UNIVERSITY OF CALIFORNIA,
IRVINE

An Asymmetric Route to Xantholipin B, Prediction of Radical Reactions with Machine Learning, and Metal-Catalyzed Carbene Insertion into Aliphatic N−H Bonds

DISSERTATION

Submitted in partial satisfaction of the requirements
For the degree of

MASTER'S

In Chemistry

By

Dora Kadish

Dissertation Committee:
Associate Dean, Professor David Van Vranken, Chair
Professor Ann Marie Carlton
Distinguished Professor Pierre Baldi

2020

Dedication

To everyone who has supported me unconditionally

Acknowledgements

I would like to thank my advisor, Professor David L. Van Vranken, for all his support, as well as Dr. Ryan Gianatassio, who continues to be an inspiration and without whom I would not have gotten as far as I have.

Table of Contents

List of Figures

Abstract

Throughout my time at UCI I have been involved in three projects: development of an asymmetric route to xantholipin B, asymmetric insertion of carbene groups into aliphatic N−H bonds, and developing a pedagogical training set for organic radical chemistry for prediction of reaction products and mechanistic pathways. During the six months that I worked for Professor Sergey Pronin, I optimized conditions for addition of a tri-*n*-butylstannylzinc reagent to a MOM-protected salicylaldehyde. In the Van Vranken group I synthesized and characterized a chiral guanidine catalyst useful for asymmetric catalysis. I then focused my efforts on the development of a prediction for prediction of step-wise organic reaction mechanisms based on machine learning. In order to improve the identification of electrophilic and nucleophilic atoms, I showed that easily calculated methyl ion affinities correlated with Mayr's solution phase reactivity parameters, E and N•s_N. I then helped to expand a training database of mechanistic radical reaction steps based on the SMIRKS formalism and test the system called *Reaction Predictor* once the training was complete.

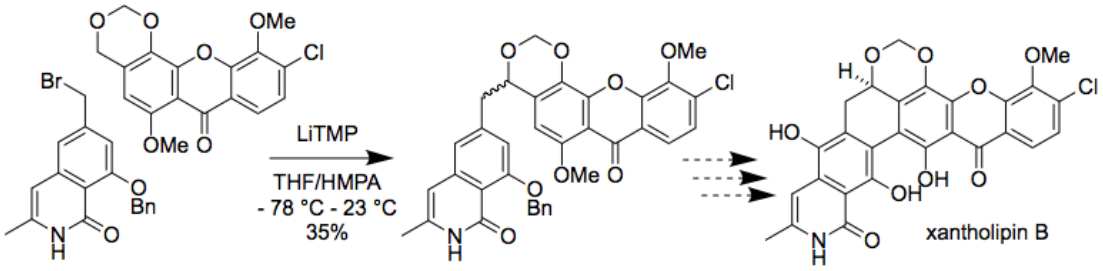
Chapter 1. Synthesis of Key Synthetic Intermediate and a Chiral Guanidine Catalyst

1.1. Synthesis of a Key Intermediate for the Synthesis of Xantholipin B

Xantholipin B was first isolated from *Streptomyces flavogriseus* in 2003.[1] Its xanthone core consists of a chlorinated δ−lactam structure with a 1,3-dioxane on the hinge region. The latter structure is one that is often found in polycyclic xanthone antibiotics.[1] Xantholipin B inhibits expression of the heat shock protein HSP47 (IC_{50} = 0.20 μM), with potential for treatment of fibrotic diseases and has modest activity against a number of bacterial and mammalian targets. Notably, xantholipin B exhibits potent activity against oral squamous carcinoma cell line KB (IC_{50} < 2 nM),[1] attracting interest as a useful synthetic target.

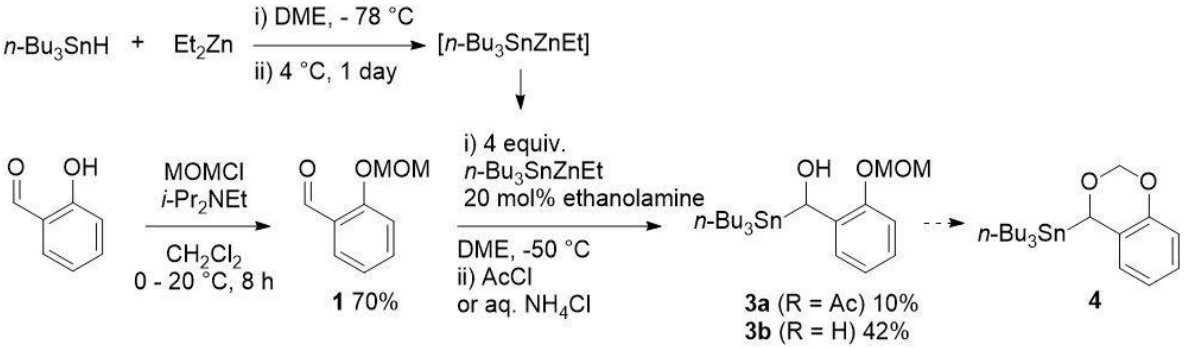
The first route toward xantholipin B was developed by Eric Kuenstner in 2018, and utilized a convergent approach (Scheme 1) to the hexacyclic core that ultimately culminated in the synthesis of the simpler dihydrophenanthrene polyketide CBS62, lacking the 1,3-dioxane ring.[2,3] The Kuenstner route to xantholipin involved a metalation of the 1,3-dioxane and capture of the racemic alkyllithium by a benzyl bromide fragment.

Scheme 1. Kuenstner route to xantholipin B



I set out to explore a strategy for asymmetric synthesis of a 4-lithio-1,3,2-benzodioxin (Scheme 2) for addition to benzylic bromides or aldehydes, ultimately leading to the asymmetric syntheses of xantholipin B, and ideally related natural products. My approach hinged upon access to a 4-stannyl-1,3,2-benzodioxin **4**, setting the stage for stereospecific Sn-Li exchange. Salicylaldehyde was used as the starting point for construction of a model system for development of the chemistry.

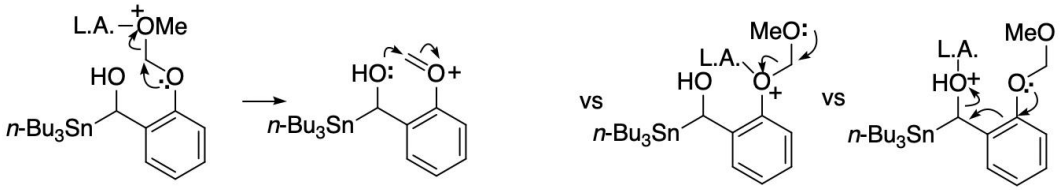
Scheme 2. Synthesis of a chiral 4-stannyl-1,3,2-benzodioxin **4**



The phenolic group was protected as the MOM ether in 70% yield. He and Falck have previously reported the generation and addition of tri-*n*-butylstannylzinc reagents and chiral amino alcohols to aldehydes, followed by protection of the alpha-stannyl alcohols.[4] The conditions indicated optimizing the temperature to be between -30 and -60 °C, at which time the reaction was reported to

proceed in up to 86% yield and high ee in the presence of chiral amino alcohols. My attempts to apply the Falck conditions to salicaldehyde derivative **1** in the presence of dimethylethanolamine at - 50 °C led to <10% yield of alpha-stannyl acetate **3a** (Scheme 2). Surprisingly, I was able to isolate a promising 42% yield of the sensitive alpha-stannyl alcohol **3b** at - 60 °C by quenching the reaction with aq. NH_4Cl. Attempts to cyclize it to the 4-stannyl-1,3,2-benzodioxin **4** were unsuccessful using various Lewis and Brønsted acids (BF_3, $FeCl_3$, H_3PO_4, etc)**,**[5, 6,7] possibly due to competing modes of ionization (Scheme 3). It might be possible to selectively generate the oxocarbenium intermediate from an MTM ether using DMTSF,[8] but given the sequential challenges of stannylzinc generation, addition, quenching, and cyclization this route has been put on hold.

Scheme 3. Alternative modes of ionization may compete with cyclization.



1.2. Insertion of Carbenes into the N−H bond of Non−Conjugated Amines

1.2.1. Background - Asymmetric Metal-Catalyzed Insertion of Carbenes into X−H Bonds

Over the past ten years metal−catalyzed insertion of carbenes, derived from diazo compounds, has emerged as a highly effective method for generation of stereogenic centers into X−H bonds.[9,10,11,12,13,14,15] Copper catalysts have been the most heavily studied for O−H and N−H insertion reactions, following seminal work by Fu and coworkers;[15, 16] but other metals, such as palladium, iron, rhodium, and silver have also been shown to be effective.

Highly efficient enantioselective (≥ 96% ee) metal−catalyzed insertion of carbenes, derived from diazo compounds, into aromatic O−H bonds has been achieved using a number of transition metal catalysts (CuCl, $Cu(OTf)_2$, $Pd(PhCN)_2Cl_2$,$Cu(MeCN)_4PF_6$, $Ag(NTf_2)$) with chiral ligands.[17,18,19,20,21] Copper(II) precatalysts are reduced to copper(I) under the conditions of the reactions.[22] Electronic structure calculations suggest that the stereogenic center is formed by protonation of a metal-free enol.[23] Enantioselective metal-catalyzed insertion of carbenes into aliphatic O−H bonds seems slightly less general.[24] High levels of asymmetric induction have been achieved using iron(II) catalysts.[25,26] Highly enantioselective insertion into HO−H has been achieved using copper catalysts;[27] DFT calculations suggest a mechanism involving the protonation of a chiral copper enolate.[23] Achiral rhodium (II) has been used in conjunction with chiral phosphoric acids for insertion into aliphatic O−H bonds.[28]

Less success has been achieved with S−H bonds, but high enantioselectivity (up to 96% ee) was obtained using an achiral rhodium (II) catalyst and a chiral phosphoric acid for insertion into aromatic and aliphatic S−H bonds using an achiral rhodium catalyst in combination with a chiral acid.[28] Chiral catalysts, however, have been less effective, particularly for insertion into aliphatic thiol S−H.[29,30]

Asymmetric insertion into N−H bonds has been achieved for anilines, carbamates and related other conjugated amines with low basicity. The earliest work on asymmetric insertion into X−H bonds can be traced to the historically important Merck synthesis of the carbapenem thienamycin through an intramolecular diastereoselective insertion of a carbene into a beta-lactam N−H.[31] No highly effective conditions for enantioselective carbene insertion into amide N−H bonds has been reported,[32] but high levels of enantioselection have been achieved for insertion into carbamates, particularly *t*-

butoxycarbamate.[33,34] In contrast to α-phenyldiazoesters, which, when used as the carbene source, resulted in high enantioselectivity, diazopropionates were shown to result in poor enantioselectivity, with respect to the final chirality of the molecule.[35] Catalytic systems with achiral rhodium catalysts and chiral acids have been particularly successful,[36, 37] presumably involving enantiotopic protonation of an achiral enolate intermediate.[38] High enantioselectivity was acquired using chiral copper(I) catalysts and achiral rhodium catalysts.

Chiral copper complexes have been highly successful for asymmetric insertion of carbenes derived from α-alkyldiazoesters into aniline N−H bonds.[39,40,41,42,43] Recently, a combination of palladium and a guanidine additive have been shown to achieve good enantioselectivity (up to 93% ee) for insertion of α-alkyldiazoesters into N−H bonds of *N*-methylanilines.[44] In contrast, highly enantioselective insertion of α-phenyldiazoesters into N−H bonds of anilines has not been achieved.[45] However, highly enantioselective insertion of α-phenyldiazoesters into N−H bonds of carbazoles was achieved by the Van Vranken group using a combination of palladium(II) and PyBOX.[46]

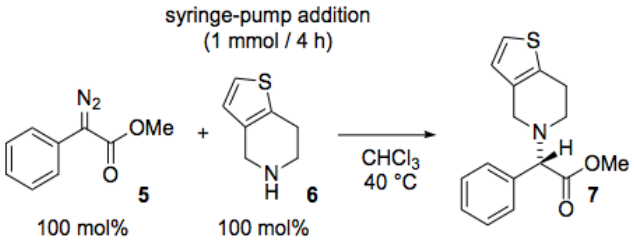
Insertion into into aliphatic N−H bonds has, so far, eluded chemists, and is an untapped archive of racemization and resolution approaches.[47] Highly basic amine substrates interfere with N−H insertion reactions, regardless of the mechanism, in two ways: 1) they compete with chiral ligands for binding to the metal, reducing the asymmetric induction and 2) they compete with chiral bases for asymmetric proton transfers, reducing both the asymmetric induction and catalytic turnover. Both of these issues interfere with obtaining a high yield and good ee. The best ee that had been achieved to date, with carbene insertion into aliphatic N−H bonds that utilized 10 mol % of a chiral copper ligand, was 27%.[48] The properties of highly basic amines that interfere with insertion reactions also lead to poor results for hetererocycles such as imidazole, benzimidazole, and indazole.[49,50,51] We set out to develop conditions for N−H insertion that would work for aliphatic amines and nucleophilic N−H heterocycles.

Scheme 4: Electron-deficient aromatic N−H heterocycles for future insertion studies

1.3. Palladium-Catalyzed Insertion of Carbenes into N−H Bonds

For typical insertion reactions, the diazo compound adds faster to the metal carbene than weakly nucleophilic substrates, (e.g., alkenes (cyclopropanation), anilines, carbamates, alcohols, activated C−H bonds). To disfavor this oligomerization, the diazo compound is added by syringe pump to keep the concentration low. It was reasoned that a highly nucleophilic substrate, such as an aliphatic amine, might outcompete the diazo compound and that a low concentration of the amine, rather than the diazo compound, might be preferred. After initial success with the inverse addition conditions, a previous member of the Van Vranken group, Stan Hiew, screened a variety of metals for carbene insertion into the N−H bond of aliphatic amines. He compared copper(I), palladium(II), and palladium(0) pre-catalysts for asymmetric insertion into the N−H bonds of thieno[3,2-*c*]piperidine (present in clopidogrel), and chose to move forward with palladium(0), which gave higher ee than Jørgensen's copper catalyst system.

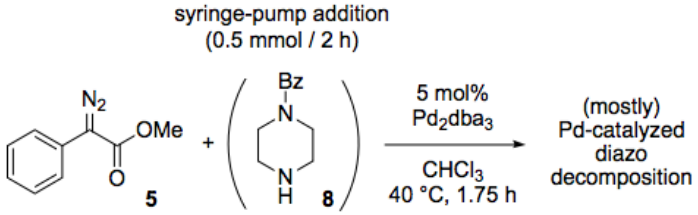
Scheme 5: Hiew screen of catalyst systems for Insertion into aliphatic N−H bonds under inverse addition conditions



conditions	Yield	ee
CuCl/NaBAr^F/SpiroBOX	58%	7%
$Pd(PhCN)_2Cl_2$/NaBAr^F/PyBOX	59%	0%
$Pd_2(dba)_3$ / MonoPhos	15%	19%

We suspected that palladium catalyzes an unwanted decomposition (presumably oligomerization)[52] of the diazo compound. After synthesizing the carbene source through diazo transfer compound **5**, I showed that palladium catalyzes complete decomposition of the diazo compound to baseline material in under two hours (eq 1). I also suspected that high concentrations of aliphatic amines could interfere in a third way — by competing with the diazo substrate for coordination to the metal and reducing catalytic turnover. Consistent with this suspicion, I showed that decomposition of the diazo compound is slowed in the presence of an amine substrate (**8**) and 35 mol % of the diazo compound was recovered after the same period of time.

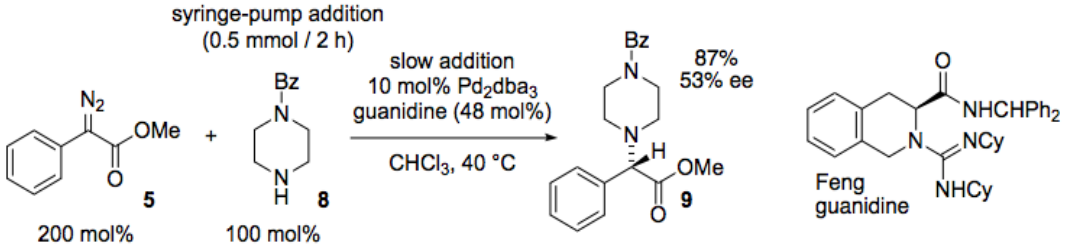
Scheme 6: The effect of amine inhibition of the catalyst



additive	recovered diazo **5**
0 mol% amine **8**	0%
100 mol% amine **8**	35%

By adding additional catalyst, Stan Hiew was able to achieve yields up to 98% and ee up to 53% with highly basic cyclic secondary amines such as *N*-benzoylpiperazine (**8**) using a chiral guanidine additive first reported by Feng without detailed synthetic procedures.[53] It isn't clear whether the guanidine is acting as a chiral ligand[53] or a chiral proton transfer agent under the conditions of the reaction, but relatively large quantities of the Feng guanidine are required for asymmetric induction.

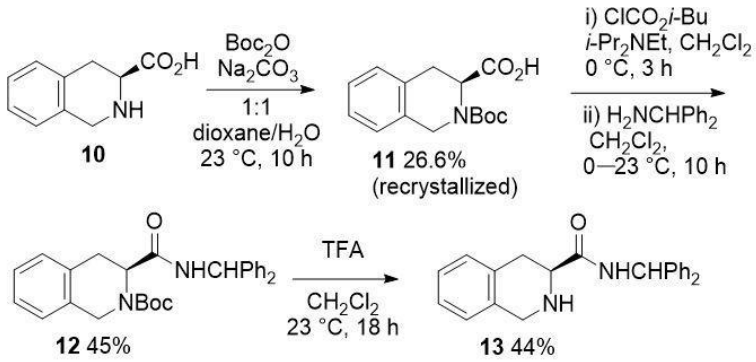
Scheme 7: Optimized dual-syringe pump setup utilizing a palladium(II) catalyst and Feng's guanidine (Hiew, unpublished work)



My first goal was to prepare more of the Feng guanidine, reproduce Hiew's results with *N*-benzoylpiperazine and then apply those conditions to electron deficient aromatic N–H heterocycles.

Hiew's route to the Feng guanidine involved a four-step synthesis with two recrystallizations (Scheme 8). Boc protection of the tetrahydroisoquinolinecarboxylic acid **10** afforded carbamate **11** in 26.6% yield after crystallization. The carboxylic acid **11** was converted to the mixed anhydride and then condensed with benzhydrylamine to afford amide **12** in 45% yield. The Boc group was removed with TFA to afford tetrahydroisoquinoline **13** in 44% yield.

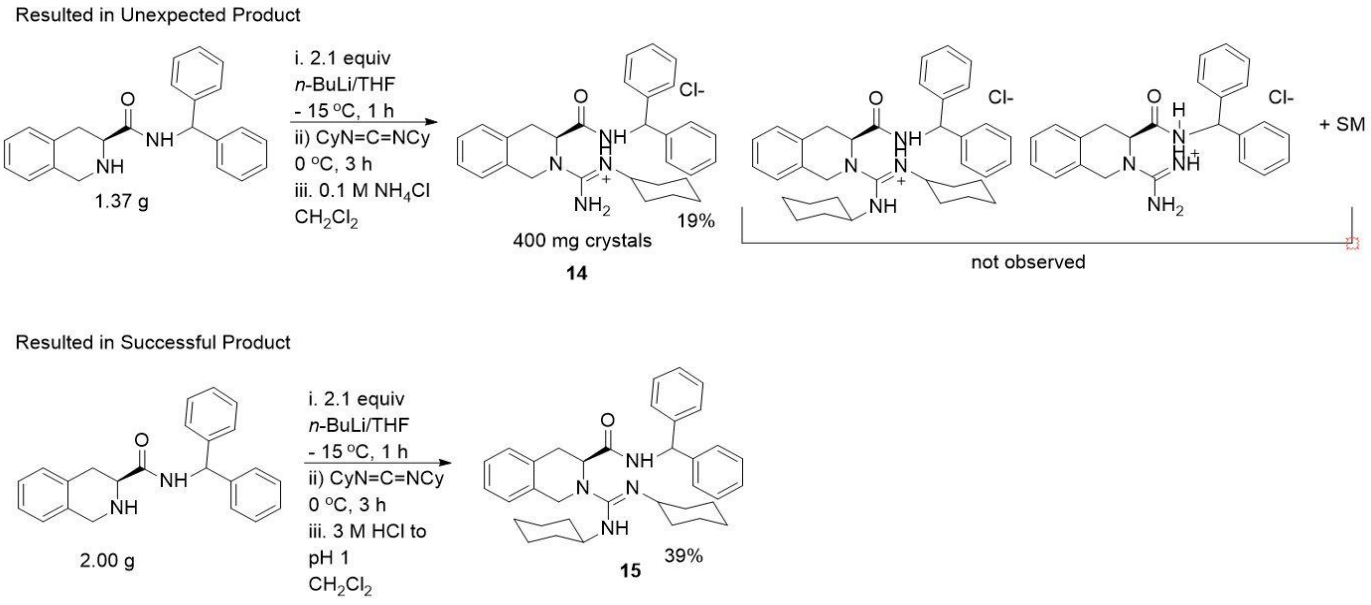
Scheme 8. Synthesis of the Feng guanidine



Conversion to the guanidine proved to be a major challenge. A previous member of the Van Vranken group employed a modified version of Feng's procedure, using sat. aq. NH_4Cl instead of aq. HCl to protonate the guanidine dianion. When I employed the Hiew procedure, adding additional water before separation, facile exchange of the guanidine cyclohexylamino groups occurred. None of the desired guanidine was observed by TLC and the only product isolated after chromatographic isolation of the guanidinium chlorides by silica gel chromatography and crystallization from acetone/CH_2Cl_2 was the mono-cyclohexylamine derivative **14** (19% yield). Exchange of guanidine pyrazolyl groups with dialkylamino groups has previously been in refluxing ethanol,[54,55] but the exchange of simple amino groups on guanidines typically requires aggressive conditions[55] Fortunately, reverting to the original Feng workup with 3 M HCl, generated the desired guanidine **15** in 39% yield. When an authentic sample of the Feng guanidine was subjected to the workup conditions: 88.0 mM NH_3, 166 mM NH_4Cl, 1:2:1 H_2O/THF/CH_2Cl_2, with no exchange of the cyclohexylamino groups being observed after 1 h.

Using material prepared by Jason Combs, a first year graduate student in the Van Vranken lab, amine **13** was converted to a dianion with 2.1 equiv *n*-BuLi and added to *N,N'*-dicyclohexylcarbodiimide, leading to complete consumption of starting material. The guanidinium salt **14•HCl** was extracted into CH_2Cl_2 and chromatographed on silica gel (as the the guanidinium salt). Guanidinium **14•HCl** was then freebased with 2 M NaOH to afford guanidine **14** in 39% yield. I have trained two undergraduates (Bill Mansiantima and An Do) who are bringing up more of the tetrahydroisoquinoline **13**.

Scheme 9: Guanidine exchange during synthesis of the guanidine[56]



While this work was in progress, another group published a highly efficient asymmetric copper-catalyzed insertion of alpha phenyl diazoesters into the N-H bonds of aliphatic substrates.[57] This led to the cessation of this project which my coworker Jason Combs has extended to tandem S-H insertion-Claisen rearrangements.

Chapter 2. Prediction of Mechanisms and Products for Organic Reactions

2.1. Background – Machine Learning and Prediction of Reaction Mechanisms.

Concurrent with my experimental work on N−H insertions, I have been working on a project to apply deep learning to the prediction of organic reaction mechanisms. The project is a collaboration with the group of Professor Pierre Baldi group in Computer Science called *Reaction Predictor*. The reactions that are provided by myself and my team are used for training the system by members of the Baldi group. Working with the Baldi group we incorporate changes to *Reaction Predictor* to improve the accuracy of prediction.

To date, machine learning has been heavily applied to the design of synthetic routes, driven by the existence of the large databases of (tens of millions) of reaction recipes in the REAXYS and SCIFINDER databases.[58,59,60,61] However, there is no database of elementary (single transition state) arrow-pushing reaction steps outside of the Baldi/Van Vranken collaboration. Without a database of elementary reaction steps, systems for prediction of reaction products have relied on expert rules,[60, 61, 62, 62] product stability[63] and quantum calculations[64] without exposing potential reactive intermediates. Machine learning has been applied to the identification of synthetic transformations.[65] Segler recently presented a system called Electro[66] that attempts to analyze transformations and extract mechanistic steps, but the results, such as S_N2 reactions on acid chlorides, are not realistic.

2.1.1. Improving the Ranking of Electrophilic and Nucleophilic Sites in Reactant Molecules

2.1.2. Reaction Predictor – Training a System to Predict Stepwise Mechanisms using Deep Learning

Problem: Identification of Top Source and Sink

In 2009, the Baldi group developed a system called *Reaction Predictor* that identifies the most reactive pairs of atoms in organic reactants and then ranks the various bonding combinations, equivalent to elementary reaction steps.[67,68,69] Reaction Predictor can search through mechanistic pathways until a target structure or target mass is identified. Highly accurate ranking of source-sink combinations has been achieved using a set of over 10,000 diverse pedagogical training reactions consisting of elementary reaction steps written in the text-based SMIRKS formalism with additional annotation to indicate changes in bonding corresponding to curved arrows. Each time the system is challenged with a new query, it identifies all reactive atoms (nucleophilic sources and electrophilic sinks) above a threshhold value and then ranks all possible combinations of nucleophilic source atoms attacking electrophilic sink atoms.

Since its initial development, *Reaction Predictor* has improved immensely. The ranking power of the system has improved in terms of predicting the correct mechanism from having an 82% success rate to a 91% success rate. The utility has suffered from the long search times associated with large reactants, mainly due to the system's inability to accurately rank source atoms and sink atoms. In the simple example of acetophenone and pentanamine, three reactions were predicted. The first has the primary nitrogen acting as a base, converting the ketone into an enolate. The second one has the lone pairs from the nitrogen attacking directly into the ketone, the most likely mechanism. The third prediction, however, shows the pi system from the benzene ring abstracting a proton from the primary amine – a reaction that could never occur. In looking at the main source sites, the nitrogen is correctly labeled as a source, but so is the benzene ring, and the lone pairs from the carbonyl oxygen. The sink sites are also problematic, as every proton is labeled as a possible sink.

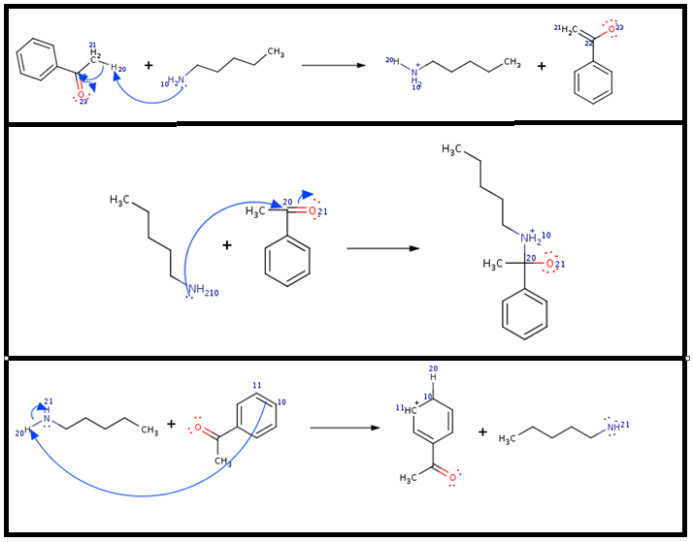


Figure 1A: Top three predicted mechanisms from two starting materials

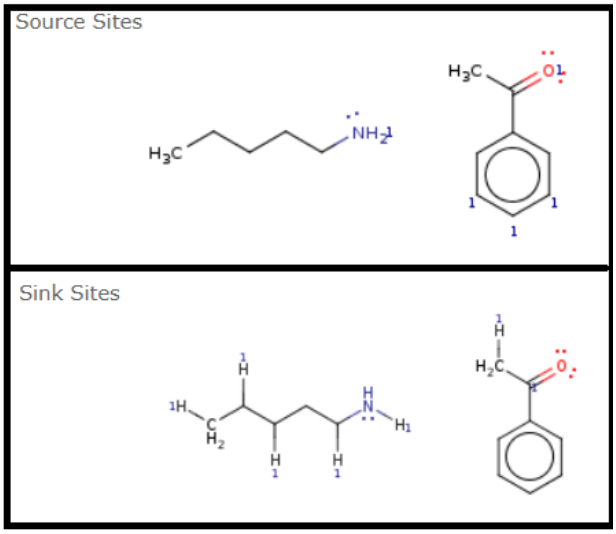


Figure 1B: Main source and sink sites

The main challenge was the lack of a complete data set that quantifies the reactivity of the key functional groups in organic chemistry. In order create a quantitative data set for source and sink reactivity, the nucleophilicity and electrophilicity parameters of Herbert Mayr were examined. The Mayr tables consist of an extensive range of parameters for hundreds of nucleophiles and electrophiles. Developed over the course of Professor Mayr's entire career, these tables were seen as a method to provide the initial data necessary to develop a range of reactivity for nucleophiles and electrophiles. Aaron Mood, had previously worked to demonstrate that readily calculated methyl ion affinity values, methyl cation affinity (MCA, eq 1)) and methyl anion affinity (MAA, eq 2) correlate with the solution phase kinetic reactivity parameters, which are only available from painstaking laboratory experiments.

$$^{-}Nu: \quad ^{+}CH_4 \xrightarrow{\Delta E} Nu{-}CH_3 \qquad MCA = -\Delta E \qquad (1)$$

$$^{+}Nu \quad :^{-}CH_3 \xrightarrow{\Delta E} Nu{-}CH_3 \qquad MAA = -\Delta E \qquad (2)$$

Mood had previously constrained many of his product geometries to reflect interactions with conjugated pi systems and then incorrectly determined free energies from these constrained geometries, which were not stationary points. I re-calculated MCAs based on unconstrained geometries and compared MCAs based on constrained energies, unconstrained free energies, and unconstrained electronic energies, ultimately leading to the published work.[70] Much of the following two sections on methyl ion affinity are taken from two publications – one published in the *Journal of Organic Chemistry* and the other being revised –on which I was a co-author.

2.2. Correlation of Calculated Methyl Cation Affinity with Mayr Nucleophilicity Parameters versus Correlation of Nucleophilicity Scales with Reaction Rate Constants

The concept of a nucleophilicity scale is controversial. Mayr has expressed skepticisim in a 2008 review where he stated, "a general nucleophilicity scale could not exist."[71] These constant selectivity relationships had been previously studied. The Swain-Scott equation (eq 3) is an example of the earliest quantification of nucleophile reactivities, primarily applicable to S_N2 reactions of organic compounds[72]:

$$\log k_X / k_{H2O} = sn_X \quad (3)$$

In this equation, n_X is the parameter that is characteristic of the nucleophilic reagent, and s is is a parameter dependent on the electrophile CH_3-X, where X was a halogen. While the Swain-Scott equation brought some initial success in showing a linear relationship between nucleophilic reagents, including solvents, with alkyl halides, a more general equation was later presented by Edwards (eq 4).[73]

$$\log k_X / k_{H2O} = \beta H + \alpha E_n \quad (4)$$

This equation strives to incorporate inorganic complexes, nucleophilic aromatic substitution reactions, and carbonyl addition reactions into the correlation but was applied to few examples. The basicity, represented as H, is related to the pK_{Ha} of the nucleophile. E_n is a measure of the 2-electron oxidation potential of the nucleophile. Both H and E_n are relative to water. Unfortunately, the parameters β and α are laid out to be factors that compare the sensitivity of the electrophile to the properties of the nucleophile.

While the equations above allowed for some success in measuring the relationship between molecules, Mayr stumbled upon a new method using carbocations. Initially looking to control Lewis-acid catalyzed additions of alkyl halides to alkenes, Mayr et al expanded their work with benzhydrylium cations with olefins.[74]

They found that the carbon-carbon-bond-forming steps were rate-determining, and that the relative reactivities of the alkenes did not depend on the reactivity of the benzhydrilium ions. However, the idea of constant selectivity relationships was shown to not be as generally applicable as initially desired, and a nucleophile-specific sensitivity parameter was developed in order to account for the observation that the reactivity of certain alkenes were altered based on the electrophiles than others.[79] This equation is one that is still in use today:

$$\log \boldsymbol{k}_{20^\circ} = s_N s_E (E + N) \quad (5)$$

The parameters E and N relate to log-scale electrophilicity and nucleophilicity; these values can be plotted on a scale to understand relative reactivity order.[75, 76] The other two parameters, s_N and s_E, are nucleophile- and electrophile-specific. The electrophile-dependent parameter s_E, however, is often close to unity, particularly for addition to sp^2 electrophiles, so the Mayr equation is frequently expressed in a simplified form (eq 6).

$$\log \boldsymbol{k}_{20^\circ} = s_N (E + N) \quad (6)$$

The principal value of the Mayr equation is that it establishes scales for nucleophilicity and electrophilicity that correlate with log-scale reactivity. Sadly, the expansion of the limited parameter set is likely to be limited by the retirement of Herbert Mayr, whose group determined rate constants for a large number of reactions. The Van Vranken group set out to determine if other types of data could be correlated with the Mayr reactivity parameters, and perhaps serve as a tool to predict parameters for which no kinetic data was yet available. That was motivation for my work.

Most organic chemists rationalize differences in nucleophilicity using pK_a values which are readily available from titrations. We attempted to correlate pK_{aH} values in water and DMSO (from Evans[77] and Bordwell[78]) with log k calculated from Mayr parameters: log $k = s_N \bullet (E + N)$. As many unique functional groups as possible were plotted against the predicted log k values for the nucleophilic attack against bis(4-methoxyphenyl)methylium. Unfortunately, pK_{aH} are not available for many of the nucleophiles in the Mayr database. The correlation was poor. As shown in Figure 5 below, the correlation between these different functional groups and the predicted R^2 value is poor (0.09 for H_2O and 0.35 for DMSO). Even if the correlation were good, the range of measurable pK_{aH} values is limited to approximately 15 orders of magnitude. Equilibrium pK_as can not be readily obtained for species with strongly acidic protons and or with highly reactive conjugate bases.

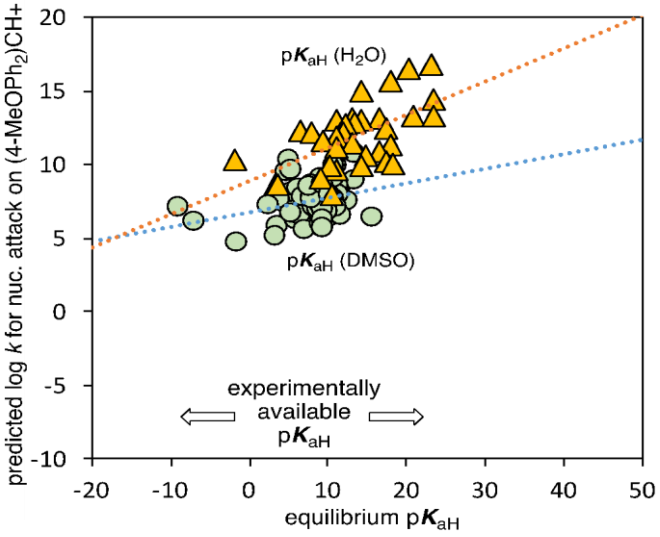


Figure 2: Correlation between pK_{aH} values and nucleophilicity is poor.

2.3. Solvation Improves Correlation of MCA* with Nucleophilicity.

Zipse and coworkers have shown that gas phase methyl cation affinities correlate with log k for within limited functional group classes.[79] We set out to explore whether methyl cation affinity would correlate with Mayr $N \bullet s_N$ across a broad range of functional group classes.

The correlation between MCA and Mayr $N \bullet s_N$ is particularly poor when both neutral and anionic nucleophiles are compared (Figure 6). When COSMO(∞) was used in the calculation of methyl anion affinity, the MCA gave surprisingly good linear correlation with Mayr $N \bullet s_N$ across both anionic and neutral nucleophiles. Going forward, we define MCA* as a methyl anion affinity calculated with a solvation model to distinguish it from the traditional notion of a gas phase methyl cation affinity (MCA). For many methyl adducts, the newly formed bond to the methyl group is not aligned with the HOMO of the nucleophile. We tested whether MCA*s would correlate better with Mayr $N \bullet s_N$ if the newly formed bond were constrained to align with the HOMO, usually a pi system. As shown in Figure

6B, constraints were applied to nucleophiles containing acyclic conjugated pi systems in the trajectory of the HOMO. These constraints were applied at a 90° torsion angle against the atoms in the nucleophile HOMO. The difference in the MCA*, however, turned out to be merely 2 kJ/mol, and therefore, constraints were not applied for the other MCA* calculations.

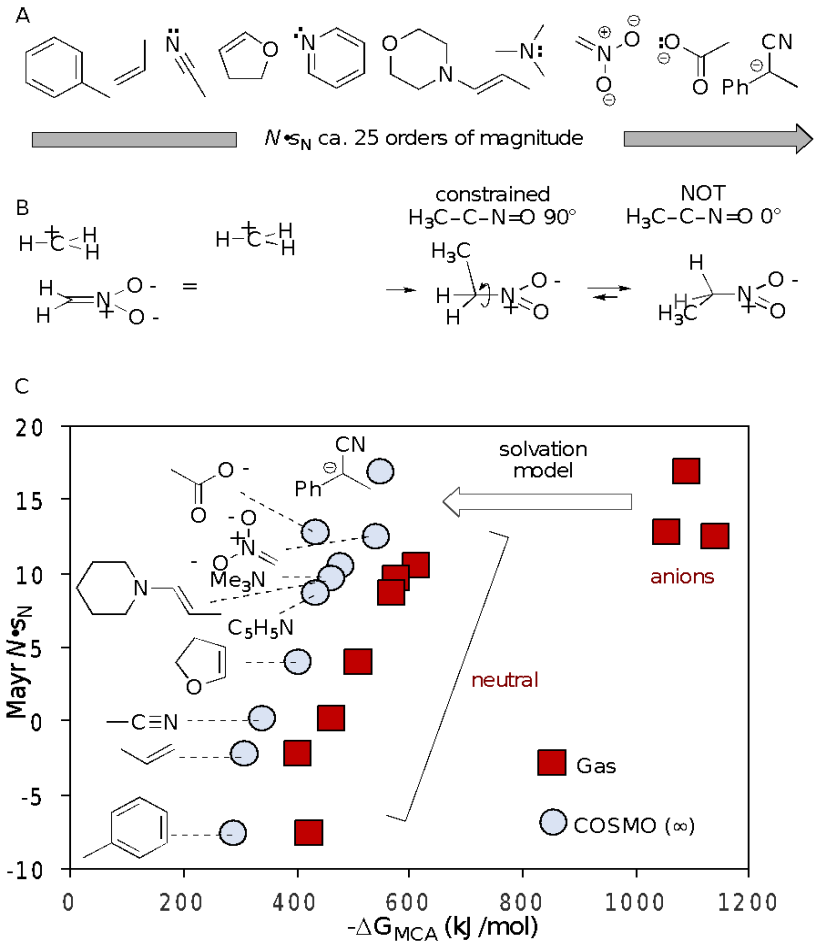
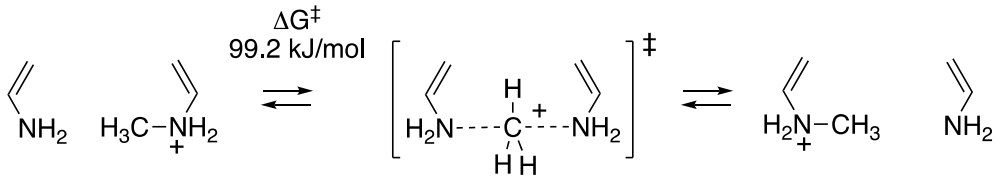
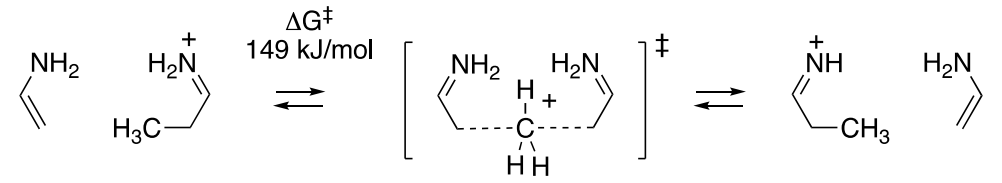


Figure 3: For a subset of ten molecules, it was determined that, while adding solvation improved the linear correlation between functional groups, adding constraints did not greatly affect the linear relationship.

Mayr and coworkers have shown that methyl cation affinity often gives erroneous prediction of regiochemistry in the reactivity of ambident nucleophiles.[80] Instead, they showed that (identity) intrinsic barriers for methyl transfer offer a more reliable clue to regiochemistry. For example, the intrinsic barrier for reaction at the nitrogen end of an enamine (eq 7) is lower than the intrinsic barrier for reaction at the carbon end (eq 8) at the MP2/6-311+G(2d,p) level of theory, in agreement with experiment. Generally, intrinsic barriers suggest that, for ambident nucleophiles, the atom further to the right of the periodic table will have the lower intrinsic barrier, and will react faster under kinetically controlled conditions. Intrinsic barriers are predictive of regiochemistry in kinetically controlled alkylations of aniline, enamine, nitronate anion, thiocyanate anion, cyanate anion, pyridone anion, and phenoxide anion. Intrinsic barriers predict the wrong regiochemistry in kinetically controlled alkylations of enolates, cyanide anion, nitrite anion, amides, and sulfinates.



(7)



(8)

We set out to test whether intrinsic barriers for a broad range of nucleophiles, both ambident and unident, would correlate with Mayr N•s_N. Intrinsic barriers calculated at the PBE0/def2-TZVP level of theory showed little correlation with Mayr N•s_N (Figure 7 below), suggesting that MCA* would offer a more promising approach to prediction of reactivity, while demanding a cautious application to the predicted reactivity of ambident nucleophiles where there is no information about regioselectivity.

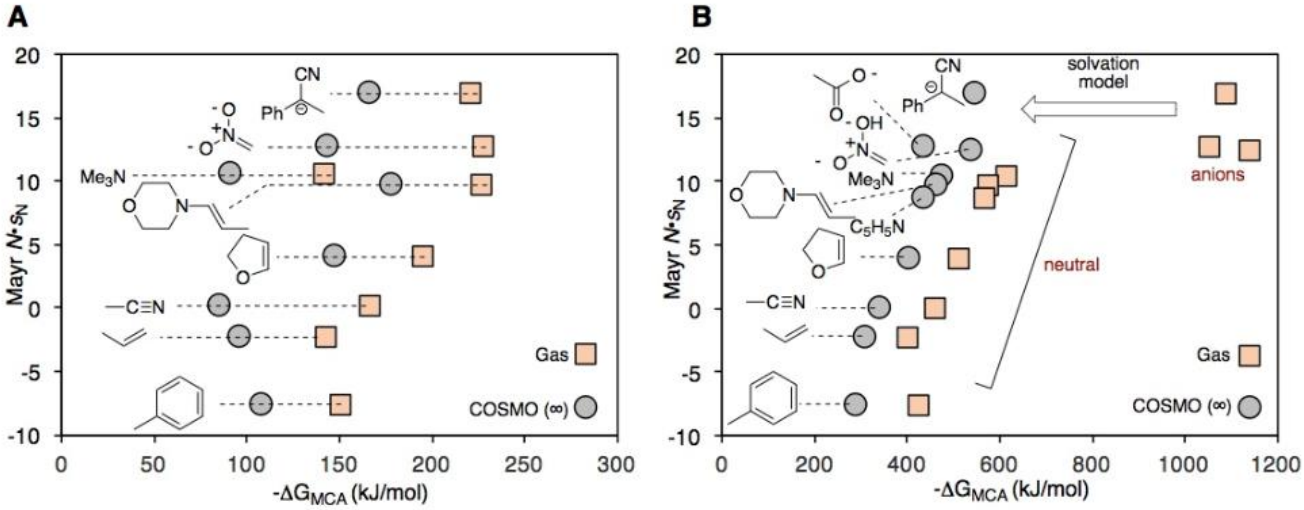


Figure 4. Intrinsic barriers for identity methyl transfer correlate poorly with Mayr N•s_N

Figure 7 shows two approaches to correlating energy values with Mayr N•s_N. The first, Figure 7A, involves measuring the intrinsic barriers as a method of measuring relative reactivity, but the method was proven to be negative. The correlation between intrinsic barriers and Mayr N•s_N, at least within the small data set, was shown to be poor. Figure 7B shows the correlation of MCA* with Mayr N•s_N. As in Figure 7A, energy measurements were carried out in both the gas phase, and with solvation. While proving to be poor in the gas phase, the correlation between MCA* and Mayr N•s_N was shown to improve greatly in linearity once solvation was added.

To better explore the correlation between MCA* and nucleophilicity, the MCA* of 98 diverse nucleophiles were calculated. Of these 98, several were measured with an anionic heteroatom acting as

a localized source of nucleophilicity, such as phenolate. In spite of the implied resonance present in said nucleophiles, the MCA* was calculated with the negative charge [primarily] located at the oxygen. Several of these anionic nucleophiles were measured as part of the set, and at first, the line appeared to contain excessive noise. Without letting this observation create a deterrence, I separated the anionic nucleophiles and generated an anionic nucleophile-specific trendline. With two functions in place, the correlation was perceptively more linear, with a final R^2 of 0.84 for the anionic heteroatoms and 0.88 for the carbon-based nucleophiles.

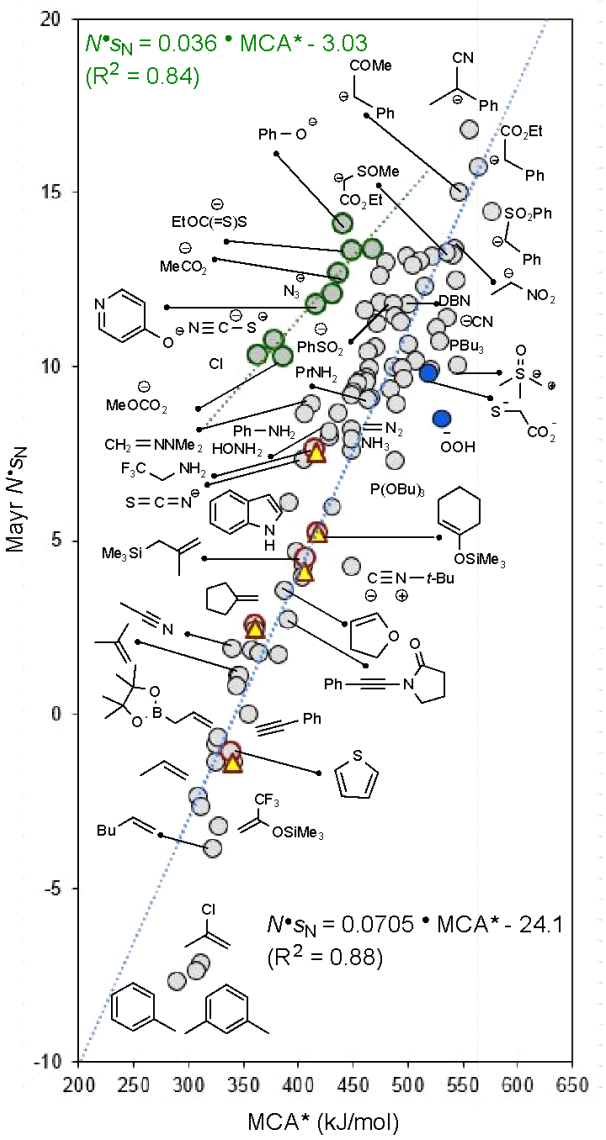


Figure 5: Distinctive correlation between MCA* and Mayr $N•s_N$.

Due to the degree of linearity seen from both trendlines, I continued by extrapolating said lines and applying them to canonical functional groups, such as ethane, that could never be measured in solution. Determining the primary reactive atom was fairly straightforward, but the interesting aspect came into play when calculating MCA* values for the atoms that do not represent the most reactive part. Various MCA* values were calculated in order to see whether or not said calculation method could be used in order to quantitatively distinguish between the reactivity of different atoms on a

molecule. If a molecule has two reactive points, they are usually a pi bond and a lone pair. In order to show which of these groups the methyl cation is adding to, it was important to constrain the angles to certain trajectories. From a certain angle close to 180^{o}, the LUMO of the methyl cation would be largely interacting with the lone pair of the functional group, as with a nitrile. From an angle closer to 90^{o}, however, the LUMO of the methyl cation would be largely interacting with the pi system of the functional group. MCA* values of a variety of canonical functional groups were measured through the extrapolation of the two functions established above.

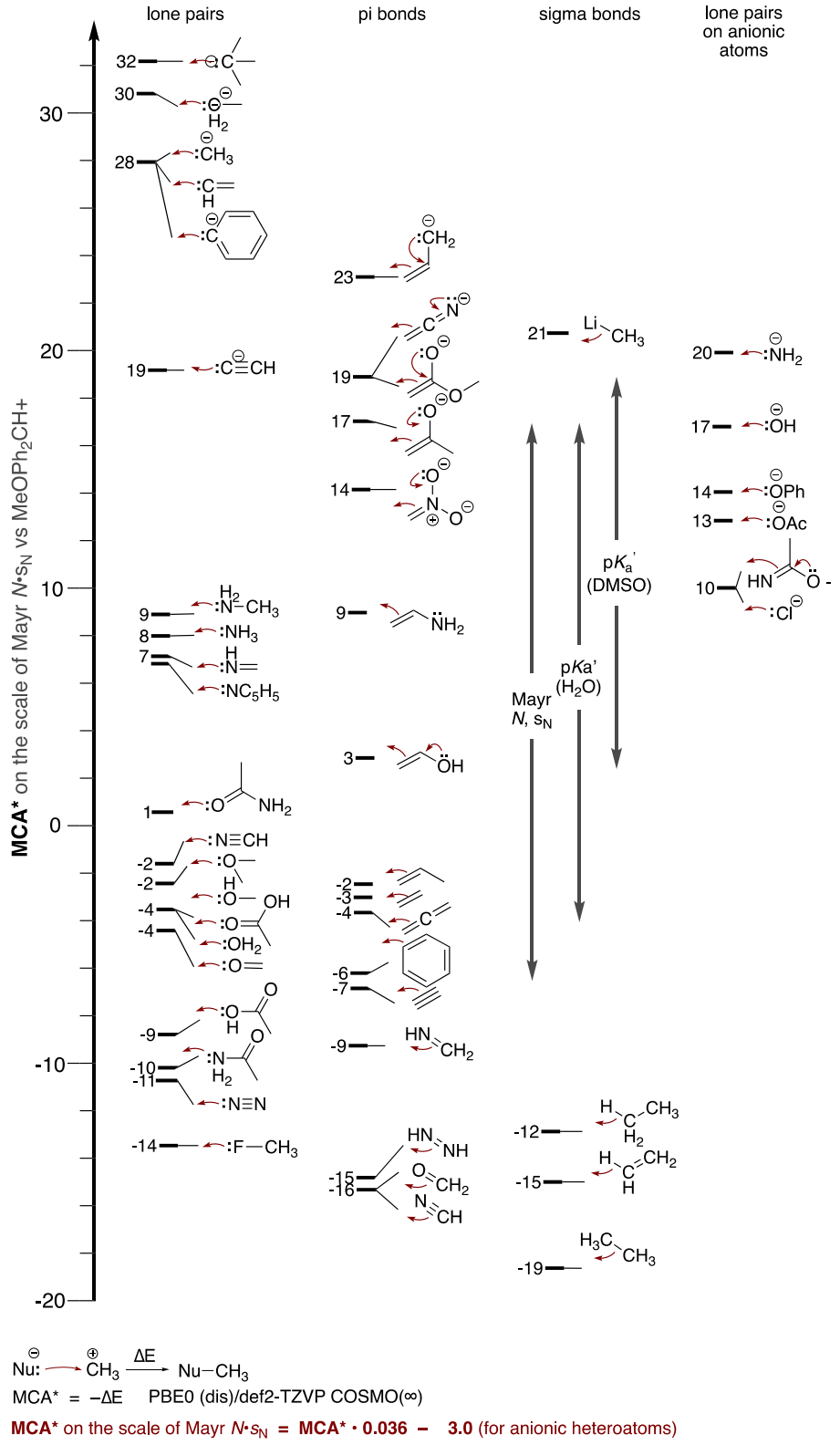


Figure 6: Extrapolation to canonical functional groups.

Some of the functional groups in Figure 7 correspond to molecular orbitals that are not the most reactive, for example, pi bonds instead of lone pairs, sigma bonds instead of pi bonds. In the figure below, the nitrile functional group is portrayed with two possible modes of attack: from the pi bond, and via the lone pair (Figure 8A). A series of measurements were made at varying trajectories in order to

determine the minimum energy. As the angle increased to 180°, the energy also increased, suggesting that the local minimum resides at a smaller trajectory between the methyl cation and the HOMO of the incoming nucleophile (Figure 8B).

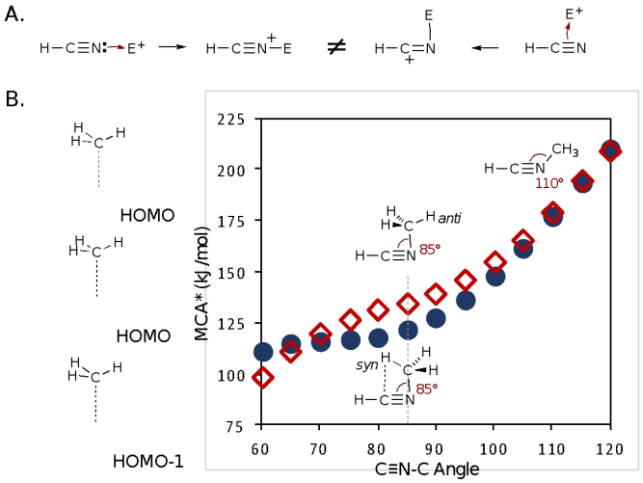


Figure 7: Varying trajectories and corresponding MCA* energies for hydrogen cyanide.

Not all canonical functional groups required measuring the trajectory within a range of bond angles. Some, however, required measurements at both the most reactive atom, and one that would be considered less reactive. When looking at the ambident reactivity of an amine functional group, it should be noted that both the lone pair of the nitrogen atom and the pi system have the potential to act as the nucleophile. When the MCA* of both of these atoms was calculated, the carbonyl oxygen was shown to have a much higher MCA* value than that of the sp2-nitrogen. This difference can be explained due to the resonance structure of acetamide. The structure on the left shows a lone pair on nitrogen, which is known to be more reactive than an unconjugated pi system. The structure on the right, however, shows the negative charge on oxygen, highlighting another reactive atom.

Figure 8: Resonance structures of acetamide.

The range of canonical functional groups, when using a methyl cation as the reference electrophile, was shown to span 50 orders of magnitude. While an impressive range, it is important to note that had a less reactive electrophile been chosen, such as cinnamonitrile, the range would have been considerably broader. Choosing a more reactive electrophile would have led to a much narrower range of reactivity. While this observation leads to the conclusion that the relative and absolute nucleophilicities change based on the electrophile involved, the rank order remains the same.

2.4. Correlation of Calculated Methyl Anion Affinity with Mayr Electrophilicity Parameters[81]

Organic chemistry is taught using a canonical set of functional groups. To a first order approximation, prediction of polar chemical reactivity rests on the deceptively simple act of quantifying the nucleophilicity and electrophilicity of every functional group that is present in the reactants. The ability to independently quantify nucleophilicity and electrophilicity largely eluded organic chemists until the pioneering work of Mayr and coworkers.[1] In a heroic body of work, Mayr's team has shown that solution-phase nucleophilicity and electrophilicity can be independently quantified using a log scale, allowing useful predictions of reaction rate constants using the equation $\log k_{20^\circ} = s_N s_E (E + N)$, where E and N are electrophilicity and nucleophilicity parameters, respectively, which quantify log-scale electrophilicity and nucleophilicity.[2,3] The variable s_N is a nucleophile-specific sensitivity parameter and s_E is close to unity.

The success of the Mayr equation centers around a focus on reactions that form bonds to carbon atoms, not, for example, Cs—O or F—Ge bonds; but a focus on carbon is not a significant limitation in the field of organic chemistry. The current set of electrophilicity parameters spans about 33 orders of magnitude and does not encompass highly reactive nor unreactive groups that serve as the pedagogical foundation for the field of organic chemistry: for example, *t*-butyl carbocation, ester carbonyls, amide carbonyls, acid chlorides, imines, alkyl halides, and carbon-carbon bonds.

Methyl anion affinity (MAA, Figure 1A) is related to the Mayr E parameter: $-\Delta E \propto$ Mayr E and is proportional to log k from the Mayr equation when s_E is 1. In early work, Mayr and Houk showed that calculated methyl anion affinity (AM1 H_f, gas phase) gives good linear correlation with the Mayr electrophilicity (Mayr E) for various cations: carbenium, iminium, and oxonium ions (Figure 1B).[4,5] It is promising that diverse cationic electrophiles fall on the same line. Subsequent work focused on individual functional groups but no attempt was made to correlate MAA with Mayr E for both neutral and cationic electrophiles. At higher levels of theory (MP2 and B3LYP), calculated methyl anion affinities also correlate linearly with Mayr electrophilicity for other classes of electrophiles: diarylallyl,[6,7] and azacarbenium ions.[8] Mayr and coworkers have also shown that neutral electrophiles such as ketones,[9] Michael acceptors.[10] and nitroarenes[11] give good linear correlation between calculated MAAs and Mayr E. In theory, the common slope of these lines should correlate with $1/s_N$ for a methyl anion.

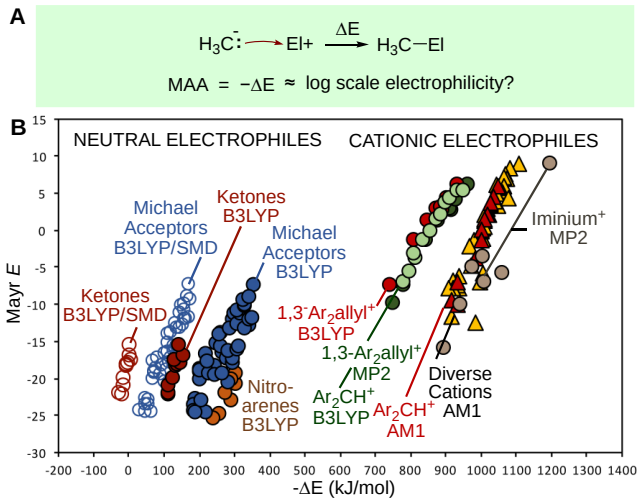


Figure 1. Previous correlations of MAA with Mayr E. (**A**) MAA is defined as the negative of the energy change for reaction of H_3C^- with electrophiles so that higher affinity correlates intuitively with higher energy, and higher kinetic electrophilicity. (**B**) Calculated MAA vs Mayr E generated lines that were dramatically different for neutral and cationic electrophiles. At higher levels of theory, different functional groups give different linear correlations.

At all levels of theory, the correlation between MAA and Mayr *E* are linear with similar slopes but it is sobering to see that, even at high levels of theory, different classes of functional groups such as ketones, Michael acceptors, nitroarenes, benzhydryl cations, allyl cations, and azacarbenium ions, fall on *different* lines, suggesting that it may be impossible to broadly correlate MAA with Mayr *E* across all of the canonical functional groups in organic chemistry. The SMD solvation model (DMSO) improves the linear correlation but did not lead to convergence of the lines for ketones and enones (Figure 1B). It is unclear if any theoretical treatment can be used to correlate the calculated methyl anion affinity with the solution phase electrophilicity of canonical organic functional groups.

2.5. Solvation Improves Correlation of MAA* with Electrophilicity. We set out to compare the COSMO(∞) solvation model with SMD(DMSO) for calculation of MAAs that would correlate with electrophilicity. For this study, we chose a functionally diverse test set of electrophilic species that span almost the entire range of experimentally determined Mayr *E* parameters (Figure 2A): pentan-2-one,[9] acrylonitrile,[10] butanal,[12] methyl vinyl ketone,[10] maleic anhydride,[13] dimethylmethyleneammonium cation,[14] tropylium cation,[14] methoxyphenylmethylium cation,[14] benzhydryl cation,[15] cumyl cation.[16] We restricted the product geometries to idealized transition state-like conformations by locking the newly formed H_3C-C bond 90° to the pi system being attacked for cumyl cation, methyl vinyl ketone, and acrylonitrile (Supporting Information). For example, methyl vinyl ketone was constrained to the *S*-cis conformation and newly formed H_3C—C bond in the product was constrained to be 90° to the original enone pi system, instead of the lower energy gauche conformation that minimizes $A_{1,3}$ strain (Figure 2B). For most electrophiles, the minimized product geometries tend to be geometrically and energetically similar to the transition state geometries.

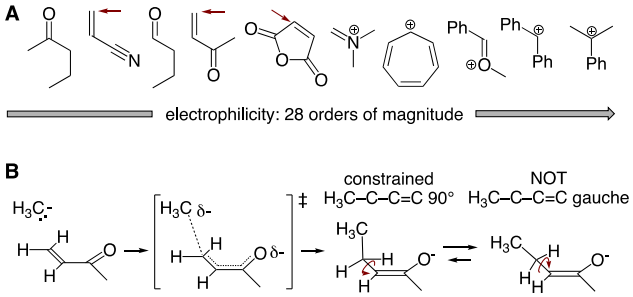


Figure 2. The test set. (**A**) Test set of electrophiles for which Mayr has reported experimental *E* parameters. Arrows show sites of attack. (**B**) MAAs were calculated by restricting products to transition state-like geometries.

The correlation between MAA and Mayr *E* is particularly poor when both neutral and cationic electrophiles are compared (Figure 3). When COSMO(∞) was used in the calculation of methyl anion affinity, the MAA gave surprisingly good linear correlation with Mayr *E* across both cationic and neutral electrophiles. Unexpectedly, on the full scale of measured Mayr *E* parameters, the SMD(DMSO) solvation was about as good as COSMO(∞) (R^2 0.95 vs 0.97, respectively) for calculation of MAAs that correlate with experimental electrophilicity although the correlation was not quite as good for ketones. Going forward, we define MAA* as a methyl anion affinity calculated with a solvation model to distinguish it from the traditional notion of a gas phase methyl anion affinity (MAA). MAA*s calculated with PBE0(disp)/def2-TZVP[17] gave about the same or better (R^2 0.97 vs. 0.96) linear correlation with Mayr electrophilicity compared to B3LYP/6-311++G(3df,2pd) but was significantly faster and was used for all subsequent calculations of MAA*s.

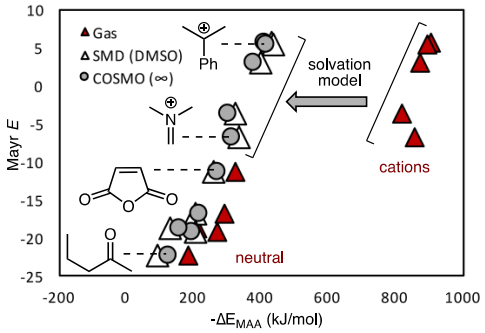


Figure 3. Solvation models lead to correlation of MAA with Mayr *E*. Inclusion of a solvation model like SMD or COSMO leads to good linear correlation between calculated MAA and Mayr *E* for both cationic and neutral functional groups.

2.5.1 MAA* Correlates with Mayr *E* Across A Broad Range of Electrophiles. Having shown that MAA*s correlate well with the Mayr *E* parameter for the test set of ten electrophiles, we set out to extend that analysis to a much fuller set of structurally diverse electrophiles. To date, Mayr and Ofial, have reported Mayr *E* parameters for over 319 electrophiles.[18] The set includes molecules from over 30 different classes of functional groups: i) empty p orbitals with π conjugation: benzylic cations,[14-16,19-23] allyl cations,[6,14] and the tropylium ion,[14] ii) empty σ* orbitals: ketones,[9] iminium

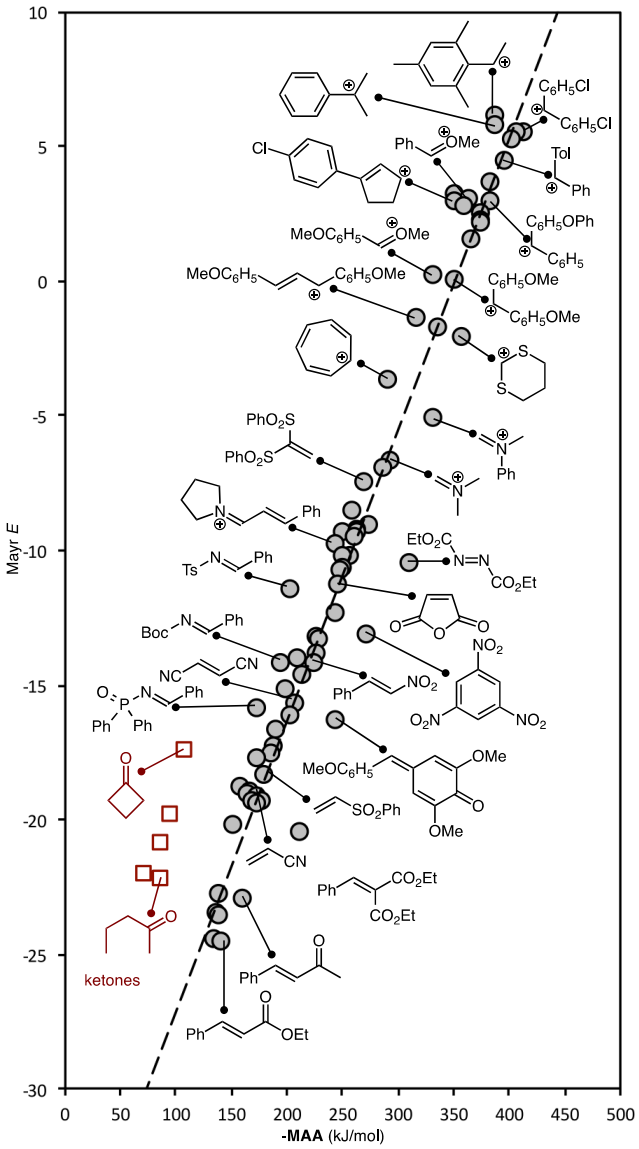


Figure 4. Correlation between experimental parameters and theory. There is good correlation between calculated MAA* (PBE0/def2-TZVP COSMO(∞)) and Mayr *E* parameter across a broad range of organic electrophiles.

ions,[14,24] oxacarbenium ions,[14] sulfacarbenium ions,[14] acyl imines,[25] tosyl imines,[25] phosphoryl imines,[25] quinone methides,[26,27] indolylmethylium ions,[28,29] alpha,beta-unsaturated ketones,[10,12,30-32] acrylates, [10] acrylamides, [10] cinnamates,[10] acrylonitriles,[10] alkylidenemalonates and related compounds,[33-35] maleates and related compounds,[13] alkylidene malonitriles,[13] nitrostyrenes,[36] cinnamyliminium ion derivatives,[37-41] benzylidene malonitriles,[42] quinones,[43,44] sulfonyl substituted ethylenes,[10,45,46] electrophilic arenes,[47-50] azocarboxylates,[51,52] and iii) empty σ* orbitals: chlorinating agents ($\sigma_{C\text{-}Cl}$*),[53] electrophilic fluorinating agents ($\sigma_{N\text{-}F}$*),[54] and electrophilic trifluoromethylsulfenylating agents ($\sigma_{X\text{-}S}$*).[55] Cationic pi-metal complexes[14,19,56,57] can be represented as empty p orbitals or empty σ*. We selected 75 of the 319 molecules (28 out of 32 functional groups) in the Mayr-Ofial database in an attempt to capture as many of the common canonical functional groups in organic chemistry as possible, excluding cationic pi-metal complexes, chlorinating, fluorinating, and sulfenylating reagents that react through S_N2-like processes as discussed below. MAA* was found to correlate linearly with the Mayr *E* parameters for nearly all of the functional groups, including the most reactive electrophiles (e.g., 1-(mesityl)ethenium ion, *E* = 6.04) and the least reactive electrophiles (e.g., cinnamonitrile, *E* = -24.60). Ultimately, it was

found that MAA* based on calculated free energies ($\mathbf{G}_{298}$) of unconstrained products gave the same correlation as MAA offering a more reproducible procedure (Figure 4, See Supporting Information). As discussed below, ketones are a notable exception to this linear correlation and were not used in the linear regression (R^2 0.96); the mean absolute error for Mayr *E* was ±1.2. Inclusion of an additional term for $(E_{HOMO}\text{-}E_{LUMO})^{-1}$ did not change R^2, using E_{HOMO} for a typical nucleophile, methylamine (Supporting Information).

Acetyl chloride and acetic anhydride do not generate stable tetrahedral intermediates at the PBE0/def2-TZVP level of theory. So we constrained the C–Cl distance to 1.800 Å in the adduct of acetyl chloride and we constrained the C–OAc distance to 1.395 Å in the adduct of acetic anhydride.[58,59]

No Mayr *E* parameter has ever been reported for an S_N2 displacement reaction at carbon although Mayr has reported *E* parameters for electrophiles in three other types of displacement reactions: chlorination by attack on Cl—C bonds, fluorination by attack on F—N bonds, and (trifluoromethyl)sulfenylation. Mayr has noted that "deviations are expected for S_N2 type reactions, where making one σ-bond is coupled with breaking another σ-bond."[14] *E* and s_E parameters can be estimated based on published kinetic data for S_N2 attack on $CH_3S^+R_2$ (*E* = -10),[2] CH_3I (*E* = -23)[2] and CH_3Br (*E* = -22)[60] in protic solvents and for CH_3I (*E* = -17),[61] in DMSO but the resulting *E* and s_E values are less reliable for calculation of rates than for other types of electrophiles.[61] Thus, it is not surprising that a plot of MAA* versus these renegade *E* parameters does not fall on the same line as the other types of electrophiles for which *E* parameters have been vetted, even after correction for translational entropy differences[62] and distortion energies.[63] It is ironic that the relative reactivity of S_N2 substrates, long studied with linear free energy relationships,[58] remains among the most pressing issues of our time in the field of physical organic chemistry.

Perhaps free carbenium ions and iminium ions give good agreement with MAA* because the experimental counterions like BF_4^- don't interact strongly with the reactants or the transition state for nucleophilic attack. Naked carbonyl compounds may end up fitting a second parallel linear function that could easily be applied to canonical electrophiles but we lack high quality Mayr *E* parameters for aldehydes and ketones at this lower end of the reactivity scale. The Mayr *E* parameters were determined in the presence of *t*-BuOH, and K^+ (which were not included in calculation of MAA*) with careful control experiments involving 18-crown-6 and phosphazene bases. To explore the issue of carbonyl activation we compared MAA*s calculated from electronic energies for various forms of cyclobutanone. The cyclobutanone•K^+•18-crown-6 adduct, modeled from the butanone crystal structure,[64] had an even lower MAA* than free cyclobutanone but the MAA* for cyclobutanone•K^+ was 31 kJ/mol higher than that for naked cyclobutanone. The MAA* for benzaldehyde•HO*t*-Bu was 21 kJ/mol higher than that of free cyclobutanone.

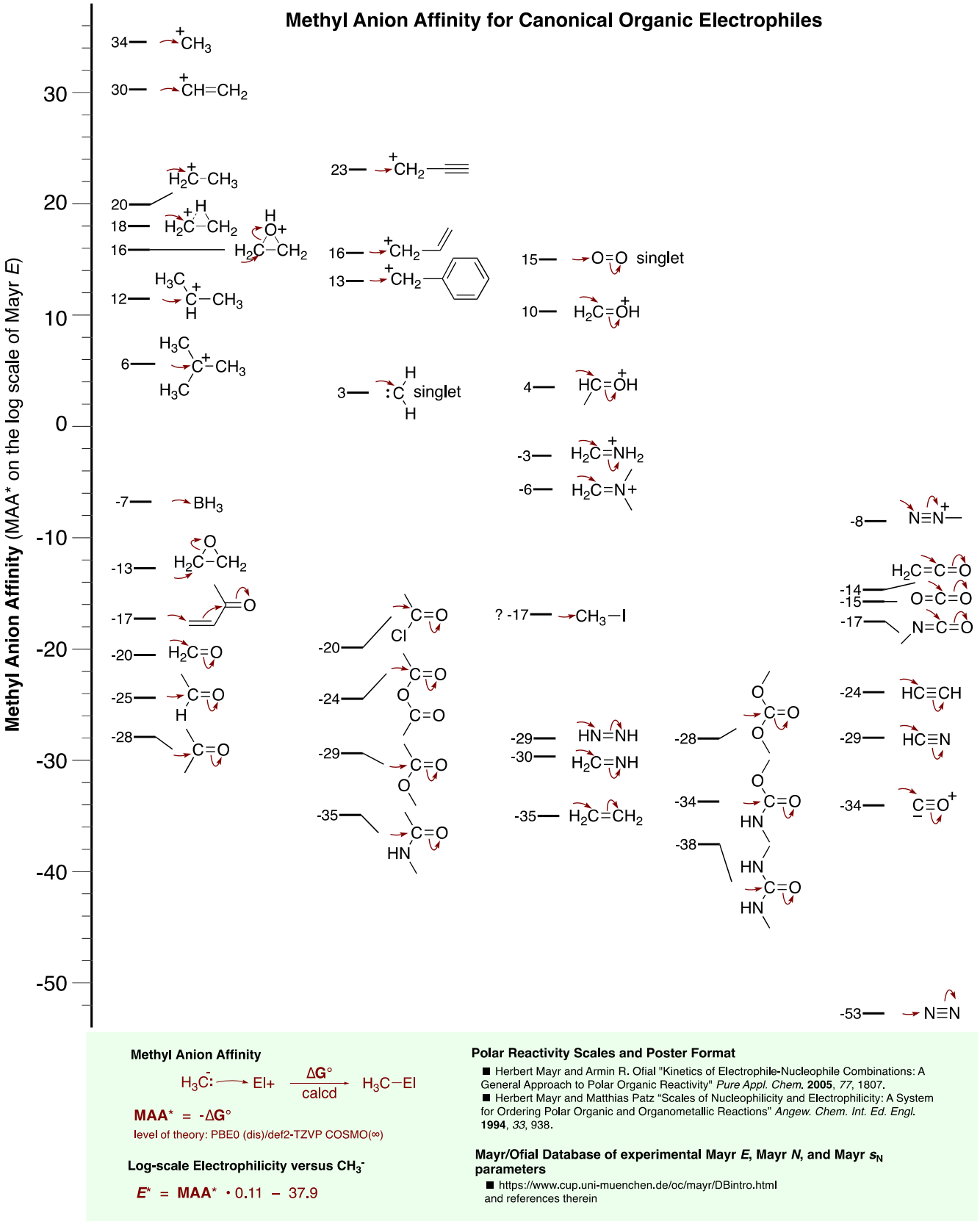


Figure 5. MAA* for the Canonical Electrophiles in Organic Chemistry. Methyl anion affinity, calculated with a solvation model, and plotted on the logarithmic Mayr *E* scale offers insight into electrophilicity.

Which value best reveals the "electrophilicity" of a carbonyl group: the MAA*, which matches the naked arrow-pushing depiction or a Mayr *E* parameter determined in the presence of species that are not depicted? After philosophical reflection, the answer is probably: both.

2.6. Quantifying the Reactivity of the Canonical Electrophiles on the Mayr Scale.

With good confidence in the linear correlation between MAA* (G_{298} from PBE0/def2-TZVP COSMO(∞)), we were then motivated to calculate MAA*s for unsubstituted canonical electrophiles (Figure 5) and convert them to the logarithmic Mayr *E* scale. On this scale, the C—C bond of ethane corresponding to the intrinsic barrier for attack on σ^*_{CC} has a value of -70. The acetylide cation and cyanide cation have values of +87 and +111, respectively. To the extent that methyl anion affinity, calculated with solvation, correlates with solution phase reactivity we begin to glimpse the breathtaking range of electrophilicity for canonical organic functional groups, as commonly depicted with arrow-pushing representations, on the scale of the Mayr *E* parameter.

2.7. Accuracy and Relevance of Methyl Anion Affinities.

If the correlation between MAA* and Mayr *E* is linear and the distribution is normal, *then* the furthest reaches of predicted electrophilicity, from $H_3C—CH_3$ to $^+C{\equiv}N$ will range from MAA* of -70 (±3) to +111(±5) on the Mayr *E* scale at a 95% confidence level. The structure of the Mayr equation, $\log k_{20^\circ} = s_N s_E (E + N)$ suggests that a truly independent scale of nucleophilicity and electrophilicity may be out of reach for some classes of polar reactions such as S_N2. Approaches involving deep learning could allow us to make predictions based on complex non-linear structure-reactivity relationships.[65-67]

What is the significance of electrophiles that can not be studied in solution? Reactions of the simplest carbenium ion, methyl cation[68] (H_3C^+), and even the supremely reactive cyanide cation ($^+C{\equiv}N$)[69,70] have been studied in gas phase experiments. Singlet methylidene,[71] and vinyl, ethyl and propargyl cations have been identified in atmospheres within our solar system and in interstellar environments.[72] Until there is a comprehensive database of *E* parameters that includes these gas phase species, it seems expedient to assess electrophilicity on the familiar Mayr scale. Allyl cation and less reactive carbenium ions have been studied in solution on our own planet under superacid conditions.[73] The ubiquitious C-C σ^* is at the bottom end of the electrophilicity scale. Understanding the kinetic electrophilicity of these species on a common scale inspires us to think about their potential reactivity on Earth and beyond.

MAA* is an economical tool for estimating the log scale electrophilicity of the canonical organic functional groups, covering a staggering range of 180 orders of magnitude. More such experimental determinations of Mayr *E* parameters are needed to explore the limits of the linear correlation with methyl anion affinity.

Chapter 3. Prediction of Radical Mechanisms

In this early stage of development we are focused on identification of products that arise from photooxidation of volatile atmospheric terpenes (isoprene, pinenes, limonene, etc) exposed to HO•, O_2, and N=O.[82] Large amounts of data are available from mass spectrometry but there are no tools for assigning masses to plausible structures.[83] Analyses are typically limited to atomic ratios such as C/O and C/N. There is no limit to the size or number of reactants in the search query, but we are focusing on small molecules at first because large search trees and large molecules can require weeks of computational time at this stage of development.

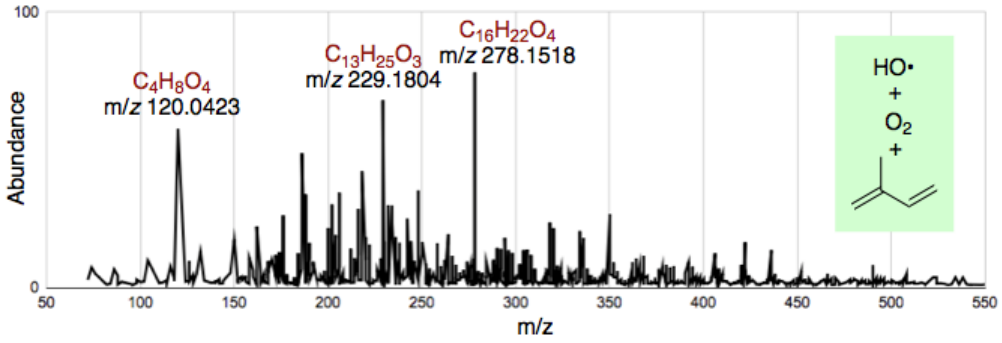


Figure 1. Experimental studies of isoprene oxidation generate products with masses that can not be readily assigned to structures. All exact masses and molecular formulae are available but not shown. [Neutral masses derived from $[M+H]^+$ from PTR-TOF]

Atmospheric chemists have been working to create databases of reaction steps with rate constants,[84, 85, 86] but the reactions in these databases are a mix of elementary and composite processes. The original training set for *Reaction Predictor*'s radical prediction module contained only about 100 arrow-pushing reactions using half arrows. I have led a team in the creation, curation, and testing of a pedagogical set of radical reactions for training *Reaction Predictor*. The team involves one of my undergraduate mentees, Romesa Khan, and a graduate student in atmospheric chemistry from the Carlton group, Mooji Boldbaatar and her mentees. We started by transcribing over 1,000 radical mechanistic steps from sophomore organic chemistry textbooks, including many redundant reactions. This process of curation and augmentation involves testing the system with reaction queries, and analyzing the set of predicted arrow-pushing steps for weaknesses in source/sink identification or arrow-pushing, and then creating new training reactions to address those weaknesses.

As shown in Fig. 2 the process of transcribing training reactions from books or research journals is easily understood by beginning students in chemistry. The process involves four steps: 1. drawing reactions using Chemdraw, 2. copying the reaction as a SMILES string, 3. mapping the atoms with the public tool *ReactionMap*[87] and 4. manually adding arrows with the SMIRKS formalism. However, in most cases, the reactants can be entered as a query into *Reaction Predictor* and the correct training will appear as one of the proposed mechanistic steps in a SMIRKS format that can be directly copied and pasted into the training set.

In theory, any user who wishes to strengthen *Reaction Predictor* in their own area of chemistry could write SMIRKS training reactions and submit them for training.

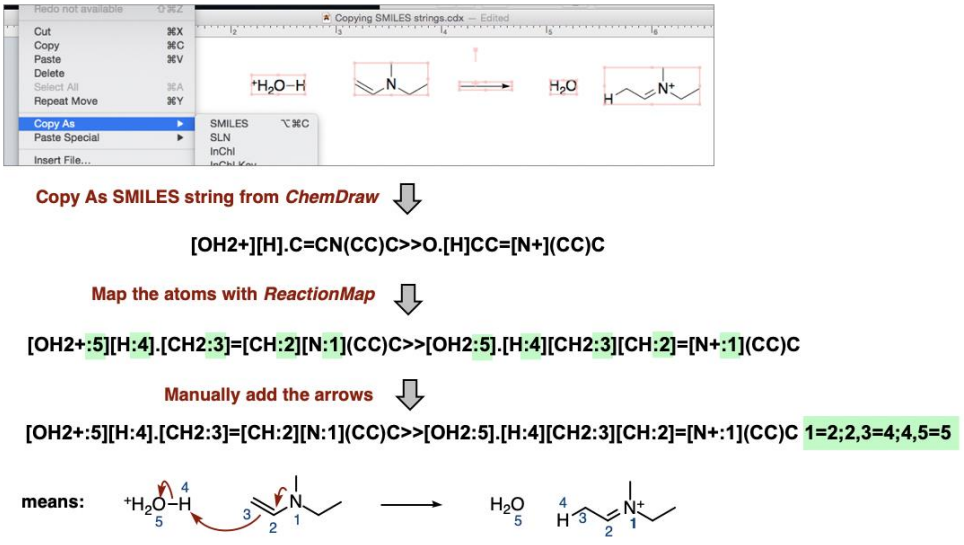


Figure 2. Stages in creation of training reactions are easily understood by students

After training, with the expanded training set we found that arrow-pushing steps from the training set were recovered in the top 5 predicted reactions with 90% accuracy (Table 1). Arrow pushing steps taken from a naïve textbook Loudon & Parise *Organic Chemistry*, 6th Ed, not used for training, were found in the top 5 predicted reactions 80% of the time. Reaction queries pushing queries taken from Chapter 1 of an advanced book, Perkins' *Radical Chemistry: The Fundamentals*, the arrow-pushing steps were found in the top 5 predicted reactions 60% of the time. This is good considering the elementary nature of the reactions. Several hundred additional radical training reactions have been taken from atmospheric isoprene chemistry.[57] As more training reactions are added from advanced books and research literature, we will continue to challenge the system and quantify the accuracy of prediction.

Table 1. Radical Step Prediction After Training with 677 Radical Reactions

Query Source	Top 5	Other
Recovered from Training Set	90% in Top 5	95% in Top 6
Loudon textbook	80% in Top 5	85% in Top 10
Perkins Book	60% in Top 5	80% in Top 10

To highlight the structural complexity addressed and the graphical output for radical reactions, we show the results of single-step prediction from a recently published article on the oxidative cleavage of tetrahydroisoquinolines (Fig. 7).[73] Among the 76 predicted steps, the expected thermolysis of the nitrite N−O bond was not in the predicted reactions, but notably the system recognized the weak N−O bond and the 4thpredicted process generated the benzylic radical, t-BuO-H and •NO in a single step. No similar reactions were present in the training set.

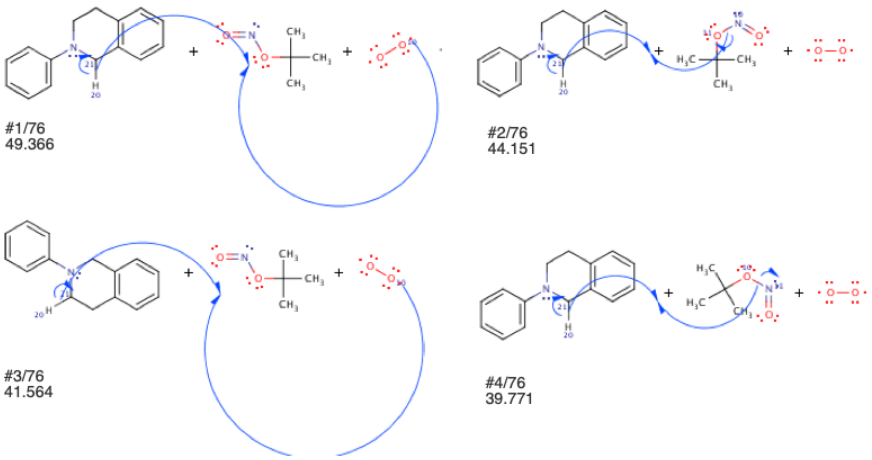


Figure 3. Prediction of steps from a recently published radical reaction.

Mooji Boldbaatar and her team have been encoding radical reactions for the atmospheric radical reactions of isoprene with the reactive atmospheric species HO•, O_2, NO, and HOO.[57] I was responsible for checking those mechanistic steps, which are being be used by Amin Tavakoli in the Baldi group to train *Reaction Predictor*.

I have also improved the system to predict known radical processes. In working with Amin Tavakoli from the Baldi group and Mooji Boldbaatar from the Carlton group, we have made progress on modifying the training parameters in order to maximize the accuracy of the system. To date, our team has created a database of over 2000 mechanistic radical steps which are being used to train Reaction Predictor to predict gas-phase atmospheric radical reactions of isoprene and other volatile plant-based terpenes such as myrcene, limonene, alpha-pinene and beta-pinene.

Chapter 4. Summary

In summary, I have worked on three projects: i. development of an asymmetric route to xantholipin B, ii. asymmetric insertion of carbene groups into aliphatic N−H bonds, iii. and developing a pedagogical training set for organic radical chemistry for prediction of reaction products and mechanistic pathways. During the six months that I worked for Professor Sergey Pronin, I optimized conditions for a tri-*n*-butylstannane addition into MOM-protected salicylaldehyde to produce a sensitive stannylalcohol. During my time in the Van Vranken group, I have made a chiral guanidine catalyst through a four-step synthetic route, and have worked on scales for electrophilicity and nucleophilicity.

Working on a collaborative project to predict reaction mechanisms I employed electronic structure calculations to demonstrate that readily methyl cation affinities calculated with a COSMO solvation model (MCA*), correlate with solution phase nucleophilic addition rates Mayr N•s_N and used it to quantify the reactivities of the common nucleophilic functional groups, from carbon-carbon bonds to t-butyl anions. Likewise I demonstrated that readily methyl anion affinities calculated with a COSMO solvation model (MAA*), correlate with solution phase electrophilic addition rates (Mayr *E*) and used it to quantify the reactivities of the common electrophilic functional groups, from carbon-carbon bonds to cyanide cations. Finally, using tools of chemoinformatics, I helped to develop an extensive database of thousands of mechanistic radical steps used to train the first dedicated variant of *Reaction Predictor* using deep learning. Reaction Predictor can now identify reactive atoms with high accuracy and will be used to identify products from complex atmospheric reactions of volatile organic compounds.

References

[1] Terui, Y.; Chen, Z.; Ando, T.; Chu, Y.; Kawashima, A. " Xanthone compound" WO2002027010A1, 2002.

[2] Kuenstner, E. J. "Synthetic Approaches to the Paxilline Indole Diterpenes and Polycyclic Xanthones" Ph.D. Thesis, UC Irvine, 2019.

[3] Fischer, C. Homung, A.; Holzenkämpfer, M.; Heizer, S.; Priefert, H.; Ulf-Dietrich, R.; Vierling, S.; Wohler, S.-E. "Novel lysolipin derivatives" WO2007079715A2, 2007.

[4] He, A.; Falck, J. R. "Synthesis of Enantioenriched α-(Hydroxyalkyl)-tri-*n*-butylstannanes." *Angew. Chem. Int. Ed.* **2008**, 120, 6688–6691.

[5] Bora, U. "An Eco-Friendly and Mild Process for Deacetylation Reactions in Water." *Asian J. Chem.* **2011**, 23, 941–942.

[6] Shcherbinin, V. A.; Konshin, V. V. "Lewis acid-mediated mono- and bis-addition of *C*-nucleophiles to 1,3-dioxolan-4-ones." *Tetrahedron Lett.* **2018,** 59, 3005−3009.

[7] Coric, I.; Kim, J. H.; Vlaar, T.; Patil, M.; Thiel, W.; List, B. "Brønsted Acid Catalyzed Asymmetric SN2-Type O-Alkylations." *Angew. Chemie. Int. Ed.* **2013**, 52, 3490–3493.

[8] Ihara, E.; Haida, N.; Iio, M.; Inoue, K. "Palladium-Mediated Polymerization of Alkyl Diazoacetates To Afford Poly(alkoxycarbonylmethylene)s. First Synthesis of Polymethylenes Bearing Polar Substituents." *Macromolecules*, **2003**, *36*, 36–41.

[9] Zhu, Y.; Liu, X.; Dong, S.; Zhou, Y; Li, W.; Lin, L.; Feng, X. "Asymmetric N−H Insertion of Secondary and Primary Anilines under the Catalysis of Palladium and Chiral Guanidine Derivatives." *Angew. Chem. Int. Ed.* **2014**, 53, 1636–1640.

[10] Li, J.; Su, Z.; Wang, J.; Hu, C. "Mechanistic investigations on asymmetric N–H insertion of amines catalyzed by palladium-chiral guanidine complex." *J. Catalysis* **2018**, *364*, 426–436.

[11] Moody, C. "Enantioselective insertion of metal carbenes into N−H bonds: a potentially versatile route to chiral amine derivatives." *Angew. Chem. Int. Ed.* **2007**, 46, 9148–9150.

[12] Saito, H. "Highly enantioselective X−H (X = N, O) insertion reactions by metal complexes." *J. Synth. Org. Chem. Jpn.* **2008**, *66*, 714–715.

[13] Zhao, X.; Zhang, Y.; Wang, J. "Recent developments in copper−catalyzed reactions of diazo compounds" *ChemComm.* **2012**, 48,10162–10173.

[14] Zhu, S.-F.; Zhou, Q.-L. "Transition−Metal−Catalyzed Enantioselective Heteroatom-Hydrogen Bond Insertion Reactions" *Acc. Chem. Res.* **2012**, *45*, 1365–1377.

[15] Maier, T. C.; Fu, G. C. "Catalytic Enantioselective O−H Insertion Reactions" *J. Am. Chem. Soc.* **2006**, *128*, 4594–4595.

[16] Lee, E. C.; Fu, G. C. "Copper−Catalyzed Asymmetric N−H Insertion Reactions: Couplings of Diazo Compounds with Carbamates to Generate α−Amino Acids." *J. Am. Chem. Soc.* **2007**, 129, 12066–12067.

[17] Chen, C.; Zhu, S.-F.; Liu, B.; Wang, L.-X.; Zhou, Q.-L. "Highly Enantioselective Insertion of Carbenoids into O−H Bonds of Phenols: An Efficient Approach to Chiral α-Aryloxycarboxylic Esters" *J. Am. Chem. Soc.* **2007**, *129*, 12616–12617.

[18] Osako, T.; Panichakul, D.; Uozumi, Y. "Enantioselective Carbenoid Insertion into Phenolic O−H Bonds with a Chiral Copper(I) Imidazoindolephosphine Complex" *Org. Lett.* **2012**, *14*, 194–197.

[19] Song, X.-G.; Zhu, S.-F.; Xie, X.-L.; Zhou, Q.-L. "Enantioselective Copper−Catalyzed Intramolecular Phenolic O−H Bond Insertion: Synthesis of Chiral 2−Carboxy Dihydrobenzofurans, Dihydrobenzopyrans, and Tetrahydrobenzooxepines" *Angew. Chem. Int. Ed*. **2013**, *52*, 2555–2558.

[20] Xie, X.−L.; Zhu, S.−F.; Guo, J.-X.; Yan, C.; Zhou, Q.-L. "Enantioselective Palladium−Catalyzed Insertion of α-Aryl-α-diazoacetates into the O−H Bonds of Phenols" *Angew. Chem. Int. Ed*. **2014**, *53*, 2978–2981.

[21] Harada, S.; Tanikawa, K.; Homma, H.; Sakai, C.; Ito, T.; Nemoto, T. "Silver−Catalyzed Asymmetric Insertion into Phenolic O−H Bonds using Aryl Diazoacetates and Theoretical Mechanistic Studies" *Chem − Eur. J.* **2019**, *25*, 12058–12062.

[22] Liang, Y.; Zhou, H.; Yu, Z.-X. "Why Is Copper(I) Complex More Competent Than Dirhodium(II) Complex in Catalytic Asymmetric O−H Insertion Reactions? A Computational Study of the Metal Carbenoid O−H Insertion into Water" *J. Am. Chem. Soc.* **2009**, 131, 49, 17783–17785.

[23] Liu, Y.; Luo, Z.; Zhang, J. Z.; Xia, F. "DFT Calculations on the Mechanism of Transition−Metal−Catalyzed Reaction of Diazo Compounds with Phenols: O–H Insertion versus C–H Insertion" *J. Phys. Chem. A* **2016**, *120*, 6485–6492.

[24] Zhu, S.-F.; Chen, W.-Q.; Zhang, Q.-Q.; Mao, H.-M.; Zhou, Q.-L. "Enantioselective Copper−Catalyzed O−H Insertion of α-Diazo Phosphonates" *SYNLETT* **2011**, 919–922.

[25] Zhu, S.-F.; Cai, Y.; Mao, H.-X.; Xie, J.-H.; Zhou, Q.-L. "Enantioselective iron−catalysed O−H bond insertions" *Nature Chemistry* **2010**, *2*, 546–551.

[26] Zhu, S.-F.; Song, X.-G.; Li, Y.; Cai, Y.; Zhou, Q.-L. "Enantioselective Copper−Catalyzed Intramolecular O−H Insertion: An Efficient Approach to Chiral 2−Carboxy Cyclic Ethers" *J. Am. Chem. Soc.* **2010**, *132*,16374 –16376.

[27] Zhu, S.-F.; Chen, C.; Cai, Y; Zhou, Q.-L. "Catalytic asymmetric reaction with water: Enantioselective synthesis of alpha−Hydroxyesters by a copper−carbenoid O−H insertion reaction" *Angew. Chem. Int. Ed.* 2008, 47, 932−934.

[28] Li, M.; Chen, M.; Xu, B.; Zhu, S.-F. "Enantioselective O−H Bond Insertion of alpha−Diazoketones with Alcohols Cooperatively Catalyzed by Achiral Dirhodium Complexes and Chiral Spiro Phosphoric Acids" *Acta Chimica Sinica* **2018**, *76*, 883–889.

[29] Xu, B.; Zhu, S.-F.; Zhang, Z.-C.; Yu, Z.-X.; Maa, Y.; Zhou, Q.-L "Highly enantioselective S–H bond insertion cooperatively catalyzed by dirhodium complexes and chiral spiro phosphoric acids" *Chem. Sci.* **2014**, *5*, 1442 –1448.

[30] Galardon, E.; Roue, S.; Le Maux, P.; Simonneaux, G. "Asymmetric Cyclopropanation of Alkenes and Diazocarbonyl Insertion into S−H Bonds Catalyzed by a Chiral Porphyrin Ru(II) Complex." *Tetrahedron Lett.* **1998**, *39*, 2333–2334.

[31] Ratcliffe, R.W.; Salzmann, T.N.; Christensen, B.G. “A novel synthesis of the carbapen-2-em ring system” *Tetrahedron Lett.* **1980**, 21, 31–34.

[32] Yi, N.; Xin, G.; Hu, W.; Liu, S. “Asymmetric N−H Insertion Reaction of alpha−Diazoesters and Carbamates Co-catalyzed by Dirhodium Acetate, Sulfonic Acid and Chiral Sulfonamide Urea” *Chin. J. Org. Chem.* **2014**, 34, 107–111.

[33] Lee, E. C.; Fu, G. C. “Copper−Catalyzed Asymmetric N−H Insertion Reactions: Couplings of Diazo Compounds with Carbamates to Generate α−Amino Acids.” *J. Am. Chem. Soc.* **2007**, 129, 12066–12067.

[34] García, C.; McKervey, A.; Ye, T. “Copper-Catalyzed Asymmetric N−H Insertion Reactions: Couplings of Diazo Compounds with Carbamates to Generate α-Amino Acids” *Chem. Commun*. **1996**, 1465–1466.

[35] Xu, B.; Zhu, S.-F.; Zuo, X.-D.;Zhang, Z.-C.; Zhou, Q.-L. "Enantioselective N−H Insertion Reaction of α-Aryl-α-Diazoketones: An Efficient Route to Chiral α-Aminoketones" *Angew. Chem. Int. Ed.* **2014**, 53, 3913–3916.

[36] Xu, B.; Zhu, S.-F.; Xie, X-L.; Shen, J.-J.; Zhou, Q-L. “Asymmetric N−H Insertion Reaction Cooperatively Catalyzed by Rhodium and Chiral Spiro Phosphoric Acids” *Angew. Chem. Int. Ed.* **2011**, *50*, 11483–11486.

[37] Guo, J.-X.; Zhou, T.; Xu, B.; Zhu, S.-F.; Zhou, Q.-L.”Enantioselective synthesis of α−alkenyl α−amino acids via N–H insertion reactions” *Chem. Sci*. **2016**, *7*, 1104–1108.

[38] Wang, X.-C.; Song, X.-S.; Guo, L.-P.; Qu, D.; Xie, Z.-Z.; Verpoort, F.; Cao, J. “Mechanistic Insight into Asymmetric N–H Insertion Cooperatively Catalyzed by a Dirhodium Compound and a Spiro Chiral Phosphoric Acid” *Organometallics.* **2014**, *33*, 4042–4050.

[39] Liu, B.; Zhu, S-F.; Zhang, W.; Chen, C.; Zhou, Q−Z. “Highly Enantioselective Insertion of Carbenoids into N–H Bonds Catalyzed by Copper Complexes of Chiral Spiro Bisoxazolines.” *J. Am. Chem. Soc.* **2007**, *129*, 5834–5835.

[40] Zhu, S-F., Xu, B.; Wang, G.P.; Zhou, Q.-L. “Well−Defined Binuclear Chiral Spiro Copper Catalysts for Enantioselective N–H Insertion.” *J. Am. Chem. Soc.* **2012**, *134*, 436–442.

[41] Hou, Z.; Wang, J.; He, P.; Wang, J.; Qin, B.; Liu, X.; Lin, L.; Feng, X. “Highly Enantioselective Insertion of Carbenoids into N−H Bonds Catalyzed by Copper(I) Complexes of Binol Derivatives” *Angew. Chem. Int. Ed.* **2010**, *49*, 4763–4766.

[42] Maux, P.; Simonneaux, G. “Enantioselective insertion of carbenoids into N–H bonds catalyzed by chiral bicyclobisoxazoline copper(I) complexes“ *Tetrahedron* **2015**, *71*, 9333–9338.

[43] Song, X.-G.; Ren, Y.-Y.; Zhu, S.-F.; Zhou, Q.-L. “Enantioselective Copper−Catalyzed Intramolecular N−H Bond Insertion: Synthesis of Chiral 2−Carboxytetrahydroquinolines.” *Adv. Synth. Catal.* **2016**, *358*, 2366–2370.

[44] Zhu, Y.; Liu, X.; Dong, S.; Zhou, Y; Li, W.; Lin, L.; Feng, X. ”Asymmetric N-H insertion of secondary and primary anilines under the catalysis of palladium and chiral guanidine derivatives.” *Angew. Chem. Int. Ed.* **2014**, *53*, 1636–1640.

[45] Guo, J-X.; Zhou, T.; Xu, B.; Zhu, S-F.; Zhou, Q-L.”Enantioselective synthesis of α-alkenyl α-amino acids via N–H insertion reactions” *Chem. Sci.* **2016**, *7*, 1104–1108.

[46] Arredondo, V.; Hiew, S.C.; Gutman, E.; Premachandra, I.D.U.A.; Van Vranken, D.L. “Enantioselective Palladium-Catalyzed Carbene Insertion into the N−H Bonds of Aromatic Heterocycles” *Angew. Chem. Int. Ed.* **2017**, *56*, 4156–4159.

[47] Merli, V.; Daverio, P., Bianchi, S. “Racemization and Enantiomer Separation of Clopidogrel” **2004**, US7259261.

[48] Bachmann, S.; Fielenbach, D.; Jørgensen, K.A. “Cu(I)-Carbenoid- and Ag(I)- Lewis Acid-Catalyzed Asymmetric Intermolecular Insertion of α-Diazo Compounds into N–H Bonds.” *Org. Biomol. Chem.* **2004**, *2*, 3044–3049.

[49] Saito, H.; Morita, D.; Uchiyama, T.; Miyake, M.; Shinichi, M. “Cinchona alkaloids induce asymmetry in the insertion reaction of thermally generated carbenes into N−H bonds” *Tetrahedron Lett.* **2012**, 53, 6662–6664.

[50] Saito, H.; Uchiyama, T.; Miyake, M.; Anada, M.; Hashimoto, S.; Takabatake, T.; Miyairi, S. “Asymmetric intermolecular N−H insertion reaction of phenyldiazoacetates with anilines catalyzed by achiral dirhodium(II) carboxylates and cinchona alkaloids.” *Heterocycles* **2010**, 81, 1149–1155.

[51] Huang, W-S.; Xu, Z.; Yang, K-F.; Chen, L.; Zheng, Z-J.; Xu, L-W. “Modular construction of multifunctional ligands for the enantioselective ruthenium-catalyzed carbenoid N–H insertion reaction: an enzyme−like and substrate−sensitive catalyst system” *RSC Adv*. **2015**, 5, 46455–46463.

[52] Ihara, E.; Haida, N.; Iio, M.; Inoue, K. “Palladium-Mediated Polymerization of Alkyl Diazoacetates To Afford Poly(alkoxycarbonylmethylene)s. First Synthesis of Polymethylenes Bearing Polar Substituents.” *Macromolecules*, **2003**, *36*, 36–41.

[53] Li, J.; Su, Z.; Wang, J.; Hu, C. “Mechanistic investigations on asymmetric N–H insertion of amines catalyzed by palladium-chiral guanidine complex.” *J. Catalysis* **2018**, *364*, 426–436.

[54] Hiew, S.; Booth, M.; Mansiantima, B. M.; Yan, S.; Van Vranken, D. L. “Palladium-Catalyzed Carbene Insertion into the N–H Bond of Cyclic Non-Conjugated Amines: Improved Yields and Enantioselectivity Using Parallel Slow Addition,” Unpublished work.

[55] Bredereck, H.; Effenberger, F.; Hajek, M. “Darstellung von 1-Guanyl-pyrazol und Pyrazol yl-(l)-s-triazin. - Synthesen substituierter s-Triazine.” *Chem. Ber.* **1965**, *98*, 3178.

[56] Bernatowicz, M. S.; Wu, Y.; Matsueda, G. R. “1*H*-Pyrazole-1-carboxamidine Hydrochloride: An Attractive Reagent for Guanylation of Amines and Its Application to Peptide Synthesis.” *J. Org. Chem.* ***1992***, 57, 2497–2502.

[57] Li, M.-L.; Yu, J.-H.; Li, Y.-H.; Zhu, S.-F.; Zhou, Q.-L. “Highly enantioselective carbene insertion into N–H bonds of aliphatic amines” Science 2019, 366, 990-994.

[58] Ley, S. V.; Fitzpatrick, D. E.; Ingham, R. J.; Myers, R. M. Organic Synthesis: March of the Machines *Angew. Chem., Int. Ed.* **2015**, *54*, 3449–3464.

[59] Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. "Computer-Assisted Synthetic Planning: The End of the Beginning" *Angew. Chem., Int. Ed*. **2016**,*55*, 5904–5937.

[60] de Almeida, A. F.; Moreira, R.; Rodrigues, T. A. "Synthetic organic chemistry driven by artificial intelligence" *Nature Reviews Chemistry* **2019**, https://doi.org/10.1038/ s41570-019-0124-0.

[61] Coley, C. W.; Thomas, D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F. "A robotic platform for flow synthesis of organic compounds informed by AI planning" *Science*, **2019**, *365*, 1566.

[62] Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. "CAMEO: a program for the logical prediction of the products of organic reactions" *Pure Appl. Chem.***1990**, *62*, 1921–1932.

[63] Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Loew, P.; Marsili, M.; Saller, H.; Yuki, K. "A new treatment of chemical reactivity: Development of EROS, an expert system for reaction prediction and synthesis design" *Topics. Curr. Chem.* **1987**, *137*, 19–73.

[64] Satoh, H.; Funatsu, K. "SOPHIA, a Knowledge Base-Guided Reaction Prediction System - Utilization of a Knowledge Base Derived from a Reaction Database" *J. Chem. Inf. Model.* **1995**, *35*, 34–44.

[65] Pole, D. L.; Ando, H. Y.; Murphy, S. T. "Prediction of Drug Degradants Using DELPHI: An Expert System for Focusing Knowledge" *Mol. Pharmaceutics* **2007**, *4*, 539–549.

[66] Segler, M. H. S.; Waller, M. P. "Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction" *Chem. Eur. J.***2017**, *23*, 5966–5971.

[67] Chen, J. H.; Baldi, P. "No Electron Left-Behind: A Rule-based Expert System to Predict Chemical Reactions and Reaction Mechanisms" *J. Chem. Inf. Model.***2009**, *49*, 2034–2043.

[68] Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. "Deep learning for chemical reaction prediction." *Mol. Sys. Des. Eng.***2018**, *3*, 442–452.

[69] Kayala, M. A.; Azencotyt, C.-A.; Baldi, P. "Learning to Predict Chemical Reactions" *J. Chem. Inf. Model.* **2011**, *51*, 2209–2222.

[70] *Reaction Predictor* can be accessed at the ChemDB Chemoinformatics Portal at http://reactions.ics.uci.edu/

[71] Mayr, H.; Ofial, A.R. "Do general nucleophilicity scales exist?" *J. Phys. Org. Chem.* **2008**, *21* 584–595.

[72] Ritchie, C.D. "Nucleophilic Reactivities Toward Cations." *Acc. Chem. Res.* **1978**, 348–353.

[73] J.O.Edwards, *ibid*., **76**, 1540 (1954); **78**, 1819 (1956).

[74] Mayr, H. "Reactivity scales for quantifying polar organic reactivity:the benzhydrylium methodology." *Tetrahedron*, **2015**, *71*, 5095–5111.

[75] Chen, J. H.; Baldi, P. "No Electron Left-Behind: A Rule-based Expert System to Predict Chemical Reactions and Reaction Mechanisms" *J. Chem. Inf. Model.* **2009**, *49*, 2034–2043.

[76] Phan, T. B.; Breugst, M.; Mayr, H. "Towards a general scale of nucleophilicity?" *Angew. Chem. Int. Ed.* **2006,** *45*, 3869–3874.

[77] Evans'p*K*a table can be accessed at http://ccc.chem.pitt.edu/wipf/MechOMs/evans_pKa_table.pdf.

[78] Bordwell's p*K*a table can be accessed at https://organicchemistrydata.org/hansreich/resources/pka/.

[79] Follet, E.; Zipse, H.; Lakhdar, S.; Ofial, A. R.; Berionni, G. "Nucleophilicities and Lewis Basicities of Sterically Hindered Pyridines" *Synthesis* **2017**, *49*, 3495–3504.

[80] Mayr, H.; Breugst, M.; Ofial, A. R. "Farewell to the HSAB Treatment of Ambident Reactivity" Angew. Chem. Int. Ed. 2011, 50, 6470-6505.

[81] Mood, A.; Tavakoli, M.; Gutman, E.; Kadish, D.; Baldi, P. Van Vranken, D. L. "Methyl Anion Affinities of the Canonical Organic Functional Groups." *J. Org. Chem*, **2020**, *85*, 4096–4102.

[82] Wennberg, P. O.; Bates, K. H.; Crounse, J. D.; Doson, L. G.; McVay, R. C.; Mertens, L. A.; Nguyen, T. B.; Praske, E.; Schwantes, R. H.; Smarte, M. D.; St. Cllair, J. M.; Teng, A. P.; Zhang, X.; Seinfeld, J. H. "Gas-Phase Reactions of Isoprene and Its Major Oxidation Products" *Chem. Rev.* **2018**, *118*, 3337–3390.

[83] Nguyen, T. B.; Laskin, J.; Laskin, A.; Nizkorodov, S. A. "Nitrogen-containing organic compounds and oligomers in secondary organic aerosol formed by photooxidation of isoprene" Env. Sci. Tech. 2011, 45, 6908–6918.

[84] Jenkins, M. E.; Young, J. C.; Rickard, A. R. "The MCM v3.3.1 degradation scheme for isoprene" *Atmos. Chem. Phys.* **2015**, *15*, 11433–11459.

[85] Jenkin, M. E.; Saunders, S. M.; Pilling, M. J. "The tropospheric degradation of volatile organic compounds: a protocol for mechanism development" *Atmos. Environ*. **1997**, 31, 81–104.

[86] Saunders, S. M.; Jenkin, M. E.; Derwent, R. G.; Pilling, M. J. "Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds" *Atmos. Chem. Phys*. **2003**, *3*, 161–180.

[87] *ReactionMap* is maintained at the ChemDB Chemoinformatics Portal at http://cdb.ics.uci.edu/cgibin/reactionmap/ReactionMapWeb.py