; Fanourakis, A.; Phipps, R. J. Strategies That Utilize Ion Pairing Interactions to Exert Selectivity Control in the Functionalization of C–H Bonds. *J. Am. Chem. Soc.* **2022**, *144* (40), 18195–18211.

(14) Proctor, R. S. J.; Davis, H. J.; Phipps, R. J. Catalytic Enantioselective Minisci-Type Addition to Heteroarenes. *Science* **2018**, *360* (6387), 419–422.

(15) Farney, E. P.; Chapman, S. J.; Swords, W. B.; Torelli, M. D.; Hamers, R. J.; Yoon, T. P. Discovery and Elucidation of Counteranion Dependence in Photoredox Catalysis. *J. Am. Chem. Soc.* **2019**, *141* (15), 6385–6391.

(16) Geunes, E. P.; Meinhardt, J. M.; Wu, E. J.; Knowles, R. R. Photocatalytic Anti-Markovnikov Hydroamination of Alkenes with Primary Heteroaryl Amines. *J. Am. Chem. Soc.* **2023**, *145* (40), 21738–21744.

(17) Vojáčková, P.; Jacobsen, E. N. An Anion-Binding Approach to Enantioselective Photoredox Catalysis. *Science* **2025**, *390* (6775), 831–836.

(18) Le Saux, E.; Georgiou, E.; Dmitriev, I. A.; Hartley, W. C.; Melchiorre, P. Photochemical Organocatalytic Functionalization of Pyridines via Pyridinyl Radicals. *J. Am. Chem. Soc.* **2023**, *145* (1), 47–52.
(19) Le Saux, E.; Zanini, M.; Melchiorre, P. Photochemical Organocatalytic Benzylation of Allylic C–H Bonds. *J. Am. Chem. Soc.* **2022**, *144* (3), 1113–1118.
(20) Fuse, H.; Mitsunuma, H.; Kanai, M. Catalytic Acceptorless Dehydrogenation of Aliphatic Alcohols. *J. Am. Chem. Soc.* **2020**, *142* (9), 4493–4499.
(21) Nakao, H.; Hassan, M. M. M.; Nakamura, Y.; Toyobe, M.; Higashi, M.; Mitsunuma, H.; Kanai, M. Visible Light–Driven Stereodivergent Allylation of Cyclic Hemiacetals with Butene for Polypropionate Synthesis. *Science* **2025**, *390* (6770), 272–278.
(22) Fuse, H.; Irie, Y.; Fuki, M.; Kobori, Y.; Kato, K.; Yamakata, A.; Higashi, M.; Mitsunuma, H.; Kanai, M. Identification of a Self-Photosensitizing Hydrogen Atom Transfer Organocatalyst System. *J. Am. Chem. Soc.* **2022**, *144* (14), 6566–6574.
(23) Dénès, F.; Pichowicz, M.; Povie, G.; Renaud, P. Thiyl Radicals in Organic Synthesis. *Chem Rev* **2014**, *114* (5), 2587–2693.
(24) Rossi-Ashton, J. A.; Clarke, A. K.; Unsworth, W. P.; Taylor, R. J. K. Phosphoranyl Radical Fragmentation Reactions Driven by Photoredox Catalysis. *ACS Catal.* **2020**, *10* (13), 7250–7261.
(25) Capaldo, L.; Ravelli, D.; Fagnoni, M. Direct Photocatalyzed Hydrogen Atom Transfer (HAT) for Aliphatic C–H Bonds Elaboration. *Chem. Rev.* **2022**, *122* (2), 1875–1924.
(26) Feng, A. G.; Pinto Pereira Junior, M. V.; Miller, S. J.; Knowles, R. R. Asymmetric Hydrogen Atom Transfer. *ACS Catal.* **2026**, *16* (2), 844–865.
(27) Shen, Y.; Funez-Ardoiz, I.; Schoenebeck, F.; Rovis, T. Site-Selective α-C–H Functionalization of Trialkylamines via Reversible Hydrogen Atom Transfer Catalysis. *J. Am. Chem. Soc.* **2021**, *143* (45), 18952–18959.
(28) Occhialini, G.; Wendlandt, A. E. Kinetics, Thermodynamics, and Emergence in Stereoediting Reactions. *Acc. Chem. Res.* **2025**, *58* (14), 2255–2268.
(29) Zhang, Y.-A.; Gu, X.; Wendlandt, A. E. A Change from Kinetic to Thermodynamic Control Enables Trans-Selective Stereochemical Editing of Vicinal Diols. *J. Am. Chem. Soc.* **2022**, *144* (1), 599–605.
(30) Auclair, K.; Hu, Z.; Little, D. M.; Ortiz de Montellano, P. R.; Groves, J. T. Revisiting the Mechanism of P450 Enzymes with the Radical Clocks Norcarane and Spiro[2,5]Octane. *J. Am. Chem. Soc.* **2002**, *124* (21), 6020–6027.
(31) Xu, Y.; Shen, C.; Dong, K. Enantioselective Decarboxylative Hydrogen-Atom Transfer Reaction. *J. Am. Chem. Soc.* **2025**.
(32) Tang, L.; Shen, C.; Hao, S.; Dong, K. A Type of Chiral C2-Symmetric Arylthiol Catalyst for Highly Enantioselective Anti-Markovnikov Hydroamination. *J. Am. Chem. Soc.* **2024**, *146* (23), 16248–16256.
(33) Garwood, J. J. A.; Chen, A. D.; Nagib, D. A. Radical Polarity. *J. Am. Chem. Soc.* **2024**, *146* (41), 28034–28059.
(34) Shapiro, N.; Rauniyar, V.; Hamilton, G.; Wu, J.; Toste, D. Asymmetric Additions to Dienes Catalysed by a Dithiophosphoric Acid | Nature. *Nature* **2011**, *470*, 245–249.
(35) Hashimoto, T.; Kawamata, Y.; Maruoka, K. An Organic Thiyl Radical Catalyst for Enantioselective Cyclization. *Nature Chem* **2014**, *6* (8), 702–705.

(36) Ryss, J. M.; Turek, A. K.; Miller, S. J. Disulfide-Bridged Peptides That Mediate Enantioselective Cycloadditions through Thiyl Radical Catalysis. *Org. Lett.* **2018**, *20* (6), 1621–1625.
(37) Turek, A. K.; Sak, M. H.; Miller, S. J. Kinetic Analysis of a Cysteine-Derived Thiyl-Catalyzed Asymmetric Vinylcyclopropane Cycloaddition Reflects Numerous Attractive Noncovalent Interactions. *J. Am. Chem. Soc.* **2021**, *143* (39), 16173–16183.
(38) Mason, S. M.; Miller, S. J. Peptidyl Glycosyl Thiols and Disulfides for Enantioselective Hydrogen Atom Transfer (HAT) and Thiyl Radical Catalysis. *J. Org. Chem.* **2025**, *90* (43), 15399–15403.
(39) Miura, K.; Fugami, K.; Oshima, K.; Utimoto, K. Synthesis of Vinylcyclopentanes from Vinylcyclopropanes and Alkenes Promoted by Benzenethiyl Radical. *Tetrahedron Letters* **1988**, *29* (40), 5135–5138.
(40) Poplata, S.; Tröster, A.; Zou, Y.-Q.; Bach, T. Recent Advances in the Synthesis of Cyclobutanes by Olefin [2 + 2] Photocycloaddition Reactions. *Chem. Rev.* **2016**, *116* (17), 9748–9815.
(41) Riener, M.; Nicewicz, D. A. Synthesis of Cyclobutane Lignans via an Organic Single Electron Oxidant–Electron Relay System. *Chem. Sci.* **2013**, *4* (6), 2625–2629.
(42) Ischay, M. A.; Ament, M. S.; Yoon, T. P. Crossed Intermolecular [2 + 2] Cycloaddition of Styrenes by Visible Light Photocatalysis. *Chem. Sci.* **2012**, *3* (9), 2807.

Appendix I – General Experimental and Computational Methods

Table of Contents:

I.1 – Materials and Methods **115**
I.2 – General Calculation and Statistical Considerations **116**
I.3 – Molecular descriptor calculation workflow **117**
I.4 – Non-covalent interaction calculation workflow **122**
I.5 – References **127**

VI.6 – Materials and Methods

- ***Reaction Setup and Purification***

Unless otherwise noted, all reactions were performed at room temperature under ambient atmosphere with Teflon-coated magnetic stir bars. Photochemistry was conducted with 456 nm Kessil (blue LED) using the experimental setup as indicated below in *Photochemistry Setup*. Tetrahydrofuran (THF), diethyl ether (Et_2O), toluene (PhMe), dichloromethane (DCM), and triethylamine (NEt_3) were dried by passing the degassed solvents through activated alumina columns under an argon atmosphere. Reaction progress was monitored by thin layer chromatography (TLC) using Merck 50 pre-coated silica gel plates with F254 indicator and visualized using UV light (254 nm). Reactions were purified using either flash column chromatography with Fischer Silica Gel 60 Å, 230X 400 mesh or CombiFlash-*Rf* with TeledyneISCO RediSep Silica Columns (60 Å porosity, 230-400 mesh, 40-63 μm particle), reverse-phase flash chromatography with Teldyne-ISCO RediSep-*Rf* C18 silica columns (60 Å porosity, 230-400 mesh, 40-63 μm particle), or preparative TLC with the above Silicycle Siliaplate™ TLC plates.

All commercially available substrates were purchased from Sigma-Aldrich, Fischer Scientific, Ambeed, or Combi-Blocks and used without further purification, unless otherwise noted.

- ***Analysis and Instrumentation***

^{1}H, ^{13}C, ^{19}F, and ^{31}P nuclear magnetic resonance (NMR) spectra were recorded on either a JEOL-400 or AVB-400, AVQ-400, NEO-500, or AV-600 Bruker spectrometers operating at 400 MHz, 500 MHz, or 600 MHz at ambient temperature. When available, some ^{13}C NMR spectra were collected by ^{1}H- and/or ^{19}F-decoupled pulse programs. Chemical shift is reported in parts per million (ppm) from either $CDCl_3$ (^{1}H NMR δ: 7.26 ppm, ^{13}C NMR δ: 77.1 ppm) or DMSO-d_6 (^{1}H NMR δ: 2.50 ppm, ^{13}C NMR δ: 39.5 ppm). ^{1}H NMR multiplicities are reported as: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), with the respective coupling constants in Hz. We

thank Dr. Hasan Celik and UC Berkeley's NMR facility in the College of Chemistry (CoC NMR) for spectroscopic assistance. Instruments in the CoC-NMR are supported in part by National Institute of Health grant S10OD024998.

High resolution mass spectra were recorded by an Agilent LC-MS TOF and reported in *m/z* of the indicated ion detected. Enantiomeric excess (% *ee*) was measured by chiral high performance liquid chromatography (HPLC) with Shimadzu Prominence HPLC and the indicated Diacel ChiralCel column.

VI.7 – General Calculation and Statistical Considerations

- ***General calculation specifications***

Calculations were remotely conducted using computing clusters at the Molecular Graphics and Computation Facility (MGCF) at the University of California, Berkeley (NIH S10OD034382). For quantum-level descriptor calculations, priority was placed for time-efficient calculations. Thus, semiempirical quantum mechanical method by Grimme's extended tight binding (xTB) package was used (specifically, xtb version 6.6.1 was employed). Conformational searching was conducted with Grimme's conformer-rotamer ensemble sampling tool (CREST) which is powered by xTB-based semiempirical quantum mechanical methods (specifically, crest version 2.12 was employed). In-house Python scripts were powered by a personal, RDKit conda environment (*Python* == 3.10.12, *rdkit* == 2023.09.2, installation directions are available at RDKit's homepage, https://www.rdkit.org/docs/Install.html). Plots of the linear correlations were generated using *Microsoft Excel*.

- ***General correlation screening***

Nearly all correlations reported are not the *best* correlation identified. In practice, an in-house Python workflow titled *mangos.py* considers a *domain* of molecular descriptors and screens linear correlations to useful output (ex. *yield*, *kcal/mol* for *enantioselectivity, diastereoselectivity,* etc.). First, univariate correlations to the output are considered. Then, bivariate correlations to the output are considered, where pairs of variables with significant covariance (greater than 0.3) are ignored. These correlations are scored by (1) Pearson's R^2 metric as measure of goodness-of-fit, and (2) a *leave-one-out* cross-validated R^2 reported as LOO-R^2. For bivariate linear regression models, a *leave-two-out* cross-validated R^2 was reported as L2O-R^2. This metric was calculated using the *RepeatedKFold* function from *sklearn* where the n_splits is set accordingly for *leave-two-out*, n_repeats is set redundantly to '100' to ensure instability in the random folds do not bias the L2O-R^2 value, and the random_state is set to '1' for reproducibility. The average R^2 over each fold is defined as the L2O-R^2. After, the top-20 correlations based on the highest LOO-R^2 are recorded for further analysis. A secondary analysis invokes chemical intuition to select correlations built from the most *chemically meaningful* feature(s).

The selected correlation is then reported with the computed R^2, LOO-R^2, and the *mean absolute error* (MAE) as a redundant, quantitative measure of error. MAE is computed as the average of the *absolute value of residuals* between the predictions made by the correlation and the true, experimental value. This metric can help provide more dimension to the reported correlation by quantifying the error that *goodness-of-fit* by R^2 only qualitatively captures.

- ***General univariate threshold detection***

Univariate threshold calculations were computed using the method as described by Newman-Stonebraker, et al. *Science* **2021.** (https://github.com/sigmangroup/threshold). The thresholds were qualified by a range of metrics, including: Matthew's Correlation Coefficient (MCC), F_1, accuracy, precision, and recall. MCC captures a more holistic measure of the entire classification performance. Thus, MCC is optimized for the threshold detection.

In this work, threshold detection was conducted over yield of the host-mediated reduction of azaarenes. We define *active* substrates as those with a yield >30% (>3 turnovers of the catalyst), while *inactive* substrates are those with a yield <30% (<3 turnovers of the catalyst). A threshold for activity was also considered at 20%, but very a single example of an *inactive* substrate existed for GaPyr-host **IV** with this threshold (see *Expanded Discussion on Figure 3B*), greatly limiting the integrity of interpretation for the classification model generated as the two classes (*active* and *inactive*) are imbalanced (*11 examples* v. *1 example* with a 20% yield threshold for host **IV**).

VI.8 – Molecular descriptor calculation workflow

- ***Overview of calculation workflow***

The molecular descriptors are computed and collated via an in-house python script titled *wrangler.py*. The script operates by accepting SMILES strings as molecular input, wrangling various calculation files and outputs, and reporting collated *csv* files containing ensemble-level information for each molecule considered.

The workflow operates by (1) generating an initial conformer for each molecule, (2) optimizing the structure to a local minimum with Grimme's *xTB*,[1] (3) searching for accessible conformers with Grimme's *CREST*,[2] (4) extracting descriptors from each conformer collected *within a 1.8 kcal window* (represents >95% of the ensemble population at 298 K), and (5) exporting ensemble-level descriptors for each molecule into *csv* files. Each step is described in detail below. Examples of calculation input and output are provided for reference. Full Python scripts can be made available upon request.

- ***#1 – Generating initial conformations for each molecule***

SMILES strings were processed into 3-D coordinates by using RDKit's *AllChem* module with a preset random seed (*0xf00d*) for reproducibility to generate a reasonable starting conformation.

```
AllChem.EmbedMolecule(this_mol, randomSeed=0xf00d)
```

The atomic coordinates were then rearranged to preserve an organized substructure for later processing. This could be done manually by submitting the desired substructure in a SMILES string to *wrangler.py*, or automatically by (1) solving for the maximum substructure with RDKit's *rdFMCS* module with the settings reported below, and then (2) using RDKit's *GetSubstructMatch*() function to find the atoms of each molecule belonging to the substructure.

```
rdkit_substr = rdFMCS.FindMCS([mol for name, mol in set_mol.items()
                                    if this_title in name],
                atomCompare=rdFMCS.AtomCompare.CompareElements,
                bondCompare=rdFMCS.BondCompare.CompareAny,
                ringMatchesRingOnly=False,
                completeRingsOnly=True).smartsString
```

The position of the atoms belonging to the conserved substructure are reported into an exported *sdf* file (atomic coordinates reported via RDKit's *Chem.MolToMolBlock(<RDKit molecule object>)* function) for each of the input molecules.

An initial *xyz* file was then prepared by (1) reporting the total number of atoms at the top of the file, (2) writing each atom from the reported substructure at the top of the coordinate file, and then (3) filling in the remaining atoms from the molecule below. This file serves as the input to the remaining calculations in the workflow.

```
<number of atoms>
<charge> <molecule name>
<atom symbol>    <atom x coord>    <atom y coord>    <atom z coord>
...
```

- ***#2 – Local optimization from initial conformation***

From the initialized *xyz* file, an *xTB* calculation[3] was conducted (default level of theory at *GFN2-xTB* with implicit solvation via *–alpb water* keyword). The *xTB* job file is specific to syntax at the MGCF computing cluster. For reference, the following generic command line sequence is submitted to initiate the optimization:

```
xtb <Molecule-TITLE>.xyz --opt --chrg 0 --alpb water
```

Manual checks are coded to detect if the local geometry optimization fails. In this work, no RDKit-generated initial conformers led to calculations that would not converge within *xTB*'s default number of cycles (250 iterations). The locally optimized geometry for each molecule is exported to individual directories under the title *xtbopt.xyz*.

- ***#3 – Conformational searching with CREST***

From the locally optimized *xtbopt.xyz* files for each molecule considered, a *CREST* calculation was conducted. Because conformational searching is computationally expensive, empirical optimization the following parameters for conformation searching with *CREST* were used.

```
crest xtbopt.xyz -notop -quick -noreftopo -ewin 1.8 -chrg 0 -alpb water -T 8
```

The conformational search uses *CREST*'s built-in *quick* feature which restricts the number of metadynamics (MTD) cycles to search the accessible conformational landscape. Additionally, we restrict the extracted conformers to an energy window of *1.8 kcal* as it represents approximately 95% of the expected population at 298 K. Like the local geometry optimization, implicit solvation is assigned with – *alpb water*. The unique conformers belonging to the desired energy window are exported to *crest_conformers.xyz* for further processing.

- ***#4 – Descriptor processing and curation***

The conformers from the *CREST* calculation are each reported in the *crest_conformers.xyz* file. The in-house python script extracts each conformer, writes the respective *xyz* file into individual directories, and resubmits an *xTB* calculation to access organized and consistent output. The parameters of this redundant *xTB* calculation are identical to those of the initial, local geometry optimization in *Step #2* (default level of theory at *GFN2-xTB* with implicit solvation via *–alpb water* keyword). This generated *xtbopt.xyz* file is subjected to descriptor processing and curation.

Based on literature precedent [4–6]and analysis of accessible descriptors from RDKit,[7] Grimme's *xTB*,[1] and Paton's *DBSTEP* package,[8] we compute and collate the following descriptors for each conformer calculated. Some descriptors are derived from secondary calculations on collected descriptors. The equations used to compute these descriptors (ex. *electronegativity, electrophilicity, molecular hardness, molecular softness*) are included below.

Source *Descriptors*

From xTB output relevant bond lengths, angles, and dihedrals for atoms belonging to the assigned molecular substructure, molecular energy (*GFN2/xTB)*, molecular energies associated with electronic corrections (G_{solv}, G_{sasa}, G_{elec}, G_{hb}, G_{shift}), molecular weight, HOMO, LUMO, molecular dipole, molecular polarizability, xTB-computed atomic charges

additional processing to compute …

electronegativity $= -(LUMO + HOMO)/2$

hardness $= (LUMO - HOMO)/2$

softness $= 1/hardness$

electrophilicity $= \frac{electronegativity^2}{2*(hardness)}$

From RDKit occupied molecular volume, van der Waals molecular volume, occupied molecular ovality, van der Waals molecular ovality, total van der Waals

molecular surface area, polar van der Waals molecular surface area, apolar van der Waals surface area.

From DBSTEP buried volumes at the following radii: 2, 2.5, 3, 3.5, 4, 5, 6, and 7 Å for each atom in the assigned molecular substructure; sterimol across the principal axis set by the first two atoms in the assigned molecular substructure: L, Bmin, Bmax.

Supplementary Table VI-1. List of molecular descriptors computed and curated by *wrangler.py*.

To access the *xTB output* dataset from the *xtbopt.xyz* file, the following keywords were invoked to identify the location of the corresponding descriptor. Atomistic descriptors were only collected if the atoms belonged to the initial, computed substructure – *see xTB*-computed atomic charges, bond lengths, bond angles, dihedral angles.

Keyword	***Descriptor(s)***
TOTAL ENERGY	**molecular energy**
-> *Gsolv*	**electronic energy correction, G_{solv}**
-> *Gelec*	**electronic energy correction, G_{elec}**
-> *Gsasa*	**electronic energy correction, G_{sasa}**
-> *Ghb*	**electronic energy correction, G_{hb}**
-> *Gshift*	**electronic energy correction, G_{shift}**
(HOMO)	**HOMO**
(LUMO)	**LUMO**
covCN	***xTB*-computed atomic charges**
molecular mass	**molecular weight**
molecular dipole	**molecular dipole**
Mol. a(0)	**molecular polarizability**
Bond Distances	**bond lengths**
Selected bond angles	**bond angles**
Selected dihedral angles	**dihedral angles**

Supplementary Table VI-2. List of keywords used to delimit the *xTB*-derived descriptors.

Once all descriptors have been collated and computed, the dataset is exported to a custom *yaml* file as a curated data placeholder (*yaml* available via the package, *PyYAML, https://github.com/yaml/pyyaml*).

- ***#5 – Ensemble generation and csv export***

Ensembles are generated for each molecule considered in the workflow. In total, *five* ensembles are computed – *ground state, Boltzmann averaged, minimum value, maximum value,*[9] *and minimum volume conformer.* In this work, all reported descriptor datasets were from the *minimum volume conformer* ensemble because of its (1) chemical relevance to host-guest binding principles, and (2) apparent correlations to experimental outcomes (*yield, diastereoselectivity*). In practice, each ensemble is generated in the following manner:

<u>*Ground state*</u> Accessible conformers calculated from *CREST* are ranked by increasing energy in the *crest_conformers.xyz* output. This relative ranking is preserved upon extraction into individual conformer *xyz* files. Therefore, the *ground state* conformer is denoted with the tag *C_0*. Each *yaml* dataset corresponding to a *ground state* conformer can be collected and exported into a *csv* file marked with the same tag *C_0*.

<u>*Boltzmann averaged*</u> For each molecule, the entire set of conformers are collected. The rel. energy from the *ground state* (from the lowest energy conformer) is computed in *Hartrees* and then converted into *J/mol* with the factor $2.626 * 10^6$.

$$rel.energy = 2.626 * 10^6(\boldsymbol{this\ energy} - gd.state\ energy)$$

Then, the *Boltzmann weight* is computed with the formula:

$$Boltzmann\ weight = e^{-rel.\ energy/(RT)}$$

where $R = 8.314$ *J/mol K* and $T = 298$ K.

The *Boltzmann averaged value* for each descriptor is then computed by (1) converting the *Boltzmann weight* to the proportion of the ensemble populated by this conformer …

$$\boldsymbol{this\ proportion} = \frac{\boldsymbol{this\ weight}}{sum\ of\ all\ weights}$$

… then (2) summing over each conformer's descriptor weighted by its calculated *proportion*. The *csv* file is marked with the tag *AV_XTB* to indicate the *Boltzmann averaged* ensemble was derived from xTB-calculated molecular energies.

Minimum volume The entire set of conformers is collected for each molecule and the molecular volume (as defined by RDKit's *van der Waals molecular volume* descriptor, see *#4*) is extracted. The *minimum* molecular volume was recorded and the descriptors of the conformer corresponding to this molecule were collected and exported into a *csv* file marked with the tag *MINVOL-N1*.

Minimum For each molecule, the entire set of conformers are collected. The minimum value of each descriptor across the set of conformers is extracted and exported into a *csv* file marked with the tag *MIN*.

Maximum For each molecule, the entire set of conformers are collected. The maximum value of each descriptor across the set of conformers is extracted and exported into a *csv* file marked with the tag *MAX*.

Caution is encouraged when considering descriptors belonging to the *minimum* or *maximum* ensembles. Unlike *ground state* or *minimum volume* ensembles, there is no guarantee that any *minimum* or *maximum*-valued descriptor belongs to the same conformer. As such, the chemical meaning of a *minimum*-value or *maximum*-value descriptor should only be invoked in a standalone context (ex. univariate threshold analysis, univariate correlations).

VI.9 – Non-covalent interaction calculation workflow

- ***Overview of calculation workflow***

The workflow borrows heavily from the technique reported by Orlandi and coworkers.[10] An extended discussion on the theory behind the non-covalent interaction calculation is included in the main text and the supplementary information of Orlandi's report. Key takeaways include:

(1) The calculation involves a coordinate scan along a highly constrained geometry between two interacting species. The *computed interaction* is reported as the difference in the energy between each component, extracted from standalone calculations, and the minimum energy along the coordinate scan. The specific constraints are described in detail below.
(2) The non-covalent interaction computed is not necessarily a *specific, quantitative* representation of the interactions within the complexes probed, but rather a *qualitative* representation that can be *quantitatively compared* across substrates within the same system analyzed.
(3) Stacked π- π complexes are energetically less-favored than slip-stacked π -π complexes. However, stacked π- π complexes linearly correlate to the preferred slip-stacked π- π complexes. As stacked π- π complexes are easier to compute due to the increased symmetry (limits the number of conformations to consider), stacked π- π

complexes are exclusively considered when computing the non-covalent interaction. Orlandi and coworkers specifically cite work by Wheeler to justify this choice.[11]

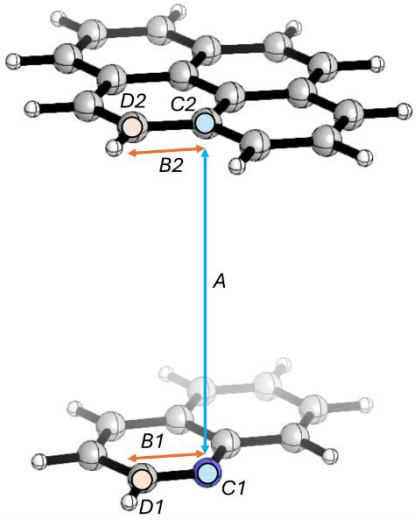
Because the nature of the stacked π- π complexes cannot be characterized in this work, we approximate the nature of the proposed complex using the following probes: **pyridine** is used in place of the reactive, pyridine-borane adducts used as reductants; **pyrene** is used in place of the complex, computationally intractable host wall; **quinolines** are used as represented by the proposed cationic, dihydroquinoline intermediates. Any *ternary complexes* are computed as the sum of separate, *binary* complexes to avoid excess complexity in the coordinate scan calculation.

The workflow operates by (1) generating geometrically constrained, coordinate scan input files specific to each molecule considered, (2) computing the coordinate scan, and (3) extracting the calculated non-covalent interaction energy. Specific details for each step are described below.

- ***#1a – Generating coordinate scan input files – curating the xyz file***

In Orlandi's initial report, the authors use Gaussian16-powered quantum mechanical calculations to evaluate the interaction energy. However, to avoid adverse computational costs in the exploratory nature of these non-covalent interactions, we elected to conduct the workflow using Grimme's *xTB* package instead. In this manner, the input files required for each calculation involve (1) an *xyz* file containing *each* component of the interaction complex under interrogation, (2) a *scan.inp* file containing the specifications for the geometric constraints and the coordinate scan calculation, and (3) a *job* file to submit the *xTB* calculation to the remote computing cluster at the MGCF. The generation of each of these files was automatically completed with an in-house Python script titled *herder*.

The *xyz* files containing each molecule in the complex (defined as *probe* and *substrate*) were assembled through a series of linear transformations. The following atom assignments are made to each *probe* and *substrate* of the complex: **principal atom, secondary atom, normal axis, secondary axis** (**Figure SX).** The **principal atom *(C1, C2)*** refers to the sites that are aligned along the axis scanned during the calculation. The **secondary atom *(D1, D2)*** refers to sites adjacent to the **principal atom** that are aligned to create the symmetry of the stacked π- π complex generated. The **normal axis *(A)*** refers to the axis perpendicular to the π-plane of each *probe* and *substrate* (note – these are aligned in the below figure). The **secondary axis *(B1, B2)*** refers to the rotation of the π-plane of each *probe (B2)* and *substrate (B1)* that undergo alignment to create the symmetry of the stacked π- π complex.



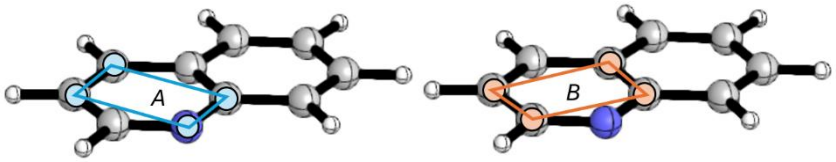
Supplementary Fig. VI-1. General features used to generate the coordinate scan input *xyz* file. The terms defining *A, B1/2, C1/2,* and *D1/2* are described in detail above.

The linear transformations involving the following steps. First, **normal axes** defined by the planes surrounding the **principal atoms** of each *probe* and *substrate* were computed. The rotation matrix required to transform the **normal axis** of the *probe* onto the **normal axis** of the *substrate* was computed using a reported formula. The coordinates of the *probe* were transformed by a dot product with the computed rotation matrix. Second, the orientation of the *probe* and *substrate* planes were twisted to achieve the desired symmetry of the stacked π- π complex. From each **principal atom**, an adjacent secondary atom was selected to form a **secondary axis**. The angle between the **secondary axes** of the *substrate* and *probe* molecules was computed and the *probe* molecule coordinates were transformed with the *apply()* function from the *Rotation* object, instantiated by *from_rotvec()* (*Rotation* is an object accessible from *scipy* in the *spatial.transform* package).[12] Finally, the aligned *probe* was centered onto the **principal atom** on the *substrate* by displacing each coordinate by its relative distance to the **principal atom** of the *probe*. Then, the *probe* atoms were displaced along the *substrate* **principal axis** by 7.5 Å, corresponding to the initial distance of the coordinate scan calculation. The aligned *probe* coordinates are then appended to the *substrate* coordinates and exported into an *xyz* file distinguished by the *substrate*, *probe*, and *non-covalent interaction* computed.

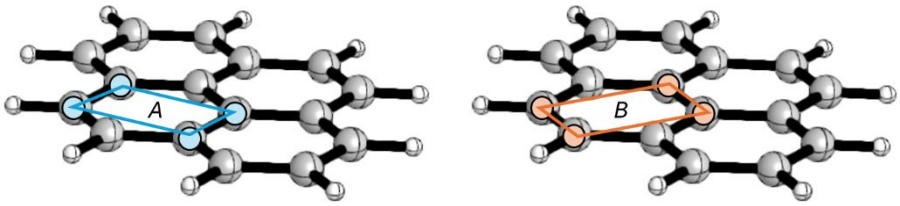
- ***#1b – generating coordinate scan input files – curating the scan.inp specifications***

The *scan.inp* file declares to the *xTB path* calculation how to constrain the input *xyz* file during the coordinate scan. These constraints required specific atomic declarations, so individual *substrate* and *probe* molecules are curated beforehand to maintain common substructures. This allows analogous atomic constraints during the calculation without requiring manual selection of the specific atoms each calculation. The constraints maintain the symmetry of the stacked π- π complex as the π- π complex constricts during the coordinate scan. The constraints within each

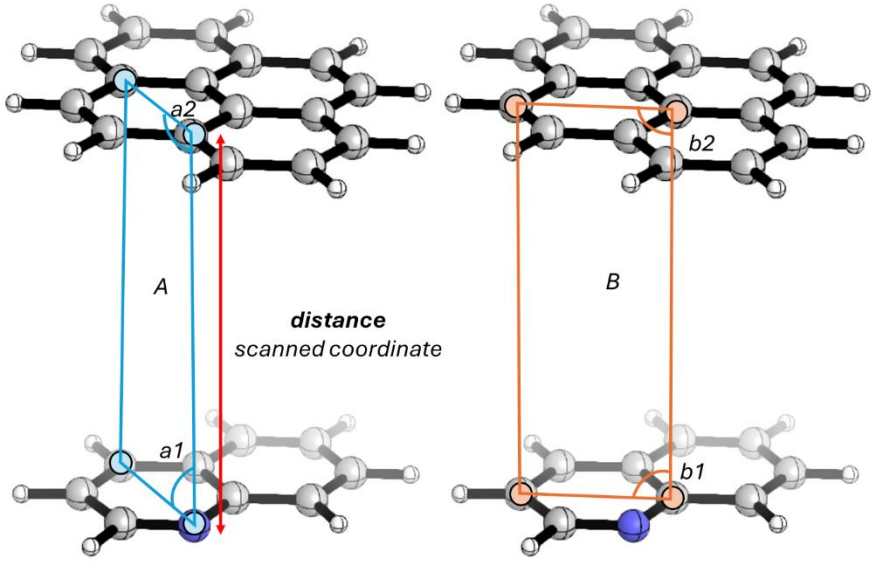
molecule of the complex and across the complex are detailed in the following figures as exemplified by a model quinoline *substrate* and pyrene *probe* complex.



Supplementary Fig. VI-2. General constraints along *substrate*. *A* and *B* refer to dihedral angles (set by the points marked in the *substrate*) constrained to 0°.



Supplementary Fig. VI-3. General constraints along the *probe*. *A* and *B* refer to dihedral angles (set by the points marked in the *probe*) constrained to 0°



Supplementary Fig. VI-4. General constraints across the *substrate-probe* complex. *A* and *B* refer to a dihedral angles (set by the four points marked across the *substrate-probe* complex) constrained to 0^0, while *a1, a2, b1,* and *b2* refer to angles constrained to 90°.

These constraints are assigned to the specific atoms of each *substrate* and *probe* in the *xyz* input file via the in-house Python script. Altogether, these constraints are specified into a *scan.inp* file

which describes to the *xTB* calculation how to proceed with the coordinate scan. As *xTB* computes coordinate scans by a *relaxed fit*, a large *force constant = 10* was found to be most suitable to avoid full, geometry optimization at each step of the coordinate scan. The coordinate scan computed varies the *distance* between the *probe* and *substrate* **principal atoms** between the distances of 7.5 Å to 2.5 Å in 200 steps (see *$scan* declaration).

```
$constrain
 force constant = 10
 distance: <substrate principal atom>, <probe principal atom>, 7.5
 <additional constraints>
$scan
 1: 7.5, 2.5, 200
$end
```

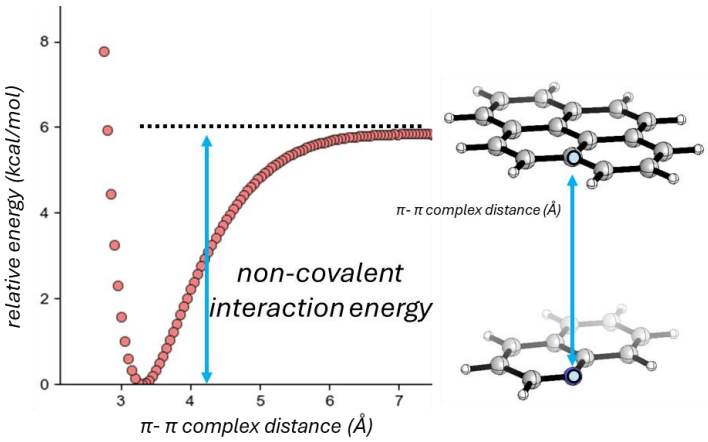
- ***#2 – calculating the coordinate scan***

With the coordinate scan files generated, the *xTB* job can be completed. This calculation is conducted using *xTB*'s path calculation as declared and specified by the *–input scan.inp* keyword. The energy calculations along each step are computed at the default *GFN2/xTB* level of theory with implicit solvation via *–alpb water* keyword.

```
<complex_calculated>.xyz --opt --chrg <complex_charge> --input scan.inp --alpb water > scan_out.txt
```

- ***#3 – extracting the non-covalent interaction energy***

The calculation produces output to *scan_out.txt* which reports the energies computed at each step in the coordinate scan. These energies are extracted by the in-house Python script, exported to a customized *csv* file for storage, and analyzed to identify the *minimum* energy along the coordinate scan. This energy corresponds to the *most stable* π- π complex. The distance between the two π systems and the interaction energy of this complex are recorded. A visualization of the computed interaction is shown below, where the *relative energy to the most stable complex* of each scanned coordinate is plotted against the corresponding *π- π complex distance*.



Supplementary Fig. VI-5. An example of the computed interaction v. π- π complex distance.

The *non-covalent interaction energy* is computed by taking the difference between the *calculated non-covalent interaction energy* and the total separate energies of both the *probe* and *substrate* molecules (calculated at the same level of theory). For a set of molecules, these calculations are exported into a *csv* file as a descriptor specified by the *probe* and the specific interaction computed. These descriptors were collated and employed when screening correlations.

VI.10 – References

(1) 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.

(2) 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.

(3) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended TIGHT-BINDING Quantum Chemistry Methods. *WIREs Comput Mol Sci* **2021**, *11* (2), e1493.

(4) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190.

(5) Van Dijk, L.; Haas, B. C.; Lim, N.-K.; Clagg, K.; Dotson, J. J.; Treacy, S. M.; Piechowicz, K. A.; Roytman, V. A.; Zhang, H.; Toste, F. D.; Miller, S. J.; Gosselin, F.; Sigman, M. S. Data Science-Enabled Palladium-Catalyzed Enantioselective Aryl-Carbonylation of Sulfonimidamides. *J. Am. Chem. Soc.* **2023**, *145* (38), 20959–20967.

(6) Liles, J. P.; Rouget-Virbel, C.; Wahlman, J. L. H.; Rahimoff, R.; Crawford, J. M.; Medlin, A.; O'Connor, V. S.; Li, J.; Roytman, V. A.; Toste, F. D.; Sigman, M. S. Data Science Enables the Development of a New Class of Chiral Phosphoric Acid Catalysts. *Chem* **2023**, *9* (6), 1518–1537.

(7) Greg Landrum; Paolo Tosco; Brian Kelley; Ricardo Rodriguez; David Cosgrove; Riccardo Vianello; sriniker; gedeck; Gareth Jones; NadineSchneider; Eisuke Kawashima; Dan Nealschneider; Andrew Dalke; Matt Swain; Brian Cole; Samo Turk; Aleksandr Savelev; Alain Vaucher; Maciej Wójcikowski; Ichiru Take; Vincent F. Scalfani; Rachel Walker; Kazuya Ujihara; Daniel Probst; guillaume godin; Axel Pahl; Juuso Lehtivarjo; Francois Berenger; jasondbiggs; strets123. Rdkit/Rdkit: 2024_03_1 (Q1 2024) Release, 2024.
(8) Guilian Luchini; Patterson, T.; Paton, R. Patonlab/DBSTEP: Release 1.1.0, 2023.
(9) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308.
(10) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
(11) Wheeler, S. E. Local Nature of Substituent Effects in Stacking Interactions. *J. Am. Chem. Soc.* **2011**, *133* (26), 10262–10274.
(12) 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 Contributors; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; De Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* **2020**, *17* (3), 261–272.

Appendix II – Supplementary Information for Chapter II

Portions of this appendix were adapted from the Supplementary Information of:

Treacy, S.†; Smith, A. L.†; et al. *J. Am. Chem. Soc.* **2024,** *146*, 29792–29800. (DOI)

Heafner, L. Smith, A. L.†; Craescu-Valentina, C.†, et al. *Chem* **2025,** *11*, 102368. (DOI)

Table of Contents:

II.1 – Synthesis of Supramolecular Assemblies **129**
II.2 – Standard Synthetic Procedures **131**
II.3 – Kinetics Experiments **132**
II.4 – Experiment Replication and Expanded Controls **135**
II.5 – Example NMR Spectra for Catalytic Reactivity and Kinetics **138**
II.6 – Extended Discussion on Figure II-2C **140**
II.7 – Extended Discussion on Figure II-3B **143**
II.8 – Extended Discussion on Figure II-4A **146**
II.9 – Extended Discussion on Figure II-4B **149**
II.10 – Extended Discussion on Figure II-4C **152**
II.11 – Extended Discussion on Figure II-5A **155**
II.12 – Extended Discussion on Figure II-5B **158**
II.13 – Extended Discussion on Figure II-5C **160**
II.14 – Extended Discussion on Figure II-5D **166**
II.15 – Substrate and Product Standards Synthesis **172**
II.16 – Characterization Data of Novel Compounds **174**
II.17 – NMR Spectra **177**
II.18 – XYZ Coordinates **196**
II.19 – References **231**

VI.11 – Synthesis of Supramolecular Assemblies

Catalysts **I** and **II** were synthesized according to literature procedures.[1]

Catalyst **I** $Ga_4L^1{}_6$ (L = *N*,*N*'-(naphthalene-1,5-diyl)bis(2,3-di(λ^1-oxidaneyl)benzamide))

^{1}H NMR (500 MHz, MeOD) δ 8.07 (d, *J* = 7.8 Hz, 12H), 7.82 (d, *J* = 8.5 Hz, 12H), 7.28 (d, *J* = 6.5 Hz, 12H), 7.00 (t, *J* = 8.2 Hz, 12H), 6.70 (d, *J* = 5.6 Hz, 12H), 6.42 (t, *J* = 7.8 Hz, 12H).

Catalyst **II** $Al_4L^1{}_6$ (L = *N*,*N*'-(naphthalene-1,5-diyl)bis(2,3-di(λ^1-oxidaneyl)benzamide))

^{1}H NMR (500 MHz, MeOD) δ 8.01 (d, *J* = 7.7 Hz, 12H), 7.81 (d, *J* = 8.6 Hz, 12H), 7.22 (dd, *J* = 8.3, 1.7 Hz, 12H), 7.01 (t, *J* = 8.2 Hz, 12H), 6.62 (dd, *J* = 7.3, 1.7 Hz, 12H), 6.39 (dd, *J* = 8.3, 7.3 Hz, 12H).

Catalyst **III** was synthesized according to a modified literature procedure.[1]

$(NMe_4)_{12}$ $[In_4L^1{}_6]$ (0.5 mmol scale) Using air-free Schlenk technique, 6 equiv. **L^1** (*N*,*N*'-(naphthalene-1,5-diyl)bis(2,3-dihydroxybenzamide)) and 12 equiv. recrystallized NMe_4Br were combined in a 250 mL round bottom flask with 150 mL degassed MeOH to form a suspension. A 1M degassed solution of KOH in MeOH was then added dropwise with stirring until all components became soluble. This solution was allowed to stir for 10 minutes and then taken up in a syringe and added to a separate 500 mL flask containing a stirring degassed solution of $In(acac)_3$ in 100 MeOH. The solution was then warmed to 40 ºC and stirred overnight. The solution was then allowed to cool to room temperature and the host-guest complex was isolated by filtration under ambient atmosphere and washed with Et_2O to furnish a pale-yellow solid. (1.84 g, 94%)

^{1}H NMR (500 MHz, MeOD) δ 8.04 (d, *J* = 7.9 Hz, 12H), 7.86 (d, *J* = 8.6 Hz, 12H), 7.19 (dd, *J* = 8.2, 1.8 Hz, 12H), 7.05 (t, *J* = 8.1 Hz, 12H), 6.60 (dd, *J* = 7.4, 1.8 Hz, 12H), 6.29 (t, *J* = 7.8 Hz, 12H).

Catalyst **IV** was synthesized according to a modified literature procedure.[2]

$K_{12}Ga_4L^2{}_6$ (0.1 mmol scale) 4 equiv. $Ga(acac)_3$ and a stir bar were added a 30 mL septum capped scintillation vial and placed under inert atmosphere. 10 mL of degassed MeOH was added to form a solution which was gently stirred. Then, 6 equiv. **L^2** (*N*,*N*'-(pyrene-1,6-diyl)bis(2,3-dihydroxybenzamide), 2 equiv. of recrystallized NMe_4Br, and a stir bar were added to a septum capped 20 mL scintillation vial and placed in an inert atmosphere. 5 mL of degassed MeOH was added to form a suspension which was gently stirred. A degassed 0.1 M solution of KOH in MeOH was added dropwise until all components became soluble. The entire solution was then added dropwise to the solution of $Ga(acac)_3$ which was stirred. The 30 mL septum capped scintillation vial was placed in a silicone oil bath and heated to 30 ºC for 16 hours. At this time, 10 mL of degassed acetone was added to vial which solubilizes any resulting precipitate. The solution was then stirred an additional 24 hours. Then, using positive pressure of N_2 gas and an outlet needle the solution was concentrated to approximately 3 mL or until a yellow precipitate begins to form in the bulk solution. Then 15 mL degassed acetone was added to the stirring solution forming a fluffy yellow precipitate. The vial was left to stir for 1 hour at room temperature and then placed in a centrifuge. After spinning down the solid, the supernatant was removed with a syringe needle.

The product was placed under vacuum to remove excess solvent. The vial was then transferred to a N_2 purgebox and the deep yellow powder was transferred to another tared vial. (266 mg, 71%)

^{1}H NMR (500 MHz, MeOD) δ 8.75 (d, *J* = 8.5 Hz, 12H), 8.54 (d, *J* = 9.4 Hz, 12H), 7.73 (d, *J* = 9.5 Hz, 12H), 7.49 (d, *J* = 8.6 Hz, 12H), 7.31 (dd, *J* = 8.2, 1.7 Hz, 12H), 6.70 (dd, *J* = 7.3, 1.7 Hz, 12H), 6.39 (dd, *J* = 8.3, 7.3 Hz, 12H).

VI.12 – Standard Synthetic Procedures

- ***Standard Reaction Procedure***

Inside an N_2 purgebox, 2.5 mmol of supramolecular host was added to a 1-dram scintillation vial with stir bar and followed by 0.5 mL of degassed H_2O. 25 mmol quinoline substrate was then added to the solution followed by 75 mmol of pyridine-borane derivative. The vial was then capped and placed in a heating block and stirred at 50 ºC for 16 hours. Immediately after opening, the reaction was quenched by the addition of 100 mL of a 0.1M NEt_4Cl solution followed by 200 mL of 3M KOH in water. The aqueous suspension was then transferred to a pipette column containing ~10 g of silica gel. Pressurized air was then applied until the column reached dryness. The vial was then rinsed with 2 mL dichloromethane three times which was also transferred to the pipette column. The pipette columned was placed inside a scintillation vial and flashed with pressurized air. This was repeated in between 2 ~6 mL additions of dichloromethane. The eluent was concentrated to dryness under reduced pressure. 25 mmol of mesitylene was then added and the mixture was dissolved in $CDCl_3$ for analysis by ^{1}H NMR.

- ***Standard Kinetics Procedure***

Due to the limited solubility of the reactants in aqueous conditions and dramatic changes in reaction rates upon the addition of cosolvents, we elected to quench the reactions at early timepoints in degassed water to obtain kinetics via initial rates.

Inside an N_2 purgebox, 1.25 mmol of supramolecular host was added to a 1-dram scintillation vial with stir bar and followed by 0.5 mL of degassed H_2O. 12.5 mmol 2-Me-quinoline was then added to the solution followed by 37.5 mmol of pyridine-borane. This procedure was repeated 7 times in groups to provide 4 time points run in duplicate. For the 1-minute time point, 100 mL of a 0.1M NEt_4Cl solution in H_2O was added to stop the reaction before the vials were removed from the glovebox. The remaining vials were then capped and placed in a heating block at 3 minutes and stirred at 30 ºC for 10, 20, and 30 minutes. Immediately after opening, the reaction was quenched by the addition of 100 mL of a 0.1M NEt_4Cl solution followed by 200 mL of 3M KOH in water. The aqueous suspension was then transferred to a pipette column containing ~10 g of silica gel. Pressurized air was then applied until the column reached dryness. The vial was then rinsed with 2 mL dichloromethane three times which was also transferred to the pipette column. The pipette columned was placed inside a scintillation vial and flashed with pressurized air. This was repeated in between 2 ~6 mL additions of dichloromethane. The eluent was concentrated to dryness under

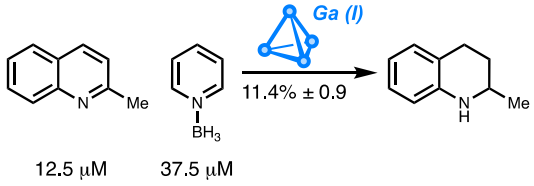
reduced pressure. 12.5 mmol of mesitylene was then added and the mixture was dissolved in $CDCl_3$ for analysis by ^1H NMR.

- ***Standard Encapsulation Procedure***

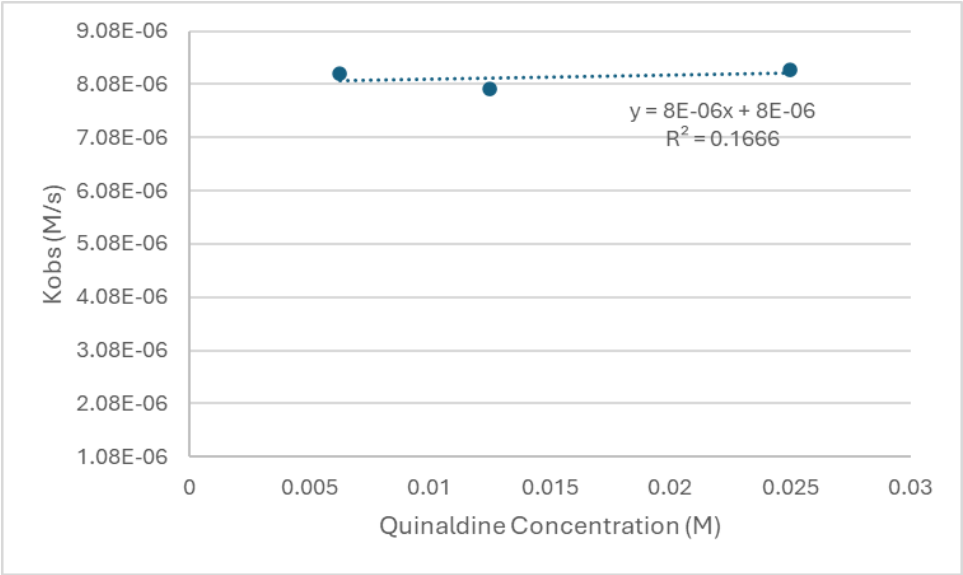
Inside an N_2 purgebox, 2.5 mmol of supramolecular host was added to a 1-dram scintillation vial with stir bar and followed by 0.5 mL of degassed D_2O. 25 mmol of guest substrate was then added to the solution and transferred to an NMR tube. A septum cap was affixed to the NMR tube and the sample was analyzed via NMR spectroscopy.

VI.13 – Kinetics Experiments

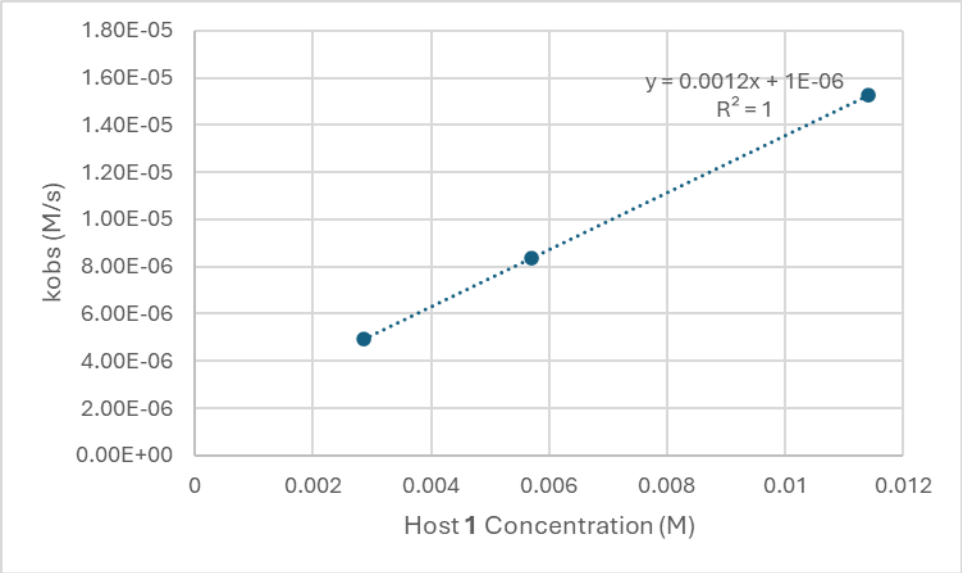
All kinetics experiments were conducted according to the reaction scheme below. The catalyst loading, determined via three independent preparations of catalyst in D_2O with sodium tosylate as an internal standard was determined to be 11.4 ± 0.9%. Determination of the solubility of quinaldine (21 mM) and pyridine-borane (33 mM) in water revealed that the catalytic reaction is under saturated conditions until high conversions to the product. Concentrations were adjusted accordingly to reflect conditions conducive to the relationship between reactant concentration and initial rates.



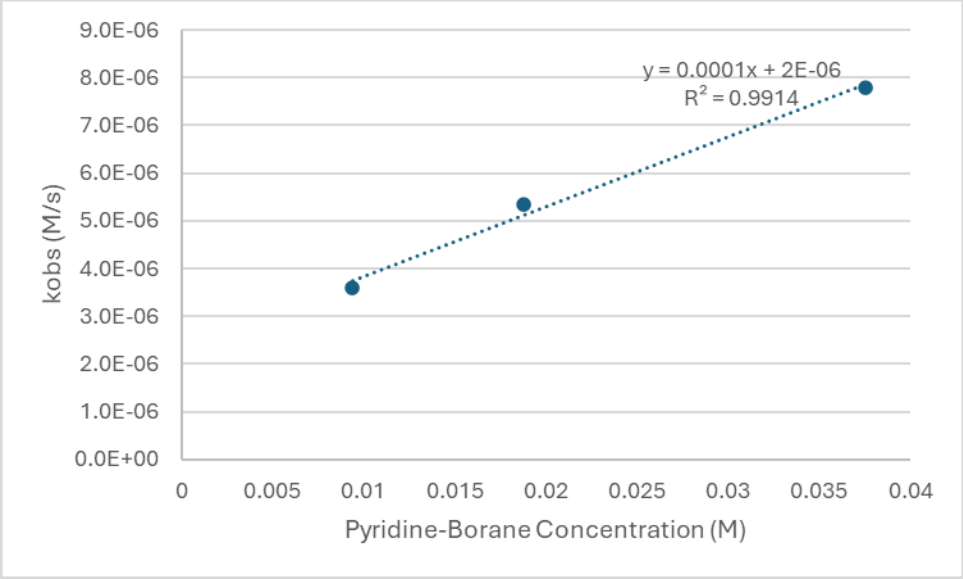
Supplementary Fig. VI-6 | Standard reaction for kinetics profiling.



Supplementary Fig. VI-7 | Rate Dependence on Quinaldine Concentration. Each point represents an average of duplicate experiments.



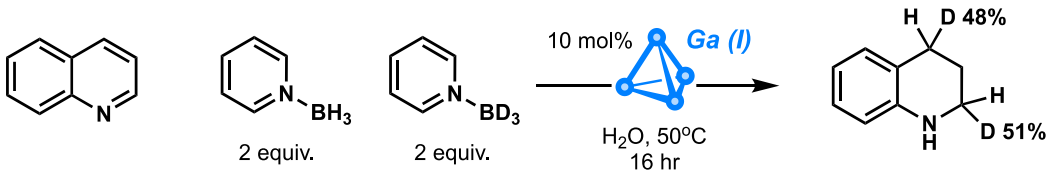
Supplementary Fig. VI-8 | Rate Dependence on Host **I** Concentration. Each point represents an average of duplicate experiments.



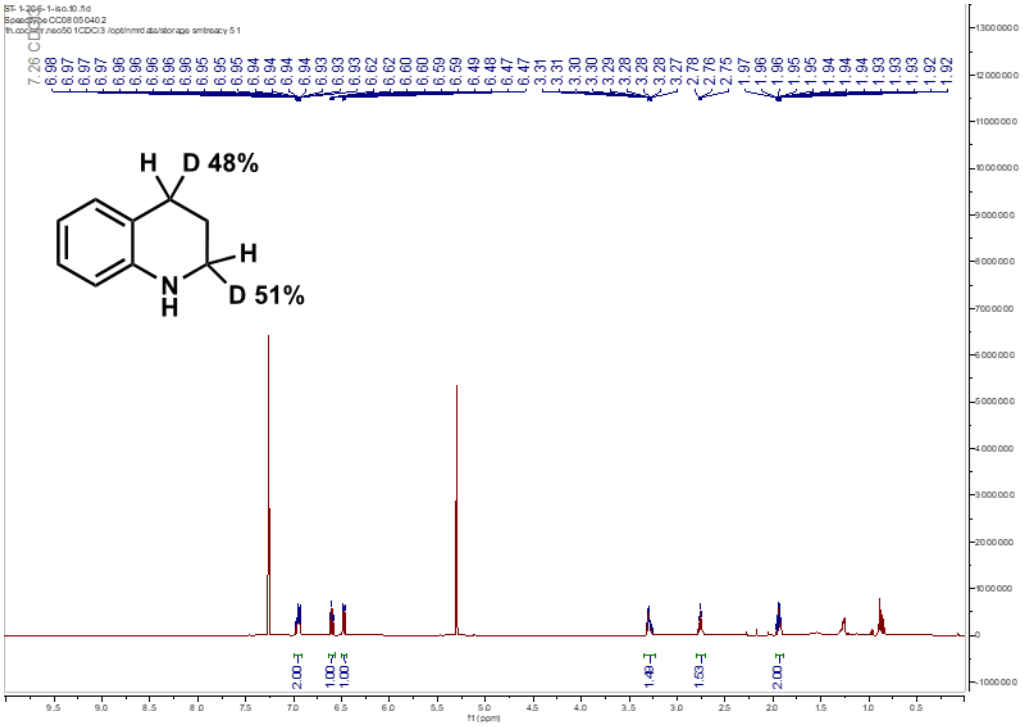
Supplementary Fig. VI-9 | Rate Dependence on Pyridine-Borane Concentration. Each point represents an average of duplicate experiments.

- ***Kinetics Isotope Effect Competition Experiment***

A modification to **Standard Reaction Procedure.** 2 equivalents of $PyBH_3$ and 2 equivalents of $PyBD_3$ were combined with supramolecular host **I** and quinoline and heated in water at 50 °C for 16 hrs. The compound was isolated via silica gel flash column chromatography and analyzed by ^1H NMR. The spectrum reveals attenuated integrations for the C-2 and C-4 positions of the product 1,2,3,4-tetrahydroquinoline according to literature precedent.[3] These integrations show that proton and deuteron incorporation occur at approximately the same rate at both the 2- and 4- positions.



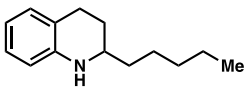
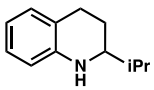
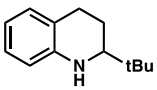
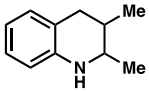
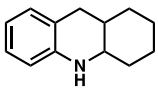
Supplementary Fig. VI-10 | Kinetic isotope effect competition experiment.



Supplementary Fig. VI-11 | KIE competition experiment between $PyBH_3$ and $PyBD_3$ furnished isotopologues of 1,2,3,4-tetrahydroquinoline.

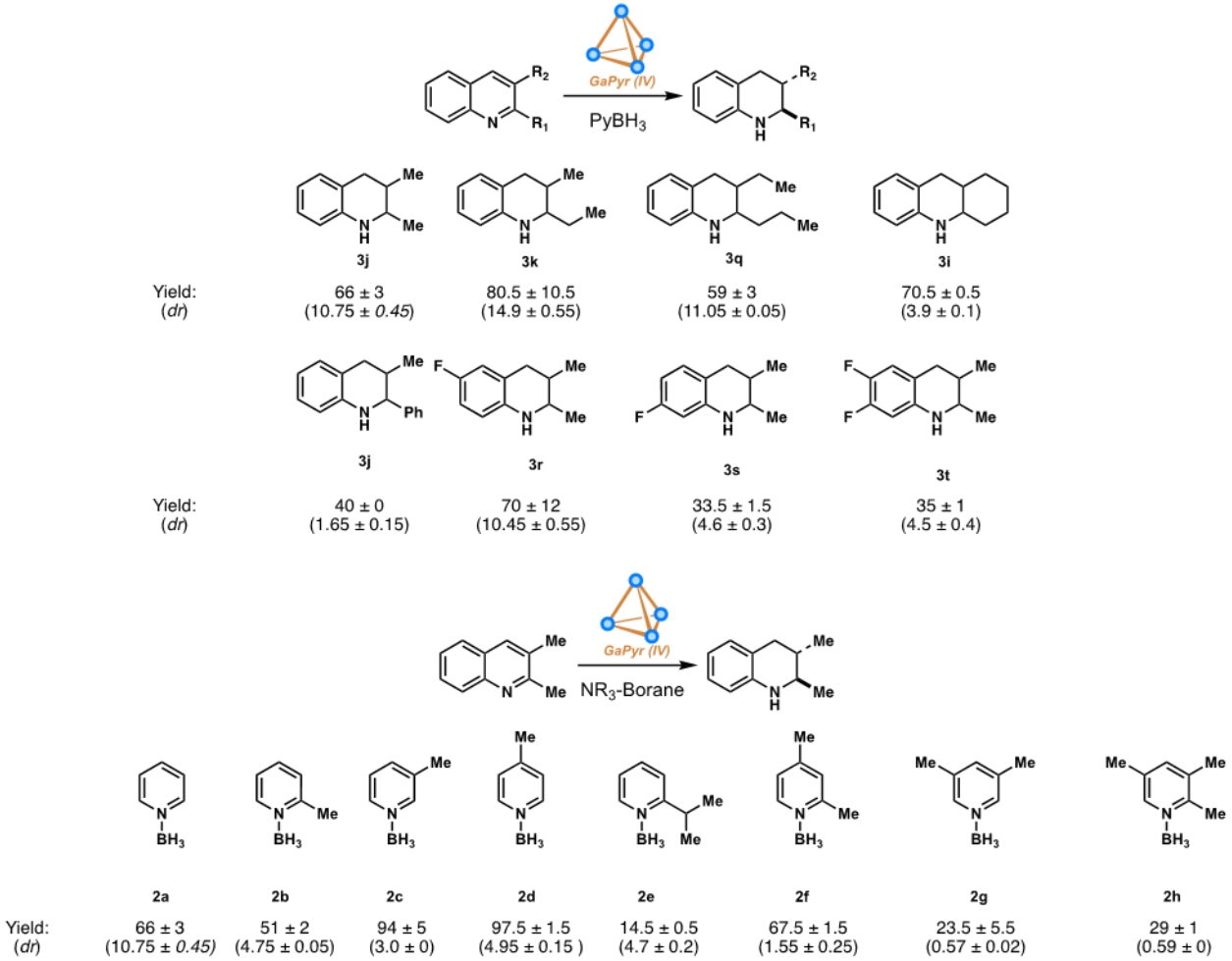
VI.14 – Experiment Replication and Expanded Controls

When setting out to apply data science to this novel reaction, we first assessed the reproducibility of our results under the standard reaction conditions in duplicate for some substrates both high and low yielding for the yield prediction across the catalyst series. The datapoints generated for the yield prediction are the average of the duplicates.

	 3e	 3f	 3g	 3j	 3i
Ga (I)	9%, ± 3	12%, ± 0	0%, 0%, ± 0	54.5% ± 1.5	43% ± 2
Al (II)	4, ± 1	5.5%, ± 0.5	0%, 0%, ± 0	49.5%, ± 5.5	27.5%, ± 2.5
In (III)	7%, ± 1	35.5%, ± 0.5	0%, 0%, ± 0	95%, ± 1.5	44.5% ± 3.5
GaPyr (IV)	17.5%, ± 1.5	89.5% ± 0.5	44% ± 2	66%, ± 3	70.5% ± 0.5

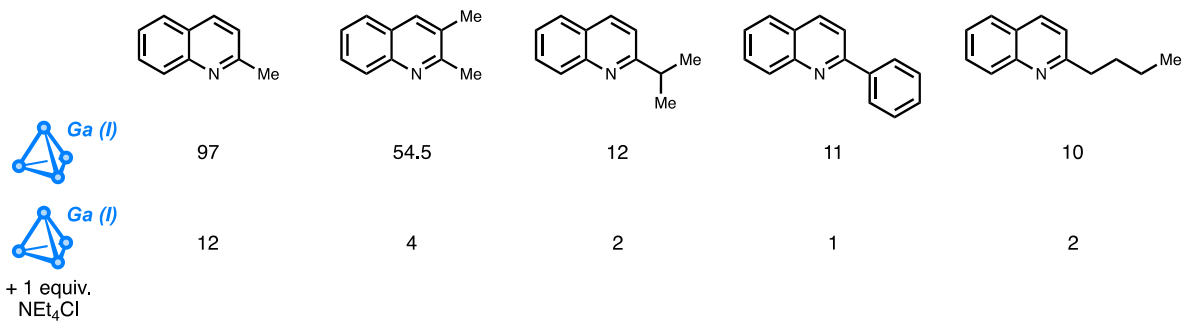
Supplementary Table VI-3 | Duplicate yield data for each catalyst (**I-IV**).

To ensure more robust representations for our model for diastereoselectivity, all reactions leading to a mixture of diastereomers were conducted in duplicate. The yield and diastereoselectivity of each reaction is presented below to show the variance introduced by experimental error.



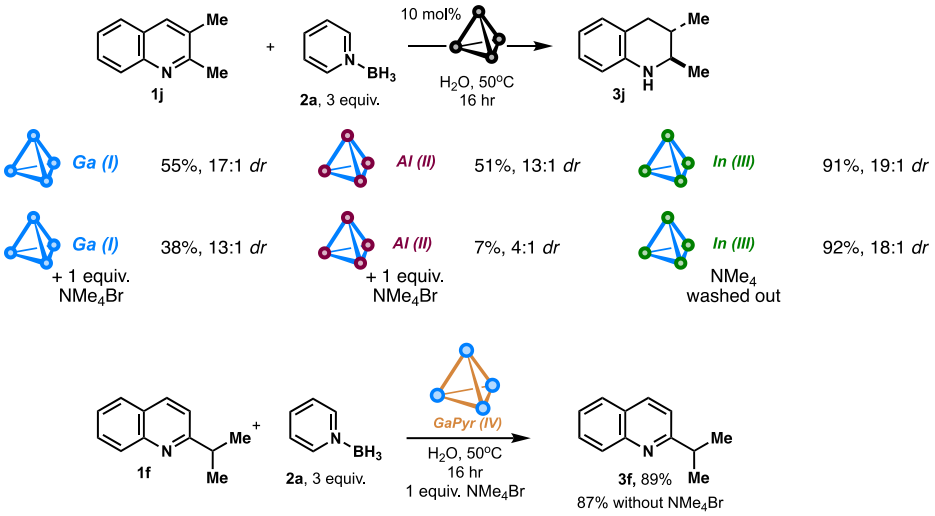
Supplementary Table VI-4 | Reactions run in duplicate with catalyst IV to assess consistency of yield and diastereoselectivity.

Furthermore, due to the background reduction of quinaldine that occurs with the addition of NEt_4Cl with Host **I**, we decided to evaluate other examples in the scope to reinforce the inhibitory effect of a strongly binding guest on the catalytic reaction. Across a series of quinolines, we note significant inhibition of the reduction reaction.

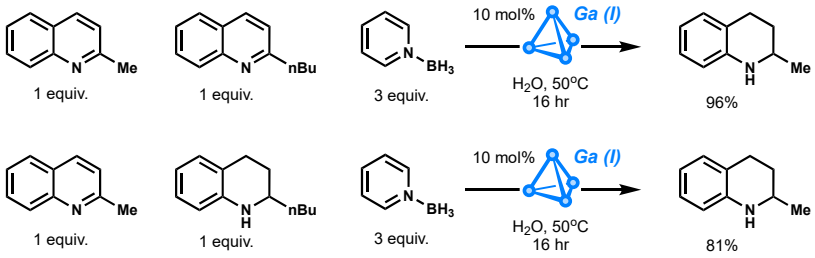


Supplementary Table VI-5 | Effect of NEt_4Cl inhibition across substrate scope with catalyst **I.**

K_{12} $In_4L^1{}_6$ was separately prepared according to a literature procedure.[1] This species was found to be equally reactive within experimental error with regard to both reactivity and diastereoselectivity for the reduction of 2,3-dimethylquinoline. Catalysts **I** and **II** are effectively inhibited by the addition of NMe_4Br, but catalysts **III** and **IV** are not. Previous studies have demonstrated a rate accelerating effect of tetramethylammonium salts with host **I** when product expulsion is rate determining,[4] but little to no effect was seen on catalytic reactivity with the addition of NMe_4Br on catalyst activity.



Supplementary Table VI-6 | Effect of added NMe_4Br on catalyst yield and diastereoselectivity.

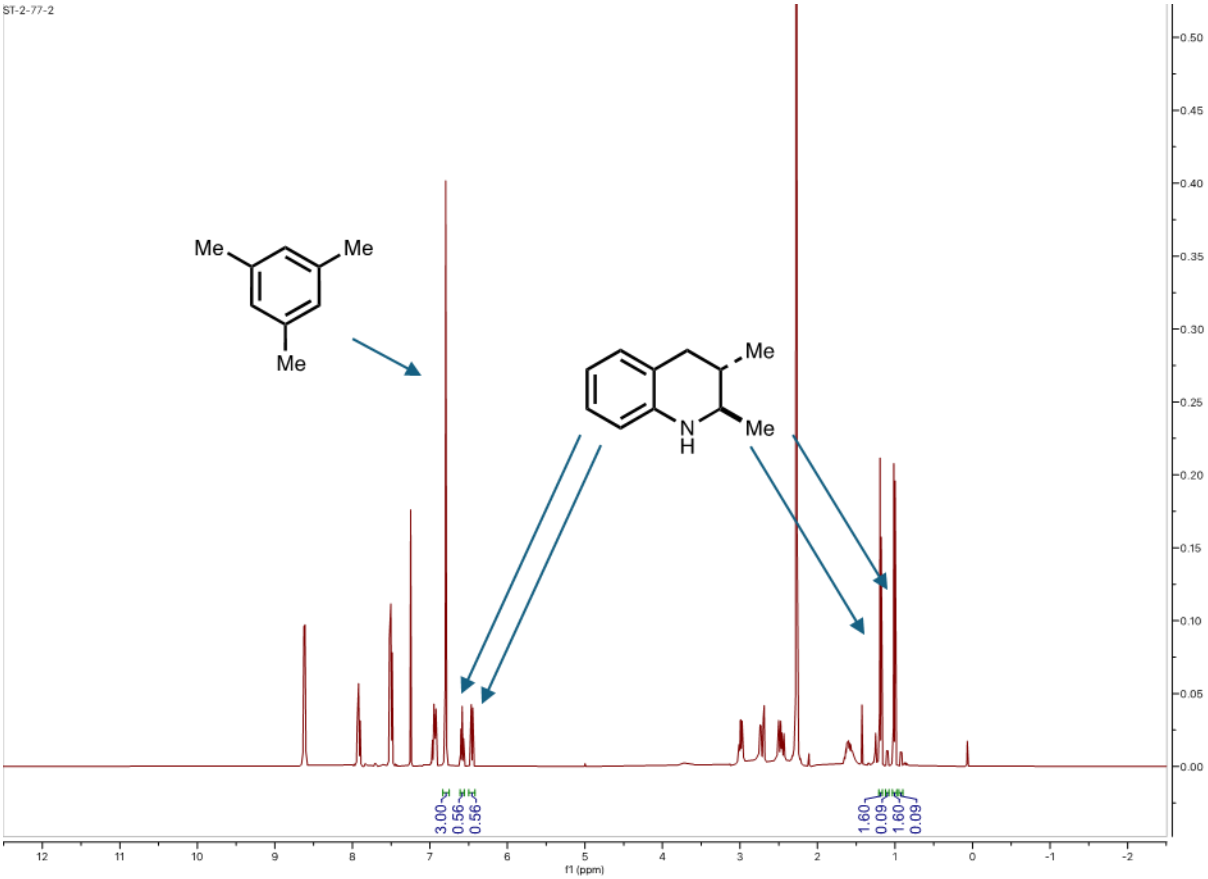


Supplementary Fig. VI-12 | Testing for substrate and product inhibition.

VI.15 – Example NMR Spectra for Catalytic Reactivity and Kinetics

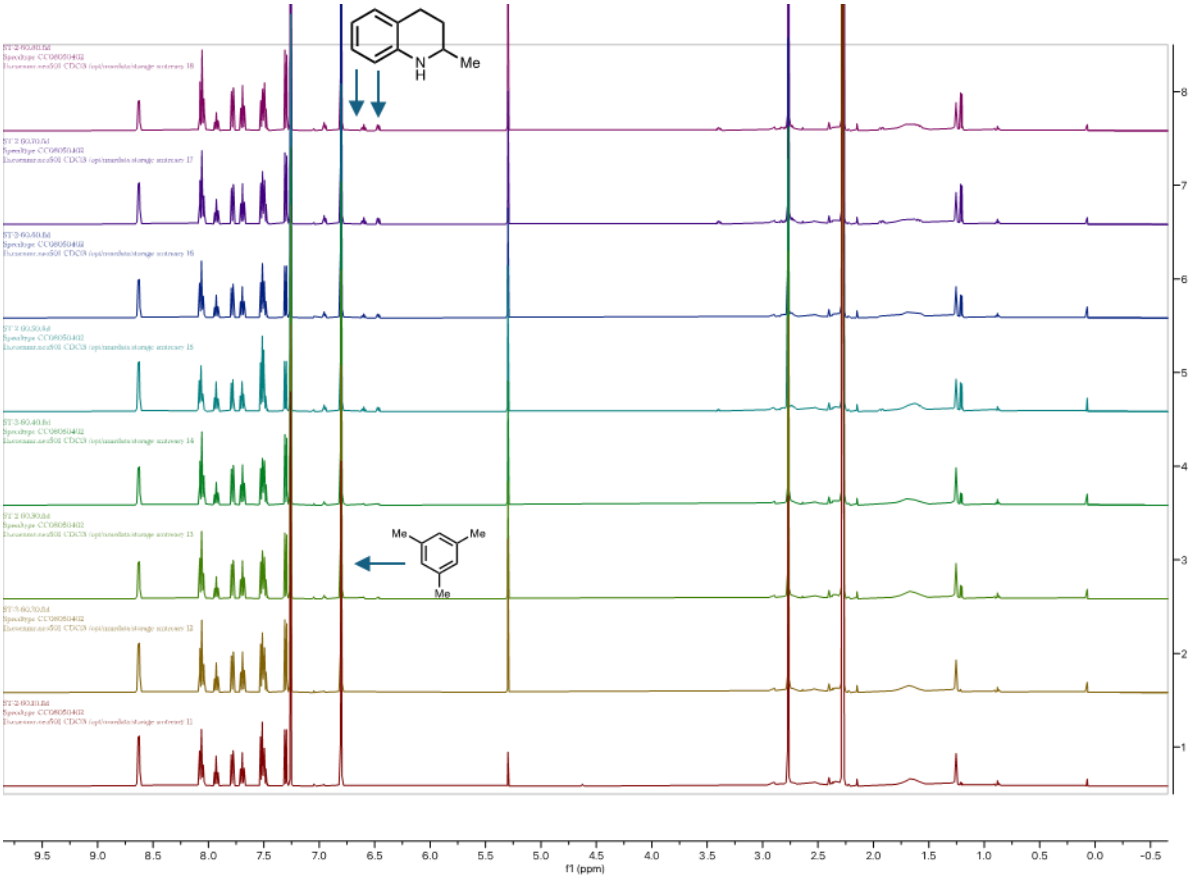
- ***Example Spectra for Yield and Diastereoselectivity Determination***

Spectral assignment for the syn- and anti-diastereomers assigned according to literature precedent.[5]



Supplementary Fig. VI-13 | Example ^{1}H NMR spectrum for determination of yield and diastereoselectivity.

- ***Example Spectra for Yield Determination in Kinetics***

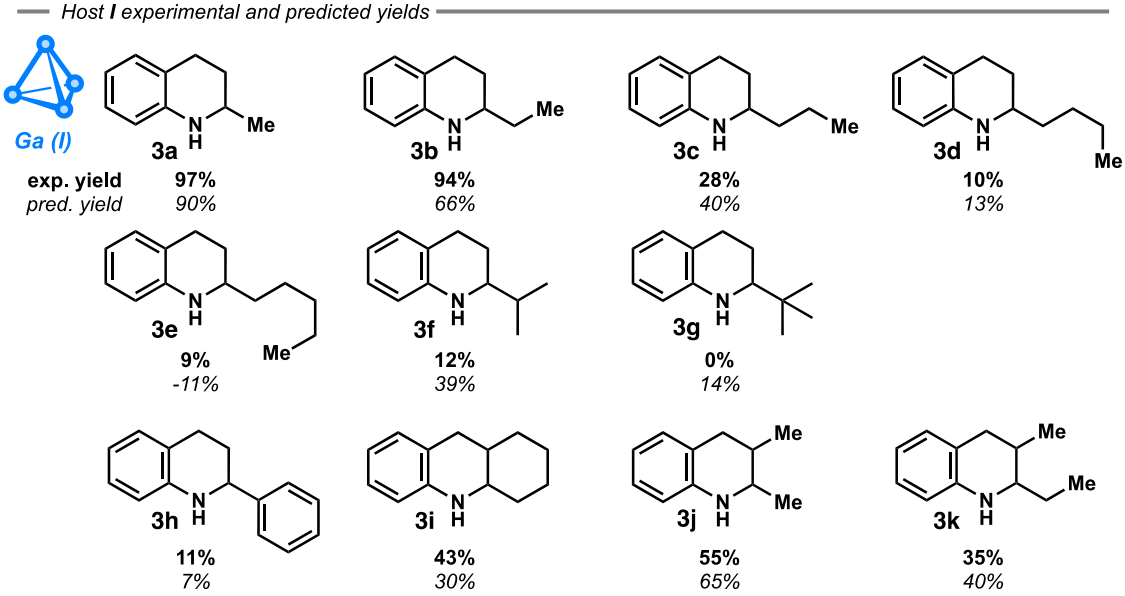


Supplementary Fig. VI-14 | Example ^{1}H NMR spectrum for kinetics analysis. Spectra 1-2, 3-4, 5-6, and 7-8 are duplicates of one another at time points t = 1 min, 10 min, 20 min, and 30 min.

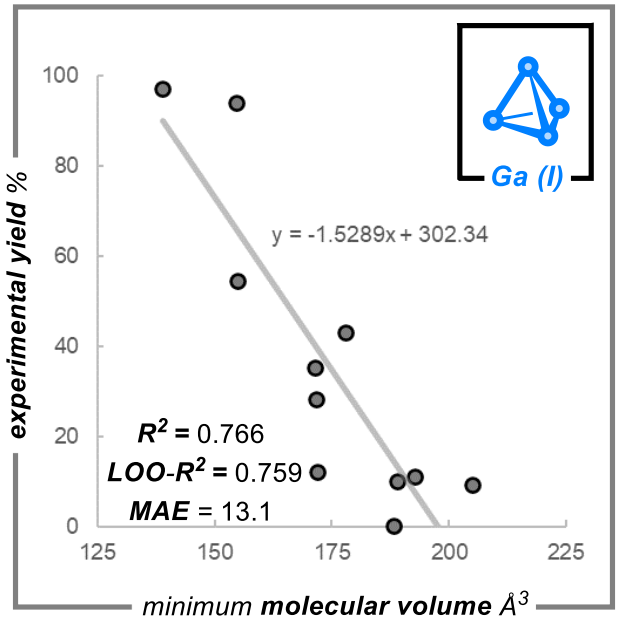
VI.16 – Extended Discussion on Figure II-2C

- ***Full dataset reported for host I-catalyzed azaarene reduction***

The yield prediction model built from *minimum molecular volume* of the systematic selection of quinoline substrates reduced by host **I** produced the following predictions.

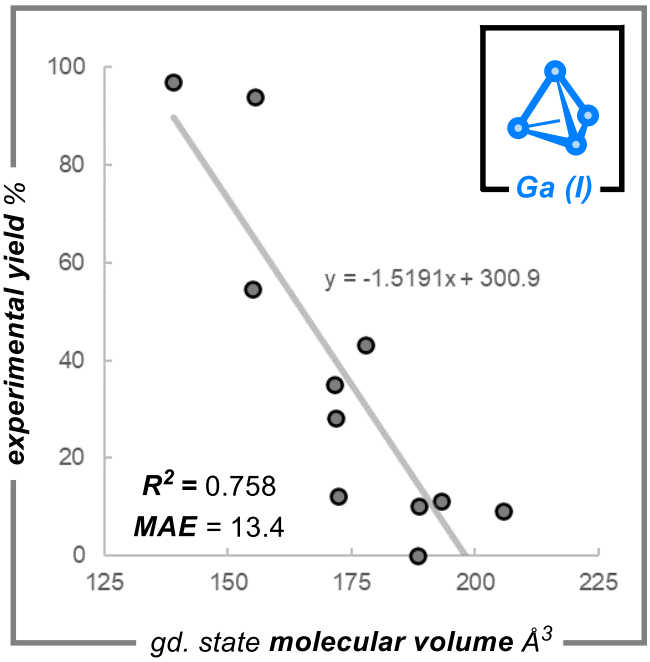


Supplementary Table VI-7 | Host I catalyzed azaarene reduction yields (**bold**) and predicted yields (*italics*) based on *minimum molecular volume* of the quinoline substrate.

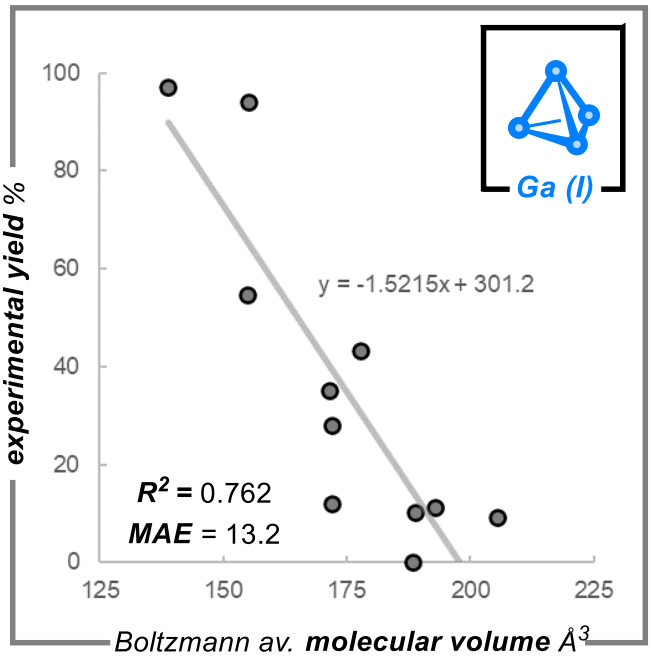


Supplementary Fig. VI-15 | Linear correlation between quinoline *minimum molecular volume* and host I catalyzed azaarene reduction yield.

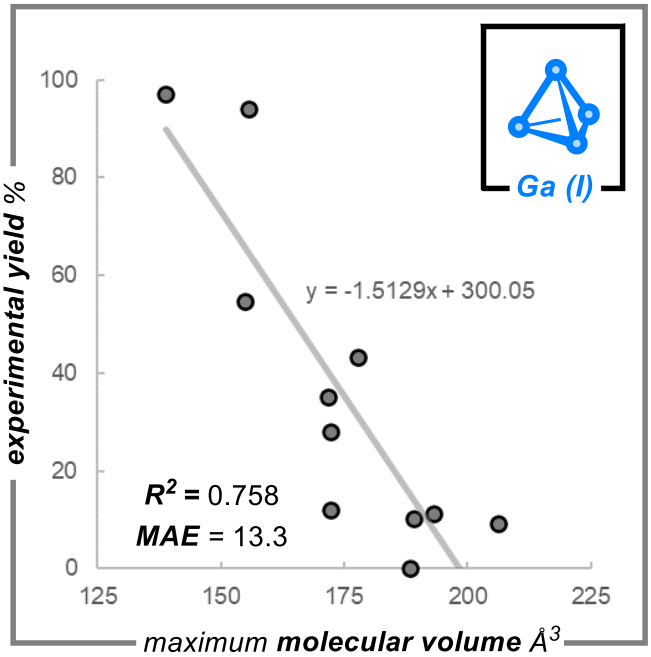
The correlation reported relied on *minimum molecular volume* and invoked known chemical principles, such as *constrictive binding* and *size-exclusion*, to justify unimolecular guest binding behavior. However, experimental analysis strongly supports the secondary binding event of the reductant to form the *reactive, ternary complex* is rate determining. Thus, we invoke the *minimum molecular volume* of the quinoline substrate serves as a meaningful representation of the bound guest's capacity to provide adequate space for the second, binding reductant. While this mechanistic interpretation does not require *minimum molecular volume* of the bound guest, *minimum molecular volume* is statistically the most significant volume descriptor across the ensembles calculated and initially considered (see **Molecular descriptor calculation workflow** for full description of all ensembles). The other *ground state, Boltzmann averaged,* and *maximum molecular volumes* display marginally worse performance than *minimum molecular volume* as shown below. As such, we only invoked correlations built from descriptors belonging to the *minimum molecular volume* ensemble in the remainder of the paper.



Supplementary Fig. VI-16 | Linear correlation between quinoline *ground state molecular volume* and host I catalyzed azaarene reduction yield.



Supplementary Fig. VI-17 | Linear correlation between quinoline *Boltzmann averaged molecular volume* (based on *xTB*-calculated energies) and host **I** catalyzed reduction yield.

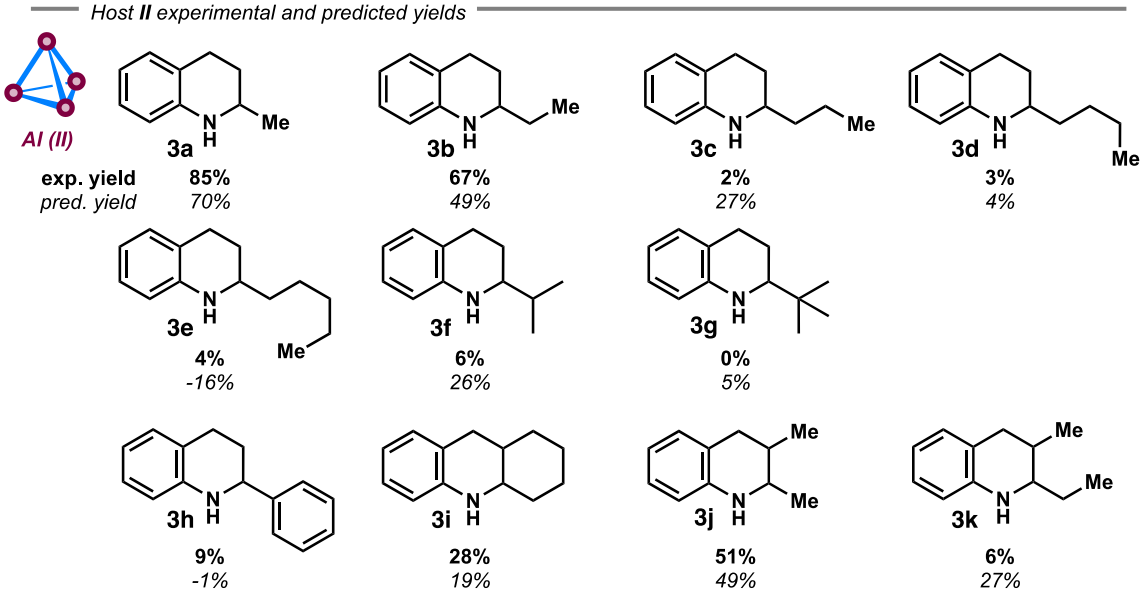


Supplementary Fig. VI-18 | Linear correlation between quinoline *maximum molecular volume* and host **I** catalyzed azaarene reduction yield.

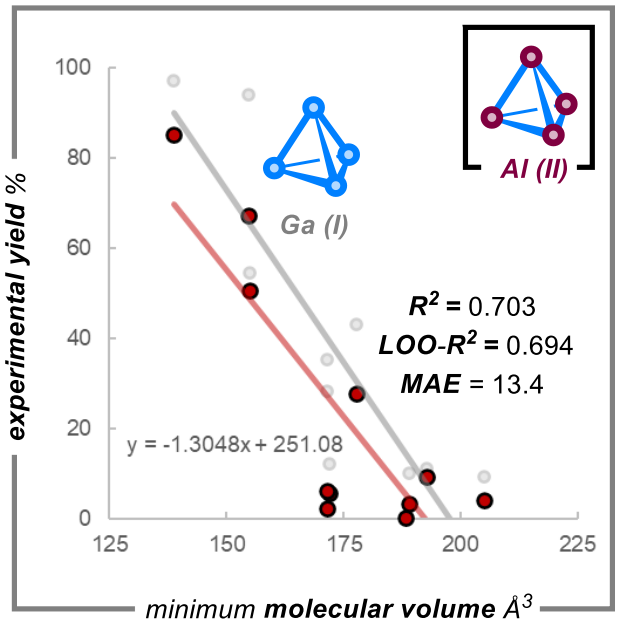
VI.17 – Extended Discussion on Figure II-3B

- ***Full dataset reported for host II-catalyzed azaarene reduction***

The yield prediction model built from *minimum molecular volume* of the systematic selection of quinoline substrates reduced by host **II** produced the following predictions.



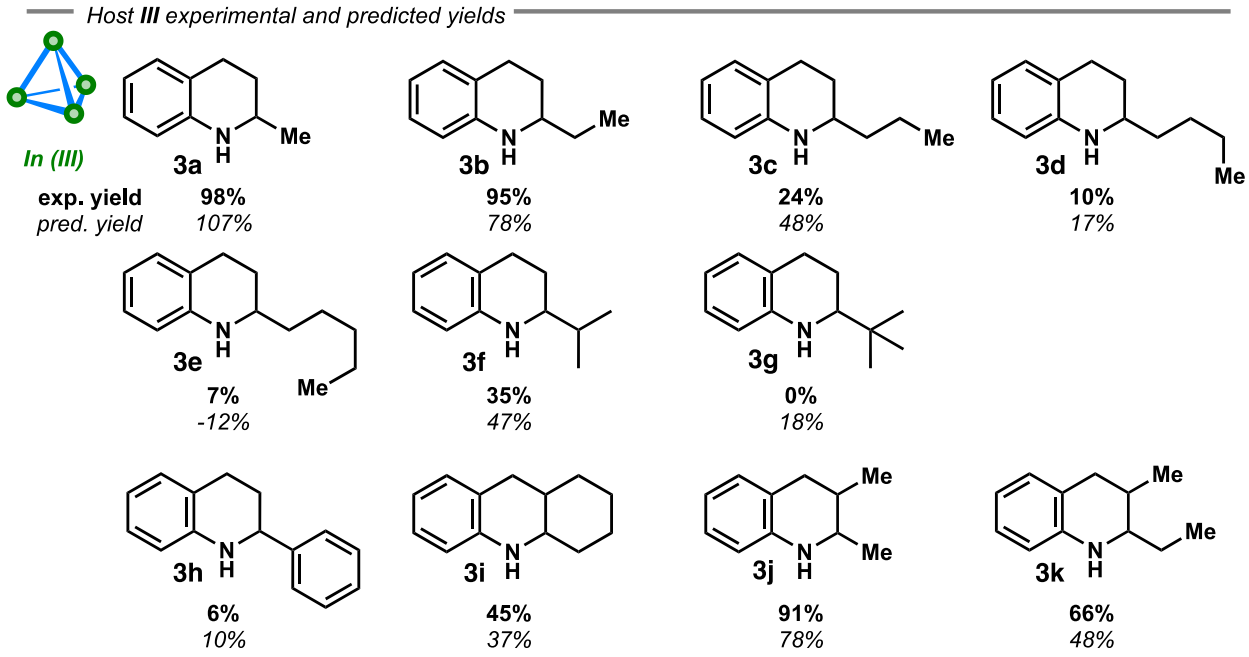
Supplementary Table VI-8 | Host **II** catalyzed azaarene reduction yields (bold) and predicted yields (*italics*) based on *minimum molecular volume* of the quinoline substrate.



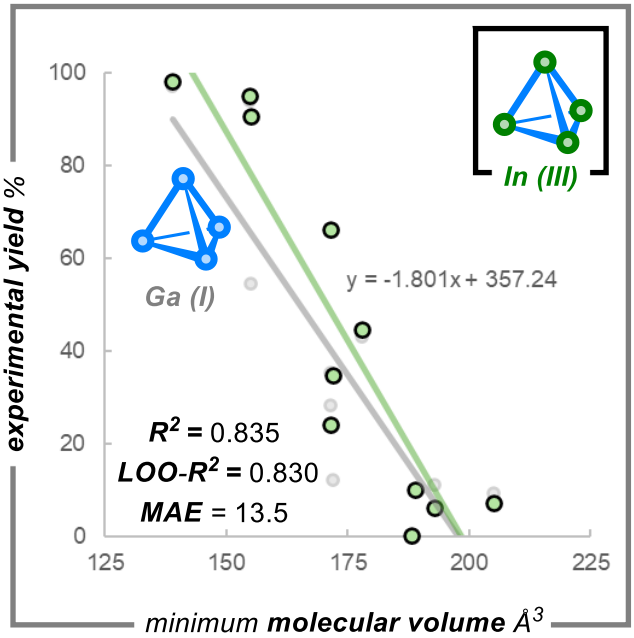
Supplementary Fig. VI-19 | Linear correlation between quinoline *minimum molecular volume* and host **II** catalyzed azaarene reduction yield.

- ***Full dataset reported for host III-catalyzed azaarene reduction***

The yield prediction model built from *minimum molecular volume* of the systematic selection of quinoline substrates reduced by host **III** produced the following predictions.



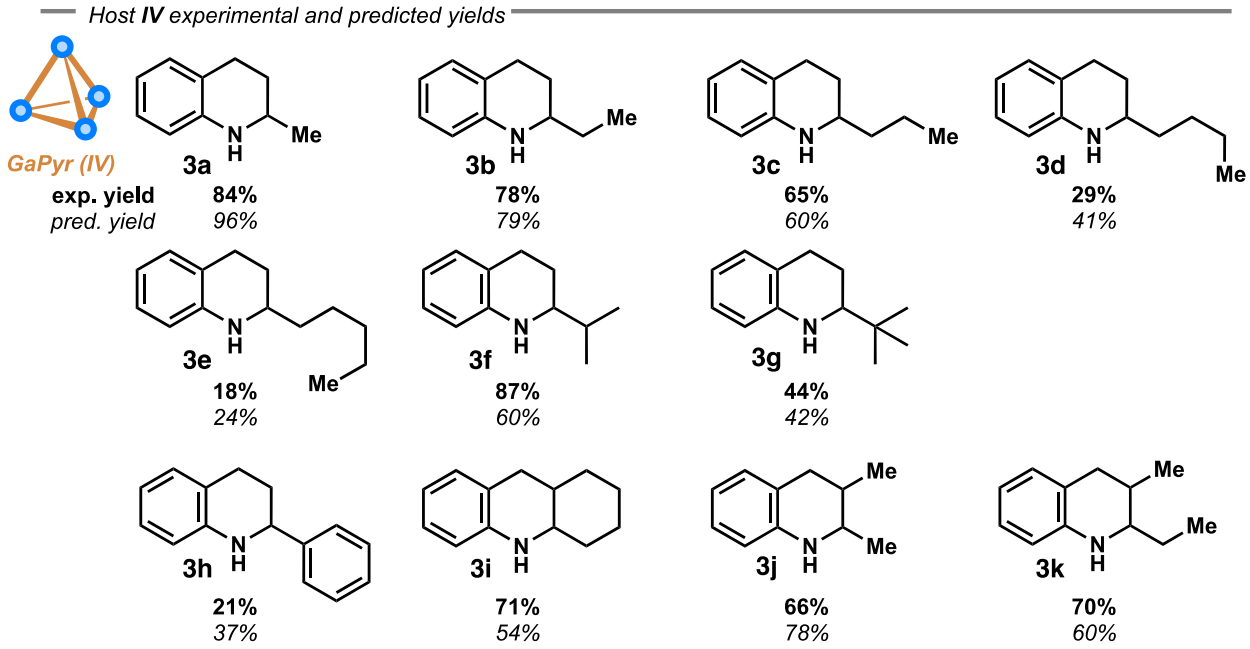
Supplementary Fig. VI-20 | Host **III** catalyzed azaarene reduction yields (bold) and predicted yields (italics) based on minimum molecular volume of the quinoline substrate.



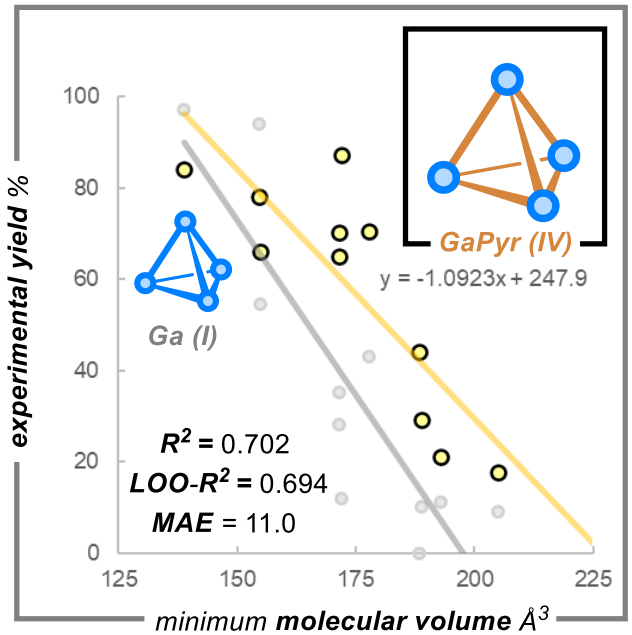
Supplementary Fig. VI-21 | Linear correlation between quinoline *minimum molecular volume* and host **III** catalyzed azaarene reduction yield.

- ***Full dataset reported for host IV-catalyzed azaarene reduction***

The yield prediction model built from *minimum molecular volume* of the systematic selection of quinoline substrates reduced by host **IV** produced the following predictions.



Supplementary Table VI-9 | Host **IV** catalyzed azaarene reduction yields (bold) and predicted yields (*italics*) based on *minimum molecular volume* of the quinoline substrate.

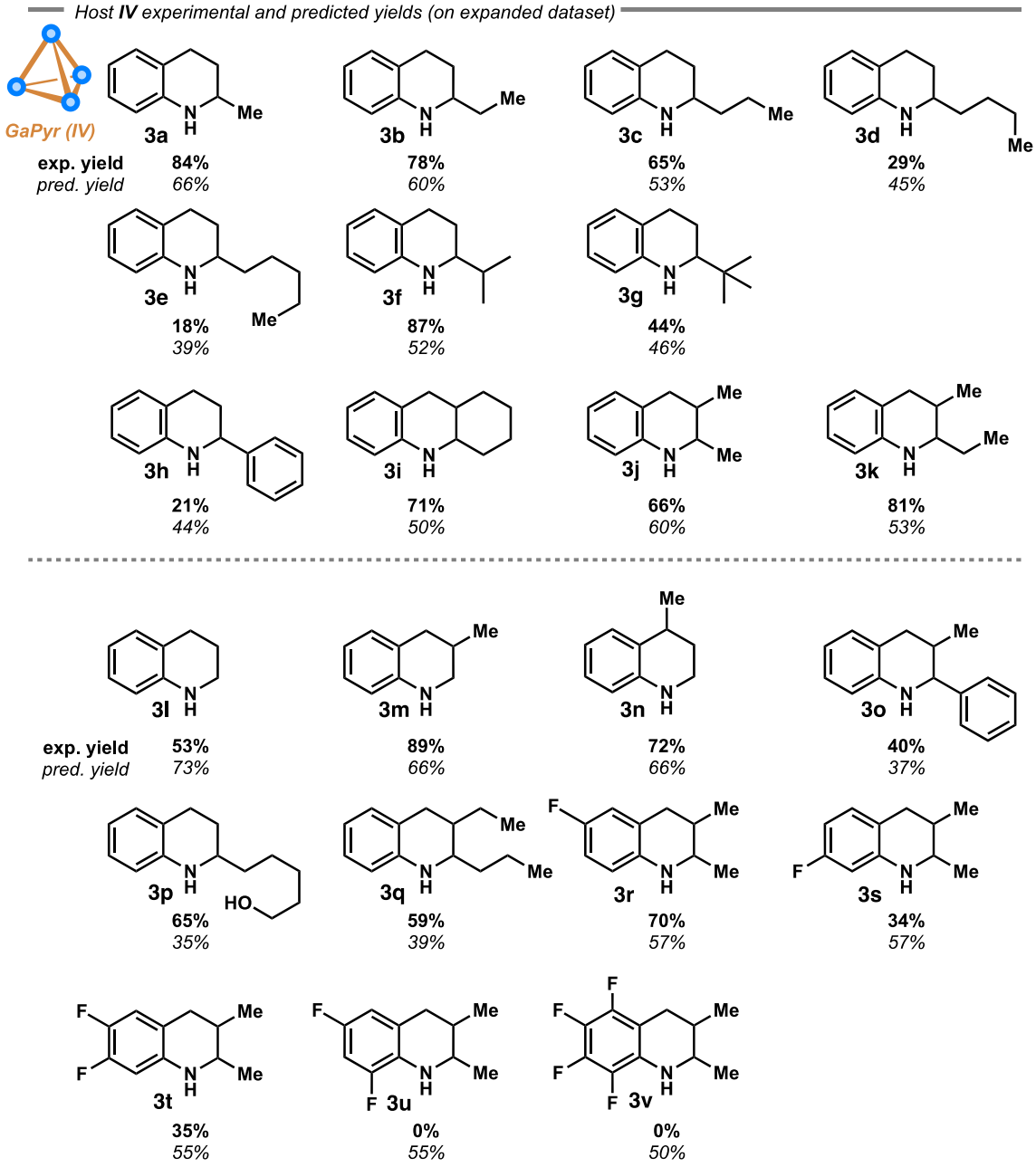


Supplementary Fig. VI-22 | Linear correlation between quinoline *minimum molecular volume* and host **IV** catalyzed azaarene reduction yield.

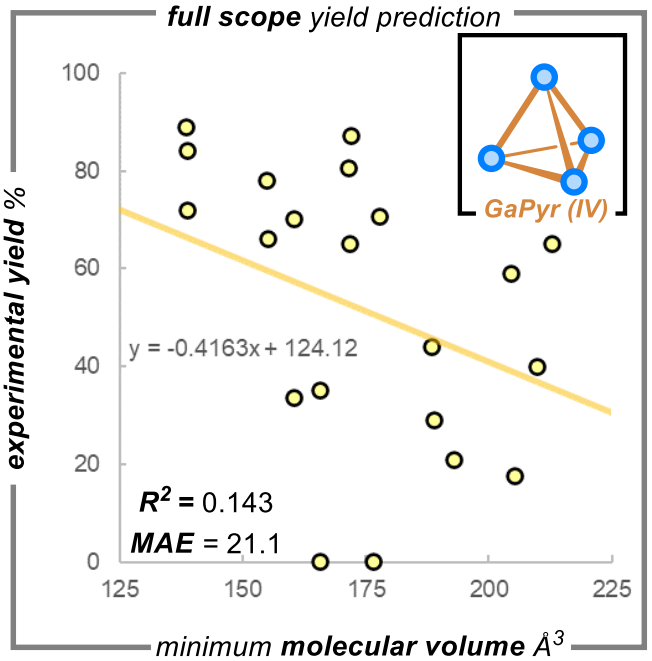
VI.18 – Extended Discussion on Figure II-4A

- ***Full dataset reported for host IV-catalyzed azaarene reduction on expanded substrate scope***

The yield prediction model built from *minimum molecular volume* from the expanded dataset of quinoline substrates produced the following predictions.



Supplementary Table VI-10 | Host **IV** catalyzed azaarene reduction yields (bold) and predicted yields (*italics*) based on *minimum molecular volume* of the quinoline substrate.

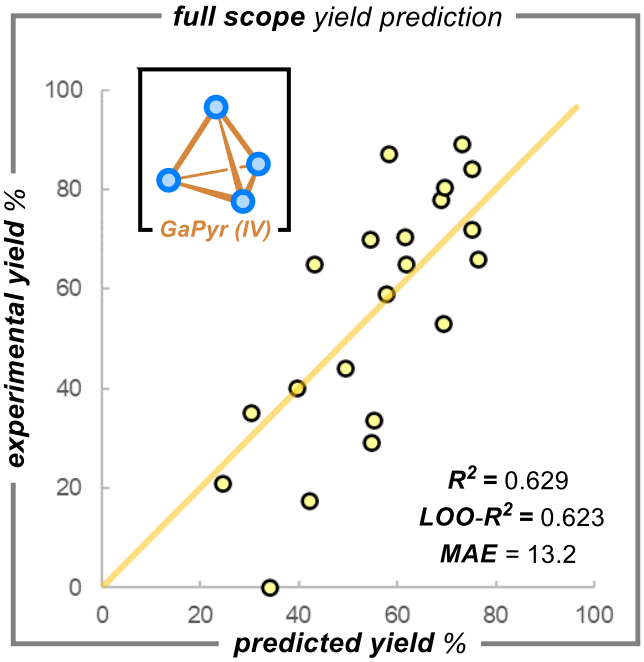


Supplementary Fig. VI-23 | Linear correlation between quinoline *minimum molecular volume* and host **IV** catalyzed azaarene reduction yield.

- ***Bivariate correlations to host IV-catalyzed azaarene reduction yield***

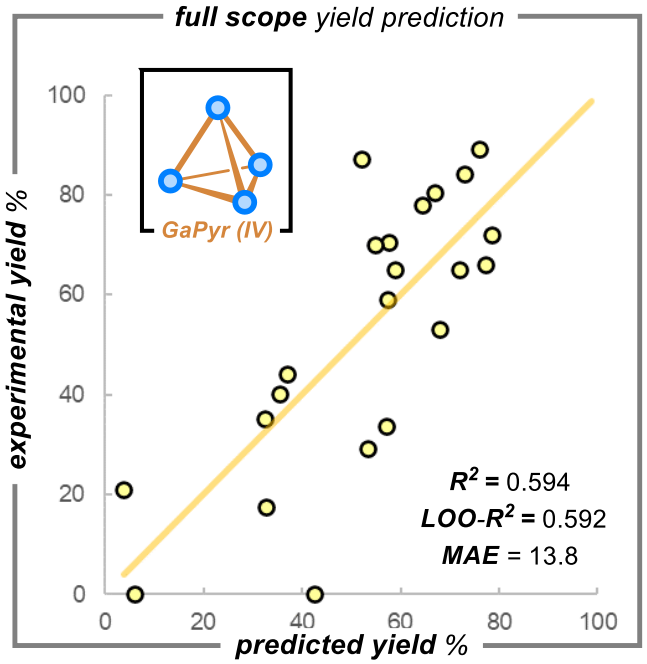
Because *minimum molecular volume* displayed little correlation to reduction yield in the expanded quinoline dataset, we initially considered a series of bivariate correlations to yield where one variable remained *minimum molecular volume*. Two correlations emerged across the full dataset, but the performance was underwhelming (attenuated *LOO*-R^2). These correlations are built from (1) *minimum molecular volume* and *electrophilicity* and (2) *minimum molecular volume* and *molecular hardness* (each electronic descriptor is a function of calculated *HOMO* and *LUMO* levels, see **Molecular Descriptor Calculation Workflow** for details). The overall correlations drawn from these descriptors are included below for reference.

The failure to identify a suitable yield correlation built from only two descriptors where one remained *minimum molecular volume* motivated us to explore how the dataset itself may be better represented as separate classes of quinolines.



***pred.yield** = -0.38x(**mol. vol.**) -7.77x(**electrophil.**)*

Supplementary Fig. VI-24 | Linear correlation between quinoline *minimum molecular volume* and *electrophilicity* to host **IV** catalyzed azaarene reduction yield.



***pred.yield** = -0.27x(**mol. vol.**) +475.2x(**hardness**)*

Supplementary Fig. VI-25 | Linear correlation between quinoline *minimum molecular volume* and *hardness* to host **IV** catalyzed azaarene reduction yield.

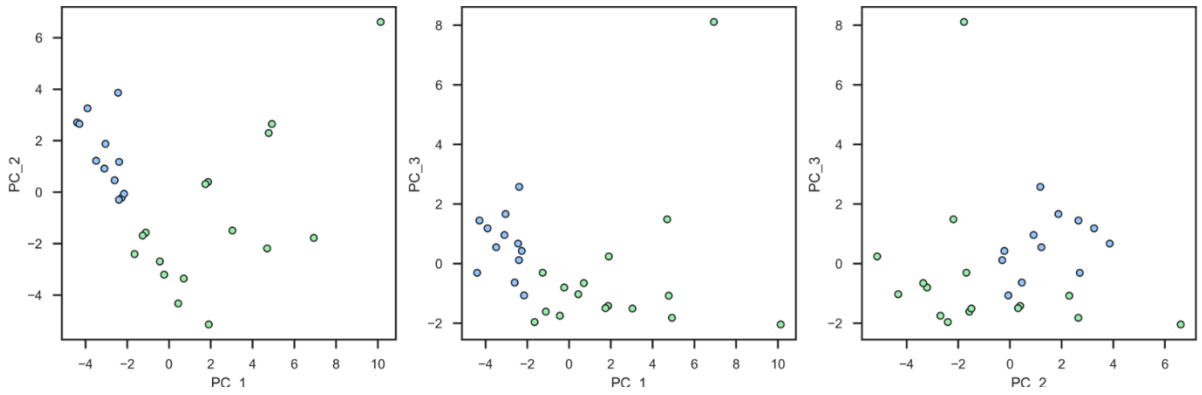
VI.19 – Extended Discussion on Figure II-4B

- ***Overview of clustering algorithm***

The clustering algorithm is conducted through an in-house Python workflow titled *bananas.py*. This workflow first reads the *csv* of the ensemble subjected to analysis, preprocesses the data by (1) normalizing all features with *sklearn*'s *preprocessing.StandardScaler()* object, and (2) dropping features that lack chemical significance like *number of atoms* and *name*, then reduces the dimensionality of the space to a *principal component*-reduced chemical space (corresponding to 90% of the total variance of the original chemical space). This is accomplished with *sklearn*'s *PCA()* object with the *fit_transform*() method.

The resulting *principal component*-reduced space is screened for statistically relevant number of clusters via *sklearn'*s *k*-means clustering object, *cluser.KMeans()*. The *random_state* attribute is set to '*1'* for reproducible clustering. The workflow then ranks the computed clusters against their relative separation from each other using the *silhouette score* metric, as computed by *sklearn*'s *metrics.silhouette_score*(). The numbers of clusters showing the largest separation between each cluster in the *principal component*-reduced chemical space (the largest set of *silhouette scores*) are selected as the most meaningful numbers of clusters.

In this work, the clusters were computed from the *principal component-reduced minimum molecular volume* ensemble from the entire dataset of quinolines considered. These clusters can be visualized across the first *three principal components* as shown below. While the values of these *principal components* are not chemically meaningful, they are comprised of a linear combination of chemically meaningful descriptors. More in-depth analysis of individual components can help to assign some meaning. However, we discuss a strategy later to identify a more meaningful representation of chemical space separated by the clusters.

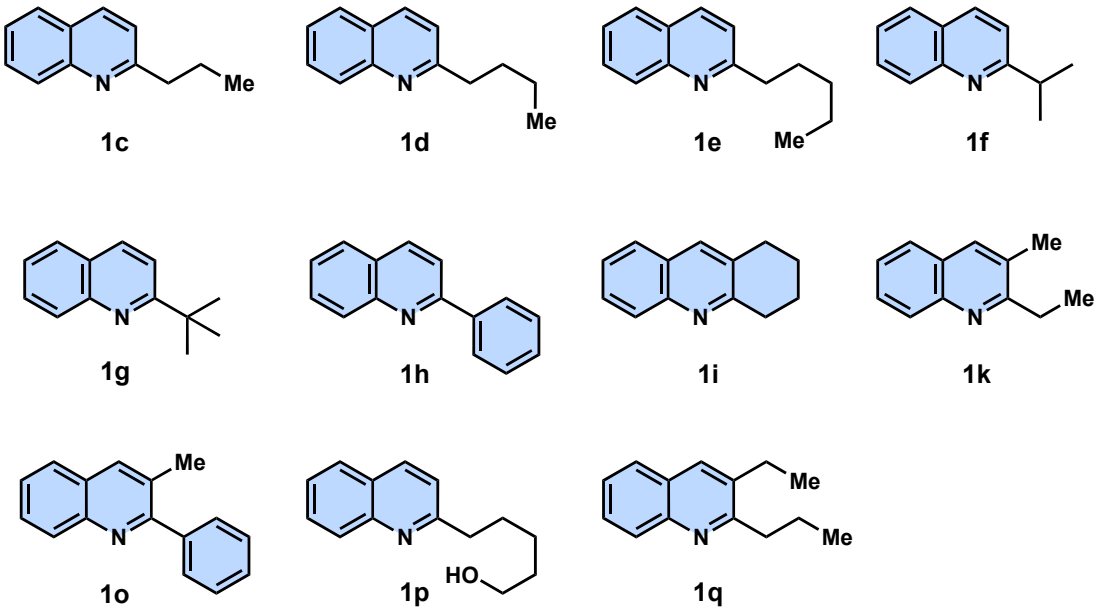


Supplementary Fig. VI-26 | Visualization of each cluster (#1 – blue, #2 – green) over the first *three principal components* (PC) generated from the *minimum molecular volume* ensemble calculated.

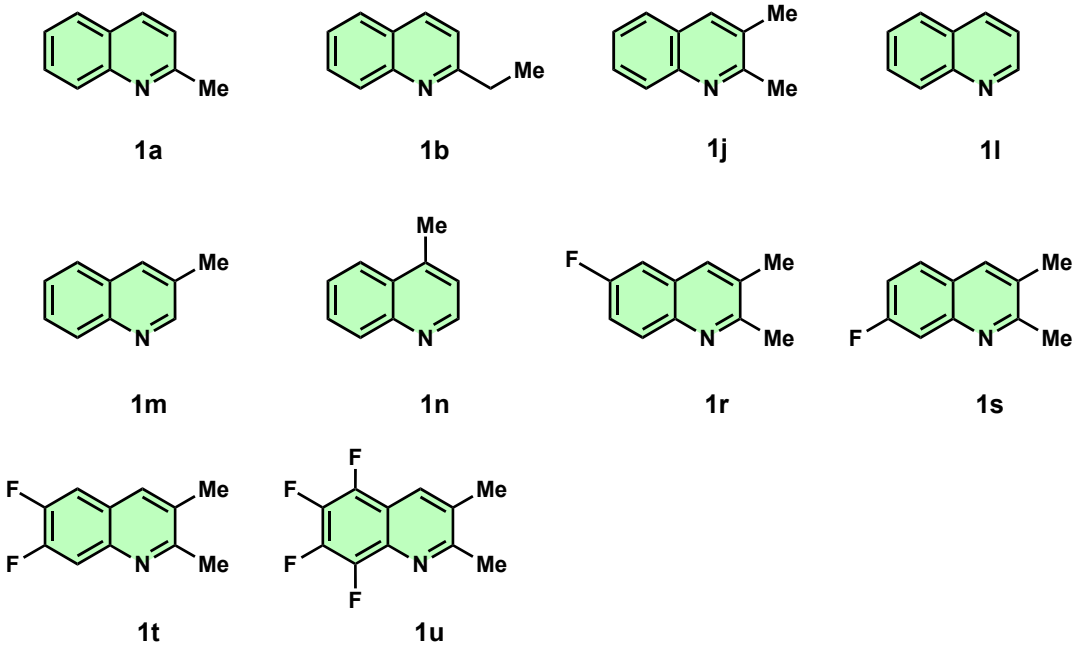
- ***Full cluster assignments over the expanded quinoline dataset***

The *k*-means clustering generated *two* clusters with the following assignments over the dataset.

Cluster #1 *assignments*



Cluster #2 *assignments*



Supplementary Fig. VI-27 | Cluster assignments made by the *k*-means clustering algorithm over the full *minimum molecular volume* ensemble chemical space.

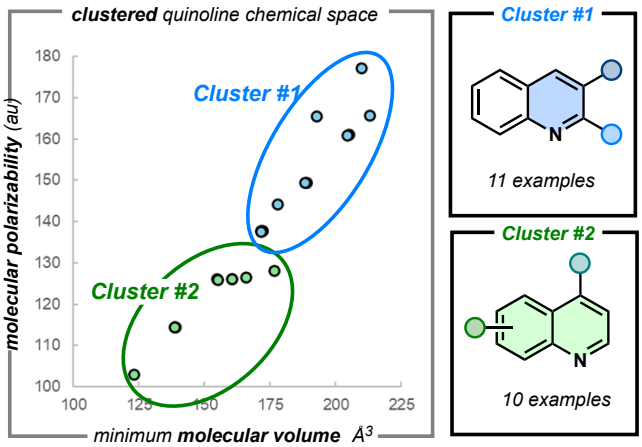
- ***Analysis of chemically meaningful features separating each cluster***

In this work, the most meaningful number of clusters considered was *two classes* of substrates. This preserved the integrity of subsequent linear correlations while providing a statistically meaningful opportunity to distinguish two sets of substrates. To assist in the analysis, we considered what features are significantly separated by the *clustering* calculation. Here, we report the most *separated* variables as determined by the *silhouette score*, a metric relating how separated two clusters are from each other (higher score imply larger separation).

```
apolar_vdw_sasa with a silhouette score of ... 0.672
total_vdw_sasa with a silhouette score of ... 0.633
xtb_polariz with a silhouette score of ... 0.633
mol_vdw_ovality with a silhouette score of ... 0.595
mol_occ_ovality with a silhouette score of ... 0.594
mol_vdw_volume with a silhouette score of ... 0.52
mol_occ_volume with a silhouette score of ... 0.496
xtb_mol_dipole with a silhouette score of ... 0.164
```

Supplementary Table VI-11 | Molecular descriptors most separated by the clusters calculated.

The *two clusters* computed in this work appear to be most separated with respect to *volume* and *surface area* descriptors (ex. *apolar van der Waals surface area* – "apolar_vdw_sasa" and *mol_vdw_volume* – "van der Waals molecular volume"). However, some electronic descriptors also emerge as a significant features as well (ex. *molecular dipole* – "xtb_mol_dipole"). As many of the sterics-associated descriptors are collinear, representing the two clusters with respect to both their steric and electronic electronic desciptors could provide a more robust chemical representation of the reactive substrates under analysis. In our report, we use a representation built with the more physically tangible *molecular dipole*. We show the more statistically separated *molecular polarizability* analogous chemical space below for reference.



Supplementary Fig. VI-28 | Chemically meaningful, clustered chemical space built from *minimum molecular volume* and *molecular polarizability*.

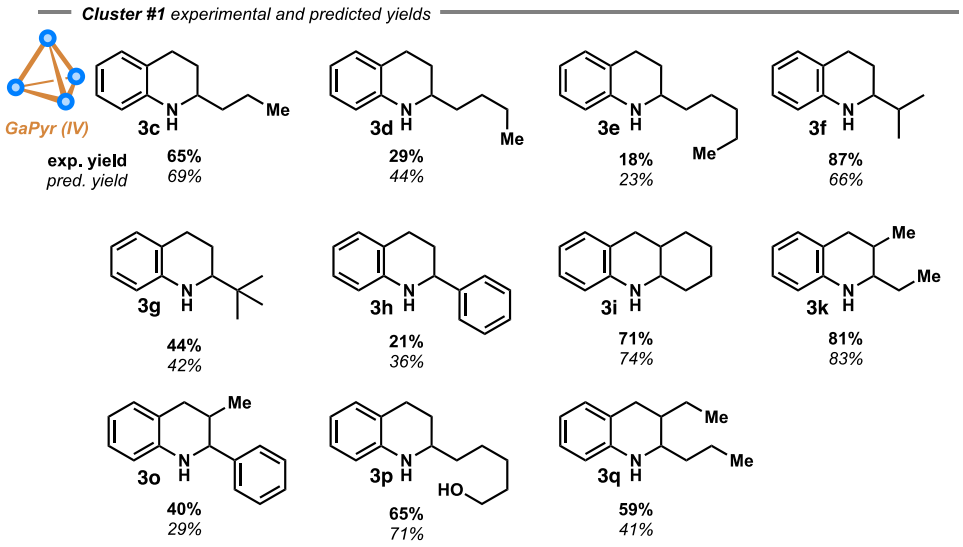
VI.20 – Extended Discussion on Figure II-4C

Following clustering, *minimum molecular volume* continued to be unsuitable for the yield prediction task (*Cluster #1*, R^2 = 0.290 and *Cluster #2,* R^2=0.44). As reported, we anticipated the underlying reason of *minimum molecular volume*'s poor correlation relates to its inability to convey more information regarding *molecular shape*. The cavity and aperture of the host are not dimensionless. They also have a complimentary shape to their volume and area. As such, we expect utilizing features that can suitably convey the bound guest's *molecular shape* will enhance the performance of the yield prediction models (which indicate the capacity to access the *reactive, ternary complex*).

We screened bivariate correlations where one variable is always *minimum molecular volume*. The best correlations emerging across both clusters were then interrogated to identify chemically meaningful interpretations. In combination with *minimum molecular volume*, a few electronic features proved suitable for the bivariate correlation to reduction yield (notably *electronegativity, electrophilicity, hardness, softness,* and *molecular dipole*). Specifically, *molecular dipole* uniquely conveys the overall organization of individual dipoles across the molecule. Because *molecular dipole* is impacted by the orientation of bonds in space, we believe it serves as the most meaningful calculated feature to invoke a *molecular shape* interpretation with the model.

- ***Cluster #1 yield prediction task and proposed interpretation***

The yield prediction model built from *minimum molecular volume* and *molecular dipole* from quinoline substrates in Cluster #1 produced the following predictions.

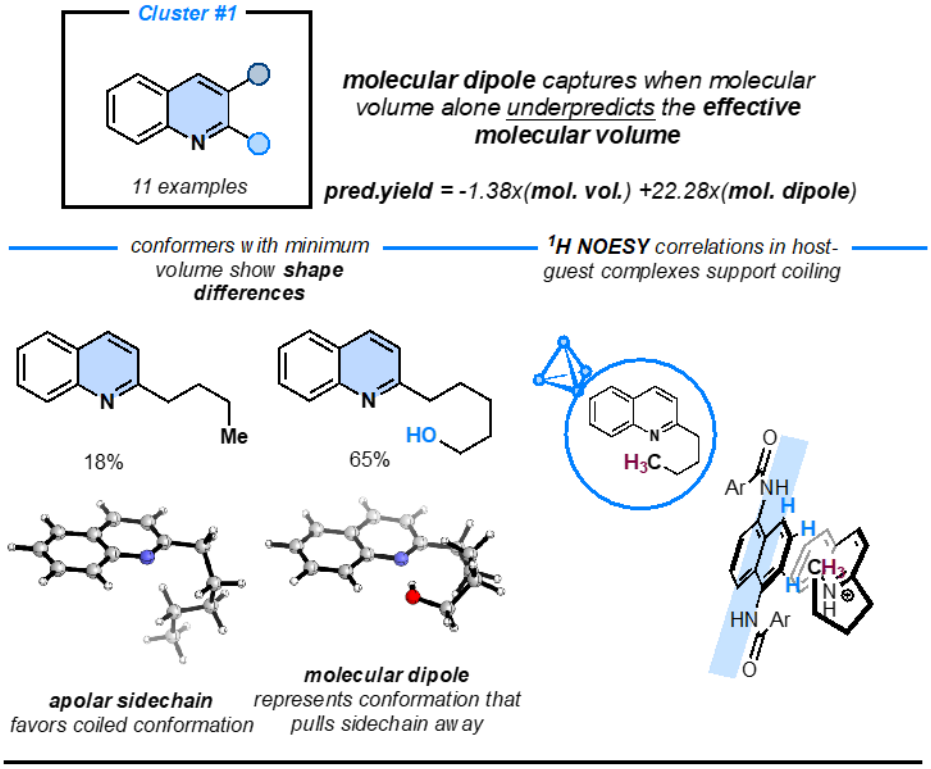


Supplementary Table VI-12 | Host **IV** catalyzed azaarene reduction yields (**bold**) and predicted yields (*italics*) based on *minimum molecular volume* and *molecular dipole* of cluster #1 substrates.

.

The model generated indicates that larger *molecular dipoles* correlated to improved yields in the host-catalyzed reduction. The correlation suggests *molecular dipole* formally correlates to a feature that *reduces* the effective molecular volume of the bound quinoline substrate, facilitating subsequent binding to the *reactive, ternary complex*. This effect is most pronounced when comparing 2-pentylquinoline with 5-(quinolin-2-yl)pentan-1-ol, an analog with a distal alcohol. Through inspection of each substrate's conformer with the *minimum molecular volume*, we can see the substrate without the pendant alcohol displays a coiled nonpolar sidechain, while the substrate with the pendant alcohol instead displays an intramolecular hydrogen bonding interaction that organizes the sidechain away from the reactive sites. Such preorganization of groups is reflected by the greater *molecular dipole* of the substrate with the pendant alcohol.

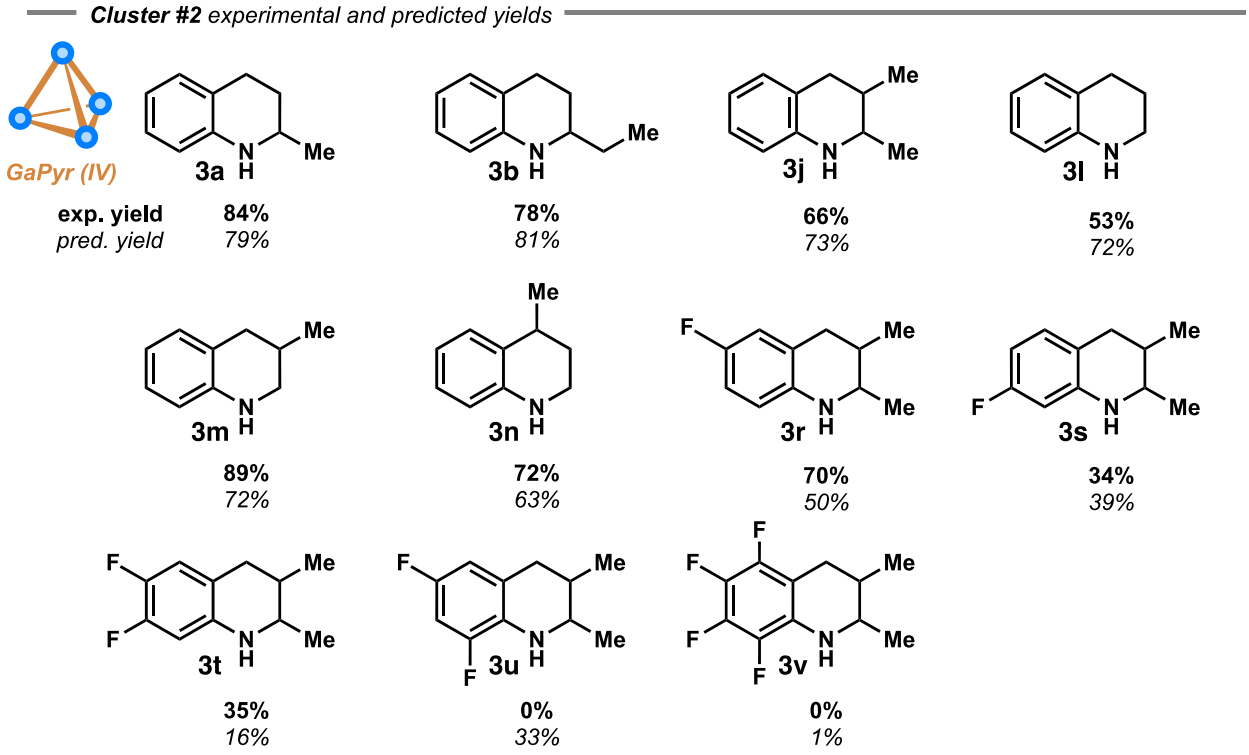
This conformational analysis is further supported by the isolated binding behavior of each substrate with host **I**. By ^{1}H NMR, 2-pentylquinoline was observed to strongly encapsulate in host **I**, displaying strong NOE-correlations between the methyl protons of the side chain and the aperture *C-H* bonds. In contrast, 5-(quinolin-2-yl)pentan-1-ol displays reversible binding, indicating the side chain is not in a well-defined conformation at the aperture, avoiding occlusion of reductant binding to react.



Supplementary Fig. VI-29 | Proposed analysis of the chemical meaning behind *molecular dipole* descriptor in the yield prediction model for cluster **I** substrates.

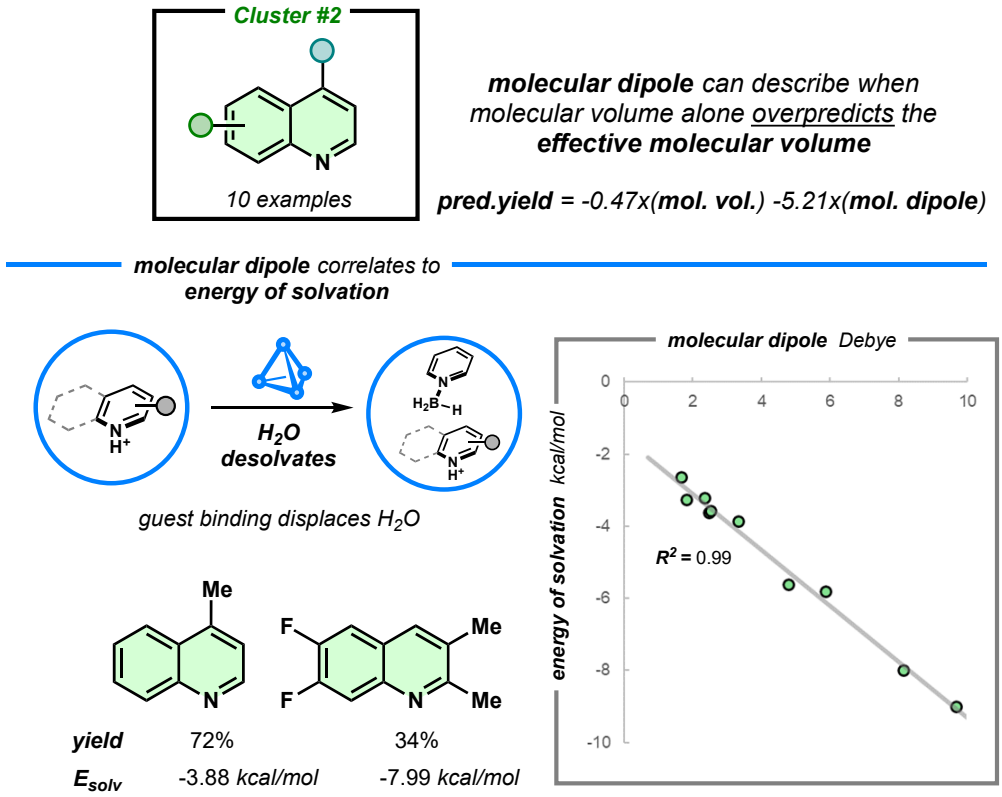
- ***Cluster #2 yield prediction task and proposed interpretation***

The yield prediction model built from *minimum molecular volume* and *molecular dipole* from quinoline substrates in Cluster #2 produced the following predictions.



Supplementary Table VI-13 | Host **IV** catalyzed azaarene reduction yields (**bold**) and predicted yields (*italics*) based on *minimum molecular volume* and *molecular dipole* of cluster #2 substrates.

In contrast to cluster #1, the model generated by cluster #2 indicates larger *molecular dipole* correlates to lower yields in the host-catalyzed reduction. The correlation suggests *molecular dipole* formally correlates to a feature that *increases* the effective molecular volume of the bound quinoline substrate, inhibiting subsequent binding to the *reactive, ternary complex*. Inspection of the representative substrates in cluster #2 reveals many quinolines are decorated with polar substituents not bound to flexible side chains. We anticipate these polar substituents recruit solvent water molecules to the quinoline undergoing reduction. While some of these molecules are likely displaced upon encapsulation into the host cavity, some water molecules may remain associated. To undergo reduction, these water molecules must be replaced with reductant. Such desolvation should be energetically more unfavorable for more polar substrates, and thus with larger *molecular dipoles*. Indeed, across the entire cluster #2, *molecular dipole* correlates excellently with the energy of solvation (R^2 = 0.98) as calculated by *xTB* from implicit solvation electronic corrections (see **Molecular descriptor calculation workflow** for details).



Supplementary Fig. VI-30 | Proposed analysis of the chemical meaning behind *molecular dipole* descriptor in the yield prediction model for cluster #2 substrates.

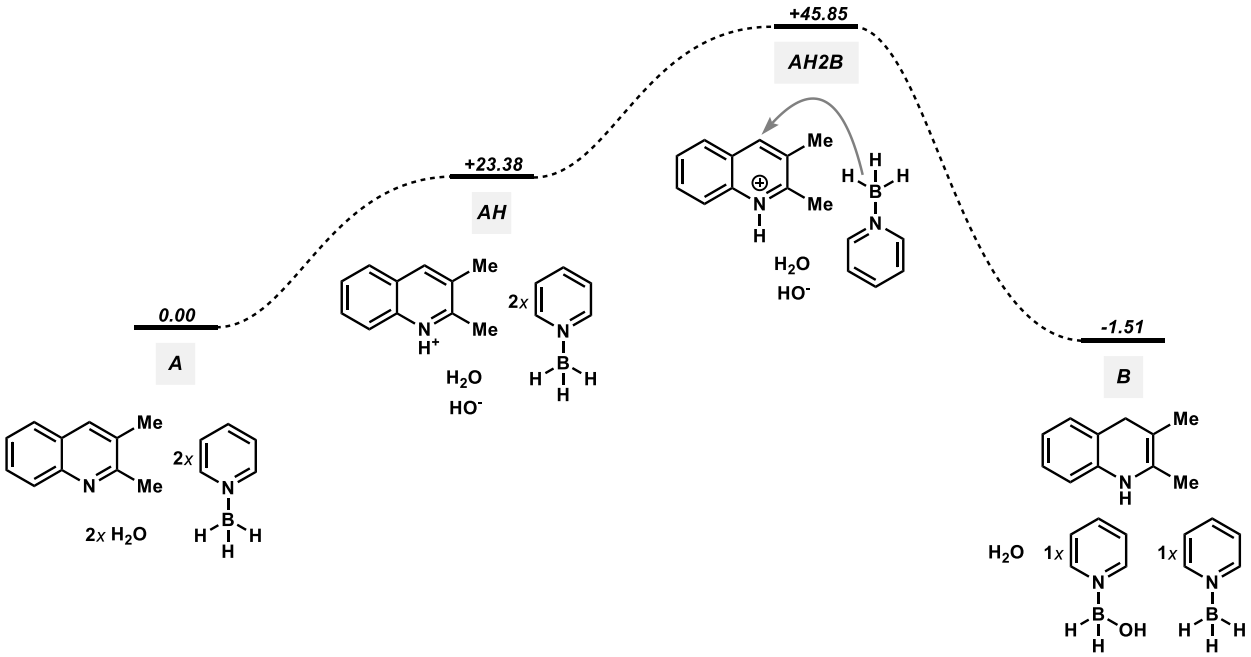
VI.21 – Extended Discussion on Figure II-5A

- ***DFT-level potential energy surface calculation of the reported azaarene reduction***

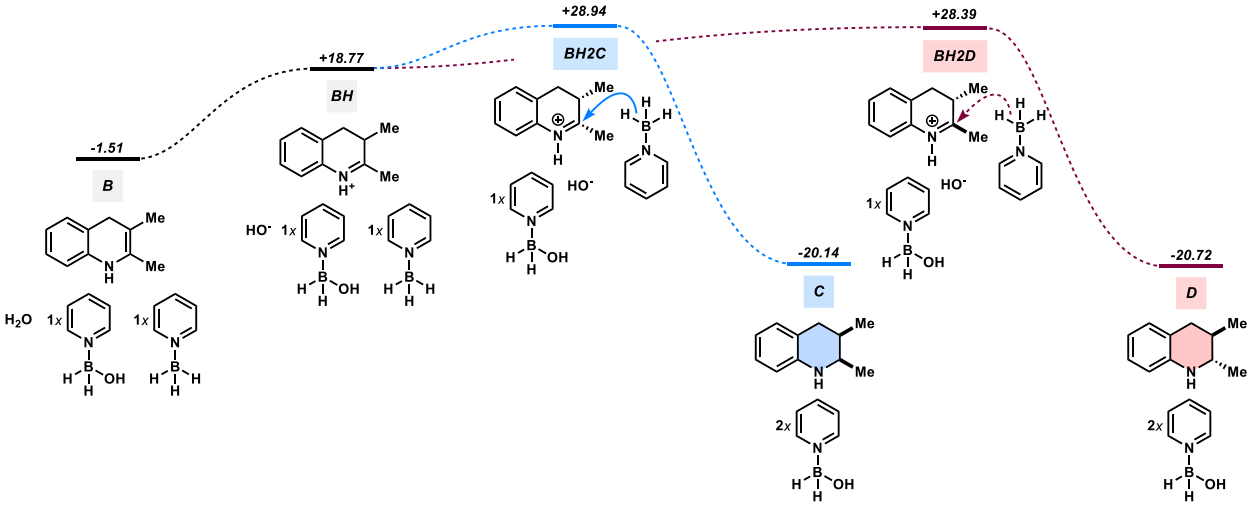
The host-mediated reduction yields diastereoselective outcomes in 2,3-disubstituted quinolines. This reduction is compared to a solution-phase reduction that requires the addition of exogenous acetic acid to drive the reduction as the background reduction in the absence of host is too slow to detect.[18] In these alternative conditions, the reduction also displays attenuated diastereoselectivity. For the reduction of 2,3-dimethylquinoline **1j,** 3.2:1 anti/syn is observed. This is compared to the more selective, catalyst **IV** mediated reduction in 10:1 anti/syn. To validate the acetic acid-mediated reduction proceeds through intermediates comparable to the proposed reduction mediated the supramolecular host, we computed a theoretical potential energy surface for the reduction.

Each structure in the proposed transformation were first optimized at the *GFN2/xTB* level of theory with implicit solvation by the *–aplb water* keyword. Then, these structures were subjected to optimization and frequency calculation at the high level of theory with Gaussian16, *wB97XD/6-311++g(d,p)/smd=water*. This level of theory was employed to characterize related azaarenes undergoing similar bond-forming transformations. The full computed surface is shown below.

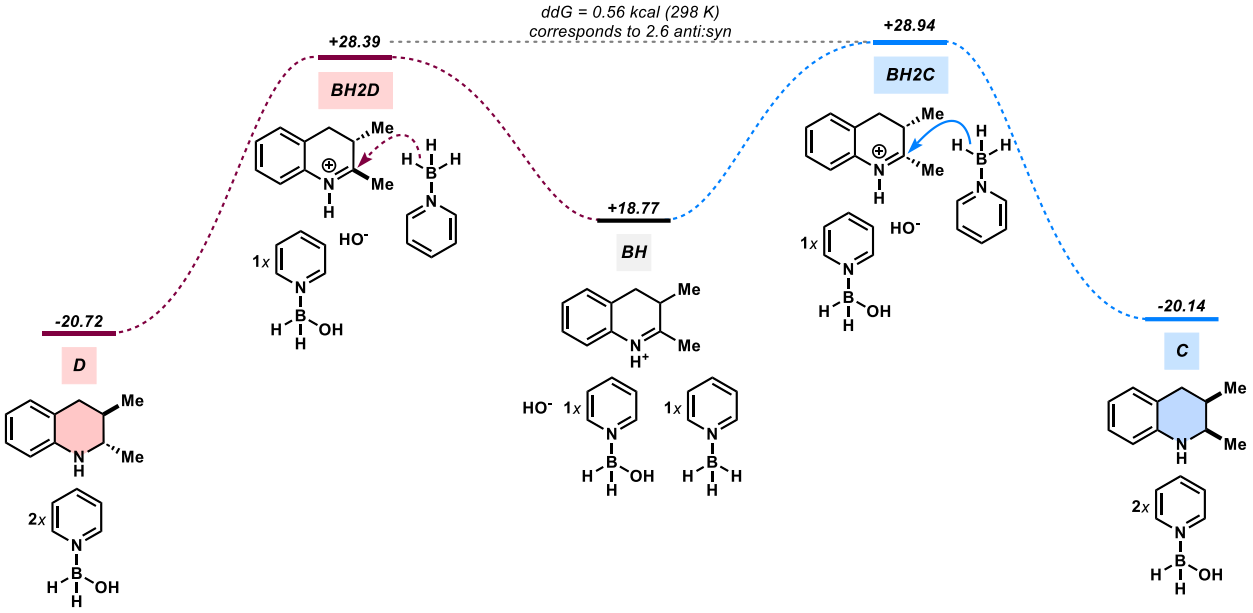
Each state is labeled with the frequency-corrected free energy at the *wB97XD/6-311++g(d,p)/smd=water* level of theory. Based on the frequency calculation, all ground states possess *zero* imaginary frequencies, while all transition states considered possess only a *single* imaginary frequency. The *xyz* coordinates of structure can be referenced in *XYZ coordinates*.



Supplementary Fig. VI-31 | Computed energy surface of the first reduction event. The states *A, AH,* and *B* refer to ground states. The state *AH2B* refers to a transition state.



Supplementary Fig. VI-32 | Computed energy surface of the second, stereodetermining reduction event. The states *B, BH, D,* and *C* refer to ground states. The states *BH2C* and *BH2D* refer to the transition states between the respective ground states.



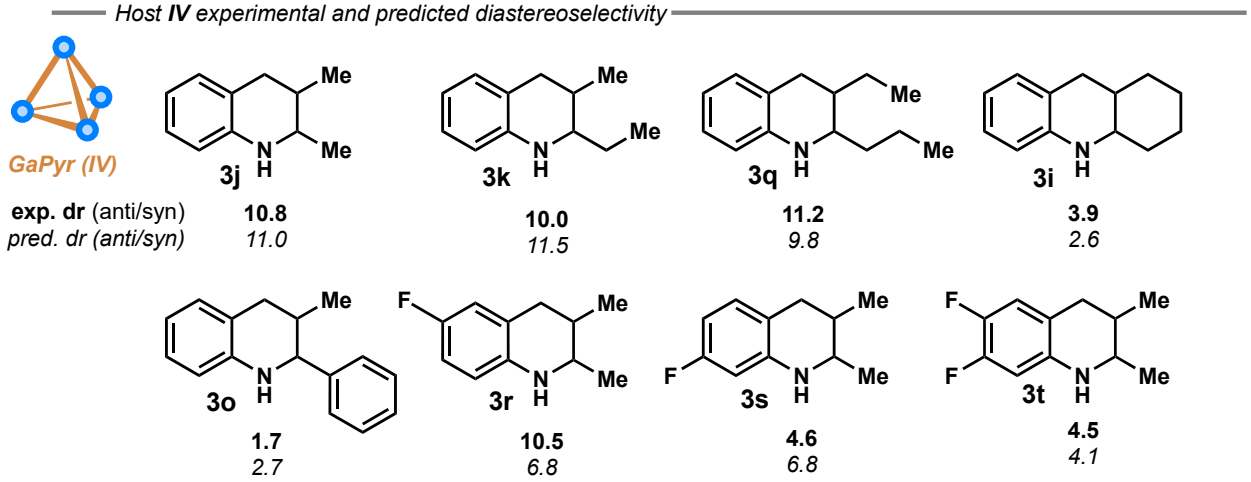
Supplementary Fig. VI-33 | Microscopic analysis of the two stereodetermining reduction pathways. The states *BH, D,* and *C* refer to ground states. The states *BH2D* and *BH2C* refer to the transition states between the respective ground states.

The solution state calculation at the high level of theory indicates an intrinsic preference for the *anti* diastereomer (state *D*) relative to the *syn* diastereomer (state *C*) by 0.56 *kcal/mol*. At 298 K, this energy corresponds to 2.6:1 *anti/syn*, which is approximately the selectivity observed by acetic acid (measured 3.2:1 *anti/syn* selectivity). This calculation supports the expectation that the specific microenvironment within the host cavity is responsible for the observed diastereoselectivity in the host-mediated reduction.

- ***Full dataset reported for the diastereoselective host IV-catalyzed azaarene reduction***

The diastereoselectivity prediction model built from a function of the calculated *electronegativity* and *sterimol BMin* of the *minimum volume conformer* of the quinoline substrates produced the following predictions for host **IV**-catalyzed reduction. The predictions were made in *kcal/mol* (termed *pred. ddG*) and converted to the *predicted diastereomeric ratio* (termed *pred. dr*) using the formula:

$$pred.dr = e^{pred.\ ddG/(0.00198*323)}, \text{ based on } \Delta G = -RTln(K) \text{ at } 50^{o}C.$$



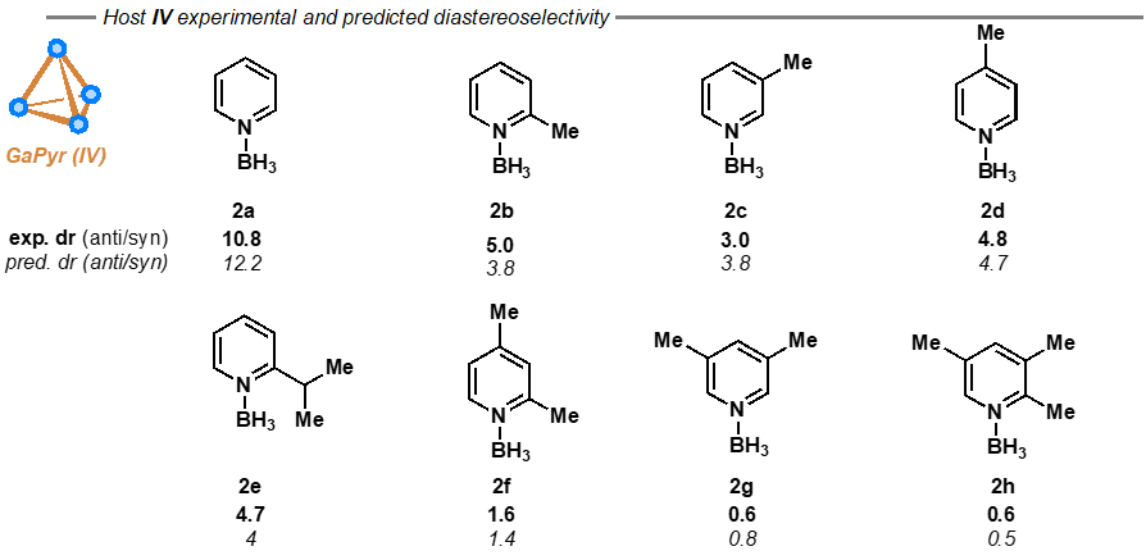
Supplementary Table VI-14 | Host **IV** catalyzed azaarene reduction diastereoselectivity (bold) and predicted diastereoselectivity (*italics*) based on the calculated *electronegativity* and *sterimol BMin* of the *minimum volume conformer* of the quinoline substrate.

VI.22 – Extended Discussion on Figure II-5B

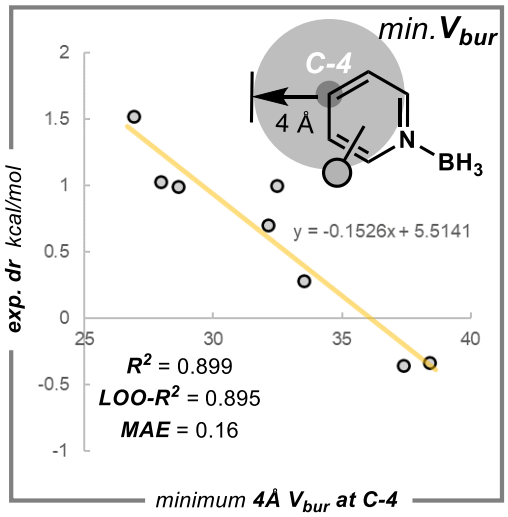
- ***Full dataset reported for the diastereoselective host IV-catalyzed azaarene reduction***

The diastereoselectivity prediction model built from the calculated *4Å buried volume at C-4* of the pyridine-borane reductants produced the following predictions for host **IV**-catalyzed reduction of 2,3-dimethylquinoline. The predictions were made in *kcal/mol* (termed *pred. ddG*) and converted to the *predicted diastereomeric ratio* (termed *pred. dr*) using the formula:

$$pred.dr = e^{pred.\ ddG/(0.00198*323)}, \text{ based on } \Delta G = -RTln(K) \text{ at } 50^{o}C.$$



Supplementary Table VI-15 | Host **IV** catalyzed azaarene reduction diastereoselectivity (**bold**) and predicted diastereoselectivity (*italics*) based on the calculated *4Å buried volume at C-4* of the reductant.



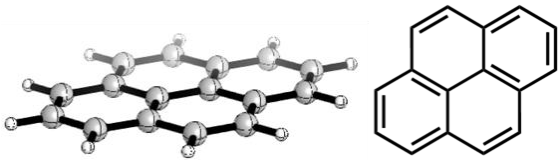
Supplementary Fig. VI-34 | Linear correlation between the experimental diastereoselectivity and the calculated *4Å buried volume at C-4* of the reductant. Data is from host IV catalyzed reduction of 2,3-dimethylquinoline with various pyridine boranes.

VI.23 – Extended Discussion on Figure II-5C

- ***Non-covalent interaction calculation specifications***

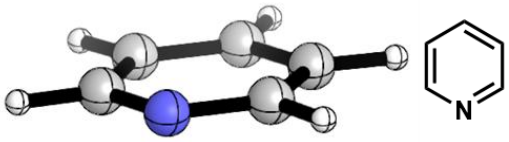
From **Figure II-5C.** non-covalent interactions across the proposed, stereodetermining ternary reactive complexes, the *substrates* in these complexes refer to the corresponding quinolines undergoing reduction. The *probes* in these complexes refer to either the (1) **pyrene**, or (2) **pyridine** molecules. The *xyz* coordinates of each *probe* molecule used are reported below. Their corresponding figures that utilize data from these *probes* is reported for reference.

Pyrene



C	1.4175049917	-1.2262237518	-0.0003203124
C	0.7077398055	-0.0000030912	-0.0005121592
C	-0.7077237363	-0.0000186832	-0.0004927716
C	-1.4175300082	-1.2262154265	-0.0003352170
C	-0.6764491735	-2.4454583653	-0.0010864128
C	0.6764810748	-2.4455134947	-0.0010591708
C	-2.8130949459	-1.2018497659	0.0007045561
C	-3.4977835384	-0.0000042001	0.0015301933
C	-2.8130240131	1.2018888130	0.0009643807
C	2.8130389993	-1.2017976495	0.0007596653
C	3.4978005480	0.0000410136	0.0015539822
C	2.8130365408	1.2018017324	0.0009026338
C	1.4175185847	1.2262216468	-0.0003022004
C	-1.4175227898	1.2261909641	-0.0002841722
C	-0.6764729036	2.4454578213	-0.0012081026
C	0.6764826310	2.4454644618	-0.0011618471
H	1.2285750667	-3.3743483748	-0.0014553131
H	-1.2284393264	-3.3743717629	-0.0015800166
H	-3.3563423066	-2.1355927358	0.0010268649
H	-4.5778286710	0.0001425249	0.0029451408
H	-3.3562697943	2.1356545776	0.0017127253
H	3.3563030853	-2.1355068459	0.0010571733
H	4.5777956905	-0.0000156475	0.0027029979
H	3.3561800974	2.1355065212	0.0016274681
H	-1.2284913904	3.3743749248	-0.0019493632
H	1.2284929184	3.3743710277	-0.0019530667

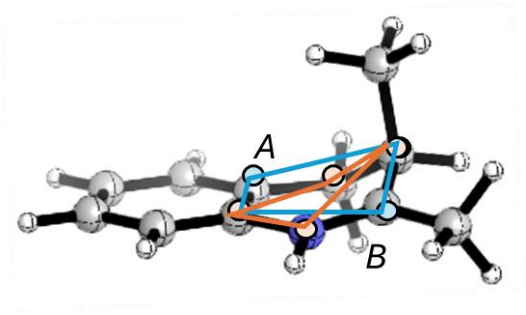
Pyridine



N	1.3748666514	-0.0003236518	0.0000080474
C	0.6905473498	-1.1342172230	0.0000149379
C	-0.6949378688	-1.1918345581	-0.0000330744
C	-1.4020050354	0.0003452543	-0.0000330656
C	-0.6942997109	1.1921647404	0.0000097178
C	0.6910998602	1.1338635108	-0.0000276945
H	1.2883051834	-2.0358573451	0.0000448878
H	-1.2026643411	-2.1429616019	0.0002945401
H	-2.4817879897	0.0007002806	0.0000504610
H	-1.2016555467	2.1435351224	0.0003167161
H	1.2893451935	2.0352473613	0.0000059560

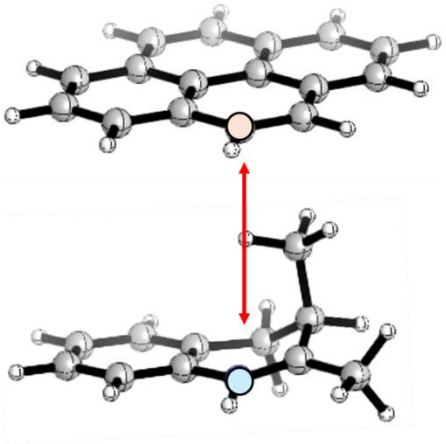
The *xyz* coordinates of the corresponding *pro-syn* and *pro-anti* dihydroquinolines are included for reference in **XYZ Coordinates.**

The reported non-covalent interactions were computed using a few alternative constraints to preserve the integrity of the scanned π- π complex. These constraints include: (1) the *substrate* dihydroquinoline π system is frozen rather than set to be fully planar to avoid any unrealistic distortions and (2) the pyrene *probe* π system is frozen as well to avoid any distortions that reduce π- π interactions. An example of the frozen dihedrals is shown below.



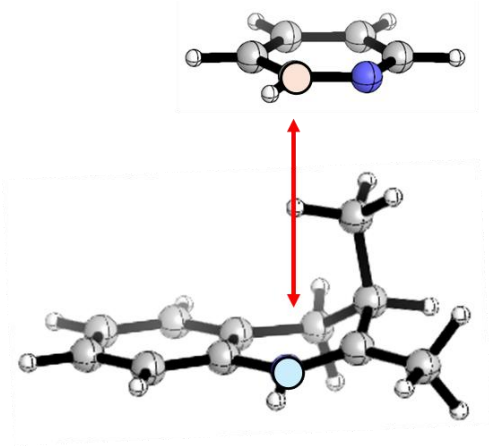
Supplementary Fig. VI-35 | The dihydroquinoline *substrate* considered in the calculated non-covalent interactions uses *frozen dihedrals* marked with *A* and *B* rather than dihedrals constrained to 0°

The reported non-covalent interactions involved combination of interactions across two *substrate-probe complexes* – specifically between one *substrate-pyrene probe* complex and one *substrate-pyridine probe* complex. With respect to the *pyrene probe*, the most favorable, dispersive interactions could be captured when the *pyrene probe* was placed in a conformation reaching along the *substrate*. Therefore, this conformation was considered.



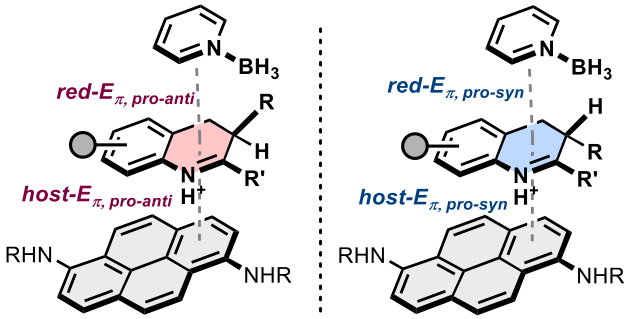
Supplementary Fig. VI-36 | An example of one *substrate-pyrene probe complex.*

With respect to the *pyridine probe*, the conformation corresponding to the interaction leading to the productive reduction event at *C*-2 of the dihydroquinoline *substrate* was considered. An example of the *substrate-probe complex* is shown for reference.



Supplementary Fig. VI-37 | An example of one *substrate-pyridine probe complex*.

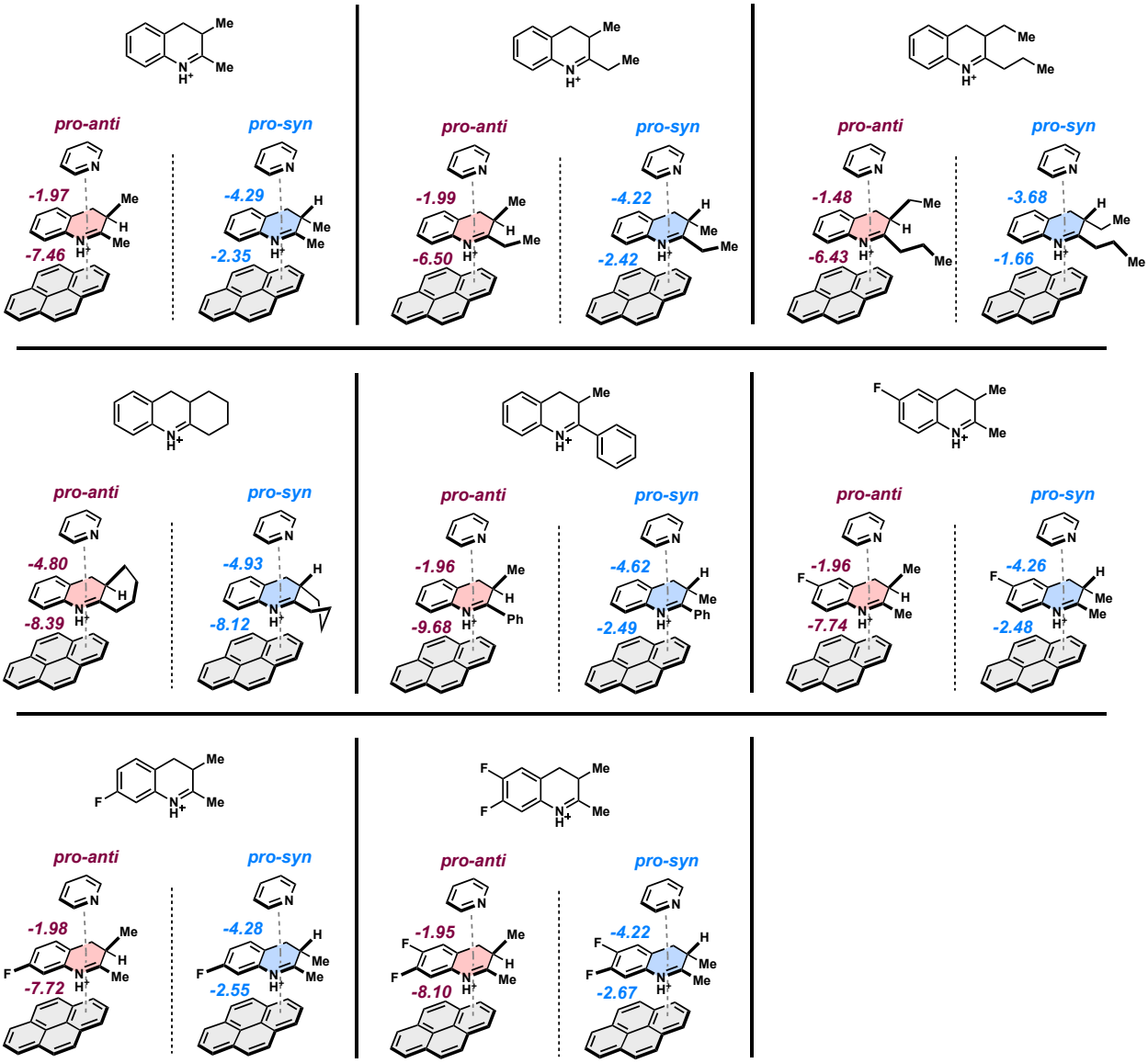
Conformations were classified as *pro-anti* or *pro-syn* depending on the *probe* and the *substrate-probe* complex. These definitions are made if they resemble the following complexes.



Supplementary Fig. VI-38 | Defining *pro-anti* and *pro-syn* for the *pyrene* and *pyridine probes*.

- ***Full dataset calculated for the non-covalent interactions reported***

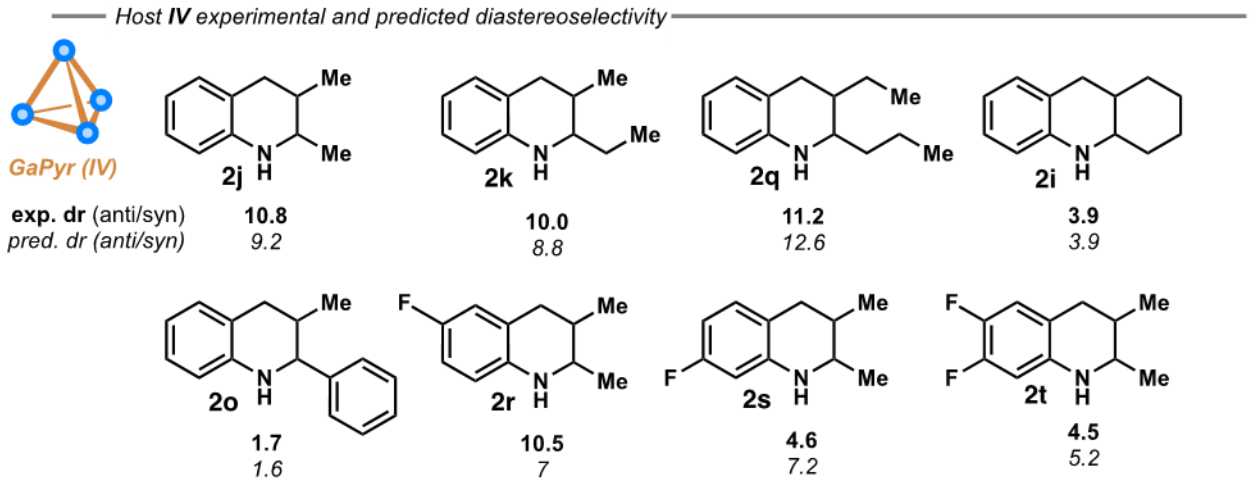
The calculated non-covalent interactions in proposed *pro-anti* and *pro-syn* complexes for the respective azaarenes are reported below. Each interaction was computed separately using the methods described above.



Supplementary Table VI-16 | Calculated non-covalent interactions between the putative cationic, dihydroquinoline intermediates and the *pyrene* and *pyridine* probes, reported in *kcal/mol*.

The diastereoselectivity prediction model built from a function between the competing non-covalent interactions in *pro-anti* and *pro-syn* complexes calculated for the quinoline substrates above produced the following predictions for host **IV**-catalyzed reduction. The predictions were made in *kcal/mol* (termed *pred. ddG*) and converted to the *predicted diastereomeric ratio* (termed *pred. dr*) using the formula:

$$pred.dr = e^{pred.\ ddG/(0.00198*323)}$$, based on $\Delta G = -RTln(K)$ at 50°C.

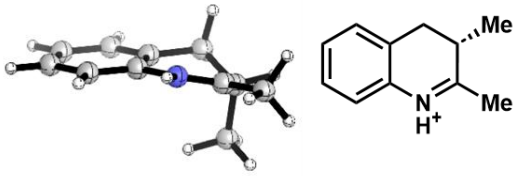


Supplementary Table VI-17 | Host **IV** catalyzed azaarene reduction diastereoselectivity (**bold**) and predicted diastereoselectivity (*italics*) based on the competing non-covalent interactions in *pro-anti* and *pro-syn* complexes reported.

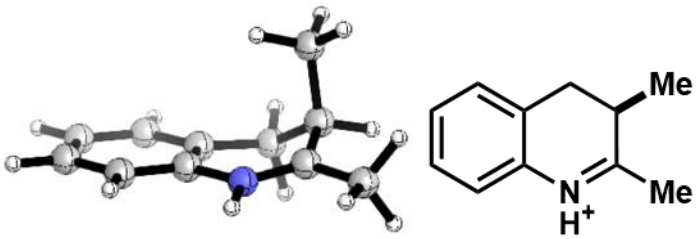
VI.24 – Extended Discussion on Figure II-5D

- ***Non-covalent interaction calculation specifications***

From **Figure II-5D.** non-covalent interactions across the reductant-controlled, proposed, stereodetermining reactive complexes, the *substrates* in these complexes refer to the corresponding pyridine-boranes reducing the quinoline. These are approximated as the respective pyridines to reduce computational complexity during the coordinate scan. The *probe* in these complexes refers to the **cationic dihydroquinoline** undergoing the stereodetermining reduction (2,3-dimethyl-3,4-dihydroquinolin-1-ium, **1j**). Both ***pro-syn*** and ***pro-anti*** complexes are included below.



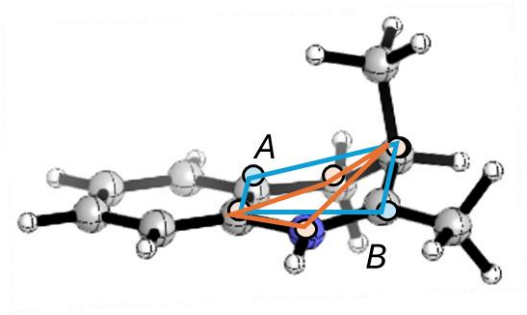
C	-1.8570169171	-0.5346127543	-0.1819652029
N	-0.7655936120	-1.2275118402	-0.1369145542
C	0.5226418929	-0.6754596462	-0.0689778603
C	1.6015190770	-1.4789673538	0.2787838661
C	2.8665362231	-0.9235355997	0.3393798953
C	3.0478280686	0.4197098293	0.0460127016
C	1.9679834727	1.2135913214	-0.3089550852
C	0.6922022513	0.6786292788	-0.3686469927
C	-0.5119498130	1.4667017429	-0.7885115203
C	-1.7724284229	0.9524495037	-0.0851812135
C	-3.1629013988	-1.1924614834	-0.2767321183
C	-1.7521601687	1.3114830504	1.4123293440
H	-3.1132096727	-2.2586727031	-0.0774099734
H	-3.8774042654	-0.7199083996	0.3943235969
H	-3.5567504708	-1.0455095404	-1.2843608378
H	-0.8121019743	-2.2463108655	-0.1347006681
H	1.4540002627	-2.5224612508	0.5087768833
H	3.7111119372	-1.5357664890	0.6121971788
H	4.0356025118	0.8505431020	0.0930830005
H	2.1191573723	2.2561691358	-0.5412082273
H	-0.6263427452	1.3745072257	-1.8729446256
H	-0.3700050693	2.5231491850	-0.5651342124
H	-2.6582017473	1.4046551861	-0.5346319299
H	-0.8654257024	0.9086946650	1.8949831333
H	-1.7383499725	2.3920162114	1.5191896607
H	-2.6332005483	0.9300787988	1.9198922118



C	-1.8568039805	-0.5349987738	0.1818364830
N	-0.7652703416	-1.2276448833	0.1362823800
C	0.5229064770	-0.6754372178	0.0683690065
C	1.6019260174	-1.4788127675	-0.2790746618
C	2.8669877580	-0.9232931246	-0.3390173537
C	3.0479777039	0.4200310955	-0.0458718172
C	1.9678767194	1.2138372450	0.3086960097
C	0.6921956891	0.6786942761	0.3682287534
C	-0.5120500250	1.4663422795	0.7886269325
C	-1.7727712967	0.9521359530	0.0856023719
C	-3.1624558574	-1.1934152723	0.2769909244
C	-1.7536865579	1.3125725873	-1.4115523837
H	-3.8762250891	-0.7237208146	-0.3969271508
H	-3.1117542415	-2.2603250644	0.0811179477
H	-3.5579323335	-1.0432168685	1.2834698035
H	-0.8116162624	-2.2464555078	0.1338632464
H	1.4546408983	-2.5222297791	-0.5094300698
H	3.7116222795	-1.5354491415	-0.6117188860
H	4.0356874173	0.8510991243	-0.0925150547
H	2.1190225049	2.2563166647	0.5413018409
H	-0.3704122719	2.5228969241	0.5655846260
H	-0.6259657437	1.3737221697	1.8730870043
H	-2.6584174557	1.4037165222	0.5359242490
H	-2.6350345226	0.9314792086	-1.9188632903
H	-1.7402020687	2.3932185133	-1.5173648012
H	-0.8672188482	0.9104511779	-1.8952095620

The *xyz* coordinates of the corresponding pyridine borane reductants, simplified to the analogous pyridine, are included for reference in **XYZ Coordinates.**

The reported non-covalent interactions were computed using a similar alternative constraints to preserve the integrity of the scanned π- π complex as described above. Here, the *probe* dihydroquinoline π system is frozen rather than set to be fully planar to avoid any unrealistic distortions.



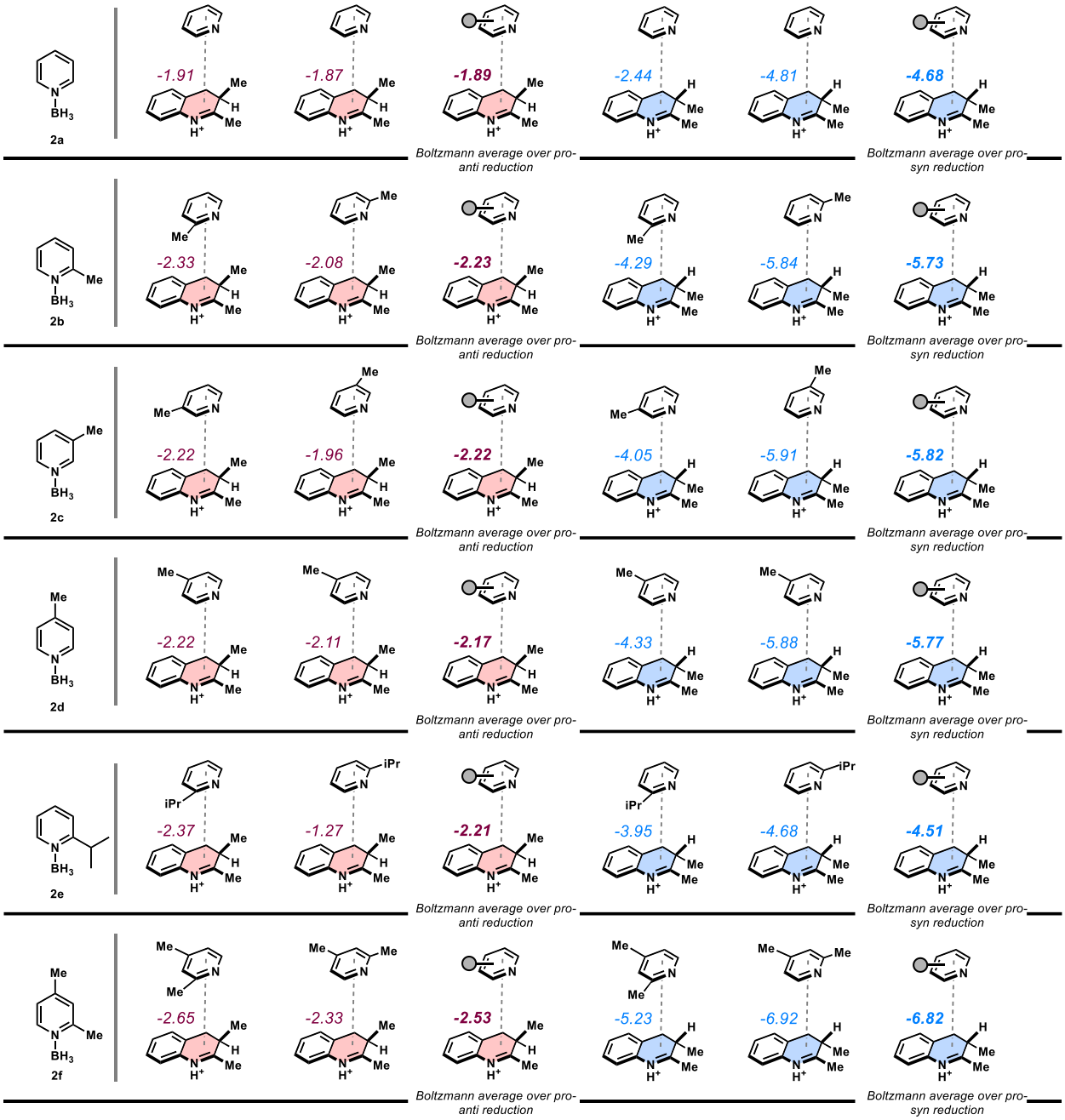
Supplementary Fig. VI-39 | For example, the *pro-anti* dihydroquinoline *probe* in the calculated non-covalent interactions uses *frozen dihedrals* marked with *A* and *B* rather than dihedrals constrained to 0°.

Consequentially, the coordinate scan that is computed is not exactly set between two parallel π systems. While this still permits the calculation of the non-covalent interaction in the stacked π- π complex, the calculation becomes sensitive to the trajectory of the *probe*'s incoming π system. This is apparent in the below calculations where formally symmetric conformations generate distinct non-covalent interactions, owing to the subtle changes in the trajectory into the nonplanar, dihydroquinoline *probes*.

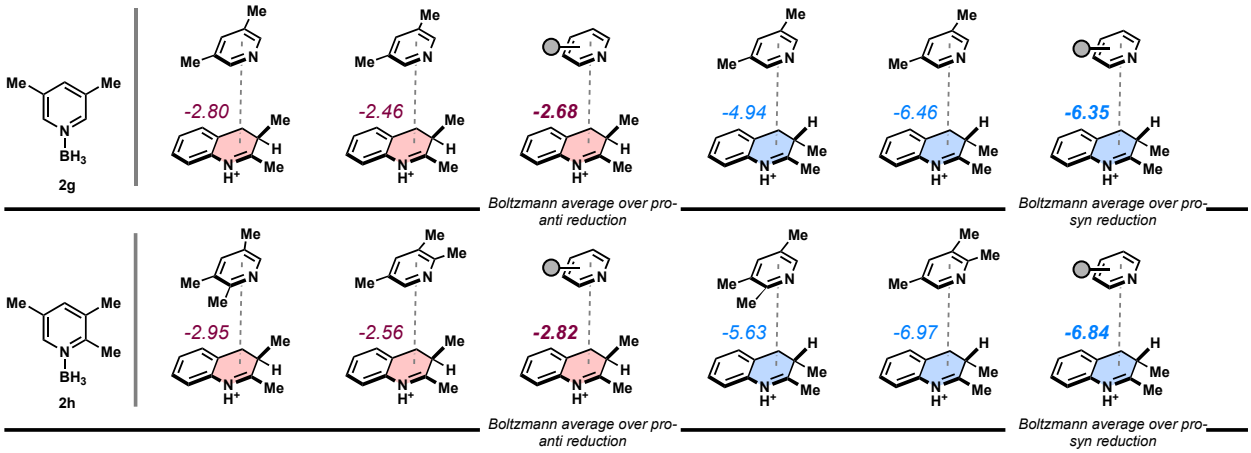
In this report, the non-covalent interactions between a series of pyridine borane reductants and the model 2,3-dimethylquinoline is considered. Because some of the pyridine borane reductants are not symmetrically substituted, two conformers are considered where the *N* atom is aligned above the *C*-2 position of the dihydroquinoline, where the reduction occurs. To account for the instability in the calculation and the inability to distinguish either trajectory in the real transformation, we computed the *Boltzmann averaged* interaction over each conformer between the *pro-anti* and *pro-syn* complexes. These *Boltzmann averaged* non-covalent interactions were used in the diastereoselectivity prediction tasks for the dataset generated by varying the pyridine borane reagent.

- ***Full dataset calculated for the non-covalent interactions reported***

For each reductant screened, the calculated non-covalent interactions in the proposed *pro-anti* and *pro-syn* complexes for the respective azaarenes are reported below.



... continued onto next page ...



Supplementary Fig. VI-40 | Calculated non-covalent interactions between the putative cationic, dihydroquinoline *probes* and *pyridine* substrates, reported in *kcal/mol*.

The diastereoselectivity prediction model built from a function of the *Boltzmann averaged* non-covalent interaction computed between the reductant and the dihydroquinoline *probe* in the *pro-anti* complex produced the following predictions. These predictions were made in *kcal/mol* (*pred. ddG*) and converted to the *predicted diastereomeric ratio* (*pred. dr*) using the formula:

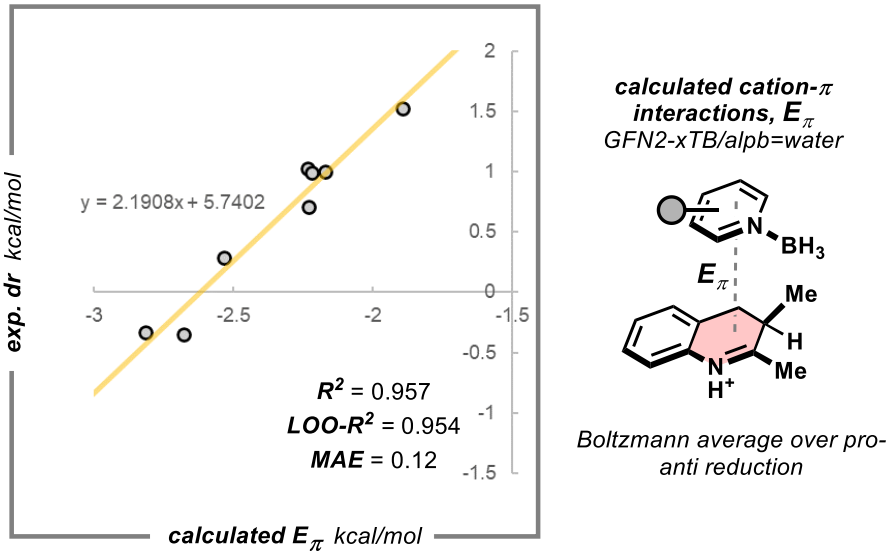
$$pred.dr = e^{pred.\ ddG/(0.00198*323)}, \text{ based on } \Delta G = -RTln(K) \text{ at } 50^{o}\text{C}.$$

*Host **IV** experimental and predicted diastereoselectivity*

	2a	2b	2c	2d
exp. dr (anti/syn)	**10.8**	**5.0**	**3.0**	**4.8**
pred. dr (anti/syn)	*12.2*	*3.8*	*3.8*	*4.7*

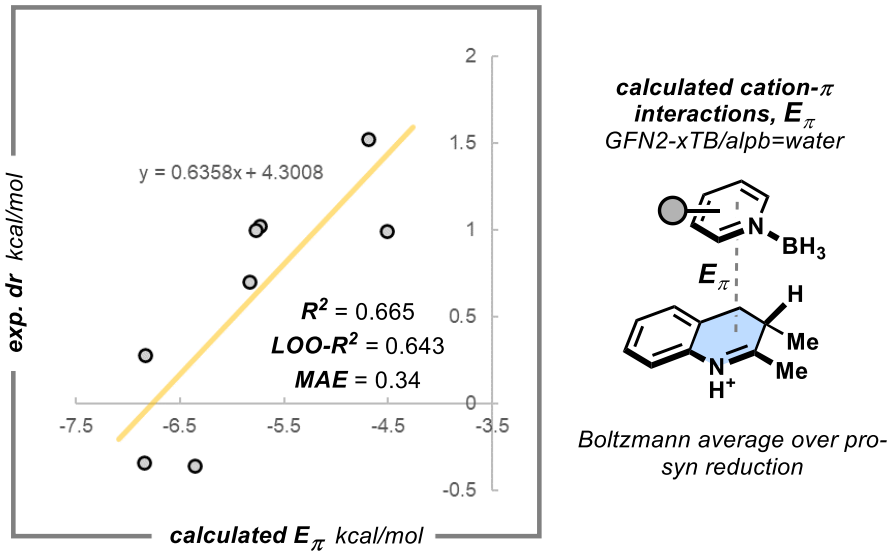
	2e	2f	2g	2h
	4.7	**1.6**	**0.6**	**0.6**
	4	*1.4*	*0.8*	*0.5*

Supplementary Table VI-18 | Host IV catalyzed azaarene reduction diastereoselectivity (bold) and predicted diastereoselectivity (*italics*) based on the *Boltzmann averaged pro-anti* non-covalent interaction between the reductant and the dihydroquinoline *probe*.



Supplementary Fig. VI-41 | Linear correlation between the experimental diastereoselectivity and the computed *Boltzmann averaged* non-covalent interactions in the *pro-anti* reduction complex. Data is from host IV catalyzed reduction of 2,3-dimethylquinoline with various pyridine boranes.

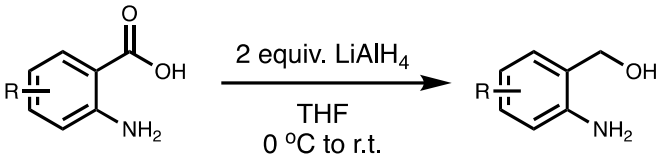
The *Boltzmann averaged pro-syn* interactions computed also provide a meaningful correlation to the experimental diastereoselectivity but perform statistically worse than the *pro-anti* counterparts. For reference, these are reported below.



Supplementary Fig. VI-42 | Linear correlation between the experimental diastereoselectivity and the computed *Boltzmann averaged* non-covalent interactions in the *pro-syn* reduction complex. Data is from host IV catalyzed reduction of 2,3-dimethylquinoline with various pyridine boranes.

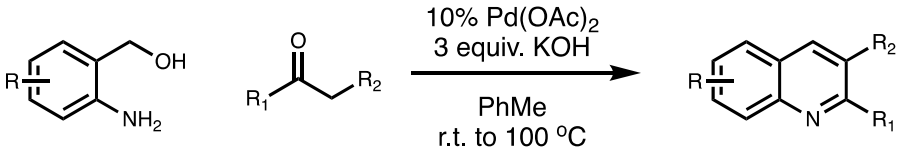
VI.25 – Substrate and Product Standards Synthesis

- ***Procedures for Quinoline Starting Material Synthesis:***



2-amino-benzyl alcohol synthesis, adapted from a literature procedure.[19]

To a stirring, degassed suspension of $LiAlH_4$ (30 mmol) in dry THF (40 mL) was added a dry degassed THF solution (10 mL) of 2-amino-benzoic acid (15 mmol) at 0 °C. The solution was allowed to warm to room temperature and stirred for 3 hours. The reaction was again cooled to 0 °C and then 1 mL of water, followed by 1 mL of 5% KOH and 3 mL of water. The reaction was warmed to room temperature and Na_2SO_4 was added. The suspension was filtered and concentrated to dryness and used in the next step without further purification.



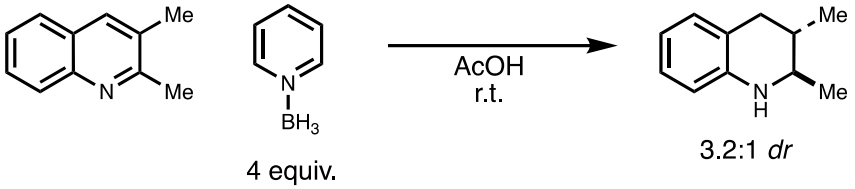
Standard Quinoline Synthesis, adapted from a literature procedure.[20]

2-amino-benzyl alcohol, (5 mmol) KOH, (15 mmol) and $Pd(OAc)_2$ were added to a round bottom flask followed by 15 mL of toluene. Then ketone (10 mmol) was added to the solution and the suspension was heated to 100 °C for 16 hours. The resulting dark solution and suspension were concentrated to dryness and then extracted with ~10 mL DCM and filtered. This solution was loaded directly onto a silica gel column and purified.

- ***Synthesis of Novel Product Standards General Procedure:***

0.3 mmol of quinoline substrate and 0.03 mmol $NiCl_2$ hydrate was dissolved in 600 mL of MeOH and cooled to 0 ºC. While stirring, $NaBH_4$ (1.2 mmol) was then added in portions and the reaction was warmed to room temperature and stirred for three hours. The reaction was filtered and loaded directly onto a silica column for purification.

- ***$PyBH_3$ Reduction of 2,3-dimethylquinoline in AcOH***



Adapted from a literature procedure.[18]

2,3-dimethylquinoline (0.3 mmol) was dissolved in 600 mL AcOH and stirred. Pyridine-Borane (1.2 mmol) was added and stirred for 6 hours. The reaction was concentrated to dryness, redissolved in DCM, and loaded onto a silica gel column for purification.

Yield: 51% (NMR) *dr* 3.2:1 anti:syn

- ***General Procedure for the Synthesis of Pyridine-Borane Derivatives***

Adapted from a literature procedure.[21]

Pyridine derivative (5 mmol) was dissolved in 10 mL of tetrahydrofuran. $NaBH_4$ (10 mmol) and $KHCO_3$ (20 mmol) was then added to the solution and the resulting suspension was cooled with an ice bath. 500 mL of H_2O was then added dropwise with stirring resulting in vigorous gas evolution. The solution was allowed to warm to room temperature and stir for 16 hours. The suspension was filtered and the filtrate was concentrated under reduced pressure. The resulting material was dissolved in DCM and loaded directly onto a silica column for purification eluting with DCM.

VI.26 – Characterization Data of Novel Compounds

Substrates **1a**, **1h**, **1l**, **1m** and **1n** were used without further purification from commercial suppliers.

Substrates **1b-g**, **1i-l**, and **1o-1u** were prepared according to the general quinoline synthesis. **1b-g**, **1i-1q** were previously characterized.[5,22,23]

Pyridine-boranes **2a-g** have been previously characterized.[24–26]

Products **3a-3i**, **3i-3q** were previously characterized. [5,22,23]

1r- 6-fluoro-2,3-dimethylquinoline
Prepared from (2-amino-5-fluorophenyl)methanol and 2-butanone according to standard quinoline synthesis.
Yield 33%
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.97 (dd, *J* = 9.1, 5.3 Hz, 1H), 7.77 (s, 1H), 7.37 (ddd, *J* = 9.1, 8.4, 2.8 Hz, 1H), 7.30 (dd, *J* = 9.0, 2.9 Hz, 1H), 2.66 (s, 3H), 2.44 (d, *J* = 1.0 Hz, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 161.15, 159.19, 158.44 (d, *J* = 2.7 Hz), 143.65, 134.75 (d, *J* = 5.1 Hz), 131.16, 130.79 (d, *J* = 9.2 Hz), 128.09 (d, *J* = 10.0 Hz), 118.52 (d, *J* = 25.6 Hz), 109.87 (d, *J* = 21.6 Hz), 23.57, 19.76.
^{19}F NMR (471 MHz, $CDCl_3$) δ -115.13 (q, *J* = 7.9 Hz).
HRMS (ESI+) [$C_{11}H_{11}FN$] m/z calculated 176.08700, found 176.08607

1s - 7-fluoro-2,3-dimethylquinoline
Prepared from (2-amino-4-fluorophenyl)methanol and 2-butanone according to standard quinoline synthesis.
Yield 36%
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.81 (s, 1H), 7.68 (dd, *J* = 8.9, 6.1 Hz, 1H), 7.62 (dd, *J* = 10.4, 2.5 Hz, 1H), 7.23 (td, *J* = 8.6, 2.6 Hz, 1H), 2.67 (s, 3H), 2.43 (s, 2H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 163.60, 161.63, 160.34, 147.35 (d, *J* = 12.4 Hz), 135.31, 129.51 (d, *J* = 2.7 Hz), 128.73 (d, *J* = 9.7 Hz), 124.53, 116.13 (d, *J* = 25.3 Hz), 112.16 (d, *J* = 20.3 Hz), 23.71, 19.63.
^{19}F NMR (471 MHz, $CDCl_3$) δ -111.46.
HRMS (ESI+) [$C_{11}H_{11}FN$] m/z calculated 176.08700, found 176.08668

1t - 6,7-difluoro-2,3-dimethylquinoline
Prepared from (2-amino-4,5-difluorophenyl)methanol and 2-butanone according to standard quinoline synthesis.
Yield 18%
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.80 (s, 1H), 7.18 – 7.09 (m, 2H), 2.71 (s, 3H), 2.46 (d, *J* = 1.0 Hz, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 160.15 (d, *J* = 11.7 Hz), 159.10 (d, *J* = 12.3 Hz), 158.14, 157.19, 134.61 (d, *J* = 2.4 Hz), 132.70, 128.86, 105.83 (dd, *J* = 21.4, 4.9 Hz), 104.21 (dd, *J* = 29.4, 22.8 Hz), 23.74, 19.88.
^{19}F NMR (471 MHz, $CDCl_3$) δ -112.17 (q, *J* = 6.5 Hz), -121.16 (t, *J* = 9.0 Hz).
HRMS (ESI+) [$C_{11}H_{10}F_2N$] m/z calculated 194.07758, found 194.07730

1u - 5,6,7,8-tetrafluoro-2,3-dimethylquinoline
Prepared from (2-amino-2,4,5,6-tetrafluorophenyl)methanol and 2-butanone according to standard quinoline synthesis.
Yield 6%
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 (s, 1H), 2.74 (s, 3H), 2.50 (s, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 127.71, 23.95, 20.00.
^{19}F NMR (376 MHz, $CDCl_3$) δ -151.97 (dd, *J* = 19.3, 15.4 Hz), -152.40 – -153.75 (m), -156.53 (t, *J* = 18.7 Hz), -159.28 (t, *J* = 19.4 Hz).
HRMS (ESI+) [$C_{11}H_{18}F_4N$] m/z calculated 230.05874, found 230.05746

2h - 2,3,5-trimethylpyridine-borane
Prepared from 2,3,5-trimethylpyridine according to the general synthesis of pyridine-boranes.
Yield 78%
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.46 (s, 1H), 7.47 (s, 1H), 2.91 – 1.98 (m, 12H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 153.77, 146.72, 141.38, 134.26, 131.69, 19.79, 18.31, 17.93.

3p - 5-(1,2,3,4-tetrahydroquinolin-2-yl)pentan-1-ol
Compound **1p** was synthesized according to a literature procedure.[23]
Substrate **2p** was synthesized via an adapted literature procedure.[27]
Compound **1p**, (5 mmol) was combined with .5 mmol 10% Pd/C was dissolved in degassed methanol and placed under an atmosphere of H_2. The solution was heated to 50 °C and stirred for 16 hours. The solution was placed under a N_2 atmosphere, then filtered through celite, rinsing with excess MeOH. The substrate was then purified by silica column chromatography.
Yield 23% The remaining mass balance was the incompletely reduced pentenyl quinolinol.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.05 – 6.84 (m, 1H), 6.60 (td, *J* = 7.4, 1.2 Hz, 0H), 6.56 – 6.41 (m, 0H), 3.66 (t, *J* = 6.6 Hz, 1H), 3.25 (dtd, *J* = 9.5, 6.2, 2.9 Hz, 0H), 2.89 – 2.66 (m, 1H), 1.96 (dddd, *J* = 12.7, 5.6, 4.1, 2.9 Hz, 1H), 1.68 – 1.55 (m, 1H), 1.51 (dt, *J* = 11.2, 4.4 Hz, 1H), 1.47 – 1.39 (m, 1H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 144.72, 129.39, 126.85, 121.55, 117.14, 114.24, 63.03, 51.64, 36.77, 32.82, 28.19, 26.51, 26.04, 25.64.
HRMS (ESI+) [$C_{14}H_{22}NO$] m/z calculated 220.16959, found 220.17000

The diastereomers of variously substituted dimethyl-F_n-quinolines were assigned by analogy from the parent 2,3-dimethylquinoline. The major, anti-diastereomer, was selected from the mixture of diastereomers for NMR assignment.

3r - 6-fluoro-2,3-dimethyl-1,2,3,4-tetrahydroquinoline
Prepared from 6-fluoro-2,3-dimethylquinoline via reduction via the general procedure.
Yield 52%, 1.8:1 *dr*
^{1}H NMR (500 MHz, $CDCl_3$) Major diastereomer δ 6.85 (ddt, *J* = 7.6, 5.1, 1.1 Hz, 1H), 6.28 (tdd, *J* = 8.5, 3.7, 2.5 Hz, 1H), 6.15 (dd, *J* = 10.8, 2.4 Hz, 1H), 3.75 (s, 1H), 3.44 (qd, *J* = 6.6, 3.1 Hz, 1H), 2.82 (dd, *J* = 15.9, 5.2 Hz, 1H), 2.53 – 2.33 (m, 1H), 2.07 – 1.94 (m, 1H), 1.10 (d, *J* = 6.6 Hz, 3H), 0.91 (d, *J* = 6.9 Hz, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 163.23, 145.19 (d, *J* = 10.6 Hz), 130.69 (d, *J* = 9.7 Hz), 115.50 (d, *J* = 2.6 Hz), 103.37 (d, *J* = 21.5 Hz), 100.03 (d, *J* = 24.5 Hz), 49.91, 33.30, 30.53, 18.13, 14.38.

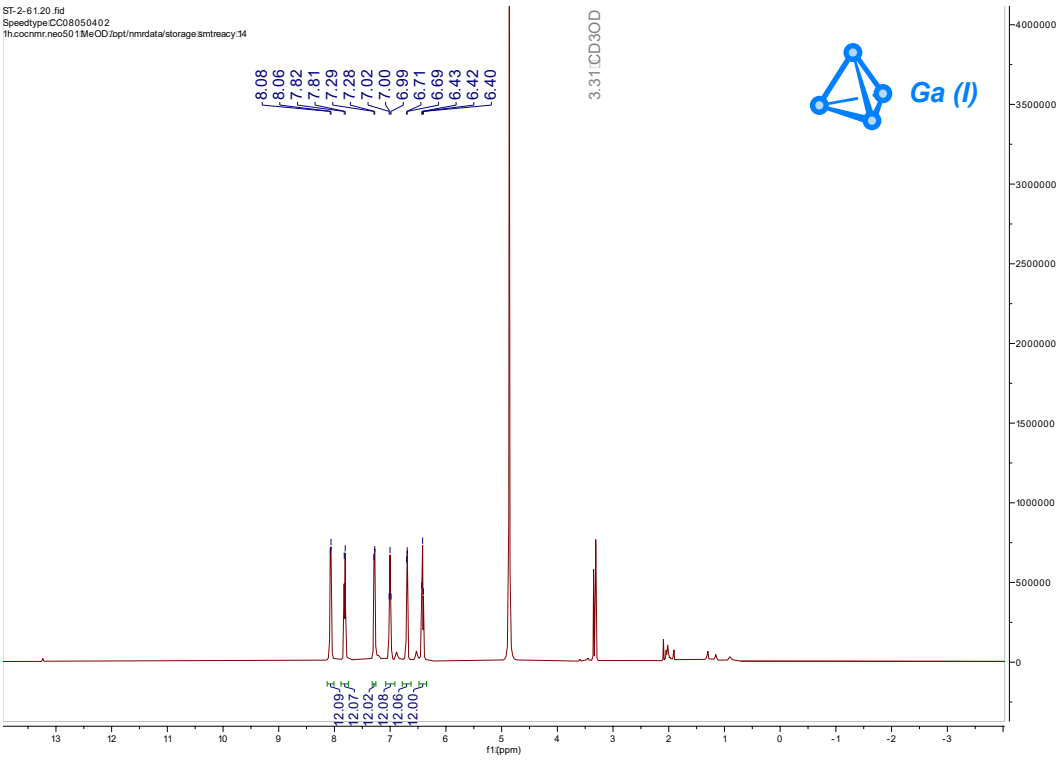
^{19}F NMR (471 MHz, $CDCL_3$) ∆ -117.16 (Q, *J* = 8.5 Hz), -117.33 (Q, *J* = 8.8 Hz).
HRMS (ESI+) [$C_{11}H_{15}FN$] m/z calculated 180.11830, found 180.11832

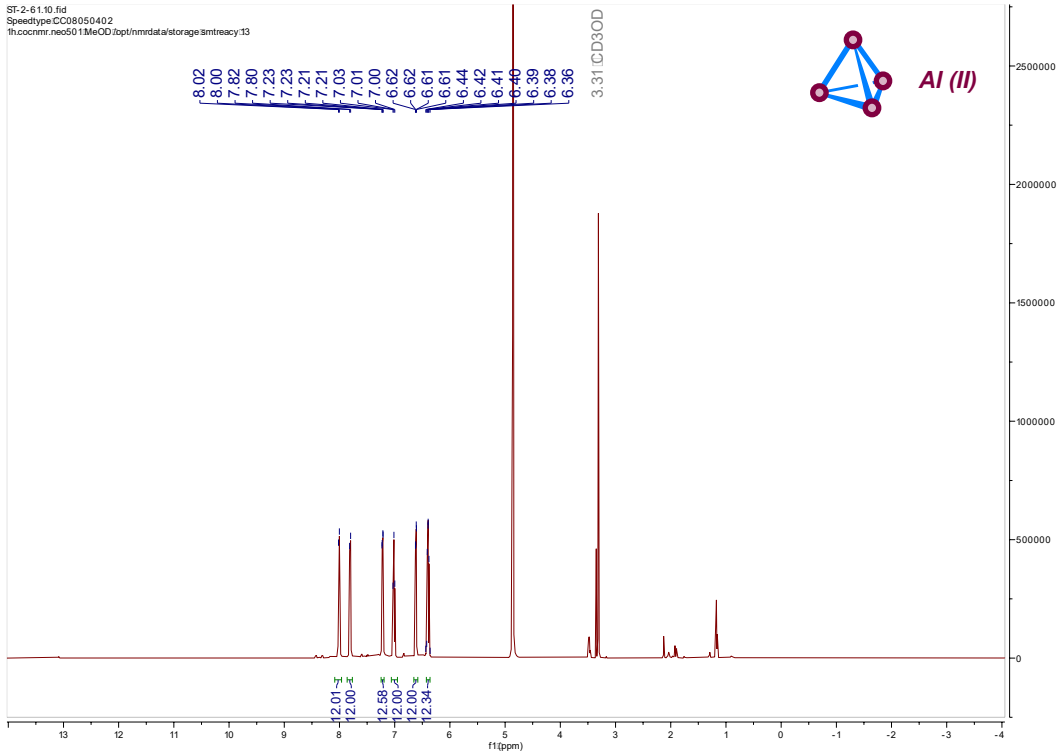
3s - 7-fluoro-2,3-dimethyl-1,2,3,4-tetrahydroquinoline
Prepared from 7-fluoro-2,3-dimethylquinoline via reduction via the general procedure.
Yield 42% 1.2:1 *dr*
^{1}H NMR (500 MHz, $CDCl_3$) Major diastereomer δ 6.71 – 6.64 (m, 2H), 6.43 – 6.37 (m, 1H), 3.35 (tdd, *J* = 6.3, 3.7, 2.7 Hz, 1H), 2.77 – 2.63 (m, 1H), 2.56 – 2.34 (m, 1H), 1.97 – 1.85 (m, 1H), 1.11 (d, *J* = 6.6 Hz, 3H), 0.92 (d, *J* = 6.9 Hz, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 156.59, 141.06, 123.07 – 122.01 (m), 115.53 (d, *J* = 21.6 Hz), 114.87 (d, *J* = 7.5 Hz), 113.28 (dd, *J* = 22.3, 4.3 Hz), 47.46, 33.48, 30.02, 18.14, 14.28.
^{19}F NMR (471 MHz, $CDCl_3$) δ -128.16 – -128.38 (m), -128.52 (dd, *J* = 13.2, 9.3 Hz).
HRMS (ESI+) [$C_{11}H_{15}FN$] m/z calculated 180.11830, found 180.11828

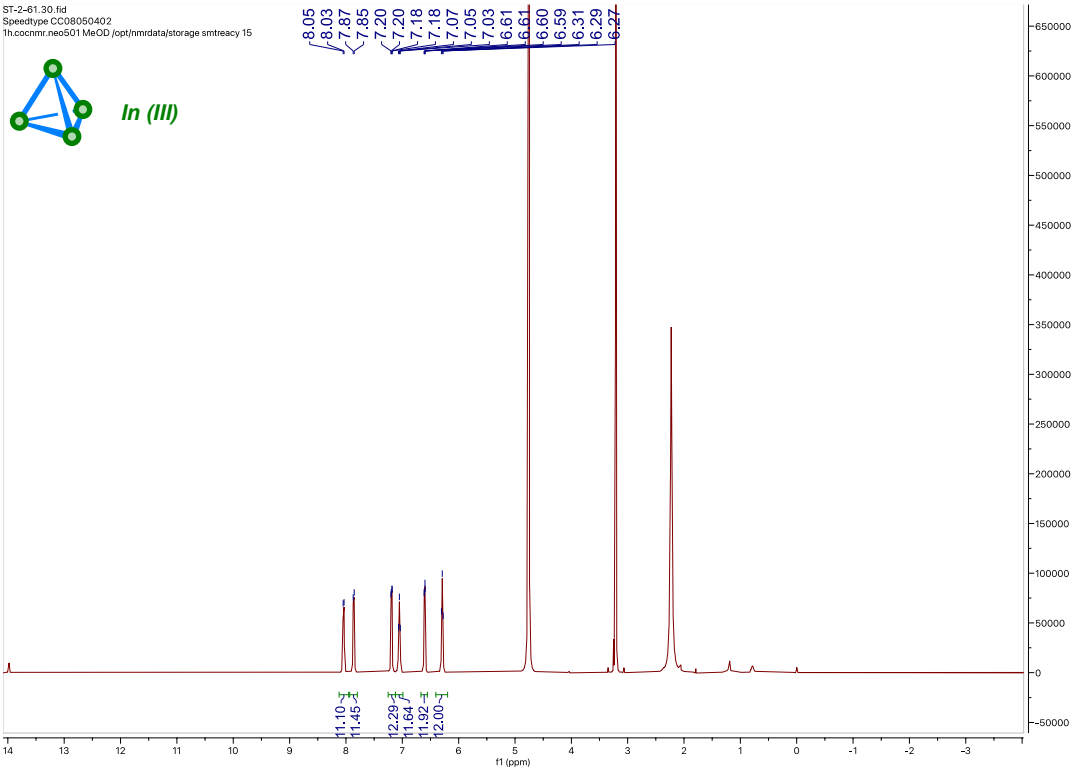
3t - 6,7-difluoro-2,3-dimethyl-1,2,3,4-tetrahydroquinoline
Prepared from 6,7-difluoro-2,3-dimethylquinoline via reduction via the general procedure.
Yield 23% 1.2:1 *dr*
^{1}H NMR (500 MHz, $CDCl_3$) δ 6.79 – 6.65 (m, 1H), 6.23 (dd, *J* = 12.1, 7.0 Hz, 1H), 3.60 (s, 1H), 3.40 (qd, *J* = 6.6, 3.0 Hz, 1H), 2.64 (dd, *J* = 16.2, 4.8 Hz, 1H), 2.44 – 2.34 (m, 1H), 2.03 – 1.93 (m, 1H), 1.09 (d, *J* = 6.5 Hz, 3H), 0.90 (d, *J* = 6.8 Hz, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 150.01, 143.62 (d, *J* = 11.9 Hz), 141.69, 140.30, 117.71 (d, *J* = 17.4 Hz), 101.96 (d, *J* = 20.1 Hz), 49.92, 33.31, 30.30, 18.08, 14.21.
^{19}F NMR (471 MHz, $CDCl_3$) δ -141.50 – -142.19 (m), -153.66 – -153.88 (m), -153.98 (ddd, *J* = 22.3, 10.9, 6.8 Hz).
HRMS (ESI+) [$C_{11}H_{14}F_2N$] m/z calculated 198.10880, found 198.10894

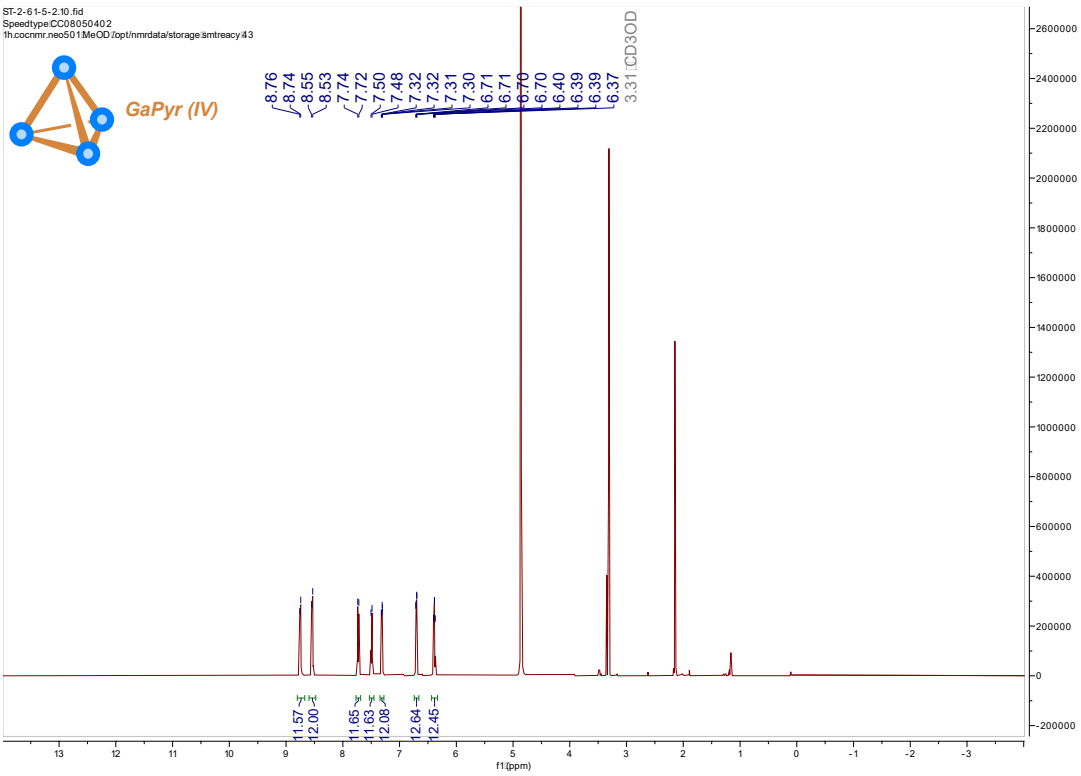
VI.27 – NMR Spectra

- ***Host NMR Spectra***

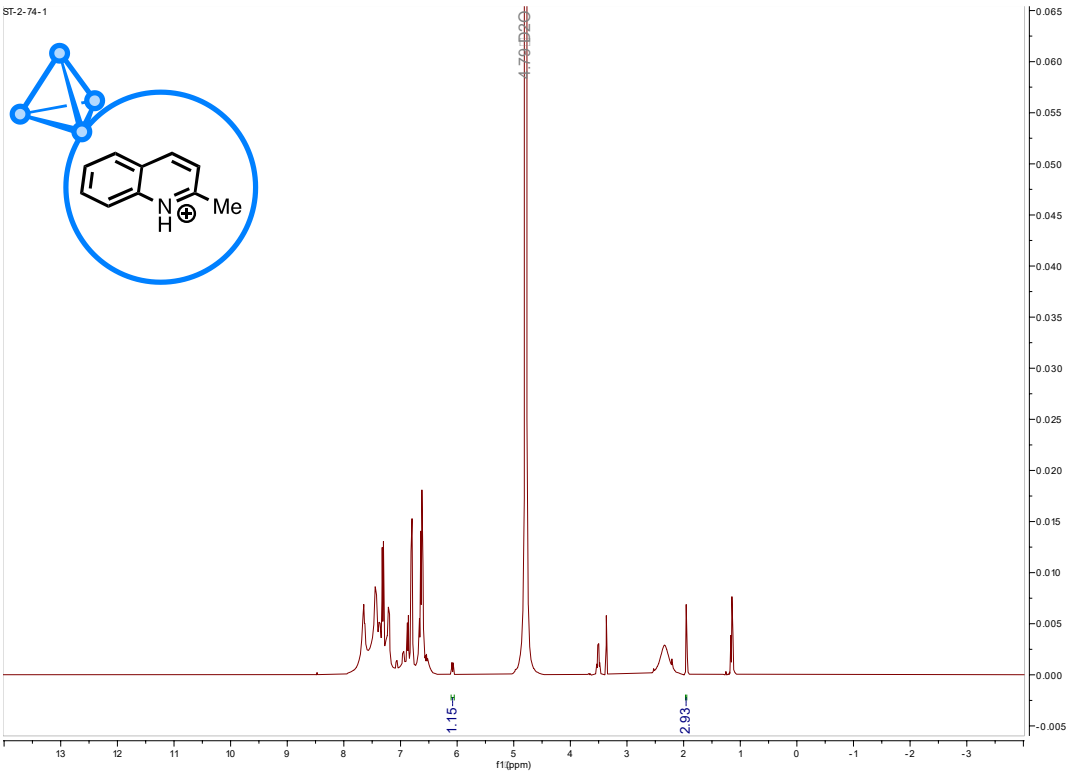


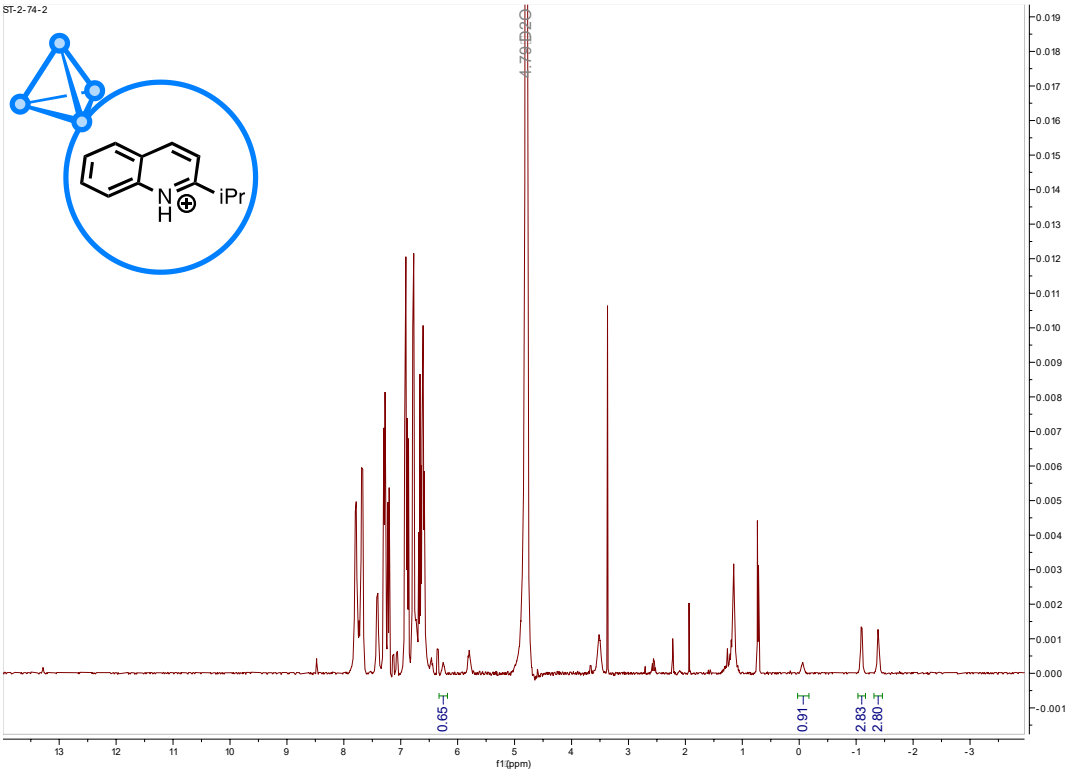


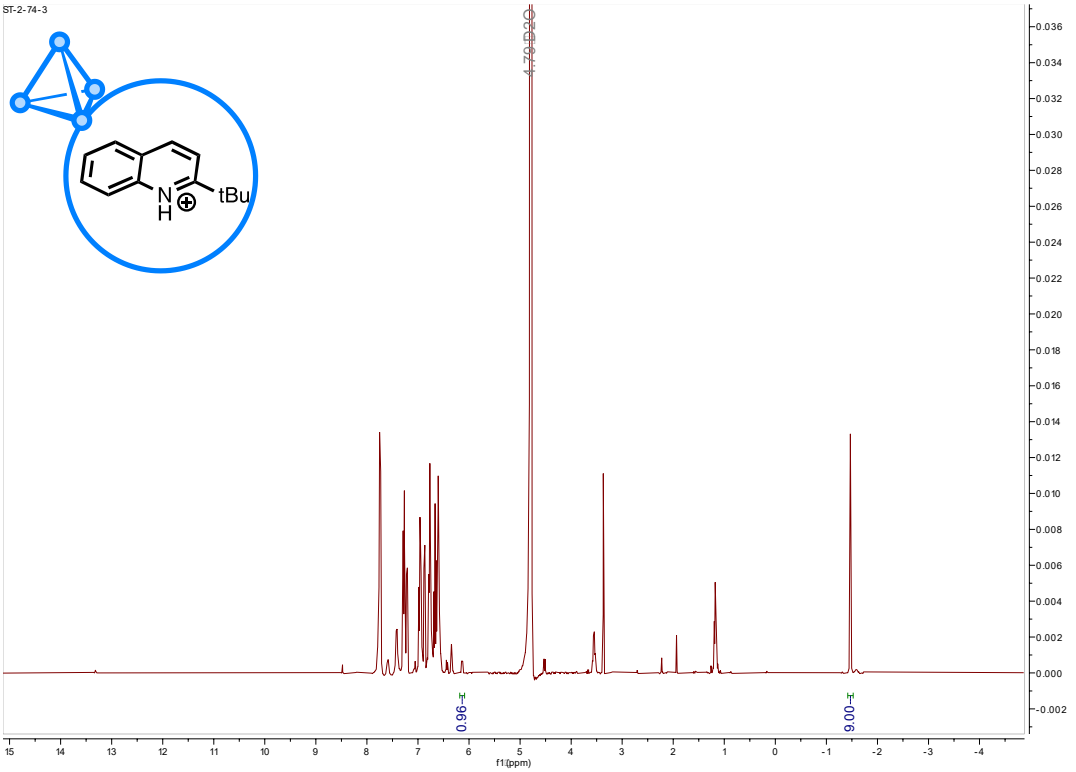

ST-2-61.30.fid
Speedtype CC08050402
1h.cocnmr.neo501 MeOD /opt/nmrdata/storage smtreacy 15
In (III)
8.05
8.03
7.87
7.85
7.20
7.18
7.07
7.05
7.03
6.61
6.60
6.59
6.31
6.29
6.27
11.10
11.45
12.29
11.64
11.92
12.00
f1 (ppm)

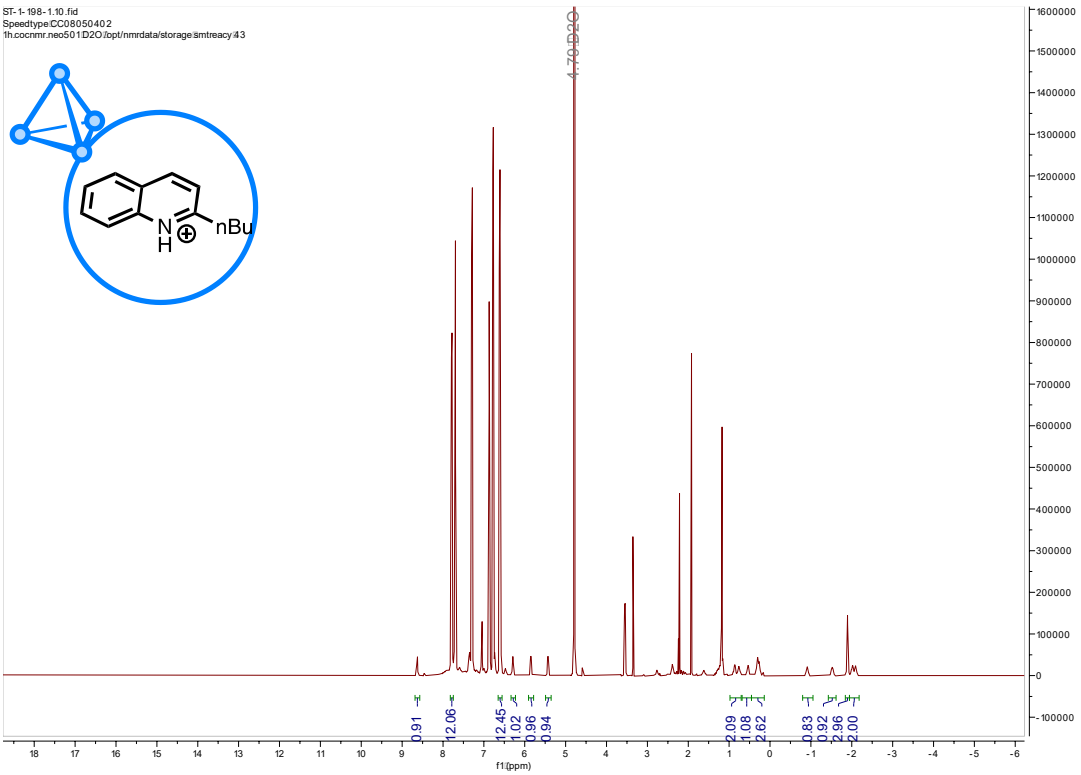

ST-2-61-5-2.10.fid
Speedtype CC08050402
1h.cocnmr.neo501 MeOD /opt/nmrdata/storage smtreacy 43
GaPyr (IV)
8.76
8.74
8.55
8.53
7.74
7.72
7.50
7.48
7.32
7.31
7.30
6.71
6.71
6.70
6.70
6.40
6.39
6.39
6.37
3.31 CD3OD
11.57
12.00
11.65
11.63
12.08
12.64
12.45
f1 (ppm)

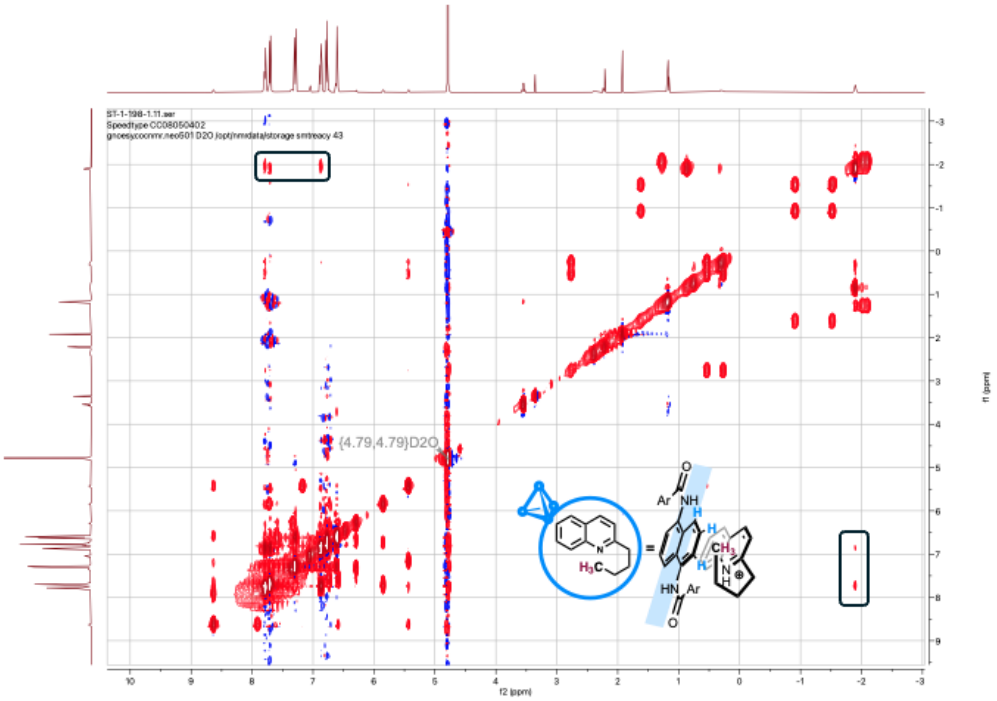
- ***Encapsulation Studies NMR Spectra***

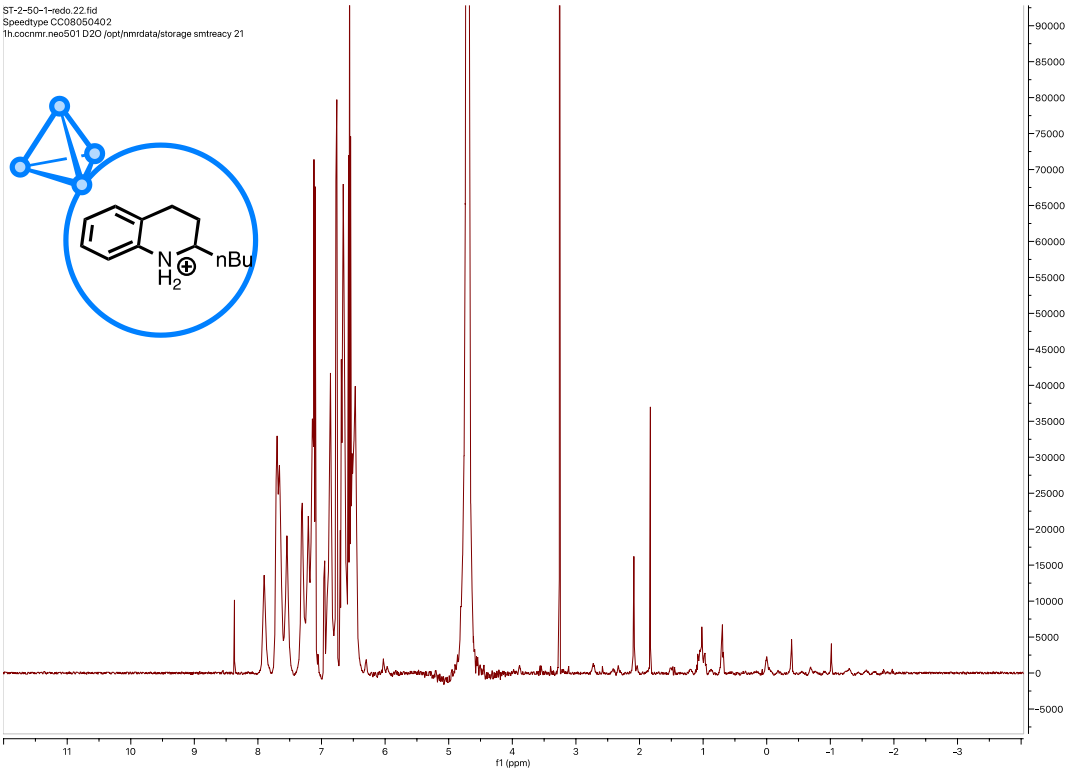


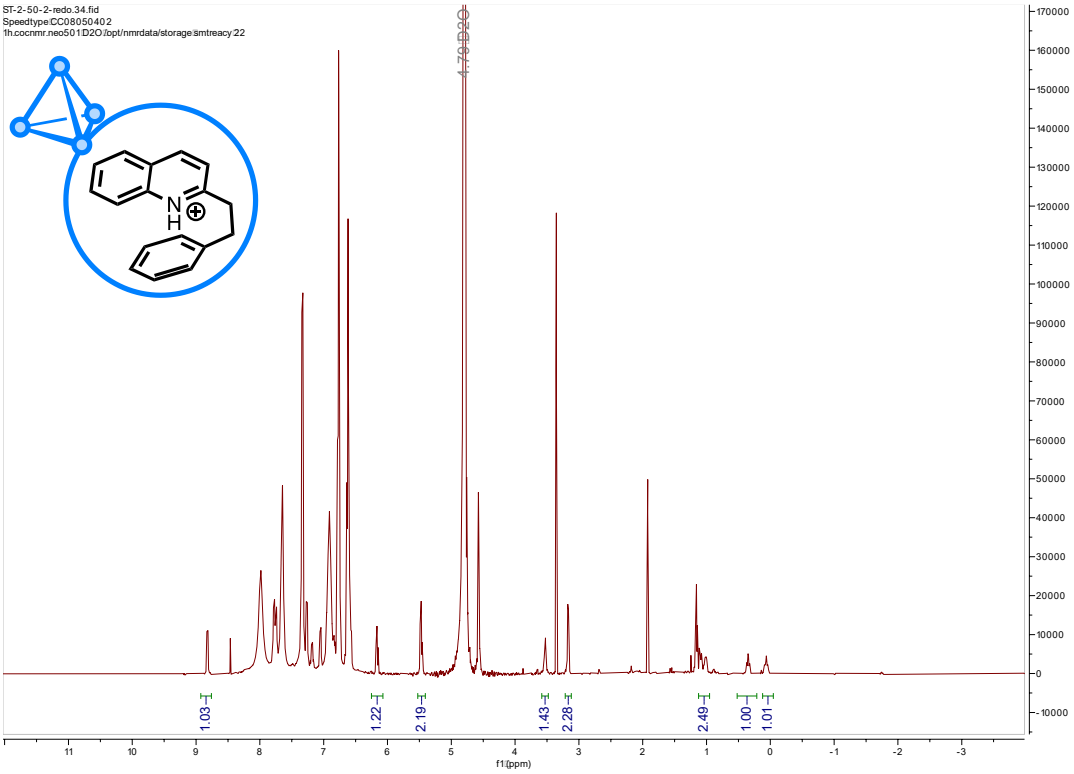



ST-2-74-3
4.70 D2O
tBu
0.96
9.00
f1 (ppm)

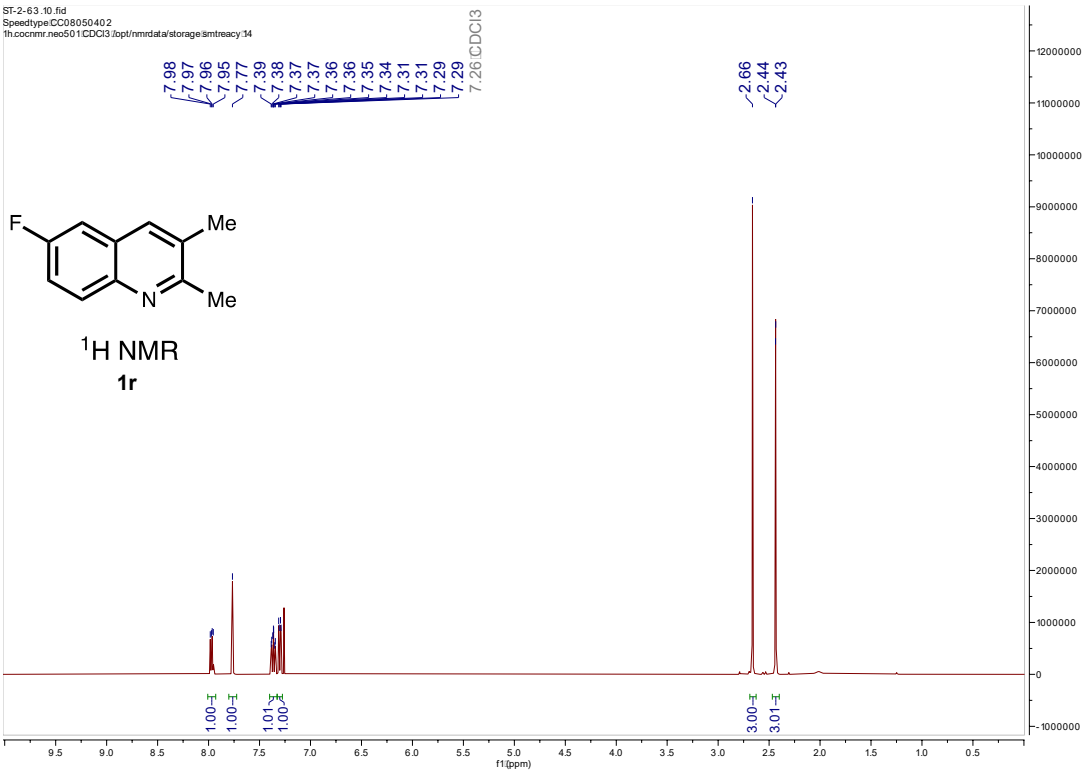

ST-1-198-1.10.fid
Speedtype CC08050402
1h.cocnmr.neo501 D2O /opt/nmrdata/storage/smtreacy/43
4.70 D2O
nBu
0.91
12.06
12.45
1.02
0.96
0.94
2.09
1.08
2.62
0.83
0.92
2.96
2.00
f1 (ppm)

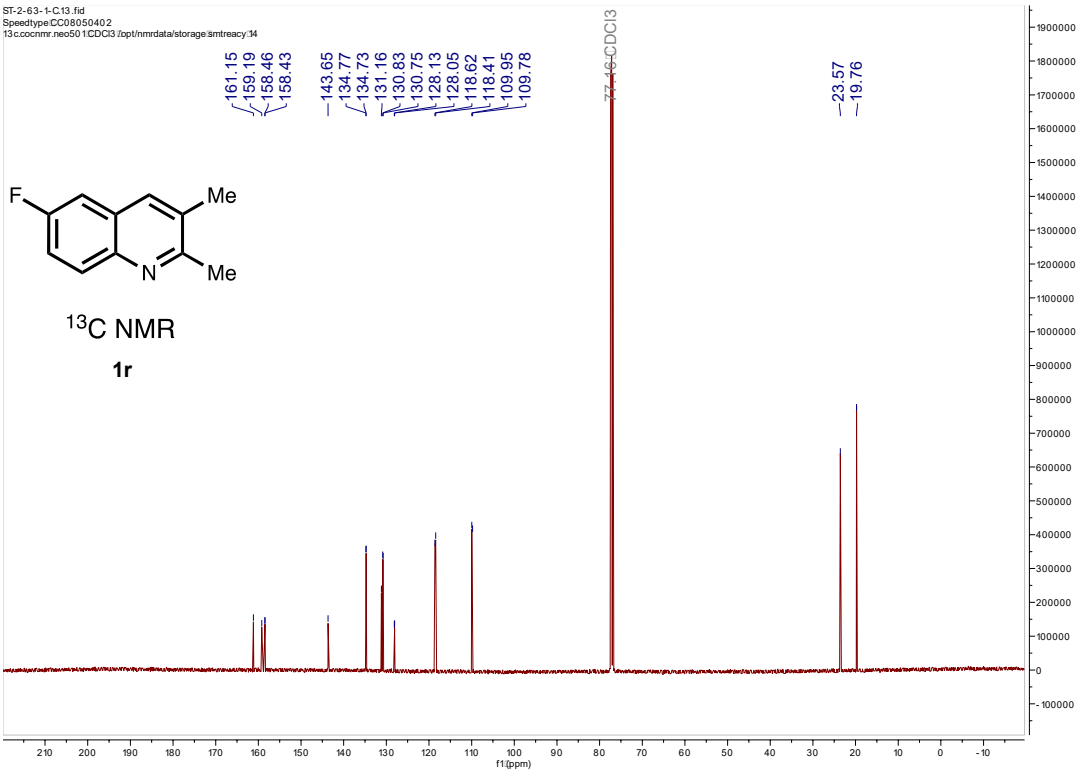

{4.79,4.79}D2O
f2 (ppm)
f1 (ppm)

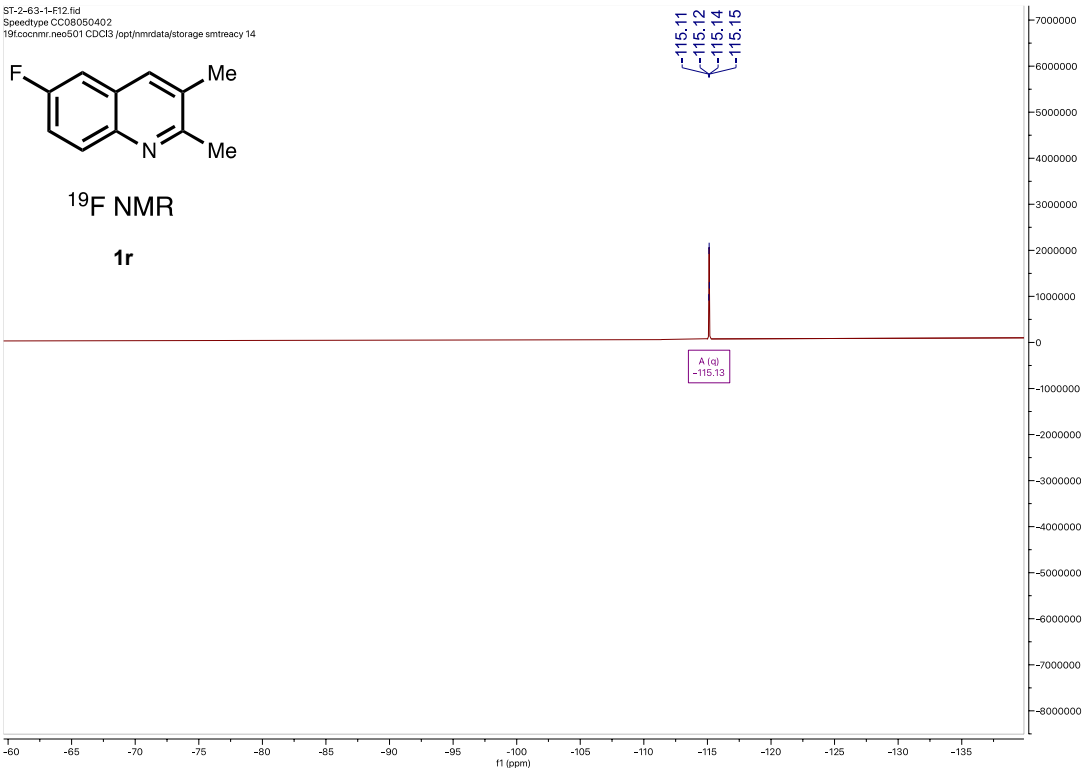

ST-2-50-1-redo.22.fid
Speedtype CC08050402
1h.cocnmr.neo501 D2O /opt/nmrdata/storage smtreacy 21
nBu
f1 (ppm)

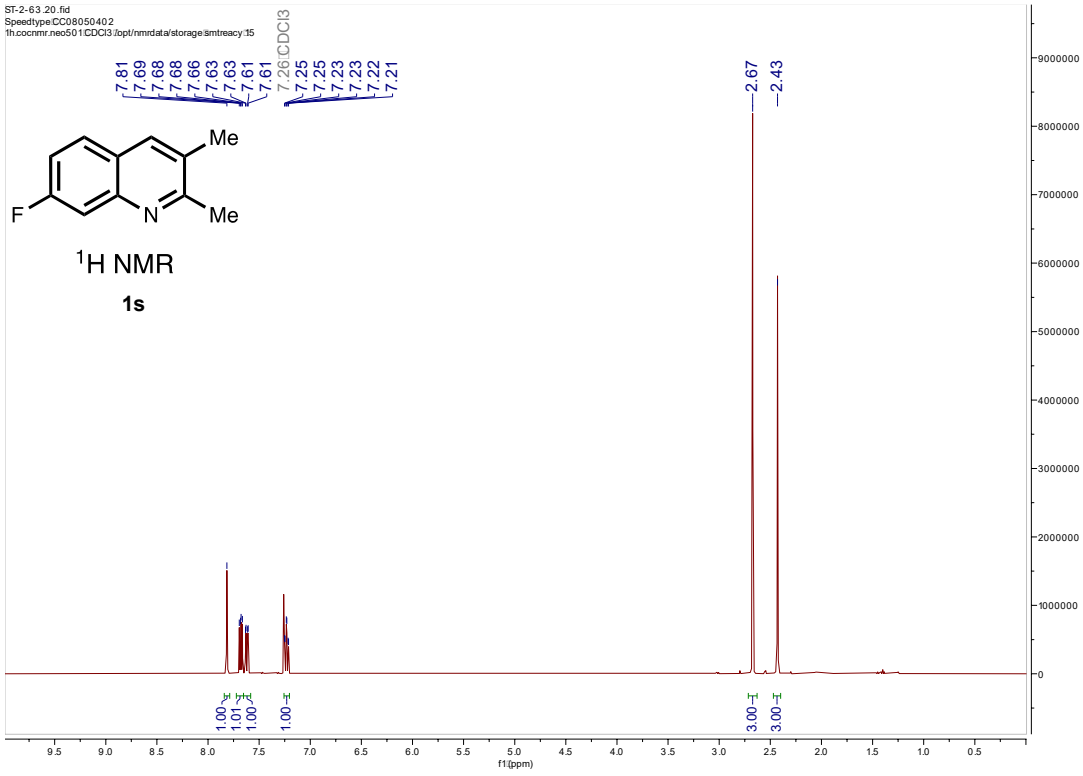

ST-2-50-2-redo.34.fid
Speedtype:CC08050402
1h.cocnmr.neo501D2O:/opt/nmrdata/storage/smtreacy:22
4.79:D2O
1.03
1.22
2.19
1.43
2.28
2.49
1.00
1.01
f1 (ppm)

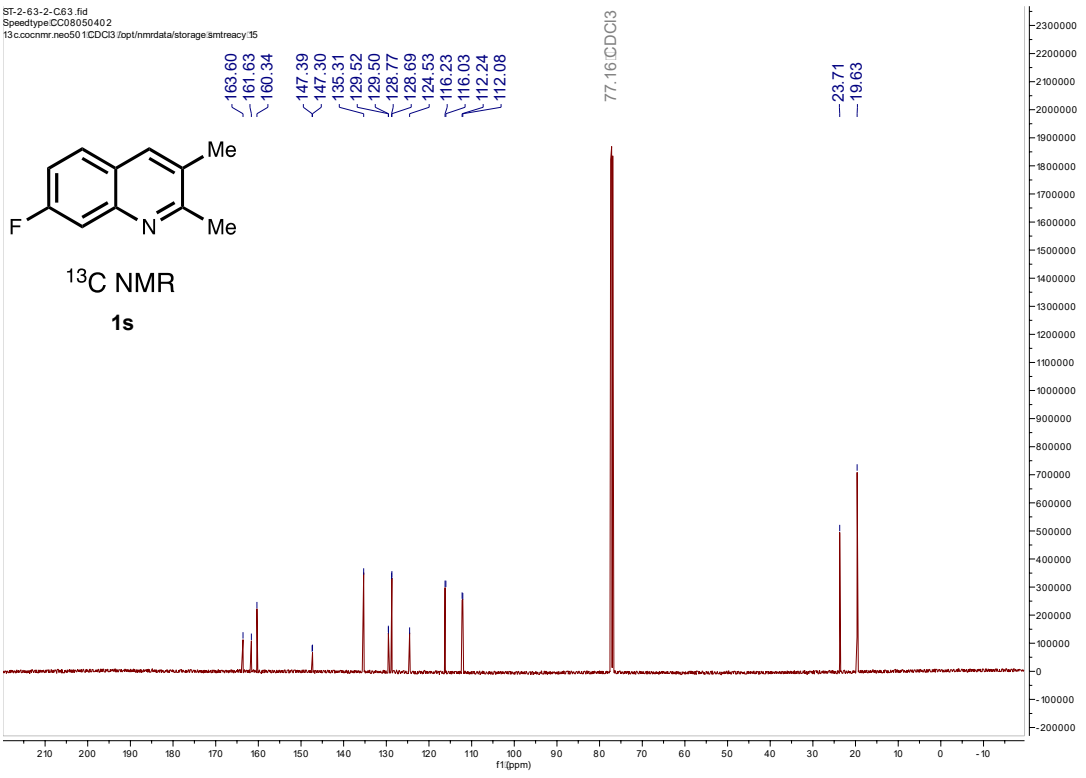
- ***Novel Products NMR Spectra***

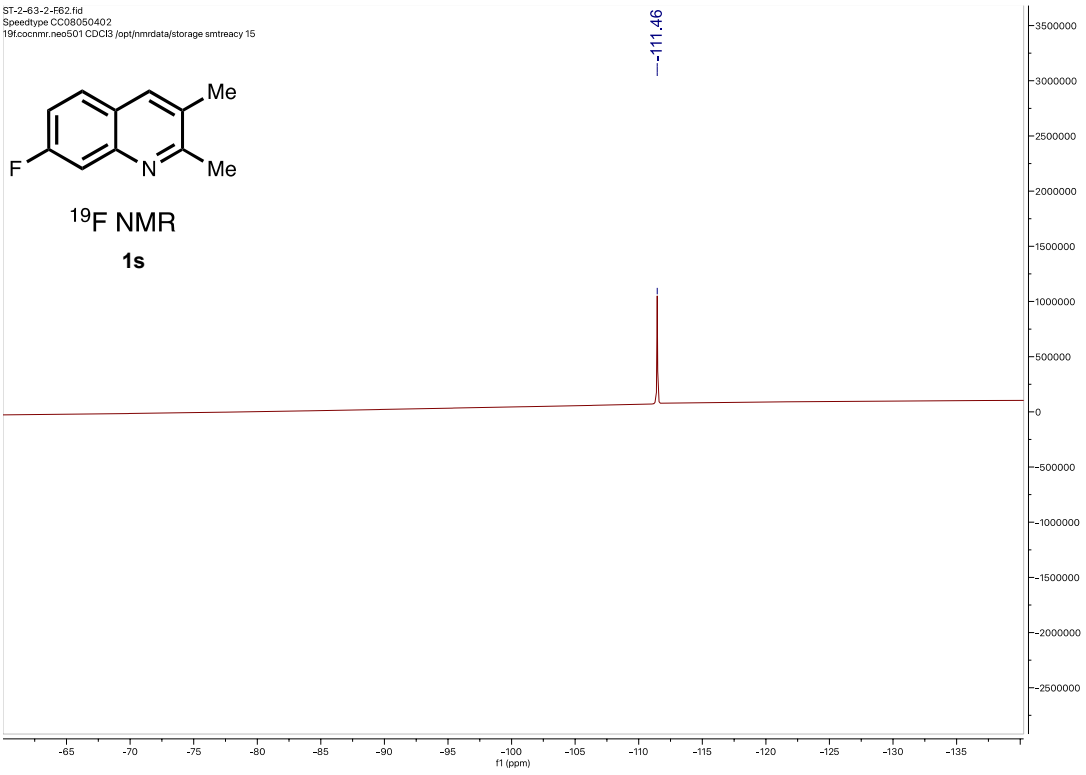


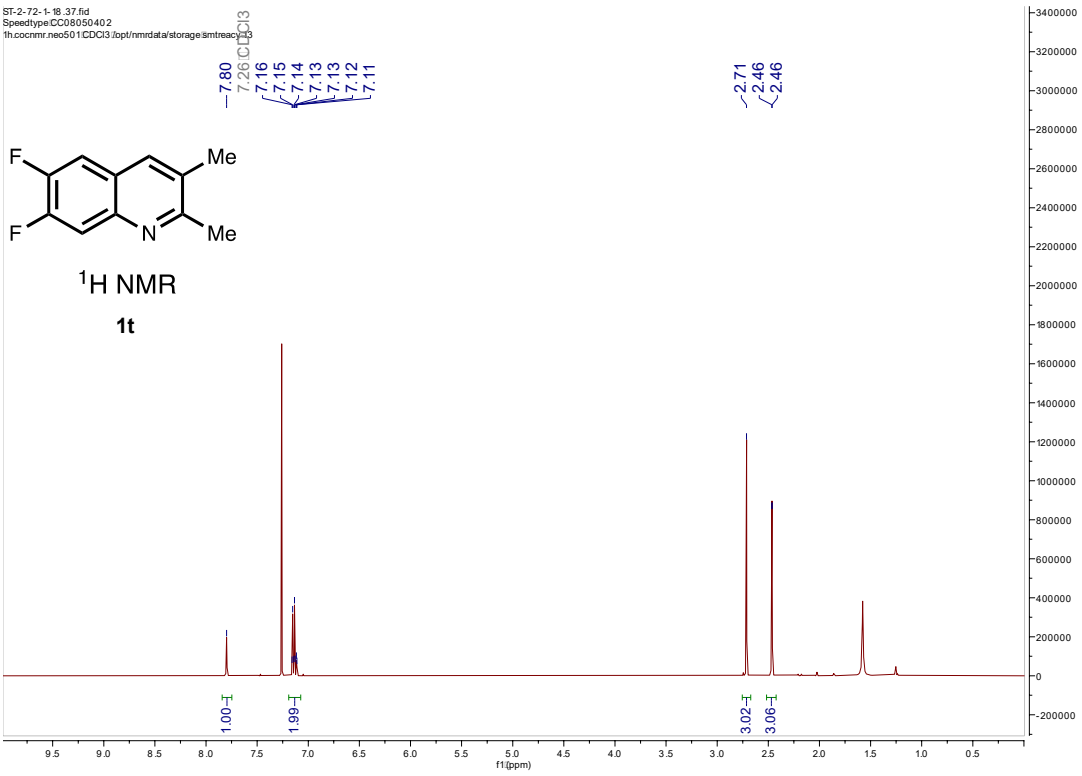


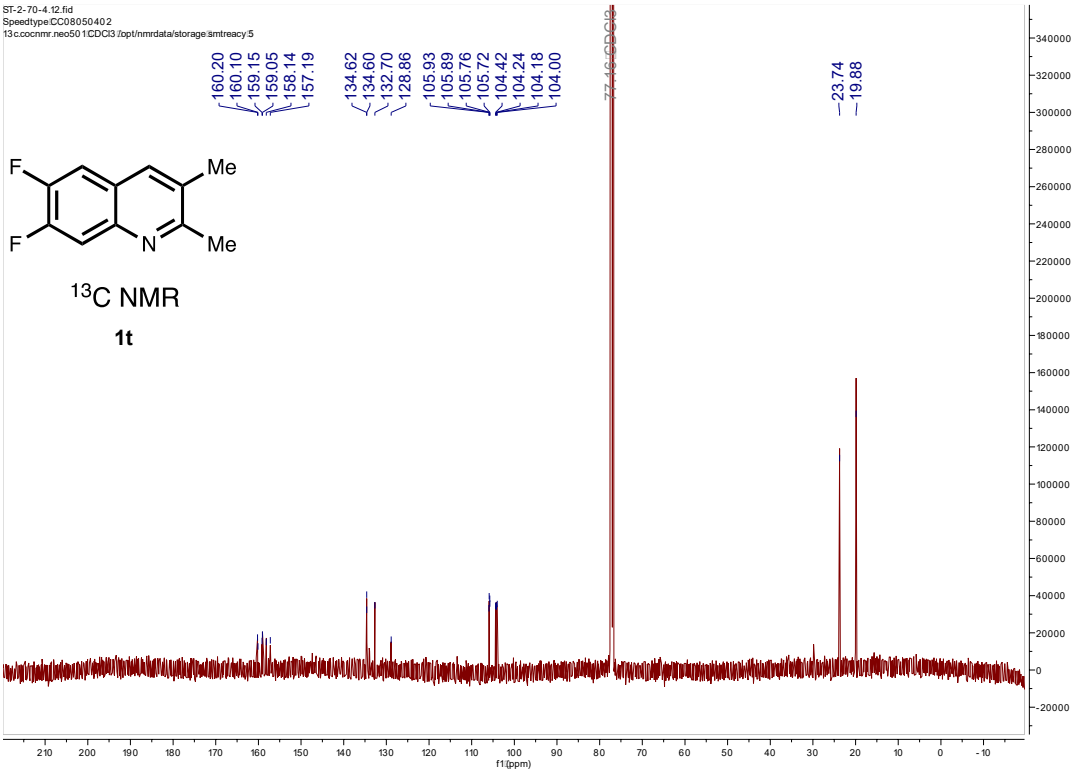

ST-2-63-1-F12.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 14
F
Me
N
Me
19F NMR
1r
-115.11
-115.12
-115.14
-115.15
A (q)
-115.13
f1 (ppm)

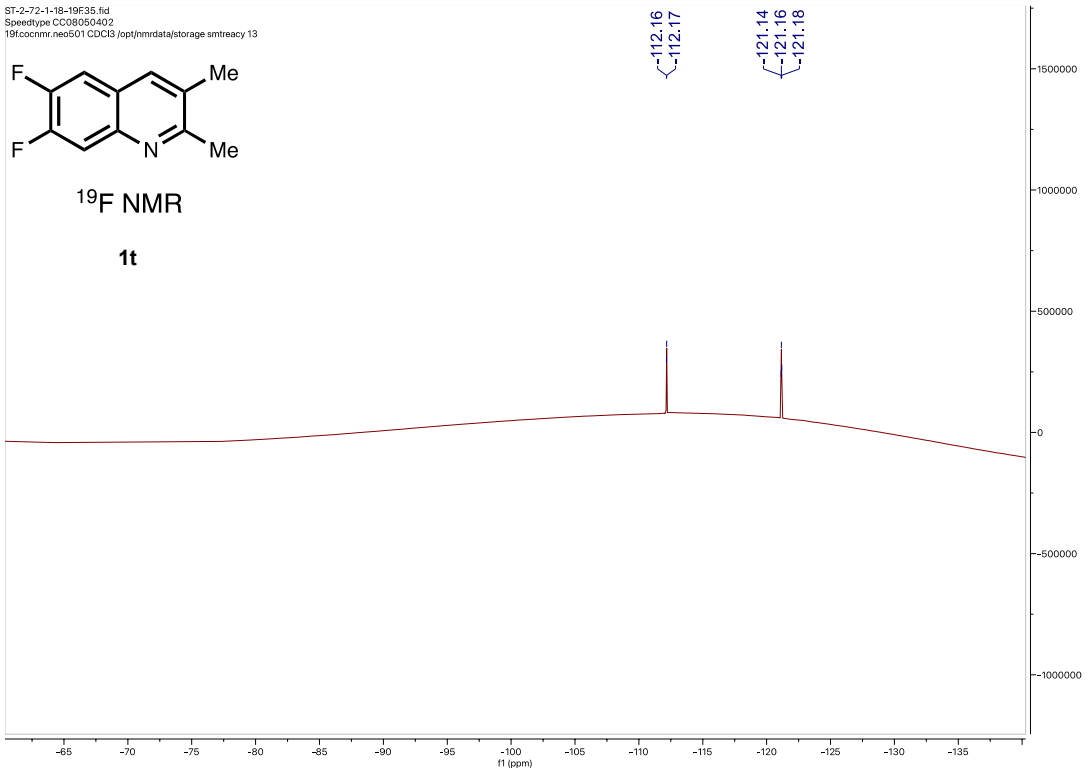

ST-2-63.20.fid
Speedtype CC08050402
1h.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 15
7.81
7.69
7.68
7.68
7.66
7.63
7.63
7.61
7.61
7.26 CDCl3
7.25
7.25
7.23
7.23
7.22
7.21
2.67
2.43
Me
F
N
Me
1H NMR
1s
1.00
1.01
1.00
1.00
3.00
3.00
f1 (ppm)

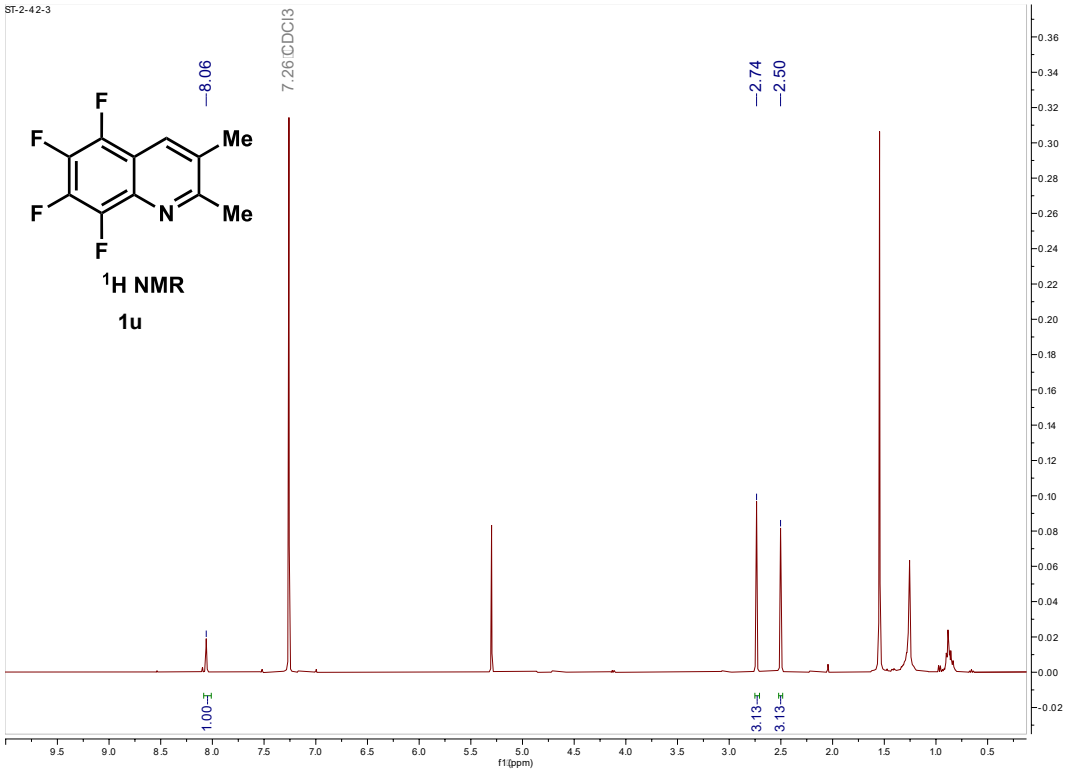

ST-2-63-2-C63.fid
Speedtype CC08050402
13c.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 15
163.60
161.63
160.34
147.39
147.30
135.31
129.52
129.50
128.77
128.69
124.53
116.23
116.03
112.24
112.08
77.16 CDCl3
23.71
19.63
Me
F
N
Me
13C NMR
1s
f1 (ppm)

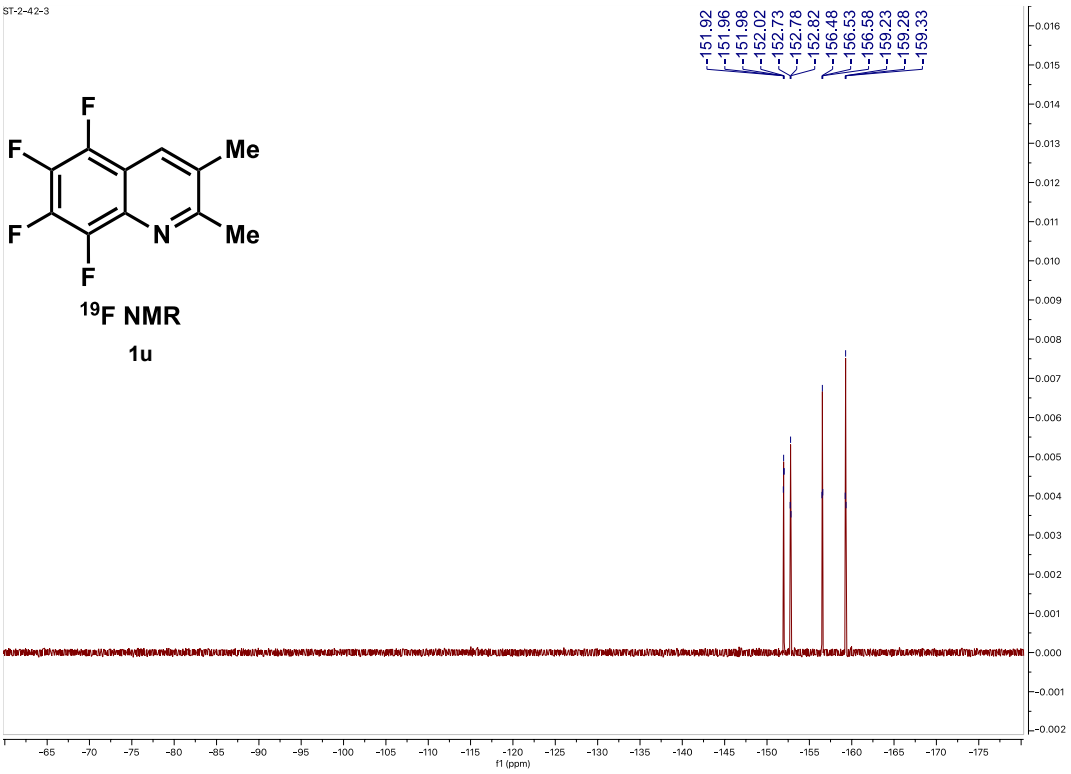

ST-2-63-2-F62.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 15
-111.46
Me
F
N
Me
19F NMR
1s
f1 (ppm)

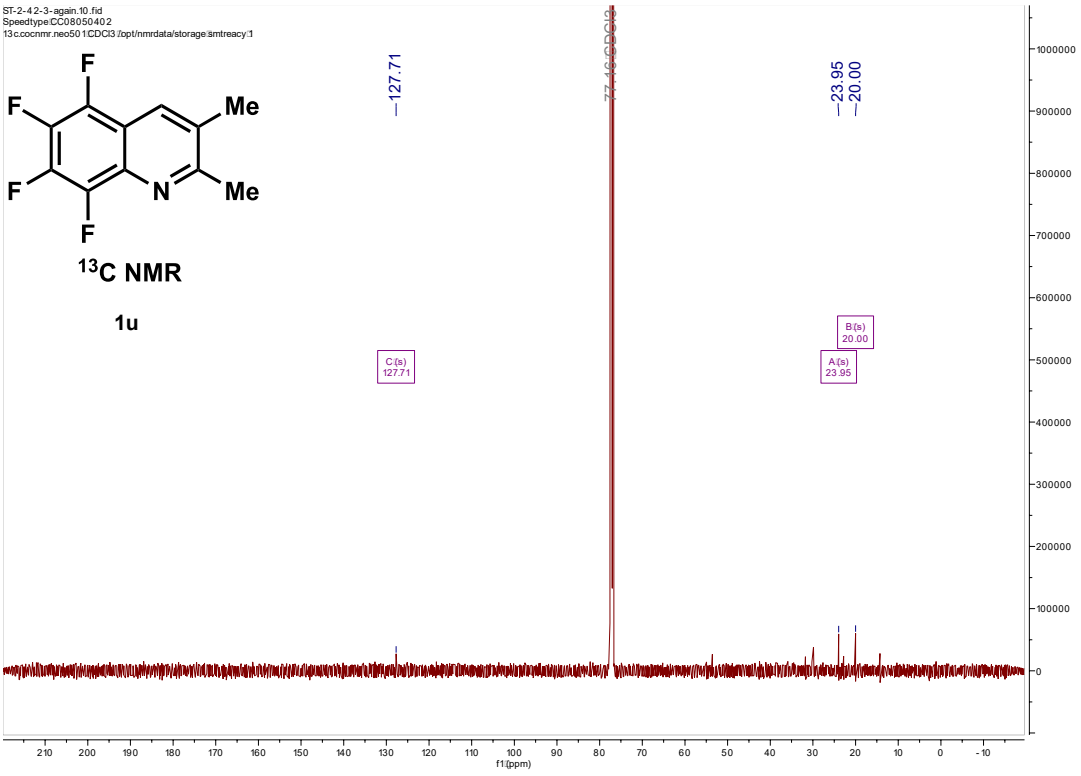

ST-2-72-1-18_37.fid
Speedtype:CC08050402
1h.cocnmr.neo501:CDCl3:/opt/nmrdata/storage:smtreacy:13
7.80
7.26 CDCl3
7.16
7.15
7.14
7.13
7.13
7.12
7.11
2.71
2.46
2.46
F
Me
F
N
Me
1H NMR
1t
1.00
1.99
3.02
3.06
f1 (ppm)

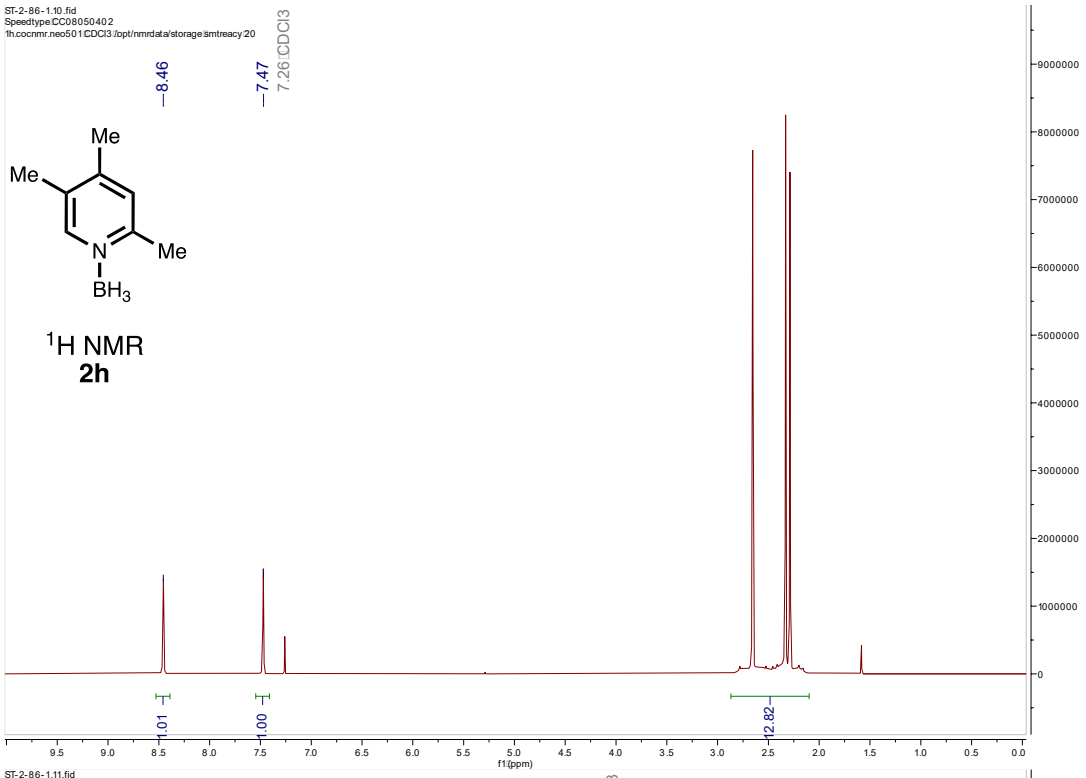

ST-2-70-4_12.fid
Speedtype:CC08050402
13c.cocnmr.neo501:CDCl3:/opt/nmrdata/storage:smtreacy:5
160.20
160.10
159.15
159.05
158.14
157.19
134.62
134.60
132.70
128.86
105.93
105.89
105.76
105.72
104.42
104.24
104.18
104.00
77.16 CDCl3
23.74
19.88
F
Me
F
N
Me
13C NMR
1t
f1 (ppm)

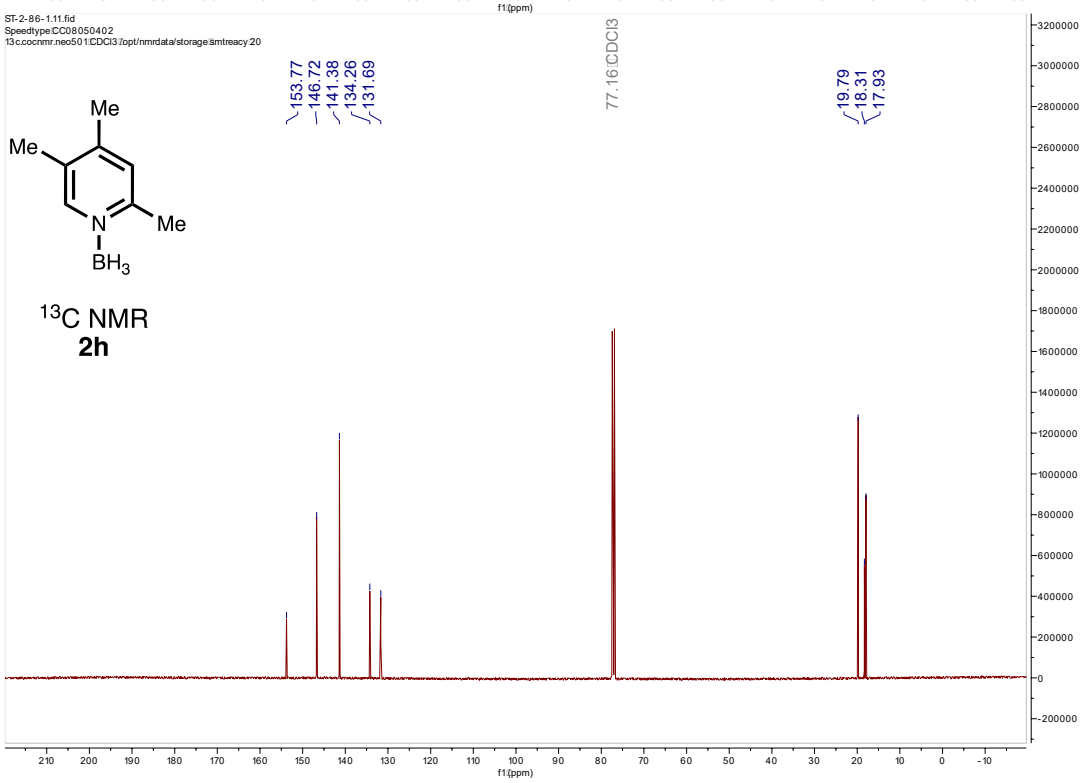

ST-2-72-1-18-19F35.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 13
F
F
N
Me
Me
19F NMR
1t
-112.16
-112.17
-121.14
-121.16
-121.18
1500000
1000000
500000
0
-500000
-1000000
-65
-70
-75
-80
-85
-90
-95
-100
-105
-110
-115
-120
-125
-130
-135
f1 (ppm)

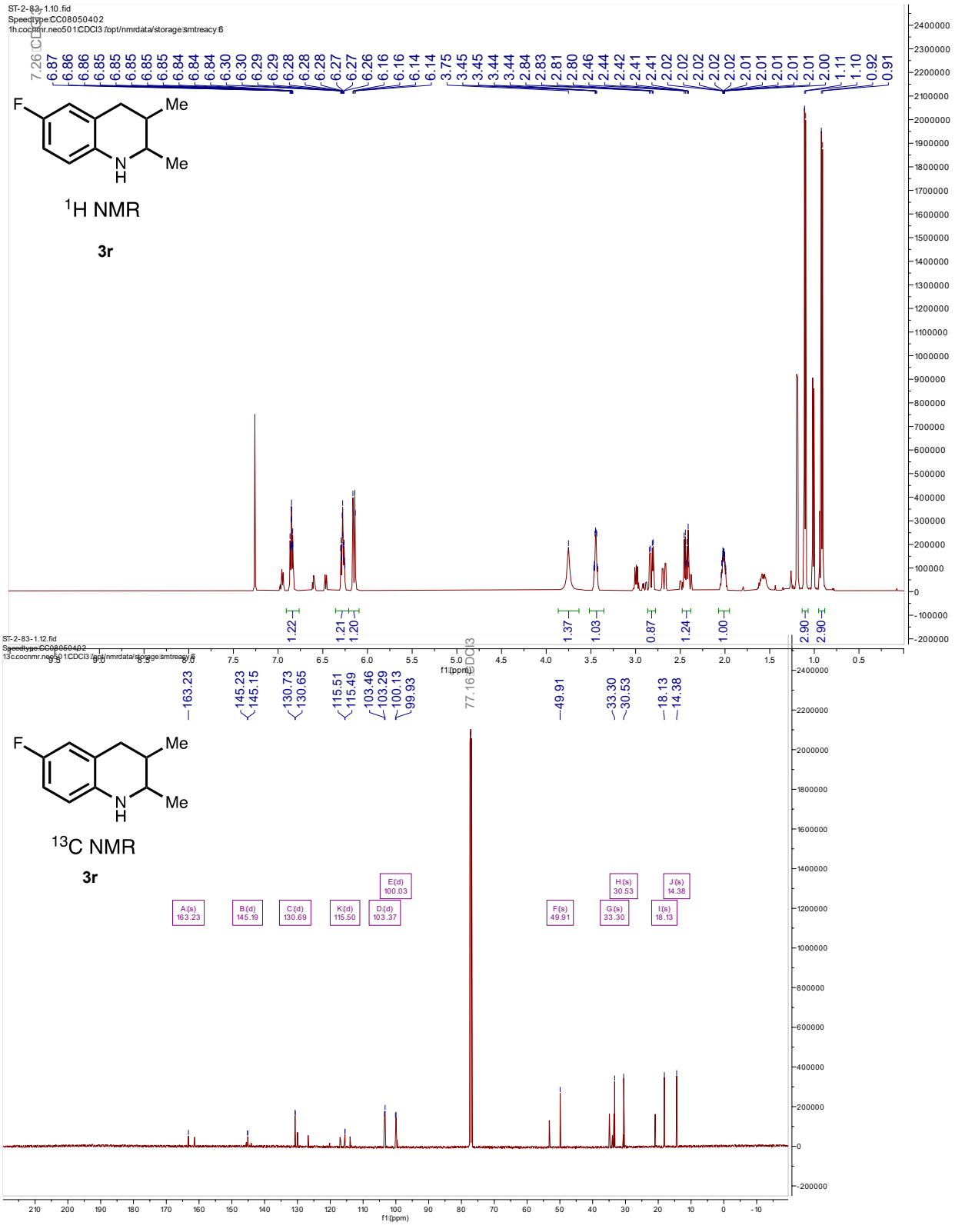

ST-2-42-3
F
F
F
F
N
Me
Me
1H NMR
1u
8.06
7.26 CDCl3
2.74
2.50
1.00
3.13
3.13
0.36
0.34
0.32
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

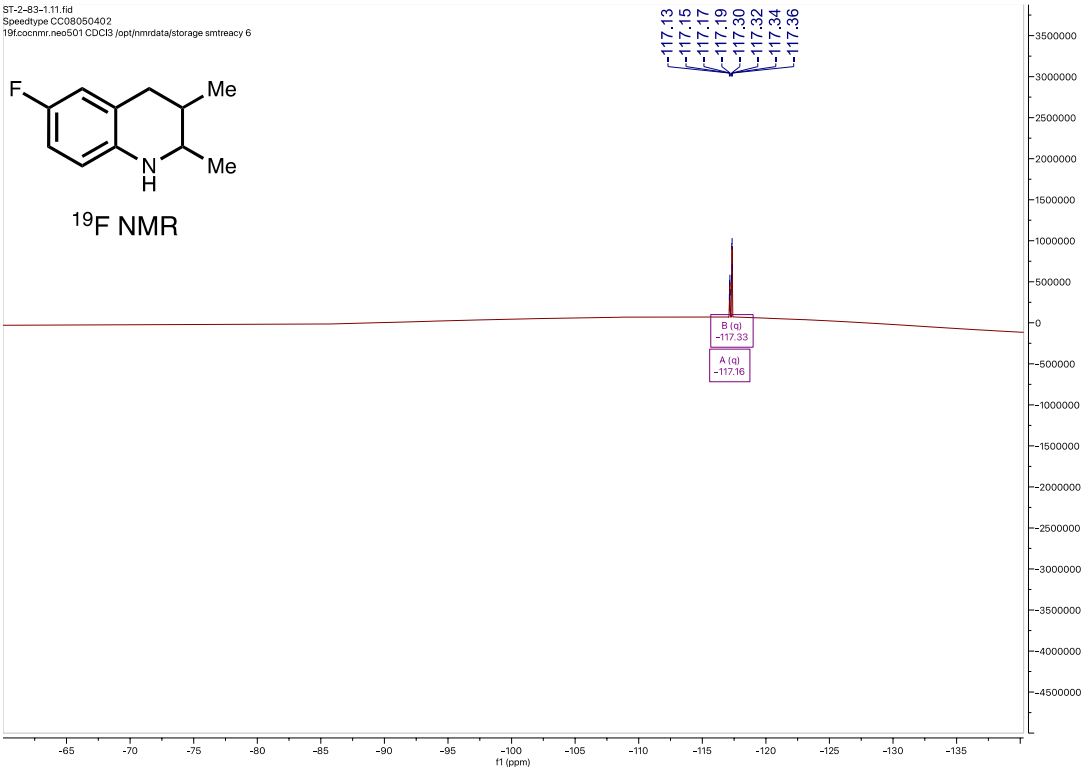

ST-2-42-3
F
F
Me
F
N
Me
F
19F NMR
1u
-151.92
-151.96
-151.98
-152.02
-152.73
-152.78
-152.82
-156.48
-156.53
-156.58
-159.23
-159.28
-159.33
f1 (ppm)

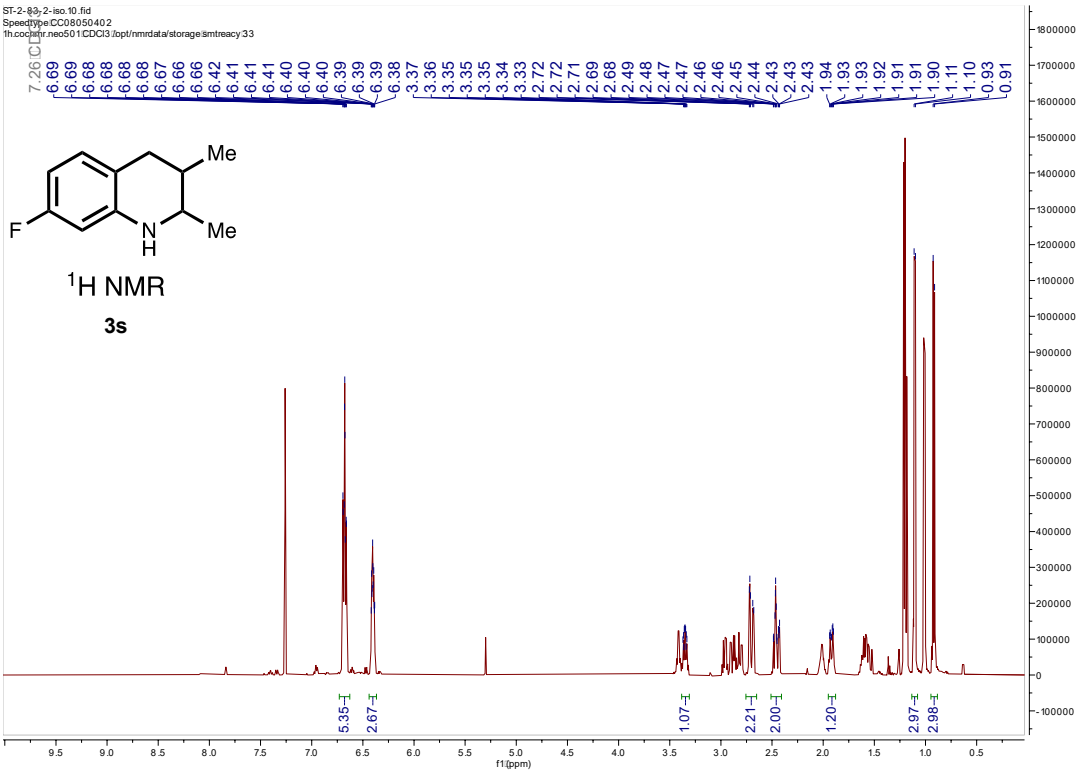

ST-2-42-3-again.10.fid
F
F
Me
F
N
Me
F
13C NMR
1u
127.71
77.16 CDCl3
23.95
20.00
C (s)
127.71
B (s)
20.00
A (s)
23.95
f1 (ppm)

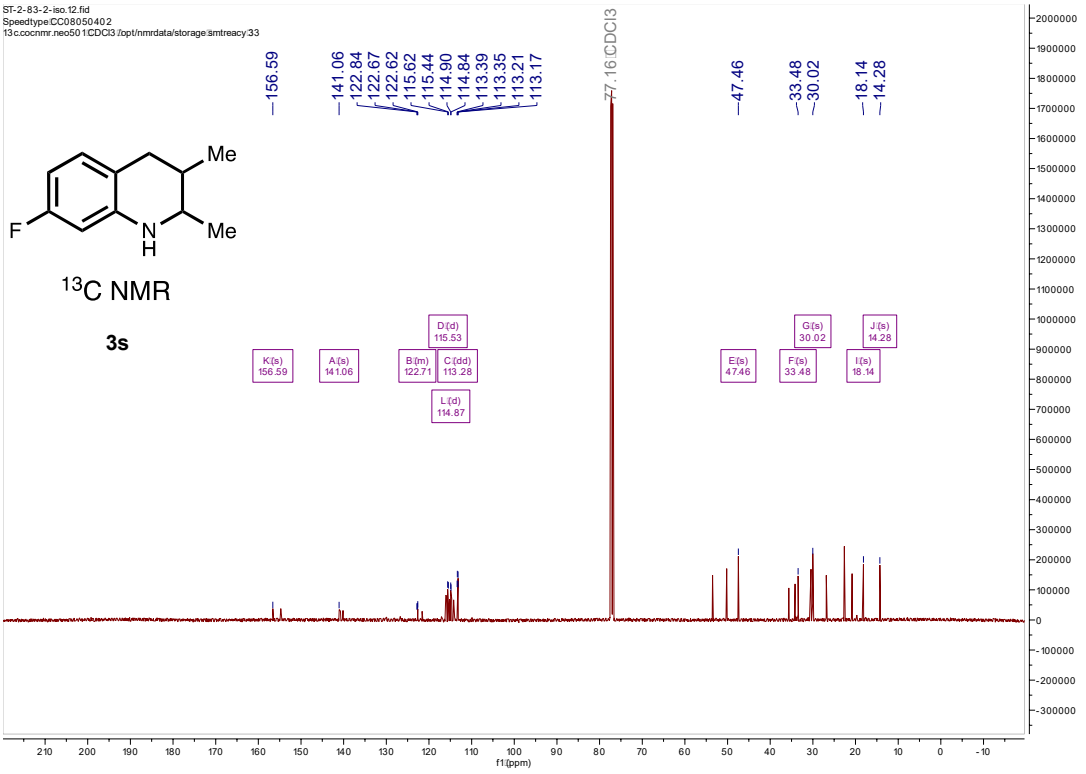

ST-2-86-1.10.fid
Speedtype:CC08050402
1h.cocnmr.neo501:CDCl3:/opt/nmrdata/storage:smtreacy:20
8.46
7.47
7.26:CDCl3
Me
Me
Me
N
BH3
1H NMR
2h
1.01
1.00
12.82
f1(ppm)

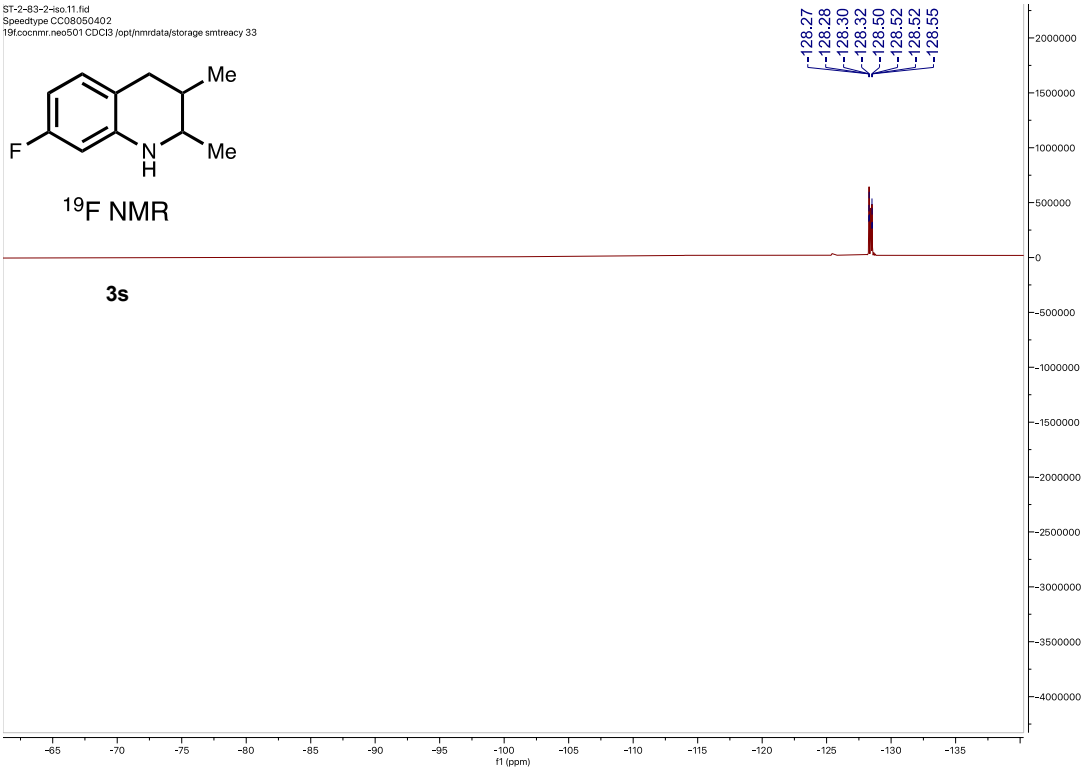

ST-2-86-1.11.fid
Speedtype:CC08050402
13c.cocnmr.neo501:CDCl3:/opt/nmrdata/storage:smtreacy:20
153.77
146.72
141.38
134.26
131.69
77.16:CDCl3
19.79
18.31
17.93
Me
Me
Me
N
BH3
13C NMR
2h
f1(ppm)

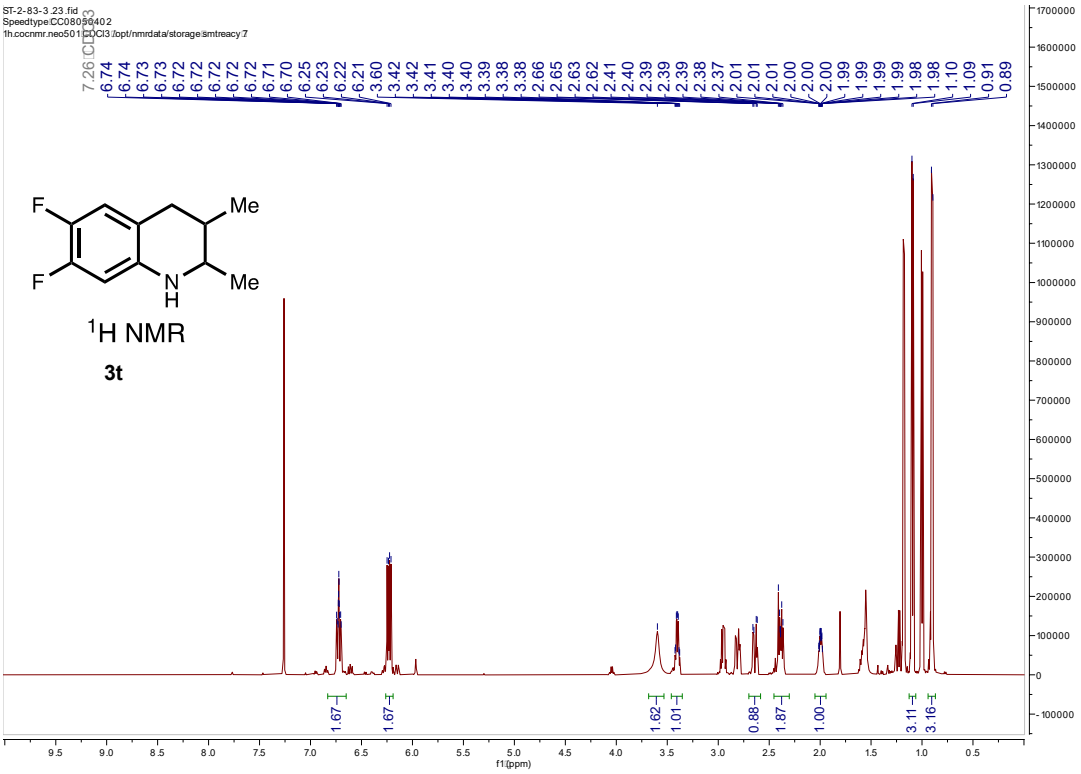

F
Me
N
H
Me
1H NMR
3r
13C NMR
3r

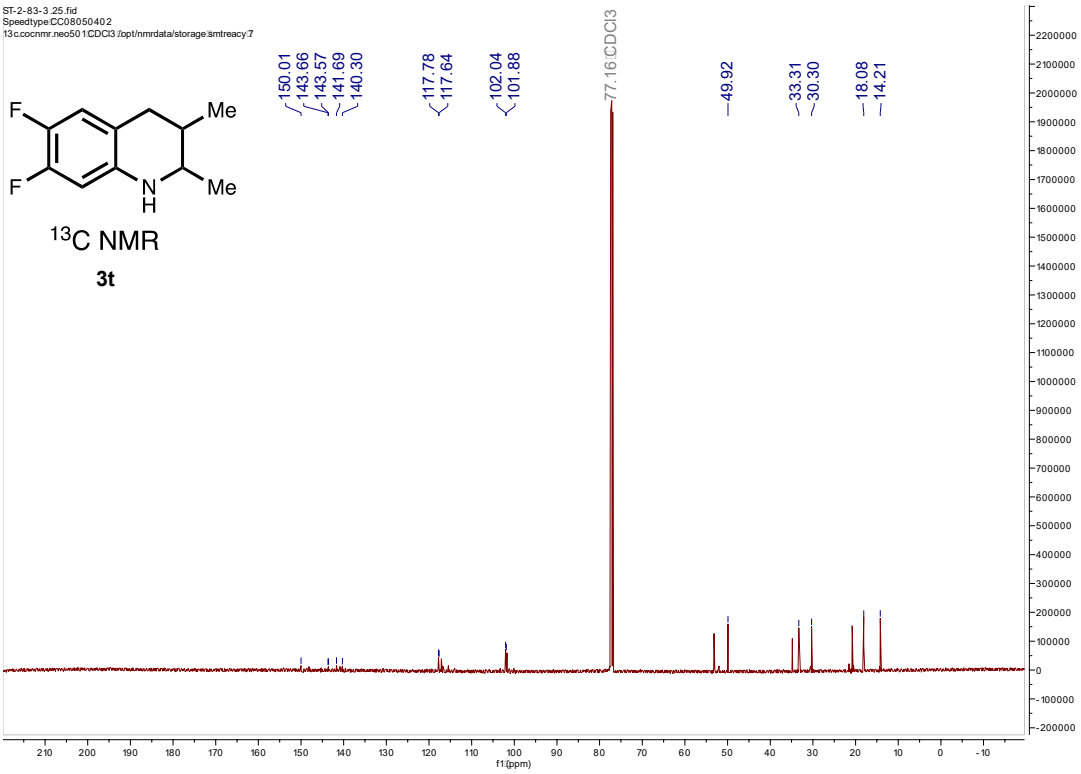

ST-2-83-1.11.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 6
F
Me
N
H
Me
19F NMR
117.13
117.15
117.17
117.19
117.30
117.32
117.34
117.36
B (q)
-117.33
A (q)
-117.16
f1 (ppm)


ST-2-83-2-iso.10.fid
Speedtype CC08050402
1h.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 33
7.26 CDCl3
F
Me
N
H
Me
1H NMR
3s
5.35
2.67
1.07
2.21
2.00
1.20
2.97
2.98
f1 (ppm)

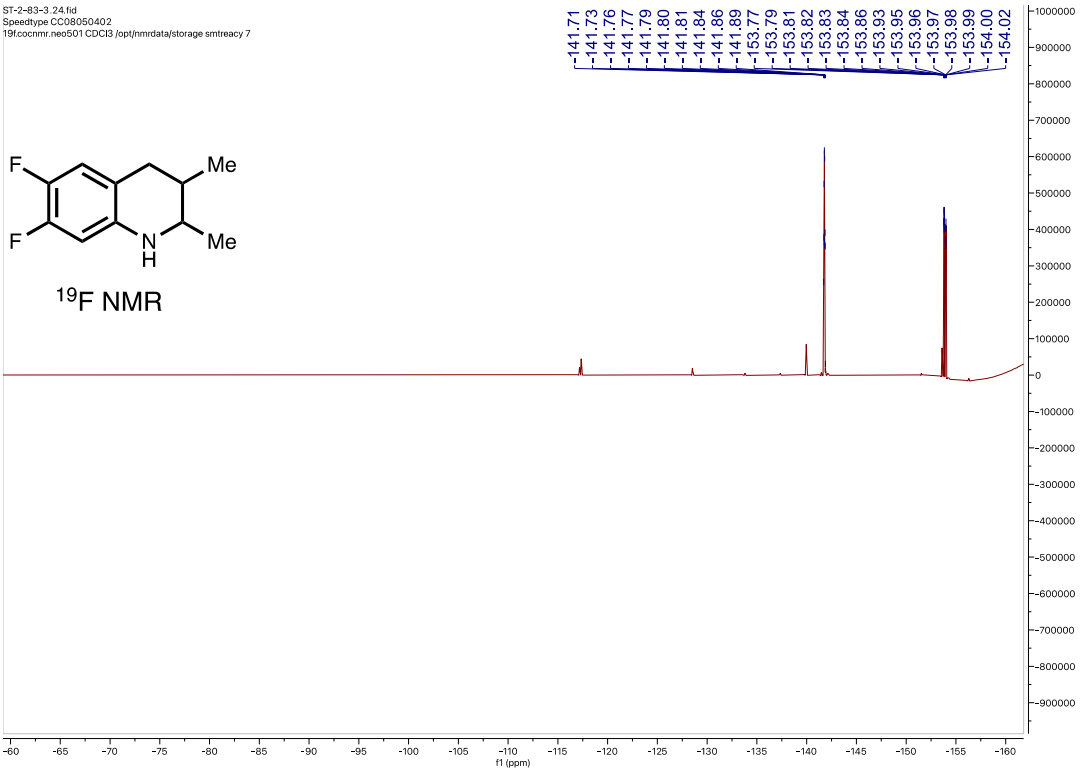

ST-2-83-2-iso.12.fid
Speedtype CC08050402
13c.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 33
156.59
141.06
122.84
122.67
122.62
115.62
115.44
114.90
114.84
113.39
113.35
113.21
113.17
77.16 CDCl3
47.46
33.48
30.02
18.14
14.28
Me
Me
F
N
H
13C NMR
3s
K (s) 156.59
A (s) 141.06
B (m) 122.71
D (d) 115.53
C (dd) 113.28
L (d) 114.87
E (s) 47.46
F (s) 33.48
G (s) 30.02
I (s) 18.14
J (s) 14.28
f1 (ppm)

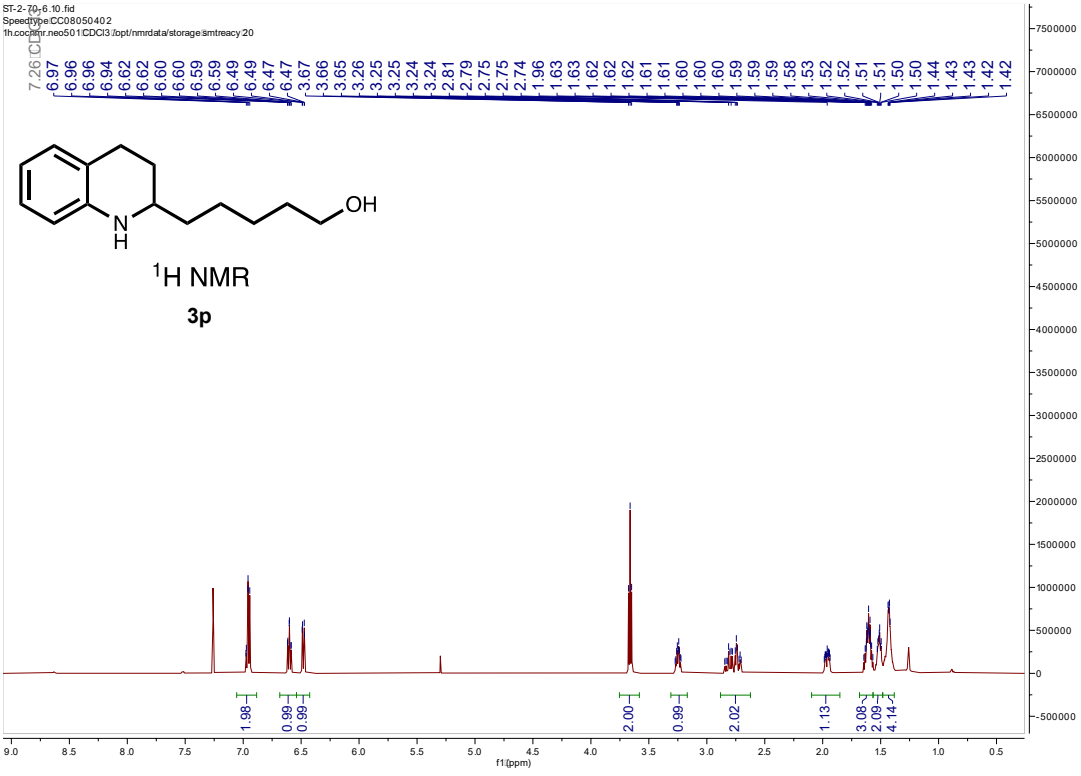

ST-2-83-2-iso.11.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 33
128.27
128.28
128.30
128.32
128.50
128.52
128.52
128.55
Me
Me
F
N
H
19F NMR
3s
f1 (ppm)

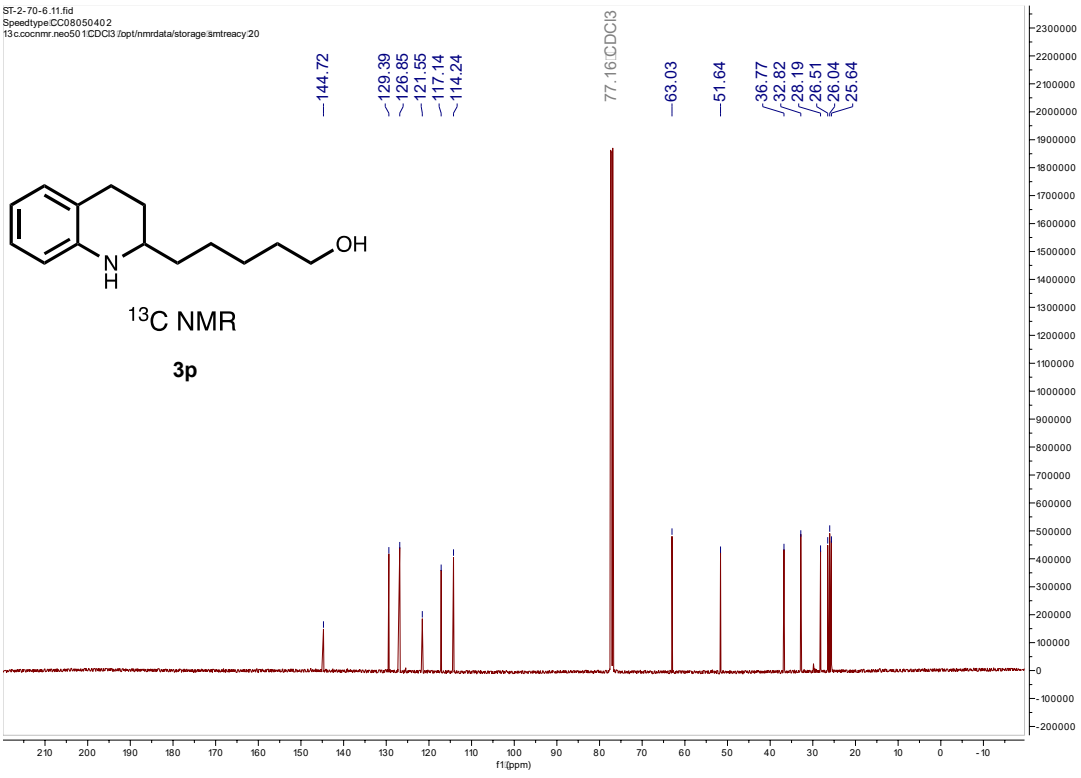

ST-2-83-3.23.fid
Speedtype:CC08050402
1h.cocnmr.neo501:CDCl3:/opt/nmrdata/storage/smtreacy:7
7.26:CDCl3
F
F
Me
N
H
Me
1H NMR
3t
1.67
1.67
1.62
1.01
0.88
1.87
1.00
3.11
3.16
f1(ppm)


ST-2-83-3.25.fid
Speedtype:CC08050402
13c.cocnmr.neo501:CDCl3:/opt/nmrdata/storage/smtreacy:7
150.01
143.66
143.57
141.69
140.30
117.78
117.64
102.04
101.88
77.16:CDCl3
49.92
33.31
30.30
18.08
14.21
F
F
Me
N
H
Me
13C NMR
3t
f1(ppm)

3t


ST-2-83-3_24.fid
Speedtype CC08050402
19f.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 7
F
Me
F
N
H
Me
19F NMR
-141.71
-141.73
-141.76
-141.77
-141.79
-141.80
-141.81
-141.84
-141.86
-141.89
-153.77
-153.79
-153.81
-153.82
-153.83
-153.84
-153.86
-153.93
-153.95
-153.96
-153.97
-153.98
-153.99
-154.00
-154.02
f1 (ppm)

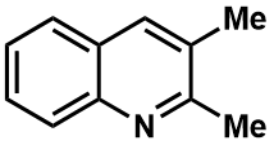

ST-2-70-6_10_fid
Speedtype CC08050402
1h.cocnmr.neo501 CDCl3 /opt/nmrdata/storage smtreacy 20
7.26 CHCl3
6.97
6.96
6.96
6.94
6.62
6.62
6.60
6.60
6.59
6.59
6.49
6.49
6.47
6.47
3.67
3.66
3.65
3.26
3.25
3.25
3.24
3.24
2.81
2.79
2.75
2.75
2.74
1.96
1.63
1.63
1.62
1.62
1.62
1.61
1.61
1.60
1.60
1.60
1.59
1.59
1.59
1.59
1.58
1.53
1.52
1.52
1.51
1.51
1.50
1.50
1.44
1.43
1.43
1.42
1.42
N
H
OH
1H NMR
3p
1.98
0.99
0.99
2.00
0.99
2.02
1.13
3.08
2.09
4.14
f1 (ppm)


ST-2-70-6.11.fid
Speedtype:CC08050402
13c.cocnmr.neo501:CDCl3:/opt/nmrdata/storage:smtreacy:20
144.72
129.39
126.85
121.55
117.14
114.24
77.16:CDCl3
63.03
51.64
36.77
32.82
28.19
26.51
26.04
25.64
OH
N
H
13C NMR
3p
f1(ppm)

VI.28 – XYZ Coordinates

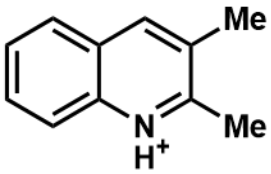
The following *xyz* coordinates were used in the calculation of the non-covalent interaction descriptors as described above. By default, these coordinates correspond to the *xTB*-optimized *ground state conformer* (*GFN2-xTB/alpb=water*) following conformational searching with *CREST (quick, GFN2-xTB/alpb=water*) as detailed in **Molecular descriptor calculation workflow**. For the DFT-optimized structures, the coordinates instead correspond to the *Gaussian16-optimized structure (wB97XD/6-311++g(d,p)/smd=water)*.

- ***From Extended Discussion on Figure II-5A – DFT-optimized structures***



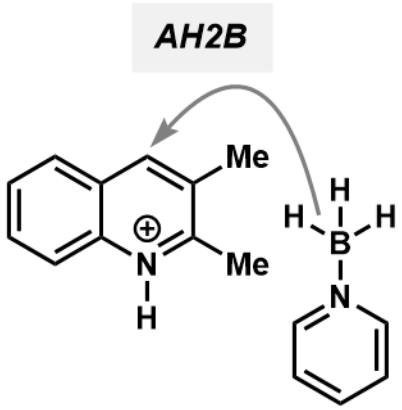
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-480.524277826**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-480.368207**

C	-3.15202200	0.68164900	0.00000700
C	-1.98282500	1.39718600	0.00000800
C	-0.73442000	0.72780000	0.00000000
C	-0.70126600	-0.68542400	-0.00000400
C	-1.92288200	-1.40159700	-0.00000600
C	-3.11940900	-0.73151600	0.00000000
H	0.51367200	2.49562500	0.00000800
H	-4.10628200	1.19617000	0.00001300
H	-1.99515100	2.48240400	0.00001300
C	0.50705600	1.40927400	0.00000000
H	-1.88909800	-2.48568500	-0.00001200
H	-4.05036000	-1.28754400	-0.00000200
C	1.61055700	-0.72015100	-0.00001400
C	1.68218600	0.71181700	-0.00000800
N	0.47442700	-1.38188600	-0.00001000
C	2.88013000	-1.51723900	0.00001600
H	3.48651700	-1.27838500	-0.87916700
H	3.48624000	-1.27869700	0.87948000
H	2.66099500	-2.58449100	-0.00020000
C	3.01359100	1.40576100	-0.00000500
H	3.60256800	1.12952100	0.87985600
H	3.60245300	1.12974100	-0.88001400
H	2.88328000	2.48919300	0.00013500



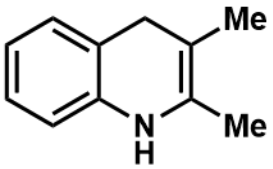
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-480.98809583**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-480.817552**

C	-3.15980300	0.69756700	-0.00032900
C	-1.99409600	1.41793600	-0.00035400
C	-0.74821900	0.74756400	-0.00017000
C	-0.74395400	-0.65764400	0.00004400
C	-1.93913900	-1.39615900	0.00006400
C	-3.13000700	-0.71439900	-0.00012200
H	0.49943000	2.50421400	-0.00034500
H	-4.11442200	1.20973400	-0.00046800
H	-2.00376700	2.50189200	-0.00051200
C	0.49766400	1.41879200	-0.00018600
H	-1.90205400	-2.47964200	0.00022400
H	-4.06185700	-1.26723900	-0.00010900
C	1.65002600	-0.67528700	0.00022200
C	1.68982000	0.74006700	0.00000600
N	0.47305700	-1.29056300	0.00023100
C	2.88215400	-1.50448400	0.00049600
H	3.48343200	-1.26245900	-0.87979000
H	3.48272600	-1.26283600	0.88137900
H	2.65721700	-2.57046000	0.00019100
C	3.01377100	1.44350000	-0.00005500
H	3.60039300	1.16927400	0.88117300
H	3.59993000	1.16990300	-0.88179600
H	2.87020500	2.52400900	0.00035700
H	0.46806400	-2.30717100	0.00037700



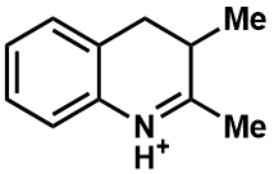
Imaginary frequency:
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-755.902105527**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-755.619301**

C	-2.44574400	-1.09668200	1.31537600	H	-3.21874600	2.81838000	0.45604300
C	-1.18984300	-1.21240400	1.90197300	H	-1.21034600	2.49119900	1.85773500
C	-0.16598400	-1.77039000	1.17920200	C	1.09432600	0.28776700	-1.55816800
C	-1.59836500	-2.12013300	-0.66072300	H	-1.24449300	0.84899100	-2.80215900
C	-2.64960900	-1.55679900	0.01803500	H	-3.24592200	1.99329700	-1.89403600
H	-3.26264100	-0.64613000	1.86550900	C	2.28641000	0.25722000	-0.63387300
H	-0.99768700	-0.85220200	2.90300200	C	2.19902400	0.72003900	0.62413900
H	0.84100500	-1.86830000	1.56076400	H	1.38088200	0.70298900	-2.53241900
H	-1.68088400	-2.49446200	-1.67180700	N	1.01349500	1.28239300	1.10897300
H	-3.61216100	-1.47474200	-0.46694300	H	1.02919700	1.70484100	2.02355000
N	-0.37331300	-2.22997900	-0.08351400	C	3.50552000	-0.36620400	-1.24514400
B	0.77172000	-2.75437300	-0.86414300	H	3.84172200	0.22028400	-2.10897800
H	0.57019000	-3.15669600	-1.95764500	H	4.34540300	-0.45978800	-0.55732300
H	0.77352100	-0.74596100	-1.78933400	H	3.26759200	-1.36833500	-1.62395100
H	1.82424800	-2.73546400	-0.32720700	C	3.28197400	0.69980100	1.66065700
C	-2.35035200	2.31792900	0.04142900	H	4.20180000	0.24990800	1.29488400
C	-1.22530600	2.13997300	0.83074500	H	3.50733600	1.71740700	1.99583700
C	-0.09768600	1.48839000	0.31620200	H	2.94837900	0.13378700	2.53672000
C	-0.10103400	1.02311100	-1.00481800				
C	-1.24288200	1.21767200	-1.78035100				
C	-2.36682900	1.85837100	-1.27464000				



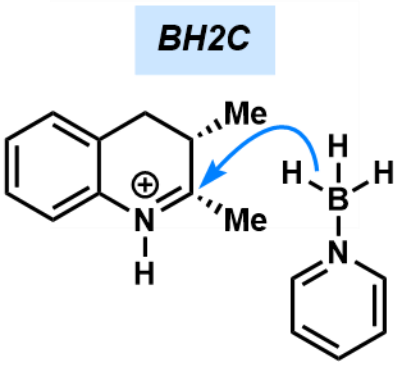
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-481.721013632**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-481.542469**

C	3.15296300	-0.67588800	0.24859200
C	1.96942300	-1.38623200	0.07389000
C	0.76130000	-0.73404400	-0.15700400
C	0.74451600	0.66510300	-0.18375000
C	1.92859400	1.38647300	-0.00548700
C	3.12623300	0.71629500	0.20162100
H	-0.51642400	-2.42537900	0.13845900
H	4.08550700	-1.20175200	0.41901200
H	1.97972800	-2.47180200	0.10719500
C	-0.52481500	-1.47782100	-0.41009800
H	1.89930900	2.47162000	-0.02474000
H	4.03999900	1.28465000	0.33747900
C	-1.67883900	0.67127400	-0.04929700
C	-1.74279100	-0.66750300	-0.02159700
N	-0.46885400	1.32025100	-0.38691600
C	-2.78886200	1.63818000	0.23145500
H	-2.91117100	2.32461700	-0.61282000
H	-3.74172500	1.14241000	0.40360400
H	-2.55272600	2.24497400	1.11223500
C	-2.96654600	-1.47519700	0.29588600
H	-2.77619000	-2.13871700	1.14753100
H	-3.84186900	-0.87040100	0.53308800
H	-3.22549000	-2.12253200	-0.55124100
H	-0.45125800	2.30734600	-0.17173700
H	-0.59276800	-1.75063100	-1.47492200



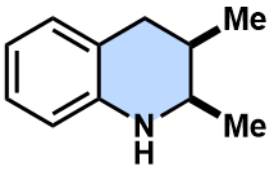
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-482.190171387**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-481.996738**

C	3.20597500	-0.65052400	0.07597700
C	2.03122500	-1.38606400	-0.04553400
C	0.79984500	-0.74552500	-0.12081600
C	0.78643400	0.64474600	-0.06584400
C	1.94499200	1.39942400	0.05062900
C	3.16479600	0.73993900	0.12218400
H	-0.44944600	-2.47140200	0.10714100
H	4.15793900	-1.16532500	0.13520400
H	2.06612500	-2.46945100	-0.08412100
C	-0.50750800	-1.46187300	-0.30127000
H	1.88746800	2.48185700	0.08700700
H	4.08010000	1.31183200	0.21589400
C	-1.62167800	0.74540000	0.02957400
C	-1.66085500	-0.71309700	0.37232300
N	-0.46874900	1.30933000	-0.14769900
C	-2.84892900	1.56074300	-0.05109000
H	-3.47669700	1.17174200	-0.85812700
H	-3.41350700	1.43783700	0.87748500
H	-2.63764300	2.61553100	-0.22257800
C	-3.00582200	-1.37507600	0.09156200
H	-2.95472800	-2.42129800	0.39854200
H	-3.81867600	-0.90085700	0.64330700
H	-3.23908200	-1.34324400	-0.97596500
H	-0.44204200	2.31127400	-0.32372700
H	-0.72122800	-1.55718400	-1.37327800
H	-1.48818800	-0.71517600	1.46094600



Imaginary frequency:
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-757.126635126**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-756.818087**

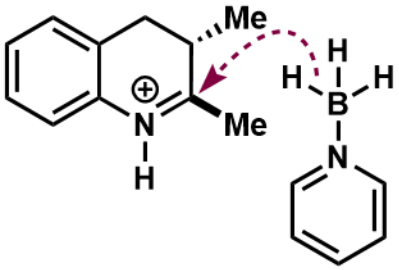
C	2.18213900	2.98918400	-0.15656700	H	2.92082100	-1.29306700	1.60126600
C	2.56444200	1.78893400	-0.74330200	H	-0.58220000	-0.49195700	1.34533200
C	1.96898500	0.57641800	-0.39722000	H	1.91589200	-0.76542800	-2.06011700
C	0.95979100	0.57727400	0.58021600	H	2.04371400	-2.82472900	-0.90809800
C	0.58804200	1.78453800	1.18819500	C	-2.88240800	-0.96837800	-0.65871900
C	1.18614900	2.97818000	0.81853100	C	-3.66244100	-0.66879500	0.42882300
H	3.46655100	-0.70571800	-1.23867300	C	-3.61755000	0.61905800	0.95689800
H	2.65768700	3.91838500	-0.44926600	C	-2.78814000	1.57485400	0.38214300
H	3.34695000	1.78332900	-1.49704200	C	-2.01417200	1.22272000	-0.69668400
C	2.38547000	-0.71298900	-1.07114200	N	-2.06612600	-0.03191900	-1.21158100
H	-0.18825700	1.77232600	1.94708300	H	-4.22675000	0.87440600	1.81542400
H	0.87578100	3.90206000	1.29484800	H	-2.86410400	-1.94555900	-1.12080200
C	0.50682400	-1.78363900	0.13009300	H	-4.29652600	-1.43363600	0.85507700
C	1.98241500	-1.94737400	-0.25710500	H	-2.72778700	2.58152500	0.77130600
N	0.35448500	-0.61346100	0.98050500	H	-1.33326600	1.90849500	-1.18133200
C	-0.12395000	-2.99255500	0.79519300	B	-1.16872700	-0.40794900	-2.33328700
H	-0.03487900	-3.86774600	0.14712600	H	-1.31830400	-1.48483100	-2.79700800
H	0.35437900	-3.21519000	1.75235500	H	-0.02448400	-1.63374400	-0.83014100
H	-1.18781400	-2.81273000	0.97994800	H	-0.38875300	0.40718100	-2.68664700
C	2.90942600	-2.16300300	0.93820700				
H	2.61374900	-3.03396700	1.52853700				
H	3.93259300	-2.32858400	0.58943900				



wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-482.947863795**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-482.744006**

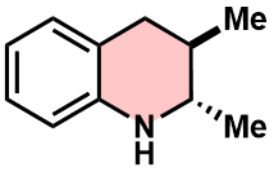
C	3.04861000	0.90101900	0.25586400
C	1.78040300	1.47013300	0.27354100
C	0.63293400	0.72338400	0.01151400
C	0.76875600	-0.64310300	-0.28977700
C	2.04773700	-1.21886900	-0.30224000
C	3.17388200	-0.45663200	-0.03210000
H	-0.99927100	1.52394000	1.13527700
H	3.92434500	1.50484300	0.46492800
H	1.66782100	2.52676800	0.50090100
C	-0.73587500	1.35867900	0.08256200
H	2.14496700	-2.27632000	-0.53013700
H	4.15241900	-0.92488200	-0.04725600
C	-1.65913500	-0.96567100	-0.17227300
C	-1.82016300	0.50572700	-0.57892100
N	-0.34211200	-1.41711200	-0.62933300
C	-1.86791700	-1.23184000	1.31904700
H	-1.15214300	-0.67636600	1.93221200
H	-2.87719200	-0.95507000	1.63392500
H	-1.73307000	-2.29724500	1.52667000
C	-3.21312800	1.05172400	-0.28800800
H	-3.98660100	0.42407300	-0.74100900
H	-3.40594900	1.10539300	0.78767100
H	-3.31844000	2.06209800	-0.69372000
H	-0.70948900	2.34866300	-0.38283000
H	-1.66215600	0.53826900	-1.66331400
H	-2.39803000	-1.54835700	-0.72971100
H	-0.18905600	-2.40332300	-0.46352700

BH2D



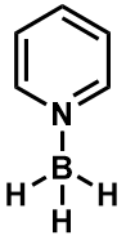
Imaginary frequency:
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-757.127243105**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-756.818981**

C	-2.07507500	2.70411100	-0.13932900	H	3.53523500	-0.36632400	-1.54825100
C	-0.86407300	2.53835200	-0.80337800	H	0.48298200	-0.55820300	2.10461900
C	0.10872000	1.67000000	-0.31973400	H	1.24308700	0.71697000	-1.85773500
C	-0.16534500	0.96639400	0.85334000	H	2.77830000	1.68842700	0.58920800
C	-1.36697900	1.12788700	1.53471800	C	-0.89216000	-2.42117400	0.36661600
C	-2.32411500	1.99944600	1.03377300	C	-2.20480500	-2.17272000	0.70923200
H	1.75273300	2.36138800	-1.50269600	C	-2.98573800	-1.39605000	-0.13648300
H	-2.82030900	3.38441400	-0.53526000	C	-2.42578800	-0.89382200	-1.30441400
H	-0.66547900	3.08915300	-1.71721600	C	-1.10584900	-1.17255400	-1.58576500
C	1.40741200	1.43290100	-1.04360600	N	-0.35325500	-1.92899900	-0.76451100
H	-1.54969600	0.56206400	2.44197400	H	-4.01861900	-1.18410000	0.11152600
H	-3.26397200	2.12331500	1.55905000	H	-0.23506300	-3.01868800	0.98440800
C	1.96588100	-0.19455500	0.78285500	H	-2.60022100	-2.58337900	1.62830700
C	2.50325400	0.89615900	-0.12242200	H	-2.99686900	-0.28287600	-1.99021200
N	0.78483000	0.04879600	1.35167900	H	-0.61097200	-0.79819500	-2.47164900
C	2.90452400	-1.05535500	1.56931900	B	1.16723400	-2.12650100	-1.04811300
H	3.63168000	-1.53995900	0.92122600	H	1.56226200	-3.16013800	-0.57393800
H	3.43876800	-0.41026100	2.27432400	H	1.72019200	-1.19827700	-0.37447800
H	2.35526200	-1.81483600	2.12797800	H	1.41953400	-1.91018100	-2.20434800
C	3.74500100	0.49216300	-0.90359300				
H	4.05633700	1.32496900	-1.53799500				
H	4.57845600	0.24154200	-0.24522600				



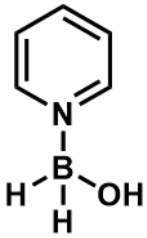
wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-482.94842921**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-482.744919**

C	-3.24182400	-0.68121000	0.01774300
C	-2.03970700	-1.37975500	0.03989100
C	-0.80857700	-0.72677200	0.02278200
C	-0.78810400	0.67684500	-0.03183500
C	-1.99957300	1.38308000	-0.05045000
C	-3.21240200	0.71123600	-0.02500600
H	0.64323800	-1.85466500	1.11196500
H	-4.18596000	-1.21370700	0.03380300
H	-2.04877800	-2.46574800	0.07503200
C	0.48458300	-1.50444000	0.08340800
H	-1.97539100	2.46817800	-0.08837600
H	-4.13736300	1.27799300	-0.03944600
C	1.64333000	0.69306700	0.32039100
C	1.70575600	-0.68618600	-0.34387600
N	0.41959100	1.36469400	-0.11705900
C	2.98989300	-1.43841900	-0.01077300
H	3.87028800	-0.95955500	-0.44530100
H	3.13382600	-1.49680000	1.07397900
H	2.94171600	-2.46084400	-0.39674700
H	0.40477700	-2.40234200	-0.53691400
H	1.66006700	-0.52424400	-1.42854300
H	0.35653300	2.32103500	0.20554400
C	2.83134200	1.58329500	-0.00566600
H	2.97396800	1.66027700	-1.08828100
H	2.67388400	2.59001400	0.39331100
H	3.74769800	1.19057900	0.43883500
H	1.61605400	0.53252900	1.41088500



wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-274.929410571**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-274.837567**

C	0.21550100	1.15905000	-0.01324000
C	-1.16464200	1.19548800	0.00362100
C	-1.87035400	-0.00017700	0.01210000
C	-1.16429700	-1.19568900	0.00361400
C	0.21578800	-1.15885900	-0.01323400
N	0.89448800	0.00020400	-0.02093500
H	-2.95366300	-0.00033300	0.02268900
H	0.81384800	2.06022800	-0.02283400
H	-1.66991500	2.15211500	0.00726300
H	-1.66931800	-2.15244800	0.00723300
H	0.81440200	-2.05987200	-0.02283300
B	2.49078900	0.00003300	0.01920300
H	2.82388600	-0.00448700	1.19153900
H	2.86671800	1.01008500	-0.54102600
H	2.86669700	-1.00576600	-0.54867100



wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-350.207213974**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-350.108794**

C	-0.44261100	1.24557600	-0.00388200
C	-1.80015800	1.03210700	0.14725600
C	-2.28159500	-0.26917600	0.12860500
C	-1.38485400	-1.31698100	-0.04497800
C	-0.04378600	-1.02916100	-0.19538300
N	0.41645500	0.23155000	-0.17449200
H	-3.34033300	-0.46597000	0.24763600
H	-0.01156400	2.23794400	0.01014100
H	-2.46005000	1.87866800	0.28276400
H	-1.71522200	-2.34689700	-0.06790600
H	0.69510700	-1.80731500	-0.34006000
B	2.01094200	0.50424900	-0.39816600
H	2.15295500	1.69295800	-0.18251000
H	2.21871000	0.20204900	-1.56360900
O	2.78183000	-0.29229900	0.55345600
H	2.95387900	-1.16932900	0.20846500

H_2O

wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-76.4462016631**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-76.443075**

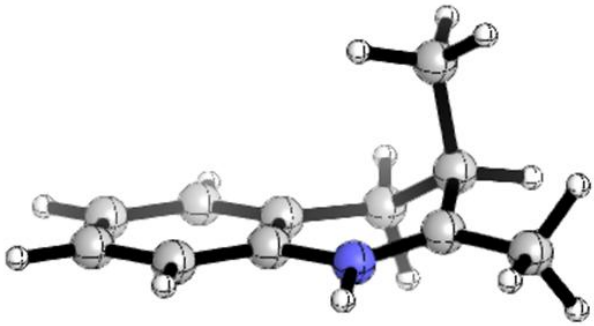
O	0.00000000	0.11769900	0.00000000
H	0.75952700	-0.47080200	0.00000000
H	-0.75952700	-0.47078800	0.00000000

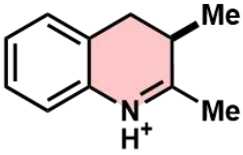
HO^-

wB97XD/6-311++g(d,p)/smd=water calculated energy (Hartrees): **-75.949118531**
Sum of electronic and thermal free energies corrected energy (Hartrees): **-75.956479**

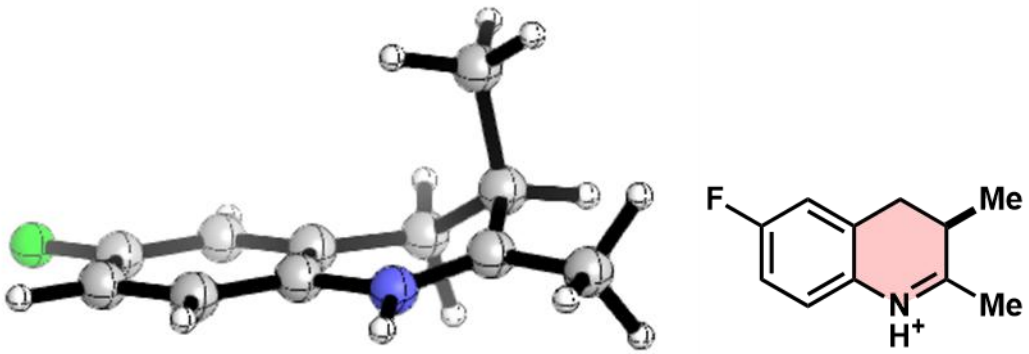
O	0.00000000	0.00000000	0.10661400
H	0.00000000	0.00000000	-0.85291000

- ***From Figure II-5C – pro-anti dihydroquinoline substrates***

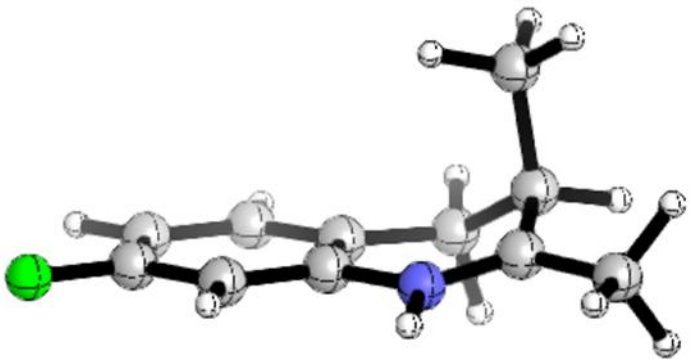


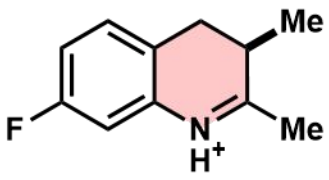


C	-1.8568039805	-0.5349987738	0.1818364830
N	-0.7652703416	-1.2276448833	0.1362823800
C	0.5229064770	-0.6754372178	0.0683690065
C	1.6019260174	-1.4788127675	-0.2790746618
C	2.8669877580	-0.9232931246	-0.3390173537
C	3.0479777039	0.4200310955	-0.0458718172
C	1.9678767194	1.2138372450	0.3086960097
C	0.6921956891	0.6786942761	0.3682287534
C	-0.5120500250	1.4663422795	0.7886269325
C	-1.7727712967	0.9521359530	0.0856023719
C	-3.1624558574	-1.1934152723	0.2769909244
C	-1.7536865579	1.3125725873	-1.4115523837
H	-3.8762250891	-0.7237208146	-0.3969271508
H	-3.1117542415	-2.2603250644	0.0811179477
H	-3.5579323335	-1.0432168685	1.2834698035
H	-0.8116162624	-2.2464555078	0.1338632464
H	1.4546408983	-2.5222297791	-0.5094300698
H	3.7116222795	-1.5354491415	-0.6117188860
H	4.0356874173	0.8510991243	-0.0925150547
H	2.1190225049	2.2563166647	0.5413018409
H	-0.3704122719	2.5228969241	0.5655846260
H	-0.6259657437	1.3737221697	1.8730870043
H	-2.6584174557	1.4037165222	0.5359242490
H	-2.6350345226	0.9314792086	-1.9188632903
H	-1.7402020687	2.3932185133	-1.5173648012
H	-0.8672188482	0.9104511779	-1.8952095620

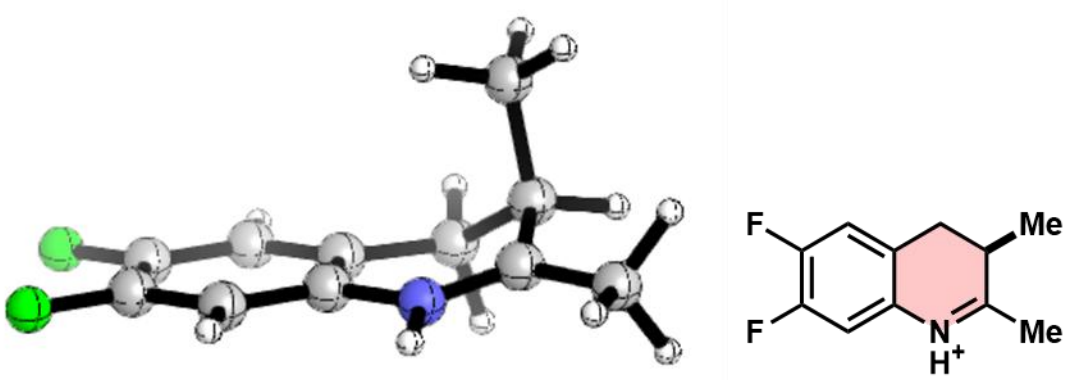


C	-2.3362393524	-0.4198665562	0.1727613627
N	-1.3109868584	-1.2096183721	0.1307131073
C	0.0217901079	-0.7767465343	0.0747011048
C	1.0253469964	-1.6749257369	-0.2688762300
C	2.3391095414	-1.2473389619	-0.3184748520
C	2.6119240236	0.0745225846	-0.0126909693
C	1.6269278674	0.9811048048	0.3381790651
C	0.3099141443	0.5571091205	0.3826427497
C	-0.8187698460	1.4510443429	0.7949035625
C	-2.1174693563	1.0530363618	0.0851061270
F	3.8990989178	0.5020138689	-0.0568666613
C	-3.6944323302	-0.9579506537	0.2549175268
C	-2.0606528330	1.4149315672	-1.4110867181
H	1.8958642360	1.9981235521	0.5720856481
H	0.7888115525	-2.6996378295	-0.5056113612
H	3.1421733002	-1.9138752747	-0.5841426545
H	-1.4508945567	-2.2200741575	0.1210041951
H	-4.0753130515	-0.7896229134	1.2646152726
H	-4.3634143229	-0.4122914735	-0.4075268671
H	-3.7421299072	-2.0215941883	0.0416576726
H	-2.9599586991	1.5819772166	0.5335021452
H	-1.2129684256	0.9362649028	-1.8942156681
H	-1.9495929830	2.4902307774	-1.5126298979
H	-2.9711585969	1.1174420063	-1.9223548886
H	-0.5801779122	2.4899040692	0.5732055323
H	-0.9451028683	1.3691907923	1.8784728219

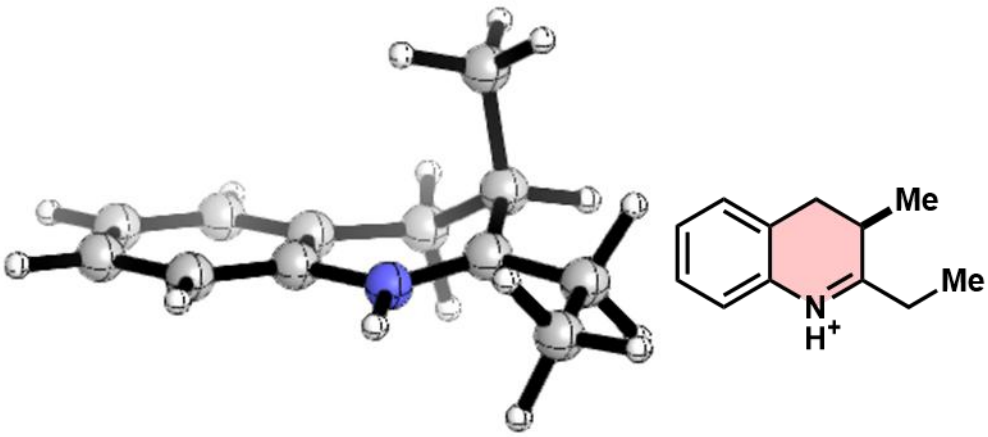




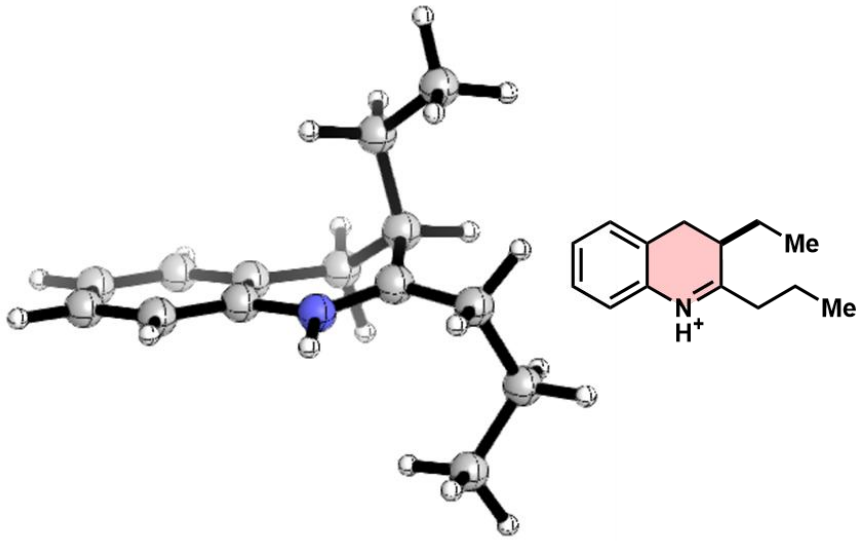
C	2.1740814922	-0.7152426398	-0.1798865150
N	0.9834213466	-1.2231453330	-0.2302561584
C	-0.1989947867	-0.4706022821	-0.1777360689
C	-1.4076601224	-1.1101592675	0.0678591347
C	-2.5495520930	-0.3344786973	0.1063868946
C	-2.5233516243	1.0336405497	-0.0962533945
C	-1.3064605318	1.6469885743	-0.3464330356
C	-0.1335639076	0.9096525044	-0.3883364081
C	1.2021644082	1.5147379737	-0.6966470407
C	2.3238941441	0.7548561350	0.0192173687
F	-3.7383329649	-0.9451314081	0.3503919567
C	2.2823182025	1.0125004253	1.5375977022
C	3.3574689436	-1.5701314515	-0.2676393638
H	-3.4406864371	1.5971064357	-0.0573502111
H	-1.2737254346	2.7117777263	-0.5103857283
H	-1.4735655684	-2.1728735191	0.2343856884
H	1.2201129764	2.5623289845	-0.4010612529
H	1.3576526910	1.4776475502	-1.7788126857
H	3.2933777108	1.0848817480	-0.3573105936
H	2.4381301722	2.0710461897	1.7220146399
H	1.3182788373	0.7308724619	1.9527589807
H	3.0621407947	0.4591946372	2.0520472070
H	3.1302341797	-2.6240610081	-0.1379413037
H	4.1109367771	-1.2568586126	0.4523087518
H	3.8131537277	-1.4326757242	-1.2508158972
H	0.8667236813	-2.2341999389	-0.2973237763



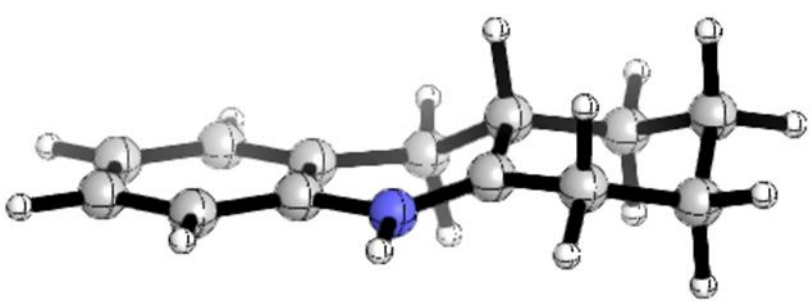
C	-2.5956792063	-0.5631306444	0.1883933088
N	-1.4733000193	-1.2116110423	0.2361463949
C	-0.2103152330	-0.6074881153	0.1756901328
C	0.9125779902	-1.3892670918	-0.0676626322
C	2.1492970354	-0.7763513475	-0.1166358316
C	2.2623536656	0.5943536222	0.0811341823
C	1.1414877028	1.3669531206	0.3283217849
C	-0.1093437875	0.7733184809	0.3765827799
C	-1.3584003004	1.5398022858	0.6817353912
C	-2.5692136588	0.9123919256	-0.0177972531
F	3.4769401513	1.1812339037	0.0328524467
F	3.2502679924	-1.5217346526	-0.3558375573
C	-3.8702596188	-1.2693686175	0.2845439307
C	-2.5127778415	1.1524204814	-1.5389158455
H	1.2634939506	2.4266319322	0.4815508906
H	0.8502137393	-2.4535775725	-0.2267075986
H	-1.4801038086	-2.2295027993	0.3061109945
H	-3.7753604930	-2.3414503308	0.1423573164
H	-4.5933485441	-0.8578635621	-0.4167327947
H	-4.2905460827	-1.0896278072	1.2770743948
H	-3.4880769315	1.3595538216	0.3642944048
H	-1.5942787672	0.7543162970	-1.9614887556
H	-3.3588048357	0.6938921062	-2.0414233747
H	-2.5426327307	2.2206682848	-1.7300893001
H	-1.2510527050	2.5779153624	0.3729843137
H	-1.5084062361	1.5354949429	1.7649993115

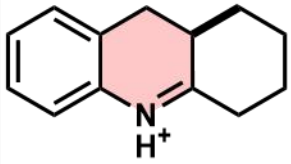


C	-1.6294307742	-0.0670972624	-0.1568173425
N	-0.7280774558	0.8575586608	-0.1028917755
C	0.6529602426	0.6183101473	-0.0511428943
C	1.5190410078	1.6436433432	0.3079709666
C	2.8792244832	1.3965887732	0.3478556776
C	3.3644987745	0.1391157836	0.0217765569
C	2.4949483281	-0.8770646093	-0.3441441397
C	1.1291385979	-0.6528811358	-0.3821486991
C	0.1348924751	-1.6907424681	-0.8080924873
C	-1.1991137836	-1.4959868615	-0.0792487585
C	-1.0664096517	-1.8562083437	1.4119389254
C	-3.0658474967	0.2564737023	-0.2338663714
C	-3.4264340613	1.7307614339	-0.0855945252
H	-0.7866427239	-2.9015728990	1.5020020546
H	-0.2975988574	-1.2544292918	1.8894825596
H	-2.0058357092	-1.7087806287	1.9363732150
H	-1.9606519853	-2.1402715314	-0.5219883966
H	-3.4287929528	-0.1162857975	-1.1980166691
H	-3.5892182033	-0.3398834416	0.5187649039
H	-4.5062460323	1.8408340778	-0.1243986929
H	-3.0927151526	2.1249035818	0.8712595474
H	-3.0119845234	2.3283770330	-0.8938746260
H	-1.0136378005	1.8362885023	-0.0814117026
H	1.1352229742	2.6188852554	0.5628360482
H	3.5613098692	2.1824543897	0.6299657704
H	4.4258508439	-0.0507919779	0.0522476106
H	2.8817787459	-1.8505689058	-0.6014644119
H	-0.0146756964	-1.6071332496	-1.8888778739
H	0.5208653329	-2.6894891129	-0.6091470059

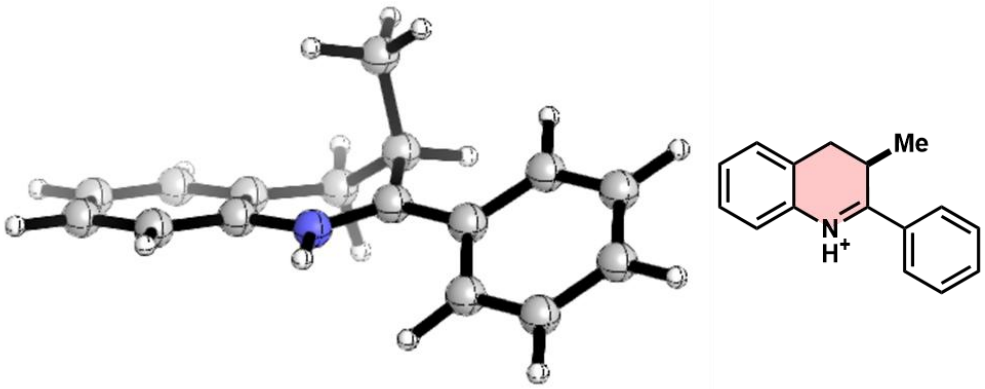


C	-1.2553092169	0.0205407983	0.3007659999
N	-0.2664940792	0.6751536172	0.8183329713
C	1.0762478487	0.4990069381	0.4531932437
C	2.0891851037	1.0094657769	1.2552116421
C	3.4061966820	0.8486050833	0.8641814310
C	3.7014573414	0.1946793353	-0.3224677082
C	2.6846114446	-0.3042383795	-1.1231169854
C	1.3606709497	-0.1640693928	-0.7444029705
C	0.2073623989	-0.6449894978	-1.5730646264
C	-0.9592163175	-1.0687880367	-0.6712870572
C	-1.7305999206	-2.9128151766	0.9488423340
C	-0.5905464685	-2.3578653724	0.1045198935
C	-2.6441200934	0.3756724134	0.6447315552
C	-3.2401561657	1.2721616084	-0.4688317567
C	-2.5510640536	2.6273591793	-0.5592934064
H	-2.6275105945	-3.0497071902	0.3494923128
H	-1.4476141560	-3.8841107487	1.3470132472
H	-1.9616524680	-2.2656922806	1.7905270349
H	-0.3040079490	-3.1042219096	-0.6371796412
H	0.2782200765	-2.1698380872	0.7367023977
H	-1.8435377280	-1.2810548866	-1.2747727037
H	-2.6953326700	0.9157240365	1.5902306977
H	-3.2532359141	-0.5253166234	0.7065608561
H	-3.1771878977	0.7607076330	-1.4291860003
H	-4.2948975639	1.4194075226	-0.2387257491
H	-3.0530876450	3.2439190232	-1.3007466174
H	-1.5103566063	2.5315183682	-0.8590761982
H	-2.5986712707	3.1514730891	0.3921409275
H	-0.4462926811	1.3860026402	1.5272733589
H	1.8528782758	1.5161690928	2.1775268656
H	4.2018732732	1.2346163391	1.4808084938
H	4.7291217085	0.0724211261	-0.6263872712
H	2.9239208126	-0.8060365068	-2.0474720280
H	0.5128065888	-1.4805179416	-2.2010754315
H	-0.1093602747	0.1649575093	-2.2378189429



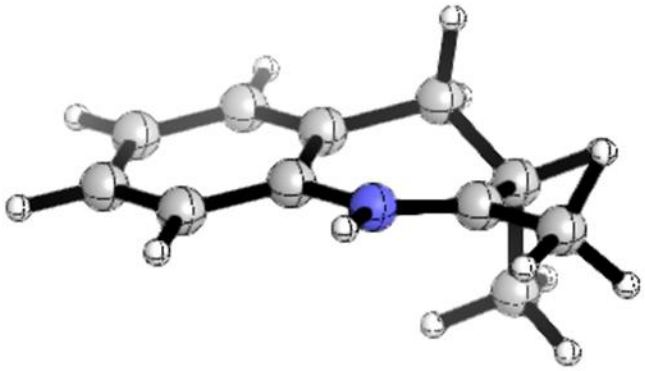


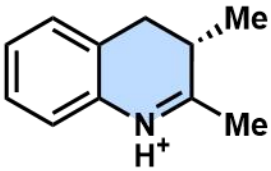
C	-1.0901709802	0.6827218892	0.0622312878
N	0.0467427803	1.2695497588	-0.1413363695
C	1.2983864687	0.6399458008	-0.0803716787
C	2.4531001507	1.4077736289	0.0100302419
C	3.6795110850	0.7747961457	0.0983347299
C	3.7464455727	-0.6104655451	0.0922438763
C	2.5906291686	-1.3697253860	-0.0093730608
C	1.3530551259	-0.7545802623	-0.0955430795
C	0.0717643707	-1.5073608982	-0.2856188627
C	-1.1242928348	-0.7792504043	0.3436975327
C	-2.3208587509	1.4957738767	0.1420732183
C	-3.5721306553	0.7565472273	-0.3418879606
C	-3.6550292264	-0.6163826545	0.3116325096
C	-2.4405182220	-1.4424496560	-0.0908388629
H	-1.0540003278	-0.8563670702	1.4392529901
H	-2.2017299650	2.4408419519	-0.3895617235
H	-2.4640471420	1.7353565697	1.2024664201
H	-4.4531956694	1.3493648800	-0.0997384806
H	-3.5355616035	0.6427820535	-1.4261306616
H	-4.5622155674	-1.1266954237	-0.0120704152
H	-3.7018271942	-0.5140582846	1.3968544538
H	-2.4952102478	-2.4372303607	0.3500663978
H	-2.4362317774	-1.5625097970	-1.1758246020
H	0.0711172879	2.2793152359	-0.2852109573
H	2.3947585034	2.4847418470	0.0205129889
H	4.5827323031	1.3589840075	0.1711372850
H	4.7041530027	-1.1016722783	0.1622322582
H	2.6523563975	-2.4465404006	-0.0246390677
H	-0.1004094582	-1.6167975713	-1.3616161316
H	0.1533384670	-2.5104346312	0.1301110174



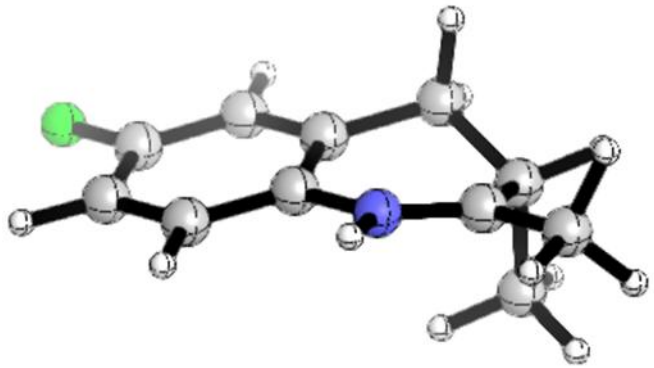
C	-0.5287935165	-0.1945944831	-0.0494073329
N	0.3242937283	0.7760817692	0.1217308535
C	1.7132005416	0.6359243788	0.0786192266
C	2.5275381727	1.6577518627	0.5563958985
C	3.9008272655	1.5151751238	0.4894973061
C	4.4528243046	0.3671627817	-0.0604946676
C	3.6360381349	-0.6443824874	-0.5435078173
C	2.2587160859	-0.5260474776	-0.4759369566
C	1.3113734651	-1.5570273389	-1.0073175529
C	0.0183304086	-1.5828199853	-0.1842560722
C	0.2881383798	-2.1266258773	1.2303908138
C	-1.9503089797	0.0725867589	-0.0503864082
C	-2.8738388935	-0.9412796626	0.2458977836
C	-4.2279319143	-0.6744138494	0.2572133853
C	-4.6950306032	0.5957356130	-0.0481040004
C	-3.7955932871	1.6037938100	-0.3660118303
C	-2.4391848744	1.3525788608	-0.3621243838
H	-0.6064062718	-2.0937947746	1.8455130089
H	1.0676708846	-1.5504498587	1.7218510464
H	0.6193287369	-3.1584443172	1.1577110670
H	-0.7070374364	-2.2253385256	-0.6841536445
H	-2.5354808138	-1.9318971452	0.4990998883
H	-4.9258650933	-1.4596879268	0.5011560469
H	-5.7545963765	0.7976149593	-0.0460575995
H	-4.1584245032	2.5872797665	-0.6201361734
H	-1.7686132886	2.1452887405	-0.6548759657
H	-0.0171670661	1.7152371689	0.3195148999
H	2.0921597703	2.5469571211	0.9845305909
H	4.5421589142	2.2976722276	0.8619881662
H	5.5247245385	0.2594570813	-0.1128424156
H	4.0735199249	-1.5313125714	-0.9741503658
H	1.7735339105	-2.5430445220	-0.9891062910
H	1.0845912070	-1.3200175830	-2.0514068671

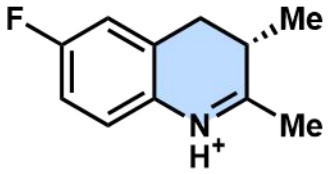
- ***From Fig. II-5C – pro-syn dihydroquinoline substrates***



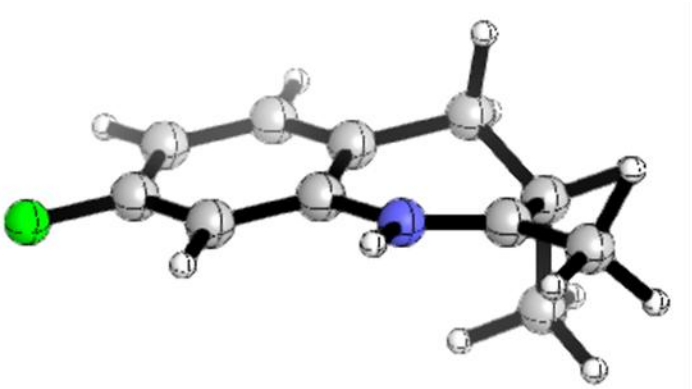


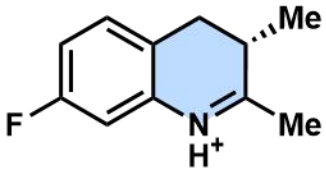
C	-1.8570169171	-0.5346127543	-0.1819652029
N	-0.7655936120	-1.2275118402	-0.1369145542
C	0.5226418929	-0.6754596462	-0.0689778603
C	1.6015190770	-1.4789673538	0.2787838661
C	2.8665362231	-0.9235355997	0.3393798953
C	3.0478280686	0.4197098293	0.0460127016
C	1.9679834727	1.2135913214	-0.3089550852
C	0.6922022513	0.6786292788	-0.3686469927
C	-0.5119498130	1.4667017429	-0.7885115203
C	-1.7724284229	0.9524495037	-0.0851812135
C	-3.1629013988	-1.1924614834	-0.2767321183
C	-1.7521601687	1.3114830504	1.4123293440
H	-3.1132096727	-2.2586727031	-0.0774099734
H	-3.8774042654	-0.7199083996	0.3943235969
H	-3.5567504708	-1.0455095404	-1.2843608378
H	-0.8121019743	-2.2463108655	-0.1347006681
H	1.4540002627	-2.5224612508	0.5087768833
H	3.7111119372	-1.5357664890	0.6121971788
H	4.0356025118	0.8505431020	0.0930830005
H	2.1191573723	2.2561691358	-0.5412082273
H	-0.6263427452	1.3745072257	-1.8729446256
H	-0.3700050693	2.5231491850	-0.5651342124
H	-2.6582017473	1.4046551861	-0.5346319299
H	-0.8654257024	0.9086946650	1.8949831333
H	-1.7383499725	2.3920162114	1.5191896607
H	-2.6332005483	0.9300787988	1.9198922118



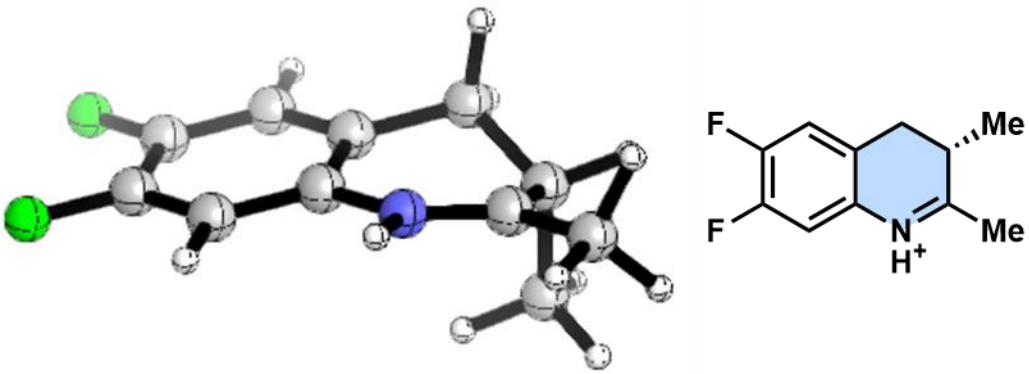


C	-2.3364956542	-0.4194341192	-0.1722021472
N	-1.3114559777	-1.2095267297	-0.1310274525
C	0.0214731374	-0.7769587985	-0.0752929515
C	1.0249256311	-1.6750924680	0.2685690998
C	2.3387203085	-1.2475100604	0.3183818066
C	2.6116304261	0.0743089556	0.0126284399
C	1.6267003419	0.9808651755	-0.3386089426
C	0.3097551027	0.5568566200	-0.3834412542
C	-0.8190127711	1.4505225638	-0.7959775431
C	-2.1169730987	1.0534012800	-0.0844122893
F	3.8987082815	0.5019653455	0.0573655466
C	-3.6949195427	-0.9569744453	-0.2539955776
C	-2.0577085095	1.4148791965	1.4117507107
H	1.8956954429	1.9978942898	-0.5724121533
H	0.7883975853	-2.6997996580	0.5053432831
H	3.1416493338	-1.9141057452	0.5843157436
H	-1.4516523935	-2.2199366282	-0.1216636055
H	-4.0743604448	-0.7918034875	-1.2647740002
H	-3.7434221071	-2.0199246565	-0.0377341488
H	-4.3644148249	-0.4088183221	0.4058628685
H	-2.9597427487	1.5829574366	-0.5315525205
H	-1.2108790976	0.9337153433	1.8941498806
H	-2.9686467545	1.1196794777	1.9236055611
H	-1.9436689995	2.4898709633	1.5133143938
H	-0.9465399352	1.3669499225	-1.8792710630
H	-0.5800978409	2.4896722122	-0.5761095215

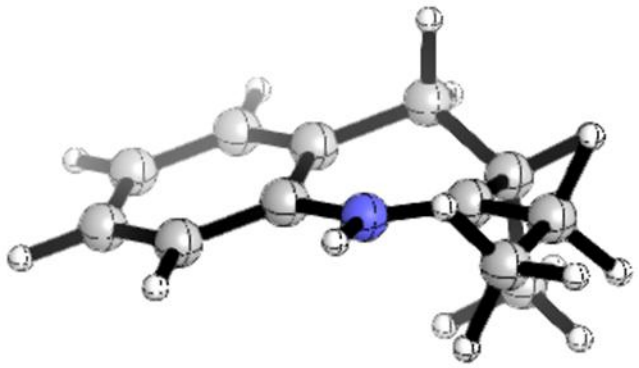


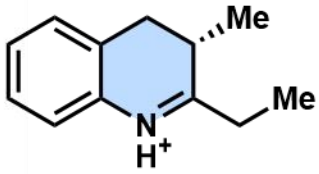


C	2.1739661979	-0.7154456468	0.1800861181
N	0.9831053031	-1.2230668830	0.2293870537
C	-0.1991178857	-0.4704163014	0.1769518010
C	-1.4079051296	-1.1099037611	-0.0684121047
C	-2.5499157299	-0.3343964603	-0.1064737660
C	-2.5235475434	1.0336686134	0.0963968902
C	-1.3065902018	1.6470429797	0.3461593876
C	-0.1336064441	0.9098109437	0.3877354092
C	1.2020990627	1.5148345187	0.6962271194
C	2.3242936112	0.7546906122	-0.0185298128
F	-3.7387118385	-0.9450415605	-0.3498412932
C	2.2842260647	1.0124313954	-1.5369135015
C	3.3570544181	-1.5706515438	0.2684068820
H	-3.4409038159	1.5972075026	0.0581307386
H	-1.2740624661	2.7118055570	0.5101251629
H	-1.4735032614	-2.1726534469	-0.2349887970
H	1.3569788854	1.4781864325	1.7784970444
H	1.2202395416	2.5624170022	0.4003255971
H	3.2934946825	1.0843831857	0.3590668613
H	1.3204125934	0.7314691912	-1.9531181371
H	2.4408794961	2.0708639624	-1.7212138864
H	3.0641562656	0.4586350815	-2.0506473283
H	4.1121967939	-1.2559842089	-0.4491148849
H	3.8105522760	-1.4353718238	1.2528397057
H	3.1300197723	-2.6242144842	0.1362118025
H	0.8665403092	-2.2341911404	0.2958864495

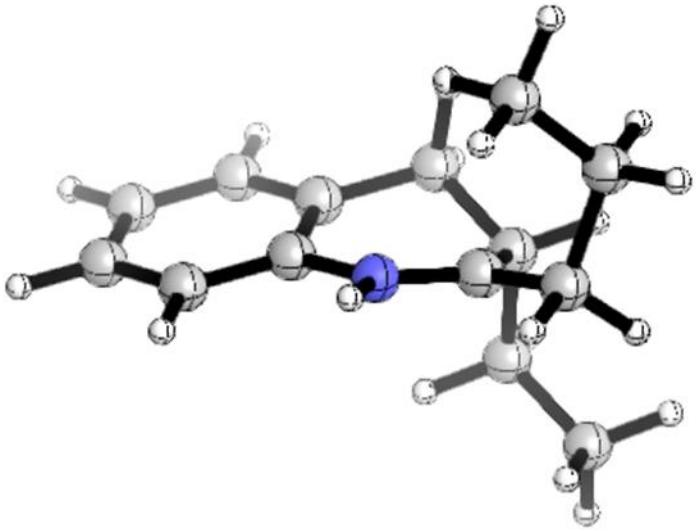


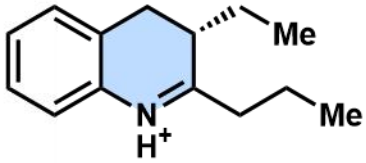
C	-2.5958160760	-0.5630355585	-0.1879717933
N	-1.4734649966	-1.2115576320	-0.2361434359
C	-0.2104236516	-0.6076174554	-0.1758647182
C	0.9124514606	-1.3894024310	0.0675496714
C	2.1490905091	-0.7763493140	0.1166061158
C	2.2621632944	0.5943837180	-0.0810721386
C	1.1413272263	1.3669504226	-0.3284402122
C	-0.1094356872	0.7731837092	-0.3769020786
C	-1.3585724879	1.5393324537	-0.6825509302
C	-2.5690272421	0.9125350448	0.0181871000
F	3.4768158851	1.1810747187	-0.0326796898
F	3.2501682765	-1.5216419162	0.3557209023
C	-3.8704372387	-1.2691651602	-0.2841496114
C	-2.5112661390	1.1527921485	1.5391566268
H	1.2631339474	2.4266456875	-0.4817110671
H	0.8501507775	-2.4537066872	0.2266296800
H	-1.4808124996	-2.2294208630	-0.3065164953
H	-3.7752080797	-2.3416853015	-0.1450221344
H	-4.5923472482	-0.8595156117	0.4195188882
H	-4.2925501082	-1.0865464620	-1.2754093241
H	-3.4880578952	1.3599176161	-0.3632591789
H	-1.5929820303	0.7538645995	1.9613836598
H	-2.5398577652	2.2210876527	1.7299708490
H	-3.3575103539	0.6954023045	2.0422841044
H	-1.2512348858	2.5779223848	-0.3753771091
H	-1.5090121183	1.5333117052	-1.7657812022



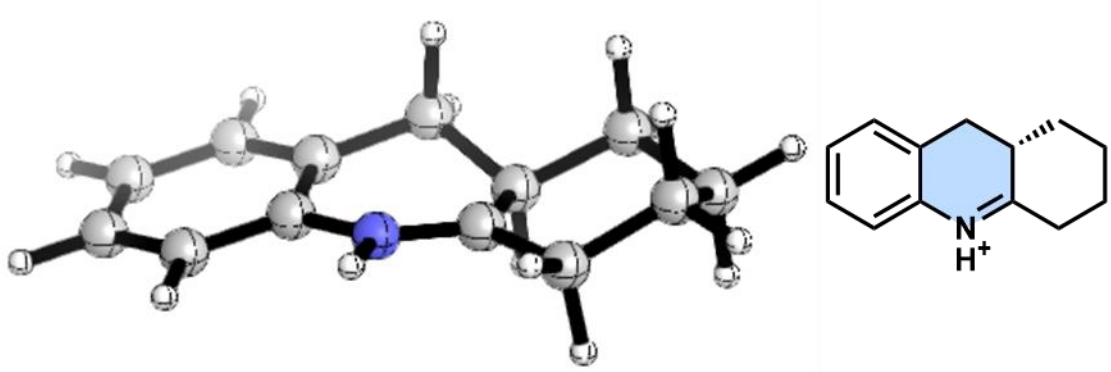


C	-1.6295073817	-0.0672649213	0.1560256595
N	-0.7283959427	0.8577028809	0.1031554119
C	0.6526773764	0.6188019857	0.0519218674
C	1.5187421685	1.6441365401	-0.3072435996
C	2.8788600853	1.3968781875	-0.3477905471
C	3.3640833800	0.1390448440	-0.0227537697
C	2.4945563528	-0.8770204232	0.3436966288
C	1.1288776161	-0.6524373211	0.3829238602
C	0.1343509555	-1.6896242286	0.8099893387
C	-1.1985847153	-1.4960509449	0.0789151999
C	-1.0630582000	-1.8568769619	-1.4118194888
C	-3.0661246388	0.2555483583	0.2318558418
C	-3.4273323341	1.7298678209	0.0855589842
H	-0.2897300806	-1.2588033626	-1.8866987327
H	-0.7881140169	-2.9035572492	-1.5010219335
H	-1.9999368031	-1.7047496025	-1.9394556321
H	-1.9605487011	-2.1405189484	0.5206335455
H	-3.4301848240	-0.1189900462	1.1948051302
H	-3.5882111708	-0.3399370701	-0.5224446095
H	-3.0149908197	2.3263237902	0.8957172448
H	-3.0920778270	2.1259601816	-0.8698848775
H	-4.5073198694	1.8389501164	0.1224039730
H	-1.0147703560	1.8362159876	0.0818832298
H	1.1349284774	2.6195415693	-0.5616175846
H	3.5608960772	2.1828659420	-0.6298221863
H	4.4253891546	-0.0511957036	-0.0540778806
H	2.8814734938	-1.8506724512	0.6005755952
H	0.5205697416	-2.6888231078	0.6140165372
H	-0.0167826092	-1.6031559321	1.8903501315

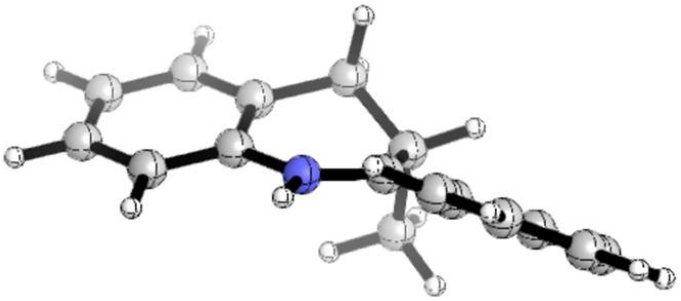


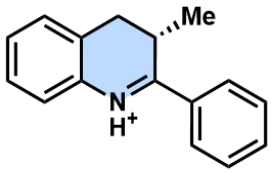


C	-1.2572467834	-0.0166634536	0.2991424089
N	-0.2699443477	-0.6729123311	0.8175596093
C	1.0732224407	-0.4992302715	0.4529358037
C	2.0850602912	-1.0090775966	1.2566722763
C	3.4024747532	-0.8500929770	0.8662513523
C	3.6991120183	-0.1988847889	-0.3215363358
C	2.6833304512	0.2994127358	-1.1239021390
C	1.3590343204	0.1612944706	-0.7457726957
C	0.2065077826	0.6426708515	-1.5752539700
C	-0.9582898917	1.0714037090	-0.6734582467
C	-1.7191314532	2.9138241454	0.9533006184
C	-0.5844268287	2.3595145438	0.1013170218
C	-2.6468123293	-0.3705543367	0.6410973156
C	-3.2381590057	-1.2725931744	-0.4706462860
C	-2.5489783509	-2.6283128937	-0.5515220879
H	-1.4326771957	3.8838320471	1.3522284059
H	-2.6194037142	3.0532120045	0.3595445003
H	-1.9459986329	2.2647287581	1.7946214439
H	0.2880495600	2.1702407402	0.7279973541
H	-0.3019790439	3.1065058523	-0.6412458304
H	-1.8422339302	1.2861372978	-1.2766166272
H	-2.7005275922	-0.9067561601	1.5886558752
H	-3.2568290577	0.5301535851	0.6973232863
H	-4.2938943113	-1.4185063636	-0.2444022515
H	-3.1708908305	-0.7657518802	-1.4331296952
H	-2.6004617348	-3.1476267382	0.4022949299
H	-1.5071283117	-2.5343696448	-0.8478555878
H	-3.0482423415	-3.2485361024	-1.2917870064
H	-0.4519050633	-1.3830498916	1.5266270101
H	1.8475105261	-1.5140124753	2.1796386803
H	4.1973523166	-1.2359378386	1.4840261221
H	4.7271036086	-0.0782047992	-0.6249506858
H	2.9237513280	0.7993583670	-2.0489723836
H	-0.1126488298	-0.1682943018	-2.2375993138
H	0.5134233736	1.4758623266	-2.2055662832



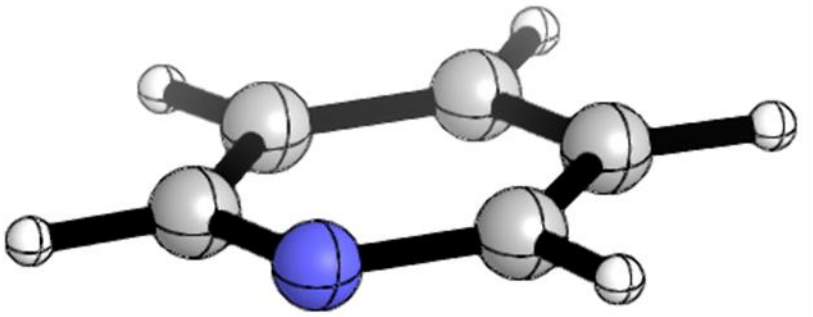
C	-1.0901456759	0.6828356335	-0.0622445056
N	0.0468309555	1.2696238868	0.1412749180
C	1.2984792696	0.6400238816	0.0803071226
C	2.4532335024	1.4077057809	-0.0104611810
C	3.6794794891	0.7745558963	-0.0988550317
C	3.7463382484	-0.6106724678	-0.0921672320
C	2.5905746833	-1.3698004372	0.0100096569
C	1.3530895586	-0.7545334869	0.0959457507
C	0.0715873454	-1.5069794435	0.2862260675
C	-1.1241328845	-0.7790932234	-0.3439350089
C	-2.3208147704	1.4958610142	-0.1415383784
C	-3.5719539215	0.7562833024	0.3423864653
C	-3.6553361585	-0.6163247724	-0.3118226547
C	-2.4405597167	-1.4425916587	0.0895902170
H	-1.0530126425	-0.8561036469	-1.4394547270
H	-2.4643230881	1.7359155437	-1.2018251641
H	-2.2016960505	2.4407351720	0.3904748696
H	-3.5349167577	0.6420529316	1.4266393005
H	-4.4531206653	1.3492643140	0.1008892114
H	-3.7027976706	-0.5135352958	-1.3970008404
H	-4.5623762651	-1.1268063320	0.0123006752
H	-2.4363104297	-1.5638993953	1.1744155933
H	-2.4952791744	-2.4368553811	-0.3524310431
H	0.0712689853	2.2793856319	0.2851885765
H	2.3949946841	2.4846584802	-0.0213626416
H	4.5827198842	1.3585976697	-0.1722057713
H	4.7039618148	-1.1020005013	-0.1623658895
H	2.6522513390	-2.4466189057	0.0253498749
H	-0.1009522438	-1.6153516808	1.3623140563
H	0.1530132189	-2.5105092383	-0.1285237597

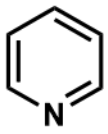




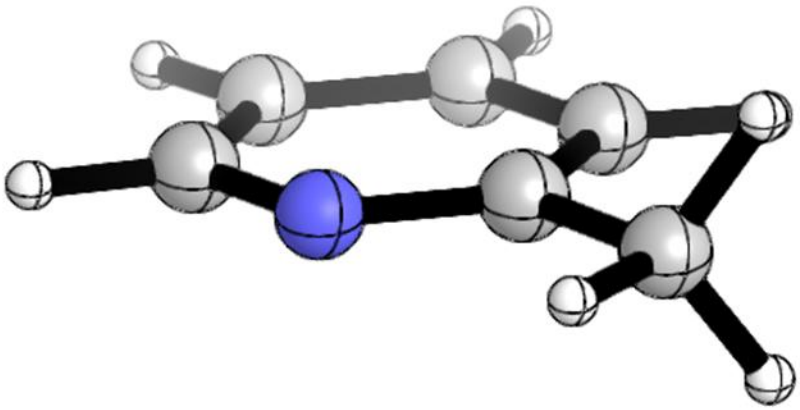
C	-0.5288712194	-0.1945715162	0.0482694148
N	0.3242477751	0.7760712686	-0.1228769820
C	1.7131456067	0.6358094885	-0.0790816538
C	2.5278140751	1.6571762762	-0.5572757833
C	3.9010657028	1.5146211573	-0.4893482327
C	4.4526807999	0.3671435980	0.0621369980
C	3.6355776488	-0.6440247213	0.5453707078
C	2.2583034752	-0.5258333284	0.4765849708
C	1.3106269969	-1.5567827843	1.0074292458
C	0.0183438602	-1.5827539295	0.1832184572
C	0.2896402753	-2.1259571600	-1.2313545068
C	-1.9503499969	0.0726059382	0.0496653566
C	-2.8740087215	-0.9413432338	-0.2459323438
C	-4.2281356178	-0.6745972396	-0.2561444831
C	-4.6951026574	0.5955494675	0.0493921759
C	-3.7955006432	1.6037553990	0.3663663195
C	-2.4390804831	1.3526301634	0.3615180324
H	0.6178331769	-3.1587507347	-1.1590840379
H	1.0716356903	-1.5513986355	-1.7208005845
H	-0.6036680858	-2.0900365315	-1.8480755625
H	-0.7072591640	-2.2256586631	0.6822472918
H	-2.5357311141	-1.9319229164	-0.4994138703
H	-4.9262263943	-1.4600088723	-0.4991910099
H	-5.7547002018	0.7972735980	0.0484320010
H	-4.1581995578	2.5872915549	0.6204908493
H	-1.7683479022	2.1454857779	0.6535019529
H	-0.0170703344	1.7152417026	-0.3208346950
H	2.0927195922	2.5459549202	-0.9865721408
H	4.5426604154	2.2967016604	-0.8622573595
H	5.5245431080	0.2595036625	0.1153832312
H	4.0727670239	-1.5305649400	0.9771198332
H	1.0828103720	-1.3195698397	2.0512380537
H	1.7728231472	-2.5427980577	0.9898300938

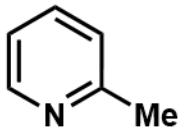
- ***From Figure II-5D – pyridine substrates***



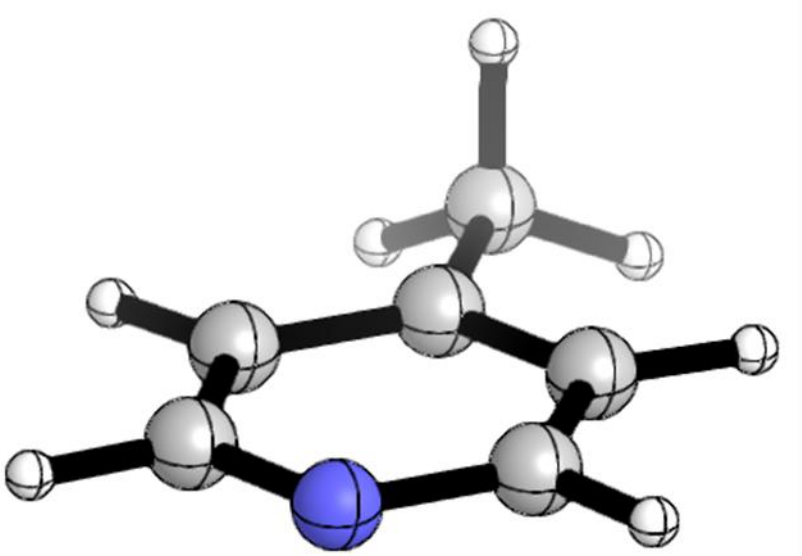


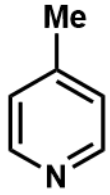
N	1.3748666514	-0.0003236518	0.0000080474
C	0.6905473498	-1.1342172230	0.0000149379
C	-0.6949378688	-1.1918345581	-0.0000330744
C	-1.4020050354	0.0003452543	-0.0000330656
C	-0.6942997109	1.1921647404	0.0000097178
C	0.6910998602	1.1338635108	-0.0000276945
H	1.2883051834	-2.0358573451	0.0000448878
H	-1.2026643411	-2.1429616019	0.0002945401
H	-2.4817879897	0.0007002806	0.0000504610
H	-1.2016555467	2.1435351224	0.0003167161
H	1.2893451935	2.0352473613	0.0000059560



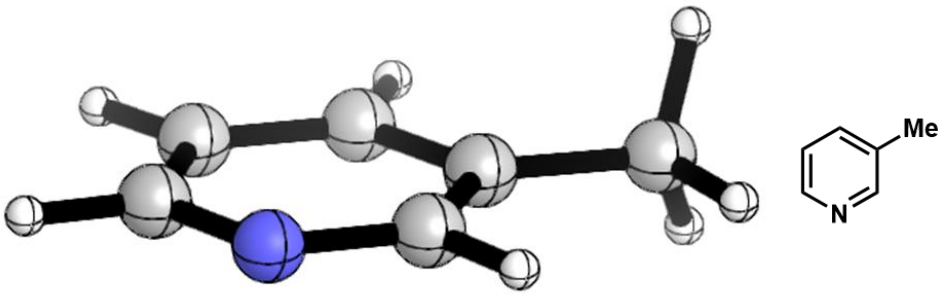


N	0.2965618720	1.1580296343	-0.0001904830
C	0.8991640055	-0.0258893331	-0.0002357599
C	0.1783606442	-1.2201649039	-0.0001447161
C	-1.2030723651	-1.1716983208	0.0000465448
C	-1.8276309264	0.0677303561	0.0001523756
C	-1.0271324554	1.1971105742	0.0000570632
C	2.3949835609	-0.0292943403	0.0001812297
H	2.7684343661	-0.5548456768	-0.8761771960
H	2.7674669939	-0.5404246356	0.8855204078
H	2.7630507645	0.9914519907	-0.0075361722
H	0.7015762542	-2.1639047908	-0.0002170456
H	-1.7856647875	-2.0812212774	0.0000691462
H	-2.9022281529	0.1543660434	0.0002252038
H	-1.4586659310	2.1897189346	0.0000866513

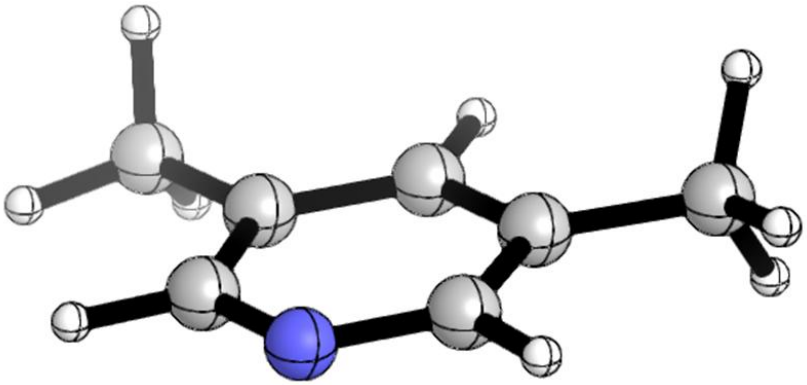


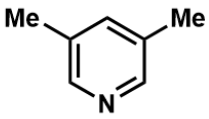


N	-1.8366398635	-0.0004947650	-0.0047636709
C	-1.1474105676	1.1313376983	-0.0015104868
C	0.2356301013	1.1876443483	0.0050959124
C	0.9620180923	0.0005042893	0.0087833439
C	0.2364795937	-1.1871485956	0.0051010742
C	-1.1464375546	-1.1318453182	-0.0016253296
C	2.4559102650	0.0002150355	-0.0057327928
H	-1.7422858238	2.0352628171	-0.0034413307
H	0.7406296995	2.1411588782	0.0091323738
H	2.8464704979	-0.8766440438	0.5037759861
H	2.8086767992	-0.0225033982	-1.0364953821
H	2.8468289691	0.8976284657	0.4661797315
H	0.7422596995	-2.1402408577	0.0091263203
H	-1.7405197913	-2.0362168256	-0.0025773066

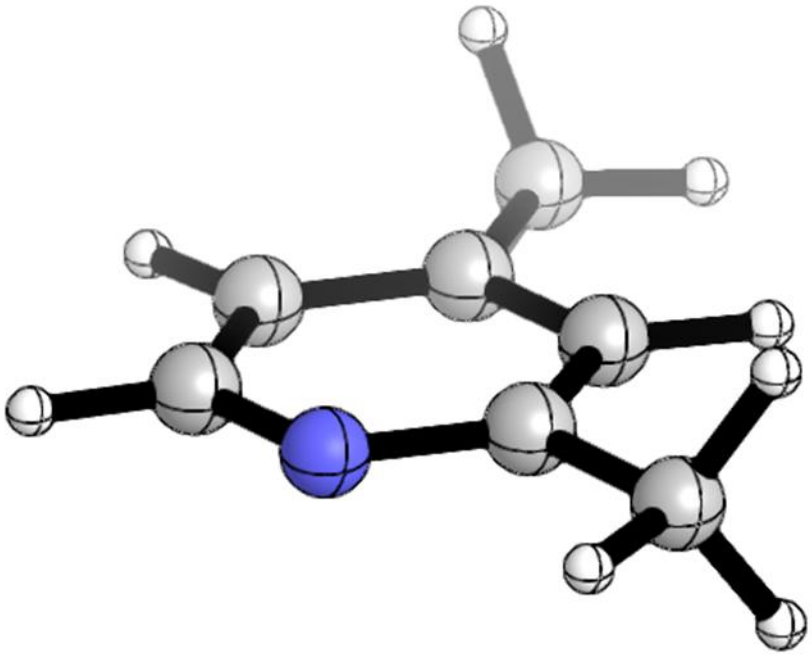


N	-1.1433280850	-1.1808048299	0.0001835295
C	0.1804379396	-1.1573687537	-0.0008126584
C	0.9418968420	0.0064233233	-0.0011856030
C	0.2479844405	1.2134410791	-0.0007499863
C	-1.1357584110	1.2025983282	0.0002165809
C	-1.7847795137	-0.0230243197	0.0007834506
C	2.4363811819	-0.0294805857	0.0009030645
H	0.6575187700	-2.1293700509	-0.0015469479
H	2.8203885763	0.4384756523	0.9057752529
H	2.8271460822	0.5224367132	-0.8518533948
H	2.7978093044	-1.0528793529	-0.0450713143
H	0.7906286296	2.1478746908	-0.0014563628
H	-1.7006126402	2.1212900535	0.0002598372
H	-2.8644176162	-0.0885344489	0.0014139875

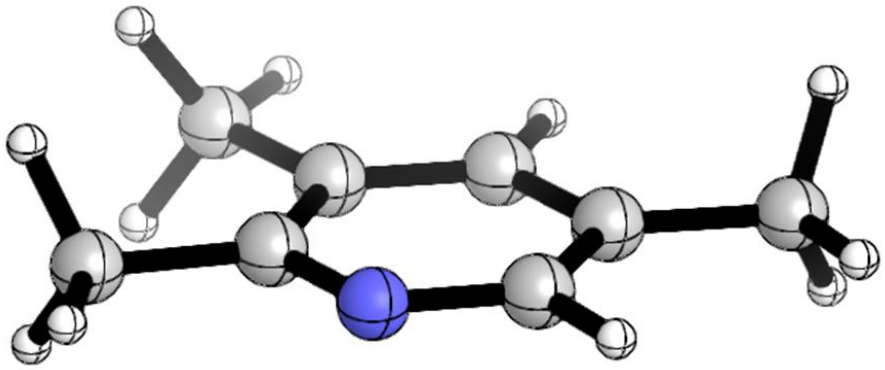


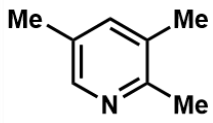


N	-0.0002109418	-1.7575738010	0.0002113217
C	1.1350269132	-1.0775952559	0.0042502919
C	1.2053309374	0.3114381474	0.0035475302
C	0.0001421356	1.0035854027	-0.0007666622
C	-1.2051039030	0.3119021068	-0.0045400919
C	-1.1350794879	-1.0773308705	-0.0042957888
C	2.5146729312	1.0343075910	-0.0018633587
C	-2.5146833368	1.0343378206	0.0028497126
H	2.0348504383	-1.6794735861	0.0079105079
H	3.3390323548	0.3389537427	0.1294673316
H	2.5423643509	1.7732531044	0.7963532504
H	2.6473156933	1.5578000670	-0.9475309088
H	0.0004856713	2.0849799639	-0.0012938729
H	-2.5265348105	1.8061826794	-0.7637987132
H	-2.6672952954	1.5172632093	0.9670677711
H	-3.3357549033	0.3450067206	-0.1733688501
H	-2.0351808775	-1.6787584807	-0.0079908739

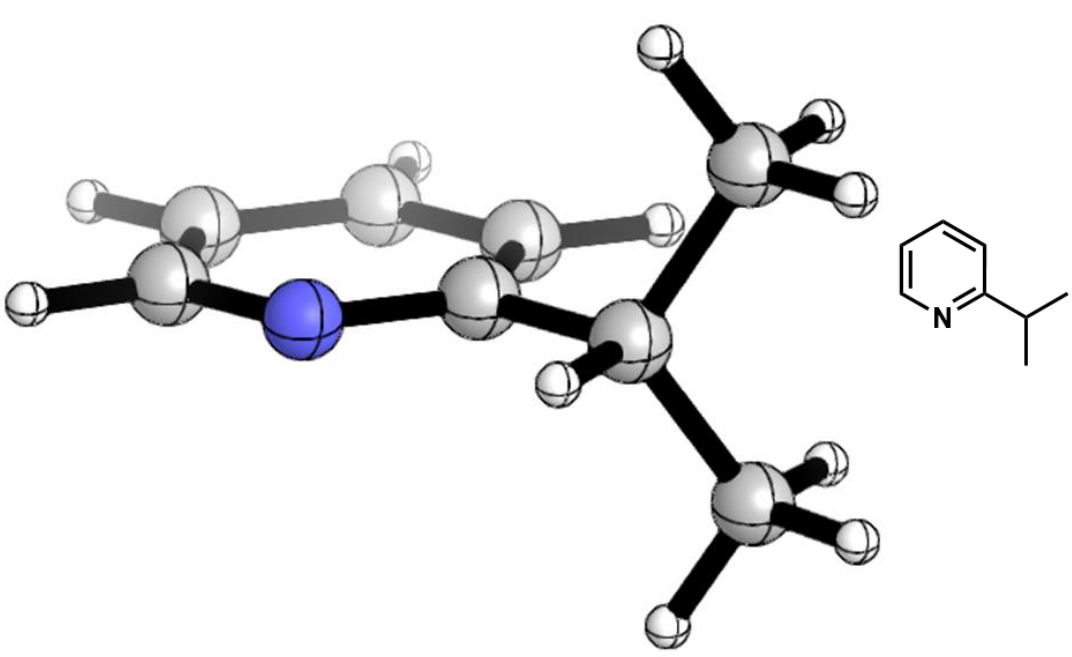


N	1.2348883147	1.0193743475	-0.0006376346
C	1.1714166942	-0.3074388730	-0.0005650484
C	-0.0429921772	-0.9905136534	-0.0001817437
C	-1.2349628698	-0.2827453254	0.0008086502
C	-1.1512068639	1.1092219228	0.0008041306
C	0.0970881766	1.7000483284	-0.0001065166
C	2.4698349952	-1.0520344513	0.0004753788
C	-2.5625539923	-0.9703163340	-0.0003641630
H	2.5393685745	-1.6817937304	0.8850584214
H	3.2950555576	-0.3473696728	-0.0048871277
H	2.5356125361	-1.6917614207	-0.8771426137
H	-0.0463559157	-2.0707888151	0.0002353289
H	-2.4416881239	-2.0492791549	0.0332985968
H	-3.1493320096	-0.6523389151	0.8593659995
H	-3.1183186207	-0.7047910168	-0.8980490068
H	-2.0445260900	1.7140677439	0.0015087401
H	0.2053497101	2.7772403370	-0.0009030501





N	-0.1893139420	1.5382948333	-0.0003555754
C	-1.0826023214	0.5577476777	0.0001063220
C	-0.7063884198	-0.7934861465	-0.0002432474
C	0.6485945670	-1.0822414869	-0.0008021546
C	1.5917598498	-0.0584301199	-0.0009266732
C	1.1001714179	1.2385409568	-0.0007502144
C	-2.5223192428	0.9645152912	0.0009104414
C	-1.7325058897	-1.8813830728	0.0000039078
C	3.0584819651	-0.3534149695	0.0012817679
H	-2.5897749584	2.0476686499	-0.0000996838
H	-3.0278021866	0.5717034818	-0.8793303763
H	-3.0263574327	0.5731065317	0.8828039680
H	-1.2547271147	-2.8574878200	-0.0008793210
H	-2.3689741514	-1.8025011570	0.8798898918
H	-2.3706776685	-1.8018040727	-0.8785427615
H	0.9792829671	-2.1123462952	-0.0012686329
H	3.3214742203	-0.9661686393	-0.8592395653
H	3.3295074637	-0.9048069081	0.9002924627
H	3.6353950068	0.5667721763	-0.0323911287
H	1.7707621674	2.0890566989	-0.0012901540



N	0.5991388900	-1.2431014123	-0.0000145985
C	-0.1805818031	-0.1666584926	-0.0007041288
C	0.3477684611	1.1226631566	-0.0013561356
C	1.7209252411	1.2870038586	-0.0008302807
C	2.5297986253	0.1603444120	0.0002197229
C	1.9126153284	-1.0791567405	0.0004821907
C	-1.6662711490	-0.4208006567	0.0004172753
C	-2.3139223641	0.1576081128	1.2604065655
C	-2.3168458612	0.1553940541	-1.2590443088
H	-1.8000188442	-1.5064734422	0.0015837614
H	-2.2188445064	1.2404470120	1.2825834786
H	-3.3706668430	-0.0957511178	1.2829090089
H	-1.8406079119	-0.2493545410	2.1508886738
H	-3.3753276543	-0.0911523460	-1.2745369428
H	-2.2148735146	1.2374170689	-1.2873549762
H	-1.8510857359	-0.2591322902	-2.1499732483
H	-0.3053890995	1.9807019709	-0.0019862709
H	2.1552040342	2.2759779298	-0.0010947946
H	3.6048896723	0.2407207970	0.0009521365
H	2.4914849794	-1.9937644449	0.0011072270

VI.29 – References

(1) Nguyen, Q. N. N.; Xia, K. T.; Zhang, Y.; Chen, N.; Morimoto, M.; Pei, X.; Ha, Y.; Guo, J.; Yang, W.; Wang, L.-P.; Bergman, R. G.; Raymond, K. N.; Toste, F. D.; Tantillo, D. J. Source of Rate Acceleration for Carbocation Cyclization in Biomimetic Supramolecular Cages. *J. Am. Chem. Soc.* **2022**, *144* (25), 11413–11424.

(2) Johnson, D. W.; Raymond, K. N. The Self-Assembly of a $[Ga_4\mathbf{L}_6]^{12-}$ Tetrahedral Cluster *Thermodynamically* Driven by Host−Guest Interactions. *Inorg. Chem.* **2001**, *40* (20), 5157–5161.

(3) Adak, S.; Braley, S. E.; Brown, M. K. Photochemical Reduction of Quinolines with γ-Terpinene. *Org. Lett.* **2024**, *26* (1), 401–405.

(4) Fiedler, D.; Bergman, R. G.; Raymond, K. N. Supramolecular Catalysis of a Unimolecular Transformation: Aza-Cope Rearrangement within a Self-Assembled Host. *Angewandte Chemie International Edition* **2004**, *43* (48), 6748–6751.

(5) Wu, J.; Wang, C.; Tang, W.; Pettman, A.; Xiao, J. The Remarkable Effect of a Simple Ion: Iodide-Promoted Transfer Hydrogenation of Heteroaromatics. *Chemistry A European J* **2012**, *18* (31), 9525–9529.

(6) 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.

(7) 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.

(8) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended TIGHT-BINDING Quantum Chemistry Methods. *WIREs Comput Mol Sci* **2021**, *11* (2), e1493.

(9) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190.

(10) Van Dijk, L.; Haas, B. C.; Lim, N.-K.; Clagg, K.; Dotson, J. J.; Treacy, S. M.; Piechowicz, K. A.; Roytman, V. A.; Zhang, H.; Toste, F. D.; Miller, S. J.; Gosselin, F.; Sigman, M. S. Data Science-Enabled Palladium-Catalyzed Enantioselective Aryl-Carbonylation of Sulfonimidamides. *J. Am. Chem. Soc.* **2023**, *145* (38), 20959–20967.

(11) Liles, J. P.; Rouget-Virbel, C.; Wahlman, J. L. H.; Rahimoff, R.; Crawford, J. M.; Medlin, A.; O'Connor, V. S.; Li, J.; Roytman, V. A.; Toste, F. D.; Sigman, M. S. Data Science Enables the Development of a New Class of Chiral Phosphoric Acid Catalysts. *Chem* **2023**, *9* (6), 1518–1537.

(12) Greg Landrum; Paolo Tosco; Brian Kelley; Ricardo Rodriguez; David Cosgrove; Riccardo Vianello; sriniker; gedeck; Gareth Jones; NadineSchneider; Eisuke Kawashima; Dan Nealschneider; Andrew Dalke; Matt Swain; Brian Cole; Samo Turk; Aleksandr Savelev; Alain Vaucher; Maciej Wójcikowski; Ichiru Take; Vincent F. Scalfani; Rachel Walker; Kazuya Ujihara; Daniel Probst; guillaume godin; Axel Pahl; Juuso Lehtivarjo; Francois Berenger; jasondbiggs; strets123. Rdkit/Rdkit: 2024_03_1 (Q1 2024) Release, 2024.

(13) Guilian Luchini; Patterson, T.; Paton, R. Patonlab/DBSTEP: Release 1.1.0, 2023.

(14) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate Classification of Phosphine Ligation State and Reactivity in Cross-Coupling Catalysis. *Science* **2021**, *374* (6565), 301–308.
(15) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
(16) Wheeler, S. E. Local Nature of Substituent Effects in Stacking Interactions. *J. Am. Chem. Soc.* **2011**, *133* (26), 10262–10274.
(17) 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 Contributors; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; De Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* **2020**, *17* (3), 261–272.
(18) Kikugawa, Y.; Saito, K.; Yamada, S. Reduction of Heterocycles with Pyridine: Borane in Acetic Acid. *Synthesis* **1978**, *1978* (06), 447–448.
(19) Jia, M.-Q.; You, S.-L. N-Heterocyclic Carbene-Catalyzed Enantioselective Intramolecular N-Tethered Aldehyde–Ketone Benzoin Reactions. *ACS Catal.* **2013**, *3* (4), 622–624.
(20) Cho, C. S.; Ren, W. X. A Recyclable Palladium-Catalyzed Modified Friedländer Quinoline Synthesis. *Journal of Organometallic Chemistry* **2007**, *692* (19), 4182–4186.
(21) Kulkarni, A. Water-Promoted, Open-Flask Synthesis of Amine-Boranes: 2-Methylpyridine-Borane (2-Picoline-Borane). *Org. Synth.* **2017**, *94*, 332–345.
(22) Wang, T.; Zhuo, L.-G.; Li, Z.; Chen, F.; Ding, Z.; He, Y.; Fan, Q.-H.; Xiang, J.; Yu, Z.-X.; Chan, A. S. C. Highly Enantioselective Hydrogenation of Quinolines Using Phosphine-Free Chiral Cationic Ruthenium Catalysts: Scope, Mechanism, and Origin of Enantioselectivity. *J. Am. Chem. Soc.* **2011**, *133* (25), 9878–9891.
(23) Jean, A.; Blanchet, J.; Rouden, J.; Maddaluno, J.; De Paolis, M. Stereoselective Access to Heteroarylmethylene-Substituted Pyrrolidines: Fully Organocatalytic Mannich–Hydroamination Reactions. *Chem. Commun.* **2013**, *49* (16), 1651.
(24) Pang, Y.; He, Q.; Li, Z.-Q.; Yang, J.-M.; Yu, J.-H.; Zhu, S.-F.; Zhou, Q.-L. Rhodium-Catalyzed B–H Bond Insertion Reactions of Unstabilized Diazo Compounds Generated *in Situ* from Tosylhydrazones. *J. Am. Chem. Soc.* **2018**, *140* (34), 10663–10668.

(25) Ryschkewitsch, G. E.; Birnbaum, E. R. Amine Boranes. III. Propanolysis of Pyridine Boranes. *Inorg. Chem.* **1965**, *4* (4), 575–578.
(26) Taylor, N. P.; Gonzalez, J. A.; Nichol, G. S.; García-Domínguez, A.; Leach, A. G.; Lloyd-Jones, G. C. A Lewis Base Nucleofugality Parameter, N^{F}_{B}, and Its Application in an Analysis of MIDA-Boronate Hydrolysis Kinetics. *J. Org. Chem.* **2022**, *87* (1), 721–729.
(27) Yerri, J.; Dias, J.; Nimmakayala, M. R.; Razafindrainibe, F.; Courageux, C.; Gastellier, A.; Jegoux, J.; Coisne, C.; Landry, C.; Gosselet, F.; Hachani, J.; Goossens, J.; Dehouck, M.; Nachon, F.; Baati, R. Chemoselective Hydrogenation of 6-Alkynyl-3-fluoro-2-pyridinaldoximes: Access to First-in-Class 6-Alkyl-3-Fluoro-2-pyridinaldoxime Scaffolds as New Reactivators of Sarin-Inhibited Human Acetylcholinesterase with Increased Blood–Brain Barrier Permeability. *Chemistry A European J* **2020**, *26* (65), 15035–15044.

Appendix III – Supplementary Information for Chapter III

Portions of this appendix were adapted from the Supplementary Information of:

Smith, A. L.; Toste, F. D. *J. Am. Chem. Soc.* **2026**, *148*, 2792-2800. (DOI)

Table of Contents:

III.1 – Specific Comments on Computational Methods **235**
III.2 – Detailed Statistical Analyses **250**
III.3 – Regression Performance over Structurally Diverse Datasets **254**
III.4 – References **288**

VI.30 – Specific Comments on Computational Methods

- ***Specific Calculation Details***

Visualizations of *active site-based* buried volume were generated using SambVca 2.1.[56]

- ***Specific Statistical Methods***

In this report, we compiled 54 reactions from literature reports (ref. *dataset_rxn.xlsx*). To accommodate comparisons between *substituent*-based (corresponding to the aryl bromide surrogates, *dataset_substituent.csv*) and *active site*-based (*dataset_active-site.csv*) descriptors, **all datasets were restricted to 3,3'-aryl substituted BAOCs**. Literature reported transformations including heteroatom or alkyl substituents were discarded from the analysis. Only the data tabulated in *dataset_rxn.xlsx* were considered in the analyses reported.

The small nature of each reaction dataset (majority are 10 or fewer data points) will necessarily result in overfitting of models with too high dimensionality. To avoid this altogether, **we exclusively deploy univariate linear regression models to build correlations** between reported enantioselectivity (in kcal/mol) and the corresponding molecular representation. We do not aim to draw conclusions with respect to the specific atomic identity of each descriptor nor to the specific ensemble. Therefore, we pool our analyses drawn across sets of atoms and ensembles that are shared by descriptors (ex. buried volumes from 2-7 Å are individually considered when building models, but the best performing across all calculated buried volumes is recorded for the specific dataset). This results in a collection of 20 *unique sets* of descriptors where 5 are *active site*-based and 15 are *substituent*-based. The full results of univariate linear regressions drawn by from each unique set of descriptors is tabulated in the *dataset_scores*.xlsx supplementary file. The univariate linear regressions are scored with respect to the correlation drawn from the full dataset, including:

R^2, mean absolute error (MAE), p-value, and LOO-R^2. Most metrics were extracted using *sklearn's* built-in modules (ex. *sklearn.metrics.mean_absolute_error* and *sklearn.model_selection.KFold*).

- ***Molecular Descriptor Calculation Workflow***

  ◊ *Substituent-based descriptor generation – dataset_substituent.csv*

The specific calculation workflow to generate aryl bromide (substituent-based) descriptors operates by (1) producing an initial geometry from SMILES strings with RDKit's *AllChem* module (random seed = *0xf00d*), (2) optimization of the initial geometry with Grimme's xTB, (3) searching for accessible conformers with Grimme's CREST within a 1.8 kcal/mol window, (4) additional single-point energy calculations with Gaussian16 at the B3LYP/6-31+(d) level of theory using the following additional keywords (emp = gd3 nmr = giao pop=orbitals=1 pop=npa pop=(chelpg, readradii) pop=Hirshfeld scf=xqc) where the van der Waals radii for each element is implemented as in DBSTEP package by Luchini, G. Paton, R. et al. **2022.**,[58] and (5) extraction and export of ensemble-level descriptors for each aryl bromide into *csv* files. A numerical index is assigned to each processed SMILES string and tabulated in *dataset_substituent.csv*.

#1 – Generating initial conformations for each aryl bromide

The aryl bromide geometries were instantiated as described above. The selected SMILES strings were inspired by the reported aryl bromide dataset by Kariofillis, S. Doyle, A. et al. *J. Am. Chem. Soc.* **2022.**[59] and augmented with additional aryl bromides as needed from the collated datasets (ref. *dataset_rxn.xlsx*)

#5 – Extracting and export ensemble-level descriptors

The calculated representations for each aryl bromide are shown in *dataset_substituent.csv*. The ensembles computed for each aryl bromide are tagged at the end of each descriptor, including: (1) xTB-based ground state (written as 'C_0'), (2) DFT-based ground state (written as 'GD'), (3) xTB-based Boltzmann average (written as 'AV_XTB'), (4) DFT-based Boltzmann average (written as 'AV_DFT'), (5) minimum value across all conformers (written as 'MIN'), and (6) maximum value across all conformers (written as 'MAX').

◊ *Active site-based descriptor generation – dataset_active-site.csv*

The specific calculation workflow to generate BAOC (active site-based) descriptors operates by (1) producing an initial geometry from the preset 3D-configuration of the following chiral scaffolds, (2) optimization of the initial geometry with Grimme's xTB, (3) searching for accessible conformers with Grimme's CREST within a 1.8 kcal/mol window, and (4) extraction and export of ensemble-level descriptors for each BAOC into *csv* files, where **each molecule is labeled BAOC_NUM (BAOC – the scaffold used, NUM – the substituent analogous to the corresponding aryl bromide in *dataset_substituent.csv*).**

#1 – Generating initial conformations for each BAOC

The BAOC geometries were instantiated using an in-house Python script that substitutes a placeholder atom (labeled 'Bq') with the xTB-based ground state conformation of the corresponding aryl bromide to produce the desired 3,3'-substituted BAOC. It should be noted that instantiation by SMILES with RDKit's *AllChem* module does not conserve the axial chirality present in the chiral backbone of the BAOCs considered in this work.

The following coordinates (*templates*) were used as the initial geometries for each BAOC prior to substitution with the corresponding aryl bromide. Each *template* is labeled BAOC_H or BAOC_CB corresponding to the acid or conjugate base of the 'BAOC' scaffold.

CPA_H

C	0.88872103	0.21932295	-0.06941144	C	-1.75259838	0.44928712	5.90668080
C	0.84438745	-1.04602043	-0.56397060	C	-0.61070021	1.03556500	5.45993875
C	0.15635054	-2.05673797	0.12984923	H	-2.10526226	-2.21656835	2.20807746
C	-0.41251901	-1.79256914	1.37864309	C	-2.48717322	-1.94010194	3.16863953
C	-0.28528609	-0.51846150	1.93665607	C	-3.49669410	-1.22925601	5.66006057
C	0.25442852	0.51934442	1.14338424	H	-2.18183698	0.76055103	6.83607351
H	0.47488569	-3.55143998	-1.37579900	C	-4.08902661	-2.20949537	4.93147238
H	1.32972158	-1.27241451	-1.49031170	C	-3.57888180	-2.56879471	3.67246164
C	0.04129354	-3.33852158	-0.42104807	H	-3.89267319	-0.96039765	6.61700800
C	-1.08838666	-2.80835741	2.06365319	H	-4.95071379	-2.71207311	5.31853144
C	-1.19153804	-4.04337078	1.51153934	H	-4.05451921	-3.34294996	3.10737478
C	-0.62046008	-4.31131051	0.25600640	O	0.09825209	1.88279642	1.56881937
H	-1.52313530	-2.61148652	3.02131728	O	1.26734287	1.07578404	3.86519640
H	-1.71083123	-4.81831535	2.03566391	P	1.22400167	2.32929139	2.74037692
H	-0.70772198	-5.28923769	-0.16934927	O	0.77988403	3.66284025	3.40816851
C	-0.04547529	0.63091660	4.24351296	O	2.77592537	2.53955076	2.05377380
C	-0.74409242	-0.26246005	3.39973966	**Bq**	**0.16213058**	**2.30997111**	**6.41467517**
C	-1.86169898	-0.92950826	3.90723706	**Bq**	**1.74728441**	**1.48735299**	**-0.95703114**
C	-2.36772027	-0.57017082	5.15938874	H	3.44641870	2.33837582	2.71071165

CPA_CB

C	0.88872103	0.21932295	-0.06941144	C	-1.75259838	0.44928712	5.90668080
C	0.84438745	-1.04602043	-0.56397060	C	-0.61070021	1.03556500	5.45993875
C	0.15635054	-2.05673797	0.12984923	H	-2.10526226	-2.21656835	2.20807746
C	-0.41251901	-1.79256914	1.37864309	C	-2.48717322	-1.94010194	3.16863953
C	-0.28528609	-0.51846150	1.93665607	C	-3.49669410	-1.22925601	5.66006057
C	0.25442852	0.51934442	1.14338424	H	-2.18183698	0.76055103	6.83607351
H	0.47488569	-3.55143998	-1.37579900	C	-4.08902661	-2.20949537	4.93147238
H	1.32972158	-1.27241451	-1.49031170	C	-3.57888180	-2.56879471	3.67246164
C	0.04129354	-3.33852158	-0.42104807	H	-3.89267319	-0.96039765	6.61700800
C	-1.08838666	-2.80835741	2.06365319	H	-4.95071379	-2.71207311	5.31853144
C	-1.19153804	-4.04337078	1.51153934	H	-4.05451921	-3.34294996	3.10737478
C	-0.62046008	-4.31131051	0.25600640	O	0.09825209	1.88279642	1.56881937
H	-1.52313530	-2.61148652	3.02131728	O	1.26734287	1.07578404	3.86519640
H	-1.71083123	-4.81831535	2.03566391	P	1.22400167	2.32929139	2.74037692
H	-0.70772198	-5.28923769	-0.16934927	O	0.77988403	3.66284025	3.40816851
C	-0.04547529	0.63091660	4.24351296	O	2.77592537	2.53955076	2.05377380
C	-0.74409242	-0.26246005	3.39973966	**Bq**	**0.16213058**	**2.30997111**	**6.41467517**
C	-1.86169898	-0.92950826	3.90723706	**Bq**	**1.74728441**	**1.48735299**	**-0.95703114**
C	-2.36772027	-0.57017082	5.15938874				

H8CPA_H

C	1.05675548	-0.32764938	-0.20390985	C	-2.68839117	-0.72992458	6.51579344
C	0.76294549	-1.47926168	-0.82687915	H	-0.90661858	1.07051109	6.95692914
C	-0.15569110	-2.34191501	-0.16688326	C	-3.93030588	-1.19095171	5.80042639
C	-0.55989132	-2.18698660	1.14988397	C	-3.55306900	-2.33788876	4.94966386
C	-0.13598241	-0.99391124	1.85309706	H	-2.29920680	-1.54374743	7.08033733
C	0.51673754	-0.06302231	1.03233817	H	-4.65419109	-1.47692682	6.46226855
H	-0.26512416	-4.29669153	-0.25273780	H	-4.42444156	-2.72765415	4.52161626
H	1.17740304	-1.69952963	-1.79500815	O	0.64667776	1.30490660	1.47632517
C	-0.82954913	-3.47881900	-0.78116895	O	1.92796997	0.31573179	3.41333118
C	-1.49147072	-3.30669446	1.65652416	P	1.99806916	1.55701846	2.31533927
C	-2.64357698	-3.27016367	0.89572001	O	1.98615715	2.93854950	3.00813870
C	-2.37292832	-3.62145231	-0.54749713	O	3.35091009	1.47270992	1.28512306
H	-0.99251440	-4.33473126	1.73435166	**Bq**	**1.57736296**	**2.11804104**	**5.77599746**
H	-3.36436983	-3.99071962	1.28896536	**Bq**	**2.08034395**	**0.89643042**	**-1.03486345**
H	-3.10210194	-3.05527600	-1.24887752	H	4.14515012	1.45389442	1.78727855
C	0.59356571	0.20673592	4.01461605	H	-2.90030850	0.09953238	7.08895552
C	-0.41270430	-0.56364133	3.41023978	H	-4.27281971	-0.35456025	5.15285331
C	-1.58029944	-0.92787588	4.21886930	H	-2.98379683	-3.08094032	5.63985315
C	-1.63453958	-0.36359275	5.53994718	H	-3.46021300	-1.23841207	3.18164831
C	-0.70585928	0.59296811	6.00029896	H	-1.77180955	-3.11452194	2.58348113
C	0.37131349	0.88407670	5.25829992	H	-2.97010404	-2.25874398	1.04353975
H	-2.37093774	-2.58908791	3.18243573	H	-2.50880093	-4.70743073	-0.65395651
C	-2.73568605	-1.87037185	3.72636221	H	-0.65132573	-3.55524044	-1.87310972

H8CPA_CB

C	1.05675548	-0.32764938	-0.20390985	C	-2.68839117	-0.72992458	6.51579344
C	0.76294549	-1.47926168	-0.82687915	H	-0.90661858	1.07051109	6.95692914
C	-0.15569110	-2.34191501	-0.16688326	C	-3.93030588	-1.19095171	5.80042639
C	-0.55989132	-2.18698660	1.14988397	C	-3.55306900	-2.33788876	4.94966386
C	-0.13598241	-0.99391124	1.85309706	H	-2.29920680	-1.54374743	7.08033733
C	0.51673754	-0.06302231	1.03233817	H	-4.65419109	-1.47692682	6.46226855
H	-0.26512416	-4.29669153	-0.25273780	H	-4.42444156	-2.72765415	4.52161626
H	1.17740304	-1.69952963	-1.79500815	O	0.64667776	1.30490660	1.47632517
C	-0.82954913	-3.47881900	-0.78116895	O	1.92796997	0.31573179	3.41333118
C	-1.49147072	-3.30669446	1.65652416	P	1.99806916	1.55701846	2.31533927
C	-2.64357698	-3.27016367	0.89572001	O	1.98615715	2.93854950	3.00813870
C	-2.37292832	-3.62145231	-0.54749713	O	3.35091009	1.47270992	1.28512306
H	-0.99251440	-4.33473126	1.73435166	**Bq**	**1.57736296**	**2.11804104**	**5.77599746**
H	-3.36436983	-3.99071962	1.28896536	**Bq**	**2.08034395**	**0.89643042**	**-1.03486345**
H	-3.10210194	-3.05527600	-1.24887752	H	-2.90030850	0.09953238	7.08895552
C	0.59356571	0.20673592	4.01461605	H	-4.27281971	-0.35456025	5.15285331
C	-0.41270430	-0.56364133	3.41023978	H	-2.98379683	-3.08094032	5.63985315
C	-1.58029944	-0.92787588	4.21886930	H	-3.46021300	-1.23841207	3.18164831
C	-1.63453958	-0.36359275	5.53994718	H	-1.77180955	-3.11452194	2.58348113
C	-0.70585928	0.59296811	6.00029896	H	-2.97010404	-2.25874398	1.04353975
C	0.37131349	0.88407670	5.25829992	H	-2.50880093	-4.70743073	-0.65395651
H	-2.37093774	-2.58908791	3.18243573	H	-0.65132573	-3.55524044	-1.87310972
C	-2.73568605	-1.87037185	3.72636221				

SPA_H

C	0.86573410	-0.01036456	-0.71014664	C	-0.46933428	0.04093369	3.63126643
C	0.10487565	-1.24513412	-0.95545406	C	-1.03938635	-1.10514976	3.56563073
C	-0.70368703	-1.83206320	0.12093243	C	-1.36473563	-1.55441775	4.75331014
C	-0.49966716	-1.36687077	1.32933725	C	-1.27945639	-0.59823783	5.86475679
C	0.21141474	-0.30054358	1.31385474	C	-0.58290812	0.68360394	5.67649387
C	0.74464034	0.60910081	0.59005943	H	-1.70642798	-0.84133397	6.81528271
H	0.12550678	-1.70398870	-1.92185296	O	1.00056279	1.98567139	0.98967822
C	0.03971552	0.93081201	4.39580632	O	1.11349408	1.85057309	4.04785739

P	1.48985157	2.66760391	2.53593917	C	-1.69690959	-3.05379402	4.49543800
O	0.82296117	4.07281443	2.58163134	H	-1.43625640	-3.71293808	5.29699212
O	3.19476636	2.79890953	2.52478642	H	-2.73934275	-3.18352095	4.29195820
Bq	**-0.46758848**	**1.86190885**	**6.99224420**	C	-2.49231898	-2.24401820	1.63288756
Bq	**1.85607529**	**0.71982044**	**-1.98252796**	H	-3.21844556	-2.85178577	2.13114329
H	3.48662565	3.24452592	3.32343791	H	-2.95887781	-1.31420646	1.38254181
C	-1.84336661	-2.87421179	0.32109299	C	-0.80166238	-3.25849468	3.19324390
H	-1.44081348	-3.84940355	0.49956435	H	-0.95724873	-4.17324920	2.66040523
H	-2.52879775	-2.94870743	-0.49715792	H	0.23187620	-3.20924589	3.4657738
C	-1.20772958	-1.99268820	2.43030199				

SPA_CB

C	0.86573410	-0.01036456	-0.71014664	O	0.82296117	4.07281443	2.58163134
C	0.10487565	-1.24513412	-0.95545406	O	3.19476636	2.79890953	2.52478642
C	-0.70368703	-1.83206320	0.12093243	**Bq**	**-0.46758848**	**1.86190885**	**6.99224420**
C	-0.49966716	-1.36687077	1.32933725	**Bq**	**1.85607529**	**0.71982044**	**-1.98252796**
C	0.21141474	-0.30054358	1.31385474	C	-1.84336661	-2.87421179	0.32109299
C	0.74464034	0.60910081	0.59005943	H	-1.44081348	-3.84940355	0.49956435
H	0.12550678	-1.70398870	-1.92185296	H	-2.52879775	-2.94870743	-0.49715792
C	0.03971552	0.93081201	4.39580632	C	-1.20772958	-1.99268820	2.43030199
C	-0.46933428	0.04093369	3.63126643	C	-1.69690959	-3.05379402	4.49543800
C	-1.03938635	-1.10514976	3.56563073	H	-1.43625640	-3.71293808	5.29699212
C	-1.36473563	-1.55441775	4.75331014	H	-2.73934275	-3.18352095	4.29195820
C	-1.27945639	-0.59823783	5.86475679	C	-2.49231898	-2.24401820	1.63288756
C	-0.58290812	0.68360394	5.67649387	H	-3.21844556	-2.85178577	2.13114329
H	-1.70642798	-0.84133397	6.81528271	H	-2.95887781	-1.31420646	1.38254181
O	1.00056279	1.98567139	0.98967822	C	-0.80166238	-3.25849468	3.19324390
O	1.11349408	1.85057309	4.04785739	H	-0.95724873	-4.17324920	2.66040523
P	1.48985157	2.66760391	2.53593917	H	0.23187620	-3.20924589	3.46577384

NTP_H

C	3.98429469	3.00088863	0.02996965	C	5.27812030	-3.10743376	-3.86549947
C	4.92294732	3.74336534	-0.61411701	H	4.39868379	-4.36506860	-1.69401930
C	5.74901175	3.15553864	-1.58786263	C	5.68776518	-2.33380478	-4.90291544
C	5.67506601	1.78283145	-1.83963015	C	5.82105970	-0.94378380	-4.74615370
C	4.77908089	0.99719130	-1.11069723	H	5.17736564	-4.16492239	-3.99382016
C	3.84274276	1.63941460	-0.26881273	H	5.91059779	-2.78566283	-5.84688046
H	6.72301518	4.99077212	-2.12227409	H	6.14461231	-0.34435797	-5.57132186
H	5.02882660	4.78292326	-0.38388050	O	2.72051327	0.89500778	0.23166492
C	6.65782195	3.93915065	-2.30866663	O	4.56565474	-0.83044984	1.18724495
C	6.50824289	1.20736787	-2.80530576	P	3.11906672	-0.05417778	1.56517456
C	7.38076394	1.98435638	-3.49519542	O	1.98444905	-1.08594639	1.82888527
C	7.45632009	3.36485120	-3.24414698	S	3.65711276	0.07530743	4.32612366
H	6.45742697	0.15633404	-2.99933065	O	2.92805417	-1.19678200	4.42605519
H	8.01551904	1.54304019	-4.23494272	O	5.10110833	-0.17840163	4.42727976
H	8.14838613	3.96765691	-3.79421054	C	3.15039185	1.15497660	5.64746704
C	4.56003577	-1.39062882	-0.13570118	F	1.82391251	1.38803818	5.55454320
C	4.81765057	-0.55141710	-1.24354041	F	3.82011772	2.32353995	5.55566832
C	5.12145400	-1.13910605	-2.47386078	F	3.42283451	0.57220186	6.83433207
C	4.98756654	-2.52170290	-2.62774351	N	3.31201863	0.81348873	2.82276129
C	4.55793840	-3.31670433	-1.55094468	**Bq**	**2.94113477**	**3.75364254**	**1.24573233**
C	4.35257423	-2.76297714	-0.32685332	**Bq**	**3.82325063**	**-3.77517258**	**1.02524760**
H	5.64275208	0.70247995	-3.44303408	H	2.00896555	-1.63580194	2.61543331
C	5.54178973	-0.35682744	-3.55516990				

NTP_CB

C	4.79198602	1.88538068	-0.85935804	C	7.91470108	-2.20345228	-0.57062066
C	5.95704748	2.43169448	-1.29680640	C	7.75767237	-4.66288577	0.71708661
C	7.15592486	1.70122679	-1.22460411	H	5.79602725	-4.82859741	2.50450603
C	7.17843083	0.44124273	-0.62035785	C	8.69252617	-4.44708920	-0.24313359
C	6.00164625	-0.08415867	-0.08138252	C	8.77195628	-3.20423073	-0.89389750
C	4.78358441	0.59068031	-0.32400792	H	7.70269731	-5.61285206	1.20641442
H	8.33244965	3.19735011	-2.21443472	H	9.37373093	-5.22906330	-0.50652797
H	5.96047500	3.42069424	-1.70517837	H	9.51302949	-3.04612712	-1.64935356
C	8.33900423	2.23289382	-1.75109774	O	3.54200852	-0.08426296	-0.06524292
C	8.37937863	-0.27302717	-0.54677580	O	4.50947008	-0.45112615	2.42961455
C	9.51499157	0.25844100	-1.06530903	P	3.11906672	-0.05417778	1.56517456
C	9.49452078	1.52482563	-1.67387367	O	1.98444905	-1.08594639	1.82888527
H	8.40216332	-1.23669485	-0.08231321	S	2.08109920	1.61756020	3.57937798
H	10.43201509	-0.29013594	-1.01025181	O	1.36674041	0.42664151	4.06013618
H	10.39630494	1.93401388	-2.07915178	O	3.22962849	1.88691216	4.45576492
C	5.19989185	-1.58104496	1.87228181	C	0.98077988	3.01663368	3.59642882
C	6.04764808	-1.39269553	0.75703146	F	-0.07427896	2.76920229	2.79136452
C	6.94677856	-2.40696428	0.41911679	F	1.63700219	4.11063225	3.15479612
C	6.86557130	-3.64252438	1.06717268	F	0.54510454	3.23116153	4.85605765
C	5.89180661	-3.86134602	2.05712356	N	2.63295455	1.34582487	1.98384811
C	5.07745123	-2.84888973	2.45580324	**Bq**	**3.28594852**	**2.80917161**	**-0.96594761**
H	7.97868146	-1.26021735	-1.07173379	**Bq**	**3.86272420**	**-3.13195460**	**3.71167446**

IDP_H

C	-3.17854294	7.95331558	-1.94949586	C	2.14799759	-1.62706065	1.88193523
C	-1.88642624	7.61978754	-2.18593766	C	2.12655088	-3.31712472	-0.72621603
C	-1.40628911	6.36669095	-1.79136500	C	2.32098782	-3.03116440	1.83719744
C	-2.27049844	5.46234594	-1.16616982	C	1.58990157	-1.02599590	3.01526710
C	-3.59865234	5.82135649	-0.93557283	C	2.67139001	-3.75161899	0.50447285
C	-4.04457022	7.04564678	-1.31594688	C	2.63701321	-3.81145498	-1.93119052
H	0.56811569	6.69448921	-2.56112228	C	2.18271035	-3.76634285	3.02137171
H	-3.54123945	8.91480066	-2.24865270	C	1.36849512	-1.76361130	4.13615588
H	-1.23348840	8.31349842	-2.67402766	C	3.53314564	-4.85296955	0.50786516
C	-0.06912299	6.01975624	-2.03073786	C	3.54305659	-4.82415835	-1.92378426
C	-1.79892347	4.20095130	-0.79467246	C	2.49948744	-5.13073741	3.06417512
H	-4.26336481	5.13226142	-0.45997432	C	1.70595726	-3.12879531	4.17169018
H	-5.06241408	7.32076179	-1.13280032	H	0.94077777	-1.29896460	4.99982352
C	-0.41366849	3.94809302	-0.89387509	C	3.97348874	-5.38531889	-0.70751214
C	0.42136733	4.83592104	-1.58444590	C	3.97397655	-5.41547126	1.71228329
C	-2.78866946	3.11320532	-0.27596171	H	3.92755953	-5.19979394	-2.84953120
C	-3.89045638	3.47689427	0.50570894	C	2.35081105	-5.82572142	4.21981125
C	-2.60452084	1.74473004	-0.58937833	H	2.85945844	-5.62587640	2.18590147
C	-4.20885396	4.82499955	0.71420991	C	1.56145887	-3.86485302	5.35524127
C	-4.67367959	2.48115935	1.09887790	C	4.85348977	-6.47542764	-0.70434512
C	-3.34577540	0.76079882	0.08100478	C	4.82649707	-6.47208739	1.69393072
C	-5.27803789	5.15854735	1.48035499	H	3.64028863	-5.01221540	2.64508559
H	-3.60886349	5.59081730	0.27037325	C	1.87858647	-5.18438503	5.37807142
C	-5.77246706	2.84806706	1.88636749	H	2.59344859	-6.86822948	4.24948892
C	-4.35732529	1.12286196	0.91201998	H	1.20130658	-3.38735105	6.24273697
C	-6.06953872	4.15929408	2.07183926	C	5.27153520	-7.00725103	0.47327987
H	-5.52068019	6.18945978	1.63561929	H	5.19746589	-6.88926706	-1.62950788
H	-6.37909348	2.09483080	2.34261068	H	5.16228842	-6.89903359	2.61560317
H	-4.92239703	0.37138505	1.42152797	H	1.76818520	-5.74064318	6.28493128
H	-6.91264012	4.43281157	2.67214155	H	5.94538942	-7.83960016	0.47136302
O	-1.69156509	1.38282233	-1.63660430	N	0.26475667	0.12788300	-0.23226063
O	0.13816418	2.80021397	-0.23884333	O	0.81516711	1.38520907	-2.34660952
P	-0.08051363	1.39359634	-1.13668075	O	2.22040513	-0.30268788	-1.75437799
P	1.52201532	-0.81135435	-0.52187867	**Bq**	**2.12078803**	**4.42446805**	**-1.85795242**
O	2.60177182	-0.82339202	0.77564207	**Bq**	**-2.97704031**	**-0.95103567**	**-0.17428909**
O	1.01821389	-2.40328222	-0.73087447	**Bq**	**2.08478709**	**-3.12617068**	**-3.46672729**

Bq	**1.19528054**	**0.69952595**	**2.99261676**	H	-0.32761974	-0.08551477	0.54462519

IDP_CB

C	-3.17854294	7.95331558	-1.94949586	C	1.87858647	-5.18438503	5.37807142
C	-1.88642624	7.61978754	-2.18593766	H	2.59344859	-6.86822948	4.24948892
C	-1.40628911	6.36669095	-1.79136500	H	1.20130658	-3.38735105	6.24273697
C	-2.27049844	5.46234594	-1.16616982	C	5.27153520	-7.00725103	0.47327987
C	-3.59865234	5.82135649	-0.93557283	H	5.19746589	-6.88926706	-1.62950788
C	-4.04457022	7.04564678	-1.31594688	H	5.16228842	-6.89903359	2.61560317
H	0.56811569	6.69448921	-2.56112228	H	1.76818520	-5.74064318	6.28493128
H	-3.54123945	8.91480066	-2.24865270	H	5.94538942	-7.83960016	0.47136302
H	-1.23348840	8.31349842	-2.67402766	N	0.26475667	0.12788300	-0.23226063
C	-0.06912299	6.01975624	-2.03073786	O	0.81516711	1.38520907	-2.34660952
C	-1.79892347	4.20095130	-0.79467246	O	2.22040513	-0.30268788	-1.75437799
H	-4.26336481	5.13226142	-0.45997432	**Bq**	**2.12078803**	**4.42446805**	**-1.85795242**
H	-5.06241408	7.32076179	-1.13280032	**Bq**	**-2.97704031**	**-0.95103567**	**-0.17428909**
C	-0.41366849	3.94809302	-0.89387509	**Bq**	**2.08478709**	**-3.12617068**	**-3.46672729**
C	0.42136733	4.83592104	-1.58444590	**Bq**	**1.19528054**	**0.69952595**	**2.9926167**
C	-2.78866946	3.11320532	-0.27596171				
C	-3.89045638	3.47689427	0.50570894				
C	-2.60452084	1.74473004	-0.58937833				
C	-4.20885396	4.82499955	0.71420991				
C	-4.67367959	2.48115935	1.09887790				
C	-3.34577540	0.76079882	0.08100478				
C	-5.27803789	5.15854735	1.48035499				
H	-3.60886349	5.59081730	0.27037325				
C	-5.77246706	2.84806706	1.88636749				
C	-4.35732529	1.12286196	0.91201998				
C	-6.06953872	4.15929408	2.07183926				
H	-5.52068019	6.18945978	1.63561929				
H	-6.37909348	2.09483080	2.34261068				
H	-4.92239703	0.37138505	1.42152797				
H	-6.91264012	4.43281157	2.67214155				
O	-1.69156509	1.38282233	-1.63660430				
O	0.13816418	2.80021397	-0.23884333				
P	-0.08051363	1.39359634	-1.13668075				
P	1.52201532	-0.81135435	-0.52187867				
O	2.60177182	-0.82339202	0.77564207				
O	1.01821389	-2.40328222	-0.73087447				
C	2.14799759	-1.62706065	1.88193523				
C	2.12655088	-3.31712472	-0.72621603				
C	2.32098782	-3.03116440	1.83719744				
C	1.58990157	-1.02599590	3.01526710				
C	2.67139001	-3.75161899	0.50447285				
C	2.63701321	-3.81145498	-1.93119052				
C	2.18271035	-3.76634285	3.02137171				
C	1.36849512	-1.76361130	4.13615588				
C	3.53314564	-4.85296955	0.50786516				
C	3.54305659	-4.82415835	-1.92378426				
C	2.49948744	-5.13073741	3.06417512				
C	1.70595726	-3.12879531	4.17169018				
H	0.94077777	-1.29896460	4.99982352				
C	3.97348874	-5.38531889	-0.70751214				
C	3.97397655	-5.41547126	1.71228329				
H	3.92755953	-5.19979394	-2.84953120				
C	2.35081105	-5.82572142	4.21981125				
H	2.85945844	-5.62587640	2.18590147				
C	1.56145887	-3.86485302	5.35524127				
C	4.85348977	-6.47542764	-0.70434512				
C	4.82649707	-6.47208739	1.69393072				
H	3.64028863	-5.01221540	2.64508559				

#4 – Extracting and export ensemble-level descriptors

The calculated representations for each BAOC are shown in *dataset_active-site.csv*. The buried volume representation is reported across the indicated distances at either '1O' or '2H', corresponding to either the heteroatom ('O' for CPA, H8CPA, and SPA; 'N' for IDP, NTP) and the acidic proton ('H') respectively.

Unique to the BAOC dataset is *calc.* pK_a for each ensemble generated. The *calc.* pK_a of each BAOC is computed from an xTB-based benchmark of the corresponding ensemble-level *deprotonation energy* (ΔG_{deprot}, in Hartrees) between calculated acid ('H') and conjugate base ('CB') of a selected array of experimentally determined acidities in MeCN. The xTB-based benchmark was validated with a higher-level DFT-based benchmark and an experimental ground truth (observed acidity in acetonitrile from Kutt, A. et al. *Eur. J. Org. Chem.* **2021**[60]). Each BAOC computed was organized into one of two bins relating to the general characteristics of the acidic proton: (1) phosphate-like, or (2) sulfonamide-like. In this work, phosphate-like acids include CPA, H8CPA, SPA, and IDP. The sulfonamide-like acids include NTP.

Phosphate-like calc. pK_a *conversion*

Ground state ('C_0') calc. $p\mathbf{K_a} = -0.142830176 * \Delta G_{deprot} - 9.073037818$

Boltzmann avg. ('AV_XTB') calc. $p\mathbf{K_a} = -0.143257524 * \Delta G_{deprot} - 9.136612018$

Minimum ('MIN') calc. $p\mathbf{K_a} = -0.142830176 * \Delta G_{deprot} - 9.073037818$

Maximum ('MAX') calc. $p\mathbf{K_a} = -0.096629251 * \Delta G_{deprot} - 1.88123327$

Sulfonamide-like calc. pK_a *conversion*

Ground state ('C_0') calc. $p\mathbf{K_a} = -0.24500443 * \Delta G_{deprot} - 26.55425678$

Boltzmann avg. ('AV_XTB') calc. $p\mathbf{K_a} = -0.245285358 * \Delta G_{deprot} - 26.60609466$

Minimum ('MIN') calc. $p\mathbf{K_a} = -0.24500443 * \Delta G_{deprot} - 26.55425678$

Maximum ('MAX') calc. $p\mathbf{K_a} = -0.24275215 * \Delta G_{deprot} - 26.39632654$

Altogether, the ensembles computed for each BAOC are tagged at the end of each descriptor, including: (1) xTB-based ground state (written as 'C_0'), (2) xTB-based Boltzmann average (written as 'AV_XTB'), (3) minimum value across all conformers (written as 'MIN'), and (4) maximum value across all conformers (written as 'MAX'). It should be noted that DFT-based ensembles (like 'GD' and 'AV_DFT') are not computed for the BAOCs because of the excessive computational cost of even single-point energies on the large molecules. xTB-based descriptors are solely extracted and deployed in the reported correlations herein.

- ***Visualizations of Buried Volume with SambVca 2.1***

Throughout the main text, *active site-based* buried volume was visualized with SambVca 2.1[56] to contextualize changes in confinement at the active site as it reflects the observed stereoselectivity. As the buried volumes are computed from ensembles of chiral acids, we instead use the ground state (lowest energy from xTB single point after conformational searching, *vida supra*) conformation identified for visualization. The coordinates for these ground state conformers are printed below corresponding to their label in the main text:

TRIP-**CPA**

O	-0.7540616335	1.9108734094	-0.9706461343	C	3.9374362949	0.8680360386	2.3668841196
H	-0.7162905563	2.8492358236	-0.7914107526	C	3.2797579958	2.0433646796	3.0914824315
C	-2.6258202481	-3.0724675463	-1.2528755066	C	5.2677394949	0.4979963160	3.0307238593
C	-1.2378920936	-3.0625392624	-0.9563457865	H	7.8318755977	2.9167399086	-0.9161336908
C	-0.7057181199	-1.9855836146	-0.1893266030	H	7.8924302804	4.6826995569	-0.8961798571
C	-1.5227464830	-0.9249937587	0.1282095721	H	7.3025522051	3.8027689494	0.5137543317
H	-4.2271304059	-4.1584612946	-2.1834192058	H	5.9758694317	3.9169940288	-2.2341619833
H	-4.4842596240	-2.0046402975	-1.0813509619	H	4.1638194157	5.1543078608	-1.0644572508
C	-3.1662175347	-4.1548349387	-1.9786021032	H	5.1044202446	5.1395866762	0.4251959943
C	-0.4332980277	-4.1024382363	-1.4674380753	H	5.7037114482	6.0189870099	-0.9824271730
C	-0.9849271273	-5.1281202496	-2.1842763354	H	4.8474423222	2.0192512479	-2.7951482640
C	-2.3654880657	-5.1666219349	-2.4276751097	H	2.9241732410	-1.1187910935	-2.1394364720
H	0.6318047117	-4.0761156823	-1.2965194425	H	1.6992654303	0.8407118635	-3.0248244910
H	-0.3553683528	-5.9153224071	-2.5717668469	H	2.2139253997	-0.2504265850	-4.3158031143
H	-2.7855937496	-5.9907740115	-2.9849590284	H	3.0800249549	1.2619766113	-4.0425972047
C	1.5013857968	-0.8945257620	-0.0618748726	H	4.3889855720	-1.5693278999	-4.0534003102
C	0.7061403485	-1.9771021223	0.2409501314	H	5.2855645636	-0.0775733739	-3.7653343986
C	1.2663601934	-3.0643663487	0.9716709947	H	5.3478317790	-1.3831460062	-2.5813592879
C	2.6575743465	-3.0564286867	1.2525872640	H	3.2614518170	0.0125301062	2.4531756482
C	3.4405433477	-1.9566097697	0.8498258048	H	3.0570949173	1.7643911214	4.1194300887
C	2.8811854660	-0.8634960839	0.2387384659	H	3.9297612928	2.9144153582	3.1096635859
H	-0.5824837963	-4.1169912227	1.3071296354	H	2.3516133347	2.3075399733	2.5934956040
C	0.4845068783	-4.1282754336	1.4685708746	H	5.7323920879	-0.3435568818	2.5210985160
C	3.2226931472	-4.1421025968	1.9542174148	H	5.1050598930	0.2247395063	4.0711589647
H	4.5006202013	-1.9659953557	1.0560077785	H	5.9597457492	1.3363482043	2.9997897139
C	2.4433587412	-5.1749893497	2.3927805154	H	5.2558662790	2.9445951096	1.3285878062
C	1.0602501008	-5.1570909495	2.1613526989	C	-3.6916799425	0.3084609663	0.1031947684
H	4.2853536914	-4.1313211775	2.1490180409	C	-3.9578385184	0.6477003497	1.4328544405
H	2.8825218416	-6.0012115913	2.9320067554	C	-4.6585981455	1.8143191626	1.7019021151
H	0.4481954314	-5.9625016753	2.5394084641	C	-5.1007928051	2.6497815023	0.6900853099
O	-1.0177795420	0.1409251039	0.8562449147	C	-4.8479857592	2.2818105586	-0.6220664563
O	0.9497441820	0.1979129071	-0.7185718475	C	-4.1516086599	1.1252195503	-0.9375604392
P	0.0226788077	1.1404074409	0.1840277506	C	-7.2542830580	3.9011579631	0.4182810672
O	0.6532972526	2.0113178216	1.1649442801	C	-5.8456175348	3.9216958431	1.0150666283
C	-2.8945568863	-0.8982811208	-0.2112667398	C	-5.0650906158	5.1432674836	0.5250909445
C	-3.4292227010	-1.9871016216	-0.8508262991	C	-3.5620228183	-0.2578168227	2.5767958622
C	3.6737141785	0.3366051177	-0.1107487811	C	-2.6929511339	0.4753614269	3.5997566208
C	3.9262825680	0.6369561002	-1.4520335734	C	-4.8169829648	-0.8346967939	3.2394055676
C	4.6476946453	1.7809761691	-1.7602493599	C	-3.9413855062	0.7714132468	-2.3938983693
C	5.1259417698	2.6284096722	-0.7749901504	C	-3.2524507766	1.8977350686	-3.1664200229
C	4.8792883633	2.3013994601	0.5488434296	C	-5.2862147014	0.4233198173	-3.0399461962
C	4.1588152523	1.1711524828	0.9023686524	H	-7.8039506780	4.7908653028	0.7182829012
C	7.3218334851	3.8143808545	-0.5730282204	H	-7.2149487487	3.8740131912	-0.6677864003
C	5.9023288697	3.8700126131	-1.1414445525	H	-7.7983377179	3.0245970059	0.7633813311
C	5.1729245977	5.1255598618	-0.6594616600	H	-5.9393938792	3.9930853566	2.1045489723
C	3.4839319516	-0.2824853008	-2.5684644372	H	-4.9689100302	5.1314369792	-0.5578858719
C	2.5625618924	0.4415457220	-3.5514649704	H	-5.5765244245	6.0590723592	0.8140286324
C	4.7046949777	-0.8629525262	-3.2882953319	H	-4.0670631772	5.1517226726	0.9579570628

H	-4.8694424559	2.0809452980	2.7271422741	H	-3.2949892267	-0.1080945610	-2.4574228383
H	-2.9808223312	-1.0960635350	2.1821678414	H	-3.0907117502	1.5905899001	-4.1977828599
H	-2.3630525222	-0.2179046771	4.3708754231	H	-3.8531037888	2.8037051503	-3.1734612237
H	-3.2440778419	1.2805666757	4.0792011640	H	-2.2886917966	2.1199089586	-2.7174810870
H	-1.8157014356	0.8955019078	3.1138517360	H	-5.9473018740	1.2867110210	-3.0438819121
H	-4.5384581742	-1.5248916575	4.0329257208	H	-5.7792287882	-0.3773678832	-2.4930984967
H	-5.4267562486	-0.0442304497	3.6700284813	H	-5.1375107723	0.0992808700	-4.0678465872
H	-5.4185918035	-1.3719601688	2.5088289461	H	-5.2046585083	2.9116461398	-1.4223582060

TRIP-**H$_8$CPA**

O	-0.7816535231	1.9713876825	-0.9542249497	C	7.4836828150	3.7170209284	-0.6046609816
H	-0.7416644414	2.9082309604	-0.7676000783	C	6.0751054676	3.8059191833	-1.1954876428
C	-2.5239251641	-2.9317259397	-1.3896780840	C	5.3836835277	5.0989607674	-0.7586925831
C	-1.1727567527	-2.9146714372	-1.0528644209	C	3.5250771407	-0.2894396469	-2.5507768149
C	-0.6934565569	-1.8977035916	-0.2153929690	C	2.6137067931	0.4394902767	-3.5396358292
C	-1.5294256511	-0.8446685672	0.1340181276	C	4.7308242269	-0.9037416575	-3.2679449819
H	-3.9035100561	-3.6575364757	-2.8483596591	C	3.9540820882	0.9726021560	2.3613076149
H	-4.4043095063	-1.9345625400	-1.2096687871	C	3.3112157212	2.1846423076	3.0369063014
C	-3.1283882400	-4.0555759540	-2.1899570156	C	5.2627221097	0.5919328269	3.0613432950
C	-0.2167112066	-3.9017858471	-1.6700395803	H	7.4486301495	3.7352965784	0.4816474858
C	-0.9124524417	-5.1842325661	-2.1100118064	H	7.9669980742	2.7934286920	-0.9161413351
C	-2.0969230136	-4.8320655946	-2.9997814989	H	8.0886889290	4.5562527466	-0.9419577148
H	0.2211650070	-3.4166188850	-2.5502161214	H	6.1655601417	3.8202079990	-2.2878965031
H	-0.2061416539	-5.8198525344	-2.6473805728	H	5.9500623651	5.9643112874	-1.0974244002
H	-2.5540157030	-5.7352038629	-3.4087951789	H	4.3820732880	5.1515380508	-1.1795317736
C	1.5125303663	-0.8069888235	-0.0642796899	H	5.3003270034	5.1448842025	0.3240632548
C	0.7043287464	-1.8866597402	0.2733571348	H	4.9732735991	1.9526603607	-2.8145738319
C	1.2199780819	-2.9170909509	1.0711859541	H	2.9515933359	-1.1082130321	-2.1066238105
C	2.5769516498	-2.9140278746	1.3842091178	H	2.2594641813	-0.2514525498	-4.3025590957
C	3.3823553941	-1.8714101021	0.9443353391	H	3.1416753157	1.2522410801	-4.0324502035
C	2.8647989893	-0.7785238847	0.2636969637	H	1.7541167572	0.8503455381	-3.0162023621
H	-0.5336326665	-4.1663796753	1.0039051243	H	4.3979107165	-1.6214423434	-4.0152126876
C	0.2944834102	-3.9412576593	1.6746832395	H	5.3200618710	-0.1388557032	-3.7673039058
C	3.2166539654	-4.0393576121	2.1544970666	H	5.3715923434	-1.4186932219	-2.5551041660
H	4.4381077387	-1.8866441987	1.1722740828	H	3.2554634984	0.1371128309	2.4591838786
C	2.2128326316	-4.8595589823	2.9560253020	H	3.9733107280	3.0466392036	3.0231737015
C	1.0261606357	-5.2177633735	2.0715288797	H	2.3884629512	2.4408964417	2.5250901576
H	3.7215562687	-4.6938872615	1.4352200963	H	3.0805599883	1.9495471631	4.0740855454
H	2.6955823157	-5.7614155045	3.3374796802	H	5.7189126027	-0.2727407640	2.5834748150
H	0.3424429959	-5.8844709503	2.6004071899	H	5.0740420374	0.3506795625	4.1054200508
O	-1.0119638834	0.2058542261	0.8817952699	H	5.9737990986	1.4137994460	3.0226893083
O	0.9411418205	0.2708439942	-0.7360890752	H	5.3648175775	2.9685033508	1.2889531486
P	0.0192468983	1.2016803067	0.1866896314	C	-3.7199645341	0.3428420019	0.0952535187
O	0.6641855634	2.0769997000	1.1555863542	C	-4.0218946666	0.6312944732	1.4296707418
C	-2.8709689993	-0.8219550430	-0.2389386185	C	-4.7755594445	1.7591032991	1.7208941072
C	-3.3551913941	-1.9080599719	-0.9537005072	C	-5.2403929153	2.6037134375	0.7272102875
H	3.9862397064	-3.6348421199	2.8152746334	C	-4.9528901940	2.2851937957	-0.5906041705
H	1.8615421193	-4.2784266050	3.8110730332	C	-4.2006091769	1.1704827831	-0.9275489055
H	1.3794957835	-5.7372641401	1.1776437441	C	-7.4516431387	3.7504611048	0.4707992161
H	-0.1424430200	-3.4899009446	2.5730071814	C	-6.0489232200	3.8294243447	1.0768914057
H	0.6079752826	-4.1257813811	-0.9951470001	C	-5.3303310210	5.1008456569	0.6204883181
H	-1.2635232944	-5.7357774768	-1.2346890690	C	-3.6077459865	-0.2878288172	2.5563549870
H	-1.7498683383	-4.2220763639	-3.8363698107	C	-2.7573536538	0.4467806686	3.5939644109
H	-3.6211807519	-4.7386781259	-1.4890848600	C	-4.8491920692	-0.9072170705	3.2061263496
C	3.7039294671	0.3831452345	-0.1043335626	C	-3.9539525047	0.8706401259	-2.3903479844
C	3.9866227447	0.6387366716	-1.4492663345	C	-3.2996808466	2.0467027054	-3.1173747007
C	4.7528563720	1.7479207788	-1.7765823059	C	-5.2734453402	0.4894476152	-3.0694445035
C	5.2499125528	2.6028476071	-0.8074261515	H	-8.0467235260	4.6060735166	0.7832440647
C	4.9745956167	2.3193620304	0.5207273654	H	-7.4043119754	3.7452028910	-0.6153087154
C	4.2076606766	1.2261898825	0.8926030087	H	-7.9533108029	2.8415639217	0.7960019845

H	-6.1518879154	3.8702905773	2.1671792673	H	-4.5551493375	-1.6123276305	3.9808783238
H	-5.2327244466	5.1222808187	-0.4621779019	H	-5.4747665301	-0.1419940877	3.6588982253
H	-5.8873577833	5.9820921564	0.9319745292	H	-3.2705019042	0.0210822975	-2.4677188056
H	-4.3343137631	5.1480164421	1.0554007435	H	-3.9271258670	2.9338708025	-3.0897051847
H	-5.0116317630	1.9869774157	2.7501369267	H	-2.3434539082	2.2793044101	-2.6581330777
H	-3.0065472312	-1.1038135824	2.1458046825	H	-3.1280497642	1.7854380161	-4.1598032224
H	-1.8815859959	0.8816178565	3.1187405599	H	-5.9659843683	1.3277949268	-3.0641863785
H	-2.4256825309	-0.2485317704	4.3626980360	H	-5.7462765187	-0.3420699570	-2.5513784851
H	-3.3228722536	1.2412267139	4.0745252095	H	-5.0957162087	0.1975648935	-4.1025112529
H	-5.4408283439	-1.4373213012	2.4623410621	H	-5.3264479419	2.9214976536	-1.3781075177

TRIP-**SPA**

O	0.0106838280	-2.7176960361	0.3754401899	H	-8.7390925443	-1.7057461178	-0.3662776174
H	-0.0184622542	-2.9524349232	-0.5520728821	H	-9.1656817633	-3.4081448329	-0.1600471139
C	2.3761614115	3.3022424442	0.2910511844	H	-8.2663097918	-2.5628485275	1.1002208736
C	1.0930958299	2.9259841646	-0.0477522122	H	-7.2730216615	-3.1860717694	-1.7192045162
C	0.7004833168	1.5905825158	-0.0738576584	H	-6.4298974123	-4.3629377317	0.9698111114
C	1.6761589118	0.6125447854	-0.1675572201	H	-7.3307175920	-5.1926564804	-0.2992977919
H	4.3932552436	2.5499460454	0.4261088842	H	-5.6576163563	-4.6968670812	-0.5792148416
C	-1.7017948634	0.6022088928	0.0514893821	H	-5.8268978609	-1.6382829143	-2.5510416990
C	-0.7193551053	1.5645138419	0.2107822338	H	-3.3335635603	1.1459107448	-2.3292076092
C	-1.1031306566	2.8723237493	0.4895597180	H	-2.5054983713	-0.2431827492	-4.2002803955
C	-2.3797743755	3.3270072328	0.2293051037	H	-3.5254067238	-1.5868116333	-3.6722787618
C	-3.3671239067	2.3729231054	0.0503953902	H	-2.2082976067	-1.0237593877	-2.6410146861
C	-3.0377795577	1.0045999434	0.0343376246	H	-4.5288874083	1.3543121524	-4.4254110192
H	-4.3955687189	2.6496807183	-0.1297218813	H	-5.5431295539	-0.0393779915	-4.0694044609
O	1.2989056859	-0.7190683220	-0.2550587515	H	-5.7017706869	1.4307948509	-3.1070896677
O	-1.3156553866	-0.7099941764	-0.1786927193	H	-3.4495324595	0.5255601102	2.2984408154
P	0.0175363683	-1.1667492522	0.6333612307	H	-2.5118430537	-1.7376428154	2.3767516471
O	0.0561322503	-0.7343909127	2.0199893344	H	-3.0450303820	-1.2447958580	3.9799535596
C	3.0166751034	0.9983422099	-0.0879680933	H	-4.0197707393	-2.3795374737	3.0392164191
C	3.3584197321	2.3298568210	0.2091564381	H	-5.9949603254	-0.7194486709	3.4367725265
C	2.2468986311	4.6545050625	0.9487724842	H	-5.8670091613	0.9203882761	2.7979171481
H	2.3718131327	5.4635196852	0.2250756566	H	-4.9376713965	0.4489042880	4.2240630536
H	2.9624659456	4.7989279869	1.7566693206	H	-5.9916552602	-2.1443842786	1.6642506916
C	0.0001028250	3.8799539153	0.3417680222	C	4.0762529017	-0.0358057754	-0.1941377372
C	-2.2368335460	4.7976541383	-0.0789852709	C	4.4979364349	-0.7334457772	0.9385943413
H	-2.9422267247	5.1412161983	-0.8342392415	C	5.4922727659	-1.6908603380	0.8033329005
H	-2.3660950034	5.4056705409	0.8191770192	C	6.0600714778	-1.9830117045	-0.4266826885
C	0.7776266087	4.6107676834	1.4676610661	C	5.6125685449	-1.2910055217	-1.5411713821
H	0.3826179761	5.6036389861	1.6776142097	C	4.6329567511	-0.3144220570	-1.4442146266
H	0.7440921967	4.0185288757	2.3829863986	C	6.6125602633	-4.4068554767	-0.1353994945
C	-0.7616365412	4.8726206223	-0.5761702758	C	7.1363322164	-3.0333678083	-0.5598081266
H	-0.3601219960	5.8836303516	-0.5277851686	C	8.3741128232	-2.6495755440	0.2536095312
H	-0.7194158288	4.5295666976	-1.6111815769	C	3.8927109372	-0.4651424099	2.2966115385
C	-4.0932985504	-0.0232447555	-0.1450903684	C	3.2655809895	-1.7352689000	2.8733417832
C	-4.5278687104	-0.3677033734	-1.4259587167	C	4.9452648761	0.1086838771	3.2484523775
C	-5.4913105690	-1.3565238778	-1.5635198609	C	4.1296599111	0.3993116867	-2.6787158413
C	-6.0336250256	-2.0023875576	-0.4641192144	C	3.1405898645	-0.5106951191	-3.4127294609
C	-5.5855576028	-1.6450539446	0.7981910967	C	5.2595055418	0.8385517555	-3.6104206747
C	-4.6208389235	-0.6656755659	0.9765753044	H	5.7307167757	-4.6707887482	-0.7151341823
C	-8.3966184672	-2.6611299151	0.0254996160	H	7.3739678555	-5.1675912801	-0.2952959072
C	-7.0841432913	-3.0713701658	-0.6455595181	H	6.3417562097	-4.4088797279	0.9173039221
C	-6.5945721621	-4.4156714982	-0.1033413332	H	7.4281192224	-3.0916157603	-1.6147905838
C	-3.9256427565	0.2831329548	-2.6507423416	H	9.1668678803	-3.3793793521	0.1026465217
C	-2.9808342045	-0.7064870333	-3.3377867035	H	8.7397657317	-1.6718967748	-0.0529404823
C	-4.9941112844	0.7854323968	-3.6228818153	H	8.1417779964	-2.6107148041	1.3148444837
C	-4.1439797266	-0.3175448623	2.3675623822	H	5.8264424385	-2.2248952455	1.6793937022
C	-3.3857501543	-1.4976969342	2.9787146418	H	3.0936327472	0.2740169570	2.1875920012
C	-5.3114971186	0.1074954075	3.2613290397	H	2.5124301610	-2.1231499322	2.1906693002

H	2.7752836895	-1.5100879488	3.8170616257	H	3.6352995931	-1.4230046667	-3.7395896862
H	4.0139363292	-2.5053896905	3.0453915036	H	2.3220103236	-0.7824846405	-2.7502233116
H	4.4920528306	0.3452442885	4.2084248337	H	6.0115570683	1.3970081510	-3.0575672493
H	5.7493628503	-0.6042974247	3.4148793564	H	4.8631985795	1.4778716758	-4.3966975262
H	5.3752925484	1.0187280813	2.8348268338	H	5.7394435111	-0.0145030504	-4.0828937787
H	3.5908417333	1.2978812163	-2.3596358947	H	6.0402564617	-1.5233889698	-2.5052813763
H	2.7316027347	-0.0044694405	-4.2849709920				

CPA-**2**

O	-0.6907074372	1.9178833086	-0.8025394478	C	5.0405080588	0.4482453451	3.3769120372
H	-0.6110244469	2.8366427485	-0.5516338421	C	2.6765479184	-0.2302176314	3.0872924964
C	-2.6357992625	-3.2532832990	-1.2748601653	H	7.7816992786	4.3043349215	-1.6266332504
C	-1.2496586265	-3.1993946455	-0.9739604934	H	7.0092438049	2.8930666159	-2.3447152040
C	-0.7521760394	-2.1080159354	-0.1976381241	H	7.8003872181	2.7555455867	-0.7773127631
C	-1.5926356935	-1.0532039742	0.0683268038	H	5.5655881984	5.6012116718	-1.8275417536
H	-4.2066309179	-4.4057597374	-2.1765421154	H	4.8129126532	4.1802287526	-2.5461863127
H	-4.5139708884	-2.2074022392	-1.1784264784	H	4.1003085245	4.9178674552	-1.1139092391
C	-3.1458135603	-4.3666807215	-1.9739725836	H	5.3552954211	4.8933187327	1.0935258127
C	-0.4147077538	-4.2160338996	-1.4815316120	H	6.8924827838	4.0180330583	1.2237497343
C	-0.9360924750	-5.2694583690	-2.1813943126	H	6.7902119023	5.4907154644	0.2672843915
C	-2.3161072636	-5.3615452561	-2.4088496775	H	4.9617270631	1.9694247541	-2.4139167078
H	0.6510388173	-4.1506058750	-1.3268458801	H	5.0285031042	-1.1282797444	-4.0804768307
H	-0.2815190177	-6.0365873615	-2.5674047913	H	5.5264387025	0.4760076446	-3.5913273120
H	-2.7132500518	-6.2088309645	-2.9478648196	H	5.9004065796	-0.9034760446	-2.5631664135
C	1.4508954067	-1.0140921535	0.0758311065	H	3.1452304957	1.4199757016	-3.5917623190
C	0.6379182726	-2.1056036816	0.2914568501	H	1.8498993821	0.5184123112	-2.8107994969
C	1.1798545055	-3.2451889698	0.9578549025	H	2.6153053320	-0.1192293722	-4.2711224685
C	2.5759046580	-3.2993814097	1.2030454356	H	3.9147504980	-2.3492207290	-1.6413379708
C	3.3791739453	-2.1901549231	0.8695176700	H	3.3110573888	-2.2532866210	-3.2912152141
C	2.8317206391	-1.0351433469	0.3754432560	H	2.2642236334	-1.7962049561	-1.9673850512
H	-0.6959706863	-4.2487157123	1.2961112110	H	2.9592398076	2.0160365030	4.3241775112
C	0.3748632288	-4.3083098579	1.4164688851	H	2.2535560377	2.4024744458	2.7547157929
C	3.1229006763	-4.4475759388	1.8122662858	H	3.8712978343	2.9717387822	3.1765229445
H	4.4437873441	-2.2393596461	1.0497146433	H	5.8003436398	1.2102111673	3.2247536968
C	2.3204650591	-5.4803805718	2.2081487072	H	5.4197506288	-0.4920376019	2.9809456245
C	0.9326847642	-5.3973940055	2.0275730626	H	4.8830261587	0.3309395387	4.4470682748
H	4.1903685808	-4.4846009027	1.9763860393	H	3.0329049496	-1.2449258639	2.9501923122
H	2.7457882450	-6.3543605874	2.6787701054	H	1.7504757876	-0.0900389169	2.5382406599
H	0.3021997214	-6.2001680878	2.3798363468	H	2.4518015077	-0.1051994690	4.1456681864
O	-1.1601877154	0.0089123514	0.8475322815	H	5.1312506907	2.7362037683	1.7252714982
O	0.9397499218	0.1242924658	-0.5358098047	C	-3.7139070792	0.2041402036	-0.1242085755
P	-0.0471822033	1.0275150915	0.3458866804	C	-4.2142773903	0.5196502773	1.1465227848
O	0.5039880593	1.8130182446	1.4406438393	C	-4.7840140092	1.7765263293	1.3396443842
C	-2.9468774686	-1.0419808324	-0.3409115837	C	-4.9225313555	2.6952918965	0.3210419461
C	-3.4635789499	-2.1649464853	-0.9289843845	C	-4.4770988504	2.3219083100	-0.9376641518
C	3.6480380145	0.1726180468	0.1200221108	C	-3.8472349681	1.1127276626	-1.1963204082
C	4.0392251916	0.4875243922	-1.1949029440	C	-4.4998166364	5.1385422919	0.1368834066
C	4.7190721245	1.6804136566	-1.4058091157	C	-5.5375408312	4.0729730675	0.5176290975
C	5.1191883855	2.5143427604	-0.3766248080	C	-6.7761524504	4.2059897794	-0.3793738958
C	4.7883850988	2.1309508943	0.9072303671	C	-5.9698704550	4.3279069079	1.9642336214
C	4.0161740904	1.0063556355	1.1905444892	C	-4.2217751986	-0.3690080335	2.3982918917
C	7.2022121095	3.4094179405	-1.4085376589	C	-5.6722539528	-0.4745911772	2.9082914025
C	5.8957028331	3.7835514532	-0.6947004185	C	-3.3316999013	0.2968774402	3.4603523878
C	5.0397733663	4.6745971945	-1.6056721360	C	-3.7345702645	-1.8137199844	2.2495812035
C	6.2514409127	4.5904210538	0.5570885739	C	-3.4516372453	0.8551508374	-2.6625389439
C	3.8073187387	-0.3354294108	-2.4729018162	C	-2.2032487037	-0.0106172365	-2.8922246977
C	5.1492751736	-0.4716015727	-3.2213365858	C	-3.1662441164	2.1800879849	-3.3978715012
C	2.7929666222	0.4222142154	-3.3437560323	C	-4.6565120243	0.1855721766	-3.3430609875
C	3.2901364214	-1.7694307949	-2.3148691542	H	-3.6123888265	5.0345279354	0.7580011000
C	3.7168256181	0.8216708118	2.6877288838	H	-4.9116398024	6.1346369447	0.2863125374
C	3.1763673676	2.1435154270	3.2656422483	H	-4.2002243845	5.0471000846	-0.9035530067

H	-7.5018637325	3.4335404878	-0.1332995171	H	-4.2838083410	-2.3477420054	1.4797111215
H	-7.2430747388	5.1779377212	-0.2334972867	H	-3.8992594902	-2.3299946690	3.1951232237
H	-6.5188544581	4.1083727224	-1.4303763694	H	-2.6727784492	-1.8630447423	2.0360988927
H	-6.7317441257	3.6172037985	2.2765133538	H	-1.3855821658	0.3118902715	-2.2568845759
H	-5.1219418828	4.2664293280	2.6424318879	H	-1.8803919504	0.0997930329	-3.9263376867
H	-6.3920519945	5.3277554999	2.0456014830	H	-2.3877151248	-1.0653472743	-2.7320360049
H	-5.1257831371	2.0398739925	2.3237332933	H	-2.7316277913	1.9674071302	-4.3721345490
H	-6.0947439596	0.4876814498	3.1783691605	H	-2.4565800168	2.7810423907	-2.8346052726
H	-6.3038920998	-0.9186917683	2.1411575952	H	-4.0677186257	2.7587845320	-3.5809533660
H	-5.7078471846	-1.1095424582	3.7911481200	H	-4.4609429934	0.0569193817	-4.4057897412
H	-3.7017647716	1.2783077848	3.7427874338	H	-5.5495714906	0.7963274460	-3.2253138248
H	-2.3213163967	0.4138457233	3.0744473052	H	-4.8514741136	-0.7914218045	-2.9104451941
H	-3.2925828491	-0.3207174645	4.3555267216	H	-4.6232450313	3.0205562507	-1.7399320425

CPA-**3**

O	-0.8087594802	1.6907641545	-1.0662589676	C	4.3541117606	0.9156502144	1.0939158563
H	-0.6593329571	2.6345016381	-1.0497546539	C	6.9667851347	3.5737996731	-1.8168691100
C	-2.5402401689	-3.0849346732	-1.4779033747	C	5.9072118249	3.8881755943	-0.7511406170
C	-1.1847449100	-3.1448348947	-1.0596289963	C	6.6198824863	4.4697027175	0.4727884705
C	-0.6759407385	-2.1028942557	-0.2355845789	C	4.9301267169	4.9463323866	-1.2827044546
C	-1.4845539044	-1.0338220384	0.0786558405	H	6.5113967674	3.2133829417	-2.7349861378
H	-4.0931994617	-4.0766189503	-2.5814085370	H	7.6505571417	2.8095618067	-1.4530136549
H	-4.3601445766	-1.9437796184	-1.4383008878	H	7.5410025679	4.4682520148	-2.0497937023
C	-3.0560953469	-4.1239567324	-2.2818493289	H	7.3441412514	3.7654513399	0.8756938263
C	-0.3822042840	-4.2103604885	-1.5188904561	H	7.1544588148	5.3732814433	0.1857754657
C	-0.9095448951	-5.1935583362	-2.3094538450	H	5.9070352340	4.7337787195	1.2501942210
C	-2.2613642929	-5.1601534682	-2.6828477905	H	4.1552283377	5.1433860146	-0.5450048040
H	0.6612303956	-4.2376927608	-1.2448707102	H	5.4581192623	5.8753267443	-1.4886852446
H	-0.2835071899	-6.0029188439	-2.6548037665	H	4.4501603408	4.6168920210	-2.1999886199
H	-2.6629513677	-5.9510261216	-3.2990157215	H	4.4382372800	2.3675993270	-2.4477007229
C	1.5147384657	-1.0171049220	-0.0560224383	H	3.1588887406	0.3509601849	-2.0050671385
C	0.7201637527	-2.0998903530	0.2490704959	H	4.3150243006	0.5173696512	2.0972624652
C	1.2460545226	-3.1497992377	1.0512781791	H	5.6068942647	2.5223761384	1.6629030041
C	2.6033440590	-3.0786500784	1.4624898095	C	-3.6307540749	0.2422372180	-0.0509941669
C	3.3788507371	-1.9588270447	1.1073311121	C	-3.6929954077	0.7720001697	1.2318036196
C	2.8553994943	-0.9195354680	0.3766595521	C	-4.4466499981	1.9019497407	1.4942461119
H	-0.5848703409	-4.2691016707	1.2286656386	C	-5.1602374862	2.5460820110	0.4909174745
C	0.4597374928	-4.2327393037	1.4975688267	C	-5.0949784758	2.0044053256	-0.7913418840
C	3.1365685486	-4.1236396251	2.2471697553	C	-4.3503829543	0.8748626135	-1.0607315991
H	4.4080299046	-1.9133232946	1.4326847719	C	-5.9599675266	4.2500965653	2.1962510168
C	2.3574040494	-5.1761750974	2.6360246605	C	-5.9899178123	3.7967380380	0.7343007131
C	1.0040858793	-5.2214917631	2.2694043758	C	-7.4494466144	3.5133557707	0.3512887059
H	4.1744719085	-4.0674401297	2.5421248685	C	-5.4367010529	4.9374451352	-0.1317740476
H	2.7725399530	-5.9711826538	3.2377012214	H	-4.9472087218	4.4902208524	2.5112047200
H	0.3905401504	-6.0443126051	2.6052802809	H	-6.3656612910	3.4831095603	2.8519602993
O	-0.9631592662	0.0095576981	0.8345392432	H	-6.5683107639	5.1452446353	2.3103725976
O	0.9635366100	0.0303121855	-0.7840950423	H	-7.5407149652	3.2487499408	-0.6984570047
P	0.0377508313	1.0076310692	0.0959431815	H	-7.8406239004	2.6901702893	0.9456706990
O	0.6984694602	1.9311905854	1.0053091900	H	-8.0623622356	4.3933973225	0.5347158908
C	-2.8245879451	-0.9471079965	-0.3600528342	H	-5.4924198100	4.6953675510	-1.1894541357
C	-3.3314867930	-1.9804502789	-1.1099871997	H	-6.0066204239	5.8483015037	0.0397009263
C	3.6416092994	0.2855542781	0.0810681226	H	-4.3956604523	5.1287308578	0.1202504031
C	3.6902814352	0.8315805595	-1.1972252202	H	-4.4643536520	2.2734863960	2.5056885750
C	4.4238905546	1.9743220621	-1.4419215636	H	-3.1561133424	0.2917649530	2.0356877483
C	5.1310023052	2.6206688398	-0.4307091242	H	-4.3043874822	0.4882773435	-2.0686685312
C	5.0815454545	2.0635823097	0.8412949093	H	-5.6315692531	2.4713718585	-1.6039591634

CPA-**4**

O	0.7992872245	-2.4663014575	-0.9684329973	C	-2.3069664818	-0.8853486631	-3.7648909720
H	0.7052500556	-3.4076013807	-0.8293558521	H	-6.1741564150	-1.9201353853	2.4972436751
C	2.7055275363	2.5360537964	-1.0209151531	H	-5.8961014886	-0.2319491007	2.0769060229
C	1.2998575237	2.5206330248	-0.8256590465	H	-5.3916066620	-0.8426122739	3.6552775670
C	0.7142464399	1.4279185211	-0.1223544987	H	-3.4298878386	-0.6151267439	2.1832429752
C	1.5065127795	0.3592143215	0.2294023568	H	-4.1722407384	-3.5230766711	2.7229397115
H	4.3687954547	3.6394339874	-1.8119007856	H	-2.5533335245	-2.9196578487	2.3471802451
H	4.5483491198	1.4662679260	-0.7316320531	H	-3.3733382999	-2.4033408003	3.8240586585
C	3.2959446882	3.6323952626	-1.6839859673	H	-5.3583530994	-3.5207223852	0.8373114989
C	0.5333265353	3.5716961131	-1.3713730408	H	-5.6016649113	-4.0058996104	-1.5514304175
C	1.1343145099	4.6113243860	-2.0257103109	H	-4.6474735334	-2.4842879777	-3.2203433087
C	2.5287261268	4.6535728842	-2.1689299857	H	-2.7678635336	0.6353300287	-2.3383905636
H	-0.5412732194	3.5428938114	-1.2778724518	H	-5.0060647784	-0.3505125827	-4.1612696691
H	0.5336523562	5.4069001015	-2.4409265794	H	-5.1531149792	0.9205127554	-2.9475078239
H	2.9871122570	5.4884818955	-2.6778647030	H	-4.0888793868	1.1458019310	-4.3393661817
C	-1.4990422136	0.3430678328	-0.1757472102	H	-1.9030722500	-0.1728889625	-4.4816781121
C	-0.7245576654	1.4152092497	0.2069301436	H	-1.4843973523	-1.3024373946	-3.1891742627
C	-1.3326166238	2.4885739534	0.9198105549	H	-2.7886382853	-1.6894881343	-4.3157139484
C	-2.7405086549	2.4806661684	1.0995127301	C	3.6789827215	-0.8663219969	0.3511606276
C	-3.4958705898	1.3939471517	0.6161608428	C	3.8500846100	-1.2135497194	1.6955814218
C	-2.8969859441	0.3117988295	0.0234970561	C	4.5427253111	-2.3757239177	2.0057670999
H	0.4903597069	3.5252692296	1.4105741760	C	5.0602707597	-3.1790990391	1.0091242032
C	-0.5854102065	3.5375654790	1.4954875321	C	4.9064561442	-2.8171048823	-0.3146436619
C	-3.3515394893	3.5526198747	1.7835305676	C	4.2180760185	-1.6636710743	-0.6665136893
H	-4.5681358537	1.4044775950	0.7444505274	C	5.4830692036	-0.9320891263	-2.6832985374
C	-2.6026573263	4.5714655055	2.3009754782	C	4.1014234136	-1.2937849023	-2.1289547942
C	-1.2065682014	4.5527448988	2.1689104903	C	3.4713637139	-2.4168581754	-2.9546386503
H	-4.4254372193	3.5422417309	1.9013343162	C	3.3589402959	-0.3229372251	2.8138009364
H	-3.0770151933	5.3867923557	2.8267419966	C	4.5531206133	0.2518425073	3.5822484838
H	-0.6209745800	5.3464062848	2.6086316649	C	2.4150529097	-1.0738358103	3.7544808683
O	0.9520464285	-0.7233694391	0.8938823334	H	6.1473310875	-1.7924988425	-2.6546675659
O	-0.9031284366	-0.7366327340	-0.8145325492	H	5.9356905527	-0.1351835167	-2.0973361151
P	-0.0499152565	-1.7013941975	0.1375616065	H	5.3983790252	-0.5969873595	-3.7148738653
O	-0.7589220415	-2.5760018904	1.0596795265	H	3.4551525240	-0.4171068716	-2.2253365996
C	2.9001780962	0.3418004378	-0.0056028188	H	4.0752915651	-3.3201531829	-2.9263866269
C	3.4789841796	1.4433551451	-0.5816013360	H	2.4804486736	-2.6460575254	-2.5734333367
C	-3.6660583054	-0.8766053795	-0.4116511862	H	3.3783985994	-2.1025846507	-3.9922827304
C	-3.8204901219	-1.1365144396	-1.7774255433	H	5.3321832238	-3.4423206005	-1.0848821642
C	-4.5194773506	-2.2693993798	-2.1699911718	H	5.5944045617	-4.0829732475	1.2643549520
C	-5.0609887902	-3.1263829087	-1.2322662280	H	4.6838761714	-2.6564242523	3.0382965557
C	-4.9195839892	-2.8514752105	0.1133930698	H	2.8062645032	0.5173554515	2.3841403106
C	-4.2223257007	-1.7317338846	0.5466368296	H	5.2106265511	0.7976684883	2.9083786915
C	-5.4788792380	-1.0902416487	2.5995617493	H	4.2074099854	0.9340542671	4.3559623703
C	-4.1059690824	-1.4609811390	2.0291185323	H	5.1286193082	-0.5401220002	4.0551105782
C	-3.5161478244	-2.6581801579	2.7759108586	H	2.9298223668	-1.8852015806	4.2629083550
C	-3.2961418100	-0.1863813354	-2.8307390922	H	2.0234685696	-0.3932347236	4.5078137059
C	-4.4606511744	0.4186862058	-3.6203177699	H	1.5798257504	-1.4883915822	3.1952122712

CPA-**5**

O	-0.6354591842	0.8048173314	-0.9967802787	C	-1.9914083227	-6.2163667089	-2.8528437705
H	-0.6267964992	1.7390620463	-0.7935601610	H	0.8208830822	-5.1599762024	-1.2946598568
C	-2.4172701762	-4.1591332259	-1.6612002528	H	0.0228407940	-6.9601529275	-2.7461028827
C	-1.0819233472	-4.1576005269	-1.1801998470	H	-2.3288171221	-7.0232632861	-3.4863706625
C	-0.6625940682	-3.1037170338	-0.3172754496	C	1.5005690468	-2.0119230060	0.1362097550
C	-1.5204063200	-2.0544399770	-0.0802089017	C	0.6790243649	-3.1049925548	0.2981558145
H	-3.8734307567	-5.2167110014	-2.8318778305	C	1.1427576010	-4.2110636937	1.0672538199
H	-4.2873927822	-3.0990209241	-1.7084783566	C	2.4829123152	-4.2067427453	1.5347132813
C	-2.8498767643	-5.2189978179	-2.4856864496	C	3.3054817190	-3.0932085420	1.2734494681
C	-0.2112280768	-5.1813770919	-1.6087878019	C	2.8264043983	-1.9850004728	0.6227000929
C	-0.6565703498	-6.1850361683	-2.4239616810	H	-0.7269952847	-5.2787240041	1.1187132110

C	0.3081764406	-5.2911393531	1.4235619086	H	3.4407446348	1.7122050565	3.7145769181
C	2.9549927622	-5.3100247146	2.2764803096	H	1.9566372332	1.1056044183	2.9696945970
H	4.3275764871	-3.1046008758	1.6221056249	H	5.3544549027	-1.5057689785	3.2834285964
C	2.1304614184	-6.3571910939	2.5766669342	H	4.5169093147	-0.9878264870	4.7495537215
C	0.7915575036	-6.3371958653	2.1597817936	H	5.4907281868	0.1628311369	3.8323564597
H	3.9812699229	-5.3013784964	2.6141316311	C	-3.6719317845	-0.8234511580	-0.3559344076
H	2.4979322813	-7.1967129094	3.1480218272	C	-4.0009274058	0.0248176560	-1.4201725368
H	0.1394116715	-7.1548148246	2.4288950529	C	-4.7352666718	1.1701758929	-1.1592740526
O	-1.1217888398	-1.0094490433	0.7388992322	C	-5.1586568510	1.5121480452	0.1190397805
O	1.0362191036	-0.9028684315	-0.5587640716	C	-4.8431835755	0.6406211504	1.1488943941
P	-0.0114889112	0.0123966391	0.2332540483	C	-4.1100190758	-0.5184154625	0.9342032855
O	0.4757562389	0.8645507507	1.3080388833	C	-3.6088277996	-0.2840496339	-2.8489898656
C	-2.8346749100	-2.0187244189	-0.5989243546	C	-4.8609190796	-0.6100560970	-3.6693709159
C	-3.2729378607	-3.0874997762	-1.3379549795	C	-2.8295082354	0.8630255239	-3.4945481480
C	3.6475958981	-0.7712968554	0.4195956171	C	-5.9432174119	2.7967385881	0.3295954958
C	4.0740515297	-0.4226485093	-0.8653148730	C	-6.3297232074	3.0244920462	1.7959771402
C	4.8124822888	0.7357324864	-1.0388332525	C	-7.1099610859	4.3302997337	1.9500264502
C	5.1515169678	1.5656354896	0.0225610944	C	-8.3861141848	4.2688472189	1.1140724735
C	4.7352934209	1.1848718590	1.2879330315	C	-8.0109685012	4.0600787680	-0.3508686385
C	3.9873720303	0.0357651978	1.5076172350	C	-7.1453983805	5.2256162441	-0.8273699536
C	3.7981928533	-1.3096582196	-2.0592992147	C	-5.8756543408	5.2939549215	0.0199501192
C	3.0124616707	-0.5633799316	-3.1387072237	C	-6.2499409834	5.5028829841	1.4850322370
C	5.1117863788	-1.8617337659	-2.6206339911	C	-5.0907409280	3.9916947859	-0.1281279328
C	5.9517925271	2.8308044779	-0.2416631761	C	-7.2359053899	2.7522065954	-0.5013617451
C	6.2451775885	3.6270942385	1.0356106718	C	-3.8657777334	-1.4561660768	2.0950723212
C	7.0440898492	4.8896459036	0.7111828697	C	-5.1951602208	-2.0514136201	2.5696747020
C	6.2441209822	5.7746976408	-0.2414970280	C	-3.1403433343	-0.7550324176	3.2446924425
C	5.9631519406	4.9965581523	-1.5244329122	H	-2.9581335531	-1.1626399009	-2.8570103674
C	7.2846259366	4.6028311655	-2.1829033429	H	-5.5156768913	0.2556868798	-3.7349383013
C	8.0892504180	3.7286847508	-1.2220598592	H	-4.5833264556	-0.9083380889	-4.6781569775
C	8.3712181973	4.5067442129	0.0606761736	H	-5.4199715442	-1.4222863390	-3.2101009825
C	7.2957133499	2.4653830063	-0.8927473684	H	-3.4233622111	1.7721338250	-3.5450204582
C	5.1600968917	3.7390403298	-1.1965610646	H	-1.9274748572	1.0654313665	-2.9244220351
C	3.5787134997	-0.3104683109	2.9216603567	H	-2.5437319656	0.5886174557	-4.5080732576
C	2.8115309019	0.8357432465	3.5825338495	H	-4.9848561103	1.8147193756	-1.9876722718
C	4.8135411848	-0.6838617320	3.7479457449	H	-5.4293075992	3.0714613154	2.4125551716
H	3.1933724705	-2.1616829757	-1.7353515408	H	-6.9500377485	2.1970845706	2.1483192664
H	2.7728424911	-1.2369048866	-3.9593925473	H	-7.3719233048	4.4664489866	3.0060062496
H	2.0844633180	-0.1793103246	-2.7225843412	H	-8.9522019856	5.1963842486	1.2271221888
H	3.5856966424	0.2701612525	-3.5372612303	H	-9.0183795303	3.4465292329	1.4573655438
H	4.9112546342	-2.5466677671	-3.4420263746	H	-8.9221589467	4.0097972631	-0.9586658029
H	5.7457972998	-1.0596923158	-2.9903541823	H	-7.7016467643	6.1618389609	-0.7400862116
H	5.6565624476	-2.3990144680	-1.8470583509	H	-6.8840023397	5.0894516071	-1.8796448874
H	5.1335812862	0.9923363075	-2.0365699644	H	-5.2533800262	6.1294760473	-0.3218990240
H	6.8199004021	3.0117567417	1.7317295815	H	-6.8004208852	6.4393936462	1.6012437853
H	5.3069480809	3.9123558834	1.5163836518	H	-5.3456290237	5.5674464518	2.0945848528
H	7.2385679544	5.4374278339	1.6410268560	H	-4.7928680862	3.8655300726	-1.1702084346
H	6.8069421166	6.6830431082	-0.4696080571	H	-4.1792017284	4.0334275917	0.4735107612
H	5.3029995898	6.0718529090	0.2270726805	H	-7.0025775913	2.5899195811	-1.5546888011
H	5.3847736739	5.6233584142	-2.2136360860	H	-7.8500060699	1.9125322121	-0.1661348730
H	7.8555532313	5.4991493866	-2.4368870569	H	-5.1708731393	0.8516348580	2.1524011341
H	7.0907042993	4.0567719415	-3.1093790522	H	-3.2354391858	-2.2832385154	1.7563528870
H	9.0374787532	3.4449698445	-1.6941013142	H	-5.6938887798	-2.5656368936	1.7504589701
H	8.9499143773	5.4054289979	-0.1658101811	H	-5.0213579999	-2.7656207041	3.3717251219
H	8.9600821268	3.8930719485	0.7465926489	H	-5.8587660164	-1.2739613461	2.9402490762
H	7.1295183040	1.8948803093	-1.8076659111	H	-2.2106887906	-0.3173267096	2.8893416793
H	7.8667026688	1.8326688005	-0.2084716780	H	-2.9075041648	-1.4727637468	4.0287468703
H	4.9298991500	3.2053174929	-2.1198744283	H	-3.7533490397	0.0324397539	3.6760248324
H	4.2112933314	4.0109395324	-0.7275356454				
H	4.9879192688	1.7901686850	2.1413626068				
H	2.9107453658	-1.1761252285	2.8940364250				
H	2.4542779847	0.5233410492	4.5617405346				

VI.31 – Detailed Statistical Analyses

The statistical analyses reported within the main text were limited to comparisons between only a few pairs of descriptors. We show here additional analyses that support the expectation that *active site-based* buried volume is an *advantageous*, *best-in-general* representation relative to other features, including: (1) expanded confidence intervals (ref. **Fig. 2** from main text), (2) average statistical performance for other representations of confinement across structural diversity (ref. **Fig. 3** from main text), (3) functional diversity (ref. **Fig. 4** from main text), (4) mechanistic diversity (ref. electronic control in **Fig. 5** from main text), and (5) univariate regression models for each dataset built from *active site-based* buried volume.

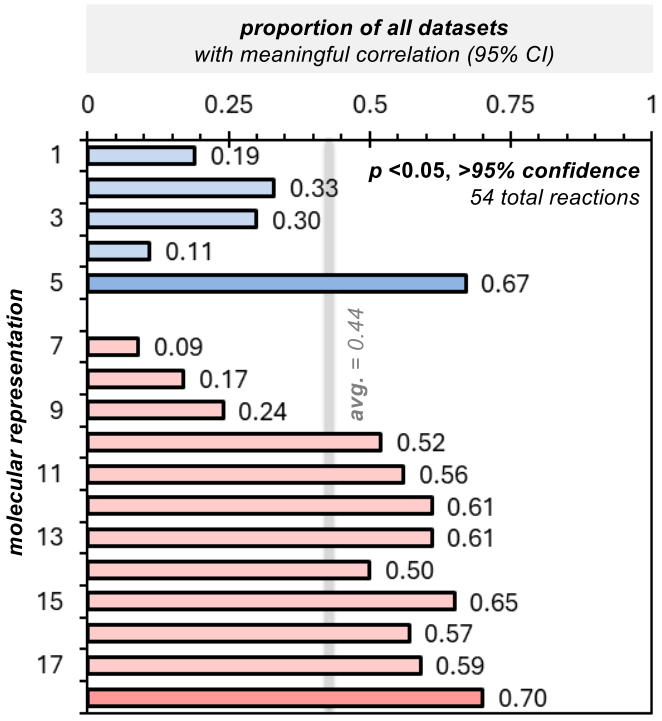
- ***Expanded Confidence Intervals for Meaningful Correlations***

Initially, *active site-based* buried volume was assigned as the *most meaningful* representation as it captured the greatest proportion of *meaningful* univariate linear regressions (based on *p*-value, ref. *dataset_scores.xlsx*) across all datasets considered (54 total reactions). In the main text (**Fig. 2B**), this analysis was conducted with the standard 95% confidence and 99% confidence. We report a complete analysis (up to 99.9% confidence) below for all descriptors considered.

Across the analyses, we observe qualitatively consistent (if not more robust) conclusions at increasingly strict confidence levels. Even at the most strict confidence, *active site-based* buried volume captures the greatest proportion of *meaningful* univariate regressions, and a much greater proportion compared to the average descriptor altogether.

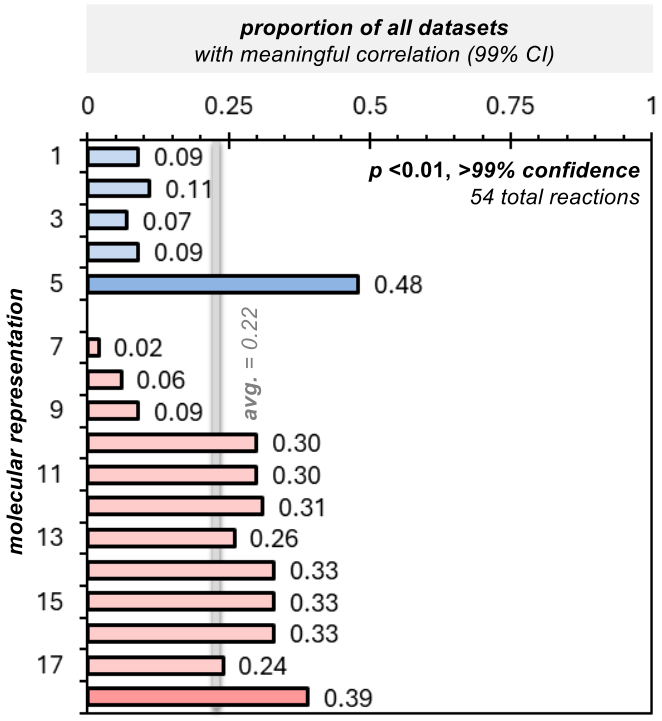
	confidence level				
molecular descriptor	**95%**	**99%**	**99.50%**	**99.90%**	
active site-based sterimol_L	0.19	0.09	0.09	0.04	1
active site-based sterimol_Bmin	0.33	0.11	0.06	0.02	2
active site-based sterimol_Bmax	0.3	0.07	0.02	0.02	3
active site-based xtb_pKa	0.11	0.09	0.06	0	4
active site-based bur_vol aggregate	**0.67**	**0.48**	**0.41**	**0.26**	5
					6
substituent-based sterimol_L	0.09	0.02	0.02	0	7
substituent-based sterimol_Bmin	0.17	0.06	0.06	0.04	8
substituent-based sterimol_Bmax	0.24	0.09	0.09	0.06	9
substituent-based xtb_atom_charge	0.52	0.3	0.22	0.07	10
substituent-based isotropic_nmr	0.56	0.3	0.2	0.09	11
substituent-based anisotropic_nmr	0.61	0.31	0.24	0.11	12
substituent-based mulliken_charge	0.61	0.26	0.15	0.11	13
substituent-based nbo_charge	0.5	0.33	0.2	0.04	14
substituent-based chelpg_charge	0.65	0.33	0.2	0.07	15
substituent-based hirshfeld_charge	0.57	0.33	0.22	0.11	16
substituent-based cm5_charge	0.59	0.24	0.19	0.09	17
substituent-based bur_vol	0.7	0.39	0.35	0.19	18
average	0.44	0.22	0.16	0.08	

Supplementary Fig. VI-43 | Indexed and tabulated proportions of meaningful regressions across all datasets by each descriptor considered, from 95% to 99.9% confidence.

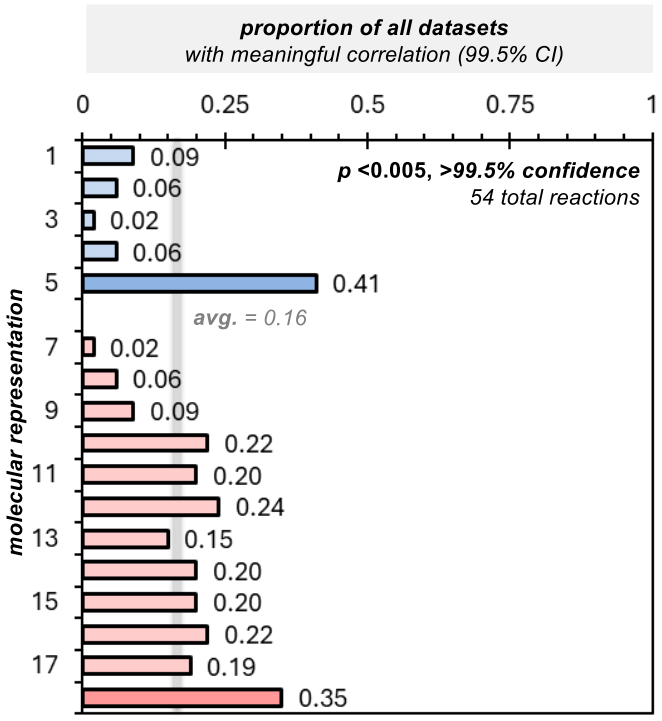


Supplementary Fig. VI-44 | Proportion of datasets (54 total reactions) captured with a *meaningful* univariate linear regression model by the corresponding descriptors (at 95% confidence).

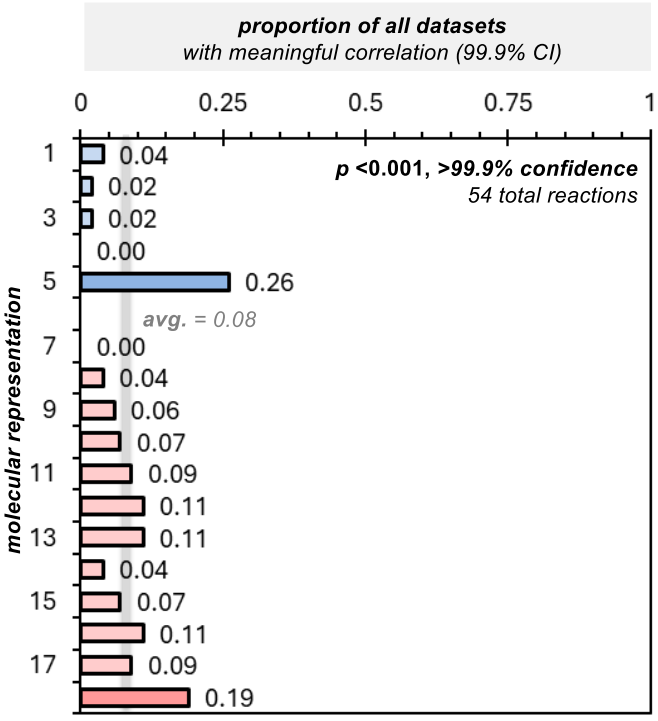
Supplementary Fig. VI-45 | Proportion of datasets (54 total reactions) captured with a *meaningful* univariate linear regression model by the corresponding descriptors (at 99% confidence).



Supplementary Fig. VI-46 | Proportion of datasets (54 total reactions) captured with a *meaningful* univariate linear regression model by the corresponding descriptors (at 99% confidence).



Supplementary Fig. VI-47 | Proportion of datasets (54 total reactions) captured with a *meaningful* univariate linear regression model by the corresponding descriptors (at 99.5% confidence).



Supplementary Fig. VI-48 | Proportion of datasets (54 total reactions) captured with a *meaningful* univariate linear regression model by the corresponding descriptors (at 99.9% confidence).

VI.32 – Regression Performance over Structurally Diverse Datasets

Active site-based buried volume and *substituent-based* buried volumes were explicitly juxtaposed in the main text (**Fig. 3**). The two are distinguished solely by their anticipated capacity to reflect *structural diversity* as it is reflected by the construction of chiral acid organocatalysts. We identify three prominent structural motifs (substituent, chiral backbone, and active site) in BAOCs. As *substituent-based* representations should reflect changes in the substituent, we purposefully group datasets that display changes primarily to this motif alone and denote these as *limited diversity*. The remaining datasets display changes at other motifs as well and we denote these as *significant diversity*.

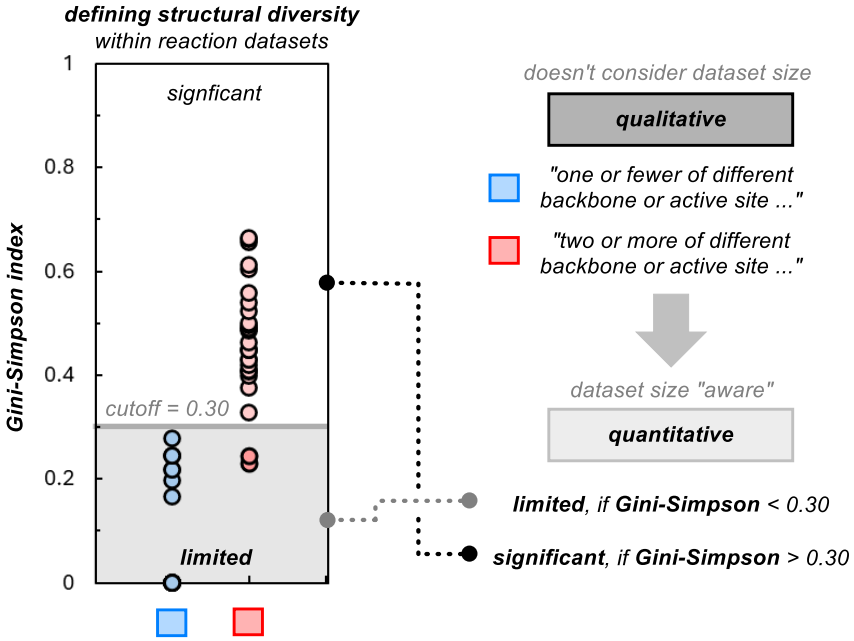
The qualitative distinction of *limited* and *significant diversity* are distinguished by whether there are more than one examples of BAOCs with structural motifs changed at the chiral backbone or active site (relative to the majority). As this qualitative definition fails to reflect dataset size, we invoke a *quantitative* distinction defined by the Gini-Simpson index. The index is essentially *the likelihood of the two samples drawn from the dataset bearing different structural motifs* (as defined by a different chiral backbone or active site). We labeled each dataset with its corresponding *qualitative* definition and then compared them to their *quantitative* Gini-Simpson index. Excellent

agreement (F1 = 0.96) between the two categories was observed with a univariate classification at Gini-Simpson = 0.30 to distinguish *limited* and *significant diversity*. This *quantitative* analysis only *misclassified* two datasets – R-**X** and R-**XXIII**. Each dataset possesses two instances of a BAOC with a different chiral backbone, although the datasets are relatively large (15 ex. and 14 ex. respectively, compared to the average size of 11). As such, we agree with the classification that such sparse diversity should be considered *limited diversity*.

Altogether, the full classification follows:

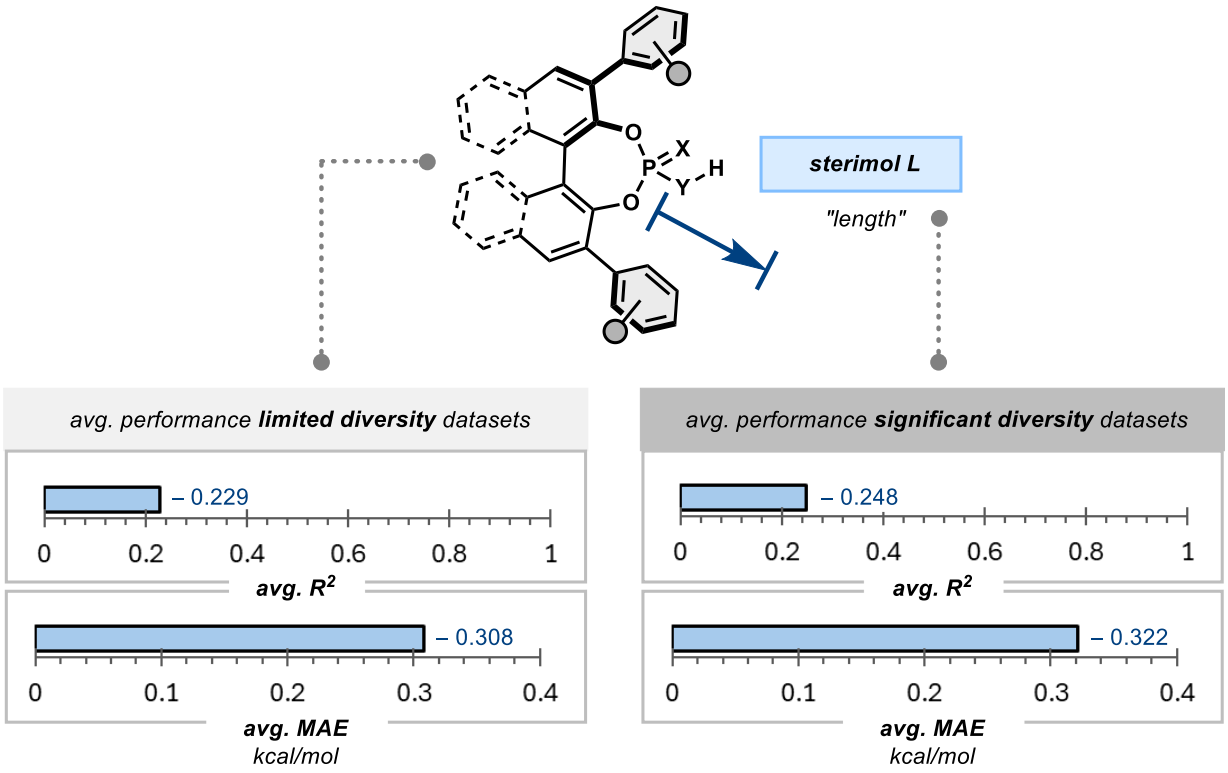
Limited diversity – R-**I** through R-**XXVII**.

Significant diversity – R-**XXVIII** through R-**LIV**.

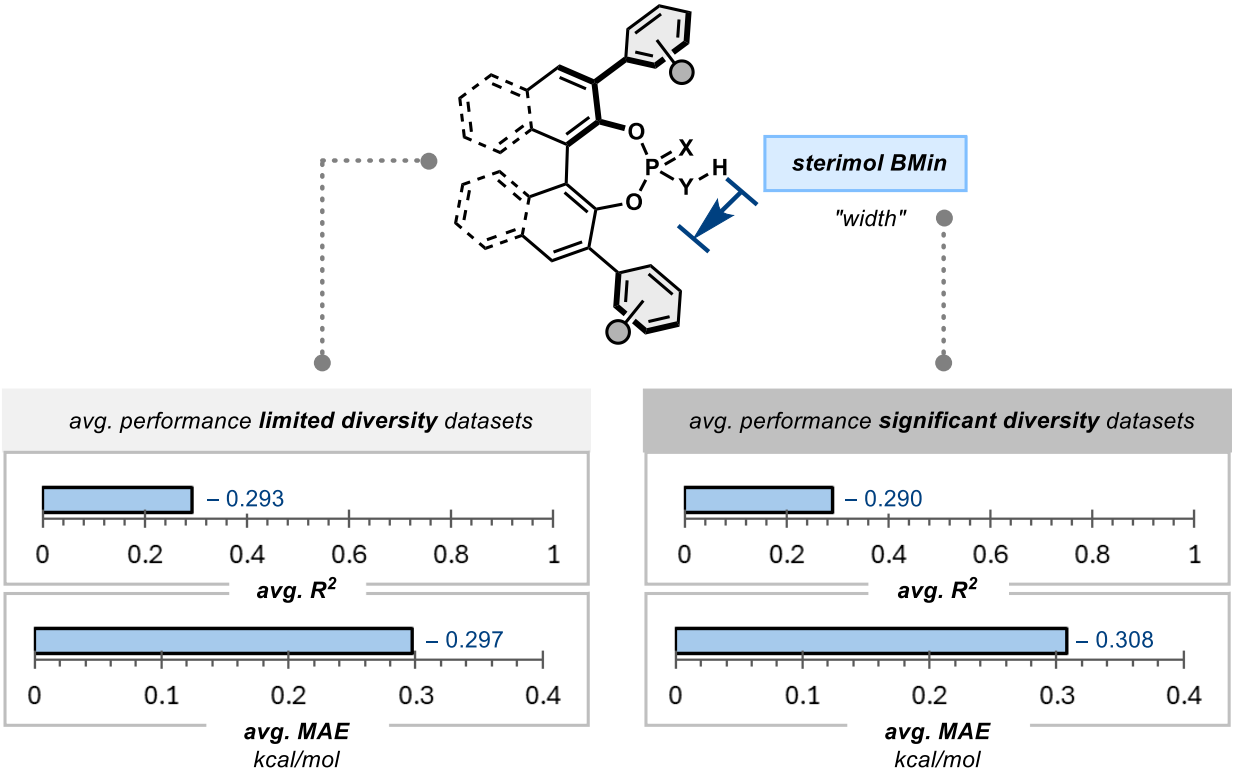


Supplementary Fig. VI-49 | Univariate classification of *structural diversity* with calculated Gini-Simpson index. Classifier defines *limited diversity* if the corresponding Gini-Simpson index is less than 0.30, and *significant diversity* if it is greater than 0.30 (F1 = 0.96).

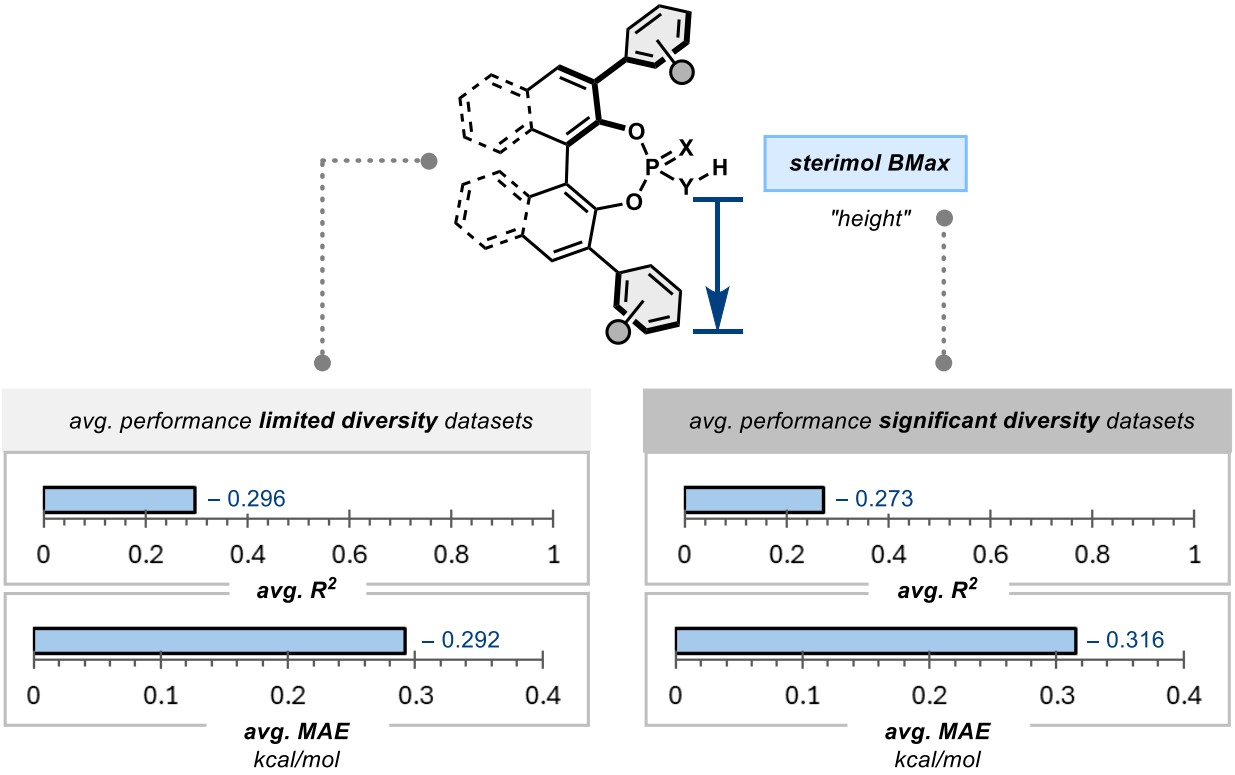
In the main text, only *active site-based* and *substituent-based* buried volumes were showcased in their performance across the *limited* and *significant diversity* datasets. We showcase other competing representations of confinement below, highlighting their inadequate performance and the relative power of *active site-based* buried volume to accommodate BAOC structural diversity in relationships to stereoselectivity.



Supplementary Fig. VI-50 | Average performance (R^2 and MAE in *kcal/mol*) across *limited diversity* (left) and *significant diversity* datasets by *active site-based* sterimol L.



Supplementary Fig. VI-51 | Average performance (R^2 and MAE in *kcal/mol*) across *limited diversity* (left) and *significant diversity* datasets by *active site-based* sterimol BMin.



Supplementary Fig. VI-52 | Average performance (R^2 and MAE in *kcal/mol*) across *limited diversity* (left) and *significant diversity* datasets by *active site-based* sterimol BMax.

- ***Regression Performance over Functionally Diverse Datasets***

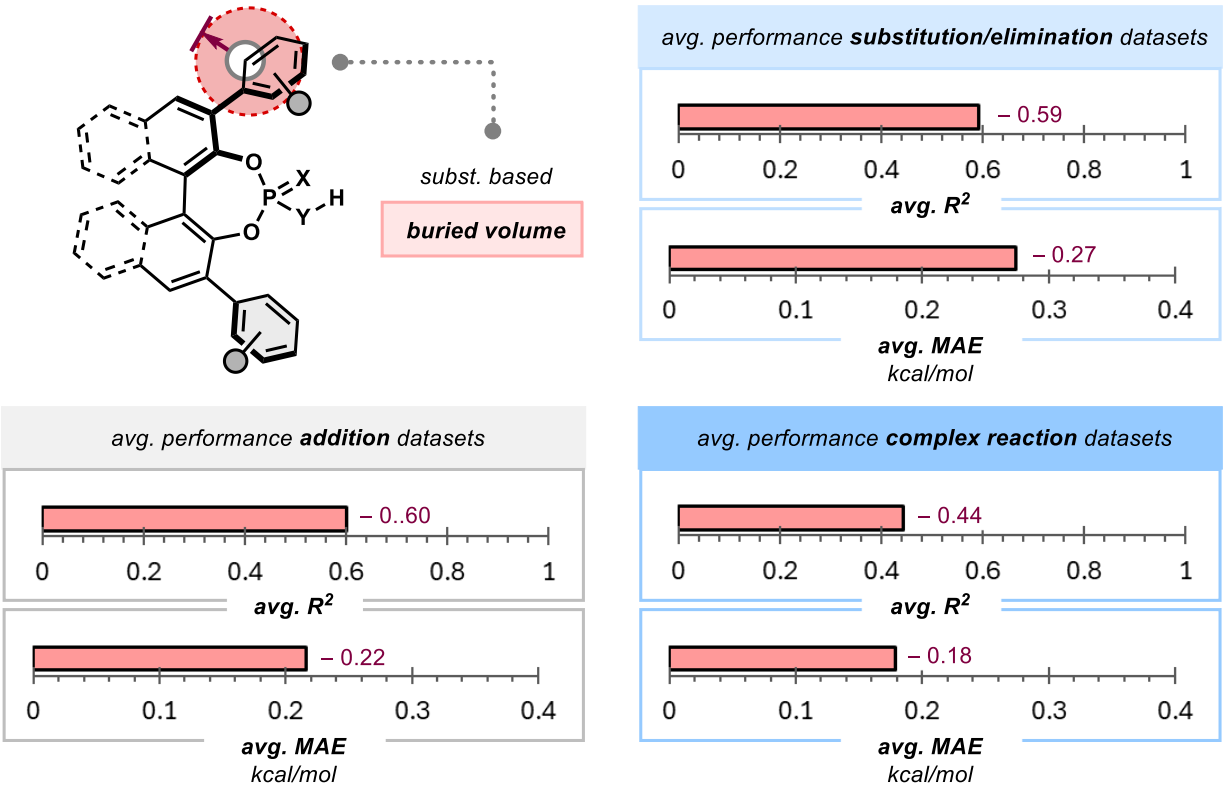
In the main text, only *active site-based* buried volume was showcased in its performance across the three classifications of *functional diversity – acid-mediated additions, acid-mediated substitutions and eliminations,* and *complex/multicomponent reaction* datasets. These categories are defined as described in the text and contain the following datasets accordingly:

Acid-mediated additions – R-**II**, R-**III**, R-**IV**, R-**VII**, R-**VIII**, R-**IX**, R-**X**, R-**XII**, R-**XV**, R-**XVII**, R-**XVIII**, R-**XIX**, R-**XX**, R-**XXIV**, R-**XXVI**, R-**XXVIII**, R-**XXIX**, R-**XXX**, R-**XXXII**, R-**XXXIII**, R-**XXXVIII**, R-**XXXIX**, R- **XLIII**, R-**XLV**, R-**XLVIII**, R-**LIII**, R-**LIV**

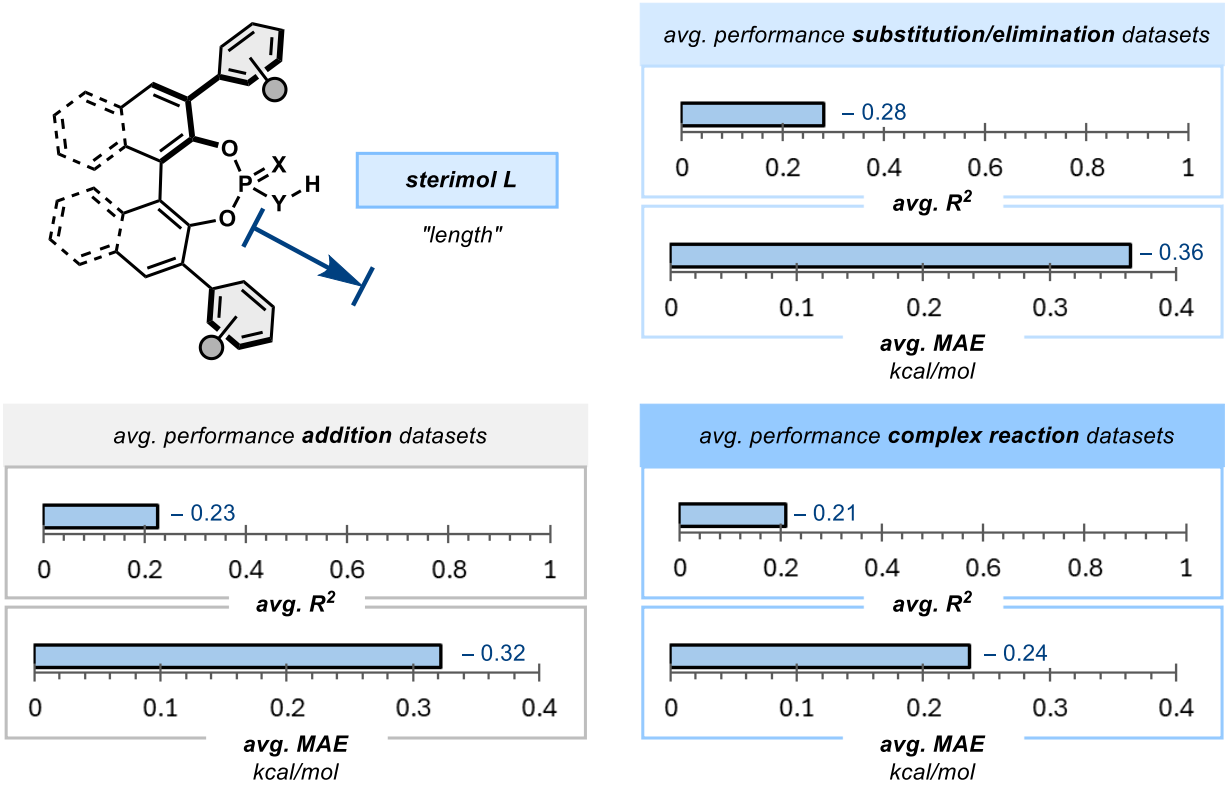
Acid-mediated substitutions/eliminations – R-**I**, R-**VI**, R-**XI**, R-**XXI**, R-**XXII**, R-**XXV**, R-**XXVII**, R-**XXXI**, R-**XXXIV**, R-**XLII**, R-**XLIV**, R-**XLVI**, R-**XLVII**, R-**L**, R-**LII**

Complex/multicomponent reactions – R-**V**, R-**XIII**, R-**XIV**, R-**XVI**, R-**XXIII**, R-**XXXV**, R-**XXXVI**, R-**XXXVII**, R-**XL**, R-**XLI**, R-**XLIX**, R-**LI**

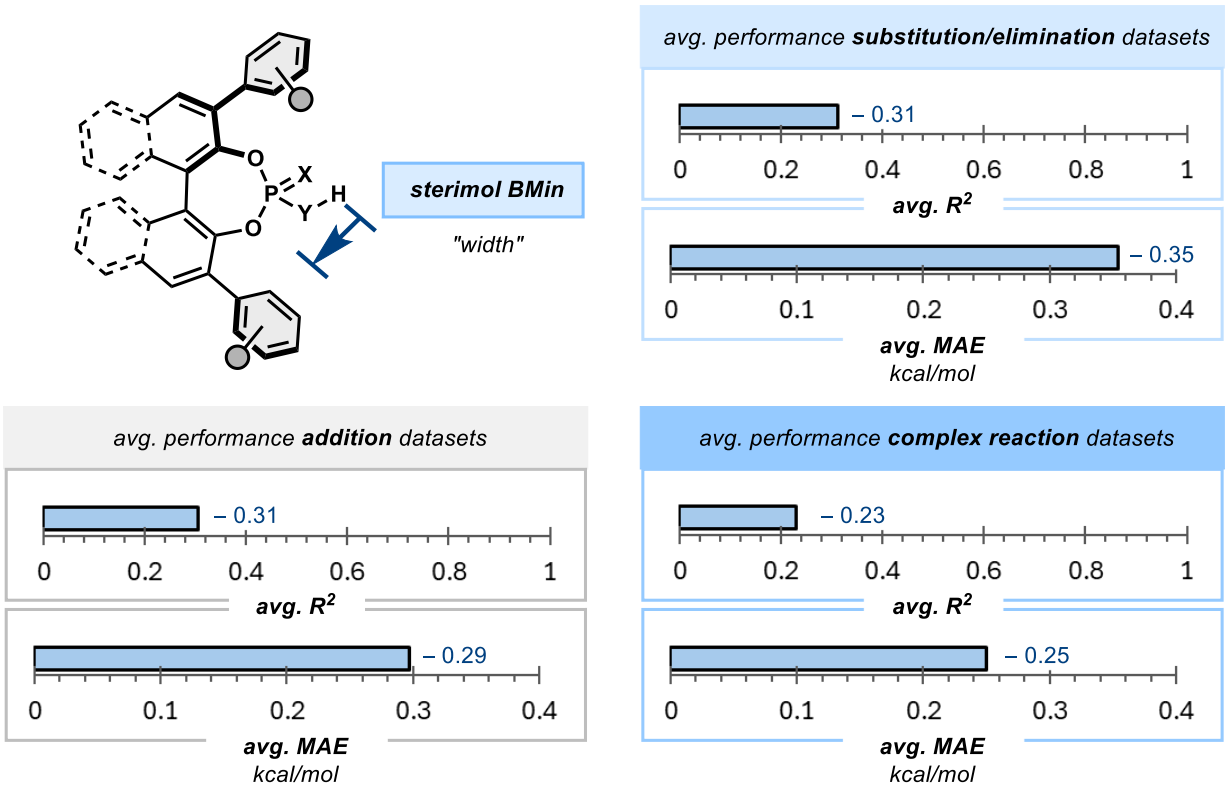
Below, we showcase other representations of confinement (both *active site-based* and *substituent-based* representations). These results highlight the utility of *active site-based* buried volume to accommodate BAOC functional diversity in relationships to stereoselectivity.



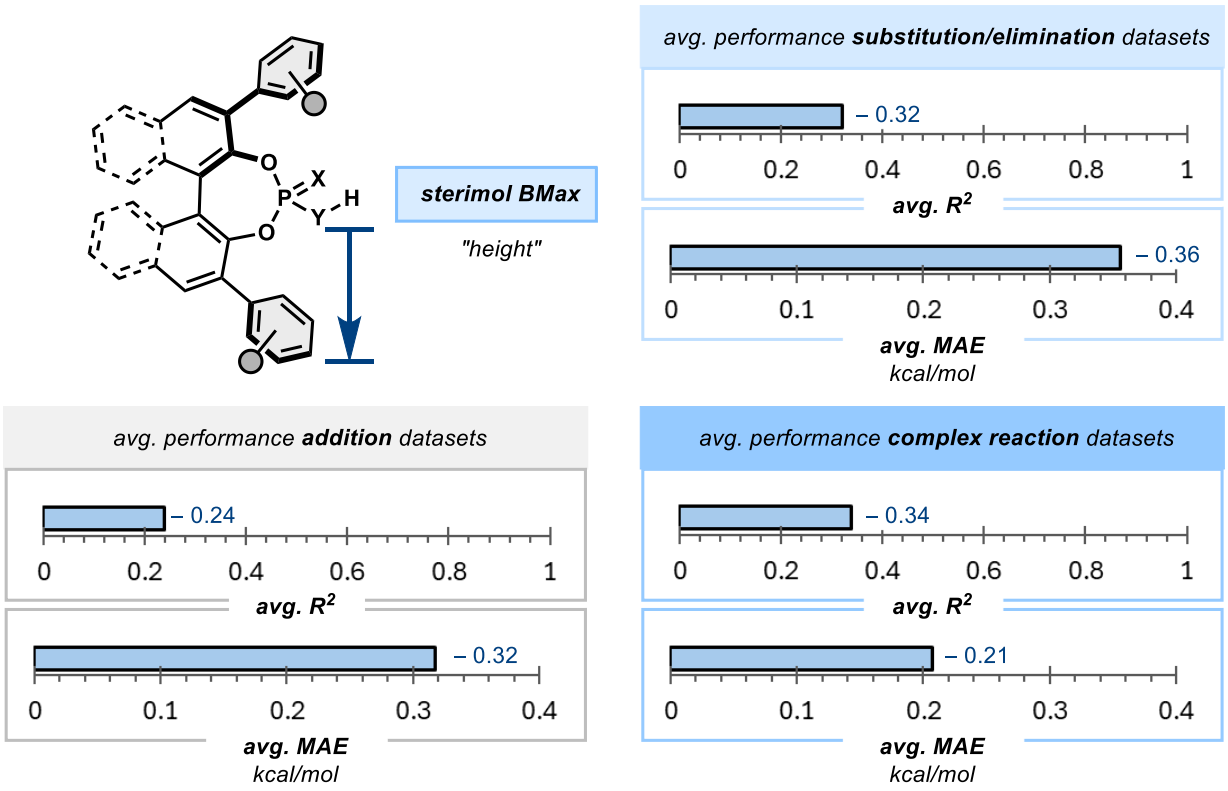
Supplementary Fig. VI-53 | Average performance (R^2 and MAE in *kcal/mol*) across *acid-mediated addition* (bottom left), *acid-mediated substitutions and eliminations* (top right), and *complex/multicomponent reaction* (bottom right) datasets by *substituent-based* buried volume.



Supplementary Fig. VI-54 | Average performance (R^2 and MAE in *kcal/mol*) across *acid-mediated addition* (bottom left), *acid-mediated substitutions and eliminations* (top right), and *complex/multicomponent reaction* (bottom right) datasets by *active site-based* sterimol L.



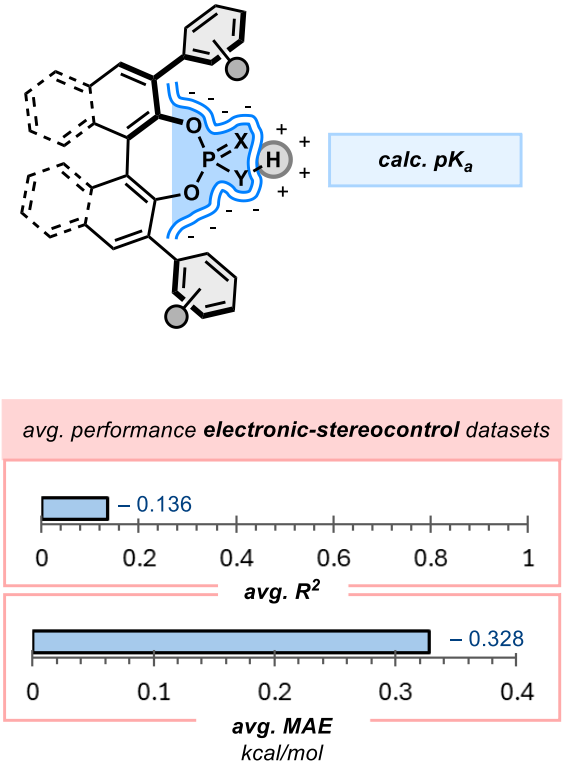
Supplementary Fig. VI-55 | Average performance (R^2 and MAE in *kcal/mol*) across *acid-mediated addition* (bottom left), *acid-mediated substitutions and eliminations* (top right), and *complex/multicomponent reaction* (bottom right) datasets by *active site-based* sterimol BMin.



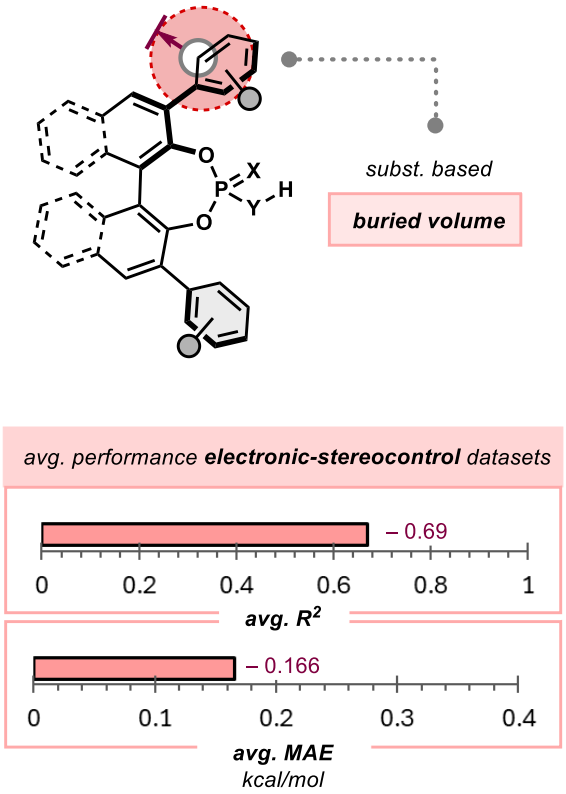
Supplementary Fig. VI-56 | Average performance (R^2 and MAE in *kcal/mol*) across *acid-mediated addition* (bottom left), *acid-mediated substitutions and eliminations* (top right), and *complex/multicomponent reaction* (bottom right) datasets by *active site-based* sterimol BMax.

- ***Regression Performance over Electronic-Controlled Datasets***

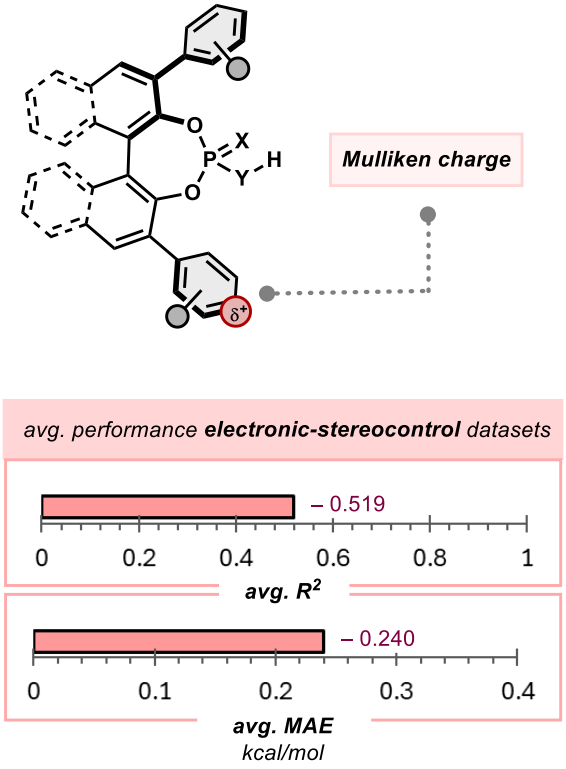
In the main text, *substituent-based* ChelpG, *substituent-based* NBO charge, and *active site-based* buried volume were showcased in their performance across three datasets operating by electronic stereocontrol – R-**XI,** R-**XXII**, R-**XXVII**. We showcase the remaining electronic representations below. Of note, the three datasets are each *limited diversity*. As such, we show that buried volume (whether at the *substituent* or *active site*) are each a suitable representation to capture the underlying electronic effects.



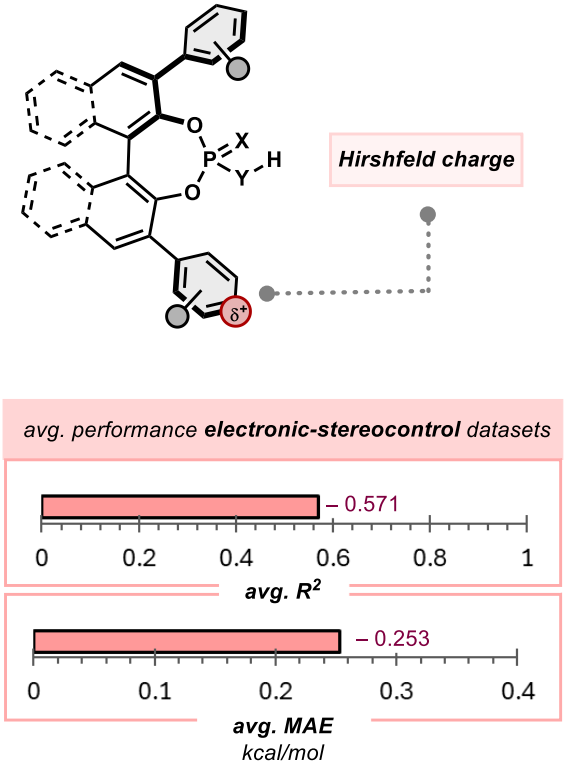
Supplementary Fig. VI-57 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *calc.* pK_a.



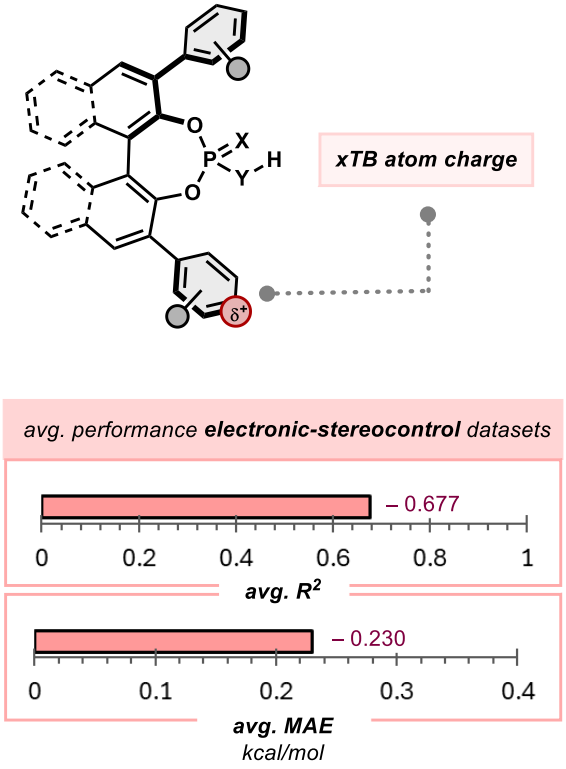
Supplementary Fig. VI-58 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* buried volume.



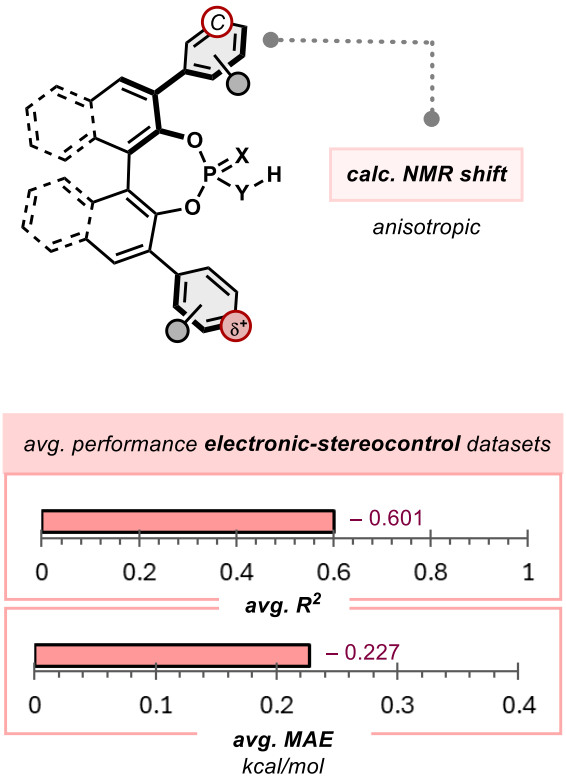
Supplementary Fig. VI-59 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* Mulliken charge.



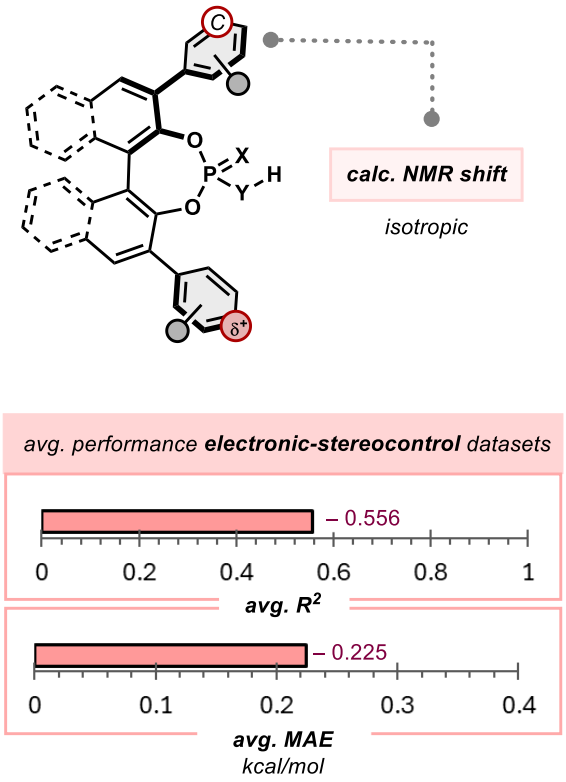
Supplementary Fig. VI-60 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* Hirshfeld.



Supplementary Fig. VI-61 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* xTB atom charge.



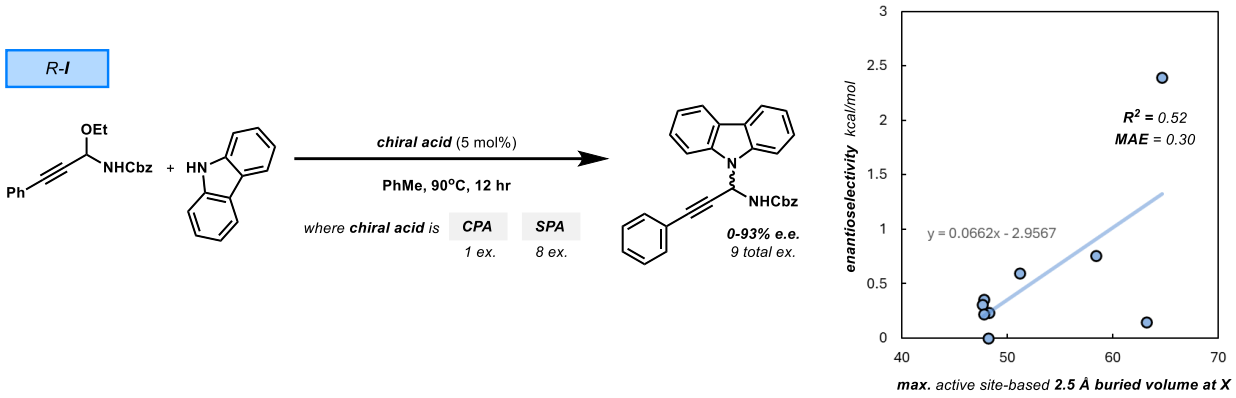
Supplementary Fig. VI-62 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* calculated anisotropic NMR shift.



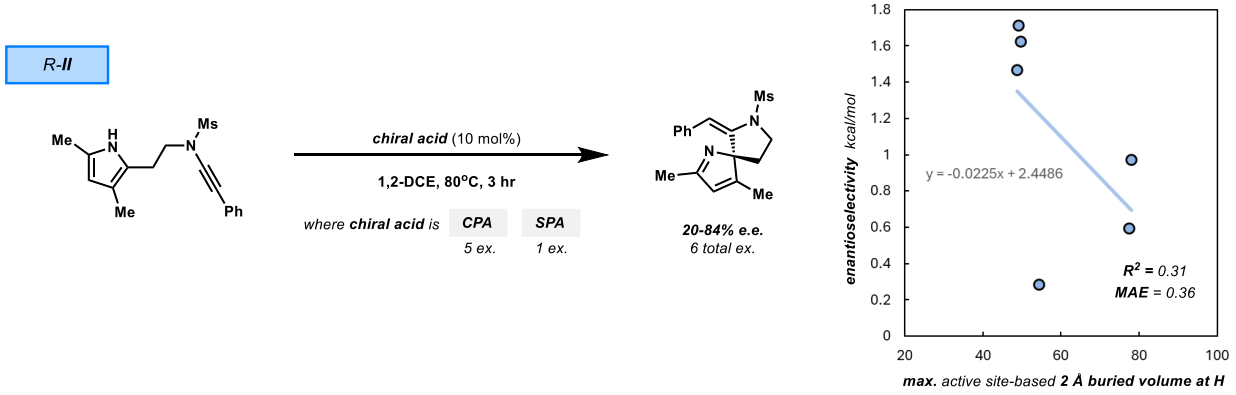
Supplementary Fig. VI-63 | Average performance (R^2 and MAE in *kcal/mol*) across *electronic stereocontrolled* datasets (R-**XI**, R-**XXII**, and R-**XXVII**) by *substituent-based* calculated isotropic NMR shift

- ***Univariate Regression Models from Active Site-based Buried Volume***

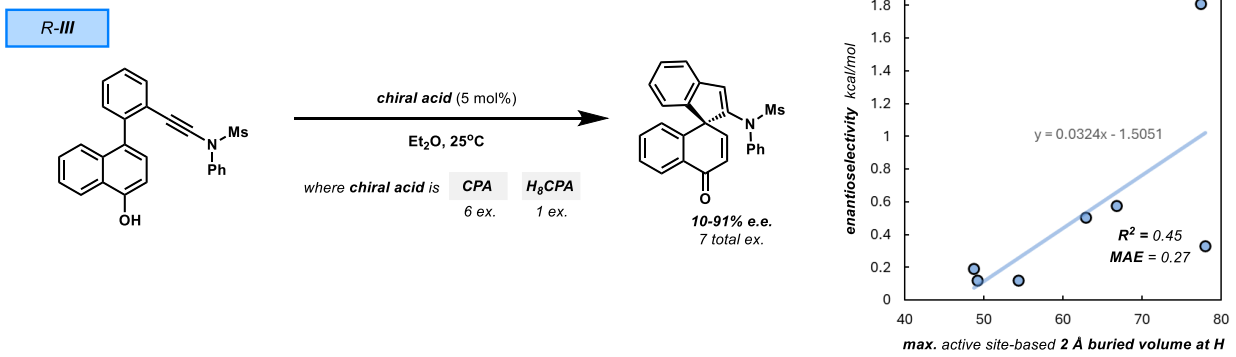
The following correlations between the selected datasets and *active site-based* buried volume are invoked through the main text and the SI. For clarity (in transformation, dataset, and correlation), we show all 54 correlations below to *active site-based* buried volume. The source for each dataset is tabulated in *dataset_rxn.xlsx* for reference under the corresponding key.



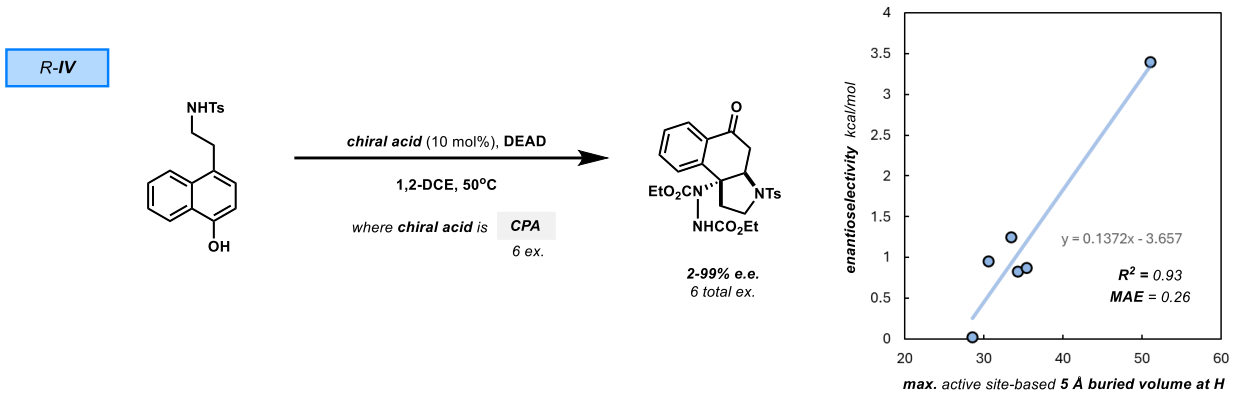
Supplementary Fig. VI-64 | R-**I** correlation drawn by *active site-based* buried volume.



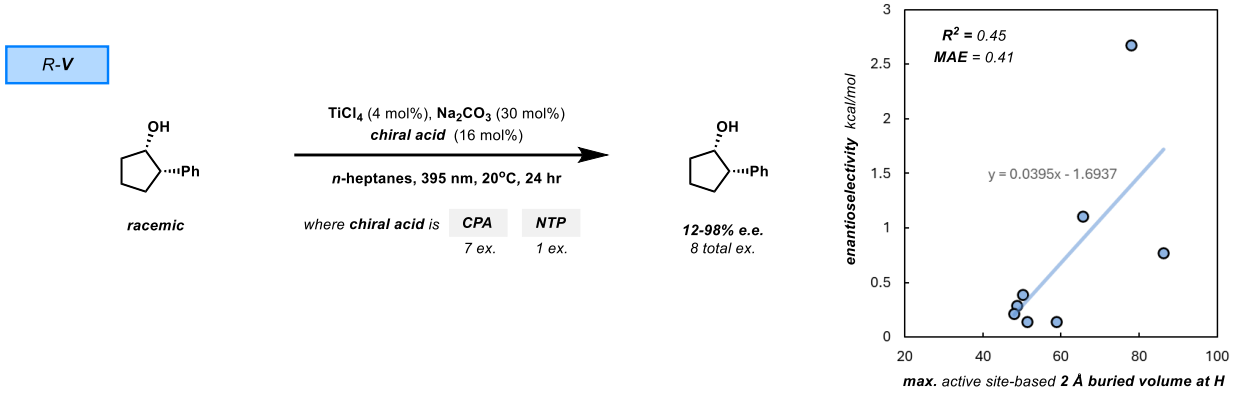
Supplementary Fig. VI-65 | R-**II** correlation drawn by *active site-based* buried volume.



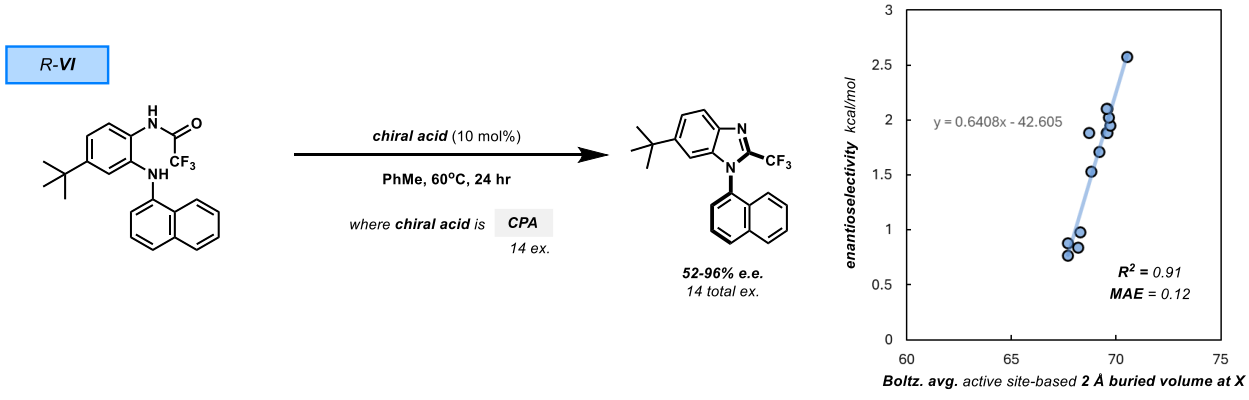
Supplementary Fig. VI-66 | R-**III** correlation drawn by *active site-based* buried volume.



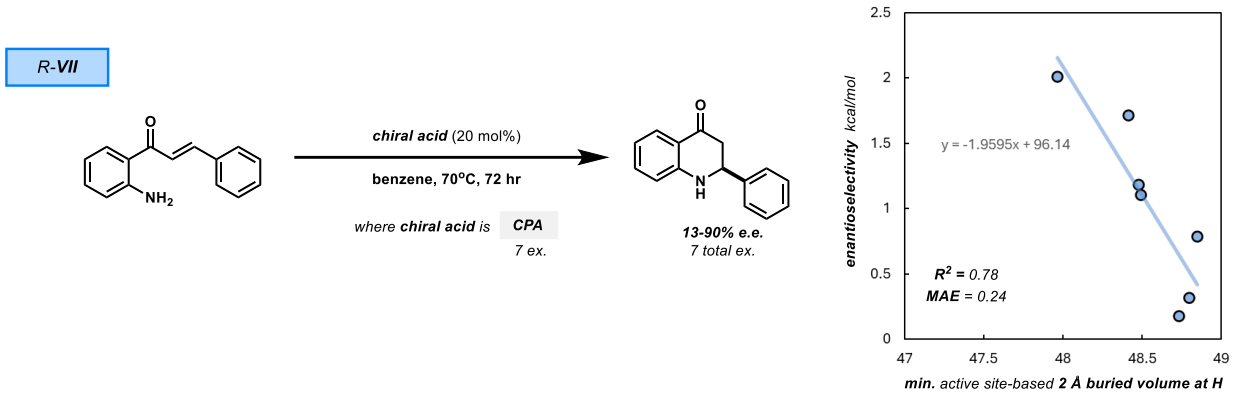
Supplementary Fig. VI-67 | R-**IV** correlation drawn by *active site-based* buried volume.



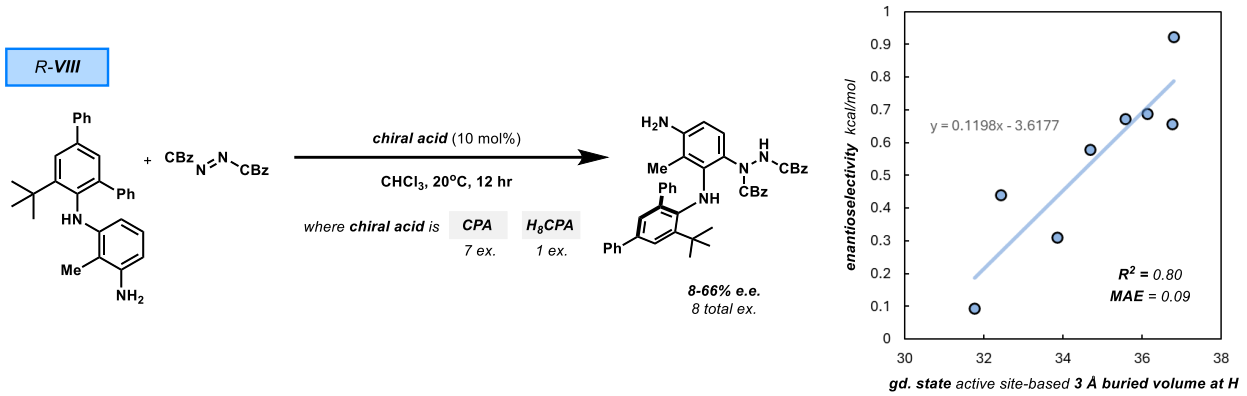
Supplementary Fig. VI-68 | R-V correlation drawn by *active site-based* buried volume.



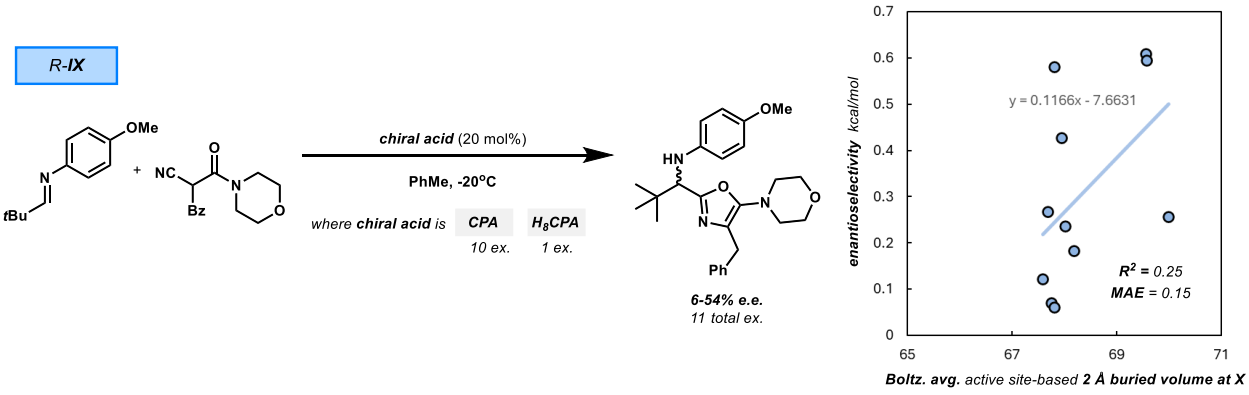
Supplementary Fig. VI-69 | R-VI correlation drawn by *active site-based* buried volume.



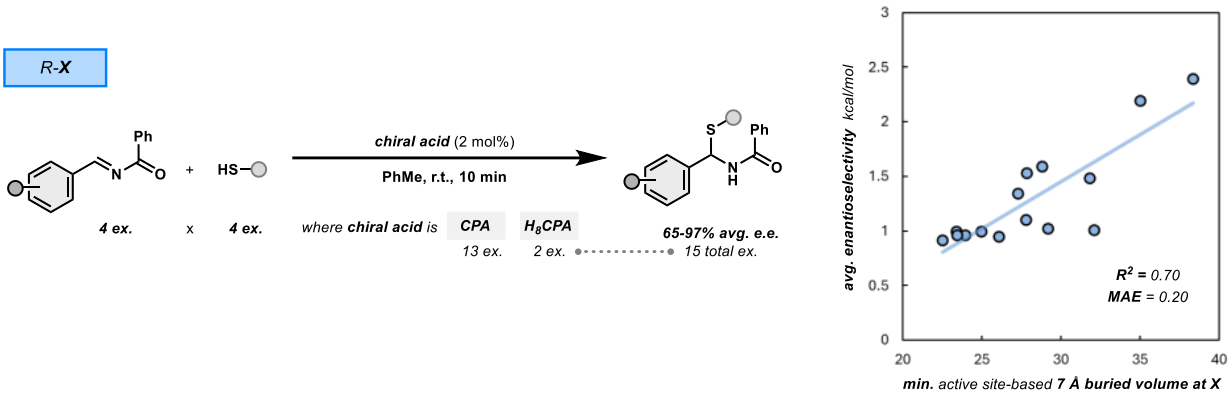
Supplementary Fig. VI-70 | R-VII correlation drawn by *active site-based* buried volume.



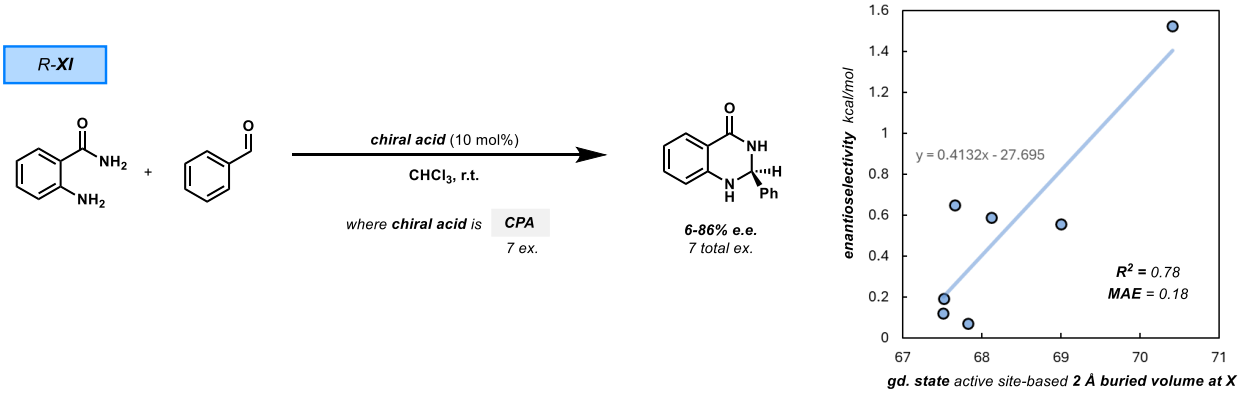
Supplementary Fig. VI-71 | R-VIII correlation drawn by *active site-based* buried volume.



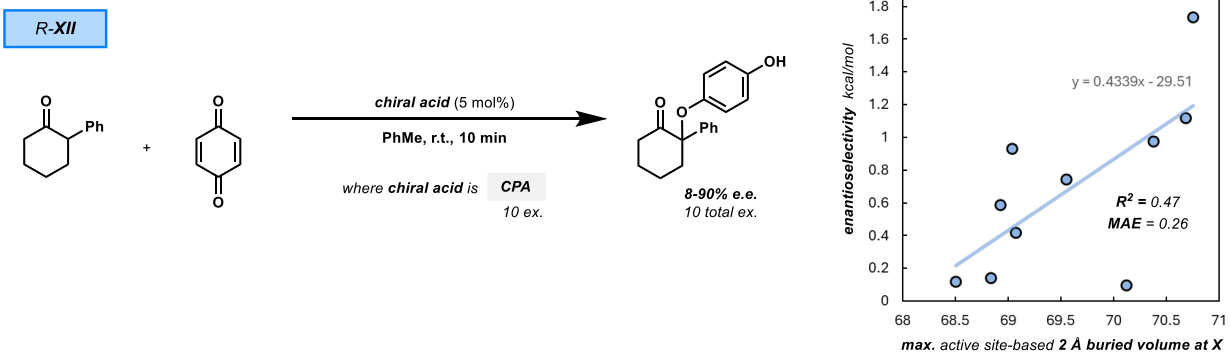
Supplementary Fig. VI-72 | R-IX correlation drawn by *active site-based* buried volume.



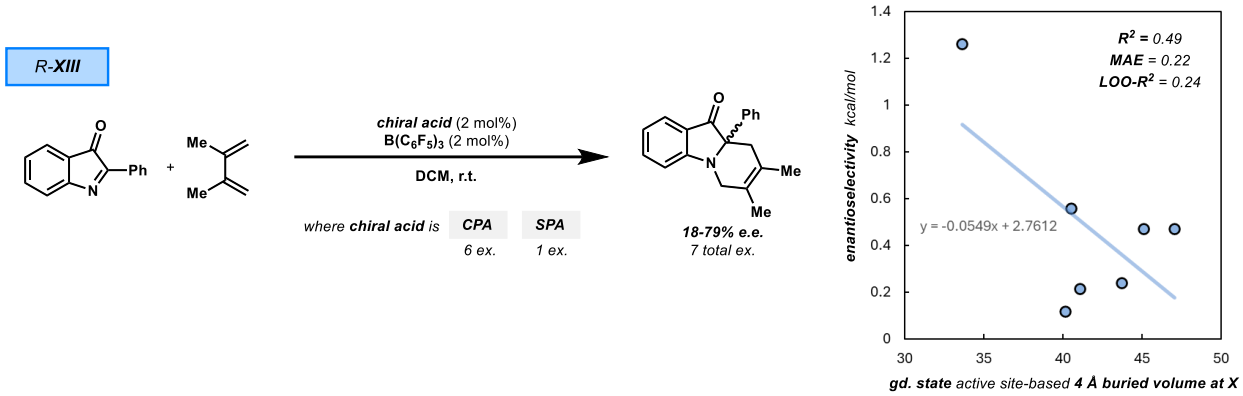
Supplementary Fig. VI-73 | R-X correlation drawn by *active site-based* buried volume.



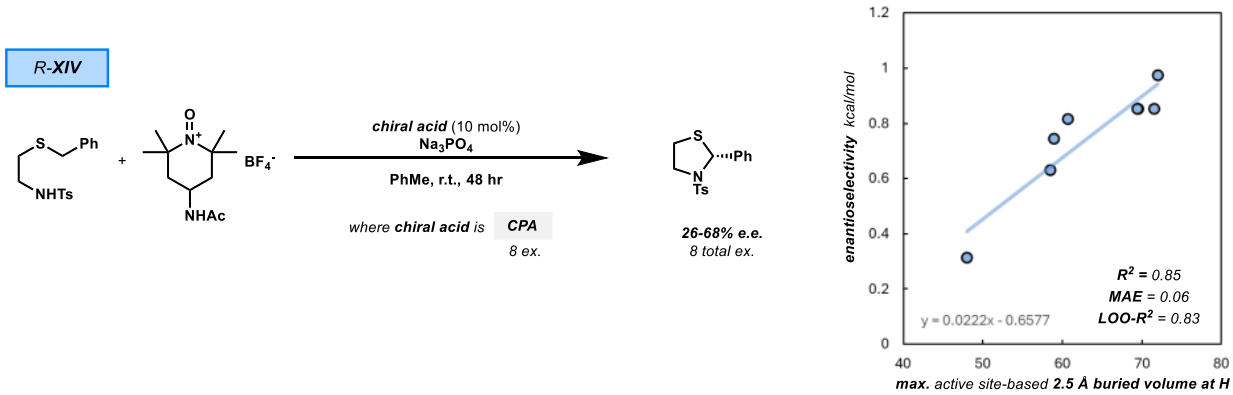
Supplementary Fig. VI-74 | R-XI correlation drawn by *active site-based* buried volume.



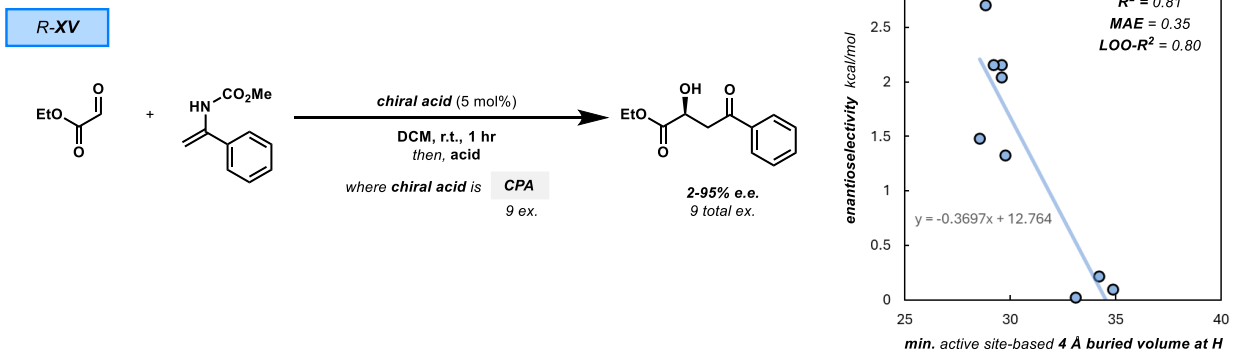
Supplementary Fig. VI-75 | R-XII correlation drawn by *active site-based* buried volume.



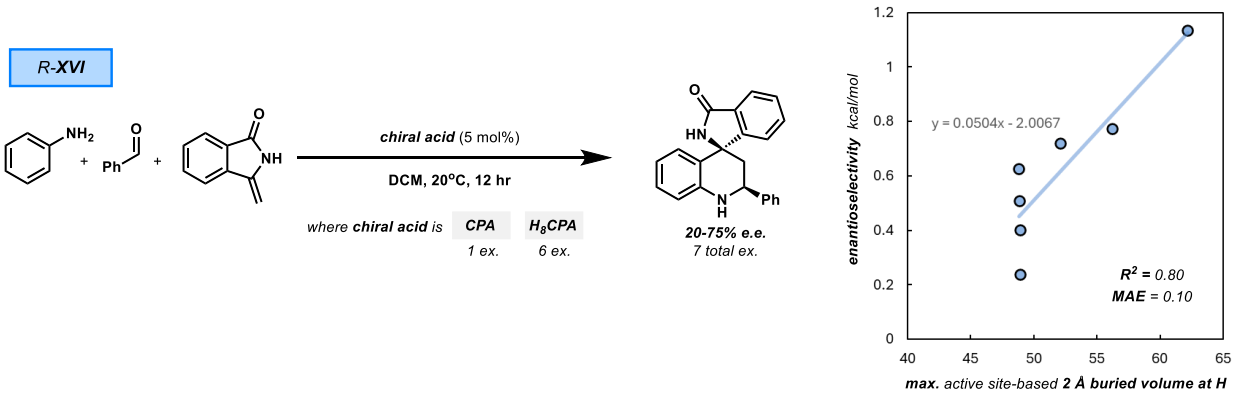
Supplementary Fig. VI-76 | R-XIII correlation drawn by *active site-based* buried volume.



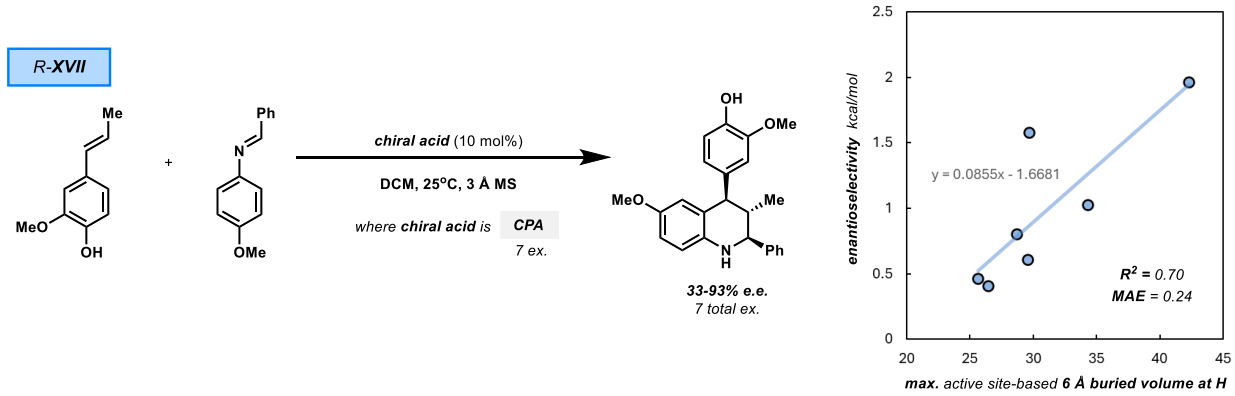
Supplementary Fig. VI-77 | R-XIV correlation drawn by *active site-based* buried volume.



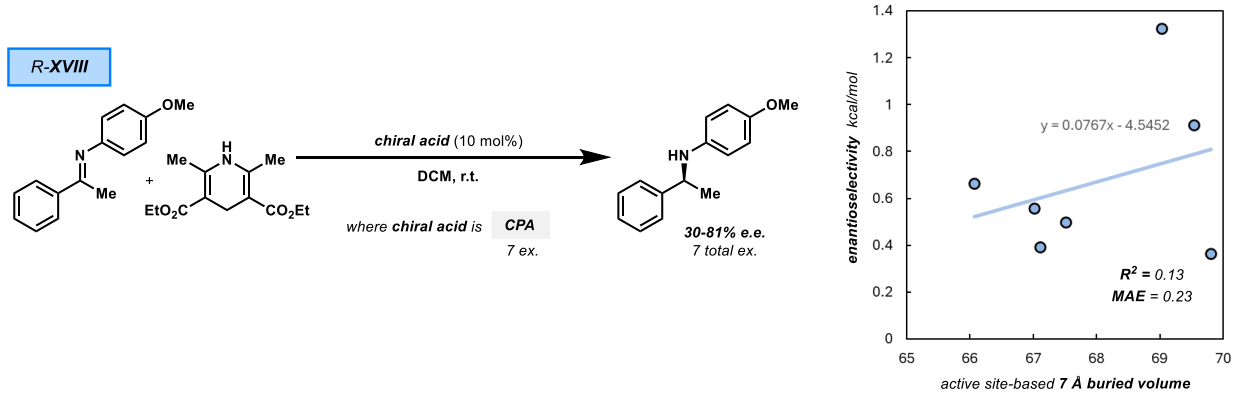
Supplementary Fig. VI-78 | R-XV correlation drawn by *active site-based* buried volume.



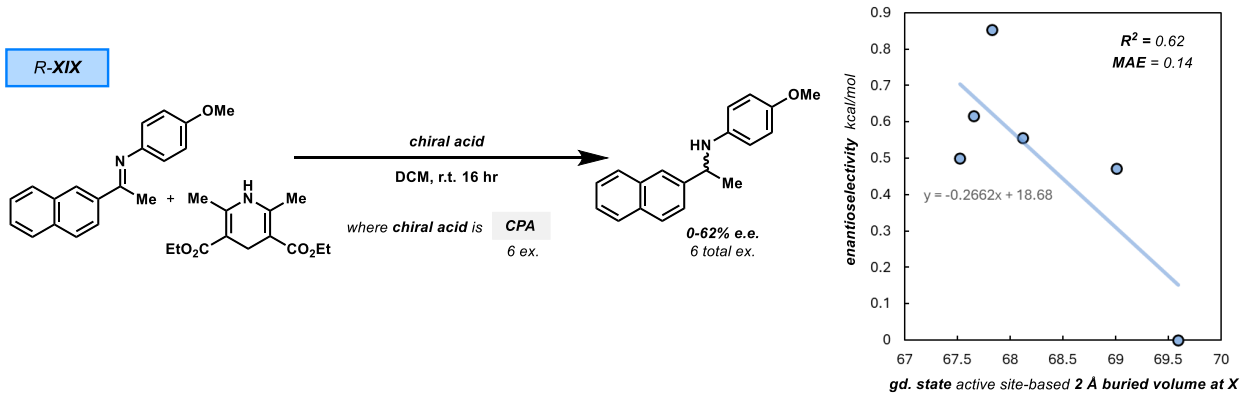
Supplementary Fig. VI-79 | R-XVI correlation drawn by *active site-based* buried volume.



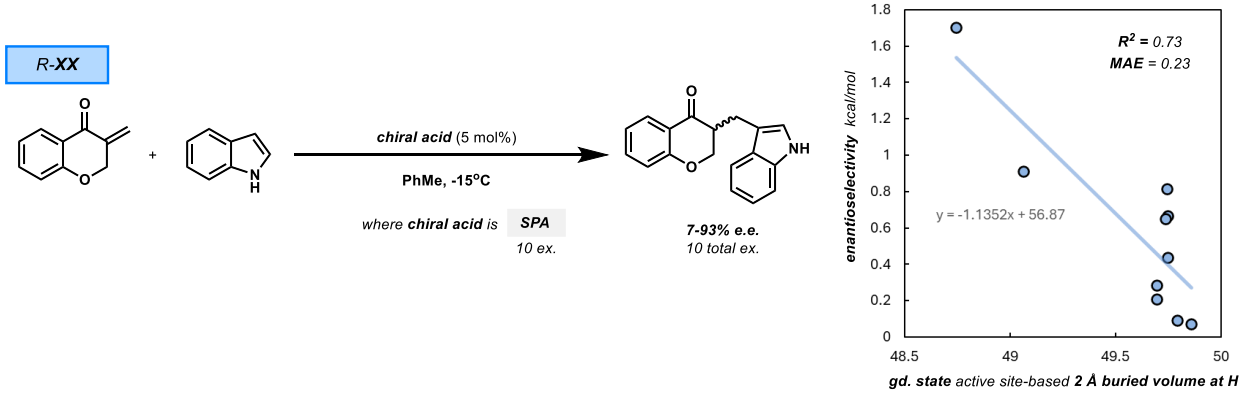
Supplementary Fig. VI-80 | R-XVII correlation drawn by *active site-based* buried volume.



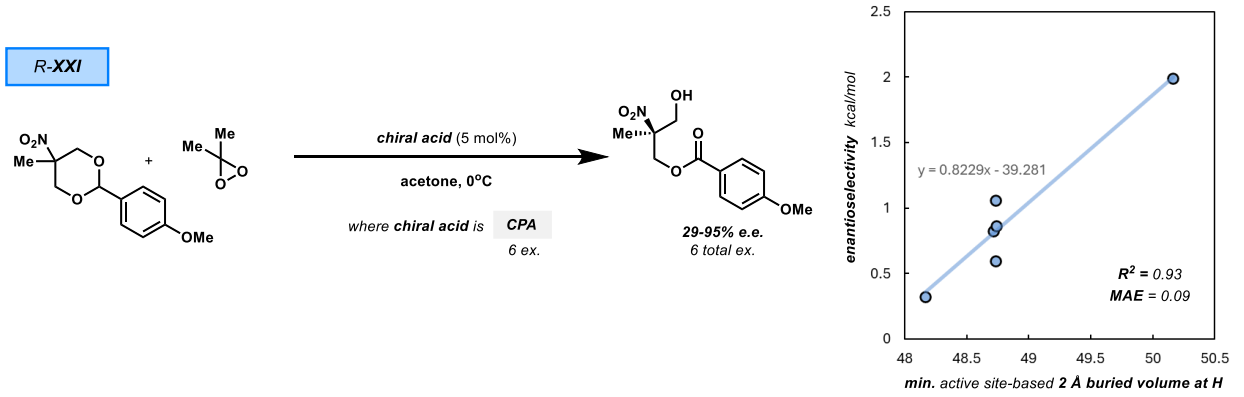
Supplementary Fig. VI-81 | R-XVIII correlation drawn by *active site-based* buried volume.



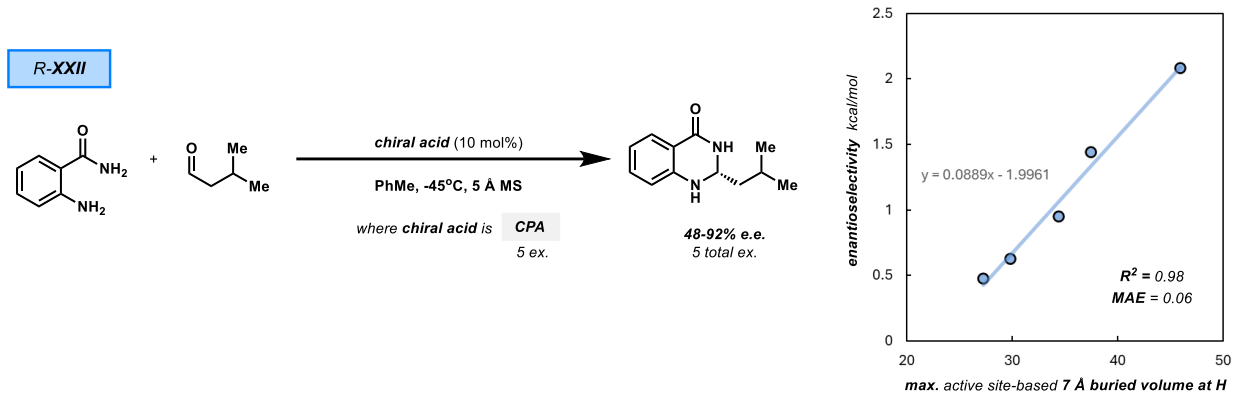
Supplementary Fig. VI-82 | R-XIX correlation drawn by *active site-based* buried volume.



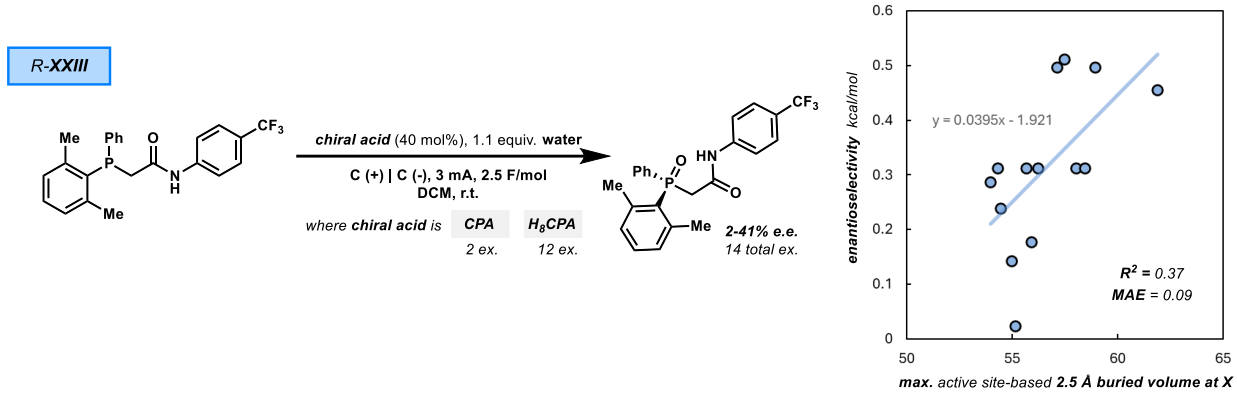
Supplementary Fig. VI-83 | R-XX correlation drawn by *active site-based* buried volume.



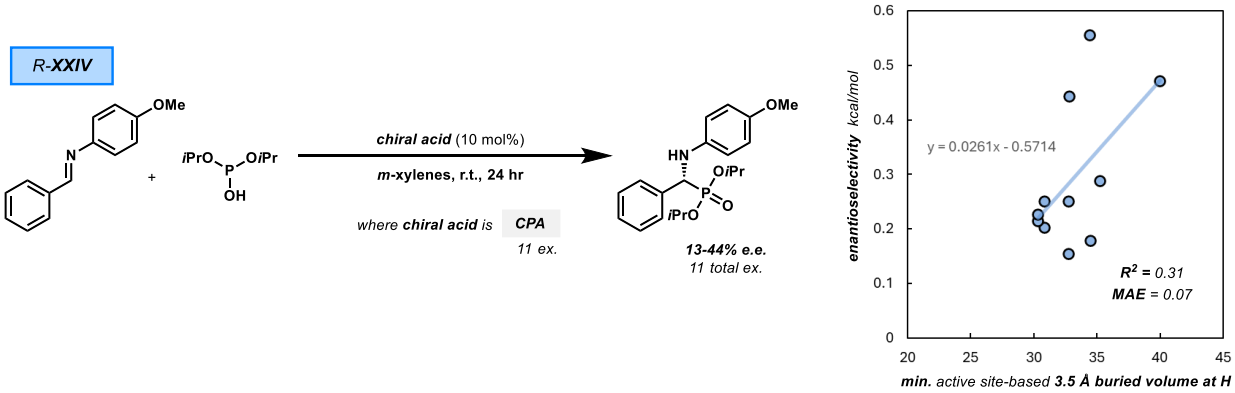
Supplementary Fig. VI-84 | R-XXI correlation drawn by *active site-based* buried volume.



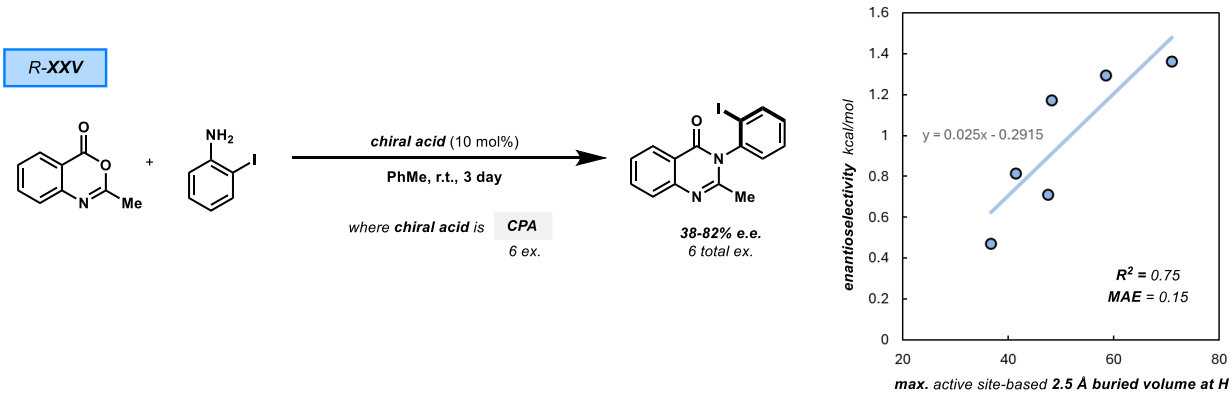
Supplementary Fig. VI-85 | R-XXII correlation drawn by *active site-based* buried volume.



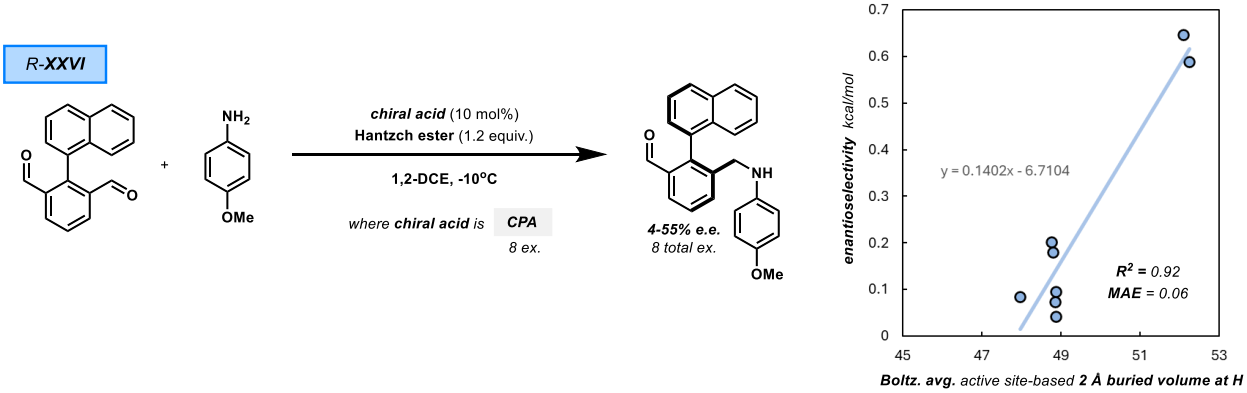
Supplementary Fig. VI-86 | R-XXIII correlation drawn by *active site-based* buried volume.



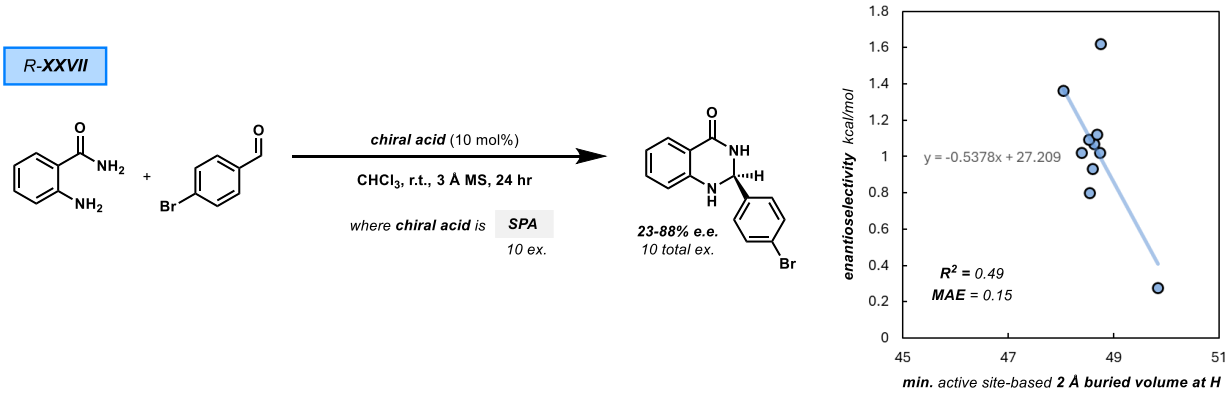
Supplementary Fig. VI-87 | R-XXIV correlation drawn by *active site-based* buried volume.



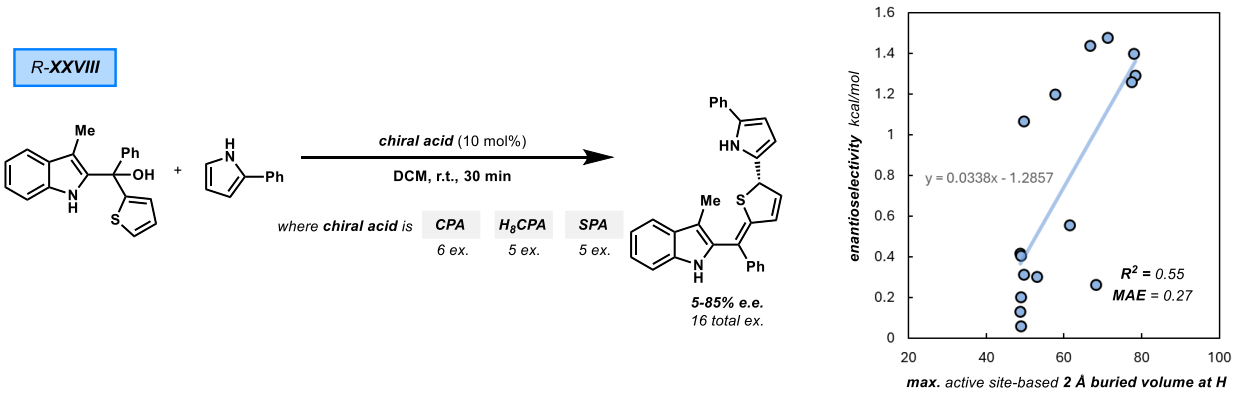
Supplementary Fig. VI-88 | R-XXV correlation drawn by *active site-based* buried volume.



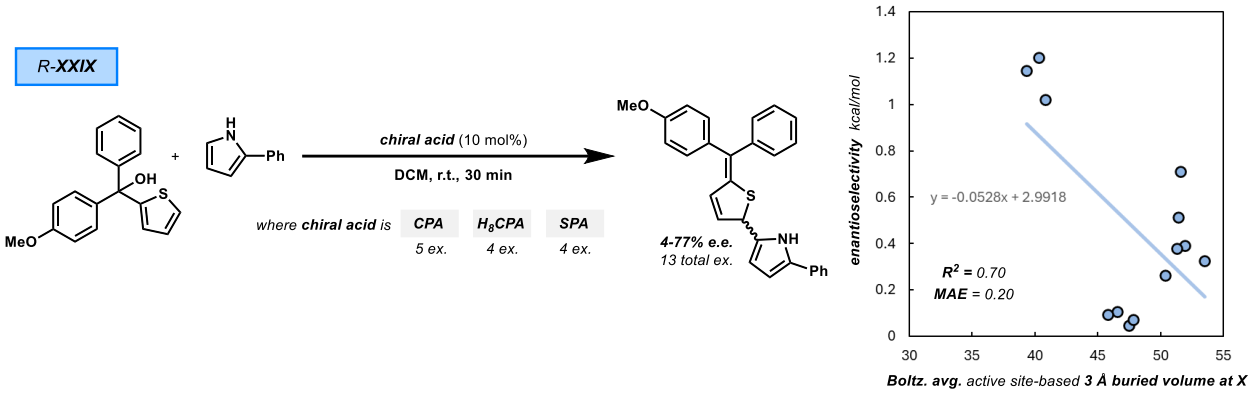
Supplementary Fig. VI-89 | R-XXVI correlation drawn by *active site-based* buried volume.



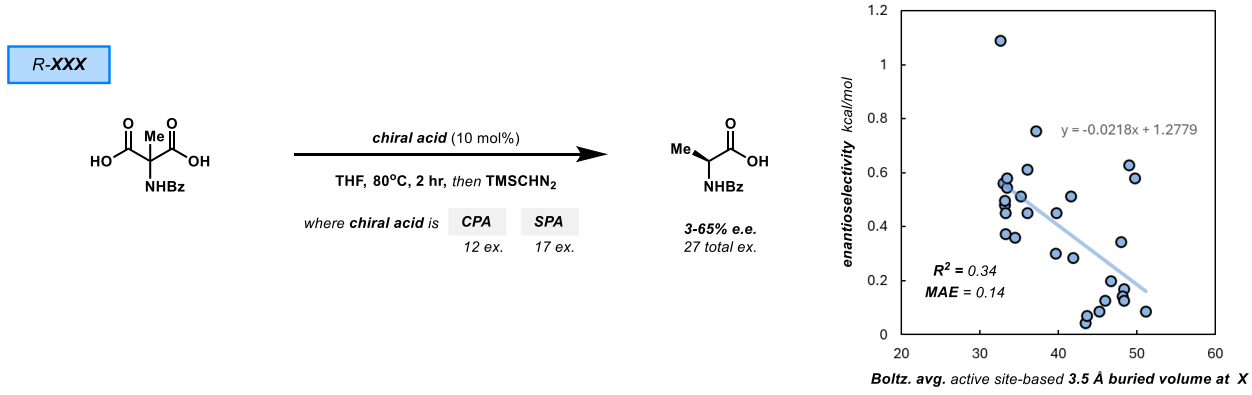
Supplementary Fig. VI-90 | R-XXVII correlation drawn by *active site-based* buried volume.



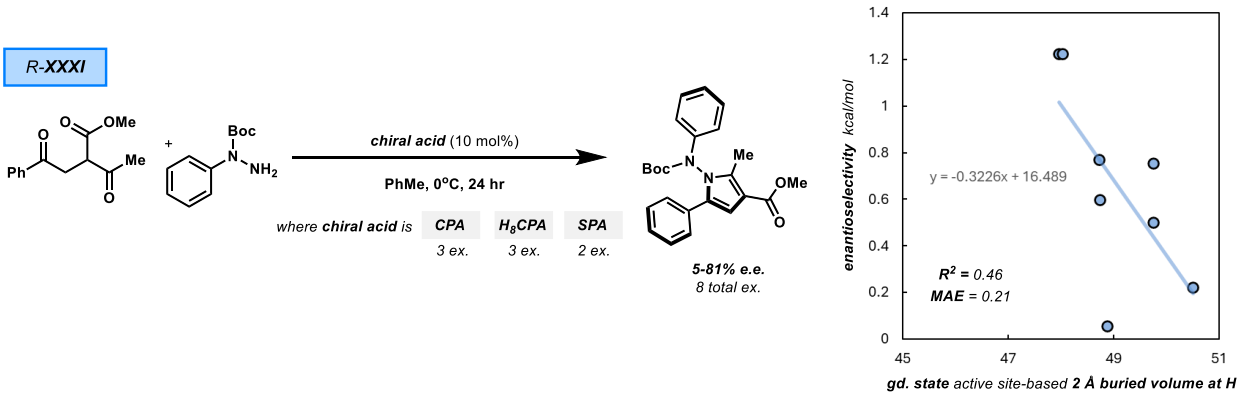
Supplementary Fig. VI-91 | R-XXVIII correlation drawn by *active site-based* buried volume.



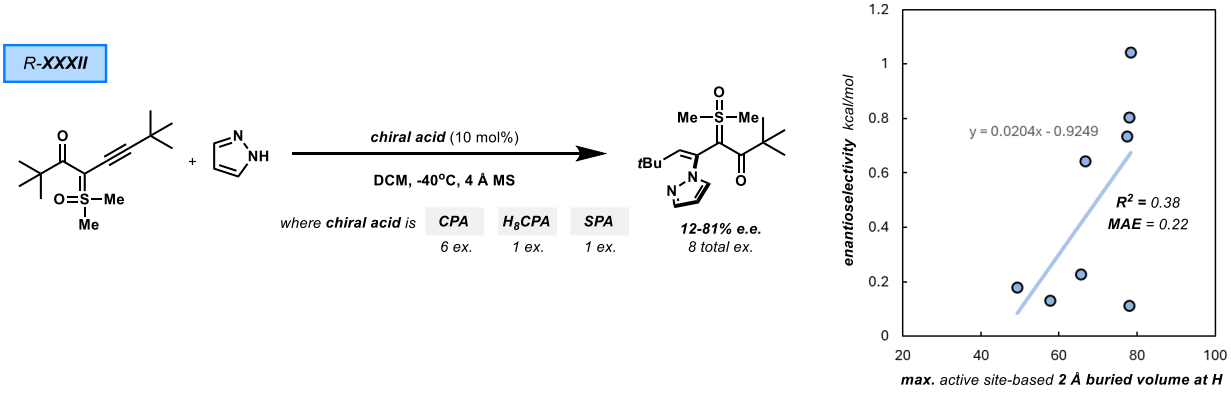
Supplementary Fig. VI-92 | R-XXIX correlation drawn by *active site-based* buried volume.



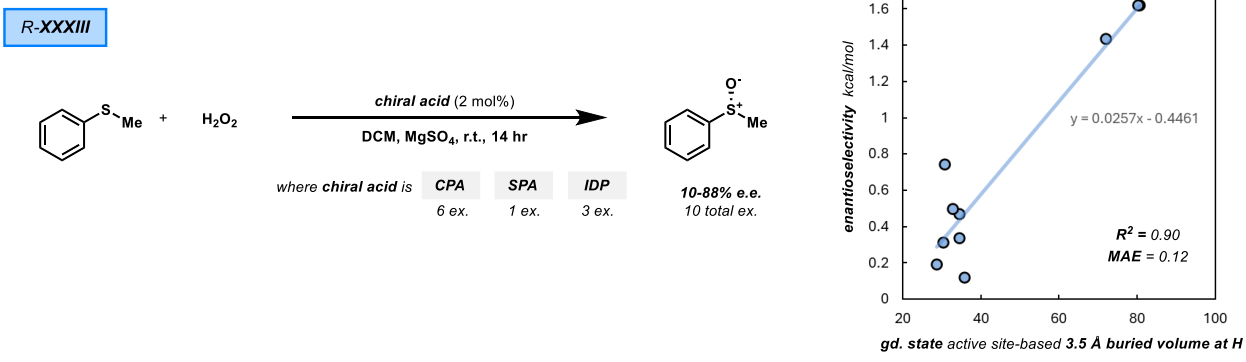
Supplementary Fig. VI-93 | R-XXX correlation drawn by *active site-based* buried volume.



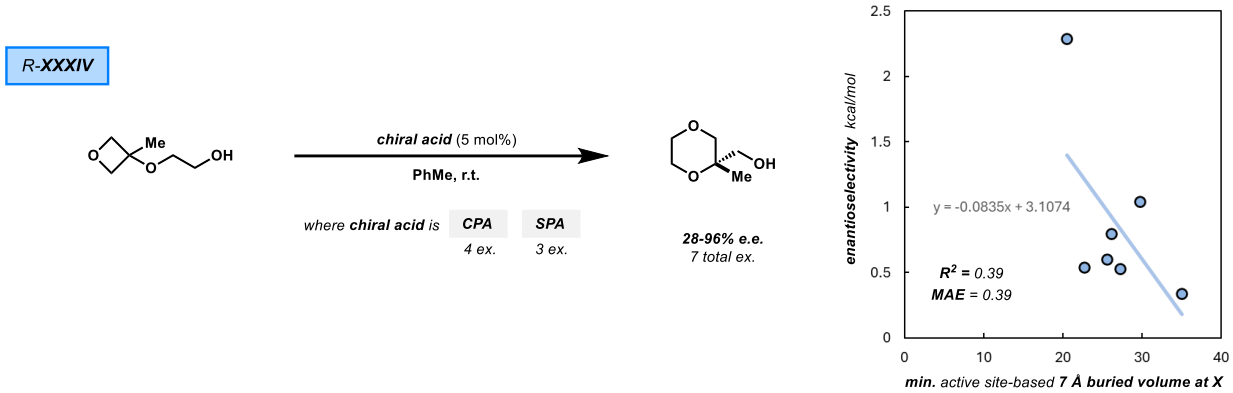
Supplementary Fig. VI-94 | R-XXXI correlation drawn by *active site-based* buried volume.



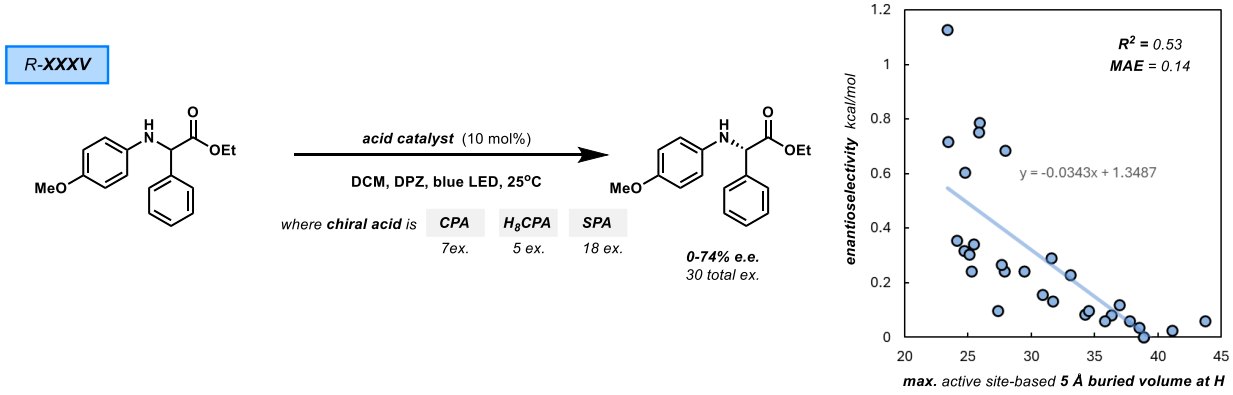
Supplementary Fig. VI-95 | R-XXXII correlation drawn by *active site-based* buried volume.



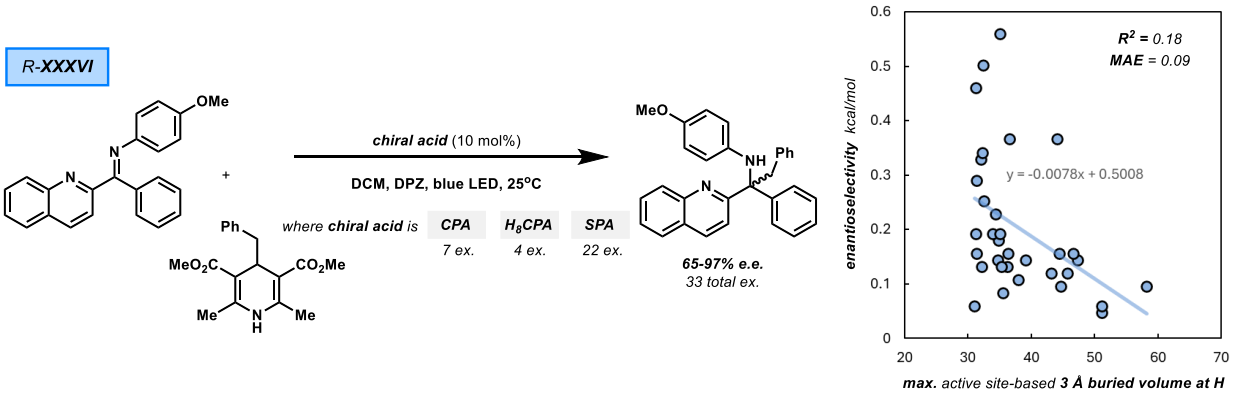
Supplementary Fig. VI-96 | R-XXXIII correlation drawn by *active site-based* buried volume.



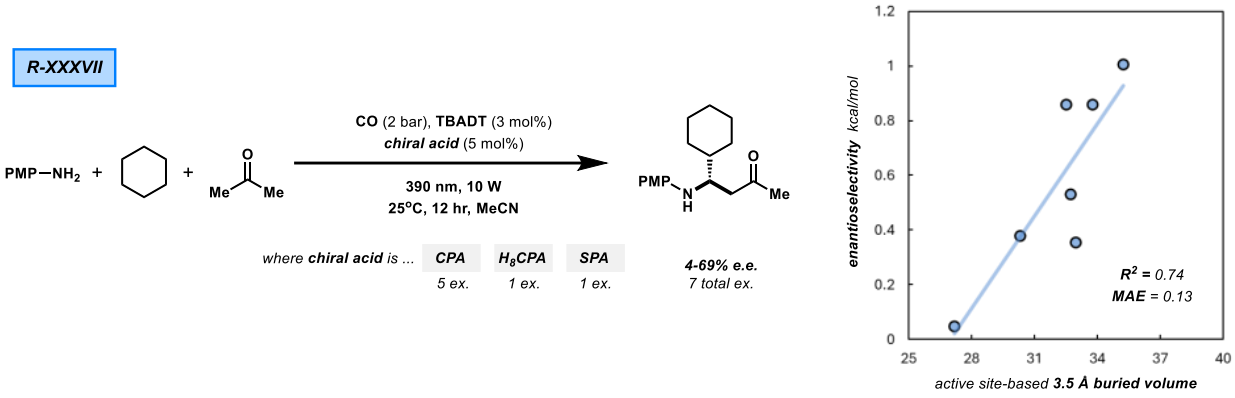
Supplementary Fig. VI-97 | R-XXXIV correlation drawn by *active site-based* buried volume.



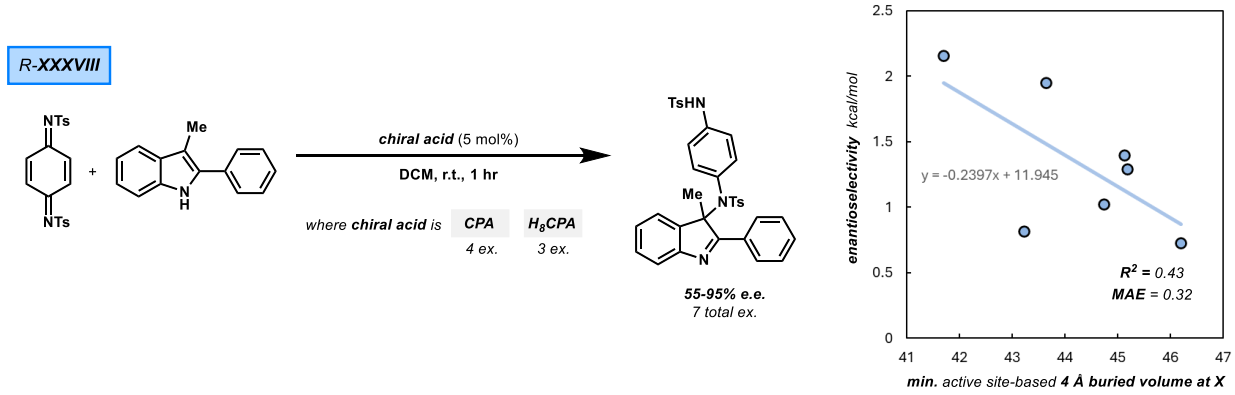
Supplementary Fig. VI-98 | R-XXXV correlation drawn by *active site-based* buried volume.



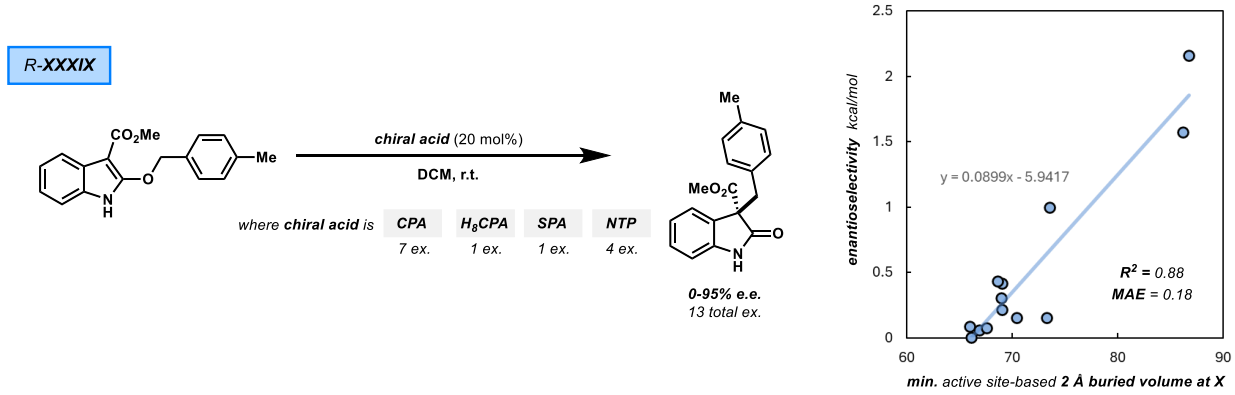
Supplementary Fig. VI-99 | R-XXXVI correlation drawn by *active site-based* buried volume.



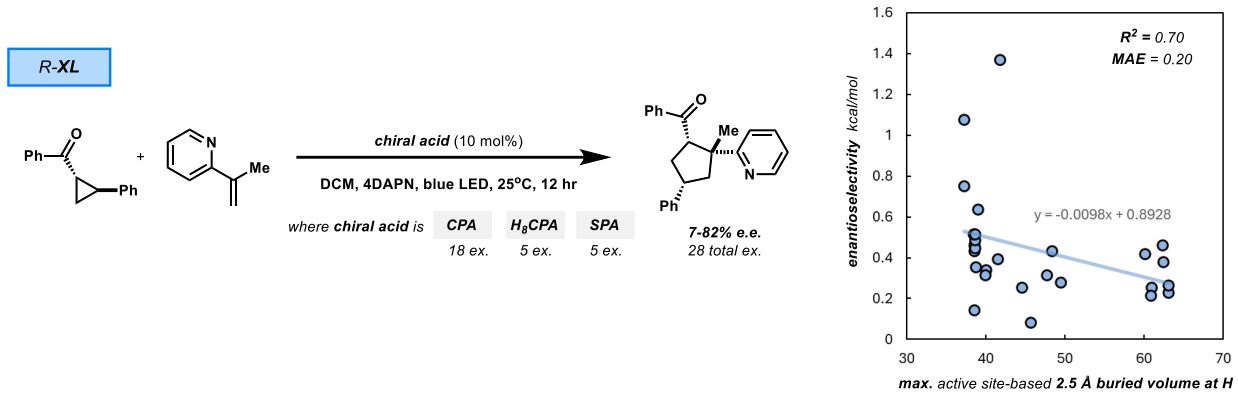
Supplementary Fig. VI-100 | R-XXXVII correlation drawn by *active site-based* buried volume.



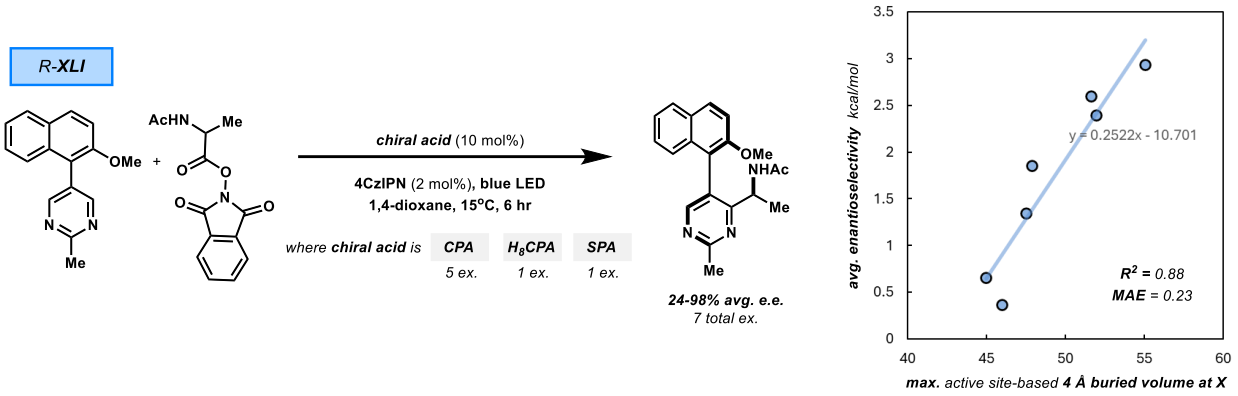
Supplementary Fig. VI-101 | R-XXXVIII correlation drawn by *active site-based* buried volume.



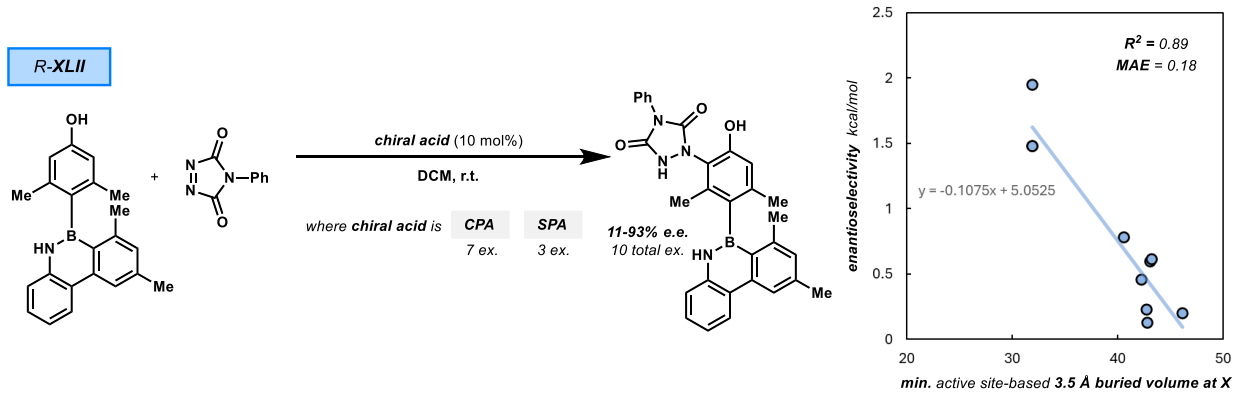
Supplementary Fig. VI-102 | R-XXXIX correlation drawn by *active site-based* buried volume.



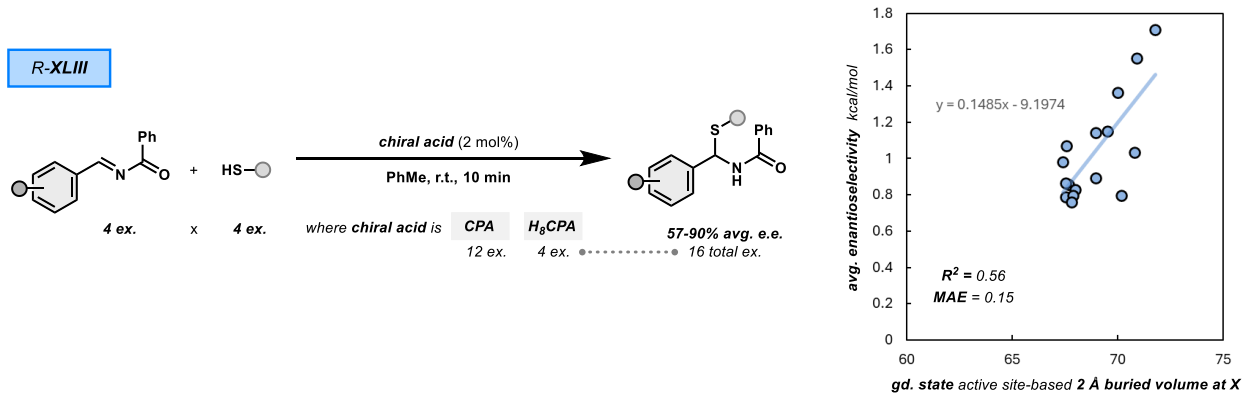
Supplementary Fig. VI-103 | R-XL correlation drawn by *active site-based* buried volume.



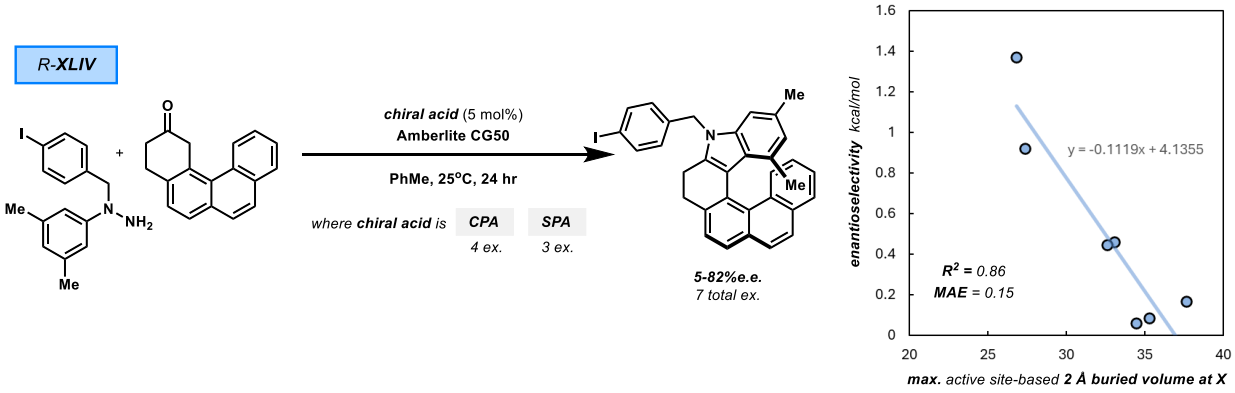
Supplementary Fig. VI-104 | R-XLI correlation drawn by *active site-based* buried volume.



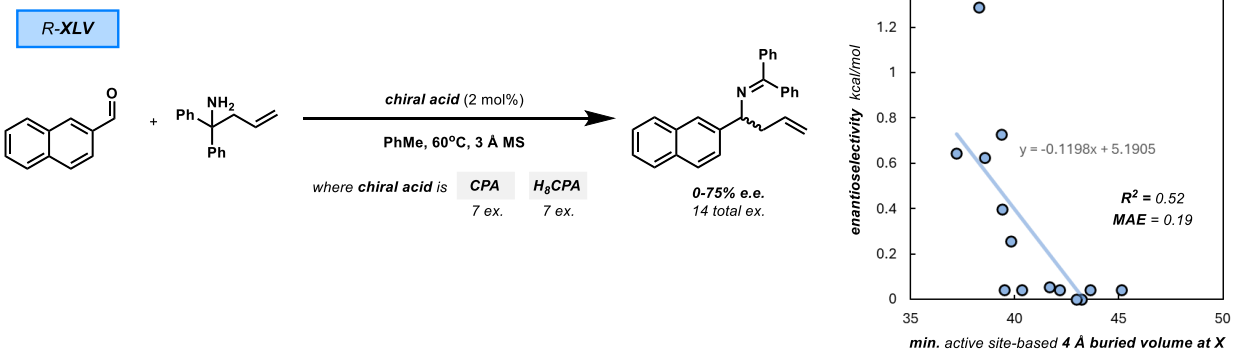
Supplementary Fig. VI-105 | R-XLII correlation drawn by *active site-based* buried volume.



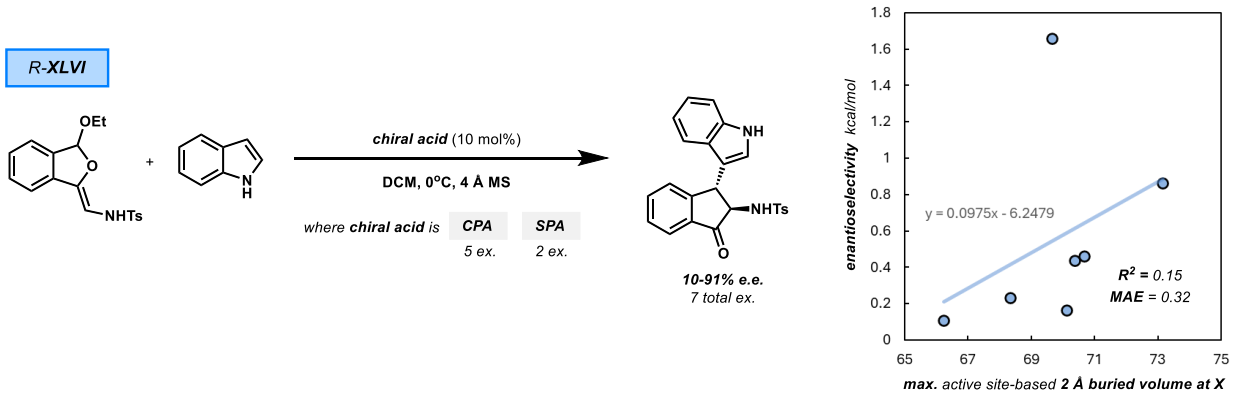
Supplementary Fig. VI-106 | R-XLIII correlation drawn by *active site-based* buried volume.



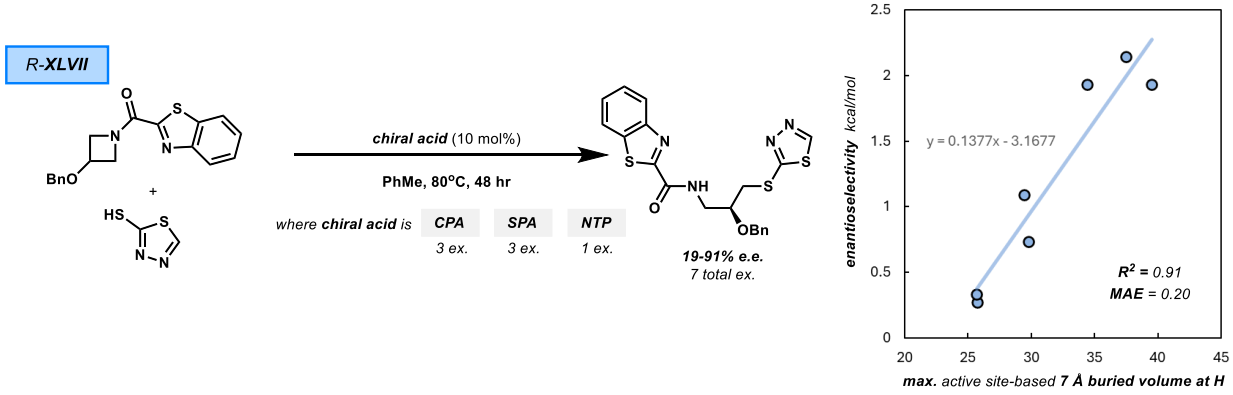
Supplementary Fig. VI-107 | R-XLIV correlation drawn by *active site-based* buried volume.



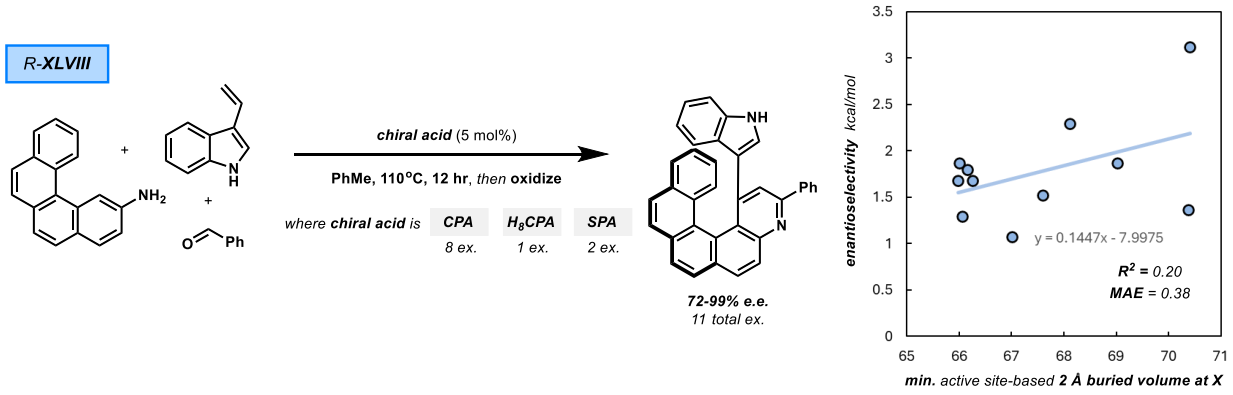
Supplementary Fig. VI-108 | R-XLV correlation drawn by *active site-based* buried volume.



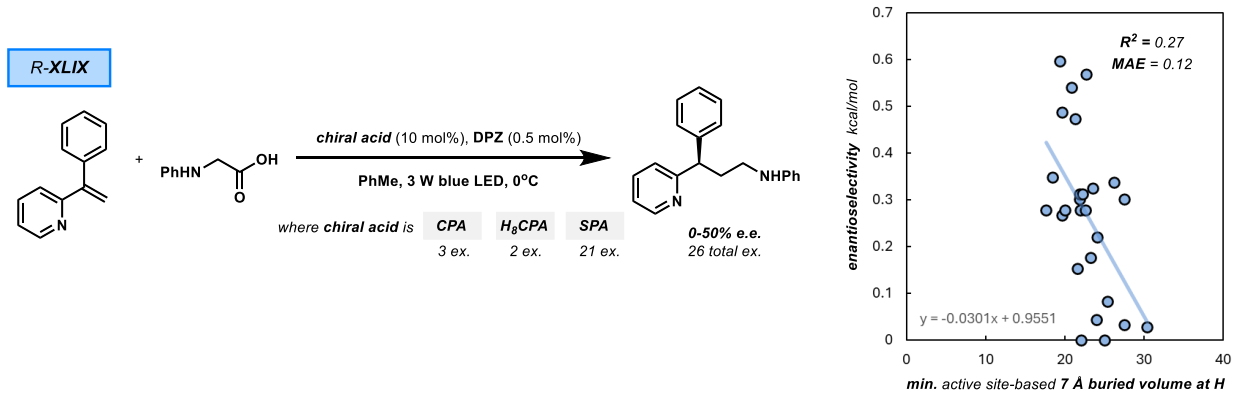
Supplementary Fig. VI-109 | R-XLVI correlation drawn by *active site-based* buried volume.



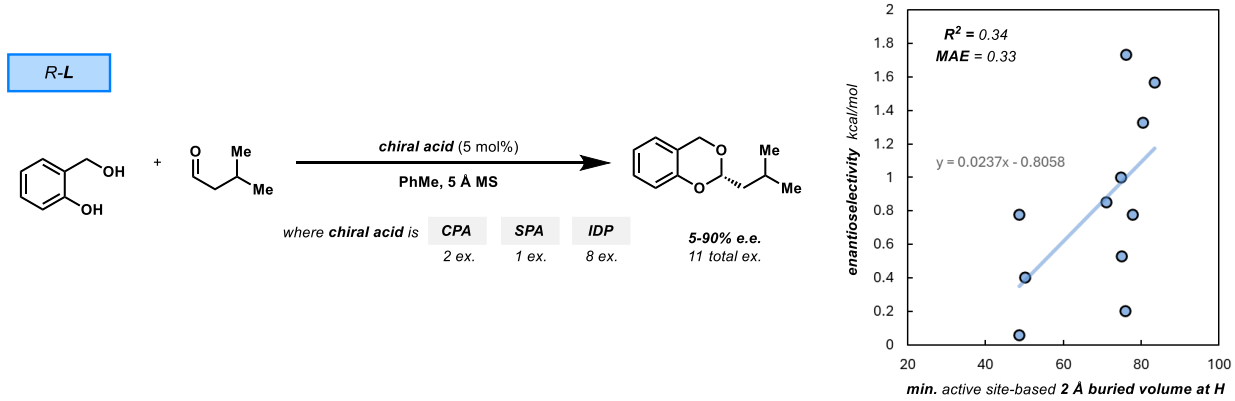
Supplementary Fig. VI-110 | R-XLVII correlation drawn by *active site-based* buried volume.



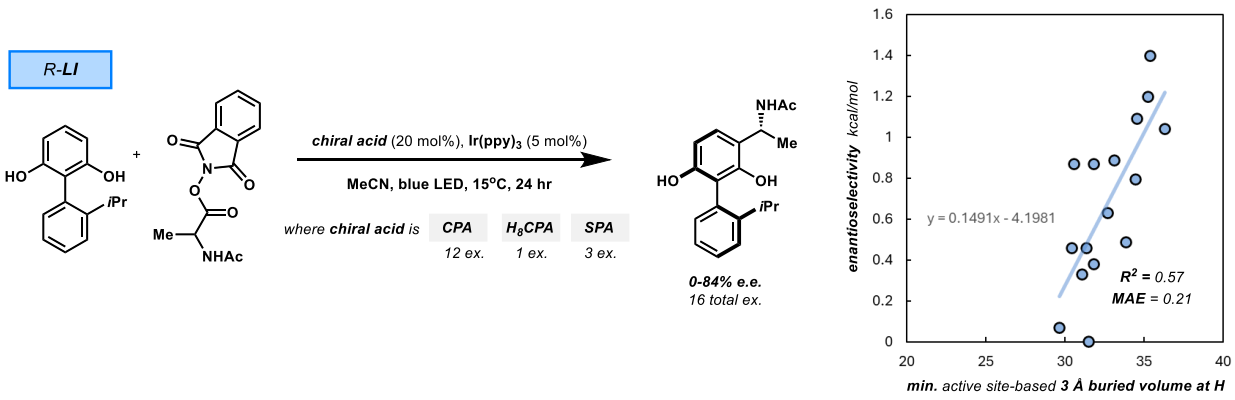
Supplementary Fig. VI-111 | R-XLVIII correlation drawn by *active site-based* buried volume.



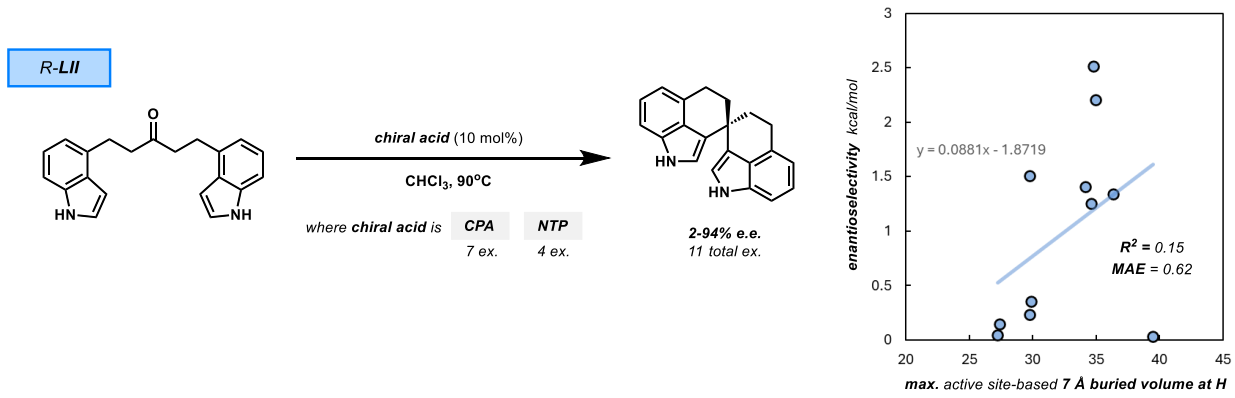
Supplementary Fig. VI-112 | R-XLIX correlation drawn by *active site-based* buried volume.



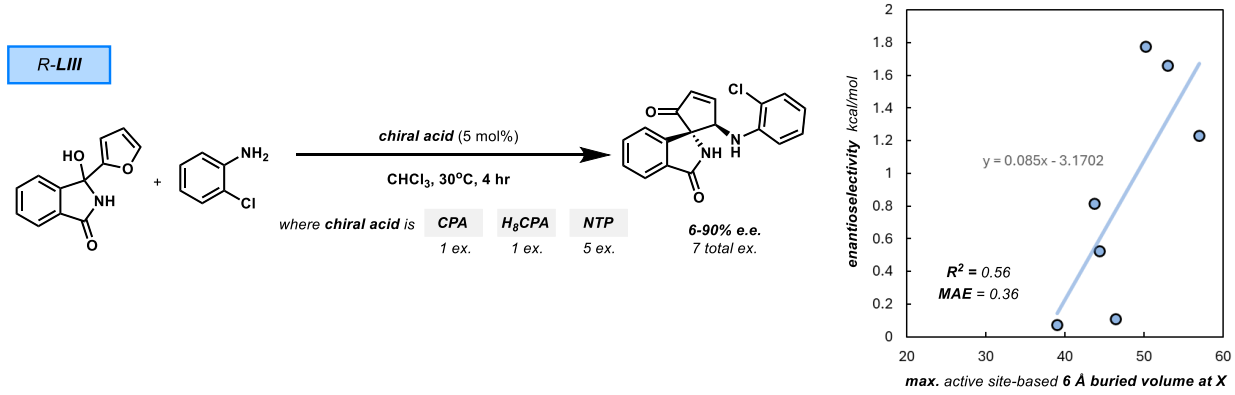
Supplementary Fig. VI-113 | R-L correlation drawn by *active site-based* buried volume.



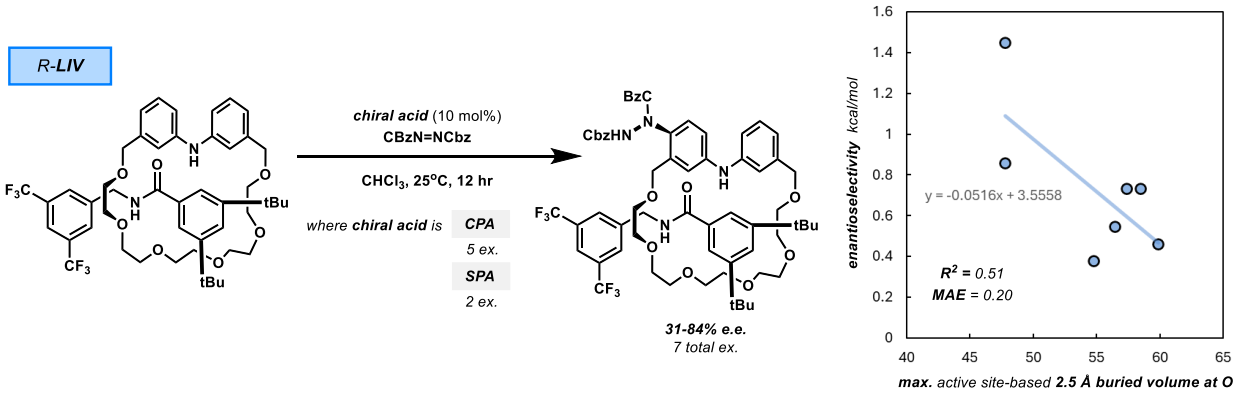
Supplementary Fig. VI-114 | R-LI correlation drawn by *active site-based* buried volume.



Supplementary Fig. VI-115 | R-LII correlation drawn by *active site-based* buried volume.



Supplementary Fig. VI-116 | R-LIII correlation drawn by *active site-based* buried volume.



Supplementary Fig. VI-117 | R-LIV correlation drawn by *active site-based* buried volume.

VI.33 – References

(1)Li, X.; Duan, M.; Yu, P.; Houk, K. N.; Sun, J. Organocatalytic Enantioselective Dearomatization of Thiophenes by 1,10-Conjugate Addition of Indole Imine Methides. *Nat Commun* **2021**, *12* (1), 4881.
(2)Wang, Y.; Wang, S.; Shan, W.; Shao, Z. Direct Asymmetric N-Propargylation of Indoles and Carbazoles Catalyzed by Lithium SPINOL Phosphate. *Nat Commun* **2020**, *11* (1), 226.
(3)Zhang, Y.-Q.; Chen, Y.-B.; Liu, J.-R.; Wu, S.-Q.; Fan, X.-Y.; Zhang, Z.-X.; Hong, X.; Ye, L.-W. Asymmetric Dearomatization Catalysed by Chiral Brønsted Acids via Activation of Ynamides. *Nat. Chem.* **2021**, *13* (11), 1093–1100.
(4)Xia, Z.-L.; Zheng, C.; Xu, R.-Q.; You, S.-L. Chiral Phosphoric Acid Catalyzed Aminative Dearomatization of α-Naphthols/Michael Addition Sequence. *Nat Commun* **2019**, *10* (1), 3150.
(5)Zheng, W.-F.; Chen, J.; Qi, X.; Huang, Z. Modular and Diverse Synthesis of Amino Acids via Asymmetric Decarboxylative Protonation of Aminomalonic Acids. *Nat. Chem.* **2023**, *15* (12), 1672–1682.
(6)Wei, Y.; Sun, F.; Li, G.; Xu, S.; Zhang, M.; Hong, L. Enantioselective Synthesis of N–N Amide–Pyrrole Atropisomers via Paal–Knorr Reaction. *Org. Lett.* **2024**, *26* (12), 2343–2348.
(7)Wu, F.; Zhang, Y.; Zhu, R.; Huang, Y. Discovery and Synthesis of Atropisomerically Chiral Acyl-Substituted Stable Vinyl Sulfoxonium Ylides. *Nat. Chem.* **2024**, *16* (1), 132–139.
(8)Wen, L.; Ding, J.; Duan, L.; Wang, S.; An, Q.; Wang, H.; Zuo, Z. Multiplicative Enhancement of Stereoenrichment by a Single Catalyst for Deracemization of Alcohols. *Science* **2023**, *382* (6669), 458–464.
(9)Liao, S.; Čorić, I.; Wang, Q.; List, B. Activation of H2O2 by Chiral Confined Brønsted Acids: A Highly Enantioselective Catalytic Sulfoxidation. *J. Am. Chem. Soc.* **2012**, *134* (26), 10765–10768.
(10)Kwon, Y.; Li, J.; Reid, J. P.; Crawford, J. M.; Jacob, R.; Sigman, M. S.; Toste, F. D.; Miller, S. J. Disparate Catalytic Scaffolds for Atroposelective Cyclodehydration. *J. Am. Chem. Soc.* **2019**, *141* (16), 6698–6705.
(11)Saito, K.; Moriya, Y.; Akiyama, T. Chiral Phosphoric Acid Catalyzed Asymmetric Synthesis of 2-Substituted 2,3-Dihydro-4-Quinolones by a Protecting-Group-Free Approach. *Org. Lett.* **2015**, *17* (13), 3202–3205.
(12)Yang, W.; Sun, J. Organocatalytic Enantioselective Synthesis of 1,4-Dioxanes and Other Oxa-Heterocycles by Oxetane Desymmetrization. *Angewandte Chemie International Edition* **2016**, *55* (5), 1868–1871.
(13)Gu, Z.; Zhang, L.; Li, H.; Cao, S.; Yin, Y.; Zhao, X.; Ban, X.; Jiang, Z. Deracemization through Sequential Photoredox-Neutral and Chiral Brønsted Acid Catalysis. *Angewandte Chemie International Edition* **2022**, *61* (49), e202211241.
(14)Song, X.; Zhang, Y.; Li, Y.; Zhao, X.; Yin, Y.; Ban, X.; Jiang, Z. Catalytic Asymmetric Synthesis of Azaarene-Functionalized Tertiary Amines and α-Amino Acid Derivatives from E/Z-Ketimine Mixtures via Enantioselective Radical Coupling. *ACS Catal.* **2023**, *13* (9), 6396–6402.

(15)Ding, W.-W.; He, Z.-Y.; Sayed, M.; Zhou, Y.; Han, Z.-Y.; Gong, L.-Z. Enantioselective Synthesis of β- and α-Amino Ketones through Reversible Alkane Carbonylation. *Nat. Synth* **2024**, *3* (4), 507–516.
(16)Subba, P.; Sahoo, S. R.; Khajuria, C.; Singh, V. K. Enantioselective Aminative Dearomatization of Indoles via Electrophilic 1,6-Addition of p-Quinone Diimides (p-QDIs). *Org. Lett.* **2024**, *26* (23), 4932–4937.
(17)Yu, H.; Hu, L.; Zhang, J.; Yang, Q.; Lu, G.; Xu, T. Organocatalytic Highly Enantioselective Formal [1,3] Sigmatropic Rearrangement of Indole Alkyl Ethers and Mechanistic Insight. *ACS Catal.* **2024**, 9207–9216.
(18)Wei, W.; Li, C.; Fan, Y.; Chen, X.; Zhao, X.; Qiao, B.; Jiang, Z. Catalytic Asymmetric Redox-Neutral [3+2] Photocycloadditions of Cyclopropyl Ketones with Vinylazaarenes Enabled by Consecutive Photoinduced Electron Transfer. *Angewandte Chemie International Edition* **2024**, *63* (30), e202406845.
(19)Liang, D.; Chen, J.-R.; Tan, L.-P.; He, Z.-W.; Xiao, W.-J. Catalytic Asymmetric Construction of Axially and Centrally Chiral Heterobiaryls by Minisci Reaction. *J. Am. Chem. Soc.* **2022**, *144* (13), 6040–6049.
(20)Yang, J.; Zhang, J.-W.; Bao, W.; Qiu, S.-Q.; Li, S.; Xiang, S.-H.; Song, J.; Zhang, J.; Tan, B. Chiral Phosphoric Acid-Catalyzed Remote Control of Axial Chirality at Boron–Carbon Bond. *J. Am. Chem. Soc.* **2021**, *143* (33), 12924–12929.
(21)Ye, Z.; Xie, W.; Wang, D.; Liu, H.; Yang, X. Atroposelective Synthesis of Diarylamines via Organocatalyzed Electrophilic Amination. *ACS Catal.* **2024**, *14* (7), 4958–4967.
(22)Yue, T.; Wang, M.-X.; Wang, D.-X.; Masson, G.; Zhu, J. Brønsted Acid Catalyzed Enantioselective Three-Component Reaction Involving the α Addition of Isocyanides to Imines. *Angewandte Chemie International Edition* **2009**, *48* (36), 6717–6721.
(23)Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631.
(24)Rueping, M.; Antonchick, A. P.; Sugiono, E.; Grenader, K. Asymmetric Brønsted Acid Catalysis: Catalytic Enantioselective Synthesis of Highly Biologically Active Dihydroquinazolinones. *Angewandte Chemie International Edition* **2009**, *48* (5), 908–910.
(25)Shevchenko, G. A.; Oppelaar, B.; List, B. An Unexpected α-Oxidation of Cyclic Ketones with 1,4-Benzoquinone by Enol Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (33), 10756–10759.
(26)Kötzner, L.; Webber, M. J.; Martínez, A.; De Fusco, C.; List, B. Asymmetric Catalysis on the Nanoscale: The Organocatalytic Approach to Helicenes. *Angewandte Chemie International Edition* **2014**, *53* (20), 5202–5205.
(27)Rueping, M.; Antonchick, A. P. Catalytic Asymmetric Aminoallylation of Aldehydes: A Catalytic Enantioselective Aza-Cope Rearrangement. *Angewandte Chemie International Edition* **2008**, *47* (52), 10090–10093.

(28)Zhao, Q.; Li, Y.; Zhang, Q.-X.; Cheng, J.-P.; Li, X. Catalytic Asymmetric Aza-Diels–Alder Reaction of Ketimines and Unactivated Dienes. *Angewandte Chemie International Edition* **2021**, *60* (32), 17608–17614.
(29)Biswas, S.; Kubota, K.; Orlandi, M.; Turberg, M.; Miles, D. H.; Sigman, M. S.; Toste, F. D. Enantioselective Synthesis of N,S-Acetals by an Oxidative Pummerer-Type Transformation Using Phase-Transfer Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (2), 589–593.
(30)Terada, M.; Soga, K.; Momiyama, N. Enantioselective Activation of Aldehydes by Chiral Phosphoric Acid Catalysts in an Aza-Ene-Type Reaction between Glyoxylate and Enecarbamate. *Angewandte Chemie International Edition* **2008**, *47* (22), 4122–4125.
(31)Li, G.; Yao, Y.; Wang, Z.; Zhao, M.; Xu, J.; Huang, L.; Zhu, G.; Bao, G.; Sun, W.; Hong, L.; Wang, R. Switchable Skeletal Rearrangement of Dihydroisobenzofuran Acetals with Indoles. *Org. Lett.* **2019**, *21* (11), 4313–4317.
(32)Wang, Z.; Sheong, F. K.; Sung, H. H. Y.; Williams, I. D.; Lin, Z.; Sun, J. Catalytic Enantioselective Intermolecular Desymmetrization of Azetidines. *J. Am. Chem. Soc.* **2015**, *137* (18), 5895–5898.
(33)Mo, N.-F.; Zhang, Y.; Guan, Z.-H. Highly Enantioselective Three-Component Povarov Reaction for Direct Construction of Azaspirocycles. *Org. Lett.* **2022**, *24* (35), 6397–6401.
(34)He, L.; Bekkaye, M.; Retailleau, P.; Masson, G. Chiral Phosphoric Acid Catalyzed Inverse-Electron-Demand Aza-Diels–Alder Reaction of Isoeugenol Derivatives. *Org. Lett.* **2012**, *14* (12), 3158–3161.
(35)Li, C.; Shao, Y.-B.; Gao, X.; Ren, Z.; Guo, C.; Li, M.; Li, X. Enantioselective Synthesis of Chiral Quinohelicenes through Sequential Organocatalyzed Povarov Reaction and Oxidative Aromatization. *Nat Commun* **2023**, *14* (1), 3380.
(36)Hoffmann, S.; Seayad, A. M.; List, B. A Powerful Brønsted Acid Catalyst for the Organocatalytic Asymmetric Transfer Hydrogenation of Imines. *Angewandte Chemie International Edition* **2005**, *44* (45), 7424–7427.
(37)Rueping, M.; Sugiono, E.; Azap, C.; Theissmann, T.; Bolte, M. Enantioselective Brønsted Acid Catalyzed Transfer Hydrogenation: Organocatalytic Reduction of Imines. *Org. Lett.* **2005**, *7* (17), 3781–3783.
(38)Li, Y.-P.; Li, Z.-Q.; Zhou, B.; Li, M.-L.; Xue, X.-S.; Zhu, S.-F.; Zhou, Q.-L. Chiral Spiro Phosphoric Acid-Catalyzed Friedel–Crafts Conjugate Addition/Enantioselective Protonation Reactions. *ACS Catal.* **2019**, *9* (7), 6522–6529.
(39)Meng, S.-S.; Tang, W.-B.; Zheng, W.-H. Catalytically Enantioselective Synthesis of Acyclic α-Tertiary Amines through Desymmetrization of 2-Substituted 2-Nitro-1,3-Diols. *Org. Lett.* **2018**, *20* (3), 518–521.
(40)Yin, Y.; Dai, Y.; Jia, H.; Li, J.; Bu, L.; Qiao, B.; Zhao, X.; Jiang, Z. Conjugate Addition–Enantioselective Protonation of N-Aryl Glycines to α-Branched 2-Vinylazaarenes via Cooperative Photoredox and Asymmetric Catalysis. *J. Am. Chem. Soc.* **2018**, *140* (19), 6083–6087.
(41)Cheng, X.; Vellalath, S.; Goddard, R.; List, B. Direct Catalytic Asymmetric Synthesis of Cyclic Aminals from Aldehydes. *J. Am. Chem. Soc.* **2008**, *130* (47), 15786–15787.

(42)Kim, J. H.; Čorić, I.; Vellalath, S.; List, B. The Catalytic Asymmetric Acetalization. *Angewandte Chemie International Edition* **2013**, *52* (16), 4474–4477.
(43)Mao, K.; Liu, C.; Wang, Y.; Gu, C.; Putziger, J. M.; Cemalovic, N. I.; Muniz, C.; Qi, Y.; Lin, S. Dynamic Kinetic Resolution of Phosphines with Chiral Supporting Electrolytes. *Nature* **2025**, *643* (8074), 1288–1296.
(44)Sanocki, M.; Russell, H. C.; Handjaya, J.; Reid, J. P. Relative Generality and Risk: Quantitative Measures for Broad Catalyst Success. *ACS Catal.* **2024**, *14* (22), 16849–16860.
(45)Moon, J.; Shin, E.; Kwon, Y. Enantioselective Desymmetrization of Biaryls via Cooperative Photoredox/Brønsted Acid Catalysis and Its Application to the Total Synthesis of Ancistrobrevolines. *J. Am. Chem. Soc.* **2025**, *147* (15), 12800–12810.
(46)Zhao, H.-W.; Jiang, F.; Chen, S.; Hu, J.; Xiang, S.-H.; Ding, W.-Y.; Lu, W.; Tan, B. Organocatalytic Asymmetric Construction and Application of Axially Chiral Spiro-Bisindoles. *Angewandte Chemie International Edition* **2025**, *64* (12), e202422951.
(47)Guo, M. C.; Miller, S. J. Catalyst–Substrate Pairings for Carbocyclic and Heterocyclic Systems in Atroposelective Quinazolinone Synthesis. *ACS Catal.* **2024**, *14* (23), 17226–17232.
(48)Yu, L.; Tang, Y.; Nie, X.; Cai, Y. Stereoselective Access to Spiro-Isoindolinone Scaffolds via Catalytic Asymmetric Aza-Piancatelli Rearrangement. *Org. Lett.* **2024**, *26* (36), 7667–7671.
(49)Wang, Y.; Song, R.-P.; Li, X.-Y.; Chen, W.-L.; Tian, Y.; Zhang, S.-H.; Shao, Y.-D.; Cheng, D.-J. Catalytic Asymmetric Reductive Amination for Axially Chiral Aryl Aldehydes via Desymmetrization/Kinetic Resolution Cascade. *Org. Lett.* **2024**, *26* (34), 7161–7165.
(50)Huang, D.; Li, X.; Xu, F.; Li, L.; Lin, X. Highly Enantioselective Synthesis of Dihydroquinazolinones Catalyzed by SPINOL-Phosphoric Acids. *ACS Catal.* **2013**, *3* (10), 2244–2247.
(51)Tang, M.; Zhou, J.; Xie, W.; Ren, J.; Ye, Z.; Gu, H.; Yang, X. Catalytic Enantioselective Synthesis of Mechanically Planar Chiral Rotaxanes by Organocatalyzed Desymmetrization. *Chem* **2025**, 102694.
(52)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.
(53)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.
(54)*Gaussian 16 | Gaussian.com*.
(55)Greg Landrum; Paolo Tosco; Brian Kelley; Ricardo Rodriguez; David Cosgrove; Riccardo Vianello; sriniker; gedeck; Gareth Jones; NadineSchneider; Eisuke Kawashima; Dan Nealschneider; Andrew Dalke; Matt Swain; Brian Cole; Samo Turk; Aleksandr Savelev; Alain Vaucher; Maciej Wójcikowski; Ichiru Take; Vincent F. Scalfani; Rachel Walker; Kazuya Ujihara; Daniel Probst; guillaume godin; Axel Pahl; Juuso Lehtivarjo; Francois Berenger; jasondbiggs; strets123. Rdkit/Rdkit: 2024_03_1 (Q1 2024) Release, 2024.

(56)Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11* (10), 872–879.
(57)Treacy, S. M.; Smith, A. L.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science. *J. Am. Chem. Soc.* **2024**, *146* (43), 29792–29800.
(58)Guilian Luchini; Patterson, T.; Paton, R. Patonlab/DBSTEP: Release 1.1.0, 2023.
(59)Kariofillis, S. K.; Jiang, S.; Żurański, A. M.; Gandhi, S. S.; Martinez Alvarado, J. I.; Doyle, A. G. Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources. *J. Am. Chem. Soc.* **2022**, *144* (2), 1045–1055.
(60)Kütt, A.; Tshepelevitsh, S.; Saame, J.; Lõkov, M.; Kaljurand, I.; Selberg, S.; Leito, I. Strengths of Acids in Acetonitrile. *European Journal of Organic Chemistry* **2021**, *2021* (9), 1407–1419.

Appendix IV – Supplementary Information for Chapter IV

Portions of this appendix were adapted from the Supplementary Information of:

Smith, A. L.; Le, C.; Toste, F. D. *Manuscript in revision.*

Table of Contents:

IV.1 – Specific Comments on Materials and Methods 293
IV.2 – Detailed Statistical Analyses 298
IV.3 – Experimental Benchmarks and Datasets 326
IV.4 – Empirical Bias Powers Discovery of Stereoselective Solutions 340
IV.5 – Synthetic Protocols 353
IV.6 – References 398
IV.7 – NMR Spectra 403

VI.34 – Specific Comments on Materials and Methods

- ***Photochemistry Setup for R-IV***

1-dram vials were placed immediately below the aperture of a 456 nm Kessil blue LED lamp and mounted atop solid dry ice submerged in acetone (**Supplementary Fig. IV-1**). If needed, the samples were adjusted with forceps to maximize irradiation. Every 60-90 minutes, the dry ice-acetone bath was replenished. While some frosting on the surface of the glass was observed, duplicate samples displayed consistent reaction outcomes (ex. yield, enantioselectivity).

Supplementary Fig. VI-118 | Photochemistry setup for irradiation of two samples at 456 nm light.

- ***Specific Calculation Details***

Consistent with our earlier approach, we generate initial coordinates for BAOCs leveraging *templates* to ensure the axial chirality is conserved. *Templates* for traditional BAOCs (ex. CPA, NTP) are included in our previous report. We include below *templates* for P2-SIA scaffold reported in the main text. Because P2-SIA (in the acid form) is chiral at *S*, we explicitly compute two configurations corresponding to (*R,R,R)* and (*R,S,R)* respectively. Experimental nor statistical data can identify if one diastereomer is *more meaningful* than the other. As such, we report statistical analyses from P2SIARH-based descriptors.

P2SIAR_H

C	0.88872103	0.21932295	-0.06941144	C	9.79174259	0.14347407	4.25135016
C	0.84438745	-1.04602043	-0.56397060	H	10.48726066	2.16843730	4.07170821
C	0.15635054	-2.05673797	0.12984923	C	7.80827371	-0.39363981	2.98454174
C	-0.41251901	-1.79256914	1.37864309	O	7.08596767	1.14452863	1.20639943
C	-0.28528609	-0.51846150	1.93665607	C	8.81282143	-0.78025541	3.87507398
C	0.25442852	0.51934442	1.14338424	C	10.78901497	-0.24080752	5.15548133
H	0.47488569	-3.55143998	-1.37579900	C	6.61076996	-1.34463961	2.70369997
H	1.39663801	0.98865577	-0.61262457	C	8.83874585	-2.07508568	4.40487105
H	1.32972158	-1.27241451	-1.49031170	H	11.54202915	0.45959904	5.45103777
C	0.04129354	-3.33852158	-0.42104807	C	10.80080544	-1.50201620	5.65712483
C	-1.08838666	-2.80835741	2.06365319	C	5.30792731	-0.84168104	2.48490716
C	-1.19153804	-4.04337078	1.51153934	C	6.80983851	-2.72688058	2.67397549
C	-0.62046008	-4.31131051	0.25600640	O	5.01450113	0.52310161	2.82451555
H	-1.52313530	-2.61148652	3.02131728	O	4.59767860	1.54530126	0.39190617
H	-1.71083123	-4.81831535	2.03566391	C	9.81533954	-2.42903220	5.27777444
H	-0.70772198	-5.28923769	-0.16934927	H	8.08965588	-2.78490475	4.12214952
C	-0.04547529	0.63091660	4.24351296	H	11.56588300	-1.79233095	6.34653715
C	-0.74409242	-0.26246005	3.39973966	C	4.30359634	-1.68372506	1.98959464
C	-1.86169898	-0.92950826	3.90723706	C	5.77136924	-3.56284344	2.25469425
C	-2.36772027	-0.57017082	5.15938874	C	8.04103218	-3.27946646	3.04353152
C	-1.75259838	0.44928712	5.90668080	H	9.83505748	-3.42033368	5.68006205
C	-0.61070021	1.03556500	5.45993875	C	4.53304140	-3.01822077	1.87250494
H	-2.10526226	-2.21656835	2.20807746	H	3.35188675	-1.27840337	1.71586090
C	-2.48717322	-1.94010194	3.16863953	C	5.97717295	-4.94680130	2.21174322
C	-3.49669410	-1.22925601	5.66006057	H	8.84018689	-2.64307725	3.36174161
H	-2.18183698	0.76055103	6.83607351	C	8.21984568	-4.62357890	2.99757172
H	-0.13648659	1.79679550	6.04346289	H	3.76417791	-3.65898440	1.49406801
C	-4.08902661	-2.20949537	4.93147238	C	7.17683935	-5.46612183	2.57727572
C	-3.57888180	-2.56879471	3.67246164	H	5.18888964	-5.59525211	1.89079283
H	-3.89267319	-0.96039765	6.61700800	H	9.16115688	-5.04465519	3.28294085
H	-4.95071379	-2.71207311	5.31853144	H	7.32945379	-6.52467979	2.54499629
H	-4.05451921	-3.34294996	3.10737478	P	5.52491837	1.60673818	1.63997482
O	0.09825209	1.88279642	1.56881937	N	5.50752626	3.23310912	2.28538225
O	1.26734287	1.07578404	3.86519640	O	0.77988403	3.66284025	3.40816851
P	1.22400167	2.32929139	2.74037692	O	3.77276434	3.59913422	4.40076305
S	3.91217621	3.65185072	2.73742746	N	2.81244369	2.54449837	2.03761734
C	8.85282447	1.80183506	2.79393917	H	8.85556596	2.79176716	2.38779524
C	9.77119849	1.44955244	3.73194988	**Bq**	**3.49740551**	**5.50863214**	**2.08902325**
C	7.91360335	0.86322542	2.34661645	H	3.07672256	2.02061957	1.22785955

P2SIAS_H

C	0.88872103	0.21932295	-0.06941144	C	-0.41251901	-1.79256914	1.37864309
C	0.84438745	-1.04602043	-0.56397060	C	-0.28528609	-0.51846150	1.93665607
C	0.15635054	-2.05673797	0.12984923	C	0.25442852	0.51934442	1.14338424

H	0.47488569	-3.55143998	-1.37579900	O	7.08596767	1.14452863	1.20639943
H	1.39663801	0.98865577	-0.61262457	C	8.81282143	-0.78025541	3.87507398
H	1.32972158	-1.27241451	-1.49031170	C	10.78901497	-0.24080752	5.15548133
C	0.04129354	-3.33852158	-0.42104807	C	6.61076996	-1.34463961	2.70369997
C	-1.08838666	-2.80835741	2.06365319	C	8.83874585	-2.07508568	4.40487105
C	-1.19153804	-4.04337078	1.51153934	H	11.54202915	0.45959904	5.45103777
C	-0.62046008	-4.31131051	0.25600640	C	10.80080544	-1.50201620	5.65712483
H	-1.52313530	-2.61148652	3.02131728	C	5.30792731	-0.84168104	2.48490716
H	-1.71083123	-4.81831535	2.03566391	C	6.80983851	-2.72688058	2.67397549
H	-0.70772198	-5.28923769	-0.16934927	O	5.01450113	0.52310161	2.82451555
C	-0.04547529	0.63091660	4.24351296	O	4.59767860	1.54530126	0.39190617
C	-0.74409242	-0.26246005	3.39973966	C	9.81533954	-2.42903220	5.27777444
C	-1.86169898	-0.92950826	3.90723706	H	8.08965588	-2.78490475	4.12214952
C	-2.36772027	-0.57017082	5.15938874	H	11.56588300	-1.79233095	6.34653715
C	-1.75259838	0.44928712	5.90668080	C	4.30359634	-1.68372506	1.98959464
C	-0.61070021	1.03556500	5.45993875	C	5.77136924	-3.56284344	2.25469425
H	-2.10526226	-2.21656835	2.20807746	C	8.04103218	-3.27946646	3.04353152
C	-2.48717322	-1.94010194	3.16863953	H	9.83505748	-3.42033368	5.68006205
C	-3.49669410	-1.22925601	5.66006057	C	4.53304140	-3.01822077	1.87250494
H	-2.18183698	0.76055103	6.83607351	H	3.35188675	-1.27840337	1.71586090
H	-0.13648659	1.79679550	6.04346289	C	5.97717295	-4.94680130	2.21174322
C	-4.08902661	-2.20949537	4.93147238	H	8.84018689	-2.64307725	3.36174161
C	-3.57888180	-2.56879471	3.67246164	C	8.21984568	-4.62357890	2.99757172
H	-3.89267319	-0.96039765	6.61700800	H	3.76417791	-3.65898440	1.49406801
H	-4.95071379	-2.71207311	5.31853144	C	7.17683935	-5.46612183	2.57727572
H	-4.05451921	-3.34294996	3.10737478	H	5.18888964	-5.59525211	1.89079283
O	0.09825209	1.88279642	1.56881937	H	9.16115688	-5.04465519	3.28294085
O	1.26734287	1.07578404	3.86519640	H	7.32945379	-6.52467979	2.54499629
P	1.22400167	2.32929139	2.74037692	P	5.52491837	1.60673818	1.63997482
S	3.91217621	3.65185072	2.73742746	N	5.50752626	3.23310912	2.28538225
C	8.85282447	1.80183506	2.79393917	O	0.77988403	3.66284025	3.40816851
C	9.77119849	1.44955244	3.73194988	O	3.77276434	3.59913422	4.40076305
C	7.91360335	0.86322542	2.34661645	N	2.81244369	2.54449837	2.03761734
C	9.79174259	0.14347407	4.25135016	H	8.85556596	2.79176716	2.38779524
H	10.48726066	2.16843730	4.07170821	H	6.30737966	3.82642970	2.37596498
C	7.80827371	-0.39363981	2.98454174	**Bq**	**3.49740551**	**5.50863214**	**2.08902325**

P2SIA_CB

C	0.88872103	0.21932295	-0.06941144	C	-2.48717322	-1.94010194	3.16863953
C	0.84438745	-1.04602043	-0.56397060	C	-3.49669410	-1.22925601	5.66006057
C	0.15635054	-2.05673797	0.12984923	H	-2.18183698	0.76055103	6.83607351
C	-0.41251901	-1.79256914	1.37864309	H	-0.13648659	1.79679550	6.04346289
C	-0.28528609	-0.51846150	1.93665607	C	-4.08902661	-2.20949537	4.93147238
C	0.25442852	0.51934442	1.14338424	C	-3.57888180	-2.56879471	3.67246164
H	0.47488569	-3.55143998	-1.37579900	H	-3.89267319	-0.96039765	6.61700800
H	1.39663801	0.98865577	-0.61262457	H	-4.95071379	-2.71207311	5.31853144
H	1.32972158	-1.27241451	-1.49031170	H	-4.05451921	-3.34294996	3.10737478
C	0.04129354	-3.33852158	-0.42104807	O	0.09825209	1.88279642	1.56881937
C	-1.08838666	-2.80835741	2.06365319	O	1.26734287	1.07578404	3.86519640
C	-1.19153804	-4.04337078	1.51153934	P	1.22400167	2.32929139	2.74037692
C	-0.62046008	-4.31131051	0.25600640	S	3.91217621	3.65185072	2.73742746
H	-1.52313530	-2.61148652	3.02131728	C	8.85282447	1.80183506	2.79393917
H	-1.71083123	-4.81831535	2.03566391	C	9.77119849	1.44955244	3.73194988
H	-0.70772198	-5.28923769	-0.16934927	C	7.91360335	0.86322542	2.34661645
C	-0.04547529	0.63091660	4.24351296	C	9.79174259	0.14347407	4.25135016
C	-0.74409242	-0.26246005	3.39973966	H	10.48726066	2.16843730	4.07170821
C	-1.86169898	-0.92950826	3.90723706	C	7.80827371	-0.39363981	2.98454174
C	-2.36772027	-0.57017082	5.15938874	O	7.08596767	1.14452863	1.20639943
C	-1.75259838	0.44928712	5.90668080	C	8.81282143	-0.78025541	3.87507398
C	-0.61070021	1.03556500	5.45993875	C	10.78901497	-0.24080752	5.15548133
H	-2.10526226	-2.21656835	2.20807746	C	6.61076996	-1.34463961	2.70369997

C	8.83874585	-2.07508568	4.40487105
H	11.54202915	0.45959904	5.45103777
C	10.80080544	-1.50201620	5.65712483
C	5.30792731	-0.84168104	2.48490716
C	6.80983851	-2.72688058	2.67397549
O	5.01450113	0.52310161	2.82451555
O	4.59767860	1.54530126	0.39190617
C	9.81533954	-2.42903220	5.27777444
H	8.08965588	-2.78490475	4.12214952
H	11.56588300	-1.79233095	6.34653715
C	4.30359634	-1.68372506	1.98959464
C	5.77136924	-3.56284344	2.25469425
C	8.04103218	-3.27946646	3.04353152
H	9.83505748	-3.42033368	5.68006205
C	4.53304140	-3.01822077	1.87250494
H	3.35188675	-1.27840337	1.71586090
C	5.97717295	-4.94680130	2.21174322
H	8.84018689	-2.64307725	3.36174161
C	8.21984568	-4.62357890	2.99757172
H	3.76417791	-3.65898440	1.49406801
C	7.17683935	-5.46612183	2.57727572
H	5.18888964	-5.59525211	1.89079283
H	9.16115688	-5.04465519	3.28294085
H	7.32945379	-6.52467979	2.54499629
P	5.52491837	1.60673818	1.63997482
N	5.50752626	3.23310912	2.28538225
O	0.77988403	3.66284025	3.40816851
O	3.77276434	3.59913422	4.40076305
N	2.81244369	2.54449837	2.03761734
H	8.85556596	2.79176716	2.38779524
Bq	**3.49740551**	**5.50863214**	**2.08902325**

- ***Specific Statistical Methods***

Regression analyses were conducted with respect to active site-based buried volume (ref. *dataset_active-site.csv*) for 55 literature-reported reaction datasets. 54 of these datasets are compiled in our previous report[1]**.** The additional dataset corresponds to R-**II**[2].

Parameters for regression models were fitted using either NumPy's *polyfit* function (*linear* – degrees = 1, *quadratic* – degrees = 2) or SciPy's *curve_fit* function with the respective initial guesses (**Supplementary Table 1**). Models with the largest R^2 were used in further analyses.

Regression	***Equation***	***Constraints***	***Initial Guess***
Linear	$y = ax + b$	*n/a*	*n/a*
Quadratic	$y = ax^2 + bx + c$	*n/a*	*n/a*
Logarithmic	$y = A\log(x) + c$	n/*a*	$A = \frac{\max(e.e.) - \min(e.e.)}{\log\left(\frac{\max(x)}{\min(x)}\right)}$ $c = \min(e.e.)$
Exponential growth	$y = Ae^{k(x-x_0)}$	$A \geq 0$ $k \geq 0$	$A = \max(e.e.) - \min(e.e.)$ $k = \frac{1}{\sigma_x}$ $c = \min(e.e.)$
Exponential decay	$y = Ae^{-k(x-x_0)}$	$A \geq 0$ $k \geq 0$	$A = \max(e.e.) - \min(e.e.)$ $k = \frac{1}{\mathrm{std}(x)}$ $c = \min(e.e.)$
Gaussian	$y = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$A \geq 0$	$A = \max(e.e.)$ $\mu = \mathrm{median}(x)$ $\sigma = \mathrm{std}(x)$
Lorentzian	$y = \frac{A}{1 + \frac{(x-x_0)^2}{\gamma^2}}$	$A \geq 0$	$A = \max(e.e.)$ $x_0 = \mathrm{median}(x)$ $\gamma = \mathrm{std}(x)$
Piecewise linear	$y = \begin{cases} a_1(x-x_0) + b, & x \leq x_0 \\ a_2(x-x_0) + b, & x > x_0 \end{cases}$	*n/a*	$b = \max(e.e.)$ $x_0 = x\ at\ \max(e.e.)$

Supplementary Table VI-19 | Parameters, constraints, and initial guesses for regression. 'y' refers to *enantiomeric excess* in kcal/mol (*e.e.*). 'x' refers to the corresponding active site buried volume.

VI.35 – Detailed Statistical Analyses

- ***Distribution of the Most Stereoselective BAOCs***

In the main text, **Fig. 2B** describes an analysis juxtaposing the distribution of the most stereoselective BAOCs (n = 34) against all BAOCs (n = 174) tested in the 50+ collated literature datasets (e.g. R-**II** from **Supplementary Table 2**). We define the distribution of the BAOCs by their corresponding Boltzmann averaged active site-based buried volumes locally (3 Å) and distally (7 Å) to represent possible effects and interactions. Then, Kolmogorov-Smirnov test (KS test, with *SciPy's* ks_2samp built-in function) was deployed to detect if the two distributions are comparable or distinct. In each case, the test did not detect significant differences (at 95% confidence) between the distributions of the *most* stereoselective BAOCs relative to any BAOCs tested (3 Å – KS statistic = 0.20, $p = 0.19$; 7 Å – KS statistic = 0.23, $p = 0.08$).

- ***Regression Model Selection and Performance***

As described in *General Statistical Methods*, 55 datasets compiled from the literature were analyzed with an array of linear and nonlinear models. 54 of these datasets are reported in our earlier report (Smith, A. et al. *J. Am. Chem. Soc.* **2026.**[1]). The additional dataset corresponds to R-**II** from **Fig. 2H** in the main text and is derived from Wu, S. et al. *Nature* **2025.**[2] and tabulated below for reference (**Supplementary Table 2**).

LABEL	CATALYST	EE	DDG
(R)-C1	CPAH_5	-27	0.26739
(R)-C2	CPAH_27	-24	0.236398
(R)-C3	CPAH_46	-12	0.116455
(R)-C4	CPAH_97	-3	0.028982
(R)-C5	CPAH_934	-7	0.067715
(R)-C6	SPAH_5	23	0.226176
(R)-C7	SPAH_27	23	0.226176
(R)-C8	SPAH_46	25	0.246673
(R)-C9	SPAH_29	28	0.277838
(R)-C10	SPAH_97	63	0.716046
(R)-C11	SPAH_7001	75	0.939661
(R)-C12	SPAH_934	78	1.009599

Supplementary Table VI-20 | R-**II** dataset from Wu, S. et al. *Nature* **2025.**[2] (sheet name is 'FFFF'). 'LABEL' refers to the name of the catalyst in the reported dataset. 'CATALYST' refers to the *internal* label for the catalyst present in *dataset_active-site.csv*. 'EE' refers to the reported enantioselectivity in % *e.e.*, while 'DDG' is the same but in kcal/mol.

- ***Performance of Regression Models***

Based on the principle of *complementarity* as it relates to confinement, catalysts that are either *more* or *less* confined to the perfect *complement* should furnish less selective solutions. This implies that the relationship between selectivity and a domain reflecting complementarity is nonlinear and non-monotonic. However, it is not guaranteed for a given dataset that chemical space was sufficiently explored to reveal such behavior. As a result, simple linear gradients can and do approximate the relationship between selectivity and confinement. To address whether regression models that can accommodate and detect such complementarity are indeed advantageous, we fit and recorded the *best*-performing regression models for each of the datasets considered. The *best*-performing regression model is defined with largest R^2 (across descriptors). The aggregate performance is recorded below for each of candidate (**Supplementary Table 3**).

Regression	***Equation***	***Linear?***	***Monotonic?***	***avg. R^2***	***avg. MAE (kcal/mol)***
Linear	$y = ax + b$	*Y*	*Y*	0.57	0.210
Quadratic	$y = ax^2 + bx + c$	***N***	***N***	**0.75**	**0.147**
Logarithmic	$y = A\log(x) + c$	***N***	*Y*	0.57	0.208
Exponential growth	$y = Ae^{k(x-x_0)}$	***N***	*Y*	0.51	0.224
Exponential decay	$y = Ae^{-k(x-x_0)}$	***N***	*Y*	0.51	0.224
Gaussian	$y = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$	***N***	***N***	**0.76**	**0.140**
Lorentzian	$y = \frac{A}{1 + \frac{(x-x_0)^2}{\gamma^2}}$	***N***	***N***	**0.75**	**0.148**
Piecewise linear	$y = \begin{cases} a_1(x-x_0)+b, & x \le x_0 \\ a_2(x-x_0)+b, & x > x_0 \end{cases}$	***Y***	***N***	**0.70**	**0.179**

Supplementary Table VI-21 | Average performance of linear and nonlinear, monotonic and non-monotonic regression models across the 55 datasets considered herein (baseline predictors, avg. MAE with *mean from dataset* = 0.373 kcal/mol, with *median from dataset* = 0.353 kcal/mol).

- ***Criteria for Selection of Gaussian Regression***

By inspection, it is clear nonlinear and non-monotonic regressions are superior relative to other models (see **Supplementary Table 3** where quadratic, Gaussian, Lorentzian, and piecewise linear regression each furnish avg. $R^2 > 0.70$). Between these three models, we deploy chemical intuition and statistical inquiry to identify whether one is superior. By inspection, quadratic regression and piecewise linear regression (Sabatier-type volcano plots) are recognized as *not* chemically meaningful. Once the domain exceeds the bounds of the training set, the model diverges to $-\infty$, which chemically refers to perfect enantioselectivity in the opposite sense.

On the other hand, Guassian and Lorentzian models furnish analogous non-monotonic fits but display tails that approach zero outside the domain, a chemically meaningful phenomenon indicating a loss of stereoinduction in these domains. Qualitatively, Lorentzian curves dampen slower than Gaussian and can accommodate a larger baseline for stereoinduction relative to Gaussian (accordingly, Lorentzian – avg. $R^2 = 0.75$, Gaussian – avg. $R^2 = 0.72$). This marginal improvement in performance is persistent in many datasets, but only a handful are very meaningful (defined by AIC weights, Burnham, K. Anderson, D. Model Selection and Multimodel Inference. *Springer* **2004.**[3]). The *likelihood of a model (L) given the dataset* is defined as the "evidence" for a model to be the *best* model within the set of considered models. According to Burnham, et al., this is proportional to the formula below:

$$L \propto e^{-\Delta_i/2}, \text{ where } \Delta_i = \mathrm{AICc}_i - \mathrm{AICc}_{min} \text{ and } \mathrm{AICc}_i = 2\ln(\mathrm{MSE}_i) + k + \frac{2k(k+1)}{n_i - k - 1}$$

such that AIC_{min} is the minimum AIC value across the set of models considered, MSE_i and n_i correspond to the mean squared error and dataset size for the model, while k represents the number of fitted parameters (note, $k = 3$ for both Lorentzian and Gaussian regression models).

Likelihood (L) can be normalized and converted to an interpretable *probability of a model given the dataset* (w_i) that the model is the *best* model within the set of considered models.

$$w_i = \frac{e^{-\Delta_i/2}}{\sum_1^n e^{-\Delta n/2}} \text{ for all } n \text{ models in the set of models considered.}$$

We employ the cutoff of *"10 times more probable to be the best model"* as sufficient evidence of improved performance. As cautioned by Burnham, et al. these metrics are not to be confused as *p*-values and cannot be converted to some metric for hypothesis-based statistical analysis. From this analysis, we can detect only 4 datasets are advantageously represented by Lorentzian fit, while 6 datasets are advantageously represented by Gaussian fit (see **Supplementary Table 4** for tabulated Gaussian, 'G' and Lorentzian, 'L' AICc values, weights across all 55 datasets).

RXN	**AICc-G.**	**wt.-G**	**AICc-L**	**wt.-L**
AA	-95.939	0.7404	-93.844	0.2596
B	-42.339	0.4064	-43.097	0.5936
BB	-134.83	0.4777	-135.009	0.5223
BBB	-6.509	0.6457	-5.309	0.3543
BBBB	-8.727	0.6843	-7.18	0.3157
C	-26.147	0.6247	-27.167	0.3753
CC	-17.13	0.6195	-18.108	0.3805
CCC	-23.176	0.5609	-23.665	0.4391
CCCC	**-36.562**	**0.0175**	**-44.634**	**0.9825**

D	-27.811	0.7753	-25.335	0.2247	*JJJ*	-17.555	0.8778	-13.614	0.1222
DD	-7.576	0.557	-7.118	0.443	*K*	-31.67	0.1217	-35.628	0.8783
DDD	-15.538	0.6562	-16.831	0.3438	*LLL*	-22.83	0.3062	-24.456	0.6938
DDDD	-19.124	0.6254	-18.099	0.3746	***MMM***	**-11.029**	**0.957**	**-4.863**	**0.043**
EE	**-42.403**	**0.9466**	**-36.651**	**0.0534**	*NNN*	-7.118	0.4962	-7.088	0.5038
EEE	-50.538	0.2124	-53.16	0.7876	***O***	**-23.389**	**0.9342**	**-18.085**	**0.0658**
EEEE	-8.196	0.723	-10.119	0.277	*OOO*	-22.014	0.8649	-18.309	0.1351
FF	-72.388	0.4037	-73.166	0.5963	*Q*	-24.587	0.6003	-25.4	0.3997
FFF	-10.116	0.7463	-12.272	0.2537	*QQQ*	-16.587	0.6918	-14.966	0.3082
FFFF	-37.679	0.7641	-35.327	0.2359	***S***	**-40.372**	**0.9792**	**-32.653**	**0.0208**
G	**-24.42**	**0.0024**	**-36.466**	**0.9976**	*SSS*	-94.141	0.6397	-95.29	0.3603
GG	-4.173	0.6443	-2.984	0.3557	*T*	-5.521	0.8965	-1.194	0.1035
GGG	-38.24	0.1477	-34.736	0.8523	*UUU*	-12.334	0.5141	-12.447	0.4859
HH	**-30.481**	**0.9642**	**-23.897**	**0.0358**	*V*	2.525	0.26	0.433	0.74
HHH	-26.117	0.5294	-26.352	0.4706	*VVV*	-56.905	0.8864	-52.79	0.1136
I	-103.343	0.6012	-102.522	0.3988	*WWW*	-43.927	0.2098	-46.579	0.7902
II	-24.067	0.5795	-23.425	0.4205	*YYY*	-38.879	0.5616	-39.374	0.4384
III	**-12.905**	**0.0066**	**-22.954**	**0.9934**	*ZZ*	-44.927	0.6287	-45.978	0.3713
J	-15.415	0.515	-15.535	0.485	*ZZ-TEST*	-29.995	0.7107	-28.193	0.2893
JJ	-36.3	0.4929	-36.243	0.5071	***ZZZ***	**-27.14**	**0.9989**	**-13.508**	**0.0011**

Supplementary Table VI-22 | Tabulated corrected AIC values and weights for best Gaussian ('G') and Lorentzian ('L') regression models trained on active site-based buried volume across the reported datasets. Each dataset ('RXN') is defined by its '*sheet name*' in *dataset_rxn.xlsx*[1]

Accordingly, we select Gaussian regression as the model for our analyses, owing to its superior regression performance and longstanding familiarity with chemists.

- ***Evidence against Null Considerations for Modeling Approach***

Although Gaussian regression models with active site-based buried volume perform exceptionally well across the 55 datasets considered herein, it can be argued its advantageous performance is a consequence of (i) its dimensionality, or (ii) its non-monotonic nature. We provide evidence supporting this *specific* utility of buried volume with Gaussian regression.

First, we can consider the counterargument – *Gaussian regression only performs well because of its additional dimensionality*. Above, we describe in detail the performance of alternative regression models for the datasets considered herein. In summary, only nonlinear models with non-monotonic behavior capture *complementarity* and thus we observe advantageous numerical fits, regardless of the model. Critically, other nonlinear models like *exponential growth* and *decay* also possess more parameters for fitting (n = 3 like Gaussian) but even underperform simpler

alternatives like 2-parameter *linear* regression (linear – avg. $R^2 = 0.57$, exponential growth – avg. $R^2 = 0.51$, exponential decay – avg. $R^2 = 0.51$).

Second, we consider the counterargument – *any Gaussian regression, regardless of the feature, will furnish meaningful fits because of its nonlinear, non-monotonic reflection of complementarity*. In our earlier report,[1] we investigated in detail the unique utility of active site-based buried volume to reflect confinement of BAOCs as it relates to stereoselectivity in a range of reactivity. We highlight here that other familiar, competing representations of confinement do *not* capture complementarity as meaningfully as active site-based buried volume, even when leveraging Gaussian regression.

Additional representations of confinement at the BAOC active site include sterimol-based features, which collectively describe geometric attributes of the active site (ex. sterimol_L 'length', sterimol_BMin 'width', and sterimol_BMax 'height'). Using the same numerical fitting approach described earlier, we quickly recognize these descriptors are not adequate reflections of stereoselectivity with Gaussian regression (Gaussian; sterimol_L – avg. $R^2 = 0.46$, sterimol_BMin – avg. $R^2 = 0.48$, sterimol_BMax – avg. $R^2 = 0.45$).

Alternatively, we can consider leveraging buried volume features from different positions along the BAOC scaffold. Previously, we juxtaposed the utility of active site-based buried volume against substituent-based buried volume. Under the same numerical fitting approach, we see substituent-based buried volume does display improved utility (relative to previously reported linear – avg. $R^2 = 0.56$), but still greatly underperforms relative to active site-based (Gaussian with substituent-based buried volume – avg. $R^2 = 0.64$).

Besides confinement, acidity of BAOCs is an essential property with known relevance to reactivity and selectivity. We represent this property using *calculated* pK_a as benchmarked in our earlier report (ref. Smith, et al. *J. Am. Chem. Soc.* **2026.**). But, Gaussian regression models trained only on calculated $p\mathrm{K_a}$ of the BAOCs furnish poor fits (avg. $R^2 = 0.34$, avg. MAE = 0.279 kcal/mol). Instead, we wondered if bivariate Gaussian regression models constructed with buried volume *and* calculated $p\mathrm{K_a}$ would generate relationships substantially more useful than buried volume alone. Here, we trained a 5-parameter Gaussian regression model (see equation below) and invoked AIC weights to decide if it furnishes an advantageous fit.

$$y = Ae^{-\frac{(x_1-\mu_1)^2}{2{\sigma_1}^2}} e^{-\frac{(x_2-\mu_2)^2}{2{\sigma_2}^2}}$$

We observed the bivariate Gaussian regression model trained on buried volume *and* calculated $p\mathrm{K_a}$ is not especially advantageous in any dataset. Instead, the univariate Gaussian regression model trained *only* on buried volume ('*VBur'*) is substantially advantageous relative to both (i) bivariate Gaussian regression ('*bivar'*) and (ii) univariate Gaussian regression with calculated $p\mathrm{K_a}$ ('*pKa*') in 34 datasets (see below).

RXN	AIC-VBur.	wt. VBur	AIC-pKa	wt. pKa	AIC-bivar	wt. bivar
AA	-95.939	0.249	-70.744	0	-98.149	0.751
B	-42.339	0.98	-13.503	0	-34.595	0.02
BB	-134.83	0.721	-130.879	0.1	-132.04	0.179
BBB	-6.509	0.972	0.582	0.028	48.234	0
BBBB	-8.727	0.997	3.186	0.003	37.059	0
C	-26.147	0.812	-23.158	0.182	-16.26	0.006
CC	-17.13	0.999	-1.997	0.001	35.181	0
CCC	-23.176	0.968	-16.31	0.031	-8.178	0.001
CCCC	-36.562	1	-15.861	0	-13.335	0
D	-27.811	0.999	2.973	0	-12.949	0.001
DD	-7.576	0.983	0.554	0.017	42.263	0
DDD	-15.538	0.954	-9.455	0.046	32.862	0
DDDD	-19.124	0.945	-13.417	0.054	-4.123	0.001
EE	-42.403	0.969	-31.393	0.004	-35.251	0.027
EEE	-50.538	0.856	-29.251	0	-46.975	0.144
EEEE	-8.196	0.677	-6.714	0.323	45.138	0
FF	-72.388	0.522	-70.504	0.203	-71.108	0.275
FFF	-10.116	0.989	-1.11	0.011	37.504	0
FFFF	-37.679	0.993	-24.84	0.002	-27.278	0.005
G	-24.42	1	5.527	0	30.677	0
GG	-4.173	0.999	10.547	0.001	51.826	0
GGG	-38.24	0.999	-24.086	0.001	-14.106	0
HH	-30.481	0.999	-10.834	0	-15.479	0.001
HHH	-26.117	1	10.026	0	-8.53	0
I	-103.343	0.933	-94.665	0.012	-97.694	0.055
II	-24.067	0.996	-13.051	0.004	-2.11	0
III	-12.905	0.996	-1.742	0.004	25.135	0
J	-15.415	0.372	-16.466	0.628	6.776	0
JJ	-36.3	0.736	-28.024	0.012	-34.158	0.252
JJJ	-17.555	0.999	-3.284	0.001	38.434	0
K	-31.67	1	-5.863	0	-4.831	0
LLL	-22.83	0.997	-11.001	0.003	28.477	0
MMM	-11.029	0.999	4.117	0.001	41.427	0
NNN	-7.118	0.948	-1.281	0.051	7.828	0.001
O	-23.389	1	8.515	0	22.609	0
OOO	-22.014	1	-4.246	0	33.323	0
Q	-24.587	0.999	-1.716	0	-10.122	0.001
QQQ	-16.587	0.835	-13.348	0.165	0.595	0
S	-40.372	0.983	-16.843	0	-32.234	0.017
SSS	-94.141	0.94	-85.198	0.011	-88.232	0.049
T	-5.521	0.633	-4.429	0.367	46.584	0
UUU	-12.334	0.957	-5.877	0.038	-1.895	0.005

V	2.525	0.838	5.805	0.162	50.411	0
VVV	-56.905	0.541	-54.09	0.132	-55.9	0.327
WWW	-43.927	0.241	-46.213	0.755	-36.013	0.005
YYY	-38.879	0.982	-23.265	0	-30.875	0.018
ZZ	-44.927	0.975	-34.551	0.005	-37.085	0.019
ZZ-TEST	-29.995	0.886	-17.45	0.002	-25.868	0.113
ZZZ	-27.14	0.995	5.14	0	-16.513	0.005

Supplementary Table VI-23 | Tabulated corrected AIC values and weights for Gaussian models trained on active site-based buried volume *only* ('Vbur'), calculated pK_a *only* ('pka'), or both features ('bivar'). The datasets are marked akin to **Supplementary Table 4.**

- ***Numerical Fits for each Literature-derived Dataset with Gaussian Regression***

We report each Gaussian fit for literature-derived datasets into a table containing (1) the selected buried volume representation, (2) full dataset fit (R^2, MAE) and (3) fitted parameters with computed uncertainty (A, μ, σ with ±1 SE, approximated by the fitting algorithm). To summarize, the best-performing Gaussian fits with buried volume representations furnish an avg. R^2 = 0.76, avg. MAE = 0.140 kcal/mol.

sheet	descriptor	R^2	MAE	A	$\pm SE_A$	μ	$\pm SE_\mu$	σ	$\pm SE_\sigma$
AA	bur_vol_70A 1 O MIN	0.58	0.124	0.706	0.096	21.038	0.395	2.297	0.413
AAAA	bur_vol_40A 2 H MAX	0.86	0.078	1.562	0.198	54.142	2.246	13.807	3.062
B	bur_vol_20A 2 H AV_XTB	0.84	0.163	1.426	0.24	52.537	1.171	1.968	0.669
BB	bur_vol_35A 2 H MAX	0.17	0.089	0.251	0.053	27.115	14.029	15.558	11.456
BBB	bur_vol_20A 1 O C_0	0.75	0.192	1.386	0.306	70.417	1.307	1.465	0.812
BBBB	bur_vol_70A 2 H MAX	0.91	0.147	1.942	0.235	41.121	0.564	4.152	0.538
C	bur_vol_20A 2 H AV_XTB	0.56	0.212	1.367	0.329	50.278	0.181	0.802	0.123
CC	bur_vol_25A 1 O MAX	0.88	0.095	1.119	0.156	56.945	0.212	2.058	0.302
CCC	bur_vol_60A 1 O MAX	0.85	0.161	1.422	0.151	36.817	0.623	3.159	0.553
CCCC	bur_vol_70A 2 H C_0	0.95	0.039	0.733	0.062	35.031	0.385	3.53	0.359
D	bur_vol_20A 1 O AV_XTB	0.97	0.089	2.698	0.188	68.109	0.068	0.919	0.041
DD	bur_vol_30A 1 O MIN	0.82	0.185	2.22	0.251	49.302	0.164	1.587	0.207
DDD	bur_vol_20A 2 H MIN	0.93	0.109	1.513	0.396	48.598	0.01	0.049	0.012
DDDD	bur_vol_20A 2 H MIN	0.51	0.162	1.209	0.225	48.111	0.727	1.137	0.634
E	bur_vol_70A 1 O MAX	0.33	0.409	1.66	0.638	33.849	2.693	7.612	3.505
EE	bur_vol_30A 2 H MAX	0.95	0.112	13.627	9.156	76.532	0.284	3.951	0.552
EEE	bur_vol_60A 1 O MAX	0.89	0.088	1.309	0.205	32.49	0.264	1.12	0.143
EEEE	bur_vol_20A 2 H C_0	0.61	0.192	1.282	0.268	49.432	0.087	0.509	0.104
FF	bur_vol_20A 1 O MIN	0.14	0.194	0.623	0.179	70.991	5.664	4.601	4.654
FFF	bur_vol_20A 2 H C_0	0.74	0.147	1.356	0.338	49.548	0.1	0.46	0.088
FFFF	bur_vol_25A 1 O AV_XTB	0.8	0.113	1.15	0.168	50.567	0.206	1.752	0.201

G	bur_vol_50A 1 O C_0	0.99	0.045	4.004	0.52	53.575	0.312	4.763	0.195
GG	bur_vol_40A 1 O MAX	0.91	0.199	2.892	0.273	54.301	1.676	5.322	1.254
GGG	bur_vol_30A 2 H MAX	0.95	0.035	0.91	0.028	64.586	1.421	14.519	1.851
H	bur_vol_20A 2 H AV_XTB	0.95	0.187	20.221	39.86	50.676	0.192	0.752	0.315
HH	bur_vol_25A 2 H AV_XTB	0.95	0.104	1.993	0.173	39.221	0.096	1.009	0.065
HHH	bur_vol_50A 1 O MIN	0.98	0.105	2.629	0.125	34.096	0.036	0.673	0.054
I	bur_vol_20A 2 H C_0	0.6	0.099	0.58	0.107	50.123	0.51	0.846	0.305
II	bur_vol_25A 2 H C_0	0.81	0.092	0.772	0.072	40.348	0.722	2.05	0.663
III	bur_vol_20A 2 H MAX	0.92	0.109	1.871	0.314	62.287	1.425	6.984	1.175
J	bur_vol_20A 2 H MIN	0.79	0.143	1.222	0.158	48.007	0.446	0.53	0.356
JJ	bur_vol_25A 1 O MIN	0.61	0.091	0.649	0.121	55.973	0.603	2.963	0.565
JJJ	bur_vol_70A 1 O MAX	0.98	0.079	2.16	0.095	39.318	0.559	5.593	0.522
K	bur_vol_40A 2 H C_0	0.96	0.051	4.75	2.029	50.265	1.101	7.519	0.627
LLL	bur_vol_70A 2 H MIN	0.93	0.061	1.244	0.122	25.791	0.192	2.99	0.284
MMM	bur_vol_25A 1 O AV_XTB	0.91	0.132	4.245	1.097	55.509	0.064	1.119	0.123
NNN	bur_vol_70A 2 H C_0	0.2	0.359	2.138	0.312	29.889	2.566	7.129	4.003
O	bur_vol_25A 1 O C_0	0.98	0.076	6030.774	5072.488	69.942	0.083	3.281	0.164
OOO	bur_vol_25A 2 H MIN	0.94	0.056	2.889	0.531	47.505	0.429	5.758	0.347
PPP	bur_vol_20A 2 H AV_XTB	0.87	0.08	0.846	0.332	48.961	0.111	0.157	0.066
Q	bur_vol_20A 2 H C_0	0.9	0.136	1.652	0.157	84.8	7.155	20.848	4.28
QQQ	bur_vol_20A 2 H C_0	0.68	0.247	1.432	0.315	48.744	0.566	0.625	0.336
RRR	bur_vol_20A 2 H MIN	0.94	0.08	2.016	0.274	50.005	0.574	0.959	0.469
S	bur_vol_20A 1 O AV_XTB	0.9	0.141	2.322	0.19	70.539	0.461	1.871	0.323
SSS	bur_vol_70A 2 H MIN	0.29	0.112	0.383	0.079	18.342	3.713	5.436	2.923
T	bur_vol_20A 2 H MIN	0.84	0.187	2.15	0.328	48.094	0.108	0.374	0.107
TTT	bur_vol_25A 1 O MAX	1	0.018	6.44	0.639	67.453	0.075	4.593	0.16
UUU	bur_vol_60A 1 O MAX	0.48	0.293	1.433	0.372	57.248	2.598	10.118	2.406
V	bur_vol_25A 2 H MIN	0.48	0.383	1.601	0.442	38.007	0.191	0.498	0.175
VVV	bur_vol_20A 1 O MAX	0.5	0.084	0.569	0.098	72.393	0.469	2.288	0.518
WWW	bur_vol_40A 1 O MIN	0.52	0.08	0.426	0.054	43.047	0.376	2.419	0.423
YYY	bur_vol_30A 2 H C_0	0.64	0.184	1.05	0.103	35.934	0.588	2.897	0.509
ZZ	bur_vol_25A 1 O C_0	0.52	0.145	1.364	0.105	58.925	1.258	5.962	1.462
ZZ-TEST	bur_vol_70A 1 O MIN	0.61	0.225	1.959	0.288	38.353	5.706	11.566	4.331
ZZZ	bur_vol_70A 1 O MAX	0.95	0.154	2.951	0.311	34.485	0.135	1.935	0.152

Supplementary Table VI-24 | Complete regression metrics, parameters, and uncertainty for all literature-derived datasets with Gaussian regression over active site-based buried volume.

We also report cross-validated performance of Gaussian fits for each dataset into a table containing (1) the selected buried volume representation, (2) full dataset fit (R^2, MAE) and (3) leave-one-out cross-validated dataset fit (LOO- R^2, LOO-MAE). To summarize, the best-performing cross-validated Gaussian fits furnish an avg. $R^2 = 0.67$, avg. MAE = 0.149 kcal/mol.

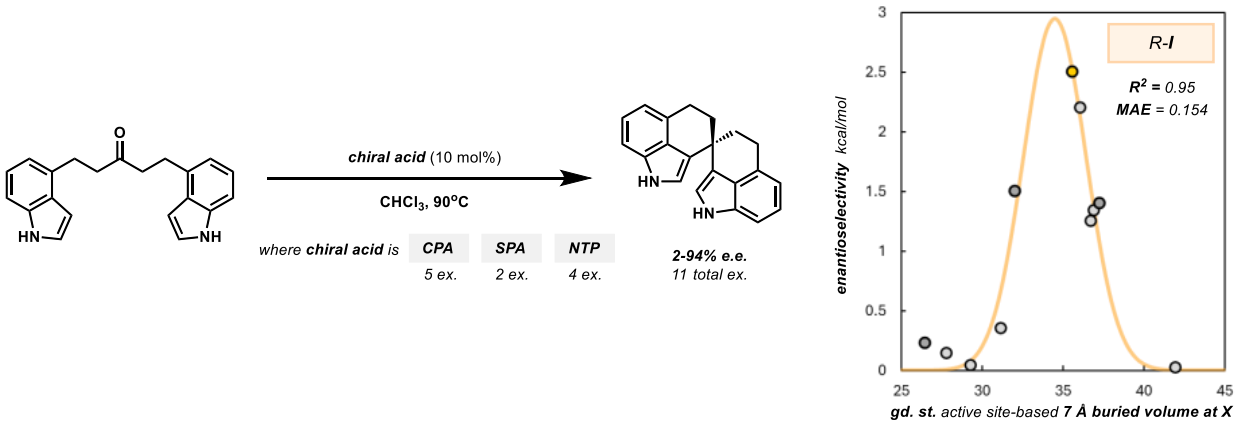
sheet	descriptor	R^2	MAE	LOO-R^2	LOO-MAE
AA	bur_vol_60A 1 O MAX	0.54	0.133	0.41	0.147
AAAA	bur_vol_30A 2 H MAX	0.84	0.091	0.25	0.221
B	bur_vol_20A 2 H AV_XTB	0.84	0.163	0.77	0.197
BB	bur_vol_40A 2 H MAX	0.16	0.088	0.04	0.094
BBB	bur_vol_20A 2 H C_0	0.17	0.274	-1.08	0.428
BBBB	bur_vol_70A 2 H AV_XTB	0.88	0.153	0.44	0.372
C	bur_vol_20A 2 H AV_XTB	0.56	0.212	-0.23	0.314
CC	bur_vol_30A 1 O MAX	0.76	0.104	0.48	0.166
CCC	bur_vol_70A 1 O C_0	0.81	0.168	0.57	0.244
CCCC	bur_vol_35A 1 O MAX	0.92	0.051	0.93	0.054
D	bur_vol_20A 1 O AV_XTB	0.97	0.089	0.7	0.268
DD	bur_vol_30A 1 O MIN	0.82	0.185	0.03	0.41
DDD	bur_vol_30A 2 H MIN	0.9	0.103	0.6	0.21
DDDD	bur_vol_20A 2 H MAX	0.05	0.19	-0.28	0.235
E	bur_vol_20A 2 H MAX	0.21	0.384	-0.64	0.577
EE	bur_vol_30A 2 H MAX	0.95	0.112	0.83	0.203
EEE	bur_vol_60A 1 O MAX	0.89	0.088	0.75	0.133
EEEE	bur_vol_60A 1 O C_0	0.15	0.21	-0.44	0.28
FF	bur_vol_70A 2 H MAX	0.12	0.148	-0.05	0.165
FFF	bur_vol_70A 2 H MAX	0	0.248	-0.83	0.347
FFFF	bur_vol_40A 1 O MIN	0.73	0.113	0.51	0.163
G	bur_vol_25A 1 O MAX	0.99	0.056	0.85	0.175
GG	bur_vol_20A 2 H C_0	0.87	0.259	0.62	0.464
GGG	bur_vol_30A 2 H MAX	0.95	0.035	0.9	0.052
H	bur_vol_20A 1 O MAX	0.9	0.259	0.07	0.84
HH	bur_vol_25A 2 H C_0	0.92	0.145	0.84	0.199
HHH	bur_vol_50A 1 O MIN	0.98	0.105	0.9	0.226
I	bur_vol_20A 2 H C_0	0.6	0.099	0.51	0.114
II	bur_vol_30A 2 H C_0	0.77	0.092	0.54	0.131
III	bur_vol_20A 2 H MAX	0.92	0.109	0.8	0.195
J	bur_vol_20A 2 H MIN	0.79	0.143	0.63	0.192
JJ	bur_vol_25A 1 O MIN	0.61	0.091	0.37	0.127
JJJ	bur_vol_70A 2 H MAX	0.98	0.074	0.95	0.116
K	bur_vol_40A 2 H C_0	0.96	0.051	0.91	0.087
LLL	bur_vol_70A 1 O MIN	0.92	0.06	0.57	0.143
MMM	bur_vol_20A 1 O AV_XTB	0.83	0.149	0.59	0.275
NNN	bur_vol_20A 2 H AV_XTB	0	0.374	-0.33	0.439
O	bur_vol_25A 1 O C_0	0.98	0.076	0.85	0.211
OOO	bur_vol_35A 2 H MIN	0.93	0.059	0.69	0.13
PPP	bur_vol_20A 2 H AV_XTB	0.87	0.08	0.55	0.137
Q	bur_vol_35A 2 H MIN	0.89	0.118	0.85	0.153

QQQ	bur_vol_25A 2 H MIN	0.55	0.172	-0.21	0.33
RRR	bur_vol_20A 2 H MAX	0.77	0.17	-0.44	0.429
S	bur_vol_25A 1 O AV_XTB	0.83	0.165	0.77	0.192
SSS	bur_vol_70A 2 H MIN	0.29	0.112	0.1	0.127
T	bur_vol_70A 1 O C_0	0.74	0.263	-0.04	0.521
TTT	bur_vol_70A 2 H MAX	0.98	0.064	-0.36	0.381
UUU	bur_vol_60A 1 O MAX	0.48	0.293	0.07	0.405
V	bur_vol_25A 2 H C_0	0.32	0.335	-0.44	0.523
VVV	bur_vol_25A 2 H MIN	-0.11	0.094	-0.29	0.112
WWW	bur_vol_35A 1 O AV_XTB	0.46	0.079	0.06	0.107
YYY	bur_vol_35A 2 H C_0	0.62	0.187	0.48	0.218
ZZ	bur_vol_25A 1 O C_0	0.52	0.145	0.24	0.18
ZZ-TEST	bur_vol_25A 1 O C_0	0.42	0.244	0.24	0.278

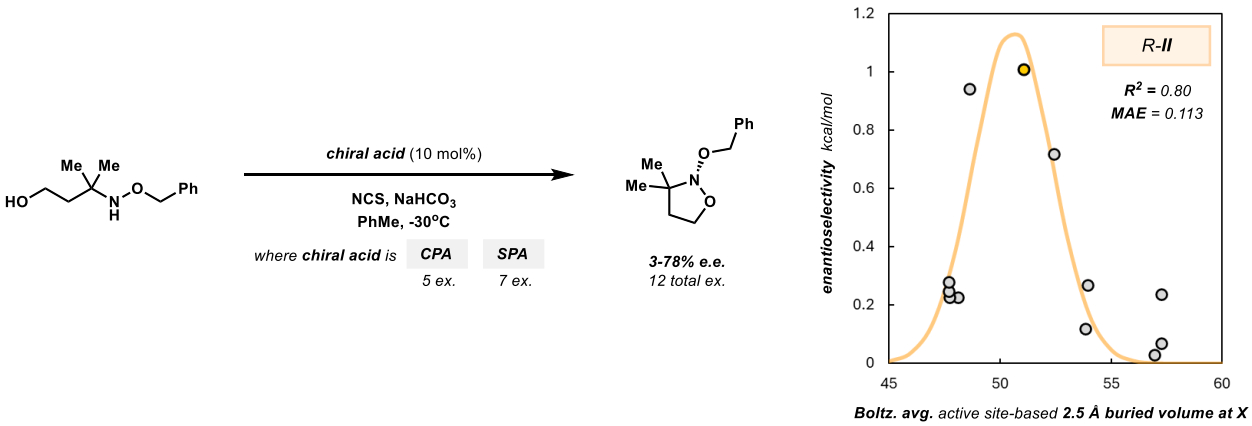
Supplementary Table VI-25 | Complete cross-validated regression metrics for all literature-derived datasets with Gaussian regression with active site-based buried volume.

- ***Visualizing each Gaussian Regression Model***

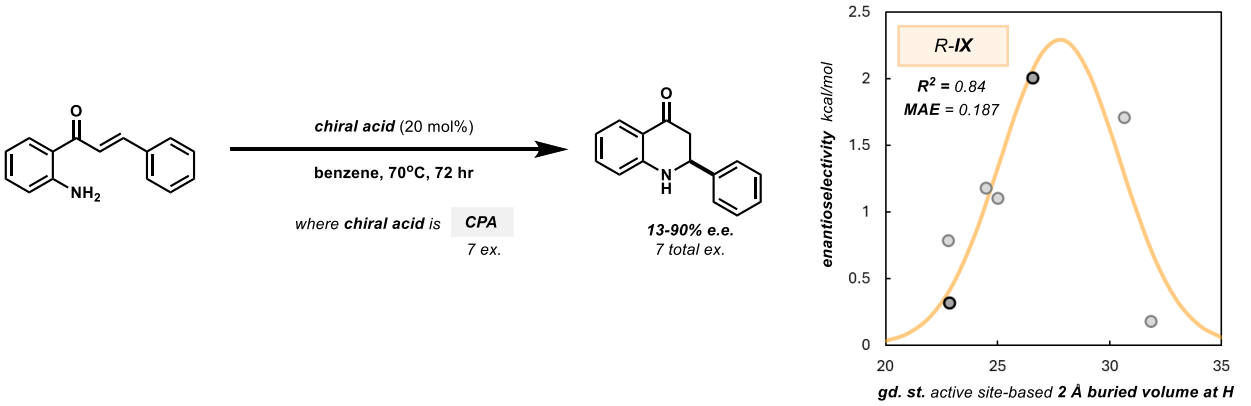
Below, we visualize the best Gaussian regression model with active site-based buried volume for each literature-derived dataset. The datasets are identified by their respective sheet name consistent with the **Supplementary Tables** and our previously compiled report.[1]



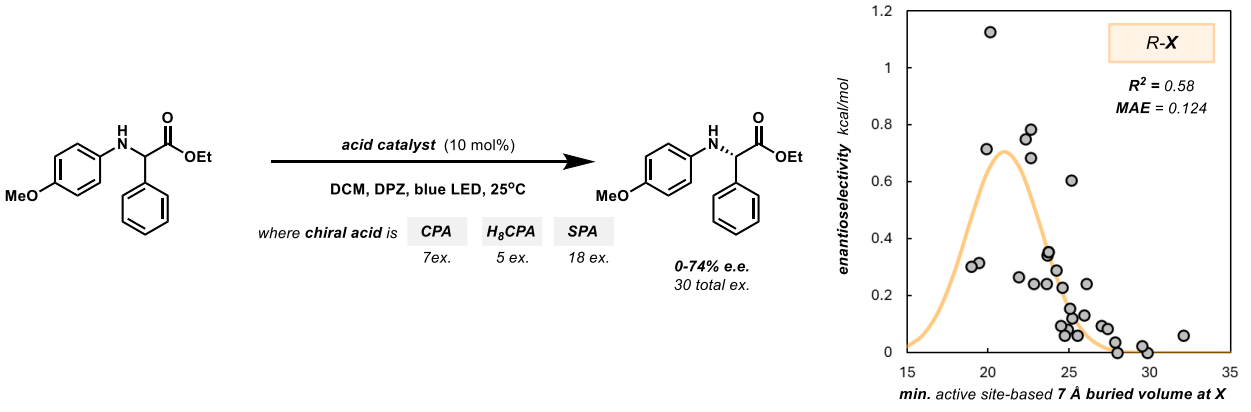
Supplementary Fig. VI-119 | R-**I** (sheet = *ZZZ*, A = 2.95, μ = 34.5, σ = 1.93).[4]



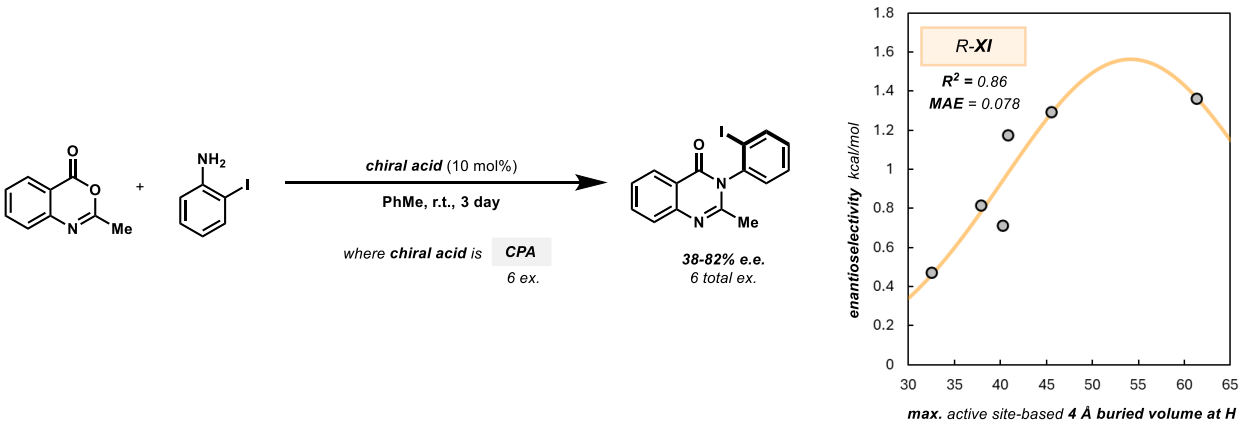
Supplementary Fig. VI-120 | R-**II** (sheet = *FFFF*, A = 1.15, μ = 50.6, σ = 1.75).[2]



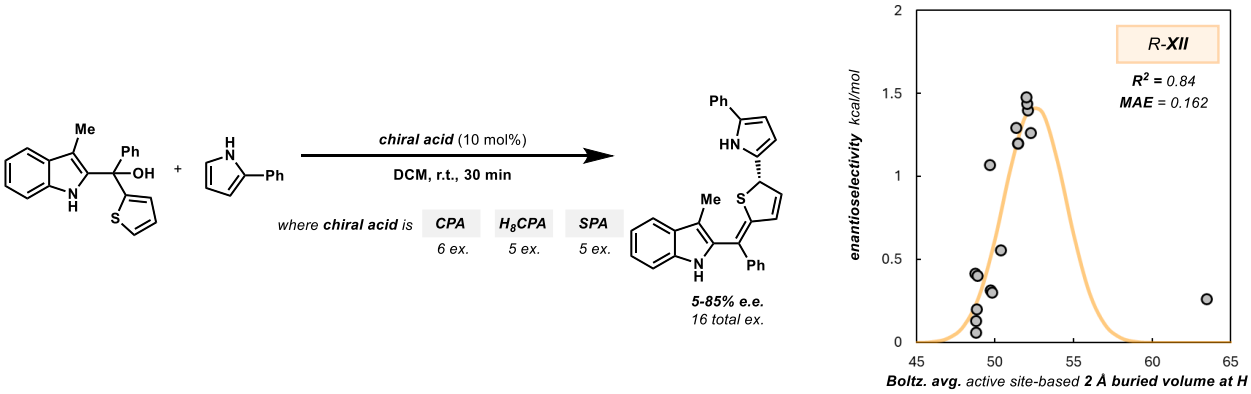
Supplementary Fig. VI-121 | R-**IX** (sheet name = *T*, A = 2.15, μ = 48.1, σ = 0.37).[5]



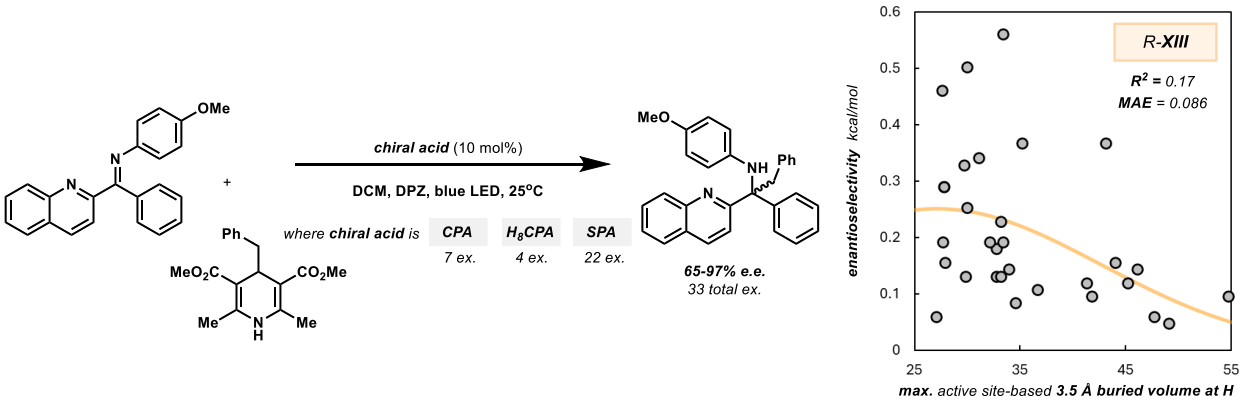
Supplementary Fig. VI-122 | R-**X** (sheet name = *AA*, A = 0.71, μ = 21.0, σ = 2.30).[6]



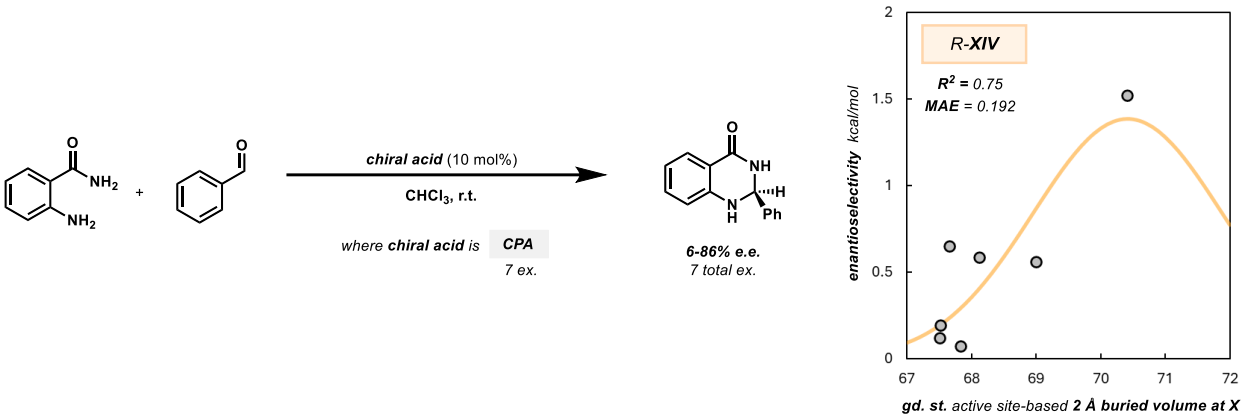
Supplementary Fig. VI-123 | R-**XI** (sheet name = *AAAA*, A = 1.56, μ = 54.1, σ = 13.8).[7]



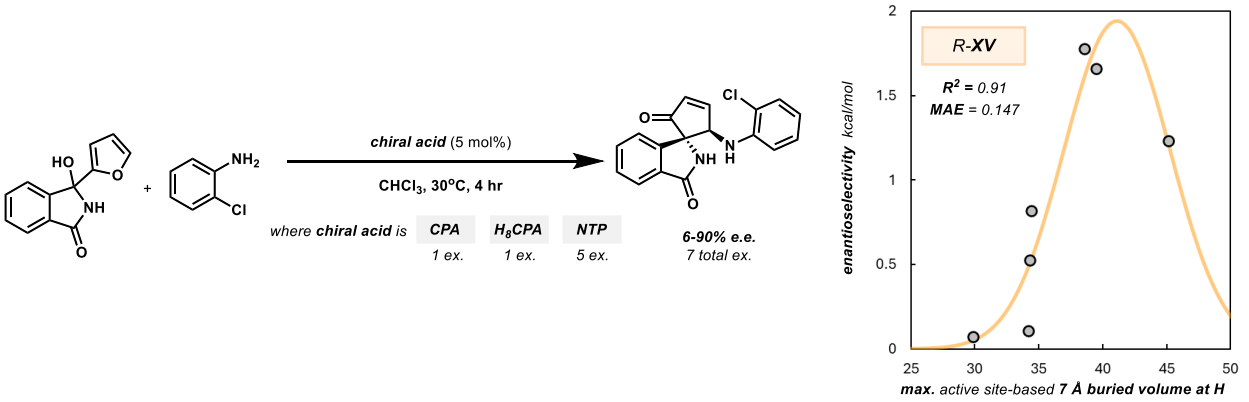
Supplementary Fig. VI-124 | R-**XII** (sheet name = *B*, A = 1.43, μ = 52.5, σ = 1.97).[8]



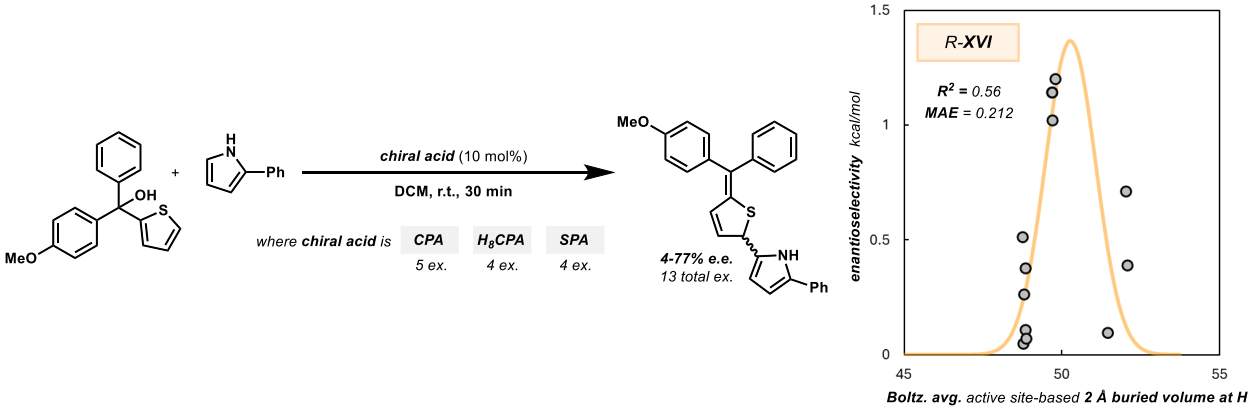
Supplementary Fig. VI-125 | R-**XIII** (sheet name = *BB*, A = 0.251, μ = 27.1, σ = 15.6).[9]



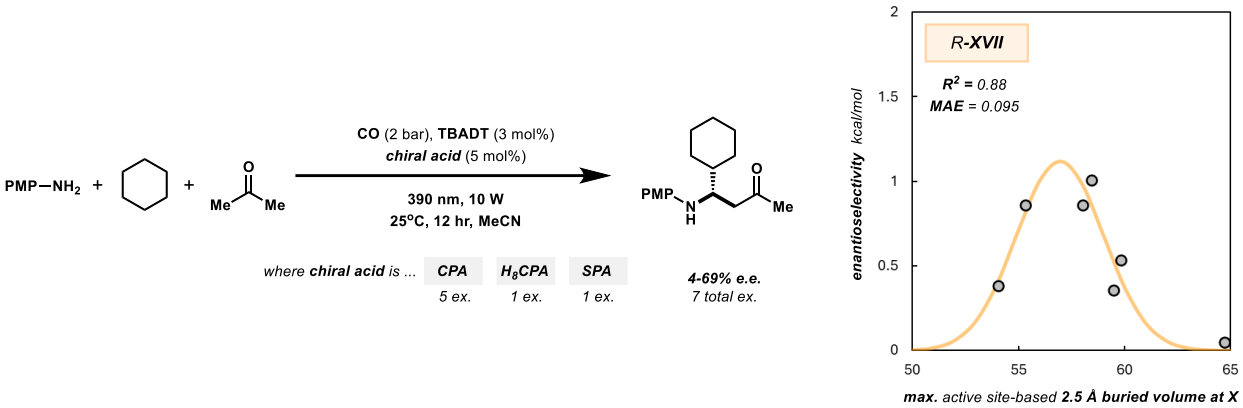
Supplementary Fig. VI-126 | R-**XIV** (sheet name = *BBB*, A = 1.39, μ = 70.4, σ = 1.46).[10]



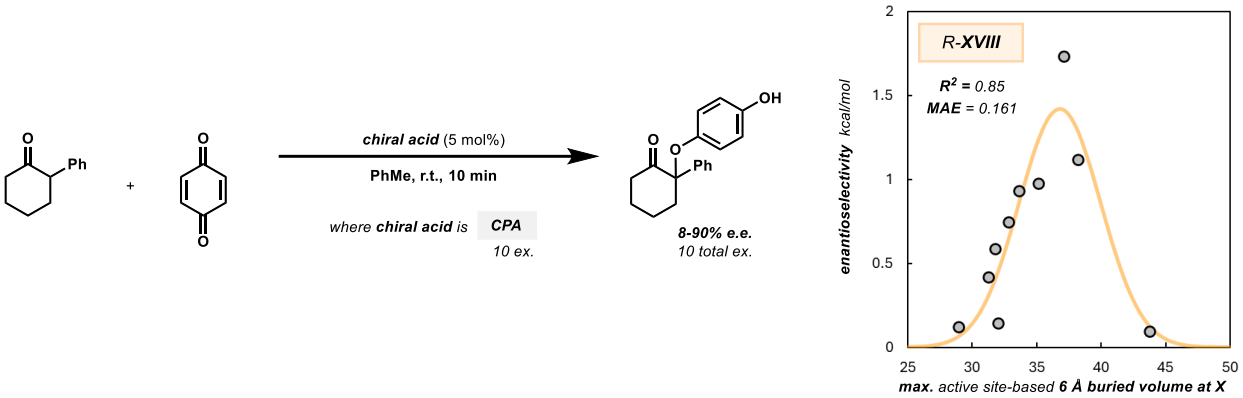
Supplementary Fig. VI-127 | R-**XV** (sheet name = *BBBB*, A = 1.94, μ = 41.1, σ = 4.15).[11]



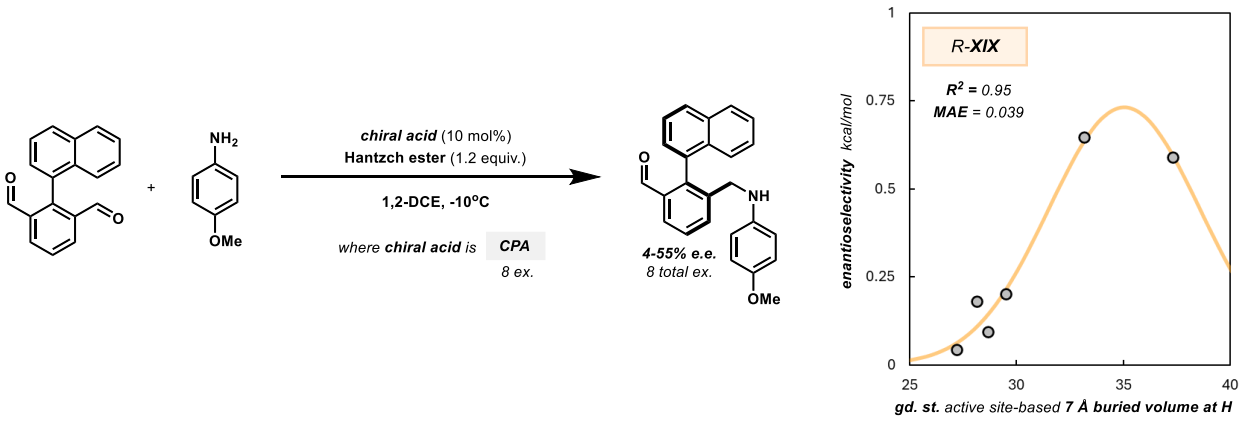
Supplementary Fig. VI-128 | R-**XVI** (sheet name = *C*, A = 1.37, μ = 50.3, σ = 0.80).[8]



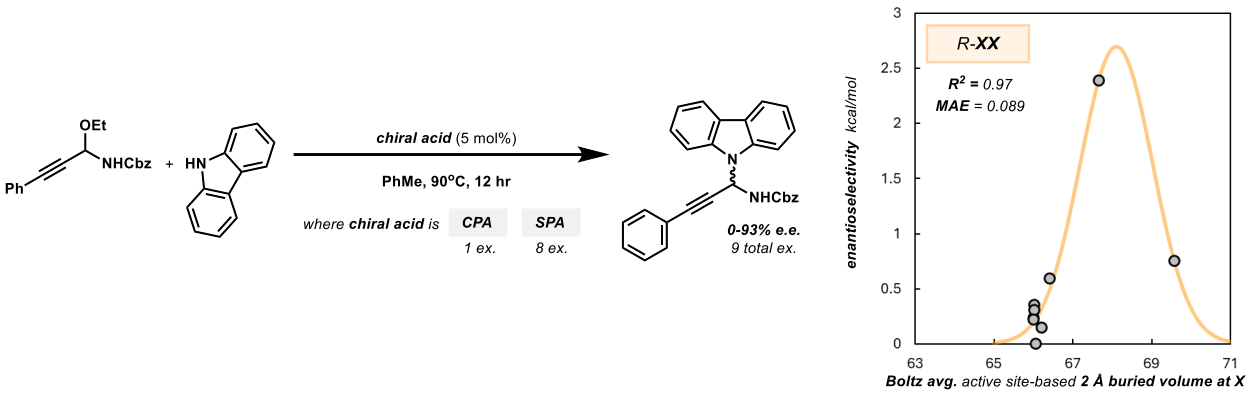
Supplementary Fig. VI-129 | R-**XVII** (sheet name = *CC*, A = 1.12, μ = 56.9, σ = 2.06).[12]



Supplementary Fig. VI-130 | R-**XVIII** (sheet name = *CCC*, A = 1.42, μ = 36.8, σ = 3.16).[13]

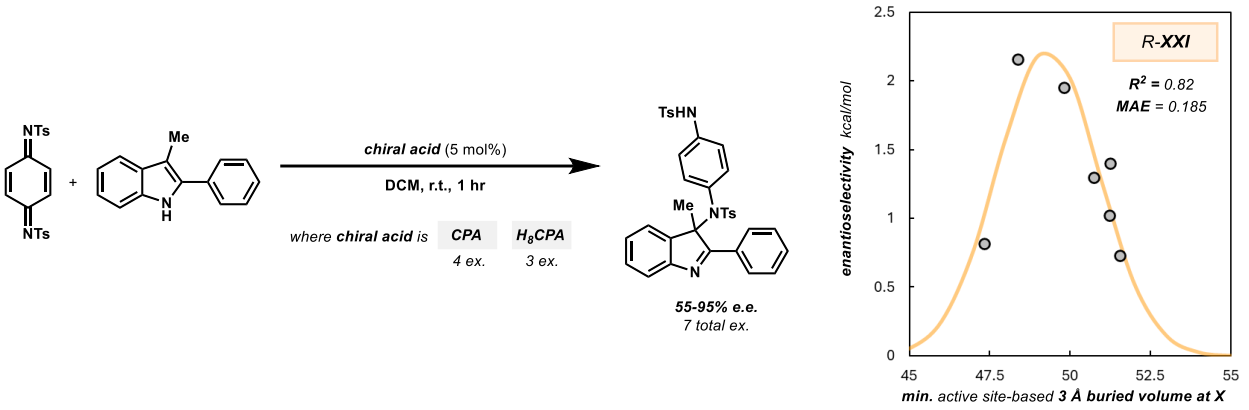


Supplementary Fig. VI-131 | R-**XIX** (sheet name = *CCCC*, A = 0.732, μ = 35.0, σ = 3.53).[14]

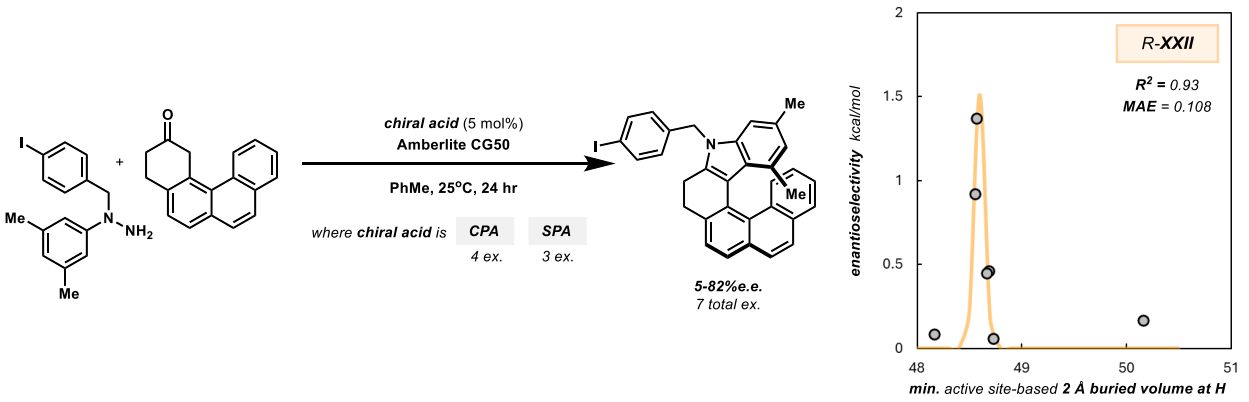


Supplementary Fig. VI-132 | R-**XX** (sheet name = *D*, A = 2.70, μ = 68.1, σ = 0.92).[15]

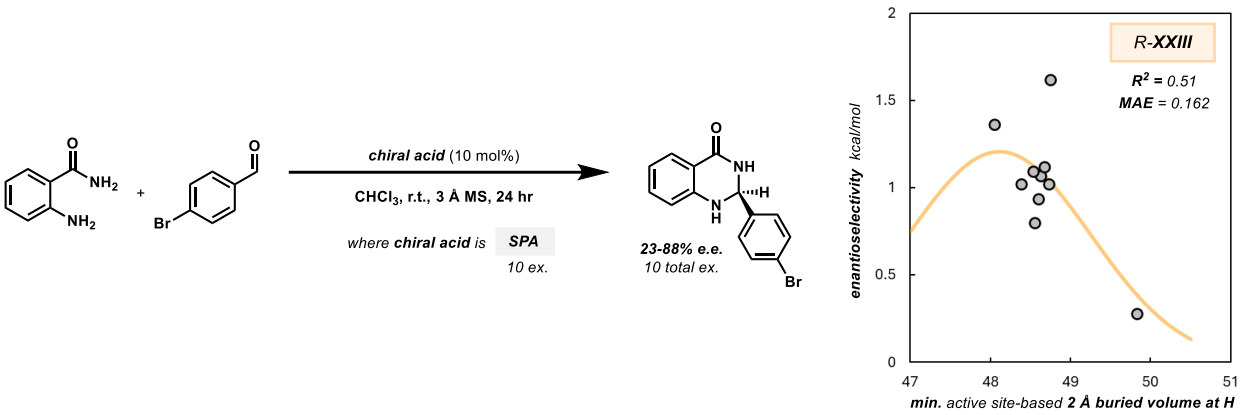
\



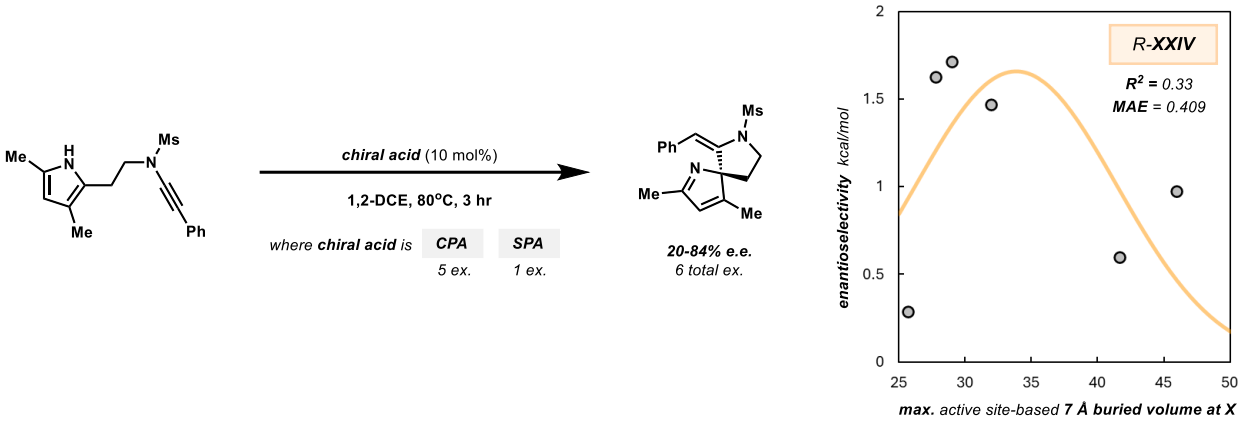
Supplementary Fig. VI-133 | R-**XXI** (sheet name = *DD*, A = 2.22, μ = 49.3, σ = 1.59).[16]



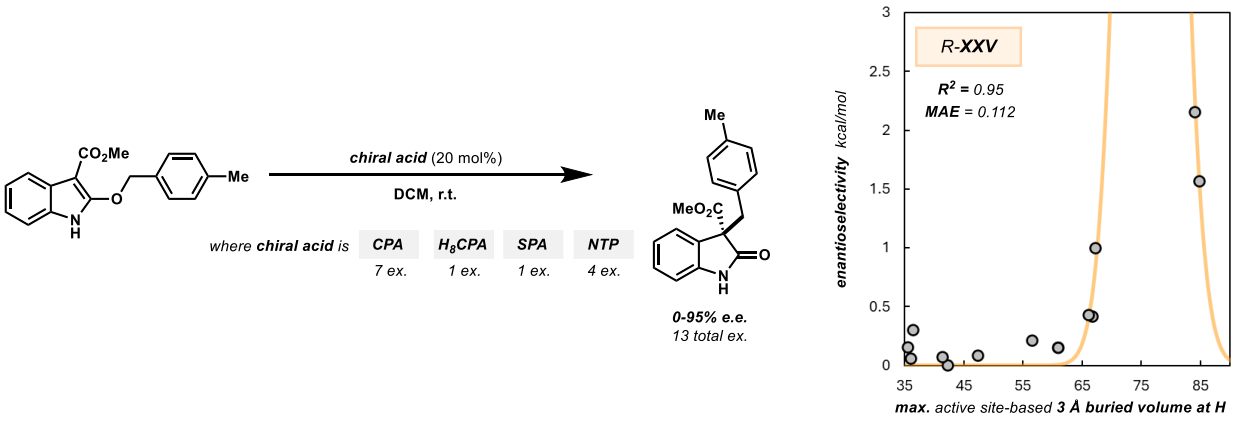
Supplementary Fig. VI-134 | R-**XXII** (sheet name = *DDD*, A = 1.51, μ = 48.6, σ = 0.05).[17]



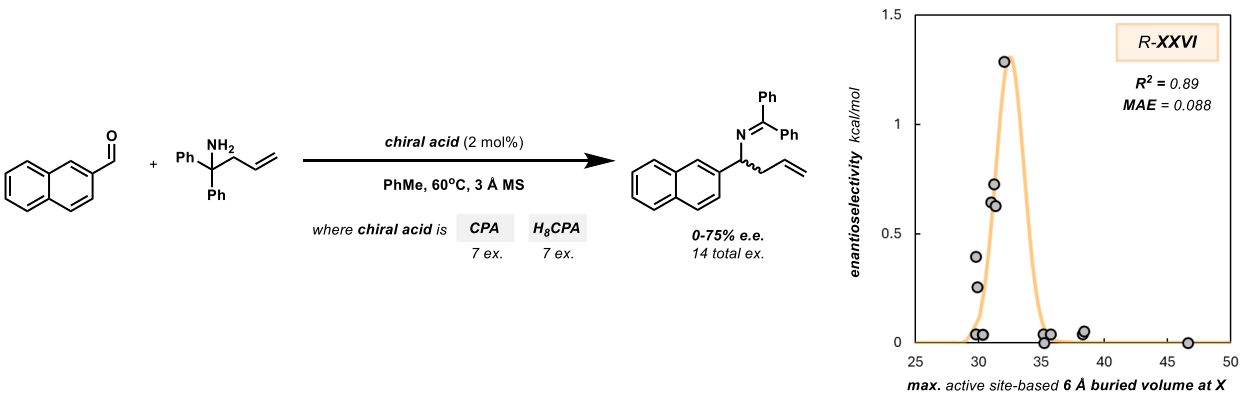
Supplementary Fig. VI-135 | R-**XXIII** (sheet name = *DDDD*, A = 1.21, μ = 48.1, σ = 1.14).[18]



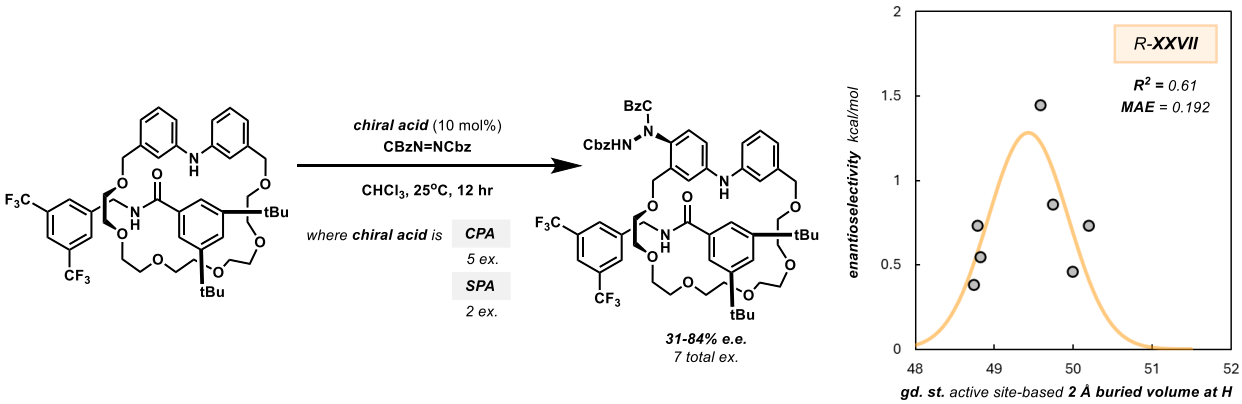
Supplementary Fig. VI-136 | R-**XXIV** (sheet name = *E*, A = 1.66, μ = 33.8, σ = 7.61).[19]



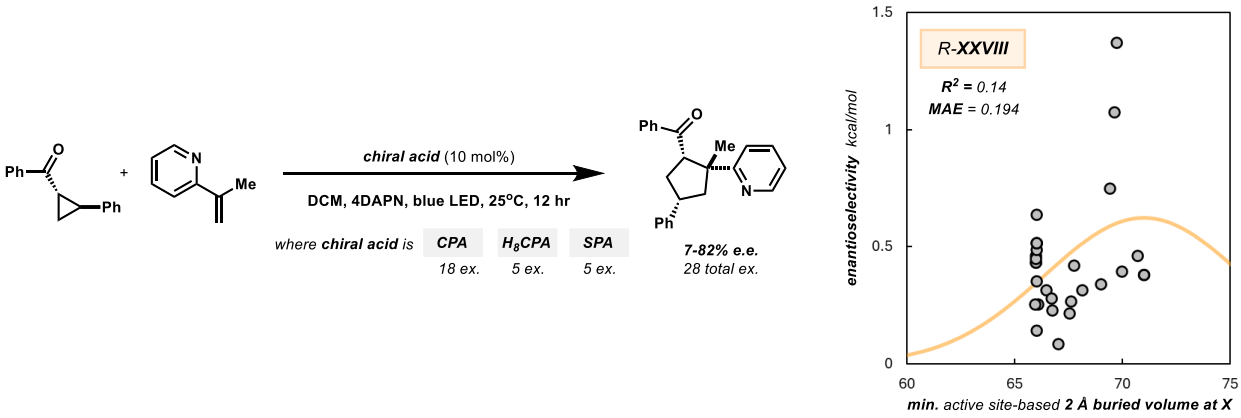
Supplementary Fig. VI-137 | R-**XXV** (sheet name = *EE*, A = 13.6, μ = 76.5, σ = 3.95).[20]



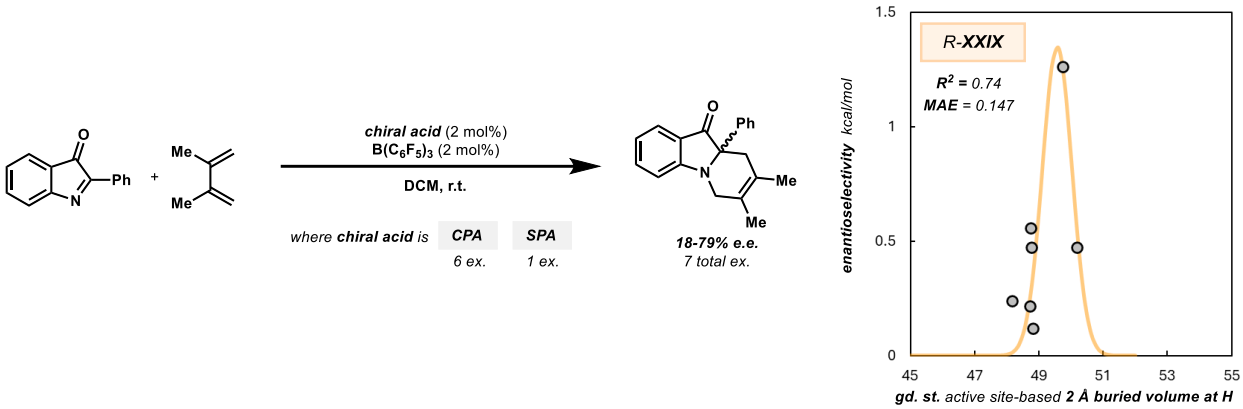
Supplementary Fig. VI-138 | R-**XXVI** (sheet name = *EEE*, A = 1.31, μ = 32.4, σ = 1.12).[21]



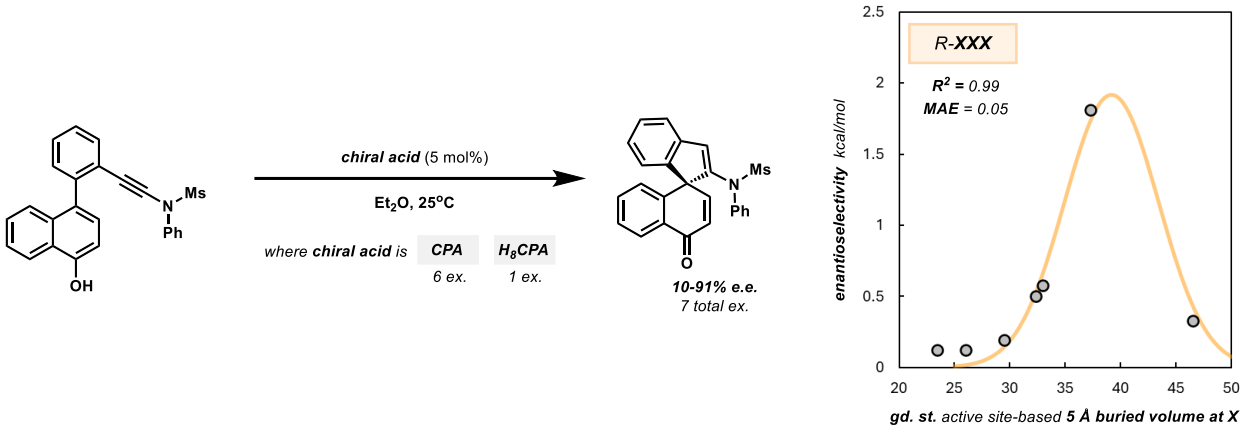
Supplementary Fig. VI-139 | R-**XXVII** (sheet name = *EEEE*, A = 1.28, μ = 49.4, σ = 0.51).[22]



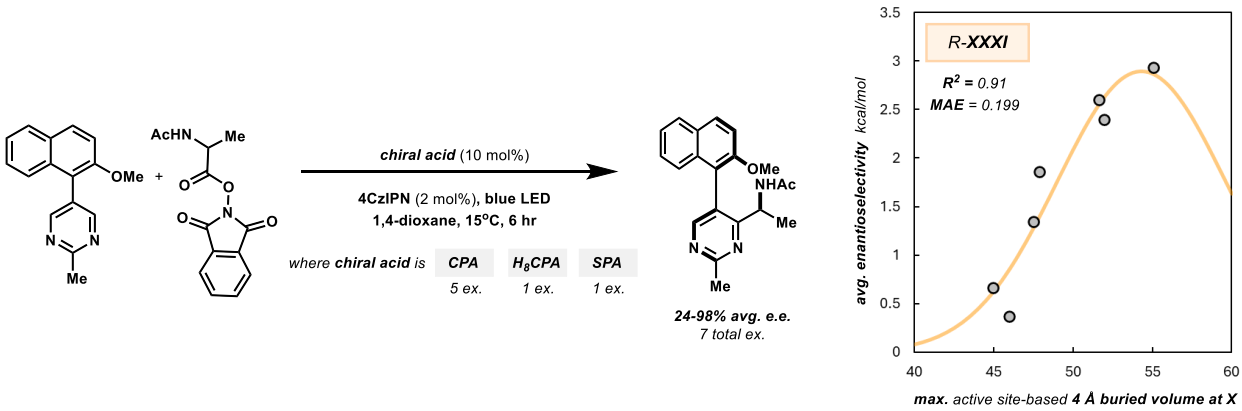
Supplementary Fig. VI-140 | R-**XXVIII** (sheet name = *FF*, A = 0.62, μ = 71.0, σ = 4.60).[23]



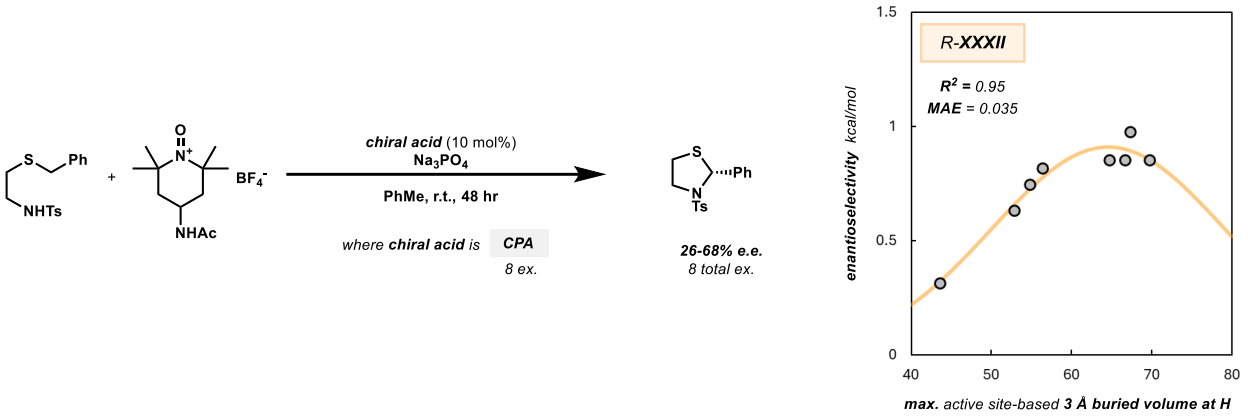
Supplementary Fig. VI-141 | R-**XXIX** (sheet name = *FFF*, A = 1.35, μ = 49.5, σ = 0.46).[24]



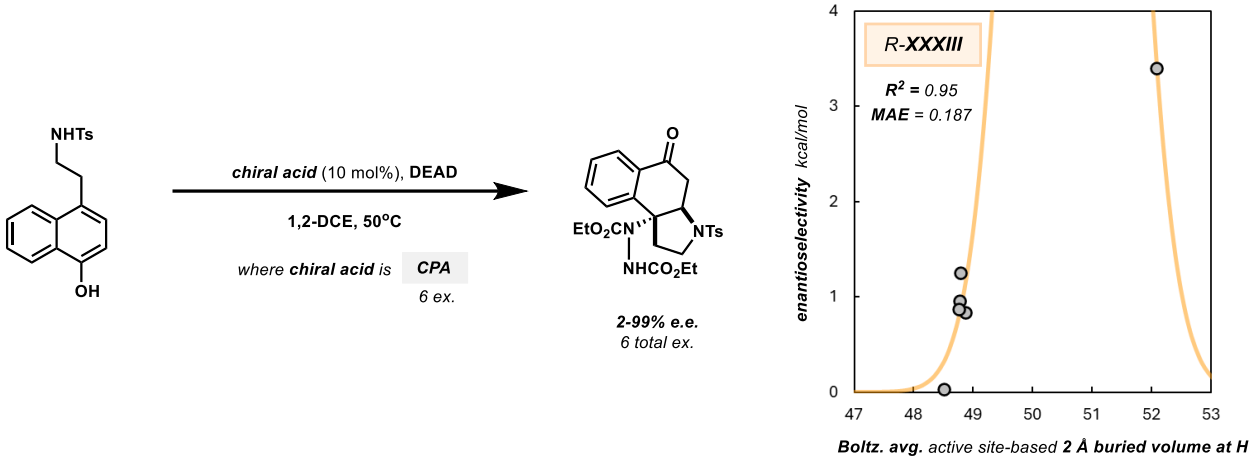
Supplementary Fig. VI-142 | R-**XXX** (sheet name = *G*, A = 4.00, μ = 53.6, σ = 4.76).[19]



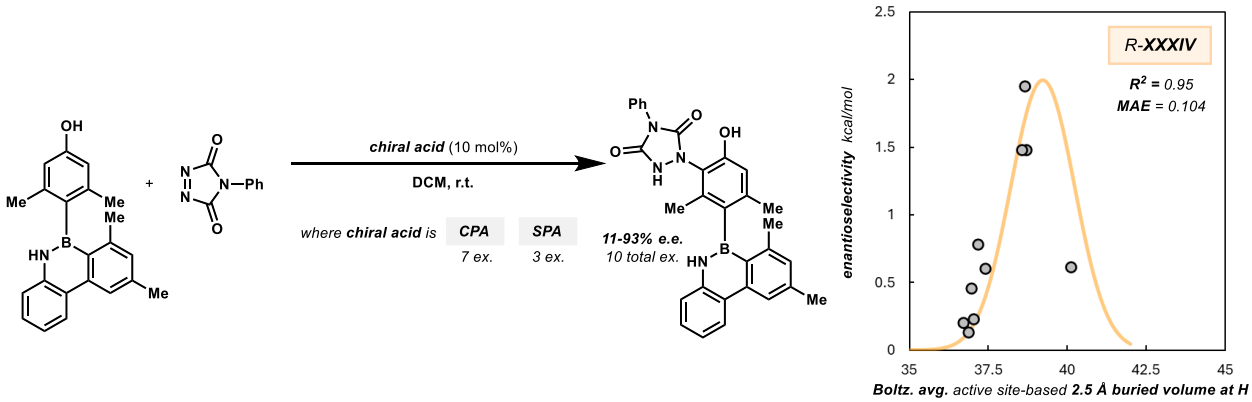
Supplementary Fig. VI-143 | R-**XXXI** (sheet name = *GG*, A = 2.89, μ = 54.3, σ = 5.32).[25]



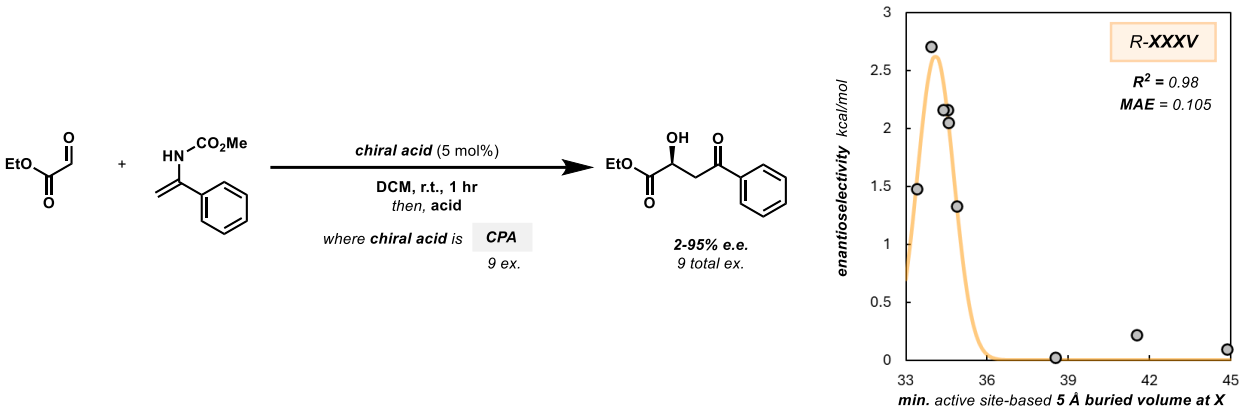
Supplementary Fig. VI-144 | R-**XXXII** (sheet name = *GGG*, A = 0.91, μ = 64.6, σ = 14.5).[26]



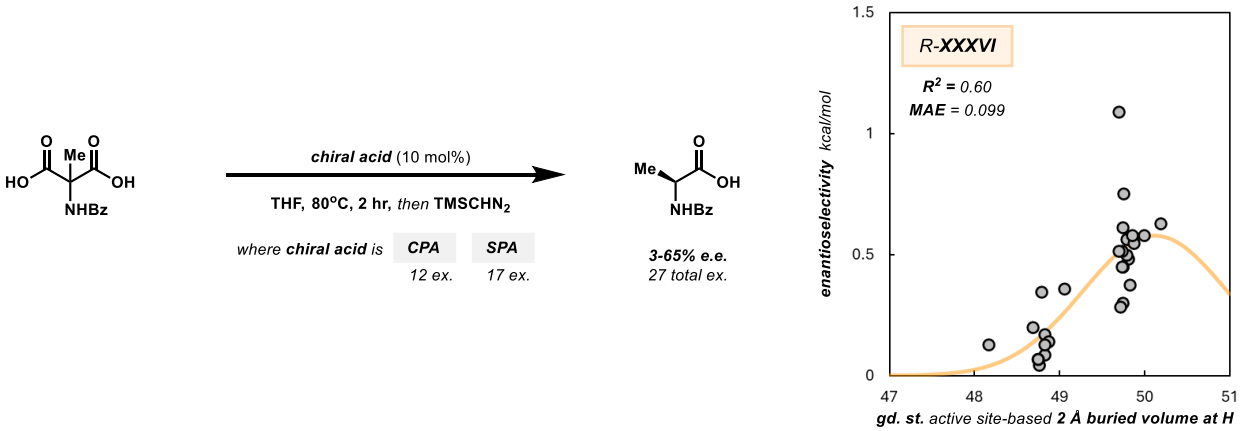
Supplementary Fig. VI-145 | R-**XXXIII** (sheet name = *H*, A = 20.2, μ = 50.7, σ = 0.75).[27]



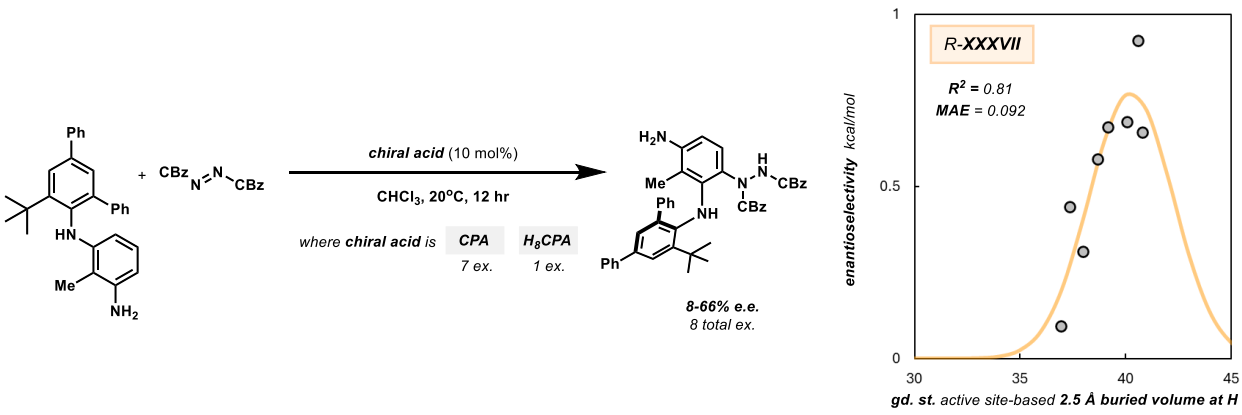
Supplementary Fig. VI-146 | R-**XXXIV** (sheet name = *HH*, A = 1.99, μ = 39.2, σ = 1.01).[28]



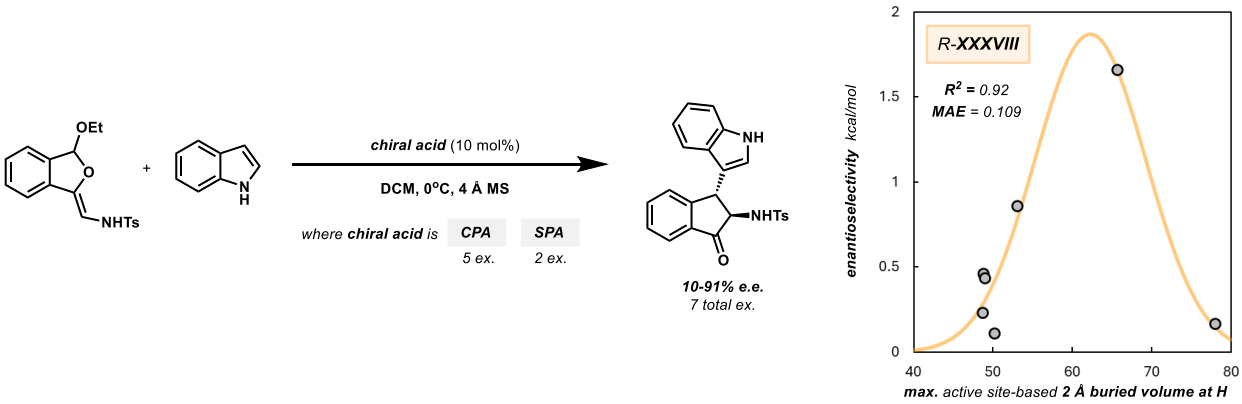
Supplementary Fig. VI-147 | R-**XXXV** (sheet name = *HHH*, A = 2.63, μ = 34.1, σ = 0.67).[29]



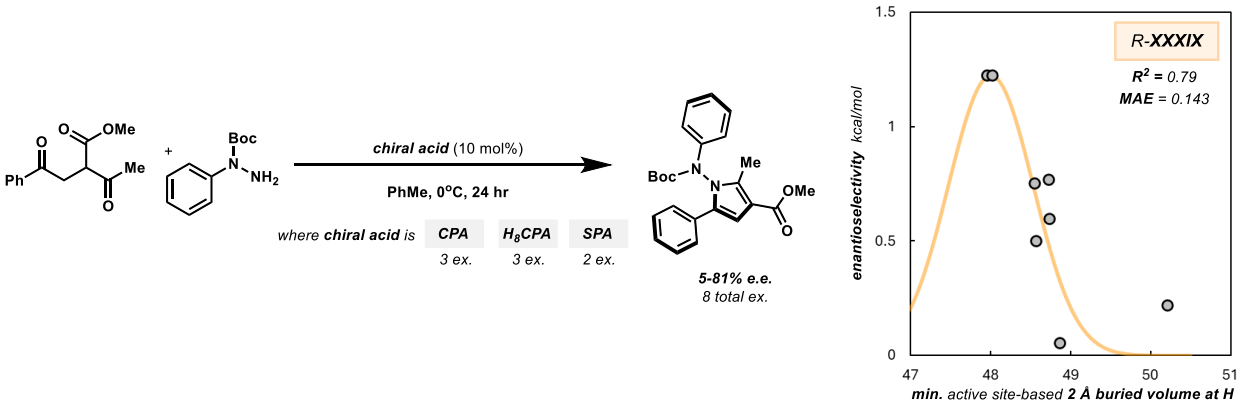
Supplementary Fig. VI-148 | R-**XXXVI** (sheet name = *I*, A = 0.58, μ = 50.1, σ = 0.85).[30]



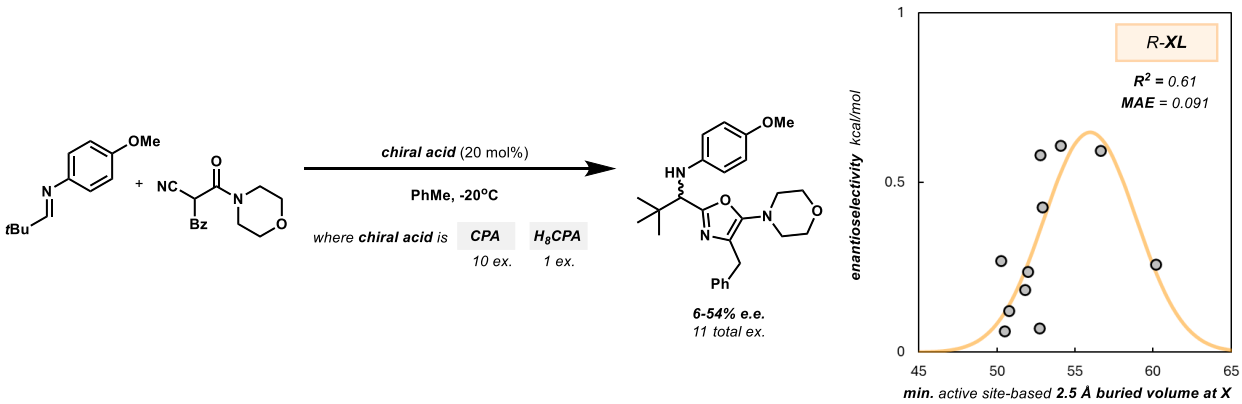
Supplementary Fig. VI-149 | R-**XXXVII** (sheet name = *II*, A = 0.77, μ = 40.3, σ = 2.05).[31]



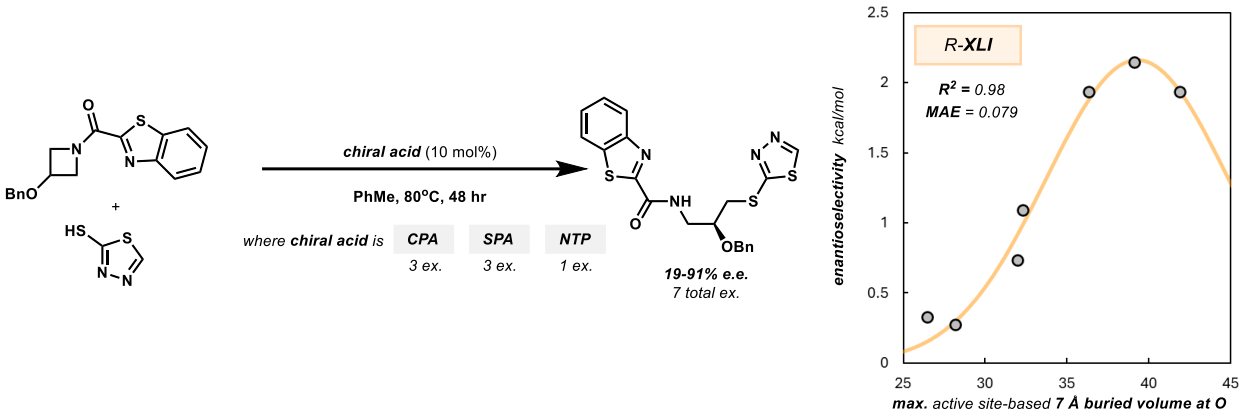
Supplementary Fig. VI-150 | R-**XXXVIII** (sheet name = *III*, A = 1.87, μ = 62.3, σ = 6.98).[32]



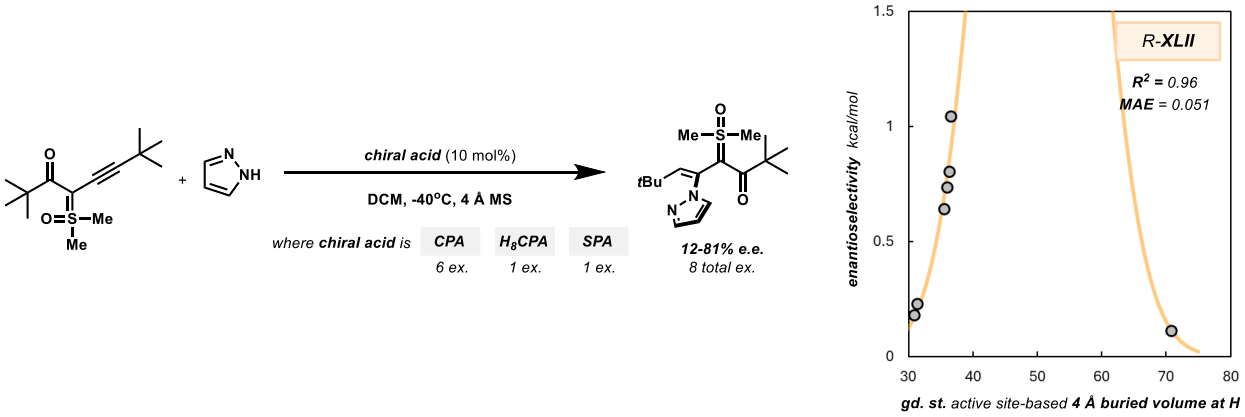
Supplementary Fig. VI-151 | R-**XXXIX** (sheet name = *J*, A = 1.22, μ = 48.0, σ = 0.53).[33]



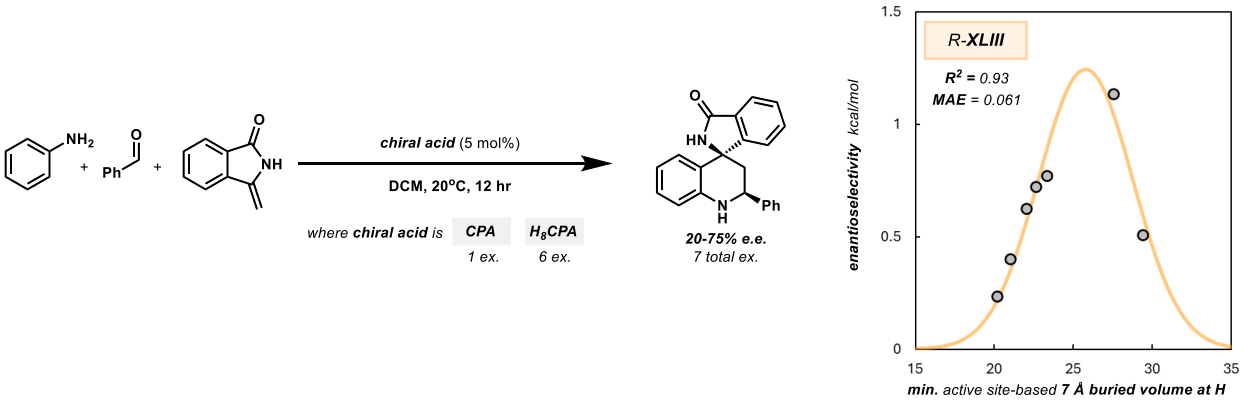
Supplementary Fig. VI-152 | R-**XL** (sheet name = *JJ*, A = 0.64, μ = 56.0, σ = 2.97).[34]



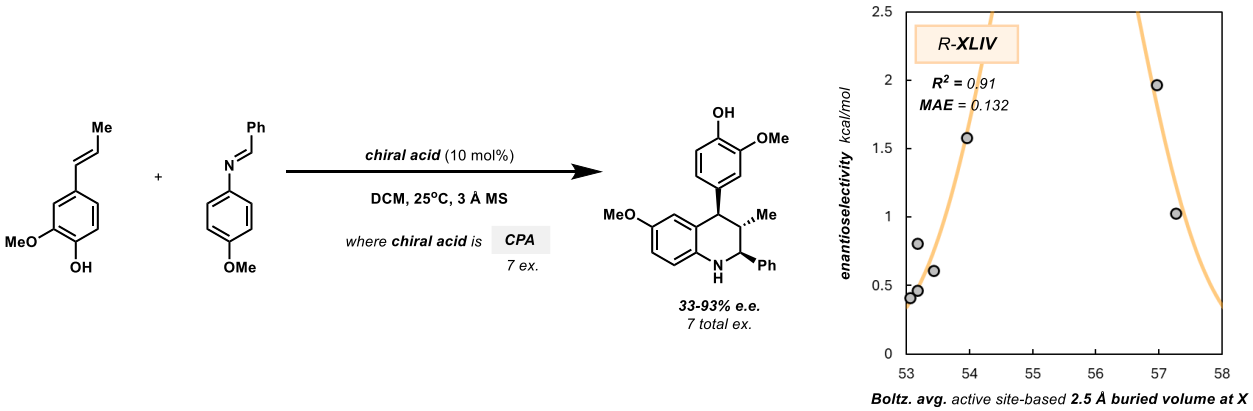
Supplementary Fig. VI-153 | R-**XLI** (sheet name = *JJJ*, A = 2.16, μ = 39.3, σ = 5.59).[35]



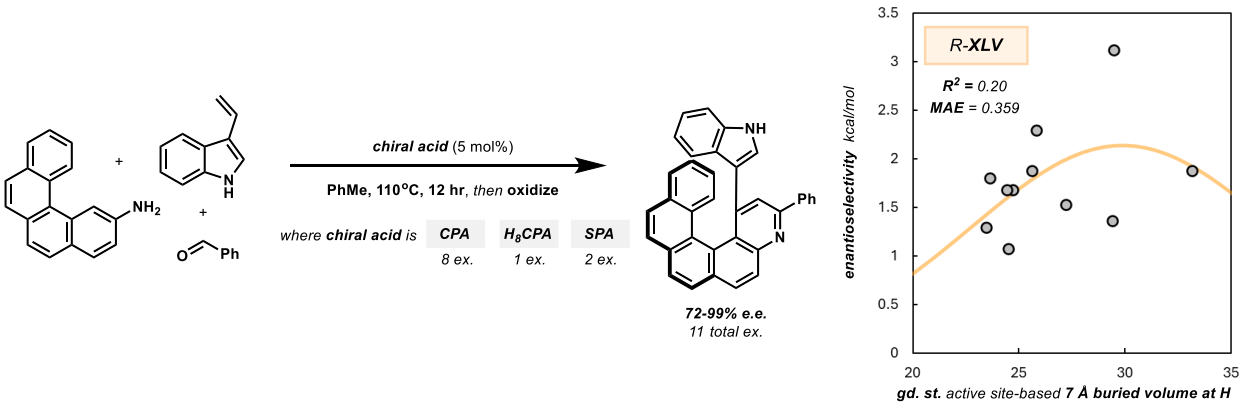
Supplementary Fig. VI-154 | R-**XLII** (sheet name = *K*, A = 4.75, μ = 50.3, σ = 7.52).[36]



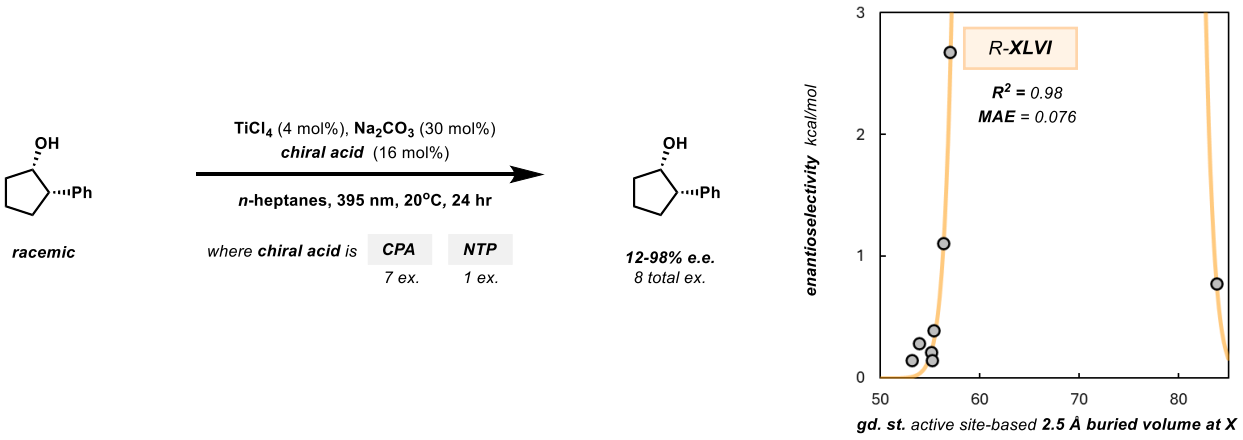
Supplementary Fig. VI-155 | R-**XLIII** (sheet name = *LLL*, A = 1.24 μ = 25.8, σ = 2.99).[37]



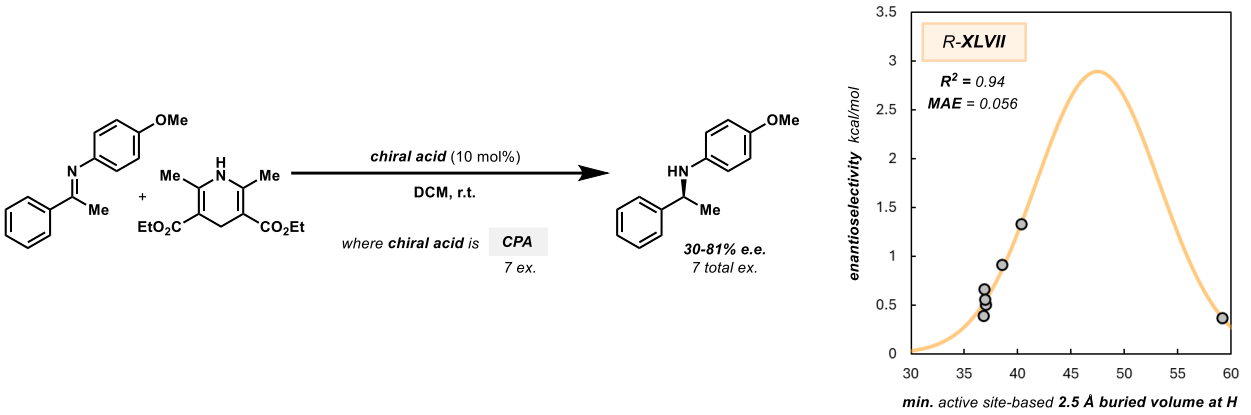
Supplementary Fig. VI-156 | R-**XLIV** (sheet name = *MMM*, A = 4.25, μ = 55.5, σ = 1.12).[38]



Supplementary Fig. VI-157 | R-**XLV** (sheet name = *NNN*, A = 2.14, μ = 29.9, σ = 7.13).[39]

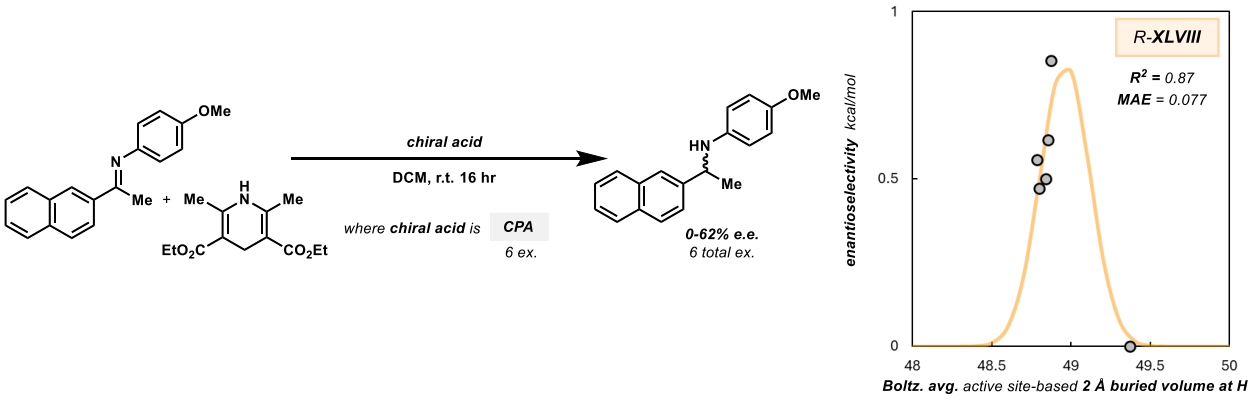


Supplementary Fig. VI-158 | R-**XLVI** (sheet name = *O*, A = 6e3, μ = 69.9, σ = 3.28).[40]

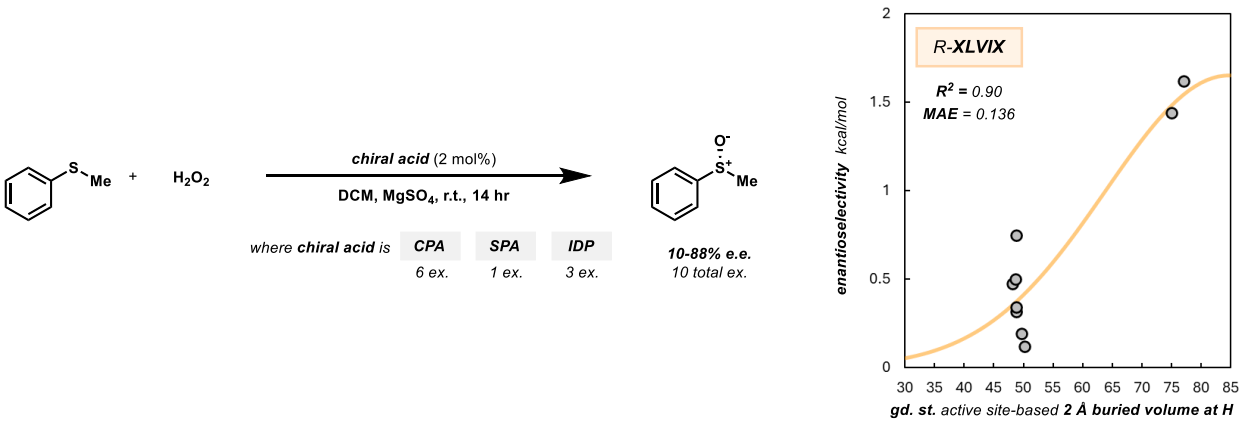


Supplementary Fig. VI-159 | R-**XLVII** (sheet name = *OOO*, A = 2.89, μ = 47.5, σ = 5.76).[41]

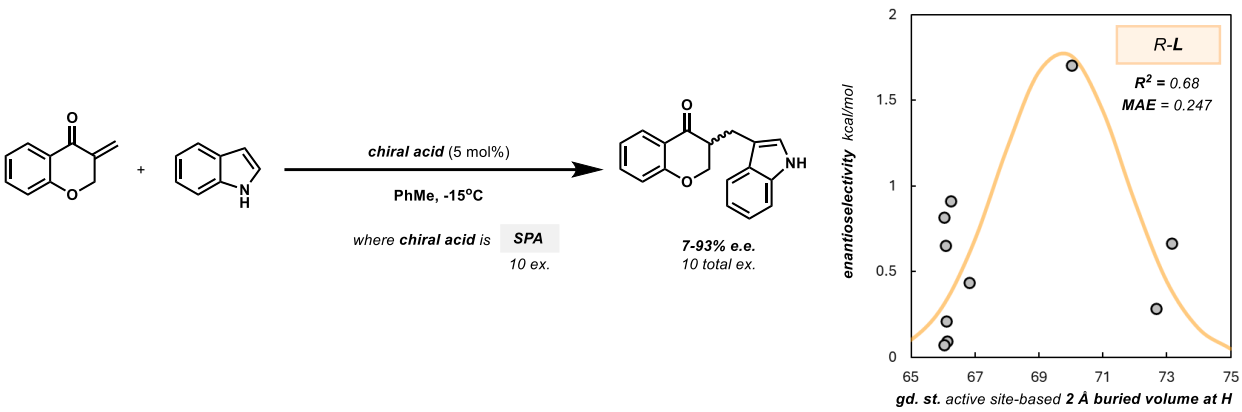
\



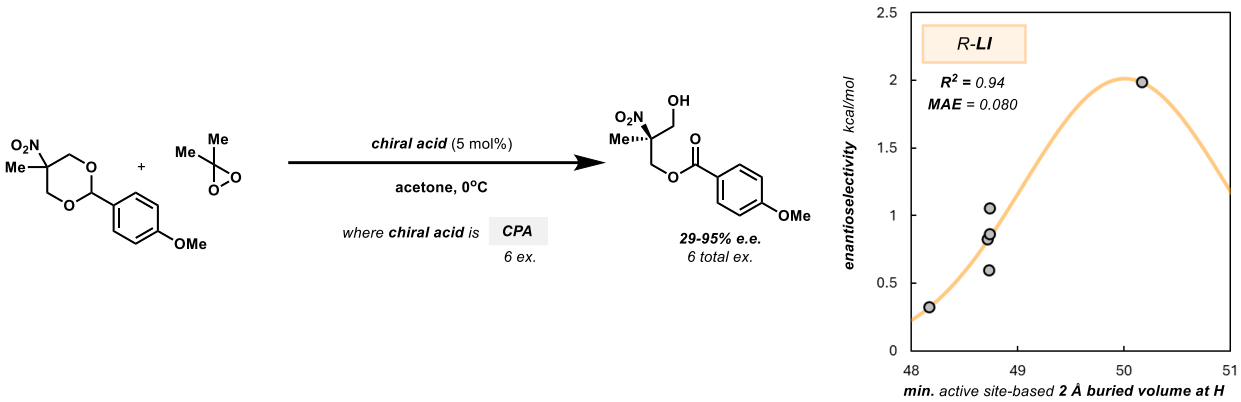
Supplementary Fig. VI-160 | R-**XLVIII** (sheet name = *PPP*, A = 0.85, μ = 49.0, σ = 0.16).[42]



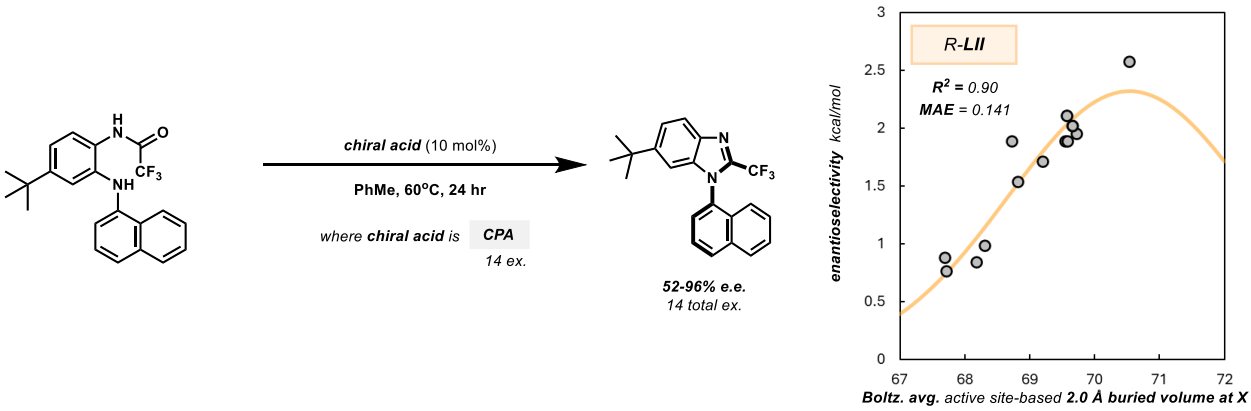
Supplementary Fig. VI-161 | R-**XLVIX** (sheet name = *Q*, A = 1.65, μ = 84.8, σ = 20.8).[43]



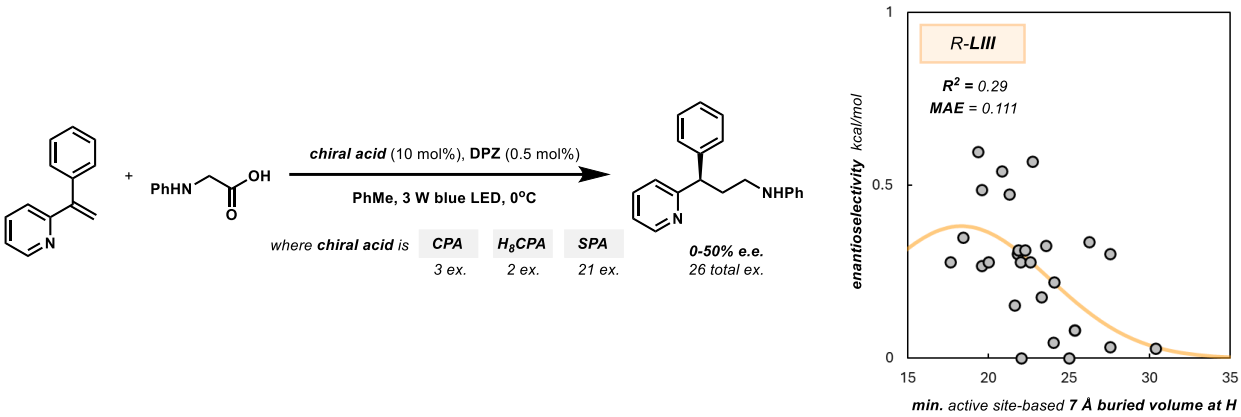
Supplementary Fig. VI-162 | R-**L** (sheet name = *QQQ*, A = 1.43, μ = 48.7, σ = 0.63).[44]



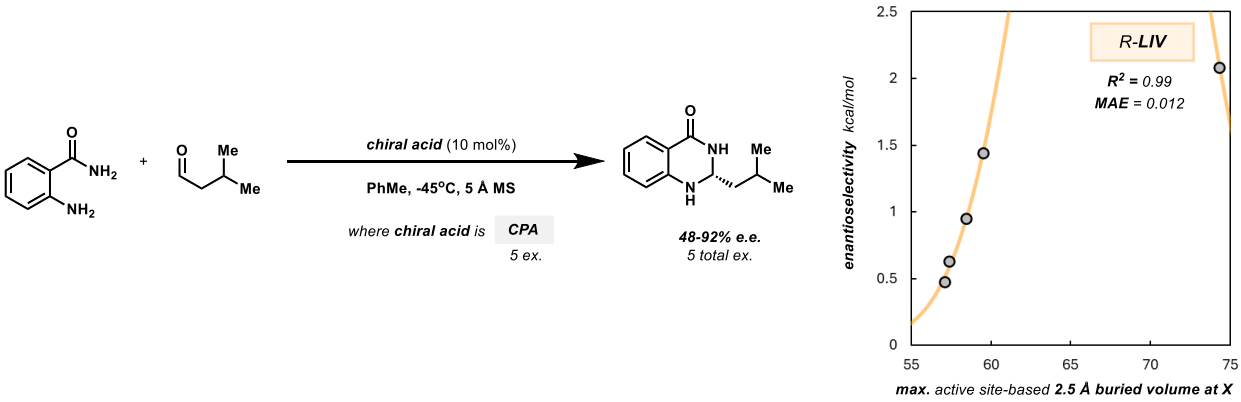
Supplementary Fig. VI-163 | R-**LI** (sheet name = *RRR*, A = 2.02, μ = 50.0, σ = 0.96).[45]



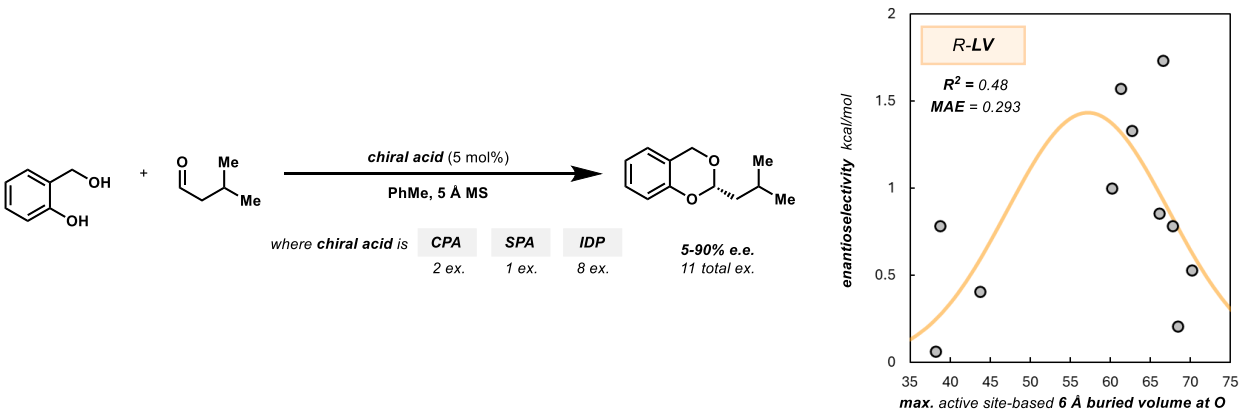
Supplementary Fig. VI-164 | R-**LII** (sheet name = *S*, A = 2.32, μ = 70.5, σ = 1.87).[46]



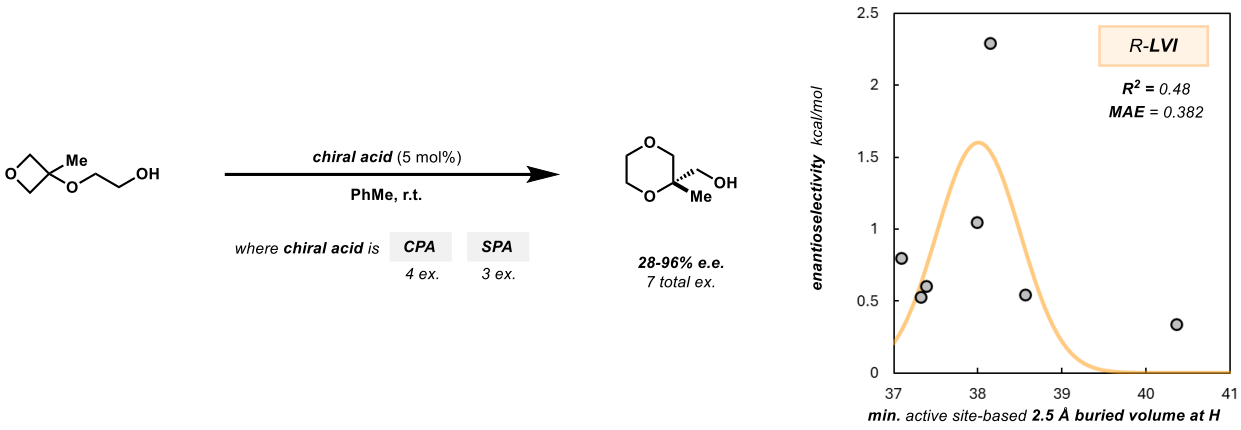
Supplementary Fig. VI-165 | R-**LIII** (sheet name = *SSS*, A = 0.38, μ = 18.3, σ = 5.44).[47]



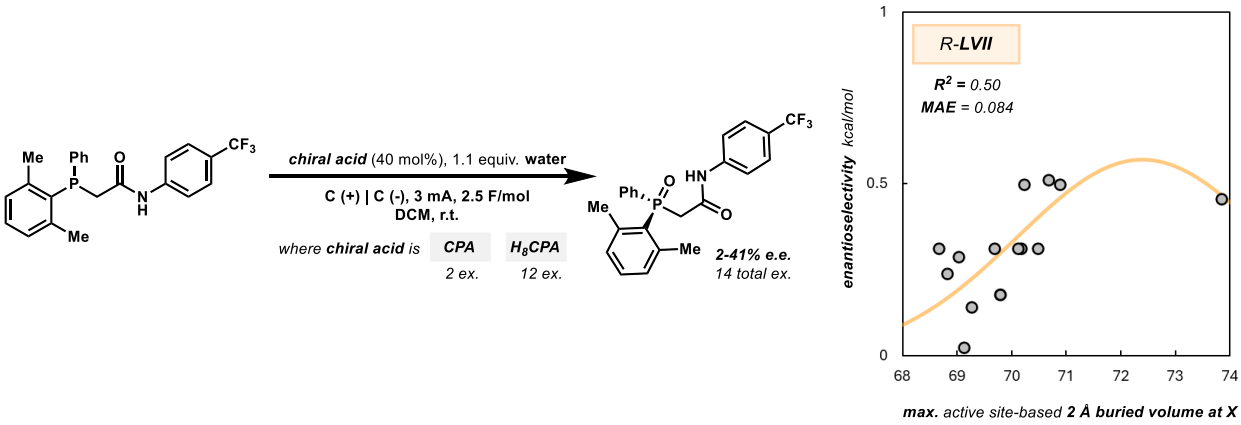
Supplementary Fig. VI-166 | R-**LIV** (sheet name = *TTT*, A = 6.44, μ = 67.6, σ = 4.59).[48]



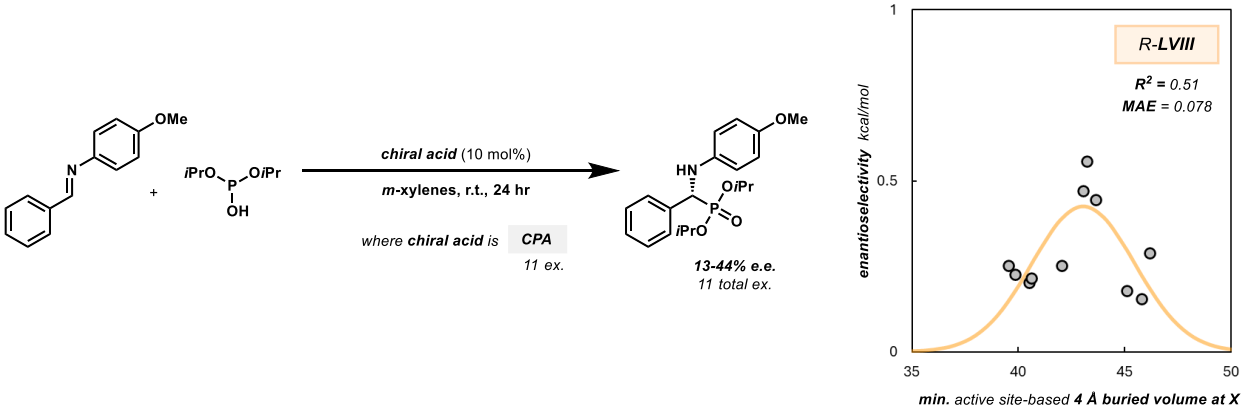
Supplementary Fig. VI-167 | R-**LV** (sheet name = *UUU*, A = 1.43, μ = 57.2, σ = 10.1).[49]



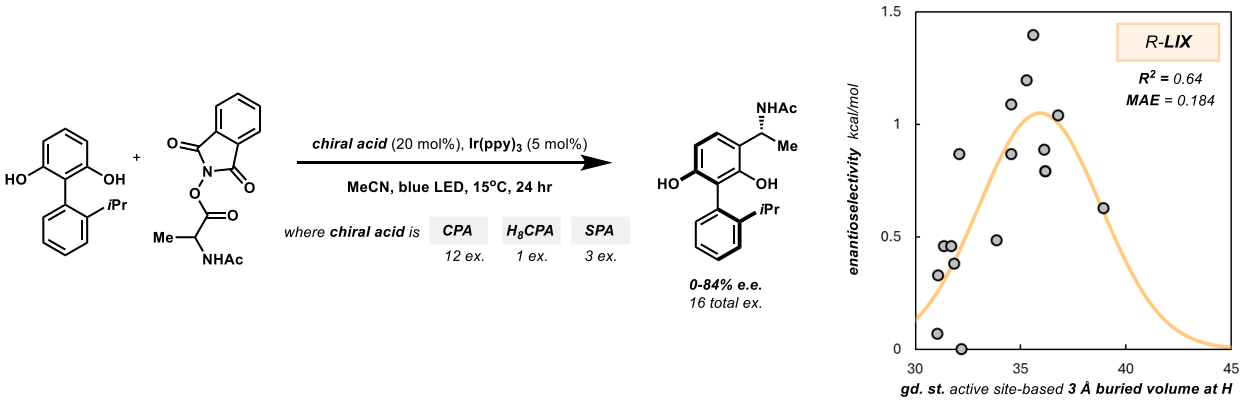
Supplementary Fig. VI-168 | R-**LVI** (sheet name = *V*, A = 1.60, μ = 38.0, σ = 0.50).[50]



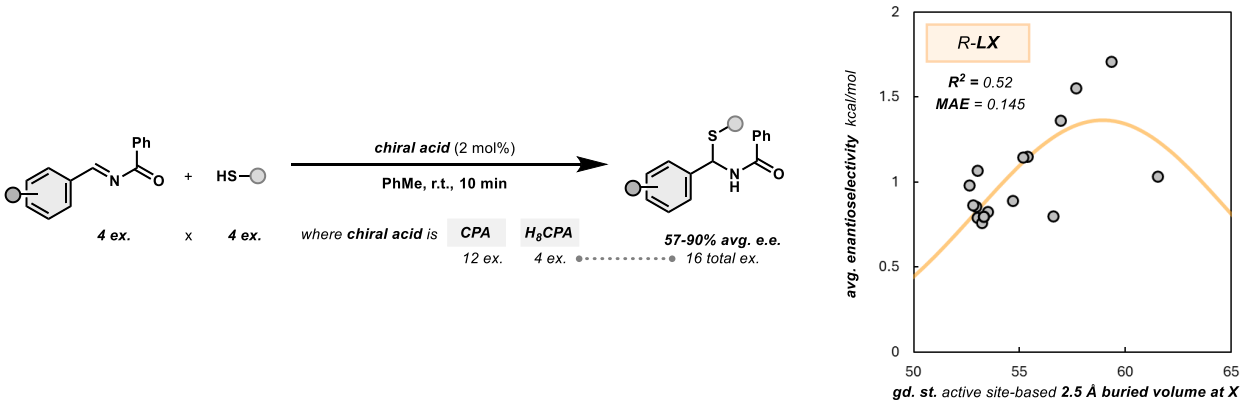
Supplementary Fig. VI-169 | R-**LVII** (sheet name = *VVV*, A = 0.57, μ = 72.4, σ = 2.29).[51]



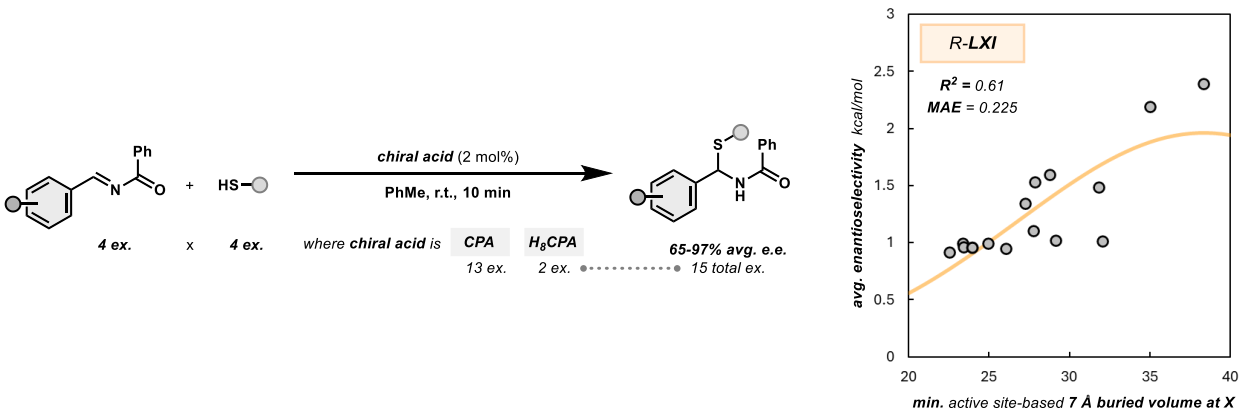
Supplementary Fig. VI-170 | R-**LVIII** (sheet name = *WWW*, A = 0.43, μ = 43.0, σ = 2.42).[52]



Supplementary Fig. VI-171 | R-**LIX** (sheet name = *YYY*, A = 1.05, μ = 35.9, σ = 2.90).[53]



Supplementary Fig. VI-172 | R-**LX** (sheet name = *ZZ*, A = 1.36, μ = 58.9, σ = 5.96).[54]



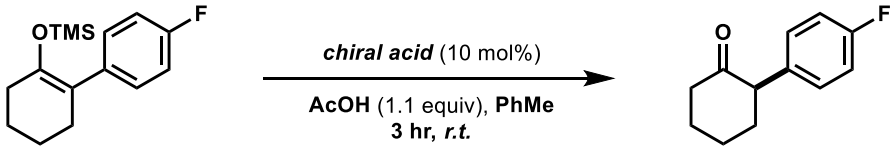
Supplementary Fig. VI-173 | R-**LXI** (sheet name = *ZZ-TEST*, A = 1.96, μ = 38.4, σ = 11.6).[54]

VI.36 – Experimental Benchmarks and Datasets

Benchmark reactions (R-**III** through R-**VII**, ref. **Fig. 3D**) were intentionally selected to span modes of enantioinduction across axial and point chiral induction, thermal and photochemical reactivity, and intra- and intermolecular transformations. Each transformation was previously reported, including a small set of existing data for traditional BAOCs (ex. CPA, NTP, IDP). For each benchmark, the indicated P2-SIAs were deployed. All experimental data is respectively tabulated for reference.

- ***R-III – Enantioselective Silyl Enol Protonation***

The enantioselective protonation reported herein is based on the transformation reported by Cheon, Y., Yamamoto, H. *J. Am. Chem. Soc.* **2008.**[55] (**Supplementary Scheme 1**).



Supplementary Scheme 1 | Enantioselective protonation of model silyl enol ether.

◊ *Experimental Protocol for R-**III***

To a 1-dram vial was charged **silyl enol ether** (100 uL, 0.4 M in DCM, 40 umol, 1 equiv) and acid catalyst (100 uL, 0.02 M in DCM, 2 umol, 0.05 equiv). The solution was dried under positive nitrogen flow, and the residue was charged with 200 uL acetic acid, AcOH (0.22 M in PhMe, 44 umol, 1.1 equiv). The solution was briefly agitated and then left to stand for 3 hours at ambient conditions. After, the reaction mixture was quenched with 10 uL of triethylamine. A small aliquot was removed with a glass pipette tip, diluted with 500 uL of 5% iPrOH in hexanes, filtered through a PTFE syringe filter. The enantioenrichment of **ketone** was recorded by HPLC (ChiralPak OJH, 98:2 Hex:iPrOH, flow rate = 1 mL/min, enantiomers elute = 20.1 min, 32.0 min, λ = 254 nm). The remaining reaction mixture was charged 100 umL of 1,3,5-trifluorobenzene (0.1 M in $CDCl_3$) as internal standard (**^{19}F NMR** δ -106.3). The yield of **ketone** can be calculated from the ^{19}F NMR spectrum of the mixture.

The product **ketone** has spectral characterization consistent with the literature.[56]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.09 (ddd, *J* = 8.4, 5.4, 2.5 Hz, 2H), 7.06 – 6.97 (m, 2H), 3.67 – 3.52 (m, 1H), 2.60 – 2.39 (m, 2H), 2.26 (ddt, *J* = 14.0, 5.7, 2.6 Hz, 1H), 2.17 (dddt, *J* = 7.3, 5.9, 3.6, 1.7 Hz, 1H), 2.07 – 1.91 (m, 2H), 1.91 – 1.74 (m, 2H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -116.10.

◊ *Synthetic Access into Substrate for R-**III***

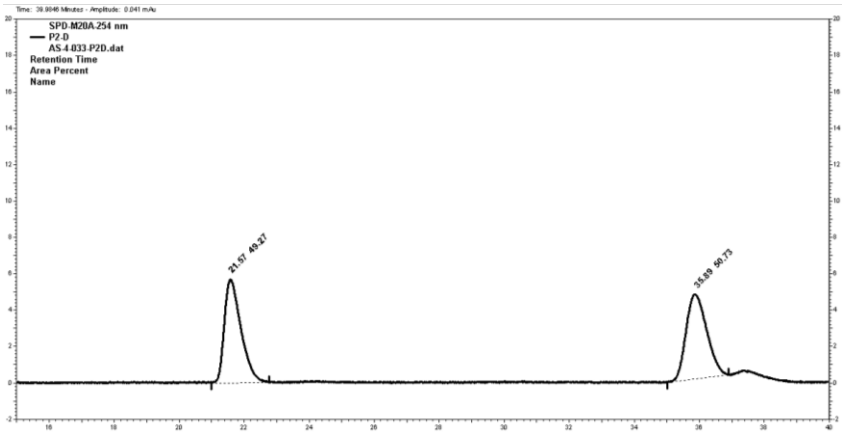
The starting material **silyl enol ether** was obtained according to the protocol by Cheon, Y. et al**.** The prepared substrate displays consistent spectral characterization.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 – 7.26 (m, 2H), 6.94 (t, *J* = 8.9 Hz, 2H), 2.41 – 2.23 (m, 2H), 2.15 (t, *J* = 5.8 Hz, 2H), 1.79 – 1.62 (m, 4H), -0.05 (s, 9H).

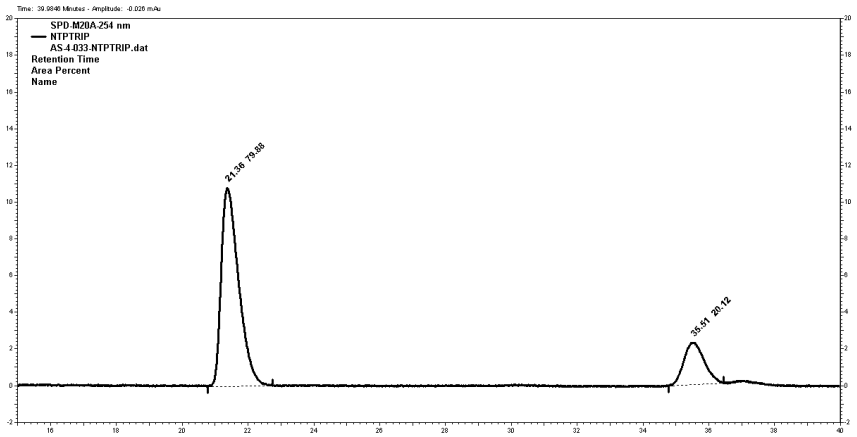
^{19}F NMR (376 MHz, $CDCl_3$) δ -117.39.

◊ *HPLC Traces for Product of R-**III***

Ex. of racemic **ketone** from R-**III**, collected from *racemic* conversion with P2-**3.**



Ex. of enantioenriched **ketone** from R-**III**, collected from reaction with (*S*)-TRIP-**NTP.**



◊ *Experimental Dataset for R-**III***

The complete dataset for R-**III** is printed below (**Supplementary Table 8**). In **Fig. 3D**, the complete dataset was subjected to regression analysis. Later, in **Fig. 3F**, we investigated the coherence across the two subsets, defined by traditional BAOCs (e.g. CPAs, NTPs) and unfamiliar

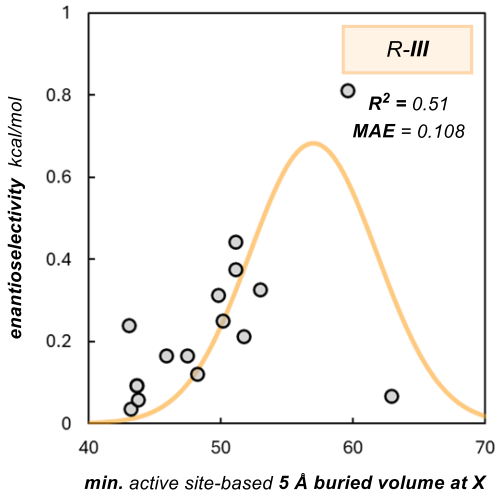
BAOCs. The model presented in **Fig. 3F** was trained *only* on the traditional BAOCs, then *tested* on the out-of-training-set, unfamiliar BAOCs (P2-**1** through P2-**8**).

LABEL	CATALYST	EE	DDG
CPA-C	CPAH_97	8	0.094316
NTP-A	NTPH_0	10.14	0.119701
NTP-B	NTPH_29	27.1	0.326979
NTP-C	NTPH_97	59.76	0.811034
NTP-D	NTPH_934	5.63	0.066303
NTP-E	NTPH_27	30.88	0.375536
NTP-F	NTPH_52	17.86	0.212387
P2-1	P2SIARH_1	20	0.2385
P2-2	P2SIARH_0	14	0.165788
P2-3	P2SIARH_10	3	0.035303
P2-4	P2SIARH_46	-5	0.05887
P2-5	P2SIARH_97	36	0.443378
P2-6	P2SIARH_29	26	0.313056
P2-7	P2SIARH_16	14	0.165788
P2-8	P2SIARH_37	21	0.25078

Supplementary Table VI-26 | Experimental dataset for R-**III**.

◊ *Gaussian Regression Models for R-**III***

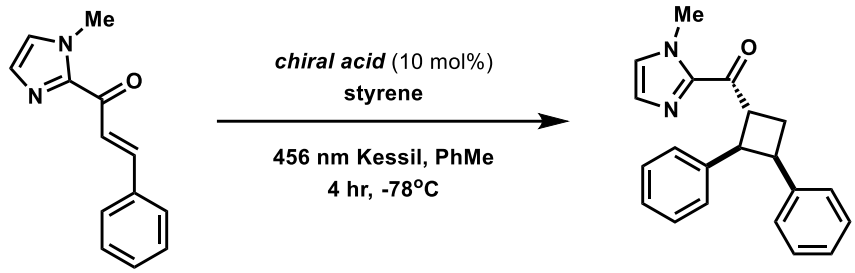
As described above, **Fig. 3D** presents the Gaussian regression performance for the full dataset. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume.



Supplementary Fig. VI-174 | R-**III** from **Fig. 3D** (A = 0.68, μ = 57.0, σ = 4.81).

- ***R-IV – Enantioselective Photochemical [2+2]***

The enantioselective, photochemical [2+2] reported herein is based on the transformation first reported by Sherbrook, E. Yoon, T. et al. *Chem. Sci.* **2020.**[57] and later rendered stereoselective in Sherbrook, E. Yoon, T. et al. *Nat. Comm.* **2021.**[58]



◊ *Experimental Protocol for R-**IV***

Two 1-dram vials were charged with **enone** (100 uL, 0.25 M in DCM, 25 umol, 1 equiv) and acid catalyst (100 uL, 0.025 M in DCM, 2.5 umol, 0.1 equiv). The vials were sealed with Teflon septa and the solution was concentrated under positive nitrogen flow. To the remaining residue was charged styrene (500 uL, 0.5 M in PhMe, 250 umol, 10 equiv). The 1-dram vials were further sealed with electrical tape and irradiated with 456 nm Kessil blue LED for 4 hours atop a dry ice/acetone cooling bath (see **Supplementary Fig. 1** for further details). After, the solutions were removed from the cooling bath, filtered over a short silica plug eluting with ethyl acetate, and concentrated in vacuo to furnish a crude residue containing **cyclobutane.** The residue was resuspended in 400 uL of CDCl3 and a small aliquot was removed with a glass pipette tip, diluted with 500 uL of 5% iPrOH in hexanes, and filtered through a PTFE syringe filter. The enantioenrichment of the major diastereomer (ref. Sherbrook, E. Yoon, T. et al. Nat. Comm. 2021.[58]) of **cyclobutane** was recorded by HPLC (ChiralPak ODH, 95:5 Hex:iPrOH, flow rate = 1 mL/min, enantiomers elute = 14.7 min, 24.5 min, $\lambda = 254$ nm). The remaining solution was charged with 100 uL of 1,4-dinitrobenzene (0.1 M in CDCl3) as internal standard (^{1}H NMR δ: 8.42). The yield and diastereoselectivity of **cyclobutane** can be calculated from the recorded ^{1}H NMR spectrum of the mixture.

Owing to the anticipated variability of irradiation, temperature, and potential deleterious, uncatalyzed [2+2] reactions, each acid catalyst was screened at least twice (up to four times). No individual trial deviated by more than $\pm$2% from the avg. The average over all trials is reported.

The product **cyclobutane** is observed as a mixture of diastereomers. Each diastereomer is already characterized by Sherbrook, et al. and they were not resolved in this work. Instead, only diagnostic ^{1}H signals around δ 5.0 and 4.5 were used for the respective analyses.

◊ *Synthetic Access into Substrate for R-**IV***

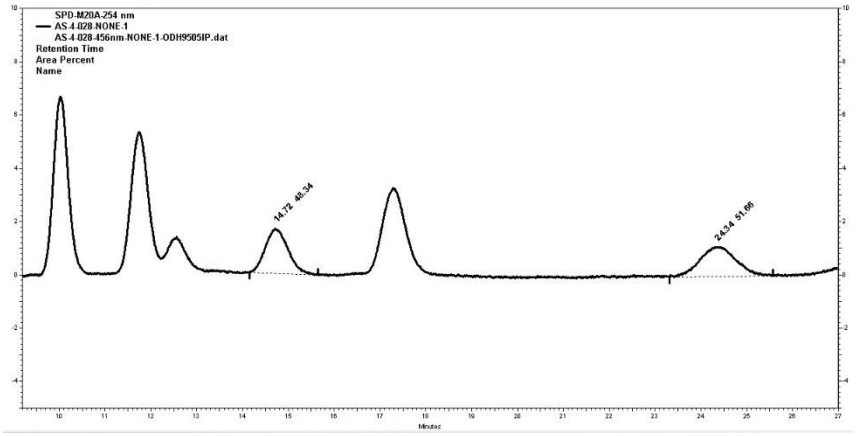
The starting material **enone** was obtained according to the protocol by Sherbrook, E. et al.[57] The prepared substrate displays consistent spectral characterization.

^{1}H NMR (600 MHz, $CDCl_3$) δ 8.08 (d, *J* = 16.0 Hz, 1H), 7.83 (d, *J* = 16.0 Hz, 1H), 7.73 – 7.67 (m, 2H), 7.40 (dd, *J* = 5.1, 1.9 Hz, 3H), 7.22 (s, 1H), 7.08 (s, 1H), 4.10 (s, 3H).

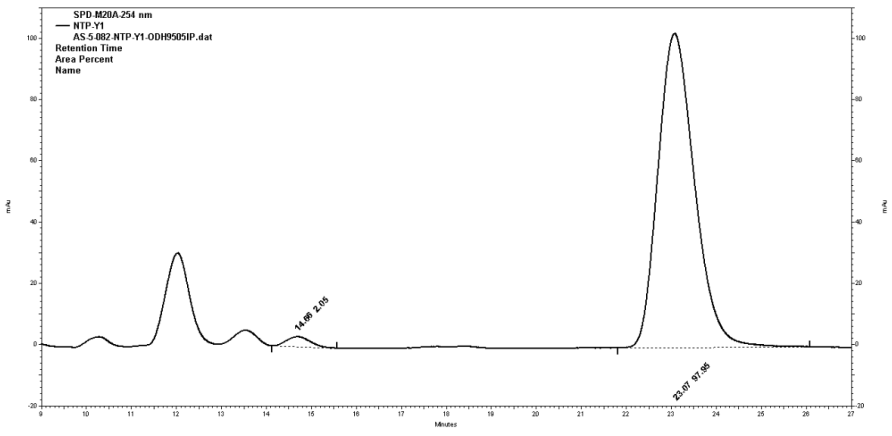
^{13}C NMR (126 MHz, $CDCl_3$) δ 180.53, 144.05, 143.45, 134.95, 130.47, 129.32, 128.85, 128.77, 127.27, 122.79, 36.40.

◊ *HPLC Traces for Product of R-**IV***

Ex. of racemic **cyclobutane**, collected from background, photochemical [2+2]**:**



Ex. of enantioenriched **cyclobutane**, collected from reaction with NTP-**7:**



◊ *Experimental Datasets for R-**IV***

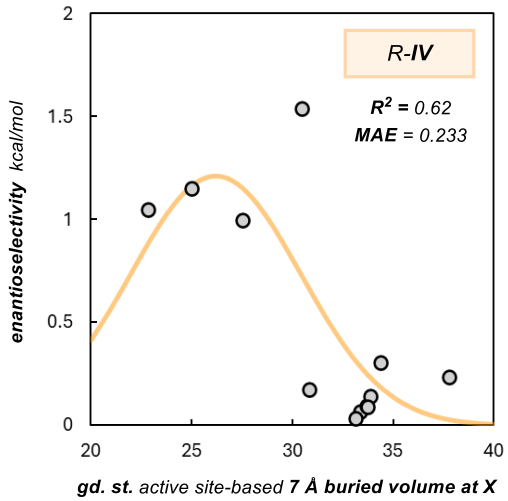
The complete dataset for R-**IV** is printed below (**Supplementary Table 9**, avg. DDG untruncated from obs. *e.e.* in 2+ trials as described in the experimental). In **Fig. 3D**, the model was fit to a dataset containing *only* the original literature-reported BAOCs and P2-SIAs (until P2-**8** in the table). Later in **Fig. 4C**, we investigated the predictive utility of Gaussian regression models with active site-based buried volume. Here, the model was fit to a dataset containing *only* the original literature-reported BAOCs (CPA-**8**, **Ph-**CPA, NTP-**8**, and **Ph-**NTP). It was tested *only* on six new, traditional BAOCs (beginning at TRIP-**CPA** to the end of the table).

LABEL	CATALYST	EE	DDG
NTP-8	NTPH_10	96.29	1.537836
Ph-NTP	NTPH_0	85.75	0.994974
Ph-CPA	CPAH_0	87.42	1.046745
CPA-8	CPAH_10	90.15	1.147146
P2-1	P2SIARH_1	8	0.062134
P2-2	P2SIARH_0	4	0.031017
P2-3	P2SIARH_10	12	0.093451
P2-4	P2SIARH_46	18	0.141038
P2-5	P2SIARH_97	29	0.231392
P2-6	P2SIARH_29	22	0.173336
P2-7	P2SIARH_16	-11	0.085597
P2-8	P2SIARH_37	37	0.301032
TRIP-CPA	CPAH_97	43.37	0.359949
CPA-9	CPAH_52	57.72	0.510153
ArF-CPA	CPAH_41	44.99	0.37555
Mesit-NTP	NTPH_29	13.64	0.106374
Anth-NTP	NTPH_27	11.62	0.090465
NTP-9	NTPH_52	98.25	1.832864

Supplementary Table VI-27 | Experimental dataset for R-**IV.**

◊ *Gaussian Regression Models for R-**IV***

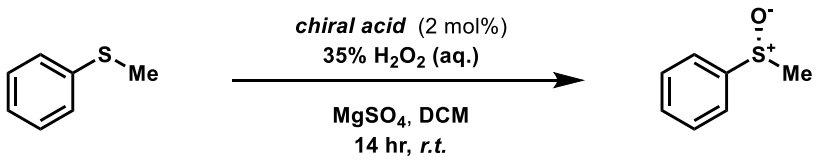
As described above, **Fig. 3D** presents the Gaussian regression performance for the dataset containing only literature-reported BAOCs and P2-SIAs. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume.



Supplementary Fig. VI-175 | R-**IV** from **Fig. 3D** (A = 1.21, μ = 26.2, σ = 4.23).

- ***R-V – Enantioselective Thioether Oxidation***

The enantioselective thioether oxidation reported herein is based on the transformation first reported by Liao, S. List, et al. *J. Am. Chem. Soc.* **2012.**[43] and later optimized by Schwengers, S. List, B. et al. *J. Am. Chem. Soc.* **2021.**[59]



.

◊ *Experimental Protocol for R-V*

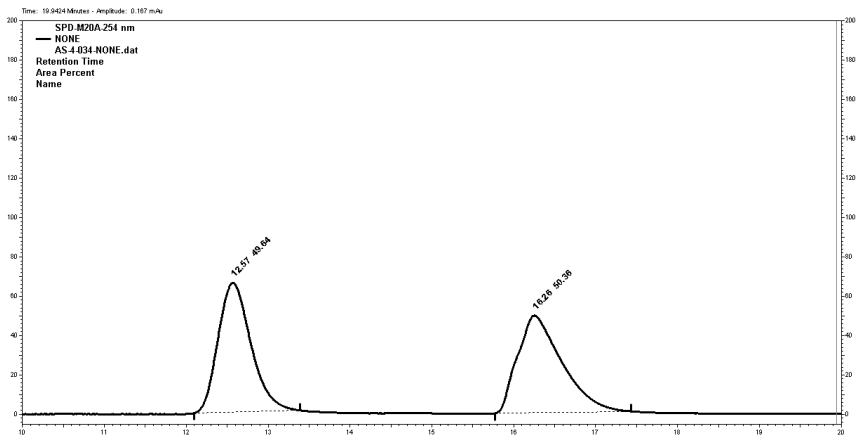
To a 1-dram vial was added magnesium sulfate (30 mg, 250 umol, 2.5 equiv) and a Teflon coated stir bar. The vial was charged with commercial **thioanisole** (500 uL, 0.2 M in DCM, 100 umol, 1.0 equiv), acid catalyst (500 uL, 4 mM in DCM, 2 umol, 0.02 equiv), and hydrogen peroxide (35% aq., 10 uL, 105 umol, 1.05 equiv). The solution was stirred for 14 hours at ambient conditions. After, the reaction mixture was filtered over a small silica plug, concentrated under positive nitrogen flow, and then further dried under vacuum. The residue was resuspended in 400 uL of CDCl3 and a small aliquot was removed with a glass pipette tip, diluted with 500 uL of 5% iPrOH in hexanes, and filtered through a PTFE syringe filter. The enantioenrichment of **sulfoxide** was recorded by HPLC (ChiralPak ODH, 90:10 Hex:iPrOH, flow rate = 1 mL/min, enantiomers elute = 12.6 min, 16.3 min, λ = 254 nm). The remaining solution was charged with 100 uL of 1,4-dinitrobenzene (0.1 M in CDCl3) as internal standard (1H NMR δ: 8.42). The yield of **sulfoxide** was calculated from diagnostic signals in the ^{1}H NMR spectrum of the crude reaction mixture.

◊ *Synthetic Access into Substrate for R-**V***

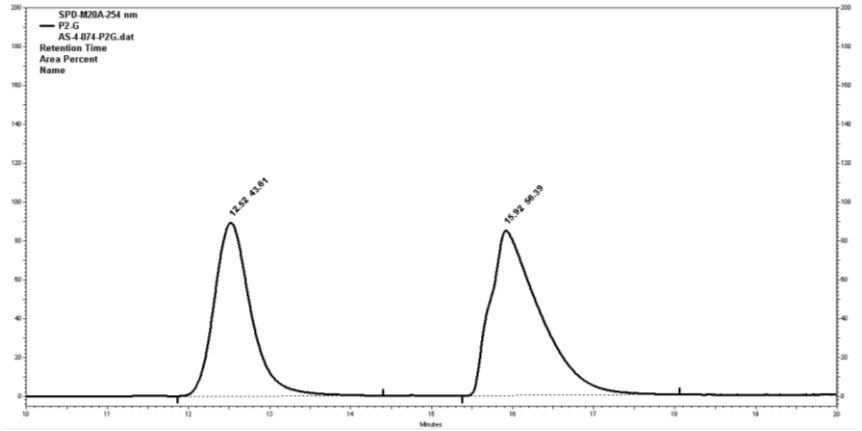
Model substrate, **thioanisole**, was commercially available.

◊ *HPLC Traces for Product of R-**V***

Ex. of racemic **sulfoxide**, collected from background oxidation without acid:



Ex. of enantioenriched **sulfoxide**, collected from reaction with P2-**5**:



◊ *Experimental Datasets for R-**V***

The complete dataset for R-**V** is printed below (**Supplementary Table 10**). In **Fig. 3D**, the model was fitted to the complete dataset below. Only data for P2-SIAs was experimentally collected. The remaining data was derived from the former literature report indicated earlier.

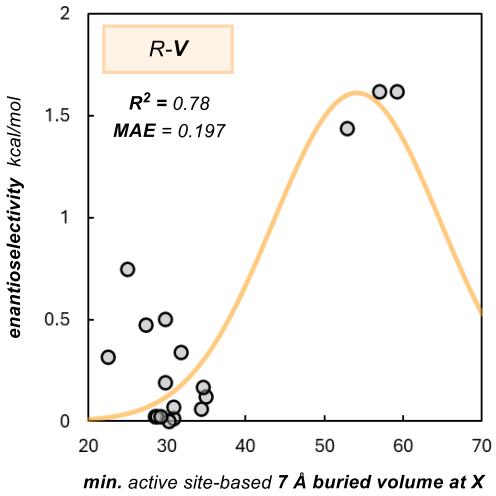
LABEL	CATALYST	EE	DDG
4a	CPAH_0	26	0.313056
4b	CPAH_46	38	0.47064

4c	CPAH_97	10	0.118037
4d	CPAH_27	28	0.338436
4e	CPAH_43	40	0.498391
4g	CPAH_23	56	0.74448
5c	SPAH_97	16	0.189859
6a	IDPH_29	84	1.436618
6b	IDPH_6000	88	1.618487
6c	IDPH_27	88	1.618487
P2-1	P2SIARH_1	2	0.023532
P2-2	P2SIARH_0	1	0.011765
P2-3	P2SIARH_10	2	0.023532
P2-4	P2SIARH_46	0	0
P2-5	P2SIARH_97	13	0.153806
P2-6	P2SIARH_29	6	0.07067
P2-7	P2SIARH_16	2	0.023532
P2-8	P2SIARH_37	5	0.05887

Supplementary Table VI-28 | Experimental dataset for R-**V.**

◊ *Gaussian Regression Models for R-V*

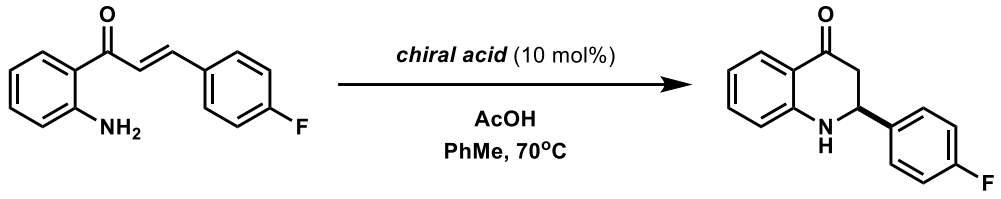
As described above, **Fig. 3D** presents the Gaussian regression performance for the dataset containing the complete dataset. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume for each model.



Supplementary Fig. VI-176 | R-**V** from **Fig. 3D** (A = 1.61, μ = 54.1, σ = 10.6).

- ***R-VI – Enantioselective aza-Michael Addition***

The enantioselective *aza*-Michael addition reported herein is based on the transformation reported by Saito, K. Akiyama, T. et al. *Org. Lett.* **2015.**[5]



◊ *Experimental Protocol for R-**VI***

To a 1-dram vial was charged **enone** (50 uL, 0.2 M in PhMe, 10 umol, 1 equiv) and acid catalyst (50 uL, 0.02 M in PhMe, 1 umol, 0.1 equiv). The vial was sealed and heated to 70ºC for 48 hours. After, the solution was cooled to room temperature, quenched with 10 uL of triethylamine, charged with 400 uL of CDCl3 and 100 uL of 1,3,5-trifluorobenzene (0.1 M in CDCl3, 10 umol) as internal standard (19F NMR δ: -106.3). The yield of aza-Michael adduct X was calculated from the recorded ^{19}F NMR spectrum. Then, the NMR sample was recovered and a 200 uL aliquot was removed, loaded onto a preparative TLC plate, and developed with 100% PhMe. The product **aza-Michael adduct** was carefully separated from any remaining starting material enone and charged with 500 uL of 5% iPrOH in hexanes (by volume). The suspension was drawn into a syringe and filtered through a PTFE syringe filter. The enantioenrichment was recorded by HPLC (ChiralPak ADH, 94:6 Hex:iPrOH, flow rate = 1 mL/min, enantiomers elute = 22.9 min, 30.8 min, λ = 254 nm). The product **aza-Michael adduct** spectral characterization is consistent with the literature.

^{1}H NMR (600 MHz, $CDCl_3$) δ 7.88 (dd, *J* =7.9, 1.9 Hz, 1H), 7.47-7.42 (m, 2H), 7.36 (ddd, *J* = 8.5, 7.0, 1.6 Hz, 1H), 7.25 (t, *J* = 7.6 Hz, 1H), 7.16 (dd, *J* = 14.4, 7.4 Hz, 1H), 7.09 (t, *J* = 8.5 Hz, 2H), 6.84-6.78 (m, 1H), 6.72 (d, *J* = 8.2 Hz, 1H), 2.94-2.64 (m, 2H)

^{19}F NMR (565 MHz, $CDCl_3$) δ -113.35.

◊ *Synthetic Access into Substrate for R-**VI***

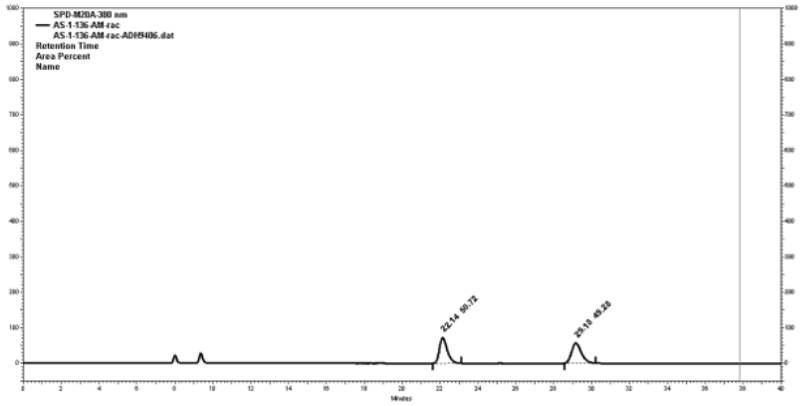
The starting material **enone** was obtained according to the protocol by Saito, K. et al**.**[5] The prepared substrate displays consistent spectral characterization.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.91-7.81 (m, 1H), 7.79-7.50 (m, 4H), 7.33 (t, J = 7.7 Hz, 1H), 7.11 (t, J = 8.6 Hz, 2H), 6.85-6.71 (m, 2H).

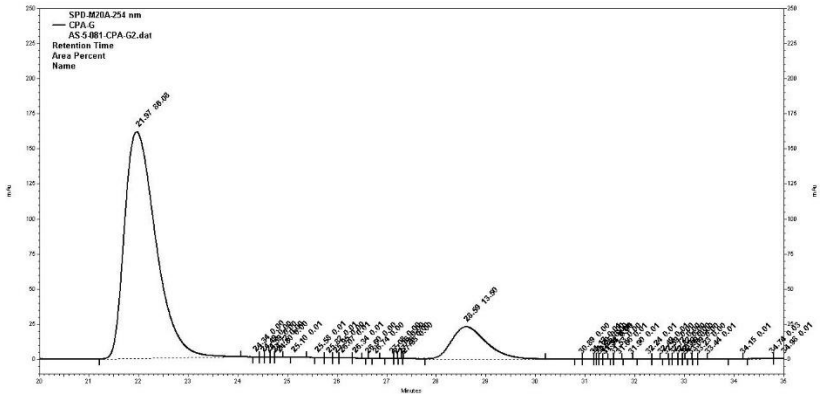
^{19}F NMR (376 MHz, $CDCl_3$) δ -108.87.

- ***HPLC Traces for Product of R-VI***

Ex. of racemic **aza-Michael adduct**, collected with trifluoroacetic acid***:***



Ex. of enantioenriched **aza-Michael adduct**, collected with TRIP-**CPA**:



◊ *Experimental Datasets for R-**VI***

The complete dataset for R-**VI** is printed below (**Supplementary Table 11**). In **Fig. 3D**, the model was fitted to the complete dataset below.

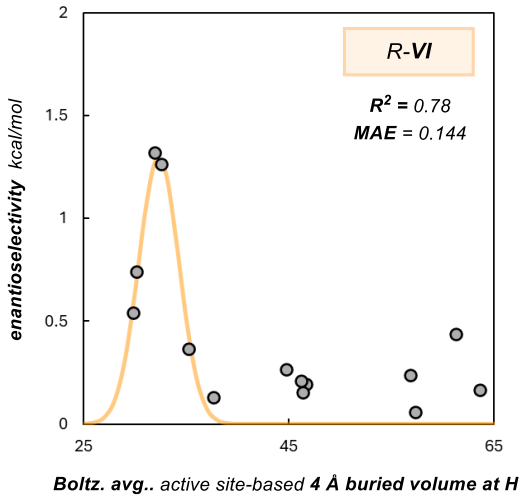
LABEL	CATALYST	EE	DDG
ArF-CPA	CPAH_41	72.89	1.262856
CPA-6	CPAH_46	26.01	0.362911
CPA-8	CPAH_52	49.55	0.740671
Mesit-CPA	CPAH_29	74.71	1.317362
Ph-CPA	CPAH_0	37.56	0.538373
TRIP-CPA	CPAH_97	9.42	0.128797
P2-1	P2SIARH_1	14	0.192113
P2-2	P2SIARH_0	11	0.150564
P2-3	P2SIARH_10	19	0.262198
P2-4	P2SIARH_46	12	0.164379
P2-5	P2SIARH_97	31	0.436974
P2-6	P2SIARH_29	4	0.054558
P2-7	P2SIARH_16	15	0.206038

P2-8	P2SIARH_37	17	0.23402

Supplementary Table VI-29 | Experimental dataset for R-**VI.**

◊ *Gaussian Regression Models for R-**VI***

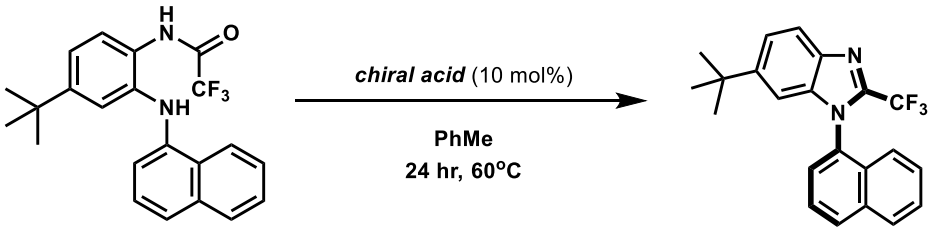
As described above, **Fig. 3D** presents the Gaussian regression performance for the dataset containing the complete dataset. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume.



Supplementary Fig. VI-177 | R-**VI** from **Fig. 3D** (A = 1.30, μ = 32.3, σ = 1.96).

▪ ***R-VII – Atroposelective Cyclodehydration***

The atroposelective cyclodehydration reported herein is based on the transformation reported by Kwon, Y. Miller, S. et al. *J. Am. Chem. Soc.* **2019.**[46] and later revisited by Liles, J. Rouget-Virbel, C. Sigman, M. et al. *Chem.* **2023.**[60]



◊ *Experimental Protocol for R-**VII***

To a 1-dram vial was charged **aniline** (50 uL, 0.2 M in PhMe, 10 umol, 1 equiv) and acid catalyst (50 uL, 0.02M in PhMe, 1 umol, 0.1 equiv). The solution was heated to 60°C for 24 hours. After, the solution was cooled to room temperature, the PhMe was removed under positive nitrogen flow, and the crude residue was further dried under vacuum. The residue was resuspended in 400 uL

CDCl3 and charged with 100 uL of triethylamine (0.1 M in CDCl3) and 100 uL of 1,3,5-trifluorobenzene (0.1 M in CDCl3, 10 umol) as internal standard (19F NMR δ: -106.3). The yield was calculated from the recorded 19F NMR spectrum. Then, the NMR sample was recovered and a 200 uL aliquot was removed, loaded onto a preparative TLC plate, and developed with 100% PhMe. The product **benzimidazole** was carefully separated from any remaining **aniline** and charged with 500 uL of 5% iPrOH in hexanes (by volume). The suspension was drawn into a syringe and filtered through a PTFE syringe filter. The enantioenrichment was recorded by HPLC (ChiralPak ADH, 96:4 Hex:iPrOH, flow rate = 1 mL/min, enantiomers elute = 6.7 min, 8.3 min, λ = 300 nm). The product **benzimidazole** characterization consistent with the literature.

^{1}H NMR (600 MHz, $CDCl_3$) δ 8.17 (d, J = 8.3 Hz, 1H), 8.05 (dd, J = 8.6, 4.3 Hz, 2H), 7.75-7.66 (m, 2H), 7.65-7.58 (m, 3H), 7.13-7.09 (m, 1H), 6.85 (d, J = 1.7 Hz, 1H), 1.24 (s, 9H).

^{19}F NMR (565 MHz, $CDCl_3$) δ -61.48.

◊ *Synthetic Access into Substrate for R-**VII***

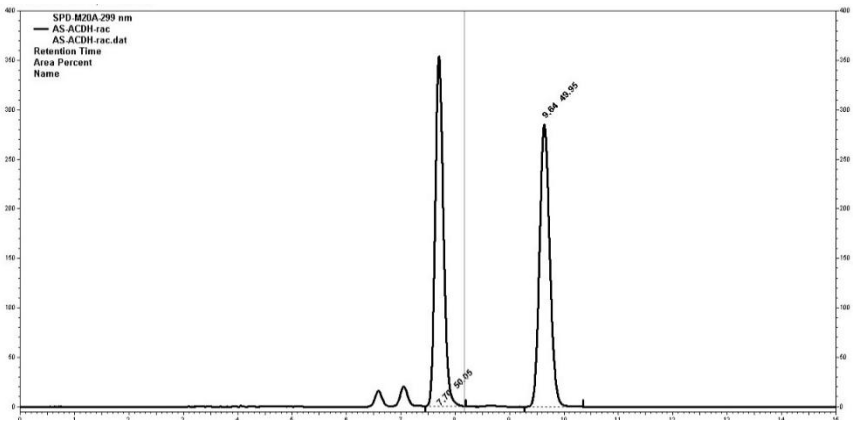
The **aniline** was obtained according to the protocol reported by Liles, J. et al. *Chem* **2023.**[60] The prepared substrate displays consistent spectral characterization.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.38 (s, 1H), 8.11-8.03 (m, 2H), 7.97-7.89 (m, 1H), 7.63-7.56 (m, 2H), 7.53 (d, J = 8.2 Hz, 1H), 7.37-7.31 (m, 2H), 7.28 (d, J = 2.2 Hz, 1H), 6.64 (dd, J = 7.6, 1.1 Hz, 1H), 1.33 (s, 9H).

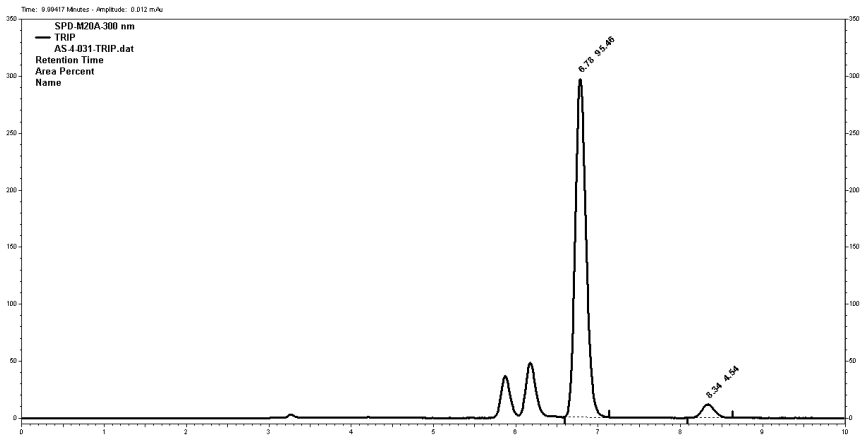
^{19}F NMR (376 MHz, $CDCl_3$) δ -75.93.

◊ *HPLC Traces for Product of R-**VII***

Ex. of racemic **benzimidazole**, collected with trifluoroacetic acid*:*



Ex. of enantioenriched **benzimidazole**, collected with TRIP-**CPA:**



◊ *Experimental Datasets for R-VII*

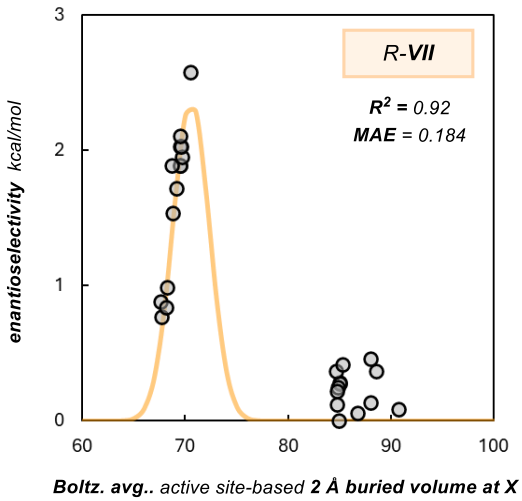
The complete dataset for R-**VII** is printed below (**Supplementary Table 12**). In **Fig. 3D**, the model was fitted to the complete dataset below. Only data for P2-SIAs was experimentally collected. The remaining data is tabulated as derived from the original report (Kwon, et al.[46]).

LABEL	CATALYST	EE	DDG
B1	CPAH_2	58	0.876755
B2	CPAH_116	52	0.762773
B3	CPAH_102	82	1.531022
B4	CPAH_39	89	1.881887
B5	CPAH_321	89	1.881887
B6	CPAH_38	56	0.837541
B7	CPAH_46	63	0.981248
B8	CPAH_612	86	1.711713
B9	CPAH_121	89	1.881887
B10	CPAH_97	91	2.021644
B11	CPAH_5215	90	1.94845
B12	CPAH_7002	91	2.021644
B13	CPAH_934	92	2.103042
B14	CPAH_27	96	2.575368
P2-1	P2SIARH_1	21	0.282128
P2-2	P2SIARH_0	20	0.268312
P2-3	P2SIARH_10	9	0.119436
P2-4	P2SIARH_46	4	0.052967
P2-5	P2SIARH_97	6	0.079504
P2-6	P2SIARH_29	27	0.366423
P2-7	P2SIARH_16	18	0.24085
P2-8	P2SIARH_37	10	0.132792
P2-9	P2SIARH_157	27	0.366423
P2-10	P2SIARH_225	33	0.453726
P2-11	P2SIARH_123	0	0
P2-12	P2SIARH_23	30	0.409642
P2-13	P2SIARH_11	16	0.213592

Supplementary Table VI-30 | Experimental dataset for R-**VII.**

◊ *Gaussian Regression Models for R-**VII***

As described above, **Fig. 3D** presents the Gaussian regression performance for the dataset containing the complete dataset. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume for each model.



Supplementary Fig. VI-178 | R-**VII** from **Fig. 3D** (A = 2.36, μ = 70.5, σ = 1.79).

VI.37 – Empirical Bias Powers Discovery of Stereoselective Solutions

The abstraction of *complementarity* with active site-based buried volume and Gaussian regression promises the discovery of optimal stereoselective solutions. Regardless, synthetic chemists prioritize the evaluation of specific catalysts. These *privileged* and empirically biased choices are often collected without any prior evidence or expectations. We leverage this empirical data to train Gaussian regression models with the corresponding active site-based buried volume (*training*). Often, the model then highlights regions of chemical space that promise optimized stereoselective solutions. Below, we highlight with both experimental and simulated analyses the utility of this approach to anticipate and accurately predict the stereoselectivity of substantially less popular, but more selective BAOCs (*testing*).

To summarize the results, across the nine datasets considered, the empirically popular *training* data generate Gaussian regression models that generalize well to the less popular, *testing* BAOCs (avg. *test* $R^2 = 0.73$, avg. *test* MAE = 0.249 kcal/mol) which include more stereoselective solutions (avg. improvement over training data = +0.45 kcal/mol)

- ***Selection Rate of Top-10 Most Popular Substituents***

Synthetic chemists consider *familiar* catalysts that are frequently stereoselective when they approach reaction development. In the case of BAOCs, *empirical bias* emerges with respect to the *privileged* 3,3'-substituents that routinely furnish stereoselective outcomes, regardless of the active site or backbone. As these groups are time-consuming to install (3-4 synthetic steps required), bias that minimizes the synthetic burden of preparing candidate BAOCs emerges.

We imagined this empirical bias can be approximated by the choices made by researchers in the 55 literature-reported datasets considered in this work. Across all the reactions, 100 unique 3,3'-substituents were deployed. Any single 3,3'-substituent occurred in about 8% of the datasets

(approximately 4 out of 55 datasets). However, a few are especially popular choices, occurring in a disproportionate number of reaction datasets. We assign these to be the *empirically preferred* BAOCs and expect them to be selected for evaluation, regardless of modeling objectives. As a result, we use these *empirical* selections as the training set for Gaussian regression models. The top-10 substituents are printed below with their index, SMILES representation, and usage rate.

INDEX	SMILES	USAGE (%)
97	CC(C)c1cc(C(C)C)c(Br)c(C(C)C)c1	77.8
27	Brc1c2ccccc2cc2ccccc12	77.8
46	FC(F)(F)c1cc(Br)cc(C(F)(F)F)c1	53.7
43	Brc1cc2ccccc2c2ccccc12	51.9
5	Brc1cccc2ccccc12	42.6
11	Brc1ccc2ccccc2c1	37.0
29	Cc1cc(C)c(Br)c(C)c1	35.2
16	Brc1ccc(-c2ccccc2)cc1	33.0
934	Brc1c(C2CCCCC2)cc(C2CCCCC2)cc1C1CCCCC1	33.0
0	Brc1ccccc1	33.0

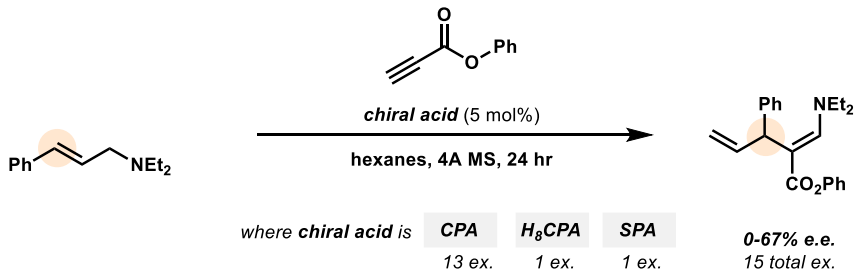
Supplementary Table VI-31 | Top-10 most popular 3,3'-substituents within the 55 literature-reported datasets considered herein.

In the below analyses, we aim to build predictive models using the *least* amount of empirical data. When appropriate (sufficient samples to train a Gaussian regression model), we use either top-3, top-5, or top-10 most popular *splits* to define the training set. These correspond to:

Top-3 =97, 27, 46
Top-5 =97, 27, 46, 43, 5
Top-10 = 97, 27, 46, 43, 5, 11, 29, 16, 934, 0

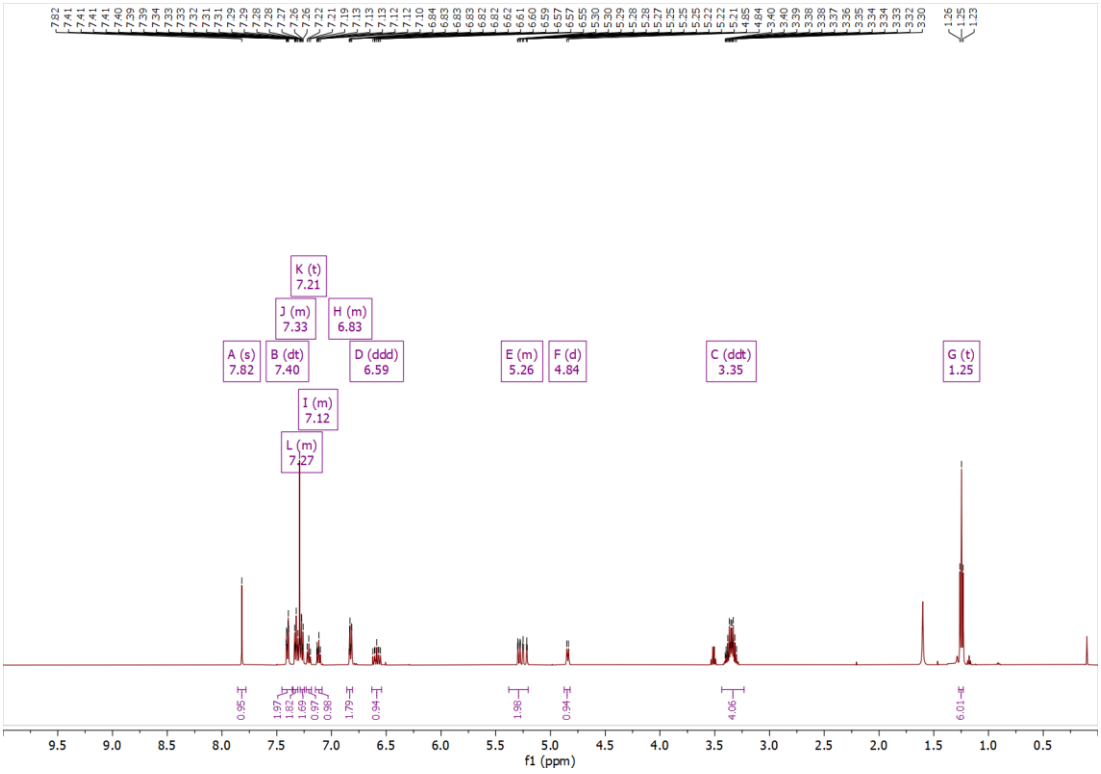
- ***R-VIII – Enantioselective aza-Claisen rearrangement***

The enantioselective *aza*-Claisen rearrangement is a previously unreported transformation mediated by BAOCs, furnishing the corresponding enamine in good stereoselectivity and perfect diastereoselectivity (up to 67% *e.e.*, >20:1 *d.r.*).



◊ *Experimental Protocol for R-**VIII***

To a 1-dram vial was charged 4Å molecular sieves, **chiral acid** (2.5 umol, 0.1 equiv.), **allyl amine** (125 uL, 0.2 M in hexanes, 25 umol, 1.0 equiv), and **propargyl ester** (125 uL, 0.2 M in hexanes, 25 umol, 1.0 equiv.). The reaction was left for 24 hours. After, the solution was filtered through a 1:1 plug of basic alumina and celite, eluting with diethyl ether. The filtered reaction mixture was concentrated to a crude residue (note – yield can be calculated by ^{1}H NMR with 0.014 M in $CDCl_3$ solution of 1,4-dioxane as internal standard; the product **enamine** characterization is reported below). The enantioenrichment was recorded by HPLC (ChiralPak IB, 98:2 Hex:iPrOH, flow rate = 0.5 mL/min, enantiomers elute = 19.9 min, 23.0 min).



^{1}H NMR (500 MHz, $CDCl_3$) δ 7.82 (s, 1H), 7.40 (dt, *J* = 8.3, 1.1 Hz, 2H), 7.35 – 7.31 (m, 2H), 7.29 – 7.25 (m, 2H), 7.21 (t, *J* = 7.3 Hz, 1H), 7.15 – 7.09 (m, 1H), 6.86 – 6.81 (m, 2H), 6.59 (ddd, *J* = 17.4, 10.0, 7.6 Hz, 1H), 5.38 – 5.20 (m, 2H), 4.84 (d, *J* = 7.6 Hz, 1H), 3.35 (ddt, *J* = 18.5, 14.4, 7.2 Hz, 4H), 1.25 (t, *J* = 7.2 Hz, 6H).

◊ *Experimental Dataset for R-**VIII***

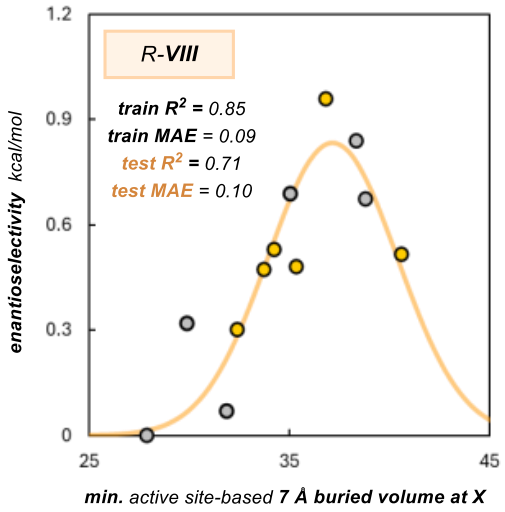
The complete dataset for R-**VIII** is printed below (**Supplementary Table 14**). In **Fig. 4E**, the model was fitted only to empirical data (containing substituents corresponding to the top-10 most popular – *bolded*). All data was experimentally collected in duplicate. The DDG values are computed from the avg. observed *e.e.*

CATALYST	EE	DDG
CPAH_97	**52.5**	**0.690747**
SPAH_97	**26.5**	**0.321532**
CPAH_934	**61**	**0.839628**
CPAH_116	11	0.13081
CPAH_27	**6**	**0.071148**
H8CPAH_934	**51.5**	**0.674514**
CPAH_41	19.5	0.233949
CPAH_7001	67	0.960223
CPAH_29	**0**	**0**
CPAH_321	25	0.302504
CPAH_9001	38.5	0.480757
CPAH_23	10	0.118835
CPAH_9010	38	0.47382
CPAH_9011	41	0.515926
sCPAH_9012	42	0.530234

Supplementary Table VI-32 | Experimental dataset for R-**VIII.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**VIII***

As described above, **Fig. 4D** presents the Gaussian regression performance for the subset of data containing the *bolded* BAOCs. Below, we report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that anticipates a *complementary* solution improving over the training data by +0.12 kcal/mol (pred. +0.06 kcal/mol).



Supplementary Fig. VI-179 | R-**VIII** from **Fig. 4D** (A = 0.835, μ = 37.2, σ = 3.24). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-I – Zhao, Q. et al. Angew. Chem. Int. Ed. 2024.***

  ◊ *Reported Dataset for R-**I***

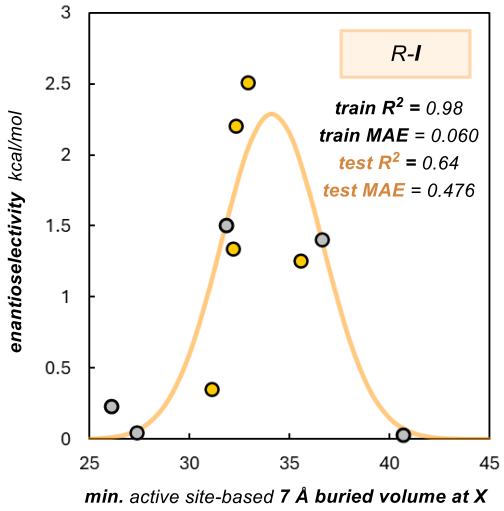
The complete dataset for R-**I** is printed below (**Supplementary Table 15**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-3 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
CPAH_46	**3**	**0.043294**
CPAH_27	**78**	**1.508166**
CPAH_122	91	2.203774
CPAH_9006	94	2.507501
NTPH_27	**75**	**1.403692**
NTPH_43	70	1.251263
SPAH_52	10	0.144755
SPAH_27	**16**	**0.232834**
CPAH_9007	73	1.339884
NTPH_16	24	0.353138
NTPH_97	**2**	**0.028858**

Supplementary Table VI-33 | Experimental dataset for R-**I.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**I***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that anticipates a *complementary* solution improving over the training data by +1.00 kcal/mol (pred. +0.52 kcal/mol).



Supplementary Fig. VI-180 | R-**I** from **Fig. 4C** (A = 2.29, μ = 34.1, σ = 2.51). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-II – Wu, S. et al. Nature 2025.***

  ◊ *Reported Dataset for R-**II***

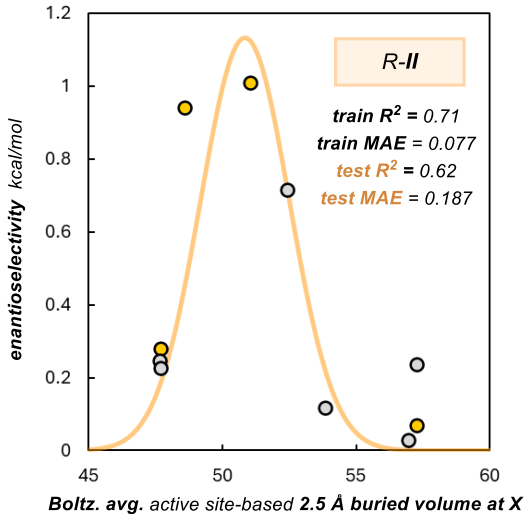
The complete dataset for R-**II** is printed below (**Supplementary Table 16**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-5 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
CPAH_5	**-27**	**0.26739**
CPAH_27	**-24**	**0.236398**
CPAH_46	**-12**	**0.116455**
CPAH_97	**-3**	**0.028982**
CPAH_934	-7	0.067715
SPAH_5	**23**	**0.226176**
SPAH_27	**23**	**0.226176**
SPAH_46	**25**	**0.246673**
SPAH_29	28	0.277838
SPAH_97	**63**	**0.716046**
SPAH_7001	75	0.939661
SPAH_934	78	1.009599

Supplementary Table VI-34 | Experimental dataset for R-**II.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**II***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that anticipates a *complementary* solution improving over the training data by +0.29 kcal/mol (pred. +0.42 kcal/mol).



Supplementary Fig. VI-181 | R-**II** (A = 1.13, μ = 50.8, σ = 1.66). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-X – Li, X. et al. Nat. Comm. 2021.***

  ◊ *Reported Dataset for R-**X***

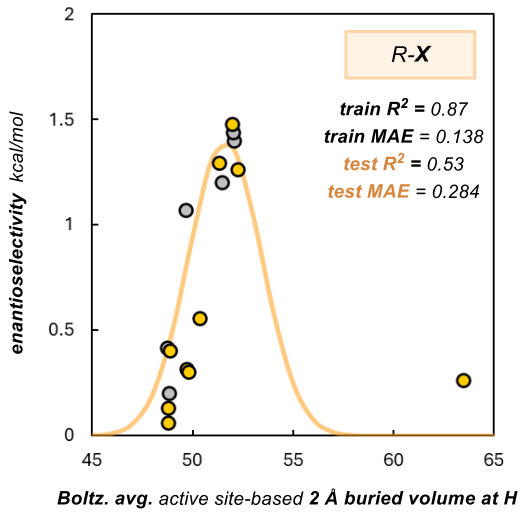
The complete dataset for R-**X** is printed below (**Supplementary Table 17**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-3 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
CPAH_97	**83**	**1.397753**
CPAH_7002	80	1.292434
CPAH_934	79	1.260459
CPAH_5	11	0.129932
CPAH_27	**34**	**0.416563**
CPAH_43	5	0.05887
H8CPAH_97	**84**	**1.436618**
H8CPAH_934	85	1.477769
H8CPAH_5	33	0.403312
H8CPAH_27	**17**	**0.201953**
H8CPAH_43	22	0.263115
SPAH_97	**77**	**1.200339**
SPAH_934	44	0.555544
SPAH_5	26	0.313056
SPAH_27	**72**	**1.067776**
SPAH_43	25	0.300474

Supplementary Table VI-35 | Experimental dataset for R-**X.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**X***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.11 kcal/mol.



Supplementary Fig. VI-182 | R-**X** (A = 1.40, μ = 51.6, σ = 1.83). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-XVIII – Shevchenko, G. Angew. Chem. Int. Ed. 2018.***

  ◊ *Reported Dataset for R-**XVIII***

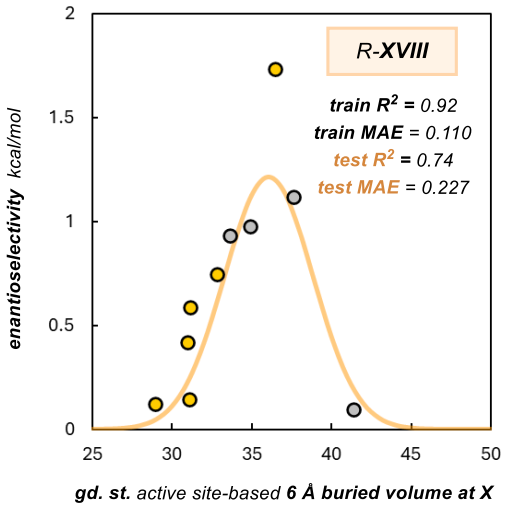
The complete dataset for R-**XVIII** is printed below (**Supplementary Table 18**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-5 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
CPAH_0	10	0.118037
CPAH_38	46	0.585049
CPAH_39	56	0.74448
CPAH_97	**8**	**0.094316**
CPAH_23	12	0.141854
CPAH_27	**74**	**1.118167**
CPAH_5	**66**	**0.932685**
CPAH_11	34	0.416563
CPAH_43	**68**	**0.97539**
CPAH_69	90	1.731955

Supplementary Table VI-36 | Experimental dataset for R-**XVIII.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**XVIII***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.61 kcal/mol (pred. +0.09 kcal/mol).



Supplementary Fig. VI-183 | R-**XVIII** (A = 1.21, μ = 36.1, σ = 2.78). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-XXV – Yu, H. et al. ACS Catal. 2013.***

  ◊ *Reported Dataset for R-**XXV***

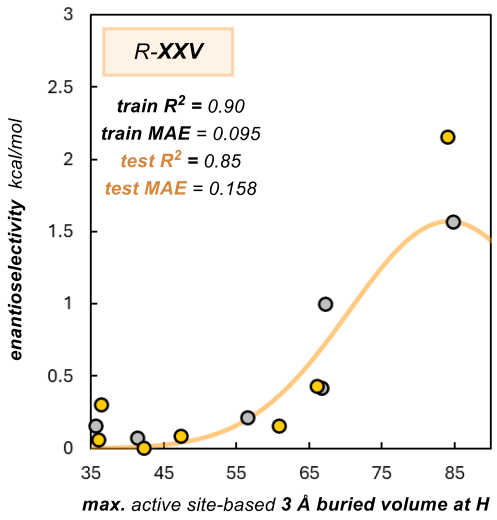
The complete dataset for R-**XXV** is printed below (**Supplementary Table 19**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-3 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
SPAH_43	-7	0.082485
CPAH_43	25	0.300474
CPAH_247	0	0
CPAH_453	5	0.05887
CPAH_27	**13**	**0.153806**
CPAH_46	**6**	**0.07067**
CPAH_97	**34**	**0.416563**
CPAH_934	35	0.429917
H8CPAH_97	**18**	**0.214089**
NTPH_27	**69**	**0.997556**
NTPH_97	**87**	**1.568268**
NTPH_43	13	0.153806
NTPH_934	95	2.154952

Supplementary Table VI-37 | Experimental dataset for R-**XXV.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**XXV***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.58 kcal/mol.



Supplementary Fig. VI-184 | R-**XXV** (A = 1.57, μ = 84.2, σ = 13.7). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-XXXIV – Yang, et al. J. Am. Chem. Soc. 2021.***

  ◊ *Reported Dataset for R-**XXXIV***

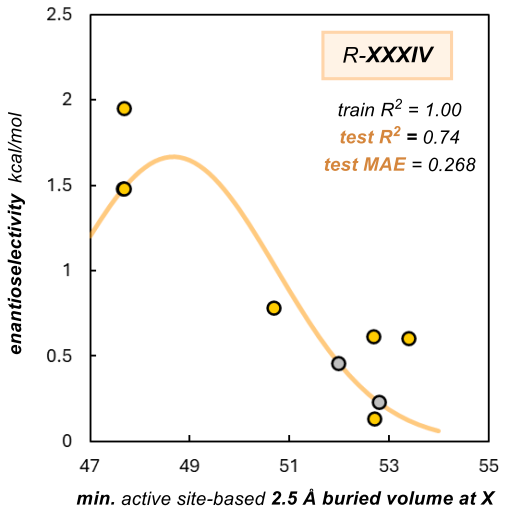
The complete dataset for R-**XXXIV** is printed below (**Supplementary Table 20**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-10 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
CPAH_11	**-37**	**0.45695**
CPAH_23	58	0.779337
CPAH_52	-47	0.600059
CPAH_88	-48	0.615252
CPAH_41	17	0.201953
CPAH_0	**-19**	**0.22627**
CPAH_8	11	0.129932
SPAH_1	85	1.477769
SPAH_11	**85**	**1.477769**
SPAH_23	93	1.950971

Supplementary Table VI-38 | Experimental dataset for R-**XXXIV.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**XXXIV***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.47 kcal/mol.



Supplementary Fig. VI-185 | R-**XXXIV** (A = 1.66, μ = 48.7, σ = 2.07). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-XXXV – Terada, M. et al. Angew. Chem. Int. Ed. 2008.***

  ◊ *Reported Dataset for R-**XXXV***

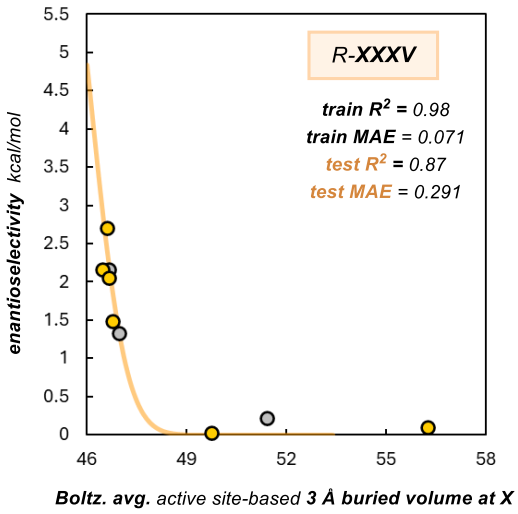
The complete dataset for R-**XXXV** is printed below (**Supplementary Table 21**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-10 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e*.

CATALYST	EE	DDG
CPAH_0	**95**	**2.154952**
CPAH_2	95	2.154952
CPAH_10	94	2.044684
CPAH_23	98	2.702906
CPAH_247	85	1.477769
CPAH_11	**81**	**1.325865**
CPAH_88	2	0.023532
CPAH_190	8	0.094316
CPAH_27	**18**	**0.214089**

Supplementary Table VI-39 | Experimental dataset for R-**XXXV.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

◊ *Gaussian Regression Models for R-**XXXV***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.55 kcal/mol (pred. +0.20 kcal/mol).



Supplementary Fig. VI-186 | R-**XXXV** (A = 7.01, μ = 45.1, σ = 1.01). Empirical training set is *gray*, while less popular testing set is *gold*.

- ***R-XLIII – Mo, N. et al. Org. Lett. 2012.***

  ◊ *Reported Dataset for R-**XLIII***

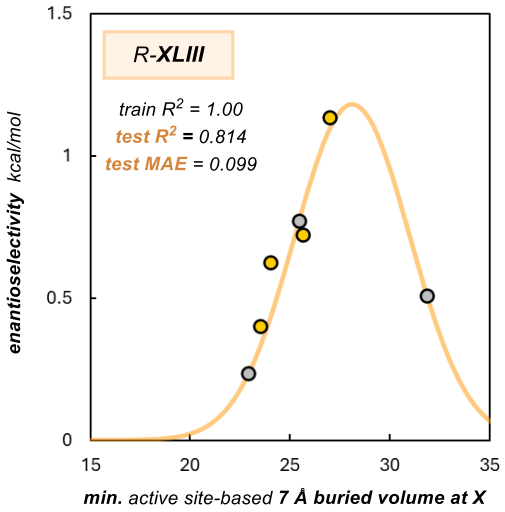
The complete dataset for R-**XLIII** is printed below (**Supplementary Table 23**). Here, the model was fitted only to empirical data (containing substituents corresponding to the top-10 most popular – *bolded*). All data was recorded as reported. The DDG values are computed from the rep. *e.e.*

CATALYST	EE	DDG
H8CPAH_0	**20**	**0.236082**
H8CPAH_8	49	0.624243
H8CPAH_2	33	0.399224
H8CPAH_11	**58**	**0.771439**
H8CPAH_27	**41**	**0.50727**
H8CPAH_65	75	1.133007
CPAH_65	55	0.720106

Supplementary Table VI-40 | Experimental dataset for R-**XLIII.** Empirically popular BAOCs are *bolded*, while less popular BAOCs are *normal*.

- ***Gaussian Regression Models for R-XLIII***

The Gaussian regression performance below is trained only on the subset of data containing the indicated *bolded* BAOCs above. We report R^2, MAE, numerical fit, and specific active site-based buried volume for this fit. Then, we overlay the remaining data from less popular BAOCs, demonstrating the empirical data trains a predictive model that helps anticipate a *complementary* solution improving over the training data by +0.36 kcal/mol (pred. +0.33 kcal/mol).



Supplementary Fig. VI-187 | R-**XLIII** (A = 1.18, μ = 28.1, σ = 2.88). Empirical training set is *gray*, while less popular testing set is *gold*.

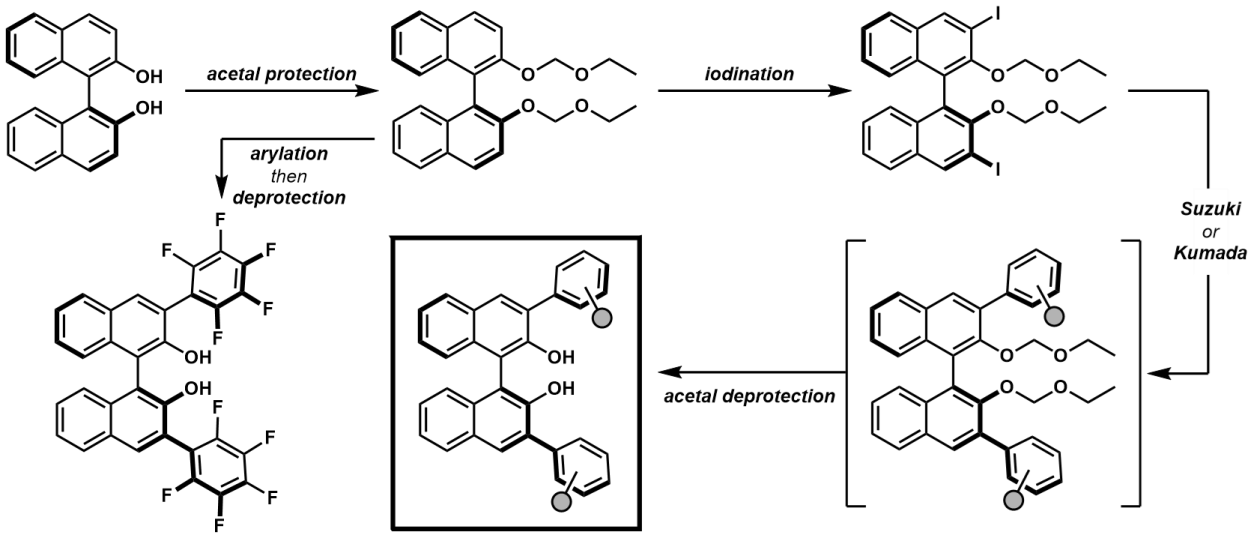
VI.38 – Synthetic Protocols

- ***Overview of BAOC Syntheses***

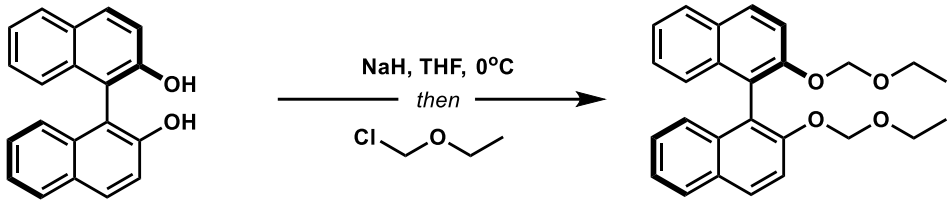
The experimental data reported herein was collected using Brønsted acid organocatalysts (BAOCs) prepared in house. Reported BAOCs (chiral phosphoric acids, CPA; *N*-triflylphosphoramide, NTP) were prepared using well-known routes documented throughout the literature (see below, **Supplementary Fig. 71**). The structurally unprecedented BAOC reported in this work, denoted P2-SIA, was prepared by the phosphorylation of sulfonimidamides (SIA). The SIA substrates were prepared based on protocols established by Willis and coworkers[61] and updated by van Dijk, L. et al.[62] (ref. **Supplementary Fig. 72)**.

- ***Synthesis of 3,3'-Substituted BINOLs***

3,3'-substituted BINOL derivatives were generally prepared by either (a) Kumada coupling between iodinated derivative and the corresponding Grignard reagent or (b) Suzuki coupling between iodinated derivative and the corresponding boronic acid.



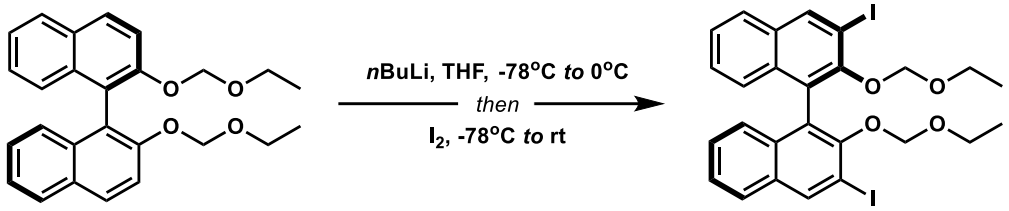
Supplementary Fig. VI-188. Access to 3,3'-substituted BINOLs utilized in this work.



(*R*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene, S-1: To a flame-dried 500 mL round-bottom flask equipped with a large Teflon coated stir bar was added NaH, 60% dispersion in mineral oil (4.30 g, 108 mmol, 2.15 equiv). The flask was equipped with a rubber septum, evacuated and purged with nitrogen. To the flask was added 150 mL of THF (0.33 M). The flask was then placed into an ice bath and to the solution was slowly added 1,1′-binaphthyl-2,2′-diol (14.3 g, 50.0 mmol, 1 equiv) portion wise. The flask was removed from the ice bath and warmed to room temperature stirring for 30 minutes. Then, the suspension was charged with chloromethyl ethyl ether (10.7 mL, 115 mmol, 2.3 equiv) and stirred for an additional 60 minutes at room temperature.

After, the reaction was slowly quenched with 100 mL of sat. $NaHCO_3$ (aq) and stirred for 10-15 minutes to consume unreacted chloromethyl ethyl ether. The suspension was transferred to a separatory funnel and extracted with 3x50 mL ethyl acetate. The organic fractions were pooled, washed with 2x100 mL sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish the **title compound** as a thick, clear oil (19.9 g, 49.9 mmol, 99%) that crystallized under vacuum. The material was used without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.95 (d, *J* = 8.3 Hz, 2H), 7.87 (d, *J* = 8.2 Hz, 2H), 7.60 (d, *J* = 9.0 Hz, 2H), 7.34 (ddd, *J* = 8.1, 6.7, 1.3 Hz, 2H), 7.21 (ddd, *J* = 8.1, 6.6, 1.3 Hz, 2H), 5.13 (d, *J* = 6.9 Hz, 2H) 3.43-3.29 (m, 4H), 1.00 (t, *J* = 7.1 Hz, 6H).



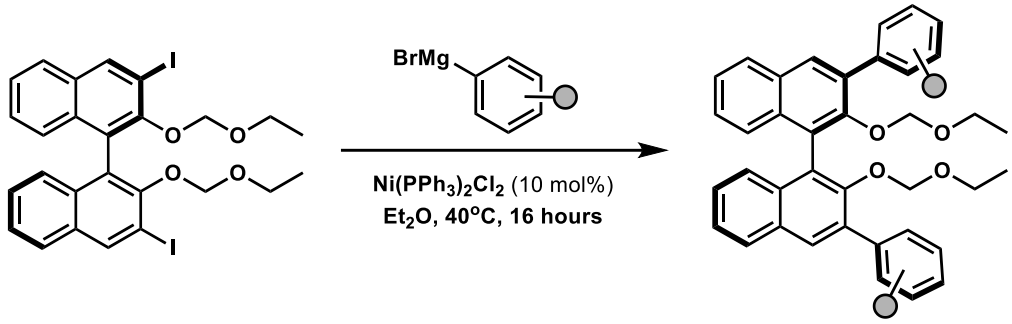
(*R*)-3,3'-diiodo-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene, S-2: To a flame dried 500 mL round bottom flask equipped with a Teflon coated stir bar was added 2,2'-bis(ethoxymethoxy)-1,1'binaphthalene (5.00 g, 12.4 mmol, 1.0 equiv) and THF (100 mL). The solution was placed in a dry ice/acetone bath and cooled to -78°C. To the solution was added *n*BuLi (14.9 mL, 2.5 M in hexanes, 37.3 mmol, 3.0 equiv) dropwise. The solution was stirred 30 minutes at -78°C and then warmed to 0°C for 2 hours. After, the solution was cooled back to -78°C and iodine (9.46 g, 37.3 mmol, 3.0 equiv) was added in one portion. The solution was removed from the cooling bath and allowed to warm to room temperature to stir for at least an additional 4 hours.

The solution was then carefully quenched by the addition of sat. NaS_2O_3 (aq) to scavenge excess iodine. The suspension was transferred to a separatory funnel and extracted with 3x50 mL of ethyl acetate. The organic fractions were pooled, washed with 2x50 mL of sat. NaCl (aq), dried over Na_2SO_4 and concentrated *in vacuo* to furnish a crude oil containing the **title compound**.

The crude oil was subjected to column chromatography (0-100% EtOAc/Hex) to furnish the **title compound** as a clear oil with mono-iodinated contamination. The oil was sonicated in a suspension of 100% hexanes to furnish a fine white suspension. The suspension was filtered over a fritted flask, and the solid was collected and dried *in vacuo* to furnish the **title compound** (4.35 g, 6.70 mmol, 54%).

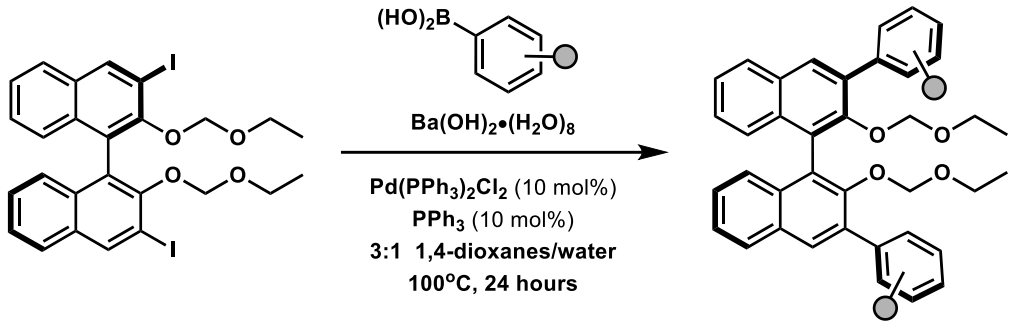
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.54 (s, 2H), 7.84 – 7.73 (m, 2H), 7.42 (ddd, *J* = 8.1, 6.9, 1.2 Hz, 2H), 7.29 (ddd, *J* = 8.2, 6.7, 1.3 Hz, 2H), 7.15 (dd, *J* = 8.5, 1.1 Hz, 2H), 4.88 (d, *J* = 5.7 Hz, 2H), 4.71 (d, *J* = 5.7 Hz, 2H), 3.07 (dq, *J* = 9.4, 7.0 Hz, 2H), 2.69 (dq, *J* = 9.3, 7.0 Hz, 2H), 0.66 (t, *J* = 7.0 Hz, 6H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 152.26, 140.01, 133.80, 132.23, 127.20, 126.85, 126.48, 126.08, 125.81, 98.13, 92.67, 65.04, 14.40.



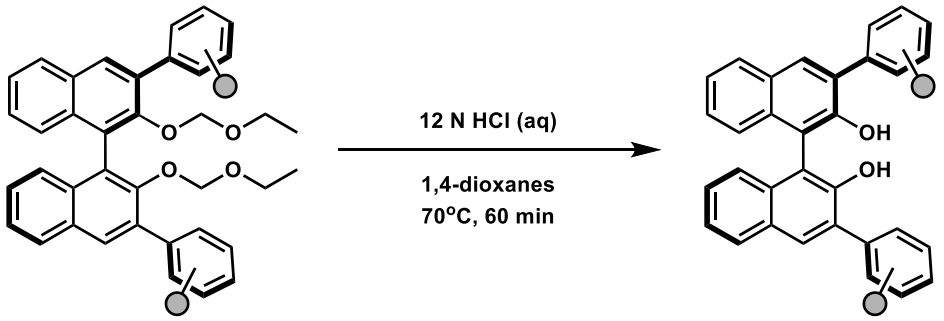
General protocol A – Kumada coupling to furnish 3,3'-substitution: To a flame dried 100 mL round bottom flask equipped with a Teflon coated stir bar was added 2,2'-bis(ethoxymethoxy)-3,3'-diiodo-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv) and bistriphenylphosphine nickel (II) chloride (100 mg, 153 umol, 0.10 equiv). The flask was equipped with a rubber septum, evacuated and purged with nitrogen. To the purged flask was added diethyl ether (15 mL, 0.1 M) to furnish a dark gray suspension. To the suspension was slowly added a **Grignard solution** (7.71 mmol, 6.0 equiv) (**note** – exothermic, solution turns a clear, red-brown solution). The suspension was lowered into an oil bath and heated at 40°C overnight.

After, the solution was cooled to room temperature and quenched with 25 mL of sat. NH_4Cl (aq). The suspension was transferred to a separatory funnel and extracted with 3x20 mL of DCM. The organic fractions were pooled, washed with 50 mL of sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish a yellow, crude oil containing the **title compound**. The crude oil was subjected to column chromatography (0-100% EtOAc/Hex) to furnish the **title compound** as a sticky, white solid.



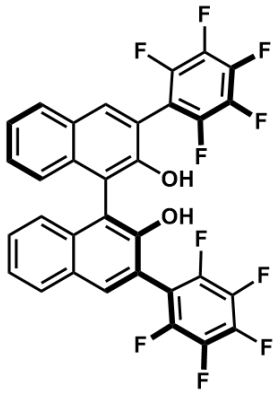
General protocol B – Suzuki coupling to furnish 3,3'-substitution: To a 250 mL round bottom flask was added a Teflon coated stir bar, 2,2'-bis(ethoxymethoxy)-3,3'-diiodo-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv), **boronic acid** (6.11 mmol, 4.0 equiv), barium hydroxide octahydrate (2.41 g, 7.64 mmol, 5.0 equiv), bistriphenylphosphine palladium (II) chloride (107 mg, 153 umol, 0.1 equiv), and triphenylphosphine (40 mg, 153 umol, 0.1 equiv). The flask was evacuated and purged under a nitrogen atmosphere. In a separate container was prepared a 3:1 solution of 1,4-dioxane and water. The solution was sparged with nitrogen for at least 30 minutes, then 30 mL of the solution was transferred to the round bottom flask containing the solids. The suspension was lowered into an oil bath and heated at 100°C for at least 16 hours.

After, the solution was cooled to room temperature and quenched with 100 mL of sat. NH_4Cl (aq). The suspension was transferred to a separatory funnel and extracted with 3x50 mL of DCM. The organic fractions were pooled, washed with 100 mL of sat. NaCl (aq), filtered over a plug of silica, dried over Na_2SO_4, and concentrated *in vacuo* to furnish a pale orange crude oil containing the **title compound**. The crude oil was subjected to column chromatography (0-100% EtOAc/Hex) to furnish the **title compound** as a pale, yellow oil.



General protocol C – acetal deprotection to furnish 3,3'-substituted BINOL: To a 100 mL round bottom flask equipped with a Teflon coated stir bar was added 3,3'-**arylated**-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (1.0 equiv), 1,4-dioxane (0.1 M), and conc. HCl (12 N, 1.0 M). The flask was sealed with a rubber septa and the solution was heated to 70°C for at least 1 hour. After, the solution was quenched over a mixture of ice and sat. $NaHCO_3$ (aq) solution. Once

the suspension no longer evolved carbon dioxide, the suspension was transferred to a separatory funnel and extracted with 3x20 mL ethyl acetate. The organic fractions were pooled, washed with 20 mL sat. NaCl (aq), dried over Na_2SO_4 and concentrated *in vacuo* to furnish a crude solid containing the **title compound**. The crude solid was subjected column chromatography (0-50% EtOAc/Hex) to furnish the **title compound**.

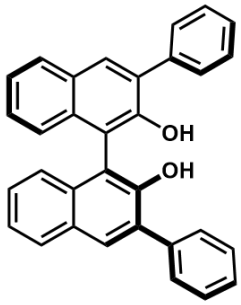


(*S*)-3,3'-bis(perfluorophenyl)-[1,1'-binaphthalene]-2,2'-diol, S-3: To a flame-dried 25 mL round bottom flask equipped with a Teflon coated stirbar was added (*S*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (400 mg, 1.0 mmol, 1.0 equiv). The flask was evacuated with nitrogen and charged with THF (8 mL). The solution was cooled to -78ºC and then charged with 2.5 M n-BuLi in hexanes (1.2 mL, 3.0 mmol, 3.0 equiv), warmed to 0ºC, and stirred for 1 hr. The solution was cooled back to -78ºC, charged with hexafluorobenzene (0.8 mL, 7.0 mmol, 7.0 equiv), and allowed to warm to room temperature stirring overnight.

The solution was then carefully quenched by the addition of sat. NH4Cl. The mixture was transferred to a separatory funnel and diluted with 50 mL of EtOAc. The organic layer was separated and the aqueous layer was extracted with 3x25 mL of EtOAc. The pooled organic fractions were then washed with 2x50 mL of brine solution, dried over Na2SO4 and concentrated in vacuo. The crude residue was subjected to column chromatography (0-50% EtOAc/Hex) to furnish the title compound (350 mg, 0.48 mmol, 48%) as a white solid. Spectral characterization is in accordance with literature reports.[63]

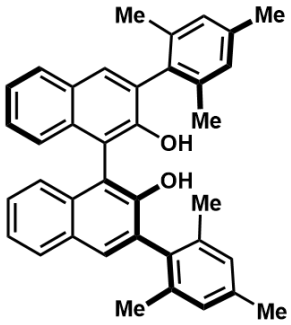
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.96 (s, 2H), 7.93 – 7.84 (m, 2H), 7.38 (dqd, J = 8.3, 6.9, 1.4 Hz, 4H), 7.17 (d, J = 1.4 Hz, 1H), 5.23 (s, 2H).

^{19}F NMR (471 MHz, $CDCl_3$) δ -139.81 (ddd, J = 146.2, 22.9, 7.9 Hz), -154.58 (t, J = 20.9 Hz), -162.41 (dtd, J = 59.0, 22.1, 8.2 Hz).



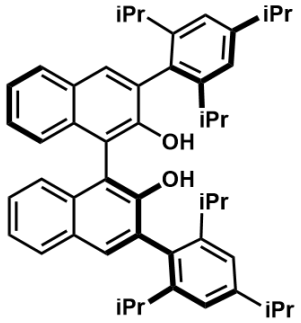
(*S*)-3,3'-diphenyl-[1,1'-binaphthalene]-2,2'-diol, S-4: According to **General Procedure A** followed by **General Procedure C** with (S)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv) and a commercial solution of phenyl magnesium bromide. Isolated the **title compound** as a white solid (62 mg, 0.14 mmol, 9%). Spectral characterization is in accordance with literature reports.[64]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.08 (s, 2H), 8.01 – 7.93 (m, 2H), 7.80 (dd, *J* = 8.3, 1.3 Hz, 4H), 7.54 (t, *J* = 7.7 Hz, 4H), 7.49 – 7.39 (m, 4H), 7.36 (ddd, *J* = 8.2, 6.7, 1.4 Hz, 2H), 7.33 – 7.28 (m, 2H), 5.45 (s, 2H).



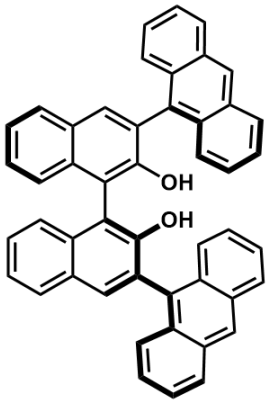
(*S*)-3,3'-dimesityl-[1,1'-binaphthalene]-2,2'-diol, S-5 According to **General Procedure A** followed by **General Procedure C** with (S)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (620 mg, 0.97 mmol, 1.0 equiv) and a commercial solution of 2,4,6-trimethylphenyl magnesium bromide. Isolated the **title compound** as a white solid (29 mg, 0.06 mmol, 6%). Spectral characterization is in accordance with literature reports.[64]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87 (d, J = 8.0 Hz, 2H), 7.74 (s, 2H), 7.38 (ddd, J = 8.1, 6.8, 1.3 Hz, 2H), 7.31 (ddd, J = 8.2, 6.8, 1.4 Hz, 2H), 7.23 (d, J = 1.3 Hz, 2H), 7.01 (s, 4H), 4.99 (s, 2H), 2.34 (s, 6H), 2.14 (s, 6H), 2.07 (s, 6H).



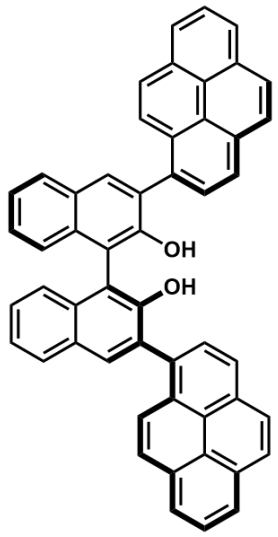
(*S*)-3,3'-bis(2,4,6-triisopropylphenyl)-[1,1'-binaphthalene]-2,2'-diol, S-6: According to **General Procedure A** followed by **General Procedure C** with (*S*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (327 mg, 0.50 mmol, 1.0 equiv) and a commercial solution of 2,4,6-triisopropylphenyl magnesium bromide. Isolated the **title compound** as a white solid (94 mg, 0.14 mmol, 27%). Spectral characterization is in accordance with literature reports.[65]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.91 (d, *J* = 8.0 Hz, 1H), 7.81 (s, 1H), 7.46 – 7.33 (m, 3H), 7.17 (d, *J* = 6.9 Hz, 2H), 4.96 (s, 1H), 3.00 (p, *J* = 6.9 Hz, 1H), 2.89 (p, *J* = 6.9 Hz, 1H), 2.73 (p, *J* = 6.8 Hz, 1H), 1.36 (d, *J* = 6.9 Hz, 6H), 1.24 (d, *J* = 6.8 Hz, 3H), 1.14 (dd, *J* = 9.0, 6.9 Hz, 6H), 1.07 (d, *J* = 6.9 Hz, 3H).



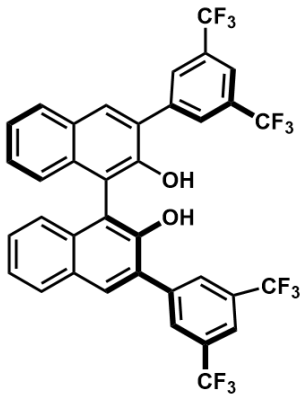
(*S*)-3,3'-di(anthracene-9-yl)-[1,1'-binaphthalene]-2,2'-diol, S-8 According to **General Procedure B** followed by **General Procedure C** with (*S*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (327 mg, 0.50 mmol, 1.0 equiv) and 9-anthracenyl boronic acid. Isolated the **title compound** as a white solid (26 mg, 0.04 mmol, 8%). Spectral characterization is in accordance with literature reports.[66]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.59 (s, 1H), 8.12 – 8.04 (m, 3H), 8.00 – 7.86 (m, 5H), 7.84 – 7.75 (m, 2H), 7.53 – 7.34 (m, 14H), 7.31 (ddt, *J* = 8.3, 6.8, 1.5 Hz, 1H), 7.15 (dd, *J* = 8.4, 1.2 Hz, 1H).



(*S*)-3,3'-di(pyren-9-yl)-[1,1'-binaphthalene]-2,2'-diol, S-9: According to **General Procedure B** followed by **General Procedure C** with (S)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv) and 1-pyrenyl boronic acid. Isolated the **title compound** as a white solid (504 mg, 0.63 mmol, 41%). Spectral characterization is in accordance with literature reports.[67]

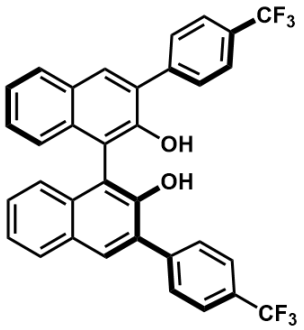
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.32 – 7.93 (m, 22H), 7.62 – 7.42 (m, 6H), 5.40 – 5.20 (m, 2H). (note – rotamers lead to complex array of peaks).



(*S*)-3,3'-bis(3,5-bis(trifluoromethyl)phenyl)-[1,1'-binaphthalene]-2,2'-diol, S-10: According to **General Procedure B** followed by **General Procedure C** with (*S*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv) and 3,5-bistrifluoromethylphenyl boronic acid. Isolated the **title compound** as a white solid (284 mg, 0.40 mmol, 26%). Spectral characterization is in accordance with literature reports.[68]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.23 (s, 4H), 8.11 (s, 2H), 8.00 (d, J = 8.1 Hz, 2H), 7.91 (s, 2H), 7.45 (dt, J = 24.0, 7.1 Hz, 4H), 7.23 (d, J = 8.4 Hz, 2H), 3.70 (s, 2H).

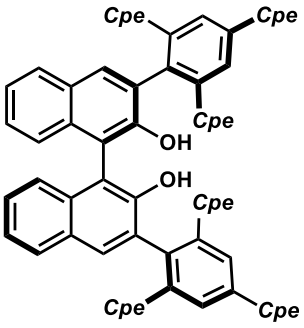
^{19}F NMR (376 MHz, $CDCl_3$) δ -62.70.



(*S*)-3,3'-bis(4-trifluoromethylphenyl)-[1,1'-binaphthalene]-2,2'-diol, S-11: According to **General Procedure B** followed by **General Procedure C** with (*S*)-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (1.00 g, 1.53 mmol, 1.0 equiv) and 4-trifluoromethylphenyl boronic acid. Isolated the **title compound** as a white solid (285 mg, 0.50 mmol, 33%). Spectral characterization is in accordance with literature reports.[68]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 (s, 2H), 8.01 – 7.92 (m, 2H), 7.87 (d, *J* = 8.1 Hz, 4H), 7.74 (d, *J* = 8.1 Hz, 4H), 7.44 (ddd, *J* = 8.1, 6.8, 1.3 Hz, 2H), 7.37 (ddd, *J* = 8.2, 6.8, 1.4 Hz, 2H), 7.23 (dd, *J* = 8.4, 1.2 Hz, 2H), 5.32 (s, 2H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -62.39.



(*R*)-3,3'-bis(2,4,6-tricyclopentylphenyl)-[1,1'-binaphthalene]-2,2'-diol, S-12: According to **General Procedure A** followed by **General Procedure C** with *(R)*- 2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (513 mg, 784 umol, 1.0 equiv) and 2-bromo-1,3,5-tricyclopentylbenzene, which was prepared according to literature protocol from benzene, bromocyclopentane, and bromine in two steps.[69] Isolated the **title compound** as a white solid (146 mg, 172 umol, 22%). Spectral characterization is in accordance with literature reports.[70]

For 1,3,5-tricyclopentylbenzene:

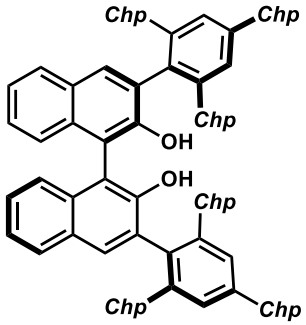
^{1}H NMR (300 MHz, $CDCl_3$) δ 7.00 (s, 2H), 3.51 (tt, *J* = 9.3, 7.4 Hz, 2H), 2.92 (m, 1H), 2.08 (m, 6H), 1.88 – 1.65 (m, 13H), 1.59 – 1.54 (m, 5H).

For 2-bromo-1,3,5-tricyclopentylbenzene:

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.00 (s, 2H), 3.51 (tt, *J* = 9.3, 7.4 Hz, 2H), 2.92 (m, 1H), 2.08 (m, 6H), 1.88 – 1.65 (m, 13H), 1.59 – 1.54 (m, 5H).

*For **title compound:***

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.87 (d, *J* = 8.0 Hz, 2H), 7.76 (s, 2H), 7.37 (ddd, *J* = 8.1, 6.8, 1.3 Hz, 2H), 7.29 (ddd, *J* = 8.2, 6.7, 1.3 Hz, 2H), 7.24 – 7.21 (m, 2H), 7.15 (q, *J* = 1.7 Hz, 4H), 4.89 (s, 2H), 3.03 (h, *J* = 7.8 Hz, 2H), 2.84 (p, *J* = 8.8 Hz, 2H), 2.80 – 2.72 (m, 2H), 2.16 – 2.05 (m, 4H), 1.93 – 1.60 (m, 32H), 1.54 – 1.32 (m, 11H).



(*R*)-3,3'-bis(2,4,6-tricycloheptylphenyl)-[1,1'-binaphthalene]-2,2'-diol, S-13: According to **General Procedure A** followed by **General Procedure C** with *(R)*- 2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene (507 mg, 775 ummol, 1.0 equiv) and 2-bromo-1,3,5-tricycloheptylbenzene, which was prepared according to literature protocol from benzene, bromocycloheptane, and bromine in two steps.[69] Isolated the **title compound** as a white solid (343 mg, 338 mmol, 44%). Spectral characterization is in accordance with literature reports.[70]

1,3,5-tricycloheptylbenzene:

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.80 (s, 3H), 2.74 – 2.50 (m, 3H), 2.06 – 1.48 (m, 40H).

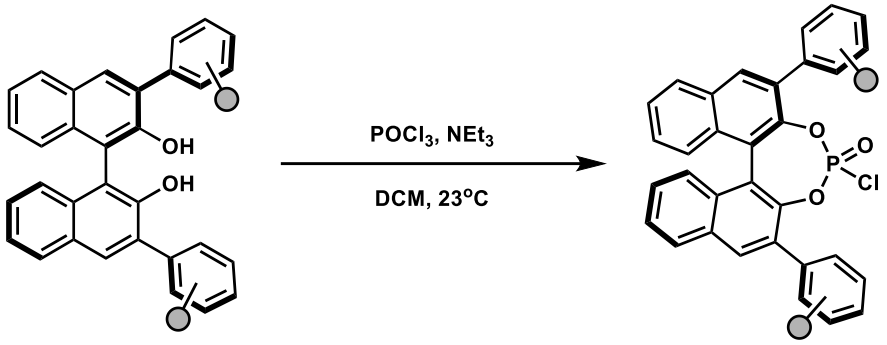
For 2-bromo-1,3,5-tricycloheptylbenzene:

^{1}H NMR (500 MHz, $CDCl_3$) δ 6.87 (s, 2H), 3.23 (tt, *J* = 6.7, 3.5 Hz, 2H), 2.58 (tq, *J* = 10.5, 3.2 Hz, 1H), 2.06 – 1.45 (m, 42H).

*For **title compound:***

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (d, *J* = 8.0 Hz, 2H), 7.70 (s, 2H), 7.35 (dt, *J* = 8.1, 4.0 Hz, 2H), 7.28 (d, *J* = 3.2 Hz, 4H), 7.01 (d, *J* = 8.6 Hz, 4H), 4.84 (s, 2H), 2.68 (tt, *J* = 10.6, 3.6 Hz, 2H), 2.59 – 2.51 (m, 2H), 2.50 – 2.35 (m, 2H), 2.03 – 1.93 (m, 4H), 1.93 – 1.75 (m, 10H), 1.73 – 1.49 (m, 56H), 1.39 – 1.32 (m, 8H), 1.24 (d, *J* = 9.7 Hz, 2H), 1.21 – 1.00 (m, 4H).

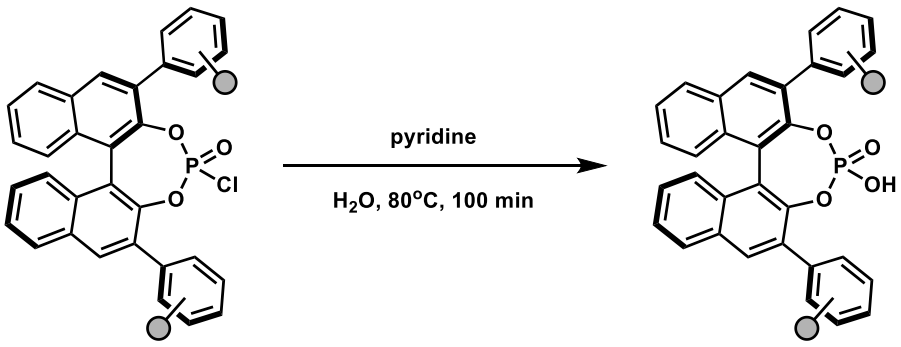
- ***Synthesis of Phosphoryl Chlorides***



General protocol D – phosphorylation of 3,3'-substiuted BINOL: To a flame dried 20 mL reaction tube equipped with a Teflon coated stir bar was added 3,3'-**aryl**-[1,1'-binaphthalene]-2,2'-diol (1.0 equiv) and dichloromethane (0.2 M). The reaction tube was equipped with a Teflon septum and charged with phosphorous oxychloride (2.0 equiv, 1.65 g/mL) and triethylamine (3.0 equiv, 0.73 g/mL) (note – triethylamine addition is exothermic). The solution was then stirred overnight at room temperature. After, the solution was transferred to a separatory funnel, quenched with sat. $NaHCO_3$ (aq), and extracted with 3x20 mL DCM. The organic fractions were pooled, washed with sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish a foamy white solid containing the **title compound**.

If necessary, the isolated **phosphoryl chloride** can be purified by column chromatography (0-100% EtOAc/Hex) to furnish the **title compound** as a white solid.

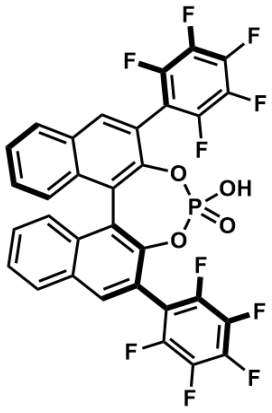
- ***Synthesis of Chiral Phosphoric Acids, CPA***



General protocol E – hydrolysis of phosphoryl chloride to furnish chiral phosphoric acid, CPA: To a flame dried 20 mL reaction tube equipped with a Teflon coated stir bar was added 3,3'-**arylated**-[1,1'-binaphthalene]-2,2'-diyl phosphoryl chloride (1.0 equiv), pyridine (1.0 M), and water (1.0 M). The reaction tube was sealed and heated to at least 80°C for over 100 minutes. After, the solution was transferred to a separatory funnel, quenched with 20 mL of 1 N HCl (aq), and

extracted with 3x20 mL DCM. The organic fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo* to furnish a foamy white solid containing the **title compound**.

The crude white solid was subjected to column chromatography (Yamazen 7g, 25 mL/min, 0-10% MeOH/DCM) to furnish a mixed salt of the **title compound**. The isolated mixed phosphate was resuspended in 3-5 mL of dichloromethane and 3-5 mL of 3 N HCl (aq). The suspension was vortexed for 10 seconds and allowed to completely settle for 5 minutes. The dichloromethane layer was separated by pipette and concentrated to furnish a foamy, white solid containing pure, **title phosphoric acid**.

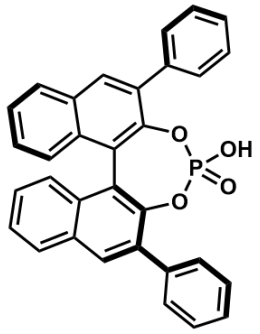


***(S)*-4-hydroxy-2,6-bis(perfluorophenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, ArF-CPA:** According to **General Procedure D** followed by **General Procedure E** from (*S*)-3,3'-bis(perfluorophenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-3** (62 mg, 100 umol). Spectral characterization is in accordance with literature reports.[63]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.08 – 7.96 (m, 4H), 7.61 (ddd, J = 8.1, 6.5, 1.3 Hz, 2H), 7.56 – 7.42 (m, 4H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -138.64 (dd, J = 22.8, 6.7 Hz), -140.95 (dd, J = 22.8, 8.2 Hz), 154.33 (t, J = 20.6 Hz), -161.89 (td, J = 21.6, 7.9 Hz), -162.89 (td, J = 21.7, 8.3 Hz).

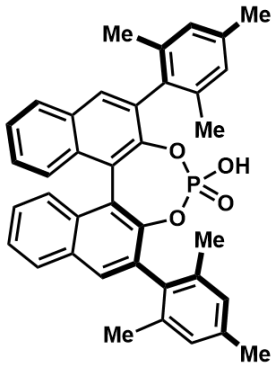
^{31}P NMR (162 MHz, $CDCl_3$) δ 4.78.



***(S)*-4-hydroxy-2,6-diphenyldinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, Ph-CPA:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-diphenyl-[1,1'-binaphthalene]-2,2'-diol, **S-4** (44 mg, 100 umol). Spectral characterization is in accordance with literature reports.[64]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.99 (s, 2H), 7.95 (d, *J* = 8.2 Hz, 2H), 7.65 – 7.53 (m, 4H), 7.49 (ddd, *J* = 8.2, 6.6, 1.3 Hz, 2H), 7.37 (d, *J* = 8.5 Hz, 2H), 7.34 – 7.29 (m, 2H), 7.26 (d, *J* = 6.5 Hz, 2H), 7.17 (t, *J* = 7.3 Hz, 2H).

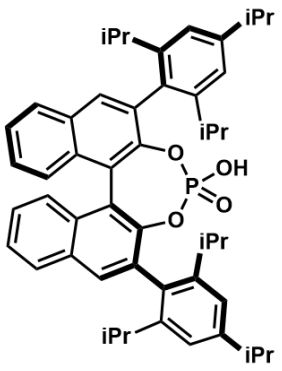
^{31}P NMR (162 MHz, $CDCl_3$) δ 3.17.



***(S)*-4-hydroxy-2,6-bis(2,4,6-trimethylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, Mesit-CPA:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-bis(2,4,6-trimethylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-5** (53 mg, 100 umol). Spectral characterization is in accordance with literature reports.[64]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.92 (d, *J* = 8.2 Hz, 2H), 7.78 (s, 2H), 7.49 (t, *J* = 7.4 Hz, 2H), 7.43 – 7.28 (m, 4H), 6.78 (d, *J* = 15.1 Hz, 4H), 2.13 (s, 6H), 2.05 (s, 6H), 1.96 (s, 6H).

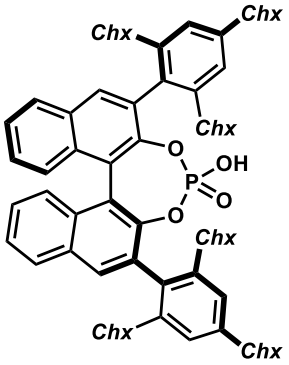
^{31}P NMR (162 MHz, $CDCl_3$) δ 3.79.



***(S)*-4-hydroxy-2,6-bis(2,4,6-triisopropylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, TRIP-CPA:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-bis(2,4,6-triisopropylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-6** (69 mg, 100 umol). Spectral characterization is in accordance with literature reports.[65]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87 (d, *J* = 8.2 Hz, 2H), 7.79 (s, 2H), 7.49 (ddd, *J* = 8.1, 5.9, 2.1 Hz, 2H), 7.31 (dd, *J* = 6.0, 1.4 Hz, 4H), 6.92 (dd, *J* = 14.3, 1.8 Hz, 4H), 2.82 (hept, *J* = 6.9 Hz, 2H), 2.59 (q, *J* = 6.8 Hz, 2H), 2.50 (dq, *J* = 13.5, 6.8 Hz, 2H), 1.21 (d, *J* = 6.9 Hz, 12H), 0.98 (d, *J* = 6.7 Hz, 12H), 0.89 (d, *J* = 6.8 Hz, 6H), 0.65 (d, *J* = 6.7 Hz, 6H).

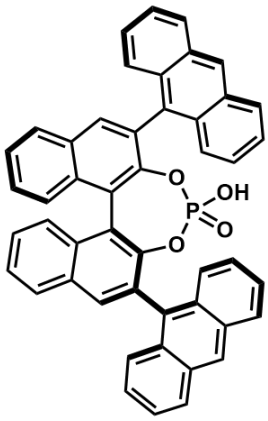
^{31}P NMR (162 MHz, $CDCl_3$) δ 3.1.



***(S)*-4-hydroxy-2,6-bis(2,4,6-tricyclohexylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, TCYP-CPA:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-bis(2,4,6-tricyclohexylphenyl)-[1,1'-binaphthalene]-2,2'-diol (93 mg, 100 umol). Spectral characterization is in accordance with literature reports.[71]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.84 (d, *J* = 8.2 Hz, 2H), 7.71 (s, 2H), 7.43 (ddd, *J* = 8.1, 6.4, 1.5 Hz, 2H), 7.24 (d, *J* = 6.7 Hz, 4H), 6.96 – 6.83 (m, 4H), 2.42 (d, *J* = 11.9 Hz, 2H), 2.12 (d, *J* = 12.1 Hz, 4H), 1.96 – 1.79 (m), 1.79 – 1.34 (m), 1.35 – 1.10 (m), 1.10 – 0.46 (m). **note** – alkyl integrations are noisy due to large number of peaks.

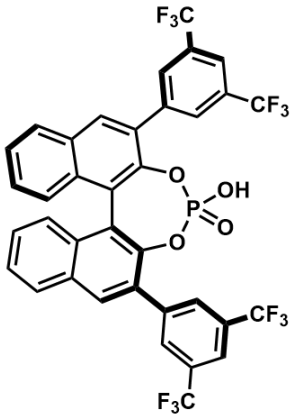
^{31}P NMR (162 MHz, $CDCl_3$) δ 2.09.



***(S)*-4-hydroxy-2,6-di(9-anthracenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, Anth-CPA:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-di(9-anthracenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-8** (64 mg, 100 umol). Spectral characterization is in accordance with literature reports.[66]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.15 (s, 2H), 7.98 (d, *J* = 8.5 Hz, 4H), 7.85 (d, *J* = 8.5 Hz, 2H), 7.69 (d, *J* = 8.5 Hz, 2H), 7.66 – 7.55 (m, 6H), 7.51 (dtd, *J* = 8.4, 3.5, 1.4 Hz, 4H), 7.34 (ddd, *J* = 8.2, 6.6, 1.2 Hz, 2H), 7.22 (ddd, *J* = 8.8, 6.6, 1.3 Hz, 2H), 7.15 – 7.04 (m, 4H).

^{31}P NMR (162 MHz, $CDCl_3$) δ 1.07.

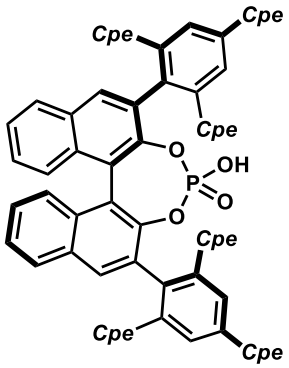


***(S)*-4-hydroxy-2,6-bis(3,5-trifluoromethylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine, CPA-S14:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-bis(3,5-bis(trifluoromethyl)phenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-10** (71 mg, 100 umol). Spectral characterization is in accordance with literature reports.[68]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.01 (d, *J* = 9.6 Hz, 8H), 7.58 (d, *J* = 4.5 Hz, 4H), 7.50 – 7.32 (m, 4H).

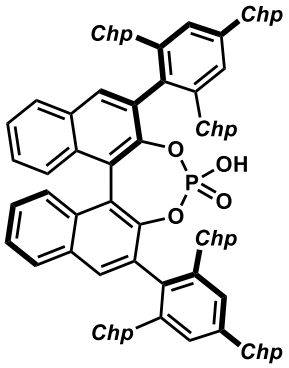
^{19}F NMR (376 MHz, $CDCl_3$) δ -62.96.

^{31}P NMR (162 MHz, $CDCl_3$) δ 5.30.



***(R)*-4-hydroxy-2,6-bis(2,4,6-tricyclopentylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, CPA-6:** According to **General Procedure D** followed by **General Procedure E** with (*R*)-3,3'-bis(2,4,6-tricyclopentylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-12** (115 mg, 135 umol). Spectral characterization is in accordance with literature reports.[70]

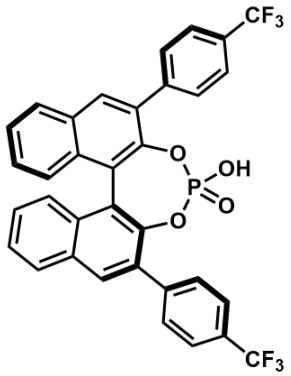
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.91 (d, *J* = 8.2 Hz, 2H), 7.81 (s, 2H), 7.54 – 7.45 (m, 2H), 7.37 – 7.30 (m, 4H), 7.06 (d, *J* = 1.7 Hz, 2H), 7.01 (d, *J* = 1.9 Hz, 2H), 2.94 (ddd, *J* = 9.9, 7.3, 2.5 Hz, 2H), 2.72 (d, *J* = 8.6 Hz, 2H), 2.65 (t, *J* = 8.1 Hz, 2H), 2.05 (q, *J* = 6.3 Hz, 4H), 2.01 – 1.92 (m, 2H), 1.89 – 1.73 (m, 8H), 1.70 – 1.64 (m, 4H), 1.62 – 1.57 (m, 10H), 1.42 – 1.20 (m, 12H).



***(R)*-4-hydroxy-2,6-bis(2,4,6-tricycloheptylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, CPA-7:** According to **General Procedure D** followed by **General Procedure E** with (*R*)-3,3'-bis(2,4,6-tricycloheptylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-13** (288 mg, 280 umol). Spectral characterization is in accordance with literature reports.[70]

^{1}H NMR (600 MHz, $CDCl_3$) δ 7.87 (d, *J* = 8.2 Hz, 2H), 7.75 (s, 2H), 7.46 (s, 2H), 7.33 – 7.27 (m, 2H), 6.92 (d, *J* = 1.8 Hz, 2H), 6.87 (d, *J* = 1.8 Hz, 2H), 2.60 (tt, *J* = 10.7, 3.5 Hz, 2H), 2.55 – 2.41 (m, 2H), 2.37 (ddt, *J* = 10.4, 6.3, 3.2 Hz, 2H), 2.05 – 1.97 (m, 2H), 1.92 (ddt, *J* = 14.4, 11.0, 3.8 Hz, 4H), 1.85 – 1.71 (m, 9H), 1.71 – 1.34 (m, 38H), 1.34 – 1.02 (m, 19H), 0.95 (dtt, *J* = 20.1, 10.6, 5.2 Hz, 4H).

^{31}P NMR (243 MHz, $CDCl_3$) δ 1.65.

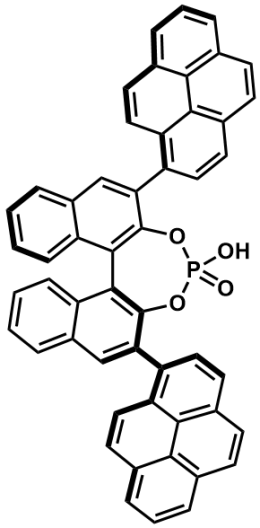


***(S)*-4-hydroxy-2,6-bis(4-trifluoromethylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine, CPA-8:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-bis(4-trifluoromethylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-11** (230 mg, 400 umol). Spectral characterization is in accordance with literature reports.[68]

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.07 – 7.94 (m, 4H), 7.68 (d, *J* = 7.9 Hz, 4H), 7.54 (td, *J* = 8.3, 4.3 Hz, 6H), 7.43 – 7.33 (m, 4H).

^{19}F NMR (471 MHz, $CDCl_3$) δ -62.56.

^{31}P NMR (202 MHz, $CDCl_3$) δ 3.76.

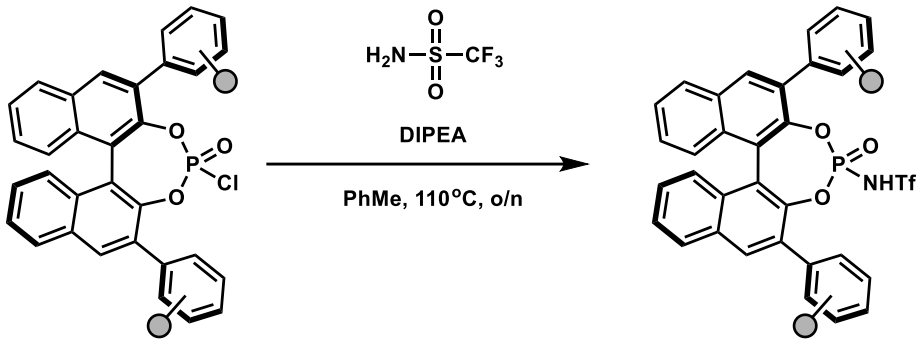


***(S)*-4-hydroxy-2,6-di(1-pyrenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide, CPA-9:** According to **General Procedure D** followed by **General Procedure E** with (*S*)-3,3'-di(1-pyrenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-11** (69 mg, 100 umol). Spectral characterization is in accordance with literature reports.[67]

^{1}H NMR (400 MHz, d_6-acetone) δ 8.48 – 7.80 (m, 22H), 7.87 – 7.35 (m, 6H).

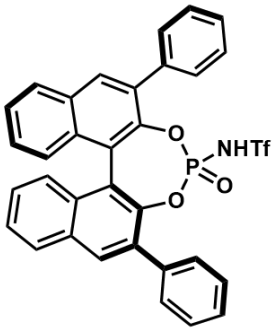
^{31}P NMR (162 MHz, d_6-acetone) δ 1.03 (br).

- ***Synthesis of N-Triflylphosphoramides, NTP***



General protocol F – substitution of phosphoryl chloride to furnish *N*-triflyl phosphoramides, NTP: To a flame dried 20 mL reaction tube equipped with a Teflon coated stir bar was added 3,3'-**arylated**-[1,1'-binaphthalene]-2,2'-diyl phosphoryl chloride (1.0 equiv), trifluoromethanesulfonamide (2.0 equiv), and toluene (0.1 M). To the suspension was charged diisopropylethylamine (4.0 equiv). The reaction tube was sealed and heated to 110°C overnight. After, the solution was transferred to a separatory funnel, quenched with 20 mL of 1 N HCl (aq), and extracted with 3x20 mL DCM. The organic fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo* to furnish a crude solid containing the **title compound**.

The crude solid was subjected to column chromatography 0-10% MeOH/DCM) to furnish a mixed salt of the **title compound**. The isolated salt was resuspended in 3-5 mL of dichloromethane and 3-5 mL of 3 N HCl (aq). The suspension was vortexed for 10 seconds and allowed to completely settle for 5 minutes. The dichloromethane layer was separated by glass pipette and concentrated to furnish a foamy, white solid containing pure, **title phosphoramide**.

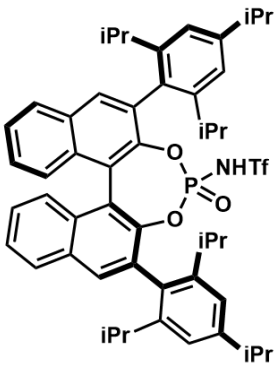


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-diphenyldinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, Ph-NTP: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-diphenyl-[1,1'-binaphthalene]-2,2'-diol, **S-4** (44 mg, 100 umol). Spectral characterization is in accordance with literature reports.[72]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 (d, *J* = 14.4 Hz, 2H), 7.98 (t, *J* = 8.8 Hz, 2H), 7.72 – 7.58 (m, 4H), 7.58 – 7.27 (m, 12H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -77.88.

^{31}P NMR (162 MHz, $CDCl_3$) δ -3.18.

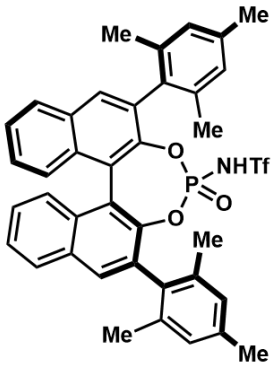


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-bis(2,4,6-triisopropylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, TRIP-NTP: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-bis(2,4,6-triisopropylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-6** (69 mg, 100 umol). Spectral characterization is in accordance with literature reports.[72]

^{1}H NMR (400 MHz, d_6-DMSO) δ 8.00 (t, *J* = 7.5 Hz, 2H), 7.86 (d, *J* = 14.7 Hz, 2H), 7.49 – 7.35 (m, 2H), 7.27 (s, 2H), 7.14 – 6.74 (m, 6H), 2.98 – 2.71 (m, 4H), 1.24 – 1.15 (m, 18H), 1.12 (d, *J* = 6.8 Hz, 6H), 1.00 (d, *J* = 6.7 Hz, 6H), 0.80 (dd, *J* = 18.6, 6.8 Hz, 6H). **Note –** 2H of the isopropyl groups are likely buried within the residual solvent signal.

^{19}F NMR (470 MHz, $CDCl_3$) δ -77.18.

^{31}P NMR (162 MHz, $CDCl_3$) δ -0.26.

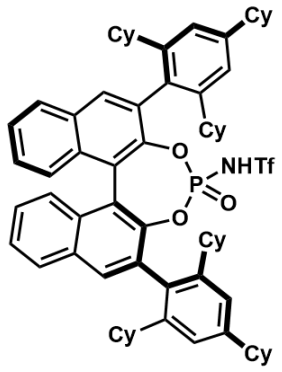


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-bis(2,4,6-trimethylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, Mesit-NTP: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-bis(2,4,6-trimethylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-5** (53 mg, 100 umol). Spectral characterization is in accordance with literature reports.[73]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.94 (d, *J* = 8.1 Hz, 2H), 7.88 – 7.82 (m, 2H), 7.52 (tt, *J* = 6.4, 2.3 Hz, 2H), 7.36 – 7.29 (m, 4H), 7.05 – 6.86 (m, 4H), 2.30 (d, *J* = 6.4 Hz, 6H), 2.23 (s, 3H), 2.11 (s, 3H), 2.03 (s, 6H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -78.55.

^{31}P NMR (162 MHz, $CDCl_3$) δ 0.10.

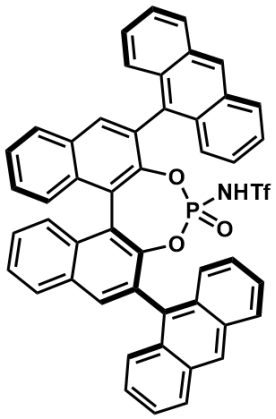


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-bis(2,4,6-tricyclohexylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, TCYP-NTP: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-bis(2,4,6-tricyclohexylphenyl)-[1,1'-binaphthalene]-2,2'-diol (93 mg, 100 umol). Spectral characterization is in accordance with literature reports.[20]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.93 (dd, *J* = 16.0, 9.9 Hz, 4H), 7.55 (t, *J* = 7.5 Hz, 2H), 7.34 – 7.26 (m, 4H), 7.14 – 7.06 (m, 3H), 6.99 (s, 1H), 2.57 – 2.45 (m, 2H), 2.32 (t, *J* = 11.4 Hz, 3H), 2.22 – 2.04 (m, 24H), 1.94 – 1.80 (m, 11H), 1.80 – 1.29 (m, 41H), 1.21 – 0.61 (m, 18H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -77.44.

^{31}P NMR (162 MHz, $CDCl_3$) δ -2.53.

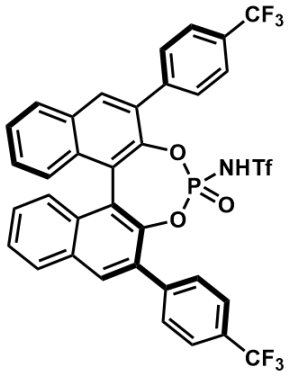


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-di(9-anthracenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, Anth-NTP: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-di(9-anthracenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-8** (64 mg, 100 umol). Spectral characterization is in accordance with literature reports.[74]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.43 (d, *J* = 13.6 Hz, 2H), 8.12 (d, *J* = 6.3 Hz, 2H), 8.06 – 7.85 (m, 5H), 7.80 (d, *J* = 8.7 Hz, 2H), 7.74 – 7.46 (m, 9H), 7.46 – 7.28 (m, 6H), 7.08 (dd, *J* = 12.4, 7.9 Hz, 2H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -79.04.

^{31}P NMR (162 MHz, $CDCl_3$) δ -1.13.

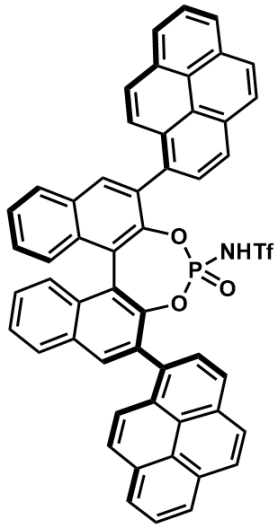


1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-bis(4-trifluoromethylphenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, NTP-8: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-bis(4-trifluoromethylphenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-11** (230 mg, 400 umol). Spectral characterization is in accordance with literature reports.[58]

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.13 (s, 1H), 8.10 – 7.99 (m, 3H), 7.77 (d, *J* = 8.0 Hz, 2H), 7.69 – 7.59 (m, 7H), 7.56 (ddd, *J* = 8.2, 6.2, 1.6 Hz, 1H), 7.52 – 7.32 (m, 4H).

^{19}F NMR (471 MHz, $CDCl_3$) δ -62.87, -63.26, -78.30.

^{31}P NMR (202 MHz, $CDCl_3$) δ -4.83.



1,1,1-trifluoro-*N*-((*S*)-4-oxido-2,6-di(1-pyrenyl)dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide, NTP-9: According to **General Procedure D** followed by **General Procedure F** with (*S*)-3,3'-di(1-pyrenyl)-[1,1'-binaphthalene]-2,2'-diol, **S-11** (69 mg, 100 umol). Spectral characterization is in accordance with literature reports.[75]

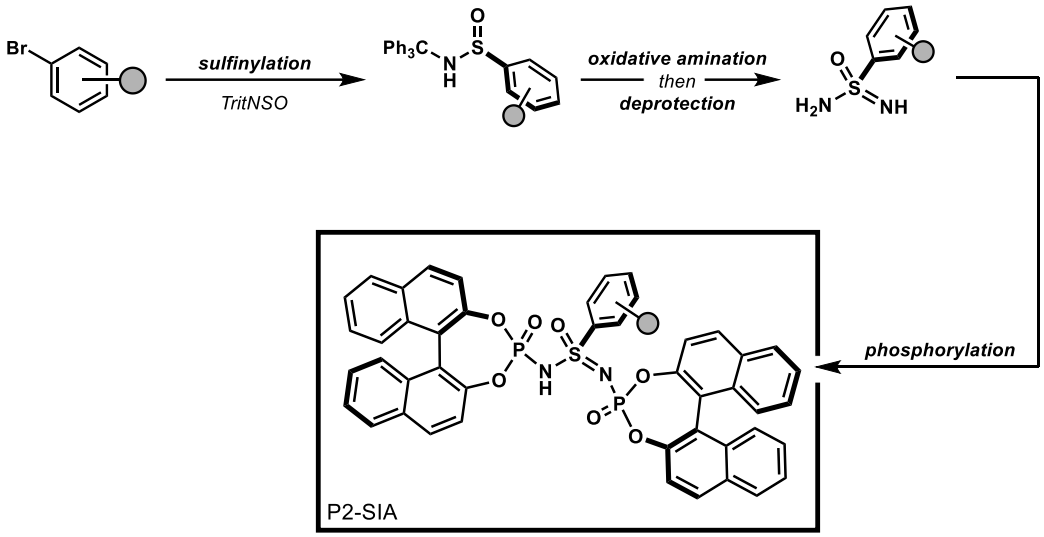
^{1}H NMR (400 MHz, d_6-acetone) δ 8.53 – 7.79 (m, 22H), 7.75 – 7.58 (m, 4H), 7.57 – 7.42 (m, 2H).

^{19}F NMR (376 MHz, d_6-acetone) δ -80.25, -80.88.

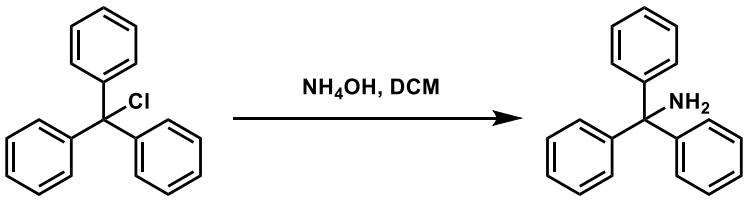
^{31}P NMR (162 MHz, d_6-acetone) δ 0.86, 0.34.

- ***Synthesis of P2-SIA***

Synthetic access into the P2-SIA scaffold is summarized below. Modular substitution at *S* is achieved by arylation of the divergent synthon TritNSO readily accessed in two steps from trityl chloride.



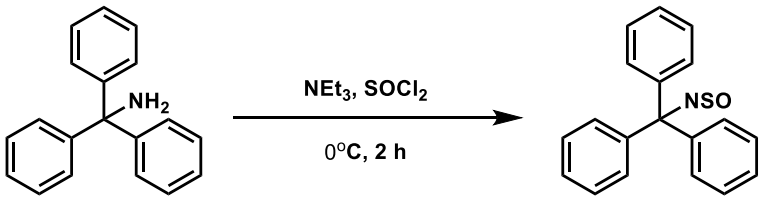
Supplementary Fig. VI-189. Overview of synthetic access into the P2-SIA scaffold.



Tritylamine, S-12: To a flame-dried 1 L round bottom flask was added DCM (250 mL) and ammonium hydroxide solution (500 mL). The suspension was cooled to 0°C in an ice bath. To the flask was then added triphenylmethyl chloride (50 g, 1.0 equiv., 180 mmol) in 5 g portions (**note** – significant exothermic reaction occurs). After addition, the biphasic solution was allowed to warm to room temperature and stirred *vigorously* overnight. The solution was then transferred to a 2 L separatory funnel and the DCM layer was collected. The aqueous layer was then washed with 1x100 mL of DCM. The organic fractions were pooled, washed with 3x250 mL of water, dried with sodium sulfate, and concentrated *in vacuo* to furnish the **title compound** (44 g, 171 mmol, 95% yield) as a white solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.30-7.21 (m, 15H), 2.32 (br, 2H)

^{13}C NMR (101 MHz, $CDCl_3$) δ 148.67, 128.18, 127.96, 126.62, 66.27.

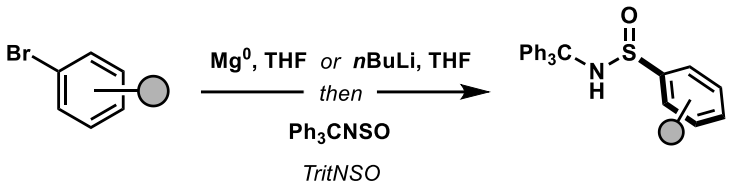


TritNSO

(Tritylimino)-l4-sulfanone, TritNSO, S-13: Adapted from the protocol reported by Davies, et al.[61] To a flame dried 250 mL round bottom flask was charged tritylamine (10.0 g, 1.0 equiv., 39 mmol), diethyl ether (140 mL), and triethylamine (11.7 mL, 2.2 equiv., 84 mmol) in one portion. The suspension was cooled in an ice bath. While stirring vigorously, thionyl chloride (3.0 mL, 1.1 equiv, 42 mmol) was added dropwise (note – white slurry formed). After addition, the suspension was allowed to stir for 2 h at the temperature.

After stirring, the suspension was diluted with 100 mL of diethyl ether, filtered over a pad of sodium sulfate and the solid was washed with an additional 2x100 mL of diethyl ether. The organic fractions were pooled and concentrated *in vacuo* to furnish the **title compound** (11.07 g, 36 mmol, 94% yield) as a yellow oil that solidifies to a pale-yellow solid upon standing.

^{13}C NMR (101 MHz, $CDCl_3$) δ 144.41, 138.69, 138.23, 127.69, 81.18



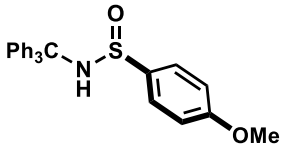
TritNSO

General protocol G – arylation of TritNSO to furnish *S*-aryl sulfinamides: Adapted from the protocol reported by Davies, et al. To a flame dried 20 mL scintillation vial containing a Teflon coated stir bar was charged **S-12** (1.00 mmol, 1.0 equiv) and THF (10 mL, 0.1 M). The solution was cooled, charged with the corresponding aryl nucleophile (1.10 mmol, 1.1 equiv), removed from the ice bath, and stirred at room temperature for an additional five minutes.

Note – the aryl nucleophile is usually a freshly prepared solution of aryl Grignard in THF (0.5 M) by mixing the corresponding aryl bromide (1.10 mmol, 1.1 equiv) with ground magnesium turnings (30 mg, 1.25 mmol, 1.25 equiv). Alternatively, it was prepared from the lithium-halogen exchange with *n*-butyllithium (1.0 equiv) and a solution of the corresponding aryl bromide (1.1 mmol, 1.1 equiv) in THF (5 mL) at -78ºC, stirring for 1 hr.

After, the solution was transferred to a separatory funnel, diluted with ethyl acetate (20 mL) and washed with sat. ammonium chloride (20 mL). The organic layer was separated, and the aqueous layer was extracted with an additional 20 mL of ethyl acetate. The organic fractions were pooled, washed with sat. sodium chloride (20 mL), dried over sodium sulfate, and concentrated *in vacuo*

to furnish a crude oil containing the **title compound**. The crude oil was subjected to flash column chromatography (0-100% EtOAc/Hex) to furnish the **title compound**.



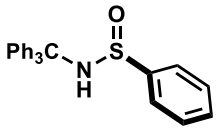
4-methoxy-*N*-tritylbenzenesulfinamide, S-14: Prepared according to **General Procedure G** with 4-methoxy-1-bromobenzene (138 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (162 mg, 390 µmol, 39%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.58 (d, *J* = 8.8 Hz, 2H), 7.39 – 7.23 (m, 15H), 6.96 (d, *J* = 8.9 Hz, 2H), 5.15 (s, 1H), 3.83 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.80, 144.84, 138.63, 129.65, 128.13, 127.55, 127.40, 114.47, 73.48, 55.62.

HRMS (ESI+) Expected $[M+H]^+$ = 414.1522, **Observed $[M+H]^+$** = 414.5208.

HRMS (ESI+) Expected $[M+Na]^+$ = 436.1342, **Observed $[M+Na]^+$** = 436.1341.



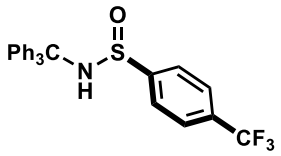
***N*-tritylbenzenesulfinamide, S-15:** Prepared according to **General Procedure G** with bromobenzene (115 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (211 mg, 550 µmol, 55%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.72 – 7.67 (m, 2H), 7.48 (m, 3H), 7.42 – 7.37 (m, 6H), 7.37 – 7.33 (m, 6H), 7.33 – 7.28 (m, 3H), 5.23 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 147.28, 144.64, 130.83, 129.59, 129.01, 128.06, 127.51, 125.65, 73.57.

HRMS (ESI+) Expected $[M+Na]^+$ = 406.1236, **Observed $[M+Na]^+$** = 406.1239.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 366.1316, **Observed $[M-H_2O]^+$** = 366.1310.



4-trifluoromethyl-*N*-tritylbenzenesulfinamide, S-16: Prepared according to **General Procedure G** with 4-trifluoromethyl-1-bromobenzene (154 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (303 mg, 670 µmol, 67%).

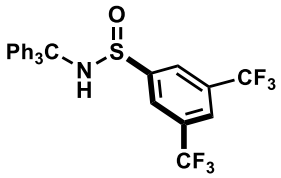
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (d, *J* = 8.2 Hz, 2H), 7.74 (d, *J* = 8.3 Hz, 2H), 7.45 – 7.29 (m, 15H), 5.28 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 151.42, 144.44, 133.04, 129.65, 128.29, 127.81, 126.44, 126.14, 123.75, 73.95.

^{19}F NMR (376 MHz, $CDCl_3$) δ -62.66.

HRMS (ESI+) Expected $[M+Na]^+$ = 474.1110, **Observed $[M+Na]^+$** = 474.1116.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 433.1190, **Observed $[M-H_2O]^+$** = 433.1194.



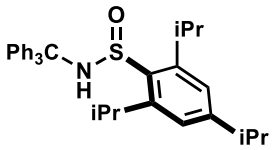
3,5-bis(trifluoromethyl)-*N*-tritylbenzenesulfinamide, S-17: Prepared according to **General Procedure G** with 3,5-bis(trifluoromethyl)-1-bromobenzene (190 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (288 mg, 550 µmol, 55%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.93 (d, *J* = 1.7 Hz, 2H), 7.83 (s, 1H), 7.29 – 7.18 (m, 15H), 5.24 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 150.66, 144.12, 132.68, 129.52, 128.41, 127.99, 126.38, 124.84, 122.89, 74.10.

^{19}F NMR (376 MHz, $CDCl_3$) δ -62.68.

HRMS (ESI+) Expected $[M+Na]^+$ = 542.0989, **Observed $[M+Na]^+$** = 542.0987.



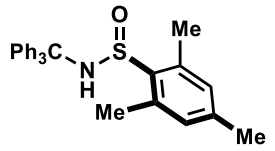
2,4,6-triisopropyl-*N*-tritylbenzenesulfinamide, S-18: Prepared according to **General Procedure G** with 1,3,5-tri(*iso*-propyl)-2-bromobenzene (278 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (431 mg, 850 µmol, 85%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 77.45 – 7.27 (m, 15H), 7.06 (s, 2H), 5.61 (s, 1H), 2.88 (p, *J* = 6.9 Hz, 1H), 1.24 (dd, *J* = 9.9, 4.3 Hz, 18H). **note** – restricted rotation between the *o*-isopropyl groups to the sulfinamide display significant broadening at the signals, seen in broad singlet representing the 12 H of the methyl groups, while the remaining 2 H of the methine groups cannot be detected under routine characterization.

^{13}C NMR (101 MHz, $CDCl_3$) δ 151.51, 144.98, 140.25, 129.58, 128.03, 127.99, 127.43, 73.94, 34.36, 28.49, 24.91, 23.90, 23.87. **note** – restricted rotation between the *o*-isopropyl groups to the sulfinamide produces observable diastereotopicity to the chiral S(IV) group, seen with very narrow splitting at 23.81 and 23.78 signals where only a single signal would be expected under full symmetry.

HRMS (ESI+) Expected $[M+H]^+$ = 510.2825, **Observed $[M+H]^+$** = 510.2827.

HRMS (ESI+) Expected $[M+Na]^+$ = 532.2645, **Observed $[M+Na]^+$** = 532.2644.



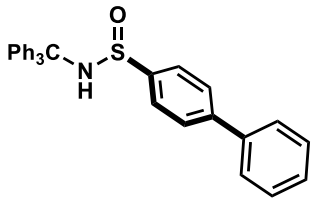
2,4,6-trimethyl-*N*-tritylbenzenesulfinamide, S-19: Prepared according to **General Procedure G** with 1,3,5-tri(methyl)-2-bromobenzene (168 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (217 mg, 510 µmol, 51%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 – 7.36 (m, 5H), 7.36 – 7.27 (m, 9H), 6.86 (s, 2H), 5.55 (s, 1H), 2.50 (s, 6H), 2.29 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 144.82, 140.41 (overlapping signal), 136.09, 130.89, 129.60, 128.05, 127.50, 73.48, 21.12, 19.65.

HRMS (ESI+) Expected $[M+H]^+$ = 426.1886, **Observed $[M+H]^+$** = 426.1880.

HRMS (ESI+) Expected $[M+Na]^+$ = 448.1706, **Observed $[M+Na]^+$** = 448.1701.



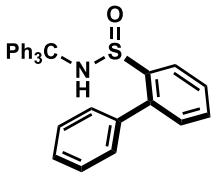
***N*-trityl-[1,1'-biphenyl]-4-sulfinamide, S-20:** Prepared according to **General Procedure G** with 4-bromobiphenyl (256 mg, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (224 mg, 490 μmol, 49%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 – 7.74 (m, 2H), 7.73 – 7.68 (m, 2H), 7.65 – 7.59 (m, 2H), 7.51 – 7.30 (m, 17H), 5.35 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 146.01, 144.69, 143.86, 139.92, 129.64, 129.01, 128.12, 127.74, 127.57, 127.32, 126.23, 73.64.

HRMS (ESI+) Expected $[M+Na]^+$ = 482.1549, **Observed $[M+Na]^+$** = 482.1544.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 442.1629, **Observed $[M-H_2O]^+$** = 442.1626.



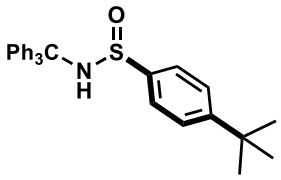
***N*-trityl-[1,1'-biphenyl]-2-sulfinamide, S-21:** Prepared according to **General Procedure H** with 2-bromobiphenyl (190 μL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (195 mg, 420 μmol, 42%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.25 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.46 (dtd, *J* = 24.9, 7.4, 1.4 Hz, 2H), 7.24 – 7.18 (m, 2H), 7.14 – 7.06 (m, 12H), 6.82 (dt, *J* = 6.7, 1.6 Hz, 6H), 4.74 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 145.98, 144.00, 140.70, 138.76, 130.84, 130.75, 129.81, 129.59, 128.29, 128.15, 127.63, 127.53, 127.31, 123.59, 73.51.

HRMS (ESI+) Expected $[M+H]^+$ = 460.1730, **Observed $[M+H]^+$** = 460.1734.

HRMS (ESI+) Expected $[M+Na]^+$ = 482.1549, **Observed $[M+Na]^+$** = 482.1550.



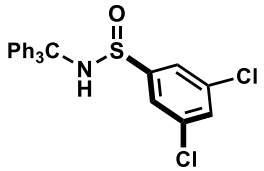
4-*tert*-butyl-*N*-tritylbenzenesulfinamide, S-22: Prepared according to **General Procedure H** with 4-*tert*-butyl-1-bromobenzene (191 µL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (236 mg, 540 µmol, 54%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.50 (d, *J* = 33.8 Hz, 2H), 7.44 (s, 2H), 7.36 – 7.23 (m, 15H), 5.16 (s, 1H), 1.24 (d, *J* = 5.5 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 154.44, 144.86, 144.21, 129.70, 128.15, 128.06, 127.57, 126.16, 125.61, 73.58, 35.05, 31.37.

HRMS (ESI+) Expected $[M+Na]^+$ = 462.1862, **Observed $[M+Na]^+$** = 462.1867.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 422.1937, **Observed $[M-H_2O]^+$** = 422.1938.

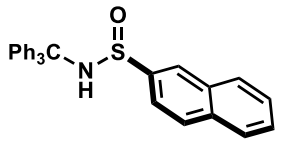


3,5-dichloro-*N*-tritylbenzenesulfinamide, S-23: Prepared according to **General Procedure H** with 3,5-dichloro-1-bromobenzene (249 mg, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (233 mg, 560 µmol, 56%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.46 (d, *J* = 1.8 Hz, 2H), 7.43 (d, *J* = 1.9 Hz, 1H), 7.42 – 7.26 (m, 14H), 5.29 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 150.72, 144.29, 135.93, 131.08, 129.53, 128.33, 127.85, 124.11, 73.90.

HRMS (ESI+) Expected $[M+Na]^+$ = 474.0462, **Observed $[M+Na]^+$** = 474.0459.



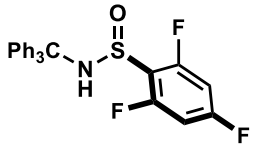
***N*-tritylnaphthalene-2-sulfinamide, S-24:** Prepared according to **General Procedure H** with 2-bromonaphthalene (227 mg, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (146 mg, 340 μmol, 34%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.25 (d, *J* = 1.8 Hz, 1H), 7.98 – 7.86 (m, 3H), 7.64 – 7.53 (m, 3H), 7.49 – 7.41 (m, 6H), 7.41 – 7.30 (m, 9H), 5.26 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 144.74, 144.34, 134.42, 132.93, 129.70, 129.14, 128.94, 128.20, 127.97, 127.89, 127.64, 127.20, 125.87, 122.04, 73.72.

HRMS (ESI+) Expected $[M+Na]^+$ = 456.1393, **Observed $[M+Na]^+$** = 456.1384.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 416.1467, **Observed $[M-H_2O]^+$** = 416.1466.



2,4,6-trifluoro-*N*-tritylbenzenesulfinamide, S-25: Prepared according to **General Procedure H** with 2,4,6-trifluoro-1-bromobenzene (130 μL, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (164 mg, 380 μmol, 38%).

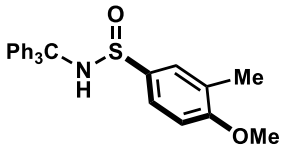
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.26 (m, 15H), 6.66 (t, *J* = 8.3 Hz, 2H), 6.12 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 164.00, 160.04, 144.33, 129.06, 128.85, 128.35, 127.64, 101.66, 73.17.

^{19}F NMR (376 MHz, $CDCl_3$) δ -102.10 (p, *J* = 8.6 Hz), -109.17 (td, *J* = 8.3, 2.6 Hz).

HRMS (ESI+) Expected $[M+Na]^+$ = 460.0959, **Observed $[M+Na]^+$** = 460.0954.

HRMS (ESI+) Expected $[M-H_2O]^+$ = 420.1034, **Observed $[M-H_2O]^+$** = 420.1027.



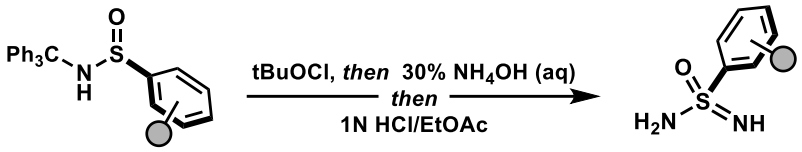
4-methoxy-3-methyl-*N*-tritylbenzenesulfinamide, S-26: Prepared according to **General Procedure H** with 4-methoxy-3-methyl-1-bromobenzene (221 mg, 1.10 mmol, 1.10 equiv) as the aryl bromide, furnishing the **title compound** (123 mg, 290 μmol, 29%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 (dd, *J* = 8.5, 2.4 Hz, 1H), 7.46 – 7.40 (m, 6H), 7.40 – 7.30 (m, 9H), 6.90 (d, *J* = 8.6 Hz, 1H), 5.12 (s, 1H), 3.89 (s, 3H), 2.24 (s, 3H).

^{13}C NMR (151 MHz, $CDCl_3$) δ 159.84, 144.83, 137.87, 129.51, 128.02, 127.69, 127.43, 127.40, 124.57, 109.91, 73.30, 55.58, 16.35.

HRMS (ESI+) Expected $[M+H]^+$ = 428.1679, **Observed $[M+H]^+$** = 428.1682.

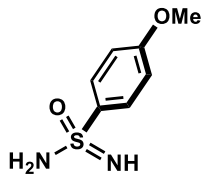
HRMS (ESI+) Expected $[M+Na]^+$ = 450.1498, **Observed $[M+Na]^+$** = 450.1503.



General protocol H – amination and deprotection to furnish sulfonimidamide, SIA: To a flame dried 20 mL scintillation vial containing a Teflon coated stir bar was charged the corresponding ***S*-aryl sulfinamide** (1.0 equiv) and THF (0.1 M with respect to the sulfinamide). To the solution was charge *tert*-butyl hypochlorite (1.1 equiv). The solution was stirred at room temperature for 15 minutes and then charged with 40% ammonium hydroxide (1.0 M with respect to the sulfinamide) in one portion. The suspension was stirred at room temperature for at least an additional 2 hours.

After, the suspension was transferred to a separatory funnel, rinsing the reaction vial with equal parts ethyl acetate and water. The solution was diluted with ethyl acetate (20 mL) and washed with sat. ammonium chloride (20 mL). The organic layer was separated, washed with sat. sodium chloride (20 mL), dried over sodium sulfate, and concentrated *in vacuo* to furnish a crude residue containing trityl-protected sulfonimidamide. The residue was transferred to another 20 mL scintillation vial containing a Teflon coated stir bar using DCM (5 mL) and then charged with 1N HCl in ethyl acetate (3 equiv. with respect to the sulfinamide). The solution was stirred at room temperature overnight to furnish a cloudy, yellow suspension.

The acidic suspension was diluted with hexanes and carefully transferred to a centrifuge tube (note – if no suspension is visible, then the solution was concentrated and reconstituted in hexanes before transferring to the centrifuge tube). The suspension was centrifuged (1,000 rpm for 5 min) and the filtrate was decanted. The pellet was washed with hexanes and the process was repeated one more time. To the crude HCl salt containing the **title compound** was charged sparing volumes of DCM and triethylamine. The suspension was then immediately subjected to flash column chromatography (0-100% EtOAc/Hex) to furnish the **title compound.**

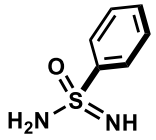


4-methoxybenzene sulfonimidamide, S-27: Prepared according to **General Procedure H** with **4-methoxy-*N*-tritylbenzenesulfinamide, S-14** (162 mg, 390 umol) as the aryl sulfinamide, furnishing the **title compound** (30 mg, 160 μmol, 41%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.97 (d, *J* = 8.9 Hz, 2H), 6.97 (d, *J* = 8.9 Hz, 2H), 3.87 (s, 3H).

^{13}C NMR (151 MHz, $CDCl_3$) δ 162.74, 135.57, 128.72, 114.15, 55.64.

HRMS (ESI+) Expected $[M+H]^+$ = 187.0536, **Observed $[M+H]^+$** = 187.0543.

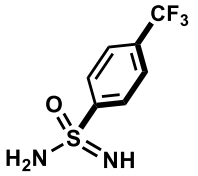


Benzene sulfonimidamide, S-28: Prepared according to **General Procedure H** with ***N*-tritylbenzenesulfinamide, S-15** (211 mg, 550 umol) as the aryl sulfinamide, furnishing the **title compound** (21 mg, 135 μmol, 30%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 (d, *J* = 7.0 Hz, 2H), 7.60 – 7.54 (m, 1H), 7.51 (t, *J* = 7.6 Hz, 2H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 143.92, 132.36, 129.06, 126.51.

HRMS (ESI+) Expected $[M+H]^+$ = 157.0430, **Observed $[M+H]^+$** = 157.0437.



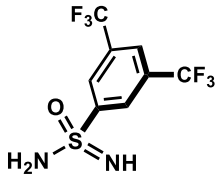
4-trifluoromethylbenzene sulfonimidamide, S-29: Prepared according to **General Procedure H** with **4-trifluoromethyl-*N*-tritylbenzenesulfinamide, S-16** (170 mg, 380 umol) as the aryl sulfinamide, furnishing the **title compound** (38 mg, 170 μmol, 45%).

^{1}H NMR (400 MHz, d_6-DMSO) δ 8.07 (d, *J* = 8.1 Hz, 2H), 7.87 (d, *J* = 8.1 Hz, 2H).

^{13}C NMR (101 MHz, d_6-DMSO) δ 150.75, 130.88, 126.79, 125.73, 123.79.

^{19}F NMR (376 MHz, d_6-DMSO) δ -61.25.

HRMS (ESI+) Expected $[M+H]^+$ = 225.0304, **Observed $[M+H]^+$** = 225.0310.



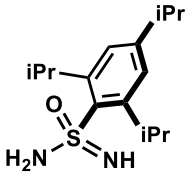
3,5-bis(trifluoromethyl)-benzene sulfonimidamide, S-30: Prepared according to **General Procedure H** with **3,5-bis(trifluoromethyl)-*N*-tritylbenzenesulfinamide, S17** (214 mg, 410 umol) as the aryl sulfinamide, furnishing the **title compound** (39 mg, 130 μmol, 32%).

^{1}H NMR (400 MHz, *d*$_6$-DMSO) δ 8.50 (s, 2H), 8.33 (s, 1H).

^{13}C NMR (101 MHz, *d*$_6$-DMSO) δ 149.83, 130.72, 126.67, 124.69, 122.90.

^{19}F NMR (376 MHz, *d*$_6$-DMSO) δ -61.45.

HRMS (ESI+) Expected [M+H]$^+$ = 293.0178, **Observed [M+H]$^+$** = 293.0188.

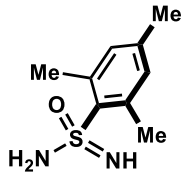


2,4,6-triisopropylbenzene sulfonimidamide, S-31: Prepared according to **General Procedure H** with **2,4,6-triisopropyl-*N*-tritylbenzenesulfinamide, S-18** (272 mg, 530 umol) as the aryl sulfinamide, furnishing the **title compound** (21 mg, 75 μmol, 14%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 7.12 (s, 2H), 4.31 (p, *J* = 6.8 Hz, 2H), 2.87 (p, *J* = 6.9 Hz, 1H), 1.19 (dd, *J* = 6.9, 1.8 Hz, 18H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 149.98, 147.83, 140.69, 122.78, 33.31, 28.58, 24.69, 23.63.

HRMS (ESI+) Expected [M+H]$^+$ = 283.1839, **Observed [M+H]$^+$** = 283.1848.

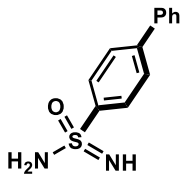


2,4,6-trimethylbenzene sulfonimidamide, S-32: Prepared according to **General Procedure H** with **2,4,6-trimethyl-*N*-tritylbenzenesulfinamide, S-19** (217 mg, 510 umol) as the aryl sulfinamide, furnishing the **title compound** (64 mg, 320 μmol, 63%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 6.94 (s, 2H), 2.70 (s, 6H), 2.29 (s, 3H).

^{13}C NMR (151 MHz, CDCl$_3$) δ 141.30, 138.81, 137.63, 131.95, 22.97, 20.82.

HRMS (ESI+) Expected [M+H]$^+$ = 199.0900, **Observed [M+H]$^+$** = 199.0902.

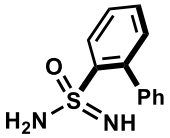


[1,1'-biphenyl]-4-sulfonimidamide, S-33: Prepared according to **General Procedure H** with ***N*-trityl-[1,1'-biphenyl]-4-sulfinamide, S-20** (224 mg, 490 umol) as the aryl sulfinamide, furnishing the **title compound** (37 mg, 160 μmol, 33%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.11 (d, *J* = 8.5 Hz, 2H), 7.72 (d, *J* = 8.5 Hz, 2H), 7.64 – 7.58 (m, 2H), 7.54 – 7.44 (m, 2H), 7.45 – 7.38 (m, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 145.56, 142.48, 139.54, 129.19, 128.58, 127.87, 127.48, 127.25.

HRMS (ESI+) Expected $[M+H]^+$ = 233.0743, **Observed $[M+H]^+$** = 223.0752.

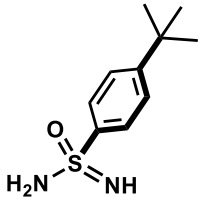


[1,1'-biphenyl]-2-sulfonimidamide, S-34: Prepared according to **General Procedure H** with ***N*-trityl-[1,1'-biphenyl]-2-sulfinamide, S-21** (195 mg, 420 umol) as the aryl sulfinamide, furnishing the **title compound** (58 mg, 250 μmol, 59%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.21 (dt, *J* = 8.1, 1.5 Hz, 1H), 7.57 – 7.52 (m, 1H), 7.52 – 7.49 (m, 2H), 7.45 (ddd, *J* = 11.4, 6.4, 1.7 Hz, 4H), 7.30 (dd, *J* = 7.6, 1.4 Hz, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 142.37, 139.59, 139.33, 132.31, 131.68, 129.69, 128.42, 128.28, 127.78, 127.53.

HRMS (ESI+) Expected $[M+H]^+$ = 233.0743, **Observed $[M+H]^+$** = 223.0753.

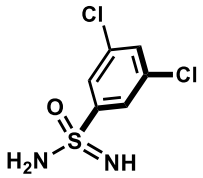


4-*tert*-butylbenzene sulfonimidamide, S-35: Prepared according to **General Procedure H** with **4-*tert*-butyl-*N*-tritylbenzenesulfinamide, S-22** (385 mg, 875 umol) as the aryl sulfinamide, furnishing the **title compound** (74 mg, 350 μmol, 40%).

^{1}H NMR (500 MHz, d_6-DMSO) δ 7.83 (d, *J* = 8.5 Hz, 2H), 7.52 (d, *J* = 8.5 Hz, 2H), 1.30 (s, 9H).

^{13}C NMR (126 MHz, d_6-DMSO) δ 153.77, 143.90, 125.83, 125.24, 34.66, 30.95.

HRMS (ESI+) Expected $[M+H]^+$ = 213.1056, **Observed $[M+H]^+$** = 213.1064.

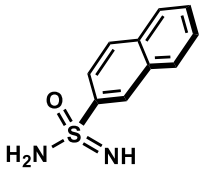


3,5-dichlorobenzene sulfonimidamide, S-36: Prepared according to **General Procedure H** with **3,5-dichloro-*N*-tritylbenzenesulfinamide, S-23** (233 mg, 560 umol) as the aryl sulfinamide, furnishing the **title compound** (49 mg, 220 μmol, 39%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93 (d, *J* = 1.8 Hz, 2H), 7.54 (t, *J* = 1.8 Hz, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 146.72, 135.91, 132.32, 125.20.

HRMS (ESI+) Expected $[M+H]^+$ = 224.9651, **Observed $[M+H]^+$** = 224.9654.

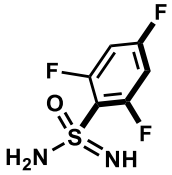


2-naphthalene sulfonimidamide, S-37: Prepared according to **General Procedure H** with ***N*-tritylnaphthalene-2-sulfinamide, S-24** (146 mg, 340 umol) as the aryl sulfinamide, furnishing the **title compound** (13 mg, 20 μmol, 6%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.48 (d, *J* = 1.8 Hz, 1H), 8.13 – 8.02 (m, 2H), 7.99 (td, *J* = 8.5, 1.9 Hz, 2H), 7.64 (tt, *J* = 7.0, 5.2 Hz, 2H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 143.88, 133.61, 131.74, 128.92, 128.45, 127.93, 127.69, 127.20, 125.57, 122.96.

HRMS (ESI+) Expected $[M+H]^+$ = 207.0587, **Observed $[M+H]^+$** = 207.0590.



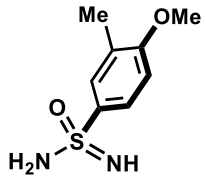
2,4,6-trifluorobenzene sulfonimidamide, S-38: Prepared according to **General Procedure H** with **2,4,6-trifluoro-*N*-tritylbenzenesulfinamide, S-25** (164 mg, 375 umol) as the aryl sulfinamide, furnishing the **title compound** (32 mg, 150 μmol, 41%).

^{1}H NMR (400 MHz, *d*$_6$-DMSO) δ 7.29 (t, *J* = 8.9 Hz, 2H).

^{13}C NMR (101 MHz, *d*$_6$-DMSO) δ 162.76, 158.83, 122.36, 101.78.

^{19}F NMR (376 MHz, *d*$_6$-DMSO) δ -104.32 (p, *J* = 9.4 Hz), -105.06 (t, *J* = 9.6 Hz).

HRMS (ESI+) Expected $[M+H]^+$ = 211.0147, **Observed $[M+H]^+$** = 211.0156.

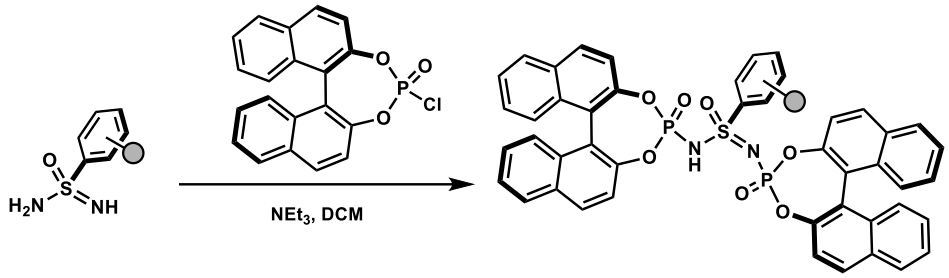


4-methoxy-3-methylbenzene sulfonimidamide, S-39: Prepared according to **General Procedure H** with **4-methoxy-3-methyl-*N*-tritylbenzenesulfinamide, S-26** (531 mg, 1.22 mmol) as the aryl sulfinamide, furnishing the **title compound** (111 mg, 550 μmol, 45%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87 (dd, *J* = 8.6, 2.5 Hz, 1H), 7.80 (d, *J* = 2.4 Hz, 1H), 6.87 (d, *J* = 8.7 Hz, 1H), 3.89 (s, 3H), 2.26 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 128.94, 127.84, 126.35, 109.53, 55.75, 16.35.

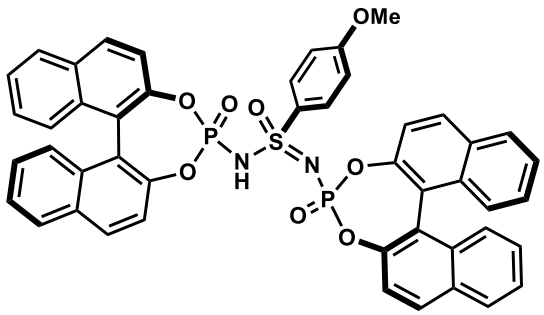
HRMS (ESI+) Expected $[M+H]^+$ = 201.0692, **Observed $[M+H]^+$** = 201.0698.



General protocol I – synthesis of diphosphoryl sulfonimidamide, P2-SIA: To a flame dried 20 mL scintillation vial containing a Teflon coated stir bar was charged the corresponding ***S*-aryl sulfonimidamide** (1.0 equiv), phosphoryl chloride (2.0 equiv), and DCM (0.1 M). The solution was charged with triethylamine (4.0 equiv) and stirred at room temperature overnight.

After, the solution was transferred to a separatory funnel, diluted with DCM (20 mL), and washed with 3N HCl (20 mL). The organic fraction was separated, dried over sodium sulfate, and concentrated *in vacuo* furnishing a crude white solid containing the **title compound**. The solid was purified by column chromatography (0-100% EtOAc/Hex) to furnish a white solid containing a mixture of phosphoramide salts. The isolated solid was resuspended in 3-5 mL DCM and vortexed with 3-5 mL 3N HCl for 10 seconds. The DCM layer was removed by pipette and concentrated *in vacuo* to furnish the **title compound.**

Note – ethyl acetate impurities in isolated P2-SIAs proved *exceptionally* challenging to remove. Characterized materials also display the following signals corresponding to ethyl acetate in d_6-DMSO (^{1}H NMR δ: 1.17 (t, 3H), 1.99 (s, 3H), 4.03 (q, 2H); ^{13}C NMR δ: 14.40, 20.68, 59.74, 170.31). These are ignored during formal analysis.



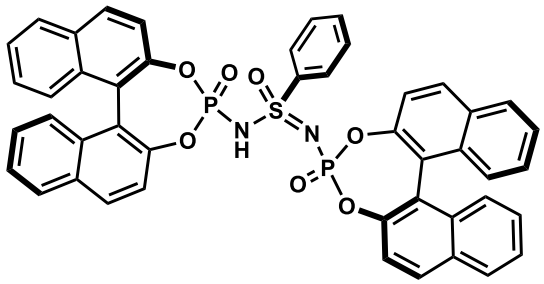
***(R,R)*-4-methoxy-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-1:** Prepared according to **General Procedure I** with **4-methoxybenzene sulfonimidamide, S-27** as the aryl sulfonimidamide, furnishing the **title compound** (49%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.16 – 8.01 (m, 7H), 7.86 (d, J = 8.8 Hz, 1H), 7.67 (d, J = 8.6 Hz, 3H), 7.56 – 7.43 (m, 6H), 7.42 – 7.30 (m, 5H), 7.27 – 7.18 (m, 4H), 6.73 (d, J = 8.9 Hz, 2H), 3.72 (s, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 160.61, 148.77, 148.67, 148.58, 147.56, 147.52, 147.49, 147.45, 139.85, 131.73, 131.70, 131.65, 130.95, 130.90, 130.80, 130.51, 130.36, 130.20, 128.61, 128.58, 128.54, 127.56, 126.63, 126.60, 126.46, 126.14, 126.10, 125.26, 125.16, 125.11, 122.47, 122.12, 121.93, 121.69, 121.67, 121.65, 121.19, 121.17, 121.11, 121.10, 112.77, 55.32.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.87 (d, *J* = 33.8 Hz).

HRMS (ESI-) Expected [M-H]$^-$ = 845.1282, **Observed [M-H]$^-$** = 845.1272.



***(R,R)-N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-2:** Prepared according to **General Procedure I** with **benzene sulfonimidamide, S-28** as the aryl sulfonimidamide, furnishing the **title compound** (26%).

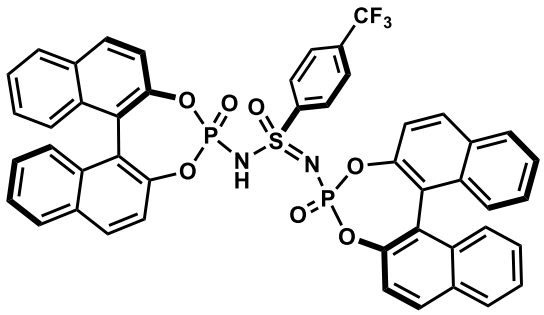
^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.14 (dd, *J* = 8.9, 2.1 Hz, 2H), 8.12 – 8.01 (m, 5H), 7.86 (d, *J* = 8.8 Hz, 1H), 7.83 – 7.77 (m, 2H), 7.68 (d, *J* = 8.7 Hz, 1H), 7.55 – 7.43 (m, 6H), 7.42 – 7.30 (m, 6H), 7.30 – 7.23 (m, 3H), 7.21 (dd, *J* = 8.7, 4.3 Hz, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 149.22, 149.11, 149.02, 148.06, 147.96, 147.94, 147.88, 132.19, 132.14, 132.11, 131.41, 131.37, 131.27, 130.98, 130.86, 130.82, 130.67, 129.05, 129.00,

128.25, 127.09, 127.06, 126.92, 126.60, 126.56, 126.08, 125.72, 125.63, 125.58, 122.90, 122.52, 122.37, 122.12, 122.09, 121.61, 121.55.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.96 (d, *J* = 44.3 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 815.1176, **Observed [M-H]$^-$** = 815.1170.



***(R,R)*-4-trifluoromethyl-*N,N′*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-3:** Prepared according to **General Procedure I** with **4-trifluoromethylbenzene sulfonimidamide, S-29** as the aryl sulfonimidamide, furnishing the **title compound** (51%).

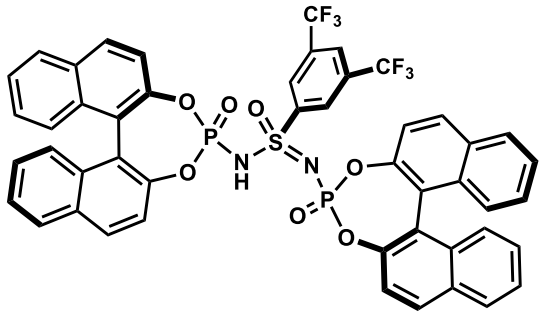
^{1}H NMR (400 MHz, *d*$_6$-DMSO) δ 8.10 (dt, *J* = 16.2, 8.3 Hz, 6H), 8.00 (d, *J* = 8.3 Hz, 1H), 7.90 (d, *J* = 8.2 Hz, 2H), 7.79 (d, *J* = 8.9 Hz, 1H), 7.69 (d, *J* = 8.8 Hz, 1H), 7.59 – 7.41 (m, 8H), 7.35 (h, *J* = 3.6 Hz, 5H), 7.27 – 7.13 (m, 4H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 151.14, 148.47, 147.27, 131.69, 131.61, 130.93, 130.79, 130.60, 130.48, 130.22, 128.63, 128.60, 128.52, 126.69, 126.60, 126.09, 125.33, 125.24, 125.17, 124.94, 124.91, 122.34, 121.91, 121.78, 121.56, 121.15, 121.07.

^{19}F NMR (376 MHz, *d*$_6$-DMSO) δ -61.12.

^{31}P NMR (162 MHz, *d*$_6$-DMSO) δ 2.37 (d, *J* = 4.7 Hz), 2.17 (d, *J* = 4.6 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 883.1050, **Observed [M+H]$^-$** = 883.1076.



***(R,R)*-3,5-bistrifluoromethyl-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2] dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-4:** Prepared according to **General Procedure I** with **3,5-bis(trifluoromethyl)-benzene sulfonimidamide, S-30** as the aryl sulfonimidamide, furnishing the **title compound** (52%).

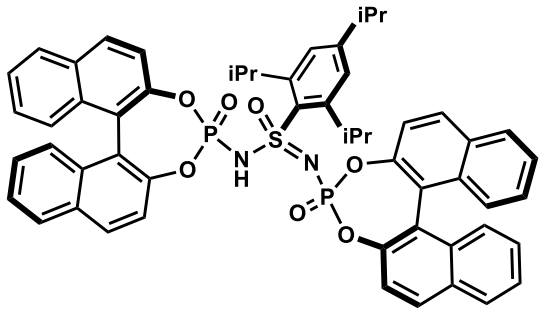
^{1}H NMR (400 MHz, *d*$_6$-DMSO) δ 8.45 (s, 2H), 8.19 – 8.03 (m, 8H), 8.00 (d, *J* = 8.2 Hz, 1H), 7.76 (d, *J* = 8.8 Hz, 1H), 7.68 (d, *J* = 8.8 Hz, 1H), 7.60 – 7.43 (m, 7H), 7.36 (dd, *J* = 15.0, 8.1 Hz, 5H), 7.22 (d, *J* = 8.6 Hz, 2H), 7.17 (d, *J* = 8.6 Hz, 2H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 149.85, 148.24, 148.15, 147.08, 131.67, 131.60, 131.55, 131.00, 130.76, 130.69, 130.63, 130.28, 130.19, 129.92, 128.64, 128.54, 126.76, 126.64, 126.14, 126.07, 125.37, 125.26, 123.94, 121.92, 121.77, 121.61.

^{19}F NMR (376 MHz, *d*$_6$-DMSO) δ -61.30.

^{31}P NMR (162 MHz, *d*$_6$-DMSO) δ 2.61 (d, *J* = 3.9 Hz), 2.33 (d, *J* = 3.9 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 951.0924, **Observed [M-H]$^-$** = 951.0984.



***(R,R)*-2,4,6-triisopropyl-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-5:** Prepared according to **General Procedure I** with **2,4,6-triisopropylbenzene sulfonimidamide, S-31** as the aryl sulfonimidamide, furnishing the **title compound** (39%).

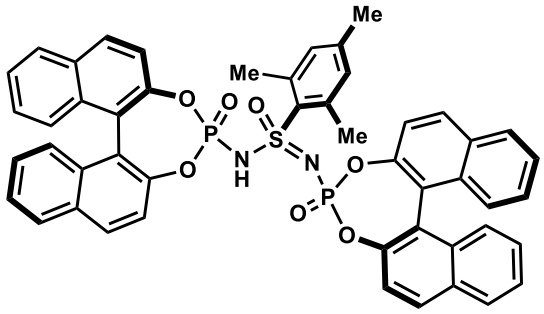
^{1}H NMR (400 MHz, *d*$_6$-DMSO) δ 8.09 (d, *J* = 8.8 Hz, 1H), 8.07 – 7.95 (m, 6H), 7.92 (d, *J* = 8.8 Hz, 1H), 7.62 (d, *J* = 8.8 Hz, 1H), 7.42 (dt, *J* = 20.1, 8.5 Hz, 5H), 7.29 (qd, *J* = 8.1, 4.5 Hz, 6H),

7.23 – 7.11 (m, 4H), 6.98 (s, 2H), 4.56 – 4.40 (m, 2H), 2.89 – 2.76 (m, 1H), 1.17 (dd, *J* = 6.9, 2.1 Hz, 6H), 0.95 (d, *J* = 6.7 Hz, 6H), 0.91 (d, *J* = 6.7 Hz, 6H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 149.27, 148.76, 147.54, 147.47, 131.75, 131.70, 131.62, 130.90, 130.71, 130.50, 130.42, 130.30, 130.09, 128.59, 128.55, 128.47, 126.58, 126.40, 126.34, 126.17, 126.13, 126.03, 125.23, 125.07, 125.02, 122.67, 122.47, 122.07, 121.93, 121.78, 121.76, 121.73, 121.71, 121.02, 121.01, 33.25, 28.13, 24.50, 24.32, 23.74, 23.65.

^{31}P NMR (162 MHz, *d*$_6$-DMSO) δ 3.52 (d, *J* = 8.7 Hz), 2.77 (d, *J* = 8.7 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 941.2585, **Observed [M-H]$^-$** = 941.2625.



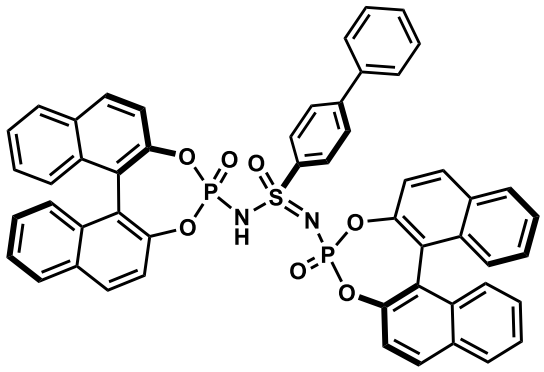
***(R,R)*-2,4,6-trimethyl-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-6:** Prepared according to **General Procedure I** with **2,4,6-trimethylbenzene sulfonimidamide, S-32** as the aryl sulfonimidamide, furnishing the **title compound** (50%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.12 (d, *J* = 8.9 Hz, 1H), 8.09 – 7.89 (m, 6H), 7.60 – 7.51 (m, 1H), 7.50 – 7.43 (m, 4H), 7.38 (d, *J* = 8.7 Hz, 1H), 7.35 – 7.26 (m, 4H), 7.21 – 7.15 (m, 3H), 6.80 (s, 2H), 2.48 (s, 6H), 2.18 (s, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 148.62, 147.68, 138.70, 136.87, 131.71, 131.57, 130.91, 130.74, 130.49, 130.42, 130.16, 128.54, 126.57, 126.46, 126.34, 126.14, 126.07, 125.21, 125.10, 125.02, 120.99, 22.87, 20.32.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 2.70 (d, *J* = 11.1 Hz), 2.12 (d, *J* = 11.1 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 857.1646, **Observed [M-H]$^-$** = 857.1615.



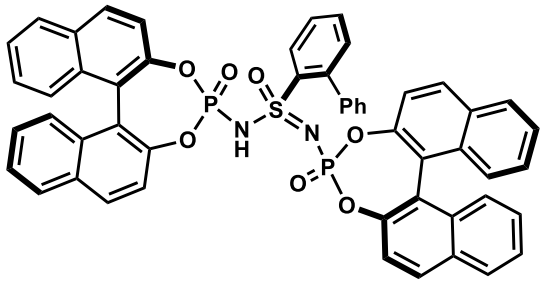
***(R,R)-N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-[1,1'-biphenyl]-4-sulfonimidamide, P2-7:** Prepared according to **General Procedure I** with **[1,1'-biphenyl]-4-sulfonimidamide, S-33** as the aryl sulfonimidamide, furnishing the **title compound** (38%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.17 – 8.05 (m, 6H), 8.05 – 7.99 (m, 1H), 7.84 (dd, *J* = 11.4, 8.7 Hz, 3H), 7.74 – 7.69 (m, 1H), 7.65 – 7.60 (m, 2H), 7.57 – 7.45 (m, 10H), 7.44 – 7.31 (m, 6H), 7.25 (d, *J* = 8.6 Hz, 1H), 7.21 (dd, *J* = 8.7, 5.3 Hz, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 142.42, 139.56, 132.18, 132.10, 131.37, 130.99, 129.45, 129.05, 127.36, 127.11, 126.93, 126.79, 126.57, 126.49, 122.36, 122.11, 121.56.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.99, 1.79.

HRMS (ESI+) Expected [M-H]$^-$ = 891.1489, **Observed [M-H]$^-$** = 891.1556.



***(R,R)-N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-[1,1'-biphenyl]-2-sulfonimidamide, P2-8:** Prepared according to **General Procedure I** with **[1,1'-biphenyl]-2-sulfonimidamide, S-34** as the aryl sulfonimidamide, furnishing the **title compound** (54%).

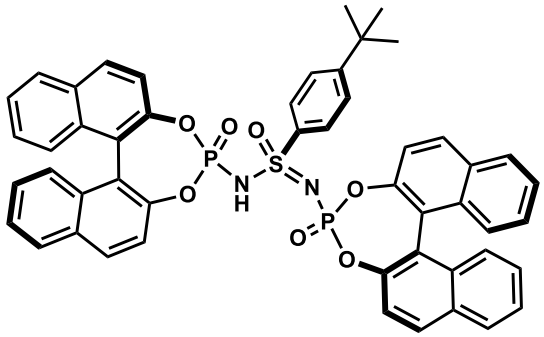
^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.18 – 8.11 (m, 5H), 8.10 – 8.04 (m, 4H), 7.92 (d, *J* = 8.2 Hz, 1H), 7.58 – 7.44 (m, 10H), 7.42 – 7.29 (m, 8H), 7.27 – 7.23 (m, 4H), 7.12 (dd, *J* = 7.6, 1.4 Hz, 1H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 149.24, 149.05, 148.96, 148.02, 147.96, 147.93, 147.86, 145.52, 141.43, 139.91, 132.56, 132.21, 132.17, 132.02, 131.44, 131.39, 131.33, 131.18, 131.15, 131.00, 130.90, 130.86, 130.62, 130.44, 130.01, 129.12, 129.06, 129.02, 128.94, 127.25, 127.18,

127.09, 127.02, 126.95, 126.88, 126.68, 126.61, 126.55, 126.51, 125.74, 125.67, 125.50, 122.87, 122.43, 122.18, 122.11, 121.67, 121.57, 121.46.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.44, 1.35.

HRMS (ESI+) Expected [M-H]$^-$ = 891.1489, **Observed [M-H]$^-$** = 891.1507.



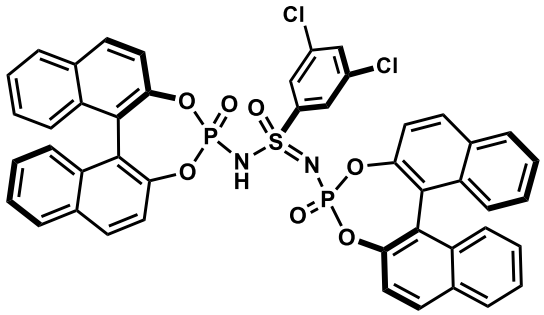
***(R,R)*-4-*tert*-butyl-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-9:** Prepared according to **General Procedure I** with **4-*tert*-butylbenzene sulfonimidamide, S-35** as the aryl sulfonimidamide, furnishing the **title compound** (59%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.14 (d, *J* = 8.9 Hz, 2H), 8.11 – 8.08 (m, 1H), 8.07 (dd, *J* = 8.8, 2.7 Hz, 3H), 8.04 (dd, *J* = 8.2, 1.3 Hz, 1H), 7.87 (d, *J* = 8.9 Hz, 1H), 7.67 – 7.62 (m, 3H), 7.56 – 7.45 (m, 6H), 7.34 (tdd, *J* = 8.5, 6.8, 1.2 Hz, 5H), 7.22 (ddt, *J* = 16.9, 8.8, 3.7 Hz, 6H), 1.23 (s, 9H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 153.13, 148.74, 148.65, 148.62, 148.54, 147.52, 147.45, 147.38, 144.77, 131.72, 131.66, 130.97, 130.92, 130.82, 130.57, 130.54, 130.42, 130.23, 128.60, 128.55, 126.67, 126.63, 126.52, 126.49, 126.15, 126.13, 126.10, 125.40, 125.30, 125.20, 125.16, 124.63, 122.40, 122.06, 121.89, 121.66, 121.63, 121.18, 121.11, 34.54, 30.93.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.94, 1.78.

HRMS (ESI+) Expected [M-H]$^-$ = 871.1802, **Observed [M-H]$^-$** = 871.1855.



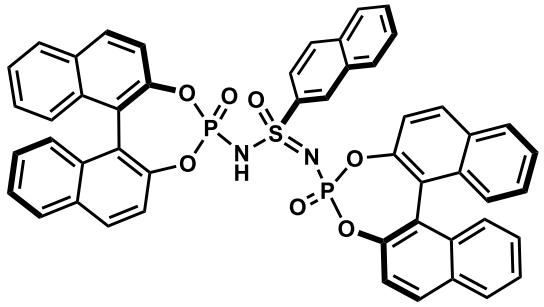
***(R,R)*-3,5-dichloro-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-10:** Prepared according to **General Procedure I** with **3,5-dichlorobenzene sulfonimidamide, S-36** as the aryl sulfonimidamide, furnishing the **title compound** (45%).

^{1}H NMR (500 MHz, d_6-DMSO) δ 8.18 – 8.11 (m, 4H), 8.09 (ddd, *J* = 7.9, 6.2, 1.3 Hz, 2H), 8.03 (dd, *J* = 8.4, 1.3 Hz, 1H), 7.82 (d, *J* = 8.9 Hz, 1H), 7.75 (d, *J* = 1.9 Hz, 2H), 7.68 (d, *J* = 8.8 Hz, 1H), 7.58 – 7.47 (m, 7H), 7.40 – 7.32 (m, 5H), 7.28 – 7.21 (m, 2H), 7.19 (dd, *J* = 8.6, 1.1 Hz, 2H).

^{13}C NMR (126 MHz, d_6-DMSO) δ 150.39, 148.39, 148.31, 148.24, 147.18, 133.56, 131.70, 131.66, 131.60, 131.03, 130.99, 130.95, 130.83, 130.70, 130.63, 130.56, 130.28, 129.91, 128.71, 128.59, 126.71, 126.66, 126.59, 126.54, 126.19, 126.15, 126.09, 125.31, 125.23, 124.65, 124.57, 122.06, 121.75, 121.60, 121.53, 121.05.

^{31}P NMR (202 MHz, d_6-DMSO) δ 1.69 (d, *J* = 7.4 Hz).

HRMS (ESI+) Expected $[M-H]^-$ = 883.0397, **Observed $[M-H]^-$** = 883.0429.



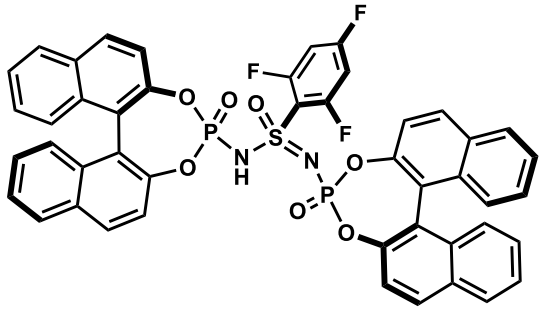
***(R,R)-N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)naphthalene-2-sulfonimidamide, P2-11:** Prepared according to **General Procedure I** with **2-naphthalene sulfonimidamide, S-37** as the aryl sulfonimidamide, furnishing the **title compound** (19%).

^{1}H NMR (500 MHz, d_6-DMSO) δ 8.19 (s, 1H), 8.12 (dd, *J* = 8.9, 3.9 Hz, 4H), 8.06 (d, *J* = 8.3 Hz, 2H), 8.05 – 7.98 (m, 1H), 7.91 – 7.86 (m, 2H), 7.84 (d, *J* = 8.8 Hz, 1H), 7.76 (d, *J* = 8.8 Hz, 1H), 7.73 (dd, *J* = 8.7, 0.8 Hz, 1H), 7.67 – 7.60 (m, 1H), 7.60 – 7.53 (m, 2H), 7.53 – 7.46 (m, 4H), 7.46 – 7.40 (m, 3H), 7.39 – 7.30 (m, 4H), 7.27 – 7.23 (m, 1H), 7.23 – 7.15 (m, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 148.75, 148.67, 148.61, 147.43, 144.67, 133.44, 131.70, 131.63, 131.40, 130.99, 130.90, 130.81, 130.55, 130.51, 130.43, 130.20, 128.79, 128.65, 128.57, 128.54, 127.71, 127.58, 127.51, 126.74, 126.65, 126.61, 126.49, 126.08, 125.25, 125.14, 122.92, 122.45, 121.84, 121.63, 121.17, 121.08.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 2.00, 1.77.

HRMS (ESI+) Expected [M-H]$^-$ = 865.1333, **Observed [M-H]$^-$** = 865.1325.



***(R,R)*-2,4,6-trifluoro-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-12:** Prepared according to **General Procedure I** with **2,4,6-trifluorobenzene sulfonimidamide, S-38** as the aryl sulfonimidamide, furnishing the **title compound** (31%).

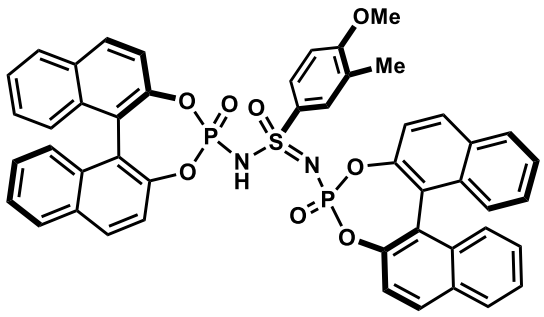
^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 8.20 – 8.04 (m, 6H), 7.94 (dt, *J* = 8.2, 0.8 Hz, 1H), 7.67 – 7.58 (m, 2H), 7.58 – 7.51 (m, 2H), 7.50 – 7.46 (m, 4H), 7.41 – 7.29 (m, 5H), 7.24 (ddd, *J* = 8.7, 2.1, 1.2 Hz, 2H), 7.17 (dd, *J* = 8.6, 1.0 Hz, 2H), 7.11 (t, *J* = 8.9 Hz, 2H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 148.44, 148.35, 148.32, 148.23, 147.26, 147.20, 147.15, 147.09, 131.69, 131.64, 131.52, 130.96, 130.90, 130.71, 130.55, 130.09, 128.59, 128.47, 127.92, 126.67, 126.61, 126.54, 126.46, 126.13, 126.10, 126.07, 126.04, 125.25, 125.14, 122.14, 121.77, 121.56, 121.52, 121.50, 121.01, 120.99, 120.87, 101.18.

^{19}F NMR (471 MHz, *d*$_6$-DMSO) δ -97.47 (t, *J* = 9.9 Hz), -100.28 (p, *J* = 9.6 Hz).

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 3.31 (d, *J* = 11.1 Hz), 1.88 (d, *J* = 11.1 Hz).

HRMS (ESI+) Expected [M-H]$^-$ = 869.0984, **Observed [M-H]$^-$** = 869.0894.



***(R,R)*-4-methoxy-3-methyl-*N,N'*-bis(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonimidamide, P2-13:** Prepared according to **General Procedure I** with **4-methoxy-3-methylbenzene sulfonimidamide, S-39** as the aryl sulfonimidamide, furnishing the **title compound** (53%).

^{1}H NMR (500 MHz, *d*$_6$-DMSO) δ 7.94 (dd, *J* = 9.0, 1.8 Hz, 2H), 7.91 – 7.81 (m, 5H), 7.69 (d, *J* = 8.8 Hz, 1H), 7.45 (d, *J* = 8.8 Hz, 1H), 7.37 – 7.26 (m, 8H), 7.21 – 7.12 (m, 5H), 7.06 – 6.98 (m, 4H), 6.51 (d, *J* = 8.7 Hz, 1H), 3.55 (s, 3H), 1.77 (s, 3H).

^{13}C NMR (126 MHz, *d*$_6$-DMSO) δ 159.22, 149.10, 149.01, 147.96, 147.90, 132.19, 132.12, 131.36, 131.27, 130.98, 130.80, 130.64, 129.04, 128.28, 127.09, 126.94, 126.59, 126.55, 125.72, 125.61, 125.15, 122.91, 122.58, 122.38, 122.12, 121.64, 109.35, 55.96, 16.38.

^{31}P NMR (202 MHz, *d*$_6$-DMSO) δ 1.99, 1.69.

HRMS (ESI+) Expected [M-H]- = 859.1438, **Observed [M-H]-** = 859.1499.

VI.39 – References

(1) Smith, A. L.; Toste, F. D. Stereoselective Generalizations over Diverse Sets of Chiral Acids Enabled by Buried Volume. *J. Am. Chem. Soc.* **2026**, jacs.5c20342.

(2) Wu, S.; Chen, P.; Duan, M.; Jiang, P.-Y.; Zhou, Q.; Xiang, S.-H.; Houk, K. N.; Tan, B. Controlling Pyramidal Nitrogen Chirality by Asymmetric Organocatalysis. *Nature* **2025**, *647* (8091), 897–905.

(3) *Model Selection and Multimodel Inference*; Burnham, K. P., Anderson, D. R., Eds.; Springer: New York, NY, 2004.

(4) Zhao, H.-W.; Jiang, F.; Chen, S.; Hu, J.; Xiang, S.-H.; Ding, W.-Y.; Lu, W.; Tan, B. Organocatalytic Asymmetric Construction and Application of Axially Chiral Spiro-Bisindoles. *Angewandte Chemie International Edition* **2025**, *64* (12), e202422951.

(5) Saito, K.; Moriya, Y.; Akiyama, T. Chiral Phosphoric Acid Catalyzed Asymmetric Synthesis of 2-Substituted 2,3-Dihydro-4-Quinolones by a Protecting-Group-Free Approach. *Org. Lett.* **2015**, *17* (13), 3202–3205.

(6) Gu, Z.; Zhang, L.; Li, H.; Cao, S.; Yin, Y.; Zhao, X.; Ban, X.; Jiang, Z. Deracemization through Sequential Photoredox-Neutral and Chiral Brønsted Acid Catalysis. *Angewandte Chemie International Edition* **2022**, *61* (49), e202211241.

(7) Guo, M. C.; Miller, S. J. Catalyst–Substrate Pairings for Carbocyclic and Heterocyclic Systems in Atroposelective Quinazolinone Synthesis. *ACS Catal.* **2024**, *14* (23), 17226–17232.

(8) Li, X.; Duan, M.; Yu, P.; Houk, K. N.; Sun, J. Organocatalytic Enantioselective Dearomatization of Thiophenes by 1,10-Conjugate Addition of Indole Imine Methides. *Nat Commun* **2021**, *12* (1), 4881.

(9) Song, X.; Zhang, Y.; Li, Y.; Zhao, X.; Yin, Y.; Ban, X.; Jiang, Z. Catalytic Asymmetric Synthesis of Azaarene-Functionalized Tertiary Amines and α-Amino Acid Derivatives from E/Z-Ketimine Mixtures via Enantioselective Radical Coupling. *ACS Catal.* **2023**, *13* (9), 6396–6402.

(10) Rueping, M.; Antonchick, A. P.; Sugiono, E.; Grenader, K. Asymmetric Brønsted Acid Catalysis: Catalytic Enantioselective Synthesis of Highly Biologically Active Dihydroquinazolinones. *Angewandte Chemie International Edition* **2009**, *48* (5), 908–910.

(11) Yu, L.; Tang, Y.; Nie, X.; Cai, Y. Stereoselective Access to Spiro-Isoindolinone Scaffolds via Catalytic Asymmetric Aza-Piancatelli Rearrangement. *Org. Lett.* **2024**, *26* (36), 7667–7671.

(12) Ding, W.-W.; He, Z.-Y.; Sayed, M.; Zhou, Y.; Han, Z.-Y.; Gong, L.-Z. Enantioselective Synthesis of β- and α-Amino Ketones through Reversible Alkane Carbonylation. *Nat. Synth* **2024**, *3* (4), 507–516.

(13) Shevchenko, G. A.; Oppelaar, B.; List, B. An Unexpected α-Oxidation of Cyclic Ketones with 1,4-Benzoquinone by Enol Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (33), 10756–10759.

(14) Wang, Y.; Song, R.-P.; Li, X.-Y.; Chen, W.-L.; Tian, Y.; Zhang, S.-H.; Shao, Y.-D.; Cheng, D.-J. Catalytic Asymmetric Reductive Amination for Axially Chiral Aryl Aldehydes via Desymmetrization/Kinetic Resolution Cascade. *Org. Lett.* **2024**, *26* (34), 7161–7165.

(15) Wang, Y.; Wang, S.; Shan, W.; Shao, Z. Direct Asymmetric N-Propargylation of Indoles and Carbazoles Catalyzed by Lithium SPINOL Phosphate. *Nat Commun* **2020**, *11* (1), 226.

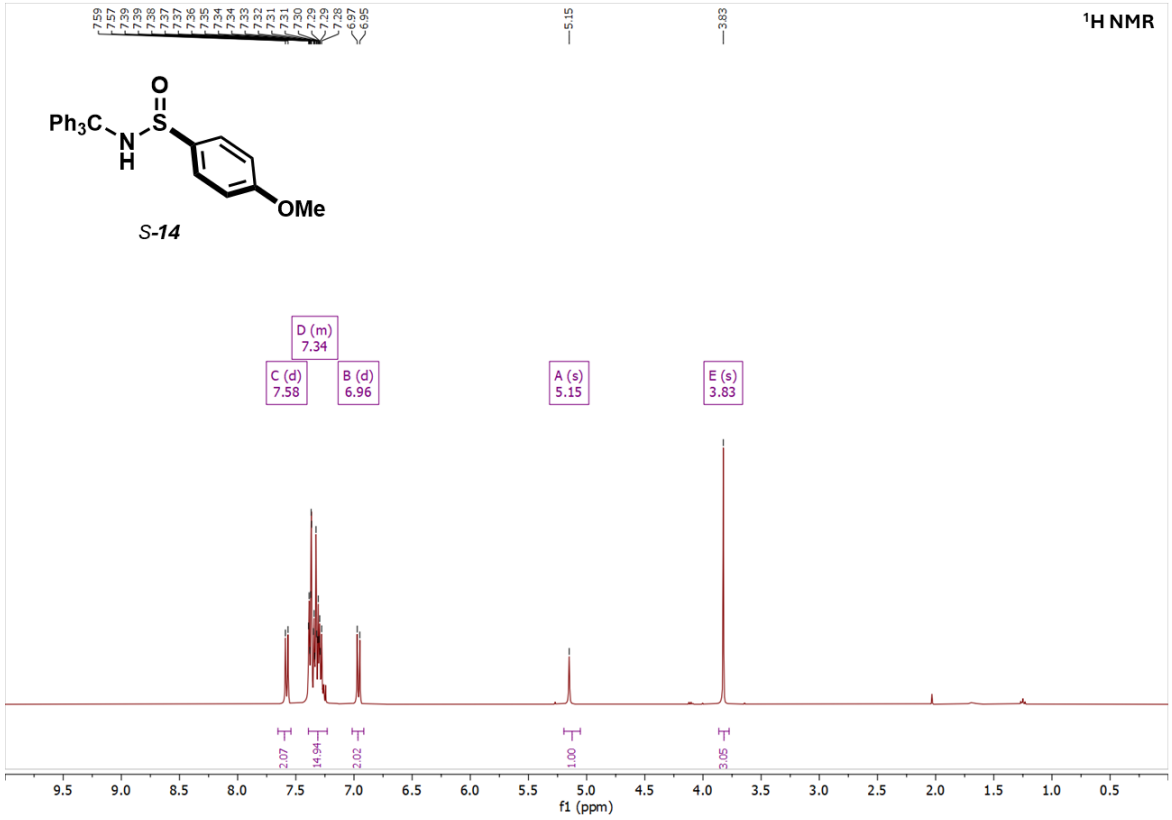
(16) Subba, P.; Sahoo, S. R.; Khajuria, C.; Singh, V. K. Enantioselective Aminative Dearomatization of Indoles via Electrophilic 1,6-Addition of p-Quinone Diimides (p-QDIs). *Org. Lett.* **2024**, *26* (23), 4932–4937.
(17) Kötzner, L.; Webber, M. J.; Martínez, A.; De Fusco, C.; List, B. Asymmetric Catalysis on the Nanoscale: The Organocatalytic Approach to Helicenes. *Angewandte Chemie International Edition* **2014**, *53* (20), 5202–5205.
(18) Huang, D.; Li, X.; Xu, F.; Li, L.; Lin, X. Highly Enantioselective Synthesis of Dihydroquinazolinones Catalyzed by SPINOL-Phosphoric Acids. *ACS Catal.* **2013**, *3* (10), 2244–2247.
(19) Zhang, Y.-Q.; Chen, Y.-B.; Liu, J.-R.; Wu, S.-Q.; Fan, X.-Y.; Zhang, Z.-X.; Hong, X.; Ye, L.-W. Asymmetric Dearomatization Catalysed by Chiral Brønsted Acids via Activation of Ynamides. *Nat. Chem.* **2021**, *13* (11), 1093–1100.
(20) Yu, H.; Hu, L.; Zhang, J.; Yang, Q.; Lu, G.; Xu, T. Organocatalytic Highly Enantioselective Formal [1,3] Sigmatropic Rearrangement of Indole Alkyl Ethers and Mechanistic Insight. *ACS Catal.* **2024**, 9207–9216.
(21) Rueping, M.; Antonchick, A. P. Catalytic Asymmetric Aminoallylation of Aldehydes: A Catalytic Enantioselective Aza-Cope Rearrangement. *Angewandte Chemie International Edition* **2008**, *47* (52), 10090–10093.
(22) Tang, M.; Zhou, J.; Xie, W.; Ren, J.; Ye, Z.; Gu, H.; Yang, X. Catalytic Enantioselective Synthesis of Mechanically Planar Chiral Rotaxanes by Organocatalyzed Desymmetrization. *Chem* **2025**, 102694.
(23) Wei, W.; Li, C.; Fan, Y.; Chen, X.; Zhao, X.; Qiao, B.; Jiang, Z. Catalytic Asymmetric Redox-Neutral [3+2] Photocycloadditions of Cyclopropyl Ketones with Vinylazaarenes Enabled by Consecutive Photoinduced Electron Transfer. *Angewandte Chemie International Edition* **2024**, *63* (30), e202406845.
(24) Zhao, Q.; Li, Y.; Zhang, Q.-X.; Cheng, J.-P.; Li, X. Catalytic Asymmetric Aza-Diels–Alder Reaction of Ketimines and Unactivated Dienes. *Angewandte Chemie International Edition* **2021**, *60* (32), 17608–17614.
(25) Liang, D.; Chen, J.-R.; Tan, L.-P.; He, Z.-W.; Xiao, W.-J. Catalytic Asymmetric Construction of Axially and Centrally Chiral Heterobiaryls by Minisci Reaction. *J. Am. Chem. Soc.* **2022**, *144* (13), 6040–6049.
(26) Biswas, S.; Kubota, K.; Orlandi, M.; Turberg, M.; Miles, D. H.; Sigman, M. S.; Toste, F. D. Enantioselective Synthesis of N,S-Acetals by an Oxidative Pummerer-Type Transformation Using Phase-Transfer Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (2), 589–593.
(27) Xia, Z.-L.; Zheng, C.; Xu, R.-Q.; You, S.-L. Chiral Phosphoric Acid Catalyzed Aminative Dearomatization of α-Naphthols/Michael Addition Sequence. *Nat Commun* **2019**, *10* (1), 3150.
(28) Yang, J.; Zhang, J.-W.; Bao, W.; Qiu, S.-Q.; Li, S.; Xiang, S.-H.; Song, J.; Zhang, J.; Tan, B. Chiral Phosphoric Acid-Catalyzed Remote Control of Axial Chirality at Boron–Carbon Bond. *J. Am. Chem. Soc.* **2021**, *143* (33), 12924–12929.
(29) Terada, M.; Soga, K.; Momiyama, N. Enantioselective Activation of Aldehydes by Chiral Phosphoric Acid Catalysts in an Aza-Ene-Type Reaction between Glyoxylate and Enecarbamate. *Angewandte Chemie International Edition* **2008**, *47* (22), 4122–4125.

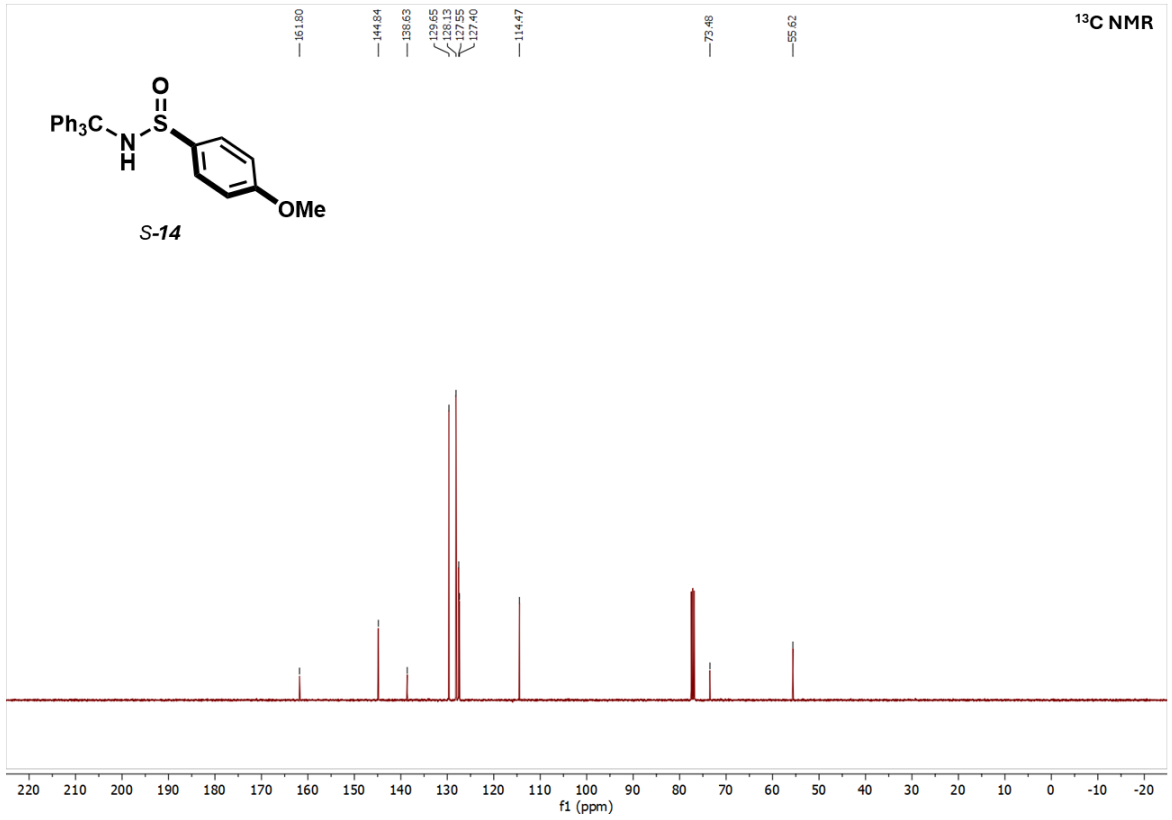
(30) Zheng, W.-F.; Chen, J.; Qi, X.; Huang, Z. Modular and Diverse Synthesis of Amino Acids via Asymmetric Decarboxylative Protonation of Aminomalonic Acids. *Nat. Chem.* **2023**, *15* (12), 1672–1682.
(31) Ye, Z.; Xie, W.; Wang, D.; Liu, H.; Yang, X. Atroposelective Synthesis of Diarylamines via Organocatalyzed Electrophilic Amination. *ACS Catal.* **2024**, *14* (7), 4958–4967.
(32) Li, G.; Yao, Y.; Wang, Z.; Zhao, M.; Xu, J.; Huang, L.; Zhu, G.; Bao, G.; Sun, W.; Hong, L.; Wang, R. Switchable Skeletal Rearrangement of Dihydroisobenzofuran Acetals with Indoles. *Org. Lett.* **2019**, *21* (11), 4313–4317.
(33) Wei, Y.; Sun, F.; Li, G.; Xu, S.; Zhang, M.; Hong, L. Enantioselective Synthesis of N–N Amide–Pyrrole Atropisomers via Paal–Knorr Reaction. *Org. Lett.* **2024**, *26* (12), 2343–2348.
(34) Yue, T.; Wang, M.-X.; Wang, D.-X.; Masson, G.; Zhu, J. Brønsted Acid Catalyzed Enantioselective Three-Component Reaction Involving the α Addition of Isocyanides to Imines. *Angewandte Chemie International Edition* **2009**, *48* (36), 6717–6721.
(35) Wang, Z.; Sheong, F. K.; Sung, H. H. Y.; Williams, I. D.; Lin, Z.; Sun, J. Catalytic Enantioselective Intermolecular Desymmetrization of Azetidines. *J. Am. Chem. Soc.* **2015**, *137* (18), 5895–5898.
(36) Wu, F.; Zhang, Y.; Zhu, R.; Huang, Y. Discovery and Synthesis of Atropisomerically Chiral Acyl-Substituted Stable Vinyl Sulfoxonium Ylides. *Nat. Chem.* **2024**, *16* (1), 132–139.
(37) Mo, N.-F.; Zhang, Y.; Guan, Z.-H. Highly Enantioselective Three-Component Povarov Reaction for Direct Construction of Azaspirocycles. *Org. Lett.* **2022**, *24* (35), 6397–6401.
(38) He, L.; Bekkaye, M.; Retailleau, P.; Masson, G. Chiral Phosphoric Acid Catalyzed Inverse-Electron-Demand Aza-Diels–Alder Reaction of Isoeugenol Derivatives. *Org. Lett.* **2012**, *14* (12), 3158–3161.
(39) Li, C.; Shao, Y.-B.; Gao, X.; Ren, Z.; Guo, C.; Li, M.; Li, X. Enantioselective Synthesis of Chiral Quinohelicenes through Sequential Organocatalyzed Povarov Reaction and Oxidative Aromatization. *Nat Commun* **2023**, *14* (1), 3380.
(40) Wen, L.; Ding, J.; Duan, L.; Wang, S.; An, Q.; Wang, H.; Zuo, Z. Multiplicative Enhancement of Stereoenrichment by a Single Catalyst for Deracemization of Alcohols. *Science* **2023**, *382* (6669), 458–464.
(41) Hoffmann, S.; Seayad, A. M.; List, B. A Powerful Brønsted Acid Catalyst for the Organocatalytic Asymmetric Transfer Hydrogenation of Imines. *Angewandte Chemie International Edition* **2005**, *44* (45), 7424–7427.
(42) Rueping, M.; Sugiono, E.; Azap, C.; Theissmann, T.; Bolte, M. Enantioselective Brønsted Acid Catalyzed Transfer Hydrogenation: Organocatalytic Reduction of Imines. *Org. Lett.* **2005**, *7* (17), 3781–3783.
(43) Liao, S.; Čorić, I.; Wang, Q.; List, B. Activation of H2O2 by Chiral Confined Brønsted Acids: A Highly Enantioselective Catalytic Sulfoxidation. *J. Am. Chem. Soc.* **2012**, *134* (26), 10765–10768.
(44) Li, Y.-P.; Li, Z.-Q.; Zhou, B.; Li, M.-L.; Xue, X.-S.; Zhu, S.-F.; Zhou, Q.-L. Chiral Spiro Phosphoric Acid-Catalyzed Friedel–Crafts Conjugate Addition/Enantioselective Protonation Reactions. *ACS Catal.* **2019**, *9* (7), 6522–6529.
(45) Meng, S.-S.; Tang, W.-B.; Zheng, W.-H. Catalytically Enantioselective Synthesis of Acyclic α-Tertiary Amines through Desymmetrization of 2-Substituted 2-Nitro-1,3-Diols. *Org. Lett.* **2018**, *20* (3), 518–521.

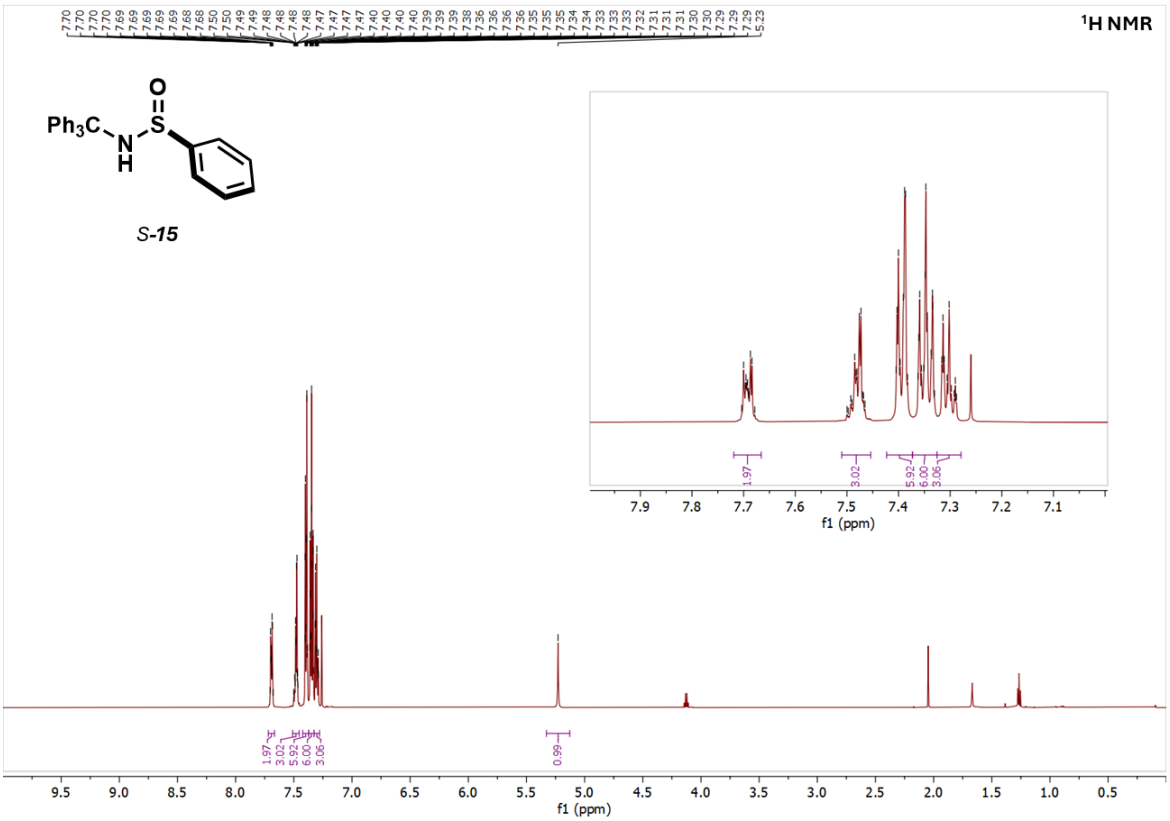
(46) Kwon, Y.; Li, J.; Reid, J. P.; Crawford, J. M.; Jacob, R.; Sigman, M. S.; Toste, F. D.; Miller, S. J. Disparate Catalytic Scaffolds for Atroposelective Cyclodehydration. *J. Am. Chem. Soc.* **2019**, *141* (16), 6698–6705.
(47) Yin, Y.; Dai, Y.; Jia, H.; Li, J.; Bu, L.; Qiao, B.; Zhao, X.; Jiang, Z. Conjugate Addition–Enantioselective Protonation of N-Aryl Glycines to α-Branched 2-Vinylazaarenes via Cooperative Photoredox and Asymmetric Catalysis. *J. Am. Chem. Soc.* **2018**, *140* (19), 6083–6087.
(48) Cheng, X.; Vellalath, S.; Goddard, R.; List, B. Direct Catalytic Asymmetric Synthesis of Cyclic Aminals from Aldehydes. *J. Am. Chem. Soc.* **2008**, *130* (47), 15786–15787.
(49) Kim, J. H.; Čorić, I.; Vellalath, S.; List, B. The Catalytic Asymmetric Acetalization. *Angewandte Chemie International Edition* **2013**, *52* (16), 4474–4477.
(50) Yang, W.; Sun, J. Organocatalytic Enantioselective Synthesis of 1,4-Dioxanes and Other Oxa-Heterocycles by Oxetane Desymmetrization. *Angewandte Chemie International Edition* **2016**, *55* (5), 1868–1871.
(51) Mao, K.; Liu, C.; Wang, Y.; Gu, C.; Putziger, J. M.; Cemalovic, N. I.; Muniz, C.; Qi, Y.; Lin, S. Dynamic Kinetic Resolution of Phosphines with Chiral Supporting Electrolytes. *Nature* **2025**, *643* (8074), 1288–1296.
(52) Sanocki, M.; Russell, H. C.; Handjaya, J.; Reid, J. P. Relative Generality and Risk: Quantitative Measures for Broad Catalyst Success. *ACS Catal.* **2024**, *14* (22), 16849–16860.
(53) Moon, J.; Shin, E.; Kwon, Y. Enantioselective Desymmetrization of Biaryls via Cooperative Photoredox/Brønsted Acid Catalysis and Its Application to the Total Synthesis of Ancistrobrevolines. *J. Am. Chem. Soc.* **2025**, *147* (15), 12800–12810.
(54) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631.
(55) Cheon, C. H.; Yamamoto, H. A Brønsted Acid Catalyst for the Enantioselective Protonation Reaction. *J. Am. Chem. Soc.* **2008**, *130* (29), 9246–9247.
(56) Zhou, H.; Zhang, P.; List, B. The Silicon–Hydrogen Exchange Reaction: Catalytic Kinetic Resolution of 2-Substituted Cyclic Ketones. *Synlett* **2021**, *32*, 1953–1956.
(57) Sherbrook, E. M.; Jung, H.; Cho, D.; Baik, M.-H.; Yoon, T. P. Brønsted Acid Catalysis of Photosensitized Cycloadditions. *Chem. Sci.* **2020**, *11* (3), 856–861.
(58) Sherbrook, E. M.; Genzink, M. J.; Park, B.; Guzei, I. A.; Baik, M.-H.; Yoon, T. P. Chiral Brønsted Acid-Controlled Intermolecular Asymmetric [2 + 2] Photocycloadditions. *Nat Commun* **2021**, *12* (1), 5735.
(59) Schwengers, S. A.; De, C. K.; Grossmann, O.; Grimm, J. A. A.; Sadlowski, N. R.; Gerosa, G. G.; List, B. Unified Approach to Imidodiphosphate-Type Brønsted Acids with Tunable Confinement and Acidity. *J. Am. Chem. Soc.* **2021**, *143* (36), 14835–14844.
(60) Liles, J. P.; Rouget-Virbel, C.; Wahlman, J. L. H.; Rahimoff, R.; Crawford, J. M.; Medlin, A.; O'Connor, V. S.; Li, J.; Roytman, V. A.; Toste, F. D.; Sigman, M. S. Data Science Enables the Development of a New Class of Chiral Phosphoric Acid Catalysts. *Chem* **2023**, *9* (6), 1518–1537.
(61) Davies, T. Q.; Hall, A.; Willis, M. C. One-Pot, Three-Component Sulfonimidamide Synthesis Exploiting the Sulfinylamine Reagent N-Sulfinyltritylamine, TrNSO. *Angewandte Chemie International Edition* **2017**, *56* (47), 14937–14941.
(62) Van Dijk, L.; Haas, B. C.; Lim, N.-K.; Clagg, K.; Dotson, J. J.; Treacy, S. M.; Piechowicz, K. A.; Roytman, V. A.; Zhang, H.; Toste, F. D.; Miller, S. J.; Gosselin, F.; Sigman, M. S.

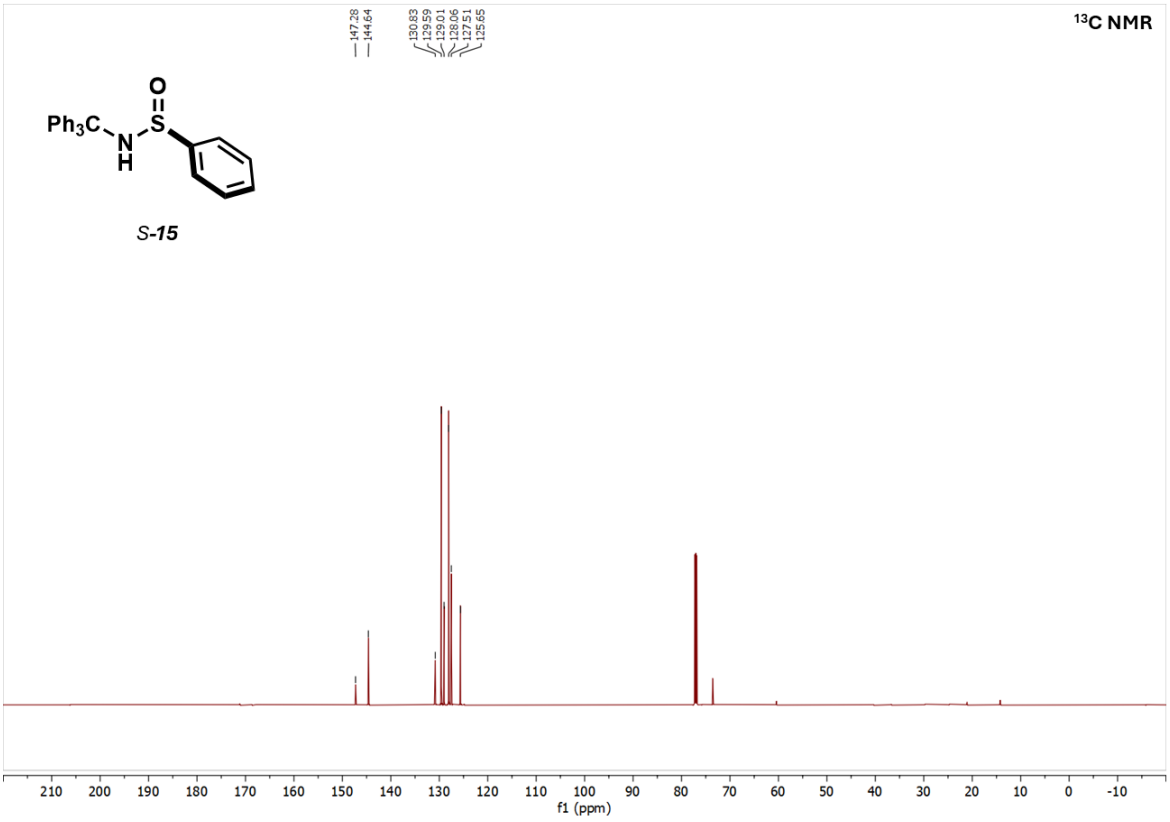
Data Science-Enabled Palladium-Catalyzed Enantioselective Aryl-Carbonylation of Sulfonimidamides. *J. Am. Chem. Soc.* **2023**, *145* (38), 20959–20967.
(63) Momiyama, N.; Nishimoto, H.; Terada, M. Chiral Brønsted Acid Catalysis for Enantioselective Hosomi−Sakurai Reaction of Imines with Allyltrimethylsilane. *Org. Lett.* **2011**, *13* (8), 2126–2129.
(64) Yamanaka, M.; Itoh, J.; Fuchibe, K.; Akiyama, T. Chiral Brønsted Acid Catalyzed Enantioselective Mannich-Type Reaction. *J. Am. Chem. Soc.* **2007**, *129* (21), 6756–6764.
(65) Klussmann, M.; Ratjen, L.; Hoffmann, S.; Wakchaure, V.; Goddard, R.; List, B. Synthesis of TRIP and Analysis of Phosphate Salt Impurities. *Synlett* **2010**, *2010*, 2189–2192.
(66) Romanov-Michailidis, F.; Guénée, L.; Alexakis, A. Enantioselective Organocatalytic Iodination-Initiated Wagner–Meerwein Rearrangement. *Org. Lett.* **2013**, *15* (22), 5890–5893.
(67) Shin, N. Y.; Ryss, J. M.; Zhang, X.; Miller, S. J.; Knowles, R. R. Light-Driven Deracemization Enabled by Excited-State Electron Transfer. *Science* **2019**, *366* (6463), 364–369.
(68) Akiyama, T.; Morita, H.; Itoh, J.; Fuchibe, K. Chiral Brønsted Acid Catalyzed Enantioselective Hydrophosphonylation of Imines: Asymmetric Synthesis of α-Amino Phosphonates. *Org. Lett.* **2005**, *7* (13), 2583–2585.
(69) Salvi, L.; Davis, N. R.; Ali, S. Z.; Buchwald, S. L. A New Biarylphosphine Ligand for the Pd-Catalyzed Synthesis of Diaryl Ethers under Mild Conditions. *Org. Lett.* **2012**, *14* (1), 170–173.
(70) McLean, L. A.; Watson, A. J. B. An Alternative Synthesis of Cycloalkyl-Substituted CPA Catalysts and Application in Asymmetric Protonation Reactions. *European Journal of Organic Chemistry* **2021**, *2021* (35), 4943–4945.
(71) Rauniyar, V.; Wang, Z. J.; Burks, H. E.; Toste, F. D. Enantioselective Synthesis of Highly Substituted Furans by a Copper(II)-Catalyzed Cycloisomerization–Indole Addition Reaction. *J. Am. Chem. Soc.* **2011**, *133* (22), 8486–8489.
(72) Nakashima, D.; Yamamoto, H. Design of Chiral N-Triflyl Phosphoramide as a Strong Chiral Brønsted Acid and Its Application to Asymmetric Diels−Alder Reaction. *J. Am. Chem. Soc.* **2006**, *128* (30), 9626–9627.
(73) Zhang, W.; Sun, J.; Lian, Z.; Song, R.; Yang, D.; Lv, J. Enantioselective Friedel–Crafts Reaction of 2-Alkynyphenols with Aromatic Ethers by Chiral Brønsted Acid Catalysis. *J. Org. Chem.* **2022**, *87* (14), 9100–9111.
(74) Franchino, A.; Martí, À.; Echavarren, A. M. H-Bonded Counterion-Directed Enantioselective Au(I) Catalysis. *J. Am. Chem. Soc.* **2022**, *144* (8), 3497–3509.
(75) Martí, À.; Montesinos-Magraner, M.; Echavarren, A. M.; Franchino, A. H-Bonded Counterion-Directed Catalysis: Enantioselective Gold(I)-Catalyzed Addition to 2-Alkynyl Enones as a Case Study. *European Journal of Organic Chemistry* **2022**, *2022* (38), e202200518.

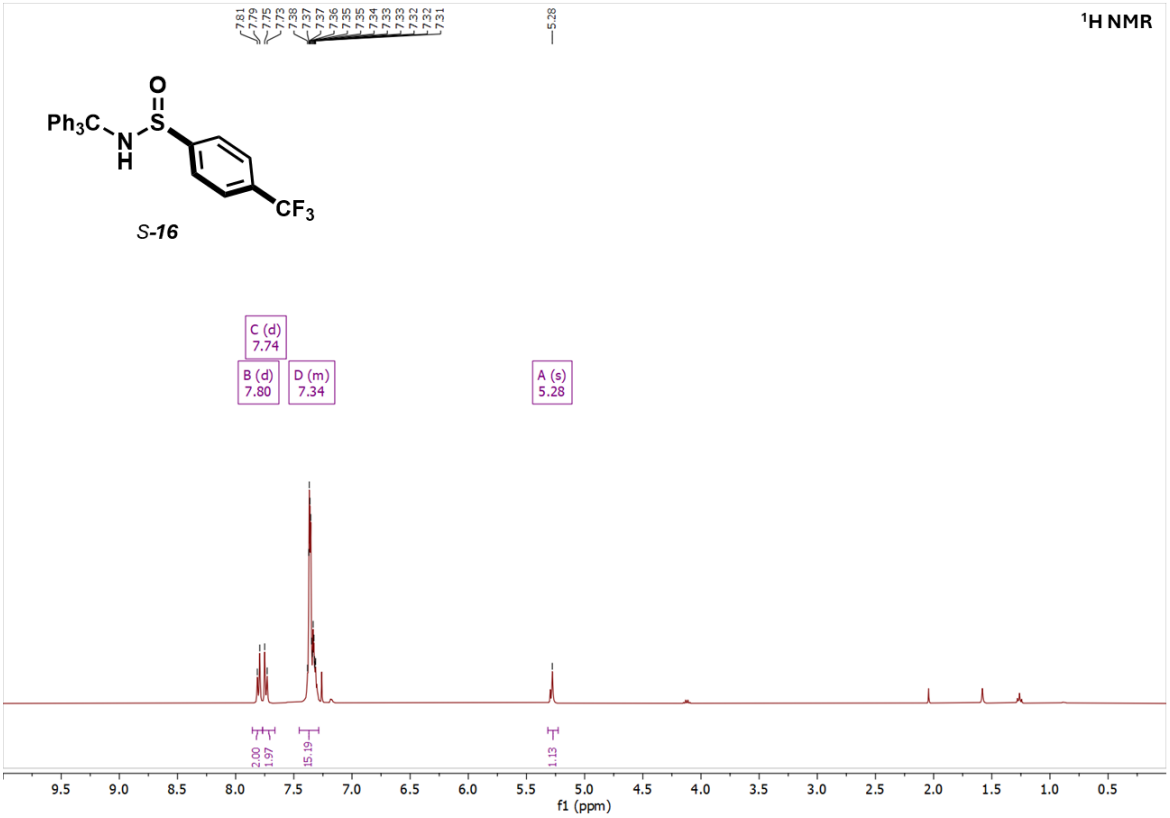
VI.40 – NMR Spectra

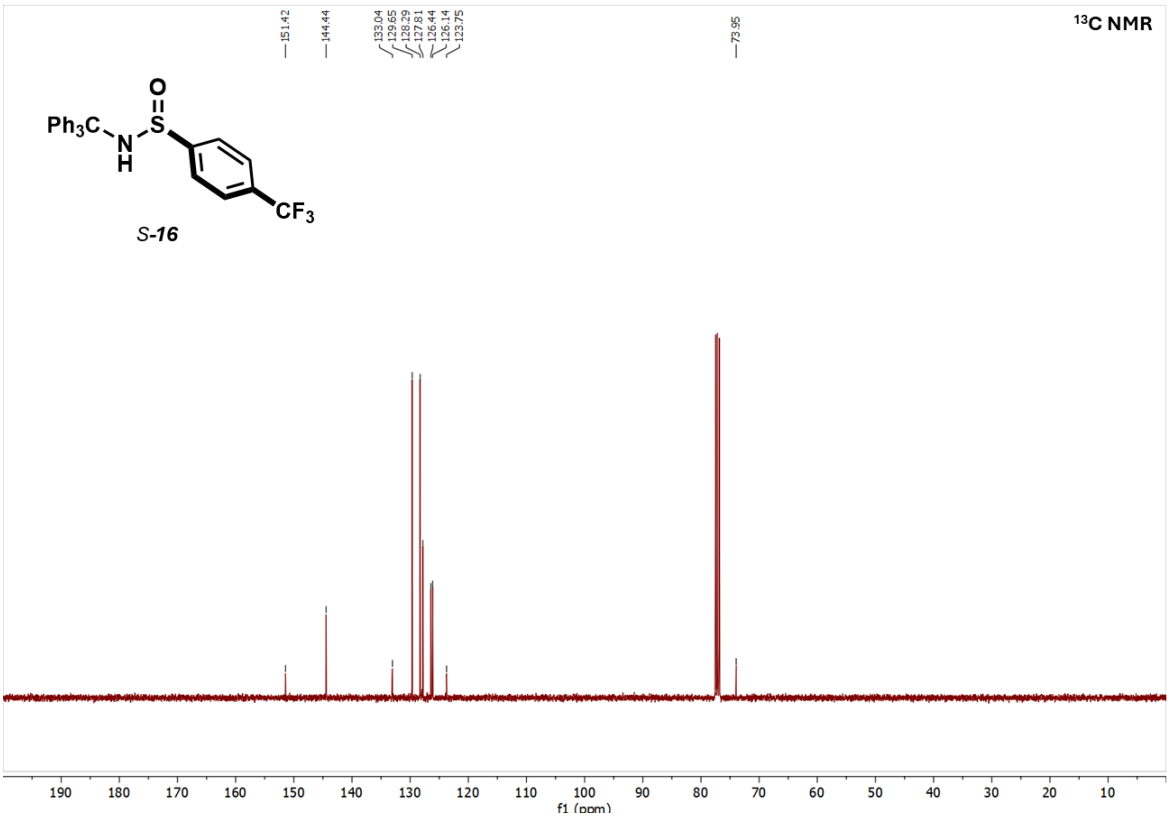


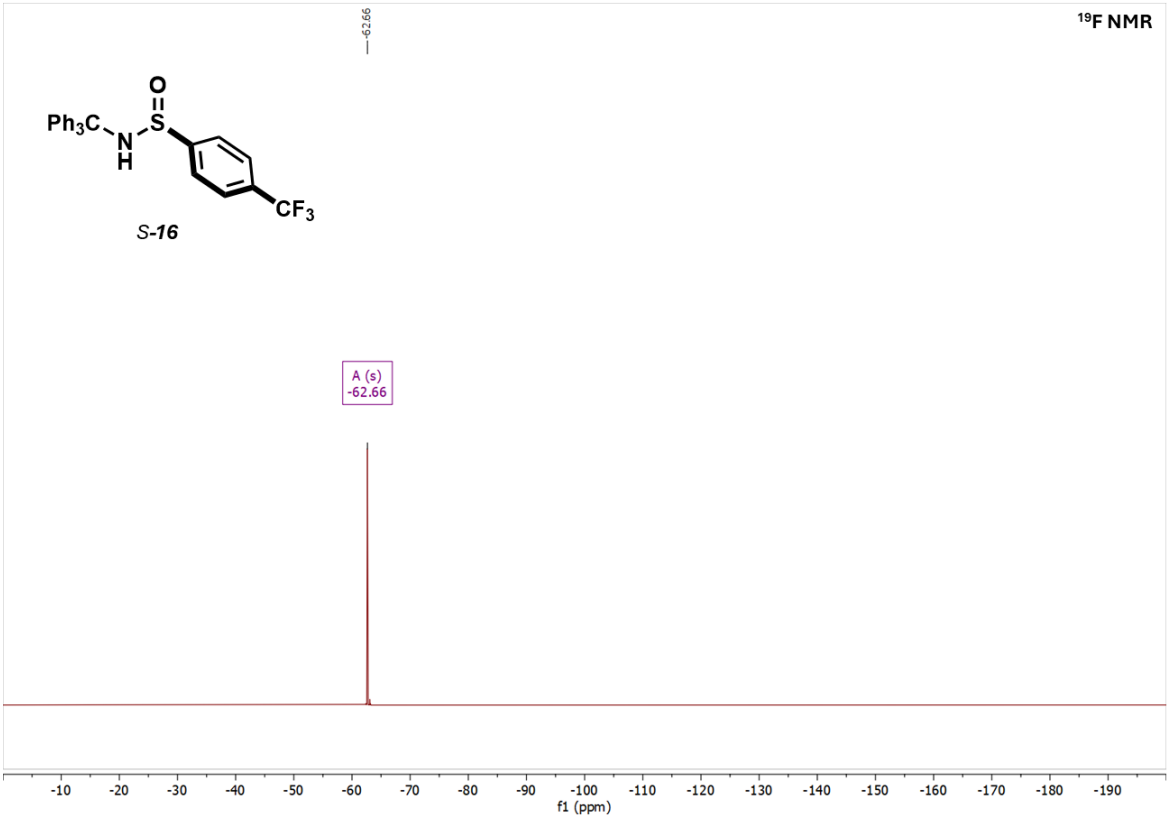


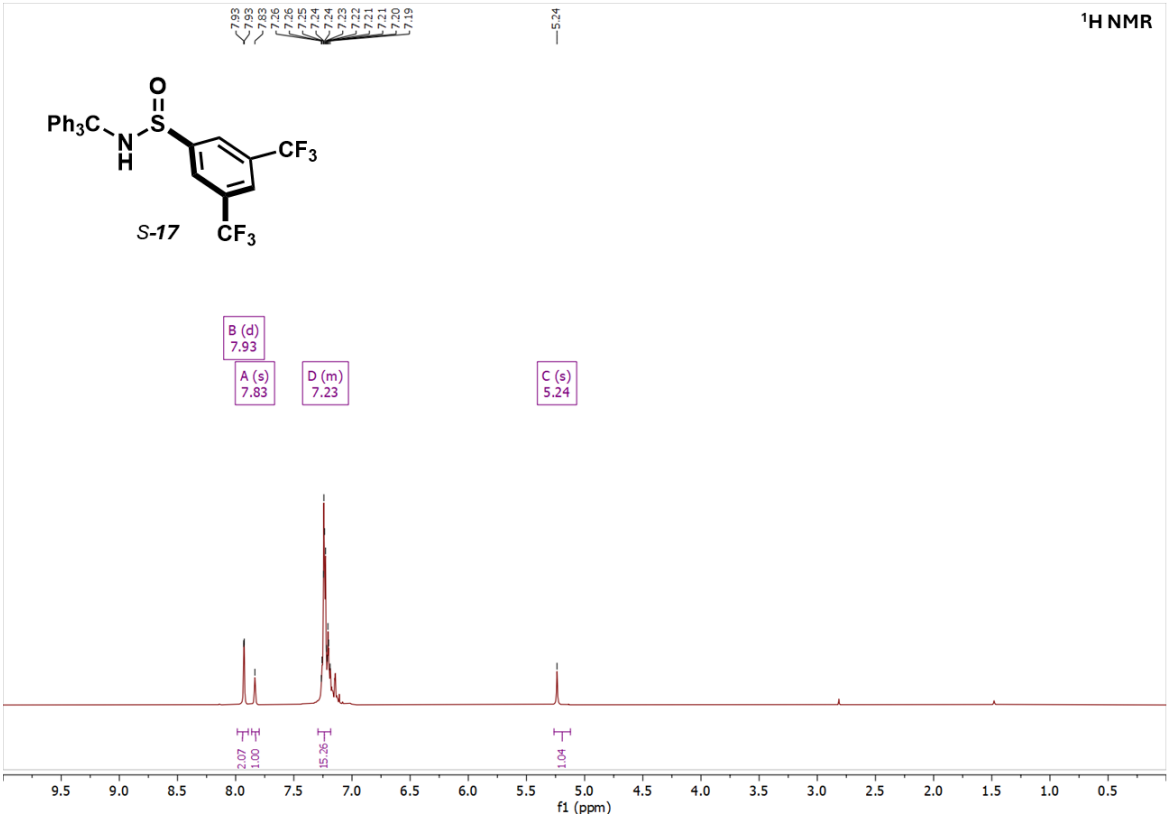

^{1}H NMR
Ph3C
N
H
O
S
S-15
1.97
3.02
5.92
6.00
3.06
0.99
f1 (ppm)

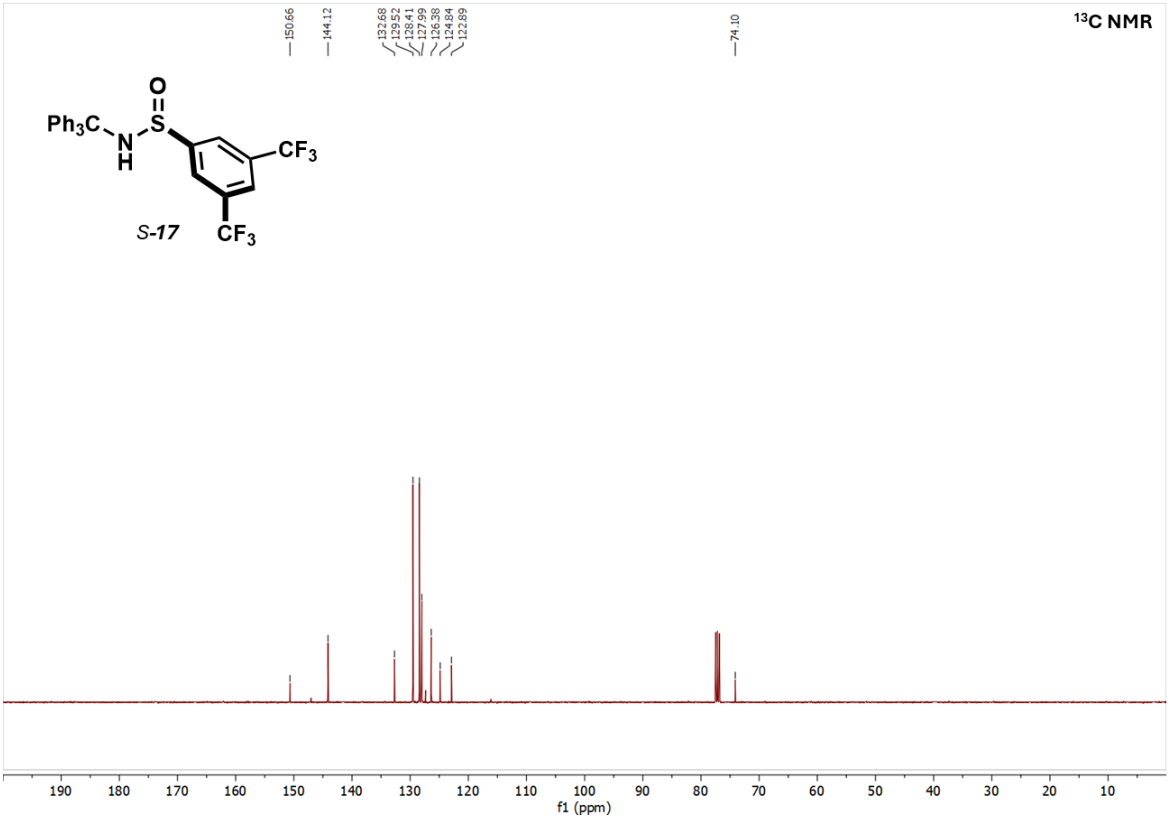

^{13}C NMR
147.28
144.64
130.83
129.59
129.01
128.06
127.51
125.65
Ph3C
N
H
O
S
S-15
f1 (ppm)

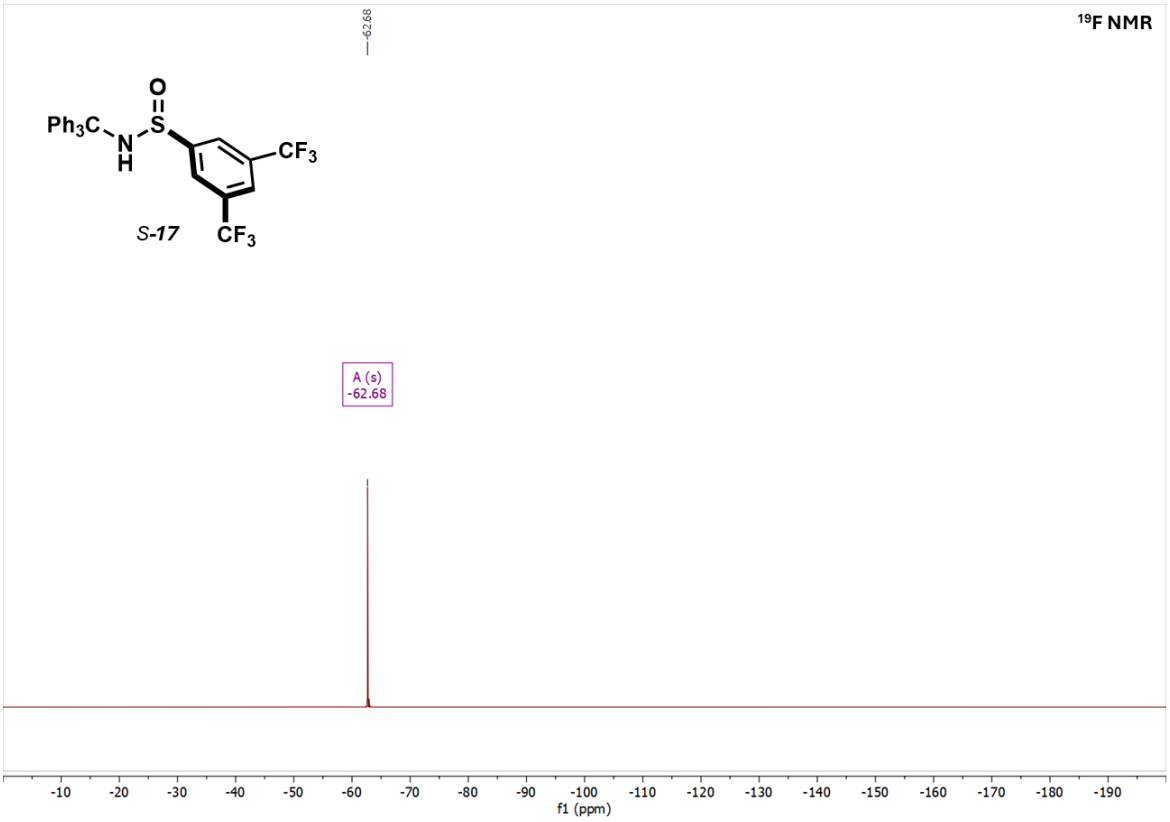

^{1}H NMR
7.81
7.79
7.75
7.73
7.38
7.37
7.37
7.36
7.35
7.35
7.34
7.33
7.33
7.32
7.32
7.31
5.28
Ph3C
N
H
O
S
CF3
S-16
C (d)
7.74
B (d)
7.80
D (m)
7.34
A (s)
5.28
2.00
1.97
15.19
1.13
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)


^{13}C NMR
151.42
144.44
133.04
129.65
128.29
127.81
126.44
126.14
123.75
73.95
Ph3C
N
H
O
S
CF3
S-16
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)

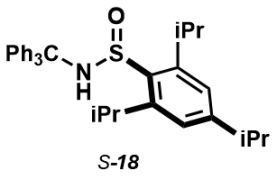

^{19}F NMR
-62.66
Ph3C
N
H
O
S
CF3
S-16
A (s)
-62.66
-10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)


^{1}H NMR
7.93
7.93
7.83
7.26
7.25
7.24
7.24
7.23
7.22
7.21
7.21
7.20
7.19
5.24
Ph3C
N
H
O
S
CF3
CF3
S-17
B (d)
7.93
A (s)
7.83
D (m)
7.23
C (s)
5.24
2.07
1.00
15.26
1.04
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

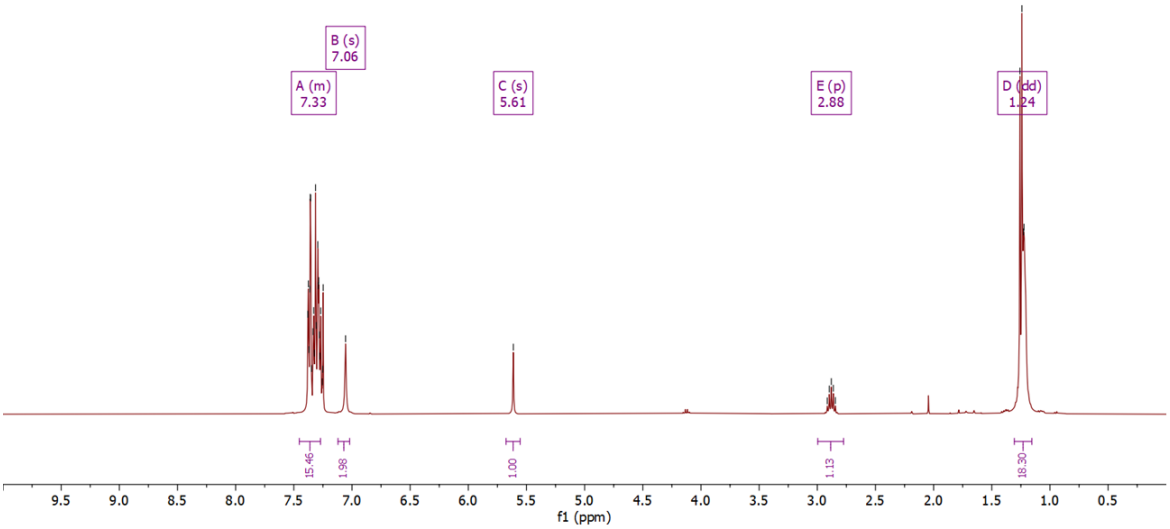

^{13}C NMR
150.66
144.12
132.68
129.52
128.41
127.99
126.38
124.84
122.89
74.10
Ph3C
N
H
S
O
CF3
CF3
S-17
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)


^{19}F NMR
-62.68
Ph3C
N
H
S
O
CF3
CF3
S-17
A (s)
-62.68
-10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)

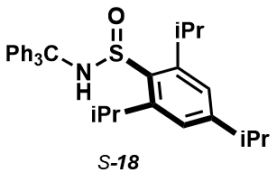
^{1}H NMR



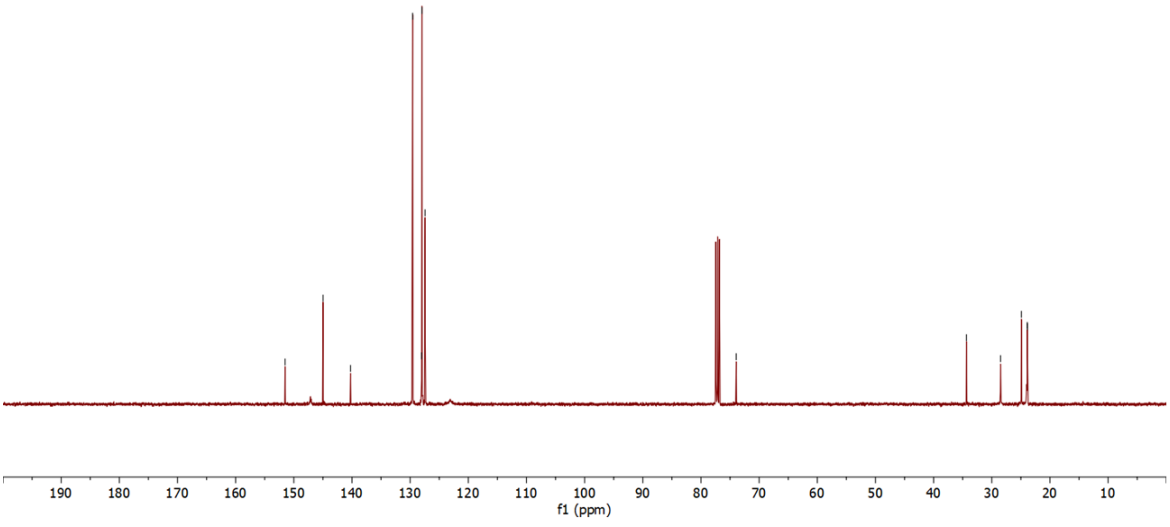
S-18

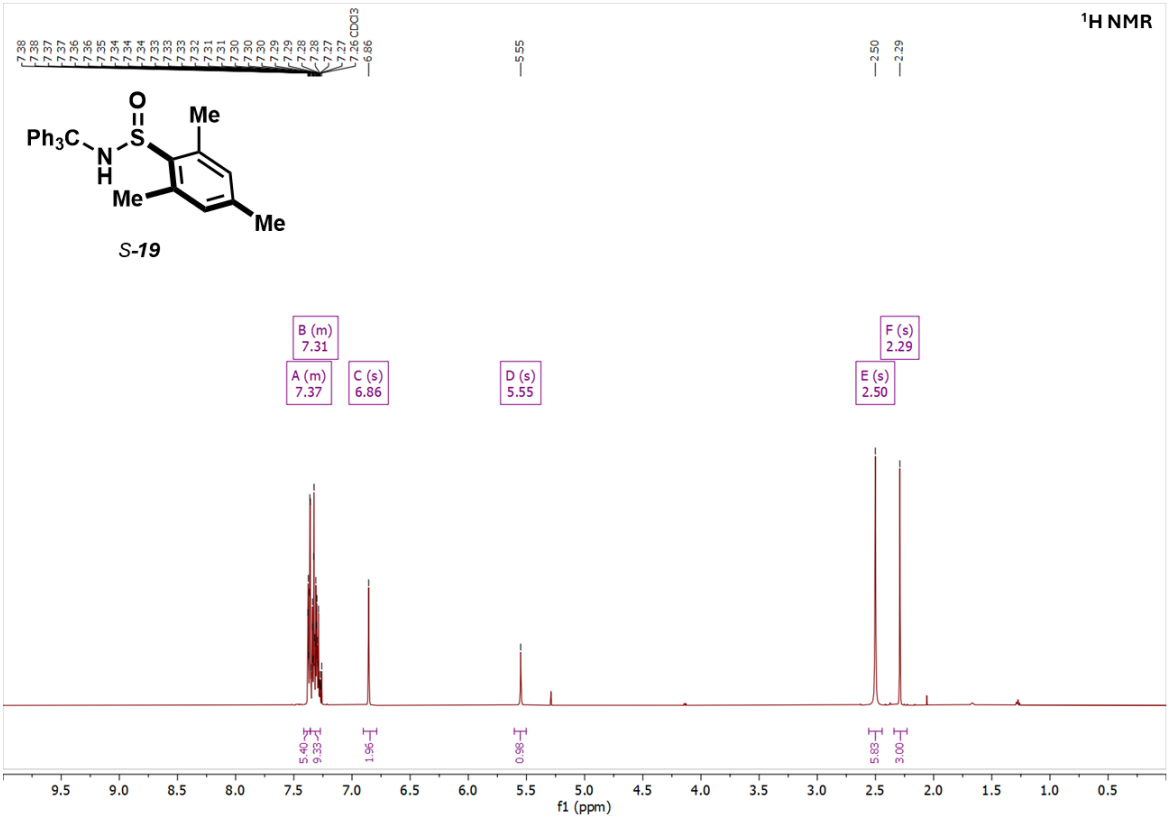


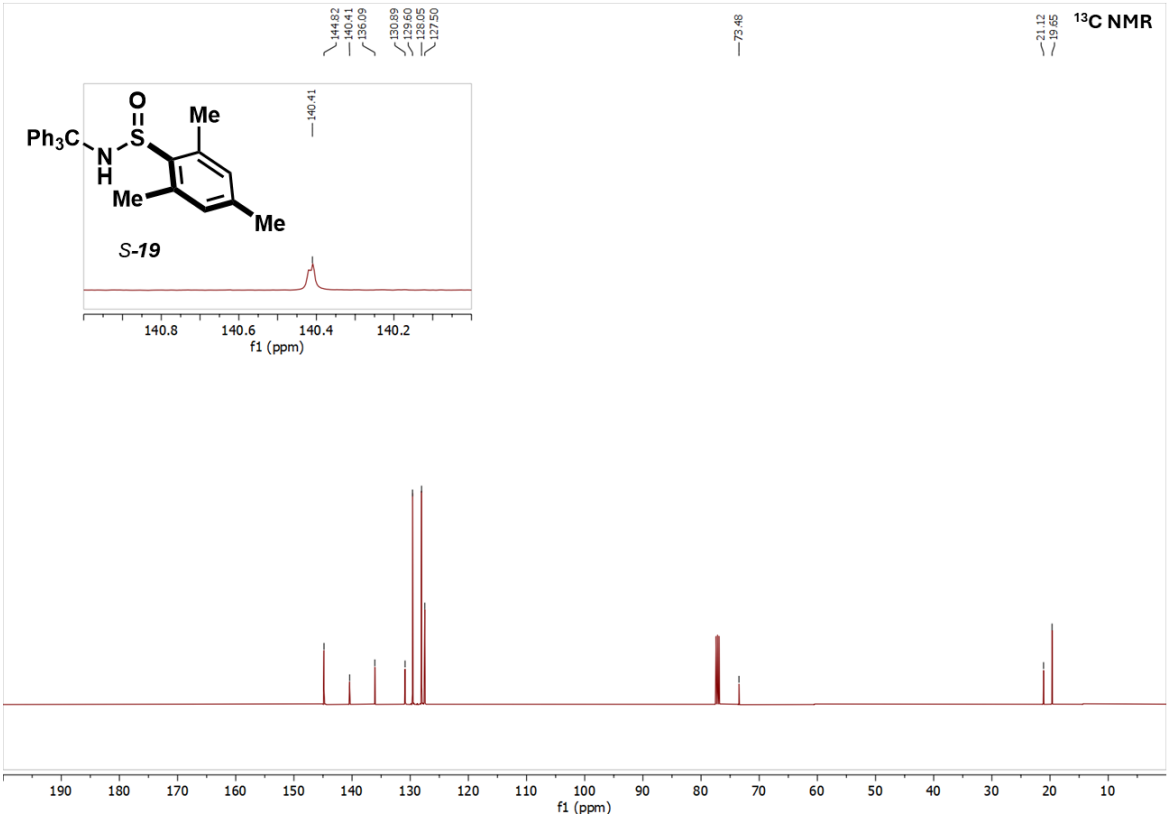
^{13}C NMR

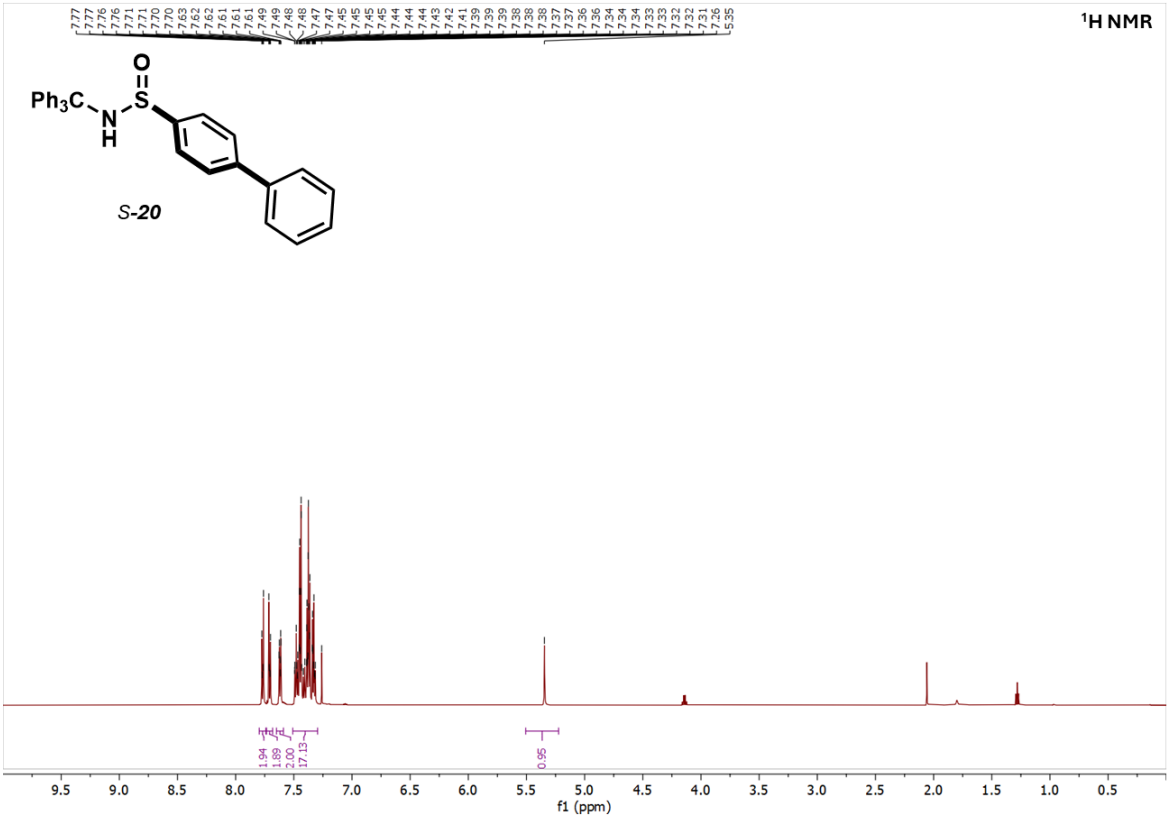


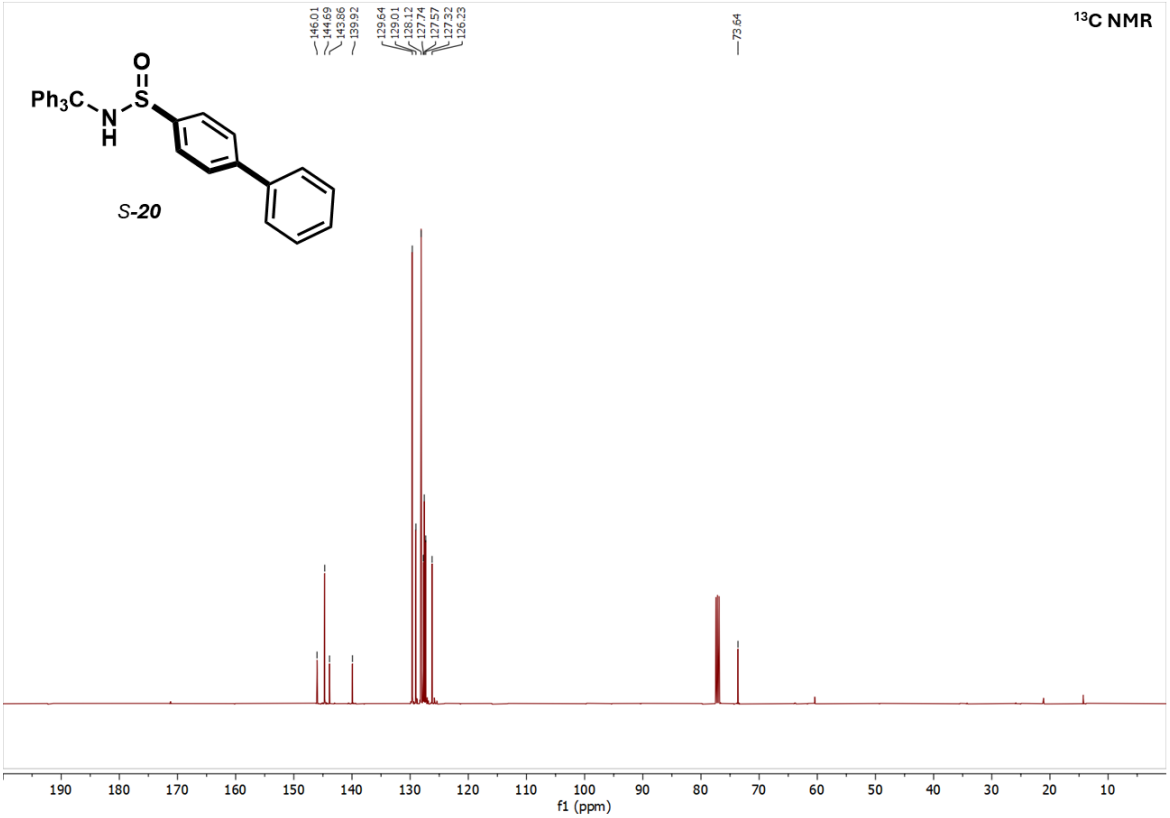
S-18

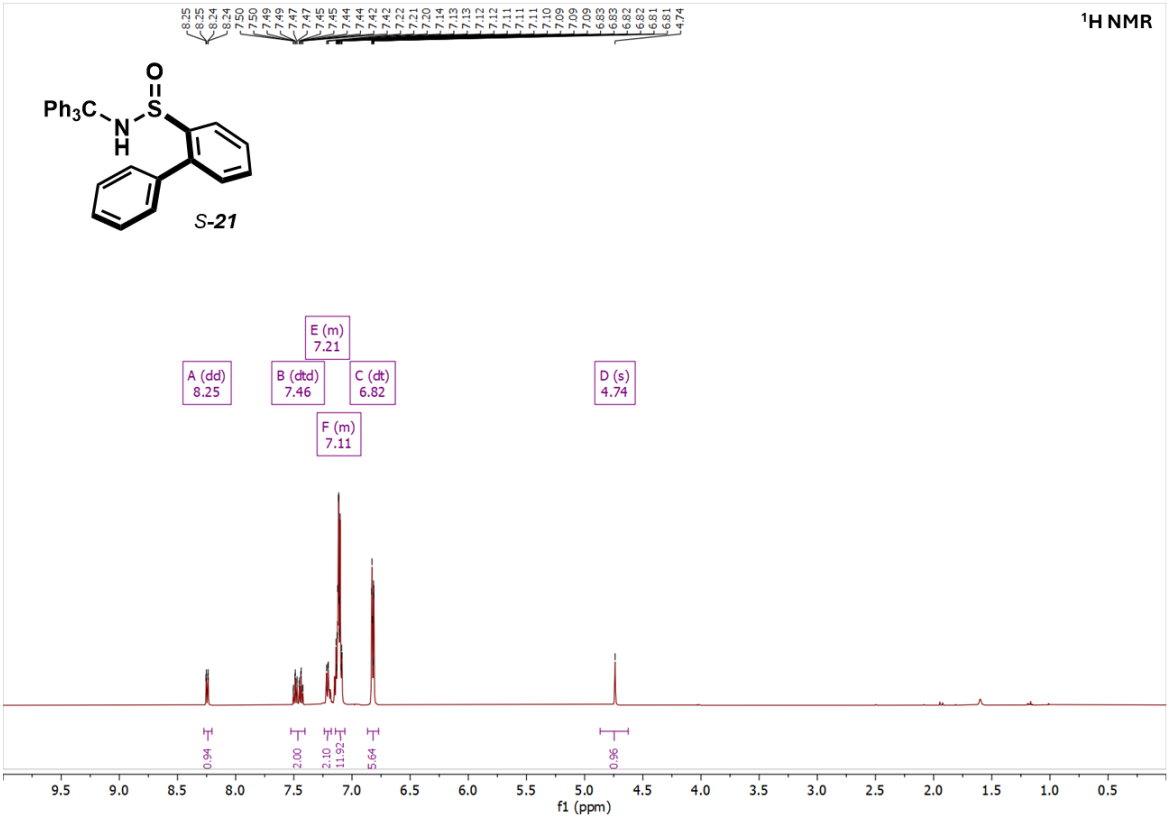


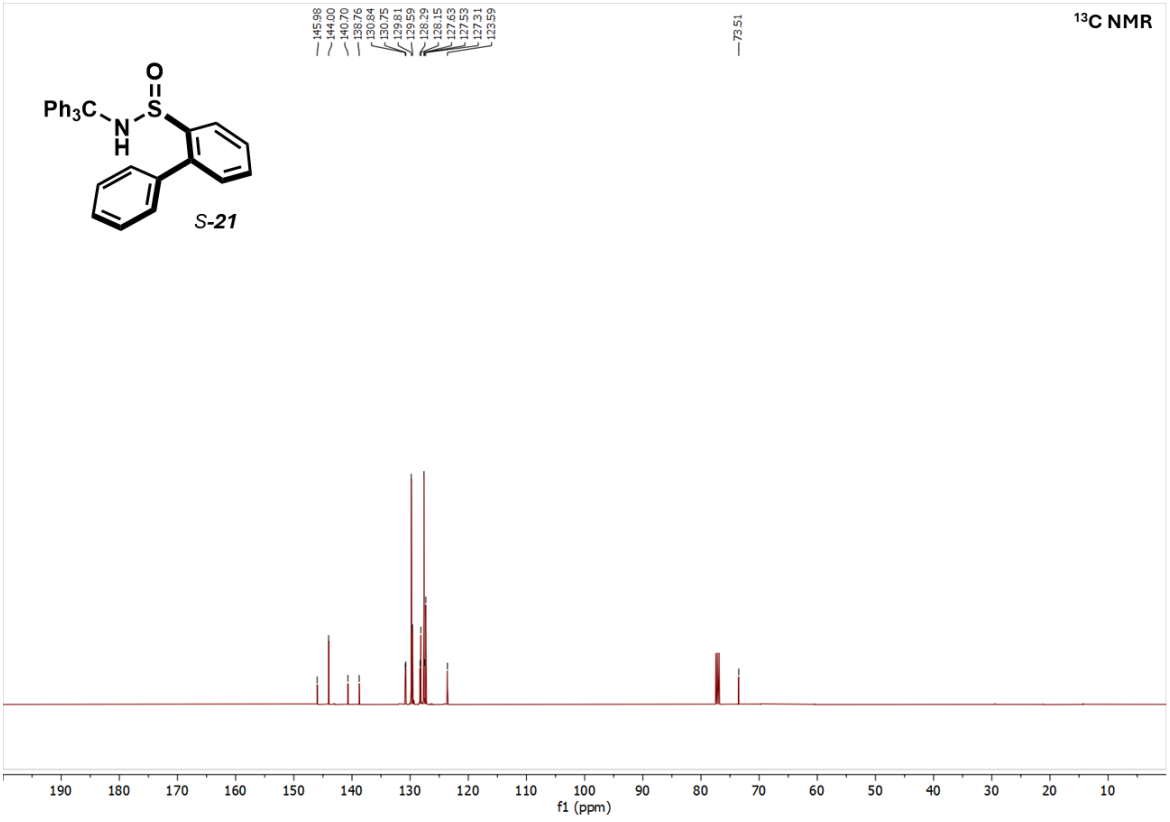

^{1}H NMR
S-19
B (m) 7.31
A (m) 7.37
C (s) 6.86
D (s) 5.55
F (s) 2.29
E (s) 2.50
f1 (ppm)

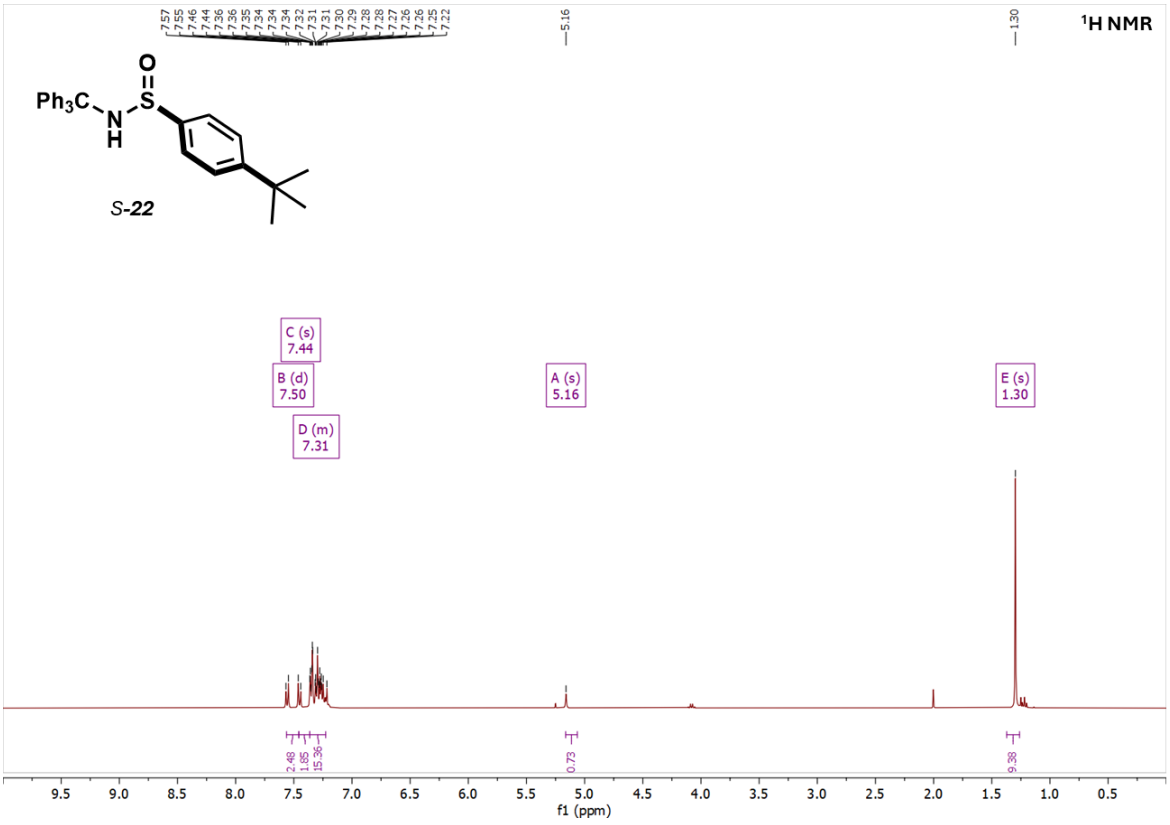

^{13}C NMR
S-19
f1 (ppm)

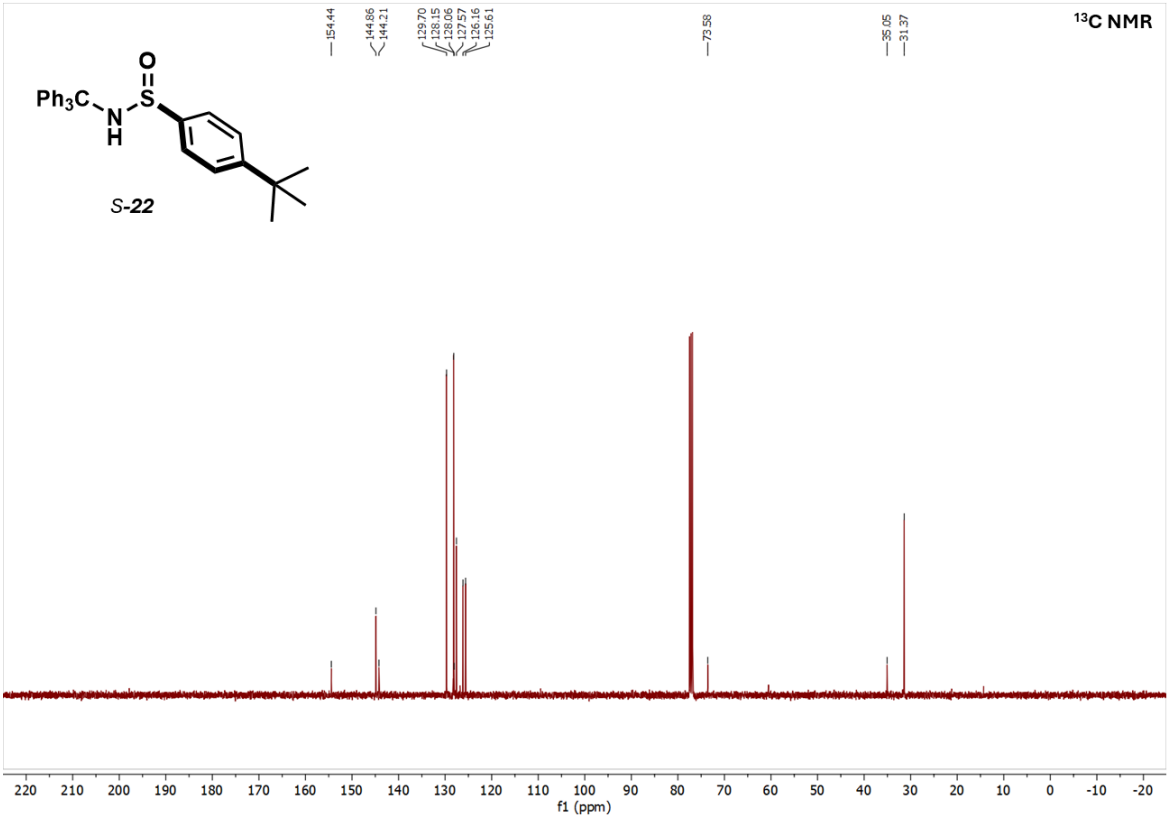

^{1}H NMR
Ph3C
N
H
O
S
S-20
1.94
1.89
2.00
17.13
0.95
f1 (ppm)

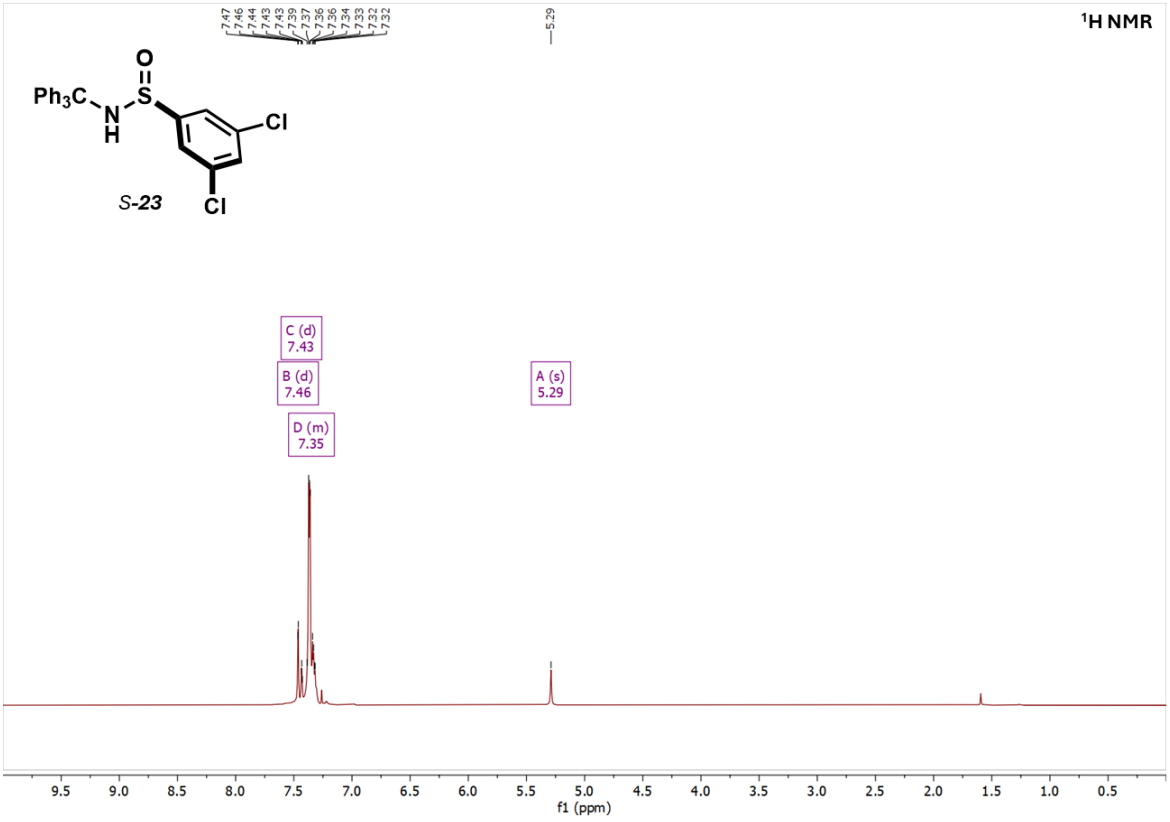

^{13}C NMR
Ph3C
N
H
O
S
S-20
146.01
144.69
143.86
139.92
129.64
129.01
128.12
127.74
127.57
127.32
126.23
73.64
f1 (ppm)

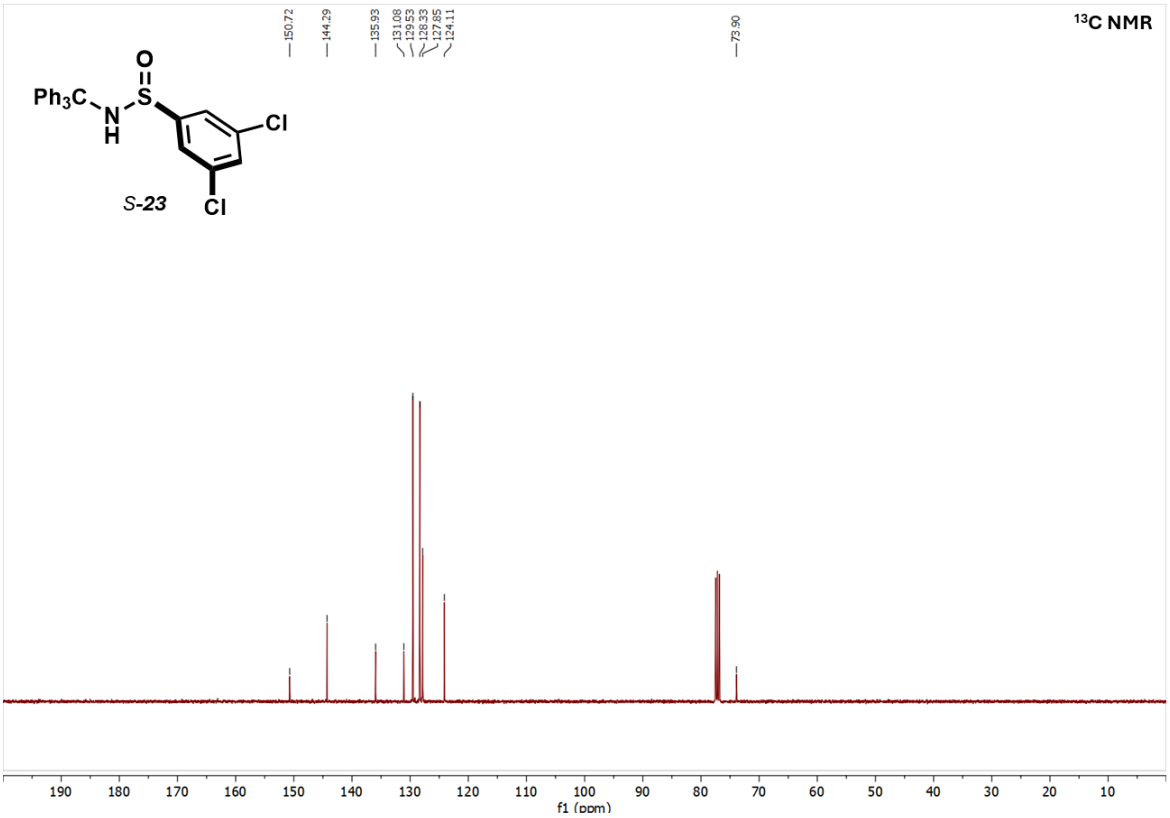

^{1}H NMR
Ph3C
N
H
O
S
S-21
A (dd) 8.25
B (dtd) 7.46
C (dt) 6.82
D (s) 4.74
E (m) 7.21
F (m) 7.11
0.94
2.00
2.10
11.92
5.64
0.96
f1 (ppm)

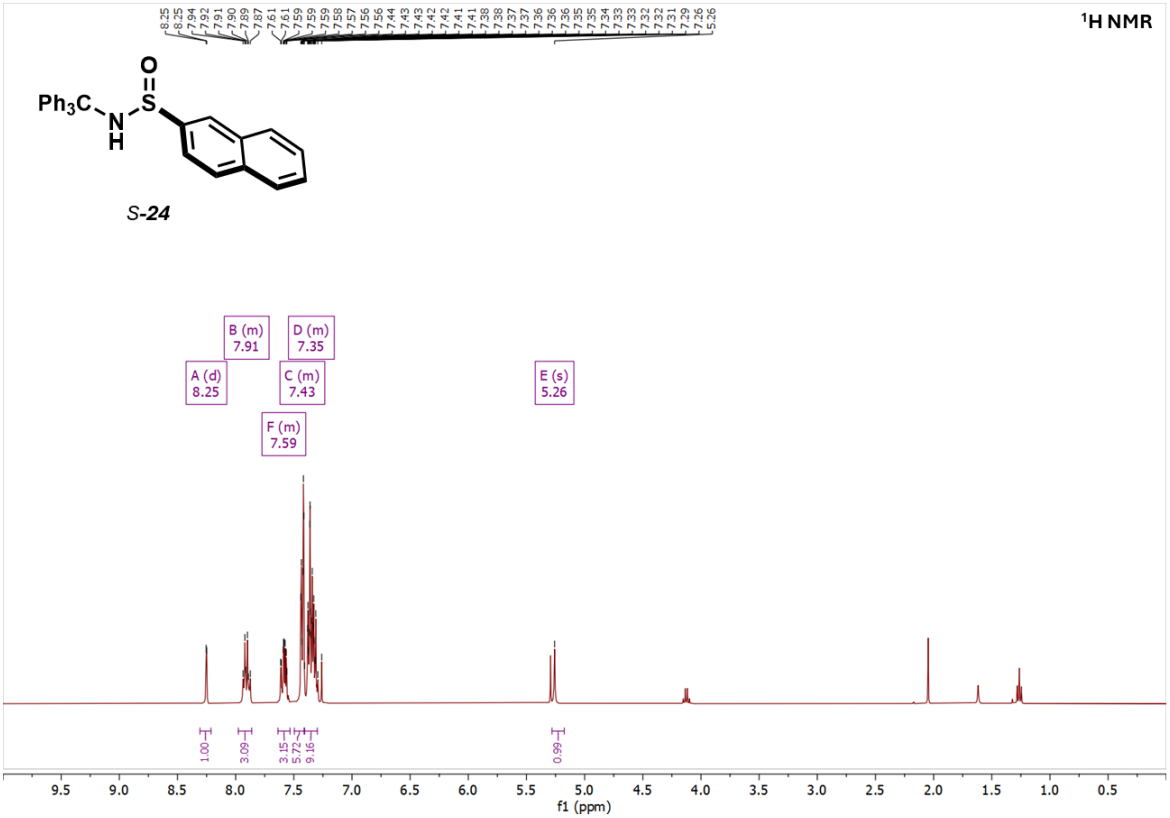

^{13}C NMR
Ph3C
N
H
O
S
S-21
f1 (ppm)

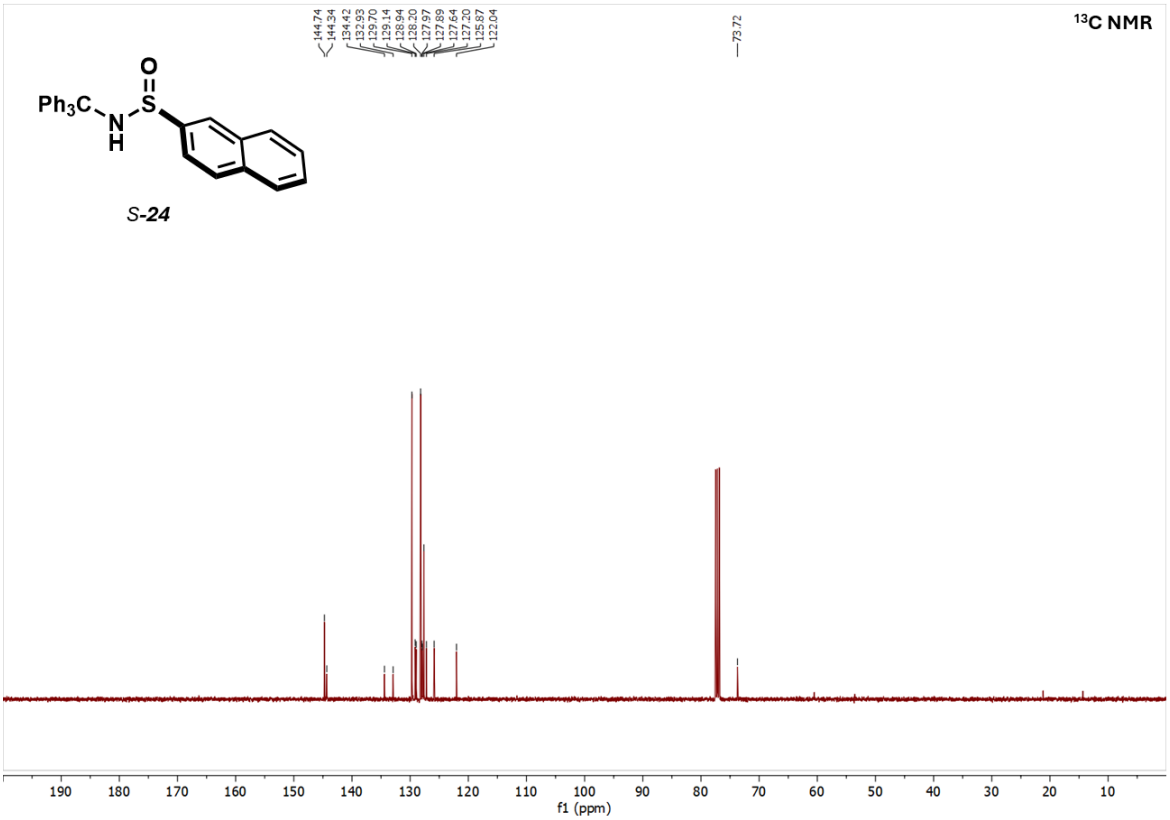

^{1}H NMR
S-22
C (s) 7.44
B (d) 7.50
D (m) 7.31
A (s) 5.16
E (s) 1.30
f1 (ppm)

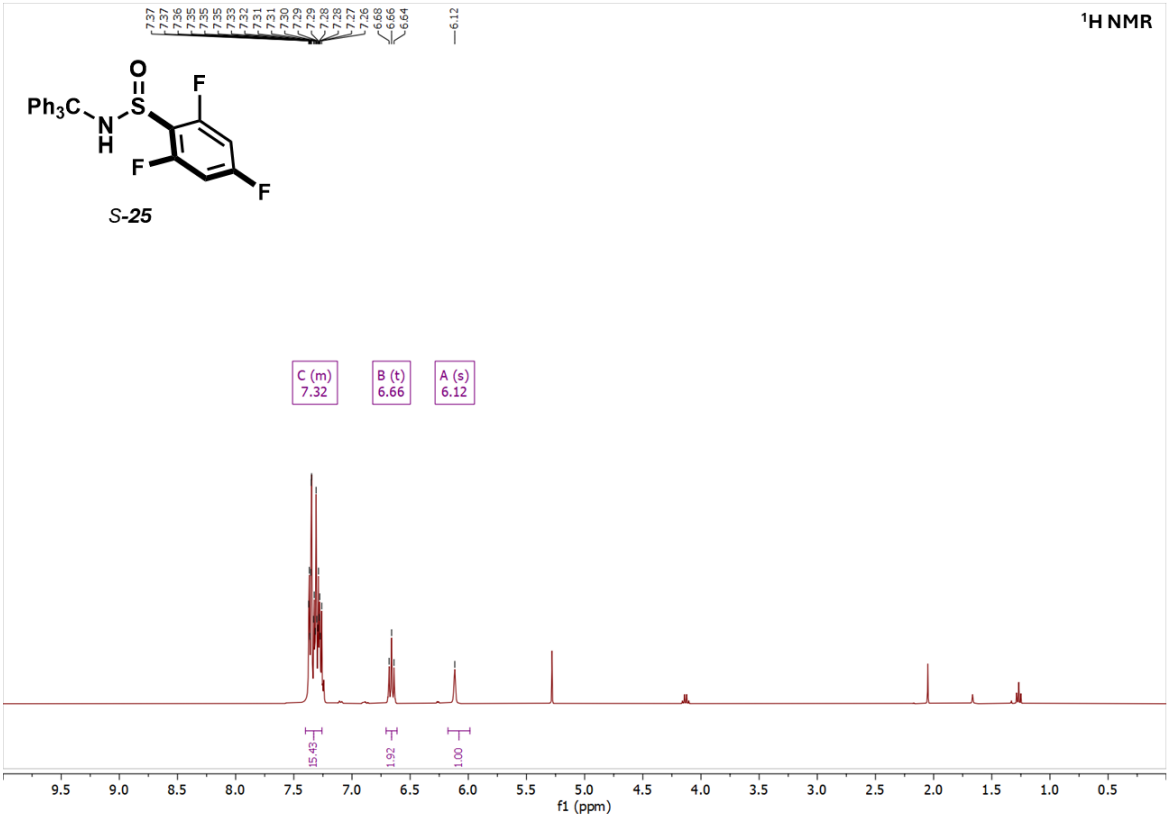

^{13}C NMR
S-22
f1 (ppm)

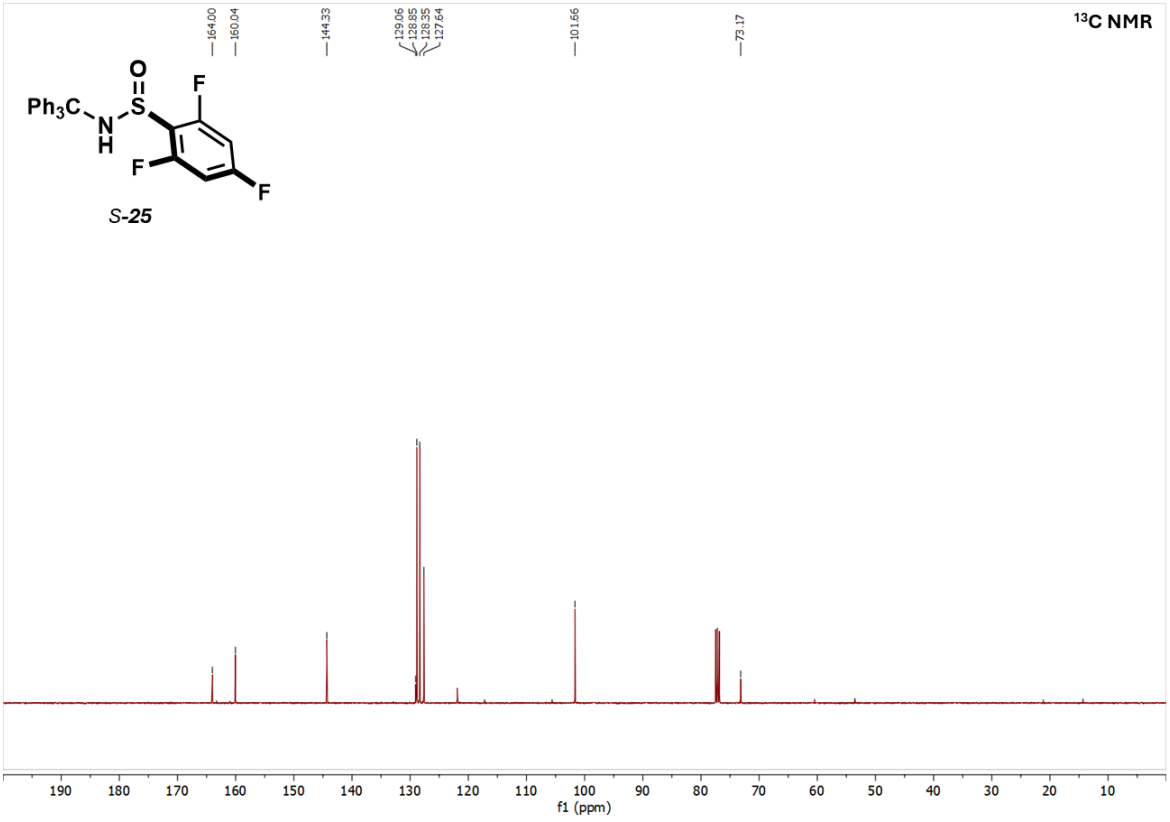

^{1}H NMR
7.47
7.46
7.44
7.43
7.43
7.39
7.37
7.36
7.36
7.34
7.33
7.32
7.32
5.29
Ph3C
N
H
O
S
Cl
Cl
S-23
C (d) 7.43
B (d) 7.46
D (m) 7.35
A (s) 5.29
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

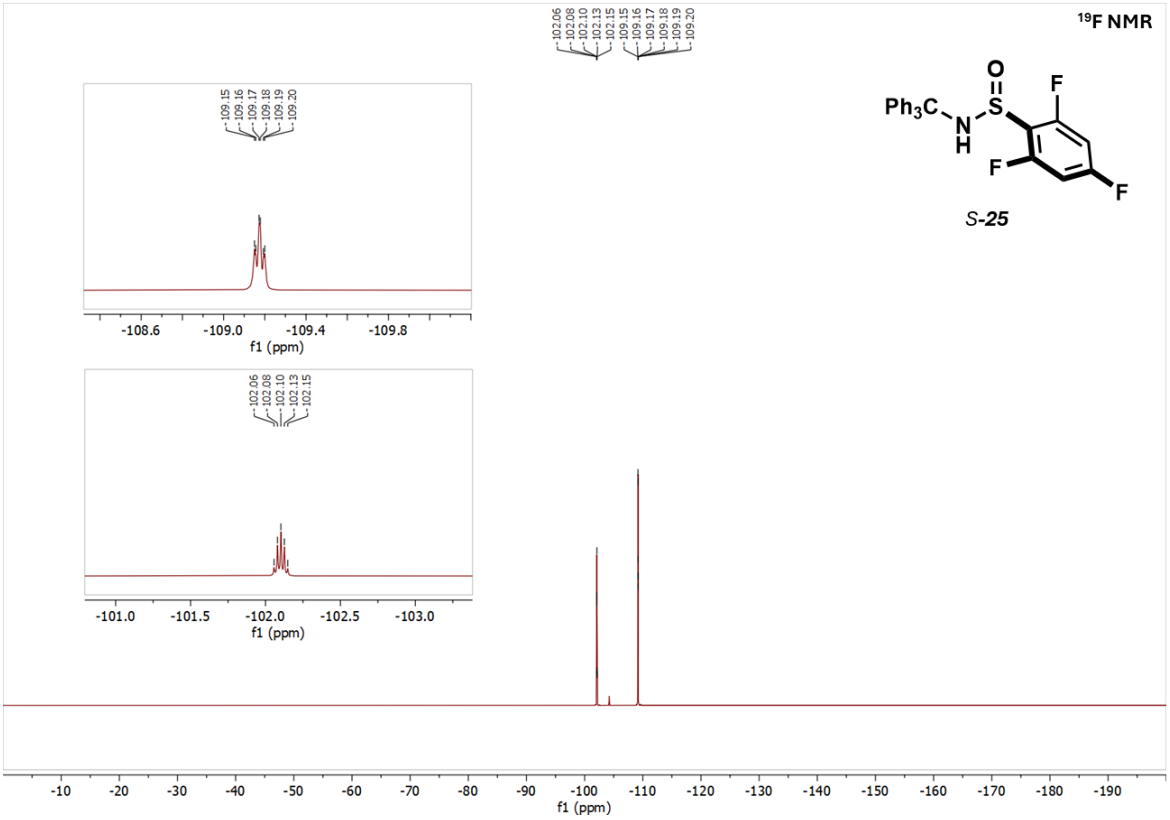

^{13}C NMR
150.72
144.29
135.93
131.08
129.53
128.33
127.85
124.11
73.90
Ph3C
N
H
O
S
Cl
Cl
S-23
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

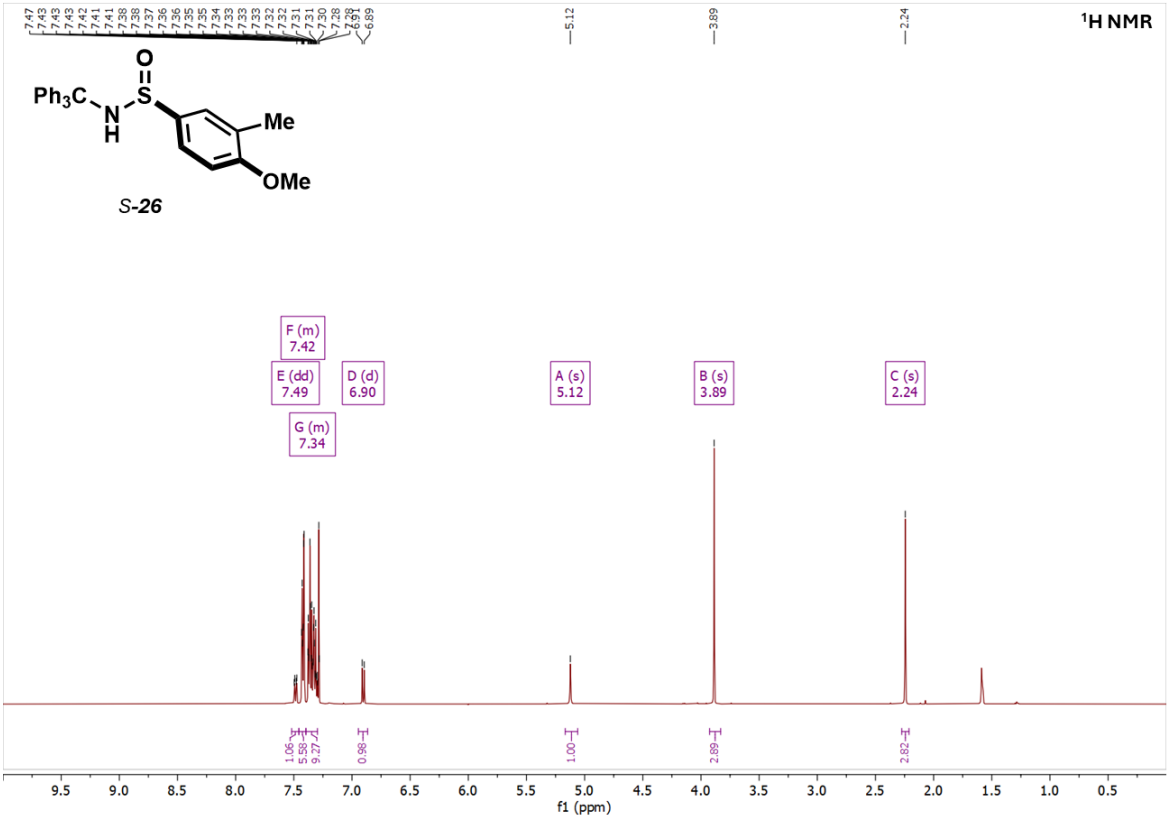

^{1}H NMR
Ph3C
N
H
O
S
S-24
B (m) 7.91
D (m) 7.35
A (d) 8.25
C (m) 7.43
E (s) 5.26
F (m) 7.59
1.00
3.09
3.15
5.72
9.16
0.99
f1 (ppm)

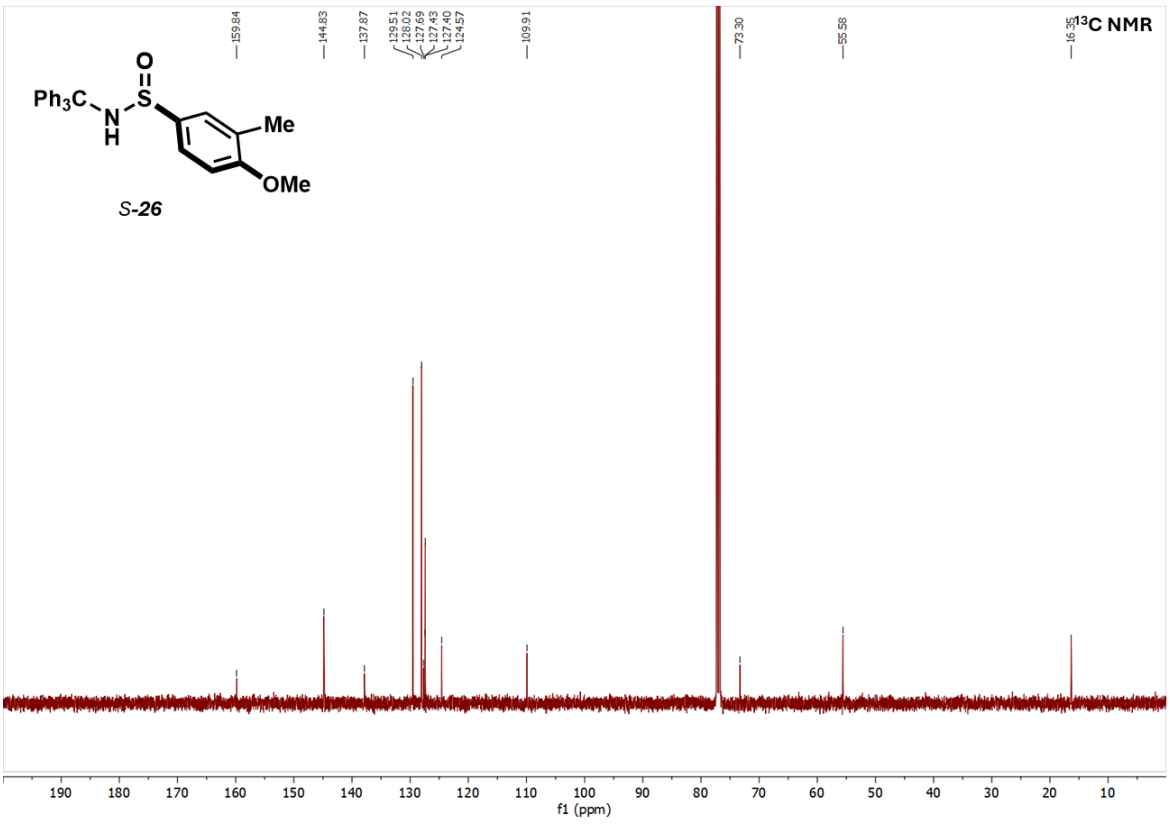

^{13}C NMR
Ph3C
N
H
O
S
S-24
73.72
f1 (ppm)

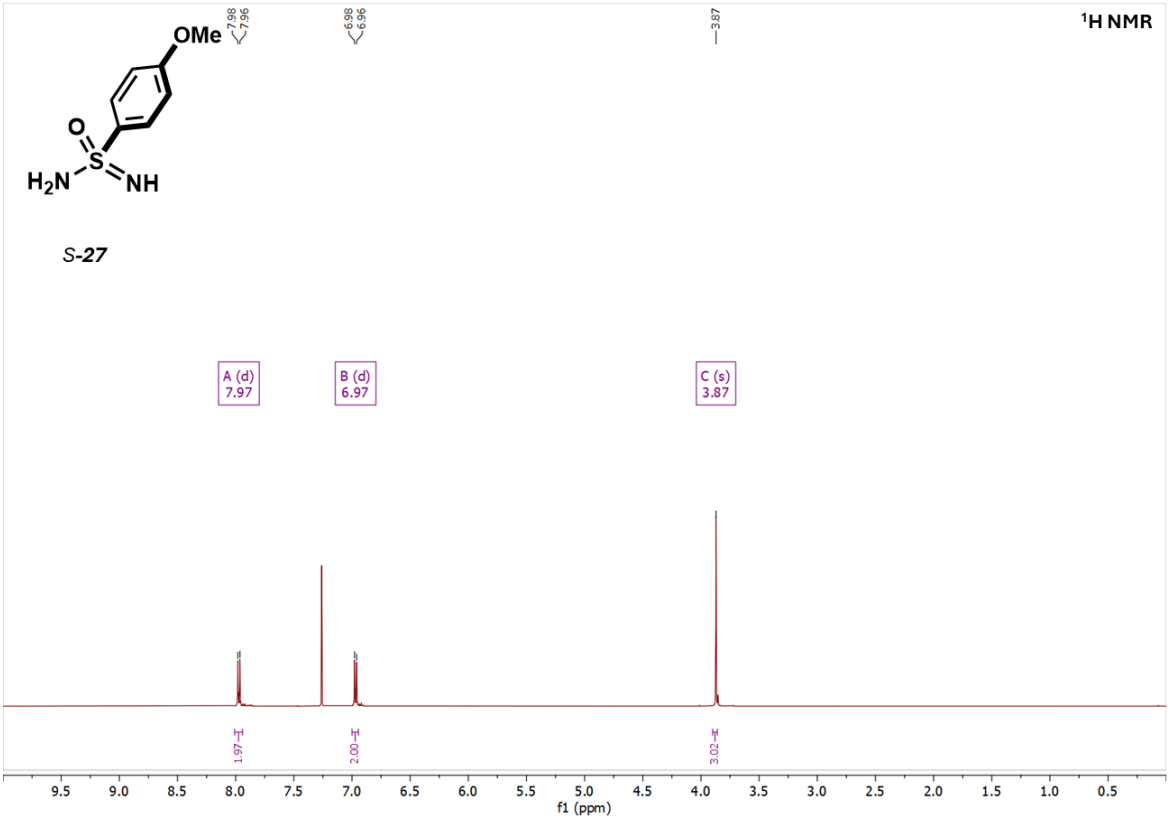

¹H NMR
S-25
C (m) 7.32
B (t) 6.66
A (s) 6.12
15.43
1.92
1.00
f1 (ppm)

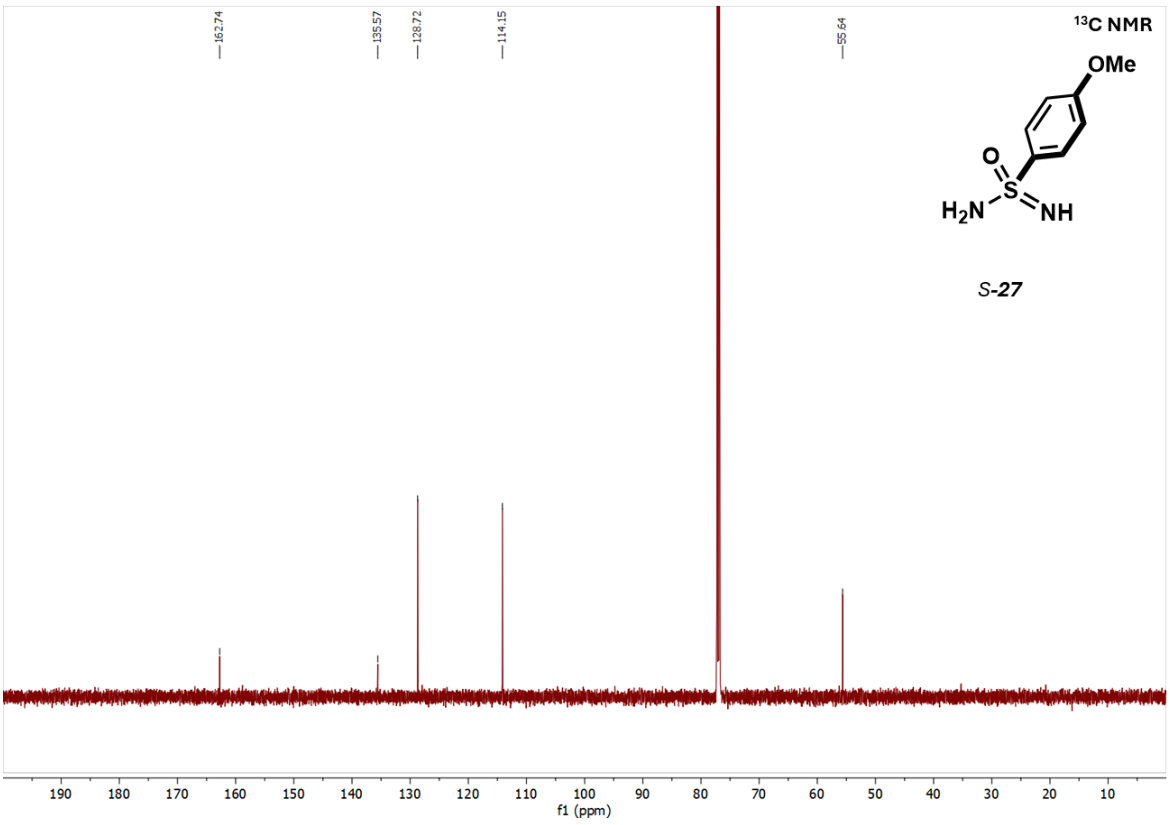

¹³C NMR
S-25
164.00
160.04
144.33
129.06
128.85
128.35
127.64
101.66
73.17
f1 (ppm)

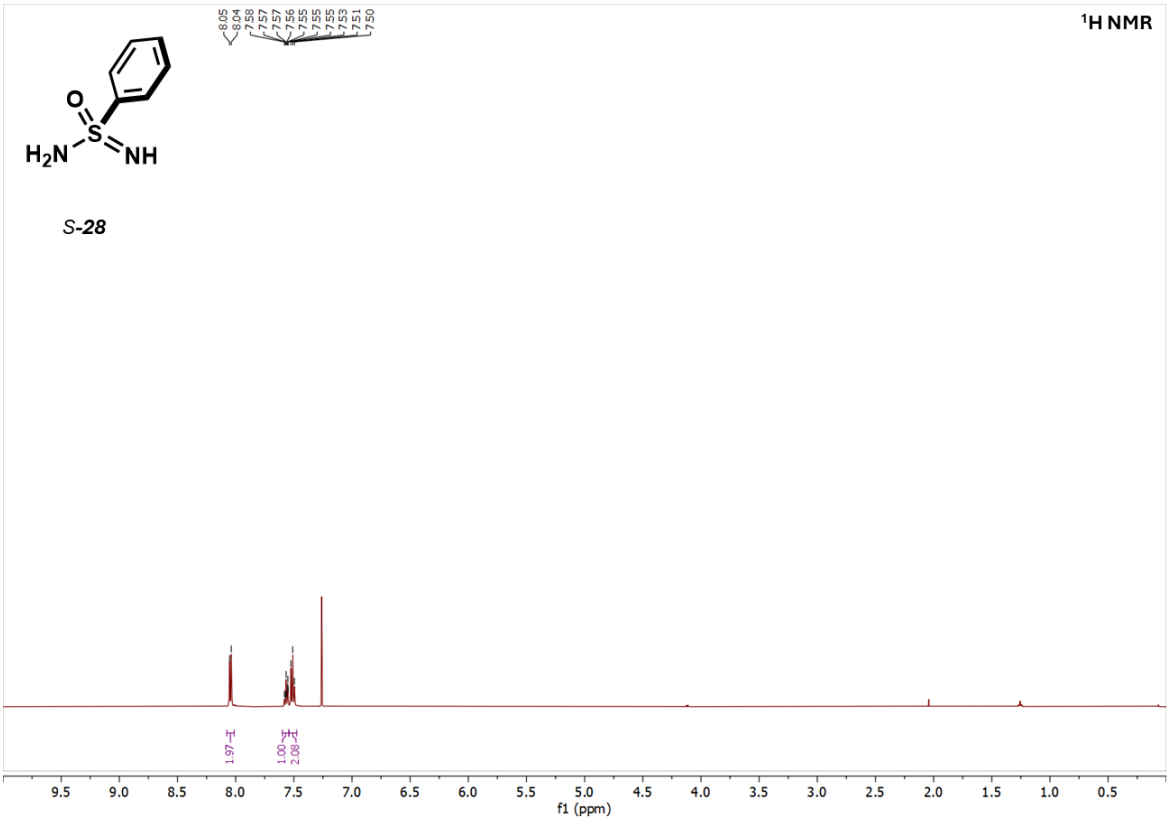

19F NMR
S-25
f1 (ppm)

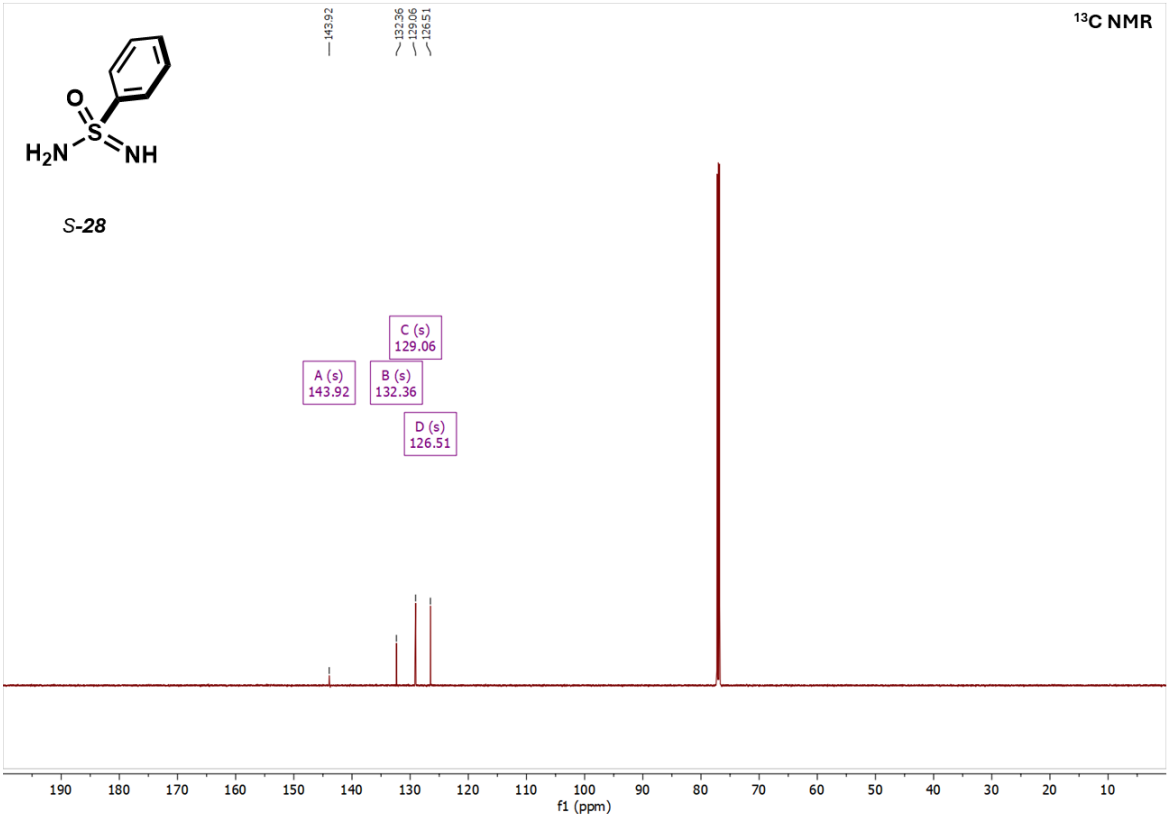

1H NMR
S-26
F (m) 7.42
E (dd) 7.49
D (d) 6.90
G (m) 7.34
A (s) 5.12
B (s) 3.89
C (s) 2.24
f1 (ppm)

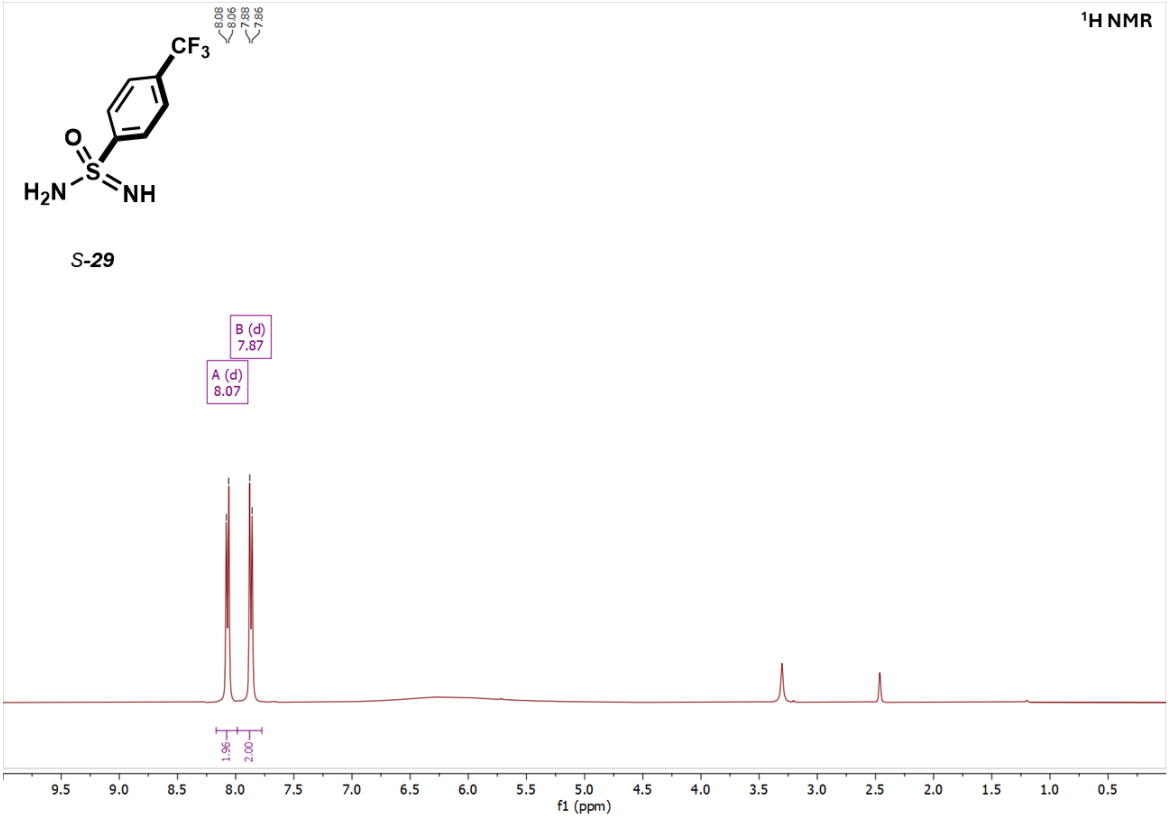

159.84
144.83
137.87
129.51
128.02
127.69
127.43
127.40
124.57
109.91
73.30
55.58
16.35
13C NMR
Ph3C
N
H
S
O
Me
OMe
S-26
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

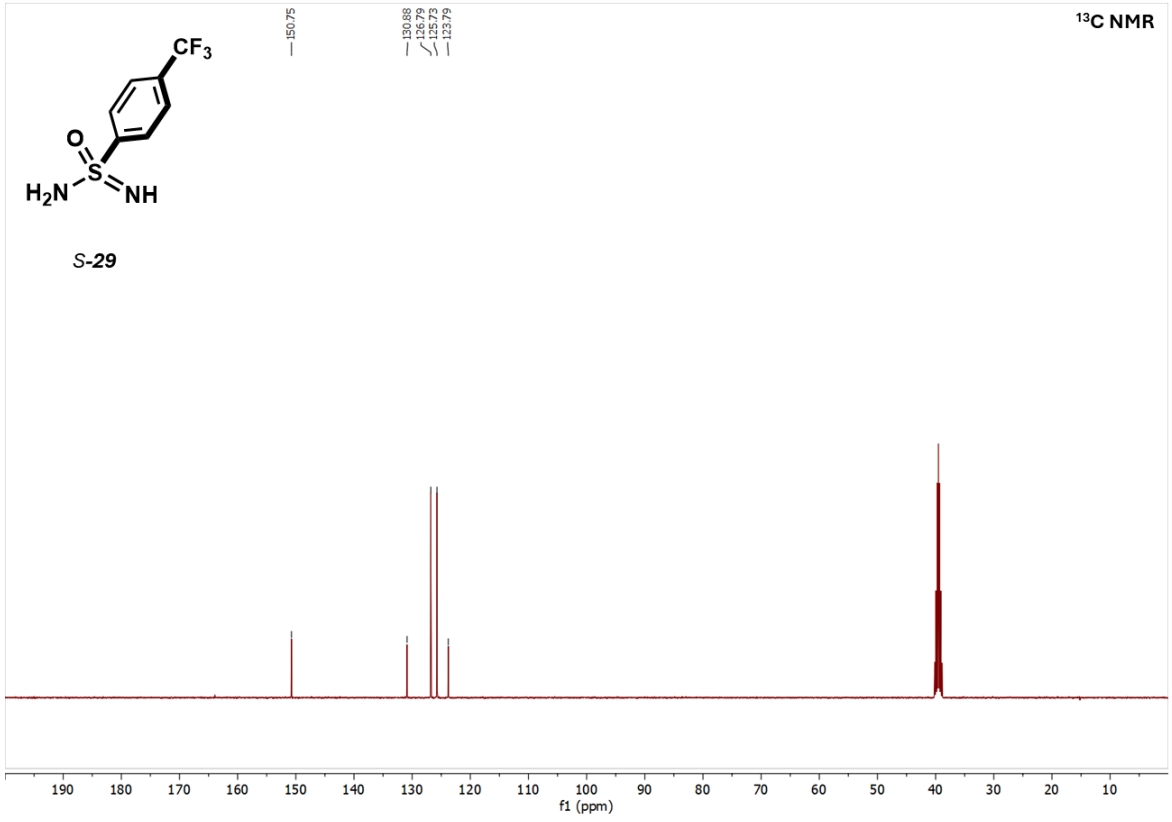

7.98
7.96
6.98
6.96
3.87
1H NMR
OMe
O
S
H2N
NH
S-27
A (d)
7.97
B (d)
6.97
C (s)
3.87
1.97
2.00
3.02
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

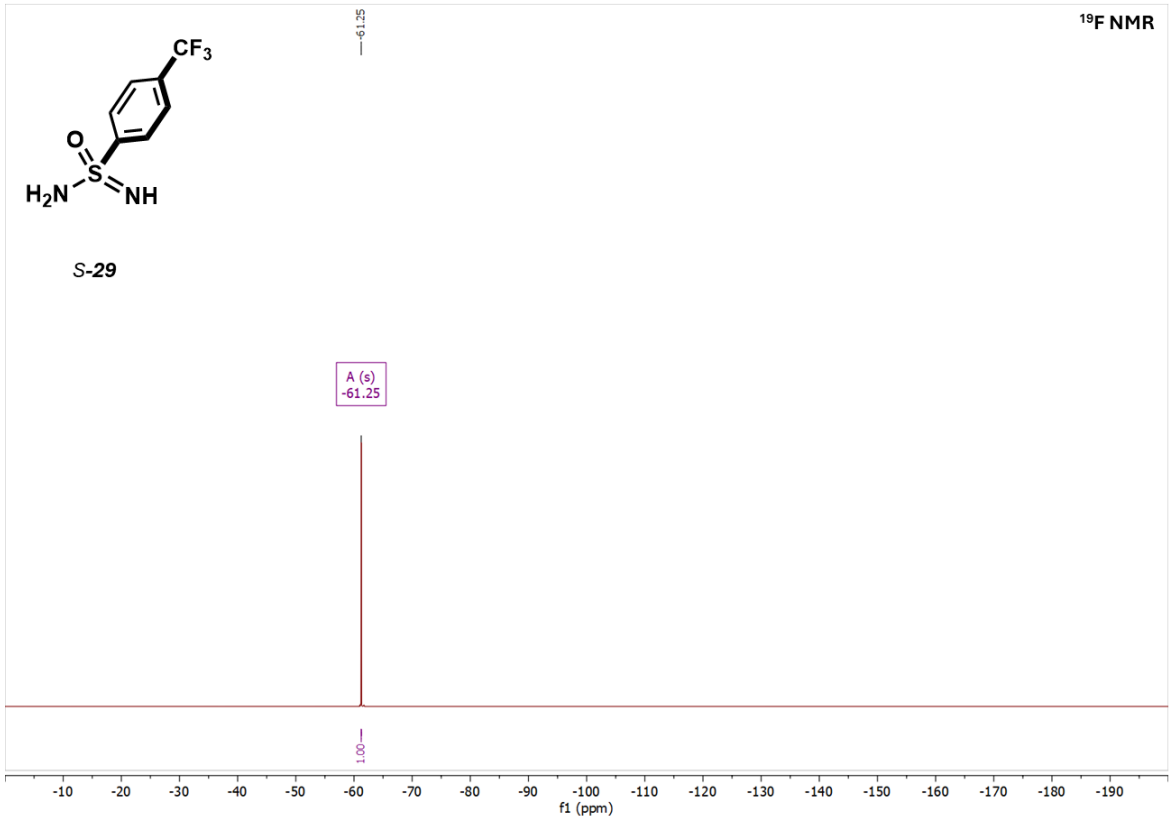

162.74
135.57
128.72
114.15
55.64
13C NMR
OMe
O
S
H2N
NH
S-27
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)

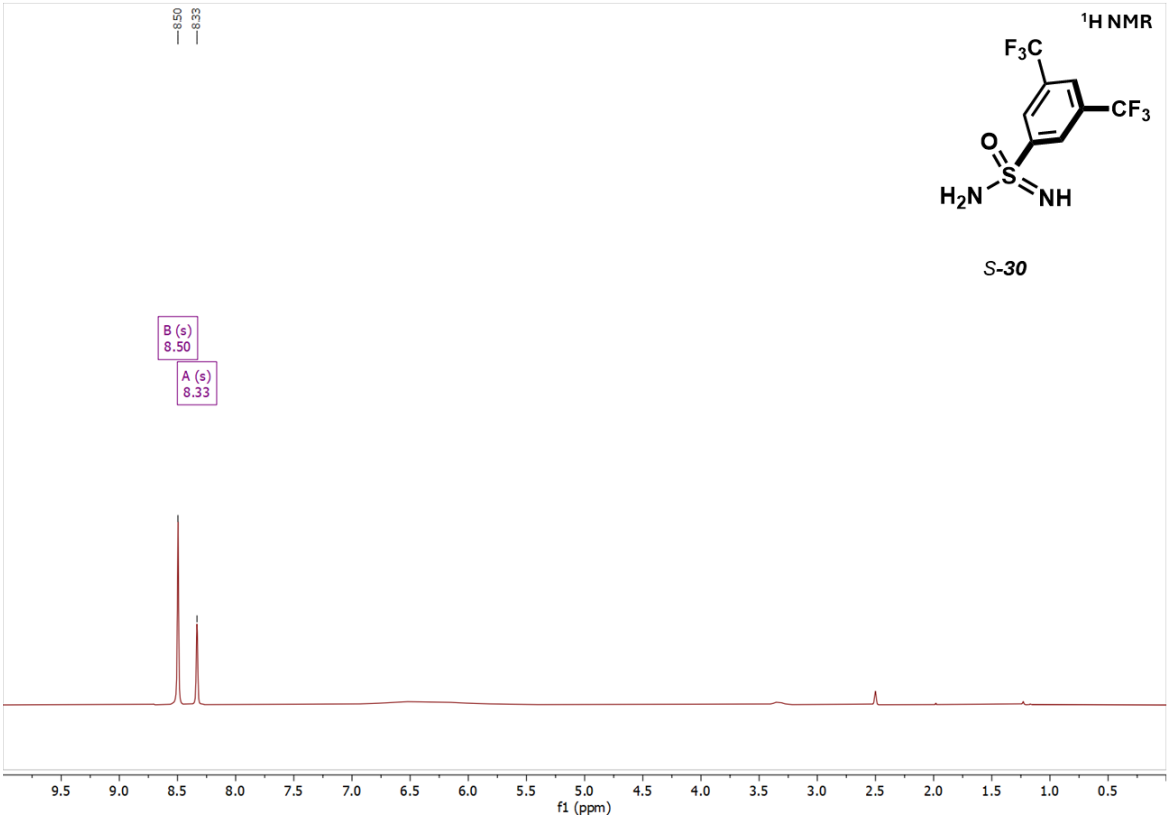

8.05
8.04
7.58
7.57
7.57
7.56
7.55
7.55
7.55
7.53
7.51
7.50
1H NMR
O
S
H2N
NH
S-28
1.97
1.00
2.08
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

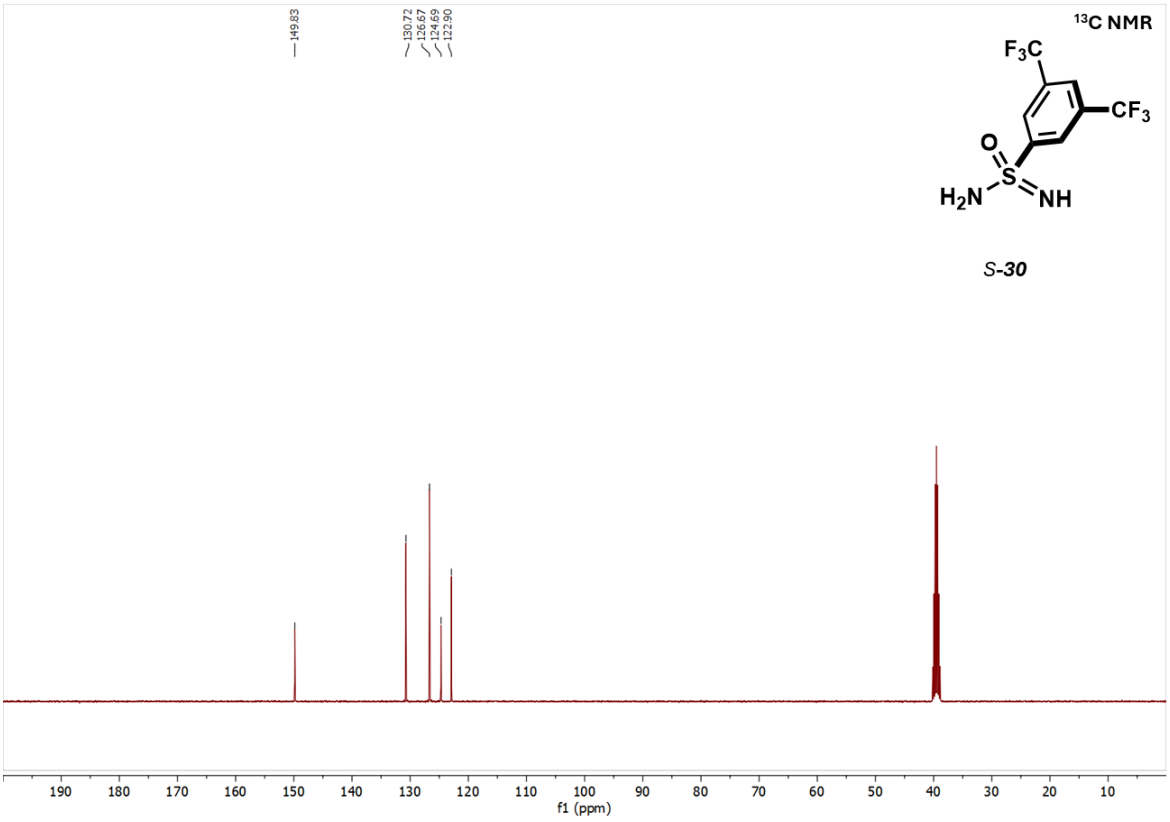

^{13}C NMR
143.92
132.36
129.06
126.51
S-28
C (s) 129.06
A (s) 143.92
B (s) 132.36
D (s) 126.51
f1 (ppm)

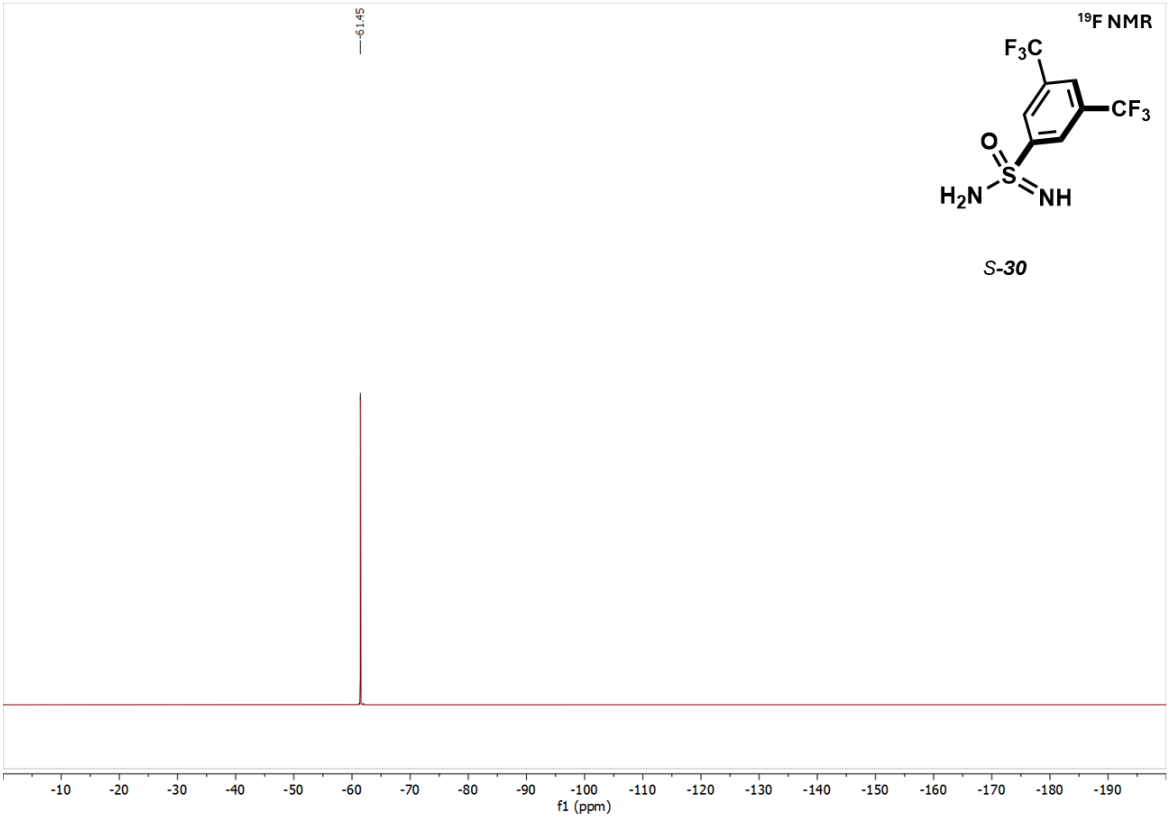

^{1}H NMR
8.08
8.06
7.88
7.86
S-29
B (d) 7.87
A (d) 8.07
1.96
2.00
f1 (ppm)


^{13}C NMR
150.75
130.88
126.79
125.73
123.79
CF3
O
S
H2N
NH
S-29
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

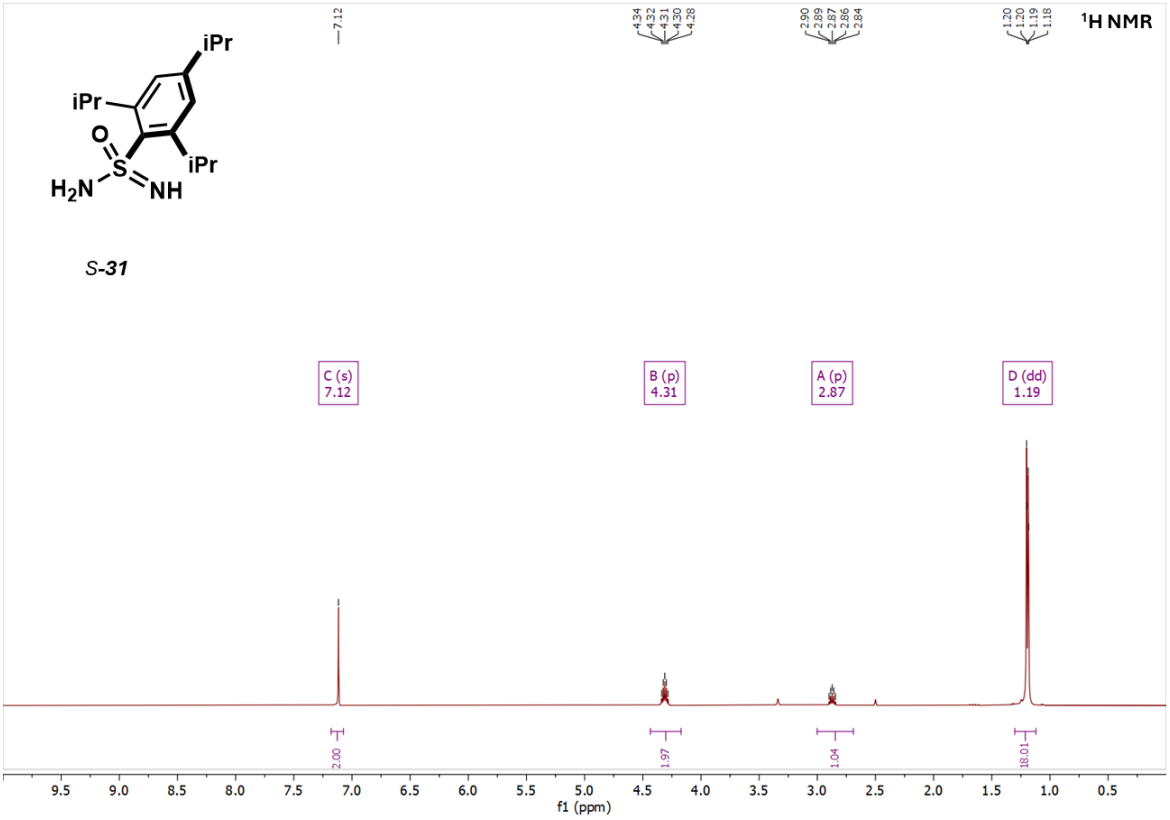

^{19}F NMR
-61.25
CF3
O
S
H2N
NH
S-29
A (s)
-61.25
1.00
-10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)


8.50
8.33
^{1}H NMR
F_3C
CF_3
O
S
H_2N
NH
S-30
B (s)
8.50
A (s)
8.33
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

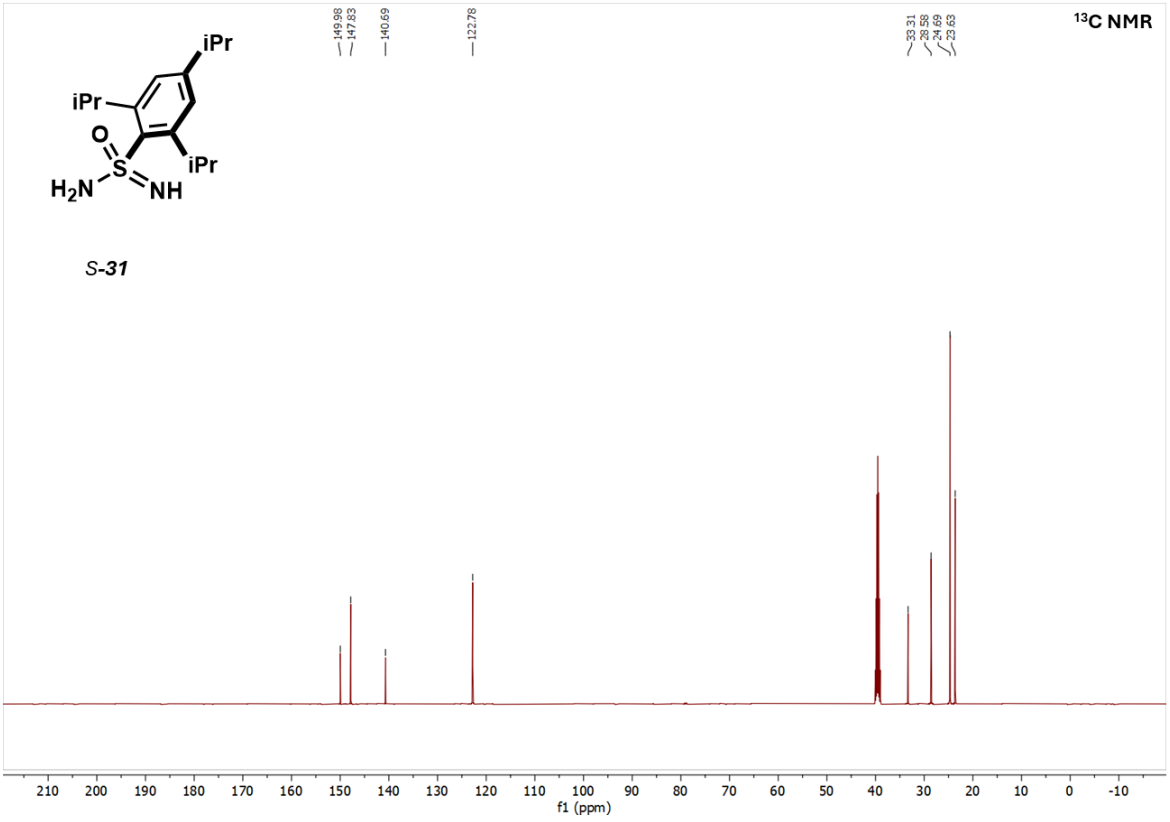

149.83
130.72
126.67
124.69
122.90
^{13}C NMR
F_3C
CF_3
O
S
H_2N
NH
S-30
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)


19F NMR
-61.45
F3C
CF3
O
S
H2N
NH
-10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)

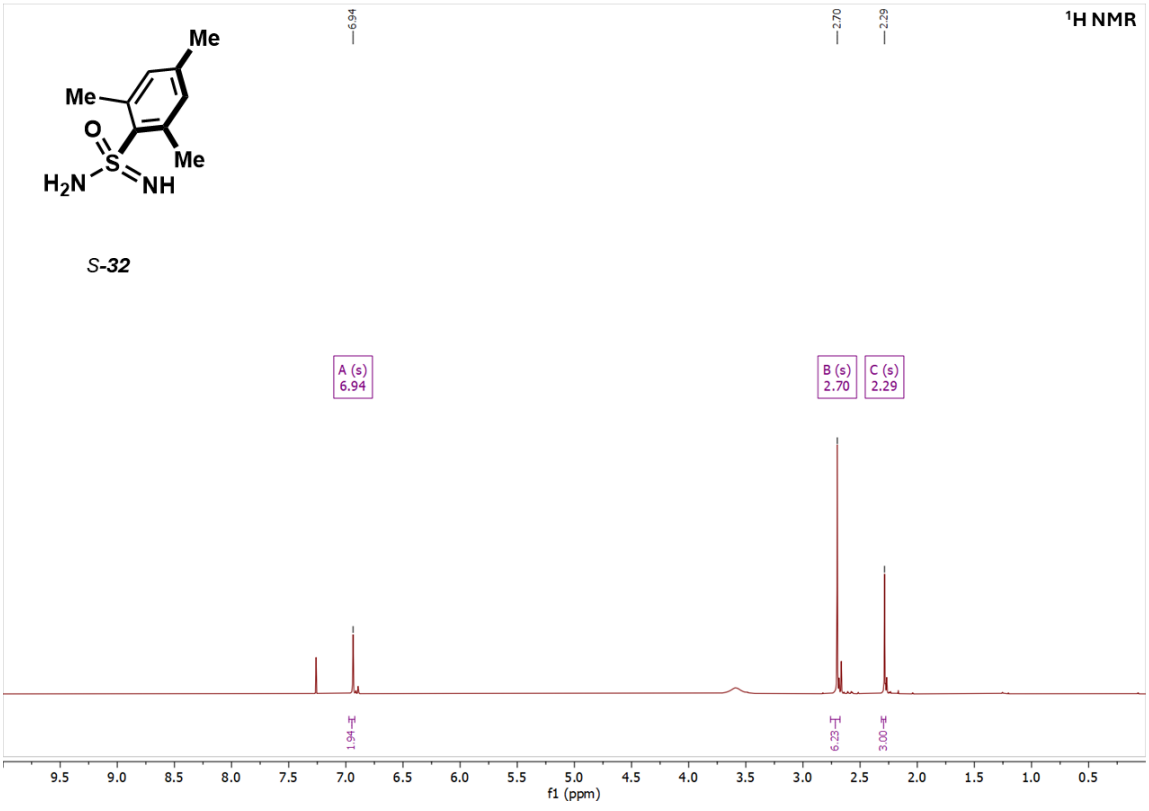
S-30


1H NMR
7.12
4.34 4.32 4.31 4.30 4.28
2.90 2.89 2.87 2.86 2.84
1.20 1.20 1.19 1.18
iPr
iPr
iPr
O
S
H2N
NH
C (s) 7.12
B (p) 4.31
A (p) 2.87
D (dd) 1.19
2.00
1.97
1.04
18.01
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

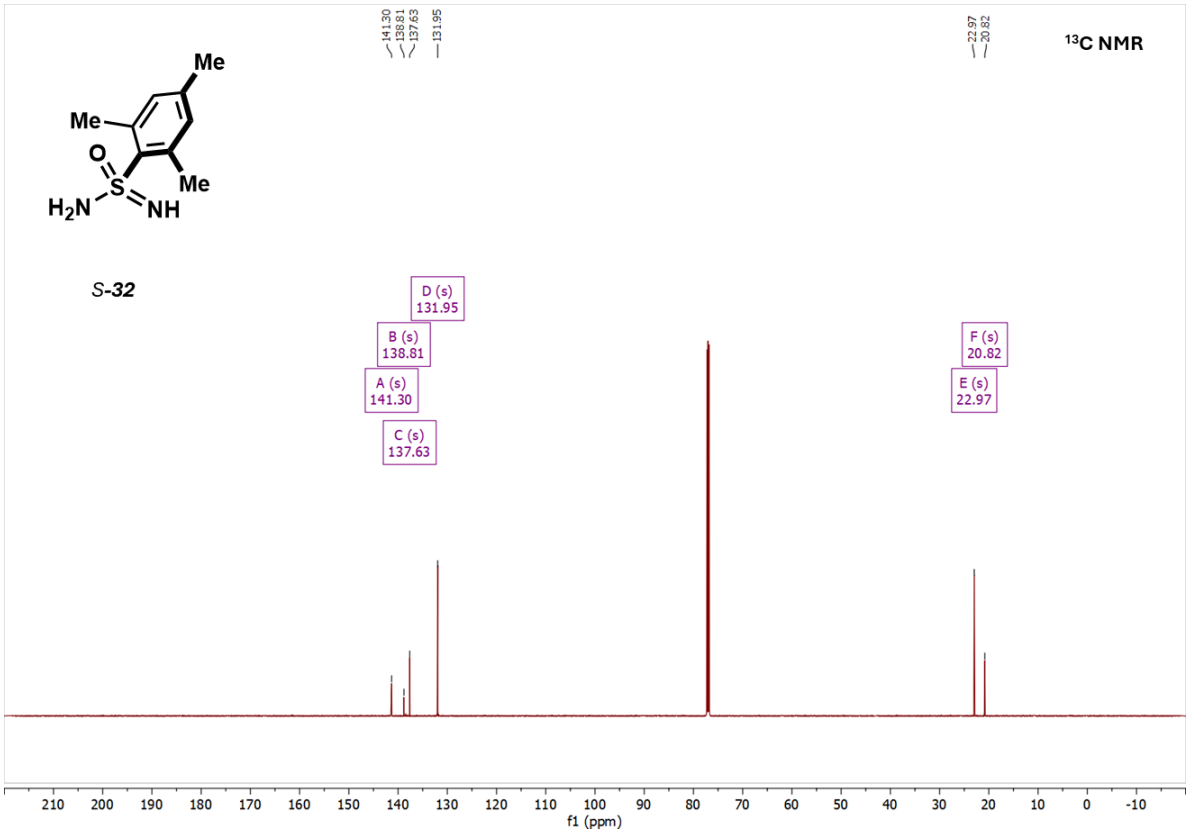
S-31

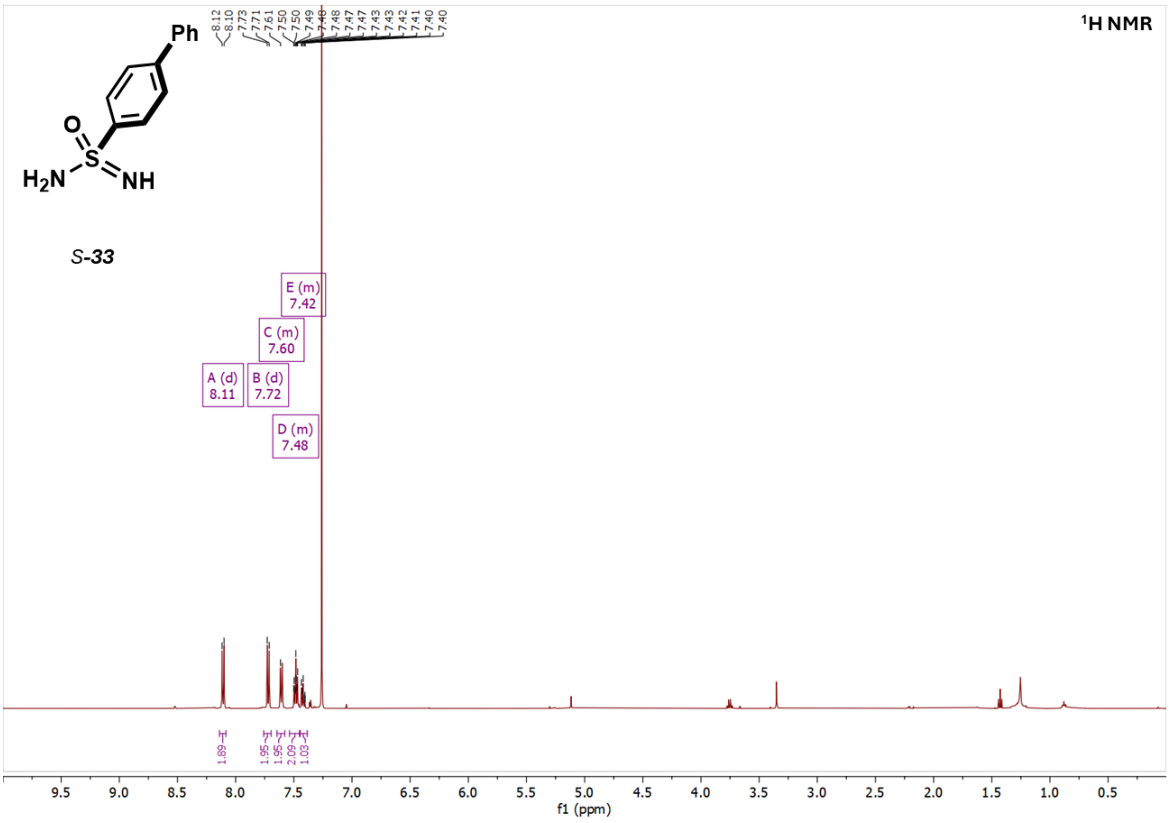

iPr
iPr
iPr
O
S
H2N
NH
149.98
147.83
140.69
122.78
33.31
28.58
24.69
23.63
13C NMR
f1 (ppm)

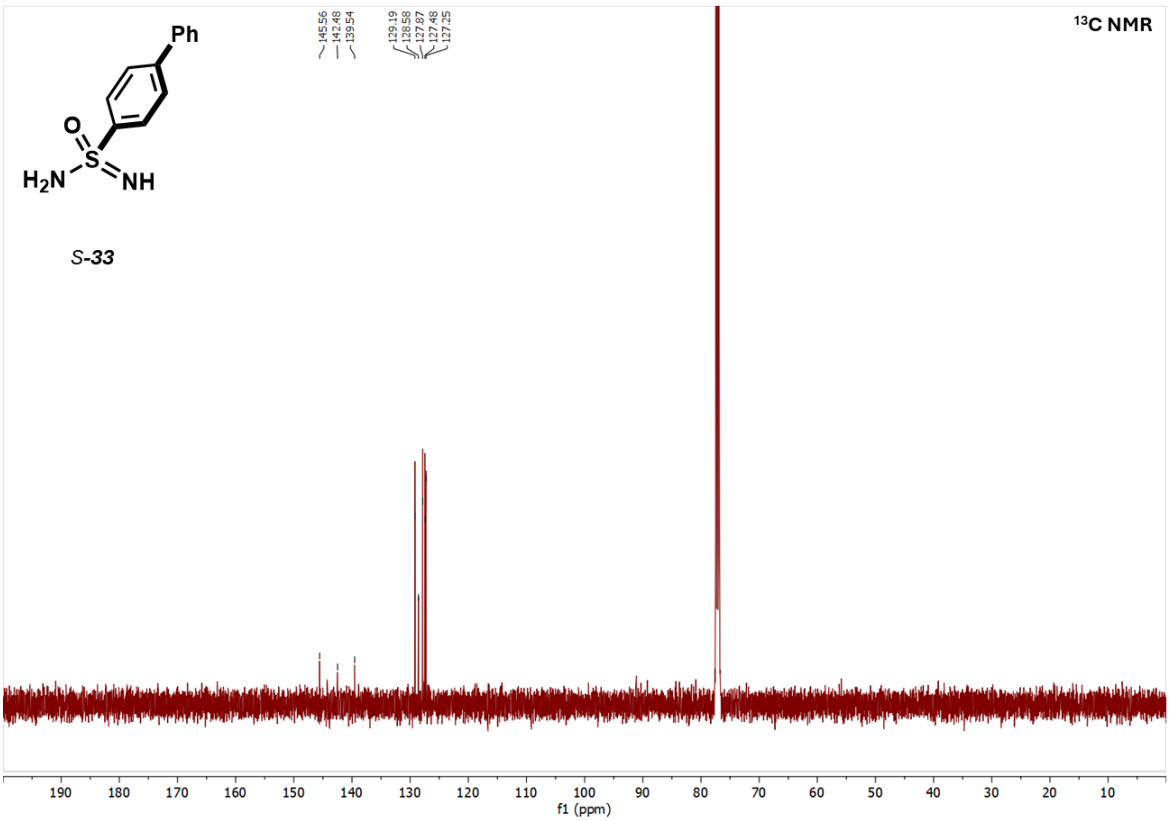
S-31

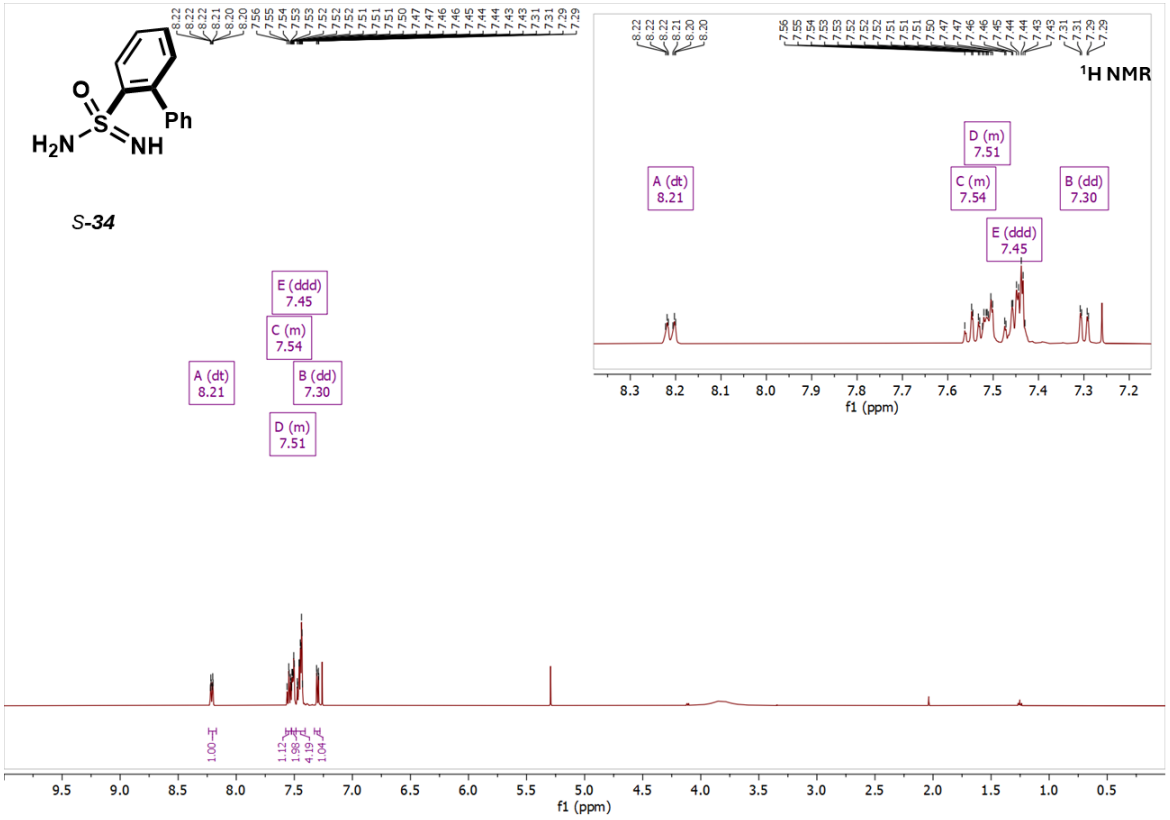

Me
Me
Me
O
S
H2N
NH
6.94
2.70
2.29
1H NMR
A (s)
6.94
B (s)
2.70
C (s)
2.29
1.94
6.23
3.00
f1 (ppm)

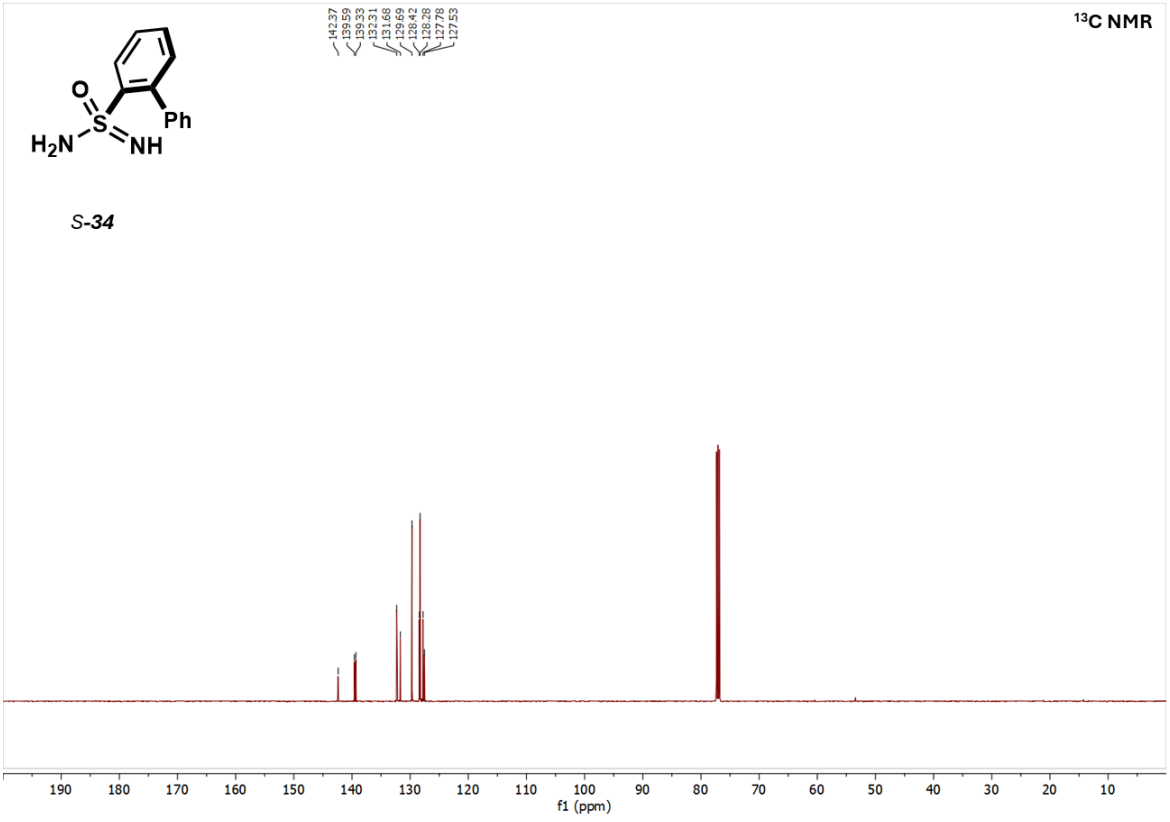
S-32

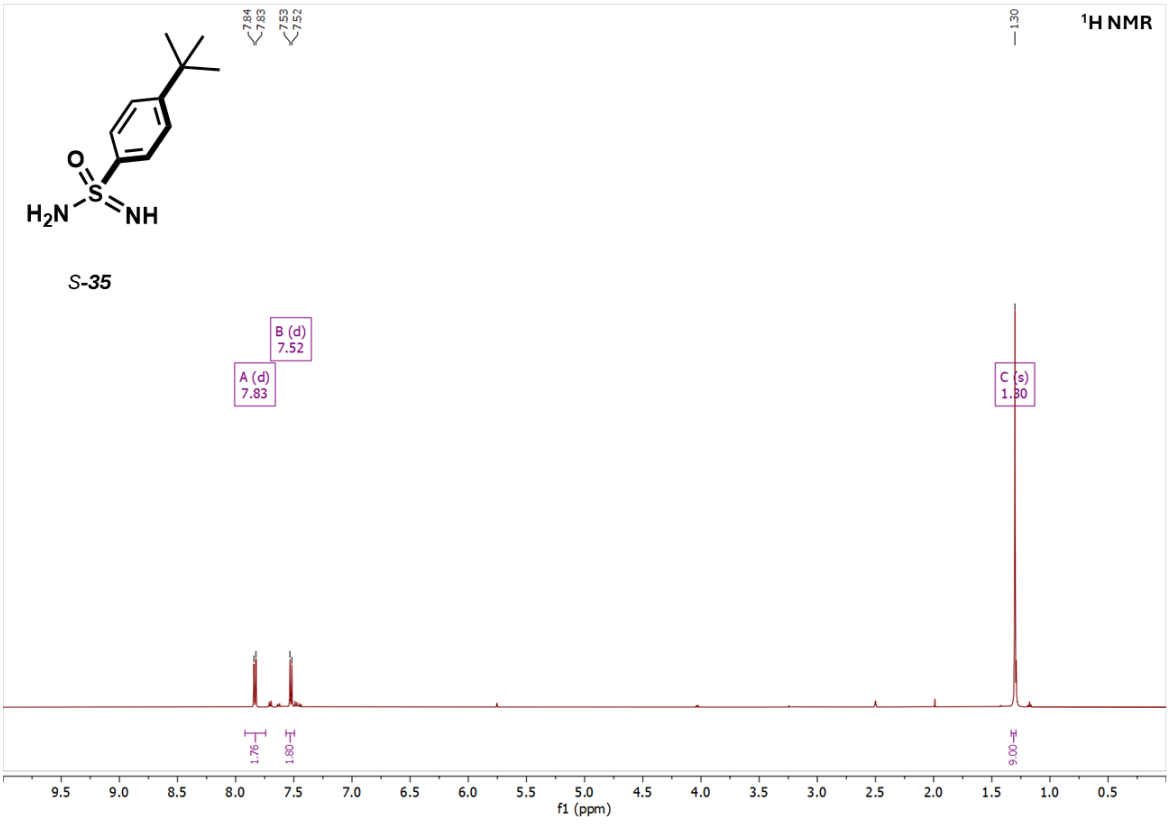

141.30
138.81
137.63
131.95
22.97
20.82
13C NMR
Me
Me
Me
O
S
H2N
NH
S-32
D (s)
131.95
B (s)
138.81
A (s)
141.30
C (s)
137.63
F (s)
20.82
E (s)
22.97
f1 (ppm)

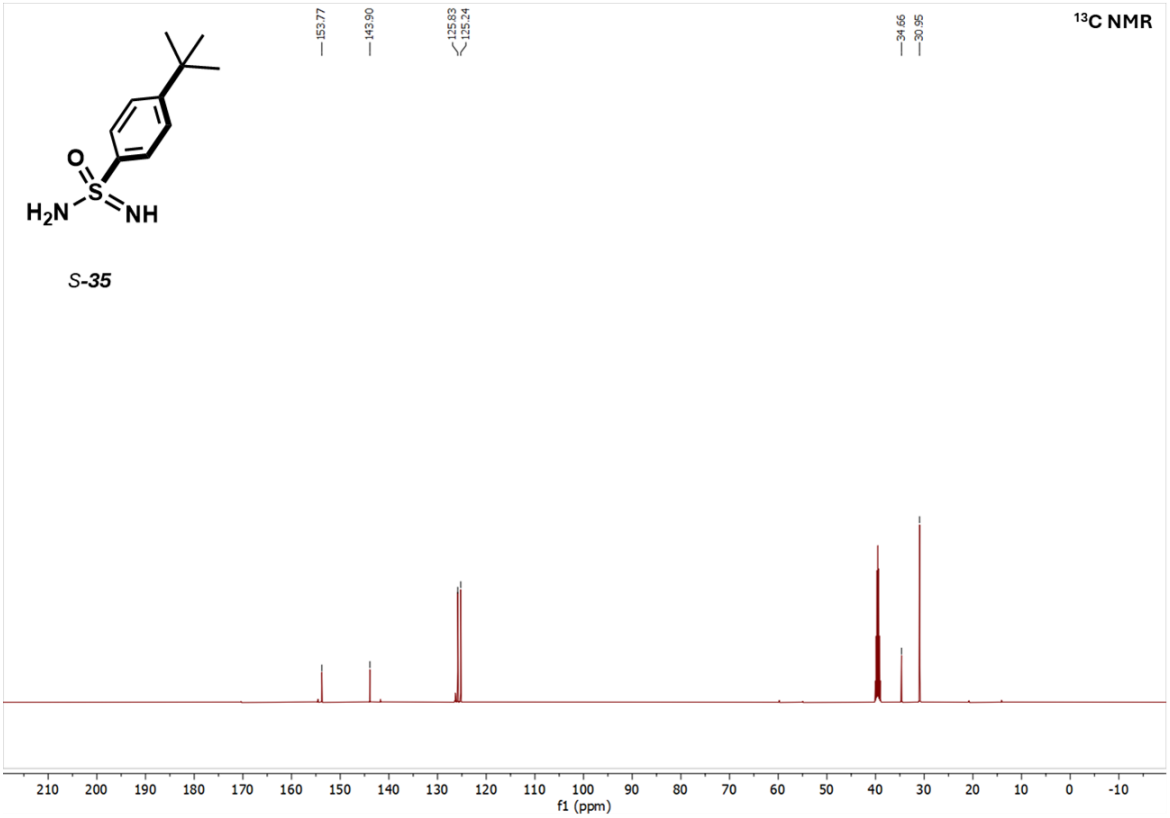

1H NMR
Ph
O
S
H2N
NH
S-33
E (m)
7.42
C (m)
7.60
A (d)
8.11
B (d)
7.72
D (m)
7.48
1.89
1.95
1.95
2.09
1.03
f1 (ppm)

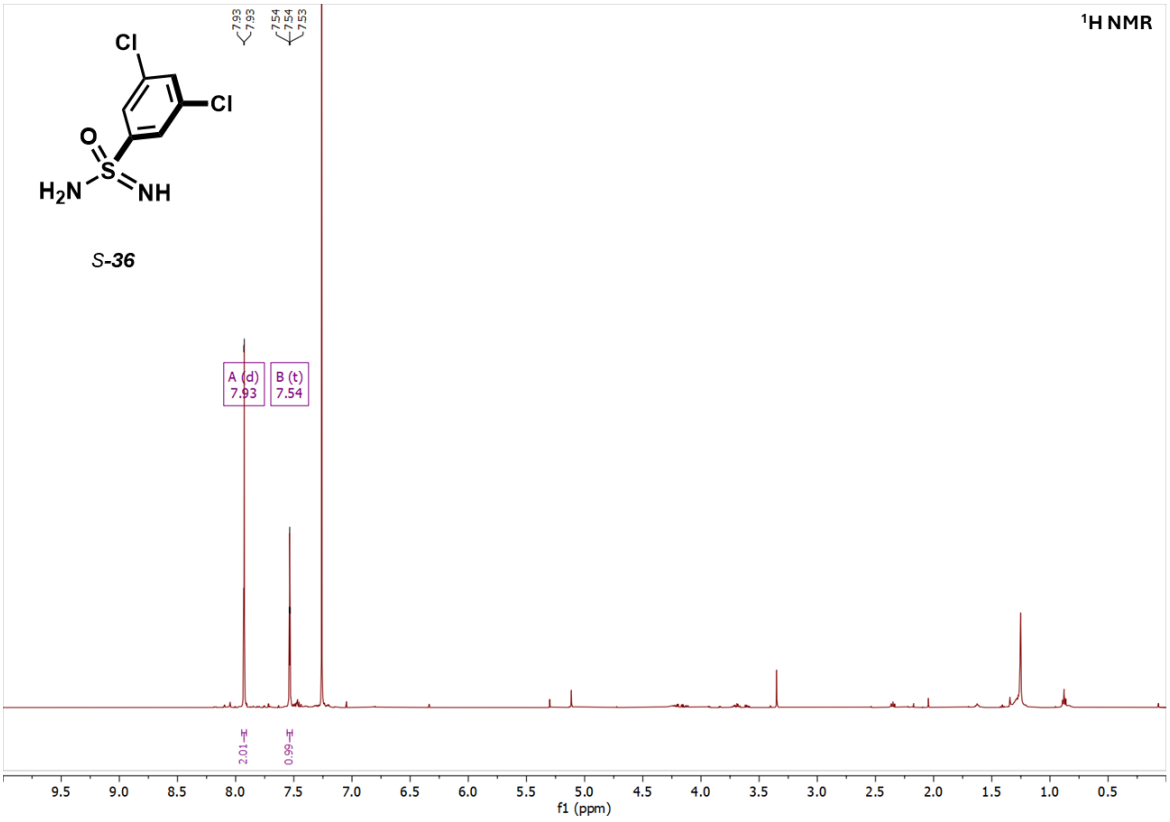

Ph
O
S
H2N
NH
S-33
145.56
142.48
139.54
129.19
128.58
127.87
127.48
127.25
13C NMR
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

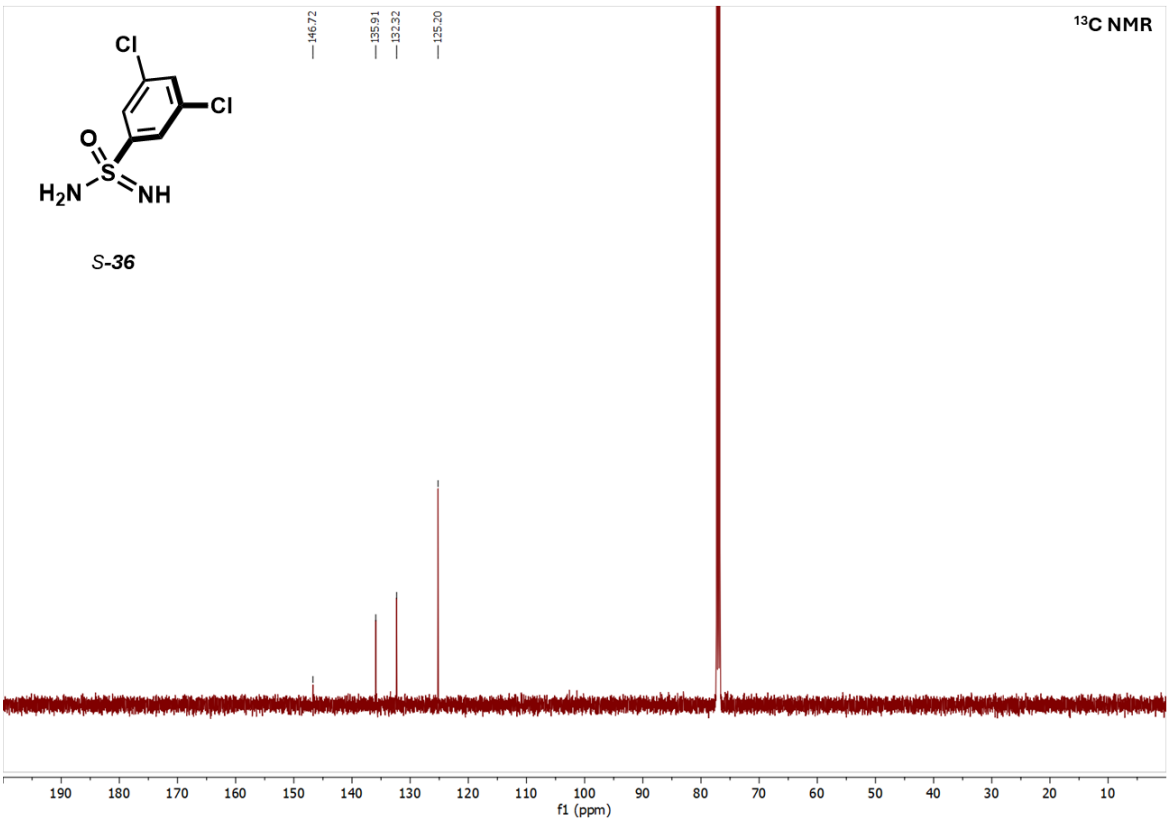

O
S
H2N
NH
Ph
S-34
1H NMR
A (dt)
8.21
B (dd)
7.30
C (m)
7.54
D (m)
7.51
E (ddd)
7.45
1.00
1.12
1.98
4.19
1.04
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

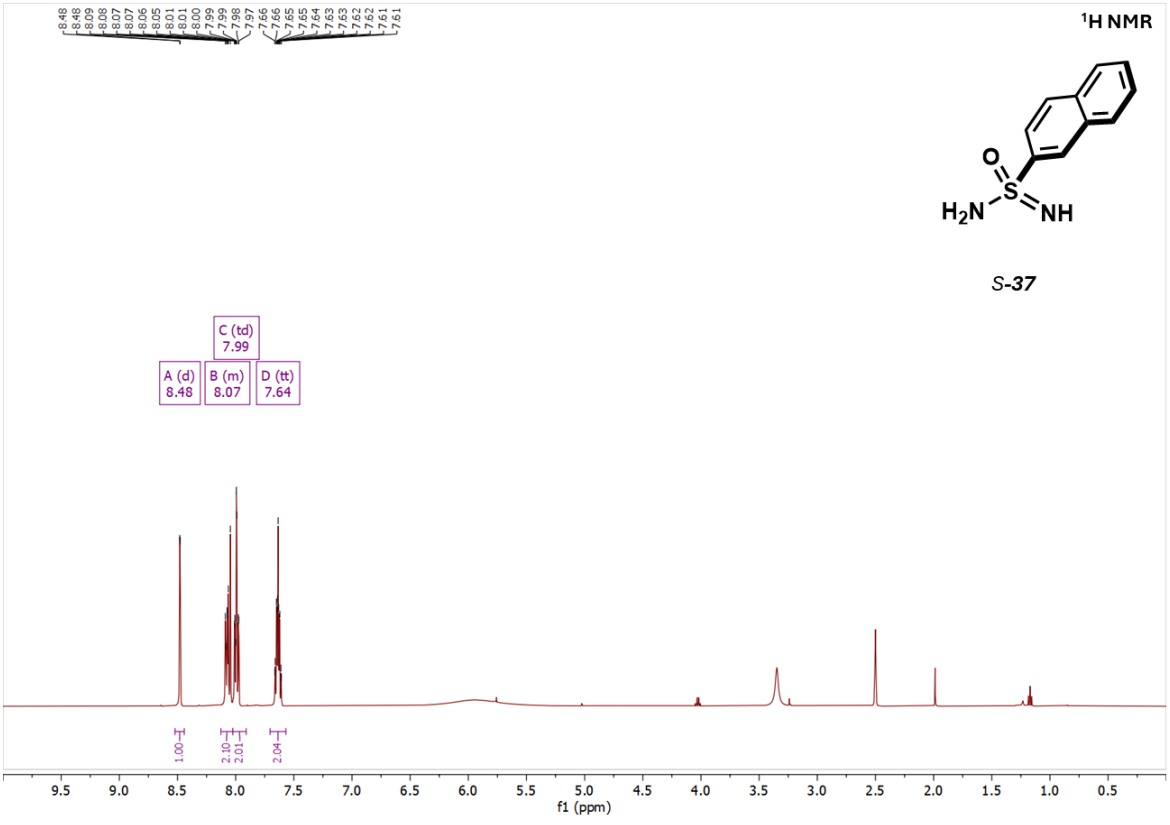

^{13}C NMR
142.37
139.59
139.33
132.31
131.68
131.49
128.42
128.28
127.78
127.53
O
S
Ph
H_2N
NH
S-34
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

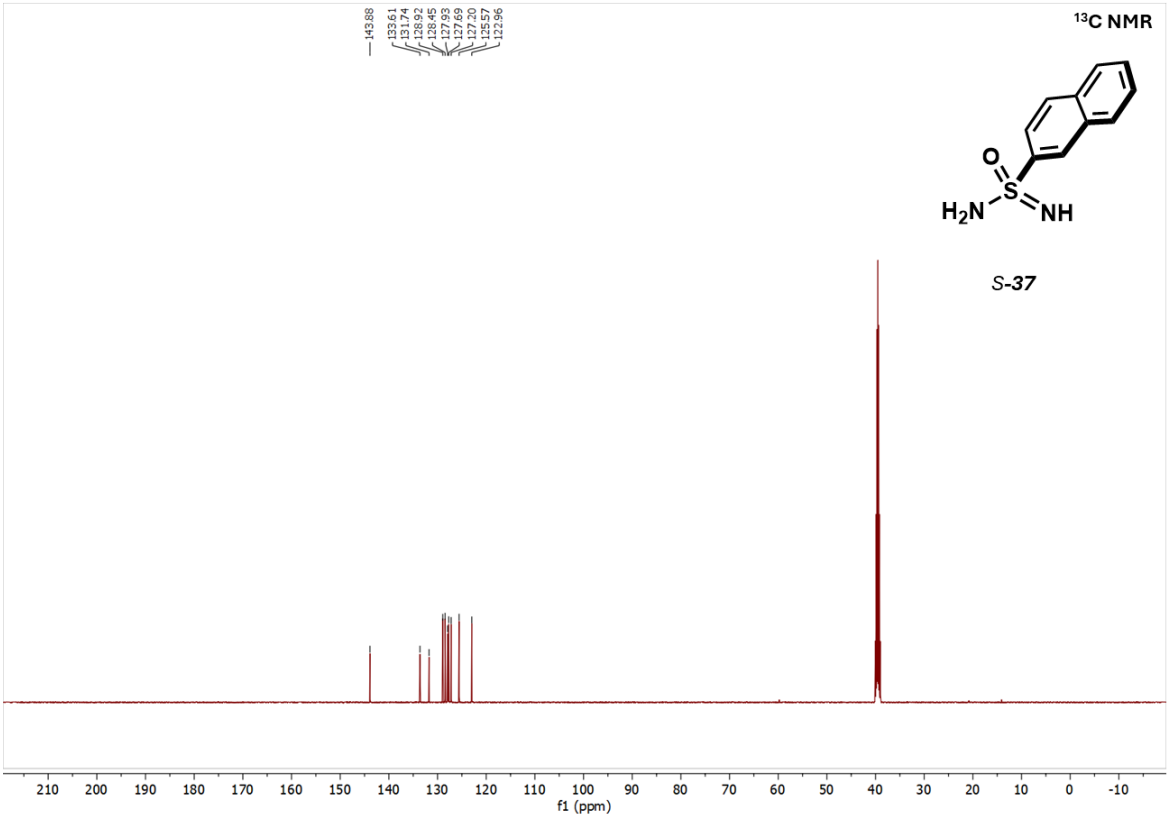

^{1}H NMR
7.84
7.83
7.53
7.52
1.30
O
S
H_2N
NH
S-35
B (d)
7.52
A (d)
7.83
C (s)
1.30
1.76
1.80
9.00
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

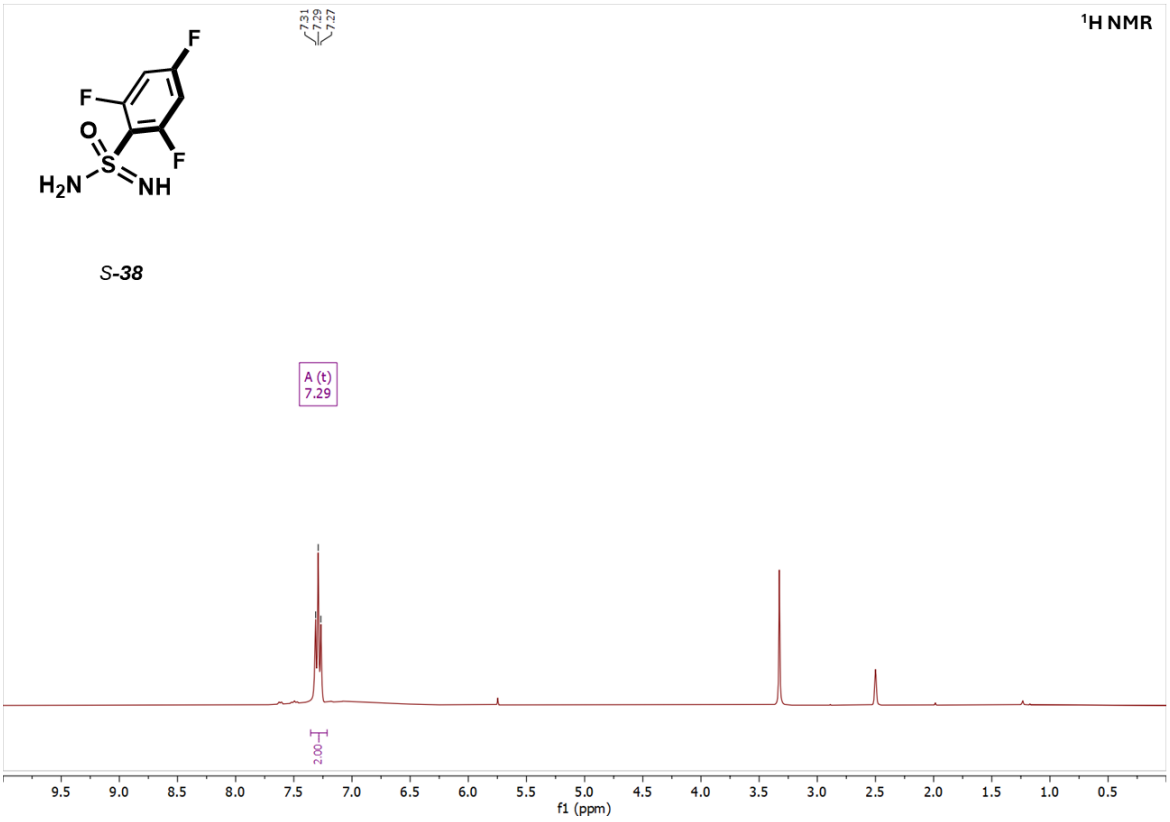

153.77
143.90
125.83
125.24
34.66
30.95
13C NMR
O
S
H2N
NH
S-35
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

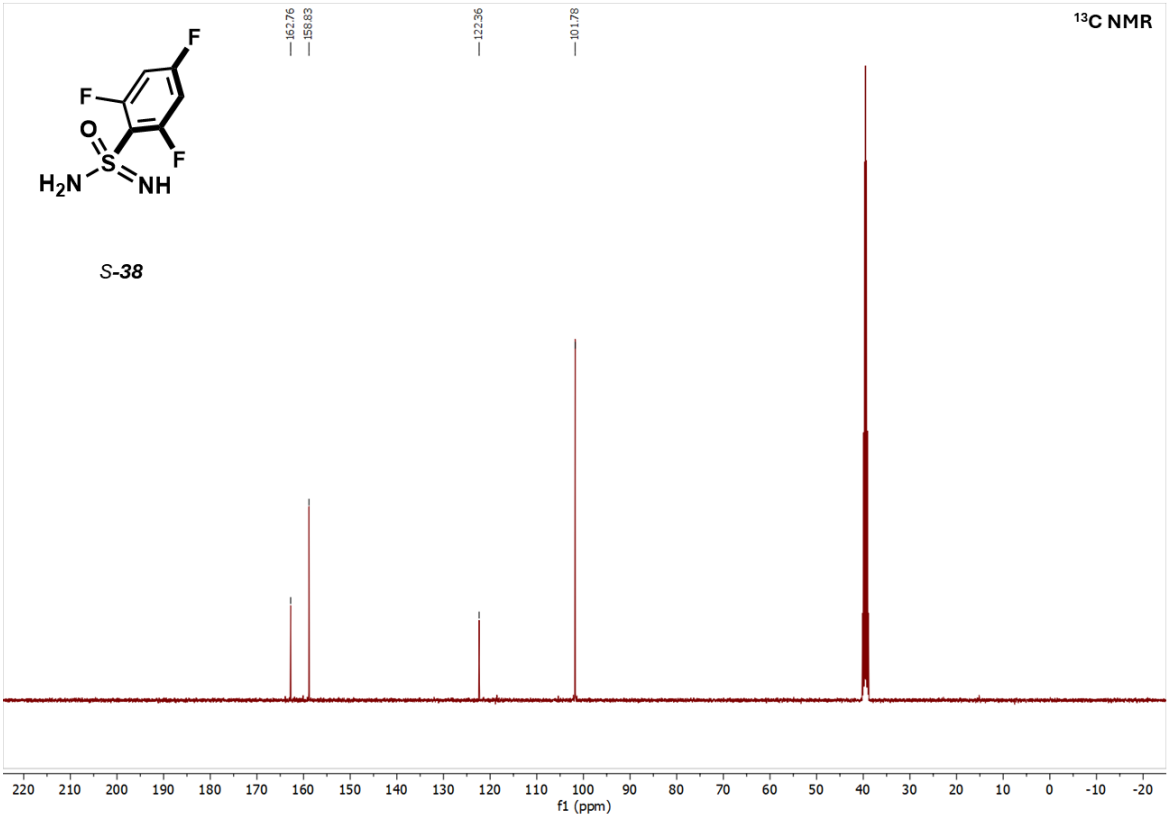

7.93
7.93
7.54
7.54
7.53
1H NMR
Cl
Cl
O
S
H2N
NH
S-36
A (d)
7.93
B (t)
7.54
2.01
0.99
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

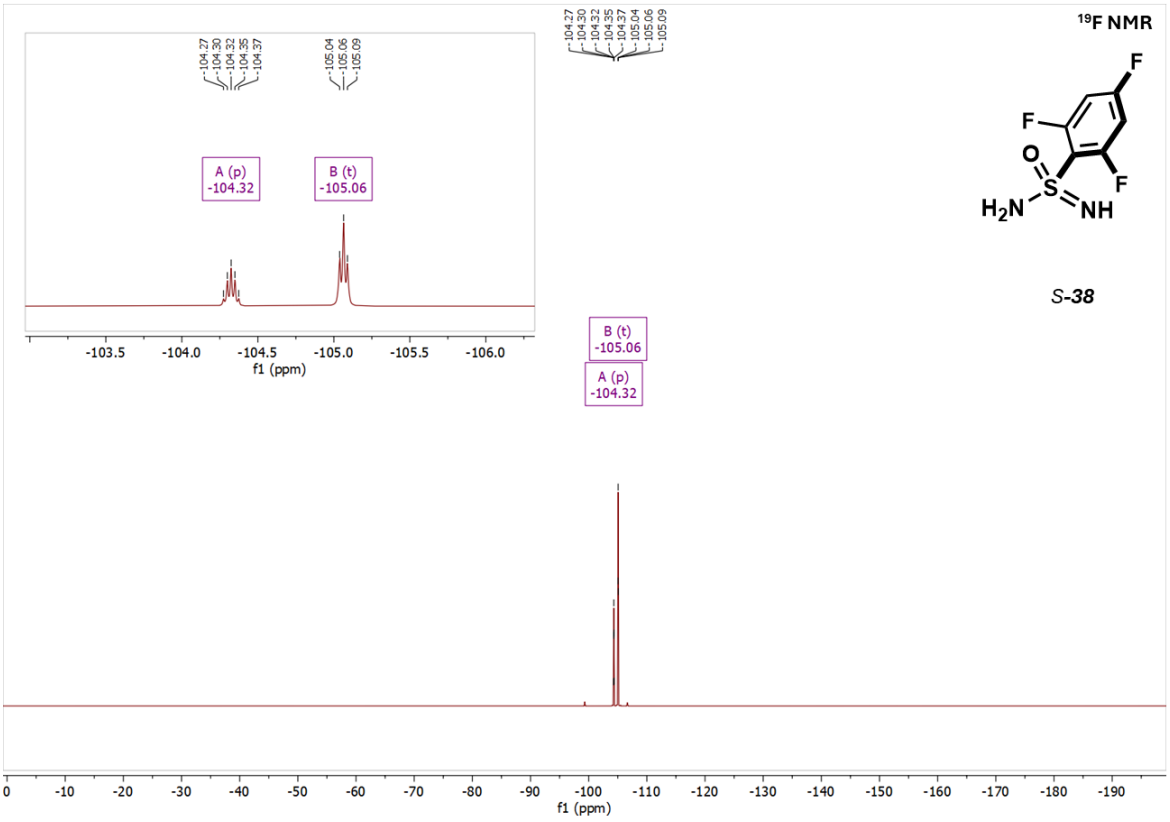

13C NMR
146.72
135.91
132.32
125.20
Cl
Cl
O
S
H2N
NH
S-36
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

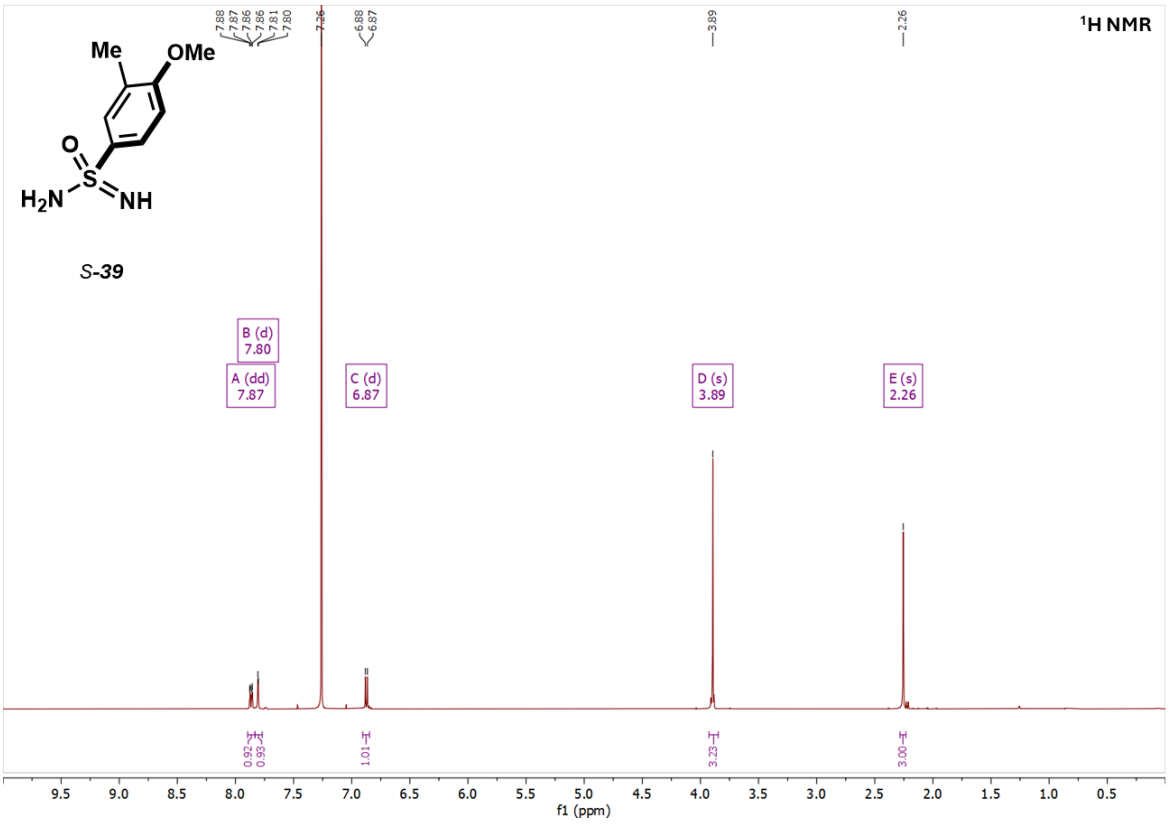

1H NMR
8.48 8.48 8.09 8.08 8.07 8.07 8.06 8.05 8.01 8.01 8.00 7.99 7.99 7.98 7.97 7.66 7.66 7.65 7.65 7.64 7.63 7.63 7.62 7.62 7.61 7.61
O
S
H2N
NH
S-37
C (td) 7.99
A (d) 8.48
B (m) 8.07
D (tt) 7.64
1.00
2.10
2.01
2.04
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

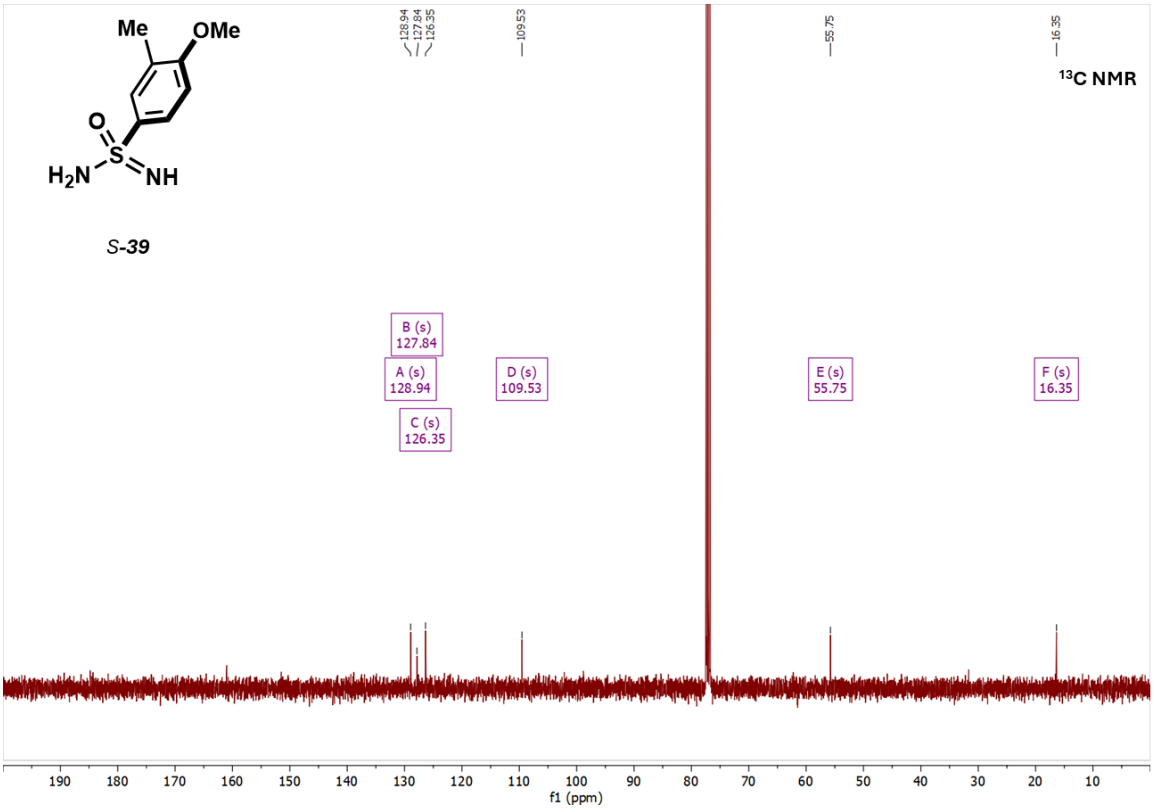

143.88
133.61
131.74
128.92
128.45
127.93
127.69
127.20
125.57
122.96
13C NMR
O
S
H2N
NH
S-37
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

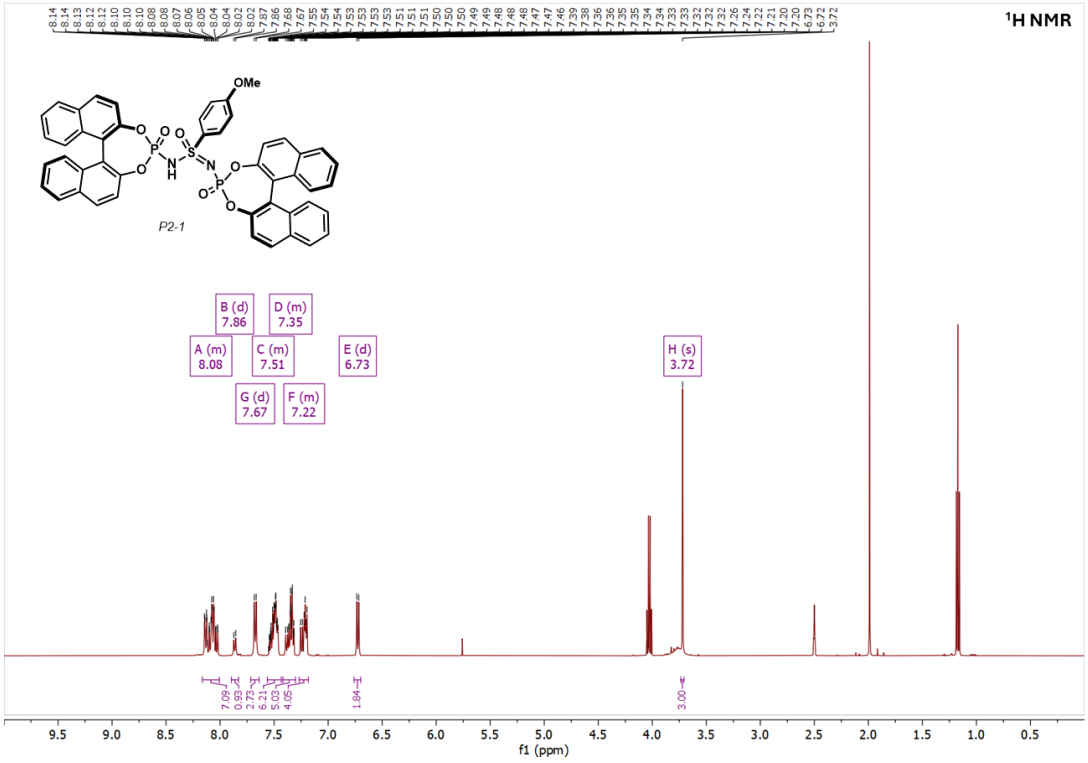

7.31
7.29
7.27
1H NMR
F
F
O
S
F
H2N
NH
S-38
A (t)
7.29
2.00
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

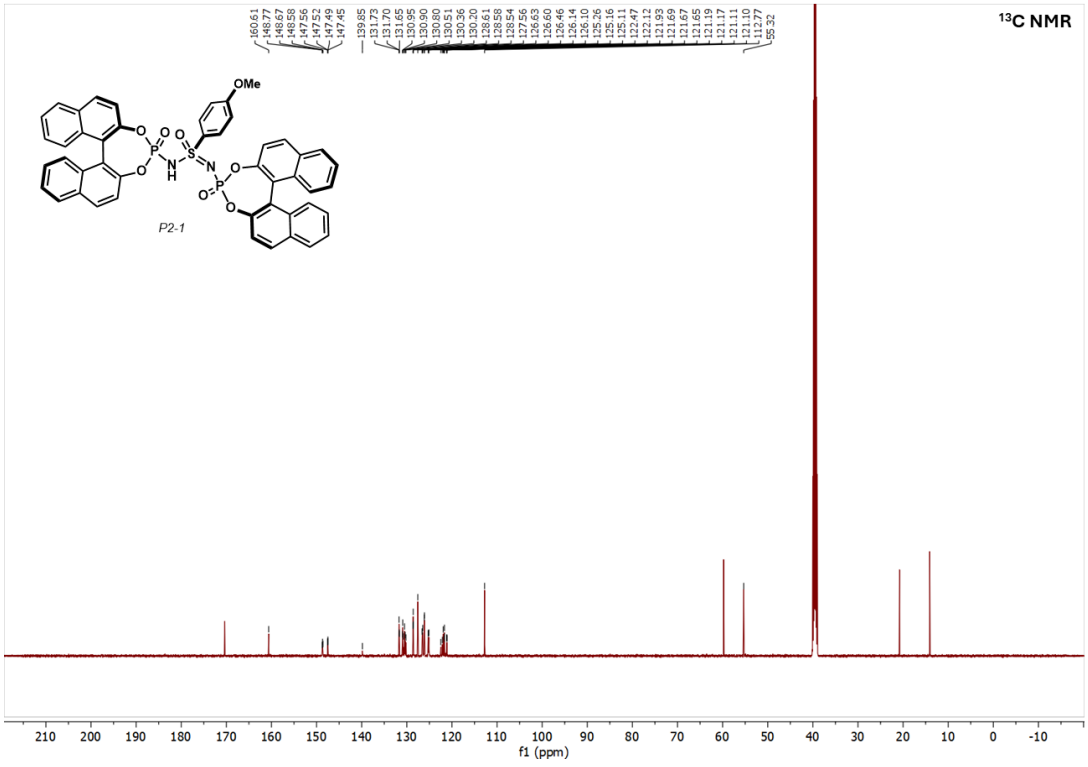

^{13}C NMR
162.76
158.83
122.36
101.78
F
F
O
S
F
H_2N
NH
S-38
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20
f1 (ppm)

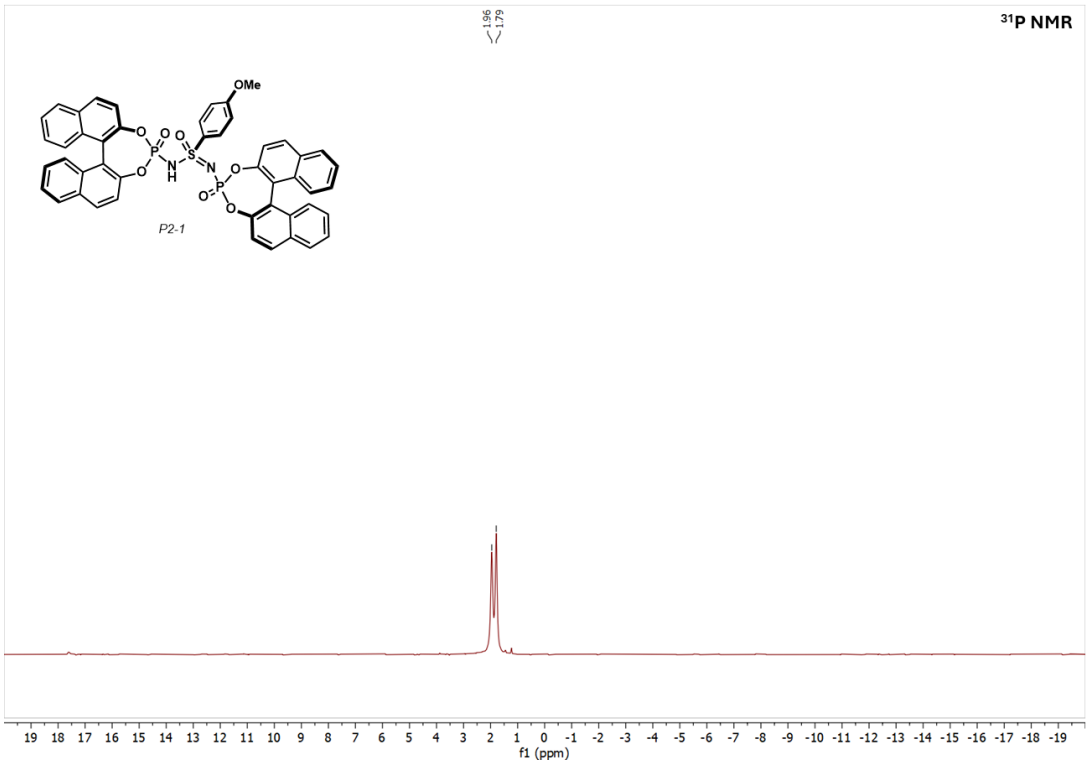

^{19}F NMR
-104.27
-104.30
-104.32
-104.35
-104.37
-105.04
-105.06
-105.09
A (p)
-104.32
B (t)
-105.06
-103.5 -104.0 -104.5 -105.0 -105.5 -106.0
f1 (ppm)
F
F
O
S
F
H_2N
NH
S-38
0 -10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)

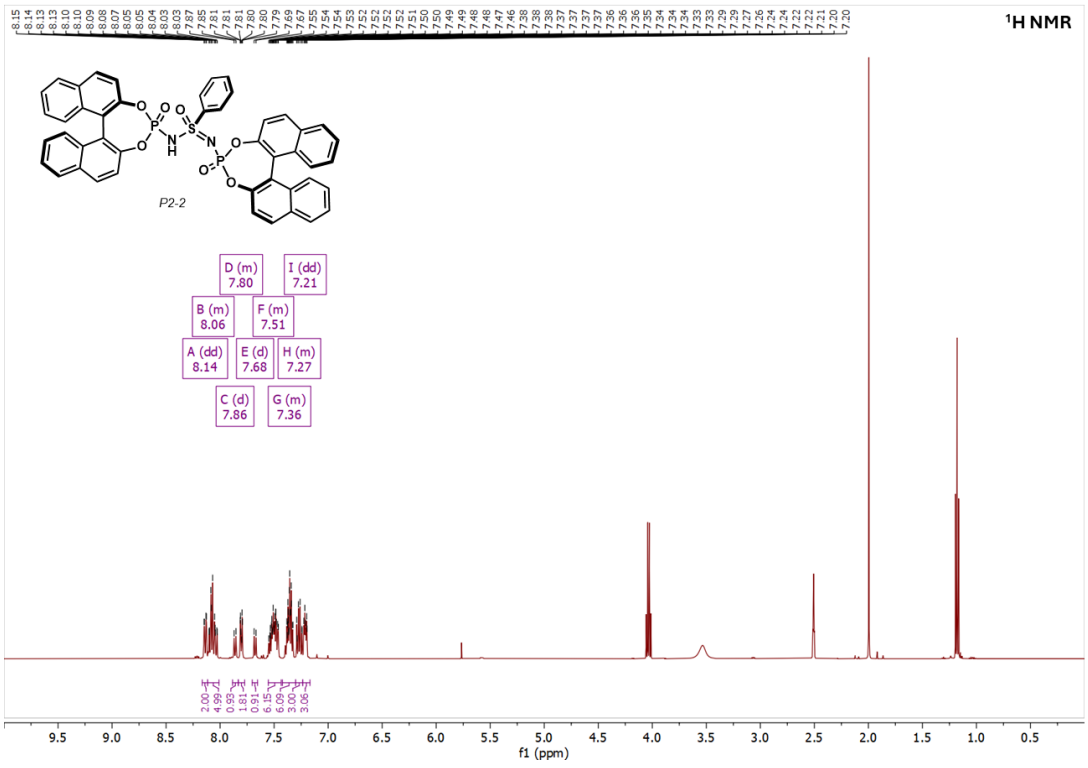

Me
OMe
O
S
H_2N
NH
S-39
^{1}H NMR
7.88
7.87
7.86
7.86
7.81
7.80
7.26
6.88
6.87
3.89
2.26
B (d)
7.80
A (dd)
7.87
C (d)
6.87
D (s)
3.89
E (s)
2.26
0.92
0.93
1.01
3.23
3.00
f1 (ppm)

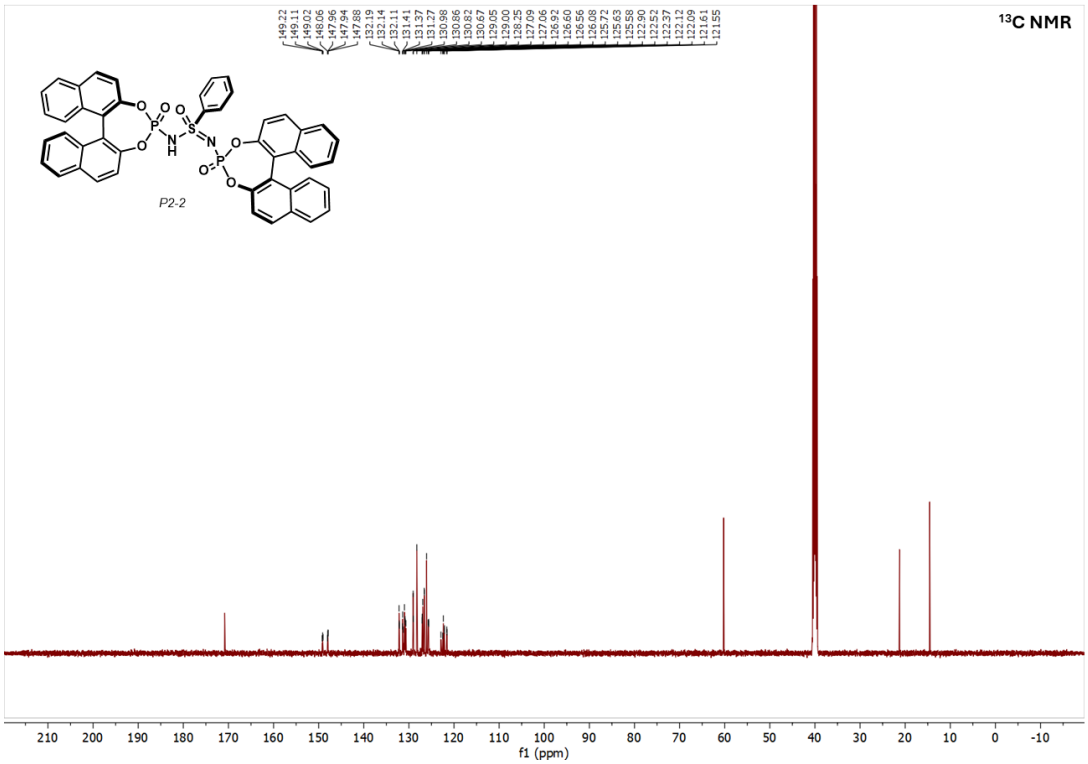

Me
OMe
O
S
H_2N
NH
S-39
^{13}C NMR
128.94
127.84
126.35
109.53
55.75
16.35
B (s)
127.84
A (s)
128.94
C (s)
126.35
D (s)
109.53
E (s)
55.75
F (s)
16.35
f1 (ppm)

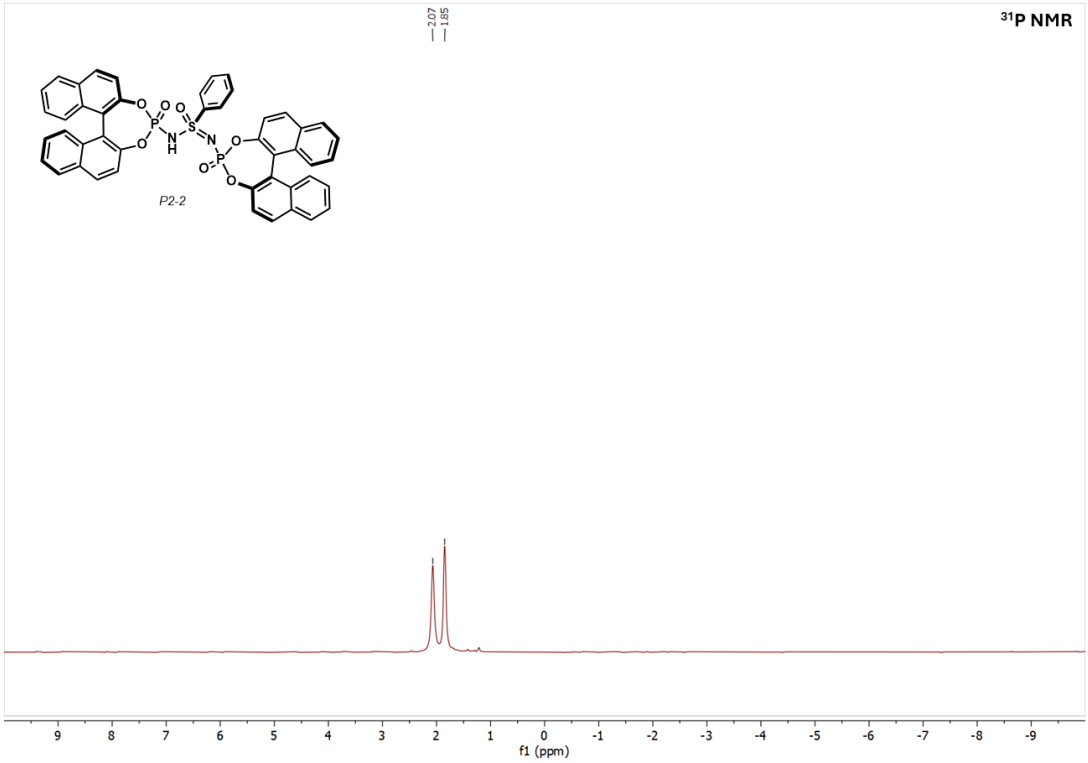

¹H NMR
OMe
P2-1
B (d) 7.86
D (m) 7.35
A (m) 8.08
C (m) 7.51
E (d) 6.73
H (s) 3.72
G (d) 7.67
F (m) 7.22
7.09
0.93
2.73
6.21
4.05
1.84
3.00
f1 (ppm)

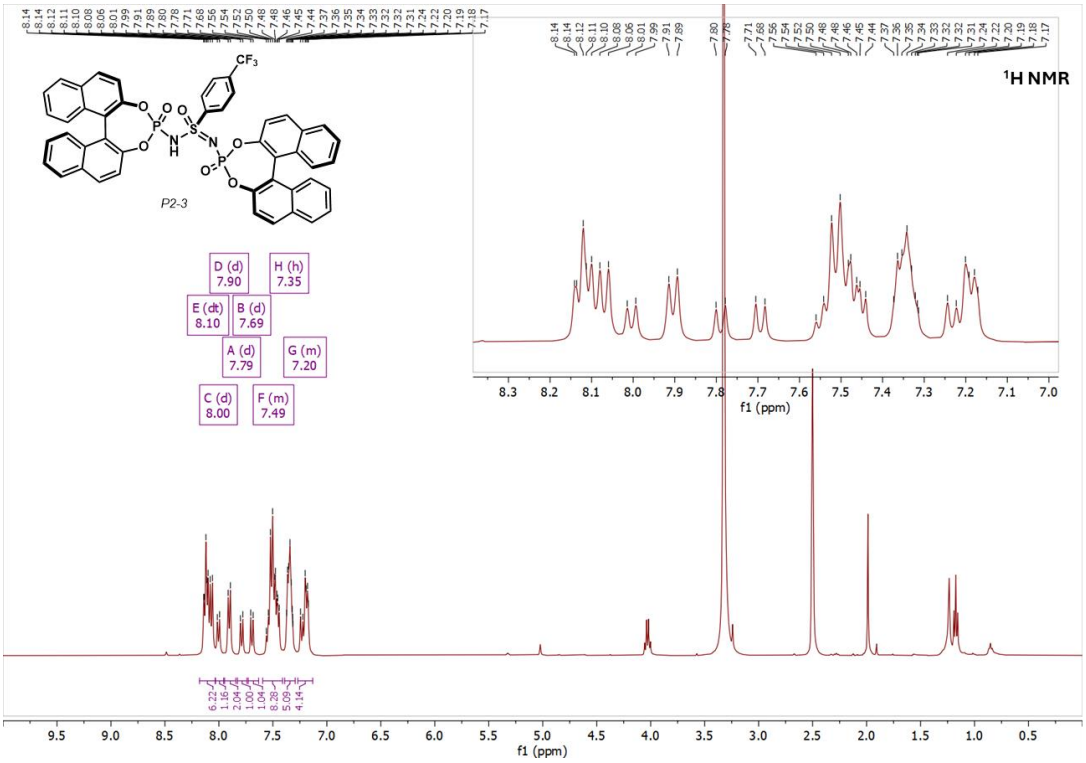

¹³C NMR
OMe
P2-1
f1 (ppm)

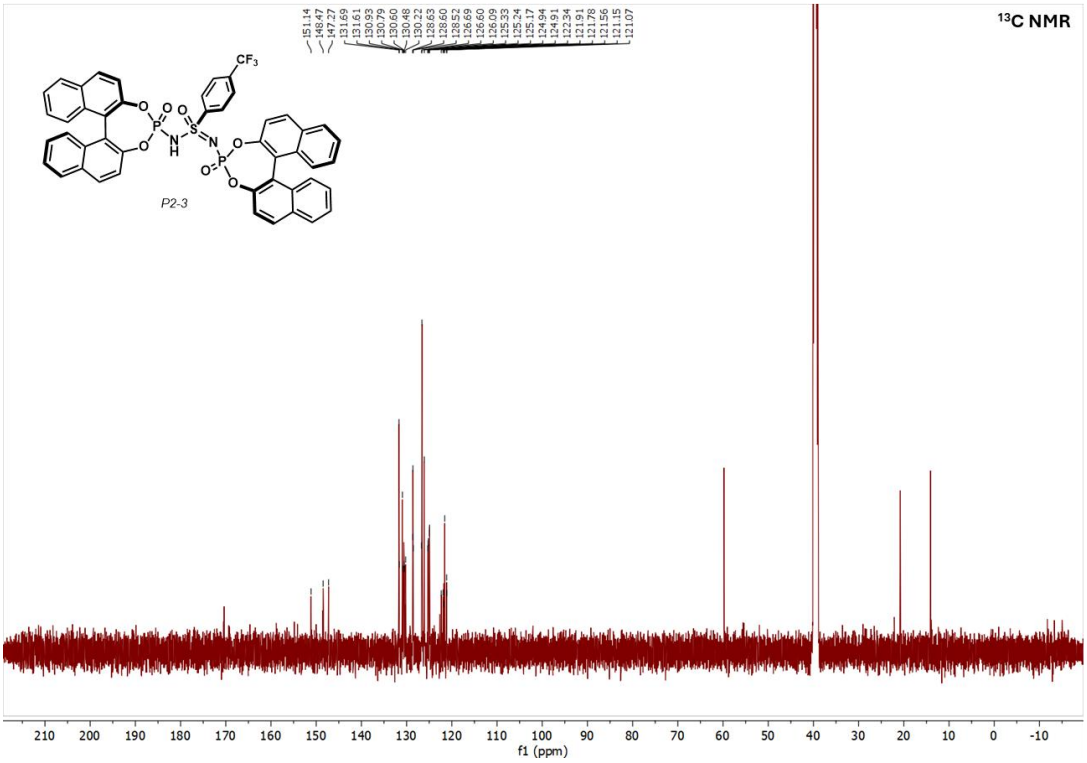

31P NMR
196
179
OMe
P2-1
19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10 -11 -12 -13 -14 -15 -16 -17 -18 -19
f1 (ppm)

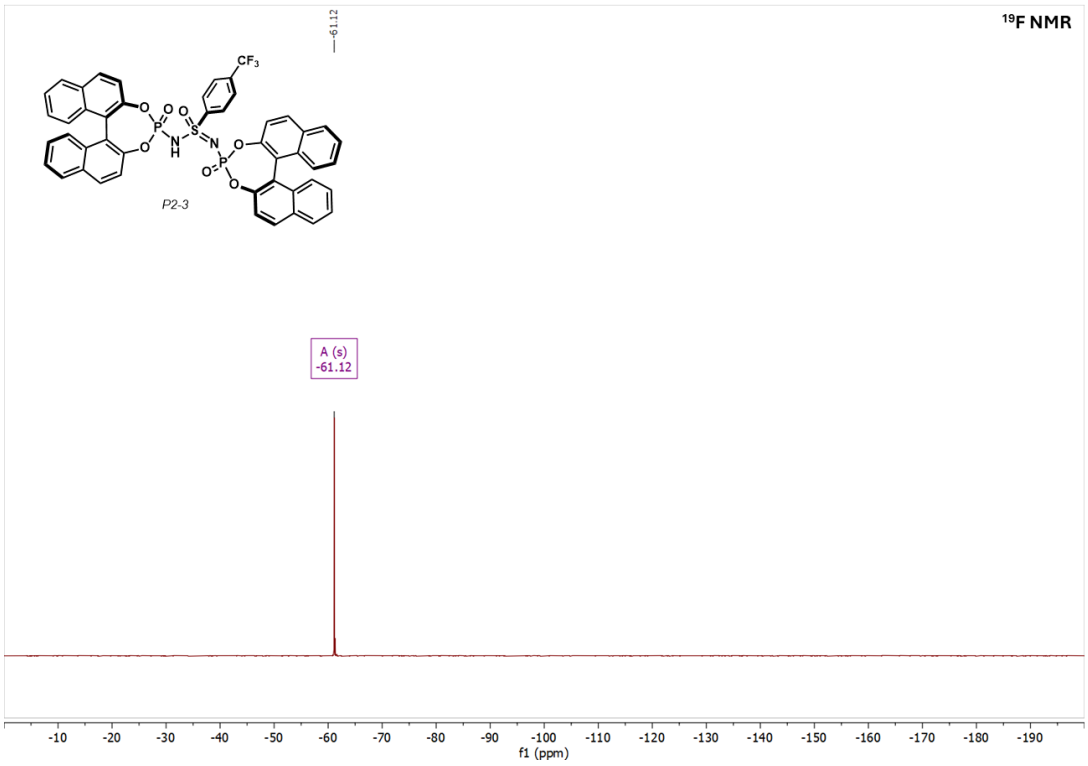

1H NMR
P2-2
D (m) 7.80
I (dd) 7.21
B (m) 8.06
F (m) 7.51
A (dd) 8.14
E (d) 7.68
H (m) 7.27
C (d) 7.86
G (m) 7.36
2.00
4.99
0.93
1.81
0.91
6.15
6.09
3.00
3.06
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

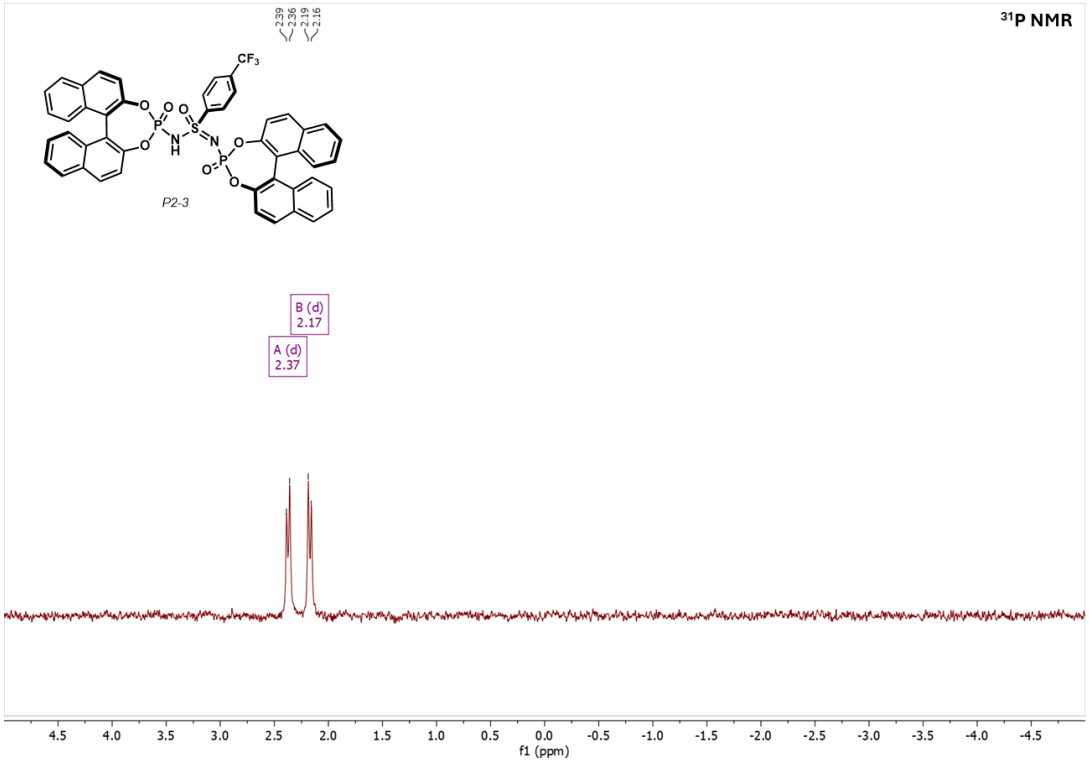

^{13}C NMR
P2-2
f1 (ppm)

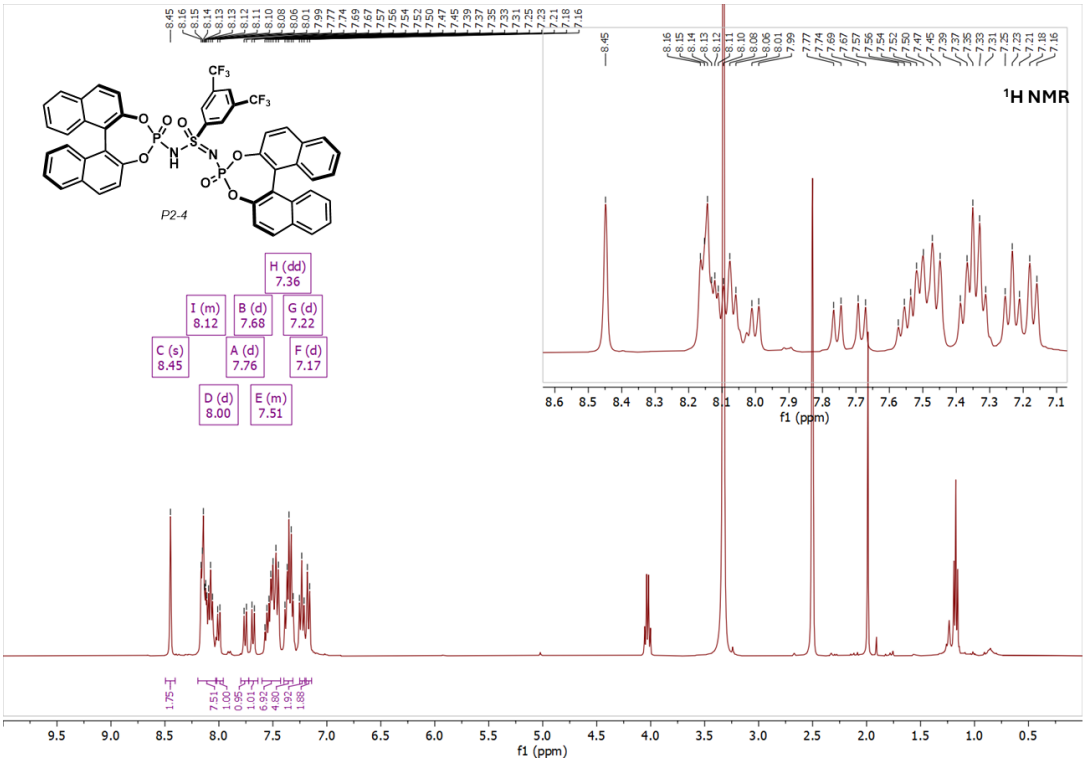

^{31}P NMR
2.07
1.85
P2-2
f1 (ppm)

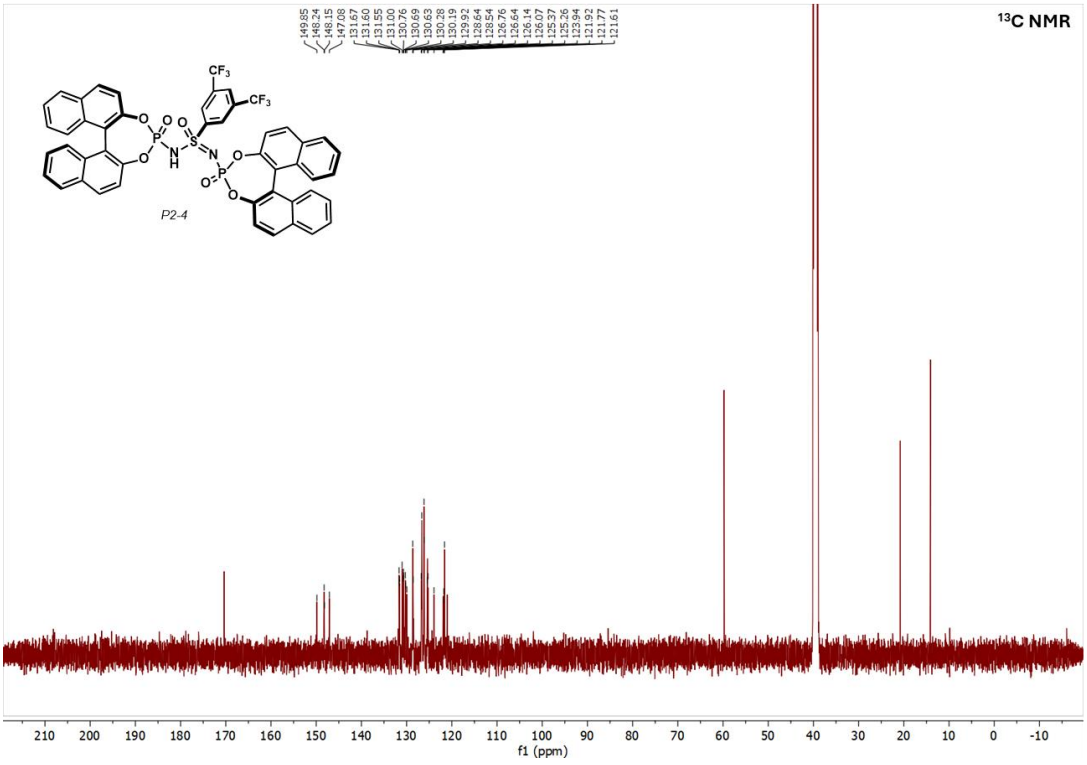

^{1}H NMR
P2-3
D (d) 7.90
H (h) 7.35
E (dt) 8.10
B (d) 7.69
A (d) 7.79
G (m) 7.20
C (d) 8.00
F (m) 7.49
f1 (ppm)

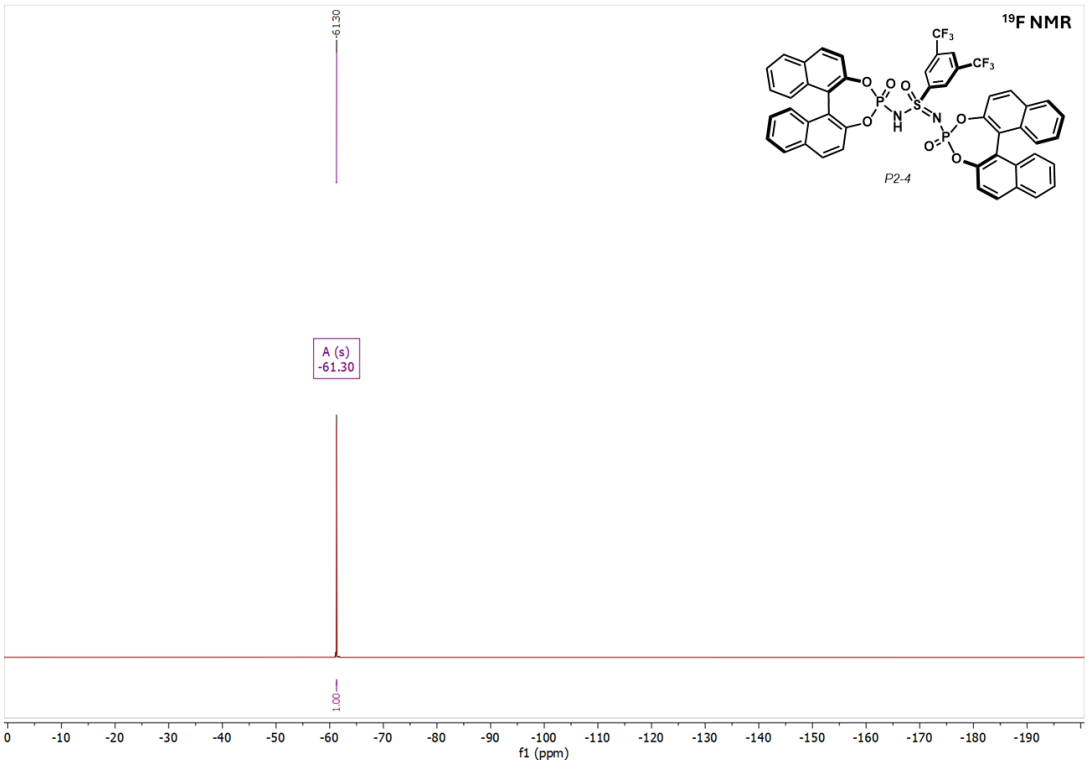

^{13}C NMR
P2-3
f1 (ppm)

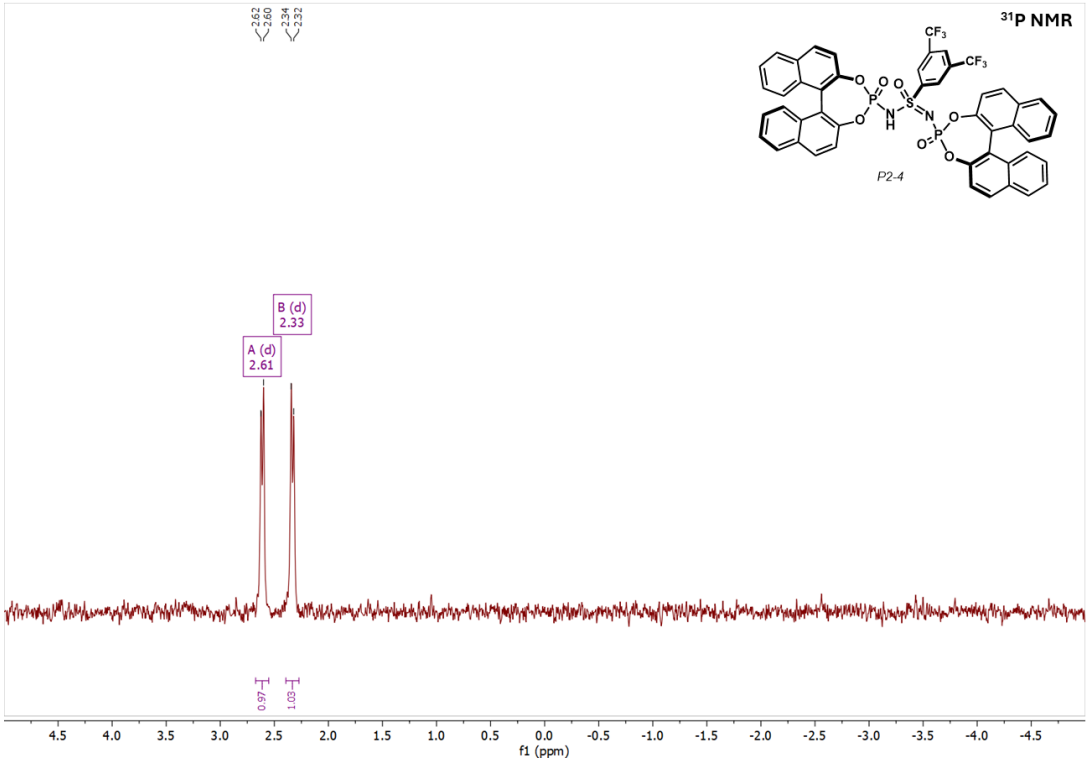

19F NMR
-61.12
P2-3
A (s)
-61.12
-10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190
f1 (ppm)

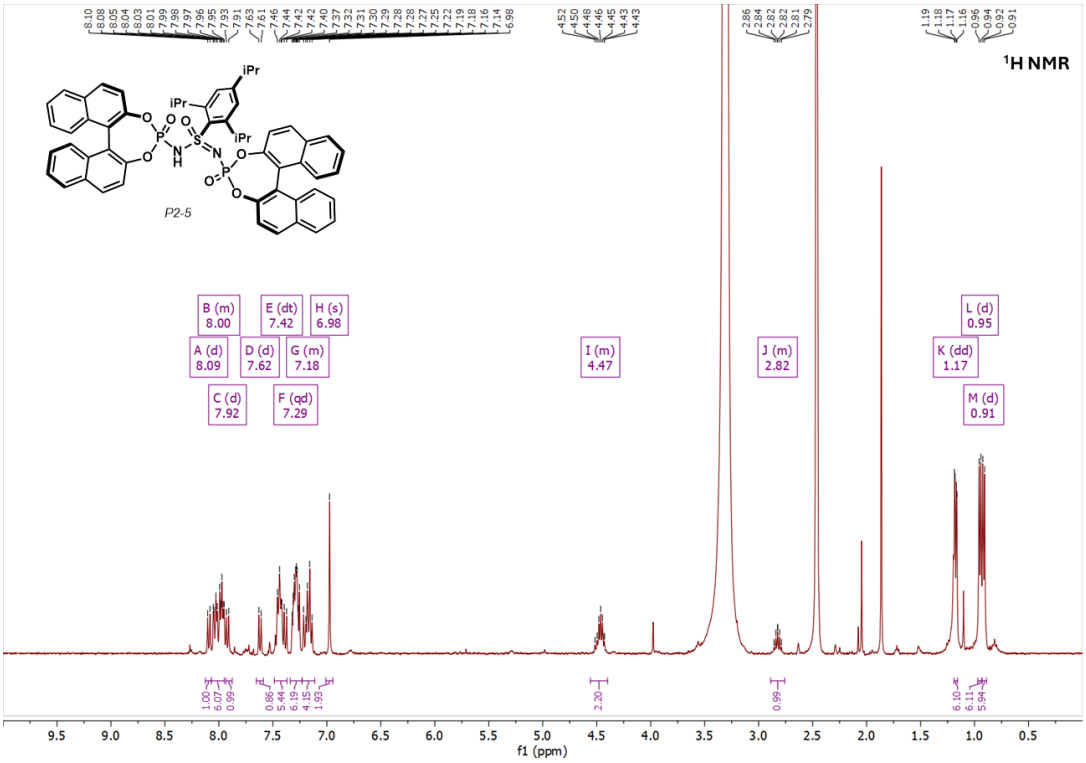

31P NMR
2.39
2.36
2.19
2.16
P2-3
B (d)
2.17
A (d)
2.37
4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0 -1.5 -2.0 -2.5 -3.0 -3.5 -4.0 -4.5
f1 (ppm)

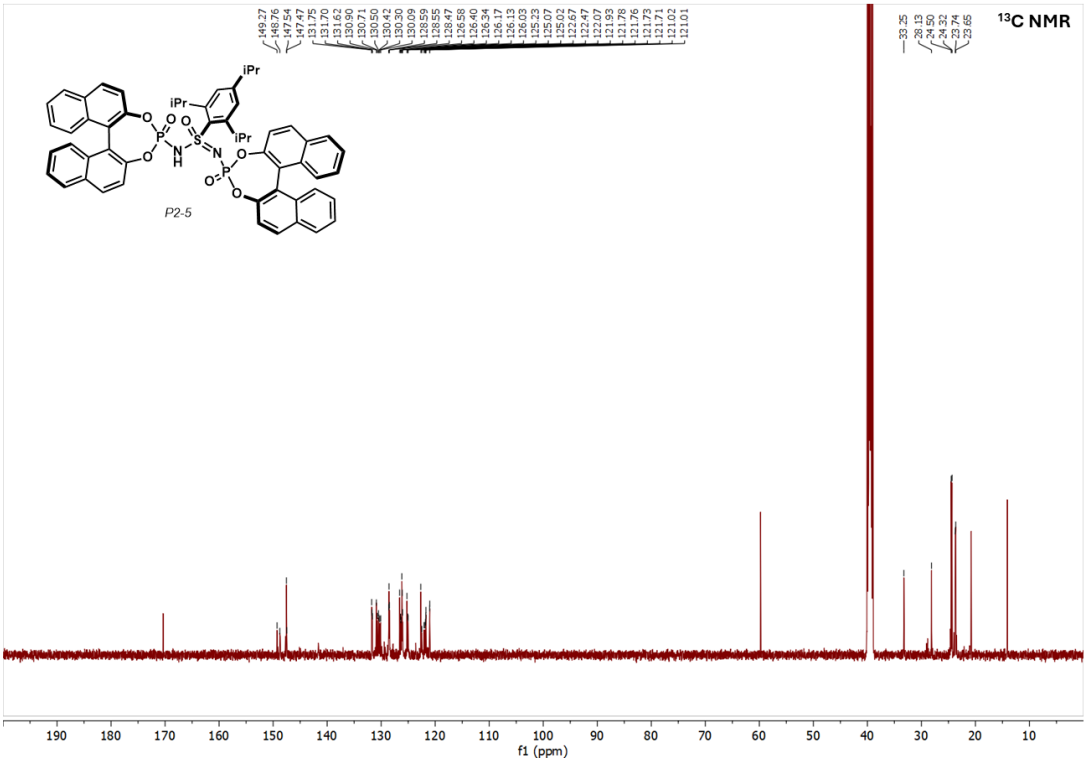

^{1}H NMR
P2-4
H (dd) 7.36
I (m) 8.12
B (d) 7.68
G (d) 7.22
C (s) 8.45
A (d) 7.76
F (d) 7.17
D (d) 8.00
E (m) 7.51
f1 (ppm)

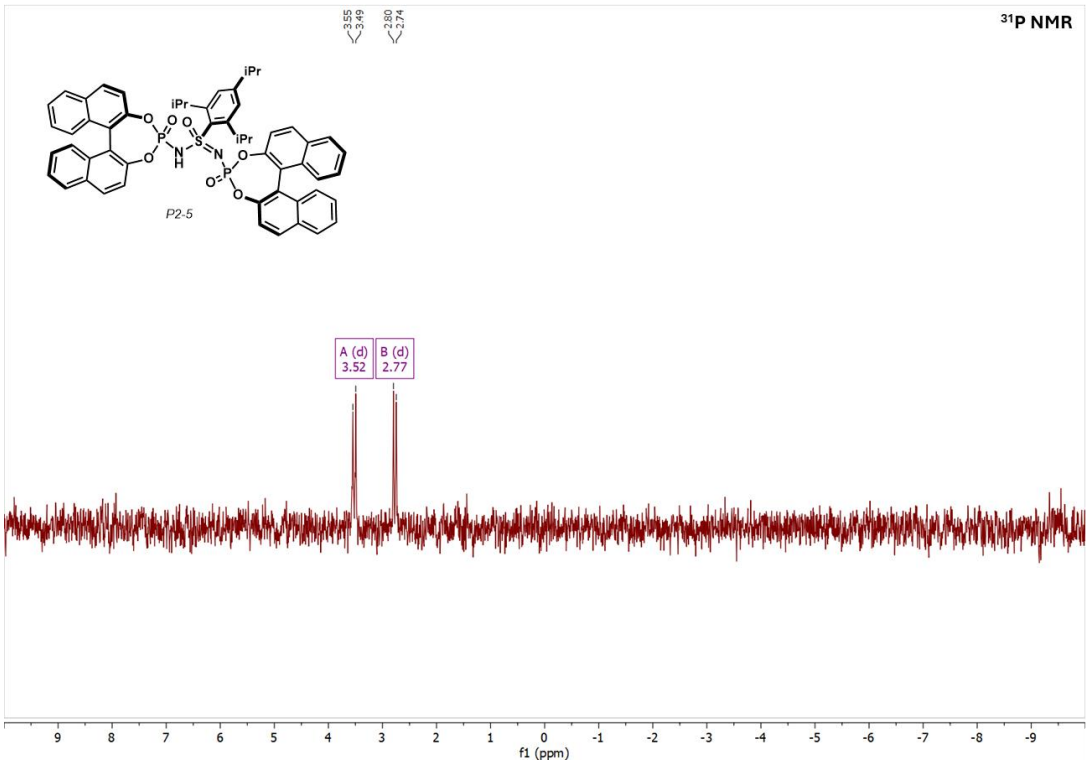

^{13}C NMR
P2-4
f1 (ppm)

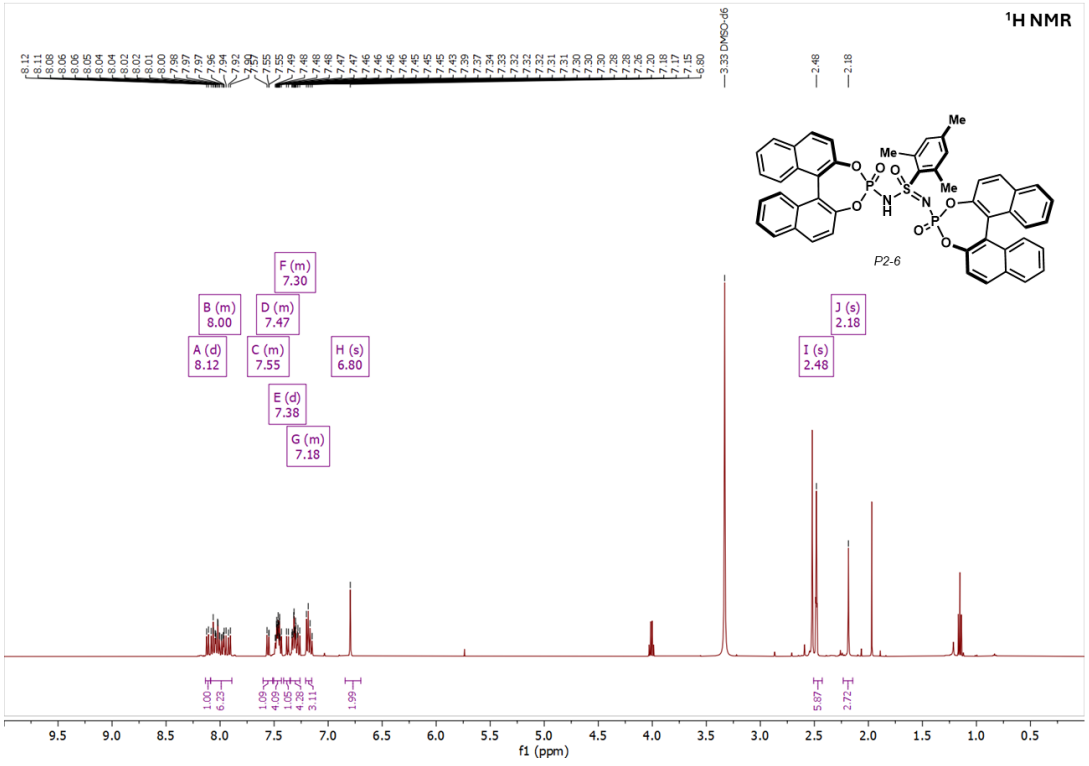

19F NMR
-61.30
A (s)
-61.30
P2-4
1.00
f1 (ppm)

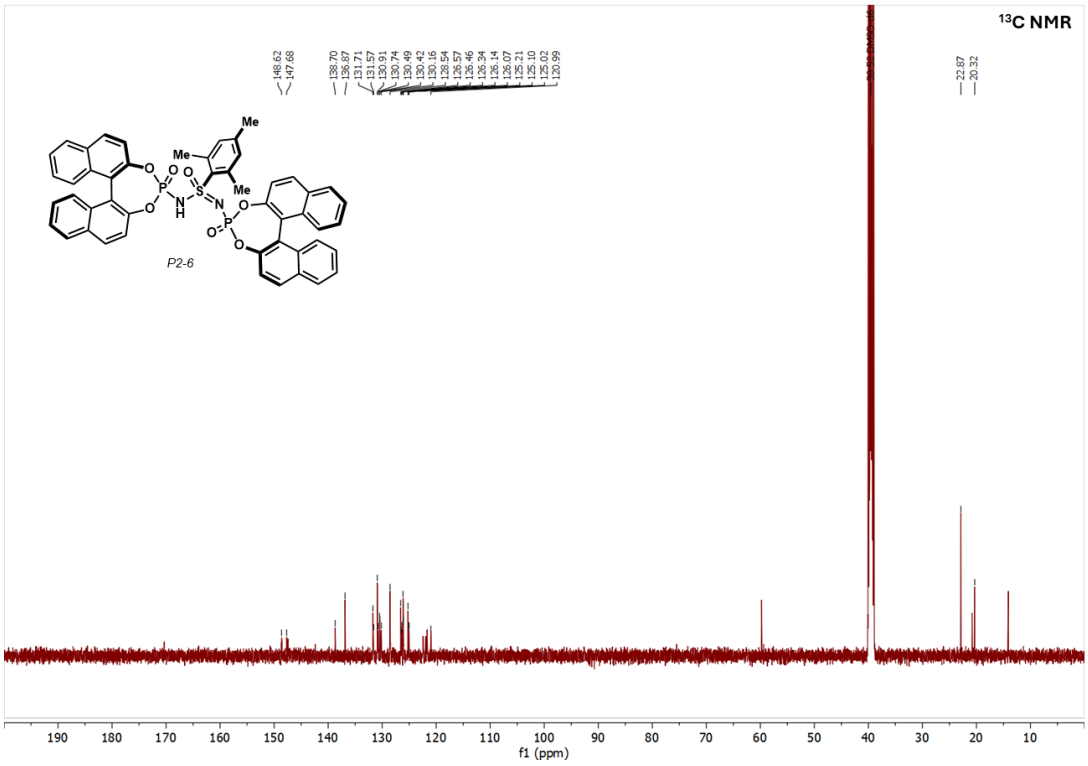

31P NMR
2.62
2.60
2.34
2.32
B (d)
2.33
A (d)
2.61
P2-4
0.97
1.03
f1 (ppm)

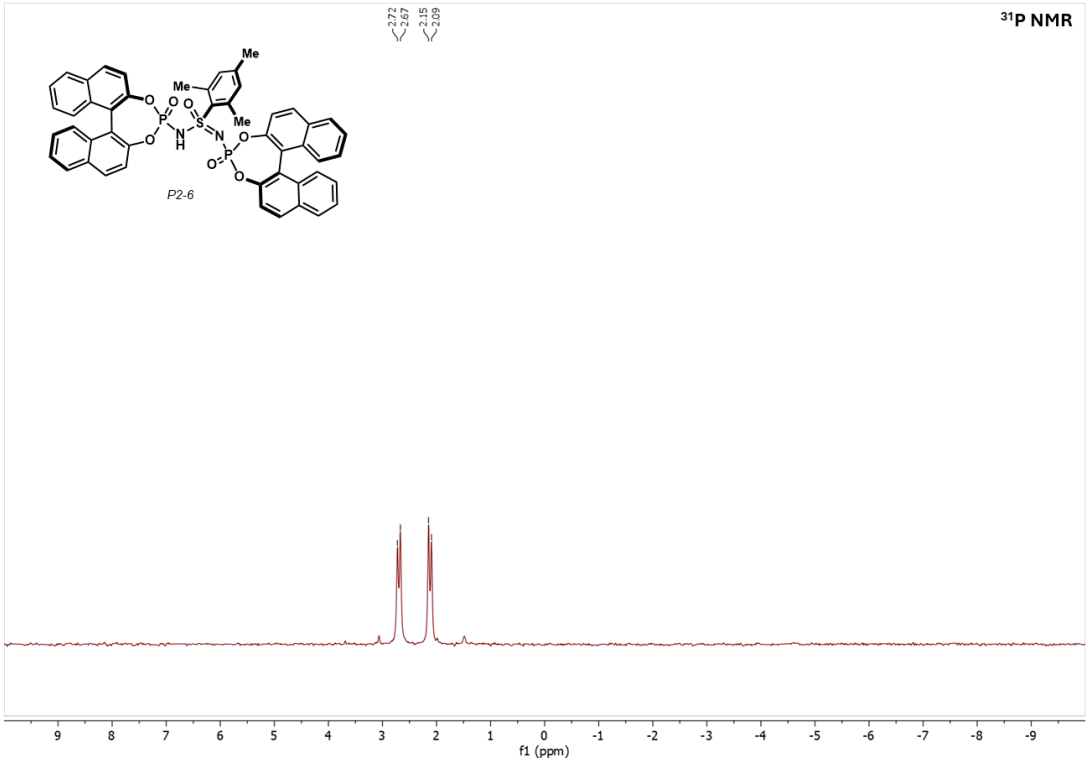

^{1}H NMR
iPr
iPr
iPr
P2-5
B (m) 8.00
E (dt) 7.42
H (s) 6.98
A (d) 8.09
D (d) 7.62
G (m) 7.18
C (d) 7.92
F (qd) 7.29
I (m) 4.47
J (m) 2.82
L (d) 0.95
K (dd) 1.17
M (d) 0.91
f1 (ppm)

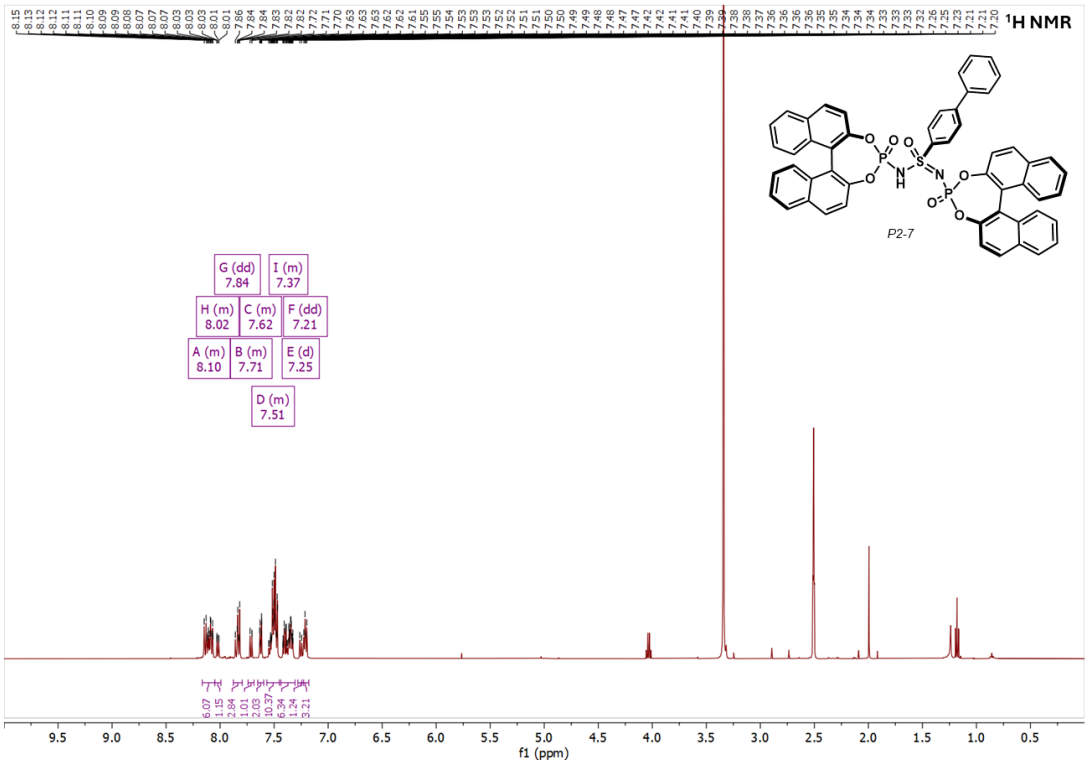

^{13}C NMR
iPr
iPr
iPr
P2-5
f1 (ppm)

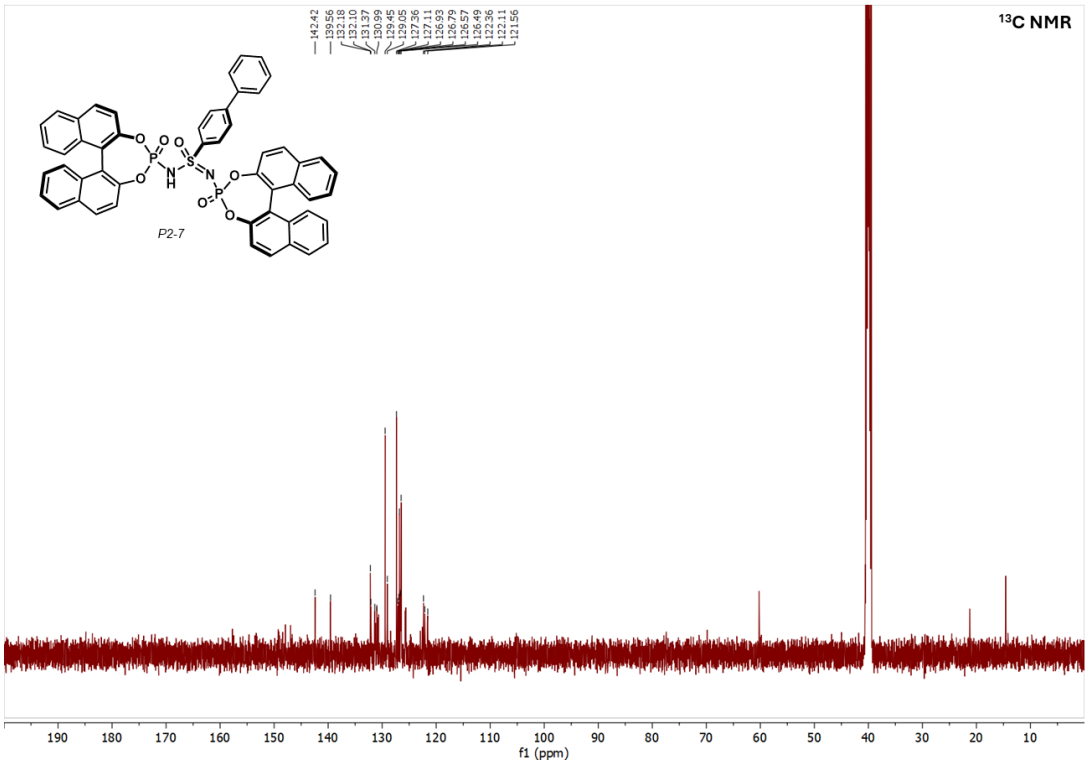

31P NMR
3.55
3.49
2.80
2.74
iPr
iPr
iPr
P2-5
A (d)
3.52
B (d)
2.77
f1 (ppm)

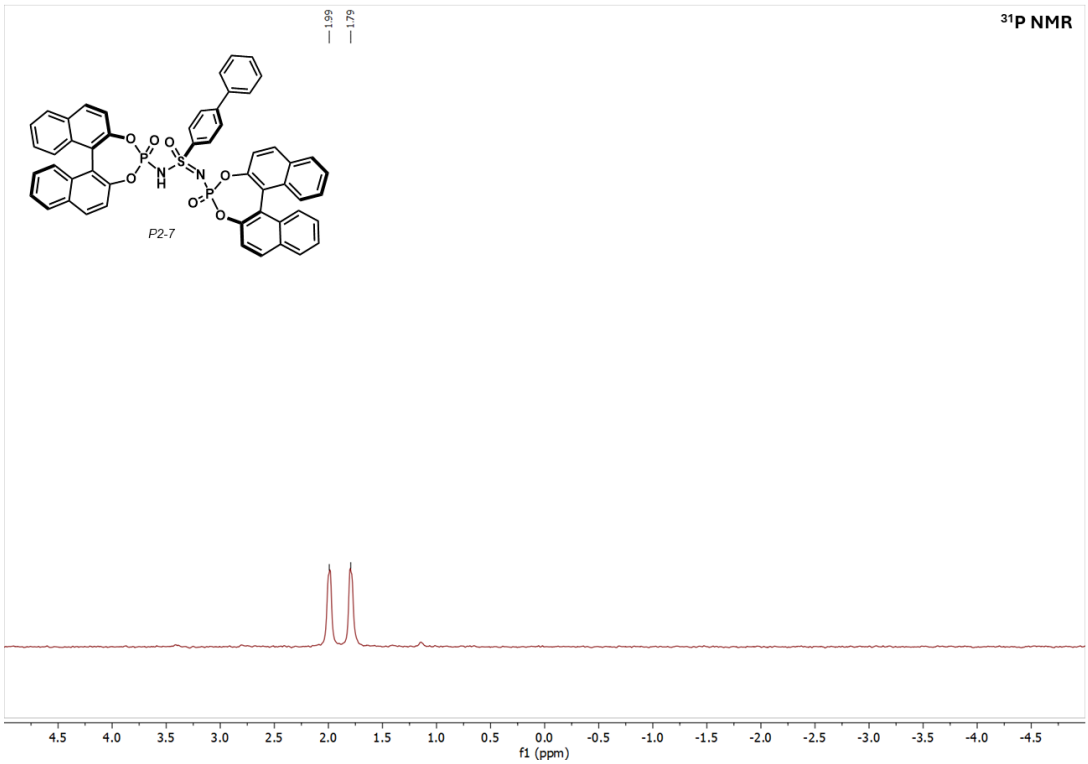

1H NMR
3.33 DMSO-d6
2.48
2.18
Me
Me
Me
P2-6
F (m)
7.30
B (m)
8.00
D (m)
7.47
A (d)
8.12
C (m)
7.55
H (s)
6.80
E (d)
7.38
G (m)
7.18
J (s)
2.18
I (s)
2.48
f1 (ppm)

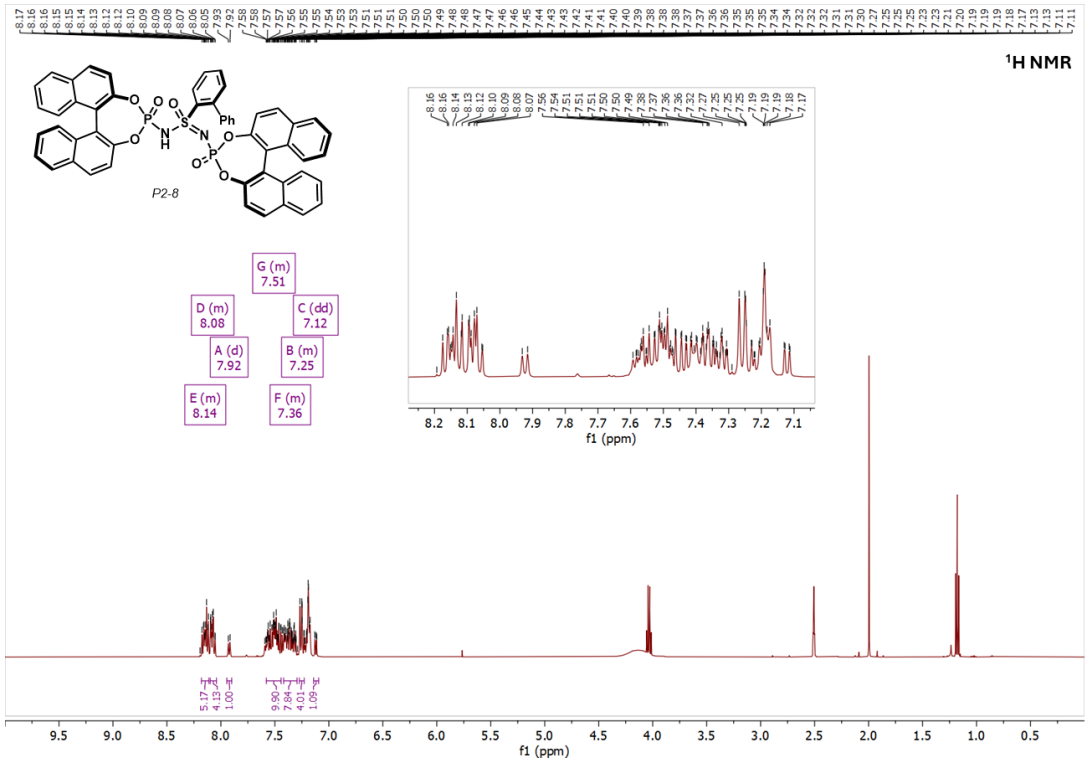

13C NMR
148.62
147.68
138.70
136.87
131.71
131.57
130.91
130.74
130.49
130.42
130.16
128.54
126.57
126.46
126.34
126.14
126.07
125.21
125.10
125.02
120.99
22.87
20.32
Me
Me
Me
P2-6
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10
f1 (ppm)

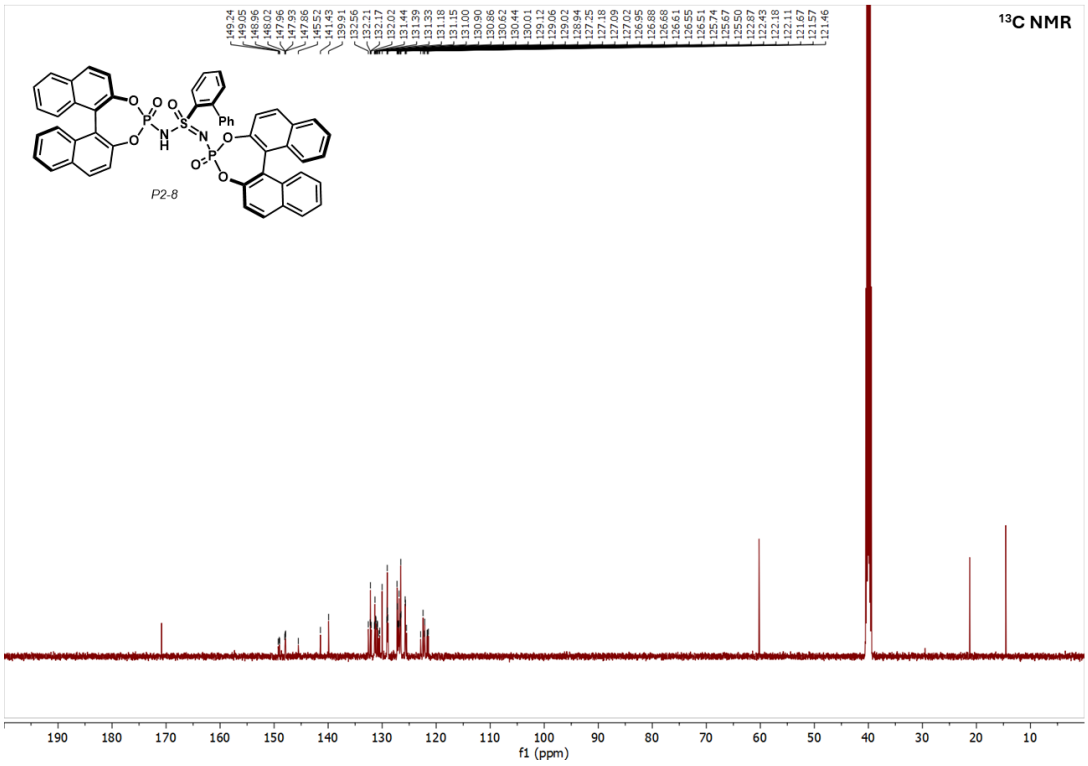

31P NMR
2.72
2.67
2.15
2.09
Me
Me
Me
P2-6
9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 -9
f1 (ppm)

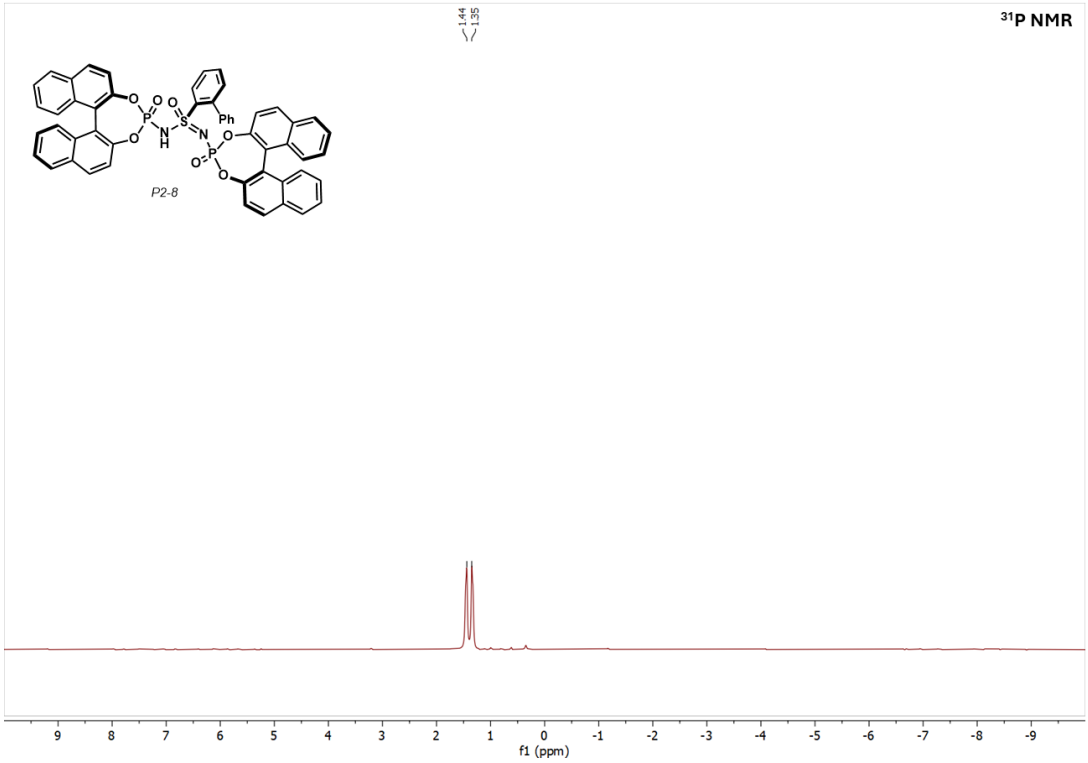

¹H NMR
P2-7
G (dd) 7.84
I (m) 7.37
H (m) 8.02
C (m) 7.62
F (dd) 7.21
A (m) 8.10
B (m) 7.71
E (d) 7.25
D (m) 7.51
f1 (ppm)

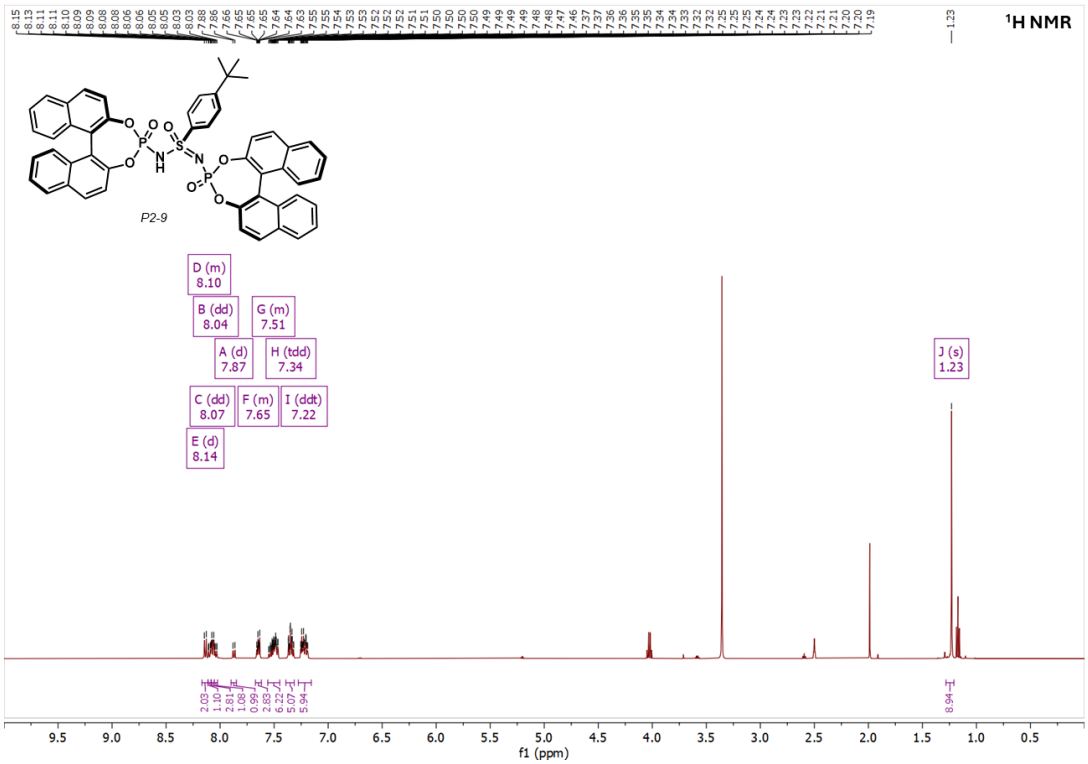

¹³C NMR
P2-7
f1 (ppm)

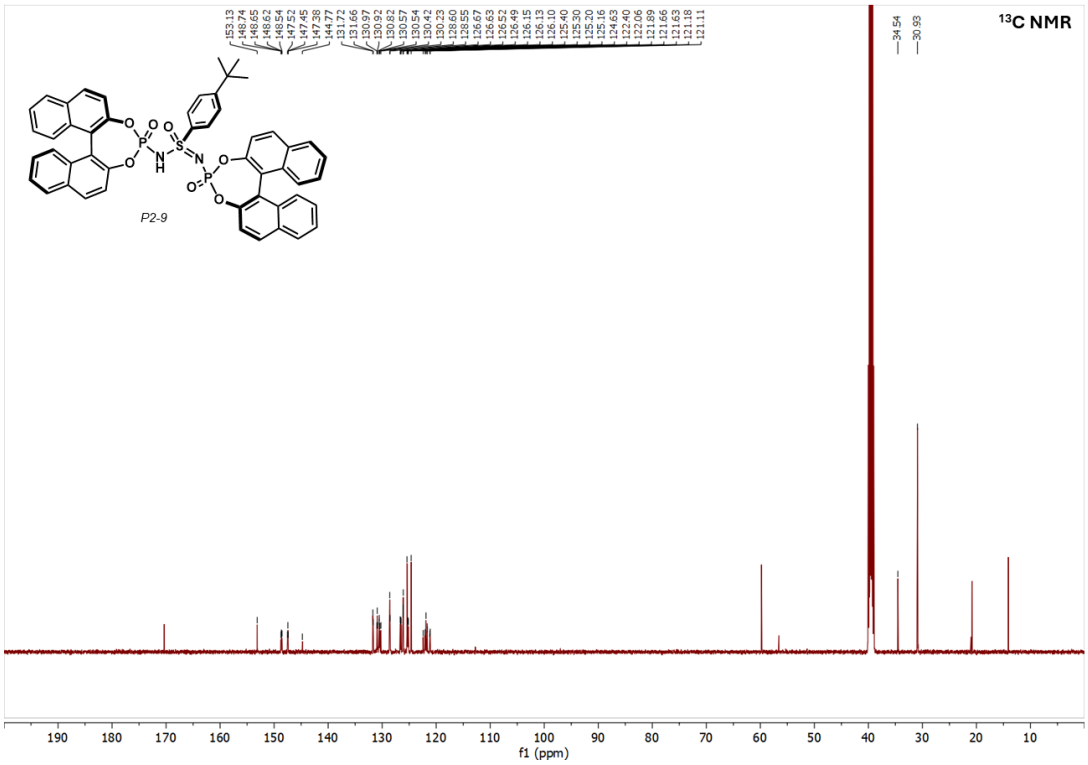

31P NMR
199
179
P2-7
f1 (ppm)

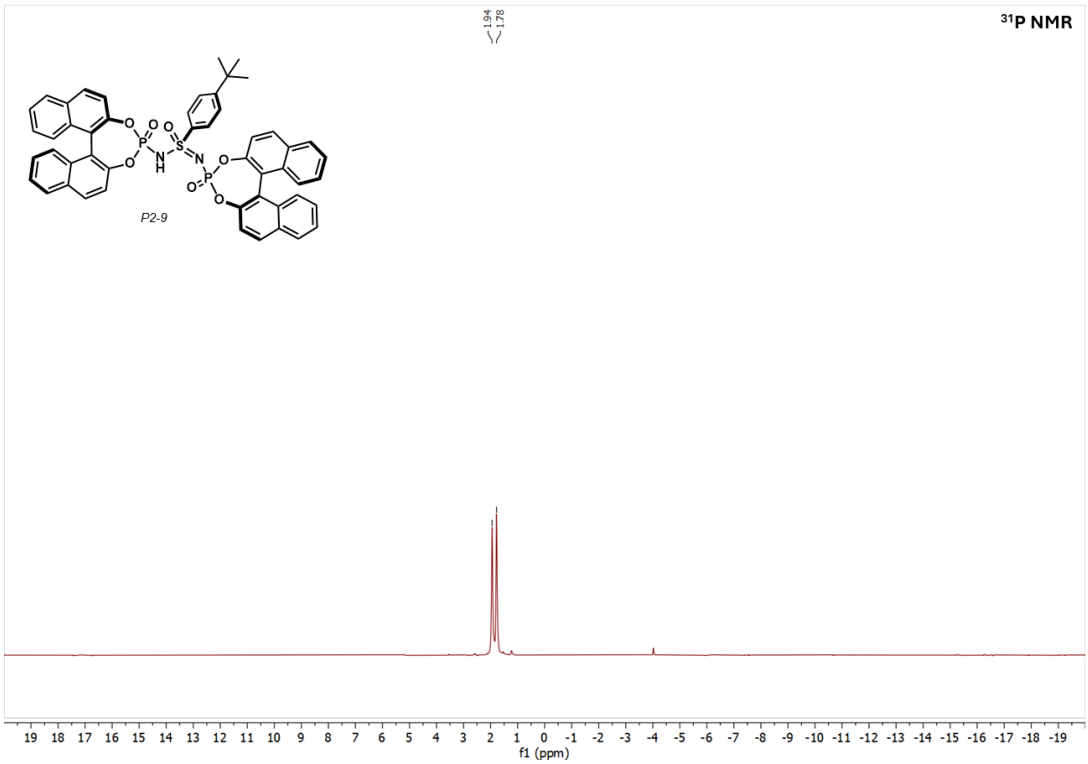

1H NMR
P2-8
G (m) 7.51
D (m) 8.08
C (dd) 7.12
A (d) 7.92
B (m) 7.25
E (m) 8.14
F (m) 7.36
f1 (ppm)

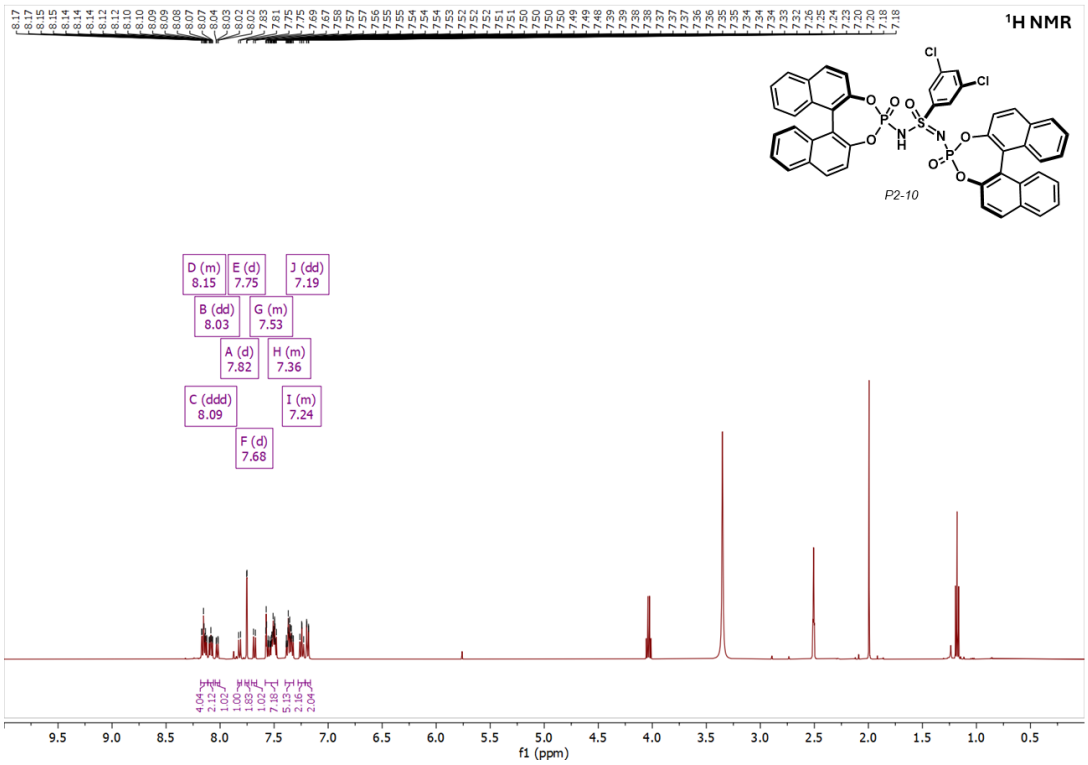

13C NMR
P2-8
f1 (ppm)

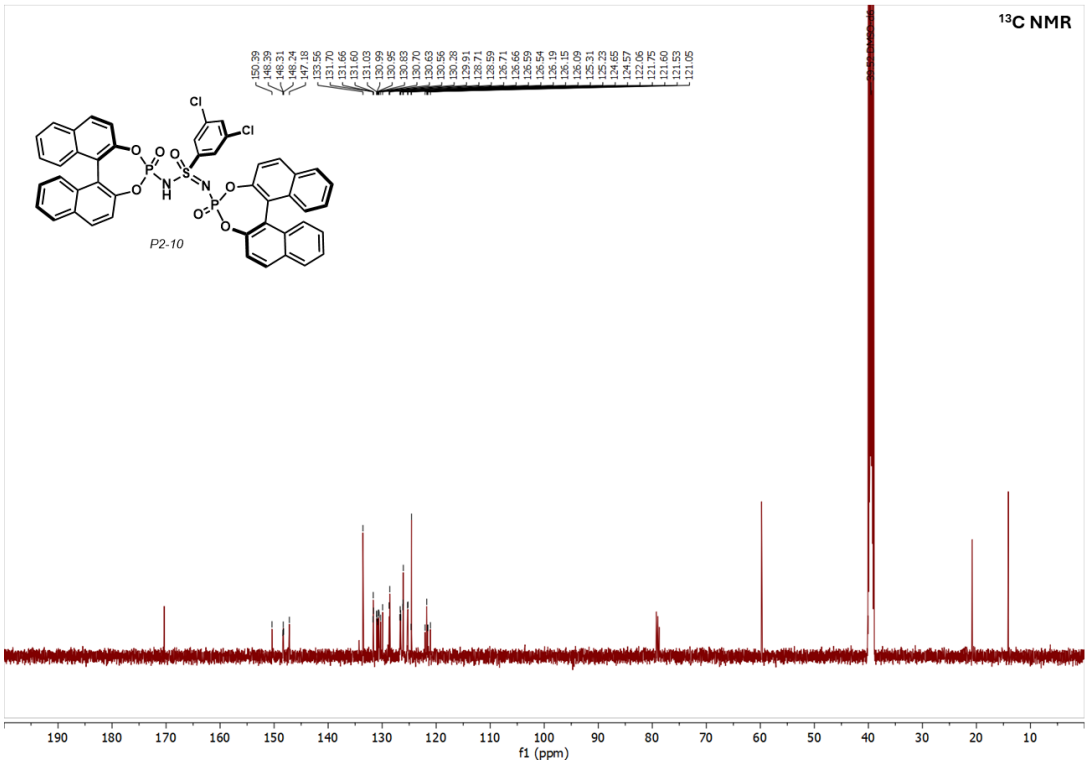

31P NMR
1.44
1.35
P2-8
f1 (ppm)

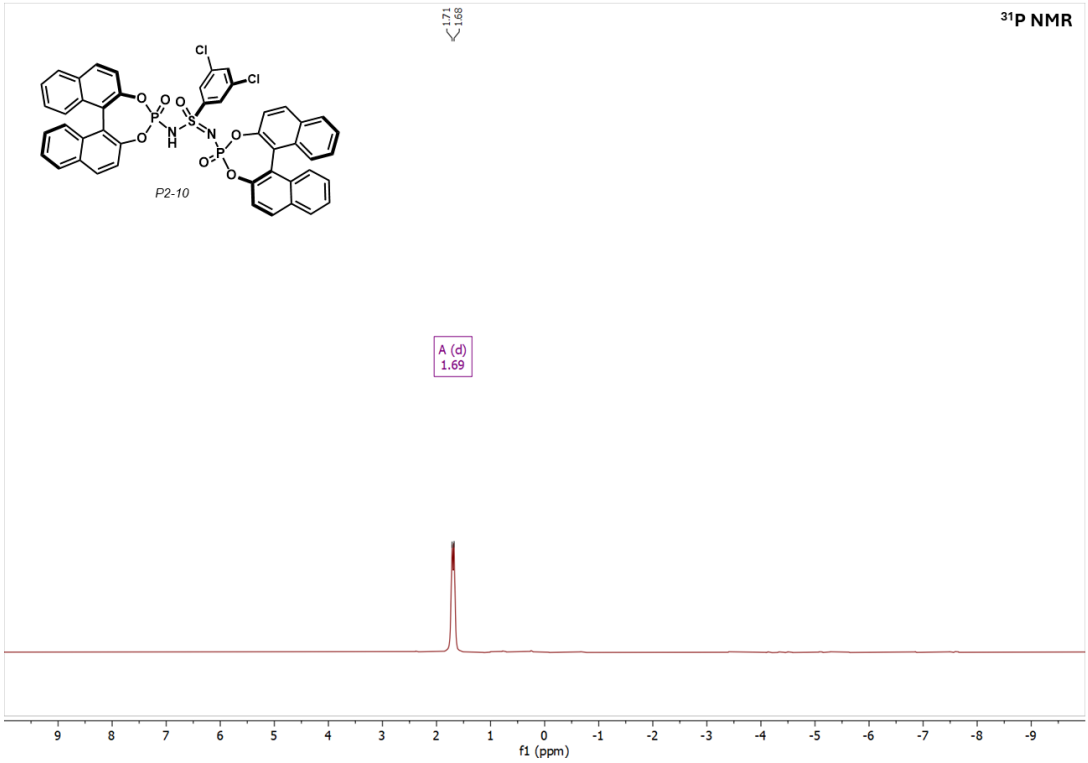

¹H NMR
P2-9
D (m) 8.10
B (dd) 8.04
G (m) 7.51
A (d) 7.87
H (tdd) 7.34
C (dd) 8.07
F (m) 7.65
I (ddt) 7.22
E (d) 8.14
J (s) 1.23
f1 (ppm)

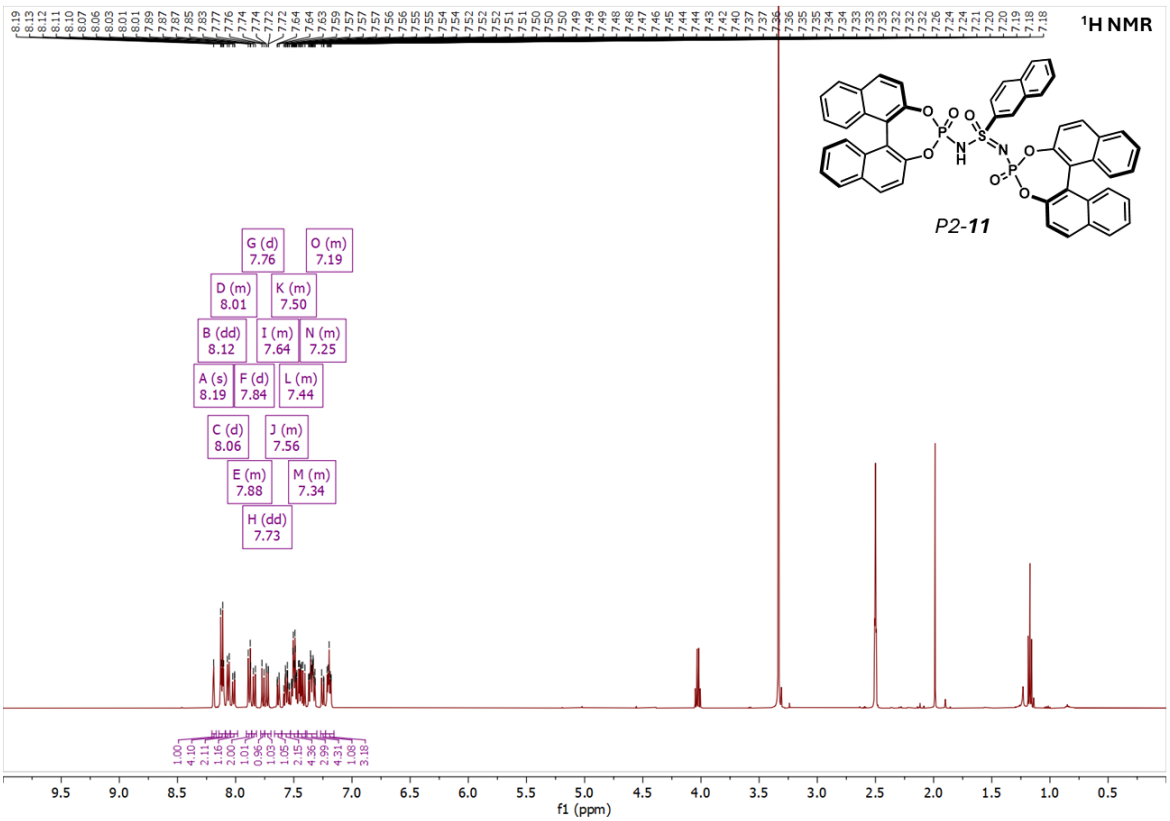

¹³C NMR
P2-9
f1 (ppm)

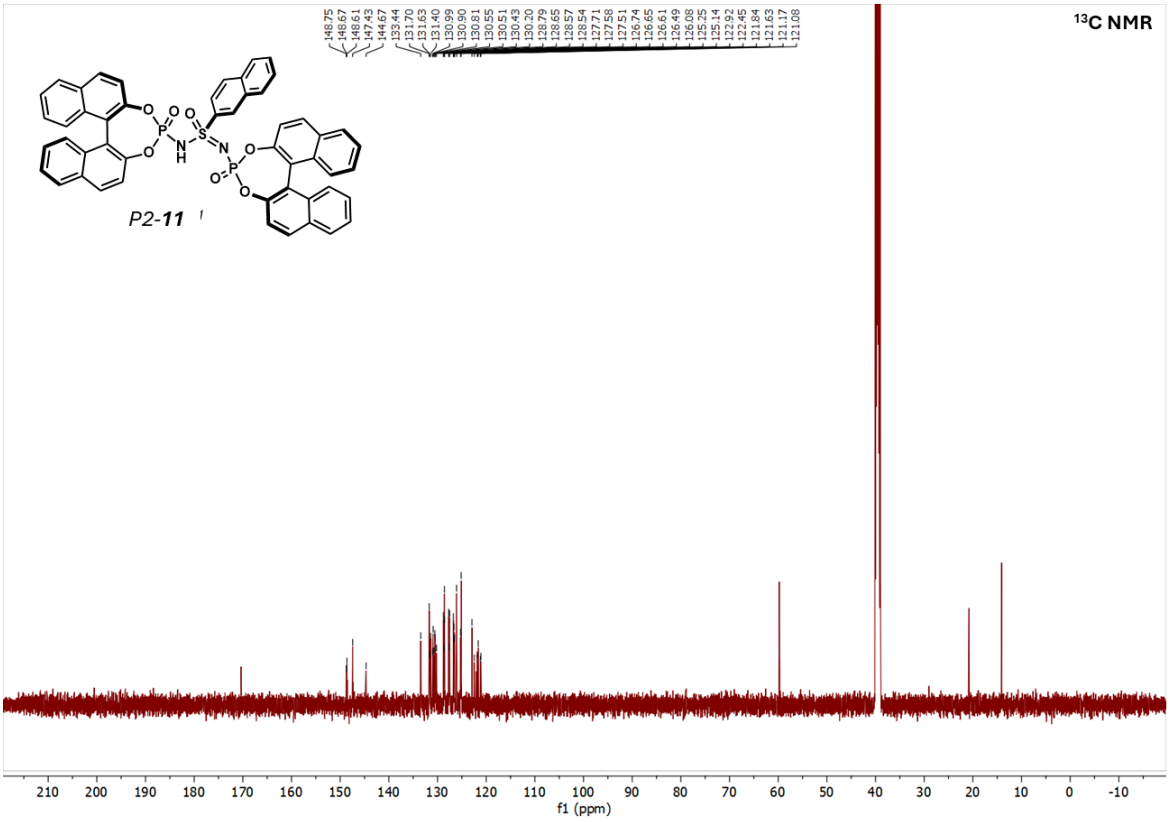

31P NMR
194
178
P2-9
f1 (ppm)

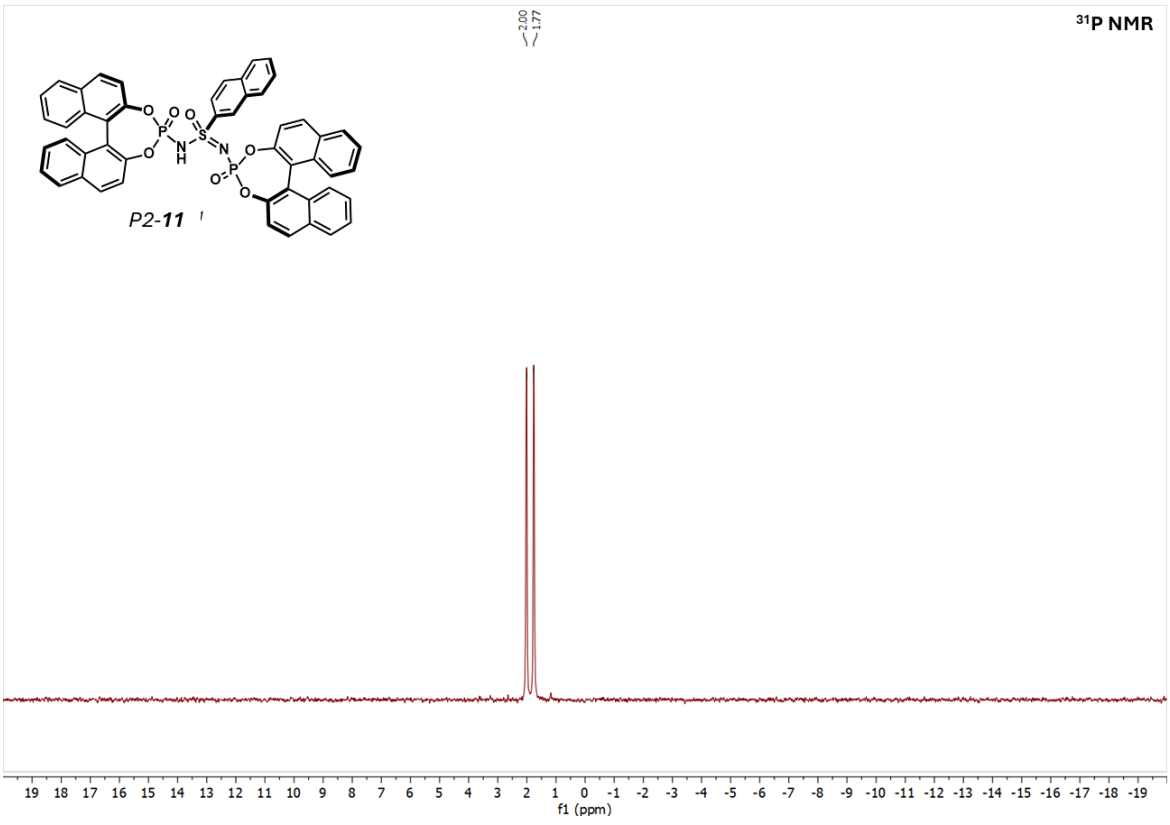

1H NMR
P2-10
D (m) 8.15
E (d) 7.75
J (dd) 7.19
B (dd) 8.03
G (m) 7.53
A (d) 7.82
H (m) 7.36
C (ddd) 8.09
I (m) 7.24
F (d) 7.68
f1 (ppm)

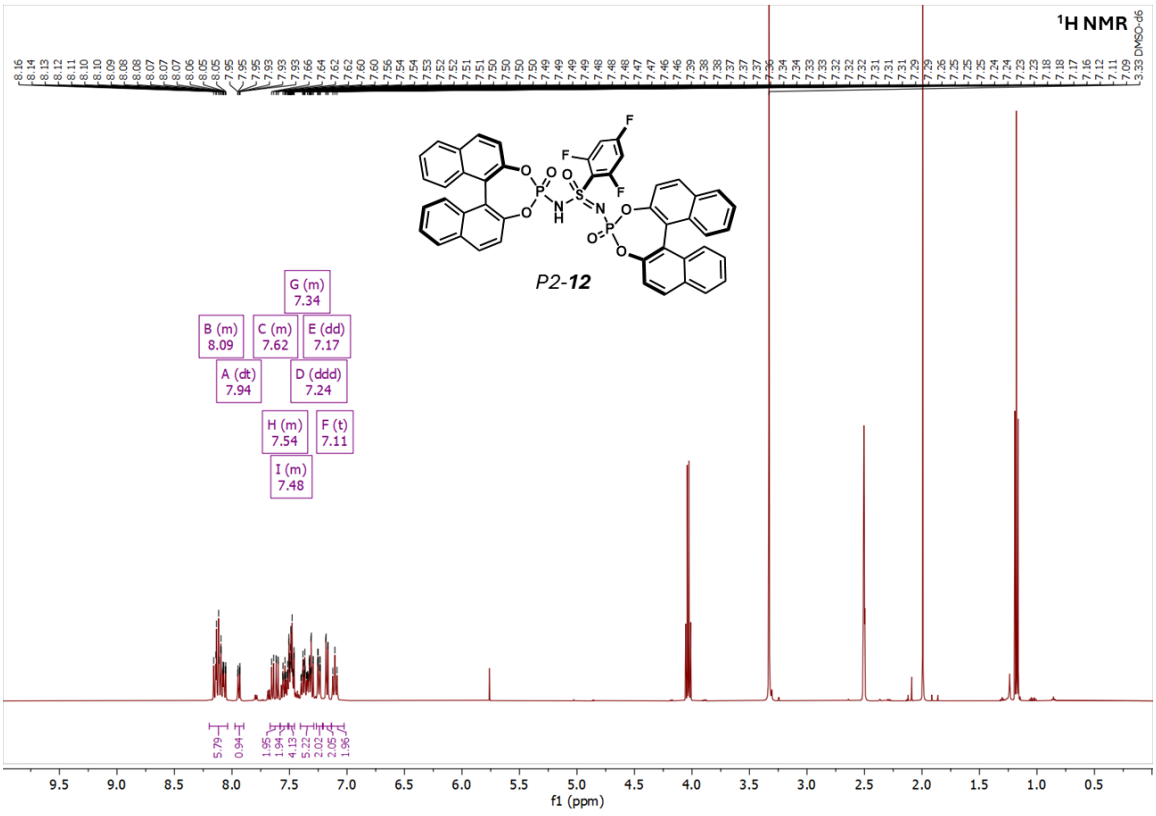

^{13}C NMR
P2-10
f1 (ppm)

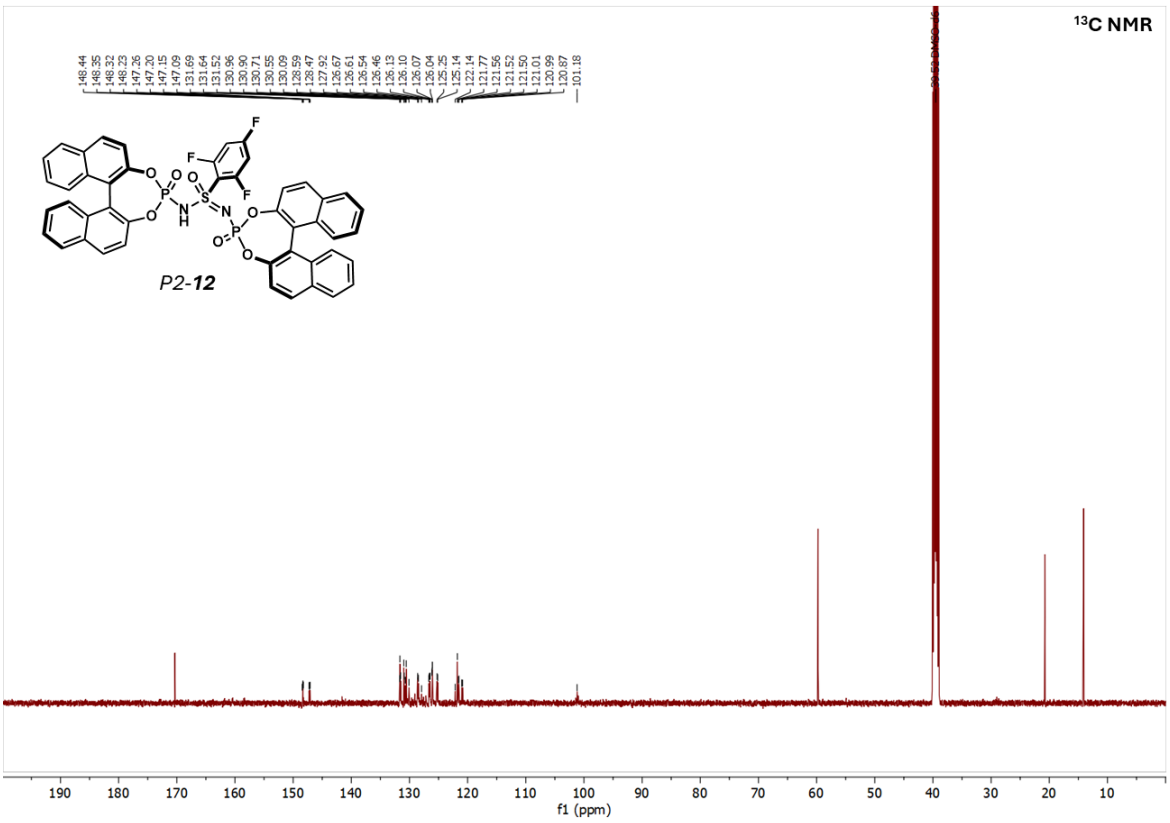

^{31}P NMR
1.71
1.68
P2-10
A (d)
1.69
f1 (ppm)

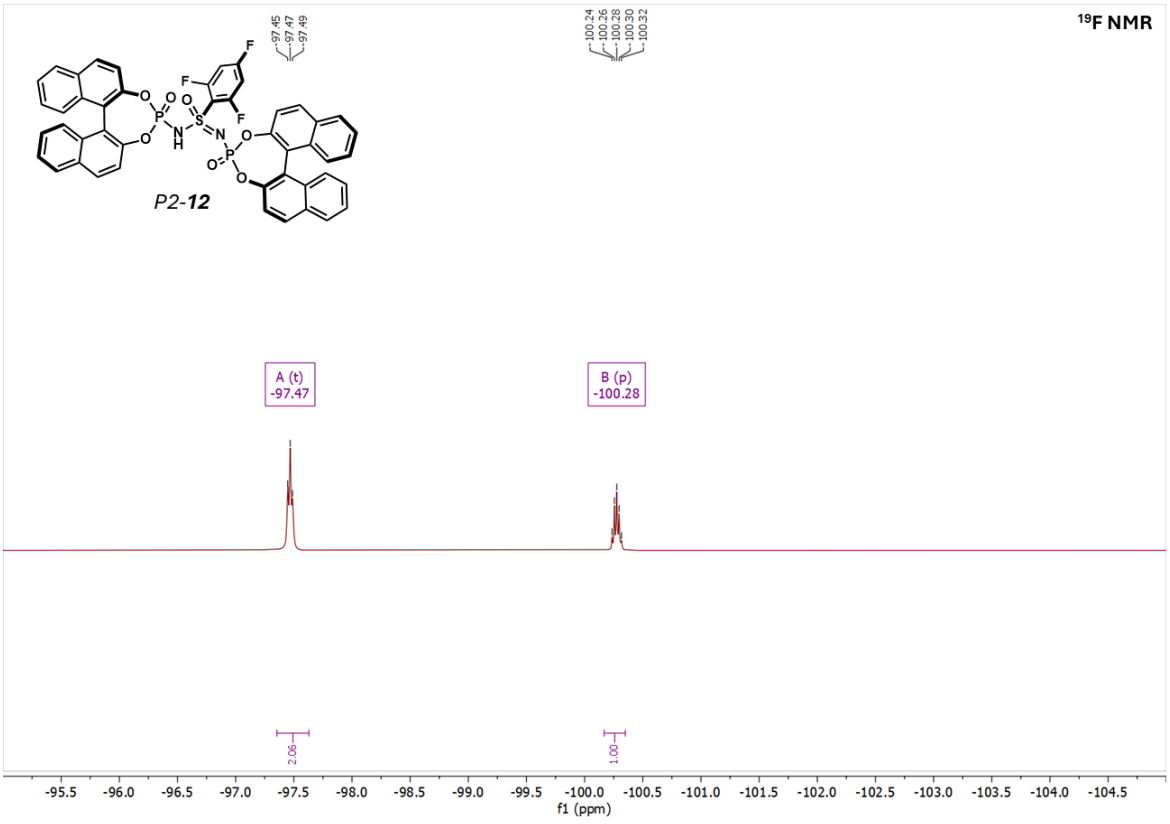

^{1}H NMR
P2-11
G (d) 7.76
O (m) 7.19
D (m) 8.01
K (m) 7.50
B (dd) 8.12
I (m) 7.64
N (m) 7.25
A (s) 8.19
F (d) 7.84
L (m) 7.44
C (d) 8.06
J (m) 7.56
E (m) 7.88
M (m) 7.34
H (dd) 7.73
f1 (ppm)

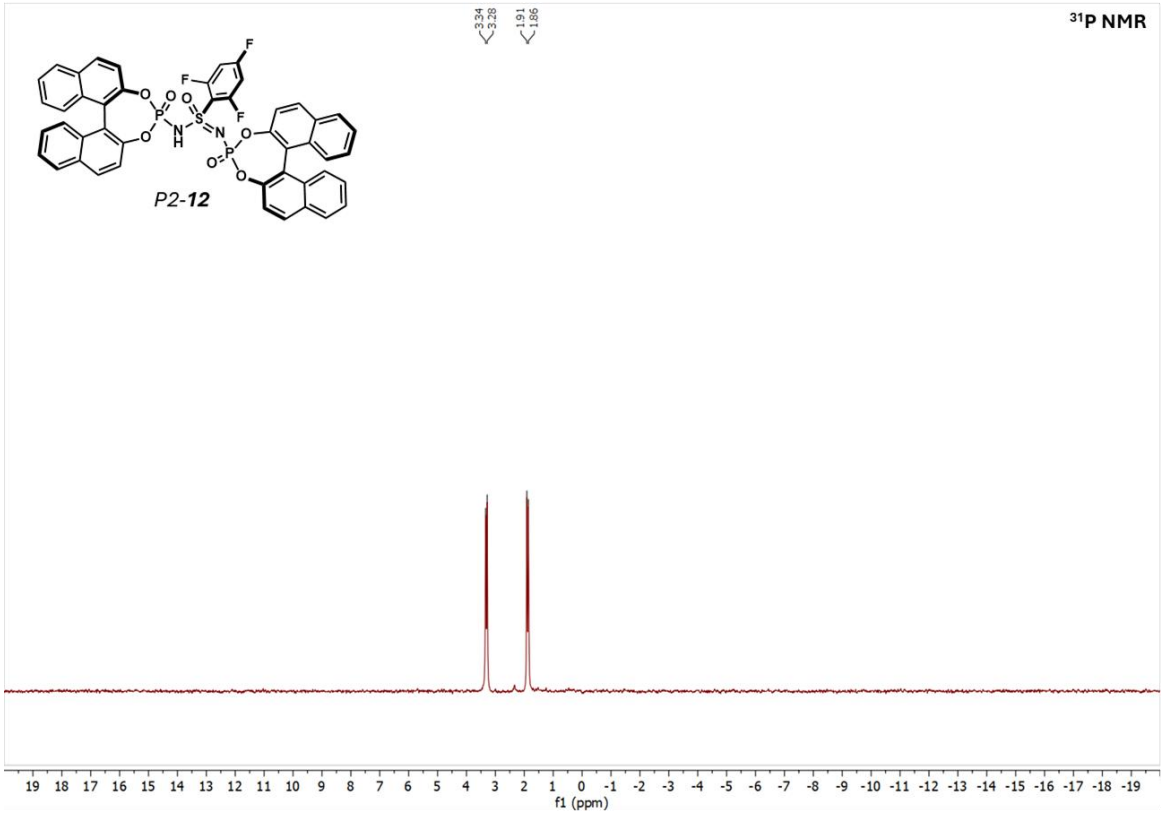

^{13}C NMR
P2-11
f1 (ppm)

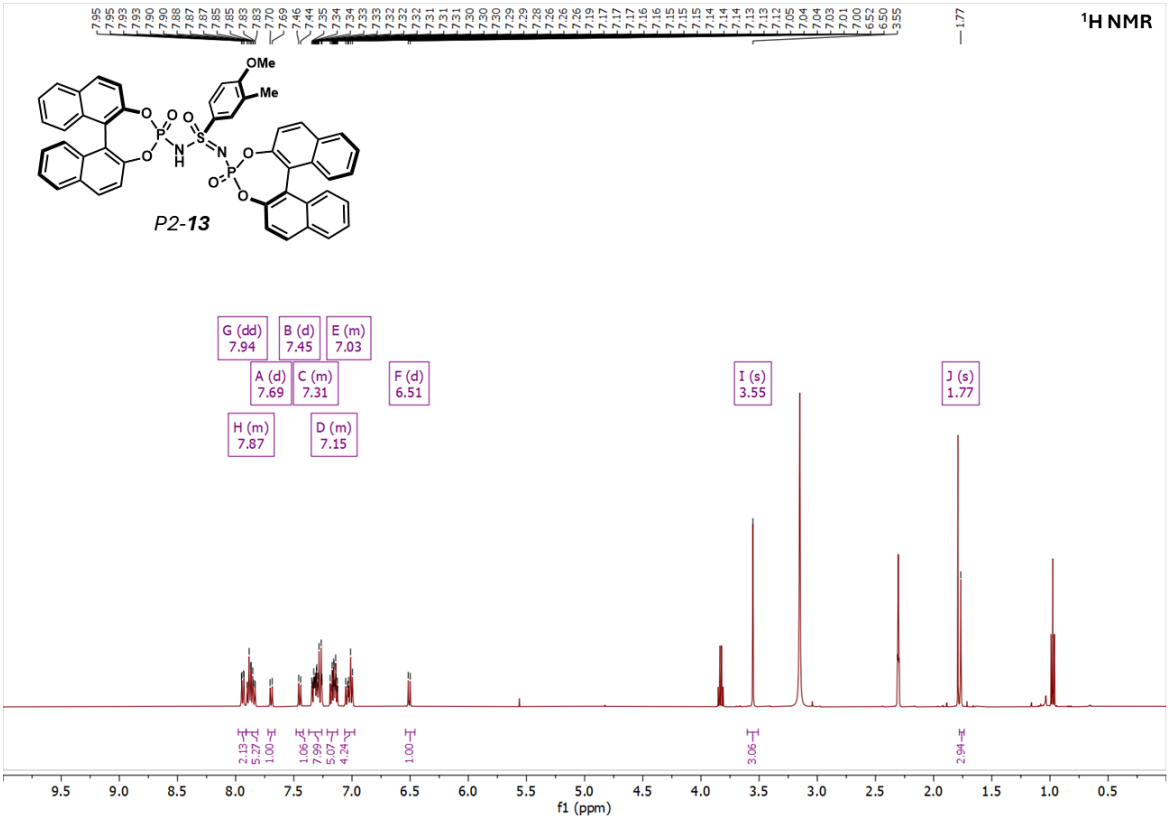

31P NMR
P2-11
f1 (ppm)

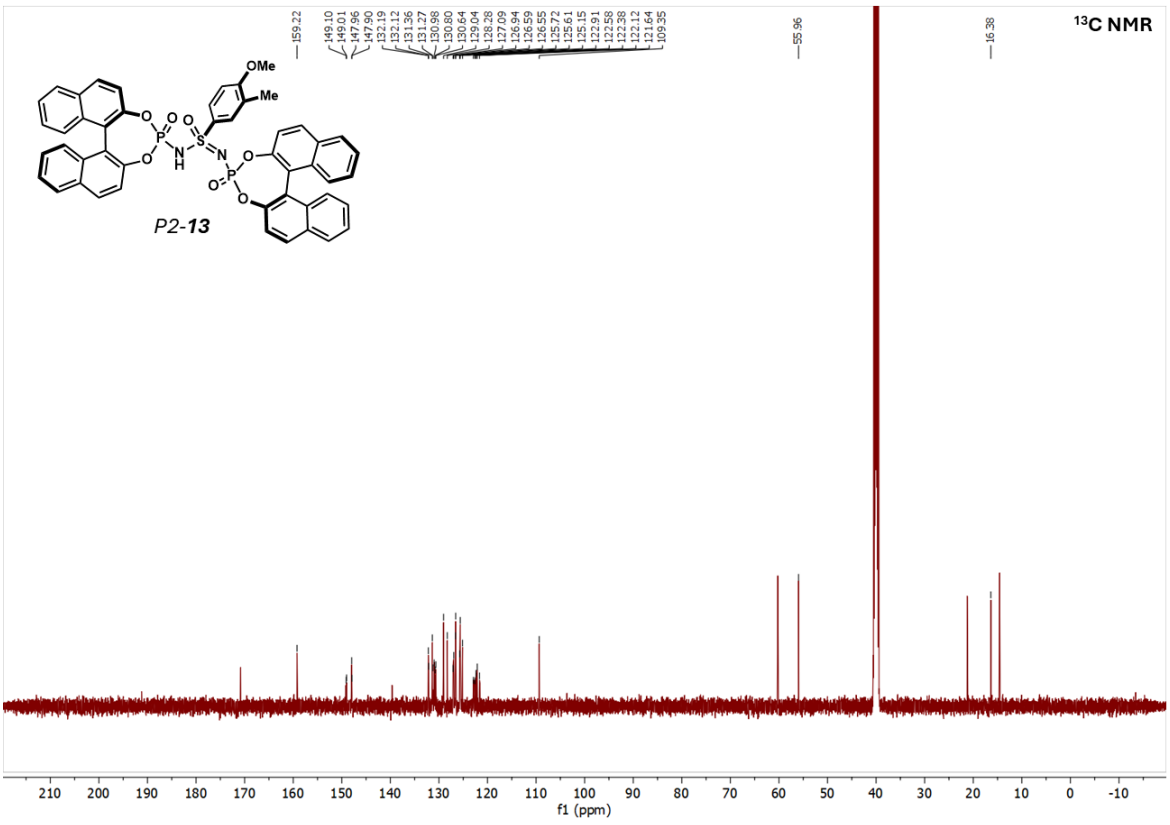

1H NMR
P2-12
G (m) 7.34
B (m) 8.09
C (m) 7.62
E (dd) 7.17
A (dt) 7.94
D (ddd) 7.24
H (m) 7.54
F (t) 7.11
I (m) 7.48
f1 (ppm)

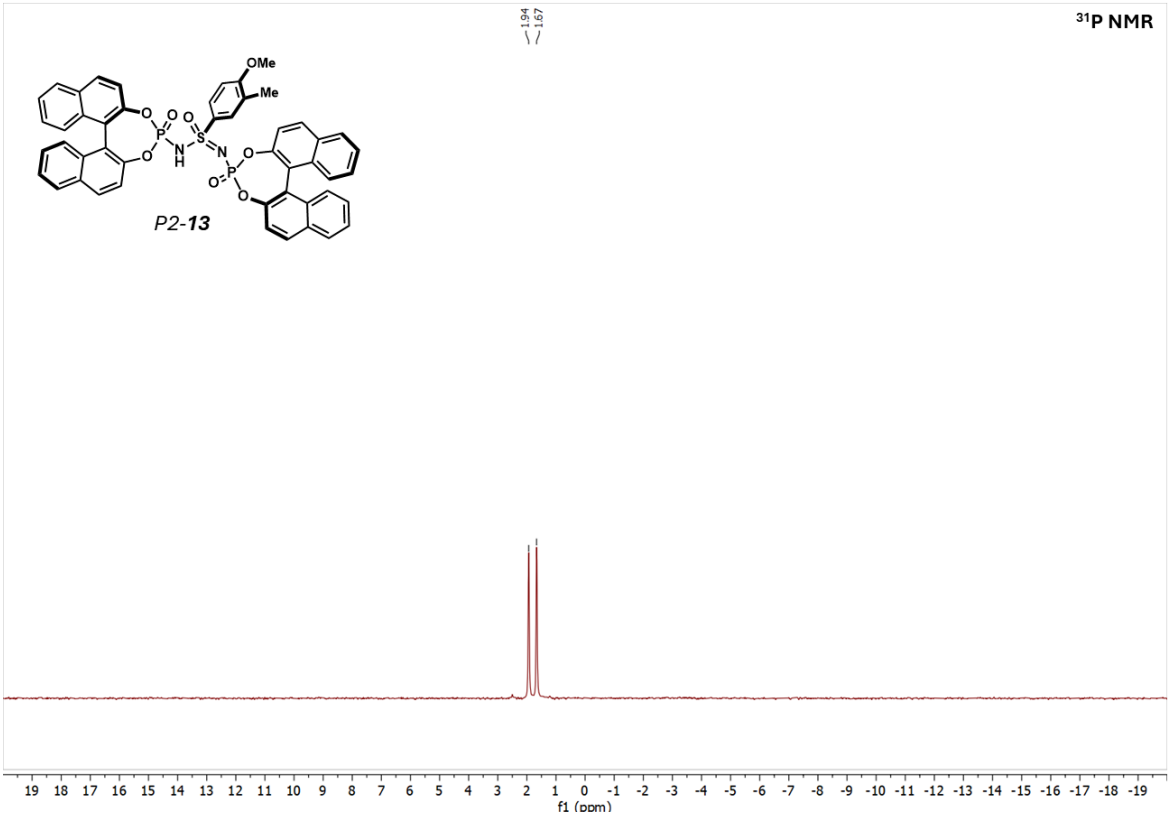

^{13}C NMR
148.44
148.35
148.32
148.23
147.26
147.20
147.15
147.09
131.69
131.64
131.52
130.96
130.90
130.71
130.55
130.09
128.59
128.47
127.92
126.67
126.61
126.54
126.46
126.13
126.10
126.07
126.04
125.25
125.14
122.14
121.77
121.56
121.52
121.50
121.01
120.99
120.87
101.18
P2-12
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)


^{19}F NMR
-97.45
-97.47
-97.49
-100.24
-100.26
-100.28
-100.30
-100.32
P2-12
A (t)
-97.47
B (p)
-100.28
2.06
1.00
-95.5
-96.0
-96.5
-97.0
-97.5
-98.0
-98.5
-99.0
-99.5
-100.0
-100.5
-101.0
-101.5
-102.0
-102.5
-103.0
-103.5
-104.0
-104.5
f1 (ppm)


31P NMR
P2-12
f1 (ppm)


1H NMR
P2-13
G (dd) 7.94
B (d) 7.45
E (m) 7.03
A (d) 7.69
C (m) 7.31
F (d) 6.51
I (s) 3.55
J (s) 1.77
H (m) 7.87
D (m) 7.15
f1 (ppm)


^{13}C NMR
159.22
149.10
149.01
147.96
147.90
132.19
132.12
131.36
131.27
130.98
130.80
130.64
129.04
128.28
127.09
126.94
126.59
126.55
125.72
125.61
125.15
125.01
122.58
122.38
122.12
121.64
109.35
55.96
16.38
OMe
Me
P2-13
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


^{31}P NMR
1.94
1.67
OMe
Me
P2-13
19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10 -11 -12 -13 -14 -15 -16 -17 -18 -19
f1 (ppm)

Appendix V – Supplementary Information for Chapter V

Portions of this appendix were adapted from the Supplementary Information of:

Smith, A. L.; Wei, N.; Toste, F. D. *Manuscript in preparation.*

Table of Contents:

V.1 – Specific Comments on Materials and Methods **453**
V.2 – Extended Discussion on Fig. V-2 ..... **455**
V.3 – Extended Discussion on Fig. V-3 ..... **459**
V.4 – Extended Discussion on Fig. V-4 ..... **468**
V.5 – Extended Discussion on Fig. V-5 ..... **476**
V.6 – Synthetic Methods **482**
V.7 – References **499**
V.8 – NMR Spectra **501**

VI.41 – Specific Comments on Materials and Methods

Literature-derived datasets were previously compiled and reported. These were used as reported without additional catalysts calculations or processing (see **Appendix III-IV** for details).

- ***General parameters for Bayesian optimization across literature-derived datasets***

Bayesian optimization campaign over literature-derived datasets sought to quickly select the *best* possible choice (max. DDG, kcal/mol) from each reported set of catalysts (see **Fig. V-2** in the main text). Simulations were conducted with specific sets of representation for the BAOC: random (also reflects one-hot encoding, OHE), active site-based buried volume, sterimol, and calculated pKa. These representations were deployed from the complete *category*. For example, sterimol includes sterimol L, BMin, and BMax across various ensembles (as previously reported). Each of these values were provided to the optimizer, without any additional down selection.

The Bayesian optimizer employed a Gaussian process regressor (GPR) *surrogate model* (Matérn kernel, υ = 2.5 with additive noise kernel, 10^{-5}) with target normalization enabled and hyperparameter optimization disabled. At each evaluation, the *acquisition function* (expected improvement, EI with exploration, ξ = 0.01) ranked the most promising available candidates, (max. EI). Models were initialized with a single random point. Performance on each dataset was measured from the aggregate behavior across 100 random simulations. Overall metrics recorded included hit rate (ex. did the model select the *best* option), EI *at* hit (ex. did the model expect the

best option to be promising), and EI *after* hit (ex. did the model expect remaining options to be promising after selecting the *best* possible option).

- ***General approach for rediscovery campaigns***

To promote engagement with the approach, rediscovery campaigns leveraged the open-source EDBO+ Bayesian optimization package reported by the Doyle group rather than in-house Bayesian optimization scripts (see **Fig. V-3** in the main text, web-app available).[1] The package was used as provided from GitHub (original, 2022-based commits). The active learning campaign sought to minimize the distance between two points in space, defined by normalized Euclidean distance across all points computed. Here, a point is defined by a vectorial representation of active site-based buried volumes (2-7 Å buried volume of a specific ensemble). Target BAOCs are represented by their buried volume, such that a *successful* rediscovery corresponds to realizing a norm. Euclidean distance = 0.

For each selection, the surrogate model is provided an ensemble of *all* possible substituents (defined by diverse array of descriptors characterizing aryl bromide surrogates, previously reported by our group in **Appendix III**[2]). The acquisition function then selects the top-8 most promising substituents expected to reduce the norm. Euclidean distance (note – EDBO+ leverages qEHVI,[3] expected hypervolume improvement to accommodate multi-objective optimization). These substituents are mapped onto a CPA scaffold and the active site-based buried volume features are computed (see **Appendix III** for calculation workflow[2]). After, the calculated buried volumes are returned to the model to evaluate proximity to target, update expectations, and select the next batch of targets. Batch selection was repeated up to 6 times to accommodate a strict, computational budget (each batch requires ~24 hr for calculation – optimization, conformational search, processing). This search (up to 56 choices) samples <2% of all substituents possible. Rediscovery campaigns were conducted in quadruplicate to quantify variance stemming from random initialization. Identical random starts were used when simulating rediscovery with different representations, minimizing variance in available information.

Molecular similarity is quantified by Tanimoto similarity between the target and selected substituents, as represented structurally by their aryl bromide surrogates. This similarity metric was used as implemented in RDKit.[4]

- ***General approach for in silico, active learning campaign for inverse design***

The same active learning framework (EDBO+, closed-loop calculation cycle) for the rediscovery campaigns was deployed for the nomination of catalyst candidates in chemical space. To compute P1-SOAs, aryl bromides were geometrically transformed onto the following *dummy* scaffold (dummy atom corresponds to modular substitution, **bolded**) akin to our previous workflows (see **Appendix III**, **IV**):

P1SOA_H

C	-0.18430908	3.28697332	3.13678253	H	-2.81275134	0.85279046	5.83181671
C	-1.48862612	3.60699423	2.92810405	C	-2.49031067	-0.04010723	6.32544181
C	-2.51158946	2.81814669	3.48252910	C	-1.66336148	-2.35606048	7.62320639
C	-2.19998699	1.64681756	4.17828158	H	0.66123914	-3.31387862	6.75761157
C	-0.86405098	1.26124185	4.31227731	C	-2.81733280	-1.73926050	7.98438941
C	0.14628793	2.15974734	3.90031621	C	-3.23530655	-0.56890691	7.32858923
H	-4.10059212	4.08981999	2.80409124	H	-1.35127327	-3.24581774	8.12900573
H	-1.73817127	4.46843674	2.34454256	H	-3.41287984	-2.14707453	8.77427005
C	-3.85163156	3.19554729	3.33625191	H	-4.14676004	-0.09155779	7.62234140
C	-3.22708415	0.86513741	4.71857376	O	1.50231571	1.93198560	4.31664843
C	-4.51965466	1.24933145	4.56821001	O	1.20828918	-0.50889017	3.20005917
C	-4.83519811	2.42700059	3.86961724	P	2.28279660	0.77749764	3.36996990
H	-2.99357171	-0.03410600	5.24935924	O	3.59158677	0.32129471	4.07735549
H	-5.30370828	0.65161730	4.98401971	S	3.68100507	2.68946811	1.84731368
H	-5.85814818	2.71994912	3.75712585	O	3.84981069	3.03911965	0.42993137
C	0.58903313	-0.87671408	4.44315659	O	3.10595427	3.83329513	2.56890948
C	-0.51304886	-0.12813070	4.91527016	N	2.63268033	1.34521542	1.98184962
C	-1.29960923	-0.66434509	5.93759416	H	1.93750298	-2.49280174	4.86470000
C	-0.88167492	-1.82807991	6.58892218	H	0.58964591	3.90310124	2.72900798
C	0.31704222	-2.45920461	6.21358876	**Bq**	**5.46887572**	**2.22689545**	**2.64081545**
C	1.03654799	-1.99557339	5.15786219	H	2.24682215	0.92775064	1.15914881

VI.42 – Extended Discussion on Fig. V-2

We discuss and report aggregate tabulated metrics for the Bayesian optimization campaigns across literature-derived datasets (as seen in **Fig. V-2** of the main text).

- ***Campaign convergence with respect to molecular representation deployed***

As discussed in **Specific Comments on Materials and Methods**, categories of molecular representations were used as previously reported: (1) buried volume, (2) sterimol, and (3) calculated $p\mathrm{K_a}$. Random choice (ex. without Bayesian inference) served as a statistical baseline to quantify *gain of function* by each representation with respect to the given experimental budgets, either: 2, 4, 6, 8, or 16 additional choices were permitted following initialization. Note – across the 50+ literature datasets compiled, the **average dataset size** was about **12 BAOCs**, while the **median dataset size** was about **9 BAOCs**. These results are visualized in the main text and tabulated below (see **Fig. V-2B**, **Supplementary Table V-X**).

Budget	Random	Buried volume		
	Hit rate	Avg. hit rate	SEM	*p*
2	0.320	0.438	0.035	0.0016
4	0.534	0.677	0.037	0.0035
6	0.726	0.839	0.029	0.0004
8	0.818	0.912	0.022	<0.0001
16	0.942	0.963	0.014	0.137

….

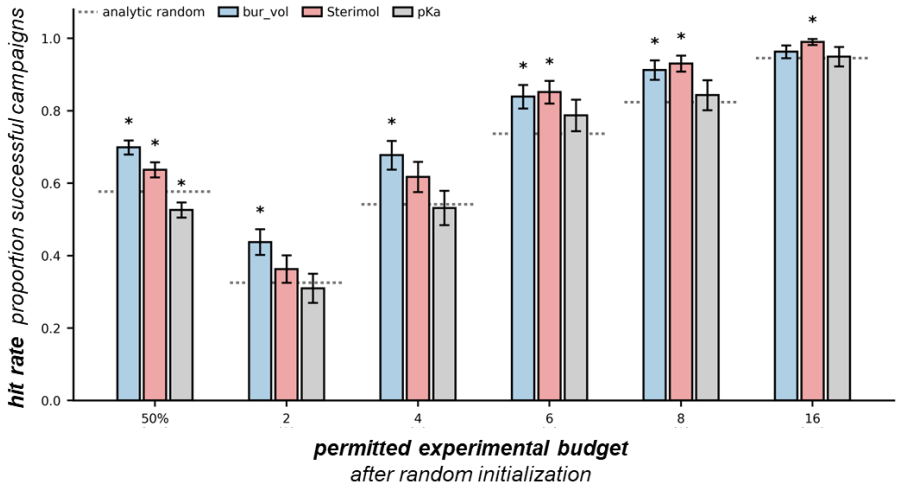
Budget	Sterimol			Calculated pKa		
	Avg. hit rate	SEM	*p*	Avg. hit rate	SEM	*p*
2	0.363	0.037	0.274	0.31	0.041	0.823
4	0.618	0.037	0.031	0.532	0.049	0.968
6	0.852	0.027	<0.0001	0.787	0.043	0.011
8	0.930	0.019	<0.0001	0.843	0.04	0.044
16	0.990	0.008	0.026	0.949	0.024	0.759

Supplementary Table VI-41 | Avg. hit rate, SEM, and *p*-value for each representation deployed across each fixed budget considered.

◊ *Proportional experimental budget recapitulates fixed budget observations*

Because each dataset has variable size, larger fixed budgets (ex. 8 and 16) may exhaust all possible options in small datasets, while smaller fixed budgets (ex. 2 and 4) may be searching a *needle in the haystack* for larger datasets. We sought to evaluate *proportional* budgets, up to 50% of the available choices, scaling with respect to size of each dataset.

Under this new exploration policy, we observe buried volume-based representations continue to furnish the most accelerated discovery of the *best* available option, while sterimol and calc. pK_a are comparatively substandard (**Supplementary Fig. V-X**, **Supplementary Table V-X**). In summary the observed utility is somewhere between fixed budget of 4-6, consistent with a 50% provided budget and the median dataset size of 9.



Supplementary Fig. VI-190 | Hit rate across all simulated campaigns v. permitted experimental budget (up to 16 additional choices after initialization). '*' refers to statistically significant difference (two-tailed *t*-test, $p < 0.05$) from random baseline (dotted gray line). Error bars correspond to ±SEM, defined from reaction-level data (ex. aggregate 100 simulations for each dataset, then compare between datasets).

Budget	Random	Buried volume		
	Hit rate	Avg. hit rate	SEM	p
50%	0.577	0.699	0.019	<0.0001

…

Budget	Sterimol			Calculated pKa		
	Avg. hit rate	SEM	p	Avg. hit rate	SEM	p
50%	0.637	0.021	0.028	0.526	0.021	0.017

Supplementary Table VI-42 | Avg. hit rate, SEM, and *p*-value for each representation deployed across a proportional budget corresponding to 50% of the dataset size.

◊ *Dimensionality-reduced representations of buried volume retain utility*

The accessible buried volume-based representation from previous reports spans a few conformational ensembles (ground state, Boltzmann avg., min., and max.) along with various radii from active site-relevant atoms (ex. 2-7 Å). As many of these representations are collinear with each other, we investigated whether similar model performance could be realized following dimensionality reduction over the domain (64-total dimensions). We conducted *principal component analysis* (as implemented in scikit-learn) over the complete active site-based buried volume domain and extracted (1) only the first component (PC1 – 82% of total variance), or (2) only the first two components (PC1-and-PC2 – 90% of total variance). Unsurprisingly, their utility in the active learning campaigns is comparable to the complete buried volume domain, highlighting low-dimensional alternatives. Across the same fixed and proportional budgets, the only-PC1 representation observes a subtle decay in relative performance, while PC1-and-PC2 representation is nearly identical with the full representation (**Supplementary Table V-X**).

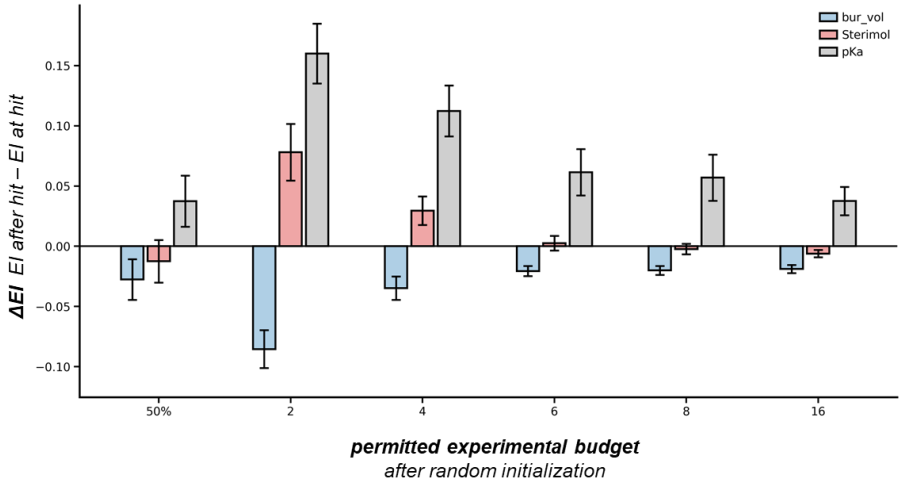
Budget	Buried volume (all)	only PC1 (82%)	only PC1 and PC2 (90%)
	Avg. hit rate	Avg. hit rate	Avg. hit rate
50%	0.699	0.639	0.684
2	0.438	0.419	0.453
4	0.677	0.643	0.677
6	0.839	0.804	0.826
8	0.912	0.867	0.888
16	0.963	0.952	0.958

Supplementary Table VI-43 | Avg. hit rate for dimensionality-reduced active site-based buried volume representations across the each experimental budget. For principal component-reduced representations, the total variance captured in the full dataset is indicated in parentheses (%).

- ***Model optimism upon selecting best available choice (ΔEI)***

Upon selection of the *best* choice, we can investigate model *optimism* by inspecting the acquisition function – expected improvement, EI. Here, if the model did not expect the *best* choice to be promising, then we expect a low EI *at* hit (meaning the EI of the *best* choice made), followed by

similar or even higher EI *after* hit (meaning the EI of the remaining choices *after* updating the model). We quantify this with a simple difference, ΔEI = EI *after* hit – EI *at* hit. When the model does not learn meaningful relationships between representation and activity, then promising observations will distort the *surrogate* model's predictions to be erroneously optimistic (ex. ΔEI > 0, suggesting it expects new points to be more promising now). If the model realizes meaningful relationships, then it should expect the *best* choice made to be the *best* choice available (ex. ΔEI < 0). We report these discussion in the main text and summarize the ΔEI metrics for each representation under each budget (fixed and proportional) below (**Supplementary Fig. V-X**, **Supplementary Table V-X**).



Supplementary Fig. VI-191 | ΔEI across all simulated campaigns v. experimental budget (fixed and proportional, as discussed). Error bars correspond to ±SEM, defined from reaction-level data (ex. aggregate 100 simulations for each dataset, then compare between datasets).

Budget	Buried volume		Sterimol		Calculated pKa	
	ΔEI	SEM	ΔEI	SEM	ΔEI	SEM
50%	-0.028	0.017	-0.013	0.018	0.037	0.021
2	-0.085	0.016	0.078	0.024	0.16	0.023
4	-0.034	0.011	0.03	0.011	0.113	0.022
6	-0.021	0.004	0.003	0.005	0.062	0.02
8	-0.02	0.003	-0.001	0.003	0.057	0.019
16	-0.019	0.003	-0.005	0.003	0.039	0.011

Supplementary Table VI-44 | Avg. **Δ**EI, SEM, and *p*-value for each representation deployed across the indicated experimental budgets.

VI.43 – Extended Discussion on Fig. V-3

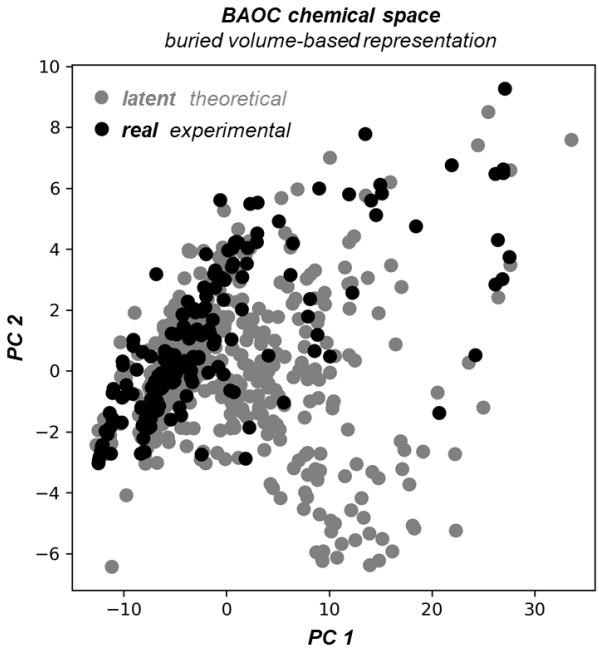
We describe in detail the generation and investigation of the latent space theoretically occupied by BAOCs as defined active site-based buried volume. Moreover, we discuss and report tabulated metrics for the rediscovery campaigns (as seen in **Fig. V-3** of the main text).

- ***Latent opportunity over theoretical BAOC chemical space***

Latent space refers to a high-dimensional representation learned by a model. Often, various machine learning applications seek to *generate* latent space embeddings from some natural input (ex. SMILES string). Here, we define the latent space of BAOCs by active site-based buried volume representation. Because we cannot compute the *true* universe of BAOCs accessible, we approximate this universe with the latent space (500 discrete points).

◊ *Latent space resembles real, enumerated population of BAOCs*

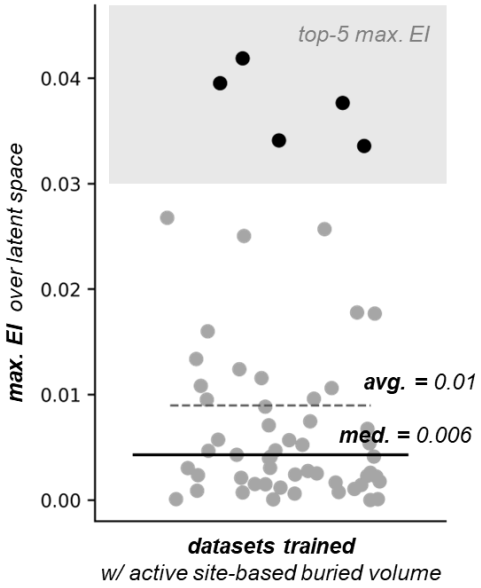
The latent space generated is defined by the first two principal components. We randomly populate with respect to the existing distribution of catalysts from datasets already calculated in this work (ex. literature-derived datasets). As such, these latent *points* strongly resemble the real domain of catalysts available while providing a more rich, lightweight reflection of *possible* BAOC opportunity (**Supplementary Fig. V-X**).



Supplementary Fig. VI-192 | Latent BAOCs (dark gray) resembles real BAOCs (black) and their distribution across BAOC chemical space, defined by active site-based buried volume.

◊ *Expected improvement from latent space is rare*

With the perspective of *possible* BAOCs furnished from the 500-point populated latent space, we can evaluate whether for each dataset the acquisition function expect outstanding opportunities beyond the scope of catalysts available (defined as large *latent* EI). As described in the main text, the average EI from across the latent space is nearly zero, on average (avg. max. *latent* EI = 0.01, median max. *latent* EI = 0.006). However, a few datasets do show larger max. *latent* EI (up to max. *latent* EI = 0.05, **Supplementary Fig. V-X**).

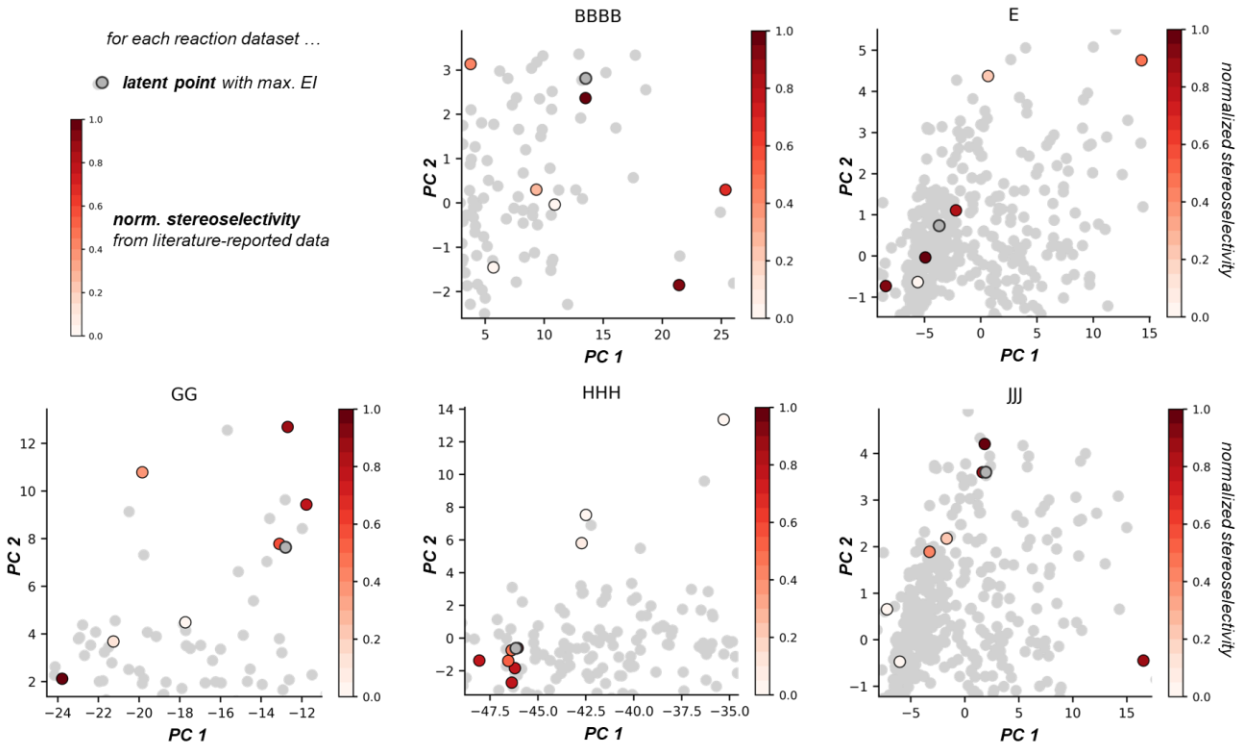


Supplementary Fig. VI-193 | Max. *latent* EI is not uniformly distributed per dataset. A few datasets show larger max. *latent* EI suggesting outstanding opportunities for optimization (avg. = mean, med. = median).

◊ *Latent opportunities are distributed in unpopulated domains near known solutions*

In **Supplementary Fig. V-X**, the top-5 datasets with max. *latent* EI are highlighted. These correspond to reaction datasets: BBBB, E, GG, HHH, JJJ (*SHEET NAME* from compiled datasets as seen in **Appendix III** and **Appendix IV**[2]). We can visualize the opportunity expected from latent space with respect to the reported catalysts (see all datasets, **Supplementary Fig. V-X**).

By inspection, we see in some cases, the expected opportunity is near the most stereoselective solution, but far from other reported catalysts. This dichotomy suggests that intentionally populating these sparse regions of chemical space could lead to the discovery of additional stereoselective BAOCs. However, we remind readers that these latent points are purely theoretical. They are simply defined by active site-based buried volume and do not correspond to tangible structures. Thus, property-driven design promises a realization of their outlined opportunity.



Supplementary Fig. VI-194 | Top-5 max. *latent* EI datasets show the identified opportunity (dark gray, outlined) is proximal to the reported stereoselective solutions (red color scale), but may lie in sparsely populated regions – far from other known examples. Reaction datasets are identified by their *SHEET NAME* from compiled datasets,[2] as seen in **Appendix III** and **Appendix IV**.

- ***In silico rediscovery campaign***

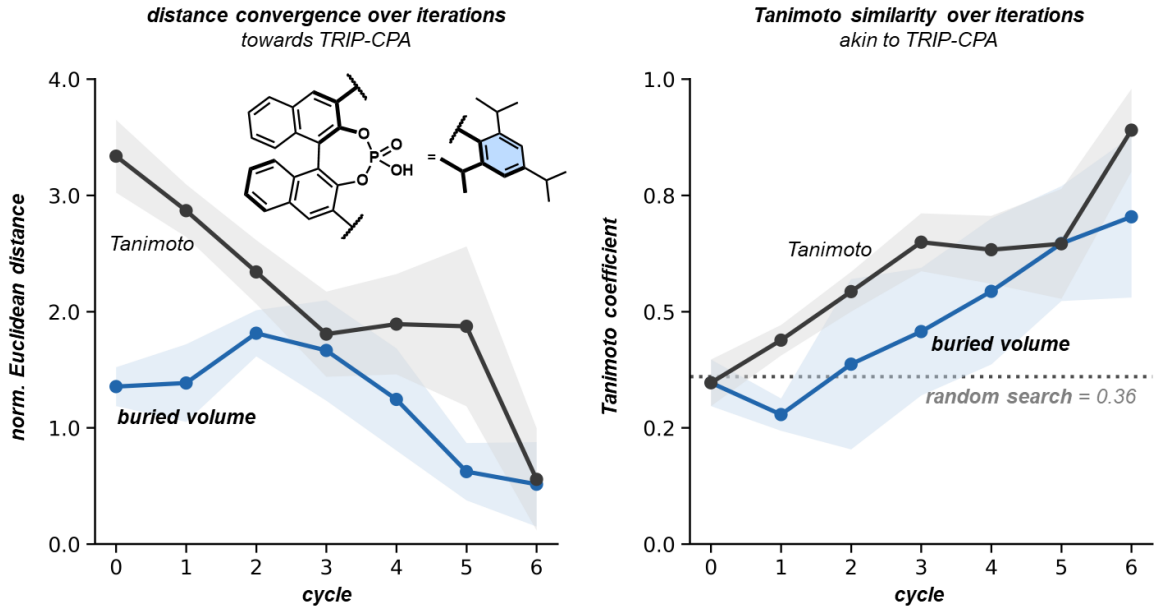
To compare the quality of *true* generative models, researchers often simulate *rediscovery* tasks. Here, a general target is declared, and the rate of convergence is compared across candidate models. In this work, we do not investigate alternative models (only use Doyle's EDBO+[2]). Instead, we leverage *rediscovery* (1) to qualify whether active site-based buried volume can guide the structural rediscovery, and (2) to quantify how fast the representation guides convergence. As described in the general methods, each rediscovery campaign uses some random initialization (batch = 8) and is repeated in quadruplicate (4 different starts). We use the same random starts between representations (ex. active site-based buried volume, Tanimoto coefficient) to enforce parity of initial information available (with respect to catalyst name as seen in *dataset_catalyst.csv,*[2] **Supplementary Table V-X**).

CS #1	CS #2	CS #3	CS #4
CPAH_1579	CPAH_2194	CPAH_1967	CPAH_1437
CPAH_2711	CPAH_2737	CPAH_1470	CPAH_999
CPAH_2019	CPAH_12	CPAH_2774	CPAH_855
CPAH_940	CPAH_1261	CPAH_1676	CPAH_1628
CPAH_2030	CPAH_1790	CPAH_1338	CPAH_2890
CPAH_2828	CPAH_1410	CPAH_962	CPAH_1218
CPAH_2499	CPAH_70	CPAH_215	CPAH_2656
CPAH_2201	CPAH_811	CPAH_2396	CPAH_818

Supplementary Table VI-45 | Specific CPAs selected for each of the four *cold starts* (random initializations) used across each *rediscovery* campaign.

◊ *TRIP-CPA rediscovery*

As seen in the main text, active site-based buried volume (specifically Boltzmann avg. ensemble) provides similar convergence to *true* structural target – TRIP-CPA when compared to campaigns using Tanimoto coefficient as representation.

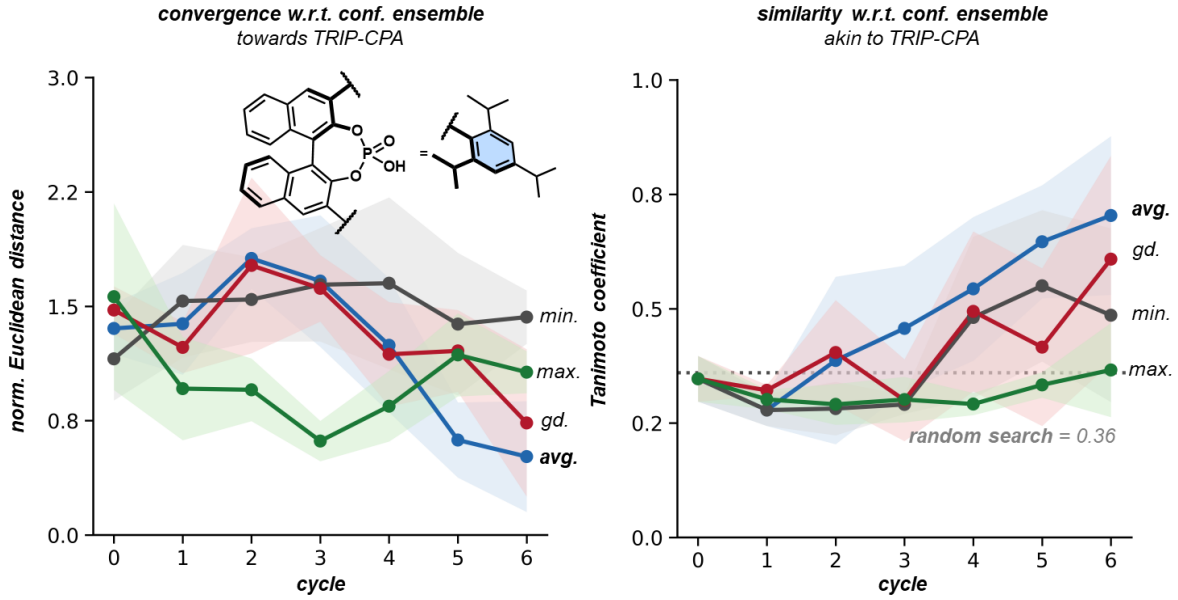


Supplementary Fig. VI-195 | Convergence to TRIP-CPA visualized by the model objective, norm. Euclidean distance (left) and Tanimoto similarity (right). The representations used are Boltzmann averaged active site-based buried volume (blue) and Tanimoto coefficient (gray). Shaded bounds correspond to ±SEM from four random initializations.

Cycle	norm. Euclidean distance				Tanimoto similarity			
	w/ Boltz. avg. buried volume		w/ Tanimoto coefficient		w/ Boltz. avg. buried volume		w/ Tanimoto coefficient	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	1.35	0.17	3.34	0.31	0.35	0.05	0.35	0.05
1	1.39	0.33	2.87	0.23	0.28	0.03	0.44	0.03
2	1.81	0.20	2.34	0.27	0.39	0.18	0.54	0.04
3	1.67	0.43	1.81	0.37	0.46	0.14	0.65	0.06
4	1.24	0.44	1.89	0.43	0.54	0.16	0.63	0.07
5	0.62	0.25	1.87	0.69	0.65	0.12	0.65	0.12
6	0.51	0.36	0.56	0.44	0.70	0.17	0.89	0.09

Supplementary Table VI-46 | Tabulated metrics for convergence to TRIP-CPA over four random initializations, seen in **Supplementary Fig. V-6.**

We observed that the choice of ensemble for buried volume representation furnishes distinct results. Overall, Boltzmann avg. ensemble furnishes the fastest *rediscovery* of TRIP-CPA.



Supplementary Fig. VI-196 | Convergence to TRIP-CPA is sensitive with respect to conformational ensemble of active site-based buried volume used. The ensembles are ground state ('gd.', red), Boltzmann average ('avg.', blue), minimum value ('min.', gray), and maximum value ('max.', green). Shaded bounds correspond to ±SEM from four random initializations.

Cycle	norm. Euclidean distance							
	w/ Boltz. avg.		w/ minimum		w/ ground state		w/ maximum	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	1.35	0.17	1.16	0.27	1.48	0.16	1.56	0.61
1	1.39	0.33	1.53	0.37	1.23	0.17	0.96	0.34
2	1.81	0.20	1.54	0.28	1.77	0.58	0.95	0.21
3	1.67	0.43	1.64	0.37	1.62	0.22	0.62	0.13
4	1.24	0.44	1.65	0.56	1.19	0.34	0.84	0.23
5	0.62	0.25	1.38	0.47	1.21	0.27	1.18	0.27
6	0.51	0.36	1.43	0.17	0.74	0.48	1.07	0.14

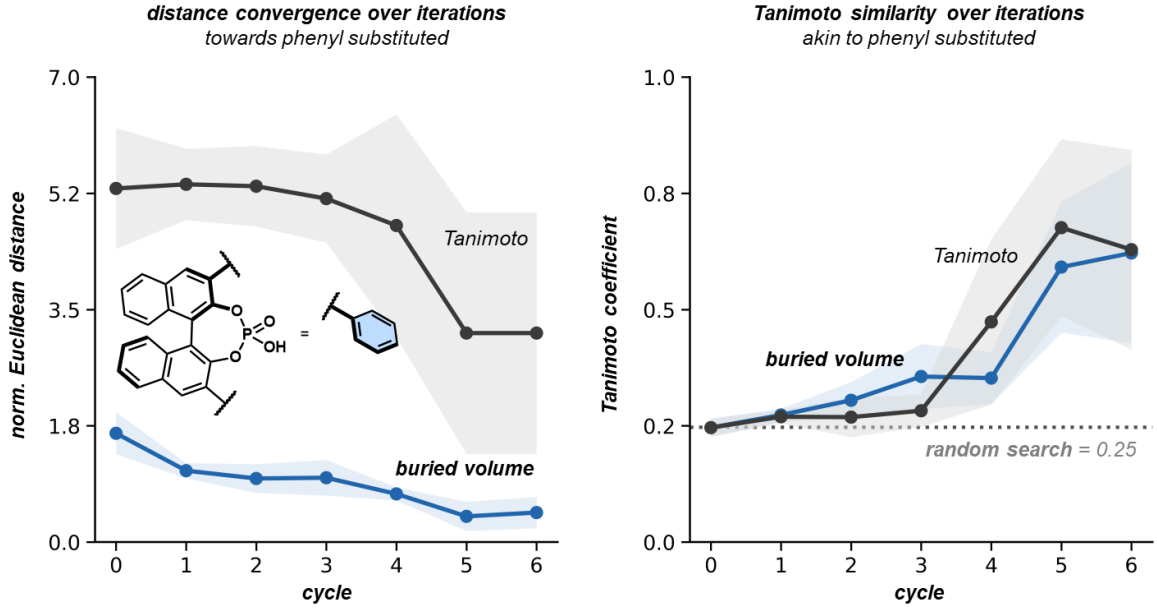
…

Cycle	Tanimoto similarity							
	w/ Boltz. avg.		w/ minimum		w/ ground state		w/ maximum	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	0.35	0.05	0.35	0.05	0.35	0.05	0.35	0.05
1	0.28	0.03	0.28	0.03	0.32	0.01	0.30	0.02
2	0.39	0.18	0.28	0.06	0.40	0.11	0.29	0.04
3	0.46	0.14	0.29	0.02	0.30	0.09	0.30	0.05
4	0.54	0.16	0.48	0.18	0.49	0.17	0.29	0.02
5	0.65	0.12	0.55	0.17	0.42	0.17	0.33	0.03
6	0.70	0.17	0.49	0.19	0.61	0.23	0.37	0.10

Supplementary Table VI-47 | Tabulated metrics for convergence to TRIP-CPA from different conformational ensembles of active site-based buried volume, over four random initializations, seen in **Supplementary Fig. V-7.**

◊ *Rediscovery of other privileged BAOC targets*

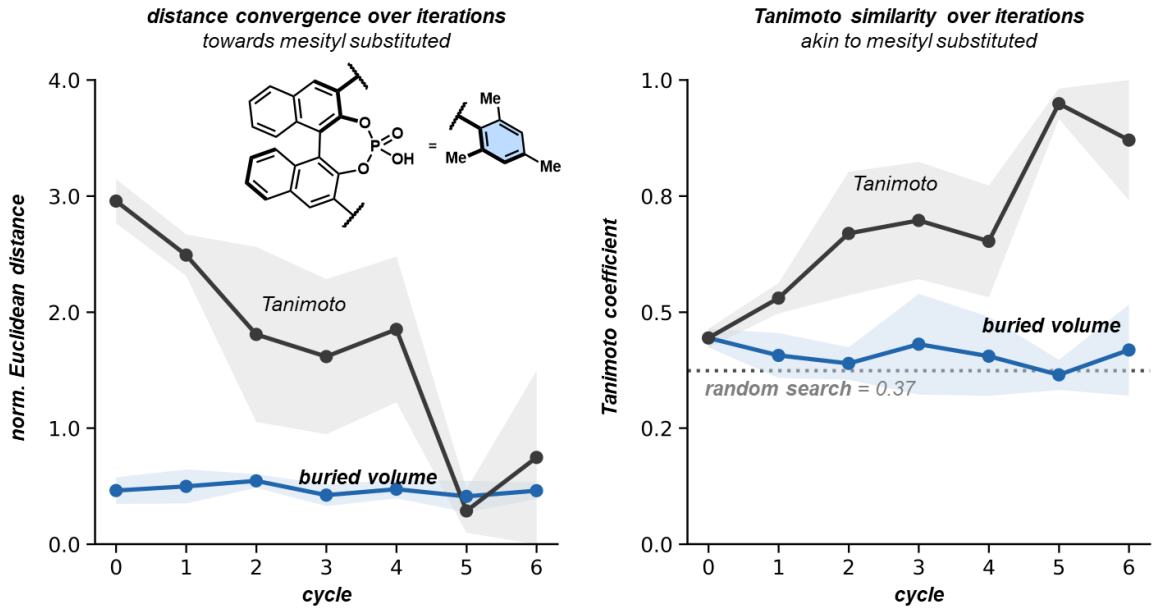
We also evaluated convergence to three other familiar CPA targets as shown below.



Supplementary Fig. VI-197 | Convergence to 3,3'-phenyl substituted CPA visualized by the model objective, norm. Euclidean distance (left) and Tanimoto similarity (right). The representations used are Boltzmann averaged active site-based buried volume (blue) and Tanimoto coefficient (gray). Shaded bounds correspond to ±SEM from four random initializations.

Cycle	norm. Euclidean distance				Tanimoto similarity			
	w/ Boltz. avg. buried volume		w/ Tanimoto coefficient		w/ Boltz. avg. buried volume		w/ Tanimoto coefficient	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	1.64	0.31	5.32	0.91	0.25	0.02	0.25	0.02
1	1.07	0.11	5.39	0.54	0.27	0.01	0.27	0.01
2	0.96	0.22	5.36	0.61	0.30	0.04	0.27	0.04
3	0.97	0.27	5.17	0.66	0.36	0.07	0.28	0.03
4	0.73	0.10	4.77	1.67	0.35	0.06	0.47	0.18
5	0.39	0.22	3.15	1.82	0.59	0.14	0.68	0.19
6	0.44	0.23	3.15	1.82	0.62	0.19	0.63	0.21

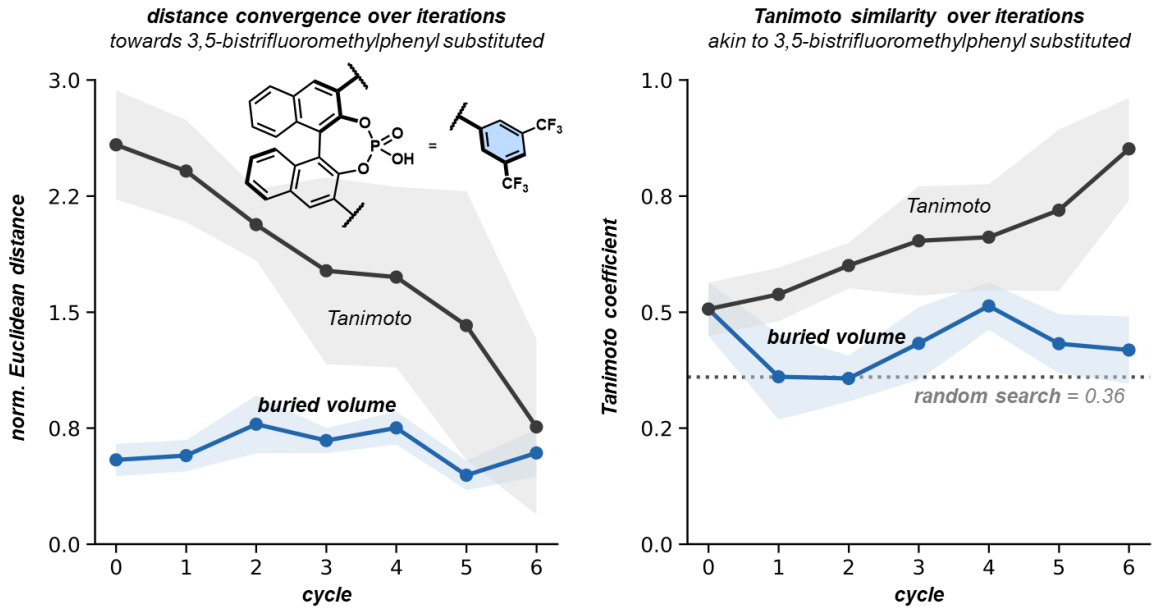
Supplementary Table VI-48 | Tabulated metrics for convergence to 3,3'-phenyl substituted CPA over four random initializations, seen in **Supplementary Fig. V-8.**



Supplementary Fig. VI-198 | Convergence to 3,3'-mesityl substituted CPA visualized by the model objective, norm. Euclidean distance (left) and Tanimoto similarity (right). The representations used are Boltzmann averaged active site-based buried volume (blue) and Tanimoto coefficient (gray). Shaded bounds correspond to ±SEM from four random initializations.

Cycle	norm. Euclidean distance				Tanimoto similarity			
	w/ Boltz. avg. buried volume		w/ Tanimoto coefficient		w/ Boltz. avg. buried volume		w/ Tanimoto coefficient	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	0.46	0.11	2.96	0.19	0.44	0.02	0.44	0.02
1	0.50	0.15	2.49	0.18	0.41	0.05	0.53	0.03
2	0.55	0.06	1.81	0.75	0.39	0.03	0.67	0.13
3	0.42	0.09	1.62	0.67	0.43	0.11	0.70	0.13
4	0.47	0.08	1.85	0.63	0.40	0.09	0.65	0.12
5	0.41	0.14	0.29	0.19	0.36	0.03	0.95	0.03
6	0.46	0.07	0.75	0.75	0.42	0.10	0.87	0.13

Supplementary Table VI-49 | Tabulated metrics for convergence to 3,3'-mesityl substituted CPA over four random initializations, seen in **Supplementary Fig. V-9.**



Supplementary Fig. VI-199 | Convergence to 3,3'-(3,5-bistrifluoromethyl)phenyl substituted CPA visualized by the model objective, norm. Euclidean distance (left) and Tanimoto similarity (right). The representations used are Boltzmann averaged active site-based buried volume (blue) and Tanimoto coefficient (gray). Shaded bounds correspond to ±SEM.

Cycle	norm. Euclidean distance				Tanimoto similarity			
	w/ Boltz. avg. buried volume		w/ Tanimoto coefficient		w/ Boltz. avg. buried volume		w/ Tanimoto coefficient	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
0	0.54	0.10	2.58	0.35	0.51	0.06	0.51	0.06
1	0.57	0.10	2.41	0.33	0.36	0.09	0.54	0.06
2	0.78	0.19	2.06	0.23	0.36	0.05	0.60	0.05
3	0.67	0.08	1.77	0.60	0.43	0.08	0.65	0.12
4	0.75	0.11	1.73	0.58	0.51	0.05	0.66	0.11
5	0.45	0.09	1.41	0.87	0.43	0.06	0.72	0.17
6	0.59	0.15	0.76	0.56	0.42	0.07	0.85	0.11

Supplementary Table VI-50 | Tabulated metrics for convergence to 3,3'-(3,5-bistrifluoromethyl)-phenyl substituted CPA over four random initializations, seen in **Supplementary Fig. V-10.**

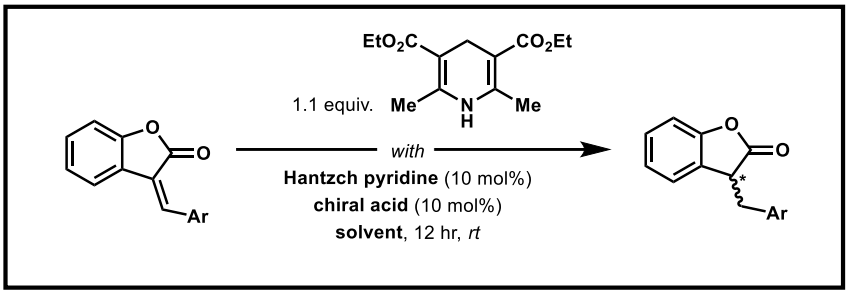
VI.44 – Extended Discussion on Fig. V-4

We describe reaction protocol, tabulate outcomes, discuss mechanistic inquiry experiments, and report on computational methods below, as seen in the main text **Fig. V-4.**

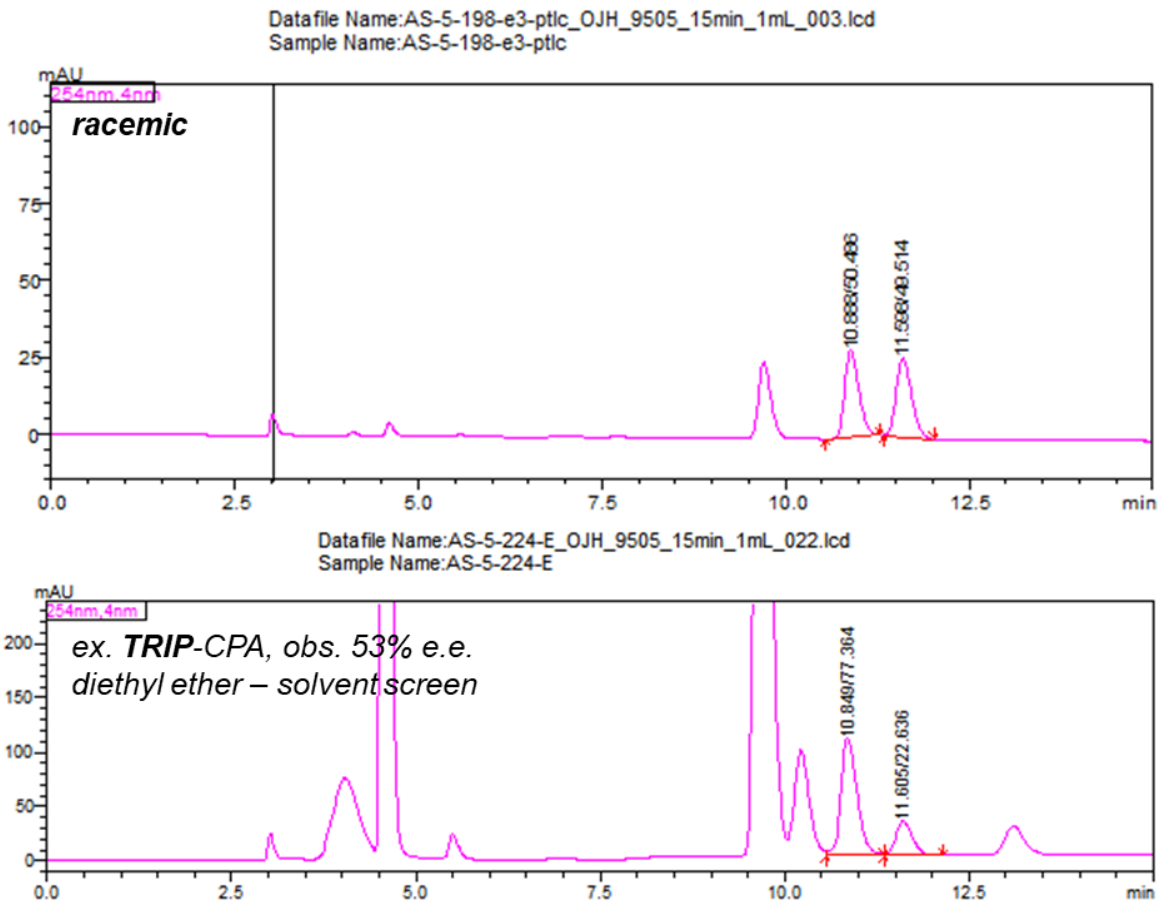
- ***Stereoselective reduction of benzylidenes***

The transformation describes the conjugate reduction of benzylidenes.

◊ *General reaction protocol*

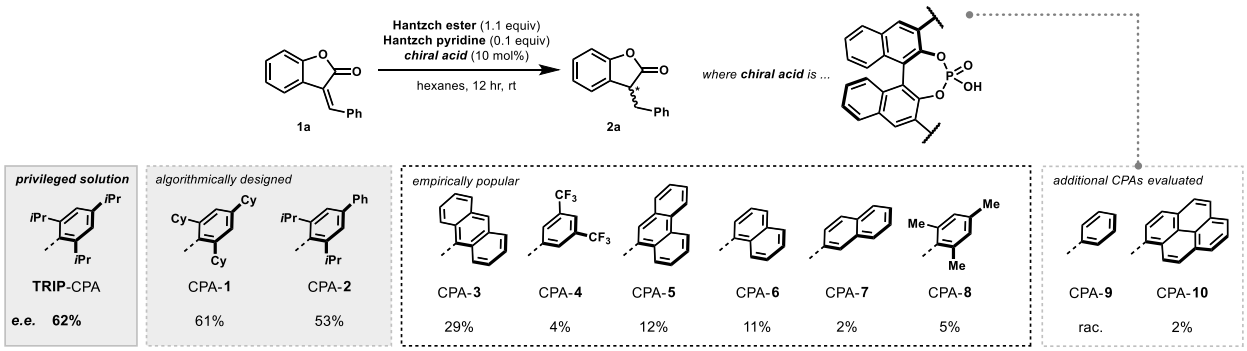


To a 1-dram vial was charged **benzylidene** (25 umol, 1.0 equiv.), **chiral acid** (10 mol%, 2.5 umol, 0.1 equiv.), and **Hantzch pyridine** (10 mol%, 2.5 umol, 0.1 equiv.) as stock solutions in DCM. The DCM was removed *in vacuo*. Then, the vial was charged **Hantzch ester** (7.0 mg, 28 umol, 1.1 equiv.) and suspended in **solvent** (250 uL, 0.1 M). The mixture was sonicated for 30 seconds to promote solvation. Then, the vial was left to react at room temperature for 16 hours. After, the **solvent** was removed *in vacuo*. The solid residue was suspended in 500 uL $CDCl_3$ and transferred to an NMR tube, charged with 50 uL of 50 mM $CDCl_3$ solution of 1,4-dinitrobenzene as internal standard (obs. 10 umol ^1H peak at δ: 8.42 (s)). **Product** *yield* was recorded from ^1H NMR analysis. The leftover residue in the reaction vial was charged with 250 uL 95:05 HPLC-grade hexanes/isopropanol. The solution was passed through a PTFE syringe filter into an HPLC vial. The **product** *enantioenrichment* was recorded from HPLC analysis (**2a** – ChiralPak OJH 95:05 hexanes/isopropanol, t_R = 10.9, 11.6 mins, 1 mL/min; see **Supplementary Fig. V-X**).



Supplementary Fig. VI-200 | Representative *rac.* and enantioenriched HPLC traces from the conjugate reduction of **1a** furnishing **2a**.

◊ *Stereoselectivity data for **2a** from tested BAOCs*



Supplementary Table VI-51 | Full stereoselectivity data for conversion of **1a** to **2a**.

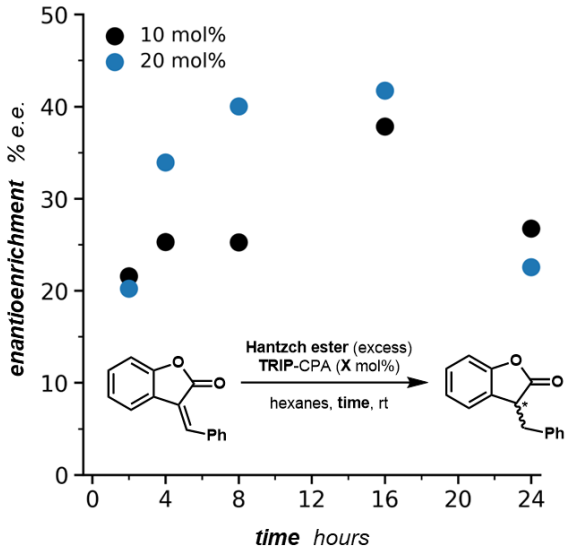
- ***Experimental mechanistic inquiry***

The reported conjugate reduction necessarily involves a protonation step, setting the observed stereocenter. As the substrate is already electrophilic, the reaction requires a catalyst that furnishes

rate acceleration to outcompete with background conjugate reduction. Below, we describe various experiments that aim to characterize this active catalyst and non-covalent interactions responsible for the observed stereoselectivity. In summary, time-course analyses, additive screens, solvent screens, and a substrate-based Hammett series support the expectation that strong electrostatic interactions are involved in productive enantioenrichment. Still, we observe active site-based buried volume remains a suitable surrogate for stereoselectivity, driving the property-based design of stereoselective alternatives relative to more familiar, empirically popular BAOCs.

◊ *Time-course (conversion of **1a**, stereoselectivity of **2a** v. time)*

At the outset of reaction development, we observed full conversion overnight, but somewhat irregular enantioenrichment. Subsequent time-course analysis conducted in the presence of excess reductant (sat. in hexanes) revealed time-dependent enantioenrichment. At varying loadings of **TRIP**-CPA, enantioenrichment of **2a** progressively increases (up to 16 hr). After, conversion to **2a** was complete and enantioerosion of **2a** was observed (after 16 hr, **Supplementary Fig. V-X**).

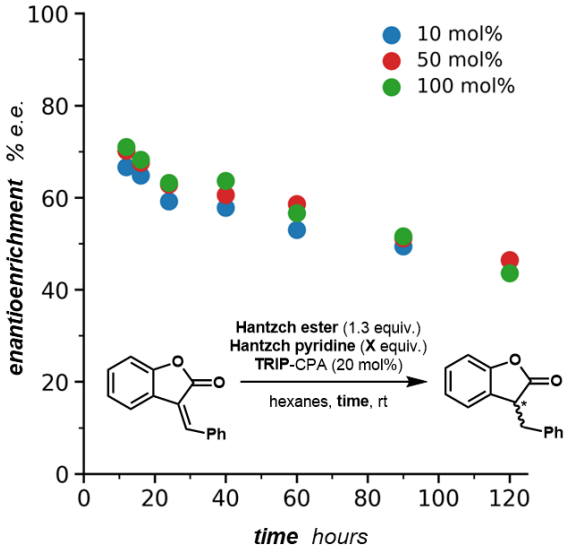


Supplementary Fig. VI-201 | Preliminary time-course analysis of conversion of **1a** to **2a** with excess reductant, where **TRIP**-CPA loading: 10 mol% (black) and 20 mol% (blue).

By inspection, we intuited that the increasing enantioenrichment reflected competing catalyzed and uncatalyzed reactions. After all, the reduction proceeds without any added acid, by virtue of the intrinsic electrophilicity of **1a**. The subsequent enantioerosion could be rationalized by the increased acidity of the generated benzofuranone relative to other esters. Deprotonation at this stereocenter can furnish an aromatic-stabilized enolate. So, we initially hypothesized that enantioerosion is likely a consequence of base-mediated epimerization by Hantzch pyridine, a byproduct from the transfer hydrogenation by Hantzch ester.

To evaluate this, time-course analysis was repeated with controlled loadings of Hatnzch ester and Hantzch pyridine (**Supplementary Fig. V-12**). Surprisingly, the addition of just 10 mol% of

Hantzch pyridine at the outset of the reaction altered the observed time-dependent enantioenrichment and appeared to attenuate any observed enantioerosion (after 16 hr). Accordingly, we concluded Hantzch pyridine additive *may* instead mitigate variability in enantioenrichment between trials by inhibiting deleterious, uncatalyzed processes.

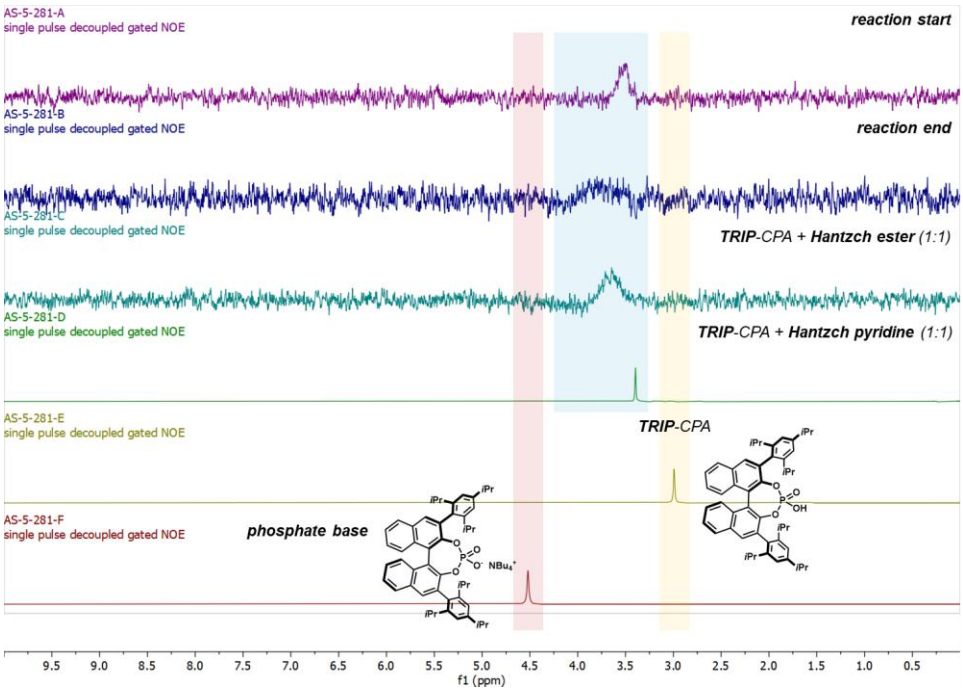


Supplementary Fig. VI-202 | Time-course analysis of conversion of **1a** to **2a** with 1.3 equiv. of Hantzch ester, 20 mol% **TRIP**-CPA, and variable loading of Hantzch pyridine: 10 mol% (blue), 50 mol% (red), and 100 mol% (green).

◊ *Catalyst speciation by ^{31}P NMR*

Additives appeared to influence enantioenrichment of **2a** in the above time-course analyses with **TRIP**-CPA. To evaluate the primary configuration of the acid catalyst during the reaction condition, $^{31}P\{^{1}H\}$ NMR were collected at the beginning and end of the reaction. These signals were then juxtaposed to the chiral acid with various additives, including: Hantzch ester (reagent), Hantzch pyridine (basic byproduct), none (acid reference), and tetrabutylammonium hydroxide (conjugate base reference, **Supplementary Fig. V-X**).

The analysis indicates that during the reaction, the chiral catalyst likely exists in a dynamic complex with either Hanztch ester and pyridine (both broad before/after reaction course).

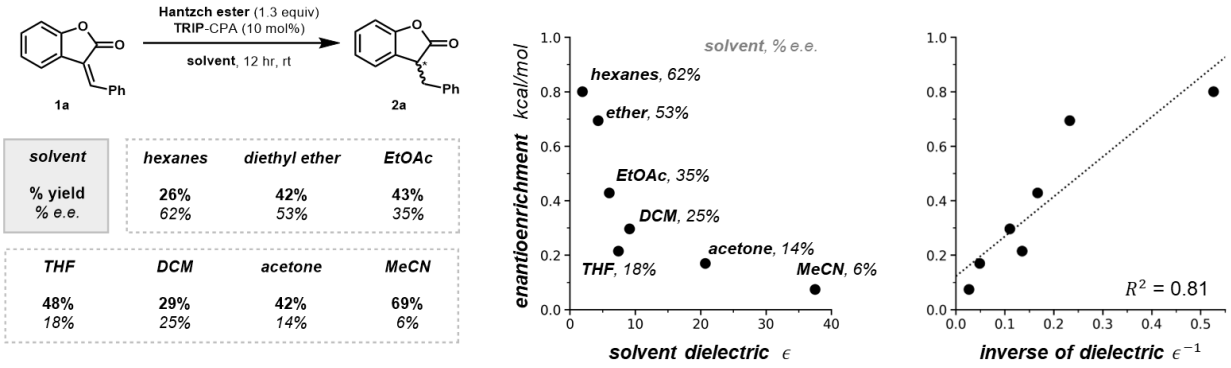


Supplementary Fig. VI-203 | $^{31}P\{^{1}H\}$ NMR analysis of **TRIP**-CPA speciation at the beginning and end of the conjugate reduction, compared to various additives and standards.

◊ *Solvent screen with TRIP-CPA*

Interrogating the activity of the complex between **TRIP**-CPA and Hantzch ester during the reaction, a solvent screen was conducted to evaluate the impact of polarity on enantioinduction (**Supplementary Fig. V-X**). Across all solvents, strong conversion of **1a** (>80%) and moderate yields of **2a** observed (approx. 30-40% yield). But, enantioenrichment was strongly correlated with the inverse of solvent dielectric ($R^2 = 0.81$). Similar solvent effects are often observed in reactions proceeding through contact ion-pairs.

According to Coulomb's law, electrostatic interactions between two interacting charges scale with the inverse of permittivity (approximated here with solvent dielectric). Based on an electric field interpretation, the linear relationship between enantioenrichment and solvent dielectric suggests solvent choice can exacerbate stereoselective, electronic interactions between competing diastereomeric transition states.[5,6] Qualitatively, solvent dielectric can be used as a surrogate for ionic stabilization. Nonpolar solvents like hexanes poorly support solvation of charged species. As a result, ionic interactions are expected to occur through closely associated charges where more pronounced stereoselective interactions may be present.

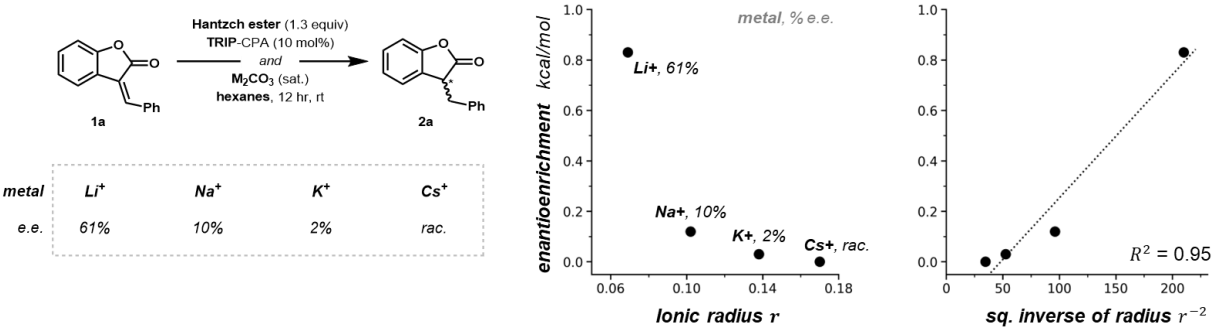


Supplementary Fig. VI-204 | Solvent screen results in the conversion of **1a** to **2a** with 1.3 equiv. of Hantzch ester and 10 mol% **TRIP**-CPA (left). Strong relationship observed ($R^2 = 0.81$) to the inverse of dielectric constant, reflecting Coulomb's law for the magnitude of electrostatic interactions (right).

◊ *Counter-cation screen with TRIP-CPA*

Based on Coulomb's law, electrostatic interactions are not only dependent on permittivity, but also the distance between interacting charges. By adding excess alkali metal carbonate to the suspension, we hypothesized we could detect a relationship to distance (*r*) as approximated by the ionic radius of the alkali metal cation. We conducted the reaction screen in hexanes with excess metal carbonate (**Supplementary Fig. V-X**). We note that the reaction *does* proceed with tetrabutylammonium phosphate salt of **TRIP**-CPA with poor enantioinduction (obs. 22% *e.e.*).

The screening data is consistent with the expected relationship to distance, as approximated by ionic radius. The narrower the ionic radius (ex. Li^+) the greater the enantioenrichment (obs. up to 61% *e.e.*). Cs^+ only furnished *rac* product. As we have not yet observed base-mediated epimerization, we do not expect poor enantioenrichment to be a consequence of erosion.

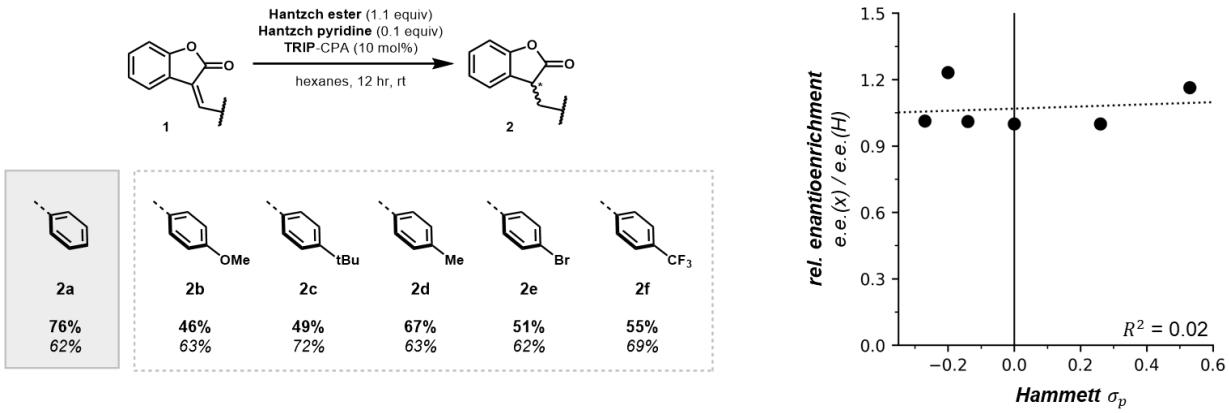


Supplementary Fig. VI-205 | Counter-cation screen results in the conversion of **1a** to **2a** with 1.3 equiv. of Hantzch ester, 10 mol% **TRIP**-CPA, and excess alkali metal carbonate in hexanes (left). Strong relationship observed ($R^2 = 0.95$) to the squared inverse of ionic radius, reflecting Coulomb's law for the magnitude of electrostatic interactions (right).

◊ *Hammett series of substituted benzylidenes*

The mechanistic insight suggests stereoselectivity is strongly dependent on electrostatic interactions between the substrate, catalyst, and reductant during the reaction. Traditionally, substrate electronic effects can be interrogated with a Hammett series. To this end, we prepared a series of *p*-substituted aryl benzylidene derivates (**1b** through **1f**). Subjection of **1a-f** to the conjugate reduction furnished the corresponding benzofuranones **2a-f** in moderate to good yields (approx. 50-70% yield) but with relatively constant enantioenrichment (avg. 65% *e.e.*, **Supplementary Fig. V-X**).

The analysis indicates that there is no significant relationship between substrate electronics (as characterized by Hammett's $\boldsymbol{\sigma_p}$) and enantioenrichment. This suggests that stereocontrol primarily stems from electrostatic control across the active acid catalyst/reductant complex in solution, with no significant charge differences from the substrate during the enantiodetermining step.

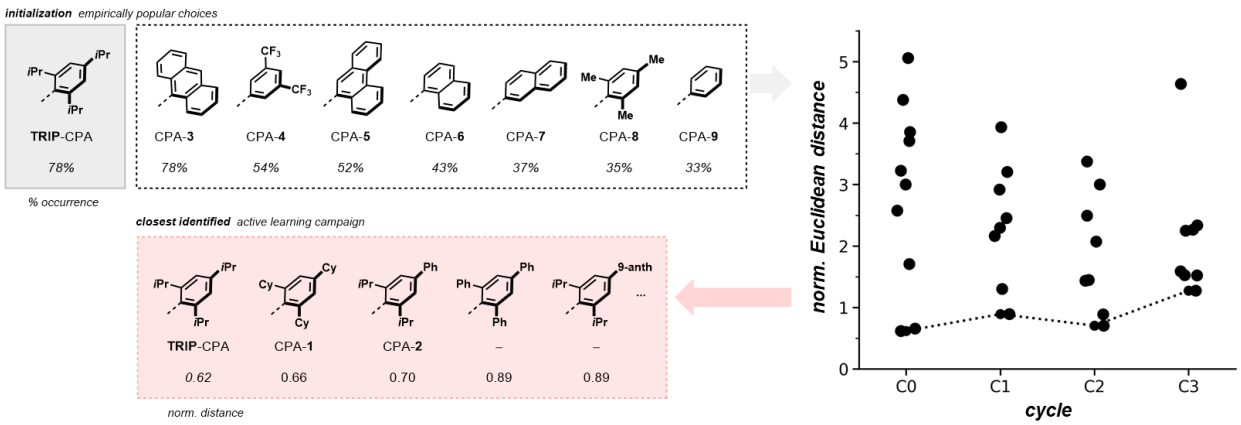


Supplementary Fig. VI-206 | Hammett analysis of substituted benzylidenes shows no apparent relationship to obs. enantioenrichment ($R^2 = 0.02$, $\boldsymbol{\sigma_p}$[7]).

▪ ***Algorithmic design of CPAs resembling TRIP-CPA***

The active learning campaign conducted sought to generalize the chemical space nearby the high-performing **TRIP**-CPA. We targeted a theoretical point just right next to **TRIP**-CPA so the model could search for similar candidates, without successfully collapsing into a *rediscovery* of **TRIP**-CPA. The campaign was initialized with the 8-most popular substituents (based on 50+ literature-derived datasets, see **Appendix IV** for further details), representing the chemical space likely accessed by synthetic chemists without additional prior knowledge or insight for catalyst design.

Immediately, the campaign displayed very shallow learning because of the proximity of **TRIP**-CPA in the initial batch (**Supplementary Fig. V-X**). Search was terminated after 24 calculations as the minimum norm. Euclidean distance in each batch continued to decay, indicating the model was no longer identifying additional, nearby points to the target. The top-five closest points were inspected and the first two were selected for experimental validation (CPA-**1** and CPA-**2**).



Supplementary Fig. VI-207 | Active learning campaign conducted to generalize stereoselective opportunity from **TRIP**-CPA in the conjugate reduction of **1a** to **2a**.

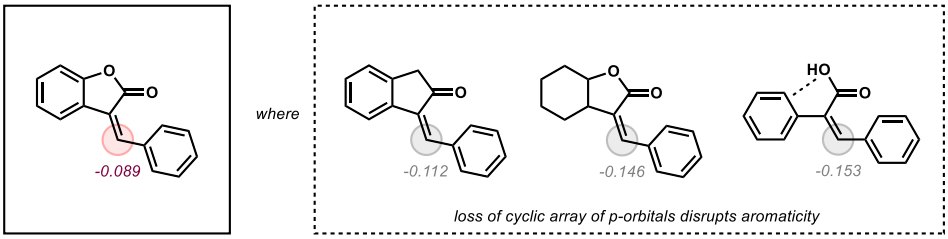
- ***Electrophilicity calculations based on partial charge***

The conjugate reduction of benzylidenes is spontaneous and occurs without the need for an activating acid catalyst. This is because the substrate is abnormally electrophilic compared to a traditional enone. We assign this behavior to significant resonance contributor that is aromatic. We deploy DFT-level calculations below to characterize this electrophilicity and provide further evidence of this expectation.

All calculations are optimized with Gaussian16 at the B3LYP/6-31+G level of theory (smd=toluene as implicit solvent).[8] Frequency calculations validated no negative vibrational modes were present in the optimized structures. The partial charges are reported as NBO (pop=nbo).

◊ *Electrophilicity as a function of aromaticity*

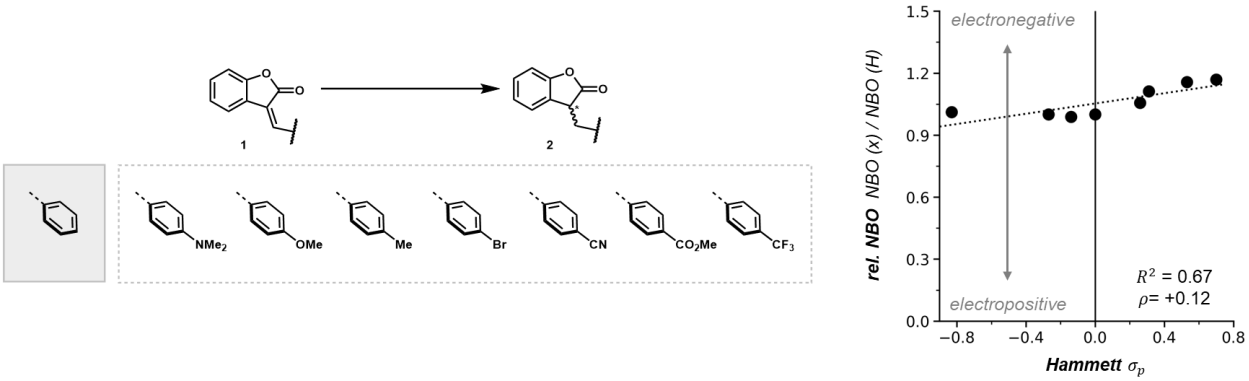
The electrophilicity of the benzylidenes deployed here are assigned to an aromatic resonance contributor. We show that disruption of this aromaticity leads to a loss in electrophilicity, even though the substrate remains an enone (ex. ester to ketone, loss of conjugation, loss of cycle, **Supplementary Fig. V-X**).



Supplementary Fig. VI-208 | Aromatic resonance contributor is disrupted when the cyclic array of *p*-orbitals are disrupted (ex. no lone pairs, no conjugation, no ring). The observed electrophilicity is substantially attenuated (more negative NBO charge).

◊ *Aromatic resonance contributor overrides substituent effects*

The aromatic resonance contributor renders the β-position of the enone substantially electrophilic. We envisioned that substituent electronic effects could compete with polarization at this site. However, aromaticity is typically viewed as a prominent stabilizing interaction. We show with a theoretical Hammett series that the aromatic resonance contributor substantially flattens substituent electronic effects at the electrophilic site (**Supplementary Fig. V-X**, ρ = +0.12).



Supplementary Fig. VI-209 | Theoretical Hammett analysis shows a weak substituent-based electronic effect on the electrophilicity of the benzylidenes under analysis ($\boldsymbol{\rho}$ = +0.12).

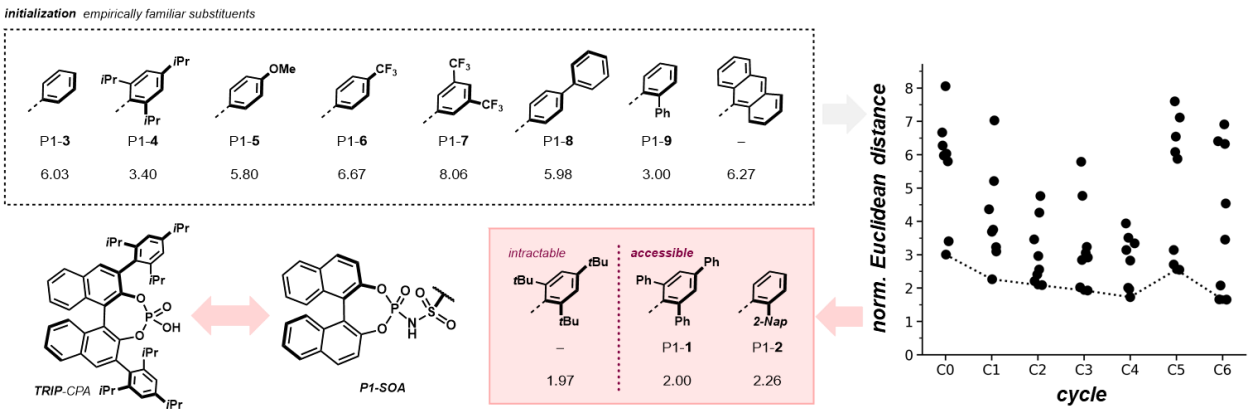
VI.45 – Extended Discussion on Fig. V-5

We describe reaction protocol, tabulate outcomes, and report on computational methods below, as seen in the main text **Fig. V-5.**

- ***Algorithmic design of P1-SOAs resembling TRIP-CPA***

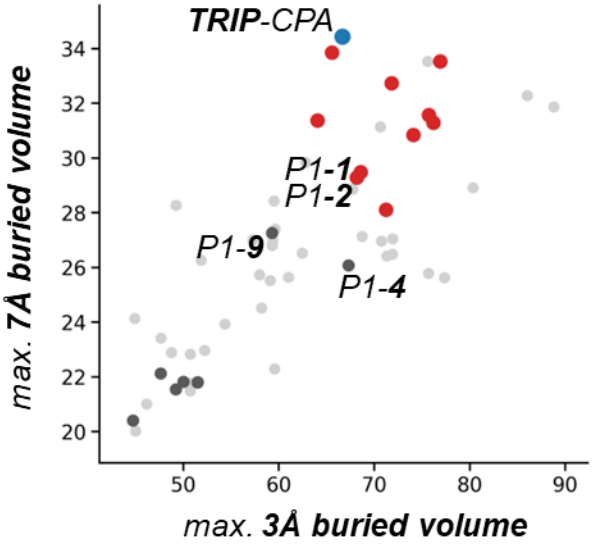
The active learning campaign sought to algorithmically design phosphorylated sulfonamides (P1-SOAs) that resembled **TRIP**-CPA with respect to active site-based buried volume. Because the targeted catalysts display chemically distinct active sites, we envisioned including calc. pK$_a$ would help to bias the search to further reflect **TRIP**-CPA as well. The calc. pK$_a$ was computed from the energy of deprotonation of each P1-SOA and compared to calibrated, experimental benchmarks resembling the active site, as previously reported.[2]

The campaign was initialized with catalysts bearing *S*-substitution familiar to BAOCs had they been traditional, 3,3'-substituents (P1-**3** through P1-**9**, **Supplementary Fig. V-X**). Over 6 cycles, learning appeared to plateau (no substituent identified to improve proximity to target). The top-10 closest P1-SOAs were inspected. The P1-**1** and P1-**2** were selected on the basis of commercial availability of the precursor aryl bromide. Notably, the model suggested 2,4,6-tri-*tert*-butylphenyl substituent but synthetic attempts to prepare this bulky sulfonamide were unsuccessful. As a result, only P1-**1** and P1-**2** were carried forward for experimental validation.



Supplementary Fig. VI-210 | Algorithmic design of P1-SOAs resembling **TRIP**-CPA.

We can visualize the selected points (P1-**1** and P1-**2**) over chemical space to showcase their relative proximity to **TRIP**-CPA (compared to P1-**3** through P1-**9**, **Supplementary Fig. V-X**).



Supplementary Fig. VI-211 | Chemical space representation showing algorithmically designed P1-SOAs and their relative proximity to **TRIP**-CPA compared to hand-picked P1-SOAs.

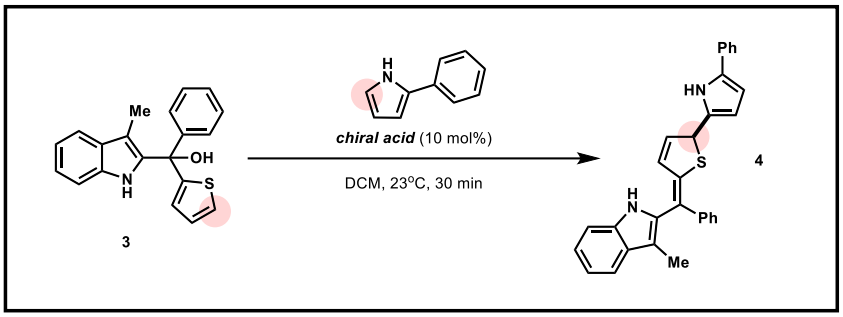
- ***Experimental evaluation of P1-SOAs***

The active learning campaign was initialized from *empirically familiar* choices (P1-**3** through P1-**9**), leading to the *algorithmically designed* choices (P1-**1** through P1-**2**). These catalysts were evaluated in a series of reported BAOC reactions.

It should be noted that the property-based model is based on expectations surrounding the relationship between buried volume and stereoselectivity. As a result, the active learning campaign **does not** consider whether it is likely that a sulfonamide-based acid would be suitable in a reaction. For example, we have evaluated P1-SOAs in the conjugate reduction of **1a** to **2a**. With less

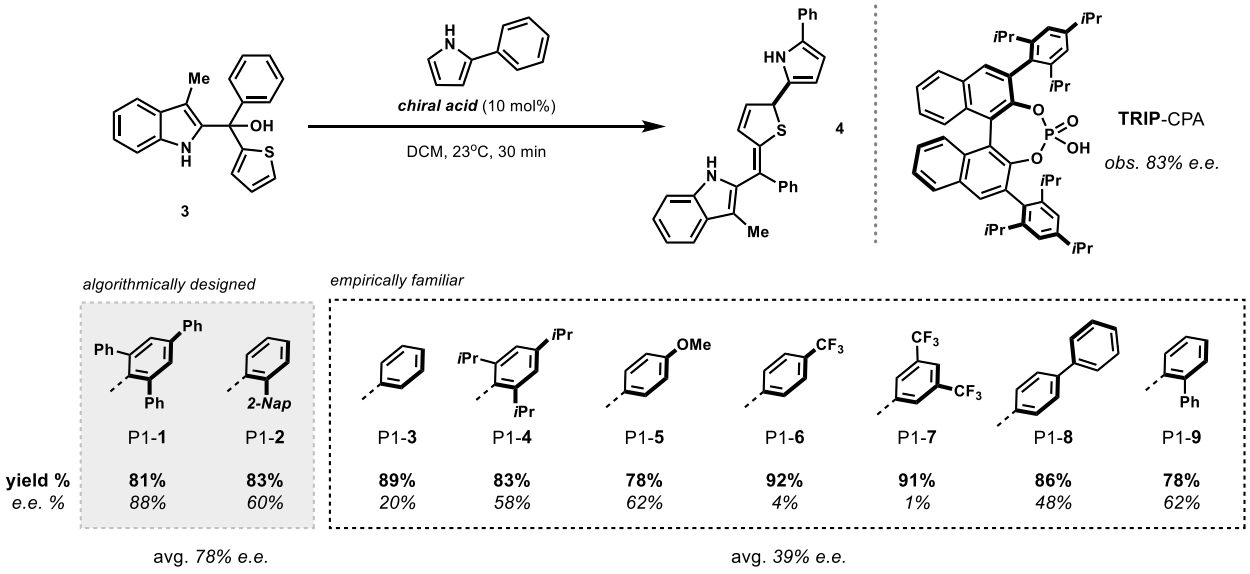
hindered sulfonamides (P1-**4**), the reaction only proceeds with poor stereoinduction and yield (obs. 30% yield, 20% *e.e.*). However, more hindered sulfonamides like P1-**1** display no catalytic activity (<10% yield, n.d. for % *e.e.*).

◊ *Friedel-Crafts Arylation*



The Friedel-Crafts arylation of **3** with 2-phenylpyrrole was originally reported by Li, X. Sun, J. et al. *Nat. Comm.* **2021.**[9] The reaction was run as described in their reported **Supplementary Table 5** (ex. 25 umol scale with 1.1 equiv. pyrrole, 10 mol% chiral acid, 0.1 M DCM, 30 min, r.t.). Yield was measured by ^{1}H NMR analysis (with 1,3,5-triisopropylbenzene as internal standard) and enantioenrichment was measured by HPLC analysis (ChiralPak ODH 70:30 hexanes/isopropanol, t_R = 9.6, 14.2 min, 1mL/min, **Supplementary Fig. V-X**).

The reaction data collected across all BAOCs is recorded below (**Supplementary Fig. V-X**). Quantitative conversion of **3** was observed in all cases.

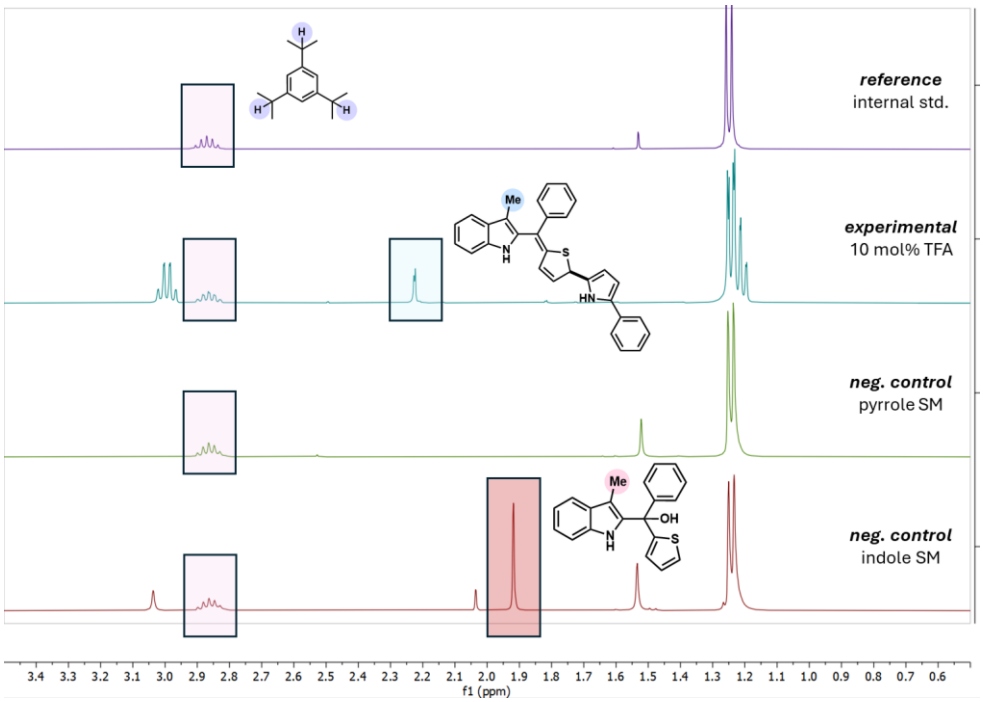


Supplementary Fig. VI-212 | Yield (by ^{1}H NMR) and enantioenrichment (by HPLC) of **4** from the Friedel-Crafts arylation across BAOCs screened in this work.

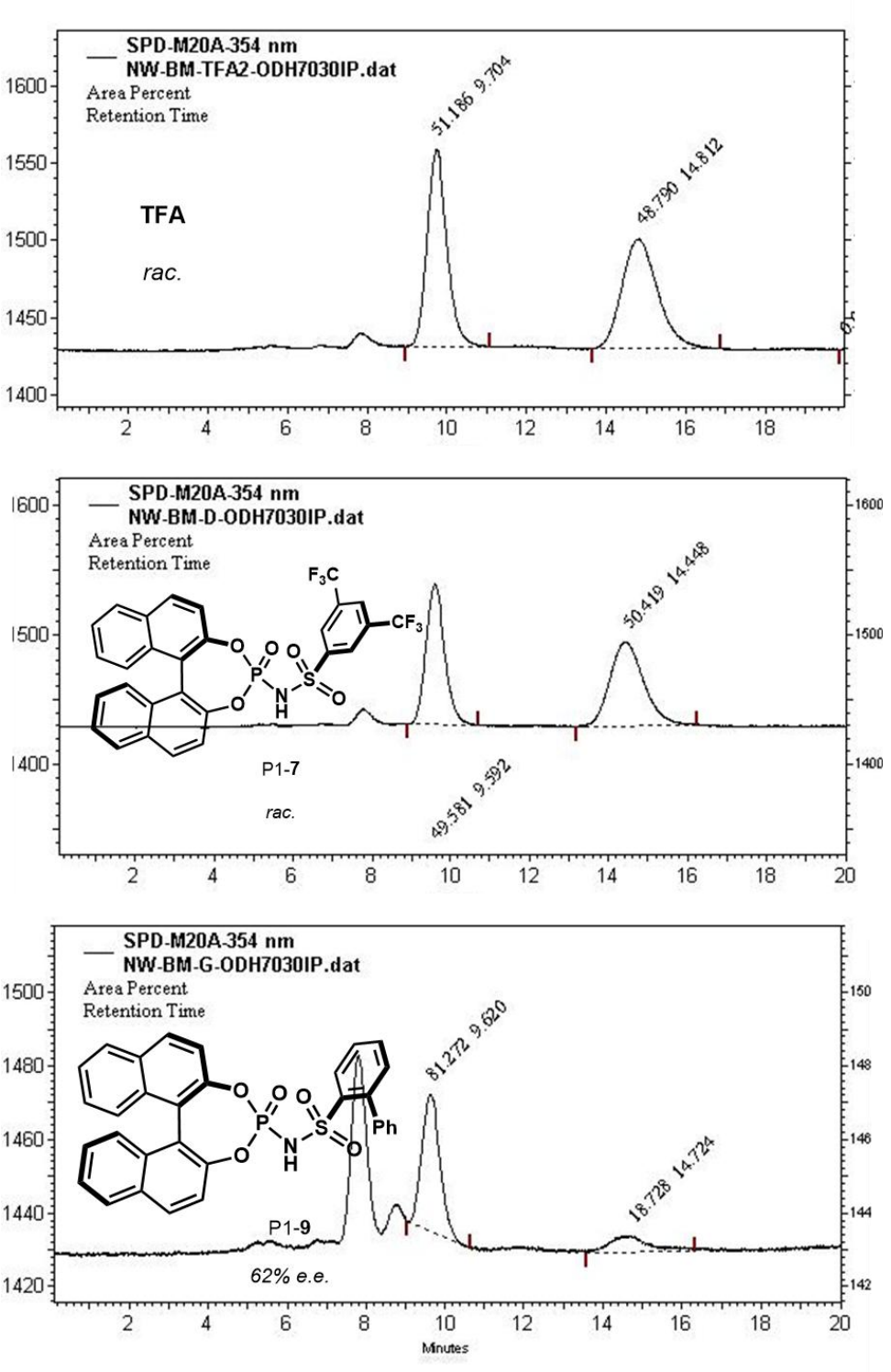
The observed ^{1}H NMR spectra for **3** and **4** are in accordance with literature report. The highlighted signal corresponds to the diagnostic signal for quantifying yield (**Supplementary Fig. V-X**).

3 – **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.21 (s, 1H), 7.55 (dq, *J* = 7.9, 0.9 Hz, 1H), 7.48 – 7.42 (m, 2H), 7.40 – 7.28 (m, 5H), 7.19 (ddd, *J* = 8.1, 7.0, 1.2 Hz, 1H), 7.12 (ddd, *J* = 7.9, 6.9, 1.1 Hz, 1H), 6.98 (dd, *J* = 5.1, 3.6 Hz, 1H), 6.85 (dd, *J* = 3.6, 1.3 Hz, 1H), 3.05 (s, 1H), 1.93 (s, 3H).

4 – **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.77 – 7.64 (m, 1H), 7.55 (dd, *J* = 7.2, 1.7 Hz, 1H), 7.49 – 7.45 (m, 2H), 7.38 – 7.33 (m, 5H), 7.31 – 7.28 (m, 3H), 7.23 – 7.20 (m, 1H), 7.16 – 7.10 (m, 2H), 6.94 (s, 1H), 6.69 (dd, *J* = 3.6, 1.1 Hz, 1H), 6.51 (dd, *J* = 3.7, 2.6 Hz, 1H), 6.40 (dd, *J* = 3.7, 2.5 Hz, 1H), 5.91 (s, 1H), 2.24 (s, 3H).



Supplementary Fig. VI-213 | Reference ^{1}H NMR spectra for yield measurement in the Friedel-Crafts arylation reported by Li, et al. *Nat. Comm.* **2021.**[9]

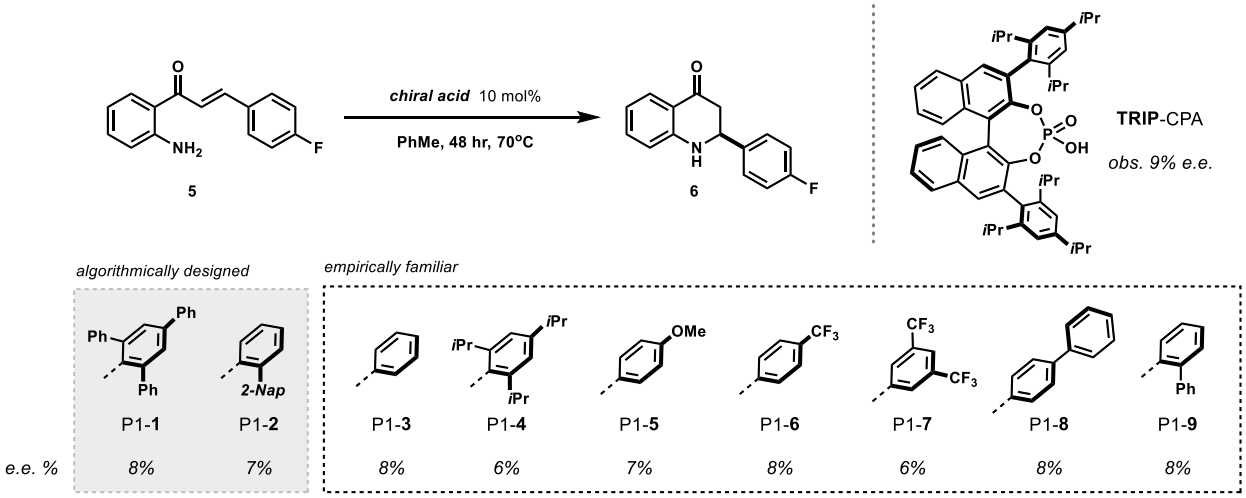


Supplementary Fig. VI-214 | Representative HPLC traces from the Friedel-Crafts arylation furnishing **4** (top – TFA, *rac.* standard; middle – P1-**7**, obs. *rac.*; bottom – P1-**9**, obs. 62% *e.e.*).

◊ *Aza-Michael addition*

The active learning campaign conducted was intentionally deployed to mimic **TRIP**-CPA in the Friedel-Crafts arylation. However, we expected that this design should also furnish *un*selective P1-SOAs when **TRIP-**CPA is also not expected to be stereoselective. This would validate the expected relationship between active site-based buried volume and stereoselectivity in a negative context.

To this end, we evaluated an intramolecular *aza*-Michael addition originally reported by Saito, K. Moriya, Y.; Akiyama, T. *Org. Lett.* **2015.**[10] In this transformation, we observed **TRIP**-CPA is completely ineffective (obs. 9% *e.e.* , see **Appendix IV** for reaction protocol and characterization). The reaction data collected across all P1-SOAs is recorded below (**Supplementary Fig. V-X**). Quantitative conversion of **3** was observed in all cases.



Supplementary Fig. VI-215 | Enantioenrichment (by HPLC) of **6** from the *aza*-Michael addition across the P1-SOAs screened in this work.

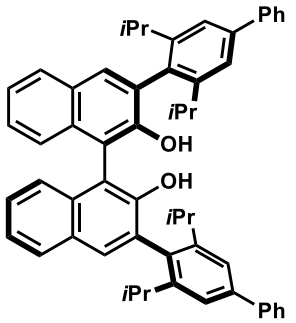
VI.46 – Synthetic Methods

We summarize synthesis of the following sets of compounds in this report: CPAs (from **Fig. V-4**), benzylidene substrates (**1a-1f**), benzofuranone products (**2a-2f**), and P1-SOAs (from **Fig. V-5**).

- ***Synthesis of CPAs***

We have previously reported the synthesis of **TRIP**-CPA, CPA-**1**, and CPA-**3** through CPA-**9**.[2]

CPA-**2** was prepared according to the sequence below, starting from the advanced intermediate – (*R*)-3,3'-diiodo-2,2'-bis(ethoxymethoxy)-1,1'-binaphthalene – prepared in two steps from commercially available (*R*)-1,1′-binaphthyl-2,2′-diol from known protocols.



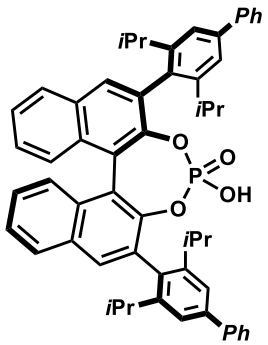
(*R*)-3,3'--bis(3,5-diisopropyl-[1,1'-biphenyl]-4-yl)-[1,1'-binaphthalene]-2,2'-diol: To a flame dried 100 mL round bottom flask equipped with a Teflon coated stir bar was added (*R*)-2,2'-bis(ethoxymethoxy)-3,3'-diiodo-1,1'-binaphthalene (107 mg, 160 umol, 1.0 equiv) and bistriphenylphosphine nickel (II) chloride (11 mg, 17 umol, 0.10 equiv). The flask was equipped with a rubber septum, evacuated and purged with nitrogen. To the purged flask was added diethyl ether (3 mL) to furnish a dark gray suspension. To the suspension was slowly added prepared **Grignard solution** (3 mL in THF, from 207 mg, 650 umol, 4.0 equiv from aryl bromide prepared according to known protocol[11]) (**note** – exothermic, solution turns a clear, red-brown solution). The suspension was lowered into an oil bath and heated at 40°C overnight.

After, the solution was cooled to room temperature and quenched with 25 mL of sat. NH_4Cl (aq). The suspension was transferred to a separatory funnel and extracted with 3x20 mL of DCM. The organic fractions were pooled, washed with 50 mL of sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish a crude oil which was carried forward without further purification.

To the flask containing the crude oil was charged 1,4-dioxane (0.1 M) and conc. HCl (12 N, 1.0 M). The flask was sealed with a rubber septa and the solution was heated to 70°C for 1 hour. After, the solution was quenched over a mixture of ice and sat. $NaHCO_3$ (aq) solution. Once the suspension no longer evolved carbon dioxide, the suspension was transferred to a separatory funnel and extracted with 3x20 mL ethyl acetate. The organic fractions were pooled, washed with 20 mL sat. NaCl (aq), dried over Na_2SO_4 and concentrated *in vacuo* to furnish a crude residue containing the **title compound**. The crude solid was subjected column chromatography (0-50% EtOAc/Hex)

to furnish the **title compound** (22 mg, 29 umol, 18% yield over two steps). Spectral characterization is consistent with literature.[11]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.96 – 7.91 (m, 2H), 7.84 (s, 2H), 7.71 – 7.66 (m, 4H), 7.54 – 7.50 (m, 4H), 7.45 – 7.43 (m, 2H), 7.41 – 7.37 (m, 4H), 7.33 (q, *J* = 4.1 Hz, 6H), 5.01 (s, 2H), 2.95 (p, *J* = 6.8 Hz, 2H), 2.79 (p, *J* = 6.8 Hz, 2H), 1.29 (d, *J* = 4.6 Hz, 6H), 1.20 (d, *J* = 6.9 Hz, 6H), 1.18 (d, *J* = 6.9 Hz, 6H), 1.12 (d, *J* = 6.9 Hz, 6H).



***(R)*-4-hydroxy-2,6-bis(3,5-diisopropyl-[1,1'-biphenyl]-4-yl)-dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide,** CPA-**2:** To a flame dried 20 mL reaction tube equipped with a Teflon coated stir bar was added **(*R*)-3,3'--bis(3,5-diisopropyl-[1,1'-biphenyl]-4-yl)-[1,1'-binaphthalene]-2,2'-diol** (22 mg, 29 umol, 1.0 equiv) and dichloromethane (3 mL). The reaction tube was equipped with a Teflon septum and charged with phosphorous oxychloride (5 uL, 60 umol, 2 equiv.) and triethylamine (12 uL, 90 umol, 3 equiv.) The solution was then stirred for 2 hr at room temperature. After, the solution was transferred to a separatory funnel, quenched with sat. $NaHCO_3$ (aq), and extracted with 3x20 mL DCM. The organic fractions were pooled, washed with sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish a crude residue that was carried forward without further purification.

To the flask containing crude phosphoryl chloride was charged pyridine (1 mL), and water (1 mL). The flask was sealed and heated to 80ºC for 2 hr. After, the solution was transferred to a separatory funnel, quenched with 20 mL of 1 N HCl (aq), and extracted with 3x20 mL DCM. The organic fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo* to furnish a foamy white solid containing the **title compound**. The crude solid was subjected to column chromatography (0-100% EtOAc/Hex) to furnish a mixed salt of the **title compound**. The isolated mixed phosphate was resuspended in 3 mL of dichloromethane and 3 mL of 3 N HCl (aq). The suspension was vortexed for 10 seconds. Then, the dichloromethane layer was separated by pipette and concentrated to furnish the **title compound** (4.1 mg, 5.0 umol, 17% yield over two steps). Spectral characterization is consistent with literature.[11]

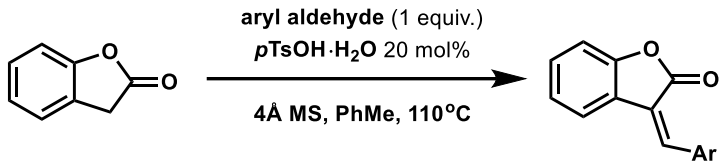
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93 (d, *J* = 8.2 Hz, 2H), 7.86 (s, 2H), 7.55 (d, *J* = 7.8 Hz, 6H), 7.43 – 7.34 (m, 8H), 7.34 – 7.28 (m, 6H), 2.64 (q, *J* = 6.7 Hz, 4H), 1.13 (d, *J* = 6.7 Hz, 6H), 1.03 (d, *J* = 6.7 Hz, 6H), 0.92 (d, *J* = 6.8 Hz, 6H), 0.85 (d, *J* = 6.9 Hz, 6H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 2.92.

- ***Preparation of benzylidenes***

Benzylidene substrates were prepared from the condensation of benzofuranone with the corresponding aryl aldehyde, according to **General Procedure A**. These substrates were used as described above in **Stereoselective reduction of benzylidenes**.

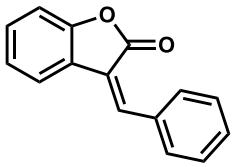
General Protocol A – condensation of benzofuranone with aryl aldehyde



Adapted from the reported literature protocol.[12] To a reaction flask was added **benzofuranone** (335 mg, 2.50 mmol, 1 equiv.), *p*-toluenesulfonic acid monohydrate (95 mg, 500 umol, 0.2 equiv.), the **corresponding aryl aldehyde** (2.50 mmol, 1 equiv.), and 4 Å molecular sieves (~250 mg). The mixture was suspended in toluene (6.3 mL, 0.4 M) and refluxed overnight.

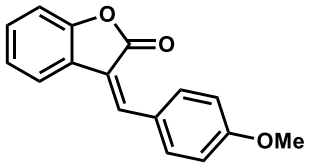
After, the solution was cooled and filtered over a fritted flask, rinsing with 1x30 mL ethyl acetate. The solution was transferred to a separatory funnel and washed with sat. $NaHCO_3$ then sat. NaCl (aq). The organic fraction was dried over Na_2SO_4, and concentrated *in vacuo* to furnish the **target benzylidene** as a crude oil. The oil was first subjected to column chromatography (0-20% EtOAc/Hex). Fractions corresponding to the product were then concentrated *in vacuo* to furnish the target compound. Further purification by recrystallization from hexanes or trituration with hot hexanes helped to resolve remaining aldehyde impurities (detected by ^{1}H NMR).

◊ *Benzylidene starting materials (1a-f)*



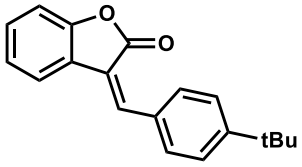
3-benzylidenebenzofuran-2(3H)-one, 1a: According to **General Protocol A** with benzaldehyde (180 uL, 2.50 mmol, 1 equiv.). Spectral characterization indicates a mixtures of geometric isomers, consistent with literature report.[12,13]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.19 (dd, *J* = 6.7, 3.0 Hz, 1H), 7.88 (s, 1H), 7.77 – 7.62 (m, 4H), 7.57 – 7.47 (m, 5H), 7.34 (tt, *J* = 7.7, 1.7 Hz, 2H), 7.21 – 7.08 (m, 2H), 7.04 (td, *J* = 7.7, 1.1 Hz, 1H).



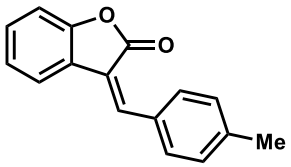
3-(4-methoxybenzylidene)benzofuran-2(3H)-one, 1b: According to **General Protocol A** with benzaldehyde (300 uL, 2.50 mmol, 1 equiv.). Spectral characterization is consistent with literature report.[12]

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.35 – 8.22 (m, 1H), 7.83 (d, *J* = 6.9 Hz, 1H), 7.70 (d, *J* = 8.6 Hz, 1H), 7.61 – 7.48 (m, 1H), 7.39 – 7.27 (m, 1H), 7.20 – 6.95 (m, 4H), 3.90 (d, *J* = 4.1 Hz, 3H).



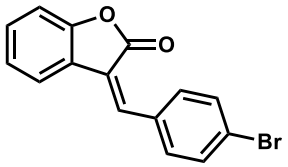
3-(4-(tert-butyl)benzylidene)benzofuran-2(3H)-one, 1c: According to **General Protocol A** with benzaldehyde (420 uL, 2.50 mmol, 1 equiv.).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (d, *J* = 6.6 Hz, 2H), 7.70 – 7.63 (m, 2H), 7.54 – 7.50 (m, 2H), 7.39 – 7.31 (m, 1H), 7.18 – 7.13 (m, 1H), 7.12 – 7.02 (m, 1H), 1.38 (s, 9H).



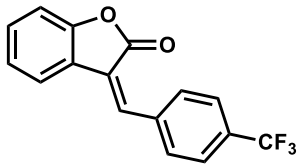
3-(4-methylbenzylidene)benzofuran-2(3H)-one, 1d: According to **General Protocol A** with benzaldehyde (300 uL, 2.50 mmol, 1 equiv.). Spectral characterization indicates a mixtures of geometric isomers, consistent with literature report.[12,13]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.13 (d, *J* = 8.0 Hz, 1H), 7.78 (dd, *J* = 8.4, 2.6 Hz, 1H), 7.66 – 7.57 (m, 1H), 7.39 – 7.28 (m, 3H), 7.21 – 6.96 (m, 2H), 2.44 (d, *J* = 7.2 Hz, 3H).



3-(4-bromobenzylidene)benzofuran-2(3H)-one, 1e: According to **General Protocol A** with benzaldehyde (460 mg, 2.50 mmol, 1 equiv.). Spectral characterization is consistent with literature report.[13]

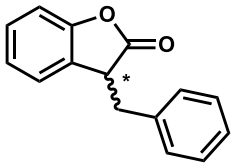
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.77 (s, 1H), 7.68 (dd, *J* = 7.9, 1.3 Hz, 1H), 7.66 – 7.62 (m, 2H), 7.58 – 7.54 (m, 2H), 7.37 (td, *J* = 7.8, 1.3 Hz, 1H), 7.16 (d, *J* = 8.0 Hz, 1H), 7.06 (td, *J* = 7.7, 1.0 Hz, 1H).



3-(4-(trifluoromethyl)benzylidene)benzofuran-2(3H)-one, 1f: According to **General Protocol A** with benzaldehyde (340 uL, 2.50 mmol, 1 equiv.). Spectral characterization is consistent with literature report.[13]

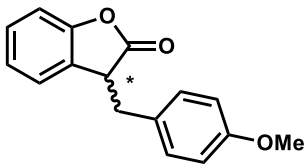
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (s, 1H), 7.82 – 7.74 (m, 4H), 7.60 (dd, *J* = 7.8, 1.3 Hz, 1H), 7.39 (td, *J* = 7.9, 1.3 Hz, 1H), 7.17 (dt, *J* = 8.1, 0.8 Hz, 1H), 7.06 (td, *J* = 7.7, 1.0 Hz, 1H).

◊ *Benzofuranone products (2a-f)*



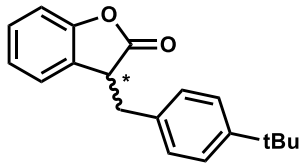
3-benzylbenzofuran-2(3H)-one, 2a: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (76% yield, 62% *e.e.* – ChiralPak OJH 95:05 hexanes/isopropanol, t_R = 10.9, 11.6 min, 1mL/min). Spectral characterization is consistent with literature report.[14]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.27 (m, 2H), 7.24 – 7.20 (m, 2H), 7.18 – 7.11 (m, 2H), 7.06 – 6.95 (m, 2H), 6.78 (d, *J* = 7.4 Hz, 1H), 3.99 (dd, *J* = 9.0, 4.9 Hz, 1H), 3.49 (dd, *J* = 13.8, 4.8 Hz, 1H), 3.02 (dd, *J* = 13.8, 9.0 Hz, 1H).



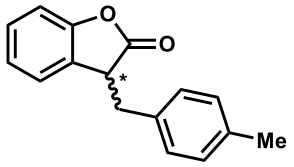
3-(4-methoxybenzyl)benzofuran-2(3H)-one, 2b: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (46% yield, 63% *e.e.* – ChiralPak ADH 95:05 hexanes/isopropanol, t_R = 7.7, 8.7 min, 1mL/min). Spectral characterization is consistent with literature report.[15,16]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.24 (dt, *J* = 7.7, 1.1 Hz, 1H), 7.07 – 7.00 (m, 4H), 6.87 – 6.83 (m, 1H), 6.83 – 6.73 (m, 2H), 3.96 (dd, *J* = 8.8, 4.8 Hz, 1H), 3.78 (s, 3H), 3.42 (dd, *J* = 13.9, 4.9 Hz, 1H), 3.00 (dd, *J* = 13.9, 8.8 Hz, 1H).



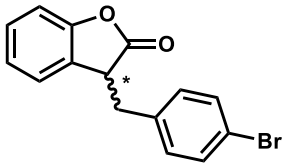
3-(4-(tert-butyl)benzyl)benzofuran-2(3H)-one, 2c: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (49% yield, 72% *e.e.* – ChiralPak ADH 98:02 hexanes/isopropanol, t_R = 6.0, 6.4 min, 1mL/min).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.53 – 7.50 (m, 1H), 7.31 – 7.28 (m, 2H), 7.11 – 7.07 (m, 2H), 7.03 (ddd, *J* = 7.5, 4.1, 1.0 Hz, 2H), 6.80 (dd, *J* = 7.5, 1.0 Hz, 1H), 3.97 (dd, *J* = 9.3, 4.7 Hz, 1H), 3.46 (dd, *J* = 13.8, 4.7 Hz, 1H), 2.97 (dd, *J* = 13.8, 9.3 Hz, 1H), 1.30 (s, 9H).



3-(4-methylbenzyl)benzofuran-2(3H)-one, 2d: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (67% yield, 63% *e.e.* – ChiralPak ADH 98:02 hexanes/isopropanol, t_R = 7.2, 8.4 min, 1mL/min). Spectral characterization is consistent with literature report.[15,16]

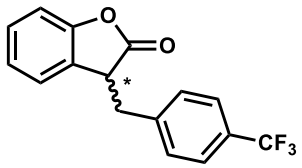
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 – 7.27 (m, 6H), 7.03 – 7.02 (m, 2H), 6.82 (dt, *J* = 8.0, 1.2 Hz, 1H), 3.98 (dd, *J* = 9.0, 4.8 Hz, 1H), 3.45 (dd, *J* = 13.8, 4.8 Hz, 1H), 2.99 (dd, *J* = 13.8, 9.0 Hz, 1H), 2.32 (s, 3H).



3-(4-bromobenzyl)benzofuran-2(3H)-one, 2e: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (51% yield, 62% *e.e.* – ChiralPak ADH 98:02 hexanes/isopropanol, t_R = 10.5, 13.1 min, 1mL/min). Spectral characterization is consistent with literature report.[15,16]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 (m, 2H), 7.27 (m, 1H), 7.04 (m, 4H), 6.89 (d, *J* = 7.5 Hz, 1H), 3.99 (dd, *J* = 8.4, 4.9 Hz, 1H), 3.41 (dd, *J* = 13.9, 4.9 Hz, 1H), 3.06 (dd, *J* = 13.9, 8.4 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 176.21, 153.77, 135.53, 131.76, 131.13, 129.19, 126.29, 124.70, 124.04, 121.25, 110.93, 44.86, 36.50.



3-(4-(trifluoromethyl)benzyl)benzofuran-2(3H)-one, 2f: According to the general reaction protocol for the stereoselective reduction of benzylidenes, with **TRIP**-CPA (55% yield, 69% *e.e.* – ChiralPak ADH 98:02 hexanes/isopropanol, t_R = 9.2, 10.9 min, 1mL/min).

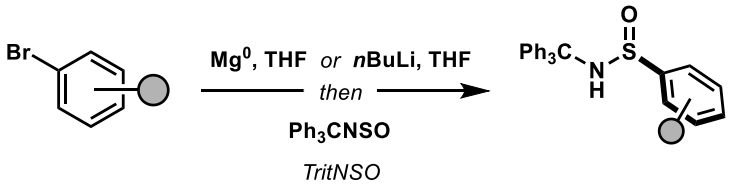
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.52 (d, *J* = 8.0 Hz, 2H), 7.27 – 7.20 (m, 3H), 7.09 – 7.00 (m, 2H), 6.93 – 6.80 (m, 1H), 4.03 (dd, *J* = 8.4, 5.0 Hz, 1H), 3.49 (dd, *J* = 13.8, 5.0 Hz, 1H), 3.16 (dd, *J* = 13.8, 8.3 Hz, 1H).

^{19}F NMR (376 MHz, $CDCl_3$) δ -62.43.

- ***Synthesis of P1-SOAs***

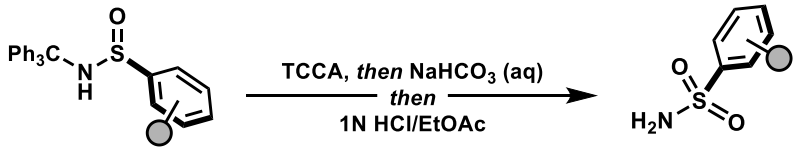
Compared to many BINOL-derived BAOCs, P1-SOAs are expediently accessed in one-step from phosphorylation of the corresponding sulfonamide. Owing to their prevalence in drug-like molecules, some sulfonamides are commercially available and many methods accessing them have been reported. In this report, sulfonamide precursors to P1-**3** and P1-**5** were commercially sourced. The remaining sulfonamides were prepared from **General Procedures B-E** below.

General Protocol B – preparation of *N*-tritylsulfinamides from *N*-tritylsulfinylamine.



According to the protocol previously reported (see **Appendix IV – General Protocol G**).

General Protocol C – preparation of sulfonamide from *N*-tritylsulfinamide

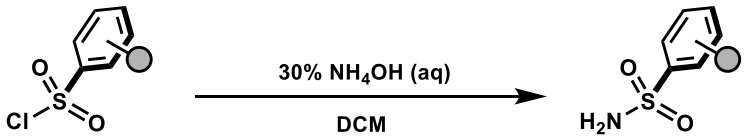


Adapted from the protocol previously reported (see **Appendix IV – General Protocol H**). To a flame dried 20 mL scintillation vial containing a Teflon coated stir bar was charged the corresponding ***S*-aryl sulfinamide** (1.0 equiv) and THF (0.1 M with respect to the sulfinamide). To the solution was charge *tert*-butyl hypochlorite (1.1 equiv). The solution was stirred at room temperature for 15 minutes and then charged with sat. $NaHCO_3$ (1.0 M with respect to the sulfinamide) in one portion. The suspension was stirred at room temperature overnight.

After, the suspension was transferred to a separatory funnel, rinsing the reaction vial with equal parts ethyl acetate and water. The solution was diluted with ethyl acetate (20 mL) and washed with sat. ammonium chloride (20 mL). The organic layer was separated, washed with sat. sodium chloride (20 mL), dried over sodium sulfate, and concentrated *in vacuo* to furnish a crude residue containing trityl-protected sulfonamide. The residue was transferred to another 20 mL scintillation vial containing a Teflon coated stir bar using DCM (5 mL) and then charged with 1N HCl in ethyl acetate (3 equiv. with respect to the sulfinamide). The solution was stirred at room temperature overnight to furnish a cloudy, yellow suspension.

The acidic suspension was diluted with hexanes and carefully transferred to a centrifuge tube (note – if no suspension is visible, then the solution was concentrated and reconstituted in hexanes before transferring to the centrifuge tube). The suspension was centrifuged (1,000 rpm for 5 min) and the filtrate was decanted. The pellet was washed with hexanes and the process was repeated one more time. To the crude HCl salt containing the **title compound** was charged sparing volumes of DCM and triethylamine. The suspension was then immediately subjected to flash column chromatography (0-100% EtOAc/Hex) to furnish the **title compound.**

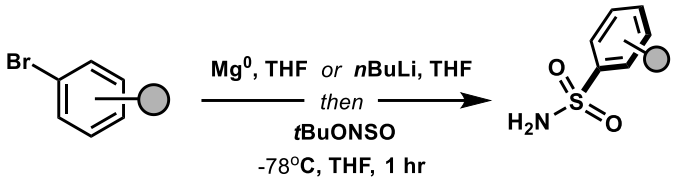
General Protocol D – aminolysis of sulfonyl chlorides.



To a reaction flask was charged **sulfonyl chloride**, DCM (0.1 M), and a Teflon coated stir bar. To the rapidly stirring solution was then added 1 volume of 30% NH_4OH (caution – exothermic). The resulting suspension was stirred vigorously overnight. After, the suspension was transferred to a separatory funnel, diluted with 50 mL DCM and washed with 2x30 mL water. The organic fraction was washed with sat. NaCl (aq), dried over Na_2SO_4, and concentrated *in vacuo* to furnish the target

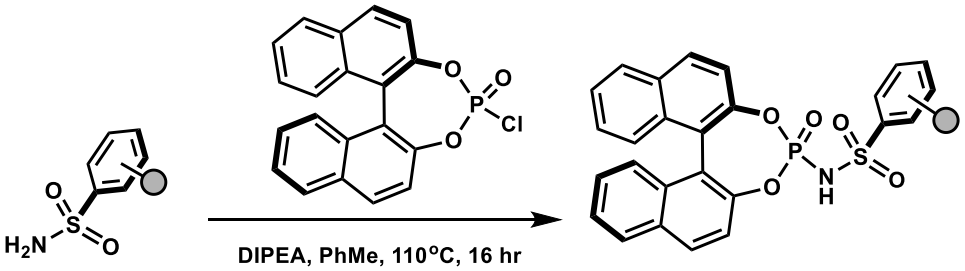
product as a white solid. If isolated in insufficient purity (based on ^{1}H NMR), the solid could be purified by column chromatography (0-100% EtOAc/Hex) for further resolution.

General Protocol E – direct access to sulfonamides from *t*BuONSO



According to the reported protocol by Willis and coworkers, primary sulfonamides can be directly prepared from the addition of an aryl organometallic to *t*BuONSO.

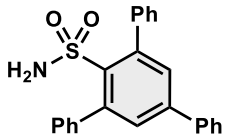
General Protocol F – phosphorylation of sulfonamides, P1-SOA synthesis



Adapted from the preparation of related, *N*-triflylphosphoramides (see **Appendix IV – General Protocol F**). To a flame dried 20 mL reaction tube equipped with a Teflon coated stir bar was added *(R)*-binaphthyl phosphoryl chloride (1.0 equiv), **corresponding sulfonamide** (2.0 equiv), and toluene (0.1 M). To the suspension was charged diisopropylethylamine (4.0 equiv). The reaction tube was sealed and heated to 110°C overnight. After, the solution was transferred to a separatory funnel, quenched with 20 mL of 1 N HCl (aq), and extracted with 3x20 mL DCM. The organic fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo* to furnish a crude solid containing the **title compound**.

The crude solid was subjected to column chromatography 0-100% EtOAc/Hex to furnish a mixed salt of the **title compound**. The isolated salt was resuspended in 3-5 mL of dichloromethane and 3-5 mL of 3 N HCl (aq). The suspension was vortexed for 10 seconds. Then, the dichloromethane layer was separated by glass pipette and concentrated to furnish the **corresponding P1-SOA**.

◊ *Preparation of precursor sulfonamide*



5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfonamide: *N*-trityl-5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfinamide was first prepared from *N*-tritylsulfinylamine (305 mg, 1.00 mmol, 1 equiv) and 2'-bromo-5'-phenyl-1,1':3',1''-terphenyl (454 mg, 1.20 mmol, 1.2 equiv.) according to **General Protocol B**, furnishing a white solid (250 mg, 410 umol, 41%).

N-trityl-5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfinamide

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.66 – 7.51 (m, 7H), 7.48 – 7.29 (m, 10H), 7.15 – 7.03 (m, 9H), 6.87 – 6.79 (m, 6H), 5.47 (s, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 144.47, 142.76, 142.21, 142.02, 139.38, 139.32, 130.57, 129.78, 129.01, 128.96, 128.54, 128.25, 128.19, 128.03, 127.60, 127.28, 126.86, 73.11.

HRMS (ESI- Expected [M+H]$^+$ = 612.2356, **Observed [M+H]$^+$** = 612.2355.

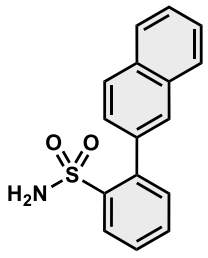
The intermediate *N*-trityl-5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfinamide was then subjected to **General Protocol C** furnishing the **title compound** (45 mg, 120 umol, 29%).

5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfonamide

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.61 – 7.44 (m, 8H), 7.44 – 7.36 (m, 7H), 7.36 – 7.31 (m, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 143.17, 142.39, 141.21, 138.64, 130.20, 129.16, 129.06, 128.62, 128.51, 128.26, 127.34.

HRMS (ESI+) Expected [M+H]$^+$ = 386.1209, **Observed [M+H]$^+$** = 386.1211.



2-(naphthalen-2-yl)benzenesulfonamide: *N*-trityl-2-(naphthalen-2-yl)benzenesulfinamide was first prepared from *N*-tritylsulfinylamine (305 mg, 1.00 mmol, 1 equiv) and 2-(2-bromophenyl)naphthalene (334 mg, 1.20 mmol, 1.2 equiv.) according to **General Protocol B**, furnishing a white solid (300 mg, 590 umol, 59%).

N-trityl-2-(naphthalen-2-yl)benzenesulfinamide

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.29 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.77 – 7.71 (m, 1H), 7.64 – 7.35 (m, 7H), 7.32 (dd, *J* = 7.4, 1.4 Hz, 1H), 7.21 (dd, *J* = 8.4, 1.8 Hz, 1H), 7.13 – 7.00 (m, 3H), 6.94 (t, *J* = 7.8 Hz, 6H), 6.80 – 6.69 (m, 6H), 4.79 (s, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 146.35, 143.87, 140.63, 136.08, 132.94, 132.55, 130.83, 129.63, 128.71, 128.35, 128.29, 127.79, 127.58, 127.50, 127.47, 127.23, 126.37, 126.29, 123.56, 73.45.

HRMS (ESI- Expected $[M+H]^+$ = 510.1886, **Observed $[M+H]^+$** = 510.1880.

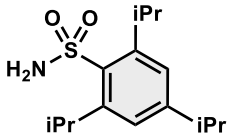
The intermediate *N*-trityl-2-(naphthalen-2-yl)benzenesulfinamide was then subjected to **General Protocol C** furnishing the **title compound** (86 mg, 300 umol, 52%).

2-(naphthalen-2-yl)benzenesulfonamide

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.20 (dd, *J* = 8.0, 1.4 Hz, 1H), 8.02 – 7.83 (m, 4H), 7.68 – 7.50 (m, 5H), 7.44 (dd, *J* = 7.5, 1.4 Hz, 1H), 4.15 (s, 2H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 140.96, 140.06, 136.38, 132.88, 132.75, 132.56, 132.30, 128.74, 128.30, 128.08, 127.91, 127.83, 127.67, 127.32, 126.95, 126.90.

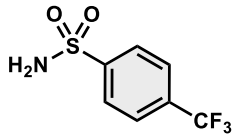
HRMS (ESI+) Expected $[M+H]^+$ = 284.0740, **Observed $[M+H]^+$** = 284.0740.



2,4,6-triisopropylbenzenesulfonamide: Prepared according to **General Protocol D** from 2,4,6-triisopropylbenzenesulfonyl chloride (3.03 g, 10.0 mmol, 1 equiv), furnishing the **title compound** (1.48 g, 5.2 mmol, 52%). Spectral characterization is consistent with literature report.[17]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.17 (s, 2H), 4.84 (d, *J* = 19.6 Hz, 2H), 4.11 (p, *J* = 6.7 Hz, 2H), 2.90 (p, *J* = 6.9 Hz, 1H), 1.29 (d, *J* = 6.8 Hz, 12H), 1.25 (d, *J* = 6.9 Hz, 6H).

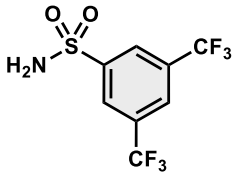
^{13}C NMR (126 MHz, $CDCl_3$) δ 152.89, 149.31, 134.98, 134.96, 123.82, 34.32, 29.93, 24.87, 23.75.



4-trifluoromethylbenzenesulfonamide: Prepared according to **General Protocol D** from 4-(trifluoromethyl)benzenesulfonyl chloride (1.00 g, 4.09 mmol, 1 equiv), furnishing the **title compound** (100 mg, 440 umol, 11%). Spectral characterization is consistent with literature report.[18]

^{1}H NMR (500 MHz, CDCl3) δ 8.09 (d, J = 8.9 Hz, 2H), 7.83 (d, J = 8.2 Hz, 2H), 4.91 (s, 2H).

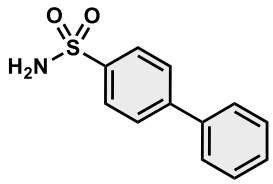
^{19}F NMR (376 MHz, CDCl3) δ -63.13.



3,5-bistrifluoromethylbenzenesulfonamide: Prepared according to **General Protocol D** from 3,5-bis(trifluoromethyl)benzenesulfonyl chloride (250 mg, 800 umol, 1 equiv), furnishing the **title compound**. Spectral characterization is consistent with literature report.[19]

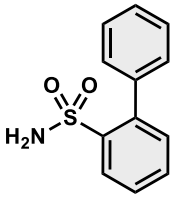
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.39 (s, 2H), 8.09 (s, 1H), 5.03 (s, 2H).

^{19}F NMR (471 MHz, $CDCl_3$) δ -62.97.



[1,1'-biphenyl]-4-sulfonamide: Prepared according to **General Protocol E** from 4-bromo-1,1'-biphenyl (233 mg, 1.00 mmol, 1 equiv), furnishing the **title compound** (20 mg, 84 umol, 8%). Spectral characterization is consistent with literature report.[20]

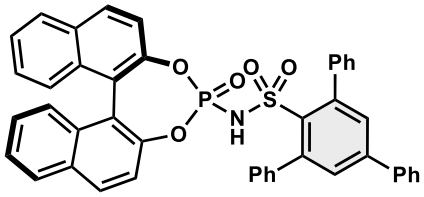
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.01 – 7.96 (m, 2H), 7.73 (d, J = 8.5 Hz, 2H), 7.51 – 7.41 (m, 5H), 4.83 (s, 2H).



[1,1'-biphenyl]-2-sulfonamide: Prepared according to **General Protocol E** from 2-bromo-1,1'-biphenyl (233 mg, 1.00 mmol, 1 equiv), furnishing the **title compound** (23 mg, 99 umol, 10%). Spectral characterization is consistent with literature report.[21]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.11 (dd, *J* = 8.0, 1.3 Hz, 1H), 7.62 – 7.56 (m, 1H), 7.51 – 7.39 (m, 6H), 7.34 (dd, *J* = 7.5, 1.4 Hz, 1H), 4.20 (s, 2H).

◊ *Phosphorylation of sulfonamide*

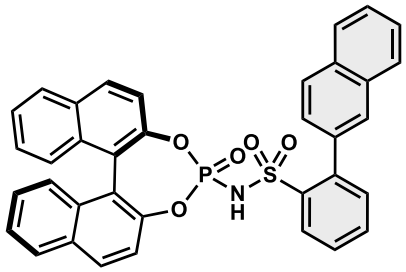


N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfonamide, P1-**1:** Prepared according to **General Protocol F**from 5'-phenyl-[1,1':3',1''-terphenyl]-2'-sulfonamide (19 mg, 49 umol, 1 equiv), furnishing the **title compound** (6 mg, 9 umol, 18%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.99 (d, *J* = 8.9 Hz, 1H), 7.95 – 7.85 (m, 3H), 7.69 – 7.64 (m, 2H), 7.60 (d, *J* = 5.5 Hz, 6H), 7.51 – 7.33 (m, 13H), 7.32 – 7.27 (m, 4H).

^{31}P NMR (202 MHz, $CDCl_3$) δ -3.00.

HRMS (ESI+) Expected $[M+H]^+$ = 716.1655, **Observed $[M+H]^+$** = 716.1650.



2-(naphthalen-2-yl)-N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzene sulfonamide, P1-**2:** Prepared according to **General Protocol F** from 2-(naphthalen-2-

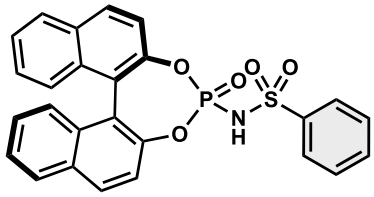
yl)benzenesulfonamide (28 mg, 100 umol, 1 equiv), furnishing the **title compound** (9 mg, 14 umol, 15%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.32 (dd, *J* = 8.1, 1.3 Hz, 1H), 7.99 (d, *J* = 9.0 Hz, 1H), 7.96 – 7.86 (m, 4H), 7.86 – 7.74 (m, 3H), 7.69 (td, *J* = 7.6, 1.4 Hz, 1H), 7.62 – 7.53 (m, 2H), 7.50 – 7.37 (m, 7H), 7.31 – 7.26 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 140.78, 138.82, 136.26, 133.21, 132.77, 132.72, 132.63, 132.16, 132.08, 131.95, 131.81, 131.69, 131.43, 129.42, 128.59, 128.51, 128.20, 128.06, 127.77, 127.65, 127.45, 127.11, 127.04, 126.95, 126.86, 126.65, 126.09, 126.00, 121.25, 121.15, 120.49, 120.39.

^{31}P NMR (202 MHz, $CDCl_3$) δ -2.74.

HRMS (ESI+) Expected $[M+H]^+$ = 614.1186, **Observed $[M+H]^+$** = 614.1184.



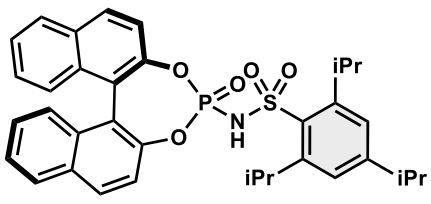
N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)benzenesulfonamide, P1-**3:** Prepared according to **General Protocol F** from benzenesulfonamide (11 mg, 67 mmol, 1 equiv), furnishing the **title compound** (5 mg, 9 umol, 14%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10 – 8.02 (m, 3H), 7.99 – 7.90 (m, 3H), 7.60 (t, *J* = 8.7 Hz, 2H), 7.53 – 7.43 (m, 5H), 7.37 – 7.27 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 146.78, 145.39, 140.34, 133.78, 132.31, 132.10, 132.02, 131.88, 131.60, 129.20, 128.75, 128.67, 127.73, 127.24, 127.18, 127.09, 127.06, 126.22, 126.18, 121.44, 120.84, 120.60.

^{31}P NMR (202 MHz, $CDCl_3$) δ -1.84.

HRMS (ESI+) Expected $[M+H]^+$ = 488.0716, **Observed $[M+H]^+$** = 488.0716.



N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-2,4,6-triisopropylbenzene sulfonamide, P1-**4:** Prepared according to **General Protocol F** from 2,4,6-

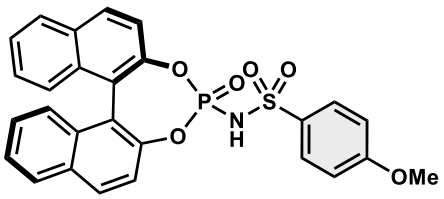
triisopropylbenzenesulfonamide (283 mg, 1.00 mmol, 1 equiv), furnishing the **title compound** (241 mg, 390 umol, 39%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.12 (d, *J* = 8.9 Hz, 1H), 8.02 – 7.90 (m, 3H), 7.77 (d, *J* = 8.9 Hz, 1H), 7.56 – 7.44 (m, 3H), 7.40 – 7.29 (m, 4H), 7.19 (s, 2H), 4.00 (p, *J* = 6.7 Hz, 2H), 2.91 (p, *J* = 6.9 Hz, 1H), 1.29 – 1.21 (m, 18H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 153.64, 152.70, 150.46, 149.17, 147.04, 146.95, 145.30, 145.23, 134.92, 133.38, 132.22, 132.19, 131.97, 131.89, 131.65, 131.36, 128.64, 128.50, 127.15, 127.03, 126.95, 126.81, 126.08, 125.95, 124.03, 123.67, 121.40, 121.39, 121.19, 121.16, 120.55, 120.53, 34.23, 29.83, 24.83, 24.76, 24.45, 23.58, 23.55, 14.19.

^{31}P NMR (202 MHz, $CDCl_3$) δ -1.45.

HRMS (ESI+) Expected $[M+H]^+$ = 614.2125, **Observed $[M+H]^+$** = 624.2131.



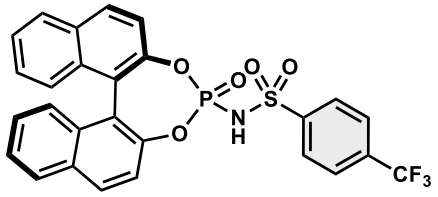
N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-4-methoxybenzene sulfonamide, P1-**5:** Prepared according to **General Protocol F** from 4-methoxybenzene sulfonamide (187 mg, 1.00 mmol, 1 equiv), furnishing the **title compound** (169 mg, 330 umol, 3%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.05 (d, *J* = 8.9 Hz, 1H), 8.00 – 7.87 (m, 3H), 7.87 – 7.77 (m, 1H), 7.62 (dd, *J* = 8.9, 1.0 Hz, 1H), 7.53 – 7.37 (m, 2H), 7.36 – 7.17 (m, 3H), 6.92 (dq, *J* = 8.0, 3.0 Hz, 2H), 3.84 (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 163.68, 162.90, 146.77, 146.68, 145.29, 145.22, 133.69, 132.18, 132.12, 131.96, 131.87, 131.68, 131.58, 131.44, 129.95, 128.59, 128.53, 127.11, 127.02, 126.94, 126.86, 126.08, 126.00, 121.30, 121.28, 121.21, 121.19, 120.81, 120.78, 120.47, 120.45, 114.20, 55.68.

^{31}P NMR (162 MHz, $CDCl_3$) δ -0.94.

HRMS (ESI+) Expected $[M+H]^+$ = 518.0822, **Observed $[M+H]^+$** = 518.0821.



N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)- 4-trifluoromethyl benzene sulfonamide, P1-**6:** Prepared according to **General Protocol F** from 4-(trifluoromethyl)benzenesulfonamide (23 mg, 100 umol, 1 equiv), furnishing the **title compound** (15 mg, 27 umol, 27%).

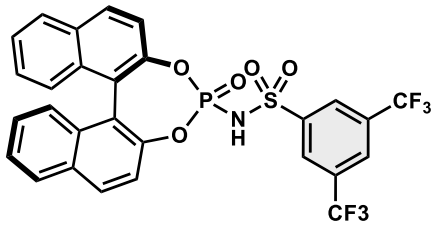
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.14 (d, *J* = 8.2 Hz, 2H), 8.09 (d, *J* = 8.8 Hz, 1H), 8.00 – 7.94 (m, 2H), 7.94 (d, *J* = 8.2 Hz, 2H), 7.74 (d, *J* = 8.1 Hz, 2H), 7.64 (d, *J* = 8.9 Hz, 1H), 7.51 (ddd, *J* = 8.1, 5.7, 2.2 Hz, 2H), 7.44 (d, *J* = 8.9 Hz, 1H), 7.31 (ddd, *J* = 8.3, 5.7, 1.4 Hz, 4H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 132.14, 131.84, 131.60, 128.63, 128.57, 128.25, 127.09, 127.02, 126.23, 126.20, 121.27, 120.53, 120.25.

^{19}F NMR (470 MHz, $CDCl_3$) δ -63.19.

^{31}P NMR (202 MHz, $CDCl_3$) δ -2.20.

HRMS (ESI+) Expected $[M+H]^+$ = 556.0590, **Observed $[M+H]^+$** = 556.0589.



N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-3,5-bistrifluoromethyl benzenesulfonamide, P1-**7:** Prepared according to **General Protocol F** from 3,5-bis(trifluoromethyl)benzenesulfonamide (59 mg, 200 umol, 1 equiv), furnishing the **title compound** (44 mg, 71 umol, 35%).

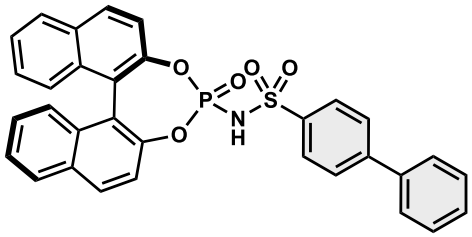
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.39 (d, *J* = 1.6 Hz, 2H), 8.08 – 7.90 (m, 5H), 7.58 – 7.42 (m, 3H), 7.36 (d, *J* = 8.9 Hz, 1H), 7.34 – 7.17 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 146.47, 146.37, 144.98, 144.91, 142.54, 133.12, 132.84, 132.56, 132.29, 132.13, 132.07, 131.83, 131.75, 131.70, 128.62, 128.59, 128.12, 127.11, 127.08, 127.06, 127.03, 126.28, 126.17, 125.58, 123.40, 121.23, 121.10, 121.08, 120.97, 120.95, 120.33, 120.30, 120.05, 120.03.

^{19}F NMR (470 MHz, $CDCl_3$) δ -62.87.

^{31}P NMR (202 MHz, $CDCl_3$) δ -2.14.

HRMS (ESI+) Expected $[M+H]^+$ = 624.0464, **Observed $[M+H]^+$** = 624.0566.



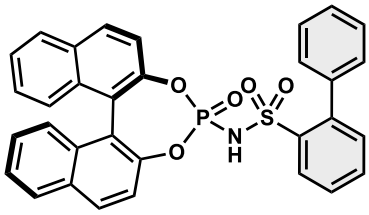
N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-[1,1'-biphenyl]-4-sulfonamide, P1-**8:** Prepared according to **General Protocol F** from [1,1'-biphenyl]-4-sulfonamide (7 mg, 28 mmol, 1 equiv), furnishing the **title compound** (6 mg, 10 umol, 35%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.12 (dd, *J* = 8.7, 4.9 Hz, 3H), 7.98 (dd, *J* = 9.1, 2.7 Hz, 2H), 7.94 (d, *J* = 8.2 Hz, 1H), 7.71 (dd, *J* = 16.5, 8.5 Hz, 3H), 7.63 – 7.55 (m, 2H), 7.50 (m, 5H), 7.46 – 7.40 (m, 1H), 7.38 – 7.28 (m, 5H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 145.20, 139.19, 132.19, 131.99, 131.93, 131.84, 131.52, 129.07, 128.64, 128.55, 128.24, 127.73, 127.43, 127.12, 127.04, 127.00, 126.12, 121.36, 120.71, 120.42.

^{31}P NMR (202 MHz, $CDCl_3$) δ -1.99.

HRMS (ESI+) Expected $[M+H]^+$ = 564.1029, **Observed $[M+H]^+$** = 564.1026.



N-(4-oxidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)-[1,1'-biphenyl]-2-sulfonamide P1-**9:** Prepared according to **General Protocol F** from [1,1'-biphenyl]-2-sulfonamide (15 mg, 64 mmol, 1 equiv), furnishing the **title compound** (10 mg, 18 umol, 28%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.25 (dd, *J* = 8.1, 1.3 Hz, 1H), 7.97 (m, 4H), 7.72 – 7.60 (m, 1H), 7.57 – 7.43 (m, 7H), 7.39 – 7.29 (m, 8H)

^{13}C NMR (126 MHz, $CDCl_3$) δ 146.59, 145.12, 140.76, 138.77, 138.69, 133.18, 132.41, 132.19, 132.11, 131.97, 131.86, 131.73, 131.45, 129.50, 129.28, 128.64, 128.53, 128.37, 128.09, 127.96, 127.12, 127.05, 126.98, 126.90, 126.11, 126.05, 121.28, 121.17, 120.70, 120.67, 120.42.

^{31}P NMR (202 MHz, $CDCl_3$) δ -2.78.

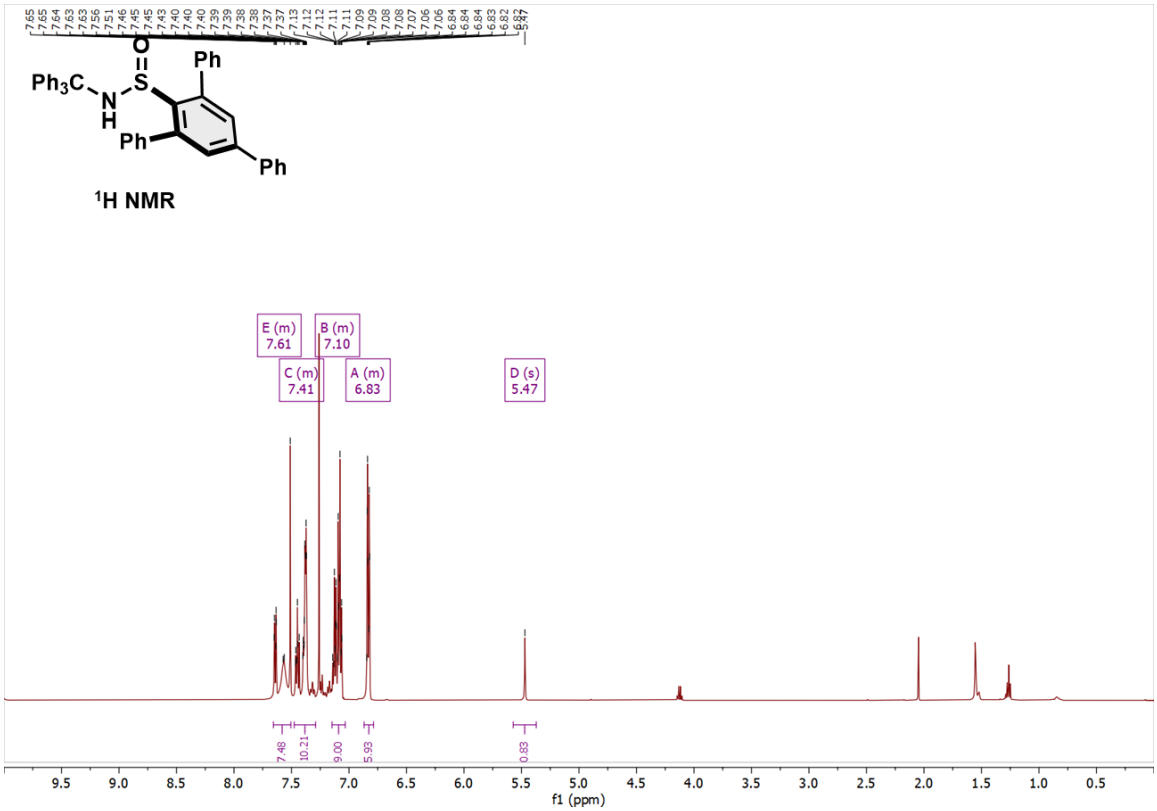
HRMS (ESI+) Expected $[M+H]^+$ = 564.1029, **Observed $[M+H]^+$** = 564.1029.

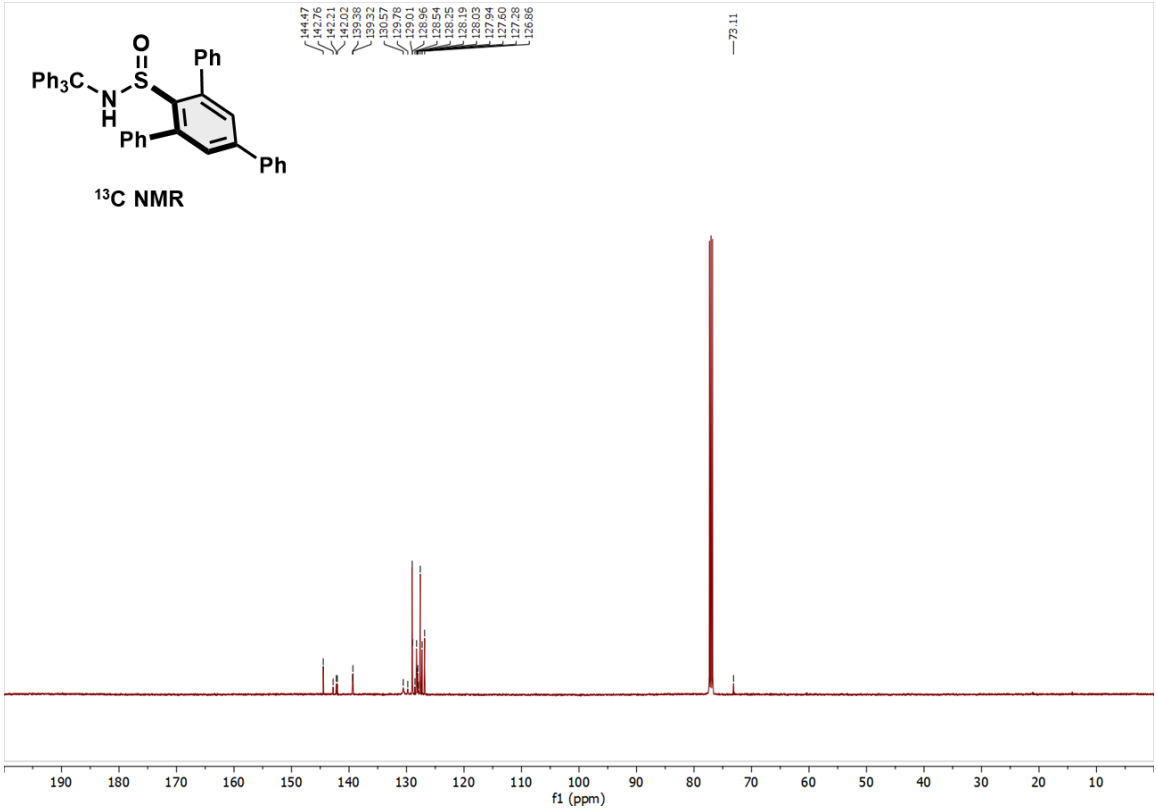
VI.47 – References

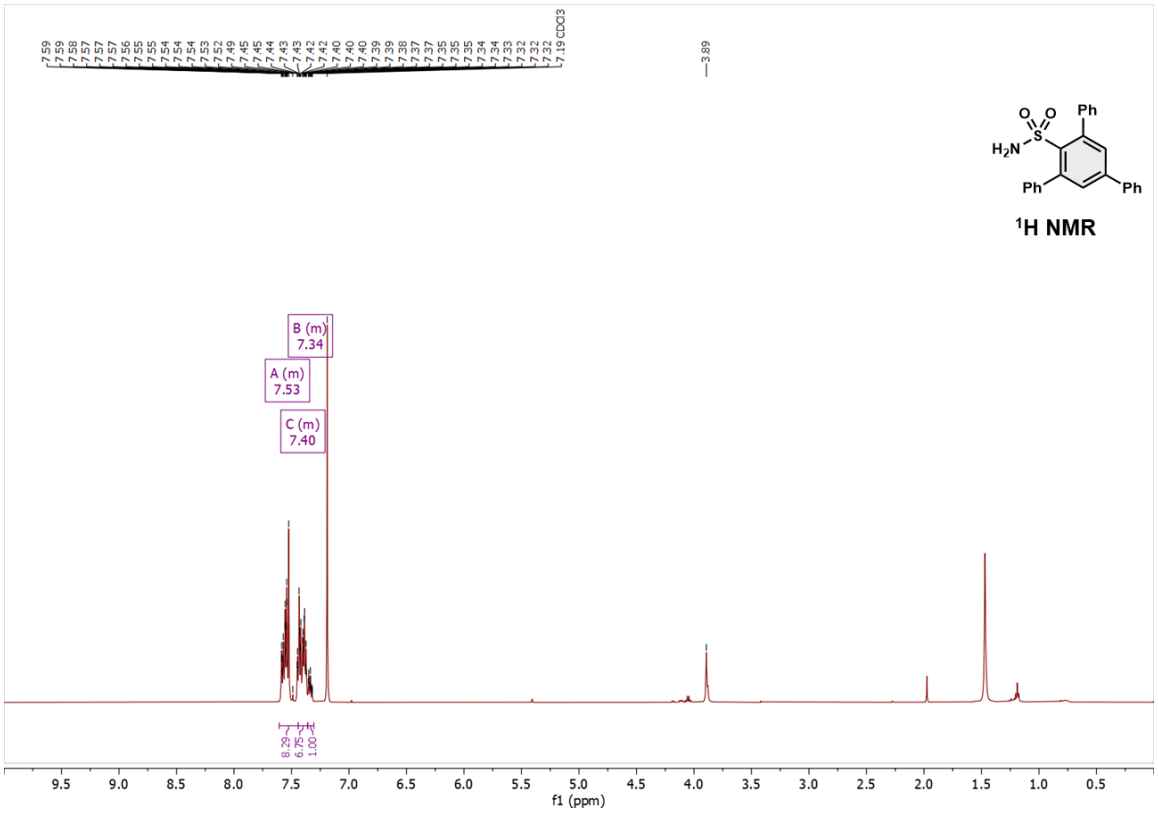
(1) Doyle-Lab-Ucla/Edboplus, 2024.
(2) Smith, A. L.; Toste, F. D. Stereoselective Generalizations over Diverse Sets of Chiral Acids Enabled by Buried Volume. *J. Am. Chem. Soc.* **2026**, jacs.5c20342.
(3) Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022**, *144* (43), 19999–20007.
(4) *RDKit*.
(5) Fried, S. D.; Boxer, S. G. Electric Fields and Enzyme Catalysis. *Annual Review of Biochemistry* **2017**, *86* (Volume 86, 2017), 387–415.
(6) Kutateladze, D. A.; Jacobsen, E. N. Cooperative Hydrogen-Bond-Donor Catalysis with Hydrogen Chloride Enables Highly Enantioselective Prins Cyclization Reactions. *J. Am. Chem. Soc.* **2021**, *143* (48), 20077–20083.
(7) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.
(8) Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016. (9)Li, X.; Duan, M.; Yu, P.; Houk, K. N.; Sun, J. Organocatalytic Enantioselective Dearomatization of Thiophenes by 1,10-Conjugate Addition of Indole Imine Methides. *Nat Commun* **2021**, *12* (1), 4881.
(10) Saito, K.; Moriya, Y.; Akiyama, T. Chiral Phosphoric Acid Catalyzed Asymmetric Synthesis of 2-Substituted 2,3-Dihydro-4-Quinolones by a Protecting-Group-Free Approach. *Org. Lett.* **2015**, *17* (13), 3202–3205.
(11) Li, J.; Grosslight, S.; Miller, S. J.; Sigman, M. S.; Toste, F. D. Site-Selective Acylation of Natural Products with BINOL-Derived Phosphoric Acids. *ACS Catal.* **2019**, *9* (11), 9794–9799.
(12) Zhang, J.-L.; Ma, R.; Zhao, H.-H.; Xu, P.-F. Enantioselective Construction of Spiro-Tetrahydroquinoline Scaffolds through Asymmetric Catalytic Cascade Reactions. *Chem. Commun.* **2022**, *58* (21), 3493–3496.
(13) Wu, Y.-L.; Wei, J.; Meng, Y.-F.; Yu, T.-T. Synthesis, Antifungal Activity, and QSAR Studies of Benzbutyrolactone Derivatives Based on α-Methylene-γ-Butyrolactone Scaffold. *Russ J Gen Chem* **2022**, *92* (6), 1085–1097.
(14) Tayama, E.; Hirano, K. Base-Promoted Aromatic [3,3] Sigmatropic Rearrangement of *N*-Acyl-*O*-Arylhydroxylamine Derivatives. *Tetrahedron* **2019**, *75* (5), 665–673.

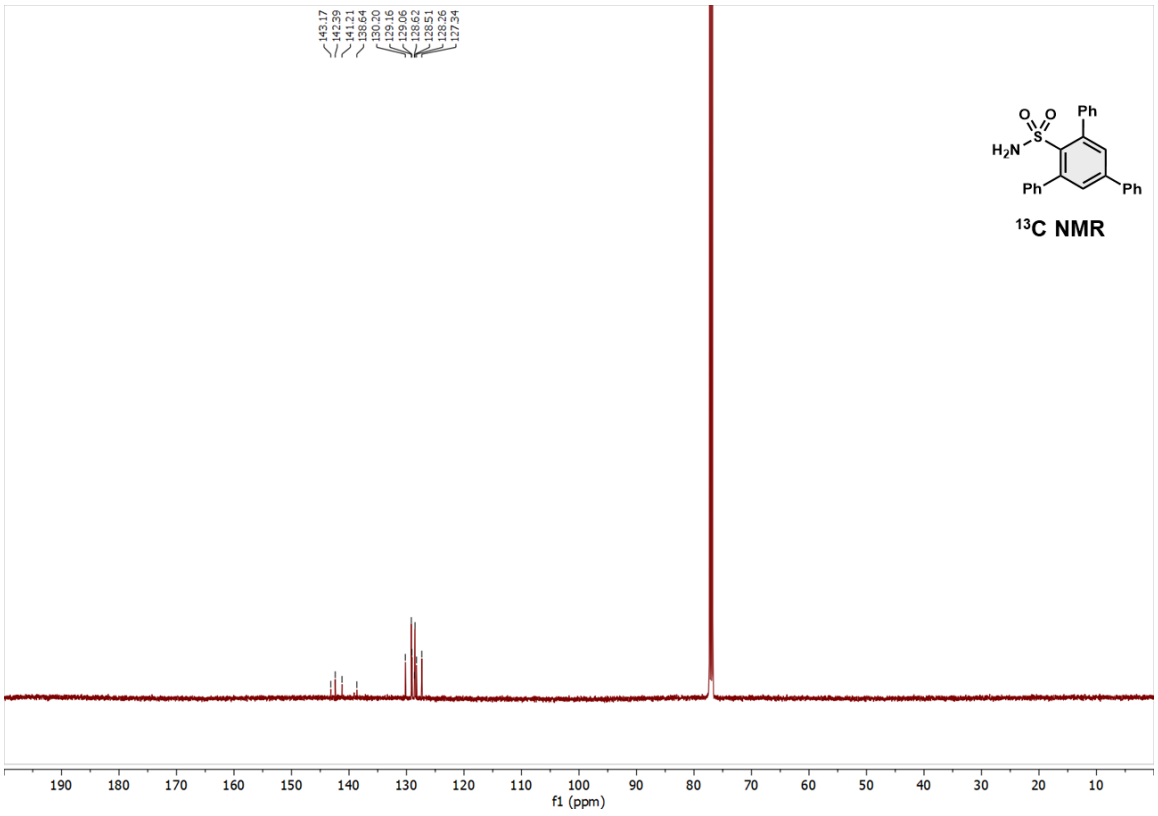
(15) Zhu, Y.; Zhang, E.; Luo, C.; Li, X.; Cheng, J.-P. Asymmetric Addition of 3-Substituted Benzofuran-2-Ones to Isatin *N*-Boc Ketimines Catalyzed by Chiral Biscinchona Alkaloid Catalyst. *Tetrahedron* **2015**, *71* (24), 4090–4097.
(16) Ohmatsu, K.; Hara, Y.; Ooi, T. In Situ Generation of Ion-Paired Chiral Ligands: Rapid Identification of the Optimal Ligand for Palladium-Catalyzed Asymmetric Allylation. *Chem. Sci.* **2014**, *5* (9), 3645–3650.
(17) Cai, H.; Liu, Q.; Gao, D.; Wang, T.; Chen, T.; Yan, G.; Chen, K.; Xu, Y.; Wang, H.; Li, Y.; Zhu, W. Novel Fatty Acid Binding Protein 4 (FABP4) Inhibitors: Virtual Screening, Synthesis and Crystal Structure Determination. *European Journal of Medicinal Chemistry* **2015**, *90*, 241–250.
(18) Taskinen, E. K.; Birnthaler, D.; Kermelj, V.; König, B. Preassembly-Controlled Radical Recombination at Bismuth: Decarboxylative C─N Coupling with Sulfonamides. *Chemistry – A European Journal* **2025**, *31* (28), e202500396.
(19) Ishihara, K.; Obayashi, R.; Gotoh, M.; Watanabe, Y.; Kobayashi, Y.; Ishihara, K.; Shioiri, T.; Matsugi, M. A Recyclable and Highly Stereoselective Multi-Fluorous Proline Catalyst for Asymmetric Aldol Reactions. *Tetrahedron Letters* **2020**, *61* (13), 151657.
(20) Davies, T. Q.; Tilby, M. J.; Skolc, D.; Hall, A.; Willis, M. C. Primary Sulfonamide Synthesis Using the Sulfinylamine Reagent N-Sulfinyl-O-(Tert-Butyl)Hydroxylamine, t-BuONSO. *Org. Lett.* **2020**, *22* (24), 9495–9499.
(21) Liu, W.; Zhao, Y.; Yi, F.; Chen, J. Palladium(II)/Copper(I)-Catalyzed Sequential CH Arylation and Oxidative CN Bond Cleavage of Aryl Sulfonamino Acids: Efficient One-Pot Synthesis of Primary Biaryl Sulfonamides. *Tetrahedron* **2016**, *72* (51), 8382–8386.

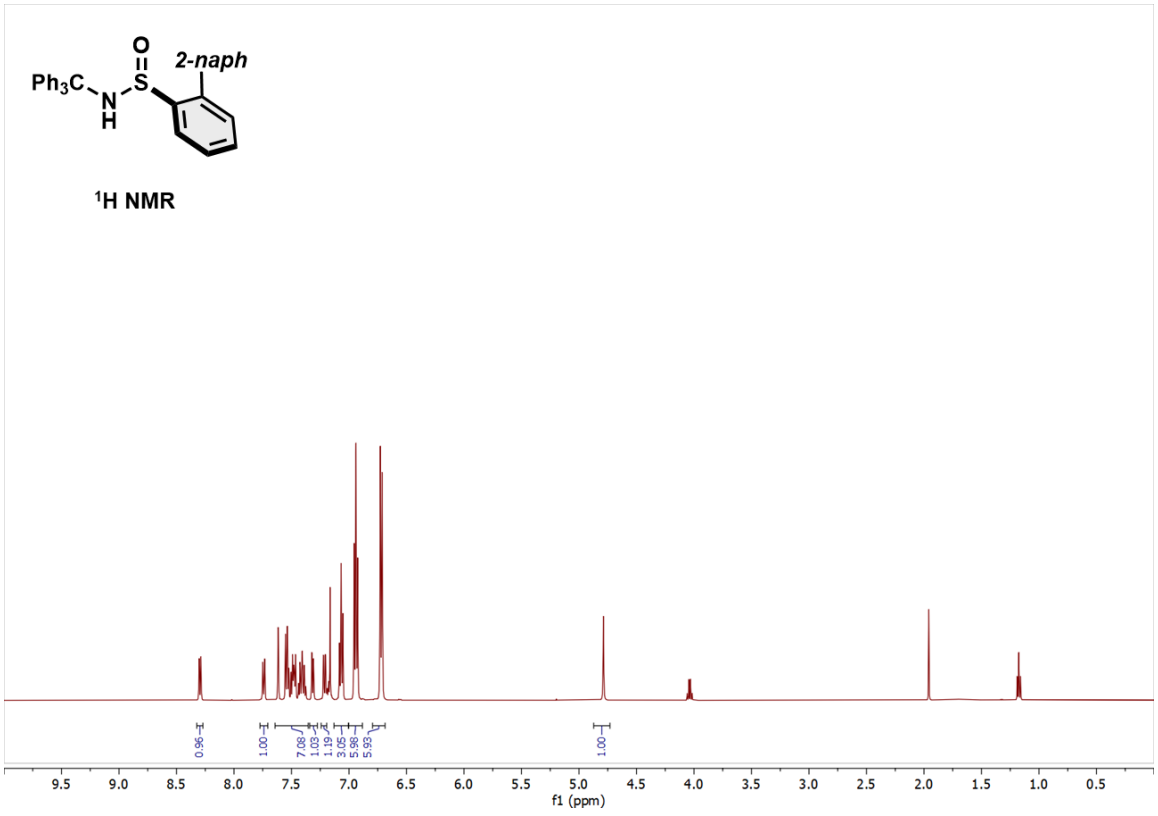
VI.48 – NMR Spectra

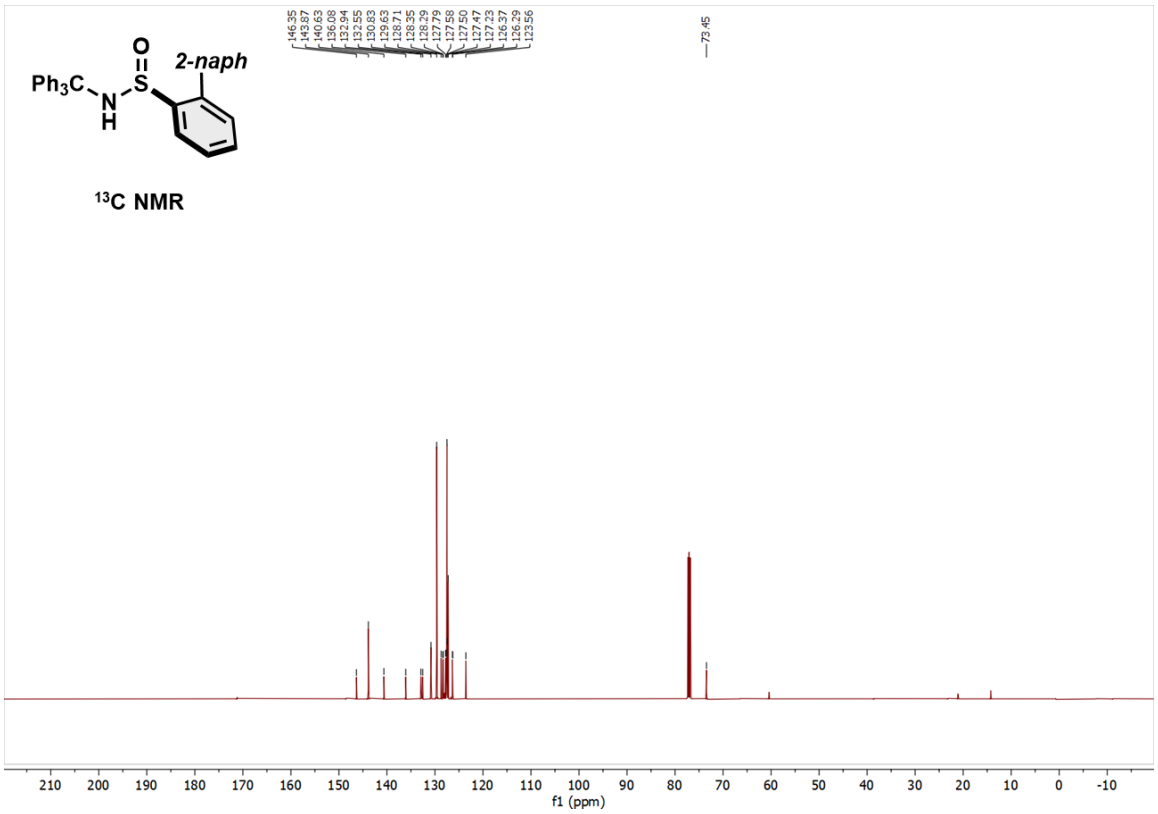


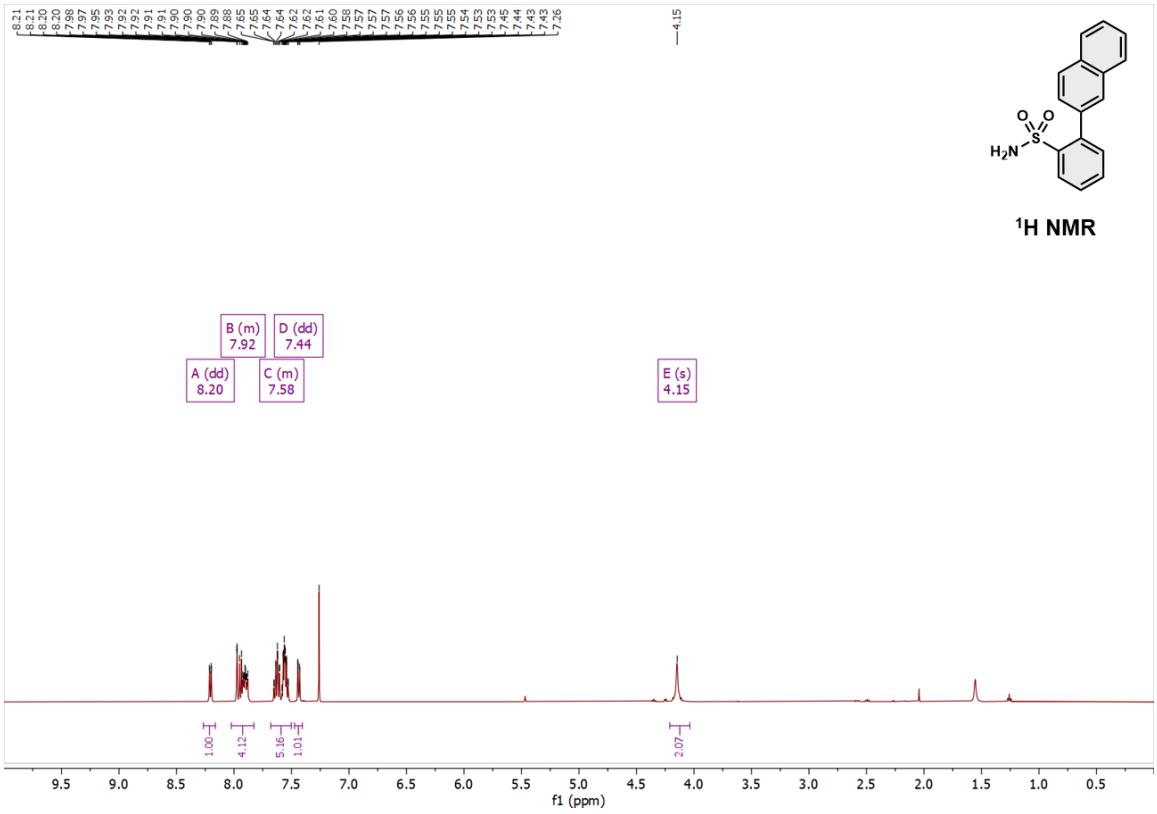


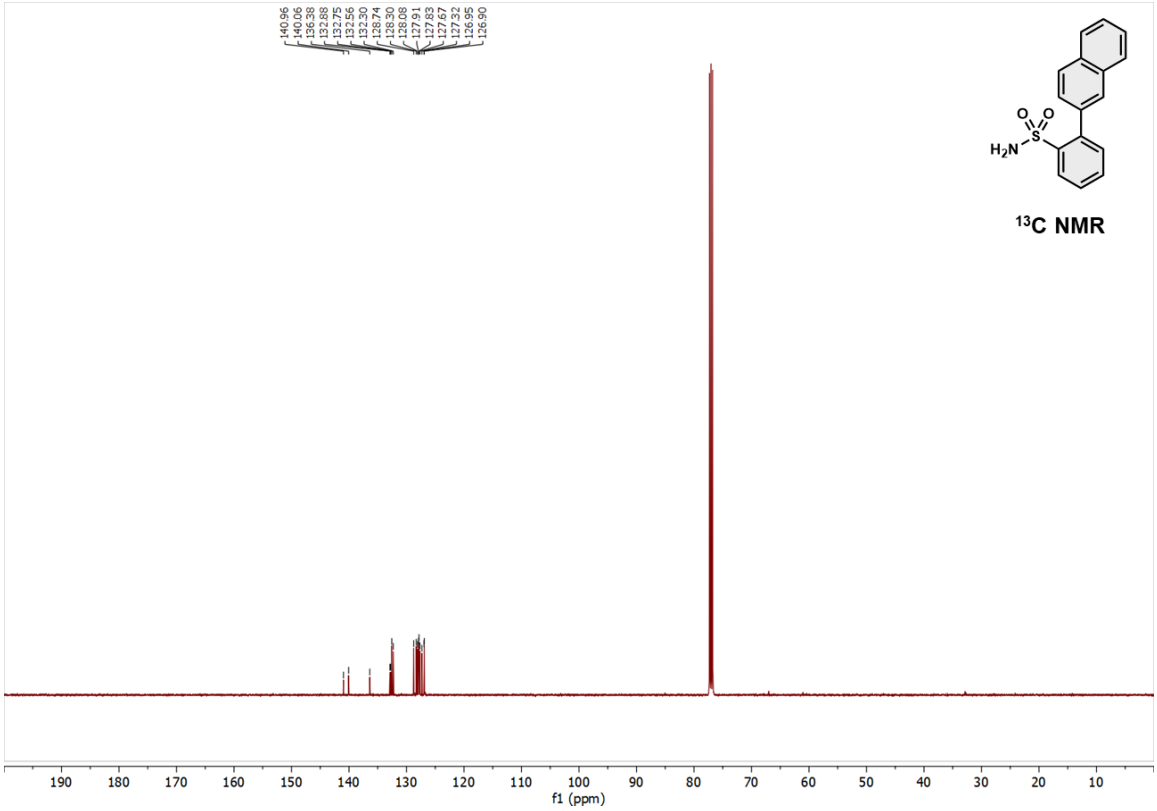

7.59
7.59
7.58
7.57
7.57
7.57
7.56
7.55
7.55
7.54
7.54
7.54
7.53
7.52
7.49
7.45
7.45
7.44
7.43
7.43
7.42
7.42
7.40
7.40
7.40
7.39
7.39
7.38
7.37
7.37
7.35
7.35
7.35
7.34
7.34
7.33
7.32
7.32
7.31
7.19 CDCl3
3.89
B (m)
7.34
A (m)
7.53
C (m)
7.40
8.29
6.75
1.00
f1 (ppm)
Ph
Ph
Ph
H2N
O
O
S
1H NMR

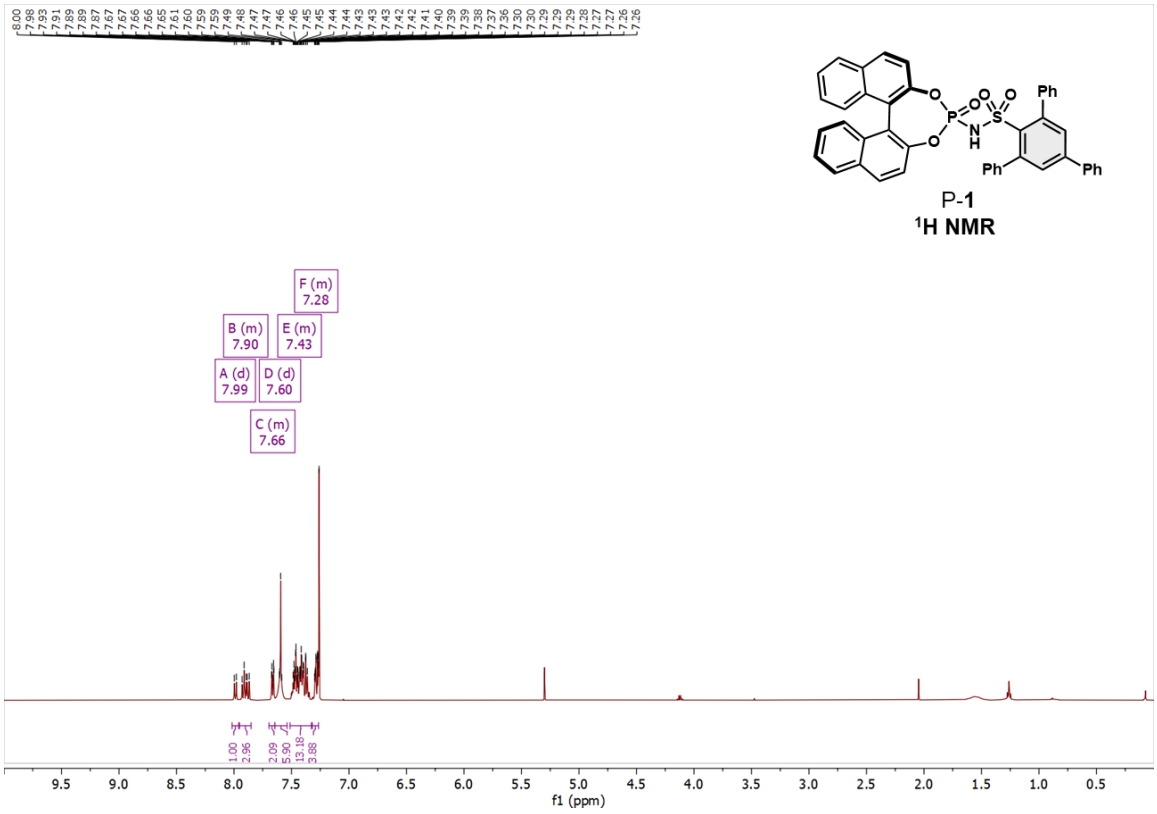

143.17
142.39
141.21
138.64
130.20
129.16
129.06
128.62
128.51
128.26
127.34
f1 (ppm)
Ph
Ph
Ph
H2N
O
O
S
13C NMR


Ph3C
N
H
O
S
2-naph
1H NMR
0.96
1.00
7.08
1.03
1.19
3.05
5.98
5.93
1.00
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)


Ph3C
N
H
O
S
2-naph
13C NMR
146.35
143.87
140.63
136.08
132.94
132.55
130.83
129.63
128.71
128.35
128.29
127.79
127.58
127.50
127.47
127.23
126.37
126.29
123.56
73.45
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

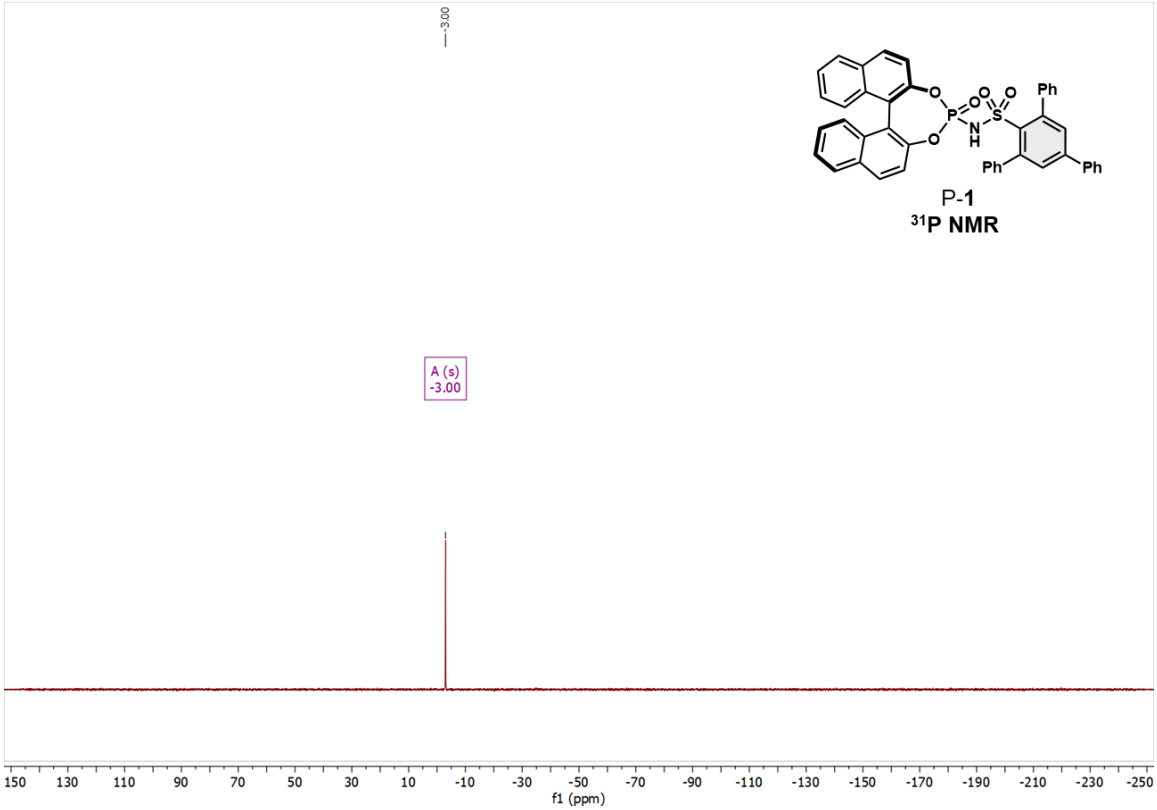

8.21
8.21
8.20
8.20
8.20
7.98
7.97
7.95
7.93
7.92
7.92
7.91
7.91
7.90
7.90
7.90
7.90
7.89
7.88
7.85
7.85
7.84
7.64
7.62
7.62
7.61
7.60
7.58
7.57
7.57
7.57
7.56
7.56
7.55
7.55
7.55
7.54
7.53
7.53
7.45
7.44
7.43
7.43
7.26
4.15
B (m)
7.92
D (dd)
7.44
A (dd)
8.20
C (m)
7.58
E (s)
4.15
1.00
4.12
5.16
1.01
2.07
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)
^{1}H NMR


140.96
140.06
135.58
133.88
133.25
132.56
132.30
128.74
128.30
128.08
127.91
127.83
127.67
127.32
126.95
126.90
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)
^{13}C NMR



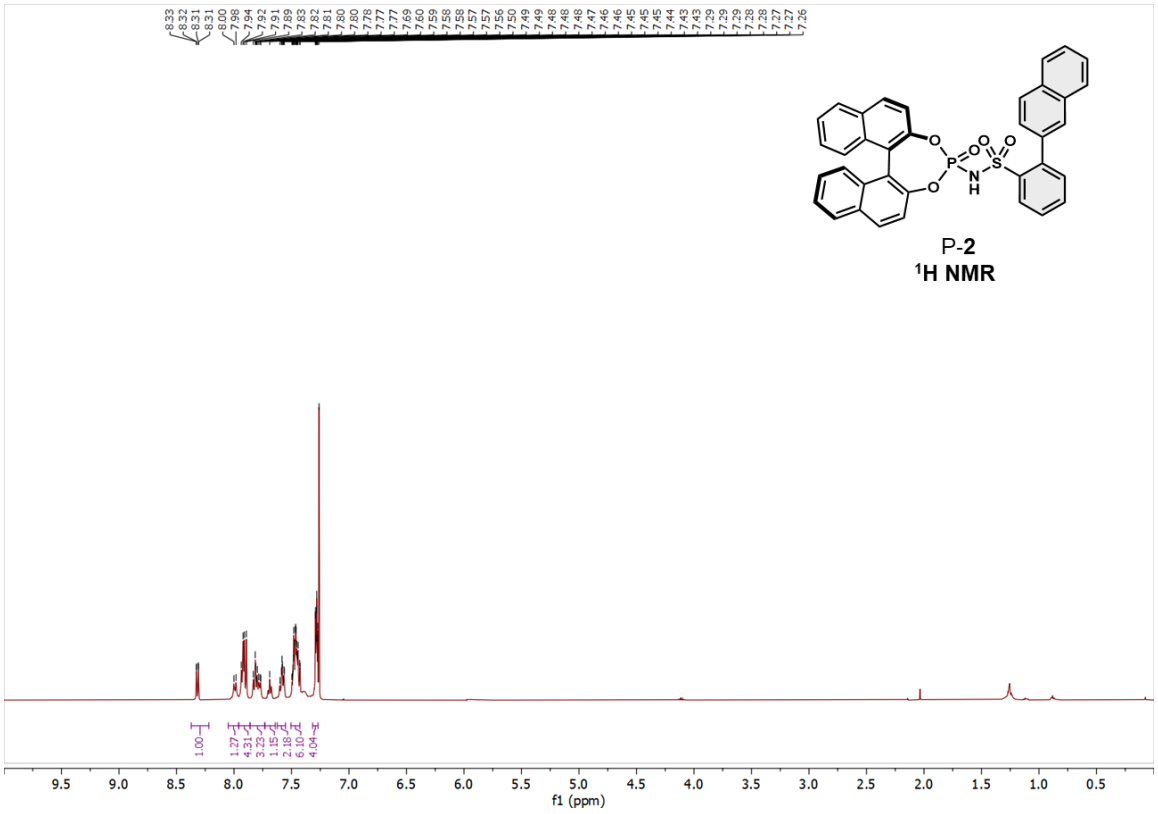
P-**1**

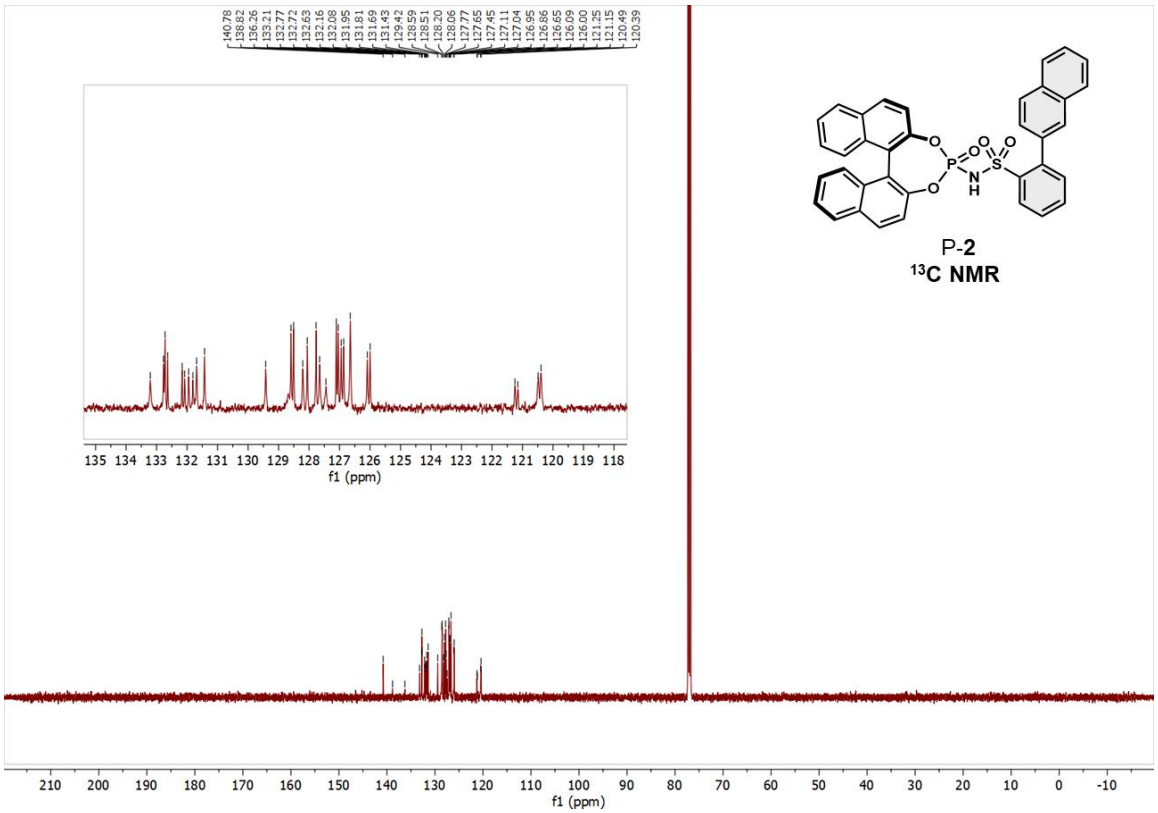
^{1}H NMR

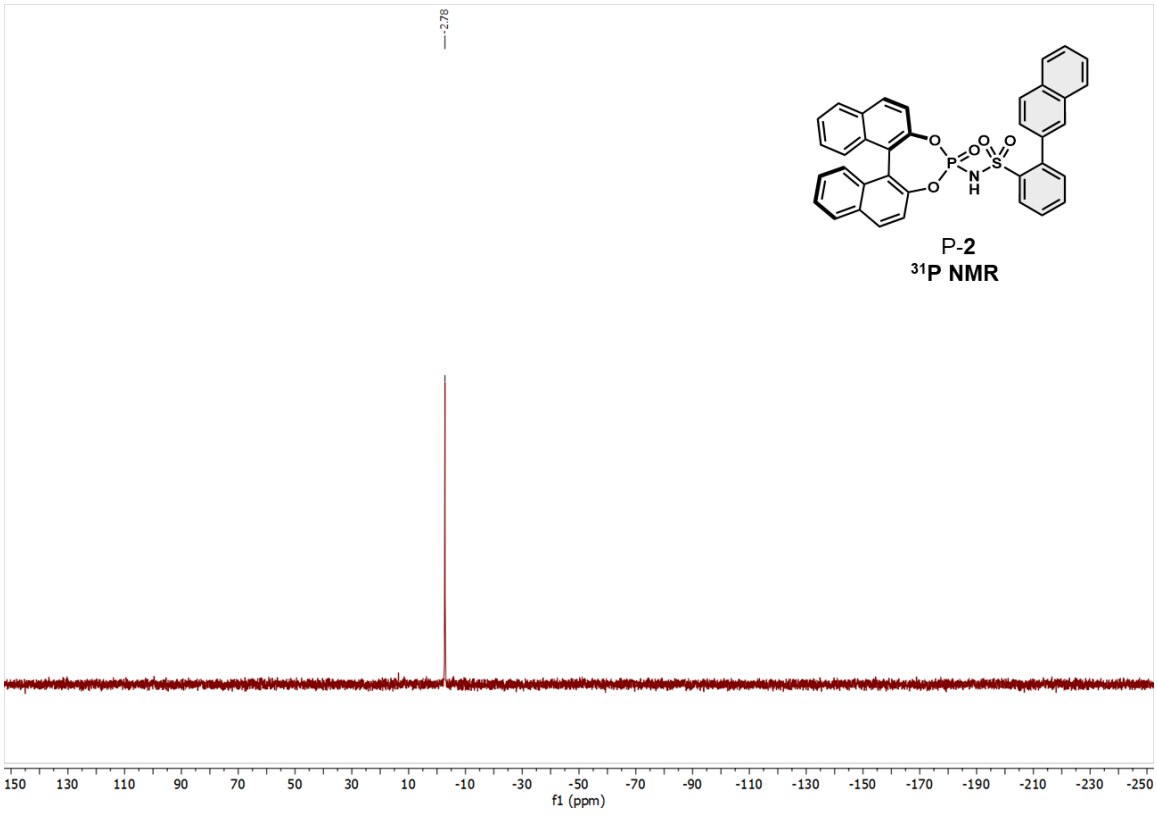


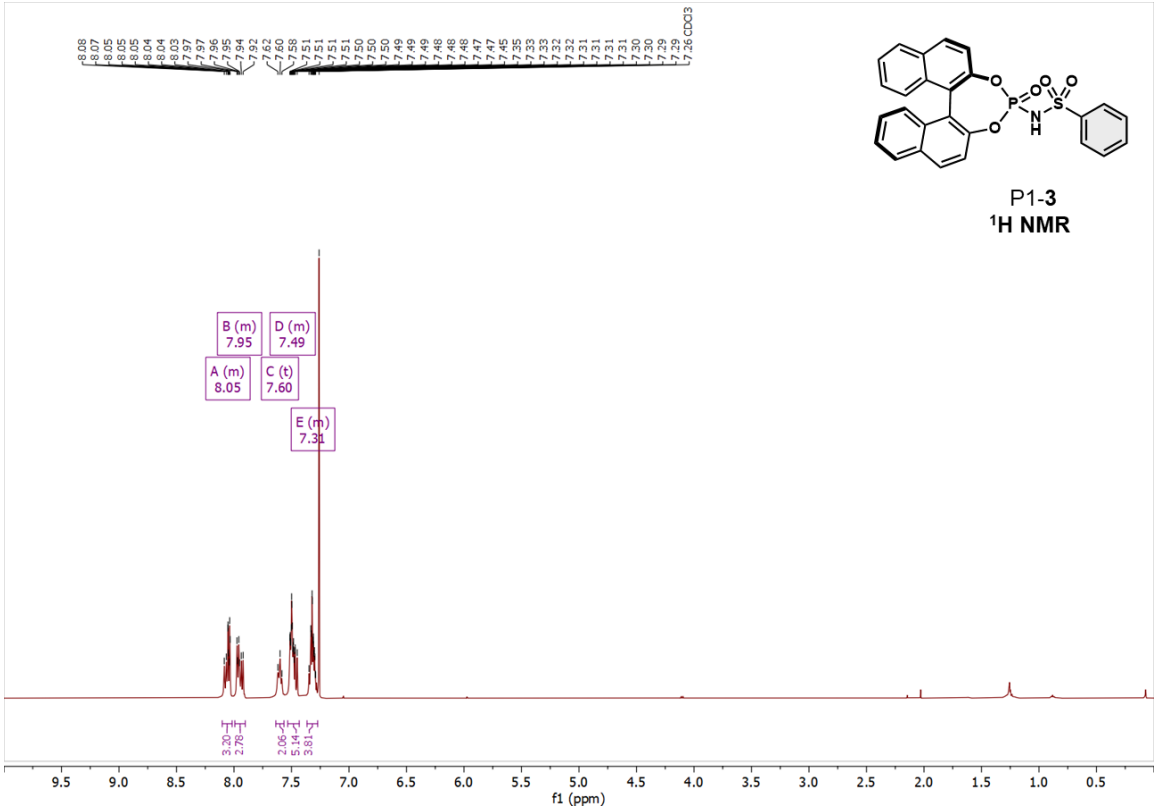
P-**1**

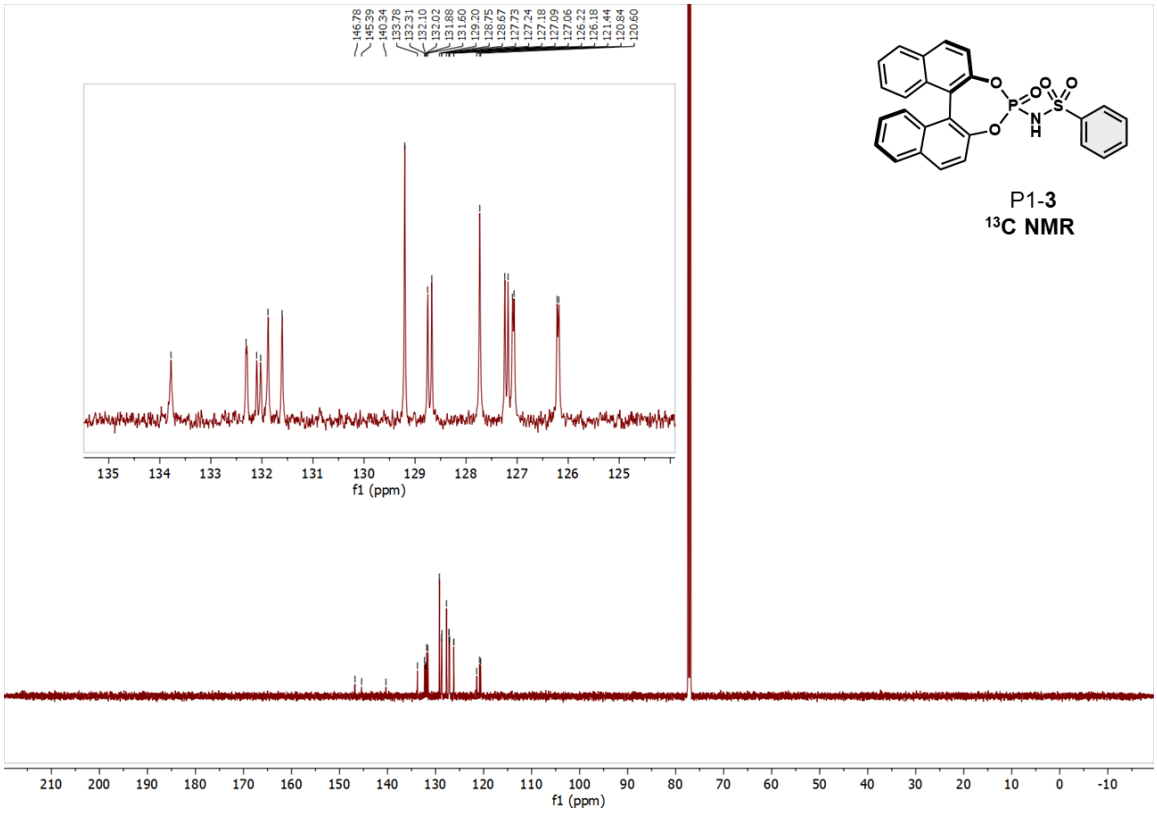
^{31}P NMR


P-2
1H NMR
f1 (ppm)

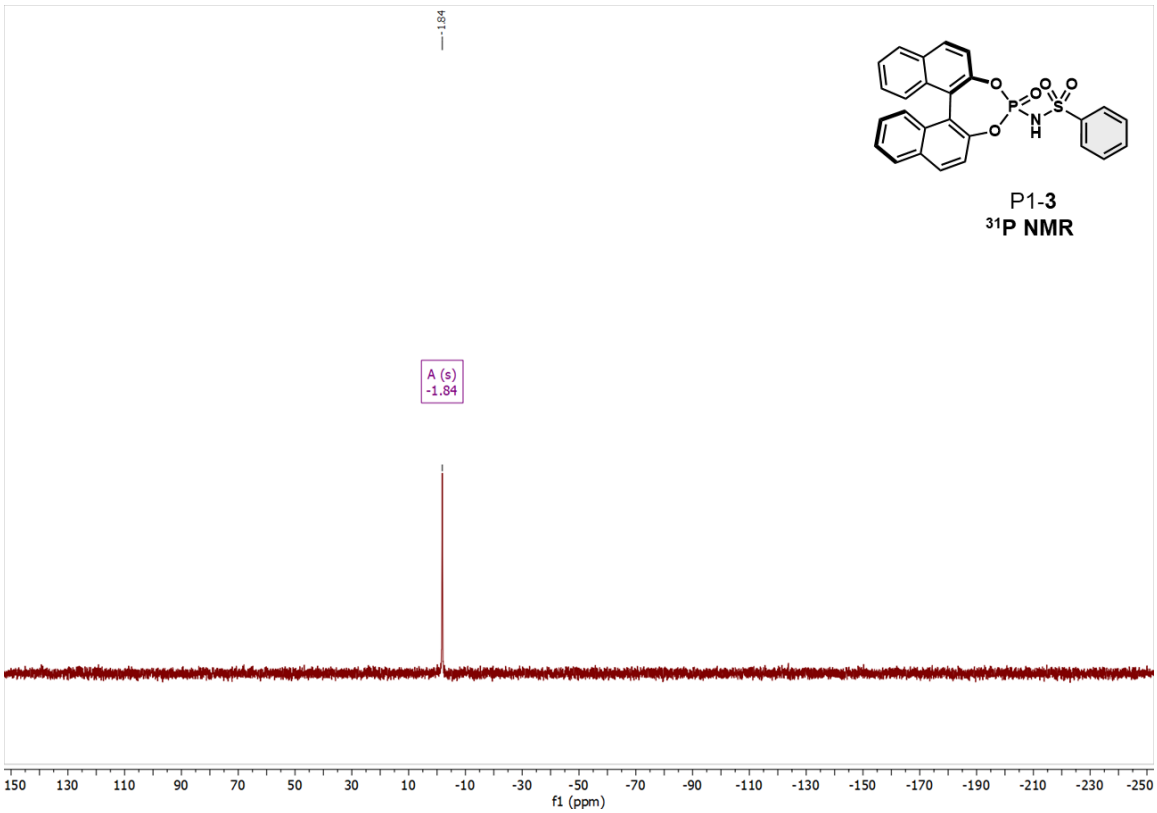

P-2
13C NMR
f1 (ppm)


-2.76
P-2
31P NMR
150 130 110 90 70 50 30 10 -10 -30 -50 -70 -90 -110 -130 -150 -170 -190 -210 -230 -250
f1 (ppm)

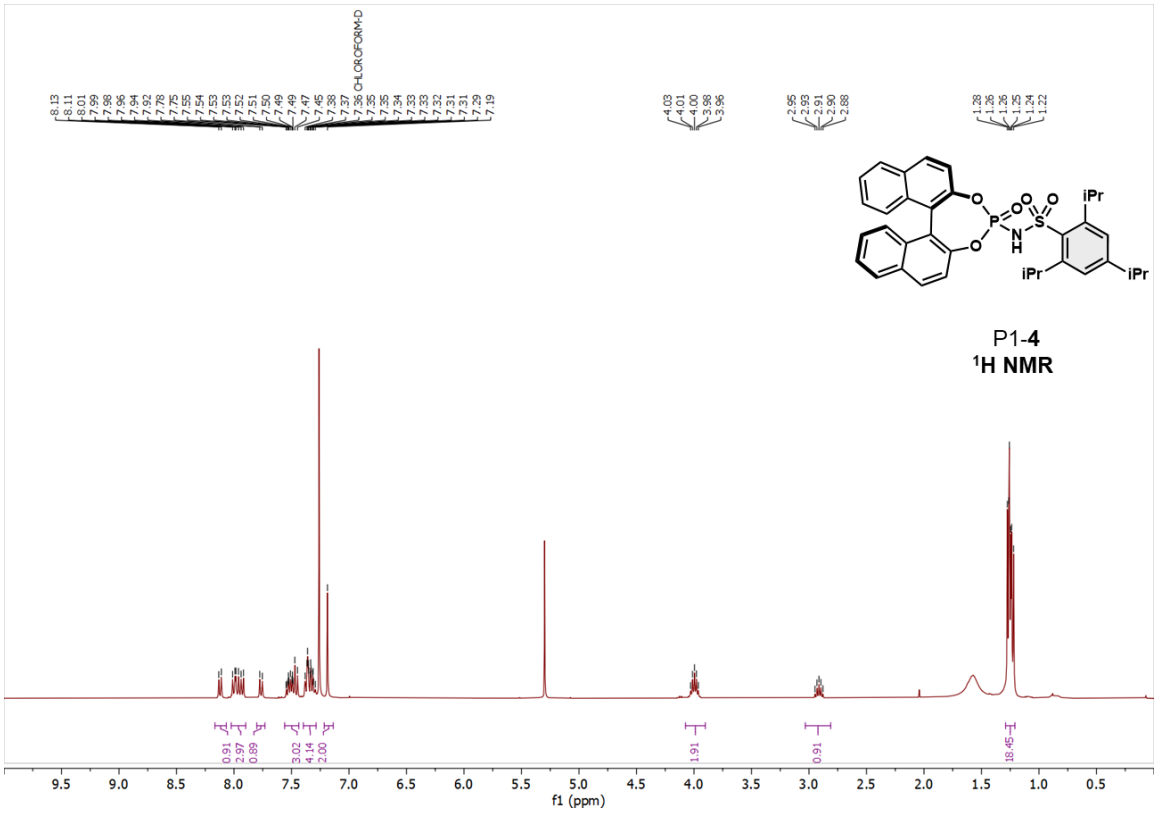

8.08 8.07 8.05 8.05 8.05 8.04 8.04 8.03 7.97 7.97 7.96 7.95 7.94 7.92 7.62 7.60 7.58 7.51 7.51 7.51 7.51 7.50 7.50 7.50 7.49 7.49 7.49 7.48 7.48 7.48 7.47 7.47 7.45 7.35 7.33 7.33 7.32 7.32 7.31 7.31 7.31 7.31 7.30 7.30 7.29 7.29 7.26 CDCl3
P1-3
1H NMR
B (m) 7.95
D (m) 7.49
A (m) 8.05
C (t) 7.60
E (m) 7.31
3.20
2.78
2.06
1.41
3.81
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

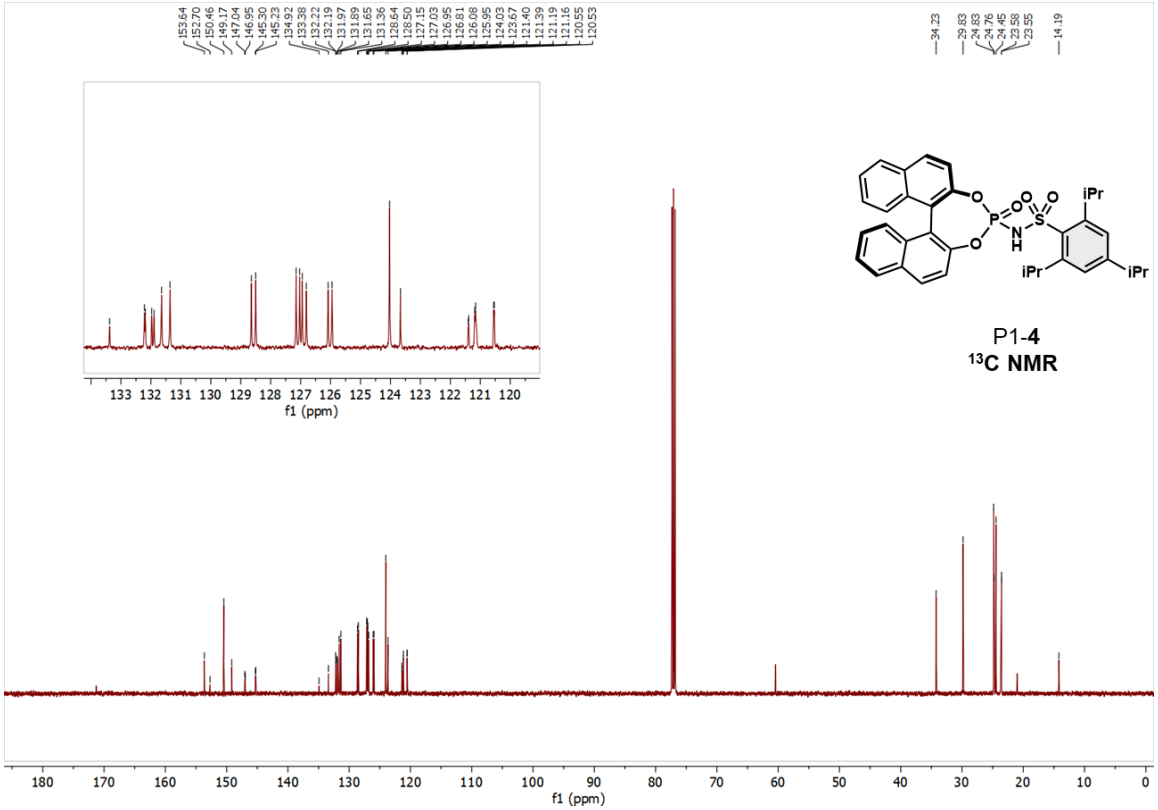


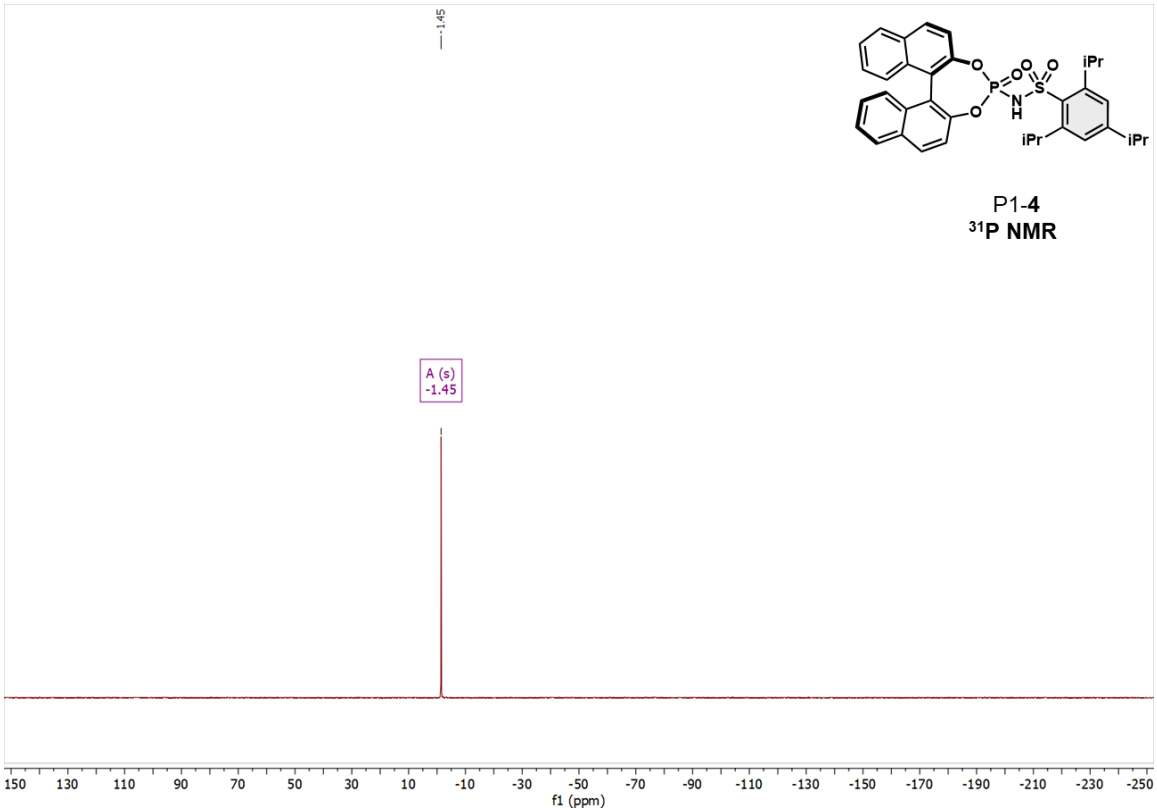
P1-**3**
^{13}C NMR



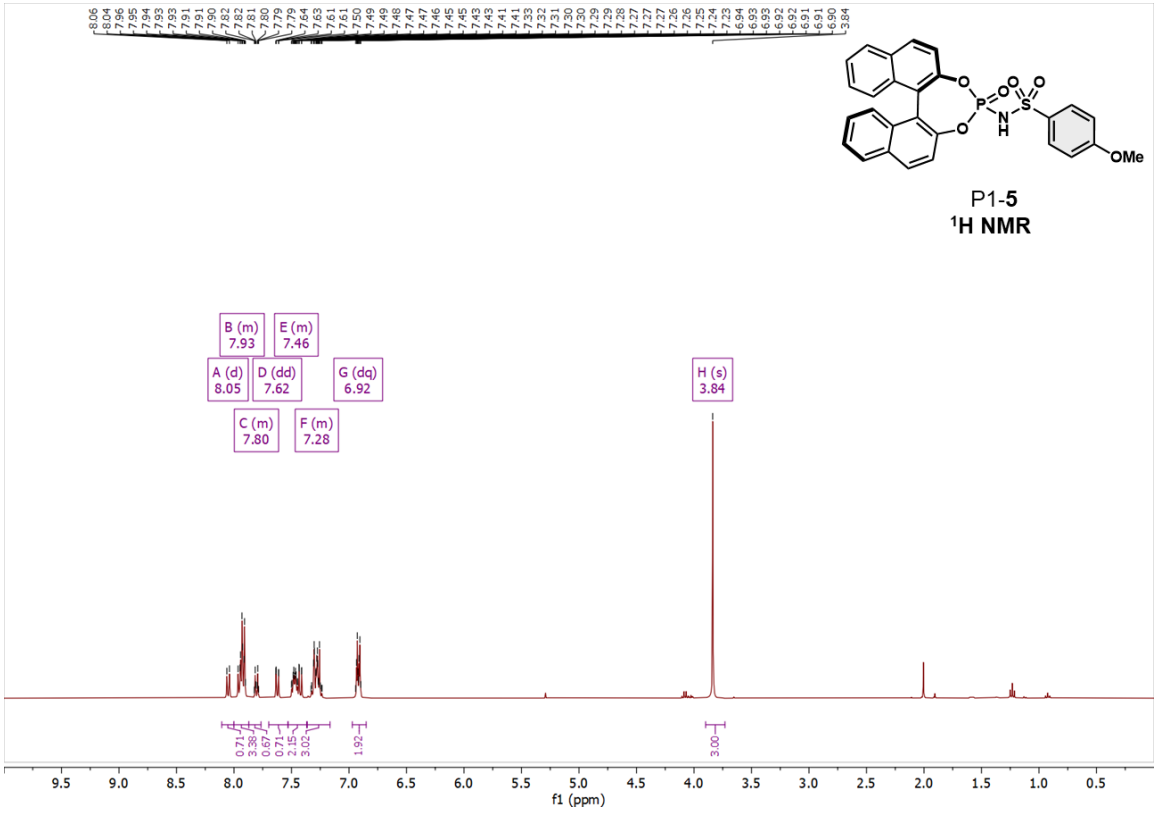
P1-**3**
^{31}P NMR


CHLOROFORM-D
iPr
iPr
iPr
P1-4
1H NMR
f1 (ppm)

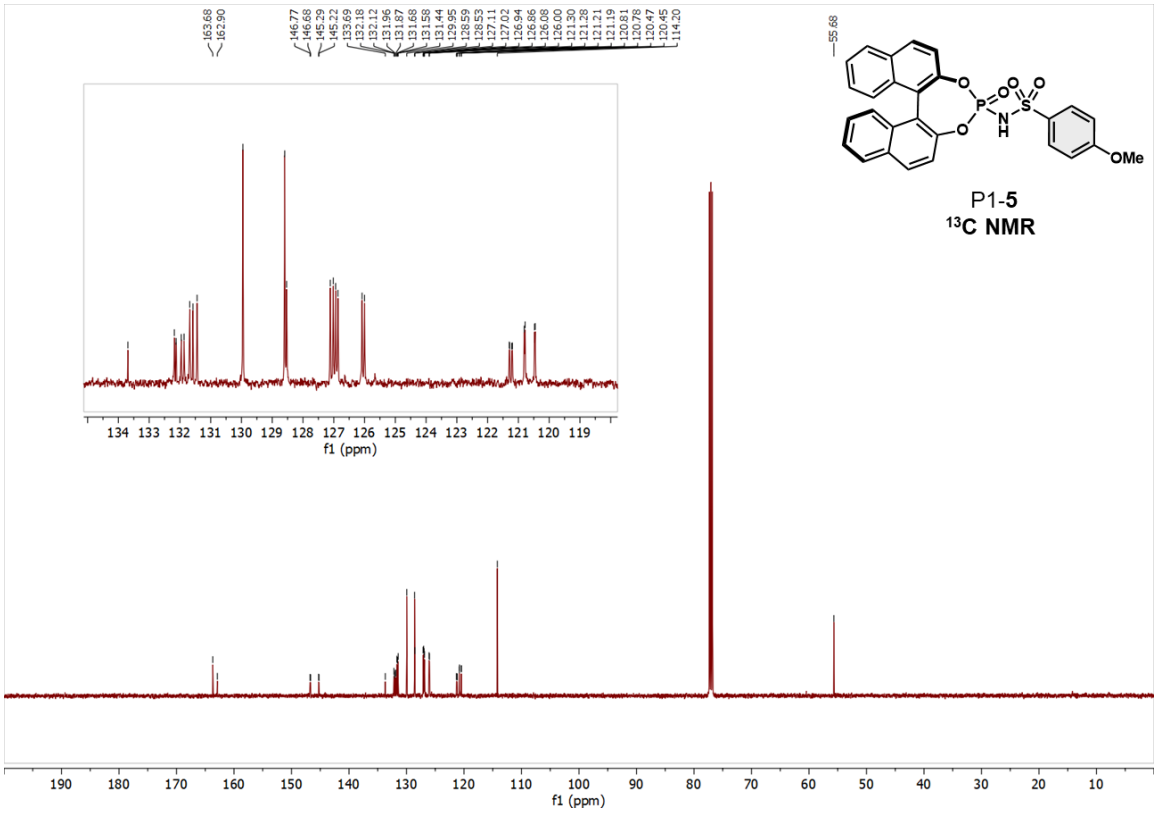

iPr
iPr
iPr
P1-4
13C NMR
f1 (ppm)

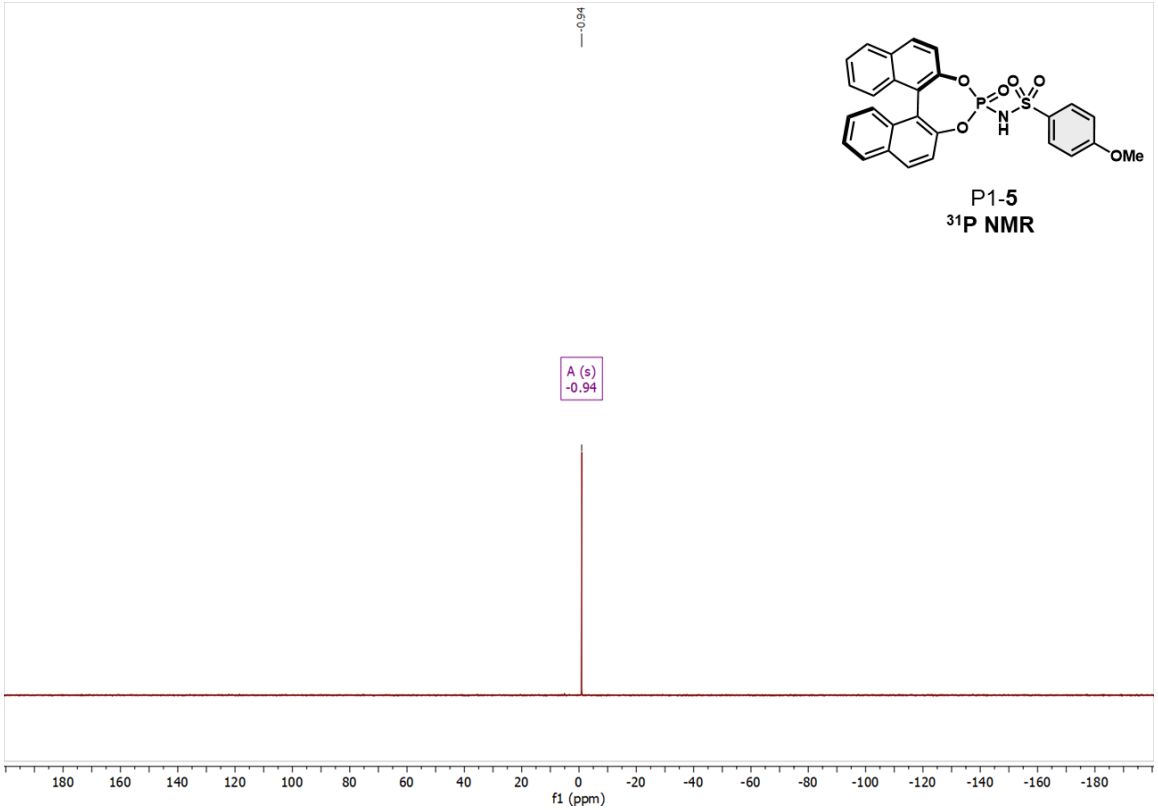


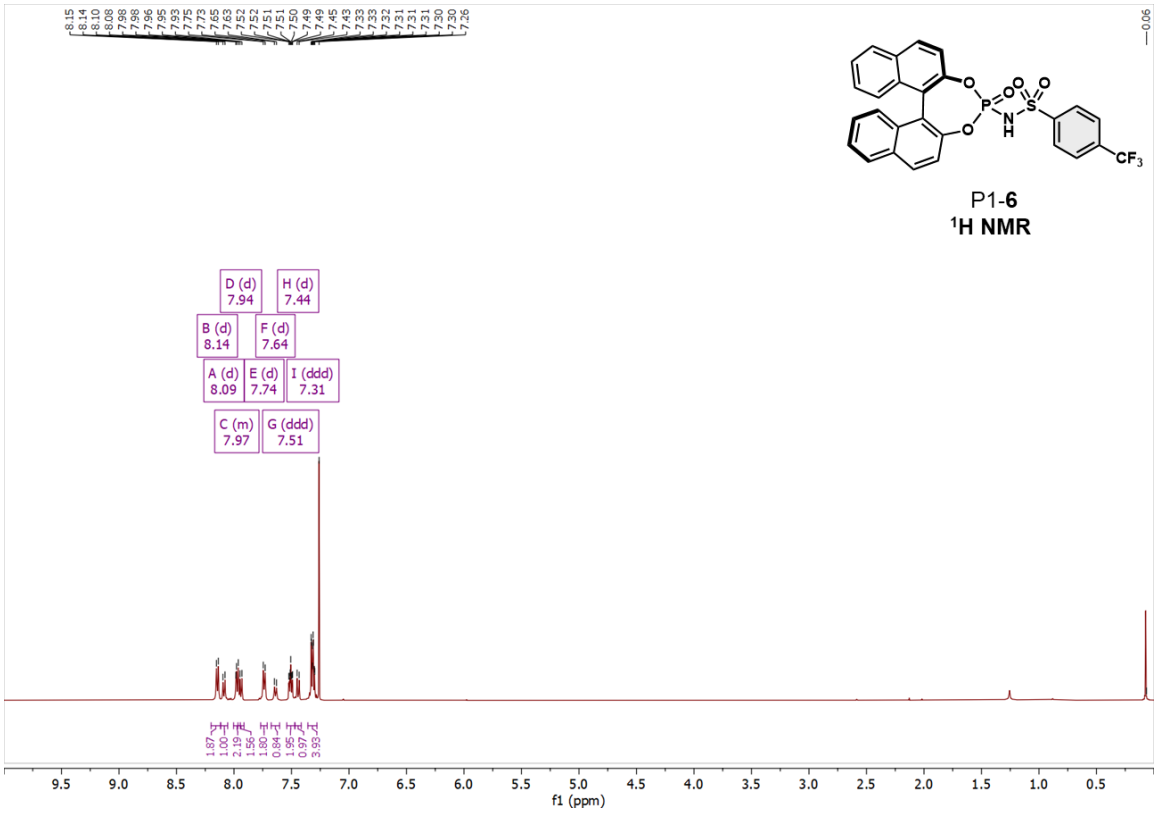
P1-**4**
^{31}P NMR

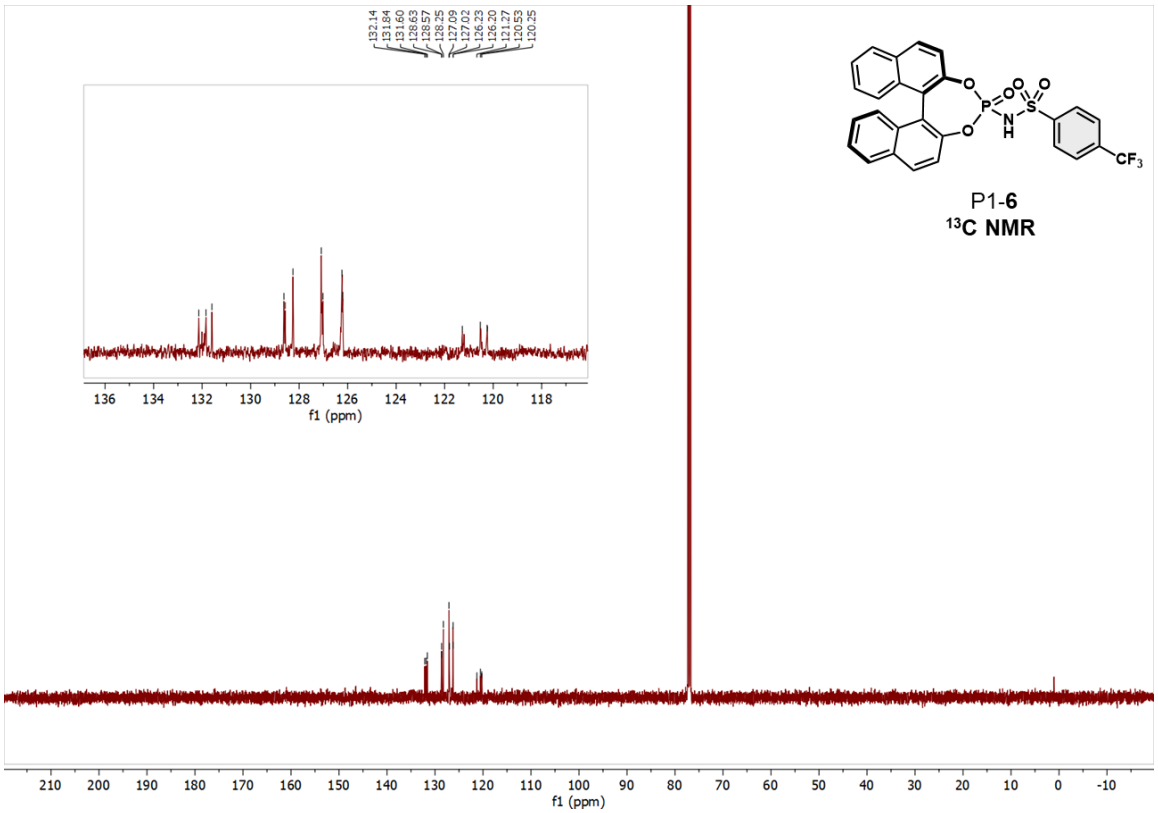


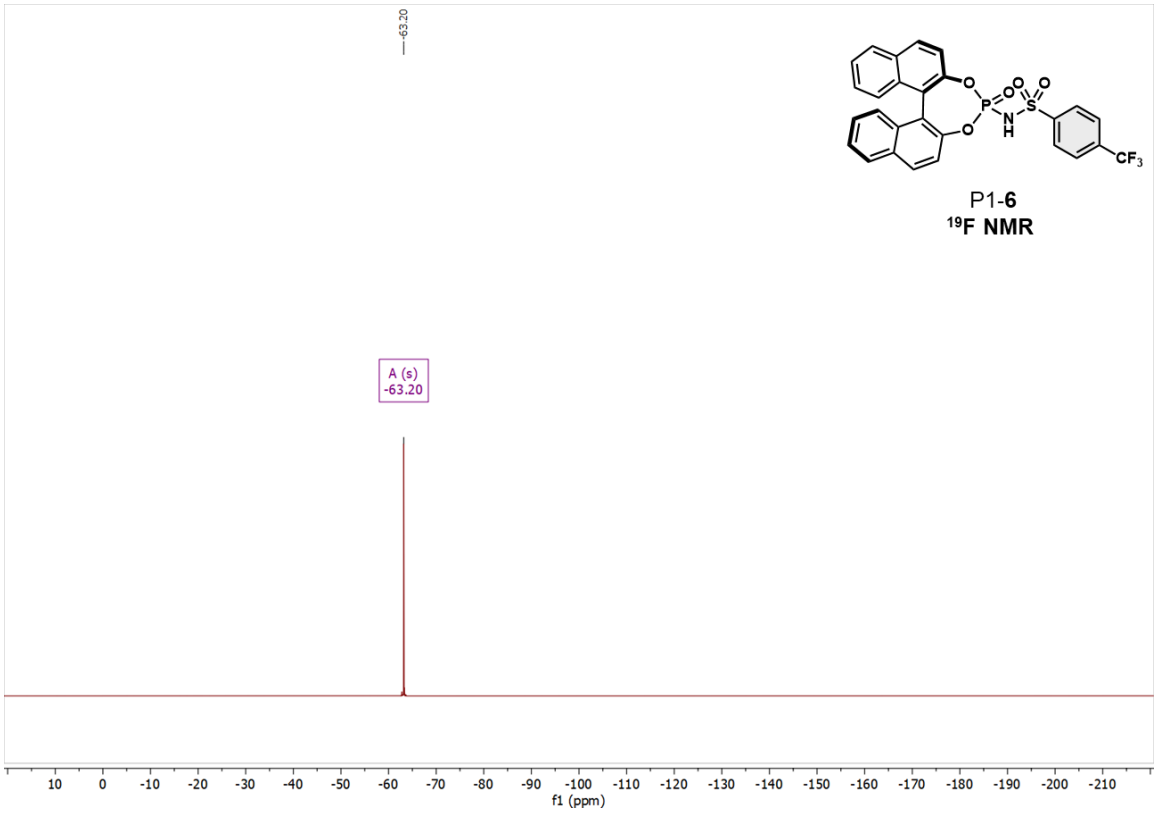
P1-**5**
^{1}H NMR


163.68
162.90
146.77
146.68
145.29
145.22
133.69
132.18
132.12
131.96
131.87
131.68
131.58
131.44
129.95
128.59
128.53
127.11
127.02
126.94
126.86
126.08
126.00
121.30
121.28
121.21
121.19
120.81
120.78
120.47
120.45
114.20
55.68
OMe
P1-5
13C NMR
f1 (ppm)
f1 (ppm)

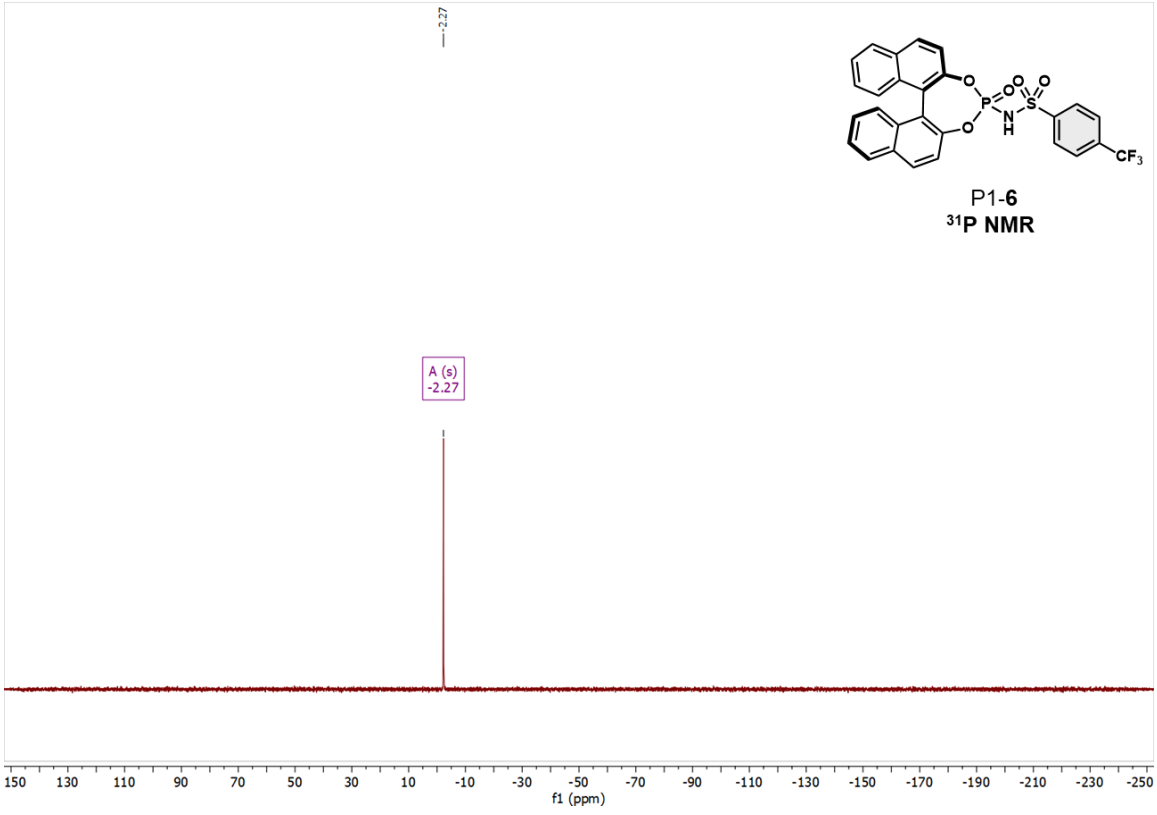

-0.94
OMe
P1-5
31P NMR
A (s)
-0.94
f1 (ppm)


P1-6
1H NMR
D (d) 7.94
H (d) 7.44
B (d) 8.14
F (d) 7.64
A (d) 8.09
E (d) 7.74
I (ddd) 7.31
C (m) 7.97
G (ddd) 7.51
f1 (ppm)

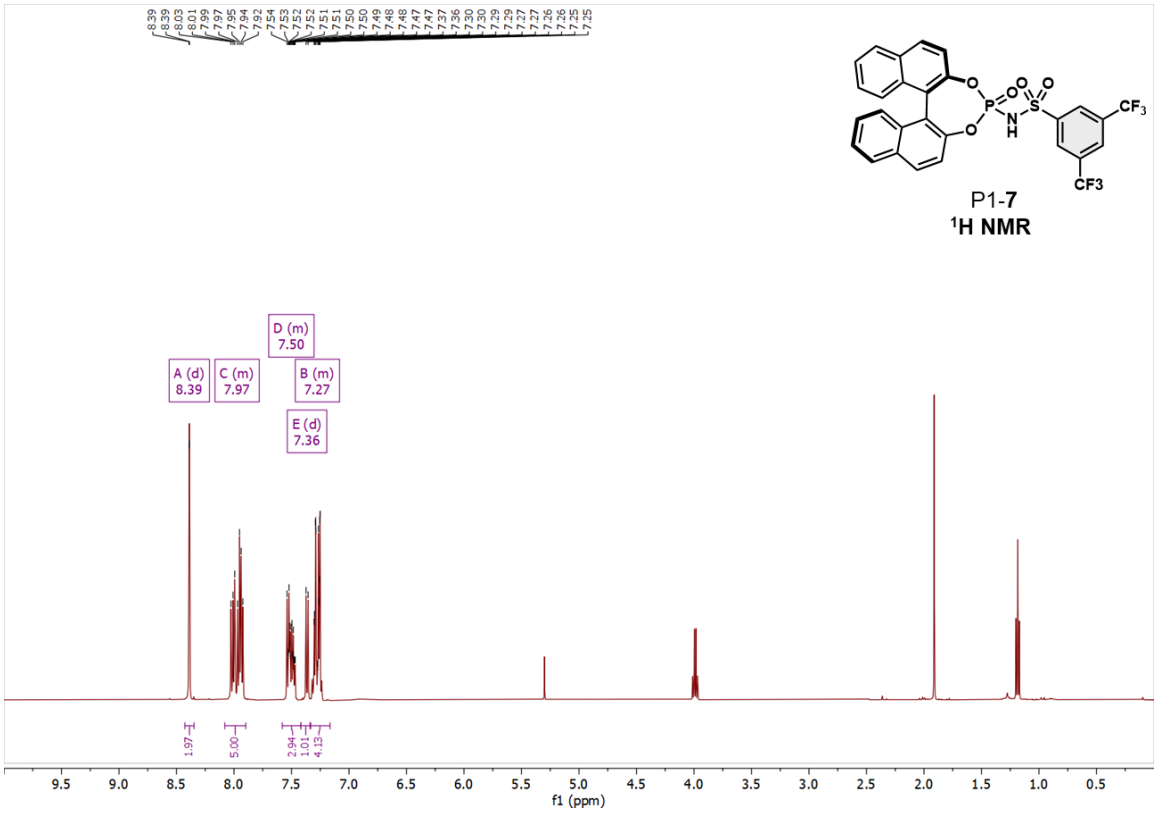

P1-6
13C NMR
f1 (ppm)

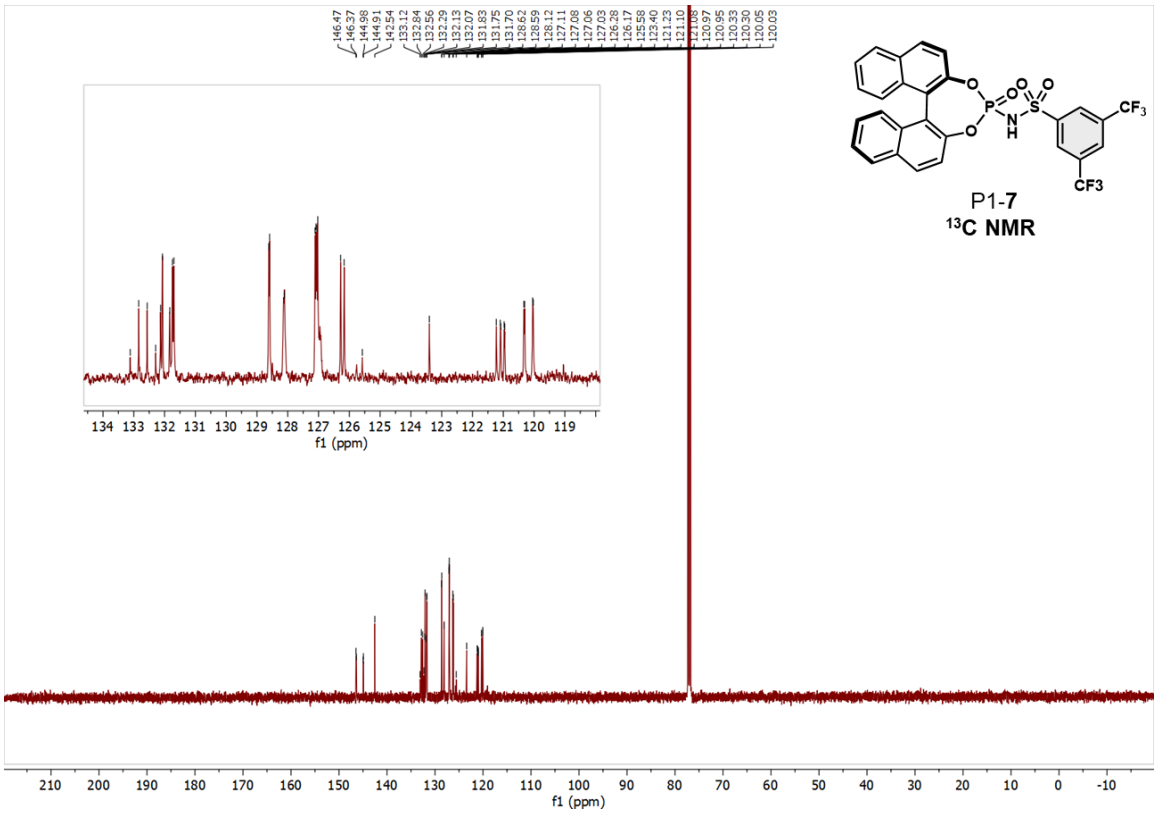


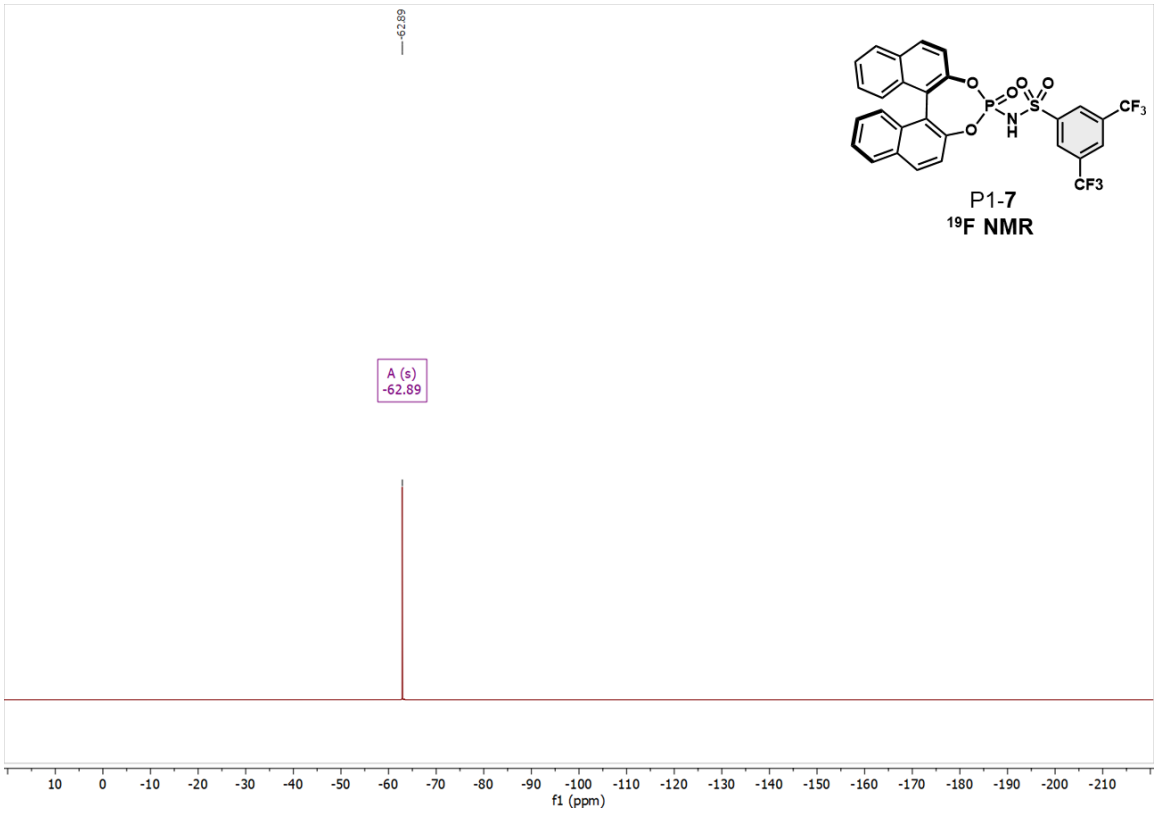
P1-**6**
^{19}F NMR



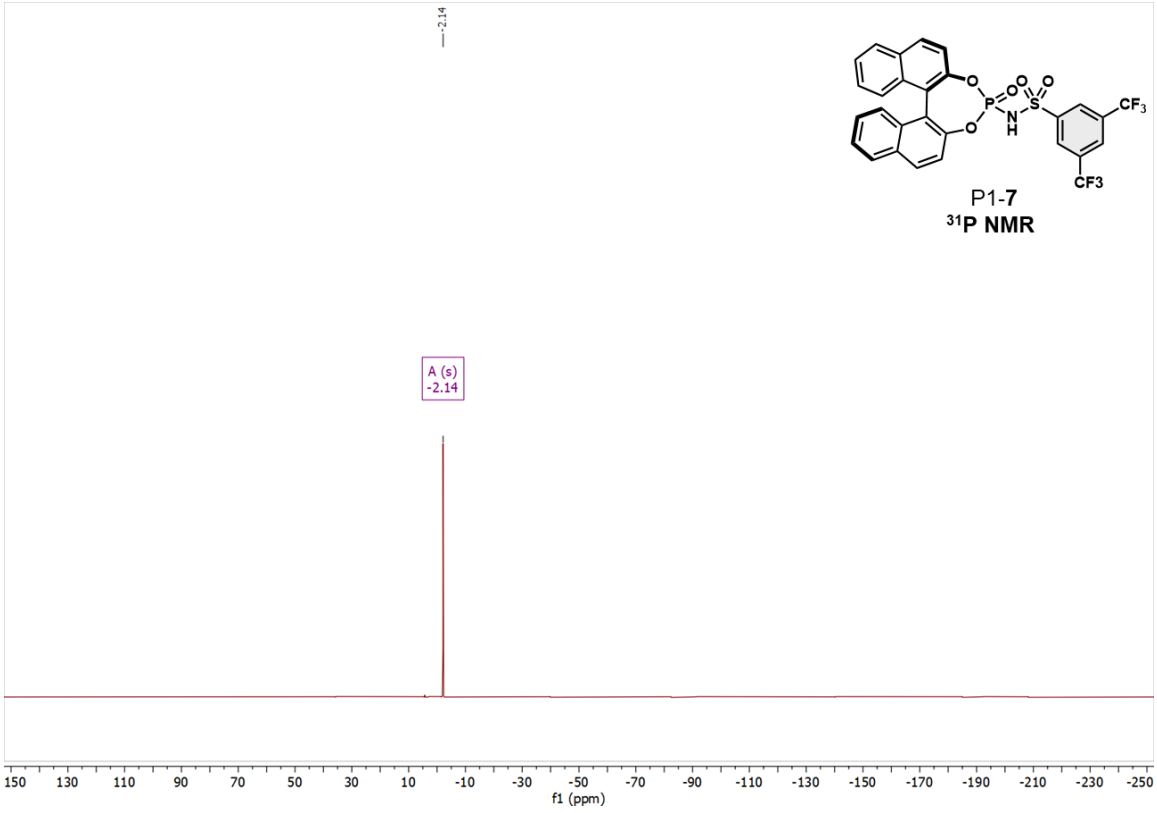
P1-**6**
^{31}P NMR


P1-7
1H NMR
A (d) 8.39
C (m) 7.97
D (m) 7.50
B (m) 7.27
E (d) 7.36
f1 (ppm)

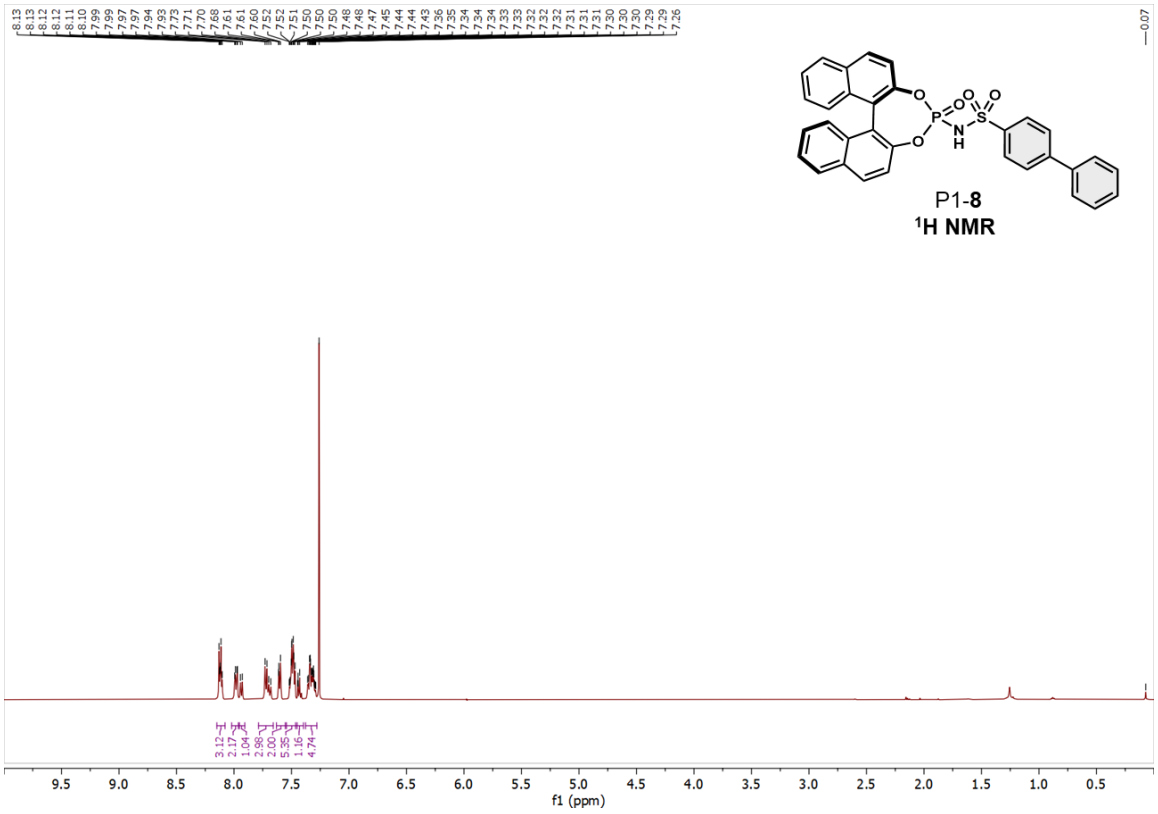

P1-7
13C NMR
f1 (ppm)

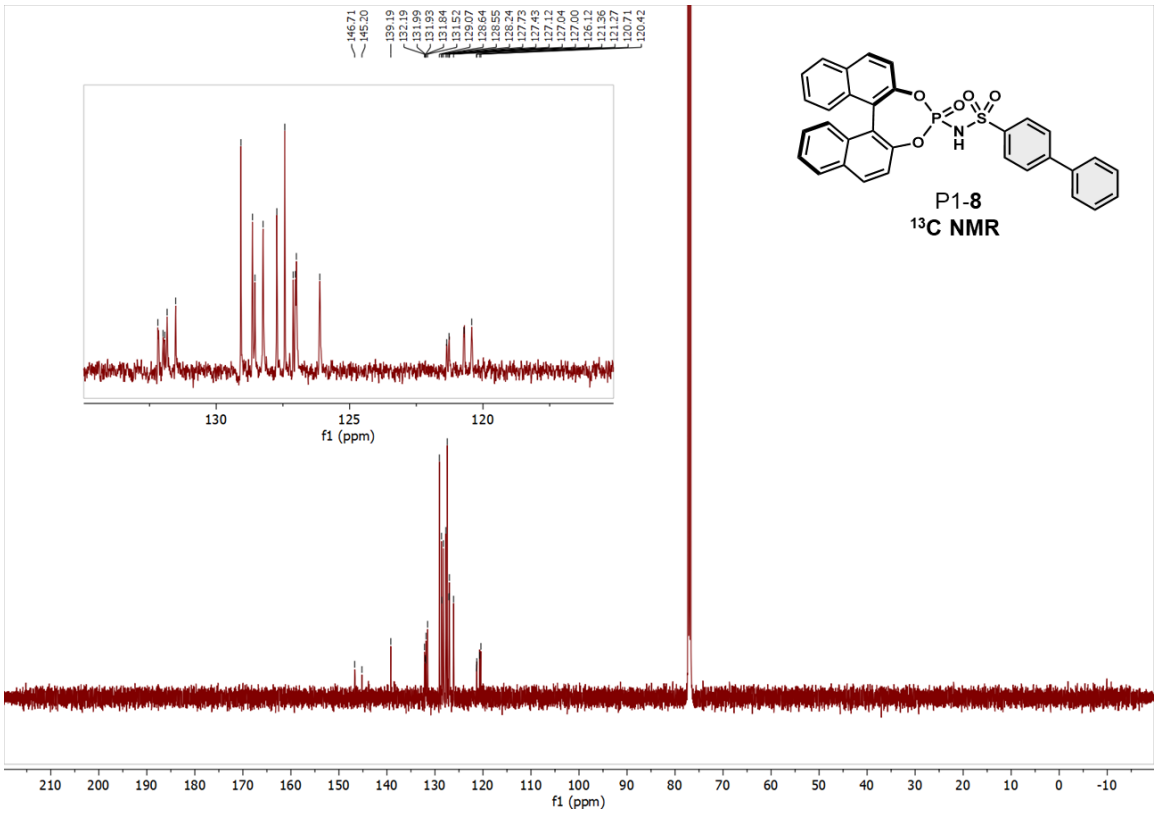


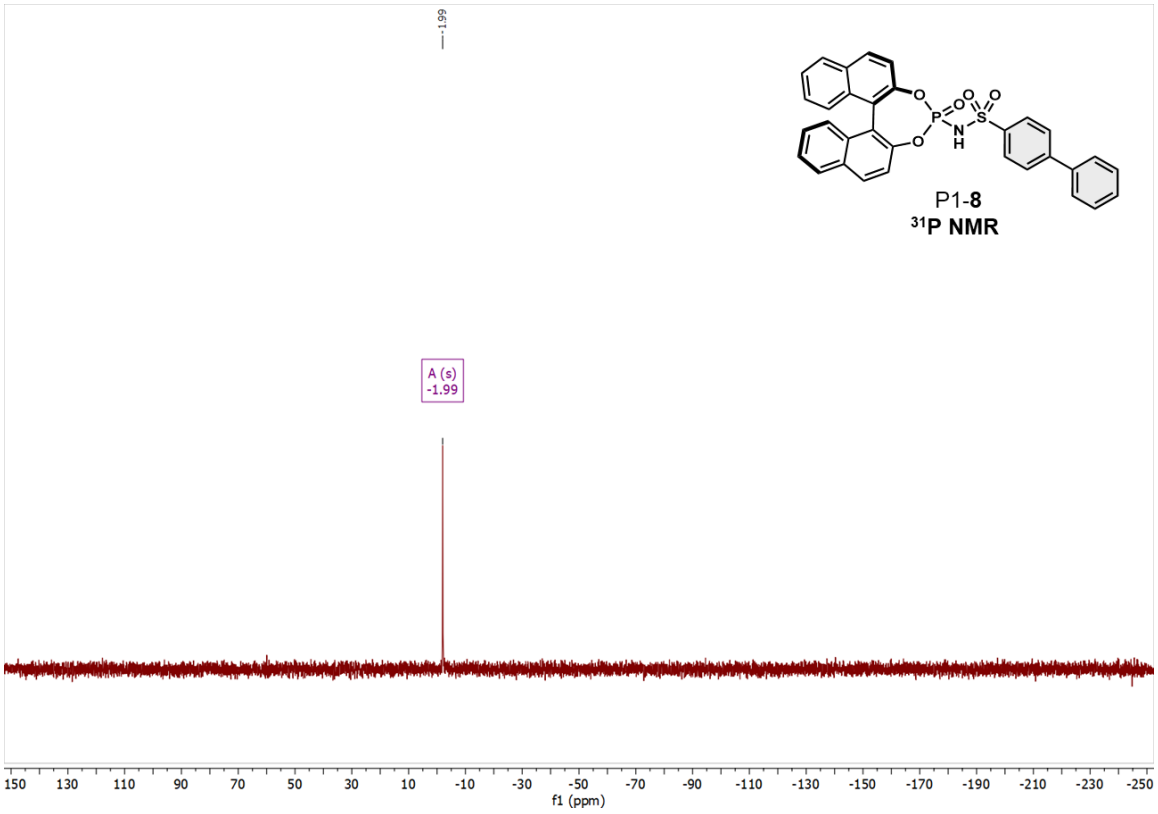
P1-**7**
^{19}F NMR

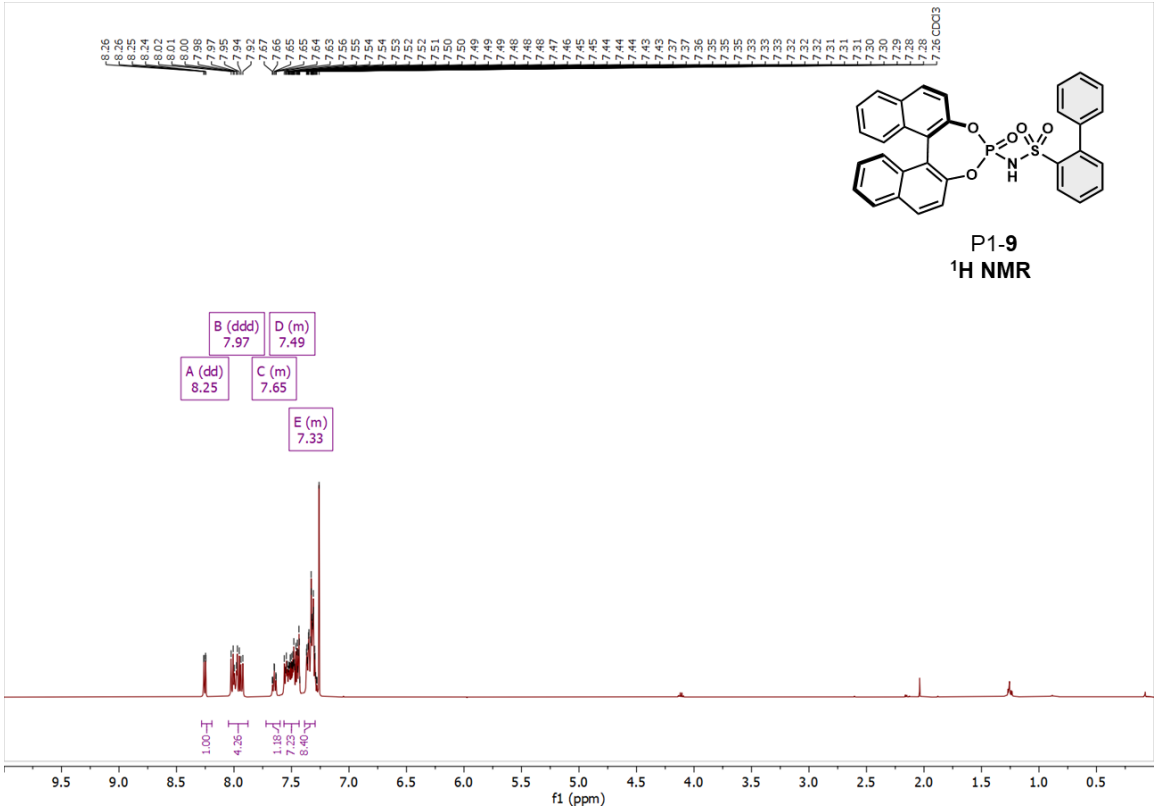


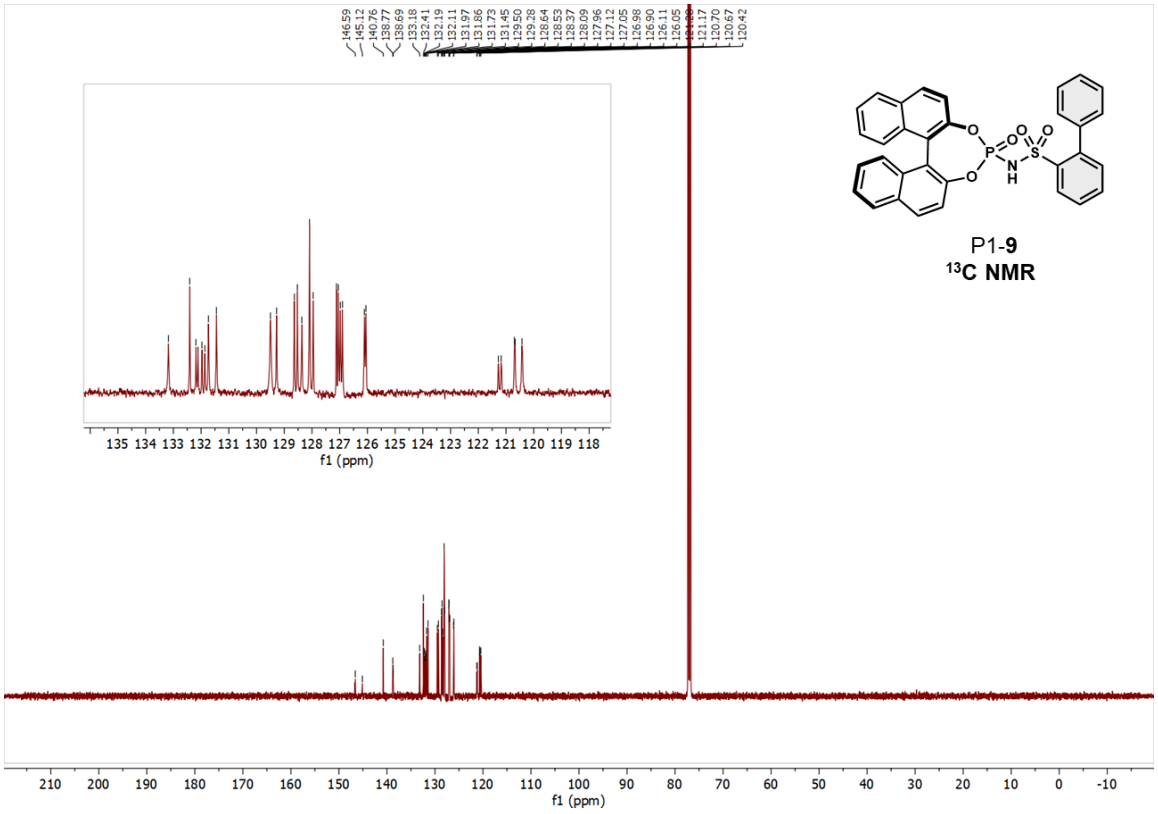
P1-**7**
^{31}P NMR

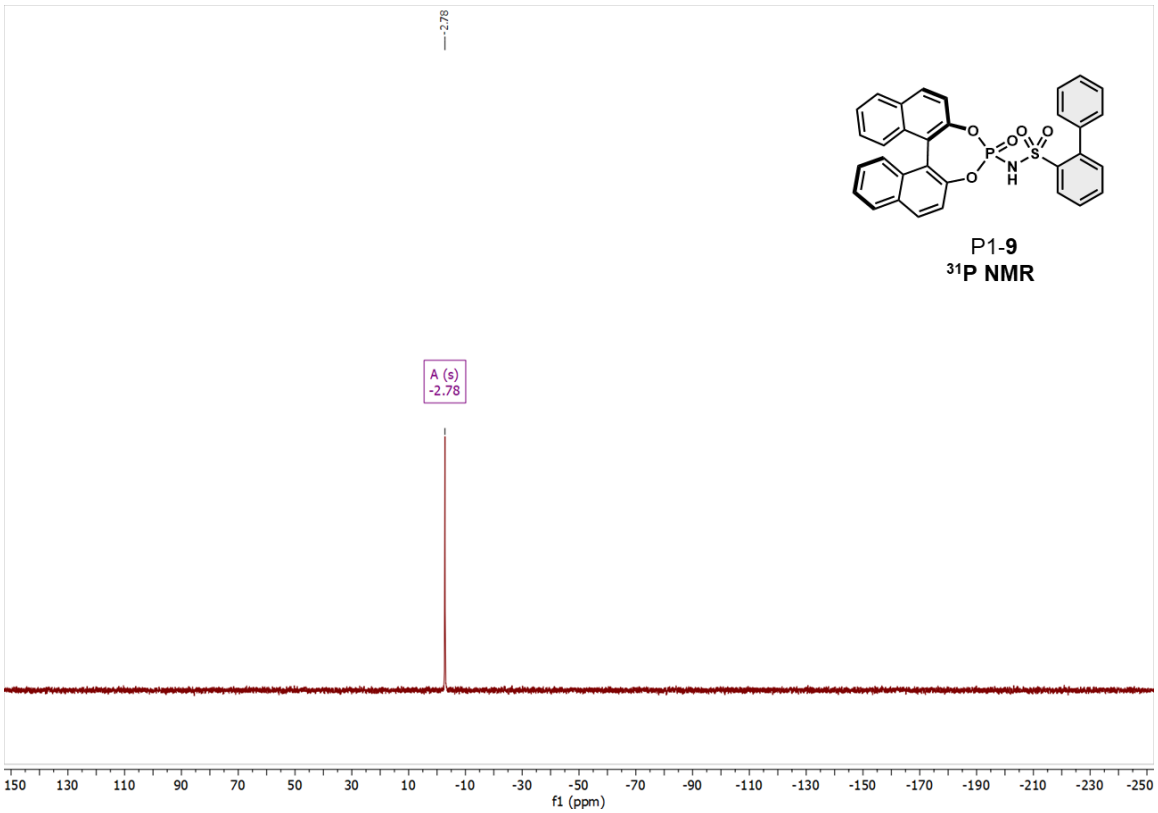

P1-8
1H NMR
f1 (ppm)


P1-8
13C NMR
f1 (ppm)


-1.99
A (s)
-1.99
P1-8
31P NMR
f1 (ppm)


7.26 CDCl3
B (ddd)
7.97
D (m)
7.49
A (dd)
8.25
C (m)
7.65
E (m)
7.33
1.00
4.26
1.18
7.23
8.40
P1-9
1H NMR
f1 (ppm)


146.59
145.12
140.76
138.77
138.69
133.18
132.41
132.19
132.11
131.97
131.86
131.73
131.45
129.50
129.28
128.64
128.53
128.37
128.09
127.96
127.12
127.05
126.98
126.90
126.11
126.05
121.37
121.20
120.67
120.42
135 134 133 132 131 130 129 128 127 126 125 124 123 122 121 120 119 118
f1 (ppm)
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)
P1-9
13C NMR


-2.78
A (s)
-2.78
150 130 110 90 70 50 30 10 -10 -30 -50 -70 -90 -110 -130 -150 -170 -190 -210 -230 -250
f1 (ppm)
P1-9
31P NMR

Appendix VI – Supplementary Information for Chapter VI

Portions of this appendix were adapted from the Supplementary Information of:

Smith, A. L.†; Moeckel, A. C.†; Tandean, E. M.; Toste, F. D. *Manuscript in revision.*

Table of Contents:

VI.1 – Specific Comments on Materials and Methods **519**
VI.2 – Reaction optimization **520**
VI.3 – Decarboxylation **525**
VI.4 – Representative reaction NMR data **539**
VI.5 – Time course data **543**
VI.6 – Mechanistic studies **545**
VI.7 – Synthetic Protocols **571**
VI.8 – Products **582**
VI.9 – Computational Data **584**
VI.10 – NMR Spectra **590**
VI.11 – References **600**

VI.49 – Specific Comments on Materials and Methods

- ***UV Vis spectroscopy***

UV-Vis spectra were recorded on a Varian Cary 50 Bio UV/Vis spectrometer with a Hellma absorption cuvette (1.00 cm path length, max. volume = 3,500 μL).

- ***Stern-Volmer experiments***

Fluorescence spectra were recorded on a Varian Cary Eclipse fluorescence spectrometer. All spectra were recorded in quartz cuvettes with a path length of 1 cm.

- ***Photochemistry Setup***

Samples for irradiation were prepared by dispensing stock solutions of every reaction compound in DCM or methanol into the reaction vials. After evaporation of the solvent the samples were brought into a nitrogen-filled glovebox and charged with the reaction solvent. All solvents used were HPLC-grade and had been sparged and dried over 3 Å or 4 Å sieves prior to use and stored in the glovebox afterwards. Samples were then capped with Teflon caps (brand: Qorpak) and

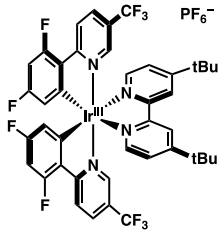
brought outside of the glovebox for irradiation. Irradiation for HAT reactions was conducted with 40 W 390 nm, 440 nm or 456 nm Kessil lamps irradiating the samples at 100% intensity from the bottom with a distance of 8 cm from the lamp. A fan (118 CFM) was placed behind the lamp and samples. Annulation reactions were irradiated with 1.2 W 275 nm LED lamps from the bottom without cooling.

Supplementary Fig. VI-1. Photochemistry setup with 440 nm Kessil lamps.

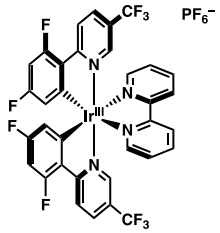
VI.50 – Reaction optimization

Reported conversions and yields are NMR-based. 1,3,5-trifluorobenzene, 1,4-dinitrobenzene, or 1,3,5-triisopropylbenzene were used as internal standards.

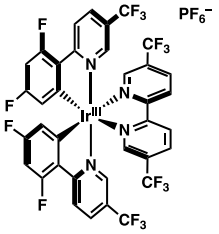
Photocatalysts



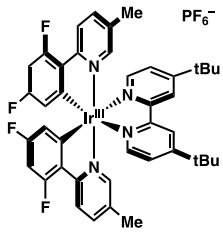
PC-1



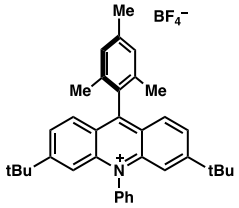
PC-2



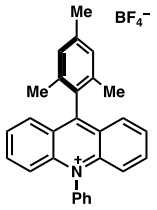
PC-3



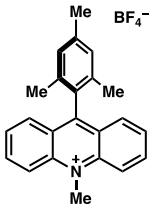
PC-4



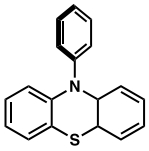
PC-5



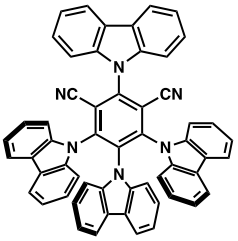
PC-6



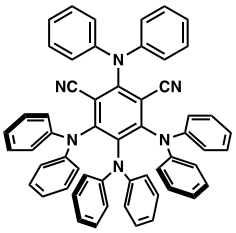
PC-7



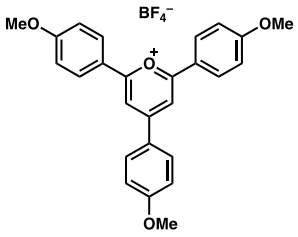
PC-8



PC-9

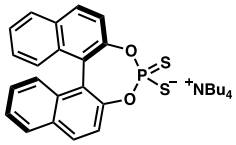


PC-10

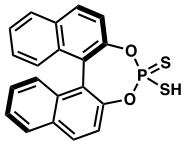


PC-11

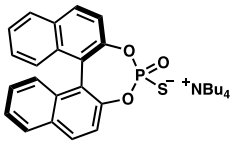
Thiyl catalysts



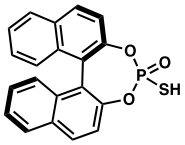
P-1



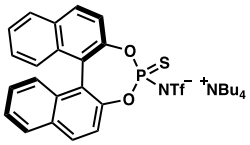
P-2



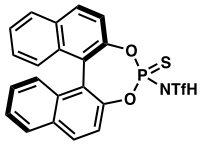
P-3



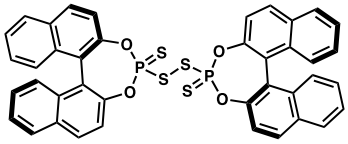
P-4



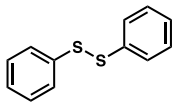
P-5



P-6



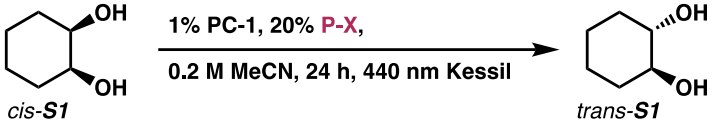
P-7



P-8

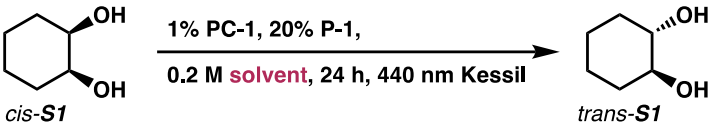
- ***Diol epimerization***

Supplementary Table VI-52. Thiyl catalyst screen for diol epimerization. With 20 μmol ***cis*-S1**.



Entry	Thiyl catalyst	Conversion	Yield
1	**P-1**	47%	23%
2	**P-2**	42%	0%
3	**P-3**	24%	0%
4	**P-4**	46%	0%
5	**P-5**	28%	4%
6	**P-6**	65%	2%
7	**P-7**	63%	0%
8	–	13%	0%

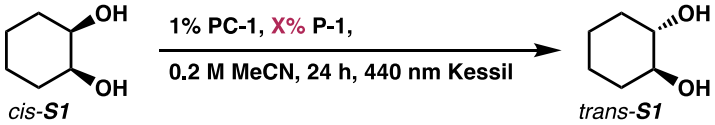
Supplementary Table VI-53. Solvent screen for diol epimerization. With 20 μmol ***cis*-S1**.



Entry	Solvent	Conversion	Yield
1	Toluene	21%	0%
2	Benzene	30%	9%
3	Chlorobenzene	29%	4%
4	Trifluorotoluene	18%	6%
5	Acetone	5%	0%
6	THF	-22%	0%
7	DCM	5%	5%
8	DMSO	33%	24%

9	*tert*-amyl alcohol	10%	2%
10	MeCN/H2O (30%)	42%	7%
11	MeCN/H2O (15%)	42%	7%

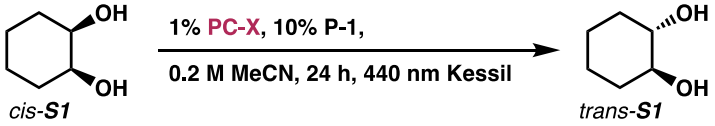
Supplementary Table VI-54. Thiyl catalyst loading screen for diol epimerization. With 20 µmol ***cis*-S1**.



Entry	Thiyl catalyst loading	Conversion	Yield
1	40%	39%	13%
2	20%	59%	31%
3	10%	64%	37%
4	5%	56%	45%[a]
5	2.5%	66%	47%[a]

[a] significant thiyl catalyst decomposition observed

Supplementary Table VI-55. Photocatalyst screen for diol epimerization. With 20 µmol ***cis*-S1**.

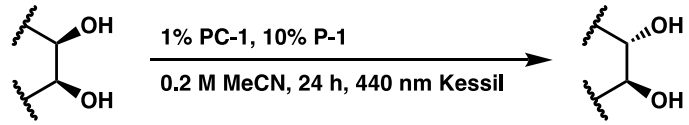


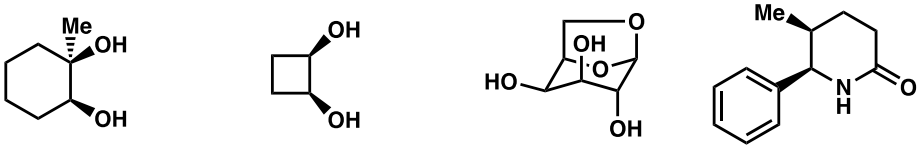
Entry	Photocatalyst	Conversion	Yield
1	PC-1	36%	33%
2	PC-2	17%	0%
3	PC-3	16%	0%
4	PC-4	14%	0%
5	PC-5	13%	0%

6	PC-6	16%	0%
7	PC-7	17%	0%
8	PC-9	13%	0%
9	PC-10	27%	0%
10[b]	PC-1	24%	0%
11	–	14%	0%

[b] no light

- ***Diol epimerization – Additional substrates***



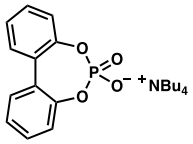


Yield	10%	27%	9%	10%
Yield b.r.s.m.	26%	45%	100%*	65%
Conversion	36%	59%	5%	15%
Literature yield	29%[2]	51%[2]	82%[2]	95%[3]

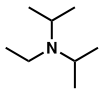
*normalized to 100% total yield

VI.51 – Decarboxylation

Bases

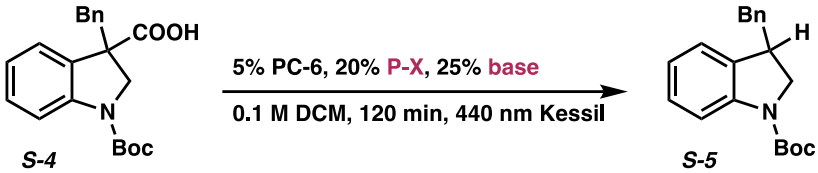


B-1



B-2

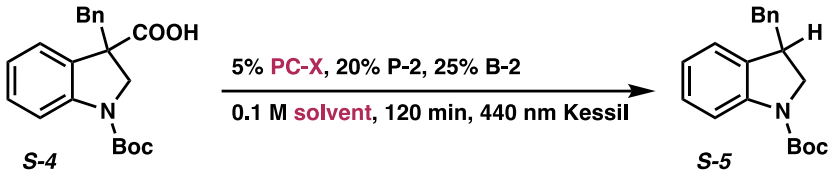
Supplementary Table VI-56. Thiyl and base catalyst screen for decarboxylation. With 10 µmol **S4**.



Entry	Thiyl catalyst	Base	Conversion	Yield
1	**P-1**	**B-1**	40%	3%
2	**P-2**		42%	0%
3	**P-3**		59%	8%
4	**P-4**		28%	4%
5	**P-5**		34%	8%
6	**P-6**		38%	12%
7	**P-7**		45%	7%
8	**P-8**		47%	5%
9	–		33%	10%
10	**P-1**	**B-2**	30%	1%
11	**P-2**		8%	0%
12	**P-3**		12%	3%
13	**P-4**		0%	8%
14	**P-5**		47%	5%

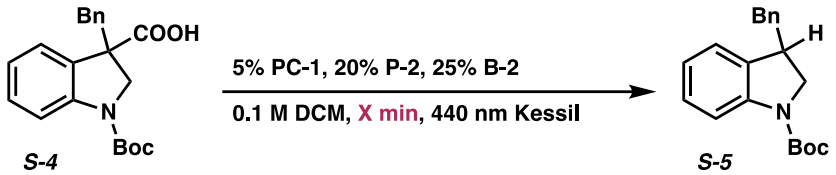
15	**P-6**		57%	17%
16	**P-7**		37%	7%
17	**P-8**		60%	16%
18	–		33%	10%
19	**P-1**		16%	2%
20	**P-2**		7%	1%
21	**P-3**		50%	2%
22	**P-4**		5%	5%
23	**P-5**		20%	1%
24	**P-6**	–	35%	3%
25	**P-7**		15%	1%
26	**P-8**		18%	4%
27	–		13%	0%

Supplementary Table VI-57. Photocatalyst and solvent screen for decarboxylation. With 10 μmol **S4**.



Entry	Solvent	Photocatalyst	Conversion	Yield
1	Toluene	**PC-3**	89%	20%
2	Benzene		90%	11%
3	MeCN		97%	18%
4	Acetone		83%	11%
5	Toluene	**PC-2**	72%	22%
6	Benzene		59%	26%

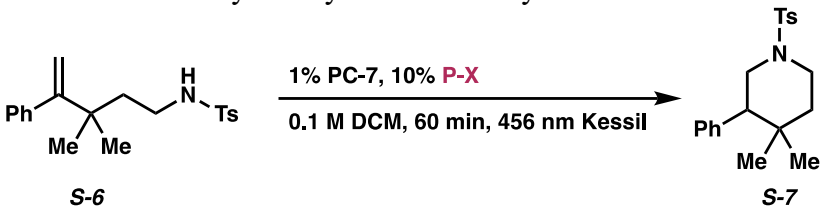
7	MeCN		58%	23%
8	MeCN/H2O (15%)		80%	5%
9	THF		46%	3%
10	Toluene		64%	3%
11	Benzene		64%	6%
12	MeCN		53%	14%
13	MeCN/H2O (15%)	**PC-9**	78%	12%
14	Acetone		66%	17%
15	THF		57%	3%
16	*tert*-amyl alcohol		34%	6%



Entry	Time	Conversion	Yield
1	30 min	20%	16%
2	45 min	61%	47%
3	60 min	72%	46%
4	75 min	73%	63%
5	90 min	77%	55%
6	105 min	79%	55%
7	120 min	82%	59%

- ***Hydroamination***

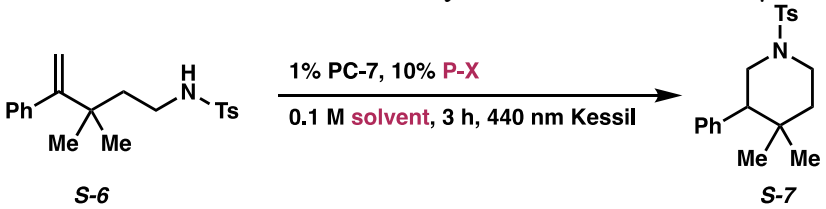
Supplementary Table VI-58. Thiyl catalyst screen for hydroamination. With 10 μmol **S6**.



S-6 *S-7*

Entry	Thiyl catalyst	Conversion	Yield
1	**P-1**	-14%	0%
2	**P-2**	-7%	36%
3	**P-3**	-14%	9%
4	**P-4**	9%	6%
5	**P-5**	4%	0%
6	**P-6**	26%	19%
7	**P-7**	35%	34%
8	**P-8**	-21%	32%

Supplementary Table VI-59. Solvent screen for hydroamination. With 10 μmol **S6**.

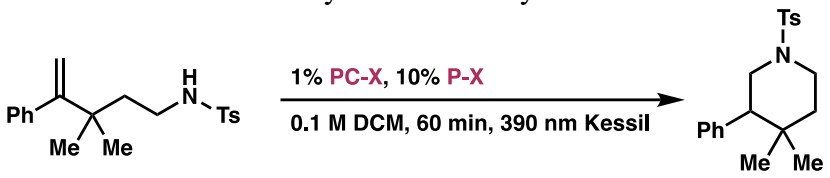


S-6 *S-7*

Entry	Thiyl catalyst	Solvent	Conversion	Yield
1	**P-2**	Toluene	99%	64%
2		Benzene	97%	66%
3		DCM	93%	64%
4		MeCN	98%	77%
5	**P-1**	Toluene	20%	10%

6	Benzene	14%	14%
7	DCM	20%	11%
8	MeCN	22%	21%

Supplementary Table VI-60. Photocatalyst screen for hydroamination. With 10 µmol **S6**.



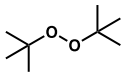
S-6 *S-7*

Entry	Thiyl catalyst	Photocatalyst	Conversion	Yield
1	P-2	**PC-1**	17%	0%
2		**PC-2**	49%	0%
3		**PC-3**	41%	0%
4		**PC-4**	28%	0%
5		**PC-5**	11%	8%
6		**PC-6**	5%	20%
7		**PC-7**	49%	6%
8		**PC-8**	27%	0%
9		**PC-9**	54%	0%
10		**PC-10**	42%	0%
11	P-1	**PC-1**	17%	0%
12		**PC-2**	49%	0%
13		**PC-3**	41%	0%
14		**PC-4**	28%	0%
15		**PC-5**	9%	1%
16		**PC-6**	53%	1%
17		**PC-7**	21%	2%

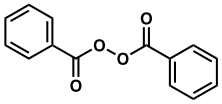
18	**PC-8**	35%	0%
19	**PC-9**	52%	0%
20	**PC-10**	26%	0%

- ***Annulation***

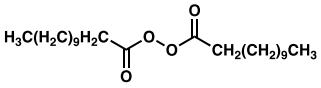
Initiators



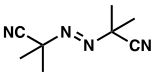
In-1



In-2

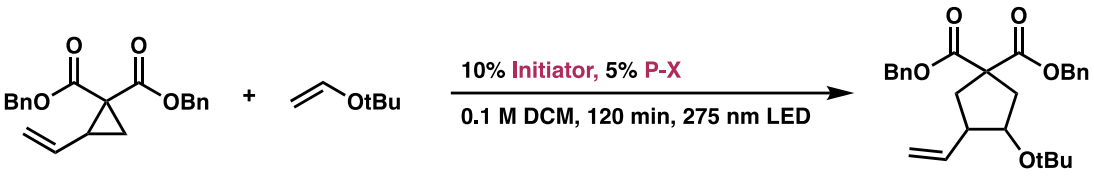


In-3



In-4

Supplementary Table VI-61. Initiator and Thiyl catalyst screen for annulation reaction. With 50 µmol **S8** and 100 µmol *tert*-butylvinyl ether.

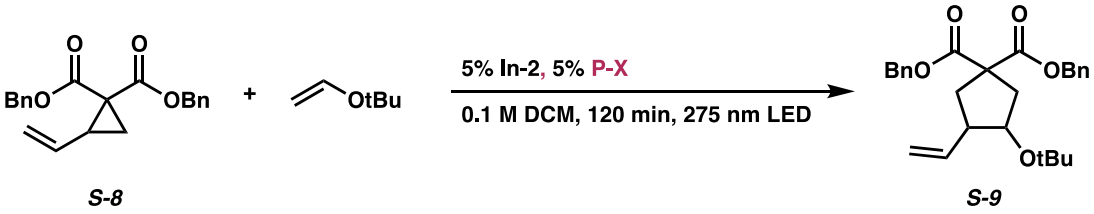


S-8 ***S-9***

Entry	Initiator	Thiyl catalyst	Conversion	Yield
1		**P-1**	88%	89%
2		**P-2**	63%	63%
3	**In-1**	**P-7**	84%	85%
4		–	2%	2%
5		**P-1**	100%	100%
6	**In-2**	**P-2**	37%	37%
7		**P-7**	83%	83%

8		–	2%	2%
9	In-3	**P-1**	74%	24%
10		**P-2**	73%	14%
11		**P-7**	34%	2%
12		–	46%	1%
13	In-4	**P-1**	76%	1%
14		**P-2**	58%	9%
15		**P-7**	80%	40%
16		–	76%	0%
17	–	**P-1**	42%	21%
18		**P-2**	54%	16%
19		**P-7**	81%	56%
20		–	15%	1%

Supplementary Table VI-62. Thiyl catalyst screen for annulation reaction. With 50 µmol **S8** and 100 µmol *tert*-butylvinyl ether.

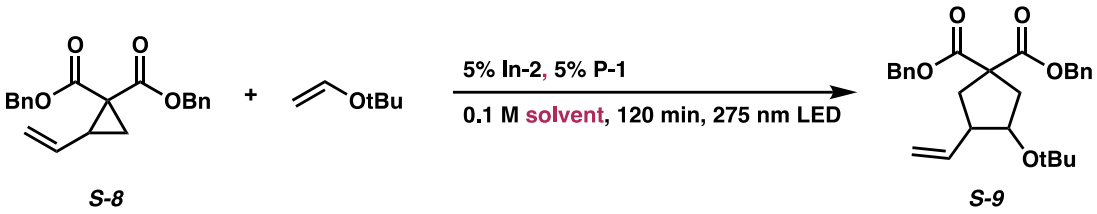


S-8 *S-9*

Entry	Thiyl catalyst	Conversion	Yield
1	**P-1**	100%	57%
2	**P-2**	83%	55%
3	**P-3**	31%	19%
4	**P-4**	53%	34%
5	**P-5**	5%	3%

6	**P-6**	6%	4%
7	**P-7**	99%	72%
8	–	3%	2%

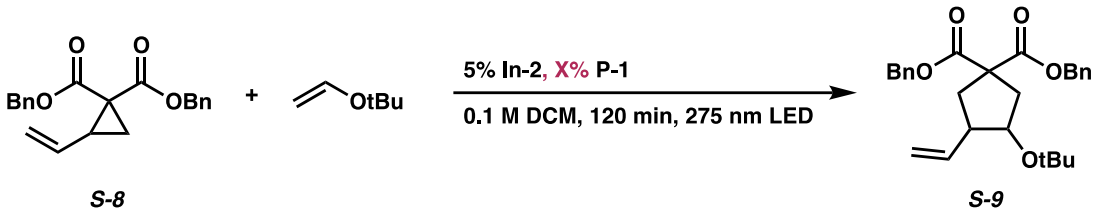
Supplementary Table VI-63. Solvent screen for annulation reaction. With 50 µmol **S8** and 100 µmol *tert*-butylvinyl ether.



S-8 *S-9*

Entry	Solvent	Conversion	Yield
1	Toluene	21%	2%
2	Benzene	12%	10%
3	Chlorobenzene	80%	72%
4	Trifluorotoluene	-6%	6%
5	Acetone	81%	59%
6	THF	99%	109%
7	DCM	98%	65%
8	DMSO	100%	107%
9	*tert*-amyl alcohol	20%	6%
10	*MeCN*	100%	88%
11	MeCN/H2O (30%)	99%	78%
12	MeCN/H2O (15%)	99%	20%

Supplementary Table VI-64. Thiyl catalyst loading screen for annulation reaction. With 50 µmol **S8** and 100 µmol *tert*-butylvinyl ether.

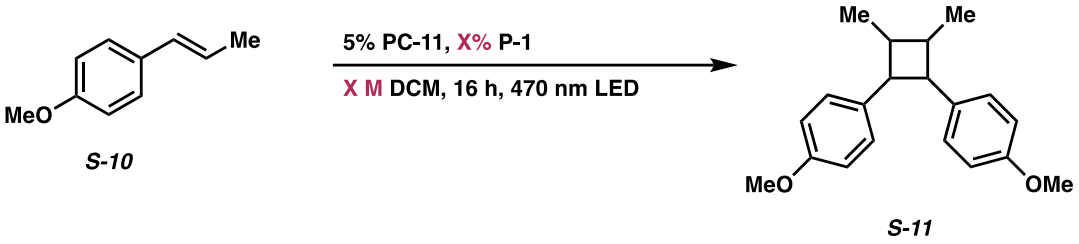


S-8 *S-9*

Entry	Thiyl catalyst loading	Conversion	Yield
1	5%	100%	119%
2	2.5%	104%	100%
3	1.3%	100%	105%
4	6,000 ppm	70%	73%
5	3,000 ppm	13%	75%
6	1,500 ppm	-3%	9%
7	750 ppm	-15%	2%

- ***[2+2] Cycloaddition***

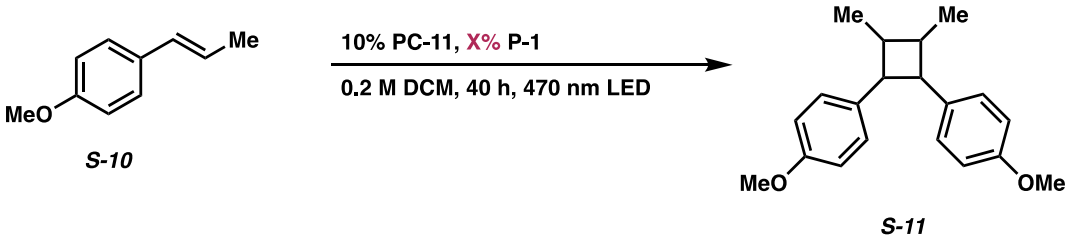
Supplementary Table VI-65. Thiyl catalyst loading and concentration screen for [2+2] cycloaddition reaction. With 12.5 μmol **S10**. Average across two trials.



S-10 *S-11*

Entry	Concentration	P-1 Loading	Yield (b.r.s.m.)	Mass balance
1	**0.2 M**	**1.25%**	24%	66%
2		**0.60%**	18%	43%
3		**0.30%**	20%	43%
4		**0.15%**	26%	44%
5		**0.08%**	23%	46%
6		**0%**	24%	46%
7	**0.1 M**	**1.25%**	17%	59%
8		**0.60%**	32%	60%
9		**0.30%**	46%	64%
10		**0.15%**	40%	59%
11		**0.08%**	40%	58%
12		**0%**	28%	56%

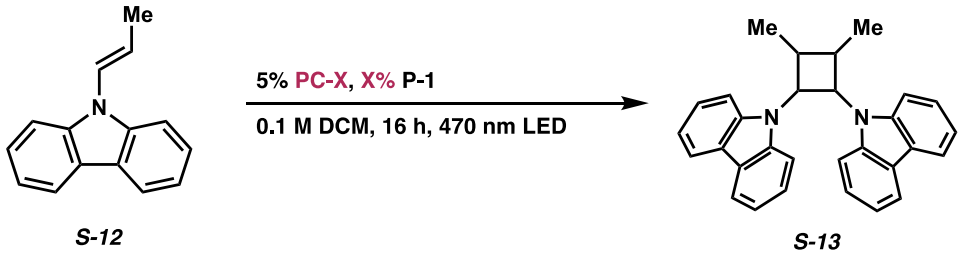
Supplementary Table VI-66. Thiyl catalyst loading screen for [2+2] cycloaddition reaction (from **Fig. 5D**). With 12.5 μmol **S10**. Average across two trials.



S-10 *S-11*

Entry	P-1 Loading	Yield (b.r.s.m.)	Mass balance
1	**0%**	20%	35%
2	**0.15%**	43%	59%
3	**0.30%**	37%	50%
4	**0.60%**	21%	41%
5	**1.3%**	24%	50%
6	**2.5%**	12%	42%
7	**5.0%**	10%	75%

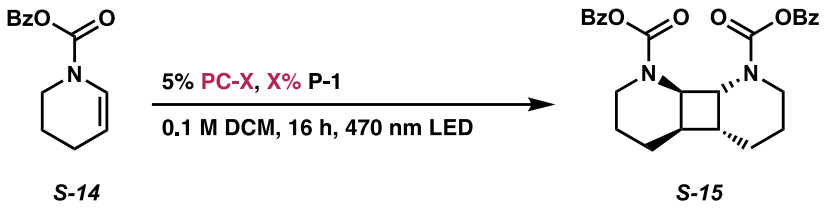
Supplementary Table VI-67. Thiyl catalyst loading screen for [2+2] cycloaddition reaction. With 12.5 μmol **S12.**



S-12 *S-13*

Entry	PC	P-1 Loading	Yield (b.r.s.m.)	Mass balance
1	PC-6	**1.2%**	0%	9%
2		**0.6%**	0%	8%
3		**0.3%**	0%	7%
6		**0%**	7%	9%
7	PC-11	**1.2%**	-	-
8		**0.6%**	0%	4%
9		**0.3%**	6%	8%
12		**0%**	20%	22%

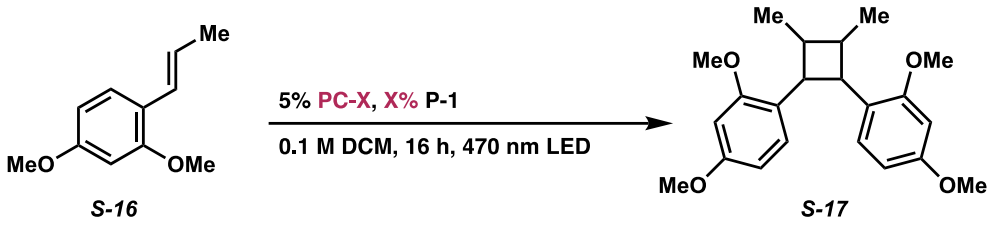
Supplementary Table VI-68. Thiyl catalyst loading screen for [2+2] cycloaddition reaction (from **Fig. 5E**). With 12.5 μmol **S14.**



S-14 *S-15*

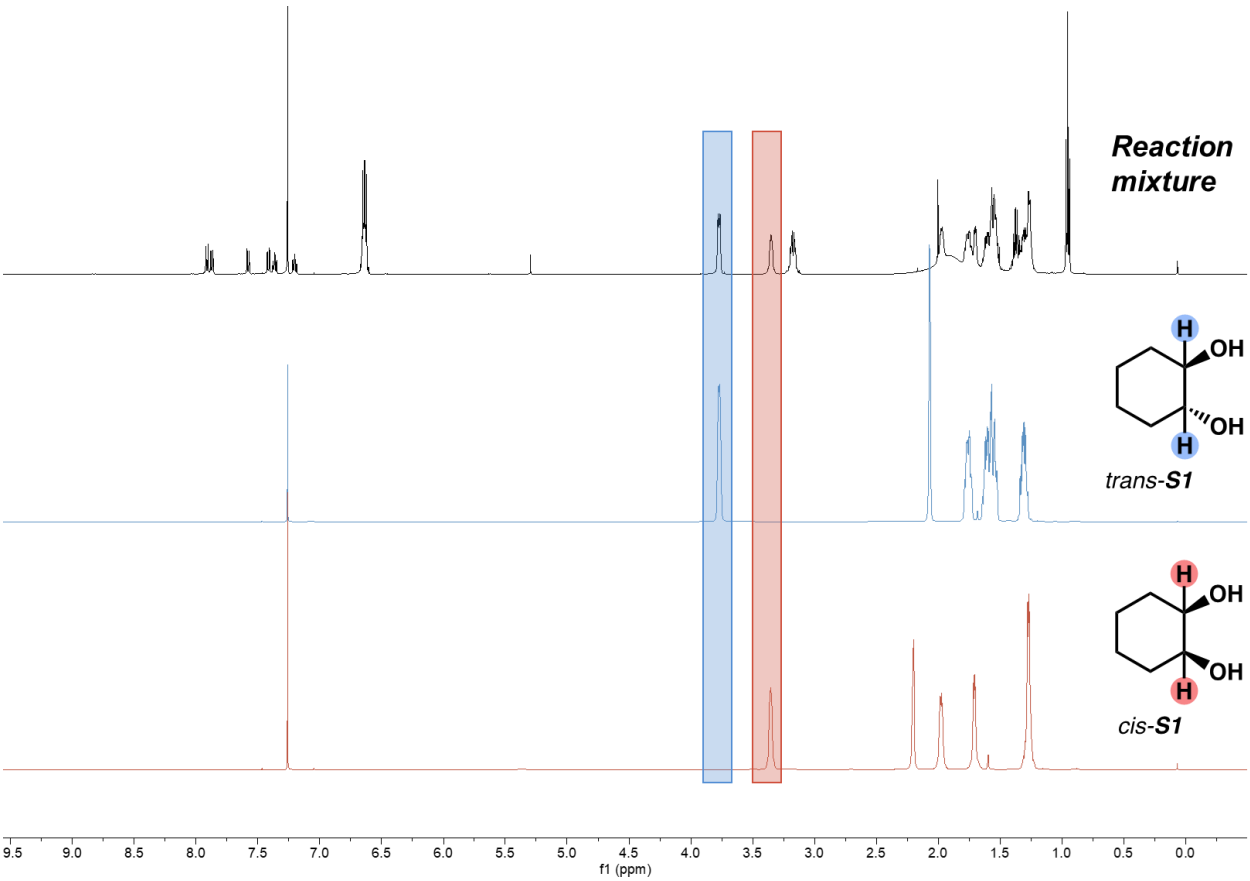
Entry	PC	P-1 Loading	Yield (b.r.s.m.)	Mass balance
1	PC-6	**1.2%**	13%	61%
2		**0.6%**	17%	67%
3		**0.3%**	26%	72%
6		**0%**	26%	68%
7	PC-11	**1.2%**	42%	42%
8		**0.6%**	40%	40%
9		**0.3%**	41%	41%
12		**0%**	41%	41%

Supplementary Table VI-69. Thiyl catalyst loading screen for [2+2] cycloaddition reaction (from **Fig. 5F**). With 12.5 μmol **S16.**

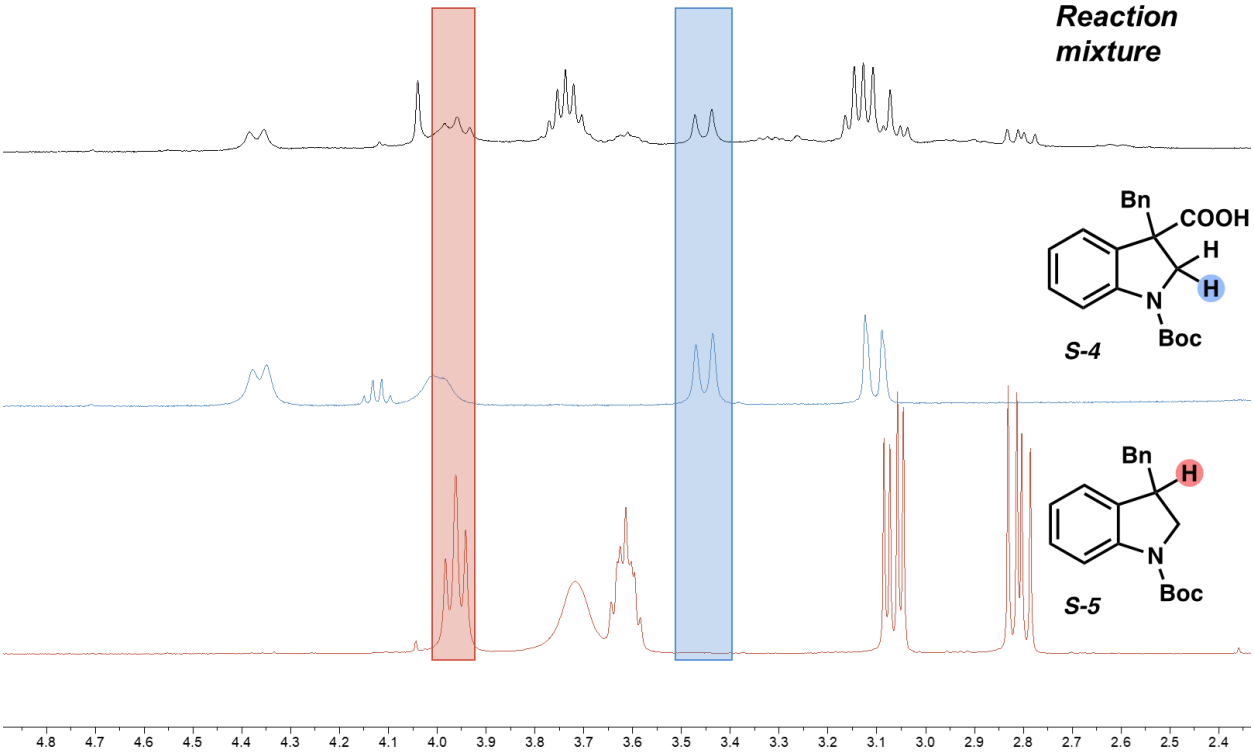


Entry	PC	P-1 Loading	Yield (b.r.s.m.)	Mass balance
1	PC-6	**1.2%**	16%	38%
2		**0.6%**	20%	45%
3		**0.3%**	28%	53%
6		**0%**	46%	65%
7	PC-11	**1.2%**	25%	58%
8		**0.6%**	32%	63%
9		**0.3%**	38%	67%
12		**0%**	25%	53%

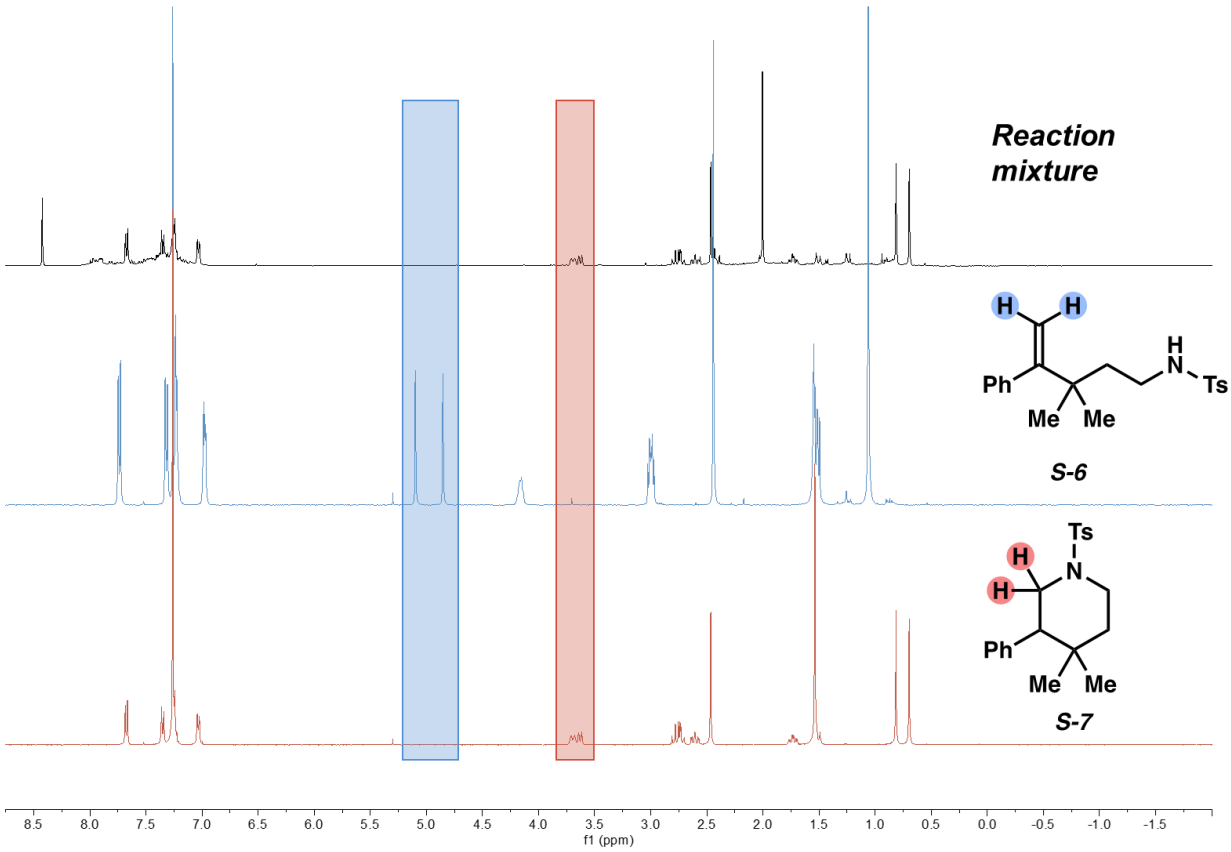
VI.52 – Representative reaction NMR data



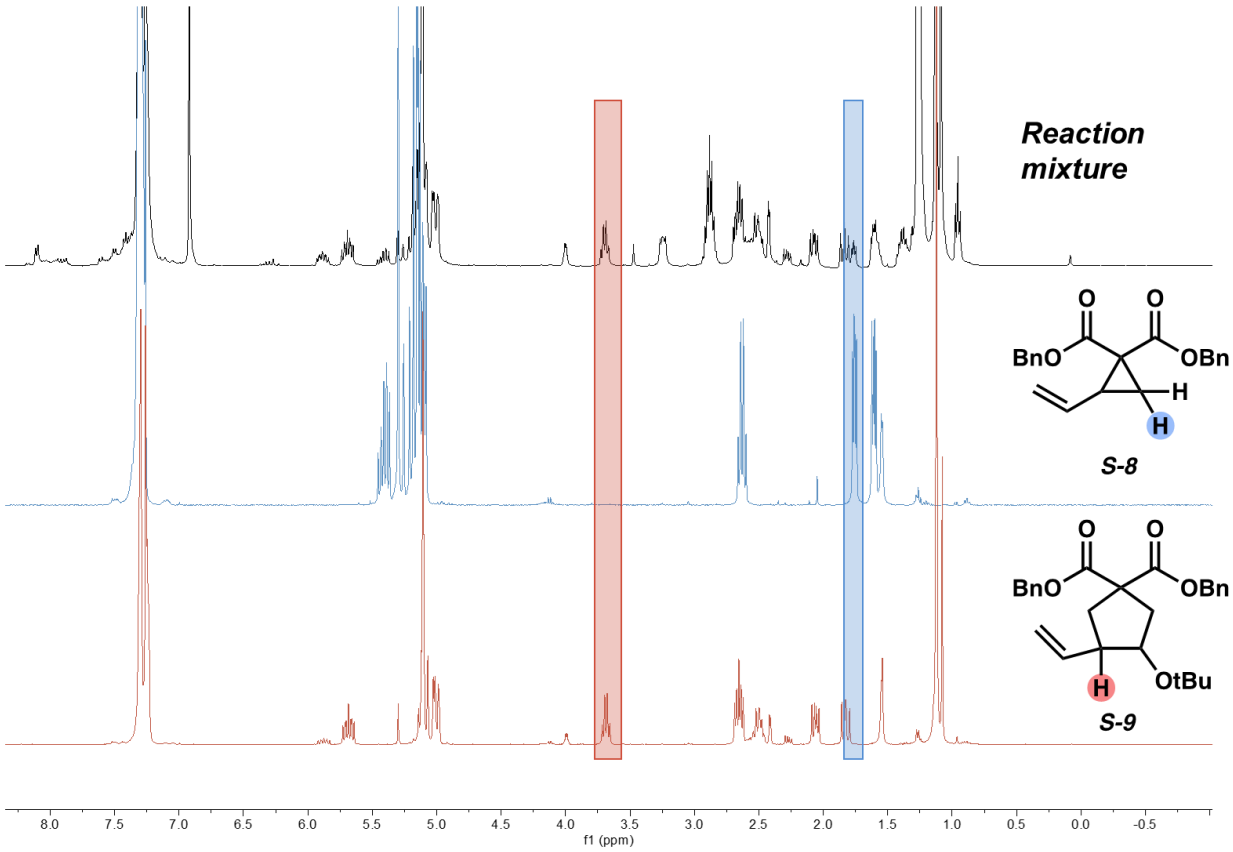
Supplementary Fig. VI-2. Overlayed NMR spectra of diol epimerization reaction mixture, ***trans*-S1** and ***cis*-S1**.



Supplementary Fig. VI-3. Overlayed NMR spectra of decarboxylation reaction mixture, **S4** and **S5**.



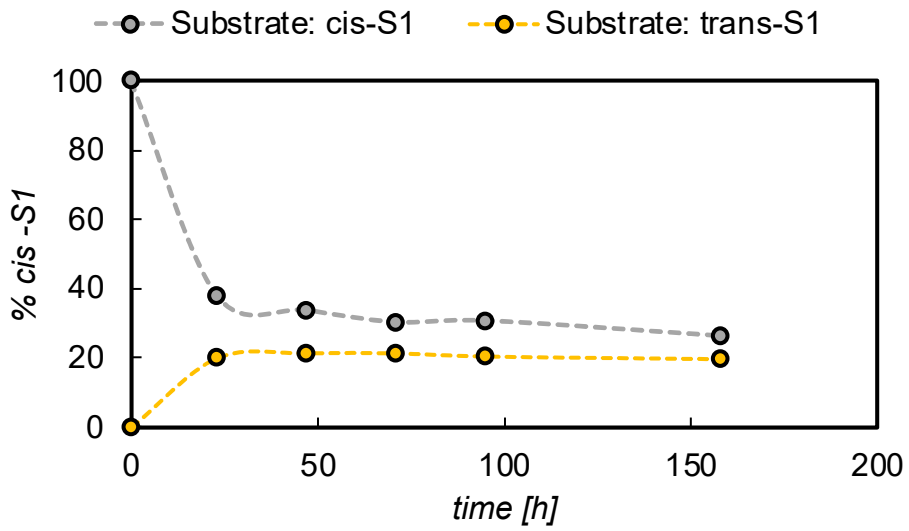
Supplementary Fig. VI-4. Overlayed NMR spectra of hydroamination reaction mixture, **S6** and **S7**.



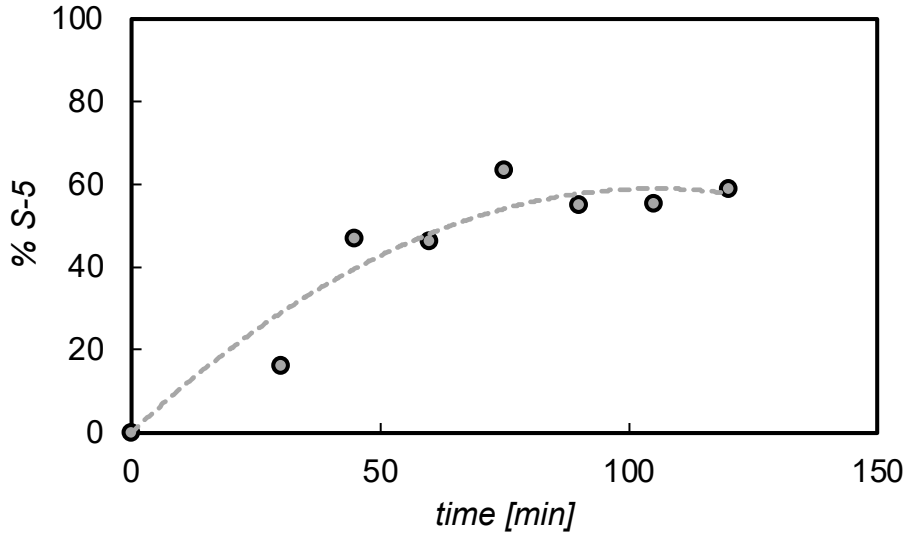
Supplementary Fig. VI-5. Overlayed NMR spectra of annulation reaction mixture, **S8** and **S9**.

VI.53 – Time course data

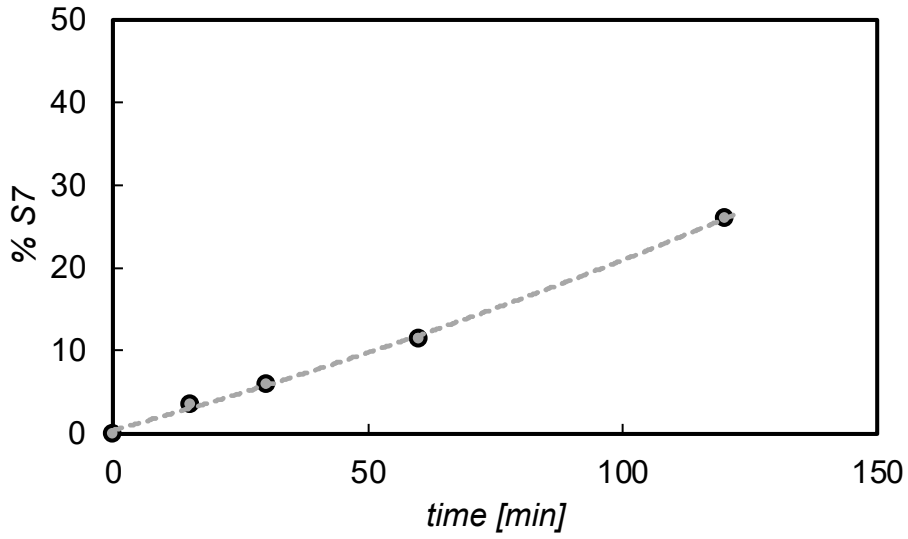
All time course data is based on NMR yields. 1,3,5-trifluorobenzene, 1,4-dinitrobenzene, or 1,3,5-triisopropylbenzene were used as internal standards.



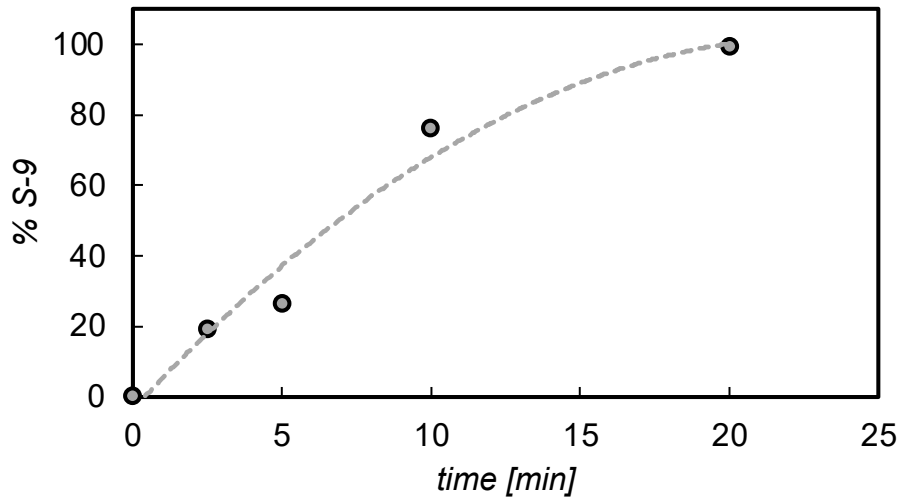
Supplementary Fig. VI-6. Time course data for diol epimerization. *Reaction conditions:* 1**% PC-1**, 10% **P-1**, 20 µmol ***cis*-S1**, 20 µmol ***trans*-S1**, MeCN (0.2 M) with 440 nm irradiation.



Supplementary Fig. VI-7. Time course data for decarboxylation. *Reaction conditions:* 5% **PC-3**, 20% **P-2**, 10 µmol **S4**, DCM (0.1 M) with 440 nm irradiation.



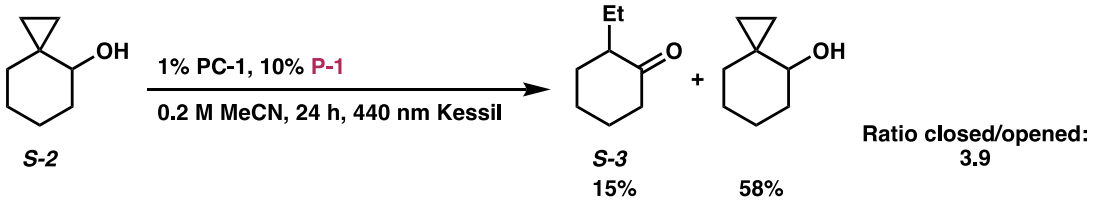
Supplementary Fig. VI-8. Time course data for hydroamination. *Reaction conditions:* 1% **PC-7**, 10% **P-2**, 10 μmol **S6**, DCM (0.1 M) with 456 nm irradiation.



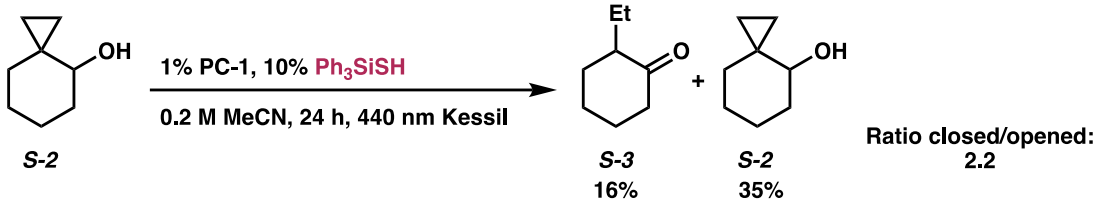
Supplementary Fig. VI-9. Time course data for annulation. *Reaction conditions:* 5% **In-2**, 5% **P-1**, 50 μmol **S8**, 100 μmol *tert*-butylvinyl ether, DCM (0.1 M) with 275 nm irradiation.

VI.54 – Mechanistic studies

- ***Radical clock experiment***

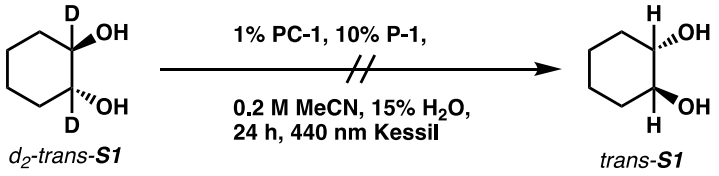


Supplementary Fig. VI-10. Radical clock experiment with **P-1** and 11 µmol of **S2**.

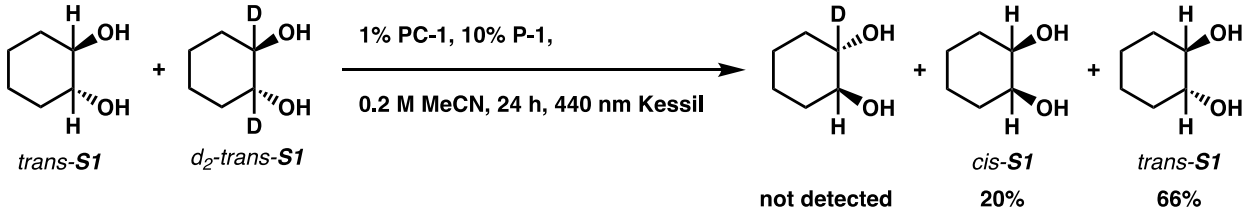


Supplementary Fig. VI-11. Radical clock experiment with Ph_3SiSH and 11 µmol of **S2**.

- ***Deuterium labeling studies***

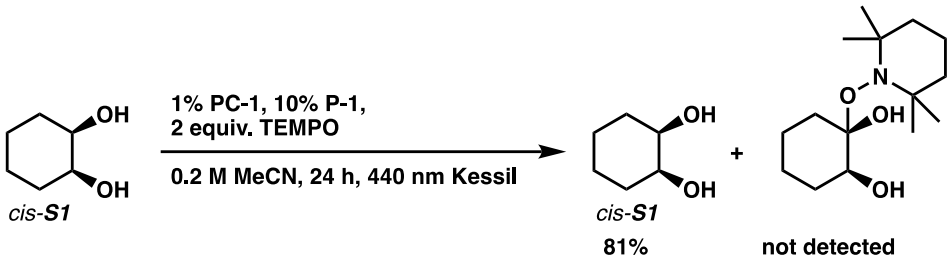


Supplementary Fig. VI-12. Deuterium labelling experiment with 20 µmol ***d2-trans*-S1.** Hydrogen atom incorporation was not observed by ^1H NMR.



Supplementary Fig. VI-13. D/H crossover experiment. % Remaining ***d*$_2$-S1** could not be determined due to overlapping ^1H NMR signals with **S1**. With 11 µmol **S1** and 11 µmol ***d*$_2$*-trans*-S1.** The crossover product was not observed by HRMS.

- ***Radical trap experiment***



Supplementary Fig. VI-14. Radical trap experiment. The TEMPO-adduct of **S1** was not observed by HRMS.

- ***Stern Volmer experiments***

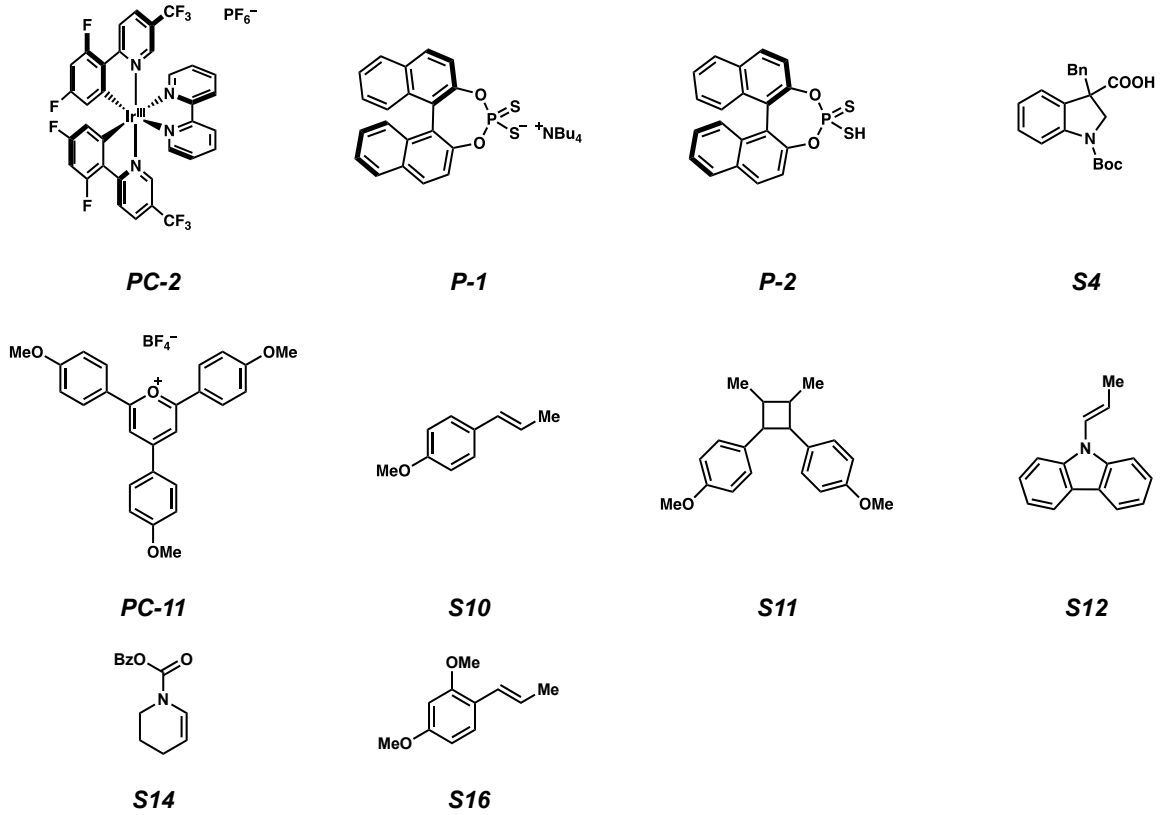
In an argon atmosphere glovebox were prepared the corresponding stock solutions of photocatalyst and substrate (2.5 mM of PC-2, 20 mM of quencher (**P-1**, **P-2**, **S4 (**with 25 mol% diethylisopropylamine)**)** in MeCN and DCM. To each of four fluorescence cuvettes was dosed photocatalyst stock solution (60 µL of PC-2 stock and added either 0 µL, 50 µL, 100 µL, or 200 µL of substrate stock solution. The solutions were diluted to 2,000 µL total volume corresponding to four samples with 0 mM, 0.5 mM, 1.0 mM, and 2.0 mM final concentration of **quencher**. Each cuvette was sealed with a Teflon-lined screen top and further secured with electrical tape. The cuvettes were removed from the glovebox, transferred to the spectrometer, and fluorescence spectra were measured (λ_{exc} = 379 nm and λ_{em} = 475-550 nm with 2.5-5 nm excitation and emission slit widths for **PC-2**.

For [2+2] reactions, a 50 mM stock solution in DCM of **PC-11** was prepared and 20 µL were added to a vial. A 0.1 M stock solution of **quencher** (**S-10**, **S-11, S-12**, **S-14**, and **S-16)** was prepared and added to the vials (0 µL, 20 µL, 40 µL, or 80 µL). The solvent was evaporated, and the vials were moved into a nitrogen-filled glovebox. 2000 µL of DCM was added to each vial and the solutions were transferred to cuvettes corresponding to four samples with 0 mM, 1.0 mM, 2.0 mM, and 4.0 mM final concentration of **quencher**. Fluorescence spectra were measured (λ_{exc} = 470 nm and λ_{em} = 550 nm with excitation slit: 20 excitation and 2.5 nm emission slit widths.

The fluorescence quenching data was converted to a Stern-Volmer quenching constant (K_{SV}) using the linear relationship, where I is the observed intensity, I_0 is the observed intensity at 0 mM of **quencher**, and [Q] is the concentration of quencher in mM:

$$\frac{I_0}{I} = 1 + K_{SV}[\mathrm{Q}]$$

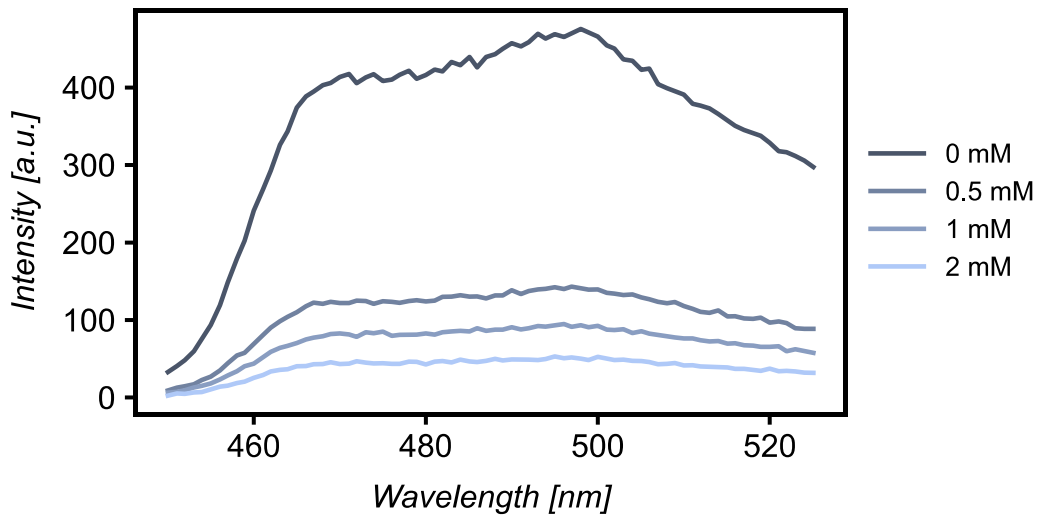
Goodness of fit is reported with R^2 and the K_{SV} is reported as the numerical average across two duplicate trials, unless otherwise noted.



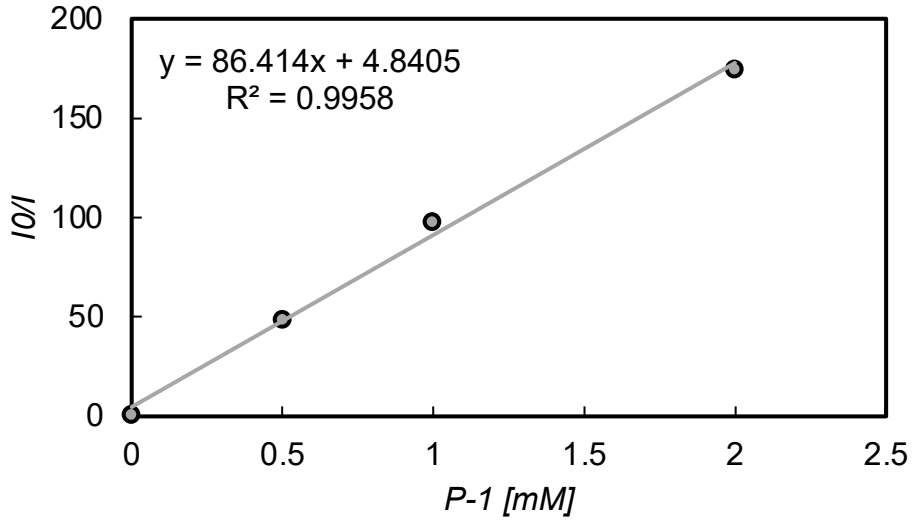
Supplementary Fig. VI-15. Compounds used in Stern–Volmer quenching experiments.

Supplementary Table VI-70. Overview of Stern–Volmer constants of **P-1**, **P-2**, **S4**, **S10** and **S11, S12, S14 and S16** in DCM and MeCN with **PC-2**, **PC-6**, or **PC-11.** Reported as numerical average across two trials.

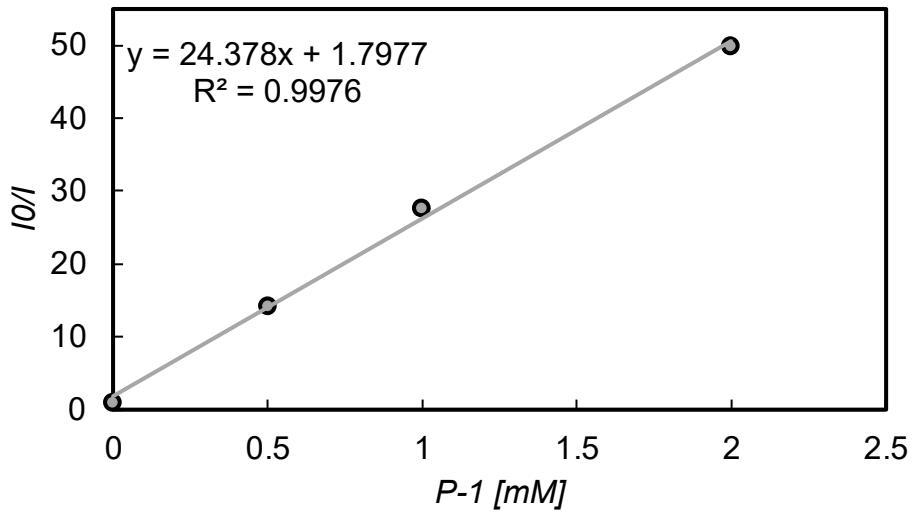
Photocatalyst	Quencher	K_{SV} in DCM [mM^{-1}]	K_{SV} in MeCN [mM^{-1}]
PC-2	**P-1**	86.4	24.4
	P-2	3.8	14.7
	S4	12.0	14.4
PC-7	**P-1**	0.76	0.51
	P-2	0.08	0.24
PC-11	**P1**	6.7	–
	S10	0.12	–
	S12	0.16	–
	S14	0.08	–
	S16	0.12	



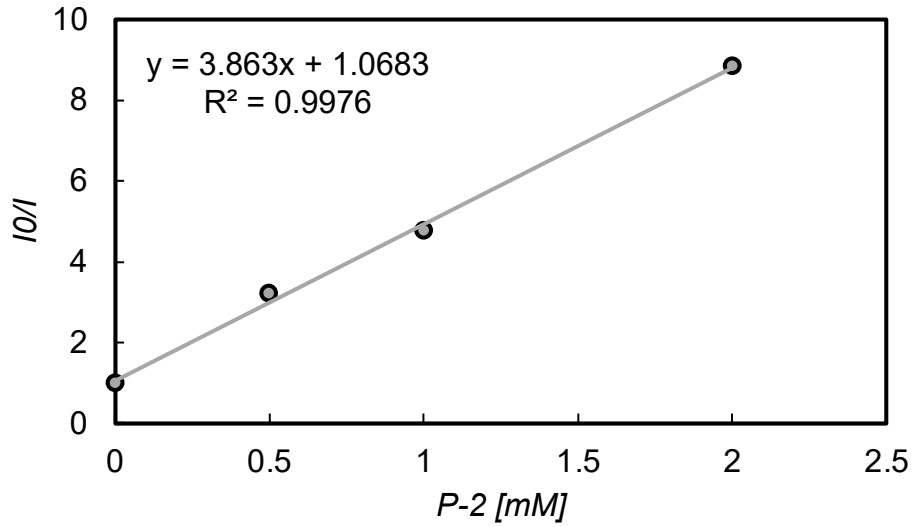
Supplementary Fig. VI-16. Luminescence quenching of **PC-2** with varying concentrations of **P-2** in DCM.



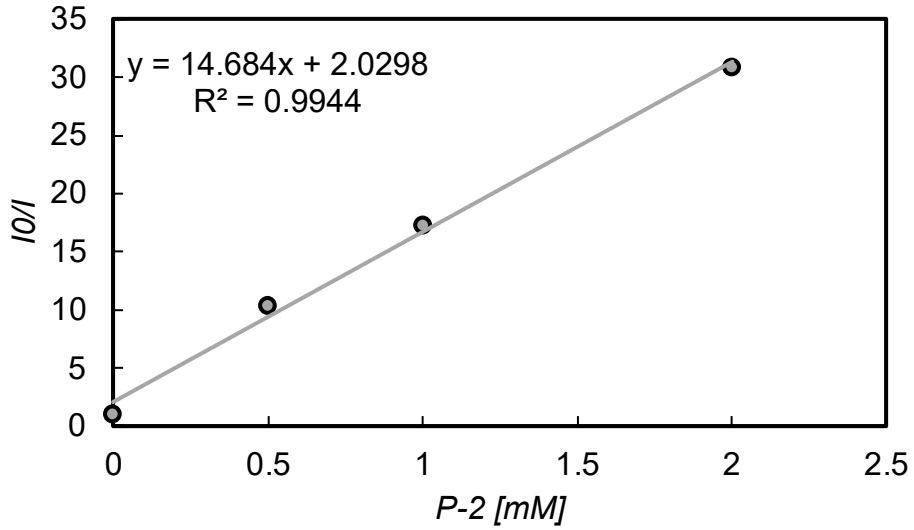
Supplementary Fig. VI-17. Stern–Volmer plot of **PC-2** with varying concentrations of **P-1** in DCM.



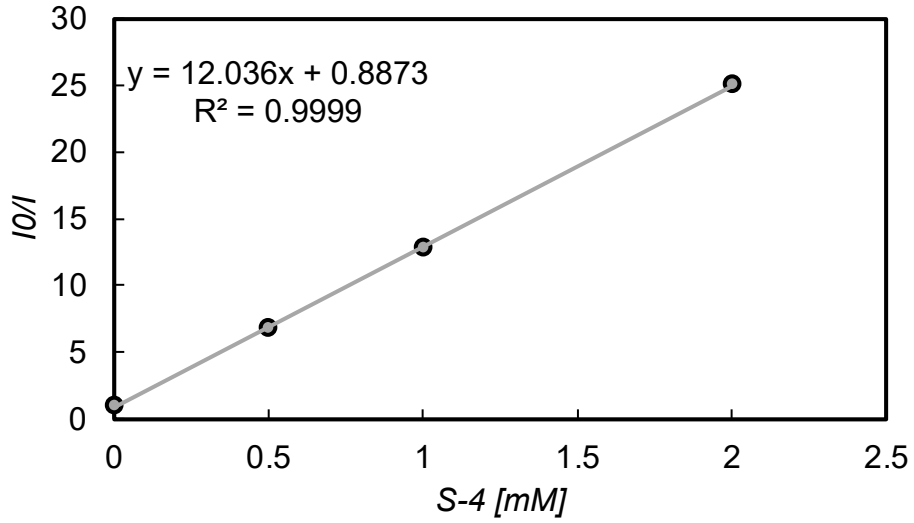
Supplementary Fig. VI-18. Stern–Volmer plot of **PC-2** with varying concentrations of **P-1** in MeCN.



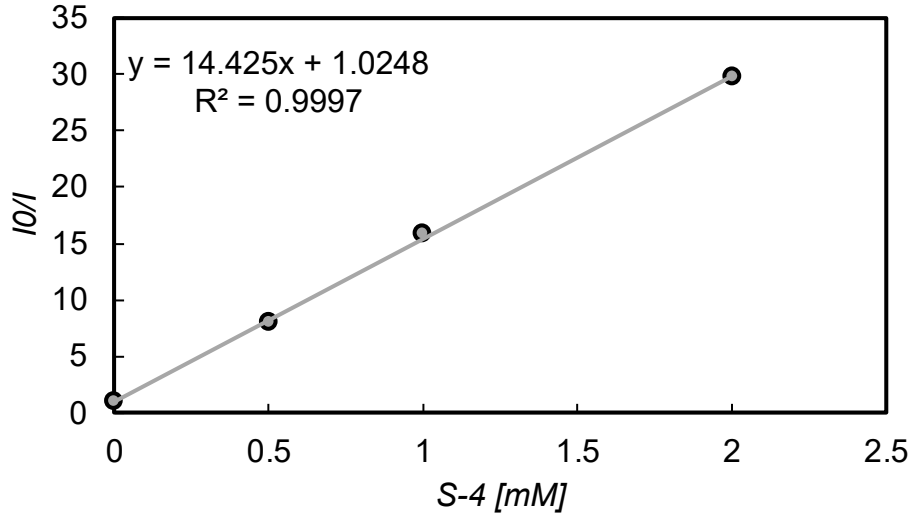
Supplementary Fig. VI-19. Stern–Volmer plot of **PC-2** with varying concentrations of **P-2** in DCM.



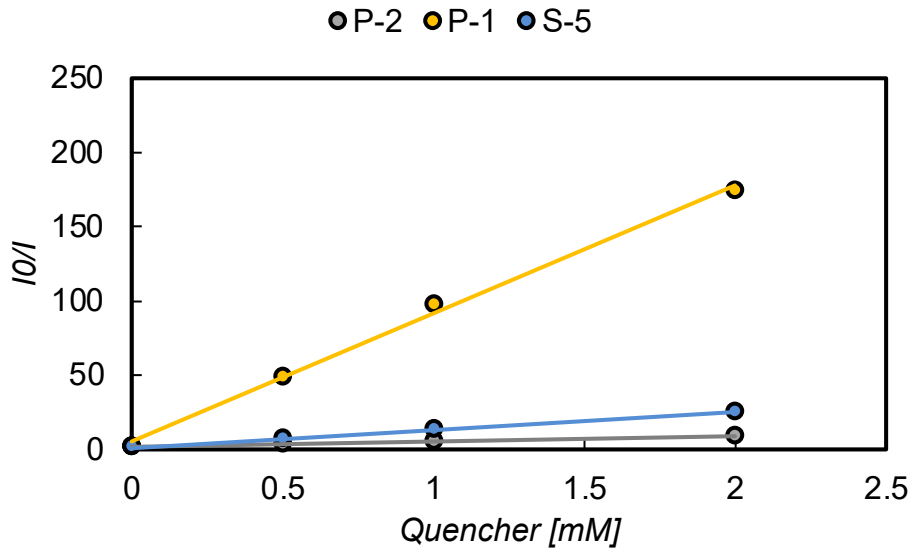
Supplementary Fig. VI-20. Stern–Volmer plot of **PC-2** with varying concentrations of **P-2** in MeCN.



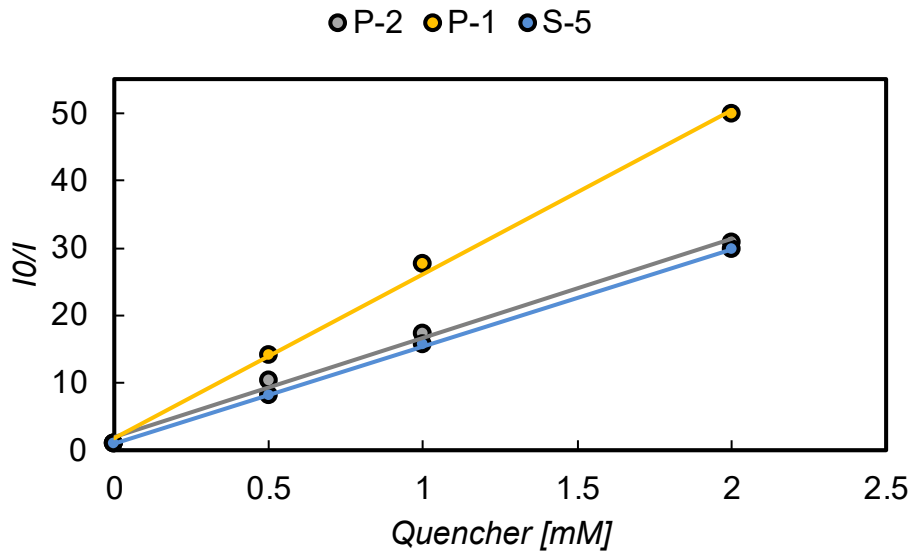
Supplementary Fig. VI-21. Stern–Volmer plot of **PC-2** with varying concentrations of **S4** in DCM with 25 mol% of DIPEA (akin to reaction cond.).



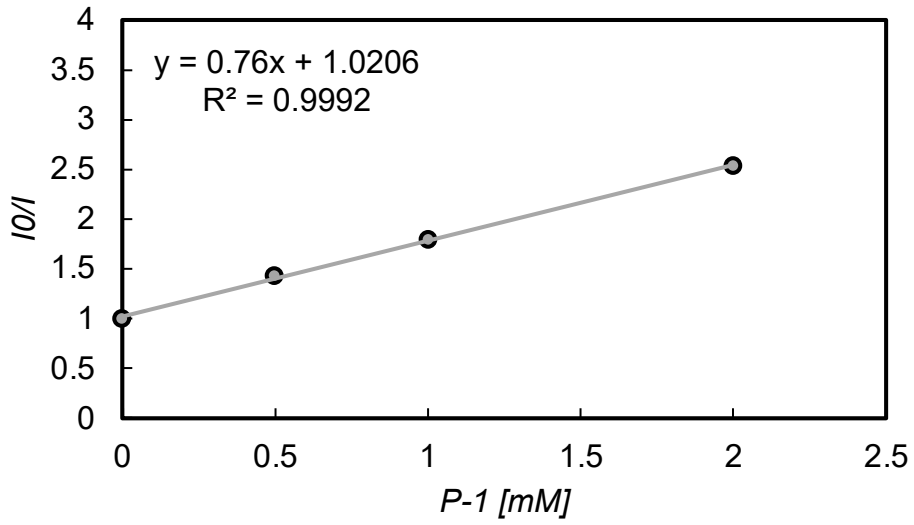
Supplementary Fig. VI-22. Stern–Volmer plot of PC-2 with varying concentrations of **S4** in MeCN with 25 mol% of DIPEA (akin to reaction cond.).



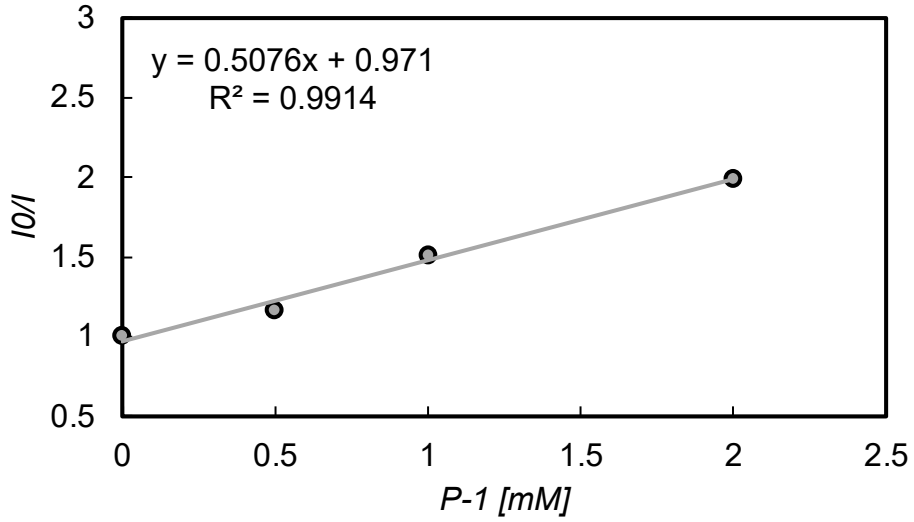
Supplementary Fig. VI-23. Overlayed Stern–Volmer plots of **PC-2** in DCM.



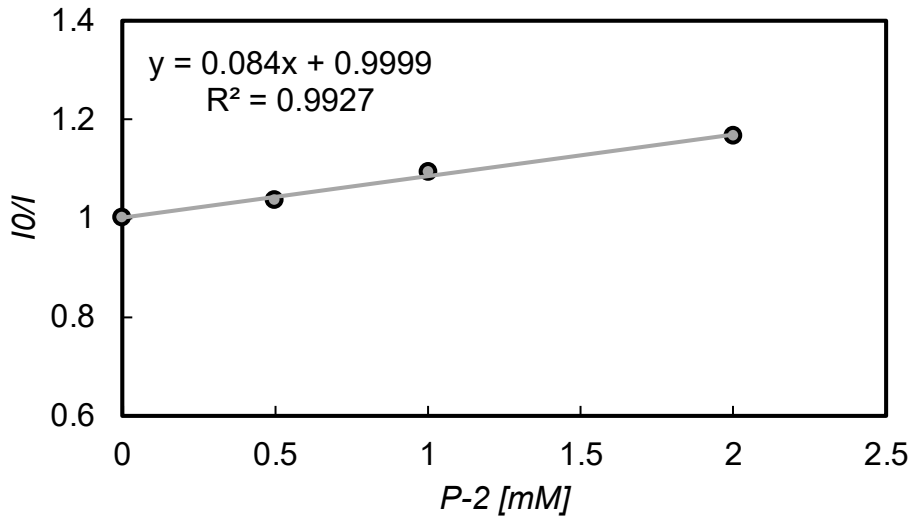
Supplementary Fig. VI-24. Overlayed Stern–Volmer plots of **PC-2** in MeCN.



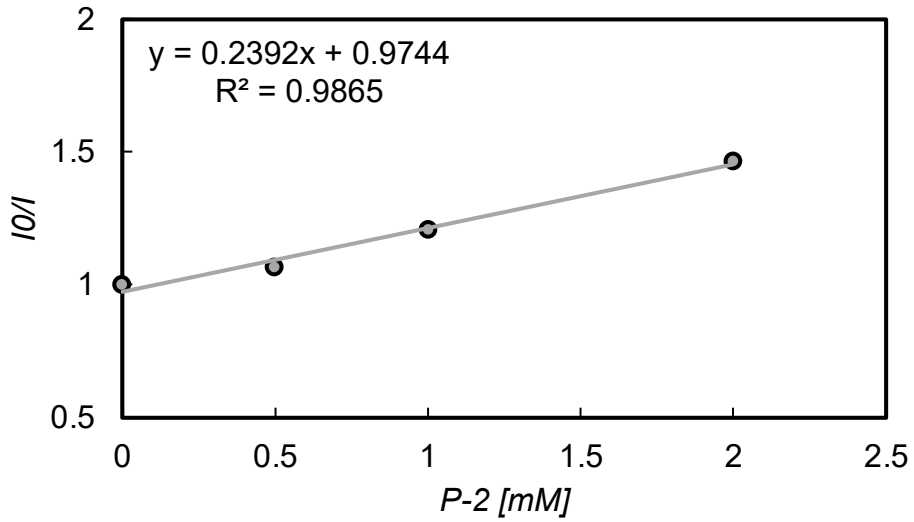
Supplementary Fig. VI-25. Stern–Volmer plot of **PC-7** with varying concentrations of **P-1** in DCM.



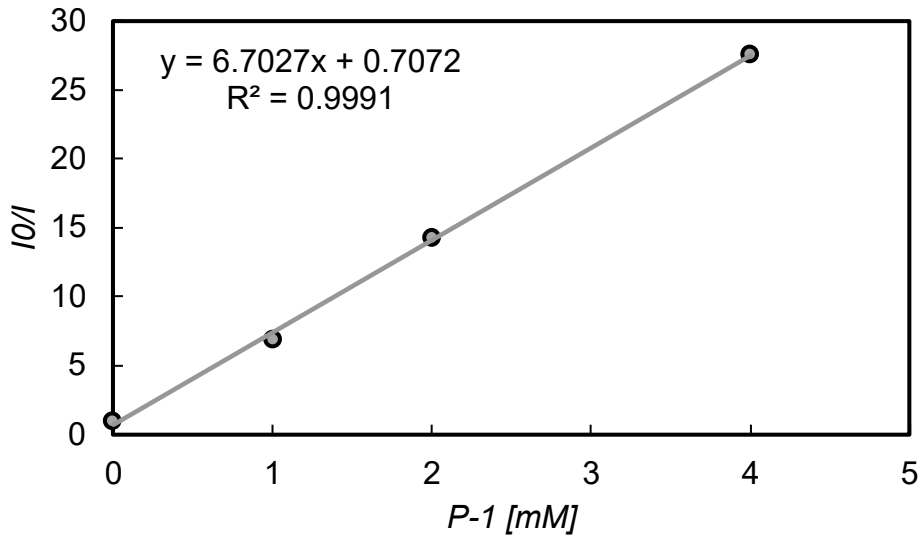
Supplementary Fig. VI-26. Stern–Volmer plot of **PC-7** with varying concentrations of **P-1** in MeCN.



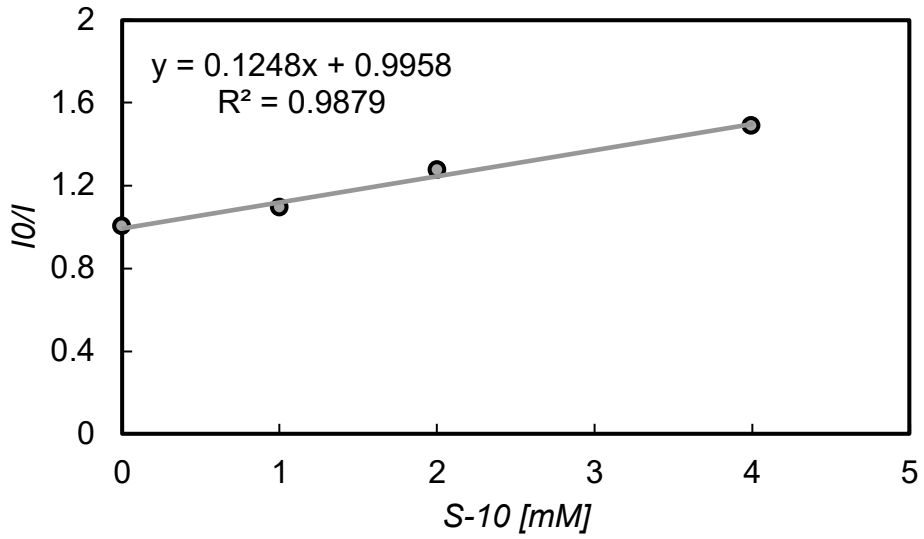
Supplementary Fig. VI-27. Stern–Volmer plot of **PC-7** with varying concentrations of **P-2** in DCM.



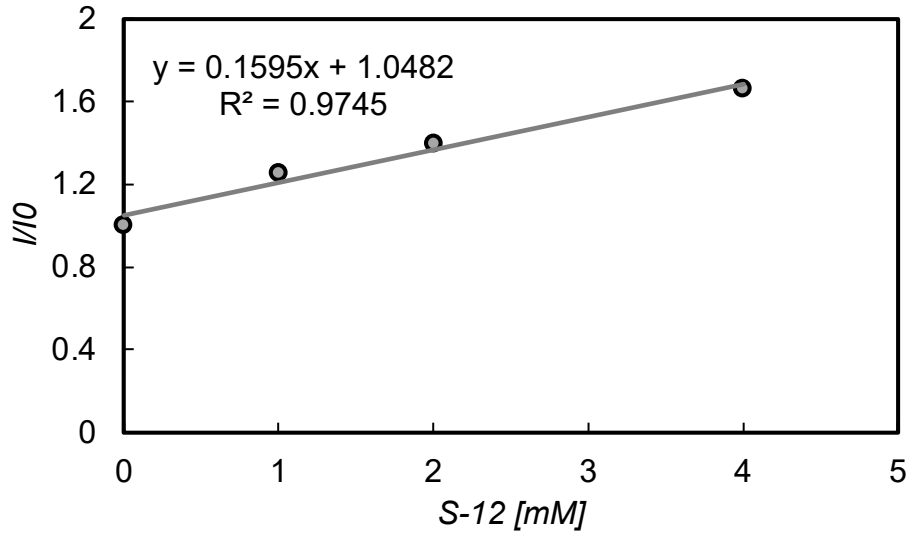
Supplementary Fig. VI-28. Stern–Volmer plot of **PC-7** with varying concentrations of **P-2** in MeCN.



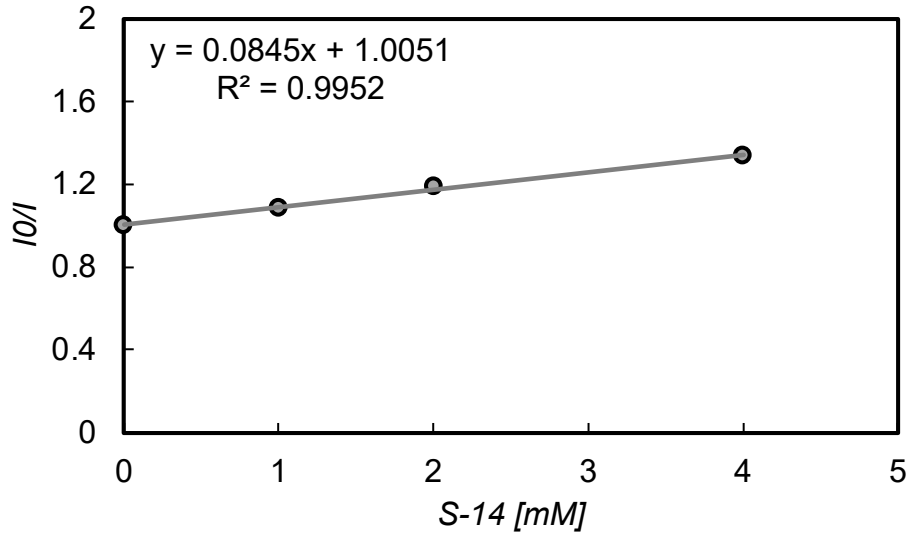
Supplementary Fig. VI-29. Stern–Volmer plot of **PC-11** with varying concentrations of **P-1** in DCM.



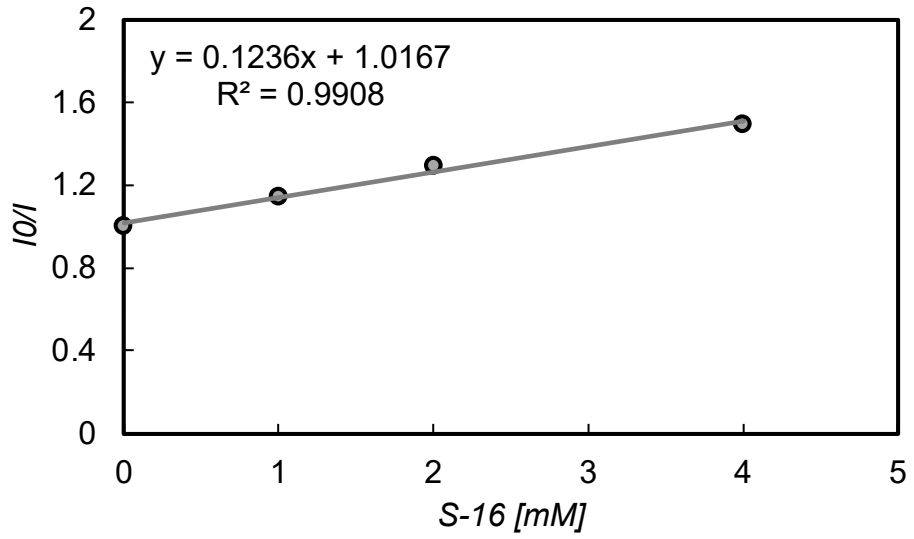
Supplementary Fig. VI-30. Stern–Volmer plot of **PC-11** with varying concentrations of **S10** in DCM.



Supplementary Fig. VI-31. Stern–Volmer plot of **PC-11** with varying concentrations of **S12** in DCM.



Supplementary Fig. VI-32. Stern–Volmer plot of **PC-11** with varying concentrations of **S14** in DCM.



Supplementary Fig. VI-33. Stern–Volmer plot of **PC-11** with varying concentrations of **S16** in DCM.

- ***UV Vis Spectroscopy***

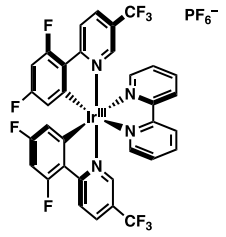
300 µL of a 5.00 mM stock solution of **PC-2** was added to the absorption cuvette. The sample was then diluted to an initial volume of 1,500 µL. An initial scan with no substrate was recorded (A_0). After, substrate was dosed from a 25.0 mM stock solution in 30 µL or 150 µL increments (0.5 equiv. and 2.5 equiv. respectively), the solution was mixed after each addition with a glass pipette, and the absorption spectrum was recorded (A_x). Tabulated for each experiment is (1) the total volume of the solution, (2) the photocatalyst (µmol, concentration in M), (3) the substrate (µmol, concentration in M), and (4) the selected wavelength (λ with substantial change in absorbance). The association constant (K_a) between the photocatalyst and substrate was numerically calculated with BindFit (https://app.supramolecular.org/bindfit/), correcting for dilution, and extracting from the following relationship, where [H] refers to the total concentration of photocatalyst and [G] refers to the total concentration of substrate (M).

$$\Delta A = \varepsilon_{\Delta HG}[\mathrm{HG}], \text{ where}$$

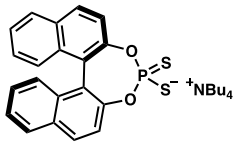
$$[\mathrm{HG}] = \tfrac{1}{2}\left(u - \sqrt{u^2 - 4[\mathrm{H}]_0[\mathrm{G}]_0}\right), \text{ and}$$

$$u = [\mathrm{G}]_0 + [\mathrm{H}]_0 + \frac{1}{K_a}$$

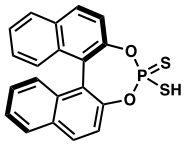
Numerical fitting provides goodness of fit metrics and propagates error within individual trials. Analyses are conducted in duplicate, and the association constant is reported as the numerical average, unless otherwise noted.



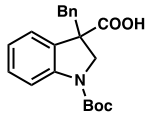
PC-2



P-1



P-2



S4

Supplementary Fig. VI-34. Compounds used in UV-Vis titration experiments.

Supplementary Table VI-71. Overview of association constants (K_a) of **P-1**, **P-2**, and **S4** with **PC-2** in DCM and MeCN. Reported as numerical average across two trials, unless indicated otherwise.

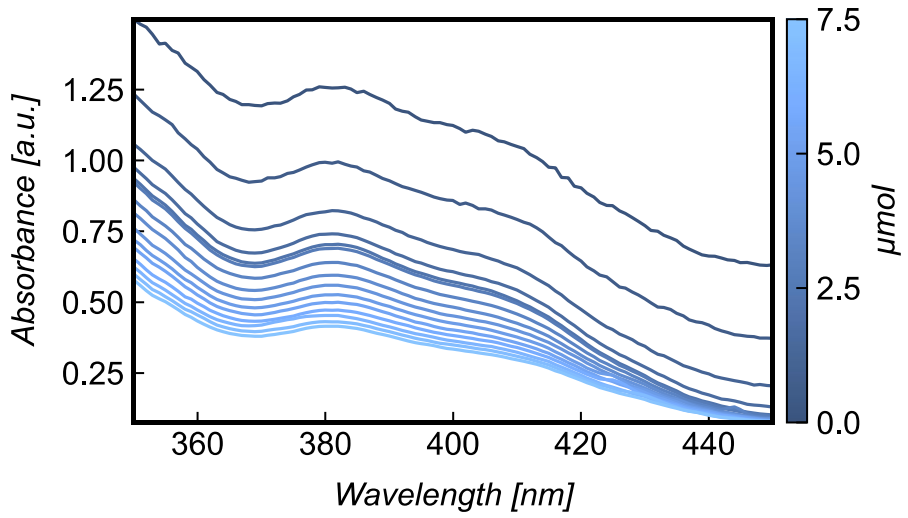
Substrate	K_a in DCM [M^{-1}]	K_a in MeCN [M^{-1}]
P-1	5500*	650
P-2	230	300
S4	100*	270

* single trial

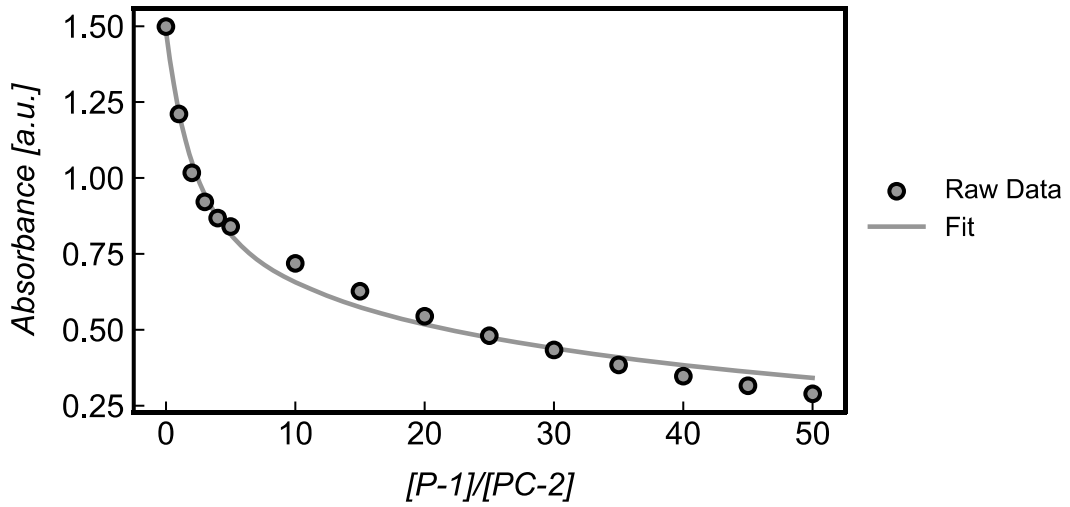
Supplementary Table VI-72. Overview of association constants (K_a) of **P-1**, **P-2**, and **S4** with **PC-6** in DCM and MeCN. Reported as numerical average across two trials, unless indicated otherwise

Substrate	K_a in DCM [M^{-1}]	K_a in MeCN [M^{-1}]
P-1	1900*	320*
P-2	280	300

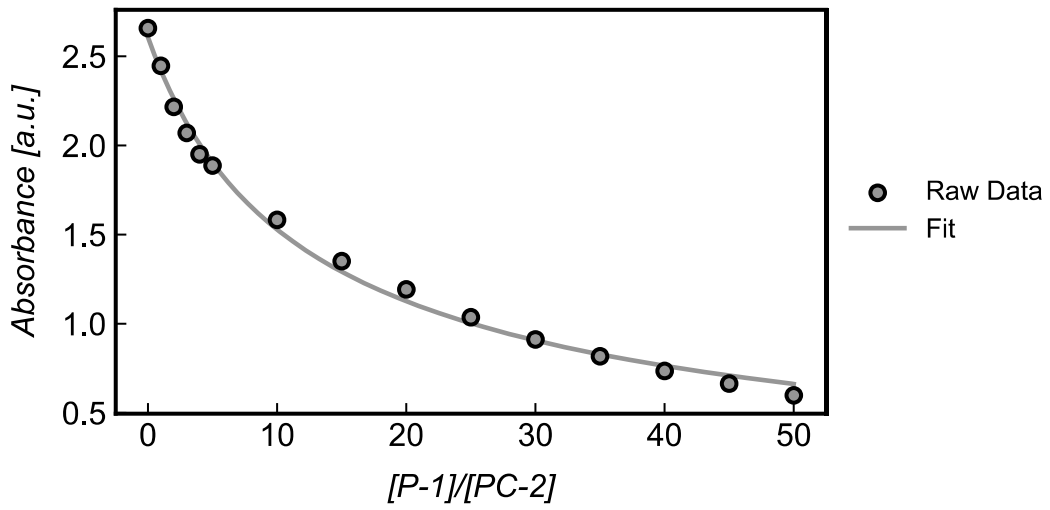
* single trial



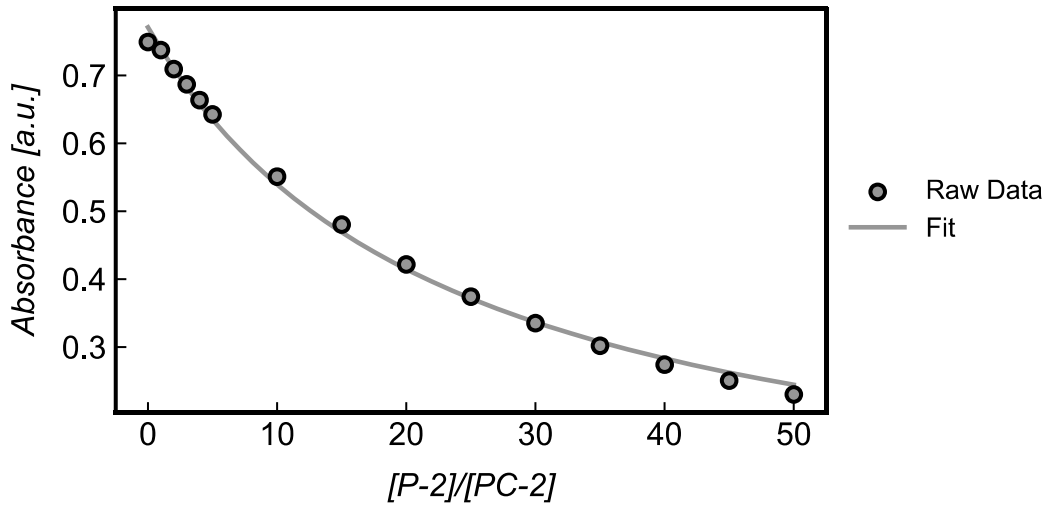
Supplementary Fig. VI-35. Absorption spectra upon the addition of increasing amounts of **P-1** in DCM.



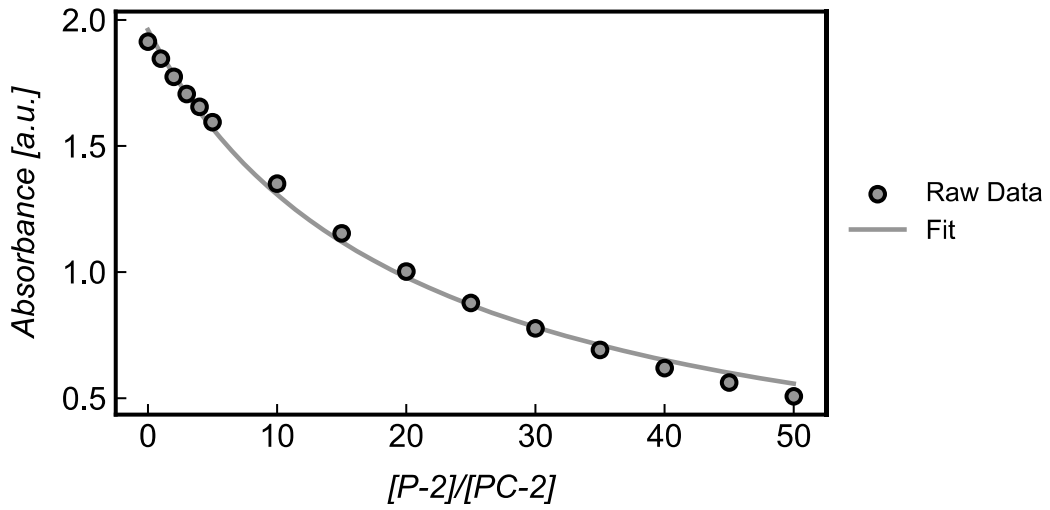
Supplementary Fig. VI-36. Binding fit of **P-1** and **PC-2** based on UV-Vis titration data in DCM. K_a = 5500 M^{-1}, R^2 = 0.96. Total volume: 1500 µL, 0.15 µmol **PC-2**, 0–7.5 µmol **P-1**, absorbance measured at 350 nm.



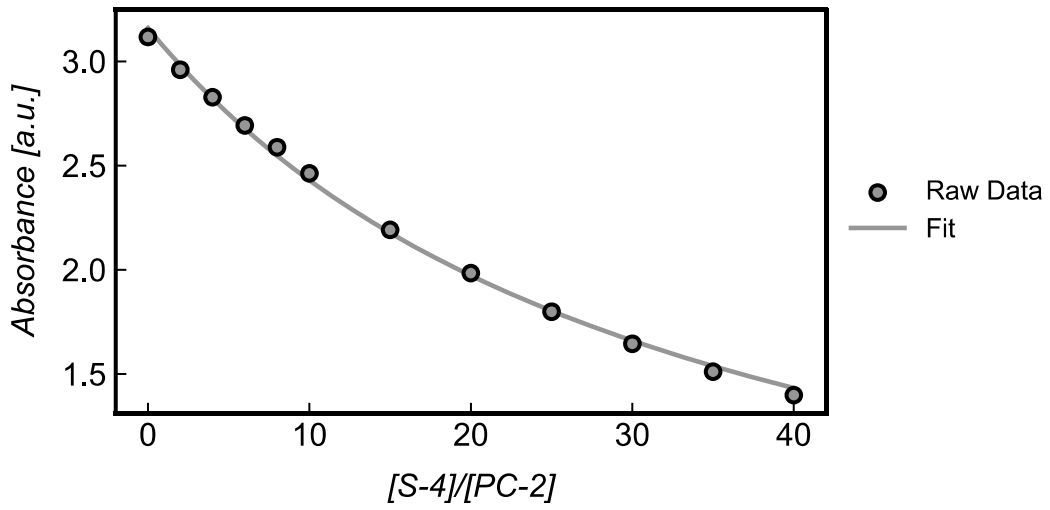
Supplementary Fig. VI-37. Binding fit of **P-1** and **PC-2** based on UV-Vis titration data in MeCN. K_a = 880 M^{-1}, R^2 = 0.98. Total volume: 1500 µL, 0.15 µmol **PC-2**, 0–7.5 µmol **P-1**, absorbance measured at 361 nm.



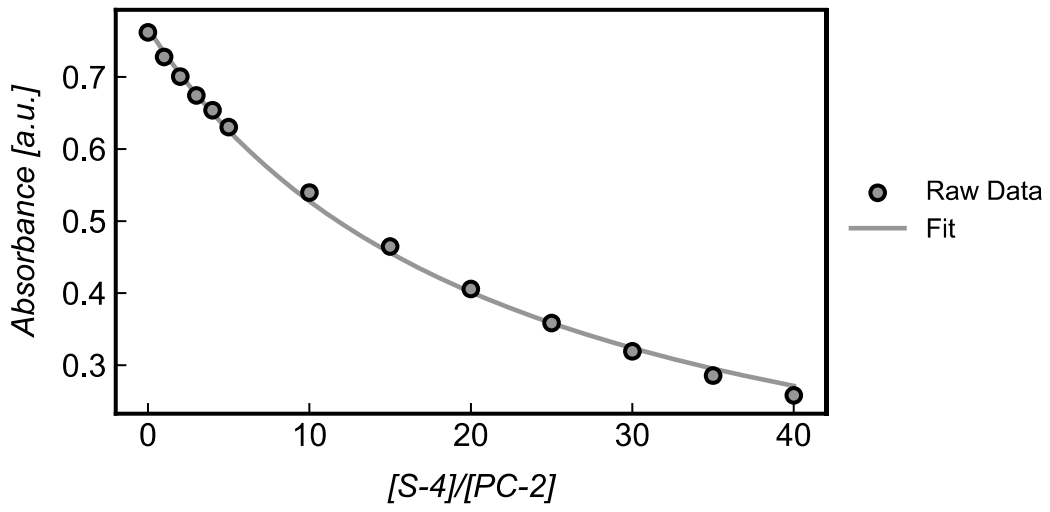
Supplementary Fig. VI-38. Binding fit of **P-2** and **PC-2** based on UV-Vis titration data in DCM. K_a = 230 M^{-1}, R^2 = 0.98. Total volume: 1500 µL, 0.15 µmol **PC-2**, 0–7.5 µmol **P-2**, absorbance measured at 350 nm.



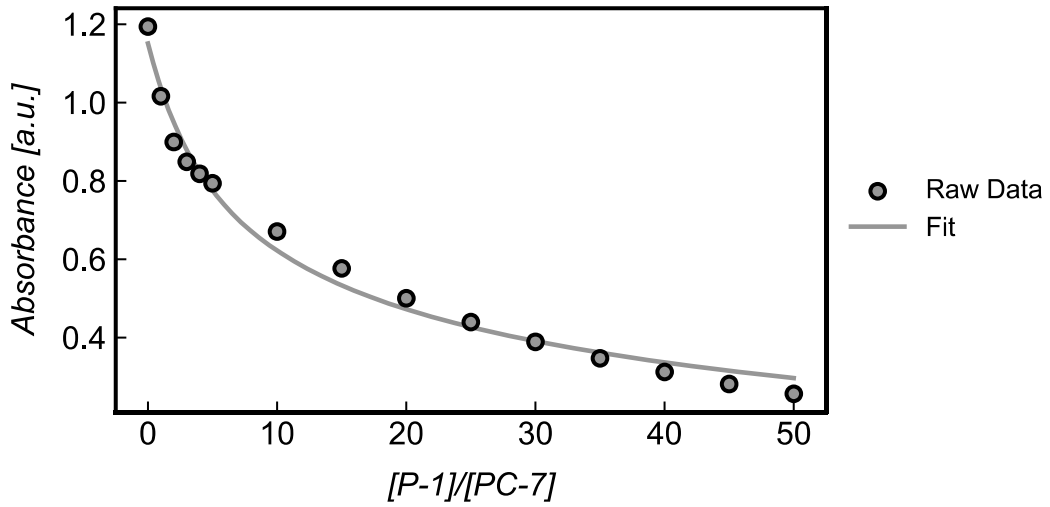
Supplementary Fig. VI-39. Binding fit of **P-2** and **PC-2** based on UV-Vis titration data in MeCN. K_a = 300 M^{-1}, R^2 = 0.98. Total volume: 1500 µL, 0.15 µmol **PC-2**, 0–7.5 µmol **P-2**, absorbance measured at 360 nm.



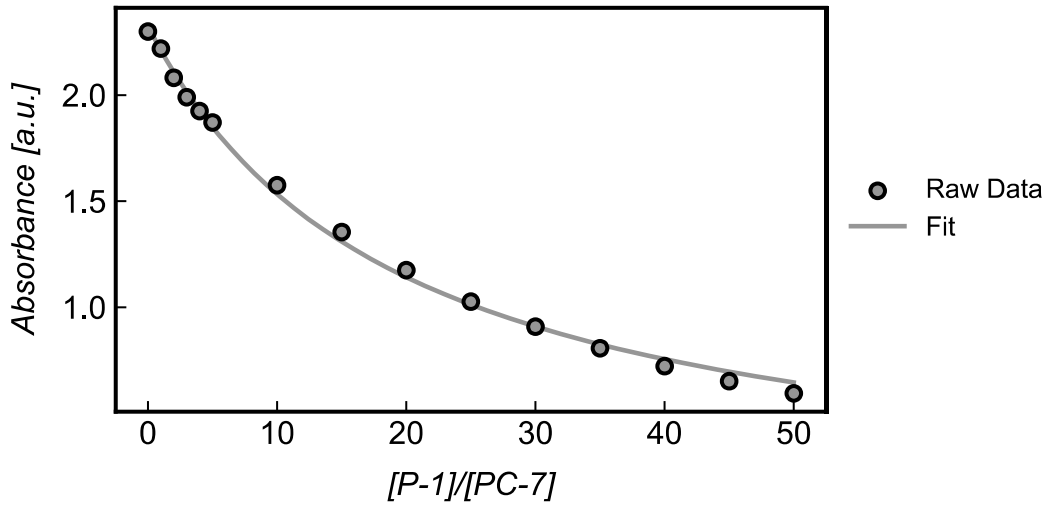
Supplementary Fig. VI-40. Binding fit of **S4** and **PC-2** with 25 mol% DIPEA based on UV-Vis titration data in DCM. K_a = 100 M^{-1}, R^2 = 0.97. Total volume: 1500 µL, 0.15 µmol **PC-2**, 0–7.5 µmol **S4**, absorbance measured at 350 nm.



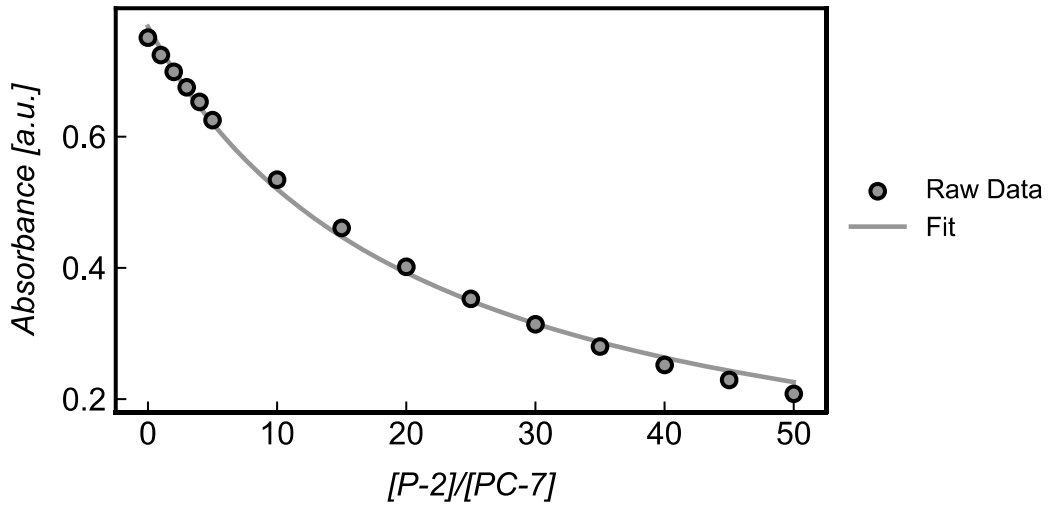
Supplementary Fig. VI-41. Binding fit of **S4** and **PC-2** based on UV-Vis titration data in MeCN. K_a = 260 M^{-1}, R^2 = 0.99. Total volume: 1200 µL, 0.15 µmol **PC-2**, 0–6.0 µmol **S4** with 25 mol% DIPEA, absorbance measured at 350 nm.



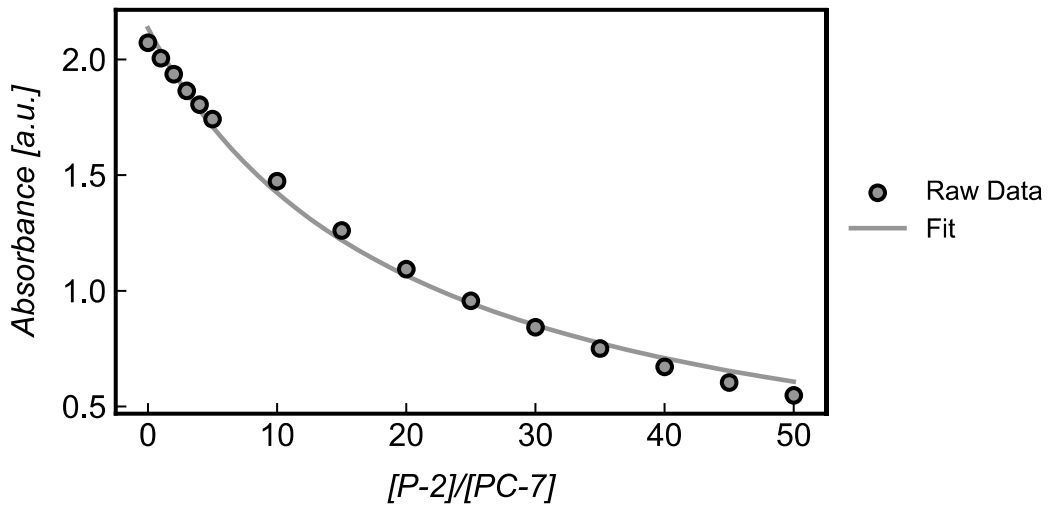
Supplementary Fig. VI-42. Binding fit of **P-1** and **PC-7** based on UV-Vis titration data in DCM. K_a = 1900 M^{-1}, R^2 = 0.96. Total volume: 1500 µL, 0.15 µmol **PC-7**, 0–7.5 µmol **P-1**, absorbance measured at 350 nm.



Supplementary Fig. VI-43. Binding fit of **P-1** and **PC-7** based on UV-Vis titration data in MeCN. K_a = 320 M^{-1}, R^2 = 0.99. Total volume: 1500 µL, 0.15 µmol **PC-7**, 0–7.5 µmol **P-1**, absorbance measured at 361 nm.



Supplementary Fig. VI-44. . Binding fit of **P-2** and **PC-7** based on UV-Vis titration data in DCM. K_a = 280 M^{-1}, R^2 = 0.98. Total volume: 1500 µL, 0.15 µmol **PC-7**, 0–7.5 µmol **P-2**, absorbance measured at 350 nm.



Supplementary Fig. VI-45. Binding fit of **P-2** and **PC-7** based on UV-Vis titration data in MeCN. K_a = 300 M^{-1}, R^2 = 0.98. Total volume: 1500 µL, 0.15 µmol **PC-7**, 0–7.5 µmol **P-2**, absorbance measured at 360 nm.

- ***Job Plot Analysis***

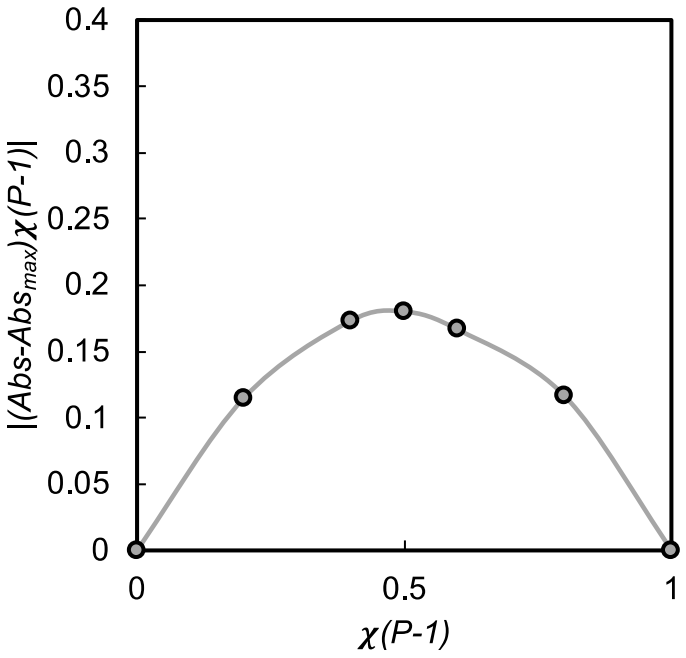
Job plot analysis was conducted using a 0.1 mM MeCN stock solution of either **PC-2** or **PC-6** and a 2.5 mM MeCN stock solution of **P-1**. Individual samples were prepared corresponding to 0, 0.2, 0.4, 0.5, 0.6, 0.8, and 1.0 mol fraction (χ) of **P-1** to photocatalyst. This was accomplished by combining the following solutions volumes:

1000, 800, 600, 500, 400, 200, 0 µL of **PC-2** or **PC-6**

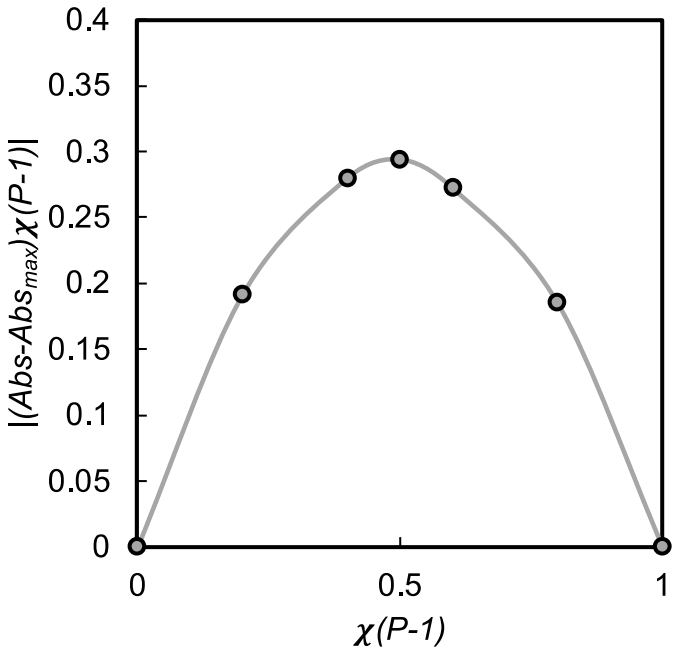
0, 8, 16, 20, 24, 32, 40 µL of **P-1**

Followed by dilution to 2,000 µL total volume. The absorption spectrum was recorded for each sample. Then, the Job metric (γ) was computed using the following formula:

$$\gamma = \left|A_i - A_{\chi=1.0}\right| * \chi$$

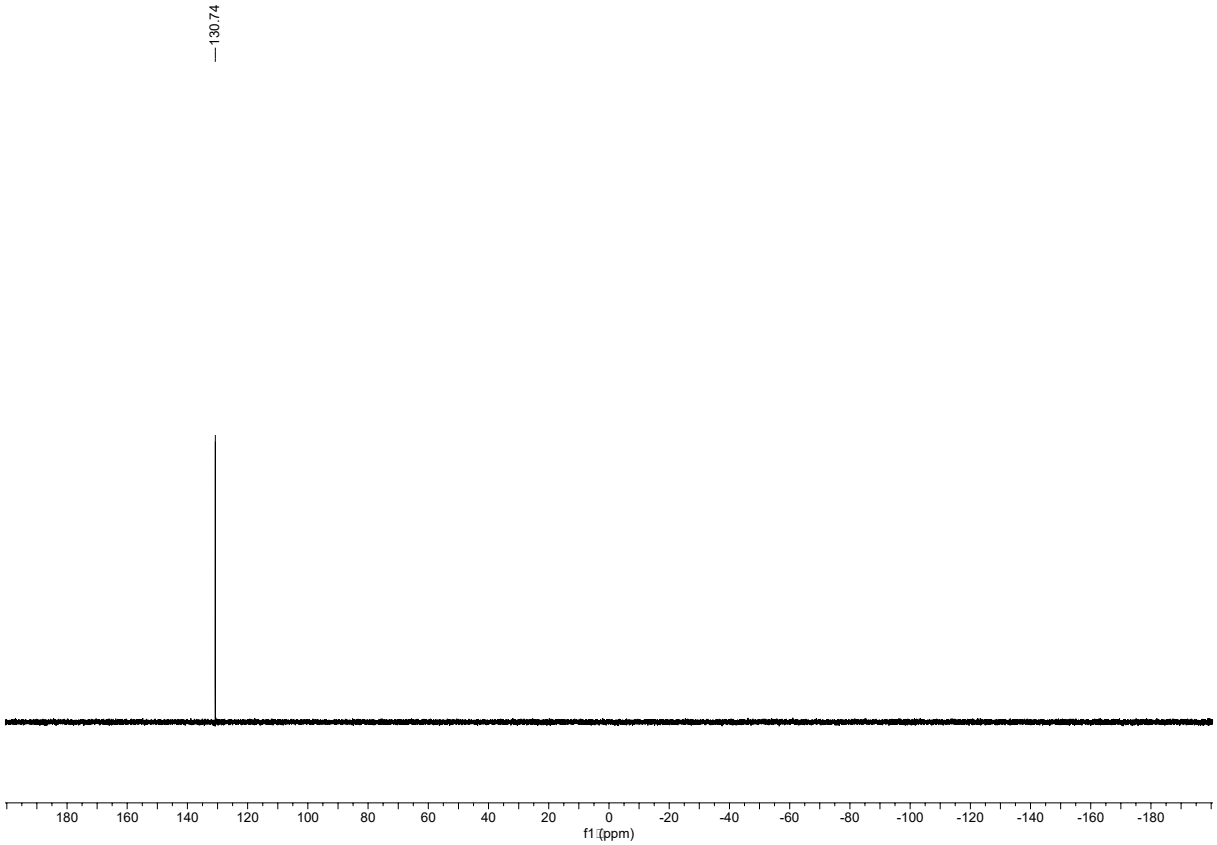


Supplementary Fig. VI-46. Job plot of **PC-2** and **P-1**.

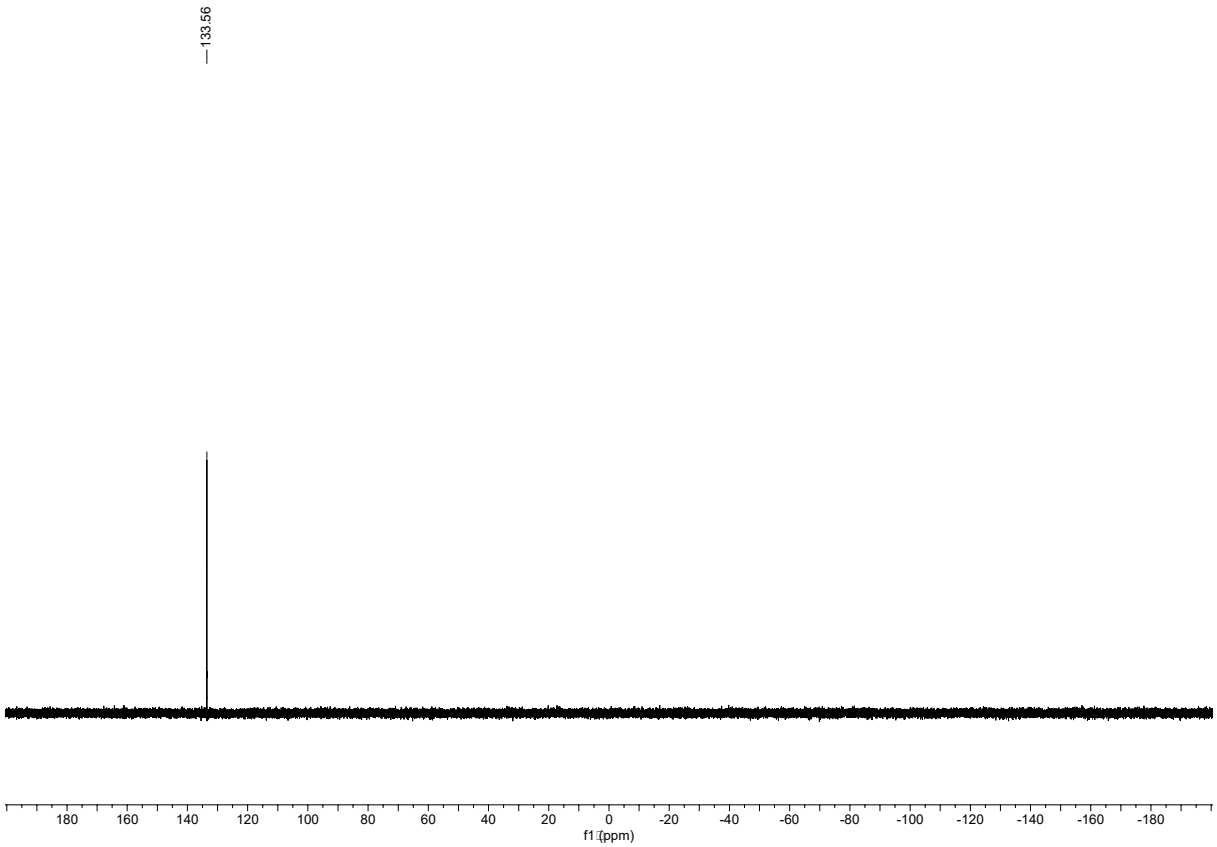


Supplementary Fig. VI-47. Job plot of **PC-7** and **P-1**.

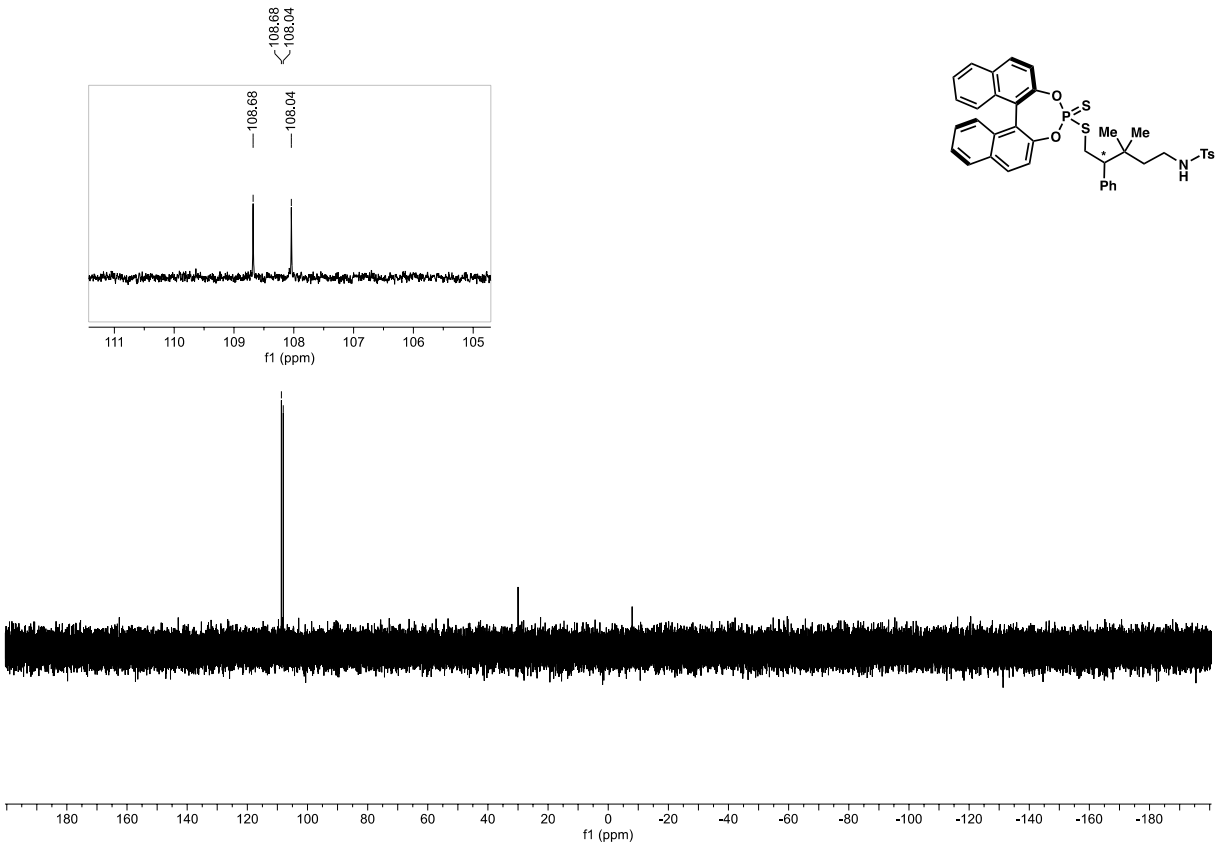
- ***^{31}P NMR Observations***



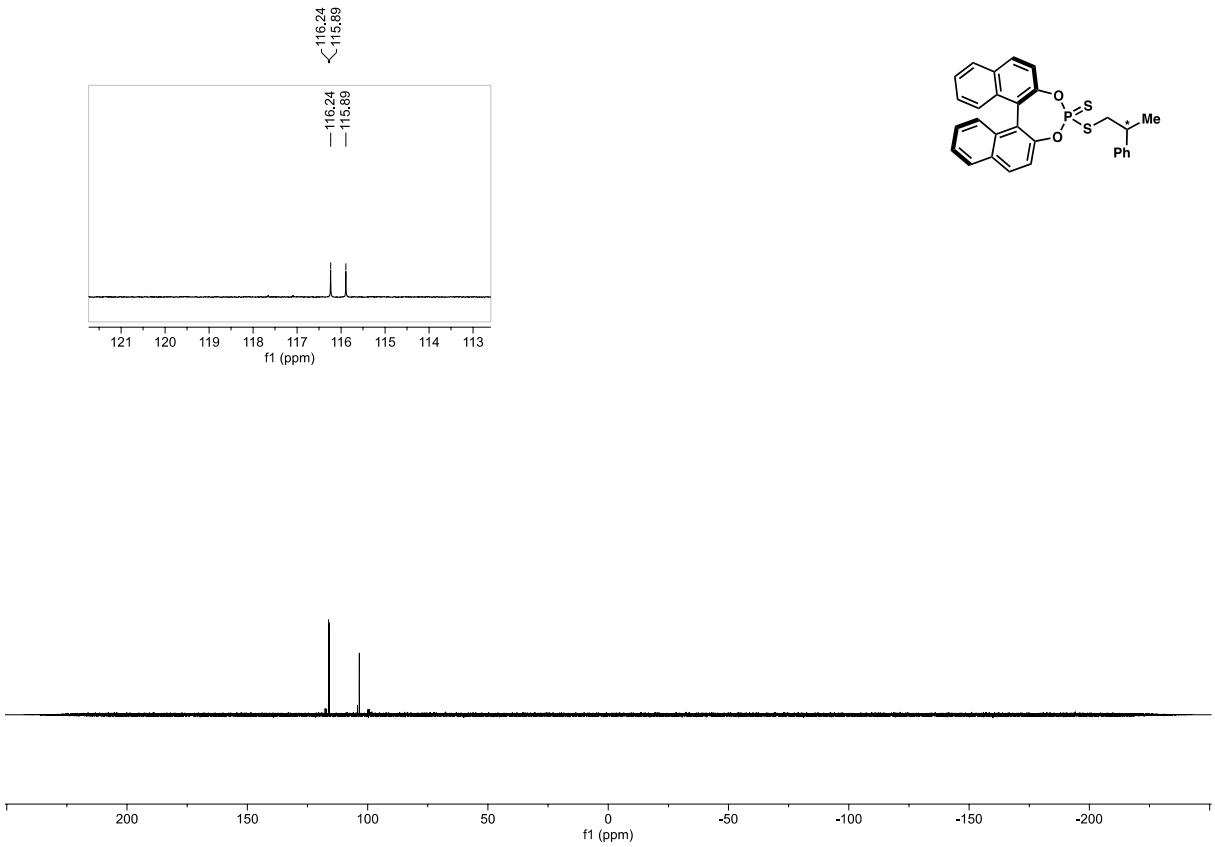
Supplementary Fig. VI-48. ^{31}P NMR of **P-2** in 400 µL DCM with 100 µL $CDCl_3$ and 1.25 equiv. DIPEA (reaction conditions for decarboxylation). $^{31}P\{^1H\}$ (162 MHz, $CDCl_3$) δ 130.74.



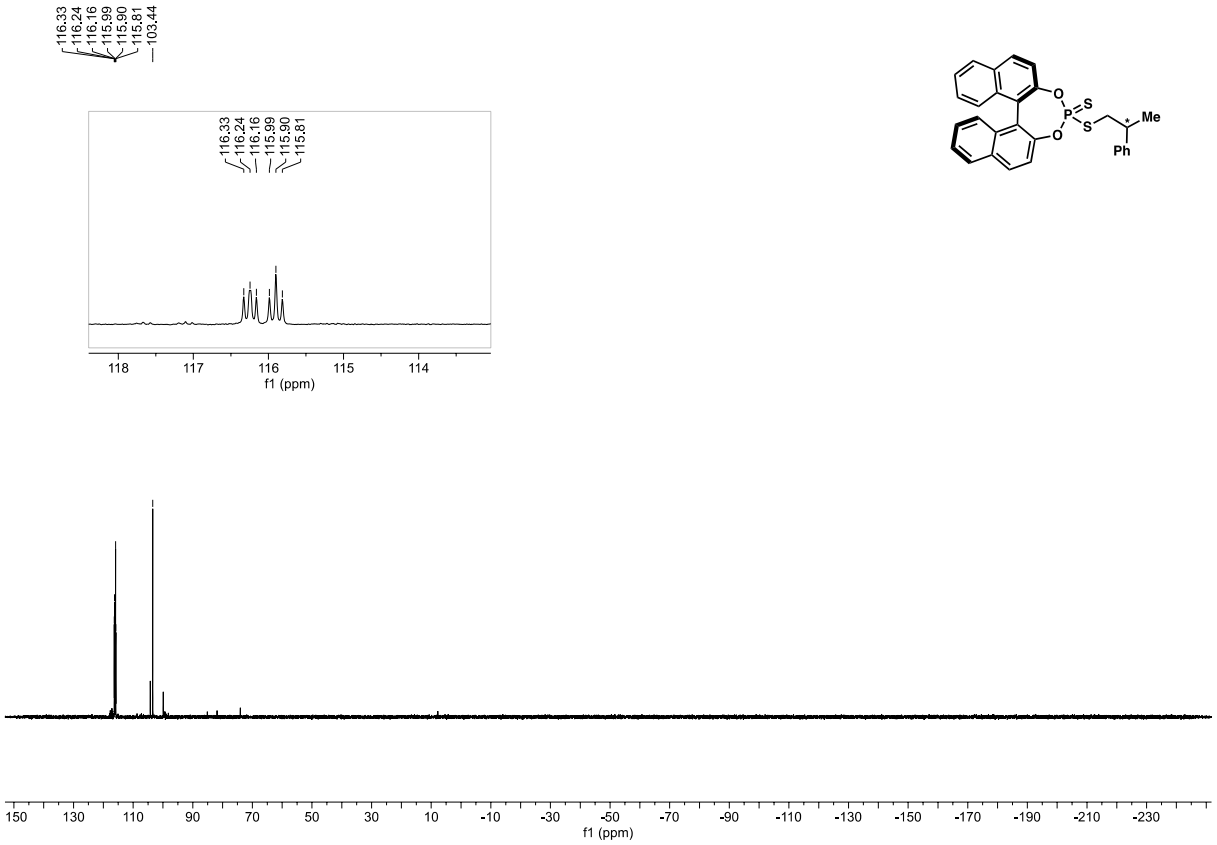
Supplementary Fig. VI-49. ^{31}P NMR of **P-1** in 400 µL MeCN with 100 µL $CDCl_3$ and 1.25 equiv. DIPEA (reaction conditions for decarboxylation). $^{31}P\{^{1}H\}$ NMR (162 MHz, $CDCl_3$) δ 133.56.



Supplementary Fig. VI-50. ^{31}P NMR of hydroamination reaction mixture with ^{1}H decoupling shows two diastereotopic signals. 5% **PC-7**, 20% **P-2**, 10 μmol **S6**, DCM (0.1 M) with 440 nm irradiation. $^{31}P\{^{1}H\}$ NMR (162 MHz, $CDCl_3$) δ 108.68, 108.04.



Supplementary Fig. VI-51. ^{31}P NMR of hydroamination reaction mixture with ^{1}H decoupling shows two diastereotopic signals. 20% **PC-7**, 1 equiv. **P-2**, 10 μmol α-methylstyrene, ***d*$_2$**-DCM (0.1 M) with 440 nm irradiation. $^{31}P\{^{1}H\}$ NMR (162 MHz, $CDCl_3$) δ 116.24, 115.89.

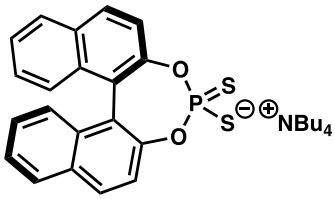


Supplementary Fig. VI-52. ^{31}P NMR of hydroamination reaction mixture shows two diastereotopic signals. 20% PC-7, 1 equiv. P-2, 10 µmol α-methylstyrene, d_2-DCM (0.1 M) with 440 nm irradiation. ^{31}P NMR (202 MHz, $CDCl_3$) δ 116.24 (t, J = 17.0 Hz), 115.90 (t, J = 17.5 Hz). $^3J_{P–H}$ values matched reported literature values. [4]

VI.55 – Synthetic Protocols

- ***Catalysts***

Terabutylammonium dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine-4-thiolate 4-sulfide (P-1)



P-2 (500 mg, 1.3 mmol, 1 equiv.) was dissolved in a minimal amount of DCM. A solution of tetrabutylammonium hydroxide in methanol (0.94 µL, 1.3 mmol, 1.4 M, 1 equiv.) was added dropwise and the solution stirred for 30 min. The solvent was then removed in vacuo. The resulting solid was dissolved in a small amount of DCM and washed with water three times to remove excess tetrabutylammonium hydroxide. After removal of the solvent, the product was dried under vacuum overnight to give **P-1** as an off-white solid.

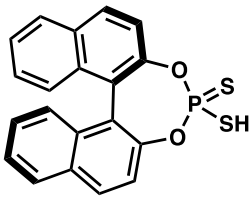
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.91 (d, *J* = 8.7 Hz, 2H), 7.87 (dd, *J* = 8.2, 1.5 Hz, 2H), 7.58 (dd, *J* = 8.9, 1.4 Hz, 2H), 7.41 (d, *J* = 8.7 Hz, 2H), 7.37 (t, *J* = 8.2 Hz, 2H), 7.24 – 7.09 (m, 2H), 3.18 – 3.04 (m, 8H), 1.56 – 1.45 (m, 8H), 1.33 (h, *J* = 7.3 Hz, 8H), 0.93 (t, *J* = 7.3 Hz, 12H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 132.76.

^{13}C NMR (126 MHz, $CDCl_3$) δ 150.25, 150.15, 132.96, 131.33, 129.40, 128.37, 127.35, 125.64, 124.59, 123.76, 123.74, 123.62, 123.59, 58.68, 24.13, 19.79, 13.85.

HRMS (ESI–): *m/z* calculated for $C_{20}H_{12}O_2PS_2^-$ $[M-H]^-$ = 363.0250,
Observed $[M-H]^-$ = 363.0247.

4-mercaptodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-sulfide (P-2)



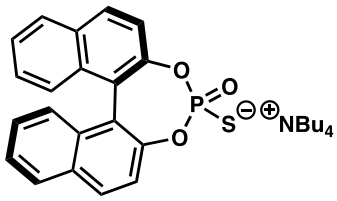
P-2 was prepared following a procedure from the literature.[5]

(R)-BINOL (500 mg, 1.75 mmol, 1 equiv.) and P_2S_5 (194 mg, 0.87 mmol, 0.5 equiv.) were placed into a flame-dried reaction tube under nitrogen. Xylenes (8.7 mL, 0.2 M) were added and the reaction mixture heated to 150 ˚C for 2 hours. After cooling to room temperature, hexanes were added. The resulting material was filtered and washed with cold hexanes three times to give **P-1** as a white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.06 (d, J = 8.9 Hz, 2H), 7.98 (dd, J = 8.2, 1.2 Hz, 2H), 7.58 (dd, J = 8.8, 1.4 Hz, 2H), 7.52 (ddt, J = 8.0, 6.8, 1.1 Hz, 2H), 7.42 (d, J = 8.5 Hz, 2H), 7.33 (ddd, J = 8.4, 6.7, 1.3 Hz, 2H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 97.12.

Tetrabutylammonium dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine-4-thiolate 4-oxide (P-3)



P-4 (100 mg, 0.25 mmol, 1 equiv.) was dissolved in a minimal amount of DCM. A solution of tetrabutylammonium hydroxide in methanol (196 µL, 0.25 mmol, 1.4 M, 1 equiv.) was added dropwise and the solution stirred for 30 min. The solvent was then removed by evaporation under vacuum. The resulting solid was dissolved in a small amount of DCM and washed with water three times to remove excess tetrabutylammonium hydroxide. After removal of the solvent, the product was dried under vacuum overnight to give **P-3** as an off-white solid.

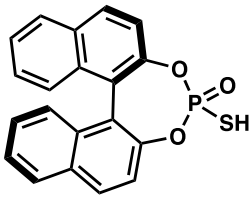
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.88 (dd, *J* = 11.3, 8.2 Hz, 4H), 7.55 (d, *J* = 8.7 Hz, 2H), 7.37 (dd, *J* = 11.7, 8.3 Hz, 4H), 7.23 – 7.16 (m, 2H), 3.03 – 2.87 (m, 8H), 1.35 (m, *J* = 16.0 Hz, 8H), 1.23 (m, 8H), 0.86 (t, *J* = 7.2 Hz, 12H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 65.89.

^{13}C NMR (126 MHz, $CDCl_3$) δ 150.24, 150.13, 132.88, 131.25, 129.39, 128.38, 127.24, 125.69, 124.64, 123.79, 123.52, 58.44, 24.01, 19.71, 13.82.

HRMS (ESI–): *m/z* calculated for $C_{20}H_{12}O_3PS^-$ $[M-H]^-$ = 363.0250,
Observed $[M-H]^-$ = 363.0227.

4-mercaptodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-oxide (P-4)



P-4 was prepared following a modified procedure from the literature.[6]

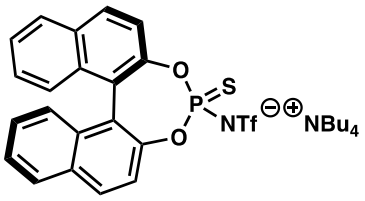
(*R*)-BINOL (2.0 g, 7.0 mmol, 1 equiv.), triethylamine (2.9 mL, 21 mmol, 3 equiv.), and dry DCM (35 mL, 0.2 M) were added to a flame dried reaction tube and placed under nitrogen. Thiophosphoryl chloride (1.4 mL, 2 equiv.) was added dropwise and the reaction mixture was stirred at room temperature for overnight. After, the solution was quenched with sat. $NaHCO_3$ (aq.) and extracted with DCM. The organic fractions were pooled and concentrated to give the phosphorochloridothioate as a white solid.

The phosphorochloridothioate (1.0 g, 2.6 mmol, 1 equiv.) was placed in a reaction tube with pyridine (5 mL). Water (5 mL) was added and the solution was stirred at 90 ˚C for 2 hours. After cooling to room temperature, the reaction mixture was extracted with DCM and washed with water. The solvent was removed in vacuo to give **P-4** as a white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.03 (dd, J = 17.2, 8.8 Hz, 2H), 7.96 (dd, J = 8.3, 5.9 Hz, 2H), 7.56 (dd, J = 22.2, 8.9 Hz, 2H), 7.49 (ddd, J = 8.3, 6.8, 1.3 Hz, 2H), 7.41 (dd, J = 8.7, 4.8 Hz, 2H), 7.31 (ddt, J = 8.3, 6.7, 1.4 Hz, 2H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 70.54.

Tetrabutylammonium (4-sulfidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)((trifluoromethyl)sulfonyl)amide (P-5)



P-6 (248 mg, 0.500 mmol, 1 equiv.) was dissolved in a minimal amount of DCM. A solution of tetrabutylammonium hydroxide in methanol (357 µL, 0.500 mmol, 1.4 M, 1 equiv.) was added dropwise and the solution stirred for 30 min. The solvent was then removed by evaporation under vacuum. The resulting solid was dissolved in a small amount of DCM and washed with water three times to remove excess tetrabutylammonium hydroxide. After removal of the solvent in vacuo, the product was dried under vacuum overnight to give **P-5** as an off-white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.23 – 7.78 (m, 4H), 7.59 (ddd, *J* = 25.1, 8.8, 1.2 Hz, 2H), 7.45 – 7.36 (m, 2H), 7.24 (ddt, *J* = 8.3, 6.7, 1.5 Hz, 2H), 3.08 (dt, *J* = 8.8, 5.5 Hz, 8H), 1.50 (ddd, *J* = 16.0, 8.6, 4.4 Hz, 8H), 1.33 (q, *J* = 7.4 Hz, 8H), 0.93 (t, *J* = 7.3 Hz, 12H).

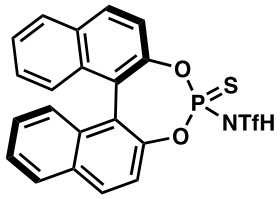
^{31}P NMR (202 MHz, $CDCl_3$) δ 66.46.

^{13}C NMR (126 MHz, $CDCl_3$) δ 149.40, 149.29, 147.58, 132.69, 132.49, 131.60, 131.41, 130.04, 128.35, 128.31, 127.22, 127.13, 125.94, 125.91, 125.02, 124.93, 123.06, 122.66, 122.46, 58.58, 23.86, 19.60, 13.61.

^{19}F NMR (471 MHz, $CDCl_3$) δ -78.66.

HRMS (ESI–): *m/z* calculated for $C_{21}H_{12}F_3NO_4PS_2^-$ $[M-H]^-$ = 493.9903,
Observed $[M-H]^-$ = 493.9880.

1,1,1-trifluoro-N-(4-sulfidodinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepin-4-yl)methanesulfonamide (P-6)



P-6 was prepared following a modified procedure from the literature.[7]

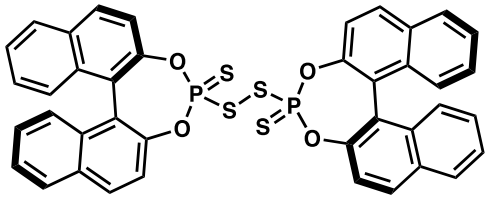
(*R*)-BINOL (5.00 g, 17.5 mmol, 1 equiv.), triethylamine (7.30 mL, 52.4 mmol, 3 equiv.), and dry DCM (100 mL, 0.2 M) were added to a flame dried reaction tube and placed under nitrogen. Thiophosphoryl chloride (1.95 mL, 1.1 equiv.) was added dropwise and the reaction mixture was stirred at room temperature overnight. After, the solution was quenched with sat. $NaHCO_3$ (aq.) and extracted with DCM. The organic fractions were pooled and concentrated to give the phosphorochloridothioate as a white solid.

The phosphorochloridothioate (1.0 g, 2.6 mmol, 1 equiv.) was dissolved toluene (5 mL, 0.5 M). $TfNH_2$ (0.78 g, 5.2 mmol, 2 equiv.), and DIPEA (1.9 mL, 10 mmol, 4 equiv.) were added and the mixture was stirred at 110 ˚C overnight. After, the mixture was cooled to. 0 ˚C and conc. HCl was added carefully until pH = 1. Water was added and the aqueous phase was extracted with EtOAc, the combined extracted dried with Na_2SO_4 and the solvent removed in vacuo. The crude material was purified by silica gel column chromatography and the combined product fractions were washed with aq. HCl. Removal of the solvent in vacuo gave **P-6** as a light brown solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.08 (t, *J* = 8.9 Hz, 2H), 8.02 – 7.94 (m, 2H), 7.59 (ddd, *J* = 9.9, 8.8, 1.2 Hz, 2H), 7.52 (ddd, *J* = 8.1, 6.5, 1.3 Hz, 2H), 7.40 – 7.35 (m, 2H), 7.35 – 7.30 (m, 2H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 59.51.

4,4'-disulfanediylbis(dinaphtho[2,1-d:1',2'-f][1,3,2]dioxaphosphepine 4-sulfide) (P-7)



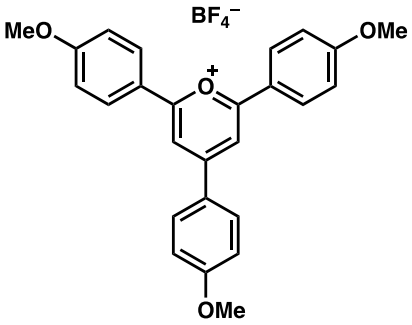
P-7 was prepared following a procedure from the literature.[8]

To a 20 mL scintillation vial was added P-2 (200 mg, 0.53 mmol, 1 equiv.) iodine (67 mg, 0.26 mmol, 0.5 equiv.) and potassium carbonate (218 mg, 1.58 mmol, 3 equiv.). DCM (2.1 mL, 0.25 M) was added, and the suspension was stirred overnight. The solution was then filtered, concentrated *in vacuo* and the crude solid containing the disulfide purified on silica (0-10% Hex/EtOAc) to furnish **P-7** as a fine yellow solid.

1H NMR (500 MHz, CDCl3) δ 8.08 (d, J = 8.9 Hz, 2H), 7.99 (d, J = 8.4 Hz, 2H), 7.91 (d, J = 8.1 Hz, 2H), 7.81 (d, J = 9.0 Hz, 2H), 7.61 (dd, J = 8.9, 4.3 Hz, 4H), 7.55 – 7.47 (m, 4H), 7.43 – 7.28 (m, 8H).

^{31}P NMR (202 MHz, $CDCl_3$) δ 103.03.

2,4,6-tris(4-methoxyphenyl)pyrylium tetrafluoroborate (PC-11)

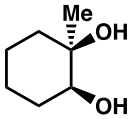


PC-11 was prepared following a procedure from the literature.[9] 4-methoxybenzaldehyde (3.0 mL, 25 mmol, 1 equiv.) and 4-methoxyacetophenone (7.5 g, 50 mmol, 2 equiv.) were dissolved in a minimal amount of toluene (2.5 mL). BF_3•OEt_2 (7.4 mL, 60 mmol, 2.4 equiv.) was added dropwise After heating the solution at 100 ˚C for 2 h, the solvent was removed in vacuo. The resulting solid was dissolved in acetone and precipitated by addition of diethyl ether. The crude material was triturated with acetone to give **PC-11** as a bright orange solid.

^{1}H NMR (400 MHz, DMSO-*d6*) δ 8.76 (s, 2H), 8.57 (d, J = 8.7 Hz, 2H), 8.46 (d, J = 8.7 Hz, 4H), 7.37 – 7.20 (m, 6H), 3.94 (s, 3H), 3.92 (s, 6H).

- ***Substrates***

cis-1-methylcyclohexane-1,2-diol (*cis*-S1a)

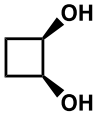


cis-**S1a** was prepared following a procedure from the literature.[2]

1-Methylcyclohezene (1.0 mL, 8.4 mmol, 1 equiv.), NMO (50% in water, 2.6 mL, 13 mmol, 1.5 equiv.) and OsO_4 (4% in water, 0.1 mL, 0.08 mmol, 0.01 equiv.) were combined in a solution of acetone (23 mL) and water (4.7 mL) and stirred overnight at room temperature. The reaction mixture was then quenched with sat. Na_2SO_3 (aq.) and extracted with DCM. The crude product was purified by silica gel column chromatography to give *cis*-**S1a** as a white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 3.40 (dd, *J* = 9.3, 4.0 Hz, 1H), 1.81 – 1.73 (m, 1H), 1.72–1.64 (m, 2H), 1.62–1.49 (m, 2H), 1.45–1.28 (m, 3H), 1.24 (s, 3H).

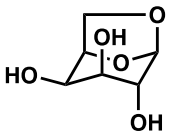
cis-cyclobutane-1,2-diol (*cis*-S1b)



cis-**S1b** was prepared following a reported procedure.[10] To 1,2-Bis(trimethylsiloxy)cyclobutene (1.0 mL, 3.0 mL, 1 equiv.) in THF (30 mL, 0.13 M) was added Pd/C (5%, 830 mg, 0.39 mmol, 0.1 equiv.) under nitrogen atmosphere. The suspension was purged with hydrogen, then placed under hydrogen atmosphere using a balloon and stirred at room temperature overnight. After removal of the catalyst using a short pad of celite, the solvent was removed in vacuo and the crude material purified by silica gel column chromatography to give *cis*-**S1b** as a colorless oil.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.33 – 4.23 (m, 2H), 2.17 – 2.06 (m, 2H), 1.93 – 1.85 (m, 2H).

D-anhydrogalactose (*cis*-S1c)

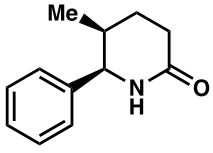


cis-**S1c** was prepared following a procedure from the literature.[11] D-Galactose pentaacetate (1.0 g, 2.6 mmol, 1 equiv.), phenol (0.84 g, 9.0 mmol, 3.5 equiv.) and *p*-toluenesulfonic acid monohydrate (15 mg, 0.077 mmol, 0.03 equiv.) were combined and stirred t 95 ˚C at 15 mmHg for 2 h. After cooling to room temperature, DCM was added and the solution was washed with NaOH (aq.) and water. After drying the combined organic fractions with Na_2SO_4 and removal of the solvent in vacuo, the intermediate 1-phenyl-tetraacetyl-β-D-galactopyranoside was recrystallized from ethanol.

In a next step, NaOH pellets (1.47 g, 36.9 mmol, 15 equiv.) were dissolved in water (30 mL, 0.08 M) were added and 1-phenyl-tetraacetyl-β-D-galactopyranoside was added. The mixture was heated to 90˚C and stirred overnight. After, the solution was cooled to 0 ˚C and acidified to pH = 3 with aq. HCl. After removal of the solvent under reduced pressure, the solid was washed with 1:1 MeOH/DCM and the filtrate was collected. After removal of the solvent in vacuo, the crude material was purified by silica gel column chromatography (2:1 Acetone/DCM) to give *cis*-**S1c** as an off-white solid.

^{1}H NMR (500 MHz, MeOD) δ 5.24 (s, 1H), 4.33 (dd, J = 7.0, 0.8 Hz, 1H), 4.29 (td, J = 4.8, 1.3 Hz, 1H), 3.97 – 3.91 (m, 1H), 3.83 (dd, J = 5.1, 1.5 Hz, 1H), 3.65 (t, *J* = 1.8 Hz, 1H), 3.53 (ddd, J = 6.7, 5.2, 1.1 Hz, 1H).

cis-5-methyl-6-phenylpiperidin-2-one (*cis*-S1d)



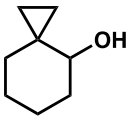
cis-**S1d** was prepared following a reported procedure. Propiophenone (2.0 g, 15 mmol, 1 equiv.), acrylamide (1.2 g, 17 mmol, 1.15 equiv.) and potassium *tert*-butoxide (1.8 g, 16 mmol, 1.05 equiv.) were combined and stirred in dioxane (20 mL, 0.75 M) for 2 h at room temperature. After, the reaction mixture was heated to 80 ˚C overnight. The reaction mixture was cooled to room temperature, diluted with DCM and quenched

with NH_4Cl. After extracting the aqueous layer with DCM, the combined extracts were dried with Na_2SO_4 and the solvent removed in vacuo to yield the dihydropyridone.

In a next step dihydropyridone (1.04 g, 5.56 mmol, 1 equiv.) and Pd/C (5%, 710 mg, 0.34 mmol, 0.06 equiv.) were added to a round bottom flask, and dioxane (9 mL) and ethanol (19 mL) were added under nitrogen atmosphere. The system was purged with hydrogen and placed under hydrogen atmosphere using a balloon. After stirring overnight, the reaction mixture was filtered through a short pad of celite with EtOAc. The crude material was purified by silica gel column chromatography to give *cis*-**S1d** as an pale yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52 – 7.32 (m, 2H), 7.32 – 7.27 (m, 1H), 7.22 – 7.15 (m, 2H), 6.15 (br, s, 1H), 4.61 (dd, *J* = 4.9, 2.3 Hz, 1H), 2.59 – 2.40 (m, 2H), 2.25 (dddd, *J* = 8.5, 6.9, 4.9, 3.6 Hz, 1H), 1.81 – 1.64 (m, 2H), 0.73 (d, *J* = 7.0 Hz, 3H).

spiro[2.5]octan-4-ol (S2)



S2 was prepared following a reported procedure.[12] The final product further purified by selective oxidation of the intermediate alkene with mCPBA.[13]

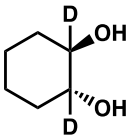
A three-necked flask was charged with diethylzinc solution (22 mL, 1.0 M, 22.4 mmol, 4 equiv.) and DCM (30 mL), and cooled to –20 ˚C. Diiodomethane (3.6 mL, 45 mmol, 8 equiv.) was added dropwise and stirred for another 10 min. A solution of 2-methylenecyclohexan-1-ol (627 mg, 5.6 mmol, 1 equiv.) in DCM (5 mL) was added dropwise and the reaction mixture was stirred at –20 ˚C for 20 min and subsequently at 0 ˚C for 5 h. The reaction was quenched by the addition of aq. NH_4Cl and the aqueous phase was extracted with DCM. The combined organic extracts were dried with Na_2SO_4 and the solvent removed in vacuo. The crude material was purified by silica gel column chromatography (Hexanes/Et_2O).

Residual alkene could not be removed by column chromatography and was therefore subjected to oxidation with *m*CPBA. The impure starting material (117 mg, 16% alkene by NMR) was dissolved in DCM (1 mL) and *m*CPBA (75%, 35 mg, 1 equiv. relative to alkene impurity) was added. After stirring at room temperature for 1 h, the reaction mixture was quenched by the addition of 1N NaOH. The aqueous phase was extracted with DCM and the combined organic

extracts were dried with Na_2SO_4. The crude material was purified by silica gel chromatography to give **S2** as a clear oil.

^{1}H NMR (500 MHz, $CDCl_3$) δ 3.19 (t, *J* = 3.7 Hz, 1H), 1.95 – 1.61 (m, 4H), 1.64 – 1.52 (m, 1H), 1.49 (d, *J* = 14.5 Hz, 2H), 0.96 (d, *J* = 9.2 Hz, 1H), 0.48 – 0.36 (m, 2H), 0.30 – 0.19 (m, 2H).

trans-cyclohexane-1,2-*d*$_2$-1,2-diol (*d2-trans*-S1)

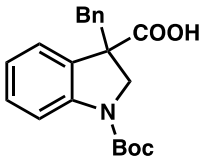


d2-trans-**S1** was prepared following a modified procedure from the literature.[14] trans-**S1** (500 mg, 4.3 mmol, 1 equiv.) was dissolved in D_2O and the solvent removed in vacuo. This procedure was repeated two more times. Then Ru/C (5%, 870 mg, 0.43 mmol, 0.1 equiv.) and D_2O (14 mL, 0.3 M) was added. The system was flushed with H_2 and placed under H_2 atmosphere using a balloon. The reaction mixture was stirred at 80 ˚C overnight and subsequently filtered through celite while still hot. The solvent was evaporated in vacuo and the crude material was purified by silica gel chromatography (3:1 hexanes/acetone) to give *d2-trans*-**S1** as a white solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 3.36 (m, 0.1 H), 1.99–1.96 (m, 2H), 1.73–1.68 (m, 2H), 1.27–1.26 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 75.48 (t), 32.89, 24.46.

3-benzyl-1-(*tert*-butoxycarbonyl)indoline-3-carboxylic acid (S4)

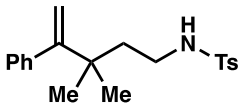


S4 was prepared following a procedure from the literature.[15] 1-(tert-butyl) 3-methyl 3-benzylindoline-1,3-dicarboxylate (920 mg, 2.5 mmol, 1 equiv.), LiOH (210 mg, 5.0 mmol, 2 equiv.), water (2.5 mL) and MeOH (2.5 mL) were combined it a round-bottom flask and stirred at 80 ˚C for 2 h. The reaction was then quenched by the addition of aq. NH_4Cl and extracted with EtOAc. The

combined organic extracts were dried with Na_2SO_4 and the solvent was removed in vacuo. The crude material was purified by silica gel chromatography (0-100% hexanes/EtOAc).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.47 (dd, J = 7.7, 1.4 Hz, 1H), 7.25 – 7.16 (m, 3H), 7.10 – 6.98 (m, 3H), 4.36 (d, J = 11.4 Hz, 1H), 4.00 (s, 1H), 3.45 (d, J = 13.7 Hz, 1H), 3.10 (d, J = 13.7 Hz, 1H), 1.52 (s, 9H).

N-(3,3-dimethyl-4-phenylpent-4-en-1-yl)-4-methylbenzenesulfonamide (S6)

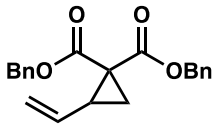


S6 was prepared in six steps from isobutyrophenone according to a reported protocol from the literature.[16]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (d, J = 8.2 Hz, 2H), 7.32 (d, J = 8.0 Hz,

2H), 7.23 (dd, J = 5.0, 2.0 Hz, 3H), 7.02 – 6.92 (m, 2H), 5.10 (d, J = 1.4 Hz, 1H), 4.85 (d, J = 1.3 Hz, 1H), 4.22 – 4.09 (m, 1H), 3.00 (ddd, J = 10.5, 8.1, 5.8 Hz, 2H), 2.44 (s, 3H), 1.62 – 1.43 (m, 4H), 1.06 (s, 6H).

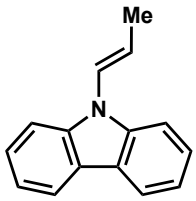
Dibenzyl 2-vinylcyclopropane-1,1-dicarboxylate (S8)



S8 was prepared following a procedure from the literature.[17] 1,4-dibromo-2-butene (2.84 g, 10.0 mmol, 1 equiv.) and NaH (60 wt%. 840 mg, 21.0 mmol, 2.1 equiv.) were suspended in THF (33 mL, 0.33 M), and di-*t*-butyl malonate (2.50 mL, 10.0 mmol, 1 equiv.) was added. The suspension was stirred at room temperature for 2 h. The reaction was the quenched by the addition of aq. NH_4Cl and the aqueous phase was extracted with EtOAc. The combined organic extracts were dried with $Na2SO_4$ and the solvent was removed in vacuo. The crude material was purified by silica gel chromatography (0-20% hexanes/EtOAc).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.36 – 7.22 (m, 10H), 5.40 (ddd, J = 17.0, 10.2, 8.3 Hz, 1H), 5.28 (dd, J = 17.1, 1.6 Hz, 1H), 5.23 – 5.06 (m, 5H), 2.63 (q, J = 8.3 Hz, 1H), 1.76 (dd, J = 7.6, 4.9 Hz, 1H), 1.64 – 1.59 (m, 1H).

(E)-9-(prop-1-en-1-yl)-9H-carbazole (S12)

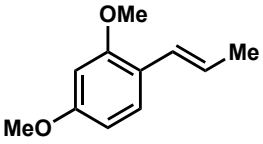


S12 was prepared following a procedure from the literature.[18] 260 mg, 64%

N-(1-Ethoxypropyl)carbazole (500 mg, 2 mmol, 1 equiv.) was dissolved in dry dioxane (2.8 mL, 0.7 M) and pyridine (1.9 mL, 23.7 mmol, 12 equiv.). Acetyl chloride (540 µL, 7.7 mmol, 3.9 equiv.) was added dropwise and the reaction mixture was stirred at 95 ˚C overnight. After cooling to room temperature, water was added and the aqueous layer was extracted with EtOAc. The combined organic extracts were washed with water three times and subsequently dried with Na_2SO_4. The solvent was removed in vacuo and the crude material was purified by silica gel column chromatography to give **S12** as a pale-yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.08 (d, *J* = 7.7 Hz, 2H), 7.75 – 7.53 (m, 2H), 7.46 (ddd, *J* = 8.4, 7.1, 1.2 Hz, 2H), 7.36 – 7.12 (m, 2H), 6.96 (dq, *J* = 14.0, 1.7 Hz, 1H), 6.14 (dq, *J* = 13.6, 6.8 Hz, 1H), 2.03 (dd, *J* = 6.8, 1.7 Hz, 3H).

(*E*)-2,4-dimethoxy-1-(prop-1-en-1-yl)benzene (S16)



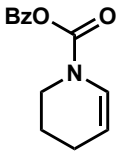
S16 was prepared following a procedure from the literature.[19] Ethyl-triphenylphosphonium iodide (3.0 g, 7.2 mmol, 1.2 equiv.) was dissolved in THF (35 mL, 0.17 M) and cooled to 0 ˚C. n-Butyllithium (2.5 M, 2.9 mL, 7.2 mmol, 1.2 equiv.) was added and the reaction mixture was stirred for another 30 min at 0 ˚C. 2,4-Dimethoxybenzaldehyde (1.0 g, 6.0 mmol, 1 equiv.) was added and the solution was allowed to warm to room temperature while stirring overnight. After the addition of aq. NH_4Cl, the aqueous phase was extracted with Et_2O, the combined organic extracts were dried with Na_2SO_4 and the solvent was removed in vacuo. The crude material was purified by silica gel column chromatography to give a mixture is cis and trans isomers.

The crude material (250 mg, 1.5 mmol, 1 equiv.) was added to a reaction tube and diluted with DCM (2.8 mL, 0.5 M). The tube was placed under nitrogen atmosphere and bis(acetonitrile)palladium (II) chloride (182 mg, 0.7 mmol, 0.5 equiv.) was added. After stirring

overnight at room temperature, the reaction mixture was filtered through a short pad of Florisil and eluted with Et_2O. After removal of the solvent in vacuo the crude material was purified by silica gel column chromatography to give **S16** as a clear oil. Spectral characterization is consistent with literature data.[20]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (d, *J* = 8.4 Hz, 1H), 6.61 (dd, *J* = 15.9, 1.8 Hz, 1H), 6.48 – 6.31 (m, 2H), 6.10 (dq, *J* = 15.8, 6.6 Hz, 1H), 3.82 (s, 3H), 3.80 (s, 3H), 1.87 (dd, *J* = 6.6, 1.8 Hz, 3H).

Benzoic 3,4-dihydropyridine-1(2*H*)-carboxylic anhydride (S14)

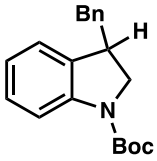


S14 was prepared in two steps from 5-amino-1-pentanol according to the literature.[21] The crude material was purified by silica gel column chromatography to give **S14** as a clear oil. Spectral characterization is consistent with the literature report.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.46 – 7.29 (m, 5H), 6.95 – 6.74 (m, 1H), 5.18 (s, 2H), 5.03 – 4.77 (m, 1H), 3.63 (d, *J* = 5.7 Hz, 2H), 2.11 – 1.97 (m, 2H), 1.91 – 1.76 (m, 2H).

VI.56 – Products

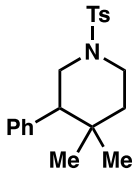
***tert*-butyl 3-benzylindoline-1-carboxylate (S5)**



S5 was isolated from the reaction mixture. Spectral characterization is consistent with literature data.[15]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (s, 1H), 7.31 (t, J = 7.4 Hz, 2H), 7.24 (d, J = 7.5 Hz, 1H), 7.19 (dd, J = 6.9, 1.8 Hz, 3H), 6.98 (d, J = 7.4 Hz, 1H), 6.91 (td, J = 7.4, 1.0 Hz, 1H), 4.02 – 3.92 (m, 1H), 3.72 (s, 1H), 3.61 (tt, J = 9.5, 5.8 Hz, 1H), 3.07 (dd, J = 13.8, 6.0 Hz, 1H), 2.81 (dd, J = 13.8, 9.0 Hz, 1H), 1.55 (s, 9H).

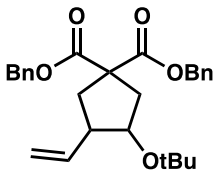
4,4-dimethyl-3-phenyl-1-tosylpiperidine (S7)



S7 was isolated from the reaction mixture. Spectral characterization is consistent with literature data.[22]

^{1}H NMR (400 MHz, **$CDCl_3$**) δ 7.74 – 7.59 (m, 2H), 7.35 (d, *J* = 7.9 Hz, 2H), 7.26 – 7.21 (m, 3H), 7.08 – 6.94 (m, 2H), 3.75 – 3.55 (m, 2H), 2.83 – 2.68 (m, 2H), 2.61 (td, *J* = 12.2, 2.8 Hz, 1H), 2.47 (s, 3H), 1.73 (td, *J* = 13.1, 4.6 Hz, 1H), 1.49 (t, *J* = 2.8 Hz, 1H), 0.81 (s, 3H), 0.70 (s, 3H).

dibenzyl 3-(*tert*-butoxy)-4-vinylcyclopentane-1,1-dicarboxylate (S9)



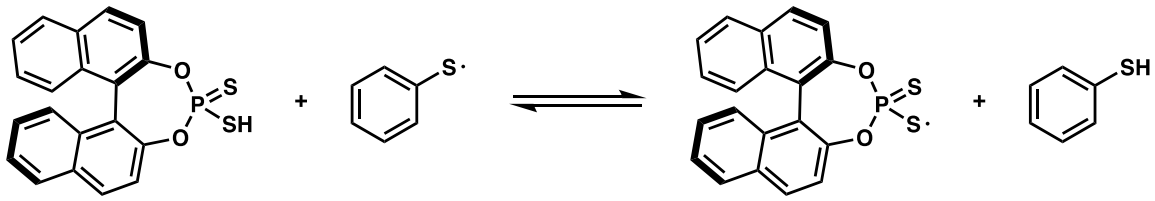
S9 was isolated from the reaction mixture. Spectral characterization is consistent with literature data.[17]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 – 7.27 (m, 6H), 7.23 (d, J = 5.2 Hz, 4H), 5.67 (ddd, J = 17.5, 10.4, 7.7 Hz, 1H), 5.18 – 4.95 (m, 6H), 3.67 (q, J = 7.9 Hz, 1H), 2.64 (dt, J = 13.5, 6.7 Hz, 2H), 2.49 (p, J = 8.3 Hz, 1H), 2.05 (dd, J = 13.7, 8.2 Hz, 1H), 1.86 – 1.74 (m, 1H), 1.21 – 1.08 (m, 9H).

VI.57 – Computational Data

DFT calculations were carried out at the UB3LYP//6-311++ G(d,p) level of theory with smd=acetonitrile implicit solvation. Structures were optimized and frequency checks were conducted to validate all structures are true minima (no negative vibrational frequencies).

The relationship $BDFE_{P\text{-}2} = E_{PhS\bullet} + E_{P\text{-}2} - E_{PhS} + E_{P\text{-}2\bullet}$ was used to determine the BDFE of P-2. using the experimental literature value of thiophenol.[23]



Supplementary Table VI-73. Calculates energies of thiyl species.

	E_{SH} (Hartree)	$E_{S\bullet}$ (Hartree)
PhS	-630.469	-629.848
P-2	-2058.33	-2057.7

Supplementary Table VI-74. Atom coordinates of PhS•.

Center Number	Atomic Number	Coordinates (Angstroms)		
		X	Y	Z
1	6	1.534908	1.214804	-0.000042
2	6	0.149558	1.221394	-0.000004
3	6	-0.575603	0	0.000029
4	6	0.149558	-1.221394	-0.000004
5	6	1.534908	-1.214804	-0.000042

6	6	2.23184	0	-0.000055
7	1	2.080249	2.151674	-0.000059
8	1	-0.395535	2.157517	0.000005
9	1	-0.395535	-2.157517	0.000005
10	1	2.080249	-2.151674	-0.000059
11	1	3.316036	0	-0.000089
12	16	-2.302279	0	0.000057

Supplementary Table VI-75. Atom coordinates of PhSH.

Center Number	Atomic Number	Coordinates (Angstroms)		
		X	Y	Z
1	6	-1.58594	1.209923	-0.000028
2	6	-0.191381	1.209742	0.000076
3	6	0.503476	-0.004707	0.000132
4	6	-0.205165	-1.212085	0.000027
5	6	-1.598818	-1.199274	0.000012
6	6	-2.296273	0.009399	-0.000045
7	1	-2.115421	2.156571	-0.000127
8	1	0.347448	2.150717	0.000045
9	1	0.324952	-2.158414	-0.000075
10	1	-2.138385	-2.140106	-0.000046
11	1	-3.380308	0.015385	-0.000093
12	16	2.293262	-0.084342	-0.000037
13	1	2.514126	1.247331	-0.00015

Supplementary Table VI-76. Atom coordinates of P-2•.

Center Number	Atomic Number	Coordinates (Angstroms)		
		X	Y	Z
1	6	-0.703696	2.586021	-1.331753
2	6	0.392288	3.387105	-1.092084
3	6	1.480279	2.92465	-0.308159
4	6	1.468947	1.58496	0.190545
5	6	0.366219	0.712993	-0.149368
6	6	-0.73977	1.281441	-0.828379
7	1	2.538253	4.800926	-0.370705
8	1	0.416438	4.400621	-1.476795
9	6	2.544155	3.792564	0.027881
10	6	2.497798	1.201597	1.079579
11	6	3.520627	2.076876	1.407423
12	6	3.55851	3.372292	0.86841
13	1	2.484445	0.217499	1.528676
14	1	4.296164	1.755145	2.092939
15	1	4.36783	4.045029	1.126632
16	6	-0.73942	-1.28196	0.828304
17	6	0.366423	-0.713112	0.14924
18	6	1.469446	-1.584695	-0.19062
19	6	1.481363	-2.924321	0.308223
20	6	0.393579	-3.38713	1.092248
21	6	-0.702725	-2.58647	1.331845
22	1	2.484255	-0.21696	-1.52898
23	6	2.498065	-1.201	-1.079767

24	6	2.545547	-3.791823	-0.027789
25	1	0.418136	-4.400601	1.47705
26	6	3.559695	-3.371217	-0.868455
27	6	3.521251	-2.0759	-1.407598
28	1	2.540098	-4.800147	0.370898
29	1	4.369262	-4.043665	-1.12665
30	1	4.296602	-1.753915	-2.093202
31	8	-1.81938	0.518205	-1.154304
32	8	-1.819374	-0.519339	1.154126
33	15	-2.947363	-0.000314	-0.000043
34	16	-3.8895	1.520789	0.825407
35	16	-3.890126	-1.521204	-0.825175
36	1	-1.543673	-2.940932	1.914233
37	1	-1.544789	2.940191	-1.914109

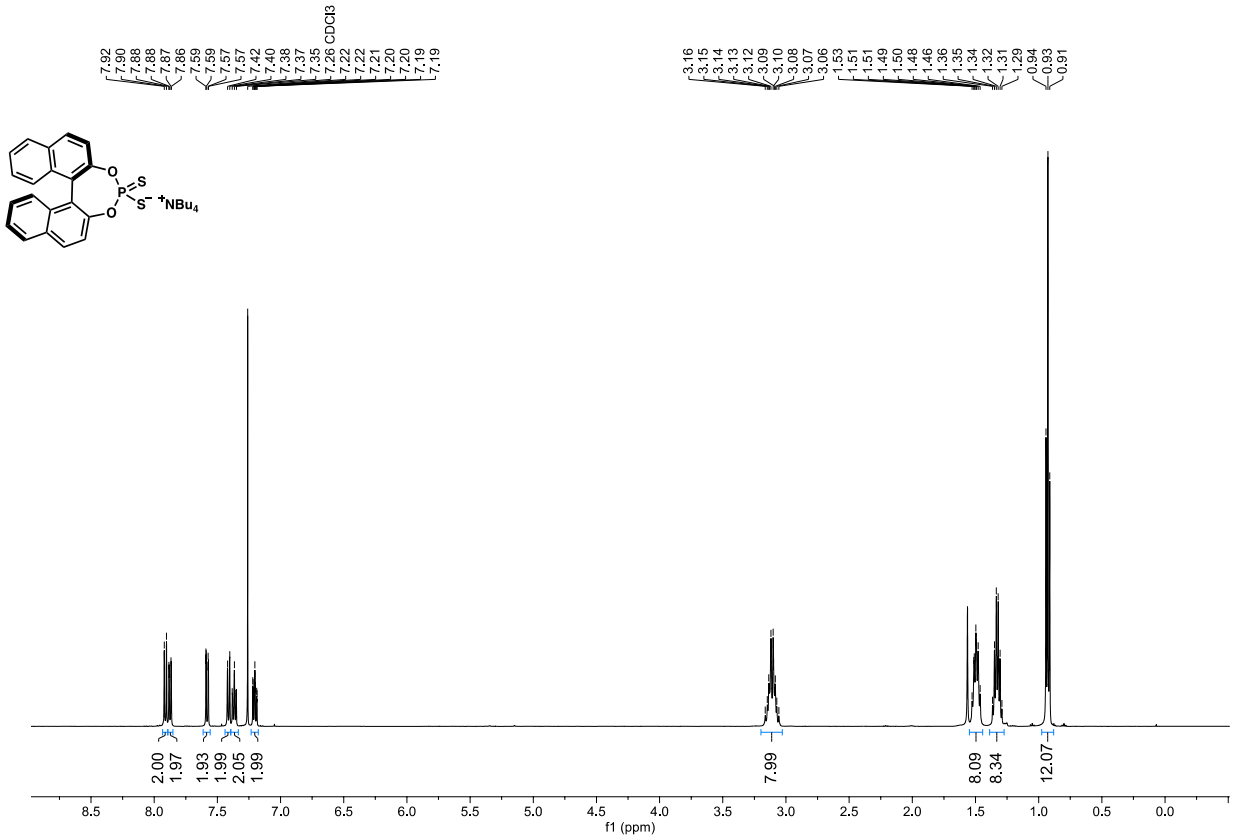
Supplementary Table VI-77. Atom coordinates of P-2.

Center Number	Atomic Number	Coordinates (Angstroms)		
		X	Y	Z
1	6	0.799235	2.515515	1.396885
2	6	-0.215307	3.394981	1.123055
3	6	-1.314243	3.0007	0.315778
4	6	-1.390747	1.654153	-0.169968
5	6	-0.369313	0.711178	0.202818
6	6	0.713514	1.196345	0.912394
7	1	-2.258805	4.942116	0.34479

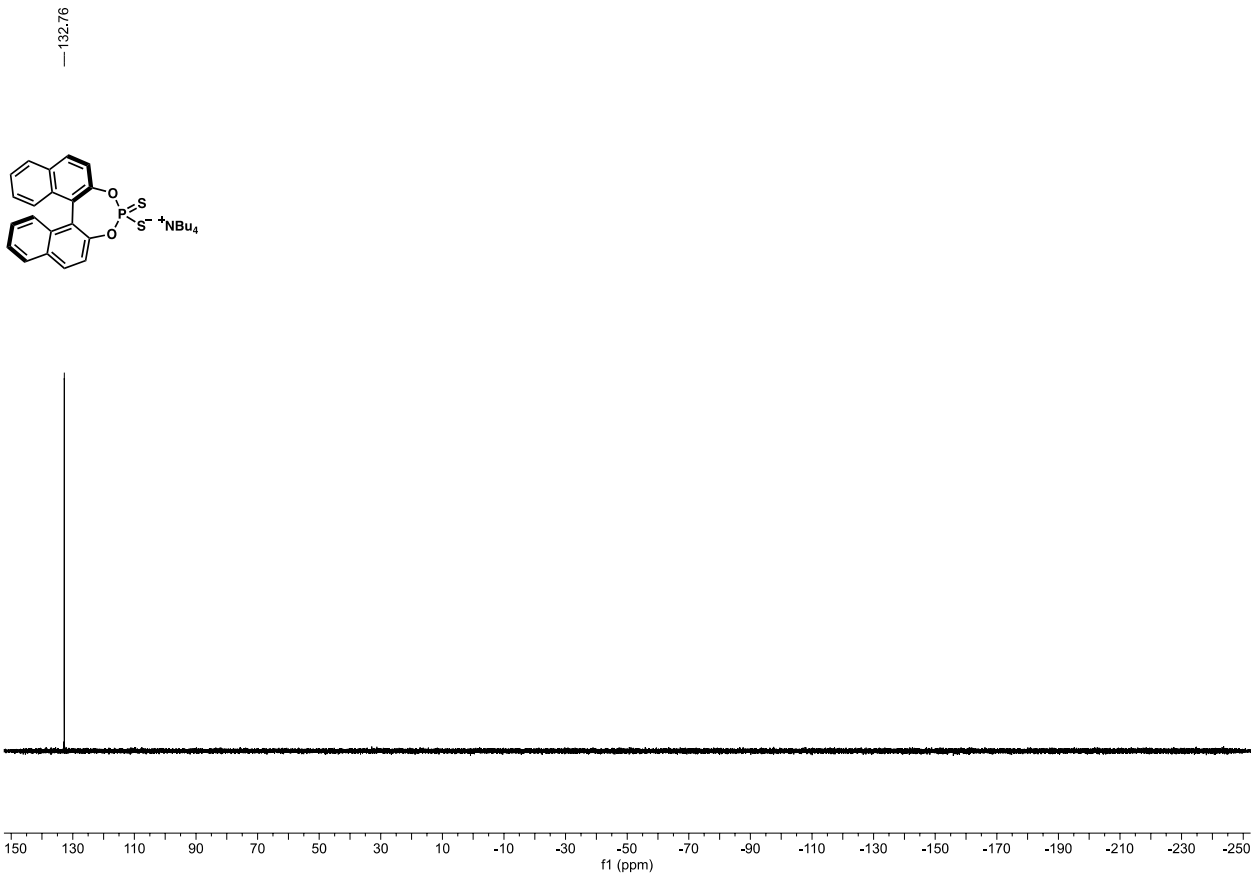
8	1	-0.174055	4.412086	1.496776
9	6	-2.326644	3.93076	-0.042108
10	6	-2.459586	1.320592	-1.045832
11	6	-3.413562	2.251064	-1.389105
12	6	-3.359057	3.56579	-0.872349
13	1	-2.516593	0.323598	-1.462028
14	1	-4.213532	1.974032	-2.066854
15	1	-4.123369	4.284914	-1.144915
16	6	0.613864	-1.307246	-0.868323
17	6	-0.432708	-0.728053	-0.17689
18	6	-1.538386	-1.577094	0.180099
19	6	-1.568632	-2.927381	-0.300674
20	6	-0.494028	-3.418093	-1.087592
21	6	0.595582	-2.629231	-1.350113
22	1	-2.568188	-0.150471	1.449689
23	6	-2.589275	-1.150292	1.036938
24	6	-2.662348	-3.766592	0.042599
25	1	-0.533761	-4.436973	-1.456572
26	6	-3.673269	-3.311921	0.854694
27	6	-3.624124	-1.995205	1.366959
28	1	-2.67432	-4.781501	-0.340636
29	1	-4.500106	-3.962508	1.116827
30	1	-4.408436	-1.648532	2.030866
31	8	1.780263	0.342753	1.230217
32	8	1.756391	-0.543307	-1.15551
33	15	2.792546	-0.19536	0.062947

34	16	3.693055	1.456457	-0.90345
35	1	4.743767	1.473661	-0.051324
36	16	3.944049	-1.586876	0.759959
37	1	1.442101	-2.993928	-1.918787
38	1	1.664395	2.805202	1.980555

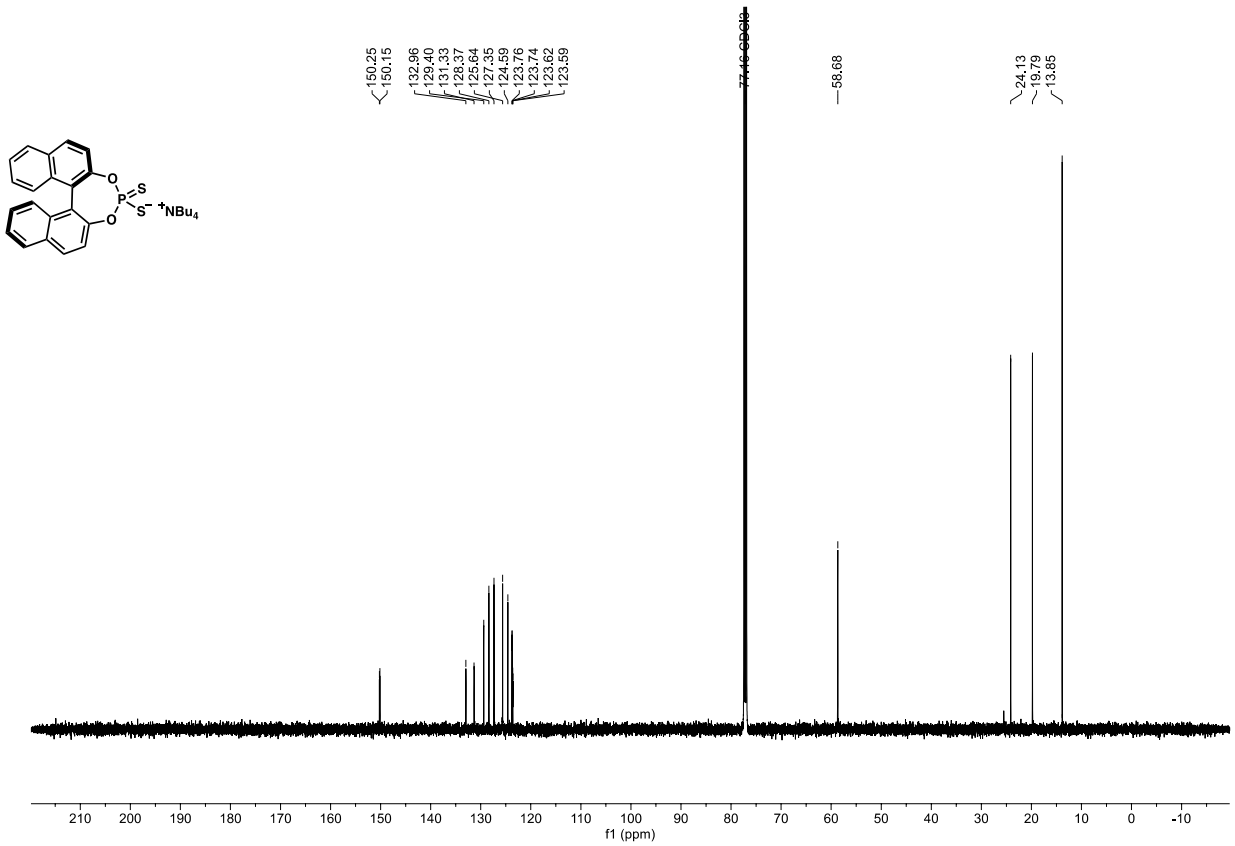
VI.58 – NMR Spectra



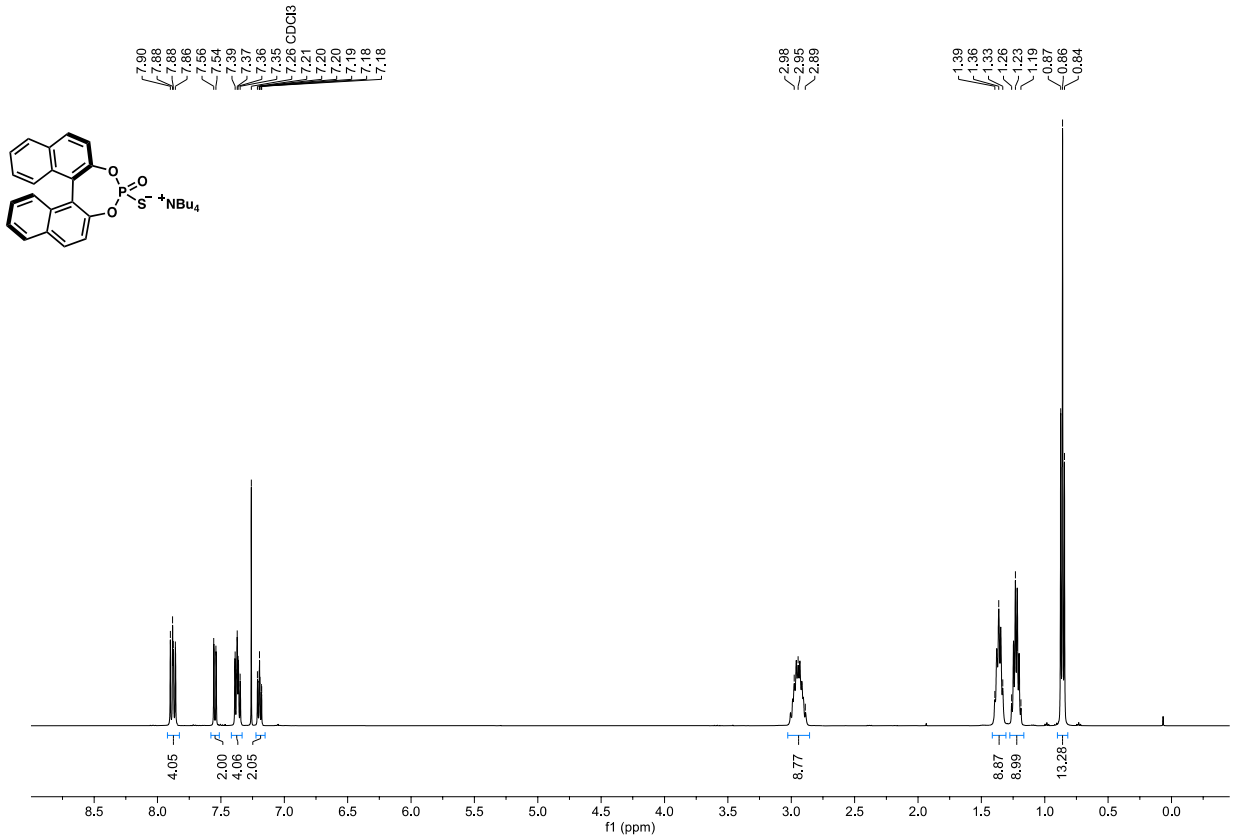
Supplementary Fig. VI-53. ^{1}H NMR of **P-1** in $CDCl_3$.



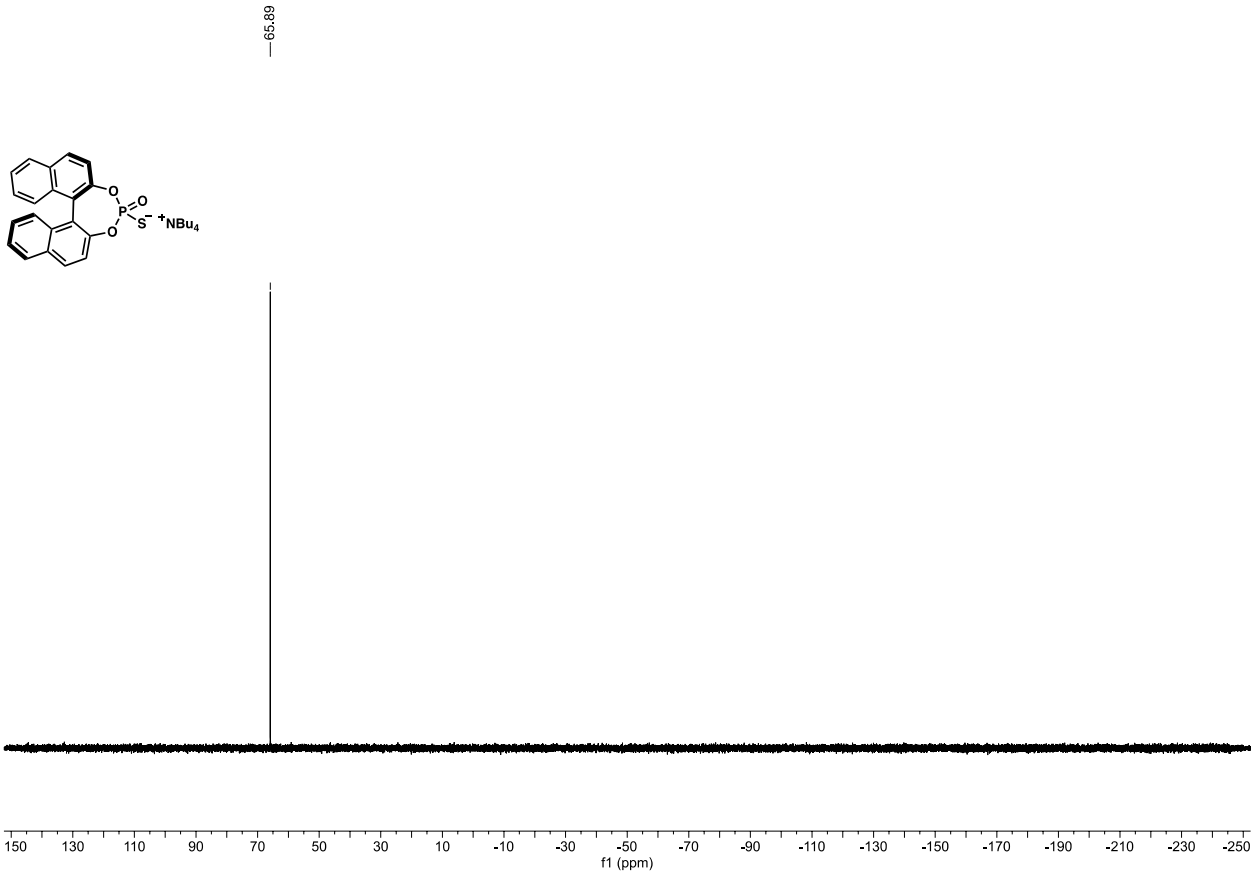
Supplementary Fig. VI-54. ^{31}P NMR of **P-1** in $CDCl_3$.



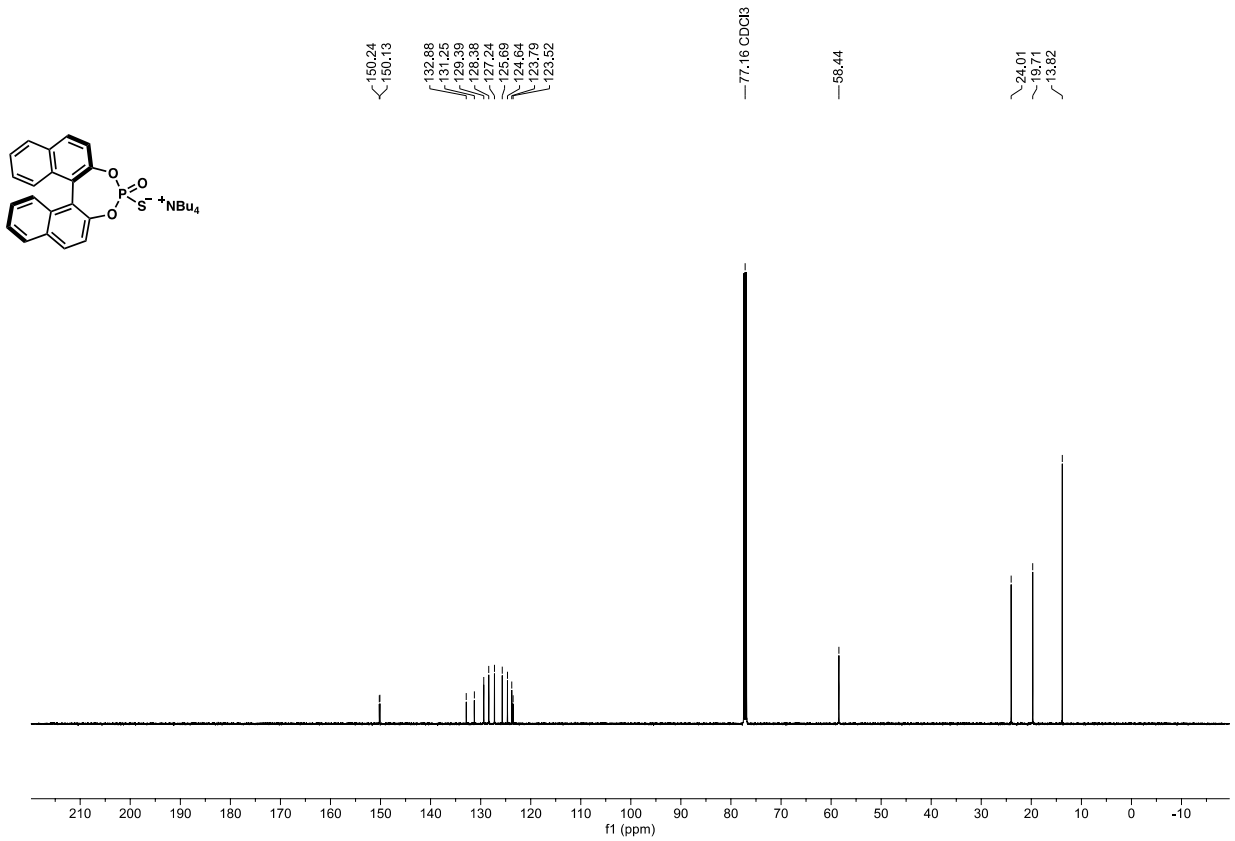
Supplementary Fig. VI-55. ^{13}C NMR of **P-1** in $CDCl_3$.



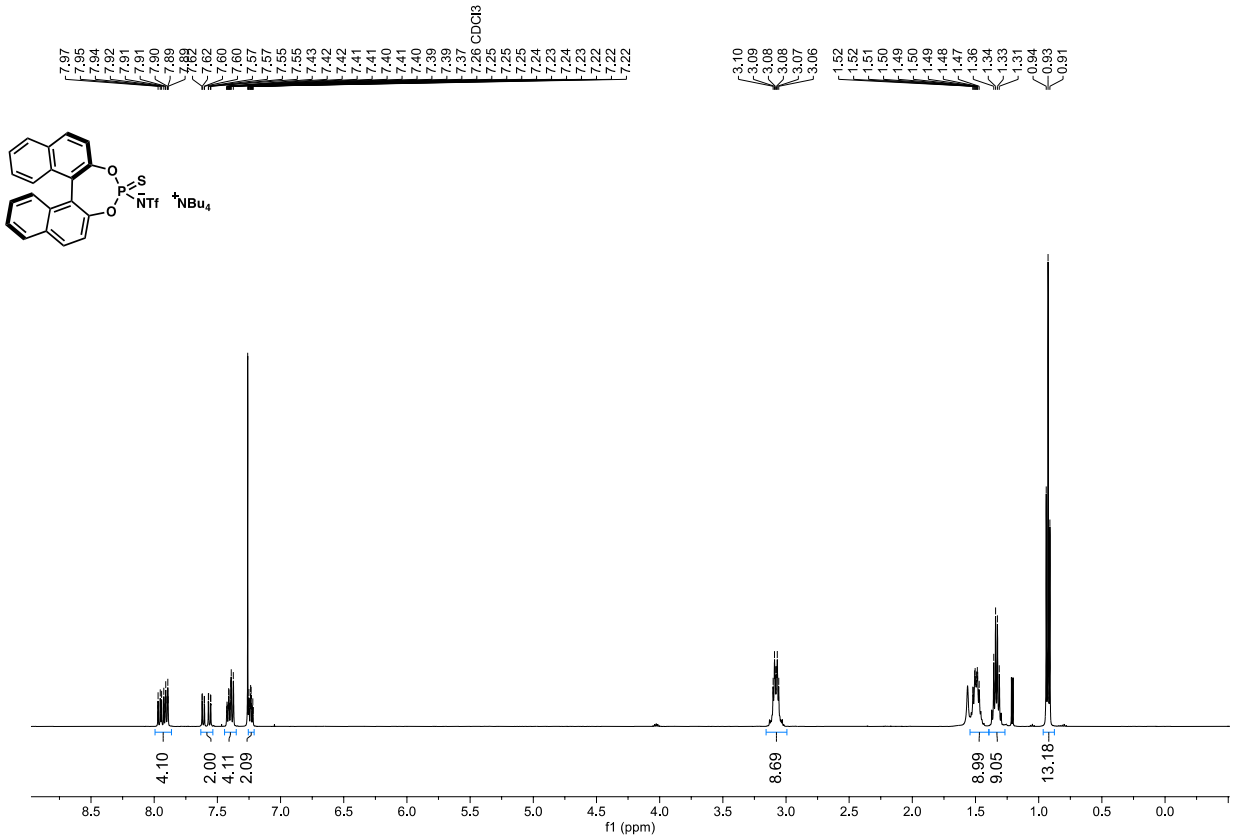
Supplementary Fig. VI-56. ^{1}H NMR of **P-3** in $CDCl_3$.



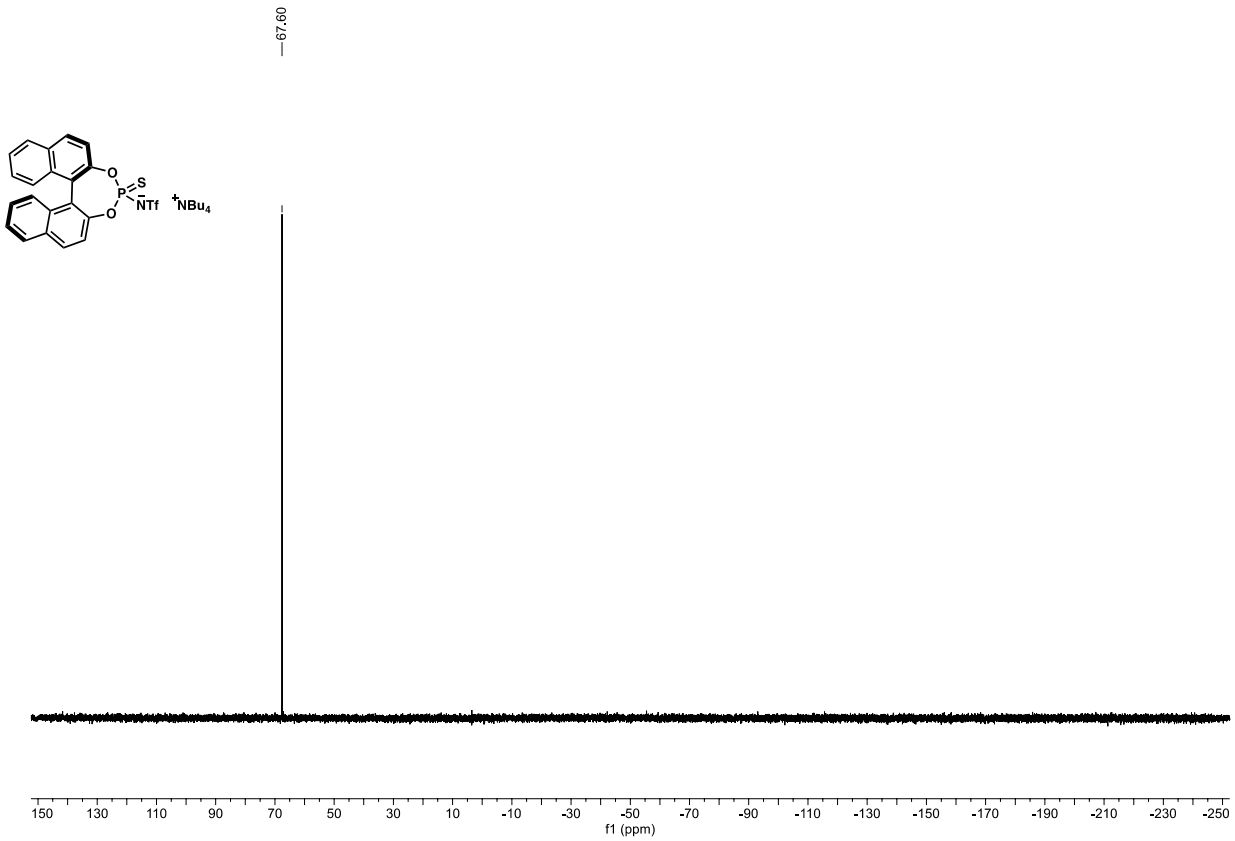
Supplementary Fig. VI-57. ^{31}P NMR of **P-3** in $CDCl_3$.



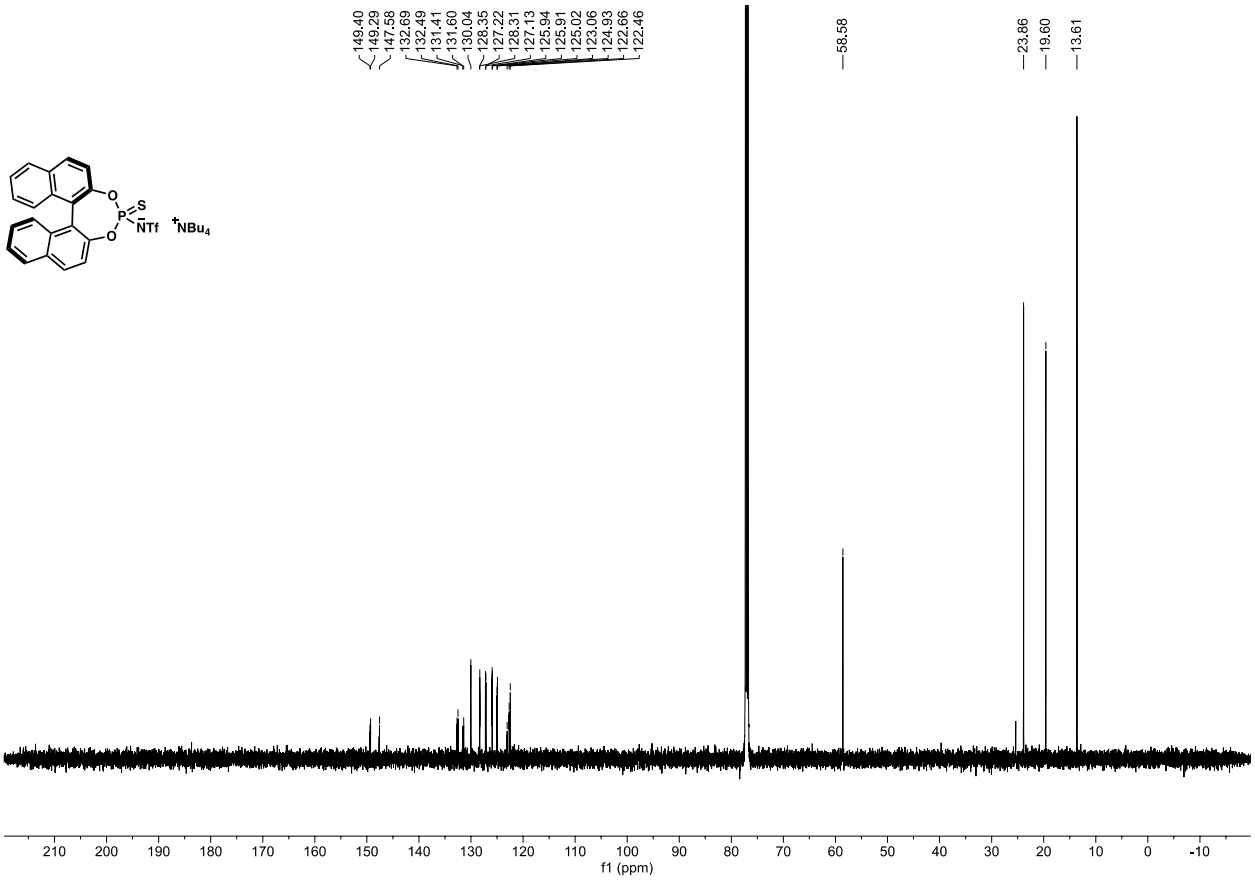
Supplementary Fig. VI-58. ^{13}C NMR of **P-3** in $CDCl_3$.



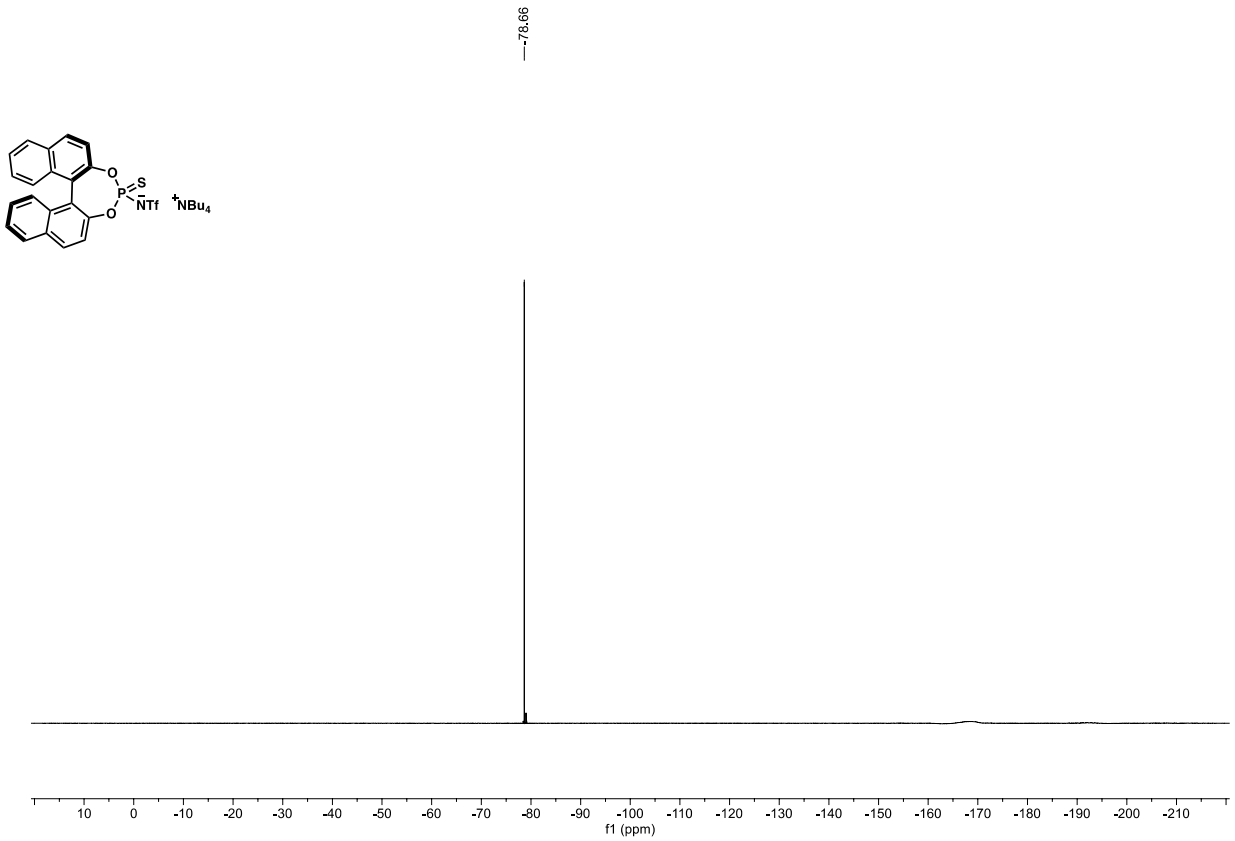
Supplementary Fig. VI-59. ^{1}H NMR of **P-5** in $CDCl_3$.



Supplementary Fig. VI-60. ^{31}P NMR of **P-5** in $CDCl_3$.



Supplementary Fig. VI-61. ^{13}C NMR of **P-5** in $CDCl_3$.



Supplementary Fig. VI-62. ^{19}F NMR of **P-5** in $CDCl_3$.

VI.59 – References

(1) *Gaussian 16 | Gaussian.com*.

(2) Zhang, Y.-A.; Gu, X.; Wendlandt, A. E. A Change from Kinetic to Thermodynamic Control Enables Trans-Selective Stereochemical Editing of Vicinal Diols. *J. Am. Chem. Soc.* **2022**, *144* (1), 599–605.

(3) Kazerouni, A. M.; Brandes, D. S.; Davies, C. C.; Cotter, L. F.; Mayer, J. M.; Chen, S.; Ellman, J. A. Visible Light-Mediated, Highly Diastereoselective Epimerization of Lactams from the Most Accessible to the More Stable Stereoisomer. *ACS Catal.* **2022**, *12* (13), 7798–7803.

(4) Kuchen, W.; Mahler, H. F. DIBENZO(d,f)-1,3,2-DIOXAPHOSPHOREPINE UND METALLKOMPLEXE VON 2,2′-BIPHENYLYLENPHOSPHORSÄUREN. *Phosphorus and Sulfur and the Related Elements* **1980**, *8* (2), 139–145.

(5) Le Saux, E.; Georgiou, E.; Dmitriev, I. A.; Hartley, W. C.; Melchiorre, P. Photochemical Organocatalytic Functionalization of Pyridines via Pyridinyl Radicals. *J. Am. Chem. Soc.* **2023**, *145* (1), 47–52.

(6) Kato, S.; Saga, Y.; Kojima, M.; Fuse, H.; Matsunaga, S.; Fukatsu, A.; Kondo, M.; Masaoka, S.; Kanai, M. Hybrid Catalysis Enabling Room-Temperature Hydrogen Gas Release from N-Heterocycles and Tetrahydronaphthalenes. *J. Am. Chem. Soc.* **2017**, *139* (6), 2204–2207.

(7) Tanabe, S.; Mitsunuma, H.; Kanai, M. Catalytic Allylation of Aldehydes Using Unactivated Alkenes. *J. Am. Chem. Soc.* **2020**, *142* (28), 12374–12381.

(8) Perlikowska, W.; Gouygou, M.; Mikolajczyk, M.; Daran, J.-C. Enantiomerically Pure Disulfides: Key Compounds in the Kinetic Resolution of Chiral PIII-Derivatives with Stereogenic Phosphorus. *Tetrahedron: Asymmetry* **2004**, *15* (22), 3519–3529.

(9) Martiny, M.; Steckhan, E.; Esch, T. Cycloaddition Reactions Initiated by Photochemically Excited Pyrylium Salts. *Chemische Berichte* **1993**, *126* (7), 1671–1682.

(10) Liang, J.; Mendonca, R. V.; Siu, M.; Tellis, J. C.; Wang, W.; Wei, B.; Chan, B. K.; Choo, E. F.; Drobnick, J. A.; Gazzard, L. J.; Heffron, T. 8-Aminoisoquinoline Compounds and Uses Thereof. WO2020072695A1, April 9, 2020.

(11) Carder, H. M.; Suh, C. E.; Wendlandt, A. E. A Unified Strategy to Access 2- and 4-Deoxygenated Sugars Enabled by Manganese-Promoted 1,2-Radical Migration. *J. Am. Chem. Soc.* **2021**, *143* (34), 13798–13805.

(12) Wübbolt, S.; Cheong, C. B.; Frost, J. R.; Christensen, K. E.; Donohoe, T. J. A Vinyl Cyclopropane Ring Expansion and Iridium-Catalyzed Hydrogen Borrowing Cascade. *Angewandte Chemie International Edition* **2020**, *59* (28), 11339–11344.

(13) Schomaker, J. M.; Pulgam, V. R.; Borhan, B. Synthesis of Diastereomerically and Enantiomerically Pure 2,3-Disubstituted Tetrahydrofurans Using a Sulfoxonium Ylide. *J. Am. Chem. Soc.* **2004**, *126* (42), 13600–13601.

(14) Moure, M. J.; Zhuo, Y.; Boons, G. J.; Prestegard, J. H. Perdeuterated and 13C-Enriched Myo-Inositol for DNP Assisted Monitoring of Enzymatic Phosphorylation by Inositol-3-Kinase. *Chem. Commun.* **2017**, *53* (92), 12398–12401.

(15) Xu, Y.; Shen, C.; Dong, K. Enantioselective Decarboxylative Hydrogen-Atom Transfer Reaction. *J. Am. Chem. Soc.* **2025**.

(16) Nguyen, T. M.; Nicewicz, D. A. Anti-Markovnikov Hydroamination of Alkenes Catalyzed by an Organic Photoredox System. *J. Am. Chem. Soc.* **2013**, *135* (26), 9588–9591.

(17) Hashimoto, T.; Kawamata, Y.; Maruoka, K. An Organic Thiyl Radical Catalyst for Enantioselective Cyclization. *Nature Chem* **2014**, *6* (8), 702–705.

(18) Roh, Y.; Gao, D.; Bauld, N. L. The Cation Radical Chain Cycloaddition Polymerization of N,3-Bis(Trans-1-2)Carbazole: The Critical Importance of Intramolecular Hole Transfer in Cation Radical Cycloaddition Polymerization. *Advanced Synthesis & Catalysis* **2001**, *343* (5), 481–489.
(19) Colomer, I.; Coura Barcelos, R.; Donohoe, T. J. Catalytic Hypervalent Iodine Promoters Lead to Styrene Dimerization and the Formation of Tri- and Tetrasubstituted Cyclobutanes. *Angewandte Chemie International Edition* **2016**, *55* (15), 4748–4752.
(20) Zhong, J.; Wang, X.; Luo, M.; Zeng, X. Chromium-Catalyzed Alkene Isomerization with Switchable Selectivity. *Org. Lett.* **2024**, *26* (15), 3124–3129.
(21) Okitsu, O.; Suzuki, R.; Kobayashi, S. Efficient Synthesis of Piperidine Derivatives. Development of Metal Triflate-Catalyzed Diastereoselective Nucleophilic Substitution Reactions of 2-Methoxy- and 2-Acyloxypiperidines. *J. Org. Chem.* **2001**, *66* (3), 809–823.
(22) Tang, L.; Shen, C.; Hao, S.; Dong, K. A Type of Chiral C2-Symmetric Arylthiol Catalyst for Highly Enantioselective Anti-Markovnikov Hydroamination. *J. Am. Chem. Soc.* **2024**, *146* (23), 16248–16256.
(23) Bordwell, F. G.; Zhang, X.-M.; Satish, A. V.; Cheng, J.-P. Assessment of the Importance of Changes in Ground-State Energies on the Bond Dissociation Enthalpies of the O-H Bonds in Phenols and the S-H Bonds in Thiophenols. *J. Am. Chem. Soc.* **1994**, *116* (15), 6605–6610.