Squaric Ester Applications as Novel Lysine Electrophiles in Molecular Probe Design

by

Jordan Sun Ho

B.S. Chemistry
University of California—Berkeley (2014)

M.S. Chemistry
University of Wisconsin—Madison (2017)

Submitted to the Department of Chemistry
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Signature of Author _____
Department of Chemistry
August 24, 2020

Certified by _____
Laura L. Kiessling
Novartis Professor of Chemistry
Thesis Supervisor

Accepted by _____
Adam Willard
Associate Professor
Graduate Officer

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Matthew D. Shoulders
Whitehead Career Development Associate Professor
Thesis Chair

Laura L. Kiessling
Novartis Professor Chemistry
Thesis Supervisor

Ronald T. Raines
Firmenich Professor of Chemistry

Squaric Esters Applications as Novel Lysine Electrophiles in Molecular Probe Design

by

Jordan S. Ho

Submitted to the Department of Chemistry
On August 24, 2020 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Abstract

Small molecule probes for biology have been instrumental in uncovering enzyme mechanisms and developing therapeutics. Covalent probes are valuable because they can irreversibly tag proteins of interest for analysis. Selective covalent proteins for lysine residues are especially valuable because they allow for fine control over biological systems. Many binding sites contain lysine residues, but current amine-reactive electrophiles in biology are largely unselective. Moreover, at almost 6% lysine is more prevalent in the proteome compared to other nucleophilic residues such as cysteine. The ability to selectively target specific lysines would open new avenues for analyzing biomolecular interactions. Current efforts have yielded compounds with high reactivity and low stability, severely limiting their utility. Squaric esters are small chemical compounds with multiple amine reactive sites that have been used extensively as a linker in organic synthesis. Additionally, squaric esters are mild electrophiles when compared to other amine reactive electrophiles used in organic chemistry. This attenuated reactivity, coupled with their high selectivity for amines suggests that substituted squaric esters may serve as novel biological probes. To this end, we characterized the reactivity and the kinetics of squaric ester reactions. We also applied squaric esters in different biological contexts to evaluate their utilities as novel lysine-reactive electrophiles. We show that squaric esters react orders of magnitude slower than other amine-reactive electrophiles commonly used in biology. We then applied squaric esters practically in the design of novel galactofuranosyltransferase 2 inhibitors, an enzyme responsible for the biosynthesis of the galactan. The galactan is a polysaccharide chain of galactofuranose residues, and it is essential for the cell wall of many bacteria, including pathogens such as *Mycobacterium tuberculosis*. We generated substituted squaric ester inhibitors that bind to galactofuranosyltransferase 2 with specificity that provided insight into a potential allosteric binding site of galactofuranosyltransferase 2. Finally using fragment-based ligand design, in efforts to expand the lysine target space of small molecular probes, we constructed a squaric ester fragment library to screen for novel ligandable lysines across entire proteomes.

Thesis Supervisor: Laura L. Kiessling
Title: Novartis Professor of Chemistry

Acknowledgements

I would not have been able to accomplish what I have without the encouragement and support of so many people. I am extremely grateful to my parents who have never stopped believing in me, and to my brother who I continually look up to as a role model. They have been instrumental in shaping who I am today, and I cannot fully express just how thankful I am for their continued love and support.

I owe a great deal to the many wonderful colleagues I was able to work with over these past six years. I consider myself blessed to be able to have worked among a community so incredibly smart as well as kind. In particular, I would like to personally thank Dr. Alex Justen, Dr. Austin Kruger, Dr. Caitlin McMahon, Dr. Cassie Jarvis, Dr. Phil Calabretta, Dr. Rob Brown, Dr. Sayaka Masuko, Dr. Spencer Brucks, Victoria Marando, and Katherine Taylor for giving me their time, advice, and friendship.

In addition to my lab mates, there are many friends I must thank deeply for helping keep my sanity as well as encouraging me toward the finish line. I have had wonderful support from those in my cohort and feel extremely fortunate to be have been able to share these experiences with, including Manar Alherech, Naomi Biok, and Kelly Suralik. Thank you to Sarah Rominger for your friendship, empathy, and constant letters of encouragement during particularly challenging times. I would also like to especially thank Nick Sanchez. Together, we were the only ones who left our tiny town to lands beyond Wisconsin and we even both pursued graduate degrees. While our similar walks of life took us to opposite coasts, your constant friendship, support, words of wisdom, and shared commiserations these last 20-some years has always helped keep me grounded.

I want to thank the University of Wisconsin-Madison and Massachusetts Institute of Technology departments of chemistry for financial support and the amazing opportunity to be in the presence of many passionate and talented scientists, and to be able to be part of cutting edge research.

I must also thank my thesis committee. I thank Professor Ron Raines, for helpful advice and feedback. I thank Matt Shoulders for being my thesis chair and for all the mentorship he has provided since moving to MIT. Lastly, I thank Laura L. Kiessling for being my thesis advisor. It has been an honor and privilege to have worked under you. I cannot thank you enough for giving me a chance; for your incredible scientific insight, encouragement, and guidance; for believing in me during the times I did not.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

3-D	three-dimensional
ABPP	activity-based protein profiling
ACN	acetonitrile
BCA	bicinchoninic acid
BSA	bovine serum albumin
CuAAc	copper-catalyzed azide-alkyne cycloaddition
DCM	dichloromethane
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
THF	tetrahydrofuran
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ESI-TOF	electrospray ionization time-of-flight
FBLD	fragment-based ligand design
gal*f*	galactofuranose
GlfT2	galactofuranosyltransferase 2
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	human immunodeficiency virus
HRMS	high resolution mass spectrometry
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl β-D-1-thiogalactopyranoside
isoTOP-ABPP	isotopic tandem orthogonal proteolysis-activity-based protein profiling
LB	lysogeny broth

LC-MS	liquid chromatography-mass spectrometry
MALDI	matrix-assisted laser desorption/ionization
MALDI-TOF	matrix-assisted laser desorption/ionization-time of flight
MS/MS	tandem mass spectrometry
M. tuberculosis/Mtb	*Mycobacterium tuberculosis*
N. brasiliensis/Nbr	*Nocardia brasiliensis*
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
PBS	phosphate-buffered saline
PEG	polyethylene glycol
PHYH	phytanoyl-CoA-dioxygenase
R. equi/Req	*Rhodococcus equi*
RT	room temperature
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TEV	tobacco etch virus
UDP	uridine diphosphate
UDP-gal*f*	uridine diphosphate galactofuranose
UMP	uridine monophosphate

Chapter 1: Covalent Functionalities in Small Molecular Probes

1.1 Introduction

Probing proteins using small chemical probes has been instrumental in studying their structure and elucidating enzyme mechanisms. These small compounds have often shown therapeutic benefits as well, as they are used to treat a variety of ailments ranging from inflammation to cancer.[1] While some probes act only transiently, one subgroup possesses a mode of action whereby they form a permanent covalent bond to their target, granting many additional benefits. In more fundamental studies, covalent bond formation and ligation can aid in studying posttranslational modification effects on proteins as well as assist in characterizing different proteins and enzymes within complex mixtures.[2–4] As a therapeutic, these covalent molecules have stronger potency than their non-covalent equivalents and retain a longer efficacy in the body as well as other pharmacodynamic and pharmacokinetic benefits.[5,6]

The crux in designing a practical covalent small molecule is the selection of an amino acid residue to target and a small electrophile to serve as the chemical "warhead" that forms the bond with said target (Figure 1.1A). Selection of a target nucleophile may seem trivial; however, a key challenge in designing covalent therapeutics is the ability to overcome nonspecific binding to nucleophiles that would produce effects that are unhelpful or even deleterious. It is not as simple as finding a reactive handle that interacts with an alcohol, thiol, or amine *in situ*, as

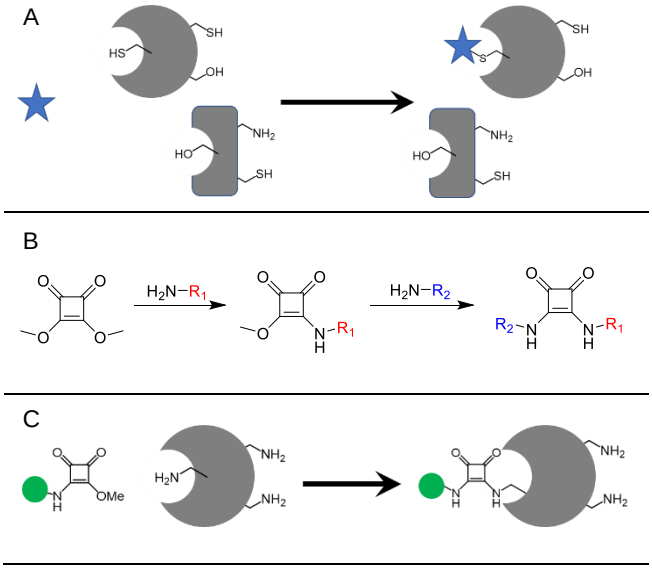


Figure 1.1 (A) Targeted covalent probes must be tailored to not only be residue specific but also specific to an amino acid among many of the same type on the protein surface. (B) Squaric esters are easily modified asymmetrically through simple stepwise addition of amines. (C) Squaric esters may be useful in developing lysine-selective covalent probes due to their innate mild electrophilicity.

biological samples and even a single protein contain multiple copies of the same amino acid that could compete for binding with the electrophile. Consequently, a significant amount of research has focused on characterizing the specificity and reactivity of different electrophiles toward the nucleophilic amino acids, such as serine, cysteine, and lysine to aid in developing more selective covalent ligands.[7–11]

The different reactivities of each amino acid have their own advantages and pitfalls that make selection very situational. For example, cysteines contain a highly nucleophilic thiol, and its scarcity of 1.4% across proteomes ranging from viral to human genomes limit the chance for off-target binding.[12] This scarcity however means that the number of applicable cases where cysteine is a valid target is also low and often requires further protein manipulation with techniques such as site-directed mutagenesis. Mutation is a viable option, but every additional perturbation increases the chance that experimental results may not properly reflect accurate protein structure and enzyme dynamics. On the other hand, lysine is another nucleophilic amino acid that appears in nature over four times that of cysteine, and offers an alternative avenue in developing covalent probes.[12,13] As such, there has been work in developing amine-directed electrophiles, such as dichlorotriazines, acyl-phosphates, and phenylboronic acids to install onto small molecular scaffolds.[11,14,15] While the increased abundance of lysine offers a larger range of utility, its prevalence coupled with its much weaker nucleophilicity make selectivity much harder to achieve.

An amine-reactive electrophile that has been arguably underrepresented in chemical biology is the squaric ester. Squaric esters have been extremely useful to link two different amines under mild conditions (Figure 1.1B). In organic synthesis, they have been used in developing complex chiral ligands for organocatalysis and linking larger macromolecules, like

polysaccharides and oligopeptides.[16–19] When choosing an amine-reactive electrophile for constructing small molecules, often times more reactive functionalities such as an activated ester are used to offset lysine's weaker nucleophilicity. As such, current lysine-specific covalent ligands offer limited utility. A less electrophilic species, like a squaric ester, may react with a proximal lysine without necessarily dictating protein-ligand interactions (Figure 1.1C). To that end, a more thorough characterization of squaric esters is warranted to evaluate their potential as an alternative lysine-specific electrophile.

1.2 Resurgence of Covalent Drugs and Probes

A small molecule that forms a covalent bond to a target protein has many advantages over their non-covalent counterparts, and this is especially visible in the context of therapeutics. To date, nearly 30% of currently marketed drugs that target enzymes act in a covalent fashion to treat a variety of diseases including cancer, cardiovascular, and gastrointestinal disorders.[1] Aspirin, one of the oldest analgesics that has become a household staple, and penicillin, one of the first widely distributed and arguably the most well-known antibiotic, both covalently inhibit their enzymatic targets.[20,21] Covalent drugs have had much success historically, yet it is only in the last few decades that interest in developing new covalent therapeutics has increased significantly.

One reason for the hesitance in actively pursuing new covalent inhibitors has been the concern over non-selective modification and other idiosyncratic toxicities that might result.[22,23] Compared to non-covalent binders, any unwanted effects and toxicities are amplified as irreversibility transforms any effect into becoming permanent rather

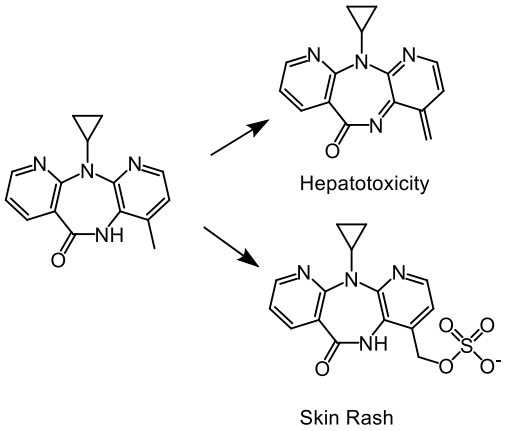


Figure 1.2 Nevirapine, used to treat HIV, is known to be processed by enzymes to form metabolites that can cause liver damage as well as skin rashes.

than transient. This irreversibility creates a considerably narrow margin for error and out of safety concerns, many discovery efforts have excluded electrophiles as potential drug candidates despite the strong historical success of covalent therapeutics.[24] However, idiosyncratic toxicities are not limited to covalent molecules as the vast majority of non-covalent drugs have been shown to cause some form of deleterious side effects through downstream metabolic processes (Figure 1.2).[25,26] In addition, covalent drugs offer numerous advantages that their non-covalent counterparts cannot offer as described below.

1.3 Covalent vs. Non-covalent Inhibitors

Covalent inhibitors can exhibit heightened potency by introducing an additional irreversible binding interaction between the ligand and its target.[27] Unlike non-covalent inhibition, covalent inhibition cannot be competed away by endogenous substrate and would require complete re-synthesis of the protein to regain function.[28] Additionally, it has been documented that non-covalent small molecules have a ceiling for their ligand efficiency, or binding affinity relative to its molecular mass, leading to a theoretical maximum in potency.[29] Ligands that rely on covalent modes of binding have been shown to exceed these limits, decreasing the effective concentration needed, and offer an avenue in developing extremely potent inhibitors. Houk *et al*. documented that while non-covalent inhibitors have a binding affinity limit of about 10pM, addition of a covalent element can increase this affinity by up to five orders of magnitude.[30]

Furthermore, covalent inhibitors can overcome drug resistant mutations whereas non-covalent inhibitors cannot.[31] The sustained inhibition by covalent compounds has been demonstrated by Hagel *et al.* using mutants of hepatitis C virus protease.[32] Because a mutation may decrease the binding affinity to a certain ligand, this changes only the rate of inhibition rather than the extent of inhibition by a covalent inhibitor. Once a covalent inhibitor forms a covalent bond with its target, it is no longer in competition with the natural substrate and

inhibition is irreversible. This unique feature contrasts with non-covalent inhibitors that are always in equilibrium with endogenous substrate *in situ*.

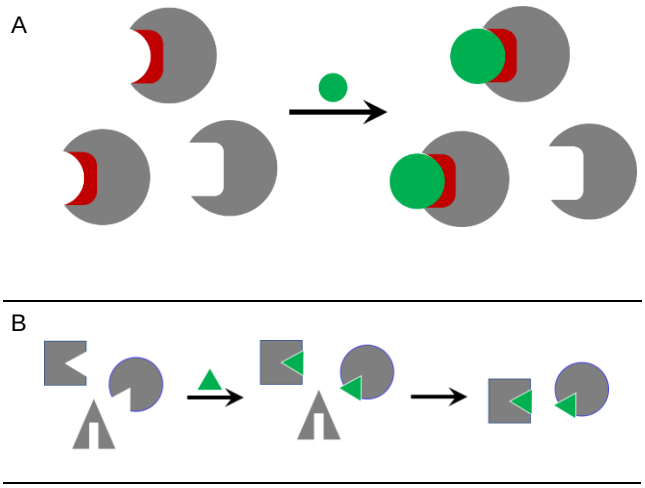


Figure 1.3 Examples of ABPP applications. (A) Covalent modification of post-translationally modified enzymes quantifies activity levels *in situ* as opposed to expression levels. (B) Proteomic identification and separation of enzyme families with similar function within complex mixtures through covalent modification.

Aside from therapeutics, covalent small molecules have been of great use in fundamental studies of protein structure and dynamics. By perturbing an enzyme with a covalent small molecule, one can directly study an enzyme's function. A large area of research using such compounds is activity-based protein profiling (ABPP). These experiments use covalent compounds to observe and probe enzymes based on activity rather than expression, typically binding to an active site only when it is active.[33] By binding only to active enzyme, covalent probes can directly monitor activity levels as opposed to expression levels. The result is a clearer picture of the biological underpinnings of a system *in situ*, unlike techniques like site-directed mutagenesis where manipulation of a protein sequence indirectly probes an enzyme through introducing loss of function (Figure 1.3A). Engagement to an enzyme only while it is active also allows for direct study of post-translational modifications of enzymes as well as uncovering allosteric sites.[4] The application of ABPP probes in proteome-wide experiments can be used to characterize unknown enzymes *in cellulo,*and thereby identify novel therapeutic targets (Figure 1.3B).[34–36] Using a covalent electrophile, enzymes of similar function can be extracted from complex cellular mixtures, even in incredibly low abundance.[37] It is evident that

ABPP probes span a wide range of biological areas that hinge on the ability of forming a covalent bond with its target.

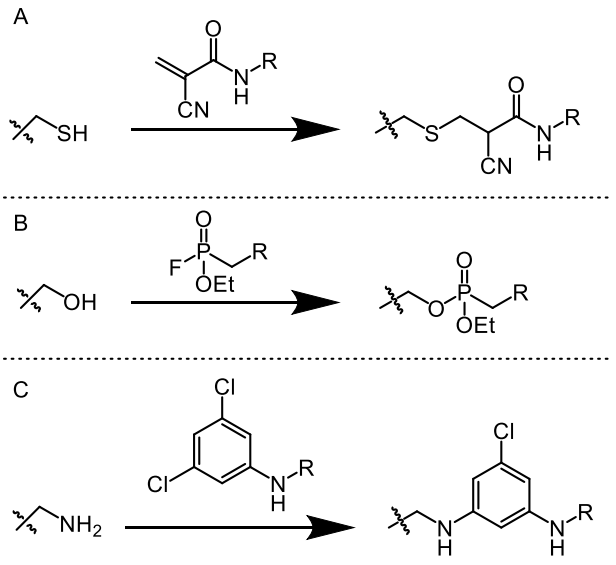


Figure 1.4 Examples of residue specific electrophiles: (A) thiol targeting cyanoacrylamides[38], (B) alcohol targeting fluorophosphonates[7], (C) amine targeting dichlorotriazines[10].

1.4 Current Landscape of Residue Specific Electrophiles

Selecting an appropriate electrophile for a given application is paramount for creating selective probes. Consequently, there has been much attention in characterizing various electrophilic moieties in the context of biological probe design. A review by Weerapana *et al*. catalogues in detail electrophiles tested against different nucleophilic residues (Figure 1.4).[39] Each nucleophilic amino acid possesses different electronic properties, each having their own strengths and weaknesses, and must be weighed carefully when selecting a target for covalent modification.

Cysteines are often important for protein structure and function, through disulfide bond formation, redox chemistry, and metal binding; therefore they have been a common choice for covalent modification.[40–43] More importantly for covalent probe design, thiolate residues are excellent nucleophiles, and therefore a weaker electrophile can be captured. Also, cysteine as an amino acid is typically low in abundance, occurring at around 1.4% in the UniProt database, and helps limit off-target binding as well.[12] Given its distinct reactivity, cysteine has the most unique electrophiles, including but not limited to, Michael acceptors, maleimides, and sulfonate esters. Michael acceptors are especially common as their covalent inhibition is reversible, helping to

ease concerns related to off-target modification. Taunton *et al.* employed cyano-acrylamides in creating reversible inhibitors to target non-catalytic cysteines (Figure 1.4A).[38] Their strategy led to potent inhibitors exhibiting long residence times across several kinases. However, the low abundance of cysteine is also a weakness for covalent probe design, as it limits the number of situations where a cysteine-targeting strategy can be applied.

Serine on the other hand offers an alternative to cysteine's scarcity as it is the most abundant nucleophilic amino acid in nature with an occurrence of 7.1%.[13] Serine is also often part of catalytic triads important for enzyme function; thus, active site serine residues can be prime candidates as target nucleophiles.[44–46] In fact, both aspirin and penicillin covalently modify a serine as their mechanism of action.[20,21] As a result, there has also been much effort in developing serine-targeting probes, to mimic the success of these mainstay therapeutics.[47,48] In addition to the β-lactams and lactones seen in penicillin and other penicillin derived compounds, various phosphonates and phosphoramidates have been explored as electrophiles (Figure 1.4B).[49,50] Verhelst *et al.* developed a platform using diphenyl phosphonates linked to oligopeptides to selectively probe endogenous proteases in an activity-dependent manner.[51] While cysteine and serine-directed electrophiles constitute most of the work evaluating reactive handles for probe design, efforts to develop probes that target other amino acids like glutamate and tyrosine have also seen success in further expanding the target space for ABPP probes.

Over the last few decades, there has been an increase in the development of lysine selective covalent probes that take advantage of amine-reactive moieties. Lysine is weakly nucleophilic, over four times more abundant than cysteine, and catalytically important to many enzyme functions.[12,13] Like serine, lysine's abundance paired with its low nucleophilicity makes designing selective lysine-targeting probes challenging. The pH of the system can also alter

reactivity significantly,[52] and in most biologically relevant environments the lysine amino group will be primarily protonated. Current efforts have yielded aryl boronic acids and dichlorotriazines that label lysine over other nucleophilic residues like cysteine, but these functionalities were observed to be highly reactive, labeling non-catalytic lysines as well (Figure 1.4C).[14,53] While still useful for larger proteomic studies, electrophiles such as these do not seem suitable for more targeted experiments. Kozarich *et al.* demonstrated that acyl phosphates could be used to selectively target protein kinases in complex cell lysates.[15] These phosphates however, are highly susceptible to hydrolysis and have poor cell permeability due to the charged phosphate group, limiting the use of this strategy to *in vitro* experiments only. The shortfalls exhibited by these electrophiles highlight the need for new alternatives for amine-reactive functionalities in the pursuit of lysine-selective protein profiling.

1.5 Applications of Squaric Esters

An amine-reactive electrophile that possesses suitable properties for covalent probe design are the squaric esters, or squarates. Squarates are small aromatic butenediones that possess two amine-reactive sites.[54] These sites can react with amines under mild conditions albeit with slow kinetics. Additionally, the electronics of squarates are such that after conjugation of one amine, the resulting squaric ester-amide is less reactive that the starting squarate.[55] This property allows for the facile synthesis of asymmetric squaramides through simple stepwise addition of amines.[56] Work from Rawal *et al*. utilized this property as well as the unique hydrogen bonding afforded by the squaramide product to synthesize chiral ligands to control

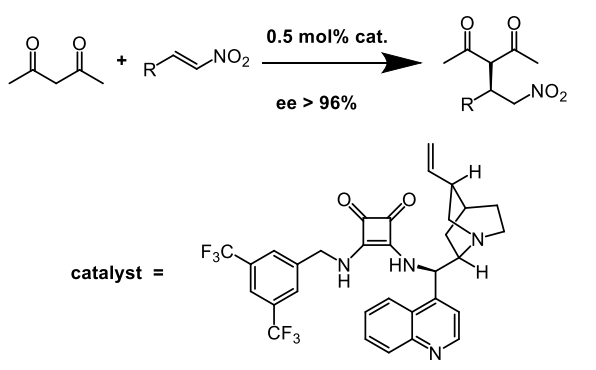


Figure 1.5 Example of an asymmetrically modified chiral squaramide ligand.

enantioselectivity to a high degree in C-C bond formation (Figure 1.5).[57,58] Squarates are unreactive with thiols and most alcohols and therefore have been useful in protein and sugar chemistry. By avoiding the many hydroxyl groups present on sugars as well as amino acids like cysteine and serine, squarates can link large, complex polysaccharides and polypeptides in a directed fashion. Consequently, squarates have been used in the synthesis of neoglycoproteins, or novel synthetic glycoproteins. Over recent years, several reported neoglycoproteins have been employed as experimental carbohydrate vaccines and ligands for studying lectins, or carbohydrate binding proteins.[59–64] Because of their reactivity profile, squarates can link together large aminoglycosides and polypeptides at specific sites while avoiding other potential nucleophilic sites, such as the hydroxyl groups on sugars or serine and cysteine residues in peptides. Kiessling *et al.* has shown that by using dimethyl squarate to link together the α-Gal epitope to an oligopeptide known to bind to the $\alpha_v\beta_3$ integrin selectively, the resulting bifunctional ligand could recruit anti-Gal antibodies to cells displaying the $\alpha_v\beta_3$ integrin (Figure 1.6).[65] These results highlight the use of squarates in a modular approach to synthesize ligands that could direct small molecules to specific cells and will be useful in the development of novel antitumor agents.

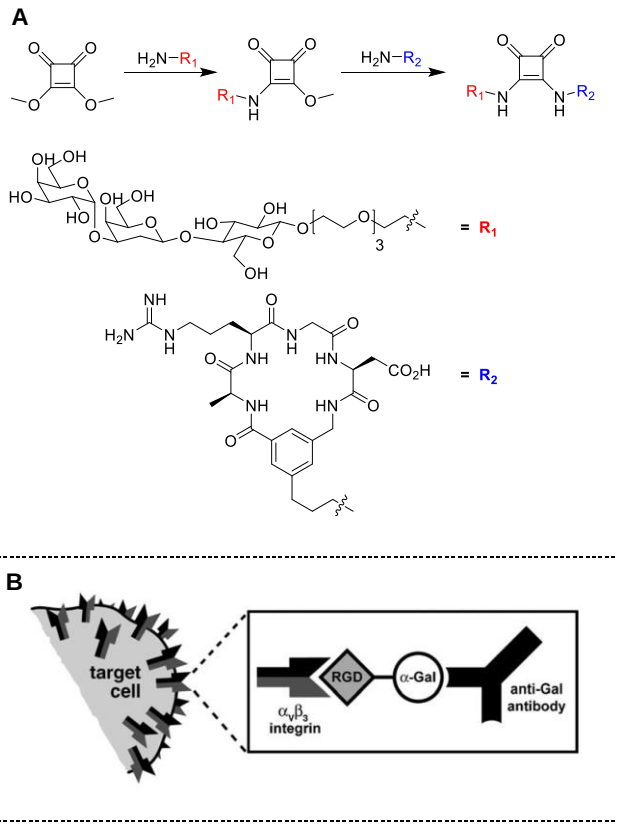


Figure 1.6 (A) Neoglycoprotein formed by conjugating α-Gal with a $\alpha_v\beta_3$-targeting peptide for (B) recruiting anti-Gal antibodies to $\alpha_v\beta_3$-displaying cells.

While chemical biology has taken advantage of the simple asymmetric modification of squarates to construct complex ligands for a variety of biological purposes, the innate reactivity of squarates in general offers another path for selectivity. Until now, the squarate's role in biology has merely been a linker between two larger macromolecules; however, the attenuated electrophilicity of squarates could be leveraged as a mild lysine-targeting electrophile for activity-based probes. Due to their mild reactivity, a squarate-containing small molecule could in theory dictate binding using a separate recognition motif, and the squarate could react to any proximal lysine after the initial binding event, without greatly perturbing the affinity of the recognition element.

1.6 Conclusions and outlook

Covalent modification of proteins presents a high risk, high reward situation that has slowly gained more and more consideration over the last few decades. Small molecules that can permanently bind to enzymes have many applications from fundamental mechanistic studies and proteomic profiling to medicinal therapeutics. Even with potential off-target binding and other idiosyncratic toxicities being amplified by forming a covalent bond, the potential benefits of more potent drugs and activity-based profiling are starting to warrant a more in-depth exploration of applying covalent binders *in vivo*. Because the electrophile can elevate the effects of off-target toxicities, there has been significant research done to find appropriate electrophilic moieties for different residues and different applications. These efforts have led to the characterization of a smorgasbord of reactive handles, especially ones that target cysteine thiols and serine alcohols.

Lysine, with its amine group, presents another potential nucleophilic site that still contains much unexplored potential. While there have been efforts in pursuing novel probes to target lysine, the current selection of electrophiles remains lacking due in part to the amine's

weak electrophilicity in aqueous media coupled with the natural abundance of lysine across the proteome. To date, electrophiles that have been used to probe lysine have been too reactive to be selective or too unstable in water to be efficacious beyond *in vitro* experiments. The simplicity of squarate coupling in developing bifunctional ligands for biology could be leveraged in forming covalent bonds to lysines *in situ*. Perhaps a slower reactivity from the electrophile can help alleviate off-target binding by allowing the molecular scaffold to retain its original binding affinity, and formation of a covalent bond would only occur when a lysine's amine is proximal to the electrophile after the initial binding event. Covalent modification of enzymes could be viewed as linking an enzyme with a smaller molecule, but under more complex conditions. With the prevalence of squarates serving as linkers to biopolymers like polysaccharides and polypeptides, it is worth further efforts in characterizing and evaluating the effectiveness of squarates as an electrophilic warhead in covalent probe design.

Chapter 2: Reactivity and Kinetics of Squaric Esters and Squaric Ester Derivatives

Contributions:
Synthesis, kinetic studies, and computational modeling were performed by Jordan Ho

2.1 Abstract

Squaric esters are amine-reactive aromatic electrophiles, composed of a cyclobutenedione core containing two vinylogous esters. The unique reactivity of squarates allows for them to serve as linchpins to link two amines asymmetrically. The prevalence of lysine residues in protein binding sites and the ability of squarates to react selectively with amines suggests that squarates could serve as useful affinity probes. To explore this possibility, we benchmarked the kinetics of squarates and squarate derivatives in comparison to other amine reactive electrophiles. We find that squarates react with amines at rates up to four orders of magnitude slower than a typical N-hydroxysuccinimidyl ester; thus, squarates have attenuated reactivity relative to their competitors, an indicator of their ability to react selectively with amines in the binding site. We further examined how changing the aromaticity of the squarate core can increase reactivity. Unlike the standard squarates, substitution of squarate oxygens with sulfur produced dithionosquarates which were highly electrophilic and dithiosquarates which could only be amidated once. Computational studies corroborated these experimental results. Our findings indicate that squaric ester-amides have an attenuated reactivity especially beneficial for developing reactive affinity-based probes.

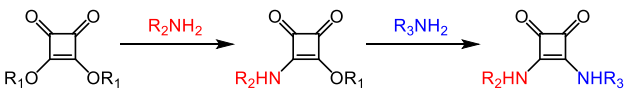
2.2 Introduction

Squarates, or squaric esters, are small electrophiles that contain a cyclobutenedione core appended to two alkoxy groups. The embedded two reactive vinylogous esters allow for sequential conjugation of two different amines under mild conditions to create asymmetric bis-amides (Scheme 2.1).[56,66] When one vinylogous ester reacts with an amine, the intermediate squaric ester-amide becomes less electrophilic than the starting material, allowing for bis-amide synthesis via stepwise addition of amines. This property has been leveraged extensively in the construction of chiral ligands for organocatalysis.[57,58,67,68]

Though many organic syntheses are optimized with highly reactive species to maximize reaction rate and yield, applications in chemical biology sometimes necessitate selectivity over reactivity. The 1,2-dione core of squarates have been shown to be a strong hydrogen bond acceptor, and the amino-selectivity of squarates has been extremely useful in linking together various carbohydrates through aminoglycosides and peptides through lysine residues.[19,65,69–71] Furthermore, the attenuated reactivity of squarates could prove advantageous when designing covalent probes to avoid off-target binding in complex biological milieu. There has been a recent increase in the pursuit of covalent therapeutics and small probes due to their high potency relative to non-covalent small molecules.[38,72–76] This development however has been limited by the need for more mild electrophiles, ones that can be reactive enough to form bonds but not reactive enough that it overpowers other binding interactions.

To best apply squarates and their derivatives towards these critical biological contexts, a precise understanding of their unique reactivity is necessary. To this end, we synthesized a panel of model squarates with varying electronic properties. We benchmarked reaction kinetics

Scheme 2.1. General scheme for sequential amidation of squarates resulting in asymmetric squaramides.



to a more commonly used amine-reactive electrophile, an N-hydroxysuccinamidyl (NHS) ester. We observed that the reactivity of the squarate core can be finely tuned over a moderate range through substituent modification, yet the most electrophilic squarate tested still reacted over 100 times slower than an NHS-ester installed on a similar scaffold. Finally, we explored sulfur substitutions to further enhance electrophilicity of the squarate core and found that replacing the carbonyl oxygen as well as the alkoxy oxygen of the squarate with sulfur drastically changed the reactivity profile depending on the location of the sulfur substitution. This work represents the most thorough characterization of squarate kinetics in the electrophilic landscape to date and demonstrates how the electronic properties of squarates and squarate derivatives drive reactivity.

2.3 Results and Discussion

Squarates react significantly slower than a conventional amine-reactive electrophile. We first investigated the reaction rate of a model squarate to benchmark kinetics relative to an NHS-ester, a commonly employed amine-reactive handle. Using dibutyl squarate **2.1**, we used ^{1}H

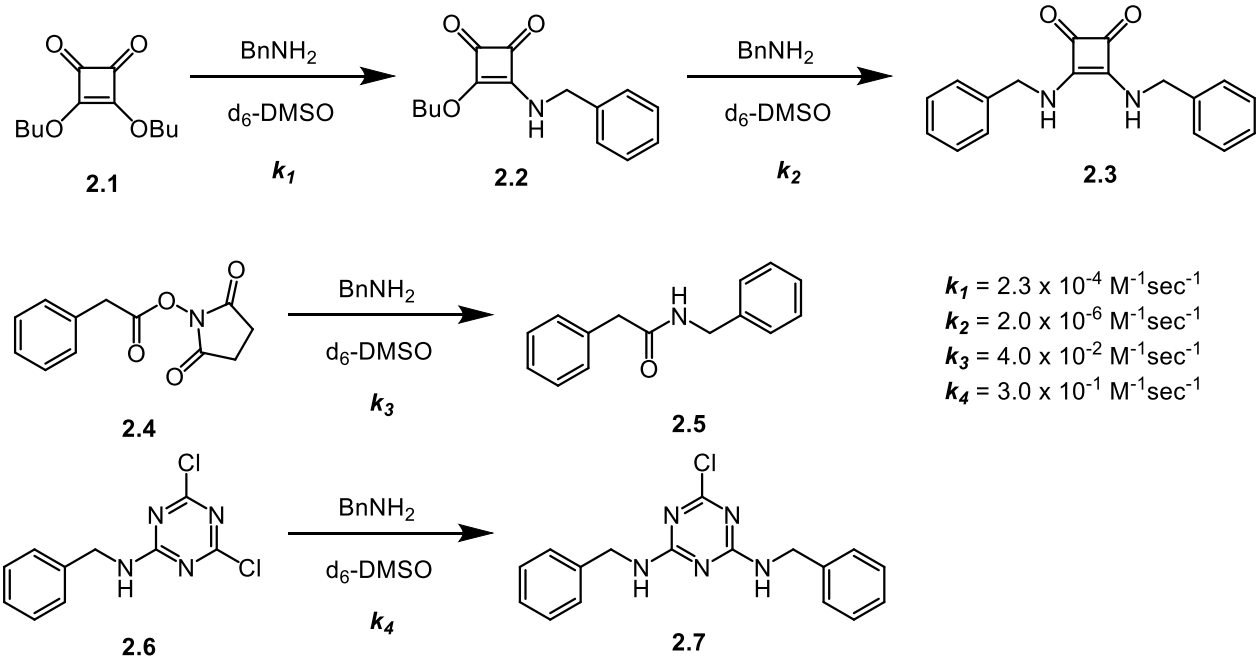


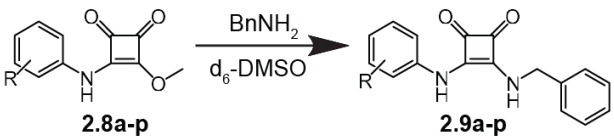
Figure 2.1. 2nd order rate constants for (A) the first and second amidation of dibutyl squarate, **2.1**, with benzylamine, (B) phenethyl NHS-ester, **2.4**, reacting with benzylamine, and (C) benzylamino-dichlorotriazine **2.7** with benzylamine.

NMR to monitor reaction progress and quantified the sequential amidations with benzylamine (k_1 & k_2) (Figure 2.1A). In d_6-DMSO, dibutyl squarate reacted with benzylamine at room temperature with a second order rate constant of 2.3×10^{-4} $M^{-1}sec^{-1}$. As expected, after the first amidation the reactivity drops significantly. Reaction with the amine introduces electron density into the squarate ring, lowering the electrophilicity of the squaric ester-amide product. Appending a second benzylamine to intermediate product **2.2** occurs at a rate of 2.0×10^{-5} $M^{-1}sec^{-1}$, two orders of magnitude slower. This significant drop in reactivity highlights how bis-amidation formation can be minimized through stoichiometric control.

Under the same conditions, an analog of **2.2** bearing an NHS-ester as the electrophile was found to react with benzylamine with a second order rate constant of 4.0×10^{-2} $M^{-1}sec^{-1}$ (Figure 2.1B). In addition, dichlorotriazines have been used in covalently binding lysines in global proteomic experiments.[74] Using a model dichlorotriazine similar to our NHS-ester, **2.6**, we found that to react even faster with a rate constant of 3.0×10^{-1} $M^{-1}sec^{-1}$ (Figure 2.1C). These electrophiles purposed for biology are reacting up to five orders of magnitude faster than their squaric ester counterpart, showcasing just how mild squaric esters can be. Even the more reactive dibutyl squarate, **2.1**, proceeds at a rate at least ~200 times slower than the NHS-ester. These data demonstrate the attenuated reactivity

R	σ_R	k_H/k_R	$Log(k_H/k_R)$
2.8a H	0	1.00	0
2.8b *p*-$(CH_3)_2N$	-0.83	0.33	-0.48
2.8c *p*-OH	-0.37	0.40	-0.40
2.8d *p*-OMe	-0.27	0.47	-0.33
2.8e *p*-Me	-0.17	0.67	-0.18
2.8f *p*-iPr	-0.15	0.93	-0.03
2.8g 3,5-$(Me)_2$	-0.07	0.89	-0.05
2.8h *p*-vinyl	-0.04	1.13	0.05
2.8i *p*-phenyl	-0.01	1.20	0.08
2.8j *3,5*-$(OMe)_2$	0.12	1.42	0.15
2.8k *p*-I	0.18	1.83	0.26
2.8l *p*-Cl	0.23	1.58	0.20
2.8m 3,5-$(CF_3)_2$	0.43	2.08	0.32
2.8n *p*-CF_3	0.54	2.75	0.44
2.8o *p*-CN	0.66	3.08	0.49
2.8p *p*-NO_2	0.78	3.00	0.48

Table 2.1. Hammett parameters **(σ_R)** of various functional groups and their effect on the reaction rate of benzylamine with substituted phenyl squaric ester-amides **2.8a-p**.



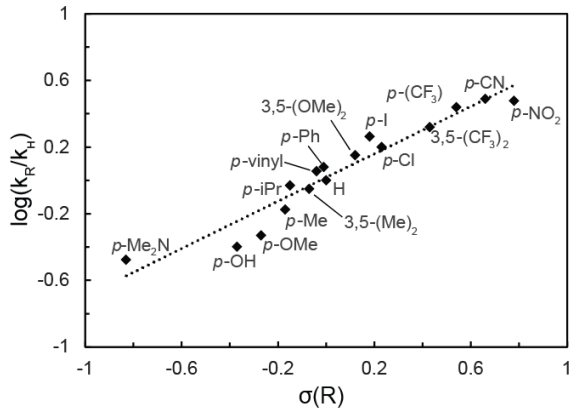


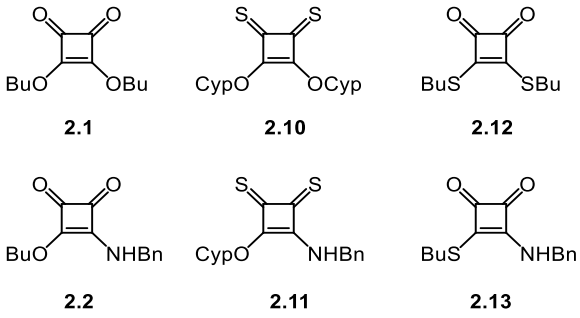
Figure 2.2. Hammett plot of monosubstituted aromatic squarates reacting with benzylamine in d_6-DMSO. Substitution constant, σ, is plotted with the log of the relative reactivity of a squarates **2.8a-p**.

of squarates, suggesting that this system could be utilized in experiments where precise reaction monitoring is required.

Electrophilicity of squaric esters can be finely tuned. Next, we explored how electronic perturbation on the squaric ester core affected its reactivity. Dimethyl squarate was reacted with aniline to give **2.8a** as the control squarate. The second order rate constant for **2.8a** reacting with benzylamine was calculated to be 7.2 x 10^{-5} $M^{-1}sec^{-1}$ reflecting the increase in electrophilicity when the π-system is extended. Compared to squaramide **2.2** tested above, this change from a benzylamine to an aniline derivative increased the rate by a factor of ~40. It should be noted however, that this faster squarate still reacts ~500 times slower to the NHS-ester tested above.

We then appended different electron withdrawing and electron donating groups around the aryl ring and observed their effect on reactivity (Table 2.1). Introduction of electron withdrawing functional groups onto the phenyl ring of **2.8a** further increased the overall electrophilicity of the compound, while electron donating groups had the opposite effect (Table 2.1). A Hammett plot of these values shows a direct relation between reactivity and substitution constant, σ (Figure 2.2). Although these changes are relatively minor, spanning about one order of magnitude, they are

Scheme 2.2. Model squarate **2.1**, dithionosquarate **2.8**, dithiosquarate **2.10**, and their respective monobenzylsquaric ester-amides of interest.



consistent with general electrophile trends where reactivity is inversely related to electron density. We observed that the difference between the most reactive squarate and the least reactive squarate tested in this group was only about one order of magnitude. These data show how changes in the structure of the delocalized π-system can moderately affect the reactivity of a squarate but not to an extent that brings in it line with other amine-reactive electrophiles, and thus such modifications are best for fine tuning squarate reactivity.

Substitution of oxygen with sulfur yields unique reactivity profiles. Lastly, in attempts to significantly perturb squarate kinetics, we explored how the replacement of oxygen atoms with sulfur affected their reactivity (Scheme 2.2). We initially synthesized the dithionosquarate derivative of compound **2.1** through treatment of Lawesson's reagent. Unfortunately, this compound was prone to quick decomposition and so we transitioned to using dicyclopentyl dithionosquarate, **2.10**, as it has been documented to be much more stable.[77] Since dicyclopentyl dithionosquarate still contains aliphatic leaving groups, we thought this small change would not perturb the overall reactivity greatly and opted to use **2.10** as the model dithionosquarate to compare to dibutyl squarate. The dithiosquarate derivative of compound **2.1** was obtained simply through treatment of squaryl chloride with butanethiol, producing dithiosquarate **2.12**.

Reacting **2.10** with benzylamine, only the bis-amide was observed. Even under conditions of excess dithionosquarate, only the disubstituted dibenzyl dithionosquaramide was observed on the time scale accessible by our assay. To better understand this dramatically enhanced reactivity, we looked to computational modeling of its electronics. In squarates **2.1** and **2.2**, there is a significant increase in electron density on the squarate carbon alpha to the nitrogen after amidation, but more importantly the electron density on the second electrophilic carbon also increases slightly, supporting that the intermediate product **2.2** is a weaker electrophile and reacts

slower (Figure 2.3). In contrast, when the same analysis is performed on the dithionosquarate core, the electrophilic carbon for the second amidation intriguingly becomes more electropositive (Figure 2.4). When this intermediate product is more electrophilic than the starting material, its amidation will compete with the starting material. This corroborated our empirical data of observing only the bis-amidated dithionosquaramide.

Electrophilicity calculations of modeled compounds and experimental results suggest dithiosquarates have a unique reactivity of only being able to be amidated once. After the first amidation of dibutylthiosquarate, the carbon for nucleophilic attack by the second amine is modeled to be extremely electronegative relative to its neighboring carbons (Figure 2.5). When the amine displaces the first thiol group and introduces electron density into the squarate core, the remaining thioester sulfur is positioned to stabilize the carbanion formed from in an iminium resonance structure with the newly introduced nitrogen (Figure 2.6). This results in the carbon adjacent to the nitrogen becoming much more electropositive relative to the carbon adjacent to the remaining sulfur. This makes nucleophilic attack to displace the remaining thiol extremely unfavorable. This was confirmed by ^{1}H NMR experiments showing that after the first amidation of compound **2.12**, the resulting product was unreactive with an additional equivalent of benzylamine. Previous studies have also shown that treating a dithiosquarate with excess amine under harsh conditions leads to nucleophilic attack at the carbonyl position eventually leading to the ring opening of the dithiosquarate.[78,79] This agrees with the theoretical calculations as the carbonyl carbons are the most electron poor. Along with empirical data, these molecular calculations characterize how replacing oxygen atoms in squarates with sulfur drastically change their reactivity. Unlike standard squarates, dithiosquarates can only react with amines once under mild conditions while dithionosquarates exclusively give the bis-dithionosquaramide.

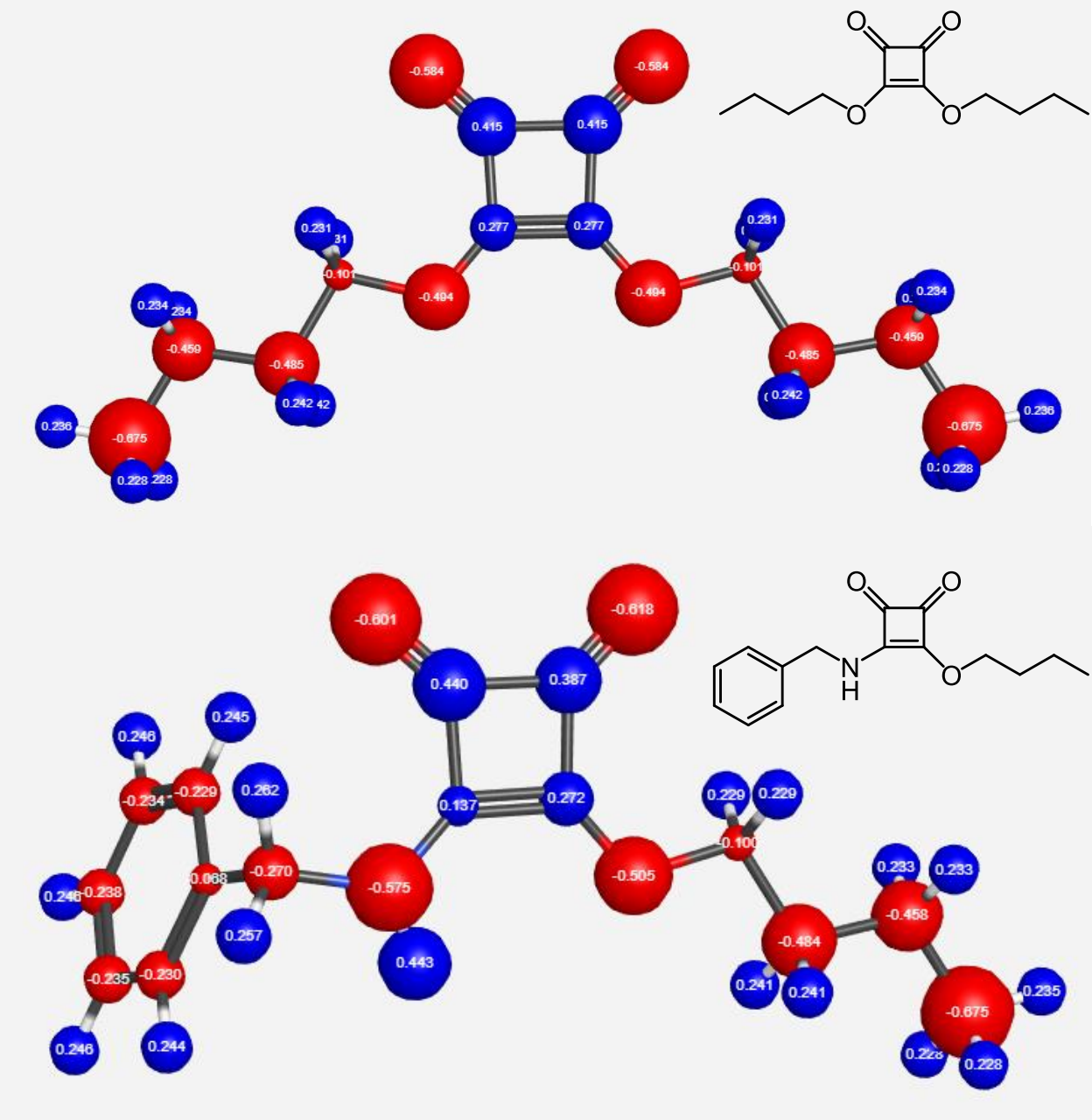


Figure 2.3. Calculated relative electron densities of dibutyl squarate **2.1** and benzylsquaramide **2.2.**

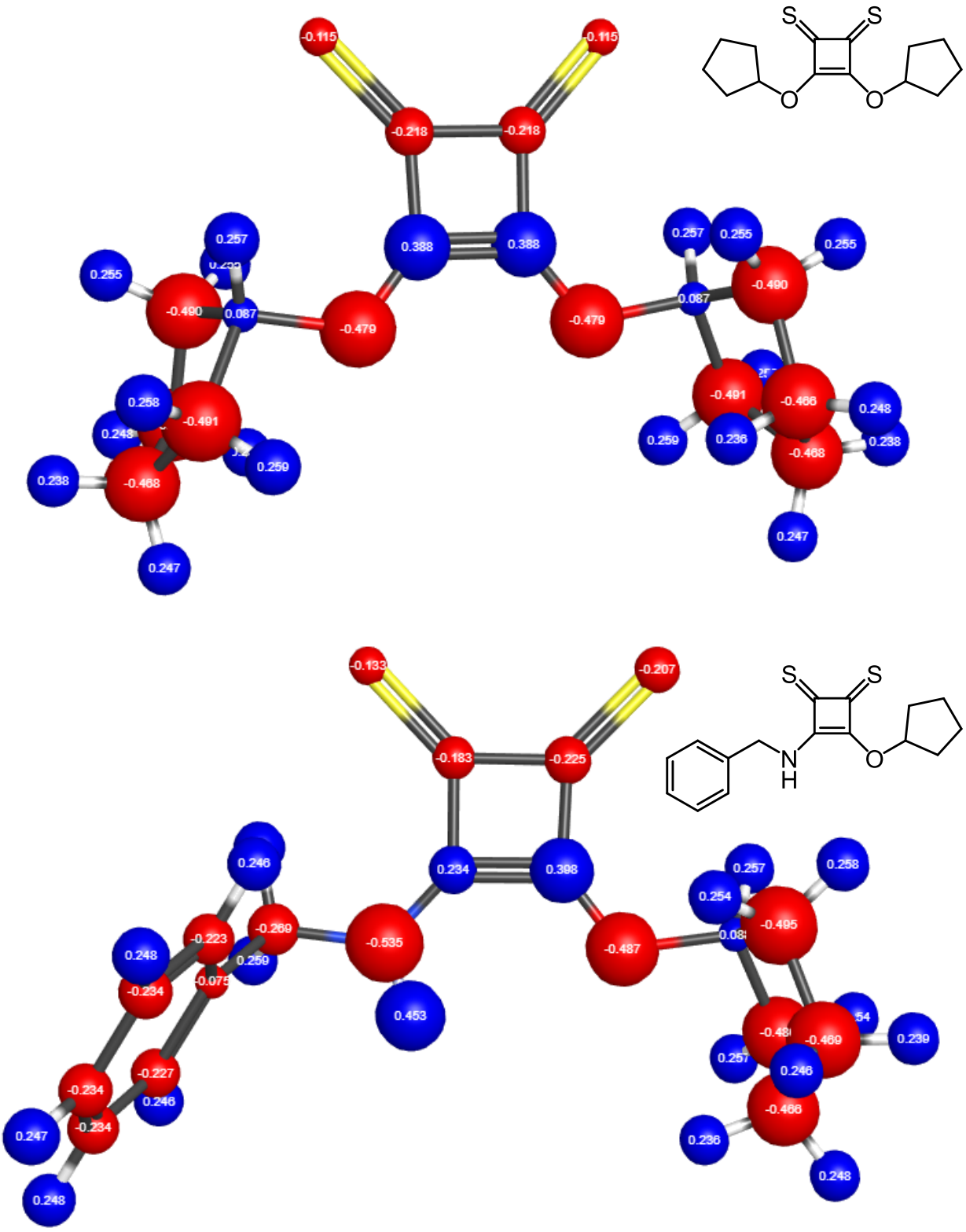


Figure 2.4. Calculated relative electron densities of dithionosquarate **2.10** and respective benzyldithionosquaramide **2.11**.

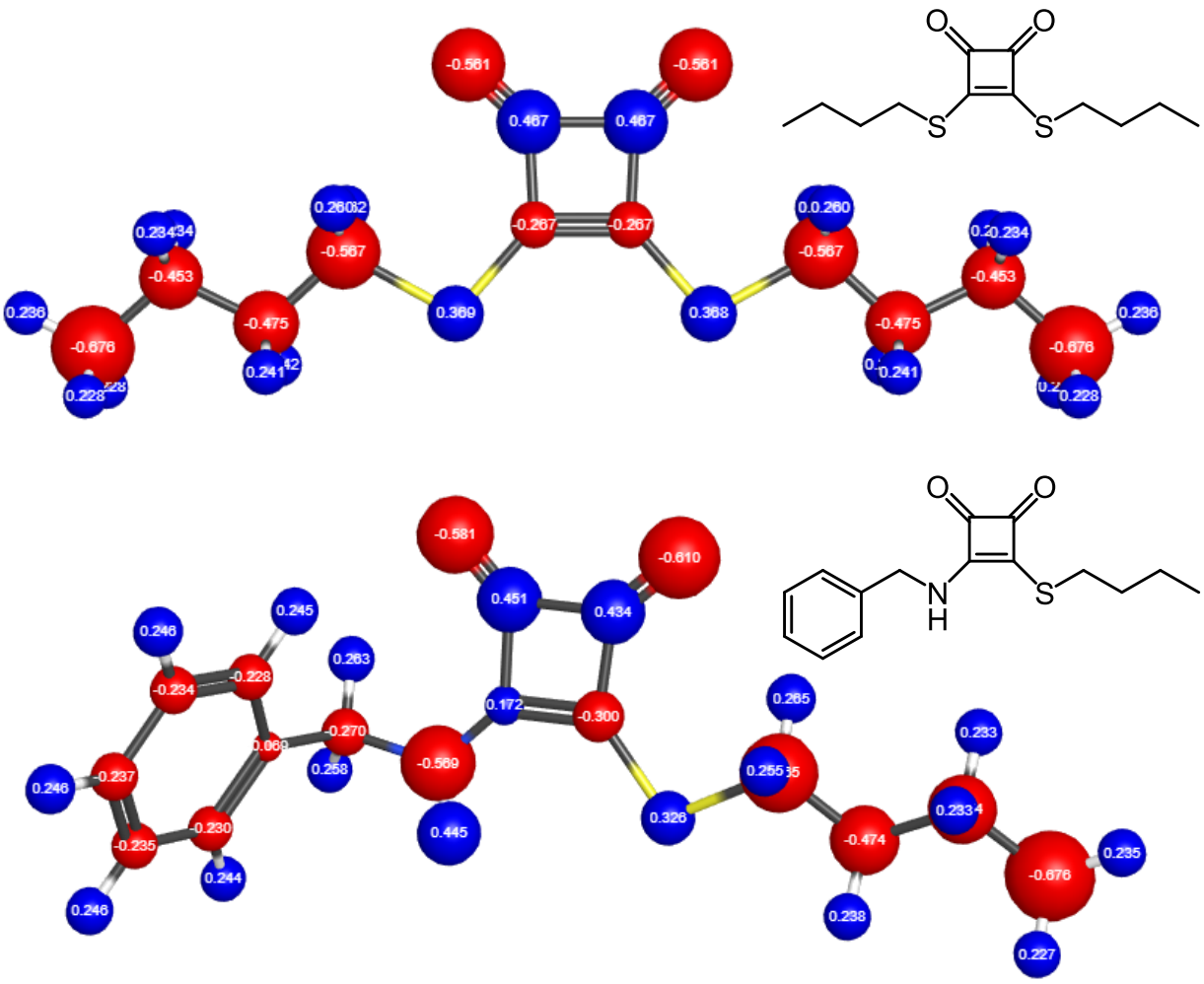


Figure 2.5. Calculated relative electron densities of dithiosquarate **2.12** and respective benzylthiosquaramide **2.13**.

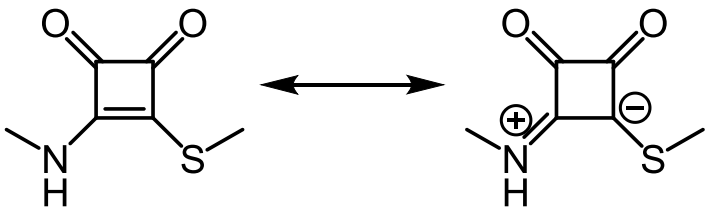


Figure 2.6. Resonance structure of monosubstituted thiosquaramide. Sulfur can help stabilize the negative charge on its neighboring carbon in a resonance form of the intermediate product, preventing the formation of the bis-amide.

While the heightened electrophilicity of the dithionosquarates may be useful in organic chemistry, it appears that they may be too reactive when considering the need for selectivity in more complex biological contexts. Furthermore, because dithiosquarates lack the ability to be bifunctionalized, it becomes much more difficult to install onto a small molecule scaffold while still maintaining a free reactive site. Squaric esters on the other hand possess a unique reactivity that may allow for the development of more selective probes. Because squaric esters have two amine-reactive sites, they can also be easily installed onto an amine-containing scaffold while still retaining a free reactive site that could be utilized in affinity-based probes.

2.4 Conclusions

Comparing squarates to other amine reactive electrophiles, it is easy to see why squarates can be overlooked when developing efficient routes in organic synthesis; however, this attenuated electrophilicity has the potential to be leveraged in developing selective affinity-based probes. Reacting up to four orders of magnitude slower than an NHS-ester, squarates may be able to help alleviate off-target covalent binding that occurs in many current amine-targeting small molecular probes. In situations where there is a complex mixture of many different chemical functionalities but reactivity with only a small subset of those compounds is desired, the installation of a squarate onto a ligand may bestow that ligand the ability to form a covalent bond without perturbing specificity of the original ligand or introducing additional off-target interactions. The kinetic studies of squarates performed here highlight how squarates could offer an alternative avenue when selecting electrophiles in designing small molecular probes.

2.5 Experimental Details

NMR kinetic experiments

Stock solutions of reagents and benzylamine in d_6-DMSO were diluted to a final concentration of 100μM, 10mM, or 50mM directly in an NMR tube for immediate 1H NMR acquisition. Ratios between the starting material squarate methyl protons and product methanol signals as well as the ratio between the methylene proton signals from the benzylamine and products (**2.8a-p**) were used to calculate the concentrations of all species over time. The inverse of concentration was plotted with time, and the slope of the best fit line fitted to data was obtained to calculate rate constants. 1H NMR spectra were recorded on a JEOL 500 MHz spectrometer.

Computational Modeling

Theoretical molecular modeling was performed using WebMO. Structure optimization and atomic electron density calculations were performed at the B3LYP/6-31G level of theory.[80]

Compound Synthesis

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Tetrahydrofuran (THF) was distilled from sodium/benzophenone, methanol (MeOH) was distilled from magnesium, and dichloromethane (DCM) was distilled from calcium hydride. Nuclear magnetic resonance spectra were recorded on a 400 MHz spectrometer (acquired at 400 MHz for ^{1}H and 100 MHz for ^{13}C), or 500 MHz spectrometer (acquired at 500 MHz for ^{1}H and 125 MHz for ^{13}C), or 600 MHz spectrometer (acquired at 600 MHz for ^{1}H and 150 MHz for ^{13}C). Chemical shifts are reported relative to tetramethylsilane or residual solvent peaks in parts per million ($CHCl_3$: ^{1}H 7.27, ^{13}C 77.23; MeOH: ^{1}H 3.31, ^{13}C 49.15; D_2O: ^{1}H 4.79; DMSO: ^{1}H 2.50, ^{13}C 39.52). High-resolution mass spectra (HRMS) were obtained on an electrospray ionization time-of-flight (ESI-TOF) mass spectrometer.

3-(benzylamino)-4-butoxycyclobut-3-ene-1,2-dione **(2.2)**

Using 1.7mL of benzylamine with 2.14g dimethylsquarate, **2.2** was purified via silica gel chromatography (1:4 ACN:DCM) and isolated in 85% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.39 (t, *J*=7.13 Hz, 2H), 7.34 (t, *J*=7.13 Hz, 1H), 7.28 (br, 2H), 6.05 (br, 0.7H, rotamer 1), 5.46 (br, 0.3H, rotamer 2), 4.84 (br, 0.6H, rotamer 2), 4.59 (br, 1.4H, rotamer 1), 4.42 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) (chemical shifts of both rotamers listed) δ 46.81, 47.40, 60.00, 60.28, 127.37, 127.46, 127.51, 128.63, 137.97, 138.40, 171.77, 172.29, 177.11, 177.80, 182.28, 182.54, 189.19, 189.58

2,5-dioxopyrrolidin-1-yl 2-phenylacetate **(2.4)**

408mg of phenylacetic acid, 345mg N-hydroxysuccinimide, and 616mg DCC was dissolved in 10mL DCM under inert atmosphere. Reaction was stirred at 27°C for 16 h. Precipitate was filtered out, and the remaining liquid was concentrated and purified via silica gel chromatography (1:4 EtOAc:DCM). **2.4** was isolated in 94% yield. Physical properties matched previously reported values.[81]

N-benzyl-4,6-dichloro-1,3,5-triazin-2-amine **(2.6)**

250mg of cyanuric chloride and 170mg K_2CO_3 stirred in 10mL THF under inert atmosphere. Solution was cooled to 0 °C. 0.14mL benzylamine was diluted in 5mL THF and slowly added to cooled cyanuric chloride solution. Reaction was allowed to come to room temperature and stirred for 16hr. Solution was concentrated and purified via silica gel chromatography (5% EtOAc in Hexanes. **2.6** was isolated in 64% yield. Physical properties matched previously reported values.[82]

General procedure of monosubstituted squaramides 2.8a-p

To a solution of 1.1mmol (156mg) dimethyl squarate in 10mL MeOH, 1.0mmol of the corresponding aniline was added and reaction was stirred at 27°C for 16 h. The resulting precipitate was filtered, washed with cold MeOH, and vacuum dried to yield the corresponding aryl squaramide, unless otherwise noted.

3-methoxy-4-(phenylamino)cyclobut-3-ene-1,2-dione **(2.8a)**

Using 92μL of aniline, **2.8a** was isolated in 69% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.38 (t, *J*=8.09 Hz, 2H), 7.28 (d, *J*=7.86 Hz, 2H), 7.18 (t, *J*=7.40 Hz, 1H) 4.51 (s, 3H). ^{13}C NMR (150 MHz, d_6-DMSO) δ 60.51, 119.57, 124.02, 129.07

3-((4-(dimethylamino)phenyl)amino-4-methoxycyclobut-3-ene-1,2-dione **(2.8b)**

Using 136mg of N,N-dimethyl-p-aniline, **2.8b** was isolated in 93% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.12 (br, 2H), 6.71 (br, 2H), 4.46 (s, 3H), 2.96 (s, 6H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 40.30, 56.53, 60.28, 112.60, 117.55, 119.93, 120.94, 121.49, 139.23, 142.35, 147.85, 177.10, 188.04

3-((4-hydroxyphenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8c)**

Using 109mg of 4-aminophenol, **2.8c** was isolated in 65% yield. ^{1}H NMR (600 MHz, MeOD) δ 7.14 (br, 2H), 6.76 (d, *J*=8.76 Hz, 2H), 4.42 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.32, 115.44, 121.79, 129.38, 154.52

3-methoxy-4-((4-methoxyphenyl)amino)cyclobut-3-ene-1,2-dione **(2.8d)**

Using 123mg of p-anisidine, **2.8d** was isolated in 77% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.22 (br, 2H), 6.92 (d, *J*=8.91 Hz, 2H), 4.51 (s, 3H), 3.84 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 55.28, 60.38, 114.23, 121.33, 130.95, 156.24

3-methoxy-4-(4-tolylamino)cyclobut-3-ene-1,2-dione **(2.8e)**

Using 107mg of p-toluidine, **2.8e** was isolated in 69% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.16 (m, 4H), 4.49 (s, 3H), 2.34 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 20.37, 60.44, 119.66, 129.46, 133.32, 135.42, 183.65

3-((4-isopropylphenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8f)**

Using 137μL of 4-isopropylaniline, **2.8f** was isolated in 19% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.22 (m, 4H), 4.50 (s, 3H), 2.90 (sept, *J*=6.93 Hz, 1H), 1.24 (d, *J*=6.95 Hz, 6H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 23.89, 32.87, 60.48, 119.73, 126.86, 135.65, 144.39

3-((3,5-dimethylphenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8g)**

Using 125μL of 3,5-dimethylaniline, **2.8g** was purified via silica gel chromatography (2:1 Hex:EtOAc) and isolated in 54%. ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.87 (s, 2H), 6.82 (s, 1H), 4.50 (s, 3H), 2.32 (s, 6H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 21.02, 60.44, 117.34, 125.62, 137.75, 138.19, 183.80

3-methoxy-4-((4-vinylphenyl)amino)cyclobut-3-ene-1,2-dione **(2.8h)**

Using 117μL of 4-vinylaniline, **2.8h** was isolated in 74% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.41 (d, *J*=8.48 Hz, 2H), 7.24 (br, 2H), 6.68 (dd, J=17.43, 10.86 Hz, 1H), 5.72 (d, *J*=17.50 Hz, 1H), 5.25 (d, *J*=10.79 Hz, 1H), 4.51 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.56, 113.42, 119.54, 126.89, 132.96, 135.86, 137.59, 183.89

3-([1,1'-biphenyl]-4-ylamino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8i)**

Using 169mg of 4-aminobiphenyl, **2.8i** was isolated in 76% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.61 (d, *J*=8.51 Hz, 2H), 7.57 (d, *J*=7.68 Hz, 2H), 7.45 (t, *J*=7.68 Hz, 2H), 7.36 (m, 3H), 4.53 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.60, 119.89, 126.32, 127.23, 127.30, 128.92, 135.70, 137.41, 139.35

3-((3,5-dimethoxyphenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8j)**

Using 168mg of 3,5-dimethoxyaniline, **2.8j** was isolated in 37% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.51 (br, 2H), 6.29 (t, *J*=1.97 Hz, 1H), 4.54 (s, 3H), 3.83 (s, 6H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 55.22, 60.60, 95.91, 97.75, 139.68, 160.85

3-((4-iodophenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8k)**

Using 219mg 4-iodoaniline, **2.8k** was isolated in 21% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.68 (d, *J*=7.86 Hz, 2H), 7.04 (br, 2H), 4.51 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.66, 88.07, 121.71, 129.08, 137.70, 184.02

3-((4-chlorophenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8l)**

Using 128mg of 4-chloroaniline, **2.8l** was isolated in 59% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.34 (d, *J*=8.66Hz, 2H), 7.23 (d, *J*=8.49 Hz, 2H), 2.51 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.63, 121.06, 128.02, 128.98, 137.02, 169.03, 183.96

3-((3,5-bis(trifluoromethyl)phenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(2.8m)**

Using 2.1g of 3,5-bis(trifluoromethyl)aniline with 1.0g of dimethylsquarate, **2.8m** was isolated in 20% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.78 (s, 2H), 7.64 (s, 1H), 4.54 (s, 3H). ^{13}C NMR (151

MHz, d_6-DMSO) δ 60.88, 116.02, 119.07, 123.92, 125.73, 125.73, 131.17 (quartet), 140.22, 169.05, 179.83, 184.38, 187.27

3-methoxy-4-((4-(trifluoromethyl)phenyl)amino)cyclobut-3-ene-1,2-dione **(2.8n)**

Using 124μL of 4-trifluoromethylaniline, **2.8n** was isolated in 50% yield. ^{1}H NMR (600 MHz, MeOD) δ 7.66 (d, *J*=8.62 Hz, 2H), 7.57 (d, *J*=8.41 Hz, 2H), 4.50 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.73, 119.27, 123.40, 123.75 (quartet), 125.19, 126.34, 141.63, 169.15, 179.55, 184.42, 187.50

4-((2-methoxy-3,4-dioxocyclobut-1-en-1-yl)amino)benzonitrile **(2.8o)**

Using 118mg of 4-aminobenzonitrile, **2.8o** was isolated in 38% yield. ^{1}H NMR (600 MHz, MeOD) δ 7.72 (d, *J*=8.79 Hz, 2H), 7.58 (d, *J*=8.47 Hz, 2H), 4.50 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.88, 105.31, 118.92, 119.19, 133.50, 142.28, 168.98, 179.94, 184.54, 187.25

3-methoxy-4-((4-nitrophenyl)amino)cyclobut-3-ene-1,2-dione **(2.8p)**

Using 138mg of 4-nitroaniline, **2.8p** was isolated in 30% yield. ^{1}H NMR (600 MHz, MeOD) δ 8.24 (d, *J*=9.23 Hz, 2H), 7.59 (d, *J*=8.66 Hz, 2H), 4.51 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) δ 60.99, 118.93, 125.28, 142.54, 144.29, 169.00, 180.37, 184.90, 187.22

3,4-bis(cyclopentyloxy)cyclobut-3-ene-1,2-dithione **(2.10)**

A solution of squaric acid (2g) and cyclopentanol (11mL) in toluene (20mL) was refluxed in a Dean-Stark apparatus for 7 hours. Solution was concentrated and purified via column chromatography (0-2%MeOH in DCM) to afford intermediate dicyclopentyl squarate. 150mg of dicyclopentyl squarate and Lawesson's reagent (250mg) was suspended in DCM (5mL) under inert atmosphere. Solution stirred at room temperature for two days. Solids were filtered, and the

remaining liquid was concentrated under vacuum and purified via column chromatography (1:1 Hex:DCM) to afford compound **2.10**. ^{1}H NMR (600 MHz, d_6-DMSO) δ 5.94 (m, 2H), 1.98 (m, 8H), 1.74 (m, 4H), 1.64 (m, 4H).

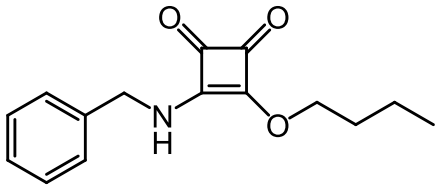
3,4-bis(butylthio)cyclobut-3-ene-1,2-dione **(2.11)**

Squaryl chloride (150mg) was dissolved in THF (10mL) under inert atmosphere. Butanethiol (0.3mL) was added dropwise, and then triethylamine (0.32mL) was added dropwise. Reaction warmed to room temperature and stirred for 6 hours before being quenched with aqueous NH_4Cl (10mL). Solution was then extracted with DCM (3x20mL) before being concentrated and purified via column chromatography (0-10% EtOAc in Hexanes) to afford compound **2.11**. ^{1}H NMR (600 MHz, d_6-DMSO) δ 3.43 (t, *J*=7.65 Hz, 4H), 1.70 (quintet, *J*=7.51 Hz, 4H), 1.40 (sextet, *J*=7.42 Hz, 4H), 0.90 (t, *J*=7.38 Hz, 6H).

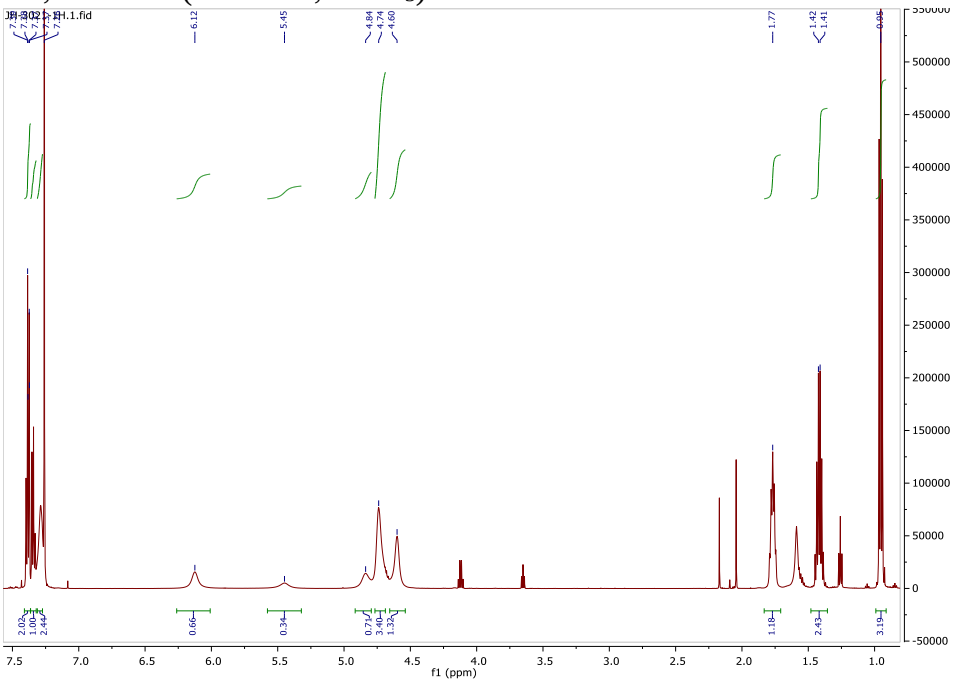
3-(benzylamino)-4-(butylthio)cyclobut-3-ene1,2-dione **(2.12)**

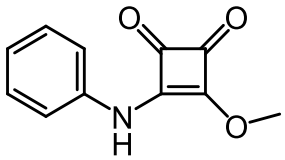
Thiosquarate **11** (33.5mg) was dissolved in DCM (1mL) in a vial. Benzylamine (14.1μL) dissolved in DCM (0.5mL) was then added slowly to thiosquarate solution and stirred at room temperature overnight. Mixture was concentrated and purified via column chromatography (25-40% EtOAc in Hexanes) to afford compound **2.12**. (600 MHz, d_6-DMSO) δ 7.39-7.24 (m, 5H), 4.73 (s, 1.1H, rotamer 1), 4.55 (s, 0.9H, rotamer 2), 4.48 (d, *J*=5.55 Hz, 0.5H rotamer 2), 4.29 (d, *J*=6.28 Hz, 0.5H rotamer 1), 3.41-3.37 (two sets of triplets, *J*=7.41 Hz, 2H), 1.69-1.60 (m, 2H), 1.42-1.33 (m, 2H), 0.90 (two sets of triplets, *J*=7.46 Hz, 3H).

2.6 NMR Spectra

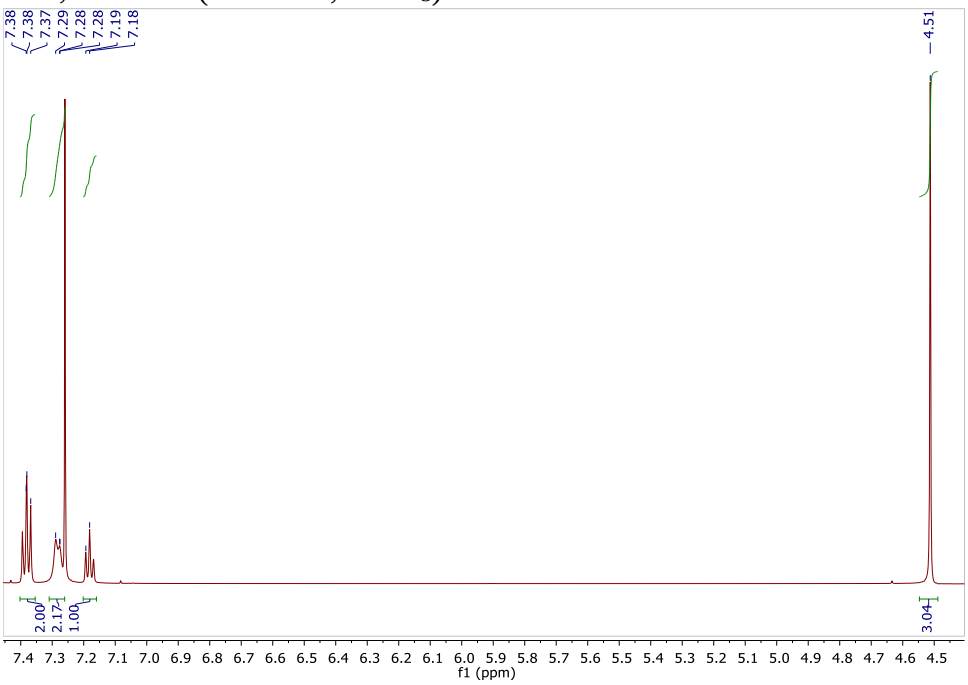


2.2, ^{1}H NMR (600 MHz, $CDCl_3$)

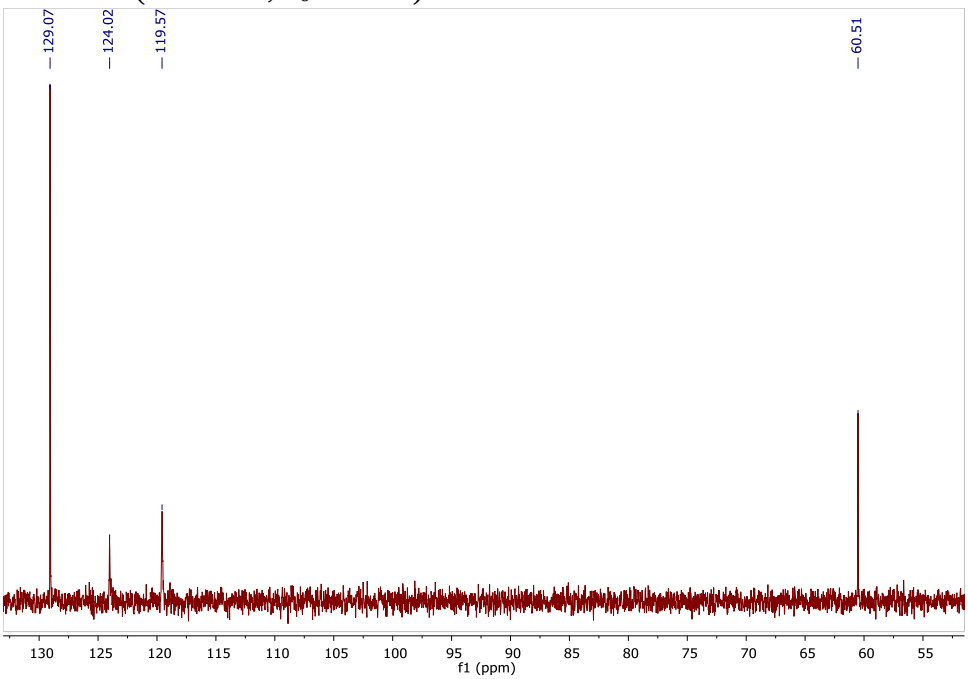


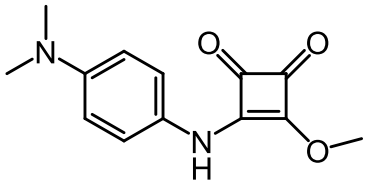


2.8a, ^{1}H NMR (600 MHz, $CDCl_3$)

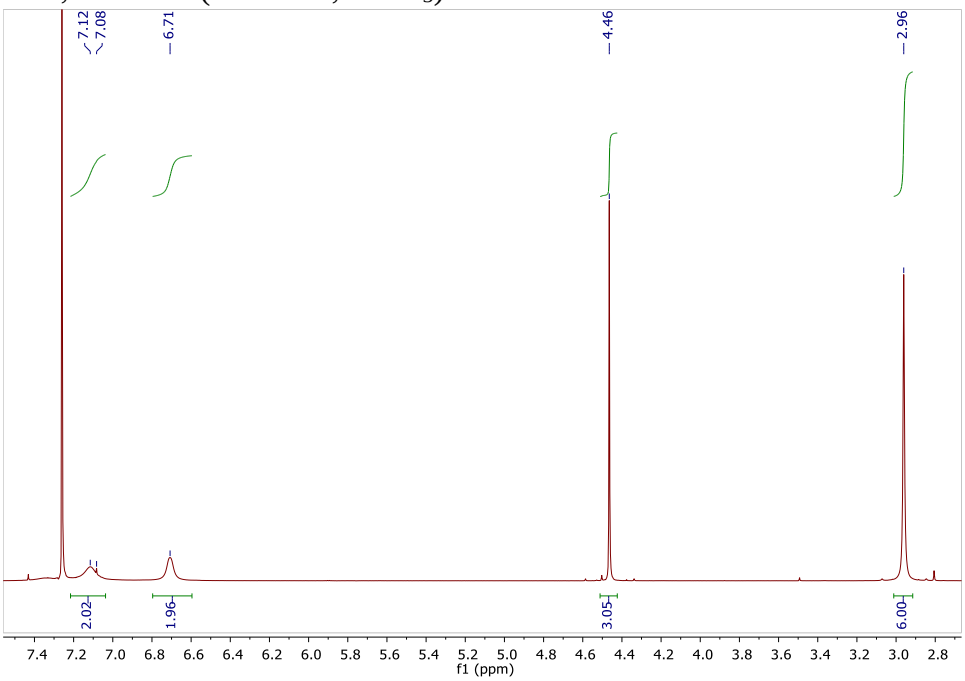


^{13}C NMR (151 MHz, d_6-DMSO)

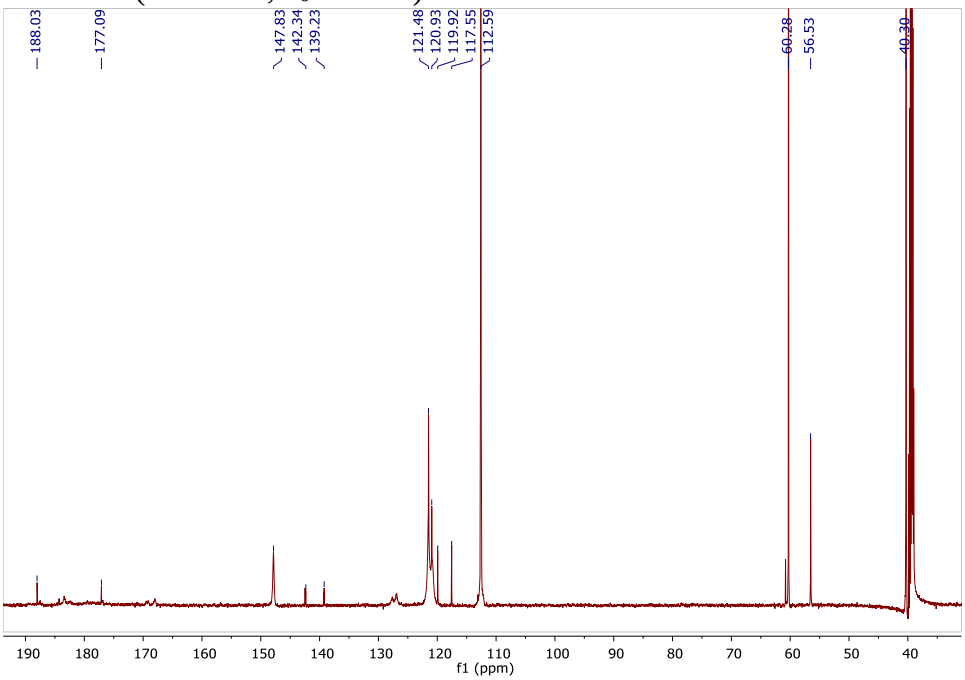


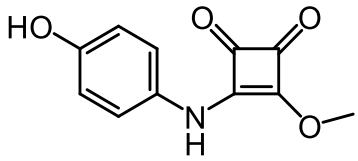


2.8b, ^{1}H NMR (600 MHz, $CDCl_3$)

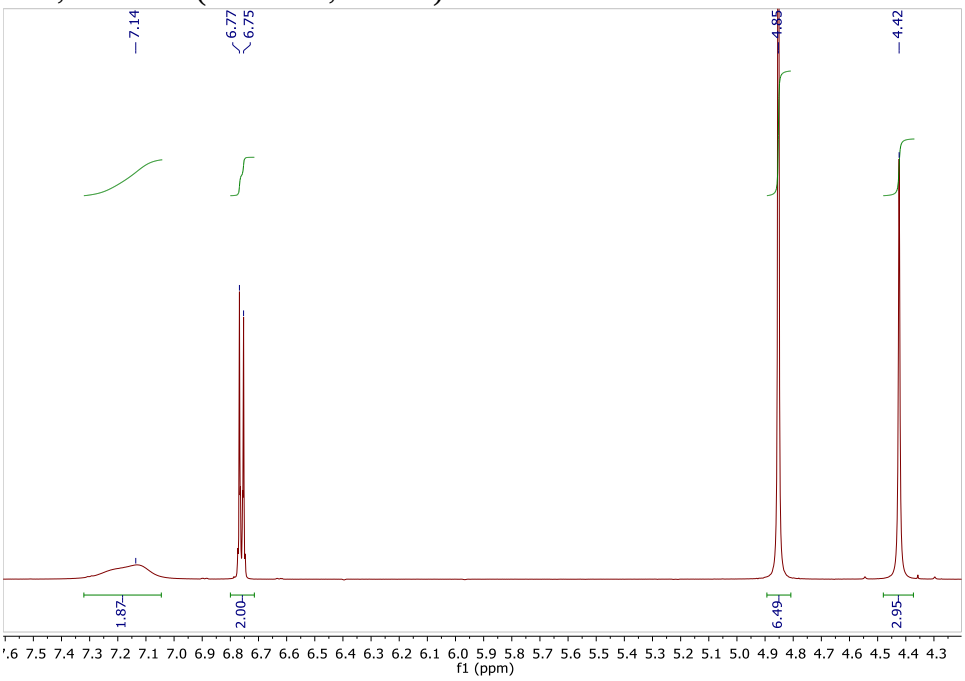


^{13}C NMR (151 MHz, d_6-DMSO)

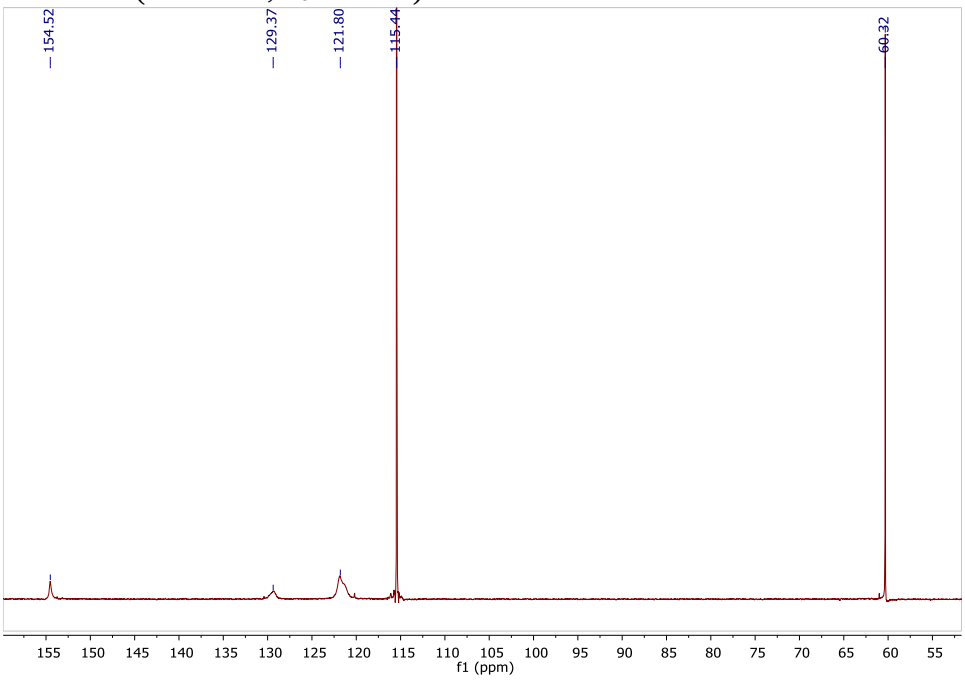


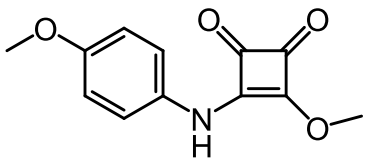


2.8c, ^{1}H NMR (600 MHz, MeOD)

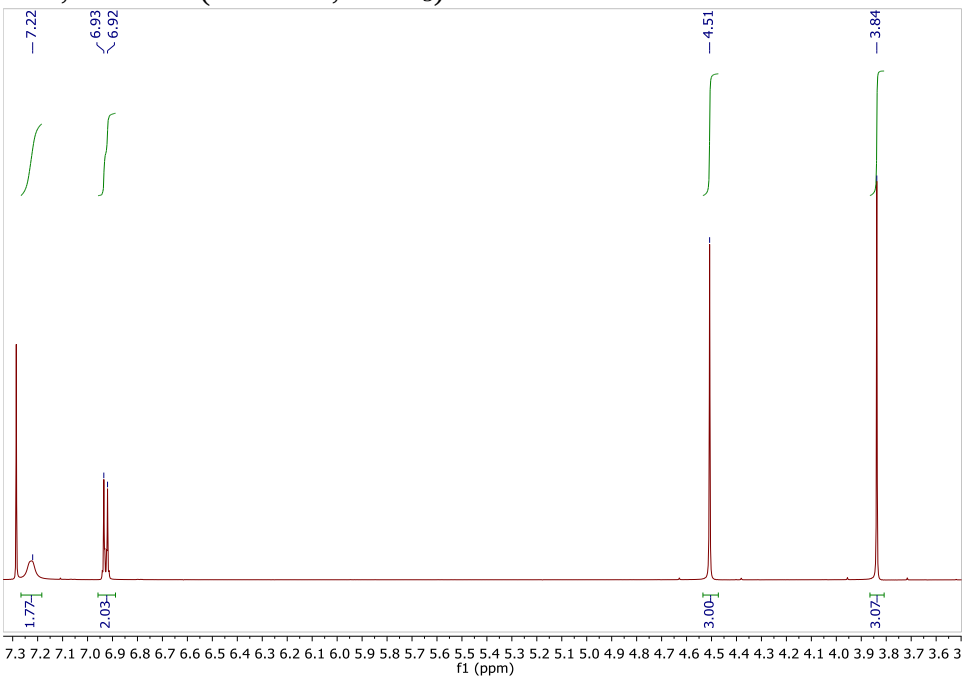


^{13}C NMR (151 MHz, d_6-DMSO)

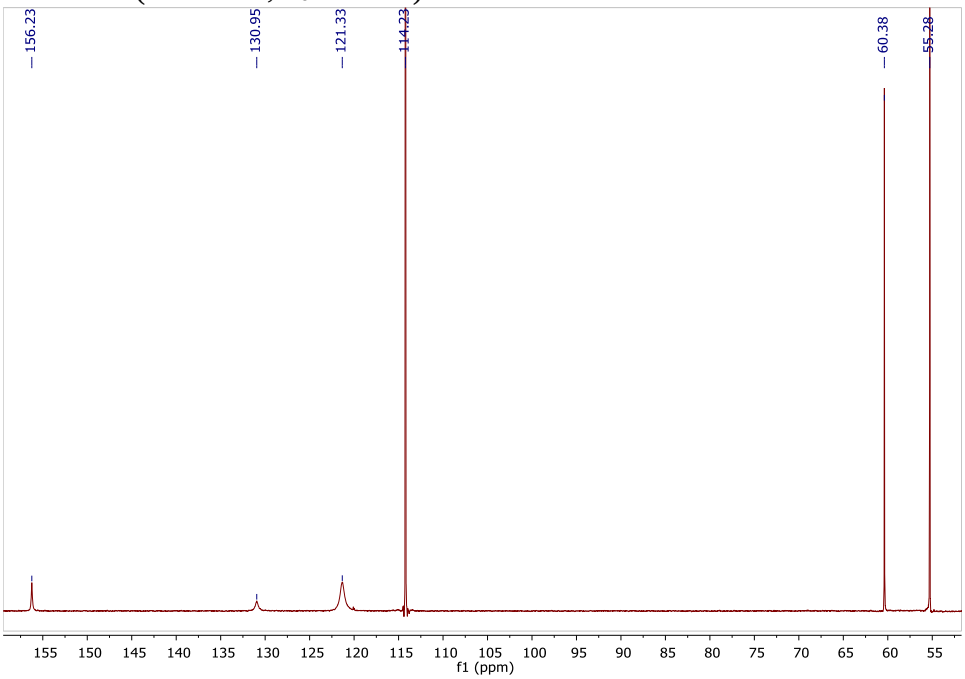


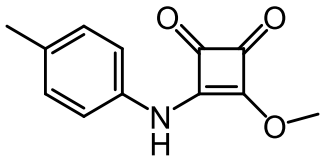


2.8d, ^{1}H NMR (600 MHz, $CDCl_3$)

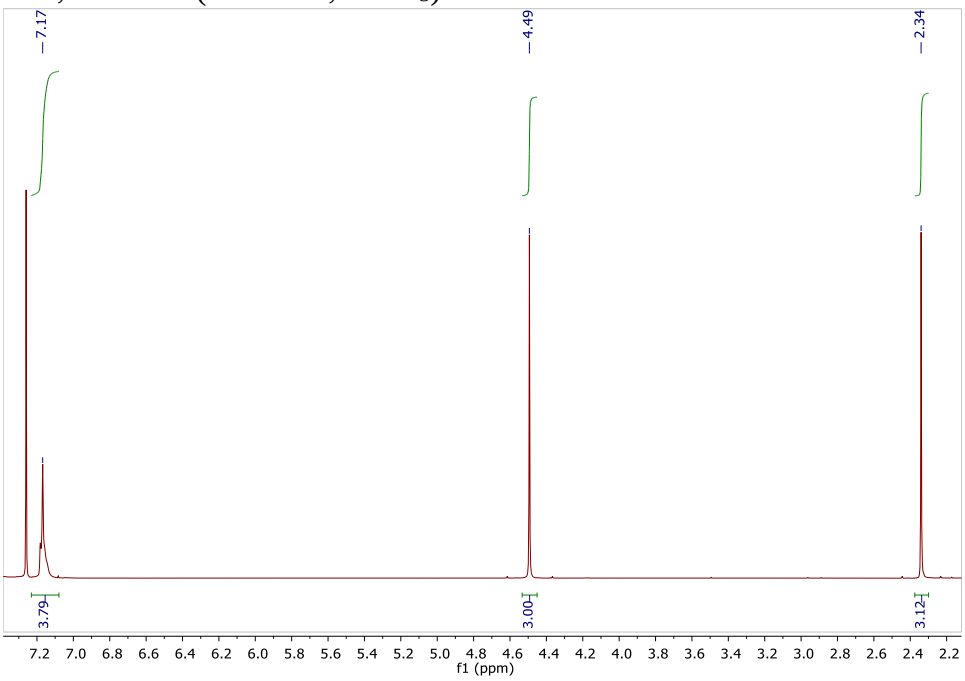


^{13}C NMR (151 MHz, d_6-DMSO)

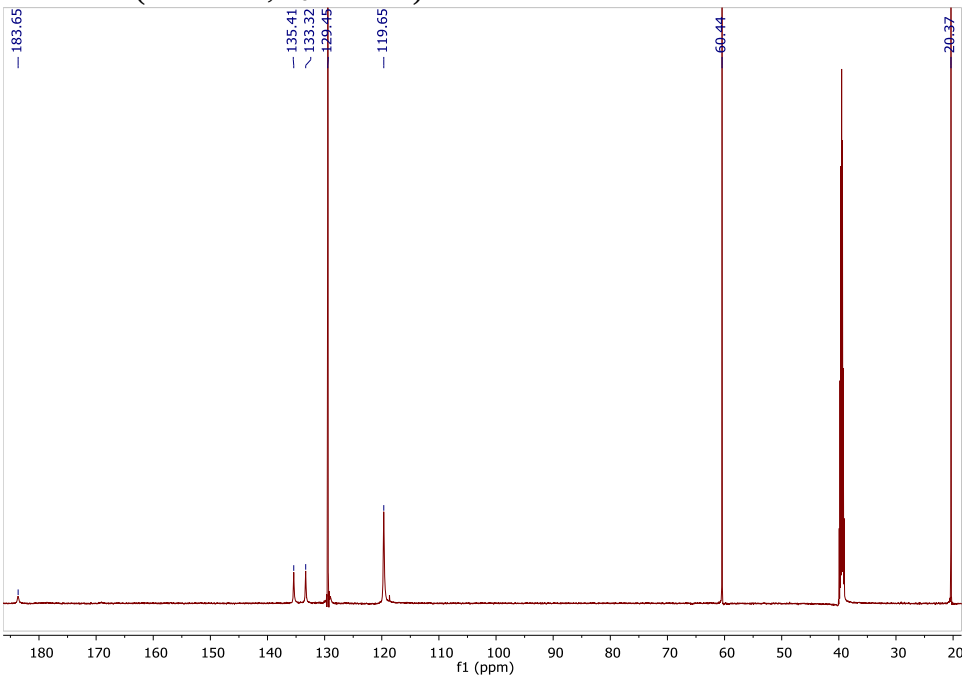


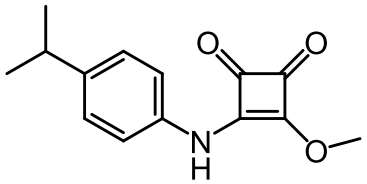


2.8e, ^{1}H NMR (600 MHz, $CDCl_3$)

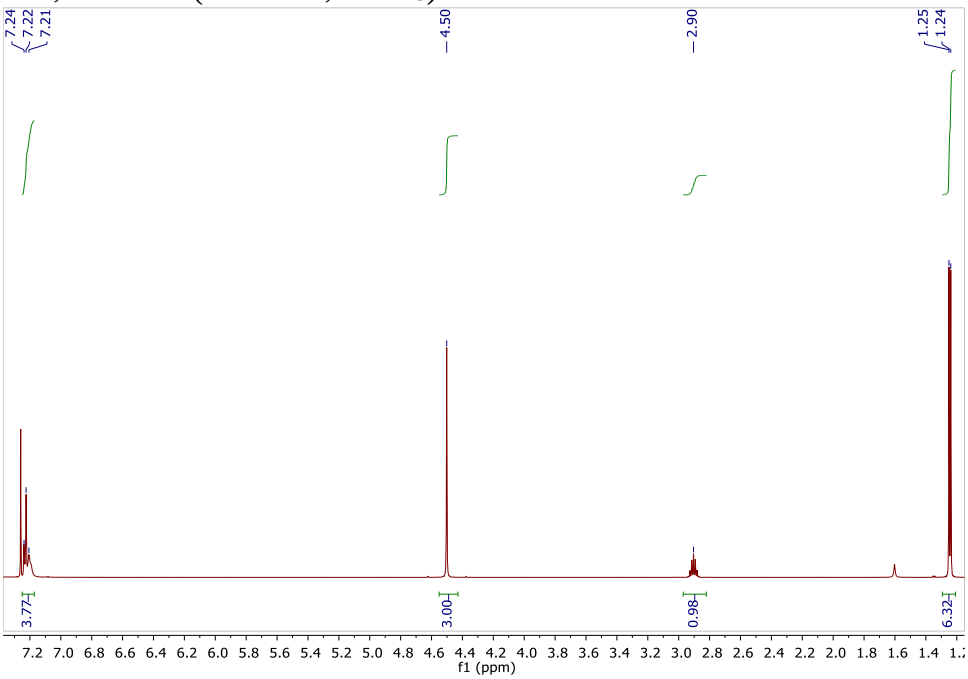


^{13}C NMR (151 MHz, d_6-DMSO)

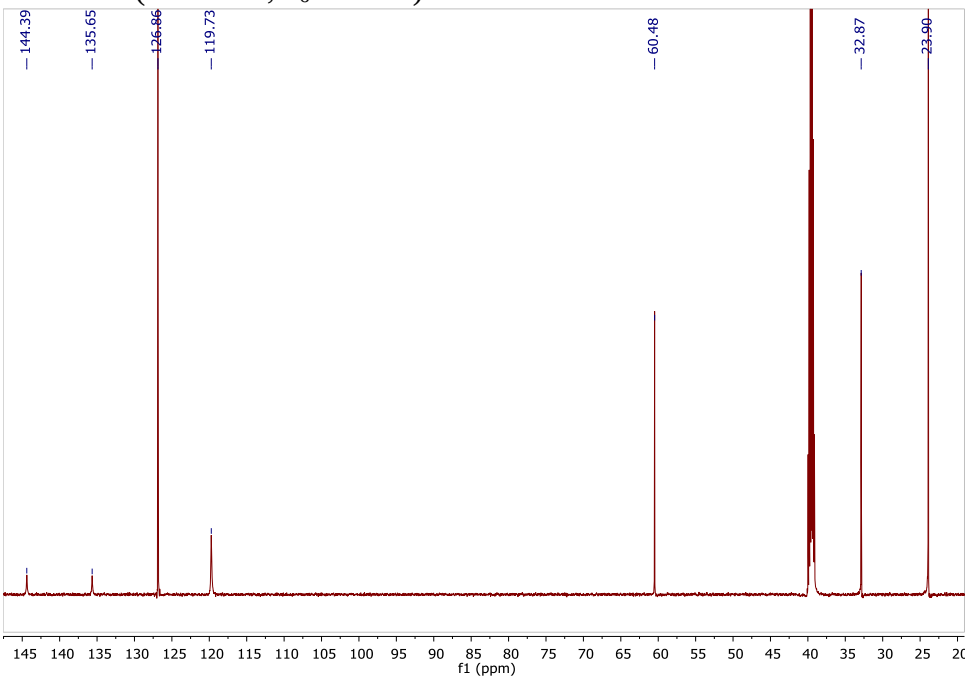


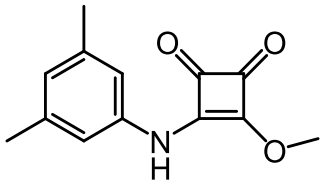


2.8f, ^{1}H NMR (600 MHz, $CDCl_3$)

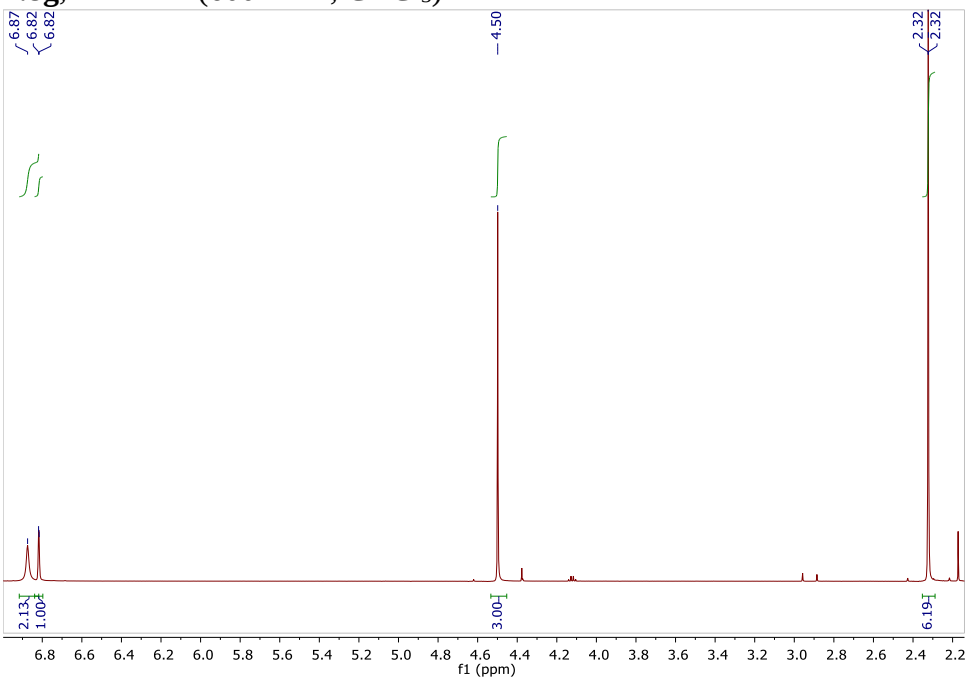


^{13}C NMR (151 MHz, d_6-DMSO)

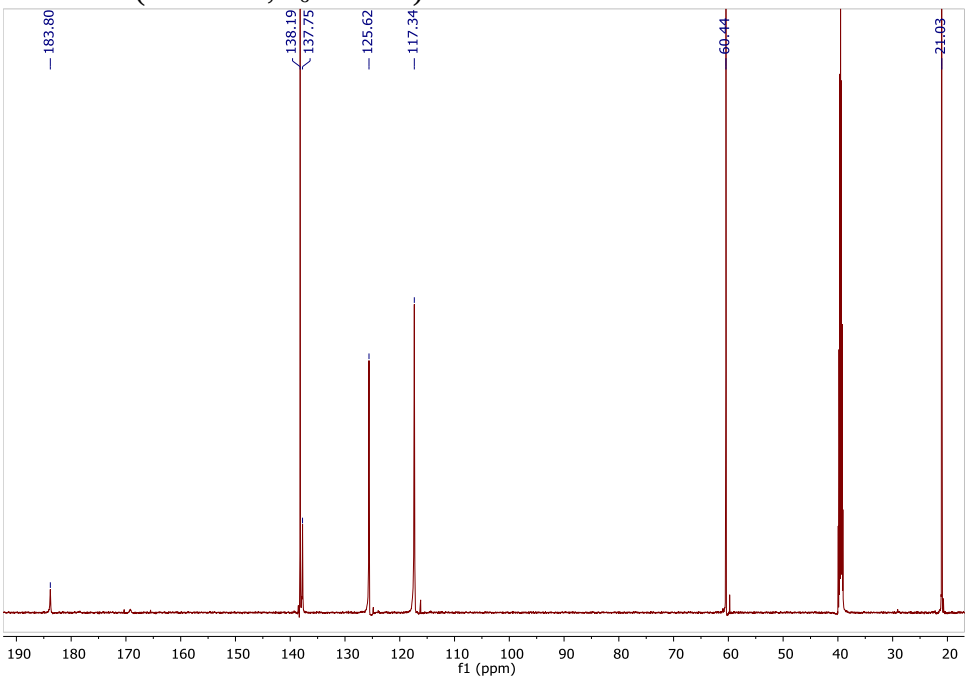


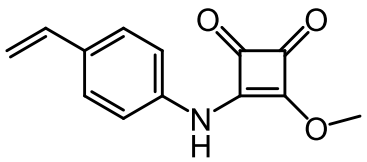


2.8g, ^{1}H NMR (600 MHz, $CDCl_3$)

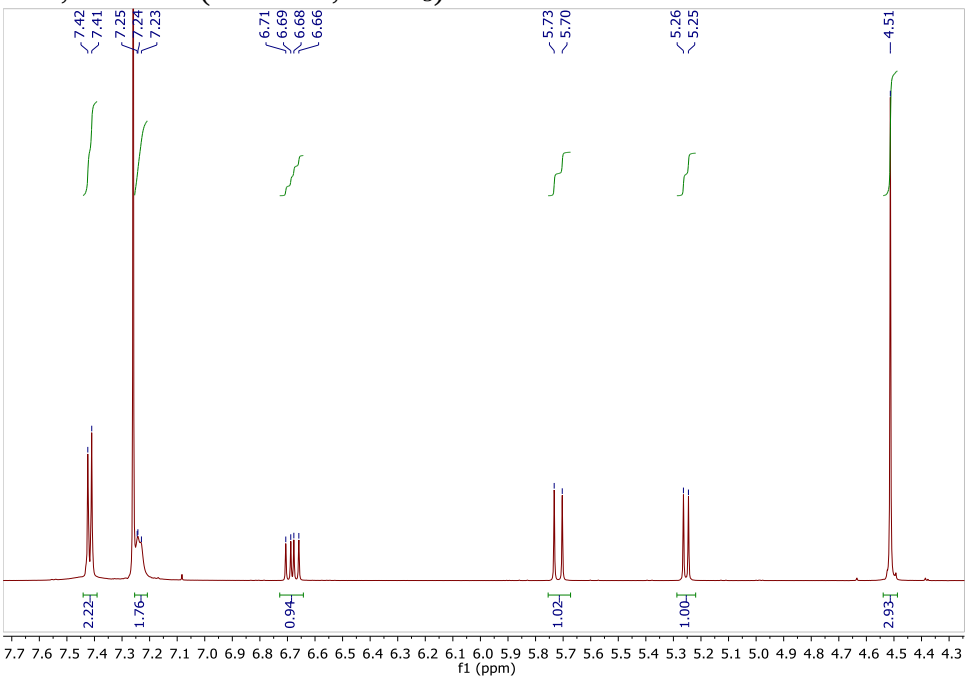


^{13}C NMR (151 MHz, d_6-DMSO)

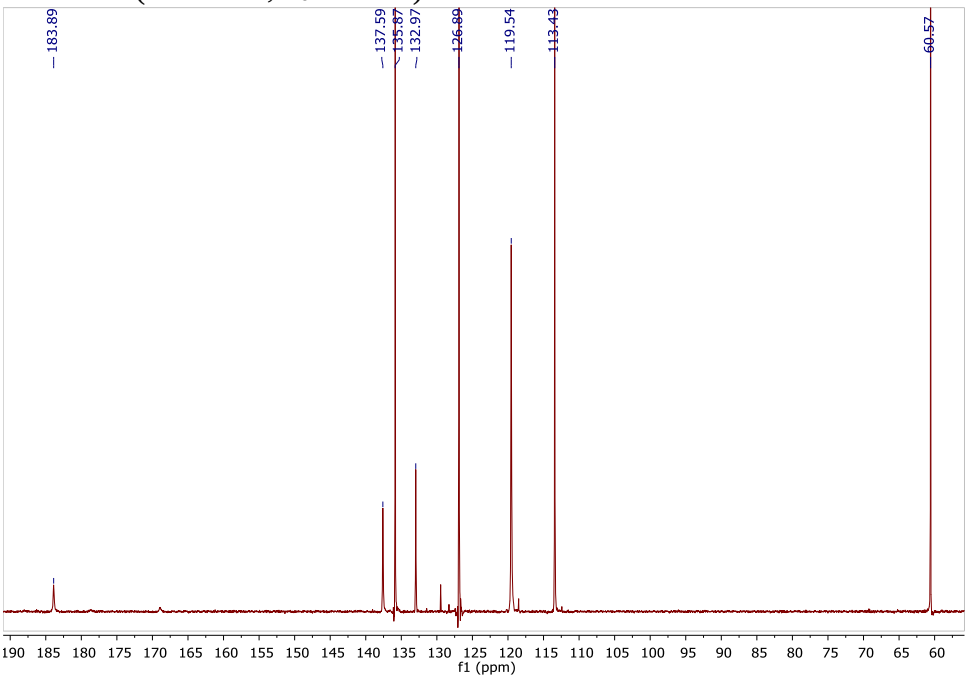


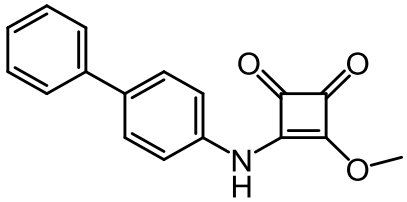


2.8h, ^{1}H NMR (600 MHz, $CDCl_3$)

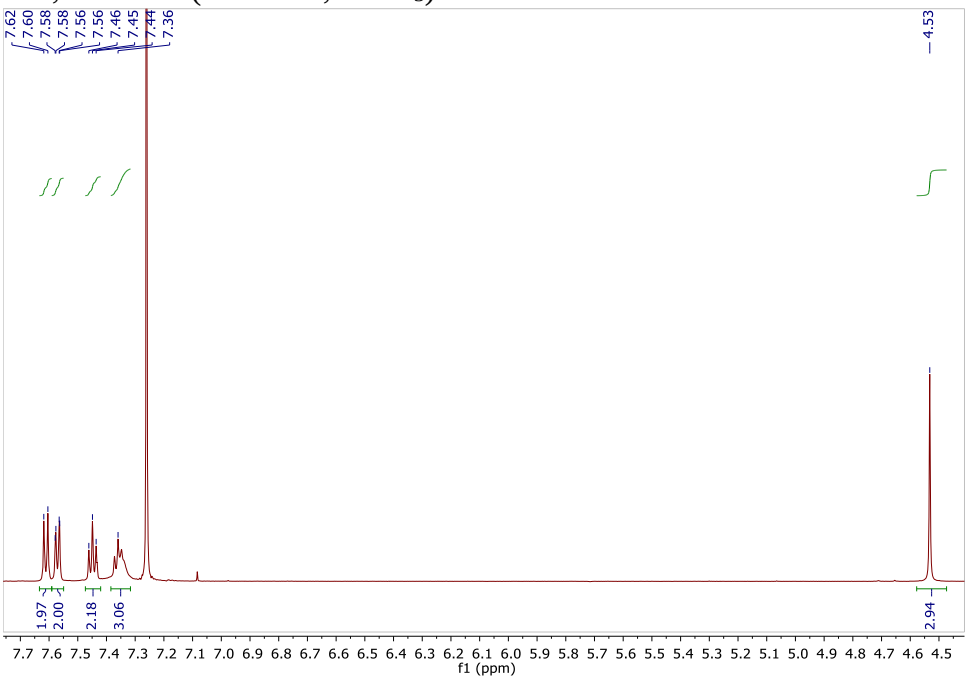


^{13}C NMR (151 MHz, d_6-DMSO)

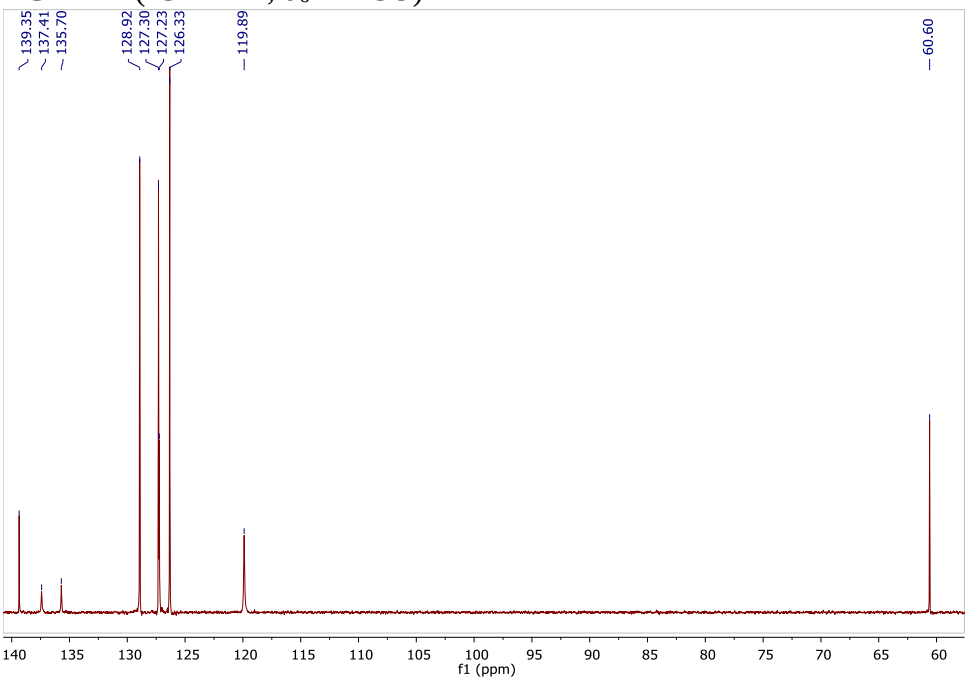


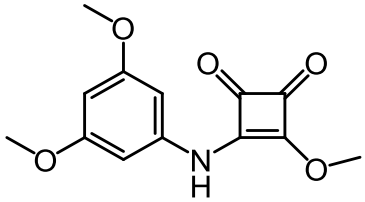


2.8i, ^{1}H NMR (600 MHz, $CDCl_3$)

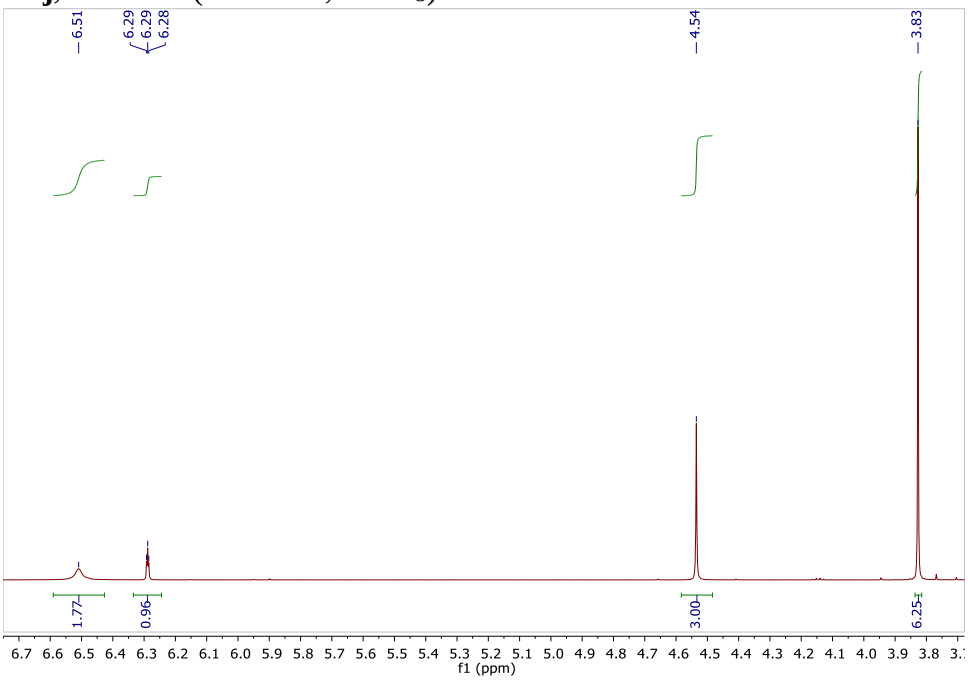


^{13}C NMR (151 MHz, d_6-DMSO)

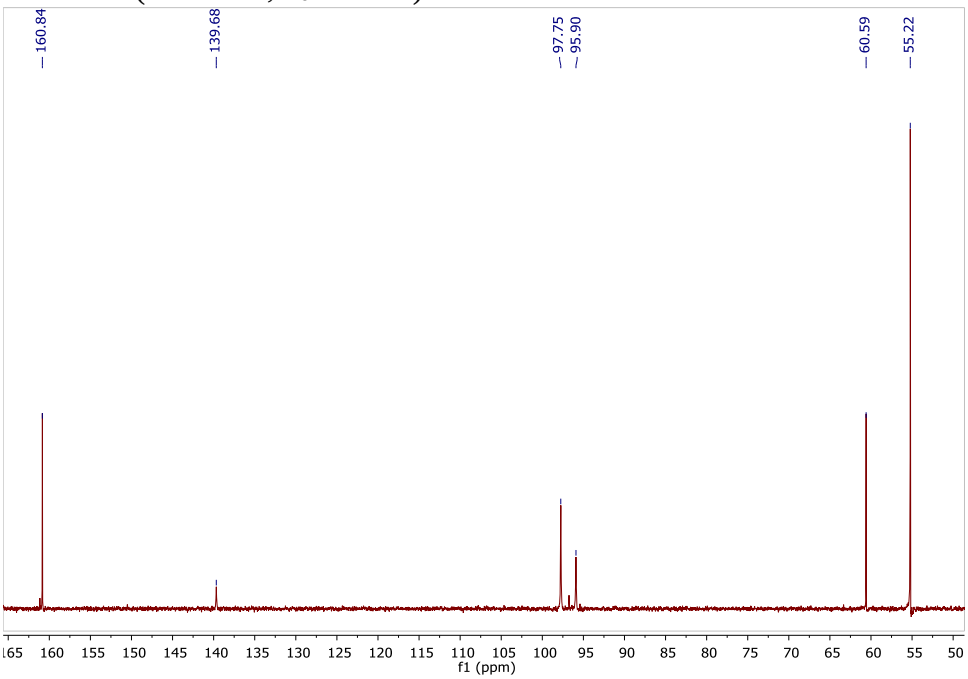


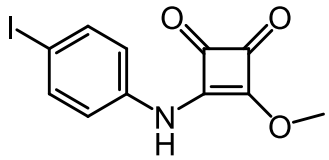


2.8j, ^{1}H NMR (600 MHz, $CDCl_3$)

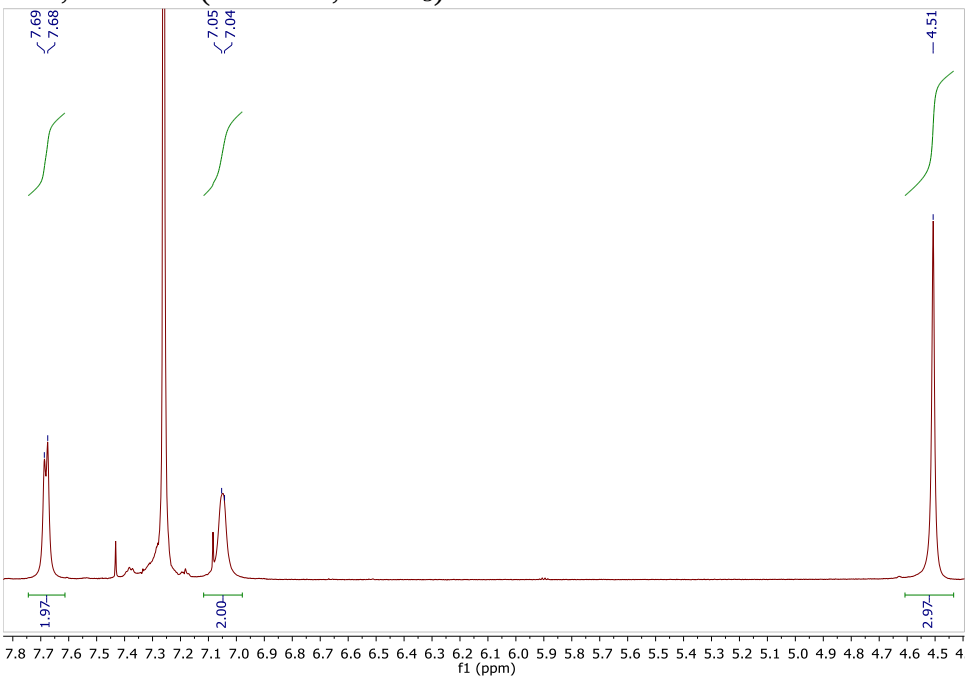


^{13}C NMR (151 MHz, d_6-DMSO)

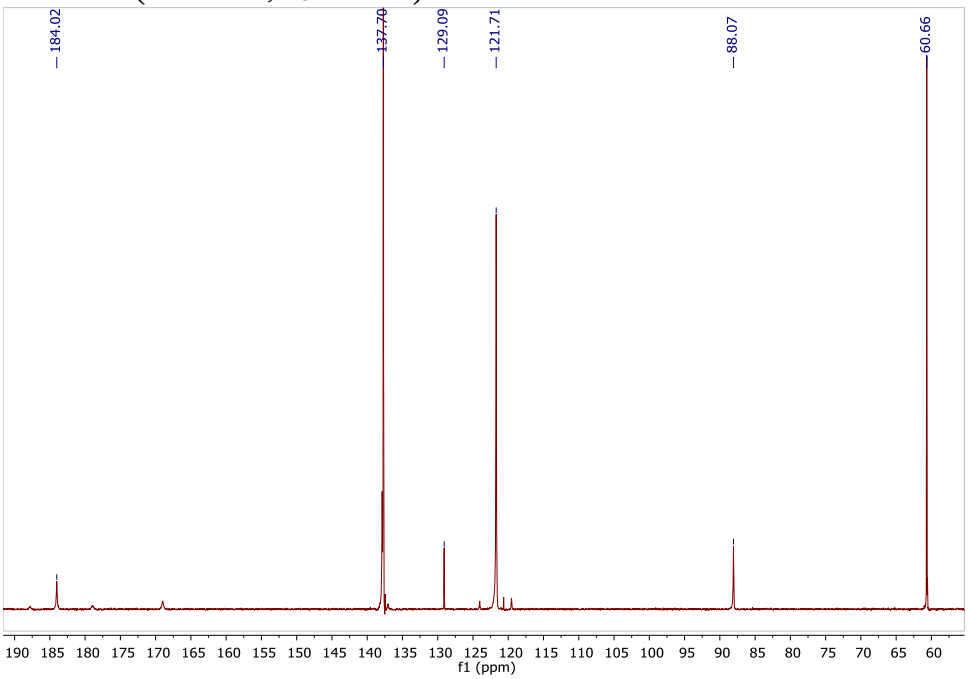


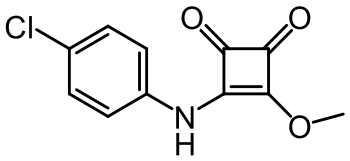


2.8k, ^{1}H NMR (600 MHz, $CDCl_3$)

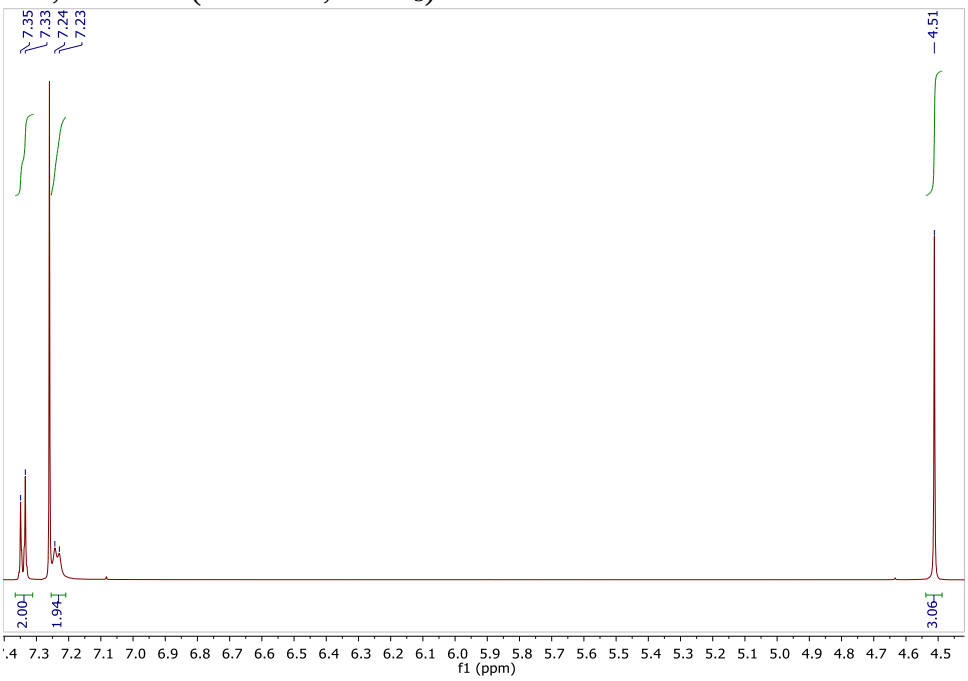


^{13}C NMR (151 MHz, d_6-DMSO)

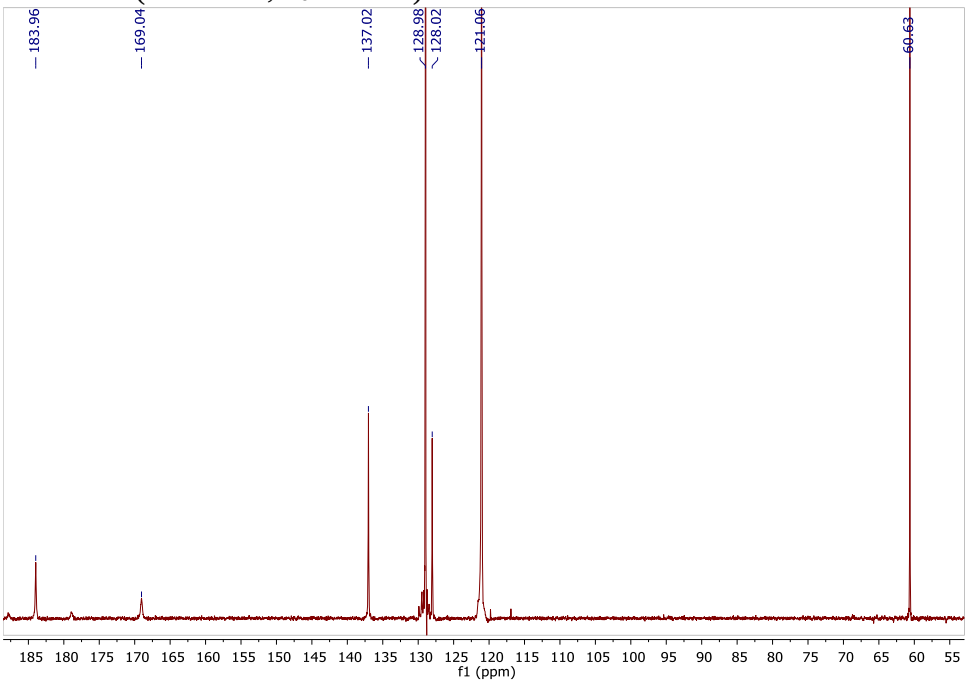


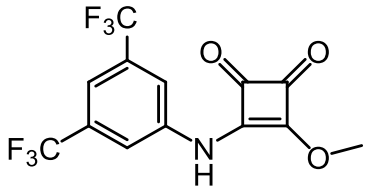


2.8l, ^{1}H NMR (600 MHz, $CDCl_3$)

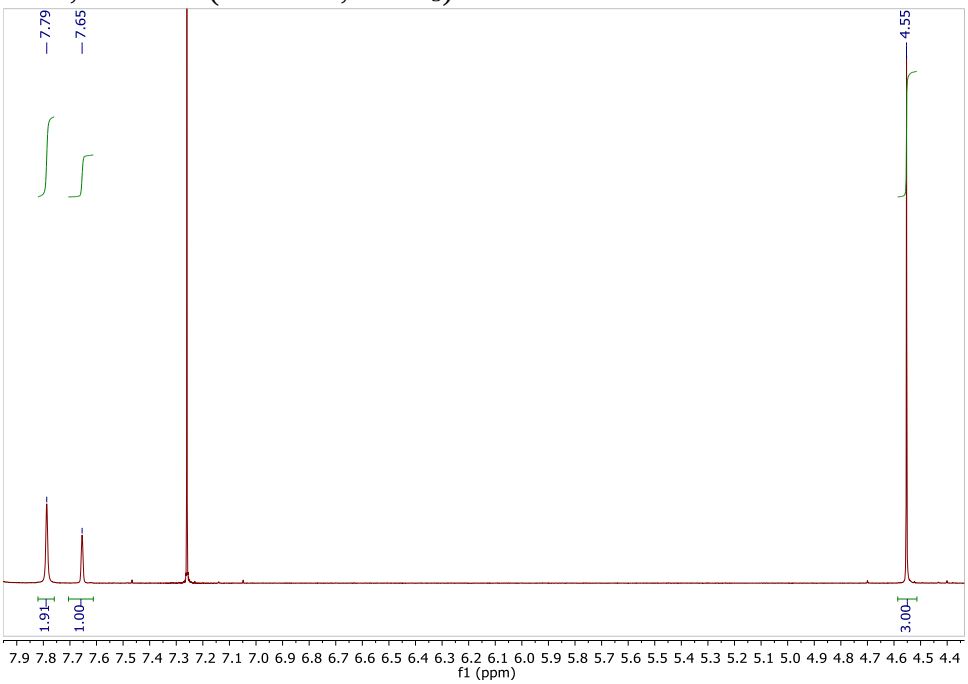


^{13}C NMR (151 MHz, d_6-DMSO)

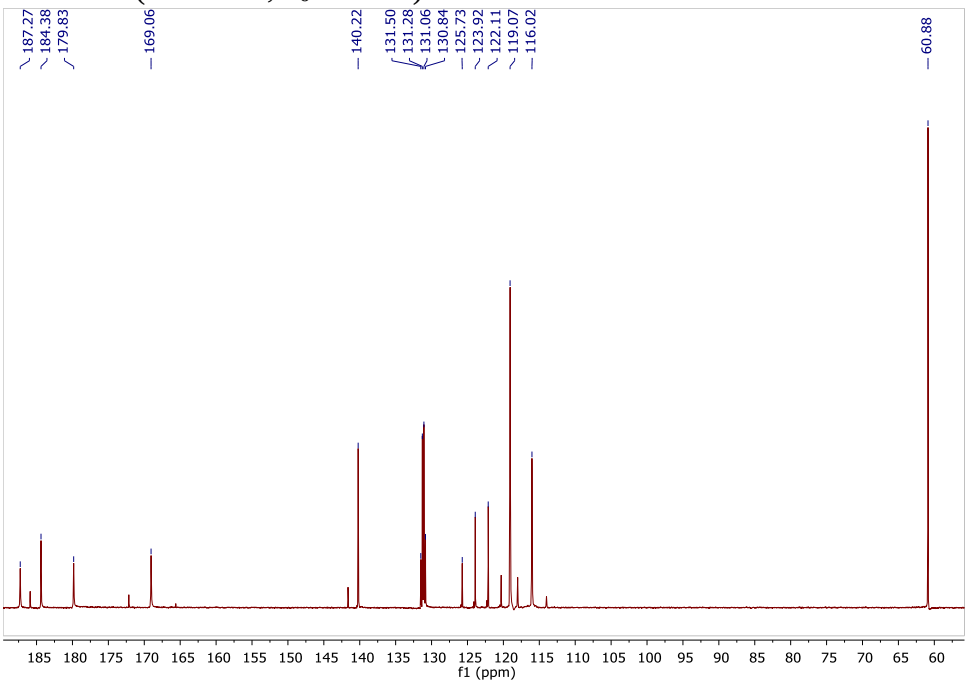


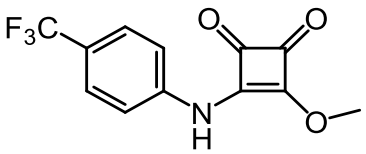


2.8m, ^{1}H NMR (500 MHz, $CDCl_3$)

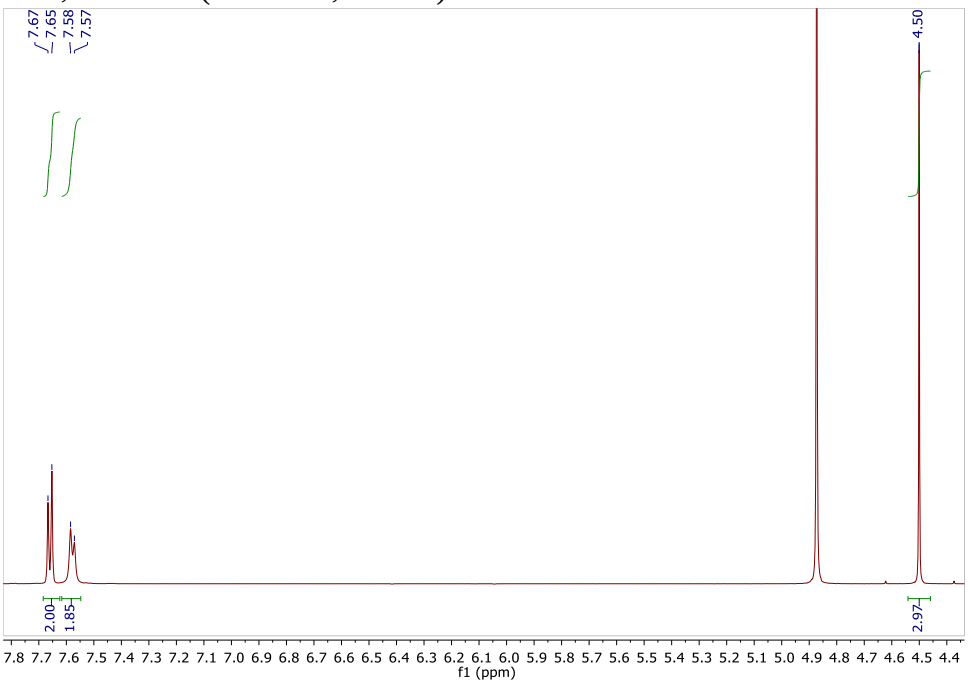


^{13}C NMR (151 MHz, d_6-DMSO)

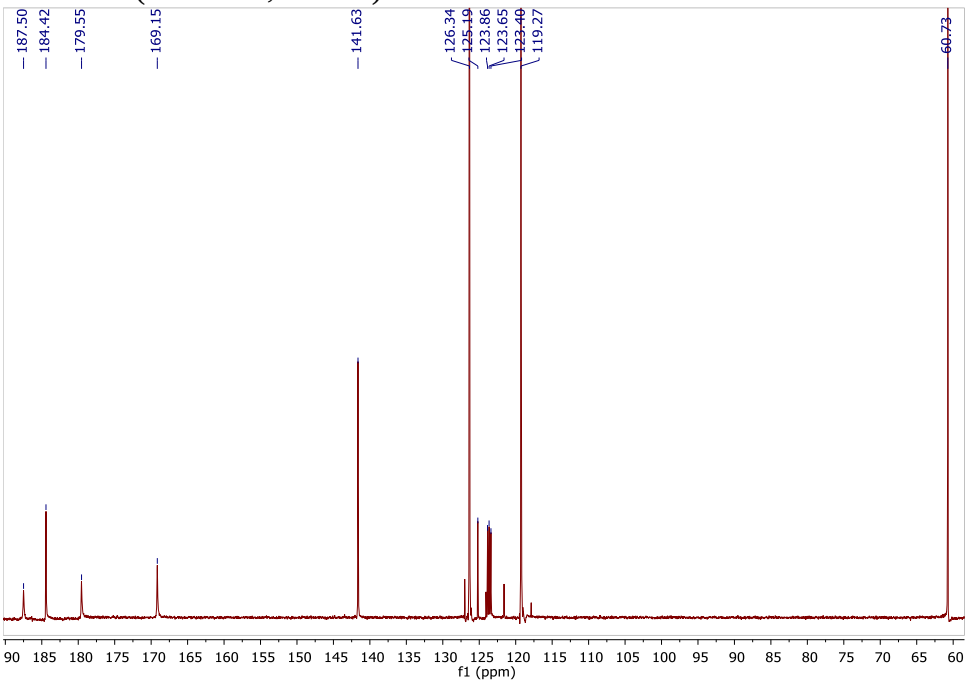


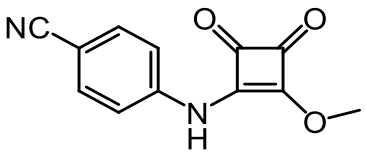


2.8n, ^{1}H NMR (600 MHz, MeOD)

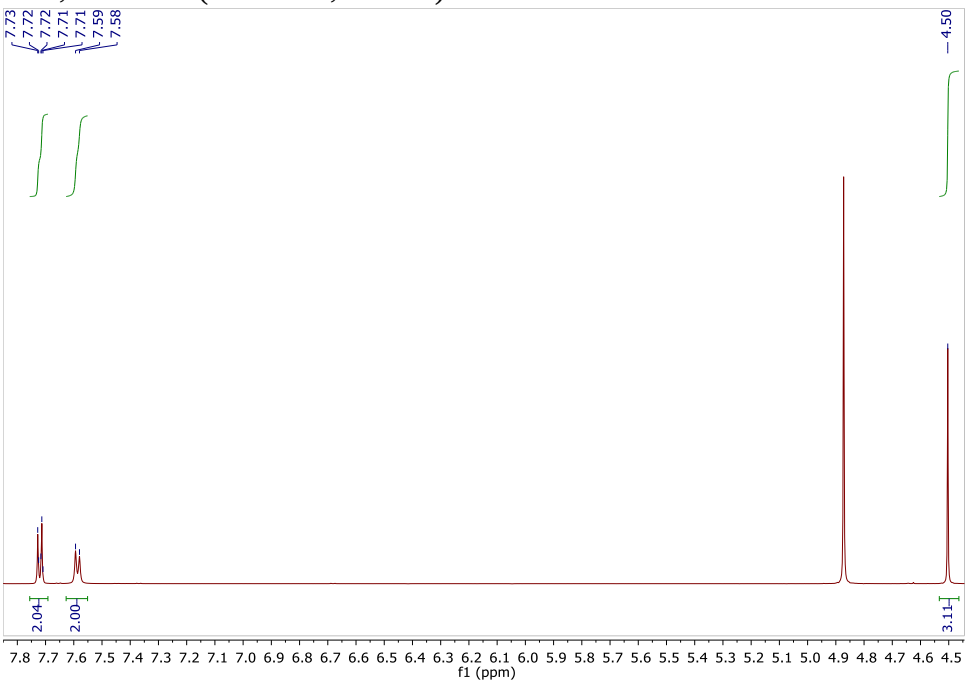


^{13}C NMR (151 MHz, MeOD)

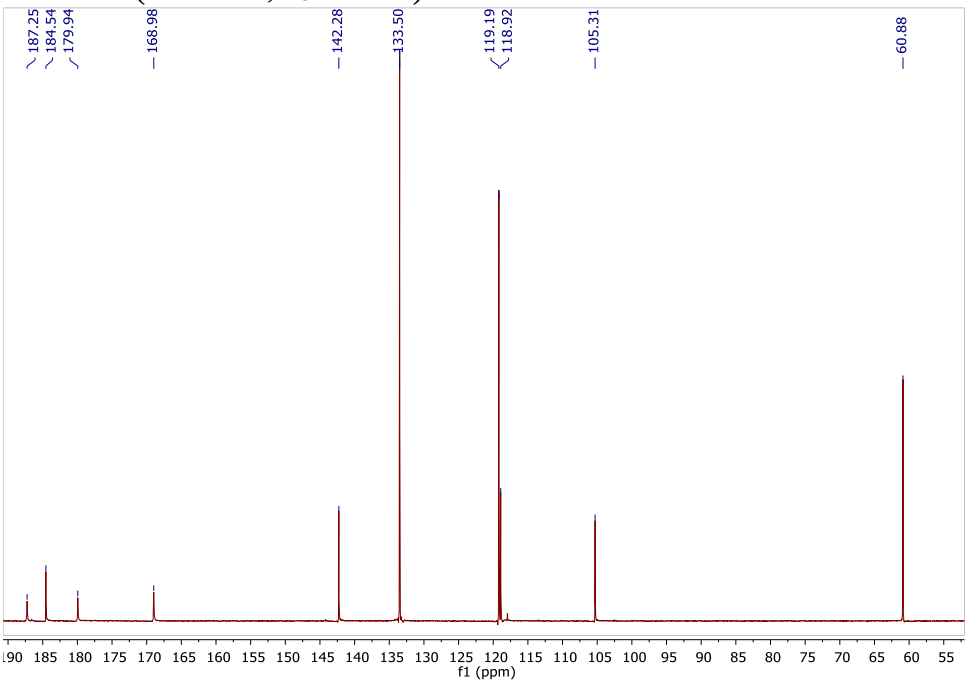


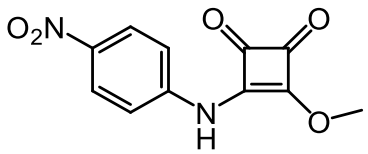


2.8o, ^{1}H NMR (600 MHz, MeOD)

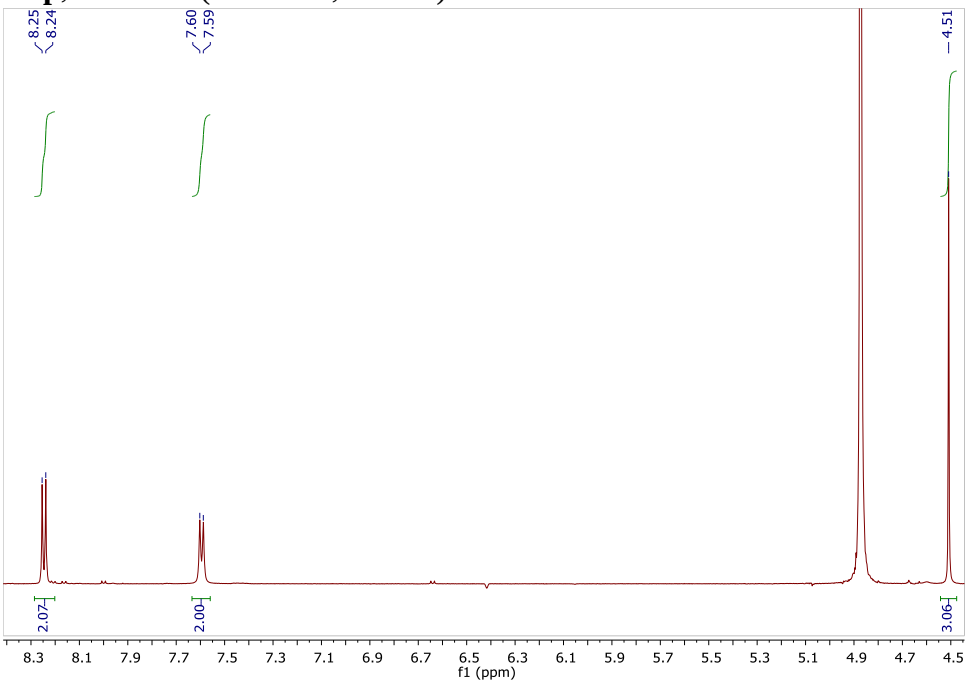


^{13}C NMR (151 MHz, d_6-DMSO)

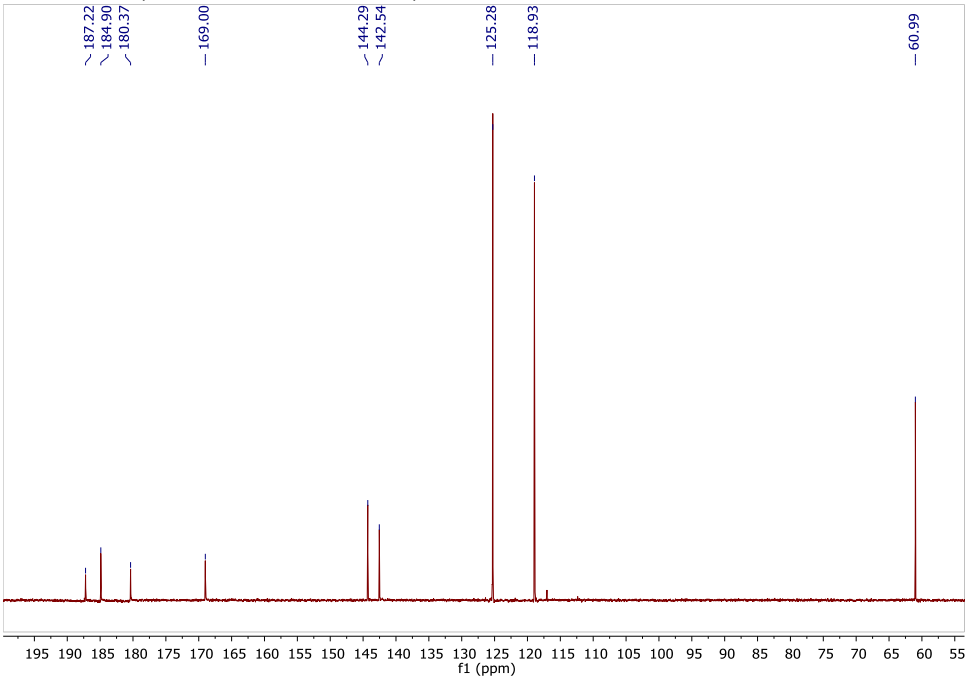


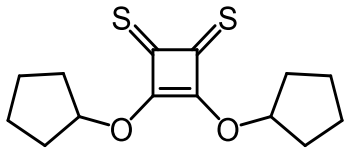


2.8p, ^{1}H NMR (600 MHz, MeOD)

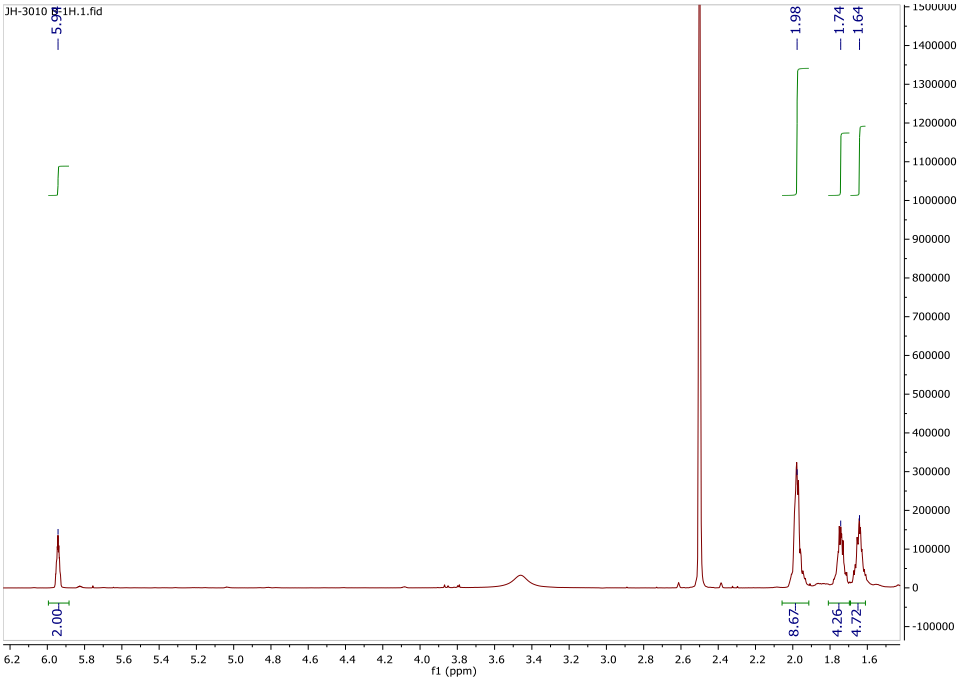


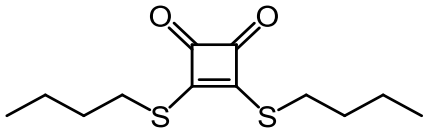
^{13}C NMR (151 MHz, d_6-DMSO)



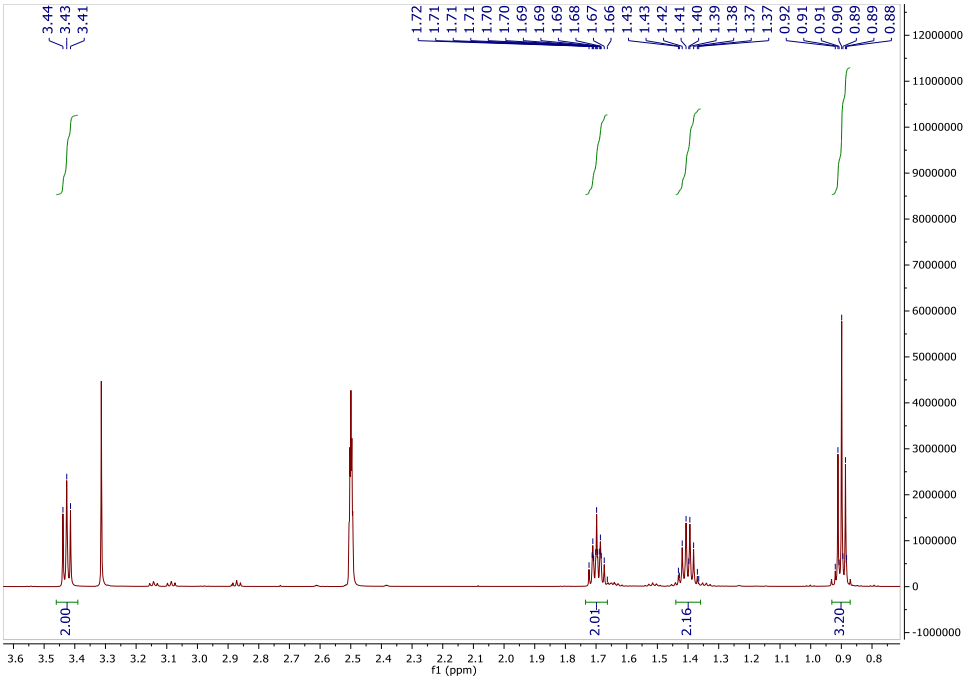


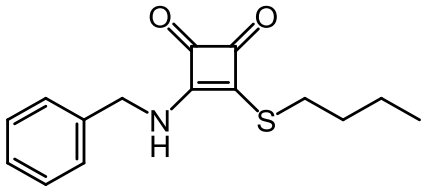
2.10, ^{1}H NMR (600 MHz,



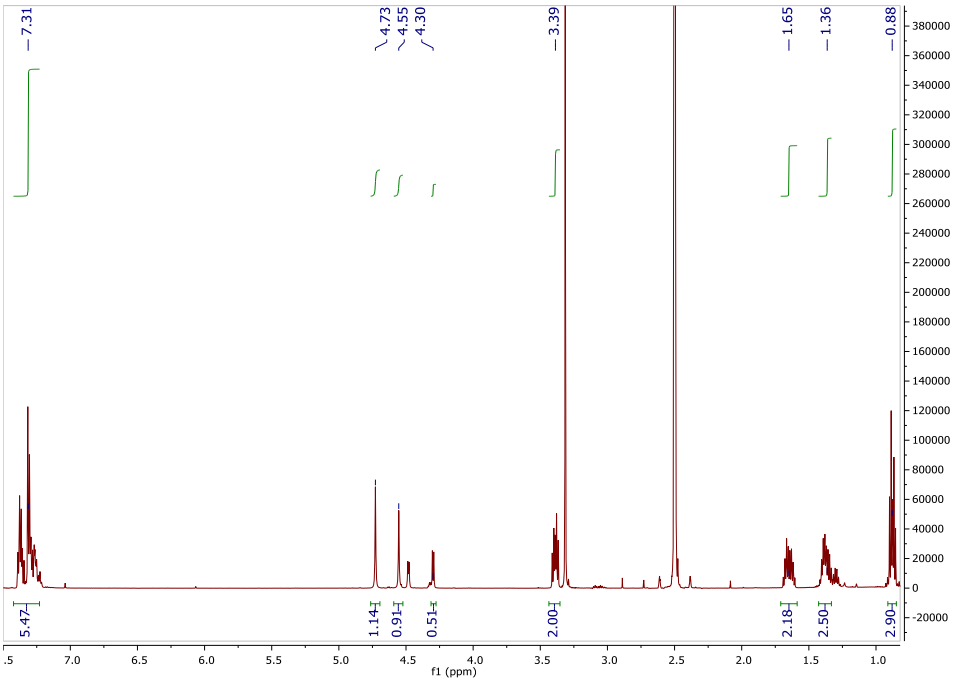


2.11, ^{1}H NMR (600 MHz,





2.12, ^{1}H NMR (600 MHz,



Chapter 3: Designing Covalent Squaric Ester Inhibitors of Galactofuranosyltransferase 2

Contributions:
GlfT2 expression and purification was performed by Dr. Alex M. Justen
Synthesis and biological studies were performed by Jordan Ho

3.1 Abstract

Small molecules that covalently modify proteins for activity-based protein profiling are a common tool to study biological systems. Modular in design, these covalent molecules typically contain an electrophilic warhead used to bind a target nucleophilic residue along a protein's surface. To date, efforts have yielded electrophiles for many nucleophilic amino acids, yet the current electrophiles designed for lysines are not as robust as those for other residues such as cysteine. The ability to target lysines would allow access to a larger pool of protein targets using activity-based protein profiling due to its higher prevalence in the proteome. However, its low nucleophilicity coupled with this higher abundance make the design of lysine specific probes challenging. Many current lysine directed electrophiles are unstable in aqueous media or too reactive to be selective to be used beyond *in vitro* experiments and global proteomic studies. Squaric esters are an amine reactive moiety that reacts much slower than traditional amine directed electrophiles, which could be advantageous for attenuating off-target labeling. Herein we demonstrate that the mild squarates can be used to design lysine specific probes. We show that squarates can specifically and irreversibly bind galactofuranosyltransferase 2, a glycosyltransferase found in Mycobacterium tuberculosis, at key functional sites when appended to a scaffold that mimics the enzyme's natural substrate. When transferred onto another scaffold that GlfT2 recognizes however, there was no observable binding interaction, highlighting the intricacy and precision that is required when rationally designing molecular probes for biology.

3.2 Introduction

The use of moluecular probes to covalently bind proteins has long been a strategy in developing potent therapeutics as well as tools for uncovering enzyme structure and function.[1,22,83–86] These covalent probes have been particularly useful in activity-based protein profiling (ABPP) to study enzymes *in vivo.*[73,87–90] ABPP probes are small molecules that are designed to covalently bind an enzyme only when it is active. Because the reactivity of ABPP probes is activity-based, they can used to compare expression levels to levels of active enzyme *in situ*, study the effects of post-translational and other epigenetic modifications, and pull down enzymes of similar function for proteomic studies.[3,91–93] As a biochemical tool, these small probes also offer a less perturbative way for directly studying an enzyme than techniques like mutagenesis.

A key element of these modular ABPP probes is an electrophile that can form a covalent bond to a specific amino acid amidst the hundreds of other amino acids that may be present across the protein surface. Initial probes used thiol reactive electrophiles to study catalytic cysteines, as its nucleophilicity often aids enzymatic function and its low prevalence decreases the chance for off-target reactivity.[8] The low abundance of cysteine however limits the potential pool of targets in the proteome for ABPP, so additional electrophiles to target other nucleophilic amino acids have been explored.[74,94–96] Lysine is a tractable target as it is more prevalent in the proteome than cysteine and can expand the scope for ABPP studies. However, its higher abundance combined with its low reactivity in aqueous environments makes lysine a more challenging nucleophile to design specific probes for. Recent efforts have found electrophiles that react with lysines, but many tend to be highly reactive and lack specificity or are prone to hydrolysis, limiting their use beyond *in vitro* systems.[97] Shannon *et al.* have shown that

dichlorotriazines can specifically target lysine residues in aqueous media, but are too reactive to use for specific labeling of lysine residues.[53]

Squaric esters could be a viable alternative for developing ABPP probes as their mild reactivity allows for easy installation onto a small molecule scaffold. Its small atomic footprint also minimizes substantially perturbing the non-covalent binding interactions of the probe. Compared to current electrophiles employed in lysine ABPP, squaric esters react much slower and are non-immunogenic.[56,98,99] Squaric esters have already been implemented as linkers in biological contexts,[17,65] so we sought to evaluate their potential as electrophiles for ABPP probes. After characterizing the specificity of squaric esters towards amines and thiols as well as their reactivity in aqueous media, we applied them in designing an ABPP probe to study the enzyme galactofuranosyltransferase 2 (GlfT2).

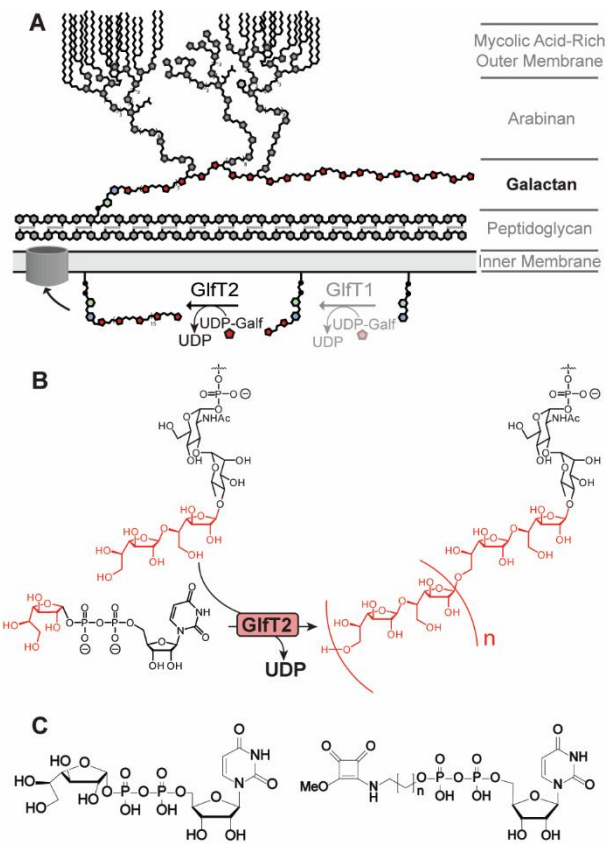


Figure 3.1. (A) Diagram of *M. tuberculosis* cell wall. (B) GlfT2 polymerizes galactan using galactofuranose from UDP-Galf in an alternating β-(1,5) β-(1,6) pattern. (C) Chemical structure of UDP-Galf and squaric ester-amide analog.

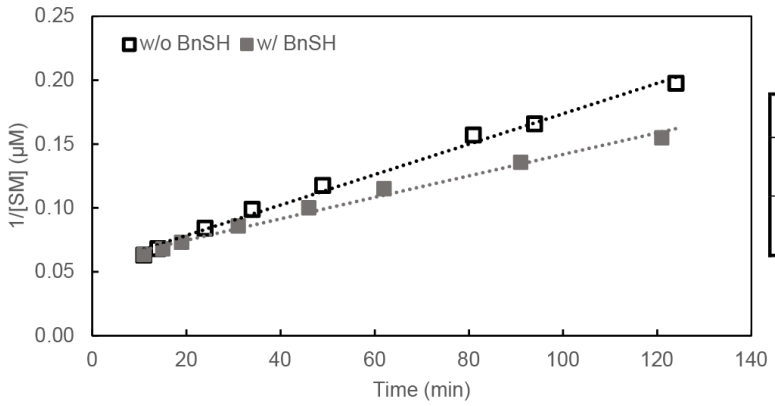
GlfT2 is a glycosyltransferase found in various bacteria including *M. tuberculosis*, the pathogen responsible for tuberculosis, and is vital for the biosynthesis of an essential layer of the bacteria's cell wall known as the galactan (Figure 3.1A).[100,101] GlfT2 transfers galactofuranose (gal*f*) from UDP-galactofuranose (UDP-gal*f*) to a growing polysaccharide chain of gal*f* residues

in alternating β-(1,5) β-(1,6) linkages (Figure 3.1B) to a consistent length of about 40 gal*f* units before it is then further elaborated by other downstream enzymes.[102,103] The resulting galactan is one of many different layers in mycobacterial cell walls that contribute to its tenacity and antibiotic resistance, and while previous mycobacterial studies have shown the galactan is necessary for bacterial survival, there is still currently no antibiotic that targets galactan biosynthesis.[104,105] More importantly, the active site of GlfT2 contains a lysine known to stabilize the pyrophosphate in UDP-gal*f*. Since the development of a GlfT2-specific ABPP probe could eventually lead to novel therapeutics, it is a prime candidate to explore selective lysine modification.

We synthesized substrate analogs of UDP-gal*f* containing squaric esters (Figure 3.1C) that were able to covalently modify GlfT2 specifically and inhibit function *in vitro*. Analysis of treated GlfT2 by mass spectrometry revealed a hydrophobic region on the opposite face of GlfT2 relative to the active site, which may serve as a proposed allosteric site used to control galactan chain length.[106]

3.3 Results and Discussion

Squarates are amine specific and react in aqueous media. We first obtained reaction rates of model squarates with benzylamine to serve as a baseline for comparison. This simplicity allowed us to measure distinct changes in the chemical shift of the benzylamine's methylene protons by ^{1}H NMR. Previous results have shown that the squaric ester moiety reacted significantly slower than other reported amine reactive handles. Furthermore, squarates do not to appear to react with thiols. When benzylmercaptan was added to compete with benzylamine only the squaramide product was observed; however, the addition of a thiol did slow the rate slightly (Figure 3.2). While arguments that the squarate's stability towards thiols is due to the formation of a reactive



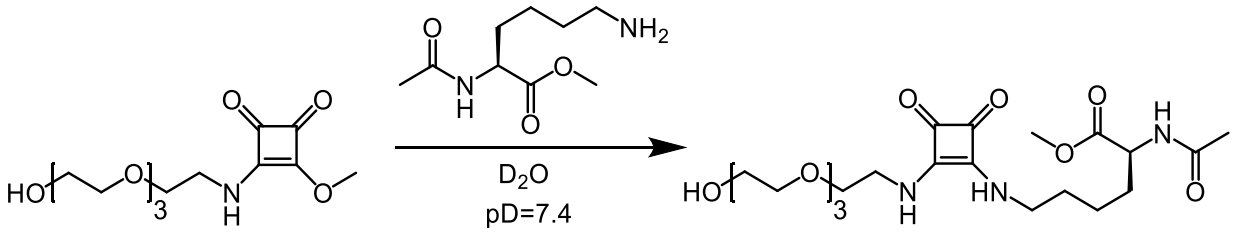
	k ($\mu M^{-1} \cdot sec^{-1}$)
w/o BnSH	1.19
w/ BnSH	0.84

Figure 3.2. Reaction rates of benzylamine with phenylsquaric ester-amide with and without the presence of benzylmercaptan.

vinylogous thioester intermediate, these data could also have been due to the change in pH from the addition of the thiol as well as additional interactions between π-systems, perturbing the electronics of the squarate's electrophilic core.

Next, for squarates to be a viable alternative as a biological electrophile, we evaluated whether squarates could still react with an amine under aqueous conditions. We synthesized a water soluble squarate, **3.1**, and monitored its reaction with N-acetyl-L-lysine methyl ester in a phosphate buffered solution of D_2O (Scheme 3.1). While there is a noticeable decrease in reaction rate, squarate **3.1** was still able to bind to the protected lysine and at a rate within an order of magnitude to the other reactions monitored above. The ability to still react with amines in aqueous media, coupled with its inertness towards more nucleophilic thiols highlight how the

Scheme 3.1. Reaction of squaramide **3.1** with N-acetyl lysine methyl ester.



3.1 **3.2**

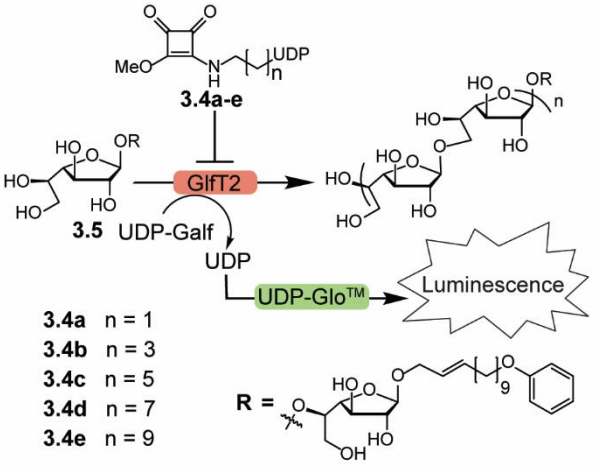


Figure 3.3. Diagram of GlfT2 activity assay. GlfT2 production of UDP during polymerization of **3.5** can be quantified using UDP-Glo™ to assess enzyme activity.[107] GlfT2 engagement with squarate analogs **3.4a-e** can be observed through inhibition of this UDP production.

unique reactivity of squarates lend themselves to be ideal electrophiles to probe lysines in more biological contexts.

Squarate Substrate Analogs Covalently Bind GlfT2. Next, to highlight the use of squarates in covalent probe design, we synthesized squaric ester analogs of UDP-gal*f* and evaluated their efficacy as covalent inhibitors. GlfT2 contains an active site lysine, K369, that could be a target for a potential ABPP probe. We generated a small library of UDP-squarates linked by an alkyl chain of varied length based off previous routes[108] and tested their inhibitory potential on a model reaction using *M. tuberculosis* GlfT2 (Figure 3.3). GlfT2 can bind to the UDP in UDP-gal*f* and transfer the galactofuranose onto a synthetic liposaccharide **3.5** modeled after the natural acceptor substrate and processively elongate it to a polysaccharide chain with a consistent average length of about +20 galactofuranose residues.[106,109]

By using a luminescence assay that quantifies GlfT2 activity via UDP production,[107] we first investigated if addition of these UDP-squarate analogs would have a notable effect on GlfT2 activity *in vitro* (Figure 3.3). Control experiments indicated preincubation of the UDP-squarates with GlfT2 was necessary for inhibition, suggesting that the covalent bond formation is not fast enough to compete with the non-covalent binding of UDP. We found that at a concentration of 100μM and preincubation time of 3 hours, the UDP-squarates with longer alkyl chains were

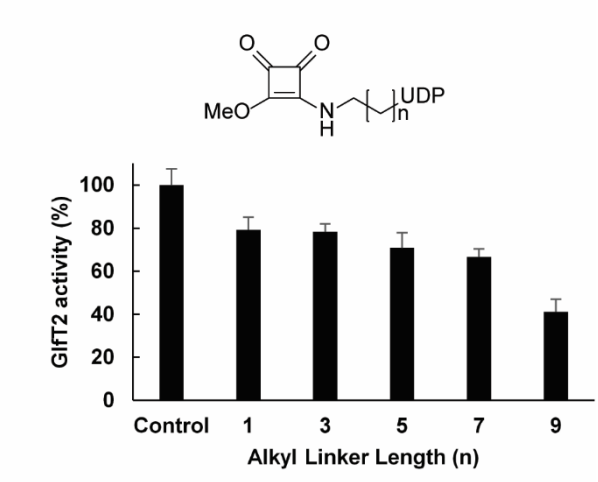


Figure 3.4. Effect of alkyl linker length on substrate analog interaction with GlfT2. Samples of GlfT2 were treated with compounds **3.4a-e** at 100μM for 3 hours before addition of UDP-gal*f*.

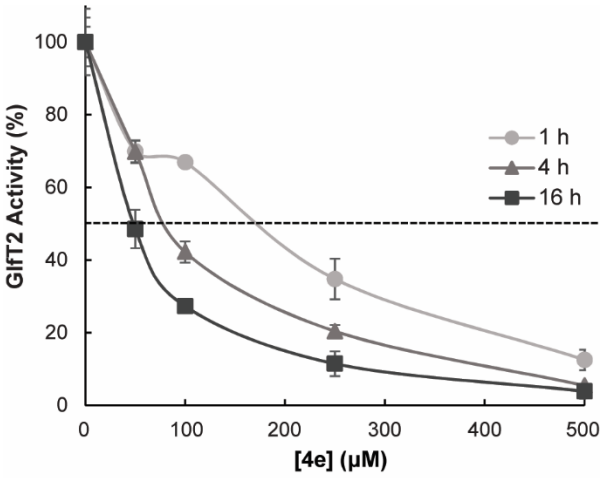


Figure 3.5. Effect of preincubation time of **3.4e** at a concentration of 100μM on GlfT2 inhibition. Y-axis depicts relative activity to no compound control in each set of conditions. Experiments were performed in triplicate. Error bars represent the standard deviation, *n*=3.

better able to inhibit GlfT2 (Figure 3.4).

Because preincubation of the probes with GlfT2 was required, we wanted to determine the time-dependence on inhibitio, a marker of covalent inhibition. As preincubation time increased, GlfT2 activity decreased with **3.4e** showing a decrease of over 40% in GlfT2 activity for a sixteen-hour preincubation compared to just one hour (Figure 3.5). This time dependence supports our hypothesis that these squarate compounds can inhibit GlfT2 through a covalent modification. Furthermore, GlfT2 inhibition by UDP-squarate **3.4e** was stable after dialysis of treated GlfT2, in contrast to uridine monophosphate (UMP) inhibition which only acts as a competitive binder to GlfT2 (Figure 3.6). This sustained inhibition by squaric ester **3.4e** further suggests that it acts via irreversible covalent inhibition.

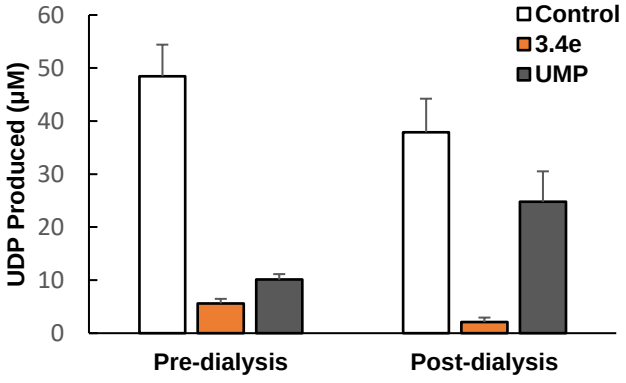


Figure 3.6. UDP production of GlfT2 after treatment with either **3.4e** or UMP. GlfT2 activity was measured before and after 4 hours of dialysis.

Squarate Analogs Specifically Target Two Conserved Sites in *Mtb* GlfT2. To confirm these UDP-squarate compounds were binding to GlfT2 not only covalently but also specifically, we analyzed GlfT2 treated with UDP-squarates using MS/MS to detect peptides modified by our UDP-squarate compounds (table 3.1). We first treated GlfT2 with squaramide **3.1**, which acted as our negative control of squarate binding lacking any recognition domain.

Analysis of trypsin digested GlfT2 treated with the PEG-squarate control showed significant modification across the surface of GlfT2 (Figure 3.7). Based on the GlfT2 structure presented by Lowary and co-workers,[110] all other unmodified lysines in the GlfT2 sequence were buried. Interestingly, covalent modification of the active site lysine, K369, was not detected. GlfT2 treated with UDP-squarate **3.4e** yielded less covalent modification. Peptides corresponding to only two lysines, K494 and K519, were found bonded to the squarate (Figure 3.7). Neither of these lysines reside in the active site; however, K494 lies just outside of the active site. Crystal structure measurements[110] estimate K494 at a ~17 Å distance from the active

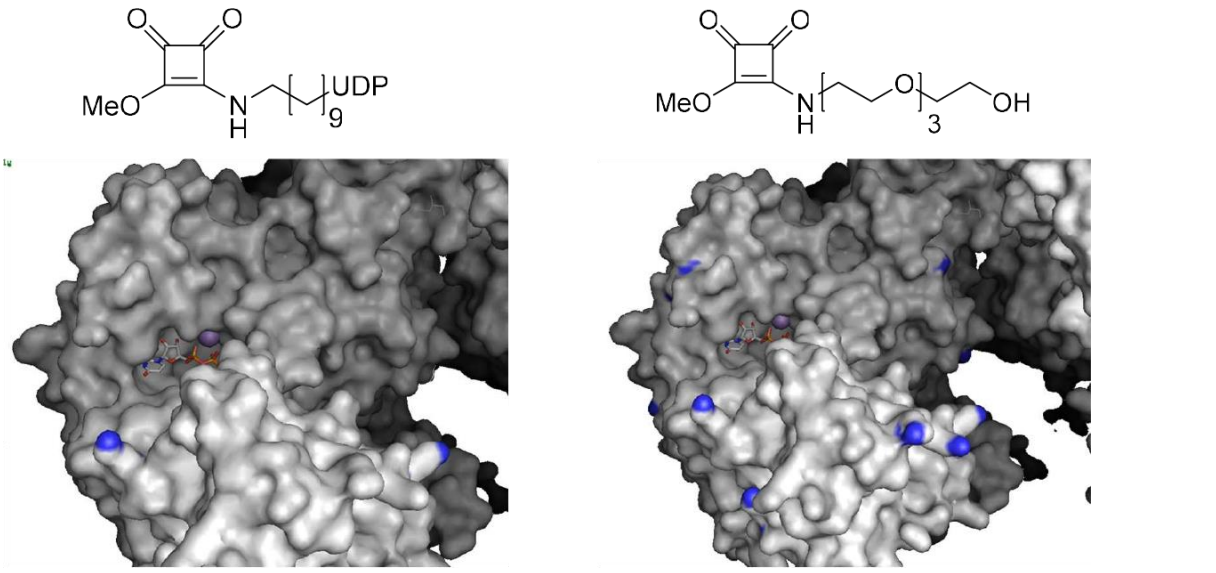


Figure 3.7. Surface models showing modified lysine residues (blue) of GlfT2 treated with compound **3.4e** and PEG_4-squaramide **3.1**. **3.4e** was only found to modify K494 and K519 whereas squarate **3.1** was observed to modify most surface accessible lysine residues.

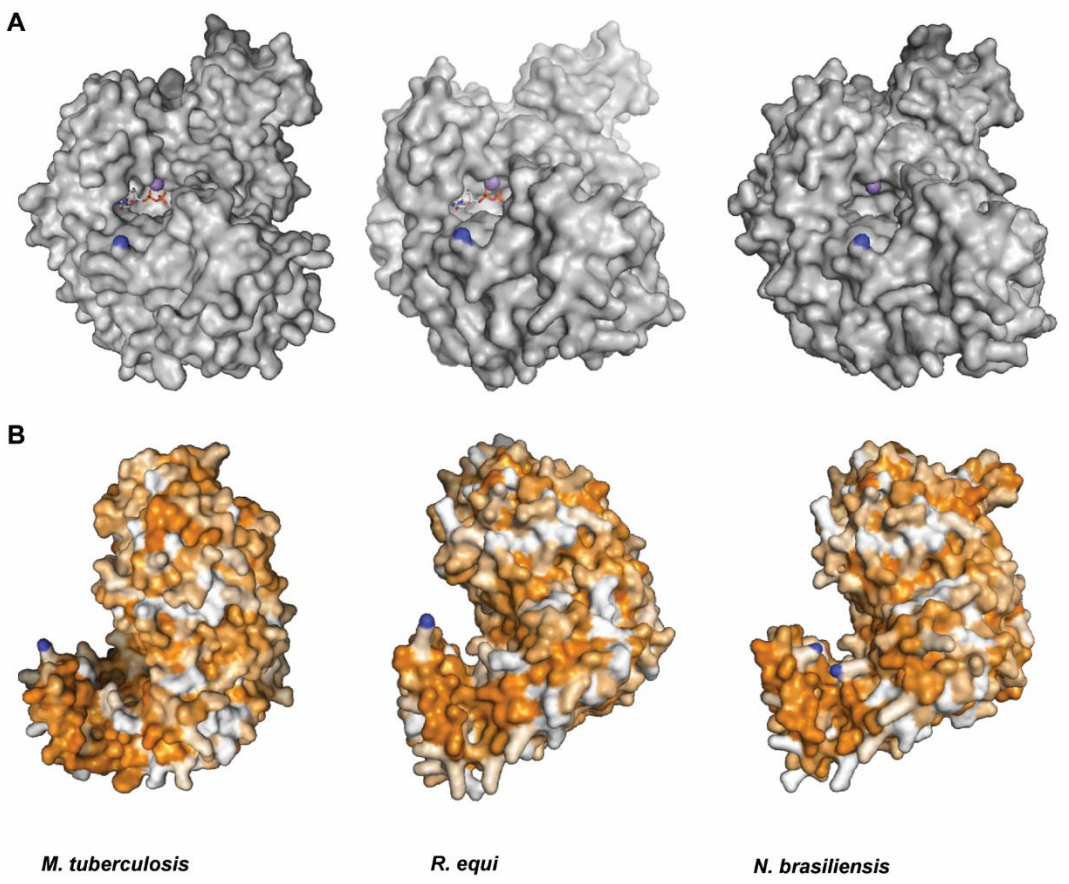


Figure 3.8. Surface models of three orthologs of GlfT2 treated with **3.5e**. R. equi and N. brasiliensis homology models based off structure of *M. tuberculosis* GlfT2 (PDB:4FIY) using SWISS-MODEL.[111] (A) Lysine modification (blue) is observed at K494, just outside of the active site. (B) Lysine modification of K519 on opposite face of GlfT2 near a groove containing a high density of hydrophobic residues (two lysines modified in *N. brasiliensis*). Hydrophobic residues shaded in orange on (B) surface models.

site pocket. We hypothesize that the UDP portion of the analogs may still be binding in the active site in an orientation that allows the alkyl squarate portion to extend out to bind K494.

Modification of K519, which is farther from the active site as well as on the opposite face of the enzyme, suggests the presence of an additional binding site between GlfT2 and UDP-squarate **3.4e**. It has been proposed that for GlfT2 to elongate the galactan to a consistent length, it utilizes an allosteric site that binds to the lipid of an acceptor substrate and only dissociates

3.1 treated		
Modified Peptide	**residue**	**hits**
KLYLEESTTNAR	K31	**10**
RWTTC**K**	K81	**1**
RWTTC**K**SVVLR	K81	**3**
WTTC**K**SVVLR	K81	**2**
T**K**ATGAR	K101	**2**
VMYEAL**K**NTDCQQILFMDDDIRLEPDSILR	K252	**6**
FA**K**APMLVGGQMLNLQEPSHLHIMGEVVDR	K283	**9**
QVAEELGQPLPLFI**K**WDDADYGLR	K375	**4**
LVVAAMHWDGP**K**AQVIGLVR	K435	**3**
ATL**K**HLACLEYSTVAIQNK	K451	**8**
IR**K**SYDAVVLPAASELPPPLHK	K494	**3**
KSYPDAVVLPAASELPPPLHK	K494	**11**
KSYPDAVVLPAASELPPPLHKNK	K494	**3**
SYPDAVVLPAASELPPPLH**K**NK	K514	**5**
N**K**AMKPPVNPLVIGYR	K516	**13**
AM**K**PPVNPLVIGYR	K519	**13**
DRA**K**MFALLWQSLR	K589	**1**
A**K**MFALLWQSLR	K589	**2**
A**K**MFALLWQSLRR	K589	**1**
IYRDALPTLSS**K**QK	K625	**3**
DALPTLSS**K**QK	K625	**2**
Q**K**WETALLPAANQEPEHG	K627	**6**
3.4e treated		
Modified Peptide	**residue**	**hits**
KSYPDAVVLPAASELPPPLHK	K494	**6**
NKAM**K**PPVNPLVIGYR	K519	**2**
AM**K**PPVNPLVIGYR	K519	**1**
3.4e + 3.5 treated		
Modified Peptide	**residue**	**hits**
IR**K**SYPDAVVLPAASELPPPLHK	K494	**2**
KSYPDAVVLPAASELPPPLHK	K494	**10**
KSYPDAVVLPAASELPPPLHKNK	K494	**1**
NKAM**K**PPVNPLVIGYR	K519	**4**
AM**K**PPVNPLVIGYR	K519	**9**

Table 3.1 Table of detected GlfT2 peptides modified by squaric esters **3.2** and **3.4e** (with and without the presence of acceptor **3.5**) after trypsin digest. Presumed modified lysine is bolded. Residue number corresponds to modified lysine and hits represents the number of isotopically distinct peptides detected with an error below 5ppm and an ions score over 20.

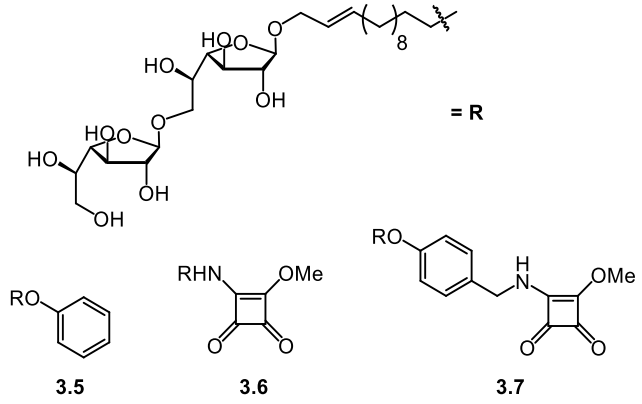
Rhodococcus equi		
Modified Peptide	**residue**	**hits**
KQYPDAVVLPSSTSLPEPSR	K495	**1**
KQYPDAVVLPSSTSLPEPSRADVGDVGLPTGK	K495	**4**
QYPDAVVLPSSTSLPEPSRADVGDVGLPTG**K**VAIATR	K526	**1**
Nocardia brasiliensis		
Modified Peptide	**residue**	**hits**
LVVASLHLPGNG**K**AMVVNTIK	K461	**6**
AMVVNTI**K**ATLK	K469	**3**
KQYPDAVILPSSTELPLASHLEVGAVAEPANPIAK	K516	**6**
KQYPDAVILPSSTELPLASHLEVGAVAEPANPIA**K**VVR	K550	**2**
QYPDAVILPSSTELPLASHLEVGAVAEPANPIA**K**VVR	K550	**5**

Table 3.2 Table of *R. equi* and *N. bras* GlfT2 peptides modified by squaric ester **3.4e** after trypsin digest. Presumed modified lysine is bolded. Residue number corresponds to modified lysine and hits represents the number of isotopically distinct peptides detected with an error below 5ppm and an ions score over 20.

when the polysaccharide reaches the appropriate length, demonstrating a processive elongation rather than a distributive elongation.[106] In this scenario, GlfT2 may be binding to the decyl linker of **3.4e** in an orientation where the squarate moiety is near K519 and can form the covalent bond. Upon close inspection of the crystal structure, K519 is situated next to a groove composed of mostly hydrophobic residues (Figure 3.6B). When the same analysis was done on GlfT2 treated with **3.4a**, very minor to no modification of any lysines was found. This finding corroborates the data from the luminescence assays performed above as increased inhibition was observed with a longer alkyl linker. The increased length of the decyl linker would make it easier for the squarate moiety to reach K494 as well as increase the binding affinity to the hydrophobic groove near K519. Surprisingly, addition of the lipid linked disaccharide **3.5** did not abate binding of **3.4e** to K519. Instead, the presence of an acceptor ligand increased the signal of modified K494 and K519, suggesting GlfT2 may undergo a conformational change into a more active form after binding to an acceptor substrate (Table 3.1).

To determine how conserved these features are to GlfT2, the same MS/MS analyses were performed on GlfT2 orthologs from *Nocardia brasiliensis (N. bras.)* and *Rhodococcus equi (R. equi.)*, after treatment with **3.4e**. Homology models were constructed of these two orthologs

Scheme 3.2. Squaramide analogs of synthetic acceptor **3.5**.



based on *M. tuberculosis* GlfT2, and we found similar modification profiles in the MS data (Table 3.2). In all cases, there was a lysine modified near the modeled active site (K495 for *R. equi* and K516 for *N. bras*) as well as a lysine on the opposite face next to a relatively hydrophobic area (K526 for *R. equi*). In the case of *N. bras.*, there were two lysines present near the hydrophobic area, K469 and K550, and both were observed to have been modified by UDP-squarate **3.4e** (Figure 3.8). Together the data demonstrate that while the squaric ester is reactive enough to bond with lysine in aqueous environments, the scaffold on which the squarate is installed can direct binding to specific lysines proximal to areas of activity on an enzyme. Furthermore, these UDP-squarate probes consistently modified GlfT2 across multiple orthologs in a hydrophobic groove distal to the active site, supporting previous suggestions of an essential allosteric site for proper GlfT2 function.

Squarate acceptor analogs do not bind to GlfT2. Based on the previous results we also synthesized analogs of a GlfT2 acceptor substrate to see if more information about GlfT2 binding dynamics could be uncovered. Analogs of acceptor **3.5** were synthesized with the squarate appended to the lipid tail, with and without the aryl ring (Scheme 3.2). GlfT2 was treated with compounds **3.6** and **3.7** under the same conditions as **3.4e**. When treated GlfT2 was analyzed by MS/MS, there was no observed covalent binding of lysine by either molecule, regardless of the presence of compound **3.5**.

Because compounds **3.6** and **3.7** are analogs of an acceptor substrate, we also tested how the addition of these squarates modified elongation by GlfT2. *M. tuberculosis* GlfT2 was treated with UDP-Galf and either compound **3.6** or **3.7**, and the resultant products were characterized by MALDI. Only compound **3.7** showed any elongation by GlfT2 and was only able to be elongated to about 6 galactofuranose units at most. However, this elongation profile fit more with a distributive mechanism rather than a processive one as the shorter length polysaccharides were detected at similar levels to unmodified **3.6** (Figure 3.9).

While squarate analogs of the donor substrate were able to bind in a specific manner and illustrate the complex dynamics of GlfT2 elongation, the same was not seen with the acceptor analogs. The fact that they could not be elongated suggests that the squarate moiety may have been installed in such a way that interferes with proper binding of the acceptor motif. If the lipid tail is important for binding, the squarate being directly attached to the alkyl chain could be sterically clashing and transfer to the disaccharide portion could alleviate that potential clash.

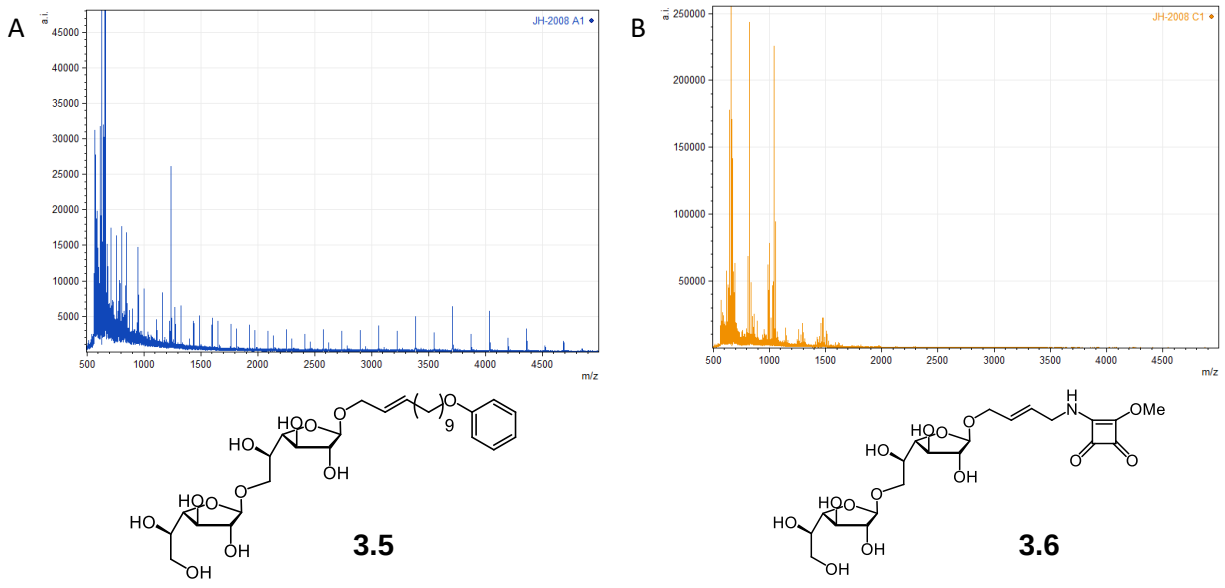


Figure 3.9. MALDI trace of elongation products of synthetic acceptor (A) **3.5** and (B) squaramide analog **3.6**. While GlfT2 elongation products of **3.5** show a processive behavior with a local maximum of m/z=3700, indicating an average extension of 19 gal*f* residues, the extension products using **3.6** as the acceptor consist of mostly +1 and +2 gal*f* chains with very minor +3 and +4 gal*f* chains.

While the data obtained from the these acceptor analogs provided little insight into GlfT2 function, these experiments highlight how precarious and sensitive rational design can be.

3.4 Conclusions

The natural abundance of lysine provides a untapped pool of targets for covalent small molecules; however, the tradeoff is that this prevalence coupled with its low nucleophilicity make the design of a specific probe challenging. Our characterization of squaric esters as mildly reactive, with tunable kinetics and selectivity, and stability in aqueous media highlight their potential for rationally designing new lysine specific probes. Squaric esters react much slower than other lysine targeting electrophiles, so binding events are heavily dictated by the scaffold it is attached to. Installation of squaric esters onto a scaffold that is recognized by an enzyme can imbue specificity to these small probes without greatly perturbing the original binding interactions when done properly.

We employed donor substrate analogs of the glycosyltransferase GlfT2 bearing a squarate to confirm specific binding at the active site and an additional hydrophobic pocket. UDP-squarates are amenable to other enzymes that use UDP-sugars for mechanistic studies, but more importantly squaric esters have the potential to create new covalent probes to explore previously untapped areas of the proteome and expand the target space for lysine-reactive small molecules.

3.5 Experimental Details

NMR kinetic experiments

Stock solutions of reagents and benzylamine in d_6-DMSO were diluted to a final concentration of 100μM, 10mM, or 50mM directly in an NMR tube for immediate 1H NMR acquisition. The inverse of compound concentrations was plotted with time, and the slope of the best fit line was

used to calculate second order rate constant *k*. 1H NMR spectra were recorded on a 500 MHz spectrometer.

Production of His_6-GlfT2

All cloning was performed using DH5α (NEB) grown in LB supplemented with kanamycin (50 μg/mL) when appropriate. Mtb GlfT2 was prepared according to previously reported methods.[106] Constructs for Nbr glfT2 (accession number: AFT98100.1) and Req glfT2 (accession number: EGD23791.1) were codon-optimized for expression in E. coli and ordered as gene blocks (IDT). Isothermal assembly was used to clone gene blocks into NdeI (NEB) linearized pET28a. Protein expression and purification: Constructs were transformed in DE3 Tuner E. coli (Novagen), and transformants were isolated on LB agar plates supplemented with kanamycin (50 μg/mL). Starter cultures were made from single colonies of transformants and grown overnight at 37°C in LB supplemented with kanamycin (Gold Biotechnology). Starter cultures were used to inoculate 1L LB cultures supplemented with kanamycin. Cultures were grown at 37°C until reaching mid-log (OD600 = ~0.6), chilled on ice for 30 mins, and protein production was induced by addition of IPTG (Gold Biotechnology) to 0.3 mM. Chilled cultures were returned to the incubator and protein expression proceeded for ~20 hours at at 16°C. Following expression, cells were harvested by centrifugation (4000 RCF, 10 mins), collected into 50 mL conical tubes and frozen until further purification.

For purification, cell pellets were thawed on ice and resuspended in lysis buffer (40 mL/liter of growth) containing 50 mM sodium phosphate (pH 7.6), 400 mM NaCl, 20 mM imidazole, 15% glycerol supplemented with 1 mg/mL lysozyme (Gold Biotechnology), 1 mM PMSF, 0.1% TX-100, 1000 U Benzonase (Sigma-Aldrich). Resuspended cells were lysed at 22 kPsi with a benchtop cell disruptor (Constant Systems, Kennesaw, GA). Cell lysate was clarified by

centrifugation (20,000 RCF, 1 hour), and supernatant was passed through a 0.22-micron filter (GE Healthcare), and applied to a 1 mL HisTrap HP column (GE Healthcare). Following loading, the column was washed with 10 column volumes of buffer A (50 mM sodium phosphate, 400 mM NaCl, 20 mM imidazole, 15% glycerol, pH 7.6), and eluted with buffer B (50 mM sodium phosphate, 400 mM NaCl, 400 mM imidazole, 15% glycerol, pH 7.6) using a linear gradient (0-100% B over 20 column volumes). Eluted fractions containing pure His6-GlfT2 were identified using SDS-PAGE, pooled, vitrified, and stored frozen at -80°C until further use.

GlfT2 activity assays

GlfT2 was diluted in buffer (100mM NaCl, 50mM HEPES, 5mM EDTA, 25μM $MgCl_2$, pH=7.4) along with acceptor **3.5**. Compounds **3.4a-e** or UMP were then added to specified concentrations and preincubated for a set amount of time before UDP-gal*f* was added. Each sample consisted of GlfT2, acceptor **3.5**, and UDP-Gal*f* at concentrations of 5nM, 40μM, and 500μM respectively in a total volume of 35μL. Samples sat for 1hr at room temp after addition of UDP-gal*f* before being split up into 10μL portions in a 384 well plate. 10μL of UDP-Glo™ working reagent was added to each well, where it was then moved to a 37°C incubator for 30 min. UDP-Glo luminescence was then read on a Tecan M1000 plate reader and compared to a control series containing known concentrations of UDP in buffer. Each set of conditions were performed in triplicate giving a total of nine measurements per condition.

GlfT2 MS/MS analyses

25μg of dialyzed GlfT2 was treated with UDP-squarate and **3.5** in buffer (100mM NaCl, 50mM HEPES, 5mM EDTA, 25μM $MgCl_2$, pH=7.4) to a final concentration of 1μM GlfT2, 1mM UDP-squarate, and 1mM **3.5**. Solution was incubated at 27°C for 16 hr. Samples were dialyzed

again to wash out unbound material, treated with trypsin, and submitted for mass spectrometry analysis. A list of detected masses was compared to a control list of trypsin digested GlfT2 peptides to extract modified peptides.

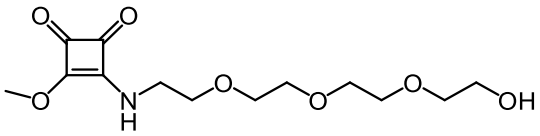
MALDI analyses of GlfT2 elongation

A sample of His6-GlfT was dialyzed twice into 2 L of 50 mM HEPES (pH 7.4), 100 mM NaCl, and 5 mM EDTA by using 10,000 molecular weight cut-off dialysis cassettes (Pierce Biotechnology). Protein was assayed by using the BCA assay (Pierce Biotechnology) with BSA as a standard. Reaction mixtures of 120-μL total volume contained final concentrations of 0.2 μM H_6GlfT2, 200 μM acceptor, 1.25 mM UDP-Galf in 50 mM HEPES, 25 mM $MgCl_2$, and 100 mM NaCl, pH 7.0 (buffer was added from 10μ concentrated stock solution). Reactions were incubated at room temperature for a specified time, and then 120 μL of 1:1 MeOH:$CHCl_3$ was added to quench the reaction. Quenched reaction mixtures were evaporated to dryness in a SpeedVac SC100 (Varian) under vacuum. Dried reaction mixtures were resuspended in 50 L of a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% trifluoroacetic acid, and 2 μL of the resuspended reaction solution was spotted onto a stainless steel target (Applied Biosystems) for analysis by MALDI-TOF MS in either linear or reflectron mode (using a Voyager DE Pro; Applied Biosystems). Mass spectra were calibrated by using external standards (angiotensin II and bovine insulin).

Compound Synthesis

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Tetrahydrofuran (THF) was distilled from sodium/benzophenone, methanol (MeOH) was distilled from magnesium, and dichloromethane (DCM) was distilled from calcium hydride. Nuclear magnetic

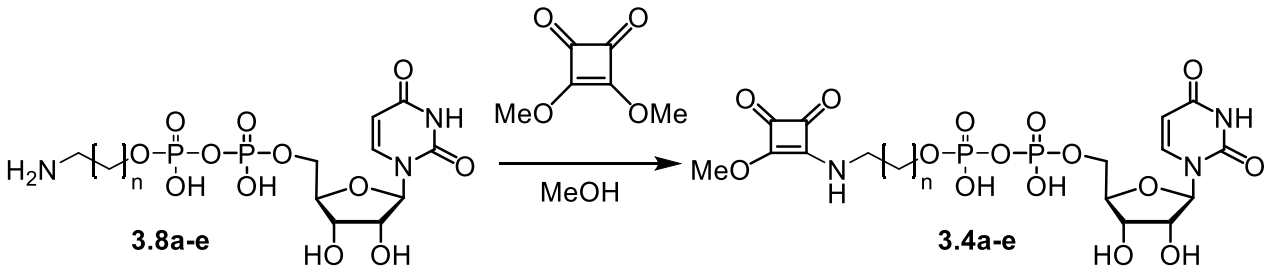
resonance spectra were recorded on a 400 MHz spectrometer (acquired at 400 MHz for ^{1}H and 100 MHz for ^{13}C), or 500 MHz spectrometer (acquired at 500 MHz for ^{1}H and 125 MHz for ^{13}C), or 600 MHz spectrometer (acquired at 600 MHz for ^{1}H and 150 MHz for ^{13}C). Chemical shifts are reported relative to tetramethylsilane or residual solvent peaks in parts per million ($CHCl_3$: ^{1}H 7.27, ^{13}C 77.23; MeOH: ^{1}H 3.31, ^{13}C 49.15; D_2O: ^{1}H 4.79; DMSO: ^{1}H 2.50, ^{13}C 39.52). High-resolution mass spectra (HRMS) were obtained on an electrospray ionization time-of-flight (ESI-TOF) mass spectrometer.



3-((2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(3.1)**

290mg of Amino-PEG4-alcohol purchased from BroadPharm was added to a solution of 200mg dimethylsquarate in 20mL MeOH. Reaction stirred at 27°C for 16 hours before being concentrated under vacuum and purified via silica gel chromatography (1:19 MeOH:DCM). Compound **3.1** was yielded in 40% yield. ^{1}H NMR (500 MHz, d_6-DMSO) δ 4.28 (s, 1.7H, rotamer 1), 4.27 (s, 1.3H, rotamer 2), 3.61 (t, *J*=5.16 Hz, 1H), 3.54-3.47 (m, 12H), 3.40 (t, *J*=5.16 Hz, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) (chemical shifts of both rotamers listed) δ 43.69, 44.24, 60.33, 60.54, 60.64, 69.65, 70.10, 70.20, 70.30, 72.76, 110.05, 120.49, 121.76, 125.63, 127.19,127.47, 127.70, 128.03, 129.34, 137.86, 139.85, 141.17, 144.33, 172.69, 173.09, 177.52, 177.83, 182,70, 182.89, 189.66, 189.85

General procedure for UDP-alkyl-squaramides **(3.4a-e)**



Using the corresponding amino alcohol, aminoalkyl-UDP intermediates **3.8a-e** were generated according to literature precedent.[108] Then to a solution of **3.8a-e** (0.5mmol) in 5mL of MeOH, dimethyl squarate (1.5mmol) was added and reaction was stirred at room temperature overnight. Reaction was concentrated under vacuum and then purified by silica gel chromatography (4:2:1 EtOAc:MeOH:1M AcOH). Fractions were concentrated under vacuum, and redissolved into minimal amount of water. Solution was further washed via C-18 SepPak cartridges, then frozen and lyophilized to afford UDP-alkyl-squaramides **3.8a-e**.

Uridine 5'-diphosphoethylamino-methyl squarate **(3.4a)**

3.4a was isolated in 36% yield before SepPak wash. ^{1}H NMR(400 MHz, D_2O) δ 7.92 (d, *J*=6.76 Hz, 1H), 5.93 (m, 2H), 4.33 (m, 5H), 4.25-4.15 (m, 3H), 4.08 (br, 2H), 3.83 (br, 1H), 3.70 (br, 1H). HRESI-MS *m/z* calc. for $[M - H]^-$ $C_{16}H_{20}N_3O_{15}P_2$: 556.0375, found: 556.0381

Uridine 5'-diphosphobutylamino-methyl squarate **(3.4b)**

3.4b was isolated in 62% yield before SepPak wash. ^{1}H NMR(400 MHz, D_2O) δ 7.97 (d, *J*=8.02 Hz, 1H), 5.96 (m, 2H), 4.36 (m, 5H), 4.28-4.15 (m, 3H), 3.97 (br, 2H), 3.62 (br, 1H), 3.48 (br, 1H), 1.77-1.55 (m, 4H). HRESI-MS *m/z* calc. for $[M-H]^-$ $C_{18}H_{24}N_3O_{15}P_2$: 584.0688, found: 584.0691

Uridine 5’-diphosphohexylamino-methyl squarate **(3.4c)**

3.4c was isolated in 46% yield before SepPak wash. ^{1}H NMR(400 MHz, D_2O) δ 7.96 (br, 1H), 5.95 (m, 2H), 4.35 (br, 5H), 4.30-4.13 (m, 3H), 3.93 (m, 2H), 3.57 (t, *J*=6.76 Hz, 1H), 3.44 (t, *J*=6.89 Hz, 1H), 1.67-1.52 (m, 4H), 1.43-1.29 (m, 4H). HRESI-MS *m/z* calc. for $[M-H]^-$ $C_{20}H_{28}N_3O_{15}P_2$: 612.1001, found: 612.1006

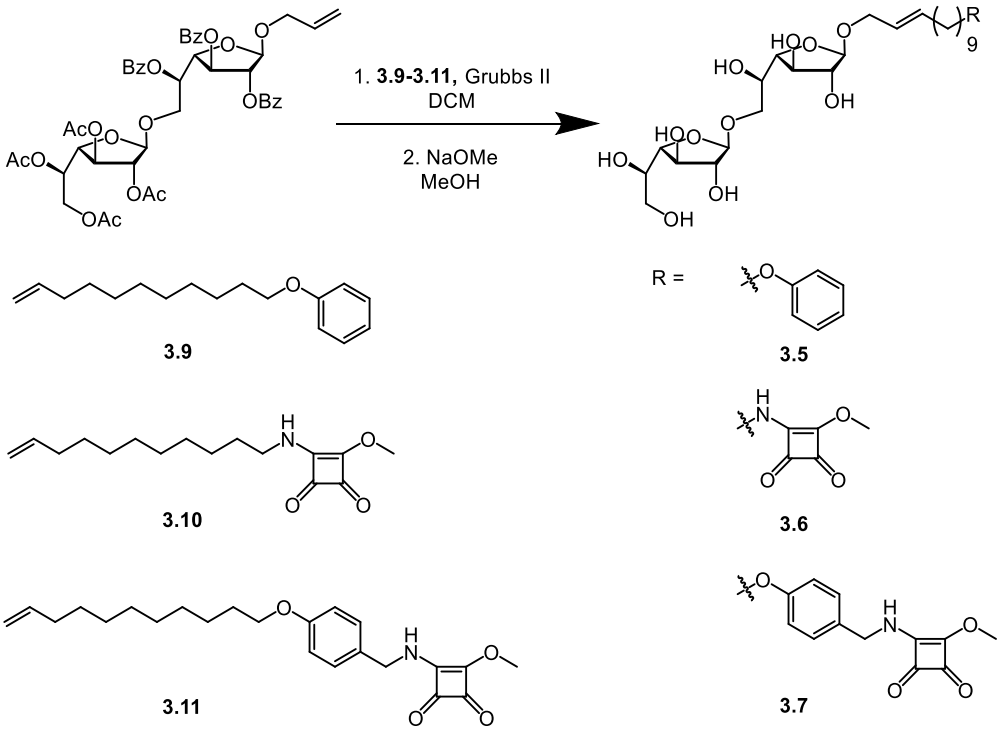
Uridine 5’-diphosphooctylamino-methyl squarate **(3.4d)**

3.4d was isolated in 50% yield before SepPak wash. ^{1}H NMR(400 MHz, D_2O) δ 7.98 (d, *J*=8.07 Hz, 1H), 5.97 (m, 2H), 4.37-4.34 (m, 5H), 4.29-4.15 (m, 3H), 3.93 (quart, *J*=6.71 Hz, 2H), 3.58 (t, *J*=7.49 Hz, 1H), 3.44 (t, *J*=7.02 Hz, 1H), 1.64-1.52 (m, 4H), 1.35-1.23 (m, 8H). HRESI-MS *m/z* calc. for $[M-H]^-$ $C_{22}H_{32}N_3O_{15}P_2$: 640.1314, found: 640.1311

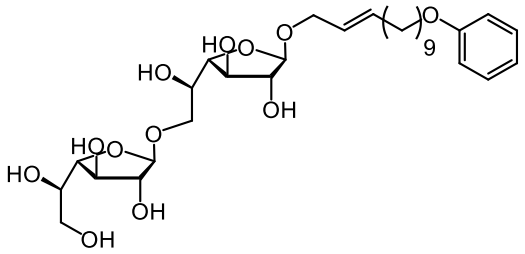
Uridine 5’-diphosphodecylamino-methyl squarate **(3.4e)**

3.4e was isolated in 23% yield before SepPak wash. ^{1}H NMR(400 MHz, D_2O) δ 7.98 (d, *J*=8.06 Hz, 1H), 5.97 (m, 2H), 4.38-4.33 (m, 5H), 4.28-4.15 (m, 3H), 3.92 (quart, *J*=6.61 Hz, 2H), 3.59 (t, *J*=6.61 Hz, 1H), 3.45 (t, *J*=6.82 Hz, 1H), 1.64-1.54 (m, 4H), 1.35-1.20 (m, 12H). HRESI-MS *m/z* calc. for $[M+H]^+$ $C_{24}H_{38}N_3O_{15}P_2$: 670.1773, found: 670.1774

General procedure for synthesis of synthetic acceptors **3.5-3.7**

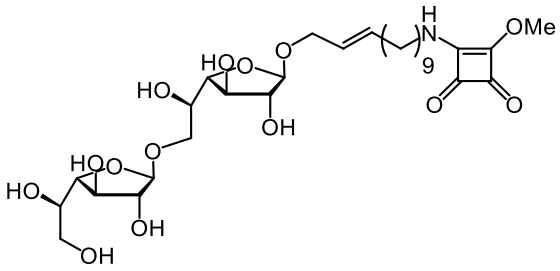


To a previously reported intermediate[65], lipids **3.9-3.11** was cross metathesized using Grubbs II in DCM before being deprotected in a sodium methoxide solution to afford acceptors **3.5-3.7**.



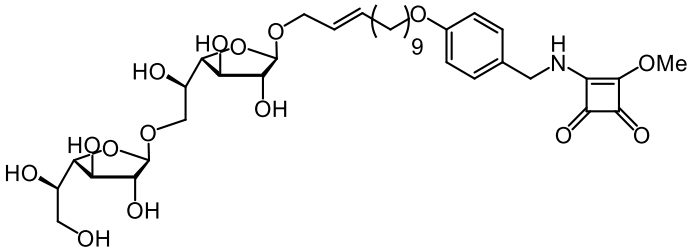
12-phenoxy-dodec-2-enyl β-D-galactofuranosyl-(1→6)-β-D-galactofuranoside **(3.5)**

3.5 was synthesized following previously reported procedures.[65]



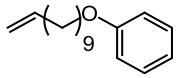
12-methoxysquaramidyl-dodec-2-enyl β-D-galactofuranosyl-(1→6)-β-D-galactofuranoside **(3.6)**

Using lipid **3.10**, synthetic acceptor **3.6** was isolated in 50% yield. ^{1}H NMR(500 MHz, MeOD) δ 5.08 (s, 1H), 4.27 (d, 3H), 4.02 (m, 1H), 3.96 (m, 1H), 3.92 (m, 2H), 3.84-3.79 (m, 2H), 3.63 (m, 3H), 3.58-3.47 (m, 4H), 3.29 (t, *J* = 7.14 Hz, 1H), 1.50 (m, 4H), 1.29-1.17 (m, 16H). ^{13}C NMR (126 MHz, MeOD) 188.66, 188.50, 183.51, 177.00, 176.44, 173.11, 173.05, 142.67, 141.14, 129.06, 128.17, 107.92, 107.77, 83.39, 81.37, 81.33, 77.33, 75.75, 70.79, 62.80, 61.43, 59.70, 57.63, 44.25, 30.55, 30.08, 29.17, 28.81, 28.02, 23.55.



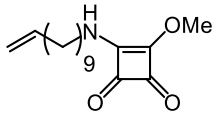
3-((4-(((E)-12-(((2R,3R,4R,5S)-5-((R)-2-(((2R,3R,4R,5S)-5-((R)-1,2-dihydroxyethyl)-3,4-dihydroxytetrahydrofuran-2-yl)oxy)-1-hydroxyethyl)-3,4-dihydroxytetrahydrofuran-2-yl)oxy)dodec-10-en-1-yl)oxy)benzyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(3.7)**

Using lipid **3.11**, synthetic acceptor **3.7** was isolated in 98% yield. ^{1}H NMR(600 MHz, MeOD) δ7.14 (m, 2H), 6.77 (m, 2H), 4.82 (m, 1H), 4.57 (m, 1H) 3.94-3.50 (m, 14H), 3.25 (s, 3H), 3.20 (m, 2H), 1.98-1.88 (m, 2H), 1.65 (m, 2H), 1.36 (m, 2H), 1.30-1.17 (m, 12H). ^{13}C NMR (151 MHz, MeOD) δ 158.94, 158.67, 128.70, 114.43, 114.29, 108.58, 83.80, 83.45, 82.00, 81.90, 81.71, 81.37, 77.51, 77.11, 69.69, 69.61, 69.22, 69.07, 67.65, 67.42, 63.08, 62.99, 29.24, 29.11, 29.06, 28.97, 25.74



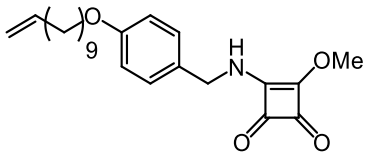
11-phenoxy-1-undecene (**3.9**)

1.0 eq. phenol was mixed in DMF with 0.66 eq. NaH (60% in oil) under inert atmosphere for 20 min. 0.5 eq. 11-bromo-undecene was then added and stirred at RT for 5 hr. Crude was purified via column chromatography (2% EtOAc in Hex) to afford **3.9** in 82% yield. Spectrometric values matched previously reported values.



3-methoxy-4-(undec-10-en-1-ylamino)cyclobut-3-ene-1,2-dione (**3.10**)

1.25 eq. of dimethyl squarate stirred with 1.0 eq. 11-amino-1-undecene in dry MeOH under N_2 for 16hr before being purified via column chromatography (3:1 diethyl ether : hexane) to afford **3.10** in 55% yield. ^{1}H NMR(500 MHz, $CDCl_3$) δ 5.80 (ddt, *J* = 10.29, 17.18, 6.49 Hz, 1H), 4.99 (ddt, *J* = 17.18, 1.68, 1.81 Hz, 1H), 4.92 (ddt, *J* = 10.29, 1.10, 0.95 Hz, 1H), 4.41 (m, 3H), 3.41 (q, *J* = 6.38 Hz, 2H), 2.03 (q, *J* = 6.04 Hz, 2H), 1.60 (p, *J* = 7.37 Hz, 2H), 1.40-1.24 (m, 12H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 189.42, 182.98, 177.52, 172.18, 139.15, 114.17, 60.47, 45.01, 33.78, 30.54, 29.40, 29.37, 29.09, 29.06, 28.89, 26.31.

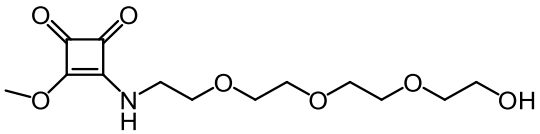


3-methoxy-4-((4-(undec-10-en-1-yloxy)benzyl)amino)cyclobut-3-ene-1,2-dione (**3.11**)

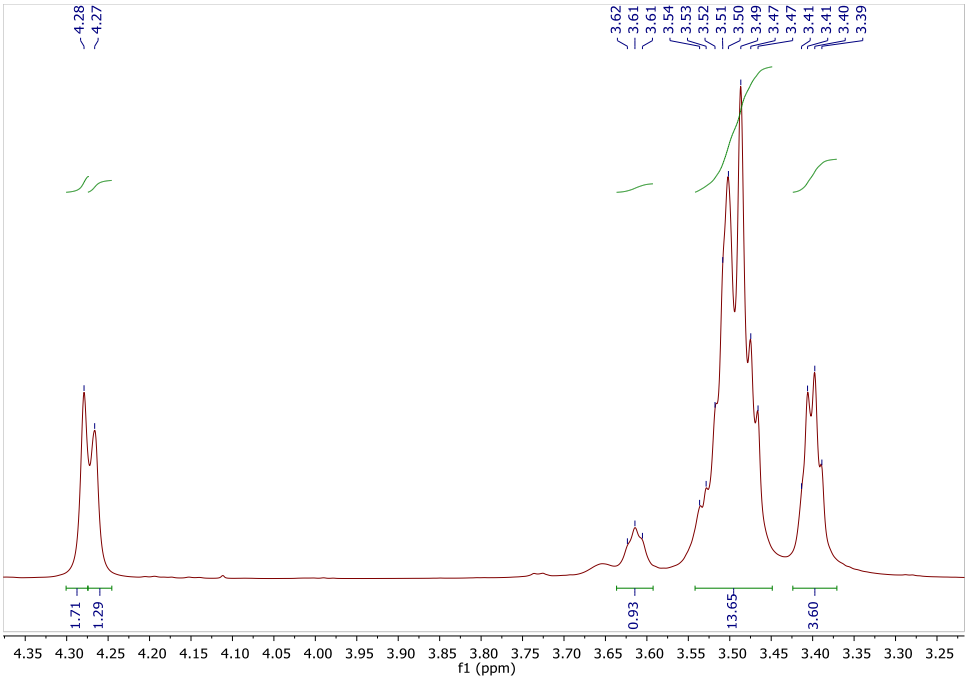
4-hydroxybenzylamine was dissolved in MeOH along with 2.5 eq. triethylamine and chilled to 0° C. TFAA was added dropwise and reaction was allowed to warm to RT and stirred for 4 hours

before concentrated and purified via column chromatography (1:1 EtOAc:Hex). The resulting amide was then dissolved in DMF with 0.66 eq. NaH (60% in oil) under inert atmosphere. Mixture stirred for 20min before 0.5eq 11-bromoundecene was added and stirred at RT for 5hrs. The reaction was then concentrated and purified via column chromatography (0-10% EtOAc in Hex). The resulting lipid was then deprotected via sodium methoxide solution and stirred with 1.1 eq. dimethyl squarate in MeOH. The crude product was then purified via column chromatography (2:3 EtOAc:Hex) to afford **3.11** in 40% yield over 4 steps. ^{1}H NMR(500 MHz, $CDCl_3$) δ 7.17 (m, 2H), 6.87 (d, *J* = 8.99 Hz, 2H), 5.80 (m, 1H), 4.98 (dm, *J* = 16.92 Hz, 1H), 4.92 (dm, *J* = 10.34 Hz, 1H), 4.75 (rotamer 1, br, 0.6 H of 2H), 4.50 (rotamer 2, br, 1.4H of 2H), 4.42 (rotamer 2, s, 2H of 3H), 4.37 (rotamer 1, s, 1H of 3H), 3.93 (t, *J* = 6.33 Hz, 2H), 2.02 (q, *J* = 7.18 Hz, 2H), 1.76 (p, *J* = 6.61 Hz, 2H), 1.43 (p, *J* = 8.50 Hz, 2H), 1.38-1.25 (m, 10H).

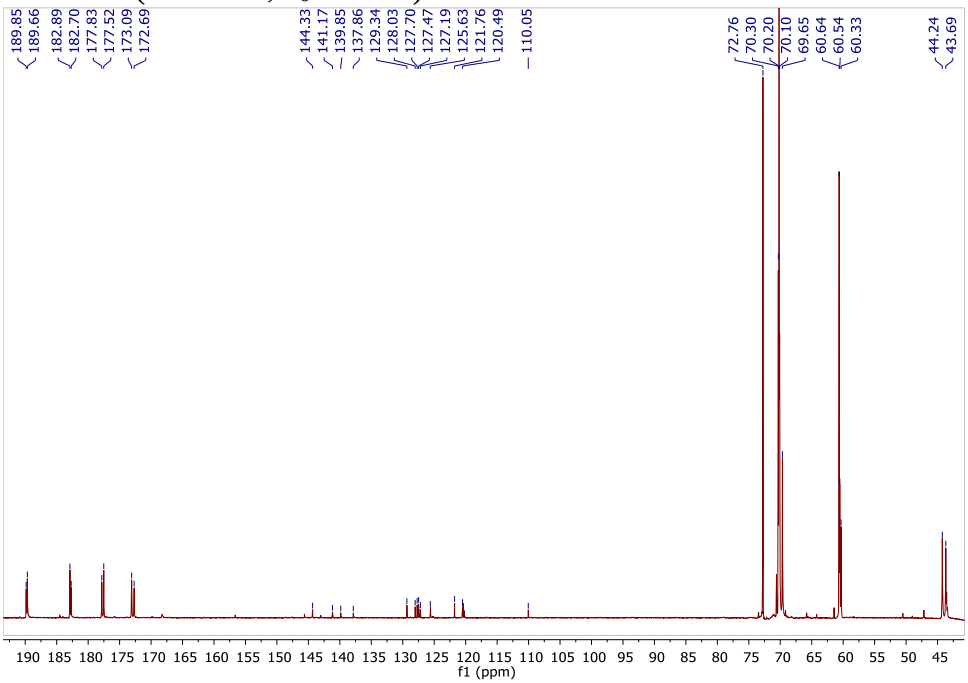
3.6 NMR Spectra

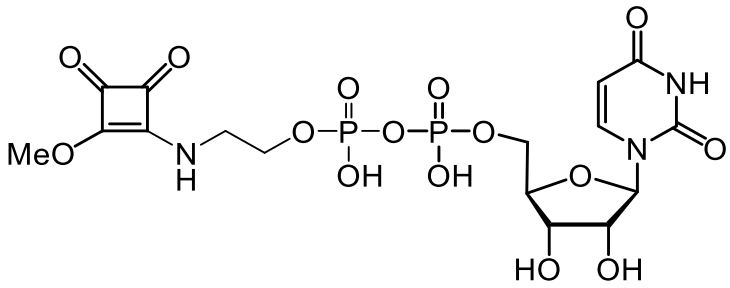


3.2, ^{1}H NMR (500 MHz, d_6-DMSO)

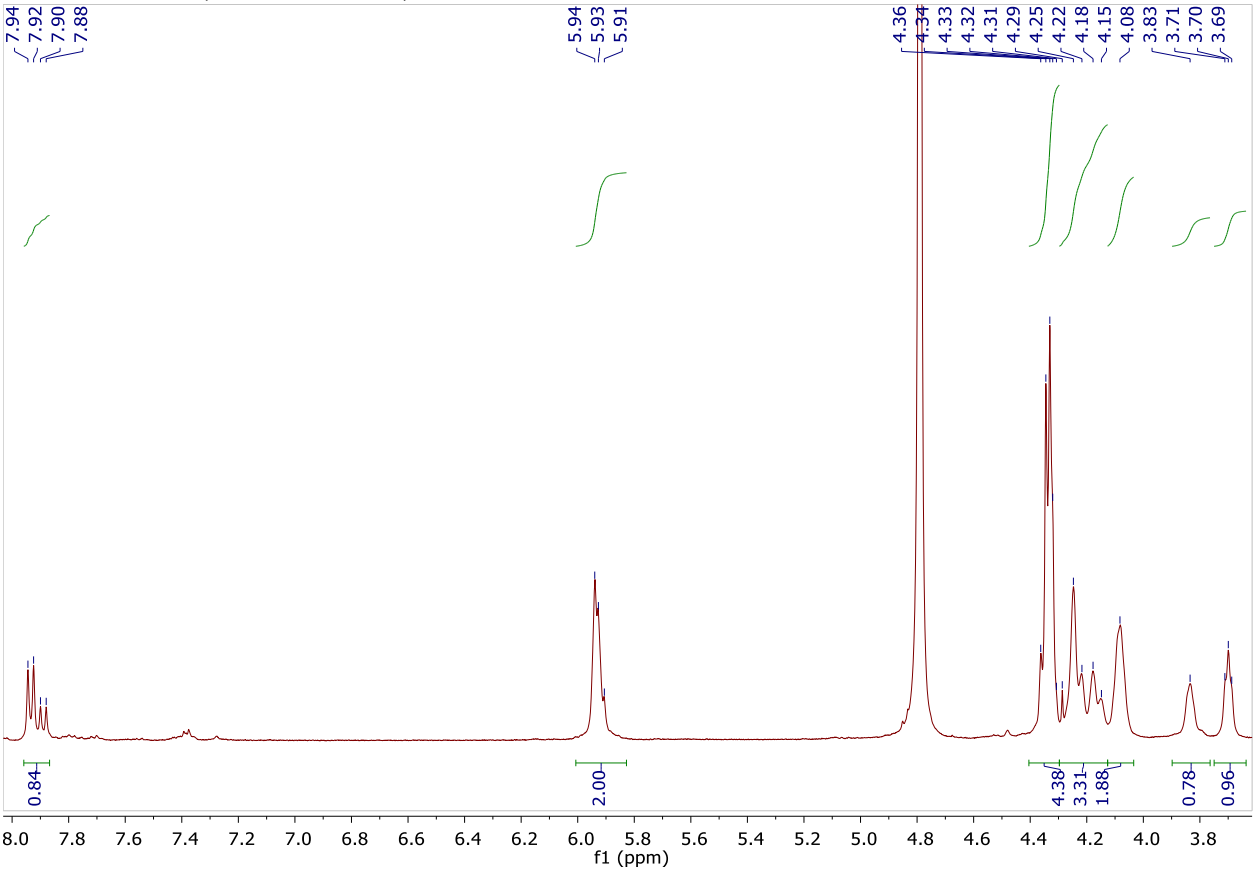


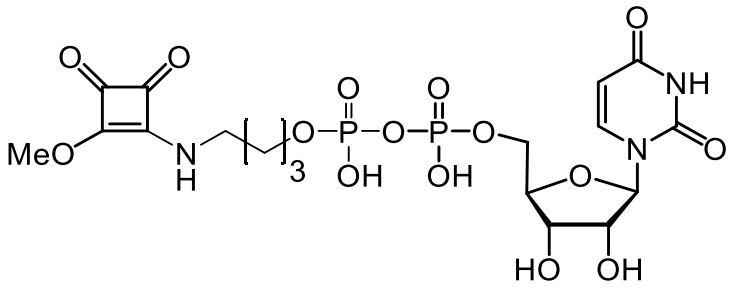
^{13}C NMR (151 MHz, d_6-DMSO)



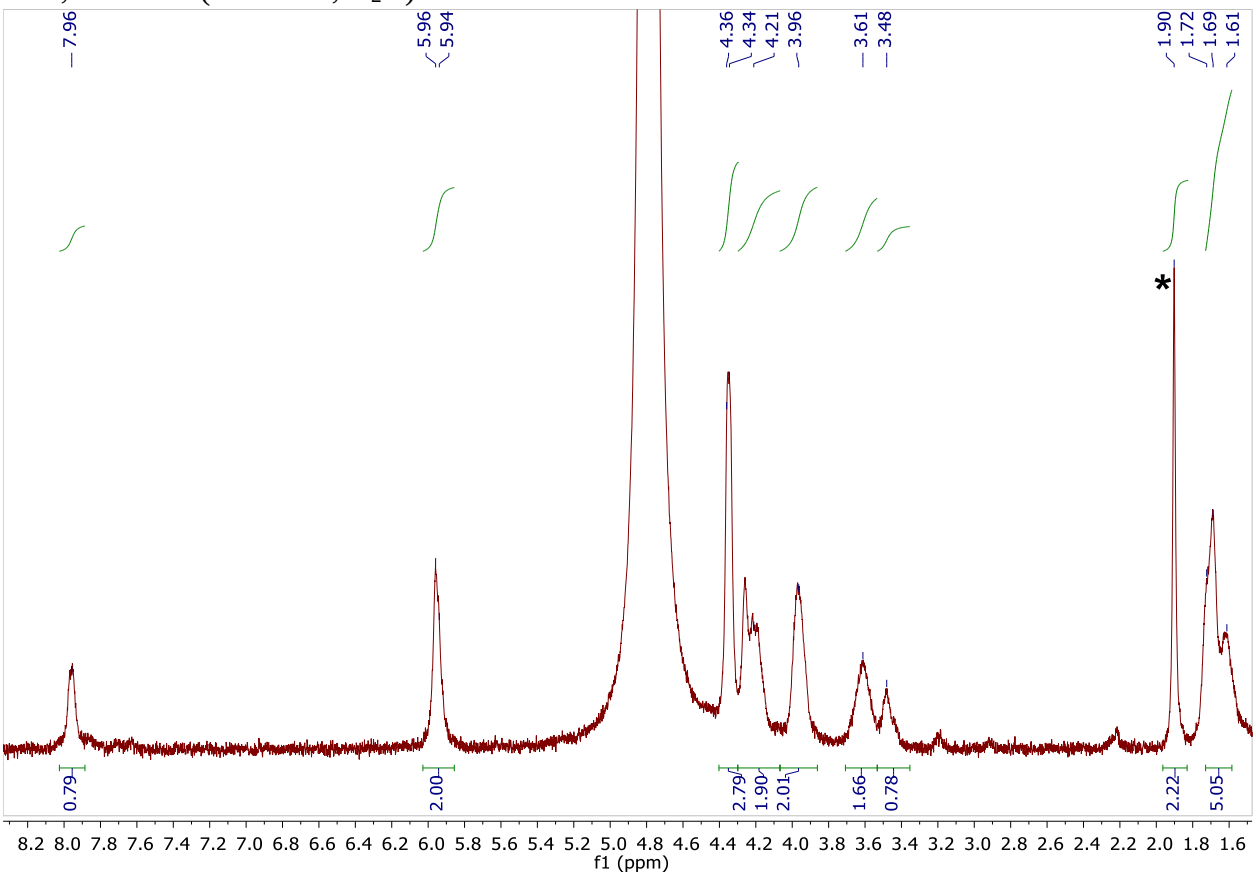


3.4a, ^{1}H NMR (400 MHz, D_2O)

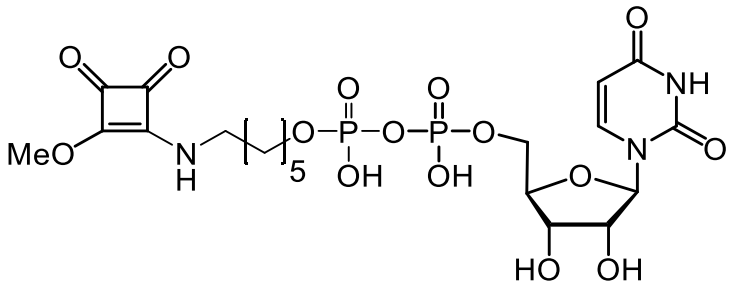




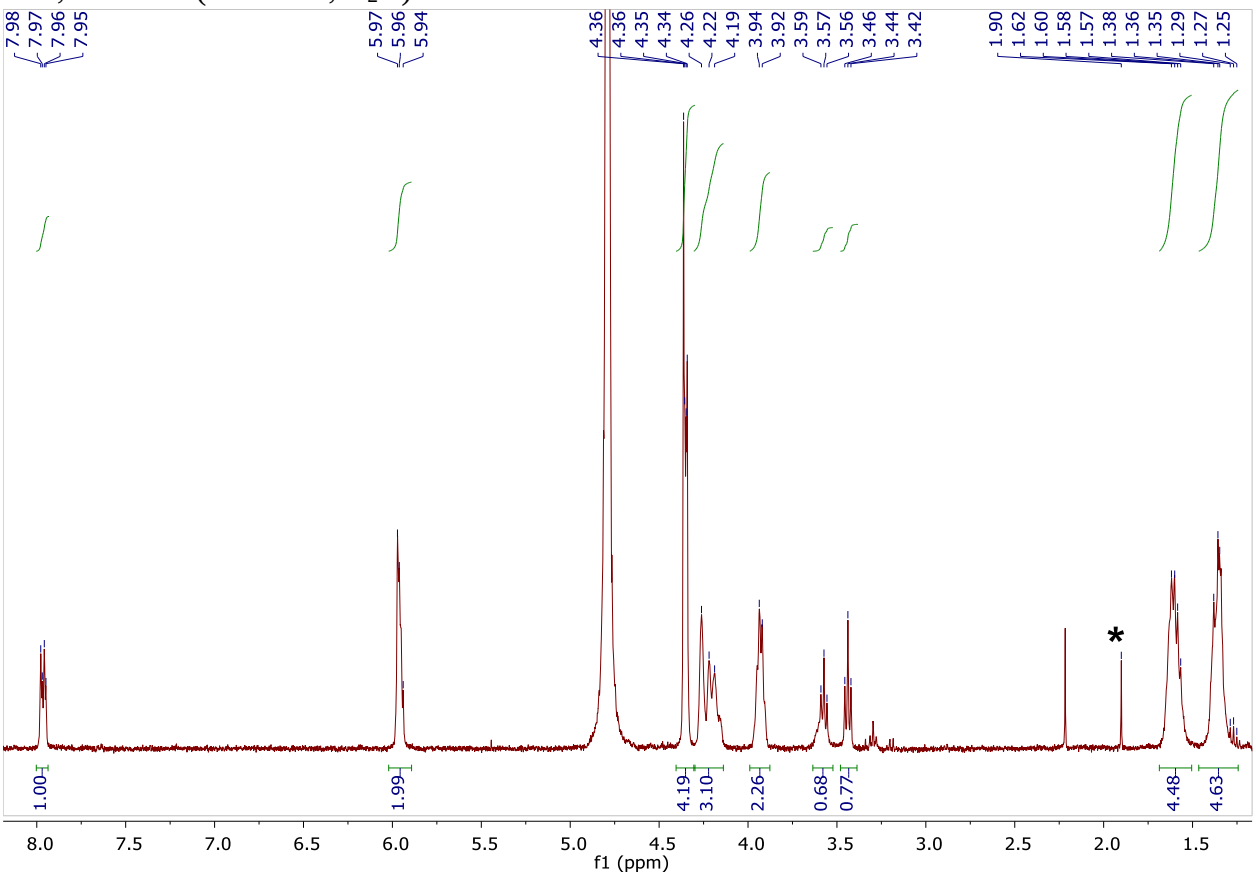
3.4b, ^{1}H NMR (400 MHz, D_2O)



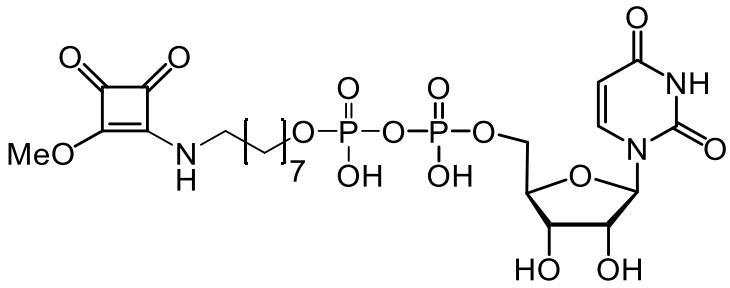
*δ 1.90 corresponds to ammonium acetate salt



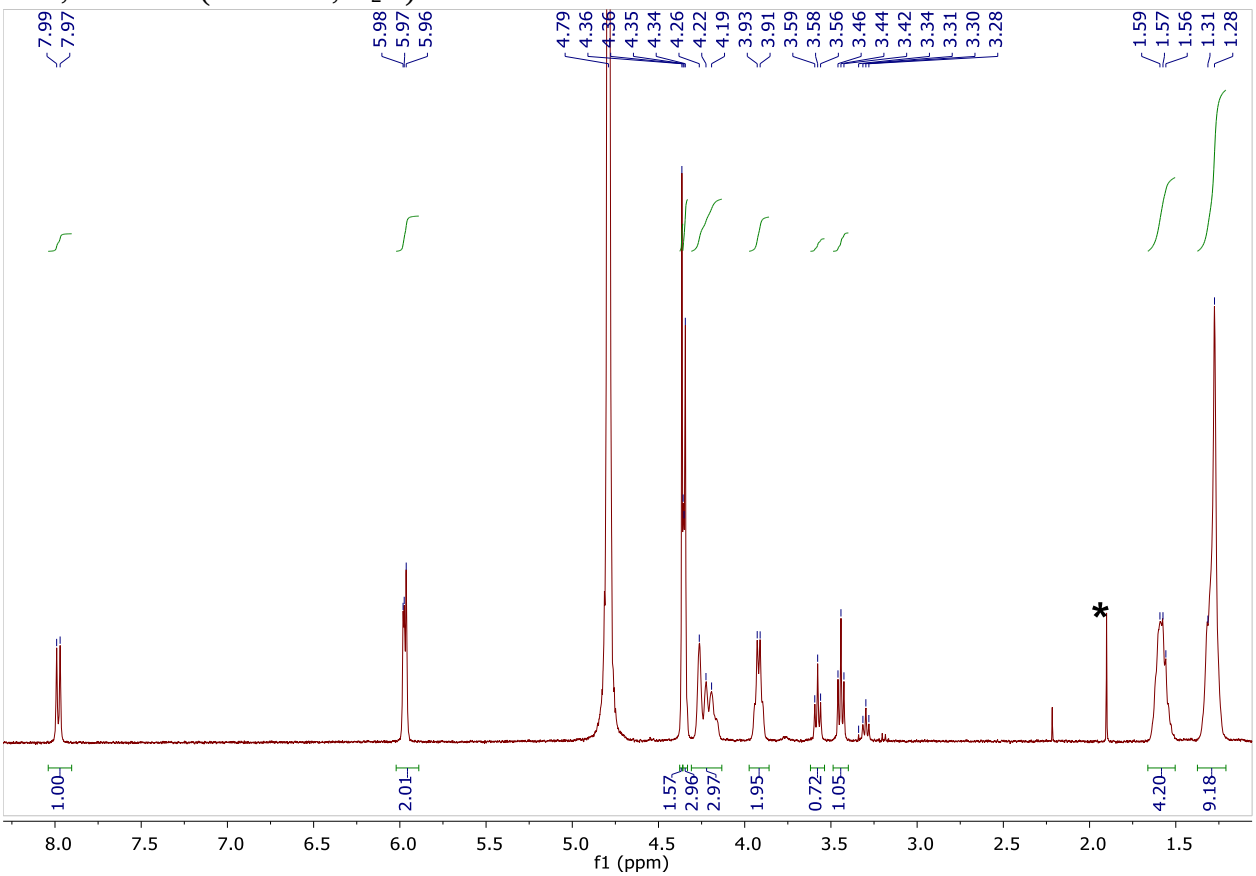
3.4c, ^{1}H NMR (400 MHz, D_2O)



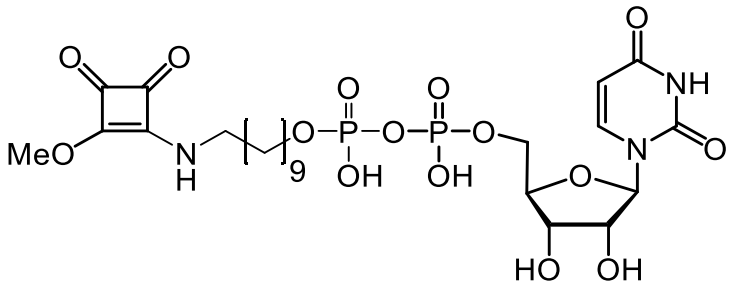
*δ 1.90 corresponds to ammonium acetate salt



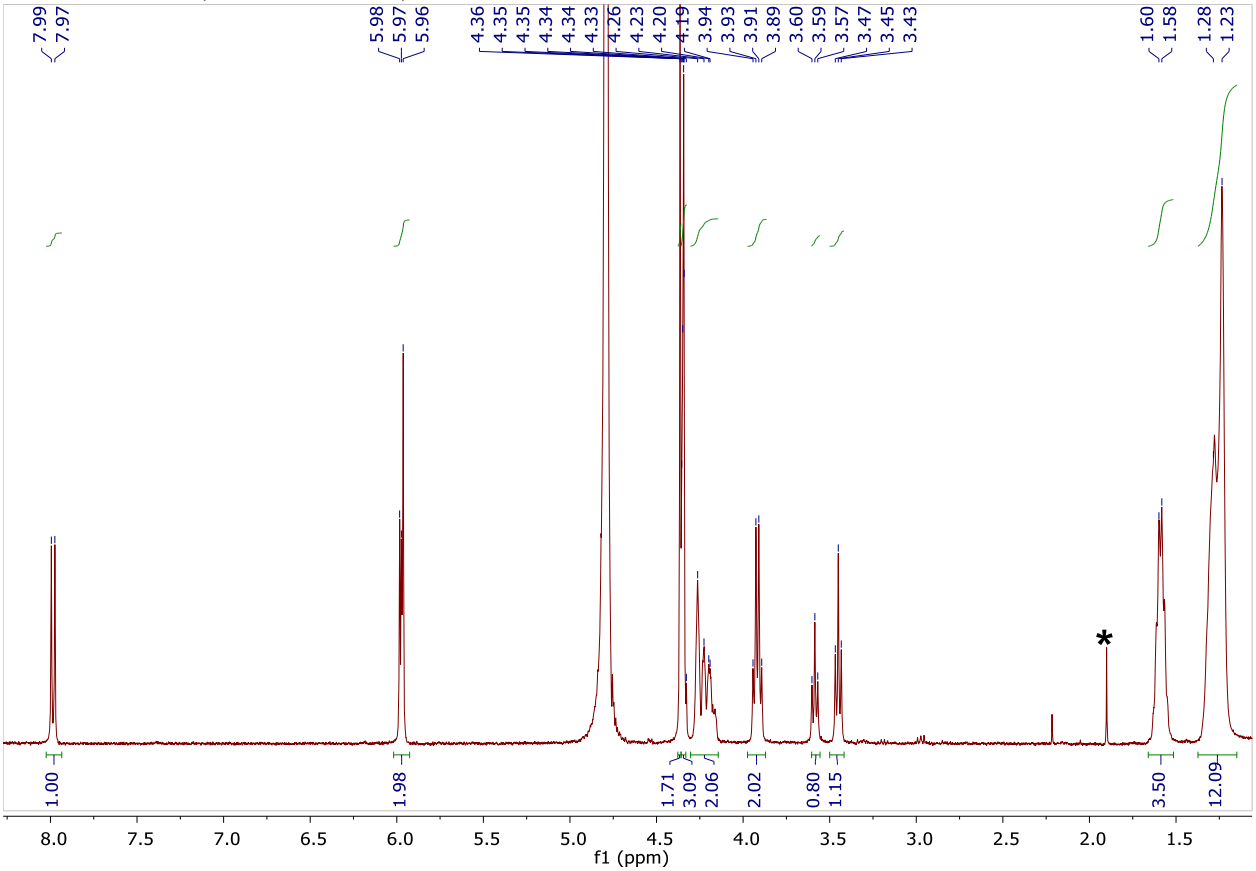
3.4d, ^{1}H NMR (400 MHz, D_2O)



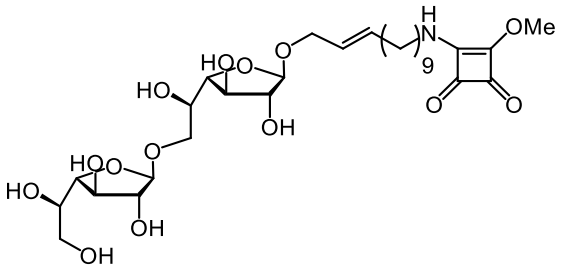
*δ 1.90 corresponds to ammonium acetate salt



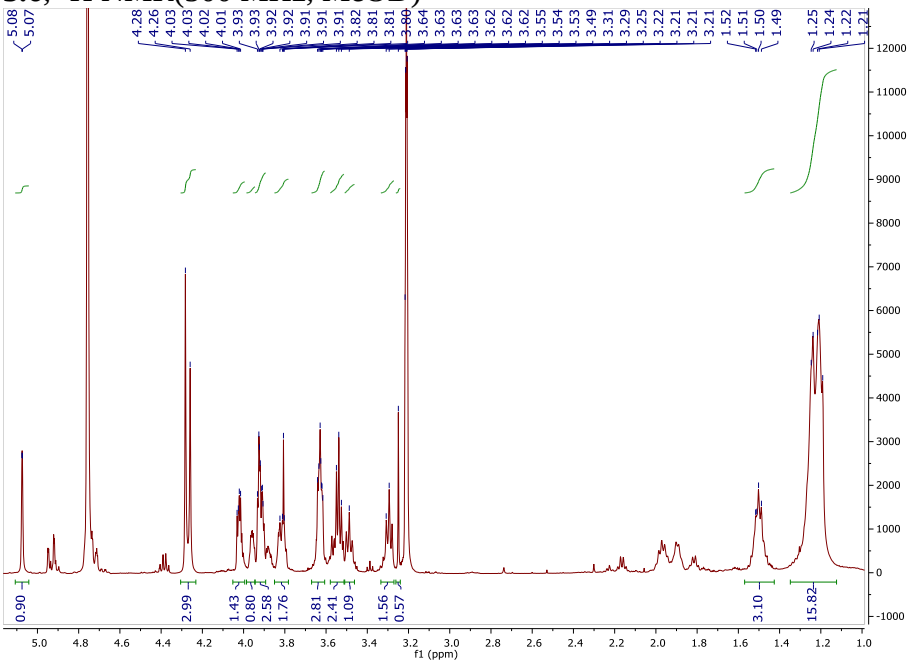
3.4e, ^{1}H NMR (400 MHz, D_2O)



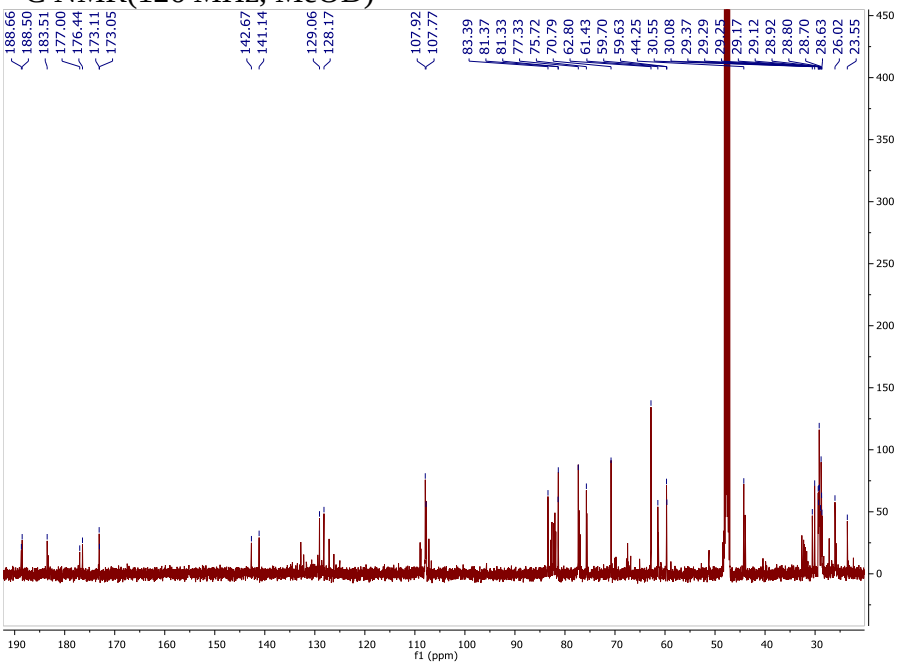
*δ 1.90 corresponds to ammonium acetate salt

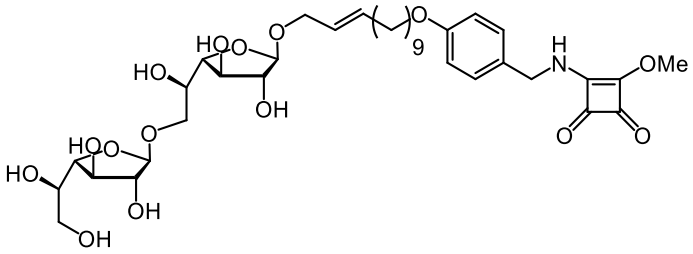


3.6, ^{1}H NMR(500 MHz, MeOD)

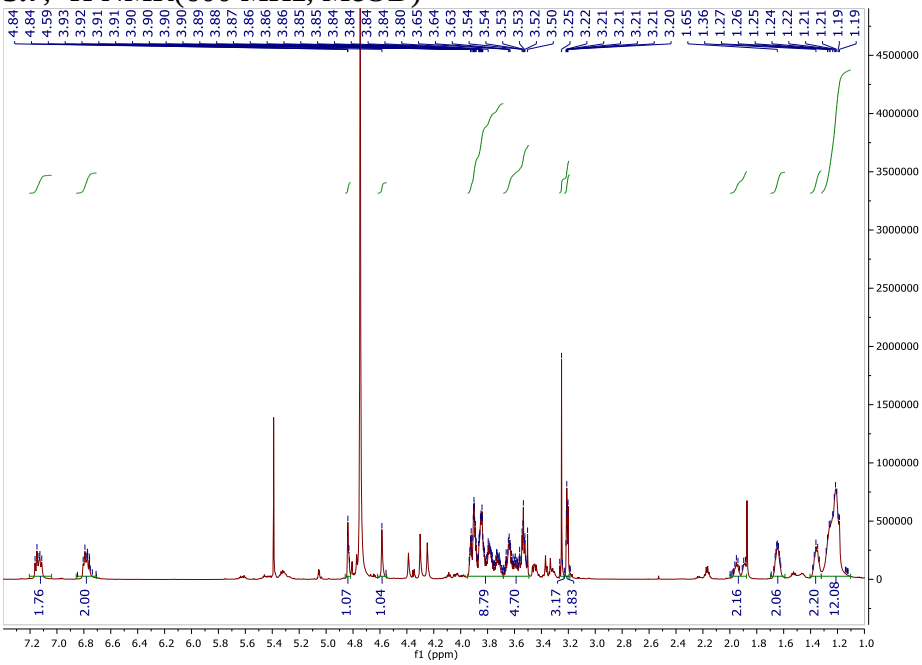


^{13}C NMR(126 MHz, MeOD)

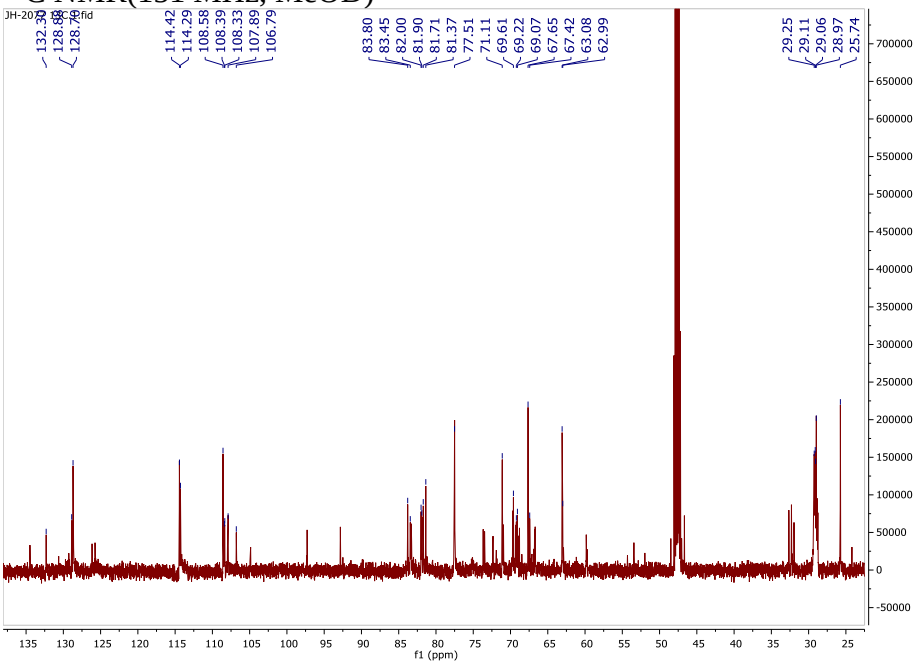


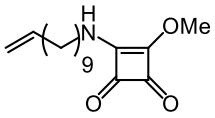


3.7, ^{1}H NMR(600 MHz, MeOD)

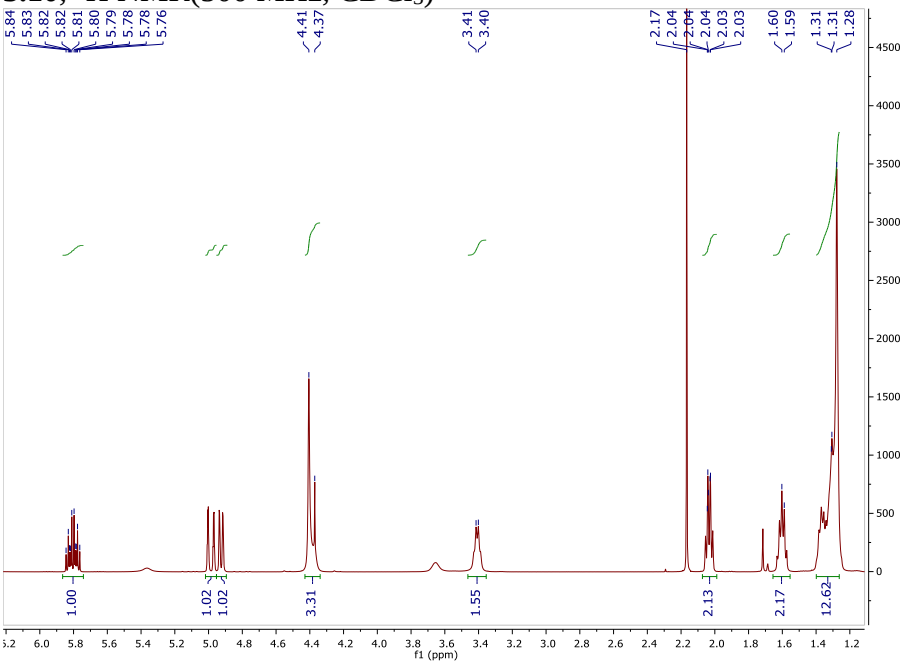


^{13}C NMR(151 MHz, MeOD)

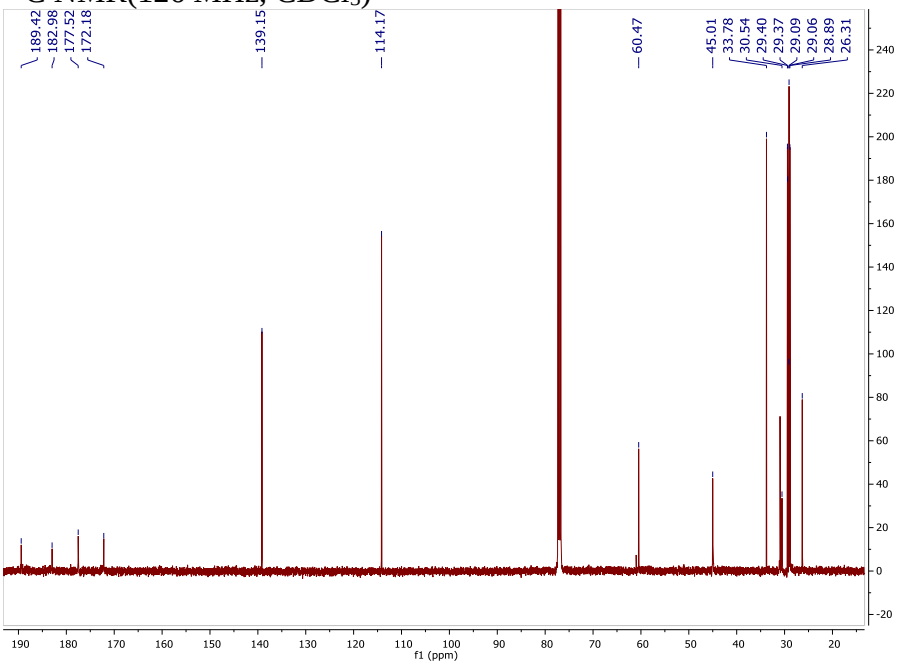


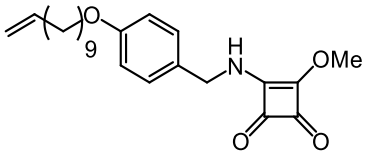


3.10, ^{1}H NMR(500 MHz, $CDCl_3$)

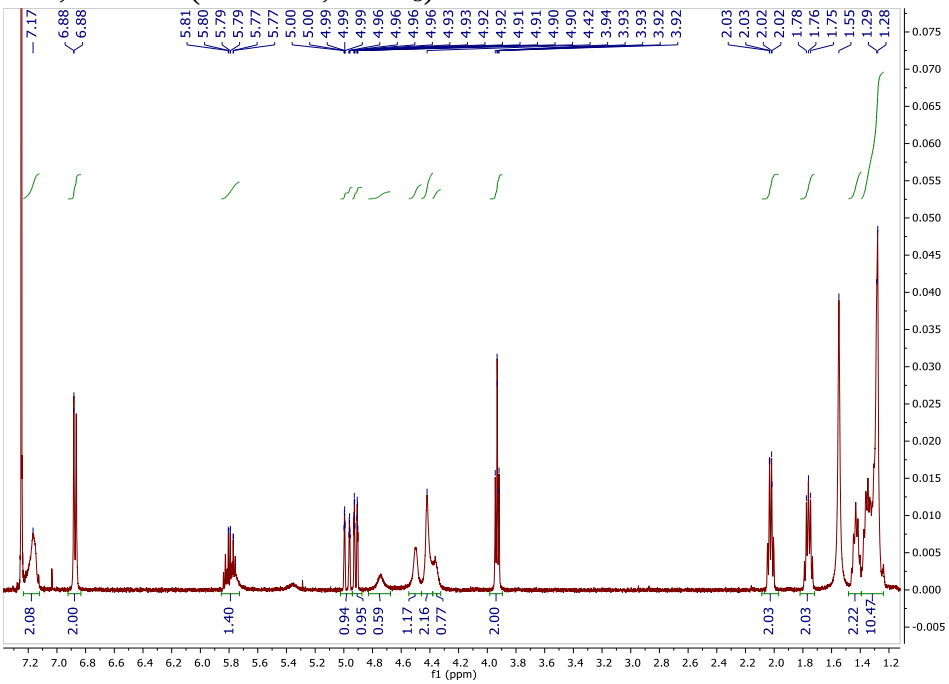


^{13}C NMR(126 MHz, $CDCl_3$)





3.11, ^{1}H NMR(500 MHz, $CDCl_3$)



Chapter 4: Synthesis of Squarate Fragments for Proteomic Screening

Contributions:
Synthesis was performed by Jordan Ho
isoTOP labeling and analysis was performed by Mikail Abbasov

4.1 Abstract

There are currently hundreds of genes in the human genome that directly contribute to disease yet have no known drug. Discovering new potential targets for future therapeutics may arguably be just as important as drug development itself. To screen for these new targets, large compound libraries have been tested against entire cell lysates in a high-throughput manner in attempts to serendipitously produce compounds that bind to a target with high potency. The issue with this strategy of induced serendipity is that the required library needed to explore all molecular complexity is so extraordinarily large that it cannot be done on any reasonable timescale. Recently, a technique known as fragment-based ligand design was has made tremendous strides in discovering new targets for future therapeutic development. Instead of attempting to tackle all molecular permutations, fragment-based ligand design emphasizes optimizing small molecular fragments with weak to moderate binding affinity and then constructing a compound by stitching these fragments together through rational design. Traditional discovery chemistry often tends to produce hits similar in structure to existing compounds, but through FBLD it is much easier to produce novel scaffolds that may interact with proteins previously deemed as "undruggable". Due to the unique structure and electronics of squaric esters, we believe that this small fragment could aid in expanding target space in lysine reactivity and ligandability within the proteome. After screening cell lysates with a library of squarate fragments, we found several unique hits, one of which includes an enzyme known as phytanoyl-CoA-dioxygenase, of PHYH. In addition to being a direct factor of a neurological disease known as Refsum's disease, PHYH is belongs in a family of oxygenases that has been deemed "undruggable" in the past.

4.2 Introduction

In addition to developing potent enzyme inhibitors as therapeutics, it is also important to expand the target space that can be addressed with novel therapeutics. Currently, much of the human and bacterial genomes still contain "undruggable" families of enzymes, many of which can contribute to disease and illness (Figure 4.1).[112–115] To expand the druggable genome, there has been much work done in developing large compound libraries and methods to validate the biological space those compounds can interact with.[116–119] This has led to novel drugs in the past; however, a downside to this approach is how large and complex combinatorial libraries must be to cover all chemical space.[120–122] Estimates of the chemical universe has been calculated to be in the vicinity of around 10^{60} compounds, yet even the largest repositories can only screen around 10^{6} compounds.[123] Not only does this not even begin to scratch the surface of potential chemical space, existing libraries tend to contain many redundancies and a lack of diversity in molecular scaffolds. With the rise of antibiotic resistance, it is more important than ever to search for new unique structures that can target previously "undruggable" enzymes.[124,125]

One approach to address this lack of diversity is through fragment-based ligand discovery, or FBLD. This technique aims to simplify molecular complexity by screening smaller fragments for weak binding as opposed to screening full elaborated molecules with high binding affinity.[115,126] The rationale behind FBLD is that by optimizing small individual binding interactions of fragments, subsequent incorporation of those fragments can produce a compound with a binding affinity comparable to the sum

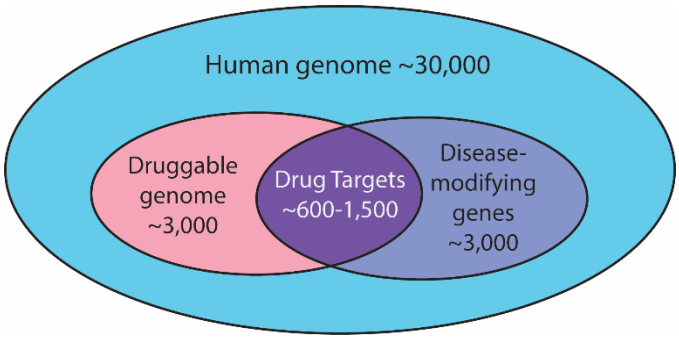


Figure 4.1. The effective number of exploitable drug targets are limited by both which genes can be targeted as well as which genes contribute to disease.

of the fragments (Figure 4.2). This reduces compound libraries into fragment libraries that contain structures of low molecular weights. Not only does this simplify computational prediction and analysis tremendously, but the fragments themselves can be thought of as building blocks that can be fitted together. Under the assumption that a larger compound is comparable to the sum of its parts, the number of structures required to cover a large target space reduces significantly. Already there has been much success in developing potent small molecule inhibitors through an FBLD approach, with several currently in phase I and phase II clinical trials.[128–130]

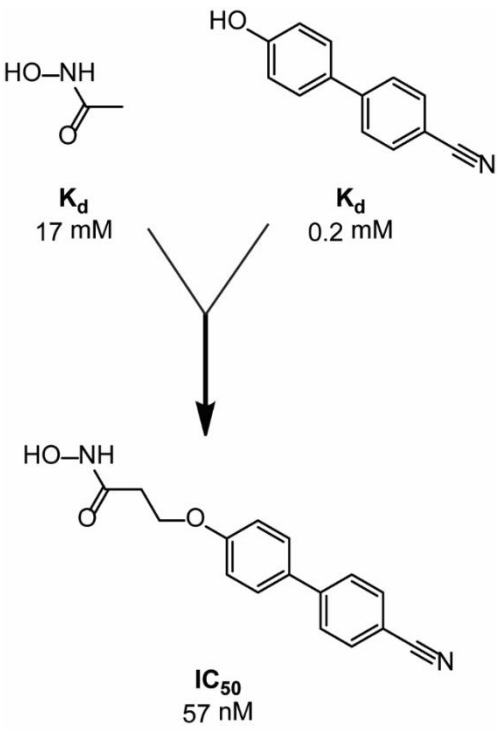
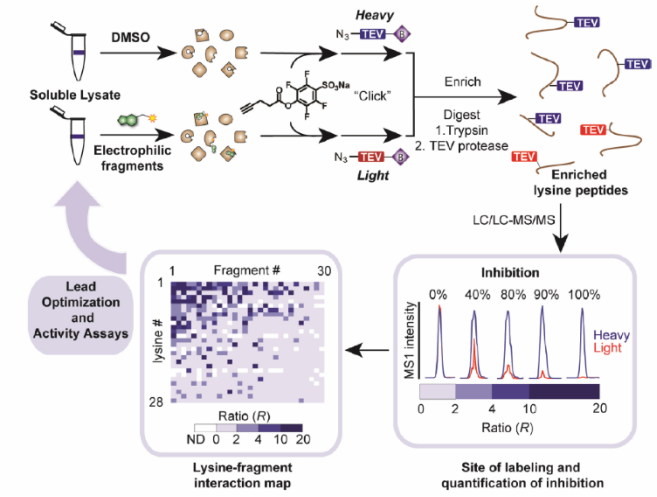


Figure 4.2. Inhibitors of matrix metalloproteinase (MMP3) with nanomolar affinity were developed by combining fragments that bound to MMP3 with millimolar affinities.[127]

We have previously shown that squarates are able to be used in the construction of targeted enzyme inhibitors, but their small molecular footprint also lend themselves to be fragments with the potential to address "undruggable" enzymes. The cyclobutenedione core offers a fragment in the form a 1,2-diketone locked into position by a cyclobutene ring. Furthermore, the two reactive vinylogous ester sites can covalently bind to proximal nucleophilic lysines enabling quantitative analysis of binding interactions. In an effort to judge the squarate's capacity to expand FBLD target space in the context of global lysine ligandability, we synthesized a library of squarates of low molecular weights and performed isotopic tandem

Scheme 4.1. isoTOP-ABPP proteomic platform for assessing lysine reactivity and ligandability.[93]



orthogonal proteolysis activity-based protein profiling (isoTOP-ABPP)[131] with this library to explore lysine reactivity in native biological systems (Scheme 4.1).

In isoTOP-ABPP, a control mixture is treated with a broad-spectrum reactive probe which modifies all nucleophilic residues of interest. These modified proteins are then conjugated via copper-catalyzed azide-alkyne cycloaddition (CuAAc) to a biotin labelled heavy isotope tobacco etch virus (TEV) tag, enriched with streptavidin, and digested. The resulting peptide fragments are then analyzed via mass spectrometry. The same is done to a sample treated with small molecule fragments but is clicked to a light isotope TEV tag instead of the heavy one. These fragments compete with the general probe and will prevent the tag from binding. This consequently lowers the signal observed for peptides containing the light tag during analysis. Thus, fragment-protein interactions can be quantified by comparing the signals of corresponding peptide pairs between the control and fragment-treated sample.

Through isoTOP-ABPP, we have found that several squarate fragments were able to covalently bind to lysines in the active site of several enzymes. One example is the modification of phytanoyl-CoA 2-dioxygenase, or PHYH, an medically relevant enzyme that is linked to a neurological condition known as Refsum's disease.[132] PHYH is also part of the family of iron-dependent oxygenases, previously thought to be "undruggable" targets.

4.3 Results and Discussion

Construction of diverse squaric ester fragment library. Due to the nature of squaric esters, a library of squaric ester fragments was synthesized through straightforward amine installation onto dimethyl squarate (Table 4.1). This allowed us to create a diverse library to evaluate using squarates in FBLD easily and quickly by simply varying the appended amine. Because many of the compounds used in previous kinetic studies were simple model squarates that fit the criteria for FBLD, we included a few of those as well. Many of the compounds contain flat aromatic groups and ring systems, but we also included more aliphatic motifs to explore more complex 3-

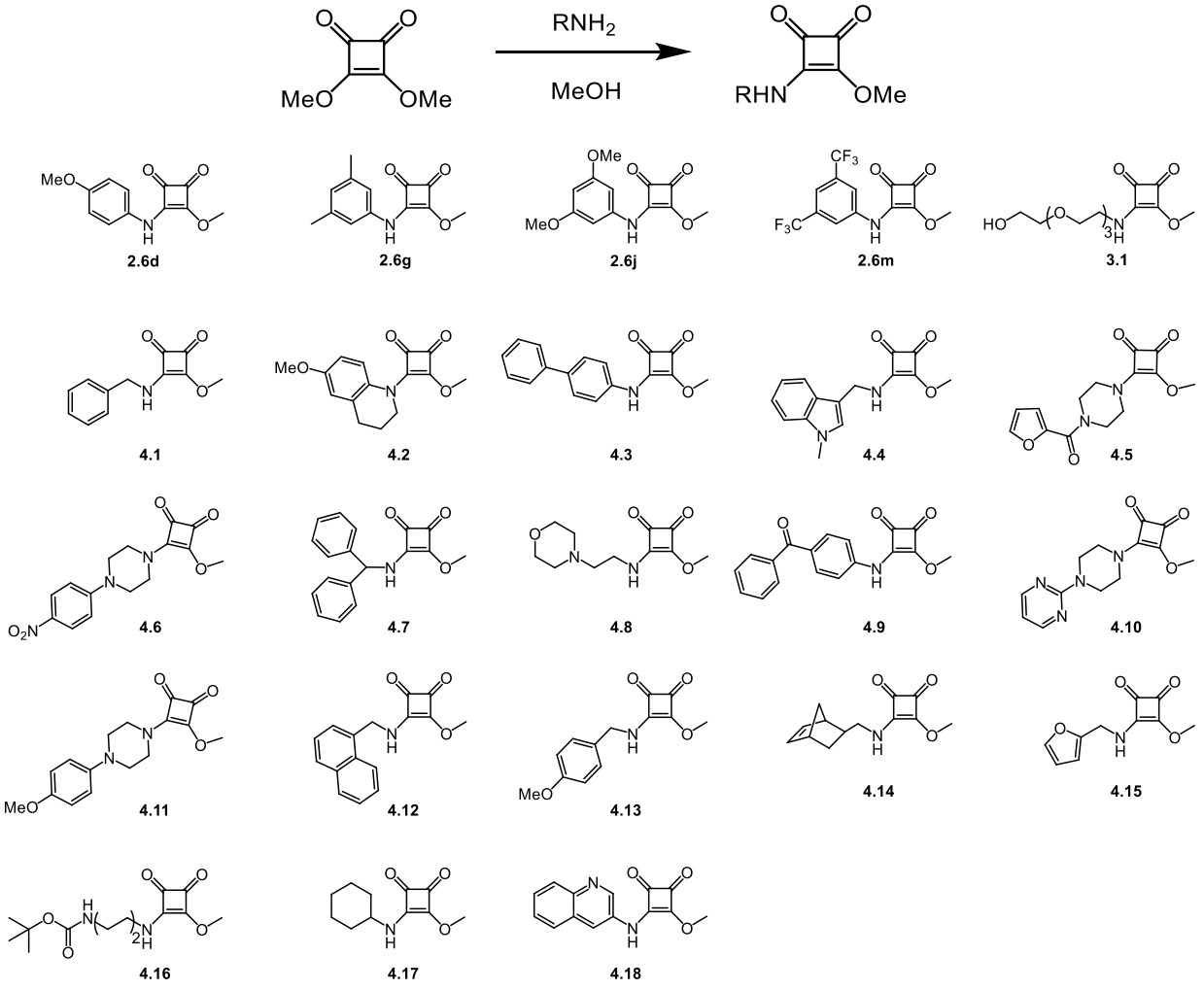


Figure 4.3. Squarate fragment library for assessing lysine reactivity and ligandability in FBLD.

D geometries as well as adding extra flexibility to fragments. In addition, we varied the hydrophobicity of the library with the inclusion of different polar groups such as morpholine and piperazine.

Initial isoTOP-ABPP screens yield several unique hits. Preliminary analysis of cell lysates treated with compounds **2.6m** and **4.2** found hits across several different enzymes, some exclusively to one compound and one that showed significant binding to both (Table 4.1). Notably, compound **2.6m** was found to have the most hits, with significant inhibition to proteins including messenger proteins, DNA repair proteins, and oxygenases. One enzyme of interest that bound tightly to a squarate fragment was phytanoyl-CoA-dioxygenase (PHYH) (Figure 4.4). With squarate **2.6m**, we observed over 95% engagement of active site lysine K120. PHYH belongs to a family of oxygenases that has been known to be "undruggable" and difficult to develop small molecule therapeutics for. Mutations in PHYH in humans can cause accumulation of phytanic acid in the system. This accumulation can then lead to a neurological disease known

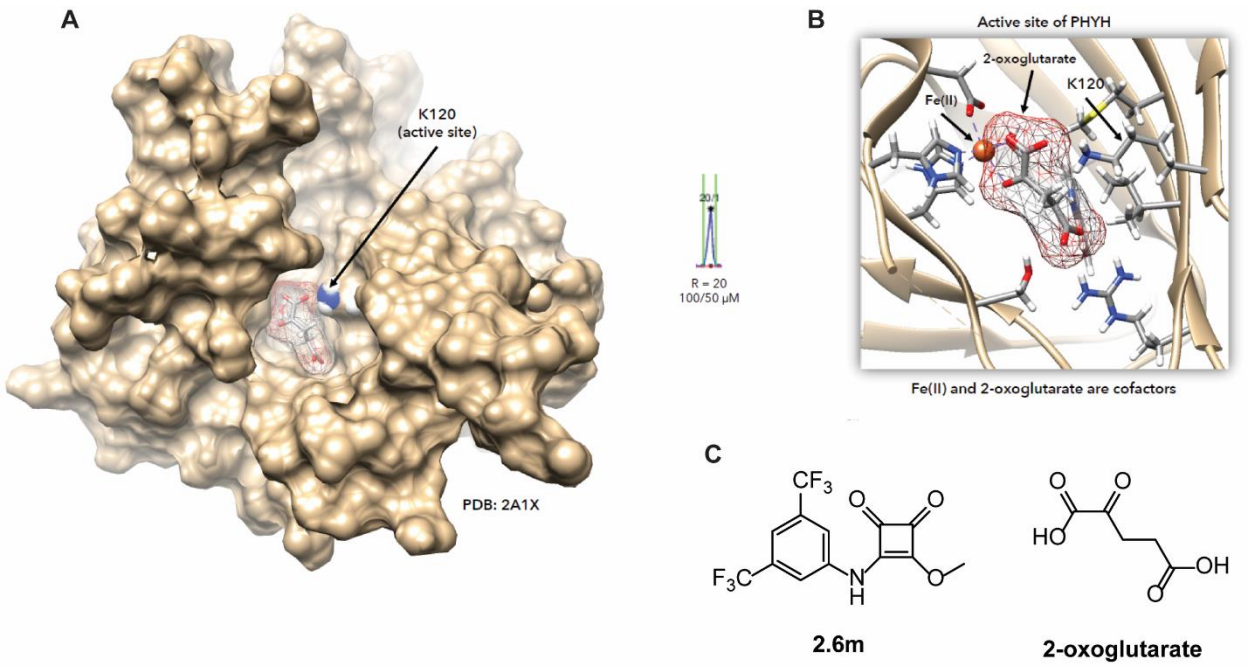


Figure 4.4. (A) Surface model of PHYH and (B) crystal structure of PHYH active site with Fe(II) and 2-oxoglutarate bound.[133] (C) Structural comparison of **2.6m** to 2-oxoglutarate.

			Spectral	4.2		2.6m	
Protein	Description	Lysine	Counts	100 µM	50 µM	100 µM	50 µM
ABCE1	ATP-binding cassette sub-family E member 1	191	46				
ACBD3	ACBD3 Golgi resident protein GCP60	117	136				
ABCF3	ATP-binding cassette sub-family F member 3	457	10				
XRCC6	X-ray repair cross-complementing protein 6	351	208				
RNF214	RING finger protein 214	582	56				
RNF214	RING finger protein 214	554	10				
XRCC4	DNA repair protein XRCC4	187	8				
TRIP13	Pachytene checkpoint protein 2 homolog	288	64				
SLC25A6	ADP/ATP translocase 3	23	22				
COMMD10	COMM domain-containing protein 10	56	86				
COPS8	COP9 signalosome complex subunit 8	14	20				
SYNJ2	Synaptojanin-2	477	62				
GBAS	Protein NipSnap homolog 2	75	142				
PHYH	Phytanoyl-CoA dioxygenase, peroxisomal	120	24				
CALM3	Calmodulin	76	18				
SIN3A	Paired amphipathic helix protein Sin3A	155	154				
AKT1	RAC-alpha serine/threonine-protein kinase	163	26				
PARVA	Alpha-parvin	265	6				
PRKAG1/2	5'-AMP-activated protein kinase subunit gamma-1/2	243/475	118/106				
CUL1	Cullin-1	459	30				
HSP90B1	Endoplasmin	404	12				
FECH	Ferrochelatase, mitochondrial	304	16				
AKR1C1/3	Aldo-keto reductase family 1 member C1/C3	270	66/82				

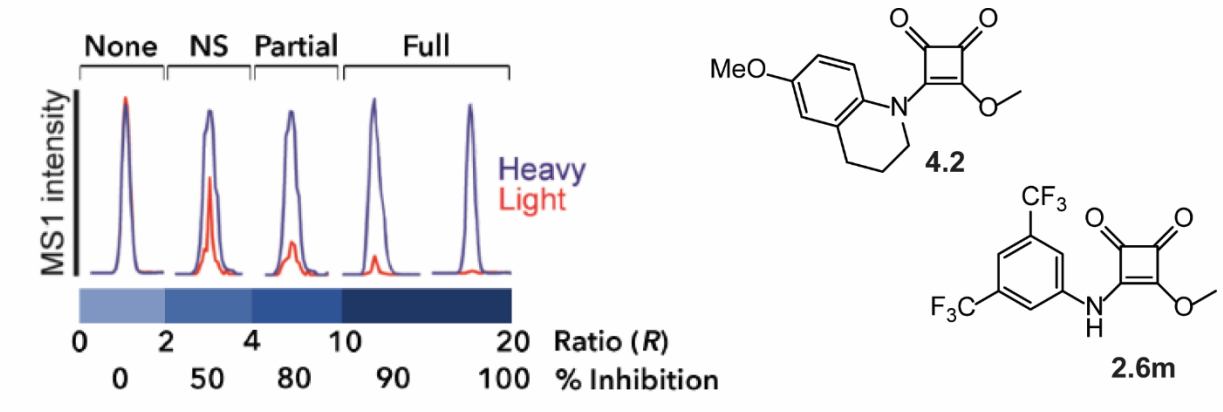


Table 4.1. Table of proteins with lysines liganded by **4.2** and **2.6m** at 100 and 50 µM. 100µM conditions were tested in triplicate and 50µM conditions were tested in duplicate. Each square represents an individual result for a given protein. Intensity of blue directly corresponds to percent inhibition by compound i.e. binding interaction between protein and ligand.

as Refsum disease which can cause deafness, ataxia, cardiac arrhythmia, and retinal degeneration.[134]

To our knowledge, while the mechanism of PHYH has been thoroughly characterized, there is currently no synthetic inhibitor or compound designed to alter PHYH function. PHYH uses 2-oxoglutarate as a co-factor which possesses a similar 1,2-diketone moiety seen in squarates. From this similarity, we hypothesize that in addition to the binding interaction afforded by the bis-trifluoromethyl phenyl ring, the diketone of the squarate is also contributing some binding interaction as well as covalently bonding to K120. The 1,2-diketone moiety not only appears in 2-oxoglutarate but many other metabolites, some necessary for important biological pathways such as the Krebs cycle. As we move forward to analyze the global reactivity of the rest of the library, we are optimistic in our efforts of uncovering novel lysines that can be targeted in developing new molecular probes.

4.4 Conclusions

Currently, there remains a large portion of the human and bacterial genome linked to diseases and have no therapeutic for. Many of these genes encode enzymes that are difficult to perturb and have been deemed "undruggable" with existing technology. In an effort to expand the target space for synthetic therapeutics, FBLD has emerged as an efficacious means to screen small molecular fragments in a high-throughput manner to uncover this new target space.

We sought to employ squaric esters in FBLD and synthesized a diverse library of squarate fragments to screen for new targets. Preliminary data gathered from two squarate fragments have yielded several novel hits. One of which was PHYH, which has been thought of as "undruggable" in the past and utilizes a cofactor that is structurally similar to the squaric ester. The 1,2-diketone in the squaric ester that we believe is attributing to the total binding interaction of the fragment is not only seen in 2-oxoglutarate, but many other important metabolites in

important biological pathways, including the Krebs cycle. Consequently, as we have only tested two squarates in our library, we have confidence in continuing to uncover novel ligandable lysines and expanding the therapeutic target space with the remainder of our library.

4.5 Experimental Details

isoTOP-ABPP sample preparation.

***In vitro* covalent fragment treatment for isoTOP-ABPP.** All compounds were made up as solutions in DMSO (100×) and were used at a final concentration of 50 μM for activated esters and 100 μM for guanidinylating agents. For each profiling sample, 0.5 ml of lysate was treated with 5 μl of the 100× compound stock solution or 5 μl of DMSO. Samples were treated with activated esters for 1 h and with guanidinylating agents for 4 h.

STP-alkyne labeling and click chemistry. For concentration-dependent reactivity measurements by isoTOP-ABPP, 0.5 ml proteome aliquots were treated at ambient temperature with 1 mM STP-alkyne **1** (5 μl of 100 mM stock in DMSO) and 0.1 mM STP alkyne (5 μl of 10 mM stock in DMSO), respectively. For competitive isoTOP-ABPP, after *in vitro* fragment treatment (detailed above), the samples were labeled for 1 h at ambient temperature with 0.1 mM STP-alkyne (5 μl of 10 mM stock in DMSO). Samples were conjugated by copper mediated azide-alkyne cycloaddition (CuAAC) to either the light (1 mM STP-alkyne or fragment treated) or heavy (0.1 mM STP-alkyne or DMSO treated) TEV tags (10 μl of 5 mM stocks in DMSO, final concentration = 100 μM) using tris(2-carboxyethyl)phosphine hydrochloride (TCEP; fresh 50× stock in water, final concentration = 1 mM), TBTA ligand (17× stock in DMSO:tbutanol 1:4, final concentration = 100 μM) and CuSO4 (50× stock in water, final concentration = 1 mM). The samples were allowed to react for 1 h at room temperature, at which point the proteins from combined light and heavy samples were precipitated by chloroform-

methanol extraction. The pellets were solubilized in PBS containing 1.2 % SDS (1 ml) with sonication and heating (5 min, 95 ºC) and any insoluble material was removed by an additional centrifugation step at ambient temperature (5,000*g*, 10 min).

Streptavidin enrichment. For each sample, 100 μl of streptavidin-agarose beads slurry (Pierce, 20349) was washed in 10 ml PBS (3×) and then resuspended in 6 ml PBS. The SDS solubilized proteins were added to the suspension of streptavidin-agarose beads and the bead mixture was rotated for 3 h at ambient temperature. After incubation, the beads were pelleted by centrifugation (2,800*g*, 3 min) and were washed (1 × 10 ml 0.2 % SDS in PBS, 2 × 10 ml PBS and 2 × 10 ml water).

Trypsin and TEV digestion. The beads were transferred to Eppendorf tubes with 1 ml PBS, centrifuged (20,000*g*, 1 min), and resuspended in PBS containing 6 M urea (500 μl). To this was added 10 mM DTT (25 μl of a 200 mM stock in water) and the beads were incubated at 65 ºC for 15 min. 20 mM iodoacetamide (25 μl of a 400 mM stock in water) was then added and allowed to react at 37 ºC for 30 min with shaking. The bead mixture was diluted with 950 μl PBS, pelleted by centrifugation (20,000*g*, 1 min), and resuspended in PBS containing 2M urea (200 μl). To this was added 1 mM CaCl2 (2 μl of a 200 mM stock in water) and trypsin (2 μg, Promega, sequencing grade in 4 μl trypsin resuspension buffer) and the samples were allowed to digest overnight at 37 oC with shaking. The beads were separated from the digest with Micro Bio-Spin columns (Bio-Rad) by centrifugation (800*g*, 30 sec), washed (2 × 1 ml PBS and 2 × 1 ml water) and then transferred to fresh Eppendorf tubes with 1 ml water. The washed beads were washed once further in 140 μl TEV buffer (50 mM Tris, pH 8, 0.5 mM EDTA, 1 mM DTT) and then resuspended in 140 μl TEV buffer. 5 μl TEV protease (80 μM stock solution) was

added and the reactions were rotated overnight at 30 oC. The TEV digest was separated from the beads with Micro Bio-Spin columns by centrifugation (8,000*g*, 3 min) and the beads were washed once with water (100 μl). The samples were then acidified to a final concentration of 5% (v/v) formic acid and stored at -80 ^{o}C prior to analysis.

Liquid-chromatography-mass-spectrometry (LC-MS) analysis of isoTOP-ABPP samples. TEV digests were pressure loaded onto a 250 μm (inner diameter) fused silica capillary columns packed with C18 resin (Aqua 5 μm, Phenomenex). The samples were analyzed by multidimensional liquid chromatography tandem mass spectrometry (MudPIT), using an LTQVelos Orbitrap mass spectrometer (Thermo Scientific) coupled to an Agilent 1200-series quaternary pump. The peptides were eluted onto a biphasic column with a 5 μm tip (100 μm fused silica, packed with C18 (10 cm) and bulk strong cation exchange resin (3 cm, SCX, Phenomenex)) in a 5-step MudPIT experiment, using 0%, 30%, 60%, 90%, and 100% salt bumps of 500 mM aqueous ammonium acetate and using a gradient of 5–100% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B: 20% water, 80% acetonitrile, 0.1% formic acid) as has been described1. Data was collected in data-dependent acquisition mode with dynamic exclusion enabled (20 s, repeat count of 2). One full MS (MS1) scan (400-1800 m/z) was followed by 30 MS2 scans (ITMS) of the nth most abundant ions.

Peptide and protein identification. The MS2 spectra were extracted from the raw file using RAW Xtractor (version 1.9.9.2; available at http://fields.scripps.edu/downloads.php). MS2 spectra were searched using the ProLuCID algorithm (publicly available at http://fields.scripps.edu/downloads.php)2 using a reverse concatenated, nonredundant variant of the Human UniProt database (release-2012_11). Cysteine residues were searched with a static modification for carboxyamidomethylation (+57.02146). For all competitive and reactivity

profiling experiments, lysine residues were searched with up to one differential modification for either the light or heavy TEV tags (+464.2491 or +470.26331, respectively). Peptides were required to have at least one tryptic terminus and to contain the TEV modification. ProLuCID data was filtered through DTASelect (version 2.0) to achieve a peptide false-positive rate below 1%.

Differential labeling analysis of residues labeled by electrophilic fragments. For analysis of the residues labeled the fragments, peptide and protein identification was conducted as detailed above with differential modification for either the light or heavy TEV tags (+464.2491 or +470.26331, respectively) allowed on lysine, arginine, aspartate, glutamate, histidine, serine, threonine, tyrosine, asparagine, glutamine and tryptophan. Cysteine was searched with a differential modification for either the light or heavy TEV tags (+413.24185 and +407.22764, respectively).

***R* value calculation and processing.** The ratios of light and heavy MS1 peaks for each unique peptide were quantified with in-house CIMAGE software3,4, using default parameters (3 MS1 acquisitions per peak and signal to noise threshold set to 2.5). For reactivity measurements by isoTOP-ABPP, the *R* value was calculated from the ratio of MS1 peak areas, comparing the 1 mM STP alkyne sample (light TEV tag) with the 0.1 mM STP alkyne sample (heavy TEV tag). For competitive isoTOP–ABPP, the *R* value was calculated from the ratio of MS1 peak areas, comparing the DMSO treated sample (heavy TEV tag) with the compound treated sample (light TEV tag). For peptides that showed a $\geqslant$ 95% reduction in MS1 peak area in both reactivity and compound treated samples a maximal ratio of 20 was assigned. Ratios for unique peptide entries are calculated for each experiment; overlapping peptides with the same modified lysine (for example, different charge states, MudPIT chromatographic steps or tryptic termini) are

grouped together and the median ratio is reported as the final ratio (*R*). The peptide ratios reported by CIMAGE were further filtered to ensure the removal or correction of low-quality ratios in each individual data set. The quality filters applied were the following: removal of half tryptic peptides; for ratios with high standard deviations from the median (90% of the median or above) the lowest ratio was taken instead of the median; removal of peptides with $R = 20$ and only a single MS2 event triggered during the elution of the parent ion; manual annotation of all the peptides with ratios of 20, removing any peptides with low quality elution profiles that remained after the previous curation steps (only done for competitive isoTOP-ABPP).

Cross-data processing for fragment screening. For compound treated samples, biological replicates of the same condition were averaged, if the standard deviation was below 60% of the mean; otherwise, for lysines with at least one *R* value <4 for a particular compound, the lowest value of the ratio set was taken. For lysines, where all *R* values for a particular compound were $\geqslant 4$, the average was reported. For peptides containing several possible modified lysines, the lysine with the highest number of quantification events was used for analysis and the remaining, redundant peptides were reported as alternative modification sites. Peptides included in the aggregate dataset (those used for further bioinformatics and statistical analyses) were required to have been quantified in 2 experiments for competitive isoTOP-ABPP. Lysines were categorized as liganded, if they had at least one ratio $R \geqslant 4$ (hit fragments). For liganded lysines with $R = 20$ for all liganding events, lysines were required to have been quantified with $R = 20$ in two separate experiments and were further required to have been quantified with R <20 in one additional experiment.

Cross-data processing for reactivity profiling. For reactivity profiling, the median of biological replicates of the same condition and cell-line was calculated. For peptides containing

several possible modified lysines, the lysine with the highest number of quantification events was used for analysis and the remaining, redundant peptides were reported as alternative modification sites. Peptides were required to be detected in at least one 1 mM vs 0.1 mM and one 0.1 mM vs 0.1 mM data set with the latter *R* value being smaller than 2.5. All ratios derived from soluble reactivity experiments were averaged. If the lysine was not detected in any soluble fraction, the *R* value from the membrane fraction was taken. Additionally, all membrane-only lysines with reactivity values were further required to have been detected in at least one 0.1 mM vs 0.1 mM membrane profiling experiment. If the final reactivity value was >10, it was set to 10. Lysines were categorized based on the *R* values (hyper-reactive: R<2; moderately-reactive: R=2-5; lowreactive: R>5).

Heatmap generation. Heat maps were generated in R (v.3.1.3) using the heatmap.2 algorithm.

DrugBank. Proteins were queried against the DrugBank database (v. 5.0.3 released on 2016-10-24; group "All") and separated into DrugBank and non-DrugBank proteins.

Functional annotation of lysines. Lysines proximal to functional sites were defined as any lysine with a Cα atom within 10 A of an annotated ligand binding site in an X-ray or NMR structure. Custom Python scripts were developed to collect relevant NMR and X-ray structures, including any co-crystallized small molecules, from the RCSB Protein Data Bank (PDB)5. The following small molecules were excluded from this analysis: MES, EDO, DTT, BME, ACR, ACY, ACE and MPD. Histograms of the frequency of functional sites for hyper-reactive, moderatelyreactive and low reactive lysines were calculated.

Analysis of lysine conservation. Sequences of all human proteins were downloaded from UniProtKB. Orthologs of human proteins were obtained using the HUGO Gene Name Consortium's database, or the DRSC Integrative Ortholog Prediction Tool, provided by Harvard

Medical School. Clustal Omega was used to generate multiple sequence alignments for each human protein and its orthologs, and in-house software was used to calculate the conservation of individual lysines. Only proteins with orthologues in all five organisms evaluated (*M. musculus*, *X. laevis*, *D. malanogaster*, *C. elegans* and *D. rerio)* were considered for the conservation analysis.

Competitive gel-based ABPP and apparent IC_{50} values. 50 μl of soluble proteome (1 mg ml^{-1}) expressing the indicated protein were treated with fragment electrophiles (1 μl of 50× stock solution in DMSO) at ambient temperature for 1 h. The indicated probe (fluorophore or alkynecontaining, 1 μl of a 500 μM solution, final concentration = 10 μM) was then added and allowed to react for an additional 1 h. CuAAC and in-gel fluorescence analysis were performed as described above. For quantification of inhibition and apparent IC50 determination, the percentage of labeling was determined by quantifying the integrated optical intensity of the bands using ImageLab 5.2.1 software (BioRad). Nonlinear regression analysis was used to determine the IC_{50} values from a dose-response curve generated using GraphPad Prism 7.

Lysine reactivity and ligandability comparison. Lysines found in both the reactivity and ligandability data sets were sorted based on their reactivity values (lower ratio indicates higher reactivity). The moving average of the percentage of total liganded lysines within each reactivity bin (step-size 200) was taken.

Compound Synthesis

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Tetrahydrofuran (THF) was distilled from sodium/benzophenone, methanol (MeOH) was distilled from magnesium, and dichloromethane (DCM) was distilled from calcium hydride. Nuclear magnetic

resonance spectra were recorded on a 400 MHz spectrometer (acquired at 400 MHz for ^{1}H and 100 MHz for ^{13}C), or 500 MHz spectrometer (acquired at 500 MHz for ^{1}H and 125 MHz for ^{13}C), or 600 MHz spectrometer (acquired at 600 MHz for ^{1}H and 150 MHz for ^{13}C). Chemical shifts are reported relative to tetramethylsilane or residual solvent peaks in parts per million ($CHCl_3$: ^{1}H 7.27, ^{13}C 77.23; MeOH: ^{1}H 3.31, ^{13}C 49.15; D_2O: ^{1}H 4.79; DMSO: ^{1}H 2.50, ^{13}C 39.52). High-resolution mass spectra (HRMS) were obtained on an electrospray ionization time-of-flight (ESI-TOF) mass spectrometer.

General procedure for the synthesis of squarate fragments 4.1-4.18

1.0 eq. amine and 1.2 eq. dimethyl squarate stirred in MeOH. Solid precipitate was collected and washed with cold MeOH. In cases where no or little precipitate was formed, reaction mixture was purified via silica chromatography.

3-(benzylamino)-4-methoxycyclobut-3-ene-1,2-dione **(4.1)**

Using 1.7mL of benzylamine with 2.14g dimethylsquarate, **4.1** was purified via silica gel chromatography (1:4 ACN:DCM) and isolated in 85% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.39 (t, J=7.13 Hz, 2H), 7.34 (t, J=7.13 Hz, 1H), 7.28 (br, 2H), 6.05 (br, 0.7H, rotamer 1), 5.46 (br, 0.3H, rotamer 2), 4.84 (br, 0.6H, rotamer 2), 4.59 (br, 1.4H, rotamer 1), 4.42 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) (chemical shifts of both rotamers listed) δ 46.81, 47.40, 60.00, 60.28, 127.37, 127.46, 127.51, 128.63, 137.97, 138.40, 171.77, 172.29, 177.11, 177.80, 182.28, 182.54, 189.19, 189.58

3-methoxy-4-(6-methoxy-3,4-dihydroquinolin-1(2H)-yl)cyclobut-3-ene-1,2-dione **(4.2)**

Using 280mg 6-methoxytetrahydroquinoline with 210mg dimethylsquarate, **4.2** was purified via silica gel chromatography (1:1 Hex:EtOAc) and isolated in 70% yield. ^{1}H NMR (500 MHz,

$CDCl_3$) δ 6.75 (br, 1H), 6.66 (s, 1H), 4.44 (s, 3H), 4.12 (br, 1H), 3.78 (s, 3H), 2.87 (t, J=6.63 Hz, 2H), 2.05 (p, J=6.10 Hz, 2H), 1.55 (s, 1H).

3-([1,1'-biphenyl]-4-ylamino)-4-methoxycyclobut-3-ene-1,2-dione **(4.3)**

Using 169mg 4-aminobiphenyl with 170mg dimethyl squarate, **4.3** was isolated in 76% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.61 (d, J=8.1 Hz, 2H), 7.57 (d, J=7.42 Hz, 2H), 7.45 (t, J=7.42 Hz, 2H), 7.36 (t, J=7.65 Hz, 2H), 7.34 (br, 1H), 4.53 (s, 3H). ^{13}C NMR (151 MHz, d_6-DMSO) 139.35, 137.41, 135.70, 128.92, 127.30, 127.23, 126.33, 119.89, 60.60.

3-methoxy-4-(((1-methyl-1H-indol-3-yl)methyl)amino)cyclobut-3-ene-1,2-dione **(4.4)**

Using 162mg of (1-mehtylindolin-3-yl)methanamine with 170mg dimethyl squarate, **4.4** was isolated in 72% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.56 (br, 1H), 7.37 (d, J=8.12 Hz, 1H), 7.32 (t, J=7.34 Hz, 1H), 7.20 (t, J=7.42 Hz, 1H), 7.08 (br, 1H), 5.80 (br, N-H rotamer 1, 0.66H), 5.43 (br, N-H rotamer 2, 0.33H), 5.04 (br, rotamer 2, 0.67H), 4.80 (br, rotamer 1, 1.33H), 4.50 (br, rotamer 1, 2H), 4.36 (br, rotamer 2, 1H), 3.82 (s, 3H).

3-(4-(furan-2-carbonyl)piperazin-1-yl)-4-methoxycyclobut-3-ene-1,2-dione **(4.5)**

Using 180mg of furoyl-piperazine with 170mg dimethyl squarate, **4.5** was isolated in 80% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.50 (s, 1H), 7.11 (s, 1H), 6.52 (s, 1H), 4.42 (s, 3H), 3.98 (br, 2H), 3.93 (br, 4H), 3.66 (br, 2H).

3-methoxy-4-(4-(4-nitrophenyl)piperazin-1-yl)cyclobut-3-ene-1,2-dione **(4.6)**

Using 207mg 4-nitrophenyl piperazine with 170mg dimethyl squarate, **4.6** was isolated in 44% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.17 (d, J=8.62 Hz, 2H), 6.88 (d, J=8.33 Hz, 2H), 4.43 (s, 3H), 4.07 (br, 2H), 3.75 (br, 2H), 3.52 (br, 4H).

3-(benzhydrylamino)-4-methoxycyclobut-3-ene-1,2-dione **(4.7)**

Using 0.17mL benzhydrylamine with 170mg dimethyl squarate, **4.7** was isolated in 39% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.38 (t, J=7.43 Hz, 4H), 7.33 (t, J=6.88 Hz, 2H), 7.24 (d, J=7.43 Hz, 4H), 6.16 (br, rotamer 1, 0.5H), 5.95 (br, rotamer 2, 0.5H), 4.32 (s, 3H).

3-methoxy-4-((2-morpholinoethyl)amino)cyclobut-3-ene-1,2-dione **(4.8)**

Using 0.13mL 4(2-aminoethyl)-morpholine with 170mg dimethyl squarate, **4.8** was purified through a silica plug (1:1 EtOAc:ACN) in 53% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 4.40 (s, 3H), 3.71 (t, J=4.48 Hz, 4H), 3.51 (br, 1H), 2.57 (t, J=5.53 Hz, 2H), 2.48 (br, 1H).

3-((4-benzoylphenyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(4.9)**

Using 197mg of 4-aminobenzophenone with 170mg dimethyl squarate, **4.9** was isolated in 42% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.87 (d, J=8.56 Hz, 2H), 7.77 (d, J=7.16 Hz, 2H), 7.60 (t, J=7.74 Hz, 1H), 7.50 (t, J=7.86 Hz, 2H), 7.40 (d, J=8.09 Hz, 2H), 4.55 (s, 3H).

3-methoxy-4-(4-(pyrimidin-2-yl)piperazin-1-yl)cyclobut-3-ene-1,2-dione **(4.10)**

Using 0.14mL N-pyrimidyl-piperazine with 170mg dimethyl squarate, **4.10** was isolated in 97% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.34 (d, J=4.67 Hz, 2H), 6.59 (t, J=4.78 Hz, 1H), 4.42 (s, 3H), 3.96 (br, 6H), 3.63 (br, 2H).

3-methoxy-4-(4-(4-methoxyphenyl)piperazin-1-yl)cyclobut-3-ene-1,2-dione **(4.11)**

Using 192mg N-(4-methoxyphenyl)-piperazine with 170mg dimethyl squarate, **4.11** was isolated in 93% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.91 (d, J=9.25 Hz, 2H), 6.86 (d, J=9.25 Hz, 2H), 4.41 (s, 3H), 4.06 (br, 2H), 3.78 (s, 3H), 3.74 (br, 2H), 3.14 (br, 4H).

3-methoxy-4-((naphthalen-1-ylmethyl)amino)cyclobut-3-ene-1,2-dione **(4.12)**

Using 0.15mL 1-naphthylene methylamine with 170mg dimethyl squarate, **4.12** was isolated in 39% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.92 (d, J=8.12 Hz, 1H), 7.87 (d, J=9.76 Hz, 1H), 7.60 (t, J=8.12 Hz, 1H), 7.56 (t, J=7.09 Hz, 1H), 7.47 (t, J=6.47 Hz, 2H), 5.98 (br, N-H rotamer 1,

0.66H), 5.47 (br, N-H, rotamer 2, 0.33H), 5.32 (br, rotamer 2, 0.66H), 5.06 (br, rotamer 1, 1.33H), 4.46 (s, rotamer 1, 2H), 4.35 (s, rotamer 2, 1H).

3-methoxy-4-((4-methoxybenzyl)amino)cyclobut-3-ene-1,2-dione **(4.13)**

Using 0.13mL 4-methoxybenzylamine with 170mg dimethyl squarate, **4.13** was isolated in 60% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.20 (br, 2H), 6.89 (d, J=8.78 Hz, 2H), 6.06 (br, N-H, rotamer 1, 0.66H), 5.43 (br, N-H, rotamer 2, .33H), 4.76 (br, rotamer 2, 0.66H), 4.52 (br, rotamer 1, 1.33H), 4.43 (br, rotamer 1, 2H), 4.39 (br, rotamer 2, 1H).

3-((((1R,2S,4R)-bicyclo[2.2.1]hept-5-en-2-yl)methyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(4.14)**

Using 123mg norbornenemethanamine with 170mg dimethyl squarate, **4.14** was isolated in 35% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 6.23 (dd, J=2.42, 5.72 Hz, 1H), 5.94 (br, 1H), 4.40 (s, 3H), 3.15 (br, 2H), 2.87 (s, 2H), 2.30 (br, 1H), 1.90 (ddd, J=3.73, 9.13, 11.68 Hz, 1H), 1.50 (d, J=8.16 Hz, 1H), 1.29 (d, J=8.16 Hz, 1H).

3-((furan-2-ylmethyl)amino)-4-methoxycyclobut-3-ene-1,2-dione **(4.15)**

Using 0.09mL furfurylamine with 170mg dimethyl squarate, **4.15** was isolated in 96% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.40 (s, 1H), 6.36 (s, 1H), 6.31 (s, 1H), 5.93 (br, N-H, rotamer 1, 0.66H), 5.50 (br, N-H, rotamer 2, 0.33H), 4.82 (br, rotamer 2, 0.66H), 4.58 (br, rotamer 1, 1.33H), 4.42 (br, 3H).

tert-butyl (2-((2-methoxy-3,4-dioxocyclobut-1-en-1-yl)amino)butyl)carbamate **(4.16)**

Using 0.38mL N-boc-putrescine with 296mg dimethyl squarate, **4.16** was purified via silica gel chromatography (2:1 → 1:1 Hex:EtOAc). ^{1}H NMR (600 MHz, $CDCl_3$) δ 4.43 (br, 3H), 3.73 (br,

rotamer 1, 0.66H), 3.47 (br, rotamer 1, 1.33H), 3.17 (q, J=6.55 Hz, 2H), 1.67 (p, J=7.15 Hz, 2H), 1.57 (br, 2H), 1.47 (s, 9H).

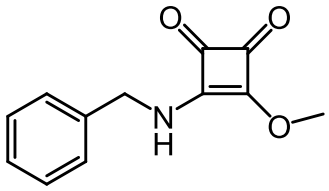
3-(cyclohexylamino)-4-methoxycyclobut-3-ene-1,2-dione **(4.17)**

Using 0.11mL cyclohexylamine with 170mg dimethyl squarate, **4.17** was isolated in 86% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 4.41 (s, 3H), 3.98 (br, rotamer 1, 0.33H), 3.50 (br, rotamer 2, 0.66H), 1.96 (br, 2H), 1.76 (br, 2H), 1.64 (m, 2H), 1.33 (m, 4H), 1.21 (m, 1H).

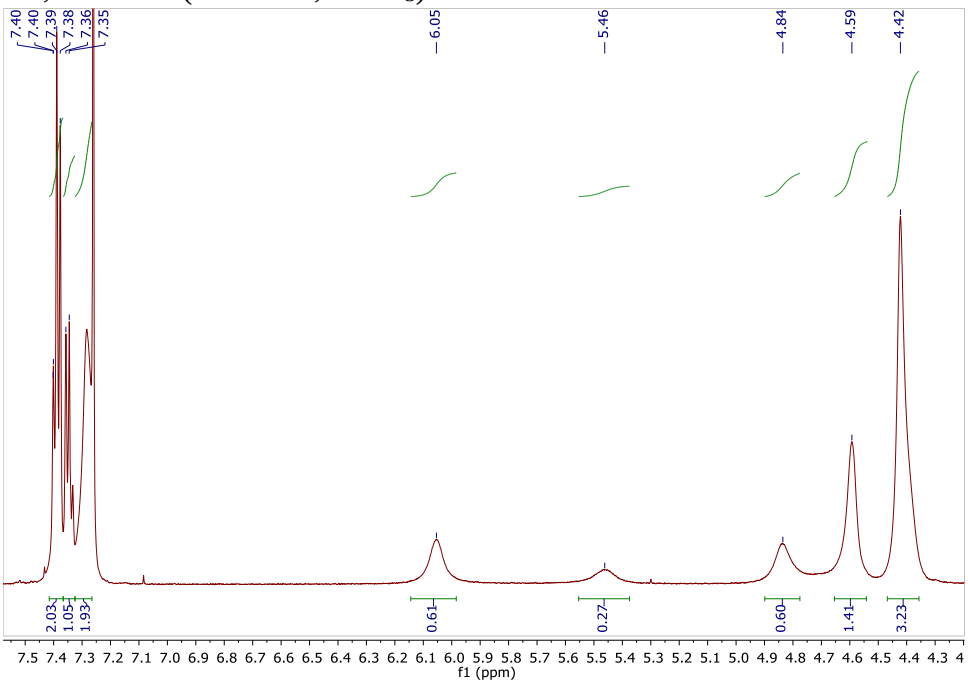
3-methoxy-4-(quinolin-3-ylamino)cyclobut-3-ene-1,2-dione **(4.18)**

Using 144mg 3-aminoquinoline with 156mg dimethyl squarate, **4.18** was isolated in 85% yield. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.88 (d, J=2.58 Hz, 1H), 8.20 (br, 1H), 8.11 (d, J=8.31 Hz, 1H), 7.86 (d, J=8.31 Hz, 1H), 7.72 (t, J=6.56 Hz, 1H), 7.62 (t, J=7.43 Hz, 1H), 4.58 (s, 3H).

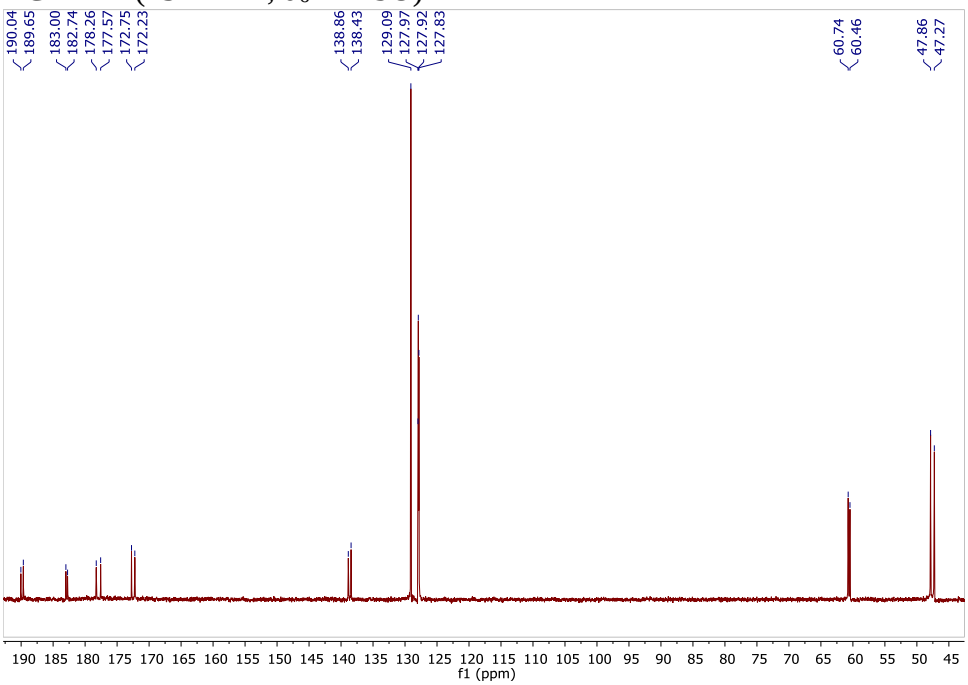
4.6 NMR Spectra

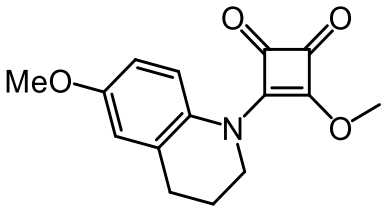


4.1, ^{1}H NMR (600 MHz, $CDCl_3$)

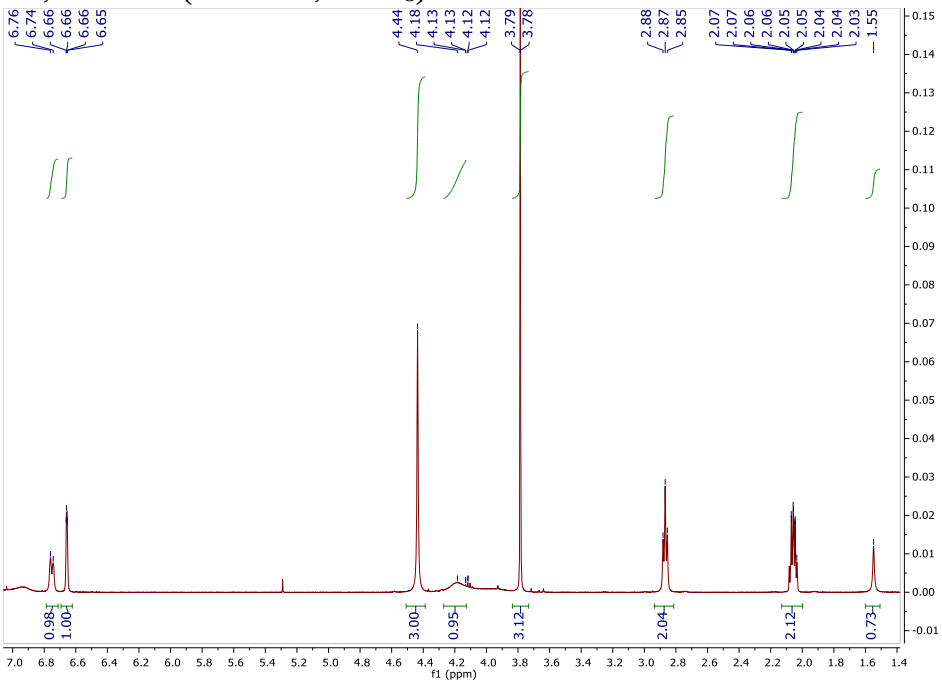


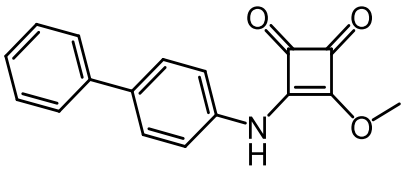
^{13}C NMR (151 MHz, d_6-DMSO)



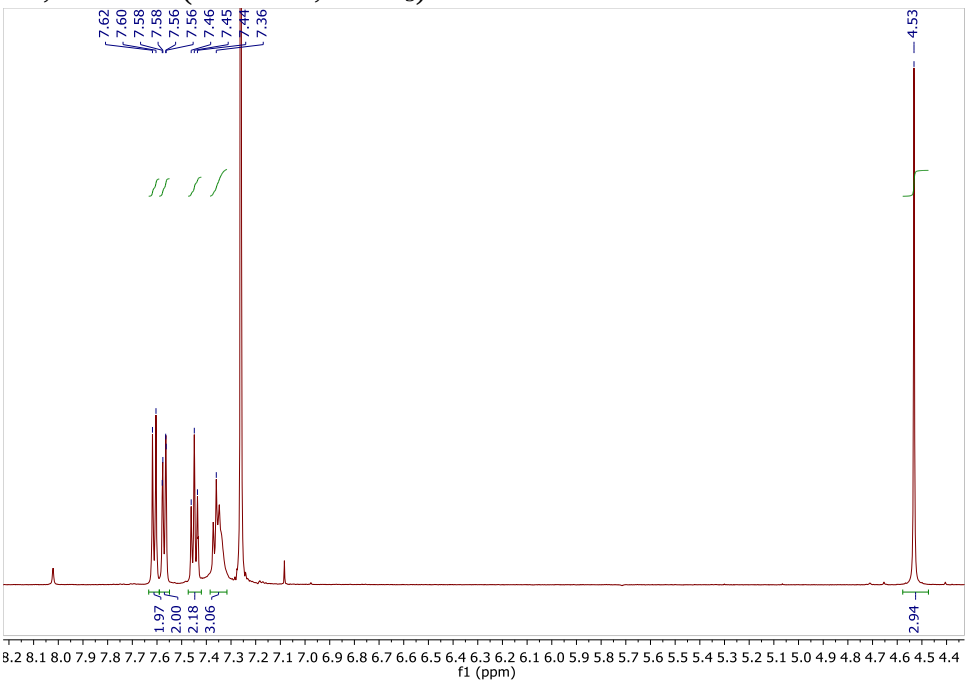


4.2, ^{1}H NMR (500 MHz, $CDCl_3$)

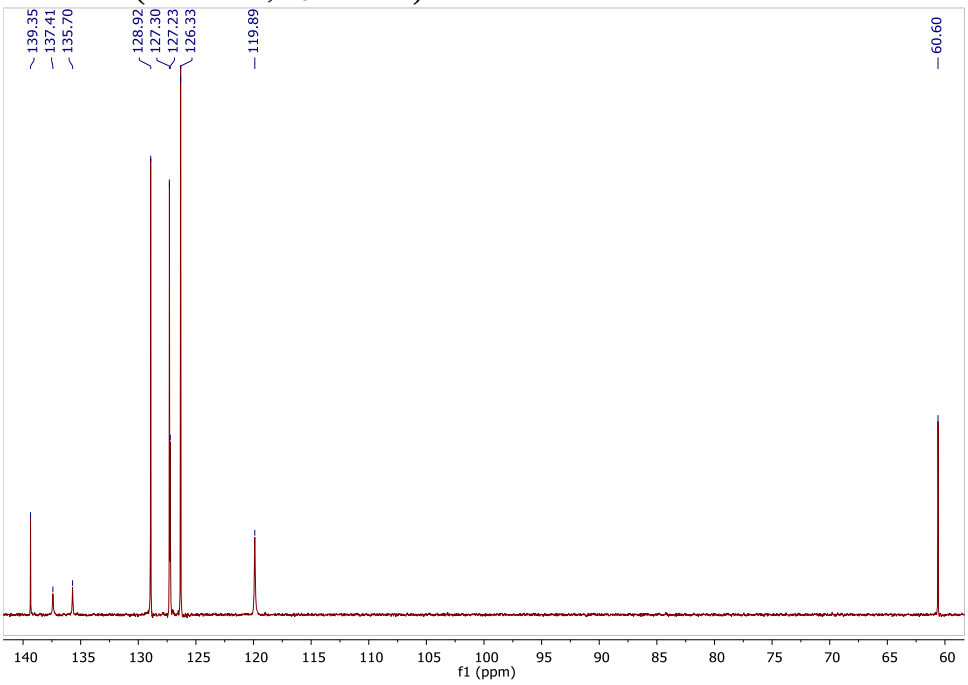


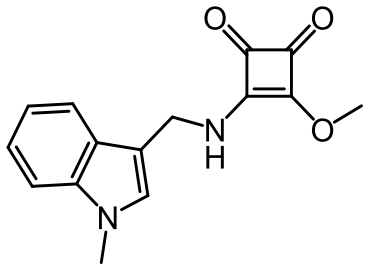


4.3, ^{1}H NMR (600 MHz, $CDCl_3$)

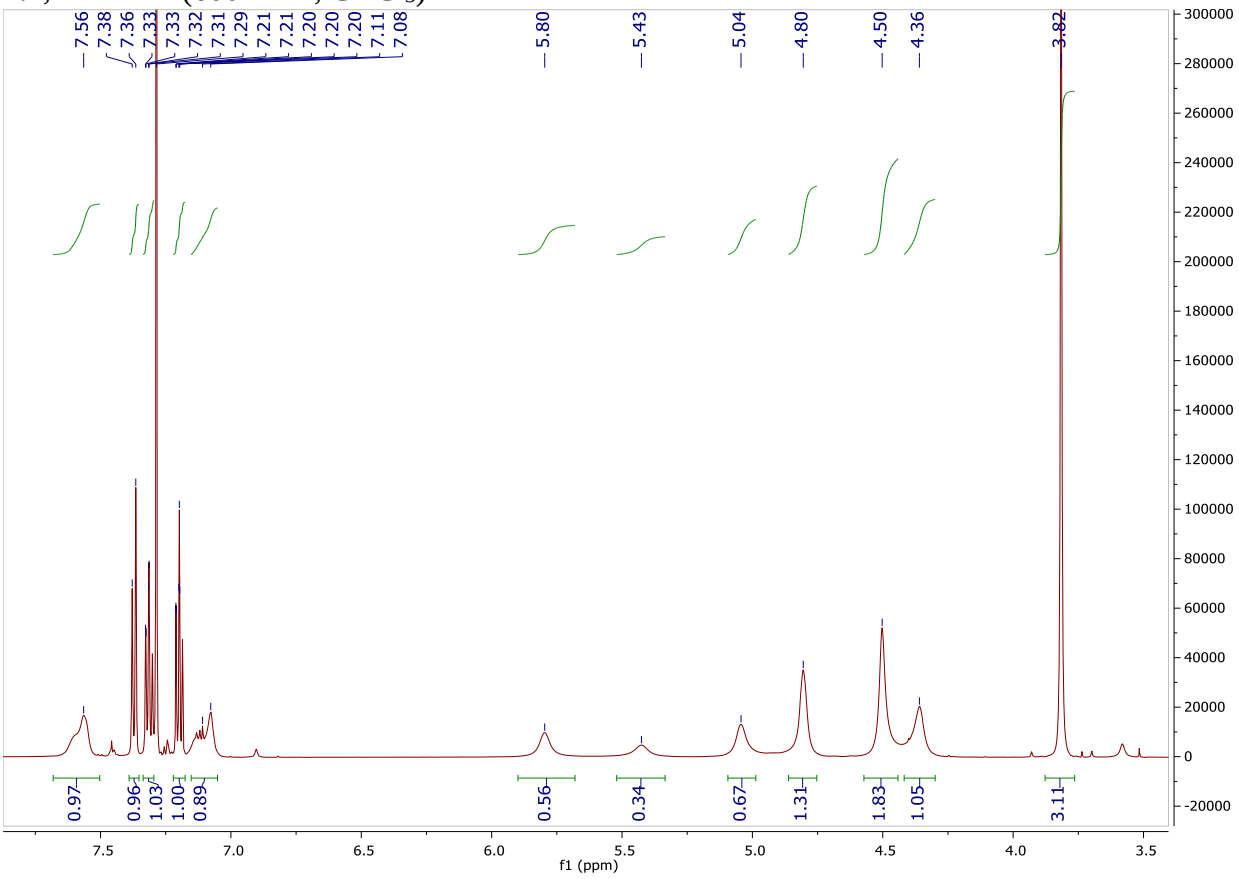


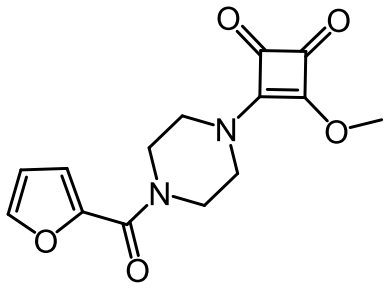
^{13}C NMR (151 MHz, d_6-DMSO)



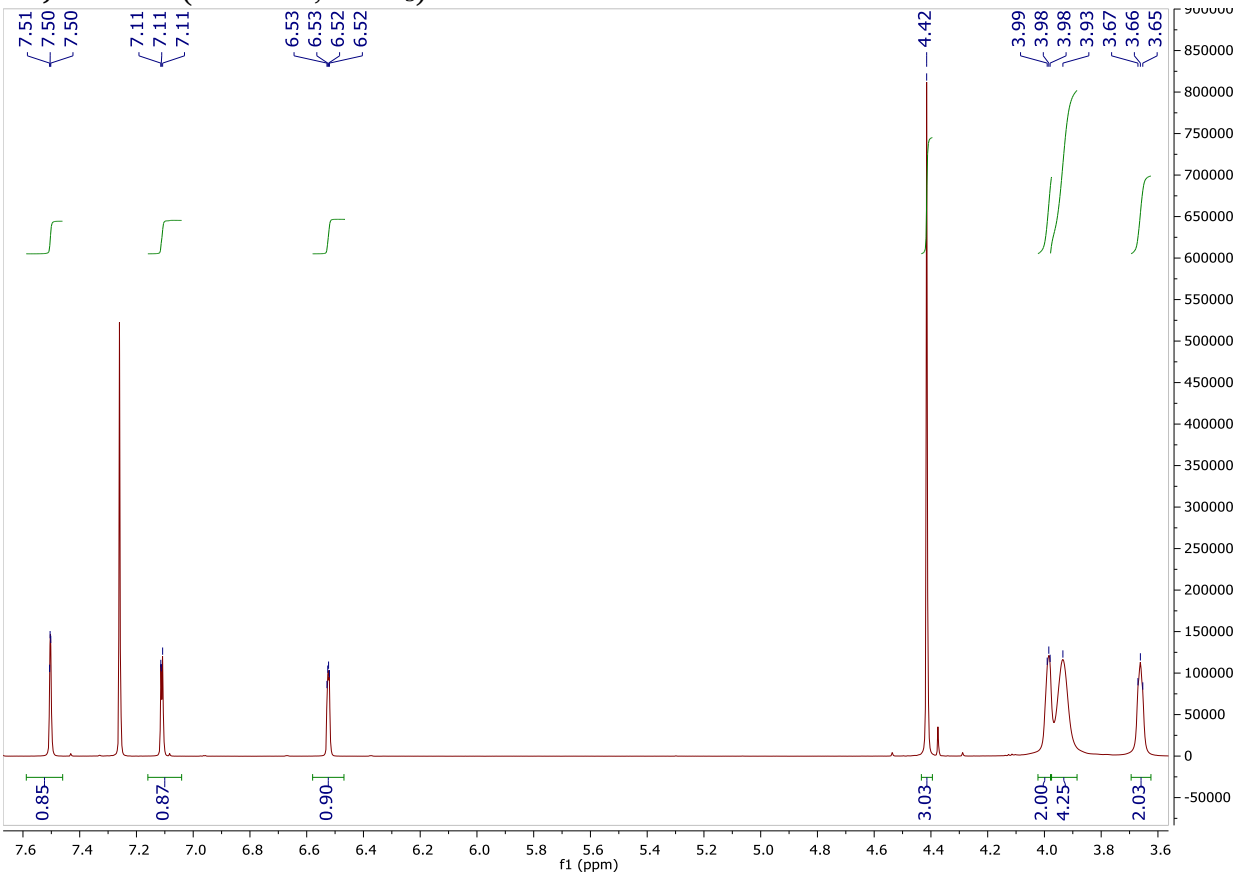


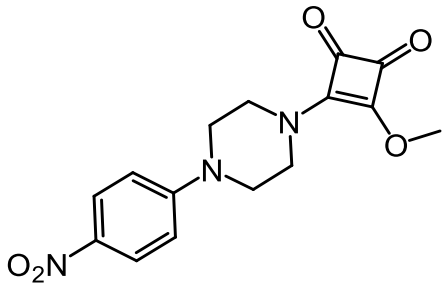
4.4, ^{1}H NMR (600 MHz, $CDCl_3$)



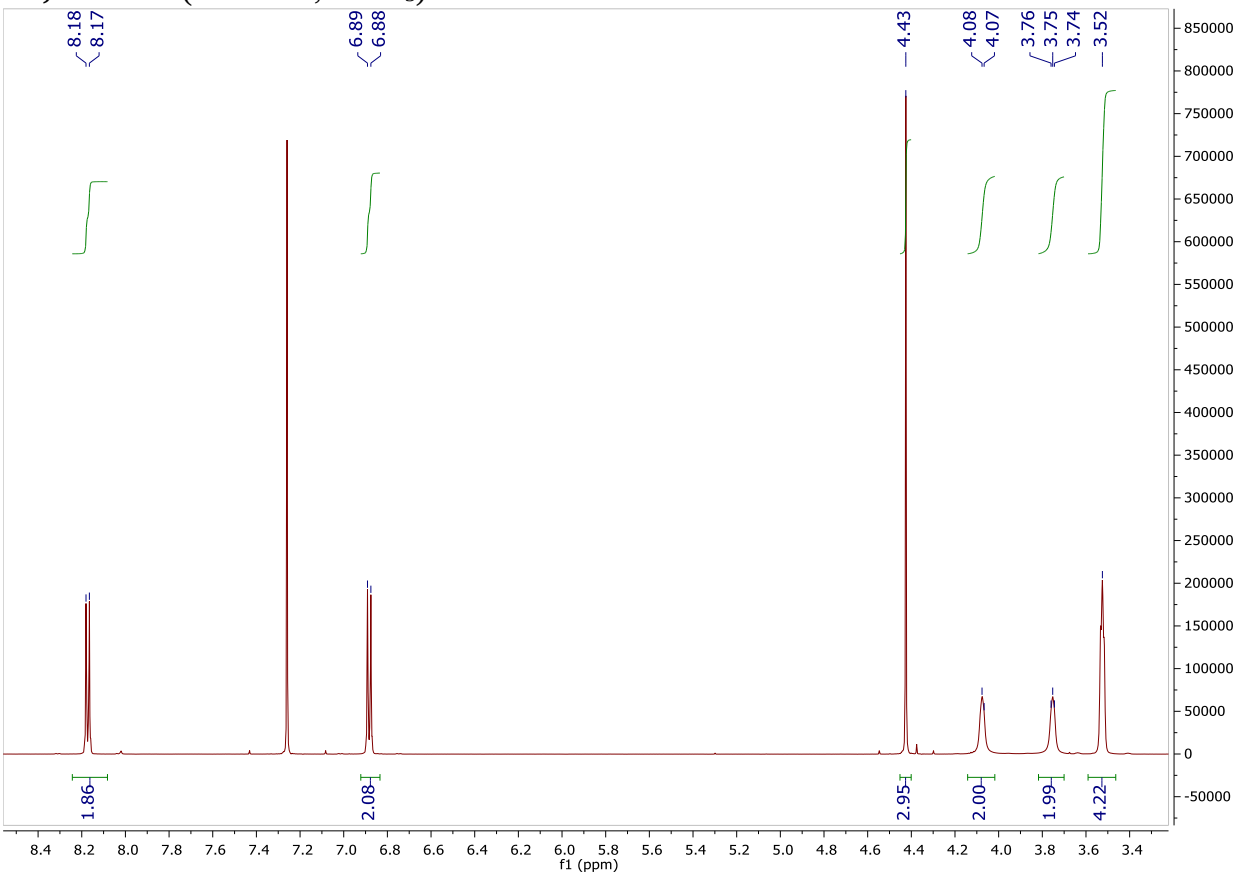


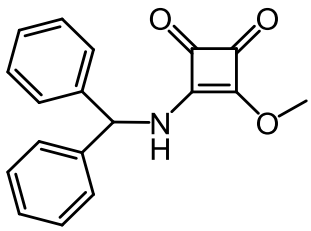
4.5, ^{1}H NMR (600 MHz, $CDCl_3$)



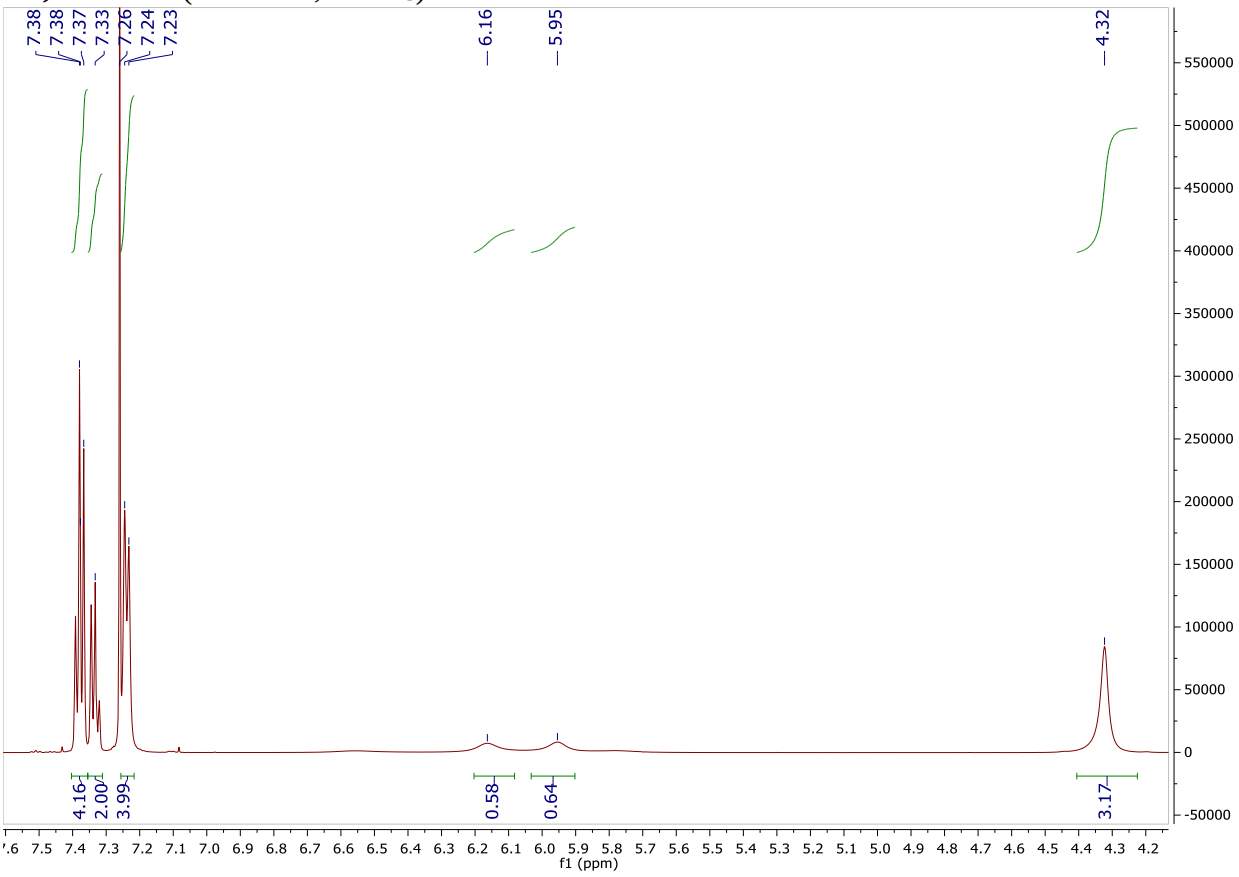


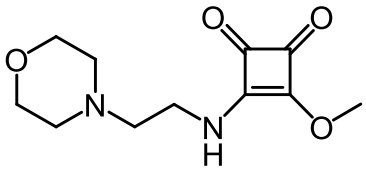
4.6, ^{1}H NMR (600 MHz, $CDCl_3$)



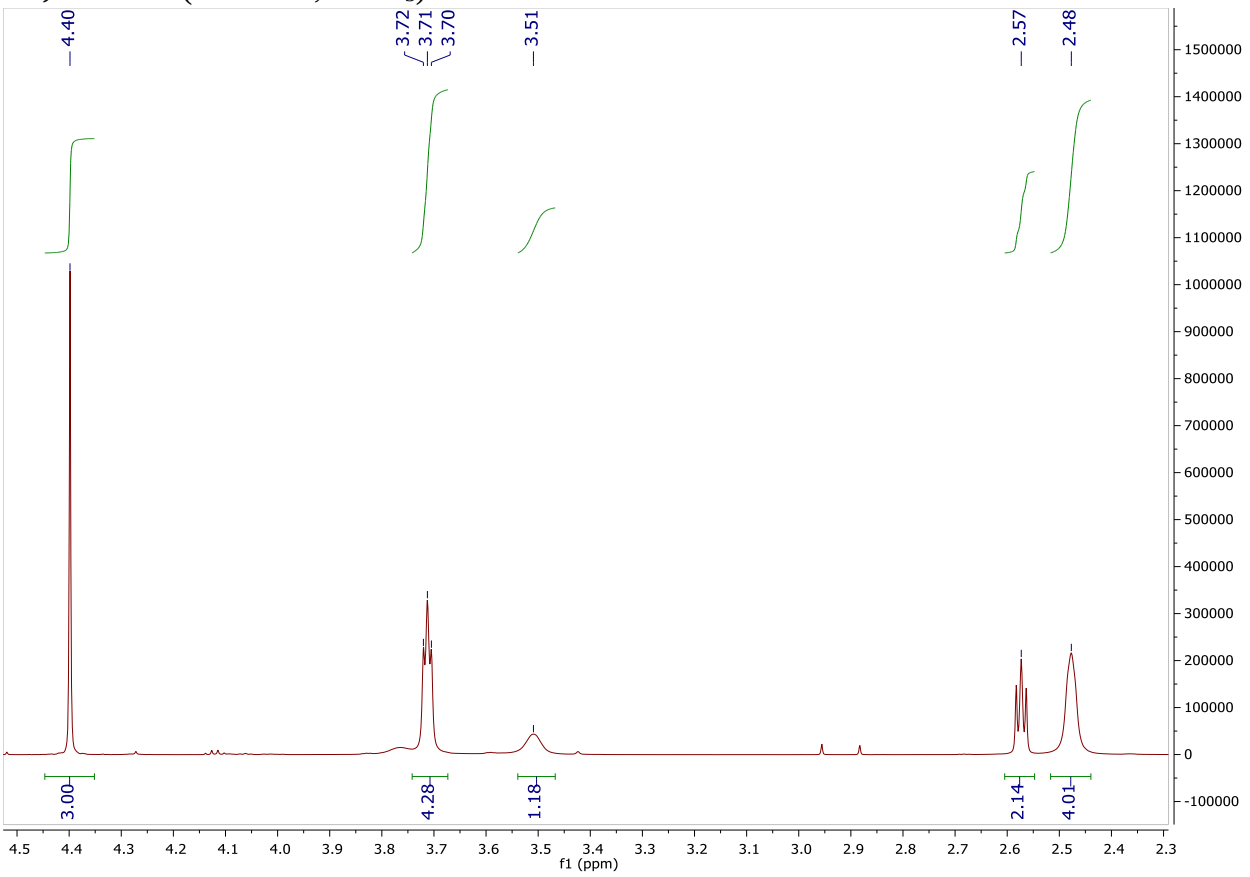


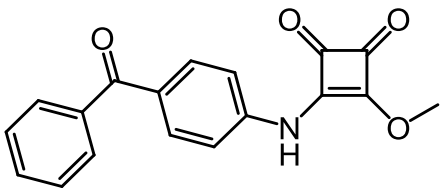
4.7, ^{1}H NMR (600 MHz, $CDCl_3$)



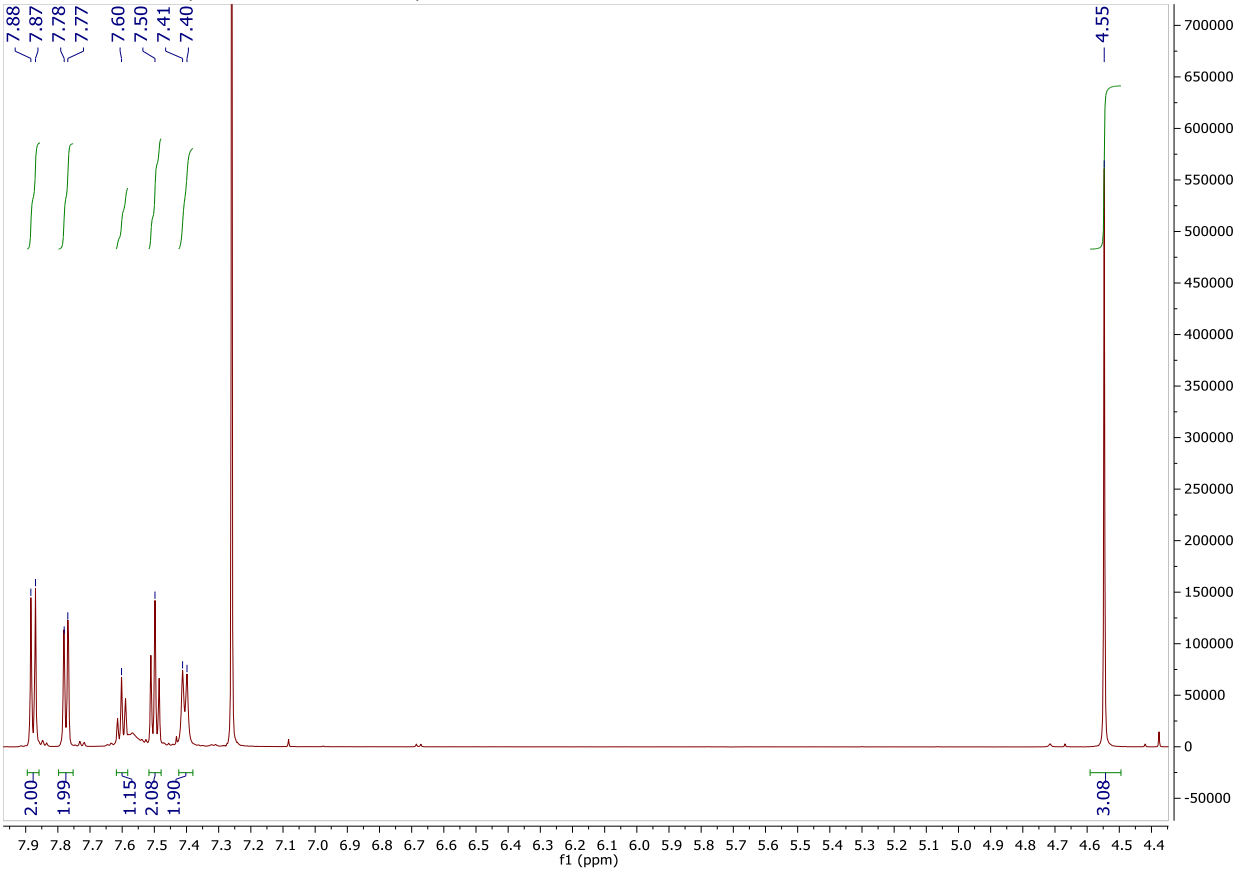


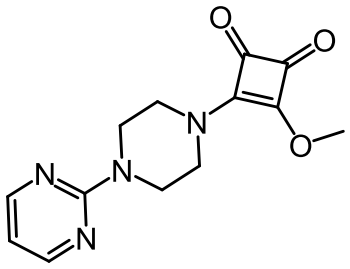
4.8, ^{1}H NMR (600 MHz, $CDCl_3$)



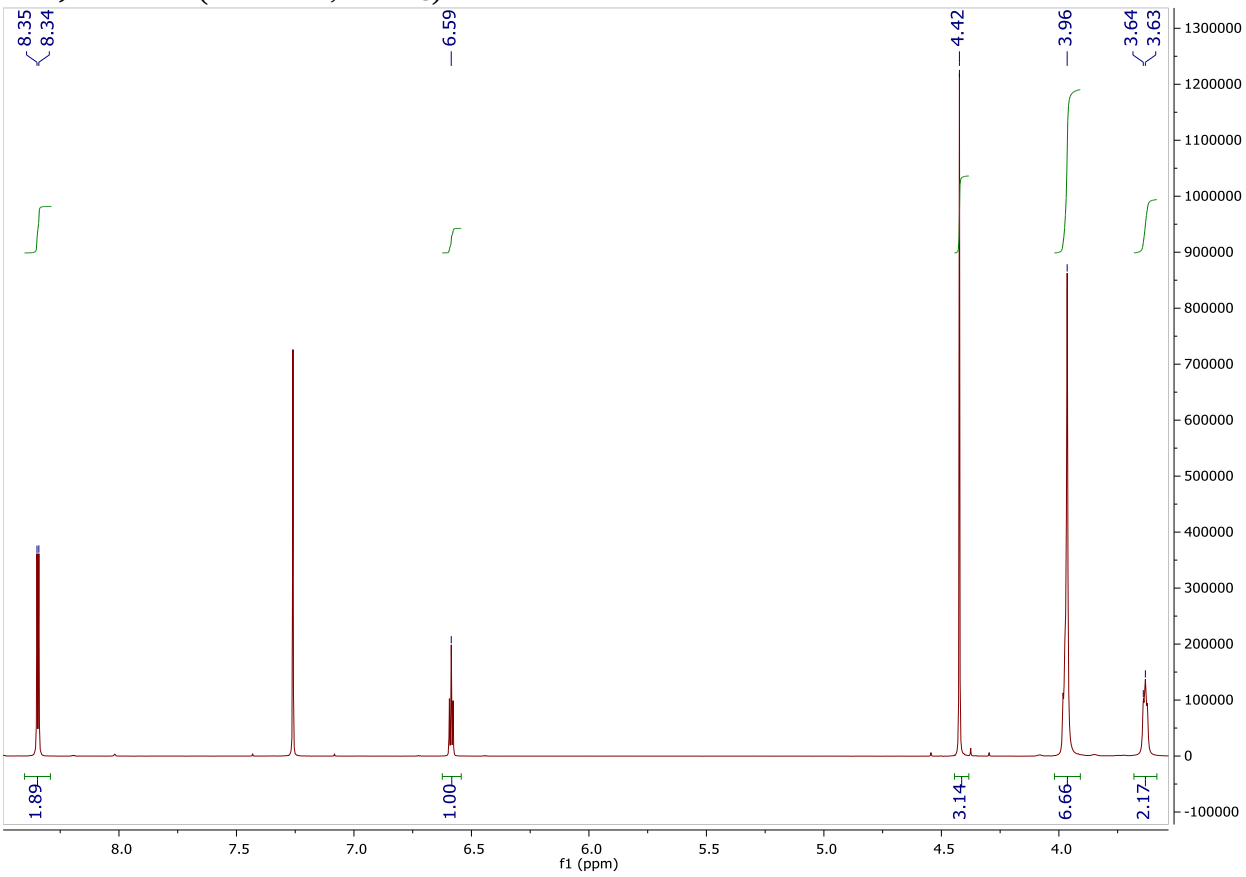


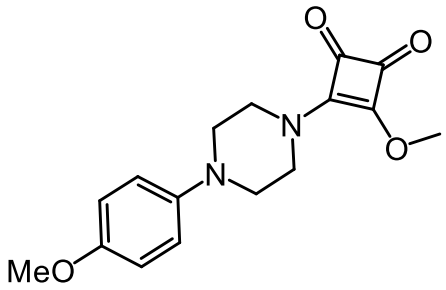
4.9, ^{1}H NMR (600 MHz, $CDCl_3$)



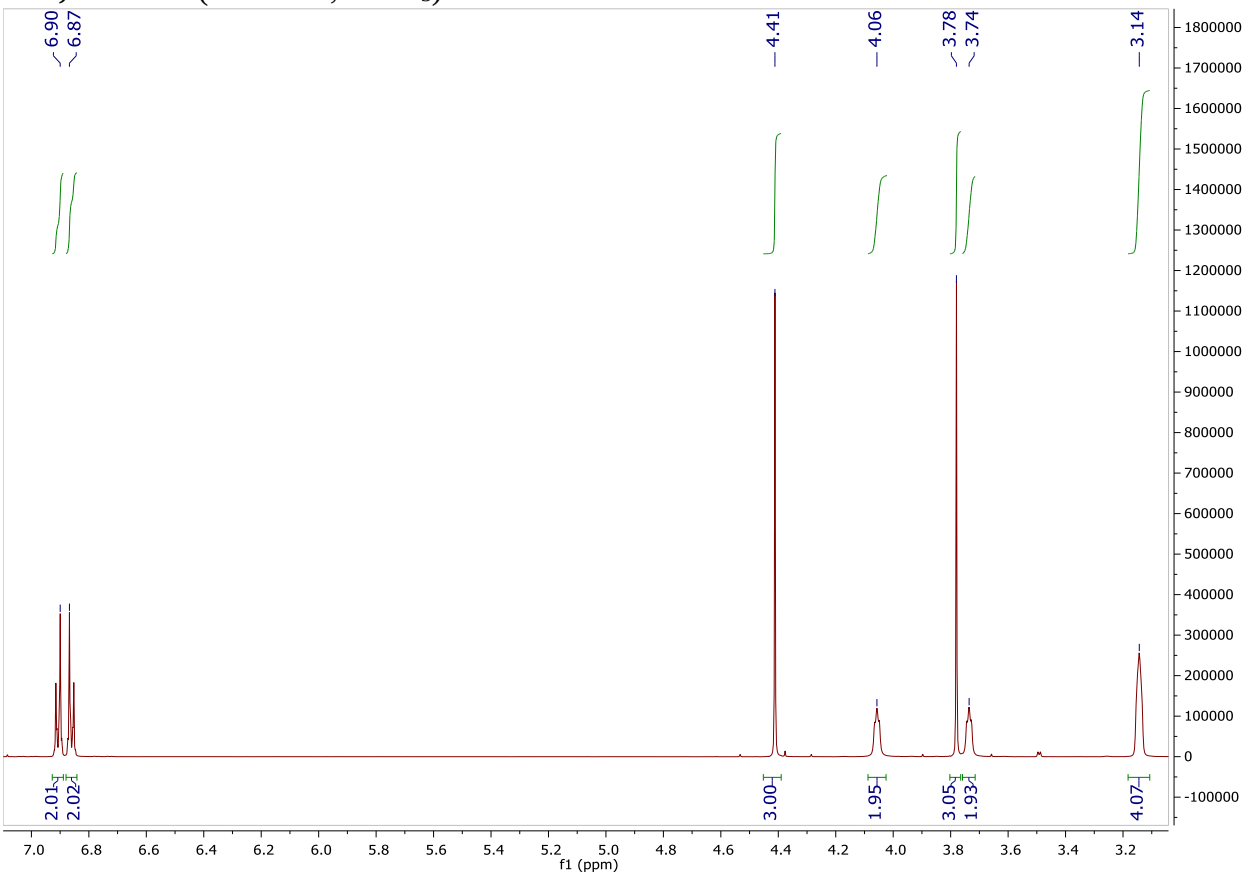


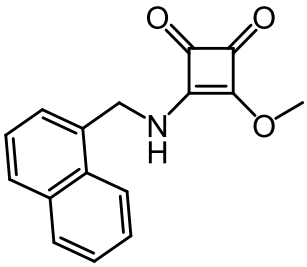
4.10, ^{1}H NMR (600 MHz, $CDCl_3$)



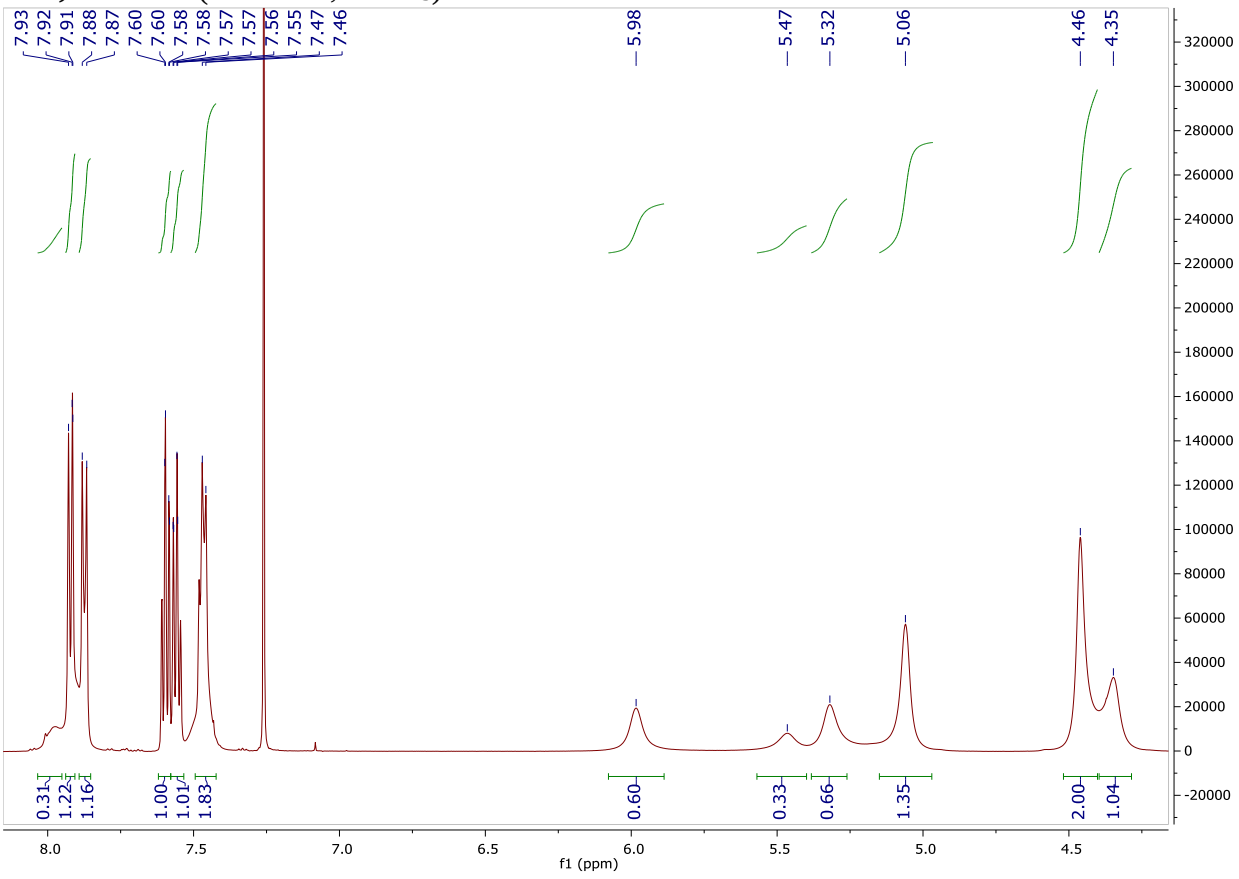


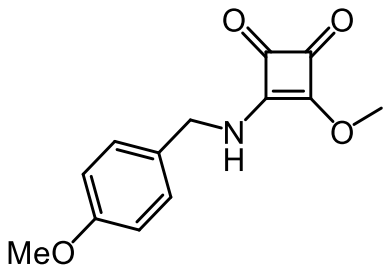
4.11, ^{1}H NMR (600 MHz, $CDCl_3$)



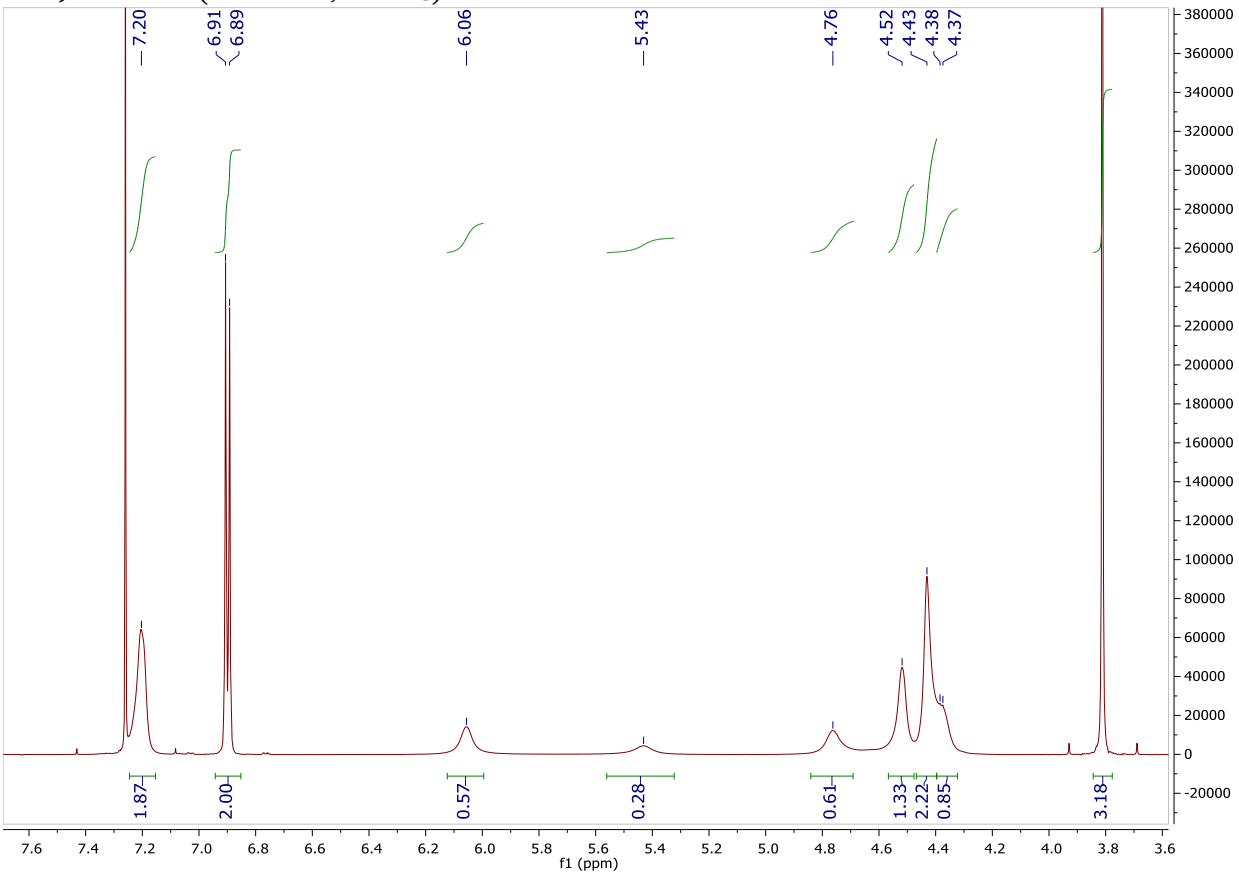


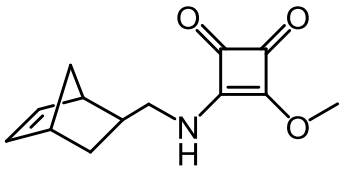
4.12, ^{1}H NMR (600 MHz, $CDCl_3$)



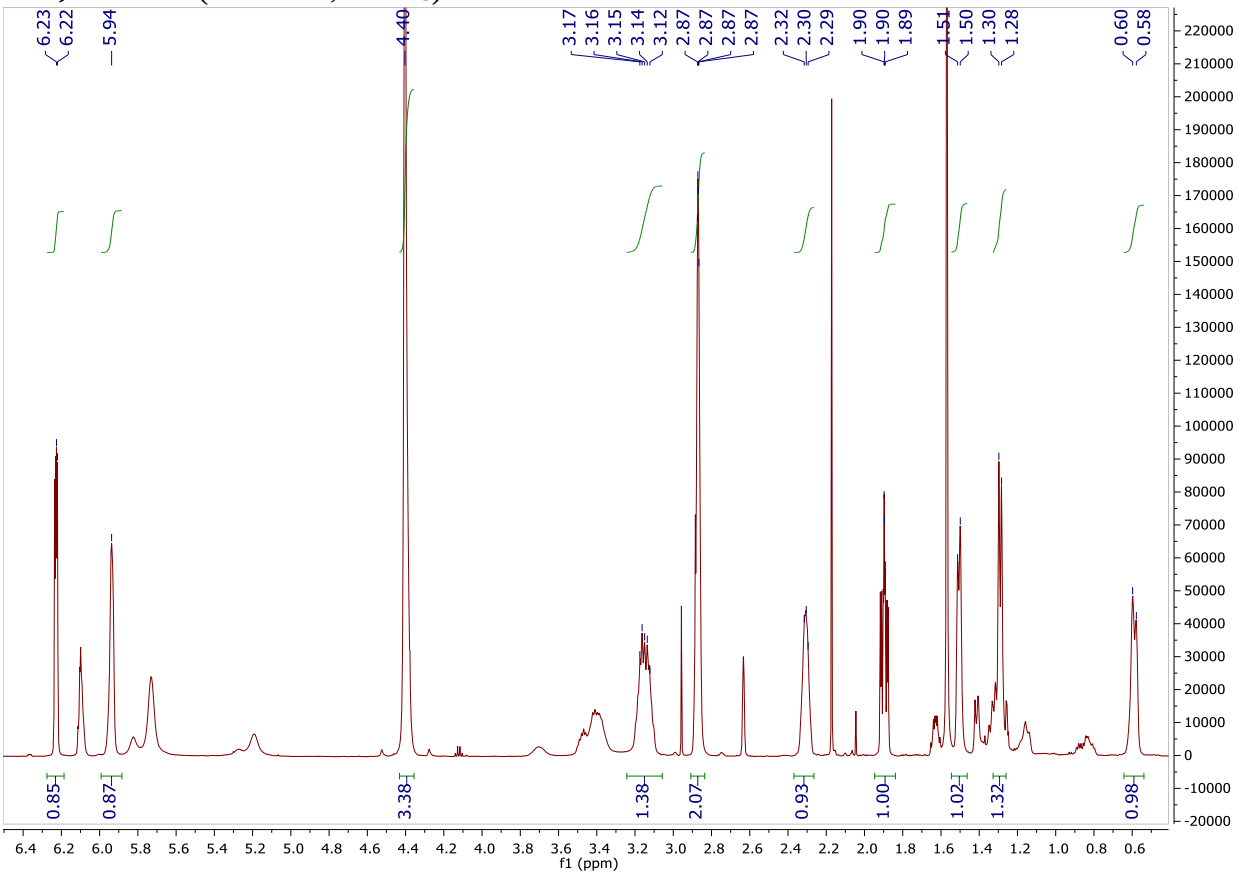


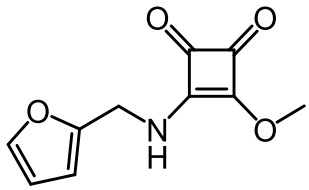
4.13, ^{1}H NMR (600 MHz, $CDCl_3$)



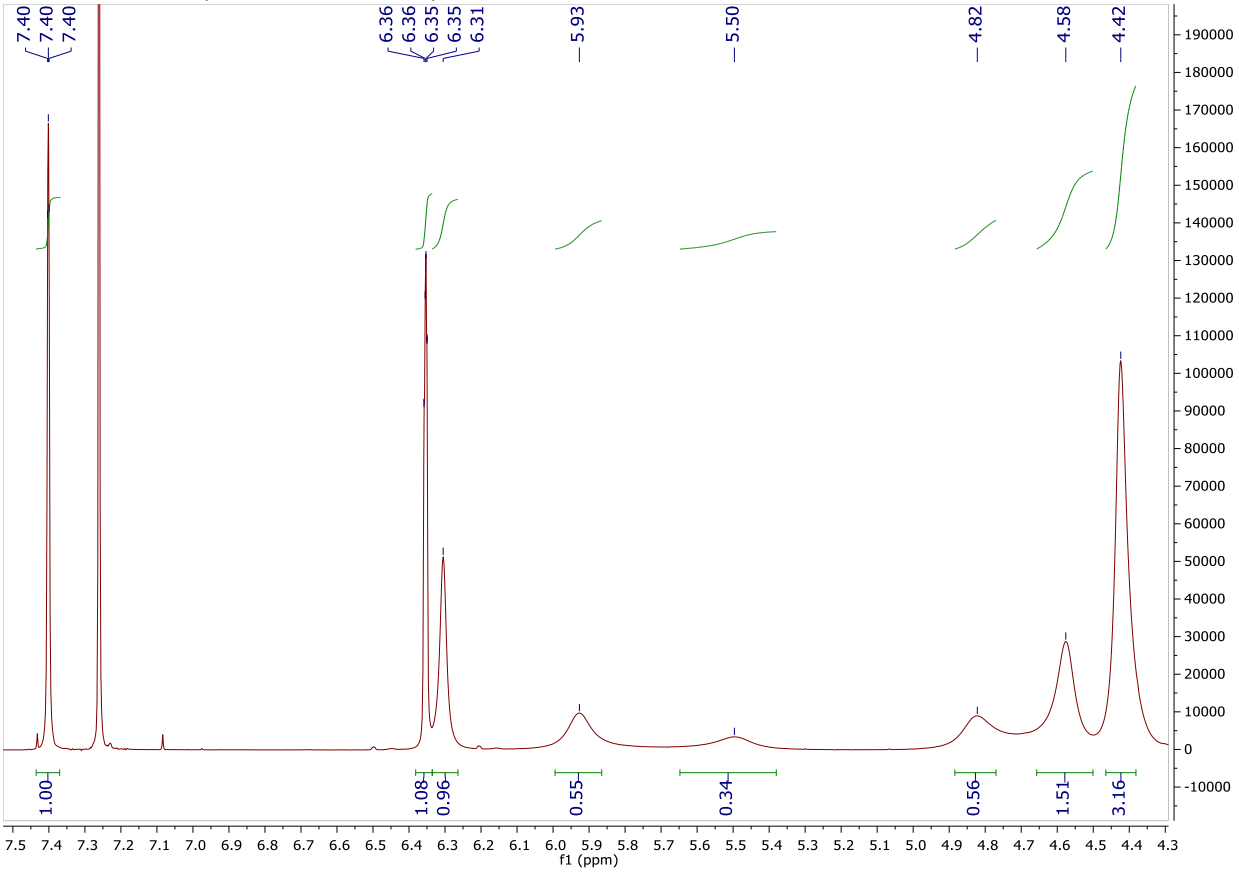


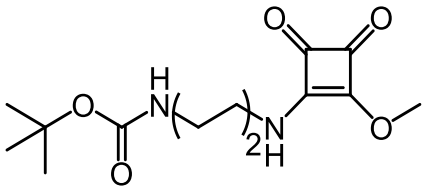
4.14, ^{1}H NMR (600 MHz, $CDCl_3$)



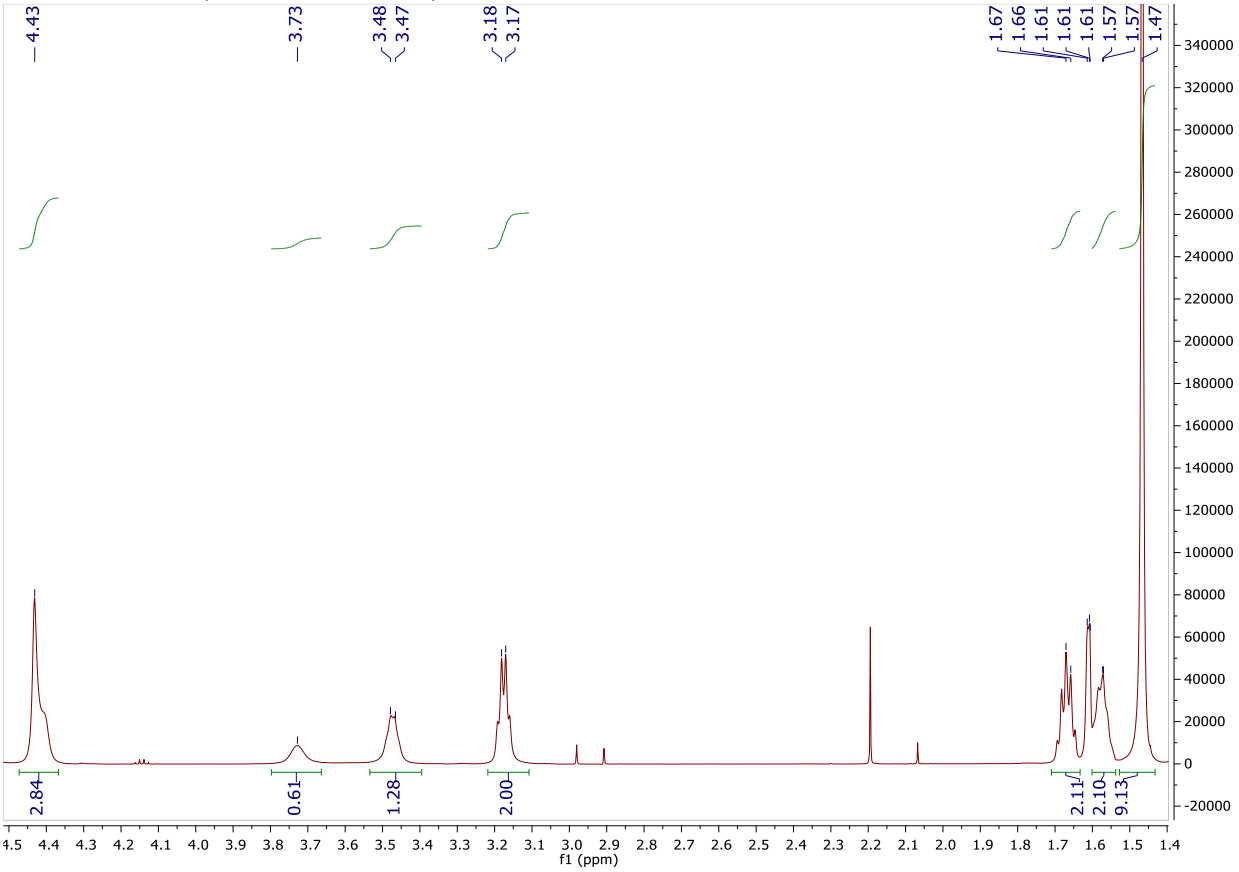


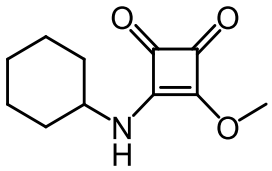
4.15, ^{1}H NMR (600 MHz, $CDCl_3$)



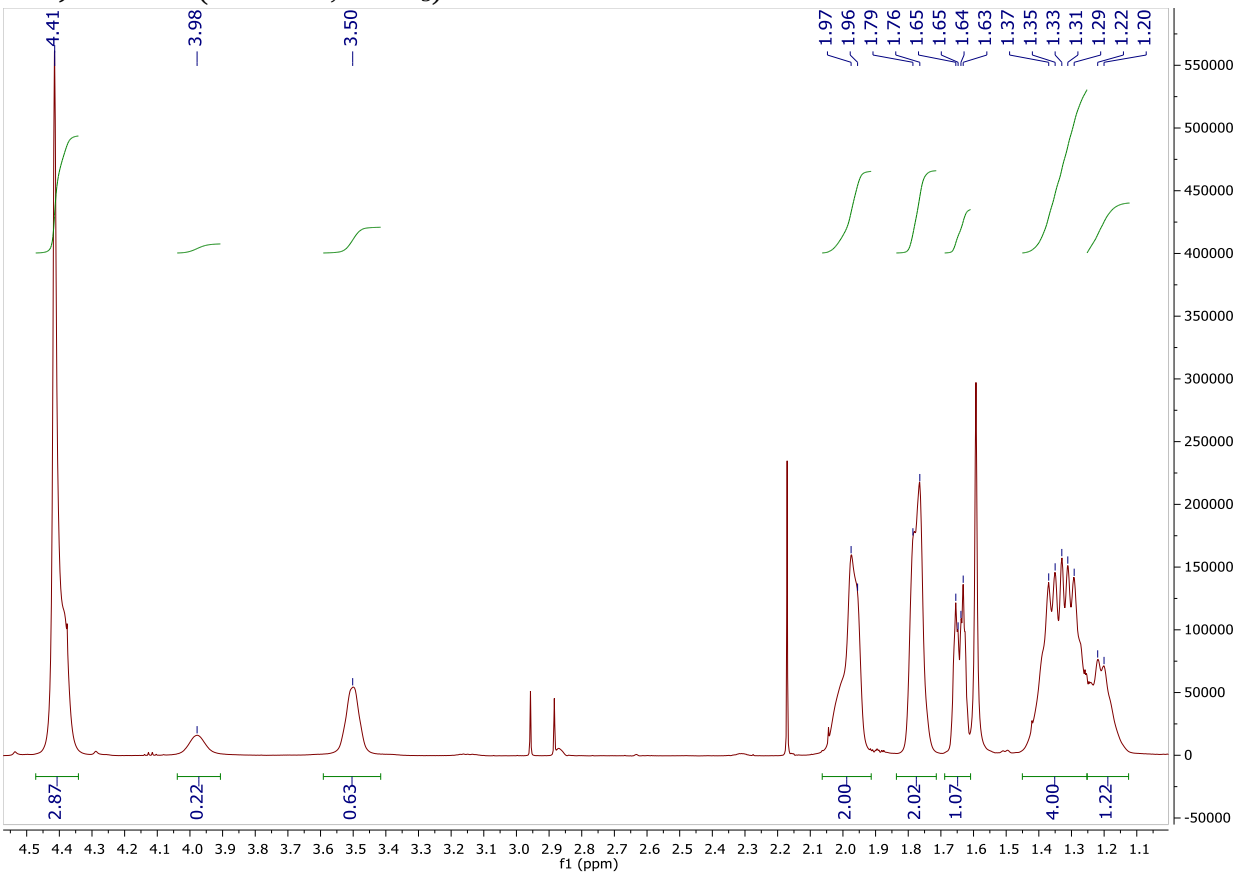


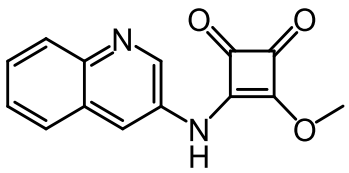
4.16, ^{1}H NMR (600 MHz, $CDCl_3$)



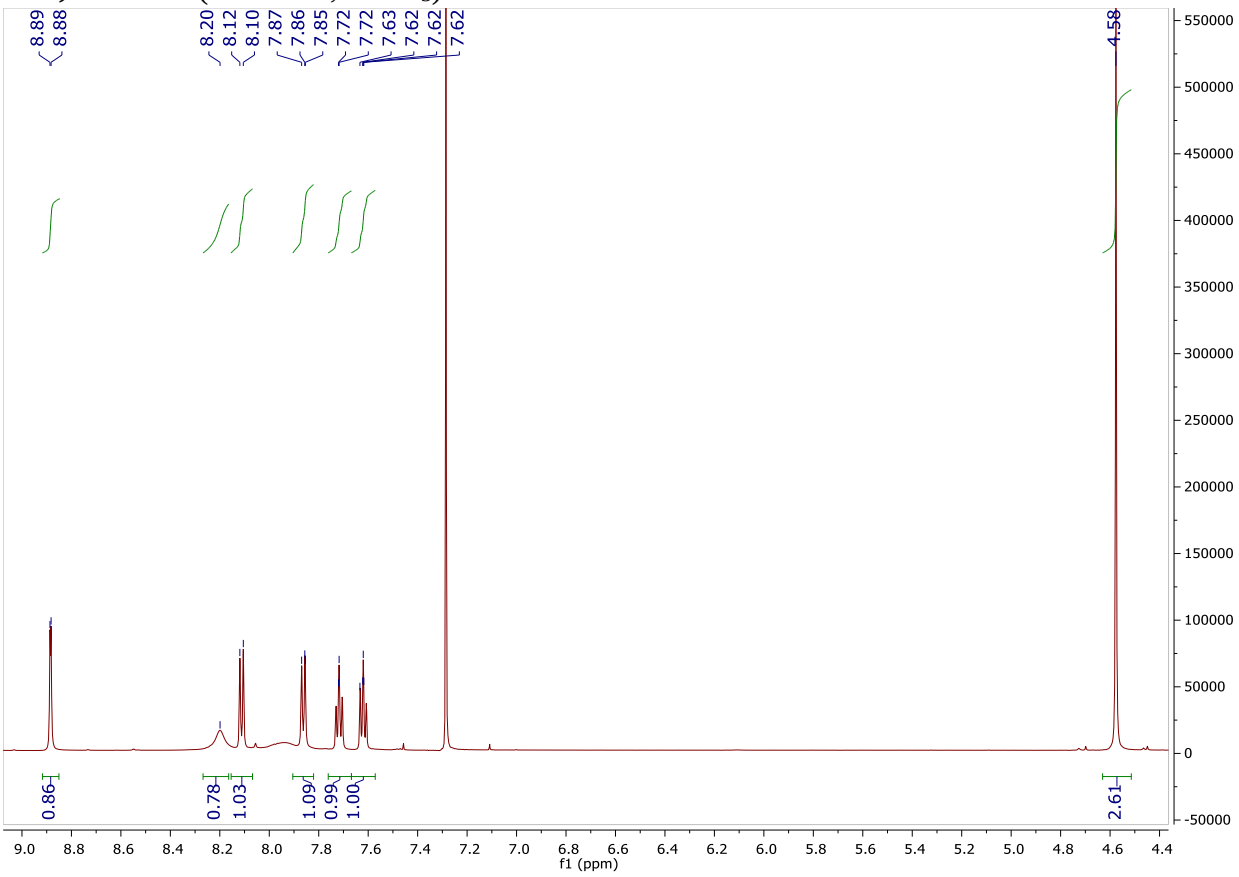


4.17, ^{1}H NMR (600 MHz, $CDCl_3$)





4.18, ^{1}H NMR (600 MHz, $CDCl_3$)



Compiled References

(1) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. (2011) The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov. 10*, 307–317.

(2) Khan, A. R.; James, M. N. G. (2008) Molecular Mechanisms for the Conversion of Zymogens to Active Proteolytic Enzymes. *Protein Sci. 7*, 815–836.

(3) Kobe, B.; Kemp, B. E. (1999) Active Site-Directed Protein Regulation . *Nature 402*, 373–376.

(4) Adam, G. C.; Burbaum, J.; Kozarich, J. W.; Patricelli, M. P.; Cravatt, B. F. (2004) Mapping Enzyme Active Sites in Complex Proteomes. *J. Am. Chem. Soc. 126*, 1363–1368.

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

(6) González-Bello, C. (2016) Designing Irreversible Inhibitors - Worth the Effort? *ChemMedChem 11*, 22–30.

(7) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. (1999) Activity-Based Protein Profiling: The Serine Hydrolases. *Proc. Natl. Acad. Sci. U. S. A. 96*, 14694–14699.

(8) Pace, N. J.; Weerapana, E. (2013) Diverse Functional Roles of Reactive Cysteines. *ACS Chem. Biol. 8*, 283–296.

(9) Blum, G.; Von Degenfeld, G.; Merchant, M. J.; Blau, H. M.; Bogyo, M. (2007) Noninvasive Optical Imaging of Cysteine Protease Activity Using Fluorescently Quenched Activity-Based Probes. *Nat. Chem. Biol. 3*, 668–677.

(10) Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E. (2014) Investigating the Proteome Reactivity and Selectivity of Aryl Halides. *J. Am.*

Chem. Soc. 136, 3330–3333.

(11) Tsai, C. S.; Yen, H. Y.; Lin, M. I.; Tsai, T. I.; Wang, S. Y.; Huang, W. I.; Hsu, T. L.; Cheng, Y. S. E.; Fang, J. M.; Wong, C. H. (2013) Cell-Permeable Probe for Identification and Imaging of Sialidases. *Proc. Natl. Acad. Sci. U. S. A. 110*, 2466–2471.

(12) Kozlowski, L. P. (2017) Proteome-PI: Proteome Isoelectric Point Database. *Nucleic Acids Res. 45*, D1112–D1116.

(13) Moura, A.; Savageau, M. A.; Alves, R. (2013) Relative Amino Acid Composition Signatures of Organisms and Environments. *PLoS One 8*.

(14) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. (2012) Iminoboronates: A New Strategy for Reversible Protein Modification. *J. Am. Chem. Soc. 134*, 10299–10305.

(15) Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W. (2007) Functional Interrogation of the Kinome Using Nucleotide Acyl Phosphates. *Biochemistry 46*, 350–358.

(16) Karelin, A. A.; Tsvetkov, Y. E.; Paulovičová, L.; Bystrický, S.; Paulovičová, E.; Nifantiev, N. E. (2010) Synthesis of 3,6-Branched Oligomannoside Fragments of the Mannan from Candida Albicans Cell Wall Corresponding to the Antigenic Factor 4. *Carbohydr. Res. 345*, 1283–1290.

(17) Saksena, R.; Zhang, J.; Kova, P. (2005) Immunogens from a Synthetic Hexasaccharide Fragment of the O-SP of Vibrio Cholerae O : 1 , Serotype Ogawa. *16*, 187–197.

(18) Bergh, A.; Magnusson, B. G.; Ohlsson, J.; Wellmar, U.; Nilsson, U. J. (2002) Didecyl Squarate - A Practical Amino-Reactive Cross-Linking Reagent for Neoglycoconjugate Synthesis. *Glycoconj. J. 18*, 615–621.

(19) Alemán, J.; Parra, A.; Jiang, H.; Jørgensen, K. A. (2011) Squaramides: Bridging from Molecular Recognition to Bifunctional Organocatalysis. *Chem. - A Eur. J. 17*, 6890–6899.

(20) Vane, J. R.; Botting, R. M. (2003) The Mechanism of Action of Aspirin. *Thromb. Res. 110*, 255–258.

(21) Gordon, E.; Mouz, N.; Duée, E.; Dideberg, O. (2000) The Crystal Structure of the Penicillin-Binding Protein 2x from Streptococcus Pneumoniae and Its Acyl-Enzyme Form: Implication in Drug Resistance. *J. Mol. Biol. 299*, 477–485.

(22) Bauer, R. A. (2015) Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discov. Today 20*, 1061–1073.

(23) Spence, R. A.; Kati, W. M.; Anderson, K. S.; Johnson, K. A. (1995) Mechanism of Inhibition of HIV-1 Reverse Transcriptase by Nonnucleoside Inhibitors. *Science (80-.). 267*, 988–993.

(24) Lipinski, C.; Hopkins, A. (2004) For Biology and Medicine. *432*.

(25) Hinson, J. A.; Roberts, D. W. (1992) Role of Covalent and Noncovalent Interactions in Cell Toxicity: Effects on Proteins. *Annu. Rev. Pharmacol. Toxicol. 32*, 471–510.

(26) Mak, A.; Uetrecht, J. (2018) Idiosyncratic Adverse Drug Reactions. *Compr. Toxicol. Third Ed. 11–15*, 681–716.

(27) Wissner, A.; Overbeek, E.; Reich, M. F.; Floyd, M. B.; Johnson, B. D.; Mamuya, N.; Rosfjord, E. C.; Discafani, C.; Davis, R.; Shi, X.; et al. (2003) Synthesis and Structure-Activity Relationships of 6,7-Disubstituted 4-Anilinoquinoline-3-Carbonitriles. The Design of an Orally Active, Irreversible Inhibitor of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor (EGFR) and the Human Epi. *J. Med. Chem. 46*, 49–63.

(28) Sachs, G.; Shin, J. M.; Howden, C. W. (2006) Review Article: The Clinical Pharmacology of Proton Pump Inhibitors. *Aliment. Pharmacol. Ther. 23*, 2–8.

(29) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. (1999) The Maximal Affinity of Ligands. *Proc. Natl. Acad. Sci. U. S. A. 96*, 9997–10002.

(30) Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. (2009) Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem. 52*, 225–233.

(31) Yun, C. H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K. K.; Meyerson, M.; Eck, M. J. (2008) The T790M Mutation in EGFR Kinase Causes Drug Resistance by Increasing the Affinity for ATP. *Proc. Natl. Acad. Sci. U. S. A. 105*, 2070–2075.

(32) Hagel, M.; Niu, D.; St Martin, T.; Sheets, M. P.; Qiao, L.; Bernard, H.; Karp, R. M.; Zhu, Z.; Labenski, M. T.; Chaturvedi, P.; et al. (2011) Selective Irreversible Inhibition of a Protease by Targeting a Noncatalytic Cysteine. *Nat. Chem. Biol. 7*, 22–24.

(33) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. (2008) Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. *Annu. Rev. Biochem. 77*, 383–414.

(34) Greenbaum, D. C.; Baruch, A.; Grainger, M.; Bozdech, Z.; Medzihradszky, K. F.; Engel, J.; DeRisi, J.; Holder, A. A.; Bogyo, M. (2002) A Role for the Protease Falcipain 1 in Host Cell Invasion by the Human Malaria Parasite. *Science (80-.). 298*, 2002–2006.

(35) Adam, G. C.; Sorensen, E. J.; Cravatt, B. F. (2002) Proteomic Profiling of Mechanistically Distinct Enzyme Classes Using a Common Chemotype. *Nat. Biotechnol. 20*, 805–809.

(36) Adam, G. C.; Cravatt, B. F.; Sorensen, E. J. (2001) Profiling the Specific Reactivity of the Proteome with Non-Directed Activity-Based Probes. *Chem. Biol. 8*, 81–95.

(37) Alexander, J. P.; Cravatt, B. F. (2006) The Putative Endocannabinoid Transport Blocker LY2183240 Is a Potent Inhibitor of FAAH and Several Other Brain Serine Hydrolases. *J. Am. Chem. Soc. 128*, 9699–9704.

(38) Bradshaw, J. M.; McFarland, J. M.; Paavilainen, V. O.; Bisconte, A.; Tam, D.; Phan, V. T.; Romanov, S.; Finkle, D.; Shu, J.; Patel, V.; et al. (2015) Prolonged and Tunable Residence Time Using Reversible Covalent Kinase Inhibitors. *Nat. Chem. Biol. 11*, 525–531.

(39) Shannon, D. A.; Weerapana, E. (2015) Covalent Protein Modification: The Current Landscape of Residue-Specific Electrophiles. *Curr. Opin. Chem. Biol. 24*, 18–26.

(40) Sekhar, R. V.; Patel, S. G.; Guthikonda, A. P.; Reid, M.; Balasubramanyam, A.; Taffet, G. E.; Jahoor, F. (2011) Deficient Synthesis of Glutathione Underlies Oxidative Stress in Aging and Can Be Corrected by Dietary Cysteine and Glycine Supplementation. *Am. J. Clin. Nutr. 94*, 847–853.

(41) Py, B.; Fontecave, M.; Barras, F.; Ollagnier de Choudens, S. (2008) From Iron and Cysteine to Iron-Sulfur Clusters: The Biogenesis Protein Machineries. *EcoSal Plus 3*.

(42) Lill, R.; Mühlenhoff, U. (2006) Iron-Sulfur Protein Biogenesis in Eukaryotes: Components and Mechanisms. *Annu. Rev. Cell Dev. Biol. 22*, 457–486.

(43) Sevier, C. S.; Kaiser, C. A. (2002) Formation and Transfer of Disulphide Bonds in Living Cells. *Nat. Rev. Mol. Cell Biol. 3*, 836–847.

(44) Rauwerdink, A.; Kazlauskas, R. J. (2015) How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: The Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS Catal. 5*, 6153–6176.

(45) Sun, Y.; Yin, S.; Feng, Y.; Li, J.; Zhou, J.; Liu, C.; Zhu, G.; Guo, Z. (2014) Molecular

Basis of the General Base Catalysis of an α/β- Hydrolase Catalytic Triad. *J. Biol. Chem. 289*, 15867–15879.

(46) Hedstrom, L. (2002) Serine Protease Mechanism and Specificity.

(47) Staub, I.; Sieber, S. A. (2008) B-Lactams As Selective Chemical Probes for the in Vivo Labeling of Bacterial Enzymes Involved in Cell Wall Biosynthesis, Antibiotic Resistance, and Virulence. *J. Am. Chem. Soc. 130*, 13400–13409.

(48) Böttcher, T.; Sieber, S. A. (2008) B-Lactones As Privileged Structures for the Active-Site Labeling of Versatile Bacterial Enzyme Classes. *Angew. Chemie - Int. Ed. 47*, 4600–4603.

(49) Simon, G. M.; Cravatt, B. F. (2010) Activity-Based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study. *J. Biol. Chem. 285*, 11051–11055.

(50) Mahrus, S.; Craik, C. S. (2005) Selective Chemical Functional Probes of Granzymes A and B Reveal Granzyme B Is a Major Effector of Natural Killer Cell-Mediated Lysis of Target Cells. *Chem. Biol. 12*, 567–577.

(51) Serim, S.; Mayer, S. V.; Verhelst, S. H. L. (2013) Tuning Activity-Based Probe Selectivity for Serine Proteases by on-Resin “click” Construction of Peptide Diphenyl Phosphonates. *Org. Biomol. Chem. 11*, 5714–5721.

(52) Dantzman, C. L.; Kiessling, L. L. (1996) Reactivity of a 2’-Thio Nucleotide Analog. *J. Am. Chem. Soc. 118*, 11715–11719.

(53) Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E. (2014) Investigating the Proteome Reactivity and Selectivity of Aryl Halides. *J. Am. Chem. Soc. 136*, 3330–3333.

(54) Liebeskind, L. S.; Fengl, R. W.; Wirtz, K. R.; Shawe, T. T. (1988) An Improved Method for the Synthesis of Substituted Cyclobutenediones. *J. Org. Chem. 53*, 2482–2488.

(55) Quiñonero, D.; Frontera, A.; Ballester, P.; Deyà, P. M. (2000) A Theoretical Study of Aromaticity in Squaramide and Oxocarbons. *Tetrahedron Lett. 41*, 2001–2005.

(56) Tietze, L. F.; Arlt, M.; Beller, M.; Gl üsenkamp, K. -H; Jähde, E.; Rajewsky, M. F. (1991) Anticancer Agents, 15. Squaric Acid Diethyl Ester: A New Coupling Reagent for the Formation of Drug Biopolymer Conjugates. Synthesis of Squaric Acid Ester Amides and Diamides. *Chem. Ber. 124*, 1215–1221.

(57) Malerich, J. P.; Hagihara, K.; Rawal, V. H. (2008) Chiral Squaramide Derivatives Are Excellent Hydrogen Bond Donor Catalysts. *J. Am. Chem. Soc. 130*, 14416–14417.

(58) Zhu, Y.; Malerich, J. P.; Rawal, V. H. (2010) Squaramide-Catalyzed Enantioselective Michael Addition of Diphenyl Phosphite to Nitroalkenes. *Angew. Chemie - Int. Ed. 49*, 153–156.

(59) Lindhorst, T. K.; Bruegge, K.; Fuchs, A.; Sperling, O. (2010) A Bivalent Glycopeptide to Target Two Putative Carbohydrate Binding Sites on FimH. 801–809.

(60) Cai, H.; Huang, Z.; Shi, L.; Zou, P.; Zhao, Y.; Kunz, H. (2011) Synthesis of Tn / T Antigen MUC1 Glycopeptide BSA Conjugates and Their. 3685–3689.

(61) Wang, J.; Asnani, A.; Auzanneau, F. (2010) Bioorganic & Medicinal Chemistry Synthesis of a BSA-Le x Glycoconjugate and Recognition of Le x Analogues by the Anti-Le x Monoclonal Antibody SH1 : The Identification of a Non-Cross Reactive Analogue. *Bioorg. Med. Chem. 18*, 7174–7185.

(62) Tietze, L. F.; Schroter, C.; Gabius, S.; Gabius, H.; Goerlach-grawj, A. (1991) Conjugation of P-Aminophenyl Glycosides with Squaric Acid Diester to a Carrier Protein and the Use of Neoglycoprotein in the Histochemical Detection of Lectins1. 148–153.

(63) Tevyashova, A.; Sztaricskai, F.; Batta, G. (2004) Formation of Squaric Acid Amides of

Anthracycline Antibiotics . Synthesis and Cytotoxic Properties. *14*, 4783–4789.

(64) Kitov, P. I.; Bundle, D. R. (2001) Synthesis and Structure–Activity Relationships of Di- and Trisaccharide Inhibitors For. *2*.

(65) Owen, R. M.; Carlson, C. B.; Xu, J.; Mowery, P.; Fasella, E.; Kiessling, L. L. (2007) Bifunctional Ligands That Target Cells Displaying the α Vβ3 Integrin. *ChemBioChem 8*, 68–82.

(66) Hou, S. jie; Saksena, R.; Kováč, P. (2008) Preparation of Glycoconjugates by Dialkyl Squarate Chemistry Revisited. *Carbohydr. Res. 343*, 196–210.

(67) Sun, W.; Zhu, G.; Wu, C.; Hong, L.; Wang, R. (2012) An Organocatalytic Cascade Strategy for the Enantioselective Construction of Spirocyclopentane Bioxindoles Containing Three Contiguous Stereocenters and Two Spiro Quaternary Centers. *Chem. - A Eur. J. 18*, 6737–6741.

(68) Li, P.; Hu, X.; Dong, X. Q.; Zhang, X. (2016) Recent Advances in Dynamic Kinetic Resolution by Chiral Bifunctional (Thio)Urea-and Squaramide-Based Organocatalysts. *Molecules 21*, 1–14.

(69) Spjut, S.; Pudelko, M.; Hartmann, M.; Elofsson, M. (2009) Carbamate Linker Strategy in Solid-Phase Synthesis of Amino-Functionalized Glycoconjugates for Attachment to Solid Surfaces and Investigation of Protein-Carbohydrate Interactions. *European J. Org. Chem.* No. 3, 349–357.

(70) Sztaricskai, F.; Roth, E.; Andrei, M.; Pelyvás, I. F.; Herczegh, P. (2007) Application of Squaric Acid Esters in Aminodeoxy Sugar Chemistry. *Chem. Lett. 36*, 1012–1013.

(71) Sejwal, P.; Han, Y.; Shah, A.; Luk, Y. Y. (2007) Water-Driven Chemoselective Reaction of Squarate Derivatives with Amino Acids and Peptides. *Org. Lett. 9*, 4897–4900.

(72) Shin, M.; Mercier, F.; Munson, K.; Miller, M.; Hersey, S.; Sachs, G. (1993) Membrane Topology and Omeprazole Labeling of the Gastric. 2345–2355.

(73) Speers, A. E.; Adam, G. C.; Cravatt, B. F. (2003) Activity-Based Protein Profiling in Vivo Using a Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc. 125*, 4686–4687.

(74) Shannon, D. A.; Gu, C.; McLaughlin, C. J.; Kaiser, M.; van der Hoorn, R. A. L.; Weerapana, E. (2012) Sulfonyl Fluoride Analogues as Activity-Based Probes for Serine Proteases. *ChemBioChem 13*, 2327–2330.

(75) Mortenson, D. E.; Brighty, G. J.; Plate, L.; Bare, G.; Chen, W.; Li, S.; Wang, H.; Cravatt, B. F.; Forli, S.; Powers, E. T.; et al. (2018) " Inverse Drug Discovery " Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exempli Fi Ed by Aryl Fluorosulfates. *2*.

(76) Bush, K. (2012) Improving Known Classes of Antibiotics : An Optimistic Approach for the Future. *Curr. Opin. Pharmacol. 12*, 527–534.

(77) Rombola, M.; Rawal, V. H. (2018) Dicyclopentyl Dithiosquarate as an Intermediate for the Synthesis of Thiosquaramides. *Org. Lett. 20*, 514–517.

(78) Eggerding, D.; West, R. (1976) Synthesis of the Monothiosquarate and 1,2-Dithiosquarate Ions and Their Derivatives. *J. Org. Chem. 41*, 3904–3909.

(79) Schmidt, A. H. (1980) Reaktionen von Quadratsäure Und Quadratsäure-Derivaten. *Synthesis (Stuttg).* 961–994.

(80) Schmidt, J.R.; Polik, W. F. (2016) WebMO Enterprise. WebMO LLC: Holland, MI, USA 2016.

(81) Huang, L.; Kerns, R. J. (2006) Diversity-Oriented Chemical Modification of Heparin:

Identification of Charge-Reduced N-Acyl Heparin Derivatives Having Increased Selectivity for Heparin-Binding Proteins. *Bioorg. Med. Chem. 14*, 2300–2313.

(82) Singh, P.; Kaur, S.; Kumari, P.; Kaur, B.; Kaur, M.; Singh, G.; Bhatti, R.; Bhatti, M. (2018) Tailoring the Substitution Pattern on 1,3,5-Triazine for Targeting Cyclooxygenase-2: Discovery and Structure-Activity Relationship of Triazine-4-Aminophenylmorpholin-3-One Hybrids That Reverse Algesia and Inflammation in Swiss Albino Mice. *J. Med. Chem. 61*, 7929–7941.

(83) Noe, M. C.; Gilbert, A. M. (2012) *Targeted Covalent Enzyme Inhibitors*, 1st ed.; Elsevier Inc.; Vol. 47.

(84) Mah, R.; Thomas, J. R.; Shafer, C. M. (2014) Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorganic Med. Chem. Lett. 24*, 33–39.

(85) Alexander, J. P.; Cravatt, B. F. (2005) Mechanism of Carbamate Inactivation of FAAH: Implications for the Design of Covalent Inhibitors and in Vivo Functional Probes for Enzymes. *Chem. Biol. 12*, 1179–1187.

(86) Long, M. J. C.; Aye, Y. (2017) Privileged Electrophile Sensors: A Resource for Covalent Drug Development. *Cell Chem. Biol. 24*, 787–800.

(87) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. (2008) Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. *Annu. Rev. Biochem. 77*, 383–414.

(88) Suginaka, H.; Blumberg, P. M.; Strominger, J. L. (1972) Multiple Penicillin-Binding Components in Bacillus Subtilis, Bacillus Cereus, Staphylococcus Aureus, and Escherichia Coli. *J. Biol. Chem. 247*, 5279–5288.

(89) Garland, M.; Yim, J. J.; Bogyo, M. (2016) A Bright Future for Precision Medicine: Advances in Fluorescent Chemical Probe Design and Their Clinical Application. *Cell*

Chem. Biol. 23, 122–136.

(90) Abd-Elrahman, I.; Kosuge, H.; Wises Sadan, T.; Ben-Nun, Y.; Meir, K.; Rubinstein, C.; Bogyo, M.; McConnell, M. V; Blum, G. (2016) Cathepsin Activity-Based Probes and Inhibitor for Preclinical Atherosclerosis Imaging and Macrophage Depletion. *PLoS One 11*, e0160522.

(91) Wang, J.; Zhang, C.-J.; Zhang, J.; He, Y.; Lee, Y. M.; Chen, S.; Lim, T. K.; Ng, S.; Shen, H.-M.; Lin, Q. (2015) Mapping Sites of Aspirin-Induced Acetylations in Live Cells by Quantitative Acid-Cleavable Activity-Based Protein Profiling (QA-ABPP). *Sci. Rep. 5*, 7896.

(92) Jessani, N.; Niessen, S.; Wei, B. Q.; Nicolau, M.; Humphrey, M.; Ji, Y.; Han, W.; Noh, D.-Y.; Yates, J. R.; Jeffrey, S. S.; et al. (2005) A Streamlined Platform for High-Content Functional Proteomics of Primary Human Specimens. *Nat. Methods 2*, 691–697.

(93) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; et al. (2016) Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. *Nature 534*, 570–574.

(94) Xin, B.-T.; de Bruin, G.; Verdoes, M.; Filippov, D. V; van der Marel, G. A.; Overkleeft, H. S. (2014) Exploring Dual Electrophiles in Peptide-Based Proteasome Inhibitors: Carbonyls and Epoxides. *Org. Biomol. Chem. 12*, 5710–5718.

(95) Hu, Q.-Y.; Allan, M.; Adamo, R.; Quinn, D.; Zhai, H.; Wu, G.; Clark, K.; Zhou, J.; Ortiz, S.; Wang, B.; et al. (2013) Synthesis of a Well-Defined Glycoconjugate Vaccine by a Tyrosine-Selective Conjugation Strategy. *Chem. Sci. 4*, 3827–3832.

(96) Kallemeijn, W. W.; Li, K.-Y.; Witte, M. D.; Marques, A. R. A.; Aten, J.; Scheij, S.; Jiang, J.; Willems, L. I.; Voorn-Brouwer, T. M.; van Roomen, C. P. A. A.; et al. (2012) Novel

Activity-Based Probes for Broad-Spectrum Profiling of Retaining β-Exoglucosidases In Situ and In Vivo. *Angew. Chemie Int. Ed. 51*, 12529–12533.

(97) Kathman, S. G.; Xu, Z.; Statsyuk, A. V. (2014) A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases. *J. Med. Chem. 57*, 4969–4974.

(98) Wurm, F. R.; Klok, H. A. (2013) Be Squared: Expanding the Horizon of Squaric Acid-Mediated Conjugations. *Chem. Soc. Rev. 42*, 8220–8236.

(99) Saksena, R.; Zhang, J.; Kováč, P. (2005) Immunogens from a Synthetic Hexasaccharide Fragment of the O-SP of Vibrio Cholerae O:1, Serotype Ogawa. *Tetrahedron Asymmetry*.

(100) Kaur, D.; Guerin, M. E.; Skovierová, H.; Brennan, P. J.; Jackson, M. (2009) Chapter 2: Biogenesis of the Cell Wall and Other Glycoconjugates of Mycobacterium Tuberculosis. *Adv. Appl. Microbiol. 69*, 23–78.

(101) Poulin, M. B.; Lowary, T. L. (2016) Chemical Insight into the Mechanism and Specificity of GlfT2, a Bifunctional Galactofuranosyltransferase from Mycobacteria. *J. Org. Chem. 81*, 8123–8130.

(102) Besra, G. S.; Khoo, K. H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J. (1995) A New Interpretation of the Structure of the Mycolyl-Arabinogalactan Complex of Mycobacterium Tuberculosis as Revealed through Characterization of Oligoglycosylalditol Fragments by Fast-Atom Bombardment Mass Spectrometry and 1H Nuclear Magnetic Resonance . *Biochemistry 34*, 4257–4266.

(103) Brown, C. D.; Rusek, M. S.; Kiessling, L. L. (2012) Fluorosugar Chain Termination Agents as Probes of the Sequence Specificity of a Carbohydrate Polymerase. *J. Am. Chem. Soc. 134*, 6552–6555.

(104) Sassetti, C. M.; Boyd, D. H.; Rubin, E. J. (2003) Genes Required for Mycobacterial

Growth Defined by High Density Mutagenesis. *Mol. Microbiol. 48*, 77–84.

(105) Pan, F.; Jackson, M.; Ma, Y.; McNeil, M. (2001) Cell Wall Core Galactofuran Synthesis Is Essential for Growth of Mycobacteria. *J. Bacteriol. 183*, 3991–3998.

(106) May, J. F.; Splain, R. A.; Brotschi, C.; Kiessling, L. L. (2009) A Tethering Mechanism for Length Control in a Processive Carbohydrate Polymerization. *Proc. Natl. Acad. Sci. 106*, 11851–11856.

(107) Zegzouti, H.; Engel, L.; Hennek, J.; Alves, J.; Vidugiris, G.; Goueli, S. (2013) Detection of Glycosyltransferase Activities with Homogenous Bioluminescent UDP Detection Assay. *Glycobiology 23*, 1340–1341.

(108) Dykhuizen, E. C.; Kiessling, L. L. (2009) Potent Ligands for Prokaryotic UDP-Galactopyranose Mutase That Exploit an Enzyme Subsite. *Org. Lett. 11*, 193–196.

(109) Splain, R. A.; Kiessling, L. L. (2010) Bioorganic & Medicinal Chemistry Synthesis of Galactofuranose-Based Acceptor Substrates for the Study of the Carbohydrate Polymerase GlfT2. *Bioorg. Med. Chem. 18*, 3753–3759.

(110) Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K. S. (2012) Tetrameric Structure of the GlfT2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial Arabinogalactan. *J. Biol. Chem. 287*, 28132–28143.

(111) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. (2018) SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res. 46*, W296–W303.

(112) Claverie, J.-M. (2001) What If There Are Only 30,000 Human Genes? *Science (80-.). 291*, 1255 LP – 1257.

(113) Walke, D. W.; Han, C.; Shaw, J.; Wann, E.; Zambrowicz, B.; Sands, A. (2001) In Vivo Drug Target Discovery: Identifying the Best Targets from the Genome. *Curr. Opin. Biotechnol. 12*, 626–631.

(114) Drews, J. (2000) Drug Discovery: A Historical Perspective. *Science (80-.). 287*, 1960–1964.

(115) Hopkins, A. L.; Groom, C. R. (2002) The Druggable Genome. *Nat. Rev. Drug Discov. 1*, 727–730.

(116) Shoichet, B. K. (2004) Virtual Screening of Chemical Libraries. *Nature 432*, 862–865.

(117) Macarron, R. (2006) Critical Review of the Role of HTS in Drug Discovery. *Drug Discov. Today 11*, 277–279.

(118) Silverman, L.; Campbell, R.; Broach, J. R. (1998) New Assay Technologies for High-Throughput Screening. *Curr. Opin. Chem. Biol. 2*, 397–403.

(119) Oprea, T. I.; Matter, H. (2004) Integrating Virtual Screening in Lead Discovery. *Curr. Opin. Chem. Biol. 8*, 349–358.

(120) Forino, M.; Johnson, S.; Wong, T. Y.; Rozanov, D. V.; Savinov, A. Y.; Li, W.; Fattorusso, R.; Becattini, B.; Orry, A. J.; Jung, D.; et al. (2005) Efficient Synthetic Inhibitors of Anthrax Lethal Factor. *Proc. Natl. Acad. Sci. U. S. A. 102*, 9499–9504.

(121) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Woodhead, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; et al. (2005) Identification of Novel P38α MAP Kinase Inhibitors Using Fragment-Based Lead Generation. *J. Med. Chem. 48*, 414–426.

(122) Wada, C. K.; Holms, J. H.; Curtin, M. L.; Dai, Y.; Florjancic, A. S.; Garland, R. B.; Guo, Y.; Heyman, H. R.; Stacey, J. R.; Steinman, D. H.; et al. (2002) Phenoxyphenyl Sulfone

N-Formylhydroxylamines (Retrohydroxamates) as Potent, Selective, Orally Bioavailable Matrix Metalloproteinase Inhibitors. *J. Med. Chem. 45*, 219–232.

(123) Martin, Y. C. (1997) Challenges and Prospects for Computational Aids to Molecular Diversity. *Perspect. Drug Discov. Des. 7--8*, 159–172.

(124) Davies, J. (1996) Origins and Evolution of Antibiotic Resistance. *Microbiologia 12*, 9–16.

(125) Zaman, S. Bin; Hussain, M. A.; Nye, R.; Mehta, V.; Mamun, K. T.; Hossain, N. (2017) A Review on Antibiotic Resistance: Alarm Bells Are Ringing. *Cureus 9*.

(126) Hajduk, P. J.; Greer, J. (2007) A Decade of Fragment-Based Drug Design: Strategic Advances and Lessons Learned. *Nat. Rev. Drug Discov. 6*, 211–219.

(127) Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; et al. (1997) Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J. Am. Chem. Soc. 119*, 5818–5827.

(128) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; et al. (2005) An Inhibitor of Bcl-2 Family Proteins Induces Regression of Solid Tumours. *Nature 435*, 677–681.

(129) Sunesis Pharmaceuticals-Products in Development https://www.sunesis.com/products.php.

(130) Astex Pharmaceuticals Clinical Pipeline https://astx.com/research-development/clinical-pipeline/.

(131) Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. (2017) Global Profiling of Lysine Reactivity and Ligandability in the Human Proteome. *Nat. Chem. 9*, 1181–1190.

(132) Mukherji, M. (2001) Structure-Function Analysis of Phytanoyl-CoA 2-Hydroxylase Mutations Causing Refsum's Disease. *Hum. Mol. Genet. 10*, 1971–1982.

(133) McDonough, M. A.; Kavanagh, K. L.; Butler, D.; Searls, T.; Oppermann, U.; Schofield, C. J. (2005) Structure of Human Phytanoyl-CoA 2-Hydroxylase Identifies Molecular Mechanisms of Refsum Disease. *J. Biol. Chem. 280*, 41101–41110.

(134) Mihalik1, S. J.; Morrell, J. C.; Kim, D.; Sacksteder, K. A.; Watkins, P. A.; Gould, S. J. (1997) Identification of Pahx, a Refsum Disease Gene. *Nat. Genet. 17*, 189.