Computer-Guided Drug Discovery: Application to the Development of Inhibitors Targeting Metalloenzymes, Nucleic Acids and Covalent Residues

Julia Stille

Abstract

Drug discovery is known to be a long and expensive process, inspiring the development of computational methods to guide drug design and improve the efficiency of this process. One of the main computational methods used in drug discovery is molecular docking, which predicts the binding of a small molecule ligand to a macromolecular target. However, docking programs have largely been developed by modeling non-covalent interactions to protein targets and face challenges when applied to targets where binding is heavily mediated by additional interactions, such as with metalloenzymes, covalent residues and nucleic acids. This thesis presents the application of molecular docking towards the development of inhibitors for these more computationally challenging therapeutic targets.

The thesis begins with a short review of the current literature on the development and application of docking programs for metalloenzymes, covalent residues and nucleic acid targets (Chapter 1).

The second chapter then presents the application of molecular docking towards activation-induced cytidine deaminase (AID), a zinc metalloenzyme implicated in the progression of blood cancer. This chapter describes synthetic efforts to synthesize a preliminary inhibitor of AID as well as a virtual screen to identify novel inhibitors. Discouraging experimental results led us to further evaluate the accuracy of our docking program with the zinc coordination sphere in AID, which ultimately was in good agreement with higher level quantum mechanics calculations (Chapter 2).

The third chapter presents the use of virtual screening in the discovery of inhibitors against the thiamine pyrophosphate (TPP) riboswitch, a promising nucleic acid target for the development of antibacterial agents. This work led to the identification of a low micromolar binder to the TPP riboswitch, one of the first examples of a successful application of docking towards this target class (Chapter 3).

The subsequent two chapters detail the computer-aided design of covalent inhibitors for SARS-CoV-2 3CL^{pro}, a cysteine protease identified as a promising target for coronavirus therapeutics. Docking was initially employed to assess the incorporation of a covalent warhead into a non-covalent inhibitor scaffold and led to the development of several potent covalent inhibitors whose activity was assessed *in vitro* and confirmed by X-ray crystallography (Chapter 4). Following the

identification of the most promising covalent warheads, our focus then turned to the optimization of the inhibitor scaffold (Chapter 5). As the inhibitors were accessible by a multicomponent reaction, the effect of each component on potency was assessed independently and led to an optimized inhibitor with a nearly four-fold improvement in potency. An additional heterocyclic analogue with similar *in vitro* and *in cellulo* activity was also successfully developed based on bioisosteric replacement of an amide bond.

As a change of pace, the next chapter details the development of an organocatalytic asymmetric Pictet-Spengler reaction as a method for accessing tetrahydro-β-carbolines (TBHCs) (Chapter 6). TBHCs are ubiquitous among indole alkaloid natural products with well-documented bioactivity, and several synthetic drugs have been developed based on these scaffolds. As the Pictet-Spengler reaction generates chiral products, there has been significant interest in the development of asymmetric variants, however the majority of reported methods are unsuccessful with ketone substrates and can therefore only be applied to the synthesis of enantioenriched tertiary stereocenters. This chapter presents the development of an asymmetric organocatalytic Pictet-Spengler reaction for diketone substrates, enabling the formation of enantioenriched TBHCs with quaternary stereocenters.

Résumé

La découverte de médicaments est connue pour être un processus long et coûteux, ce qui a encouragé le développement de méthodes informatiques pour guider et améliorer l'efficacité de ce processus. L'une des principales méthodes informatiques utilisées dans la découverte de médicaments est l'amarrage moléculaire, qui prédit la liaison d'un ligand à une cible macromoléculaire. Cependant, les logiciels d'amarrage ont été largement développés en modélisant les interactions non-covalentes avec les cibles protéiques et rencontrent des difficultés lorsqu'ils sont appliqués à des cibles où la liaison est fortement médiée par des interactions supplémentaires – comme pour les métalloenzymes, les résidus covalents et les acides nucléiques. Cette thèse présente donc l'application de l'amarrage moléculaire au développement d'inhibiteurs pour ces cibles thérapeutiques plus difficiles à modéliser.

La thèse commence par une brève revue de la littérature sur le développement et l'application des programmes d'amarrage pour les métalloenzymes, les résidus covalents et les cibles d'acides nucléiques.

Le deuxième chapitre aborde ensuite l'application de l'amarrage moléculaire à l'enzyme AID (Activation-Induced Cytidine Deaminase), une métalloenzyme à zinc impliquée dans la progression du cancer du sang. Ce chapitre décrit les efforts de synthèse d'un inhibiteur préliminaire de l'AID ainsi qu'un criblage virtuel pour identifier de nouveaux inhibiteurs. Des résultats expérimentaux décourageants nous ont amenés à réévaluer en amont la précision de notre programme d'amarrage vis-à-vis de la sphère de coordination du zinc dans l'AID, qui s'est finalement avérée en bon accord avec des calculs de mécanique quantique.

 Le troisième chapitre présente l'utilisation du criblage virtuel dans la découverte d'inhibiteurs du riboswitch de la thiamine pyrophosphate (TPP), une cible d'acide nucléique prometteuse pour le développement d'agents antibactériens. Ces travaux ont conduit à l'identification d'une nouvelle molécule qui se lie au riboswitch TPP à une concentration micromolaire, représentant l'un des premiers exemples d'application réussie de l'amarrage moléculaire à cette classe de cibles.

Les deux chapitres suivants détaillent la conception assistée par ordinateur d'inhibiteurs covalents pour la 3CL^{pro} du SARS-CoV-2, une protéase à cystéine identifiée comme une cible prometteuse pour la thérapeutique des coronavirus. L'amarrage moléculaire a d'abord été utilisé pour évaluer

l'incorporation d'un groupe covalent dans une structure d'inhibiteur non-covalent et a mené au développement de plusieurs inhibiteurs covalents candidats dont la liaison a été évaluée *in vitro* et confirmée par cristallographie aux rayons X (chapitre 4). Après l'identification des groupes covalents les plus prometteurs, nous nous sommes concentrés sur l'optimisation de la structure chimique de ces inhibiteurs (chapitre 5). Ces derniers étant accessibles par une réaction à plusieurs composants, l'effet de chaque composant sur l'efficacité de l'inhibiteur a été évalué indépendamment, ce qui a permis d'obtenir un inhibiteur optimisé dont l'activité a été multipliée par près de quatre. Un autre analogue hétérocyclique présentant une activité *in vitro* et *in cellulo* a également été développé avec succès par remplacement bioisostérique d'un groupe amide.

Pour changer de rythme, le chapitre suivant détaille le développement d'une réaction organocatalytique asymétrique de Pictet-Spengler comme méthode d'accès aux tétrahydro-β-carbolines (TBHC) (chapitre 6). Les TBHC sont omniprésents parmi les produits naturels alcaloïdes indoliques dont la bioactivité est bien documentée, et plusieurs médicaments synthétiques ont été développés sur la base de ces structures. La réaction de Pictet-Spengler générant des produits chiraux, le développement de variantes asymétriques a suscité un grand intérêt. Cependant, la majorité des méthodes rapportées ne réussissent pas avec les substrats cétoniques et ne peuvent donc être appliquées qu'à la synthèse de stéréocentres tertiaires. Ce chapitre présente le développement d'une réaction organocatalytique asymétrique de Pictet-Spengler pour les substrats dicétoniques, permettant la formation de TBHC énantioenrichis avec des stéréocentres quaternaires.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Nicolas Moitessier, for his support throughout my time in the Moitessier lab. He was always available to offer guidance when I needed it, while also encouraging me to work independently and believe in my own abilities. I feel very thankful to have had a supervisor who not only supported me in my research goals, but also genuinely cared about my personal well-being beyond academia.

To my family – I couldn't have done this without you. I feel so fortunate to have parents that have provided me with the freedom to pursue my dreams, whatever they may be. Your support during my PhD, whether through visits to Montreal or Switzerland, or welcoming me into your home in BC, was invaluable. I'm forever grateful that I got to share so many moments during this journey with you all.

To my friends – where do I even begin?

Sharon and Anne – I truly can't imagine having gone through this experience without you both. I am so lucky to have had you as lab mates and meeting you has easily been the best part of my PhD. Words can't convey how much you guys mean to me. From sangria nights filled with laughter, lab floods, rock talks, camping trips, skating at Parc Lafontaine, and whisper yelling at the spa, the list of memories could go on forever. You both are like family to me and I can't wait for all of the amazing moments together that are yet to come. I also want to thank your wonderful families – the time I got to spend with them (visiting Sharon's family in Tel Aviv and our annual tomato sauce canning with Anne's family in Montreal) are some of my fondest memories.

Jérôme – Although you joined the lab more recently, I am so glad we overlapped. Your friendship has meant so much to me and I'm constantly inspired by your compassion, intelligence, and maturity. Thank you for all the amazing coffee shop/smoothie work sessions, koh-lanta evenings, tennis games at Jeanne Mance, and the hardest crossword ever made.

Ahmed – I think you're basically an honorary member of the Moitessier lab at this point. Although we only met recently, you feel like someone I've known forever. Thank you for being a wonderful friend, and I can't wait for more card game nights and early morning runs together!

Chris – Thank you for being a ray of sunshine in the office. Whether by bringing in Ramen or just

being the wonderfully positive person that you are, you always made the day a little brighter. I feel very thankful that I got to work with you and learn from you (and I especially appreciate your patience in teaching me all there is to know about bioassays and the plate reader!).

I would also like to thank the other Moitessier lab members, past and present – it has been a pleasure working alongside you and I am deeply appreciative of all your encouragement and support.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

3CL^{pro}	3-chymotrypsin-like protease
Ac	acetyl
ACE	angiotensin-converting enzyme
AID	activation-induced cytidine deaminase
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
AUROC	area under the ROC curve
BEDROC	Boltzmann enhanced discrimination of the ROC curve
Boc	*tert*-butyloxycarbonyl
BSA	bovine serum albumin
CA	carbonic anhydrase
CDI	carbonyl diimidazole
CN	cyano-
CoV	coronavirus
d	days
dC	deoxycytidine
DCM	dichloromethane
DIPEA	diisopropylethylamine
DMF	dimethyl formamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dU	deoxyuridine
ee	enantiomeric excess
EF	enrichment factor
ESI	electrospray ionization
Et	ethyl
EtOAc	ethyl acetate
EtOH	ethanol

FAP	fibroblast activation protein α
FDA	food and drug administration
FMN	flavin mononucleotide
HDAC	histone deacetylase
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSP	heat shock protein
I	inhibitor
IC_{50}	half maximal inhibitory concentration
KIE	kinetic isotope effect
K_M	Michaelis-Menten constant
LC-MS	liquid chromatography mass spectrometry
Me	methyl
MeCN	acetonitrile
MeOD	deuteratured methanol
MeOH	methanol
MERS	middle East respiratory syndrome
ML	machine Learning
MM	molecular mechanics
MMP	matrix metalloproteinase
M^{pro}	main protease
mRNA	messenger RNA
NMR	nuclear magnetic resonance
PDB	protein data bank
POP	prolyl oligopeptidase
PTPP	pyrithiamine pyrophosphate
QM	quantum mechanics
r.t.	room temperature
RMSD	root mean square deviation
RNA	ribonucleic acid

ROC	receiver operating characteristic
S	substrate
SARS	severe acute respiratory syndrome
SFC	supercritical fluid chromatrography
SPR	surface plasmon resonance
ssDNA	single stranded DNA
Temp.	temperature
THBC	tetrahydro-beta-carboline
THIQ	tetrahydroisoquinoline
THU	tetrahydrouridine
TIPS	triisopropylsilyl
TLC	thin layer chromatography
TPP	thiamine pyrophosphate
Ts	tosyl
Ugi-4-CR	4-component Ugi reaction
vdW	van der Waals
WHO	world health organization

List of Author Contributions

During the course of my PhD I have co-authored 2 published manuscripts and 3 manuscripts in preparation:

Publications. [‡] denotes first author or co-first author.

1. A. Labarre[‡], **J. K. Stille**, M. Burai-Patrascu, A. Martins, J. Pottel, N. Moitessier. Docking Ligands into Flexible and Solvated Macromolecules. 8. Forming New Bonds─Challenges and Opportunities. *J Chem Inf Model*. 2022, *62,* 1061-1077.

2. **J. K. Stille**[‡], J. Tjutrins[‡], G. Wang[‡], F. A. Venegas[‡], C. Hennecker, A. M. Rueda, I. Sharon, N. Blaine, C. E. Miron, S. Pinus, A. Labarre, J. Plescia, M. Burai Patrascu, X. Zhang, A. S. Wahba, D. Vlaho, M. J. Huot, T. M. Schmeing, A. K. Mittermaier, and N. Moitessier. Design, synthesis and in vitro evaluation of novel SARS-CoV-2 3CL^{pro} covalent inhibitors. *Eur. J. Med. Chem.*, 2022, *229*, 114046.

Manuscripts in Preparation. [‡] denotes first author or co-first author.

1. H. Y. Huang[‡], S. Pinus, X. Zhang, G. Wang, A. M. Rueda, Y. Souaibou, S. Huck, M. Huot, D. Vlaho, J. Pottel, F. A. Venegas, Z. Lu, C. Hennecker, **J. K. Stille**, J. Tjutrins, C. E. Miron, A. Labarre, J. Plescia, M. Burai-Patrascu, A. K. Mittermaier, and N. Moitessier. Preprint on ChemRxiv: https://doi.org/10.26434/chemrxiv-2023-r0v7t

2. C. Hennecker[‡], F. Venegas, G. Wang, **J. K. Stille**, N. Moitessier, A. Mittermaier. 2024. Manuscript in preparation.

3. **J. K. Stille**[‡], N. Takyi[‡], O. S. Ayon[‡], M. Rivera, T. Satkunarajah, J. Pottel, A. Mittermaier, M. McKeague, N. Moitessier. 2024. Manuscript in preparation.

<div align="center">

CHAPTER 1:

MOLECULAR DOCKING WITH METAL-BINDING LIGANDS, COVALENT RESIDUES AND NUCLEIC ACIDS

</div>

Contribution of Authors

All work presented in this chapter is the work of **J.K. Stille**.

1.1   Introduction

Following the development of the first docking program by Kuntz *et al.* in the 1980s,[1] the field has since expanded to include many more improved programs, and its successful application in drug discovery has been extensively reviewed.[2-4] However, docking programs have largely been developed to predict the binding of non-covalent ligands to protein targets. The aim of this review is to introduce the current state of research regarding the development and application of docking programs towards more complex targets, focusing specifically on the unique challenges posed in docking to *metalloenzymes*, *covalent residues* and *nucleic acids.*

As the use of docking programs becomes more prevalent and accessible to medicinal chemists, it is important that reviews of the field reach beyond computational chemists. In order to cater to a broader audience, this review will first introduce molecular docking, outlining the basic components of a docking program and how their performance is evaluated. The aim is to enable all users of docking programs to understand the theory behind the process, allowing them to make well-informed decisions when developing a docking protocol or analyzing docking results.

1.2   Molecular Docking

Drug discovery is known to be a long and expensive process, with the development of a single approved drug taking on average between 10-15 years at a cost of up to $6 billion.[5] The use of computational methods to guide drug design has emerged as a key strategy to help improve the efficiency of this process. One of the main computational methods used in drug discovery is molecular docking, an *in silico* method that predicts the binding of a small molecule ligand to a macromolecular target.

Figure 1.1. Schematic of molecular docking.

Molecular docking can aid the drug discovery process in several ways. By predicting the binding mode of a ligand, docking programs can help elucidate the molecular basis for a ligand's binding affinity without necessitating an experimentally determined structure (e.g., X-ray or cryo-EM) which can be challenging to obtain. Understanding how a ligand binds its target can provide valuable information that can be employed to design more potent ligands. Molecular docking also gives a prediction of a compound's binding affinity by generating a "score" that represents the strength of binding. This score can be used to prioritize the most promising ligands, thereby improving the efficiency of structure-activity relationship studies and decreasing the number of compounds that need to be experimentally synthesized or evaluated. The docking score can also be used to virtually screen a large library of compounds against a target of interest, a computational alternative to experimental high-throughput screening. In this manner, potential ligands can be identified from a virtual library of compounds by selecting those with the best docking score.

Given that more than 70 docking programs have been reported to date, this review cannot serve as an exhaustive overview of every program. Focus is instead given to the most frequently cited programs[6] with publications detailing specific optimizations and assessments relevant to the three target classes (metalloenzymes, covalent residues, nucleic acids) discussed in this review. For example, while ICM[7] and Surflex-Dock[8] are among the more commonly cited programs, no evaluation of their applicability to metalloenzymes specifically has been reported and thus are not included in the discussion of docking to metalloenzymes.

1.2.1 Pose Prediction

The first aspect of a docking program is to accurately predict the conformation of a ligand within the target binding site (i.e. its *binding pose*). Most docking programs consider the flexibility of ligands with translational, rotational and torsional freedom, resulting in a massive number of theoretical ligand conformations that increases exponentially with the number of rotatable bonds. For example, considering only the torsional freedom of a ligand with ten rotatable bonds and just three minima per bond still results in nearly 60,000 possible ligand conformations.[9] In addition to conformational flexibility, a docking program also needs to consider the six degrees of rotational and translational freedom of a ligand that describe its ability to move and rotate within the binding site. Therefore, the challenge with pose prediction is to develop a method that is able to rapidly cover many theoretically possible ligand conformations, translations and orientations to

successfully identify the most favourable pose (i.e. the predicted binding mode). In order to address this challenge, several different search algorithms have been developed and implemented into docking programs.

In contrast to algorithms that dock the entire ligand, some methods take a fragment-based approach. One such technique is *incremental construction*, used in popular docking programs like DOCK[10] and FlexX[11]. Here, the ligand is broken into smaller rigid fragments, typically by severing at all rotatable bonds. The best orientation and translation of a "base fragment" is determined, serving as an anchor on which to add subsequent fragments. This process is repeated iteratively, reconstructing the ligand in a stepwise manor. DOCK and FlexX employ rapid *matching algorithms* to position base fragments. These algorithms generate interaction sites (such as hydrogen bond accepting, hydrogen bond donating, and hydrophobic) on both the ligand fragments and the receptor, represented by points, spheres, or volumes. The algorithm then utilizes the superimposition of compatible interaction sites to position the ligand within the binding site.

Many docking programs employ *stochastic search algorithms* such as a Monte Carlo algorithm (e.g., ICM, Glide[12]) or genetic algorithm (e.g., AutoDock[13], GOLD[14], FITTED[15]). In a Monte Carlo search algorithm, new configurations are generated based on random changes to ligand translation, rotation and torsions. The binding energy of the new configurations is assessed, keeping those that are more favourable. It also uses a probabilistic algorithm to keep some less favourable configurations, such that the search algorithm is able to go "uphill" (i.e. it is able to travel over energy maxima in order to reach greater minima). Genetic algorithms are similar in that they randomly generate an initial population of ligand configurations. It then encodes these poses in "chromosomes" based on their position, orientation and conformation. It assesses the "fitness" of the initial population, keeping the best poses and applying genetic functions (i.e. mutation, cross-over, etc.) to generate a subsequent population.

For a more detailed discussion of docking search algorithms, readers are referred to the comprehensive review by Halperin *et al.*[16] A thorough description of search algorithms in lay terms, intended for a medicinal chemistry audience, can be found in a more recent review by Moitessier *et al.*[17]

1.2.2 Scoring Functions

The second aspect of docking is the estimation of binding free energy between a ligand and a macromolecular target. This is accomplished through the use of scoring functions, enabling docking programs to generate a score that approximates the theoretical binding affinity of different ligands. Additionally, scoring functions are also used to assess the many potential ligand conformations generated by the search algorithm and therefore also play a role in accurate pose prediction. The second class of scoring function (which compares different poses of the same ligand) is often a simplified version of the first class (which compares the poses of different ligands). In this review, focus will be given to the first class of scoring functions.

Scoring functions can generally be classified into four main categories: *physics-based*, *empirical*, *knowledge-based*, and *machine learning-based*. Physics-based scoring functions (e.g., DOCK-GridScore[18], GOLD-GoldScore[19], FITTED-RankScore[20]), compute binding energies with molecular mechanics (MM)-based equations that use pairwise interactions between atoms to describe intermolecular interactions such as van der Waals (vdW) and electrostatics. These scoring functions may be enhanced by including a term to account for desolvation effects.[21] As a note: physics-based scoring functions have also been referred to as *force field-based*, as they typically employ energy functions and parameters from existing force fields (e.g., AMBER, CHARMM). However, terms for desolvation are not force field-based, and therefore it has been suggested that the term physics-based is more appropriate to encompass these scoring functions.

Empirical scoring functions (e.g., GOLD-ChemScore[14], PLANTS-ChemPLP[22], Glide-GlideScore) decompose binding affinities into several different energy terms and use experimental binding affinities to determine the contribution of each term to the overall binding score. The functional form of empirical terms is often simpler and more intuitive than physics-based scoring functions. For example, the hydrogen bond term in GOLD-ChemScore uses a simple block function to assign a value between 0 and 1 to each hydrogen bond interaction based on deviations from ideal geometry. The weighted contribution of all hydrogen bond interactions (i.e. the hydrogen bond weighting coefficient) is then determined by regression analysis using co-crystallized structures with known binding affinities.

Knowledge-based scoring functions (e.g., GOLD-ASP[23]) use structural data from experimental 3-dimensional receptor-ligand structures to generate atom pair interaction types and use the relative

frequency of these interactions to determine their energetic contributions to binding affinity. Knowledge-based scoring functions operate under the theory that interactions which occur more frequently (relative to their frequency in the reference state) represent energetically more favourable interactions. Unlike empirical scoring functions, knowledge-based scoring functions are derived solely from statistical analyses of 3-dimensional structural data and do not rely on experimental binding affinities.

Machine learning (ML)-based scoring functions (e.g., RFScore,[24] NNScore[25]) predict binding affinity using "features" (or "descriptors"). Features can encompass a broad range of target/ligand properties, such as intermolecular interactions (e.g., hydrogen-bonding, hydrophobic interactions), geometric properties (e.g., shape complementarity), and ligand properties (e.g., molecular weight, number of rotatable bonds). Although the features in a ML-based scoring function may be similar to the terms in empirical scoring functions, they differ in that features are selected using ML-based techniques (i.e. inferred from experimental data) as opposed to pre-determined based on medicinal chemistry principles. Additionally, ML-based scoring functions usually contain significantly more terms than empirical scoring functions, and the contribution of terms to overall binding affinity is determined using machine learning algorithms (e.g., random forest, gradient boosting).

Physics-based

$$Score = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{solv}$$

Empirical

$$Score = c_o + c_1 S_{H-bonds} + c_2 S_{lipo} + c_3 S_{metal} + c_4 S_{rot}$$

Knowledge-based

$$Score = \sum_i^{lig} \sum_j^{prot} -k_B T ln\left[\frac{\rho_{ij}(r)}{\rho^*_{ij}}\right]$$

Machine Learning-based

$$Score = f(features)$$

Figure 1.2. Classification of scoring functions and simplifications of their forms.

Categorizing scoring functions into discrete groups helps describe different approaches to scoring receptor-ligand interactions. However, in reality, many scoring functions blend aspects from multiple categories, blurring these distinctions. For example, the functional form of the vdW, hydrogen bond and electrostatic terms in the AutoDock scoring function are physics-based (i.e. they are described using MM-based equations such as Lennard-Jones potentials and Coulomb's

law). However, the weighted contribution of each term to the overall binding score was determined empirically through regression fitting to experimental data.[26] While most terms in the empirical Glide scoring function make use of a block function (as described above for the hydrogen bonding term in ChemScore), it also includes physics-based terms for vdW and electrostatic interactions.[12, 27]

From a user's perspective, it can be instructive to understand both the functional form of the scoring function terms as well as how their contribution to the overall binding score is determined. This can be particularly helpful in understanding potential limitations or biases in a docking program's scoring function and determining whether that might pose a significant issue for your target and/or ligand(s) of interest. Even within a single docking program, several different scoring functions may be available. Therefore, a solid understanding of how a scoring function describes specific binding interactions may be critical for deciding which function to select. For example, the applicability of scoring functions that are parameterized using experimental data (e.g., empirical, knowledge-based and ML-based) are expected to rely more heavily on the selection of training data. These scoring functions may be expected to perform poorly on systems not incorporated in the training set. As physics-based scoring functions (e.g., DOCK) use MM-based equations to calculate the interaction energy between atom pairs, they may be more sensitive to the assignment of atomic properties (e.g., the assignment of partial charges is expected to significantly affect the electrostatic energy term).

For a more detailed review of scoring functions, readers are directed to selected literature sources.[28, 29]

1.2.3   Evaluating the Accuracy of Docking Programs

There are several metrics commonly used to evaluate the accuracy of docking programs. One of the most frequently employed methods is to assess pose prediction accuracy on a set of experimentally determined receptor-ligand complexes by comparing a ligand's predicted pose to the experimentally observed conformation. A cut-off root mean square deviation (RMSD) value is selected (typically RMSD < 2 Å) and the percentage of receptor-ligand complexes with a predicted pose below that cut-off is determined. These evaluations fall into two categories: self-docking and cross-docking. In self-docking, ligands are docked to their own co-crystallized receptor structure whereas cross-docking experiments use receptor structures that have been co-

crystallized with a different ligand. Differences in receptor conformations are often observed when they are bound to different ligands, as receptors themselves have some degree of flexibility and can bind ligands through induced fit or conformational selection mechanisms. For this reason, self-docking accuracy is typically higher than cross-docking since the binding site is pre-arranged for ligand docking (i.e. any conformational changes in the receptor upon ligand binding are already captured in the crystal structure and do not need to be modeled by the docking program). While self-docking accuracy is commonly reported in the literature, it is important to note that cross-docking accuracy is a more realistic evaluation of a docking program when considering its ultimate application. Docking programs are typically used to predict the binding of novel ligands with unknown binding modes and is therefore expected to be less accurate than what is observed in self-docking experiments. Docking programs will typically perform multiple runs for each ligand, resulting in more than one predicted binding pose. Therefore, self-docking experiments can evaluate accuracy based on the best RMSD across all poses (regardless of their score) or based on the RMSD of the only the best scoring pose. Differences in these values can be instructive, identifying whether the inaccuracy is a result of a *scoring failure* (i.e. a correct pose was generated, but not scored higher than an incorrect binding pose) or a *sampling failure* (the docking program failed to identify a correct pose, even across multiple runs). As a medicinal chemistry user, it is often most informative to compare the accuracy for the best scoring pose because this is most representative of how docking would be applied to virtual screening, (where compounds are typically selected on the basis of their predicted binding score).

Another important attribute for a docking program is its ability to correctly score active compounds highly and to distinguish between active and inactive compounds. This ability to discriminate between active and inactive compounds has been referred to in the literature as a docking program's *screening power*, as it describes how well a program is able to screen a library of compounds to identify active ligands. Correlation coefficients are occasionally employed to assess how well the predicted binding scores correlate with experimental binding affinities, with theoretical absolute values ranging from 0 to 1 (1 indicating perfect correlation, 0 indicating no correlation). Metrics such as enrichment factors (EF) and receiver operating characteristic (ROC) curves have also been devised to evaluate the screening power of docking programs. These metrics look at the ability of a docking program to score known active compounds higher than presumed inactive "decoy" compounds. As shown in Eq 1.1, the enrichment factor looks at how many active

compounds are retrieved within a specific percentage of top scoring compounds, relative to the total number of active compounds present in the entire library.

$$EF_{x\%} = \frac{\#actives_{top\ x\%}/\#compounds_{top\ x\%}}{\#actives_{total}/\#compounds_{total}} \tag{1.1}$$

ROC curves are plots of false positives vs true positives at different threshold values. This demonstrates how many true active compounds are identified by the docking program relative to inactive compounds at varying cutoff docking scores. The area under the ROC curve (AUROC) is a quantitative measure of how well the docking program identifies active compounds by scoring them higher than inactive compounds. A perfect docking program would have an AUROC of 1.0 while identifying active compounds through random selection would result in an AUROC of 0.5 (**Figure 1.3**).
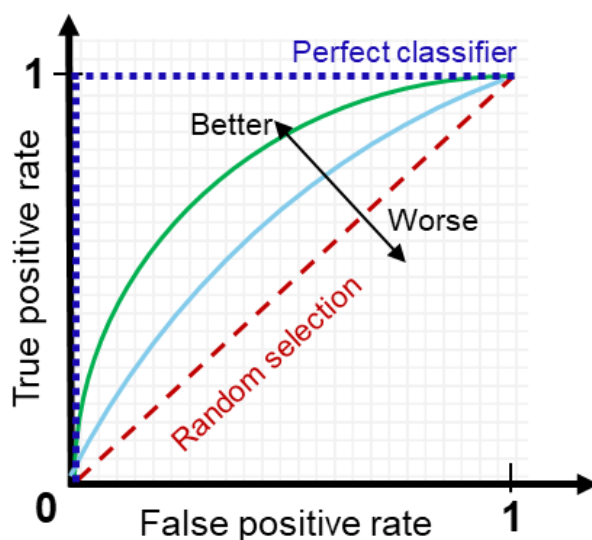
Figure 1.3. Example of ROC curves for random selection (red) and increasing AUROC values (light blue, green, dark blue = perfect).

Variations on AUROC, such as the Boltzmann enhanced discrimination of the ROC curve (BEDROC), have also been devised to quantify the ability of a docking program to discriminate between true active compounds and inactive compounds. BEDROC values emphasize "early recognition" by more heavily weighting the proportion of true positives at the beginning of the

ROC curve.

It is important to note that the observed accuracy of a docking program is specific to the dataset on which it was evaluated and should not be assumed to extend to all targets and ligands equally. It is therefore critical that a user validate the accuracy of a docking protocol using their specific target and/or ligands of interest prior to its application (i.e. in a virtual screening campaign) to get a more realistic sense of its prediction accuracy. For a more detailed description, the reader is directed to several excellent reviews regarding the comparative assessment of docking programs.[30-32]

1.3   Docking to Metalloenzymes

1.3.1   Introduction

It is well established that many proteins require the presence of a metal ion for proper function, with the metal ion serving either a structural role for proper protein folding, or a functional role in the protein's catalytic activity. Proteins that require a metal ion for catalytic function are referred to as metalloenzymes, and it is estimated that nearly 50% of all enzymes fall into this category.[33, 34] A bioinformatic study demonstrated that metalloenzymes make up a large proportion of enzymes across all six major enzyme classifications, with magnesium, zinc, iron and manganese being the most commonly observed ions.[33] Given their prevalence, it is not surprising that metalloenzymes have garnered interest as potential drug targets. Metalloenzymes have shown therapeutic relevance in the treatment of fungal infections (e.g., 14α-demethylase),[35, 36] bacterial infections (e.g., LpxC),[37] cancer (e.g., histone deacetylase - HDAC),[38-40] cardiovascular disease (e.g., angiotensin-converting enzyme - ACE),[41] and HIV (e.g., HIV-1 integrase),[42-44] among other clinical applications. The majority of metalloenzyme-targeting drugs coordinate the active site metal ion via a "metal-binding group", and several metal-binding groups have become well established, such as hydroxamic acids, terminal sulfonamides, carboxylic acids, and triazoles (**Figure 1.4**).
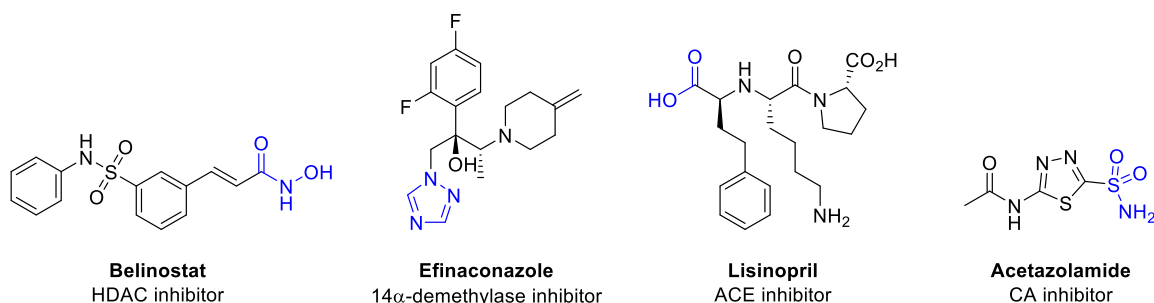
Figure 1.4. Metalloenzyme-targeting drugs, with their metal-binding group shown in blue.

Despite the documented success of metalloenzyme-targeting drugs, metalloenzymes are still considered to be relatively underexplored therapeutic targets. A recent review investigated the number of approved drugs over a 5-year period (2013-2017) and found that only 7% of new drugs targeted metalloenzymes.[45] A larger study looking at all FDA-approved drugs up to 2016 found that metalloenzyme inhibitors accounted for only 4% of approved drugs.[46] Additionally, these studies have demonstrated that the majority of FDA-approved metalloenzyme inhibitors target just three metalloenzymes, with lanosterol 14α-demethylase, ACE, and carbonic anhydrase (CA) accounting for > 80% of metalloenzyme-targeting drugs.

Beyond metalloenzymes, metal ions are also present in the binding site of additional therapeutic targets. For example, the TPP riboswitch is a promising nucleic acid target for the development of antibacterials and features two Mg^{2+} ions in its binding site that are required for ligand binding.[47] Given the central role that metal coordination plays in ligand binding to many relevant therapeutic targets, it is important that computational methods are capable of accurately modeling this interaction.

1.3.2 Challenges

Docking to metalloenzymes presents some unique challenges to established docking programs. Docking programs typically classify interactions as either bonded or non-bonded, with most describing intermolecular interactions by non-bonded terms such as hydrogen-bonding, electrostatics and vdW interactions. However, metal-ligand bonds have both bonded and non-bonded characteristics and may not be accurately modeled by purely non-bonded terms. One challenge to consider is the geometry of binding – metals can form many different geometries depending on the identity of the metal ion and its coordinating residues.[48] Additionally, there can

be a strong electrostatic interaction between a negatively charged ligand atom and a positively charged metal, complicated by the fact that the effective charge of the metal ion depends heavily on its environment and potential charge transfer with coordinating residues. Because of the strength of this interaction, small errors in charge can result in large errors of estimation of the Coulombic interaction energy.

1.3.3 Current Treatment of Metals by Docking Programs

An overview of the current treatment of metals by current docking programs is given in **Table 1.1** and covered in more detail in the sections below.

Table 1.1. Overview of metal treatment by common docking programs.

Program	Optimizations	References
AutoDock	AutoDock3 — optimized zinc parameters using a set of known MMP inhibitors AutoDock4$_{Zn}$ — automatic generation of zinc coordination sites (tetrahedral), zinc-ligand interaction described by MM-based terms for zinc coordination	[49, 50]
GOLD	Automatic generation of metal coordination sites (many geometries) GoldScore — physics-based scoring function, metals treated as h-bond donors & scored based on distance- & angle-dependent weighting of pre-computed values ChemScore — empirical scoring function, metal-ligand interaction described by distance-dependent term for metal coordination	[14, 23, 51]
DOCK	No optimization for metalloenzymes, no consideration of ideal metal coordination geometry GridScore — physics-based scoring function, metal-ligand interaction described by MM-based terms for vdW and electrostatic interactions	[10]
FlexX	Automatic generation of metal coordination surfaces (many geometries) Empirical scoring function, metal-ligand interaction described by distance- and angle-dependent term for metal coordination	[11, 52]
Glide	No consideration of ideal metal coordination sites, but user can implement metal coordination constraints (many geometries) GlideScore — empirical scoring function, metal-ligand interaction described by distance-dependent term for metal coordination (anionic ligands only) and MM-based terms for vdW, and electrostatic interactions	[12]
PLANTS	Automatic generation of metal coordination sites (tetrahedral or octahedral) Empirical scoring function, metal-ligand interaction described by distance- and angle-dependent term for metal coordination	[22]

FITTED	Automatic generation of metal coordination site Physics-based scoring function with MM-based terms for metal coordination	[53, 54]

1.3.3.1 Coordination Geometry

While some metals may display a preference for a specific coordination geometry (e.g., Mg^{2+} is frequently observed in an octahedral geometry), many metals can accommodate several different coordination numbers and geometries.[55, 56] For example, Ca^{2+} is frequently observed with coordination numbers ranging from six to eight, while Zn^{2+} displays a preference for coordination numbers between four and six.[57] Therefore, possible coordination geometries for metal ions can include tetrahedral, trigonal bipyramidal, tetrahedral, and pentagonal bipyramidal, among many others. In order to successfully predict the binding pose of ligands to metalloenzymes, docking program should ideally take into consideration the possible conformational preferences of each metal ion.

Many docking programs model metal-ligand interactions through the use of a metal coordination sphere (**Figure 1.5a**). In this manner, the programs only consider the ideal distance between the ligand and metal while ignoring the conformational preference of the metal ion. In order to address this limitation, several programs have tried to more accurately model metal binding through the use of coordination sites or surfaces to guide ligand docking towards preferred metal coordination geometries (**Figure 1.5b,c**).
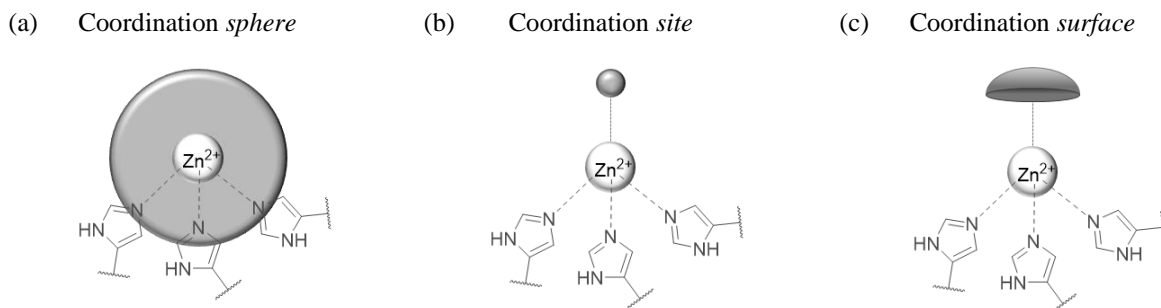
Figure 1.5. Approaches to modelling metal coordination. (a) Coordination *sphere* around the metal (AutoDock3, DOCK, Glide), (b) Coordination *site* (AutoDock4$_{Zn}$, FITTED, GOLD, PLANTS), and (c) Coordination *surface* (FlexX). *Note: some programs use a mixture of these*

approaches depending on the metal and ligand.

For example, while metal coordination geometry is not implicitly considered by the Glide docking program, it includes the option to implement a metal coordination constraint during the docking process.[58] This constraint introduces a requirement for a ligand atom to lie within a specific distance from an optimal metal coordination site for a pose to be considered, thereby taking into consideration the preferred coordination geometry of the metal ion. FITTED generates metal interaction sites based on the vector bond valence postulate (i.e. the sum of all vectors of coordinating atoms must equal zero).[59] These interaction sites are used by the matching algorithm during the conformational search to guide ligand atoms towards metal coordination sites. In the GOLD docking program, the geometry of metal coordination is determined automatically based on the number and angles of metal-coordinating residues and was originally limited to tetrahedral or octahedral geometry.[14] Acceptable geometries have since expanded to include trigonal bipyramidal for zinc ions as well as capped trigonal prism and pentagonal bipyramidal geometries for calcium.[23] The mean ligand-metal contact distances from structures in the CSD were used to generate metal coordination sites at ideal geometries and distances from the metal ion which can interact with specific atom types in a pair-wise manor (**Figure 1.5b**).

A similar approach was taken by the FlexX docking program. They previously utilized an interaction sphere around the metal ion but found it led to pose prediction failures by enabling the ligand to form excessive interactions with the metal ion.[11] To address this, they performed a statistical analysis of the PDB to generate new, more selective metal ion interaction surfaces (e.g., a capped sphere for tetrahedral binding geometries, **Figure 1.5c**).[52] They also updated the interaction surfaces of ligand functional groups to allow for bi-dentate binding. In order for a metal-ligand interaction to be considered, the interaction surface of the ligand must lie on the interaction surface of the metal and vice versa.

The original AutoDock force field does not account for any directionality of metal-ligand interactions and is modeled using an interaction sphere around the metal, with ideal bond distances set by equilibrium van der Waals distances. In 2014, they improved upon this with the development of AutoDock4$_{Zn}$, a new force field for docking to zinc metalloenzymes.[50] They performed a statistical analysis of zinc metalloenzymes in the PDB, and introduced spherical potentials for

atom types with no observed directional preference (sulfur H-bond acceptor, oxygen H-bond acceptor, nitrogen non-H-bond acceptor) with equilibrium distances set by the calculated averages (**Figure 1.5a**). The nitrogen H-bond acceptor atom type showed strong preference for tetrahedral geometry and was modeled through the use of a pseudoatom placed at the ideal tetrahedral geometry that can only form pairwise interactions with this specific atom type (**Figure 1.5b**). The PLANTS docking program uses a spherical potential with no coordination geometry preference for magnesium and calcium, while ideal metal coordination sites are generated for other common metals based on an automatic classification of either tetrahedral or octahedral binding geometry (**Figure 1.5a,b**).[22]

1.3.3.2   Scoring metal-ligand interactions

Many physics-based scoring functions (such as those found in AutoDock, DOCK, and FITTED) treat metal-ligand interactions through distance-dependent MM terms for electrostatic and/or vdW interactions. As shown in **Figure 1.6**, these terms aim to calculate the energy contribution for each metal-ligand contact, often using parameters (e.g., σ, ε, k) from existing force fields.

$$\Delta G_{vdW} = \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right]$$

$$\Delta G_{elec} = k\frac{q_i q_j}{r}$$
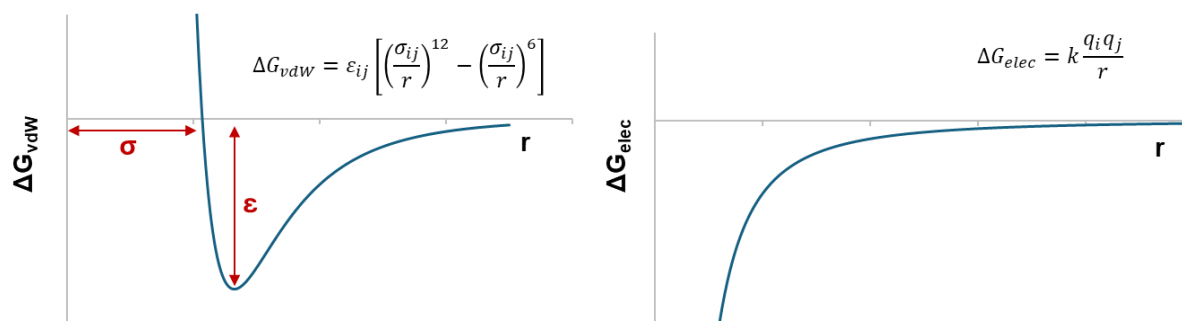
Figure 1.6. MM-based equations for vdW interactions (Lennard-Jones potential, left) and attractive electrostatic interactions (Coulomb potential, right).

The default scoring functions in DOCK and AutoDock account for metal-ligand interactions through the use of MM terms for vdW interactions, electrostatic interactions and desolvation effects. With DOCK, no specific parameterization for metals has been reported and version 6.0 was shown to perform significantly worse for metal-containing proteins.[10, 60] The parameters for zinc ions (e.g., σ, ε, q) were updated in AutoDock3.0 by optimizing the correlation between predicted and experimental binding energies for a set of matrix metalloproteinase inhibitors.[49] For

specific metal ions, FITTED and AutoDock4$_{Zn}$ have been optimized to use a single MM-based term to describe metal-ligand interactions, using a functional form similar to that shown in **Figure 1.6** for vdW interactions. With FITTED, parameters for ligand coordination to Fe, Mg and Zn were determined by fitting to quantum mechanics (QM)-derived energy curves.[53, 54] A more empirical approach was pursued with AutoDock4$_{Zn}$, where zinc coordination parameters were determined using a set of 292 crystal structures with zinc-binding ligands.[50] In the majority of cases, the MM-based terms implemented in DOCK, FITTED and AutoDock do not consider metal coordination geometry when scoring metal-ligand interactions. Notable exceptions to this include the treatment of magnesium with FITTED, which incorporates a periodic function in the Mg-binding term to model its preference for octahedral coordination geometry.[53] Although the metal interaction term for Zn and Fe does not account for metal coordination geometry, FITTED includes a composite scoring function (FITTEDScore) which enhances the physics-based interaction energy by rewarding poses that fulfill metal interaction sites.[59] As described previously, AutoDock4$_{Zn}$ uses a pseudoatom placed at ideal tetrahedral coordination geometry to model zinc coordination by nitrogen H-bond acceptor atom types. For this interaction, the MM-based interaction term depends on the distance to the pseudoatom and not the metal ion, thereby incorporating the preference for tetrahedral coordination geometry.

The majority of empirical scoring functions (e.g., PLANTS, FlexX, ChemScore, Glide) include a term for metal-ligand interactions which are typically described by a weighted pairwise interaction between a metal ion and any ligand atoms capable of forming metal interactions (e.g., metal-acceptor atom types). These often employ a block function to assign a value between 0 and 1 to each metal-ligand interaction based on deviations from ideal distances and angles (**Figure 1.7**). The contribution of metal coordination to the overall binding score (e.g., the value of c_{metal}, **Figure 1.7**) is then determined by fitting to experimental binding affinity data.

$$Score_{metal} = c_{metal} \sum f(\Delta r, \Delta \alpha)$$
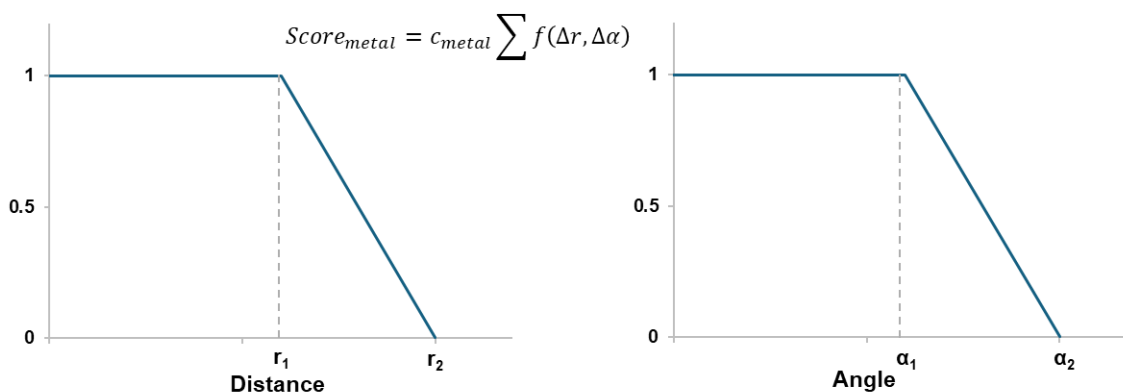
Figure 1.7. An example of distance- and angle-dependent block functions employed by empirical scoring functions (ChemScore, PLANTS, Glide).

The ChemScore scoring function implemented in GOLD uses a weighted block function that depends only on the distance between the metal ion and coordinating ligand atoms, thus ignoring any geometric preference of the metal ion.[14, 61] The scoring function in FlexX includes a metal-ligand binding term that incorporates metal coordination geometry by requiring the metal-coordinating atom to lie on the interaction surface of the metal and vice versa in order to be considered.[52] The metal-ligand interaction is scaled by both distance- and angle-dependent terms to account for deviations from ideal coordination geometries. A similar approach is used by the empirical scoring function implemented in the PLANTS docking program (ChemPLP).[22] Much like the scoring function in FlexX, ChemPLP includes a metal-ligand interaction term that is both distance- and angle-dependent. For magnesium and calcium, the distance-dependence is relative to the metal ion and is therefore akin to a spherical potential with no coordination geometry preference, while for all other metals, the distance-dependent term is relative to the generated ideal coordination site (either tetrahedral or octahedral). In both FlexX and PLANTS, the angle-dependence does not relate to the coordination geometry of the metal but rather of the coordinating ligand atom, requiring the ligand acceptor atom to be oriented appropriately such that electron-donation to the metal ion is possible.

Some docking programs employ a combination of physics-based and empirical approaches. For example, Glide describes metal coordination with both a weighted empirical term (using the same functional form as in ChemScore) as well as with weighted MM-based terms for vdW and electrostatic interactions.[12] In the original GOLD-GoldScore scoring function, metals are treated

as hydrogen bond donors and their interactions with ligands are treated in the same way as hydrogen bonds. Pairwise interactions between hydrogen bond acceptor atoms and metal ions were pre-calculated using gas phase molecular mechanics and this interaction energy is then weighted using a block function to account for any deviation from the ideal bond distances and angles in the docked pose.[19, 51] The bond distance is relative to the generated metal coordination site while the angle refers to the ideal angle of the hydrogen bond acceptor lone pair relative to the metal ion.

1.3.3.3   Water Displacement

Another challenge for scoring metal-ligand interactions is the issue of water displacement. Prior to ligand binding, it is expected that a water molecule will occupy the vacant metal coordination site and therefore ligand binding will involve displacement of this water. Most docking programs do not explicitly account for the energy of metal-bound water displacement as it will be the same for all ligands docked to the same target and will not affect the ranking of metal-binding ligands. However, it may lead to errors with ligands where water displacement may not be favourable, such as with ligands that lack a strong metal-binding group or targets with tightly bound water molecules. For example, self-docking of 1XM6 (phosphodiesterase 4b in complex with potent inhibitor Mesopram) predicted the ligand to coordinate to the zinc ion while the crystal structure shows a tightly bound water molecule that blocks ligand coordination to the metal ion (**Figure 1.8**).[62]
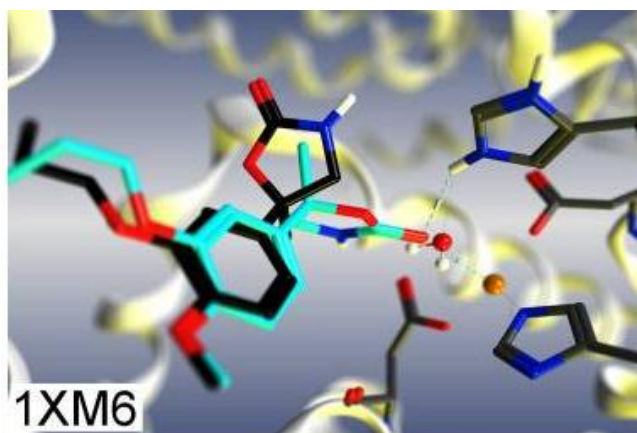
Figure 1.8. Co-crystallized structure of Mesopram (black) compared to the predicted binding pose (teal) from Corbeil *et al.*[62]

To address this issue, the docking program FITTED includes the energy of water displacement and employs an explicit water molecule, enabling ligands to dock in either a coordinated or uncoordinated manner.[54] Although slightly improved accuracies were obtained with implicit treatment of water displacement energy (69% vs 65% pose prediction accuracy), the authors maintained the use of an explicit molecule as they believe it to be a more realistic model of metal coordination.

1.3.3.4 Metal Charges and Ligand Protonation States

Due to the positive charge associated with metal ions, metal-ligand bonding contains a large electrostatic interaction that can be heavily influenced by the charges associated with the metal and the coordinating ligands. This can be a particular issue for scoring functions that describe metal-ligand interactions using MM-based electrostatic terms (such as in AutoDock3, Glide and DOCK). This issue is further complicated by the fact that many metals possess a lower atomic charge than would be expected from their formal charge due to charge transfer between the metal ion and its coordinating residues. For example, the zinc ion in MMP has a formal charge of +2, however calculations have shown that the atomic charge may actually be closer to +1.[63, 64] Similarly, Merz *et al.* found that zinc atomic charges varied from 0.43 to 0.92 depending on the coordinating residues and bound ligands.[65] In order to account for these effects, zinc atomic charges were updated in AutoDock3.0 through regression fitting on a set of MMP-ligand complexes with known binding affinity.[49] The authors found that the best results were obtained when a charge of +0.95 was used for zinc and showed a marked improvement in the accuracy of the zinc-binding group compared to the previously employed charge of +2.0e. In an attempt to recognize the charge transfer effect of coordinating residues, Shoichet *et al.* redistributed 0.2e from each metal-coordinating group to the metal ion, resulting in a charge of +1.4e for zinc.[66] They applied these metal parameters within DOCK3.5, leading to a successful retrospective and prospective virtual screening campaign. Wu *et al.* developed a new force field to describe zinc-ligand interactions by modifying the standard Coulombic pairwise electrostatic term as a means to implicitly account for polarization and charge transfer effects at short-range interactions.[67] This method was later implemented into the scoring function of the MpSDock$_{Zn}$ docking program.[68] An alternative charge-independent method is employed by FITTED and involves the use of a Lennard-Jones potential fit to QM data to describe metal coordination energies.[54] A similar approach is used

by AutoDock4$_{Zn}$, which disables the electrostatic potential for zinc by setting its partial charge to 0 and describes zinc-binding energies using MM-based vdW potentials.[50] Conversely, the empirical scoring functions in GOLD, FlexX and PLANTS contain weighted metal coordination terms that are not explicitly affected by metal charge. However, their contribution to the overall binding score is determined by fitting to experimental affinity data and therefore their accuracy may depend on the diversity of metalloenzymes used to determine the metal coordination weighting coefficients.

In addition to the metal charge, the interaction energy for metal-binding ligands is also expected to depend on the ionization state of the ligand (e.g., anionic ligands are expected to bind more strongly than neutral ligands). This is accounted for with the electrostatic term in physics-based scoring functions, making them sensitive to the assignment of ligand partial charges. Several empirical scoring functions (e.g., PLANTS and Glide) include a separate term or reward for anionic ligands to account for their increased binding strength. However, these terms still treat all anionic ligands equivalently, and are therefore expected to be insensitive to more subtle effects on ligand charges (e.g., electronegativity, inductive, or resonance effects). The consideration of ligand charges is further complicated by the fact that metal coordination can decrease the pKa of the coordinating ligand and facilitate deprotonation of functional groups that are typically neutral at physiological pH.[69, 70] A common approach to this issue has been to dock metal-coordinating groups in their anionic form (e.g., docking sulfonamides as deprotonated sulfonamidates) either in place of or in addition to their neutral form.[66, 71] A more advanced approach to ligand protonation has been employed by FITTED by fitting the metal-binding vdW potential to QM data for both protonated and deprotonated ligand states. Their method accounts for potential proton transfers to adjacent residues and automatically determines the optimal protonation state for binding based on the strength of the coordination and basicity of neighboring residues. They similarly account for the possibility of proton transfer from the coordinated water molecule when determining the energy of water displacement, modelling metal-ligand binding as shown in **Figure 1.9**.[54]
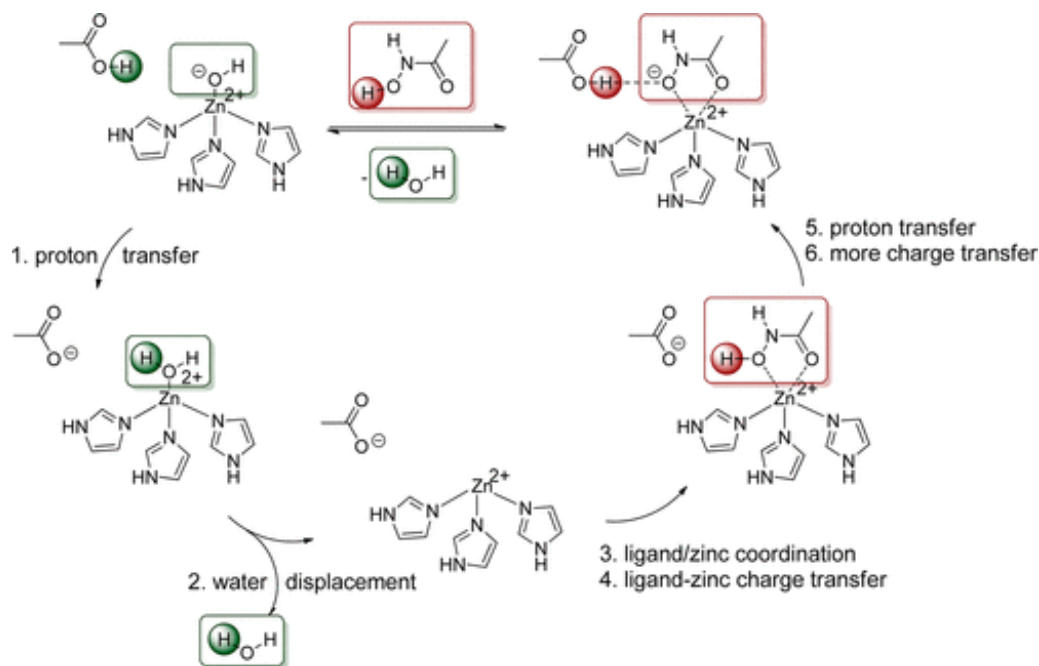
Figure 1.9. Modelling of metal coordination in FITTED[54], which explicitly accounts for water displacement and proton transfer from the ligand to adjacent basic residues (e.g., glutamic acid).

1.3.4   Comparative Assessments of Docking Programs

Despite the prevalence of metal ions in target binding sites, relatively few comparative assessments of their treatment by docking programs have been performed. While metalloenzymes are occasionally included in larger benchmarking sets, they have been the focus of very few evaluations. The reported assessments of select docking programs on metalloenzyme testing sets is summarized in **Table 1.2**. Unless otherwise stated, pose prediction accuracy was determined through self-docking evaluation, focusing on the accuracy of pose prediction and AUROC values, as described in more detail below.

Table 1.2. Assessments of docking programs with metalloenzymes.

Program	Set Size	Metals in Set	AUROC	Success Rate (%)			Reference
				RMSD cutoff	Best Score	Best RMSD	
AutoDock4	213	Ca, Mg, Mn, Zn	0.75[a]	2.0	23	48	[72]
	106	Zn	nd[b]	2.0	32	61	[73]
	292	Zn	nd[b]	2.0	36	nd[b]	[50]
AutoDock4$_{Zn}$	213	Ca, Mg, Mn, Zn	0.74[a]	2.0	18	46 %	[72]
	106	Zn	nd[b]	2.0	31	65 %	[73]
	292	Zn	nd[b]	2.0	45	nd[b]	[50]
DOCK6	213	Ca, Mg, Mn, Zn	0.74[a]	2.0	38	56	[72]
	5	Zn	0.74	nd[b]	nd[b]	nd[b]	[54]
FITTED	237	Fe, Mg, Zn	nd[b]	2.0	60	82	[53]
	5	Zn	0.83	nd[b]	nd[b]	nd[b]	[54]
FlexX	103	Many	nd[b]	2.0	45	83	[52]
Glide	106	Zn	nd[b]	2.0	50	82	[73]
GOLD	106	Zn	nd[b]	2.0	46	73	[73]
LeDock	213	Ca,Mg, Mn, Zn	0.74[a]	2.0	51	77	[72]
PLANTS	213	Ca, Mg, Mn, Zn	0.82[a]	2.0	51	80	[72]
MpSDock$_{Zn}$	106	Zn	nd[b]	2.0	49	90	[73]

[a] AUROC value determined on smaller set. [b] Not determined.

(a) Overall pose prediction

(b) Metal coordination accuracy
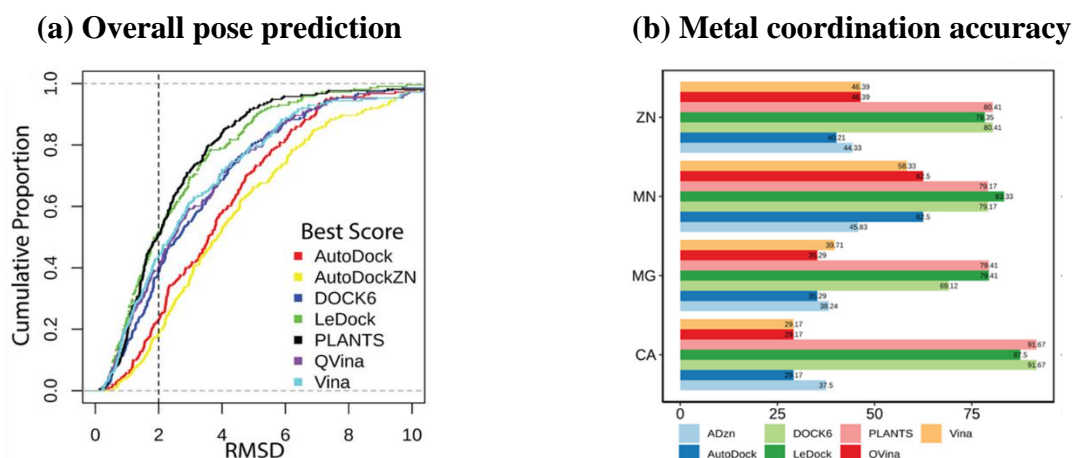
Figure 1.10. a) accuracy of select docking programs on metalloenzyme self-docking pose prediction, and b) comparison of metal coordination accuracy for different metal ions.[72]

The most comprehensive evaluation of docking programs for metalloprotein targets investigated the accuracy of seven programs on a set of 213 metalloprotein complexes featuring Ca, Mg, Mn and Zn ions.[72] As shown in **Figure 1.10a**, the comparative assessment identified PLANTS and LeDock as the most accurate programs for reproducing experimental ligand binding poses that were able to successfully predict the binding pose (RMSD < 2 Å) as the top scoring pose in 51% of cases. They also analyzed posing accuracy for metal coordination by evaluating the RMSD of the closest 3 ligand atoms to the metal ion. Using the same metric for accuracy (top scoring pose with RMSD < 2 Å), they found that PLANTS, LeDock, and DOCK6 were the most accurate at reproducing the metal coordination geometry (**Figure 1.10b**) and had similar accuracies across all four metal ions. The same study also looked at the screening power of the different docking programs using active and decoy ligand sets for three different target proteins for each of the four metal ions. They found that the screening power of PLANTS was the greatest across nearly all metal ion sets with the exception of Zn where it was second behind DOCK6. PLANTS had an average $EF_{1\%}$ between 13 and 22 for all metals (compared to theoretical maximum values of approximately 50) and an average AUROC of 0.82. Surprisingly, the pose prediction accuracy of LeDock did not translate to effective screening power as it had among the lowest $EF_{1\%}$ values across all metal ions. This result highlights the fact that good pose prediction accuracy does not necessarily confer good performance in virtual screening. The average $EF_{1\%}$ values of LeDock

ranged only from 1 to 4, and thus should be cautiously applied to virtual screening campaigns despite its pose prediction accuracy. Despite moderate AUROC and $EF_{1\%}$ values, none of the programs displayed promising correlation coefficients between docking scores and binding affinities, with the average values all falling below 0.2 (indicative of a very weak correlation).

In a benchmarking study, the developers of $MpSDock_{Zn}$ compared its pose prediction accuracy to select docking programs (AutoDock, $AutoDock4_{Zn}$, Glide, GOLD, EADock) on a set of 106 zinc metalloproteins.[73] They found that Glide, GOLD and $MpSDock_{Zn}$ were the most accurate programs for pose prediction, successfully predicting the binding pose as the top scoring pose in 50%, 46% and 49% of cases, respectively. These programs significantly outperformed AutoDock, $AutoDock4_{Zn}$, and EADock which achieved accuracies of 32%, 31%, and 28% respectively on the same test set.

It is interesting to note that in the development of $AutoDock4_{Zn}$, the authors demonstrate its significant improvement in pose prediction accuracy over AutoDock4 (36% to 45%),[50] however these accuracies have not been observed in subsequent evaluations on different testing sets.[72, 73] This may be a result of overtraining the zinc-binding parameters of $AutoDock4_{Zn}$, as the same dataset of 292 structures was used for both parameter calibration and validation. This may have resulted in weighting coefficients that perform well on that specific dataset but are not extendable to broader zinc-ligand complexes.

To assess the metal-binding optimization implemented into FlexX, Seebeck *et al.* evaluated its accuracy on a diverse set of 103 metalloprotein-ligand complexes.[52] They observed an improvement of 6% in pose prediction accuracy over their previous implementation (39% to 45%) when considering only the top scoring pose. However, they obtained very similar accuracies when considering the best RMSD out of all generated poses, suggesting that the improvements in accuracy are the result of improvements in the scoring of metal-ligand interactions.

A recent assessment of FITTED on metalloprotein docking accuracy evaluated its pose prediction accuracy on a set of 237 Zn-, Fe- and Mg-containing complexes.[53] While the average accuracies across the whole dataset were promising (60% and 82% for top-scoring and top-RMSD out of 10 poses, respectively), the results varied significantly between the different metal ions. High accuracies were obtained for zinc-ligand complexes (70%), whereas much lower accuracies were obtained for iron-ligand and magnesium-ligand complexes (55% and 35%, respectively). An

earlier evaluation on the screening power of FITTED on a set of five zinc metalloenzymes obtained an average AUROC value of 0.83, compared to 0.74 for DOCK.[54]

Several examples describing the use of docking methods in the development of metalloenzymes inhibitors have been reported. One of the earliest examples demonstrating the application of docking towards metalloenzymes was reported by Irwin *et al.*[66] DOCK was used to virtually screen a library of fragment-like compounds against zinc-β-lactamase, and experimental evaluation of 15 of the top 50 docking hits led to the identification of 5 compounds with K_i values between 2 and 120 μM. More recently, Choi *et al.* used DOCK to virtually screen a library of 400,000 compounds against mushroom tyrosinase, which contains a di-copper catalytic centre.[74] The top 60 compounds from docking were tested experimentally, leading to the identification of 21 active hits with K_i values between 29 and 35 μM. Pala *et al.* used FlexX and AutoDock4 to successfully identify a novel carbonic anhydrase (hCA II) inhibitor.[75] An initial docking library was assembled by filtering the ZINC database with a pharmacophore model based on key hCA II-ligand interactions. The resulting compounds were first docked to hCA II using FlexX, and the top 29 compounds were further subjected to docking with AutoDock4. The most promising compound was selected for experimental evaluation and displayed micromolar potency (K_i = 9.0 μM). Another virtual screening approach using FlexX led to the identification of a hydrazide-based inhibitor of HDAC8 (IC_{50} = 1.6 μM).[76] Schlimme *et al.* used GOLD to successfully identify an inhibitor for HDAC6.[77] The docking library was assembled by filtering the Maybridge compound library for hydroxamate and hydroxamate-like compounds and docked to HDAC6. The top 5 scoring compounds were tested *in vitro*, with one hydroxamate-containing compound displaying sub-micromolar activity (IC_{50} = 0.3 μM). A virtual screening campaign by Gantner *et al.* using AutoDock4$_{Zn}$ led to the identification of four compounds with nanomolar potency against a human carbonic anhydrase isoform (hCA VII).[78] However, only one of the four active compounds was predicted by docking to coordinate to the active site zinc ion.

1.3.5   Summary and Outlook

As summarized in **Table 1.2**, the accuracy for metalloenzyme pose prediction in self-docking ranges from approximately 30 to 60% depending on the program and metal ion. Although this represents an improvement on early assessments,[71] the accuracy for metalloenzymes is still on average about 5-10 % lower than observed with traditional protein-ligand complexes.[79]

The majority of studies have focused on the accuracy of pose prediction, and further assessments may be necessary to evaluate the ability of docking programs to identify active compounds for metalloprotein targets. The results so far suggest that the top programs can obtain enrichments that are comparable to non-covalent protein targets when evaluating $EF_{1\%}$ and AUROC values, however deeper analysis may be necessary to verify that high enrichments are not a result of biases in the screening set (i.e. inactive compounds without metal-binding groups). Additionally, when comparing correlation coefficients of binding affinity and docking scores, significantly worse results are observed with metalloproteins. These results suggest that while docking programs are relatively successful in modeling coordination geometry to metalloproteins, they still require further optimization in scoring metal-ligand interactions. This could stem from the challenge physics-based scoring functions face in accurately modeling charge transfer effects, as well as the more rudimentary handling of metal-ligand coordination energies by empirical scoring functions. As illustrated in **Figure 1.9**, docking programs could enhance their modeling of metal-ligand coordination by considering factors such as water displacement, charge transfers, and proton transfers.

1.4   Covalent Docking

1.4.1   Introduction

Although roughly 30% of approved drugs with enzyme targets bind their target covalently, covalent inhibition has largely been disfavoured by the pharmaceutical industry and has only recently displayed a re-emerging interest.[80, 81] This seemingly paradoxical statistic stems from the fact that many covalent drugs were discovered serendipitously (e.g., Aspirin and penicillin), with their covalent mode of action identified much later.[82, 83] Concerns about covalent drugs largely stemmed from an increase in awareness about potential toxicity issues of reactive drug metabolites, resulting in a reluctance in drug discovery programs to pursue molecules that contain reactive moieties.[81] Concerns about potential selectivity issues with covalent drugs are often raised, citing the possibility that they may also covalently modify proteins beyond the intended target, resulting in adverse off-target effects.

However, it has been shown that there are many other factors that can influence drug toxicity beyond the presence of a single reactive group in a covalent inhibitor, and covalent inhibitors are capable of achieving remarkable selectivity, even between highly homologous targets.[84]

Additionally, covalent drugs are often more potent than their non-covalent counterparts and have longer residence times due to the strength of covalent bond formation between the ligand and protein target.[85] This allows covalent drugs to be given in lower doses and with decreased dosing frequency, both of which offer advantages in terms of potential off-target toxicity effects and patient compliance.[86, 87] Covalent drugs may also offer advantages in terms of combatting drug resistance or targeting proteins previously considered "undruggable" by more traditional non-covalent binding.[88, 89]

Given the potential advantages of covalent drugs, there has been a re-emerging interest in their development. As the majority of docking programs were developed to model non-covalent protein-ligand interactions, there has also been a corresponding need to develop computational methods to accurately model the binding of covalent inhibitors.

1.4.2   Challenges

Covalent docking presents many unique challenges that must be incorporated into a docking program. Although cysteine accounts for the majority of residues targeted by existing covalent drugs, many additional residues have been shown to be targetable by covalent groups such as serine, lysine, tyrosine, histidine, threonine, glutamic acid, aspartic acid, and methionine.[90] Additionally, many different reactive ligand groups have been shown to covalently modify protein residues, such as aldehydes, Michael acceptors, epoxides, nitriles, boronic acids, among many others.[91] Modeling covalent bond formation of these warheads may result in changes in bond order, protonation states, leaving groups and stereochemistry, all of which need to be accounted for by the docking program in order to accurately model covalent binding. Docking programs typically have terms in their scoring functions to approximate the energy contributions of non-covalent interactions between the ligand and the target. However, in the case of covalent inhibition, a docking program would also ideally consider the energetic contributions of covalent bond formation in the overall docking score. While non-covalent binding is largely driven by thermodynamics, the kinetics of covalent bond formation are expected to be a major factor in the activity of covalent inhibitors.

1.4.3   Current Treatment by Docking Programs

An overview of the current treatment of covalent residues by docking programs is given in **Table**

1.3 and covered in more detail in the sections below.

Table 1.3. Overview of covalent docking programs.

Program	Optimizations	References
AutoDock	Pre-formation of covalent bond based on defined reaction patterns	[92]
CovalentDock	Covalent bond modeled by Morse potential fit to *ab initio* data with empirical correction, allowing ligand to dock either covalently or non-covalently	[93]
CovDock	Two-step docking process based on defined reaction patterns, first docking ligand non-covalently followed by sampling nucleophilic residue rotameric states to generate covalent complex. Covalent ligands scored by averaging non-covalent and covalent complexes	[94, 95]
DOCKovalent	Pre-formation of covalent bond based on pre-generated covalent adduct conformations for specific warhead classes	[96]
FITTED	Complete conformational freedom of covalent ligands, with covalent bond formation only if warhead is positioned sufficiently close to reactive residue. Implementation enables simultaneous screening of different warheads & non-covalent ligands	[53, 97, 98]
FlexX	Pre-formation of covalent bond, uses covalent warhead as base fragment for incremental search algorithm	[11]
GOLD	Pre-formation of covalent bond based on defined reaction patterns	[19]
ICM	Pre-formation of covalent bond based on defined reaction patterns	[99]
MOE	Pre-formation of covalent bond based on defined reaction patterns	
WIDOCK	Employs non-covalent docking of AutoDock with a reactivity-scaled potential (from experimental reaction kinetics data) for poses that place the warhead in close proximity to the reactive residue.	[100]

1.4.3.1   Covalent Bond Formation

Many docking programs predict the binding of covalent inhibitors through pre-formation of the covalent bond. Prior to docking, the ligand is modified to include the side chain atoms of the nucleophile with the corresponding bond order and geometry of the covalently bound state. These

side chain atoms are then used to map the ligand to the nucleophilic residue in the protein binding site during the docking process (**Figure 1.11**). This pre-formation approach is employed by the docking programs GOLD,[19] ICM,[99] and AutoDock's flexible side chain method.[92] A similar method is employed by the program FlexX, which uses the covalent warhead as the base fragment for the incremental construction search algorithm, as well as with the anchor-guided method used by DOCKovalent.[11, 96]
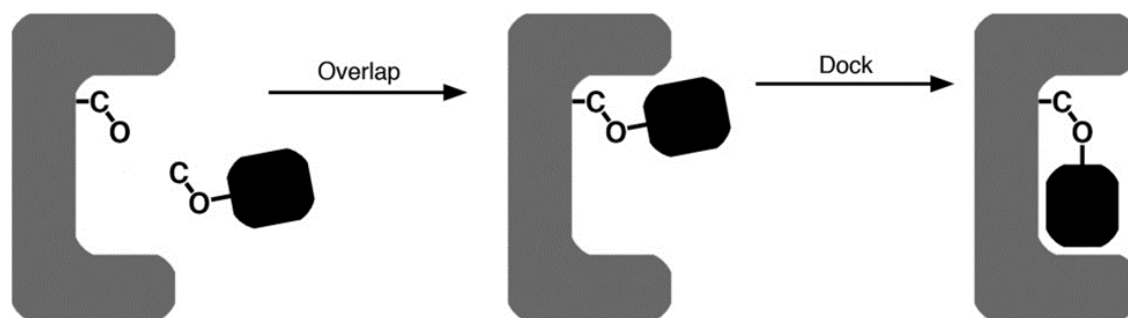
Figure 1.11. Link-atom approach to covalent docking.[92]

A downside to link-atom methods is that it forces the inhibitor to be docked covalently through the explicit pre-formation of the covalent bond. This disregards the possibility that the covalent warhead may not be appropriately positioned relative to the reactive residue to facilitate covalent bond formation and may bind more favourably without covalent bond formation. In order to address this shortcoming, additional docking methods have been devised.

As an alternative to the explicit covalent bond pre-formation employed by many docking programs, the program CovDock instead uses a two-step process to model covalent binding.[94] The ligand is first docked non-covalently, mutating the reactive residue to alanine to avoid side chain steric clashes, and with a constraint that the reactive ligand group must be within a specific distance to the reactive protein residue. The residue is then mutated back to the original residue and a rotamer library is sampled to form a covalent bond with the ligand. After covalent bond formation, bond orders and protonation states are modified as necessary and structurally refined. Covalent ligands are then scored using the averaged scores from both the non-covalent docking step and the covalent complex. By employing a two step process, CovDock accounts for the fact that an optimal

covalent inhibitor must bind favourably both before and after covalent bond formation. Although not as limiting as the direct link-atom approach, the use of a distance restraint during non-covalent docking still restricts the conformational freedom of the ligand and may inaccurately model ligand binding in instances where the most favourable binding pose results in the warhead being further from the reactive residue. Conversely, the covalent docking method implemented in the docking program FITTED allows for complete conformational freedom of the covalent ligand.[53, 97, 98] Ligands are docked non-covalently and covalent bond formation only occurs in poses where the warhead is positioned sufficiently close to the reactive residue (e.g. within 1 Å of the ideal bond length).

1.4.3.2   Binding Energy

As shown in **Figure 1.12**, the binding of covalent inhibitors typically follows a two-step process, with the inhibitor initially forming a non-covalent complex followed by subsequent covalent bond formation. Each step in the binding process has a corresponding energy barrier ($\Delta G^{\dagger}$) and free energy difference (ΔG) that contributes to the potency of a covalent inhibitor. A major limitation to the majority of existing covalent docking programs is that they do not consider the energy barrier for covalent bond formation ($\Delta G^{\dagger}_2$) despite this being a key contributor to covalent inhibitor potency. As $\Delta G^{\dagger}_2$ is expected to be different based on the reactivity of the covalent warhead, it is often stated that docking is capable of distinguishing between inhibitors of the same class but may struggle to accurately assess the potency of scaffolds featuring different covalent warheads.[101] This limitation is largely due to the fact that most docking programs are based on molecular mechanics, and modeling the energy of covalent bond formation would require the use of higher level computational methods like quantum mechanics. Additionally, the majority of covalent docking programs estimate binding affinity based on non-covalent interactions and may only minimally consider the energy contribution of the covalent bond by including torsional and angle-bending terms in the scoring function.[19]
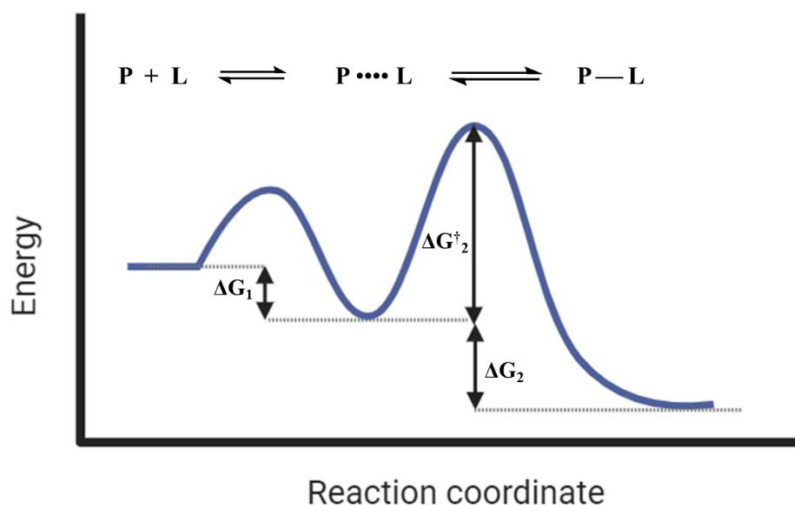
Figure 1.12. Free energy profile for the two-step binding process of covalent inhibitors.

In order to address this, Ouyang *et al.* developed the program CovalentDock, based on the existing docking program AutoDock program.[93] They added a term to the scoring function to consider the energetic contribution of the covalent bond using a Morse potential parameterized to fit QM calculations. This allows the energy of the covalent bond in the docked pose to be included in the overall ligand binding score, taking into account any deviations from ideal covalent bond lengths. While they did observe improved docking accuracy (both in pose prediction and screening power) for the two reactions implemented (serine binding to β-lactams and cysteine binding to Michael acceptors), full implementation of this method would require parameterization of all possible residue-warhead combinations and has not yet been realized.

The docking program WIDOCK aims to accounts for covalent bond formation energy ($\Delta G^{\dagger}_2$) by implementing a reactive docking method first described by Backus *et al.*[100, 102] This method employs non-covalent docking with AutoDock, incorporating a reactivity-scaled reward for poses that place the electrophilic ligand group in close proximity to the nucleophilic residue. The reactivity reward is modeled by a pseudo-Lennard Jones potential parameterized against experimental reaction kinetics data or QM calculations and is applicable to cysteine nucleophiles with a diverse set of electrophilic warheads. This method showed significantly improved screening power over the default covalent docking protocol in AutoDock when screening libraries containing diverse warhead chemotypes.

1.4.3.3 Residue and Warhead Implementation

Current covalent docking programs are now capable of docking to the majority of covalent warheads (carbonyls, Michael acceptors, epoxides, β-lactams, halides and nitriles).[103] Similarly, common nucleophilic residues such as serine, cysteine, threonine and lysine are implemented as potential covalent residues in the majority of docking programs. Many programs model covalent bond formation through the use of pre-defined and user-defined reaction patterns (e.g., GOLD, AutoDock, ICM, CovDock). This enables the user to implement any residue-ligand combination they want and can account for the formation of stereocenters upon covalent binding by generating multiple bound forms to dock. However, the manual identification of protein and ligand reactive groups can create challenges with automating docking for virtual screening applications. In the GOLD docking program, for example, the user selects a covalent bond template for their desired residue and warhead and applies it to dock ligands containing the same defined substructure. In this manner, the docking program is able to covalently dock ligands featuring the same warhead but cannot screen a library of ligands with many different warheads or a library containing both covalent and non-covalent ligands. In order to address this, the docking program FITTED was developed as the first program to implement fully automated docking for covalent ligands. Although it does not allow the user to implement custom warheads, it is able to automatically detect a large range (> 50) of pre-defined covalent groups within a ligand library, allowing libraries featuring many different warheads to be screened simultaneously. Unlike docking programs that utilise a pre-formed covalent bond approach, FITTED only covalently links the ligand when the reactive warhead is sufficiently close to the covalent residue, automatically generating the bound form while accounting for stereochemistry. FITTED is therefore capable of screening libraries containing both covalent and non-covalent ligands simultaneously.

As another approach to address the current limitations in automated covalent docking, David *et al.* developed a computational toolkit in the KNIME platform to enable automated docking of libraries featuring several different covalent warheads with the GOLD docking program.[104]

1.4.4 Assessment of Covalent Docking Programs

Several comparative assessments have been performed on covalent docking programs and are summarized in **Table 1.4**. These assessments have largely focused on cysteine and serine as the

reactive residues while covering a range of warhead chemistries. Unless otherwise stated, pose prediction accuracy was determined through self-docking evaluations.

Table 1.4. Assessments of docking programs with covalent residues.

Program	Set Size	Residues	AUROC	Success Rate (%)			Reference
				RMSD cutoff	Best Score	Best RMSD	
AutoDock4	207	Cys	nd[b]	2.0	55	75	[103]
	76	Cys	0.68[a]	2.0	nd[b]	78	[93]
	4	Cys	0.61	nd[b]	nd[b]	nd[b]	[100]
CovalentDock	76	Cys	0.97[a]	2.0	nd[b]	86	[93]
CovDock (Glide)	207	Cys	nd[b]	2.0	59	75	[103]
	330	Cys, Ser	nd[b]	2.0	57	70	[105]
	405	Cys, Ser	nd[b]	2.0	58	72	[106]
MOE	207	Cys	nd[b]	2.0	37	51	[103]
	330	Cys, Ser	nd[b]	2.0	51	81	[105]
	405	Cys, Ser	nd[b]	2.0	46	67	[106]
FITTED	175	Cys	nd[b]	2.0	56	81	[103]
	207	Cys	nd[b]	2.0	71	90	[53]
	281	Many	nd[b]	2.0	64	76	[53]
GOLD	207	Cys	nd[b]	2.0	53	65	[103]
	330	Cys, Ser	nd[b]	2.0	45	60	[105]
	405	Cys, Ser	nd[b]	2.0	46	62	[106]
ICM	207	Cys	nd[b]	2.0	62	88	[103]
	330	Cys, Ser	nd[b]	2.0	44	62	[105]
	405	Cys, Ser	nd[b]	2.0	52	75	[106]
WIDOCK	4	Cys	0.73	nd[b]	nd[b]	nd[b]	[106]

[a] AUROC value determined on smaller set. [b] Not determined.

Keserű *et al.* performed a comparative assessment of six covalent docking programs (AutoDock4, CovDock, FITTED, GOLD, ICM-Pro and MOE), evaluating their ability to reproduce experimental binding poses on a set of 207 cysteine-bound ligands.[103] They found that the majority of programs were able to achieve 50-60% accuracy in pose prediction when assessing the RMSD of the top scoring pose, using a standard cut-off of 2 Å to classify accurate (< 2 Å) and inaccurate (> 2 Å)

poses, and identified AutoDock4, FITTED and ICM-Pro as "top performers". Additionally, they observed that the accuracy of all programs improved to 65-90% when selecting the best RMSD out of 10 poses. They found that the majority of failures in AutoDock, FITTED and ICM-Pro were due to scoring errors, meaning that these programs were often able to generate the correct pose within 10 docking runs, but the correct pose was not scored as the highest. Conversely, CovDock, GOLD and MOE had the majority of their failures due to sampling errors and were not able to generate a correct pose in these instances, even over 10 docking runs. The authors also noted that most programs performed well with aldehydes, nitriles and Michael addition but struggled to reproduce correct binding poses for epoxide ligands. While this study evaluated in depth the accuracy of pose prediction, it did not include an assessment of scoring accuracy for active and inactive compounds. The authors also note that due to the nature of warhead parameterization (automatic identification by the program, not user-defined), the entire set could not be applied to FITTED at the time of publication as certain warheads and reaction mechanisms had not been incorporated into the program. In a recent optimization of their program, Labarre *et al.* implemented over 50 common warheads into FITTED, improving the accuracy from 56% (on 175 complexes as reported by Keserű) to 71% on the entire Keserű set. They also evaluated the accuracy on an extended set of 281 complexes covering a more diverse range of warheads and covalent residues, achieving an accuracy of 64% for the top-scoring ligand.

A similar study was performed by Wen *et al.* with a dataset of 330 protein-ligand complexes featuring both cysteine and serine as the covalently modified residue. They evaluated four docking programs (MOE, GOLD, CovDock and ICM) and identified CovDock and MOE as the best performing programs when considering the accuracy of the best scoring and best RMSD poses, respectively. They also recognized CovDock as having the greatest generality – it was able to obtain accurate binding poses for diverse warhead chemotypes better than the other three programs.

A recent comparative assessment was performed by Wei *et al*., to compare their hybrid docking/QM method Cov_Dox against several common docking programs.[106] They evaluated the same programs as Wen *et al.* (MOE, GOLD, CovDock and ICM) on a set of 405 cysteine- and serine-bound complexes, obtaining similar accuracies (46%, 46%, 58% and 52%, respectively). They also found that while GOLD, CovDock and ICM performed better with cysteine, MOE

obtained a 20% greater accuracy on complexes with serine as the covalent residue. While Cov_DOX obtained high accuracies on their test set (81% for the top-scoring pose), the high computational cost of the method (approx. 60-100h per run) precludes its use as a true docking program applicable to virtual screening and was therefore not included in this discussion of covalent docking programs.

In the development of CovalentDock, Ouyang *et al.* assessed its pose prediction on a set of 76 cysteine-ligand complexes and obtained an accuracy of 86% when considering the best RMSD pose, outperforming AutoDock and GOLD on this test set. They also evaluated the screening power of CovalentDock on a library of actives and decoys against an EGFR kinase domain. CovalentDock obtained an $EF_{5\%}$ of 17.25 (theoretical max. = 20) and an AUROC of 0.97, significantly outperforming AutoDock in both metrics. Despite the promising results, it is important to note that CovalentDock is only parameterized for Michael addition and lactam warheads, severely limiting its application to diverse warhead chemotypes. The screening power of CovDock was evaluated on four distinct targets using a library of known actives and generated decoys, obtaining an average BEDROC of 0.59.[95] While this value demonstrates a promising ability to identify known active compounds as top scoring compounds, it is important to note that the decoys were limited to compounds containing the same warhead(s) as the active molecules. For three out of four evaluations, the ligands all contained a single warhead type, and no more than three distinct warhead types were considered.

Scarpino *et al.* evaluated their WIDOCK docking protocol, focusing on its screening power against different targets and libraries. Against two targets (MurA and CatB) with a diverse screening library, WIDOCK obtained AUROC values of 0.78 and 0.95, compared to 0.51 and 0.56 with AutoDock. However, when testing against two different targets (OTUB2 and NUDT7) with a library featuring only acrylamide and chloroacetamide warheads, WIDOCK obtained poor AUROC values of 0.54 and 0.63 (compared to 0.74 and 0.64 for AutoDock).

Several examples describing the use of docking methods in the development of covalent inhibitors have also been reported. Virtual screening of a library of compounds containing ketone, nitrile, vinyl sulfone and thiosemicarbazone warheads against cathepsin K was performed with GOLD, leading to the discovery of three novel inhibitors with K_i values in the nanomolar range.[107] Covalent docking to ubiquitin-like poxvirus proteinase I7L was performed using ICM with a

library of ketone- and aldehyde-containing compounds.[99] Experimental evaluation of the top-scoring compounds resulted in a hit rate of nearly 20%, with 12 compounds displaying IC_{50} values below 100 μM. Several reports detailing the application of FITTED towards the discovery and optimization of covalent inhibitors for prolyl oligopeptidase (POP) and fibroblast activation protein α (FAP) have been published, leading to several distinct chemical series with both selective and dual nanomolar inhibition.[97, 98, 108-110] CovDock was successfully applied to a virtual screening campaign of boronic acid-containing compounds against the chymotrypsin-like subunit of the 20S immunoproteasome, resulting in two compounds with low micromolar affinity.[111] CovDock was also used to accurately simulate the binding mode of novel covalent inhibitors of a fibroblast growth factor receptor following incorporation of covalent warheads into a previously identified non-covalent scaffold.[112] DOCKovalent identified two novel covalent K-Ras^{G12C} binders from a virtual library of acrylamine-containing fragments, and the predicted binding pose was used to guide the optimization of an initial hit.[113] CovalentDock was used to screen a library of Michael acceptor- and β-lactam-containing compounds against SARS-CoV-2 3CL^{pro}, leading to the identification of trichostatin A which displayed low micromolar inhibitory activity.[114]

1.4.5   Summary and Outlook

As can be seen from these results, the majority of covalent docking assessments have focused on pose prediction, with comparatively fewer examples of screening power evaluations being reported in the literature. While pose prediction accuracy for covalent docking is comparable to non-covalent docking, the treatment of the covalent bond by many docking programs (e.g., automatic pre-formation, forcing the ligand to dock covalently) greatly reduces the conformational search space and therefore higher accuracies could be expected. However, high accuracies were also obtained with CovDock and FITTED, both of which consider ligand binding in the non-covalent state prior to covalent bond formation. This demonstrates that successful docking results can be obtained without covalent bond pre-formation and may more accurately represent the two-step nature of covalent binding.

Beyond pose prediction accuracy, the ability to discriminate between active and inactive compounds is also an important feature of a docking program, particularly in the context of virtual screening. The general lack of evaluation of covalent docking screening power (e.g., AUROC or EF values) makes it difficult to assess how well these programs could be expected to perform in a

virtual screening campaign. While CovDock reported promising BEDROC values for four covalent targets, their testing set was very limited in terms of warhead diversity, with many of the sets containing only a single warhead type.[95] Similarly, while AutoDock has obtained acceptable AUROC values in several retrospective screening studies, it was shown to perform very poorly on diverse screening libraries.[93, 100] The two programs that attempt to incorporate some consideration of covalent bond energy, WIDOCK and CovalentDock, both outperformed AutoDock in terms of screening power, suggesting that this aspect of covalent binding may need to be explicitly considered by docking programs in order to be applied successfully in virtual screens containing diverse ligands.

1.5   Nucleic Acids

1.5.1   Introduction

Although proteins have historically been favoured as drug targets, studies have shown that only about 2% of the human genome is translated into proteins.[115] Given that nucleic acids play important roles in many aspects of biological function (e.g., gene storage, regulation of gene expression, and protein biosynthesis), targeting disease pathways at the DNA or RNA level significantly increases the number of potential therapeutic targets, and many nucleic acids have been identified as promising drug targets for anti-cancer, antibiotic and antiviral agents. Chemotherapeutic agents often target DNA either through intercalation or major/minor groove binding to induce programmed cell death that inhibits the proliferation of cancer cells. Actinomycin D, for example, is a potent anti-tumour agent that functions by intercalation to double-stranded DNA to form a stable complex that inhibits DNA transcription.[116] Psoralen, a DNA groove binder, has been used to treat skin conditions such as psoriasis and vitiligo.[117] Other DNA structures such as G-quadruplexes have been shown to regulate cell replication through their effect on telomere maintenance, and G-quadruplex binders have been able to induce tumour cell apoptosis, making them attractive targets for anti-cancer therapeutics.[118] Nucleic acid binders have also found applications as antibacterial agents, such as with aminoglycoside antibiotics that act through the binding of ribosomal RNA.[119] Regulatory segments of messenger RNA (mRNA) known as riboswitches have also been identified as a promising new class of bacterial drug targets, leading to the discovery of Ribocil – a highly selective inhibitor of the flavin mononucleotide (FMN) riboswitch with *in vivo* antibacterial activity.[120]
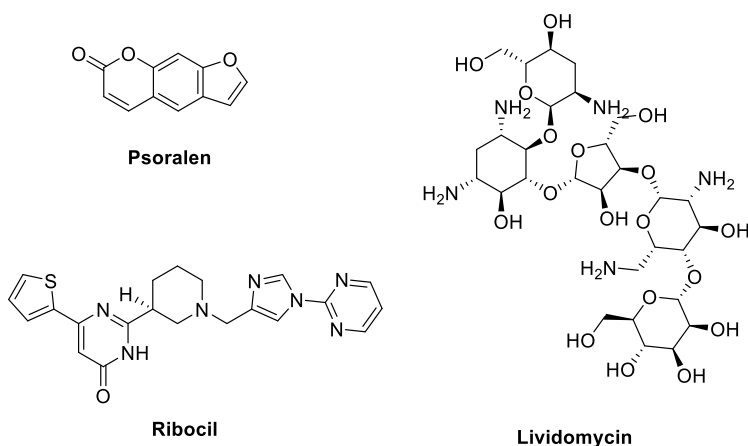
Figure 1.13. Select examples of nucleic acid-binding small molecules.

1.5.2 Challenges

Although binding to nucleic acids follows the same physical principles as binding to proteins and is mediated by many of the same interactions, docking programs developed for proteins may still face some challenges due to differences between nucleic acids and proteins. A major difference between proteins and nucleic acids is that nucleic acids are considerably more polar and solvated than proteins. Nucleic acids contain a negatively charged sugar-phosphate backbone that is often well solvated with water and counterions such as Mg^{2+} and Na^+ that may act to neutralize their charge and enable proper folding.[47] Under physiological conditions, the phosphate ions of nucleic acids are typically neutralized by counterions and water molecules, however these are often too mobile to be captured during the crystallization process and are therefore not represented in the nucleic acid structures employed by docking programs. This can result in artificial energy wells generated by high electrostatic interactions to naked phosphate ions, causing incomplete sampling of ligand conformations and scoring failures. Because of the increased polarity of nucleic acids, their binding sites typically contain more structural and polarized water molecules compared to proteins and are more likely to display water-mediated ligand binding. As an example, many aminoglycosides (see PDB: 2ESJ, Lividomycin or PDB: 1J7T, Paromomycin) have been shown to bind RNA via bridging water molecules. While solvent effects may be accurately modeled by solvent models for protein targets, they may not be applicable to nucleic acid receptors which may require the use of explicit and polarized water molecules.

Another challenge with docking to nucleic acids is accurately modeling the flexibility of the target

macromolecule. The issue of modeling receptor flexibility is not necessarily unique to nucleic acids, as proteins are also known to be flexible receptors that can bind ligands through induced fit mechanisms. However, nucleic acids are generally more flexible than proteins and can undergo large structural changes (i.e. different folding) upon small molecule binding. Therefore, incorrect modeling of the receptor flexibility or selection of the wrong fold may have a more significant impact on docking accuracy for nucleic acids.

1.5.3   Current Treatment of Nucleic Acids by Docking Programs

An overview of the current docking programs optimized for use with nucleic acid targets is given in **Table 1.5** and covered in more detail in the sections below. Several additional programs that were developed for protein-ligand docking have been applied to nucleic acid targets without optimization (e.g., GOLD, Glide) and are not included in the table but are covered in the discussion on comparative assessment of docking programs.

Table 1.5. Docking programs optimized or developed for nucleic acid-ligand docking.

Program	Optimizations/Modification	References
AutoDock4	Used with implicit shielding of phosphate charge and additional solvation parameters for nucleic acids Used with modified weighting parameters determined by training on RNA-ligand complexes Added empirical entropy term based on DNA-ligand complexes	[121-124]
DOCK6	Optimized (implicit solvent, explicit ions + water)	[125]
FITTED	Optimized (implicit shielding of phosphate charge, nucleic acid-specific water placement, introduced nucleotides as residues)	[126]
ICM	Used with modified parameters determined by training empirical scoring function on RNA-ligand complexes	[127]
MORDOR	Developed for nucleic acid docking (incorporates nucleic acid flexibility to model induced fit binding, uses physics-based scoring function)	[128]
rDock	Developed for both nucleic acid and protein docking (empirical scoring function trained on protein-ligand and RNA-ligand complexes)	[129]

RiboDock	Developed for nucleic acid docking (empirical scoring function trained on RNA-ligand complexes)	[130]
RLDock	Developed for nucleic acid docking (hybrid scoring function trained on RNA-ligand complexes)	[131]
NLDock	Developed for nucleic acid docking (knowledge-based scoring function trained on nucleic acid-ligand complexes)	[132]

1.5.3.1   Electrostatic Interactions to Phosphate

As outlined in the metalloenzyme docking section, physics-based scoring functions can encounter difficulties with highly charged species, where an incorrect assignment of effective charge can result in large errors in the calculated electrostatic interaction energy. With regards to nucleic acids, this can result in inaccuracies when modelling electrostatic interactions to backbone phosphate anions as they are neutralized to some extent by the presence of water molecules and counterions. In order to more accurately model electrostatic interactions to phosphate groups, several docking programs have implemented either implicit or explicit methods to account for the neutralization of phosphate charges under biological conditions. In their optimization of the docking program DOCK (developed for proteins), the authors hypothesized that some of the docking failures with nucleic acid targets may be caused by incorrect modeling of the highly charged backbone of RNA. In order to address this issue and optimize the DOCK docking program for use with nucleic acids, they shielded the phosphate charges by adding counterions and explicit water molecules during the docking process. They found that addition of counterions and explicit water molecules enabled more diverse ligand conformations to be generated and applied this method with a more advanced implicit solvent model to significantly improve RNA-ligand docking.[125] An alternative strategy that implicitly accounts for the presence of counterions involves increasing the charge of phosphorous by +1. This approach has been applied by several groups and resulted in a 10% improvement in pose prediction accuracy as observed by Kallert *et al*.[121, 133-135] Another strategy employed by the knowledge-based scoring function ITScoreNL (implemented in the NLDock program) was to model electrostatic interactions with the Debye-Huckel expression instead of the more commonly employed Coulomb potential as the authors consider it to be more representative of effective charges in solution.[136]

1.5.3.2   Binding Site Solvation

Another challenge associated with docking to nucleic acids is the effect of water molecules present in their binding sites. Nucleic acid binding sites are highly polar and solvated, resulting in conserved and polarized water molecules that can play a significant role in ligand binding affinity. Several ligands have been shown to interact with their nucleic acid targets through water-mediated binding (e.g., **Figure 1.14**), and therefore explicit modeling of water molecules may be critical for nucleic acid docking.
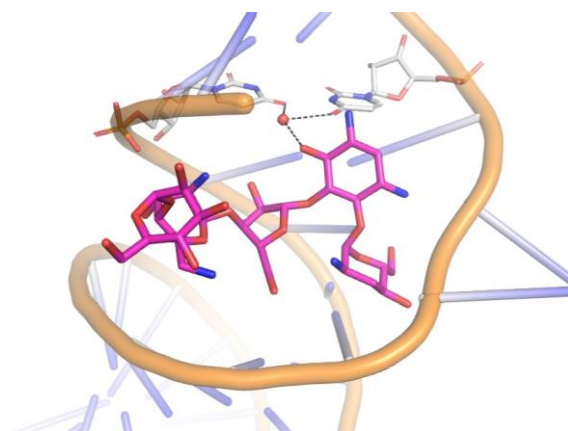
Figure 1.14. Water-mediated binding of Lividomycin (pink) to 16s-rRNA (PDB: 2ESJ).

While it has been shown that including key water molecules in the receptor structure can improve docking accuracy,[134] this is complicated by the fact that key water molecules are expected to be different for different ligands. Including explicit water molecules as a rigid part of the receptor structure reduces the conformational search space of the binding site and can bias virtual screening results towards known ligands. In order to account for explicit water molecules that may differ between ligands, one approach has been to identify key water molecules prior to docking and to treat them as displaceable during the docking process. For example, hydrated docking with FITTED [15], FlexX[137], GOLD[138] and DOCK[139] include explicit water molecules in the receptor structure that can be considered as conserved ("on") or displaced ("off") depending on each individual ligand pose. However, this requires the accurate identification of water placement in the *apo* structure prior to docking and therefore may rely on the availability of high resolution crystal structures with well-resolved water molecules. This also assumes that water molecules do not move in the presence of other ligands. Additionally, this approach typically incorporates a rough estimation of

water binding energies that may lead to challenges when applying values developed for protein receptors to nucleic acids. For example, when the displaceable water protocol of GOLD (originally developed for protein-ligand docking) was applied to a set of hydrated RNA-ligand complexes, it led to a majority of docking failures.[140] This may be a result of the fact that conserved waters in GOLD are assigned a constant entropic penalty that was determined empirically using a set of protein-ligand complexes.[138] However, when an RNA-specific displaceable water potential was developed to describe the binding of aminoglycosides to RNA, it led to improved docking accuracy when implemented into AutoDock.[134] An alternative ligand-centric approach has been employed by AutoDock, with explicit displaceable water molecules considered as part of the ligand. Much like with GOLD, this method was developed for protein-ligand docking and only two examples describing its application to nucleic acids have been reported.[141, 142] As an alternative to displaceable waters, Glide assesses hydration *after* pose prediction, adding explicit water molecules to the ligand-receptor complex.[27] Rewards or penalties are applied based on the nature and number of contacts formed by each water molecule, however these were developed in the context of protein-ligand complexes and their applicability to nucleic acid targets has not been evaluated.

Additional computational methods have been developed to predict water placement and binding energies (such as WaterMap[143], 3D-RISM[144], and JAWS[145], among many others[146]), although they were primarily developed for proteins and their applicability to nucleic acid targets has largely been overlooked. A notable exception is the recent report of a 3D-RISM derived method to predict the placement of water in G-quadruplexes, however only modest accuracies ($< 40\%$) were observed.[147] More promisingly, Wei *et al.* developed a statistical and force-field based method that was able to correctly predict the placement of more than 60% of water molecules in nucleic acid-ligand complexes.[126] This method accounts for the polarization of water molecules and has been incorporated into the FORECASTER platform (which includes FITTED) as part of its receptor preparation for nucleic acid docking, although its impact on docking accuracy has not yet been evaluated.

1.5.3.3   Flexibility of Nucleic Acids

The issue of receptor flexibility is not necessarily unique to nucleic acids, as proteins are known to bind ligands through induced-fit mechanisms that are dependent on protein flexibility. However,

nucleic acids are considerably more flexible than proteins and may adopt several distinct conformations, making the selection of an appropriate structure for docking a key factor in docking accuracy. For example, structural studies of G-quadruplexes have shown that their tertiary structure differs significantly depending on the presence of Na^+ or K^+, and the crystallized form in the presence of K^+ differs from the biologically relevant solution structure.[148] Therefore care should be taken to ensure that the experimentally-derived structure corresponds to the biologically relevant conformation, keeping in mind that nucleic acids may be susceptible to adopting alternative folds depending on the conditions and method of structure determination. The conformational flexibility of nucleic acids also plays a role in the binding of intercalators. As can be seen in **Figure 1.15**, the receptor conformation in the unbound state does not have sufficient space between base pairs to accommodate an intercalating ligand and undergoes a considerable conformational change in the bound state. Therefore, accurate docking of intercalators may require the use of an experimental structure that was obtained in complex with an intercalating ligand or computational methods that can model this flexibility.
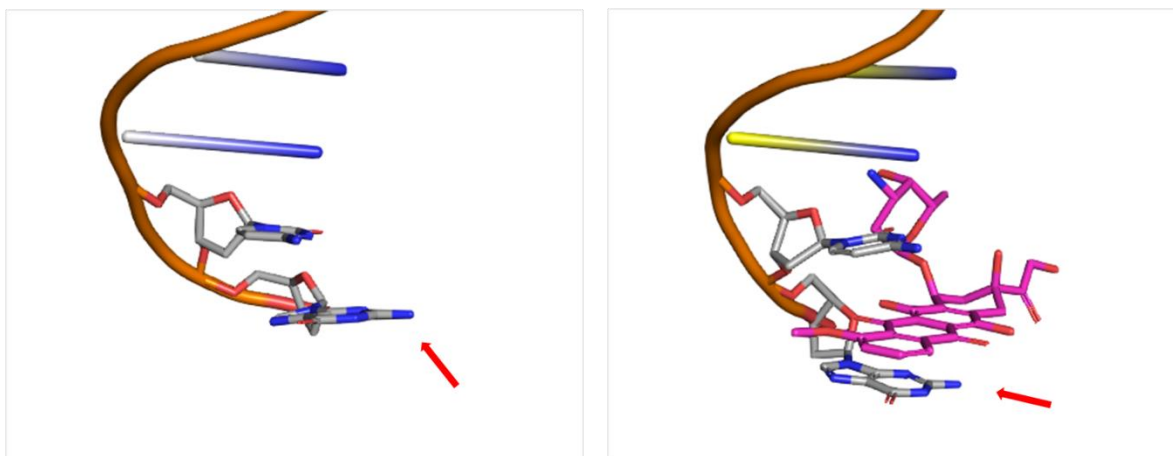
Figure 1.15. Conformational difference between an unligated DNA sequence (left, PDB: 1UQD) and the same sequence bound to intercalating agent Doxorubicin (right, PDB: 151D).

In terms of modelling receptor flexibility, methods such as ensemble docking and soft docking have been developed for proteins and applied to nucleic acid targets with varying success. As an alternative to using only a single receptor structure, ensemble docking involves docking a ligand to an ensemble of discrete receptor conformations or an ensemble-averaged receptor structure. Docking molecules into an ensemble of RNA conformations resulted in a successful virtual

screening campaign against HIV-TAR, leading to the discovery of several high affinity compounds.[149, 150] Similarly, an ensemble of receptor conformations was used to investigate the binding mode of ligands to the RNA TSL2 hairpin.[151] While ensemble docking provides an efficient means of modeling receptor flexibility, there are several limitations to its applicability. Averaged structures are only effective with conformations that are reasonably close, as the averaged structure of two distinct conformations would not represent a biologically relevant (or perhaps even accessible) receptor structure. Additionally, the flexibility that can be captured by ensemble docking is limited to what has been observed experimentally. With respect to the previous example of intercalating agents, if no structure in the ensemble contains an intercalator, then this conformational flexibility will not be modelled by the docking program. It has also been observed that the choice of conformations can be critical to the success of ensemble docking. It has been shown that the number of false positives can increase with increasing receptor conformations, suggesting the importance of using experimentally informed ensembles.[149, 152, 153] As an alternative to ensemble docking, the use of elastic potential grids has been reported to model protein flexibility and was later applied to nucleic acids.[154, 155] This approach enabled the consideration of RNA movements up to 6 Å and was twice as successful as docking to an *apo* RNA structure, although it cannot account for conformational changes involving rotational flip motions and has similar limitations to ensemble docking.[155] Soft docking involves modelling receptor flexibility by "softening" the repulsive scoring term to permit some degree of steric clash between the ligand and target, implicitly accounting for small changes in receptor conformation. Most docking programs with physics-based vdW terms can accommodate soft docking, either by manually modifying keywords in a docking parameter file (e.g., AutoDock, DOCK) or by applying softened potentials to atoms/residues selected through the graphical user interface (e.g., GOLD, Glide). Although soft docking alone cannot account for the larger conformational changes often observed with nucleic acid targets, the use of softened vdW potentials with an averaged receptor structure led to improved docking accuracy of aminoglycosides to ribosomal RNA.[134]

The docking program MORDOR was specifically developed to more accurately account for receptor flexibility in RNA docking.[128] MORDOR identifies a preliminary ligand conformation while initially treating the receptor as rigid. Once an acceptable ligand conformation has been found, energy minimization of the receptor-ligand complex is performed using the "path exploration with distance constraints" algorithm, which treats both the receptor and ligand as

flexible. The algorithm forces the ligand to move to explore the receptor surface and allows the receptor to adjust in response, effectively modelling an induced-fit binding mechanism. However, this method is considerably more computationally intensive compared to other docking approaches. The developers report a runtime ranging from 0.5 to 3 hours per ligand (depending on ligand and receptor size) which may restrict its feasibility for large-scale virtual screening applications.

1.5.3.4   Nucleic Acid-Specific Docking Programs

While binding to nucleic acids is mediated by many of the same interactions as with proteins, differences in polarity and solvation between nucleic acids and proteins may alter the relative influence of these interaction types to the overall binding energy. For example, a statistical analysis of the PDB found that hydrophobic interactions were by far the most frequently observed interactions in protein-ligand complexes while a similar analysis of 800 nucleic acid-ligand complexes found the frequency of hydrogen-bonding interactions to be equivalent to hydrophobic interactions.[156, 157] While these studies may be influenced by the availability of structural data, it is nonetheless reasonable to consider how differences between protein and nucleic acid binding sites might influence the relative importance of different intermolecular interactions. This may pose a challenge to docking programs with empirical scoring terms that have been developed for proteins and trained solely on protein-ligand binding data. Physics-based scoring functions typically lack explicit terms for aromatic interactions such as pi-stacking or cation-pi interactions. Instead, they approximate these interactions using a combination of electrostatic and vdW energies. This approach can lead to inaccuracies in predicting binding geometries and energies, as it may not fully capture the nuances of orbital and induced dipole effects.[158, 159] Considering the prevalence of these interactions in nucleic acid-ligand complexes (e.g., intercalators, G-quadruplex binders), their impact on accuracy might be more pronounced when docking to nucleic acids compared to proteins.

Some empirical parameters originally developed for protein-ligand docking have been trained on nucleic acid-ligand complexes to optimize their contribution to the overall binding score.[122, 123] While the use of optimized parameters led to improved docking accuracy with AutoDock, the parameters were user-implemented and are not included as a default setting within the docking program. Therefore, application of these parameters requires manual modifications by the user

based on values reported in the literature, limiting their user-friendliness. In order to address these issue, several nucleic acid-specific docking programs have been developed. RiboDock was developed based on the empirical scoring function of Bohm,[160] but includes additional RNA-specific terms such as guanidinium-RNA and pi-stacking interactions and was trained only on experimental RNA-ligand complexes.[130] This later evolved into rDock (applicable for both protein and nucleic acid docking), which added a physics-based vdW term and was trained on a dataset of only 15 RNA-ligand and 43 protein-ligand complexes.[129] RLDOCK was developed using a hybrid scoring function. It employs physics-based terms for vdW and electrostatic interactions and an empirical contact term for hydrogen-bonding, with the weighting coefficients for all terms determined empirically using experimental RNA-ligand complexes.[131] As mentioned previously, the program MORDOR was developed to more accurately account for receptor flexibility in RNA docking. After generating several initial ligand poses, MORDOR performs an energy minimization of the RNA-ligand complex, treating both the ligand and receptor as flexible and scoring the final complex with a physics-based scoring function.[128] The nucleic acid-ligand docking program NLDock was developed based on the implementation of the ITScoreNL scoring function into a modified version of the Mdock program.[132] ITScoreNL is a knowledge-based scoring function trained on a set of 213 nucleic acid-ligand complexes with distance-dependent atomic pair potentials as well as explicit potentials for electrostatic and stacking interactions.[136]

While several independent scoring functions for nucleic acids have been developed (e.g., DrugScore[RNA],[135] LigandRNA,[161] SPA-LN,[162] AnnapuRNA[163]), these are standalone scoring functions that are not implemented within a docking program, and therefore are not further included in this discussion.

1.5.4   Assessment of Nucleic Acid Docking

Several comparative assessments have been performed on nucleic acid docking programs and are summarized in **Table 1.6**. Unless otherwise stated, pose prediction accuracy was determined through self-docking evaluations.

Table 1.6. Assessments of nucleic acid-ligand docking.

| Program | NA-Specific | Set Size | AUROC | Success Rate (%) | | | References |
				RMSD cutoff	Best Score	Best RMSD	
AutoDock4	No	56	nd[b]	3.0	nd[b]	30	[123]
		800	nd[b]	2.0	26	61	[157]
DOCK6	No	800	nd[b]	2.0	23	43	[157]
		150	0.71[a]	2.5	35 (45[a])	nd[b]	[133]
FlexX	No	150	0.80[a]	2.5	40	nd[b]	[133]
Glide	No	56	nd[b]	3.0	nd[b]	54	[123]
		800	nd[b]	2.5	22	47	[157]
		60	0.86[a]	2.5	60	nd[b]	[140]
GOLD	No	56	0.87[a]	3.0	nd[b]	73	[123]
		60	0.73[a]	2.5	62	nd[b]	[140]
MORDOR	Yes	57	nd[b]	2.5	74	nd[b]	[128]
PLANTS	No	800	nd[b]	2.0	36	76	[157]
rDock	Yes	56	0.89[a]	3.0	nd[b]	73	[123]
		800	nd[b]	2.0	34	72	[157]
		150	0.89[a]	2.5	50	nd[b]	[133]
RLDock	Yes	800	nd[b]	2.0	8	21	[157]

[a] AUROC value determined on smaller set. [b] Not determined.

An early study by Detering and Varani evaluated the use of DOCK4 and AutoDock3 on a test set of 16 RNA-ligand complexes.[121] AutoDock correctly reproduced the binding mode of the ligand within 2.5 Å in 44% of cases when considering the top 2 scored poses. The same assessment with DOCK gave a lower accuracy of 25%. At the time of publication, neither program had been optimized for nucleic acids and the evaluation served as a proof of principle that protein-ligand docking programs could be applied to nucleic acid targets with moderate success.

A similar study in 2010 investigated the suitability of GOLD (GoldScore) and Glide for nucleic acid docking on a set of 60 RNA-ligand complexes and found they correctly reproduced the binding mode of the ligand within 2.5 Å in 62% and 60% of cases respectively when considering

the top scoring pose.[140] Interestingly, they also found that GOLD and Glide outperformed the RNA-specific programs RiboDock and DrugScore[RNA] in pose prediction, though were less successful than MORDOR. The study also looked at the influence of "dry" versus "wet" docking on a set of 25 hydrated RNA-ligand complexes using Glide. In dry docking, all crystal structure water molecules were removed prior to ligand docking, whereas wet docking kept any water molecules capable of forming hydrogen bond interactions with both the ligand and RNA. They found the accuracy of pose prediction nearly doubled for wet docking compared to dry (44% to 84%). However, it is important to note that identifying key water molecules in self-docking studies is more straightforward than in prospective virtual screening campaigns. Although significant improvements in the accuracy of self-docking was observed, it is unlikely that the inclusion of crystallographic water molecules will have as significant of an impact on cross-docking or virtual screening studies where key water molecules may differ between ligands and cannot be identified prior to docking. In a retrospective virtual screening study against the rRNA A-site, Glide was superior to GOLD in distinguishing between active and inactive compounds, although both provided good enrichments (AUROC values of 0.86 and 0.73, respectively).

A 2012 study by Chen *et al.* evaluated several different docking programs on a set of 56 RNA-ligand complexes.[123] Using a less strict criterion of identifying one pose with RMSD below 3 Å out of the top five scored poses, they obtained the highest success rates for GOLD (GoldScore) and rDock (73% and 74%, respectively). They also evaluated the ability of GOLD and rDock to distinguish between actives and inactives using multiple datasets on two different RNA targets. Both programs performed similarly, resulting in moderate to good AUROC values ranging from 0.58 to 1.0 depending on the target and dataset. Kallert *et al.* performed a comparative assessment of several docking programs (DOCK, FlexX, FRED, HYBRID, LeadIT, rDock) prior to a virtual screening campaign against the Pre-Q1 riboswitch.[133] They assessed the pose prediction accuracy on 150 RNA-ligand complexes as well as their ability to discriminate between active and inactive compounds on 8 distinct riboswitch targets. They observed that the docking programs developed for protein-ligand docking performed similarly well in pose prediction accuracy as rDock, with most falling between 40-49% when considering the top scoring pose (compared to 50% for rDock). While the docking program HYBRID had the highest accuracy (80%), the authors note that its scoring function considers the reference ligand during docking, which could explain the high accuracy observed in self-docking. In the retrospective virtual screening evaluation, DOCK

obtained the lowest average AUROC (0.71), while the highest average was obtained with rDOCK (0.89). The remaining programs all had AUROC values between 0.8-0.87.

A comprehensive evaluation of nucleic acid-ligand docking was recently published by Jiang *et al.,* evaluating the performance of eight docking programs on a dataset of 800 nucleic acid-ligand complexes (**Figure 1.16**).[157] When considering only the top scoring pose and a cut-off RMSD of 2.0 Å, they observed the highest pose prediction accuracy with PLANTS (36%), followed by rDock and LeDock (both 27%). Interestingly, RLDOCK, which was developed specifically for nucleic acid-ligand docking, was the worst performing program and obtained a pose prediction accuracy of only 8%. When considering the best RMSD pose out of all generated poses (100), PLANTS, rDock and LeDock remained as the top performing programs, obtaining pose prediction accuracies of 76%, 72% and 64%, respectively.

The docking program MORDOR obtained very high pose prediction accuracy, as reported in its development.[128] On a set of 57 RNA-ligand complexes, MORDOR was able to achieve a self-docking accuracy of 74% when using a cut-off of 2.5 Å. Impressively, 93% accuracy was obtained when considering the best RMSD pose of the top 10 scoring poses. However, the length of the calculations may have precluded its evaluation by other groups (requiring 0.5 to 3 hours per ligand, as previously stated).
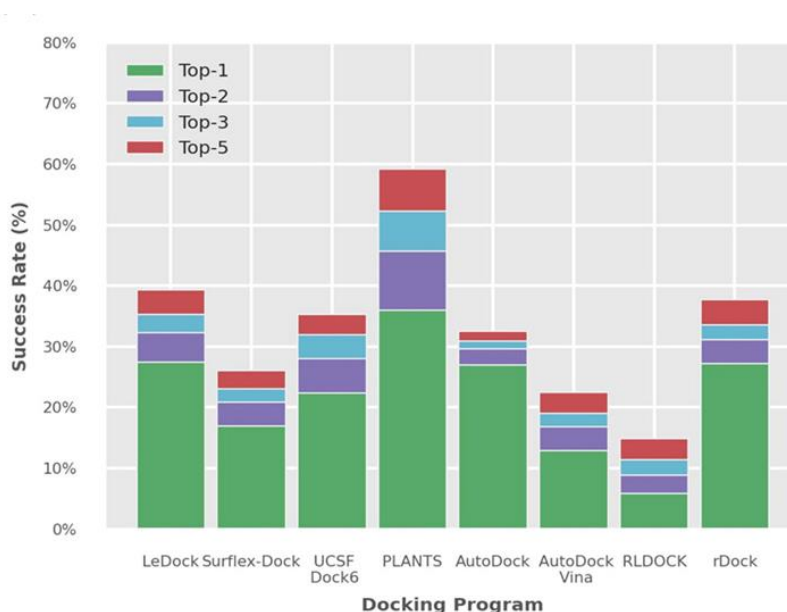
Figure 1.16. Self-docking accuracy of select docking programs against nucleic acid targets.[157]

Several examples describing the use of docking methods in the development of nucleic acid-targeting compounds have been reported. Six novel ligands for HIV-1 TAR were identified by virtual screening against an ensemble of receptor structures using ICM and displayed potent *in vitro* affinities (K_i values from 0.7 to 1.7 μM).[150] Eleven compounds with *in vitro* inhibitory activity against HIV-1 TAR were identified using a virtual screening protocol combining both DOCK and ICM.[127] Preliminary docking poses were obtained with DOCK, followed by minimization and scoring using a modified ICM scoring function (Lind *et al.* implemented an empirical scoring function trained on 13 RNA-ligand complexes in place of the default ICM scoring function). Initial virtual screening with DOCK followed by docking of the top compounds with AutoDock led to the discovery of high affinity ($K_D < 100$ μM) binders to HIV-1 Stem-Loop 3 RNA.[164] In an extension of this work, AutoDock was later applied to the discovery of inhibitors of the HIV-1 Nucleocapsid–Stem-Loop 3 RNA complex.[165]

1.5.5   Summary and Outlook

As shown in **Table 1.6**, the accuracy of nucleic acid-ligand docking pose prediction varies widely across different programs, ranging from a low of 8% with RLDOCK to a maximum accuracy of 74% with MORDOR. On average, the pose prediction accuracy when considering either the best score or best RMSD pose is lower than the values obtained in self-docking assessments of protein-ligand complexes.[79] Additionally, many of these studies were performed using an RMSD cut off greater than 2.0 Å to account for challenges associated with nucleic acid-ligand docking (e.g., increased receptor and ligand flexibility, variability in reference pose with NMR structures). While an adoption of an increased cutoff RMSD is understandable, it is worth noting when comparing these results to protein-ligand docking assessments where a 2.0 Å cutoff is almost universally employed.

The results from the comparative evaluations suggest that docking programs developed for proteins can be successfully extended to nucleic acids. For example, PLANTS, developed for protein-ligand docking, outperformed all other evaluated docking programs on Jiang's test set.[157] rDock was originally developed for RNA-ligand docking but has since been trained on both nucleic acids and proteins. rDock also performed well on Jiang's test set, further demonstrating that a docking program does not have to be limited to nucleic acids to perform well in nucleic acid-ligand docking. Similar results were obtained by Kallert and co-workers.[133] While the best results

were obtained with rDock, several other docking programs developed for protein-ligand docking performed well in both pose prediction and screening power. Additionally, the poor performance of RLDOCK with Jiang's test set demonstrates that although a program has been developed specifically for nucleic acids does not guarantee that it will outperform protein-ligand docking programs.

Even with protein targets, it is advisable to validate a docking method prior to virtual screening campaign through pose prediction assessments or by determining the AUROC on a library of active and decoy compounds. However, this may be especially worthwhile for nucleic acid targets, where the reported accuracy of various docking programs seems to vary considerably between datasets. For example, Glide displayed 60% pose prediction accuracy on a test set of 60 RNA-ligand complexes but was significantly worse (22% accuracy) on Jiang's set of 800 nucleic acid-ligand complexes. This discrepancy in results also serves as a reminder that the availability of nucleic acid-ligand complexes is significantly lower than for protein-ligand complexes. In the PDBbind (a database of biomolecular complexes from the PDB with experimental binding affinities), there is over 100 times as many protein-ligand complexes as nucleic acid-ligand complexes. This may have significant implications in both the development and evaluation of nucleic acid-ligand docking programs. As there are fewer complexes on which to train empirical scoring functions, they may be more susceptible to overtraining. This could lead to decreased accuracy when applied to more diverse targets that were not included in the training data set. Even for physics-based scoring functions that do not use empirical data, less diverse testing sets may be biased with regards to the diversity of observed intermolecular interactions and therefore may not accurately represent the docking program's performance when applied to more diverse nucleic acid targets.

In addition to the relative scarcity of nucleic acid-ligand structural data, the reduced docking accuracy may also be a result of difficulties in modeling nucleic acid-specific interactions and the effect of solvation on binding affinity. Protein-ligand docking programs may be improved by incorporating explicit terms for nucleic acid-specific interactions such as pi-stacking or cation-pi. However, these terms are implemented in the rDock scoring function and yet its pose prediction accuracy did not exceed 35% in Jiang's assessment, suggesting that incorporating these interactions as explicit terms may only modestly improve docking performance. Achieving more significant improvements in docking to nucleic acids may require the implementation of explicit,

polarized water molecules that appropriately model the observed solvent effects in nucleic acid-ligand complexes.

1.6   Conclusions and Perspectives

As summarized in this review, many docking programs have been extended to accommodate metalloenzymes, covalent residues and nucleic acids as potential targets. In order to more accurately model binding to these target classes, several docking programs have been optimized by modifying terms in their scoring functions or optimizing the treatment of key intermolecular interactions. While improvements in docking have been observed as a result of these optimizations, the results summarized in **Table 1.2**, **Table 1.4**, and **Table 1.6** demonstrate that pose prediction and screening accuracy for these targets still fall short when compared to traditional non-covalent protein-ligand docking.

With regards to metalloenzymes, the major challenges lie with scoring of the metal-ligand coordinate bond. Physics-based scoring functions could likely be improved by a more accurate representation of polarization and charge transfer effects that influence the prediction of electrostatic interaction energies, as well as more explicit treatments of water displacement and proton transfers. While docking programs with empirical scoring functions perform reasonably well in terms of pose predictions, they may struggle to differentiate between different metal-binding groups. The nuances of metal-binding energies may require a more advanced treatment of metal-ligand bonding that is not currently achieved by simpler empirical contact terms.

Comparatively, covalent docking has shown pose prediction accuracy that is on par with non-covalent docking. However, many programs are limited by their implementation of covalent bond formation. Pre-formation of the covalent bond forces ligands to dock covalently, disregarding the influence of the initial non-covalent complex on binding affinity. It also results in challenges regarding the automation of covalent docking and virtual screening of diverse warhead classes, with many programs only capable of docking covalent ligands containing a single warhead type at a time. Improvements in the applicability and user-friendliness of covalent docking programs would involve the development of programs that are capable of docking diverse covalent warheads and non-covalent compounds simultaneously. Additionally, incorporating the kinetic aspect of covalent bond formation on binding affinity would also likely lead to significant improvements of docking accuracy, particularly when comparing the predicted binding energies of ligands with

different warhead types. Ideally, covalent docking programs would consider both the affinity of the initial non-covalent complex and the reactivity of the covalent warhead. This would also likely involve some consideration of how the non-covalent complex might influence reactivity based on the positioning of the warhead relative to the reactive residue.

As summarized in **Table 1.6**, the pose prediction accuracy observed when docking to nucleic acids is considerably lower than with protein-ligand docking. Similar to what has been observed with metalloenzymes, the highly charged nature of nucleic acids may benefit from improved treatment of atomic charges, taking into account the effect of polarization and charge screening by water molecules or ions. Given the increased binding site solvation and polarity observed with nucleic acid targets as compared to proteins, nucleic acid-ligand docking may also require more explicit modeling of solvation effects. Ideally, the docking program should account for the desolvation energy of both the target and ligand and allow for water placement that differs based on the nucleic acid-ligand complex. The energetic contribution of explicit water molecules (either from water displacement or for bridging waters) should be specific to the nucleic acid environment and account for polarization and entropic effects. Docking to nucleic acids may also benefit from considerations of user-friendliness, as many of the reported results have employed user-implemented parameters to improve docking accuracy. For a less experienced docking program user (i.e. a medicinal chemist with less computational expertise), parsing out parameters from literature sources and manually implementing them into the docking program may be challenging. Using non-optimized default parameters or incorrectly implementing optimized parameters could lead to reduced docking accuracy and misleading predictions.

As a more general note regarding the discussion of docking performance, users should also be wary of conflating the relative accuracy of a docking program with absolute accuracy. For example, while an increase from 35% to 50% accuracy in pose prediction represents a significant improvement, it is imperative to recognize that an accuracy of 50% implies that a docking program predicts an incorrect binding pose in half of all cases. Additionally, the majority of these evaluations assess self-docking accuracy and not cross-docking accuracy. However, the application of docking programs towards medicinal chemistry projects generally involves docking novel ligands to a target of interest and therefore cross-docking accuracy is a significantly more relevant metric. As cross-docking accuracy is typically lower than self-docking, this must be taken into consideration when using self-docking accuracy to validate a docking protocol. Although the

predicted binding poses and scores generated by a docking program can be useful for prioritizing compounds for experimental evaluation, it is still important to recognize the inherent uncertainty in these predictions. This may be particularly relevant for docking to metalloenzymes, covalent residues and nucleic acids, where the observed accuracy is notably lower.

1.7 References

1.   Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982,** *161* (2), 269-88.

2.   Batool, M.; Ahmad, B.; Choi, S., A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019,** *20* (11).

3.   Pagadala, N. S.; Syed, K.; Tuszynski, J., Software for molecular docking: a review. *Biophys Rev* **2017,** *9* (2), 91-102.

4.   Pinzi, L.; Rastelli, G., Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci* **2019,** *20* (18).

5.   Schuhmacher, A.; Hinder, M.; von Stegmann Und Stein, A.; Hartl, D.; Gassmann, O., Analysis of pharma R&D productivity - a new perspective needed. *Drug Discov Today* **2023,** *28* (10), 103726.

6.   Sousa, S. F.; Ribeiro, A. J.; Coimbra, J. T.; Neves, R. P.; Martins, S. A.; Moorthy, N. S.; Fernandes, P. A.; Ramos, M. J., Protein-ligand docking in the new millennium--a retrospective of 10 years in the field. *Curr Med Chem* **2013,** *20* (18), 2296-314.

7.   Schapira, M.; Abagyan, R.; Totrov, M., Nuclear hormone receptor targeted virtual screening. *J Med Chem* **2003,** *46* (14), 3045-59.

8.   Jain, A. N., Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* **2003,** *46* (4), 499-511.

9.   Lorber, D. M.; Shoichet, B. K., Flexible ligand docking using conformational ensembles. *Protein Sci* **1998,** *7* (4), 938-50.

10.   Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of New Features and Current Docking Performance. *J Comput Chem* **2015,** *36* (15), 1132-1156.

11.   Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996,** *261* (3), 470-89.

12.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K., et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** *47* (7), 1739-49.

13.   Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **1998,** *19* (14), 1639-1662.

14.   Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* **2003,** *52* (4), 609-623.

15.   Corbeil, C. R.; Englebienne, P.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model* **2007,** *47* (2), 435-449.

16.   Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R., Principles of docking: An overview of

search algorithms and a guide to scoring functions. *Proteins* **2002,** *47* (4), 409-43.

17.   Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R., Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol* **2008,** *153 Suppl 1* (Suppl 1), S7-26.

18.   Balius, T. E.; Tan, Y. S.; Chakrabarti, M., DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *J Comput Chem* **2024,** *45* (1), 47-63.

19.   Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** *267* (3), 727-748.

20.   Englebienne, P.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J Chem Inf Model* **2009,** *49* (11), 2564-71.

21.   Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S., A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* **2007,** *28* (6), 1145-52.

22.   Korb, O.; Stützle, T.; Exner, T. E., Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J Chem Inf Model* **2009,** *49* (1), 84-96.

23.   GOLD User Guide. https://www.ccdc.cam.ac.uk/media/Documentation/0C5D99BC-7CC3-49B6-8319-06BEA8CA342D/GOLD_User_Guide_2020_1.pdf (accessed 2024-02-20).

24.   Ballester, P. J.; Mitchell, J. B., A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010,** *26* (9), 1169-75.

25.   Durrant, J. D.; McCammon, J. A., NNScore 2.0: a neural-network receptor-ligand scoring function. *J Chem Inf Model* **2011,** *51* (11), 2897-903.

26.   Hill, A. D.; Reilly, P. J., Scoring functions for AutoDock. *Methods Mol Biol* **2015,** *1273*, 467-74.

27.   Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006,** *49* (21), 6177-96.

28.   Liu, J.; Wang, R., Classification of current scoring functions. *J Chem Inf Model* **2015,** *55* (3), 475-82.

29.   Yan, Z.; Wang, J., Scoring Functions of Protein-Ligand Interactions. In *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery*, Dastmalchi, S.; Hamzeh-Mivehroud, M.; Sokouti, B., Eds. IGI Global: Hershey, PA, USA, 2016; pp 220-245.

30.   Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R., Comparing protein-ligand docking programs is difficult. *Proteins* **2005,** *60* (3), 325-32.

31.   Huang, N.; Shoichet, B. K.; Irwin, J. J., Benchmarking sets for molecular docking. *J Med Chem* **2006,** *49* (23), 6789-801.

32.   Jain, A. N.; Nicholls, A., Recommendations for evaluation of computational methods. *J Comput Aided Mol Des* **2008,** *22* (3-4), 133-9.

33.   Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem* **2008,** *13* (8), 1205-18.

34.   Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009,** *460* (7257), 823-30.

35.   Ellsworth, M.; Ostrosky-Zeichner, L., Isavuconazole: Mechanism of Action, Clinical Efficacy, and Resistance. *J Fungi (Basel)* **2020,** *6* (4).

36.   Monk, B. C.; Sagatova, A. A.; Hosseini, P.; Ruma, Y. N.; Wilson, R. K.; Keniya, M. V., Fungal Lanosterol 14alpha-demethylase: A target for next-generation antifungal design. *Biochim Biophys Acta Proteins Proteom* **2020,** *1868* (3), 140206.

37.   Di Leo, R.; Cuffaro, D.; Rossello, A.; Nuti, E., Bacterial Zinc Metalloenzyme Inhibitors: Recent Advances and Future Perspectives. *Molecules* **2023,** *28* (11).

38.   Moore, D., Panobinostat (Farydak): A Novel Option for the Treatment of Relapsed Or Relapsed and Refractory Multiple Myeloma. *P T* **2016,** *41* (5), 296-300.

39.   Sawas, A.; Radeski, D.; O'Connor, O. A., Belinostat in patients with refractory or relapsed peripheral T-cell lymphoma: a perspective review. *Ther Adv Hematol* **2015,** *6* (4), 202-8.

40.   Tan, J.; Cang, S.; Ma, Y.; Petrillo, R. L.; Liu, D., Novel histone deacetylase inhibitors in clinical trials as anti-cancer agents. *J Hematol Oncol* **2010,** *3*, 5.

41.   Lonn, E. M.; Yusuf, S.; Jha, P.; Montague, T. J.; Teo, K. K.; Benedict, C. R.; Pitt, B., Emerging role of angiotensin-converting enzyme inhibitors in cardiac and vascular protection. *Circulation* **1994,** *90* (4), 2056-69.

42.   Osterholzer, D. A.; Goldman, M., Dolutegravir: a next-generation integrase inhibitor for treatment of HIV infection. *Clin Infect Dis* **2014,** *59* (2), 265-71.

43.   Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J., et al., Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem* **2008,** *51* (18), 5843-55.

44.   Tsiang, M.; Jones, G. S.; Goldsmith, J.; Mulato, A.; Hansen, D.; Kan, E.; Tsai, L.; Bam, R. A.; Stepan, G.; Stray, K. M., et al., Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile. *Antimicrob Agents Chemother* **2016,** *60* (12), 7086-7097.

45.   Chen, A. Y.; Adamek, R. N.; Dick, B. L.; Credille, C. V.; Morrison, C. N.; Cohen, S. M., Targeting Metalloenzymes for Therapeutic Intervention. *Chem Rev* **2019,** *119* (2), 1323-1455.

46.   Yang, Y.; Hu, X. Q.; Li, Q. S.; Zhang, X. X.; Ruan, B. F.; Xu, J.; Liao, C., Metalloprotein Inhibitors for the Treatment of Human Diseases. *Curr Top Med Chem* **2016,** *16* (4), 384-96.

47.   Yamauchi, T.; Miyoshi, D.; Kubodera, T.; Nishimura, A.; Nakai, S.; Sugimoto, N., Roles of Mg2+ in TPP-dependent riboswitch. *FEBS Lett* **2005,** *579* (12), 2583-8.

48.   Kuppuraj, G.; Dudev, M.; Lim, C., Factors Governing Metal−Ligand Distances and Coordination Geometries of Metal Complexes. *The Journal of Physical Chemistry B* **2009,** *113* (9), 2952-2960.

49.   Hu, X.; Shelver, W. H., Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *J Mol Graph Model* **2003,** *22* (2), 115-26.

50.   Santos-Martins, D.; Forli, S.; Ramos, M. J.; Olson, A. J., AutoDock4(Zn): An Improved Auto Dock Force Field for Small-Molecule Docking to Zinc Metalloproteins. *J Chem Inf Model* **2014,** *54* (8), 2371-2379.

51.   Jones, G.; Willett, P.; Glen, R. C., Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* **1995,** *245* (1), 43-53.

52.   Seebeck, B.; Reulecke, I.; Kamper, A.; Rarey, M., Modeling of metal interaction geometries for protein-ligand docking. *Proteins-Structure Function and Bioinformatics* **2008,** *71* (3), 1237-1254.

53.   Labarre, A.; Stille, J. K.; Patrascu, M. B.; Martins, A.; Pottel, J.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 8. Forming New Bonds-Challenges and Opportunities. *J Chem Inf Model* **2022,** *62* (4), 1061-1077.

54.   Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 6. Development and application to the docking of HDACs and other zinc metalloenzymes inhibitors. *J Chem Inf Model* **2014,** *54* (1), 254-65.

55.   Lin, G. Y.; Su, Y. C.; Huang, Y. L.; Hsin, K. Y., MESPEUS: a database of metal coordination groups in proteins. *Nucleic Acids Res* **2024,** *52* (D1), D483-D493.

56.   Tang, S.; Yang, J. J., Magnesium Binding Sites in Proteins. In *Encyclopedia of Metalloproteins*, Kretsinger, R. H.; Uversky, V. N.; Permyakov, E. A., Eds. Springer New York: New York, NY, 2013; pp 1243-1250.

57.   Harding, M. M., The geometry of metal–ligand interactions relevant to proteins. *Acta Crystallographica Section D: Biological Crystallography* **1999,** *55* (8), 1432-1443.

58.   Glide 6.7 User Manual. https://gohom.win/ManualHom/Schrodinger/Schrodinger_2015-2_docs/glide/glide_user_manual.pdf (accessed 2024-02-20).

59.   Corbeil, C. R.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J Chem Inf Model* **2009,** *49* (4), 997-1009.

60.   DOCK 6.11 User Manual. https://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm (accessed 2024-02-20).

61.   Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aid Mol Des* **1997,** *11* (5), 425-445.

62.   Corbeil, C. R.; Williams, C. I.; Labute, P., Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des* **2012,** *26* (6), 775-86.

63.   Kleifeld, O.; Van den Steen, P. E.; Frenkel, A.; Cheng, F.; Jiang, H. L.; Opdenakker,

G.; Sagi, I., Structural characterization of the catalytic active site in the latent and active natural gelatinase B from human neutrophils. *J Biol Chem* **2000,** *275* (44), 34335-43.

64.	Ryde, U., Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion. *Proteins* **1995,** *21* (1), 40-56.

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

66.	Irwin, J. J.; Raushel, F. M.; Shoichet, B. K., Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry-Us* **2005,** *44* (37), 12316-12328.

67.	Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y., A Transferable Non-bonded Pairwise Force Field to Model Zinc Interactions in Metalloproteins. *J Chem Theory Comput* **2011,** *7* (2), 433-443.

68.	Bai, F.; Liao, S.; Gu, J.; Jiang, H.; Wang, X.; Li, H., An accurate metalloprotein-specific scoring function and molecular docking program devised by a dynamic sampling and iteration optimization strategy. *J Chem Inf Model* **2015,** *55* (4), 833-47.

69.	Czarny, B.; Stura, E. A.; Devel, L.; Vera, L.; Cassar-Lajeunesse, E.; Beau, F.; Calderone, V.; Fragai, M.; Luchinat, C.; Dive, V., Molecular determinants of a selective matrix metalloprotease-12 inhibitor: insights from crystallography and thermodynamic studies. *J Med Chem* **2013,** *56* (3), 1149-59.

70.	Cross, J. B.; Duca, J. S.; Kaminski, J. J.; Madison, V. S., The active site of a zinc-dependent metalloproteinase influences the computed pK(a) of ligands coordinated to the catalytic zinc ion. *J Am Chem Soc* **2002,** *124* (37), 11004-7.

71.	Hu, X.; Balaz, S.; Shelver, W. H., A practical approach to docking of zinc metalloproteinase inhibitors. *J Mol Graph Model* **2004,** *22* (4), 293-307.

72.	Cinaroglu, S. S.; Timucin, E., Comparative Assessment of Seven Docking Programs on a Nonredundant Metalloprotein Subset of the PDBbind Refined. *J Chem Inf Model* **2019,** *59* (9), 3846-3859.

73.	Bai, F.; Liao, S.; Gu, J. F.; Jiang, H. L.; Wang, X. C.; Li, H. L., An Accurate Metalloprotein-Specific Scoring Function and Molecular Docking Program Devised by a Dynamic Sampling and Iteration Optimization Strategy. *J Chem Inf Model* **2015,** *55* (4), 833-847.

74.	Choi, J.; Choi, K. E.; Park, S. J.; Kim, S. Y.; Jee, J. G., Ensemble-Based Virtual Screening Led to the Discovery of New Classes of Potent Tyrosinase Inhibitors. *J Chem Inf Model* **2016,** *56* (2), 354-367.

75.	Pala, N.; Dallocchio, R.; Dessi, A.; Brancale, A.; Carta, F.; Ihm, S.; Maresca, A.; Sechi, M.; Supuran, C. T., Virtual screening-driven identification of human carbonic anhydrase inhibitors incorporating an original, new pharmacophore. *Bioorg Med Chem Lett* **2011,** *21* (8), 2515-20.

76.	Zhang, L.; Li, M.; Feng, J.; Fang, H.; Xu, W., Discovery of a novel histone deacetylase 8 inhibitor by virtual screening. *Med Chem Res* **2012,** *21* (2), 152-156.

77.	Schlimme, S.; Hauser, A. T.; Carafa, V.; Heinke, R.; Kannan, S.; Stolfa, D. A.;

Cellamare, S.; Carotti, A.; Altucci, L.; Jung, M., et al., Carbamate prodrug concept for hydroxamate HDAC inhibitors. *ChemMedChem* **2011,** *6* (7), 1193-8.

78.   Gantner, M. E.; Gori, D. N. P.; Llanos, M. A.; Talevi, A.; Angeli, A.; Vullo, D.; Supuran, C. T.; Gavernet, L., Identification of New Carbonic Anhydrase VII Inhibitors by Structure-Based Virtual Screening. *J Chem Inf Model* **2022**.

79.   Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T., Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys* **2016,** *18* (18), 12964-75.

80.   Robertson, J. G., Mechanistic basis of enzyme-targeted drugs. *Biochemistry-Us* **2005,** *44* (15), 5561-71.

81.   Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nat Rev Drug Discov* **2011,** *10* (4), 307-17.

82.   De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A. K.; Moitessier, N., Covalent inhibitors design and discovery. *Eur J Med Chem* **2017,** *138*, 96-114.

83.   Lei, J. P.; Zhou, Y. Z.; Xie, D. Q.; Zhang, Y. K., Mechanistic Insights into a Classic Wonder Drug-Aspirin. *J Am Chem Soc* **2015,** *137* (1), 70-73.

84.   Gan, J. P.; Ruan, Q.; He, B.; Zhu, M. S.; Shyu, W. C.; Humphreys, W. G., In Vitro Screening of 50 Highly Prescribed Drugs for Thiol Adduct Formation-Comparison of Potential for Drug-Induced Toxicity and Extent of Adduct Formation. *Chem Res Toxicol* **2009,** *22* (4), 690-698.

85.   Smith, A. J.; Zhang, X.; Leach, A. G.; Houk, K. N., Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J Med Chem* **2009,** *52* (2), 225-33.

86.   Adeniyi, A. A.; Muthusamy, R.; Soliman, M. E., New drug design with covalent modifiers. *Expert Opin Drug Discov* **2016,** *11* (1), 79-90.

87.   Coleman, C. I.; Roberts, M. S.; Sobieraj, D. M.; Lee, S.; Alam, T.; Kaur, R., Effect of dosing frequency on chronic cardiovascular disease medication adherence. *Curr Med Res Opin* **2012,** *28* (5), 669-80.

88.   Huang, L.; Guo, Z.; Wang, F.; Fu, L., KRAS mutation: from undruggable to druggable in cancer. *Signal Transduct Target Ther* **2021,** *6* (1), 386.

89.   Yun, C. H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K. K.; Meyerson, M.; Eck, M. J., The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A* **2008,** *105* (6), 2070-5.

90.   Jones, L. H., Chapter Four - Design of next-generation covalent inhibitors: Targeting residues beyond cysteine. In *Annual Reports in Medicinal Chemistry*, Ward, R. A.; Grimster, N. P., Eds. Academic Press: 2021; Vol. 56, pp 95-134.

91.   Gehringer, M.; Laufer, S. A., Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J Med Chem* **2019,** *62* (12), 5673-5724.

92.   Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J., Covalent docking using autodock:

Two-point attractor and flexible side chain methods. *Protein Sci* **2016,** *25* (1), 295-301.

93.   Ouyang, X.; Zhou, S.; Su, C. T.; Ge, Z.; Li, R.; Kwoh, C. K., CovalentDock: automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J Comput Chem* **2013,** *34* (4), 326-36.

94.   Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E., Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J Chem Inf Model* **2014,** *54* (7), 1932-40.

95.   Toledo Warshaviak, D.; Golan, G.; Borrelli, K. W.; Zhu, K.; Kalid, O., Structure-based virtual screening approach for discovery of covalently bound ligands. *J Chem Inf Model* **2014,** *54* (7), 1941-50.

96.   London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermancic, P.; Bonnet, R.; Shoichet, B. K., et al., Covalent docking of large libraries for the discovery of chemical probes. *Nat Chem Biol* **2014,** *10* (12), 1066-72.

97.   De Cesco, S.; Deslandes, S.; Therrien, E.; Levan, D.; Cueto, M.; Schmidt, R.; Cantin, L. D.; Mittermaier, A.; Juillerat-Jeanneret, L.; Moitessier, N., Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells. *J Med Chem* **2012,** *55* (14), 6306-15.

98.   Lawandi, J.; Toumieux, S.; Seyer, V.; Campbell, P.; Thielges, S.; Juillerat-Jeanneret, L.; Moitessier, N., Constrained peptidomimetics reveal detailed geometric requirements of covalent prolyl oligopeptidase inhibitors. *J Med Chem* **2009,** *52* (21), 6672-84.

99.   Katritch, V.; Byrd, C. M.; Tseitin, V.; Dai, D.; Raush, E.; Totrov, M.; Abagyan, R.; Jordan, R.; Hruby, D. E., Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J Comput Aided Mol Des* **2007,** *21* (10-11), 549-58.

100.   Scarpino, A.; Petri, L.; Knez, D.; Imre, T.; Abranyi-Balogh, P.; Ferenczy, G. G.; Gobec, S.; Keseru, G. M., WIDOCK: a reactive docking protocol for virtual screening of covalent inhibitors. *J Comput Aided Mol Des* **2021,** *35* (2), 223-244.

101.   Chatterjee, P.; Botello-Smith, W. M.; Zhang, H.; Qian, L.; Alsamarah, A.; Kent, D.; Lacroix, J. J.; Baudry, M.; Luo, Y., Can Relative Binding Free Energy Predict Selectivity of Reversible Covalent Inhibitors? *J Am Chem Soc* **2017,** *139* (49), 17945-17952.

102.   Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; Gonzalez-Paez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J., et al., Proteome-wide covalent ligand discovery in native biological systems. *Nature* **2016,** *534* (7608), 570-4.

103.   Scarpino, A.; Ferenczy, G. G.; Keseru, G. M., Comparative Evaluation of Covalent Docking Tools. *J Chem Inf Model* **2018,** *58* (7), 1441-1458.

104.   David, L.; Mdahoma, A.; Singh, N.; Buchoux, S.; Pihan, E.; Diaz, C.; Rabal, O., A toolkit for covalent docking with GOLD: from automated ligand preparation with KNIME to bound protein-ligand complexes. *Bioinform Adv* **2022,** *2* (1), vbac090.

105.   Wen, C.; Yan, X.; Gu, Q.; Du, J.; Wu, D.; Lu, Y.; Zhou, H.; Xu, J., Systematic Studies on the Protocol and Criteria for Selecting a Covalent Docking Tool. *Molecules* **2019,** *24* (11).

106.   Wei, L.; Chen, Y.; Liu, J.; Rao, L.; Ren, Y.; Xu, X.; Wan, J., Cov_DOX: A Method for Structure Prediction of Covalent Protein-Ligand Bindings. *J Med Chem* **2022,** *65* (7), 5528-5538.

107.   Schroder, J.; Klinger, A.; Oellien, F.; Marhofer, R. J.; Duszenko, M.; Selzer, P. M., Docking-based virtual screening of covalently binding ligands: an orthogonal lead discovery approach. *J Med Chem* **2013,** *56* (4), 1478-90.

108.   Plescia, J.; De Cesco, S.; Patrascu, M. B.; Kurian, J.; Di Trani, J.; Dufresne, C.; Wahba, A. S.; Janmamode, N.; Mittermaier, A. K.; Moitessier, N., Integrated Synthetic, Biophysical, and Computational Investigations of Covalent Inhibitors of Prolyl Oligopeptidase and Fibroblast Activation Protein alpha. *J Med Chem* **2019,** *62* (17), 7874-7884.

109.   Plescia, J.; Dufresne, C.; Janmamode, N.; Wahba, A. S.; Mittermaier, A. K.; Moitessier, N., Discovery of covalent prolyl oligopeptidase boronic ester inhibitors. *Eur J Med Chem* **2020,** *185*, 111783.

110.   Plescia, J.; Hedou, D.; Pousse, M. E.; Labarre, A.; Dufresne, C.; Mittermaier, A.; Moitessier, N., Modulating the selectivity of inhibitors for prolyl oligopeptidase inhibitors and fibroblast activation protein-alpha for different indications. *Eur J Med Chem* **2022,** *240*, 114543.

111.   Scarpino, A.; Bajusz, D.; Proj, M.; Gobec, M.; Sosic, I.; Gobec, S.; Ferenczy, G. G.; Keseru, G. M., Discovery of Immunoproteasome Inhibitors Using Large-Scale Covalent Virtual Screening. *Molecules* **2019,** *24* (14).

112.   Wang, Y.; Dai, Y.; Wu, X.; Li, F.; Liu, B.; Li, C.; Liu, Q.; Zhou, Y.; Wang, B.; Zhu, M., et al., Discovery and Development of a Series of Pyrazolo[3,4-d]pyridazinone Compounds as the Novel Covalent Fibroblast Growth Factor Receptor Inhibitors by the Rational Drug Design. *J Med Chem* **2019,** *62* (16), 7473-7488.

113.   Nnadi, C. I.; Jenkins, M. L.; Gentile, D. R.; Bateman, L. A.; Zaidman, D.; Balius, T. E.; Nomura, D. K.; Burke, J. E.; Shokat, K. M.; London, N., Novel K-Ras G12C Switch-II Covalent Binders Destabilize Ras and Accelerate Nucleotide Exchange. *J Chem Inf Model* **2018,** *58* (2), 464-471.

114.   Wen, L.; Tang, K.; Chik, K. K.; Chan, C. C.; Tsang, J. O.; Liang, R.; Cao, J.; Huang, Y.; Luo, C.; Cai, J. P., et al., In silico structure-based discovery of a SARS-CoV-2 main protease inhibitor. *Int J Biol Sci* **2021,** *17* (6), 1555-1564.

115.   Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C., et al., The transcriptional landscape of the mammalian genome. *Science* **2005,** *309* (5740), 1559-63.

116.   Lo, Y. S.; Tseng, W. H.; Chuang, C. Y.; Hou, M. H., The structural basis of actinomycin D-binding induces nucleotide flipping out, a sharp bend and a left-handed twist in CGG triplet repeats. *Nucleic Acids Res* **2013,** *41* (7), 4284-94.

117.   Wang, M.; Yu, Y.; Liang, C.; Lu, A.; Zhang, G., Recent Advances in Developing Small Molecules Targeting Nucleic Acid. *Int J Mol Sci* **2016,** *17* (6).

118.   Collie, G. W.; Parkinson, G. N., The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem Soc Rev* **2011,** *40* (12), 5867-92.

119.   Vicens, Q.; Westhof, E., Molecular recognition of aminoglycoside antibiotics by

ribosomal RNA and resistance enzymes: an analysis of x-ray crystal structures. *Biopolymers* **2003,** *70* (1), 42-57.

120.   Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A., et al., Selective small-molecule inhibition of an RNA structural element. *Nature* **2015,** *526* (7575), 672-7.

121.   Detering, C.; Varani, G., Validation of automated docking programs for docking and database screening against RNA drug targets. *J Med Chem* **2004,** *47* (17), 4188-201.

122.   Barbault, F.; Zhang, L.; Zhang, L.; Fan, B. T., Parametrization of a specific free energy function for automated docking against RNA targets using neural networks. *Chemometrics and Intelligent Laboratory Systems* **2006,** *82* (1), 269-275.

123.   Chen, L.; Calin, G. A.; Zhang, S., Novel insights of structure-based modeling for RNA-targeted drug discovery. *J Chem Inf Model* **2012,** *52* (10), 2741-53.

124.   McElfresh, G. W.; Deligkaris, C., A vibrational entropy term for DNA docking with autodock. *Comput Biol Chem* **2018,** *74*, 286-293.

125.   Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D., DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009,** *15* (6), 1219-30.

126.   Wei, W.; Luo, J.; Waldispuhl, J.; Moitessier, N., Predicting Positions of Bridging Water Molecules in Nucleic Acid-Ligand Complexes. *J Chem Inf Model* **2019,** *59* (6), 2941-2951.

127.   Lind, K. E.; Du, Z.; Fujinaga, K.; Peterlin, B. M.; James, T. L., Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target TAR RNA. *Chem Biol* **2002,** *9* (2), 185-93.

128.   Guilbert, C.; James, T. L., Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J Chem Inf Model* **2008,** *48* (6), 1257-68.

129.   Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D., rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* **2014,** *10* (4), e1003571.

130.   Morley, S. D.; Afshar, M., Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *J Comput Aided Mol Des* **2004,** *18* (3), 189-208.

131.   Sun, L. Z.; Jiang, Y.; Zhou, Y.; Chen, S. J., RLDOCK: A New Method for Predicting RNA-Ligand Interactions. *J Chem Theory Comput* **2020,** *16* (11), 7173-7183.

132.   Feng, Y.; Zhang, K.; Wu, Q.; Huang, S. Y., NLDock: a Fast Nucleic Acid-Ligand Docking Algorithm for Modeling RNA/DNA-Ligand Complexes. *J Chem Inf Model* **2021,** *61* (9), 4771-4782.

133.   Kallert, E.; Fischer, T. R.; Schneider, S.; Grimm, M.; Helm, M.; Kersten, C., Protein-Based Virtual Screening Tools Applied for RNA-Ligand Docking Identify New Binders of the preQ(1)-Riboswitch. *J Chem Inf Model* **2022,** *62* (17), 4134-4148.

134.   Moitessier, N.; Westhof, E.; Hanessian, S., Docking of aminoglycosides to hydrated and

flexible RNA. *J Med Chem* **2006,** *49* (3), 1023-33.

135.   Pfeffer, P.; Gohlke, H., DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions. *J Chem Inf Model* **2007,** *47* (5), 1868-76.

136.   Feng, Y.; Huang, S. Y., ITScore-NL: An Iterative Knowledge-Based Scoring Function for Nucleic Acid-Ligand Interactions. *J Chem Inf Model* **2020,** *60* (12), 6698-6708.

137.   Rarey, M.; Kramer, B.; Lengauer, T., The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **1999,** *34* (1), 17-28.

138.   Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R., Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **2005,** *48* (20), 6504-15.

139.   Huang, N.; Shoichet, B. K., Exploiting ordered waters in molecular docking. *J Med Chem* **2008,** *51* (16), 4862-5.

140.   Li, Y.; Shen, J.; Sun, X.; Li, W.; Liu, G.; Tang, Y., Accuracy assessment of protein-based docking programs against RNA targets. *J Chem Inf Model* **2010,** *50* (6), 1134-46.

141.   Song, M.; Li, Y.; Gao, R.; Liu, J.; Huang, Q., De novo design of DNA aptamers that target okadaic acid (OA) by docking-then-assembling of single nucleotides. *Biosens Bioelectron* **2022,** *215*, 114562.

142.   Taliani, S.; Pugliesi, I.; Barresi, E.; Salerno, S.; Marchand, C.; Agama, K.; Simorini, F.; La Motta, C.; Marini, A. M.; Di Leva, F. S., et al., Phenylpyrazolo[1,5-a]quinazolin-5(4H)-one: a suitable scaffold for the development of noncamptothecin topoisomerase I (Top1) inhibitors. *J Med Chem* **2013,** *56* (18), 7458-62.

143.   Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A., Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci U S A* **2007,** *104* (3), 808-13.

144.   Kovalenko, A.; Hirata, F., Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem Phys Lett* **1998,** *290* (1-3), 237-244.

145.   Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Prediction of the water content in protein binding sites. *J Phys Chem B* **2009,** *113* (40), 13337-46.

146.   Nittinger, E.; Flachsenberg, F.; Bietz, S.; Lange, G.; Klein, R.; Rarey, M., Placement of Water Molecules in Protein Structures: From Large-Scale Evaluations to Single-Case Examples. *J Chem Inf Model* **2018,** *58* (8), 1625-1637.

147.   Giambasu, G. M.; Case, D. A.; York, D. M., Predicting Site-Binding Modes of Ions and Water to Nucleic Acids Using Molecular Solvation Theory. *J Am Chem Soc* **2019,** *141* (6), 2435-2445.

148.   Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B., Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Research* **2005,** *33* (14), 4649-4659.

149.   Ganser, L. R.; Lee, J.; Rangadurai, A.; Merriman, D. K.; Kelly, M. L.; Kansal, A. D.; Sathyamoorthy, B.; Al-Hashimi, H. M., High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nat Struct Mol Biol* **2018,** *25* (5), 425-434.

150.   Stelzer, A. C.; Frank, A. T.; Kratz, J. D.; Swanson, M. D.; Gonzalez-Hernandez, M. J.; Lee, J.; Andricioaei, I.; Markovitz, D. M.; Al-Hashimi, H. M., Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol* **2011,** *7* (8), 553-9.

151.   Garcia-Lopez, A.; Tessaro, F.; Jonker, H. R. A.; Wacker, A.; Richter, C.; Comte, A.; Berntenis, N.; Schmucki, R.; Hatje, K.; Petermann, O., et al., Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes. *Nat Commun* **2018,** *9* (1), 2032.

152.   Craig, I. R.; Essex, J. W.; Spiegel, K., Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *J Chem Inf Model* **2010,** *50* (4), 511-24.

153.   Xu, M.; Lill, M. A., Utilizing experimental data for reducing ensemble size in flexible-protein docking. *J Chem Inf Model* **2012,** *52* (1), 187-98.

154.   Kazemi, S.; Kruger, D. M.; Sirockin, F.; Gohlke, H., Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. *ChemMedChem* **2009,** *4* (8), 1264-8.

155.   Kruger, D. M.; Bergs, J.; Kazemi, S.; Gohlke, H., Target Flexibility in RNA-Ligand Docking Modeled by Elastic Potential Grids. *ACS Med Chem Lett* **2011,** *2* (7), 489-93.

156.   Ferreira de Freitas, R.; Schapira, M., A systematic analysis of atomic protein-ligand interactions in the PDB. *MedChemComm* **2017,** *8* (10), 1970-1981.

157.   Jiang, D.; Zhao, H.; Du, H.; Deng, Y.; Wu, Z.; Wang, J.; Zeng, Y.; Zhang, H.; Wang, X.; Wu, J., et al., How Good Are Current Docking Programs at Nucleic Acid-Ligand Docking? A Comprehensive Evaluation. *J Chem Theory Comput* **2023,** *19* (16), 5633-5647.

158.   Paton, R. S.; Goodman, J. M., Hydrogen bonding and pi-stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions. *J Chem Inf Model* **2009,** *49* (4), 944-55.

159.   Turupcu, A.; Tirado-Rives, J.; Jorgensen, W. L., Explicit Representation of Cation-pi Interactions in Force Fields with 1/r(4) Nonbonded Terms. *J Chem Theory Comput* **2020,** *16* (11), 7184-7194.

160.   Bohm, H. J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* **1994,** *8* (3), 243-56.

161.   Philips, A.; Milanowska, K.; Lach, G.; Bujnicki, J. M., LigandRNA: computational predictor of RNA-ligand interactions. *RNA* **2013,** *19* (12), 1605-16.

162.   Yan, Z.; Wang, J., SPA-LN: a scoring function of ligand-nucleic acid interactions via optimizing both specificity and affinity. *Nucleic Acids Res* **2017,** *45* (12), e110.

163.   Stefaniak, F.; Bujnicki, J. M., AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLoS Comput Biol* **2021,** *17* (2), e1008309.

164.   Warui, D. M.; Baranger, A. M., Identification of specific small molecule ligands for stem loop 3 ribonucleic acid of the packaging signal Psi of human immunodeficiency virus-1. *J Med Chem* **2009,** *52* (17), 5462-73.

165.   Warui, D. M.; Baranger, A. M., Identification of small molecule inhibitors of the HIV-1

nucleocapsid-stem-loop 3 RNA complex. *J Med Chem* **2012,** *55* (9), 4132-41.

CHAPTER 2:

DEVELOPMENT OF INHIBITORS TARGETING ACTIVATION-INDUCED CYTIDINE DEAMINASE

Contribution of Authors

J.K. Stille synthesized all compounds shown in this chapter and performed all work related to the virtual screening. Quantum mechanics calculations were performed by another graduate student (M. Burai-Patrascu). *In vitro* testing was performed by our collaborators (M. Larijani) at Memorial University of Newfoundland (and later Simon Fraser University).

2.1 Abstract

The work in this chapter describes research aimed at developing an inhibitor for activation-induced cytidine deaminase (AID), a zinc metalloenzyme implicated in the progression of leukemia and lymphoma. A synthetic route was developed to resynthesize a preliminary inhibitor to validate its *in vitro* activity and access potential analogues with improved binding. Discouragingly, the synthesized inhibitor did not display any inhibitory activity, suggesting the initial result was a false positive. A virtual screen was subsequently performed in order to identify novel inhibitory scaffolds, focusing on ligands with zinc-binding groups. Based on the predicted docking poses and scores, 10 compounds were sent for experimental evaluation, however none were active *in vitro*. These results led us to re-evaluate the zinc-binding term implemented in FITTED, which ultimately showed good accuracy with the coordination sphere of AID.

2.2 Introduction

Blood cancers encompass a diverse group of malignancies affecting the hematopoietic system and represent approximately 10% of all cancer diagnoses.[1] These conditions, including leukemia, lymphoma, and myeloma, are characterized by the uncontrolled growth and proliferation of abnormal blood cells. The current average 5-year survival rate for blood cancers is approximately 70%, however the statistics vary greatly depending on the cancer subtype (ranging from approximately 86% for chronic lymphocytic leukemia to as low as 23% for acute myeloid leukemia).[1] The high occurrence of blood cancers and their modest survival rates highlight a clear need for novel therapeutic agents to target these diseases.

2.2.1 AID as a Therapeutic Target

One protein that has been heavily implicated in the pathology of blood cancers is activation-induced cytidine deaminase (AID). AID is a member of the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of cytidine deaminases, mutating deoxycytidine (dC) to deoxyuridine (dU) in hotspot regions of single-stranded DNA. AID is primarily expressed in mature B cells and acts on immunoglobulin genes to introduce minor mutations that promote the diversification of antibodies. However, AID has also been shown to act off-target and has been implicated in the development and progression of several leukemias and lymphomas. It has been

shown to cause the mutation of tumour suppressor genes, as well as the translocation of known oncogenes to regions of less regulated expression.[2, 3] Additionally, its activity has been directly linked to the development of resistance to the chemotherapeutic drug Gleevec.[4] The level of expression of AID is strongly correlated with the prognosis of patients with lymphoma and leukemia and is therefore a promising target for the treatment of these diseases.[5]

2.2.2   Structure and Mechanism of Action

While AID was first identified in 1999,[6] it was not until nearly 20 years later that the first crystal structure was reported.[7] This is largely due to difficulties associated with the purification of AID and its tendency to form aggregates in solution, making it challenging to obtain pure AID at concentrations high enough for X-ray crystallography.[8] However, by modifying the *N*- and C-termini and introducing point mutations on non-conserved surface residues, Qiao *et al.* were able to obtain a crystal structure of biologically active human AID.[7]
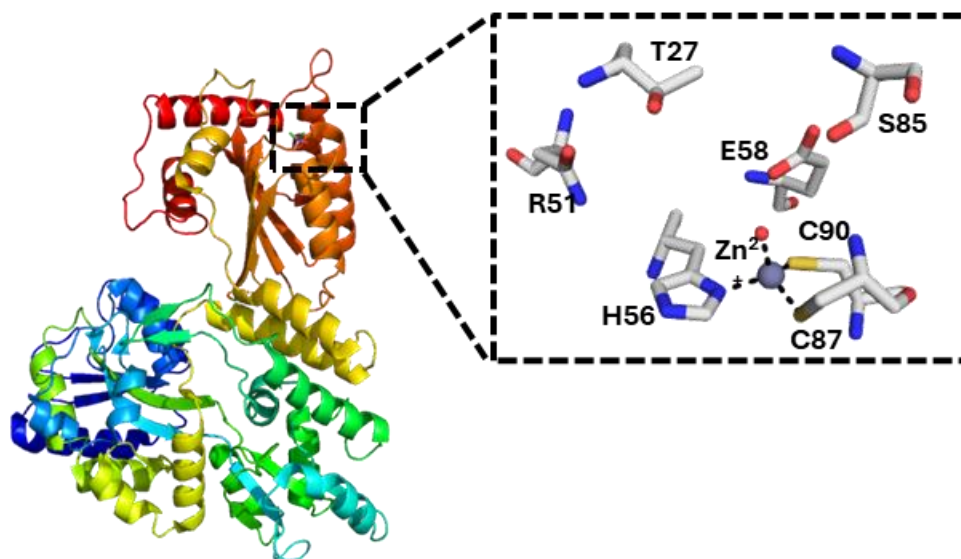

Figure 2.1. Crystal structure of AID (PDB: 5W0Z) and its key active site residues.

The active site of AID contains a zinc ion coordinated in a tetrahedral geometry by a histidine (His[56]) and two cysteine thiolate residues (Cys[87], Cys[90]), with the fourth coordination site occupied by a bound water molecule. AID is a zinc metalloenzyme, with the active site zinc ion playing a critical role in the catalytic activity of AID. The proposed mechanism for AID-mediated

deamination is based on structural studies of cytidine deaminases and shown in **Figure 2.2**.[9-11] It involves the nucleophilic attack of the zinc-bound water on deoxycytidine to form a tetrahedral intermediate which subsequently eliminates ammonia to form deoxyuridine as the product. AID contains an adjacent residue (Glu[58]) that facilitates proton transfers between water and the substrate throughout the reaction mechanism.
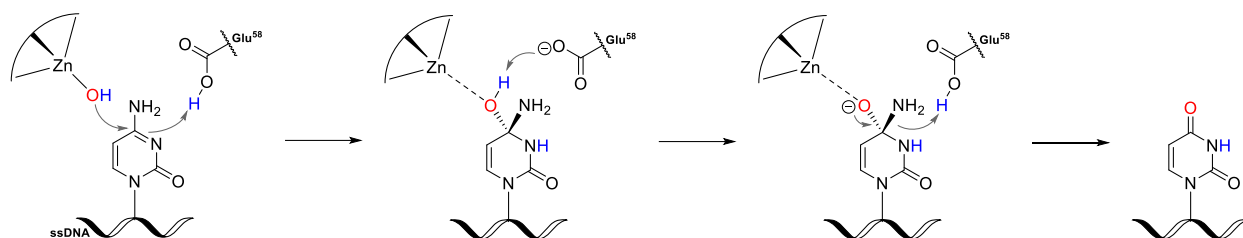

Figure 2.2. Proposed mechanism for AID-catalysed deamination of cytidine to uridine.[11]

2.2.3   Previously Reported Inhibitors

Despite the interest in AID as a promising therapeutic target for the treatment of leukemia and lymphoma, no direct inhibitors have yet been reported. Indirect inhibition of AID was observed with 17-DMAG (**Figure 2.3**) which induces proteasomal degradation of AID via inhibition of heat shock protein (HSP) 90.[12] While 17-DMAG displayed promising activity in an *in vivo* cancer model, HSP90 is known to mediate a wide range of cellular functions and therefore inhibiting HSP90 is anticipated to impact numerous processes beyond just AID expression.[13] Although early reports suggested that AID may be inhibited by known cytidine deaminase inhibitors such as tetrahydrouridine (THU), these results are controversial.[6] It is widely believed that the observed activity was the result of inhibition of a protein contaminant (*E.coli* cytidine deaminase) and not due to AID inhibition.[14, 15]
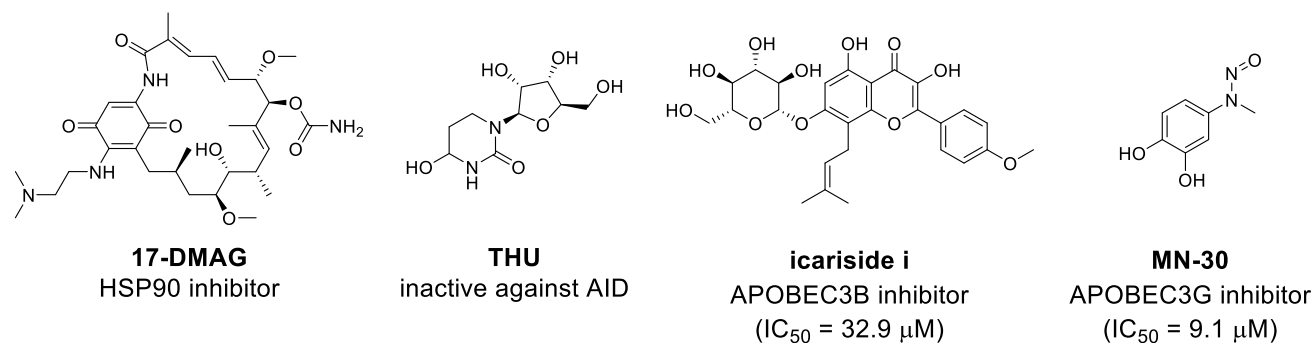
Figure 2.3. Reported inhibitors for AID & other APOBEC proteins.

Even among the APOBEC family, the number of reported inhibitors is scarce. Several small-molecule inhibitors of APOBEC3G have been reported, however they have been shown to act through covalent modification of a cysteine residue specific to APOBEC3G but not observed in other APOBEC proteins.[16, 17] A more recent experimental screening on a library of flavonoids and dihydrochalcones led to the identification of several weakly active APOBEC3B inhibitors.[18]

2.2.4   Computer-Aided Design of Metalloenzyme Inhibitors

The majority of zinc metalloenzyme-targeting drugs bind their target by forming a dative bond to the active site zinc ion, as the strength of a metal-ligand interaction can confer significant potency to an inhibitor.[19] Many zinc-binding groups have been identified, typically featuring a nitrogen, oxygen or sulfur heteroatom capable of forming a dative bond with the metal ion in either a monodentate or bidentate fashion. As shown in **Figure 2.4**, functional groups such as hydroxamic acids, thiols and sulfonamides have found widespread use in the development of metalloenzyme inhibitors. However, many other zinc-binding groups have been developed, often through efforts to expand the toolbox of metal-binding groups and improve their pharmacokinetic properties.[20-23]
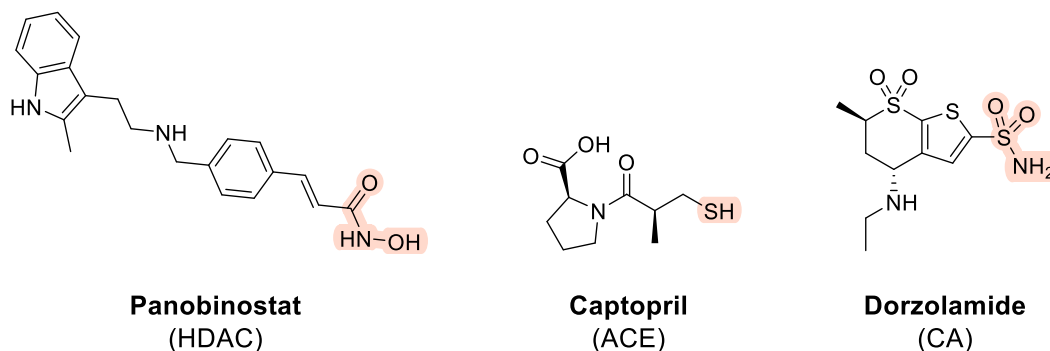
Figure 2.4. Zinc-metalloenzyme inhibitors, with metal-binding groups highlighted.

Given the central role that metal coordination plays in ligand binding to metalloenzymes, it is critical that computational methods are capable of accurately modeling this interaction. As discussed in Chapter 1, this can pose a challenge for docking programs, which were largely developed to predict binding based on non-covalent protein-ligand interactions. Metal-ligand bonds (i.e. coordination) have a partial covalent nature that may not be appropriately modeled by purely non-covalent models. Additionally, the charge on the metal can vary significantly depending on its environment, resulting in large differences in the predicted electrostatic interaction energy. In order to address some of these limitations, we previously optimized our in-house docking program FITTED for use with metalloenzymes, leading to its application in the design of HDAC inhibitors.[24, 25] The optimization included implementing a charge-independent Lennard-Jones potential to describe metal-ligand bonding, as well as considering water displacement and ligand ionization upon metal binding. With these optimizations in mind, it was envisioned that our in-house docking program FITTED could be applied toward the development of AID inhibitors.

2.3 Computer-Guided Optimization of a Potential Hit Compound

Our collaborators at Memorial University of Newfoundland identified a preliminary inhibitor of AID through an independently performed virtual screening campaign and experimental evaluation of a small library of commercially available compounds (**Figure 2.5**). As the preliminary inhibitor **2.1** displayed only very modest inhibitory activity (IC$_{50}$ = 230 μM), the objective of this project was to use FITTED to predict the binding mode of **2.1** and optimize its activity through docking-
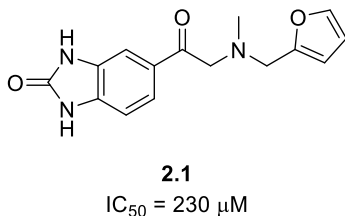
guided design.

2.1

IC$_{50}$ = 230 μM

Figure 2.5. Preliminary inhibitor of AID.

Considering the weak potency which could be resulting from impurities in the tested sample, the first step was to confirm the activity of **2.1** by compound resynthesis, as is standard practice in early drug discovery campaigns.[26] Compound resynthesis validates that the observed biological activity is due to the purported structure while simultaneously developing synthetic methodology that can be used to prepare analogues following validation.
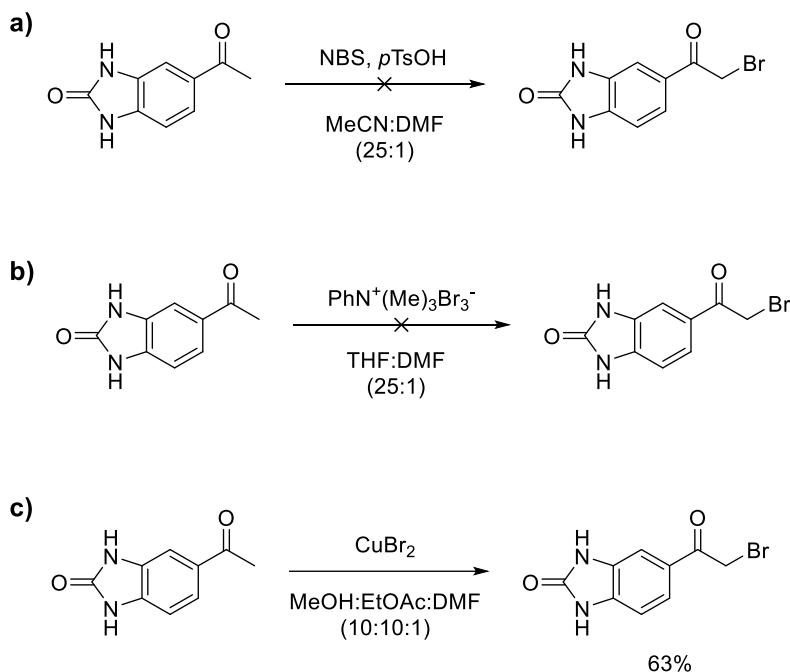
2.3.1   Compound Resynthesis

A synthetic route to access the preliminary hit compound was developed, involving five steps from readily available 4-aminoacetophenone (**Scheme 2.1**). A directed nitration reaction with bismuth nitrate and acetic anhydride afforded the desired *ortho*-nitrated intermediate **2.2**. Simultaneous acetyl deprotection and nitro reduction of **2.2.** was achieved with iron and hydrochloric acid to afford the diamino intermediate **2.3**. This was followed by cyclization with carbonyl diimidazole (CDI) to generate the desired urea functional group in intermediate **2.4**. Selective alpha-bromination of the ketone of **2.4** was performed with copper bromide to afford **2.5**, followed by an S$_N$2 reaction with *N*-methylfurfurylamine to generate the desired final product **2.1**.
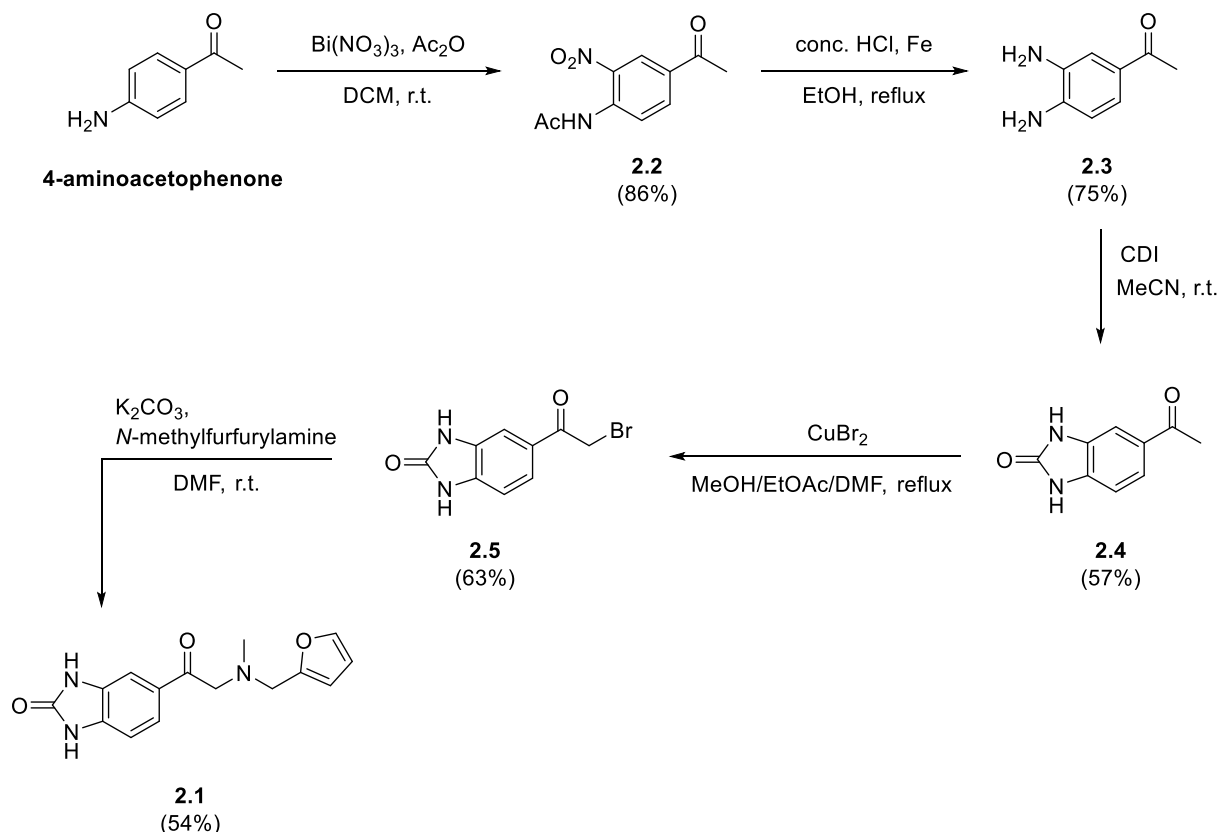
**Scheme 2.1**. Synthesis of **2.1** from **4-aminoacetophenone**.

While a synthetic route to access the preliminary hit was successfully obtained, the synthesis was not without challenges, particularly due to the diamine and urea functionalities. The diamine intermediate **2.3** was unstable to air, leading to challenges in obtaining the product in high purity. Additional challenges were posed by the poor solubility of the intermediates once the urea functionality was introduced. While this was advantageous in the synthesis of **2.4** (as impurities could be removed by simple vacuum filtration), it led to challenges in running subsequent reactions. Intermediate **2.4** was only soluble in highly polar solvents such as DMF and DMSO, however very few examples of alpha-bromination reactions have been reported in these solvents. After unsuccessful attempts with *N*-bromosuccinimide and phenyltrimethylammonium tribromide, it was found that using a minimal amount of DMF was compatible with CuBr$_2$ bromination (**Scheme 2.2**).

a)

NBS, *p*TsOH

MeCN:DMF
(25:1)

b)

PhN$^+$(Me)$_3$Br$_3^-$

THF:DMF
(25:1)

c)

CuBr$_2$

MeOH:EtOAc:DMF
(10:10:1)

63%

Scheme 2.2. Attempted bromination conditions in DMF.

Although S$_N$2 reactions are often performed in DMF, the poor solubility of both **2.4** and **2.1** for the final step led to challenges in purification. A liquid-liquid extraction was difficult due to the poor solubility of the final compound **2.1** in most organic solvents. This also led to difficulties in silica gel column chromatography, as residual DMF resulted in streaking and poor separation. This was overcome by the use of reverse phase chromatography, as the product could be loaded onto the column with DMF without any negative effects on separation, allowing the desired product to be obtained in high purity.

Given the challenges associated with the synthesis of the preliminary hit **2.1**, a more synthetically accessible amide analogue was also prepared. This analogue could be accessed in only two steps, starting from commercially available 3,4-diaminobenzoic acid (**Scheme 2.3**). Amide coupling gave intermediate product **2.6** which was then cyclized to the amide analogue product **2.7**. The urea functionality (which is responsible for the poor solubility observed in compounds **2.4**, **2.5** and **2.1**) was introduced in the final step, allowing the desired product **2.7** to be obtained in high purity by vacuum filtration as a result of the decreased solubility relative to starting material **2.6**.
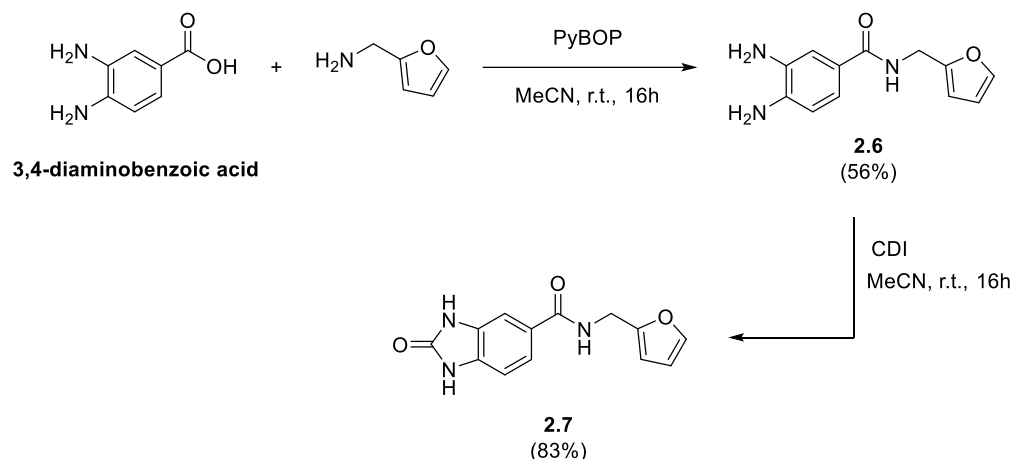
Scheme 2.3. Synthesis of amide analogue **2.7**.

2.3.2 Biological Evaluation

The compounds were evaluated by our collaborators for their inhibitory activity against AID-catalyzed deamination. They were tested at 500 µM concentration using a standard alkaline cleavage assay that quantifies the enzymatic deamination of an oligonucleotide substrate (as was used by them previously).[27, 28] Unfortunately, both **2.1** and the amide analogue **2.7** showed no activity *in vitro*.

2.1
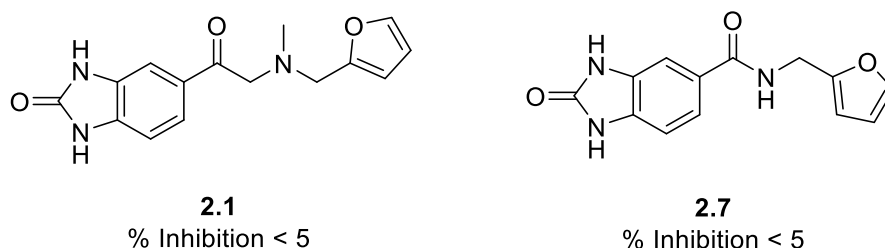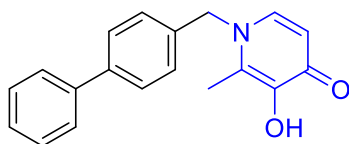% Inhibition < 5

2.7
% Inhibition < 5

Figure 2.6. *In vitro* activity of **2.1** and **2.7**.

False positives are a frustratingly common occurrence in experimental screening results and can be caused by many different mechanisms, such as assay interference, non-specific inhibition, and the presence of impurities.[29] Although the experimental results of the resynthesized **2.1** were discouraging, it demonstrates the importance of hit validation early in a drug discovery campaign to identify any potential false positives. Compound resynthesis is often considered the "gold

standard" for hit validation, particularly in instances where the initial observed activity is very weak and may be caused by impurities in the sample. The purity of commercial samples is often determined by LC-MS and may not accurately quantify the presence of impurities in the sample nor sample degradation during storage, both of which can result in false positive screening results.[26, 30, 31] Comparatively, compound resynthesis usually involves more rigorous evaluations of purity, including ^{1}H and ^{13}C NMR spectra of the product and intermediates in addition to HPLC chromatography. It is possible that the initial activity of **2.1** was due to an impurity in the commercial sample not present in the resynthesized compound. The *in vitro* testing was performed at high inhibitor concentrations to allow for the identification of weak binding inhibitors that can then be subsequently optimized to improve potency, however this means that even minor impurities may have been present at sufficiently high concentrations to affect enzyme activity.

2.4   Virtual Screening of Zinc-Binding Fragments

In light of the results with resynthesized **2.1**, we decided to perform a virtual screen to identify novel inhibitors for AID using our in-house docking program FITTED. A fragment-based drug design approach was pursued when assembling the virtual library of compounds for docking to AID. Fragment-based drug design involves the preliminary identification of small molecular fragments that display promising, though typically modest, affinity to a target of interest. Once initial fragments are identified, they are expanded to a more drug-like size by adding chemical groups to enhance binding affinity. A fragment-based strategy is particularly suited to the development of metalloenzyme inhibitors, where an optimal metal-binding group can confer considerable potency to an inhibitor. From a virtual screening standpoint, a fragment-based approach is attractive as it reduces the conformational search space of the molecule being docked, resulting in faster computations and allowing a greater chemical space to be explored in the same amount of time. As a result of these advantages, fragment-based drug design has been successfully employed towards the development of several metalloenzyme inhibitors, as shown in **Figure 2.7**.[32-34]

2.8
Anthrax lethal factor
$K_d \sim 1.8 \ \mu M$
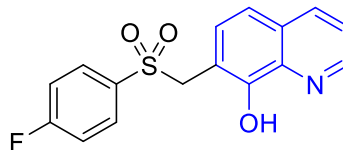
2.9
MMP-2
$IC_{50} = 3.0 \ \mu M$

Figure 2.7. Select examples of metalloenzyme inhibitors developed through a fragment-based approach, with the initial fragment shown in blue.

2.4.1 Preparation of a Virtual Library of Metal-Binding Fragments

It was envisioned that virtual screening of low molecular weight compounds could be used to identify an optimal zinc-binding fragment which could then be expanded on to form additional non-covalent interactions within the binding site of AID. Construction of the virtual library began with the ZINC database[35], selecting commercially available compounds with a molecular weight below 300 Da. The initial library was then filtered to select compounds containing zinc-binding groups based on a reported survey of zinc-binding groups in the PDB (**Figure 2.8**).[36] Compounds containing reactive functional groups (**Figure 2.8**) were removed and the remaining compounds were then clustered by similarity to select 2000 diverse compounds for virtual screening.
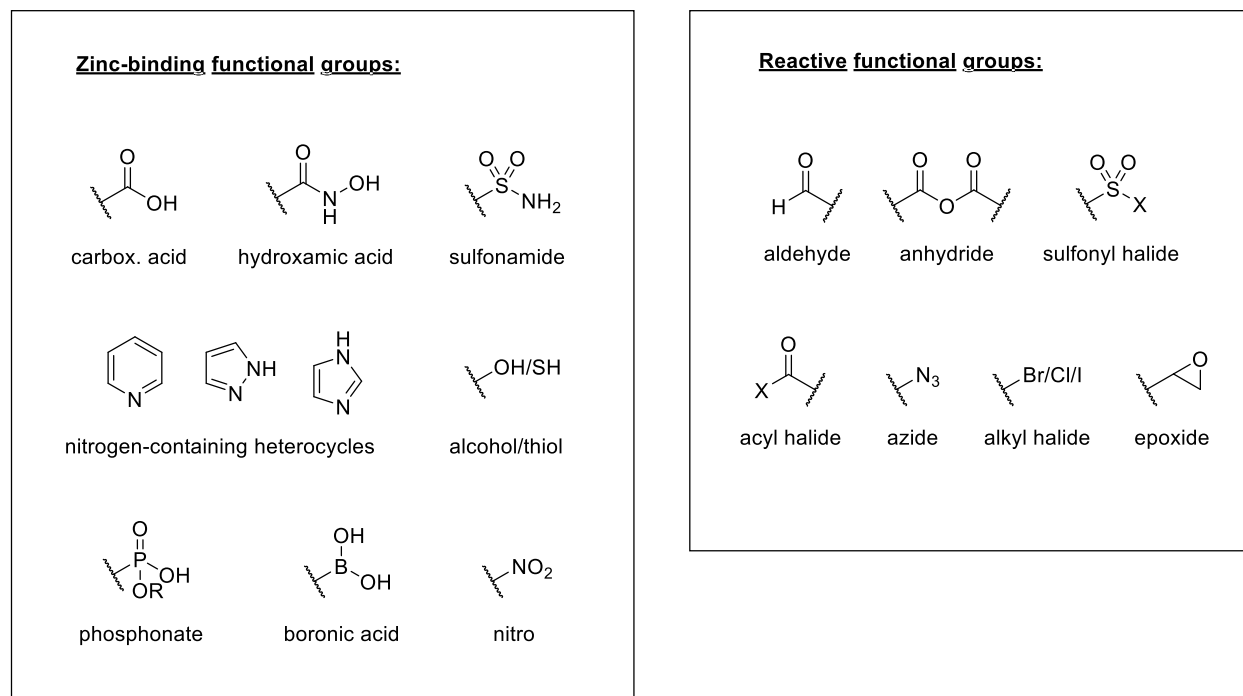
Figure 2.8. Zinc-binding functional groups and reactive functional groups used to filter the virtual screening library.

2.4.2   Virtual Screening to AID Crystal Structure

Docking was performed using the X-ray crystal structure of AID from PDB 5W0Z as it is the only available crystal structure with a catalytically active enzyme. The other three available crystal structures (5W1C, 5W0U, 5W0R) feature an E58A mutation, however this residue is known to play a key role in AID-catalyzed deamination, as shown in **Figure 2.2**, and its mutation results in a loss of enzymatic activity.[37, 38] Additionally, the energy of metal-binding ligands can be influenced by proton transfers to adjacent residues ,[39, 40] therefore the presence of the wildtype Glu[58] residue was believed to be critical for obtaining accurate docking results.

After docking, 8 compounds were selected based on docking scores as well as visual inspection of the 100 top scoring compounds. We were particularly optimistic about the predicted binding pose of an aromatic sulfonamide. As shown in **Figure 2.9**, the sulfonamide group was predicted to coordinate with the zinc ion, while the aromatic moiety appeared to favourably interact with the aromatic residues in the active site, forming a pi-stacking interaction with Tyr[114]. An additional

hydrogen bonding interaction was predicted between the sulfonamide carbonyl and the hydroxide group of Thr[27]. With this in mind, a small library of commercially available aromatic sulfonamides was assembled in a similar fashion as the previous library and docked to AID. Again, based on the most promising docking scores and visual inspection of the predicted binding poses, an additional two sulfonamides were selected for experimental evaluation.
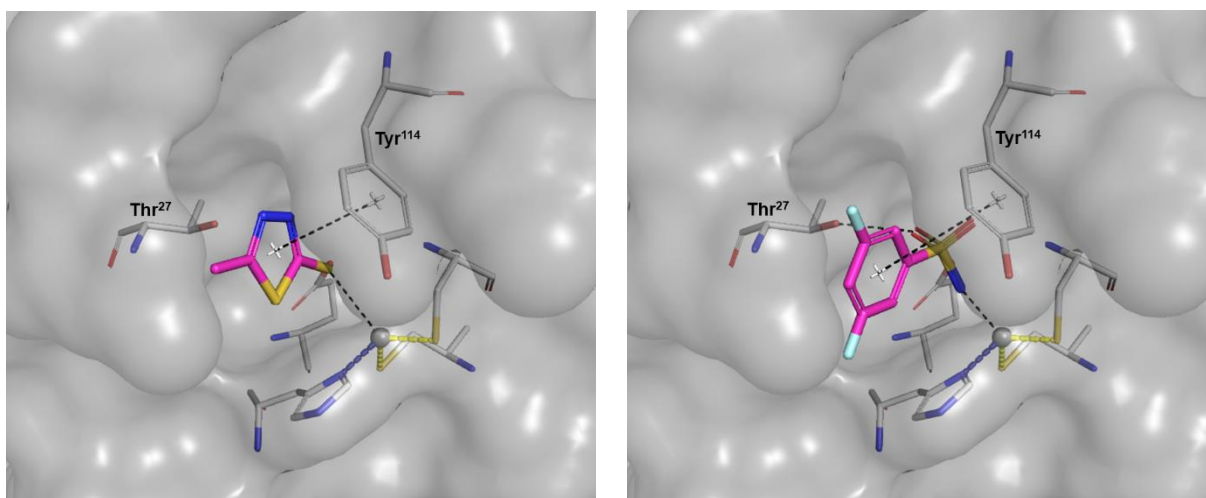

Figure 2.9. Select docking poses for **2.16** (left) and **2.19** (right).

2.4.3   Biological Evaluation of Screening Hits

Ten compounds were evaluated *in vitro* following the same procedure as with **2.1** and **2.7**, assessing their inhibitory activity at 500 µM inhibitor concentration. However, the inhibitors showed no *in vitro* inhibition of AID.
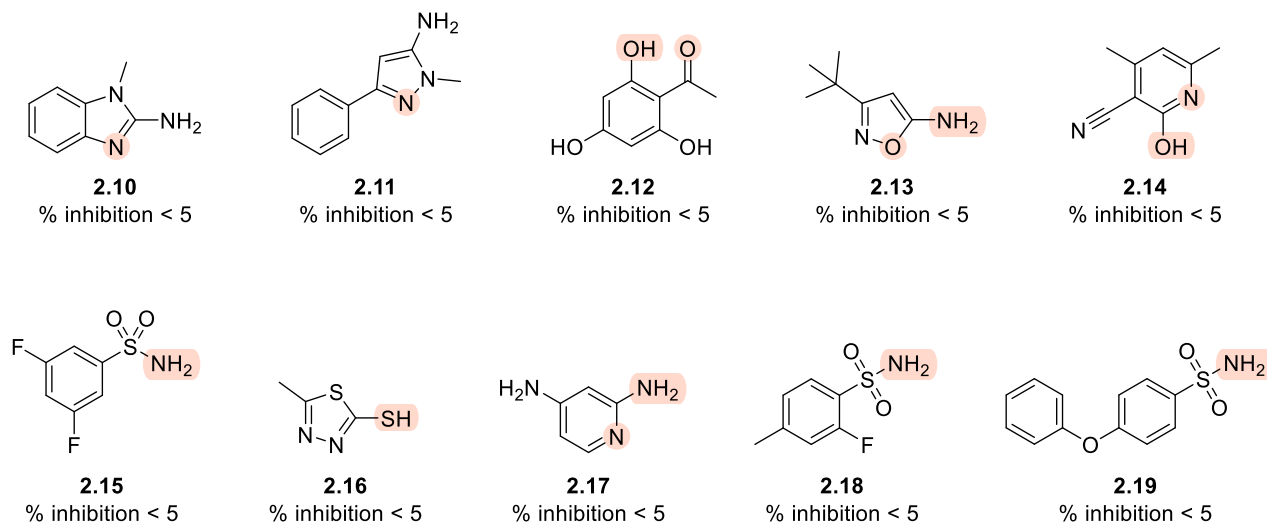
Figure 2.10. *In vitro* screening results of 10 compounds (**2.10-2.19**) identified by virtual screening, with their predicted zinc-binding groups highlighted.

Although the experimental results were discouraging, it was difficult to assess whether the results were indicative of a limitation in our computational method or influenced by the absence of experimental activity data against AID. The docking scores predicted by FITTED are not representative of absolute binding affinities, but rather relative scores that should score high affinity compounds better than inactive compounds. However, in the absence of any known inhibitors, it is difficult to assess whether our predicted high scoring ligands in **Figure 2.10** would have scored higher than a true inhibitor of AID. FITTED has been previously shown to perform well with other zinc metalloenzyme targets, successfully distinguishing between active and inactive compounds.[24] However, those studies were performed on different zinc metalloenzymes and does not necessarily guarantee the same performance with AID. Altogether, we felt that these screening results warranted further evaluation of our docking program against AID.

2.5   Validation of FITTED with AID Coordination Sphere

One consideration that we thought could be contributing to the poor results of our screening assay was the difference in the metal-binding residues of AID compared to what had been included in our previous zinc training set. AID contains a Cys_2His coordination sphere, whereas the training set for FITTED included zinc metalloenzymes that were predominantly coordinated by histidine

and aspartic acid residues. However, extensive mutation studies on carbonic anhydrase have shown that changes in the zinc coordinating residues can significantly influence enzymatic properties such as the pKa of the coordinated water, the hydrolytic activity of the enzyme and the binding affinity of known inhibitors.[41] Even when comparing similarly charged coordination spheres (e.g. His_2Asp vs. His_2Cys), the affinity of metal-binding ligands has been shown to vary.[42, 43] Therefore, we wanted to ensure that our model for metal interaction energy within FITTED was appropriate for the coordination sphere of AID.
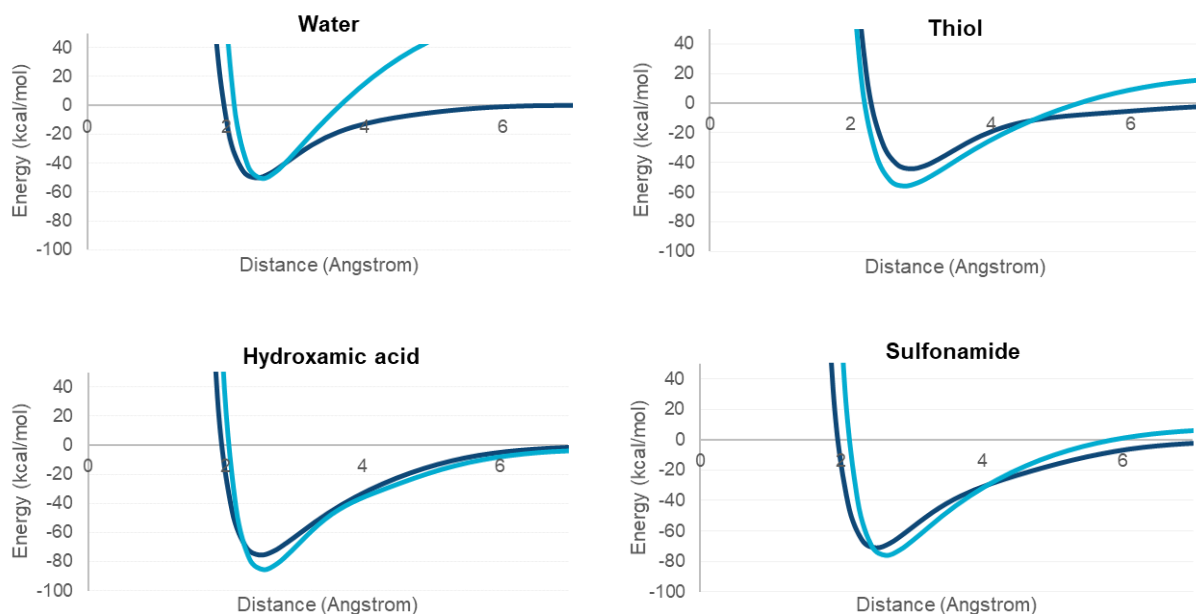

Figure 2.11. Comparison of implemented Lennard-Jones potential (dark blue) with QM-derived curve (light blue).

Following the procedure previously employed for zinc-binding optimization in FITTED, quantum mechanics (QM) calculations were used to determine the binding energy of various ligands at incremental distances from the zinc ion in a truncated AID coordination sphere.[24] As shown in **Figure 2.11**, we found that the QM curves were well reproduced by the Lennard-Jones potential previously developed and implemented in FITTED. As a note, the thiol, hydroxamic acid and sulfonamide curves are a combination of curves obtained when considering both a neutral ligand at further zinc-ligand distances and ligand ionization via proton transfer to Glu[58] at shorter zinc-ligand distances. For water, we were only concerned about the energy of water displacement and

therefore only the curve representing ligand ionization via proton transfer at short zinc-water distances was obtained (leading to the observed discrepancy in binding energy at longer distances). These results indicate that ligand binding to the zinc ion of AID is accurately modeled with FITTED, suggesting that the poor virtual screening results may not be due to a limitation of our program for this target.

2.6   Conclusion and Future Work

Although no inhibitor for AID was developed, several valuable takeaways were obtained from this work. The loss of activity following resynthesis of **2.1** demonstrates the importance of hit validation early in the drug discovery process to identify preliminary results that were due to false positives, especially in the case of commercial compound libraries. Additionally, the Lennard-Jones potential implemented in FITTED successfully reproduced the QM-derived energy curves for a truncated AID system, demonstrating the applicability of this method to accurately model metal coordination in a range of zinc metalloenzymes. Other takeaways include considerations regarding the preparation of virtual screening libraries and the selection of compounds for experimental testing. The virtual screening library was prepared using the commercially available subset of the ZINC database. However, many of these compounds are available at a prohibitively high price point and with long lead times, making them unsuitable for experimental evaluation. Based on this experience, it would likely be more efficient to use libraries provided by suppliers like ChemSpace rather than the commercially available subset of the ZINC database. Although supplier libraries are typically smaller, this will ensure that all compounds in a virtual screening library are truly commercially available. It also might be worthwhile to consider the diversity of the zinc-binding group when selecting compounds for experimental evaluation, rather than selecting compounds based predominately on the docking score. Considering the strength of metal-ligand bonds, the top scoring compounds may be more sensitive to small biases of the scoring function that could limit the diversity of high-scoring compounds.

Following the unsuccessful results of compounds **2.1**, **2.7**, and **2.10-2.19**, an alternative approach was pursued by another graduate student based on reported cytidine deaminase inhibitors. Several compounds based on transition-state analogues for cytidine deaminase were tested *in vitro* but did not display any inhibitory activity. In combination with the work reported in this chapter, these

results suggest that AID may not be well suited to inhibition with small molecules. A structural study done by Pham *et al.* suggests the presence of an ssDNA binding cleft, as can be observed with the high surface charge of AID and shown in **Figure 2.12**.[44]
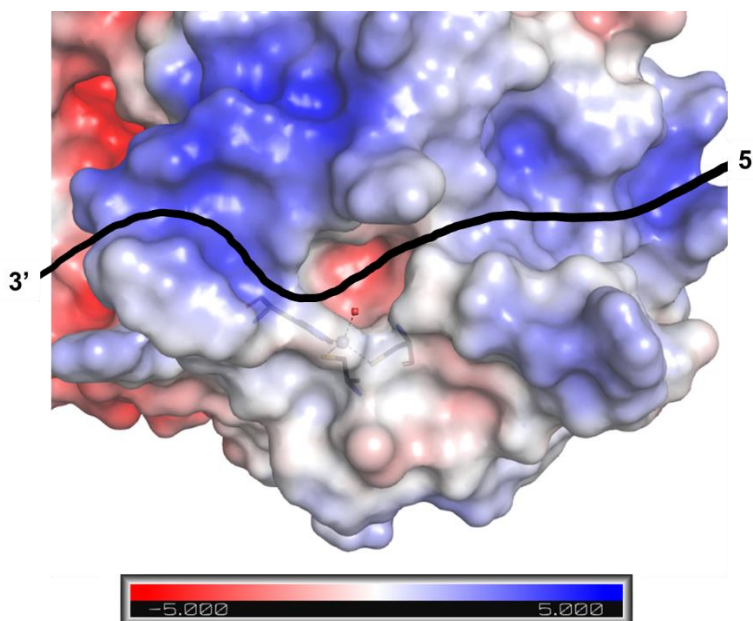

Figure 2.12. Proposed ssDNA binding cleft of AID.

The substrate affinity of AID may be more predominantly attributed to favorable interactions between the negatively charged DNA backbone and the positively charged binding cleft, rather than interactions within the cytidine binding pocket. Targeting AID with small molecules may be challenging considering the shallowness of active site, and sufficient potency may not achievable even when incorporating a zinc-binding group. This is also supported by recent reports of APOBEC inhibition using ssDNA constructs,[10, 45] suggesting that oligonucleotide-based therapeutics may be a more viable approach for the development of AID inhibitors.

2.7   Experimental

2.7.1   General

Unless otherwise specified, all solvents and reagents were purchased from commercial suppliers and used without further purification. All ^{1}H, ^{13}C and ^{19}F NMR spectra were acquired on a Bruker

400 or 500 MHz spectrometer, or a Varian 500 MHz spectrometer. Chemical shifts are reported in ppm using the residual of deuterated solvents as an internal standard. Chromatography was performed on silica gel 60 (230-40 mesh) or using the Biotage One Isolera with ZIP cartridges. High resolution mass spectrometry was performed by ESI on a Bruker Maxis Impact API QqTOF mass spectrometer at McGill University. Reversed-phase HPLC (water and MeCN or MeOH gradient) was used to verify the purity of compounds on an Agilent 1100 series instrument equipped with VWD-detector, C18 reverse column (Agilent, Zorbax Eclipse XDBC18 150 mm 4.6 mm, 5 μm), and UV detection at 254 nm.

2.7.2   Synthesis and Characterization Data

***N*-(4-acetyl-2-nitrophenyl)acetamide (2.2).** To a solution of 4-aminoacetophenone (1.0 g, 7.4 mmol) in DCM (35 mL) was added $Bi(NO_3)_3$•$5H_2O$ (10.0 g, 22.2 mmol) followed by acetic anhydride (4.2 mL, 44.4 mmol), and the reaction stirred at room temp. overnight. The reaction was neutralized with sat. $NaHCO_3$ and filtered through celite. The filtrate was extracted with DCM (x3), dried over Na_2SO_4, filtered and concentrted *in vacuo*. The crude product was further purified by column chromatography (1:1 Hex:EtOAc) to afford the desired product (1.42 g, 86% yield) as a yellow powder that was pure by ^{1}H NMR. All characterization data were in full agreement with those reported in the literature.[46]

1-(3,4-diaminophenyl)ethan-1-one (2.3). To a suspension of *N*-(4-acetyl-2-nitrophenyl) acetamide (500 mg, 2.25 mmol) in EtOH (20 mL) was added 6 M HCl (7.5 mL) and the reaction heated at reflux for 2 h. Iron powder (628 mg, 11.25 mmol) was added to the solution and refluxed for an additional hour. The reaction was cooled to room temp., neutralized with sat. $NaHCO_3$ and filtered through celite. The filtrate was diluted with EtOAc and washed with sat. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo*. The desired product (253 mg, 75% yield) was obtained as a brown powder that was used in the next step without further purification. All characterization data were in full agreement with those reported in the literature.[47]

5-acetyl-1,3-dihydro-2H-benzo[d]imidazol-2-one (2.4). To a solution of 1-(3,4-diaminophenyl) ethan-1-one (1.19 g, 7.92 mmol) in MeCN (10 mL) was added carbonyl-diimidazole (1.93 g, 11.9 mmol) and the reaction stirred at room temp. for 3 h. The precipitate was collected by filtration, rinsing with acetone, to afford the desired product (795 mg, 57% yield) as a beige powder that was

pure by NMR. All characterization data were in full agreement with those reported in the literature.[48]

5-(2-bromoacetyl)-1,3-dihydro-2H-benzo[d]imidazol-2-one (2.5). 5-Acetyl-1,3-dihydro-2H-benzo[d]imidazol-2-one (480 mg, 2.70 mmol) was dissolved in a minimal amount of DMF (1 mL) and diluted with a 1:1 mixture of EtOAc:MeOH (20 mL). To the solution was added $CuBr_2$ (1.34 g, 6.0 mmol) and the reaction stirred at reflux for 3 h. The solvent was removed *in vacuo* and the crude product suspended in acetonitrile. The precipitate was collected by vacuum filtration and rinsed with acetonitrile until the green copper salts were removed. The precipitate was rinsed with acetone and dried under vacuum to afford the desired product (436 mg, 63% yield) as a beige powder that was used in the next step without further purification. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.13 (s, 1H), 10.96 (s, 1H), 7.71 (dd, J = 8.3, 1.7 Hz, 1H), 7.51 (d, J = 1.7 Hz, 1H), 7.04 (d, J = 8.2 Hz, 1H), 4.85 (s, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 190.57, 155.43, 134.70, 129.84, 126.81, 123.31, 108.52, 108.13, 33.73. HRMS (ESI/QTOF) m/z: $[M + Na]^+$ Calcd for $C_9H_7BrN_2NaO_2^+$ 276.9583; Found 276.9579.

5-(*N*-(furan-2-ylmethyl)-N-methylglycyl)-1,3-dihydro-2H-benzo[d]imidazol-2-one (2.1). To a solution of 5-(2-bromoacetyl)-1,3-dihydro-2H-benzo[d]imidazol-2-one (100 mg, 0.39 mmol) in DMF (1.5 mL) was added *N*-methylfurfurylamine (68 μL, 0.58 mmol) and K_2CO_3 (81 mg, 0.59 mmol) and the reaction stirred at room temp. for 1 h. The reaction was filtered through filter paper, rinsing with MeOH. The filtrate was concentrated *in vacuo* and loaded directly onto a C18 column and purified by column chromatography (50% $MeOH/H_2O$ – 95% $MeOH/H_2O$). The desired fractions were combined, concentrated *in vacuo* and lyophilized to afford the desired product (60 mg, 54% yield) as a pale beige solid that was pure by NMR and HPLC. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.02 (s, 1H), 10.86 (s, 1H), 7.64 (dd, J = 8.2, 1.6 Hz, 1H), 7.60 (dd, J = 1.9, 0.9 Hz, 1H), 7.50 (d, J = 1.6 Hz, 1H), 6.98 (d, J = 8.2 Hz, 1H), 6.40 (dd, J = 3.1, 1.9 Hz, 1H), 6.30 (d, J = 3.1 Hz, 1H), 3.77 (s, 2H), 3.69 (s, 2H), 2.23 (s, 3H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 196.20, 155.45, 151.91, 142.51, 134.05, 129.55, 128.79, 122.37, 110.30, 108.88, 108.05, 107.86, 62.00, 52.69, 41.76. HRMS (ESI/QTOF) m/z: $[M + H]^+$ Calcd for $C_{15}H_{16}N_3O_3^+$ 286.1186; Found 286.1188.

3,4-diamino-*N*-(furan-2-ylmethyl)benzamide (2.6). To a solution of 3,4-diaminobenzoic acid (500 mg, 3.29 mmol) in DCM (50 mL) was added DIPEA (1.14 mL, 6.57 mmol) and furfurylamine (0.35 mL, 3.94 mmol) followed by PyBOP (2.57 g, 4.93 mmol) and the reaction stirred at room temp. for 3 h. The reaction was diluted with DCM and washed with sat. NH_4Cl, sat. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude product was purified by column chromatography (0 – 10% MeOH/DCM + 1% NH_4OH) to afford the desired product (424 mg, 56% yield) as a beige powder. ^{1}H NMR (500 MHz, dmso) δ 8.33 (t, *J* = 5.7 Hz, 1H), 7.58 – 7.49 (m, 1H), 7.07 (d, *J* = 2.0 Hz, 1H), 7.05 – 6.93 (m, 1H), 6.47 (d, *J* = 8.1 Hz, 1H), 6.37 (dd, *J* = 3.2, 1.9 Hz, 1H), 6.24 – 6.16 (m, 1H), 5.00 (s, 4H), 4.46 – 4.35 (m, 2H).

N-(furan-2-ylmethyl)-2-oxo-2,3-dihydro-*1H*-benzo[*d*]imidazole-5-carboxamide (2.7). To a solution of 3,4-diamino-*N*-(furan-2-ylmethyl)benzamide (200 mg, 0.86 mmol) in MeCN (5 mL) was added carbonyl-diimidazole (205 mg, 1.28 mmol) and the reaction stirred at room temp. overnight. The precipitate was collected by filtration, rinsing with acetone, to afford the desired product (184 mg, 83% yield) as a beige powder that was pure by NMR. ^{1}H NMR (500 MHz, dmso) δ 10.88 (s, 1H), 10.84 (s, 1H), 8.82 (t, *J* = 5.7 Hz, 1H), 7.58 – 7.52 (m, 2H), 7.47 (d, *J* = 1.6 Hz, 1H), 6.96 (d, *J* = 8.1 Hz, 1H), 6.38 (dd, *J* = 3.2, 1.8 Hz, 1H), 6.24 (d, *J* = 3.2 Hz, 1H), 4.44 (d, *J* = 5.7 Hz, 2H).

2.7.3   Computational Studies

Energy Curves. The computational modelling of ligand binding to AID was performed according to the previously reported procedure.[24] A truncated coordination sphere was employed from the PDB (5W0Z). It involved systematically moving the ligand away from the zinc ion in 0.2 Å increments and obtaining single point energy calculations at the B3LYP D3BJ/6-31G* level of theory. The acquired QM energy curves were then compared to those obtained using the optimized Lennard-Jones equation implemented in FITTED.

2.7.4   Experimental Evaluation

***In vitro* evaluation of AID inhibition.**

The standard alkaline cleavage assay for AID-mediated deamination was used to screen

compounds for inhibition of AID, using a standard seven-nucleotide bubble substrate containing the WRC motif TGC (5′-TTTGCTT-3′) as a substrate (2.5 nM) as previously reported.[49] The compounds were tested at 500 µM concentration and dissolved in water or DMSO, with the final concentration of DMSO not exceeding 1%. Positive controls with DMSO (1%) or without were employed, with no significant impact on AID deamination observed with 1% DMSO (see Supporting Information – Figure S2.1).

2.8 References

1. Canadian Cancer Statistics 2019. cancer.ca/Canadian-Cancer-Statistics-2019-EN

2. Liu, M.; Duke, J. L.; Richter, D. J.; Vinuesa, C. G.; Goodnow, C. C.; Kleinstein, S. H.; Schatz, D. G., Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **2008,** *451* (7180), 841-5.

3. Ramiro, A. R.; Jankovic, M.; Eisenreich, T.; Difilippantonio, S.; Chen-Kiang, S.; Muramatsu, M.; Honjo, T.; Nussenzweig, A.; Nussenzweig, M. C., AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* **2004,** *118* (4), 431-8.

4. Klemm, L.; Duy, C.; Iacobucci, I.; Kuchen, S.; von Levetzow, G.; Feldhahn, N.; Henke, N.; Li, Z.; Hoffmann, T. K.; Kim, Y. M., et al., The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell* **2009,** *16* (3), 232-45.

5. Kawamura, K.; Wada, A.; Wang, J. Y.; Li, Q.; Ishii, A.; Tsujimura, H.; Takagi, T.; Itami, M.; Tada, Y.; Tatsumi, K., et al., Expression of activation-induced cytidine deaminase is associated with a poor prognosis of diffuse large B cell lymphoma patients treated with CHOP-based chemotherapy. *J Cancer Res Clin Oncol* **2016,** *142* (1), 27-36.

6. Muramatsu, M.; Sankaranand, V. S.; Anant, S.; Sugai, M.; Kinoshita, K.; Davidson, N. O.; Honjo, T., Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* **1999,** *274* (26), 18470-6.

7. Qiao, Q.; Wang, L.; Meng, F. L.; Hwang, J. K.; Alt, F. W.; Wu, H., AID Recognizes Structured DNA for Class Switch Recombination. *Mol Cell* **2017,** *67* (3), 361-373 e4.

8. Begum, N. A.; Nagaoka, H.; Kobayashi, M.; Honjo, T., Chapter 18 - Molecular Mechanisms of AID Function. In *Molecular Biology of B Cells (Second Edition)*, Alt, F. W.; Honjo, T.; Radbruch, A.; Reth, M., Eds. Academic Press: London, 2015; pp 305-344.

9. Olson, M. E.; Harris, R. S.; Harki, D. A., APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem Biol* **2018,** *25* (1), 36-49.

10. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A. M.; Kurup, H. M.; Jones, K. F.; Sutton, L.; Donahue, J.; D'Aquila, R. T.; Jameson, G. B., et al., Differential Inhibition of APOBEC3 DNA-Mutator Isozymes by Fluoro- and Non-Fluoro-Substituted 2'-Deoxyzebularine Embedded in Single-Stranded DNA. *ChemBioChem* **2020,** *21* (7), 1028-1035.

11. Ko, T. P.; Lin, J. J.; Hu, C. Y.; Hsu, Y. H.; Wang, A. H.; Liaw, S. H., Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J Biol Chem* **2003,** *278* (21), 19111-7.

12. Montamat-Sicotte, D.; Litzler, L. C.; Abreu, C.; Safavi, S.; Zahn, A.; Orthwein, A.; Muschen, M.; Oppezzo, P.; Munoz, D. P.; Di Noia, J. M., HSP90 inhibitors decrease AID levels and activity in mice and in human cells. *Eur J Immunol* **2015,** *45* (8), 2365-76.

13. Taldone, T.; Gozman, A.; Maharaj, R.; Chiosis, G., Targeting Hsp90: small-molecule inhibitors and their clinical development. *Curr Opin Pharmacol* **2008,** *8* (4), 370-4.

14.   Beale, R. C.; Petersen-Mahrt, S. K.; Watt, I. N.; Harris, R. S.; Rada, C.; Neuberger, M. S., Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* **2004,** *337* (3), 585-96.

15.   Samaranayake, M.; Bujnicki, J. M.; Carpenter, M.; Bhagwat, A. S., Evaluation of molecular models for the affinity maturation of antibodies: roles of cytosine deamination by AID and DNA repair. *Chem Rev* **2006,** *106* (2), 700-19.

16.   Li, M.; Shandilya, S. M.; Carpenter, M. A.; Rathore, A.; Brown, W. L.; Perkins, A. L.; Harki, D. A.; Solberg, J.; Hook, D. J.; Pandey, K. K., et al., First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem Biol* **2012,** *7* (3), 506-17.

17.   Olson, M. E.; Li, M.; Harris, R. S.; Harki, D. A., Small-molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-amino-1,2,4-triazole-3-thiol scaffold. *ChemMedChem* **2013,** *8* (1), 112-7.

18.   Zhang, Y.-H.; Guo, X.-C.; Zhong, J.-B.; Zhong, D.-X.; Huang, X.-H.; Fang, Z.-Y.; Zhang, C.; Lu, Y.-J., Discovery of APOBEC Cytidine Deaminases Inhibitors Using a BspH1 Restriction Enzyme-Based Biosensor. *ChemistrySelect* **2022,** *7* (21), e202201456.

19.   Riccardi, L.; Genna, V.; De Vivo, M., Metal–ligand interactions in drug design. *Nature Reviews Chemistry* **2018,** *2* (7), 100-112.

20.   Chen, K.; Xu, L.; Wiest, O., Computational exploration of zinc binding groups for HDAC inhibition. *J Org Chem* **2013,** *78* (10), 5051-5.

21.   Fruhauf, A.; Meyer-Almes, F. J., Non-Hydroxamate Zinc-Binding Groups as Warheads for Histone Deacetylases. *Molecules* **2021,** *26* (17).

22.   Geurs, S.; Clarisse, D.; De Bosscher, K.; D'Hooghe, M., The Zinc-Binding Group Effect: Lessons from Non-Hydroxamic Acid Vorinostat Analogs. *J Med Chem* **2023,** *66* (12), 7698-7729.

23.   Winum, J. Y.; Scozzafava, A.; Montero, J. L.; Supuran, C. T., Design of zinc binding functions for carbonic anhydrase inhibitors. *Curr Pharm Des* **2008,** *14* (7), 615-21.

24.   Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of HDACs and other Zinc Metalloenzymes Inhibitors. *J Chem Inf Model* **2014,** *54* (1), 254-265.

25.   Mendoza-Sanchez, R.; Cotnoir-White, D.; Kulpa, J.; Jutras, I.; Pottel, J.; Moitessier, N.; Mader, S.; Gleason, J. L., Design, synthesis and evaluation of antiestrogen and histone deacetylase inhibitor molecular hybrids. *Bioorg Med Chem* **2015,** *23* (24), 7597-606.

26.   Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L., Principles of early drug discovery. *Br J Pharmacol* **2011,** *162* (6), 1239-49.

27.   Larijani, M.; Martin, A., Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol Cell Biol* **2007,** *27* (23), 8038-48.

28.   Sohail, A.; Klapacz, J.; Samaranayake, M.; Ullah, A.; Bhagwat, A. S., Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res* **2003,** *31* (12), 2990-4.

29.   Sink, R.; Gobec, S.; Pecar, S.; Zega, A., False positives in the early stages of drug discovery. *Curr Med Chem* **2010,** *17* (34), 4231-55.

30.   Hermann, J. C.; Chen, Y.; Wartchow, C.; Menke, J.; Gao, L.; Gleason, S. K.; Haynes, N. E.; Scott, N.; Petersen, A.; Gabriel, S., et al., Metal impurities cause false positives in high-throughput screening campaigns. *ACS Med Chem Lett* **2013,** *4* (2), 197-200.

31.   Thorne, N.; Auld, D. S.; Inglese, J., Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr Opin Chem Biol* **2010,** *14* (3), 315-24.

32.   Agrawal, A.; Johnson, S. L.; Jacobsen, J. A.; Miller, M. T.; Chen, L. H.; Pellecchia, M.; Cohen, S. M., Chelator fragment libraries for targeting metalloproteinases. *ChemMedChem* **2010,** *5* (2), 195-9.

33.   Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J., Recent developments in fragment-based drug discovery. *J Med Chem* **2008,** *51* (13), 3661-80.

34.   Jacobsen, J. A.; Fullagar, J. L.; Miller, M. T.; Cohen, S. M., Identifying chelators for metalloprotein inhibitors using a fragment-based approach. *J Med Chem* **2011,** *54* (2), 591-602.

35.   Irwin, J. J.; Shoichet, B. K., ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005,** *45* (1), 177-82.

36.   Kawai, K.; Nagata, N., Metal-ligand interactions: an analysis of zinc binding groups using the Protein Data Bank. *Eur J Med Chem* **2012,** *51*, 271-6.

37.   Chaudhuri, J.; Tian, M.; Khuong, C.; Chua, K.; Pinaud, E.; Alt, F. W., Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **2003,** *422* (6933), 726-30.

38.   Ramiro, A. R.; Stavropoulos, P.; Jankovic, M.; Nussenzweig, M. C., Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol* **2003,** *4* (5), 452-6.

39.   Cross, J. B.; Duca, J. S.; Kaminski, J. J.; Madison, V. S., The active site of a zinc-dependent metalloproteinase influences the computed pK(a) of ligands coordinated to the catalytic zinc ion. *J Am Chem Soc* **2002,** *124* (37), 11004-7.

40.   Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y., Zinc chelation with hydroxamate in histone deacetylases modulated by water access to the linker binding channel. *J Am Chem Soc* **2011,** *133* (16), 6110-3.

41.   Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M., Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem Rev* **2008,** *108* (3), 946-1051.

42.   Martin, D. P.; Hann, Z. S.; Cohen, S. M., Metalloprotein-inhibitor binding: human carbonic anhydrase II as a model for probing metal-ligand interactions in a metalloprotein active site. *Inorg Chem* **2013,** *52* (21), 12207-15.

43.   Kiefer, L. L.; Fierke, C. A., Functional characterization of human carbonic anhydrase II variants with altered zinc binding sites. *Biochemistry* **1994,** *33* (51), 15233-40.

44.   Pham, P.; Afif, S. A.; Shimoda, M.; Maeda, K.; Sakaguchi, N.; Pedersen, L. C.; Goodman, M. F., Structural analysis of the activation-induced deoxycytidine deaminase required

in immunoglobulin diversification. *DNA Repair (Amst)* **2016,** *43*, 48-56.

45.   Harjes, S.; Kurup, H. M.; Rieffer, A. E.; Bayarjargal, M.; Filitcheva, J.; Su, Y.; Hale, T. K.; Filichev, V. V.; Harjes, E.; Harris, R. S., et al., Structure-guided inhibition of the cancer DNA-mutating enzyme APOBEC3A. *Nat Commun* **2023,** *14* (1), 6382.

46.   Lu, Y.; Li, Y. M.; Zhang, R.; Jin, K.; Duan, C. Y., Regioselective nitration of phenyl carboxamides and primary anilines using bismuth nitrate/acetic anhydride. *Tetrahedron* **2013,** *69* (45), 9422-9427.

47.   Reinhardt, C. J.; Zhou, E. Y.; Jorgensen, M. D.; Partipilo, G.; Chan, J., A Ratiometric Acoustogenic Probe for in Vivo Imaging of Endogenous Nitric Oxide. *J Am Chem Soc* **2018,** *140* (3), 1011-1018.

48.   Diao, X.; Wang, Y.; Jiang, Y.; Ma, D., Assembly of substituted 1H-benzimidazoles and 1,3-dihydrobenzimidazol-2-ones via CuI/L-proline catalyzed coupling of aqueous ammonia with 2-iodoacetanilides and 2-iodophenylcarbamates. *J Org Chem* **2009,** *74* (20), 7974-7.

49.   Abdouni, H.; King, J. J.; Suliman, M.; Quinlan, M.; Fifield, H.; Larijani, M., Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res* **2013,** *41* (10), 5457-5468.

CHAPTER 3:

APPLICATION OF VIRTUAL SCREENING TOWARDS THE DISCOVERY OF A NOVEL TPP RIBOSWITCH BINDER

The work in this chapter is based, in part, on work in the following manuscript:

J. K. Stille[‡], N. Takyi[‡], O. S. Ayon[‡], M. Rivera, T. Satkunarajah, J. Pottel, A. Mittermaier, M. McKeague, N. Moitessier. Docking-based Virtual Screening for the Discovery of RNA-targeting Molecules - Application to the Riboswitch Binding Molecules. 2024. Manuscript in preparation

Contribution of Authors

J.K. Stille performed all work related to virtual screening in this chapter, including the development and validation of the docking protocol and preparation of the virtual screening library. Experimental evaluation of the virtual screening hits was performed by Omma S. Ayon in collaboration with the laboratory of Dr. Maureen McKeague.

3.1 Abstract

The work in this chapter describes the application of virtual screening towards the development of TPP riboswitch inhibitors. Our in-house docking program FITTED, which has been optimized for use with nucleic acids, was used to perform a virtual screening campaign of 6000 compounds against the TPP riboswitch. Based on the predicted docking scores and poses, 20 compounds were selected for experimental evaluation against the TPP riboswitch by surface plasmon resonance. Promisingly, one compound displayed dose-dependent binding to the TPP riboswitch aptamer domain, with an experimentally determined K_d of 170 μM.

3.2 Introduction

The growing resistance of pathogenic bacteria to antibiotics and the emergence of multi-drug resistant bacteria represents a major threat to human welfare.[1] The issue of antibiotic resistance has been exacerbated by the overuse and improper use of antibiotics, driving the evolutionary development of resistance.[2] Alarmingly, this has coincided with a steady decline in the development of new antibiotics and the withdrawal of many pharmaceutical companies from the antibacterial field.[3] Over the last 40 years, the majority of antibiotics have been based on existing structures and mechanisms, resulting in a low diversity of antibacterial agents.[4,5] Considering the rise of resistance to existing antibiotics, it is clear that new strategies for developing antibacterial therapeutics are urgently needed.

3.2.1 Riboswitches as Antibacterial Targets

Following their discovery in the early 2000s, riboswitches have emerged as an attractive new class of bacterial drug targets.[6,7] Riboswitches are regulatory segments of mRNA that control gene expression in response to binding a small molecule or ion. Riboswitches are composed of two parts, an aptamer-binding domain, and an expression platform. The binding of a ligand by the aptamer domain induces a conformational change in the expression platform that influences the translation or transcription of the mRNA.[8-10] Most riboswitches act in a *cis* manner in which ligand binding results in the inhibition of gene expression, either through the formation of a terminator loop (**Figure 3.1a**) or sequestration of the ribosome-binding site (**Figure 3.1b**).[11,12]
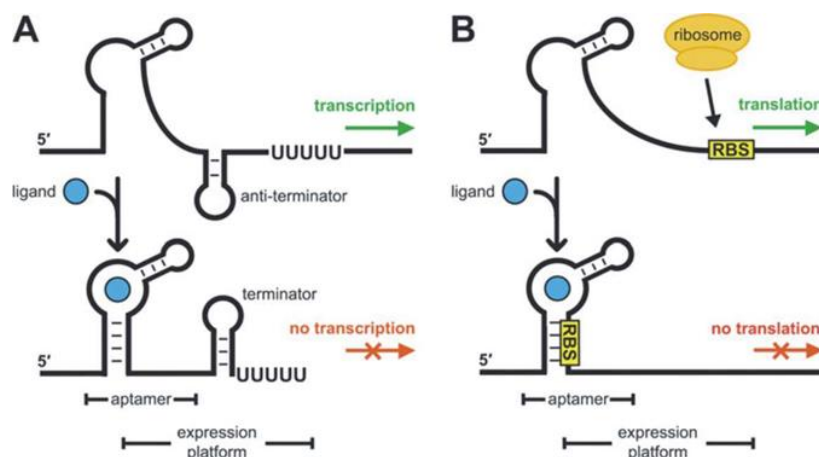
Figure 3.1. Riboswitch-mediated regulation of gene expression.[11]

This regulatory mechanism is found almost exclusively in bacteria – to date, roughly 40 classes of riboswitches have been identified in more than 6000 bacterial species.[13, 14] Riboswitches have been shown to bind their native ligands with high affinity and specificity, allowing them to selectively control gene expression even in the presence of a large number of structurally similar metabolites.[15-17]

The targeting of riboswitches by small molecules is still in its early stages, but there have already been several promising preliminary results that indicate the potential for this novel antibiotic target. It was discovered that the antimicrobial activity of roseoflavin, a chemical analogue of flavin mononucleotide (FMN), was a result of its ability to directly bind the FMN riboswitch and downregulate the expression of FMN riboswitch-regulated genes.[18] In a more recent report, a phenotypic screen by researchers at Merck led to the discovery of Ribocil, a highly selective inhibitor of the FMN riboswitch capable of inhibiting bacterial cell growth and displaying *in vivo* activity in a murine model of *E. coli* infection.[19] Importantly, Ribocil is structurally distinct from FMN, providing a promising proof-of-principle for riboswitch inhibition by synthetic small molecules.
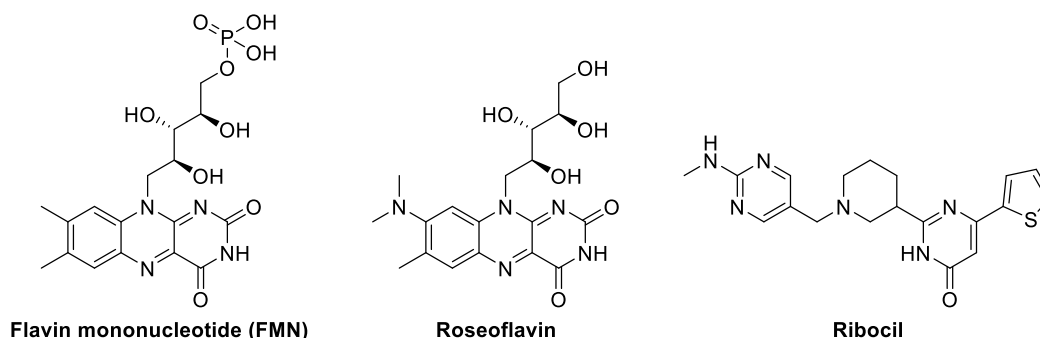
Figure 3.2. Chemical structures of the native ligand (FMN) and two inhibitors (Roseoflavin and Ribocil) for the FMN riboswitch.

3.2.2 TPP Riboswitch

Thiamine pyrophosphate (TPP) is an essential co-factor in carbohydrate and amino acid metabolism, and its availability is governed by the transport and *de novo* biosynthesis of thiamine derivatives. In bacteria, these processes have been shown to be regulated by the TPP riboswitch, such that binding of TPP is able to downregulate its own biosynthesis. This suggests that a TPP mimic could confer antibacterial activity by binding to the TPP riboswitch to induce TPP deficiency. Indeed, this has been observed with the antibacterial agent pyrithiamine pyrophosphate (PTPP), a structural analogue of TPP that was identified before the TPP riboswitch was even discovered.[20] Furthermore, the TPP riboswitch is one of the most abundant riboswitches among pathogenic bacteria and is present in 48 human pathogens, including the majority of pathogens listed on the WHO priority list (such as *S. aureus*, *K. pneumoniae*, and *E. faecium*).[21] The prevalence of the TPP riboswitch in therapeutically-relevant bacterial pathogens makes it a promising target for the development of broad-spectrum antibiotics.

The TPP riboswitch is one of the more studied riboswitches, and several crystal structures of the aptamer-binding domain in complex with TPP have been determined. As shown in **Figure 3.3** , TPP binds the *E. coli ThiM* riboswitch through two regions termed the pyrimidine sensor helix and the pyrophosphate sensor helix. Binding of both of these helices by TPP stabilizes the riboswitch in the "OFF" state, resulting in the downregulation of TPP-regulated genes. In the pyrimidine sensor helix, TPP intercalates between G42 and A43, and forms hydrogen bond interactions with G40. Binding in the pyrophosphate sensor helix is mediated by interactions between the pyrophosphate moiety and two Mg^{2+} ions, one of which is coordinated by the riboswitch residues

G78 and G60. Several reports have investigated the role of Mg^{2+} in TPP-binding, demonstrating that the presence of Mg^{2+} is necessary for TPP binding and may pre-organize the binding site toward ligand binding.[22-24]
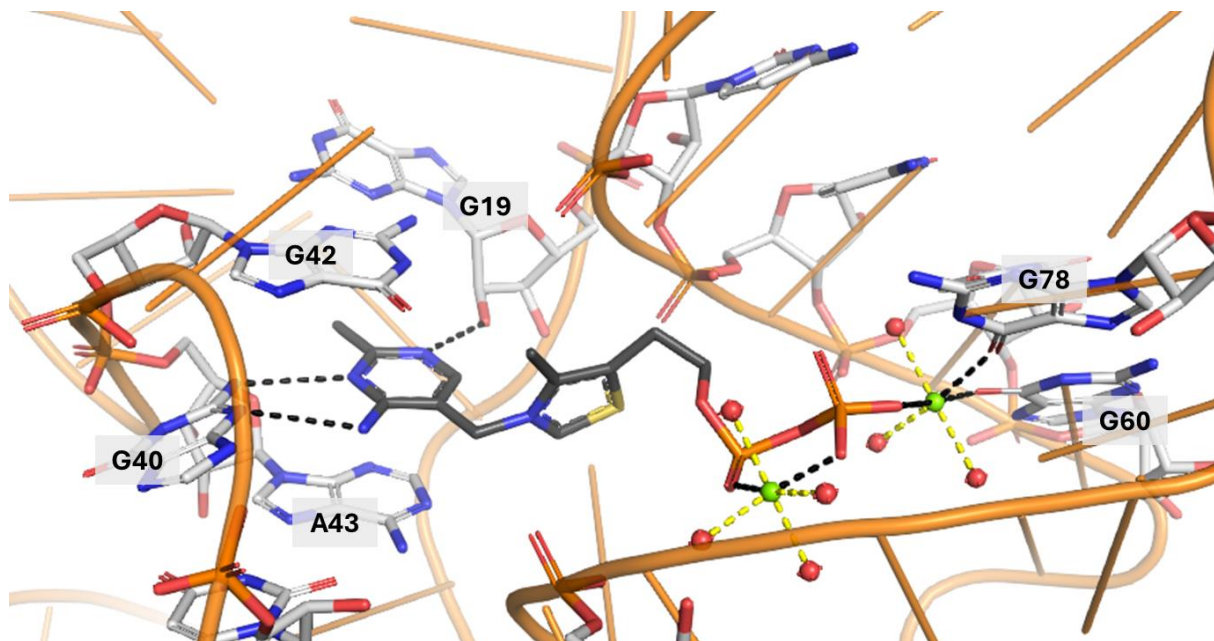
Figure 3.3. Binding site interactions of TPP (grey) with the aptamer domain of the *ThiM* riboswitch.

Due to the interest in the TPP riboswitch as a possible antibacterial target, several reports have been published describing the identification of novel binders. However, these have largely been limited to either close structural analogues of TPP or fragment-based ligands that bind only the pyrimidine sensor helix (**Figure 3.4**).
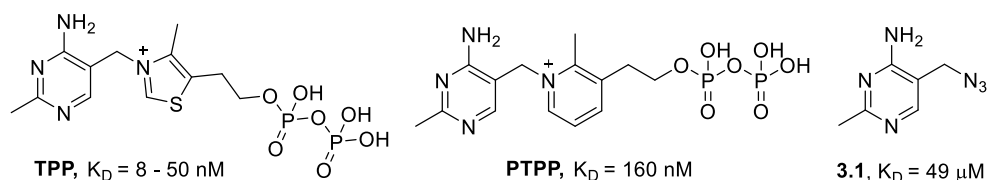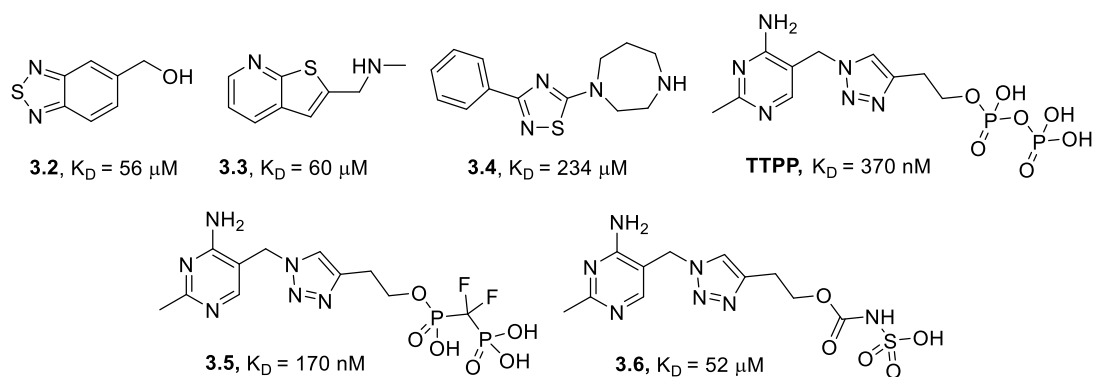
Figure 3.4. Chemical structures and binding affinities of selected TPP riboswitch ligands.[25, 26]

Biophysical screening of a fragment library against the TPP riboswitch led to the identification of several micromolar binders (**Figure 3.4** compounds **3.1, 3.2, 3.3, 3.4**).[26] However, a subsequent biochemical assay showed that none of the fragments were able to inhibit riboswitch-regulated gene expression. These fragments were shown to bind only the pyrimidine sensor helix, demonstrating that binding both the pyrimidine and pyrophosphate sensor helices may be necessary to inhibit gene expression.

As an example, Chen *et al.* screened a library of TPP analogues, and were able to maintain affinity when changing the central thiazolium ring of TPP to triazole (**Figure 3.4**, TTPP), while replacement of this heterocyclic group with aliphatic analogues led to a complete loss in potency.[25] Additionally, they explored modifications to the pyrophosphate group (**Figure 3.4** compounds **3.5** and **3.6**), with several analogues displaying moderate inhibition of riboswitch-mediated gene expression. As can be seen in **Figure 3.4**, TPP riboswitch inhibitors have largely been limited to close analogues of the native ligand, and only inhibitors featuring a trianionic pyrophosphate group (or similarly anionic groups) have shown inhibition of TPP-riboswitch regulated gene expression and sub-micromolar affinities.

3.2.3 Computer-Aided Design of Riboswitch-Targeting Drugs

In silico methods such as molecular docking has found widespread use in the development of therapeutically active inhibitors, however their application has largely been limited to protein targets. This is due to the fact that proteins have historically been favoured as drug targets, and therefore the majority of docking programs have been developed and optimized for use with

protein targets. While nucleic acid structures such as riboswitches have emerged as promising therapeutic targets, existing computational programs developed for proteins often face difficulties when applied directly to nucleic acids, limiting their use. Notably, nucleic acids are typically more flexible, polar, and solvated when compared to the more well-defined, lipophilic binding pockets of many proteins. As detailed in Chapter 1, this can result in challenges when applying docking programs developed for proteins to nucleic acid targets. As a result of these challenges, relatively few examples exist for the application of molecular docking towards riboswitch drug development. In 2011, Daldrop *et al.* used the DOCK docking program with RNA-specific parameters for van der Waals and electrostatic interaction energies to virtually screen for potential ligands for the adenine-binding riboswitch.[27] Their efforts led to the identification of four ligands with mid-micromolar affinity, and the predicted binding poses were confirmed by X-ray crystallography. More recently, a prospective virtual screen against the Pre-Q1 riboswitch led to the identification of six new ligands with binding affinities in the nanomolar to low micromolar range.[28] While no RNA-specific modifications were made to the docking programs employed, the authors used a pharmacophore model based on the binding mode of the Pre-Q1 riboswitch to its native ligand to aid in their selection of compounds for experimental testing, prioritizing compounds that reproduced the binding interactions observed with the native ligand. The experimental results indicated that the number of matching interactions between the binding pose and the pharmacophore model was more predictive of binding affinity than the docking score alone.

These examples demonstrate the possibility of employing molecular docking programs towards the development of riboswitch-targeting drugs. However, they also highlight the importance of optimizing the docking protocol, either through the optimization of the program itself or the implementation of additional compound selection criteria beyond docking score alone.

3.3 Results

3.3.1 Development of Docking Protocol

It was envisioned that our in-house docking program FITTED could be applied toward the identification of novel TPP riboswitch inhibitors by virtual screening. While FITTED was originally developed for docking to proteins, it has since undergone several optimizations for docking to nucleic acids.[29-31] In the most basic optimization, nucleotides were added to the program as

available "residues", allowing interaction sites to be generated around nucleotides that are employed during the conformational search and scoring process. Additionally, the charge on phosphate anions was minimized to account for the presence of counterions such that electrostatic interactions are not overestimated.[31] Nucleic acid binding sites are typically more exposed and solvated when compared to those in proteins, and therefore water often plays a critical role in binding. In order to account for this, we implemented "displaceable waters", which enable ligand binding through bridging water molecules or displacement of water molecules in poses in which direct ligand-nucleic acid binding is more favourable.[31]

As previously mentioned, nucleic acid structures are often highly negatively charged due to the presence of phosphate ions and will often contain metal counterions in the binding site. This is observed in the TPP riboswitch, where binding of TPP is mediated by the presence of two Mg^{2+} ions in the binding site. In order to more accurately model the binding of ligands to metals, our docking program was further optimized by fitting the molecular mechanics scoring function to more accurate quantum mechanics-derived potential energy curves.[29] For Mg^{2+} ions specifically, the interaction was also modelled by a periodic function to more accurately capture its preference for an octahedral binding geometry.
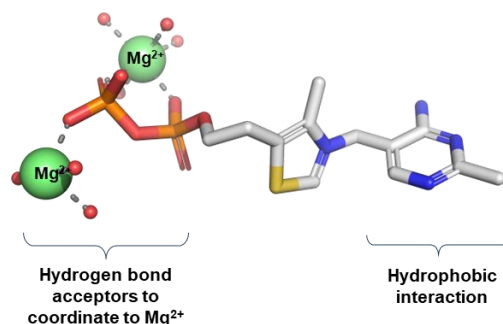
Figure 3.5. Important binding interactions included in the pharmacophore model.

To stabilize the riboswitch in the "OFF" state and provide significant binding affinity, the inhibitor should ideally bind both the pyrophosphate and pyrimidine sensor helices, mimicking the binding of the native TPP ligand. As a means of prioritizing potential inhibitors that can form both of these key interactions, a pharmacophore model was incorporated into the docking protocol. In addition to a genetic algorithm, FITTED also employs a matching algorithm during the conformational

search to guide ligand conformations towards more likely biding modes. The PROCESS module within the docking program generates interaction sites (i.e. hydrogen bond donor, acceptor, hydrophilic) within the target binding site which are used by the matching algorithm to favour ligand conformations that fulfil a certain percentage of these sites. For docking to TPP, the generated interaction sites were replaced by the key interactions shown in **Figure 3.5**. The pharmacophore interaction sites were based on the crystal structure of TPP bound to the aptamer domain of the TPP riboswitch (PDB: 2GDI). In order for a docked pose to be considered, it must contain a hydrophobic interaction within the pyrimidine sensor helix and a hydrogen-bonding interaction with at least one of the two magnesium ions of the pyrophosphate sensor helix.

To validate the computational method, it was evaluated by self-docking of TPP to PDB 2GDI, obtaining an RMSD of 1.01 Å (**Figure 3.6**). While a similar accuracy was obtained without the pharmacophore (RMSD = 1.29 Å), 2 out of 3 test runs had an RMSD > 2.5 Å, whereas all 3 with the pharmacophore had RMSD < 2.2 Å confirming that this set of interactions properly orient the docking towards more likely binding modes.
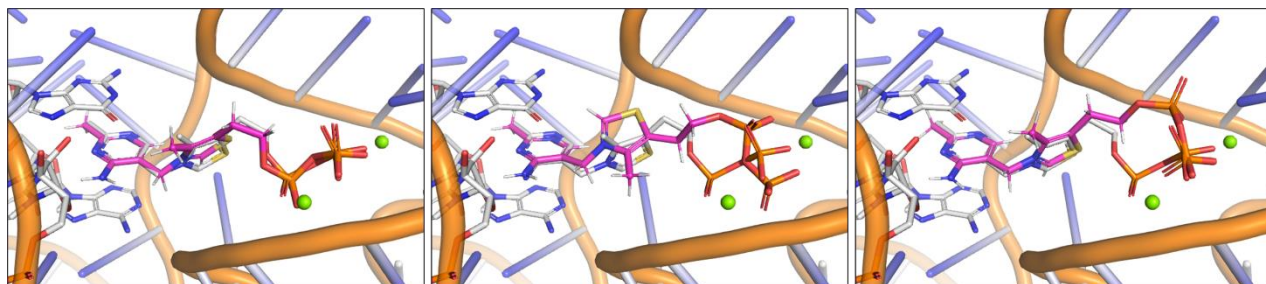
Figure 3.6. Predicted binding poses of TPP (pink) compared to crystal structure (white) when using pharmacophore constraint.
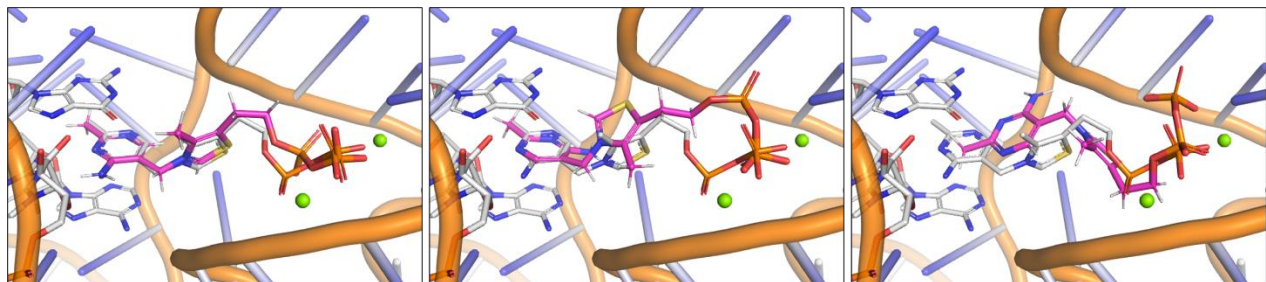
Figure 3.7. Predicted binding poses of TPP (pink) compared to crystal structure (white) without

<div align="center">pharmacophore.</div>

The self-docking results suggest that using a pharmacophore constraint could help reduce potential false negatives in the virtual screen due to incorrectly predicted binding poses (as was observed with 2 out 3 runs for TPP without the pharmacophore). The model was further evaluated on a library of known actives and decoys to assess its ability to distinguish between active and inactive compounds. Nine active molecules were selected based on the experimental binding affinity data from Chen *et al.*, selecting only molecules that showed > 25% inhibition of riboswitch-controlled gene expression (**Figure 3.8**).[32] Decoys were generated using the DUD-E database,[33] which generates decoys with similar physico-chemical properties to each active compound, and 20 decoys per active compound were selected.
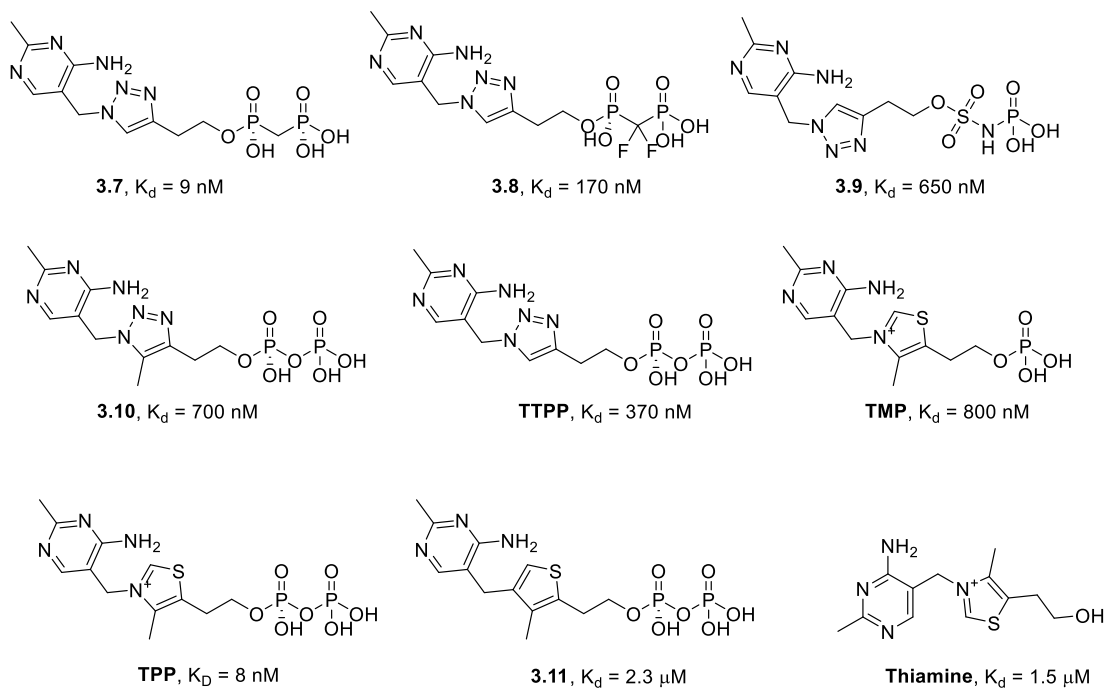
Figure 3.8. Active compounds from Chen *et al.*[32]
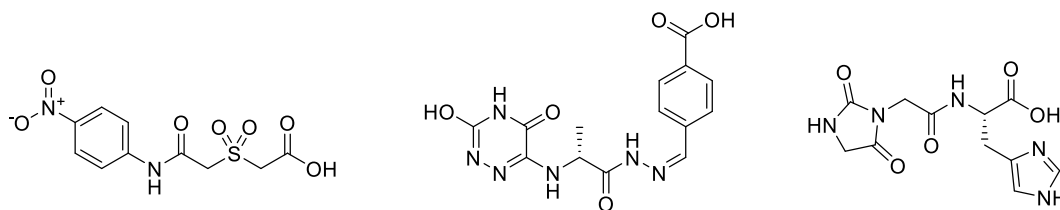
Figure 3.9. Select examples of decoy molecules generated using the DUD-E database.

Although the library of actives was limited due to the availability of experimental data, the docking method showed promising screening power against the TPP riboswitch, resulting in an excellent AUC of 0.90 (**Figure 3.10**). However, we were aware that the presence of phosphates in most of the active molecules and not in most of the decoys may represent a significant bias.
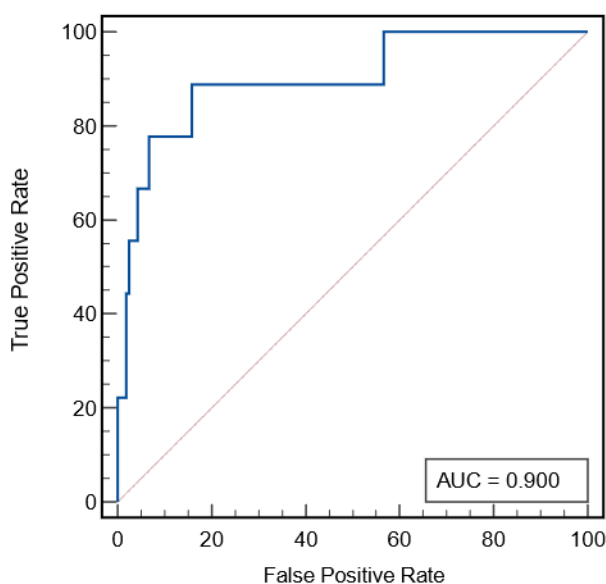
Figure 3.10. ROC curve for TPP riboswitch actives and decoys.

3.3.2   Virtual Screening and Analysis

With these validations complete, a virtual library of compounds was assembled for prospective docking to the TPP riboswitch. The initial library included compounds with known antibacterial activity from the CHEMBL repository, a manually curated database of bioactive molecules reported in the literature. The search from CHEMBL was also filtered for molecules based on

drug-like properties (between 300 - 550 Da, fewer than 10 rotatable bonds), TPP-like properties (at least 2 hydrogen bond acceptors to coordinate to magnesium of the pyrophosphate sensor helix, at least 1 hydrogen bond donor for pyrimidine sensor helix) and antibacterial activity (greater than 80% inhibition in one of the 48 pathogenic bacteria that contain the TPP riboswitch).[21] This resulted in an initial antibacterial library of 2590 compounds.
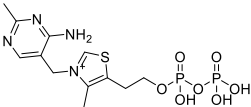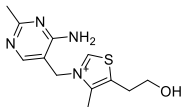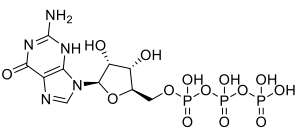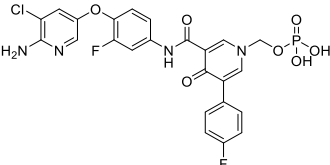
Several commercial vendors have pre-curated libraries of chemicals based on known active compounds against specific target classes. For example, the ChemSpace RNA-focused library used 170 compounds with known activity against RNA in order to select a subset of compounds (2400) from their entire screening library (3.9 M). The ChemSpace RNA-focused library was quite small (only 2400 compounds), therefore an additional RNA-focused library from Life Chemicals was employed (4452 compounds). Given the presence of two Mg^{2+} ions that mediate TPP riboswitch binding, it was also envisioned that compounds containing metal-chelating groups may be advantageous as TPP riboswitch binders. The ChemSpace metal chelator subset (130k) was therefore also incorporated in the initial library. The commercial libraries were filtered using the same drug-like and TPP-like properties as with the antibacterial library, and combined with the pre-filtered antibacterial library. Molecules containing functional groups that were too reactive to be included in potential drugs were removed and 6000 diverse compounds were selected. The library was also seeded with the previously identified 9 active compounds in order to assess how well the docking protocol was able to enrich the top scoring compounds with known actives.
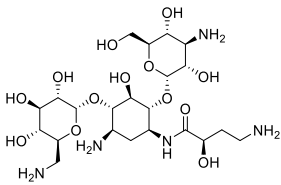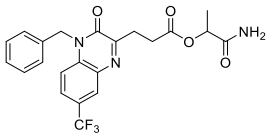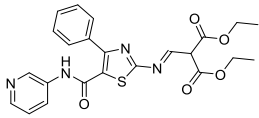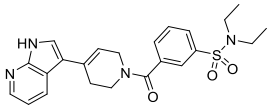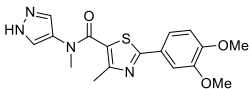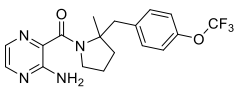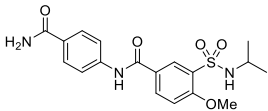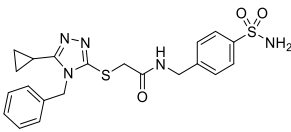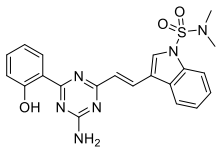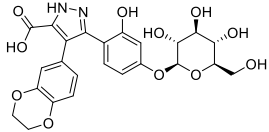
The virtual library of compounds was docked to a crystal structure of the TPP riboswitch aptamer domain bound to its native ligand (PDB: 2GDI), using the protocol developed previously for the self-docking and retrospective virtual screening studies. The docking score was used to automatically select the top 250 scoring compounds, from which 20 compounds were selected for experimental testing based on visual evaluation of key binding interactions between the docked ligand and the pyrimidine and pyrophosphate sensor helices. The docking score was also used to assess the enrichment factor of the docking protocol, evaluating the concentration of known actives within the top scoring compounds compared to their concentration in the entire library. Promisingly, 7 of the 9 known actives were ranked within the top 1% of the library, resulting in an enrichment factor of 78 (a very high enrichment factor when compared to the theoretical maximum of 100).
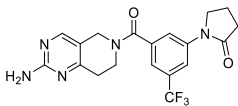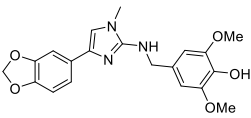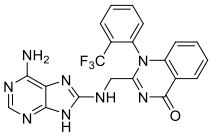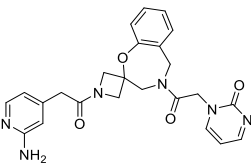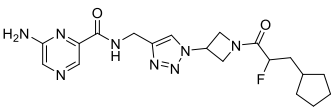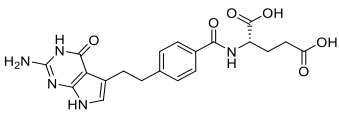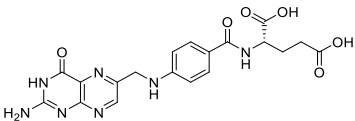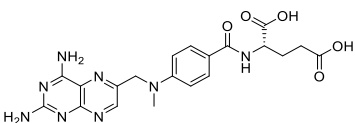
3.3.3 Biophysical Evaluation of Virtual Screening Hits

Surface plasmon resonance (SPR) was used to assess the binding of the selected compounds to the aptamer domain of the *thi*M TPP riboswitch. The biophysical method was first validated with two known ligands – the native ligand TPP and the related compound thiamine. Based on the SPR response curves, a K_d of 0.82 nM was calculated for TPP and 370 nM for thiamine. These are similar to their literature values of approximately 8 nM and 1.5 μM, respectively.[32, 34, 35] As a note, it has previously been observed that binding affinities of ligands to their corresponding aptamers tends to decrease as the length of the aptamer increases.[6, 36] Therefore, it is not unexpected that the K_d values determined for binding to the *thi*M aptamer domain are slightly lower than those reported in the literature for binding to the full length sequence (*thiM* aptamer domain + expression platform). The selected compounds were tested against the TPP riboswitch at concentrations up to 35 μM and the results shown in **Table 3.1**.

Table 3.1. Biophysical evaluation of virtual screening compounds.

Entry	ID	Structure	Activity
1	**TPP**		$K_d = 0.82$ nM
2	**Thiamine**		$K_d = 370$ nM
3	**3.12**		n.r.[a]
4	**3.13**		n.r. [a]

5	**3.14**		n.r.[a]
6	**3.15**		n.r.[a]
7	**3.16**		n.r.[a]
8	**3.17**		n.r.[a]
9	**3.18**		n.r.[a]
10	**3.19**		n.r.[a]
11	**3.20**		n.r.[a]
12	**3.21**		n.r.[a]
13	**3.22**		n.r.[a]
14	**3.23**		$K_d = 170\ \mu M$

15	**3.24**		n.r.[a]
16	**3.25**		n.r.[a]
17	**3.26**		n.r.[a]
18	**3.27**		n.r.[a]
19	**3.28**		n.r.[a]
20	**Pemetrexed**		n.r.[a]
21	**Folic Acid**		n.r.[a]
22	**Methotrexate**		n.r.[a]

[a] No response.

Out of 20 compounds tested, 19 showed no detectable response in the sensorgram, indicating that these molecules do not interact with the TPP riboswitch aptamer domain. Somewhat surprisingly, none of entries 20-22 (**pemetrexed**, **folic acid**, **methotrexate**) displayed any significant interaction with the riboswitch. Based on the structure and predicted binding pose, we thought that the aromatic moiety would form favourable interactions within the pyrimidine sensor helix and the dicarboxylic acid tail would be able to mimic the pyrophosphate group to form similar interactions

with Mg^{2+}. Similarly, neither of the phosphate-containing compounds (**3.12** and **3.13**) displayed riboswitch binding affinity.

Promisingly, one compound (**3.23**) displayed potential binding in the initial screen. Further investigation using a concentration series of **3.23** demonstrated that this compound binds to the aptamer domain in a dose-dependent manner with a K_d value of 170 μM (**Figure 3.11**). This affinity is in the same range as the best binders not featuring phosphates (**Figure 3.4**).
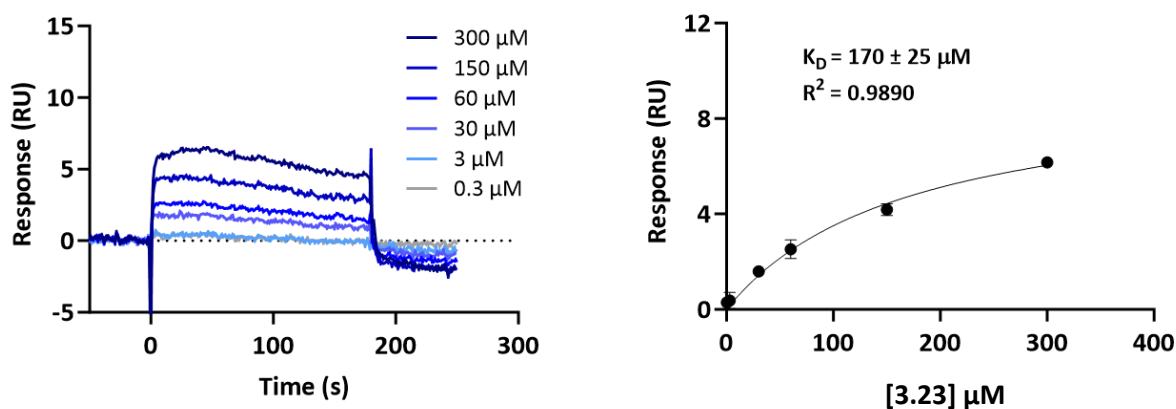

Figure 3.11. SPR sensorgram and dose-response curves for **3.23**.

From the virtual screening results, compound **3.23** is predicted to bind the pyrimidine sensor helix through its benzodioxane moiety, forming hydrogen bonds with G40 and C57 and pi-stacking with A43. The pyrazole group is predicted to form a hydrogen bond with the backbone of C74, with the carboxylate hydrogen bonding with A43 and U39. The glucose moiety is predicted to bind one of the two magnesium ions in the active site, as well as forming hydrogen bond interactions with the phosphate groups of C58 and U59.
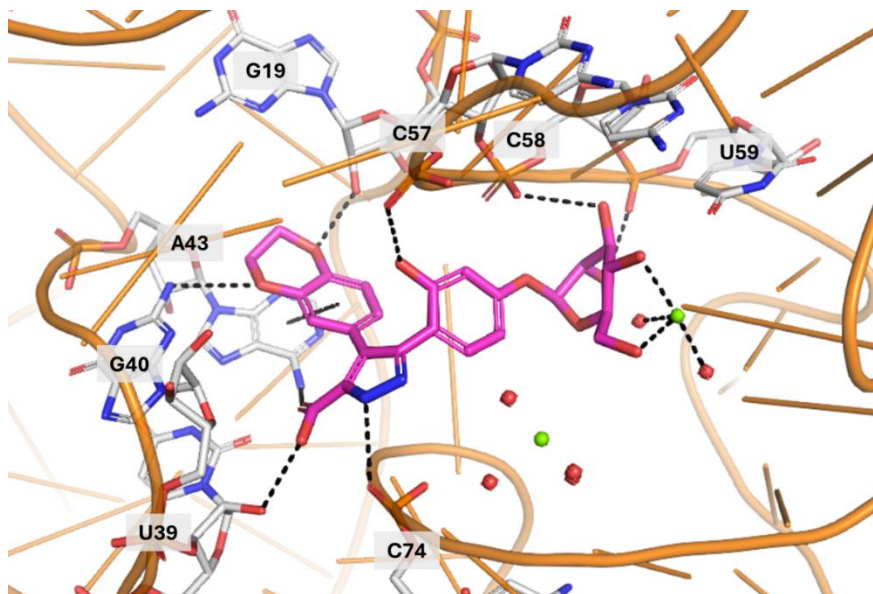
Figure 3.12. Predicted binding mode of **3.23**.

While **Figure 3.12** is a *predicted* pose from docking, the possibility of magnesium chelation by glucose is supported by literature reviews describing carbohydrate-based coordination complexes.[37, 38] Additionally, the crystal structure of mycinamicin III (PDB: 4X7Z) displays Mg^{2+} chelation by a monosaccharide group, while that of glucose isomerase (PDB: 4ZB0) shows Mn^{2+} chelation by glucose (**Figure 3.13**).
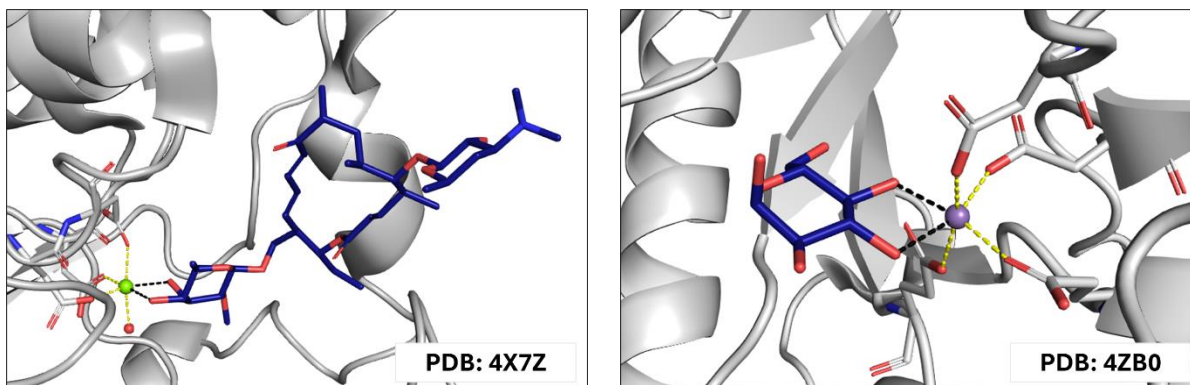

Figure 3.13. Crystal structures displaying monosaccharide-based metal chelation.

3.4 Conclusion and Future Work

This project described a successful virtual screening campaign to identify a preliminary

micromolar binder of the TPP riboswitch that is chemically distinct from the native ligand and from known binders. By using our in-house protein-based docking program FITTED that was modified for use with nucleic acids, this work demonstrates the possibility of extending *in silico* methods developed for proteins to novel nucleic acid targets. It is one of the first examples detailing the use of virtual screening to identify novel riboswitch binders and deviates from the close structural analogues of previously reported TPP inhibitors.

Although not described in this chapter, **3.23** was successfully resynthesized by another graduate student and displayed the same binding affinity for the TPP riboswitch, validating the activity observed in the preliminary biophysical screen. While the data obtained from SPR is promising, it only determines the ligand binding affinity and does not guarantee that **3.23** is binding in a similar manner to TPP to stabilize the riboswitch in the "OFF" state. Therefore, future work involving the biochemical evaluation of **3.23** on TPP riboswitch-controlled gene expression is envisioned. Additionally, the synthesis of analogues of **3.23** is currently underway in order to investigate structure-activity relationships, focusing on modifications to the benzodioxane and glucose moieties.

3.5 Methodology

3.5.1 Virtual Screening

Libraries for screening.

A virtual library of compounds was assembled for docking to the TPP riboswitch. The initial library included compounds with known antibacterial activity (from CHEMBL repository), RNA-binding motifs (from ChemSpace and Life Chemicals compound libraries), or metal-binding motifs (from ChemSpace compound library). This initial library was filtered based on drug-like properties to select compounds with fewer than 10 rotatable bonds and a molecular weight between $300 - 550$ Da using our programs SMART to compute molecular properties and REDUCE to select molecules fulfilling these criteria. Based on the binding motif of TPP, the library was also filtered to select molecules that contain at least two metal coordinating groups that would be necessary for coordination to the magnesium ions of the pyrophosphate sensor helix. The library was also filtered to include molecules that contained at least one hydrogen bond donor that would be necessary to facilitate key hydrogen bond interactions in the pyrimidine sensor helix. Molecules containing

functional groups that were too reactive to be included in potential drugs were removed and 6,000 diverse compounds were selected using our program SELECT and ECFP4 fingerprints.

Docking and Virtual Screening

All docking calculations were performed using the x-ray crystal structure 2GDI. Our program SMART was used to prepare the libraries for docking using default parameters. PREPARE and PROCESS were used to prepare the riboswitches for docking using default parameters from crystal structure. FITTED was used for docking, with the population size determined automatically for each ligand based on torsions. The pharmacophore constraint was generated by using a 2.5 Å hydrophobic interaction bead in the pyrimidine sensor helix and a 4.0 Å hydrogen bond acceptor interaction bead at each of the two magnesium ions. The FITTED Score was used to automatically select the top 250 scoring compounds. These compounds were visually inspected, and 20 compounds were selected for experimental testing based on manual evaluation of key binding interactions between the docked ligand and the riboswitch.

3.5.2 Biophysical Evaluation

Preparation of Ligands

All compounds for experimental evaluation were purchased from ChemSpace. Molecules that were water soluble were dissolved in RNAse-free water to create 5 mM stock solutions. All other molecules were dissolved in 100% DMSO to generate 5 mM stock solutions. The stock solutions were further diluted with 1X HBS-N and 5 mM $MgCl_2$ buffer to obtain no more than 0.1% DMSO.

Surface Plasmon Resonance

All compounds were tested at concentrations up to 35 µM via surface plasmon resonance, using the aptamer domain of the TPP riboswitch immobilized on the SPR chip with a 24-mer-polyA tail. For compound **3.23**, the ligand was tested at 6 concentrations ranging from 0.3 to 300 µM in order to generate a dose-response curve. The Biacore X100 Evaluation Software version 2.0 (GE Healthcare) was used for processing and analyzing the data and a specific binding equation was used to obtain the K_d. Experimental details can be found in NAR2024 and will not be discussed herein.

3.6   References

1.   Aslam, B.; Wang, W.; Arshad, M. I.; Khurshid, M.; Muzammil, S.; Rasool, M. H.; Nisar, M. A.; Alvi, R. F.; Aslam, M. A.; Qamar, M. U., et al., Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist* **2018,** *11*, 1645-1658.

2.   Bartlett, J. G.; Gilbert, D. N.; Spellberg, B., Seven Ways to Preserve the Miracle of Antibiotics. *Clin Infect Dis* **2013,** *56* (10), 1445-1450.

3.   Plackett, B., Why big pharma has abandoned antibiotics. *Nature* **2020,** *586*, S50-S52.

4.   *Antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline.*; World Health Organization (WHO): 2019.

5.   Shi, Z.; Zhang, J.; Tian, L.; Xin, L.; Liang, C.; Ren, X.; Li, M., A Comprehensive Overview of the Antibiotics Approved in the Last Two Decades: Retrospects and Prospects. *Molecules* **2023,** *28* (4).

6.   Winkler, W.; Nahvi, A.; Breaker, R. R., Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **2002,** *419* (6910), 952-956.

7.   Panchal, V.; Brenk, R., Riboswitches as Drug Targets for Antibiotics. *Antibiotics (Basel, Switzerland)* **2021,** *10* (1).

8.   Baird, N. J.; Kulshina, N.; Ferre-D'Amare, A. R., Riboswitch function: flipping the switch or tuning the dimmer? *RNA Biol* **2010,** *7* (3), 328-32.

9.   Blouin, S.; Mulhbacher, J.; Penedo, J. C.; Lafontaine, D. A., Riboswitches: ancient and promising genetic regulators. *ChemBioChem* **2009,** *10* (3), 400-16.

10.   Breaker, R. R., Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **2012,** *4* (2).

11.   Bastet, L.; Turcotte, P.; Wade, J. T.; Lafontaine, D. A., Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. *RNA Biol* **2018,** *15* (6), 679-682.

12.   Bedard, A. V.; Hien, E. D. M.; Lafontaine, D. A., Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochim Biophys Acta Gene Regul Mech* **2020,** *1863* (3), 194501.

13.   McCown, P. J.; Corbino, K. A.; Stav, S.; Sherlock, M. E.; Breaker, R. R., Riboswitch diversity and distribution. *RNA* **2017,** *23* (7), 995-1011.

14.   Pavlova, N.; Kaloudas, D.; Penchovsky, R., Riboswitch distribution, structure, and function in bacteria. *Gene* **2019,** *708*, 38-48.

15.   Lim, J.; Winkler, W. C.; Nakamura, S.; Scott, V.; Breaker, R. R., Molecular-recognition characteristics of SAM-binding riboswitches. *Angew Chem Int Ed Engl* **2006,** *45* (6), 964-8.

16.   Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R., Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **2003,** *113* (5), 577-86.

17.	Thore, S.; Leibundgut, M.; Ban, N., Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **2006,** *312* (5777), 1208-11.

18.	Lee, E. R.; Blount, K. F.; Breaker, R. R., Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* **2009,** *6* (2), 187-94.

19.	Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A., et al., Selective small-molecule inhibition of an RNA structural element. *Nature* **2015,** *526* (7575), 672-7.

20.	Sudarsan, N.; Cohen-Chalamish, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R., Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem Biol* **2005,** *12* (12), 1325-35.

21.	Pavlova, N.; Penchovsky, R., Genome-wide bioinformatics analysis of FMN, SAM-I, glmS, TPP, lysine, purine, cobalamin, and SAH riboswitches for their applications as allosteric antibacterial drug targets in human pathogenic bacteria. *Expert Opin Ther Tar* **2019,** *23* (7), 631-643.

22.	Yamauchi, T.; Miyoshi, D.; Kubodera, T.; Nishimura, A.; Nakai, S.; Sugimoto, N., Roles of Mg2+ in TPP-dependent riboswitch. *FEBS Lett* **2005,** *579* (12), 2583-8.

23.	Padhi, S.; Pradhan, M.; Bung, N.; Roy, A.; Bulusu, G., TPP riboswitch aptamer: Role of Mg(2+) ions, ligand unbinding, and allostery. *J Mol Graph Model* **2019,** *88*, 282-291.

24.	Kumar, S.; Reddy, G., TPP Riboswitch Populates Holo-Form-like Structure Even in the Absence of Cognate Ligand at High Mg(2+) Concentration. *J Phys Chem B* **2022,** *126* (12), 2369-2381.

25.	Chen, L.; Cressina, E.; Dixon, N.; Erixon, K.; Agyei-Owusu, K.; Micklefield, J.; Smith, A. G.; Abell, C.; Leeper, F. J., Probing riboswitch-ligand interactions using thiamine pyrophosphate analogues. *Org Biomol Chem* **2012,** *10* (30), 5924-31.

26.	Cressina, E., Chen, L., Abell, C., Leeper, F.J., Smith, A.G., Fragment screening against the thiamine pyrophosphate riboswitch *thi*M. *Chem Sci* **2011,** (2), 157-165.

27.	Daldrop, P.; Reyes, F. E.; Robinson, D. A.; Hammond, C. M.; Lilley, D. M.; Batey, R. T.; Brenk, R., Novel ligands for a purine riboswitch discovered by RNA-ligand docking. *Chem Biol* **2011,** *18* (3), 324-35.

28.	Kallert, E.; Fischer, T. R.; Schneider, S.; Grimm, M.; Helm, M.; Kersten, C., Protein-Based Virtual Screening Tools Applied for RNA-Ligand Docking Identify New Binders of the preQ(1)-Riboswitch. *J Chem Inf Model* **2022,** *62* (17), 4134-4148.

29.	Labarre, A.; Stille, J. K.; Patrascu, M. B.; Martins, A.; Pottel, J.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 8. Forming New Bonds─Challenges and Opportunities. *J Chem Inf Model* **2022,** *62* (4), 1061-1077.

30.	Wei, W.; Luo, J.; Waldispuhl, J.; Moitessier, N., Predicting Positions of Bridging Water Molecules in Nucleic Acid-Ligand Complexes. *J Chem Inf Model* **2019,** *59* (6), 2941-2951.

31.	Moitessier, N.; Westhof, E.; Hanessian, S., Docking of aminoglycosides to hydrated and flexible RNA. *J Med Chem* **2006,** *49* (3), 1023-33.

32.	Agrawal, A.; Johnson, S. L.; Jacobsen, J. A.; Miller, M. T.; Chen, L. H.; Pellecchia,

M.; Cohen, S. M., Chelator fragment libraries for targeting metalloproteinases. *ChemMedChem* **2010,** *5* (2), 195-9.

33.   Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **2012,** *55* (14), 6582-94.

34.   Chang, A. L.; McKeague, M.; Liang, J. C.; Smolke, C. D., Kinetic and equilibrium binding characterization of aptamers to small molecules using a label-free, sensitive, and scalable platform. *Anal Chem* **2014,** *86* (7), 3273-8.

35.   Kulshina, N.; Edwards, T. E.; Ferre-D'Amare, A. R., Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA* **2010,** *16* (1), 186-96.

36.   Winkler, W. C.; Cohen-Chalamish, S.; Breaker, R. R., An mRNA structure that controls gene expression by binding FMN. *P Natl Acad Sci USA* **2002,** *99* (25), 15908-15913.

37.   E. Alekseev, Y.; D. Garnovskii, A.; A. Zhdanov, Y., Complexes of natural carbohydrates with metal cations. *Russ Chem Rev* **1998,** *67* (8), 649-669.

38.   Gyurcsik, B.; Nagy, L., Carbohydrates as ligands: coordination equilibria and structure of the metal complexes. *Coord Chem Rev* **2000,** *203* (1), 81-149.

CHAPTER 4:

DEVELOPMENT OF COVALENT COVID-19 ANTIVIRALS TARGETING 3CL[PRO]

The work in this chapter is based, in part, on work published in:

J. K. Stille‡ , J. Tjutrins‡ , G. Wang‡ , F. A. Venegas‡ , C. Hennecker, A. M. Rueda, I. Sharon, N. Blaine, C. E. Miron, S. Pinus, A. Labarre, J. Plescia, M. Burai Patrascu, X. Zhang, A. S. Wahba, D. Vlaho, M. J. Huot, T. M. Schmeing, A. K. Mittermaier, and N. Moitessier. Design, synthesis and in vitro evaluation of novel SARS-CoV-2 3CL[pro] covalent inhibitors. *Eur. J. Med. Chem.*, **2022**, *229*, 114046

Contribution of Authors

J.K. Stille contributed to the design of all compounds and performed all docking studies. **J.K. Stille** synthesized and characterized compounds **4.9, 4.11, 4.15, 4.16, 4.18-4.22** and all related intermediates.

All remaining compounds were synthesized and characterized by J. Tjutrins.

In vitro testing (% inhibition and IC$_{50}$) was performed by G. Wang and F.A. Venegas. All kinetic assays were performed by **J.K. Stille**.

4.1 Abstract

Severe diseases such as COVID-19 infection, as well as the previous SARS and MERS outbreaks, are the result of coronavirus infections and have demonstrated the urgent need for antiviral drugs to combat these deadly viruses. Due to its essential role in viral replication and function, $3CL^{pro}$ (main coronaviruses cysteine-protease) has been identified as a promising target for the development of antiviral drugs. Previously reported SARS-CoV $3CL^{pro}$ non-covalent inhibitors were used as a starting point for the development of covalent inhibitors of SARS-CoV-2 $3CL^{pro}$. This chapter focuses on the exploration and optimization of covalent warheads in the development of sub-micromolar covalent $3CL^{pro}$ inhibitors.

4.2 Introduction

4.2.1 Coronaviruses and COVID-19

Coronaviruses (CoV) are a large family of viruses associated with some forms of common colds (together with rhinoviruses, respiratory syncytial virus, adenoviruses and others), as well as far more serious diseases including Severe Acute Respiratory Syndrome (SARS, caused by SARS-CoV infection), which made headlines worldwide in 2002-2003 with over 700 deaths including 43 in Canada,[1] and the Middle East Respiratory Syndrome (MERS, caused by MERS-CoV infection), which was reported in Saudi Arabia in 2012 and killed over 900.[2] The outbreak of novel coronavirus (COVID-19, caused by SARS-CoV-2 infection), its numerous variants, and the discovery of animal reservoirs provide significant motivation for the development of potent therapeutics against these viruses to prevent future outbreaks.[3, 4]

SARS, MERS, and COVID-19 are respiratory illnesses characterized by fever, cough, and shortness of breath, posing significant danger to patients. The case fatality rates for those infected with SARS-CoV and MERS-CoV were estimated at about 10% and 35%, respectively.[1, 2] Estimates for SARS-CoV-2 are ranging anywhere from of 0.1 to 25% depending on the age group, the country and the stage of the pandemic.[5, 6] In contrast to SARS and MERS, COVID-19 has rapidly spread worldwide despite the severe restrictions imposed in many countries, and the official number of deaths now exceeds 7.0 million[7] (which is a well underestimated number as shown by excess mortality studies).[8]

While vaccines have been a central pillar of the efforts against the COVID-19 pandemic, therapeutics offer a complementary approach with many distinct advantages. For example, oral therapeutics tend to be easy to store and administer and need only be given to the small minority of patients suffering more serious symptoms. In contrast, a large proportion of the population must be inoculated for vaccines to be effective, and mRNA-based vaccines require complex logistics to maintain the cold chain, leading to enormous challenges in production, supply and administration. In addition, large vaccine campaigns require public compliance and amplifies the number of people suffering from adverse reactions to medication. To add to these difficulties, Pfizer announced that the immunity of their vaccine drops after about 6 months, suggesting that regular injections would be needed, further amplifying the public compliance issue and burden to public health systems.[9, 10] Importantly, vaccines primarily induce an immune response against the spike protein,[11] while future variants of concern may have mutations in this protein that could allow them to evade immunity. In contrast, antiviral therapeutics can target a wide range of proteins including viral proteases (3CL^{pro}, PL^{pro}), the RNA-dependent RNA polymerase (RdRp) and RNA helicase. Therefore, they can be equally effective against strains of the virus with mutations that escape spike-based vaccination or herd immunity. Overall, it is clear that effective therapeutics would be complementary to mass vaccination. Finally, some groups (pregnant and breastfeeding women, people with allergies, young children, immunocompromised patients or people with other conditions) may be at risk or not responsive to vaccines, and alternative treatments (e.g., oral therapeutics) must be available.[12] Consequently, major efforts from a large number of research groups focused on the development of small molecules as antivirals against SARS-CoV-2, resulting in the recent approval of Molnupiravir and Nirmaltrelvir.[13, 14]

4.2.2 3CL^{pro} as an Antiviral Target

Coronaviruses express 3-chymotrypsin-like cysteine protease (3CL^{pro}), also referred to as the main protease (M^{pro}) or nsp5 (non-structural protein 5), which features a Cys-His catalytic dyad (Cys^{145}, His^{41}) and is required for viral replication and infection. As shown in **Figure 4.1**, the life cycle of SARS-CoV-2 first involves the attachment and entry of the virus into the host cell. The viral RNA is translated into polyprotein chains which must then be cleaved by viral proteases in order to form the necessary viral proteins needed to continue the life cycle. 3CL^{pro} is responsible for the majority of polyprotein cleavage sites and therefore plays a crucial role in viral replication, making it an

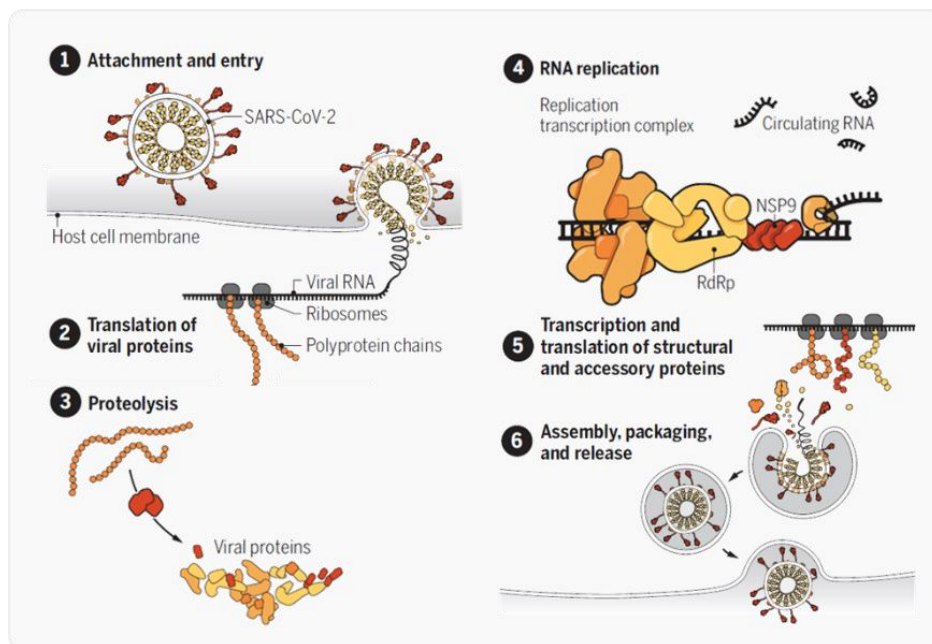attractive target for antiviral therapeutics.

Figure 4.1. Viral life cycle of SARS-CoV-2.[15]

3CL^{pro} enzymes were identified early on as attractive targets for antiviral development, resulting in several inhibitors and structures of SARS-3CL^{pro}-inhibitor complexes (e.g., PDB codes: 4TWY, 2ZU5, 2ALV [16]). The 3CL^{pro} enzymes from SARS-CoV and SARS-CoV-2 share nearly 80% sequence identity,[17, 18] suggesting that many of the lessons learned for developing SARS therapeutics can be applied to COVID-19 and to other coronaviruses (known or yet unknown).

It is also important to note that 3CL^{pro} is not limited to coronaviruses. It is also a target for the development of antivirals for several different viruses (including noroviruses,[19] enteroviruses,[20] and rhinoviruses[21]) and therefore 3CL^{pro} inhibitors could also find applications beyond the recent COVID-19 epidemic. Viruses are known to mutate rapidly and often require the use of therapeutic combinations to safeguard against drug resistance. Therefore, the development of multiple distinct inhibitors, ideally with non-overlapping resistance profiles, may be critical for preparedness against novel strains of SARS-CoV-2. Similarly, given the prevalence of related proteases among several different viruses, the development of a greater library of known protease inhibitors may serve as useful building blocks to combat future viral pandemics.

4.2.3 Covalent Inhibition

The quest for novel antivirals against SARS-CoV and, more recently, SARS-CoV-2 has been intense, and several viral enzyme inhibitors and crystal structures of enzyme-inhibitor complexes were quickly reported (eg, PDB codes: 6LU7 [22], 6M2N [23], 6XQU [24], 6WQF [25]).[26-28] The presence of a catalytic cysteine residue in the active site makes 3CL[pro] amenable to covalent inhibition, a strategy that was successfully employed following the SARS-CoV pandemic. In fact, many of the reported SARS-CoV inhibitors feature an electrophilic group, such as an α-ketoamide, epoxide, aziridine, α,β-unsaturated ester (Michael acceptor), or α-fluoroketone, which forms a covalent bond with the catalytic cysteine residue (Cys[145]), as confirmed by X-ray crystallography (e.g., PDB code: 5N19).[26] A crystal structure of the SARS-CoV-2 3CL[pro] with a covalent peptidic inhibitor bound to Cys[145] was quickly elucidated (PDB code: 6LU7). This pseudo-peptidic inhibitor, an analogue of **Rupintrivir** (tested on SARS[29] and COVID-19[30]), has been the starting point for a number of drug discovery campaigns.[31-35] **PF-0730814** was reported as a promising clinical candidate and entered clinical trials as an intravenous treatment option for COVID-19.[33, 35, 36] **PF-07321332** (nirmatrelvir) was subsequently developed as an orally bioavailable 3CL[pro] inhibitor and was authorized for emergency use in December 2021 before receiving FDA approval in May 2023 under the name **Paxlovid** (nirmatrelvir + ritonavir).[14, 37, 38]

Investigations from a group of Canadian researchers identified other warheads for this lead molecule with potential for further development.[39, 40] The identification of a potent warhead was also the focus of Hilgenfeld and co-workers.[41]

The structurally similar **GC376** was originally identified as active against a feline coronavirus[42] and more recently confirmed as a SARS-CoV-2 3CL[pro] inhibitor,[28] and structure-activity relationship studies led to improved analogues.[43] Smaller, more drug-like inhibitors such as the isatine derivative **2** have been devised for SARS-CoV.[44, 45] Jorgensen and co-workers successfully converted Perampanel, a known antiepileptic drug that is also a weak 3CL[pro] inhibitor, into potent inhibitors (e.g., **3**, **Figure 4.2**) using a combination of computational and experimental investigation.[46, 47] This series was further optimized by Jacobs *et. al.,* leading to analogues with improved metabolic stability.[48] A similar scaffold was pursued by researchers at Shionogi, resulting in the clinical candidate **S-217622** (ensitrelvir) which was approved for emergency

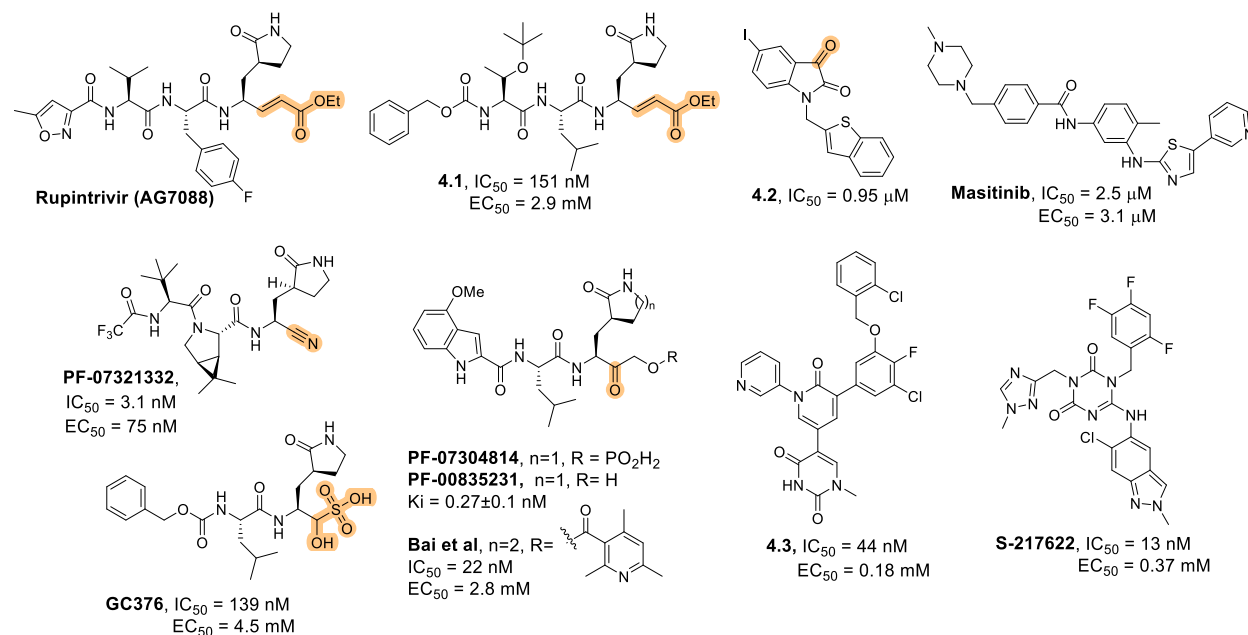authorization in Japan following promising phase 2/3 clinical trials.[49, 50]

Figure 4.2. Reported covalent SARS-CoV-2 3CL[pro] inhibitors **4.1,**[32] **GC376,**[28] **PF-07321332**,[14] **PF-0730814**[33, 35] and an analogue from Bai *et al.*[39] Reported non-covalent SARS-CoV-2 3CL[pro] inhibitors **Masitinib,**[51] **4.3,**[47] and **S-217622**.[49] Reported covalent inhibitor of SARS-CoV 3CL[pro] **4.2**.[44] Warheads for covalent binding are highlighted.

As described in our recent review,[52] covalent drugs can be extremely effective and useful pharmaceuticals, yet they have been largely ignored in most drug design endeavours and particularly in those concerning structure-based drug design. Concerns about their potential off-target reactivity and toxicity have often been raised.[53] Despite these concerns, there are many examples of covalent drugs on the market, including two of the ten most widely prescribed medications in the U.S., as well as several other common drugs like aspirin and penicillin.[52] The advantages of covalent drugs are becoming increasingly recognized: they have extremely high potencies, long residence times, and high levels of specificity.[54] Although skepticism persists, many pharmaceutical companies are embracing covalent drugs as exemplified by **Neratinib** (Nerlynx®, Pfizer) and **Afatinib** (Gilotrif®, Boehringer-Ingelheim).

4.3 Computer-Aided Design

4.3.1 Docking-Guided Optimization of a Non-Covalent Inhibitor

Many of the structure-based studies related to COVID-19 to date have employed virtual screening and machine learning techniques. Several *potential* 3CL[pro] inhibitors have been identified, however experimental verification has lagged.[55-57] As of today, much of the research has focused on peptidic substrate-like inhibitors. There is currently a need for the development of non-peptidic inhibitors with synthetically accessible scaffolds that will allow for more thorough investigations of structure-activity relationships. We thought to benefit from our team's expertise in covalent inhibition and from our software that enables automated docking and virtual screening of covalent inhibitors, which is not possible with most commercial packages.

An investigation of the crystal structure of a non-covalent inhibitor (**X77**, **Figure 4.3**) bound to 3CL[pro] of SARS-CoV-2 (PDB code: 6W63) suggested that it might be possible to modify this inhibitor by incorporating a covalent warhead in proximity to the catalytic cysteine residue. As shown in **Figure 4.3a**, the sulphur atom of Cys^{145} is positioned at 3.2 Å from the imidazole moiety. Thus, replacement of the imidazole with a covalent warhead appeared to be a promising strategy to improve the inhibitory potency of this non-covalent inhibitor (**Figure 4.3b**). Additionally, this scaffold could be prepared via a 4-component Ugi reaction, enabling a combinatorial approach that would provide an efficient synthetic method for preparing diverse analogues. This would provide a significant advantage in exploring structure-activity relationships when compared to previously reported inhibitors, as a wide range of covalent warheads could be readily incorporated into the same inhibitor scaffold. As a note, a consortium of research groups including a group at the Weizmann Institute of Science in Rehovot (Israel) took a very similar strategy although focusing primarily on non-covalent inhibitors.[58]
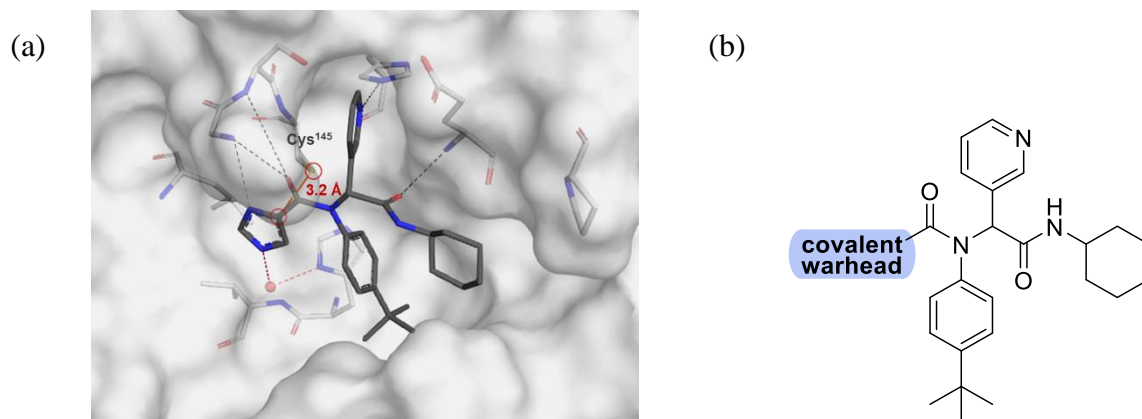
Figure 4.3. (a) Crystal structure of **X77** bound to SARS-CoV-2 3CL^pro (PDB: 6W63), with the proximity of Cys[145] to the inhibitor imidazole shown in red. (b) Design strategy for the incorporation of a covalent warhead in **X77**.

To validate the design strategy, a virtual library of modified inhibitors was prepared based on incorporation of covalent warheads that could be accessed via a traditional or modified Ugi 4 component coupling (4CC) reaction (Figure 3). These compounds were docked to 3CL^pro (PDB code: 6W63) using our docking program, FITTED.[59] The docked poses (**Figure 4.4**) suggested that many of these modified inhibitors would be able to maintain the same non-covalent interactions as the original non-covalent inhibitor while also positioning the warhead close enough to Cys[145] to facilitate the formation of a covalent bond. Based on the promising docking results with multiple warheads, a small library of analogues was synthesized for experimental testing.
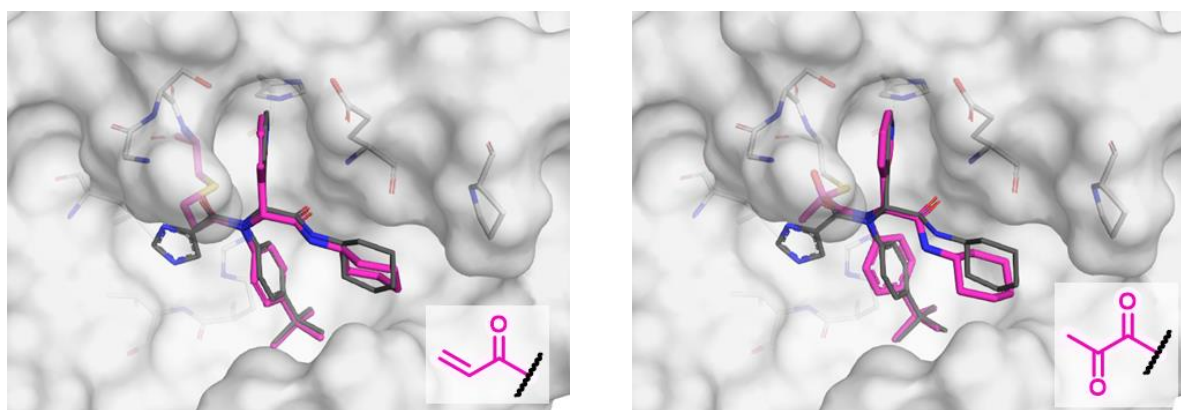
Figure 4.4. Selected docked binding modes of proposed covalent inhibitors (pink) overlaid

with the co-crystallized non-covalent inhibitor **X77** (grey).

4.4 Results

4.4.1 Synthesis of Warhead Library

Following a protocol reported by Jacobs *et al.*[60], a 4-component Ugi reaction was used to prepare analogues bearing four different classes of covalent warheads (alkene Michael acceptor, nitrile, alkyne Michael acceptor, and α-halo amide).
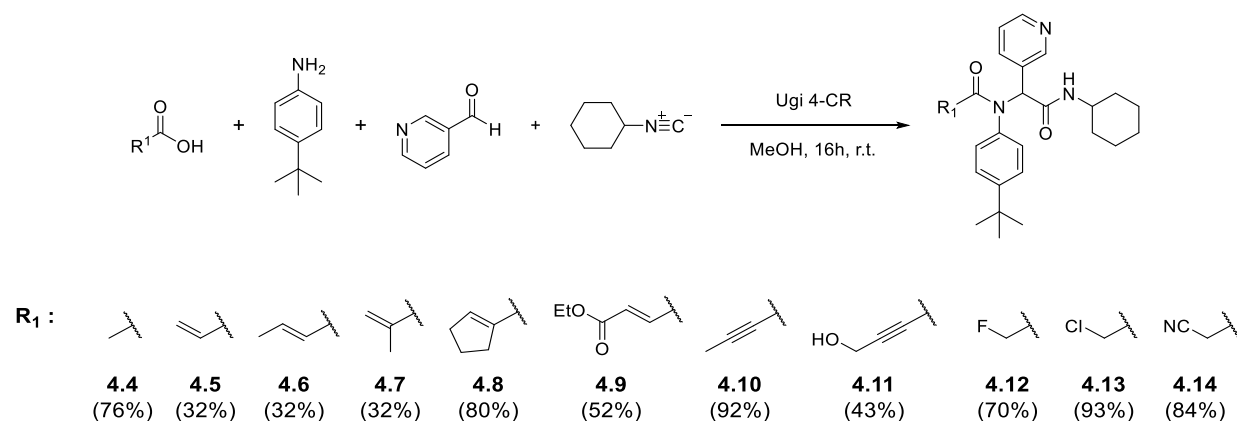

Scheme 4.1. Synthesis of inhibitors **4.4**, **4.5**, **4.6**, **4.7**, **4.8**, **4.9**, **4.10**, **4.11**, **4.12**, **4.13**, and **4.14** with yields shown in brackets.

The same conditions were applied to access the corresponding α-ketoamide, however it did not lead to any product formation. A review article by Marcaccini suggested that oxoacids have a tendency to react with the amine component in competing side reactions and that use of these starting materials may require preformation of the imine prior to addition of the oxoacid.[61] Based on this discussion, modified Ugi conditions were employed and gratifyingly led to the desired product **4.15** in moderate yield (**Scheme 4.2**). The same conditions were also applied to the synthesis of a trifluoromethyl analogue **4.16**.
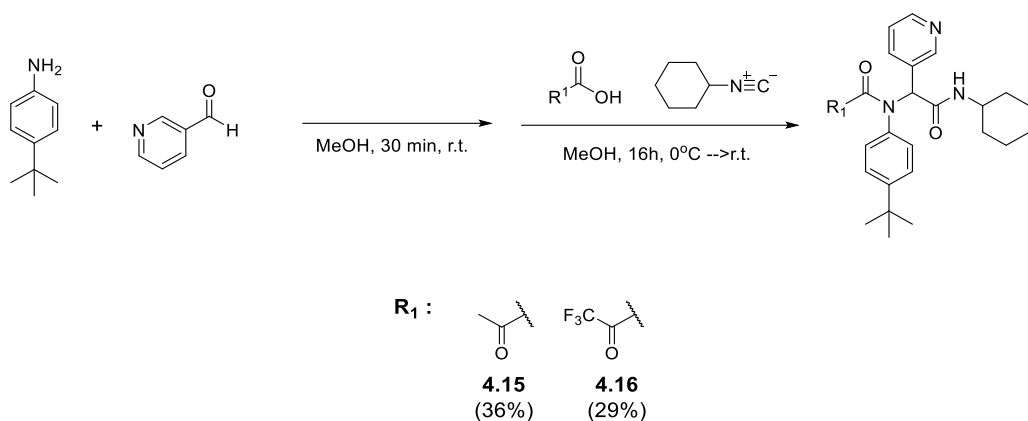
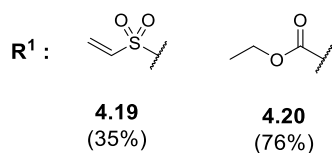**Scheme 4.2.** Synthesis of inhibitors **4.15** and **4.16**, yields shown in brackets.

A 3-component Ugi reaction using β-alanine as both the amine and carboxylic acid component was used to prepare the corresponding β-lactam analogue following a reported procedure (**Scheme 4.3**).[62]
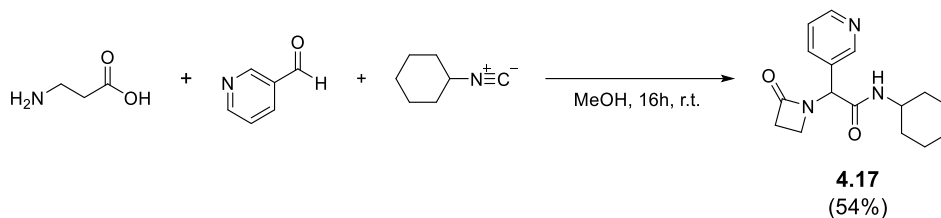
**Scheme 4.3.** Synthesis of inhibitor **4.17**, yield shown in brackets.

A modified 3-component Ugi reaction adapted from a protocol by Ding *et al.*[63] was applied to the synthesis of amine analogue **4.18**. This intermediate was then used to prepare inhibitors featuring vinyl sulfonamide (**4.19**) and ethyl carbamate (**4.20**) warheads (**Scheme 4.4**).

R^1 :

4.19
(35%)

4.20
(76%)
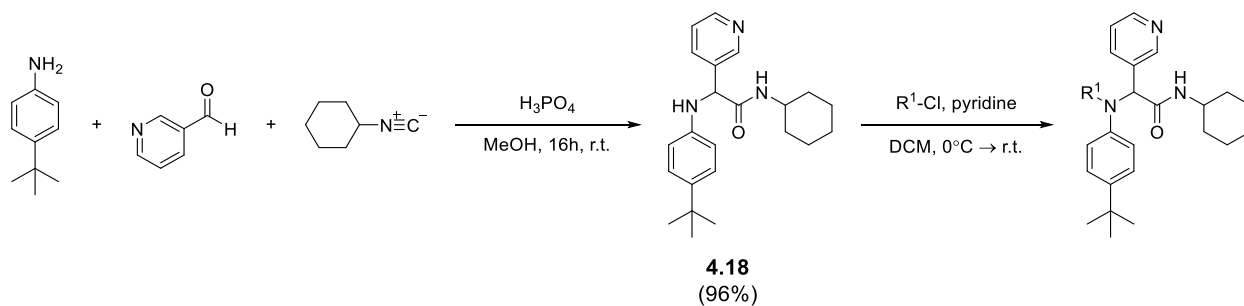
Scheme 4.4. Synthesis of intermediate **4.18**, and covalent inhibitors **4.19** and **4.20**.

The same modified 3-component Ugi protocol was used to prepare an additional nitrile analogue **4.21**, with the secondary amine starting material prepared by alkylation of 4-*t*Bu-aniline with bromoacetonitrile following reported literature conditions.[64]
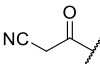

4.21
(68%)

4.22
(69%)
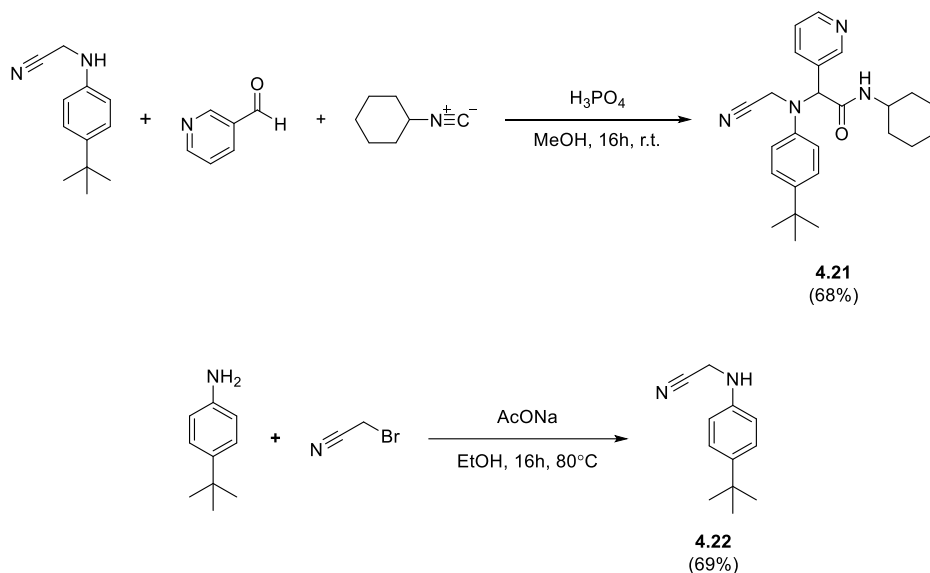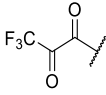
Scheme 4.5. Synthesis of nitrile analogue **4.21** and amine starting material **4.22**.

4.4.2   Biological Evaluation

Sixteen potential warheads (compounds **4.4-4.21**) and three non-covalent analogues (**X77**, **4.4** and **4.18**) were evaluated for their inhibitory potency using a fluorescence inhibition assay. The compounds were initially screened at 50 µM, and IC$_{50}$ values were subsequently determined for compounds displaying greater than 80% inhibition (**Table 4.1**).

Table 4.1. Inhibitory potency against SARS-CoV-2 3CL^{pro}. Evaluation of warheads (R^1).
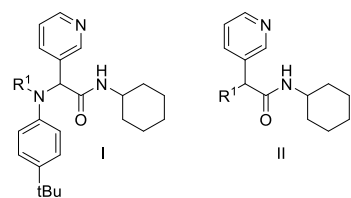
Entry	Scaffold	Compound	R^1	Inhibition (%)[a]	IC$_{50}$ (μM)
1	-	**GC376**	-	>95	0.11 ± 0.06
2	I	**X77**	-	>95	4.1 ± 1.2
3	I	**4.4**		25 ± 8	nd[b]
4	I	**4.18**		28 ± 3	nd[b]
5	I	**4.5**		84 ± 1	11.1 ± 1.5
6	I	**4.6**		63 ± 5	nd[b]
7	I	**4.7**		30 ± 1	nd[b]
8	I	**4.8**		55 ± 10	nd[b]
9	I	**4.9**		59 ± 6	nd[b]
10	I	**4.10**		> 95	5.3 ± 0.8
11	I	**4.11**		> 95	0.85 ± 0.42

12	I	**4.12**	F–CH2–C(=O)–	30 ± 7	nd[b]
13	I	**4.13**	Cl–CH2–C(=O)–	> 95	0.41 ± 0.13
14	I	**4.14**	NC–CH2–C(=O)–	74 ± 1	7.0 ± 0.2
15	I	**4.15**	α-ketoamide	92 ± 1	5.2 ± 1.2
16	I	**4.16**	F3C–C(=O)–C(=O)–	77 ± 4	12.4 ± 5.2
17	II	**4.17**	β-lactam	< 5	nd[b]
18	I	**4.19**	vinyl sulfonyl	> 95	0.42 ± 0.11
19	I	**4.20**	EtO–C(=O)–	35 ± 3	nd[b]
20	I	**4.21**	NC–CH2–CH2–	< 5	nd[b]

[a] The enzyme activity was measured with 114 nM 3CL[pro] and 50 μM of each potential inhibitor with incubation time of 30 min. [b] not determined.

We observed that known inhibitor **GC376** inhibited SARS-CoV-2 3CL[pro] with a similar potency in our assay as reported previously in the literature (IC_{50} = 110 nM vs 139 nM). Similarly, known SARS-CoV 3CL[pro] inhibitor **X77** (IC_{50} = 3.4 μM[65]) showed similar potency against 3CL[pro] of SARS-CoV-2 (IC_{50} = 4.1 μM). Gratifyingly, low micromolar to sub-micromolar potencies were also observed for the covalent analogues containing acrylamide (**4.5**), alkynylamide (**4.10**), α-chloroamide (**4.13**), α-ketoamide (**4.15**) and vinyl sulfonamide (**4.19**) warheads. Interestingly, our two most potent inhibitors **4.13** and **4.19** (IC_{50} = 0.4 and 0.5 μM) were an order of magnitude more potent than the original non-covalent hit molecule **X77**.

An increase in potency was observed when increasing the electrophilicity of the warhead – the α-chloroamide (**4.13**) was more active than the corresponding α-fluoroamide (**4.12**), and the vinyl sulfonamide (**4.19**) was more active than the corresponding acrylamide (**4.5**). However, this trend was not observed when increasing the electrophilicity of the ketoamide (**4.15**) with a CF_3 group (**4.16**), potentially due to an increase in the steric bulk and/or electrostatic properties of the warhead that are not tolerated in the active site. Similarly, while acrylamides are typically more reactive with cysteine than the corresponding alkynylamides (when tested in glutathione or cysteine binding assays)[66], the alkynylamide warhead (**4.10**) was more active against 3CL[pro]. A possible explanation could be that the sp geometry of the alkynylamide warhead positions the electrophile more favourably to the cysteine residue to facilitate covalent bond formation.

The binding pocket also seems to favour smaller warheads – any steric bulk around the acrylamide warhead resulted in a decrease in potency, regardless of electronics. As mentioned previously, a similar effect was observed when comparing the activity of inhibitor **4.16** and **4.15**. The position of the covalent bond formation also appears to influence inhibitor activity. Minimal inhibition was observed with compounds **4.20** and **4.17** which both involve the formation of a covalent bond directly with the carbonyl carbon. This carbon is positioned slightly further from the cysteine residue (3.4 Å in PDB: 6W63) and therefore covalent bond formation may result in the loss of other non-covalent interactions. Covalent bond formation appears to be equally tolerated at either the alpha position (**4.13** and **4.15**) or beta position (**4.5**, **4.10**, and **4.19**).

Another observation is the significant loss of potency when removing the heterocyclic ring of **X77** (**X77** vs. **4.4**, **4.18** in **Table 4.1**). As illustrated in **Figure 4.3**, the basic imidazole nitrogen of **X77** forms a hydrogen bond interaction with the backbone of Gly[143], an interaction that is also observed with the furan ring of **ML188** (PDB: 3V3M) or other heterocycles of the same chemical series.[60, 67] Gly[143], together with the backbone amides of Ser[144] and Cys[145], forms an oxyanion hole that contributes to the catalytic activity of this enzyme. Substitution of this heterocycle with a carbocycle of similar size but no hydrogen bonding groups (compounds **4.8** vs. **X77**) does not preserve the inhibitory potency even when this ring was converted to a warhead for covalent binding. Additionally, no difference in activity was observed when comparing the activity of compounds **4.4** and **4.18**, suggesting that the carbonyl group may not contribute significantly to the inhibitor's binding affinity.

As shown in **Figure 4.3**, **X77** interacts with the catalytic His[41] through a water-mediated hydrogen bond via a conserved water molecule. In an attempt to reproduce this interaction, longer covalent groups were designed by incorporating an ethyl ester to an acrylamide warhead (**4.9**) and by incorporating a hydroxyl group to an alkynylamide warhead (**4.11**). While **4.9** resulted in a loss of potency, **4.11** displayed a nearly 10-fold improvement in potency over the alkynylamide analogue **4.10**.

4.4.3   Confirmation of Binding Mode

To evaluate the covalent inhibition hypothesis, crystal structures of 3CL[pro] co-crystallized with **4.13** and **4.19** were obtained (**Figure 4.5**).
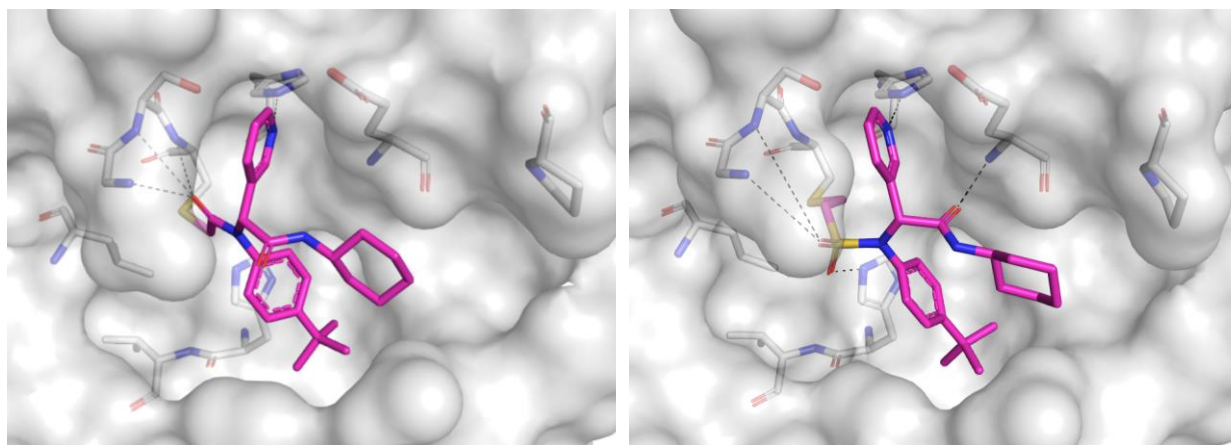

Figure 4.5. Crystal structure of 3CL[pro] bound with a) **4.13** (PDB: 7MLF), and b) **4.19** (PDB: 7MLG).

The co-crystallized structure of **4.13** shows that covalent bond formation at the alpha position results in a slight shift of the inhibitor towards the cysteine residue, resulting in the loss of a hydrogen bond interaction between the amide carbonyl of the inhibitor and the backbone of Glu[166]. This interaction was maintained in the co-crystallized structure of **4.19**, suggesting that covalent bond formation at the beta-position may be preferable for non-covalent binding affinity.

4.4.4   Evaluation of Binding Kinetics

In addition to the IC$_{50}$ values determined through biochemical assays, the binding kinetics of our

most promising covalent analogues were also evaluated experimentally. While the IC_{50} values determined through enzyme inhibition assays provide a valuable means to assess the inhibitor potency, this assessment can be complicated for covalent inhibitors. Covalent inhibitors display time-dependent inhibition, with greater inhibition (and therefore lower IC_{50} values) observed with longer incubation time, a trend that was observed when evaluating two of our most potent inhibitors (**4.13** and **4.19**, **Figure 4.6**).
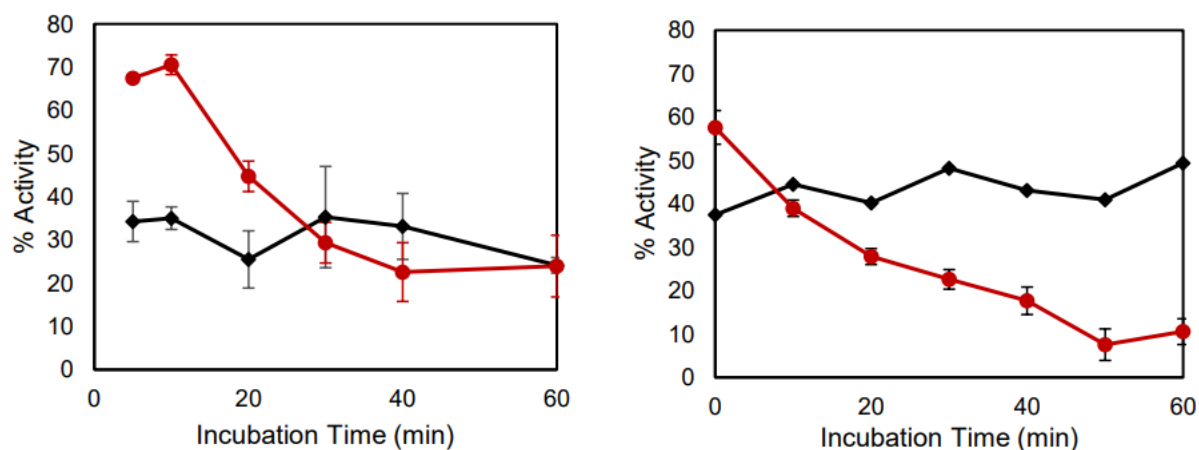
Figure 4.6. Time-dependent potency for **X77** (black) and **4.13** (left, red) and **4.19** (right, red).

However, the impact of incubation time on IC_{50} values may not be consistent among covalent analogues and therefore can complicate the assessment and prioritization of inhibitors when simply comparing IC_{50} values. This is due to the fact that most covalent inhibitors bind their target in a two-step process defined by an initial non-covalent binding mode followed by covalent bond formation (**Figure 4.7**), and improvements in IC_{50} values may be the result of improvement of either the non-covalent binding affinity (lower K_i) or more rapid covalent bond formation (greater k_{inact}).
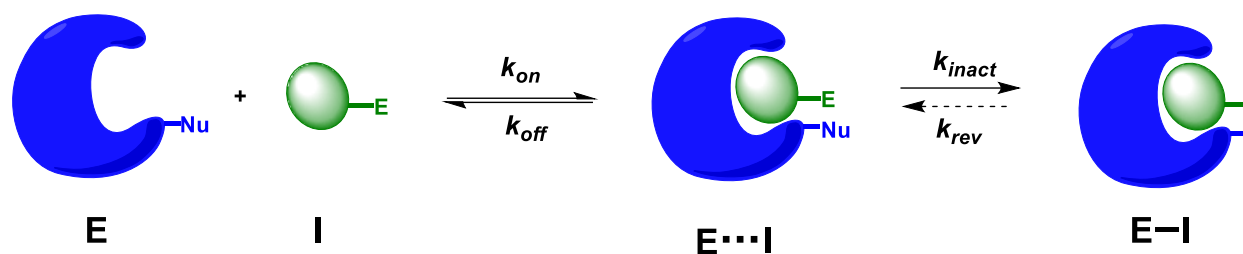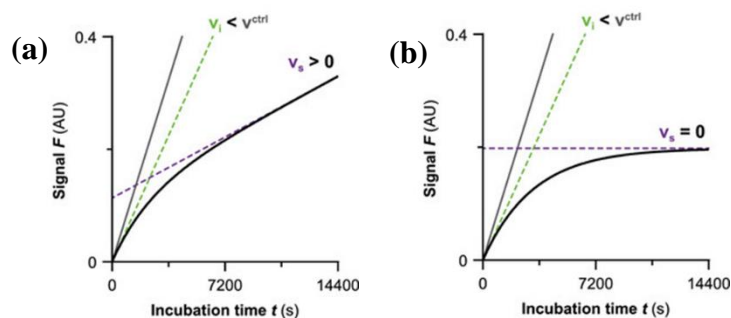
Figure 4.7. Two-step covalent binding mechanism.

When assessing covalent analogues, it is important to understand how changes to the inhibitor scaffold affect both K_i and k_{inact} as these values can often be predictive of inhibitor selectivity. Ideally, one would want an inhibitor that displays a low K_i (strong binding through protein-ligand recognition) while also having a sufficiently fast k_{inact} value, with the value of k_{inact}/K_i often being employed to prioritize covalent inhibitors in drug development campaigns.[68, 69]

Several methods exist to determine the individual values of k_{inact} and K_i for covalent inhibitors, with one of the most commonly employed methods involving the analysis of kinetic progress curves.[70] As originally proposed by Kitz & Wilson, the progress curves of covalent inhibitors will display non-linear product formation over time as a result of covalent bond formation between the inhibitor and the enzyme.[71] Two-step inhibitors will display a decrease in initial enzyme velocity v_i relative to the uninhibited control (v_i^{ctrl}) before reaching a steady-state velocity v_s (**Figure 4.8a**). In the case of irreversible inhibitors, the steady-state velocity v_s is constrained to full inhibition ($v_s = 0$) as they are expected to reach complete covalent modification of the enzyme under ideal assay conditions (**Figure 4.8b**) and with enough time. The progress curves can be described by Eq. 4.1 in order to determine the apparent first-order rate constant k_{obs} at varying inhibitor concentrations, and a secondary plot of inhibitor concentration ([I]) vs k_{obs} can then be described by Eq. 4.2 to obtain individual values of k_{inact} and K_I^{app}. It is important to note that this analysis provides a means of determining the equilibrium constant K_I and not the non-covalent dissociation constant K_i directly (Eq. 4.3). However, if the inhibitor is under rapid equilibrium between the unbound and non-covalently bound state, covalent bond formation will be rate-limiting and therefore K_I will approximate K_i.

Figure 4.8. Sample progress curves for (a) two-step reversible inhibitors and (b) two-step irreversible inhibitors. (c) Secondary plot of [I] vs k_{obs} to obtain individual values of k_{inact} and K_I.[70]

$$F = F_o + v_s t + \frac{(v_i - v_s)(1 - e^{-k_{obs}t})}{k_{obs}} \tag{4.1}$$
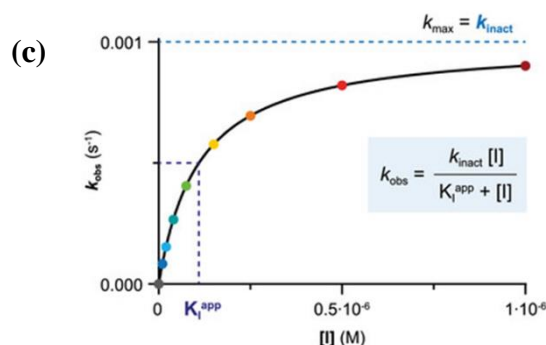
$$k_{obs} = \frac{k_{inact}[I]}{K_I{}^{app} + [I]} \tag{4.2}$$

$$K_I = \frac{(k_{off} + k_{inact})}{k_{on}} \tag{4.3}$$

The observed decrease in v_i of two-step covalent inhibitors is due to the contribution of the initial non-covalent interaction towards enzyme inhibition. In addition to the plots of k_{obs} shown in **Figure 4.8**, the non-covalent binding affinity K_i can also be determined by secondary plots of inhibitor concentration ([I]) vs v_i and fitting to Eq. 4.4, as shown in **Figure 4.9**.

Figure 4.9. Secondary plot of [I] vs v_i to determine $K_i$app.[70]
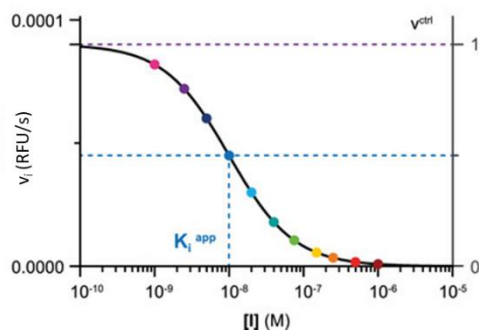
$$v_i = v_i^{min} + \frac{v^{ctrl} - v_i^{min}}{1 + \left(\frac{[I]}{K_i^{app}}\right)^h} \qquad (4.4)$$

In both cases, K_i^{app} can be converted to K_i by Eq. 4.5, correcting for substrate concentration relative to the K_M of the enzyme under the experimental conditions.

$$K_i = \frac{K_i^{app}}{1 + \frac{[S]}{K_m}} \qquad (4.5)$$

Using the method described above, reaction progress curves for inhibitors **4.10**, **4.11** and **4.19** were fit to Eq. 4.1, and secondary plots of [I] vs k_{obs} were fit to Eq. 4.2 to determine individual values of k_{inact} and K_I^{app} (**Figure 4.10a-c**). For reversible inhibitor **4.15**, the initial velocities were obtained by linear regression of progress curves, and a secondary plot of [I] vs v_i was fit to Eq. 4.4 to obtain a value of K_i^{app} (**Figure 4.10d**). In both methods, the apparent binding affinities were converted to absolute values using Eq. 4.5, using a K_m value determined experimentally (see Supporting Information – Figure S4.1).
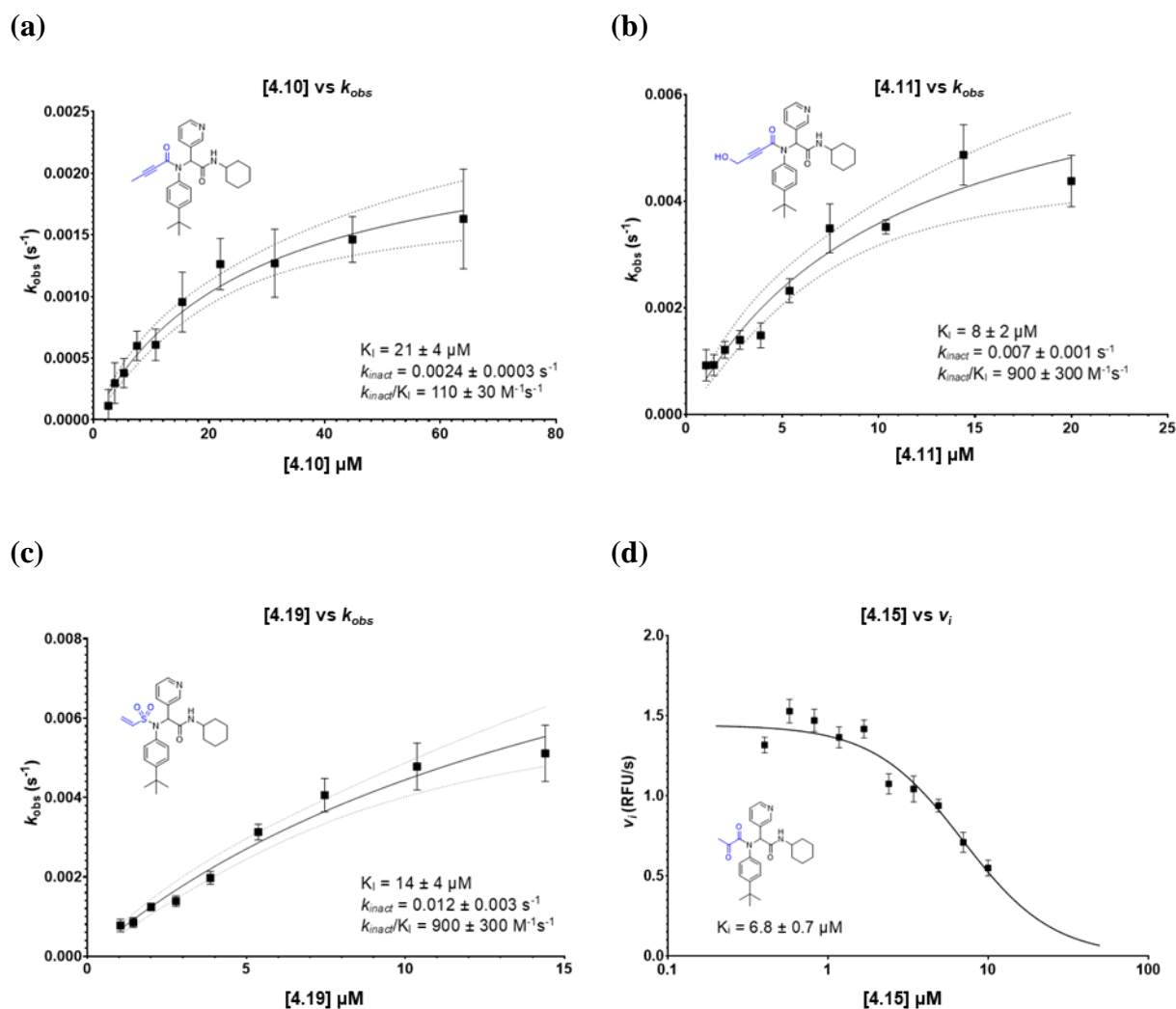
Figure 4.10. Secondary plots of [I] vs k_{obs} fit to Eq. (2) for covalent inhibitors (a) **4.10**, (b) **4.11**, and (c) **4.19**. (d) Secondary plot of [I] vs v_i fit to Eq. 4.4 for covalent inhibitor **4.15**.

The results in **Figure 4.10** suggest that inhibitors **4.11** and **4.15** have the most potent non-covalent binding affinity, while inhibitor **4.19** contains the most reactive warhead. Inhibitor **4.10** has both the least potent non-covalent binding affinity and the slowest k_{inact}, resulting in a k_{inact}/K_I value nearly an order of magnitude weaker than both **4.11** and **4.19**.

Although no k_{inact} was determined for reversible inhibitor **4.15**, the K_i value determined by fitting the initial velocity to Eq. 4.4 is very similar to the experimentally determined IC$_{50}$ at 30 min. incubation time (6.8 vs 5.2 μM). This suggests that covalent binding for this warhead does not

have a significant contribution to its potency, in line with a very low value of k_{inact} or k_{inact}/k_{rev}. The values of k_{inact} follow the expected trend based on electrophile reactivity in glutathione binding assays.[72, 73]
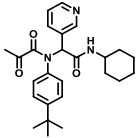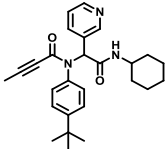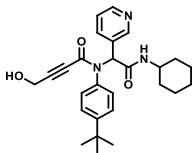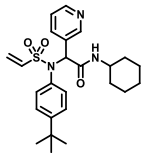
The progress curve analysis also suggests that the improvement in potency from alkynyl amide **4.10** to alkynyl amide **4.11** is a result of both an improvement in non-covalent binding affinity and covalent warhead reactivity. The hydroxyl group of **4.11** may facilitate additional hydrogen-bonding interactions within the active site (e.g., to the backbone of Thr^{26} or the sidechain of His^{41}) while also serving as an electron-withdrawing group to help activate the alkyne towards nucleophilic attack. A similar effect was observed in covalent $3CL^{pro}$ inhibitors recently reported by Wang *et al* using a similar scaffold to the initial non-covalent inhibitor **X77**.[74] They observed a significant improvement in IC_{50} when substituting their alkynylamide warhead with a chlorine, suggesting the importance of electronic effects on inhibitor potency for this warhead.

The experimentally determined values of k_{inact} and K_I were fitted to Eqs. 4.6 and 4.7 (as derived by Krippendorf *et al.*[80]) to calculate the expected IC_{50} under the previously employed experimental conditions (incubation time (t) = 30 minutes, [S] = 11.76 µM, K_M = 95 µM), and the calculated IC_{50} values were all in good agreement with the experimental values (**Table 4.2**). While the calculated IC_{50} value for **4.11** is lower than the experimentally determined value, it still falls within the margin of error for the experimental value. The calculated IC_{50} value for **4.10** is also slightly below the experimentally determined value and may represent a slight overestimation of k_{inact} or an underestimation of K_I.

$$n_{IC50} = \frac{IC_{50}(t)}{K_i * \left(1 + \frac{S}{K_M}\right) + IC_{50}(t)} \tag{4.6}$$

$$IC_{50}(t)_{calc} = K_i * \left(\frac{2 * (K_M + S) * e^{-n_{IC50}*k_{inact}*t} - S}{K_M} - 1\right) \tag{4.7}$$

Table 4.2. Experimentally determined inhibitor binding affinities and kinetics.

Structure	k_{inact} (s⁻¹)	K_I (µM)	k_{inact}/K_I (M⁻¹s⁻¹)	IC₅₀ calc (µM)	IC₅₀ exp (µM)
	n.d.	6.8 ± 0.7	n.d.	n.d.	5.2 ± 1.2
	0.0024 ± 0.0002	21 ± 4	110 ± 30	3.5	5.3 ± 0.8
	0.007 ± 0.001	8 ± 2	900 ± 300	0.48	0.8 ± 0.4
	0.013 ± 0.003	14 ± 4	900 ± 300	0.49	0.4 ± 0.1

4.5 Conclusions and Future Work

Covalent inhibition of SARS-CoV-2 3CL^pro is a promising strategy for the treatment of COVID-19. Our strategy relied on a previously reported imidazole-containing inhibitor of the similar coronavirus SARS-CoV responsible for the epidemic of SARS in the early 2000s. We first used our docking program FITTED (specifically modified to accommodate covalent inhibitors[75]) and screened a set of covalent warheads. The docked poses confirmed that replacing the imidazole ring by a reactive group should lead to potent covalent inhibition. Gratifyingly, while the imidazole of **X77** was known to be essential for the inhibitory potency, replacing it with many warheads maintained and even improved the potency, with our lead compounds **4.13** and **4.19** being an order of magnitude more potent. The covalent binding mode of **4.13** and **4.19** suggested by FITTED was

confirmed by crystallography, and covalent binding kinetics of **4.10**, **4.11**, **4.19** and **4.15** were further investigated to determine individual values of k_{inact} and K_I.

Our work, initially reported in ChemRXiv in 2020, showed the potential of covalent warhead incorporation into the scaffold of non-covalent SARS-CoV inhibitor **X77**.[76] Since our initial report, several other groups have followed the same strategy.[74, 77-79] This approach has notably led to the development of covalent inhibitor **Y180** that has shown promising *in vivo* activity against SARS-CoV-2 and features a similar core scaffold and the same ketoamide warhead of **4.15.**

Additional work for this project has focused on the optimization of our most promising covalent scaffolds by improving their non-covalent interactions within the binding site of 3CL[pro.] This work is further elaborated in Chapter 5.

4.6   Experimental

4.6.1   General

Unless otherwise specified, all solvents and reagents were purchased from commercial suppliers and used without further purification. All ^{1}H, ^{13}C and ^{19}F NMR spectra were acquired on a Bruker 400 or 500 MHz spectrometer, or a Varian 500 MHz spectrometer. Chemical shifts are reported in ppm using the residual of deuterated solvents as an internal standard. Chromatography was performed on silica gel 60 (230-40 mesh) or using the Biotage One Isolera with ZIP cartridges. High resolution mass spectrometry was performed by ESI on a Bruker Maxis Impact API QqTOF mass spectrometer at McGill University. Reversed-phase HPLC (water and MeCN or MeOH gradient) was used to verify the purity of compounds on an Agilent 1100 series instrument equipped with VWD-detector, C18 reverse column (Agilent, Zorbax Eclipse XDBC18 150 mm 4.6 mm, 5 μm), and UV detection at 254 nm. Measured purities for all tested compounds are listed in Table S3 in the supporting information.

4.6.2   Synthesis and Characterization Data

General Procedures A, B and C for 4-Component Ugi Reaction. In a 6-dram vial equipped with a stir bar, aldehyde (1.0 mmol, 1.0 eq.), aniline (1.0 mmol, 1.0 eq.) and carboxylic acid (1.0 mmol, 1.0 eq.) were combined in MeOH (4 mL). The obtained reaction mixture was stirred for 30 min.

at room temperature. Afterwards cyclohexyl isocyanide (0.9 mmol, 0.9 eq.) was added to the reaction mixture and the walls of the vial were washed with 1 mL of MeOH. The reaction mixture was continued to stir at room temperature overnight. The crude reaction mixture was evaporated in vacuo. Purification procedure A) The crude product was triturated with hexanes (5 mL) and filtered. The obtained product was further washed with hexanes (3 x 3 mL). Purification procedure B) The crude product recrystallized from $CHCl_3$/hexanes mixture, filtered and the obtained product was further washed with hexanes (3 x 3 mL). Purification procedure C) The crude product was redissolved in DCM. The obtained crude solution was deposited on silica. It was then purified using flash column chromatography using DCM/MeOH (gradient $0 \rightarrow 5\%$) as eluent.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-1H-imidazole-5-carboxamide (X77). Compound was made and purified using general procedure B, white solid 30% yield, 125 mg. ^{1}H NMR (500 MHz, MeOD) δ 8.37 (s, 1H), 8.33 (dd, J = 4.9, 1.5 Hz, 1H), 7.66–7.57 (m, 2H), 7.31 (d, J = 7.8 Hz, 2H), 7.22 (dd, J = 7.9, 4.9 Hz, 1H), 6.27 (s, 1H), 5.46 (s, 1H), 3.71 (td, J = 10.5, 9.3, 3.9 Hz, 1H), 1.93 (d, J = 12.3 Hz, 1H), 1.80–1.72 (m, 2H), 1.65 (ddt, J = 30.9, 12.9, 3.8 Hz, 2H), 1.27 (s, 12H). ^{13}C NMR (126 MHz, MeOD) δ 169.00, 152.54, 150.62, 148.24, 138.83, 136.36, 131.58, 131.07, 125.72, 123.31, 62.74, 34.16, 32.16, 32.12, 30.25, 25.22, 24.72, 24.64. HRMS (ESI/Q-TOF) m/z: $[M + Na]^+$ calculated for $C_{27}H_{33}N_5NaO_2$ 482.2526; found 482.2535.

2-(N-(4-(*tert*-butyl)phenyl)acetamido)-N-cyclohexyl-2-(pyridin-3-yl)acetamide (4.4). Compound was made and purified using general procedure A, white solid 76% yield, 280 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.46–8.43 (m, 1H), 8.42 (s, 1H), 7.41 (d, J = 8.1 Hz, 1H), 7.23 (d, J = 8.5 Hz, 2H), 7.03 (dd, J = 8.0, 4.8 Hz, 1H), 6.93 (s, 1H), 6.03 (s, 2H), 3.94–3.65 (m, 1H), 1.98 (d, J = 16.8 Hz, 1H), 1.89–1.81 (m, 4H), 1.75–1.63 (m, 2H), 1.59 (dt, J = 13.0, 4.3 Hz, 1H), 1.43–1.28 (m, 2H), 1.25 (s, 9H), 1.23–1.04 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 171.83, 168.21, 151.89, 151.44, 149.66, 138.09, 137.42, 130.87, 129.60, 126.28, 122.92, 62.42, 48.88, 34.73, 32.95 (d, J = 10.4 Hz), 31.37, 25.61, 24.87 (d, J = 6.9 Hz), 23.34. HRMS (ESI/Q-TOF) m/z: $[M + Na]^+$ calculated for $C_{25}H_{33}N_3NaO_2$ 430.2465; found 430.2464.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl) ethyl)acrylamide (4.5). Compound was made and purified using general procedure C, white solid

32% yield, 120 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.41–8.39 (m, 1H), 8.38–8.37 (m, 1H), 7.42–7.35 (m, 1H), 7.19 (d, J = 8.1 Hz, 2H), 7.01 (dd, J = 8.0, 4.8 Hz, 1H), 6.91 (s, 1H), 6.49 (d, J = 8.0 Hz, 1H), 6.33 (dd, J = 16.8, 2.0 Hz, 1H), 6.11 (s, 1H), 5.93 (dd, J = 16.8, 10.3 Hz, 1H), 5.49 (dd, J = 10.4, 2.0 Hz, 1H), 3.84–3.73 (m, 1H), 1.93 (s, 1H), 1.81 (dd, J = 13.1, 4.1 Hz, 1H), 1.67–1.53 (m, 3H), 1.37–1.26 (m, 2H), 1.22 (s, 9H), 1.17–1.06 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.04, 166.50, 151.79, 151.28, 149.45, 137.93, 136.14, 130.83, 129.93, 128.64, 128.52, 126.08, 122.85, 62.69, 48.76, 34.64, 32.86, 32.80, 31.27, 25.52, 24.83, 24.77. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{26}$H$_{33}$N$_3$NaO$_2$ 442.2465; found 442.2456.

(*E*)-*N*-(4-(*tert*-butyl)phenyl)-*N*-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-enamide (4.6). Compound was made and purified using general procedure B, pale white solid 32% yield, 126 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.46–8.41 (m, 2H), 7.44 (dt, J = 7.9, 2.0 Hz, 1H), 7.24 (d, J = 8.2 Hz, 2H), 7.04 (ddd, J = 8.0, 4.8, 0.8 Hz, 1H), 6.99–6.91 (m, 3H), 6.36 (d, J = 8.1 Hz, 1H), 6.09 (s, 1H), 5.65 (dd, J = 15.0, 1.7 Hz, 1H), 3.87–3.76 (m, 1H), 1.97 (dd, J = 11.7, 4.6 Hz, 1H), 1.92–1.82 (m, 1H), 1.72 (dd, J = 7.0, 1.7 Hz, 3H), 1.69–1.63 (m, 1H), 1.60–1.56 (m, 1H), 1.42–1.30 (m, 2H), 1.27 (s, 9H), 1.22–1.09 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.25, 166.99, 151.78, 151.24, 149.41, 143.08, 138.09, 136.62, 131.00, 129.80, 126.22, 122.88, 122.71, 62.96, 48.78, 34.75, 32.97, 32.91, 31.37, 25.62, 24.88, 24.84, 18.21. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{27}$H$_{35}$N$_3$NaO$_2$ 456.2621; found 456.2630.

N-(4-(*tert*-butyl)phenyl)-*N*-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl) methacrylamide (4.7). Compound was made and purified using general procedure B, pale white solid 68% yield, 265 mg. ^{1}H NMR (500 MHz, MeOD) δ 8.33 (d, J = 2.3 Hz, 1H), 8.31 (dd, J = 4.9, 1.6 Hz, 1H), 7.56 (dt, J = 7.9, 2.0 Hz, 1H), 7.23–7.15 (m, 3H), 7.05 (s, 2H), 6.10 (s, 1H), 5.01 (dt, J = 6.9, 1.3 Hz, 2H), 3.70 (tt, J = 10.9, 3.9 Hz, 1H), 1.90 (dd, J = 10.7, 3.8 Hz, 1H), 1.73 (s, 5H), 1.70–1.59 (m, 2H), 1.41–1.27 (m, 3H), 1.21 (s, 9H), 1.19–1.06 (m, 2H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 172.82, 167.95, 151.29, 151.19, 149.55, 140.42, 137.86, 137.64, 130.88, 129.32, 125.77, 122.97, 119.77, 63.47, 48.78, 34.63, 32.90 (d, J = 10.4 Hz), 31.32, 25.57, 24.83, 24.77, 20.42. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{27}$H$_{35}$N$_3$NaO$_2$ 456.2621; found 456.2620.

N-(4-(*tert*-butyl)phenyl)-*N*-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)cyclopent-1-

ene-1-carboxamide (4.8). Compound was made and purified using general procedure A, pale yellow solid 80% yield, 333 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.48 (d, J = 2.3 Hz, 1H), 8.46 (dd, J = 4.8, 1.7 Hz, 1H), 7.51 (dt, J = 8.0, 2.0 Hz, 1H), 7.21–7.16 (m, 2H), 7.08 (ddd, J = 7.9, 4.8, 0.8 Hz, 1H), 6.90 (d, J = 8.0 Hz, 2H), 6.28 (d, J = 8.0 Hz, 1H), 6.04 (s, 1H), 5.82 (d, J = 2.3 Hz, 1H), 3.89–3.78 (m, 1H), 2.19 (ddt, J = 7.7, 5.1, 2.5 Hz, 2H), 2.12 (tt, J = 6.7, 2.8 Hz, 2H), 2.02–1.93 (m, 1H), 1.92–1.84 (m, 1H), 1.73–1.54 (m, 4H), 1.44–1.29 (m, 2H), 1.25 (s, 9H), 1.23–1.07 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.92, 168.16, 151.57, 151.22, 149.46, 140.13, 139.09, 137.96, 137.69, 130.98, 129.50, 125.81, 122.94, 63.92, 48.74, 34.69, 33.80, 33.22, 32.95, 32.92, 31.35, 25.60, 24.84, 24.79, 23.29. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{29}$H$_{37}$N$_3$NaO$_2$ 482.2778; found 482.2781.

Ethyl (*E*)-4-((4-(*tert*-butyl)phenyl)(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)amino)-4-oxobut-2-enoate (4.9). Compound was made and purified using general procedure A, white powder 52% yield, 170 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.51–8.38 (m, 2H), 7.49 (dt, J = 8.1, 2.0 Hz, 1H), 7.23 (d, J = 8.1 Hz, 2H), 7.09 (dd, J = 8.0, 4.8 Hz, 1H), 6.85 (d, J = 15.3 Hz, 1H), 6.72 (d, J = 15.3 Hz, 1H), 6.24 (d, J = 8.1 Hz, 1H), 6.10 (s, 1H), 4.12 (q, J = 7.1 Hz, 2H), 3.80 (dtd, J = 10.8, 7.2, 4.0 Hz, 1H), 1.98–1.90 (m, 1H), 1.89–1.79 (m, 1H), 1.72–1.53 (m, 3H), 1.33 (ddd, J = 13.0, 10.0, 3.3 Hz, 1H), 1.25 (s, 9H), 1.20 (t, J = 7.1 Hz, 3H), 1.10 (ddt, J = 23.0, 15.4, 10.8 Hz, 1H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.37, 165.46, 165.04, 152.42, 150.54, 148.89, 138.83, 135.41, 133.89, 132.15, 130.95, 129.77, 126.48, 123.26, 63.08, 61.13, 49.00, 34.77, 32.90, 32.84, 31.28, 25.53, 24.86, 24.80, 14.11. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{29}$H$_{37}$N$_3$NaO$_4$ 514.2676; found 514.2691.

_N_-(4-(*tert*-butyl)phenyl)-_N_-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamide (4.10). Compound was made and purified using general procedure A, white solid 92% yield, 357 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.44 (d, J = 4.9 Hz, 1H), 8.41 (d, J = 2.3 Hz, 1H), 7.44 (d, J = 8.1 Hz, 1H), 7.21 (d, J = 8.4 Hz, 2H), 7.05 (dd, J = 8.0, 4.8 Hz, 1H), 6.97 (d, J = 8.0 Hz, 2H), 6.20 (s, 1H), 6.03 (s, 1H), 3.80 (dtd, J = 10.8, 7.2, 4.0 Hz, 1H), 1.96 (dq, J = 13.2, 4.8 Hz, 1H), 1.84 (d, J = 16.6 Hz, 1H), 1.75–1.62 (m, 5H), 1.58 (dd, J = 13.1, 4.1 Hz, 1H), 1.42–1.26 (m, 2H), 1.25 (s, 9H), 1.25–1.05 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.43, 155.41, 151.87, 151.20, 149.62, 138.15, 136.38, 130.34, 129.90, 125.70, 122.98, 92.12, 73.86, 62.23, 34.71, 32.87, 32.80, 31.31, 25.56, 24.85, 24.80. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for

$C_{27}H_{33}N_3NaO_2$ 454.2465; found 454.2458.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-4-hydroxybut-2-ynamide (4.11). To a solution of 4-*tert*-butylaniline (0.31 mmol, 0.05 mL) in MeOH was added 3-Py-carboxaldehyde (0.31 mmol, 0.03 mL) and stirred for 30 min. 4-hydroxy-butynoic acid (0.31 mmol, 31 mg) and cyclohexyl isocyanide (0.31 mmol, 0.04 mL) were added and the solution stirred at room temperature overnight. The solvent was evaporated under a stream of air, and the crude reaction mixture was suspended in a small amount of EtOAc. Hexanes was added and the precipitate was collected by filtration and further rinsed with hexanes. The precipitate was dried over vacuum to afford the desired product (60 mg, 43% yield) as a white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.49 (d, *J* = 2.3 Hz, 1H), 8.44 (dd, *J* = 4.8, 1.7 Hz, 1H), 7.44 (dt, *J* = 8.0, 2.0 Hz, 1H), 7.22 (d, *J* = 8.8 Hz, 2H), 7.06 (dd, *J* = 8.0, 4.8 Hz, 1H), 7.03 (d, *J* = 7.9 Hz, 2H), 6.17 (d, *J* = 8.0 Hz, 1H), 6.07 (s, 1H), 4.05 (s, 2H), 3.84–3.73 (m, 1H), 2.45 (s, 1H), 1.96 (dd, *J* = 12.5, 4.1 Hz, 1H), 1.84 (dd, *J* = 12.1, 4.2 Hz, 1H), 1.74–1.53 (m, 2H), 1.43–1.27 (m, 3H), 1.25 (s, 9H), 1.23–1.00 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.25, 154.75, 152.38, 151.32, 149.75, 138.19, 135.94, 130.27, 130.25, 125.80, 123.11, 92.47, 78.94, 62.16, 50.59, 49.09, 34.78, 32.88, 32.83, 31.36, 25.57, 24.92, 24.85. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for $C_{27}H_{33}N_3O_3$ 448.2595; found 448.2592.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-2-fluoroacetamide (4.12). Compound was made and purified using general procedure A, pale yellow solid 70% yield, 270 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.47 (dd, *J* = 4.9, 1.7 Hz, 1H), 8.43 (d, *J* = 2.3 Hz, 1H), 7.43 (dt, *J* = 7.9, 2.0 Hz, 1H), 7.24 (s, 2H), 7.06 (ddd, *J* = 8.0, 4.8, 0.8 Hz, 1H), 6.04 (s, 1H), 5.87 (d, *J* = 7.8 Hz, 1H), 4.64 (d, *J* = 3.1 Hz, 1H), 4.55 (d, *J* = 3.3 Hz, 1H), 3.86–3.75 (m, 1H), 2.02–1.95 (m, 1H), 1.89–1.81 (m, 1H), 1.74–1.55 (m, 3H), 1.44–1.28 (m, 2H), 1.25 (s, 9H), 1.22–1.02 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.05, 167.89, 167.48, 153.00, 151.45, 149.99, 138.13, 133.76, 129.97, 129.82, 126.64, 123.12, 78.70 (d, *J* = 178.1 Hz), 62.38, 49.16, 34.84, 32.94 (d, *J* = 9.3 Hz), 31.31, 25.57, 24.90 (d, *J* = 6.9 Hz). HRMS (ESI/Q-TOF) *m/z*: [M + Na]$^+$ calculated for $C_{25}H_{32}FN_3NaO_2$ 448.2371; found 448.2366.

N-(4-(*tert*-butyl)phenyl)-2-chloro-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)acetamide (4.13). Compound was made and purified using general procedure A, yellow solid 93%

yield, 370 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.47 (dd, J = 4.8, 1.6 Hz, 1H), 8.43 (d, J = 2.3 Hz, 1H), 7.44 (dt, J = 7.9, 2.0 Hz, 1H), 7.26 (s, 3H), 7.07 (ddd, J = 7.8, 4.9, 0.8 Hz, 1H), 5.99 (s, 1H), 5.88 (s, 1H), 3.85 (s, 2H), 3.84–3.78 (m, 1H), 1.98 (dd, J = 12.6, 4.2 Hz, 1H), 1.85 (dd, J = 12.6, 4.2 Hz, 1H), 1.74–1.63 (m, 2H), 1.59 (dt, J = 12.8, 3.8 Hz, 1H), 1.43–1.28 (m, 2H), 1.26 (s, 9H), 1.22–1.03 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.43, 167.42, 152.83, 151.27, 149.75, 138.27, 135.35, 130.28, 129.73, 126.61, 123.16, 63.16, 49.11, 42.58, 34.84, 32.97, 32.90, 31.33, 25.57, 24.91, 24.85. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{25}$H$_{32}$ClN$_3$NaO$_2$ 464.2075; found 464.2087.

N-(4-(*tert*-butyl)phenyl)-2-cyano-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)acetamide (4.14). Compound was made and purified using general procedure B, white solid 84% yield, 328 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.52–8.45 (m, 1H), 8.41 (s, 1H), 7.44–7.38 (m, 2H), 7.10–7.04 (m, 2H), 6.47 (s, 1H), 5.98 (d, J = 2.1 Hz, 1H), 5.68 (s, 1H), 3.86–3.75 (m, 1H), 3.27–3.20 (m, 2H), 1.97 (d, J = 12.6 Hz, 1H), 1.84 (d, J = 13.0 Hz, 2H), 1.68–1.56 (m, 2H), 1.40–1.28 (m, 2H), 1.25 (s, 9H), 1.23–0.99 (m, 4H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 166.87, 162.92, 152.91, 151.14, 149.87, 137.79, 135.02, 129.75, 129.52, 123.02, 113.69, 62.87, 49.04, 34.61, 32.70, 32.64, 31.04, 26.19, 25.28, 24.67, 24.60.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-2-oxopropanamide (4.15). To a solution of 4-*tert*-butylaniline (0.10 mL, 0.67 mmol, 1.0 eq.) in MeOH (2 mL) was added 3-pyridine carboxaldehyde (0.06 mL, 0.67 mmol, 1.0 eq.) and the solution stirred at room temperature for 30 min. The solution was cooled to 0 °C, and pyruvic acid (0.06 mL, 0.80 mmol, 1.2 eq.) and cyclohexyl isocyanide (0.10 mL, 0.80 mmol, 1.2 eq.) were added in quick succession. The solution was slowly warmed to room temperature and stirred overnight. The crude reaction mixture was evaporated in vacuo and purified by column chromatography (1:1 Hex:EtOAc) to afford the product (105 mg, 36%) as a white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.69 (s, 1H), 8.56 (s, 1H), 7.84 (d, J = 8.0 Hz, 1H), 7.35 (d, J = 7.7 Hz, 1H), 7.23 (d, J = 8.8 Hz, 2H), 7.03 (d, J = 8.1 Hz, 2H), 6.49 (d, J = 8.0 Hz, 1H), 6.15 (s, 1H), 3.80 (tdd, J = 10.7, 6.7, 4.0 Hz, 1H), 2.19 (s, 3H), 1.97–1.79 (m, 2H), 1.69 (ddt, J = 17.1, 13.1, 4.0 Hz, 2H), 1.59 (dt, J = 12.8, 3.9 Hz, 1H), 1.41–1.25 (m, 2H), 1.24 (s, 9H), 1.16 (dtd, J = 16.2, 13.6, 12.7, 9.9 Hz, 2H).^{13}C NMR (126 MHz, CDCl$_3$) δ 197.49, 168.14, 166.68, 152.69, 150.19, 148.63, 139.25, 134.15, 130.64, 129.82, 126.29, 123.61, 62.36, 49.22, 34.77, 32.85, 32.81, 31.26, 27.84,

25.51, 24.87, 24.81. HRMS (ESI/Q-TOF) m/z: $[M + Na]^+$ calculated for $C_{26}H_{33}N_3NaO_3$ 458.2412; found 458.2421.

N-(4-($tert$-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-3,3,3-trifluoro-2-oxopropanamide (4.16). To a solution of 4-$tert$-butylaniline (0.62 mmol, 0.1 mL) in MeOH was added 3-pyridine carboxaldehyde (0.62 mmol, 0.06 mL) and stirred for 30 min. The solution was cooled to 0 °C and trifluoropyruvic acid (0.62 mmol, 100 mg) and cyclohexyl isocyanide (0.62 mmol, 0.08 mL) were added. The solution was slowly warmed to room temp. and stirred overnight. The solvent was evaporated under a stream of air, and the crude reaction mixture was suspended in a small amount of EtOAc. Hexanes was added and the precipitate was collected by filtration and rinsed with hexanes and acetone. The precipitate was dried over vacuum to afford the desired product (87 mg, 29% yield) as a white powder. ^1H NMR (500 MHz, CDCl$_3$) δ 8.62 (s, 1H), 8.55–8.51 (m, 1H), 7.60 (d, $J = 7.9$ Hz, 1H), 7.23 (d, $J = 8.6$ Hz, 3H), 7.03 (s, 2H), 6.10 (s, 2H), 3.82 (tdt, $J = 10.9, 7.8, 3.9$ Hz, 1H), 1.90 (dd, $J = 49.5, 12.9$ Hz, 2H), 1.75–1.54 (m, 4H), 1.43–1.27 (m, 1H), 1.23 (s, 9H), 1.21–1.02 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 165.76, 162.63, 153.82, 150.12, 148.74, 139.63, 132.24, 130.28, 130.01, 126.64, 123.76, 62.36, 49.38, 34.89, 32.81, 32.79, 31.33, 31.21, 31.17, 25.49, 24.88, 24.83. ^{19}F NMR (471 MHz, CDCl$_3$) δ −74.82. HRMS (ESI/Q-TOF) m/z: $[M + Na]^+$ calculated for $C_{26}H_{30}F_3N_3NaO_3$ 512.2131; found 512.2114.

2-((cyanomethyl)amino)-N-cyclohexyl-2-(pyridin-3-yl)acetamide (4.17). In a 6-dram vial equipped with a stir bar 3-pyridinecarboxaldehyde (107 mg, 1.0 mmol, 1.0 eq.) and β-alanine (89 mg, 1.0 mmol, 1.0 eq.) were mixed together in MeOH (4 mL). The obtained solution was stirred for 30 min at room temperature. Cyclohexyl isocyanide (109 mg, 1.0 mmol, 1.0 eq.) was added to the reaction mixture and the walls of the vial were washed with 1 mL of MeOH. The obtained reaction mixture was stirred at room temperature overnight. The crude reaction mixture was evaporated in vacuo and redissolved in DCM. The obtained crude solution was deposited on silica. It was then purified using flash column chromatography using DCM/MeOH (gradient 0 → 5%) as eluent. The product was obtained as colorless oil, 156 mg 54%. ^1H NMR (500 MHz, CDCl$_3$) δ 8.65 (d, $J = 2.4$ Hz, 1H), 8.62 (dd, $J = 4.8, 1.7$ Hz, 1H), 7.75 (dt, $J = 7.9, 2.0$ Hz, 1H), 7.34 (dd, $J = 7.9, 4.8$ Hz, 1H), 6.02 (d, $J = 8.4$ Hz, 1H), 4.40 (s, 1H), 3.82–3.71 (m, 1H), 3.64 (d, $J = 17.5$ Hz, 1H), 3.45 (d, $J = 17.3$ Hz, 1H), 2.53 (s, 1H), 1.85 (ddd, $J = 17.0, 12.4, 4.4$ Hz, 2H),

1.64 (dtd, J = 25.9, 9.0, 4.7 Hz, 4H), 1.40–1.28 (m, 2H), 1.20–1.02 (m, 2H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.45, 150.57, 149.43, 135.66, 133.25, 124.24, 116.87, 63.55, 48.67, 35.50, 33.06, 33.02, 25.49, 24.83. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{15}$H$_{20}$N$_4$NaO 295.1529; found 295.1523.

N-cyclohexyl-2-(pyridin-3-yl)-2-(*p*-tolylamino)acetamide (4.18). To a solution of 4-*tert*-butylaniline (1.01 mmol, 0.16 mL) and 3-pyridinecarboxaldehyde (1.01 mmol, 0.09 mL) in MeOH (5 mL) was added cyclohexyl isocyanide (1.01 mmol, 0.12 mL) and phosphoric acid (0.2 mmol, 0.01 mL, 85%), and the solution stirred at room temperature overnight. The solvent was evaporated under a stream of air, and the crude reaction mixture was suspended in a small amount of EtOAc. Hexanes was added and the precipitate was collected by filtration and rinsed with hexanes and acetone. The precipitate was dried over vacuum to afford the desired product (353 mg, 96% yield) as a white powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.69 (dd, J = 2.4, 0.9 Hz, 1H), 8.46 (dd, J = 4.8, 1.7 Hz, 1H), 8.19 (d, J = 7.9 Hz, 1H), 7.85 (dt, J = 7.9, 2.0 Hz, 1H), 7.35 (ddd, J = 7.8, 4.8, 0.9 Hz, 1H), 7.10–7.04 (m, 2H), 6.61–6.55 (m, 2H), 6.04 (d, J = 8.1 Hz, 1H), 5.02 (d, J = 8.0 Hz, 1H), 3.56–3.47 (m, 1H), 1.74 (dd, J = 10.5, 4.8 Hz, 1H), 1.70–1.63 (m, 1H), 1.62–1.48 (m, 3H), 1.31–1.19 (m, 2H), 1.18 (s, 10H), 1.17–1.03 (m, 1H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 169.29, 148.65, 148.56, 144.47, 138.99, 135.44, 134.52, 125.39, 123.45, 112.89, 58.39, 47.58, 33.46, 32.24, 32.06, 31.38, 25.11, 24.37, 24.26. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{23}$H$_{31}$N$_3$NaO 388.2359; found 388.2352.

2-(N-(4-(*tert*-butyl)phenyl)vinylsulfonamido)-N-cyclohexyl-2-(pyridin-3-yl)acetamide (4.19). To a solution of **4.18** (0.41 mmol, 150 mg) in DCM (4 mL) was added pyridine (0.56 mmol, 0.04 mL) and the solution cooled to 0 °C. 2-chloroethanesulfonyl chloride (0.49 mmol, 0.05 mL) was added dropwise and stirred for 1 h, after which pyridine (0.56 mmol, 0.04 mL) was added, and the solution was warmed to room temp. and stirred overnight. The reaction was monitored by TLC (1:1 DCM:EtOAc). The reaction was quenched with water and extracted with DCM (x2). The combined organic layers were washed with sat. NH$_4$Cl, sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, and concentrated in vacuo. The crude product was further purified by column chromatography (0–4% (MeOH + 1%NH$_4$OH)/DCM) to afford the pure product (66 mg, 35% yield) as a white powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.35 (dd, J = 4.9, 1.6 Hz, 1H), 8.31 (s, 0H), 8.10 (d, J = 7.6 Hz, 1H), 7.36 (dt, J = 8.0, 2.0 Hz, 1H),

7.17–7.09 (m, 4H), 6.98 (dd, J = 16.5, 9.9 Hz, 1H), 6.06 (d, J = 9.9 Hz, 1H), 5.98 (d, J = 16.5 Hz, 1H), 5.81 (s, 1H), 3.57 (tdt, J = 11.0, 7.6, 3.7 Hz, 1H), 1.79–1.72 (m, 1H), 1.67 (dt, J = 12.9, 4.0 Hz, 1H), 1.64–1.47 (m, 2H), 1.33–1.19 (m, 2H), 1.17 (s, 9H), 1.16–0.92 (m, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 167.43, 150.53, 150.51, 148.97, 136.88, 136.07, 133.56, 131.92, 131.00, 127.18, 124.92, 122.94, 63.05, 47.95, 34.19, 32.07, 31.88, 30.94, 25.09, 24.41, 24.31. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{25}$H$_{33}$N$_3$NaO$_3$S 478.2135; found 478.2127.

Ethyl-(4-(*tert*-butyl)phenyl)(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)carbamate (4.20). To a solution of **4.18** (0.41 mmol, 150 mg) in DCM (8 mL) was added pyridine (0.82 mmol, 0.06 mL) and the solution was cooled to 0 °C. Ethyl chloroformate (0.49 mmol, 0.05 mL) was added dropwise and stirred at 0 °C for 1 h, then room temp. overnight. The reaction was quenched with water and extracted twice with DCM. The combined organic layers were washed with sat. NH$_4$Cl, sat. NaHCO$_3$ and brine, then dried over Na$_2$SO$_4$ and concentrated in vacuo. The crude residue was further purified by column chromatography (0–80% EtOAc/Hex) to afford the desired product (136 mg, 76% yield) as a white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.55 (s, 1H), 8.51–8.46 (m, 1H), 7.66 (dt, J = 8.1, 1.9 Hz, 1H), 7.23 (td, J = 6.0, 5.5, 2.3 Hz, 3H), 6.97 (d, J = 8.5 Hz, 2H), 6.24 (d, J = 8.1 Hz, 1H), 5.72 (s, 1H), 4.15 (qd, J = 7.1, 2.7 Hz, 2H), 3.84 (dddd, J = 14.5, 10.5, 7.9, 3.9 Hz, 1H), 1.99–1.83 (m, 2H), 1.69 (tt, J = 12.5, 3.9 Hz, 2H), 1.60 (dt, J = 12.9, 3.9 Hz, 1H), 1.43–1.30 (m, 2H), 1.26 (s, 9H), 1.22–1.08 (m, 6H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.76, 156.29, 150.83, 148.98, 147.37, 139.54, 137.15, 132.52, 128.33, 126.00, 123.58, 64.95, 62.65, 48.91, 34.67, 32.97, 32.93, 31.38, 25.57, 24.84, 24.82, 14.64. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{26}$H$_{35}$N$_3$NaO$_3$ 460.2571; found 460.2579.

2-((4-(*tert*-butyl)phenyl)(cyanomethyl)amino)-N-cyclohexyl-2-(pyridin-3-yl)acetamide (4.21). To a solution of 2-((4-(*tert*-butyl)phenyl)amino)acetonitrile (0.27 mmol, 50 mg) and 3-pyridinecarboxaldehyde (0.27 mmol, 0.03 mL) in MeOH (3 mL) was added cyclohexyl isocyanide (0.27 mmol, 0.04 mL) and phosphoric acid (0.05 mmol, 0.004 mL, 85%), and the solution stirred at room temp. overnight. The solvent was evaporated under a stream of air, and the crude reaction mixture was suspended in a small amount of EtOAc. Hexanes was added and the precipitate was collected by filtration and rinsed with hexanes and acetone. The precipitate was dried over vacuum to afford the desired product (74 mg, 68% yield) as a white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 9.07 (s, 1H), 8.61 (s, 1H), 8.32 (d, J = 7.9 Hz, 1H), 7.61 (t, J = 6.6 Hz, 1H), 7.33 (d, J = 8.6 Hz,

2H), 7.08 (d, J = 8.7 Hz, 2H), 7.03 (d, J = 7.8 Hz, 1H), 5.57 (s, 1H), 4.16 (d, J = 17.9 Hz, 1H), 4.01 (d, J = 17.8 Hz, 1H), 1.98–1.67 (m, 2H), 1.68–1.49 (m, 4H), 1.45–1.28 (m, 1H), 1.27 (s, 9H), 1.24–0.64 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 166.94, 147.84, 143.94, 126.95, 125.47, 120.77, 115.57, 65.98, 48.79, 42.14, 34.47, 32.71, 32.32, 31.42, 31.21, 25.44, 24.68, 24.61 HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{26}$H$_{33}$N$_3$NaO$_3$ 458.2412; found 458.2421.

4.6.3 *In vitro assays*

Detection of inhibitors by fluorescence spectrophotometry. For all IC$_{50}$ experiments, reactions were performed in 50 μL assay with 11.76 μM fluorescence substrate DABCYL-KTSAVLQSGFRKME-EDANS from BPS Biosciences (San Diego, CA, USA), which has been previously used for assaying 3CL proteases.

38 μL of enzyme (3CL^{pro} with His-tag, diluted in a 3CL^{pro} Protease Assay Buffer from BPS Bioscience supplemented with 1 mM DTT) were incubated for 30 min at room temp. with 2 μL of 1.25 mM compounds (diluted in DMSO; Sigma). In both cases, screenings (duplicates) and IC$_{50}$ experiments (duplicates and triplicates), 10 μL of diluted substrate (58.8 μM) were added and reactions monitored by following the fluorescence as a function of time (excitation at 360 nm, emission at 460 nm) using a Synergy™ H4 Hybrid Multi-Mode Microplate Reader (Winooski, VT, USA). Controls were (i) no inhibition: 2 μL of DMSO, (ii) positive inhibition: 2 μL GC-376 (500 μM, BPS Bioscience) diluted in distilled water and (iii) blank: 38 μL of the 3CLpro Protease Assay Buffer with 1 mM DTT was added instead of the enzyme, together with 2 μL of DMSO. In all cases the compounds were added as DMSO solutions and their final concentration in the reactions was 50 μM. The reactions ran for at least 1 h, and the linear initial slopes of the progress curves were used to calculate the reaction initial velocity in Relative Fluorescent Units in time (RFU per minute). GraphPad software was used to determine IC$_{50}$ values.

Determination of inhibitor binding kinetics. For all kinetic assays, reactions were performed in triplicate in 100 μL assay with final concentrations of 25 μM fluorescence substrate DABCYL-KTSAVLQSGFRKME-EDANS from BPS Biosciences (San Diego, CA, USA), and 50 nM 3CL^{pro}. Buffer A contained 20 mM HEPES – pH 7.5, 10 mM NaCl and 1 mM EDTA. Buffer B contained 20 mM HEPES – pH 7.5, 10 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.1% BSA.

To each well, 4 µL of inhibitor dissolved in DMSO was added to 20 µL buffer A. Subsequently, 56 µL fluorescence substrate dissolved in buffer B was added, followed by 20 µL 3CL^{pro} (with His-tag) in buffer B. Reactions were monitored by following the fluorescence as a function of time (excitation at 360 nm, emission at 460 nm) using a Synergy™ H4 Hybrid Multi-Mode Microplate Reader (Winooski, VT, USA). Controls were (i) no inhibition: 4 µL of DMSO, and (ii) blank: 4 µL DMSO and 20 µL of buffer B. The reactions ran for at least 1 h, and the first 10 minutes were used for fitting to Eq. (1) by non-linear regression. In the case of inhibitor **4.15**, the first 3 minutes were used to calculate the initial velocity by linear regression. All fitting was done using GraphPad software.

Protein crystallization and structure solution. The enzyme was buffer exchanged into 20 mM Tris pH8, 100 mM NaCl, 1 mM DTT and concentrated to 5 mg/mL and was then incubated with 450 µM of compound **4.13** for 1 h at room temp. Following incubation, the sample was filtered using a 0.22 µm filter and used for crystallization trials. Crystals were grown using the sitting drop method at 22 °C. 200 nL enzyme was mixed with 200 nL well solution (30% PEG2000 MME, 0.2 M Potassium thiocyanate) and allowed to equilibrate against 50 µL well solution. The crystals were cryo-protected using well solution supplemented with 20% ethylene glycol and flash frozen in liquid nitrogen. Data was collected at the Canadian Light Source CMCF-BM beamline and processed in space group C 1 2 1. The structure was solved in PHASER with a previously published structure of the enzyme (PDB: 6WTK) as a search model. Restraints for the covalently bonded inhibitor were generated using AceDRG in CCP4i2, and the model was refined with REFMAC5 and Coot.

4.7   References

1.   Kumalo, H. M.; Bhakat, S.; Soliman, M. E. S., Theory and Applications of Covalent Docking in Drug Discovery: Merits and Pitfalls. *Molecules* **2015,** *20* (2), 1984-2000.

2.   Oyedele, A. Q. K.; Ogunlana, A. T.; Boyenle, I. D.; Adeyemi, A. O.; Rita, T. O.; Adelusi, T. I.; Abdul-Hammed, M.; Elegbeleye, O. E.; Odunitan, T. T., Docking covalent targets for drug discovery: stimulating the computer-aided drug design community of possible pitfalls and erroneous practices. *Mol Divers* **2022**.*4*, 1879-1903

3.   Schlottau, K.; Rissmann, M.; Graaf, A.; Schön, J.; Sehl, J.; Wylezich, C.; Höper, D.; Mettenleiter, T. C.; Balkema-Buschmann, A.; Harder, T., et al., SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. *Lancet Micr.* **2020,** *1*, E218-E225.

4.   Shi, Z.; Hu, Z., A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res* **2008,** *133* (1), 74-87.

5.   Abdollahi, E.; Champredon, D.; Langley, J. M.; Galvani, A. P.; Moghadas, S. M., Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. *Can Med Assoc J* **2020,** *192* (25), E666-E670.

6.   Tatsui, G. J., *J. Pharm Soc Jpn* **1928,** *48*, 92.

7.   Organization, W. H. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/ (accessed Oct, 7).

8.   Modi, C.; Böhm, V.; Ferraro, S.; Stein, G.; Seljak, U., Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. *Nat. Comm.* **2021,** *12* (1), 2729.

9.   Stankiewicz, K. Pfizer's CEO says Covid vaccine effectiveness drops to 84% after six months. https://www.cnbc.com/2021/07/28/pfizers-ceo-says-covid-vaccine-effectiveness-drops-to-84percent-after-six-months.html (accessed Aug. 25).

10.   Ledford, H., Six months of COVID vaccines: what 1.7 billion doses have taught scientists. *Nature* **2021,** *594*, 164-167.

11.   Dai, L.; Gao, G. F., Viral targets for vaccines against COVID-19. *Nat Rev Immunol* **2021,** *21* (2), 73-82.

12.   Chodick, G.; Tene, L.; Rotem, R. S.; Patalon, T.; Gazit, S.; Ben-Tov, A.; Weil, C.; Goldshtein, I.; Twig, G.; Cohen, D., et al., The Effectiveness of the Two-Dose BNT162b2 Vaccine: Analysis of Real-World Data. *Clin Infect Dis* **2021**, *74* (3), 472-478

13.   Merck Merck and Ridgeback's Molnupiravir, an Oral COVID-19 Antiviral Medicine, Receives First Authorization in the World. https://www.merck.com/news/merck-and-ridgebacks-molnupiravir-an-oral-covid-19-antiviral-medicine-receives-first-authorization-in-the-world/ (accessed Nov. 9).

14.   Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J., et al., An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374* (6575), 1586-1593.

15.   Couzin-Frankel, J., Antiviral pills could change pandemic's course. *Science* **2021**, 799-800.

16.   Cannalire, R.; Cerchia, C.; Beccari, A. R.; Di Leva, F. S.; Summa, V., Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J Med Chem* **2020**.

17.   Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L., et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020,** *579* (7798), 270-273.

18.   Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J., et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe* **2020**, *27* (3), 325-328

19.   Rathnayake, A. D.; Kim, Y.; Dampalla, C. S.; Nguyen, H. N.; Jesri, A.-R. M.; Kashipathy, M. M.; Lushington, G. H.; Battaile, K. P.; Lovell, S.; Chang, K.-O., et al., Structure-Guided Optimization of Dipeptidyl Inhibitors of Norovirus 3CL Protease. *J Med Chem* **2020,** *63*, 11945-11963

20.   Kuo, C.-J.; Shie, J.-J.; Fang, J.-M.; Yen, G.-R.; Hsu, J. T. A.; Liu, H.-G.; Tseng, S.-N.; Chang, S.-C.; Lee, C.-Y.; Shih, S.-R., et al., Design, synthesis, and evaluation of 3C protease inhibitors as anti-enterovirus 71 agents. *Bioorg Med Chem* **2008,** *16* (15), 7388-7398.

21.   Dragovich, P. S.; Prins, T. J.; Zhou, R.; Webber, S. E.; Marakovits, J. T.; Fuhrman, S. A.; Patick, A. K.; Matthews, D. A.; Lee, C. A.; Ford, C. E., et al., Structure-Based Design, Synthesis, and Biological Evaluation of Irreversible Human Rhinovirus 3C Protease Inhibitors. 4. Incorporation of P1 Lactam Moieties as l-Glutamine Replacements. *J Med Chem* **1999,** *42* (7), 1213-1224.

22.   Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C., et al., Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020,** *582* (7811), 289-293.

23.   Su, H.-x.; Yao, S.; Zhao, W.-f.; Li, M.-j.; Liu, J.; Shang, W.-j.; Xie, H.; Ke, C.-q.; Hu, H.-c.; Gao, M.-n., et al., Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sin.* **2020,** *41* (9), 1167-1177.

24.   Kneller, D. W.; Galanie, S.; Phillips, G.; O'Neill, H. M.; Coates, L.; Kovalevsky, A., Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* **2020,** *28* (12), 1313-1320.e3.

25.   Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A., Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun* **2020,** *11* (1), 3202.

26.   Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H., An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J Med Chem* **2016,** *59* (14), 6595-6628.

27.   Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal structure of SARS-CoV-2 main protease provides a basis for design

of improved α-ketoamide inhibitors. *Science* **2020,** *368* (6489), 409-412.

28.   Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A., et al., Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun* **2020,** *11* (1), 4282.

29.   Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R., Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003,** *300* (5626), 1763-1767.

30.   Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S., et al., Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent Sci* **2020,** *6* (3), 315-331.

31.   Westberg, M.; Su, Y.; Zou, X.; Ning, L.; Hurst, B.; Tarbet, B.; Lin, M. Z., Rational design of a new class of protease inhibitors for the potential treatment of coronavirus diseases. *bioRxiv* **2020**, 2020.09.15.275891.

32.   Iketani, S.; Forouhar, F.; Liu, H.; Hong, S. J.; Lin, F.-Y.; Nair, M. S.; Zask, A.; Huang, Y.; Xing, L.; Stockwell, B. R., et al., Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors. *Nat Commun* **2021,** *12* (1), 2016.

33.   Boras, B.; Jones, R. M.; Anson, B. J.; Arenson, D.; Aschenbrenner, L.; Bakowski, M. A.; Beutler, N.; Binder, J.; Chen, E.; Eng, H., et al., Discovery of a Novel Inhibitor of Coronavirus 3CL Protease as a Clinical Candidate for the Potential Treatment of COVID-19. *bioRxiv* **2021**, 2020.09.12.293498.

34.   Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y., et al., Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J Med Chem* **2020,** *63* (21), 12725-12747.

35.   Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y., et al., Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J Med Chem* **2020,** *63* (21), 12725-12747.

36.   Pfizer's novel COVID-19 antiviral heads to clinical trials. https://cen.acs.org/pharmaceuticals/drug-discovery/Pfizers-novel-COVID-19-antiviral/98/web/2020/09 (accessed 2023-11-01).

37.   Study of PF-07321332 in healthy participants - NCT04756531. https://www.pfizer.com/science/find-a-trial/nct04756531 (accessed 2021-04-29).

38.   Pfizer's PAXLOVID™ Receives FDA Approval for Adult Patients at High Risk of Progression to Severe COVID-19. https://www.pfizer.com/news/press-release/press-release-detail/pfizers-paxlovidtm-receives-fda-approval-adult-patients (accessed 2023-11-01).

39.   Bai, B.; Belovodskiy, A.; Hena, M.; Kandadai, A. S.; Joyce, M. A.; Saffran, H. A.; Shields, J. A.; Khan, M. B.; Arutyunova, E.; Lu, J., et al., Peptidomimetic α-Acyloxymethylketone Warheads with Six-Membered Lactam P1 Glutamine Mimic: SARS-CoV-2 3CL Protease Inhibition, Coronavirus Antiviral Activity, and in Vitro Biological Stability. *J*

Med Chem **2021**, *65* (4), 2905-2925

40.   Bai, B.; Arutyunova, E.; Khan, M. B.; Lu, J.; Joyce, M. A.; Saffran, H. A.; Shields, J. A.; Kandadai, A. S.; Belovodskiy, A.; Hena, M., et al., Peptidomimetic nitrile warheads as SARS-CoV-2 3CL protease inhibitors. *RSC Med Chem* **2021,** *12* (10), 1722-1730.

41.   Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J., et al., α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J Med Chem* **2020,** *63* (9), 4562-4578.

42.   Kim, Y.; Liu, H.; Galasiti Kankanamalage, A. C.; Weerasekara, S.; Hua, D. H.; Groutas, W. C.; Chang, K.-O.; Pedersen, N. C., Reversal of the Progression of Fatal Coronavirus Infection in Cats by a Broad-Spectrum Coronavirus Protease Inhibitor. *PLoS Pathog.* **2016,** *12* (3), e1005531.

43.   Vuong, W.; Fischer, C.; Khan, M. B.; van Belkum, M. J.; Lamer, T.; Willoughby, K. D.; Lu, J.; Arutyunova, E.; Joyce, M. A.; Saffran, H. A., et al., Improved SARS-CoV-2 Mpro inhibitors based on feline antiviral drug GC376: Structural enhancements, increased solubility, and micellar studies. *Eur J Med Chem* **2021,** *222*, 113584.

44.   Chen, L.-R.; Wang, Y.-C.; Lin, Y. W.; Chou, S.-Y.; Chen, S.-F.; Liu, L. T.; Wu, Y.-T.; Kuo, C.-J.; Chen, T. S.-S.; Juang, S.-H., Synthesis and evaluation of isatin derivatives as effective SARS coronavirus 3CL protease inhibitors. *Bioorg Med Chem Lett* **2005,** *15* (12), 3058-3062.

45.   Zhang, J.; Huitema, C.; Niu, C.; Yin, J.; James, M. N. G.; Eltis, L. D.; Vederas, J. C., Aryl methylene ketones and fluorinated methylene ketones as reversible inhibitors for severe acute respiratory syndrome (SARS) 3C-like proteinase. *Bioorg Chem* **2008,** *36* (5), 229-240.

46.   Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.; Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N., et al., Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent Sci* **2021**.

47.   Zhang, C. H.; Spasov, K. A.; Reilly, R. A.; Hollander, K.; Stone, E. A.; Ippolito, J. A.; Liosi, M. E.; Deshmukh, M. G.; Tirado-Rives, J.; Zhang, S., et al., Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med Chem Lett* **2021,** *12* (8), 1325-1332.

48.   Jacobs, L.; van der Westhuyzen, A.; Pribut, N.; Dentmon, Z. W.; Cui, D.; D'Erasmo, M. P.; Bartsch, P. W.; Liu, K.; Cox, R. M.; Greenlund, S. F., et al., Design and Optimization of Novel Competitive, Non-peptidic, SARS-CoV-2 M(pro) Inhibitors. *ACS Med Chem Lett* **2023,** *14* (10), 1434-1440.

49.   Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T., et al., Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *J Med Chem* **2022,** *65* (9), 6499-6512.

50.   Mukae, H.; Yotsuyanagi, H.; Ohmagari, N.; Doi, Y.; Imamura, T.; Sonoyama, T.; Fukuhara, T.; Ichihashi, G.; Sanaki, T.; Baba, K., et al., A Randomized Phase 2/3 Study of Ensitrelvir, a Novel Oral SARS-CoV-2 3C-Like Protease Inhibitor, in Japanese Patients with

Mild-to-Moderate COVID-19 or Asymptomatic SARS-CoV-2 Infection: Results of the Phase 2a Part. *Antimicrob Agents Chemother* **2022,** *66* (10), e0069722.

51.   Drayman, N.; Jones, K. A.; Azizi, S.-A.; Froggatt, H. M.; Tan, K.; Maltseva, N. I.; Chen, S.; Nicolaescu, V.; Dvorkin, S.; Furlong, K., et al., Drug repurposing screen identifies masitinib as a 3CLpro inhibitor that blocks replication of SARS-CoV-2 in vitro. *bioRxiv* **2020**, 2020.08.31.274639.

52.   De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A.; Moitessier, N., Covalent inhibitors design and discovery. *Eur J Med Chem* **2017,** *138*, 96-114.

53.   Zhou, S.; Chan, E.; Duan, W.; Huang, M.; Chen, Y. Z., Drug bioactivation, covalent binding to target proteins and toxicity relevance. *Drug Metab Rev* **2005,** (37), 41-213.

54.   Guterman, L., Covalent Drugs Form Long-Lived Ties. *C&EN* **2011,** *89*, 19-26.

55.   Tang, B.; He, F.; Liu, D.; Fang, M.; Wu, Z.; Xu, D., AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *bioRxiv* **2020**, 2020.03.03.972133.

56.   Nguyen, D. D.; Gao, K.; Chen, J.; Wang, R.; Wei, G.-W., Potentially highly potent drugs for 2019-nCoV. *bioRxiv* **2020**, 2020.02.05.936013.

57.   Roy, R.; Sk, M. F.; Jonniya, N. A.; Poddar, S.; Kar, P., Finding Potent Inhibitors Against SARS-CoV-2 Main Protease Through Virtual Screening, ADMET, and Molecular Dynamic Simulation Studies. . *ChemRxiv* **2020,** *Preprint*.

58.   Zaidman, D.; Gehrtz, P.; Filep, M.; Fearon, D.; Prilusky, J.; Duberstein, S.; Cohen, G.; Owen, D.; Resnick, E.; Strain-Damerell, C., et al., An automatic pipeline for the design of irreversible derivatives identifies a potent SARS-CoV-2 Mpro inhibitor. *bioRxiv* **2020**, 2020.09.21.299776.

59.   Moitessier, N.; Pottel, J.; Therrien, E.; Englebienne, P.; Liu, Z.; Tomberg, A.; Corbeil, C. R., Medicinal chemistry projects requiring imaginative structure-based drug design methods. *Acc Chem Res* **2016,** *49* (9), 1646-1657.

60.   Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S. A.; Chase, P.; Eggler, A.; Dawson, E. S.; Baez-Santos, Y. M.; Tomar, S., et al., Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease. *J Med Chem* **2013,** *56* (2), 534-546.

61.   Marcaccini, S.; Torroba, T., The use of the Ugi four-component condensation. *Nat Protoc* **2007,** *2* (3), 632-9.

62.   Gedey, S.; Van der Eycken, J.; Fülöp, F., Liquid-Phase Combinatorial Synthesis of Alicyclic β-Lactams via Ugi Four-Component Reaction. *Org Lett* **2002,** *4* (11), 1967-1969.

63.   Kong, H. H.; Pan, H. L.; Ding, M. W., Synthesis of 2-Tetrazolyl-Substituted 3-Acylpyrroles via a Sequential Ugi-Azide/Ag-Catalyzed Oxidative Cycloisomerization Reaction. *J Org Chem* **2018,** *83* (20), 12921-12930.

64.   Freese, T.; Lücke, A. L.; Schmidt, C. A. S.; Polamo, M.; Nieger, M.; Namyslo, J. C.; Schmidt, A., Anionic N-heterocyclic carbenes derived from sydnone imines such as

molsidomine. Trapping reactions with selenium, palladium, and gold. *Tetrahedron* **2017,** *73* (36), 5350-5357.

65.   St John, S. E.; Mesecar, A. D. Broad spectrum non-covalent coronavirus protease inhibitors. 2018.

66.   Petri, L.; Egyed, A.; Bajusz, D.; Imre, T.; Hetényi, A.; Martinek, T.; Ábrányi-Balogh, P.; Keserű, G. M., An electrophilic warhead library for mapping the reactivity and accessibility of tractable cysteines in protein kinases. *Eur. J. Med. Chem.* **2020,** *207*, 112836.

67.   Turlington, M.; Chun, A.; Tomar, S.; Eggler, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J. S.; Dawson, E.; Saldanha, A.; Chase, P., et al., Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg. Med. Chem. Lett.* **2013,** *23* (22), 6172-6177.

68.   Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nat Rev Drug Discov* **2011,** *10* (4), 307-17.

69.   Strelow, J. M., A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS Discov* **2017,** *22* (1), 3-20.

70.   Mons, E.; Roet, S.; Kim, R. Q.; Mulder, M. P. C., A Comprehensive Guide for Assessing Covalent Inhibition in Enzymatic Assays Illustrated with Kinetic Simulations. *Curr Protoc* **2022,** *2* (6), e419.

71.   Kitz, R.; Wilson, I. B., Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. *J Biol Chem* **1962,** *237*, 3245-9.

72.   Flanagan, M. E.; Abramite, J. A.; Anderson, D. P.; Aulabaugh, A.; Dahal, U. P.; Gilbert, A. M.; Li, C.; Montgomery, J.; Oppenheimer, S. R.; Ryder, T., et al., Chemical and computational methods for the characterization of covalent reactive groups for the prospective design of irreversible inhibitors. *J Med Chem* **2014,** *57* (23), 10072-9.

73.   Martin, J. S.; MacKenzie, C. J.; Fletcher, D.; Gilbert, I. H., Characterising covalent warhead reactivity. *Bioorg Med Chem* **2019,** *27* (10), 2066-2074.

74.   Tan, B.; Sacco, M.; Tan, H.; Li, K.; Joyce, R.; Zhang, X.; Chen, Y.; Wang, J., Exploring diverse reactive warheads for the design of SARS-CoV-2 main protease inhibitors. *Eur J Med Chem* **2023,** *259*, 115667.

75.   Labarre, A.; Stille, J. K.; Patrascu, M. B.; Martins, A.; Pottel, J.; Moitessier, N., Docking Ligands into Flexible and Solvated Macromolecules. 8. Forming New Bonds horizontal line Challenges and Opportunities. *J Chem Inf Model* **2022,** *62* (4), 1061-1077.

76.   Stille, J. T., J.; Wang, G.; Venegas, F. A.; Hennecker, C.; Rueda, A. M.; Miron, C. E.; Pinus, S.; Labarre, A.; Plescia, J.; Patrascu, M. B.; Vlaho, D.; Huot, M.; Mittermaier, A. K.; Moitessier, N. , Design, Synthesis and Biological Evaluation of Novel SARS-CoV-2 3CLpro Covalent Inhibitors. *ChemRxiv* **2020**.

77.   Ma, C.; Xia, Z.; Sacco, M. D.; Hu, Y.; Townsend, J. A.; Meng, X.; Choza, J.; Tan, H.; Jang, J.; Gongora, M. V., et al., Discovery of Di- and Trihaloacetamides as Covalent SARS-CoV-2 Main Protease Inhibitors with High Target Specificity. *J Am Chem Soc* **2021,** *143* (49),

20697-20709.

78.   Quan, B. X.; Shuai, H.; Xia, A. J.; Hou, Y.; Zeng, R.; Liu, X. L.; Lin, G. F.; Qiao, J. X.; Li, W. P.; Wang, F. L., et al., An orally available M(pro) inhibitor is effective against wild-type SARS-CoV-2 and variants including Omicron. *Nat Microbiol* **2022,** *7* (5), 716-725.

79.   Yamane, D.; Onitsuka, S.; Re, S.; Isogai, H.; Hamada, R.; Hiramoto, T.; Kawanishi, E.; Mizuguchi, K.; Shindo, N.; Ojida, A., Selective covalent targeting of SARS-CoV-2 main protease by enantiopure chlorofluoroacetamide. *Chem Sci* **2022,** *13* (10), 3027-3034.

80.   Krippendorff B.F.; Neuhaus, R,; Lienau, P.; Reichel, A.; Huisinga, W. Mechanism-based inhibition: deriving K(I) and k(inact) directly from time-dependent IC(50) values. *J Biomol Screen* **2009**, *14* (8), 913-923.

CHAPTER 5:

OPTIMIZATION OF COVALENT COVID-19 ANTIVIRALS TARGETING 3CL^PRO

The work on this chapter is based, in part, on work published in:

J. K. Stille‡ , J. Tjutrins‡ , G. Wang‡ , F. A. Venegas‡ , C. Hennecker, A. M. Rueda, I. Sharon, N. Blaine, C. E. Miron, S. Pinus, A. Labarre, J. Plescia, M. Burai Patrascu, X. Zhang, A. S. Wahba, D. Vlaho, M. J. Huot, T. M. Schmeing, A. K. Mittermaier, and N. Moitessier. Design, synthesis and *in vitro* evaluation of novel SARS-CoV-2 3CL^pro covalent inhibitors. *Eur. J. Med. Chem.*, **2022**, *229*, 114046

Contribution of Authors

J.K. Stille contributed to the design of all compounds and performed all docking studies. **J.K. Stille** synthesized and characterized compounds 5.11-5.17, 5.23, 5.24, 5.31, 5.32, and all related intermediates, with the assistance of undergraduate student N. Blaine.

All remaining compounds were synthesized and characterized by J. Tjutrins and G. Wang.

In vitro testing was performed by G. Wang and F.A. Venegas. *In cellulo* testing was performed by collaborators at INRS.

5.1 Abstract

The ongoing COVID-19 pandemic has demonstrated an urgent need for the development of antiviral therapeutics as a complementary approach to vaccination, and the viral protease 3CL^{pro} has been identified as a promising therapeutic target as it is necessary for viral replication and function. 3CL^{pro} is also conserved across several virus families, including enterovirus, norovirus and rhinovirus, suggesting that inhibitors for this target could play a role in pandemic preparedness against future outbreaks. This chapter details the optimization of our initial covalent inhibitor scaffolds, focusing on optimizing their interactions with the S1, S2 and S4 subsites. Additionally, bioisosteric replacement of the S3 amide was pursued as a viable strategy towards the developed of non-peptidic 3CL^{pro} inhibitors. The optimized inhibitors were tested *in vitro* and the antiviral activity of the most promising analogues were further evaluated *in cellulo*, with several displaying low micromolar inhibition of viral replication.

5.2 Introduction

As introduced in Chapter 4, the recent COVID-19 pandemic has sparked intense efforts to develop therapeutic agents to combat this deadly disease. COVID-19 is caused by the SARS-CoV-2 virus, an RNA virus that is part of the large coronavirus family that has been associated with several previous viral outbreaks including Middle East Respiratory Syndrome (MERS, caused by MERS-CoV) and Severe Acute Respiratory Syndrome (SARS, caused by SARS-CoV). SARS-CoV-2 has resulted in more than 7 million fatalities worldwide,[1] providing significant motivation for the development of therapeutics to treat and prevent outbreaks.

While vaccines have been a central pillar of efforts against COVID-19, they are not without limitations and could be complemented by the development of antiviral therapeutics. The majority of COVID-19 vaccines have focused on eliciting an immune response to the SARS-CoV-2 spike protein, a structural protein on the surface of the virus that facilitates entry into host cells.[2] Mutations in the spike protein have been observed in several SARS-CoV-2 variants, resulting in an attenuation of vaccine efficacy.[3, 4] It has also been shown that vaccine efficacy decreases significantly within a 6 month period, suggesting that frequent vaccination against COVID-19 may be necessary.[5, 6] In order for a vaccine campaign to attain herd immunity, a large portion of the population would need to be vaccinated regularly, amplifying the number of people at risk of

adverse events following immunization.[7] Additionally, groups such as the elderly and immunocompromised, who are most at risk for severe infection, have also been shown to elicit a lower responsiveness to COVID-19 vaccination.[8, 9] In light of these limitations, the development of antiviral therapeutics (especially those that act through alternative viral targets to the spike protein) can be considered as a complementary approach to vaccines in the fight against SARS-CoV-2.

One of the main targets identified for the development of SARS-CoV-2 antivirals has been 3CL^{pro}, a cysteine protease required for viral replication. 3CL^{pro} has previously been identified as a promising antiviral target as it is conserved among the broader *pisoniverites* class which encompasses not only coronaviruses, but also a range of other viruses such as rhinovirus, norovirus and poliovirus.[10] Similar to other previously identified 3CL^{pro} structures, SARS-CoV-2 3CL^{pro} is able to achieve substrate selectivity through well-defined S1'-S4 subsites (using the Schechter-Berger nomenclature) that preferentially bind specific amino acid residues. Notably, SARS-CoV-2 3CL^{pro} has been shown to have a strong preference for substrates with glutamine and leucine at the P1 and P2 positions, respectively.[11] A common approach to the design of protease inhibitors is to apply knowledge of residue preferences to generate substrate mimics that can compete for the active site. These peptidomimetic inhibitors can then be further optimized to improve potency and/or selectivity, while also taking into account pharmacokinetic considerations such as oral bioavailability, solubility, and metabolism. This approach has been employed with SARS-CoV-2 3CL^{pro}, leading to several potent peptidomimetic inhibitors such as **5.1**,[12, 13] **5.2**,[14, 15] and **PF-07321332**[16-18] (nirmatrelvir) as shown in **Figure 5.1**.

5.1, SARS-CoV-2 3CL^{pro}
IC$_{50}$ = 34 nM

5.2, SARS-CoV-2 3CL^{pro}
IC$_{50}$ = 0.07 μM

PF-07321332, SARS-CoV-2 3CL^{pro}
IC$_{50}$ = 3.1 nM
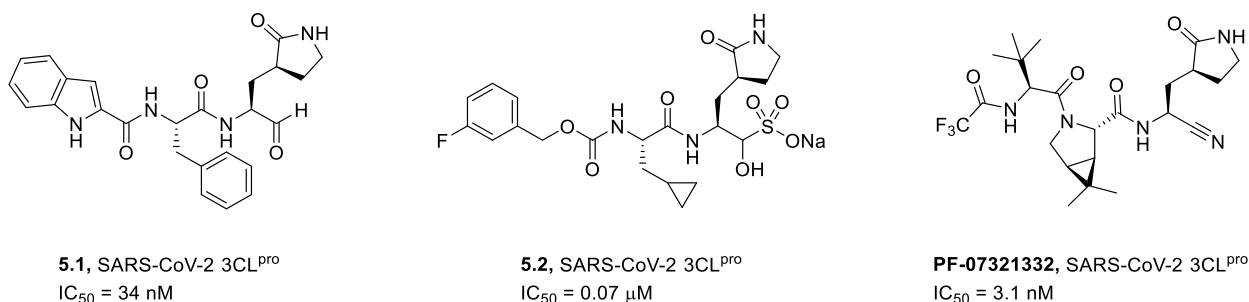
Figure 5.1. Select examples of reported peptidomimetic SARS-CoV-2 3CL^{pro} inhibitors.

Given our previous experience in the computational design of covalent inhibitors,[19-21] we were interested in applying our expertise to develop 3CL^{pro} inhibitors using a computer-guided approach. As detailed in Chapter 4, we used docking studies to support our design of incorporating a covalent warhead into a non-covalent inhibitor scaffold, leading to one of the first non-peptidomimetic inhibitors for SARS-CoV-2 3CL^{pro} with sub-micromolar potency. Additionally, the predicted covalent binding modes of our inhibitors were confirmed with experimental crystal structures that demonstrate that our inhibitor forms favourable non-covalent interactions with the S1-S4 pockets (**Figure 5.2**).
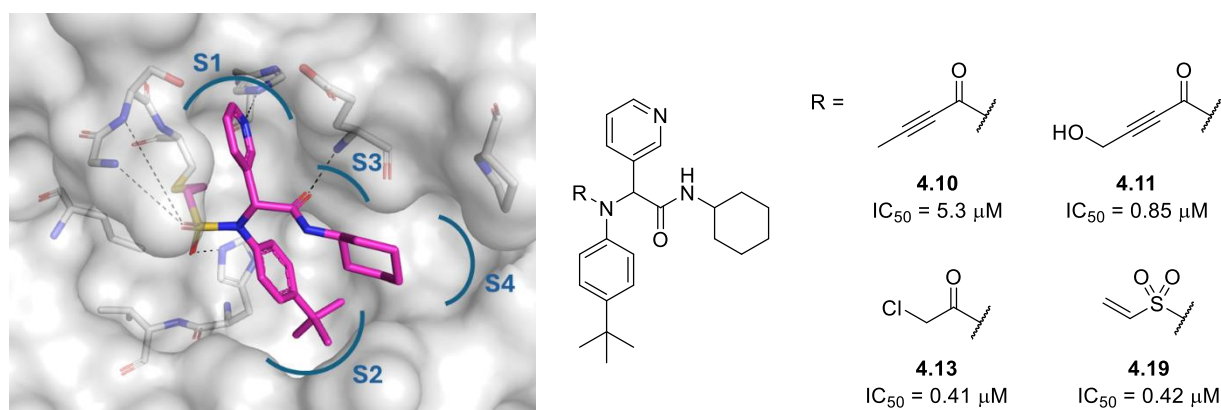
Figure 5.2. Left: crystal structure of **4.19** bound to 3CL^{pro} (PDB: 7MLG). Right: Structures and activities of our most promising covalent analogues.

While the reported SARS-CoV-2 3CL^{pro} inhibitors have demonstrated promising biological activity, the design of 3CL^{pro} inhibitors should also include considerations regarding preparedness towards the development of antiviral resistance and future viral outbreaks. A challenge with many peptidomimetic inhibitors is that they are often prepared through long linear syntheses where any changes to the inhibitor scaffold could require carrying analogues through several synthetic steps. For example, the reported synthesis of **PF-07321332** requires six steps, with the S2, S3 and S4 binding groups introduced in the first step.[17] One aspect that drew us to the scaffold of **4.19** was the modularity of its synthesis – it can be prepared via a multicomponent Ugi reaction that combines four separate starting materials (acid, amine, aldehyde and isocyanide) in only one step to generate a single product. This could be advantageous for optimizing its non-covalent binding interactions and for investigating structure-activity relationships of the S1-S4 binding pockets, as

changes to the scaffold can be rapidly accessed without requiring analogues to be carried through multiple synthetics steps. This also provides a more adaptable inhibitor scaffold that could be useful for the future design of inhibitors towards alternative 3CL^{pro} targets (i.e. in the case of antiviral resistance or novel viral outbreaks).

5.3 Optimization of Covalent Inhibitors

Following our search for an optimal covalent warhead, our attention then turned to the optimization of non-covalent interactions formed by the different R groups present in the parent scaffold. Given the modularity of the Ugi reaction used to synthesize the inhibitors, it was envisioned that different amines (R^2), aldehydes (R^3) and isocyanides (R^4) could be explored to investigate their impact on binding affinity.
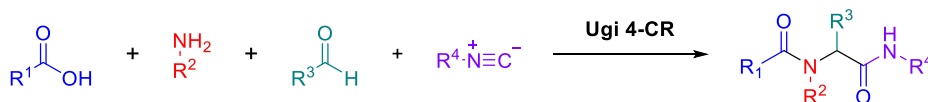
Scheme 5.1. Modularity of the Ugi-4-CR product.

5.3.1 Varying the R^2 Component

5.3.1.1 Design

The crystal structure of **4.19** bound to 3CL^{pro} shows that the amine R^2 group sits in a hydrophobic pocket, and its binding is largely mediated by hydrophobic interactions to Met^{49}, Met^{165}, and His^{41} (**Figure 5.2**). This R group occupies the enzyme's S2 site which is known to be selective for leucine, though phenylalanine and valine are also tolerated.[11] It is envisioned that further exploration of hydrophobic groups at this position could lead to analogues with improved binding. Additionally, although the interactions between **4.19** and the S2 site are largely hydrophobic, the nearby residues do contain heteroatoms capable of forming hydrogen-bond interactions (e.g., His^{41}) that could be targeted by an optimized inhibitor.
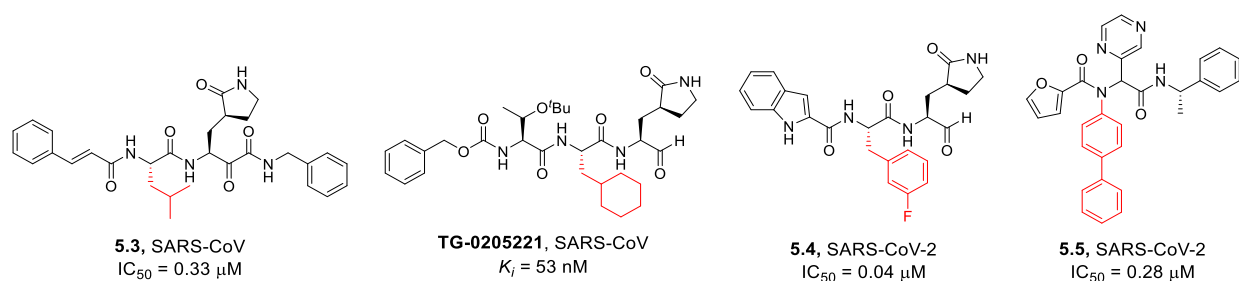
Figure 5.3. Select 3CL^{pro} inhibitors for SARS-CoV and SARS-CoV-2, with S2 binding groups shown in red.

Previously reported peptidomimetic 3CL^{pro} inhibitors for SARS-CoV and SARS-CoV-2 have similarly focused on hydrophobic groups occupying the S2 site. While many have incorporated an isobutyl group at this position, other hydrophobic groups such as cyclohexyl and benzyl derivatives have been used in potent 3CL^{pro} inhibitors (**Figure 5.3**).[13, 22, 23] As observed in SARS-CoV 3CL^{pro}, residues in the S2 site can undergo large conformational changes on inhibitor and substrate binding.[24, 25] This can be observed when comparing the crystal structures of our chemical series (PDB: 7MLF and 7MLG) to other reported inhibitors (e.g., PDB: 6Y2G). The change in conformation of Met^{49} results in a much smaller S2 binding site in 6Y2G as compared to 7MLG, suggesting that the binding mode of our chemical series may accommodate or prefer larger hydrophobic residues at this position (**Figure 5.4**). This has been observed with previous structure-activity relationship studies on the non-covalent **X77** scaffold, which found that replacement of the tert-butyl group by small substituents led to a reduction in potency and replacement of the aromatic group with a cyclopropyl substituent led to complete loss of activity.[26, 27] Concurrent to our own work on covalent inhibitors, Wang *et al.* developed a series of non-covalent inhibitors of SARS-CoV-2 3CL^{pro} based on the same scaffold.[28] As with X77 inhibition of SARS-CoV 3CL^{pro}, they observed the greatest potency with aromatic amines containing bulky *para* substituents. Their most potent inhibitor (compound **5.5**, **Figure 5.3**) contained a diphenyl group in the P2 position that enabled favourable pi-stacking interactions with His^{41}.
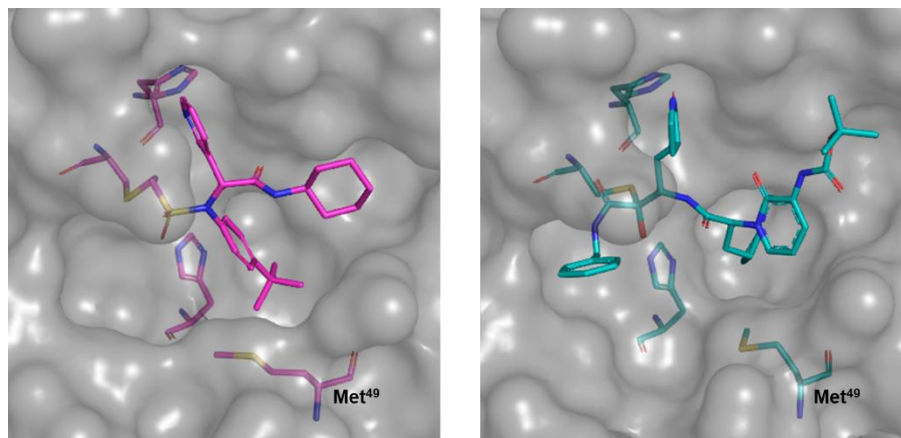
Figure 5.4. Comparison of crystal structure binding sites for 3CL^{pro} bound to inhibitor **4.19** (left, PDB 7MLG) and a peptidomimetic inhibitor (right, PDB 6Y2G).

Our design of analogues was also supported by docking studies of amine analogues of covalent inhibitor **4.10**. As shown in **Figure 5.5**, several bulky hydrophobic amine analogues were predicted to bind in a similar fashion as **4.10**, maintaining their occupation of the same S2 pocket. Promisingly, these analogues showed small improvements in docking scores, suggesting that modifications of the amine substituent could result in improved binding affinity.
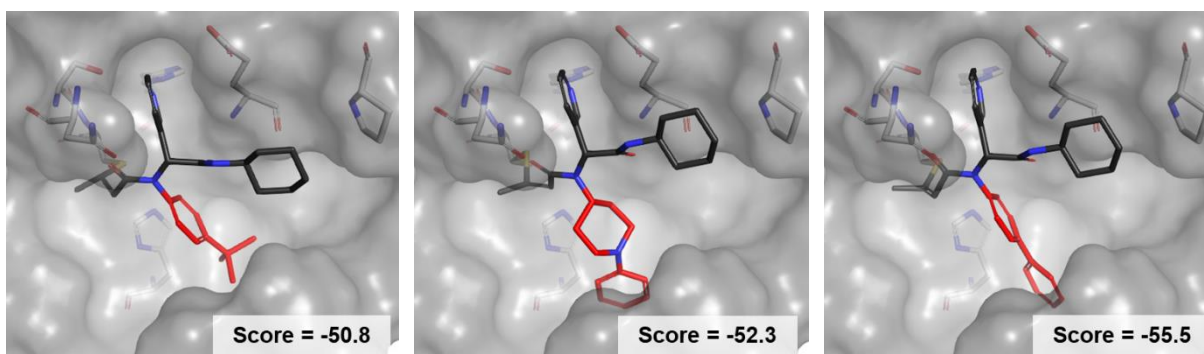
Figure 5.5. Docking poses of select amine analogues, with the amine substituent shown in red and its corresponding docking score (lower = better). Left: alkyne analogue **4.10**. Middle: phenyl piperidine analogue (**5.13**). Right: diphenyl analogue (**5.16**).

5.3.1.2   Synthesis

With these consideration in mind, a library of analogues featuring different R^2 substituents was

synthesized via a 4-component Ugi, with analogues selected based on promising docking results and the commercial availability of amine starting materials. As shown in **Scheme 5.2**, analogues with several different aromatic substituents were covered. The aromatic ring was also replaced with substituted piperidines to investigate the influence of an aliphatic ring at this position. Additionally, benzyl and napthyl derivates were prepared to investigate the effect of different aromatic groups in the R^2 position. Given the proximity of the R^2 group to His[41], two analogues featuring a substituted pyridine in place of the phenyl group were synthesized to probe the possibility of favourable hydrogen-bonding to the nitrogen atom. The analogues were synthesized primarily using the alkynyl warhead from inhibitor **4.10** (**Scheme 5.2**) as it enabled analogues to be synthesized in a single step, compared to the two steps required to access the vinyl sulfonamide analogue. An additional two analogues were synthesized by the same methods using the hydroxy-alkynyl warhead from inhibitor **4.11**.
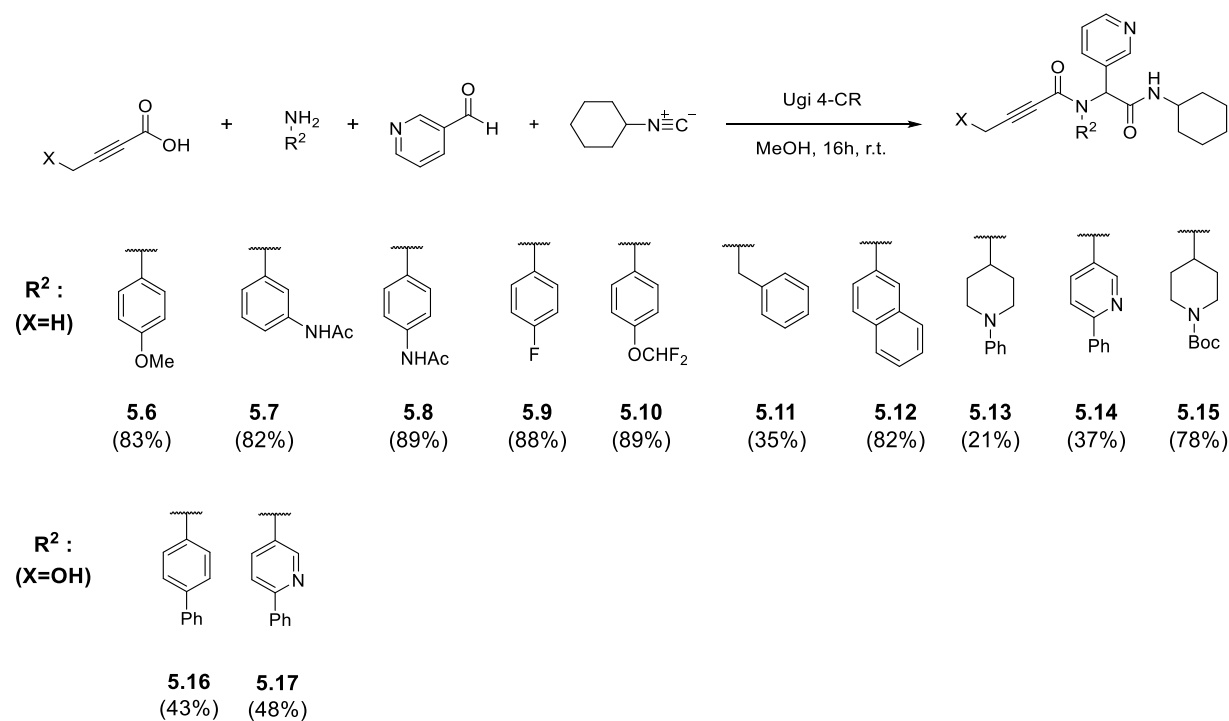
Scheme 5.2. Synthesis of R^2 analogues.

In the case of **5.13**, the amine starting material was synthesized in two steps from 4-(Boc-amino)-piperidine via Ullmann coupling followed by Boc-deprotection (**Scheme 5.3**). Additionally, the amine used in analogues **5.14** and **5.17** was synthesized in one step via Suzuki coupling of the

corresponding 5-amino-2-bromopyridine (**Scheme 5.4**).
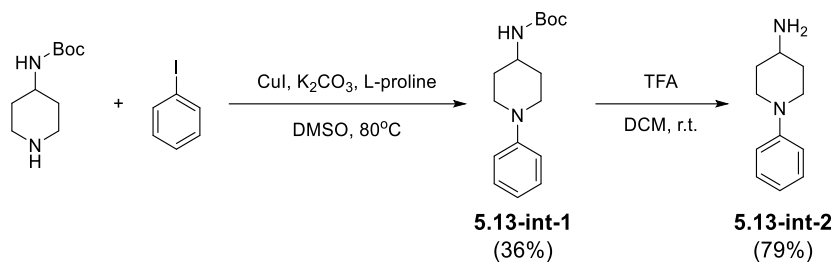
Scheme 5.3. Synthesis of amine starting material for analogue **5.13**.
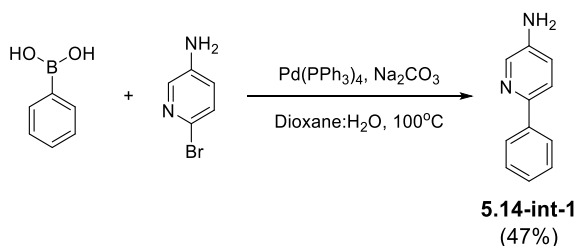
Scheme 5.4. Synthesis of amine starting material for analogues **5.14** and **5.17**.

Although promising docking results were observed with *tert*-butyl-substituted heterocyclic amines, attempts to perform the Ugi reaction with commercially available heterocycles were unsuccessful (**Scheme 5.5**). This is potentially a result of the low nucleophilicity of heterocyclic amines due to delocalization of the nitrogen lone pair, leading to slow imine formation under the experimental conditions.[29, 30] This may also be complicated by the potential for heterocyclic amidines to undergo a competing Groebke-Blackburn-Bienaymé reaction[31] to form undesired side products, and therefore heterocyclic analogues were not further pursued here.
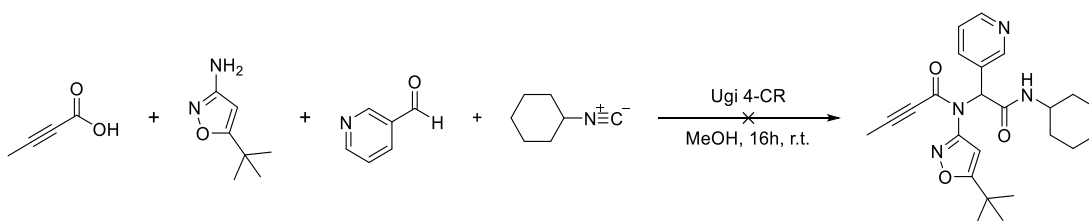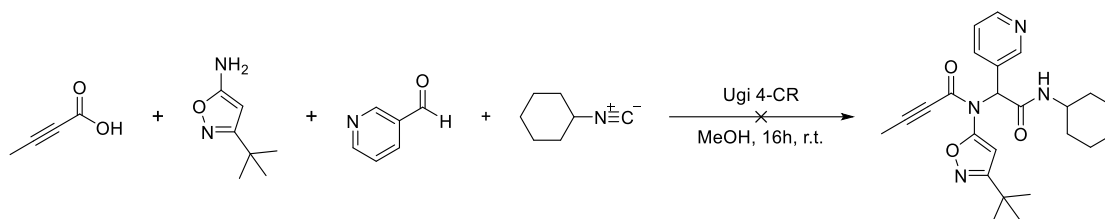
Scheme 5.5. Attempted synthesis of heterocyclic amine analogues.

5.3.1.3 *In Vitro* **Biological Evaluation**

Table 5.1. Biological evaluation of R^2 analogues
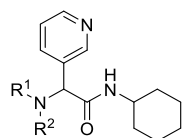
Entry	Compound	R^1	R^2	Inhibition (%)[a]	IC$_{50}$ (μM)
1	**4.10**			> 95	5.3 ± 0.8
2	**5.6**			39 ± 8	nd[b]
3	**5.7**			43 ± 3	nd[b]
4	**5.8**			23 ± 13	nd[b]
5	**5.9**			65 ± 3	nd[b]

6	**5.10**			77 ± 2	nd[b]
7	**5.11**			< 5	nd[b]
8	**5.12**			55 ± 1	nd[b]
9	**5.13**			35 ± 3	nd[b]
10	**5.14**			> 95	1.6 ± 0.7
11	**5.15**			12 ± 5[c]	nd[b]
12	**4.11**			> 95	0.85 ± 0.42
13	**5.16**			> 95	0.54 ± 0.15
14	**5.17**			> 95	1.29 ± 0.75

[a] The enzyme activity was measured with 114 nM 3CL[pro] and 50 μM of each potential inhibitor with incubation time of 30 min. [b] not determined. [c] Enzyme activity measured with 25 μM enzyme

Unfortunately, many of our attempts to replace the *tert*-butyl phenyl group proved unsuccessful. In an attempt to target one of the two methionine residues present in the R^2 binding site, a

hydrogen-bond donating acetamide substituent was introduced at both the *para* (**5.8**) and *meta* (**5.7**) positions, however no improvement in potency was observed with these modifications. Similarly, introduction of smaller substituents at the para position (**5.6, 5.9, 5.10**) also led to a decrease in inhibitor potency. Changing from a substituted phenyl group to a benzyl group (**5.11**) dramatically decreased the potency, as did changing the phenyl group to an aliphatic *N*-substituted piperidine (**5.13, 5.15**). We did see moderate improvement in potency when using a biphenyl group at the R^2 position, improving the IC_{50} from 0.85 μM (**4.11**) to 0.54 μM (**5.16**). In an attempt to target the His^{41} residue adjacent to the R^2 binding site, the biphenyl group was also modified to a pyridylphenyl group (**5.14, 5.17**). This change resulted in a slight improvement in potency when using the alkynyl warhead, decreasing the IC_{50} from 5.3 μM (**4.10**) to 1.6 μM (**5.14**). However, this improvement in potency was not maintained when applied to the hydroxyalkyne warhead of **4.11**.

5.3.2 Varying the R^3 Component

5.3.2.1 Design

The R^3 group of **4.19** occupies the enzyme's S1 site, which is known to almost exclusively bind glutamine at that position.[11] The pyridine moiety of **4.19** acts as a glutamine mimic by forming a hydrogen bond with His^{163}, an interaction that could potentially be optimized through exploration of different heterocyclic groups at this position. Previously reported covalent inhibitors of SARS-CoV 3CL^{pro} incorporated a dimethylated glutamine moiety at this position in order to prevent observed intramolecular cyclization between the amide and electrophilic warhead to an inactive form.[32] The use of lactams as glutamine mimics was pioneered by Dragovich *et al.* with their work on rhinovirus 3CL^{pro} inhibitors where it was observed that a lactam moiety maintained the desired *cis* amide configuration necessary to facilitate hydrogen bond donation of the amide nitrogen to an adjacent residue.[33] The rigidity of the cyclic lactam would also result in a decrease in the entropic penalty incurred by ligand binding, contributing to the improved binding affinity observed with the lactam analogues. On the basis of these observations, lactams have found widespread use as glutamine mimics in the development of 3CL^{pro} inhibitors, with several representative examples shown in **Figure 5.6** (e.g., GC-376, nirmatrelvir).[16, 34, 35] In addition to **X77**, several 3CL^{pro} inhibitors have been reported with heterocyclic groups occupying the S1 binding pocket. For example, the S1 site is occupied by an imidazole group with the peptidomimetic inhibitor **5.18** and

the non-covalent inhibitor **ML300** has been shown to interact with His[163] of SARS-CoV 3CL[pro] through its benzotriazole moiety (PDB: 4MDS).[36]
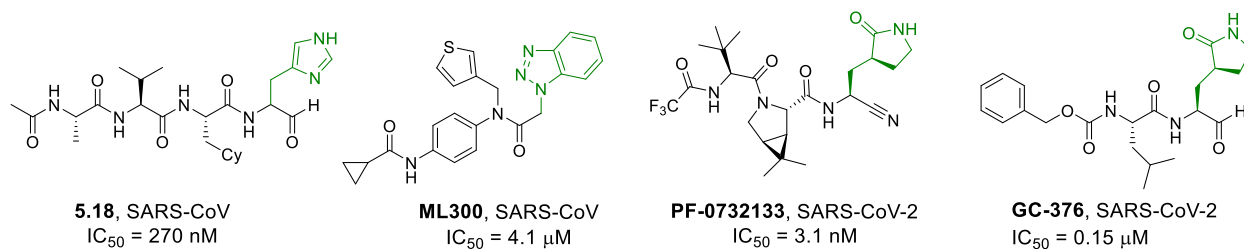
Figure 5.6. Select 3CL[pro] inhibitors, with S1 binding groups shown in green.

On the basis of these reports, it was envisioned that alternative R^3 groups in our scaffold could improve upon the binding interactions within the S2 site. While maintaining a hydrogen bond interaction with His[163] is critical, this could theoretically be accomplished with different groups whose steric or electrophilic properties may improve binding affinity. It was also envisioned that introduction of a hydrogen bond-donating atom in the R^3 group could facilitate interactions with Glu[166] and Phe[140] as is observed with the lactam group of previously reported 3CL[pro] inhibitors. To support our design of R^3 analogues, docking studies were performed. Promisingly, many analogues had predicted poses where the interaction with His[163] was maintained. Additionally, introduction of a hydrogen bond donor via pyrazole and hydantoin analogues also showed promising docking poses and scores, as shown in **Figure 5.7**.
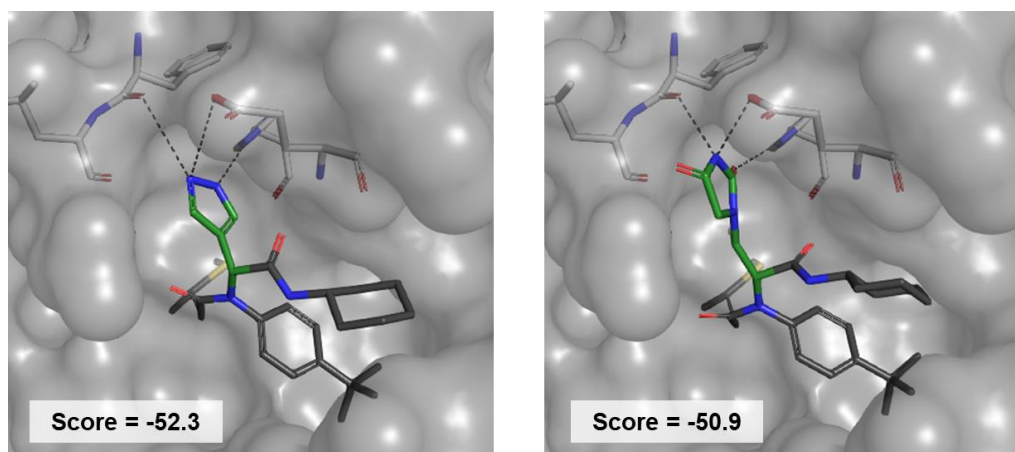
Figure 5.7. Docking poses of select aldehyde analogues, with the aldehyde substituent shown in

green and its corresponding docking score (lower = better). Left: pyrazole analogue **5.24**. Right: hydantoin analogue **5.23**.

5.3.2.2 Synthesis

Based on the promising docking results, a library of analogues featuring different aldehyde components (R^3) was synthesized via a 4-component Ugi (**Scheme 5.6**), using the same alkynyl warhead from **4.10**.
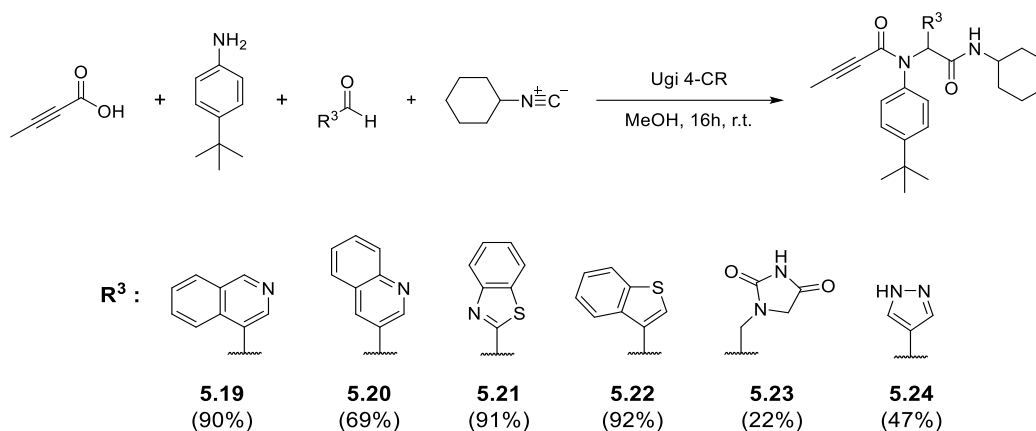
Scheme 5.6. Synthesis of R^3 analogues.

In the case of analogue **5.23**, the aldehyde component was synthesized by ozonolysis of 1-allylhydantoin (**Scheme 5.7**).
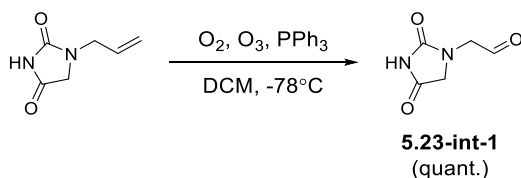
Scheme 5.7. Synthesis of aldehyde starting material for analogue **5.23**.

As a note, while many 3CL^{pro} inhibitors employ a lactam group as a glutamine mimic to occupy the S1 site, we chose not to pursue this group due to considerations of synthetic feasibility. Our inhibitors are synthesized via an Ugi reaction, and therefore are obtained as a racemic mixture of enantiomers. The use of a chiral aldehyde such as a lactam would most likely generate four
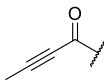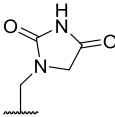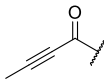
diasteromeric products in a roughly even ratio due to the poor stereocontrol typically observed with Ugi reactions.[37, 38] This would not only result in challenges in purification and characterization of the reaction products, but presumably only one of the four diastereomers would be biologically active. In light of this, we felt that the achiral hydantoin group could sufficiently act as a lactam-like glutamine mimic to serve as a proof-of-principle prior to any further exploration of the aldehyde component that may require greater synthetic efforts.

5.3.2.3 *In Vitro* Biological Evaluation

Table 5.2. Biological evaluation of R^3 analogues.
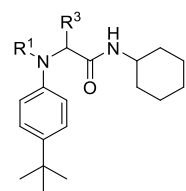
Entry	Compound	R^1	R^3	Inhibition (%)[a]	IC_{50} (μM)
1	**4.10**			> 95	5.3 ± 0.8
2	**5.19**			> 95	5.0 ± 2.3
3	**5.20**			80 ± 10	nd[b]
4	**5.21**			54 ± 15	nd[b]
5	**5.22**			57 ± 6	nd[b]

6	**5.23**		10 ± 10	nd[b]
7	**5.24**		19 ± 5	nd[b]

[a] The enzyme activity was measured with 114 nM 3CL^{pro} and 50 µM of each potential inhibitor with incubation time of 30 min. [b] not determined.

From the crystal structures of **4.13** and **4.19**, we observed that the pyridine group at R^3 forms a key hydrogen bonding interaction with His^{163}. With this in mind, attempts were made to modify the R^3 group using different heterocyclic groups that could also maintain this interaction, however all modifications have proven unsuccessful to date. More specifically, the use of a variety of aromatic heterocycles such as benzothiazole, and benzothiophene all led to loss of potency. Additionally, nearly identical activity was observed with the isoquinoline analogue **5.19** (IC$_{50}$ = 5.0 µM) compared to the initial pyridine analogue **4.10** (IC$_{50}$ = 5.3 µM). Disappointingly, our attempt to resemble the hydrogen bonding pattern of glutamine more closely by using a pyrazole (**5.24**) or hydantoin (**5.23**) moiety led to a significant loss of potency, as well.

5.3.3 Varying R^4 Component

5.3.3.1 Design

As observed in the crystal structure in **Figure 5.2**, our inhibitor scaffolds bind the S4 subsite via the cyclohexyl moiety introduced from the isocyanide starting material. Compared to the S1 and S2 subsites, the S4 subsite has been shown to accommodate a wider range of residues such as alanine, valine, phenylalanine and threonine.[11] The R^4 group of C63 partially occupies this hydrophobic pocket, primarily interacting with Pro^{168} and Gln^{189}. From the crystal structure, it appears that elongating the R^4 group could enable it to bind deeper into the S4 subsite to form additional favourable interactions with Leu^{167}, Thr^{190} and Gln^{192} as is observed in the binding mode of nirmatrelvir (PDB: 7RFW).[17] A similar approach of P4 elongation was successfully applied with **GC-376** for the related MERS-CoV 3CL^{pro} resulting in a nearly 2-fold increase in potency (**Figure 5.8**).[39]

GC-376, MERS-CoV
EC$_{50}$ = 0.9 µM

GC-813, MERS-CoV
EC$_{50}$ = 0.5 µM
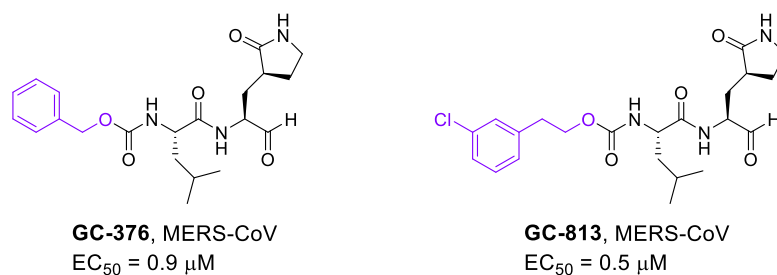
Figure 5.8. P4 elongation of MERS-CoV 3CL^{pro} inhibitor **GC-376** to **GC-813**.

Our design hypothesis was supported by docking a small library of analogues of **4.19** to 3CL^{pro}, with several displaying improved binding to the S4 subsite. Promising docking results were obtained for analogues containing a *para*-methoxy substituted homobenzyl (**5.28**) or *N*-benzyl piperidine group (**5.30**), which demonstrated additional hydrogen bond interactions with Thr^{190} and Glu^{166}, respectively (**Figure 5.9**).
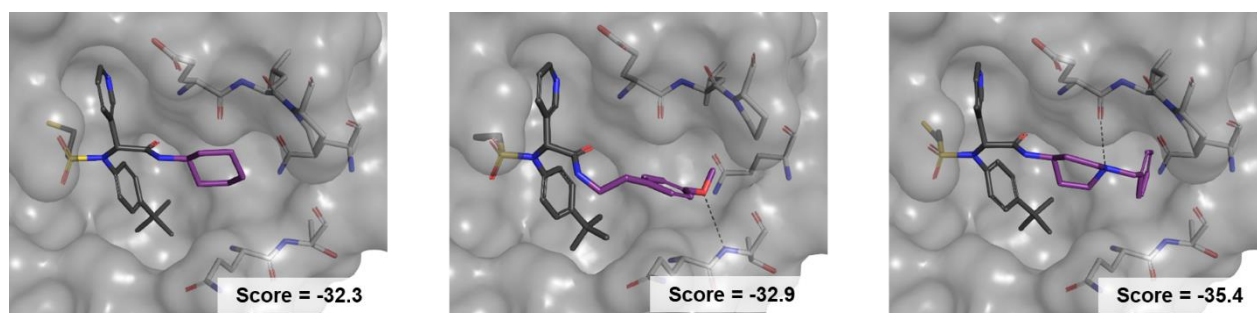
Figure 5.9. Predicted docking poses of **4.19** analogue. Left: **4.19**. Middle: para-methoxy homobenzyl derivative **5.28**. Right: *N*-benzyl piperidine derivative **5.30**.

5.3.3.2 Synthesis

A small library of analogues featuring different R^4 components was synthesized via a 4-component Ugi, using the same alkynyl warhead from **4.10** (**Scheme 5.8**).
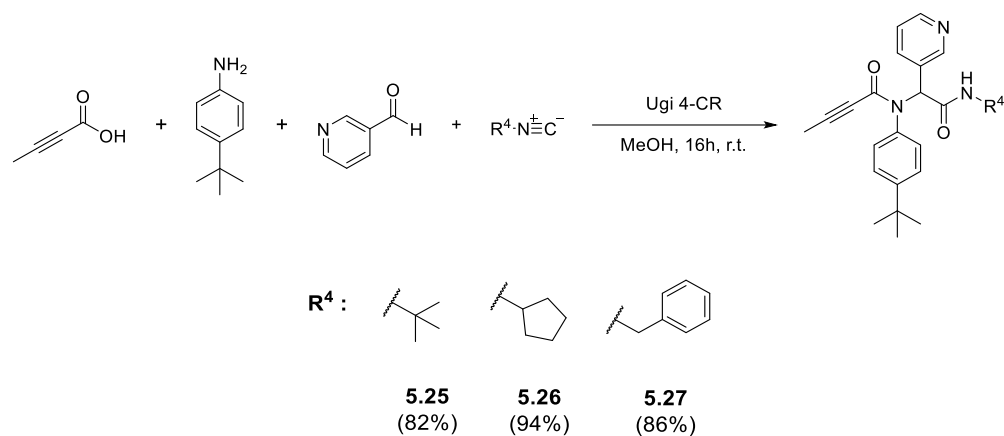
Scheme 5.8. Synthesis of R^4 analogues via 4-component Ugi.

The R^4 group was also further explored with analogues featuring the vinyl sulfonamide warhead of inhibitor **4.19**. In this case, analogues were synthesized in two steps, starting with a 3-component Ugi to access an amine intermediate. The warhead was then installed using the same conditions as with inhibitor **4.19** (**Scheme 5.9**).
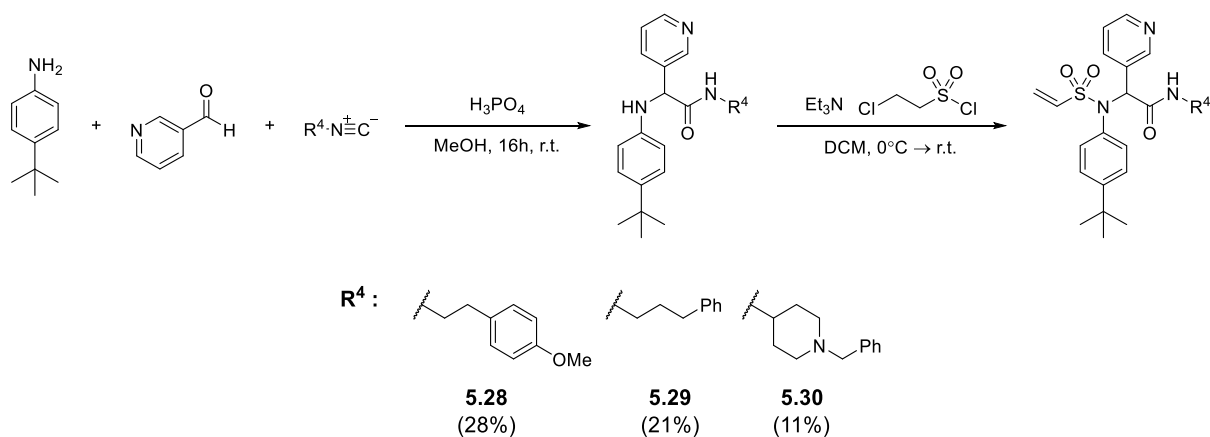
Scheme 5.9. Synthesis of R^4 analogues via 3-component Ugi, with yields over two steps.

5.3.3.3 *In Vitro* Biological Evaluation

Table 5.3. Biological evaluation of R^4 analogues.
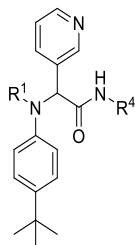
Entry	Compound	R^1	R^4	Inhibition (%)[a]	IC_{50} (µM)
1	**5.25**			93	15.0 ± 9.3
2	**5.26**			88	9.7 ± 3.8
3	**5.27**			69	> 30
4	**5.28**			> 95	0.28 ± 0.10
7	**5.29**			> 95	0.52 ± 0.16
8	**5.30**			> 95	0.22 ± 0.08

[a] The enzyme activity was measured with 114 nM 3CL^{pro} and 50 µM of each potential inhibitor with incubation time of 30 min.

Replacement of the cyclohexyl group with similar aliphatic groups such as *tert*-butyl (**5.25**), cyclopentyl (**5.26**) or benzyl (**5.27**) did not improve the potency. However, we found that the use of longer R^4 groups led to improved potency, as we predicted with docking. However, there appears to be a limit to the length of R^4 group that can be tolerated in this pocket, as a slight decrease in potency was observed with the longer propylphenyl analogue **5.29**. We also observed an improvement in potency with benzyl-substituted piperidine analogue **5.30**. Docking suggested

that piperidine, which would be protonated under physiological conditions, would be able to form a hydrogen-bonding interaction with Glu[166]. This interaction may explain the improved potency of this analogue.

5.3.4 Heterocyclic Analogue

5.3.4.1 Design

Analysis of the crystal structure of vinyl sulphonamide analogue **4.19** indicates that the cyclohexyl amide group interacts with the S3 subsite of 3CL[pro] through its carbonyl oxygen, with the amide nitrogen pointing out towards the solvent (**Figure 5.2**). The amide functional group acts only as a hydrogen bond acceptor, forming a key hydrogen bond with the backbone of Glu[166], however no interactions are mediated by the amide nitrogen. This suggested the possibility of replacing amide group with a heterocyclic bioisostere that could still maintain the hydrogen bond interaction with Glu[166].

As originally defined by Friedman, bioisosteres are atoms, functional groups or molecules that demonstrate similar biological activity.[40] Bioisosteres have found widespread use in medicinal chemistry as they can be used to maintain key ligand binding interactions while optimizing the pharmacokinetic properties of an inhibitor. Several amide bioisosteres have been devised and applied to address the pharmacokinetic limitations often observed with the amide functional group, such as their susceptibility to enzymatic hydrolysis, transporter-mediated efflux, and poor cell permeability.[41] With regards to 3CL[pro] inhibitors, it was recently shown that the low bioavailability of peptidomimetic inhibitor **YH-30** was due to amide bond hydrolysis, suggesting that the numerous amide bonds of current peptidomimetic 3CL[pro] inhibitors may be a liability for effective *in vivo* activity.[42] With this in mind, we wanted to explore the possibility of replacing the cyclohexyl amide group of our inhibitor scaffold with a heterocyclic bioisostere.

To validate this design strategy, a heterocyclic analogue of **4.19**, which featured a 1,3,4-oxadiazole in place of the amide functional group was docked to 3CL[pro] (PDB: 7MLG). Promisingly, docking suggested that the oxadiazole was able to form the same hydrogen bonding interaction with Glu[166] while maintaining the same binding mode in the other enzyme subsites (**Figure 5.10**).
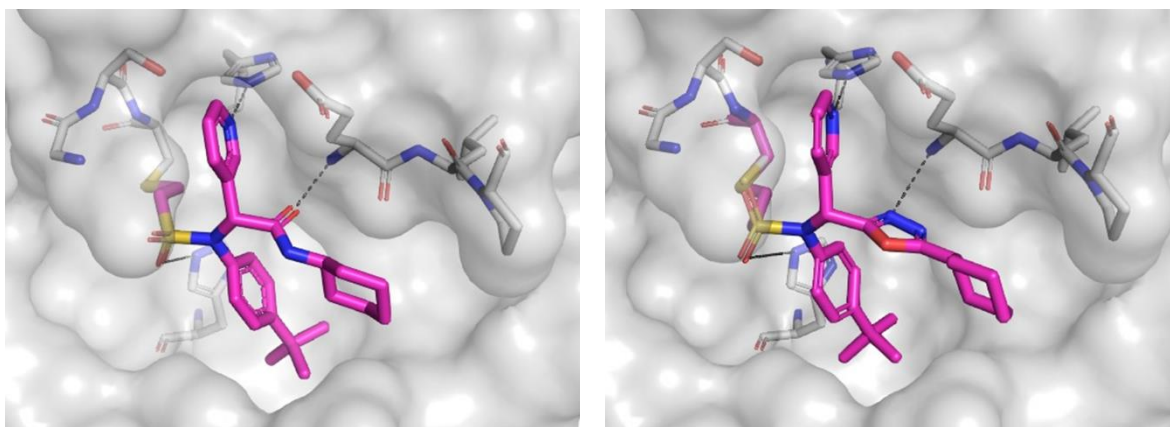
Figure 5.10. Left: co-crystal structure of **4.19** bound to 3CL[pro]. Right: predicted binding pose of a heterocyclic analogue of **4.19**.

5.3.4.2   Synthesis

Based on the promising docking results, efforts were made to synthesize an oxadiazole analogue. A literature procedure from Ramazani *et al.* suggested that the oxadiazole could be accessible via an Ugi-*aza*-Wittig reaction using *N*-(isocyanoamino)triphenylphosphorane.[43] However, running the reaction with the corresponding starting materials resulted in the formation of both the desired product and an acyl dihydrazide byproduct in low yields (**Scheme 5.10**). Further complicating matters, the desired product was inseparable from the triphenylphosphine oxide byproduct that is formed during the course of the reaction.
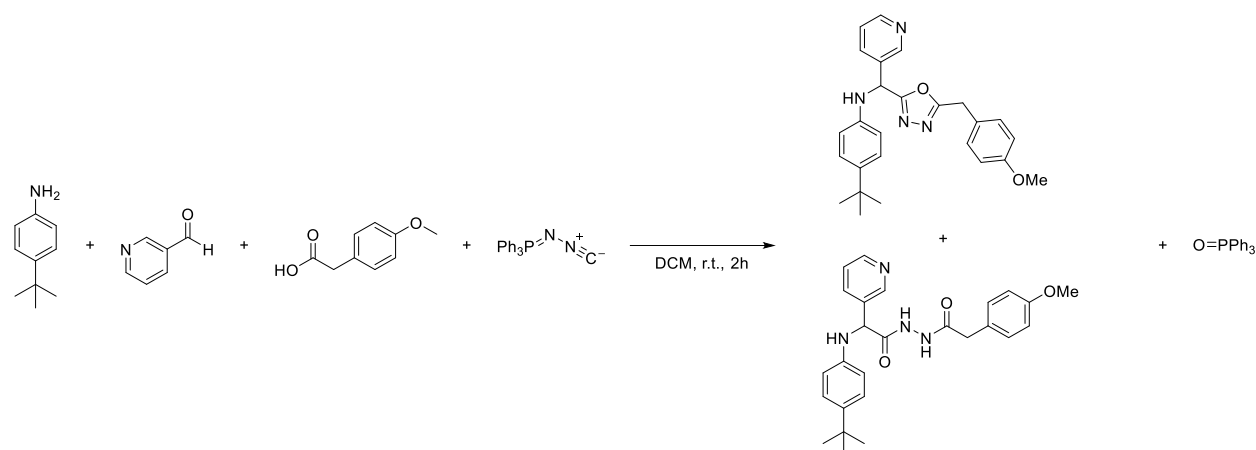
Scheme 5.10. Attempted synthesis of oxadiazole intermediate via an Ugi-*aza*-Wittig reaction.

It is important to note that the original publication utilized a secondary aliphatic amine, not a primary aromatic amine as required to access our inhibitor scaffold. A secondary amine could assist in product formation as the Mumm rearrangement is not possible with secondary amines and therefore could limit the potential for non-productive side reactions. Additionally, secondary amines form positively charged iminiums upon condensation with aldehydes, compared to the neutral imines formed with primary amines. It is possible that this could also have an effect on product formation, as the imine formed with our substrates would likely need to be protonated by the acid to be activated towards attack by the iminophosphorane. This protonation step could also be influenced by the presence of other basic nitrogens (i.e. pyridine) that could compete with the imine for protonation.

Given the challenges of accessing the oxadiazole from a multicomponent reaction, we instead chose to pursue a more linear synthesis. The hydrazide intermediate **5.31-int-2** was prepared in two steps starting with a modified 3-component Ugi, via a Lewis-acid catalysed condensation in methanol to form the intermediate methyl ester **5.31-int-1**.[44] This was then followed by a substitution reaction to afford the desired hydrazide **5.31-int-2**. Given our previous results in optimizing the R^4 position, a 4-methoxybenzyl group was installed in place of the original cyclohexyl moiety via an amide coupling reaction with CDI to form **5.31-int-3**. The acylhydrazide was subsequently cyclized to the desired oxadiazole **5.31-int-4** via an Appel-type reaction with polystyrene-supported triphenylphosphine and I_2.
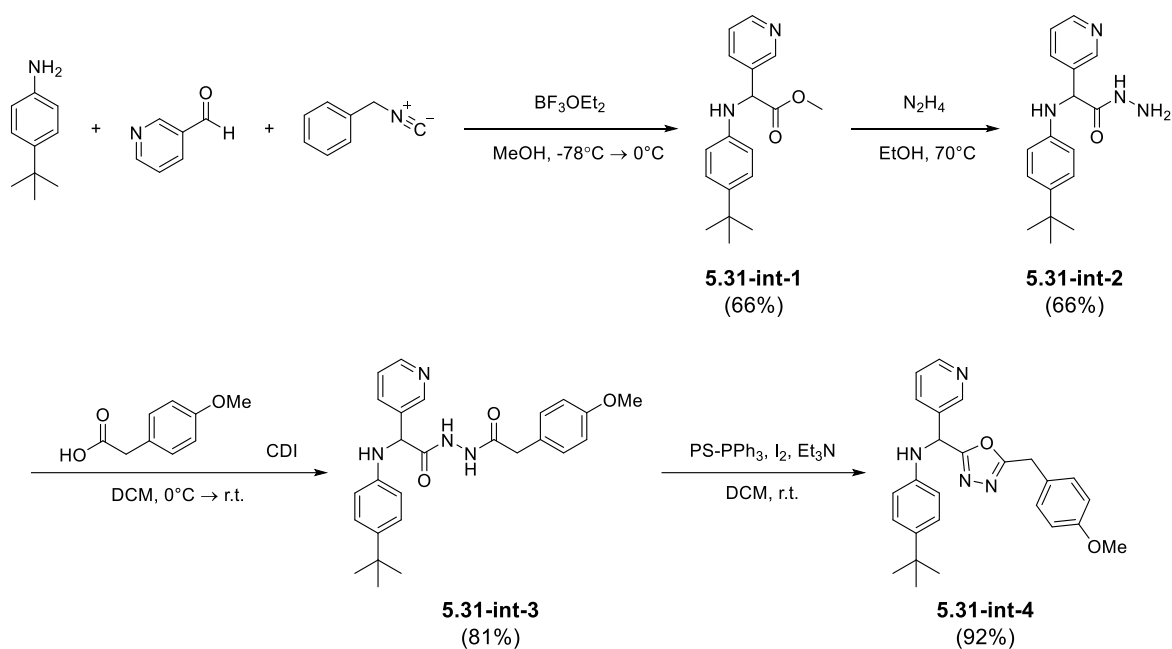
Scheme 5.11. Synthesis towards oxadiazole analogue **5.31**. Yields of intermediates given in brackets.

Although the linear synthesis involves more steps, analogues are accessible through the hydrazide intermediate **5.31-int-2** and therefore only need to be carried through two transformations to access the oxadiazole scaffold **5.31-4**. As shown in **Table 5.3**, we obtained our most potent analogues through the optimization of the R^4 group, with modifications at this position corresponding to the isocyanide starting material. However, there are a limited number of commercially available isocyanides and their synthesis can be challenging due to the volatility of intermediates and functional group compatibility issues.[45] In the synthesis of the oxadiazole analogue, modifications to the R^4 group correspond to the carboxylic acid starting material in forming the acylhydrazide intermediate **5.31-int-3**. This may be advantageous and enable further exploration of the R^4 group with substituents not accessible through an isocyanide reagent.

Initial attempts at incorporating the vinyl sulfonamide under the previously developed conditions were unsuccessful (as were further attempts under different reaction conditions). For ease of synthetic accessibility, the non-covalent furan group was instead installed at this position using commercially available 2-furoyl chloride. Although this analogue does not feature the covalent vinyl sulfonamide warhead of **4.19**, it allows us to more easily assess the effect of the oxadiazole

bioisostere on the non-covalent binding affinity.
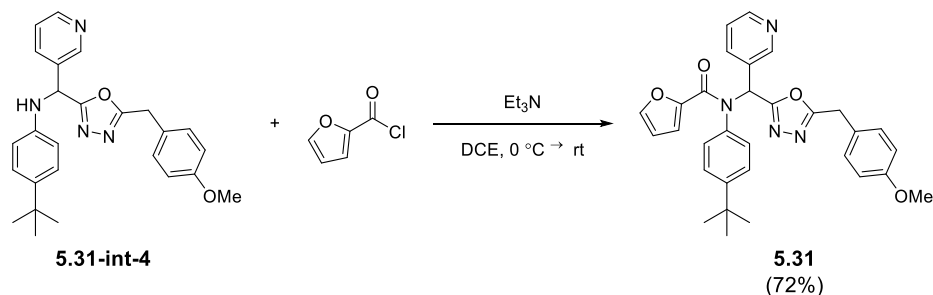
Scheme 5.12. Incorporation of furan group at the R^1 position to access **5.31**.

Additionally, the corresponding amide analogue of compound **5.32** was synthesized in order to directly evaluate the effect of the oxadiazole group on the inhibitor binding affinity. The analogue was synthesized via a traditional 4-component Ugi (**Scheme 5.13**)
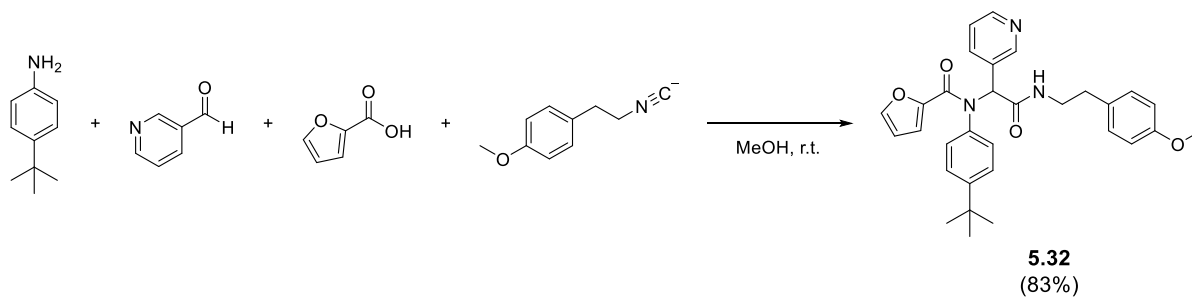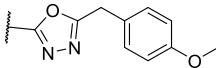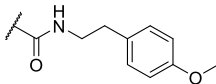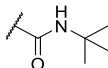
Scheme 5.13. Synthesis of amide analogue **5.32**.

5.3.4.3 *In Vitro* **Biological Evaluation**

Table 5.4. Biological evaluation of heterocylic analogue.
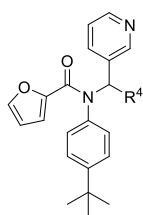
Entry	Compound	R^4	Inhibition (%)^a	IC$_{50}$ (µM)

1	**5.31**		nd[b]	7.7 ± 1.1
2	**5.32**		nd[b]	0.91 ± 0.32
3	**ML-188**		> 95	1.4 ± 0.4

[a] The enzyme activity was measured with 114 nM 3CL^{pro} and 50 µM of each potential inhibitor with incubation time of 30 min. [b] not determined.

The oxadiazole analogue **5.31** and the corresponding amide analogue **5.32** were also screened against 3CL^{pro}. It is important to note that both **5.31** and **5.32** are non-covalent inhibitors, featuring a furan ring in the R^1 position in place of a covalent warhead. Given that our covalent inhibitors are screened with an incubation time of 30 minutes and have displayed time-dependent potency typical for covalent inhibition, it is expected that these non-covalent analogues may have weaker IC$_{50}$ values. However, the IC$_{50}$ values will be more representative of binding affinity and can more accurately be compared to the K_i values determined from kinetic studies of our covalent inhibitors.

The amide analogue **5.32** showed sub-micromolar activity, and when comparing to non-covalent inhibitor **ML-188**, these results confirm that longer R^4 groups result in improved binding affinity. Although the heterocyclic analogue **5.31** displays a weaker potency compared to the amide analogue **5.32**, it still displays low micromolar binding affinity. It serves as a promising proof-of-principle that the amide bond in the original Ugi series could feasibly be replaced by a heterocyclic bioisostere without having a significant impact on activity.

5.4 *In Cellulo* Activity

Based on the promising *in vitro* results, 10 compounds were selected for *in cellulo* testing to assess their ability to inhibit viral replication. Compounds were tested against the SARS-CoV-2 Vido strain in Huh7.5 cells, assessing their effect on viral replication and cell viability at two concentrations (20 µM and 10 µM). The results are shown in **Figure 5.11**, with the viral replication shown in dark blue and red, and cell viability in light blue and orange. Ideally, inhibitors should demonstrate significant inhibition of viral replication while not affecting cell viability.
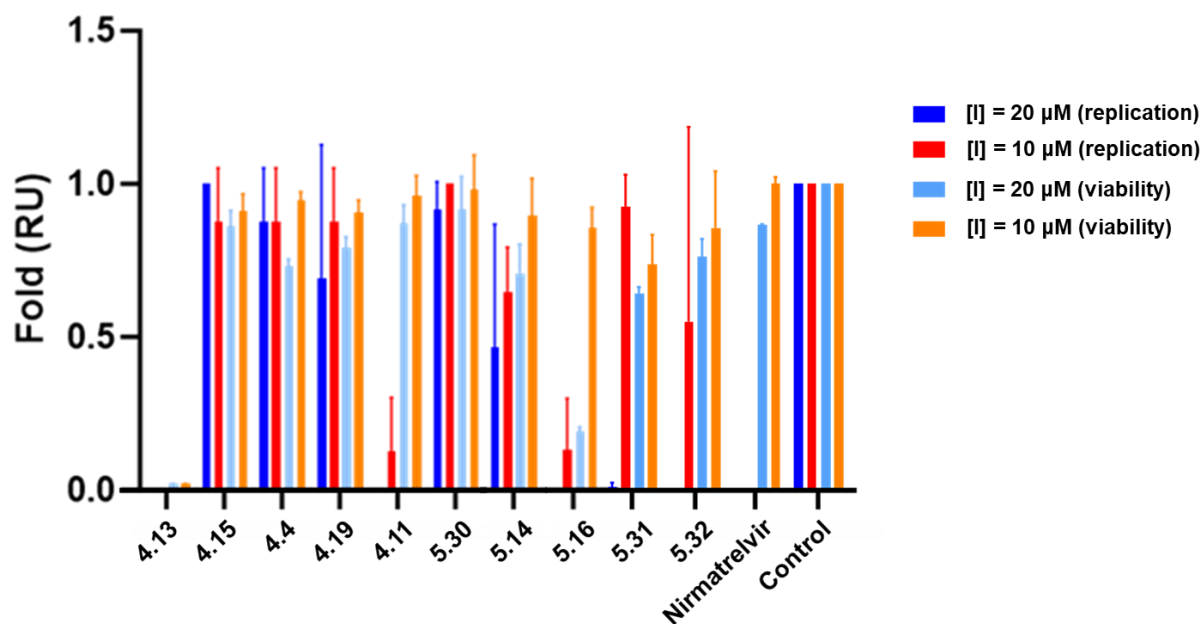
Figure 5.11. *In cellulo* evaluation of inhibitor effects on viral replication and cell viability.

The results in **Figure 5.11** show that 5 compounds (**4.13**, **4.11**, **5.16**, **5.31** and **5.32**) display complete viral inhibition at 20 μM. While inhibitor **4.13** also displayed complete viral inhibition at the lower 10 μM concentration, it also resulted in nearly complete inhibition of cell viability, indicating significant cytotoxic effects for this inhibitor. This is likely a result of the high reactivity and lower selectivity associated with the chloroamide warhead. Inhibitors featuring the hydroxyalkyne warhead (**4.11**, **5.16**) both showed significant viral inhibition at both 20 μM and 10 μM, however analogue **5.16** also displayed significant cytotoxicity at the higher concentration. However, **4.11** showed almost complete cell viability at both concentrations tested while simultaneously displaying significant viral inhibition. This suggests that while the replacement of the *tert*-butyl phenyl R^2 group with a diphenyl group afforded a more potent inhibitor *in vitro*, this resulted in a worse selectivity profile *in cellulo*. Somewhat surprisingly, neither of the inhibitors containing a vinyl sulfonamide warhead (**4.19**, **5.30**) showed *in cellulo* activity at either concentration tested despite being our most promising compounds *in vitro*. This may be due to their increased susceptibility to nucleophilic attack, perhaps resulting in their covalent modification prior to reaching 3CL^{pro}. The two non-covalent analogues **5.31** and **5.32** both showed complete viral inhibition at 20 μM but very little inhibition at 10 μM. Promisingly, the oxadiazole

analogue **5.31** displayed nearly identical *in cellulo* inhibition as the amide analogue **5.32**, suggesting that despite the difference in *in vitro* potency, bioisosteric replacement of the amide group by a heterocycle is a viable strategy for the scaffold.

The non-covalent analogue **4.4** was also tested to assess the impact of the core scaffold on cell viability, with minor cytotoxicity observed at 20 µM even in the absence of a covalent warhead. This suggests that modifications to the R^2-R^4 groups are likely necessary to reduce cytotoxicity at higher concentrations and/or improve their potency at lower concentrations to improve the selectivity index of these inhibitors.

5.5 Conclusion and Future Work

Our initial covalent scaffolds were further optimized, focusing on improving the non-covalent interactions with the S1, S2, and S4 binding pockets. While no improvements in potency were observed with modifications to the S1 binding group, changes to both the S2 and S4 binding groups resulted in improved potency. Notably, introduction of a diphenyl group at the P2 position (**5.16**) and elongation of the P4 group to bind deeper into the S4 subsite with a *para*-methoxy homobenzyl group (**5.28**) or an *N*-benzyl piperidine group (**5.30**) led to a nearly two-fold improvement in potency. Promising results were also obtained with heterocyclic analogues **5.31** which displayed low micromolar activity *in vitro*. Importantly, this represents one of the first 3CL^pro inhibitors to introduce a heterocyclic group in the P3 position, and the observed *in vitro* and *in cellulo* activity suggests that it could be a viable strategy for the design of future 3CL^pro inhibitors. Several of our inhibitors were tested for their *in cellulo* activity, with four compounds (**4.11**, **5.14**, **5.16**, and **5.32**) displaying promising antiviral activity (**Figure 5.12**).
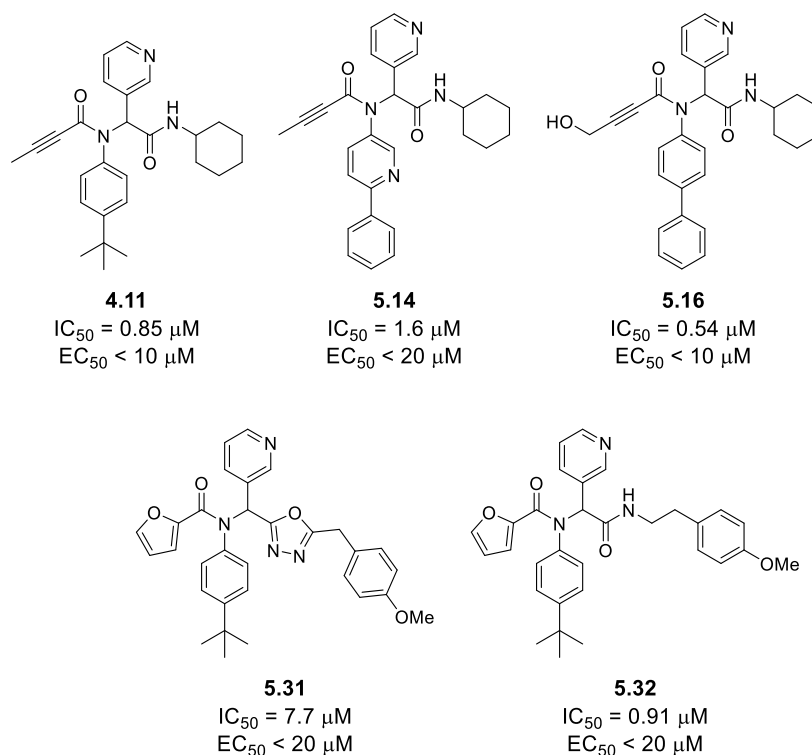
Figure 5.12. Our 3CL^{pro} inhibitors that displayed both *in vitro* and *in cellulo* activity.

Future work will involve combining our most promising optimizations into a single analogue. For example, the most promising *in cellulo* results were observed with **5.16** which contains the hydroxyalkyne warhead and optimized P2 diphenyl group. Our most promising *in vitro* results were obtained with P4 analogues **5.28** and **5.30**, however these were not active *in cellulo*, perhaps as a result of their vinyl sulfonamide warhead. Combining the best aspects of **5.16** and **5.28/5.30** may be a promising strategy to further optimize this scaffold and to assess the effect of the P4 group on *in cellulo* activity. Considering the promising *in vitro* and *in cellulo* results, work is currently underway to access a covalent analogue of **5.31**. Although the current *in cellulo* results were only obtained at two inhibitor concentrations, work is presently underway to investigate the dose dependent activity of our most promising inhibitors in order to obtain EC$_{50}$ values.

5.6 Experimental

5.6.1 General

Unless otherwise specified, all solvents and reagents were purchased from commercial suppliers and used without further purification. All ^{1}H, ^{13}C and ^{19}F NMR spectra were acquired on a Bruker

400 or 500 MHz spectrometer, or a Varian 500 MHz spectrometer. Chemical shifts are reported in ppm using the residual of deuterated solvents as an internal standard. Chromatography was performed on silica gel 60 (230-40 mesh) or using the Biotage One Isolera with ZIP cartridges. High resolution mass spectrometry was performed by ESI on a Bruker Maxis Impact API QqTOF mass spectrometer at McGill University. Reversed-phase HPLC (water and MeCN or MeOH gradient) was used to verify the purity of compounds on an Agilent 1100 series instrument equipped with VWD-detector, C18 reverse column (Agilent, Zorbax Eclipse XDBC18 150 mm 4.6 mm, 5 μm), and UV detection at 254 nm. Measured purities for all tested compounds are listed in Table SX in the supporting information.

General Procedures A, B and C for 4-Component Ugi Reaction. In a 6-dram vial equipped with a stir bar aldehyde (1.0 mmol, 1.0 eq.), aniline (1.0 mmol, 1.0 eq.) and carboxylic acid (1.0 mmol, 1.0 eq.) were combined in MeOH (4 mL). The obtained reaction mixture was stirred for 30 min. at room temperature. Afterwards cyclohexyl isocyanide (0.9 mmol, 0.9 eq.) was added to the reaction mixture and the walls of the vial were washed with 1 mL of MeOH. The reaction mixture was continued to stir at room temperature overnight. The crude reaction mixture was evaporated in vacuo. Purification procedure A) The crude product was triturated with hexanes (5 mL) and filtered. The obtained product was further washed with hexanes (3 x 3 mL). Purification procedure B) The crude product recrystallized from $CHCl_3$/hexanes mixture, filtered and the obtained product was further washed with hexanes (3 x 3 mL). Purification procedure C) The crude product was redissolved in DCM. The obtained crude solution was deposited on silica. It was then purified using flash column chromatography using DCM/MeOH (gradient $0 \rightarrow 5\%$) as eluent.

General procedure D for Synthesis of Vinyl Sulfonamide Intermediates. In a 6-dram vial equipped with a stir bar, 1 mmol (1.0 eq.) of 4-*tert*-butylaniline and 1 mmol (1.0 eq.) of 3-pyridinecarboxaldehyde were dissolved in 5 mL MeOH with 10 μL of 85% H_3PO_4. 1 mmol (1.0 eq.) of isocyanide was then added. The mixture was stirred at room temp. overnight. The reaction mixture was then concentrated in vacuo. The crude product was purified using flash column chromatography using EtOAc/hexanes (33% $\rightarrow$ 80%) as eluent.

General procedure E for Synthesis of Vinyl Sulfonamides. In a 6-dram vial equipped with a stir bar, 0.2–0.3 mmol (1.0 eq.) of the previously made acetamide (see General procedure D) was

dissolved in 5 mL of DCM with 0.7–3. 5 mmol (from 0.5 eq. to 3 eq.) of Et₃N. The mixture was cooled to 0 °C and 0.3–0.45 mmol (1.5 eq.) of 2-chloroethanesulfonyl chloride was added dropwise. The solution was stirred at 0 °C for 2 h. The solution was then diluted with 5 mL DCM and washed with 10 mL sat. NaHCO₃. The aqueous layer was extracted with 10 mL DCM and the combined organic layer was washed with 10 mL sat. NaCl solution and further dried with anhydrous Na₂SO₄. The crude product was purified using flash column chromatography using DCM/EtOAc (gradient 0% → 50%) as eluent.

5.6.2 Synthesis and Characterization Data

N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-N-(4-methoxyphenyl)but-2-ynamide (5.6). Compound was made and purified using general procedure A, grey solid 83% yield, 302 mg. ^{1}H NMR (500 MHz, CDCl₃) δ 8.50 (dd, J = 4.8, 1.7 Hz, 1H), 8.45 (d, J = 2.3 Hz, 1H), 7.48 (d, J = 8.0 Hz, 1H), 7.12 (ddd, J = 8.0, 4.9, 0.9 Hz, 1H), 6.98 (s, 2H), 6.79–6.70 (m, 2H), 6.10 (s, 1H), 6.00 (d, J = 8.0 Hz, 1H), 3.88–3.81 (m, 1H), 3.79 (s, 3H), 1.94–1.84 (m, 2H), 1.74 (s, 4H), 1.61 (s, 1H), 1.42–1.11 (m, 6H). ^{13}C NMR (126 MHz, CDCl₃) δ 167.75, 159.81, 155.97, 151.57, 149.93, 138.74, 132.10, 130.52, 123.44, 114.20, 92.59, 61.98, 55.74, 49.28, 33.21, 33.14, 25.85, 25.10, 4.39. HRMS (ESI/Q-TOF) m/z: [M + H]$^+$ calculated for C₂₄H₂₈N₃O₃ 406.21252; found 406.21250.

N-(3-acetamidophenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamide (5.7). Compound was made and purified using general procedure A, white solid 82% yield, 320 mg. ^{1}H NMR (500 MHz, CDCl₃) δ 8.42 (s, 2H), 7.96 (s, 1H), 7.83–7.76 (m, 1H), 7.49 (dt, J = 7.9, 2.1 Hz, 2H), 7.17–7.04 (m, 2H), 6.74 (d, J = 7.8 Hz, 1H), 6.34 (d, J = 8.1 Hz, 1H), 6.03 (s, 1H), 3.78–3.68 (m, 1H), 2.13 (s, 3H), 1.92 (d, J = 12.2 Hz, 1H), 1.82 (s, 1H), 1.69 (s, 3H), 1.63–1.51 (m, 2H), 1.33–1.22 (m, 2H), 1.11 (dtd, J = 35.2, 11.2, 3.6 Hz, 3H). ^{13}C NMR (126 MHz, CDCl₃) δ 168.57, 167.51, 155.06, 151.09, 149.64, 139.27, 138.81, 137.84, 130.10, 129.22, 125.98, 123.24, 121.38, 119.91, 92.30, 77.29, 77.04, 76.79, 73.69, 62.29, 49.04, 32.69, 25.40, 24.76, 24.71, 24.54, 3.99. HRMS (ESI/Q-TOF) m/z: [M + H]$^+$ calculated for C₂₅H₂₉N₄O₃ 433.22342; found 433.22345.

N-(4-acetamidophenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamide (5.8). Compound was made and purified using general procedure A, light brown solid 89% yield,

345 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.48–8.44 (m, 1H), 8.41 (s, 1H), 7.50 (dt, J = 7.6, 1.9 Hz, 1H), 7.43 (d, J = 8.4 Hz, 2H), 7.14 (dd, J = 8.0, 4.8 Hz, 1H), 7.03 (d, J = 7.5 Hz, 2H), 6.24 (s, 1H), 6.09 (s, 1H), 3.81 (dp, J = 11.0, 4.0, 3.6 Hz, 1H), 2.15–2.12 (m, 3H), 2.01–1.91 (m, 1H), 1.89–1.82 (m, 1H), 1.78–1.64 (m, 4H), 1.64–1.52 (m, 1H), 1.45–1.05 (m, 6H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.74, 167.40, 155.36, 151.06, 149.50, 138.70, 138.23, 134.05, 131.10, 130.16, 123.29, 119.30, 92.50, 77.30, 77.05, 76.79, 73.67, 61.88, 49.01, 32.77, 32.71, 25.42, 24.75, 24.70, 24.57, 4.00. HRMS (ESI/Q-TOF) m/z: [M + H]$^+$ calculated for C$_{25}$H$_{29}$N$_4$O$_3$ 433.22342; found 433.22352.

N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-N-(4-fluorophenyl)but-2-ynamide (5.9). Compound was made and purified using general procedure A, light brown solid 88% yield, 311 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.49 (dd, J = 4.8, 1.6 Hz, 1H), 8.44 (d, J = 2.3 Hz, 1H), 7.46 (dt, J = 8.0, 2.0 Hz, 1H), 7.11 (dt, J = 10.7, 5.3 Hz, 2H), 6.91 (t, J = 8.6 Hz, 2H), 6.10 (s, 1H), 6.02 (s, 1H), 3.82 (dtd, J = 10.8, 7.3, 4.0 Hz, 1H), 2.04–1.93 (m, 1H), 1.89–1.82 (m, 1H), 1.72 (s, 3H), 1.67–1.56 (m, 2H), 1.45–1.29 (m, 2H), 1.27–1.04 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.24, 163.25, 161.27, 155.15, 151.13, 149.73, 138.05, 134.72, 134.69, 132.76, 132.69, 130.05, 123.21, 115.69, 115.51, 92.45, 73.58, 61.42, 49.01, 32.78, 32.72, 25.43, 24.77, 24.71, 3.89.

N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-N-(4-(difluoromethoxy)phenyl)but-2-ynamide (5.10). Compound was made and purified using general procedure A, light brown solid 89% yield, 355 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.51 (dd, J = 4.8, 1.6 Hz, 1H), 8.46 (d, J = 2.3 Hz, 1H), 7.50 (dt, J = 8.0, 2.0 Hz, 1H), 7.20–7.10 (m, 2H), 7.03–6.93 (m, 2H), 6.50 (t, J = 73.4 Hz, 1H), 6.10 (s, 1H), 5.93 (d, J = 8.1 Hz, 1H), 3.83 (dtd, J = 10.8, 7.7, 7.3, 4.0 Hz, 1H), 1.98 (s, 1H), 1.87 (dd, J = 12.1, 4.0 Hz, 1H), 1.76–1.71 (m, 4H), 1.70–1.57 (m, 2H), 1.45–1.28 (m, 2H), 1.24–1.03 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.53, 155.46, 151.37, 150.03, 138.52, 132.78, 130.44, 123.65, 119.71, 115.88, 92.99, 73.95, 61.89, 49.41, 33.17, 33.11, 25.81, 25.14, 25.09, 4.30, 4.16. ^{19}F NMR (471 MHz, CDCl$_3$) δ −81.48 (d, J = 73.5 Hz). HRMS (ESI/Q-TOF) m/z: [M + H]$^+$ calculated for C$_{24}$H$_{26}$N$_3$O$_3$F$_2$ 442.19367; found 442.19361.

N-benzyl-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)propiolamide (5.11). Pyridine-3-carboxaldehyde (0.09 mL, 0.93 mmol) and benzylamine (0.10 mL, 0.93 mmol) were stirred in MeOH (3.7 mL) and for 30 min at rt. 2-Butynoic acid (78.5 mg, 0.93 mmol) and cyclohexyl

isocyanide (0.10 mL, 0.84 mmol) were added and the solution stirred at room temp. overnight. The reaction was concentrated in vacuo and the crude product was dissolved in EtOAc washed with water and brine. The resulting aqueous phases were extracted three times for residual compound. The combined organic phases were washed with brine and dried over Na_2SO_4. The crude product was purified using column chromatography with a gradient of mixture of (MeOH + 1% NH_4OH) and DCM (0% → 5%). The reaction was further purified using column chromatography with a gradient of EtOAc in hexanes (0% → 100%). White powder (114 mg, 35% yield). Rf (5% (MeOH + 1% NH_4OH) in DCM) = 0.39. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.57–8.49 (m, 0.4H), 8.43 (t, J = 2.9 Hz, 1.6H), 7.75 (dt, J = 8.1, 1.9 Hz, 0.8H), 7.64 (d, J = 7.9 Hz, 0.2H), 7.25–7.16 (m, 3H), 7.11 (dd, J = 7.4, 2.0 Hz, 1.4H), 7.03 (dd, J = 6.7, 2.8 Hz, 0.3H), 6.14 (s, 0.3H), 6.02 (d, J = 8.0 Hz, 0.7H), 5.73 (d, J = 8.1 Hz, 0.3H), 5.64 (s, 0.7H), 5.06 (d, J = 16.2 Hz, 0.7H), 4.79 (d, J = 16.2 Hz, 0.7H), 4.58 (d, J = 15.3 Hz, 0.3H), 4.52 (d, J = 15.3 Hz, 0.3H), 3.75–3.64 (m, 1H), 2.04 (d, J = 9.0 Hz, 0.7H), 1.98 (s, 2.1H), 1.85–1.77 (m, 2H), 1.71–1.52 (m, 4H), 1.37–1.23 (m, 2H), 1.21–0.98 (m, 2H). ^{13}C NMR (126 MHz, $CDCl_3$, ∗: major isomer) δ 167.05, 166.83∗, 156.12, 150.98∗, 150.72, 149.83∗, 149.60, 137.46∗, 137.30∗, 137.19, 136.84, 130.54∗, 130.44, 128.92∗, 128.67, 128.25∗, 127.75∗, 127.70, 127.43, 123.43∗, 123.26, 92.78, 91.19, 73.52, 73.15, 64.66∗, 60.61, 52.48, 48.81, 47.03∗, 32.78, 32.72, 32.66, 25.57, 25.46, 24.80, 24.76, 4.39∗, 4.27. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for $C_{24}H_{27}N_3NaO_2$ 412.1995, found 412.1995.

N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-N-(naphthalen-2-yl)but-   2-ynamide (5.12) Pyridine-3-carboxaldehyde (0.09 mL, 0.93 mmol) and 2-naphthalenamine (0.13 mL, 0.93 mmol) were added with MeOH (3.7 mL) and stirred for 30 min. at room temperature. 2-butynoic acid (78.5 mg, 0.93 mmol) and cyclohexyl isocyanide (0.10 mL, 0.84 mmol) were added and the solution stirred at room temperature overnight. The reaction was monitored by TLC in 1:1 EtOAc/Hex. A stream of air was used to evaporate the MeOH. EtOAc was added to dissolve the impurities and the mixture was sonicated. Hexanes was added and vacuum filtration was used to collect the precipitate and hexanes was used to rinse. The white powder precipitate was dried using the vacuum filtration and the product (293mg, 82% yield) was obtained. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.52 (s, 1H), 8.43 (d, J = 4.8 Hz, 1H), 7.76 (dd, J = 20.0, 7.7 Hz, 2H), 7.67 (d, J = 8.9 Hz, 2H), 7.53 (d, J = 8.6 Hz, 1H), 7.49 (t, J = 7.2 Hz, 2H), 7.14 (s, 1H), 7.05 (dd, J = 8.2, 4.7 Hz,

1H), 6.16 (s, 1H), 6.12 (d, $J = 8.1$ Hz, 1H), 3.85 (d, $J = 12.8$ Hz, 1H), 2.69 (s, 2H), 2.17 (s, 4H), 1.76 – 1.64 (m, 2H), 1.62 (s, 1H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 207.12, 167.31, 155.41, 150.54, 149.00, 138.93, 136.39, 133.14, 132.88, 130.69, 129.73, 128.73, 128.40, 128.03, 127.77, 127.07, 126.65, 123.42, 92.63, 77.41, 77.16, 76.91, 73.92, 62.24, 49.14, 32.95, 32.87, 31.72, 31.07, 25.59, 25.41, 24.91, 24.85, 4.08. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{27}$H$_{27}$N$_3$NaO$_2$ 448.1995, found 448.1986.

_N_-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-_N_-(1-phenylpiperidin-4-yl)but-2-ynamide (5.13) Pyridine-3-carboxaldehyde (0.03 mL, 0.35 mmol) and 5.13-INT-2 (62 mg, 0.35 mmol) were added with MeOH (1.4 mL) and stirred for 30 min. at room temperature. 2-butynoic acid (30 mg, 0.35 mmol) and cyclohexyl isocyanide (0.04 mL, 0.32 mmol) were added and the solution stirred at room temperature overnight. The reaction was monitored by TLC and the MeOH was evaporated under air stream. The crude product was purified using column chromatography from 20-100% EtOAc in hexanes. The column was loaded by dissolving the crude products in DCM. Fractions 18-27 were collected. A white powder product (34 mg, 21% yield) was obtained and confirmed by NMR. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.59 – 8.50 (m, 2H), 7.74 (dt, $J = 8.0, 2.0$ Hz, 1H), 7.29 (dd, $J = 8.1, 4.9$ Hz, 1H), 7.25 – 7.22 (m, 1H), 6.96 – 6.89 (m, 2H), 6.85 (tt, $J = 7.3, 1.0$ Hz, 1H), 6.34 (d, $J = 8.1$ Hz, 1H), 5.08 (s, 1H), 4.53 – 4.41 (m, 1H), 3.80 (dddd, $J = 18.7, 12.0, 6.1, 3.2$ Hz, 2H), 3.72 (ddt, $J = 12.5, 4.7, 2.5$ Hz, 1H), 2.81 (dtd, $J = 31.4, 12.3, 2.6$ Hz, 2H), 2.25 (qd, $J = 12.3, 4.4$ Hz, 1H), 2.01 (s, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.85, 155.27, 150.82, 149.26, 149.13, 135.50, 131.77, 129.21, 123.52, 120.02, 116.80, 91.32, 77.29, 77.24, 77.03, 76.78, 73.39, 60.86, 58.48, 49.73, 49.43, 48.73, 32.66, 32.54, 31.89, 31.57, 30.55, 30.27, 29.03, 25.47, 24.64, 24.61, 22.70, 14.21, 14.13, 4.24, 4.18. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{28}$H$_{34}$N$_4$NaO$_2$ 481.2592, found 481.2596.

_N_-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-_N_-(6-phenylpyridin-3-yl)but-2-ynamide (5.14) Pyridine-3-carboxaldehyde (0.03 mL, 0.27 mmol) and 5.14-int-1 (46 mg, 0.27 mmol) were added with MeOH (1.1 mL) and stirred for 30 min. at room temperature. 2-butynoic acid (23 mg, 0.27 mmol) and cyclohexyl isocyanide (0.03 mL, 0.24 mmol) were added and the solution stirred at room temperature for 48 hrs. The reaction was monitored by TLC. A stream of air was used to evaporate the MeOH. EtOAc was added to dissolve the impurities and the mixture was sonicated. Hexanes was added and vacuum filtration was used to collect the precipitate and

hexanes was used to rinse. The white powder precipitate was dried using the vacuum filtration and the product (40 mg, 36% yield) was obtained. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.56 – 8.43 (m, 2H), 8.25 (s, 1H), 8.00 – 7.91 (m, 2H), 7.72 (d, J = 8.3 Hz, 1H), 7.65 (d, J = 8.3 Hz, 1H), 7.53 – 7.38 (m, 4H), 7.11 (ddd, J = 7.9, 4.8, 0.8 Hz, 1H), 6.18 (s, 1H), 5.85 (d, J = 8.1 Hz, 1H), 3.88 – 3.77 (m, 1H), 2.03 – 1.96 (m, 1H), 1.86 (d, J = 12.8 Hz, 1H), 1.71 (s, 3H), 1.69 – 1.56 (m, 1H), 1.43 – 1.30 (m, 2H), 1.27 – 1.02 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.20, 156.91, 155.09, 151.50, 151.37, 150.36, 139.15, 138.20, 137.93, 134.17, 129.90, 129.64, 128.96, 127.11, 123.53, 119.90, 93.29, 73.69, 61.35, 49.23, 32.94, 32.88, 25.55, 24.91, 24.84, 4.11. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{28}$H$_{28}$N$_4$NaO$_2$ 475.2104, found 475.2106.

tert-butyl 4-(N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamido)piperidine-1-carboxylate (5.15) Pyridine-3-carboxaldehyde (0.09 mL, 0.93 mmol) and tert-butyl 4-aminopiperidine-1-carboxylate (187 mg, 0.93 mmol) were added with MeOH (3.7 mL) and stirred for 30 min at room temperature. 2-Butynoic acid (79 mg, 0.93 mmol) and cyclohexyl isocyanide (0.10 mL, 0.84 mmol) were added and the solution stirred at room temperature for 48 h. The reaction was monitored by TLC. The crude product was purified using column chromatography from 18-100% EtOAc in hexanes. The column was loaded by dissolving the crude products in DCM. Fractions 23-39 were collected. A white powder product (319 mg, 79% yield) was obtained and confirmed by NMR. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.54 (dd, J = 4.8, 1.6 Hz, 1H), 8.52 (d, J = 2.4 Hz, 1H), 7.78 (d, J = 8.0 Hz, 1H), 7.68 (d, J = 8.0 Hz, 1H), 7.26 (s, 3H), 6.39 (s, 1H), 4.98 (s, 1H), 4.44 (s, 1H), 4.24 (s, 3H), 3.83 – 3.72 (m, 0H), 2.74 (d, J = 40.7 Hz, 2H), 2.05 (s, 3H), 2.03 (s, 1H), 1.44 (s, 8H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 171.11, 167.86, 156.61, 155.17, 154.58, 154.45, 154.32, 151.50, 150.25, 149.72, 149.28, 148.99, 136.29, 135.26, 134.52, 131.49, 123.63, 123.45, 123.27, 91.41, 79.97, 79.58, 77.34, 77.28, 77.08, 76.83, 73.34, 67.41, 62.87, 60.94, 60.37, 58.40, 48.99, 48.70, 32.92, 32.61, 32.48, 31.86, 30.55, 28.46, 28.37, 25.44, 25.35, 24.73, 24.60, 24.56, 22.67, 21.04, 14.19, 14.10, 4.24, 4.13, 1.00. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{27}$H$_{34}$N$_4$NaO$_4$ 505.2785, found 505.2806.

N-([1,1'-biphenyl]-4-yl)-N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-4-hydroxybut-2-ynamide (5.16) To a solution of 4-aminobiphenyl (85 mg, 0.50 mmol) in MeOH (2.0 mL) was added 3-pyridine carboxaldehyde (0.05 mL, 0.50 mmol) and stirred at room temp. for 30 min. 4-Hydroxy-2-butynoic acid (50 mg, 0.50 mmol) was added followed by cyclohexyl isocyanide (0.06

mL, 0.50 mmol), and the reaction stirred at room temp. overnight. The solvent was removed under a stream of air then suspended in EtOAc and hexanes, then filtered to collect precipitate to afford the desired product (100 mg, 43%) as white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.54 (d, J = 2.3 Hz, 1H), 8.44 (dd, J = 4.9, 1.6 Hz, 1H), 7.53 (dd, J = 7.5, 1.6 Hz, 2H), 7.51 – 7.38 (m, 5H), 7.38 – 7.31 (m, 1H), 7.24 – 7.17 (m, 2H), 7.07 (dd, J = 8.0, 4.8 Hz, 1H), 6.17 (d, J = 8.0 Hz, 1H), 6.14 (s, 1H), 4.08 (d, J = 1.4 Hz, 2H), 3.87 – 3.75 (m, 1H), 1.97 (d, J = 12.4 Hz, 1H), 1.85 (d, J = 12.4 Hz, 1H), 1.64 (ddd, J = 39.6, 20.5, 8.7 Hz, 2H), 1.42 – 1.28 (m, 2H), 1.23 – 1.02 (m, 2H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.24, 154.63, 151.36, 149.85, 141.65, 139.72, 138.15, 137.65, 131.20, 130.19, 129.00, 127.98, 127.40, 127.16, 123.32, 92.77, 78.67, 62.07, 50.56, 49.19, 32.88, 32.84, 25.55, 24.93, 24.86. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{29}$H$_{29}$N$_3$NaO$_3$ 490.2101, found 490.2116.

N-(2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-4-hydroxy-N-(6-phenylpyridin-3-yl)but-2-ynamide (5.17) To a solution of 6-phenyl-3-pyridinamine (85 mg, 0.50 mmol) in MeOH (2.0 mL) was added 3-pyridine carboxaldehyde (0.05 mL, 0.50 mmol) and stirred at room temp. for 30 min. 4-Hydroxy-2-butynoic acid (50 mg, 0.50 mmol) was added followed by cyclohexyl isocyanide (0.06 mL, 0.50 mmol), and the reaction stirred at room temp. overnight. The solvent was removed under a stream of air and the crude residue purified by column chromatography (20 – 100% EtOAc/Hex) to afford the desired product (112 mg, 48%) as a pale beige powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.56 (d, J = 2.3 Hz, 1H), 8.45 (dd, J = 4.7, 1.6 Hz, 1H), 8.30 (s, 1H), 7.93 (dd, J = 8.1, 1.6 Hz, 2H), 7.79 (d, J = 8.7 Hz, 1H), 7.65 (d, J = 8.3 Hz, 1H), 7.45 (tt, J = 10.5, 6.6 Hz, 4H), 7.11 (dd, J = 8.0, 4.8 Hz, 1H), 6.23 (s, 1H), 6.18 (d, J = 7.4 Hz, 1H), 4.08 (s, 2H), 3.88 – 3.73 (m, 1H), 3.06 (s, 1H), 1.96 (d, J = 12.7 Hz, 1H), 1.83 (d, J = 12.4 Hz, 1H), 1.77 – 1.52 (m, 4H), 1.41 – 1.23 (m, 3H), 1.22 – 0.98 (m, 2H), 0.87 (q, J = 6.6 Hz, 1H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.08, 157.22, 154.48, 151.50, 151.39, 150.26, 139.38, 138.03, 137.97, 133.79, 129.92, 129.75, 128.98, 127.16, 123.64, 120.07, 93.56, 78.40, 61.45, 50.54, 49.32, 32.88, 32.84, 25.53, 24.95, 24.86. HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ calculated for C$_{29}$H$_{29}$N$_3$NaO$_3$ 491.2054, found 491.2.

$tert$-butyl (1-phenylpiperidin-4-yl)carbamate (5.13-int-1) Phenyl iodide (250 mg, 1.23 mmol, 1.0 eq.), amine (368 mg, 1.84 mmol, 1.5 eq.), K$_2$CO$_3$ (339 mg, 2.45 mmol, 2.0 eq.), CuI (23 mg, 0.12 mmol, 0.1 eq.) and L-proline (28 mg, 0.25 mmol, 0.2 eq.) were combined in DMSO (8 mL)

and stirred at 80 °C overnight. The reaction mixture was diluted with dH$_2$O and extracted with EtOAc (x3). The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The crude product was further purified by column chromatography (3 – 25% EtOAc/Hexanes) to afford the desired product (123 mg, 36%) as a white powder. ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.26 – 7.22 (m, 2H), 6.95 – 6.90 (m, 2H), 6.84 (t, *J* = 7.3 Hz, 1H), 4.47 (s, 1H), 3.60 (d, *J* = 12.8 Hz, 2H), 2.89 – 2.76 (m, 2H), 2.04 (s, 2H), 1.54 (qd, *J* = 11.4, 4.0 Hz, 2H), 1.46 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 155.33, 151.45, 129.25, 119.82, 116.78, 79.54, 48.87, 47.92, 32.56, 28.57.

1-phenylpiperidin-4-amine (5.13-int-2) *tert*-Butyl (1-phenylpiperidin-4-yl)carbamate (123 mg, 0.45 mmol, 1.0 eq.) was dissolved in DCM (0.34 mL) and TFA (0.34 mL, 4.45 mmol, 10.0 eq.) and stirred at room temp. for 2 h. The reaction was diluted with DCM and washed with 0.1 M NaOH (x 3). The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo* to afford the desired product in 79% yield. The characterization data were in full agreement with those reported in literature.[46]

6-phenylpyridin-3-amine (5.14-int-1) A microwave vial was charged with 5-amino-2-bromopyridine (500 mg, 2.89 mmol, 1.0 eq.), phenyl boronic acid (705 mg, 5.78 mmol, 2.0 eq.), and Pd(PPh$_3$)$_4$ (347 mg, 0.3 mmol, 10 mol%). The vial was sealed and purged with argon. Dioxane (16 mL) was added and purged with argon, followed by a 2 M solution of Na$_2$CO$_3$ (4 mL). The reaction was stirred at 100 ºC overnight. The reaction was diluted with water and extracted three times with EtOAc. The combined organics were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered and concentrated *in vacuo*. The crude product was further purified by column chromatography (10 – 100% EtOAc/Hex) to afford the desired product as a yellow solid in 47 % yield. Characterization data are in full agreement with those reported in the literature.[47] ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.18 (dd, *J* = 2.9, 0.7 Hz, 1H), 7.89 (dd, *J* = 8.3, 1.3 Hz, 2H), 7.54 (dd, *J* = 8.4, 0.7 Hz, 1H), 7.43 (dd, *J* = 8.3, 7.0 Hz, 2H), 7.37 – 7.30 (m, 1H), 7.05 (dd, *J* = 8.4, 2.9 Hz, 1H).

***N*-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-1-(isoquinolin-4-yl)-2-oxoethyl)but-2-ynamide (5.19).** Compound was made and purified using general procedure A, pale white solid 90% yield, 392 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 9.11 (d, *J* = 0.7 Hz, 1H), 8.24 (s, 1H), 8.00– 7.94 (m, 2H), 7.82–7.73 (m, 1H), 7.65 (ddd, *J* = 7.9, 6.9, 1.0 Hz, 1H), 7.00 (d, *J* = 8.1 Hz, 2H),

6.95 (s, 1H), 6.86 (s, 2H), 5.62 (d, *J* = 8.1 Hz, 1H), 3.89 (dtd, *J* = 10.8, 7.8, 7.2, 3.9 Hz, 1H), 2.01–1.94 (m, 1H), 1.93 (s, 1H), 1.65 (s, 6H), 1.31 (d, *J* = 18.7 Hz, 2H), 1.15 (s, 9H), 1.10–0.97 (m, 3H). ^{13}C NMR (126 MHz, CDCl3) δ 167.30, 155.31, 153.07, 151.22, 144.79, 135.59, 134.71, 131.56, 129.62, 128.48, 127.90, 127.41, 124.89, 124.21, 122.31, 91.66, 73.65, 57.76, 48.95, 34.33, 32.74, 32.68, 30.99, 25.32, 24.73, 24.63, 3.79, 3.63. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for $C_{31}H_{36}N_3O_2$ 482.28020; found 482.28025.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(quinolin-3-yl)ethyl)but-2-ynamide (5.20). Compound was made and purified using general procedure A, pale white solid 69% yield, 298 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.70 (d, *J* = 2.2 Hz, 1H), 8.07 (dd, *J* = 8.4, 1.0 Hz, 1H), 8.00 (d, *J* = 2.4 Hz, 1H), 7.73–7.70 (m, 1H), 7.65 (dd, *J* = 8.2, 1.4 Hz, 1H), 7.52 (ddd, *J* = 8.1, 6.8, 1.2 Hz, 1H), 7.21 (d, *J* = 8.8 Hz, 2H), 7.01 (d, *J* = 8.0 Hz, 2H), 6.27 (s, 1H), 6.23–6.18 (m, 1H), 3.95–3.83 (m, 1H), 2.05–1.99 (m, 1H), 1.92–1.84 (m, 1H), 1.78–1.58 (m, 6H), 1.47–1.35 (m, 3H), 1.23 (s, 9H), 1.21–1.10 (m, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.79, 155.80, 152.21, 151.92, 147.77, 138.75, 136.61, 130.53, 130.17, 129.23, 128.49, 127.69, 127.51, 127.29, 126.06, 92.59, 74.16, 49.31, 35.00, 33.22, 33.15, 31.54, 25.85, 25.15, 25.11, 23.05, 14.51, 4.35. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for $C_{31}H_{36}N_3O_2$ 482.28020; found 482.28007.

N-(1-(benzo[*d*]thiazol-2-yl)-2-(cyclohexylamino)-2-oxoethyl)-N-(4-(*tert*-butyl)phenyl)but-2-ynamide (5.21). Compound was made and purified using general procedure A, pale white solid 91% yield, 398 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 9.61 (d, *J* = 7.9 Hz, 1H), 8.11–8.00 (m, 1H), 7.95–7.84 (m, 1H), 7.56–7.51 (m, 1H), 7.47 (d, *J* = 1.2 Hz, 1H), 7.46–7.35 (m, 2H), 7.29–7.22 (m, 2H), 6.15 (d, *J* = 1.6 Hz, 1H), 4.04–3.94 (m, 1H), 1.99 (d, *J* = 1.6 Hz, 3H), 1.91–1.64 (m, 4H), 1.55–1.34 (m, 6H), 1.24 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 171.84, 169.60, 163.31, 158.61, 152.14, 148.14, 135.47, 134.80, 126.83, 126.70, 126.56, 126.18, 124.24, 123.56, 122.32, 121.64, 121.47, 50.84, 49.26, 34.67, 32.77, 32.67, 31.72, 31.59, 31.37, 25.94, 25.14, 24.84, 24.68, 24.60, 13.39, 4.08. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for $C_{29}H_{34}N_3O_2S$ 488.23662; found 488.23672.

N-(1-(benzo[*b*]thiophen-3-yl)-2-(cyclohexylamino)-2-oxoethyl)-N-(4-(*tert*-butyl)phenyl)but-2-ynamide (5.22). Compound was made and purified using general procedure A, grey solid 92%

yield, 402 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.86–7.82 (m, 1H), 7.77–7.71 (m, 1H), 7.50 (s, 1H), 7.45–7.34 (m, 2H), 7.14–7.08 (m, 2H), 6.82 (s, 2H), 6.65 (s, 1H), 6.04 (d, J = 8.2 Hz, 1H), 3.88 (dddd, J = 14.5, 10.5, 7.9, 3.9 Hz, 1H), 2.01–1.91 (m, 1H), 1.92–1.87 (m, 1H), 1.79–1.60 (m, 6H), 1.47–1.28 (m, 3H), 1.23 (s, 9H), 1.21–1.07 (m, 2H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.45, 155.48, 151.34, 139.47, 138.29, 136.01, 129.42, 129.35, 128.26, 125.05, 124.60, 124.58, 122.81, 121.41, 91.83, 73.83, 56.30, 48.74, 34.53, 32.83, 32.79, 31.19, 25.49, 24.79, 24.75, 3.93. HRMS (ESI/Q-TOF) m/z: [M + H]$^+$ calculated for C$_{30}$H$_{35}$N$_2$O$_2$S 487.24138; found 487.24187.

N-(4-(*tert*-butyl)phenyl)-N-(1-(cyclohexylamino)-3-(2,5-dioxoimidazolidin-1-yl)-1-oxopropan-2-yl)but-2-ynamide (5.23). To a solution of 4-*tert*-butylaniline (0.56 mmol, 0.09 mL) in MeOH was added 2-(2,4-dioxoimidazolidin-1-yl)acetaldehyde (0.56 mmol, 80 mg) and stirred for 30 min. Butynoic acid (0.56 mmol, 47 mg) and cyclohexyl isocyanide (0.56 mmol, 0.07 mL) were added and the solution stirred at room temp. overnight. The resulting white precipitate was collected by filtration and rinsed with hexanes. The precipitate was dried over vacuum to afford the desired product (58 mg, 22% yield) as a white precipitate. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.35 (s, 1H), 7.39 (d, J = 8.5 Hz, 2H), 7.18 (d, J = 8.5 Hz, 2H), 6.87 (d, J = 8.0 Hz, 1H), 5.24 (dd, J = 9.3, 5.3 Hz, 1H), 4.07 (d, J = 18.0 Hz, 1H), 3.77 (d, J = 18.0 Hz, 1H), 3.70 (dd, J = 14.0, 5.3 Hz, 1H), 3.28 (dd, J = 13.9, 9.4 Hz, 1H), 1.89 (t, J = 16.9 Hz, 2H), 1.70 (s, 5H), 1.65–1.54 (m, 0H), 1.44–1.28 (m, 11H), 1.22 (q, J = 10.5 Hz, 4H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.49, 168.02, 156.59, 155.84, 152.37, 135.31, 128.88, 126.21, 93.11, 73.57, 56.15, 52.78, 48.72, 42.35, 34.89, 32.75, 32.71, 31.38, 25.59, 24.78, 24.74, 4.08. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{26}$H$_{34}$N$_4$O$_4$ 489.2472; found 489.2475.

2-(2,4-dioxoimidazolidin-1-yl)acetaldehyde (5.23-int-1) A solution of 1-allylhydantoin (1.43 mmol, 200 mg) in DCM (10 mL) and MeOH (10 mL) was cooled to -78 ºC. The solution was purged with nitrogen for 10 min., then O$_3$ was bubbled through the solution until a consistent blue colour was maintained. Nitrogen was bubbled through again until the blue colour dissipated, Resin-bound PPh$_3$ (717 mg, 3 mmol/g) was added, and the solution warmed to room temperature and stirred overnight. The resin was filtered through celite, rinsed with DCM, and the filtrate was concentrated *in vacuo* to afford the desired product (210 mg, 103% yield) as a white crystalline powder which was used directly in the next step without further purification.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclohexylamino)-2-oxo-1-(1*H*-pyrazol-4-yl)ethyl)propiolamide (5.24). 1*H*-pyrazole-4-carbaldehyde (100 mg, 1.04 mmol) and 4-*tert*-butylaniline (0.17 mL, 1.04 mmol) were stirred in MeOH (4.16 mL) and for 30 min at room temp. A white precipitate was observed. But-2-ynoic acid (87.4 mg, 1.04 mmol) and cyclohexyl isocyanide (0.12 mL, 0.936 mmol) were added and the solution stirred at room temp. overnight. Upon addition of the isocyanide, the white precipitate dissolved. A stream of air was used to evaporate the MeOH. EtOAc was added to dissolve the impurities and the mixture was sonicated. The product was finally triturated with hexanes and EtOAc. Vacuum filtration was used to collect the precipitate. The white powder precipitate was dried using the vacuum filtration and the product (186 mg, 47% yield) was obtained. Rf (5% (MeOH + 1% NH$_4$OH) in DCM): 0.45. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.63–8.39 (m, 2H), 7.50 (s, 2H), 7.28 (s, 1H), 7.00–6.93 (m, 2H), 6.38 (d, J = 8.1 Hz, 1H), 6.09 (s, 1H), 3.78 (tdd, J = 10.4, 7.1, 4.0 Hz, 1H), 1.96–1.82 (m, 2H), 1.69 (d, J = 8.5 Hz, 4H), 1.59 (dq, J = 13.0, 4.0 Hz, 1H), 1.41–1.31 (m, 1H), 1.28 (s, 9H), 1.24–1.12 (m, 2H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.63, 155.30, 151.92, 136.43, 134.76, 129.41, 126.03, 125.75, 114.45, 92.11, 73.85, 55.72, 48.82, 34.78, 32.86, 32.80, 31.37, 25.61, 24.80, 4.07. HRMS (ESI/Q-TOF) m/z: [M+Na]¬+ calculated for C25H32N4NaO2 443.2417, found 443.2426.

N-(4-(*tert*-butyl)phenyl)-N-(2-(*tert*-butylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamide (5.25). Compound was made and purified using general procedure A, white solid 82% yield, 300 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.45 (dd, J = 4.8, 1.5 Hz, 1H), 8.43 (d, J = 2.3 Hz, 1H), 7.43 (dt, J = 8.1, 2.0 Hz, 1H), 7.25–7.20 (m, 2H), 7.05 (dd, J = 8.0, 4.8 Hz, 1H), 6.97 (d, J = 8.0 Hz, 2H), 6.09 (s, 1H), 5.96 (s, 1H), 1.68 (s, 3H), 1.36 (s, 9H), 1.26 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.54, 155.43, 151.94, 151.27, 149.67, 138.25, 136.37, 130.29, 129.91, 125.77, 122.95, 92.21, 73.90, 62.66, 51.98, 34.76, 31.35, 28.75. HRMS (ESI/Q-TOF) m/z: [M + Na]$^+$ calculated for C$_{25}$H$_{31}$N$_3$NaO$_2$ 428.2308; found 428.2307.

N-(4-(*tert*-butyl)phenyl)-N-(2-(cyclopentylamino)-2-oxo-1-(pyridin-3-yl)ethyl)but-2-ynamide (5.26). Compound was made and purified using general procedure A, white solid 94% yield, 354 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.46 (dd, J = 4.9, 1.6 Hz, 1H), 8.44 (d, J = 2.3 Hz, 1H), 7.50–7.44 (m, 1H), 7.23 (d, J = 8.8 Hz, 2H), 7.07 (dd, J = 8.0, 4.8 Hz, 1H), 6.97 (d, J = 8.0 Hz, 2H), 6.20 (s, 1H), 6.02 (s, 1H), 4.23 (h, J = 6.8 Hz, 1H), 2.02 (dd, J = 13.0, 6.2 Hz, 1H), 1.95 (q, J = 2.7 Hz, 1H), 1.68 (s, 3H), 1.64–1.56 (m, 4H), 1.49–1.42 (m, 1H), 1.39–1.34 (m,

1H), 1.26 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 167.92, 155.47, 151.97, 151.06, 149.54, 138.34, 136.39, 130.32, 129.87, 125.79, 123.06, 92.26, 73.85, 62.23, 51.89, 34.76, 33.08, 33.01, 31.34, 23.87, 23.84, 4.06. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for C$_{26}$H$_{32}$N$_3$O$_2$ 418.24890; found 418.24854.

N-(2-(benzylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-N-(4-(*tert*-butyl)phenyl)but-2-ynamide (5.27). Compound was made and purified using general procedure A, pale yellow solid 86% yield, 340 mg. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.43 (d, *J* = 4.7 Hz, 1H), 8.40 (t, *J* = 1.6 Hz, 2H), 7.45 (d, *J* = 7.8 Hz, 1H), 7.33–7.16 (m, 6H), 7.05 (dd, *J* = 8.0, 4.8 Hz, 1H), 6.94 (d, *J* = 8.1 Hz, 2H), 6.66 (d, *J* = 8.5 Hz, 1H), 6.05 (s, 1H), 4.55–4.43 (m, 2H), 1.66 (s, 3H), 1.25 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 168.38, 155.47, 151.96, 151.08, 149.59, 138.41, 137.91, 136.33, 130.18, 129.93, 128.85, 127.84, 127.64, 125.78, 123.11, 92.27, 73.83, 62.35, 44.02, 34.74, 31.65, 31.33, 4.04. HRMS (ESI/Q-TOF) *m/z*: [M + H]$^+$ calculated for C$_{28}$H$_{30}$N$_3$O$_2$ 440.23325; found 440.23303.

2-(N-(4-(*tert*-butyl)phenyl)vinylsulfonamido)-N-(4-methoxyphenethyl)-2-(pyridin-3-yl)acetamide (5.28). Intermediate compound was made and purified using General procedure D. White solid with 42% yield, 176 mg. *Rf* = 0.2 (2:1 EtOAc:hexanes). ^{1}H NMR (500 MHz, CDCl$_3$): δ 8.66 (d, *J* = 2.3 Hz, 1H), 8.60 (dd, *J* = 4.9, 1.6 Hz, 1H), 7.60 (dt, *J* = 7.9, 1.8 Hz, 1H), 7.27 (d, *J* = 8.6 Hz, 2H), 6.93 (d, *J* = 8.5 Hz, 2H), 6.84 (t, *J* = 6.2 Hz, 1H), 6.77 (d, *J* = 8.5 Hz, 2H), 6.60 (d, *J* = 8.6 Hz, 2H), 4.74 (d, *J* = 2.3 Hz, 1H), 4.28 (s, 1H), 3.80 (s, 3H), 3.56 (ddt, *J* = 46.9, 13.6, 6.9 Hz, 2H), 2.73 (ddt, *J* = 59.3, 14.1, 6.7 Hz, 2H), 1.31 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$): δ 170.25, 158.27, 149.90, 149.08, 143.78, 142.65, 134.86, 134.70, 130.38, 129.71, 126.25, 123.94, 114.05, 113.67, 62.23, 55.26, 40.66, 34.72, 34.03, 31.49. HRMS (ESI) *m/z*: [M + Na]+ calculated for C$_{26}$H$_{31}$N$_3$NaO$_2$ 440.2308; found 440.2324.

Final compound was made and purified using General procedure E. White solid with 66% yield, 61 mg. *Rf* = 0.38 (1:1 EtOAc:DCM). ^{1}H NMR (500 MHz, CDCl$_3$): δ 8.49 (dd, *J* = 4.7, 1.6 Hz, 1H), 8.35 (d, *J* = 2.3 Hz, 1H), 7.22–7.18 (m, 3H), 7.08–7.04 (m, 3H), 6.94 (d, *J* = 8.6 Hz, 2H), 6.83 (d, *J* = 8.7 Hz, 2H), 6.11 (d, *J* = 16.6 Hz, 1H), 6.02 (t, *J* = 5.8 Hz, 1H), 5.94 (d, *J* = 9.9 Hz, 1H), 5.80 (s, 1H), 3.80 (s, 3H), 3.59 (ddt, *J* = 29.2, 13.2, 6.5 Hz, 2H), 2.87–2.74 (m, 2H), 1.26 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.34, 159.85, 149.90, 149.08, 143.83, 142.63, 140.01,

134.83, 134.69, 129.64, 126.26, 123.97, 121.08, 114.24, 113.67, 112.05, 62.31, 55.15, 40.40, 35.61, 34.02, 31.48. HRMS (ESI) *m/z*: [M + Na]+ calculated for $C_{28}H_{33}N_3NaO_4S$ 530.2084; found 530.2087.

2-(*N*-(4-(*tert*-butyl)phenyl)vinylsulfonamido)-*N*-(3-phenylpropyl)-2-(pyridin-3-yl)acetamide (5.29). Intermediate compound was made and purified using General procedure D. Pale yellow solid with 26% yield, 105 mg. *Rf* = 0.16 (1:1 EtOAc:hexanes). ^{1}H NMR (500 MHz, CDCl3) *δ* 8.71 (d, *J* = 2.3 Hz, 1H), 8.62 (dd, *J* = 4.8, 1.6 Hz, 1H), 7.75 (dt, *J* = 7.9, 2.0 Hz, 1H), 7.33 (dd, *J* = 7.8, 4.7 Hz, 1H), 7.28–7.25 (m, 3H), 7.22–7.19 (m, 1H), 7.11 (d, *J* = 6.7 Hz, 2H), 6.89 (t, *J* = 6.1 Hz, 1H), 6.64 (d, *J* = 8.6 Hz, 1H), 4.76 (s, 1H), 4.32 (s, 1H), 3.36 (q, *J* = 7.2 Hz, 2H), 2.59 (td, *J* = 8.3, 7.6, 2.4 Hz, 2H), 1.88–1.81 (m, 2H), 1.30 (s, 9H). ^{13}C NMR (126 MHz, CDCl3) *δ* 170.35, 149.93, 148.96, 143.76, 142.70, 141.20, 134.96, 134.82, 128.50, 128.34, 126.30, 126.06, 124.00, 113.67, 62.29, 39.19, 34.02, 33.20, 31.47, 31.13. HRMS (ESI) *m/z*: [M + H]+ calculated for $C_{26}H_{32}N_3O$ 402.2540; found 402.2535.

Final compound was made and purified using General procedure E. White solid with 80% yield, 103.0 mg. *Rf* = 0.50 (1:1 EtOAc:DCM). ^{1}H NMR (500 MHz, CDCl3): δ 8.52 (d, *J* = 4.2 Hz, 1H), 8.39 (s, 1H), 7.34 (dt, *J* = 8.0, 2.0 Hz, 1H), 7.31–7.26 (m, 3H), 7.22 (d, *J* = 8.5 Hz, 2H), 7.17 (d, *J* = 6.9 Hz, 2H),7.12 (dd, *J* = 8.0, 4.8 Hz, 1H), 7.04 (d, *J* = 8.6 Hz, 2H), 6.84 (dd, *J* = 16.6, 9.9 Hz, 1H), 6.12 (d, *J* = 16.5 Hz, 1H), 6.02 (s, 1H), 5.95 (d, *J* = 9.9 Hz, 1H), 5.82 (s, 1H), 3.38 (dt, *J* = 7.3, 5.4 Hz, 2H), 2.67 (td, *J* = 7.4, 1.3 Hz, 2H), 1.89 (p, *J* = 7.4 Hz, 2H), 1.26 (s, 9H). ^{13}C NMR (126 MHz, CDCl3) δ 168.64, 152.27, 151.26, 150.03, 141.06, 138.01, 135.58, 133.12, 131.45, 130.30, 128.54, 128.34, 127.54, 126.13, 126.04, 123.06, 65.45, 39.83, 34.62, 33.23, 31.18, 30.99. HRMS (ESI) *m/z*: [M + H]+ calculated for $C_{28}H_{34}N_3O_3S$ 493.2315; found 492.2313.

N-(1-benzylpiperidin-4-yl)-2-(*N*-(4-(*tert*-butyl)phenyl)vinylsulfonamido)-2-(pyridin-3-yl)acetamide (5.30). Intermediate compound was made and purified using General procedure D. Pale yellow solid with 25% yield, 346 mg. *Rf* = 0.31 (5% MeOH in DCM). ^{1}H NMR (500 MHz, CDCl3) δ 8.71 (d, *J* = 2.4 Hz, 1H), 8.61 (dd, *J* = 4.8, 1.6 Hz, 1H), 7.75 (dt, *J* = 7.9, 2.0 Hz, 1H), 7.36–7.29 (m, 4H), 7.28–7.22 (m, 3H), 6.78 (d, *J* = 8.3 Hz, 1H), 6.62 (d, *J* = 8.7 Hz, 2H), 4.75 (d, *J* = 2.5 Hz, 1H), 4.27 (d, *J* = 2.6 Hz, 1H), 3.92–3.77 (m, 1H), 3.48 (s, 2H), 2.77 (dd, *J* = 24.3, 11.3 Hz, 2H), 2.13 (dq, *J* = 11.3, 2.1 Hz, 2H), 1.90 (dddd, *J* = 16.2, 11.7, 4.0, 2.0 Hz, 2H), 1.70

(bs, 1H), 1.51 (qd, J = 11.3, 4.0 Hz, 1H), 1.41 (qd, J = 11.1, 4.0 Hz, 1H), 1.30 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 169.70, 149.94, 149.00, 143.76, 142.74, 134.93, 134.77, 129.06, 128.23, 127.07, 126.25, 123.97, 113.71, 62.99, 62.47, 52.11, 52.00, 46.66, 34.02, 32.02, 31.77, 31.47. HRMS (ESI) m/z: [M + H]+ calculated for C$_{29}$H$_{37}$N$_4$O 457.2962; found 457.1959.

methyl 2-((4-(*tert*-butyl)phenyl)amino)-2-(pyridin-3-yl)acetate (5.31-int-1). To a solution of 4-*tert*-butylaniline (1.0 mL, 6.30 mmol) in MeOH (15 mL) was added 3-pyridinecarboxaldehyde (0.6 mL, 6.30 mmol) and stirred at room temp. for 30 min. The reaction was cooled to -78 ℃ and cyclohexyl isocyanide (0.78 mL, 6.30 mmol) was added, followed by BF$_3$OEt$_2$ (2.4 mL, 18.9 mmol) dropwise. The reaction was stirred at -78 ℃ for 5 min then at 0 ℃ for 1 h. A solution of sat. NaHCO$_3$ was added and stirred for 1 h at room temp. The reaction was diluted with dH2O and extracted three times with EtOAc. The combined organics were washed with sat. NH$_4$Cl, sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$. The organics were filtered and concentrated *in vacuo*. The product was further purified by crashing out of EtOAc:Hexanes and obtained by vacuum filtration to afford the desired product (1.88 g, 66 % yield) as a white powder that was pure by ^{1}H and ^{13}C NMR. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.78 (d, J = 2.3 Hz, 1H), 8.56 (dd, J = 4.8, 1.7 Hz, 1H), 7.81 (dt, J = 7.7, 1.7 Hz, 1H), 7.30 – 7.26 (m, 1H), 7.16 (d, J = 8.6 Hz, 1H), 6.50 (d, J = 8.7 Hz, 2H), 5.09 (d, J = 5.2 Hz, 1H), 4.91 (d, J = 5.2 Hz, 1H), 3.75 (s, 3H), 1.24 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 171.74, 149.75, 149.56, 143.25, 141.53, 134.69, 133.95, 126.28, 123.90, 113.34, 58.99, 53.24, 34.02, 31.59. HRMS (ESI /QTOF?) m/z: [M + H]$^+$ calculated for C$_{18}$H$_{23}$N$_2$O$_2$ 299.17540 ; found 299.17499.

2-((4-(*tert*-butyl)phenyl)amino)-2-(pyridin-3-yl)acetohydrazide (5.31-int-2). 5.31-int-1 (500 mg, 1.68 mmol) was dissolved in EtOH (10 mL) and hydrazine monohydrate (0.41 mL, 8.4 mmol) was added. The reaction stirred at 70 ℃ overnight. TLC showed some remaining starting material so an additional aliquot of hydrazine (0.1 mL, 2.1 mmol) was added and stirred at 70 ℃ for 2 h. The reaction was diluted with dH$_2$O and extracted three times with DCM. The combined organics were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered and concentrated *in vacuo*. The product was further purified by crashing out of EtOAc and Hexanes and was obtained by vacuum filtration to afford the desired product (330 mg, 66 % yield) as a white powder that was pure by ^{1}H and ^{13}C NMR. ^{1}H NMR (400 MHz, DMSO-*d$_6$*) δ 9.54 (d, J = 4.0 Hz, 1H), 8.69 (d, J = 2.2 Hz, 1H), 8.46 (dd, J = 4.8, 1.6 Hz, 1H), 7.87 (dt, J = 7.9, 2.0 Hz, 1H), 7.36 (ddd, J = 7.8, 4.7,

0.9 Hz, 1H), 7.15 – 6.94 (m, 2H), 6.68 – 6.45 (m, 2H), 6.10 (d, J = 8.5 Hz, 1H), 5.01 (d, J = 8.5 Hz, 1H), 4.32 (d, J = 4.1 Hz, 2H), 1.18 (s, 9H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 169.65, 148.67, 148.65, 144.36, 139.07, 135.21, 134.66, 125.42, 123.46, 112.95, 57.00, 33.46, 31.38. HRMS (ESI /QTOF?) m/z: [M + H]$^+$ calculated for $C_{17}H_{23}N_4O$ 299.18664 ; found 299.18631.

2-((4-(tert-butyl)phenyl)amino)-*N'*-(2-(4-methoxyphenyl)acetyl)-2-(pyridin-3-yl) acetohydrazide (5.31-int-3). To a solution of 4-methoxyphenylacetic acid (140 mg, 0.84 mmol) in DCM (10 mL) was added carbonyldiimidazole (149 mg, 0.92 mmol) and the reaction stirred at room temp. for 30 min. **5.31-int-2** (250 mg, 0.84 mmol) was added and the reaction stirred at room temp. overnight. The reaction was concentrated *in vacuo*, suspended in cold DCM and filtered to obtain the desired product (305 mg, 81% yield) as a white powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.47 (s, 1H), 10.19 (s, 1H), 8.78 – 8.67 (m, 1H), 8.48 (dd, J = 4.8, 1.6 Hz, 1H), 7.94 (dt, J = 7.9, 2.0 Hz, 1H), 7.37 (ddd, J = 7.9, 4.8, 0.9 Hz, 1H), 7.19 (d, J = 8.7 Hz, 2H), 7.09 (d, J = 8.7 Hz, 2H), 6.85 (d, J = 8.6 Hz, 2H), 6.64 (d, J = 8.7 Hz, 2H), 6.15 (d, J = 8.4 Hz, 1H), 5.14 (d, J = 8.4 Hz, 1H), 3.71 (s, 3H), 3.37 (s, 2H), 1.19 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 169.34, 169.21, 158.01, 148.89, 148.80, 144.39, 139.23, 134.88, 134.67, 130.01, 127.47, 125.43, 123.47, 123.43, 113.65, 112.98, 56.94, 55.02, 33.49, 31.39, 31.33. HRMS (ESI /QTOF) m/z: [M + H]$^+$ calculated for $C_{26}H_{31}N_4O_3$ 447.23907 ; found 447.23897.

4-(tert-butyl)-*N*-((5-(4-methoxybenzyl)-1,3,4-oxadiazol-2-yl)(pyridin-3-yl)methyl)aniline (5.31-int-4). To a suspension of PS-PPh3 (~3 mmol/g, 210 mg, 0.62 mmol) in anhydrous DCM (7.5 mL) was added I$_2$ (157 mg, 0.62 mmol). The mixture was stirred at room temp. for 5 min. then cooled to 0 °C. Et$_3$N (0.17 mL, 1.24 mmol) was added and stirred for 5 min at 0 °C. **5.31-int-3** (140 mg, 0.31 mmol) was added and the reaction stirred at room temp. overnight. The reaction was filtered and diluted with sat. NaHCO3 and sat. Na$_2$S$_2$O$_3$ and extracted three times with DCM. The combined fractions were washed with brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The crude product was further purified by column chromatography (30 – 100% EtOAc/Hex) to afford the desired product (122 mg, 92 % yield) as a yellow oil. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.76 (s, 1H), 8.59 (s, 1H), 7.79 (dt, J = 8.0, 1.8 Hz, 1H), 7.29 (dd, J = 8.2, 4.6 Hz, 1H), 7.18 (d, J = 8.7 Hz, 2H), 7.14 (d, J = 8.6 Hz, 2H), 6.84 (d, J = 8.6 Hz, 2H), 6.60 (d, J = 8.7 Hz, 2H), 5.85 – 5.77 (m, 1H), 4.73 (s, 1H), 4.13 (d, J = 16.0 Hz, 1H), 4.06 (d, J = 15.9 Hz, 1H), 3.78 (s, 3H), 1.25 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 166.76, 166.09, 159.17, 150.17, 149.02,

142.97, 142.38, 134.86, 133.64, 129.93, 126.34, 125.42, 124.09, 114.46, 113.73, 55.41, 53.21, 34.06, 31.56, 31.06. HRMS (ESI /QTOF?) m/z: $[M + Na]^+$ calculated for $C_{26}H_{28}N_4O_2Na$ 451.21045 ; found 451.20914.

N-(4-(tert-butyl)phenyl)-N-((5-(4-methoxybenzyl)-1,3,4-oxadiazol-2-yl)(pyridin-3-yl)methyl)furan-2-carboxamide (5.31). To a solution of **5.31-int-4** (50 mg, 0.12 mmol) in anhydrous DCE (1.5 mL) at 0 ℃ was added Et$_3$N (19.5 μL, 0.14 mmol), followed by a solution of 2-furoyl chloride (13.8 μL, 0.14 mmol) in DCE (0.5 mL) dropwise. The reaction stirred at 0 ℃ for 1 h then was diluted with dH$_2$O and extracted three times with EtOAc. The combined organics were washed with sat. NH$_4$Cl, sat. NaHCO$_3$, and brine, dried over Na$_2$SO$_4$, filtered and concentrated *in vacuo*. The crude product was purified by column chromatography (100% DCM → 50% DCM/EtOAc). The combined fractions were concentrated *in vacuo* and triturated with acetone to afford the desired product (44 mg, 72% yield) as a pale yellow powder. ^{1}H NMR (400 MHz, CDCl$_3$) δ 9.26 (d, J = 2.5 Hz, 1H), 8.53 (dd, J = 4.8, 1.6 Hz, 1H), 8.37 (dt, J = 8.5, 1.9 Hz, 1H), 7.33 – 7.27 (m, 2H), 7.22 (d, J = 3.7 Hz, 1H), 7.02 (d, J = 8.7 Hz, 2H), 6.89 (d, J = 8.6 Hz, 2H), 6.69 (d, J = 8.7 Hz, 2H), 6.51 (d, J = 8.7 Hz, 2H), 6.37 (dd, J = 3.7, 1.7 Hz, 1H), 5.99 (s, 1H), 4.08 (d, J = 15.9 Hz, 1H), 3.99 (d, J = 15.9 Hz, 1H), 3.75 (s, 3H), 1.18 (s, 9H).^{13}C NMR (126 MHz, CDCl$_3$) δ 179.17, 166.49, 164.37, 158.89, 150.76, 149.73, 148.63, 147.71, 141.86, 139.73, 136.90, 130.64, 129.50, 125.98, 125.64, 123.45, 122.39, 115.01, 114.29, 112.88, 66.33, 55.39, 33.96, 31.52, 30.86. HRMS (ESI /QTOF?) m/z: $[M + H]^+$ calculated for $C_{31}H_{31}O_4N_4$ 523.23398 ; found 523.23392.

N-(4-(tert-butyl)phenyl)-N-(2-((4-methoxyphenethyl)amino)-2-oxo-1-(pyridin-3-yl)ethyl) furan-2-carboxamide (5.32). A solution of N-(4-methoxyphenethyl)formamide (223 mg, 1.24 mmol) and Et$_3$N (0.6 mL, 4.34 mmol) in DCM (3 mL) was cooled to -78 ℃. POCl$_3$ (0.17 mL, 1.86 mmol) was added dropwise and the reaction stirred at -78 ℃ for 3 h. The reaction was poured into sat. NaHCO$_3$ and extracted three times with DCM. The combined organic layers were washed with dH$_2$O and brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The crude product was further purified by silica plug, eluting with DCM, and concentrating *in vacuo* to afford the desired product (1-(2-isocyanoethyl)-4-methoxybenzene, 172 mg, 86% yield) as a colourless oil that was used directly in the next step. To a solution of 4-*tert*-butylaniline (0.08 mL, 0.52 mmol) in MeOH (1.5 mL) was added 3-pyridine carboxaldehyde (0.05 mL, 0.52 mmol) and stirred at room temp.

for 15 min. Furoic acid (58 mg, 0.52 mmol) was added followed by a solution of 1-(2-isocyanoethyl)-4-methoxybenzene (85 mg, 0.52 mmol) in MeOH (0.5 mL). The reaction stirred at room temp. overnight. The reaction was diluted with sat. $NaHCO_3$ and extracted three times with EtOAc. The combined organics were washed with sat. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo*. The crude product was further purified by column chromatography (50% DCM/EtOAc → 100% EtOAc) to afford the desired product (221 mg, 83% yield) as a beige crystalline powder. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.46 (dd, J = 4.8, 1.6 Hz, 1H), 8.44 (d, J = 2.0 Hz, 1H), 7.46 (dt, J = 8.0, 2.0 Hz, 1H), 7.40 (dd, J = 1.8, 0.7 Hz, 1H), 7.25 – 7.20 (m, 2H), 7.12 – 6.99 (m, 3H), 6.80 (s, 2H), 6.75 – 6.70 (m, 2H), 6.42 (t, J = 5.9 Hz, 1H), 6.17 (dd, J = 3.6, 1.7 Hz, 1H), 6.09 (s, 1H), 5.39 (dd, J = 3.7, 0.7 Hz, 1H), 3.73 (s, 3H), 3.67 – 3.47 (m, 2H), 2.86 – 2.72 (m, 2H), 1.28 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 168.72, 159.84, 158.37, 152.67, 151.52, 149.75, 146.34, 145.18, 138.37, 136.75, 130.79, 130.31, 130.03, 129.92, 126.28, 122.94, 117.38, 114.12, 111.39, 63.80, 55.30, 41.26, 34.84, 34.76, 31.40. HRMS (ESI /QTOF?) m/z: [M + Na]$^+$ calculated for $C_{31}H_{33}N_3O_4Na$ 534.2363 ; found 534.2371.

5.6.3 *In vitro* assays

Detection of inhibitors by fluorescence spectrophotometry. As reported in Chapter 4.

5.6.4 *In cellulo* assays

Drug tests in SARS-CoV-2 model. Huh7.5 cells were plated in 24 well-plate at 1 x 105 cell / mL and cultured over night at 37°C. Cells are infected or not by SARS-CoV-2 virus during 2 hours at 37°C at MOI of 0.05. After virus infection, the cells were treated with drugs at concentrations 10 or 20 μM or an equal volume of DMSO was added in negative control. After 48 hours of treatment the cells' supernatant was collected and viral titration was determined.

SARS-CoV-2 viral titration. Viral titrations were monitored by plaque assay method. In brief, 10-fold serial dilutions of supernatant containing an unknown amount of SARS-CoV-2 were absorbed on a monolayer of susceptible Vero E6 cells for 2 hours. After viral absorption, the supernatant was removed and a solid overlay Eagle's Minimal Essential Medium (EMEM, Wisent) containing 0.8% carboxymethylcellulose (SIGMA) was applied to the cells for 3 days. After fixation with 10 % of formaldehyde, cells were wash and coloured by 5% of crystal violet.

Cell viability assay. Cell viability was monitored by a mitochondrial activity test using methyl-thiazolyl-tetrazolium (MTT, Sigma). After 48 hours of drug treatment, cells were treated with 1 mg/mL of MTT during 3 hours at 37°C. The supernatants were removed and the MTT crystals were dissolved using DMSO and a shaker. The absorbance was read with a Tecan Spark at 570 nm with reference at 650 nm.

5.7 References

1.   Organization, W. H. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/ (accessed Oct, 7).

2.   Dai, L.; Gao, G. F., Viral targets for vaccines against COVID-19. *Nat Rev Immunol* **2021,** *21* (2), 73-82.

3.   Lau, J. J.; Cheng, S. M. S.; Leung, K.; Lee, C. K.; Hachim, A.; Tsang, L. C. H.; Yam, K. W. H.; Chaothai, S.; Kwan, K. K. H.; Chai, Z. Y. H., et al., Real-world COVID-19 vaccine effectiveness against the Omicron BA.2 variant in a SARS-CoV-2 infection-naive population. *Nat Med* **2023,** *29* (2), 348-357.

4.   Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C. O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J. A., et al., mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **2021,** *592* (7855), 616-622.

5.   Ledford, H., Six months of COVID vaccines: what 1.7 billion doses have taught scientists. *Nature* **2021,** *594* (7862), 164-167.

6.   Lin, D. Y.; Gu, Y.; Wheeler, B.; Young, H.; Holloway, S.; Sunny, S. K.; Moore, Z.; Zeng, D., Effectiveness of Covid-19 Vaccines over a 9-Month Period in North Carolina. *N Engl J Med* **2022,** *386* (10), 933-941.

7.   Abraham, N.; Spruin, S.; Rossi, T.; Fireman, B.; Zafack, J.; Blaser, C.; Shaw, A.; Hutchings, K.; Ogunnaike-Cooke, S., Myocarditis and/or pericarditis risk after mRNA COVID-19 vaccination: A Canadian head to head comparison of BNT162b2 and mRNA-1273 vaccines. *Vaccine* **2022,** *40* (32), 4663-4671.

8.   Galmiche, S.; Luong Nguyen, L. B.; Tartour, E.; de Lamballerie, X.; Wittkop, L.; Loubet, P.; Launay, O., Immunological and clinical efficacy of COVID-19 vaccines in immunocompromised populations: a systematic review. *Clin Microbiol Infect* **2022,** *28* (2), 163-177.

9.   Genc Bahce, Y.; Acer, O.; Ozudogru, O., Effectiveness of Inactivated and mRNA COVID-19 Vaccines Against SARS-CoV-2 Infection, Severe Disease and Mortality in the Geriatric Population. *Curr Microbiol* **2023,** *80* (6), 206.

10.   Bafna, K.; Cioffi, C. L.; Krug, R. M.; Montelione, G. T., Structural similarities between SARS-CoV2 3CL(pro) and other viral proteases suggest potential lead molecules for developing broad spectrum antivirals. *Front Chem* **2022,** *10*, 948553.

11.   Xiong, M.; Su, H.; Zhao, W.; Xie, H.; Shao, Q.; Xu, Y., What coronavirus 3C-like protease tells us: From structure, substrate selectivity, to inhibitor design. *Med Res Rev* **2021,** *41* (4), 1965-1998.

12.   Dai, W.; Jochmans, D.; Xie, H.; Yang, H.; Li, J.; Su, H.; Chang, D.; Wang, J.; Peng, J.; Zhu, L., et al., Design, Synthesis, and Biological Evaluation of Peptidomimetic Aldehydes as Broad-Spectrum Inhibitors against Enterovirus and SARS-CoV-2. *J Med Chem* **2022,** *65* (4), 2794-2808.

13.   Dai, W.; Zhang, B.; Jiang, X. M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F., et al., Structure-based design of antiviral drug candidates targeting the SARS-CoV-2

main protease. *Science* **2020,** *368* (6497), 1331-1335.

14.   Vuong, W.; Fischer, C.; Khan, M. B.; van Belkum, M. J.; Lamer, T.; Willoughby, K. D.; Lu, J.; Arutyunova, E.; Joyce, M. A.; Saffran, H. A., et al., Improved SARS-CoV-2 M(pro) inhibitors based on feline antiviral drug GC376: Structural enhancements, increased solubility, and micellar studies. *Eur J Med Chem* **2021,** *222*, 113584.

15.   Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A., et al., Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun* **2020,** *11* (1), 4282.

16.   Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y., et al., Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J Med Chem* **2020,** *63* (21), 12725-12747.

17.   Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J., et al., An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021,** *374* (6575), 1586-1593.

18.   Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **2020,** *368* (6489), 409-412.

19.   De Cesco, S.; Deslandes, S.; Therrien, E.; Levan, D.; Cueto, M.; Schmidt, R.; Cantin, L. D.; Mittermaier, A.; Juillerat-Jeanneret, L.; Moitessier, N., Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells. *J Med Chem* **2012,** *55* (14), 6306-15.

20.   Plescia, J.; De Cesco, S.; Patrascu, M. B.; Kurian, J.; Di Trani, J.; Dufresne, C.; Wahba, A. S.; Janmamode, N.; Mittermaier, A. K.; Moitessier, N., Integrated Synthetic, Biophysical, and Computational Investigations of Covalent Inhibitors of Prolyl Oligopeptidase and Fibroblast Activation Protein alpha. *J Med Chem* **2019,** *62* (17), 7874-7884.

21.   Plescia, J.; Dufresne, C.; Janmamode, N.; Wahba, A. S.; Mittermaier, A. K.; Moitessier, N., Discovery of covalent prolyl oligopeptidase boronic ester inhibitors. *Eur J Med Chem* **2020,** *185*, 111783.

22.   Yang, S.; Chen, S. J.; Hsu, M. F.; Wu, J. D.; Tseng, C. T.; Liu, Y. F.; Chen, H. C.; Kuo, C. W.; Wu, C. S.; Chang, L. W., et al., Synthesis, crystal structure, structure-activity relationships, and antiviral activity of a potent SARS coronavirus 3CL protease inhibitor. *J Med Chem* **2006,** *49* (16), 4971-80.

23.   Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J., et al., alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J Med Chem* **2020,** *63* (9), 4562-4578.

24.   Muramatsu, T.; Takemoto, C.; Kim, Y. T.; Wang, H.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S., SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc Natl Acad Sci U S A* **2016,** *113* (46), 12997-13002.

25.   Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J., et al., Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol* **2005,** *3* (10), e324.

26.   Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S. A.; Chase, P.; Eggler, A.; Dawson, E. S.; Baez-Santos, Y. M.; Tomar, S., et al., Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease. *J Med Chem* **2013,** *56* (2), 534-46.

27.   St John, S. E.; Mesecar, A. D. Broad spectrum non-covalent coronavirus protease inhibitors. 2018.

28.   Kitamura, N.; Sacco, M. D.; Ma, C.; Hu, Y.; Townsend, J. A.; Meng, X.; Zhang, F.; Zhang, X.; Ba, M.; Szeto, T., et al., Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *J Med Chem* **2022,** *65* (4), 2848-2865.

29.   Gutierrez, C. D.; Bavetsias, V.; McDonald, E., TiCl(OPr) and NaBH(OAc): an efficient reagent combination for the reductive amination of aldehydes by electron-deficient amines. *Tetrahedron Lett* **2005,** *46* (20), 3595-3597.

30.   Schäfer, G.; Reber, S.; Ahmetovic, M.; Fleischer, T.; Leuenberger, D.; Blumer, N.; Udry, M.; Abele, S., Robust and Scalable Reductive Amination Protocol for Electron-Poor Heterocyclic Amines Using EtSiH/TFA as Reducing Agent. *Synthesis-Stuttgart* **2023,** *55* (09), 1328-1336.

31.   Boltjes, A.; Domling, A., The Groebke-Blackburn-Bienayme Reaction. *Eur J Chem* **2019,** *2019* (42), 7007-7049.

32.   Zhang, H. Z.; Zhang, H.; Kemnitzer, W.; Tseng, B.; Cinatl, J., Jr.; Michaelis, M.; Doerr, H. W.; Cai, S. X., Design and synthesis of dipeptidyl glutaminyl fluoromethyl ketones as potent severe acute respiratory syndrome coronovirus (SARS-CoV) inhibitors. *J Med Chem* **2006,** *49* (3), 1198-201.

33.   Dragovich, P. S.; Prins, T. J.; Zhou, R.; Webber, S. E.; Marakovits, J. T.; Fuhrman, S. A.; Patick, A. K.; Matthews, D. A.; Lee, C. A.; Ford, C. E., et al., Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements. *J Med Chem* **1999,** *42* (7), 1213-24.

34.   Stubbing, L. A.; Hubert, J. G.; Bell-Tyrer, J.; Hermant, Y. O.; Yang, S. H.; McSweeney, A. M.; McKenzie-Goldsmith, G. M.; Ward, V. K.; Furkert, D. P.; Brimble, M. A., P(1) Glutamine isosteres in the design of inhibitors of 3C/3CL protease of human viruses of the Pisoniviricetes class. *RSC Chem Biol* **2023,** *4* (8), 533-547.

35.   Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y., et al., Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* **2020,** *30* (8), 678-692.

36.   Turlington, M.; Chun, A.; Tomar, S.; Eggler, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J. S.; Dawson, E.; Saldanha, A.; Chase, P., et al., Discovery of N-(benzo[1,2,3]triazol-

1-yl)-N-(benzyl)acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg Med Chem Lett* **2013,** *23* (22), 6172-7.

37.   Lambruschini, C.; Moni, L.; Banfi, L., Diastereoselectivity in Passerini Reactions of Chiral Aldehydes and in Ugi Reactions of Chiral Cyclic Imines. *Eur J Org Chem* **2020,** *2020* (25), 3766-3778.

38.   Moni, L.; Banfi, L.; Basso, A.; Carcone, L.; Rasparini, M.; Riva, R., Ugi and Passerini reactions of biocatalytically derived chiral aldehydes: application to the synthesis of bicyclic pyrrolidines and of antiviral agent telaprevir. *J Org Chem* **2015,** *80* (7), 3411-28.

39.   Galasiti Kankanamalage, A. C.; Kim, Y.; Damalanka, V. C.; Rathnayake, A. D.; Fehr, A. R.; Mehzabeen, N.; Battaile, K. P.; Lovell, S.; Lushington, G. H.; Perlman, S., et al., Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element. *Eur J Med Chem* **2018,** *150*, 334-346.

40.   Friedman, H. L., Influence of isosteric replacements upon biological activity. *NASNRS* **1951,** *206*, 295-358.

41.   Kumari, S.; Carmona, A. V.; Tiwari, A. K.; Trippier, P. C., Amide Bond Bioisosteres: Strategies, Synthesis, and Successes. *J Med Chem* **2020,** *63* (21), 12290-12358.

42.   Konno, S.; Kobayashi, K.; Senda, M.; Funai, Y.; Seki, Y.; Tamai, I.; Schakel, L.; Sakata, K.; Pillaiyar, T.; Taguchi, A., et al., 3CL Protease Inhibitors with an Electrophilic Arylketone Moiety as Anti-SARS-CoV-2 Agents. *J Med Chem* **2022,** *65* (4), 2926-2939.

43.   Ramazani, A.; Rezaei, A., Novel one-pot, four-component condensation reaction: an efficient approach for the synthesis of 2,5-disubstituted 1,3,4-oxadiazole derivatives by a Ugi-4CR/aza-Wittig sequence. *Org Lett* **2010,** *12* (12), 2852-5.

44.   Groebke Zbinden, K.; Banner, D. W.; Ackermann, J.; D'Arcy, A.; Kirchhofer, D.; Ji, Y. H.; Tschopp, T. B.; Wallbaum, S.; Weber, L., Design of selective phenylglycine amide tissue factor/factor VIIa inhibitors. *Bioorg Med Chem Lett* **2005,** *15* (3), 817-22.

45.   Patil, P.; Ahmadian-Moghaddam, M.; Dömling, A., Isocyanide 2.0. *Green Chemistry* **2020,** *22* (20), 6902-6911.

46.   Jean, L.; Rouden, J.; Maddaluno, J.; Lasne, M. C., Palladium-mediated arylation of 3-aminopiperidines and 3-aminopyrrolidines. *J Org Chem* **2004,** *69* (25), 8893-902.

47.   Yang, J. F.; Liu, S. J.; Zheng, J. F.; Zhou, J. R., Room-Temperature Suzuki-Miyaura Coupling of Heteroaryl Chlorides and Tosylates. *Eur J Org Chem* **2012,** *2012* (31), 6248-6259.

CHAPTER 6:

ASYMMETRIC ORGANOCATALYTIC PICTET-SPENGLER REACTION WITH DIKETONE SUBSTRATES

Contribution of Authors

All work presented in this chapter is the work of **J.K. Stille**. It was performed in the laboratory of Dr. Jieping Zhu at Ecole Polytechnique Fédérale de Lausanne as a visiting doctoral student from February to July 2022. Catalyst 6.1 was previously synthesized and characterized by R. Andres.

6.1 Abstract

Tetrahydro-β-carbolines (TBHCs) are prevalent among many indole alkaloid natural products with well-documented bioactivity, and several synthetic drugs have been devised based on these scaffolds. However, the majority of asymmetric Pictet-Spengler reactions to access TBHCs are unsuccessful with ketone substrates and can only be applied to the synthesis of tertiary TBHC scaffolds. To address this limitation, this chapter presents the development of an asymmetric organocatalytic Pictet-Spengler reaction for diketone substrates, enabling the formation of enantioenriched TBHCs with quaternary stereocenters. The experimental conditions were optimized, using a chiral urea-based catalyst and carboxylic acid co-catalyst system. When evaluated on various diketone substrates, the optimized reaction conditions achieved high yields (51-90%) and high enantioselectivities (75-95%), as well as good functional group tolerance.

6.2 Introduction

6.2.1 Pictet-Spengler Reaction

The Pictet-Spengler reaction was first discovered in 1911 and described the reactivity of β-phenylethylamine with formaldehyde methyl acetal in the presence of hydrochloric acid to form 1,2,3,4-tetrahydroisoquinoline.[1] The reaction was subsequently extended to tryptamine and acetaldehyde, generating 1-methyl-1,2,3,4-tetrahydro-β-carboline as the product.[2] The Pictet-Spengler reaction can generally be described as the reaction of phenylethylamine or tryptamine derivatives with carbonyl groups to form tetrahydroisoquinoline (THIQ) or tetrahydro-β-carboline (THBC) products, respectively (**Scheme 6.1**). However, many variants of the Pictet-Spengler reaction have since been developed, expanding the scope of potential substrates to include sulfonamides,[3] carbamates,[4] hydroxylamines,[5] and alcohols[6], to name a few.[7, 8]
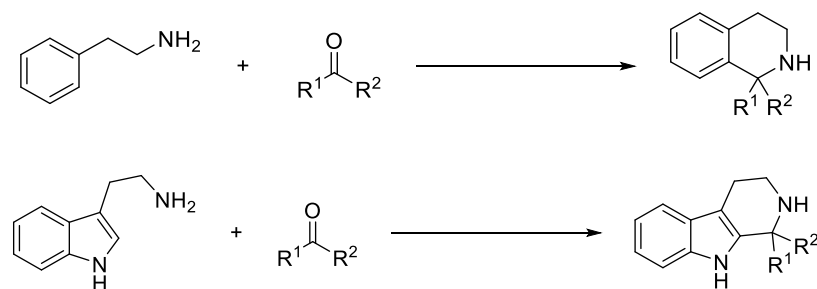
Scheme 6.1. Pictet-Spengler reaction to generate tetrahydroisoquinoline (top) and tetrahydro-β-

carbolines (bottom).

The mechanism of the Pictet-Spengler reaction has been widely studied and is generally considered to follow one of two mechanistic pathways (**Scheme 6.2**).[9-12] The reaction begins by the formation of an imine through condensation of the amine and carbonyl starting materials. Intermediate I then undergoes nucleophilic attack, either via the C-2 position (path a) to form the tetrahydro-β-carboline intermediate II or via the C-3 position (path b) via spiroindolization to form *aza*-spiroindolenine intermediate III. The intermediate formed through path b can then undergo C-C bond migration to generate the same intermediate II formed through path a, which ultimately undergoes rearomatization to form the final product IV.
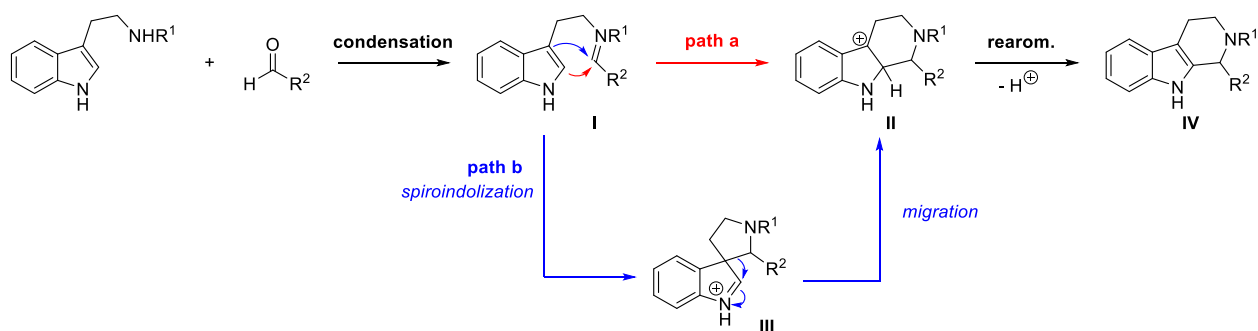
Scheme 6.2. Mechanism of the Pictet-Spengler reaction.

Since its discovery more than 100 years ago, the Pictet-Spengler reaction remains as one of the main synthetic strategies employed to access THIQ and THBC scaffolds. These scaffolds are found in many indole and isoquinoline alkaloids, making the Pictet-Spengler reaction a key transformation in the total synthesis of these classes of natural products, as noted by several reviews detailing its application in this field.[13, 14] Indole and isoquinoline alkaloids are widespread in nature and are produced by a vast range of organisms, including plants, fungi, and animals. Their bioactivity has been well-documented, with many of these alkaloid natural products displaying potent therapeutic activity.[15, 16] For example the indole alkaloid ajmaline was first isolated from the roots of *Raowulfina serpentina*[17] and has since found widespread use as an antiarrhythmic agent in the treatment of several cardiovascular disorders.[18, 19] The isoquinoline alkaloid neferine, found in the seeds of the lotus plant *Nelumbo nucifera*, has been shown to display anti-inflammatory[20] and anti-cancer[21, 22] effects. Inspired by the therapeutic potential of indole and

isoquinoline natural products, several synthetic drugs have been conceived based on these scaffolds, such as the FDA-approved drugs tadalafil and etodolac. In both these examples, a Pictet-Spengler reaction is employed as a key reaction to access their tricyclic core, demonstrating its significance in medicinal chemistry applications.[23]
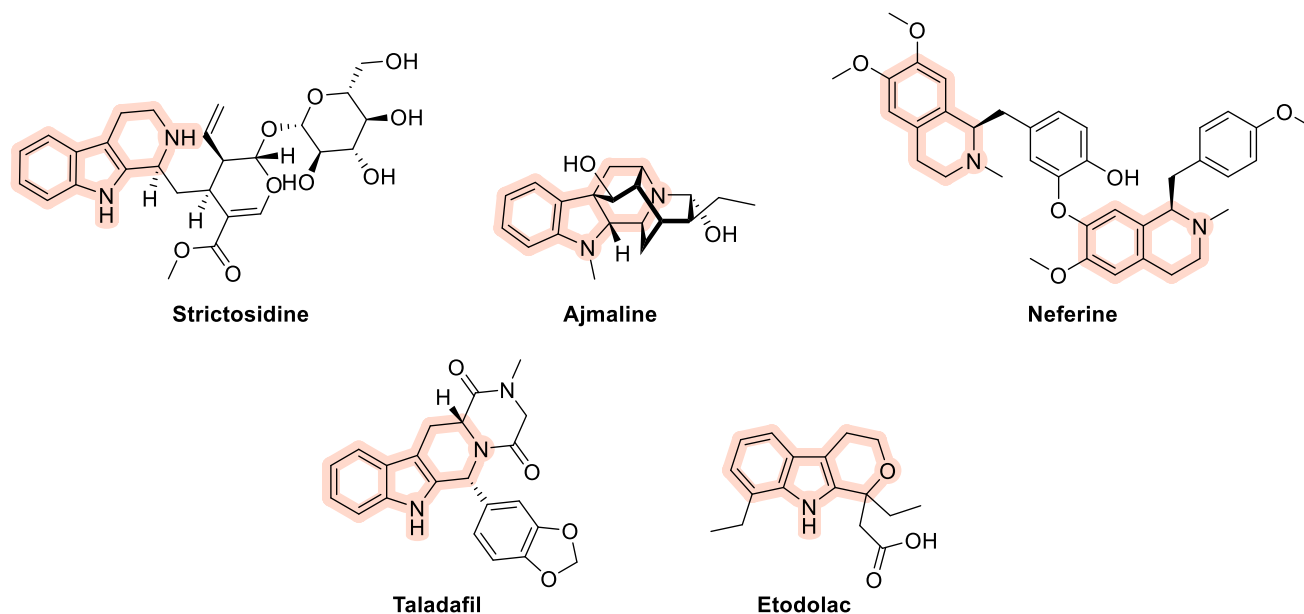
Figure 6.1. Examples of indole and isoquinoline natural products featuring THIQ and THBC scaffolds (highlighted), and two examples of approved drugs based on similar scaffolds.

6.2.2 Asymmetric Variants of the Pictet-Spengler Reaction

As the reaction results in the formation of a chiral centre, there has been significant interest in the development of asymmetric Pictet-Spengler reactions. Stereocontrol is an important aspect of synthetic chemistry, particularly when applied to the synthesis of therapeutic agents. Currently, more than 50% of drugs on the market are chiral. While many chiral drugs have been administered as racemic mixtures, there are several distinct advantages to enantiopure drugs. From an efficacy standpoint, it is expected that enantiomers will bind to their therapeutic target with differing binding affinities. Consequently, enantiopure drugs are expected to display greater potency as they contain only the active enantiomer. The undesired enantiomer may not only lack therapeutic activity, but could also be responsible for undesired side effects, as has been observed with several chiral drugs (e.g., naproxen, ethambutol).[24] In addition to differing binding affinities, enantiomers

may also display different pharmacokinetic profiles.[25] Following guidelines established by the FDA in 1992, thorough pharmacokinetic assessment of racemic drugs may necessitate the separate characterization of each enantiomer individually, significantly increasing the experimental workload.[26] It is therefore advantageous to have synthetic methods that allow access to enantiopure (or enantioenriched) products as this can enable the selective formation of the desired enantiomer without unnecessary formation of the undesired enantiomer.

The first asymmetric Pictet-Spengler reaction was reported by Taylor and Jacobsen in 2004, employing a chiral thiourea catalyst and *in situ* formation of an *N*-acyliminium intermediate (**Scheme 6.3**). Since this report, many other asymmetric variants have been reported, predominantly using chiral (thio)urea, squaramide and phosphoric acid catalysts.[27-30]
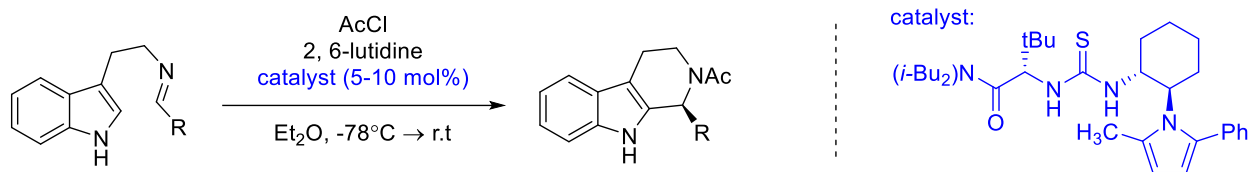
Scheme 6.3. Asymmetric Pictet-Spengler first reported by Taylor and Jacobsen.

Almost all of the developed asymmetric Pictet-Spengler reactions are limited to aldehyde substrates and therefore limited to the formation of a tertiary stereocenter. Even among racemic Pictet-Spengler reactions, examples using ketones are uncommon due their decreased reactivity towards imine formation and subsequent nucleophilic attack of the ketimine intermediate. To circumvent this issue, activated ketones such as α-ketoamides and α-ketoesters have successfully been employed in racemic Pictet-Spengler reaction, however there are still very few examples of catalytic enantioselective methods for ketone substrates. At the time of this work, these examples were limited to those shown in **Scheme 6.4** and **Scheme 6.5**.[31-34]
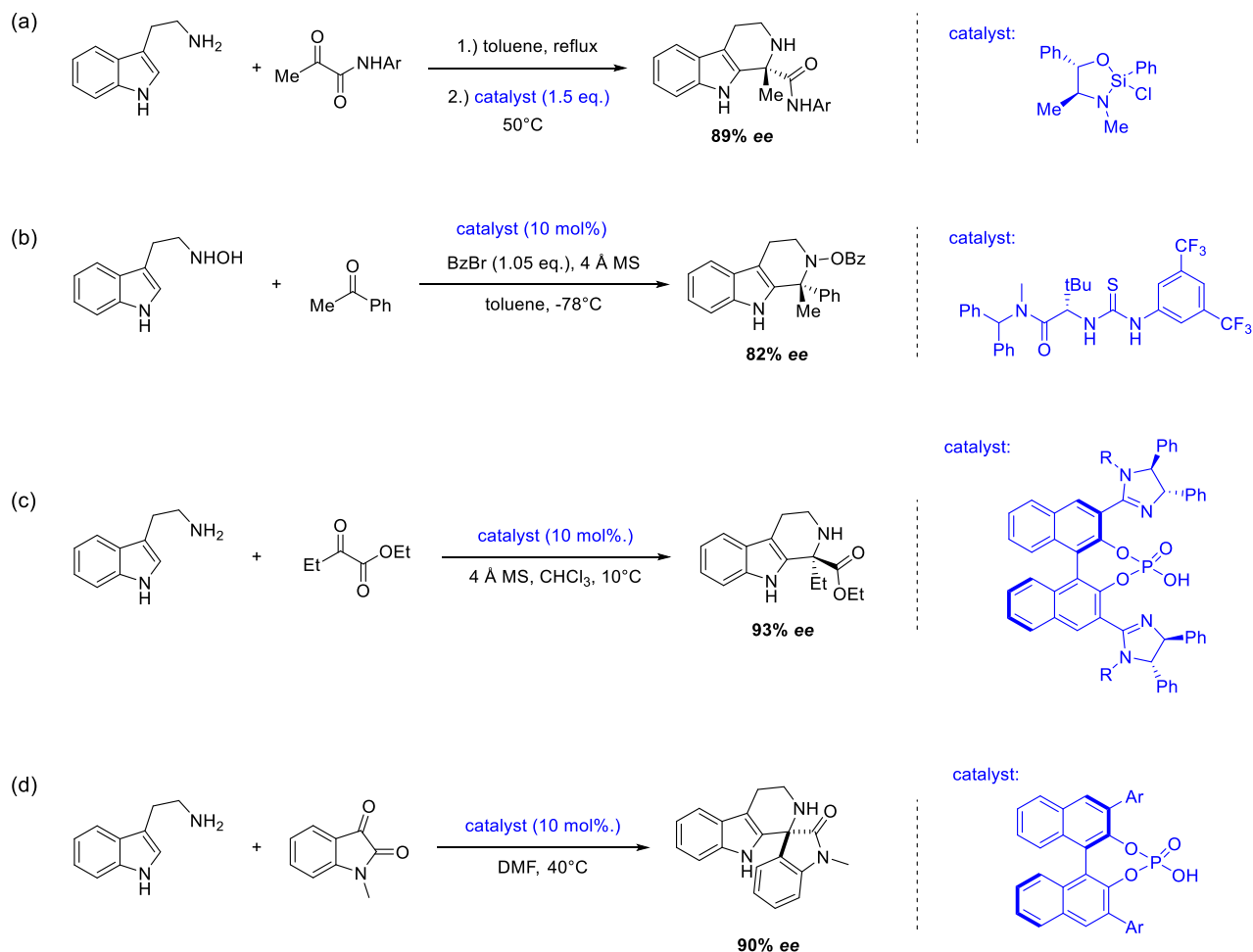
Scheme 6.4. Reported enantioselective reactions of ketone substrates using a) α-ketoamides[31]. b) ketones with *N*-hydroxytryptamine[32]. c) α-ketoesters[33], and d) isatins[34].

The Zhu group was initially interested in the development of an asymmetric Pictet-Spengler reaction as a means to access the 1,1-disubstituted TBHC core of arborisidine, as reported by Andres in 2020, leading to the development of an asymmetric Pictet-Spengler with 2,3-pentanedione (**Scheme 6.5**).[35]
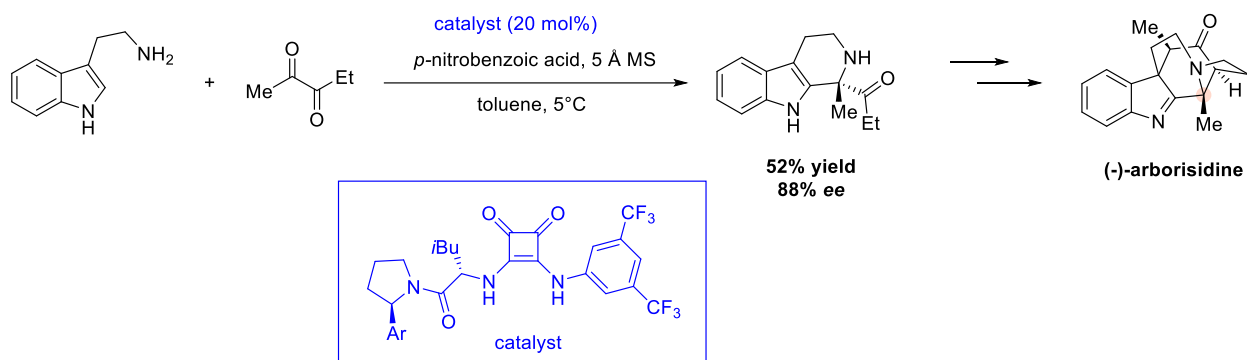
Scheme 6.5. Asymmetric Pictet-Spengler reaction with 2,3-pentandione in the synthesis of (-)-arborisidine.[35]

Further catalyst optimization by Zhu *et al.* led to the identification of urea-based catalyst **6.1** which displayed high yields and enantioselectivities for ketoamides (as reported in a recent publication[36]) as well as with ketoesters and diketones.[37]
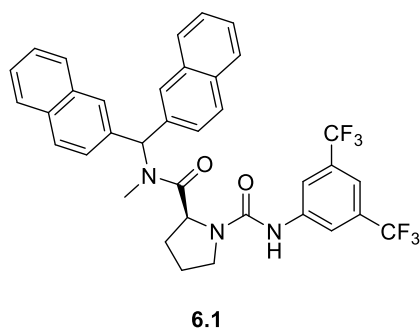
Figure 6.2. Optimized catalyst **6.1** for diketone and ketoamide substrates.
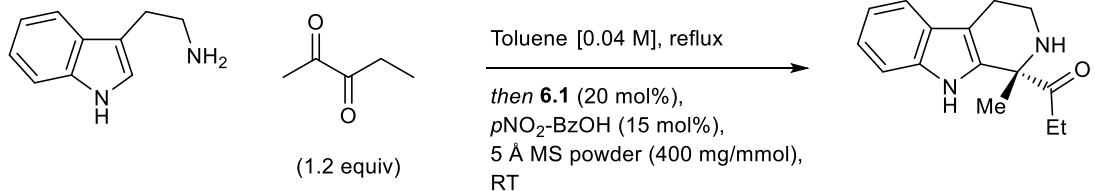
As shown in **Scheme 6.4**, most ketone-based Pictet-Spengler reactions employ ketoamides or ketoesters as substrates. The development of a protocol with diketones could provide additional synthetic versatility in the formation of 1,1-disubstituted TBHC scaffolds. The objective of this work is to optimize the reaction conditions for an asymmetric Pictet-Spengler reaction using optimized catalyst **6.1** and evaluate its selectivity on a diverse diketone substrate scope.

6.3 Results

6.3.1 Reaction Optimization

Prior to exploring the substrate scope of the reaction, the effect of several different reaction conditions on enantioselectivity was explored and the results are summarized in **Table 6.1.** 2,3-Pentanedione was used as the substrate for all optimization reactions, as the diketone product had previously been synthesized and characterized by Andres in his synthesis of (-)-arborisidine.[35] The results of reaction condition optimization are summarized in **Table 6.1**, with the enantiomeric excess (*ee*) determined by supercritical fluid chromatography (SFC).

Table 6.1. Optimization of reaction conditions.

Entry	Deviations from the conditions above	Yield [%]	ee [%]
1	None	n.d.	79
2	Toluene/trifluorotoluene 1/1	n.d.	73
3	Toluene/fluorobenzene 1/1	n.d.	73
4	Toluene/chlorobenzene 1/1	n.d.	74
5	Toluene/*m*xylene 1/1	n.d.	80
6	Toluene/mesitylene 1/1	n.d.	80
7	Toluene/*n*hexane 5/1	n.d.	79
8	Toluene/cyclohexane 5/1	n.d.	80
9	*o*CN-BzOH	n.d.	79
10	*p*CF$_3$-BzOH	n.d.	78
11	*o*NO$_2$-BzOH	n.d.	78
12	*o*I-BzOH	n.d.	75
13	*o*F-BzOH	n.d.	87
14	*m*Cl-BzOH	n.d.	78
15	*o*OH-BzOH	n.d.	85
16	*m*OH-BzOH	n.d.	72
17	*p*OH-BzOH	n.d.	74
18	F$_5$C$_6$CO$_2$H	n.d.	78
19	*m*xylene, *o*CF$_3$BzOH	n.d.	80
20	*o*F-BzOH, 4 Å MS powder	n.d.	85
21	*o*xylene, *o*F-BzOH	n.d.	81
22	*m*xylene, *o*F-BzOH	n.d.	89
23	*p*xylene, *o*F-BzOH	n.d.	90
24	mesitylene, *o*F-BzOH	n.d.	78
25	Toluene/*n*hexane 5/1, *o*F-BzOH	n.d.	82
26	*p*xylene, *o*F-BzOH, reflux 0h	n.d.	89
27	*p*xylene, *o*F-BzOH, reflux 2h	n.d.	90
28	*p*xylene, *o*F-BzOH, reflux 24h	n.d.	77

All the reactions were performed on a 15 mg scale. The reaction was stopped after 24 h in order to determine the *ee* even if full conversion wasn't reached. n.d. = not determined.

Several reaction conditions were screened for their effect on enantioselectivity, including acid, solvent, molecular sieves, and reflux time. The solvent (entries 2-8) had a modest effect on enantioselectivity, with minor improvements observed when using *m*-xylene or mesitylene (entries 5 & 6). Further solvent screening displayed a slight improvement in enantioselectivity when employing *p*-xylene as the solvent (entry 23). A substantial decrease in enantioselectivity was observed when more polar solvents such as trifluorotoluene, fluorobenzene and chlorobenzene were introduced (entries 2-4). These results follow the expected trend, as catalysis with chiral

(thio)ureas is known to be more selective in nonpolar solvents of low dielectric strength due to the increased favourability of ion pairing.[38]

The largest improvement in selectivity was observed when screening the acids, with *o*OH-BzOH and *o*F-BzOH having the largest influence (entries 13 & 15). It appears that the presence of a small hydrogen-bond acceptor in the *ortho* position improves the enantioselectivity, while larger substituents at this position (e.g., nitro, iodo, and CF_3) did not result in increased selectivity. There appears to be a weak correlation between acidity and selectivity, with the most weakly acidic reagents (*m*OH-BzOH and *p*OH-BzOH) resulting in the lowest selectivity. However, the most acidic reagent ($F_5C_6CO_2H$) did not improve the enantioselectivity despite its increased acidity and the presence of an *ortho*-fluoro group, suggesting there is possibly a limit to these trends. The acid would be expected to influence the enantioselectivity both by its influence on ion pairing (electronically and sterically) as well as its effect on background reactivity without contribution from the chiral thiourea, and thus benzoic acid substituents may have conflicting influences on these two critical factors.

Similar enantioselectivities were observed when the reaction was run entirely at room temperature and omitting the reflux time, while a decrease in selectivity was observed when the amine and diketone were refluxed for 24 h (entries 26-28). This is perhaps due to the formation of side products during reflux that could interfere with the hydrogen bond network in the *ee*-determining step. When the reaction was monitored in d^8-toluene, no significant difference in regioselectivity was observed whether the imine was pre-formed under reflux or if the reaction was run entirely at room temperature (**Table 6.2**). Given that imine preformation did not result in higher regio- or enantioselectivity, all subsequent catalytic Pictet-Spengler reactions were run at room temperature.

Table 6.2. Influence of reaction conditions on regioselectivity.
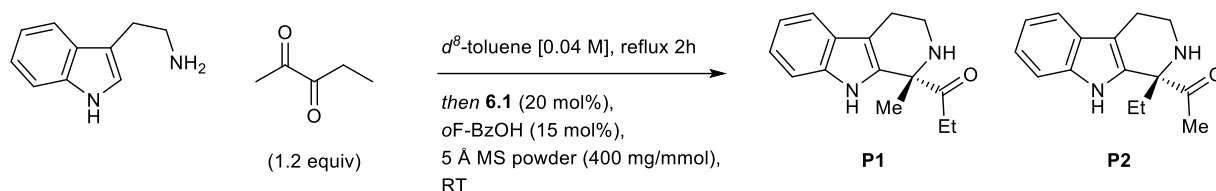
Entry	Deviations from the conditions above	P1:P2	ee [%]
1	None	1.00:0.37	n.d.
2	No reflux	1.00:0.33	n.d.

n.d. = not determined

6.3.2 Synthesis of Diketones

Given the relatively few commercially available diketones (limited to 2,3-butanedione, 2,3-pentanedione and 3,4-hexanedione), a small library of diketones was prepared in order to evaluate the substrate scope of the reaction. Synthetic methodology to access a variety of diketones was developed and applied to the synthesis of several symmetric and unsymmetric diketones. Previous results on similar substrates showed that bulky substituents (e.g. *i*Pr, *t*Bu, phenyl) at the α-position resulted in poor yields and lower enantioselectivities, so the substrate scope was limited to diketones with either a methyl or methylene group at this position.[36, 37] The substrate scope aimed to encompass diketones featuring a wide range of functional groups in order to assess their stability under the reaction conditions and their influence on enantioselectivity. Similarly, the synthesis of diketones of varying steric bulk were pursued to investigate their role in the observed enantioselectivity.

Initial attempts at symmetric diketone synthesis involved the reaction of an organocuprate (generated from the corresponding Grignard reagent) with oxalyl chloride (**Scheme 6.6**). However, this resulted in low product yield due to the suspected formation of the homocoupling product which is reported to form more readily with alkyl organocuprates.[39] Additionally, considering the low polarity of the product, finding conditions for chromatographic separation from other non-polar side products was challenging and therefore alternative methods to prepare the diketones were pursued.
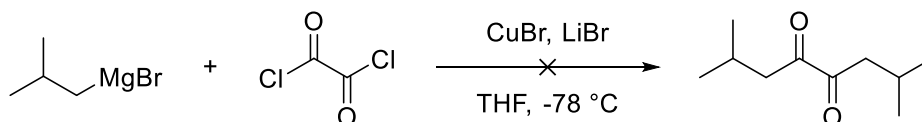
Scheme 6.6. Attempted synthesis of diketones via organocuprate addition to oxalyl chloride.

Instead, synthesis of several symmetric diketones was achieved via an acyloin condensation reaction of the corresponding aldehydes. As shown in **Scheme 6.7**, diketones **6.2**, **6.3**, **6.4**, and **6.5** were prepared in two steps via an acyloin condensation followed by oxidation of the resulting α-hydroxyketone by DMP. Acyloin condensation was also attempted with the bulkier *tert*-butylacetaldehyde, however no reaction was observed and only led to the recovery of starting material. In the case of diketone **6.5**, the starting aldehyde was also synthesized via oxidation of the commercially available alcohol (**Scheme 6.8**).
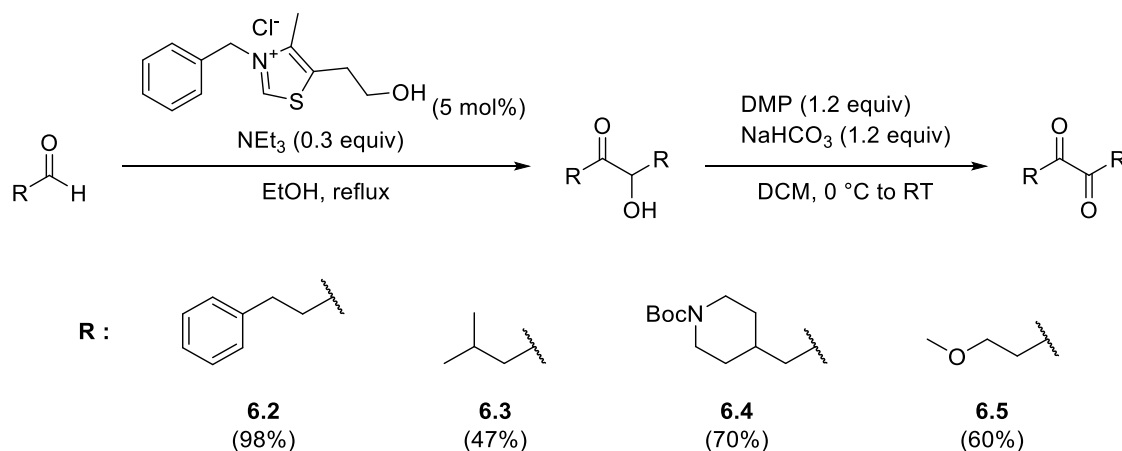
Scheme 6.7. Synthesis of symmetric diketones via an acyloin reaction and subsequent oxidation. Yields over two steps given in brackets.
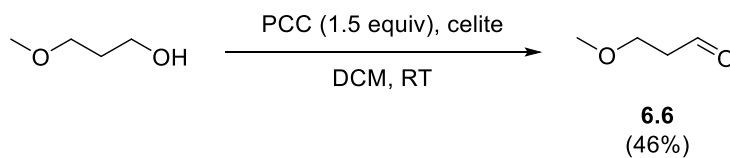
Scheme 6.8. Synthesis of aldehyde starting material **6.6** for diketone **6.5**.

Similarly, unsymmetric diketone **6.8** was prepared in two steps, starting with a crossed acyloin reaction using a 3:1 ratio of acetaldehyde to cinnamaldehyde in order to favour the crossed acyloin products (while also producing 1 equiv. of the homo-acyloin reaction of acetaldehyde).[40] As either aldehyde can react as the nucleophilic acyloin, a mixture of desired crossed acyloin products (**6.7a** and **6.7b**) was obtained. The resulting α-hydroxyketones were combined and subsequently oxidized to afford the corresponding diketone as a single product.
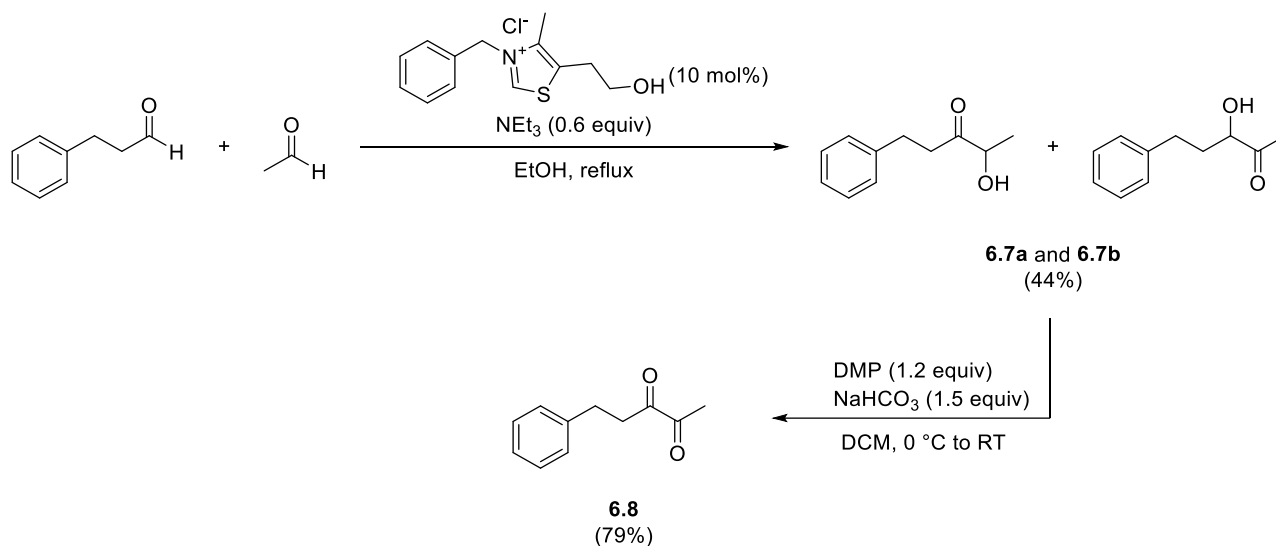
Scheme 6.9. Synthesis of **6.8** via cross-acyloin reaction and oxidation.

While acyloin condensation and subsequent oxidation were successful at obtaining several symmetric and one unsymmetric diketone in high yields, the methodology was limited by the availability of aldehyde starting materials. Therefore, an alternative approach to the synthesis of unsymmetric diketones was pursued via alkyne oxidation.

Several alkyne substrates were accessible from hex-5-yn-1-ol, starting with isomerization of the alkyne to afforded hex-4-yn-1-ol (**6.9**). This intermediate was then functionalized to introduce OTIPS, acetal, and nitrile functional groups (as shown in **Scheme 6.11**).
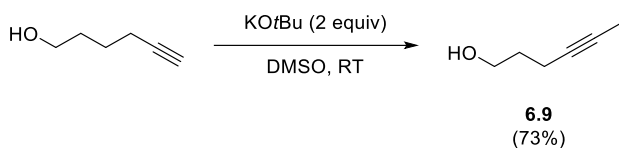
Scheme 6.10. Alkyne isomerization to access internal alkyne.
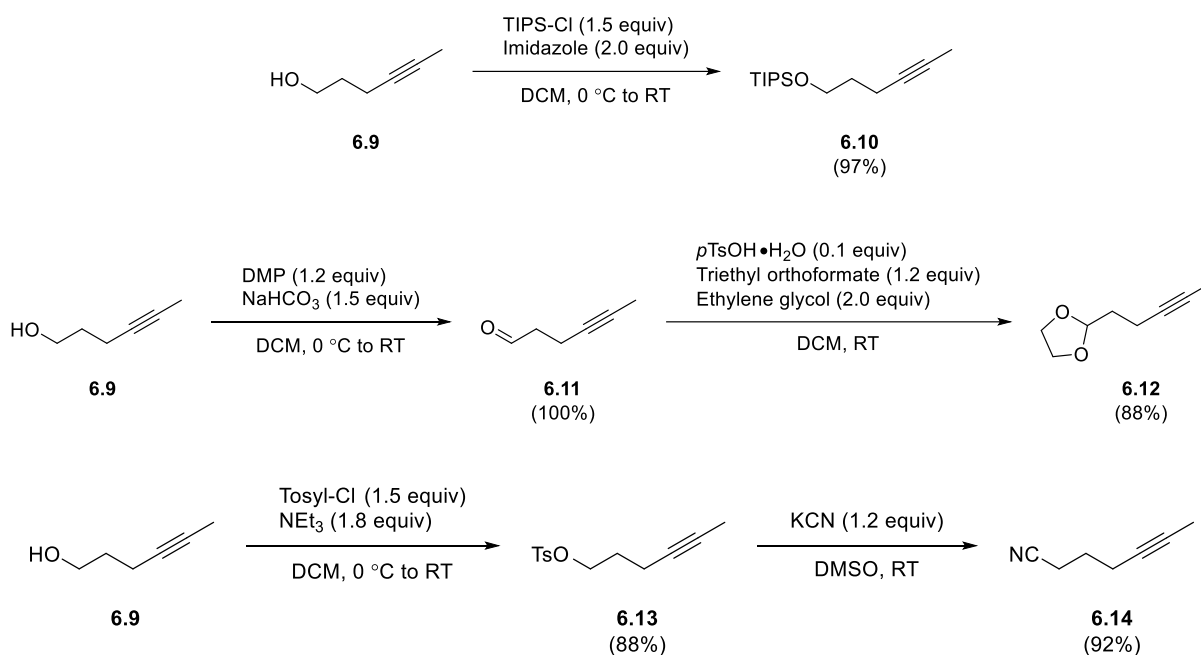
Scheme 6.11. Functionalization of **6.9** to incorporate TIPS (**6.10**), acetal (**6.12**), and nitrile

(**6.14**).

The unsymmetric diketones (**6.15**, **6.16**, and **6.17**) were then prepared by Ruthenium-catalysed oxidation of the corresponding alkyne (**Scheme 6.12**).

6.15
(98%)

6.16
(47%)

6.17
(60%)
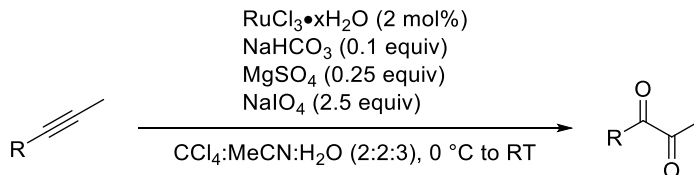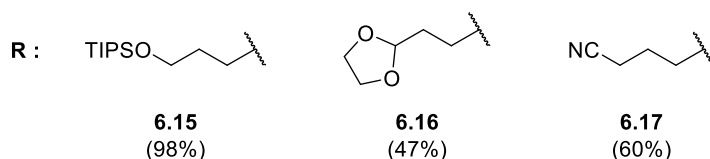
Scheme 6.12. Synthesis of unsymmetric diketones **6.15**, **6.16** and **6.17** by Ru-catalysed oxidation.

6.3.3 Synthesis and Characterization of Racemic Pictet-Spengler Products

In order to assess the enantioselectivity of the optimized reaction conditions on the diketone substrates, the racemic products were first synthesized via a racemic Pictet-Spengler reaction. For each diketone substrate, the racemic products were fully characterized and chromatographic conditions to separate the two enantiomers were screened.

6.18
(72%)

6.19
(96%)

6.20
(76%)
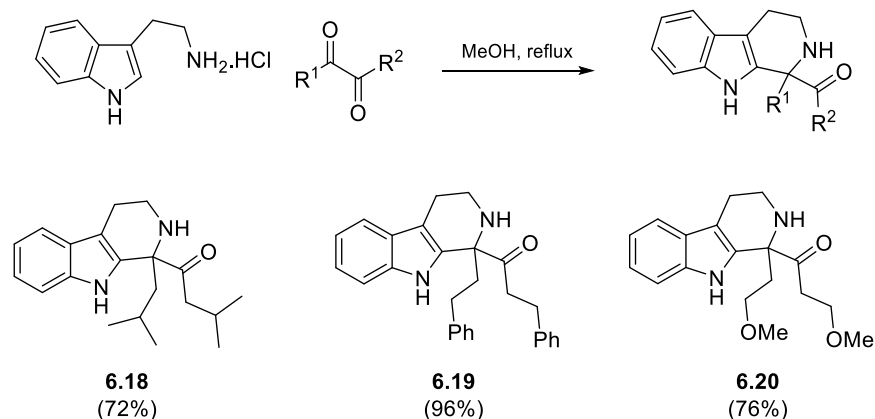
Scheme 6.13. Racemic Pictet-Spengler reactions.

For the substrates containing isopropyl, homobenzyl and methoxy-ethyl substituents, the racemic Pictet-Spengler reaction was run with tryptamine-HCl in refluxing methanol to afford the desired products in good yields (**Scheme 6.13**). For the Pictet-Spengler reaction of 2,3-butanedione and 3,4-hexanedione, tryptamine was used with stoichiometric *p*TsOH in MeOH (**Scheme 6.14**, conditions 2), while acid-sensitive substrates (*N*-Boc-piperidine, O-TIPS, acetal) were performed with stoichiometric benzoic acid in DCM (**Scheme 6.14**, conditions 3).
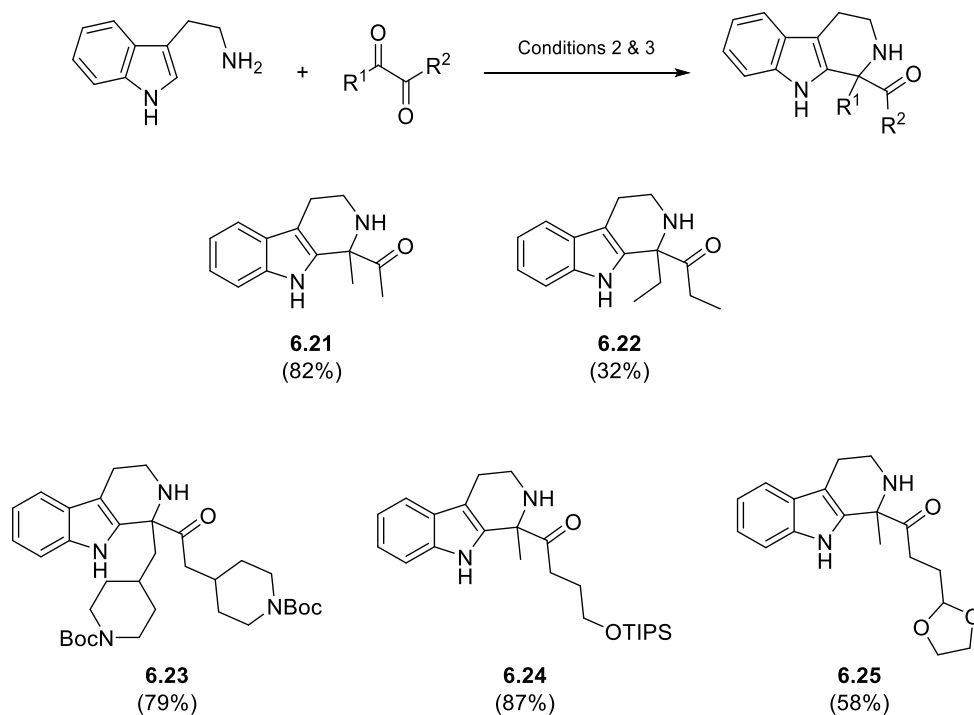
Scheme 6.14. Racemic Pictet-Spengler reaction with tryptamine.

The racemic products were analyzed by SFC in order to determine the adequate column and mobile phase for separation. In most cases, separation was obtained using an isocratic mobile phase of between 5 and 20% MeOH in supercritical CO_2, eluting for up to 10 minutes. In the case of **6.21, 6.22, 6.24** and **6.25**, 1% Et_3N was required as a mobile phase additive in order to achieve adequate separation of enantiomers. An example chromatogram is shown in **Figure 6.3**, while the rest are supplied in the supporting information for this chapter.

Signal:	MWD1A,Sig=270,4 Ref=360,100					
RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.827	MM m	0.38	767.41	118.34	50.53	
3.070	MM m	1.43	751.34	18.59	49.47	
		Sum	1518.75			
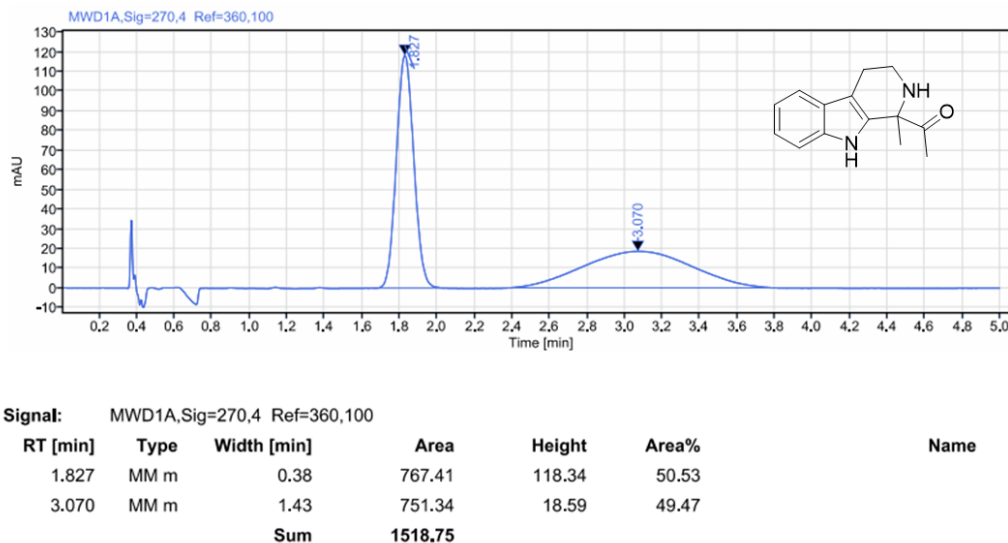
Figure 6.3. SFC chromatogram for **6.21**, eluting with 19% (1% NEt$_3$ / MeOH) in supercritical CO$_2$. (See SI for other chromatograms).

6.3.4 Synthesis and Evaluation of Enantioselective Pictet-Spengler Products

The optimized conditions for the enantioselective reaction were employed with the prepared diketone substrates as summarized in **Scheme 6.15**.

6.26
90% yield, 76% *ee*

6.27
88% yield, 94% *ee*

6.28
42% yield, 96% *ee* (RT)
87% yield, 95% *ee* (40 °C)

6.29
72% yield, 92% *ee*

6.30
51% yield, 90% *ee*

6.31
28% yield, 96% *ee* (RT)
70% yield, 93% *ee* (40 °C)

6.32
71% yield, 80% *ee*

6.33
68% yield, 75% *ee*
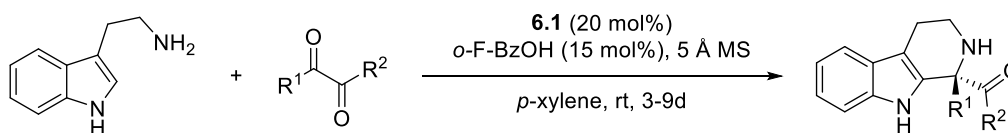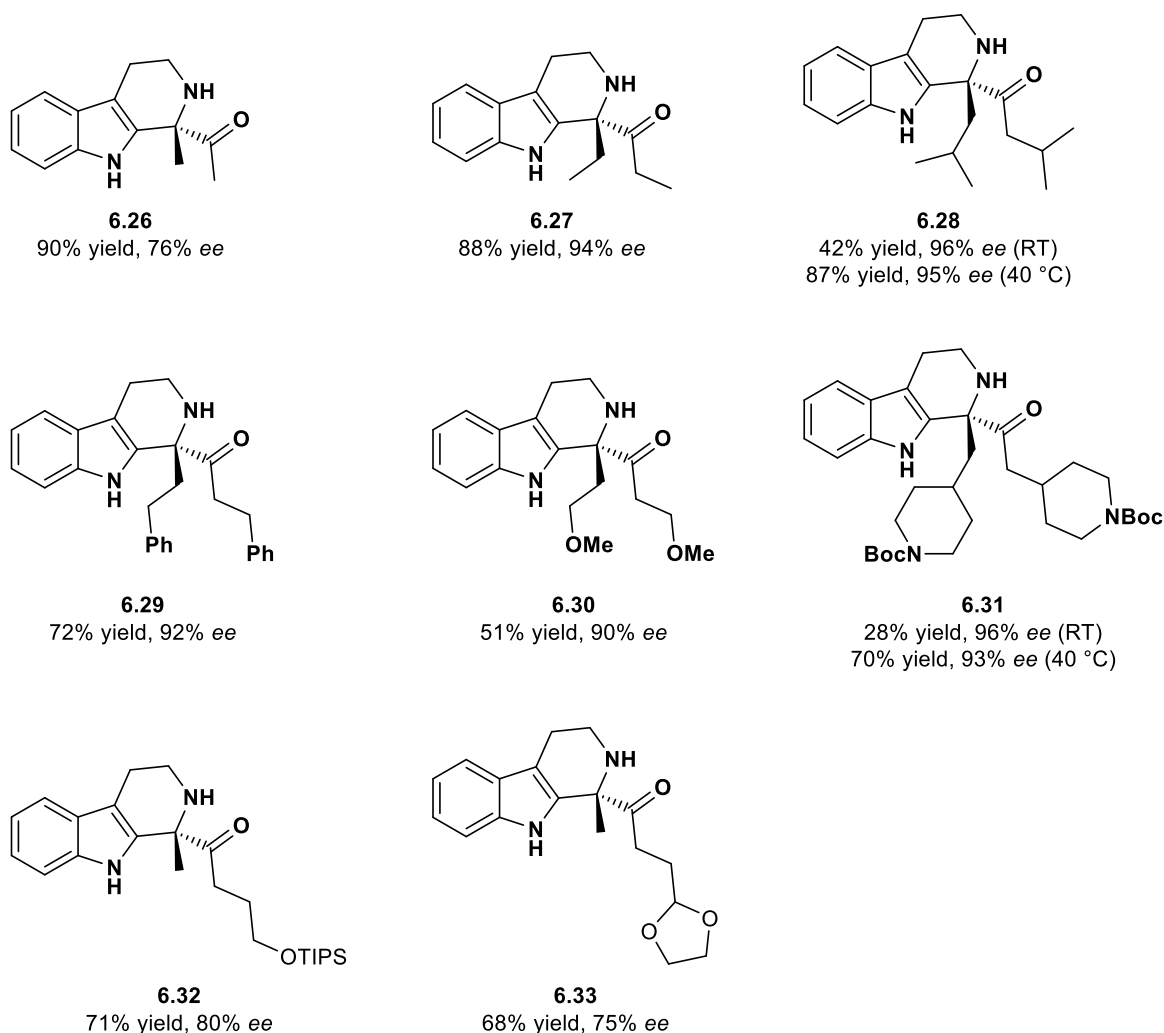
Scheme 6.15. Substrate screening of the asymmetric organocatalytic PS reaction.

Under the optimized reaction conditions, the desired products were obtained in moderate yields and good enantioselectivity. In the case of the *i*Pr (**6.28**) and *N*-Boc-piperidine (**6.31**) substituents, higher temperatures were employed to obtain the products in higher yields while still maintaining high enantioselectivity at 40 °C. A variety of functional groups were well tolerated under the reaction conditions, including ethers (**6.30**), TIPS-protected alcohol (**6.32**), acetal (**6.33**), and *N*-Boc group (**6.31**). The reaction conditions also tolerated the presence of a methoxy group at the beta position (**6.30**) without observing the formation of any elimination product.

The enantioselectivity appears to increase with increasing steric bulk of the diketone, with 2,3-butanedione (**6.26**) resulting in the lowest enantioselectivity. Similarly, the presence of a methyl

group at the R^1 position regardless of the steric bulk of the group at the R^2 position, results in similar stereoselectivity. This suggests that the R^1 group may play a greater role in enantioselectivity as compared to R^2. The higher enantioselectivity observed with 2,3-pentanedione during reaction optimization (see **Table 6.1**) may be a result of diminishing enantioselectivity at longer reaction times. The optimization reactions were analysed after 24h, regardless of conversion, while the substrate scope reactions were monitored for tryptamine conversion and worked up either when full conversion was observed or when conversion appeared to have plateaued. Jacobsen observed product inhibition with a similar thiourea/benzoic acid co-catalyst system.[11] He proposed that the basicity of the tryptamine product competes with the imine for ion pairing with the co-catalyst system, leading to decreasing reaction rate over time and slow catalyst turnover. A similar effect may be occurring under our reaction conditions, explaining the long reaction times (3-9d), even with relatively high catalyst loading. This may also contribute to the lower enantioselectivity if it causes the background racemic reaction to occur at a faster rate relative to the catalysed reaction. It may be worthwhile to pursue conditions which involve the pre-formation of the imine, followed by addition of Boc$_2$O along with the catalyst system in order to protect the amine product and avoid catalyst inhibition.

6.4   Conclusion and Future Work

This chapter detailed the optimization and evaluation of an organocatalytic enantioselective Pictet-Spengler reaction with diketone substrates. Using a previously identified urea-based catalyst, the preliminary conditions were optimized by screening different acid co-catalysts. The use of *ortho*-fluorobenzoic acid resulted in a 10% increase in enantioselectivity, emphasizing the critical role it plays in this catalytic system. A solvent screen was also performed, with improved enantioselectivites observed with less polar solvents (e.g. xylene, *n*-hexane), as expected for ion-pairing catalysis. When evaluated on various diketone substrates, the optimized reaction conditions achieved high yields (51-90%) and high enantioselectivities (75-95%). The reaction could also tolerate a range of different functional groups and was performed as a simple one-pot procedure, highlighting the ease and versatility of this method towards the synthesis of 1,1-disubstituted TBHC scaffolds.

Future experimental work is currently ongoing by students in the Zhu lab to expand on the substrate scope. For example, while diketone **6.17** was synthesized, it has not yet been evaluated in the

Pictet-Spengler reaction. Additionally, the unsymmetric diketones were all limited to those with a R^1 methyl group and should be further explored to include groups of varying size. Further experimental work could also include a more thorough evaluation of how reaction conditions influence the observed regioselectivity. As shown in **Table 6.*1*** and **Table 6.*2***, the focus of reaction optimization was primarily on improving stereoselectivity. However, improved yields may be obtained for unsymmetric diketones by improving regioselectivity.

While the experimental work in this chapter was performed in the Zhu lab, it could be applicable to computational work in the Moitessier research group. We recently reported on the development of Virtual Chemist, a computational platform that aims to predict the stereoselectivity of asymmetric transformations.[41] Future work for this project could involve a computational investigation into the mechanism of this reaction and its enantioselectivity-determining transition state, using the experimental data obtained in this chapter. Once a transition state has been proposed, it could be implemented into Virtual Chemist, allowing different catalysts and substrates for this transformation to be evaluated *in silico*.

Jacobsen developed an organocatalytic enantioselective Pictet-Spengler reaction of aldehydes using a similar carboxylic acid/thiourea co-catalysis[28] and later investigated the mechanism with both experimental and computational studies.[11] A primary kinetic isotope effect (KIE) was observed at the tryptamine C2 position, implying that the rate-determining and enantioselectivity-determining step of the reaction is rearomatization. Computational modelling predicted transition state structures for this step where CH-π interactions stabilize the transition state leading to the major enantiomer (**Figure 6.4**).
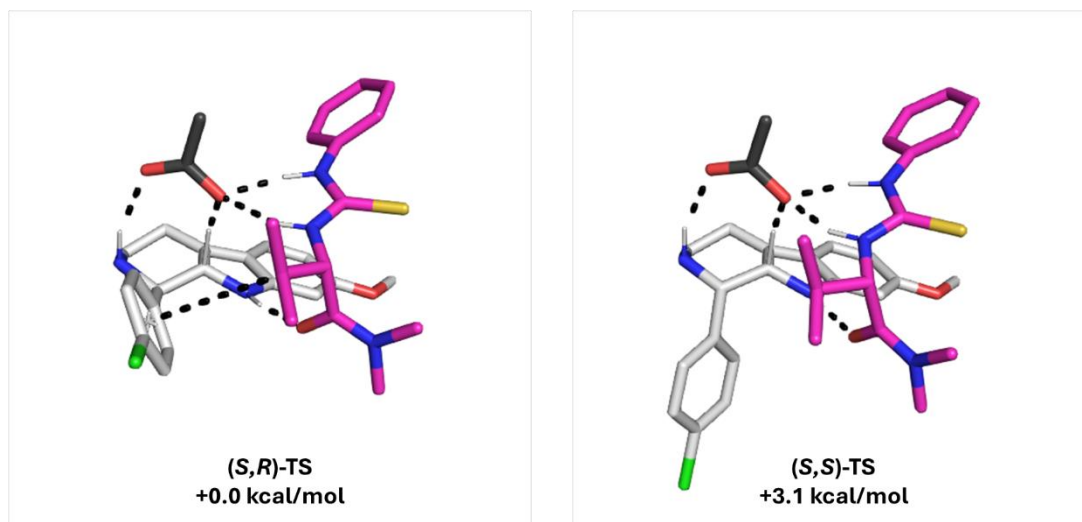
Figure 6.4. Proposed transition state from Jacobsen.

Andres *et al.* observed similar results in kinetic experiments with ketoamide substrates, which showed a KIE of 2.3 at the tryptamine C2 position.[36] Given the similarity in catalysts and the observed KIE, it is likely that rearomatization is also the enantioselectivity-determining step with catalyst **6.1** and the diketone substrates. With this in mind, work is currently underway to obtain accurate parameters to model the carbocation in the enantioselectivity-determining transition state.

6.5 Experimental

6.5.1 General

Reagents and solvents were purchased from commercial sources and preserved under argon. More sensitive compounds were stored in a desiccator or in the glovebox if required. Reagents were used as received without further purification. All reactions were performed under argon (or nitrogen) and stirring unless otherwise noted. When needed, glassware was dried at least overnight in an oven (170 °C) or under vacuum with a heat gun (650 °C). Solvents indicated as dry were either purchased as such or distilled prior to use. Flash column chromatography was performed using Silicycle SiliaFlash® P60 230-400 mesh. Reactions were monitored using Merck Kieselgel 60F$_{254}$ aluminum or glass backed plates. TLCs were revealed by UV fluorescence (254 nm) then one of the following: KMnO$_4$, phosphomolybdic acid, ninhydrin, *p*-anisaldehyde, vanillin.

NMR spectra were recorded on a Brüker AvanceIII-400 spectrometer at room temperature. IR

spectra were recorded in a Jasco FT/IR-4100 spectrometer outfitted with a PIKE technology MIRacle[TM] ATR accessory as neat films compressed onto a Zinc Selenide window or a Perkin Elmer Spectrum BX FT-IR. The spectra are reported in cm^{-1}. Abbreviations used are: w (weak), m (medium), s (strong) and br (broad). Mass spectra were determined with a Waters ACQUITY H-class UPLC/MS ACQ-SQD by electron ionization (EI positive and negative) or a Finnigan TSQ7000 by electrospray ionization (ESI$^+$). The accurate masses were done by the mass spectrometry service of EPFL by ESI-TOF using a QTOF Ultima from Waters. Melting points were determined using a Stuart SMP30 or Büchi B-540. Optical rotations α_D were measured with a Jasco P-2000 polarimeter (589 nm). Enantiomeric excesses were determined with a 1260 Infinity II SFC System from Agilent using chiral stationary phase columns by comparing the samples with the corresponding racemic samples, column and elution details specified in each entry.

General procedure 1 for the synthesis of diketones 6.2, 6.3, 6.4, and 6.5. To a suspension of 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride (5 mol%) in EtOH [1 M] was added aldehyde (1.0 equiv) and NEt$_3$ (0.3 equiv). The reaction was heated to reflux and stirred overnight. The reaction mixture was then partitioned between H$_2$O and AcOEt and the phases were separated. The aqueous layer was extracted 2 times with AcOEt, the combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and evaporated *in vacuo*. The crude products were used in the subsequent step without further purification.

The crude hydroxyketone was dissolved in DCM [0.3 M] and anhyd. NaHCO$_3$ (1.2 equiv) was added. The reaction was cooled to 0 °C and DMP (1.2 equiv) was added portionwise. The reaction stirred at 0°C for 1 h then warmed to room temperature over 30 minutes. The reaction was diluted with PE and filtered over celite. The filtrate was quenched with sat. NaHCO$_3$ and Na$_2$S$_2$O$_3$ then extracted three times with AcOEt. The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and evaporated *in vacuo*. The crude product was then purified by flash column chromatography (PE/Acetone 9/1 or PE/Et$_2$O 98/2, depending on product polarity) to afford the desired diketones.

General procedure 2 for the synthesis of diketones 6.15, 6.16, and 6.17. To a solution of alkyne (1.0 equiv) in CCl$_4$:MeCN:H$_2$O [0.25 M] was added NaHCO$_3$ (0.1 equiv), MgSO$_4$ (0.25 equiv) and RuCl$_3$•xH$_2$O (2 mol%). The reaction was cooled to 0 °C and NaIO$_4$ (2.5 equiv) was added

portionwise. The reaction was warmed to room temperature and stirred for 2 h. The reaction was diluted with sat. NaHCO$_3$ and extracted three times with DCM. The combined organic layers were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered, and evaporated *in vacuo*. The crude diketones were then purified by flash column chromatography (2-20% Et$_2$O/pentane or 100% DCM, depending on product polarity).

General procedure 3 for the racemic Pictet-Spengler reaction: A solution of tryptamine hydrochloride (1.0 equiv) and the α-diketone (1.1 equiv) in MeOH [0.25 M] was stirred under reflux overnight. An additional portion of the α-diketone derivative (0.9 equiv) was added and the reaction stirred at reflux for 24 h. The reaction was then quenched with a saturated aqueous NaHCO$_3$ solution, diluted with AcOEt and the phases were separated. The aqueous layer was extracted 3 times with AcOEt, and the combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and evaporated *in vacuo*. The crude product was then purified by flash column chromatography (DCM + 1% (4% NH$_3$/MeOH) to afford the pure desired product.

General procedure 4 for the racemic Pictet-Spengler. To a solution of tryptamine (1.0 equiv) in MeOH [0.25M] was added the α-diketone derivative (1.1 equiv) and *p*TsOH (1.0 equiv.) and the reaction stirred at reflux overnight. An additional portion of the α-diketone derivative (0.9 equiv.) was added and the reaction stirred at reflux for 16 h. The reaction was quenched with a saturated aqueous NaHCO$_3$ solution and extracted three times with DCM. The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The crude product was then purified by flash chromatography (PE/Acetone 10/1 to 8/2 or DCM + 4% (4% NH$_3$/MeOH) depending on the product polarity) to afford the pure desired product.

General procedure 5 for the racemic Pictet-Spengler reaction. A solution of the tryptamine (1.0 equiv), the α-diketone derivative (1.2 equiv) and BzOH (1.0 equiv) in DCE [0.25 M] was stirred under reflux until the reaction was complete. The reaction was then quenched with a saturated aqueous Na$_2$CO$_3$ solution, diluted with AcOEt and the phases were separated. The aqueous layer was extracted 3 times with AcOEt, and the combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered and evaporated under vacuo. The crude product was then purified by flash column chromatography (DCM/acetone 20/1 to 6/4) to afford the pure desired product.

General procedure 6 for the catalytic enantioselective Pictet-Spengler reaction. A vial was

charged with oF-BzOH (15 mol%), tryptamine (1.0 equiv), catalyst (20 mol%) and 5 Å MS powder (400 mg/mmol). *P*-xylene [0.04 M] was added, followed by the α-diketone derivative (1.2 equiv) and the reaction stirred at the indicated temperature until the reaction was complete. The reaction mixture was then directly purified by flash column chromatography (DCM/Acetone 20/1 to 6/4 depending on the product polarity) to afford the pure desired product.

6.5.2   Synthesis and Characterization Data

1,6-diphenylhexane-3,4-dione (6.2). Compound 6.2 (1.12 g, 4.2 mmol, 98% yield) was prepared according to general procedure 1. All characterization data were in full agreement with those reported in the literature.[42]

2,7-dimethyloctane-4,5-dione (6.3). Compound 6.3 (883 mg, 5.2 mmol, 47% yield) was prepared according to general procedure 1. Yellow oil. Rf = 0.33 (99/1 PE/Et$_2$O). ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.61 (d, J = 6.9 Hz, 4H), 2.11 (dh, J = 13.5, 6.7 Hz, 2H), 0.93 (d, J = 6.7 Hz, 12H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 200.08, 44.74, 24.40, 22.71. HRMS (EI/QqQ) m/z: [M]+ Calcd for C$_{10}$H$_{18}$O$_2^+$ 170.1307; Found 170.1571. IR (v_{max}, cm^{-1}) 2960 (m), 2874 (w), 1710 (s), 1467 (m), 1369 (m), 1035 (m), 757 (m)

di-*tert*-butyl 4,4'-(2,3-dioxobutane-1,4-diyl)bis(piperidine-1-carboxylate (6.4). Compound 6.4 (1.04 g, 2.3 mmol, 70% yield) was prepared according to general procedure 1. Yellow powder. Rf = 0.29 (9/1 PE/Acetone). M.p. = 113.1 – 115.7 °C. ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.06 (s, 2H), 2.74 (t, J = 12.8 Hz, 2H), 2.67 (d, J = 6.7 Hz, 2H), 1.96 (dtq, J = 14.9, 6.9, 3.5 Hz, 1H), 1.63 (d, J = 12.8 Hz, 2H), 1.45 (s, 9H), 1.15 (qd, J = 12.3, 4.4 Hz, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 198.78, 154.92, 79.58, 42.33, 32.05, 31.71, 28.60. HRMS (nanochip-ESI/LTQ-Orbitrap) m/z: [M + H]+ Calcd for C$_{24}$H$_{41}$N$_2$O$_6^+$ 453.2959; Found 453.2960. IR (v_{max}, cm^{-1}) 2975 (w), 2925 (w), 2852 (w), 1691 (s), 1475 (s), 1422 (m), 1365 (m), 1312 (m), 1281 (m), 1244 (m), 1126 (s), 1127 (w)

1,6-dimethoxyhexane-3,4-dione (**6.5**). Compound 6.5 (480 mg, 2.76 mmol, 60 % yield) was prepared according to general procedure 1. Yellow oil. Rf = 0.34 (9/1 PE/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 3.68 (t, J = 6.1 Hz, 4H), 3.32 (s, 6H), 2.99 (t, J = 6.1 Hz, 4H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 197.81, 66.99, 58.94, 36.52. HRMS (ESI/QTOF) m/z: [M + Na]$^+$ Calcd for

$C_8H_{14}NaO_4^+$ 197.0784; Found 197.0785. IR (ν_{max}, cm^{-1}) 2983 (w), 2928 (w), 2892 (w), 1713 (s), 1454 (w), 1388 (m), 1193 (w), 1113 (s), 969 (w), 690 (w)

3-methoxypropanal (6.6). To a suspension of PCC (13.5 g, 62.7 mmol, 1.5 equiv) and celite (13.5 g) in DCM (80 mL) was added 3-methoxy-1-propanol (4.0 mL, 41.8 mmol, 1.0 equiv), and the reaction stirred at room temperature for 1 h. The crude reaction mixture was filtered through a silica plug and rinsed with DCM. The filtrate was concentrated *in vacuo* to afford the desired aldehyde (1.7 g, 19.3 mmol, 46% yield) as a clear oil that was used in the subsequent step without further purification. All characterization data were in full agreement with those reported in the literature.[43]

4-hydroxy-1-phenylpentan-3-one (6.7a) and 3-hydroxy-5-phenylpentan-2-one (6.7b). To a solution of hydrocinnamaldehyde (2.5 mL, 19.0 mmol, 1.0 equiv) and acetaldehyde (3.2 mL, 57.0 mmol, 3.0 equiv) in EtOH (100 mL) was added 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride (509 mg, 10 mol%) and NEt$_3$ (1.6 mL, 11.4 mmol, 0.6 equiv), and the reaction stirred at reflux overnight. The reaction was diluted with dH$_2$O and extracted three times with DCM. The combined organics were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The reaction was further purified by column chromatography (8:2 PE:EtOAc) to afford the desired product (1.50 g, 8.4 mmol, 44% yield) as a ~1:1 mixture of regioisomers. All characterization data were in full agreement with those reported in the literature.[44, 45]

5-phenylpentane-2,3-dione (6.8). The mixture of hydroxy-ketones **6.7a** and **6.7b** (1.40 g, 7.9 mmol, 1.0 equiv) was dissolved in DCM (20 mL) and cooled to 0 °C. NaHCO$_3$ (996 mg, 11.9 mmol, 1.5 equiv) was added followed by DMP (4.0 g, 9.4 mmol, 1.2 equiv) and the reaction was warmed to room temperature and stirred for 1 h. The reaction was diluted with Et$_2$O and filtered through celite. The filtrate was quenched with sat. NaHCO$_3$ and sat. Na$_2$S$_2$O$_3$ and extracted three times with Et$_2$O. The combined organics were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo* to afford the pure diketone (1.10 g, 6.2 mmol, 79% yield) as a yellow oil. All characterization data were in full agreement with those reported in the literature.[46]

Hex-4-yn-1-ol (6.9). To a solution of hex-5-yn-1-ol (10.0 mL, 90.7 mmol, 1.0 equiv) in DMSO

(150 mL) was added KO*t*Bu (20.4 g, 181.4 mmol, 2.0 equiv) and the reaction stirred at room temperature for 3 h. The reaction was quenched with 2 M HCl and extracted three times with Et_2O. The combined organic layers were washed with sat. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo*. The product was further purified by flash column chromatography (4/1 PE/EtOAc) to afford the desired product (6.5 g, 66.2 mmol, 73% yield) as a clear oil. All characterization data were in full agreement with those reported in the literature.[47]

(hex-4-yn-1-yloxy)triisopropylsilane (6.10). To a solution of hex-4-yn-1-ol **6.9** (1.47 g, 15.0 mmol, 1.0 equiv) in DCM (30 mL) was added imidazole (2.0 g, 30.0 mmol, 2.0 equiv). The reaction was cooled to 0 °C and TIPS-Cl (4.8 mL, 22.5 mmol, 1.5 equiv) was added. The reaction was warmed to room temperature and stirred for 6 h. The reaction was concentrated *in vacuo* then diluted with dH_2O and extracted three times with Et_2O. The combined organic layers were with sat. NH_4Cl, dH_2O and brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the desired product (3.8 g, 14.9 mmol, 97% yield) as a clear oil. The product was carried to the subsequent step without any further purification, and all characterization data were in full agreement with those reported in the literature.[48]

hex-4-ynal (6.11). To a solution of hex-4-yn-1-ol (1.97 g, 20.0 mmol, 1.0 equiv) in DCM (100 mL) was added $NaHCO_3$ (2.50 g, 30.0 mmol, 1.5 equiv) and the reaction was cooled to 0 °C. DMP (10.2 g, 24.0 mmol, 1.2 equiv) was added portionwise and the reaction was slowly warmed to room temperature, stirring for 1 h. The reaction was quenched with sat. $NaHCO_3$ and sat. $Na_2S_2O_3$, and the aqueous phase was extracted three times with DCM. The combined organic layers were washed with sat. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the desired product (1.9 g, 20.0 mmol, 100% yield) as a clear oil that was used in the subsequent step without further purification.

2-(pent-3-yn-1-yl)-1,3-dioxolane (6.12). To a solution of aldehyde **6.11** (1.9 g, 20.0 mmol, 1.0 equiv) in DCM (100 mL) was added ethylene glycol (2.2 mL, 40.0 mmol, 2.0 equiv), triethyl orthoformate (4.0 mL, 24.0 mmol, 1.2 equiv) and *p*TsOH•H_2O (380 mg, 2.0 mmol, 0.1 equiv), and the reaction stirred at room temperature for 2 h. The reaction was quenched with sat. $NaHCO_3$ and extracted three times with DCM. The combined organic layers were washed with brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the desired acetal (2.46 g, 17.5 mmol,

88% yield) as a clear oil. All characterization data was in full agreement with those reported in the literature.[49]

hex-4-yn-1-yl 4-methylbenzenesulfonate (6.13). To a solution of hex-4-yn-1-ol (5.0 g, 50.0 mmol, 1.0 equiv) in DCM (200 mL) a t0 °C was added tosyl chloride (14.3 g, 75.0 mmol, 1.5 equiv) followed by NEt$_3$ (12.5 mL, 90.0 mmol, 1.8 equiv) dropwise and the reaction stirred at room temperature overnight. The reaction was diluted with dH$_2$O and the aqueous phase was extracted three times with DCM. The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo*. The product was further purified by flash column chromatography (8/2 PE/Et$_2$O) to afford the desired product (11.1 g, 44.0 mmol, 88% yield) as a clear oil. All characterization data was in full agreement with those reported in the literature.[50]

hept-5-ynenitrile (6.14). To a solution of tosylate **6.13** (5.0 g, 19.8 mmol, 1.0 equiv) in DMSO was added KCN (1.56 g, 24.0 mmol, 1.2 equiv) and the reaction stirred at room temperature overnight. The reaction was quenched with sat. Na$_2$CO$_3$ and extracted three times with Et$_2$O. The combined organic layers were washed with sat. NaHCO$_3$ and brine, dried over Na$_2$SO$_4$, filtered, and concentrated *in vacuo* to afford the desired nitrile (1.96 g, 18.3 mmol, 92% yield) as a clear oil that was used in the subsequent step without further purification.

6-((tri-isopropylsilyl)oxy)hexane-2,3-dione (6.15). Compound 6.15 (840 mg, 2.9 mmol, 21% yield) was prepared according to general procedure 2 and purified by column chromatography (2% Et$_2$O/pentane). Yellow oil. Rf = 0.82 (9/1 PE/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 3.71 (t, J = 6.0 Hz, 2H), 2.85 (t, J = 7.1 Hz, 2H), 2.33 (s, 3H), 1.88 – 1.79 (m, 2H), 1.15 – 0.95 (m, 21H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 199.49, 197.58, 62.36, 32.58, 26.65, 23.89, 18.13, 12.09. HRMS (ESI/QTOF) m/z: [M + Na]$^+$ Calcd for C$_{15}$H$_{30}$NaO$_3$Si$^+$ 309.1856; Found 309.1859. IR (v_{max}, cm^{-1}) 2942 (s), 2892 (w), 2866 (s), 1714 (s), 1463 (m) 1384 (w), 1354 (m), 1106 (s), 1068 (m), 996 (w), 946 (m), 920 (w), 881 (s), 782 (m), 679 (s), 658 (m)

5-(1,3-dioxolan-2-yl)pentane-2,3-dione (6.16). Compound 6.16 (590 mg, 3.4 mmol, 34% yield) was prepared according to general procedure 2 and purified by column chromatography (8:2 pentane:Et$_2$O). Yellow oil. Rf = 0.37 (9/1 PE/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.94 (t, J = 3.6 Hz, 1H), 3.93 – 3.76 (m, 4H), 2.79 (t, J = 6.9 Hz, 2H), 2.31 (s, 3H), 2.08 (td, J = 6.9, 3.6 Hz, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 198.96, 197.64, 102.95, 65.14, 29.71, 28.25, 23.89. HRMS

(Sicrit plasma/LTQ-Orbitrap) m/z: [M + H$_{-1}$]$^+$ Calcd for C$_8$H$_{11}$O$_4^+$ 171.0652; Found 171.0648. IR (ν$_{max}$, cm^{-1}) 2957 (w), 2889 (w), 1712 (s), 1412 (w), 1353 (m), 1137 (s), 1080 (m), 1033 (m), 944 (w), 884 (w)

5,6-dioxoheptanenitrile (6.17). Compound 6.17 (1.02 g, 7.3 mmol, 52% yield) was prepared according to general procedure 2 and purified by column chromatography (100% DCM). Yellow oil. *Rf* = 0.19 (8/2 PE/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.95 (t, *J* = 7.0 Hz, 2H), 2.45 (t, *J* = 7.1 Hz, 2H), 2.36 (s, 3H), 1.98 (p, *J* = 7.0 Hz, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 197.47, 196.79, 119.02, 34.21, 23.69, 19.10, 16.62. HRMS (Sicrit plasma/LTQ-Orbitrap) m/z: [M + H]$^+$ Calcd for C$_7$H$_{10}$NO$_2^+$ 140.0706; Found 140.0701. IR (ν$_{max}$, cm^{-1}) 2359 (m), 2342 (m), 1711 (s), 1423 (w), 1355 (m), 1079 (w), 1041 (w), 584 (w), 536 (w)

1-(1-isobutyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-3-methylbutan-1-one (6.18). Compound 6.18 (336 mg, 1.08 mmol, 72% yield) was prepared according to general procedure 3. Brown oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.31 (s, 1H), 7.49 (dd, *J* = 7.8, 1.1 Hz, 1H), 7.34 (dt, *J* = 8.0, 1.0 Hz, 1H), 7.16 (ddd, *J* = 8.2, 7.0, 1.3 Hz, 1H), 7.08 (ddd, *J* = 8.1, 7.1, 1.1 Hz, 1H), 3.32 (dt, *J* = 13.0, 4.5 Hz, 1H), 2.97 (ddd, *J* = 13.0, 8.3, 4.7 Hz, 1H), 2.84 – 2.66 (m, 3H), 2.61 (dd, *J* = 18.3, 6.3 Hz, 1H), 2.10 (dp, *J* = 13.4, 6.7 Hz, 1H), 1.95 (dd, *J* = 14.2, 6.1 Hz, 1H), 1.79 (dd, *J* = 14.2, 6.3 Hz, 1H), 1.69 (dp, *J* = 13.0, 6.5 Hz, 1H), 0.90 (t, *J* = 6.9 Hz, 6H), 0.85 (d, *J* = 6.7 Hz, 3H), 0.77 (d, *J* = 6.6 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 215.61, 136.21, 133.30, 127.13, 121.94, 119.38, 118.31, 111.19, 110.80, 67.15, 47.49, 46.90, 41.64, 24.61, 24.55, 23.92, 23.81, 22.91, 22.83, 22.72. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for C$_{20}$H$_{29}$N$_2$O$^+$ 313.2274; Found 313.2284. IR (ν$_{max}$, cm^{-1}) 3395 (m), 2955 (s), 2870 (m), 2841 (w), 1699 (s), 1464 (s), 1366 (m), 1298 (m), 1165 (m), 1142 (m), 741 (s)

1-(1-phenethyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-3-phenylpropan-1-one (6.19). Compound 6.19 (393 mg, 0.96 mmol, 96% yield) was prepared according to general procedure 3. Brown oil. *Rf* = 0.54 (4/1 PE/EtOAc). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.28 (s, 1H), 7.51 (dd, *J* = 7.8, 1.1 Hz, 1H), 7.38 – 7.32 (m, 1H), 7.31 – 7.02 (m, 12H), 3.38 – 3.17 (m, 2H), 3.11 – 2.91 (m, 2H), 2.89 – 2.81 (m, 2H), 2.78 – 2.68 (m, 2H), 2.48 (ddd, *J* = 9.2, 6.3, 2.2 Hz, 2H), 2.27 (ddd, *J* = 13.9, 10.1, 6.7 Hz, 1H), 2.13 (ddd, *J* = 13.8, 9.7, 6.9 Hz, 1H), 1.81 (s, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 214.75, 141.31, 141.21, 136.48, 132.42, 128.64, 128.59, 128.55, 128.44, 127.06, 126.27,

126.23, 122.22, 119.58, 118.45, 111.58, 111.31, 67.10, 41.63, 41.18, 40.29, 30.32, 29.77, 22.86. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for $C_{28}H_{29}N_2O^+$ 409.2274; Found 409.2282. IR (ν_{max}, cm^{-1}) 3410 (w), 3060 (w), 3026 (w), 2922 (m), 2843 (w), 1701 (s), 1603 (w), 1495 (m), 1454 (s), 1298 (m), 744 (s), 699 (s)

3-methoxy-1-(1-(2-methoxyethyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)propan-1-one (6.20). Compound 6.20 (19 mg, 0.06 mmol, 76% yield) was prepared according to general procedure 3. Brown oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.50 (s, 1H), 7.55 – 7.47 (m, 1H), 7.33 (dt, *J* = 8.1, 1.0 Hz, 1H), 7.16 (ddd, *J* = 8.1, 7.0, 1.3 Hz, 1H), 7.09 (ddd, *J* = 8.0, 7.1, 1.1 Hz, 1H), 3.61 (td, *J* = 6.3, 2.3 Hz, 2H), 3.45 – 3.08 (m, 11H), 3.03 – 2.90 (m, 2H), 2.81 – 2.63 (m, 2H), 2.33 (ddd, *J* = 14.7, 8.8, 4.6 Hz, 1H), 2.13 – 2.02 (m, 1H). ^{13}C NMR (101 MHz, CDCl$_3$ δ 214.57, 136.44, 132.41, 127.13, 122.10, 119.43, 118.44, 111.78, 111.27, 69.21, 67.78, 66.58, 58.95, 58.85, 41.44, 38.70, 38.28, 22.79. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for $C_{18}H_{25}N_2O_3^+$ 317.1860; Found 317.1853. IR (ν_{max}, cm^{-1}) 3336 (w), 2924 (m), 2892 (m), 1703 (m), 1454 (m), 1387 (m), 1298 (m), 1193 (m), 1110 (s), 1009 (w), 743 (m)

1-(1-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)ethan-1-one (6.21). Compound 6.21 (351 mg, 1.54 mmol, 82% yield) was prepared according to general procedure 4. All characterization data was in full agreement that reported in the literature.[35]

1-(1-ethyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)propan-1-one (6.22). Compound 6.22 (154 mg, 0.60 mmol, 32% yield) was prepared according to general procedure 4. Beige powder. *Rf* = 0.48 (PE/acetone 6/4). M.p. = 52.6 – 54.6 °C. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.35 (s, 1H), 7.49 (dd, *J* = 7.9, 1.1 Hz, 1H), 7.34 (dt, *J* = 8.1, 1.0 Hz, 1H), 7.16 (ddd, *J* = 8.2, 7.0, 1.3 Hz, 1H), 7.09 (ddd, *J* = 8.1, 7.1, 1.1 Hz, 1H), 3.34 (ddd, *J* = 13.0, 4.9, 3.8 Hz, 1H), 3.02 – 2.82 (m, 2H), 2.82 – 2.64 (m, 3H), 2.03 (dq, *J* = 14.7, 7.4 Hz, 1H), 1.96 – 1.84 (m, 1H), 1.76 (s, 2H), 1.02 (t, *J* = 7.2 Hz, 3H), 0.83 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 216.87, 136.35, 133.03, 127.03, 122.03, 119.43, 118.34, 111.23, 111.19, 67.15, 41.66, 32.71, 31.92, 22.97, 8.30, 7.66. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for $C_{16}H_{21}N_2O^+$ 257.1648; Found 257.1654. IR (ν_{max}, cm^{-1}) 3390 (w), 2970 (w), 2934 (w), 1698 (m), 1458 (m), 1298 (m), 1128 (m), 1087 (m), 1009 (m), 898 (m), 858 (m), 741 (s), 730 (s), 717 (s), 701 (m).

***tert*-butyl-4-((1-(2-(1-(*tert*-butoxycarbonyl)piperidin-4-yl)acetyl)-2,3,4,9-tetrahydro-1H-**

pyrido[3,4-b]indol-1-yl)methyl)piperidine-1-carboxylate (6.23). Compound 6.23 (94 mg, 0.16 mmol, 79% yield) was prepared according to general procedure 5. Beige foam. Rf = 0.52 (8/2 PE/Acetone). M.p. = 97.4 – 99.2 C. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.25 (s, 1H), 7.53 – 7.45 (m, 1H), 7.36 – 7.29 (m, 1H), 7.17 (ddd, J = 8.1, 7.0, 1.3 Hz, 1H), 7.09 (ddd, J = 8.1, 7.0, 1.1 Hz, 1H), 3.95 (d, J = 50.4 Hz, 5H), 3.30 (dt, J = 12.8, 4.5 Hz, 1H), 2.96 (ddd, J = 12.8, 8.1, 4.7 Hz, 1H), 2.87 – 2.43 (m, 10H), 2.03 – 1.74 (m, 3H), 1.69 – 1.48 (m, 5H), 1.43 (d, J = 5.4 Hz, 18H), 1.20 – 0.90 (m, 5H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 214.52, 154.95, 154.84, 136.25, 132.54, 127.08, 122.24, 119.60, 118.45, 111.29, 111.15, 79.50, 79.46, 66.82, 45.30, 44.51, 41.61, 33.78, 33.04, 32.30, 32.08, 31.45, 28.59, 28.58, 22.83. IR (v_{max}, cm^{-1}) 3325 (w), 2975 (w), 2922 (m), 2848 (w), 1688 (s),1675 (s), 1427 (s), 1366 (m), 1283 (m), 1247 (m), 1165 (s), 751 (m)

1-(1-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-4-((triisopropylsilyl)oxy)butan-1-one (6.24). Compound 6.24 (374 mg, 0.87 mmol, 87% yield) was prepared according to general procedure 5. Brown oil. Rf = 0.32 (20/1 DCM/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.25 (s, 1H), 7.49 (d, J = 7.8 Hz, 1H), 7.33 (d, J = 8.0 Hz, 1H), 7.16 (t, J = 7.5 Hz, 1H), 7.09 (t, J = 7.4 Hz, 1H), 3.67 (t, J = 6.1 Hz, 2H), 3.30 (dt, J = 13.0, 4.7 Hz, 1H), 2.99 (dq, J = 18.9, 6.8 Hz, 2H), 2.82 (dt, J = 18.8, 7.2 Hz, 1H), 2.77 – 2.70 (m, 2H), 1.82 – 1.71 (m, 2H), 1.56 (s, 3H), 1.08 – 0.97 (m, 21H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 215.36, 136.35, 134.13, 127.08, 122.12, 119.53, 118.45, 111.18, 110.19, 63.58, 62.42, 41.53, 33.77, 26.99, 26.58, 22.87, 18.20, 18.13, 12.09, 12.06. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for C$_{25}$H$_{41}$N$_2$O$_2$Si$^+$ 429.2932; Found 429.2926. IR (v_{max}, cm^{-1}) 3394 (w), 2942 (s), 2865 (s), 1703 (m), 1462 (m), 1366 (w), 1297 (m), 1108 (s), 883 (m), 741 (m), 682 (m)

3-(1,3-dioxolan-2-yl)-1-(1-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)propan-1-one (6.25). Compound 6.25 (181 mg, 0.58 mmol, 58 % yield) was prepared according to general procedure 5. Brown oil. Rf = 0.20 (7/3 DCM/Acetone). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.27 (s, 1H), 7.49 (dq, J = 7.0, 0.8 Hz, 1H), 7.34 (dt, J = 8.1, 1.0 Hz, 1H), 7.16 (ddd, J = 8.2, 7.1, 1.3 Hz, 1H), 7.09 (ddd, J = 8.0, 7.1, 1.0 Hz, 1H), 4.88 (t, J = 4.4 Hz, 1H), 3.96 – 3.86 (m, 2H), 3.85 – 3.76 (m, 2H), 3.30 (dt, J = 13.0, 4.8 Hz, 1H), 3.10 – 2.95 (m, 2H), 2.90 – 2.67 (m, 3H), 2.02 – 1.88 (m, 2H), 1.57 (s, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 214.48, 136.36, 134.05, 127.07, 122.16, 119.55, 118.47, 111.20, 110.23, 103.54, 65.05, 63.58, 41.52, 31.55, 27.61, 26.60, 22.88. HRMS (ESI/QTOF) m/z: [M + H]$^+$ Calcd for C$_{18}$H$_{23}$N$_2$O$_3$$^+$ 315.1703; Found 315.1711. IR (v_{max}, cm^{-1})

3387 (m), 3330 (m), 2933 (m), 2890 (m), 1708 (s), 1454 (m), 1366 (m), 1298 (m), 1137 (s), 1076 (m), 1010 (s), 944 (m), 745 (s)

(*S*)-1-(1-Methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)ethan-1-one (6.26).

Compound 6.26 (81 mg, mmol, 90% yield, 76% *ee*) was prepared according to general procedure 6 at room temperature for 72 h. All characterization data were in full agreement with those reported in the literature.[35] Brown oil. $[\alpha]_D^{26}$ = -12.8 (*c* 1.0, CHCl₃).

(*S*)-1-(1-Ethyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)propan-1-one (6.27).

Compound 6.27 (90 mg, 0.35 mmol, 88% yield, 94% *ee*) was prepared according to general procedure 6 at room temperature for 72 h. All characterization data were in full agreement with the racemic compound. Brown oil. $[\alpha]_D^{26}$ = -37.4 (*c* 1.0, CHCl₃).

(*S*)-1-(1-isobutyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-3-methylbutan-1-one (6.28).

Compound 6.28 (52 mg, 0.17 mmol, 42% yield, 96% *ee*) was prepared according to general procedure 6 at room temperature for 9 d. All characterization data were in full agreement with the racemic compound. Brown oil. $[\alpha]_D^{26}$ = -22.9 (*c* 1.0, CHCl₃). Compound 6.28 (109 mg, 0.35 mmol, 87% yield, 95% *ee*) was also prepared according to general procedure 6 at 40 ℃ for 7d. All characterization data were in full agreement with the racemic compound. Brown oil. $[\alpha]_D^{26}$ = -27.8 (*c* 1.0, CHCl₃).

(*S*)-1-(1-phenethyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-3-phenylpropan-1-one (6.29).

Compound 6.29 (117 mg, 0.29 mmol, 72% yield, 92% *ee*) was prepared according to general procedure 6 at room temperature for 5 d. All characterization data were in full agreement with the racemic compound. Brown oil. $[\alpha]_D^{26}$ = 6.4 (*c* 1.0, CHCl₃).

(*S*)-3-methoxy-1-(1-(2-methoxyethyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl) propan-1-one (6.30).

Compound 6.30 (65 mg, 0.21 mmol, 51% yield, 90% *ee*) was prepared according to general procedure 6 at room temperature for 5 d. All characterization data were in full agreement with the racemic compound. Brown oil. $[\alpha]_D^{26}$ = -35.2 (*c* 1.0, CHCl₃).

tert-butyl (S)-4-((1-(2-(1-(tert-butoxycarbonyl)piperidin-4-yl)acetyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)methyl)piperidine-1-carboxylate (6.31).

Compound 6.31 (67 mg, 0.11 mmol, 28% yield, 96% *ee*) was prepared according to general procedure 6 at room temperature for

9 d. All characterization data were in full agreement with the racemic compound. Compound 6.31 (167 mg, 0.28 mmol, 70% yield, 93% *ee*) was prepared according to general procedure 6 at 40 ℃ for 7 d. All characterization data were in full agreement with the racemic compound.

(S)-1-(1-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)-4-((triisopropylsilyl)oxy)butan-1-one (6.32). Brown oil. Compound 6.32 (151 mg, 0.35 mmol, 71% yield, 80% *ee*) was prepared according to general procedure 6 at room temperature for 5 d. All characterization data were in full agreement with the racemic compound. $[\alpha]_D^{26} = -7.1$ (*c* 1.0, CHCl$_3$).

(S)-3-(1,3-dioxolan-2-yl)-1-(1-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indol-1-yl)propan-1-one (6.33). Brown oil. Compound 6.33 (106 mg, 0.34 mmol, 68% yield, 75% *ee*) was prepared according to general procedure 6 at room temperature for 5 d. All characterization data were in full agreement with the racemic product. $[\alpha]_D^{26} = -1.3$ (*c* 1.0, CHCl$_3$).

6.6 References

1.   Pictet, A.; Spengler, T., On the formation of isochinolin-derivatives through the development of methylial on phenyl-aether, phenyl-alanin and tyrosin. *Ber Dtsch Chem Ges* **1911,** *44*, 2030-2036.

2.   Tatsui, G. J., *J. Pharm. Soc. Jpn.* **1928,** *48*, 92.

3.   Subba Reddy, B. V.; Swain, M.; Reddy, S. M.; Yadav, J. S.; Sridhar, B., Gold-Catalyzed Domino Cycloisomerization/Pictet–Spengler Reaction of 2-(4-Aminobut-1-yn-1-yl)anilines with Aldehydes: Synthesis of Tetrahydropyrido[4,3-b]indole Scaffolds. *J Org Chem* **2012,** *77* (24), 11355-11361.

4.   Scharf, M. J.; List, B., A Catalytic Asymmetric Pictet-Spengler Platform as a Biomimetic Diversification Strategy toward Naturally Occurring Alkaloids. *J Am Chem Soc* **2022,** *144* (34), 15451-15456.

5.   Plate, R.; Van Hout, R. H. M.; Behm, H.; Ottenheijm, H. C. J., Synthesis of 2-hydroxy-3-(ethoxycarbonyl)-1,2,3,4-tetrahydro-.beta.-carbolines from N-hydroxytryptophans. An approach to the eudistomin series. *J Org Chem* **1987,** *52* (4), 555-560.

6.   Wünsch, B.; Zott, M., Tricyclic benzomorphan analogues by intramolecular Oxa-Pictet-Spengler reaction. *Tetrahedron: Asymmetry* **1993,** *4* (11), 2307-2310.

7.   Calcaterra, A.; Mangiardi, L.; Delle Monache, G.; Quaglio, D.; Balducci, S.; Berardozzi, S.; Iazzetti, A.; Franzini, R.; Botta, B.; Ghirga, F., The Pictet-Spengler Reaction Updates Its Habits. *Molecules* **2020,** *25* (2).

8.   Ingallina, C.; D'Acquarica, I.; Delle Monache, G.; Ghirga, F.; Quaglio, D.; Ghirga, P.; Berardozzi, S.; Markovic, V.; Botta, B., The Pictet-Spengler Reaction Still on Stage. *Curr Pharm Des* **2016,** *22* (12), 1808-50.

9.   Bailey, P. D., Direct Proof of the Involvement of a Spiro Intermediate in the Pictet-Spengler Reaction. *J Chem Res* **1987,** (6), 202-203.

10.   Bailey, P. D., On the Stereochemistry of the Pictet-Spengler Reaction. *Tetrahedron Lett* **1987,** *28* (43), 5181-5184.

11.   Klausen, R. S.; Kennedy, C. R.; Hyde, A. M.; Jacobsen, E. N., Chiral Thioureas Promote Enantioselective Pictet-Spengler Cyclization by Stabilizing Every Intermediate and Transition State in the Carboxylic Acid-Catalyzed Reaction. *J Am Chem Soc* **2017,** *139* (35), 12299-12309.

12.   Zheng, C.; Xia, Z. L.; You, S. L., Unified Mechanistic Understandings of Pictet-Spengler Reactions. *Chem* **2018,** *4* (8), 1952-1966.

13.   Heravi, M. M.; Zadsirjan, V.; Malmir, M., Application of the Asymmetric Pictet-Spengler Reaction in the Total Synthesis of Natural Products and Relevant Biologically Active Compounds. *Molecules* **2018,** *23* (4).

14.   Stöckigt, J.; Antonchick, A. P.; Wu, F. R.; Waldmann, H., The Pictet-Spengler Reaction in Nature and in Organic Chemistry. *Angew Chem Int Ed* **2011,** *50* (37), 8538-8564.

15.   Dey, P.; Kundu, A.; Kumar, A.; Gupta, M.; Lee, B. M.; Bhakta, T.; Dash, S.; Kim, H.

S., Analysis of alkaloids (indole alkaloids, isoquinoline alkaloids, tropane alkaloids). *Recent Advances in Natural Products Analysis* **2020**, 505-567.

16.   Olofinsan, K.; Abrahamse, H.; George, B. P., Therapeutic Role of Alkaloids and Alkaloid Derivatives in Cancer Management. *Molecules* **2023,** *28* (14).

17.   Siddiqui, S.; Siddiqui, R. H. J. J. I. C. S., Chemical examination of the roots of Rauwolfia serpentina Benth. *J Indian Chem. Soc.* **1931,** *8*, 667-680.

18.   Cotoi, S.; Repolski, M.; Georgescu, C.; Bratu, A., The effect of ajmaline in atrial flutter. *Acta Cardiol* **1977,** *32* (2), 99-103.

19.   Manz, M.; Mletzko, R.; Jung, W.; Luderitz, B., Electrophysiological and haemodynamic effects of lidocaine and ajmaline in the management of sustained ventricular tachycardia. *Eur Heart J* **1992,** *13* (8), 1123-8.

20.   Zhong, Y.; He, S.; Huang, K.; Liang, M., Neferine suppresses vascular endothelial inflammation by inhibiting the NF-kappaB signaling pathway. *Arch Biochem Biophys* **2020,** *696*, 108595.

21.   Xie, J.; Chen, M. H.; Ying, C. P.; Chen, M. Y., Neferine induces p38 MAPK/JNK1/2 activation to modulate melanoma proliferation, apoptosis, and oxidative stress. *Ann Transl Med* **2020,** *8* (24), 1643.

22.   Deng, G. L.; Zeng, S.; Ma, J. L.; Zhang, Y.; Qu, Y. L.; Han, Y.; Yin, L.; Cai, C. J.; Guo, C.; Shen, H., The anti-tumor activities of Neferine on cell invasion and oxaliplatin sensitivity regulated by EMT Snail signaling in hepatocellular carcinoma. *Sci Rep-Uk* **2017,** *7*.

23.   Baumann, M.; Baxendale, I. R.; Ley, S. V.; Nikbin, N., An overview of the key routes to the best selling 5-membered ring heterocyclic pharmaceuticals. *Beilstein J Org Chem* **2011,** *7*, 442-495.

24.   Tamatam, R.; Shin, D., Asymmetric Synthesis of US-FDA Approved Drugs over Five Years (2016-2020): A Recapitulation of Chirality. *Pharmaceuticals (Basel)* **2023,** *16* (3).

25.   Drayer, D. E., Pharmacodynamic and pharmacokinetic differences between drug enantiomers in humans: an overview. *Clin Pharmacol Ther* **1986,** *40* (2), 125-33.

26.   FDA, U., Development of New Stereoisomeric Drugs. 1992.

27.   Seayad, J.; Seayad, A. M.; List, B., Catalytic asymmetric Pictet-Spengler reaction. *J Am Chem Soc* **2006,** *128* (4), 1086-7.

28.   Klausen, R. S.; Jacobsen, E. N., Weak Bronsted Acid-Thiourea Co-catalysis: Enantioselective, Catalytic Protio-Pictet-Spengler Reactions. *Org Lett* **2009,** *11* (4), 887-890.

29.   Qi, L.; Hou, H.; Ling, F.; Zhong, W. J. O.; chemistry, b., The cinchona alkaloid squaramide catalyzed asymmetric Pictet–Spengler reaction and related theoretical studies. *Org Biomol Chem* **2018,** *16* (4), 566-574.

30.   Biswas, A., Organocatalyzed Asymmetric Pictet-Spengler Reactions. *ChemistrySelect* **2023,** *8* (3), e202203368.

31.   Bou-Hamdan, F. R.; Leighton, J. L., Highly enantioselective Pictet-Spengler reactions with alpha-ketoamide-derived ketimines: access to an unusual class of quaternary alpha-amino amides. *Angew Chem Int Ed* **2009,** *48* (13), 2403-6.

32.   Lynch-Colameta, T.; Greta, S.; Snyder, S. A., Synthesis of aza-quaternary centers via Pictet-Spengler reactions of ketonitrones. *Chem Sci* **2021,** *12* (17), 6181-6187.

33.   Nakamura, S.; Matsuda, Y.; Takehara, T.; Suzuki, T., Enantioselective Pictet-Spengler Reaction of Acyclic alpha-Ketoesters Using Chiral Imidazoline-Phosphoric Acid Catalysts. *Org Lett* **2022,** *24* (4), 1072-1076.

34.   Badillo, J. J.; Silva-Garcia, A.; Shupe, B. H.; Fettinger, J. C.; Franz, A. K., Enantioselective Pictet-Spengler Reactions of Isatins for the Synthesis of Spiroindolones. *Tetrahedron Lett* **2011,** *52* (43), 5550-5553.

35.   Andres, R.; Wang, Q.; Zhu, J., Asymmetric Total Synthesis of (-)-Arborisidine and (-)-19-epi-Arborisidine Enabled by a Catalytic Enantioselective Pictet-Spengler Reaction. *J Am Chem Soc* **2020,** *142* (33), 14276-14285.

36.   Andres, R.; Wang, Q.; Zhu, J., Catalytic Enantioselective Pictet-Spengler Reaction of alpha-Ketoamides Catalyzed by a Single H-Bond Donor Organocatalyst. *Angew Chem Int Ed Engl* **2022,** *61* (19), e202201788.

37.   Unpublished data from the Zhu group.

38.   Brak, K.; Jacobsen, E. N., Asymmetric ion-pairing catalysis. *Angew Chem Int Ed Engl* **2013,** *52* (2), 534-61.

39.   Babudri, F.; Fiandanese, V.; Marchese, G.; Punzi, A., A direct access to α-diones from oxalyl chloride. *Tetrahedron Lett* **1995,** *36* (40), 7305-7308.

40.   Heck, R.; Henderson, Alistair P.; Köhler, B.; Rétey, J.; Golding, Bernard T., Crossed Acyloin Condensation of Aliphatic Aldehydes. *Eur J Org Chem* **2001,** *2001* (14), 2623-2627.

41.   Burai Patrascu, M.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N., From desktop to benchtop with automated computational workflows for computer-aided design in asymmetric catalysis. *Nat Catal* **2020,** *3* (7), 574-584.

42.   Stetter, H.; Dambkes, G., Preparative Use of 1,3-Thiazolium Salt Catalyzed Formation of Acyloin and Benzoin .3. New Method for Preparation of Substituted Enol Trimethylsilyl Ethers of 1,2-Cyclopentanedione. *Synthesis-Stuttgart* **1980,** (4), 309-310.

43.   Juric, S.; Kronja, O., Reactivity of some tertiary chlorides with olefinic and methoxy neighboring groups. A case of extended pi, n-participation. *J Phys Org Chem* **2002,** *15* (8), 556-560.

44.   Mathies, A. K.; Mattson, A. E.; Scheidt, K. A., Thieme Chemistry Journal Awardees-Where are They Now? Intermolecular Cross-Acyloin Reactions by Fluoride-Promoted Additions of O-Silyl Thiazolium Carbinols. *Synlett* **2009,** (3), 377-383.

45.   Xin, L. H.; Bausch, C. C.; Johnson, J. S., Mechanism and scope of the cyanide-catalyzed cross silyl benzoin reaction. *J Am Chem Soc* **2005,** *127* (6), 1833-1840.

46.   Fredrich, S.; Gostl, R.; Herder, M.; Grubert, L.; Hecht, S., Switching Diarylethenes Reliably in Both Directions with Visible Light. *Angew Chem Int Ed* **2016,** *55* (3), 1208-1212.

47.   Seo, S. Y.; Yu, X. H.; Marks, T. J., Intramolecular Hydroalkoxylation/Cyclization of Alkynyl Alcohols Mediated by Lanthanide Catalysts. Scope and Reaction Mechanism. *J Am Chem Soc* **2009,** *131* (1), 263-276.

48.   Nguyen, J.; Duncan, N.; Lalic, G., Direct -Selective Cross-Coupling of Alkenyl Gold Complexes with Alkyl Electrophiles. *Eur J Org Chem* **2016,** *2016* (35), 5803-5806.

49.   Furstner, A.; Guth, O.; Rumbo, A.; Seidel, G., Ring closing alkyne metathesis. Comparative investigation of two different catalyst systems and application to the stereoselective synthesis of olfactory lactones, azamacrolides, and the macrocyclic perimeter of the marine alkaloid nakadomarin A. *J Am Chem Soc* **1999,** *121* (48), 11108-11113.

50.   Zheng, W. F.; Xu, Q. J.; Kang, Q., Rhodium/Lewis Acid Catalyzed Regioselective Addition of 1,3-Dicarbonyl Compounds to Internal Alkynes. *Organometallics* **2017,** *36* (12), 2323-2330.

CHAPTER 7:

CONCLUSIONS AND FUTURE WORK

As introduced in **Chapter 1**, molecular docking is a valuable computational tool that can aid in the development of therapeutic agents. One advantage of molecular docking is that it enables libraries of compounds to be screened *in silico* (i.e. virtual screening), from which only the most promising compounds are selected for experimental evaluation. Docking can also be particularly advantageous in structure-activity relationship studies, as it can be used to predict how changes to an inhibitor's scaffold might affect its binding conformation and affinity. Similar to virtual screening, this allows synthetic efforts to focus on only the most promising analogues, improving the efficiency of the drug development process. Although more than 70 docking programs have been developed, the majority of these programs were developed with a focus on non-covalent binding to protein targets. However, this represents a relatively narrow therapeutic target space, considering that many protein targets contain a metal ion or covalent residue within their binding site which may not be accurately modeled by purely non-covalent interactions. Additionally, many nucleic acid-based targets have been identified, raising the question of whether docking programs developed for proteins can be extended to nucleic acids.

The aim of this thesis has been to demonstrate the application of docking towards these more diverse therapeutic targets, where binding may be mediated by metal coordination, covalent bond formation, or via interactions with nucleic acids. As discussed in Chapter 1, our in-house docking program FITTED has undergone several optimizations to improve its applicability and accuracy when applied to metalloenzymes, covalent residues and nucleic acids. Several assessments on relevant testing sets have demonstrated its improved accuracy for these target classes when considering the accuracy of pose prediction and its ability to discriminate between active and inactive compounds. However, one could argue that the true success of a docking program would be its successful application towards the development of novel biologically active compounds. With this in mind, the subsequent chapters in this thesis detail the application of FITTED towards the development of inhibitors targeting AID (a zinc-metalloenzyme), the TPP riboswitch (an RNA drug target), and SARS-CoV-2 3CL^{pro} (a cysteine protease amenable to covalent inhibition). While

this led to the successful identification of a preliminary hit compound against the TPP riboswitch and potent inhibitors against SARS-CoV-2 3CL^{pro}, several challenges and limitations were encountered during the course of this work. These experiences could provide valuable insights for guiding future research efforts.

As discussed in **<u>Chapter 2</u>**, initial efforts towards the development of inhibitors against AID focused on the resynthesis of a previously identified inhibitor of AID, with the plan of using docking to then aid in the design of optimized analogues. However, upon resynthesis, the initial inhibitor did not display any *in vitro* inhibitory activity. Although this was a disappointing result, it demonstrated the importance of resynthesizing commercial samples to confirm biological activity. This may be especially true for preliminary hit compounds that are only weakly active, as impurities may be present at relatively high concentrations when testing at such high inhibitor concentrations. The preliminary hit compound had been identified through an independently performed virtual screening campaign and serves as another important reminder that false positives are also possible for preliminary hits identified by molecular docking. In an attempt to identify a novel preliminary inhibitor for AID, a virtual screen was performed using FITTED. Considering the presence of a zinc ion in the active site of AID and the strength often afforded by metal-ligand bonds, we began by virtually screening a library of fragment-sized compounds with known zinc-binding groups, with the intention of first identifying a promising zinc-binding group and then subsequently optimizing it via a fragment-growth strategy. However, none of the fragments selected from the virtual screening results displayed any inhibitory activity *in vitro*. By comparing the binding energies calculated by FITTED with those obtained with higher level quantum mechanics calculations, we demonstrated that FITTED was accurately modelling metal-ligand interactions with the specific zinc ion and coordination sphere found in AID. This suggested that the lack of activity observed from our screening library was not merely a result of incorrect modelling of the zinc-ligand bond in FITTED. Considerations for future work on this project, or in the development of metalloenzyme inhibitors in general, might include a more explicit consideration of metal-binding group diversity. It is possible that no active compounds were identified from our virtual screen because the selected compounds did not cover a diverse enough molecular space. While overall ligand diversity was considered in the assembly of the virtual screening library, it may be beneficial to ensure that a wide range of metal-binding groups are equally represented in future metalloenzyme screening libraries. Similarly, it may also be

advantageous to consider the diversity of metal-binding groups when selecting compounds for experimental evaluation as opposed to selecting based predominately on the predicted docking score. Additional work on this target was subsequently carried out by another graduate student who investigated the application of transition state analogues as potential AID inhibitors, however these also failed to inhibit AID. Considering these results, as well as structural studies of AID that demonstrate the presence of a positively charged DNA binding cleft, AID may be more suited to oligonucleotide-based inhibitors as opposed to more traditional small molecules.

Although our attempts at the computer-aided development of inhibitors was unsuccessful for AID, virtual screening against the TPP riboswitch successfully led to the identification of a micromolar binder, as presented in **<u>Chapter 3</u>**. A pharmacophore model was employed to guide the docking process towards binding poses that were able to satisfy key intermolecular interactions, which improved the docking accuracy in self-docking of TPP and resulted in a promising AUROC value when applied towards a retrospective virtual screen using known active compounds. This docking protocol was then applied to virtually screen a library of compounds against the TPP riboswitch, and 20 compounds were selected for experimental evaluation based on their docking score and visual inspection of the top ranking binding poses. Compounds were tested for their binding affinity to the TPP aptamer domain using surface plasmon resonance (SPR), with one compound displaying mid-micromolar affinity. Considering the results obtained with the preliminary AID inhibitor discussed in Chapter 2 (where the resynthesized compound no longer displayed *in vitro* activity), subsequent work on this project focused on confirming the activity observed with our commercial inhibitor sample. The preliminary hit compound was successfully resynthesized by another graduate student and tested on TPP, validating the activity observed with the commercial sample. Many of the previously reported inhibitors of the TPP riboswitch have focused on close analogues of the native ligand that feature a pyrophosphate (or similarly anionic) group. The results obtained in Chapter 3 suggest that alternative groups such as the monosaccharide moiety in compound **3.23** could mediate similar intermolecular interactions as the pyrophosphate group of TPP. These findings could help instruct the design of more drug-like inhibitors of the TPP riboswitch than have been previously reported. Future work for this project may involve the optimization of initial hit compound **3.23** through docking-guided structure-activity relationship studies. This could include modifying the glucose moiety, which is predicted to bind the pyrophosphate sensor helix, to investigate whether similar or improved affinities are observed with

other monosaccharide-like groups. Similarly, modifications to the benzodioxane group, which is predicted to bind the pyrimidine sensor helix, could confer greater affinity by optimizing hydrogen-bonding interactions with G40 as observed with TPP. However, it is important to keep in mind that the accuracy of pose prediction for docking to nucleic acids is low, and although the predicted binding pose of compound **3.23** (and related analogues) could be instructive for designing improved analogues, further experimental validation is also crucial. Although we have confirmed by SPR that the preliminary hit **3.23** binds to the aptamer domain of the TPP riboswitch, this does not provide any validation regarding its binding mode nor its inhibitory activity. The most reliable method to validate the predicted binding pose of **3.23** would be to obtain an experimental co-crystal structure. Experimental evaluation of the effect of **3.23** on TPP riboswitch-mediated gene expression could also serve as further validation. Such an evaluation would indicate whether the preliminary hit binds to both the pyrimidine and pyrophosphate sensor helices, thereby stabilizing the riboswitch in the off state. So far, our attention has been directed towards identifying a preliminary binder for the aptamer domain of the TPP riboswitch. However, the ultimate objective of this project is to develop an inhibitor for the TPP riboswitch and will therefore require more thorough investigations on how any preliminary hit compounds affect riboswitch-mediated gene expression.

Chapters 4 and 5 then focused on the application of docking towards the development of covalent inhibitors against SARS-CoV-2 3CL^{pro}. As detailed in **Chapter 4**, our initial efforts were aimed at the incorporation of a covalent warhead into the scaffold of a previously reported non-covalent inhibitor for SARS-CoV 3CL^{pro}. The docked poses predicted by FITTED supported the design strategy, suggesting that a covalent bond could be favourably formed between the reactive cysteine residue in the binding site of 3CL^{pro} while still maintaining key non-covalent interactions. A library of covalent analogues was synthesized and evaluated *in vitro*, with several analogues displaying low micromolar or sub-micromolar potencies. Considering the two-step nature of targeted covalent inhibitors, further *in vitro* evaluations were performed on the most promising inhibitors to obtain their initial non-covalent binding affinity (K_i) and rate of covalent bond formation (k_{inact}). Additionally, the work presented in Chapter 4 demonstrated that our in-house docking program FITTED was able to successfully predict the binding mode of novel covalent inhibitors, as validated by X-ray crystallography. Unlike many other docking programs, covalent docking in FITTED does not involve covalent bond pre-formation and only forms the covalent adduct if the inhibitor is

sufficiently close to the reactive residue in its non-covalent state. The results obtained with our 3CL^{pro} inhibitors further demonstrate that accurate docking predictions can be obtained without pre-formation of the covalent bond. However, it is equally important to note the limitations in covalent docking with FITTED. Currently, there is no consideration of warhead reactivity in the predicted docking score and therefore analogues with different warheads may be predicted to have very similar scores. For example, the docking score calculated in FITTED would not be able to differentiate between an α-fluoroketone and an α-chloroketone, despite the latter being considerably more active *in vitro*. As a result of this limitation, docking could only be effectively applied to support our design strategy but could not be confidently applied to prioritize specific warheads. An area of future work may be to incorporate warhead reactivity into our docking program, taking into account the initial non-covalent binding mode and the intrinsic reactivity of different warheads. The data obtained in Chapter 4 may be applicable to this aim, as we obtained IC$_{50}$ values for many analogues that differ only by their covalent warhead, and kinetic data (i.e. K_i and k_{inact} values) were obtained for several analogues, as well.

In **Chapter 5**, our attention then turned to the optimization of our initial covalent inhibitors by docking-guided modifications to the P1-P4 moieties. A slight improvement in potency was observed when modifying the P2 tert-butyl group to a phenyl substituent, while no improvements were observed with any of the P1 analogues. The greatest improvements in potency were observed with the P4 analogues. As predicted by docking, increasing the length of the P4 group could allow the inhibitor to better occupy the S4 site. This prediction was validated with several analogues *in vitro*, where modifying the cyclohexyl group to a longer homobenzyl or benzyl piperidine group led to a 2-fold improvement in IC$_{50}$ values. While bioisosteric replacement of the P3 amide group to an oxadiazole did not improve the inhibitor potency, low micromolar potency was still maintained and is one of the first 3CL^{pro} inhibitors to introduce a heterocyclic group in the P3 position. The observed *in vitro* and *in cellulo* activity of this analogue suggests that a P3 heterocycle could be a viable strategy for the design of future 3CL^{pro} inhibitors to address any potential pharmacokinetic limitations related to the P3 amide group. Ten of our most promising inhibitors were subsequently tested in a cell-based assay against SARS-CoV-2, with five displaying complete viral inhibition at 20 μM. However, inhibitor **4.13**, which features an α-chloroketone warhead, also displayed complete inhibition of cell viability at this concentration, indicating significant cytotoxic effects for this analogue. Somewhat surprisingly, none of the

inhibitors featuring a vinyl sulfonamide warhead were active *in cellulo* despite being the most potent analogues *in vitro*. As observed in Chapter 4, the vinyl sulfonamide analogue had a rate of covalent bond formation nearly twice that determined for the hydroxyalkyne analogue and five times greater than the alkyne analogue. The poor *in cellulo* results may be due to this increased reactivity, as the vinyl sulfonamide may be more susceptible to deactivation by other nucleophiles that are present *in cellulo* or may have decreased aqueous stability. It may be advantageous for future covalent inhibitors to evaluate the stability of the warhead under physiological conditions when prioritizing warheads. For example, the vinyl sulfonamide was selected as the warhead in the synthesis of P4 analogues based on its *in vitro* potency, however the warhead displayed poor *in cellulo* activity. As a result, we were not able to assess the effects of the P4 modifications on viral inhibition. Additional work on this project is currently underway to address this, focusing on the synthesis of analogues that incorporate the most promising modifications (considering both *in vitro* and *in cellulo* results) into the same scaffold. Further work for this project also entails a more thorough evaluation of *in cellulo* activity to determine EC_{50} and CC_{50} values for the most promising inhibitors. The current *in cellulo* results only provide a more rudimentary assessment of inhibition, as only two inhibitor concentrations were tested. As a result, it is difficult to assess more subtle differences between different analogues. For example, the P3 amide inhibitor **5.32** and oxadiazole analogue **5.31** both show complete inhibition at 20 μM but minimal inhibition at 10 μM, suggesting that their EC_{50} values are likely somewhere between these two values. However, neither showed significant cytotoxicity at the two inhibitor concentrations tested and therefore little information can be extracted regarding their respective CC_{50} values. While the two inhibitors may have similar potencies, they may have very different selectivity indices (e.g., the difference between their effective antiviral concentration and cytotoxic concentration) which is an important consideration in drug development.

As a change of pace, **<u>Chapter 6</u>** describes research performed as a visiting doctoral student in the laboratory of Dr. Jieping Zhu. This chapter focused on the development of an organocatalytic enantioselective Pictet-Spengler reaction for diketone substrates via a chiral urea/benzoic acid co-catalyst system. A preliminary reaction optimization led to improved enantioselectivities by modifying the solvent and the acid co-catalyst. Following reaction optimization, a small library of symmetric and unsymmetric diketones were synthesized to evaluate the substrate scope of the reaction. High enantioselectivities (*ee* > 90%) were obtained for 5 out of 8 substrates and the

reaction showed good functional group tolerance. However, lower enantioselectivities (*ee* < 80%) were obtained for diketones bearing a methyl substituent. This was unexpected considering the enantioselectivity obtained during reaction optimization with 2,3-pentanedione (*ee* = 89%), which also contains a methyl substituent. One consideration for this result is that the optimization reactions were all stopped after 24 hours in order to determine the *ee* regardless of whether full conversion had been reached. Considering that most of the substrate scope reactions were performed with significantly longer reaction times (3-9 days), it is possible that this discrepancy in results is due to decreasing enantioselectivity over time. The Pictet-Spengler product may be competing with the imine intermediate for ion pairing with the co-catalyst system, which may influence the enantioselectivity if it causes a background racemic reaction to occur at a faster rate relative to the catalysed reaction. Future experimental work for this project could involve the investigation of enantioselectivity with respect to reaction time to see if such an effect is occurring. It may also be worthwhile to develop reaction conditions that could protect or mask the basicity of the Pictet-Spengler product (i.e. through *in situ* Boc protection). This may enable lower catalyst loading and/or decreased reaction times while potentially improving the enantioselectivity. Additionally, future work for this project could involve a computational investigation into the reaction mechanism and its enantioselectivity-determining transition state. Work is currently underway to incorporate this reaction into our computational platform for Asymmetric Catalyst Evaluation (ACE), which would allow for new catalysts to be evaluated *in silico*. Similar to how molecular docking has been applied to screen inhibitors and select only the most promising compounds for experimental testing, the use of ACE for this transformation could allow us to further optimize the reaction *in silico*, selecting only the best catalyst analogues for experimental evaluation of enantioselectivity.

SUPPORTING INFORMATION FOR CHAPTER 2

Characterization Data

^{1}H NMR of **Compound 2.2**
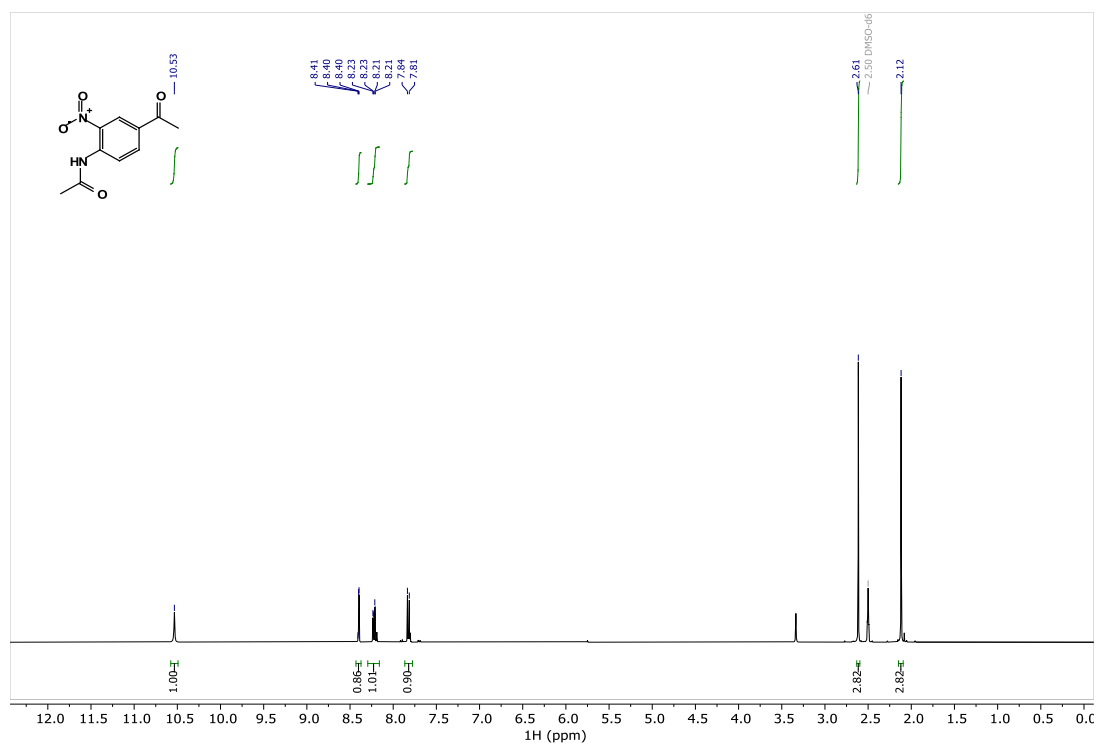
^{1}H NMR of **Compound 2.3**
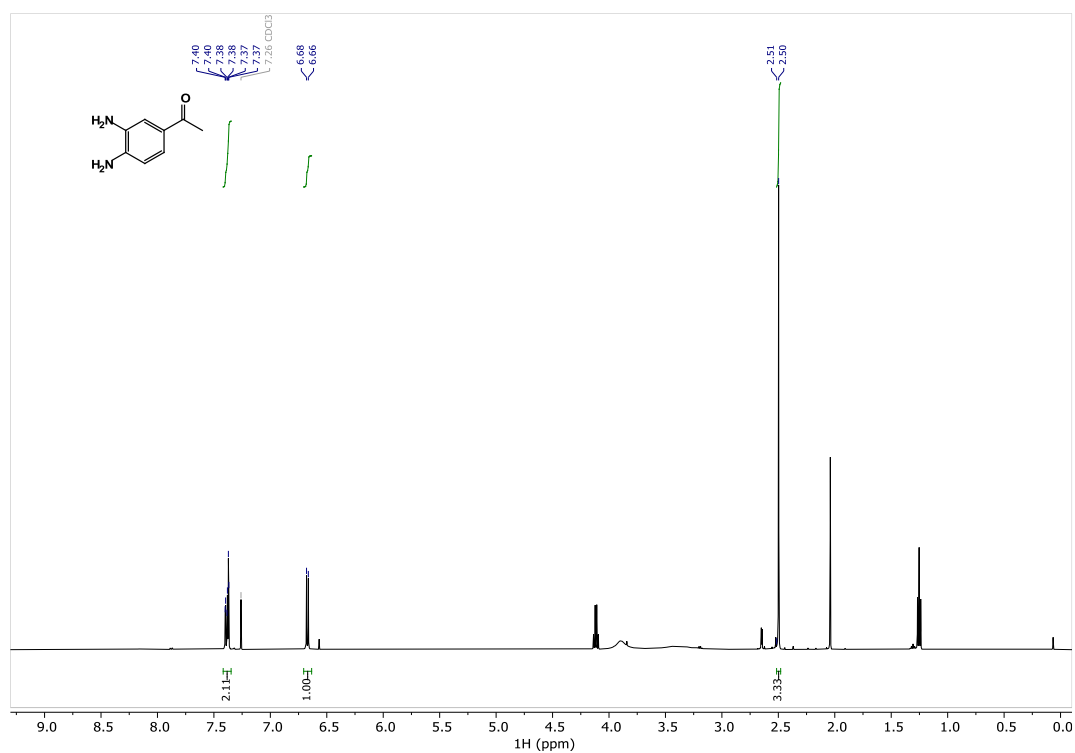
^{1}H NMR of **Compound 2.4**
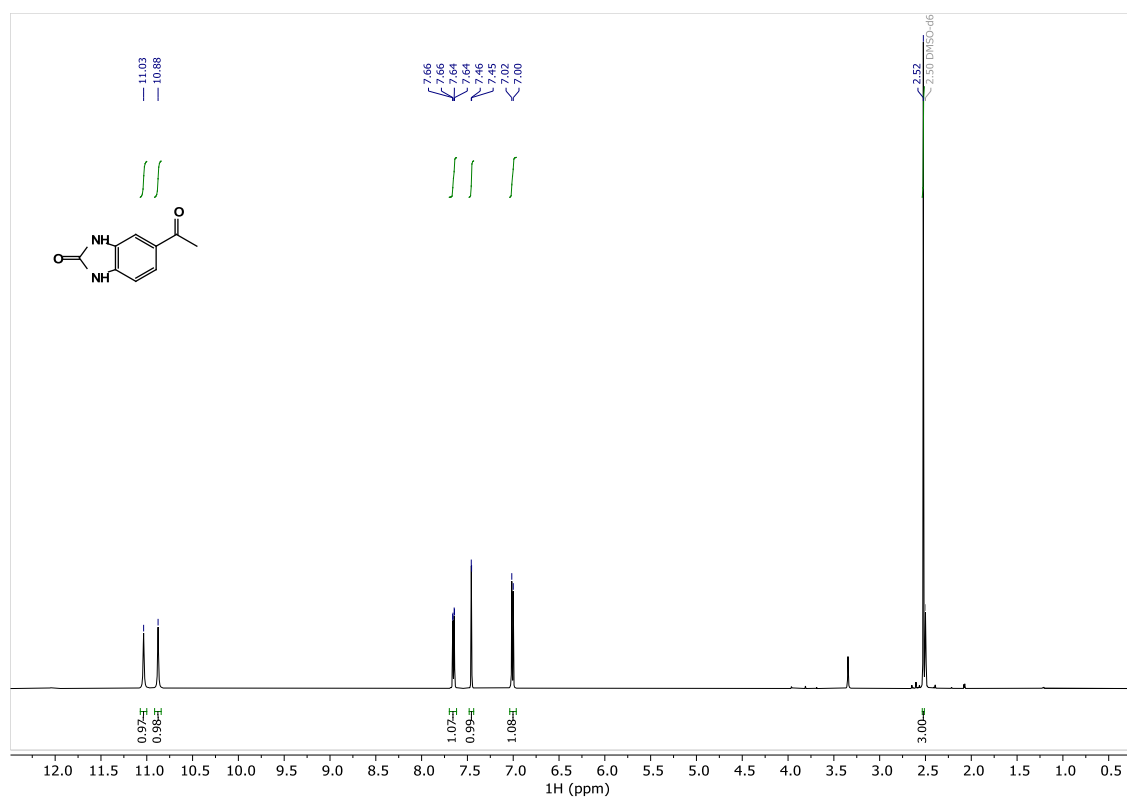
^{1}H and ^{13}C NMR of **Compound 2.5**
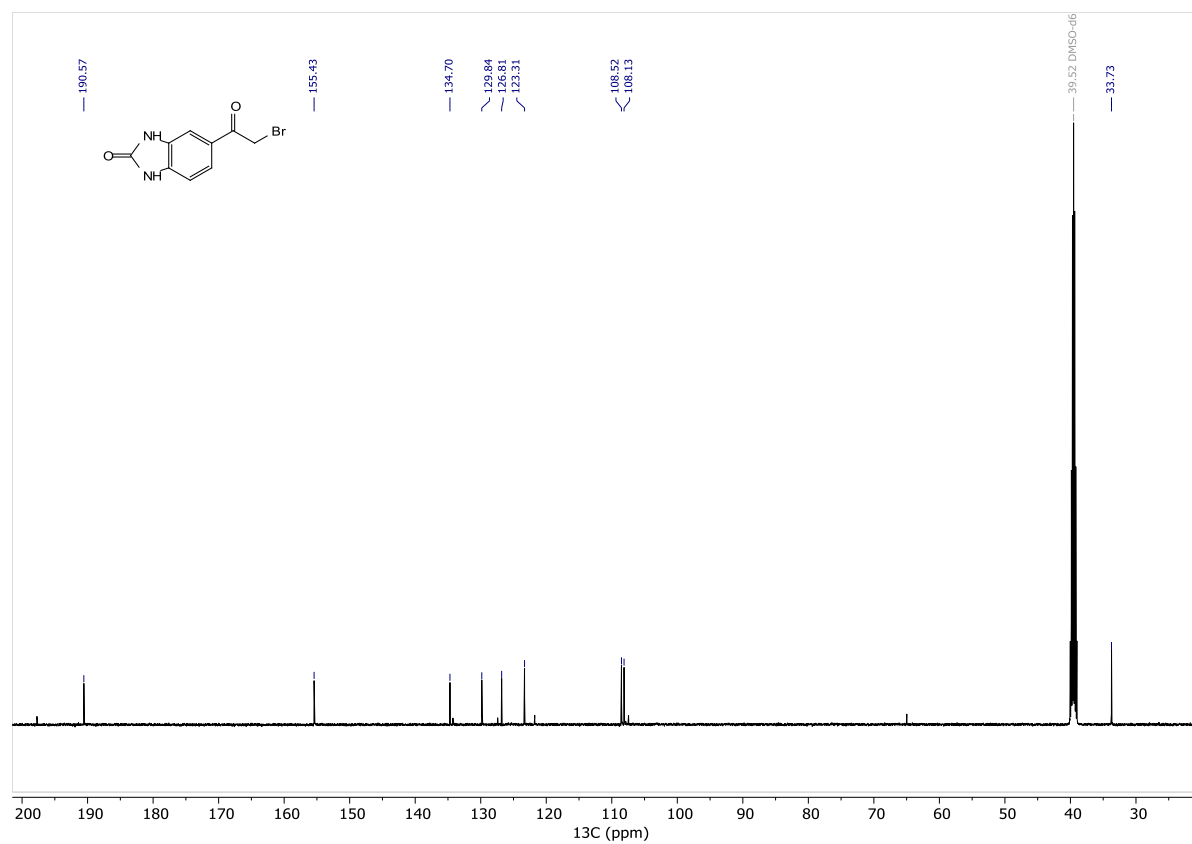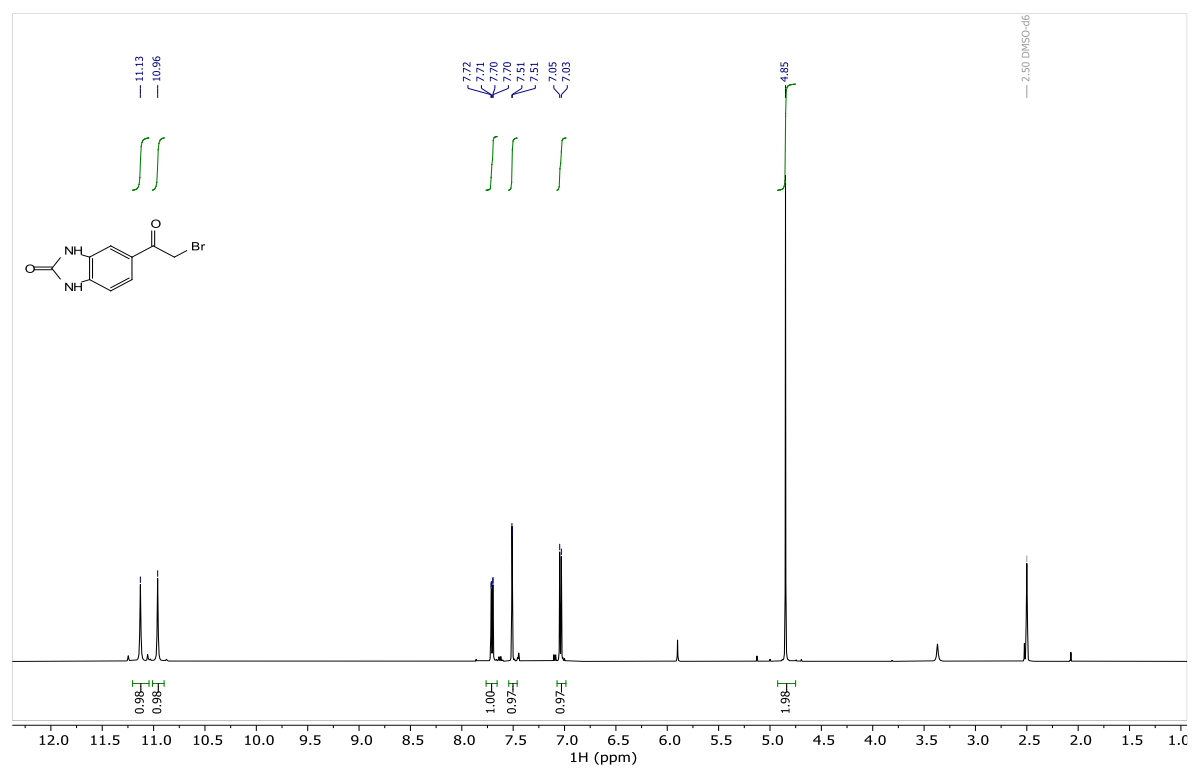
¹H and ¹³C NMR of **Compound 2.1**
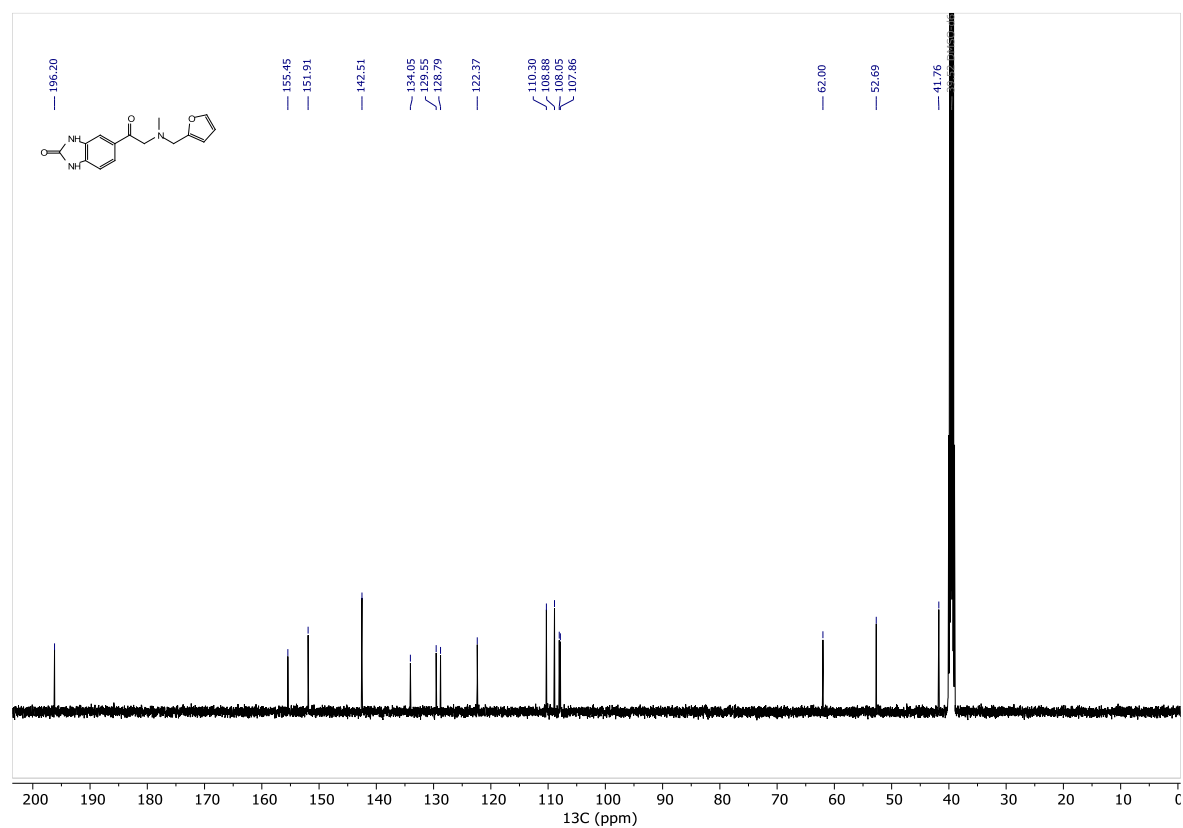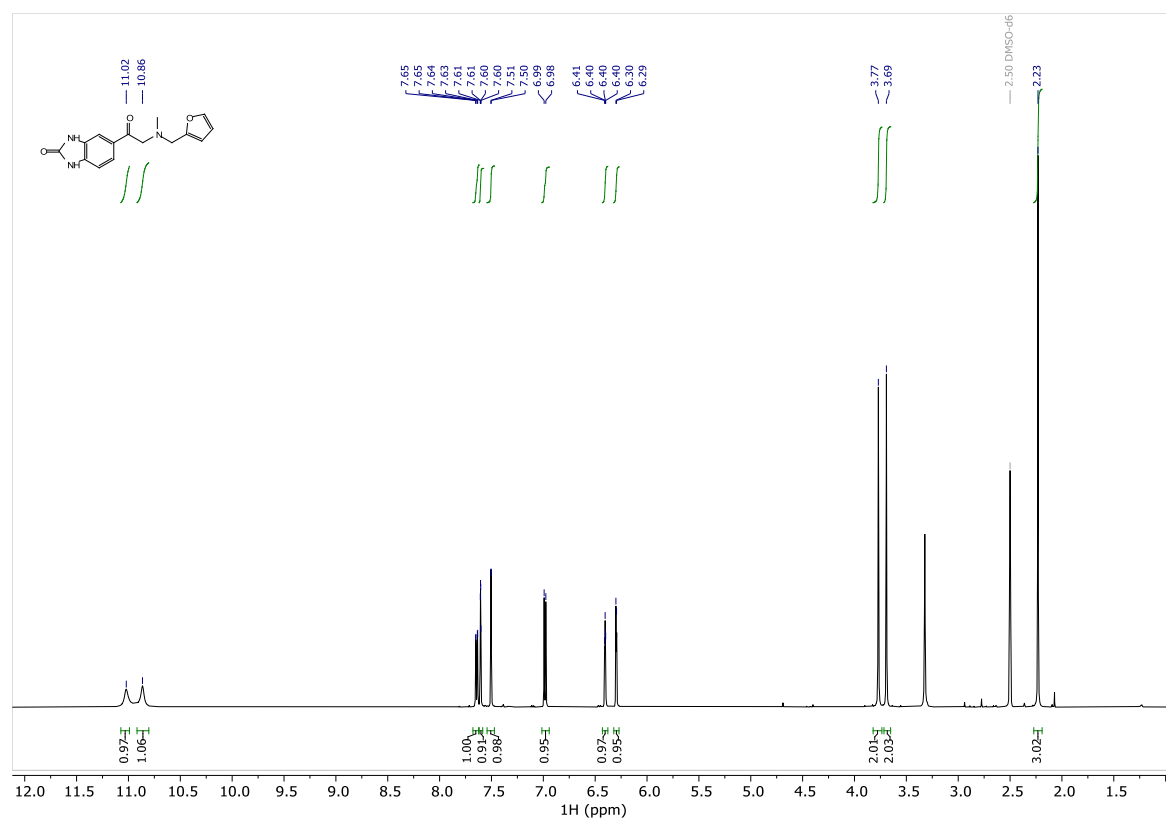
HPLC trace of **Compound 2.1**
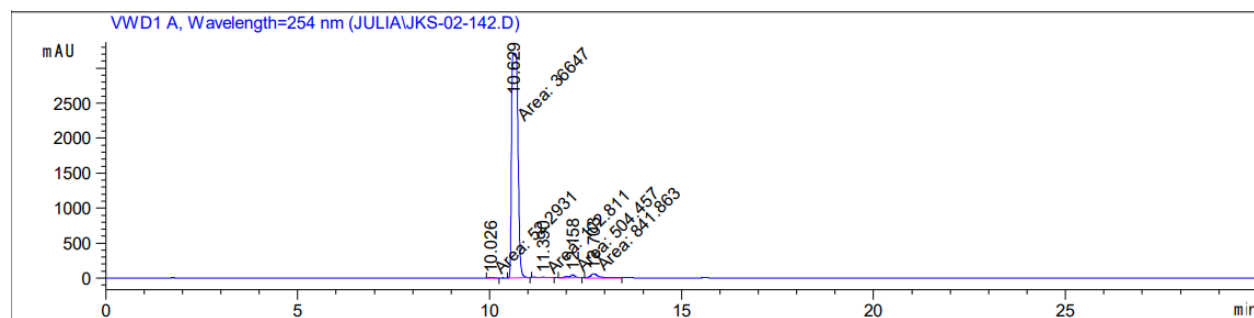
Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	10.026	MM	0.1100	52.29306	7.92340	0.1371
2	10.629	MM	0.1901	3.66470e4	3212.66895	96.0643
3	11.390	MM	0.2378	102.81062	7.20653	0.2695
4	12.158	MM	0.1916	504.45691	43.88692	1.3224
5	12.703	MM	0.2521	841.86322	55.64804	2.2068

^{1}H NMR of **Compound 2.6**
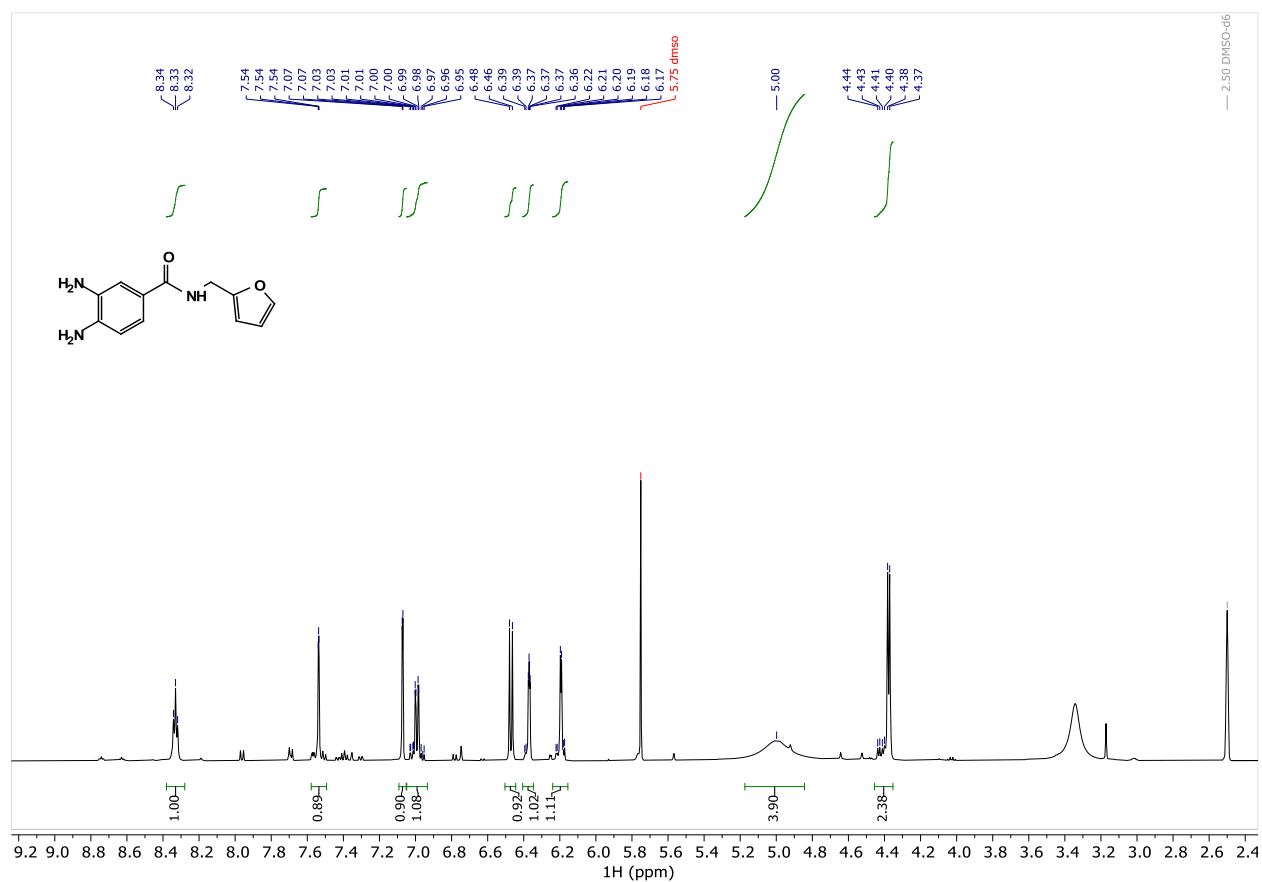
^{1}H NMR of **Compound 2.7**
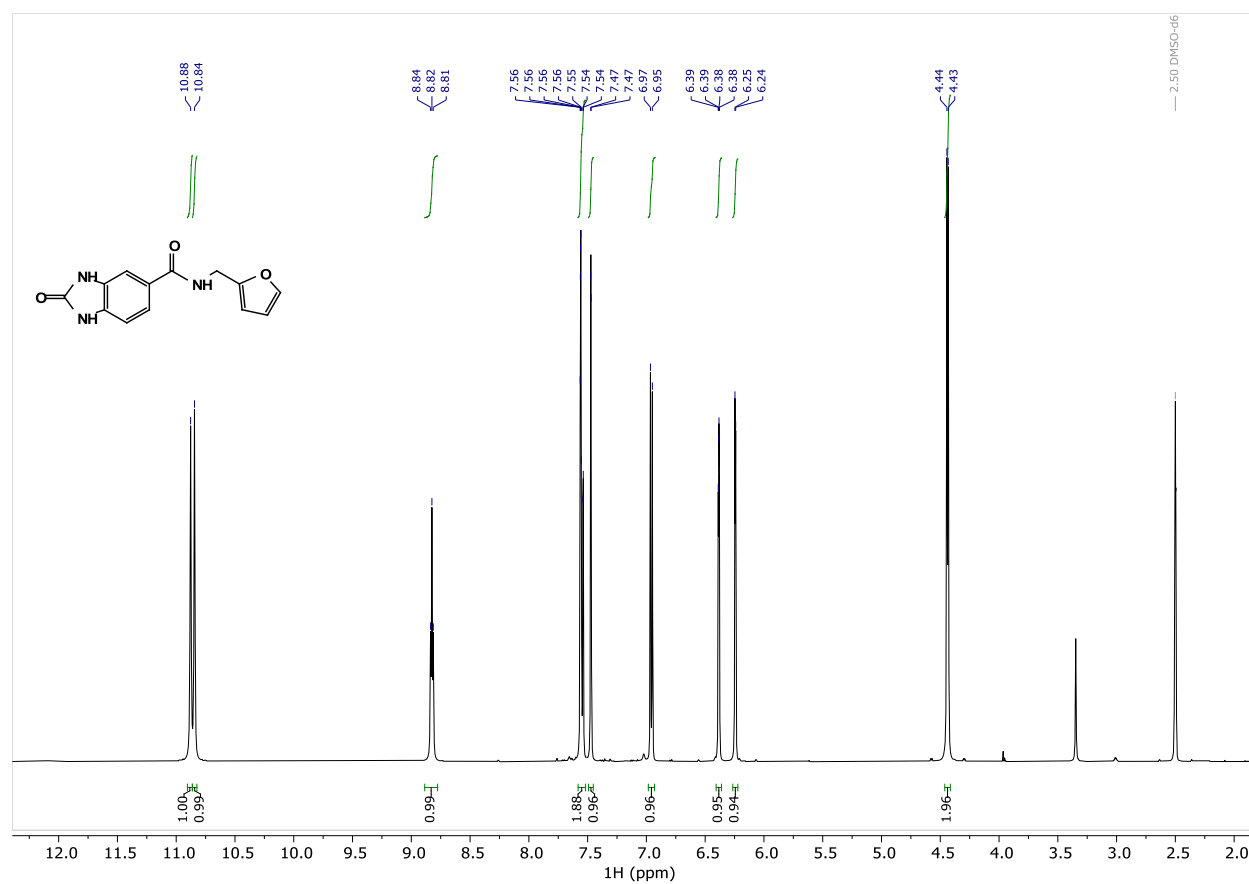
In vitro testing of compounds against AID

The standard alkaline cleavage assay for AID-mediated deamination was used to screen compounds for inhibition of AID, using bacterially-expressed and purified GST-AID. A standard seven-nucleotide bubble substrate containing the WRC motif TGC (5′-TTTGCTT-3′) as a substrate (2.5 nM) as previously reported. The compounds were tested at 500 µM concentration and dissolved in water or DMSO, with the final concentration of DMSO not exceeding 1%.
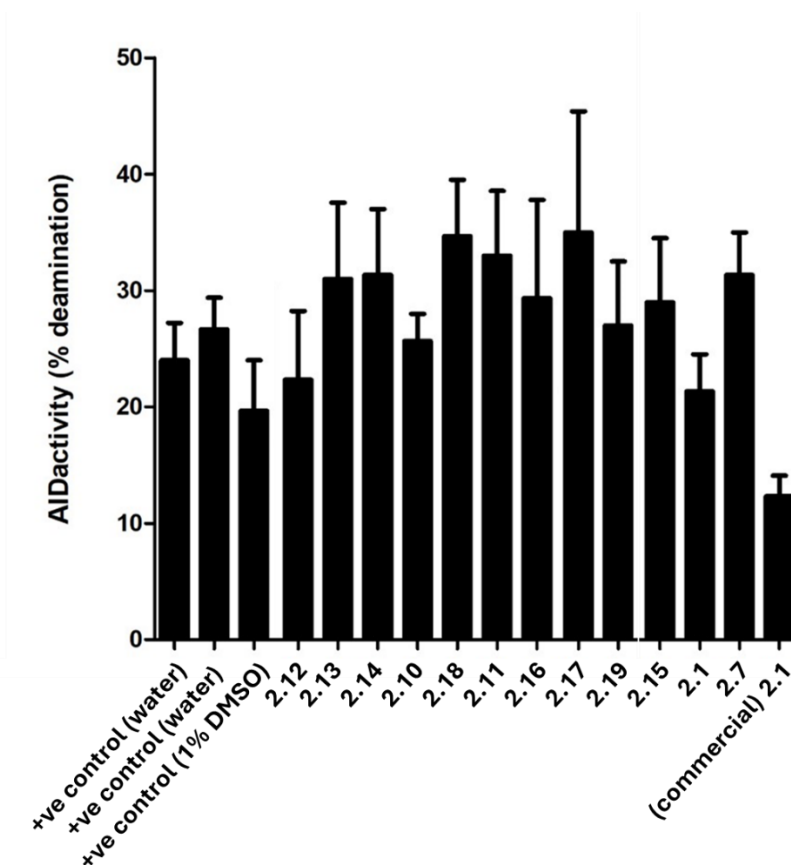
Figure S2.1. Deamination activity of compounds 2.1, 2.7, 2.10-2.19, in comparison with +ve controls and the initial commercial sample of 2.1

SUPPORTING INFORMATION FOR CHAPTER 4

Dose-Response Curves
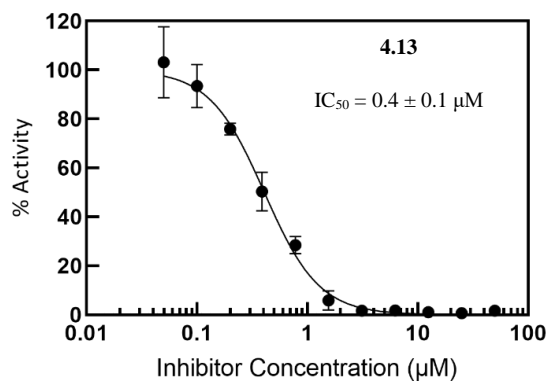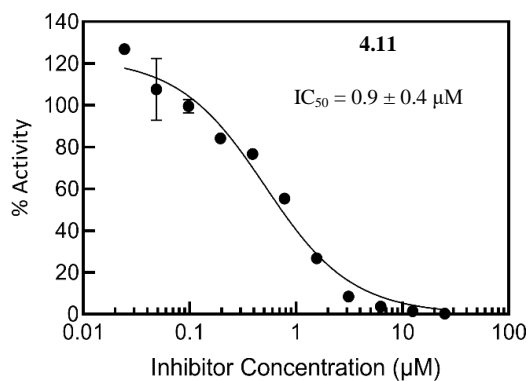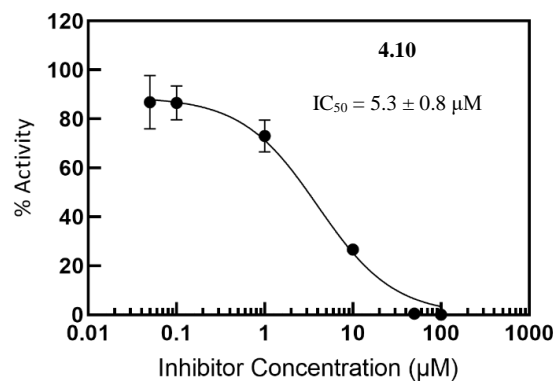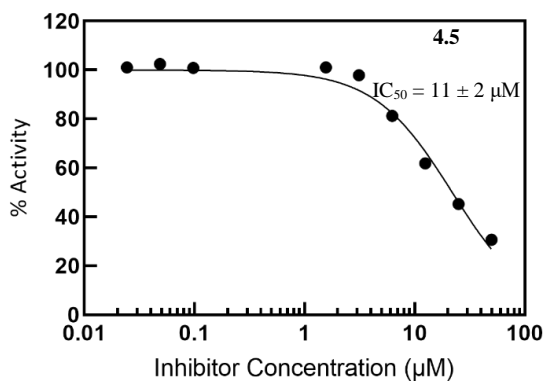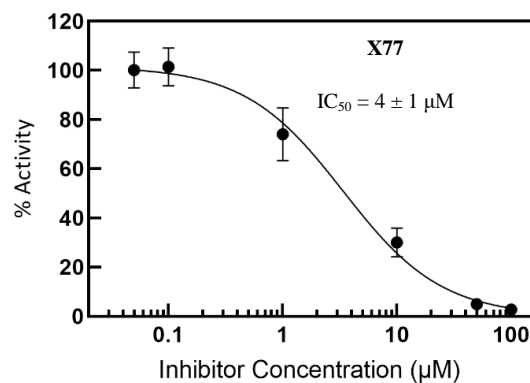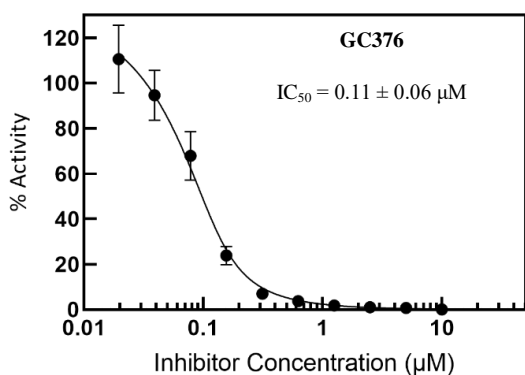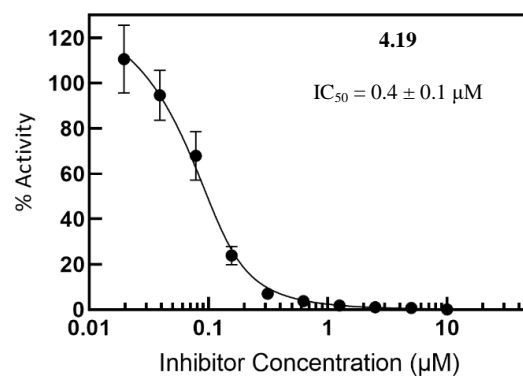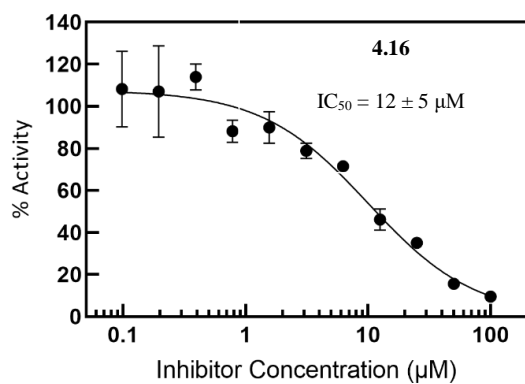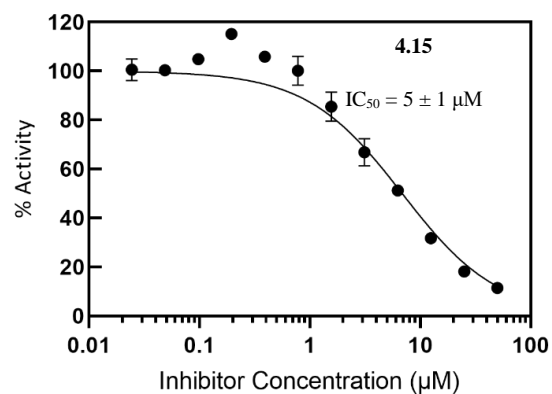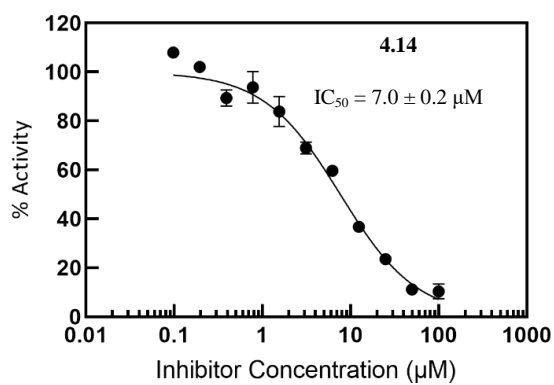
K$_M$ Determination

Reaction progress curves were determined by FRET enzymatic assays initiated by addition of 3CL^{pro} (50 nM) over a range of substrate concentrations (3 μM – 200 μM). The initial velocity for each substrate concentration was determined by linear regression of the first 20 minutes of the reaction progress curves. A plot of initial velocity was plotted as a function of substrate concentration and fitted to the Michaelis-Menten equation by non-linear regression to obtain values of K$_m$ and V$_{max}$.
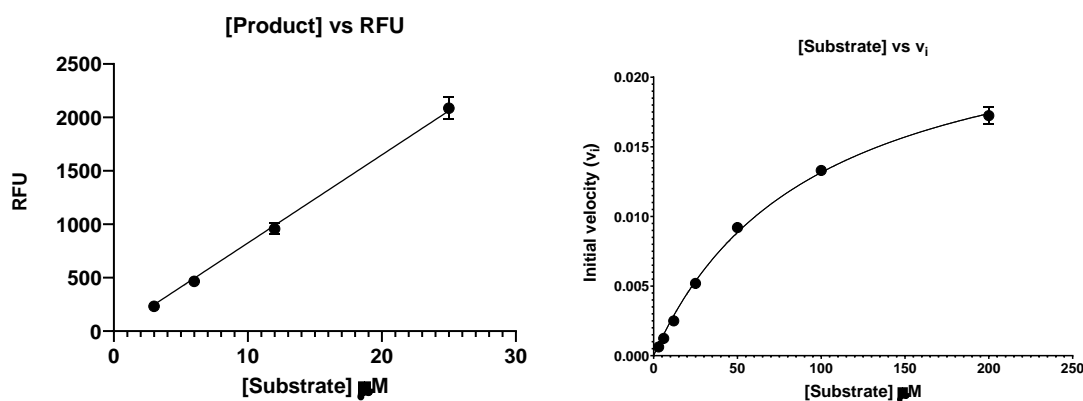
Figure S4.1. K$_M$ determination for 3CL^{pro}

Crystallography

Table S4.1. Statistics for crystallography data and structure refinement.

	3CL^{pro} – 4.13	3CL^{pro} – 4.19
Wavelength (Å)	1.521	0.97918
Resolution range	38.38 - 2.6 (2.693 - 2.6)	48.06 - 2.5 (2.59 - 2.5)
Space group	C 1 2 1	C 1 2 1
Unit cell	116.81 53.54 45.33 90 98.917 90	113.45 53.318 45.4745 90 101.876 90
Total reflections	16512 (1645)	17032 (1708)
Unique reflections	8544 (862)	9105 (901)
Multiplicity	1.9 (1.9)	2.9 (3.1)
Completeness (%)	93.18 (48.82)	97.59 (98.79)

257

Mean I/sigma(I)	6.00 (0.45)	9.37 (1.14)
Wilson B-factor	71.23	54.62
R-merge	0.04975 (1.3)	0.03943 (0.2481)
R-meas	0.07036 (1.838)	0.05576 (0.3509)
R-pim	0.04975 (1.3)	0.03943 (0.2481)
CC1/2	0.996 (0.679)	0.997 (0.777)
CC*	0.999 (0.899)	0.999 (0.935)
Reflections used in refinement	8072 (435)	9104 (901)
Reflections used for R-free	417 (20)	434 (40)
R-work	0.2146	0.2242 (0.3611)
R-free	0.2677	0.2680 (0.3928)
Number of non-hydrogen atoms	2417	2407
macromolecules	2347	2347
ligands	30	32
solvent	40	28
Protein residues	304	304
RMS(bonds)	0.013	0.012
RMS(angles)	1.67	1.65
Ramachandran favored (%)	96.36	96.36
Ramachandran allowed (%)	3.31	3.64
Ramachandran outliers (%)	0.33	0
Rotamer outliers (%)	0	0.38
Clashscore	1.5	3.83
Average B-factor	65.04	63.73
macromolecules	65.15	63.38
ligands	79.53	107.2
solvent	47.9	43.37

HPLC Compound Purity

Table S3. Purity of Biologically Tested Compounds

Entry	Compound	Retention time (min.)[a]	Purity (%)[b]
1	**X77**	15.45	99.1
2	**4.4**	14.52	95.6
3	**4.5**	15.18	98.9
4	**4.6**	15.53	99.8
5	**4.7**	15.34	95.2
6	**4.8**	16.06	95.1
7	**4.9**	16.00	99.5
8	**4.10**	15.25	94.2
9	**4.11**	14.52	98.2
10	**4.12**	15.03	94.7
11	**4.13**	15.37	95.6
12	**4.14**	15.08	90.9
13	**4.15**	15.34	97.7
14	**4.16**	15.32	97.7
15	**4.17**	12.05	93.1
16	**4.18**	15.77	96.7
17	**4.19**	15.76	91.0
18	**4.20**	15.80	99.8
19	**4.21**	15.71	98.5

[a] Conditions: (gradient of 95% water, 5% MeOH or MeCN, 1 mL/min). [b] UV detection at 254 nm.

NMR Spectra

Compound **X77** ^{1}H and ^{13}C NMR

JEV-1-Frit-MeOD.1.fid
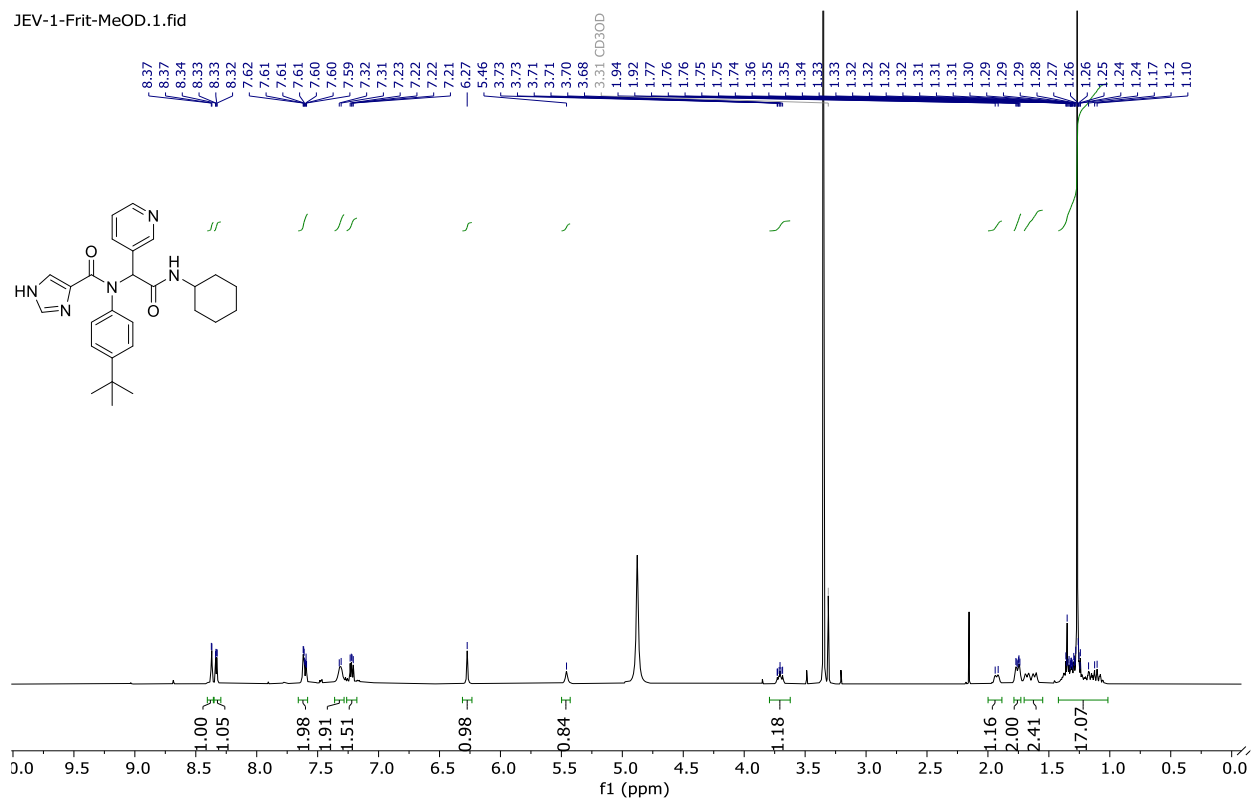
JEV-1-Frit-MeOD.2.fid
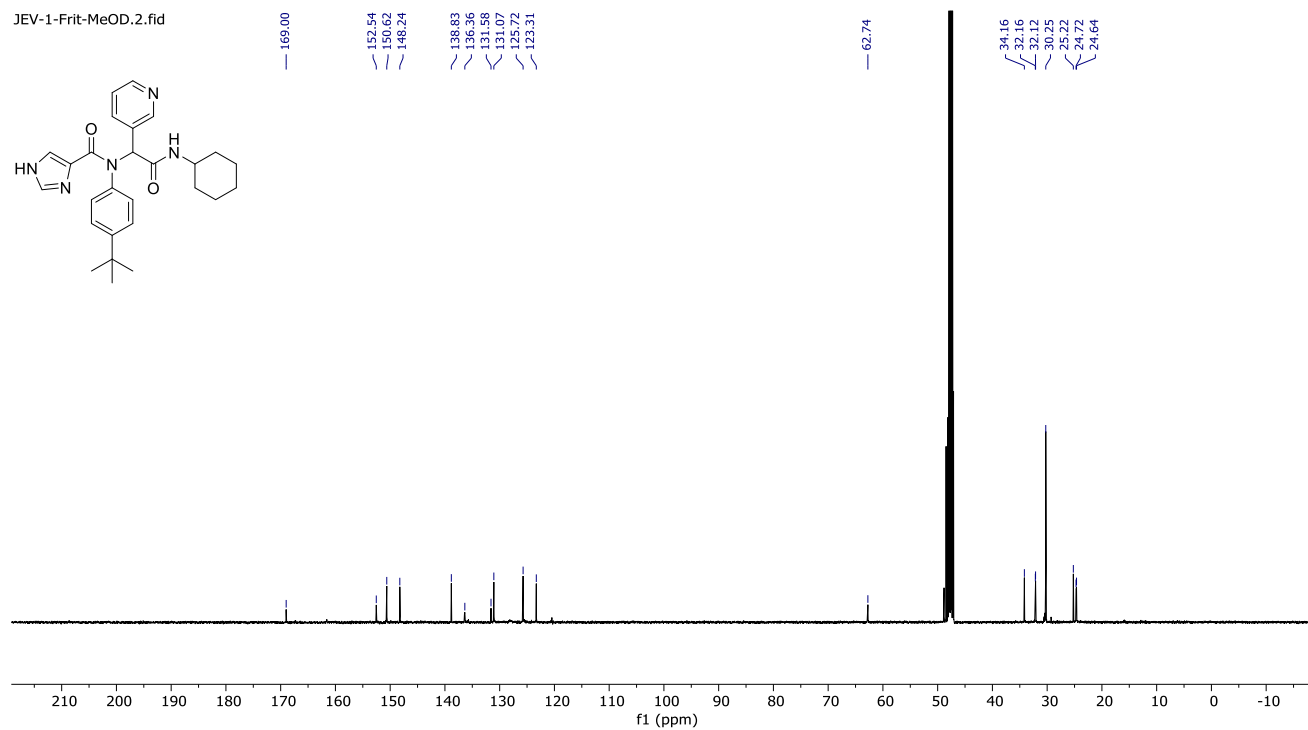
Compound 4.4 ¹H and ¹³C NMR

JT-29.2.fid
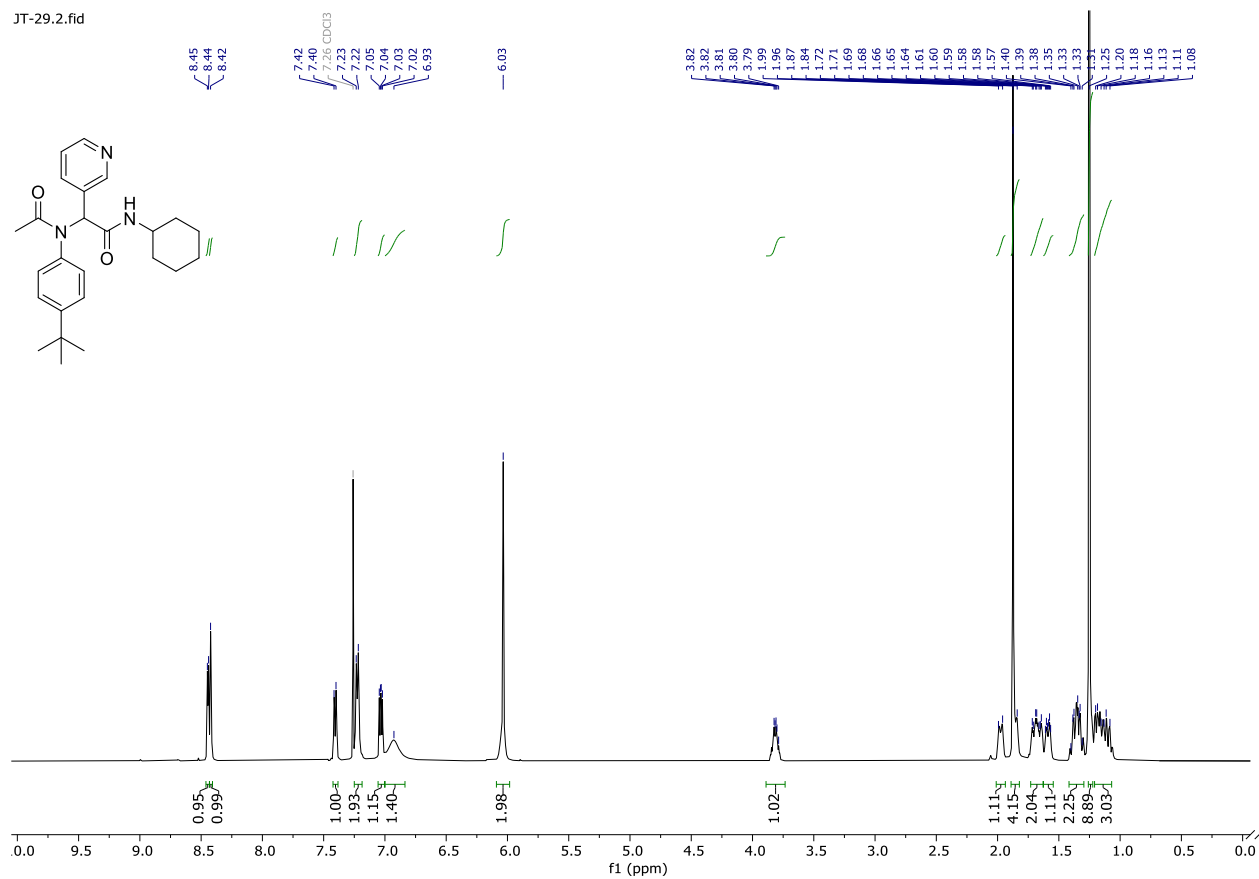
JT-29.4.fid
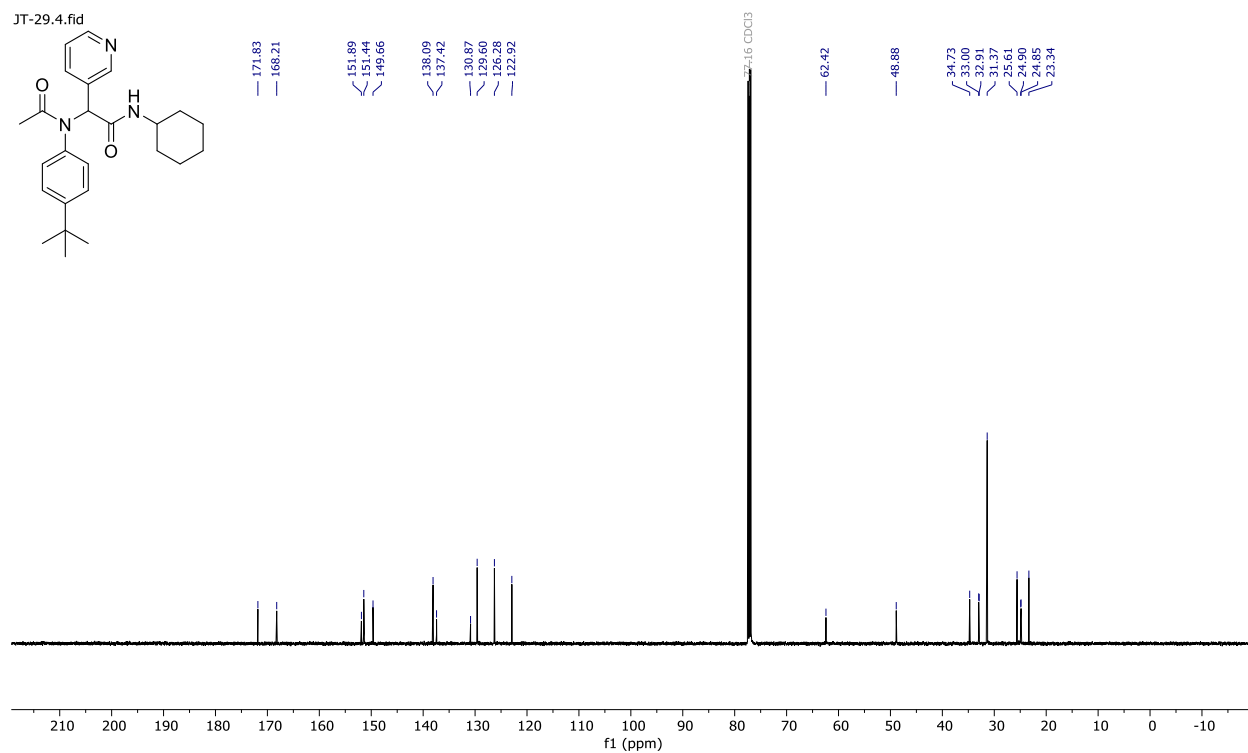
Compound 4.5 ^{1}H and ^{13}C NMR

JT-21.1.fid

JT-21.2.fid
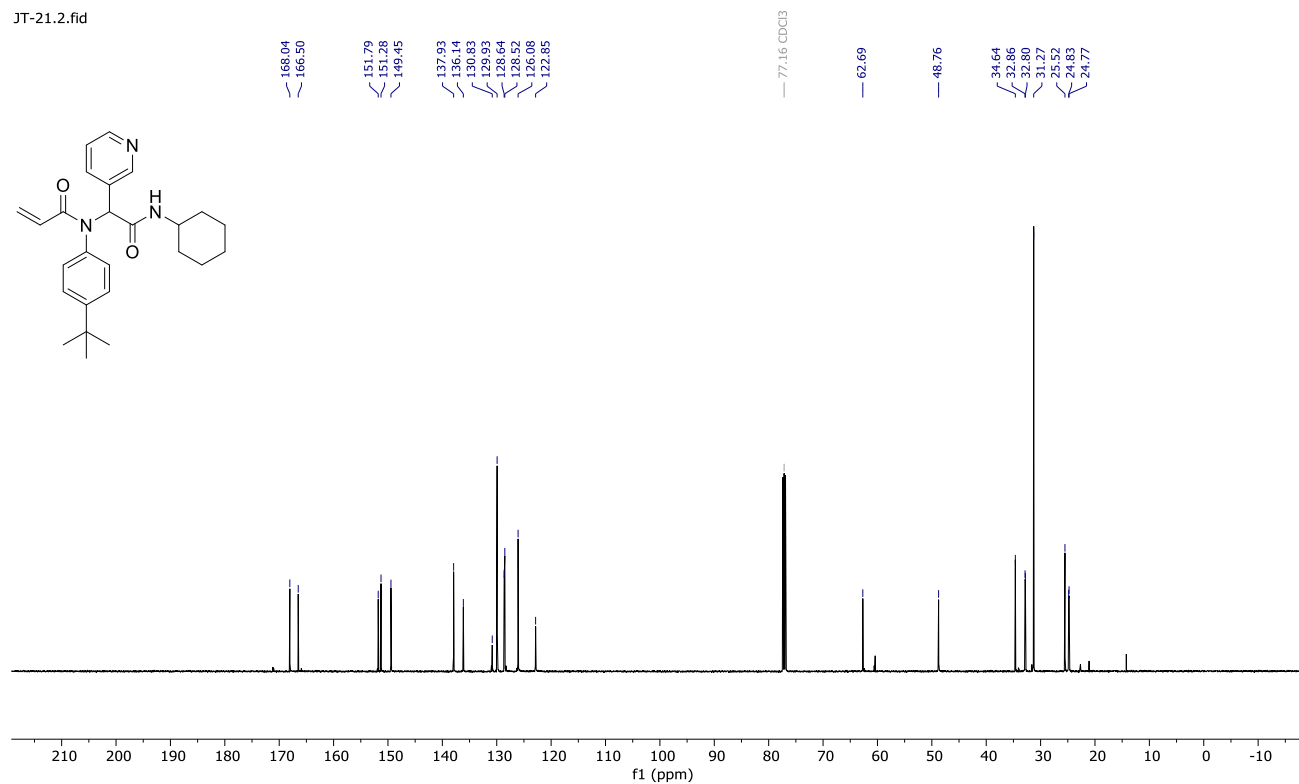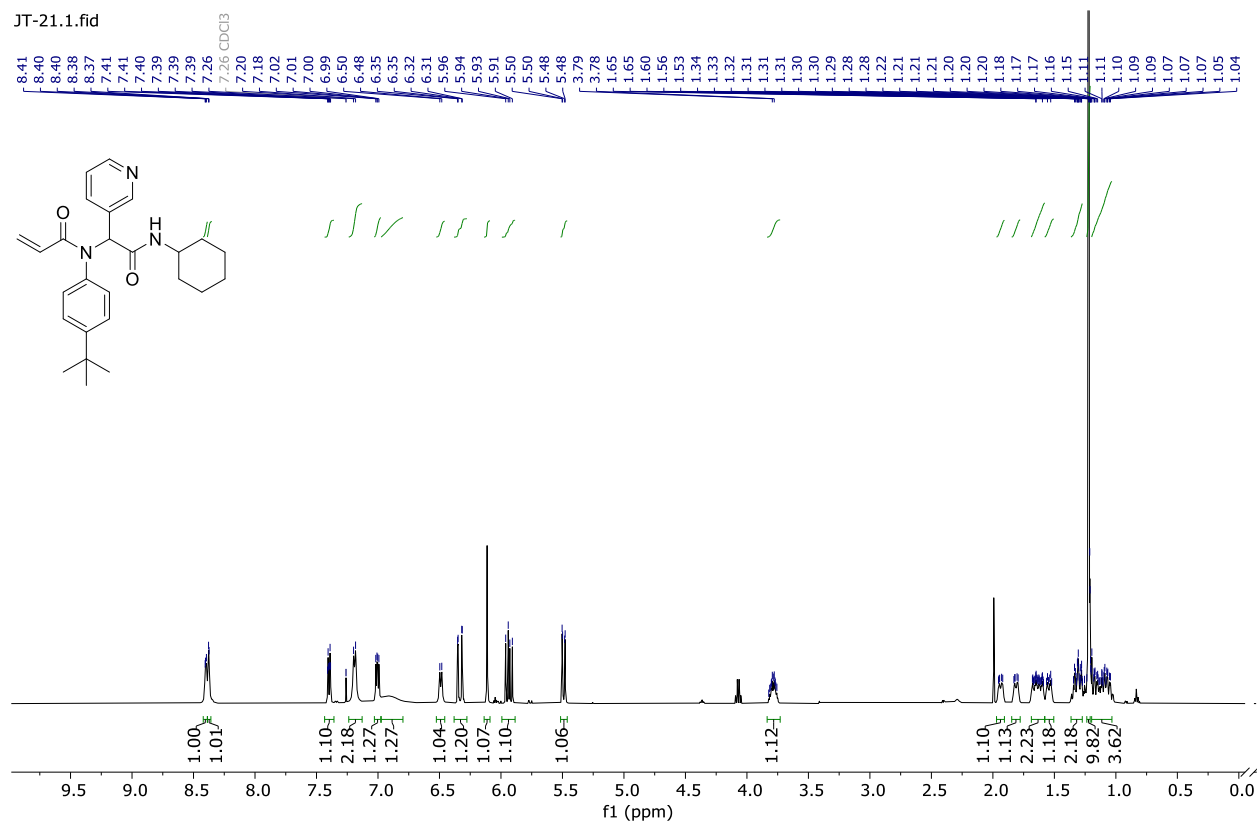
Compound 4.6 ^{1}H and ^{13}C NMR

JT-3.5.fid

JT-3.6.fid
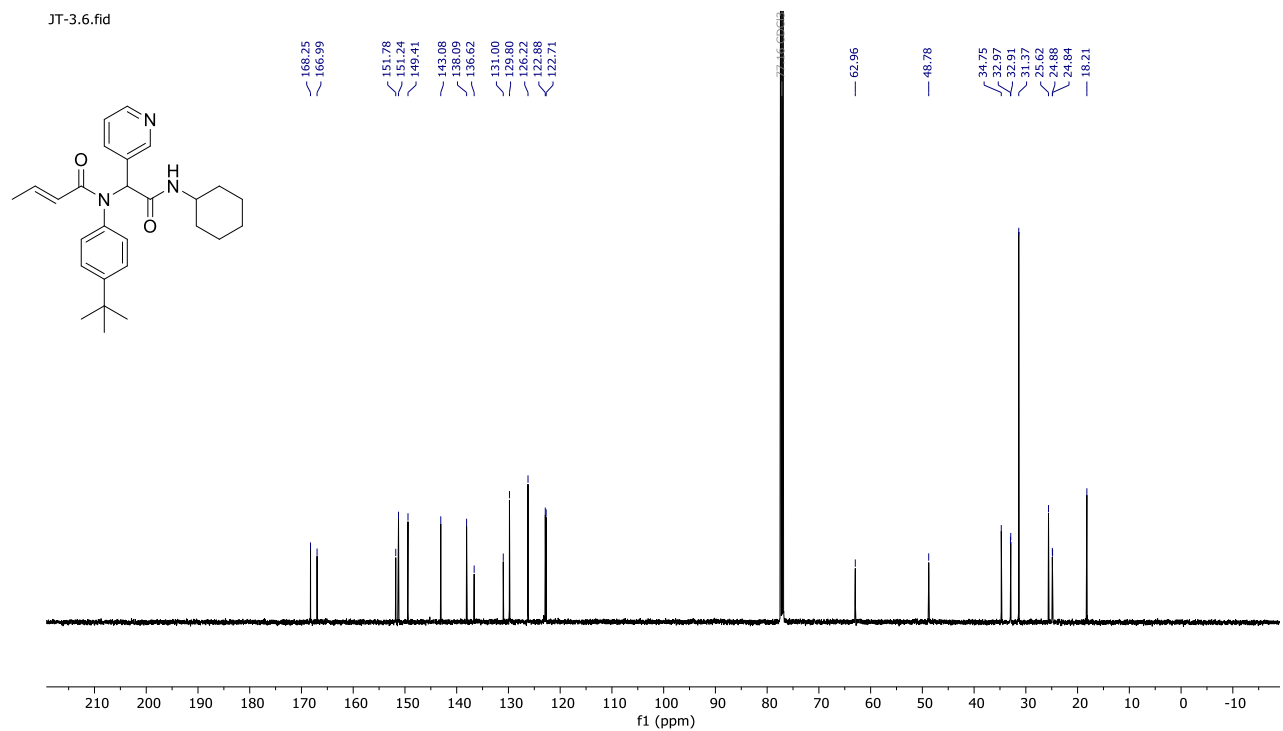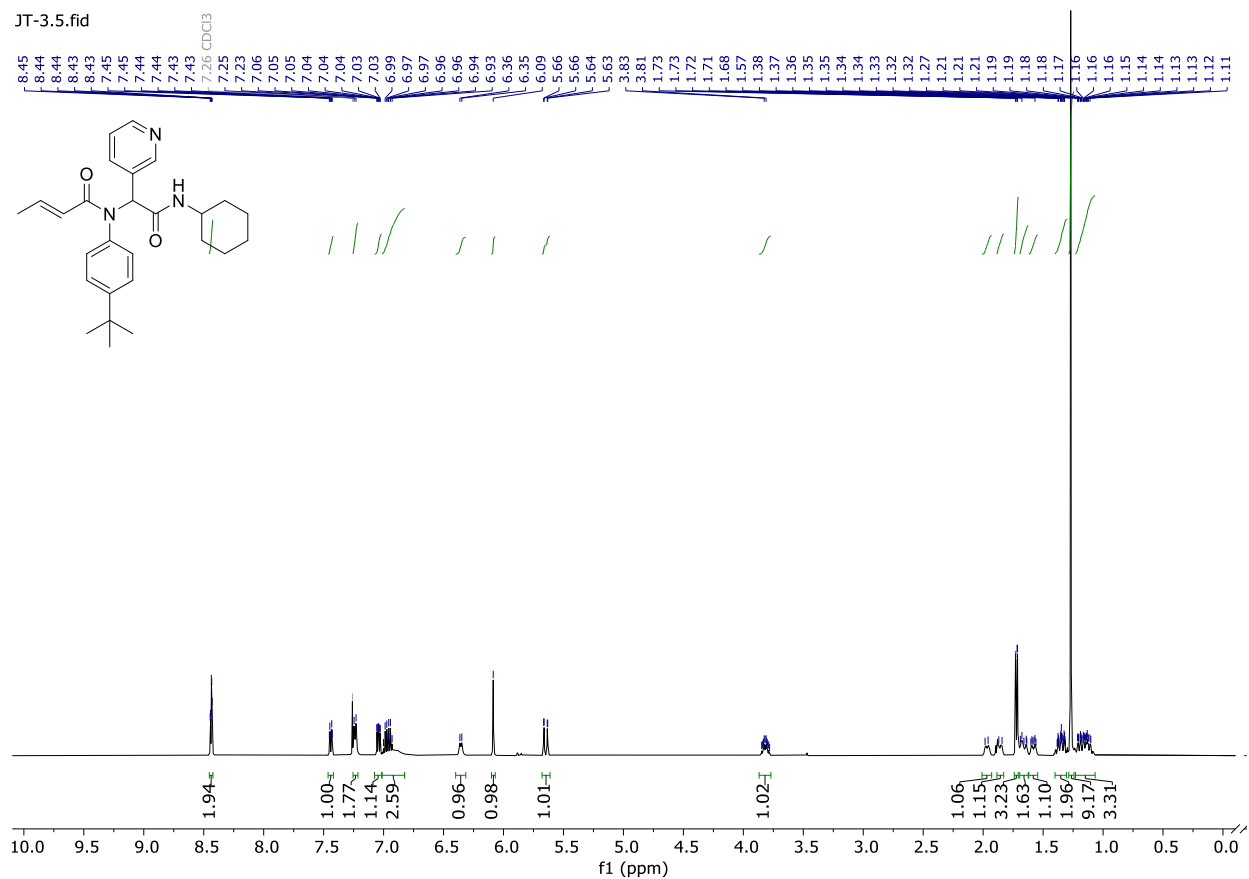
Compound 4.7 ^{1}H and ^{13}C NMR

JT-4.4.fid
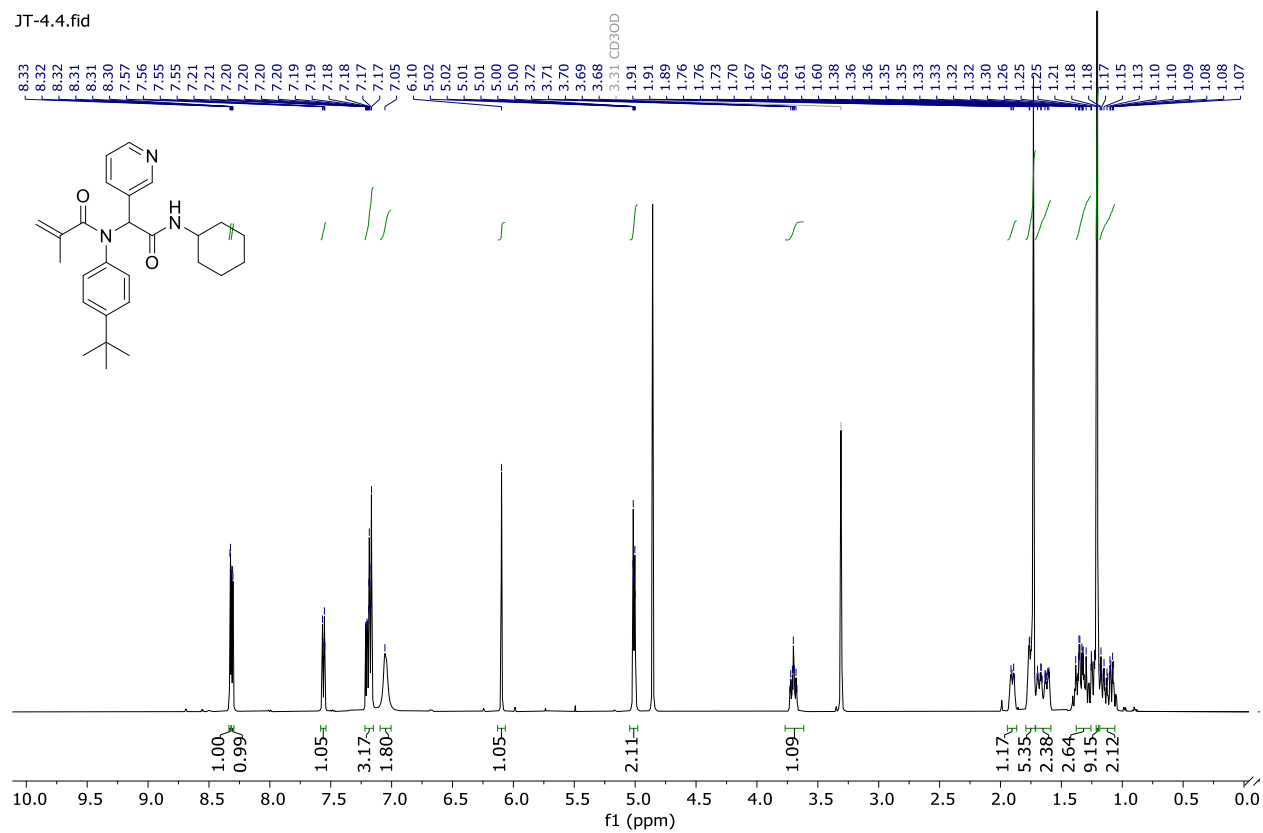
JT-4.10.fid
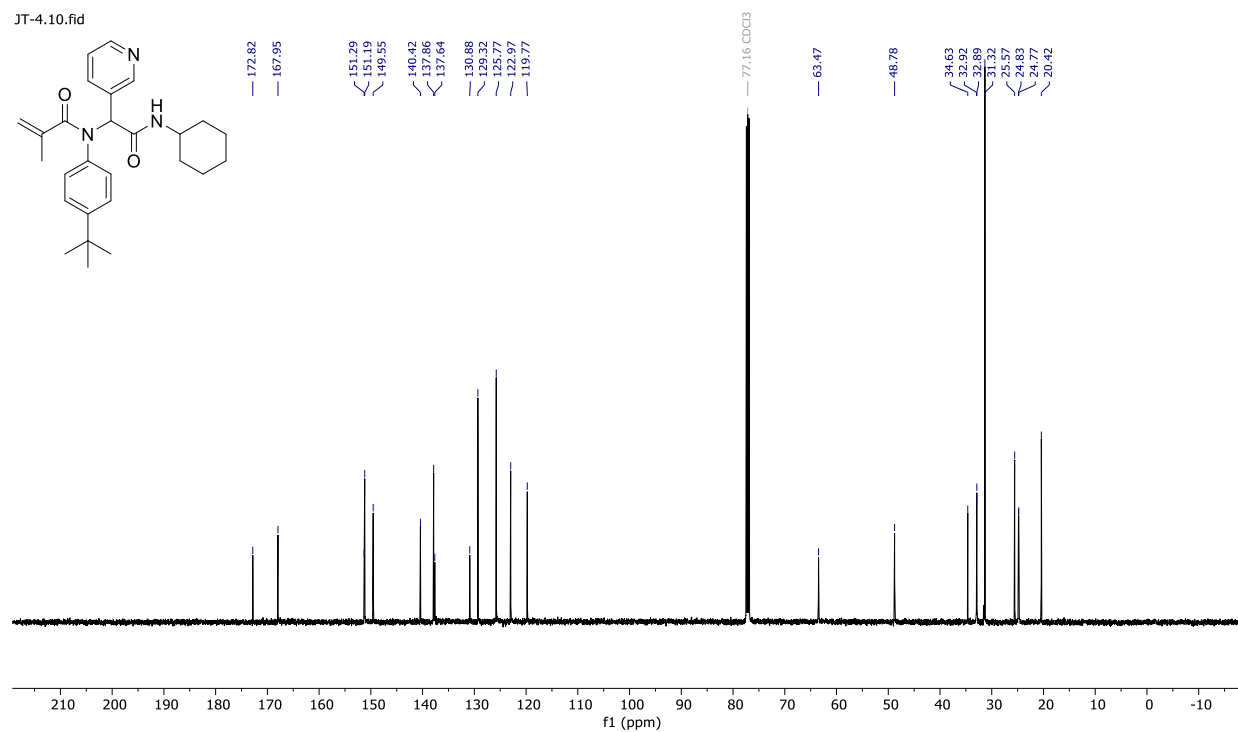
Compound 4.8 ¹H and ¹³C NMR

JT-14.11.fid

JT-14.12.fid
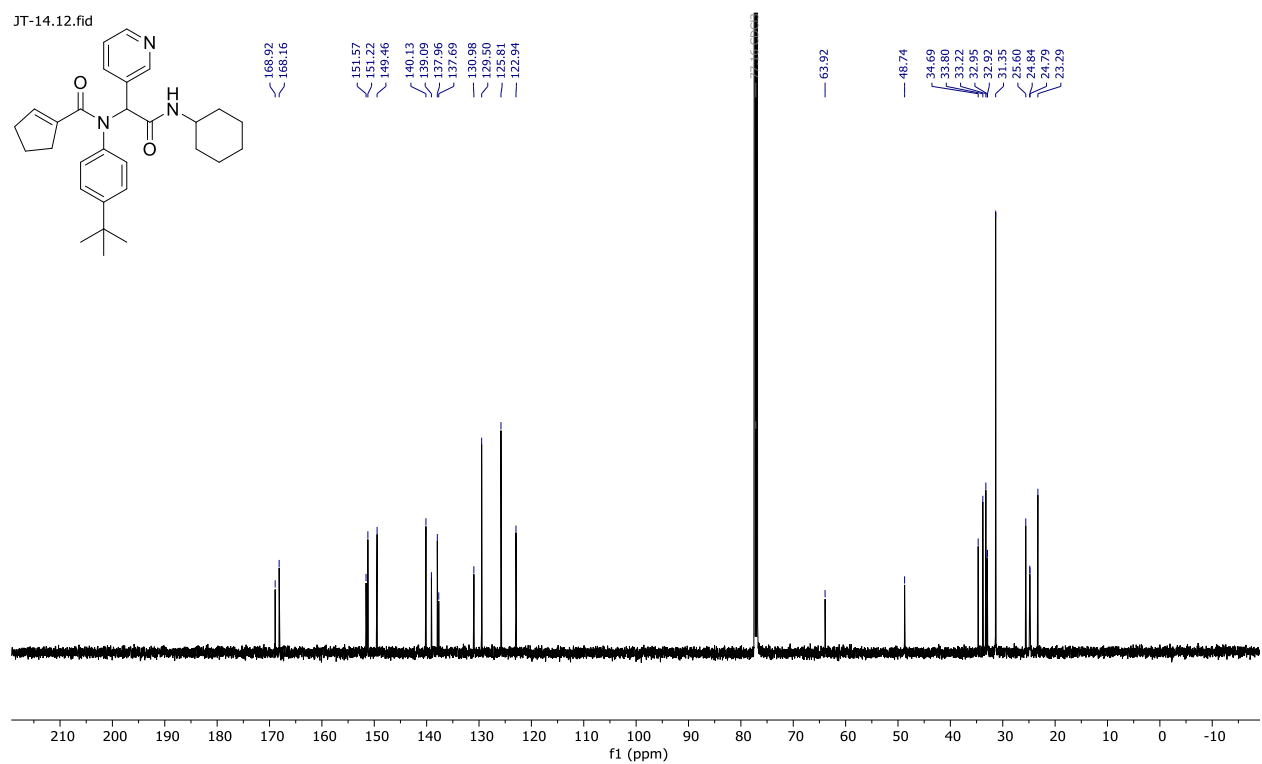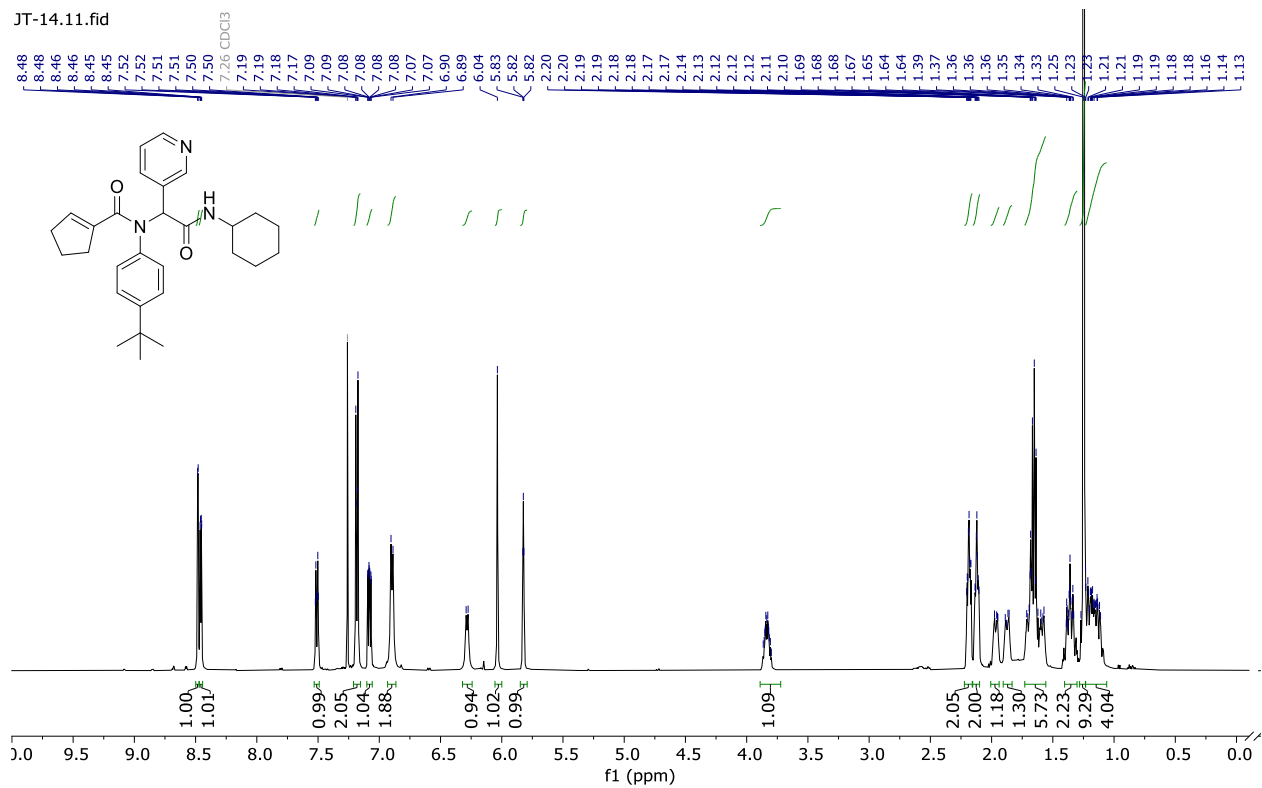
Compound 4.9 ^{1}H and ^{13}C NMR
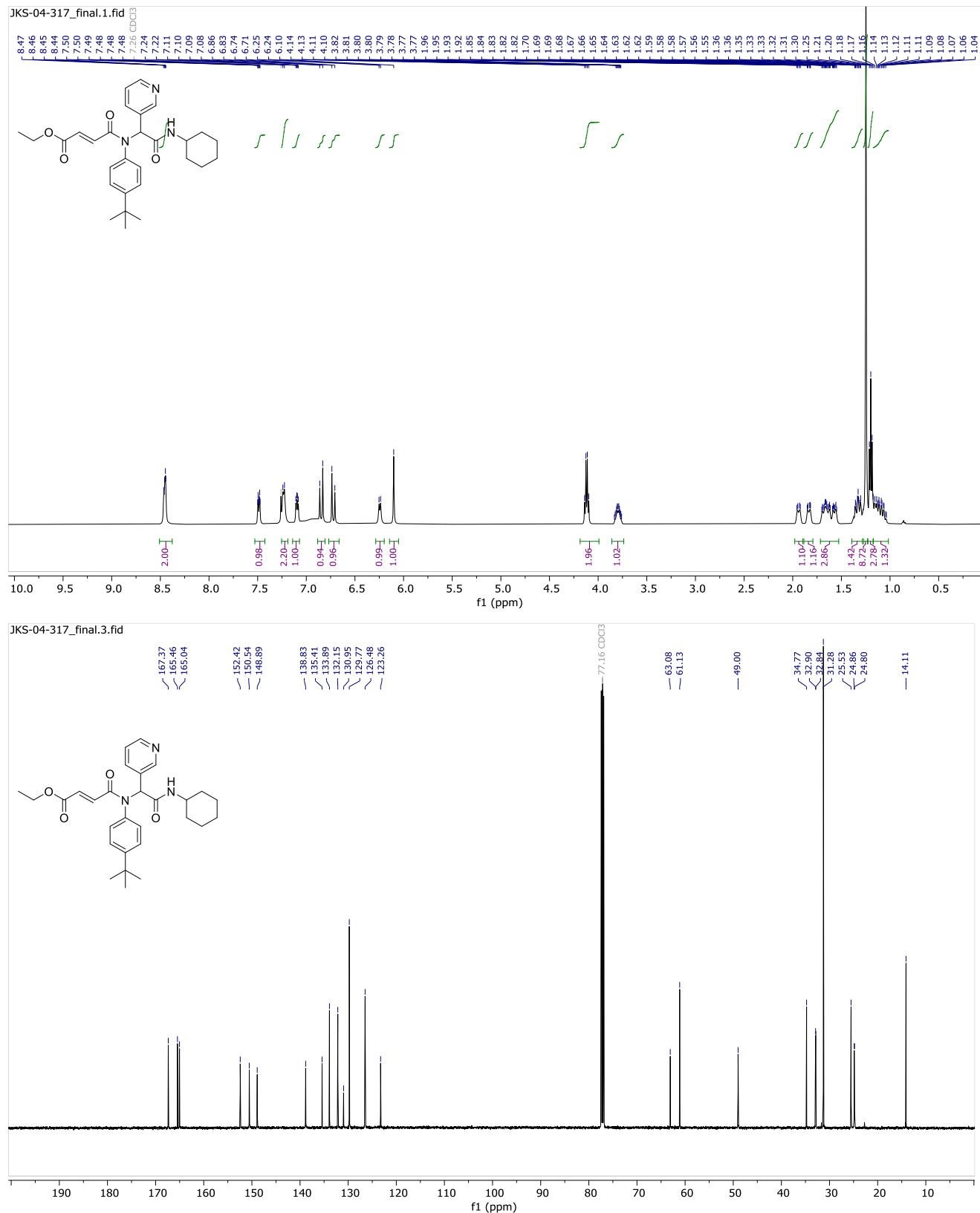
Compound 4.10 ¹H and ¹³C NMR

JT-5.5.fid

JT-5.6.fid
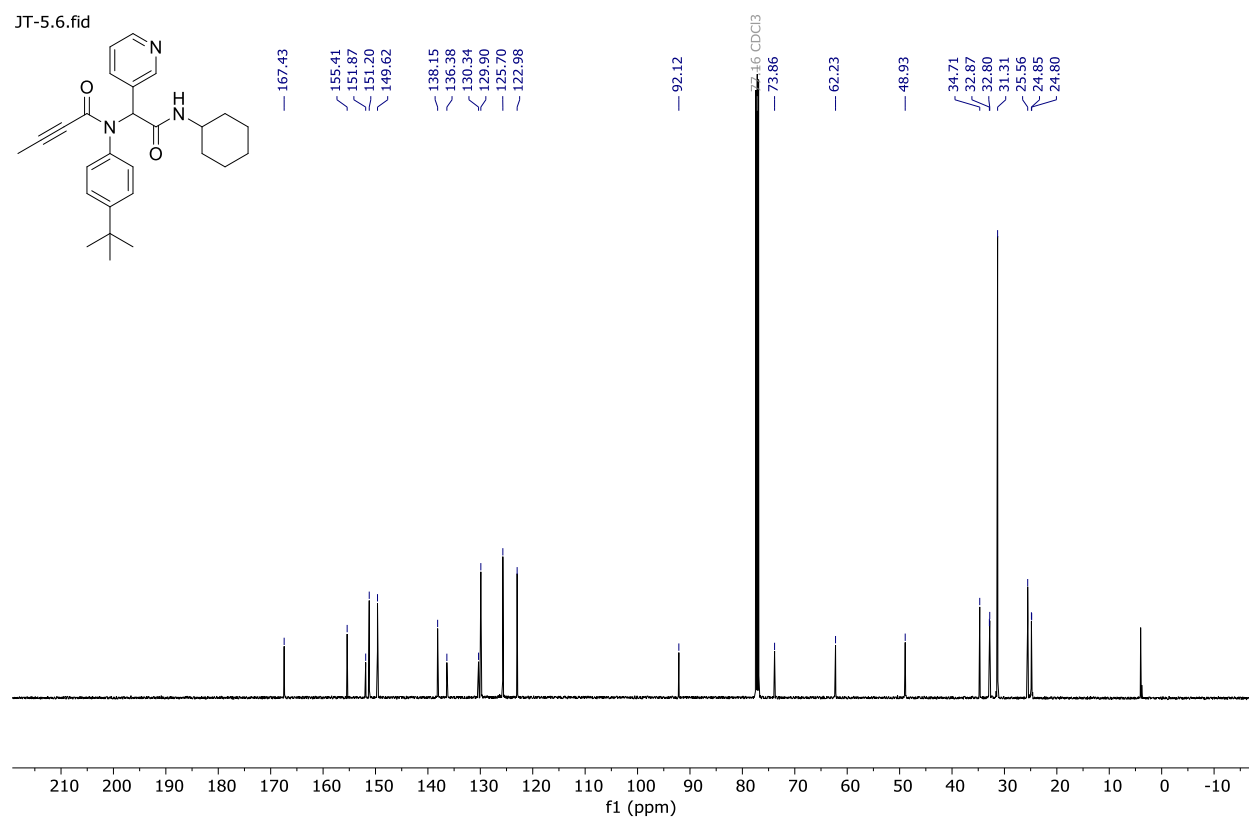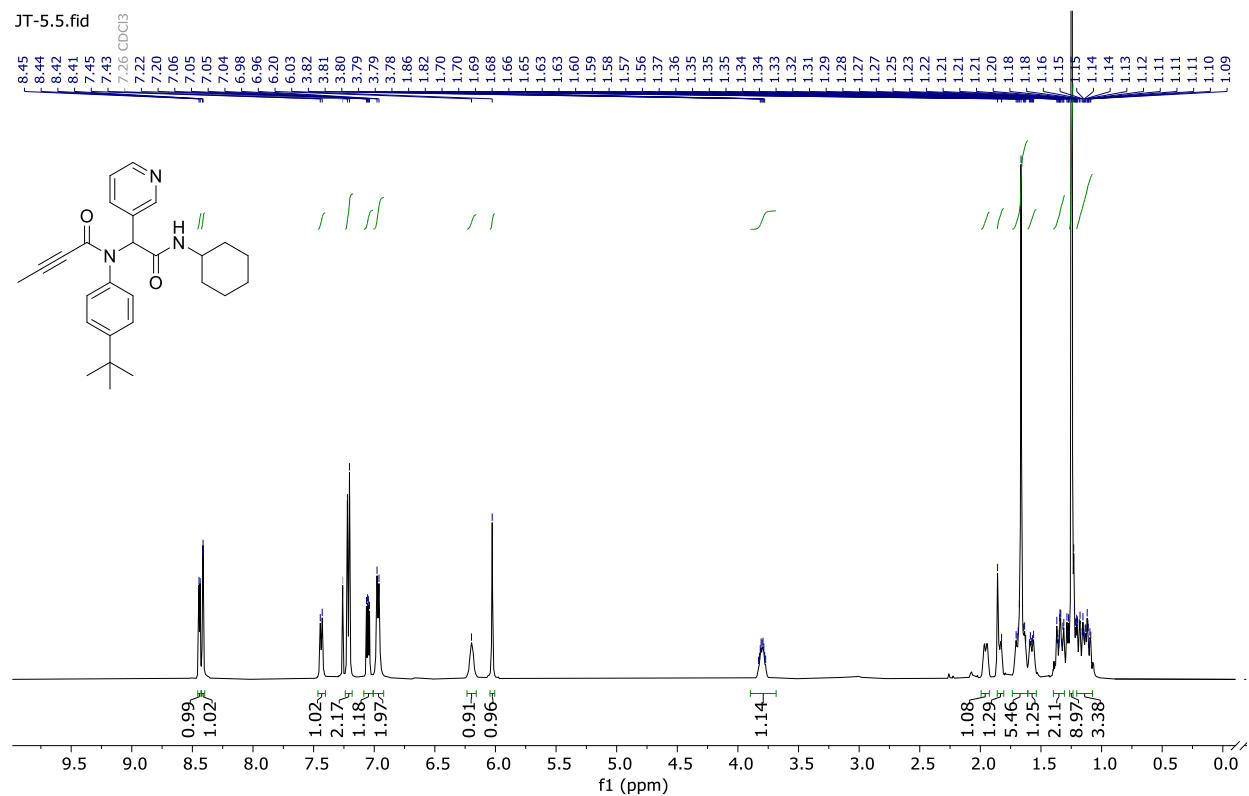
Compound 4.11 ^{1}H and ^{13}C NMR
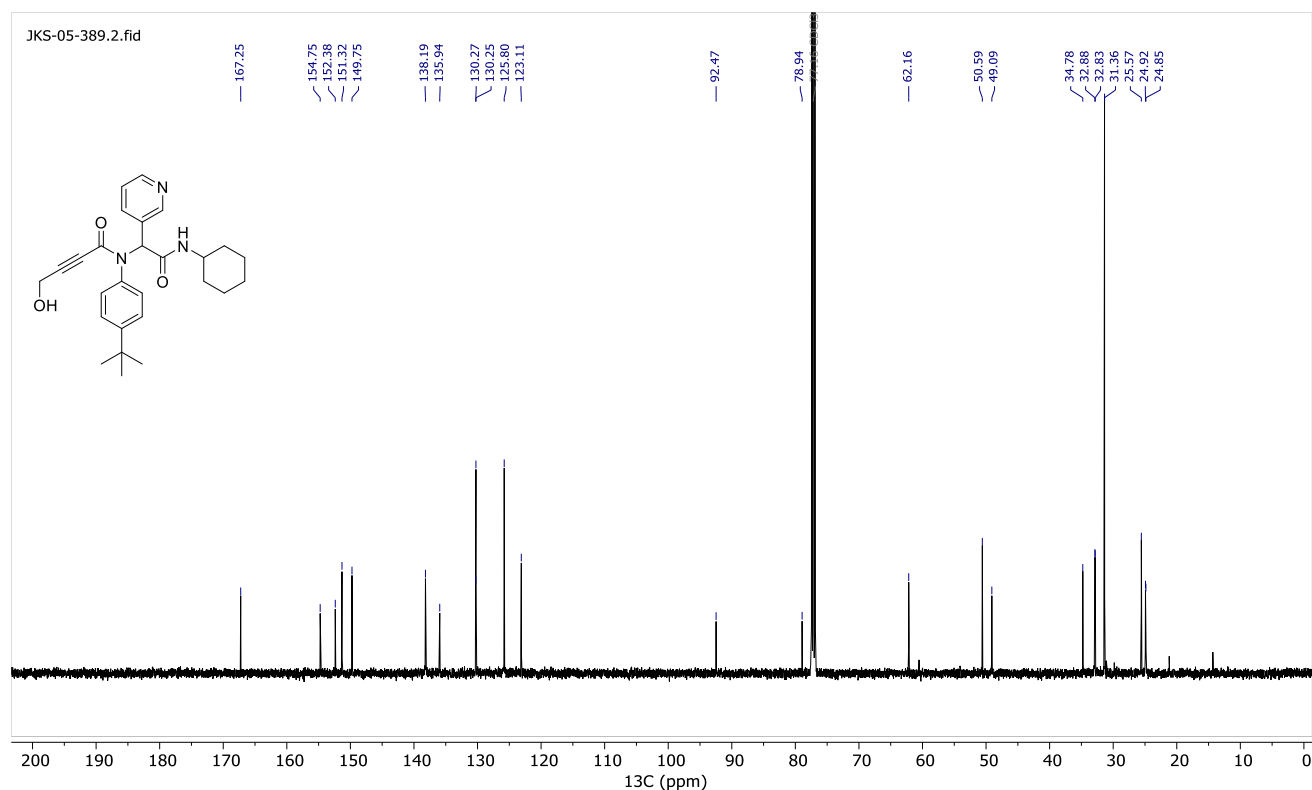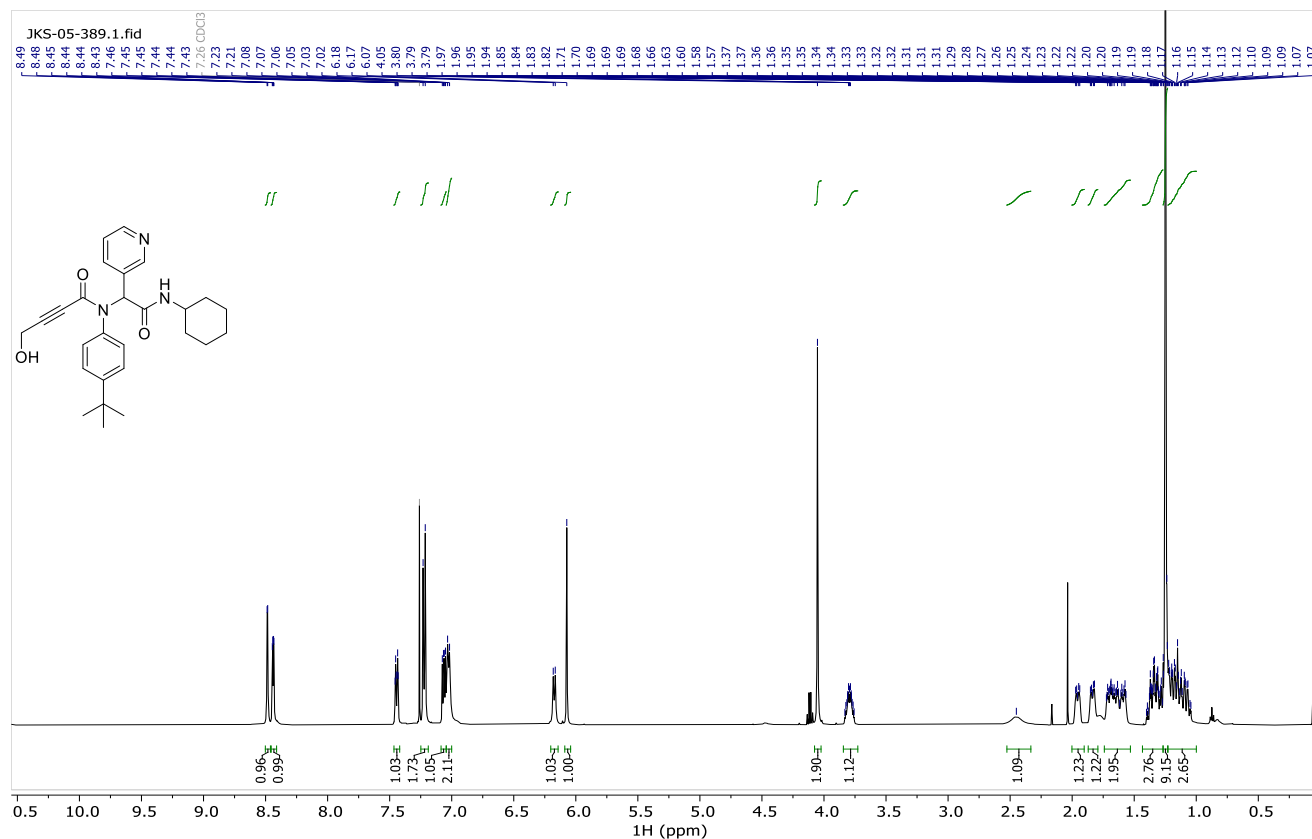
Compound 4.12 ^{1}H and ^{13}C NMR

JT-45.10.fid

JT-45.12.fid
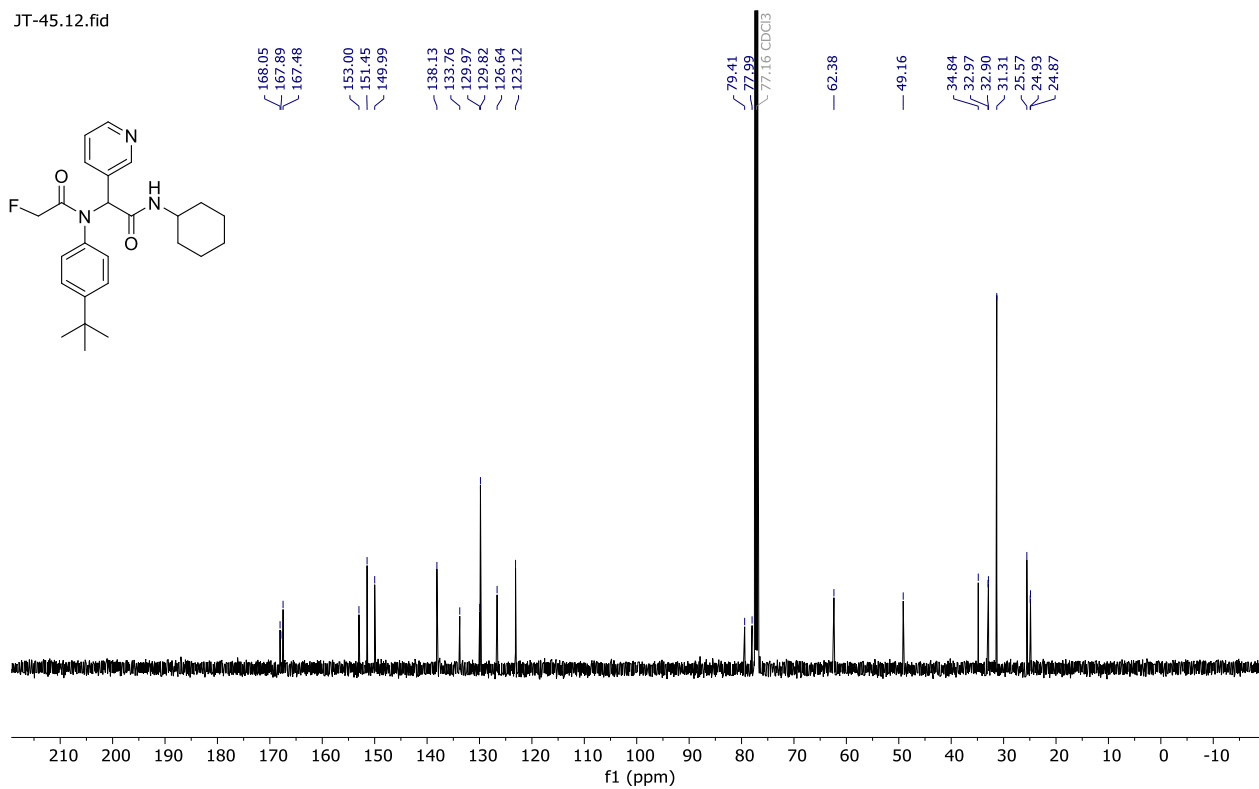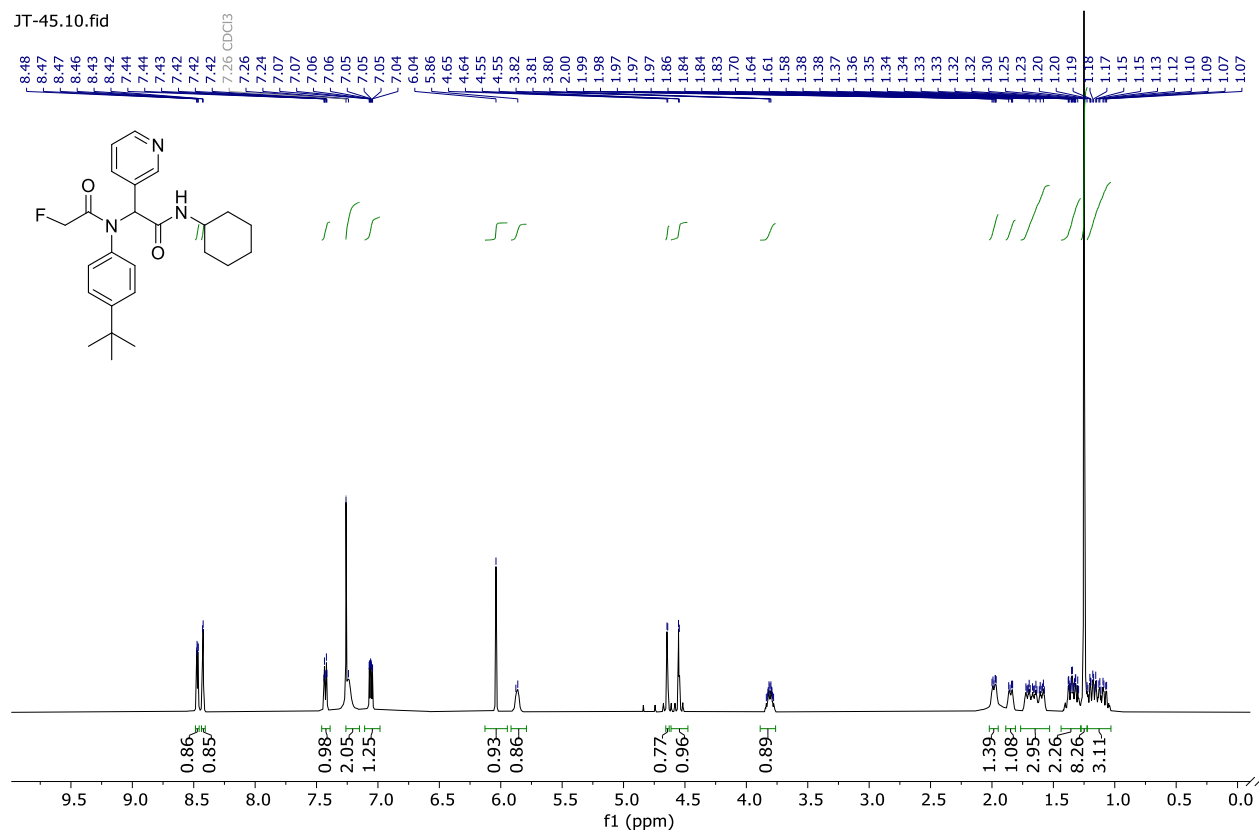
Compound 4.13 ^{1}H and ^{13}C NMR

JT-11.11.fid

JT-11.12.fid
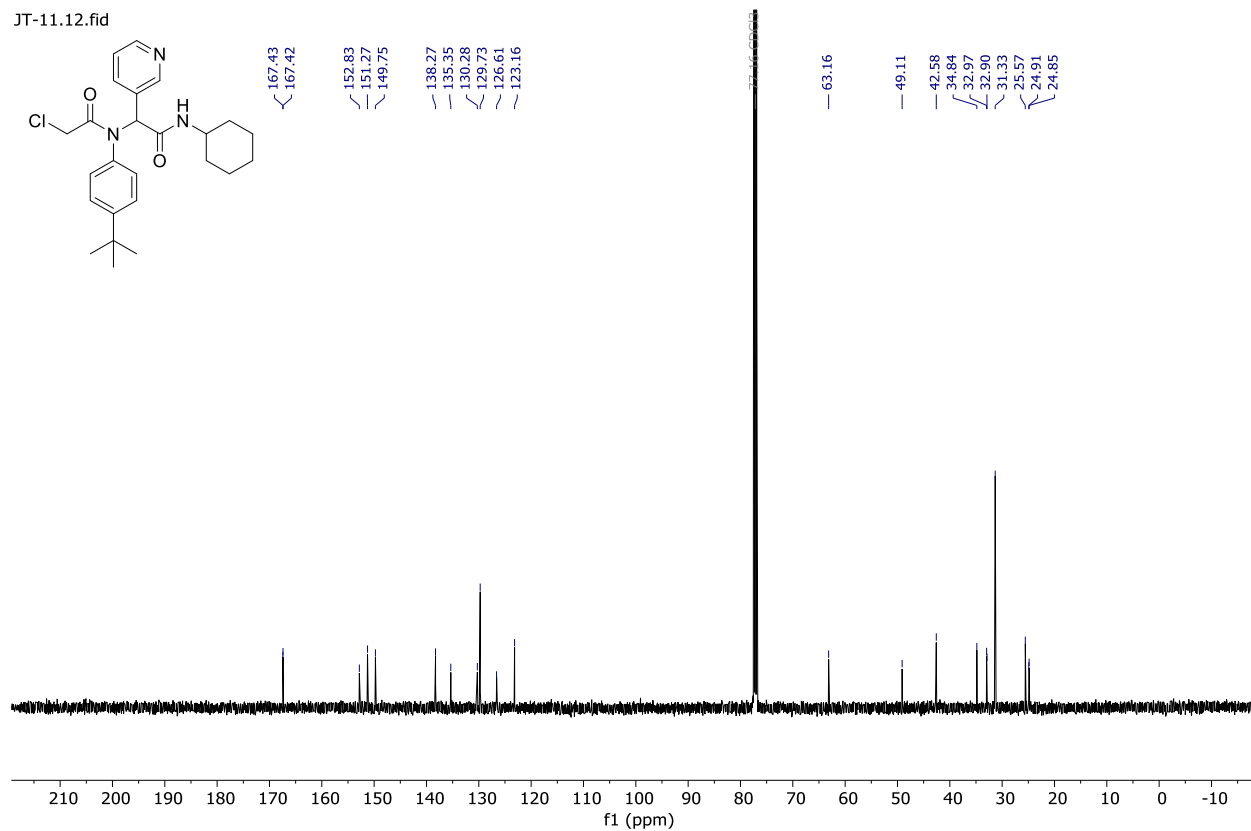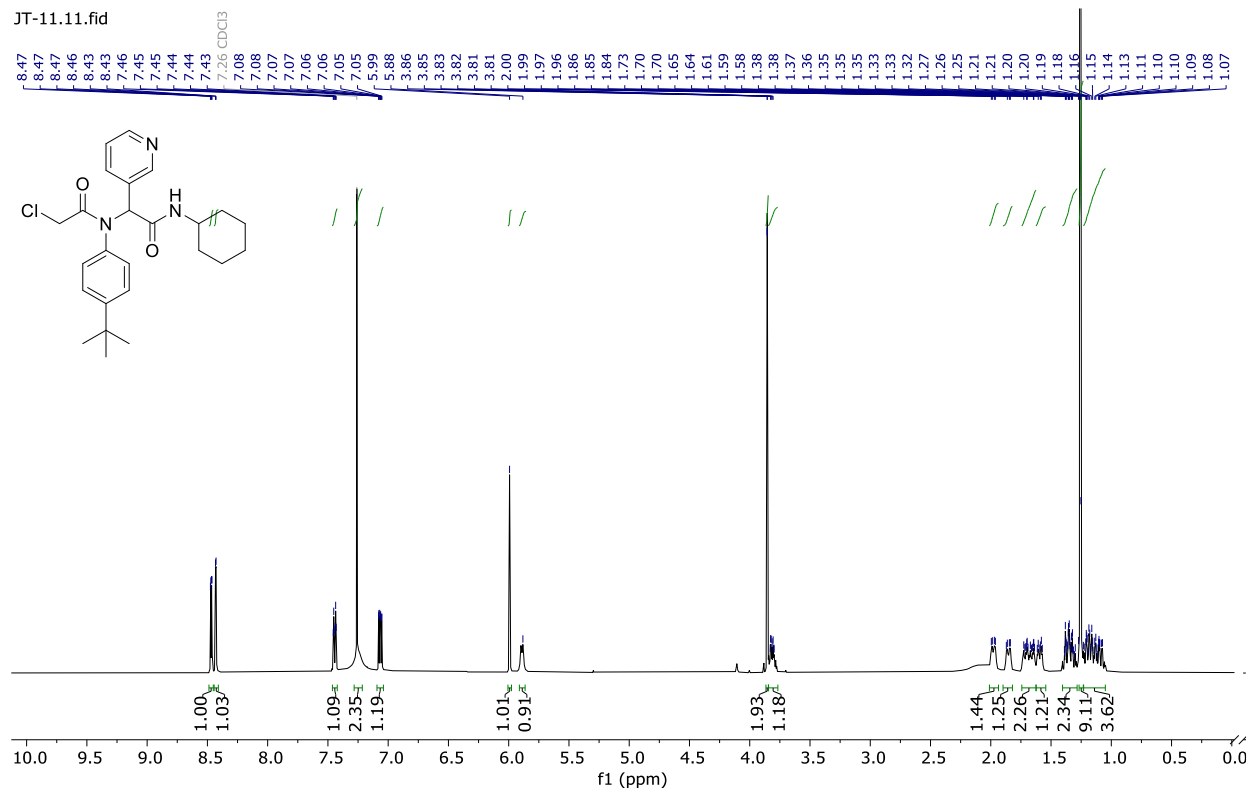
Compound 4.14 ^{1}H and ^{13}C NMR
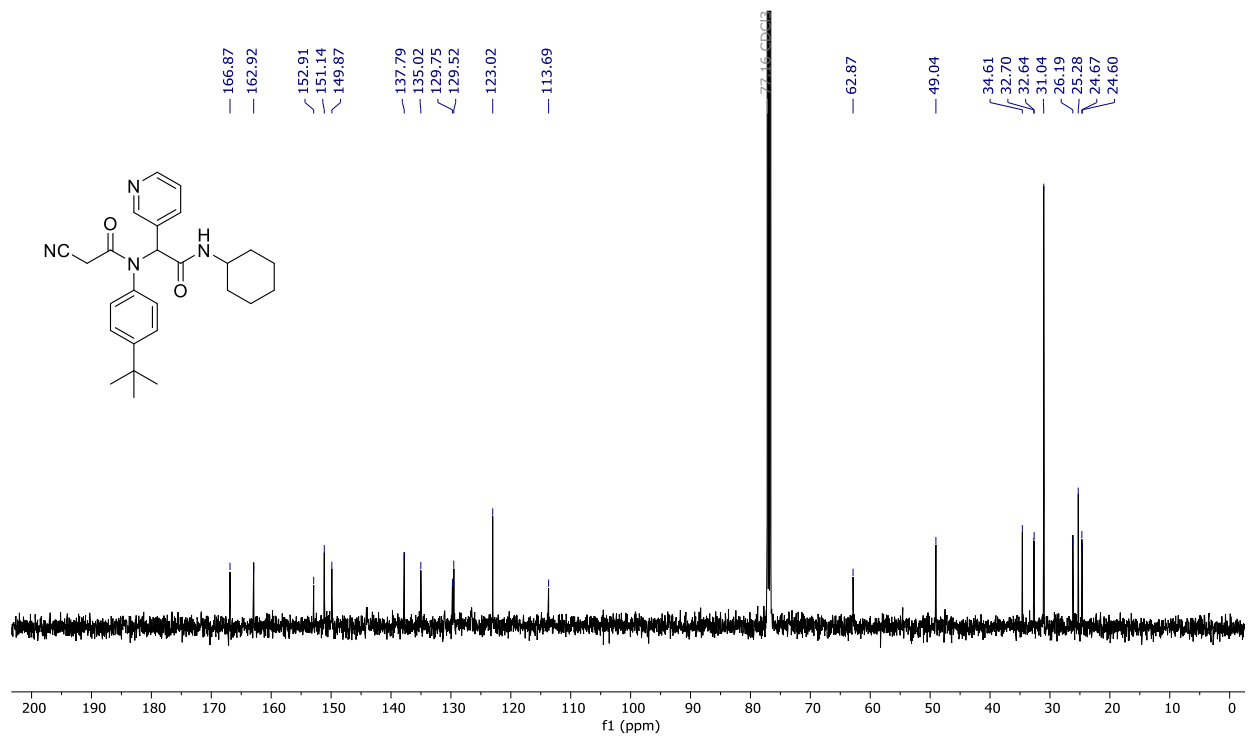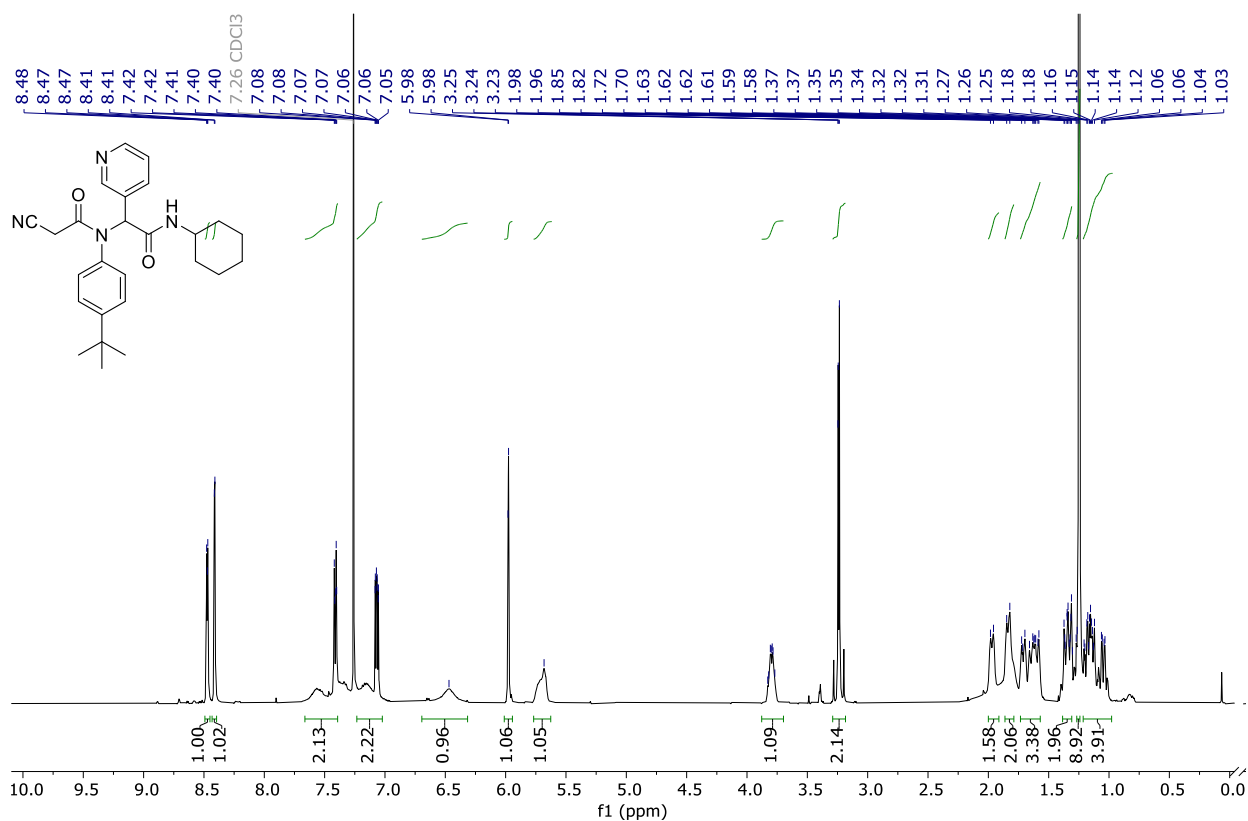
Compound 4.15 ¹H and ¹³C NMR

JKS-04-290_2.1.fid
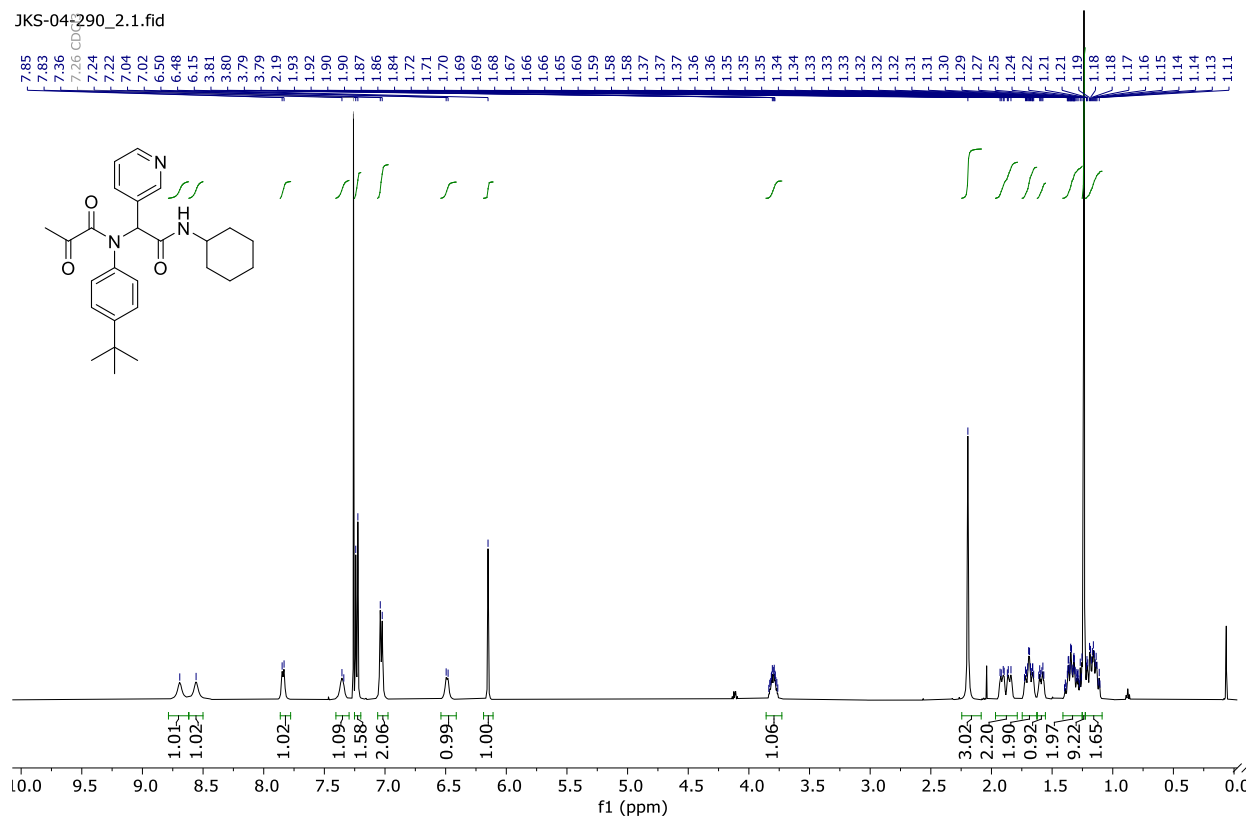
JKS-04-290_final.3.fid

Compound 4.16 ^{1}H and ^{13}C NMR

Compound 4.16 ^{19}F NMR

JKS-05-356.4.fid

−74.82

f1 (ppm)
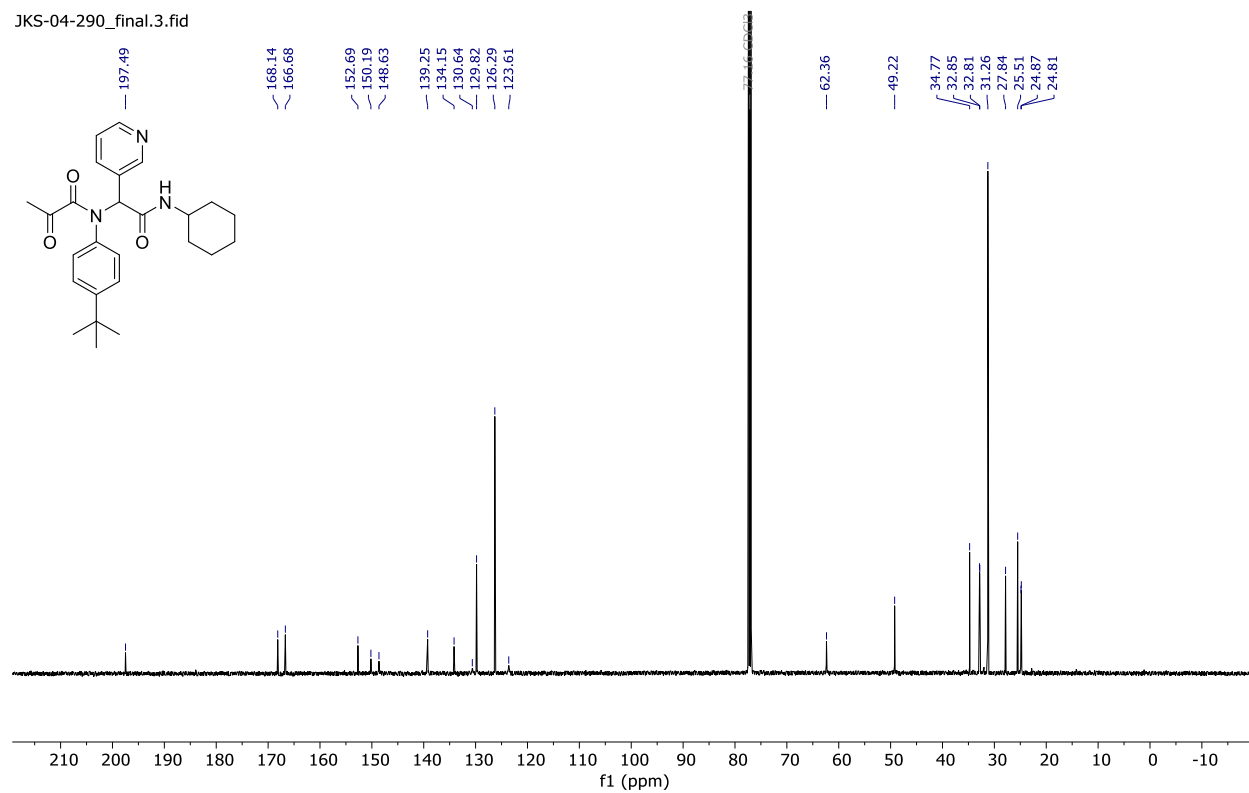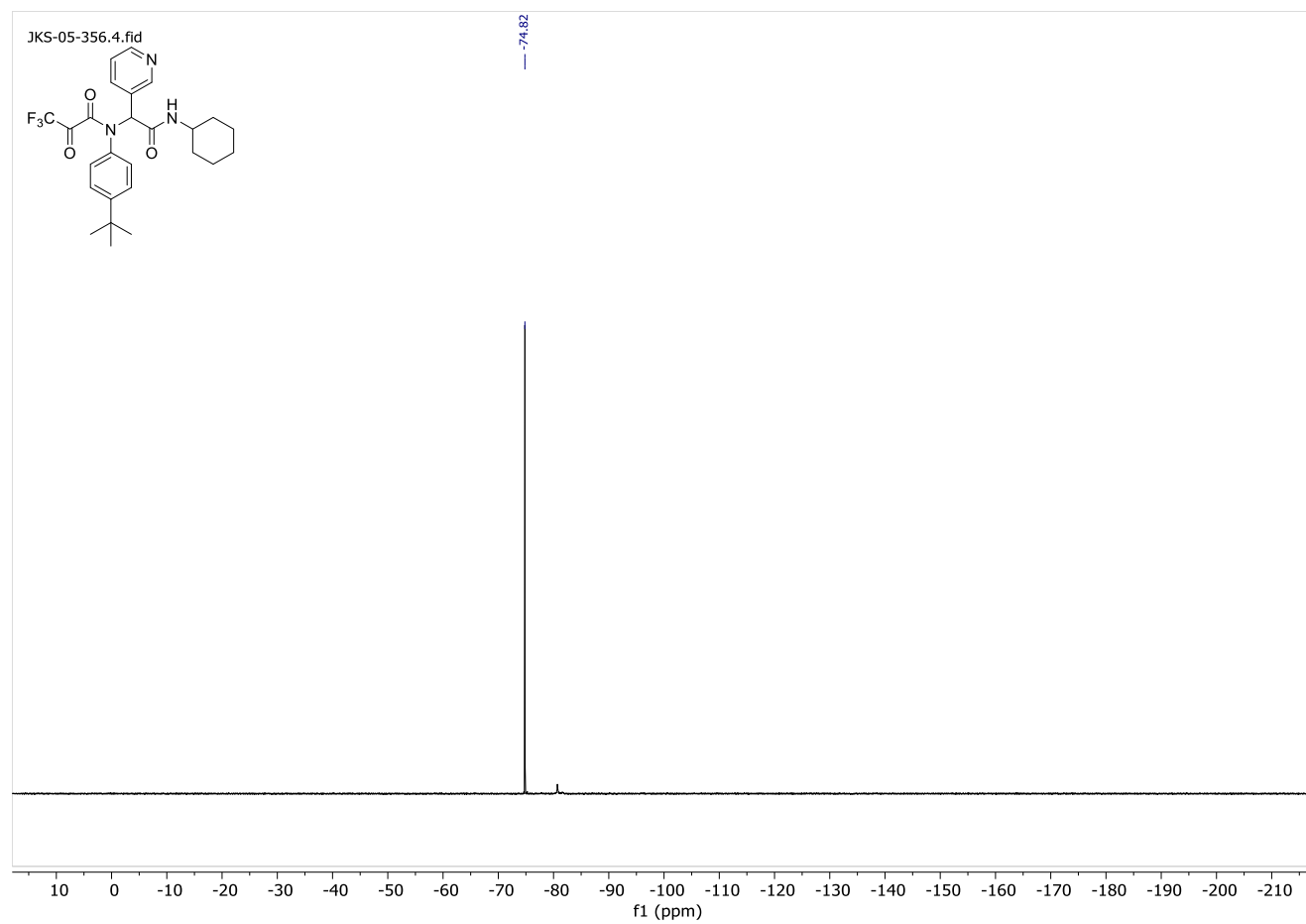
Compound 4.17 ^{1}H and ^{13}C NMR

JT-13.31.fid
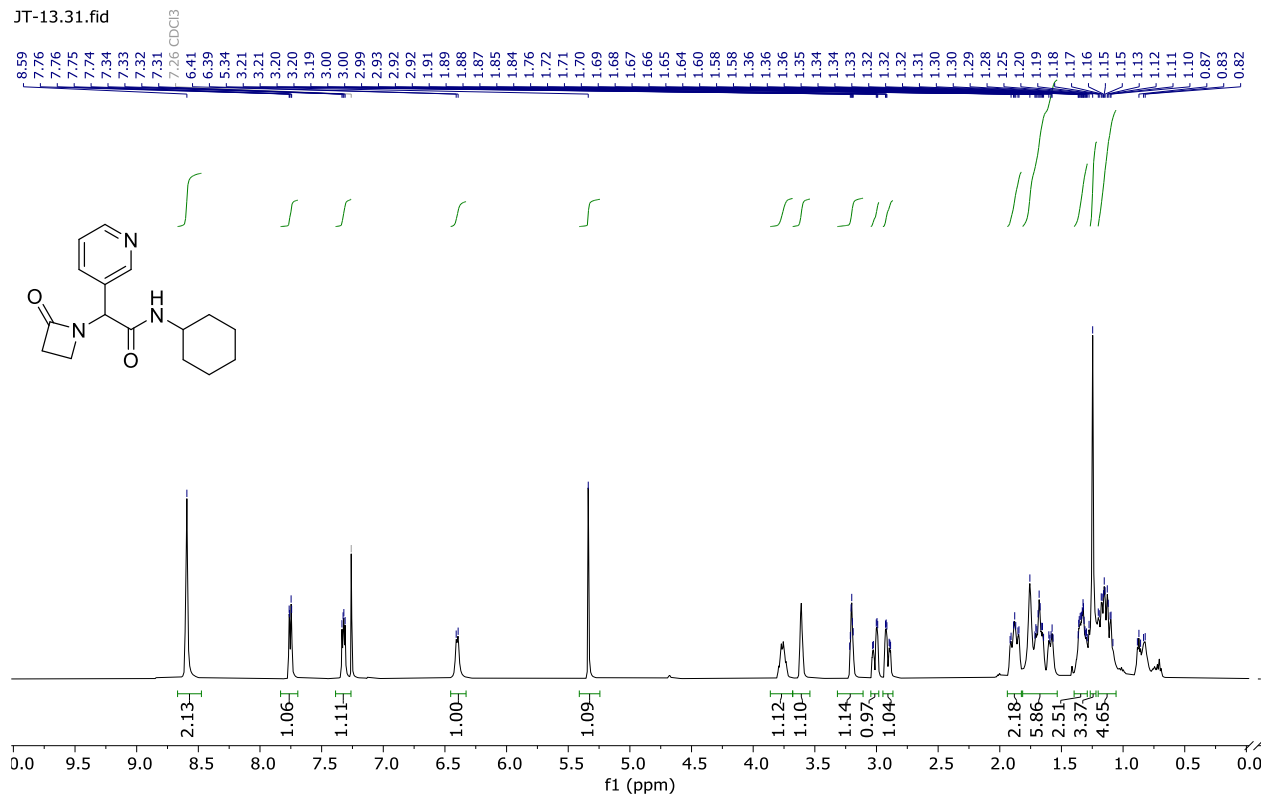
JT-13.32.fid
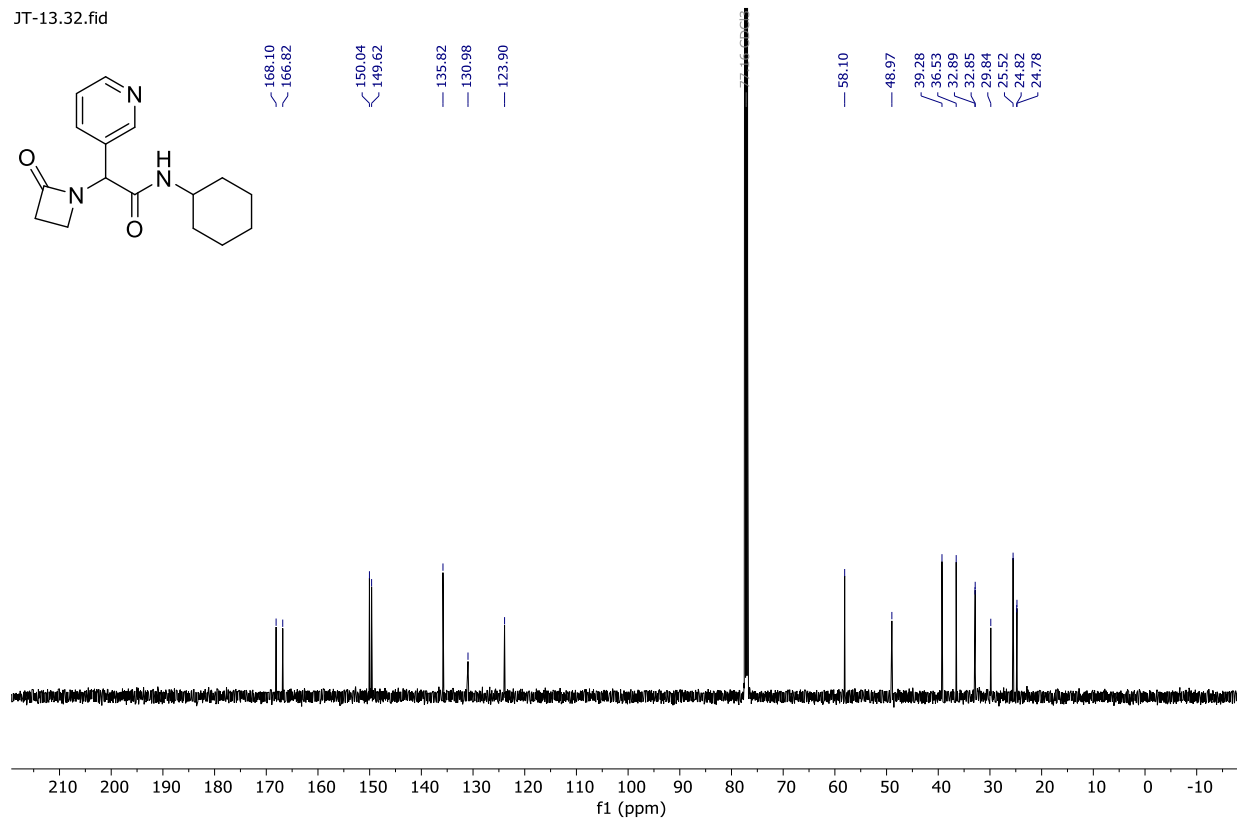
Compound 4.18 ^{1}H and ^{13}C NMR
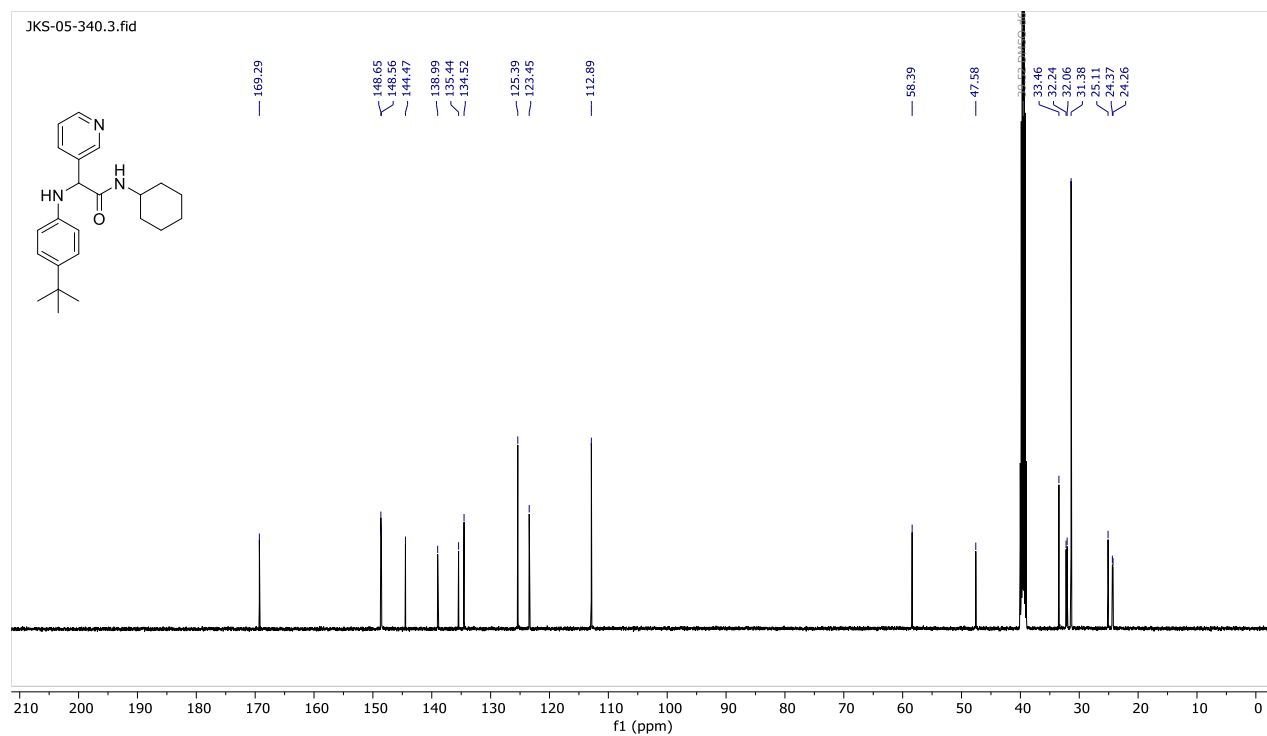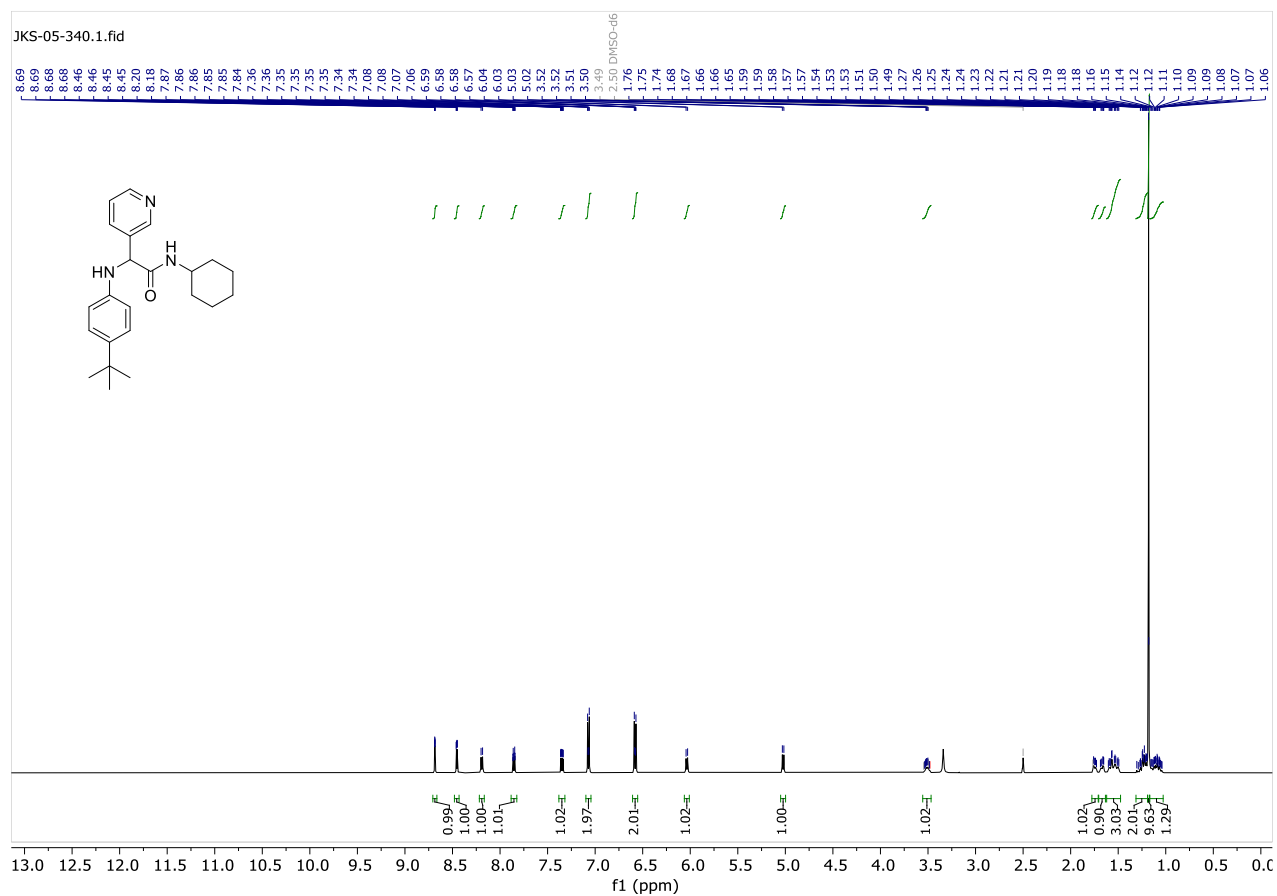
Compound 4.19 ¹H and ¹³C NMR

JKS-05-337.1.fid

JKS-05-337.3.fid
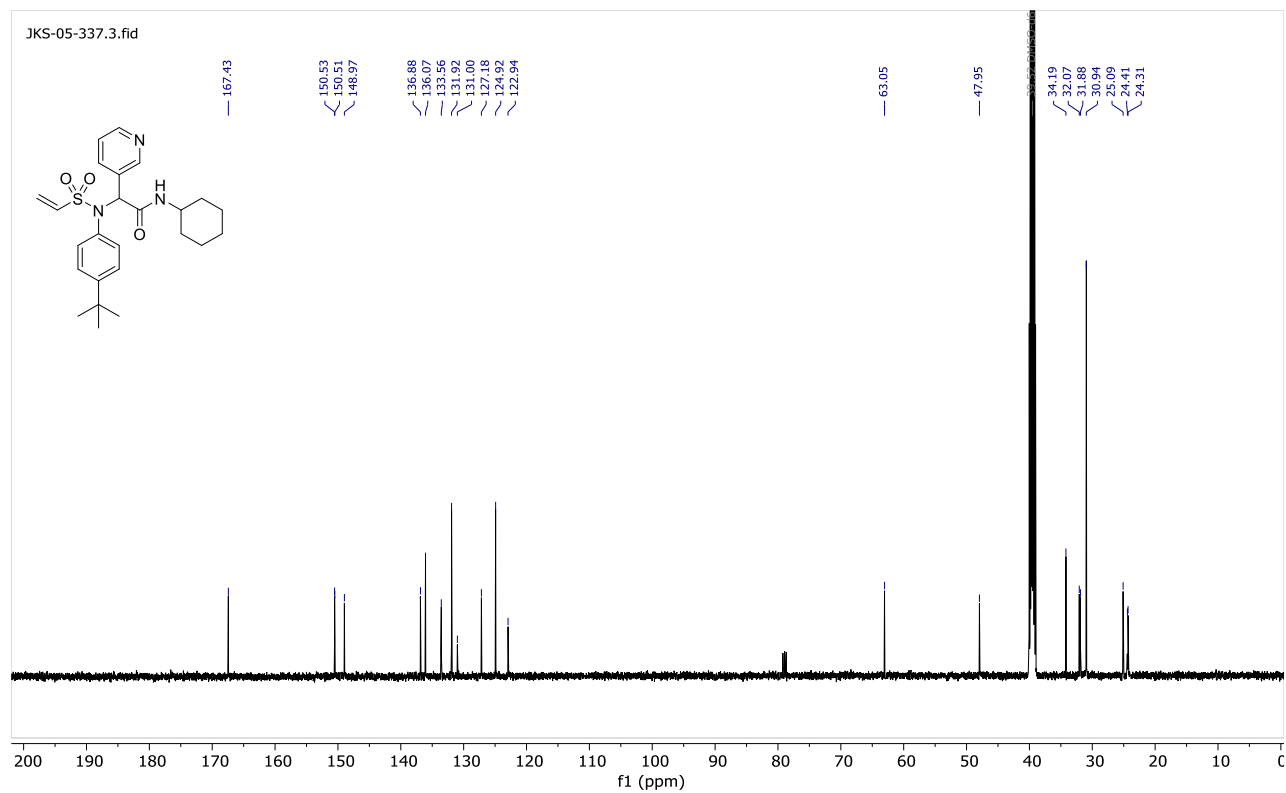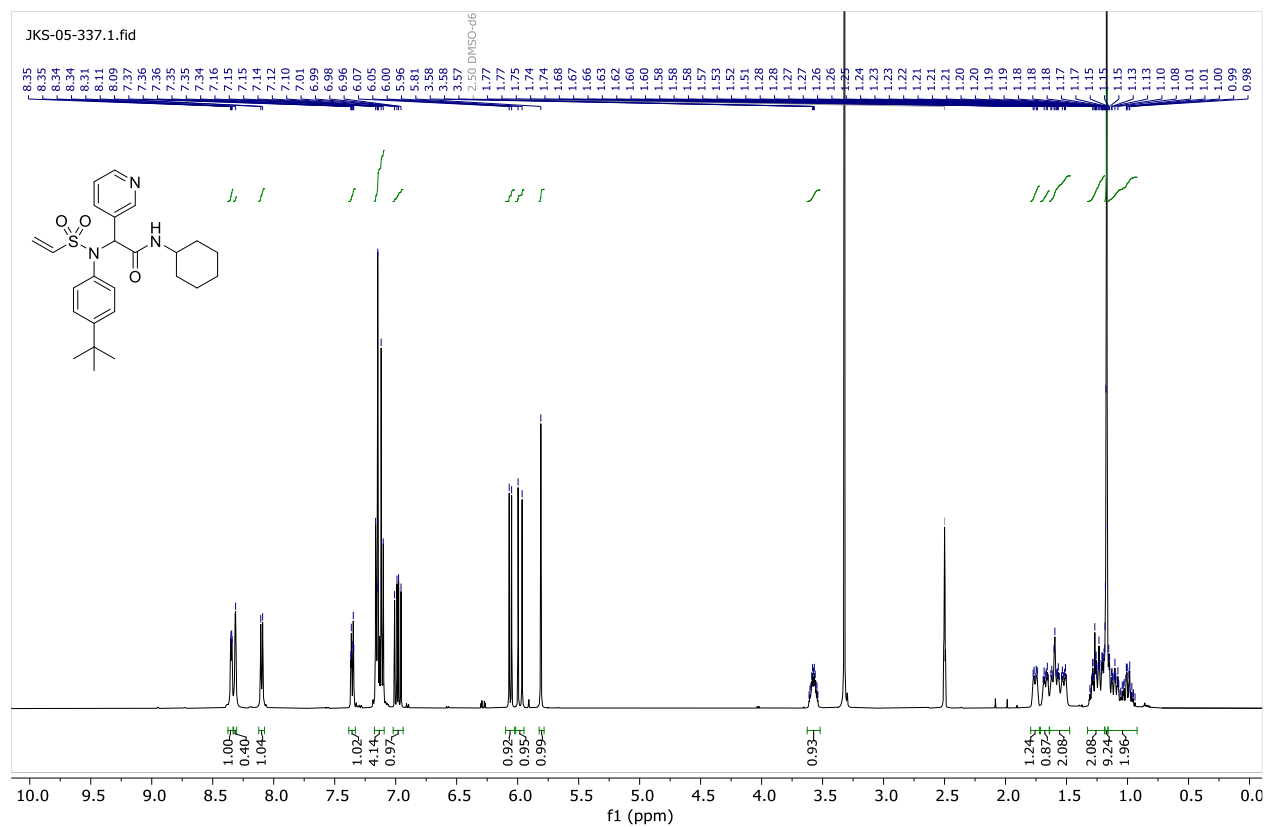
Compound 4.20 ^{1}H and ^{13}C NMR
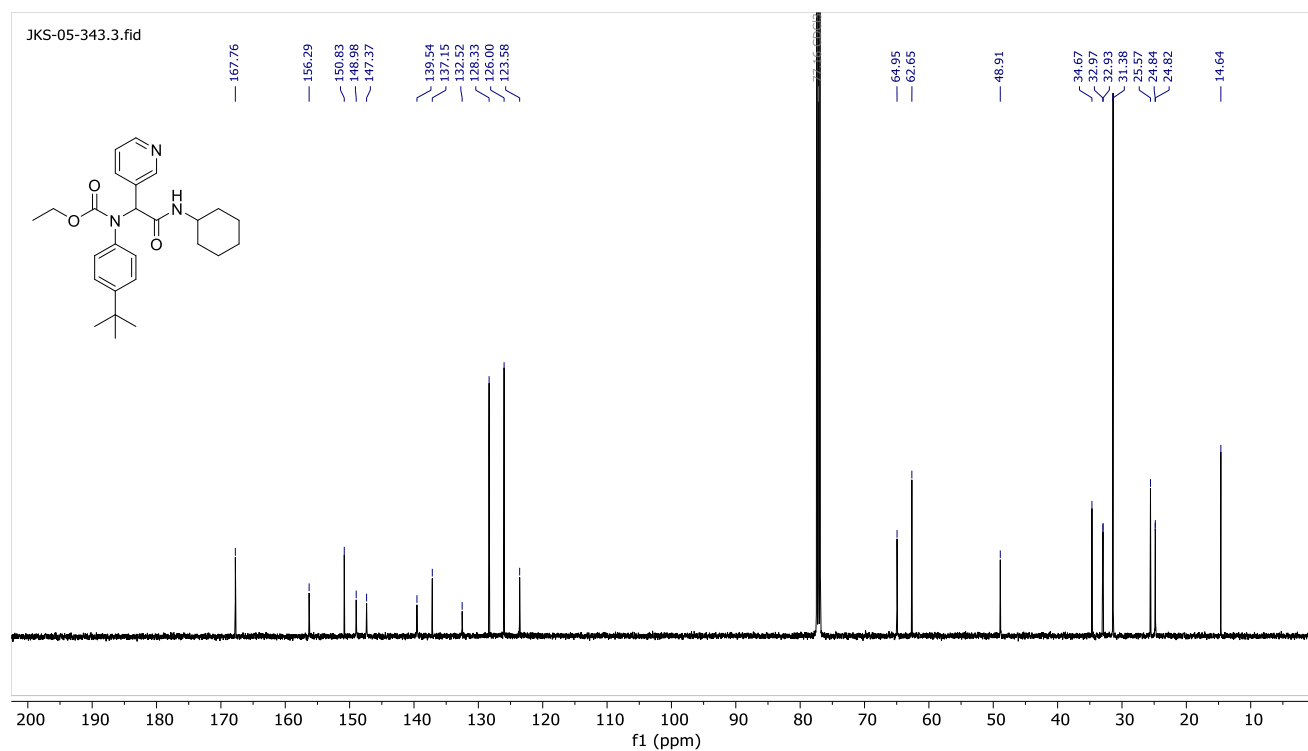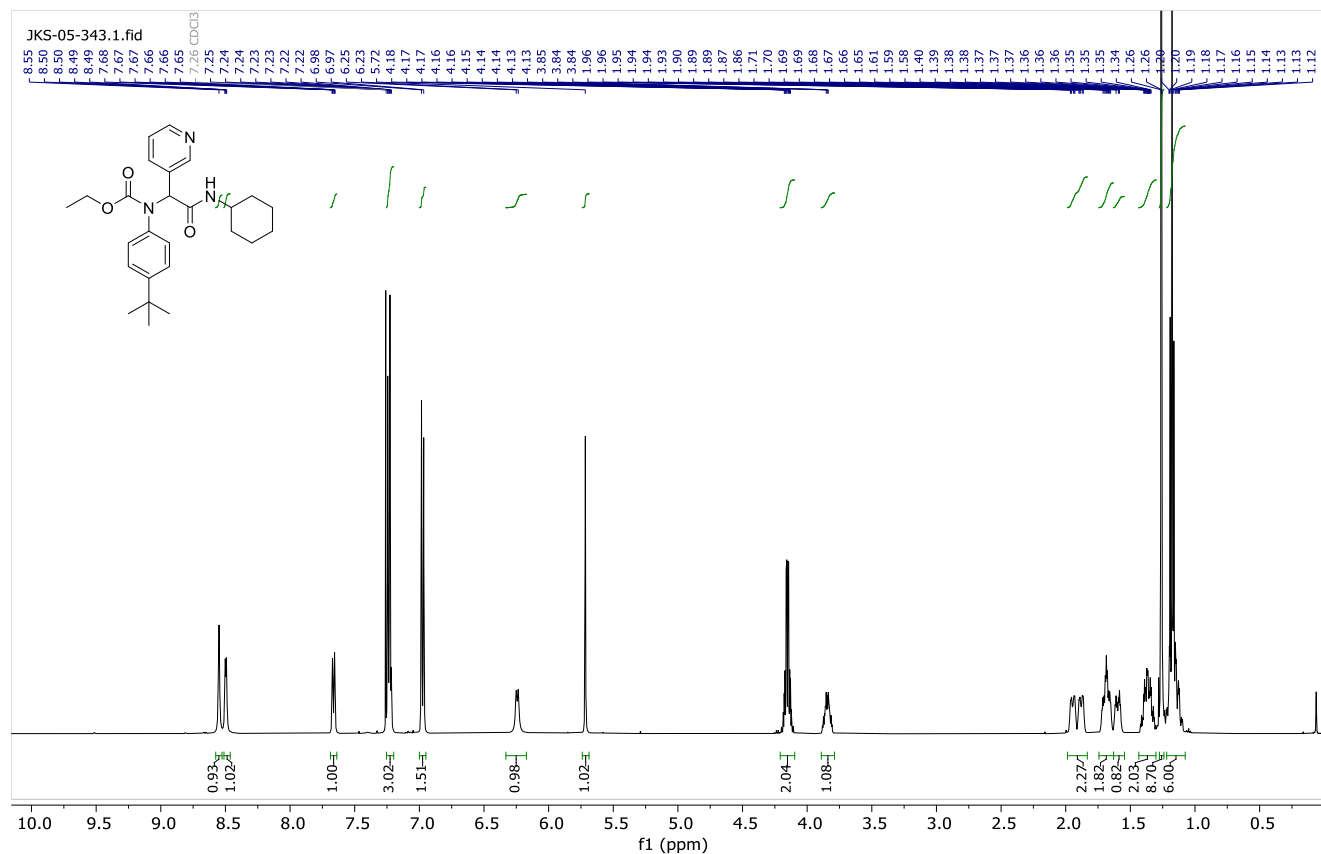
Compound 4.21 ^{1}H and ^{13}C NMR
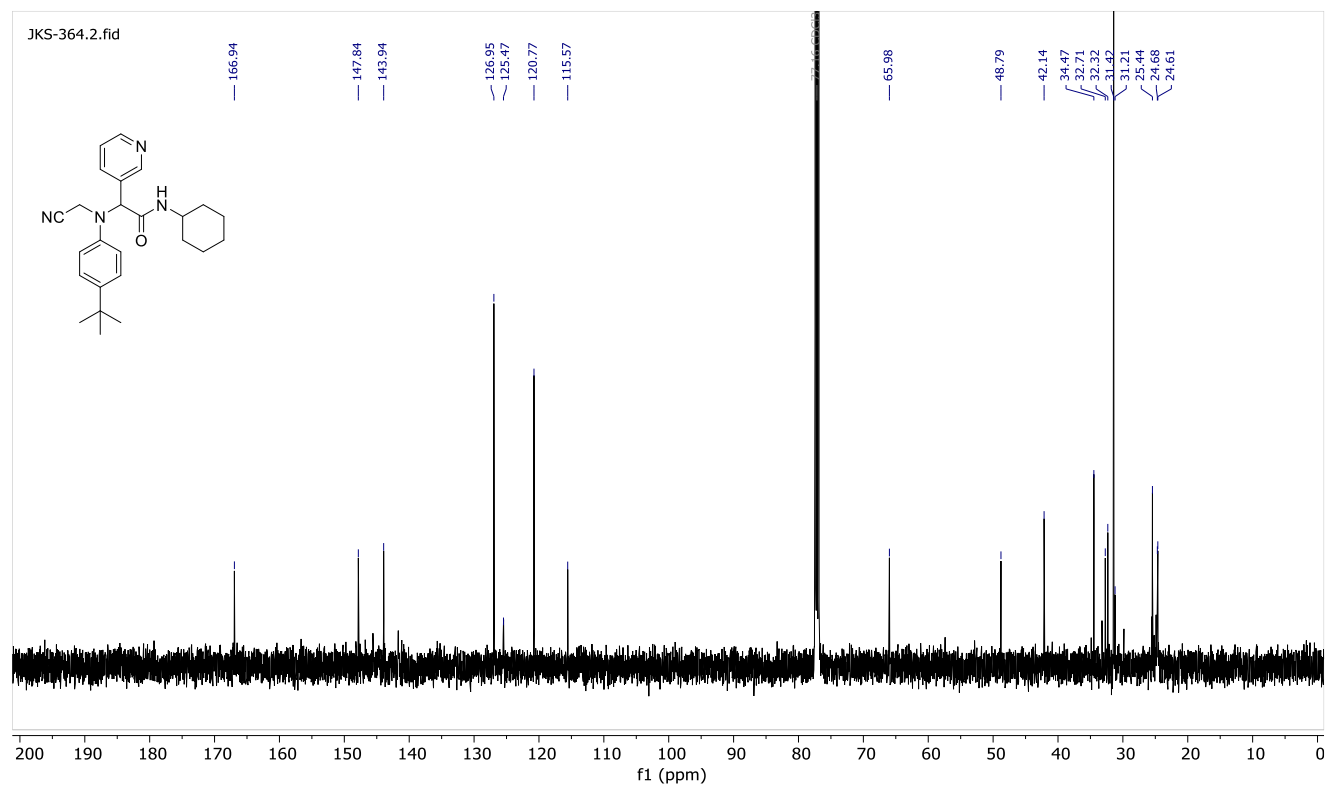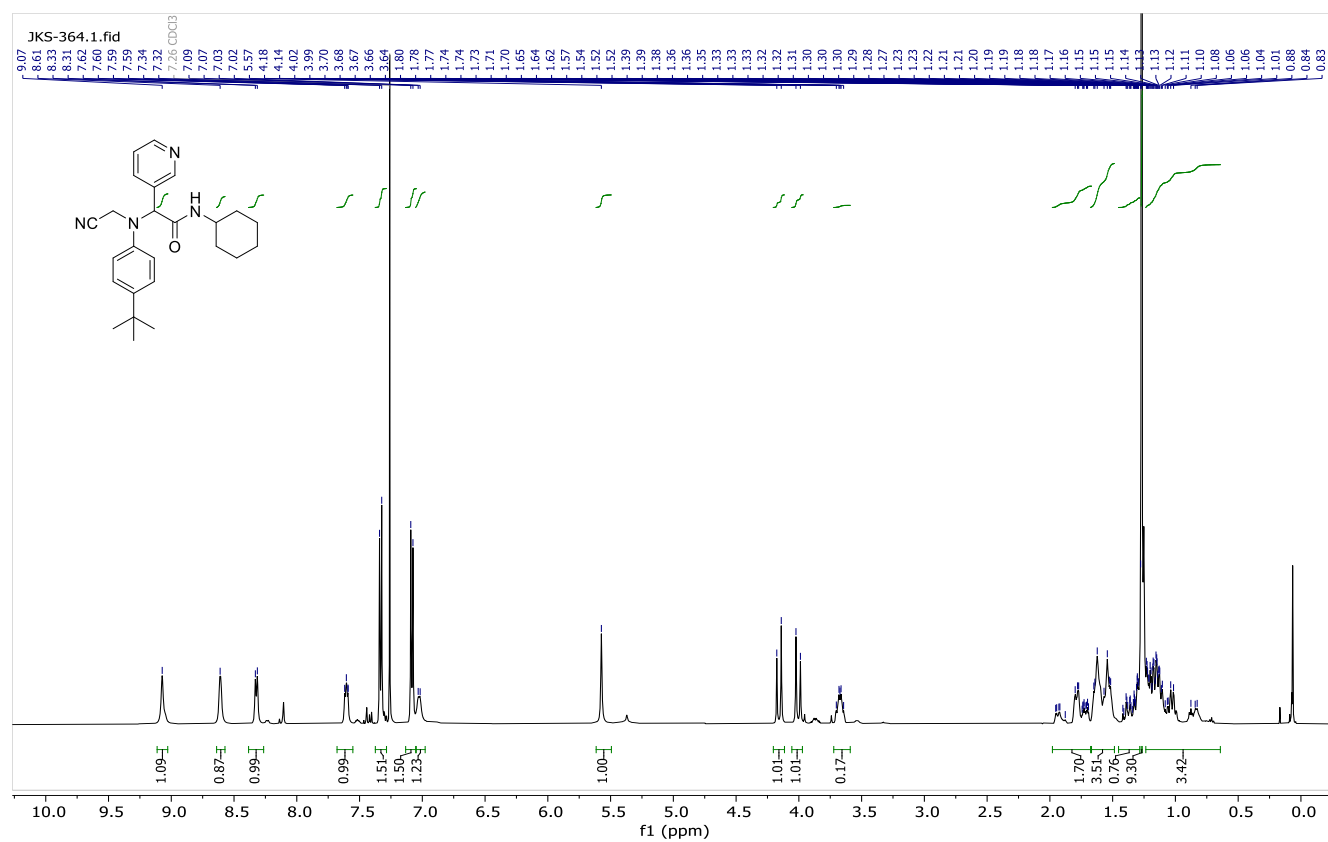
SUPPORTING INFORMATION FOR CHAPTER 5

Dose-Response Curves
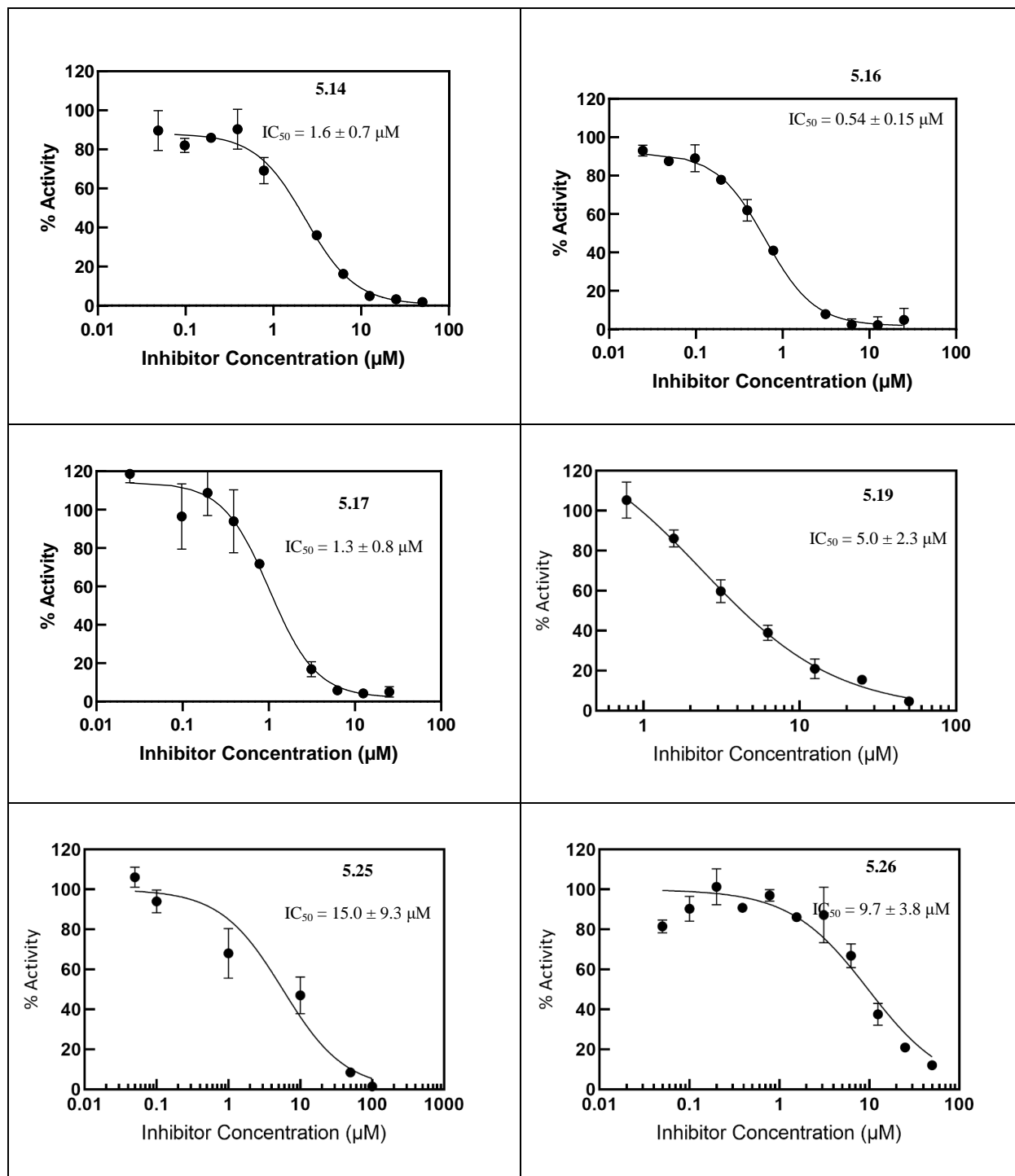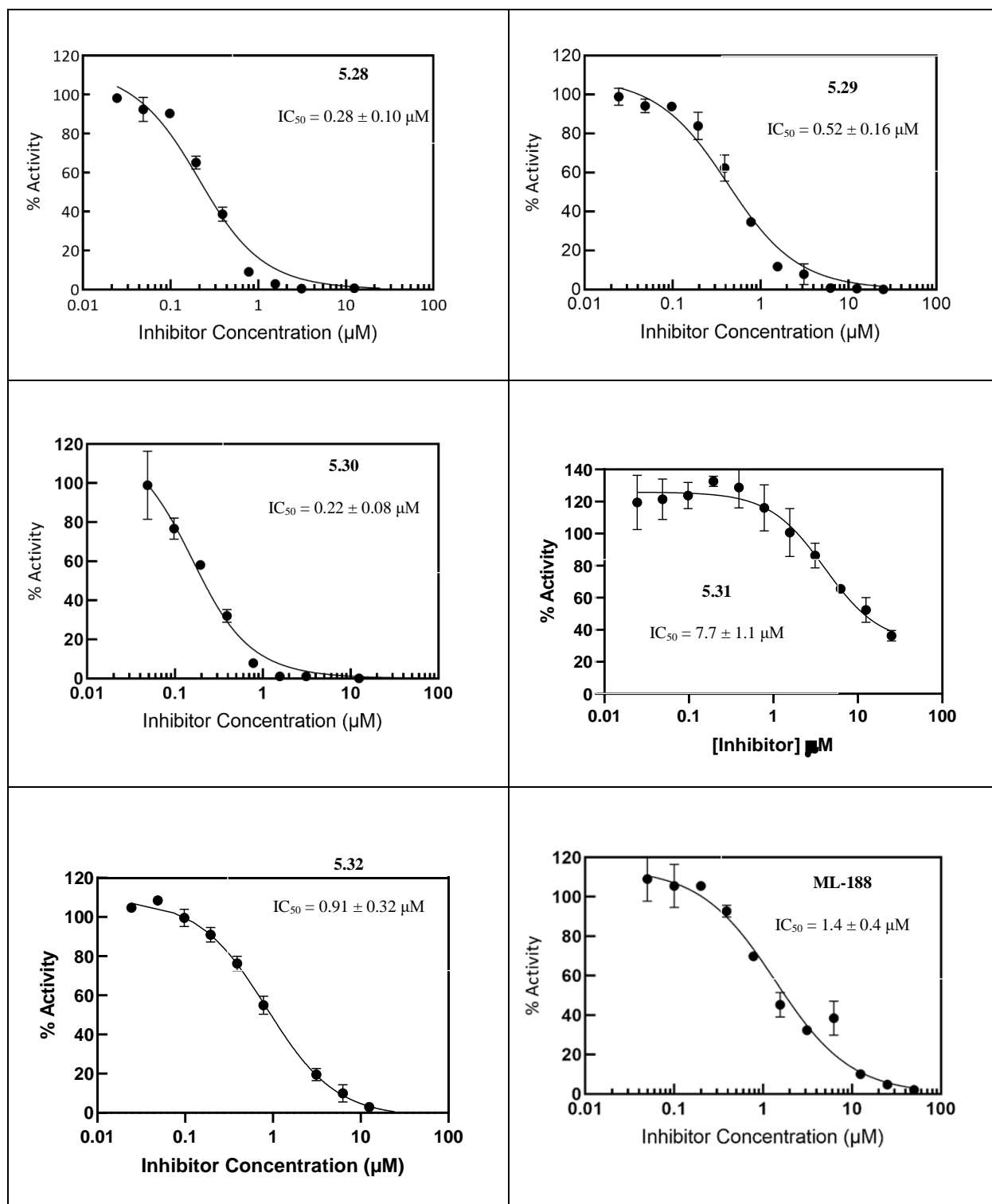
HPLC Compound Purity

Table S4.1. Purity of Biologically Tested Compounds

Entry	Compound	Retention time (min.)[a]	Purity (%)[b]
1	5.6	14.01	97.5
2	5.7	12.96	92.3
3	5.8	15.88	96.1
4	5.9	15.67	95.8
5	5.10	14.52	98.0
6	5.11	14.07	95.9
7	5.12	14.69	95.5
8	5.13	14.75	92.6
9	5.14	14.45	95.8
10	5.15	14.21	99.5
11	5.16	14.3	96.8
12	5.17	13.5	88.1
16	5.19	16.80	99.5
17	5.20	16.30	95.1
18	5.21	15.56	95.7
19	5.22	17.41	91.8
20	5.23	14.61	99.5
21	5.24	14.84	97.6
22	5.25	14.69	97.2
23	5.26	15.1	96.0

24	5.27	14.53	97.5
25	5.28	15.39	99.4
28	5.29	15.84	94.5
29	5.30	11.6	87.8
36	5.31[c]	9.07	97.9
37	5.32[c]	17.9	97.5
38	ML-188	15.09	98.9

[a] Conditions: (gradient of 95% water, 5% MeOH or MeCN, 1 mL/min). [b] UV detection at 254 nm. [c] Conditions: (isocratic elution, 1:1 water:MeCN, 1 mL/min)

NMR Spectra

Compound 5.6 ^{1}H and NMR

JT-69.1.fid

JT-69.2.fid
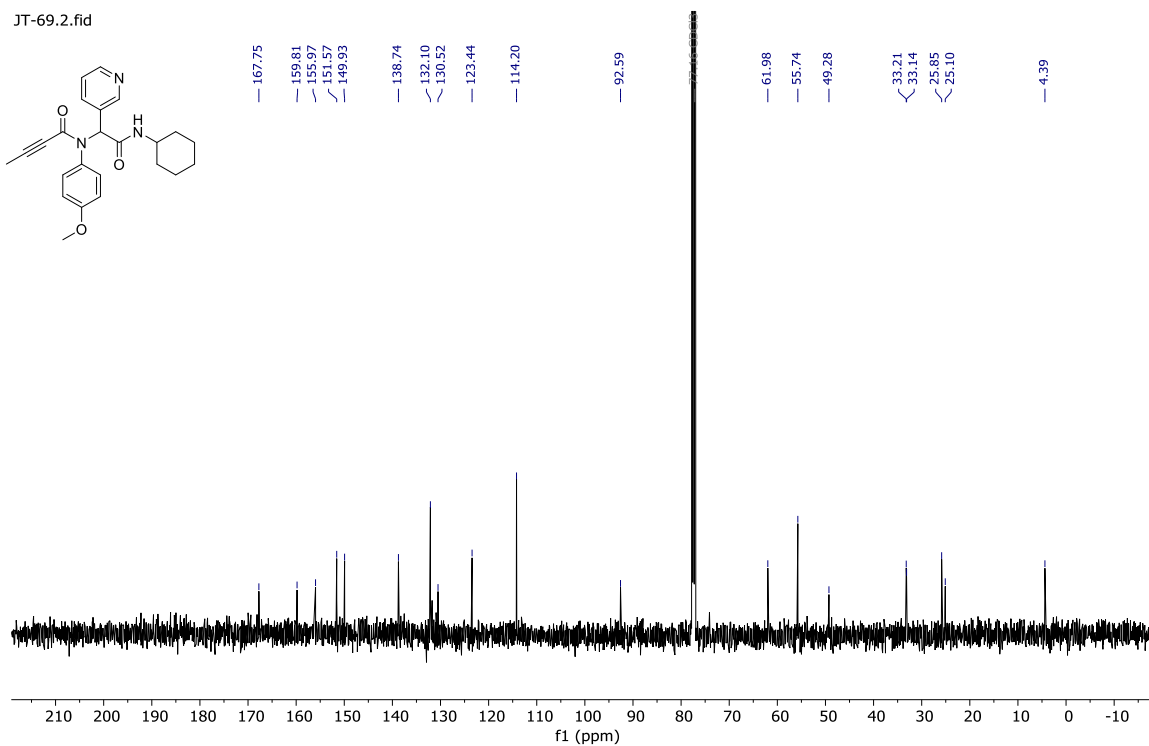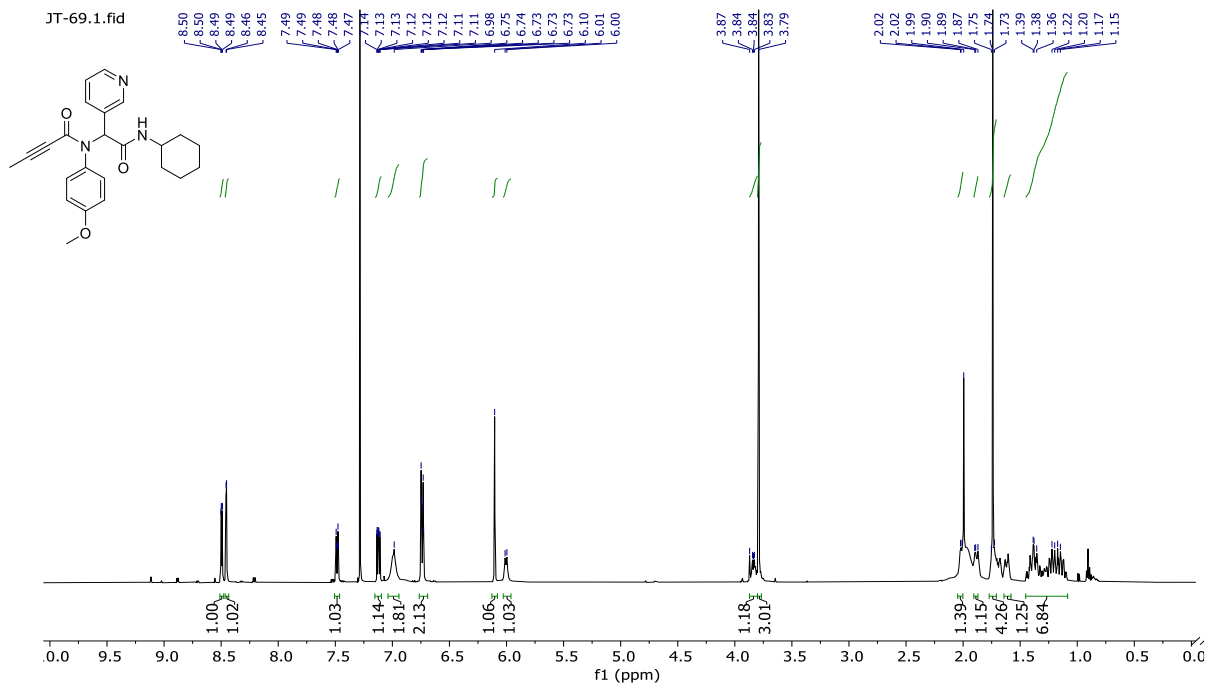
Compound 5.7 ¹H and ¹³C NMR

JT-71.3.fid
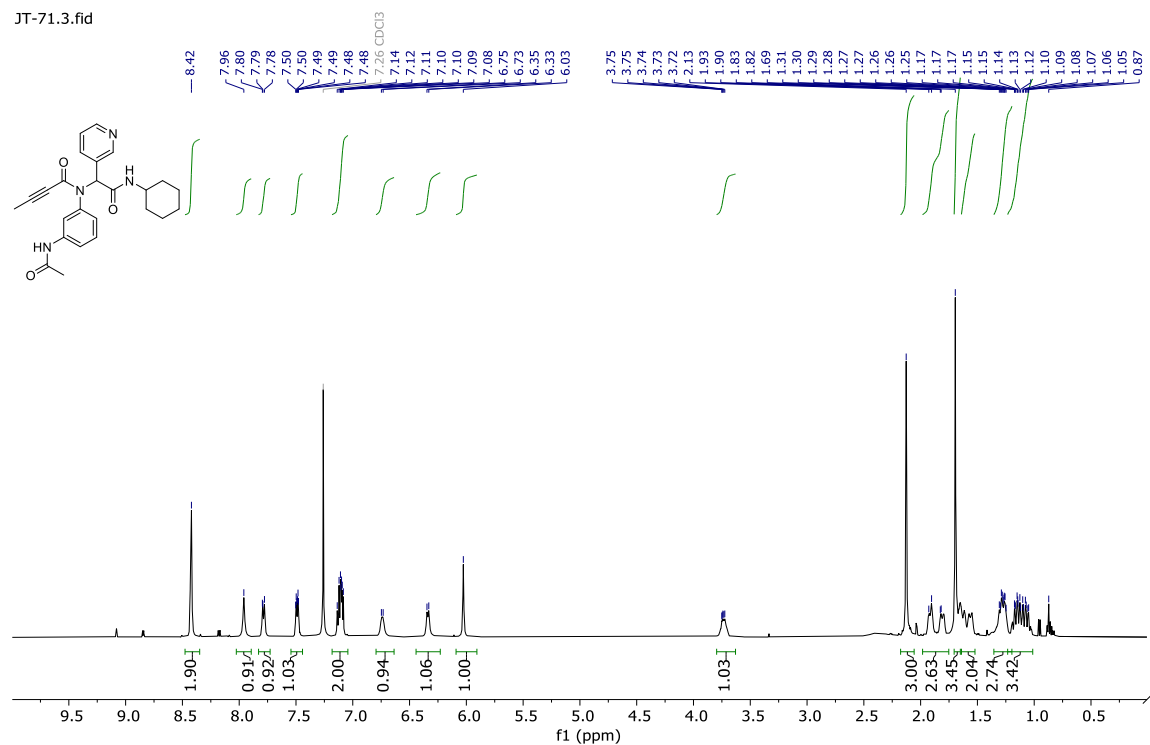
JT-71.22.fid
Moitessier
1d_C13 CDCl3 E:\\ tjutrins 11
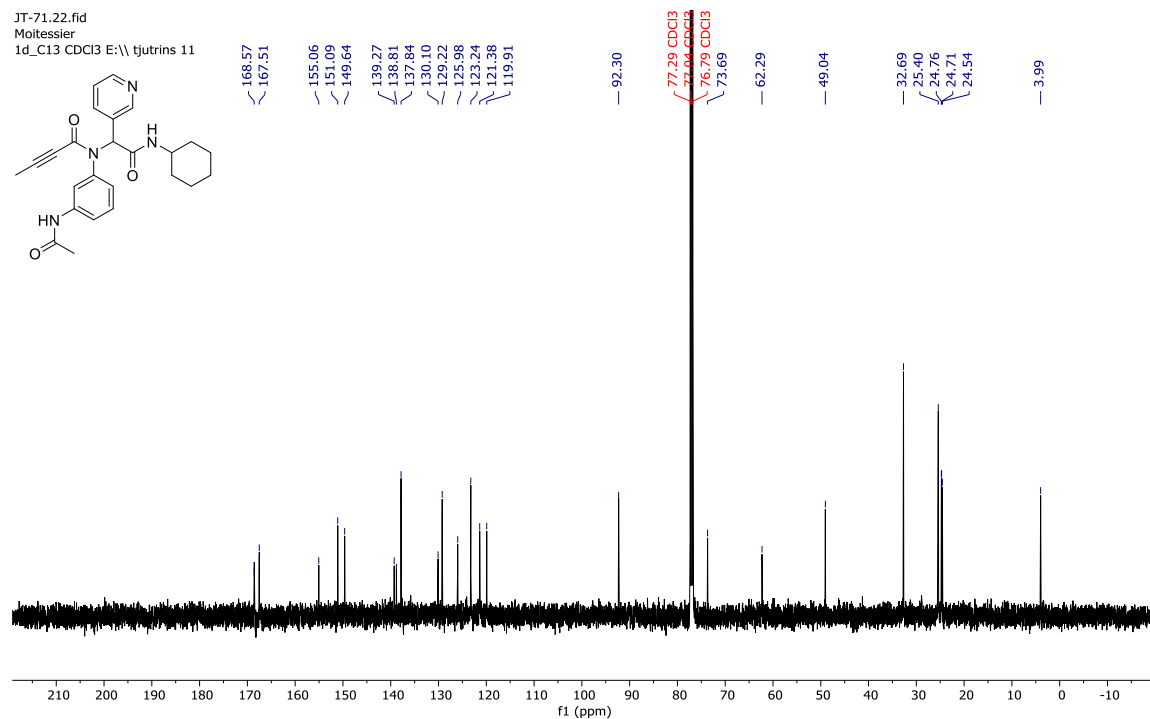
Compound 5.8 ^{1}H and ^{13}C NMR
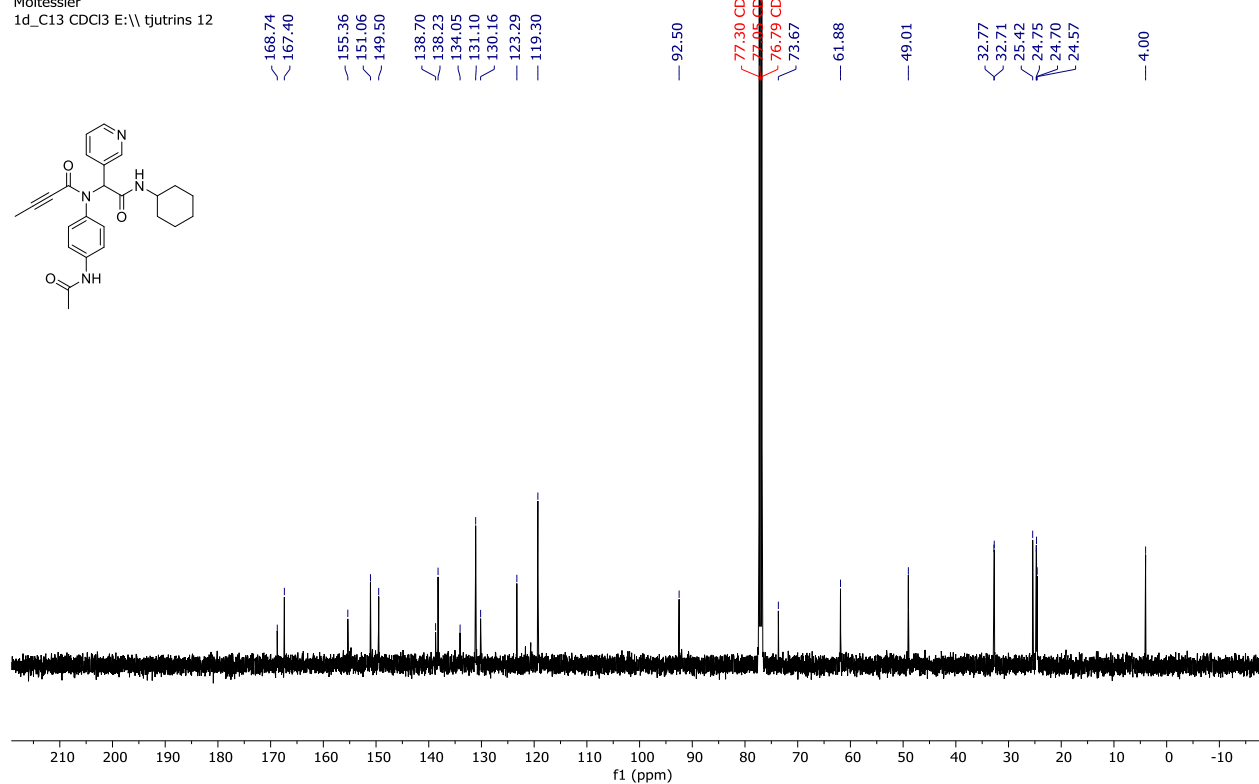
Compound 5.9 ^{1}H and ^{13}C NMR
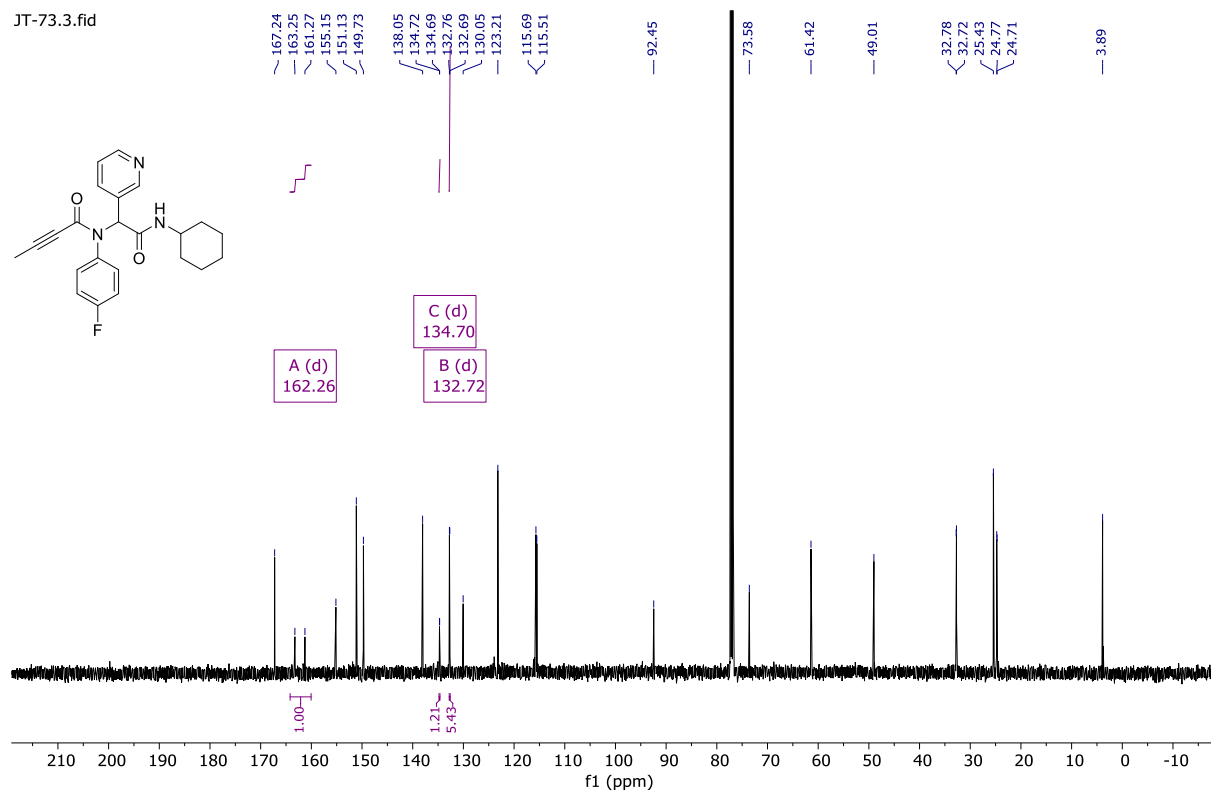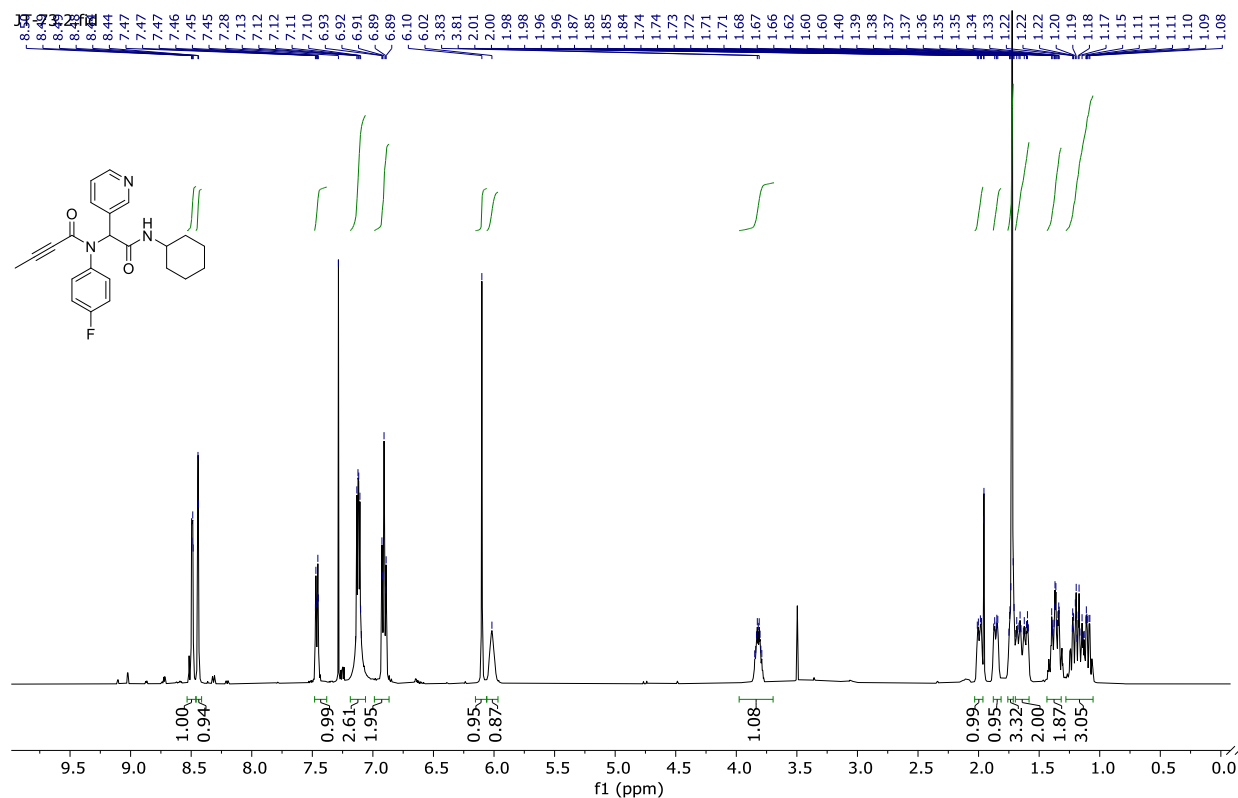
Compound 5.10 ^{1}H and ^{13}C NMR
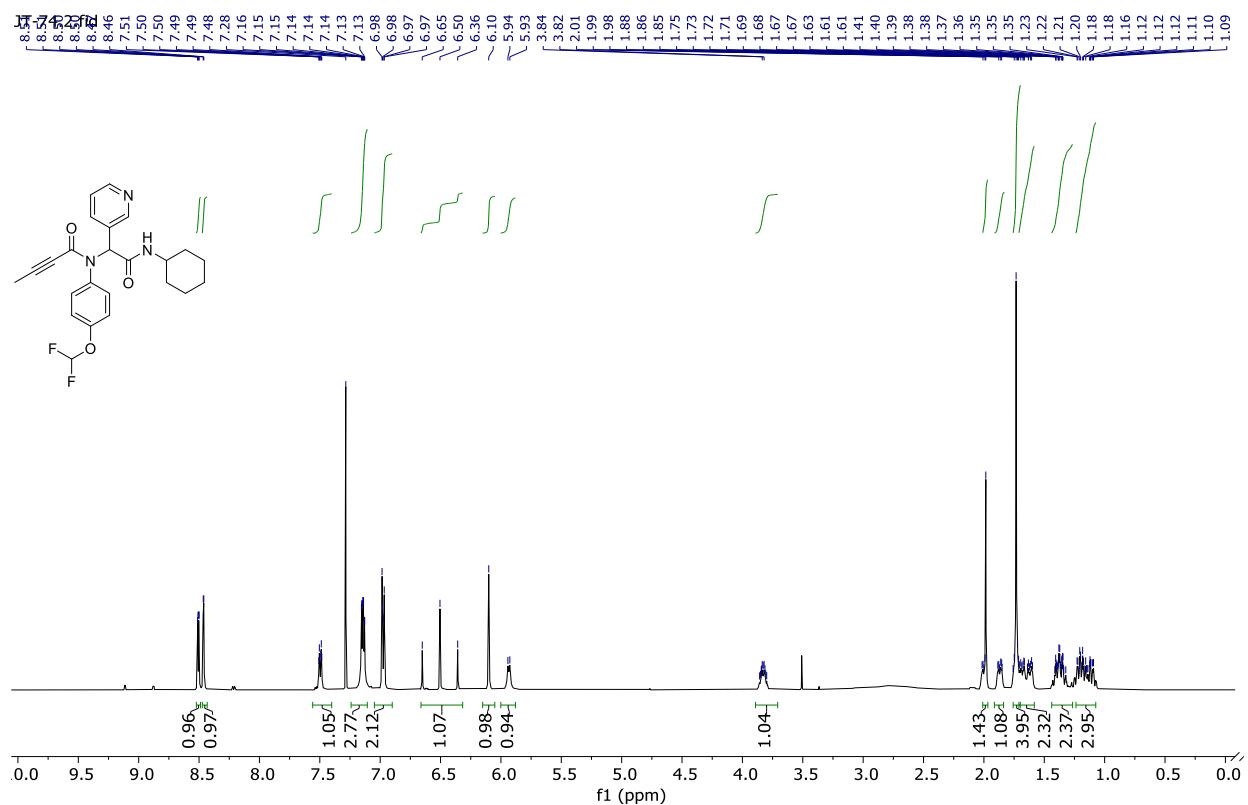
Compound 5.11 ^{1}H and ^{13}C NMR
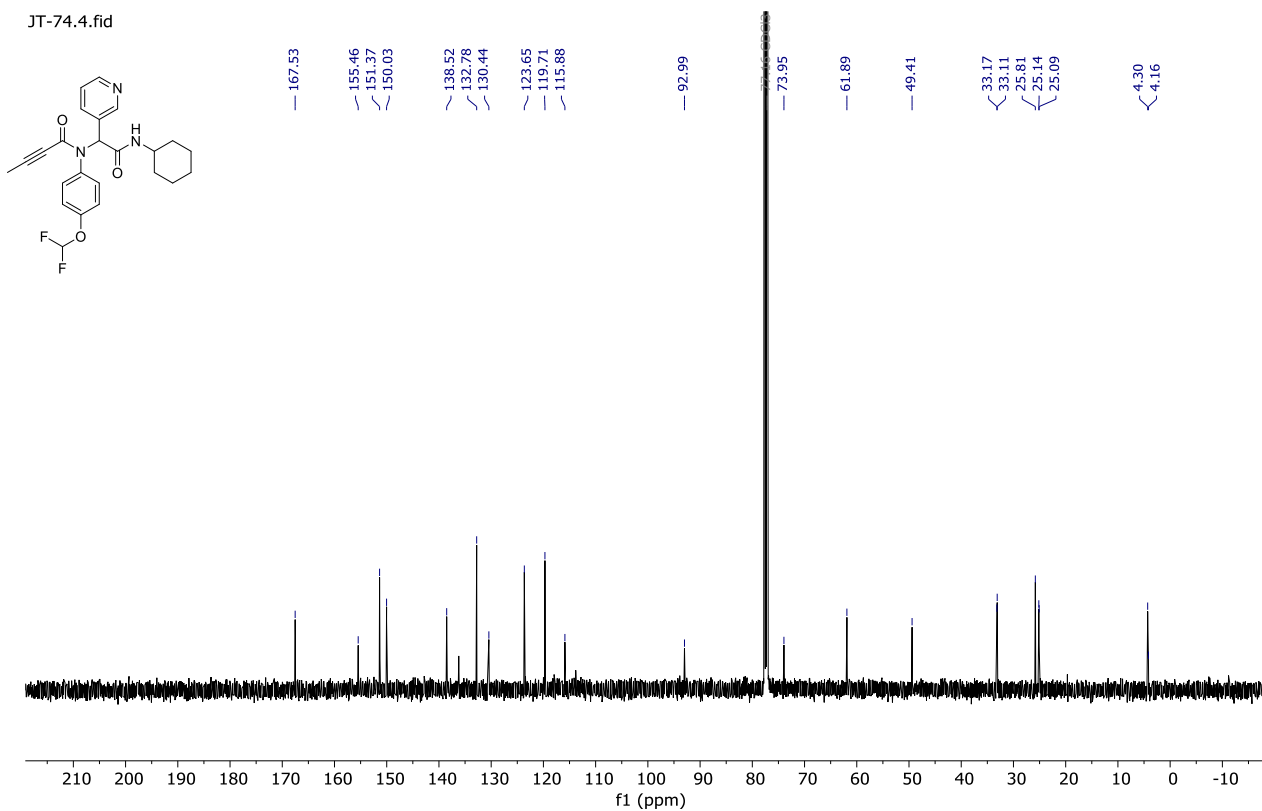
Compound 5.12 ^{1}H and ^{13}C NMR
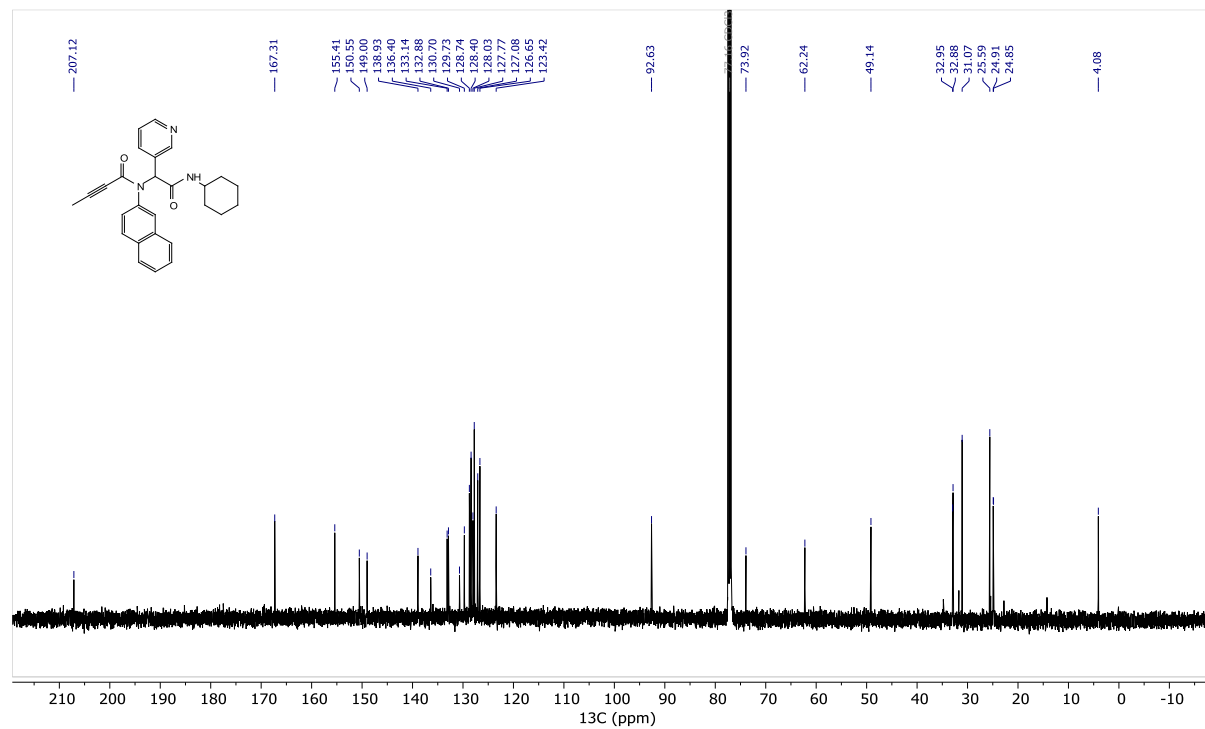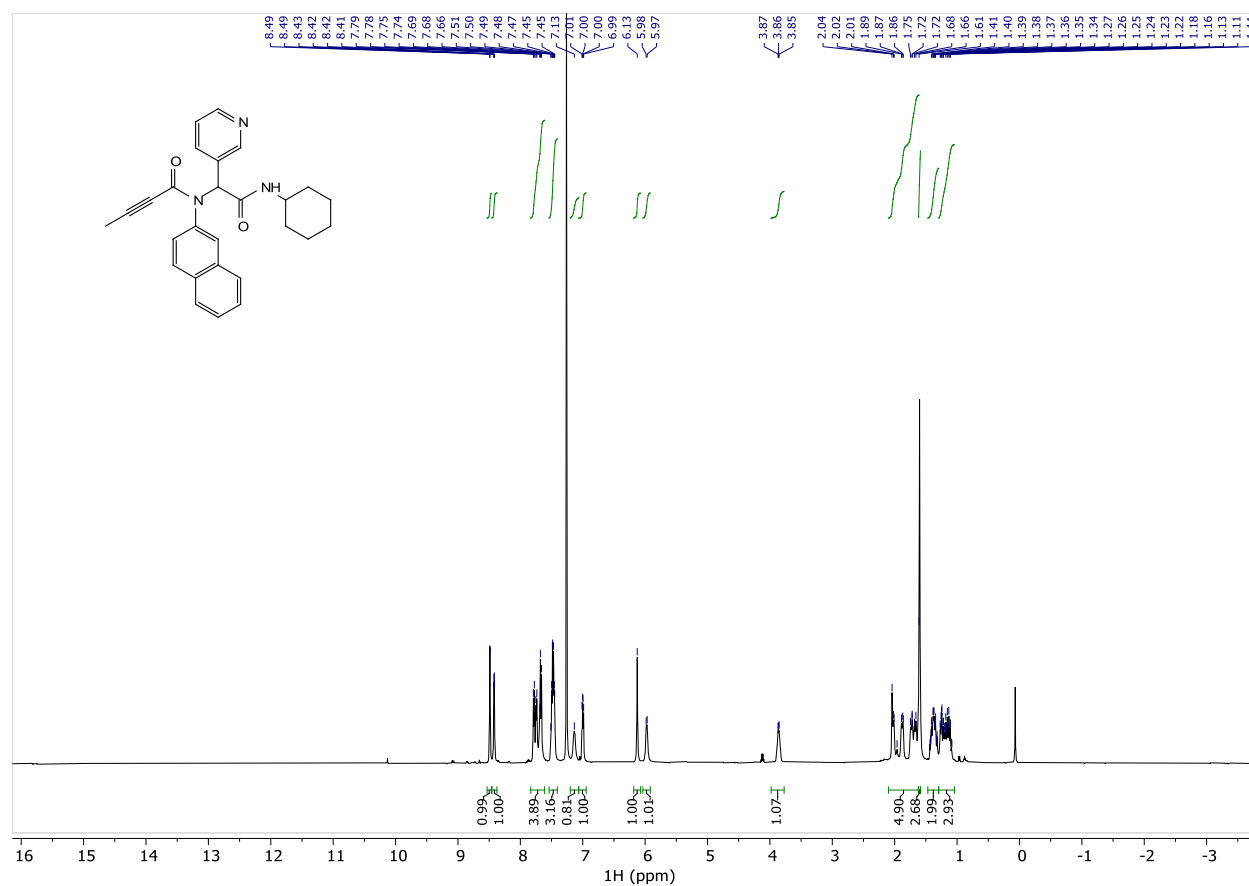
Compound 5.13 ^{1}H and ^{13}C NMR

NB-37 ^{1}H NMR

NB-37 ^{13}C NMR
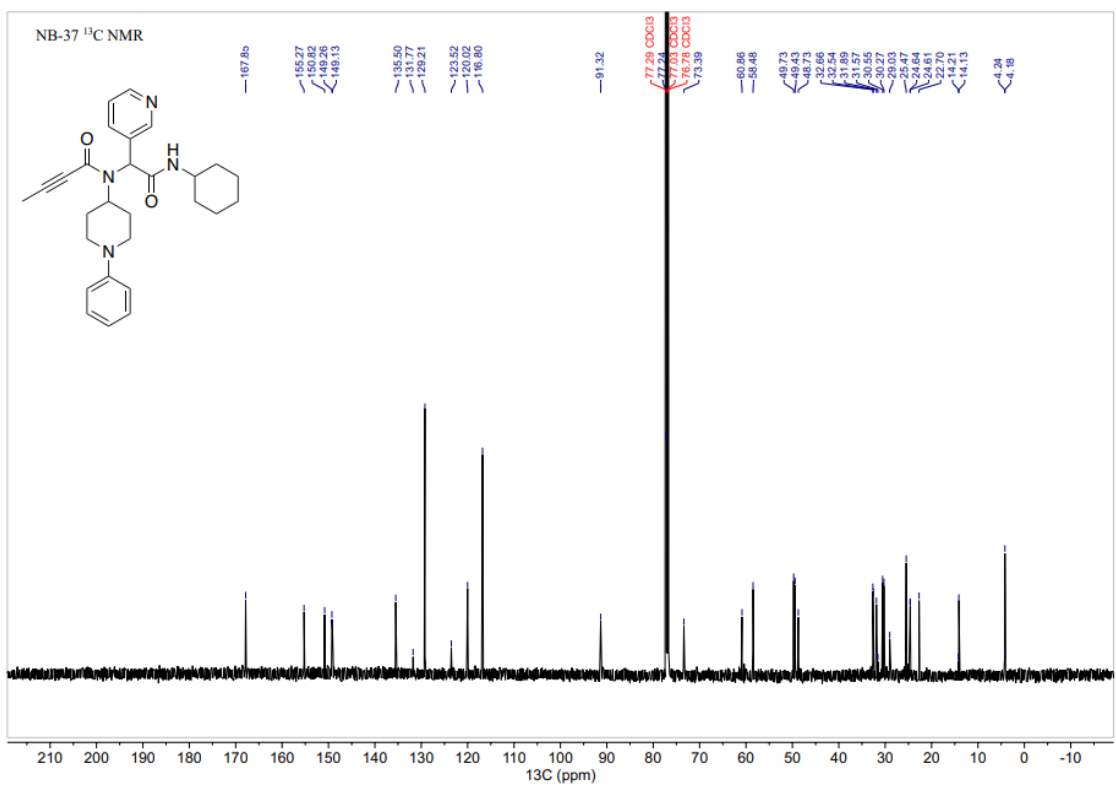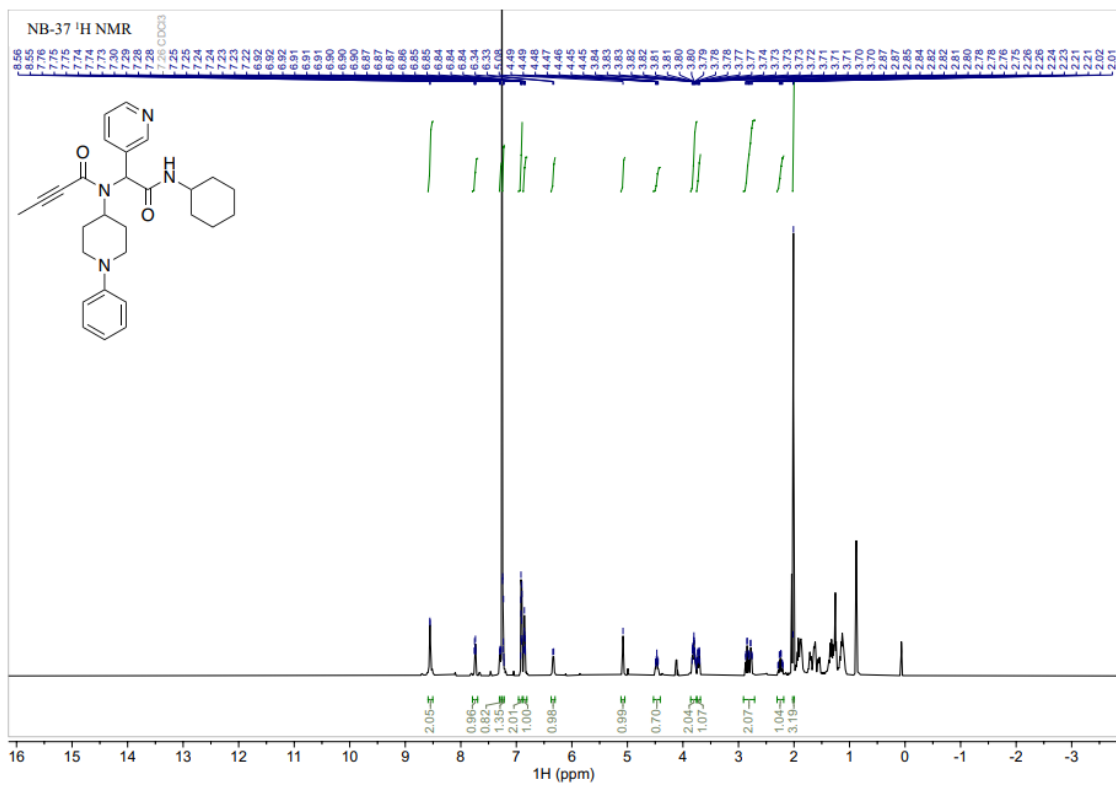
Compound 5.14 ^{1}H and ^{13}C NMR

Compound 5.15 ^{1}H and ^{13}C NMR
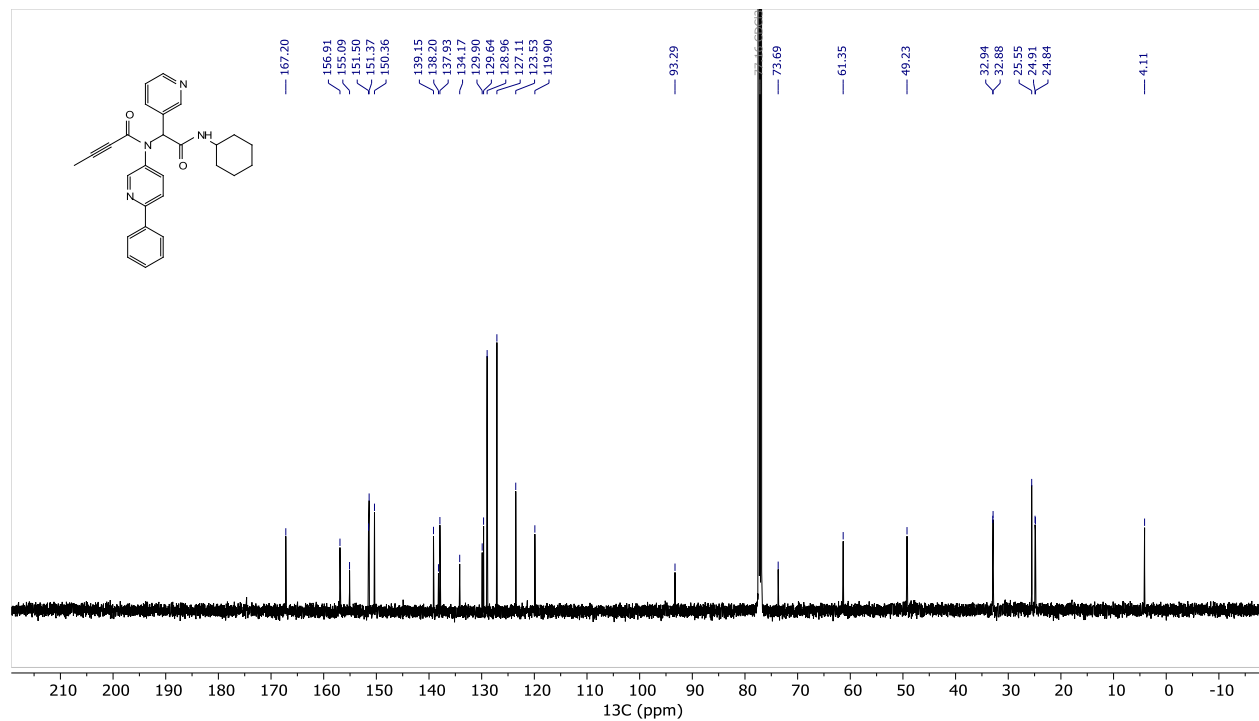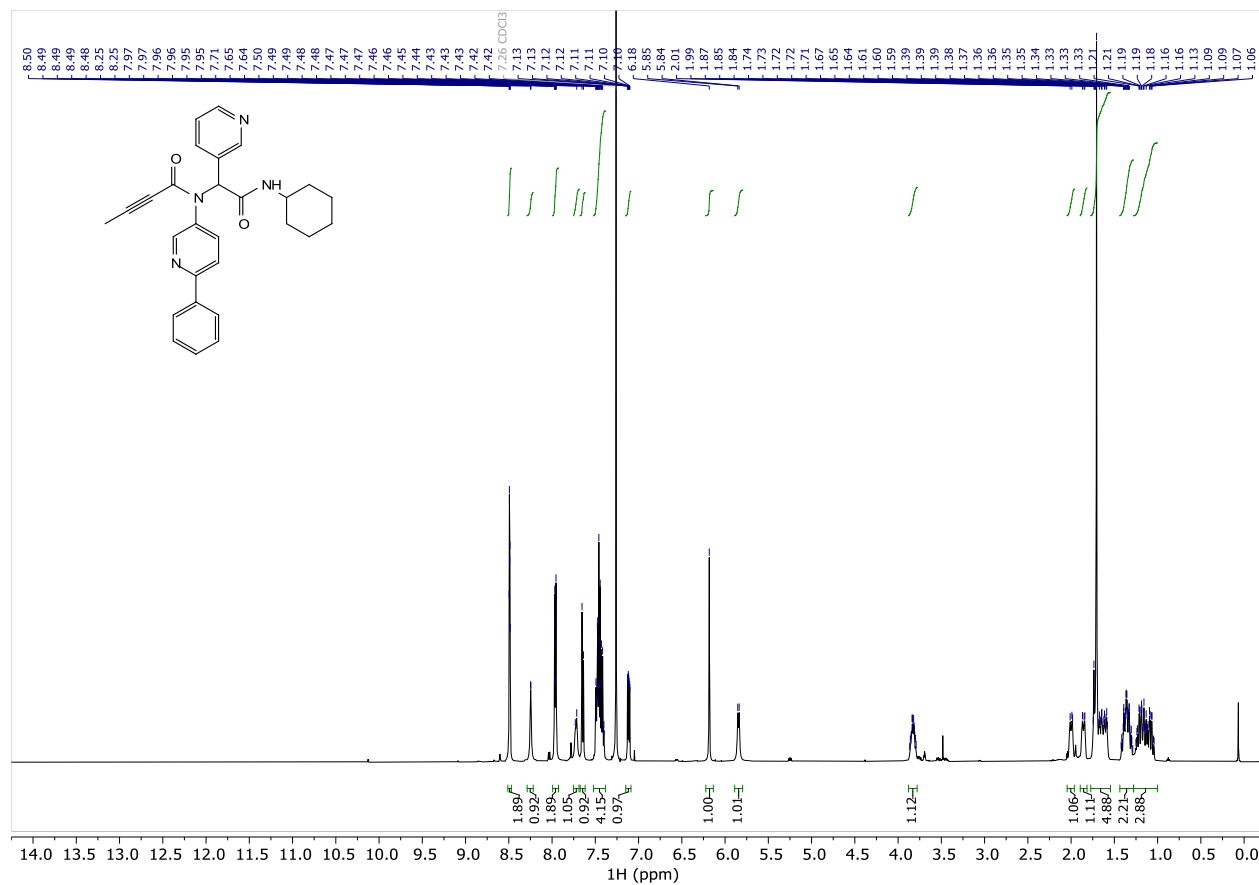
Compound 5.16 ^{1}H and ^{13}C NMR

Compound 5.17 ^{1}H and ^{13}C NMR
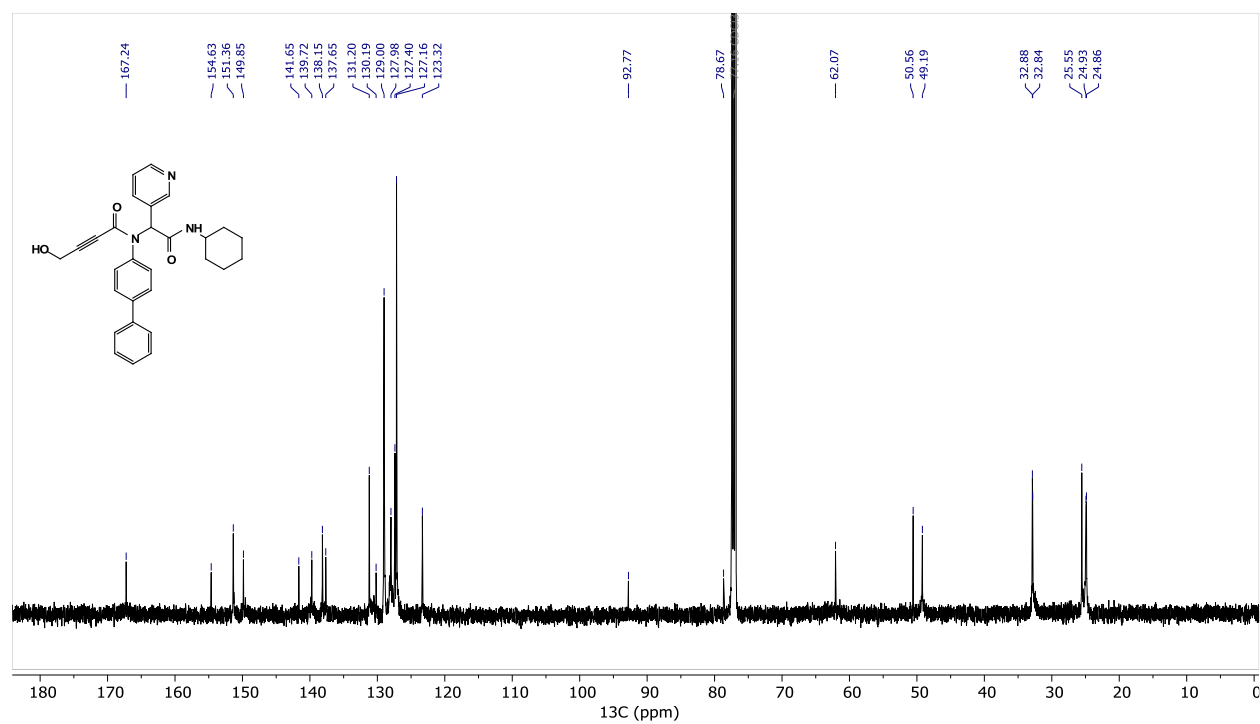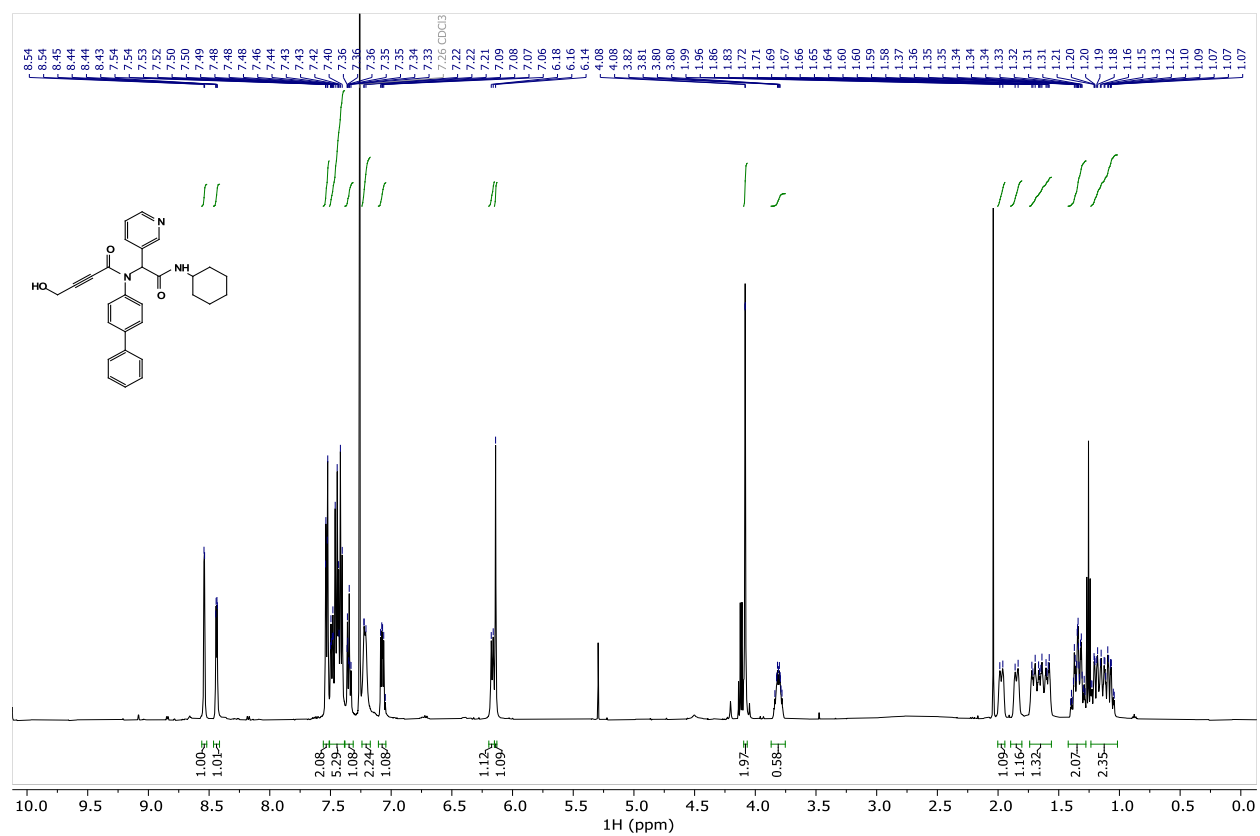
Compound 5.13-int-1 ^{1}H and ^{13}C NMR

20210811_NB-35-1_PROTON_01

F (m)	E (t)
7.24	6.84

D (m)
6.93

K (s)
4.47

G (d)
3.60

H (m)
2.84

A (s)
2.04

J (s)
1.46

I (qd)
1.54

1H (ppm)

NB-35.1.fid

13C (ppm)
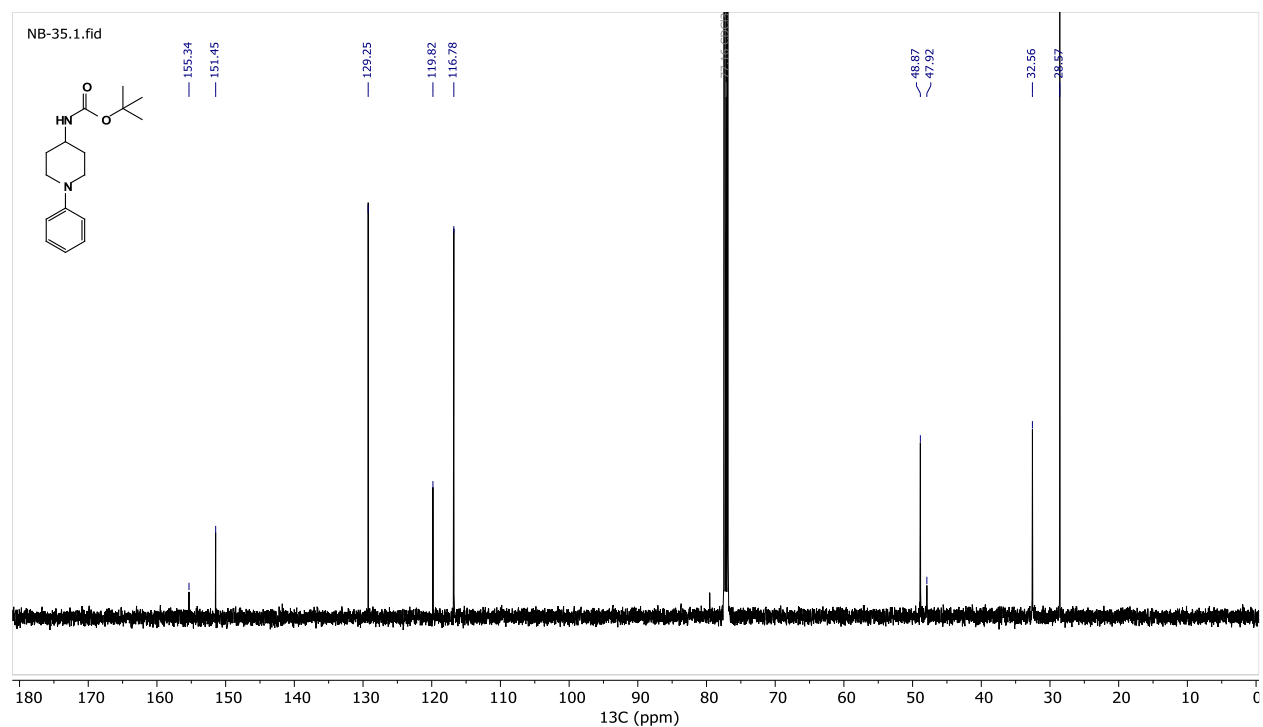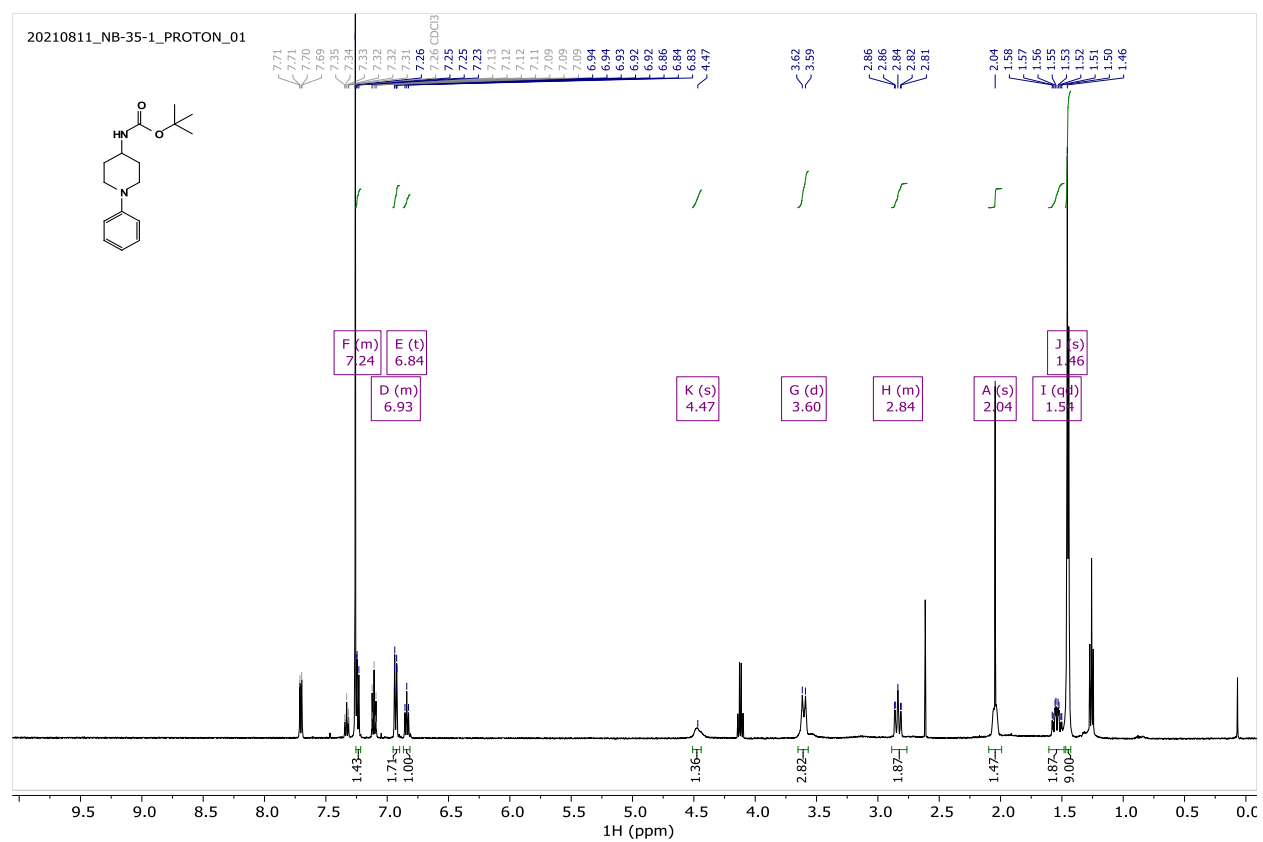
Compound 5.19 ^{1}H and ^{13}C NMR

JT-75.2.fid

JT-75.3.fid
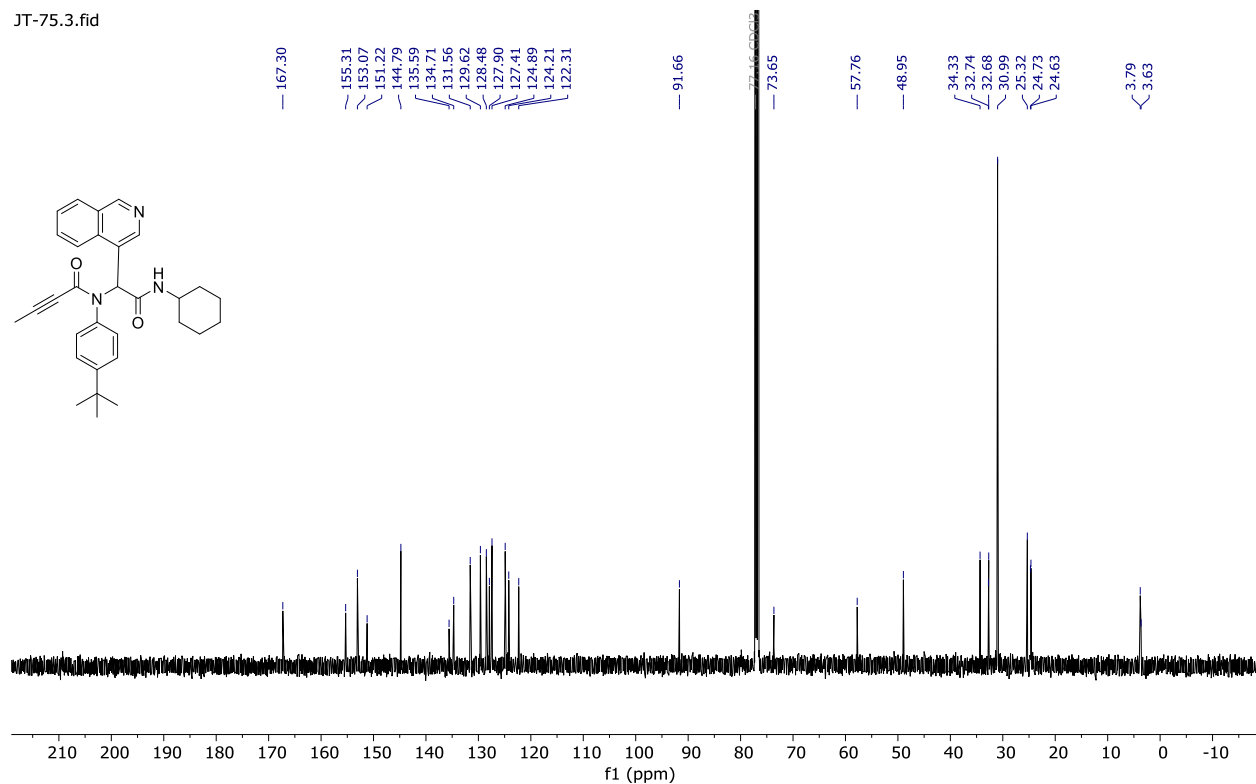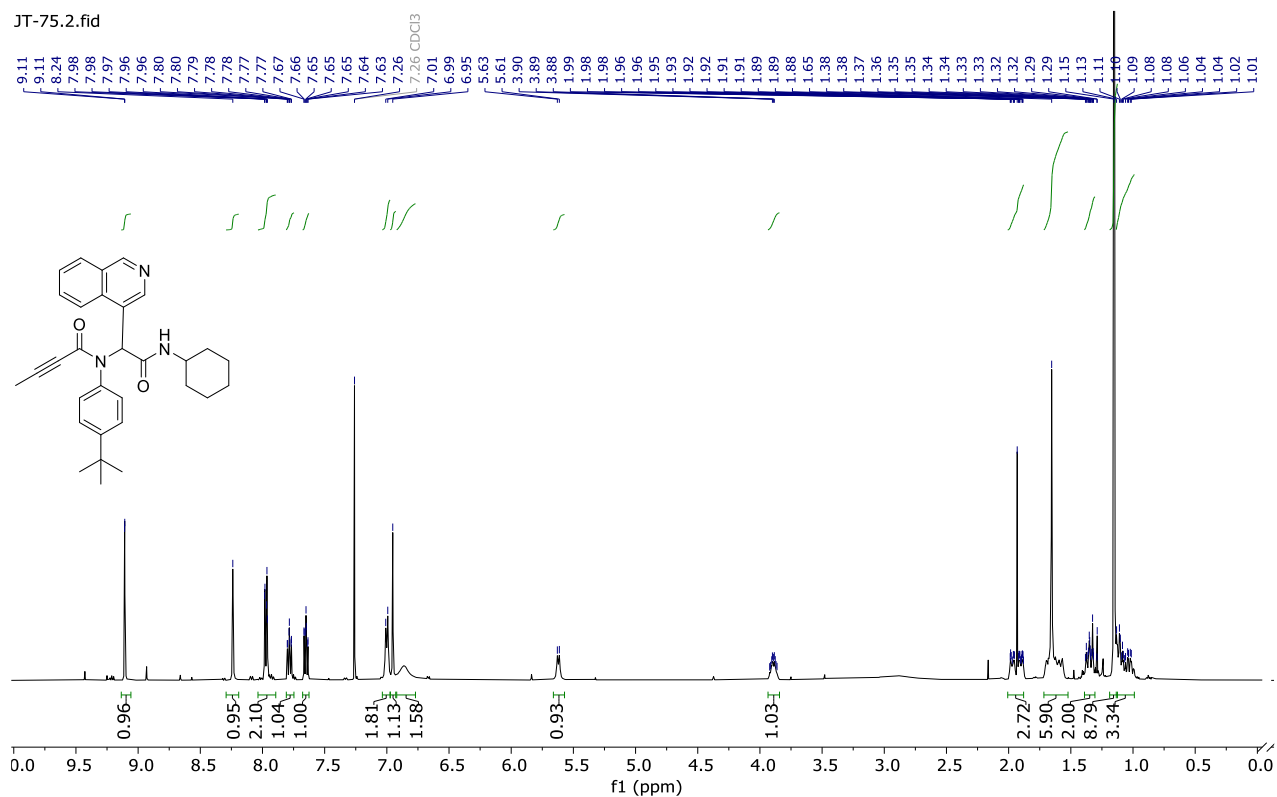
Compound 5.20 ¹H and ¹³C NMR

JT-78.2.fid

JT-78.3.fid
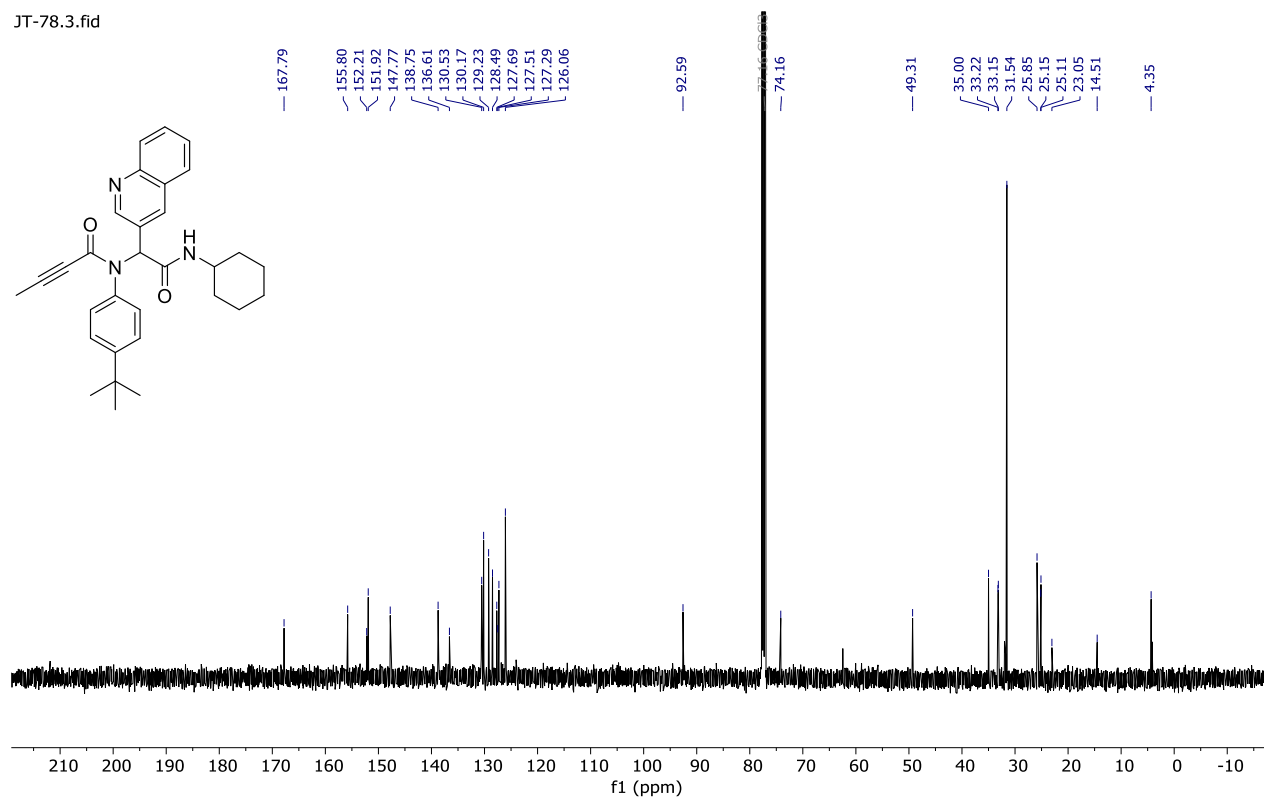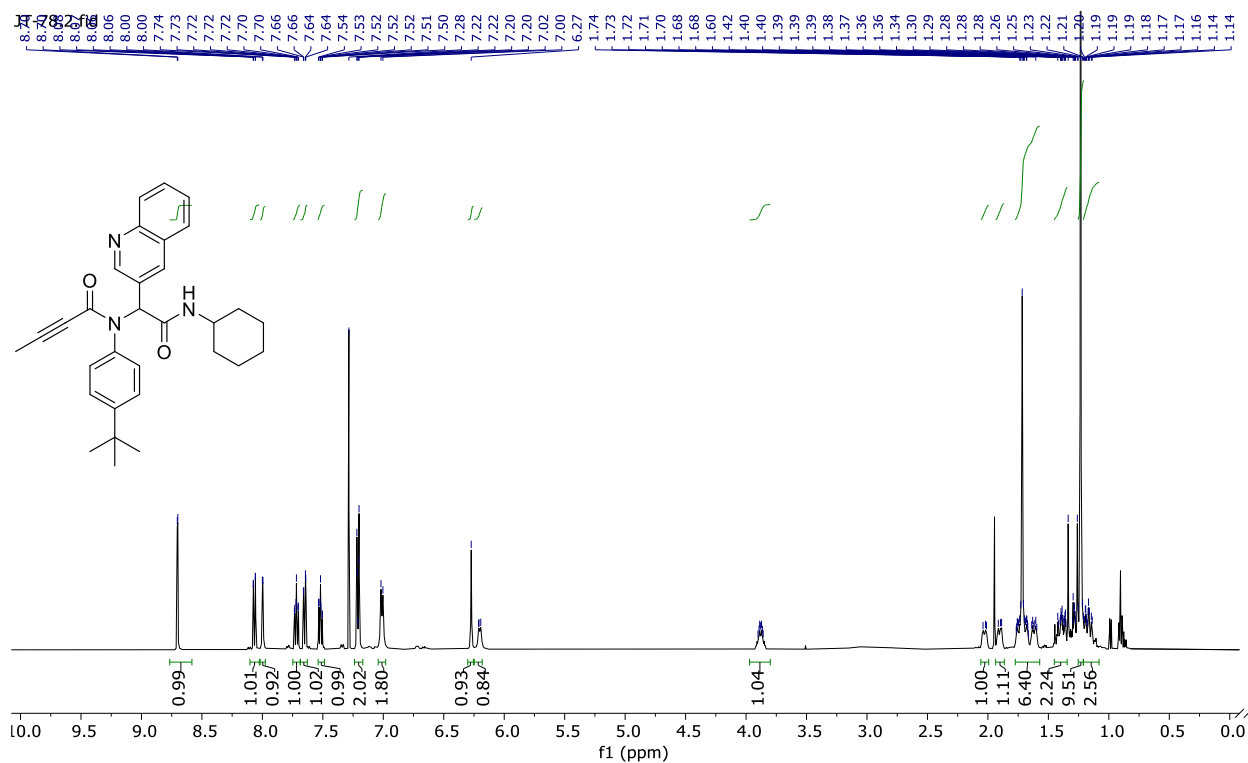
Compound 5.21 ^{1}H and ^{13}C NMR
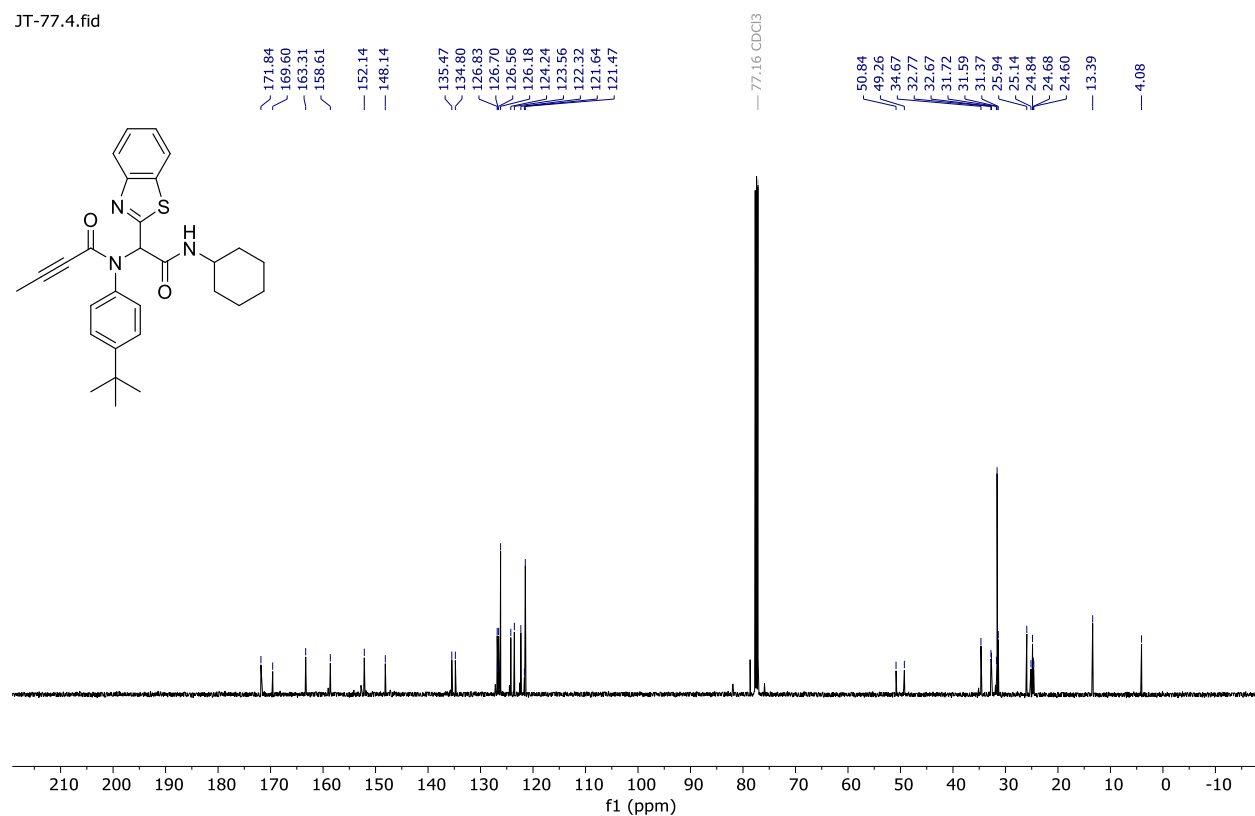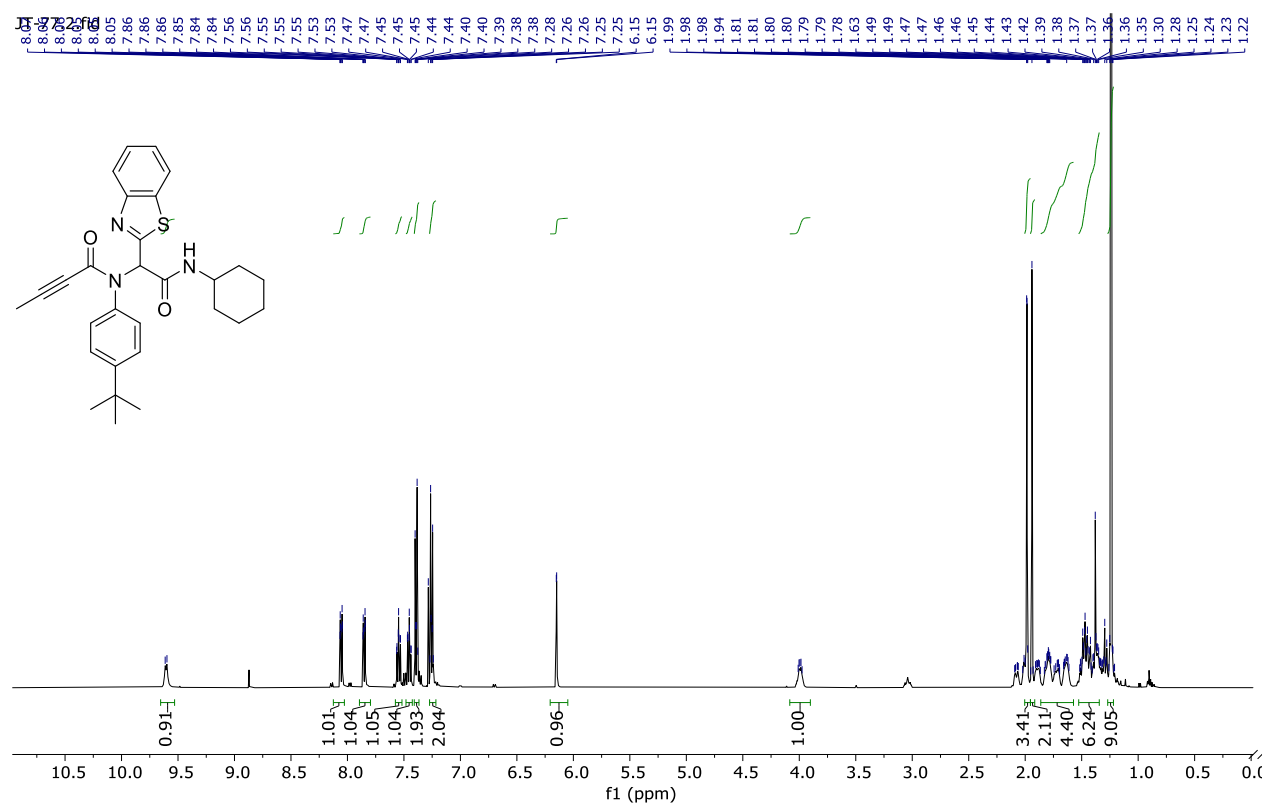
Compound 5.22 ^{1}H and ^{13}C NMR
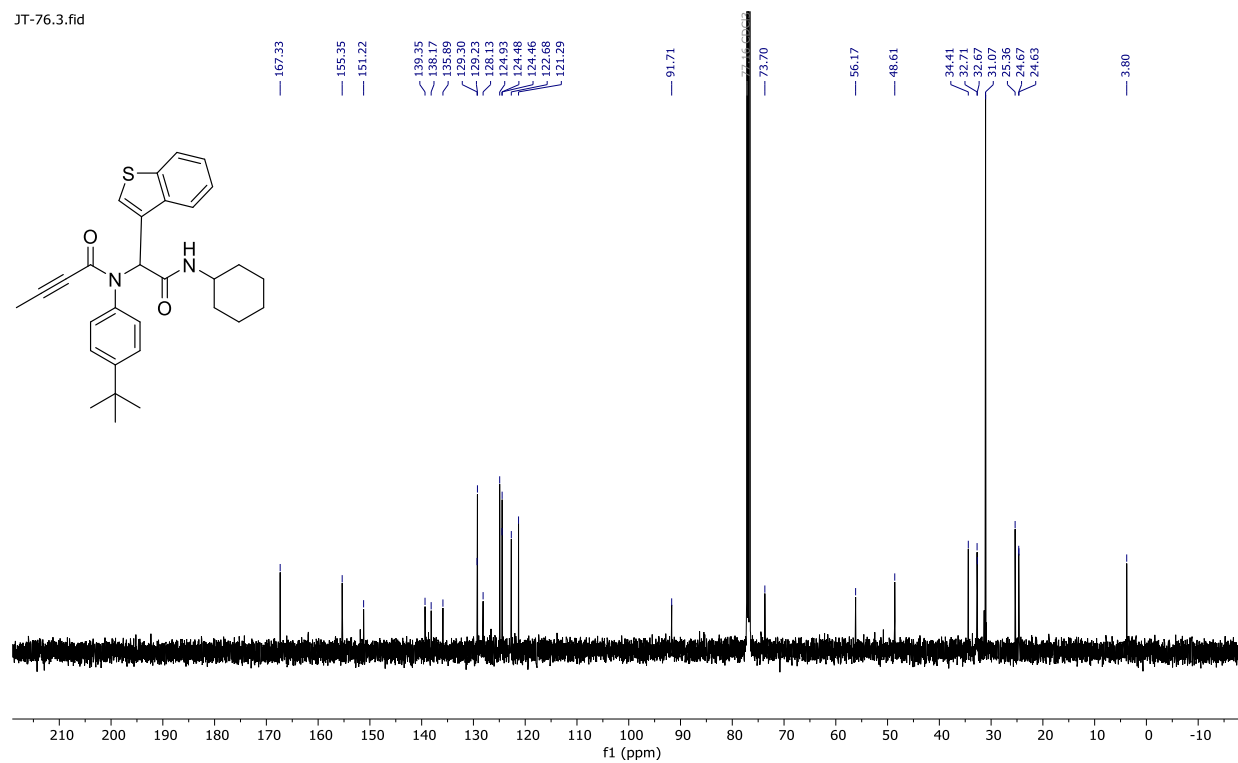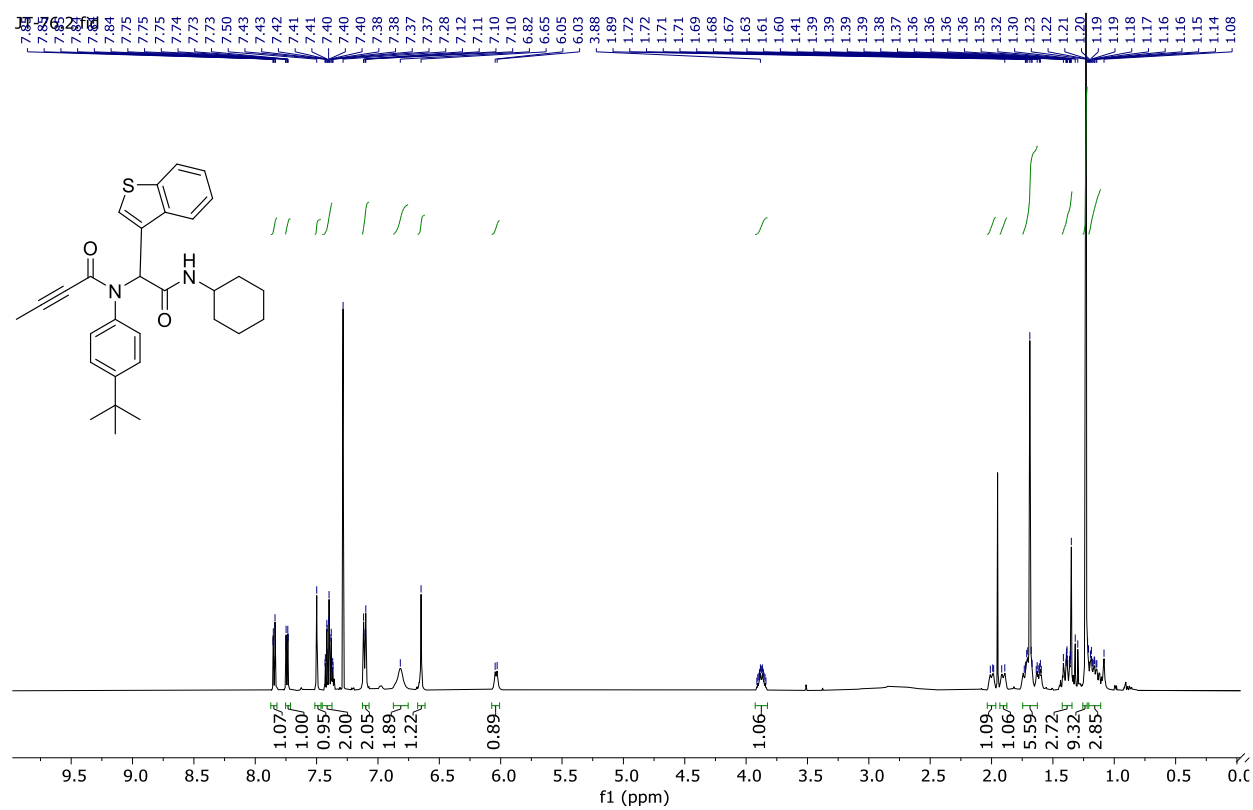
Compound 5.23 ^{1}H and ^{13}C NMR

JKS-05-385_PROTON_01

JKS-05-385.1.fid
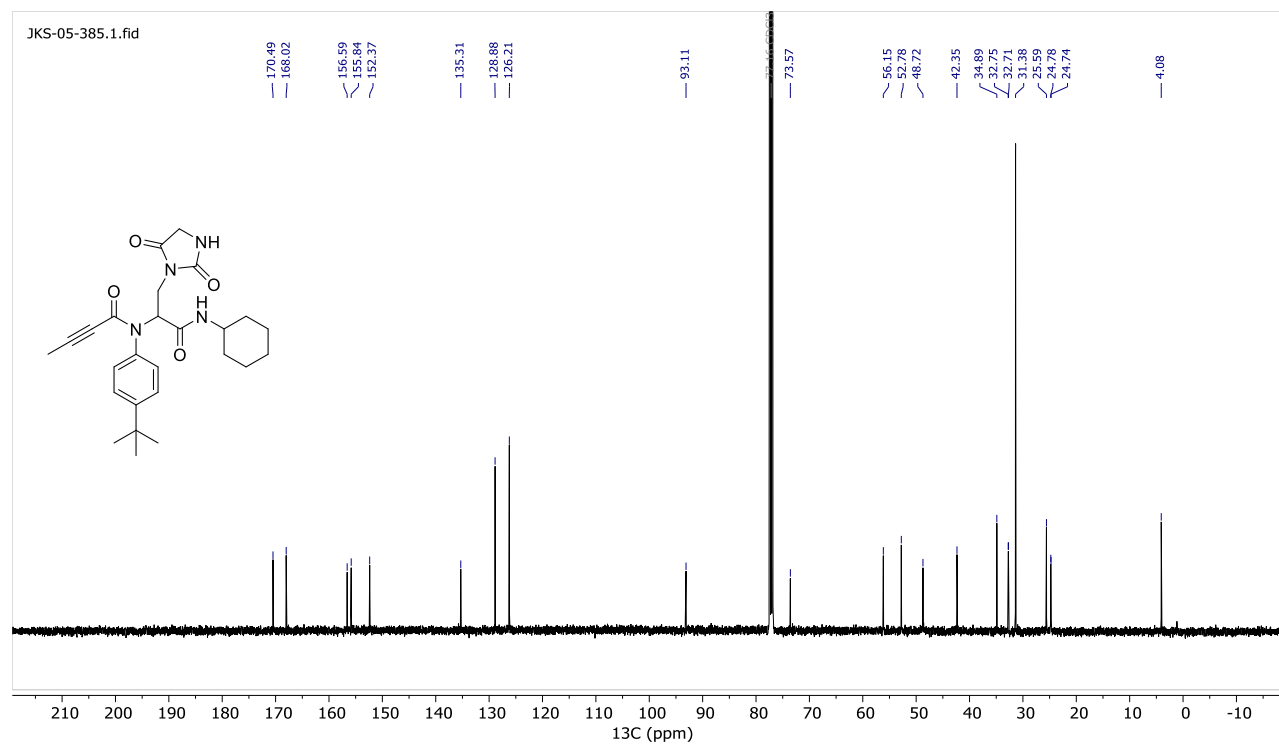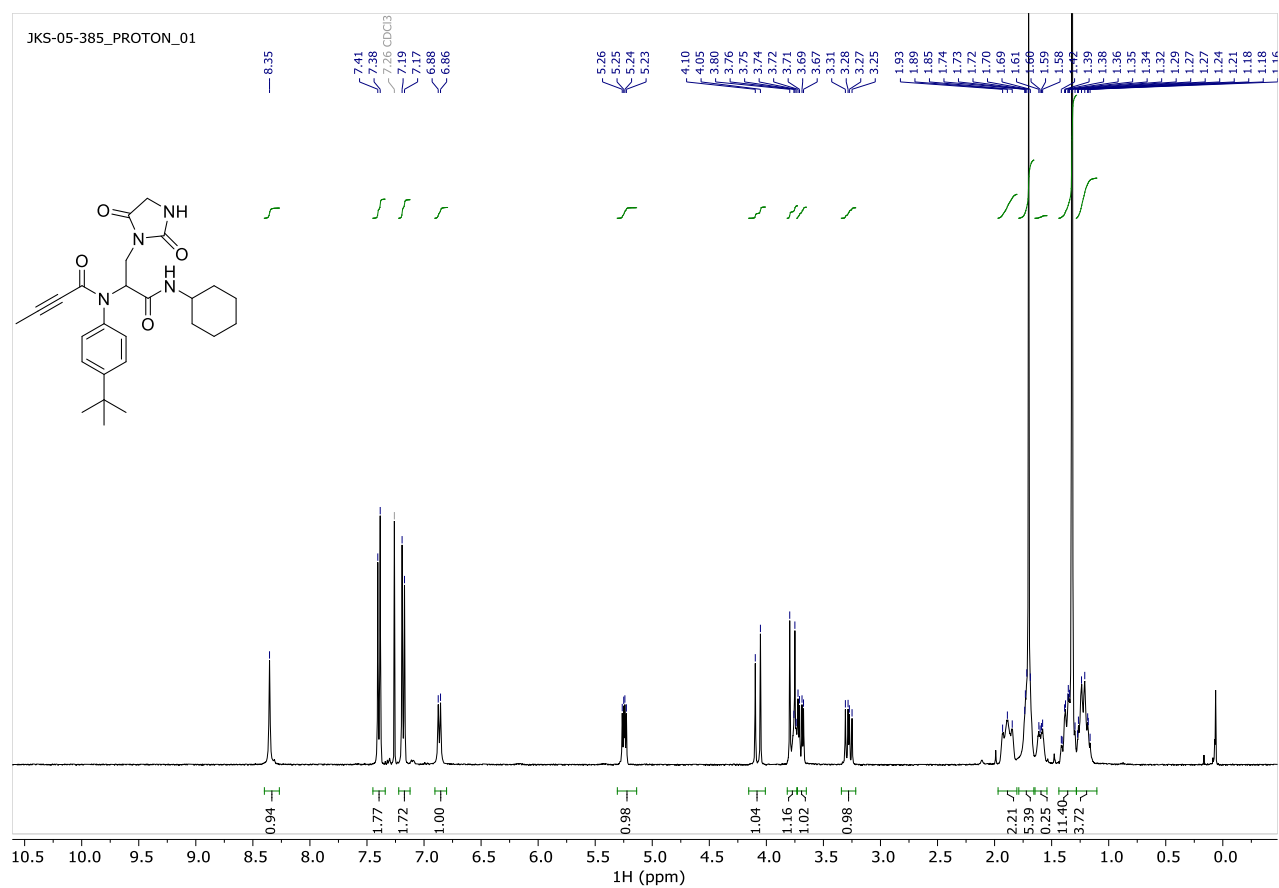
Compound 5.24 ¹H and ¹³C NMR
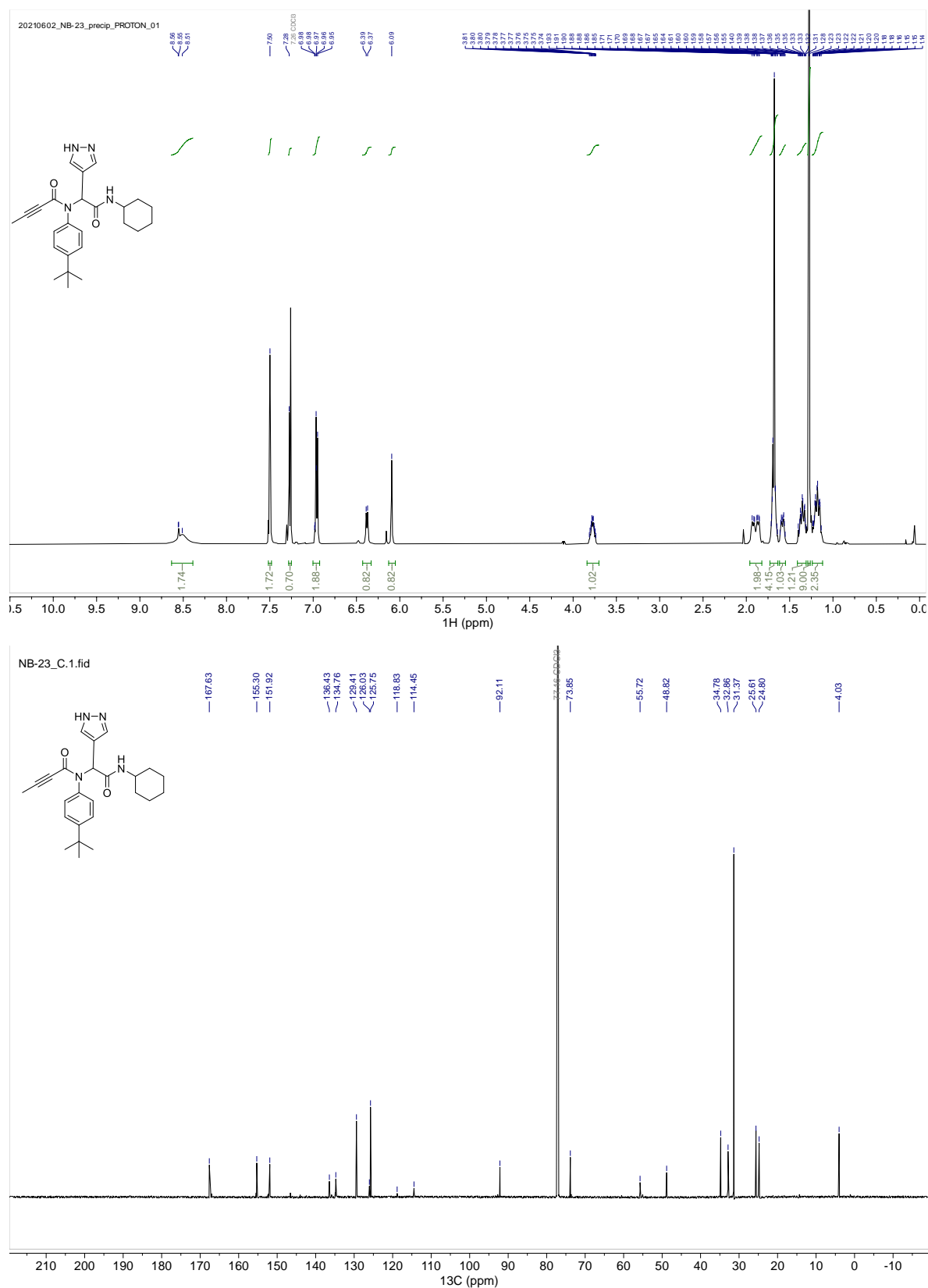
Compound 5.25 ^{1}H and ^{13}C NMR

JT-19.4.fid
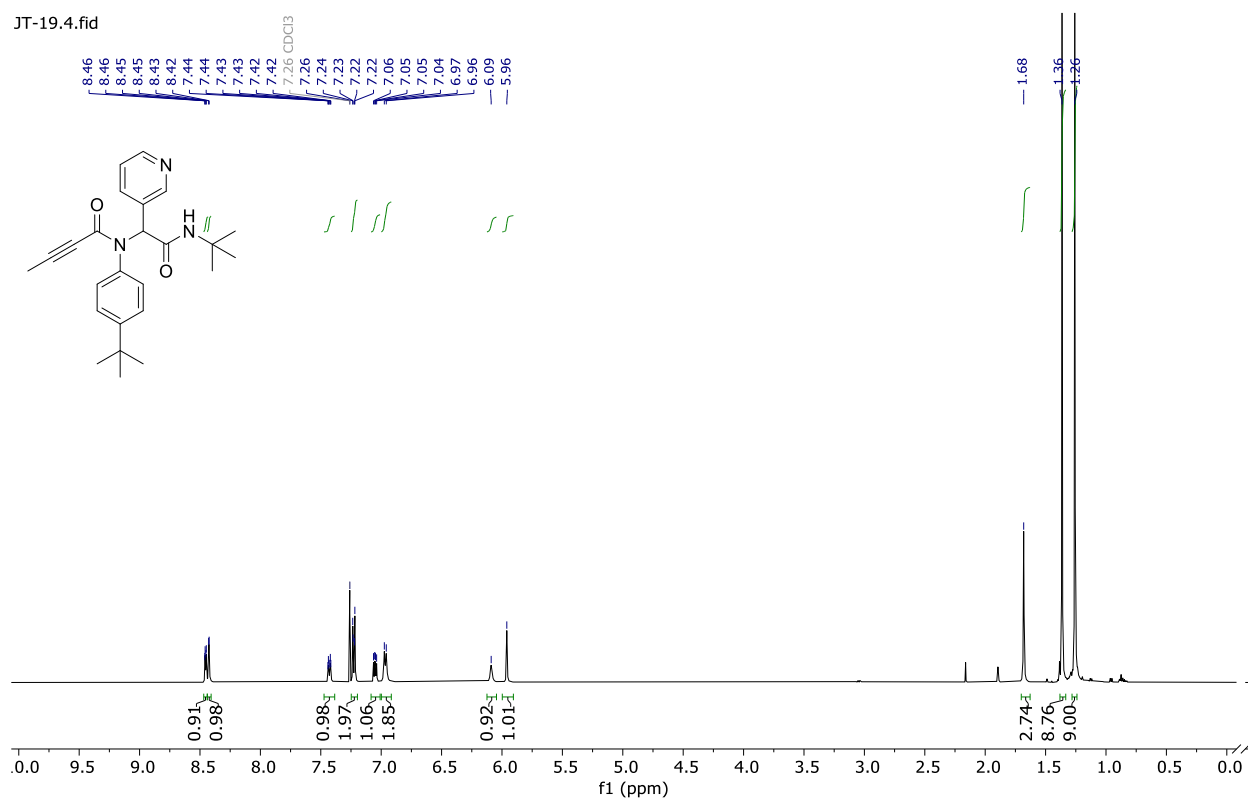
JT-19.2.fid
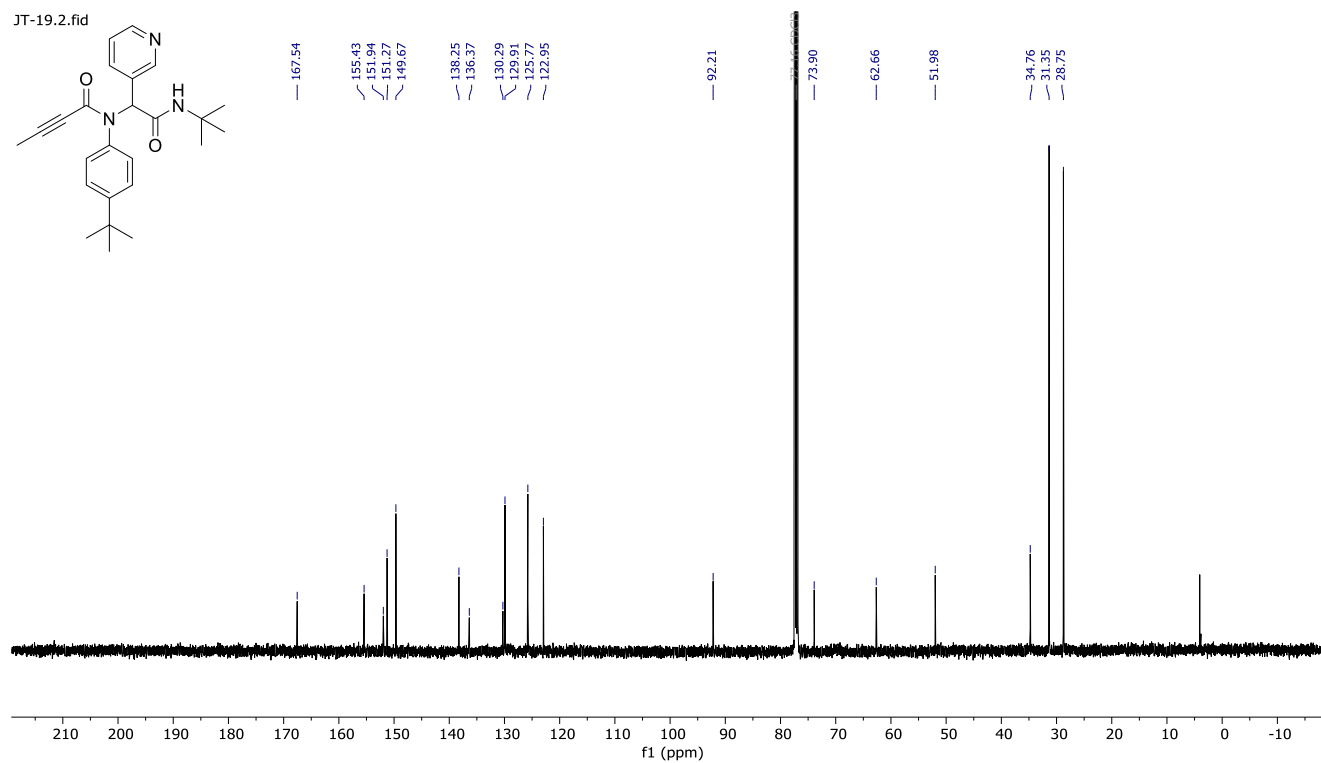
Compound 5.26 ^{1}H and ^{13}C NMR

JT-64.2.fid
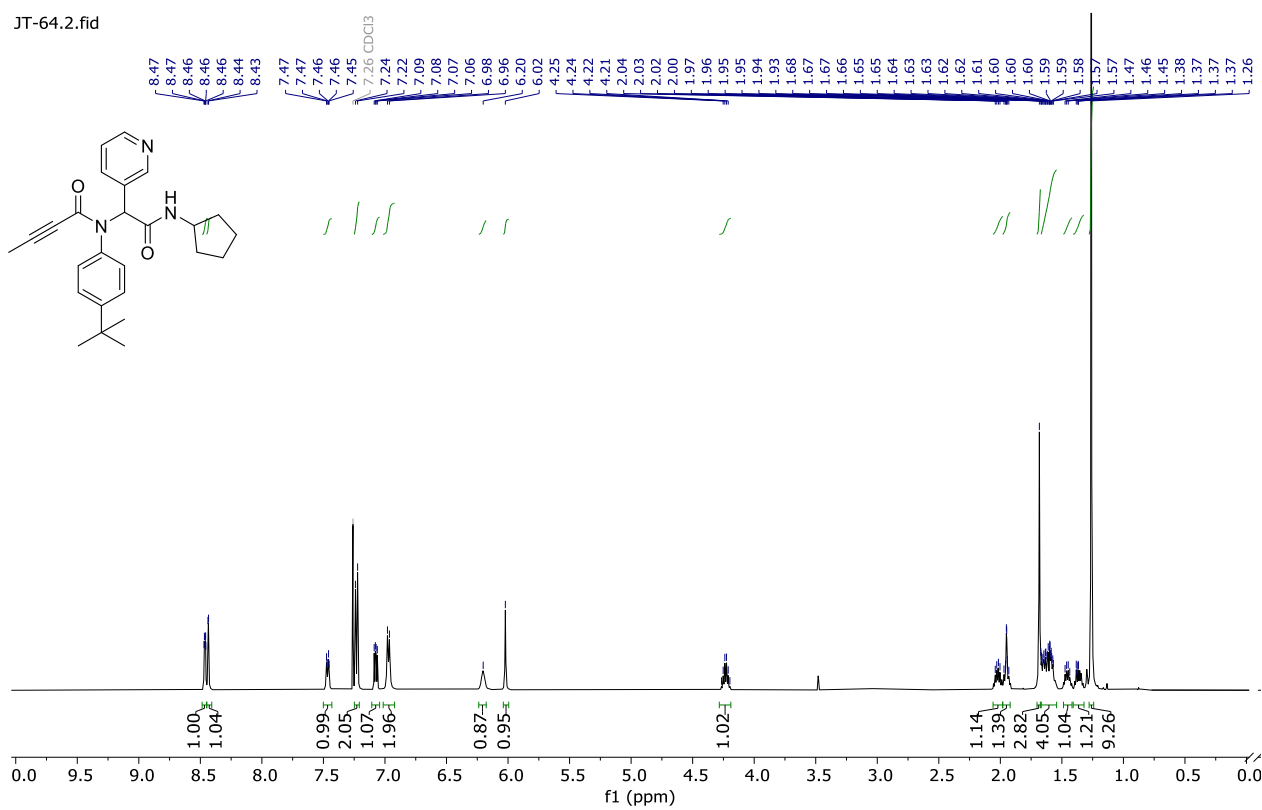
JT-64.3.fid
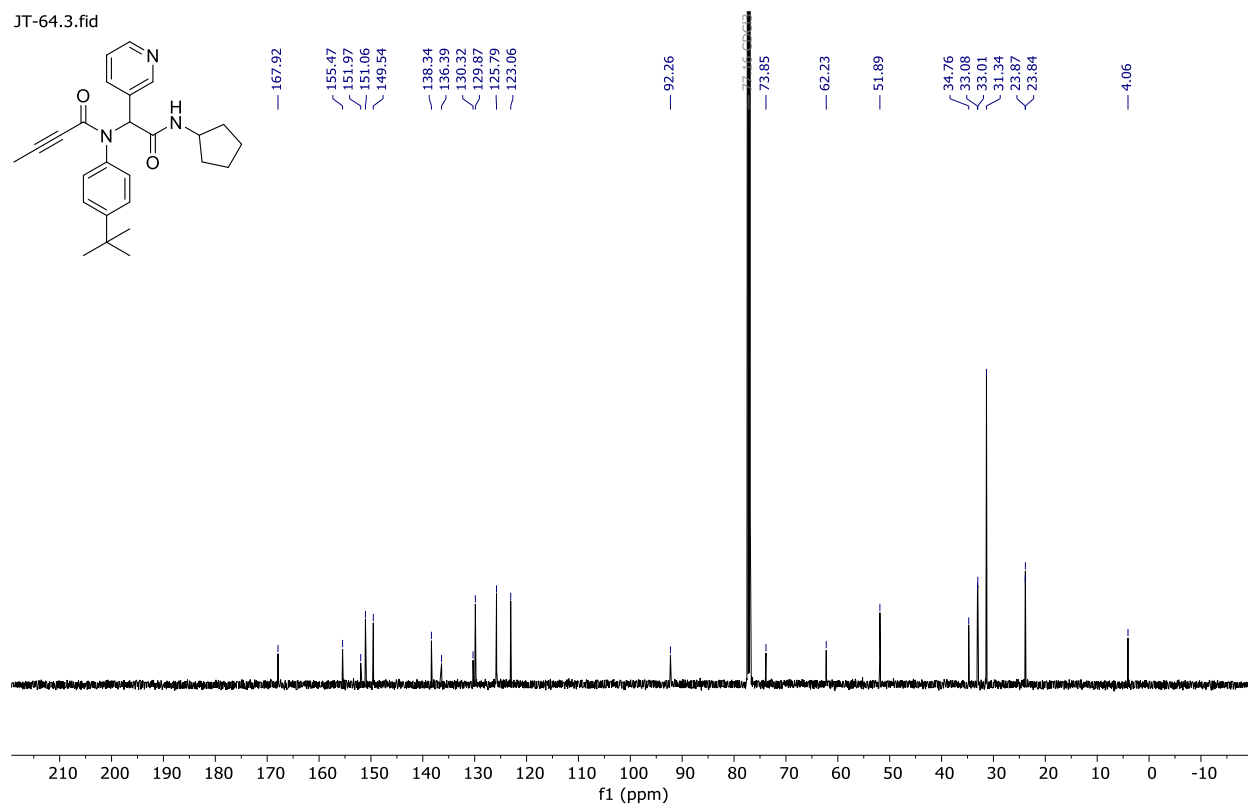
Compound 5.27 ^{1}H and ^{13}C NMR

JT-65.2.fid
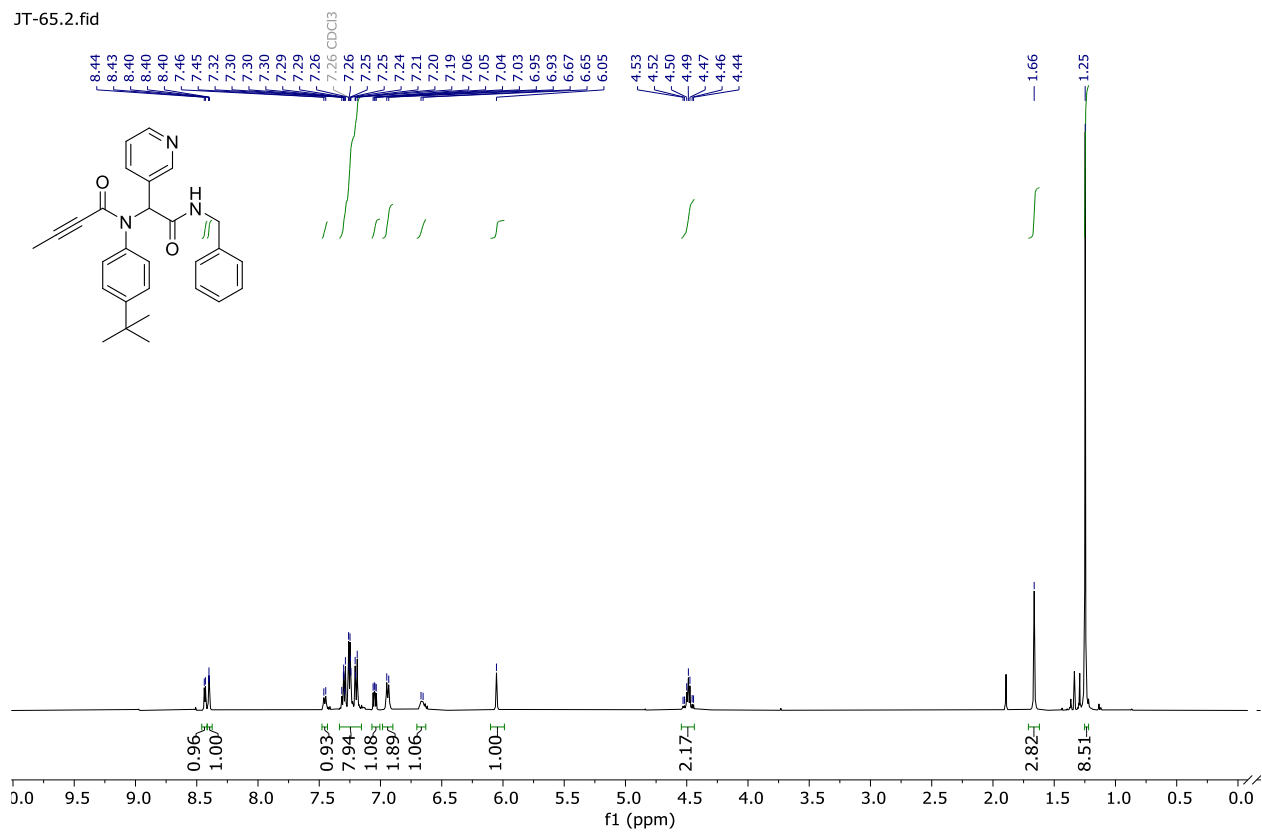
JT-65.3.fid
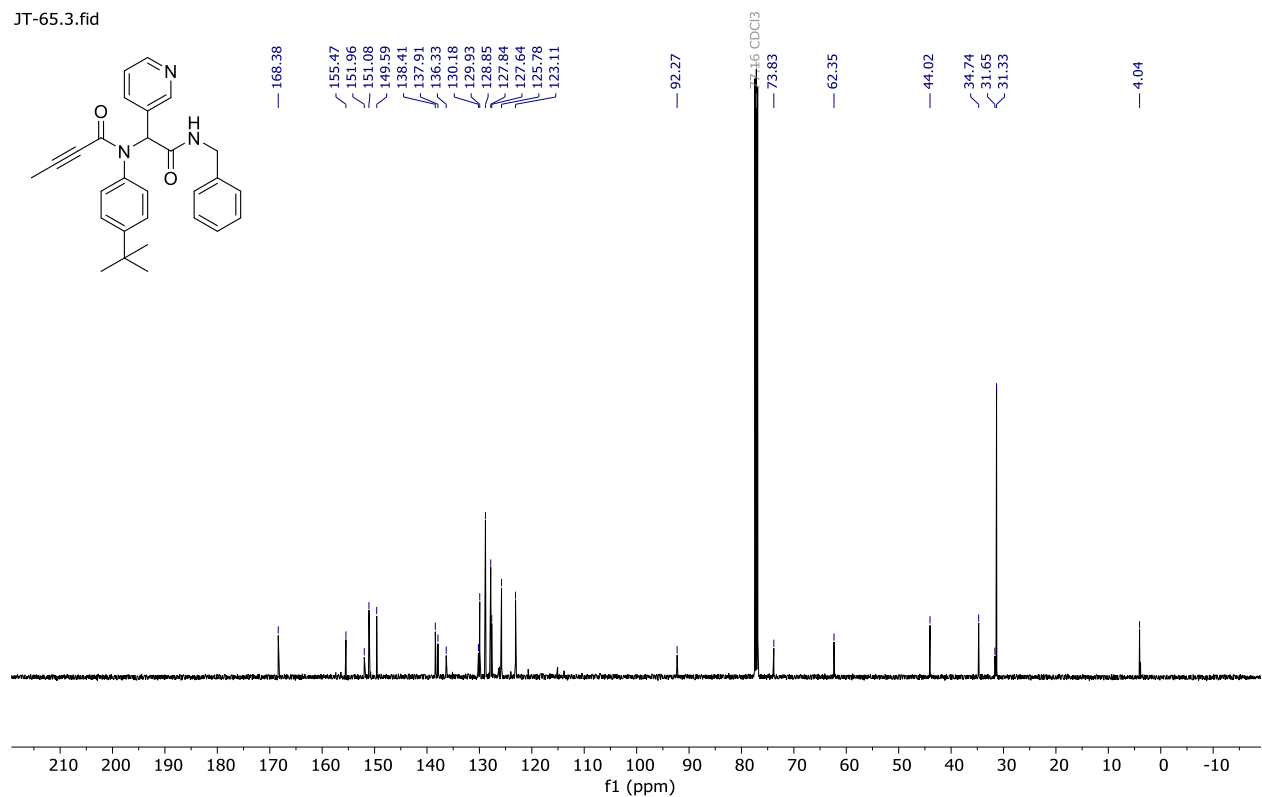
Compound 5.28 ^{1}H and ^{13}C NMR
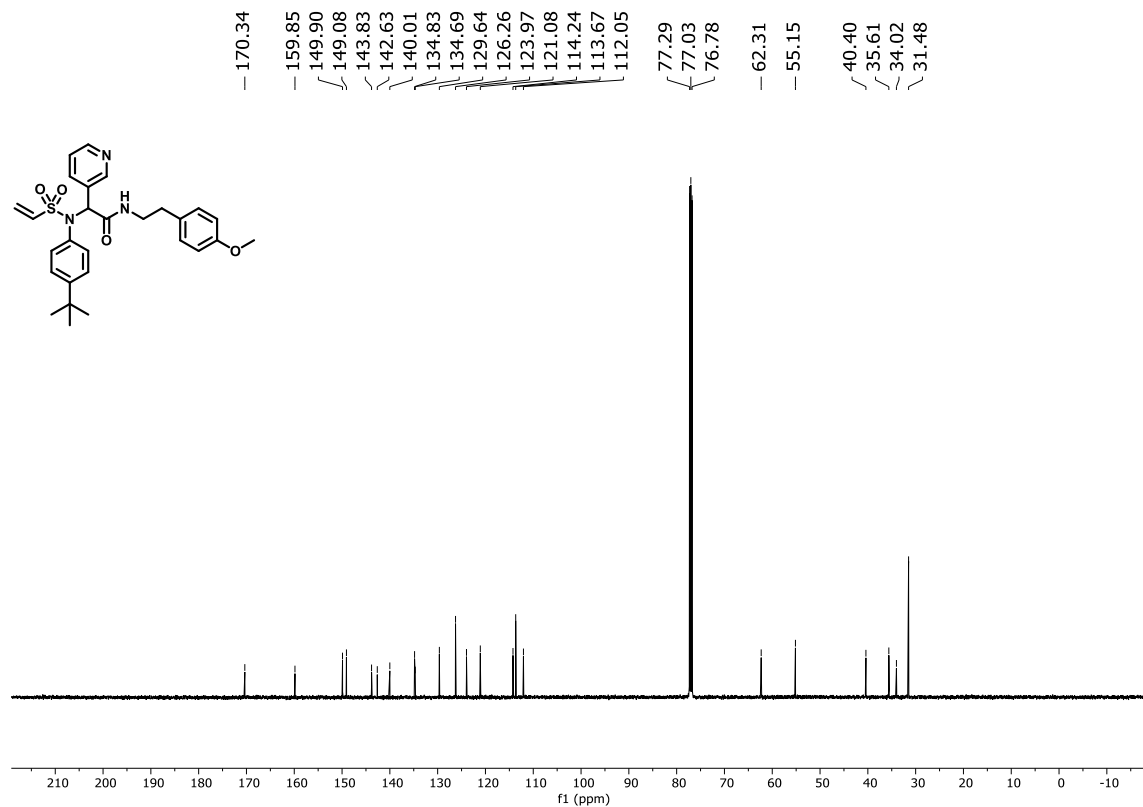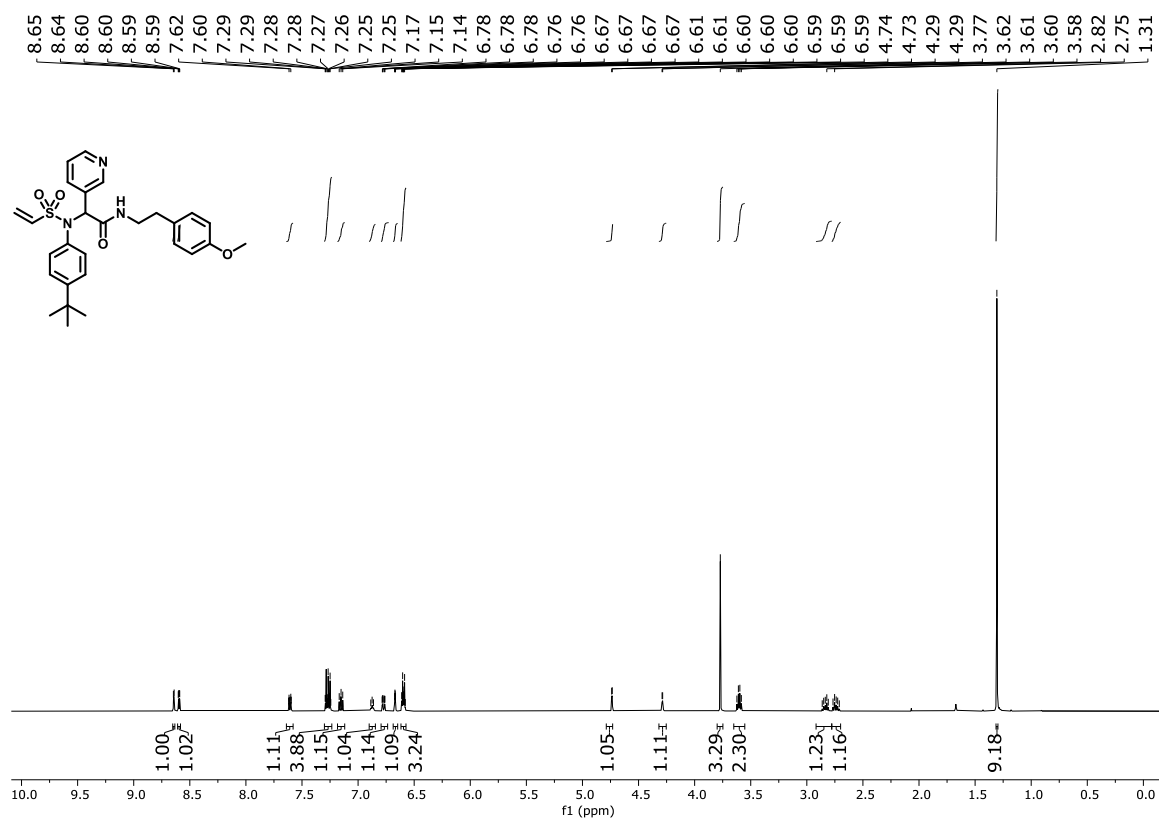
Compound 5.29 ¹H and ¹³C NMR
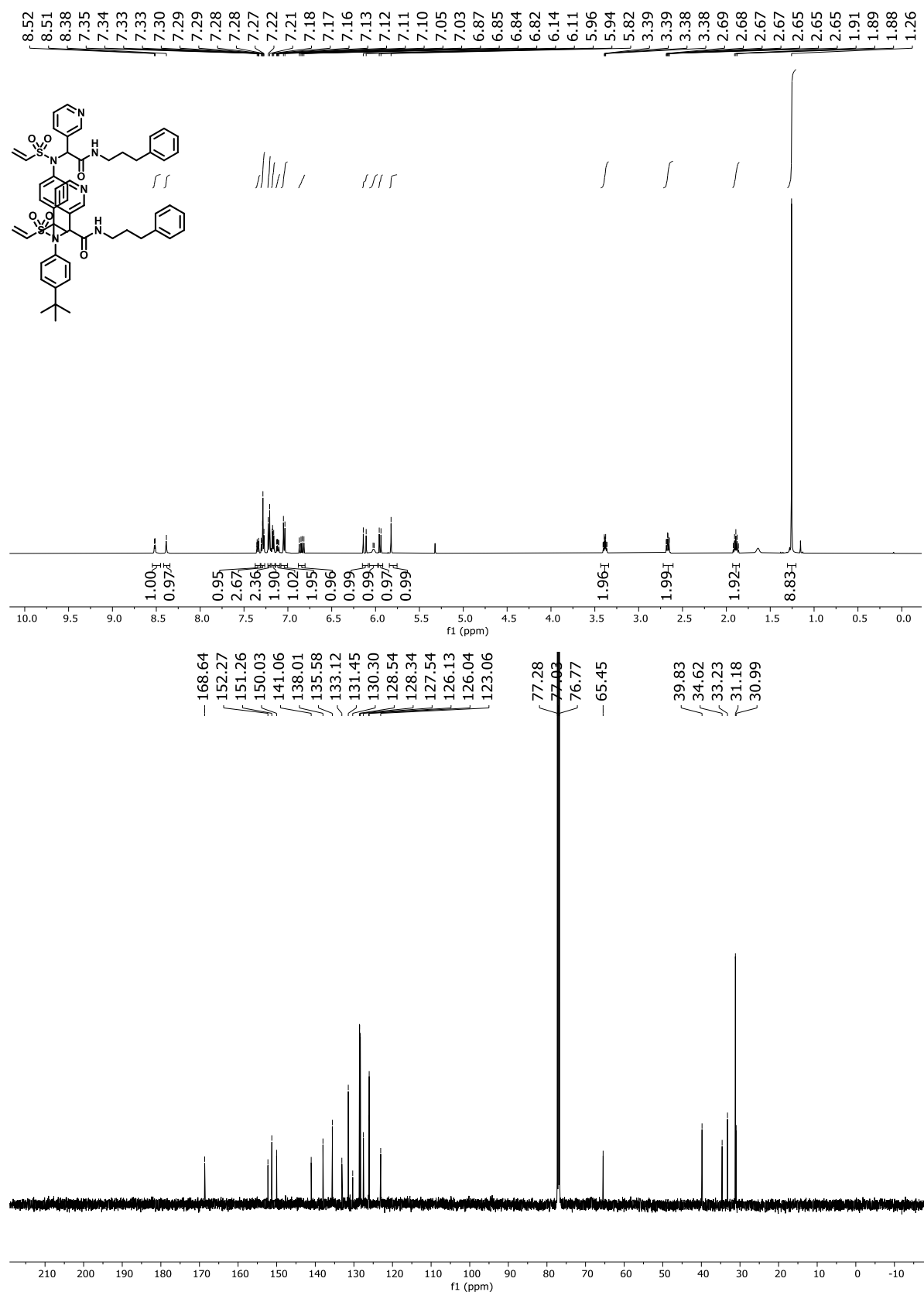
Compound 5.30 ¹H and ¹³C NMR
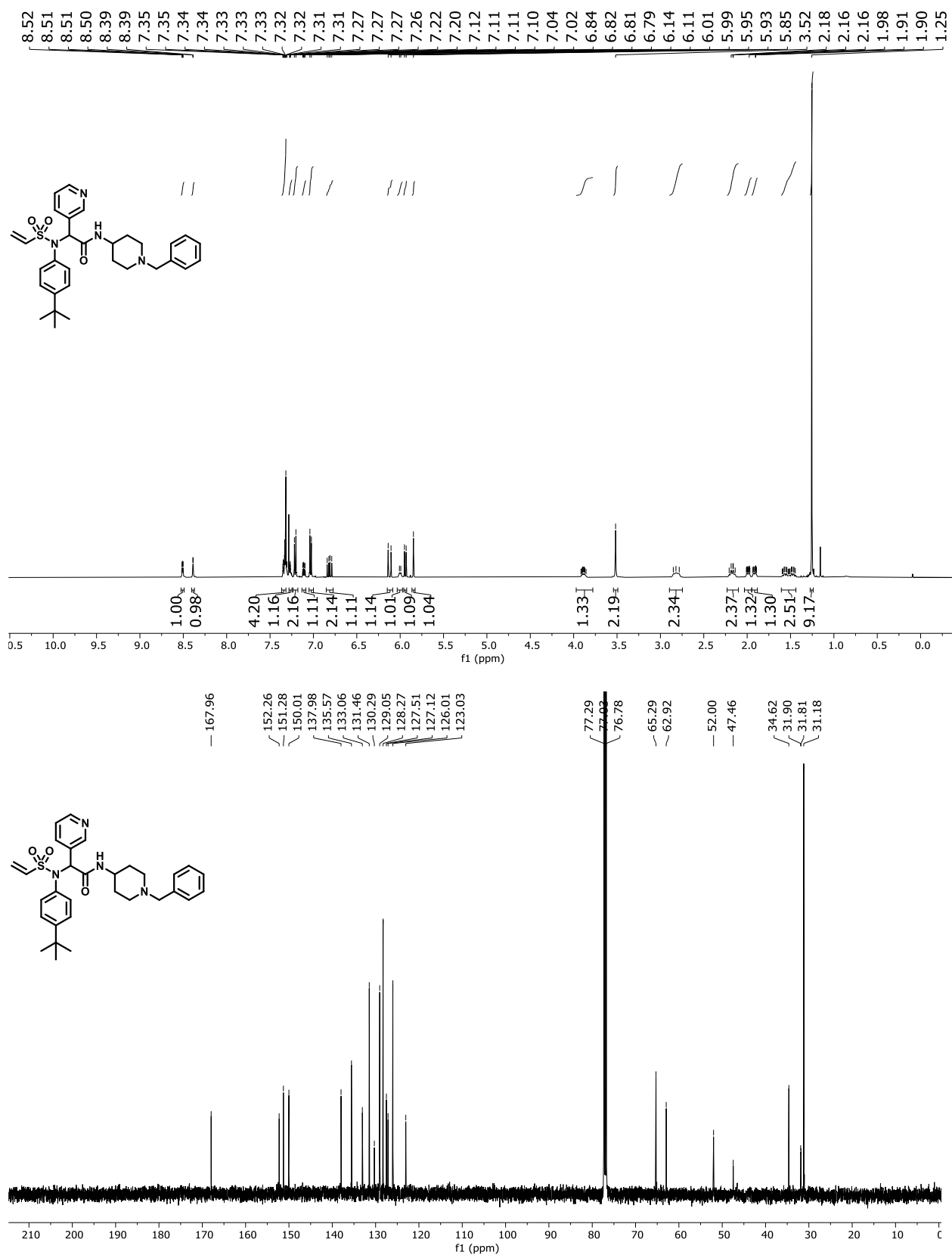
Compound 5.31-int-1 ^{1}H and ^{13}C NMR
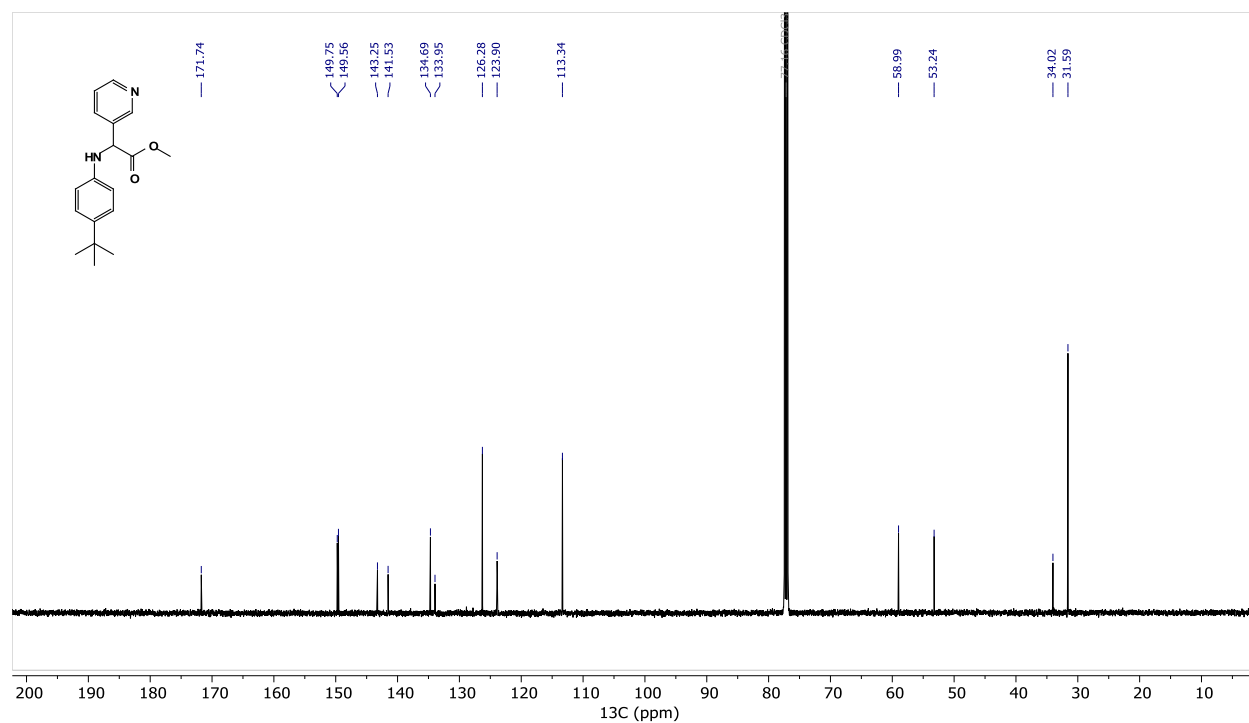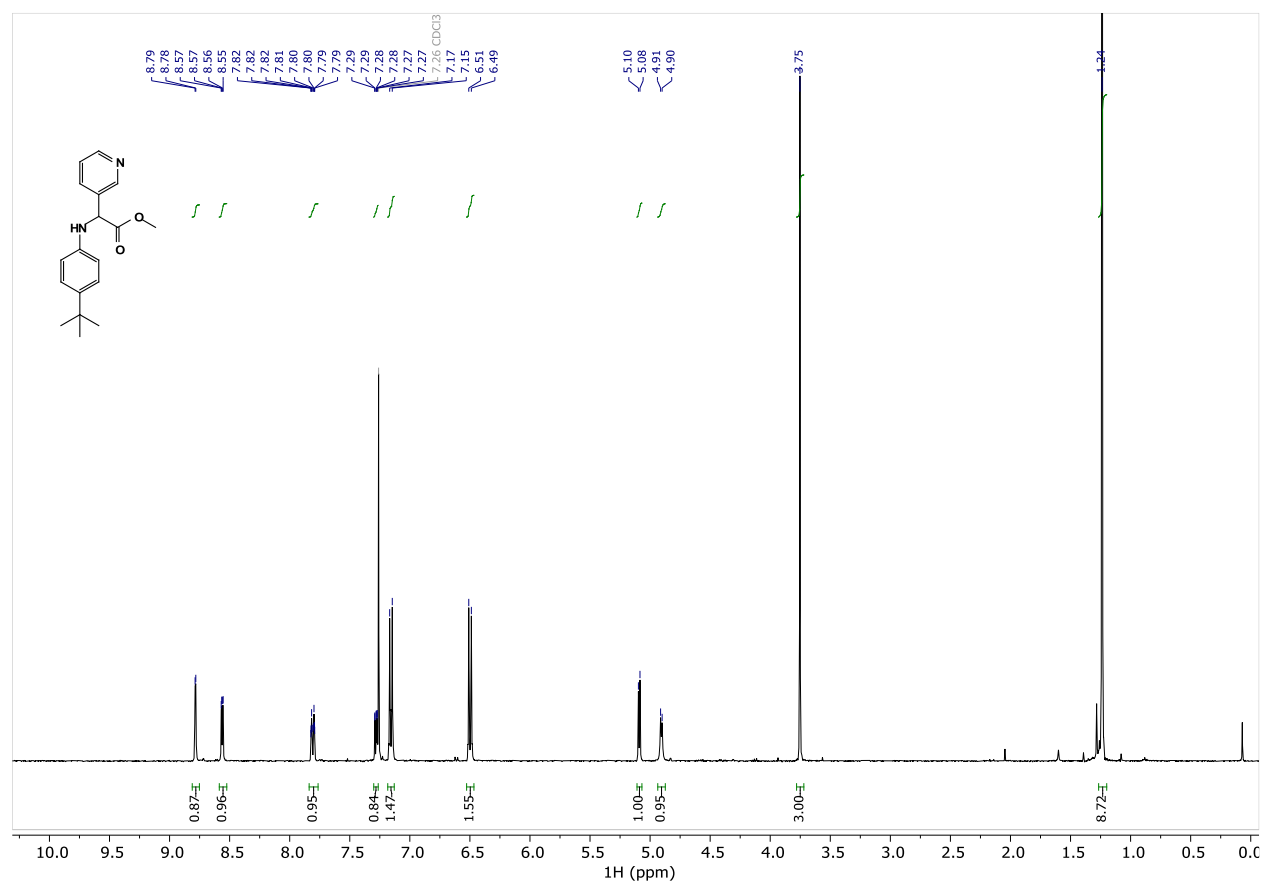
Compound 5.31-int-2 ^{1}H and ^{13}C NMR
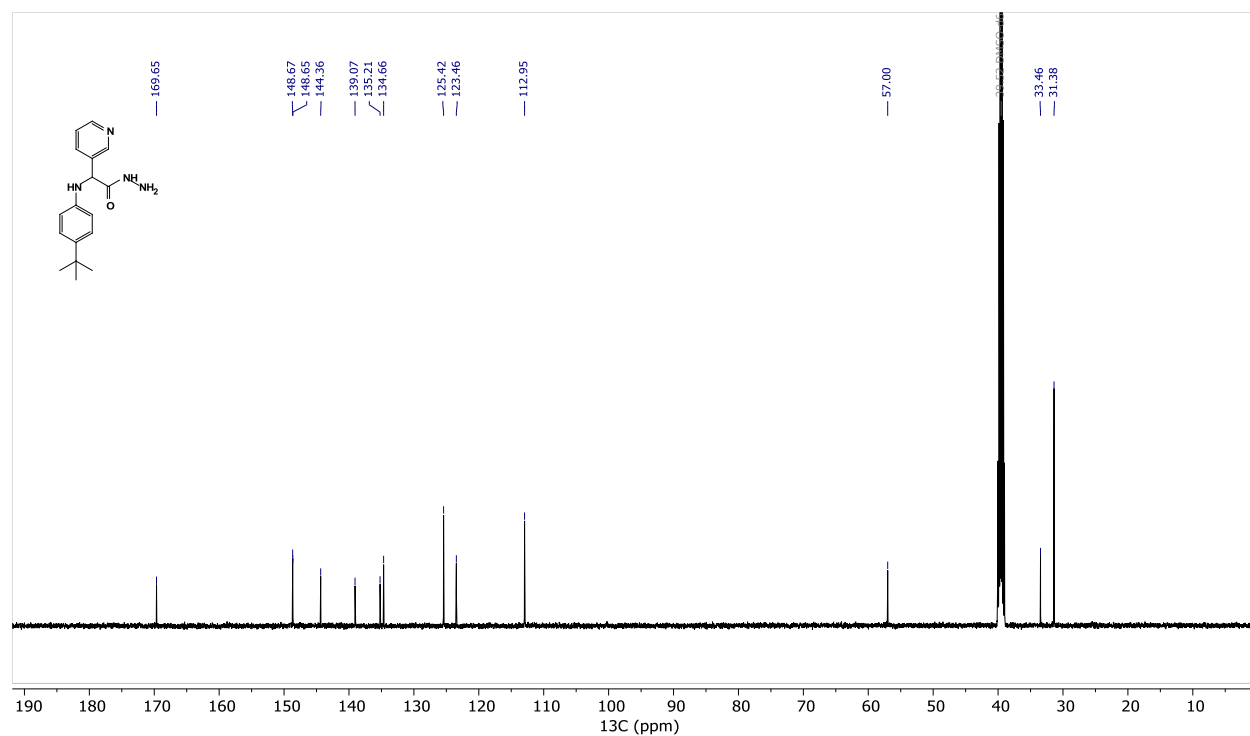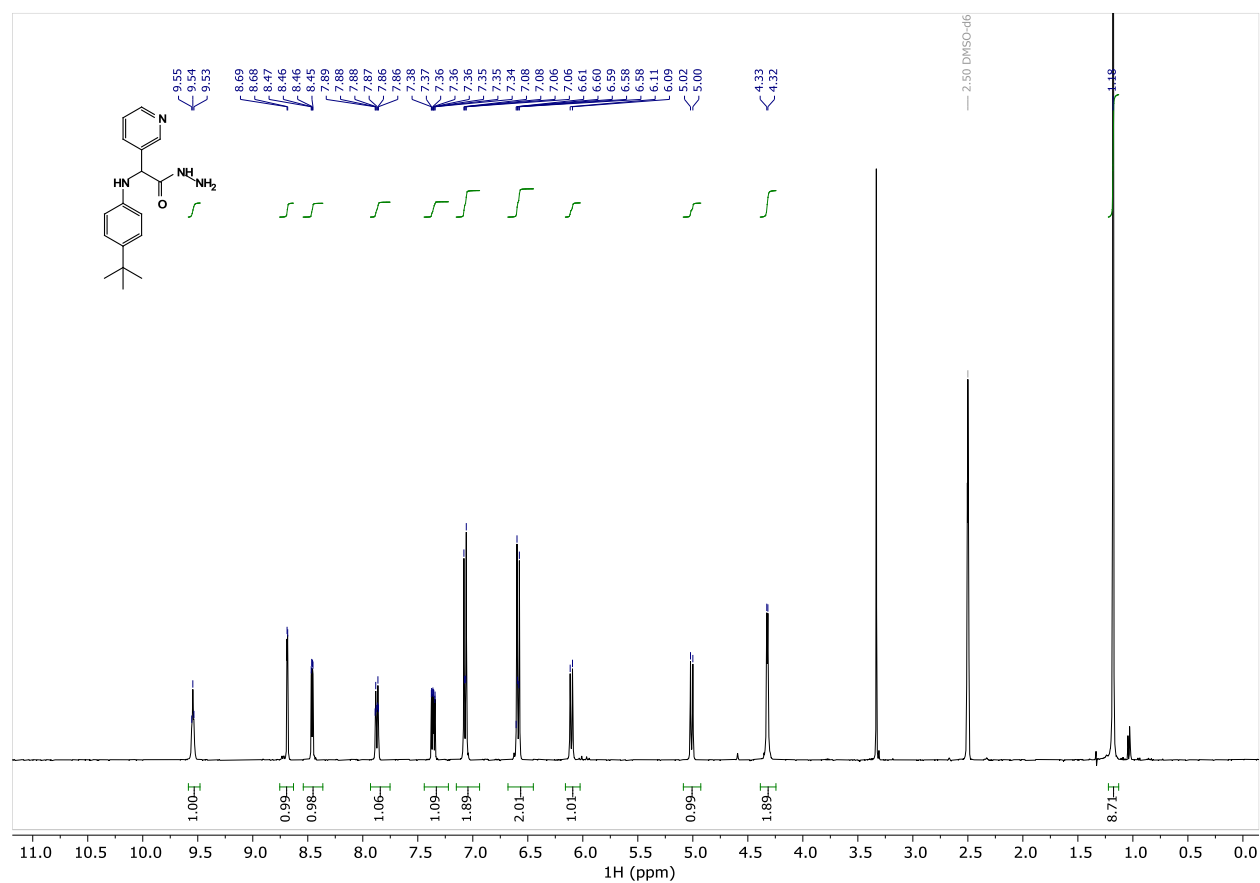
Compound 5.31-int-3 ^{1}H and ^{13}C NMR
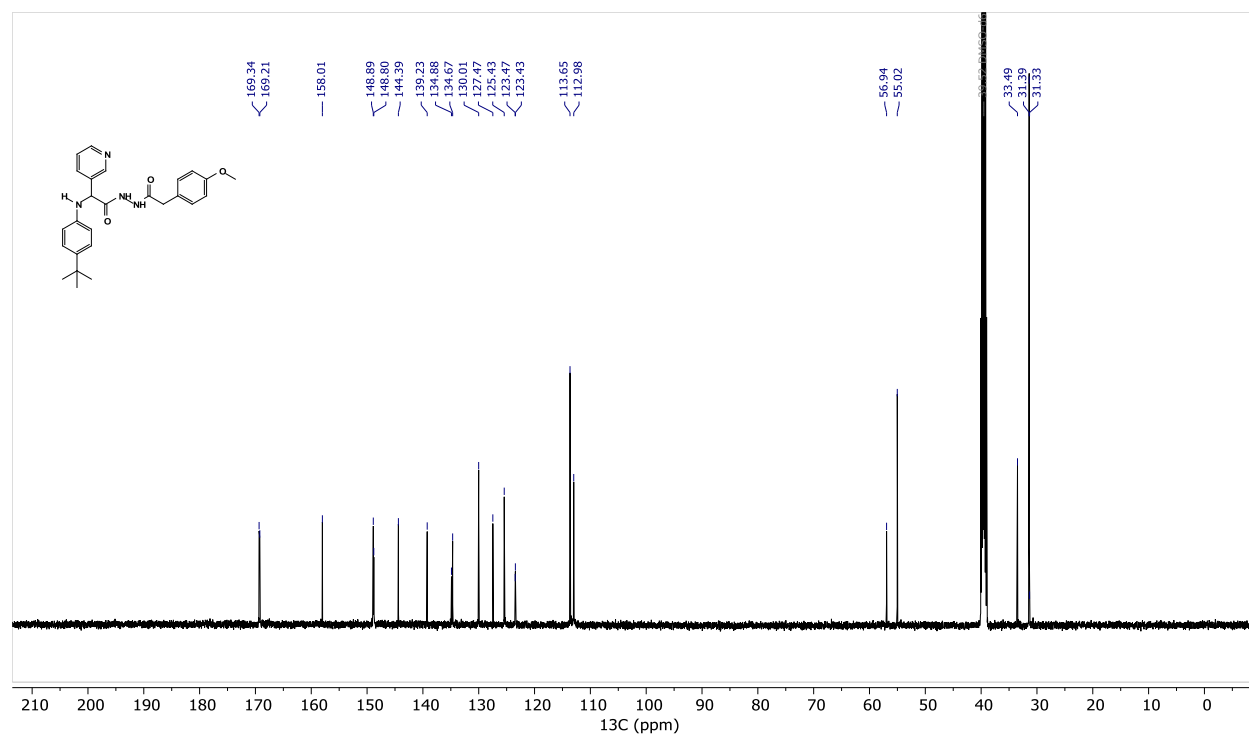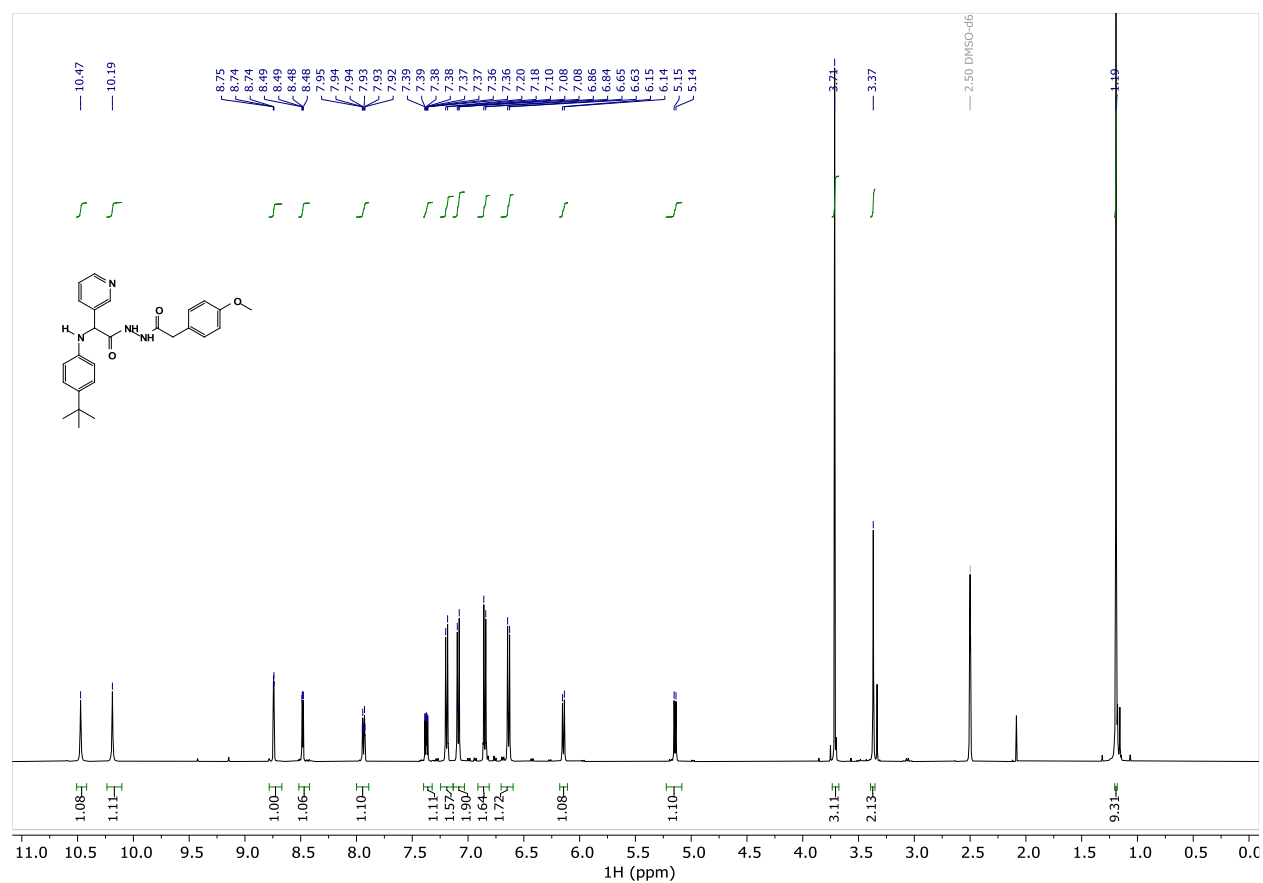
Compound 5.31-int-4 ^{1}H and ^{13}C NMR
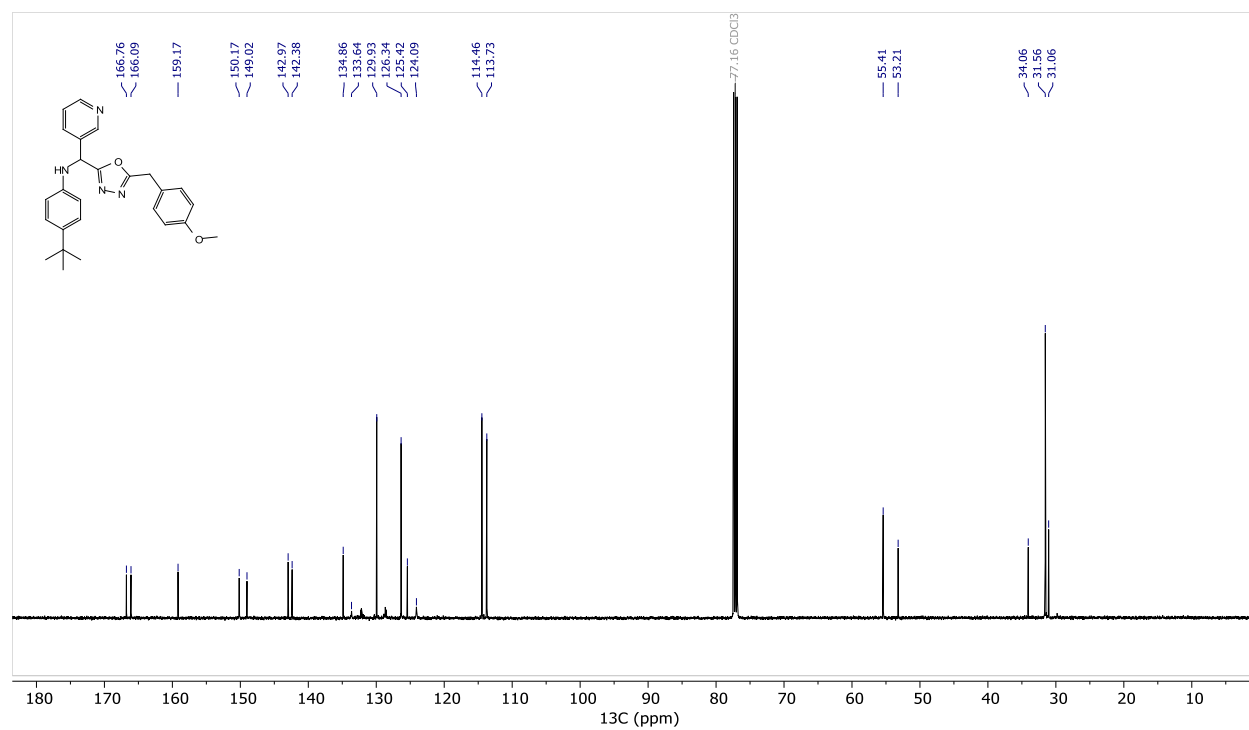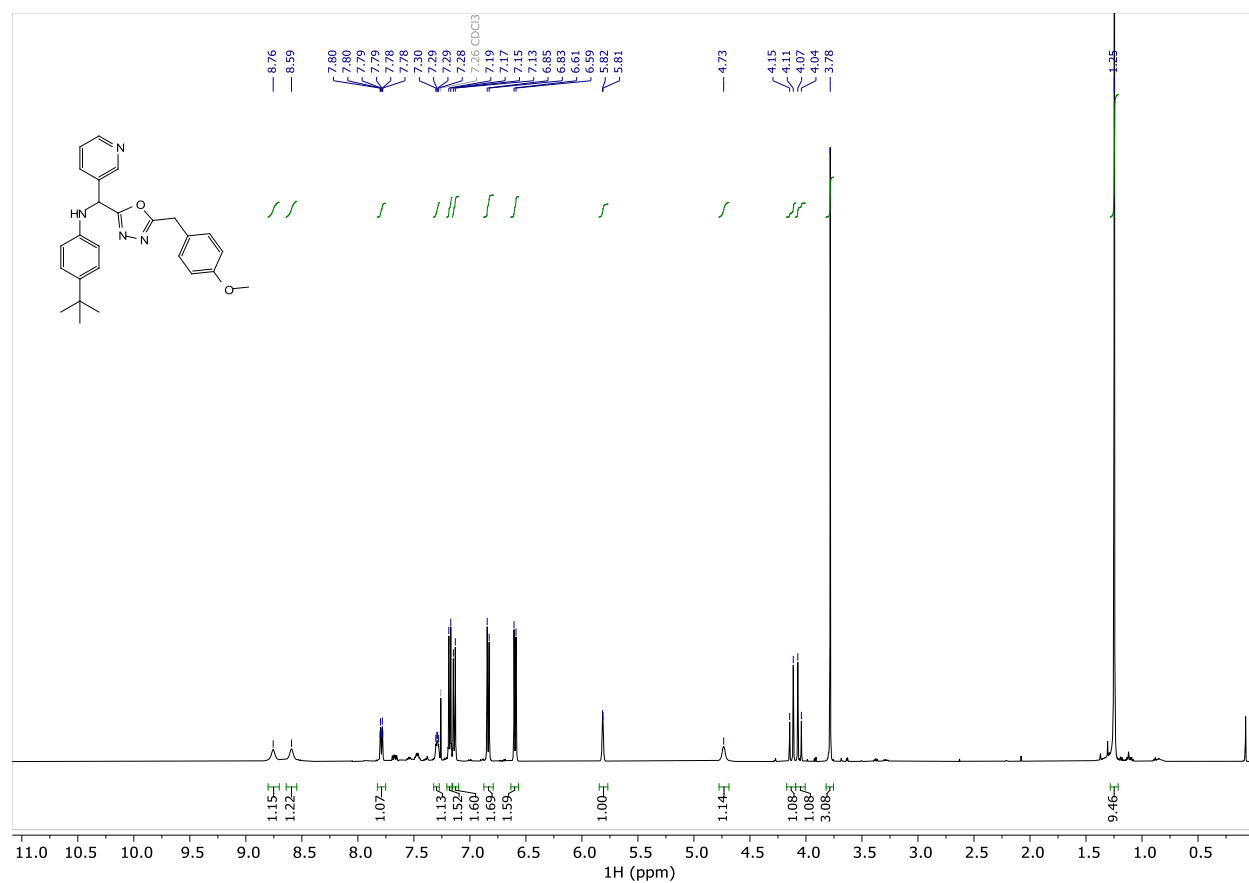
Compound 5.31 ^{1}H and ^{13}C NMR
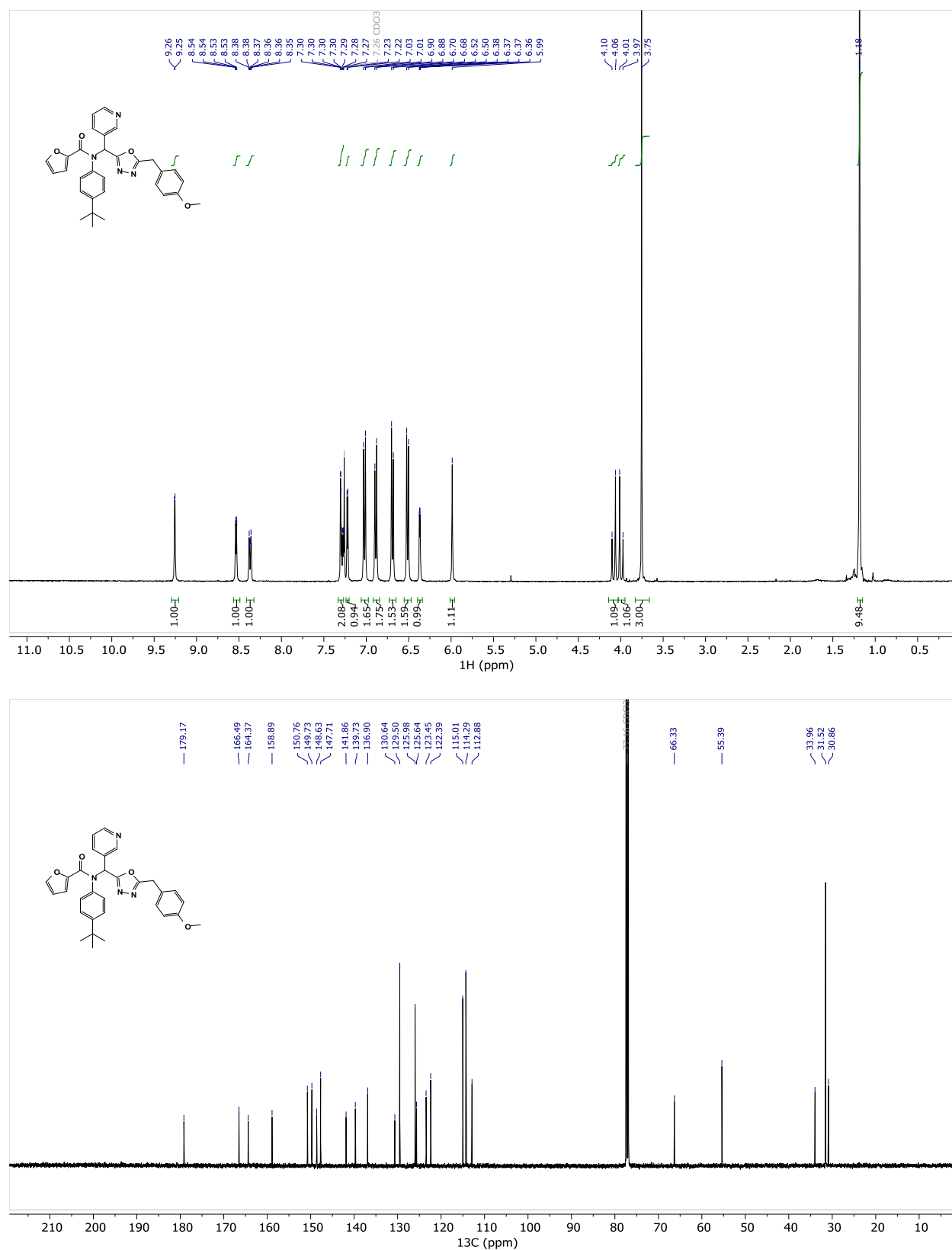
Compound 5.32 ¹H and ¹³C NMR
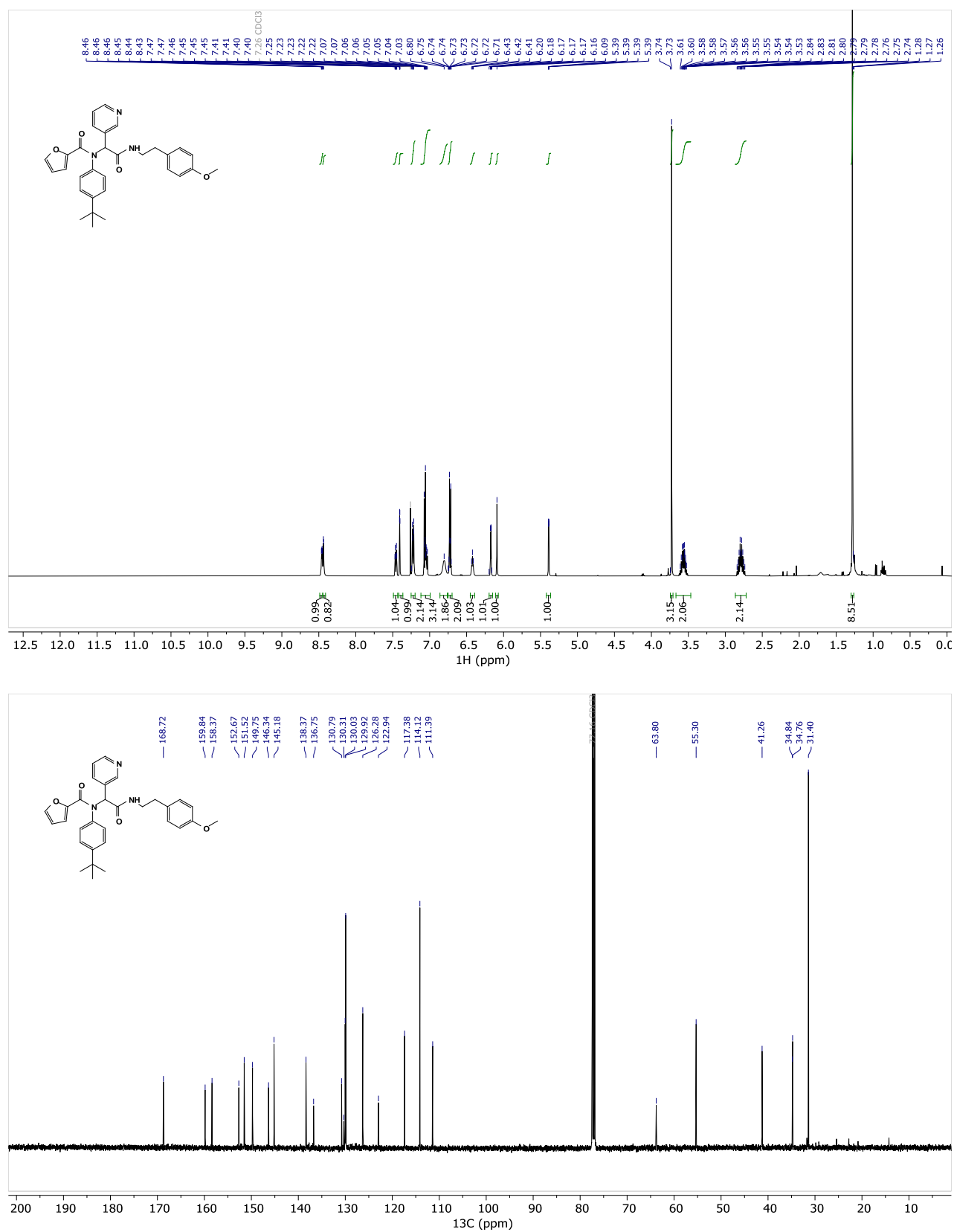
Compound ML188 ^{1}H and ^{13}C NMR

JT-52.10.fid
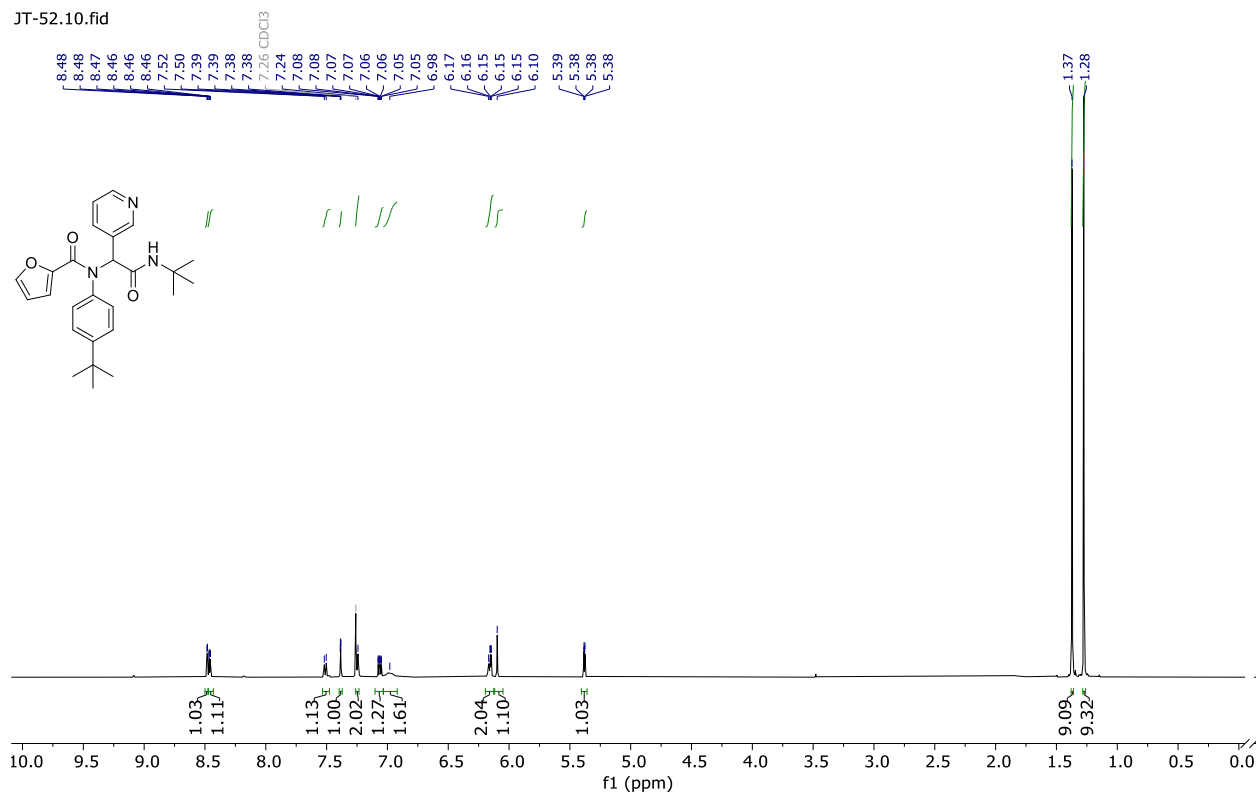
JT-52.11.fid
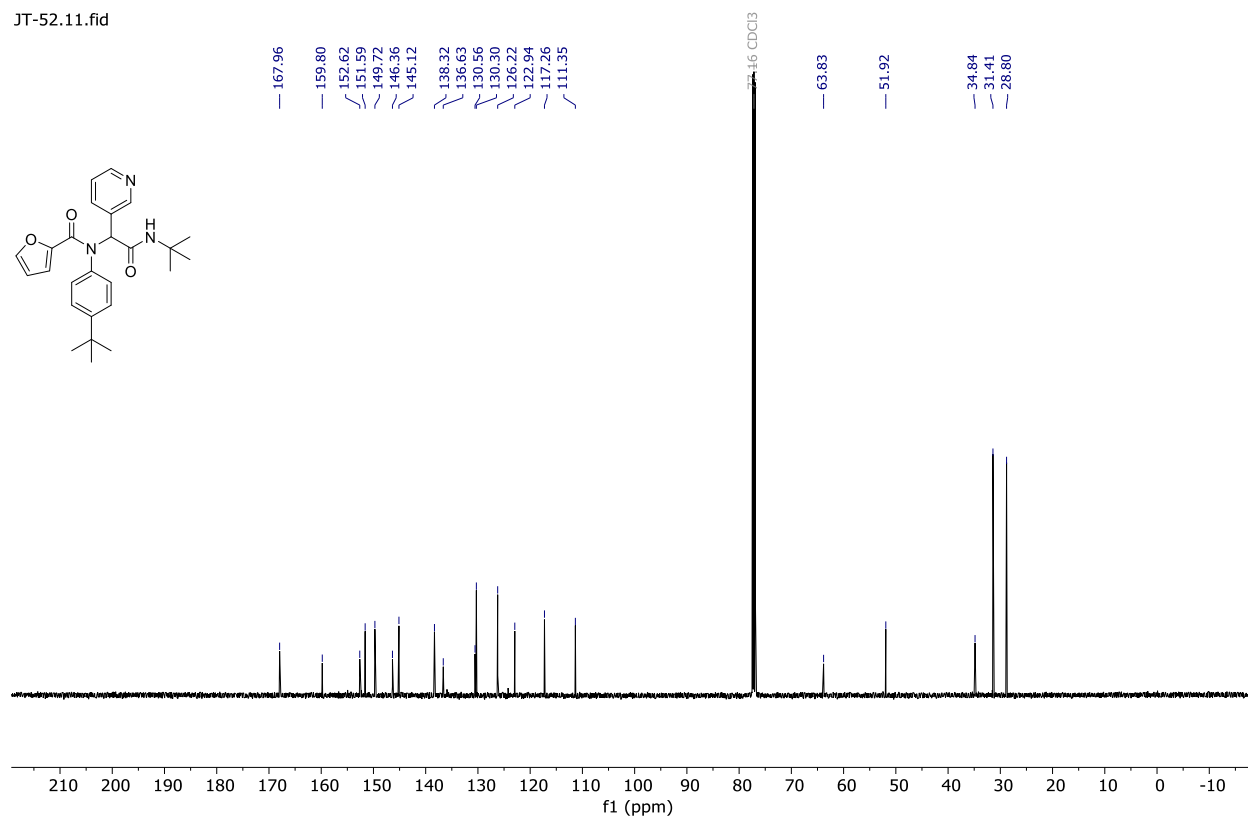
SUPPORTING INFORMATION FOR CHAPTER 6

NMR Spectra

Compound 6.3 ^{1}H and ^{13}C NMR
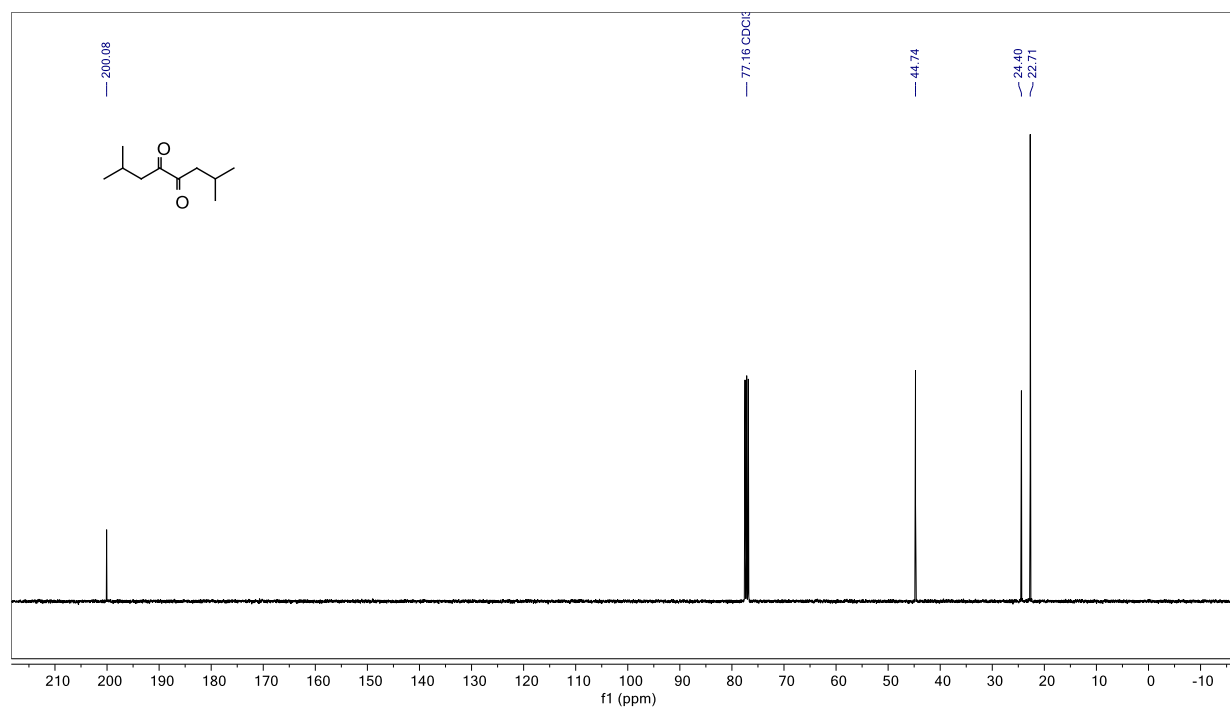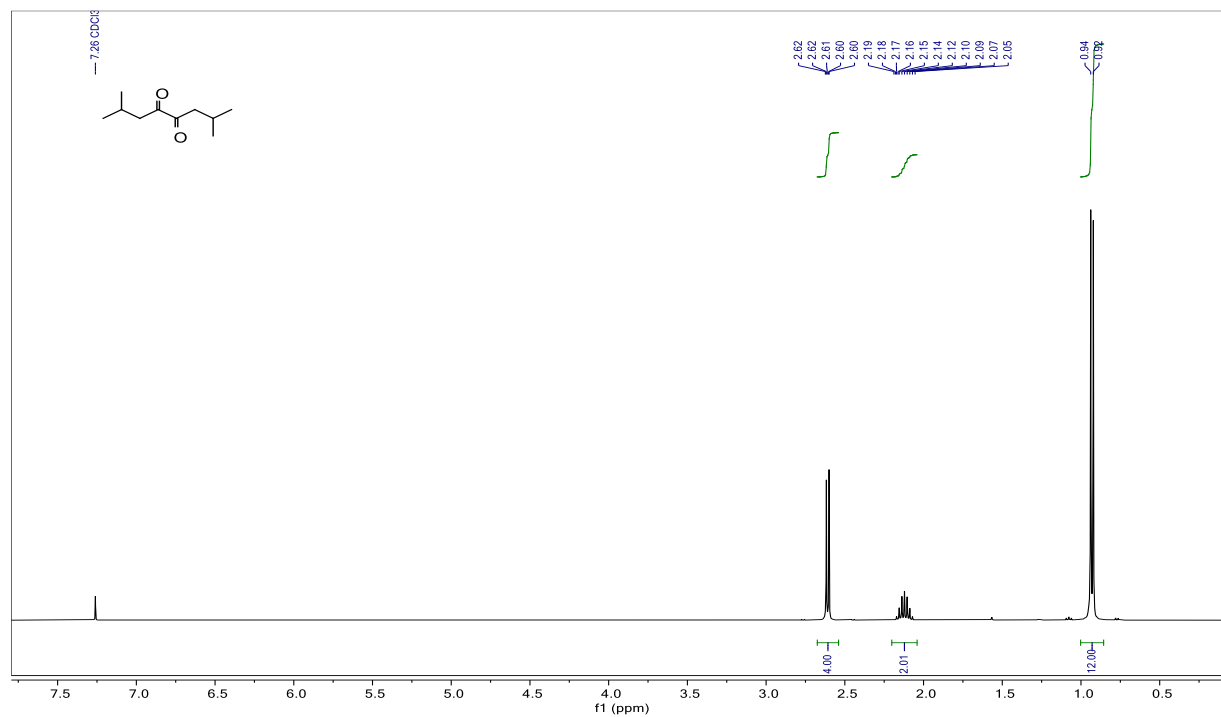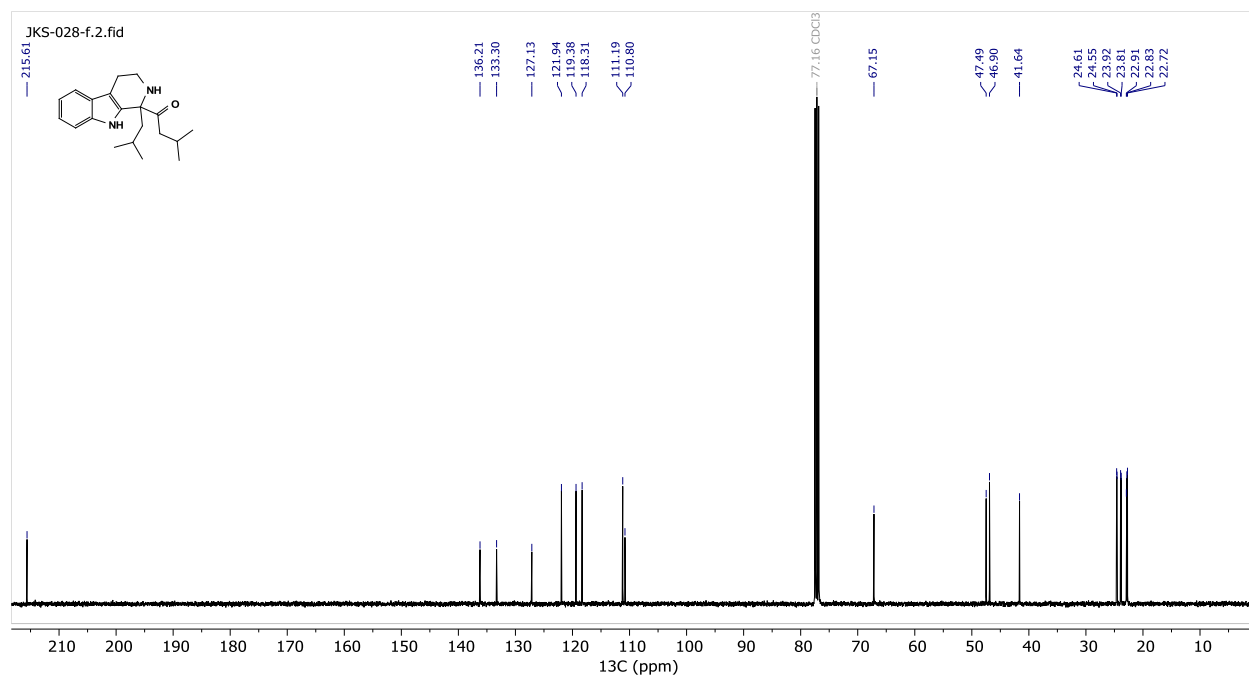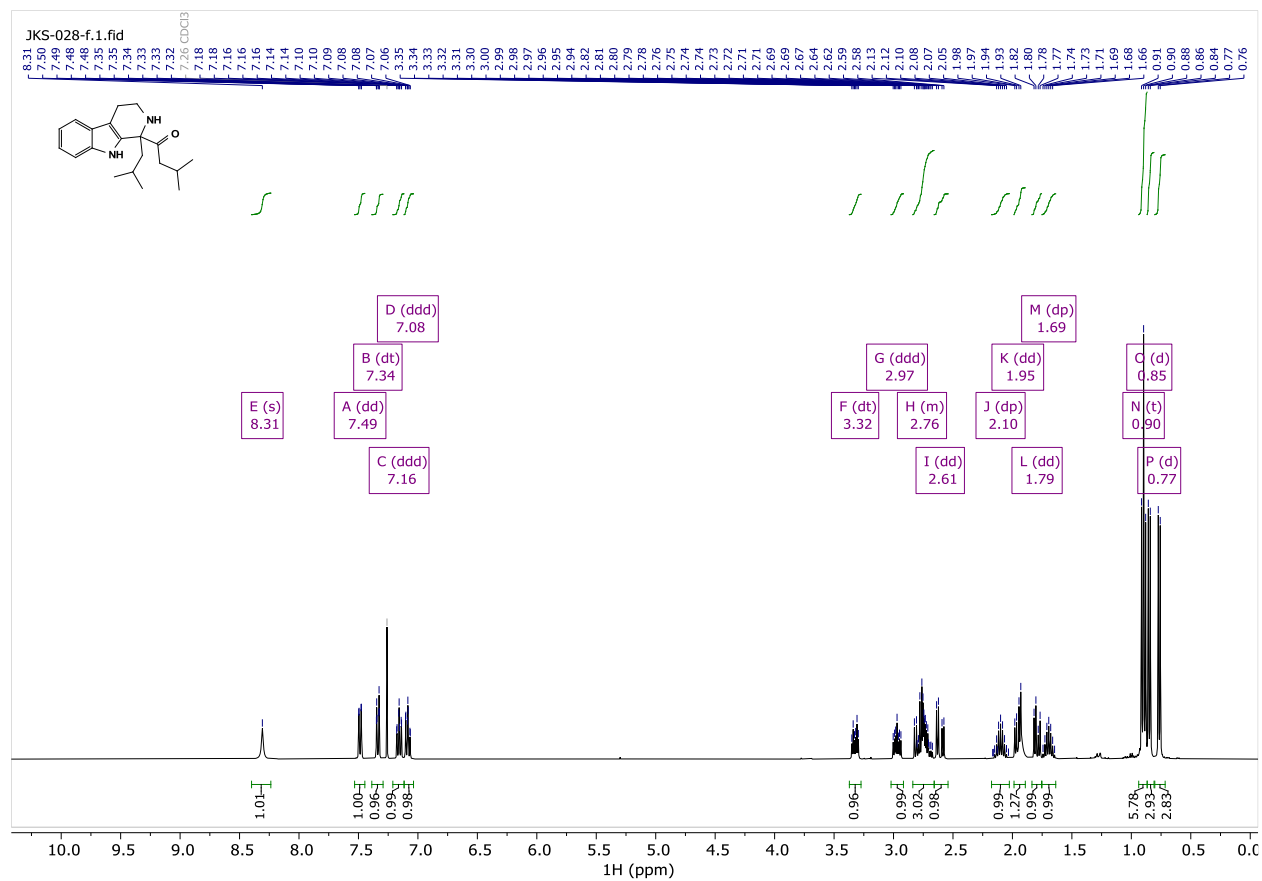
Compound 6.4 ^{1}H and ^{13}C NMR
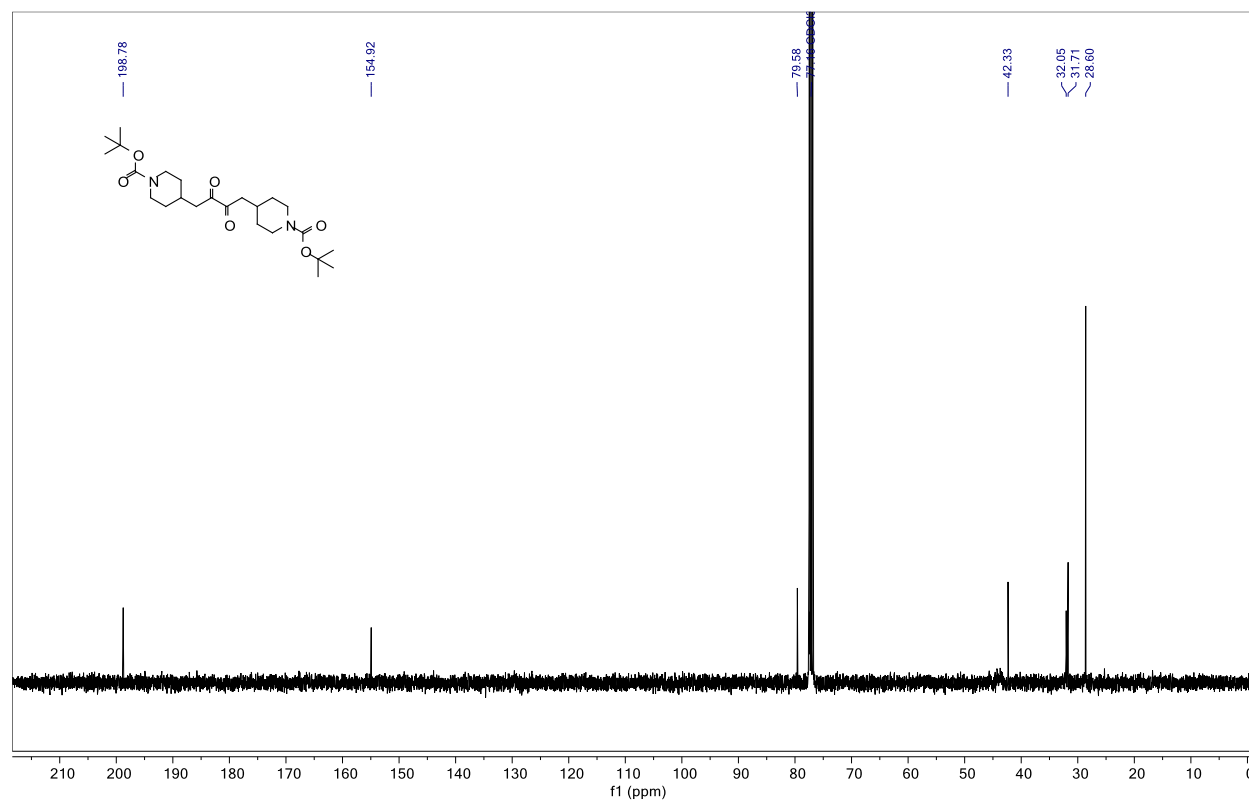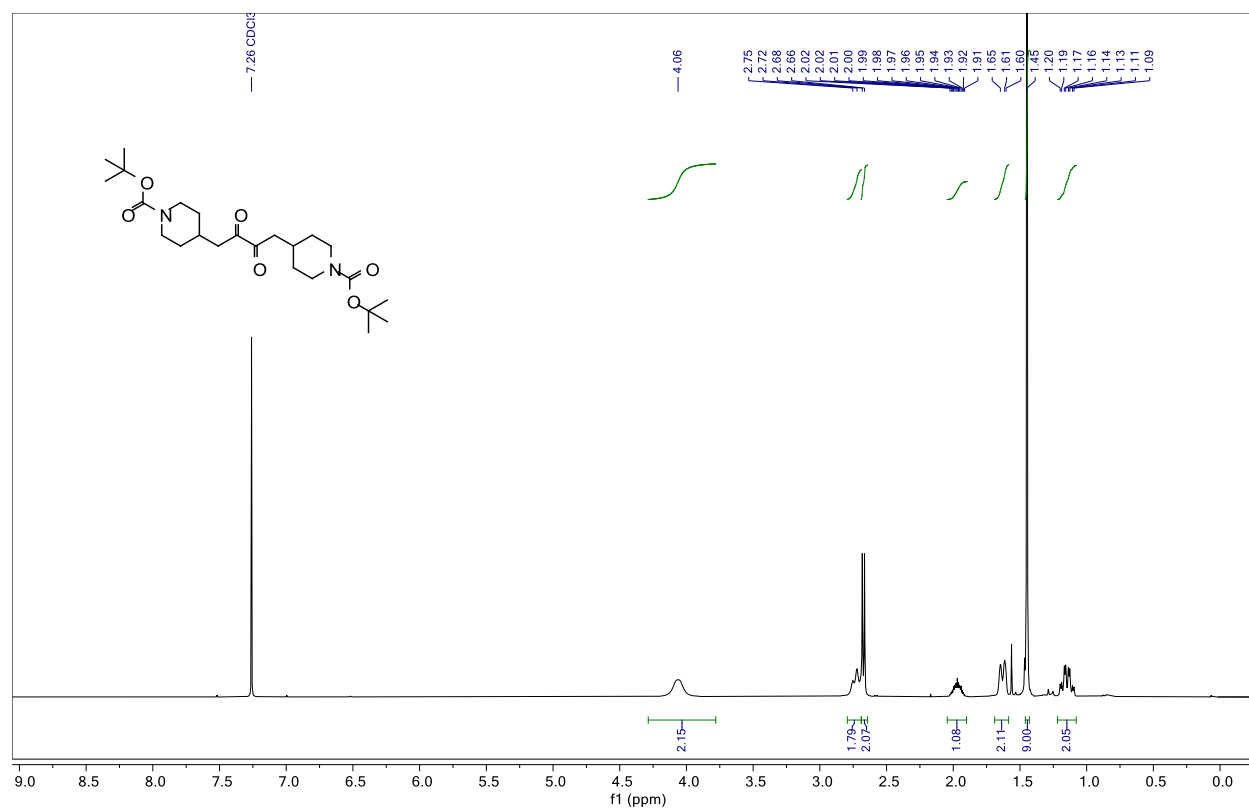
Compound 6.5 ^{1}H and ^{13}C NMR

Compound 6.15 ¹H and ¹³C NMR
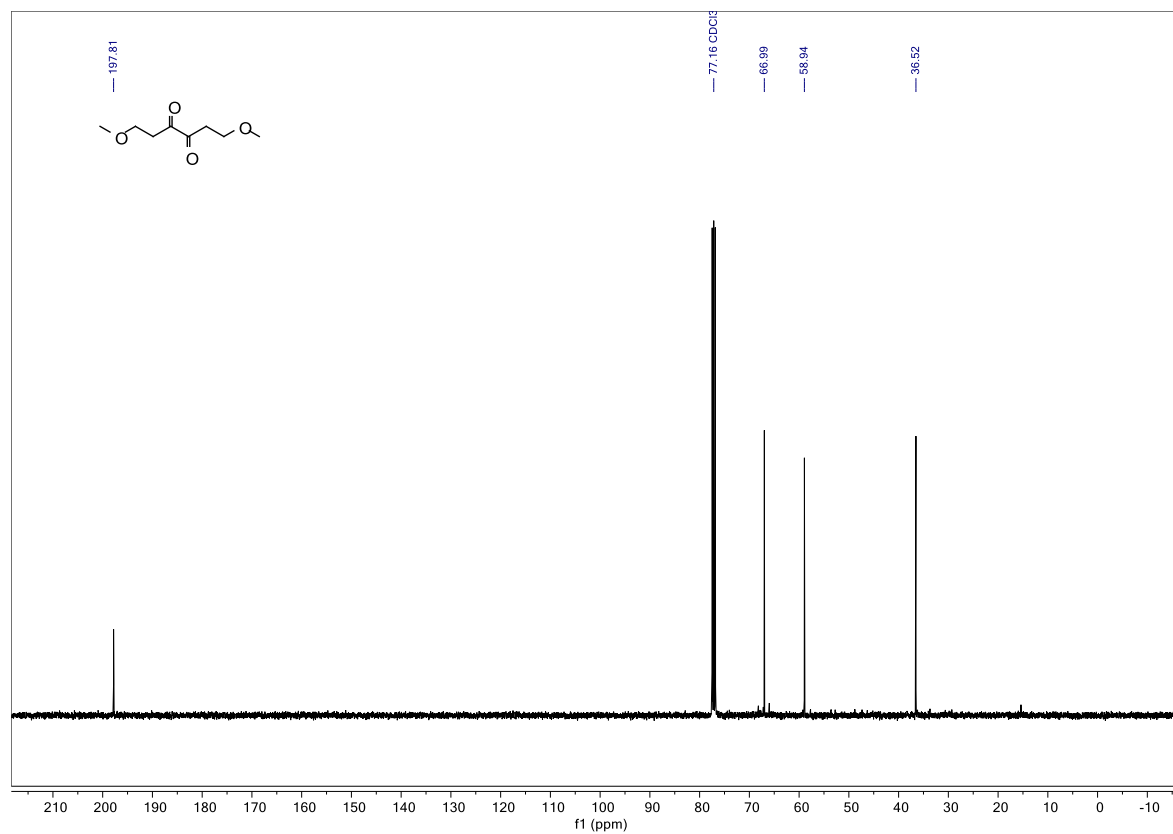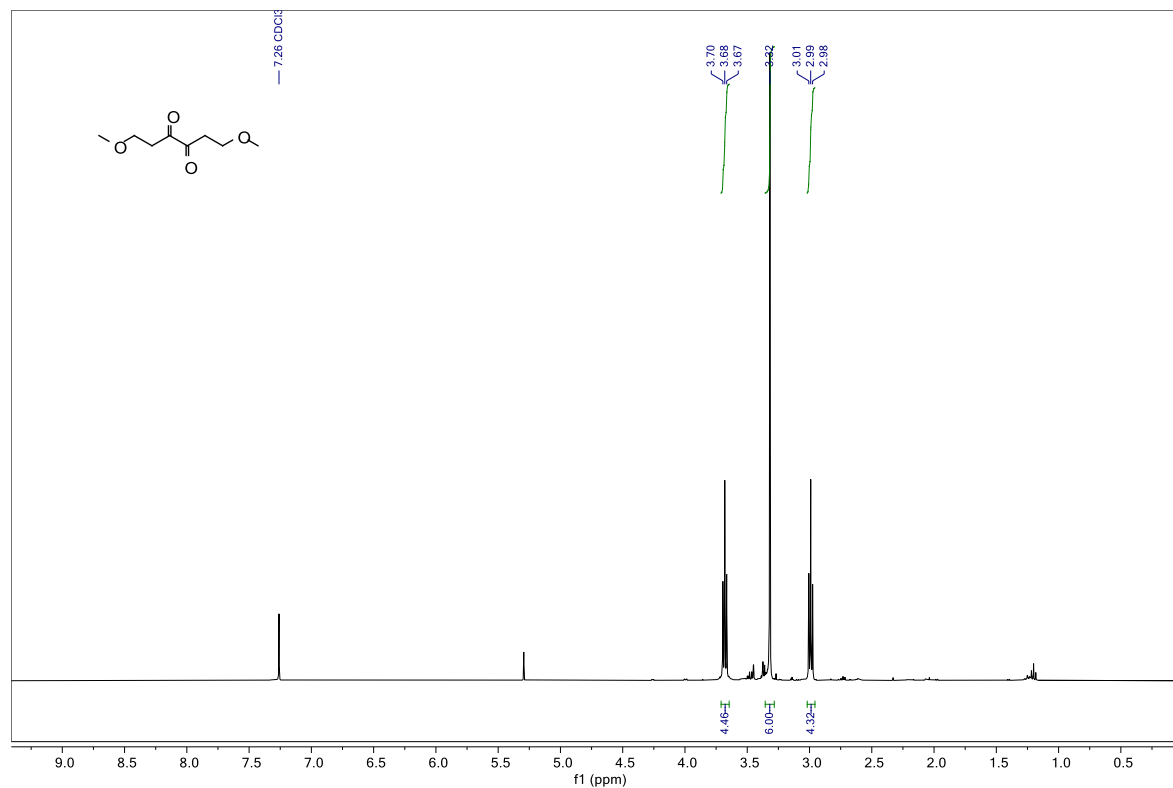
Compound 6.15 ^{1}H and ^{13}C NMR
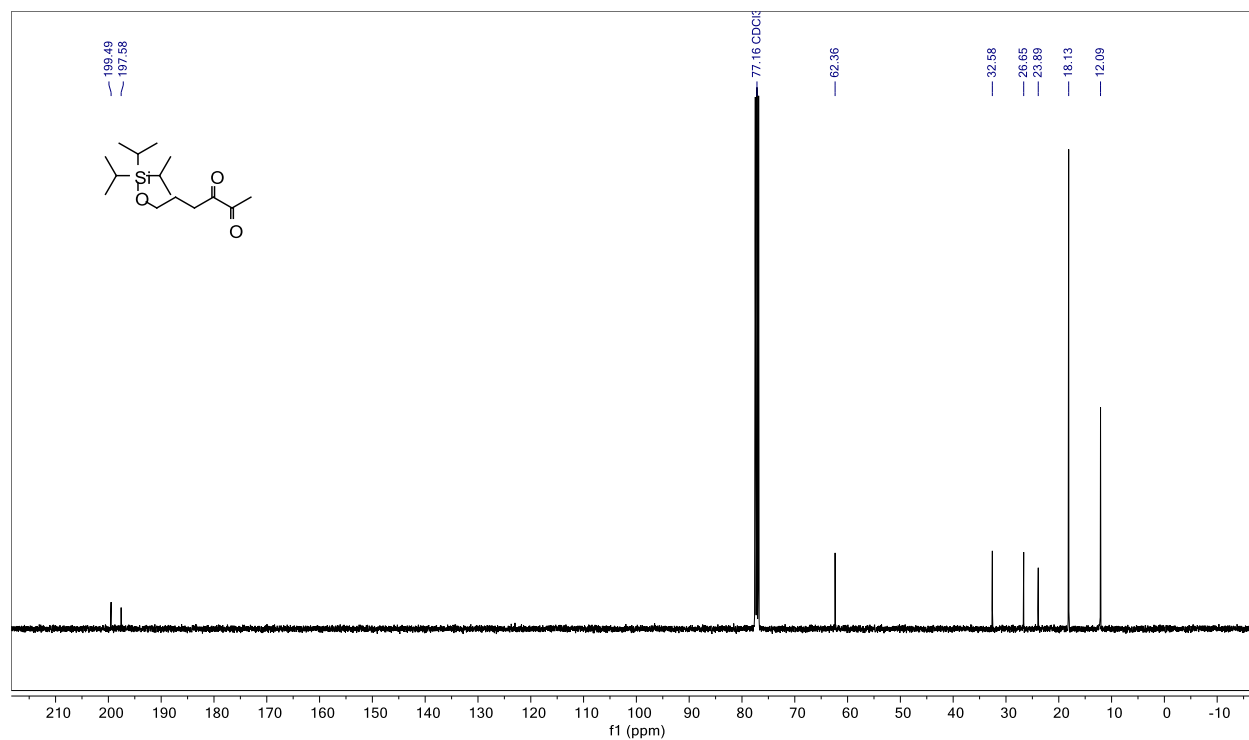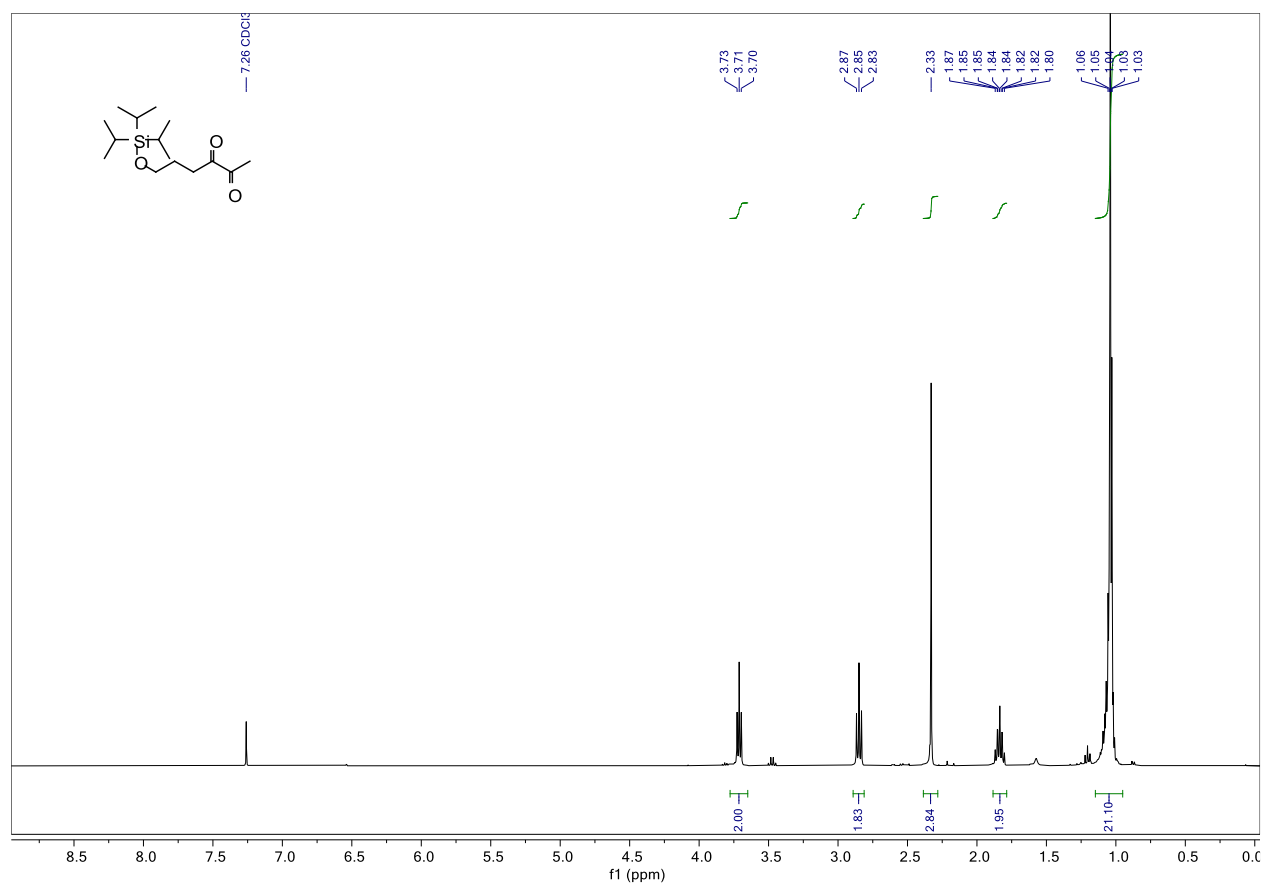
Compound 6.16 ^{1}H and ^{13}C NMR
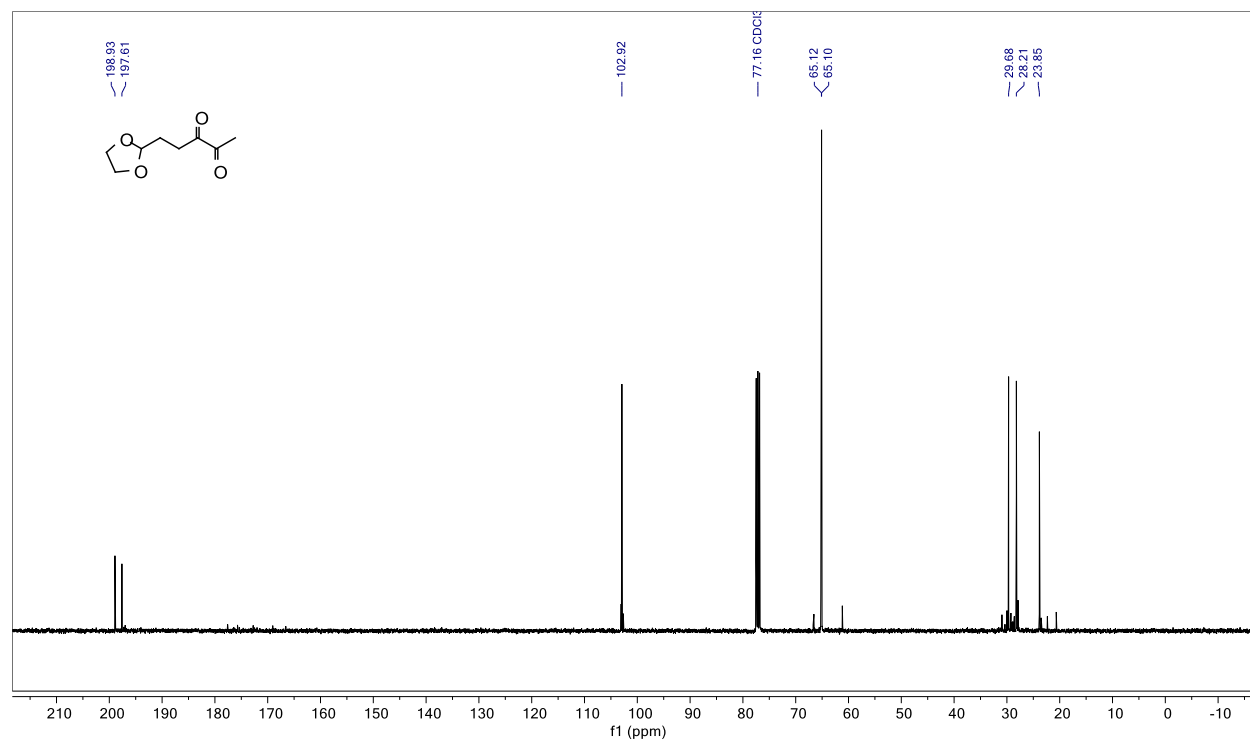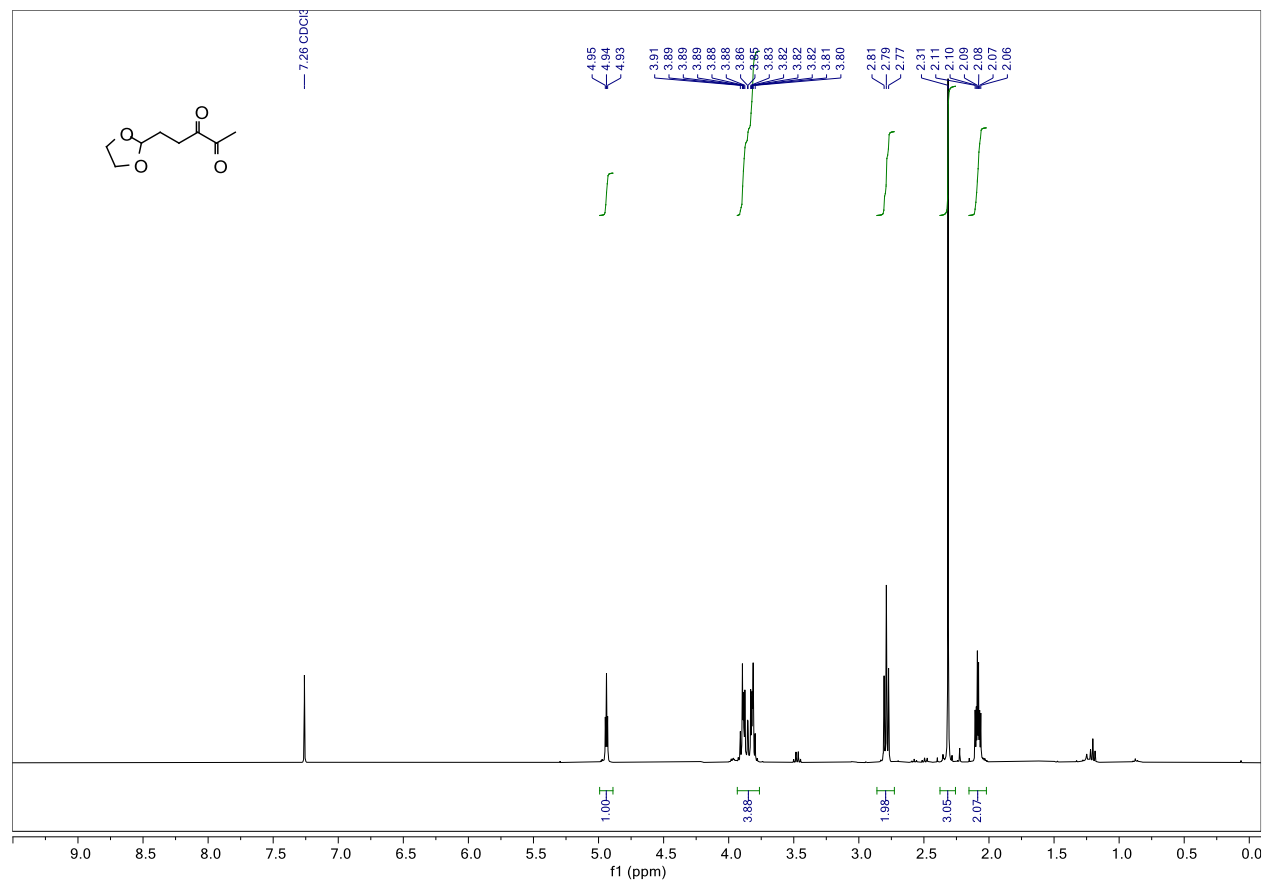
Compound 6.17 ^{1}H and ^{13}C NMR
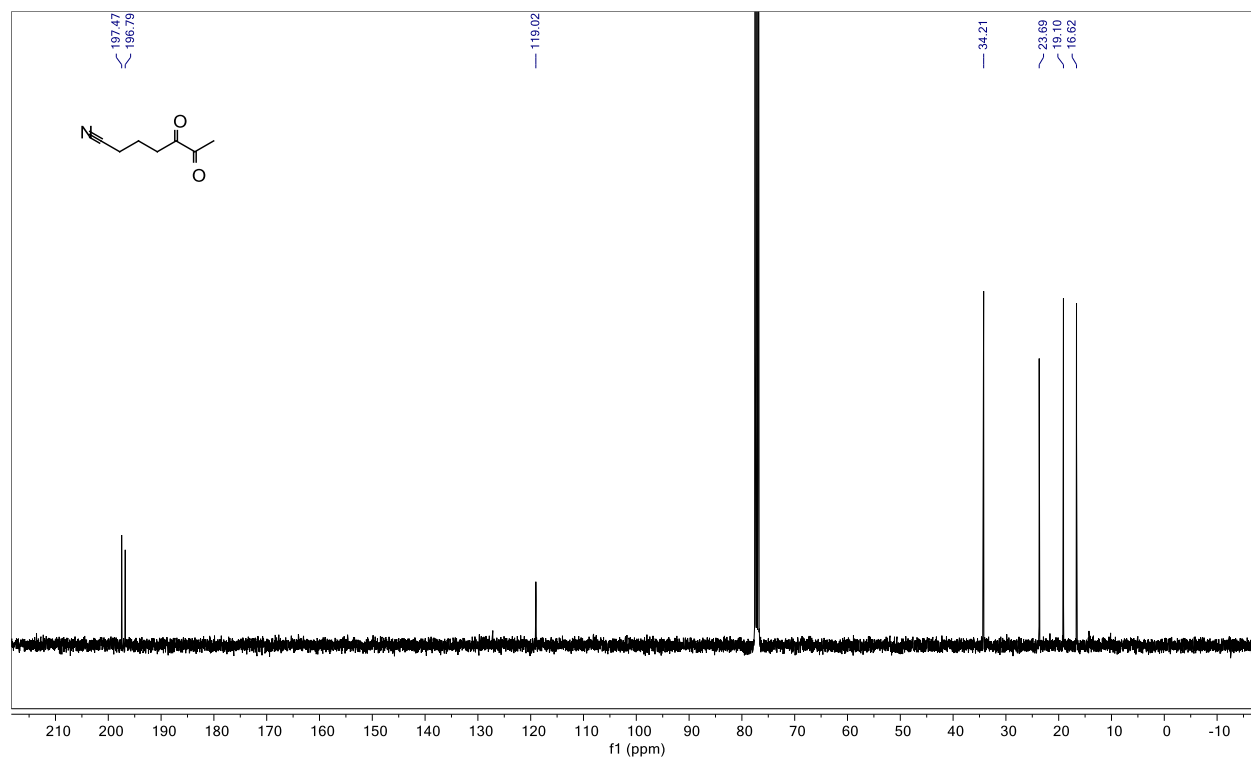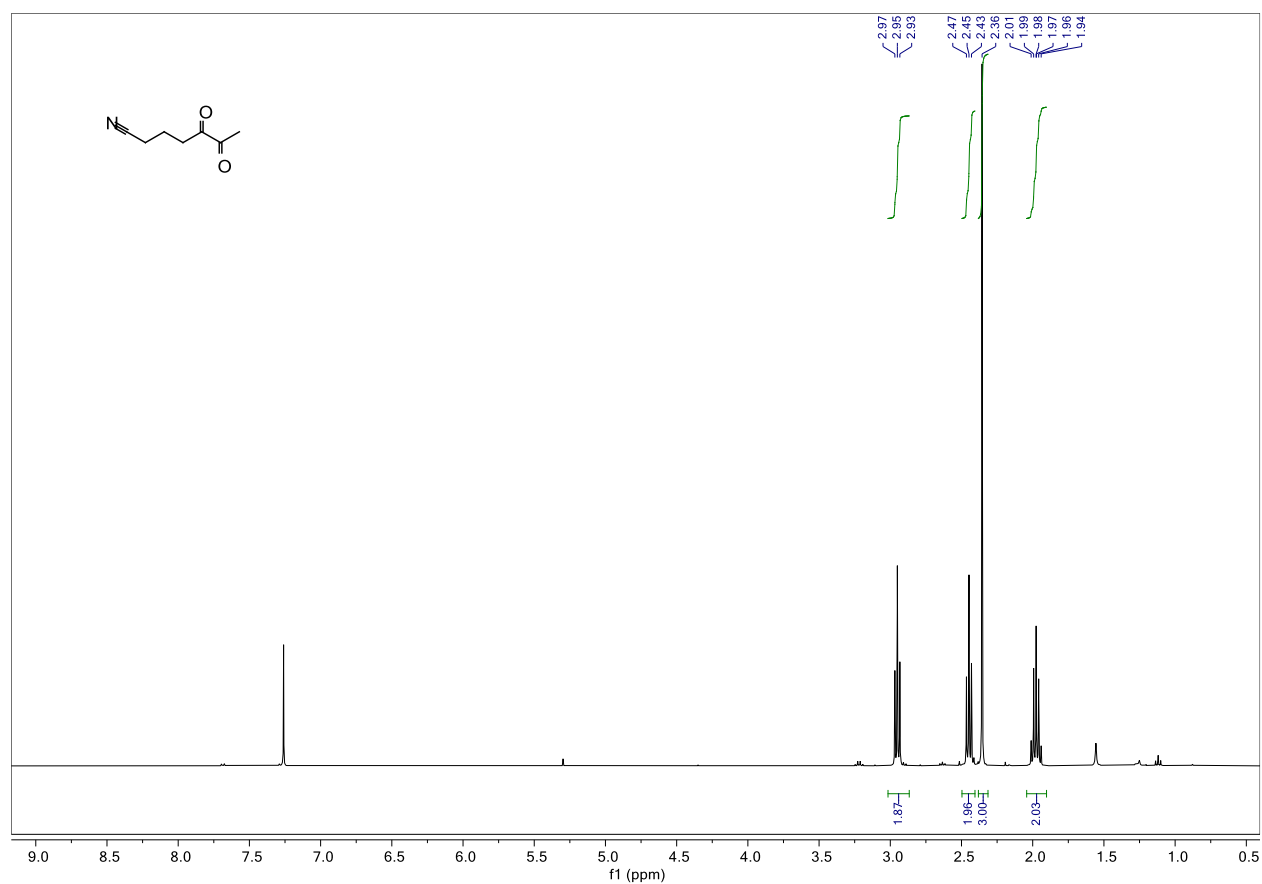
Compound 6.18 ^{1}H and ^{13}C NMR

Compound 6.19 ^{1}H and ^{13}C NMR
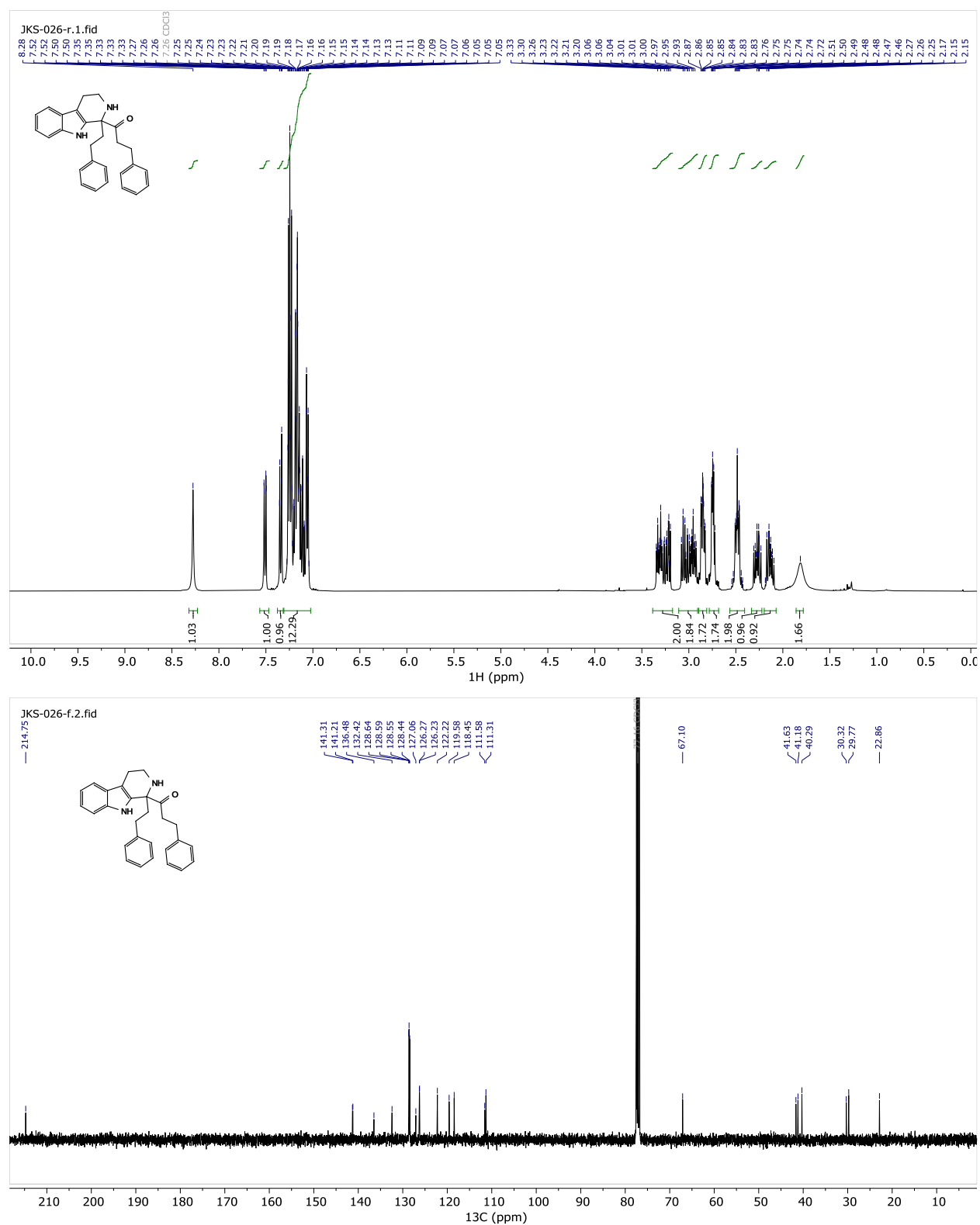
Compound 6.20 ^{1}H and ^{13}C NMR
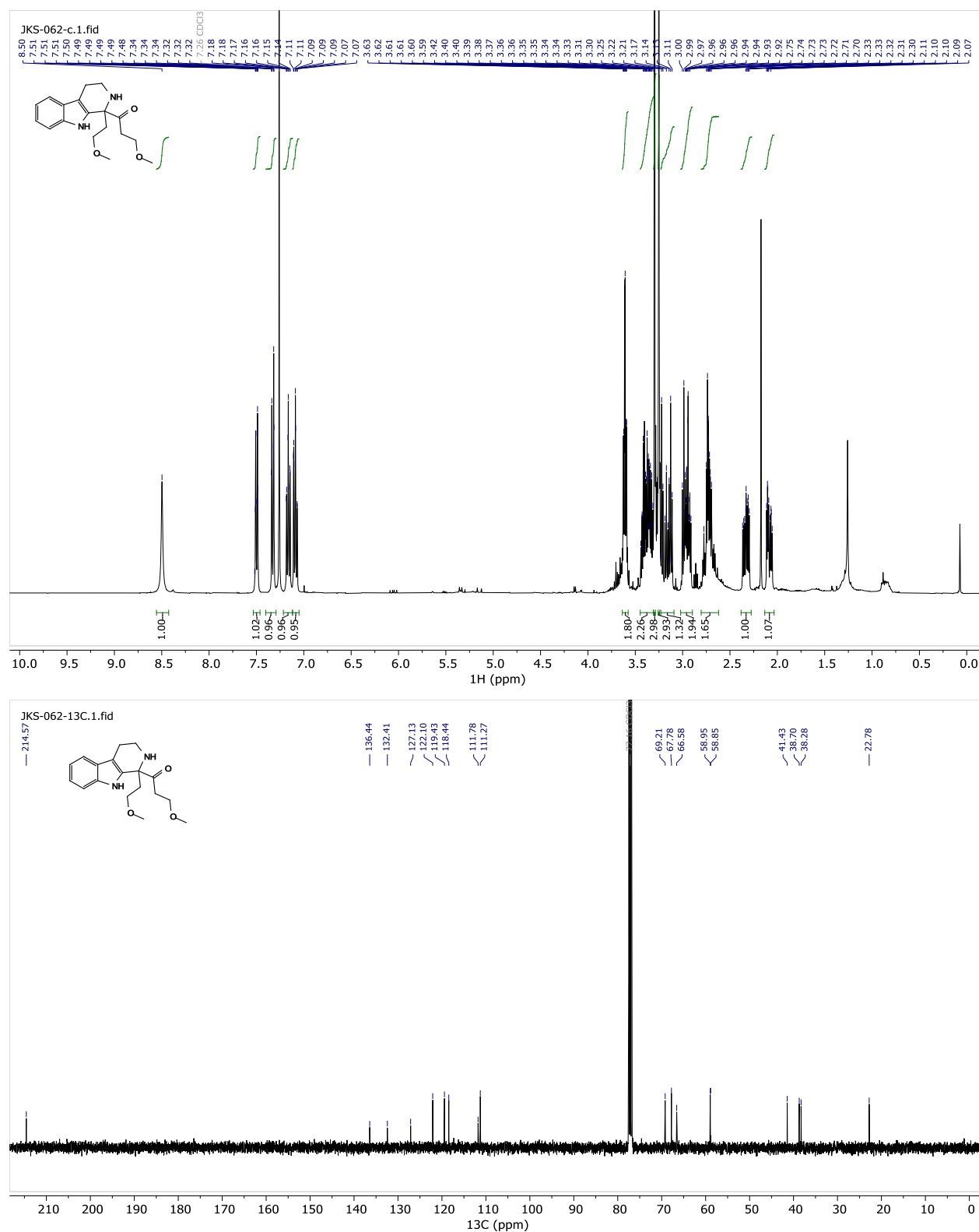
Compound 6.21 ¹H NMR
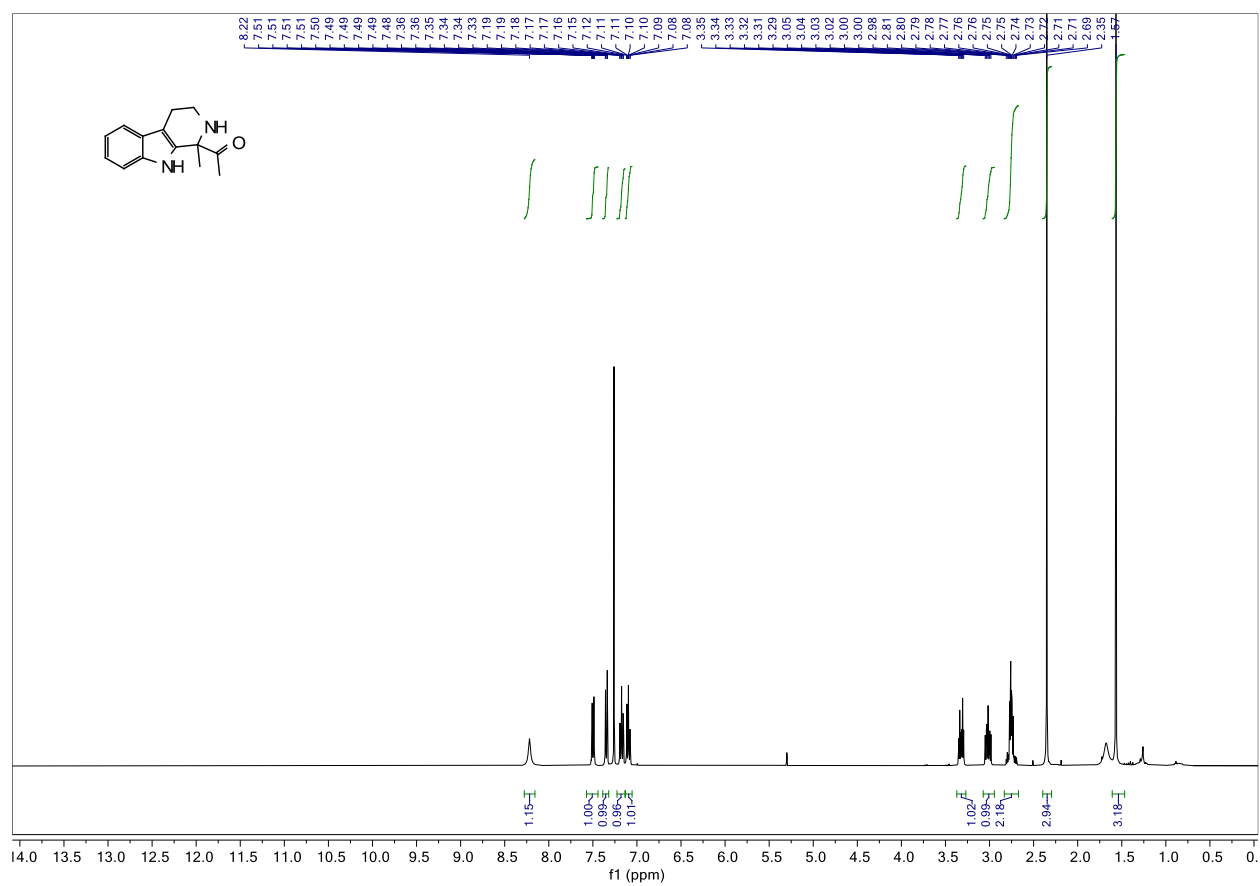
Compound 6.22 ^{1}H and ^{13}C NMR
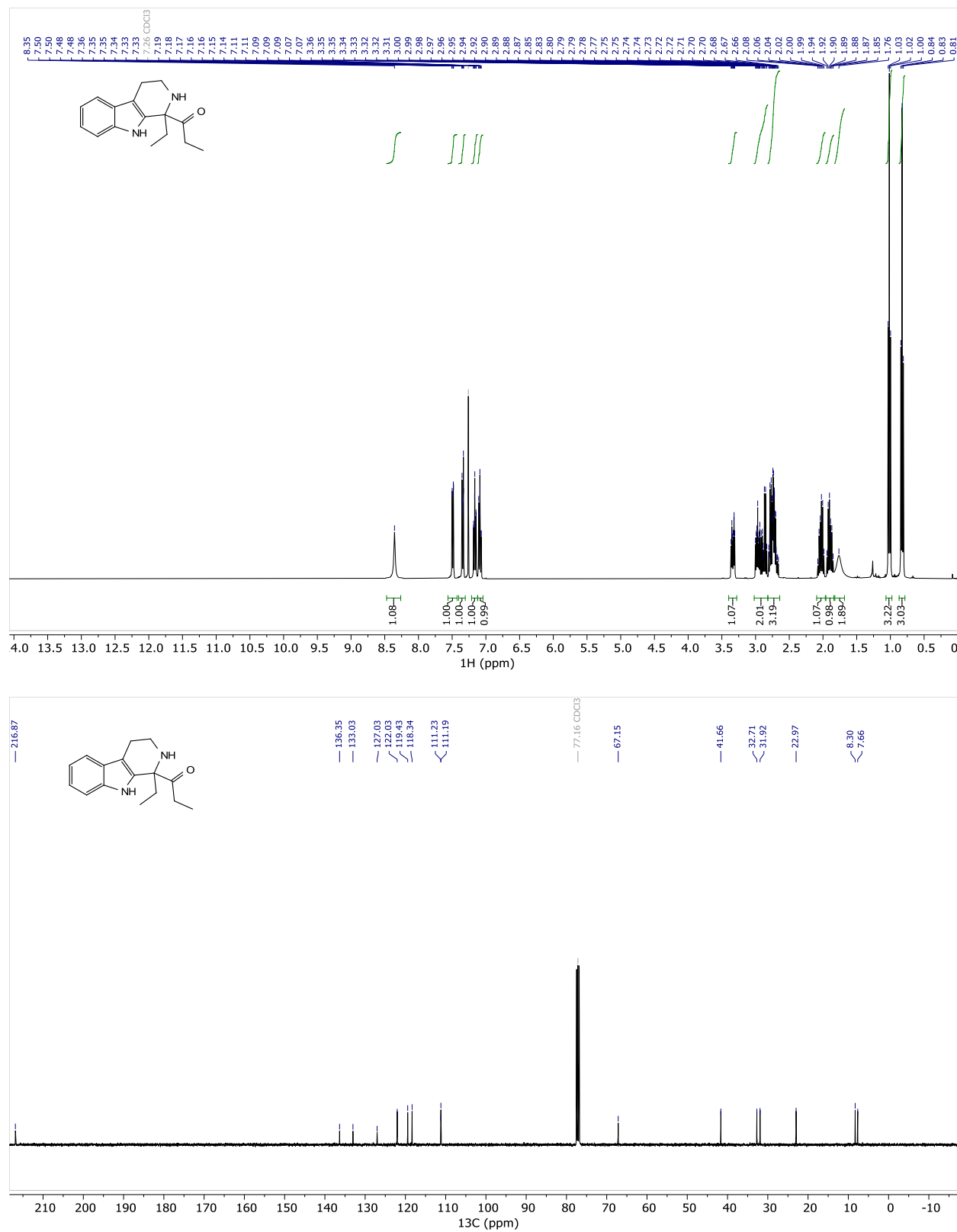
Compound 6.23 ¹H and ¹³C NMR
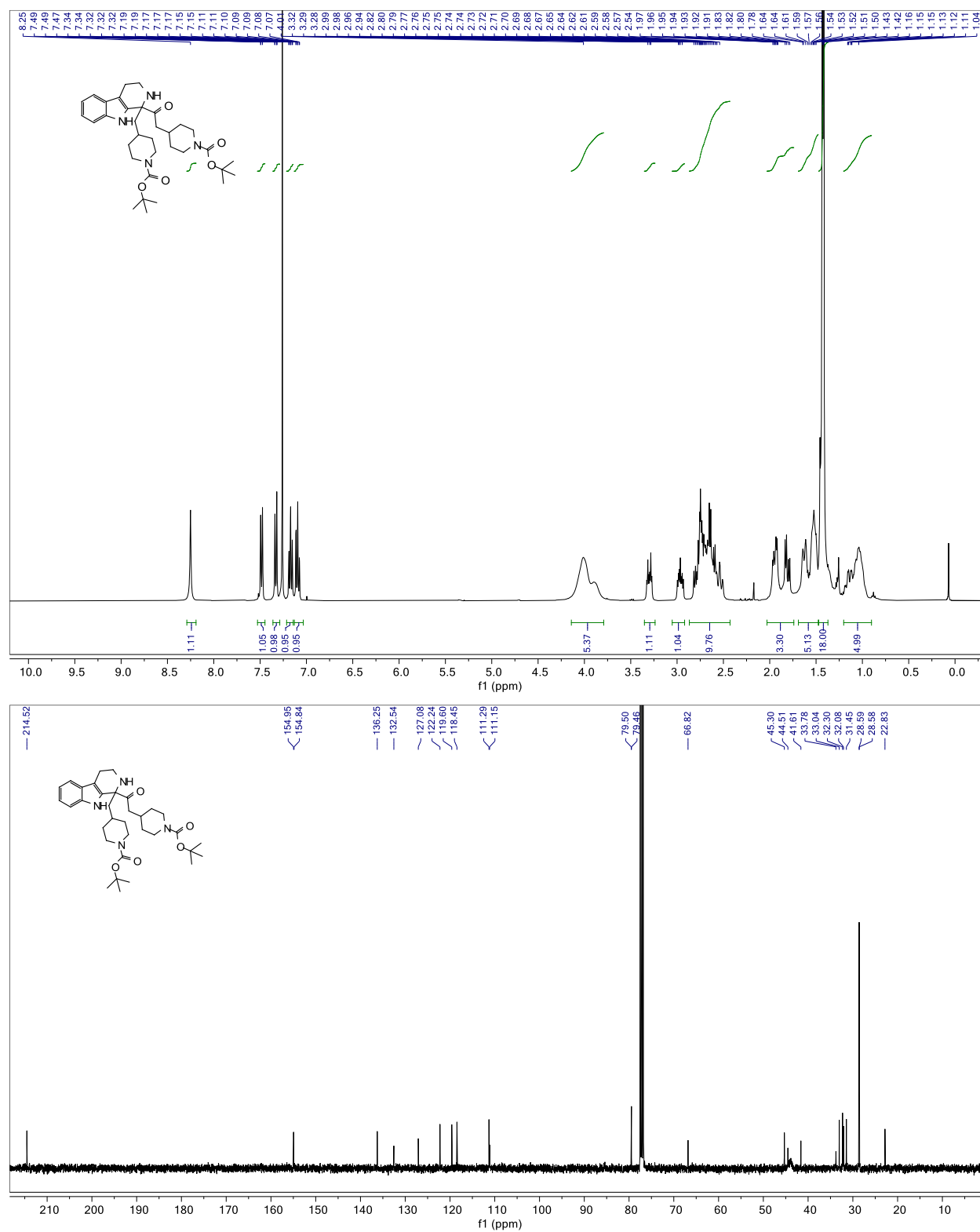
Compound 6.24 ^{1}H and ^{13}C NMR

Compound 6.25 ^{1}H and ^{13}C NMR
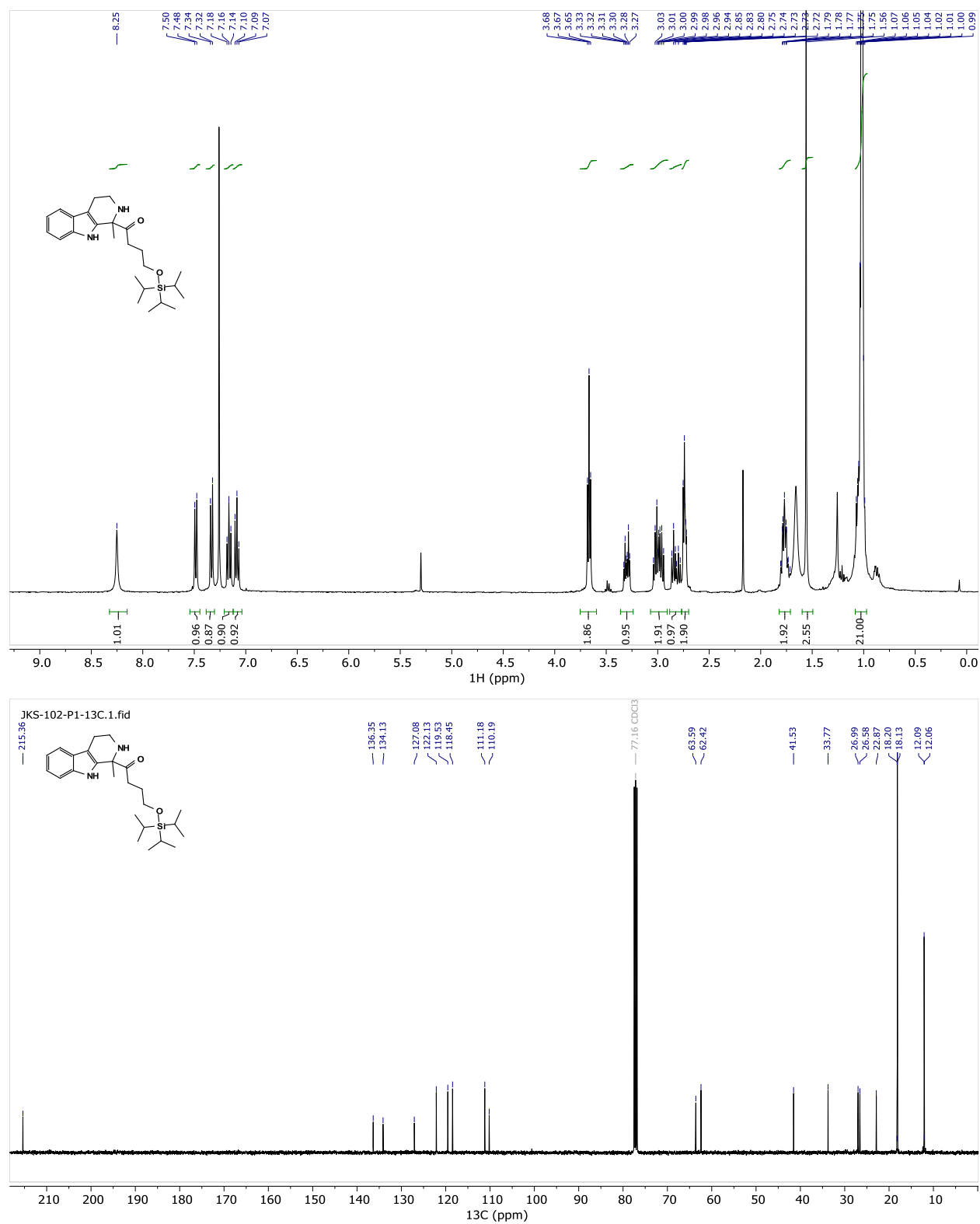
SFC Chromatograms

SFC Chromatogram for 6.26 (enantioselective) and 6.21 (racemic)

SFC: IG column, 2 mL/min, 19% (1% NEt$_3$ / MeOH) in supercritical CO2, λ = 270 nm.

Instrument:	SFC		**Injection date:**	2022-07-24 13:07:49+02:00
Inj. volume:	2.000 µL		**Location:**	2
Acq. method:	Run IG 19% MeOH 5 minutes.amx		**Type:**	Sample
Processing method:	*Manual Integration.pmx		**Column:**	
Manually modified:	Manual Integration			
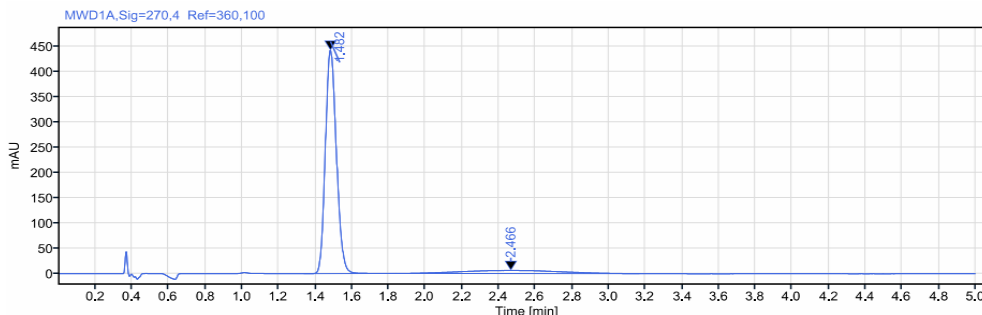
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.482	MM m	0.50	1861.19	443.72	88.76	
2.466	MM m	1.33	235.63	6.14	11.24	
		Sum	**2096.82**			

Instrument:	SFC		**Injection date:**	2022-02-24 18:29:09+01:00
Inj. volume:	2.000 µL		**Location:**	3
Acq. method:	Run IG 19% MeOH 5 minutes.amx		**Type:**	Sample
Processing method:	*Manual Integration.pmx		**Column:**	IG
Manually modified:	Manual Integration			

Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.827	MM m	0.38	767.41	118.34	50.53	
3.070	MM m	1.43	751.34	18.59	49.47	
		Sum	**1518.75**			

SFC Chromatogram for 6.27 (enantioselective) and 6.22 (racemic)

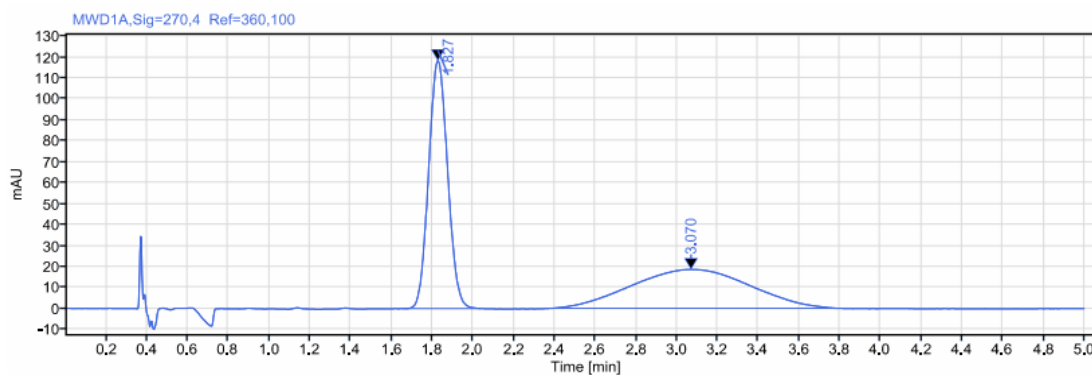SFC: IB column, 2 mL/min, 12% (1% NEt_3 / MeOH) in supercritical CO2, $\lambda = 270$ nm.

Sample name:	JKS-059		Operator:	SYSTEM
Instrument:	SFC		Injection date:	2022-07-24 13:23:27+02:00
Inj. volume:	2.000 µL		Location:	4
Acq. method:	Run IB 12% MeOH 5 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			
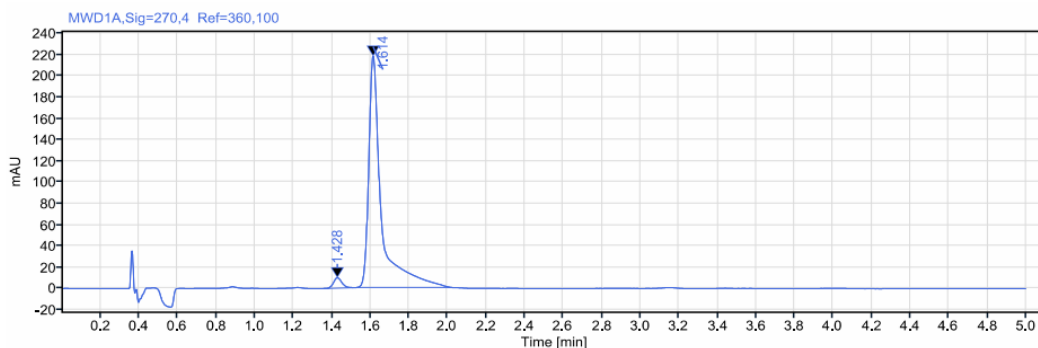
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.428	MM m	0.16	28.74	9.66	2.81	
1.614	MM m	0.49	993.88	217.50	97.19	
		Sum	1022.62			

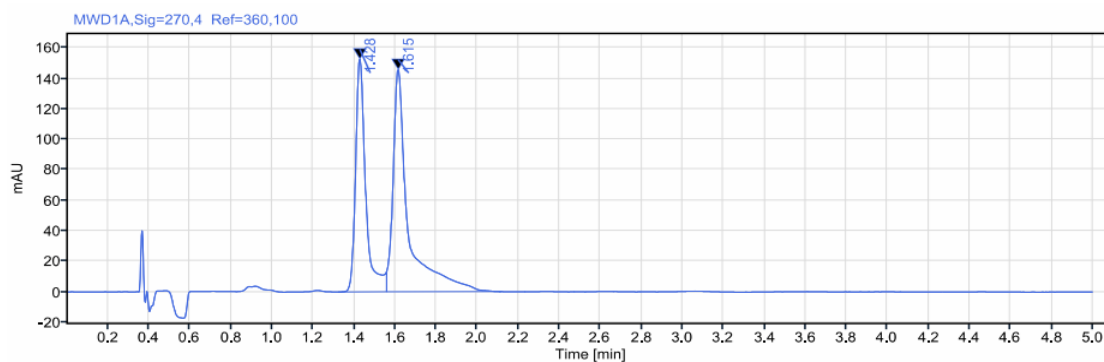| Instrument:	SFC		Injection date:	2022-07-24 13:15:45+02:00
Inj. volume:	2.000 µL		Location:	3
Acq. method:	Run IB 12% MeOH 5 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			

Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.428	MM m	0.24	526.64	152.72	41.13	
1.615	MM m	0.56	753.82	145.72	58.87	
		Sum	1280.46			

SFC Chromatogram for 6.28 RT (enantioselective) and 6.18 (racemic)

SFC: IJ column, 2 mL/min, 10% MeOH in supercritical CO_2, $\lambda = 214$ nm.

Instrument:	SFC		Injection date:	2022-05-30 16:08:01+02:00
Inj. volume:	2.000 µL		Location:	3
Acq. method:	Run IJ 10% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	IJ
Manually modified:	Manual Integration			

Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.206	MM m	0.34	69.16	6.16	1.94	
1.507	MM m	1.31	3500.38	411.35	98.06	
		Sum	3569.54			

Instrument:	SFC		Injection date:	2022-05-30 15:33:36+02:00
Inj. volume:	2.000 µL		Location:	2
Acq. method:	Run IJ 10% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	IJ
Manually modified:	Manual Integration			

Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.059	MM m	0.49	1635.10	204.09	49.31	
1.554	MM m	1.44	1681.14	126.54	50.69	
		Sum	3316.23			

SFC Chromatogram for 6.28 40 ºC (enantioselective)

SFC: IJ column, 2 mL/min, 10% MeOH in supercritical CO2, λ = 214 nm.

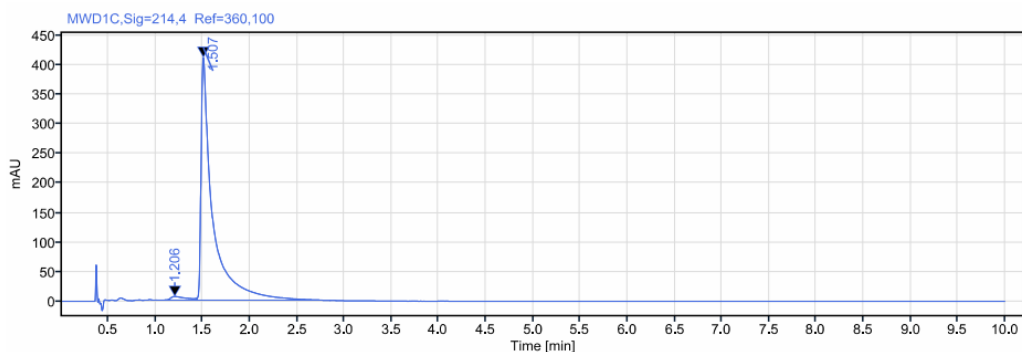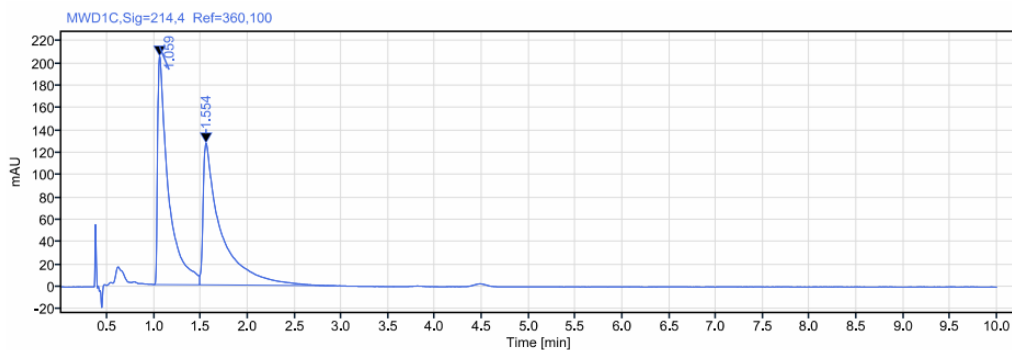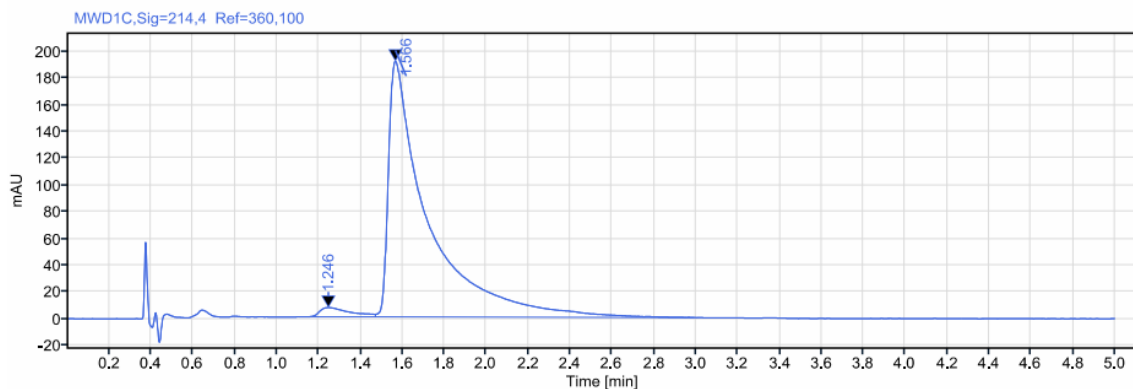| Instrument:	SFC	Injection date:	2022-07-24 12:11:22+02:00
Inj. volume:	2.000 µL	Location:	5
Acq. method:	Run IJ 10% MeOH 5 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	
Manually modified:	Manual Integration		

Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
1.246	MM m	0.36	70.44	6.99	2.73	
1.566	MM m	1.52	2513.26	191.68	97.27	
		Sum	2583.71			

SFC Chromatogram for 6.31 40 ºC (enantioselective) and 6.23 (racemic)

SFC: IG column, 2 mL/min, 15% MeOH in supercritical CO2, λ = 214 nm.

Instrument:	SFC		Injection date:	2022-07-24 11:36:26+02:00
Inj. volume:	2.000 µL		Location:	2
Acq. method:	Run IG 15% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			
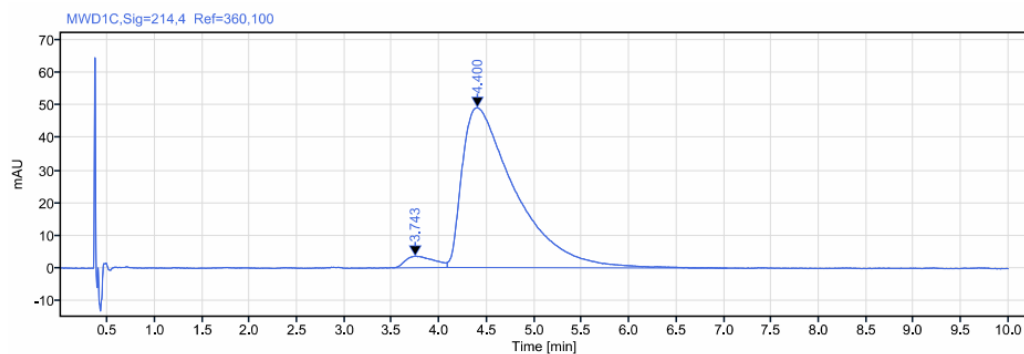
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
3.743	MM m	0.58	74.90	3.56	3.75	
4.400	MM m	2.95	1920.22	49.15	96.25	
		Sum	1995.12			

Instrument:	SFC		Injection date:	2022-07-24 11:23:45+02:00
Inj. volume:	4.000 µL		Location:	1
Acq. method:	Run IG 15% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			
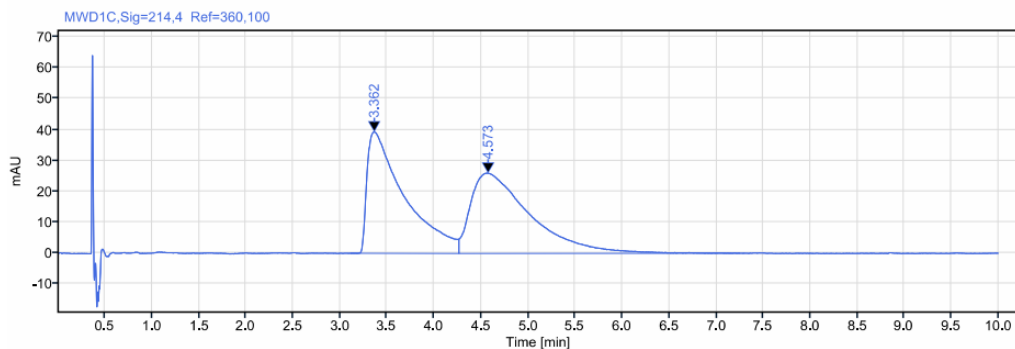
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
3.362	MM m	1.15	1089.52	39.38	48.83	
4.573	MM m	2.99	1141.62	26.05	51.17	
		Sum	2231.14			

SFC Chromatogram for 6.29 (enantioselective) and 6.19 (racemic)

SFC: IG column, 2 mL/min, 15% MeOH in supercritical CO2, λ = 214 nm.

| Instrument:	SFC	Injection date:	2022-05-10 15:02:06+02:00
Inj. volume:	2.000 µL	Location:	3
Acq. method:	Fenggang - Run IG 15% MeOH 15 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	IG
Manually modified:	Manual Integration		
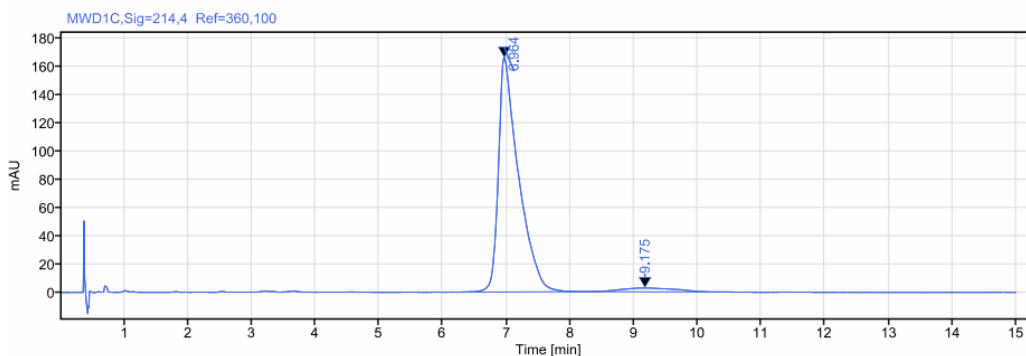
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
6.964	MM m	1.98	3666.67	166.50	95.80	
9.175	MM m	2.01	160.57	2.72	4.20	
		Sum	**3827.23**			

Instrument:	SFC	Injection date:	2022-05-10 14:08:59+02:00
Inj. volume:	2.000 µL	Location:	1
Acq. method:	Fenggang - Run IG 15% MeOH 15 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	IG
Manually modified:	Manual Integration		
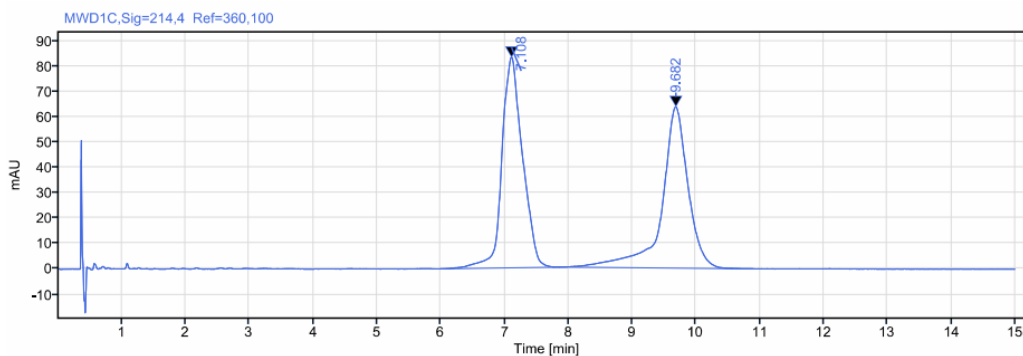
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
7.108	MM m	1.96	1874.70	83.39	50.64	
9.682	MM m	2.97	1827.23	63.88	49.36	
		Sum	**3701.93**			

SFC Chromatogram for 6.30 (enantioselective) and 6.20 (racemic)

SFC: IG column, 2 mL/min, 15% MeOH in supercritical CO2, λ = 214 nm

| Instrument: | SFC | | Injection date: | 2022-05-17 19:22:27+02:00 |
Inj. volume:	2.000 µL		Location:	1
Acq. method:	Run IG 15% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	IG
Manually modified:	Manual Integration			

MWD1C,Sig=214,4 Ref=360,100
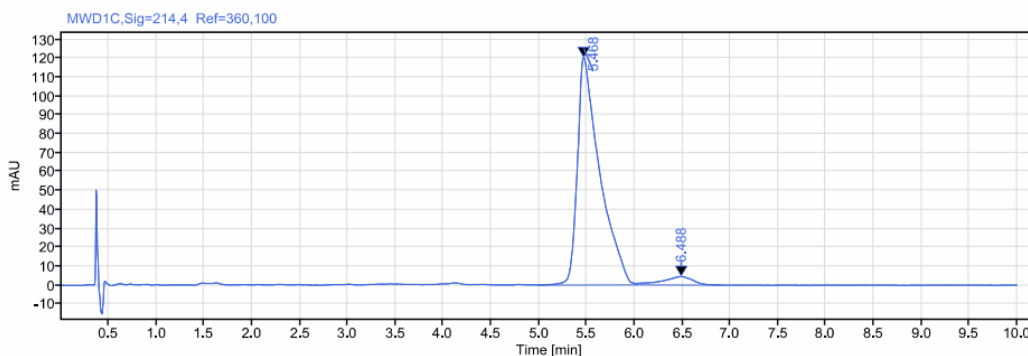
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
5.468	MM m	1.03	1956.71	120.20	95.22	
6.488	MM m	0.89	98.17	4.45	4.78	
		Sum	**2054.88**			

Instrument:	SFC		Injection date:	2022-05-16 10:30:40+02:00
Inj. volume:	2.000 µL		Location:	1
Acq. method:	Run IG 15% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	IG
Manually modified:	Manual Integration			

MWD1C,Sig=214,4 Ref=360,100
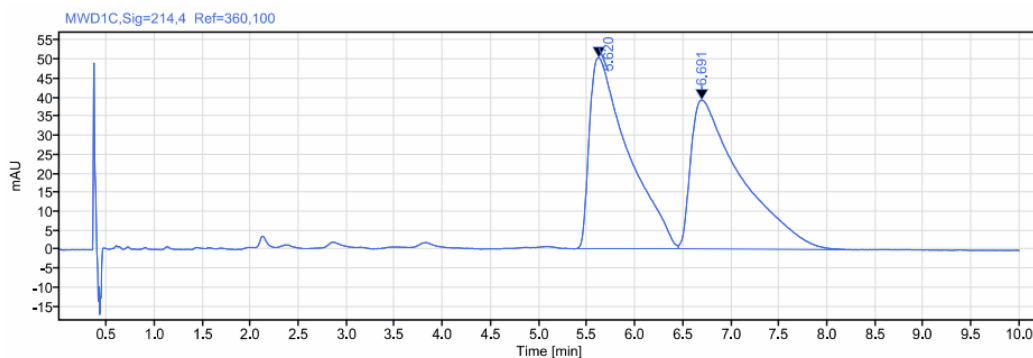
Signal: MWD1C,Sig=214,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
5.620	MM m	1.07	1410.92	50.32	49.75	
6.691	MM m	1.87	1424.87	39.20	50.25	
		Sum	**2835.79**			

SFC Chromatogram for 6.32 (enantioselective) and 6.24 (racemic)

SFC: IB column, 2 mL/min, 6% (1% NEt₃ / MeOH) in supercritical CO2, λ = 270 nm.

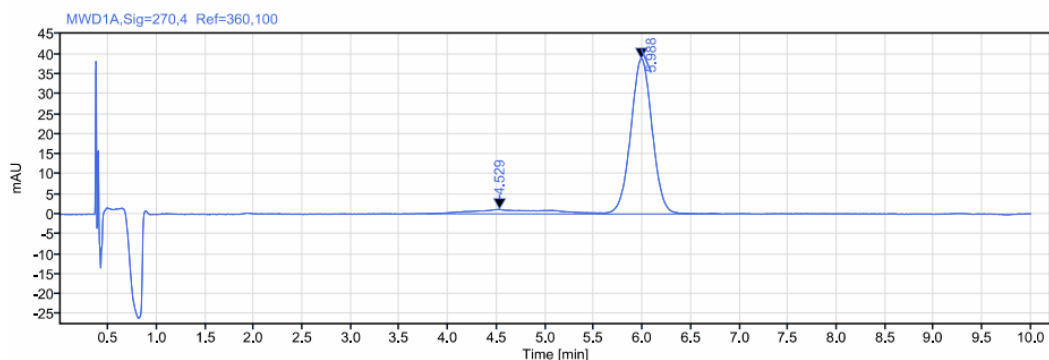| Instrument:	SFC	Injection date:	2022-07-19 18:44:08+02:00
Inj. volume:	2.000 µL	Location:	1
Acq. method:	Run IB 6% MeOH 10 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	
Manually modified:	Manual Integration		

Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
4.529	MM m	1.82	64.70	1.15	9.78	
5.988	MM m	1.20	596.53	38.76	90.22	
		Sum	661.22			

Instrument:	SFC	Injection date:	2022-07-19 18:16:58+02:00
Inj. volume:	2.000 µL	Location:	1
Acq. method:	Run IB 6% MeOH 10 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	
Manually modified:	Manual Integration		
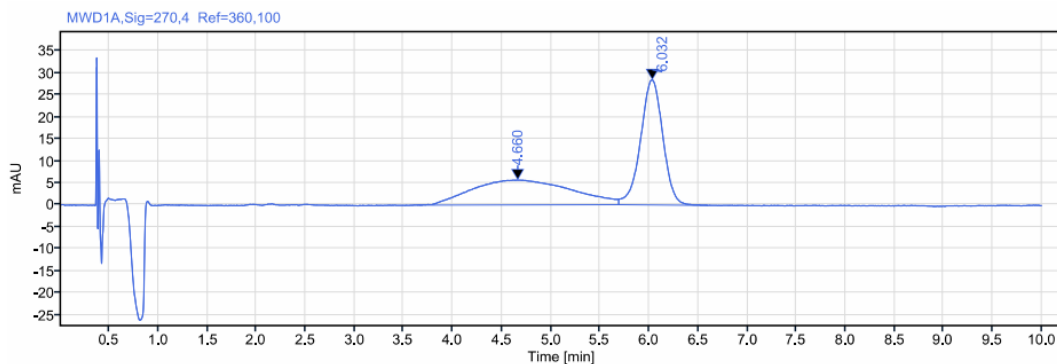
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
4.660	MM m	2.05	400.06	5.67	47.02	
6.032	MM m	0.91	450.79	28.48	52.98	
		Sum	850.85			

SFC Chromatogram for 6.33 (enantioselective) and 6.25 (racemic)

SFC: IB column, 2 mL/min, 12% (1% NEt$_3$ / MeOH) in supercritical CO2, λ = 270 nm.

| Instrument: | SFC | | Injection date: | 2022-07-19 17:05:36+02:00 |
Inj. volume:	2.000 µL		Location:	2
Acq. method:	Run IB 12% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			

MWD1A,Sig=270,4 Ref=360,100
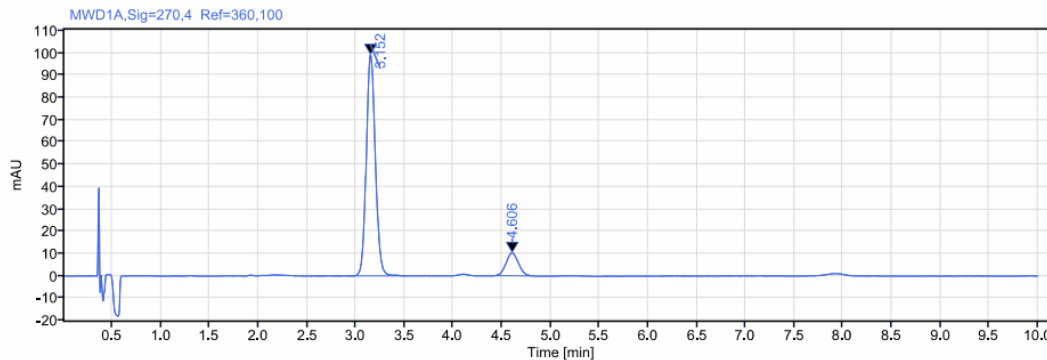
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
3.152	MM m	0.61	641.81	99.14	87.69	
4.606	MM m	0.38	90.11	10.23	12.31	
		Sum	731.92			

Instrument:	SFC		Injection date:	2022-07-19 16:49:14+02:00
Inj. volume:	2.000 µL		Location:	2
Acq. method:	Run IB 12% MeOH 10 minutes.amx		Type:	Sample
Processing method:	*Manual Integration.pmx		Column:	
Manually modified:	Manual Integration			

MWD1A,Sig=270,4 Ref=360,100
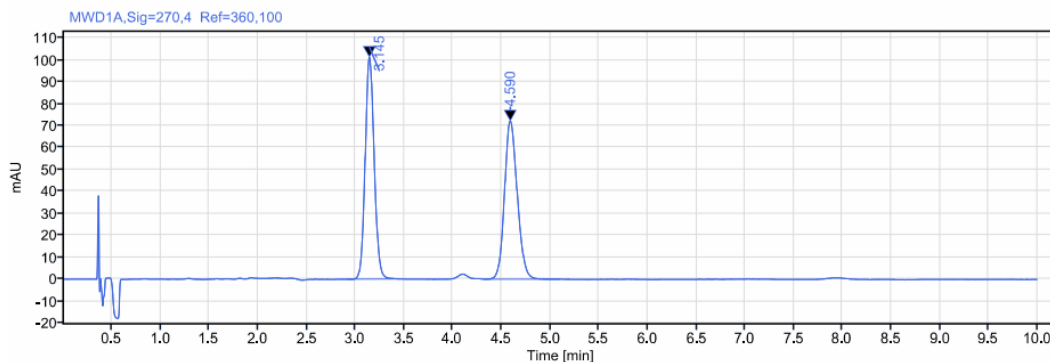
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
3.145	MM m	0.54	652.33	101.27	49.83	
4.590	MM m	0.78	656.73	72.14	50.17	
		Sum	1309.06			

SFC Chromatogram for Table 6.1 Racemic and Entry 23

SFC: IB column, 2 mL/min, 12% MeOH (1% NEt$_3$ / MeOH) in supercritical CO2, λ = 214 nm.

| Instrument:	SFC	Injection date:	2022-02-24 17:56:18+01:00
Inj. volume:	2.000 µL	Location:	1
Acq. method:	Run IB 12% MeOH 5 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	IB
Manually modified:	Manual Integration		
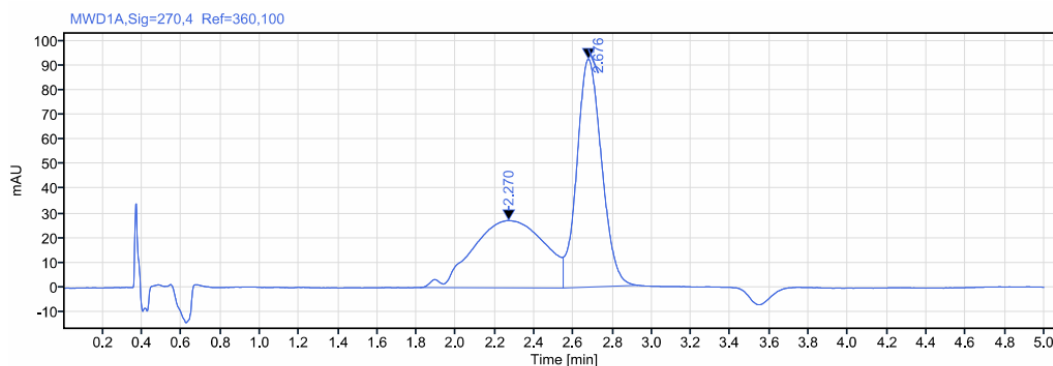
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
2.270	MM m	0.73	696.51	27.32	46.09	
2.676	MM m	0.38	814.63	92.51	53.91	
		Sum	1511.14			

Instrument:	SFC	Injection date:	2022-03-04 16:03:10+01:00
Inj. volume:	2.000 µL	Location:	D1F-A3
Acq. method:	Run IB 12% MeOH 5 minutes.amx	Type:	Sample
Processing method:	*Manual Integration.pmx	Column:	IB
Manually modified:	Manual Integration		
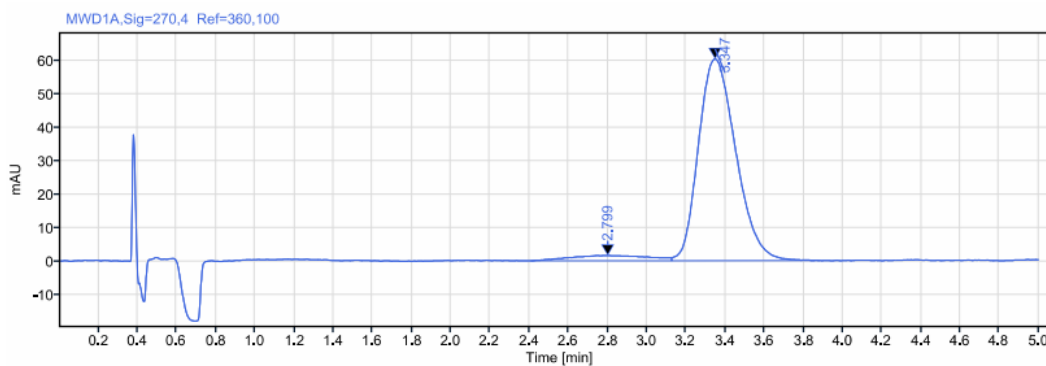
Signal: MWD1A,Sig=270,4 Ref=360,100

RT [min]	Type	Width [min]	Area	Height	Area%	Name
2.799	MM m	0.72	44.00	1.57	5.20	
3.347	MM m	0.75	802.15	60.46	94.80	
		Sum	846.16			