Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Kimberly R. Sharp November 20, 2020

Computational and Experimental Studies of Selective Rhodium-Catalyzed Allylic C–H Functionalization through π-Allyl Intermediates and Computational Development of Macrocycle Leads for Undruggable Targets

By

Kimberly R. Sharp

Simon Blakey, Ph.D.

Advisor

Chemistry

Simon Blakey, Ph.D.

Advisor

Cora MacBeth, Ph.D.

Committee Member

David Zureick-Brown, Ph.D.

Committee Member

2020

Computational and Experimental Studies of Selective Rhodium-Catalyzed Allylic C–H Functionalization through π-Allyl Intermediates and Computational Development of Macrocycle Leads for Undruggable Targets

By

Kimberly R. Sharp

Simon Blakey, Ph.D.
Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Chemistry

2020

Abstract

Computational and Experimental Studies of Selective Rhodium-Catalyzed Allylic C–H Functionalization through π-Allyl Intermediates and Computational Development of Macrocycle Leads for Undruggable Targets
By Kimberly R. Sharp

Allylic C–H functionalization methodology has been developed as an atom-economical way of accessing π-allyl intermediates to form allylic products. Recent advances in the Blakey group have resulted in the development of a regioselective amination methodology tolerating the use of amines with a single electron-withdrawing group as the nucleophile, implying greater synthetic utility. An efficient synthesis of Sensipar using this amination procedure with an allylbenzene derivative was proposed. Optimization studies on a model system were then pursued to provide a yield double the previously disclosed results. Unfortunately, applying these conditions with the target amine required for this synthetic plan, resulted in oxidation of the substrate, indicating incompatibility. Following a study of functional group tolerance of the allylic coupling partners, where low yields were observed, an iridium-catalyzed branched-selective method was developed. Further catalyst design within the Blakey group resulted in the development of an asymmetric rhodium indenyl catalyst capable of performing an allylic C–H amidation with high enantio- and regioselectivity. Computational studies were undertaken to investigate the source of these selectivities. Through the development of a full energy profile, it was determined that the two transition states of the formation of the π-allyl complex distinguished selectivity through the alignment of the allylbenzene to the catalyst; the energetic difference matched the experimentally observed enantiomeric ratio. The regioselectivity was determined by the steric interaction of the amide source to the π-allyl intermediate in the C–N bond forming step. Determining potential directions for future C–H functionalization methodology led to a computational investigation of leads for "undruggable" protein targets, those with binding surfaces rather than a well-defined binding pocket. The semi-rigid nature of macrocycles has been implicated as a way of attaining affinity for these binding surfaces. Due to challenges in attaining structural data of target interactions from which to design drugs, high throughput screening of biologically synthesized macrocyclic peptides has been utilized to identify hits. Problems arise with the bioavailability of these structures due to their large number of amide linkages, indicating a need for more structurally diverse macrocycles. Using a macrocyclic peptide which displayed high affinity to a given target as a template of functionality for these non-peptide macrocycles can be accomplished through machine learning techniques. A starting set for this computation has been developed using randomly connected elements of a fragment library, with particular attention to ensuring their druglike nature. Machine learning computations on this set should then result in targets for these undruggable proteins and aims for future synthesis.

Computational and Experimental Studies of Selective Rhodium-Catalyzed Allylic C–H Functionalization through π-Allyl Intermediates and Computational Development of Macrocycle Leads for Undruggable Targets

By

Kimberly R. Sharp

Advisor: Simon Blakey, Ph.D.

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Chemistry

2020

Acknowledgments

I would like first and foremost to thank Professor Blakey for his constant mentorship, advice, and affirmation throughout my time in his lab. Not only has he helped me develop into a competent chemist, but he has helpfully reminded me of my ability. Dr. MacBeth has also been an enormously supportive source of advice; being able to meet with her to develop plans for summers and the future has been greatly helpful. I would also like to thank DZB for his constant empathy for students and remarkably dry wit.

The prodigious efforts of Taylor to teach me how to actually do chemistry are also what has enable me to develop from a freshman in awe of a rotavap to a researcher capable of working independently. Her continually helpful advice, her willingness to edit any writing on short notice (including many parts of this thesis), and her friendship have been a key part of my lab experience and my success. Caitlin and Amaan have also been excellent assistant mentors, constantly able to provide advice and answer any question I've thrown at them as well as being excellent examples of how chemists think about problems.

While a large part of working in lab is the actual research, the people I've been able to work alongside have made it a wonderful experience. My lab experience has been a key part of the time I've spent at Emory, and I will greatly miss the people I've had the pleasure to work with.

Finally, I would like to thank my roommates for dealing with me during the insanity of this semester, and especially Maryam Khaqan for teaching me how to use LaTeX, which made this thesis a far easier task.

Contents

List of Figures

List of Tables

Chapter 1

Allylic Amination

1.1 Introduction

Transition-metal catalyzed allylic substitution reactions provide a means of forming complex functionality from simple allylic precursors . Formation of a metal π-allyl intermediate provides convenient access to a number of bond forming reactions. Most notably, Tsuji and Trost accessed these π-allyl intermediates using a palladium catalyst, and disclosed compatibility with a wide range of nucleophiles capable of generating C–C and C –heteroatom bonds.[1] As this reaction requires a leaving group on the allyl moiety, stoichiometric byproducts are formed,[2] a problem solved by incorporating C–H functionalization techniques. Directly accessing a π-allyl complex from a C–H bond allows for compatibility with a wider range of substrates and greatly increases the overall atom efficiency.

C–H functionalization methods have been incorporated into this chemistry successfully starting from pre-oxidized starting materials, e.g. alkynes, allenes, and conjugated dienes, using rhodium catalysis (Figure 1.1). While this methodology is able to avoid the use of a previously necessary leaving group, its utility is nevertheless decreased due to their pre-oxidized substrates. Further advances in using unactivated alkenes to form π-allyl

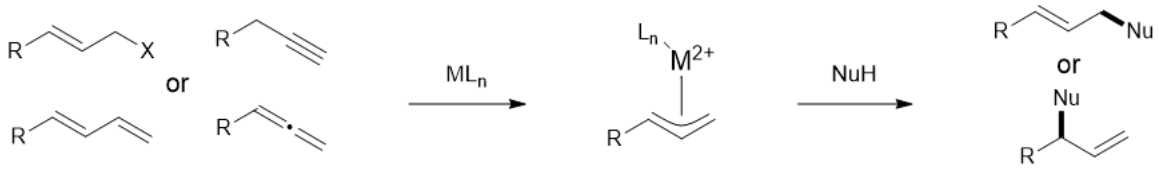


Figure 1.1: General Reaction Scheme for Pre-Oxidized Alkenes

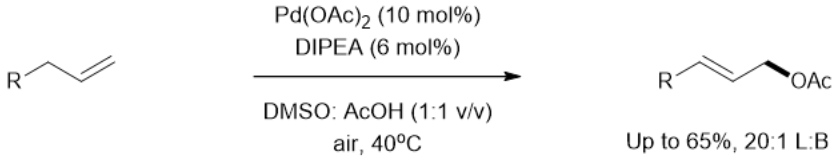


Figure 1.2: Palladium-Catalyzed C–H Oxygenation

complexes came from White and coworkers with their disclosure of a palladium-catalyzed C–H oxygenation, showing significant regioselectivity for the linear product[3] as shown in Figure 1.2. This method has been expanded to include a range of oxygen, nitrogen, and carbon nucleophiles, but relies on highly activated nucleophiles,[4] a significant limitation for applications to synthesis. Cossy and coworkers were able to expand this chemistry by using an amine source with only a single electron withdrawing group to accomplish an allylic transformation through a rhodium-catalyzed π-allyl intermediate;[5] however, this reaction was intramolecular, as shown in Figure 1.3. Expanding these efficient allylic transformations to achieve more general utility in synthetic strategy represents a significant advancement.

Cossy's amination methodology was expanded to intermolecular allylic C–H activation through the use of a rhodium (III) pentamethylcyclopentadienyl (Cp*) catalyst[6] by Dr. Jacob Burman, a previous member of the Blakey group (Figure 1.4). This methodology displayed high regioselectivity for the linear product, required no pre-oxidation, and required

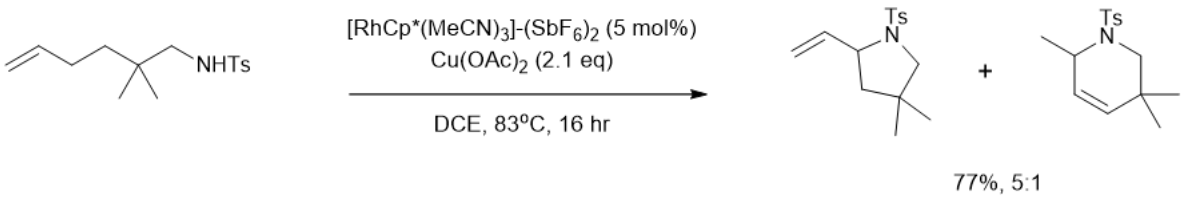


Figure 1.3: Rhodium-Catalyzed Intramolecular C–H Amination

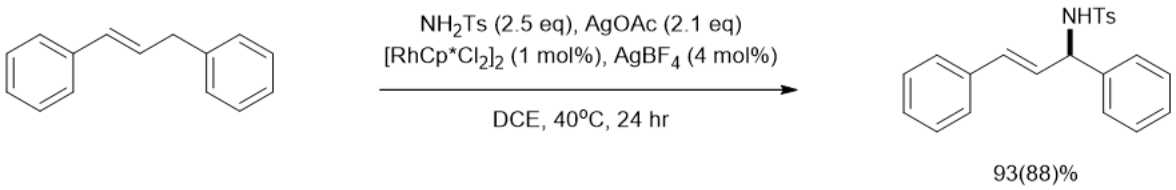


Figure 1.4: Rhodium-Catalyzed Intermolecular C–H Amination

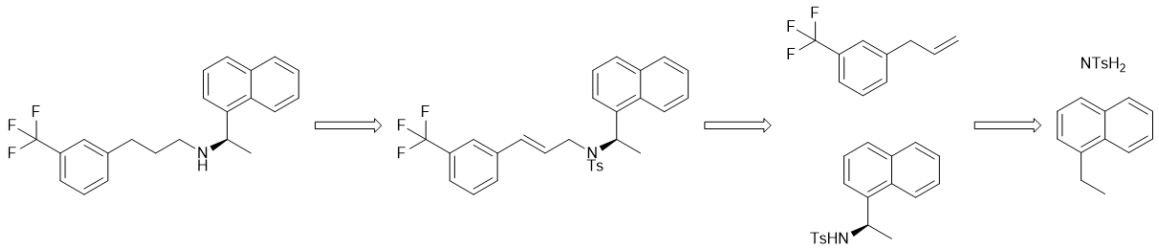


Figure 1.5: Retrosynthetic Analysis of Sensipar

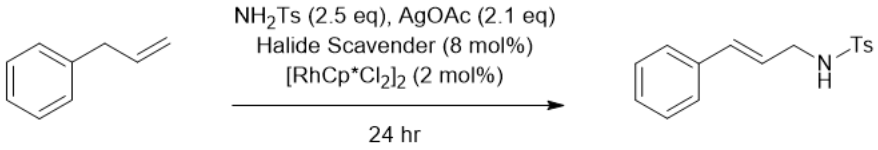
only one electron withdrawing group on the amine source, indicating key improvements for applicability to synthesis.

A proposed synthesis of Sensipar was intended to capitalize upon this advance. The structure indicated a potential linear allylic amination prior to hydrogenation could be used to generate a key linkage from a commercially available substituted allylbenzene and a protected amine. C–H functionalization techniques can be used to efficiently form this amine from commercially available starting materials using nitrene chemistry pioneered by Davies and coworkers.[7] Here, enantioselectively aminating the benzylic position of ethylnaphthalene should yield the desired amine in one step. This synthetic strategy, which can be accomplished in two amination steps and followed by hydrogenation, is shown in Figure 1.5. Utilizing these two C–H functionalization methods would allow for a conveniently short synthesis.

1.2 Results and Discussion

While this amination method was shown to be compatible with terminal alkenes, specifically allylbenzene, only a moderate yield was seen given the conditions used were instead optimized

Table 1.1: Optimization Table of Amination Conditions on Allylbenzene



entry	Halide Scavenger	solvent	Temp (°C)	% yield
1	$AgBF_4$	DCE	40	54
2	$AgBF_4$	DCE	60	61
3	$AgBF_4$	DCE	80	62
4	$AgSbF_6$	DCE	40	28
5	$AgSbF_6$	DCE	60	37
6	$AgSbF_6$	DCE	80	29
7	$AgNTf_2$	DCE	40	31
8	$AgNTf_2$	DCE	60	47
9	$AgNTf_2$	DCE	80	55
10	$AgBAr^F_4$	DCE	40	n.d.
11	$AgBAr^F_4$	DCE	60	11
12	$AgBAr^F_4$	DCE	80	15
13	$AgBF_4$	DCM	40	46
14	$AgBF_4$	THF	60	70
15	$AgBF_4$	trifluorotoluene	80	35
16	$AgBF_4$	trifluorotoluene	100	70
17	$AgBF_4$	chlorobenzene	80	72
18	$AgBF_4$	chlorobenzene	100	67
19	$AgBF_4$	p-dioxane	80	64
20	$AgBF_4$	p-dioxane	100	81

on diphenylpropene, an internal alkene. Identifying conditions to obtain a more synthetically viable yield is important for demonstrating the efficacy of the reaction; thus, a variety of halide scavengers, solvents, and temperatures were tested using a placeholder amine—p-toluenesulfonamide—to determine the optimal conditions. The conditions tested are shown in Table 1.1. A significant increase in yield over literature conditions was seen following the change in solvent and temperature; given that previous reports had highlighted the importance of chlorinated solvents such as DCM and DCE, the high yield found in p-dioxane was unexpected.

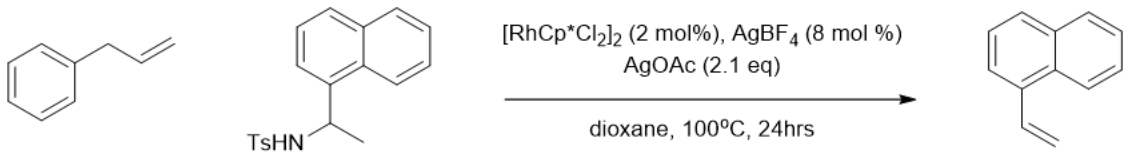


Figure 1.6: Elimination of Ethylnapthylamine to form Vinylnaphthalene

Armed with these improved conditions for amination with a simple amine, the commercially available racemic naphthylethylamine was protected with a tosyl group and tested. Unfortunately, this reaction led to the elimination the amine in naphthylethylamine to form vinylnaphthalene (Figure 1.6), preventing the amination from occurring. As these conditions are necessary for the catalytic cycle,[8] the reaction cannot be modified to avoid this deactivation, and the required substrates for the proposed synthetic scheme were deemed incompatible with the amination methodology.

However, the high yields following the optimization of the amination conditions for allylbenzene indicated utility of this disclosed method for functionalizing terminal alkenes. Exploring the functional group tolerance displayed by this method can provide a better understanding of its potential applications. To this end, a substrate scope was explored with various substituted allylbenzene structures, the results of which are displayed in Figure 1.7. Unfortunately, all substitutions led to a decrease in yield: electron withdrawing, electron donating, and sterically encumbered variants were not well tolerated.

Given the limited success of this method on terminal alkenes, further methodology was pursued to achieve these transformations with better tolerance. This work informed an iridium-catalyzed method developed by Amaan Kazerouni and Taylor Nelson, graduate students in the Blakey group: these conditions were able to successfully aminate terminal alkenes, and showed significant regioselectivity for amination at the branched position.[9] Tolerance for electron withdrawing substituents and other functionality was displayed, with these reactions able to achieve similarly high yield.

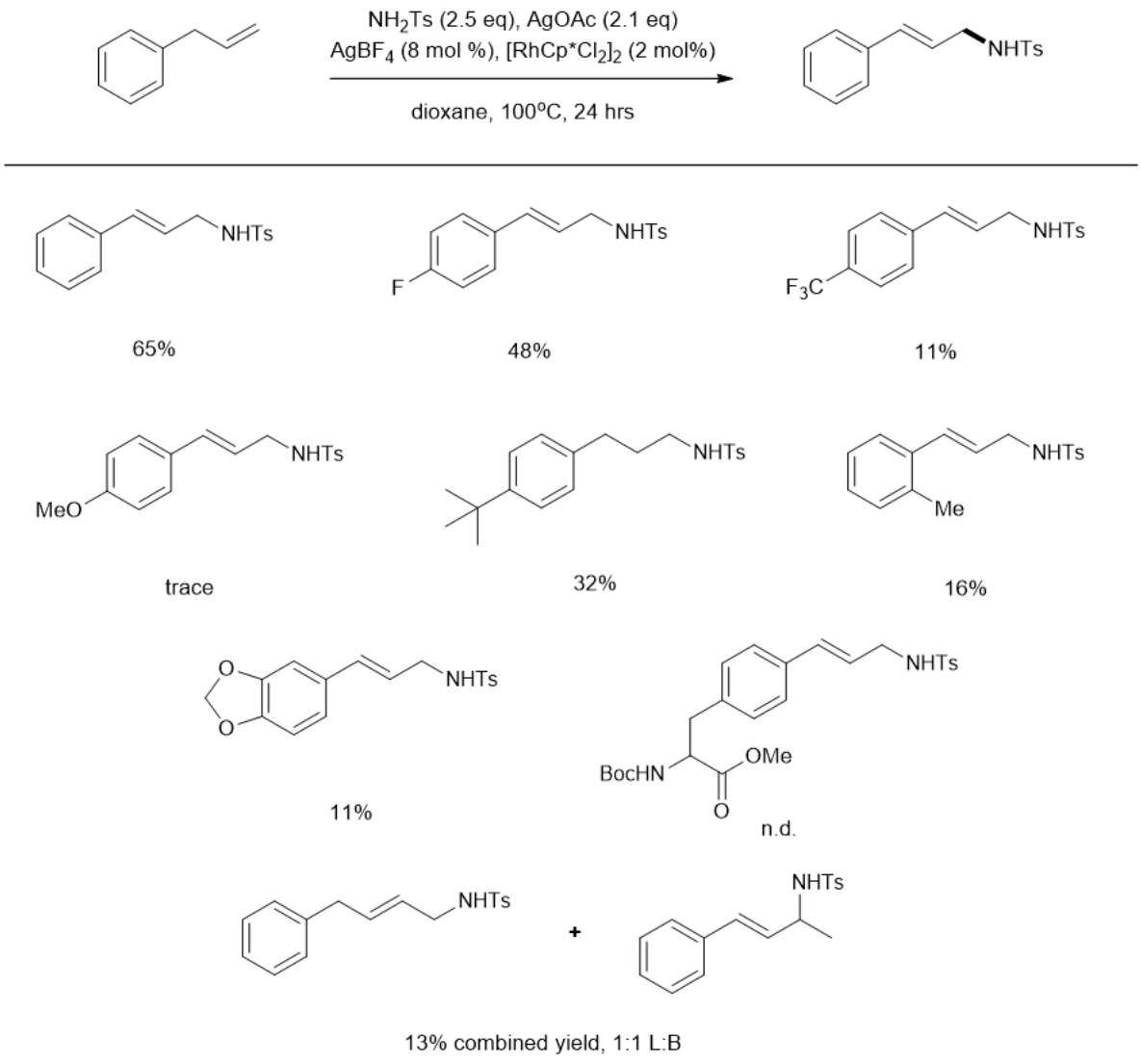


Figure 1.7: Substrate Scope of Substituted Allylbenzene Structures

1.3 Conclusions

While the synthesis of Sensipar was unable to be accomplished through the proposed scheme, optimizing the amination of terminal alkenes using a rhodium Cp* catalyst demonstrated a yield high enough for synthetic viability. Introducing additional functionality, whether electron withdrawing or electron donating, led to significantly decreased yield. This focus on terminal alkenes contributed to the development of an iridium Cp* catalyzed amination method, selective for the branched regioisomer.

Chapter 2

Mechanism of Allylic Amidation

2.1 Introduction

Despite the successes of rhodium Cp* catalysts for allylic C–H amination, the reaction necessarily results in a racemic product, posing a barrier to its use in the synthesis of natural products or pharmaceutically relevant compounds. An analysis of the mechanism revealed that the reaction proceeds through an S_N1 type substitution, occurring off the metal center;[8] asymmetric catalysis would thus fail to induce enantioselectivity. Developing a catalyst capable of achieving allylic functionalization through a different mechanism should allow an enantioselective allylic amination method to be developed.

Replacing the Cp* ligand with an indenyl ligand represents a significant change in potential reactivity due to the indenyl effect: the increased stability of the η^3 coordination mode decreases the energetic barrier for ring slippage (Figure 2.1), allowing for coordination

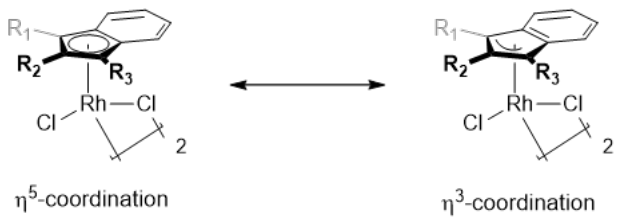


Figure 2.1: Ring Slip on Indenyl Ligand

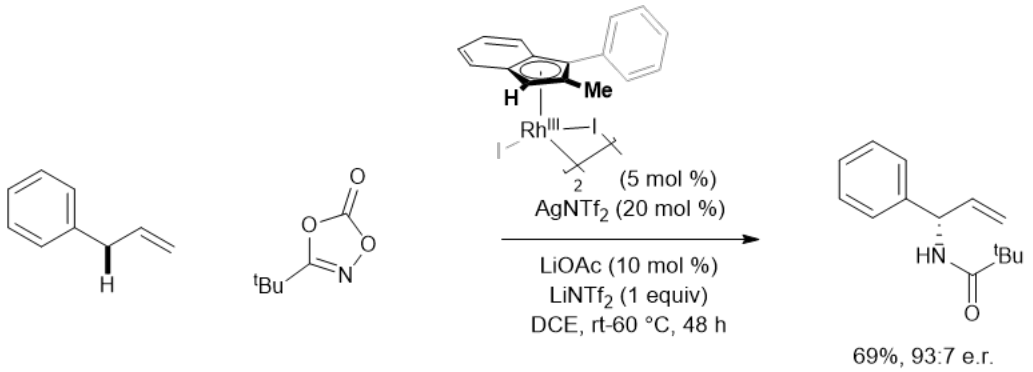


Figure 2.2: Successful Enantioselective Amidation of Allylbenzene with 2-methyl 3-phenyl Indene Catalyst

of additional substrates.[10] Significant improvements in reaction rate in addition to changes in reactivity have been observed when compared to Cp ligands in similar chemistry;[11] using these ligands thus represents a promising change to an effective yet mechanistically distinct reaction. Both electronic and steric asymmetry were easily induced in these indenyl ligands through formation of 2-methyl 3-phenyl indene.

For more applicability to natural product synthesis, this reaction was developed for the formation of amide bonds using dioxazolones as a nitrogen source; precedent for allylic functionalization with dioxazolones exists from our group,[12] as well as by the Rovis[13] and Glorius[13] groups. Using the asymmetric indene catalyst to amidate allylbenzene with dioxazolone provided excellent enantioselectivity and regioselectivity, as well as significant yield (Figure 2.2). While mechanistic investigations performed by both the Blakey group and Rovis group indicated that there is likely formation of a π-allyl intermediate as the rate determining step, computational work was needed to provide evidence for a full mechanism, with special attention paid to the source of this selectivity.

2.2 Results and Discussion

Density functional theory (DFT) calculations were performed in collaboration with the Baik group to generate a full energy profile of this reaction. Beginning the calculations with the

presumed π-allyl formation, the free energy was determined of each component: the proposed active rhodium catalyst, the association of the alkene of the allylbenzene, then the transition state of the formation of a π-allyl complex through a concerted metalation-deprotonation process.

The calculated activation energy for π-allyl formation was consistent with it being rate limiting, providing further evidence for this mechanism. Additionally, the asymmetric ligand impacted the preferred orientation of the allylbenzene for this step: a difference of 1.1 $kcal/mol$ was calculated, consistent with the observed enantiomeric ratio. The two transition states are shown in figure. Given that the energetic barrier of the reverse reaction is prohibitively high for any significant reversibility to occur at room temperature, the enantioselectivity is likely set at this step.

Following this step, energy of coordination of dioxazolone to this complex was calculated; the *tert*-butyl substituent on the dioxazolone had a significant steric effect on the stability of these arrangements, preferring to be positioned opposite the phenyl of the allylbenzene. After the dioxazolone releases carbon dioxide to form a nitrene, interconversion is still possible, indicating that the regioselectivity isn't set until reductive coupling to form the C–N bond. The full energy profile can be seen in Figure 2.3.

2.3 Conclusions

Computational evidence suggests that the enantioselectivity of the reaction is set at the concerted metalation-deprotonation step, then conserved throughout the reaction, while the regioselectivity was determined by the position of the dioxazolone relative to the allylbenzene. Using this understanding of the importance of steric impact on the mechanism, further catalyst development can be pursued to achieve increased yield and utility from similar allylic functionalization reactions.

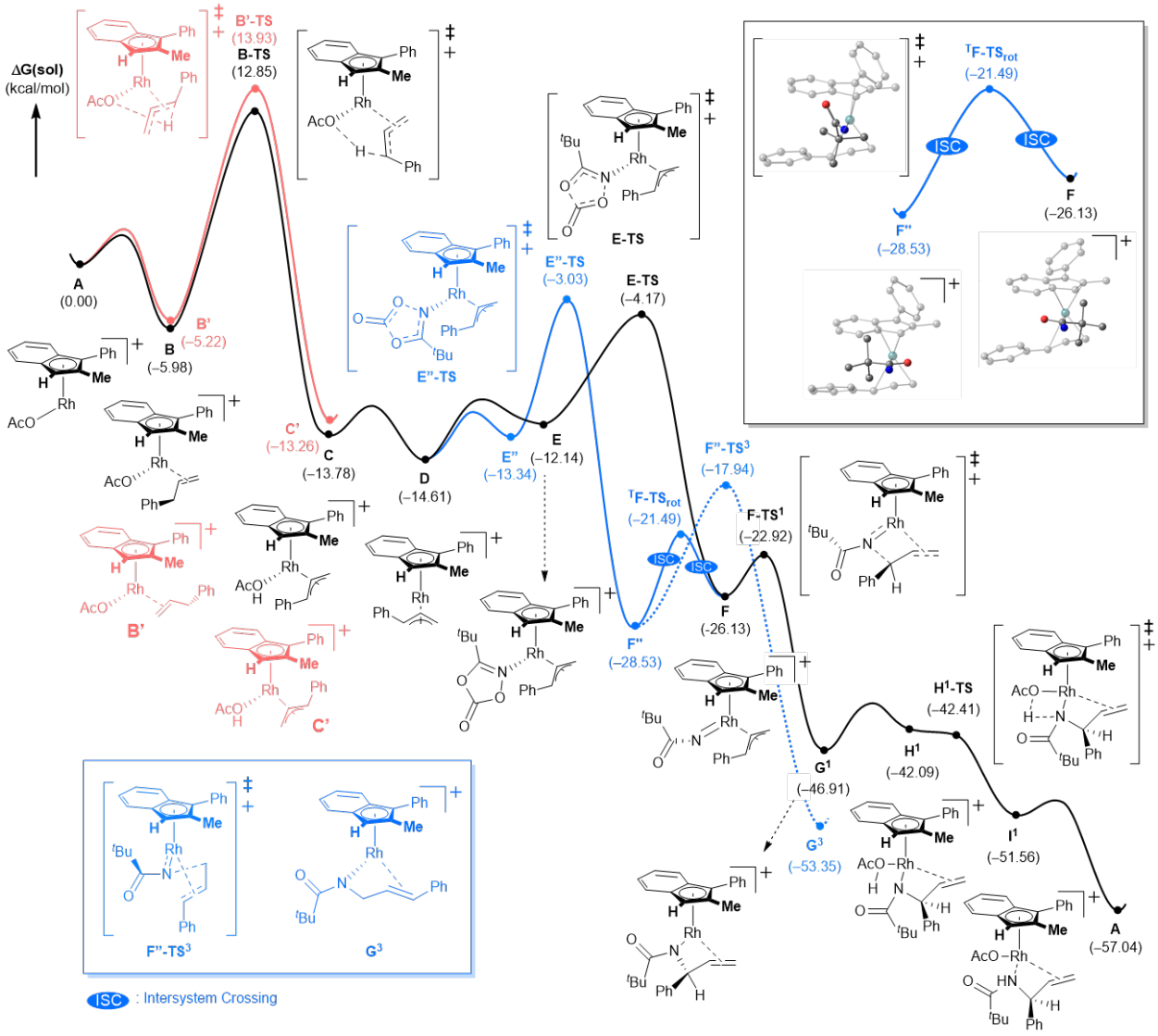


Figure 2.3: Energy Profile of Enantioselective Amidation

Chapter 3

Development of Structurally Diverse Macrocycles for Undruggable Targets

3.1 Introduction

Traditional drug development techniques focus on proteins that contain a clear binding site for small molecules, such as G-protein coupled receptors, ion channels, and several classes of enzymes.[14] This relatively straightforward functionality allows computational studies and crystallography to provide information for the development of drug molecules based on the binding sites functionality. Important potential targets, such as transcription factors, intracellular protein-protein interactions, and protein-nucleic acid interactions, have less well-defined binding sites, displaying instead general binding surfaces. Due to the difficulty in designing appropriate small molecules for these interactions, they have been termed "undruggable."[15] One emerging method for targeting these interactions is with macrocycles, molecules with at least one ring of 12 members or larger. The ring structure provides an entropic advantage and convenient pre-ordering while maintaining flexibility, allowing noncovalent interactions to provide a large binding area.[16] Macrocyclic properties exist in an optimal middle ground between biological compounds and small molecules,

making them desirable for drug development.

Due to common challenges in attaining crystal structures that clarify the interaction of molecules with the binding surfaces in these targets, there is little information available to intentionally design a macrocycle that displays significant affinity. Instead, high throughput screening of a range of macrocycles to experimentally identify those that associate to the binding site can be performed; for this process to be feasible, these macrocycles are limited to macrocyclic peptides, which can be conveniently biologically synthesized from a source DNA or RNA segment. Libraries comprised of these segments can span a huge range of chemical space, containing references to up to 10^{14} macrocyclic peptides.[15] By performing several rounds of synthesis and analysis, those showing the highest affinity for the selected site can be identified as "hits" and used as a starting place for drug development.[15]

Though macrocyclic peptide libraries are able to highlight potential interactions, the high number of amide bonds imposes restrictions on their utility. The highly polar nature of these bonds limits membrane permeability, especially in comparison to small molecules. Additionally, while more stable than their linear counterparts, the amide bonds in cyclic peptides are still susceptible to both hydrolytic and proteolytic cleavage. Developing macrocycles with more variation in structural motifs will yield molecules capable of targeting "undruggable" interactions and yield drugs with pharmaceutically useful bioavailability.

As accessing interaction data directly from crystal structures is not always possible, macrocyclic peptides recognized as hits from the existing affinity selection process will act as a template for desired functionality. Computational methods can be utilized to not only recognize which non-peptide structures provide similar binding interactions as a given template peptide, but also to make modifications on a given macrocycle to achieve greater binding similarity. Starting with a randomly generated druglike starting set, two machine learning techniques can be achieved to employ this goal. A neural network will be trained to quantify similarity between macrocycles based on characteristics important for docking, and an evolutionary algorithm will modify a set of non-peptide macrocycles to more closely

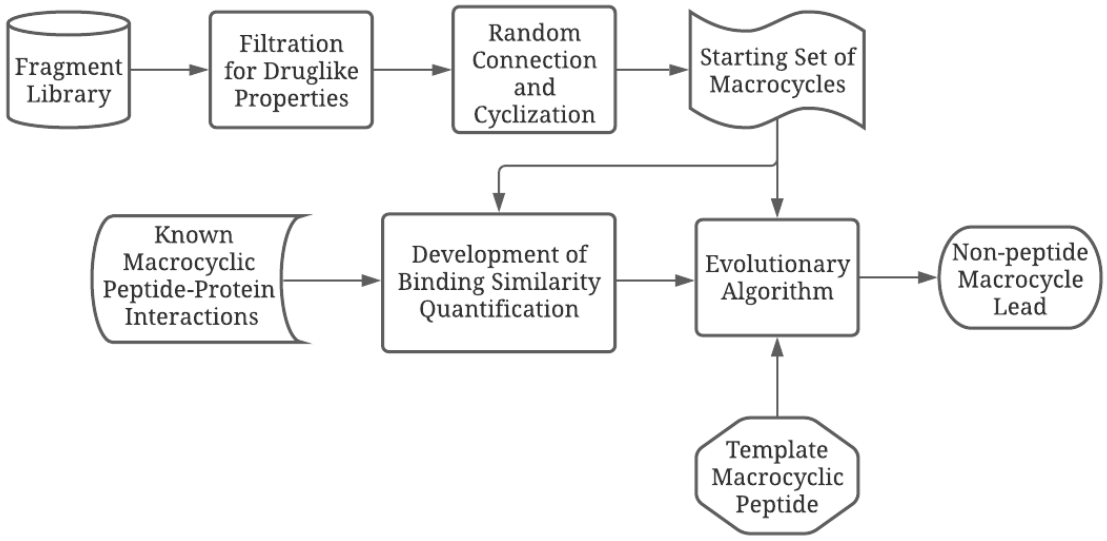


Figure 3.1: Project Overview

approximate the specific macrocyclic peptide. This process should result in a workflow able to generate a structurally appealing macrocycle with high affinity to the target without direct knowledge of its structure. Figure 3.1 illustrates the process of the overall project.

3.2 Results and Discussion

Given the starting set of non-peptide macrocycles informs both machine learning steps, a method of generating one was developed in collaboration with Sophia Xu, an undergraduate member of the Blakey group. While existing macrocyclic peptide libraries have been built using defined reaction routes,[17] changes made later in the machine learning process would undo the efforts to restrict the macrocycles to synthetically convenient structures; accessing a wider range of functionality from the beginning is then more of a priority. To generate this set, a fragment library was developed, then fragments were randomly connected and cyclized to yield structurally and functionally diverse macrocycles. The fragment library was formed by sorting an existing library intended for use in fragment-based drug discovery projects to ensure the substituents had properties indicative of oral bioavailability; this sorting was

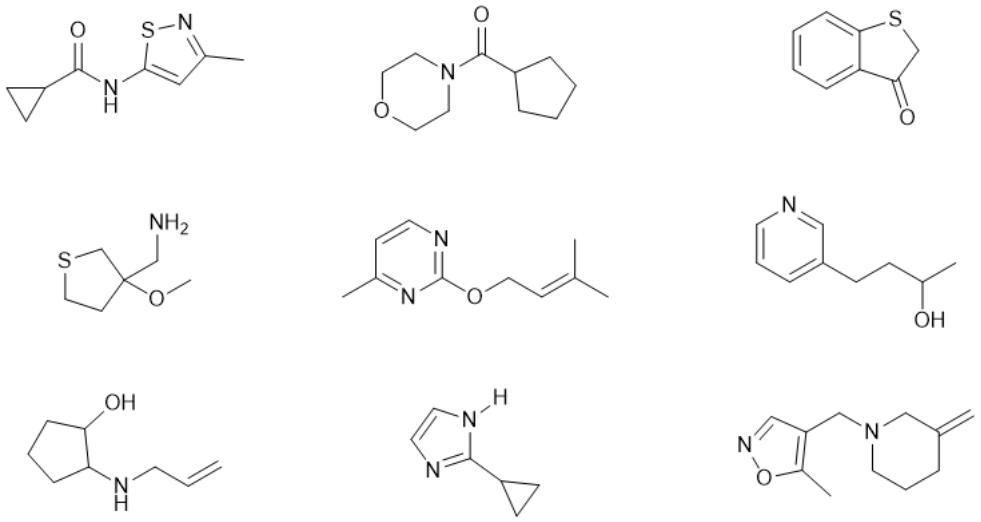


Figure 3.2: Randomly Selected Fragments

based on Lipinski's Rule of Three (RO3), an analogue of his commonly used Rule of Five for determining the pharmaceutical relevance of drugs.[18] Factors such as the number of rotatable bonds and the hydrogen bond donors and acceptors are considered;[19] these values were calculated for each fragment in the source library, and those that were outside the RO3 bounds were discarded. This resulting set consisted of 2,252 fragments, an appropriately sized library to ensure structural diversity: von Itzstein and coworkers have determined that after sorting based on RO3 criteria, a library of 2,000 fragments had a diversity metric of within 1% of more than 200,000 measured fragments.[20] A randomly selected subset of these fragments can be seen in Figure 3.2, and the library's distribution along RO3 and other useful criteria is illustrated in Figure 3.3. It's clear that significant diversity exists within these screened criteria.

Combining these fragments to form macrocycles requires both randomly selecting members of the fragment library to build onto the macrocycle as well as forming new bonds at randomly generated locations on each fragment. When both the fragment and site are selected in one step, disjointed combinations can occur, with further bonds formed between an already connected molecular piece. This problem can be avoided by partitioning the two steps: selecting first a specified number of fragments, then proceeding to travel in one

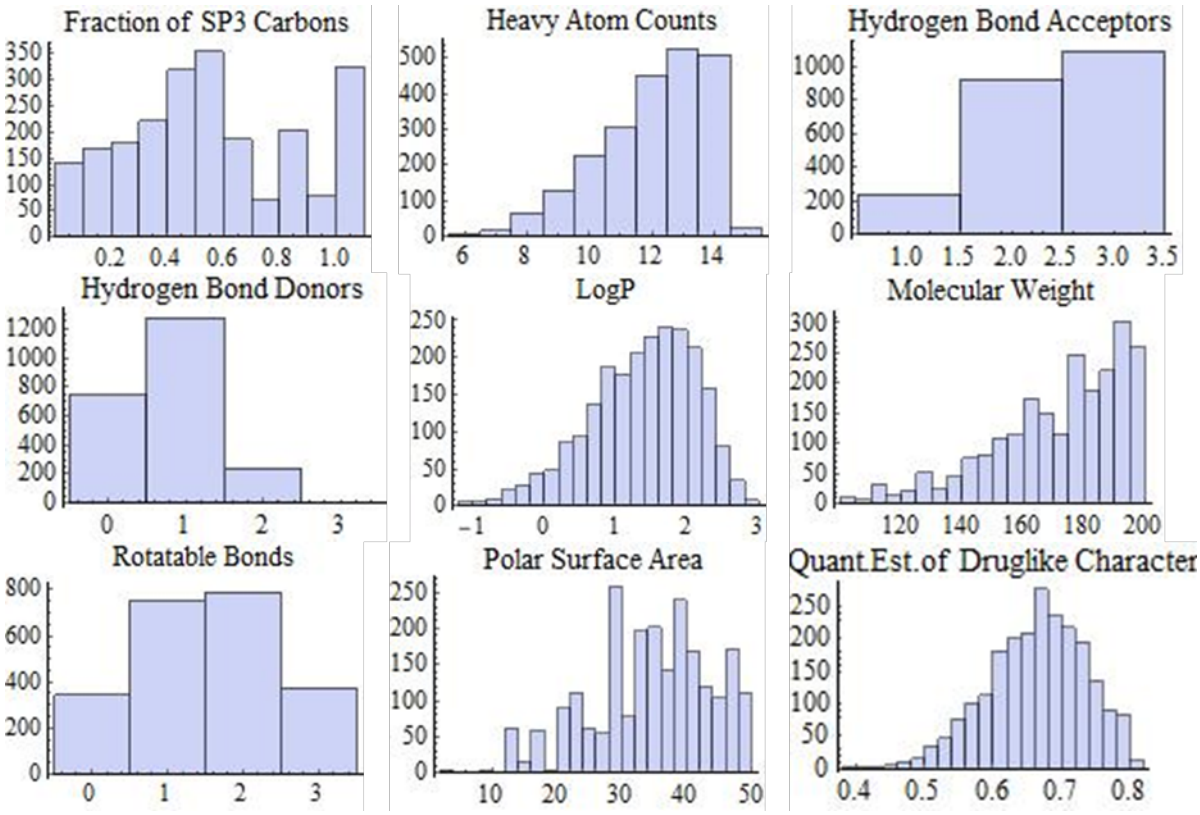


Figure 3.3: Fragment Library Distribution Along Predictive Criteria for Druglike Nature

direction down this list to add a fragment to the growing molecule at one end. This process can be best performed with a stack; fixing one end prevents side additions, while the other takes two fragments to be "popped," combined, then the result can "pushed" back onto the stack.

To respect valencies of involved atoms, open binding sites were defined as those with hydrogens attached: any site that was either implicitly or explicitly recognized to contain a hydrogen can instead be replaced by a bond to another element without disrupting the structure. Determining the ideal number of fragments to connect for formation of macrocycles became the next step; while the necessary ring size of twelve or more imposed a lower constraint of three fragments, determining an upper limit became important. While more fragments permit larger rings and a greater variety of cyclizations, the additional bulk can interfere with the druglike character important for developing an appropriate starting set.

Given that macrocyclization interferes with the predictive ability of the RO5 criteria,[21] the pre-cyclized linear molecules were studied. Molecules of three, four, and five fragments were formed and these characteristics were calculated and analyzed for each; selected linear molecules for each group are shown in Figure 3.4, and relative distributions of their RO5 analyses can be shown in Figure 3.5. While no set had a perfect distribution within ideal criteria, the five fragment molecules generally exceeded both the ideal number of hydrogen bond donors and hydrogen bond acceptors, indicating that those molecules were simply too large. Three and four fragment molecules were then focused on for future development and cyclization.

Following a challenging process of defining cyclizations of rings size twelve or greater, these randomly generated intramolecular bonds resulted in significant variation among macrocycles formed from the same linear molecule. As can be seen in Figure 3.6, different placements of this intramolecular bond affect which functional groups are constrained by the ring, changing how this macrocycle would interact with a target. However, given

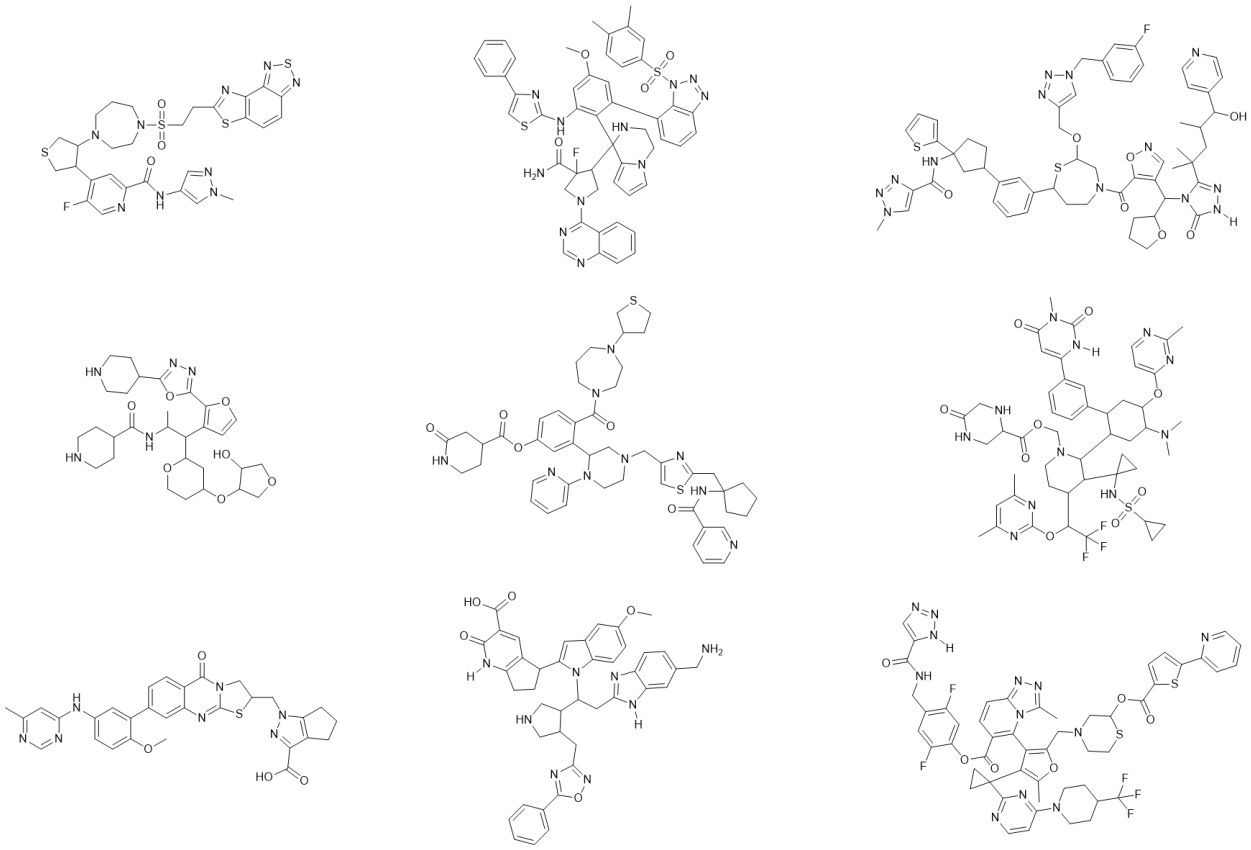


Figure 3.4: Randomly Selected Examples of 3, 4, and 5 Fragment Molecules

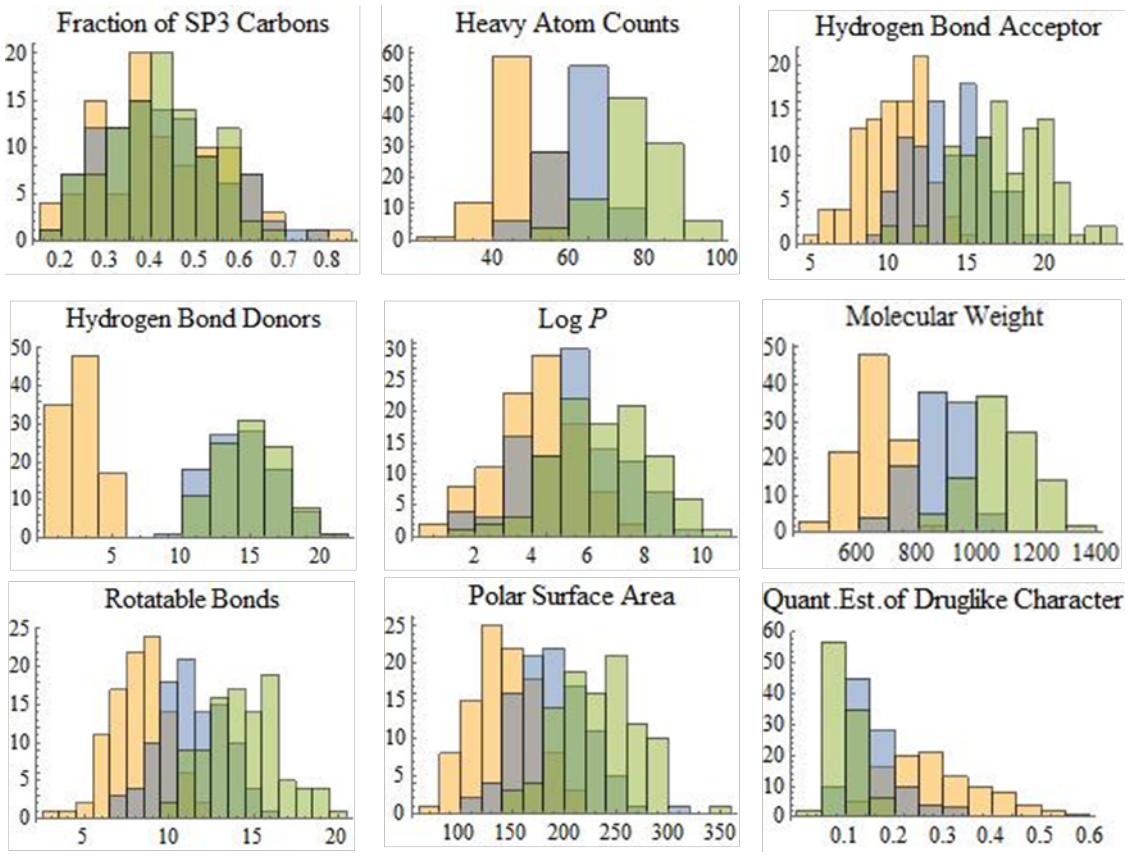


Figure 3.5: Connected Molecule Distributions Along Predictive Criteria for Druglike Nature

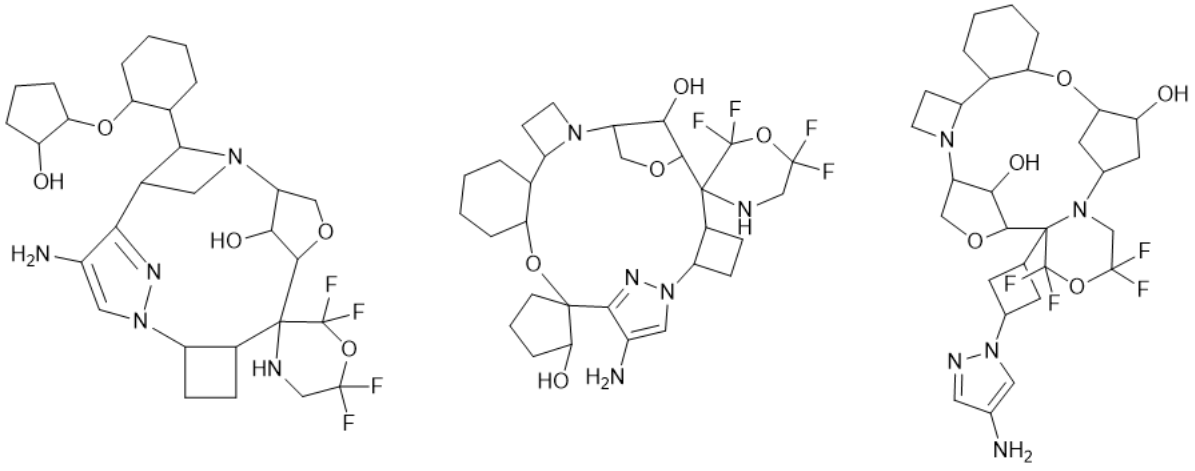


Figure 3.6: Possible Connections for a Three Fragment Macrocycle

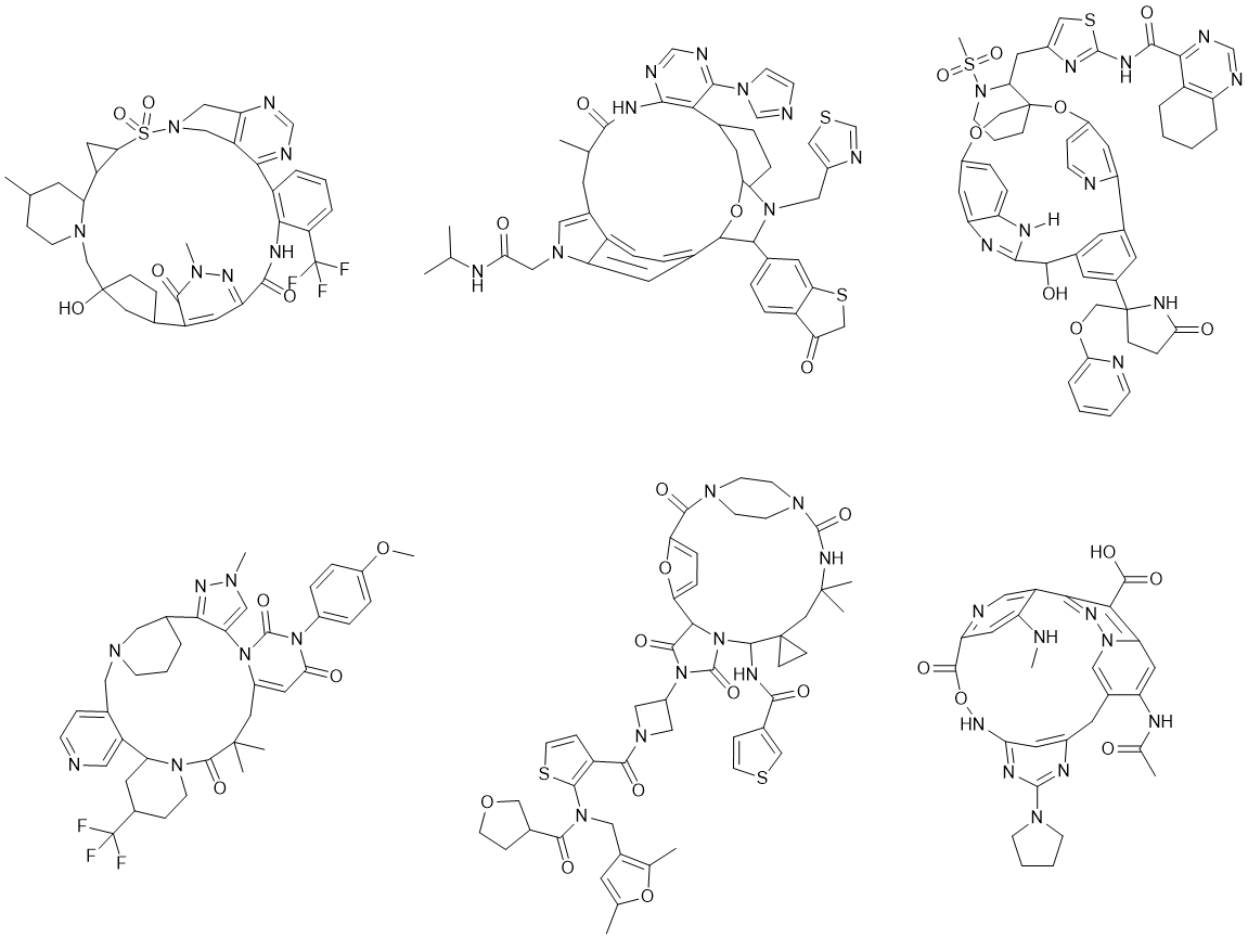


Figure 3.7: Example Macrocycles from 3 and 4 Fragment Molecules

that an evolutionary algorithm with the ability to make random changes to bonds, atoms, stereochemistry, and functional groups will be run on this starting set, selecting only one of these possibilities will not limit final possibilities. This combination of macrocycles of three and four fragments can thus be utilized as a basis for machine learning algorithms to generate tailored bioavailable non-peptide macrocycles. Examples of connected macrocycles can be seen in Figure 3.7.

3.3 Conclusions and Future Work

A method for generating a starting set of appropriately sized macrocycles from a modified library of druglike fragments has been developed, and can be easily altered in response to

feedback from future machine learning steps. As this project is ongoing, attention must now be paid to developing an evolutionary algorithm. This process consists of scoring elements of this starting set to focus on the most relevant, then introducing random modification of elements and rescoring them until a predetermined level is reached.[22] Here, the scoring system should reflect a similarity measure to the template macrocyclic peptide; modifications will ensure greater similarity, and a level at which the algorithm can end should be determined.

Developing this scoring system represents a computationally intense area of research: a neural network will be employed to learn from known data to produce a generalized similarity calculation. This known data, otherwise known as a training set, will be developed using existing crystal structures of macrocyclic peptide and protein interactions. Using the few known structures and docking software, the free energy of association to the target of both the macrocyclic peptide and a set of fragment-based macrocycles can be compared, representing a similarity value. From analyzing macrocycles with both similar and dissimilar binding energies, salient characteristics key to binding can be identified, and used for a generalized metric.

Overall, this approach should result in an easily repeatable technique for identifying pharmaceutically useful macrocyclic leads for "undruggable" targets. Incorporating the ease of high throughput screening for biologically synthesized compounds with computational methods will avoid a lengthy synthetic process to individually test macrocycles of unknown affinity. Identifying high affinity molecules without knowing the structure of the target represents a significant advance in drug development techniques.

Appendix A

Supplemental Information

A.1 General Information

Reagents purchased from commercial sources (Sigma Aldrich, Oakwood, Alfa Aesar, Fluka, and Fischer scientific) were used as received, without further purification. All reactions were performed under nitrogen atmosphere using oven-dried glassware using standard Schlenk technique, unless otherwise stated. Anhydrous dichloromethane (DCM), tetrahydrofuran (THF), and toluene were obtained through passage through alumina using a *GlassContours* solvent system. All other solvents used were used as received from commercial suppliers without further purification.

$AgBF_4$, $AgSbF_6$, $AfNTf_2$, $AgBAr^F_4$, AgOAc, and $[RhCp^*Cl_2]_2$ were stored and weighed in a nitrogen-filled glovebox. $[RhCp^*Cl_2]_2$ was synthesized using previously reported methods.

^1H and ^{13}C Nuclear Magnetic Resonance (NMR) was performed on a Varian Inova 600 spectrometer (600 MHz ^1H, 151 MHz ^{13}C), a Bruker 600 spectrometer (600 MHz ^1H, 151 MHz ^{13}C),a Varian Inova 500 spectrometer (500 MHz ^1H, 126 MHz ^{13}C), and aVarian Inova 400 spectrometer (400 MHz ^1H, 100 MHz ^{13}C) at room temperature in $CDCl_3$ with internal $CHCl_3$ as the reference (7.26 ppm for ^1H, 77.16 ppm for ^{13}C), unless otherwise stated. Chemical shifts (δ values) were reported in parts per million (ppm).

Analytical thin layer chromatography (TLC) was performed on precoated glass backed Silicycle SiliaPure 0.25 mm silica gel 60 plates and visualized with UV light or ethanolic p-anisaldehyde. Flash column chromatography was performed using Silicycle SiliaFlash F60 silica gel (40-63 μm) on a Biotage IsoleraOne system. Preparatory TLC was performed on precoated glassbacked Silicycle SiliaPure 1.0 mm silica gel 60 plates.

A.2 Allylic Amination

General Procedure: Optimization Table

$[RhCp^*Cl_2]_2$ (2 mol%, 0.0026 mmol), halide scavenger (8 mol%, 0.0104 mmol), AgOAc (2.1 eq, 0.273 mmol), and NH_2Ts (2.5 eq, 0.325 mmol) were weighted out into an oven-dried 4 mL vial with a stir bar in the glove box. The vial was then fitted with a cap and septum and removed. Allylbenzene (1 eq, 0.13 mmol) and nonane (internal standard, 0.5 eq, 0.065 mmol) were added via microliter syringe, followed by 0.65 mL of solvent. The reaction mixture was then heated and stirred for 24 hours, at which point a 50 μL aliquot was taken via microliter syringe and filtered into a 2 mL vial through celite with DCM. The sample was then run and analyzed through prepared calibration curves to calculate a yield.

Gas Chromatography

Calibration Curves can be seen in Figures A.1, A.2, and A.3.

Tosyl Protection of 1-(naphthalen-1-yl)ethan-1-amine (S1)

S1, p-toluenesulfonamide chloride (1 eq, 4.6 mmol), and triethylamine (1 eq, 4.6 mmol) were added to a 50 mL oven-dried round bottom flask fitted with a stir bar, then capped with a septum. 25 mL of DCM were added and the mixture was heated to 45°C and allowed to run overnight. The following morning the reaction was quenched with 35 mL of water. An extraction was then performed in DCM (3 x 15 mL) then dried using sodium sulfate. The reaction mixture was then filtered and concentrated under reduced pressure. Column chromotagraphy was run in Hex:EtOAc 0→30% to yield 1.256 g of

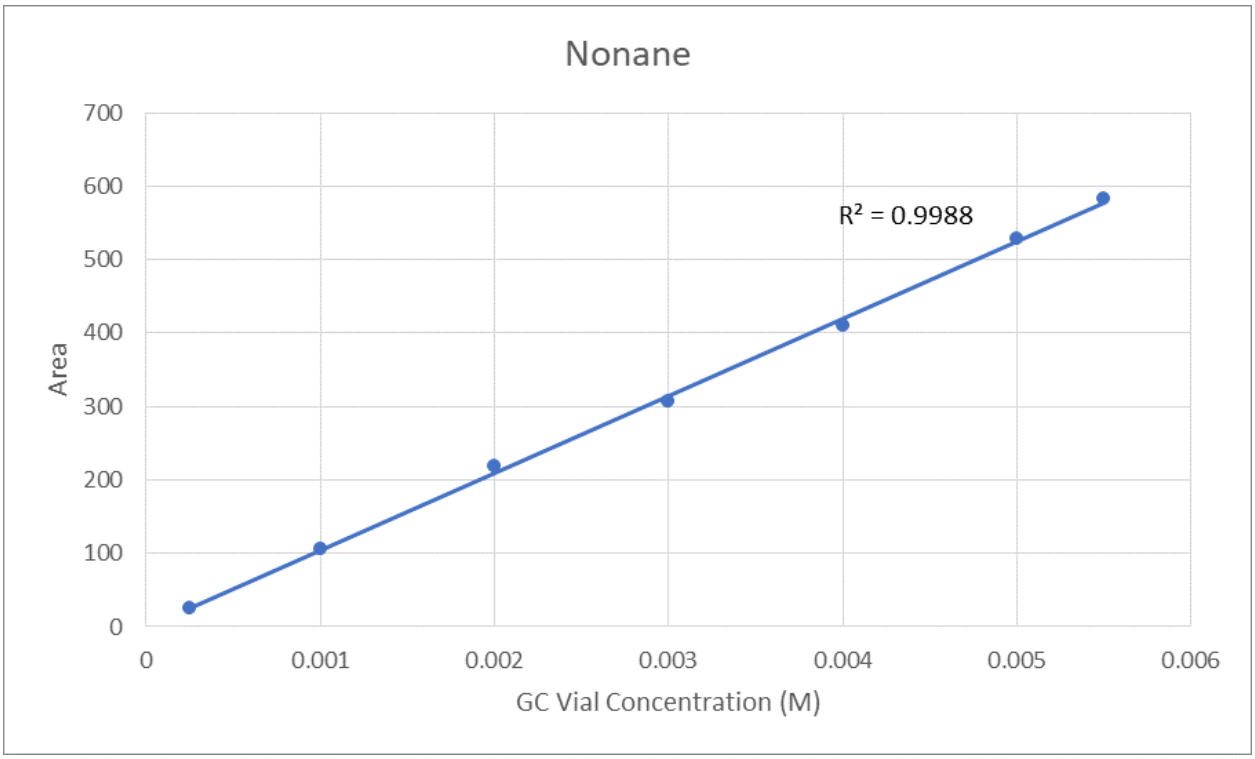


Figure A.1: Calibration Curve of Nonane

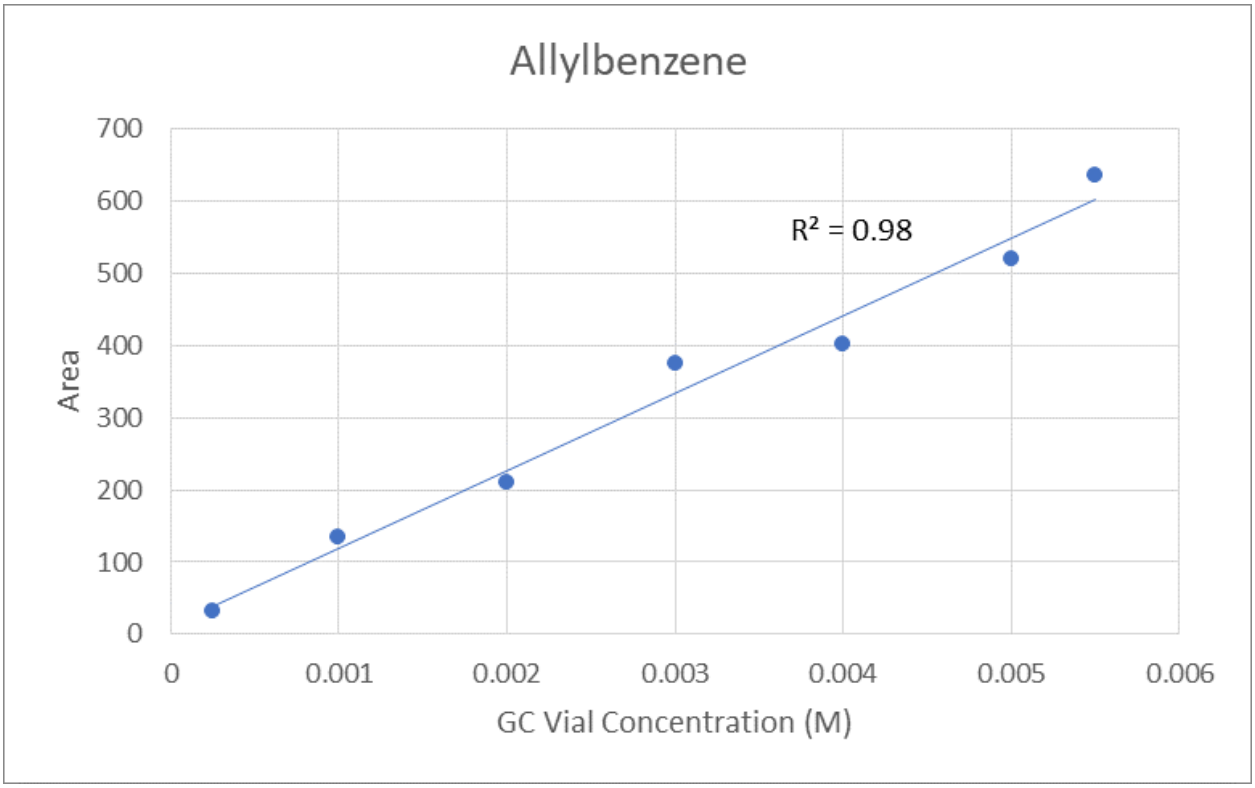


Figure A.2: Calibration Curve of Allylbenzene

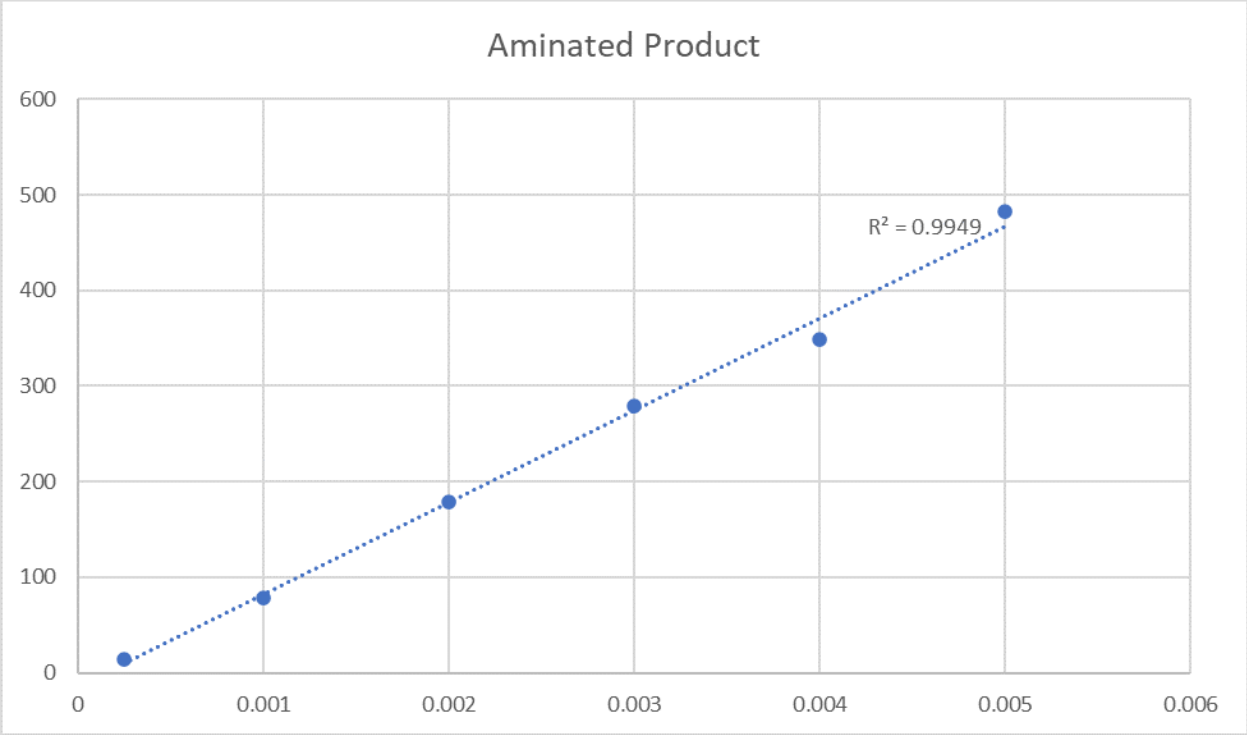


Figure A.3: Calibration Curve of Aminated Allylbenzene

4-methyl-N-(1-(naphthalen-1-yl)ethyl)benzenesulfonamide, 84% yield.

Amination of Allylbenzene with 4-methyl-N-(1-(naphthalen-1-yl)ethyl)benzenesulfonamide (S2)

$[RhCp^*Cl_2]_2$ (2 mol%, 0.0026 mmol), $AgBF_4$ (8 mol%, 0.0104 mmol), AgOAc (2.1 eq, 0.273 mmol), and **S2** (2.5 eq, 0.325 mmol) were weighted out into an oven-dried 4 mL vial with a stir bar in the glove box. The vial was then fitted with a septum and cap and removed. Allylbenzene (1 eq, 0.13 mmol) was added via microliter syringe, followed by 0.65 mL of chlorobenzene. A preperatory TLC was run in 8%:12%:80% mixture of EtOAc: DCM: Hex (v/v/v). By NMR, a 13% yield of vinyl naphthalene was calculated.

Oxidation of S2 with $AgBF_4$

S2 (2.5 eq, 0.125 mmol) and $AgBF_4$ (0.20 eq, 0.01 mmol) were added to an oven-dried 4 mL vial in the glovebox. A stir bar was added, then the vial was fitted with a septum and cap and removed. 0.25 mL of dioxane were then added via syringe. The reaction was monitored by TLC, and elimination to form vinylnaphthalene was confirmed by NMR.

Preparation of Aminated Allylbenzene Derivatives for Isolation

General Procedure

$[RhCp^*Cl_2]_2$ (2 mol%, 0.005 mmol), halide scavenger (8 mol%, 0.02 mmol), AgOAc (2.1 eq, 0.525 mmol), and NH_2Ts (2.5 eq, 0.625 mmol) were weighted out into an oven-dried 4 mL vial with a stir bar in the glove box. The vial was then fitted with a cap and septum and removed. The allylbenzene derivative was then added via microliter syringe, followed by 1.25 mL of dioxane. The reaction mixture was then allowed to stir for 24 hours at 100C. The following day, the mixture was filtered through celite with ethyl acetate and a crude NMR was taken, followed by column chromatography.

N-cinnamyl-4-methylbenzenesulfonamide

Prepared using the General Procedure. A hand column was run in EtOAc: Hex 0%→100%, with product coming off at 20:80. The fractions were concentrated under reduced pressure to yield product (0.0467 g, 65% yield). NMR values were in accordance with published values.

(E)-N-(3-(4-fluorophenyl)allyl)-4-methylbenzenesulfonamide

Prepared using the General Procedure. A column was run in EtOAc: Hex 0%→100%, with product coming off at 20:80. Two fractions were split, and product was recovered using a prep TLC. The material was concentrated under reduced pressure to yield a combined total of 0.0365 g of product, 48% yield.

(E)-4-methyl-N-(4-phenylbut-3-en-2-yl)benzenesulfonamide

Prepared using the General Procedure. Column chromatography was performed in EtOAc: Hex 0%→50%. An inseparable mixture of regioisomers in a 1.5:1 ratio by NMR were isolated for a combined total of 0.01 g, or 13% yield.

(E)-4-methyl-N-(3-(4-(trifluoromethyl)phenyl)allyl)benzenesulfonamide

Prepared using the General Procedure. Column chromatography was performed in EtOAc: Hex 0%→80%. Fractions were concentrated under reduced pressure to yield product (0.0096 g, 11% yield).

(E)-4-methyl-N-(3-(o-tolyl)allyl)benzenesulfonamide

Prepared using the general procedure. Column chromatography was performed in EtOAc: Hex 0%→80%. Fractions were concentrated under reduced pressure to yield product (0.0119 g, 16% yield).

methyl(E)-2-((tert-butoxycarbonyl)amino)-3-(4-(3-((4-methylphenyl)sulfonamido)prop-1-en-1-yl)phenyl)propanoate

Solids were weighed out in the glovebox according to the general procedure. A solution of methyl 3-(4-allylphenyl)-2-((tert-butoxycarbonyl)amino)propanoate and 1.25 mL of dioxane was made in an oven-dried 4 mL vial under nitrogen. This solution was then added to the solids and the reaction was stirred and heated at 100 C for 24 hours. A crude NMR was taken and no evidence of product was seen.

(E)-N-(3-(4-methoxyphenyl)allyl)-4-methylbenzenesulfonamide

Prepared using the general procedure. Column chromatography was performed in EtOAc: Hex 0%→75%. By NMR, trace amounts of product were isolated.

(E)-N-(3-(benzo[d][1,3]dioxol-5-yl)allyl)-4-methylbenzenesulfonamide

Prepared using the general procedure. A column was performed in EtOAc: Hex 0%→100%, with product coming off at 40:60. Fractions were further separated using prep TLC in 50% petroleum ether: 50% ether. It was run twice to eventually yield 0.009 g of product, representing 11% yield.

A.3 Computational Studies

All calculations were conducted using DFT as implemented in the Jaguar 9.1 suite of ab initio quantum chemistry programs with Becke's three-parameter exchange functional B3LYP including Grimme's D3 dispersion correction levels of theory. Geometry optimizations were proceeded using the 6-31G** basis set. Rhodium was represented using the Los Alamos LACVP basis that includes relativistic effective core potentials. The energies of

the optimized structures were reevaluated by additional single point calculations on each optimized geometry using the same functional and Dunning's correlation consistent triple-ζ basis set cc-pVTZ(-f) which includes a double set of polarization functions.

Analytical vibrational frequencies within the harmonic approximation were calculated using the 6-31G** basis to confirm proper convergence to well-defined minima or saddle points on the potential energy surface. Solvation energies were calculated using a self-consistent reaction field (SCRF) approach based on accurate numerical solutions of the Poisson-Boltzmann equation and were performed with the 6-31G** basis at the optimized gas phase geometry with the dielectric constant of $\epsilon = 10.3$ for 1,2-dichloroethane (1,2-DCE). As is the case for all continuum models, the solvation energies are subject to empirical parametrization of the atomic radii that are used to generate the solute surface. The standard set of optimized radii in Jaguar was used for H (1.150 $\mathring{A}$), C (1.900 $\mathring{A}$), N (1.600 $\mathring{A}$), O (1.600 $\mathring{A}$), and Rh (1.464 $\mathring{A}$). The Gibbs free energies in solution phase G(sol) were computed with the following protocol.

$$G(Sol) = G(gas) + G^{solv} \tag{A.1}$$

$$G(gas) = H(gas) - TS(gas) \tag{A.2}$$

$$H(gas) = E(SCF) + ZPE \tag{A.3}$$

$$\Delta E(SCF) = \sum_{products} E(SCF) + \sum_{reactants} E(SCF) \tag{A.4}$$

$$\Delta G(sol) = \sum_{products} G(sol) + \sum_{reactants} G(sol) \tag{A.5}$$

G(gas) is the free energy in gas phase; G^{solv} is the free energy of solvation; H(gas) is the enthalpy in gas phase; T is the temperature (313.15K); S(gas) is the entropy in gas phase; E(SCF) is "raw" electronic energy as computed from the SCF procedure which is the self-consistent field energy, and ZPE is the zero point energy. The entropy we refer to is specifically vibrational/rotational/translational entropy of the solute(s), and the entropy of

the solvent is implicitly comprised in the continuum solvation model.

A.4 Macrocycle Starting Set

General

Molecular manipulation and storage was done through rdkit packages as accessed through Anaconda Powershell Prompt (Anaconda3). Simplified molecular-input line-entry system (SMILES) codes were used for fragment and molecule storage, and SMILES arbitrary target specification (SMARTS) codes were used for pattern recognition for filtering and reaction definition. All coding was done in python, using packages os, random, and rdkit.

Fragment Library

The Life Chemicals General Fragment Library was downloaded and filtered according to the Rule of Three criteria as shown in Table. Results were processed through Wolfram Mathematica for data visualization.

Properties	RO3
Molecular Weight (MW)	≤ 300
Log P	≤ 3
Hydrogen Bond Donors (HBD)	≤ 3
Hydrogen Bond Acceptor (HBA)	≤ 3
Rotatable Bonds (ROTB)	≤ 3
Polar Surface Area (PSA)	≤ 60

Random Connections

A method for building a stack of fragments to be combined to form a molecule was defined, taking the number of fragments as an input. The fragment library was imported as a list, then a stack was defined as a double-ended queue and elements at randomly selected indices were pushed onto the stack. Once this had been performed a specified number of times as indicated by the input value, the randomly generated fragment stack was returned.

Another method was defined taking the stack as an input. The last two elements were popped, with the bond formed somewhere randomly between these two fragments.

Using rdkit.Chem.rdchemreactions, SMARTS codes were used to define reactions taking an undefined atom, i.e. not defined as any atom type, with hydrogens attached and forming a bond with another undefined atom with hydrogens attached. Separate reactions were defined for all possible reactions, then the generated possible results were added to a single list. One element from that list was then randomly selected, and pushed back onto the stack. This process was repeated until the stack had only one element, which was then popped and passed as the result of the method.

These two methods were combined into a new method taking the number of fragments per molecule and the number of molecules generated as the arguments, and a list of possible molecules were returned.

Macrocyclization

Defining a reaction to cyclize these molecules required SMARTS codes, as the atom indicies were defined by the order of the SMILES code not the placement on the molecule (i.e. the addBond method couldn't be used). Defining the code to run a reaction to only create rings of size 12 or greater proved to be impossible; the alternate solution is to run random intramolecular reactions, then use the SMARTS codes to select from this list those containing a ring of size 12 or greater. The reaction definition from the random fragment connection was modified to provide intramolecular products, then the products were combined into a list. Every item in the list was then compared to the SMARTS criteria and a new list was defined to contain only those for which this comparison was true. A macrocycle from this list was then randomly selected as one to be returned from the method.

Bibliography

(1) Trost, B. M. Metal catalyzed allylic alkylation: its development in the Trost Laboratories. *Tetrahedron* **2015**, *71*, 5708-5733.

(2) Engelin, C. J.; Fristrup, P. Palladium Catalyzed Allylic C–H Alkylation: A Mechanistic Perspective. *Molecules* **2011**, *16*, 951-969.

(3) Chen, M. S.; White, M. C. Sulfoxide-Promoted, Catalytic Method for the Regioselective Synthesis of Allylic Acetates from Monosubstituted Olefins via C–H Oxidation. *J. Am. Chem. Soc.* **2004**, *126*, 1346-1347.

(4) Howell, J. M.; Liu, W.; Young, A. J.; White, M. C. General Allylic C–H Alkylation with Tertiary Nucleophiles. *J. Am. Chem. Soc.* **2014**, *136*, 5750-5754.

(5) Cochet, T.; Bellosta, V.; Roche, D.; Ortholand, J.-Y.; Grenier, A.; Cossy, J. Rhodium(III)-catalyzed allylic C–H bond amination. Synthesis of cyclicamines fromx-unsaturatedN-sulfonylamines. *Chem. Commun.* **2012**, *48*, 10745-10747.

(6) Burman, J. S.; Blakey, S. B. Regioselective Intermolecular Allylic C–H Amination of Disubstituted Olefins via Rh-π-allyl Intermediates. *Angew. Chem. Int. Ed.* **2017**, *56*, 13666-13669.

(7) Reddy, R. P.; Davies, H. M. L. Dirhodium Tetracarboxylates Derived from Adamantylglycine as Chiral Catalysts for Enantioselective C–H Aminations. *Org. Lett.* **2006**, *8*, 5013-5016.

(8) Harris, R. J.; Park, J.; Nelson, T. A. F.; Iqbal, N.; Salgueiro, D. C.; Bacsa, J.; MacBeth, C. E.; Baik, M.-H.; Blakey, S. B. The Mechanism of Rhodium-Catalyzed Allylic C–H Amination. *J. Am. Chem. Soc.* **2020**, *142*, 5842-5851.

(9) Kazerouni, A. M.; Nelson, T. A. F.; Chen, S. W.; Sharp, K. R. Regioselective Cp*Ir(III)-Catalyzed Allylic C–H Sulfamidation ofAllylbenzene Derivatives. *J. Org. Chem.* **2019**, *84*, 13179-13185.

(10) Calhorda, M. J.; Romao, C. C.; Veiros, L. F. The Nature of the Indenyl Effect. *Chem. Eur. J.* **2002**, *8*, 868-875.

(11) Trost, B. M.; Ryan, M. C. Indenylmetal Catalysis in Organic Synthesis. *Angew. Chem. Int. Ed.* **2016**, *56*, 2862-2879.

(12) Burman, J. S.; Harris, R. J.; Farr, C. M. B. F.; Simon, S. B. Rh(III) and Ir(III)Cp* Complexes Provide Complementary Regioselectivity Profiles in Intermolecular Allylic C–H Amidation Reactions. *ACS Catal.* **2019**, *9*, 5474-5479.

(13) Knecht, T.; Mondal, S.; Ye, J.-H.; Das, M.; Glorius, F. Intermolecular, Branch-Selective, and Redox-Neutral Cp*Ir^{I}II-Catalyzed Allylic C–H Amidation. *Angew. Chem. Int. Ed.* **2019**, *58*, 7117-7121.

(14) Makley, L. N.; Gestwicki, J. E. Expanding the Number of "Druggable" Targets: Nonenzymes and Protein-Protein Interactions. *Angew. Chem. Int. Ed.* **2013**, *81*, 22-32.

(15) Huang, Y.; Wiedmann, M. M.; Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **2019**, *119*, 10360-10391.

(16) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discovery* **2008**, *7*, 608-624.

(17) Saha, I.; Dang, E. K.; Svatunek, D.; Houk, K. N.; Harran, P. G. Computational generation of an annotated gigalibraryof synthesizable, composite peptidic macrocycles. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 24679-24690.

(18) Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discovery* **2013**, *12*, 644-645.

(19) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235-249.

(20) Shi, Y.; von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24*, 2838.

(21) Tyagi, M.; Begnini, F.; Poongavanam, V.; Doak, B.; Kihlberg, J. Drug Syntheses Beyond the Rule of 5. *Chemistry* **2020**, *26*, 49-88.

(22) Al-Sahaf, H.; Bi, Y.; Chen, Q.; Lensen, A.; Mei, Y.; Sun, Y.; Tran, B.; Xue, B.; Zhang, M. A Survey on Evolutionary Machine Learning. *J. R. Soc. N. Z.* **2019**, *49*, 205-228.