Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a 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 in whole or in part in all forms of media, now or hereafter now, 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.

Signature:

Mengfei Xu March 22, 2022

Computational Development of Macrocyclic Drug Hits and Synthetic Strategy towards Peptide Macrocyclization

By

Mengfei Xu

Simon Blakey, Ph.D.

Advisor

Chemistry

Simon Blakey, Ph.D.

Advisor

John Heemstra, Jr., Ph.D.

Committee Member

Didem Uca, Ph.D.

Committee Member

2022

Computational Development of Macrocyclic Drug Hits and Synthetic Strategy towards Peptide Macrocyclization

By

Mengfei Xu

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

2022

Abstract

Computational Development of Macrocyclic Drug Hits and Synthetic Strategy towards Peptide Macrocyclization

By Mengfei Xu

Protein-protein interactions (PPIs) are often considered "undruggable" targets because they are difficult to inhibit by conventional small molecule drugs. Macrocyclic peptides have emerged as a promising class of therapeutic agents because of their capacity to resemble PPIs, along with several important pharmacological advantages. In particular, macrocyclization is considered as a strategy to improve the oral bioavailability of drug candidates. Therefore, it has also been applied to non-peptide molecules to enhance their druggability. Herein, we will describe our idea and efforts to develop machine learning models to identify potential non-peptide macrocyclic drug hits from a randomly generated library of macrocyclic compounds. This will minimize the time and cost of purchasing, synthesizing, and screening a large library of molecules and accelerate the current drug development process. In addition, we will report on our progress towards peptide macrocyclization via cobalt catalysis. This new approach will diversify the toolkit through the use of an environmentally friendly and earth-abundant transition metal. The future integration of the two divisions, computational and synthetic chemistry, will certainly advance our understanding of the synthesis and applications of macrocycles in the pharmaceutical industry.

Computational Development of Macrocyclic Drug Hits and Synthetic Strategy towards Peptide Macrocyclization

By

Mengfei Xu

Simon Blakey, Ph.D.

Advisor

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

2022

Acknowledgments

Four years of college seemed to pass by in a flash, but I encountered many problems during these years. However, because of the support of many people, I persevered with my love for chemistry research. Therefore, I would like to express my sincere gratitude to these people.

First of all, I would like to thank my research supervisor, Professor Simon Blakey, for allowing me to join this group. This experience helped me realize my love for research and led to my decision to pursue graduate school. Of course, I also appreciate your constant guidance, recognition, and encouragement during my two years in the group, especially during the application season. I would also like to thank Professor John Heemstra, who has taught me three different classes. Your passionate and inspiring teaching style in CHEM 203 drew me into the world of organic chemistry. Also, deep thanks to Professor Uca, with whom I studied German for one year. Your kindness has been a great support in my life, particularly during the tough times of COVID.

I also need to mention all the past and present members of the Blakey lab. Chris, you were a great mentor in helping me not only learn all the techniques but also develop my chemistry mindset and understand the process of research. It was great to be in the same lab room with you and my semi-mentor Michael as you both made me feel included when I first entered the lab after COVID. Michael, I always enjoyed chatting with you, not only about chemistry but also about school life and random things, which was a big relief when I felt stressed. Christina, I appreciate your patience with me while I was using the solid-phase peptide synthesizer. Also, a big thank you to Eleda, who arranged different activities for me to fit in with the group and was a great emotional supporter! I also want to thank Kim, who was the first person I met in the lab and coached me as a peer mentor during my first project.

Last but not least, I have to thank my best friend, Anna Cai, from the Davies group. Thank you for being an excellent listener to my concerns and excitement about research. Thank you also for your encouragement and affirmation when I felt unsure and doubted myself.

It is because of all of you that I have been able to come this far and stay on my path towards more chemistry. It was a pleasure to meet you all and I will miss my time in the Blakey lab and at Emory very much.

Table of Contents

List of Figures

List of Schemes

List of Tables

Chapter ONE

Introduction to Macrocycles and Drug Development

Direct interactions between proteins have become a focus in drug discovery because of their importance in regulating biological systems. However, difficulties in developing small molecules that inhibit these protein-protein interactions (PPIs) have led them to be considered challenging or even "undruggable" targets.[1] Peptides are superior to small molecules as drug candidates for their possibility to engage large shallow protein surfaces through multiple hydrogen bonding interactions, which enables to them resemble the functionality of PPIs. Despite both being robust natural agonists/antagonists, linear and cyclic peptides are different in terms of their success in pharmaceutical applications.

Macrocyclic peptides, with at least one ring containing twelve or more atoms, offer several pharmacological advantages over linear peptides. First, the cyclic structure provides a degree of pre-organization while retaining sufficient flexibility to maximize binding interactions between key functional sites without significant entropic loss.[2] Due to intramolecular hydrogen bonding, macrocycles may also adopt an inside-out conformation to accommodate their immediate environment leading to significantly improved passive membrane permeability and aqueous solubility. In addition, hydrolytic enzymes are unable to process macrocyclic peptides as their active sites are naturally designed for linear substrates. Thus, macrocyclic peptides have better metabolic stability compared to their linear counterparts. Other advantages of macrocyclic peptides include their high potency and specificity; both can be attributed to the large surface area of the ring system. Currently, over forty cyclic peptides, including vancomycin and daptomycin,[3] (**Figure 1**) have been approved by the FDA for clinical use, and many others have been documented as successful therapeutic agents.

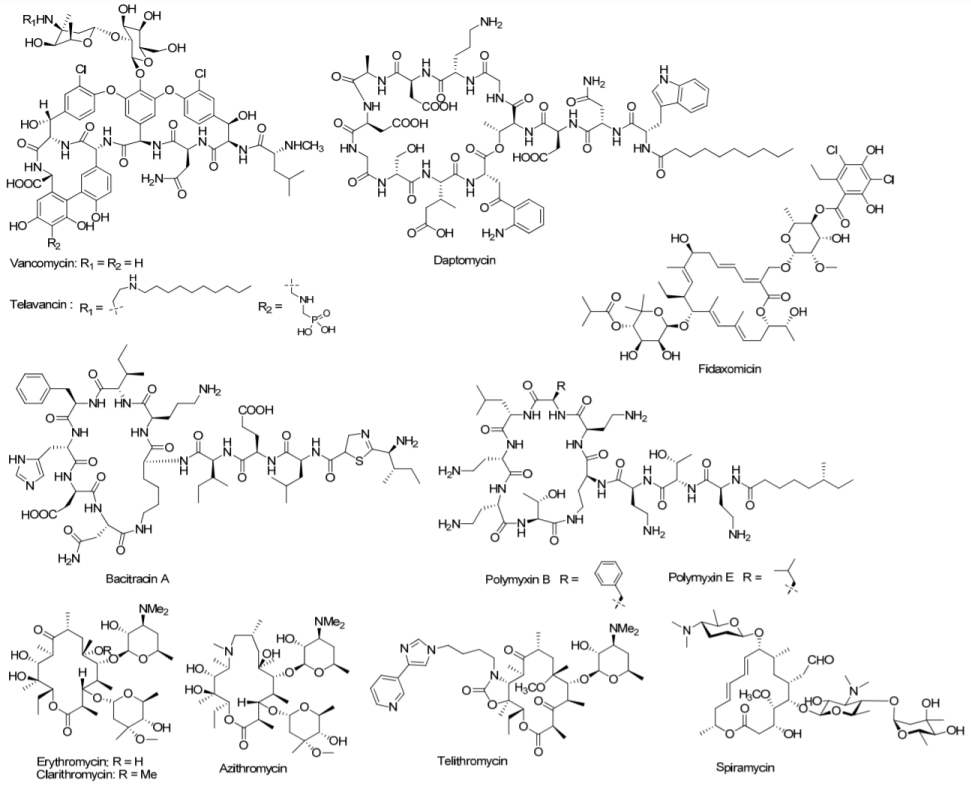


Figure 1: Examples of clinically used macrocyclic compounds.[4]

Similar to small molecules, macrocyclic peptide drugs are typically identified via *de novo* drug discovery, which involves a hit-to-lead stage and lead optimization. The "hit" is a compound that has the desired activity, albeit usually a poor interaction with the target, out of a large library of molecules being screened. After identifying the hits, the results are confirmed and evaluated. Additionally, several analogs of the hits are analyzed as well to find the ones with the best affinities, a process known as hit expansion. The overall goal of lead optimization is to synthesize the "lead" compound, which has not only improved potency but also better physiochemical properties, to move forward into the clinic.

Two primary methods, high-throughput screening (HTS) and fragment-based drug discovery (FBDD), are used to design these macrocyclic drug candidates.[5] In HTS, the initial hit identification requires a large library for compound screens. Libraries of macrocyclic peptides are available through the rapid linkage of orthogonally protected amino acid building blocks. This mechanism is established from DNA and RNA sequences which can be translated in vivo and/or

in vitro into amino-acid-containing oligomers. These innovative methods for building DNA- or RNA-encoded libraries can be integrated with HTS to screen a largely expanded library (10^7 – 10^{13}) of highly diverse peptides. The second strategy, FBDD, starts with the detection of a small library of fragments that can bind to a specific protein target. The library contains molecules with smaller molecular weights and therefore lower binding affinity, but also more desired interactions and better pharmaceutical properties than those in the HTS library. Selected fragments are then linked to improve potency. FBDD libraries are also easier to maintain and screen for their smaller size.

Although progress has been made in increasing the size of libraries or developing potential interactions, the utility of macrocyclic molecules in medicinal chemistry remains limited. The primary reason is that macrocycles are not regarded as common drug-like molecules because most of them do not adhere to Lipinski's Rule of Five (Ro5),[6] a widely used standard for assessing druggability. The rule predicts a poor oral bioavailability if a molecule has a molecular mass of more than 500 Da, an octanol-water partition coefficient exceeding 5, and more than 5 hydrogen-bond donors and 10 hydrogen-bond acceptors. Interestingly, excluding natural products and substrates of transporters, 16% of oral drugs do not follow Ro5, and 6% violate two or more property criteria.[7] This discrepancy stimulated investigation and led to the origination of the concept, "beyond" the Ro5 (bRo5),[8] referring to the drug candidates that do not meet Ro5 in a strict manner but may still be bioavailable. These drugs can usually modulate novel targets that have difficult binding sites, which spurs interest in the potential of bRo5 chemical space. In particular, macrocyclization is recognized as a method to induce bioavailability and advance drug-like properties. This approach is also applied to modify non-peptidic drug candidates towards similar ends, especially because they do not have as many unstable amide bonds as typical peptides. In response to the limitations outlined above, my two projects have centered on furthering the knowledge of macrocyclization through the construction of both types of macrocyclic compounds.

Chapter TWO

Computational Construction of Macrocyclic Libraries Towards Drug Design

In this chapter, we will discuss a workflow design to computationally score a library of macrocycles based on a given target. The model is expected to compare the macrocycles with a cyclic peptide known to interact with the target. We have achieved the construction of macrocyclic libraries by applying a strategy similar to FBDD. This project was conducted in collaboration with Kimberly R. Sharp, a former undergraduate in the Blakey lab.

2.1 In Silico Research in Drug Development

With the exponential growth of protein targets and expansion of library sizes, the drawbacks of traditional screening methods, such as the cost of purchasing libraries and the inefficiency of identification, have become apparent. In a period of rapid development in in silico research, virtual screening has promoted drug discovery to a new level. Designated for "hits" discovery, virtual screening can be combined with HTS and FBDD to computationally predict binding information for large initial libraries and reduce the number of compounds to be tested.

This implementation of virtual screening has stimulated a "gold-rush" of machine learning in drug discovery, making it possible to screen and process libraries at a much larger scale. Pioneering this field, Bhat and coworkers reported a model in 2019 that aimed to perform computational hit expansion and lead optimization, referred to as "PathFinder".[9] The method started with breaking down a known lead molecule to its respective starting materials, and a specific region (R-group) could be explored and replaced through a one-step reaction (**Figure 2.1a**). Enumeration of the R-groups generated a new library, and these compounds were filtered based on their properties. Finally, the resulting molecules were docked computationally followed by a machine learning model to predict the potency (**Figure 2.1b**). Although this workflow was

able to avoid the expensive Free Energy Methods (FEP) potency calculations, it still required direct docking. However, docking is not always a choice since the crystal structure of the target is not always available.

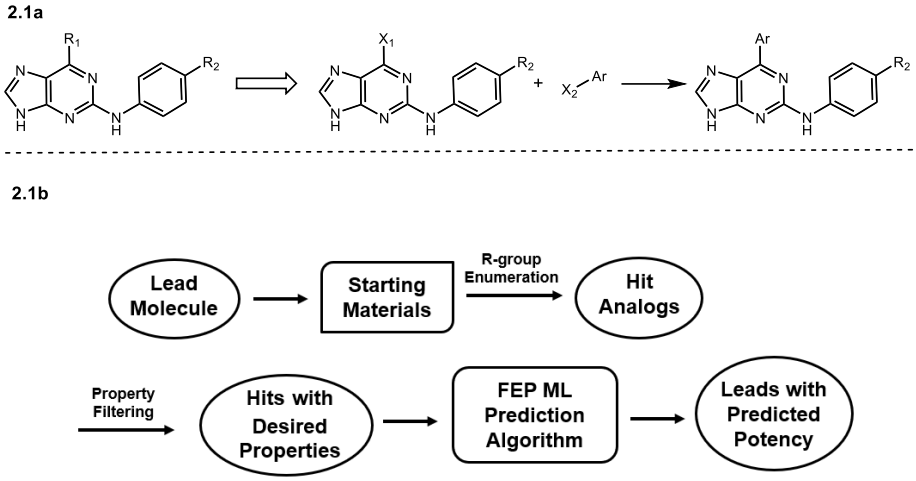


Figure 2.1: PathFinder workflow to enumerate hit analogs and select potential leads.

Another model for computational generation of peptidic macrocycle libraries, Composite Peptide Macrocycle Generator, was published by Harran and coworkers.[10] They chose to formulate several heterocycles into amino acids and convert them into oligomers. These were bound to a template to enumerate the macrocyclic products, which were then filtered based on desired properties (**Figure 2.2**). All the connections were based on known reactions. Therefore, with this method, a Giga ($> 2 \times 10^9$) library of synthesizable macrocycles was produced. This process stretched the size of computational libraries but restricted further use of machine learning in structure improvement as it could easily destroy the synthesizability of the molecules.

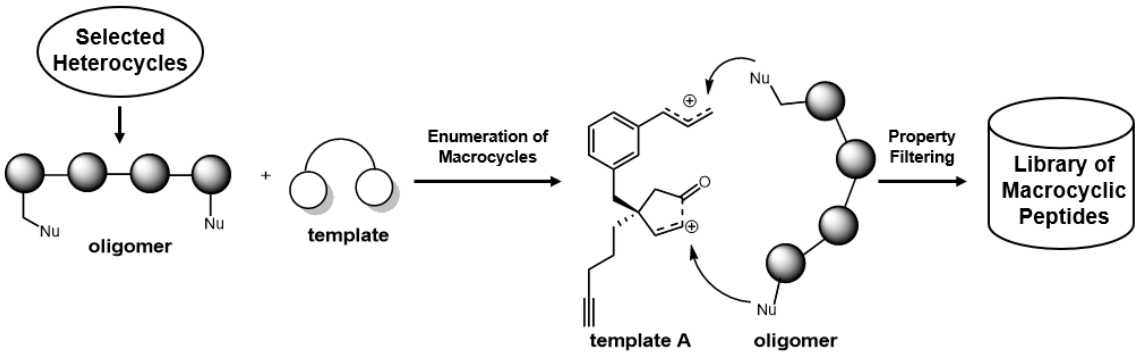


Figure 2.2: Composite Peptide Macrocycle Generator workflow to enumerate macrocyclic peptides.

Inspired by PathFinder, we decided to develop two machine learning models to produce non-peptide macrocycles with similar functionality using cyclic peptides recognized as hits for specific proteins.[11] Starting from a library of macrocycles, we could create a neural network model that would concentrate on binding affinity calculations, comparing properties important for docking, and selecting the compounds with similar interactions to the target as the cyclic peptide. An evolutionary algorithm would employ the molecules selected by the previous model as a training set and learn to generate a weighted evaluation system on the similarity between macrocycles and the peptide without structural information. It could then adjust and improve the structure of a set of macrocycles based on the scoring. In the end, the compounds being suggested could be synthesized, and in vitro testing would be conducted to confirm their properties. The starting set of macrocycles in the workflow, illustrated in **Figure 2.3**, was implemented in Python using the RDKit cheminformatics toolkit. The robust open package would allow the development of machine learning algorithms with chemical information and functionality supported, e.g., substructure searching, editing molecules, 2D depictions, conformational analysis, etc.

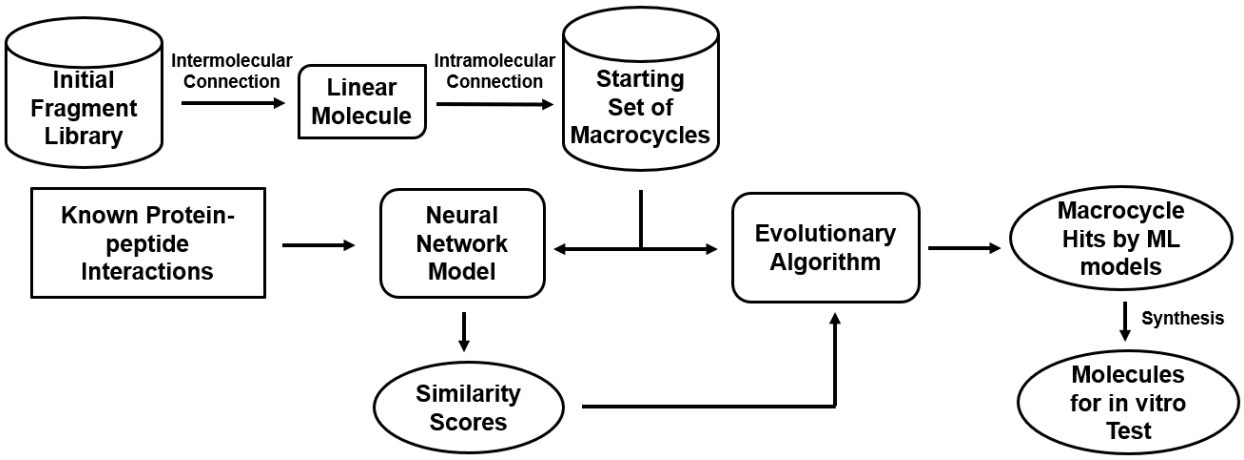


Figure 2.3: Workflow for the computational development of macrocyclic hits.

2.2 Results and Discussion

Integrating the idea of FBDD, building an initial fragment library could begin with an online small molecule database. These molecules needed to be fragmented and filtered according to the desired properties to produce a new fragment library. Alternatively, a commercially available fragment library could be obtained directly with a subsequent property filtering. Working with a molecule database was much more complicated, requiring excellent knowledge of molecular fragmentation, storage of fragments, and data cleaning. The resulting fragments were also difficult to control. Hence, we chose to start with an existing free library designed for FBDD projects, Life Chemicals general fragment library (14,947 fragments). Most of the fragments included follow Rule of Three (Ro3),[12] an analog of Ro5 tailored to fragments in FBDD to ensure the final molecules are likely to be orally available. Two other factors that influence fragment library design are size and diversity. Shi and von Itzstein investigated the diversity and size relationship, concluding that only about 2,000 fragments are required to attain the same level of diversity as an FBDD library with more than 200,000 fragments.[13]

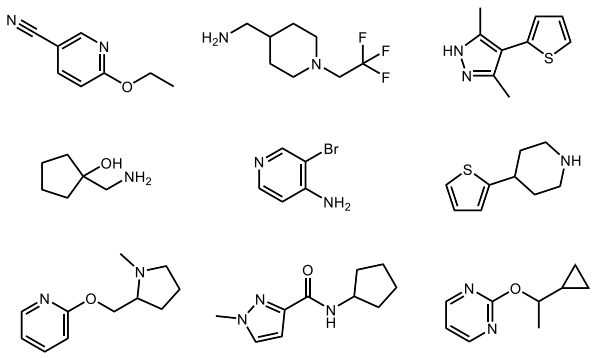


Figure 2.4: Randomly selected fragments from the filtered fragment library.

Combining the two pieces of information, the filtering criteria were set to molecular weight (MW) ≤ 200 Da, octanol-water partition coefficient (logP) ≤ 3, number of hydrogen-bond donors (HBD) ≤ 3, number of hydrogen-bond acceptors (HBA) ≤ 3, number of rotatable bonds (ROTB ≤ 3), polar surface area (PSA) ≤ 50, and heavy atom counts (HAC) ≤ 20. For MW and PSA, limits stricter than Ro3 were implemented to narrow down the eligible candidates to around 2,000 with the least adjustments on the rules. The additional upper limit on HAC was derived from the typical property of fragments.[12] Data analysis on the filtered fragments also included fraction of sp^3 (FSP3) and quantitative estimate of drug properties (QED)[7] (**Figure 2.5**). QEDs were chosen to evaluate the fragments and, later, the linear and cyclic molecules, based on the properties of amine-acid coupling products reported by Mahjour *et al.*[14] The metric could range from zero to one, with a larger value suggesting more favorable drug properties.[7] The ranges of properties for the final 2,252 fragments in the library are: MW(Da) = 100.01 – 199.99, LogP = -1.16 – 2.96, HBD = 0 – 3, HBA = 1 – 3, ROTB = 0 – 3, PSA = 3.24 – 49.93, FSP3 = 0.00 – 1.00, HAC = 6 – 15, QED = 0.39 – 0.81.

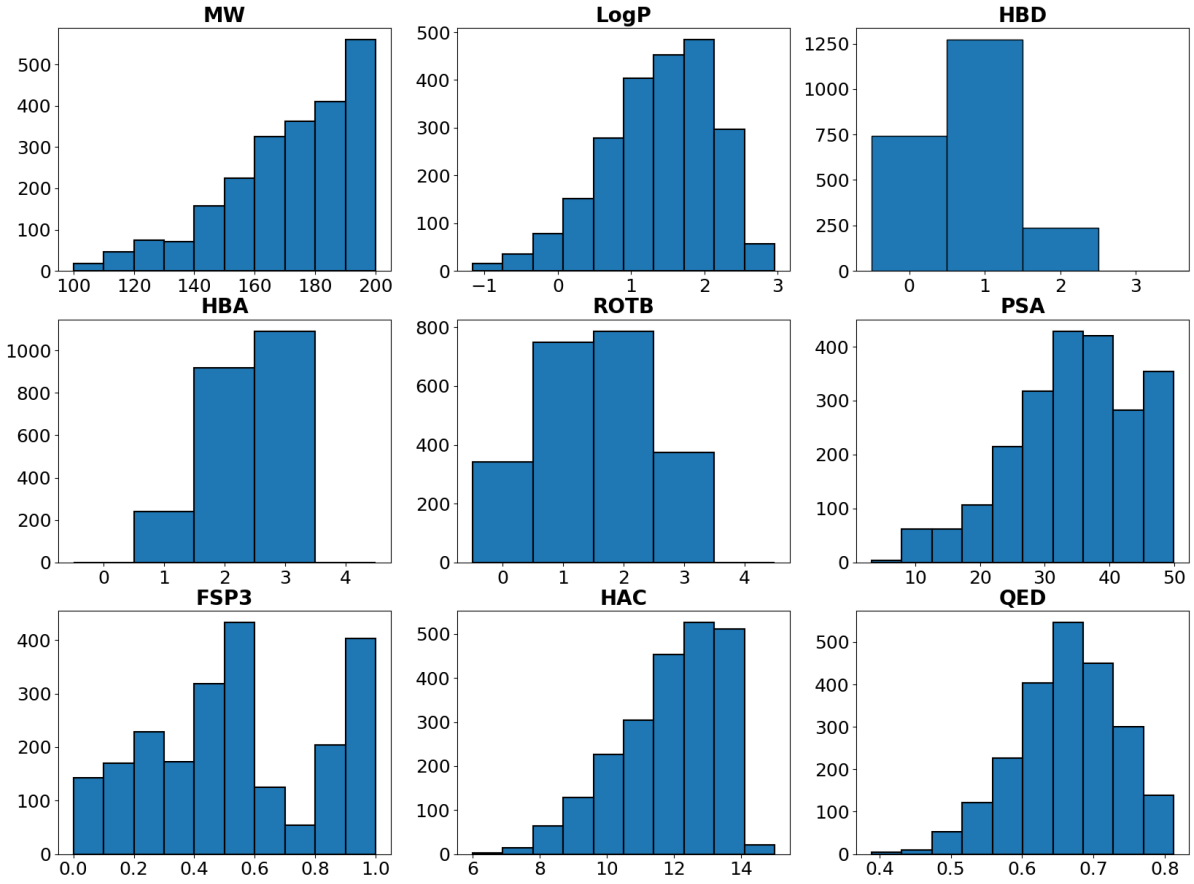


Figure 2.5: Distribution of molecular properties in the filtered fragment library (total = 2,252): MW (Da) (171.36 ± 21.82), LogP (1.38 ± 0.75), HBD (0.78 ± 0.62), HBA (2.38 ± 0.67), ROTB (1.53 ± 0.94), PSA (34.48 ± 9.29), FSP3 (0.54 ± 0.30), HAC (12.06 ± 1.71), QED (0.66 ± 0.07).

With the initial library, fragments were randomly chosen and assembled into linear molecules, which were subjected to intramolecular cyclization to produce macrocycles. To form a linear molecule, a defined number of fragments were selected, and their open binding sites, i.e. the ones with hydrogens attached, were identified. They were then connected by replacing the hydrogen on one fragment with a bond to an element on another fragment. The direction of the molecule elongation was fixed to prevent side additions. This process formed linear molecules with inputs of the initial fragment library, number of fragments, and number of molecules desired.

Next, we focused on determining the appropriate number of fragments to form the macrocycles. Considering that a macrocycle requires at least twelve atoms on its ring, molecules

with three fragments start to exhibit chances of macrocyclization. However, the upper bound was more arbitrary. While more fragments meant more variability, the resulting molecules would also be more complex and bulky, rendering them less drug-like. To visualize the differences, linear molecule libraries with three, four, and five fragments were generated, analyzed, and compared with each other. **Figure 2.6** shows examples from each group; **Figure 2.7** shows the distribution of their properties. The five-fragment ones generally have very large HBAs and HBDs, indicating they were simply too big. They also have a very low average QED (0.10 ± 0.03) with a narrow range (all < 0.2), indicating a low possibility to be therapeutically successful. Recognizing this, molecules with three and four fragments were passed on for any potential cyclization followed by a ring size check. In the end, macrocycle libraries with moderate sizes (three-fragment macrocycle = 15,160; four-fragment macrocycle = 34,616) were generated, each from 100 linear molecules. Representative molecules of successful macrocycles can be found in **Figure 2.8**.

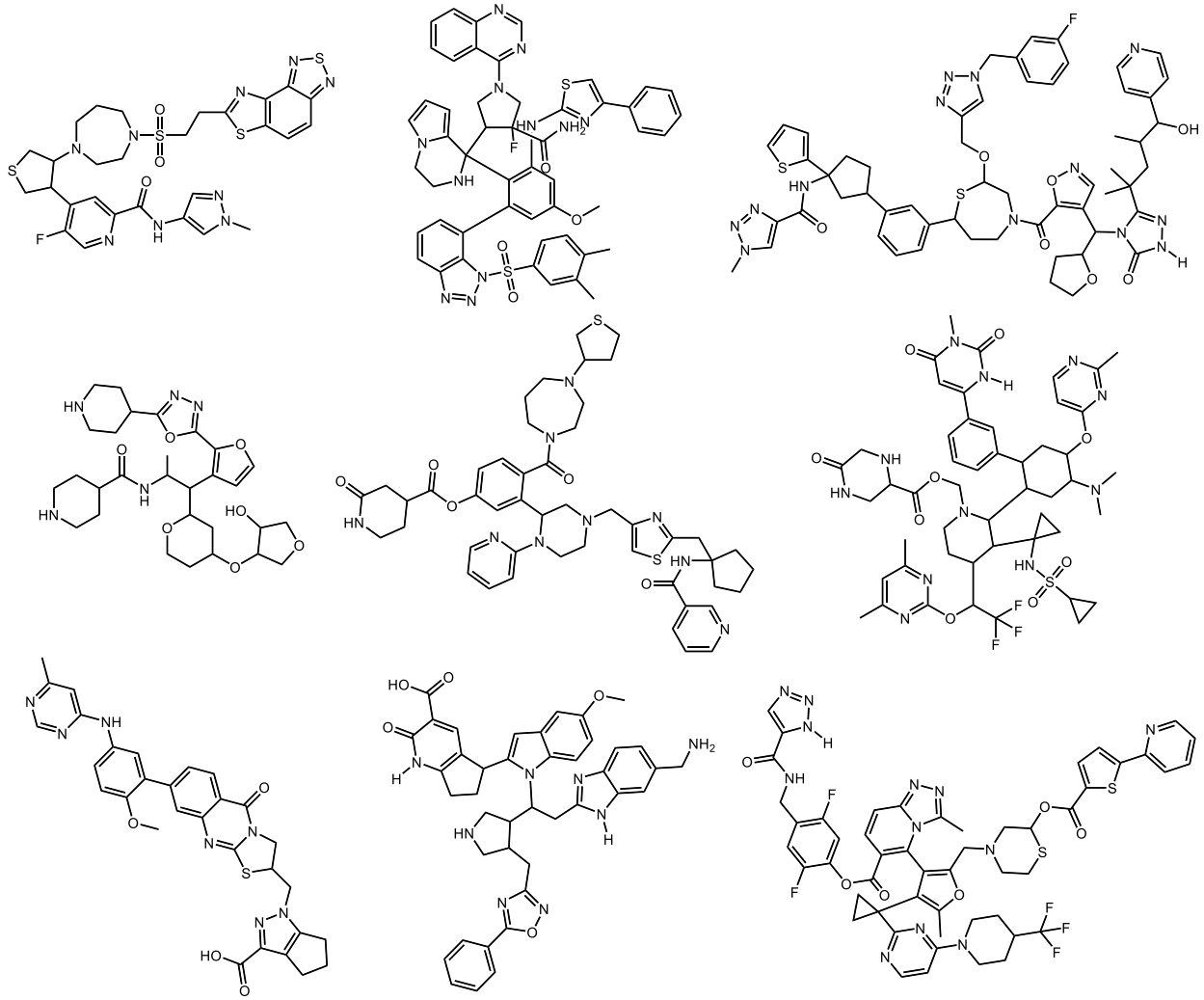


Figure 2.6: Representative linear molecules connected with three to five fragments, from left to right.

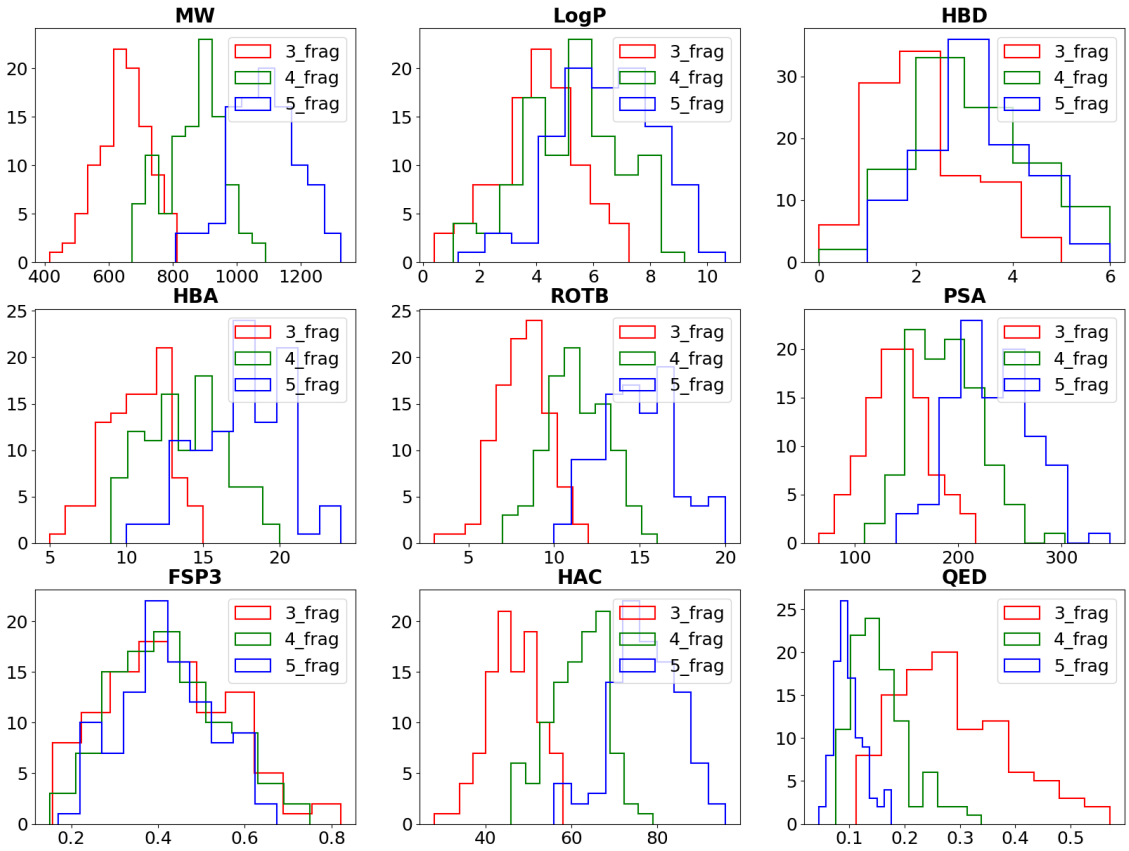


Figure 2.7: Distribution of properties of linear fragment-connected molecules (for each fragment number, total = 100).

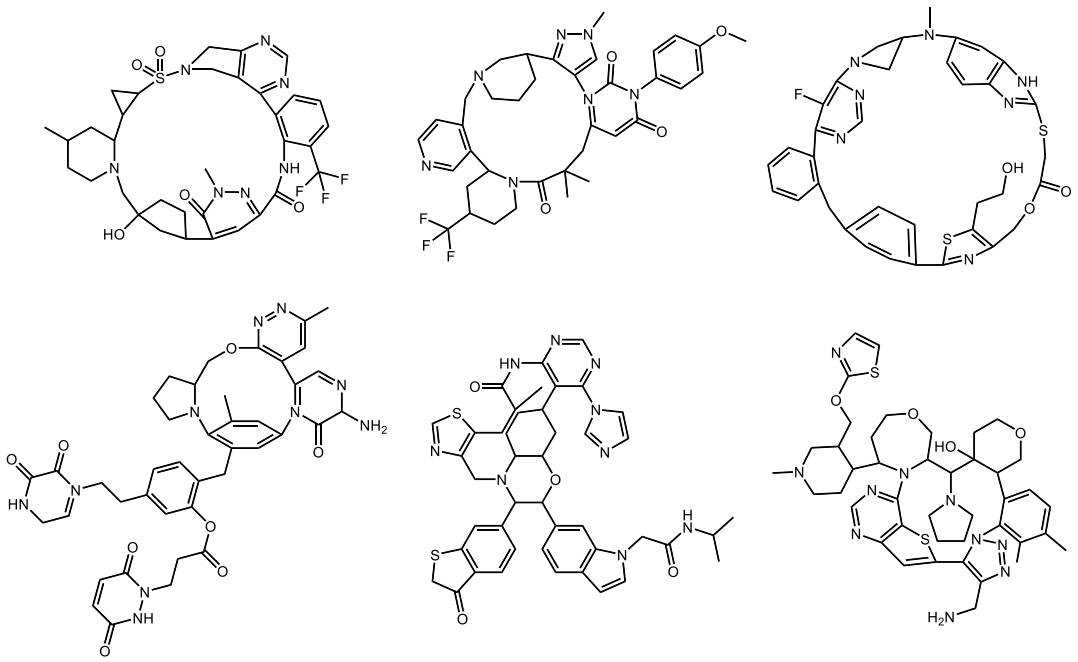


Figure 2.8: Representative macrocycles with three (top) and four fragments.

All aforementioned efforts only considered two-dimensional macrocyclic structures. Macrocycles might have very different properties when they adopted three-dimensional conformations, as mentioned in the first chapter. Therefore, using 2D structures to predict similarity and understand the interaction between the target and drug candidates was unreliable. Enumeration of all possible stereoisomers for each macrocycle could not only address this concern but also help increase the library sizes by almost 100-fold (three-fragment stereoisomers = 1,044,996; four-fragment isomers = 4,333,735). An example of the three-fragment stereoisomer is shown in **Figure 2.9**. Another round

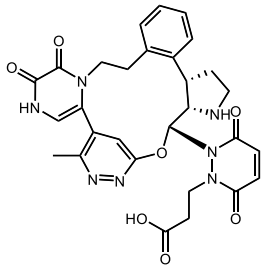


Figure 2.9: An example of final generated stereoisomers.

of parameter analysis returned similar property distributions for the four-fragment stereoisomers to those of the corresponding two-dimensional macrocycles. This was not the case for the three-fragment macrocycles. The QEDs of the three-fragment stereoisomers were more left-skewed (average of two-dimensional macrocycles = 0.20 ± 0.06; average of stereoisomers = 0.40 ± 0.09), reflecting their improved drug-likeness (**Figure 2.10**). These random libraries of stereoisomers will ultimately serve as candidates for the algorithm to select potential hits and suggest bioavailable non-peptide macrocycles.

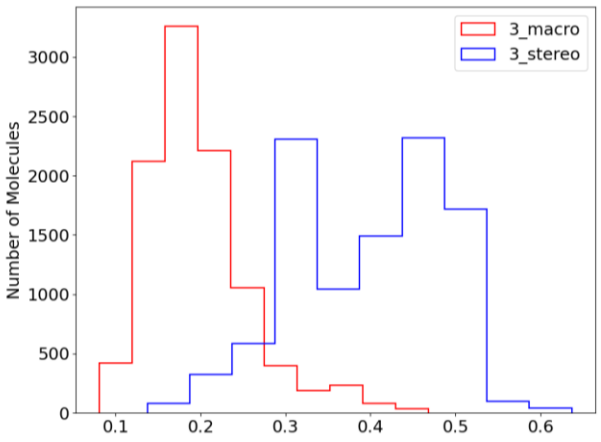


Figure 2.10: Distribution of QED of three-fragment 2D macrocycles and stereoisomers (for each group, total = 10,000).

2.3 Conclusion and Future Directions

A complete roadmap for the random generation of candidate libraries of stereospecific macrocycles has been conceived as follows: commercially available drug-like fragments are linked to form a set of linear molecules; intramolecular cyclization and enumeration of stereoisomers induce macrocycles with chiral centers. Although the current library sizes are only in the millions, this method leaves the chances of building an even larger library by raising the set number of linear molecules. In the future, descriptors such as Tanimoto index[13] should be utilized to confirm the structural diversity of the initial fragment library in addition to relying solely on the library size.

With an appropriate library in hand, the emphasis has turned to the development of the algorithms. As discussed, the neural network will use a known PPI as a reference to score the macrocycles and identify potential hits from the library. Some of the outstanding non-peptide macrocycles will be docked to the target and the computational data will be used as feedback for

the model to learn and improve its performance. After the iterative study, the model is expected to make reasonable judgments about the similarities. These answers, along with the template peptide, will then become the training set for the evolutionary algorithm, which will develop an evaluative system with minimal information on existing crystal structures or protein-peptide relationships. This algorithm needs to estimate the similarity of macrocycles and peptides by adjusting the metrics and the weights of each parameter continuously until it achieves similar results as the neural network model. After establishing the rating system, the algorithm can modify the macrocycle structure based on the scores to improve the similarity. It can even be taught to recognize and edit unfeasible connections within the molecules before producing the final answers. In this way, even without knowing the specific crystal structure and binding information, the algorithm can still recommend new bioactive macrocycles that are likely to be synthesizable from known effective peptides.

Although computationally demanding, this workflow will quickly retrieve a huge macrocycle library, find the molecule that may interact with a given target in relatively simple steps, then adjust the macrocycle based on scoring to provide the best product. With the aid of machine learning and virtual screening, it reduces the calculation time and cost spent on docking and avoids the complicated process of synthesizing and testing each unknown macrocycle in the laboratory, offering more potential in the field of drug design and development.

Chapter THREE

Synthetic Peptide Macrocyclization via 1,2-Carboamidation

In this chapter, we will discuss the accomplishment of an ongoing cobalt-catalyzed peptide macrocyclization project. The model system was an intermolecular coupling, and the condition was adapted onto a linear peptide. This project was conducted in collaboration with Dr. Christopher D. Poff, my graduate mentor in the Blakey lab.

3.1 Peptide Macrocyclization through C-H functionalization

As noted, macrocyclization has been a common method to improve the drug-like properties of peptides. Specifically, peptide coupling-based lactamization and disulfide bond formation dominated the field of peptide cyclization for almost thirty years.[15] In addition, because of the labile peptide bonds, most traditional reactions involving peptides need to be undertaken with protected residues. But with the development of solid-phase peptide synthesis (SPPS), lactamization can happen at specific deprotected residues, leading to chemoselective macrocyclization. However, with the advent of a new generation of peptide therapeutics, the inaptness of these methods to diversity generation renders them less favorable. In contrast, metal-catalyzed processes based on C-H functionalization are gaining more popularity because of their variations in ring-forming linkages including arylation, olefination, alkynylation, and alkylation of peptide side chains and pre-functionalized termini.

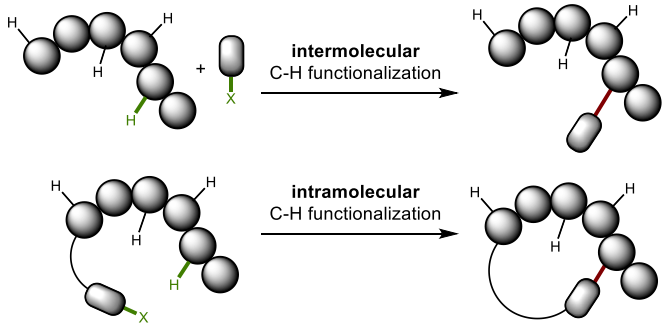
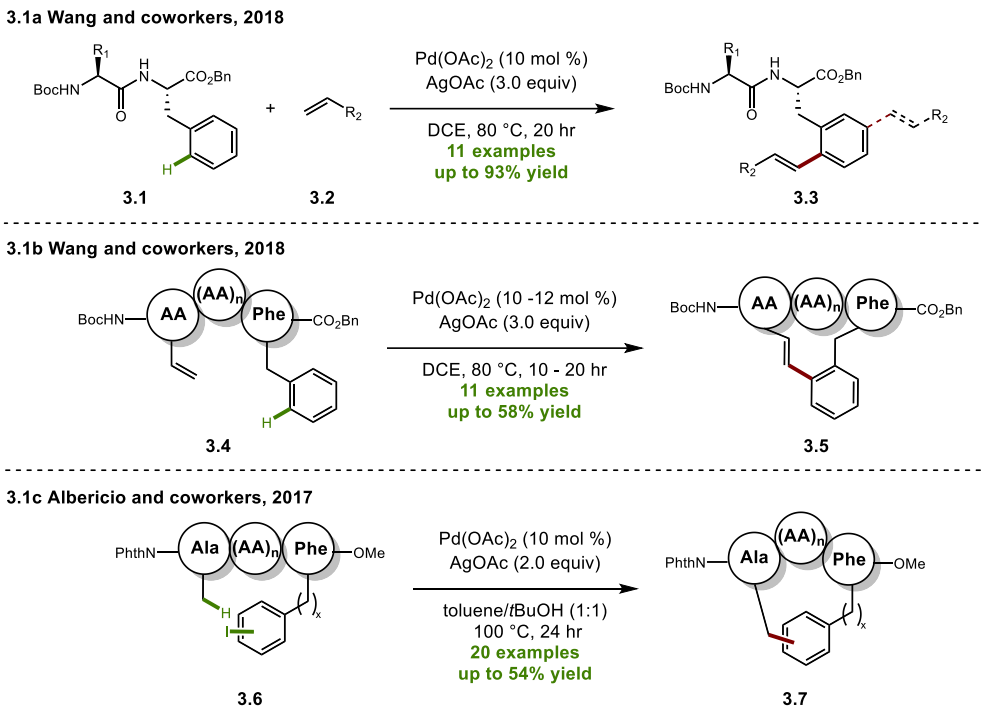


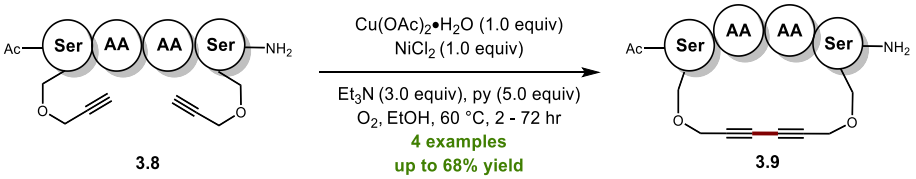
Figure 3.1: Peptide macrocyclization via C-H functionalization. (Figured created from Rivera, *et al.*)[15]



Scheme 3.1: Palladium-catalyzed peptide macrocyclization strategies.

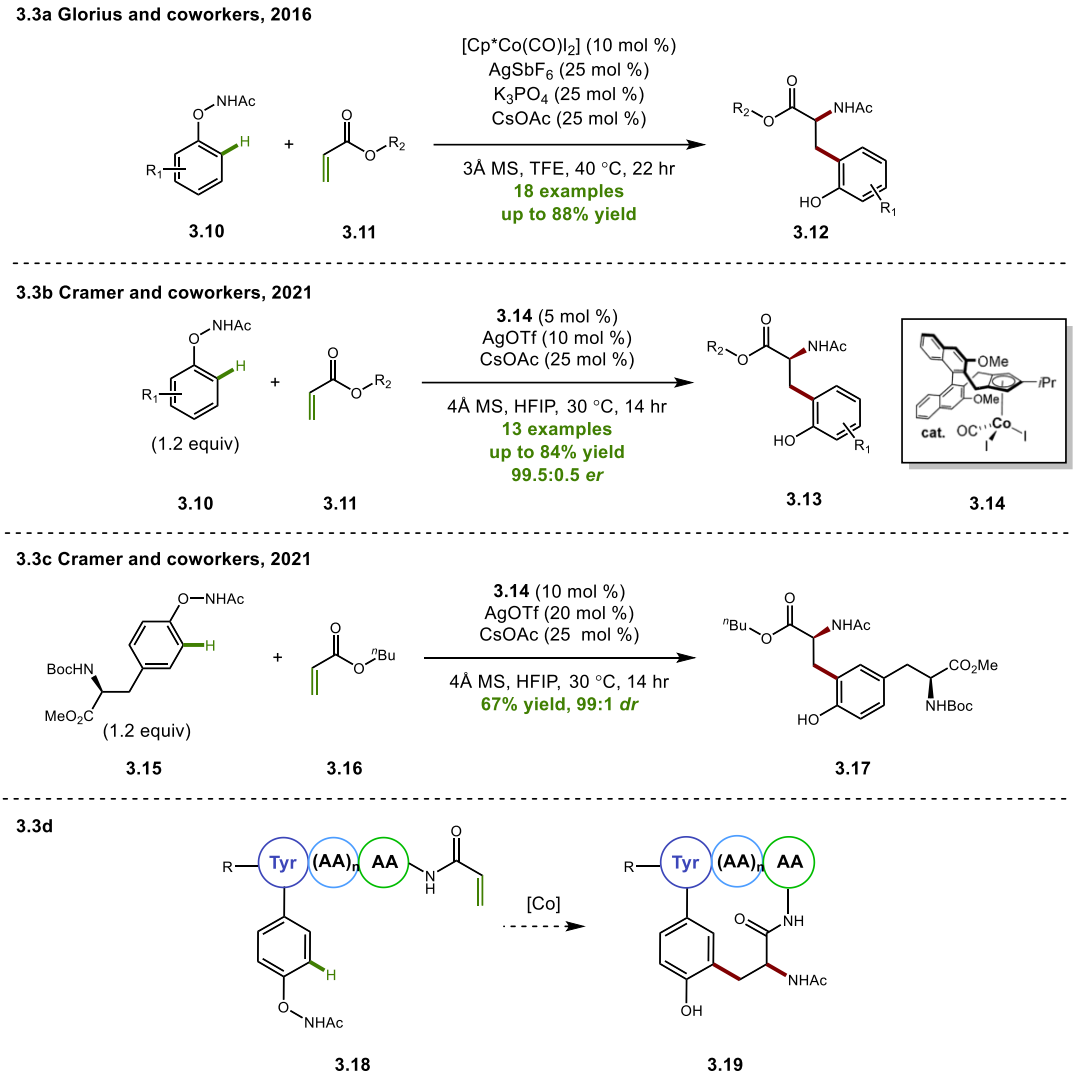
Direct functionalization of the C-H bond, especially via transition metal catalysis, has evolved to be a powerful tool in organic synthesis as it can access the typically unreactive C-H bonds selectively, reducing the synthetic length and avoiding large amounts of undesired by-products.[16] When this is applied to complex peptides with multiple C-H bonds, a site-selective intramolecular coupling can result in macrocyclization (**Figure 3.1**). In this field, the dominant and most successful strategies have been involved with palladium catalysts. For example, Wang and coworkers reported a palladium-catalyzed C(sp^2)-H olefination of phenylalanine in dipeptide substrates **3.1** (**Scheme 3.1a**).[17] They then applied the generation of the aryl-alkene crosslinks to a linear peptide **3.4** with one residue end bearing a terminal olefin and the other end being phenylalanine (**Scheme 3.1b**). Palladium-catalyzed C(sp^3)-H functionalization methods have also been investigated heavily. In 2017, a cyclization method was developed by Albericio and coworkers through activating the primary β-C(sp^3)-H of N-terminal alanine and then coupling it with the terminal iodo-phenylalanine intramolecularly to produce the macrocycle **3.7** (**Scheme 3.1c**).[18]

In fact, while palladium is utilized in the majority of transition-metal catalyzed peptide macrocyclizations, it has also been exploited in the entire field of C-H functionalization, along with several other precious transition metals including iridium, rhodium, and ruthenium.[16] Unfortunately, these metals are both expensive and generally more toxic. Therefore, environmentally friendly C-H activation is usually achieved through earth-abundant and less toxic first-row transition metals such as manganese, iron, cobalt, and copper. Copper catalysis is the most studied in this group for peptide macrocyclization; one example is shown in **Scheme 3.2**.[19] However, this method, along with many other precedents, requires a stoichiometric amount of both copper and nickel, which is far less attractive than actual copper catalysis. And very few examples reported incorporated the other earth-abundant transition metals.



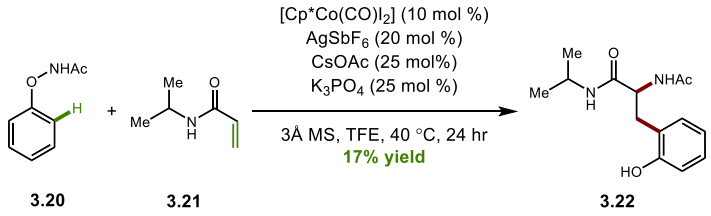
Scheme 3.2: Copper-catalyzed peptide macrocyclization through diyne coupling.

Aiming for an alternative strategy, we were inspired by a cobalt-catalyzed 1,2-carboamidation of acrylates **3.11** reported by the Glorius group (**Scheme 3.3a**).[20] Later, the Cramer group achieved an enantioselective version of this work with a different cobalt catalyst **3.14** (**Scheme 3.3b**).[21] In addition, they were able to incorporate a tyrosine-derived substrate **3.15** as the coupling partner (**Scheme 3.3c**). Preliminary success on the two amino ester moieties laid the foundation for our proposed macrocyclization method. Specifically, we aimed to build a linear peptide **3.18** with a tyrosine hydroxamate on the C-terminus. The N-terminus would be an acrylamide, an analog of the acrylate. In this case, a side chain-end peptide cyclization would proceed through an intramolecular coupling.



Scheme 3.3: Cobalt-catalyzed intermolecular 1,2-carboamidation and proposed intramolecular peptide macrocyclization.

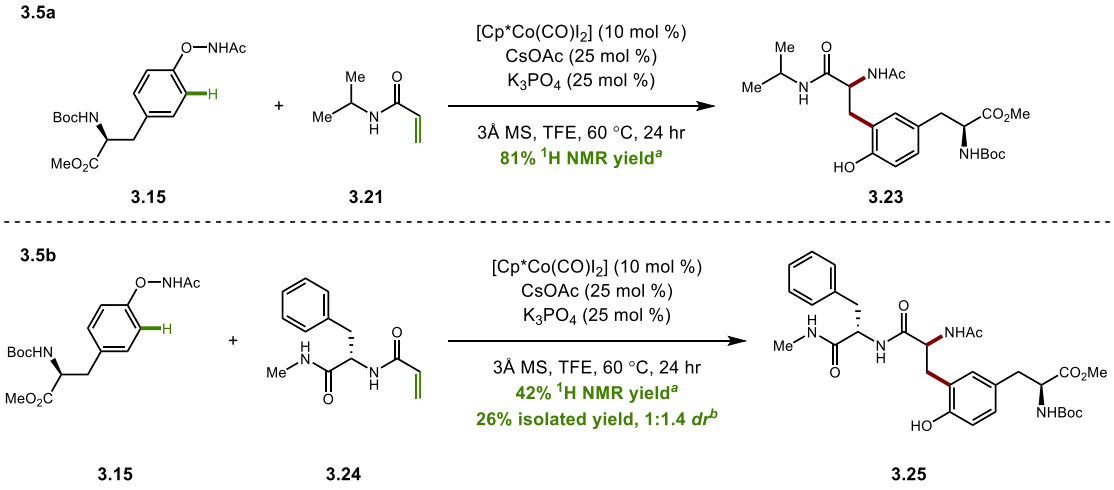
3.2 Results and Discussion



Scheme 3.4: Test reaction of 1,2-carboamidation of acrylamides.

Since no previous cobalt-catalyzed coupling work exists for acrylamides, my graduate mentor Chris conducted an initial test reaction on phenoxyacetamide **3.20**

and an alkyl acrylamide **3.21** under conditions the Glorius group reported.[20] Under these conditions, he obtained the desired product in a 17% yield (**Scheme 3.4**). Albeit a low yield, the result demonstrated the feasibility of our idea. While we started to improve the yield, we found that tyrosine hydroxamate **3.15** behaved differently from the simple phenoxyacetamide **3.20**. This prompted Chris to optimize the intermolecular coupling using **3.15** and **3.21**. By removing the silver and increasing the temperature to 60 °C, Chris obtained the desired product in an 81% yield (**Scheme 3.5a**). With the optimized reaction conditions, we aimed to apply this model system to a more complex peptide-like substrate prior to the attempt of macrocyclization (**Scheme 3.5b**). Therefore, we prepared a phenylalanine diamide **3.24** from the commercially available phenylalanine. The same reaction condition afforded **3.25** as a ~1:1 mixture of diastereomers.

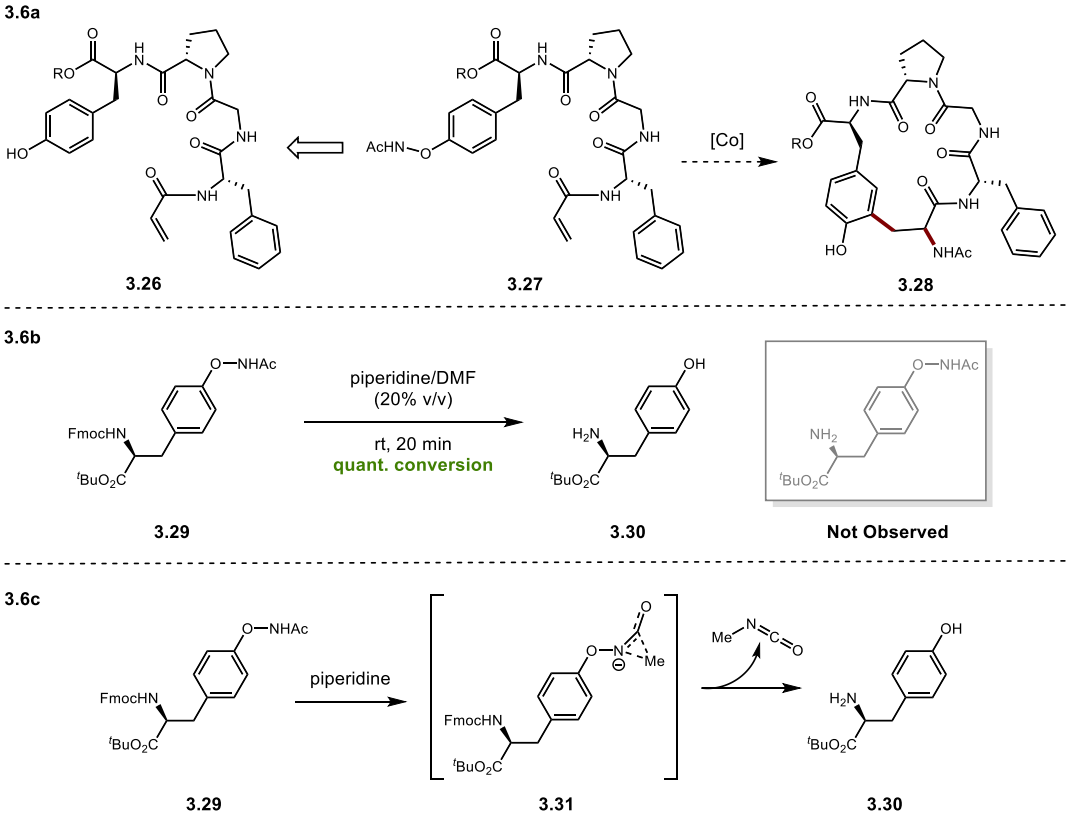


[a]Yield determined by ^{1}H NMR integration against 3,4-dinitrobenzoate and represents an average of two results.
[b]Isolated yield after two purifications.

Scheme 3.5: Optimized condition for intermolecular carboamidation of acrylamides and tyrosine substrate and its application in a dipeptide system.

While Chris focused on the intermolecular coupling, I worked on the synthesis of the linear peptide for macrocyclization. In addition to phenylalanine and the tyrosine substrate on each end, we also hoped to include a glycine-proline pair as the linker amino acids, as this would encourage the backbone turn and allow for closer proximity of the two ends. This notion led us to devise

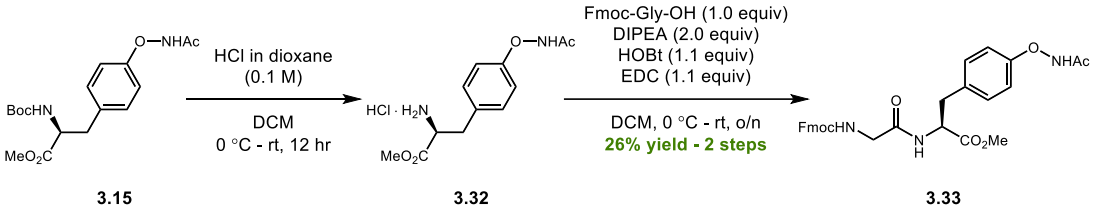
functionalized linear peptide **3.27**, derived from phenylalanine-glycine-proline-tyrosine **3.26** (**Scheme 3.6a**). Cyclization of **3.27** would also result in an 18-membered macrocycle **3.28**, one of the ring sizes that occur most commonly in macrocyclic natural products.[22]



Scheme 3.6: Tyrosine hydroxamate reactivity leads to unsuccessful synthesis of the original linear peptide target on solid-phase peptide synthesizer.

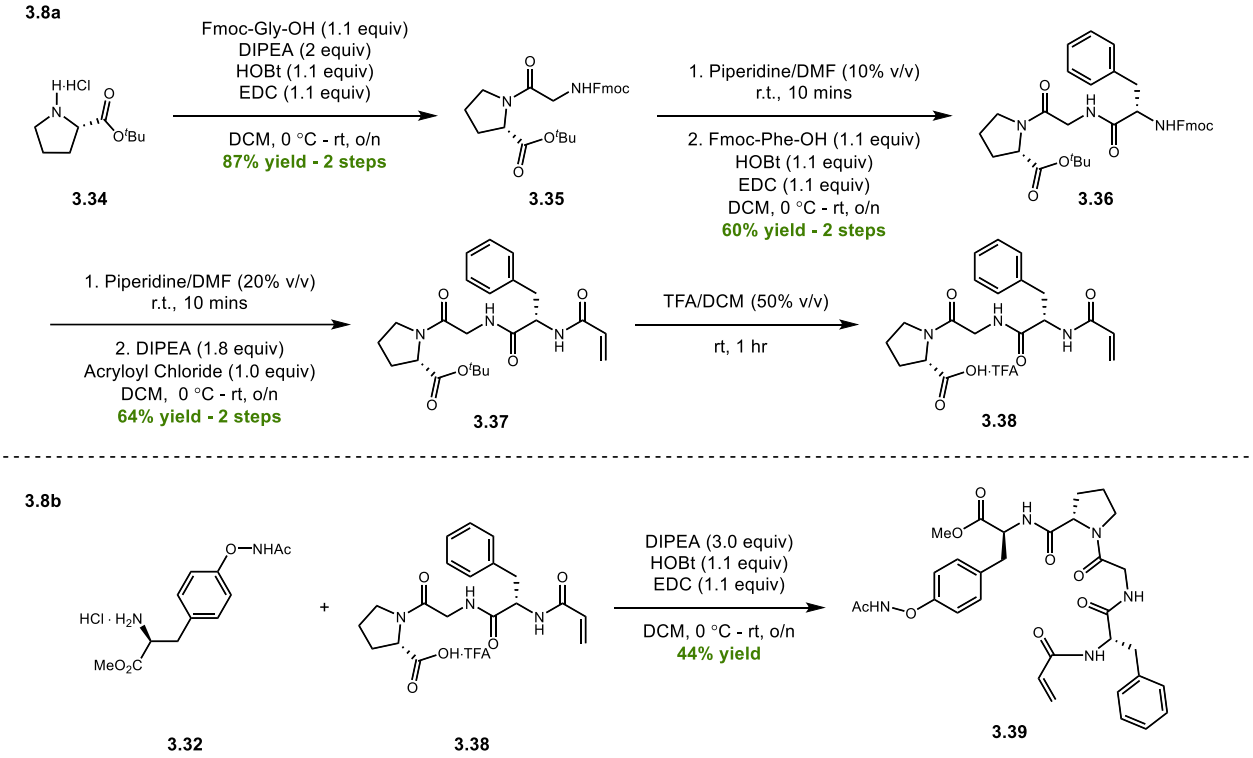
To synthesize the peptide **3.27**, we originally conceived using an automatic solid-phase peptide synthesizer. We planned to load a tyrosine hydroxamate onto a resin, build the sequence, cap the N-terminus as acrylamide, and cleave the peptide from the resin. Until this point, we had performed all intermolecular couplings with Boc-protected tyrosine hydroxamate **3.15**. However, the deprotection condition on the synthesizer is designed for Fmoc-protected amino acid, which necessitates the use of Fmoc-protected tyrosine hydroxamate **3.29**. To confirm its tolerance of SPPS, I subjected **3.29** to the Fmoc deprotection conditions (20% piperidine/DMF). Unfortunately, I only recovered tyrosine **3.30** (**Scheme 3.6b**), the product from an N-O bond cleavage along with Fmoc deprotection. After searching in literature, I proposed that the

hydroxamate group was likely undergoing a Lossen Rearrangement upon treatment of base. The reaction then gave rise to isocyanate and tyrosine **3.30** (**Scheme 3.6c**), which has also been reported to convert hydroxamate esters to alcohols.[23] Variations on temperature and/or reaction time were not able to prevent the rearrangement, which indicated a need to revise the synthetic route.



Scheme 3.7: Attempt to use Boc-protected tyrosine hydroxamate to avoid Lossen Rearrangement.

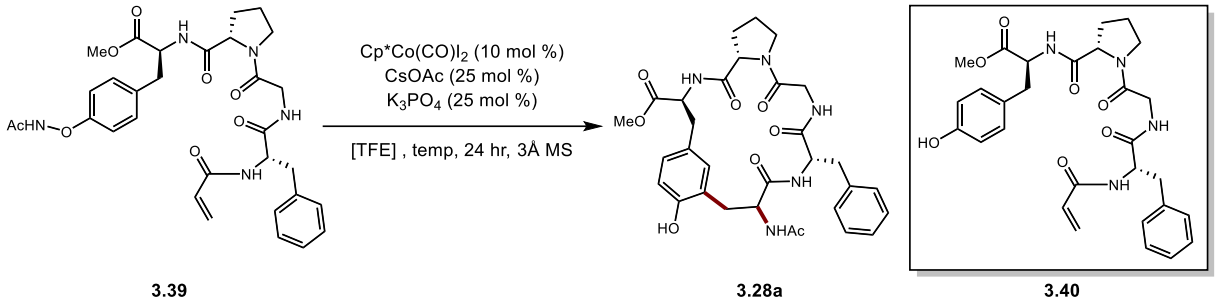
Understanding that the Lossen Rearrangement could be triggered under basic conditions, we reverted to **3.15** as our coupling partner**,** as Boc, unlike Fmoc, is removed under acid treatment. A preliminary attempt indicated the tolerance of hydroxamate in hydrochloric acid. Subsequent coupling of Fmoc-glycine-OH gave the dipeptide **3.33** with minimal sign of rearrangement under mild base condition (**Scheme 3.7**). Therefore, we believed that the final peptide could be generated with a last-step installment of the deprotected tyrosine hydroxamate **3.32** to the tripeptide (phenylalanine-glycine-proline). The synthetic route toward this tripeptide is summarized in **Scheme 3.8**. We chose proline *tert*-butyl ester **3.34** to start with for the convenient C-terminus deprotection conditions. The dimer **3.35** and trimer **3.36** were built via standard solution phase peptide synthesis.[24] The N-terminus was then acrylated to produce the corresponding acrylamide **3.37**, followed by the C-terminus deprotection. Finally, we generated the full peptide target **3.39** in a 44% yield by coupling **3.32** and **3.38**.



Scheme 3.8: Solution phase synthesis of the functionalized peptide.

With the complete linear peptide synthesized, I then applied the optimized condition of the intermolecular carboamidation to the peptide system (**Table 3.1**). **Entry 1** represents the standard conditions, which only yielded the Lossen Rearrangement product. This was expected due to the presence of potassium tribasic and cesium acetate, both being sources of base. Since intramolecular macrocyclization needs to compete with intermolecular polymerization, concentration could also affect the reactivity. However, in this case, dilution did not help push macrocyclization (**Entry 2**). Further increase in heat to the temperatures at and above the boiling point of the solvent also gave no desired product (**Entry 3 and 4**).

Table 3.1: Reaction screen on linear peptide macrocyclization

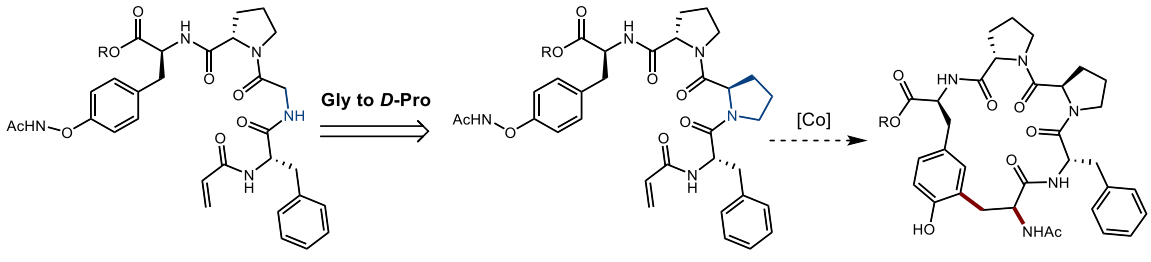


3.39 3.28a 3.40

Entry	[TFE] (M)	Temperature (°C)	Recovered **3.39** (%)	**3.40** (%)	**3.28a** (%)
1	0.1	60	46	21	0
2	0.05	60	76	24	0
3	0.05	75	74	26	0
4	0.05	90	72	27	0

Yield determined by ^{1}H NMR integration against 3,4-dinitrobenzoate.

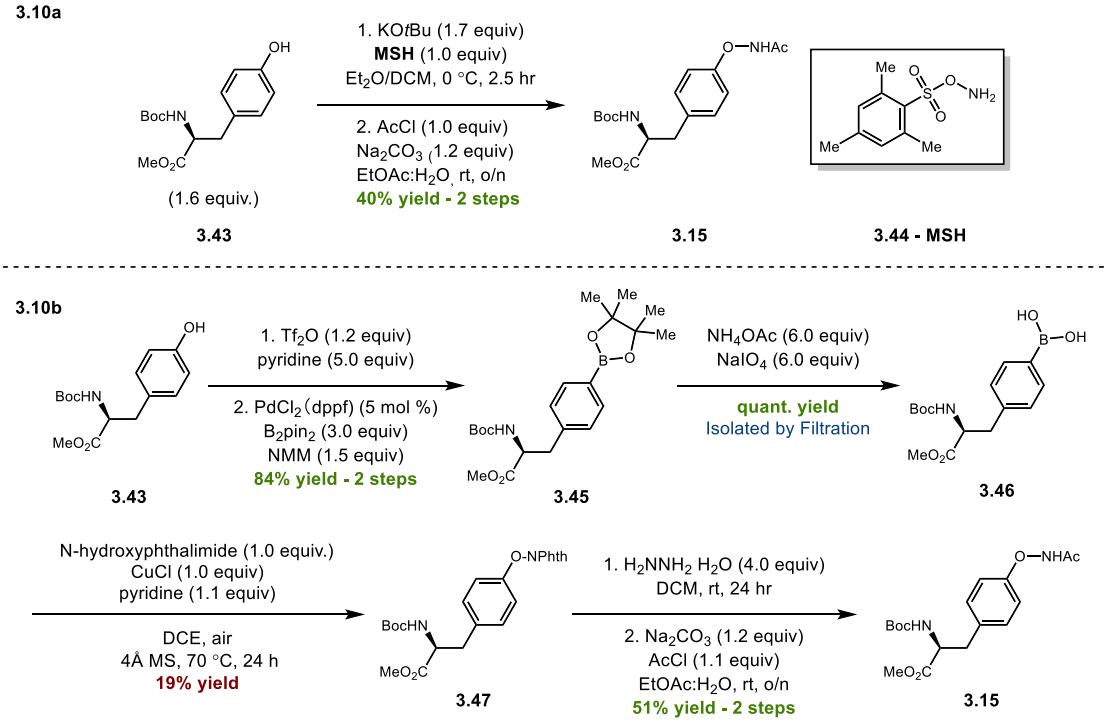
In the previously discussed work of the copper-catalyzed peptide macrocyclization, Verlinden *et al.* found that certain linear peptides were readily cyclized while others were not.[19] Considering the hard work accomplished by Chris on the optimization of the intermolecular system, we concluded that a peptide redesign should be the most reasonable solution. According to Verlinden *et al.*, a *D*-proline-*L*-proline motif can induce β-turns as they introduce a sharp bend in a polypeptide chain.[19] The replacement should reduce the flexibility of glycine and promote cyclization, while still producing the desired 18-membered macrocycle. The new peptide target **3.41** is shown in **Scheme 3.9**.



3.27 3.41 3.42

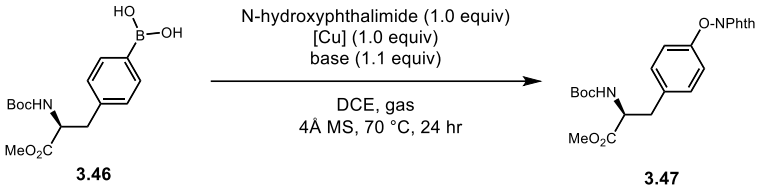
Scheme 3.9: Redesign of the linear peptide to incorporate the *D*-proline-*L*-proline linker.

Up to this point, all the tyrosine hydroxamates **3.15** and **3.29** were prepared following literature precedent,[25] which requires an amination by *O*-mesitylsulfonylhydroxylamine (MSH) **3.44** and subsequent acetylation (**Scheme 3.10a**). However, **3.44** is a known explosion hazard.[26] Since no other methods have been disclosed, Chris took the lead to develop a safer and scalable route (**Scheme 3.10b**) to access **3.15**. As shown, the current limitation of this new route is the low yield of the Chan-Evans-Lam coupling, which converts the boronic acid **3.46** to phenoxyacetamide **3.47**. Currently, I am focusing on the optimization of the Chan-Evans-Lam reaction (**Table 3.2**) while Michael Hollerbach, a graduate student in the Blakey lab, is confirming the scalability of the previous steps.



Scheme 3.10: Published method and newly designed route to access tyrosine hydroxamate.

Table 3.2: Optimization on Chan-Evans-Lam coupling



Entry	[Cu]	Gas	Base	**3.47** (%)
1	CuCl	air	pyridine	17
2	CuCl	O_2	pyridine	-
3	$Cu(OAc)_2$	O_2	pyridine	-
4	CuCl	O_2	Et_3N	-

I first followed the same reaction condition that Chris used, which was the Glorius condition to make phenoxyphthalimide from phenol (**Entry 1**).[20] Getting a similar yield, Michael and I have worked together to propose an optimization plan. First, since oxygen is required for copper oxidation in the catalytic cycle, we envision that the replacement of air with oxygen will help promote the reaction (**Entry 2**). If this gives a better result, we will then screen different copper sources, for example, copper acetate, one of the most used catalysts for Chan-Evans-Lam reactions **(Entry 3)**, and different bases such as triethylamine **(Entry 4)** to see if any would influence the reaction outcome. We expect to finish the optimization in April.

3.3 Conclusion and Future Directions

Herein, we have proposed a peptide macrocyclization strategy through cobalt-catalyzed 1,2-carboamidation of acrylamides. We have achieved the optimized condition from the intermolecular model study between tyrosine hydroxamate and alkyl acrylamides and confirmed its applicability to a dipeptide system. While synthesizing the target linear peptide substrate for macrocyclization, we discovered the incompatibility of the required tyrosine hydroxamate on the solid-phase peptide synthesizer due to the happening of Lossen Rearrangement. Fortunately, this problem was solved through solution-phase coupling. However, although we have prepared the linear peptide, the preliminary attempt on macrocyclization was not successful. Recognizing the

importance of the peptide identity in cyclization, we have designed a new peptide to incorporate a *D*-proline-*L*-proline pair to force the peptide turn. Similar to the first linear peptide, the new peptide will be generated through a last-step installment of the pre-functionalized tyrosine onto the rest of the sequence (phenylalanine-*D*-proline-*L*-proline). We also hope to develop an efficient and scalable solid-phase synthesis method for the tripeptide this time considering the time that could be spent on solution-phase synthesis.

In addition, with Chris's success in designing a new and safer method to access the tyrosine hydroxamate **3.15**, I am currently focusing on the optimization of the Chan-Evans-Lam reaction to achieve the greatest utility of the route. This will involve screening for the optimal oxygen source, catalyst, and base. We will also test the direct coupling of **3.46** with acetohydroxamic acid (AcNHOH) instead of N-hydroxyphthalimide. If this reaction were to proceed with a relatively good yield, only one step would be needed to convert **3.46** to **3.15** in replacement of the original three-step sequence.

The new peptide macrocyclization strategy could fill in the current limits of first-row earth-abundant transition metal catalysis in peptide cyclization. It also has the potential to control stereochemistry by switching to bulky ligands as the Cramer group did.[21] This selectivity will push the boundaries of the already existing chemistry to previously underexplored stereoselective macrocyclization.

Chapter FOUR

Conclusion: Integrating the Computational and Synthetic Tools

Centering around macrocycles, a promising class of drug candidates, we have reported on the progress of two projects, a machine learning platform designed to computationally generate and select non-peptide macrocyclic drugs and a strategy for the construction of peptide macrocycles via cobalt-catalyzed 1,2-carboamidation. These approaches can provide new insights in computational drug design and first-row transition metal-catalyzed peptide cyclization, respectively. Moreover, the combination of these two methods can bridge the gap between theoretical and synthetic work on macrocycles. For example, as explained in the second chapter, certain macrocycles suggested by machine learning models should be synthesized to confirm the reliability of the models. Although carboamidation is employed as a side chain-end cyclization in peptide systems in the second project, it can certainly assist in the synthesis of the proposed non-peptidic macrocycles or their analogs, as it can link aryl groups to olefin structures for intermolecular coupling or intramolecular cyclization. On the other hand, the new cyclization methodology could diversify the toolbox for generating bioactive natural products or cyclic peptides. The structures can then be resolved computationally and will provide machine learning programs with more data to learn and improve. These two ideas can be tied together to institute collaboration between theoretical and synthetic chemists.

The potential of macrocyclic compounds to modulate PPIs, as well as several pharmacological advantages, make them attractive therapeutic agents. More efficient methods for identifying macrocyclic drug hits and environmentally friendly synthetic methods could foster the drug design and development process. With these goals in mind, we also hope to see more interdisciplinary research involving both divisions. As discussed, both computational and experimental data can provide support for the other part, which will greatly contribute to our understanding of macrocyclic compounds and their application in pharmaceutics.

References

(1) Arkin, M. R.; Wells, J. A. Small-Molecule Inhibitors of Protein–Protein Interactions: Progressing towards the Dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.

(2) Sengupta, S.; Mehta, G. Macrocyclization via C–H Functionalization: A New Paradigm in Macrocycle Synthesis. *Org. Biomol. Chem.* **2020**, *18*, 1851–1876.

(3) Zorzi A, Deyle K, Heinis C. Cyclic peptide therapeutics: past, present and future. *Curr Opin. Chem. Biol.* **2017**, *38*, 24-29.

(4) Yu, X.; Sun, D. Macrocyclic Drugs and Synthetic Methodologies toward Macrocycles. *Molecules* **2013**, *18*, 6230–6268.

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

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

(7) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4*, 90–98.

(8) Matsson, P.; Doak, B. C.; Over, B.; Kihlberg, J. Cell Permeability beyond the Rule of 5. *Adv. Drug Deliv. Rev.* **2016**, *101*, 42–61.

(9) Konze, K. D.; Bos, P. H.; Dahlgren, M. K.; Leswing, K.; Tubert-Brohman, I.; Bortolato, A.; Robbason, B.; Abel, R.; Bhat, S. Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors. *J. Chem. Inf. Model.* **2019**, *59*, 3782–3793.

(10) Saha, I.; Dang, E. K.; Svatunek, D.; Houk, K. N.; Harran, P. G. Computational Generation of an Annotated Gigalibrary of Synthesizable, Composite Peptidic Macrocycles. *Proc. Natl. Acad. Sci.* **2020**, *117*, 24679–24690.

(11) Malde, A. K.; Hill, T. A.; Iyer, A.; Fairlie, D. P. Crystal Structures of Protein-Bound Cyclic Peptides. *Chem. Rev.* **2019**, *119*, 9861–9914.

(12) Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24*, 4309.

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

(14) Mahjour, B.; Shen, Y.; Liu, W.; Cernak, T. A Map of the Amine–Carboxylic Acid Coupling System. *Nature* **2020**, *580*, 71–75.

(15) Rivera, D. G.; Ojeda-Carralero, G. M.; Reguera, L.; Van der Eycken, E. V. Peptide Macrocyclization by Transition Metal Catalysis. *Chem. Soc. Rev.* **2020**, *49*, 2039–2059.

(16) Gandeepan, P.; Müller, T.; Zell, D.; Cera, G.; Warratz, S.; Ackermann, L. 3d Transition Metals for C–H Activation. *Chem. Rev.* **2018**, *119*, 2192–2452.

(17) Bai, Z.; Cai, C.; Yu, Z.; Wang, H. Backbone-Enabled Directional Peptide Macrocyclization through Late-Stage Palladium-Catalyzed Δ-C (Sp2) –H Olefination. *Angew. Chem.* **2018**, *130*, 14108–14112.

(18) Noisier, A. F.; García, J.; Ionuţ, I. A.; Albericio, F. Stapled Peptides by Late-Stage C (Sp3) – H Activation. *Angew. Chem.* **2017**, *129*, 320–324.

(19) Verlinden, S.; Geudens, N.; Martins, J.; Tourwé, D.; Ballet, S.; Verniest, G. Oxidative α, ω-Diyne Coupling as an Approach towards Novel Peptidic Macrocycles. *Org. Biomol. Chem.* **2015**, *13*, 9398–9404.

(20) Lerchen, A.; Knecht, T.; Daniliuc, C. G.; Glorius, F. Unnatural Amino Acid Synthesis Enabled by the Regioselective Cobalt(III)-Catalyzed Intermolecular Carboamination of Alkenes. *Angew. Chem. Int. Ed.* **2016**, *55*, 15166–15170.

(21) Ozols, K.; Onodera, S.; Woźniak, Ł.; Cramer, N. Cobalt (III)-Catalyzed Enantioselective Intermolecular Carboamination by C– H Functionalization. *Angew. Chem. Int. Ed.* **2021**, *60*, 655–659.

(22) Yudin, A.K. Macrocycles: lessons from the distant past, recent development, and future directions. *Chem. Sci.* **2015**, *6*, 30-49.

(23) Fier, P. S.; Maloney, K. M. Direct Conversion of Haloarenes to Phenols under Mild, Transition-Metal-Free Conditions. *Org. Lett.* **2016**, *18*, 2244–2247.

(24) Ghosh, K. C.; Duttagupta, I.; Bose, C.; Banerjee, P.; Gayen, A. K.; Sinha, S. Synthesis and Anticancer Activities of Proline-Containing Cyclic Peptides and Their Linear Analogs and Congeners. *Syn. Comm.* **2019**, *49*, 221–236.

(25) Li, B.; Lan, J.; Wu, D.; You, J. Rhodium (III)-Catalyzed Ortho-Heteroarylation of Phenols through Internal Oxidative C – H Activation: Rapid Screening of Single-Molecular White-Light-Emitting Materials. *Angew. Chem.* **2015**, *127*, 14214–14218.

(26) Bernardes, G. J.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.

Supplemental Information

1. General Information

All reactions were carried out under nitrogen atmosphere with anhydrous solvents in oven- or flame-dried glassware using standard Schlenk technique, unless otherwise stated. Anhydrous dichloromethane (DCM) and diethyl ether (Et_2O) were obtained by passage through activated alumina using a Glass Contours solvent purification system.2,2,2-trifluoroethanol (TFE) was distilled over calcium hydride (CaH_2) and stored over activated molecular sieves (MS). Solvents for workup, extraction, and column chromatography were used as received from commercial suppliers without further purification. Unless otherwise reported, compounds *L*-Phe-NHMe,[1] **3.15**,[2] **3.20**,[3] [Cp*Co(CO)I_2],[4] and **3.44**[5] were synthesized according to previously reported procedures. All other chemicals were purchased from Millipore Sigma, Strem Chemicals, Oakwood Chemicals, Alfa Aesar, or Combi-Blocks and used as received without further purification, unless otherwise stated.

^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectra were recorded on a Varian Inova 600 spectrometer (600 MHz ^{1}H, 151 MHz ^{13}C), a Bruker 600 spectrometer (600 MHz ^{1}H, 151 MHz ^{13}C), a Varian Inova 500 spectrometer (500 MHz ^{1}H, 126 MHz ^{13}C), and a Bruker400 spectrometer (400 MHz ^{1}H, 126 MHz ^{13}C) at room temperature in $CDCl_3$ (dried over activated molecular sieves) with internal $CHCl_3$ as the reference (7.26 ppm for ^{1}H, 77.16 ppm for ^{13}C), unless otherwise stated. Chemical shifts (δ values) were reported in parts per million (ppm) and coupling constants (J values) in Hz. Multiplicity was indicated using the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, qn = quintet, m = multiplet, br = broad. High resolution mass spectra (HRMS) were obtained using a Thermo Electron Corporation Finigan LTQFTMS (at the Mass Spectrometry Facility, Emory University). High Pressure Liquid Chromatography (HPLC) was performed on an Agilent 1100 series HPLC utilizing CHIRALPAK AD-H, AS-H, OD-H and OJ-H 4.6 x 150 mm analytical columns.

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, ethanolic p-anisaldehyde, ethanolic bromocresol green, or aqueous potassium permanganate ($KMnO_4$). Flash column chromatography was performed using Silicycle SiliaFlash® F60 silica gel (40-63 μm) on a Biotage Isolera One system. Preparatory TLC was performed on precoated glass-backed Silicycle SiliaPure® 1.0 mm silica gel 60 plates. We acknowledge the use of shared instrumentation provided by grants from the NIH and the NSF.

2. Construction of Libraries of Macrocycles

General

All coding, data analysis and data visualization were done in Python 3.8 with packages pandas, numpy, matplotlib, statistics, os, random, and rdkit. 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.

Initial Fragment Library

The Life Chemicals general fragment library[6] was downloaded and stored as SMILES codes. All properties were calculated using rdkit.Chem.rdMolDescriptors and rdkit.Chem.QED. Distribution of properties are summarized in **Figure S.1**. The personalized initial fragment library was filtered according to the following criteria: MW ≤ 200 Da, logP ≤ 3, HBD ≤ 3, HBA ≤ 3, ROTB ≤ 3, PSA ≤ 50, and HAC ≤ 20.

Linear Molecule

The fragment library was imported, and a given number of fragments were randomly selected to push onto a stack. The fragment stack was then defined as an input for the random connections, and the last two elements were popped to identify any atoms with hydrogens attached. The atoms were connected with any possible reactions using

rdkit.Chem.rdchemreactions. The resulting linear molecules were pushed back onto the stack as the last element, so the process could repeat with another fragment added to the linear molecule until reaching the given number of fragments for a single linear molecule. A list of random molecules was generated and stored for macrocyclization.

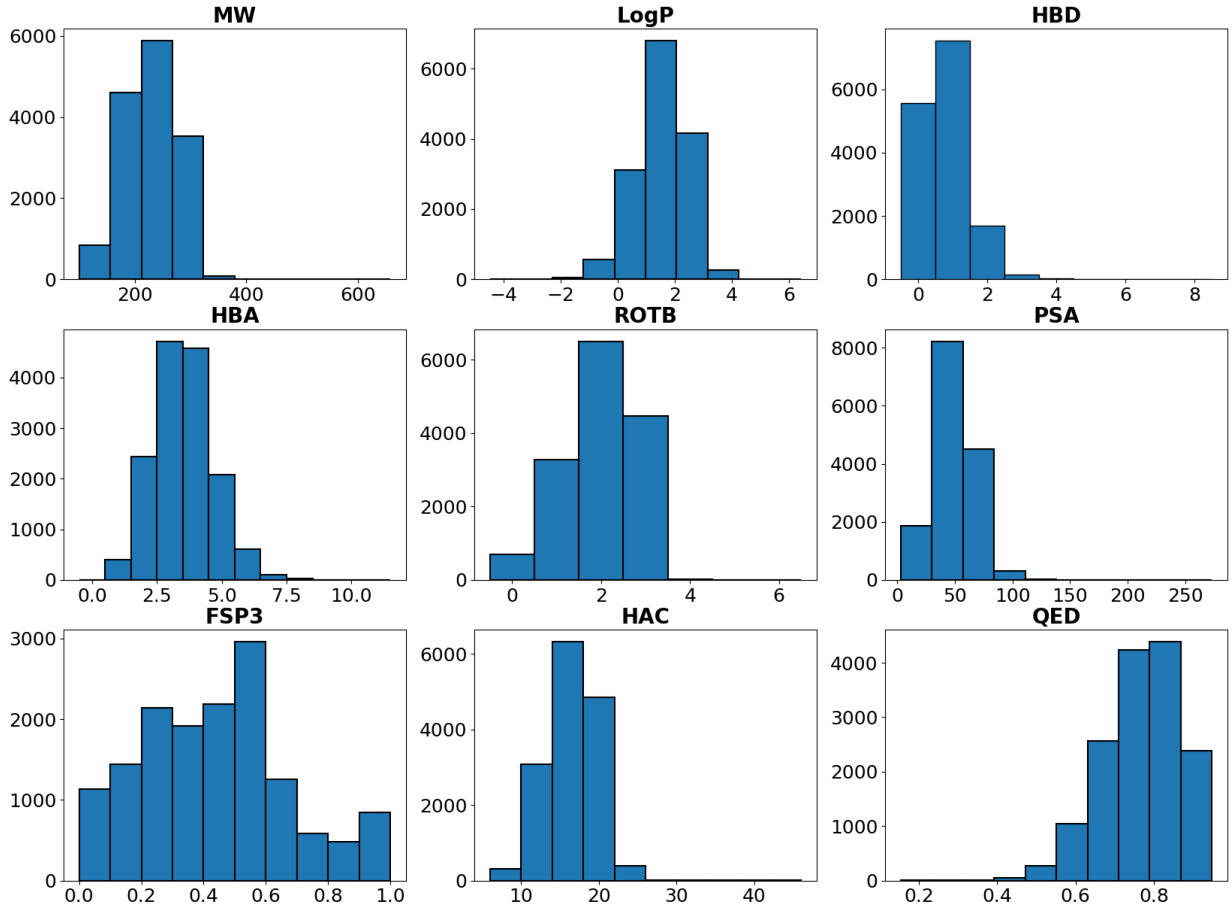


Figure S.1: Distribution of molecular properties in Life Chemicals general fragment library (total = 14,947): MW (Da) (229.68 ± 45.78), LogP (1.54 ± 0.88), HBD (0.76 ± 0.70), HBA (3.53 ± 1.18), ROTB (1.99 ± 0.84), PSA (49.64 ± 17.18), FSP3 (0.43 ± 0.24), HAC (16.03 ± 3.25), QED (0.77 ± 0.10).

Macrocyclization and Enumeration of Stereoisomers

The random fragment connection method was modified to perform intramolecular reactions. All the cycles were produced and stored as a list and passed onto a ring size check through rdkit.Chem.MolFromSmarts. Only the ones that had at least twelve atoms on their rings

were added to a new list. All stereoisomers of the list of macrocycles were enumerated using rdkit.Chem.EnumerateStereoisomers.

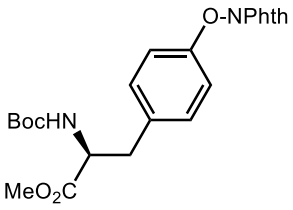
3. Experimental Section for Synthetic Peptide Macrocyclization

General Procedure 1 (Chan-Evans-Lam optimization procedure): To an oven-dried 7 mL reaction vial equipped with a stir bar was added Boc-Tyr($B(OH)_2$)-OMe **3.46** (0.1000 g, 0.31 mmol, 1.0 equiv), activated 4Å MS (70 mg, 250 mg/mmol), N-hydroxyphthalimide (0.0505 g, 0.31 mmol, 1.0 equiv), and CuCl (0.0307 g, 0.31 mmol, 1.0 equiv). The mixture was dissolved in 1,2-DCE (2.0 mL) and pyridine (0.0270 g, 0.34 mmol, 1.1 equiv). The reaction was placed into a 70 °C heating block and allowed to stir open to air or under an O_2 balloon. After 24 hours, the reaction was cooled to room temperature. The mixture was washed with 2.0 N NaOH (3 x 50 mL) and the organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo*. Purification by flash chromatography on silica (30% EtOAc in hexanes) afforded the title compound.

General Procedure 2 (Solution Phase Peptide Synthesis): To a flame-dried two neck RBF equipped with a stir bar was added the amine (1.0 equiv). DCM (0.1 M/0.25 M as indicated) was added to the flask, followed by DIPEA (2.0 equiv) if needed. On dissolution, the mixture was cooled to 0 °C in an ice bath. Under positive pressure, the free acid (1.0 or 1.1 equiv as indicated) and 1-hydroxybenzotriazole (HOBt) (1.1 equiv) were added successively, each in one portion. The suspension was allowed to stir for 15 mins, and then (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) (EDC) (1.1 equiv) was added. The reaction was stirred overnight, allowed to warm up to room temperature. The solution was washed with either cold 0.1 N HCl or 10% citric acid followed by saturated $NaHCO_3$. The aqueous layers were each extracted with DCM again. The combined organic layers were dried over Na_2SO_4, filtered,

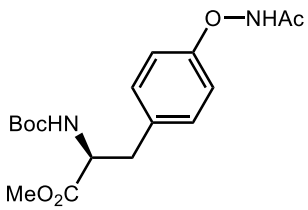
concentrated *in vacuo*. Purification by flash chromatography on silica afforded the title compound.

General Procedure 3 (Fmoc Deprotection): To a flame-dried RBF equipped with a stir bar was added the Fmoc-protected peptide (1.0 equiv). 20% or 10% piperidine/DMF (0.1 M) was added, and the reaction was allowed to stir at room temperature for 10 mins. The solvent was removed *in vacuo*, and the resulting white crystals were used directly for next coupling.



3.47

***Methyl (S)-3-(4-(acetamidooxy)phenyl)-2-((tertbutoxycarbonyl)amino)propanoate* (3.47)** – Synthesized according to **General Procedure 1** (0.0232 g, 17% yield). **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.91 (dd, J = 5.4, 3.1 Hz, 2H), 7.83 – 7.78 (m, 2H), 7.15 – 7.07 (m, 4H), 4.95 (d, J = 8.3 Hz, 1H), 4.55 (q, J = 6.6 Hz, 1H), 3.70 (s, 3H), 3.09 (dd, J = 14.0, 5.8 Hz, 1H), 3.02 (dd, J = 14.0, 6.1 Hz, 1H), 1.41 (s, 9H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 172.27, 163.06, 158.15, 155.19, 135.05, 132.60, 130.74, 128.93, 124.13, 114.87, 80.17, 54.47, 52.42, 37.59, 28.42. **HRMS** (-APCI) calculated for $C_{23}H_{23}O_7N_2$ [M-H]- 439.1511, found 439.1514.

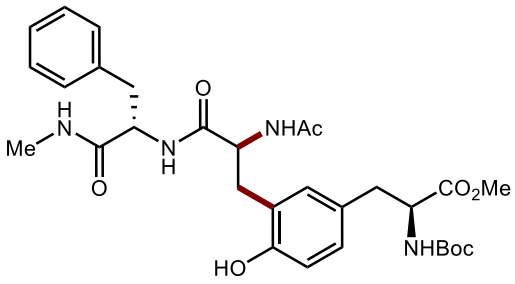


3.15

***Methyl (S)-3-(4-(acetamidooxy)phenyl)-2-((tertbutoxycarbonyl)amino)propanoate* (2.15)** – To a 25 mL flame-dried RBF equipped with a stir bar was added a solution of **3.47** (0.1925 g, 0.44 mmol, 1.0 equiv) in DCM (4.5 mL, 0.1 M). Hydrazine hydrate (0.0875 g, 1.75 mmol, 4.0 equiv) was added and the reaction stirred at room temperature. After 24 hours, $MgSO_4$ (2 scoops) was added, and the reaction continued to stir at room temperature. The reaction was filtered and the solid washed with DCM (10.0 mL) and EtOAc (10.0 mL) before solvent removal *in vacuo*. To a 10 mL flame-dried RBF equipped with a

stir bar was added Na_2CO_3 (0.0562 g, 0.53 mmol, 1.2 equiv). The crude material from the first step was dissolved in EtOAc (1.3 mL) and the solution added to the reaction vial followed by DI H_2O (0.7 mL). Acetyl chloride (0.0374 g, 0.479 mmol, 1.1 equiv) was added and the reaction allowed to stir at room temperature. After 16 hours, the reaction was quenched with saturated $NaHCO_3$ (2.5 mL). The aqueous layer was extracted with EtOAc (3 x 10 mL), and the combined organic layers dried over $MgSO_4$, filtered, and concentrated *in vacuo*. Purification by flash chromatography on silica afforded **3.15** (0.0694 g, 51% yield) as a white solid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.20 – 6.91 (m, 4H), 4.96 (s, 1H), 4.54 (d, J = 7.4 Hz, 1H), 3.71 (s, 3H), 3.06 – 2.97 (m, 2H), 2.08 (s, 3H), 1.41 (s, 9H).[3]

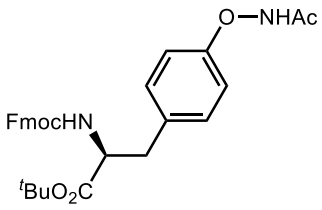
Experimental Procedures for Compound Synthesized by Mengfei (Sophia) Xu



3.25

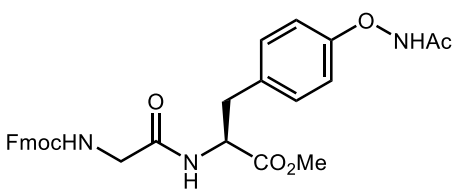
Methyl (S)-3-(3-((S)-2-acetamido-3-(((S)-1-(methylamino)-1-oxo-3-phenylpropan-2-yl)amino)-3-oxopropyl)-4-hydroxyphenyl)-2-((tert-butoxycarbonyl)amino)propanoate **(3.25)** – To an oven-dried 4 mL vial equipped with a stir bar and activated 3 Å MS in a nitrogen-filled glove box, added Boc-Tyr(OHNAc)-OMe **3.15** (0.0301 g, 0.085 mmol, 1.0 equiv), acryloyl-*L*-Phe-NHMe 3.24 (0.0197 g, 0.085 mmol, 1.0 equiv), K_3PO_4 (0.0045 g, 0.021 mmol, 25 mol %), CsOAc (0.0043 g, 0.021 mmol, 25 mol %), [Cp*Co(CO)I_2] (0.0041g, 0.009 mmol, 10 mol %), sealed with a teflon-cap and removed from the glove box. TFE (0.84 mL, 0.1 M) was added to the reaction vial. The vial was placed into a 60 °C heating block and allowed to stir for 24 hours before being removed and cooled to room temperature. After cooling, a stock solution of methyl 3,5-dinitrobenzoate (1.0 equiv) in DCM (1 mL) was added and the reaction was filtered through a silica plug and concentrated *in vacuo* before analysis by ^{1}H NMR. Purification by preparative TLC (2 x 10% Acetone/EtOAc) afforded both diastereomers (0.0130 g, 26% yield). **^{1}H NMR**

(400 MHz, $CDCl_3$) δ 7.30 – 7.16 (m, 5H), 7.08 (d, J = 7.1 Hz, 2H), 6.98 (d, J = 8.2 Hz, 2H), 6.91 – 6.78 (m, 2H), 6.77 – 6.69 (m, 1H), 4.53 – 4.47 (m, 3H), 3.72 (s, 3H), 3.20 – 2.94 (m, 2H), 2.91 (s, 2H), 2.83 – 2.67 (m, 3H), 2.01 (s, 3H), 1.40 (s, 9H). **HRMS** (+APCI) calculated for $C_{30}H_{40}N_4O_8$ $[M+H]^+$ 585.2924, found 585.2921.



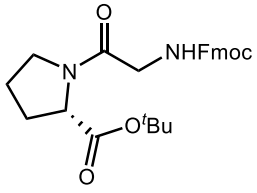
3.29

***tert-butyl (S)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-(acetamidooxy)phenyl)propanoate* (3.29)** – To a 100 mL flame-dried two neck RBF equipped with a stir bar was added Fmoc-*L*-Tyr-O^{t}Bu (4.9274 g, 10.7 mmol, 1.0 equiv). MeOH (10.7 mL) was added followed by potassium *tert*-butoxide (1.3263 g, 11.8 mmol, 1.1 equiv). The mixture was allowed to stir at room temperature for 30 mins. MeOH was removed *in vacuo*. To a 250 mL flame-dried RBF equipped with a stir bar was added the residue from last step dissolved in DCM (5 mL). The flask was cooled to 0 °C in an ice bath and the freshly prepared **3.44** (2.3070 g, 10.7 mmol, 1.0 equiv) dissolved in DCM (5.7 mL) was added. The reaction bubbled vigorously. The reaction was allowed to stir for 1 hr, DCM was removed *in vacuo* to afford the corresponding *N*-aryloxyamine. The solids were dissolved in Et_2O (54 mL) and cooled to 0 °C in an ice bath. Acetic anhydride (2.06 mL, 21.4 mmol, 2.0 equiv) was added slowly to the solution. The reaction was stirred for 3 hours, allowed to warm to room temperature. The mixture was concentrated *in vacuo* and purified by flash column chromatography (20% EtOAc/hexanes) to afford **3.29** (0.9925 g, 18% yield, 2 steps) as yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.81 – 7.74 (m, 2H), 7.64 – 7.57 (m, 2H), 7.41 (t, J = 7.4 Hz, 2H), 7.32 (t, J = 7.4 Hz, 2H), 7.19 (d, J = 8.2 Hz, 2H), 7.06 – 6.99 (m, 2H), 5.47 (d, J = 8.1 Hz, 1H), 4.57 (dt, J = 8.1, 6.1 Hz, 1H), 4.46 (dd, J = 10.6, 7.2 Hz, 1H), 4.23 (t, J = 7.1 Hz, 1H), 3.10 (d, J = 6.2 Hz, 2H), 2.29 (s, 3H), 1.43 (s, 9H).



3.33

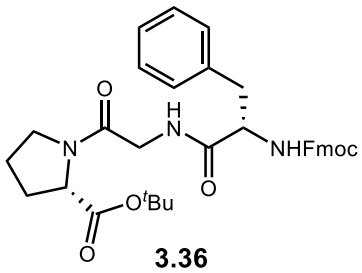
***Methyl (S)-2-(2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)acetamido)-3-(4-(acetamidooxy)phenyl)propanoate* (3.33)** – To a 100 mL flame-dried two neck RBF equipped with a stir bar was added Boc-Tyr(OHNAc)-OMe **3.15** (0.1830 g, 0.52 mmol, 1.0 equiv). The solid was dissolved in DCM (5.2 mL) and the solution was cooled to 0 ℃ in an ice bath. HCl (4.0 M) in dioxane (5.2 mL) was added dropwise. The reaction was allowed to stir at 0 ℃ for 2 hours, monitored by TLC. The solvent was removed *in vacuo* to afford *L*-Tyr(OHNAc)-OMe·HCl as brown solid. The dimer was synthesized based on **General Procedure 2**, utilizing the product from last step, DCM (5.2 mL), DIPEA (0.1344 g, 1.04 mmol, 2.0 equiv), Fmoc-Gly-OH (0.1546 g,0.52 mmol, 1.0 equiv), HOBt (0.0809 g, 0.59 mmol, 1.1 equiv) and EDC (0.1131 g, 0.59 mmol, 1.1 equiv). After stirring overnight, the solution was washed with 10% citric acid followed by saturated $NaHCO_3$. Purification by flash column chromatography (80% EtOAc/hexanes) afforded **3.33** (0.0562 g, 26% yield, 2 steps) as white solid. **^{1}H NMR** (600 MHz, DMSO-d6) δ 11.62 (s, 1H), 8.29 (d, J = 7.8 Hz, 1H), 7.89 (d, J = 7.6 Hz, 2H), 7.71 (d, J = 7.5 Hz, 2H), 7.50 (t, J = 6.2 Hz, 1H), 7.43 – 7.38 (m, 2H), 7.36 – 7.30 (m, 2H), 7.13 (d, J = 8.2 Hz, 2H), 6.89 (d, J = 8.2 Hz, 2H), 4.44 (td, J = 8.1, 5.7 Hz, 1H), 4.27 (d, J = 7.8 Hz, 2H), 4.22 (t, J = 7.1 Hz, 1H), 3.65 (dd, J = 16.9, 6.3 Hz, 1H), 3.60 (s, 3H), 3.58 – 3.50 (m, 1H), 2.96 (dd, J = 13.8, 5.7 Hz, 1H), 2.87 (dd, J = 13.8, 8.6 Hz, 1H), 1.88 (s, 3H). **HRMS** (+APCI) calculated for $C_{29}H_{29}N_3O_7Na$ $[M+Na]^+$ 554.1903, found 554.1909.



3.35

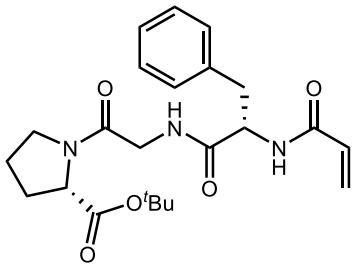
***tert-butyl (((9H-fluoren-9-yl)methoxy)carbonyl)glycyl-L-prolinate* (3.35)** – Synthesized based on **General Procedure 2** from *L*-Pro-O^{t}Bu **(3.34)** (1.0410 g, 5.0 mmol, 1.0 equiv), Fmoc-Gly-OH (1.4857 g, 5.0 mmol, 1.0 equiv), DCM (20 mL), HOBt (0.7454 g, 5.5 mmol, 1.1 equiv) and EDC (1.0506 g, 5.5 mmol, 1.1 equiv). After stirring overnight, the solution was

washed with cold 0.1 N HCl. Purification by flash column chromatography (40% EtOAc/hexanes) afforded **3.35** (1.9561 g, 87% yield) as white foam. **^{1}H NMR (**400 MHz, DMSO-d6) δ 7.89 (d, J = 7.5 Hz, 2H), 7.73 (d, J = 7.5 Hz, 2H), 7.50 (t, J = 6.1 Hz, 1H), 7.42 (t, J = 7.5 Hz, 2H), 7.32 (tt, J = 7.4, 1.2 Hz, 2H), 4.31 – 4.21 (m, 3H), 4.20 – 4.14 (m, 1H), 3.92 – 3.81 (m, 1H), 3.75 (dd, J = 17.1, 5.8 Hz, 1H), 3.58 – 3.36 (m, 2H), 2.20 – 2.05 (m, 1H), 1.96 – 1.84 (m, 2H), 1.79 (ddd, J = 11.8, 9.3, 5.2 Hz, 1H), 1.38 (s, 9H).



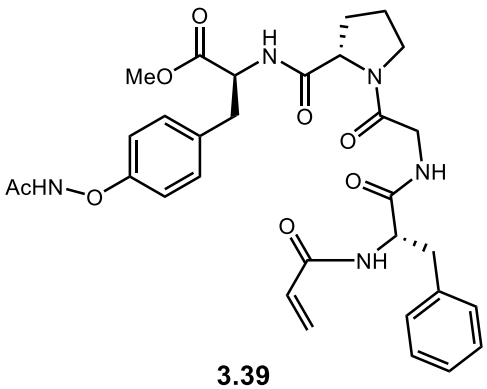
3.36

***tert-butyl (((9H-fluoren-9-yl)methoxy)carbonyl)-L-phenylalanylglycyl-L-prolinate* (3.36)** – **3.35** (0.9067g, 2.0 mmol, 1.0 equiv) was deprotected based on **General Procedure 3** with 10% piperidine/DMF (20 mL total). The trimer was then synthesized based on **General Procedure 2**, utilizing the product from deprotection, DCM (20 mL), Fmoc-Phe-OH (0.8571 g, 2.2 mmol, 1.1 equiv), HOBt (0.2970 g, 2.2 mmol, 1.1 equiv) and EDC (0.4254 g, 2.2 mmol, 1.1 equiv). After stirring overnight, the solution was washed with 10% citric acid followed by saturated $NaHCO_3$. Purification by flash column chromatography (50% EtOAc/hexanes) afforded **3.36** (0.7204 g, 60% yield, 2 steps) as white solid. **^{1}H NMR** (400 MHz, DMSO-d6) δ 8.20 (t, *J* = 5.4 Hz, 1H), 7.88 (d, *J* = 7.4 Hz, 2H), 7.71 – 7.63 (m, 3H), 7.38 – 7.23 (m, 9H), 4.32 (ddd, *J* = 11.7, 8.6, 3.6 Hz, 1H), 4.21 (dt, *J* = 9.1, 4.6 Hz, 1H), 4.06 – 3.97 (m, 1H), 3.87 (dd, *J* = 17.3, 4.7 Hz, 1H), 3.53 (q, *J* = 6.6 Hz, 2H), 3.11 – 2.97 (m, 2H), 2.78 (dd, *J* = 13.8, 11.1 Hz, 1H), 2.21 – 2.05 (m, 2H), 1.98 – 1.77 (m, 3H), 1.39 (s, 9H).



3.37

***tert-butyl acryloyl-L-phenylalanylglycyl-L-prolinate* (3.37)** – **3.36** (0.6662 g, 1.1 mmol, 1.0 equiv) was deprotected based on **General Procedure 3**. DCM (11 mL) and DIPEA (0.2559 g, 2.0 mmol, 1.8 equiv) was added to the reaction vial and the solution was cooled to 0 °C in an ice bath. Acryloyl chloride (0.0996 g, 1.1 mmol, 1.1 equiv) was added dropwise. The reaction was allowed to stir at room temperature overnight. The mixture was washed with 5% $KHSO_4$ followed by saturated $NaHCO_3$. The combined organic layers were dried over Na_2SO_4, filtered, concentrated *in vacuo*. Purification by flash column chromatography (5% MeOH/DCM) afforded **3.37** (0.3023 g, 64% yield, 2 steps) as white solid. **^{1}H NMR** (400 MHz, DMSO-d6) δ 8.38 (dd, *J* = 8.6, 4.2 Hz, 1H), 8.25 (dt, *J* = 15.7, 5.4 Hz, 1H), 7.32 – 7.10 (m, 5H), 6.24 (ddd, *J* = 17.1, 10.2, 3.4 Hz, 1H), 5.98 (dt, *J* = 17.1, 1.9 Hz, 1H), 5.52 (dt, *J* = 10.2, 1.9 Hz, 1H), 4.76 – 4.61 (m, 1H), 4.19 (dd, *J* = 8.8, 3.6 Hz, 1H), 4.08 – 3.80 (m, 2H), 3.59 – 3.46 (m, 2H), 3.07 (dd, *J* = 13.8, 3.8 Hz, 1H), 2.75 (ddd, *J* = 13.9, 10.5, 3.9 Hz, 1H), 2.25 – 2.03 (m, 2H), 1.96 – 1.70 (m, 3H), 1.40 (d, *J* = 18.8 Hz, 9H).



3.39

***Methyl (S)-3-(4-(acetamidooxy)phenyl)-2-((S)-1-(acryloyl-L-phenylalanylglycyl)pyrrolidine-2-carboxamido)propanoate* (3.39)** – **3.37** (0.3023g, 0.67 mmol, 1.0 equiv) was dissolved in DCM (3.35 mL) and TFA (3.35 mL) was added to deprotect the peptide. After stirring at room temperature for 1 hour, the solvent was removed *in vacuo* to afford acryloyl-Phe-Gly-Pro-OH·TFA as brownish oil. To a 50 mL flame-dried two neck RBF equipped with a stir bar was added Boc-Tyr(OHNAc)-OMe **3.15** (0.2365 g, 0.67 mmol, 1.0 equiv). The solid was dissolved in DCM (6.7 mL) and the solution was cooled to 0 °C in an ice bath. HCl (4.0 M) in dioxane (6.7 mL) was added slowly.

The reaction was allowed to stir at 0 °C for 2 hours, monitored by TLC. The solvent was removed *in vacuo* to afford *L*-Tyr(OHNAc)-OMe·HCl as brown solid. DCM (5 mL) was added to the flask followed by DIPEA (0.35 mL, 2.0 mmol, 3.0 equiv). On dissolution, the mixture was cooled to 0 °C in an ice bath. The TFA salt obtained above was dissolved in DCM (2 mL) and transferred to the flask. Under positive pressure added HOBt (0.1014g, 0.74 mmol, 1.1 equiv). The suspension was allowed to stir for 15 mins, and then EDC (0.1444 g, 0.74 mmol, 1.1 equiv) was added. The reaction was stirred overnight, allowed to warm up to room temperature. The solution was washed with 10% citric acid followed by saturated $NaHCO_3$. The combined organic layers were dried over Na_2SO_4, filtered, removed *in vacuo*. Purification by RP-HPLC afforded **3.39** (0.1784 g, 44% yield, 3 steps) as yellow foam. ^{1}H NMR (400 MHz, DMSO-d6) δ 11.63 (s, 1H), 8.58 – 8.07 (m, 3H), 7.31 – 7.19 (m, 4H), 7.16 (tdd, J = 7.9, 5.1, 2.1 Hz, 3H), 7.00 – 6.79 (m, 2H), 6.25 (ddd, J = 17.1, 10.2, 1.7 Hz, 1H), 5.99 (dd, J = 17.1, 2.2 Hz, 1H), 5.53 (dd, J = 10.2, 2.1 Hz, 1H), 4.10 – 3.74 (m, 2H), 3.58 (d, J = 14.0 Hz, 3H), 3.49 – 3.19 (m, 2H), 3.15 – 2.85 (m, 3H), 2.85 – 2.65 (m, 1H), 2.22 – 1.91 (m, 2H), 1.88 (s, 2H), 1.83 – 1.51 (m, 3H). **HRMS** (+APCI) $C_{31}H_{38}N_5O_8$ $[M+H]^+$ 608.2715, found 608.2719.

4. Supplementary References

(1) Nacsa, E. D.; MacMillan, D. W. C. Spin-Center Shift-Enabled Direct Enantioselective a-Benzylation of Aldehydes with Alcohols. *J. Am. Chem. Soc.* **2018**, *149*, 3322-3330.

(2) a) Li, B.; Lan, J.; Wu, D.; You, J. Rhodium(III)-Catalyzed ortho-Heteroarylation of Phenols through Internal Oxidative C–H Activation: Rapid Screening of Single-Molecular White-Light-Emitting Materials. *Angew. Chem. Int. Ed.* **2015**, *54*, 14008-14012. b) Yan, D.; Want, G.; Xiong, F.; Sun, W. Y.; Shi, Z.; Lu, Y.; Li, S.; Zhao, J. A Selenium-Catalysed Para-Amination of Phenols. *Nat. Commun.* **2018**, *9*, 1-9.

(3) Liu, G.; Shen, Y.; Zhou, Z; Lu, X. Rhodium(III)-Catalyzed Redox-Neutral Coupling of N-Phenoxyacetamides and Alkynes with Tunable Selectivity. *Angew. Chem. Int. Ed.* **2013**, *52*, 6033-6037.

(4) Sun, B.; Yoshino, T.; Matsunaga, S.; Kanai, M. A Cp^*CoI_2-dimeras a precursor for cationic Co(III)-catalysis: Application to C–H phosphoramidation of indoles. *Chem. Commun.* **2015**, *51*, 4659-4661.

(5) Bernardes, G. J.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.

(6) Life Chemicals https://lifechemicals.com/ (accessed Jun 29, 2020).