Development of Assays and Irreversible Inhibitors for ATG4B; A Key Cysteine Protease Implicated in the Process of Autophagy

by

Thanh-Giau Nguyen

B.Sc., (Hons., Chemistry), Queen's University, 2008

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Chemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Spring 2015

Approval

Name: **Thanh-Giau Nguyen**

Degree: **Doctor of Philosophy (Chemistry)**

Title: ***Development of Assays and Irreversible Inhibitors for ATG4B; A Key Cysteine Protease Implicated in the Process of Autophagy***

Examining Committee: **Chair:** Dr. Robert A. Britton
Associate Professor

Dr. Robert N. Young
Senior Supervisor
Professor

Dr. Andrew J. Bennet
Supervisor
Professor

Dr. David J. Vocadlo
Supervisor
Professor

Dr. Dipankar Sen
Internal Examiner
Professor

Dr. David Grierson
External Examiner
Professor
Department of Pharmaceutical Sciences
University of British Columbia

Date Defended/Approved: April 21, 2015

Partial Copyright Licence

Abstract

Autophagy is a biological process responsible for the sequestration and degradation of cellular components for the purposes of maintaining homeostasis. ATG4B is a cysteine protease that plays a key role in the initiation of autophagy by cleaving the protein proLC3 revealing a C-terminal glycine residue and to form LC3-I. This glycine then becomes lipidated with phosphatidylethanolamine (PE) to form LC3-II which is inserted into partial protomembrane fragments in the golgi which stimulates formation of autophagosomes. ATG4B also cleaves PE from the membrane bound LC3-II to reform and recycle LC3-I and to facilitate fusion of the autophagosomes with liposomes where the contents are subsequently digested producing energy and amino acids for protein synthesis. As cancer cells frequently “hijack” autophagy as a survival and resistance mechanism against therapy, autophagy has recently emerged as a potential therapeutic target for treating cancer.

The development of two different assay methods for measuring ATG4B activity are reported. The first, LC/MS-based, assay monitors the cleavage of proLC3 to LC3-I based on the ratio of peak heights detected. This method has the added utility of being able to detect enzyme integrity throughout the reaction. The second, FRET-based, assay involves a novel YFP-LC3B-EmGFP doubly fluorescent protein substrate developed for the purposes of large scale screening of compound libraries. The FRET-based assay was successfully used to screen 5000 compounds in the commercial LOPAC and KD2 compound libraries with an overall hit rate of 0.6% and 0.5% respectively.

Irreversible inhibitors of ATG4B were designed and synthesized based on information from the screening results as well as previous work on putative peptide substrates conducted by Dr. Nag Kumar (Young Lab, SFU). Inhibitors of the halo-methyl ketone type were developed and tested for inhibitory activity against ATG4B. Fluorescent analogues of these irreversible inhibitors were further developed and optimized for *in cellulo* potency. Kinetic analysis and digestion studies revealed the inhibitors were indeed active-site directed ATG4B inhibitors.

Keywords: Autophagy;ATG4B;Irreversible inhibition;FRET;assay;LC3B

Dedication

This is dedicated to my family and friends, all of whom shared a part in this journey.

Acknowledgements

I would like to thank Professor Robert Young for allowing me to work under his guidance and supervision.

I thank Dr. Aled Edwards and Dr. Anias Abdalin and the Structural Genomics Consortium in Toronto Ontario for providing ATG4B and its substrates.

I thank Dr. David Vocadlo, Dr. Matt Macauley and David Shen for their guidance in molecular biology, biochemistry and enzyme kinetics.

I am greatly indebted to all members of the Young Lab at Simon Fraser University, past and present, for their contributions to my work, experience, and character.

I thank Dr. Hongwen Chen for his LC/MS guidance and training.

I thank Dr. Sharon Gorski and her coworkers for the cell studies.

I thank the members of the Centre for Drug Research and Development; Dr. Tom Pfiefer, Dr. Steve Jones, Dr. Jianghong An, Dr. Nicollette Honson for their contributions to the screening effort.

Finally, I would like to thank all my family and friends for their continued support in all my endeavours.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACC	7-amino-4-carboxymethylcoumarin
ATG4B	Autophagin related enzyme 4B
CDRD	Centre for Drug Research and Development
CNS	Central nervous system
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DMF	Dimethylformamide
DMP	Dess-Martin Periodinane
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EmGFP	Emerald-green fluorescent protein
EtOAc	Ethyl acetate
FRET	Fluorescence resonance energy transfer
hex	Hexanes
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	1-Hydroxybenzotriazole
HPLC	High performance liquid chromatography
HTS	High throughput screening
KD2	Known drugs library 2
LC3B	Microtubule associated light chain 3 protein B
LC/MS	Liquid chromatography coupled mass spectrometer
LOPAC	Library of pharmaceutically active compounds
NMM	N-Methylmorpholine
NMP	N-Methyl-2-pyrrolidone
PE	Phosphotidylethanolamine
RNA	Ribonucleic acid
RNAi	RNA interference
SGC	Structural Genomics Consortium
TFA	Trifluoroacetic acid
YFP	Yellow fluorescent protein

Chapter 1. Introduction

1.1. Autophagy

Autophagy, or self-eating, is a general term used for any intracellular process involving the degradation of cytosolic components by lysosomal enzymes. The end result of autophagy is the complete digestion of the substrate into its essential components - proteins into amino acids for example. This process can be observed in all eukaryotic cells which indicates the importance of this evolutionarily conserved process is.[1,2] There are 3 different types of autophagy; macroautophagy, microautophagy, and chaperone-mediated autophagy. For the purposes of this thesis, the term autophagy will be used to refer to the process of macroautophagy. In macroautophagy, any components requiring degradation are first engulfed by an isolation membrane which begins in the Golgi apparatus as a double layer membrane fragment known as a phagophor and then matures to a vesicle known as an autophagosome. The autophagosome then fuses with lysosomes to deliver its contents for degradation to form an autophagolysosome.

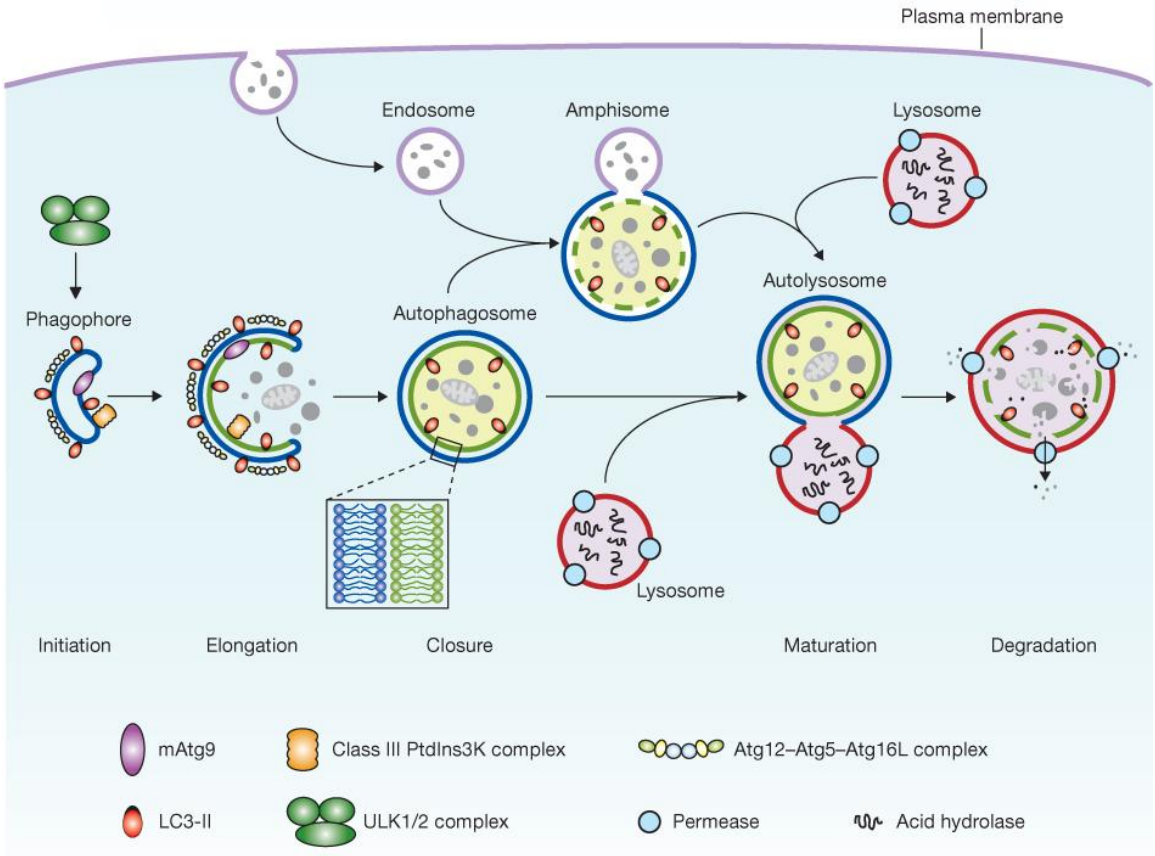


Figure 1a: Illustration of the process of mammalian autophagy, which is initiated by the formation of a phagophore, followed by elongation and expansion of the phagophore, completion and closure of the autophagosome (a double membrane structure, which surrounds a portion of the cytoplasm), maturation and fusion with an endosome and/or lysosome (known as an autolysosome), degradation and breakdown of the autophagosome inner membrane as well as cargo through acid hydrolases inside the autolysosome, and finally recycling of the resulting macromolecules through permeases. Also depicted are the ULK1 and ULK2 complexes that are required for autophagy induction, class III PtdIns3K complexes that are involved in autophagosome formation, mammalian Atg9 (mAtg9) that potentially contributes to the delivery of membrane to the forming autophagosome and two conjugation systems, the LC3-II and Atg12–Atg5–Atg16L complex, which are proposed to function during elongation and expansion of the phagophore membrane.[3,4]

In mammalian cells, autophagy is regulated by metabolic stress and nutrient availability and it is suggested that this process is essential for cellular homeostasis. Autophagy has also been linked to a number of diseases including neurodegeneration[5,6] and more recently cancer. It is proposed that certain tumours, due to the tumour microenvironment or genetic aberrations, become "addicted" to autophagy for survival and these tumour types may be the most responsive towards autophagy inhibition.[7–9] However, autophagy appears to have a dual role in cancer, which depends on the state of progression of the disease.[10–15,7,8] In the early stages of tumour progression, autophagy is up-regulated as a mechanism to survive periods of rapid growth when microvasculature has not yet been properly established.[2] This process can also help protect cancer cells from cytotoxic anticancer drugs by removing damaged cellular components or toxic proteins, thus imparting a degree of resistance to treatment. In contrast to this survival role, it has also been suggested that autophagy can serve as a mechanism to induce cell death in cancer cells primarily when the apoptosis mechanism has been disabled in an attempt to trick the cancer cells into inducing its own cell death.[16]

To date, there are no known potent and specific inhibitors of autophagy. Recent studies have used RNAi to interfere with the expression of autophagy specific proteins. Other strategies to inhibit autophagy involve the use of lysosomotropic agents, such as hydroxychloroquine (HCQ), but such non-specific agents are general inhibitors of lysosomal enzyme activity that act by raising the local pH. While lysosome targeting may be a viable strategy to inhibit autophagy, widespread and sustained systemic use of such an agent could result to a build-up of unprocessed organelles and proteins which in turn may lead to complications such as increased toxicity and phospholipidosis.[17]

There are a number of autophagy associated proteins commonly referred to as Atgs or autophagy related proteins. Atg8 (LC3) is a ubiquitin-like protein that can be conjugated to phosphatidylethanolamine (PE). In yeast, the conjugated form is referred to as Atg8-PE. The mammalian homologs of Atg8 are subdivided into two subfamilies: LC3 (microtubule-associated protein 1 light chain 3) and GABARAP ($GABA_A$ receptor-associated protein). The former is comprised of LC3A, B, B2 and C, while the latter family is comprised of GABARAP, GABARAPL1/Atg8L/GEC1 (GABAA receptor-associated protein like 1/Glandular Epithelial Cell 1), GABARAPL2/GATE-16/GEF2

(GABAA receptor-associated protein like 2/Golgi-associated ATPase enhancer of 16kDa/ganglioside expression factor 2) and GABARAPL3 (GABAA receptor-associated protein like 3). Both the LC3 and GABARAP subfamilies, with the exception of GABARAPL3, have been demonstrated to participate in phagophore biogenesis.[18,19] While LC3 and GABARAP subfamilies are commonly referred to as autophagy related proteins, there is no guarantee that they are only implicated in autophagy. For example, certain types of mitophagy (selective autophagy of mitochondria) are highly dependent on GABARAP and less dependent on LC3 proteins.[20] Nevertheless, LC3 has been the primary Atg8 homolog examined in mammalian cells and is typically characterized as an autophagosome marker.[21]

ATG9 is the only integral membrane ATG protein and is essential for autophagosome formation.[22] A unique feature of ATG9 is that it localizes to multiple discrete puncta, whereas most other ATG proteins are detected primarily in a single punctum or diffusely in the cytosol. It has been suggested that it is possible to follow this protein as an indication of ATG13 function.[23]

ATG12-ATG5 is a conjugated system of proteins that interact with ATG16L1 (in mice) to form a ~350 kDa complex that is also essential for autophagy.[24] ATG12 is conjugated to ATG5 through an isopeptide bond between the C-terminal glycine of ATG12 and the Lys^{149} residue of ATG5. This conjugation between ATG12-ATG5 is facilitated by E1-like enzymes (ubiquitin-like activating enzyme) ATG7 and ATG10.[24] It has been demonstrated that cells deficient in ATG12 are defective in LC3 modification, suggesting that the function of these two ubiquitin-like systems are closely related.[24,25] ATG12-ATG5 conjugation has been used in some studies to monitor autophagy. However, it was discovered that essentially all of the ATG5 and ATG12 proteins exist in the conjugated form and expression levels do not change under short-term starvation.[26–28] Therefore, monitoring ATG12-ATG5 conjugation may not be a useful method for following the induction of autophagy.

ATG4 is a cysteine protease and has 4 isotypes (ATG4A-D). Of the 4 homologues, only ATG4B is reported to efficiently cleave pro-LC3B.[29,30] ATG4B is implicated in more than one step of the autophagy cascade. Of the ATG4 isotypes, ATG4B is most effective at cleaving pro-LC3B (also known as MAP1LC3B and ATG8)

and does so near the N-terminus to liberate a free terminal glycine residue and provide LC3-I. LC3-I is then lipidated, at the C-terminal glycine residue, by conjugation with phosphatidylethanolamine by the action of autophagins ATG7 and ATG3 (an E2-like or ubiquitin-like conjugating enzyme) and the ATG12-ATG5 complex to form LC3-II, which is inserted into forming membranes to help initiate autophagosome formation. In addition, ATG4B cleaves the lipid-conjugated membrane bound LC3B-II to facilitate lysosomal fusion and recycling of LC3-I. While interest in ATG4B as a therapeutic target grows, no potent selective inhibitors of this protease have yet been reported.[31–33]

Considering the pivotal role of ATG4B in autophagosome formation and trafficking, a small molecule inhibitor of ATG4B could serve as a useful tool to study the effects of autophagy inhibition and to examine its specific potential as an anti-cancer therapy both alone and in combination with existing cancer treatments. While the long-term aim of our research is to develop a potent, selective and *in vivo* active inhibitor, my research has been directed to the development of assays to measure ATG4B activity *in vitro*, to screen for potential lead inhibitors and to discover and characterize irreversible inhibitors as an important first step in this effort. The next section will provide a brief summary of the type of kinetic analysis used in this investigation.

1.2. Enzyme Kinetics

ATG4B is an enzyme in the papain family of proteases and we have studied its kinetic behaviour and inhibition using kinetic analyses as first described by Leonor Michaelis and Maude Menten. Despite their pioneering efforts, they were never recognized by the Nobel Committee. This section is a summary of the work that lead to what is now known as the Michaelis-Menten equation of enzyme kinetics. This equation can define the characteristics of many enzyme-substrate systems and thus offers broad utility when investigating enzyme kinetics. However, there are cases when enzymes will not behave according to the Michaelis-Menten equation. Nevertheless, the Michaelis-Menten equation holds true for many enzyme-substrate systems and remains an important tool for investigating enzyme kinetics.

Deriving the Michaelis-Menten Equation

Chemical kinetics is the study of the rates at which chemical reactions occur. Enzyme kinetics therefore, is the study of the rates at which enzymatic reactions occur. The study began in 1902 when Adrian Brown started investigating the rate of hydrolysis of sucrose by the yeast enzyme invertase, now known as β-fructofuranosidase:

$$\text{Sucrose} + H_2O \rightarrow \text{glucose} + \text{fructose} \qquad (1)$$

Brown demonstrated that when the concentration of sucrose was much higher than the enzyme concentration, the reaction rate became independent of sucrose concentration. He proposed that the overall reaction was comprised of two elementary reactions in which the enzyme and substrate form a complex that subsequently decomposes into product and enzyme:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (2)$$

Here, E, S, ES, and P refer to enzyme, substrate, enzyme-substrate complex, and products respectively, while k_1, k_{-1}, and k_2 refer to the individual rate constants associated with each step. According to this model, when substrate levels are high enough to completely convert all the enzyme into the enzyme-substrate complex, the second step becomes rate limiting and the overall rate becomes independent of substrate concentration. The rate expression for this would be:

$$v = \frac{d[P]}{dt} = k_2 ES \qquad (3)$$

The overall rate of production of ES would then be:

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \qquad (4)$$

This equation however, could not be integrated unless two simplifying assumptions were made. In 1913, Leonor Michaelis and Maude Menten assumed that $k_{-1} >> k_2$, so the first step (formation of the ES complex) reaches equilibrium. In recognition of their pioneering work, the non-covalently bound enzyme-substrate complex is known as the

Michaelis complex. The second assumption was first proposed by G. E. Briggs and John B. S. Haldane in 1925. They assumed that the amount of Michaelis complex remains constant throughout the enzymatic reaction until the substrate is nearly exhausted. That is:

$$\frac{d[ES]}{dt} = 0 \qquad (5)$$

This is the so-called *steady state assumption*. In order to be of any use, kinetic expressions must be formulated in terms of experimentally measurable quantities. The quantities [E] and [ES] are not directly measurable, but the total enzyme concentration, $[E]_T$, is readily determined.

$$[E]_T = [E] + [ES] \qquad (6)$$

The rate equation for enzymatic reactions can be derived by combining equation (4), (5), and (6):

$$k_1([E]_T - [ES])[S] = (k_{-1} + k_2)[ES] \qquad (7)$$

Upon rearrangement, 7 becomes:

$$[ES](k_{-1} + k_2 + k_1[S]) = k_1[E]_T[S] \qquad (8)$$

By dividing both sides by k_1 and solving for [ES],

$$[ES] = \frac{[E]_T[S]}{K_M + [S]} \qquad (9)$$

Where K_M is known as the Michaels constant and is defined,

$$K_M = \frac{k_{-1} + k_2}{k_1} \qquad (10)$$

We can now go back to the original rate expression (3) and substitute for experimentally measurable quantities:

$$v_o = (\frac{d[P]}{dt})_{t=t_s} = k_2[ES] = \frac{k_2[E]_T[S]}{K_M + [S]} \qquad (11)$$

This expression refers to the initial velocity of the reaction, v_o, where t_s is the time when steady state is achieved (usually milliseconds after t = 0). Initial velocity is operationally taken as the velocity measured before ~10% of substrate has been converted to product. The use of initial velocity, rather than velocity, minimizes complicating factors such as inhibition of enzyme by product, reversible reactions, and progressive inactivation of the enzyme. The maximum velocity, or V_{max}, occurs when the substrate concentration is so high that the enzyme becomes saturated, which is to say when the enzyme is entirely in the [ES] form and therefore one can substitute [ES] by $[E]_T$ (a known quantity),

$$V_{\max} = k_2[E]_T \qquad (12)$$

By combining (11) and (12) we obtain,

$$v_o = \frac{V_{\max}[S]}{K_M + [S]} \qquad (13)$$

This expression is the Michaelis-Menten equation and is the basic equation of enzyme kinetics. The Michaelis constant has a simple operational definition. According to (13), when $K_M = [S]$, $v_o = V_{max}/2$. So K_M is the substrate concentration at which the velocity of the reaction is half of the maximum. Therefore, if an enzyme has a small K_M value, it achieves maximal catalytic efficiency at low substrate concentrations. The magnitude of K_M can vary widely depending on the nature of the enzyme and substrate.[34]

Analysis of Kinetic Data

There are several methods for determining the parameters of the Michaelis-Menten equation. We have already established that at very high substrate concentrations [S], the initial velocity v_o asymptotically approaches V_{max}. In practice however, it is very difficult to assess V_{max} accurately from plots of v_o vs [S]. V_{max} is a theoretical asymptote and in reality it would not be possible to get infinitely high

substrate concentrations. An alternative approach was proposed by Hans Lineweaver and Dean Burk which uses the reciprocal of the Michaelis-Menten equation:

$$\frac{1}{v_o} = (\frac{K_M}{V_{\max}})\frac{1}{[S]} + \frac{1}{V_{\max}} \qquad (14)$$

A plot of $1/v_o$ vs 1/[S] is referred to as the Lineweaver-Burk plot or double-reciprocal plot, where the slope of the line is K_M/V_{max}, the y-intercept is $1/V_{max}$ and the extrapolated x-intercept is $-1/K_M$. Unfortunately, this method of analysis is far from perfect. One disadvantage is that most experiments involve relatively high levels of [S] so the data points tend to crowd the left side of the graph. Likewise, for small values of [S], small errors in v_o will lead to larger errors in $1/v_o$ and hence larger errors in K_M and V_{max}. Modern day tools and sophisticated computer software can now directly analyze Michaelis-Menten plots using non-linear regression analysis to provide far more accurate calculations of the kinetic parameters.[35]

Catalytic efficiency

In the Michaelis-Menten model, k_{cat} is known as the catalytic constant ($k_{cat} = k_2$). k_{cat} can be defined as:

$$k_{cat} = \frac{V_{\max}}{[E]_T} \qquad (15)$$

This equation also represents the turnover number of the enzyme or the number of substrate molecules converted per molecule of enzyme. At very low [S], such that [S] << K_M, very little [ES] is formed, therefore $[E] \approx [E]_T$, equation 11 becomes:

$$vo \approx (\frac{k_2}{K_M})[E]_T[S] \approx (\frac{k_{cat}}{K_M})[E][S] \qquad (16)$$

k_{cat}/K_M is a second-order rate constant for the interaction between enzyme and substrate and is therefore a measure of the enzyme's catalytic efficiency.

Enzyme Inhibition

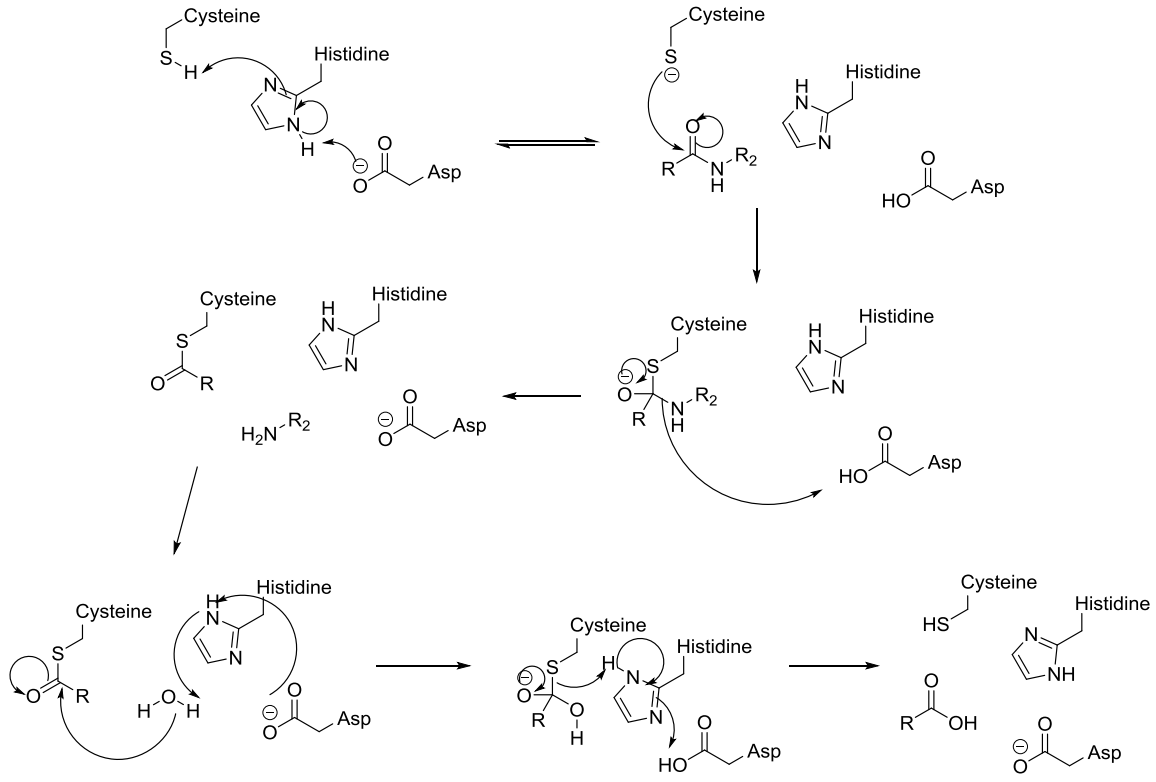
Inhibitors of an enzyme can alter (suppress) the activity of an enzyme by combining with it in such a way that alters the binding of substrate or impairs the ability of the enzyme to act without altering the substrate binding.

Many enzyme inhibitors have structures that closely resemble the substrate structure and compete for the active site, but either react slowly compared to substrate or not at all. Inhibitors are commonly used to probe the conformational or chemical nature of the enzyme's active site in an attempt to elucidate its catalytic mechanism. There are various mechanisms through which inhibitors can act. A competitive inhibitor is a substance that will compete for and bind to the enzyme active site normally reserved for substrate. Such a substance usually resembles the substrate structure enough to bind the active site, but differs enough to be unreactive (or less reactive). An uncompetitive inhibitor is a substance that will bind the enzyme-substrate complex, but not free enzyme. Mixed-mode inhibitors (often termed non-competitive inhibitors) can bind both free enzyme and the enzyme-substrate complex. Another type of inhibitors, and the major focus of this thesis, are irreversible inhibitors. Irreversible inhibitors will bind free enzyme (usually covalently) and convert it to an inactive form. As more and more irreversible inhibitor reacts with an enzyme, the more inactivated it becomes until it is eventually totally and permanently inactivated. This time-dependence must be accounted for when examining the kinetics of inhibition for irreversible inhibitors. The details of the kinetic analysis will be discussed in chapter 4.

Mechanism of Action of Cysteine Proteases

In general, the amino acids directly involved in an enzymatic reaction are referred to as the *catalytic triad* or *catalytic dyad* depending on the family of enzyme. For cysteine proteases, the catalytic nucleophile is cysteine and is generally flanked by a basic residue and sometimes an acidic residue. The catalytic cysteine and histidine residues are generally conserved and referred to as the *catalytic dyad*. A *catalytic triad* includes an acidic residue which can be aspartic or glutamic acid. ATG4B has a proposed triad consisting of Cys^{74}, Asp^{278}, and His^{280}.[36] Below is an example of a cysteine protease cleaving a peptide bond.

Scheme 1



1.3. A Medicinal Chemistry Approach

Medicinal chemistry is an interdisciplinary science which combines organic chemistry, pharmaceutical chemistry, molecular biology and biochemistry in an attempt to produce a 'drug'. A drug can be classified as any chemical agent that ellicit biological responses useful for treating disease. For example, morphine is a powerful drug that interacts with receptors in our central nervous system (CNS) and the biological response is relief of pain. The general goal of medicinal chemistry is to provide drugs for the betterment of humanity and to improve the quality of life for individuals suffering from sickness and disease.

Formulating a drug from start to finish

Mankind has used herbs and tinctures for thousands of years, but it was only since the mid 19th century that a serious effort was made to isolate and purify the active ingredients in such remedies. Since then, a large variety of biologically active compounds have had their structures determined; cocaine from coca leaves, morphine from opium and quinine from the bark of the cinchona tree to name a few. Inspired by these natural products, major synthetic efforts were conducted where chemists made literally thousands of analogues in an attempt to improve on what nature had already provided. The majority of this work was conducted with no clear reason or design, but from the results came an appreciation for tactics that generally worked and a pattern for drug development evolved. Nowadays, drug development involves several discrete and defined steps: validating a target, screening natural and non-natural compounds for biological activity, isolation and purification (or synthesis) of the active compound, structure determination (in the case of natural mixtures), derivation of structure-activity relationship studies, synthesis of analogues, determining mechanism or mode of action, pre-clinical experiments, clinical trials, and finally bringing a useful compound to the market.

Validating ATG4B as a target

In general, one of the first steps towards drug discovery is validating a target. This requires applying or developing well established, or at least validated, assay methods for enzymes or receptors of interest. In our case, upon the initiation of this project there were no published assays or known inhibitors of ATG4B. There were however, studies using nonspecific inhibitors of the lysosome such as the anti-malarial drug chloroquine, which did show inhibition of autophagy in cells. This compound, alongside more active analog LYSO5, have even been tested in human clinical trials in cancer, but with mixed results due to their weak activity and with interpretation of effects complicated by the general inhibition of lysosomal activity. In addition, there were some studies of inhibition of ATG4 and its homologs using gene knockouts in mice and siRNA in cells which have been encouraging, but not definitive as there remains a possibility that in knocking down ATG4B, one of its homologs may compensate.[37] While we were hoping to develop a selective inhibitor of ATG4B it was likely that, due to similarity, such inhibitors would also inhibit the other isotypes. Thus, the long-term plan of the project in

the Young lab was to develop ATG4B inhibitors and to ultimately characterize them for selectivity over the other ATG's and also unrelated cysteine proteases.

ATG4B is structurally interesting and possesses an auto-regulatory loop which acts to block the active site. Below is a figure to help illustrate the active site of ATG4B.

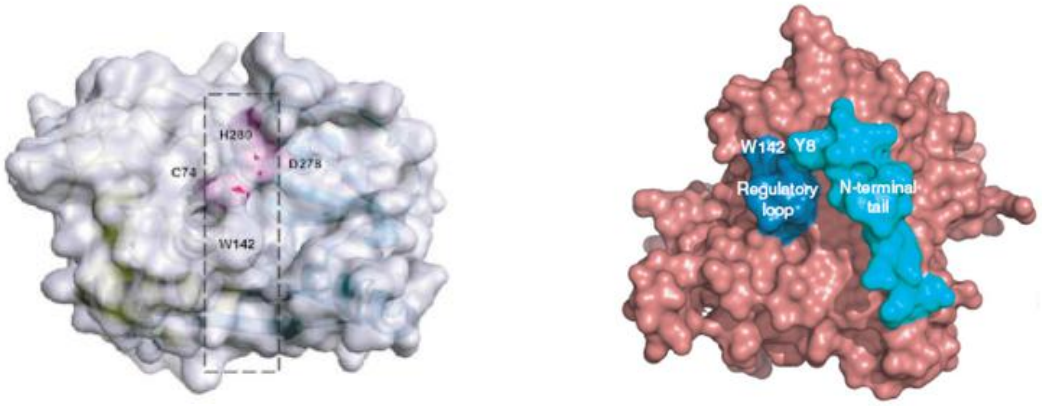


Figure 1b: Left) X-ray crystal structure of ATG4B highlighting the active-site residues Cys^{74}, His^{280}, Asp^{278}. **Right)** X-ray crystal structure of ATG4B in an inactive form highlighting a regulatory loop that blocks the active-site. Studies have demonstrated that large conformational changes occur as proLC3B binds to ATG4B.[36]

The first goal however was to develop reliable methods to monitor ATG4B activity. This would require access to ATG4B itself along with its isotypes and known substrates. There are 4 homologues of ATG4: A, B, C and D. It has been shown that ATG4B acts on 3 different substrates, GABARAP, GATE-16, and LC3.[38] While ATG4B can cleave all 3 substrates, it is the only one out of the four to process LC3 efficiently.[39] The Structural Genomics Consortium (SGC) in Toronto Ontario was kind enough to supply us with quantities of ATG4B enzyme as well as all three substrates in an effort to kick start our research project. Later in the project they also supplied us with ATG4A and more recently ATG4D. Unfortunately the initial batch of substrates appeared to have been degraded by the time they reached our lab. Nevertheless, the samples of ATG4B and LC3B were largely intact and these initial quantities were critical to allow the investigation and development of the now 3 different assay methods established in this research project. Subsequently in our lab, Dr. Suzana Kovacic expressed and purified

ATG4B and proLC3B and related substrates in quantity to help sustain our efforts and to allow high throughput screening. My contribution has been to develop and validate two of these assay methods, which were then optimized for high-throughput screening (HTS) by the CDRD (Center for Drug Research and Development). The details of which are presented in the next chapter of this thesis.

Screening compounds for biological activity

In the past, screening often involved evaluation of extracts from plants, bacteria or animals as well as pure compounds, but more recently, with the advances in computer technology, new *in silico* screening methods have emerged.[40] Libraries of synthetic molecules have become commercially available as possible sources of simple lead structures. The *in silico* screening approach generally requires X-ray crystallographic structures of the target protein, however there are also pharmacophore searching methods that do not require X-ray data. In our case we were fortunate to have several published high resolution X-ray structures for ATG4B including one where a catalytically incompetent enzyme (active site cysteine mutated to alanine) is bound to LC3.[41] For *in silico* screening, a database of known chemical structures is modelled (one at a time and with numerous low energy conformers examined for each molecule) into the active site of the enzyme. The program then calculates how each chemical structure binds to the active site, which should allow identification of potentially active compounds or "hits" as well as a ranking of calculated potencies. This approach however, does not give any indication of selectivity for the target protein. This method requires considerable computing power, but if available, can quickly screening thousands of compounds without having to do any experimental work in the laboratory. The best "hits" can then be purchased from suppliers for experimental testing with the target enzyme itself. Even with the best computing technology, the predictive models are not completely representative of true binding potency and at best improve the "hit rate" for screening on the target enzyme (increases in hit rates from 0.5% to 5-10% are often found when this is done well). Inevitably, drug discovery still remains in the hands of chemists willing to synthesize, optimize and test such compounds. Nevertheless, *in silico* screening remains a useful tool for finding leads and was useful in our case. However, the *in silico* screening that was part of our project was something I was not directly involved with, although my assays were used to follow-up hits. My personal

efforts resulted in useful assays for HTS which also resulted in identification of a number of “hit” structures.

Structure-activity relationship studies

After “hits” are verified and prioritized based on potency and estimated ease of synthesis of analogs, the next step in drug development would be to optimize any validated hits obtained from the *in silico* screening or HTS efforts into a lead compound. A lead compound should already be relatively potent, while its chemical structure should be readily amenable to structural modification through chemistry. This is largely up to the judgement of experienced medicinal chemists who will then work with the initial structure and synthesize numerous analogues in an attempt to improve on the original structure in terms of potency, selectivity for the target, and to optimize pharmaceutical properties such as chemical and metabolic stability, solubility, and oral absorption.

When dealing with chemical structures, the characteristics of the drug are defined by its functional groups. Functional groups are specific arrangements of atoms that give the overall molecule unique chemical reactivity, physical and electronic properties. A medicinal chemist must be able to recognize which functional groups are important and which ones can be modified to increase potency and selectivity. The most common way of achieving this goal is through iterative syntheses. It is at this stage of drug development that several hundred compounds might be synthesized, generally by a team of medicinal chemists, working together with biochemists to evaluate the different analogues.

Structure-activity relationship studies (SAR) are a series of iterative synthesis experiments conducted to determine which chemical modifications lead to the greatest improvements in potency and selectivity against your target protein or enzyme. Some SAR studies were used to direct the synthesis of the first generation irreversible inhibitors detailed in chapter 3 of this thesis. The lead inhibitors were developed from earlier work done by Dr. Nag Kumar and his fluorescent peptide substrates developed within our lab. These initial peptide substrates revealed a key pharmacophore that was integrated into both the fluorescent substrates and irreversible inhibitors to provide enhanced activity against ATG4B. The details of which will be discussed in the following chapter.

The need for an irreversible inhibitor

Having a small molecule irreversible inhibitor of ATG4B available can be a first step in investigating the role of ATG4B in autophagy. If it has acceptable selectivity and pharmaceutical properties it may also reveal the role of autophagy in health and disease by being able to turn autophagy off and on in living cells. Irreversible inhibitors also have the added property of covalently modifying your target of interest. This can be observed by using mass spectrometry when measuring the molecular mass of your target before and after modification. This method of analysis is detailed in the next chapter and was used to validate some of our predictions about inhibitor reactivity. The use of irreversible inhibitors can lead to an understanding of where exactly your inhibitor is binding to your target. Another common method to confirm sites of binding is by the use of *photo-affinity probes*.

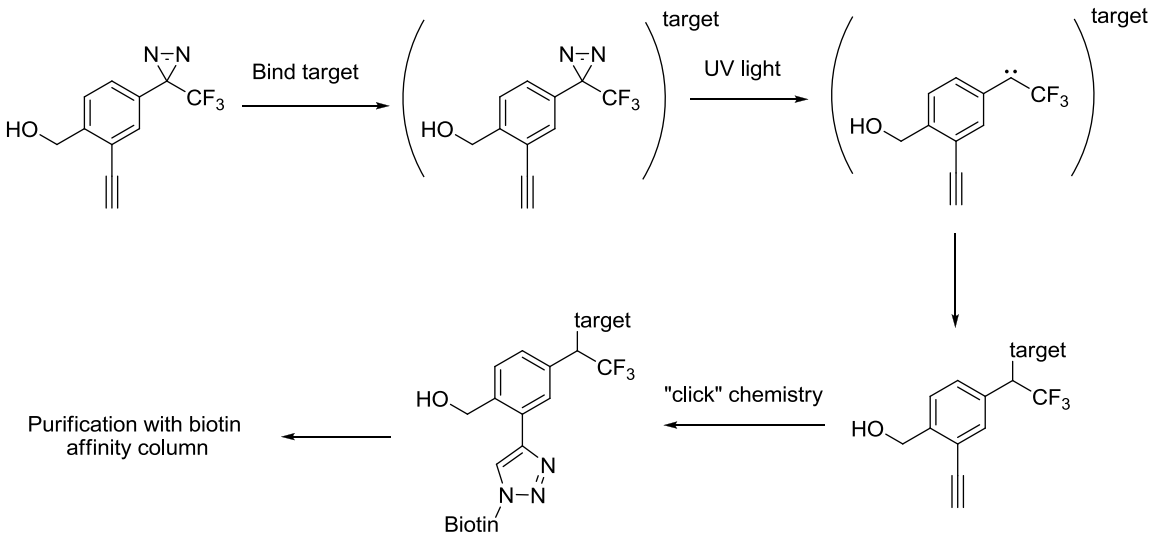


Figure 2: An example of a photo-affinity probe binding to a protein target of interest. UV irradiation promotes the formation of a reactive carbene which can then bind covalently to the target. The bound label can then be purified with biotin affinity columns.

The photo-affinity reactivity in the figure is produced by the addition of a photoreactive functionality, such as a diazirine ring, to the probe molecule at a site that does not impair binding activity. After the probe binds reversibly to its target, UV irradiation causes the diaziridine ring to lose N_2 promoting the formation of a reactive carbene which can covalently insert into various bonds proximal to the carbene at the time of generation.

The protein or peptide fragment with the bound label can be subsequently purified and analyzed for the site of attachment. This is just an example to illustrate how irreversible inhibitors can be utilized to determine where exactly they bind to their target.

It is more common and generally preferable for drugs to be reversible inhibitors as opposed to irreversible. Reversible inhibitors tend to have less side-effects, such as less build up of inactivated enzymes or off-target activities which may lead to toxicity.[42] However, irreversible inhibitors are still important tools that can be used quite effectively to observe the effects of inhibition of the target and if the irreversible inhibitor alkylates the target enzyme, protein analysis and degradation can be used to determine where exactly they bind to their target. They can also act as starting points for elaboration into reversible analogous compounds. Historically, halomethyl ketones have been one of the first compounds described to inhibit both serine and cysteine proteases.[43,44] It has also been shown that dipeptidyl chloromethyl ketones react with a wide variety of serine proteases whereas fluoromethyl ketones have been described to be more specific towards cysteine proteases.[45] This design and characterization of irreversible inhibitors of ATG4B is explored and validated in chapter 4 of this thesis.

While potency of inhibitors of enzymes can be optimized *via in vitro* enzyme assays, it is important to find compounds that inhibit the target in cells and it is desirable to have a direct measure of such inhibition within the cell. A small molecule irreversible inhibitor of ATG4B that works in cells or *in vivo* could be the basis of a type of cell assay, known as an enzyme-occupancy cell assay. Provided that the binding of the inhibitor to the protein target can be readily measured (such as by having the inhibitor radiolabeled or if it is fluorescent), incubation with cells would then allow for the enzyme (ATG4B in our case) to become labelled in cells. Competitive inhibitors could be screened for inhibition of binding of the probe inhibitor and in the case of fluorescence one could have a rapid read-out by immunoprecipitation of the labeled protein and quantification of label on a gel. Any competitive inhibitors could then be detected by the dose-dependent disappearance of fluorescently labelled ATG4B. This would not only provide information on competition, but could also directly link the phenotype of autophagy inhibition in cells with inhibition of ATG4B. Efforts towards such a compound are presented in chapter 4.

From clinic to market

The final stages of drug development may be the hardest hurdles to overcome. Many drugs make it to clinical trials only to fail due to unforeseen side-effects or toxicity. Pre-clinical tests often require years to complete and includes tests for toxicity on animal models (usually inbred strains), ADME (absorption, distribution, metabolism, excretion) optimization and a host of other tests in an attempt to ensure the safety of the drug before it ever reaches human trials. Even with the recent advances in modern technology, it is still almost impossible to fully predict human toxicity based on animal models. There are indeed cases where certain drugs are safe in the animal testing (generally in inbred rodents or dogs), but this safety does not translate to humans who are outbred and diverse.

Thiol protease inhibitors

The idea of targeting a cysteine protease, such as ATG4B, as a potential therapeutic target is not a new one. In the past decade, more than a dozen compounds have made it through various phases of clinical trials targeting various cysteine proteases.[46] There are a variety of cysteine proteases that are classified into clans and families, but generally they are categorized into 3 structurally distinct groups; papain-like (clan CA, the cathepsins), interleukin beta-converting enzyme (ICE-like, clan CD, the caspases), and picornain-like (clan PA(C), human rhinovirus 3C).

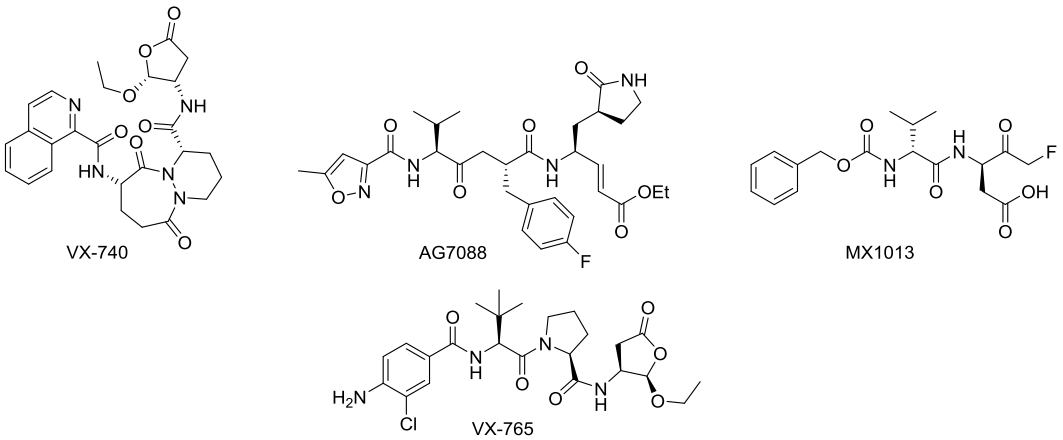


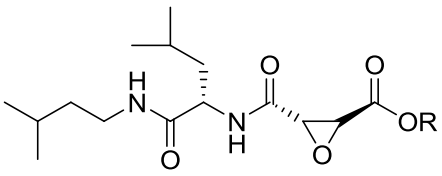
Figure 3: Structures of VX-740 (Pralnacasan), VX-765, AG7088 (Ruprintrivir) and MX1013 (Z-Val-Asp-CH_2F).

Pralnacasan (VX-740) is a masked aldehyde that was orally administered for the treatment of rheumatoid arthritis and osteoarthritis by inhibiting ICE or caspase-1. Unfortunately the phase IIb clinical trials were voluntarily discontinued due to toxicity effects seen in 9-month animal studies. Second generation caspase inhibitors VX-765 and VX-799 have been discontinued due to septic shock.[47]

Ruprintrivir (AG7088) is a Michael acceptor that targets rhinovirus 3C caspase and is currently in development for the treatment of the common cold *via* intranasal delivery. Ruprintrivir has advanced to phase II/III clinical trials, but failed to meet the FDA's approval.[48]

MX1013 is a fluoromethyl ketone in the early stages of development for the treatment of myocardial infarction. MX1013 inhibits caspases 1, 3, 6, 7, 8, and 9 with *in vivo* IC_{50}'s ranging from 5 to 20 nM.[49]

Epoxysuccinyl peptide derivatives have attracted considerable interest among the many natural cysteine protease inhibitors that have been isolated from microorganisms. They displayed little inhibitory activity against aspartyl, metallo and serine proteases while strongly inhibiting cathepsins B and L irreversibly.[50,51]



R = H; E-64c
R = Et; E-64d

Figure 4: Structures of E-64c and E-64d (Loxistatin).

Loxistatin (E-64d) is a pro-drug for E-64c and was developed for the treatment of muscular dystrophy. Development was stopped in phase III clinical trials due to adverse effects observed in rats ranging from hepatic injury to teratogenic effects on rat embryogenesis.[52,53]

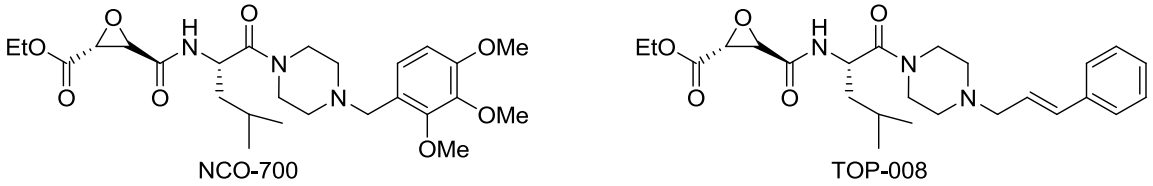


Figure 5: Structures of piperazine epoxysuccinate derivatives NCO-700 and TOP-008.

NCO-700 was initially developed as a calpain and cathepsin B inhibitor for the treatment of cardiovascular diseases such as myocardial infarction.[54] However, it was reported that NCO-700 and its derivative TOP-008 were effective anti-cancer agents when tested *in vitro* and *in vivo* against human breast and prostate tumours. NCO-700 is currently in phase 2 clinical trials as a new oncolytic drug, but the mechanism of action may not be linked to cysteine protease inhibition.[55]

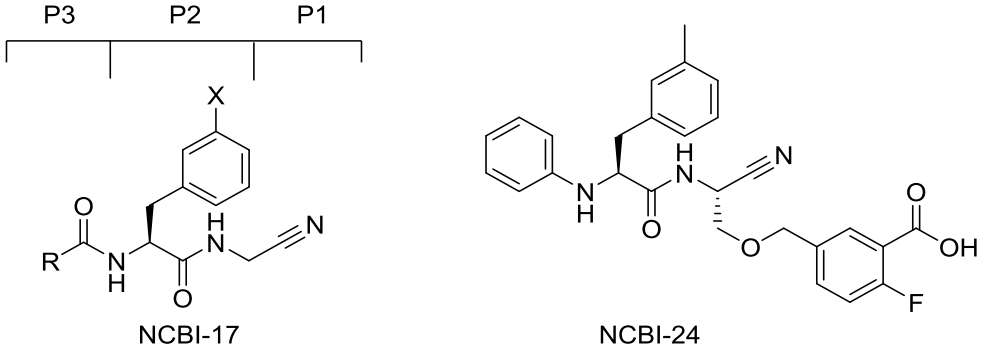
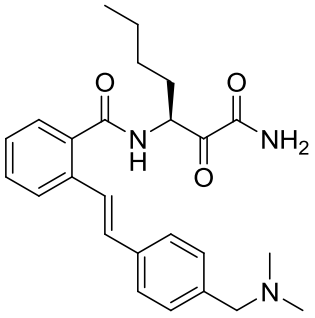


Figure 6: Structures of lead template NCBI-17 and compound NCBI-24.

Dipeptidyl nitrile compounds have recently been disclosed as cathepsin B inhibitors.[56] NCBI-17 was initially identified with an IC_{50} of 62 µM and the P2 and P3 positions were subsequently optimized using data from molecular modelling of the enzyme/inhibitor complex as well as from high-resolution X-ray crystallography. This eventually led to a new compound, NCBI-24 which displayed excellent oral bioavailability.

Another type of cysteine protease, calcium activated papain-like cysteine protease (calpain), requires Ca^{2+} for activation. Initially discovered approximately 40 years ago, the number of known calpains has increased to approximately 16, which are broadly classified and tissue-specific (calpain 1-16).[57] A series of ketoamides were developed targeting calpain-1 in order to provide a better understanding of the otherwise unknown physiological roles of the various calpain enzymes.[58–60] One compound in particular (A-705239) is being developed for the treatment of traumatic brain injury.[58]



A-705239

Figure 7: Structure of ketoamide inhibitor A-705239.

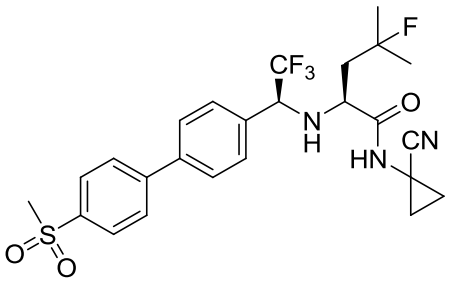


Figure 8: Structure of cathepsin K inhibitor Odanacatib

Odanacatib has recently finished phase III clinical trials for the treatment of osteoporosis. It is being developed by Merck as an inhibitor of cathepsin K, an enzyme that is implicated in the bone resorption process.[61] If approved, it would make Odanacatib the first cysteine protease drug to reach the market as a cathepsin K inhibitor for the treatment of osteoporosis after nearly 20 years in development.

In our case, it was decided that the first set of inhibitors to be investigated would be the substituted methyl ketone type. The reasons will be highlighted within chapter 3,

but a large determining factor comes from the fact that ATG4B possesses an auto-regulatory loop that acts to block the very constrained active site. It is only when a substrate binds ATG4B that conformational changes occur to lift the loop and reveal the active site of the enzyme. Even then the site is very small. For this reason, we proposed that substituted methyl ketones and possibly cyanomethylene amides would be small enough to fit within the active site of ATG4B. The body of work represented in this thesis demonstrates that small molecule inhibitors of ATG4B are possible and indeed act to block the active site of ATG4B. However, before any testing of compounds could begin, a suitable assay for activity of ATG4B needed to be developed and validated.

Chapter 2. Assay Development

2.1. Research Objective

The goal for this part of my research was to engineer a novel fluorescent protein substrate for ATG4B, which would enable high throughput screening (HTS) and ranking of inhibitors using a simple fluorescent readout. Initially there were no published assays for ATG4B suitable for screening large libraries of compounds, but subsequently a high throughput assay based on a construct of LC3 and phospholipase A2 was described.[33] More recently, a fluorescence resonance energy transfer (FRET)-based assay has been described, but with no application to high-throughput screening (HTS).[32] The work in this chapter highlights the development of two useful assays for studying ATG4B activity with the optimization of one of the assays for HTS. HTS subsequently has identified a number of interesting active compounds and chemical motifs that can be incorporated into the next generation of inhibitors described in chapters 3 and 4 of this thesis. This chapter is in large a representation of the work already published in Assay and Drug Development Technologies, April 2014, volume 12, issue number 3, page 176-189.

2.2. Results

Mass Spectrometry Assay to Measure ATG4B Activity

At the initiation of our project, in the absence of any other viable assay to monitor ATG4B activity, we developed an assay directly measuring cleavage of the proLC3B substrate by mass spectrometry. Recombinant human His_6-ATG4B and His_6-proLC3B (proLC3B) were used (proteins and cDNA were kindly supplied by Aled Edwards, SGC, Toronto), and ATG4B was shown to effectively convert proLC3B (expected mass: 16851 Da, observed mass: 16720 Da) with an observed molecular mass of 16,720.3 Da (Figure 9), to LC3-I (expected mass: 16161 Da, observed mass: 16161 Da) with a mass of

16,161.4 Da, in keeping with the predicted loss of 558.9 Da (Figure 10). The kinetics of cleavage of proLC3B was then examined in detail by mass spectrometry. For the assay, varying concentrations of either ATG4B or proLC3B were mixed, and both the disappearance of the nominal mass for proLC3B and the appearance of the expected mass of the processed protein LC3-I were monitored. Change of peak heights of the respective peaks were used to calculate rate of cleavage according to the formula: Percentage of product formation = A/(A + B) x 100, where A and B are the peak heights of LC3-I and proLC3B respectively. Concentrations were adjusted to allow the reaction to proceed over about a 1 hour period and reaction mixtures were sampled every 3–5 minutes to allow duplicate runs to be followed in parallel. The ATG4B was activated by pre-incubation for 1 hour with a suitable reducing agent (TCEP). As expected, ATG4B converted proLC3B to LC3B-I in a concentration and time-dependent manner. Reactions were all run at pH 8.0, which has been previously reported to be optimal[32] and which we also confirmed (studies by a colleague Dr. Steve Arns). Standard curves for proLC3B and LC3-I were developed using the purified proLC3B and samples of known concentrations of proLC3B that had been fully cleaved to yield LC3B-I by incubation with ATG4B at 25^{o}C over 12 hours. Using the mass spectrometric method for analysis of the enzyme kinetics, the K_M for His_6-LC3B on our His_6-ATG4B was determined as 12 ± 4.6 µM, k_{cat} as 0.98 ± 0.82 s^{-1}, and k_{cat}/K_M as 81000 s^{-1} M^{-1} (Figure 11) at an enzyme concentration of 48 nM.

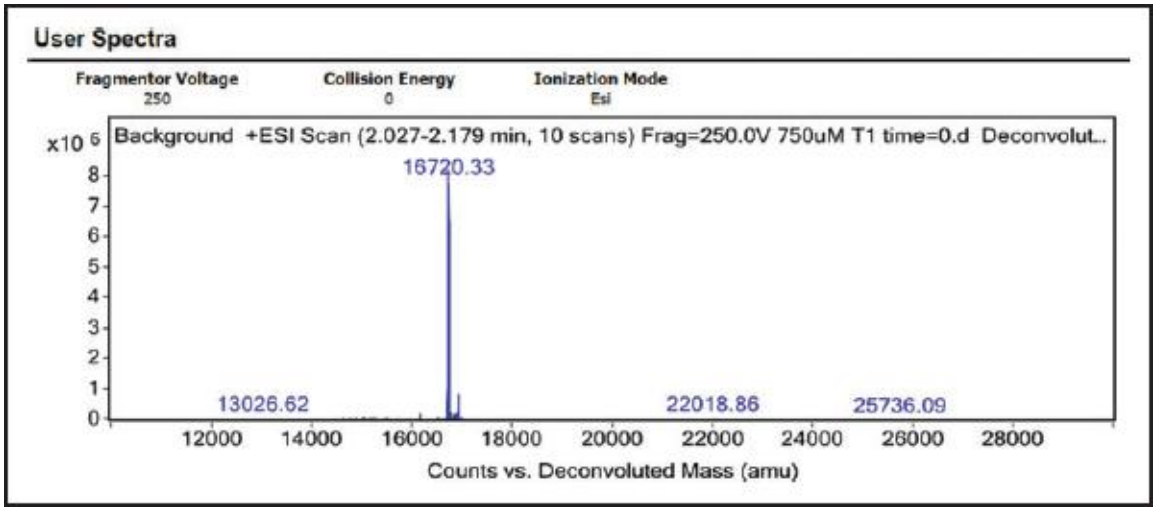


Figure 9: Deconvoluted mass spectrum of proLC3B.

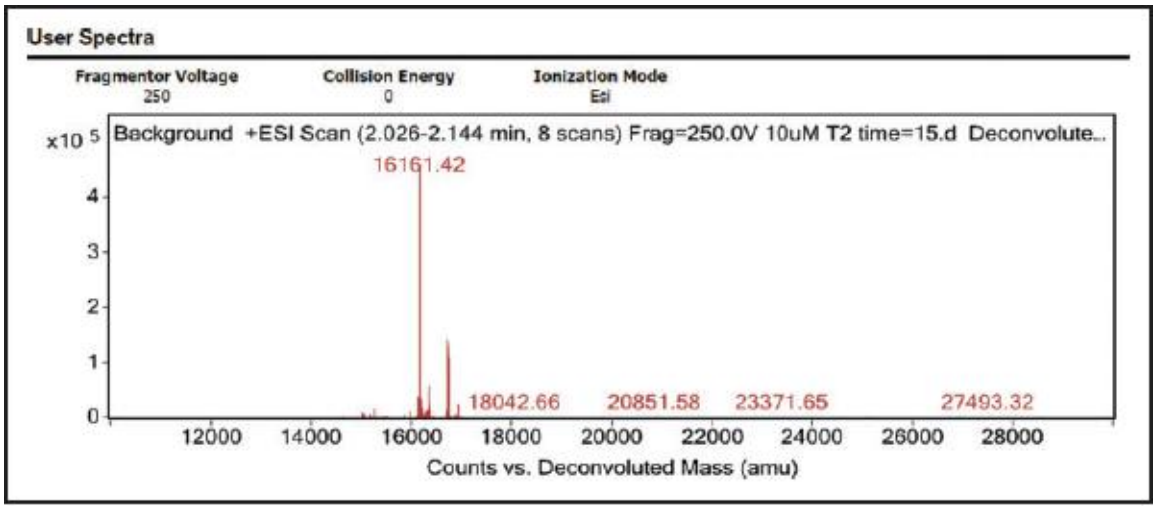


Figure 10: Deconvoluted mass spectrum of LC3-I after incubation with ATG4B. A loss of 559 Da is observed which matches the expected mass loss due to cleavage by ATG4B.

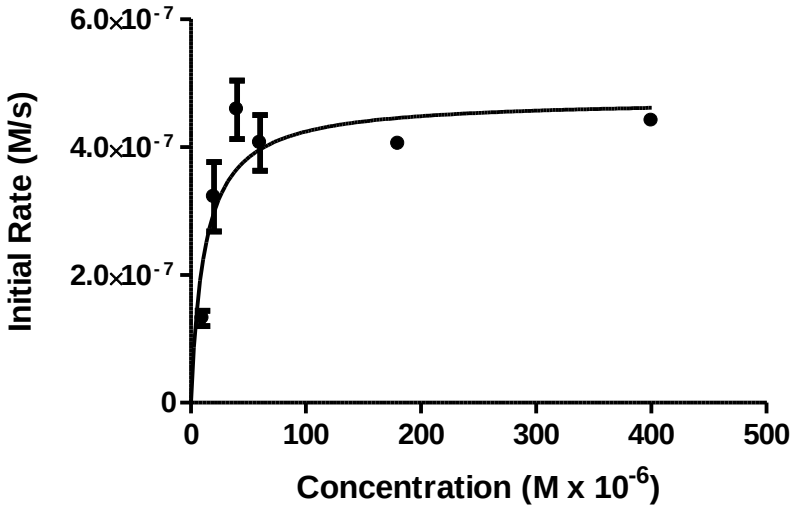


Figure 11: Michaelis-Menten plot for LC3B cleavage by 48 nM ATG4B *via* LC/MS. K_M = 12 ± 4.6 µM, k_{cat} = 0.98 ± 0.82 s^{-1}, and k_{cat}/K_M = 81,000 s^{-1} M^{-1}. Non-linear regression analysis performed by GraphPad 5 software fitting to the equation Y=Vmax*X/(Km+X). Error bars represent standard deviations from triplicate experiments.

This LC-MS assay was useful to characterize inhibitors and was not prone to interference, but was very demanding in time and protein quantities and only one compound could be analyzed per day, making it unsuitable for HTS.

The amino acid sequence of LC3B can be seen below:

MGSSHHHHHHSSGLVPRGSHMPSEKTFKQRRTFEQRVEDVRLIREQHPTKIPVIIERYK
GEKQLPVLDKTKFLVPDHVNMSELIKIIRRRLQLNANQAFFLLVNGHSMVSVSTPISEVY
ESEKDEDGFLYMVYASQETFGMKLSV

LC3B has a predicted mass of 16851 Da with an observed mass of 16720 Da which corresponds to a loss of an *N*-terminal methionine.

The amino acid sequence of ATG4B can be seen below:

MGSSHHHHHHSSGLVPRGSMDAATLTYDTLRFAEFEDFPETSEPVWILGRKYSIFTEK
DEILSDVASRLWFTYRKNFPAIGGTGPTSDTGWGCMLRCGQMIFAQALVCRHLGRDW
RWTQRKRQPDSYFSVLNAFIDRKDSYYSIHQIAQMGVGEGKSIGQWYGPNTVAQVLKK
LAVFDTWSSLAVHIAMDNTVVMEEIRRLCRTSVPCAGATAFPADSDRHCNGFPAGAEV
TNRPSPWRPLVLLIPLRLGLTDINEAYVETLKHCFMMPQSLGVIGGKPNSAHYFIGYVGE
ELIYLDPHTTQPAVEPTDGCFIPDESFHCQHPPCRMSIAELDPSIAVGFFCKTEDDFND
WCQQVKKLSLLGGALPMFELVEQQPSHLACPDVLNLSLDSSDVERLERFFDS

ATG4B has a predicted mass of 45144 Da with an observed mass of 45013 Da which corresponds to a loss of an *N*-terminal methionine.

Development of a Fluorogenic Substrate and Assay for ATG4B

An assay using active-site dependent cleavage of a small molecule fluorogenic substrate was preferred for higher throughput screening and we sought to develop such a substrate by investigating the structure–activity relationships for the enzyme on fluorogenic peptides. An enzyme binding site can be broken down into 2 different categories. S-sites denote binding sites on the enzyme and P-sites denote the corresponding substrate site that binds with the enzyme site. S and P sites are numbered corresponding to their position and proximity to the cleavage site. For example, if an enzyme cleaves its substrate's C-terminal tail beside a glycine residue,

the glycine residue is referred to as in position P1, binding to the S1 site of the enzyme. The upstream residues are denoted P2 and P3 respectively. Downstream of the cleavage sites are denoted as P1` and P2` etc. Studies have indicated that ATG4B requires a glycine at P1 and recognizes the motif phenylalanine-glycine at P2 and P1,[38] but recombinant ATG4B was initially reported to be unable to cleave small peptide fragments (such as QETFGTALA) spanning the murine LC3B cleavage site.[29] Recent reports have subsequently indicated the enzyme is capable of cleaving some small peptide-amino-trifluoromethylcoumarin (AFC) substrates such as Ac-GTFG-AFC, albeit with very low rates of turnover.[31] Other studies have shown that fluorescent self-quenching peptides such as Mca-TFGM-Dpa are cleaved by the homologous ATG4C,[1] but in our hands, recombinant ATG4B did not cleave (Mca)-TFGM-(Dpa), (Mca)-QETFGMKLS-(Dpa), or (Mca)-QETFGTALA-(Dpa) at any measureable rate. Of the many fluorogenic peptides synthesized by Dr. Nag Kumar, only a select few turned out to be moderate substrates for ATG4B. Only when the peptides contained an N-terminal acid moiety (such as N-hemisuccinyl-FG-AMC, k_{cat}/K_M = 0.18 M^{-1} s^{-1}) did they become significant substrates.

Development of FRET-LC3 substrate

Unfortunately, these fluorogenic peptide substrates required relatively high concentrations of enzyme (5 µM) to show measureable conversion. Such a high concentration would not be viable for scaling up to HTS. To develop an assay suitable for screening at lower enzyme concentrations, we devised a fluorescent analog of proLC3B. We conceptualized that an LC3B analog with appropriately chosen fluorescent proteins appended at the N and C termini might display fluorescence resonance energy transfer (FRET), which would be lost after cleavage, thus allowing us to monitor the progression of the reaction by changes in fluorescence. FRET is a phenomenon that occurs when there are two fluorescent proteins in close proximity with one another and when the excitation wavelength of one fluorescent protein overlaps with the emission wavelength of the neighboring fluorescent protein. The efficiency of FRET interactions has been investigated and it was revealed that distances greater than 10 nm would result in very poor FRET interactions.[62] A published X-ray structure of LC3B[29] indicated that the C and N termini are within about 20 Å of each other, a distance that should be consistent with a robust FRET interaction.[62] Also, both the His_6-proLC3B

construct and the LC3B-GST fusion protein[29] are known substrates for ATG4B, indicating the enzyme is largely indifferent to bulky protein elongations at either the N or C termini. Evaluation of the absorption-emission spectra, stability and availability of suitable commercial genetic constructs of possible candidate fluorescent proteins led us to choose of emerald-green fluorescent protein (Em-GFP) and yellow fluorescent protein (YFP). Em-GFP is known to be quite photo stable and absorbs with maximum at 487 nM and emits at 509 nM, while YFP absorbs at 514 nM and emits at 527 nM.[63]

The construct His_6-YFPLC3-EmGFP (FRET-LC3; Figure 12) was prepared in collaboration with Dr. Matthew Macauley and David Shen of Dr. David Vocadlo's lab at Simon Fraser University. Standard molecular biology and biochemistry techniques were used under the supervision and guidance of Dr. Macauley to prepare the FRET-LC3 substrate. Commercially available plasmids containing the DNA sequence for YFP (yellow fluorescent protein) and proLC3B, containing BamH1/Sal1 and Sal1/EcoR1 restriction sites respectively, were amplified by standard PCR (polymerase chain reaction) techniques. The LC3 DNA was obtained as a GFP-LC3 construct and the GFP part was excised to give the desired LC3 sequence. These fragments were then ligated into an expression vector containing EmGFP using the BamH1 and EcoR1 restriction sites. The final DNA sequence of the resulting expression vector can be seen in the figure below.

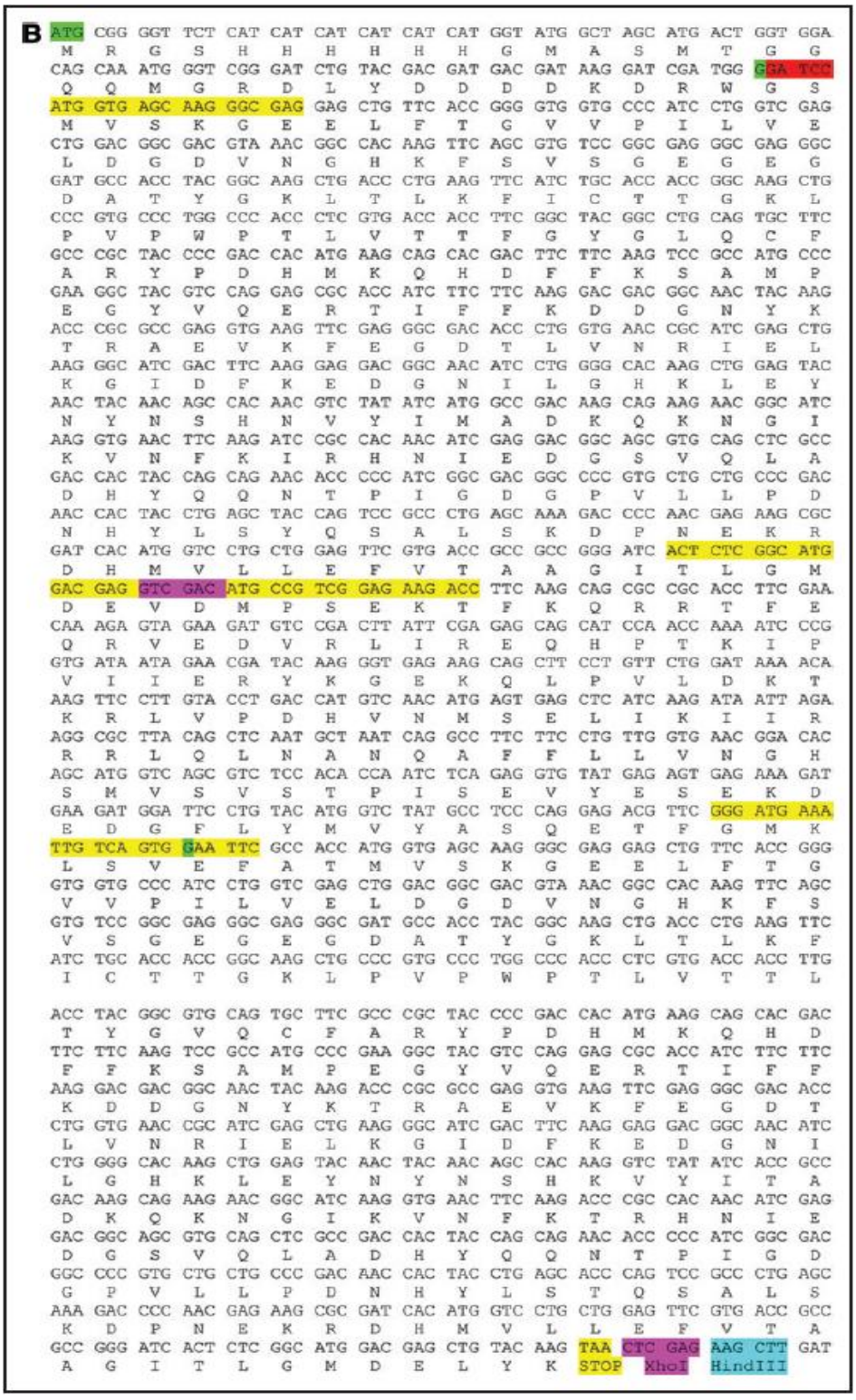


Figure 12a: Polynucleotide and amino acid sequence of the expression vector for FRET-LC3. Highlight green indicates start codon. Highlight red indicates BamH1

restriction site. Highlight yellow indicates original primer locations. Highlight purple indicates Sal1 restriction site. The stop codon is indicated at the end with Xho1 and HindIII restriction sites.

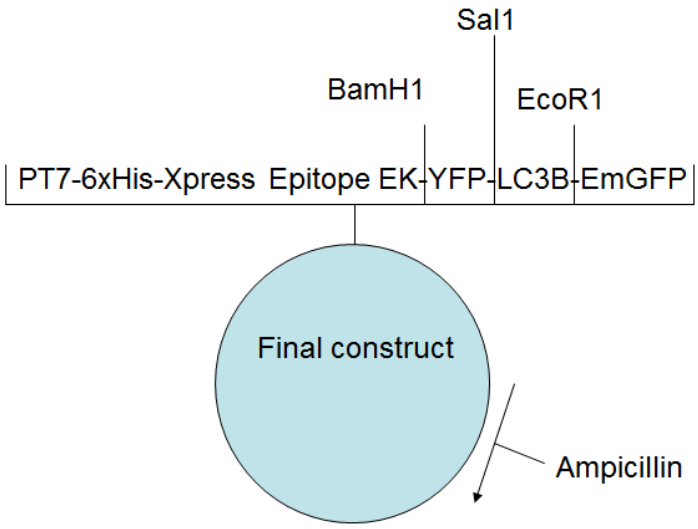


Figure 12b: Illustration of the plasmid obtained after ligating YFP and LC3B fragments into the EmGFP expression vector. Ampicillin resistance gene is also incorporated for bacterial expression.

The plasmid was transformed into a culture of *E.coli* and grown to an optical density of 600 before induction with IPTG (Isopropyl β-D-1-thiogalactopyranoside) for 4 hours at 25 ^{o}C. Cellular contents were lysed and purified by histidine trap columns. Concentration was determined by nanodrop analysis. The purified protein was essentially a single band running at an apparent molecular weight of 72 kDA on a denaturing gel, the UV spectrum of the purified protein had absorption maxima at 485 and 515 nm (Figure 13) and the protein showed an observed molecular mass of 72,902 Da corresponding to the expected molecular mass (Figure 14). When FRET-LC3 was incubated with ATG4B and the reaction followed by mass spectrometry, it was found to be rapidly and completely cleaved to yield the expected fragments of molecular masses 27,885 Da (for the cleaved EmGFP) and 45,036 Da (for the YFP-LC3 fragment; Figure 15). Observation of the fluorescence emission spectrum of FRET-LC3 showed that excitation at 485 nm led to emission with maximum at 527 nm, indicative of the expected FRET response. FRET-

LC3 was then incubated with ATG4B, and the fluorescence changes were monitored (excitation at 485 nm). The reaction was also monitored in parallel experiments by mass spectrometry. Emission at 510 nm exhibited the largest relative increase (Figure 16), but changes over time were not linear with concentration of product formed. The absorption spectrum for EmGFP includes a shoulder with significant absorption at about 490 nm,[64] and excitation near this wavelength could be expected to lead to significant, non-FRET excitation of GFP that could complicate the fluorescence changes. Thus, in initial studies with FRET-LC3, we decided to excite at a wavelength lower than 485 nm to avoid overlap with this absorption shoulder. It was found that monitoring cleavage of FRET-LC3 by ATG4B with excitation at 450 nm and observation of fluorescence changes at 510 nm gave increases in fluorescence that were linear and correlated with respect to concentration of the components of the reaction (Figure 16). Equally, correlation of concentration of FRET-LC3 and its fragments with mass spectrometric peak heights was also linear (Figure 17). Next, kinetic analyses of activity of ATG4B using FRET-LC3 substrate were performed comparing the fluorescence and mass spectrometric readout. Following processing of FRET-LC3 by ATG4B with the fluorescence readout (excitation 450 nM; emission 510 nM) gave K_M = 16 ± 13 μM (with 48 nM enzyme) and k_{cat} = 0.52 ± 0.23 s^{-1} and k_{cat}/K_M of 33,000 s^{-1} M^{-1} (Figure 18, left). Parallel assays using mass spectrometric analysis gave K_M = 9 ± 3 μM, k_{cat} = 0.25 ± 0.04 s^{-1}, and k_{cat}/K_M = 27,000 s^{-1} M^{-1} (with 48 nM enzyme; Figure 18, right). Thus, the FRET assay of ATG4B activity using YFP-LC3-EmGFP as substrate was validated and considered suitable for further optimization to support HTS.

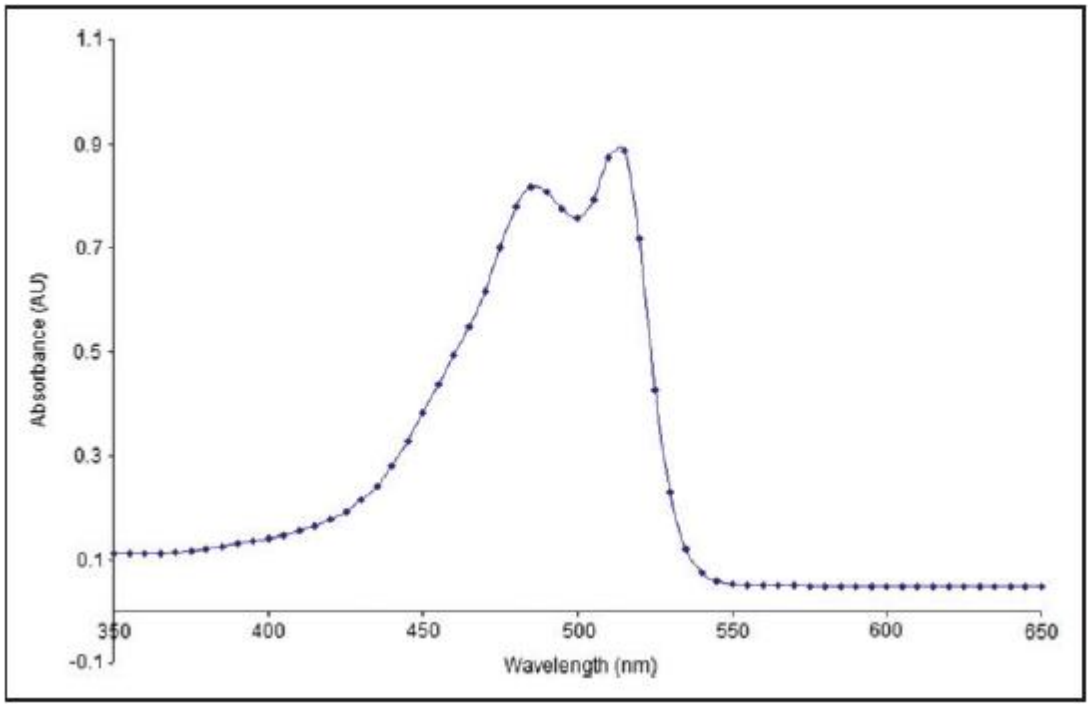


Figure 13: Absorbance spectrum of YFP-LC3B-EmGFP. A 1 mL sample of YFP-LC3B-EmGFP at 1 µM concentration was measured using a BioTek Synergy4 fluorometer in a 1 mL BioCell cuvette. $^{EmGFP}\lambda_{max}$ = 485 nm, $^{YFP}\lambda_{max}$ = 512 nm.

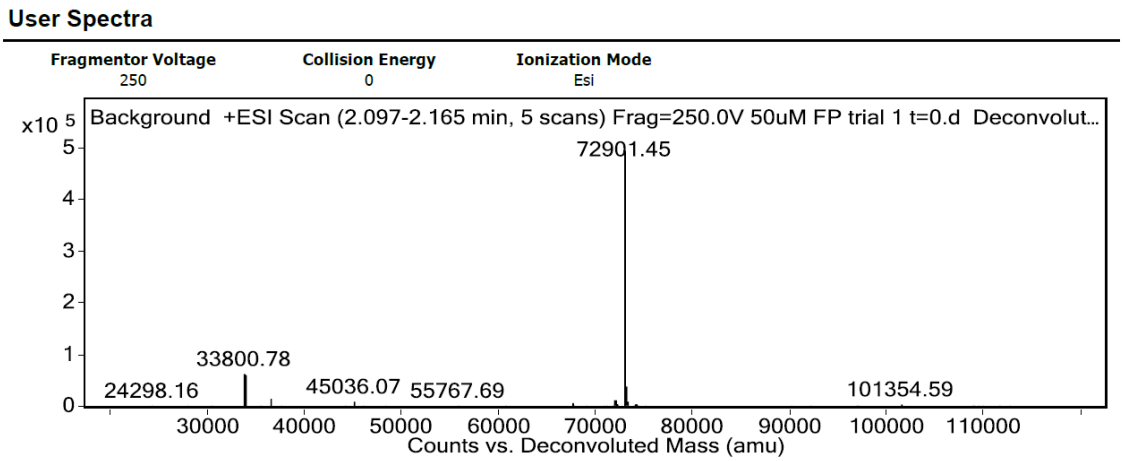


Figure 14: Deconvoluted mass spectrum of YFP-LC3B-EmGFP

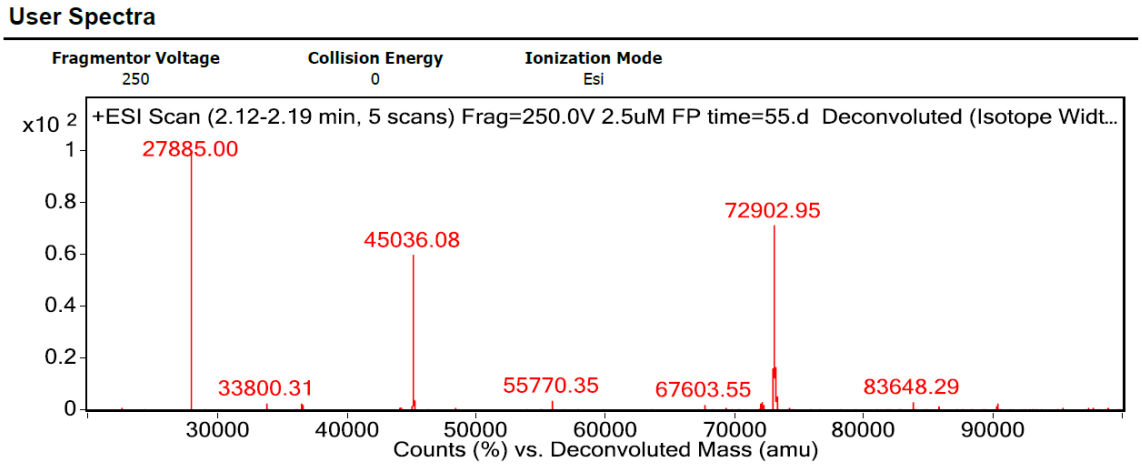


Figure 15: Deconvoluted mass spectrum of YFP-LC3B-EmGFP after incubation with ATG4B. Complete cleavage is represented when the peak at 72902.95 is no longer observable.

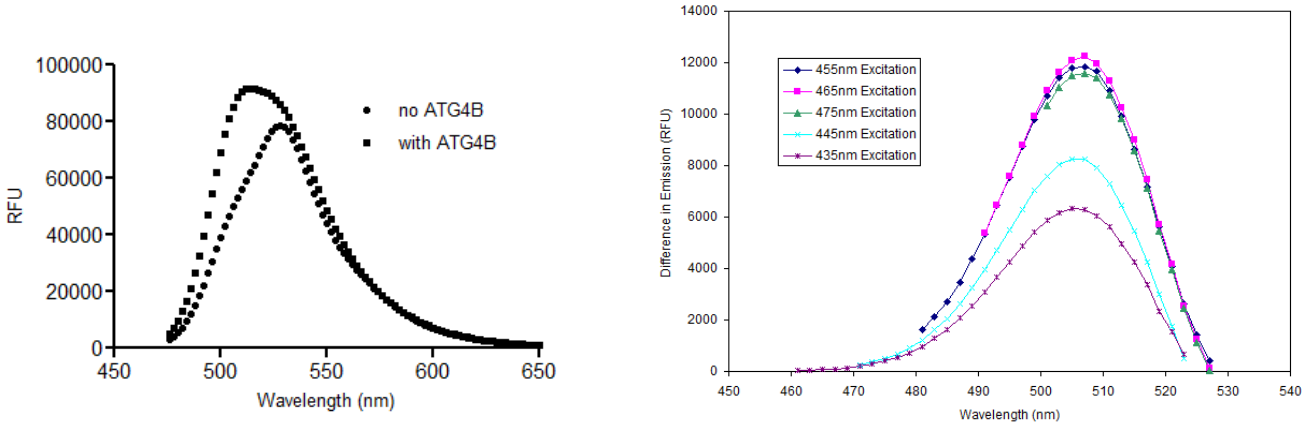


Figure 16: Fluorescence emission spectrum of YFP-LC3B-EmGFP using an excitation wavelength of 485 nm before and after treatment with ATG4B (left). The difference in RFU is plotted in the graph on the right. It can be seen that the largest changes in fluorescence emission occur around 508 nm.

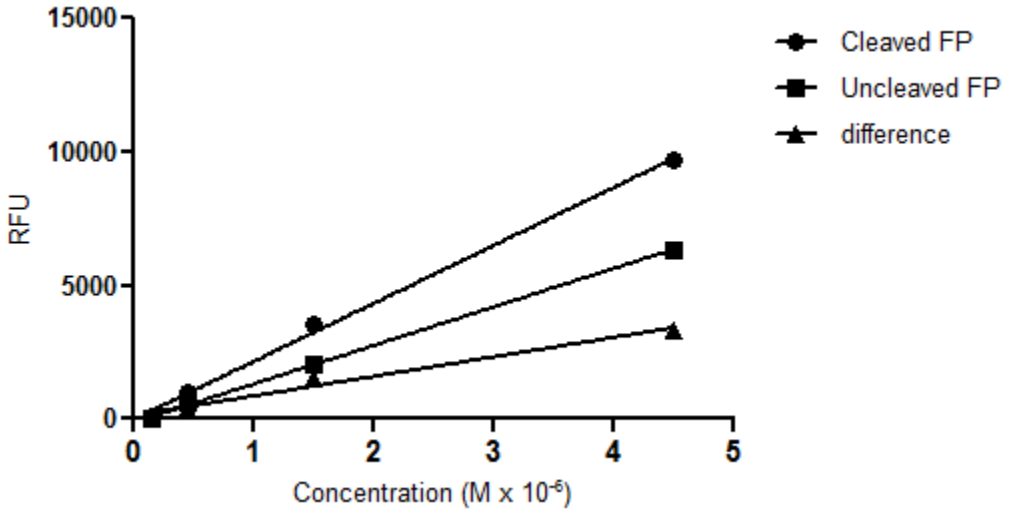


Figure 17a: A standard fluorescent curve of YFP-LC3B-EmGFP fluorescent protein. 450ex/510em fluorescence signal was measured at various concentrations of both cleaved and uncleaved fluorescent protein substrate. Cleaved protein was obtained by incubation with 12 nM ATG4B overnight at 4^{o}C. Complete hydrolysis of the substrate was confirmed *via* LC/MS analysis. A plot of concentration versus RFU (relative fluorescence units) reveals a linear correlation between concentration and RFU. The slope of the difference between cleaved and uncleaved protein samples was used as the correction factor. Cleaved FP slope = 2100 ± 77 RFU/µM (R^2 = 0.9975), uncleaved FP slope = 1440 ± 20 RFU/µM (R^2 = 0.9996), difference slope = 700 ± 70 RFU/µM (R^2 = 0.9819).

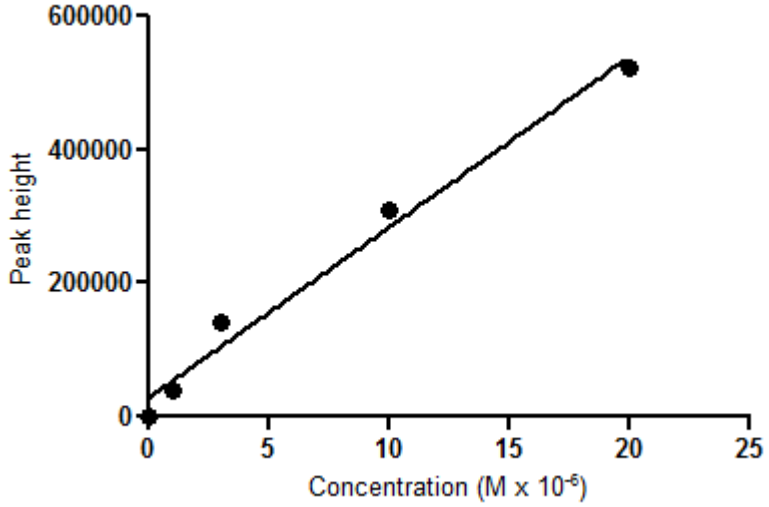


Figure 17b: A standard response curve for YFP-LC3B-EmGFP *via* LC/MS. Various concentrations of YFP-LC3B-EmGFP were analyzed by Agilent Time-of-Flight LC/MS mass spectrometer equipped with an ESI ion source. Slope = 25000 ± 2000 peak height/μM (R^2 = 0.9821).

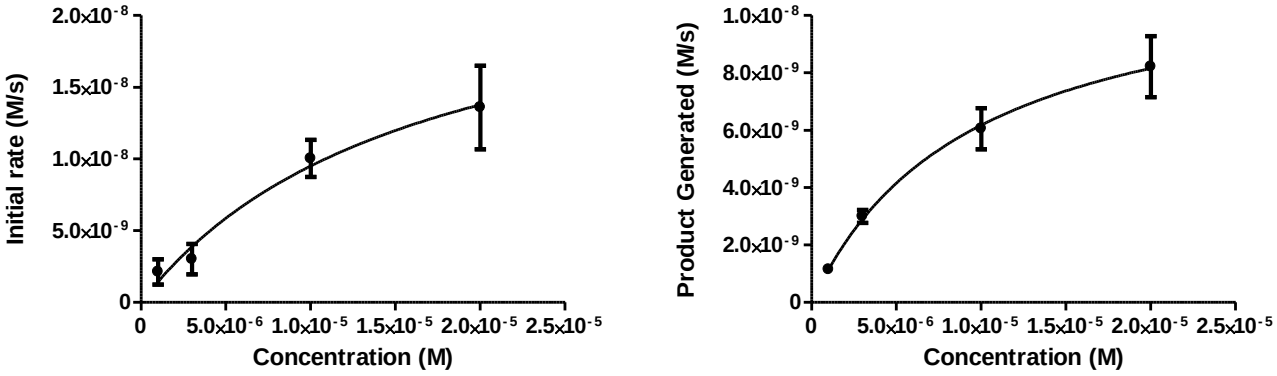


Figure 18: Michaelis-Menten plot of YFP-LC3B-EmGFP cleavage monitored by fluorescence (450ex/510em, left) and LC/MS (right). K_M = 16 ± 13 (with 48 nM enzyme) and k_{cat} = 0.52 ± 0.23 s^{-1} and k_{cat}/K_M of 33,000 s^{-1} M^{-1} *via* fluorescence. K_M = 9 ± 3 μM, k_{cat} = 0.25 ± 0.04 s^{-1}, and k_{cat}/K_M = 27,000 s^{-1} M^{-1} (with 48 nM enzyme) *via* LC/MS.

Alkylation of ATG4B by NEM

Pretreatment of ATG4B with high concentrations (20,000-fold excess) of the nonspecific thiol trapping agent *N*-ethylmaleimide (NEM) for 5 hours led to complete

inactivation of the enzyme. Mass spectrometric evaluation of the mixture revealed that the reagent reacted (putatively) with multiple thiol nucleophiles on the enzyme, giving modified ATG4B protein with mass equivalent to the addition of 13 molecules of NEM, one for each of the 13 cysteine residues present in the protein. Pre-treatment of ATG4B for 24 hours with a lower, 2,000-fold, excess of NEM led to addition of 4 Molar equivalents of NEM and this modification had little effect on the rate of the enzyme.

Time Dependence of Inhibition of ATG4B with NEM

Pre-treatment of ATG4B (12 nM) with 1mM NEM (a 83,000-fold molar excess) for 50 minutes led to complete inhibition of activity, while pre-treatment under the same conditions but for only 15 min gave partial inhibition (ca. 50%). When the substrate proLC3B (30 µM) was added after 15 minutes, the residual rate of activity was linear over the next 45 min of observation (Figure 19).

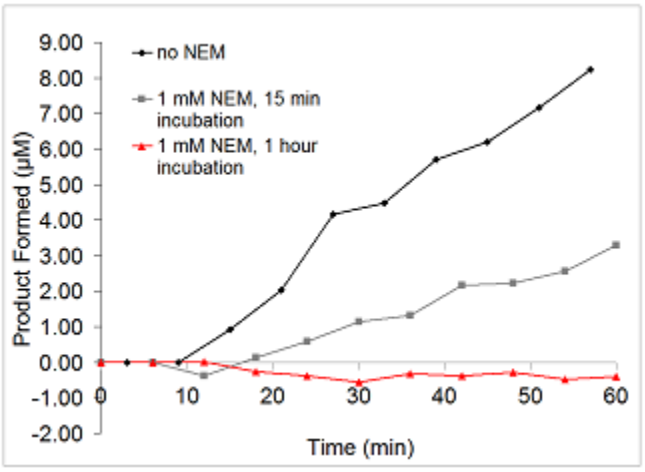


Figure 19: NEM time-dependent inhibition at 15 min and one hour.

2.3. Discussion

Assay of Enzymatic Activity of ATG4B Using Mass Spectrometry

Recombinantly expressed proLC3B gave a strong mass spectrometric signal that after deconvolution corresponds to a molecular mass of 16,720.3 Da, as predicted from the calculated molecular weight. Upon processing of proLC3B with recombinant ATG4B, a mass of 16,161.4 Da was observed, corresponding to LC3-I with the predicted

loss of 558.9 Da relative to proLC3B. The parent mass signal intensities of proLC3B and LC3-I were linear with respect to concentration, and thus useful to follow enzymatic cleavage over time with an optimized sampling time of one measure every 5 min. Kinetic analysis of the reaction gave a K_M of 12 μM and k_{cat}/K_M of 81,000 s^{-1} M^{-1} consistent with parameters determined by other modes of assay.[31]

Development of a FRET Assay for Screening for Inhibitors of ATG4B

While the enzyme activity assay based on mass spectrometric analysis is useful, it is a very low throughput assay and is therefore inappropriate for large-scale chemical library screening. Thus, the protein construct FRET-LC3 (His_6-YFP-LC3-EmGFP) was prepared and was found to be a substrate for ATG4B with kinetic parameters very similar to proLC3B itself. Therefore, FRET-LC3 can serve as a ‘‘natural’’ fluorescent substrate more amenable to HTS. The construct has UV absorption maxima at 485 and 512 nm (Figure 13) close to those reported for the individual fluorescent proteins EmGFP and YFP respectively,[63] and thus was evaluated to determine the best conditions for the greatest FRET response and to determine that fluorescence changes correlate with processing of the substrate by ATG4B. While one might expect the maximum FRET response (and therefore sensitivity) would be derived from excitation at 485 nm (with predicted emission of the EmGFP component at 509 nm, almost coincident with the absorption maximum attributed to the YFP component at 512 nm), these conditions gave nonlinear results when monitoring cleavage of FRET-LC3 by ATG4B over time. YFP has a significant absorption shoulder at 490 nm,[64] and thus excitation close to this wavelength can confound the response. After scanning a number of lower excitation frequencies, an optimal excitation wavelength of 465 nm was found. This wavelength provided maximum sensitivity while maintaining linear correlation with changing concentrations of products in the reaction. The enzyme kinetic parameters, determined by mass spectrometric analysis, were found to parallel those derived from the FRET-LC3 assay offering validation of the fidelity of the FRET assay. Optimization of the FRET assay for testing in a 384-well format, performed in collaboration with the CDRD (Centre for Drug Research and Development), showed that preferred conditions involved more physiological sodium ion concentrations of 35 mM and a higher concentration of reducing agent TCEP (5 mM) with a FRET-LC3 substrate concentration of 3.5 μM. The FRET-LC3 assay was shown to give robust data after optimization for

HTS format in 384-well plates and served to identify a small number of verified hits (>50% inhibition at 8 mM) from the 4,761 compounds tested (overall hit rate 0.5%). There was no indication that the assay identified a significant number of false positive compounds due to coloration or fluorescence, and compounds such as sennosides A and B or fluorescein, while found in the initial screen, did not confirm on retest for a dose response. Subsequent to the completion of these studies, another report of a similar FRET assay[32] and also an enzyme-coupled assay have been reported.[33] The enzyme-coupled assay of Shu et al. utilizes an LC3 construct with phospholipase A2 appended at the C-terminus and was used also for screening the commercial LOPAC (Library of pharmaceutically active compounds) compound set, as well as the Spectrum™ library, and was subsequently used in a campaign to screen 340,000 compounds.[65] This construct has the possibility of finding inhibitors of PLA2 as false positives, and thus a parallel PLA2 assay counterscreen was run on identified active hits. Shu's screen of the LOPAC compounds identified three confirmed hits (including aurintricarboxylic acid, also found in our screens), and a low rate of positive PLA2 actives was found (none of his LOPAC hits were found active in the PLA2 counterscreen, while only 1 of 14 of the hits in the Spectrum assay was active). The other hits confirmed by Shu in the LOPAC (R2751 and T7822) were not found as hits in our screen. While we did not screen the Spectrum™ compounds per se, several of the actives found by Shu were also found initially in our screen of the KD2 (known drugs 2) compound library. These included sennosides A and B and gossypol. The sennosides did not confirm in our hands (see table 1). The large-scale 340,351 compound screen conducted by Shu is reported on PubChem as multiple project iterations.[65] It is reported that 1,735 actives were identified with up to 1,085 of these showing significant PLA2 inhibitory activity and 289 actives identified with IC_{50} of <20 μM for ATG4B. No further follow-up is reported, although a second set of 16 compounds screened provided three actives with IC_{50} <20 μM.[65] The FRET assay of Li *et al*[32] uses a fluorescent protein construct cyan fluorescent protein-LC3-yellow fluorescent protein (CFP-LC3-YFP) very similar to our FRET-LC3. Their method involves excitation at 435 nm and monitoring of the ratio of emissions at 477 nm (increase) and the emission at 527 nm (decrease). There are no published results where the Li assay has been used in HTS thus far, but our FRET assay may be complementary, as irradiation is at the higher wavelengths (465–475 nm), and this may offer some advantage to avoid interference from fluorescent test compounds. From the

confirmed active inhibitors, aurintricarboxylic acid is known to be a promiscuous compound active in many screening efforts. It was nonetheless retested as an internal control, and its potency was consistent in multiple assay evaluations. The most potent compound found in the screen was Z-L-Phe-chloromethyl ketone (**40**) with an IC_{50} of 0.63 µM, but it is a potent electrophile that also shows multiple activities, including inhibition of enzymes such as aldehyde dehydrogenase 1 and thiol proteases such as cruzain.[65] The figure below is an illustration of two different examples of how FRET proteins can be used to monitor enzymatic reactions.

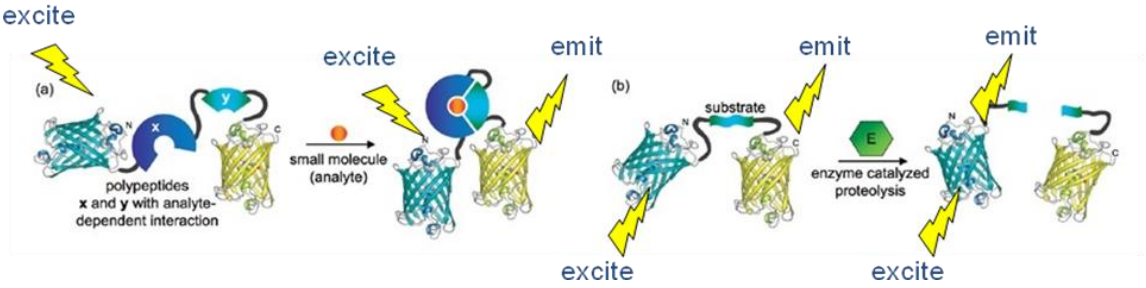


Figure 20: a) Shows how initially the FRET proteins are outside of range for a FRET interaction until a small molecule analyte binds, which brings the proteins closer together and thus illicit a FRET response. b) Shows the method used in this thesis to append FRET proteins on either end of a substrate protein and then after enzymatic cleavage to observe a FRET response decrease. Conversely, one can observe a fluorescence increase from the donor protein as the acceptor or quenching protein is cleaved.

Table 1: Top actives identified from the KD2 library screen.

Compound	% Inhibition	Confirmed*	Confirmed from fresh stock**
Metolazone	120.9	No	n.t.
Aurintricarboxylic acid	90.4	Yes	Yes
Sennoside A	84.9	No	n.t.
4-Nonylphenol	78.7	No	n.t.
Mephenesin	76.9	No	n.t.
Hexachlorophene	72.2	Yes	n.t.
Sennoside B	70.1	No	n.t.
Tannic acid	69.3	Yes	n.t.
Pyrithione zinc	68.0	Yes	n.t.
Isoniazid	65.6	No	n.t.
Vinpocetine	61.3	No	n.t.
Diethylstilbestrol	55.7	No	n.t.
Riluzole hydrochloride	55.6	No	n.t.
Fluorescein	55.1	No	n.t.
Minoxidil	54.8	No	n.t.
Gossypol	52.5	Yes	n.t.
Hypericin	50.0	Yes	Yes

Actives showing >50% inhibition are listed.

*Compounds appeared active in a concentration dependent manner when titrated ($n=1$).

**Compounds appeared active in a concentration dependent manner when fresh stock was made from powder and titrated ($n=3$).

n.t., "not tested," as these compounds were not ordered in house for further testing.

Irreversible Alkylation of ATG4B by NEM

The LC/MS assay, while low throughput, gives a direct and sensitive measure of activity and is not prone to interference with fluorescence quenching or photo bleaching, which can complicate fluorescence-monitored assays. In that one can also monitor the mass of the ATG4B enzyme at the same time. This technique allowed the observation of enzyme integrity throughout the reaction. Additionally, chemical modifications of ATG4B can be directly monitored using mass spectrometry–based methods, which may provide mechanistic insights about potential inhibitors.

The alkylation of ATG4B by NEM indicates that the active site thiol of ATG4B is less accessible than some other peripheral thiols on the protein. This is in keeping with structural observations from the literature[29] that indicate that many of the cysteines are at or near the exterior surface of the protein and that access to the active site and its cysteine thiol is limited by a blocking loop. One cannot rule out from these results that some level of structural reorganization or denaturation might occur due to alkylation of the non-active site cysteine thiol groups that in itself might affect activity or make the active site cysteine more accessible.

Figure 19 shows that inhibition of ATG4B with NEM is time dependent and that the LC3B substrate effectively protects the active site cysteine from alkylation by competition for the active site, which is characteristic for irreversible inhibitors.

Z-L-Phe-Chloromethyl Ketone is an Irreversible Inhibitor and Alkylator of ATG4B

Based on its structure, it was thought that Z-L-Phe-chloromethyl ketone (**40**, page 85) could be an irreversible inhibitor. Indeed, when ATG4B was pre-incubated with **40** (10-fold molar excess for 5 min), enzymatic activity was completely inhibited and subsequent mass spectrometric analysis of the ATG4B indicated only mono-alkylation of the inhibited enzyme (Figure 21), presumably at the active site cysteine. Further mechanistic insights as well as relevant kinetic experiments will be discussed in the next chapters.

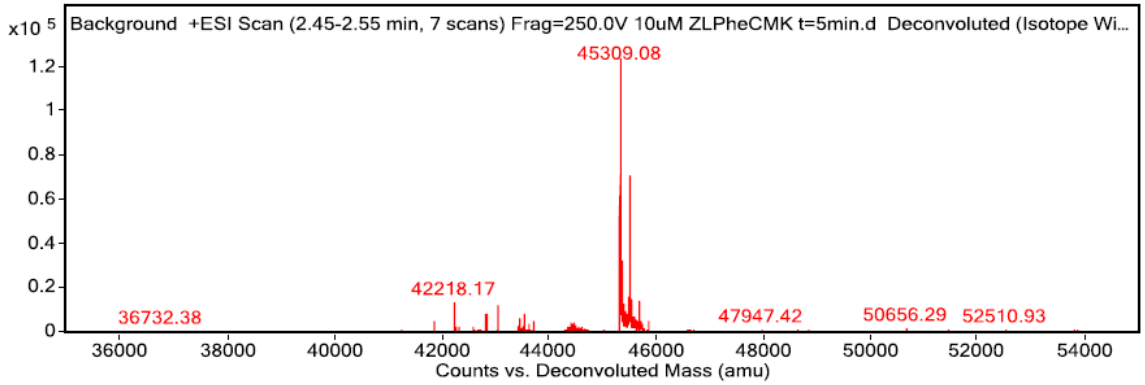


Figure 21: Deconvoluted mass spectrum of ATG4B revealing 1 molecule *Z-L*-Phe-CMK (**40**) alkylation of Atg4B. The mass addition of 297 Da corresponds to the addition of *Z-L*-Phe-CMK with the displacement of chloride.

2.4. Experimental – YFP-LC3B-EmGFP construct

2.4.1. General Methods

The high-resolution mass spectra were determined with the Agilent Technologies 1200 Series liquid chromatography (LC) system using a Zorbax 300SB-C8 Analytical Guard Column (4.6 mm x 12.5 mm, 5 µm) connected to an Agilent Technologies 6210 Time-of-Flight LCMS in positive ion mode with an ESI ion source. Proteins were analyzed by gradient elution from a mixture of acetonitrile and water with 0.1% formic acid at a flow rate of 1 mL/min. The gradient started at 10% acetonitrile for 0.5 min, increased to 60% at 2 min, and maintaining 60% until 3 min before going back to 10% at 3.1 min. The entire analysis lasted up to 5 min with the first minute diverted to waste. Proteins were eluted between 2 and 2.5 min. Proteins used were His-tagged proteins proLC3B and ATG4B, which were kind gifts from Aled Edwards (Structural Genomics Consortium (SGC), Toronto, Canada). (Mca)-TFGM-(Dpa), (Mca)-QETFGMKLS-(Dpa), or (Mca)-QETFGTALA-(Dpa) were purchased after custom synthesis from Biomer Technology (Hayward, CA).

2.4.2. Preparation of the ATG4B FRET Ligand

Protein expression and purification

DNA coding for human LC3-B (F1, residues 1363–1737) was amplified by polymerase chain reaction (PCR) from pCMV-GFP-LC3 expression vector (Cell Biolabs, Inc. Product #CBA-401) and engineered to contain an N-terminal BamH1 restriction site and a C-terminal Sal1 restriction site. DNA coding for YFP (F2, residues 1772–2479) was amplified by PCR from pCAG-YFP plasmid (Addgene; Plasmid 11180) and engineered to contain an N-terminal Sal1 site and C-terminal EcoR1 site. The resulting fragments F1 and F2 were ligated into pRSET-EmGFP expression vector (Invitrogen;cat # V353-20) using BamH1, Sal1, and EcoR1 restriction sites. The final sequence of the expression vector is YFP-LC3B-EmGFP. The resulting plasmid was transformed into *Escherichia coli* XL-10 Gold Ultracompetent cells (Stratagene; cat. # 200314). A colony of freshly transformed cells was cultured in lysogeny broth (LB) medium that contained 100 µg/mL ampicillin. Cells were grown at 37 ^{o}C overnight. The cells were then centrifuged and subjected to a miniprep protocol (Qiagen QuickLyse Miniprep Kit; cat. # 27405). The resulting plasmid was sent for sequencing to verify the expression vector. Once verified, the plasmid was then transformed into *E. Coli* BL-21 CodonPlus (DE-3)-RP cells (Stratagene; cat. # 230245). A colony of freshly transformed cells was cultured in LB medium that contained 100 µg/mL ampicillin. Cells were grown at 37 ^{o}C until A_{600} of 0.6 was reached, and expressed overnight at 25 ^{o}C using 0.5 mM IPTG. Harvested cells were resuspended in binding buffer (20 mM Na_2PO_4, 5 mM imidazole, 20 mM NaCl, pH 7.4) that also contained protease inhibitor mix (Roche cOmplete_Protease Inhibitor Cocktail EDTA free; cat. # 11697498001). Cell lysis was achieved by sonication, and the lysate was centrifuged for 1 h at 12,000 *g*. The supernatant was then subjected to protein purification using a 5 mL Fast Flow HisTrap (GE Healthcare; cat. # 17-5248-02) that was pre-equilibrated with the same binding buffer. Bound protein was then washed with 100 mL wash buffer (20 mM Na_2PO_4, 60 mM imidazole, 500 mM NaCl, pH 7.4) and finally eluted in 20 mL using an elution buffer (20 mM Na_2PO_4, 250 mM imidazole, 300 mM NaCl, pH 7.4). The eluted protein was exchanged to a Tris buffer (50 mM Tris, 100 mM NaCl, pH 8.0) and used without further purification.

DNA Ligation

A total of 3.5 µL of digested F1 and F2 (gel denaturation extracts) were incubated with 1 µL of digested pRSET-EmGFP expression vector using 1 µL 100U T4 DNA Ligase (Invitrogen) in a final volume of 10 µL. After incubating for 30 min at 25^{o}C, an additional 1 µL of T4 DNA ligase was added to the mixture, and incubation continued for an additional 30 min. The resulting plasmid obtained from transformed XL10-Gold miniprep-ultracompetent cells (QIAprep Spin Miniprep Kit, QIAGEN; cat. # 27104) was subjected to an insertion check prior to sequencing. 5 µL of the plasmid was incubated for 1 h at 37^{o}C with 1 µL of BamH1 restriction enzyme and 1 µL EcoR1 restriction enzyme in a final volume of 10 µL using a standard 10X buffer solution. Gel electrophoresis confirmed the insertion size displaying a band at the appropriate size when compared to a control plasmid that did not contain F1 and F2.

2.5. Experimental - Enzyme Kinetics

2.5.1. LC-MS Assay

Kinetics of Cleavage of ProLC3B by ATG4B

Stock solutions of LC3B (223 µM) and Atg4B (213 µM) in 50mM of Tris buffer and 500mM of NaCl at pH 8 were stored in a -80^{o}C freezer in small aliquots. Each assay utilized fresh aliquots of protein and enzyme. Aliquots were flash thawed, then stored over ice at 0^{o}C. A fresh reducing buffer consisting of TCEP (1.0mM) in a stock buffer solution of 50mM of Tris and 35mM of NaCl at pH 8 was prepared daily. ATG4B was diluted to a concentration of 5 µM in the reducing buffer and incubated for 1 h at 0^{o}C. With a final working volume of 100 µL, the assay was run with ATG4B at 12nM to 48nM and proLC3B at concentrations varying from 12 µM to 480 µM. Two samples of protein plus enzyme were prepared simultaneously by first diluting the proLC3B to the appropriate concentration and then adding the enzyme. The cleavage of proLC3B in the two samples was monitored in tandem using LC-MS. The analytical systems used included an Agilent Technologies 1200 Series LC using a Zorbax 300SB-C8 Analytical Guard Column (4.6 x 12.5 mm, 5 µm) connected to an Agilent Technologies 6210 Time-of-Flight LC-MS. Alternating injections from the two samples (1 µL) were performed

every 5 min for 1 h. Elution was performed at 1.0 mL/min using a gradient elution from 90:10 water:acetonitrile + 0.1% formic acid to 40:60 water:acetonitrile + 0.1% formic acid. Data analysis was performed using Agilent Technologies MassHunter Workstation Software Qualitative Analysis Version B.02.00 Build 2.0.197.0. The relative concentrations of proLC3B and cleaved LC3B-I were calculated by measuring the peak height of the two proteins in the deconvoluted mass spectra. K_M determination via LC-MS was performed in duplicate in 4mL HPLC vials using 350 µL glass inserts. The vials were charged with 50 µL of proLC3B at various concentrations and then diluted with the reducing buffer (50 mM Tris, 35 mM NaCl, 5 mM TCEP, pH 8.0) to bring the volume up to 90 µL. Finally, 10 µL of ATG4B at 0.48 mM, preincubated with the reducing buffer for 60 min, was added to the vials. Every 5 min, 0.5–5.0 µL was injected into a TOF LC-MS (Agilent 2610). The analysis of the m/z distribution signal resulting from the protein substrate (proLC3B) was performed by Agilent MassHunter Workstation software elucidating proteins 10,000–30,000 Da between 700 and 1800 m/z. Peak intensities were recorded for the parent substrate (16,720 Da) and cleaved protein (16,161 Da). The amount of cleavage was determined by taking the ratio of cleaved and noncleaved peak intensities and comparing with a standard curve derived from known concentrations of proLC3B and fully cleaved LC3-I. To obtain a standard solution of LC3-I, proLC3B (100 µM) in the reducing buffer was incubated with Atg4B (48 nM) overnight at 4 ^{o}C. Full cleavage was verified by mass spectral analysis. Peak intensities increased linearly with concentration, and ionization potentials did not change for the product peaks over the concentrations tested. K_M and k_{cat} values were determined by nonlinear regression analysis fitting to the equation $Y = Et \cdot k_{cat} \cdot X/(K_M + X)$.

ATG4B Labeling and Inhibition with N-ethylmaleimide

Frozen aliquots of ATG4B (20 µL, 99 µM) were freshly thawed and diluted to a concentration of 5 µM using the reducing buffer (50 mM Tris, 35 mM NaCl, 2 mM TCEP, pH 8.0). *N*-ethylmaleimide (NEM) stock solutions were prepared in dimethyl sulfoxide (DMSO) at various concentrations from 20 to 100 mM. Eppendorf tubes were then charged with 5 µL of NEM solution followed by 85 µL of the reducing buffer and 10 µL of ATG4B to ensure a final concentration of either 500 nM or 250 nM of ATG4B with no more than 5% DMSO and a final volume of 100 µL. Incubation of 500 nM of ATG4B with 1 mM of NEM took place overnight at 4 ^{o}C. Similarly, 250 nM of ATG4B was incubated

with 5 mM of NEM over 5 h at 4 ^{o}C. The enzyme samples were analyzed by TOF LC-MS (Agilent 2610). The analysis was performed using Agilent MassHunter Workstation software elucidating proteins of 20,000–60,000 Da between 700 and 1800 m/z. Incubation of 500 nM of ATG4B with 1 mM NEM led to a major peak at 45,513 Da, with the parent peak at 45,013 Da undetected. Incubation of 250 nM ATG4B with 5 mM NEM led to a major peak at 46,639 Da. These enzyme samples were then assayed for residual activity via LC-MS using proLC3B as the substrate using the method described above.

Time Dependence of Inhibition of ATG4B with NEM

A frozen stock of ATG4B (60 µM) was flash thawed and diluted to a concentration of 48 nM using the reducing buffer (50 mM Tris, 500 mM NaCl, 2 mM TCEP, pH 8.0). A frozen stock of LC3B (1 mM) substrate was flash thawed and diluted to a concentration of 60 µM. ATG4B and LC3B were allowed to incubate with the reducing buffer at 0^{o}C for 1 h prior to experiments. Testing solutions of NEM were prepared at various concentrations from a 15mM stock solution of NEM dissolved in a 9:1 buffer:DMSO solution. The dilutions were performed to ensure no more than 5% DMSO would be present in the final 50 µL working volume. Incubation of 12.5 µL of ATG4B (48 nM) with 12.5 µL of NEM (4 mM) took place for 15 and 60 min, at which point 25 µL of LC3B (60 µM) was added to the mixture. Samples were analyzed by TOF LC-MS as described above.

Determination of Optimal FRET Excitation-Emission Wavelength

Spectrophotometric experiments were performed using a BioTek Synergy 4 Fluorometer. An absorbance spectrum was obtained on a 1 mL sample of FRET-LC3 (10 µM) in order to determine the wavelength of maximum absorption (λ_{max}), which was found at 485 nm (for the EmGFP component) and 512 nm (for the YFP component, Figure 14). Emission spectra were obtained at various excitation wavelengths (435–475 nm) of both FRET-LC3 and cleaved FRET-LC3, each 10 µM. Subtraction of the FRET-LC3 spectra from the corresponding cleaved FRET-LC3 spectra gave the difference emissions at each wavelength. The greatest difference in emission between the whole and cleaved FRET-LC3 samples at 510 nm was observed to be from using an excitation wavelength of 465 nm.

K_M Determination-Fluorescent Protein Substrate *via* Fluorescence

10µL aliquots of Atg4B were flash thawed and diluted to 0.48 µM using a reducing buffer (50 mM Tris, 35 mM NaCl, pH 8.0, 5 mM TCEP). ATG4B was incubated with the reducing buffer at 0°C for 1 h. His_6-YFP-LC3B-EmGFP (FRET-LC3) was also incubated with the reducing buffer for at least 30 min prior to loading, and diluted into aliquots in such a way that 20 µL of fluorescent protein would be delivered to each well to achieve the desired substrate concentrations. 25 µL of the reducing buffer was loaded into an all black 384-well plate (Corning product #3822). 5 µL of ATG4B was then added to all of the wells, except for the negative control well. Finally, 20 µL of fluorescent protein substrate was added to the wells to bring the final volume to 50 µL, and the plate was then loaded into a Synergy 4 Fluorometer. Experiments were performed in triplicate at 25 °C with slow shaking prior to each reading. Fluorescence was scanned every 2 min for up to 2 h excitation at 450 nm and monitoring the emission at 510 nm. Raw data were replotted and analyzed using GraphPad Prism 5 software fitting to the equation: $Y = Et \cdot k_{cat} \cdot X/(K_M + X)$.

Calibration Curve – FRET-LC Fluorescence

The FRET-LC3 calibration curve was obtained by serial dilution into 384-well all-black plates to achieve four final concentrations between 0.15 µM and 4.5 µM and a final volume of 50 µL. 20 µL of the FRET-LC3 at various concentrations were manually pipetted into the plate followed by 30 µL of the reducing buffer (50 mM Tris, 35 mM NaCl, 5 mM TCEP, pH 8.0). A similar calibration curve was derived from a stock sample of FRET-LC3 (100 µM) that had been fully cleaved (verified by mass spectrometry) after incubated in assay buffer with 48 nM ATG4B at 4 °C for 18 h. An average RFU difference was calculated for each substrate concentration and then plotted using GraphPad Prism 5. Differences in fluorescence between intact substrate and cleaved substrate were compared and used as the calibration curve.

Calibration Curve – FRET-LC3 *via* LC-MS

Various concentrations of substrate (0.15 – 20 µM) were prepared in the standard reducing buffer up to a final volume of 100 µL. Each concentration was analyzed in terms of peak intensities and was found to increase linearly with concentration. Additionally, known ratios of cleaved and uncleaved substrate were

analyzed to determine if ionization potentials differed between the two cleaved peptides. Again, the peak intensities increased linearly with concentration.

K_M Determination – FRET-LC3 substrate *via* LC-MS Analysis

K_M determination *via* LC-MS was achieved as described above, except that experiments were performed in duplicate in 4 mL HPLC vials using 350 µL glass inserts. The vials were charged with 50 µL of FRET-LC3 at various concentrations and then diluted with the reducing buffer (50 mM Tris, 35 mM NaCl, 5 mM TCEP, pH 8.0) to bring the volume up to 90 µL. Finally, 10 µL of Atg4B at 0.48 µM preincubated with the reducing buffer was added to the vials. Every 5 min, 1 µL was injected into a TOF LC-MS (Agilent 2610). The analysis of the m/z distribution signal resulting from the protein substrate was performed using Agilent MassHunter Workstation software elucidating proteins of 20,000–80,000 Da between 700 and 1800 m/z. Peak intensities were recorded for the parent substrate (72,902 Da) and both cleaved peptides (27,885 and 45,036 Da). The amount of cleavage was determined by taking the ratio of cleaved and noncleaved peak intensities and by comparison with a standard curve. The ratio between the 27,884 and 72,902 Da peaks were used for the calculations. The standard curve was obtained by injecting known ratios of cleaved and noncleaved substrate. Peak intensities increased linearly with concentration, and ionization potentials did not change for the product peaks over the concentrations tested. K_M and k_{cat} values were determined by nonlinear regression analysis fitting to the equation: $Y = Et \cdot k_{cat} \cdot X/(K_M + X)$.

LC-MS Assay for Inhibition of ATG4B

ATG4B was diluted to a concentration of 240 nM in the reducing buffer (50 mM Tris, 35 mM NaCl, 5 mM TCEP, pH 8.0) and incubated for 1 h at 0 ^{o}C. With a final working volume of 100 µL, the assay was run with ATG4B at 48 nM final concentration. Inhibitor at various concentrations (in 5 µL DMSO or DMSO control) was added to 4 mL HPLC vials using 350 µL glass inserts and diluted with 25 µL of the reducing buffer, and then 20 µL of the above solution of 240 nM ATG4B was added, and the mixture was incubated at room temperature for 30 min. Finally, 50 µL of the substrate, proLC3B or FRET-LC3 at 5 µM, was added and the reaction was allowed to proceed. To obtain triple replicates, samples of the control reaction with inhibitor or two inhibitor samples

were injected alternately in parallel for each inhibitor concentration. Every 5 min, 0.5–5.0 µL was injected into a TOF LC-MS (Agilent 2610) alternating between the two parallel reactions. The analysis of the m/z distribution signal resulting from the protein substrate (proLC3B or FRET-LC3) was performed using Agilent MassHunter Workstation software elucidating proteins of 10,000–30,000 Da between 700 and 1800 m/z. Peak intensities were recorded for the parent substrate (16,720 Da) and cleaved protein (16,161 Da) for pro-LC3 and the parent substrate (72,903 Da) and two fragments (27,885 and 45,036 Da) for cleaved FRET-LC3. The amount of cleavage was determined by taking the ratio of cleaved and noncleaved peak intensities and by comparison with a control sample of proLC3 or FRET-LC3 with no inhibitor added.

Chapter 3. 1st Generation Irreversible Inhibitors

3.1. Research Objective

The objective of this part of the project was to design and synthesize a small molecule inhibitor of the cysteine protease ATG4B. We envisaged a type of halomethyl ketone that would react irreversibly with the active site cysteine residue to form a covalent bond with the enzyme. As previewed in chapter 1, there are many cysteine protease inhibitors currently in development for a variety of conditions.[46] Due to the unique properties of ATG4B's active site, we proposed that a halomethyl ketone would be small enough to react within the active site. Evidence from the HTS has already shown that small molecule inhibitors, such as Z-L-Phe-CMK, have a chance at successfully inhibiting ATG4B.

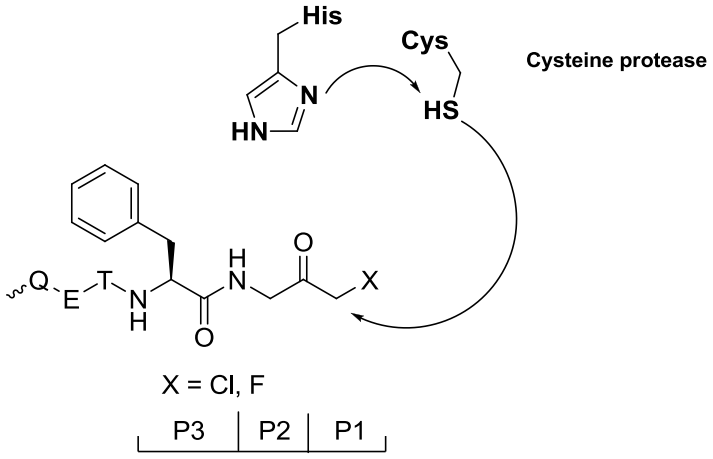
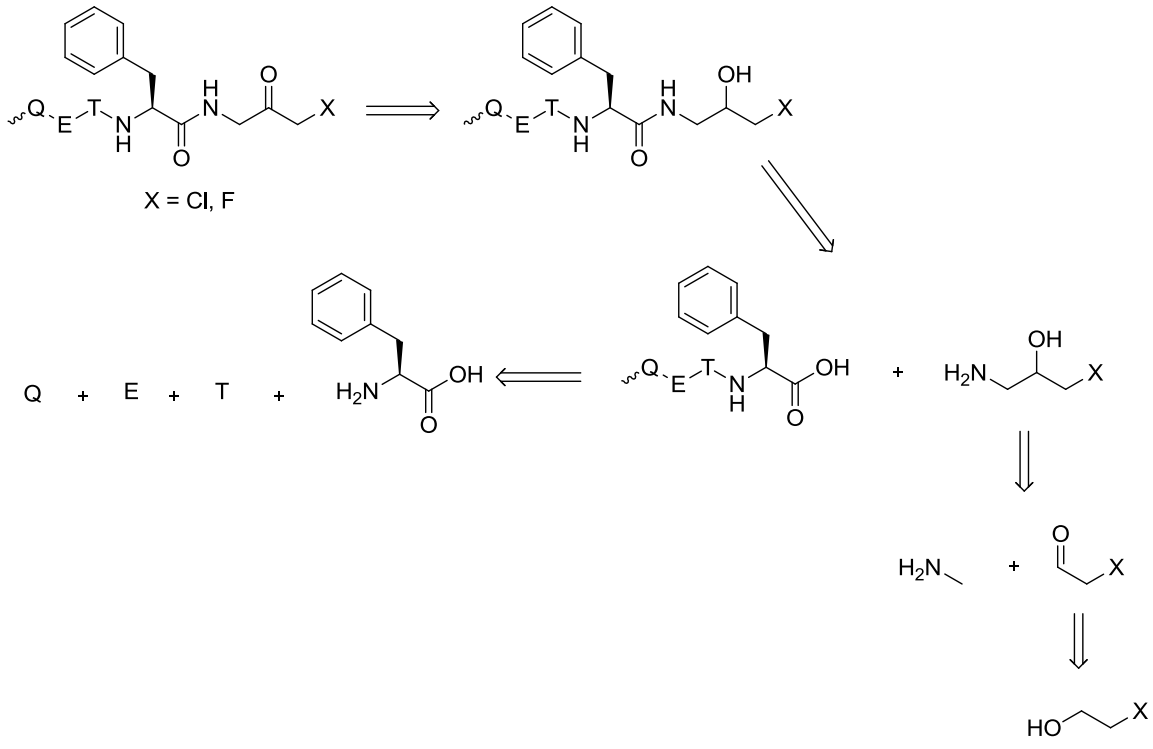


Figure 22: An illustration of the active-site residues of ATG4B with the active cysteine thiol attacking and displacing X (Cl or F) from a proposed inhibitor structure. The proposed inhibitor will mimic the cleavage site on LC3B; glycine and phenylalanine at the P2 and P3 sites respectively.

3.2. Retrosynthetic Analysis

By way of rational design, the initial inhibitors would mimic the amino acid sequence of LC3B by incorporating phenylalanine and glycine into the small molecule inhibitor as can be seen from scheme 1, below. It was proposed that one could start the synthesis with simple, commercially available starting materials such as 2-chloro or 2-fluoroethanol. The hydroxyl group could be oxidized to an aldehyde and after subsequent addition of nitromethane the result would be an aminoethanol derivative that would serve as the electrophilic "warhead" of our inhibitor, after a late stage oxidation. The oxidation of the secondary alcohol to a ketone is one of the last steps of the synthesis due to the reactive nature of halomethyl ketones.[66] From this scheme, it is possible to incorporate any number of amino acids to the inhibitor to potentially increase (or decrease) the selectivity and potency against ATG4B.

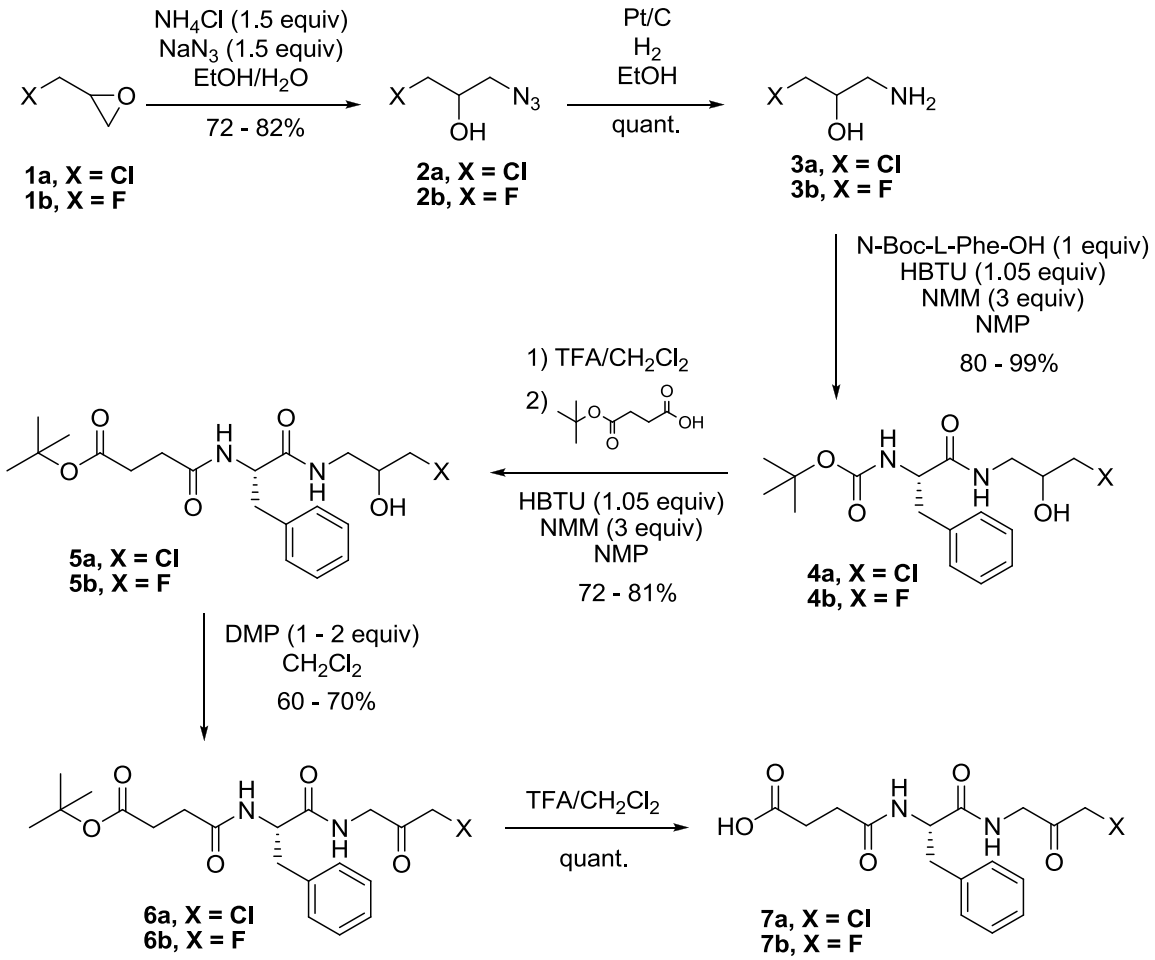
Scheme 1



3.3. Synthesis of Halomethyl ketones

Previous work within the Young Lab was focused on making synthetic fluorescent peptide substrates for ATG4B. The synthetic fluorescent peptide substrates were made to mimic the amino acid sequence of LC3B. Through this work, it was discovered that many of the synthetic substrates were not turned over by ATG4B unless the N-terminus was capped with a succinic acid moiety (unpublished work). Thus, incorporating an *N*-hemisuccinic acid amide as an N-capping group into the inhibitor design was logical to evaluate. The synthesis of the chloro and fluoromethyl ketone inhibitors is shown in scheme 2.

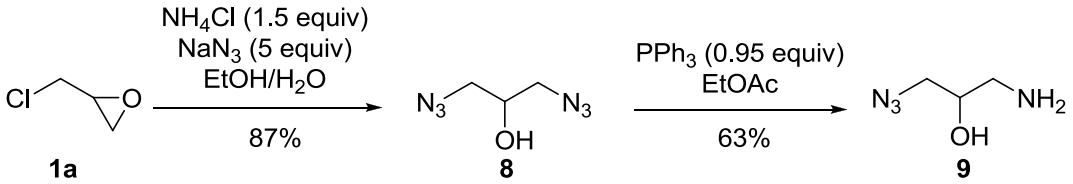
Scheme 2



The initial plan was to start with 2-fluoroethanol and perform a Swern oxidation[67] reaction to convert the alcohol moiety into an aldehyde. Combining DMSO and oxalyl chloride at low temperature successfully generated the reactive intermediate dimethylchlorosulfonium chloride, which was confirmed by the evolution of CO and CO_2 observed as bubbles in the reaction mixture. Unfortunately, after addition of 2-fluoroethanol and an acidic workup, the expected product could not be isolated. Such a small molecule is difficult to observe when performing TLC analysis. The product is not UV active and stains were difficult to interpret from crude reaction mixtures. Attempts to react the aldehyde with nitromethane *in situ* were also unsuccessful. At this time, an alternative route was proposed starting from an epoxide such as epichloro and epifluorohydrin. Epoxide ring opening with a suitable nucleophile would produce the desired secondary alcohol in 1 step as opposed to the previously proposed route involving an oxidation and subsequent addition.

Indeed starting from epichlorohydrin (**1a**) proved more successful. Treatment of epichlorohydrin with sodium azide afforded the pure ring opened azide (**2**), obtained by flash column chromatography. The azide (**2**) was then reduced to an amine group using catalytic hydrogenation with platinum on activated carbon as the catalyst and a hydrogen filled balloon delivering a positive pressure of hydrogen. This hydrogenation reaction was straight forward and progress of the reaction was monitored *via* TLC observing a disappearance of starting material with a new baseline spot forming over time. Once the starting azide was completely consumed, the reaction mixture was concentrated under reduced pressure to afford either the animochloropropanol (**3a**) or aminofluoropropanol (**3b**) as a pale yellow oil and colourless oil, respectively. The product was confirmed by NMR analysis of the crude sample and used for the next step without further purification.

Scheme 3



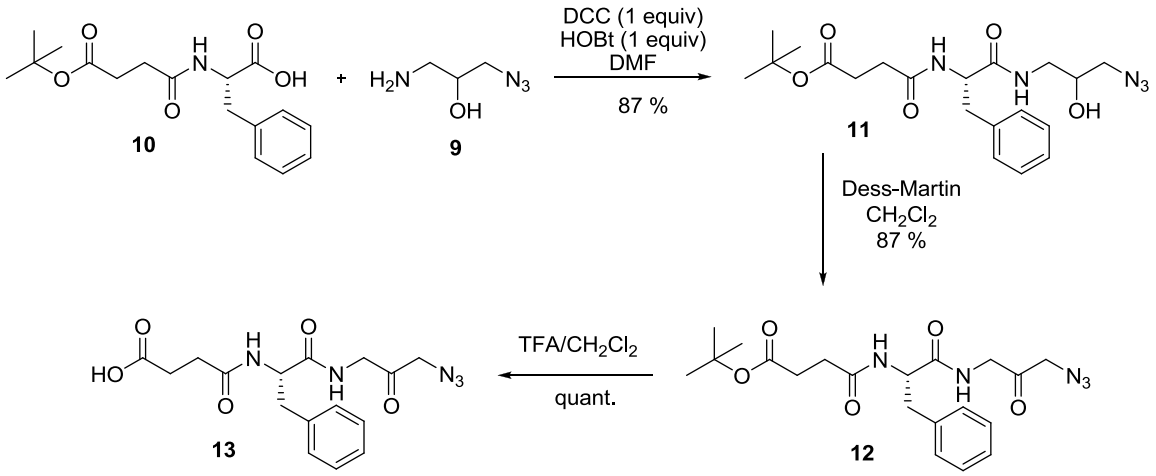
1a **8** **9**

Another inhibitor bearing an azido moiety was also proposed to potentially probe the reactivity of a series of substituted methyl-ketones. An azidomethyl ketone inhibitor could be synthesized from the same epichlorohydrin starting material. Simply treating the epichlorohydrin with a vast excess of sodium azide afforded the di-azide compound (**8**) in good yield. A selective reduction was needed for the next step to reduce one of the azide groups into an amine. This type of reaction, commonly referred to as the Staudinger reduction, typically involves the use of triphenylphosphine in a bi-phasic reaction mixture involving an organic solvent and water in the presence of an acid. As one of the azide groups is reduced, the resulting amine will become protonated and partition into the aqueous phase, preventing over reduction. This method was attempted; however the product was too soluble in the aqueous phase and could not be extracted even after making the aqueous phase basic. It turns out that a simple treatment with 0.95 equivalents of triphenylphosphine in EtOAc will afford the mono-reduced compound (**9**) in 63% yield with the rest of the by-product being the over-reduced diamine. This aminoazide compound was then carried through a similar synthetic route as the aminochloropropanol (**3a**). This can be seen in scheme 4.

Aminochloropropanol (**3a**) was then subjected to a peptide coupling reaction with *N*-Boc-phenylalanine. Many peptide coupling conditions exist and a wide array of reagents have been manufactured and optimized to suit specific needs of peptide synthesis. Various reagents were evaluated with varying degrees of success. Reagents such as EDC or HBTU can be quite useful solution phase reagents as their by-products can be readily extracted. Reagents such as DCC may be more suitable for solid phase synthesis since its by-product is a urea compound that is sometimes difficult to remove even by flash column chromatography. For this solution phase chemistry, HBTU was the most desirable reagent to use since it is easy to handle and also minimizes epimerization of amino-acids during the coupling.[68] Typical solvents for peptide synthesis involve very polar high boiling solvents such as DMF and NMP. *N*-Boc-phenylalanine was first activated by HBTU at low temperature (0 ^{o}C) in NMP for at least 15 minutes prior to the addition of the aminochloropropanol (**3a**). The product was isolated by first diluting the reaction mixture with copious amounts of EtOAc and then extracting out the NMP with a mixture of water and brine. This proved to be the most efficient method for removing NMP after the reaction is complete. The product was

finally purified by flash column chromatography to afford the coupled product (**4a**) as a white crystalline foam. The next step in the synthesis was a deprotection reaction to remove the Boc group. The standard deprotection procedure is to use a strong acid to facilitate the formation of a stable tertiary carbocation with the removal of CO_2 as a by-product of the reaction. This was achieved using TFA and the reaction was concentrated under reduced pressure and the free amine was used in the next coupling reaction without further purification. The 4-*tert*-butoxy-4-oxobutanoic acid was prepared from a ring-opening reaction between succinic anhydride and *tert*-butanol[69] and was activated by HBTU in the same type of peptide coupling reaction to form (**5a**) in good yields. It may be useful to note here that the next oxidation reaction using Dess-Martin periodinane[70,71] fails in the presence of a free acid. It is speculated that the free acid may form unusually strong I-O bonds complexing the reagent and substrate together and preventing the oxidation reaction from proceeding. After protecting the free acid as a *tert*-butyl ester the oxidation reaction proceeded smoothly with the last step of the synthesis being a simple hydrolysis reaction using a 50:50 mixture of TFA and CH_2Cl_2 to convert the *tert*-butyl ester back into a free acid.

Scheme 4

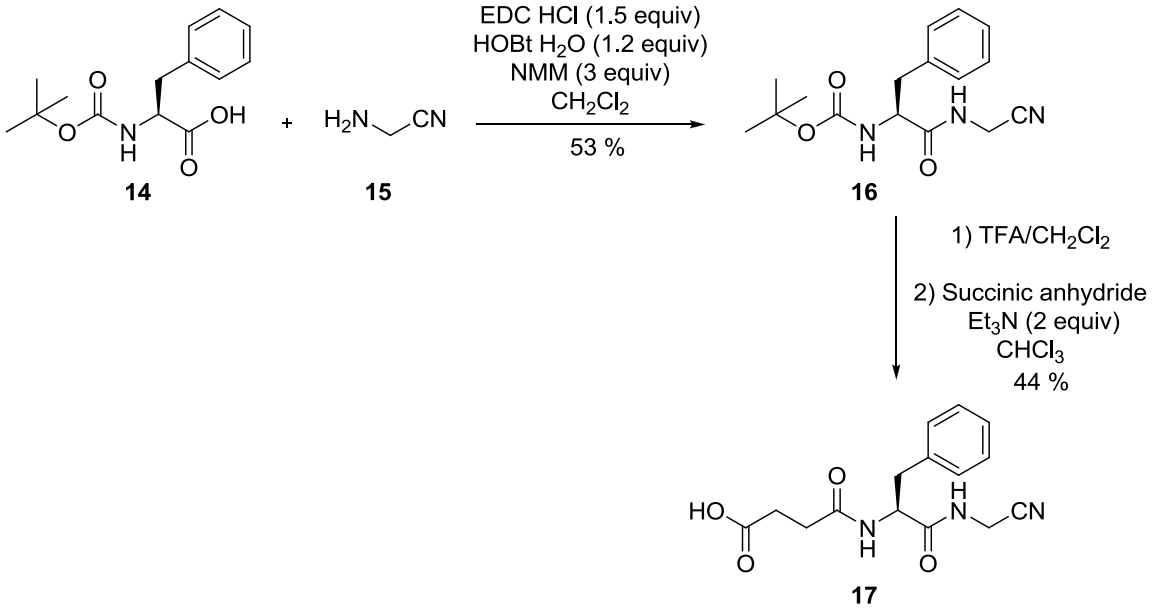


The aminoazidopropanol (**9**) was coupled to an advanced intermediate that was made available during the synthesis of the initial batch of chloro and fluoromethyl ketones. The advanced intermediate (**10**) was synthesized starting from a benzyl-ester protected *L*-phenylalanine isolated as a PTSA salt, which was coupled to a hemi-

succinic *tert*-butyl ester[69] and then deprotected using activated Palladium on charcoal under a positive pressure of H_2. The resulting compound (**10**) was then coupled to the aminoazidopropanol (**9**) using standard peptide coupling reagents. The remaining steps of the synthesis proceeded smoothly as described above to afford the final azidomethyl ketone (**13**) as a colourless oil after flash column chromatography.

3.4. Synthesis of Cyanomethyl Ketone

Scheme 5



A cyanomethyl ketone (**17**) was also synthesized with the idea that this type of "warhead" would bind covalently, but reversibly with the active site thiol in ATG4B. This compound (**17**) was obtained in 3 steps starting from the peptide coupling of phenylalanine and aminoacetonitrile using EDAC, HOBt and NMM to afford (**16**). The final compound was obtained after a Boc-deprotection reaction using TFA/CH_2Cl_2 in a 50:50 mixture for 30 min and finally by reacting the free amine with succinic anhydride in a ring opening reaction to afford (**17**) after flash column chromatography. Unfortunately both the cyano compound (**17**) and the azide compound (**13**) were not active as inhibitors of ATG4B in any of our enzyme assays (IC_{50} >> 200 µM).

Determination of $[\alpha]_D$ and Chiral Purity

Great care was taken to ensure that *L*-phenylalanine would not racemize during the key peptide coupling steps. The use of reagents such HOBt and HBTU greatly reduce the risk of racemization.[68] Nevertheless, the *D*-phenylalanine analog of compound **7b** was synthesized simply by substituting *N*-Boc-*D*-phenylalanine for *N*-Boc-*L*-phenylalanine in scheme 2. The synthesis was carried out in the same manner for the rest of the synthesis in accordance to scheme 2. The final compound, **7b-D**, was obtained in good yield and the rotation was measured in solution (MeOH) according to the experimental section 3.5.3. It was found that compound **7b** had a calculated specific rotation of -2.0^o while **7b-D** had a calculated specific rotation of $+2.5^o$. Racemic mixtures would have specific rotations of 0^o. In order to further validate the chiral purity of **7b** and **7b-D**, both pure and racemic mixtures of each of the compounds were analyzed *via* HPLC equipped with a Chiralpak IA-3 column (Diacel Corporation PartNo 80524). Initially, separation of **7b** and **7b-D** could not be achieved with normal phase solvents. However, the *tert*-butyl ester **6b** along with the D-phenylalanine analogue **6b-D**, could be separated with baseline separation achieved using a mobile phase of 80:20 isopropanol:hexanes with a run time of 15 min. Compound **6b-D** displayed a retention time of approximately 7 min while a retention time of 10 min was revealed for **6b**. HPLC analysis reveals approximately 95% enantiomeric excess for **6b** and >99% enantiomeric excess for **6b-D**.

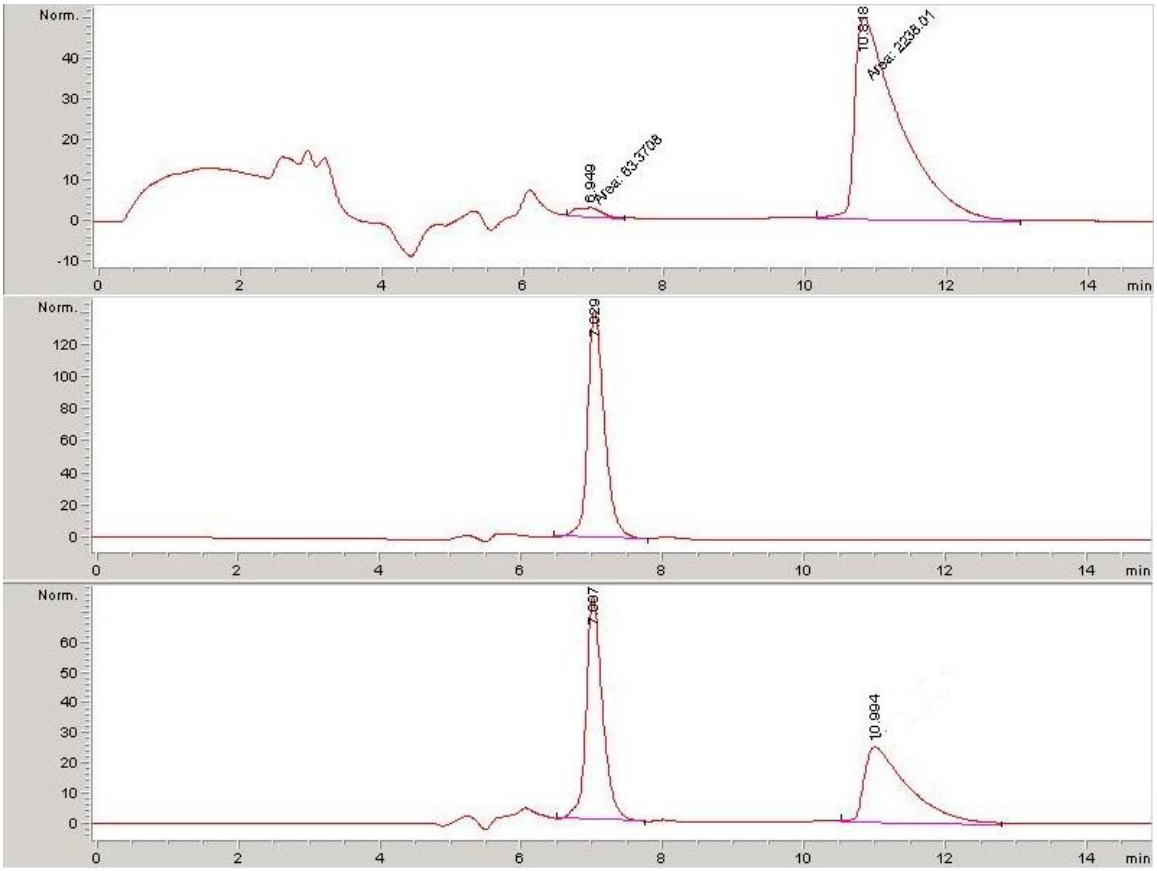


Figure 23: Chiral HPLC trace of **6b** (top), **6b-D** (middle) and a 50:50 mixture of **6b** and **6b-D** (bottom) monitoring at 254nm.

3.5. Experimental

3.5.1. General methods

^{1}H and ^{13}C NMR spectra were recorded with Bruker Avance II™ 600 MHz, Bruker Avance III™ 500 MHz or Bruker Avance III™ 400 MHz. Processing of the spectra was performed with MestRec™ software. The high-resolution mass spectra were recorded in positive ion-mode with an ESI ion source on an Agilent™ 2690 Time-of-Flight LC/MS mass spectrometer. Analytical thin-layer chromatography (TLC) was performed on aluminum plates pre-coated with silica gel 60F-254 as the absorbent. The developed plates were air-dried, exposed to UV light and/or dipped in $KMnO_4$ solution and heated. Column chromatography was performed with silica gel 60 (230-400 mesh). All HPLC analyses were performed utilizing a Dikma Technologies™ Inspire® C18

reverse-phase analytical column (4.6 × 150 mm). All HPLC purifications were carried out using an Agilent™ C18 reverse-phase semi-preparative scale column (21.2 × 250 mm).

3.5.2. General Dess-Martin Periodinane Oxidation procedure

Dess-Martin Periodinane was prepared from IBX.[70] IBX was prepared from 2-iodobenzoic acid.[71] Dess-Martin Periodinane was stored under argon at -20 ^{o}C.

The alcohol to be oxidized was first dissolved in CH_2Cl_2 at a concentration of 0.1 – 0.2 moles/L. To this stirring solution, Dess-Martin Periodinane (1 – 4 equiv) would be added and the solution would be allowed to stir for 2 – 16 hr monitoring the reaction *via* TLC. Upon complete consumption of starting material, the reaction mixture was diluted with CH_2Cl_2 and washed with a sat. solution of $Na_2S_2O_3$ at least 3 times or until the organic layer was translucent. The organic layer was then washed with brine and dried over Na_2SO_4 and purified by flash column chromatography.

3.5.3. Determination of $[\alpha]_D$

The degree of rotation for solutions was measured using a Perkin Elmer Polarimeter 341 with a cell of path length 1 dm. Compounds were dissolved in ethanol to a known concentration and loaded into the cell and placed into the polarimeter. The degree of rotation was then recorded. The specific rotation ($[\alpha]_D$) was then calculated based on the equation:

$$\alpha_D = \frac{\alpha}{l \times c} \qquad (17)$$

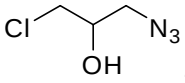
α is the measured rotation, l is the path length in decimeters, c is the concentration of the solution in g/mL.

3.5.4. Determination of Enantiomeric Excess

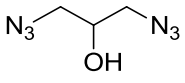
Compounds **6b** and **6b-D** were resuspended in isopropanol to a concentration of 10 mM (~4 mg dissolved in 500 µL isopropanol). A racemic sample was prepared by mixing 250 µL of each separate solution into one vial. HPLC analysis was performed

using Daicel Corporation ChiralPak IA-3 normal phase analytical column (3 µm, 4.6 x 150 mm, partno. 80524) with a mobile phase of 80:20 isopropanol:hexanes, a runtime of 15 min, a flow rate of 0.4 mL/min and an injection volume of 10 µL. Compound **6b** displayed a retention time of approximately 10 min while the retention time of **6b-D** was approximately 7 min. Comparison with the racemic sample revealed an enantiomeric excess of 95% for **6b** and >99% for **6b-D** after peak integration.

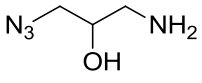
3.5.5. Characterization of compounds



1-azido-3-chloropropan-2-ol, **2a**. A 250 mL round bottom flask was charged with sodium azide (10.5 g, 0.16 moles), ammonium chloride (8.60 g, 0.16 moles), and epichlorohydrin (10.0 g, 0.100 moles) and dissolved in 150 mL of a 50:50 mixture of ethanol and water. The solution was allowed to stir at r.t. over night. The reaction mixture was then extracted 5 times with 10 mL CH_2Cl_2. The organic extract was concentrated under reduced pressure and the crude residue was purified via flash column chromatography using a 10% EtOAc in hexanes solvent system to afford pure azido chloropropanol as a colourless liquid in 72.5% yield. ^{1}H NMR spectrum agrees with that in previous reports.[72]

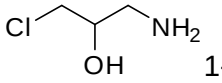


1,3-diazidopropan-2-ol, **8**. A 100 mL round bottom flask was charged with epichlorohydrin (1.00 g, 10.807 mmol), NaN_3 (3.512 g, 54.036 mmol), and NH_4Cl (0.578 g, 10.807 mmol) dissolved in H_2O (20 mL) and heated to 70^{o}C. The reaction mixture was allowed to stir at this temperature overnight before extracting with CH_2Cl_2 (4 x 10 mL). The organic extract was concentrated under reduced pressure and purified by silica plug (20% EtOAc in hexanes) to afford a colourless oil in 78% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 3.96 (s, 1H), 3.52 – 3.36 (m, 4H), 2.32 (s, 1H).[73]

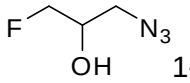


1-amino-3-azidopropan-2-ol, **9**. A 100 mL round bottom flask was charged with 1,3-diazido-2-propanol (1.10 g, 8.30 mmol) and PPh_3 (2.00 g, 7.90

mmol) dissolved in EtOAc (20 mL). After complete addition of PPh_3, the reaction mixture was allowed to stir for 2 hr before 200 µL of H_2O was added and then the reaction continued to stir overnight. The reaction mixture was then filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography (10% MeOH in CH_2Cl_2) to afford a colourless oil in 66% yield. ^{1}H NMR (400 MHz, MeOD) δ_{ppm} = 3.71 (ddd, *J* = 8.7, 7.0, 4.4 Hz, 1H), 3.36 – 3.22 (m, 3H), 2.67 (ddd, *J* = 20.5, 13.1, 5.9 Hz, 2H).[73]

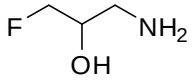


1-amino-3-chloropropan-2-ol, **3a**. In a 100 mL round bottom flask, 1-azido-3-chloropropan-2-ol (5.00 g, 36.88 mmole) was dissolved in 30 mL ethanol. Activated platinum on charcoal was then introduced into the flask (0.50 g) under argon. Following the addition of platinum, the flask was evacuated and filled with hydrogen 3 times. Hydrogen atmosphere was maintained with balloons and the mixture was allowed to stir for 5 days refilling hydrogen as needed. The reaction was monitored by disappearance of starting material via TLC. When the reaction was complete, the mixture was filtered through celite and concentrated under reduced pressure. The crude residue was used without further purification and gave essentially quantitative yields. ^{1}H NMR (600 MHz, $CDCl_3$) δ_{ppm} = 3.83 – 3.71 (m, 2H), 3.60 (d, 1H, J = 1.8 Hz), 3.58 (d, OH, J = 3 Hz), 2.94 (dd, 1H, J_{AB} = 19.2 Hz, J_{AX} = 6 Hz), 2.82 (dd, 1H, J_{AB} = 19.8 Hz, J_{BX} = 10.2 Hz).[72]

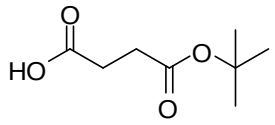


1-azido-3-fluoropropan-2-ol, **2b**. A 100 mL round bottom flask was charged with sodium azide (1.28 g, 19.70 mmole) and ammonium chloride (1.05 g, 19.70 mmole) and epifluorohydrin (1.00 g, 13.10 mmole) and dissolved in a 50/50 mixture of ethanol and water. The reaction mixture was then heated and allowed to reflux for 2 days. The reaction mixture was then cooled and extracted with dichloromethane. Complete extraction was monitored by the disappearance of the product from the aqueous layer via TLC. The dichloromethane extract was then dried over sodium sulphate and concentrated under reduced pressure. Flash column chromatography (20% EtOAc in hex, isocratic) afforded pure azidofluoropropanol in 83% yield as a colourless liquid. ^{1}H NMR (600 MHz, $CDCl_3$) δ_{ppm} = 4.54 – 4.41 (m, 2H), 4.10

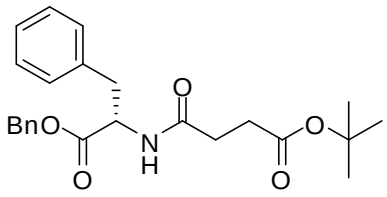
– 4.02 (m, 1H), 3.52 – 3.44 (m, 2H). ^{13}C NMR (150 MHz, CDCl3) δppm = 84 (d, J_F = 672 Hz), 69.4 (d, J_F = 78 Hz, 52.5 (d, J_F = 24 Hz).



1-amino-3-fluoropropan-2-ol, **3b**. 1-azido-3-fluoropropan-2-ol (0.70 g, 5.95 mmole) was dissolved in 15 mL ethanol in a 100 mL round bottom flask. The reaction flask was then evacuated and filled with argon. 0.2 g of activated platinum on carbon was then added slowly into the reaction flask. The flask was evacuated and filled with hydrogen. A positive pressure of hydrogen was maintained by hydrogen filled balloons and the reaction was allowed to stir at room temperature for 2 days. The reaction mixture was filtered through celite and concentrated under reduced pressure to afford the essentially pure amine product in quantitative yield as a pale yellow oil. This material was used with further purification. ^{1}H NMR (600 MHz, $CDCl_3$) δ_{ppm} = 4.54 – 4.45 (m, 1H), 4.42 – 4.33 (m, 1H), 3.88 – 3.77 (m, 1H), 2.91 (dd, 1H, J_{AB} = 19.2 Hz, J_{AX} = 6.6 Hz), 2.81 (dd, 1H, J_{AB} = 19.2 Hz, J_{BX} = 10.8 Hz). ^{13}C NMR (150 MHz, CDCl3) δppm = 84 (d, J_F = 672 Hz), 69.4 (d, J_F = 78 Hz, 43.5 (d, J_F = 24 Hz).

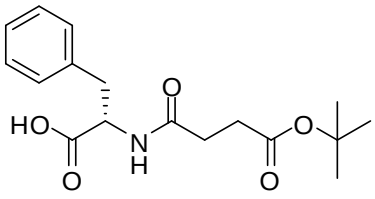


4-*tert*-butoxy-4-oxobutanoic acid. This material was made following literature procedures.[69] The ^{1}H and ^{13}C NMR spectra matches the reported values in the literature.



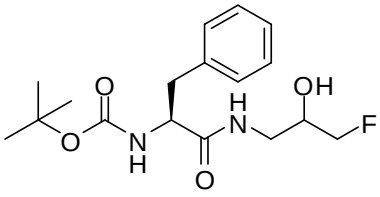
(*S*)-*tert*-butyl 4-(1-(benzyloxy)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **10a**. 4-*tert*-butoxy-4-oxobutanoic acid (0.81 g, 4.67 mmole) was dissolved in 4 mL of freshly degassed DMF in a 100 mL round bottom flask. The reaction flask was then immersed in an ice water bath and allowed to stir for 10 min. HBTU (1.86 g, 4.91 mmole) was then slowly added and the mixture allowed to stir at 0 ^{o}C for an additional 15 min. In a separate 10 mL vial, (*S*)-1-(benzyloxy)-1-oxo-3-phenylpropan-2-aminium *p*-toluene sulfonate (2.00 g, 4.67 mmole) was dissolved in 5 mL freshly degassed DMF and then added dropwise into the reaction

mixture at 0 ^{o}C. After full addition was complete, the reaction mixture was allowed to stir for an additional 10 min before 4-methylmorpholine (1.54 mL, 14.03 mmole) was added dropwise over 5 minutes. The reaction flask was removed from the ice bath and allowed to stir overnight under argon atmosphere. The reaction mixture was then evaporated under reduced pressure until most of the DMF had evaporated. The resulting residue was dissolved in 30 mL ethyl acetate and washed with a 50/50 mixture of brine and water (25 mL x 4), then HCl (50 mL x 2), $NaHCO_3$ (50 mL x 2), brine (50 mL x 2) and the organic phase was dried over sodium sulphate. Flash column chromatography afforded the pure amide product as a colourless oil in 89% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.32 – 7.27 (m, 3H), 7.24 – 7.20 (m, 2H), 7.17 – 7.13 (m, 3H), 6.97 – 6.93 (m, 2H), 6.06 (d, 1H, J = 7.6 Hz), 5.11 – 5.01 (d, 2H, J_{AB} = 12 Hz), 4.87 – 4.81 (dt, 1H, J = 5.6 Hz), 3.09 – 3.04 (dd, 1H, J_{AB} = 13.6 Hz, J_{AX} = 5.6 Hz), 3.03 – 2.99 (dd, 1H, J_{AB} = 13.6 Hz, J_{BX} = 3.6 Hz), 2.50 – 2.45 (m, 2H), 2.39 – 2.34 (m, 2H), 1.36 (s, 9H). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 172.05, 171.36, 171.20, 135.72, 135.08, 129.37, 128.62, 128.60, 128.54, 80.80, 67.24, 53.21, 37.87, 31.10, 30.60, 28.06. HRMS (ESI, TOF) m/z calculated for $C_{24}H_{29}NO_5$: 434.1943 (M+Na), found: 434.1938 m/z (M+Na).

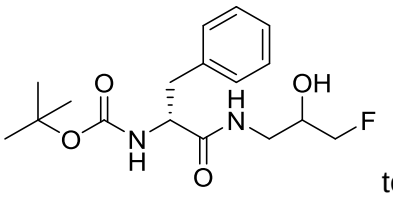


(S)-2-(4-tert-butoxy-4-oxobutanamido)-3-phenylpropanoic acid, **10**. The benzyl ester was removed via platinum catalyzed hydrogenation. Benzyl ester **10a** (2.173 g, 5.280 mmol) was dissolved in 20 mL ethanol under inert atmosphere. Activated platinum on carbon (0.300 g) was added to the reaction mixture slowly under argon atmosphere. Upon complete addition of the platinum, the reaction flask was evacuated and refilled with H_2 3 times. A positive pressure of H_2 gas was maintained by H_2 filled balloons. The reaction was allowed to stir over the course of 1 weekend and completion was monitored by disappearance of starting material via TLC. The reaction mixture was then filtered through a pad of celite and concentrated under reduced pressure. The crude material was purified via flash column chromatography (5% methanol in CH_2Cl_2) to afford the free acid **10** in 74% yield as an off white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.36 – 7.29 (m, 3H), 7.22 – 7.18 (m, 2H), 6.31 – 6.27 (d, 1H, J = 7.6 Hz), 4.87 (q, 1H, J = 6.4 Hz), 3.28 – 3.21 (dd, 1H, J_{AB}

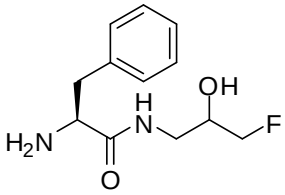
= 14 Hz, J_{AX} = 5.6 Hz), 3.17 – 3.10 (dd, 1H, J_{AB} = 14 Hz, J_{BX} = 6.4 Hz), 2.65 – 2.45 (m, 4H), 1.45 (s, 9H). $_{13}$C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 174.01, 172.33, 172.31, 135.73, 129.35, 128.67, 127.22, 81.17, 53.38, 37.19, 31.04, 30.62, 28.04. HRMS (ESI, TOF) m/z calculated for $C_{17}H_{23}NO_5$: 344.1474 (M+Na), found: 344.1457 m/z (M+Na).



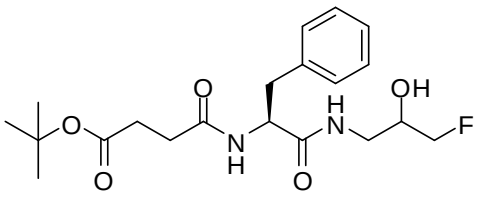
tert-butyl (2S)-1-(3-fluoro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **4b**. A 100 mL round bottom flask was charged with Boc-L-Phe-OH (0.500 g, 1.884 mmol) dissolved in 5 mL *N*-methyl-2-pyrrolidone (NMP). 1 equivalent of *N*-methylmorpholine (NMM, 0.207 mL, 1.884 mmol) was then added. HBTU (0.750 g, 1.970 mmol) was then added before immersing the flask in an ice-water bath. The reaction mixture was allowed to stir at 0^{o}C for 15 min. In a separate vial, the aminofluoropropanol **3b** (0.175 g, 1.884 mmol) was dissolved in 1 mL NMP and checked with pH paper to ensure basicity. This aminofluoropropanol solution was then added drop-wise to the reaction at 0 ^{o}C. The mixture was allowed to stir at 0 ^{o}C for an additional 15 min before adding another 2 equivalents of NMM (0.414 mL, 3.770 mmol). The ice-bath was removed and the reaction was allowed to warm up to room temp and stirred 16 hr. The reaction was diluted with 50 mL EtOAc and then washed with a mixture of 50:50 brine:H_2O (3 x 20 mL), then brine (3 x 20 mL). The organic layer was then dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified *via* automatic flash column chromatography using a gradient from 15:85 EtOAc:hexanes to 100% EtOAc to give **4b** as a colourless oil in quantitative yield. ^{1}H NMR (500 MHz, DMSO) δ_{ppm} = 7.94 (s, 1H), 7.35 – 7.11 (m, 5H), 6.92 (dd, *J* = 2.57, 8.35 Hz, 1H), 5.22 (d, *J* = 5.02 Hz, 1H), 4.40 – 4.06 (m, 3H), 3.68 (s, 1H), 3.19 – 3.04 (m, 2H), 2.95 – 2.89 (m, 1H), 2.83 – 2.63 (m, 1H), 1.28 (s, 9H). ^{13}C NMR (126 MHz, DMSO) δ_{ppm} = 172.36, 170.80, 155.66, 138.64 (d, *J* = 10.08 Hz), 129.64, 128.46, 126.63 (d, *J* = 1.26 Hz), 86.13 (d, *J* = 7.56 Hz), 84.80 (d, *J* = 8.82 Hz), 78.47, 68.59 (d, *J* = 3.78 Hz), 68.46 (d, *J* = 6.3 Hz), 60.22, 56.32 (d, *J* = 3.78 Hz), 38.01, 37.90, 31.16, 28.59, 21.23, 14.55. mp = 105 – 107^{o}C. HRMS (ESI TOF) m/z calculated for $C_{17}H_{25}FN_2O_4$: 363.1696 (M+Na), found: 363.1695 m/z (M+Na).



tert-butyl (2R)-1-(3-fluoro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **4b-D**. This compound was obtained following the same procedure for the compound above, **4b**. Purification *via* flash column chromatography afforded **4b-D** as a colourless oil in 95% yield. ^{1}H NMR (500 MHz, DMSO) δ_{ppm} = 7.82 (s, 1H), 7.34 – 7.11 (m, 5H), 6.95 (dd, *J* = 2.57, 8.35 Hz, 1H), 5.24 (d, *J* = 5.02 Hz, 1H), 4.40 – 4.06 (m, 3H), 3.68 (s, 1H), 3.25 – 3.14 (m, 2H), 2.95 – 2.89 (m, 1H), 2.83 – 2.63 (m, 1H), 1.25 (s, 9H). ^{13}C NMR (126 MHz, DMSO) δ_{ppm} = 171.22, 170.09, 157.77, 138.24 (d, *J* = 10.04 Hz), 129.64, 128.46, 126.57 (d, *J* = 1.24 Hz), 86.07 (d, *J* = 7.54 Hz), 84.75 (d, *J* = 8.8 Hz), 78.47, 68.51 (d, *J* = 3.74 Hz), 68.46 (d, *J* = 6.3 Hz), 60.22, 56.32 (d, *J* = 3.78 Hz), 38.01, 37.90, 31.16, 28.59, 21.23, 14.55. HRMS (ESI TOF) m/z calculated for $C_{17}H_{25}FN_2O_4$: 363.1696 (M+Na), found: 363.1694 m/z (M+Na).

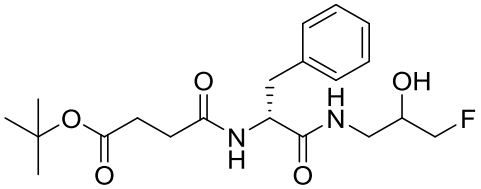


(2S)-2-amino-*N*-(3-fluoro-2-hydroxypropyl)-3-phenylpropanamide, **4d**. The N-Boc protected compound (0.640 g, 1.880 mmol) was dissolved in 5 mL of a 1:2 mixture of trifluoroacetic acid:dichloromethane and was allowed to stir at r.t. for 1 hr. The reaction mixture was then evaporated under reduced pressure, resuspended in dichloromethane and evaporated under reduced pressure again. This resuspension and evaporation was repeated 3 times alternating between dichloromethane and methanol. After the final resuspension and evaporation of dichloromethane, the crude material was allowed to sit under high vacuum for 1 hr and was used without further purification.



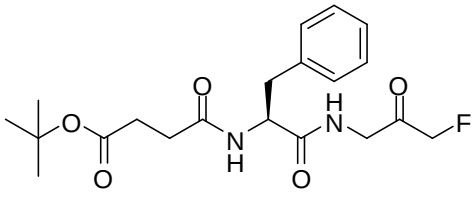
tert-butyl 4-((2S)-1-(3-fluoro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **5b**. In a

separate 100 mL round bottom flask, *tert*-butyl hydrogen butanediolate (0.327 g, 1.880 mmol) and N-methylmorpholine (0.200 mL, 1.880 mmol) was dissolved in 4 mL of *N*-methyl-2-pyrollidinone. HBTU coupling reagent (0.75 g, 1.970 mmol) was then added and the reaction mixture was allowed to stir at 0 ^{o}C for 15 min. **4d** free amine was resuspended in 1 mL *N*-methylpyrrolidinone in a separate 20 mL vial. N-methylmorpholine was added to **4d** until the solution was basic according to pH paper. The basic solution was then added drop wise to the reaction mixture stirring at 0 ^{o}C. The reaction was allowed to stir for an additional 30 min before adding N-methylmorpholine (0.400 mL, 3.760 mmol), removing the ice-water bath, and allowing the reaction to warm up to r.t. over night. The following morning, the reaction mixture was diluted with 50 mL EtOAc and washed with a 50:50 mixture of brine:water 3 times followed by just brine 3 times. The organic layer was collected, dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified *via* flash column chromatography (20% EtOAc/hexanes to 100% EtOAc) to give **5b** as a colourless oil in 81% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.32 – 7.25 (m, 3H), 7.22 – 7.16 (m, 2H), 6.89 (t, 0.5H, rotomer, J = 4.8 Hz), 6.81 (t, 0.5H, rotomer, J = 6 Hz), 6.10 (t, 1H, J = 7.2 Hz), 4.72 – 4.64 (m, 1H), 4.32 (dd, 1H, J = 5.2 Hz), 4.20 (dd, 1H, J = 5.2 Hz), 4.10 (q, 1H, J = 7.2 Hz), 3.94 – 3.78 (m, 1H), 3.52 – 3.34 (m, 1H), 3.26 – 3.12 (m, 2H), 3.10 – 3.04 (m, 1H), 2.73 – 2.64 (m, 1H), 2.49 – 2.41 (m, 1H), 2.37 – 2.29 (m, 2H), 1.4 (s, 4.5H, rotomer), 1.39 (s, 4.5H, rotomer). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 172.91 (d, J = 0.8 Hz), 172.64, 172.31, 172.06 (d, J = 4.4 Hz), 136.42, 129.25 (d, J = 1.9 Hz), 128.84 (d, J = 1.7 Hz), 127.19, 85.05 (d, J = 27.2 Hz), 83.36 (d, J = 27 Hz), 81.44, 54.46 (d, J = 8.9 Hz), 42.34 (t, J = 5.8 Hz), 37.44 (d, J = 15.4 Hz), 31.16 (d, J = 1.4 Hz), 30.58 (d, J = 1.8 Hz), 28.07 (d, J = 1.5 Hz). HRMS (ESI TOF) m/z calculated for $C_{20}H_{29}FN_2O_5$: 419.1958 m/z (M+Na), found: 419.1934 m/z (M+Na).

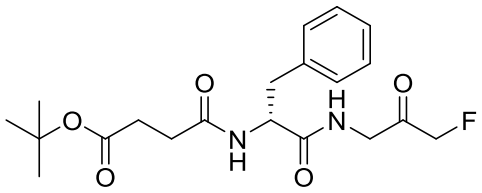


tert-butyl 4-(2R)-1-3-fluoro-2-hydroxypropylamino-1-oxo-3-phenylpropan-2-ylamino-4-oxobutanoate, **5b-D**. This compound was obtained in accordance with the procedure for the above compound, **5b**. Flash column chromatography afforded pure **5b-D** in 84% yield as a colourless oil. ^{1}H

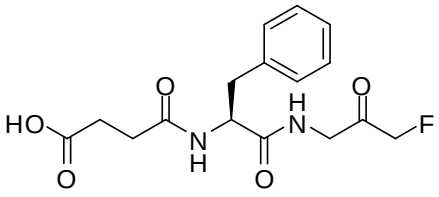
NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.39 – 7.27 (m, 3H), 7.22 – 7.15 (m, 2H), 6.92 (t, 0.5H, rotomer, J = 4.8 Hz), 6.79 (t, 0.5H, rotomer, J = 6 Hz), 6.15 (t, 1H, J = 7.2 Hz), 4.71 – 4.65 (m, 1H), 4.28 (dd, 1H, J = 5.2 Hz), 4.20 (dd, 1H, J = 5.2 Hz), 4.11 (q, 1H, J = 7.2 Hz), 3.95 – 3.79 (m, 1H), 3.50 – 3.32 (m, 1H), 3.26 – 3.13 (m, 2H), 3.11 – 3.01 (m, 1H), 2.73 – 2.64 (m, 1H), 2.49 – 2.41 (m, 1H), 2.37 – 2.29 (m, 2H), 1.4 (s, 4.5H, rotomer), 1.39 (s, 4.5H, rotomer). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 173.12 (d, J = 0.8 Hz), 172.58, 172.31, 172.06 (d, J = 4.4 Hz), 136.42, 129.25 (d, J = 1.9 Hz), 128.84 (d, J = 1.7 Hz), 127.19, 85.05 (d, J = 27.2 Hz), 83.36 (d, J = 27 Hz), 81.44, 54.46 (d, J = 8.9 Hz), 42.34 (t, J = 5.8 Hz), 37.44 (d, J = 15.4 Hz), 31.16 (d, J = 1.4 Hz), 30.58 (d, J = 1.8 Hz), 28.07 (d, J = 1.5 Hz). HRMS (ESI TOF) m/z calculated for $C_{20}H_{29}FN_2O_5$: 419.1958 m/z (M+Na), found: 419.1944 m/z (M+Na).



(S)-*tert*-butyl 4-(1-(3-fluoro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **6b**. The alcohol (1.00 g, 2.522 mmol) was dissolved in 12 mL CH_2Cl_2. Dess-Martin Periodinane (prepared from lit. ref.) (3.20 g, 7.566 mmol) was then added to the reaction mixture and allowed to stir overnight at r.t. The reaction mixtured turned opaque white and was diluted with 100 mL CH_2Cl_2. The solution was then washed with 2 x 50 mL sat. aqueous sodium thiosulfate solution followed by brine (2 x 50 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. Flash column chromatography (40% EtOAc – 100% EtOAc in hex) afforded pure compound in 70% yield as a colourless oil. ^1H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.36 – 7.23 (m, 5H), 7.09 (s, 0.2H, rotomer), 6.90 (s, 0.8H, rotomer), 6.18 (d, 0.8H, rotomer, J = 8 Hz), 6.08 (d, 0.2H, rotomer, J = 8 Hz), 5.01 (s, 0.2H, rotomer), 4.98 (s, 0.8H, rotomer), 4.9 (s, 0.2H, rotomer), 4.86 (s, 0.8H, rotomer), 4.79 (q, 1H, J = 8 Hz), 4.42 – 4.32 (m, 1H), 4.18 – 4.11 (m, 1H), 3.22 – 3.11 (m, 2H), 2.74 – 2.66 (m, 1H), 2.58 – 2.46 (m, 1H), 2.42 – 2.34 (m, 2H), 1.43 (s, 9H). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 172.70, 171.98, 171.34, 136.41, 129.23, 128.78, 127.13, 85.41, 83.58, 81.21, 54.17, 46.54 (d, J = 2.2 Hz), 37.46, 31.20, 30.61, 28.05. HRMS (ESI TOF) m/z calculated for $C_{20}H_{27}FN_2O_5$: 417.1802 m/z (M+Na), 433.1541 m/z (M+K), found: 417.1787 m/z (M+Na), found: 433.1535 m/z (M+K).

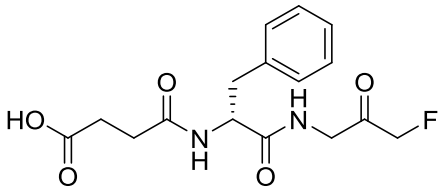


tert-butyl (R)-4-(1-(3-fluoro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **6b-D**. This compound was obtained in accordance with the procedure above for **6b**. The final compound **6b-D** was obtained after flash column chromatography to yield a colourless oil in 74% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.36 – 7.23 (m, 5H), 7.09 (s, 0.2H, rotomer), 6.90 (s, 0.8H, rotomer), 6.18 (d, 0.8H, rotomer, J = 8 Hz), 6.08 (d, 0.2H, rotomer, J = 8 Hz), 5.01 (s, 0.2H, rotomer), 4.98 (s, 0.8H, rotomer), 4.9 (s, 0.2H, rotomer), 4.86 (s, 0.8H, rotomer), 4.79 (q, 1H, J = 8 Hz), 4.42 – 4.32 (m, 1H), 4.18 – 4.11 (m, 1H), 3.22 – 3.11 (m, 2H), 2.74 – 2.66 (m, 1H), 2.58 – 2.46 (m, 1H), 2.42 – 2.34 (m, 2H), 1.43 (s, 9H). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 172.70, 171.98, 171.34, 136.41, 129.23, 128.78, 127.13, 85.41, 83.58, 81.21, 54.17, 46.54 (d, J = 2.2 Hz), 37.46, 31.20, 30.61, 28.05. HRMS (ESI TOF) m/z calculated for $C_{20}H_{27}FN_2O_5$: 417.1802 m/z (M+Na), 433.1541 m/z (M+K), found: 417.1788 m/z (M+Na), found: 433.1532 m/z (M+K).

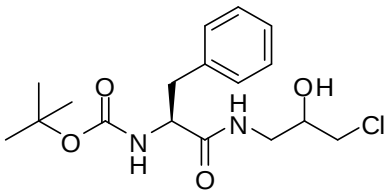


(S)-4-(1-(3-fluoro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoic acid, **7b**. *t*-butyl ester (0.241 g, 0.611 mmol) was dissolved in 3 mL CH_2Cl_2. 3 mL of trifluoroacetic acid was then added into the reaction mixture drop-wise at room temperature. The reaction was monitored by disappearance of starting material via TLC. The reaction was complete after 3 h and was then concentrated under reduced pressure. The crude residue was purified *via* prep-HPLC equipped with a Phenomenex 5μ C18 100Å 150 x 21.20 mm column and DAD monitoring 254nm. Mobile phase consisted of 30:70 ACN:H_2O for 1 min, up to 50:50 ACN:H_2O over the next 9 min, holding at 50:50 for 1 min, then returning to 30:70 ACN:H_2O for the last minute using a flow-rate of 20mL/min. Fractions containing compound were collected and concentrated under reduced pressure to remove organic solvents. Water was removed by lyophilisation overnight to afford compound as a fluffy white solid in 82% yield. ^{1}H NMR (400 MHz, acetone-D6) δ_{ppm} = 7.67 (br t, 1H), 7.53 (d,

1H, J = 6.4 Hz), 7.32 – 7.26 (m, 4H), 7.23 – 7.18 (m, 1H), 5.15 (s, 0.1H, rotomer), 5.12 (s, 0.9H, rotomer), 5.05 (s, 0.1H, rotomer), 5.03 (s, 0.9H, rotomer), 4.76 – 4.70 (m, 1H), 4.18 – 4.07 (m, 2H), 3.23 (dd, 1H, J = 4 Hz), 2.94 (dd, 1H, J = 6.8 Hz), 2.66 – 2.40 (m, 4H). ^{13}C NMR (126 MHz, Acetone-D6) δ_{ppm} = 201.24, 173.50, 171.69, 137.82, 129.26, 128.20, 126.39, 85.02, 83.59, 65.22, 54.41, 45.69, 37.37, 30.13, 14.71. mp = 120 – 124^{o}C. HRMS (ESI, TOF) m/z calculated for C16H19FN2O5: 361.1176 m/z (M+Na), found: 361.1174 m/z (M+Na) = -2.0^{o} (c = 5x10^{-3} g/mL, EtOH).

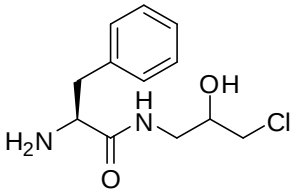


(R)-4-(1-(3-fluoro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoic acid, **7b-D**. This compound was obtained in accordance to the procedure above for **7b**. Compound **7b-D** was isolated *via* prep-HPLC with conditions similar to that for compound **7b**. Subsequent lyophilisation afforded pure **7b-D** in 75% yield as a fluffy white solid. ^{1}H NMR (400 MHz, Acetone-D6) δ_{ppm} = 7.64 (br t, 1H), 7.51 (d, 1H, J = 6.4 Hz), 7.28 – 7.22 (m, 4H), 7.20 – 7.18 (m, 1H), 5.10 (s, 1H), 4.98 (s, 1H), 4.75 – 4.65 (m, 1H), 4.15 – 4.05 (m, 2H), 3.25 - 3.15 (m, 1H), 3.00 – 2.85 (m, 1H), 2.60 – 2.35 (m, 2H). ^{13}C NMR (126 MHz, Acetone-D6) δ_{ppm} = 201.25, 172.50, 171.55, 135.89, 130.26, 128.24, 126.40, 85.00, 83.49, 65.12, 55.41, 45.71, 38.38, 31.13, 15.74. mp = 122 – 126^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{16}H_{19}FN_2O_5$: 361.1176 m/z (M+Na), found: 361.1166 m/z (M+Na) = +2.5^{o} (c = 5x10^{-3} g/mL, EtOH).

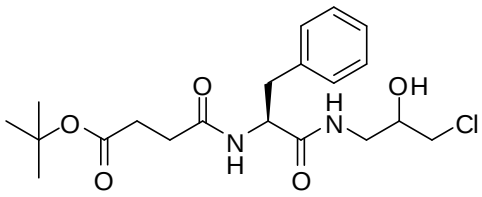


tert-butyl (2-S)-1-(3-chloro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **4a**. A 250 mL round bottom flask was charged with HBTU (7.500 g, 19.788 mmol) and *N*-boc-protected phenylalanine (5.000 g, 18.846 mmol) dissolved in 80 mL *N*-methylpyrrolidinone. This mixture was allowed to stir in an ice-water bath for 15 min. Aminochloropropanol **3a** was dissolved in 1 mL NMP in a separate vial and then added to the reaction mixture drop-wise. After an additional 15

min of stirring, N-methylmorpholine (NMM, 6.10 mL, 55.5 mmol) was added. Upon complete addition of NMM, the ice-water bath was removed and the reaction mixture was allowed to stir at r.t. overnight. The following morning the reaction was diluted with 100 mL EtOAc and washed with 50:50 brine:H_2O (3 x 30 mL), 1 M aqueous HCl (3 x 20 mL), then brine (3 x 30 mL). The organic phase was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography using EtOAc/hexanes to afford a colourless oil in 60 % yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.42 – 7.15 (m, 5H), 6.40 (m, 1H), 5.09 (s, 1H), 4.35 (s, 1H), 3.94 – 3.71 (m, 1H), 3.59 – 3.20 (m, 4H), 3.08 (m, 2H), 1.43 (d, *J* = 1.2 Hz, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ_{ppm} = 172.98, 171.17, 136.38, 129.26, 128.80, 127.14, 70.84, 60.40, 46.20 (d, *J* = 8.4 Hz), 43.04 (d, *J* = 8.8 Hz), 38.37 (d, *J* = 11.4 Hz), 28.24 (d, *J* = 5.1 Hz), 24.83, 21.05, 14.19. mp = 98 – 105^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{17}H_{25}ClN_2O_4$: 379.1401 m/z (M+Na), found: 379.1398 m/z (M+Na).

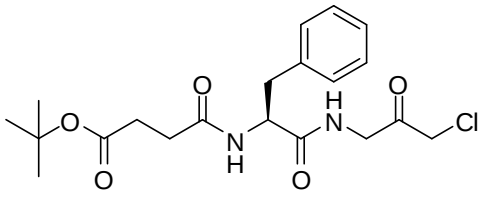


(2S)-2-amino-N-(3-chloro-2-hydroxypropyl)-3-phenylpropanamide, **4c**. **4a** was deprotected using a mixture of TFA/CH_2Cl_2 as described above for **4d**, and was used without further purification.

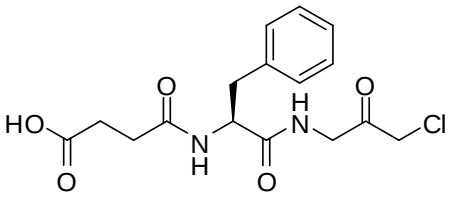


tert-butyl 4-(2-*S*)-1-(3-chloro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylamino-4-oxobutanoate, **5a**. Conditions are the same as for compound **5b** detailed above. Flash column chromatography (30% EtOAc – 100% EtOAc gradient in hexanes) afforded **5a** as a colourless oil in 72% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.28 – 7.12 (m, 5H), 6.84 (t, 0.5H, rotomer, J = 4 Hz), 6.74 (t, 0.5H, rotomer, J = 4 Hz), 5.98 (t, 1H, J = 8 Hz), 3.83 – 3.72 (m, 1H), 3.52 – 3.38 (m, 1H), 3.36 – 3.28 (m, 2H), 3.25 – 3.10 (m, 2H), 3.05 – 2.97 (m, 1H), 2.71 – 2.61 (m, 1H), 2.45 – 2.35 (m, 1H), 2.33 – 2.20 (m, 2H), 1.36 (d, 9H, J = 4 Hz). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 172.97, 136.40, 136.32, 129.26, 129.23, 128.90, 127.28, 127.24,

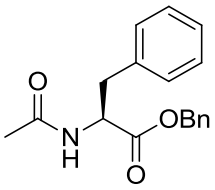
71.07, 71.03, 54.43, 54.28, 46.21, 45.82, 43.50, 43.35, 38.61, 37.44, 37.30, 31.23, 30.60, 28.10, 28.09. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{29}ClN_2O_5$: 435.1663 m/z (M+Na), 451.1402 m/z (M+K), found: 435.1647 m/z (M+Na), found: 451.1389 m/z (M+K).



(*S*)-*tert*-butyl 4-(1-(3-chloro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **6a**. The chloro-alcohol (0.148 g, 0.358 mmol) was dissolved in 3 mL CH_2Cl_2 in a 25 mL round bottom flask. Dess-Martin Periodinane (0.456 g, 1.075 mmol) was then added into the reaction solution and a pale yellow colour started to form. The reaction mixture was allowed to stir at r.t. overnight. Complete conversion of starting material was monitored via TLC. The reaction mixture was then filtered and diluted with 20 mL CH_2Cl_2 and washed with a concentrated aqueous solution of sodium thiosulfate (3 x 5 mL). In some cases an emulsion would form. This was simply treated with addition of brine to the extraction mixture and the emulsion separated. The organic layer was then washed with 1 M $NaHCO_3$ (3 x 5 mL) and 1 M HCl (3 x 5 mL), brine (3 x 5 mL), and dried over Na_2SO_4. The crude residue was purified via flash column chromatography (55% EtOAc in hexanes) to afford pure compound in 60% yield as a white foam. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.37 – 7.23 (m, 5H), 7.02 (s, 1H), 6.13 (d, 1H, J = 8 Hz), 4.78 (q, 1H, J = 8 Hz), 4.34 (dd, 1H, J_{AB} = 19.2 Hz, J_{AX} = 5.6 Hz), 4.15 (dd, 1H, J_{AB} = 18.8 Hz, J_{BX} = 5.2 Hz), 4.13 (s, 2H), 3.18 (m, 2H), 2.77 – 2.67 (m, 1H), 2.55 – 2.46 (m, 1H), 2.40 – 2.32 (m, 2H), 1.43 (s, 9H). ^{13}C NMR (100 MHz, $CDCl_3$) δ_{ppm} = 198.06, 172.76, 172.03, 171.43, 136.38, 129.22, 128.80, 127.15, 81.28, 54.14, 47.18, 46.32, 37.36, 31.20, 30.61, 28.06. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{27}ClN_2O_5$: 433.1506 m/z (M+Na), 449.1246 m/z (M+K), found: 433.1495 m/z (M+Na), found: 449.1235 m/z (M+K).

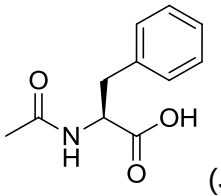


(*S*)-4-(1-(3-chloro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoic acid, **7a**. Hydrolysis of the *tert*-butyl ester to form the free acid was achieved as described above for compound **7b**. The crude residue was titriated with toluene to induce precipitate formation. Pure **7a** was afford after filtration as a white solid in 75% yield. ^{1}H NMR (400 MHz, Acetone-D6) δ_{ppm} = 7.75 (br t, 1H), 7.54 (d, 1H, J = 7.6 Hz), 7.31 – 7.26 (m, 4H), 7.24 – 7.18 (m, 1H), 4.74 – 4.68 (m, 1H), 4.42 (s, 2H), 4.18 (dd, 1H, J_{AB} = 18.4 Hz, J_{AX} = 5.6 Hz), 4.12 (dd, 1H, J_{AB} = 18.4 Hz, J_{BX} = 5.2 Hz), 3.23 (dd, 1H, J = 5.2 Hz), 2.95 (dd, 1H, J = 8.8 Hz), 2.64 – 2.40 (m, 4H). ^{13}C NMR (100 MHz, Acetone-D6) δ_{ppm} = 197.90, 173.46, 171.67, 171.65, 137.83, 129.25, 128.21, 126.40, 54.46, 54.43 (rotomer), 46.95, 46.89 (rotomer), 46.85 (rotomer), 37.31, 37.27 (rotomer), 30.15, 30.10 (rotomer), 28.79. mp = 109 – 113^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{16}H_{19}ClN_2O_5$: 377.0880 m/z (M+Na), found: 377.0866 m/z (M+Na) = -10.0^{o} (c 2.1x10^{-3} g/mL, MeOH).

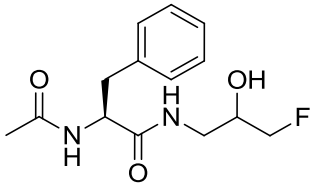


(*S*)-benzyl 2-acetamido-3-phenylpropanoate, **18**. (S)-1-(benzyloxy)-1-oxo-3-phenylpropan-2-aminium *p*-toluene sulfonate (0.100 g, 0.233 mmol) was dissolved in CH2Cl2 (2 mL). To this solution, acetic anhydride (0.044 mL, 0.467 mmol) and then diisopropylethylamine (0.122 mL, 0.701 mmol) were added and the reaction mixture was allowed to stir at r.t. until the starting material was completely consumed as monitored by TLC. After the reaction was complete the mixture was concentrated under reduced pressure. The crude residue was purified by flash column chromatography (60% EtOAc in hexanes) to give **18** as a colourless oil in 95% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.38 – 7.31 (m, 3H), 7.31 – 7.26 (m, 2H), 7.23 – 7.17 (m, 3H), 6.99 (dt, *J* = 4.4, 3.3 Hz, 2H), 6.00 (d, *J* = 7.3 Hz, 1H), 5.12 (q, *J* = 12.1 Hz, 2H), 4.91 (dt, *J* = 7.9, 5.8 Hz, 1H), 3.19 – 2.99 (m, 2H), 1.95 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ_{ppm} = 171.52, 169.64, 129.31, 128.63, 128.59, 128.56, 128.55, 127.07, 67.28,

53.17, 37.82, 23.12. HRMS (ESI, TOF) m/z calculated for $C_{18}H_{19}NO_3$: 320.1263 m/z (M+Na), found: 320.1204 m/z (M+Na).

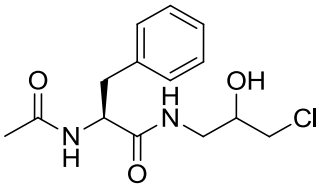


(*S*)-2-acetamido-3-phenylpropanoic acid, **19**. A 250 mL round bottom flask was charged with benzyl-ester **18** (0.660 g, 2.219 mmol) and dissolved in ethanol (10 mL). The reaction flask was vacuum purged with nitrogen and 10% Platinum on activated charcoal (0.150 g) was then added portion-wise. The reaction flask was then sealed with a rubber septum and purged with a hydrogen filled balloon. A positive pressure of hydrogen was supplied by a balloon and the reaction was allowed to stir at r.t. overnight. Complete reaction was monitored by disappearance of starting material *via* TLC and the reaction mixture was then filtered through a celite pad and concentrated under reduced pressure to afford a colourless oil that was used without further purification.

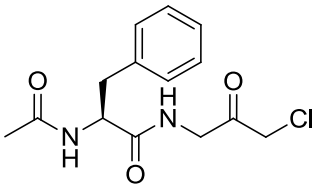


(2*S*)-2-acetamido-N-(3-fluoro-2-hydroxypropyl)-3-phenylpropanamide, **20**. A 250 mL round bottom flask was charged with N-acyl-L-phenylalanine (0.45 g, 2.10 mmol) and dissolved in freshly degassed DMF (10 mL). HBTU (0.86 g, 2.28 mmol) was then added and the reaction was placed in an ice-water bath and allowed to stir at 0^{o}C for 15 min. In a separate vial, aminofluoropropanol **3b** (0.200 g, 2.28 mmol) was dissolved in 2 mL freshly degassed DMF and was added drop-wise to the stirring solution. After 15 min, NMM (0.70 mL, 6.50 mmol) was added to the reaction mixture and the ice-bath was removed. The reaction was allowed to stir at r.t. overnight. The reaction was then diluted with EtOAc (50 mL) and washed with a mixture of brine/H_2O (50:50, 5 x 20 mL), saturated $NaHCO_3$ (3 x 20 mL), 1 M HCl (2 x 20 mL), and brine (3 x 20 mL). The organic layer was collected and dried over Na_2SO_4 and filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography (5% CH_2Cl_2 in MeOH) to afford the product in 52% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 7.61 – 7.53 (m, 1H), 7.40 – 7.13 (m, 5H), 4.79 – 4.71 (m, 1H), 4.58 – 4.51(m, 1H), 4.49 – 4.41 (m, 1H), 4.27 – 4.16 (m, 1H), 3.61 – 3.12 (m, 4H), 3.08 – 2.98 (m, 2H), 2.00 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ_{ppm} = 173.82, 173.71 (rotomer), 172.34, 172.31 (rotomer), 136.47, 136.39, 129.24, 129.21, 128.59, 127.05, 85.29, 85.12, 85.07, 83.59, 83.43, 83.38, 38.68, 34.07, 33.41, 32.82, 26.32, 26.13, 26.00, 22.89. HRMS (ESI, TOF) m/z calculated for $C_{14}H_{19}FN_2O_3$: 282.1380 m/z (M+), 305.1277 m/z (M+Na), found: 283.1422 m/z (M+), found: 305.1259 m/z (M+Na).

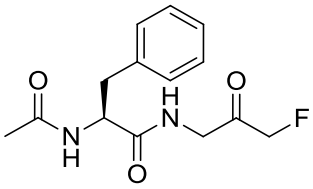


(2*S*)-2-acetamido-N-(3-chloro-2-hydroxypropyl)-3-phenylpropanamide, **21**. Similar to the procedure above, a 100 mL round bottom flask was charged with N-acyl-L-phenylalanine (**19**) (0.200 g, 0.965 mmol) and was dissolved in freshly degassed DMF (5 mL). HBTU (0.4 g, 1.00 mmol) was then added and the mixture was placed in an ice-water bath for 15 min. The aminochloropropanol **3a** (0.10 g, 1.00 mmol) was dissolved in 2 mL freshly degassed DMF in a separate vial and added drop-wise to the stirring solution at 0^{o}C. After an additional 15 min, NMM (0.320 mL, 2.90 mmol) was added and the ice-water bath was removed. The reaction was allowed to stir overnight at r.t. The work up procedure and purification was the same as the above for **20**, which afforded the desired product **21** in 43% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 7.33 (m, 3H), 7.25 (m, 2H), 6.69 (m, 1H), 6.44 (m, 1H), 4.70 (m, 1H), 3.84 (m, 1H), 3.40 (m, 6H), 3.06 (m, 2H), 2.01 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 172.24, 172.21 (rotomer), 170.45, 170.37 (rotomer), 136.37, 136.25, 129.23, 129.19, 128.77, 128.75, 127.21, 127.19, 70.35, 54.90, 46.26, 43.00, 38.65, 23.14. HRMS (ESI, TOF) m/z calculated for $C_{14}H_{19}ClN_2O_3$: 321.0982 m/z (M+Na), found: 321.0969 m/z (M+Na).

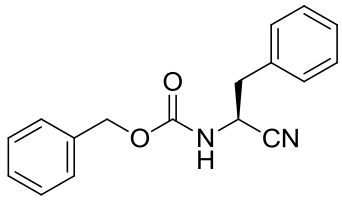


(*S*)-2-acetamido-N-(3-chloro-2-oxopropyl)-3-phenylpropanamide, **22**. The secondary alcohol **21** (0.070 g, 0.24 mmol) was oxidized

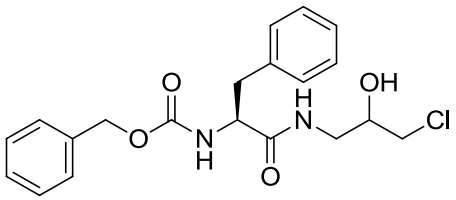
using Dess-Martin Periodinane (0.40 g, 0.950 mmol) in accordance to the general oxidation procedure. The crude residue was purified by flash column chromatography using 4% MeOH in CH_2Cl_2 to afford a white solid in 54% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 8.33 (s, 1H), 7.33 – 7.25 (m, 5H), 6.69 (s, 1H), 4.90 (m, 1H), 4.70 (s, 2H), 3.84 (s, 2H), 3.40 (dd, 1H), 3.06 (dd, 1H), 2.01 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 187.23, 172.24, 172.21 (rotomer), 170.45, 170.37 (rotomer), 136.37, 136.25, 129.23, 129.19, 128.77, 128.75, 127.21, 127.19, 54.90, 46.26, 43.00, 38.65, 23.14.HRMS (ESI TOF) m/z calculated for $C_{14}H_{17}ClN_2O_3$: 319.0825 m/z (M+Na), found: 319.0813 m/z (M+Na).



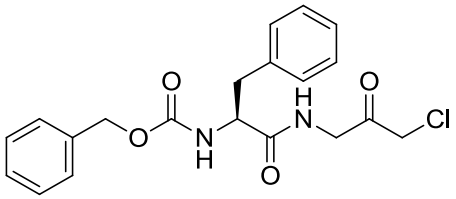
(*S*)-2-acetamido-N-(3-fluoro-2-oxopropyl)-3-phenylpropanamide, **23**. The secondary alcohol **20** (0.100 g, 0.356 mmol) was oxidized using Dess-Martin Periodinane (0.23 g, 0.54 mmol) in accordance to the general oxidation procedure. The crude residue was purified by flash column chromatography using 5% MeOH in CH2Cl2 to afford a pale yellow oil in 35 % yield. ^{1}H NMR (500 MHz, DMSO) δ_{ppm} = 8.23 (m, 1H), 8.04 (m, 1H), 7.34 – 7.14 (m, 5H), 5.17 (s, 1H), 5.07 (s, 1H), 4.36 – 4.28 (m, 1H), 4.19 (s, 2H), 4.03 – 3.92 (m, 2H), 1.88 – 1.82 (m, 2H). HRMS (ESI TOF) m/z calculated for $C_{14}H_{17}FN_2O_3$: 303.1121 m/z (M+Na), found: 303.1108 m/z (M+Na).



(*S*)-benzyl 1-cyano-2-phenylethylcarbamate, **24**. This compound was obtained according to a literature procedure.[69] All spectroscopic data match the reported data.

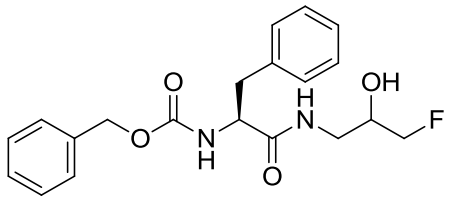


benzyl (2*S*)-1-(3-chloro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **25**. A 100 mL round bottom flask was charged with N-Cbz-L-phenylalanine (0.300 g, 1.00 mmol) and dissolved in freshly degassed DMF (10 mL). HBTU (0.400 g, 1.00 mmol) was then added to the reaction mixture which was stirring at 0°C. In a separate vial, the aminochloropropanol (0.120 g, 1.00 mmol) was dissolved in freshly degassed DMF (2 mL) and added drop-wise to the stirring solution. After 15 min, NMM (0.330 mL, 3.00 mmol) was added and the ice-water bath was removed. The reaction mixture was allowed to stir at r.t. overnight. The following morning the reaction was diluted with EtOAc (50 mL) and washed with a 50:50 mixture of brine:H_2O (3 x 20 mL), 1M HCl (2 x 20 mL), then brine (3 x 20 mL). The organic layer was dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford a white solid in 80% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 7.41 – 7.32 (m, 8H), 7.23 – 7.20 (m, 2H), 6.14 (s, 1H), 5.32 (s, 1H), 4.41 (s, 1H), 3.80 (d, *J* = 40.7 Hz, 1H), 3.59 – 2.99 (m, 8H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 172.44, 172.35 (rotomer), 135.92, 129.26, 129.24, 128.89, 128.86, 128.60, 128.36, 128.17, 128.15, 127.29, 70.74, 70.50 (rotomer), 67.31, 56.51, 46.32 (rotomer), 46.29, 43.05 (rotomer), 42.98, 1.03. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{23}ClN_2O_4$: 390.1346 m/z (M+), 413.1244 m/z (M+Na), 429.0983 m/z (M+K), found: 391.1316 m/z (M+), found: 413.1235 m/z (M+Na), found: 429.0977 m/z (M+K).

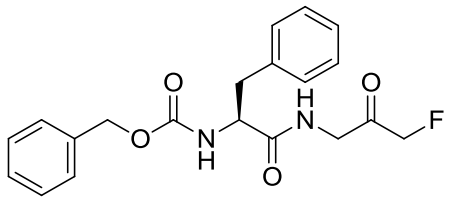


(*S*)-benzyl 1-(3-chloro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **26**. The secondary alcohol (**25**) (0.110 g, 0.26 mmol) was oxidized using Dess-Martin Periodinane (0.45 g, 1.060 mmol) in accordance to the general oxidation procedure. The crude material was purified by flash column chromatography (50% EtOAc in hexanes) to afford an off-white solid in 46% yield. ^{1}H

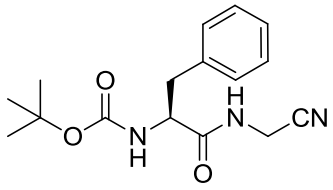
NMR (500 MHz, $CDCl_3$) δ_{ppm} = 7.41 – 7.30 (m, 8H), 7.21 (d, *J* = 7.1 Hz, 2H), 6.51 (s, 1H), 5.32 (s, 1H), 5.11 (s, 2H), 4.50 (s, 1H), 4.39 – 4.20 (m, 2H), 4.10 (s, 2H), 3.12 (d, *J* = 6.7 Hz, 2H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 197.87, 171.26, 136.08, 135.98, 129.22, 128.82, 128.58, 128.29, 128.09, 127.22, 67.25, 56.16, 47.12, 46.08, 38.28. mp = 94 - 96°C. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{21}ClN_2O_4$: 411.1088 m/z (M+Na), 427.0827 m/z (M+K), found: 411.1068 m/z (M+Na), found: 427.0800 m/z (M+K).



benzyl (2*S*)-1-(3-fluoro-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **27**. A 100 mL round bottom flask was charged with *N*-Cbz-*L*-phenylalanine (0.200 g, 0.668 mmol) and dissolved in anhydrous NMP (8 mL). HBTU (0.270 g, 0.700 mmol) was then added and the reaction mixture was placed in an ice-water bath and allowed to stir for 15 min. In a separate vial, the aminofluoropropanol (0.060 g, 0.700 mmol) was dissolved in 2 mL NMP before addition to the reaction mixture. After the aminofluoropropanol was added, the reaction was allowed to stir for an additional 15 min at 0°C before NMM (0.220 mL, 2.00 mmol) was added. The reaction was then allowed to stir at r.t. overnight. The following morning the reaction was diluted with EtOAc (50 mL) and washed with a 50:50 mixture of brine:H_2O (4 x 20 mL), then brine (3 x 20 mL). The organic layer was dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified by flash column chromatography (30% EtOAc – 100% EtOAc in hexanes) to afford an off-white solid in 92% yield. ^1H NMR (500 MHz, DMSO) δ_{ppm} = 8.09 (s, 1H), 7.52 (dd, *J* = 8.4, 5.0 Hz, 1H), 7.40 – 7.14 (m, 10H), 5.25 (dd, *J* = 5.0, 3.4 Hz, 1H), 4.95 (s, 2H), 4.39 – 4.15 (m, 3H), 3.83 – 3.54 (m, 1H), 3.23 – 3.03 (m, 2H), 3.05 – 2.88 (m, 1H), 2.81 – 2.69 (m, 1H). HRMS (ESI, TOF) m/z calculated for $C_{20}H_{23}FN_2O_4$: 374.1642 m/z (M+), 392.1986 m/z (M+NH4), found: 375.1744 m/z (M+H), found: 392.1996 m/z (M+NH4).

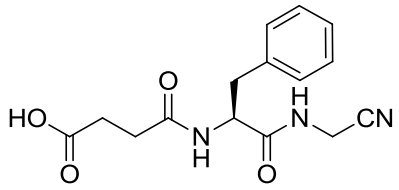


(*S*)-benzyl 1-(3-fluoro-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **28**. The secondary alcohol **27** (0.220 g, 0.60 mmol) was oxidized according to the general oxidation procedure using Dess-Martin Periodinane (0.380 g, 0.90 mmol). The crude residue was purified by flash column chromatography (30% EtOAc – 100% EtOAc in hexanes) to afford an off-white solid in 76% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 7.40 – 7.29 (m, 7H), 7.28 – 7.25 (m, 1H), 7.21 (d, *J* = 7.2 Hz, 2H), 6.48 (s, 1H), 5.32 (s, 1H), 5.11 (s, 2H), 4.95 (s, 1H), 4.85 (s, 1H), 4.51 (s, 1H), 4.29 (dd, *J* = 47.0, 19.1 Hz, 2H), 3.12 (d, *J* = 6.7 Hz, 2H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 201.61, 201.43, 171.21, 136.12, 136.02, 129.23, 128.79, 128.57, 128.27, 128.07, 127.18, 85.14, 83.68, 67.21, 56.16, 46.54, 38.35. mp = 123 – 125^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{21}FN_2O_4$: 372.1485 m/z (M+), 395.1383 m/z (M+Na), found: 373.1561 m/z (M+H), found: 395.1377 m/z (M+Na).

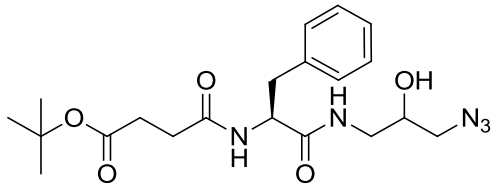


(*S*)-*tert*-butyl 1-(cyanomethylamino)-1-oxo-3-phenylpropan-2-ylcarbamate, **16**. To a solution of Boc-L-phenylalanine (2.650 g, 10.0 mmol), EDC-HCl (3.00 g, 15.00 mmol), HOBt monohydrate (1.840 g, 12.0 mmol) in freshly distilled CH_2Cl_2 (40 mL), N-methylmorpholine (4.40 mL, 40.0 mmol) and aminoacetonitrile hydrochloride (1.120 g, 12.10 mmol) was added. The resulting reaction mixture was allowed to stir at r.t. for 24 hrs. After which, the reaction mixture was concentrated under reduced pressure and any excess base was removed *in vacuo*. The residue was resuspended in CH_2Cl_2 (50 mL) and washed with aqueous HCl (0.1 M, 2 x 30 mL), saturated sodium bicarbonate (2 x 30 mL), and brine (2 x 30 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was recrystallized from EtOAc and hexanes to afford a white solid in 53% yield. 1H NMR (400 MHz, CDCl3) δppm = 7.26 (m, 5H), 6.37 (s, 1H), 4.82 (s, 1H), 4.35 (q, *J* = 8 Hz, 1H), 4.06 (dd, *J* = 8 Hz, 16 Hz, 2H), 3.09 (d, *J* = 4 Hz, 2H), 1.42 (s,

9H). HRMS (ESI, TOF) m/z calculated for $C_{16}H_{21}N_3O_3$: 303.1583 m/z (M+), found: 304.1651 m/z (M+H). mp = 138 – 143°C. IR (film) = 2932, 2253, 1604 cm^{-1}.

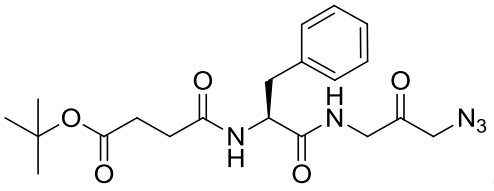


(*S*)-4-(1-(cyanomethylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoic acid, **17**. The previous compound **16** (0.100 g, 0.329 mmol) was dissolved in a 50:50 mixture of TFA:CH_2Cl_2 (1 mL) and allowed to stir at r.t. for 30 min. The reaction mixture was then concentrated under reduced pressure and placed under high vacuum for 1 hr. The resulting oil was resuspended in anhydrous chloroform (1 mL) and stirred with triethylamine (0.090 mL, 0.640 mmol) and succinic anhydride (0.035 g, 0.35 mmol) at r.t. for 24 hrs. The reaction mixture was then diluted with CH_2Cl_2 (10 mL), washed with aqueous HCl (0.1 M, 2 x 15 mL) then brine (2 x 15 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography (10% MeOH in CH_2Cl_2) to afford a white solid in 44% yield. ^{1}H NMR (400 MHz, MeOD) δ_{ppm} = 7.24 (m, 5H), 4.57 (dd, *J* = 8 Hz, 12 Hz, 1H), 4.11 (s, 2H), 3.19 (dd, *J* = 4 Hz, 16 Hz, 1H), 2.90 (dd, *J* = 12 Hz, 16 Hz, 1H), 2.51 (m, 4H). HRMS (ESI, TOF) m/z calculated for $C_{15}H_{17}N_3O_4$: 303.1219 m/z (M+), found: 304.1279 m/z (M+H). mp = 148 – 149°C. IR (film) = 2932, 1707, 1660 cm^{-1}.

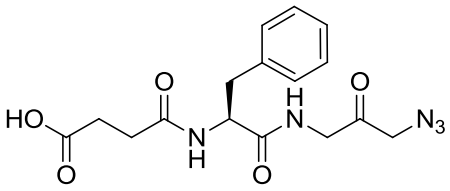


tert-butyl 4-((2*S*)-1-(3-azido-2-hydroxypropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **11**. Free acid compound **10** (0.300 g, 0.930 mmol), DCC (0.190 g, 0.930 mool) and HOBt monohydrate (0.140 g, 0.930 mmol) were dissolved in freshly degassed DMF (10 mL). Free amine compound **9** (0.120 g, 1.020 mmol) was dissolved in a separate vial and added to the reaction mixture drop-wise. The reaction mixture was allowed to stir at r.t. overnight. The reaction was then diluted with EtOAc (50 mL) and washed with a 50:50 mixture of brine:H_2O (4 x 25 mL), 1 M $NaHCO_3$ (3 x 20 mL), 1M HCl (3 x 20 mL), then

brine (3 x 20 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography (60% EtOAc in hexanes) to afford **11** as a white solid 87% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δppm = 7.42 – 7.15 (m, 5H), 6.93 (d, *J* = 29.5 Hz, 1H), 6.21 (s, 1H), 4.82 – 4.64 (m, 1H), 3.90 – 3.73 (m, 1H), 3.50 – 3.33 (m, 1H), 3.26 – 3.15 (m, 4H), 3.15 – 3.06 (m, 1H), 2.81 – 2.65 (m, 1H), 2.57 – 2.45 (m, 1H), 2.45 – 2.27 (m, 2H), 1.45 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δppm = 200.11, 172.83, 171.50, 136.23, 129.42, 128.95, 127.24, 110.10, 81.42, 55.84, 54.09, 47.27, 37.82, 31.22, 30.55, 28.01. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{29}N_5O_5$: 419.2169 m/z (M+), found: 420.2242 m/z (M+H).



(*S*)-*tert*-butyl 4-(1-(3-azido-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoate, **12**. The secondary alcohol **11** (0.330 g, 0.770 mmol) was oxidized using Dess-Martin Periodinane (0.800 g, 1.90 mmol) in accordance to the general oxidation procedure. The crude residue was purified by flash column chromatography (60% EtOAc in hexanes) to afford **12** as a white solid in 87% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δppm = 7.36 – 7.29 (m, 2H), 7.23 (d, *J* = 7.0 Hz, 2H), 7.13 (s, 1H), 6.12 (d, *J* = 7.8 Hz, 1H), 4.78 (dd, *J* = 14.5, 6.8 Hz, 1H), 4.15 (dt, *J* = 21.4, 6.9 Hz, 1H), 4.08 – 3.94 (m, 3H), 3.25 – 3.11 (m, 2H), 2.79 – 2.68 (m, 1H), 2.56 – 2.46 (m, 1H), 2.42 – 2.31 (m, 2H), 1.42 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δppm = 200.10, 172.93, 172.08, 171.47, 136.32, 129.22, 128.85, 127.22, 81.39, 55.79, 54.11, 47.37, 37.18, 31.20, 30.59, 28.06. HRMS (ESI, TOF) m/z calculated for $C_{20}H_{27}N_5O_5$: 417.2012 m/z (M+), found: 418.2093 m/z (M+H).



(*S*)-4-(1-(3-azido-2-oxopropylamino)-1-oxo-3-phenylpropan-2-ylamino)-4-oxobutanoic acid, **13**. tert-butyl-ester **12** (0.30 g, 0.720 mmol) was dissolved in CH_2Cl_2 (2 mL) and trifluoroacetic acid (1 mL) and allowed to stir

at r..t for 1 h. Upon complete consumption of starting material (*via* TLC) the reaction mixture was concentrated under reduced pressure. The resulting residue was resuspended in methanol and concentrated, resuspended again in CH_2Cl_2 and concentrated under reduced pressure. After a few rounds of resuspension and concentration, a white solid was obtained in quantitative yield. The product was used without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δppm = 7.36 – 7.29 (m, 3H), 7.27 – 7.16 (m, 2H), 4.87 (d, *J* = 7.7 Hz, 1H), 4.25 – 4.16 (m, 1H), 4.13 – 4.04 (m, 1H), 4.01 (s, 2H), 3.17 – 3.01 (m, 2H), 2.82 – 2.71 (m, 1H), 2.71 – 2.60 (m, 1H), 2.60 – 2.52 (m, 2H). ^{13}C NMR (101 MHz, $CDCl_3$) δppm = 199.49, 177.28, 173.43, 172.65, 159.74, 159.32, 135.20, 129.06, 128.92, 127.54, 116.02, 113.19, 55.85, 54.75, 47.35, 37.82, 30.31, 29.01. HRMS (ESI, TOF) m/z calculated for $C_{16}H_{19}N_5O_5$: 361.1386 m/z (M+), found: 362.1458 m/z (M+H).

Chapter 4. Irreversible Fluorescent Inhibitors

4.1. Research Objectives

Over the years, researchers have been putting more and more emphasis on generating cost-effective, sensitive and high-throughput assays for the screening of pharmaceutically relevant targets. There has also been increasing acknowledgement that isolated enzyme assays may not accurately reflect *in vivo* performance in the context of a whole cell where many other biological pathways, intracellular barriers and materials are present. Therefore, the use of cellular assays has increased as researchers strive to study compounds in a more biologically relevant context.[63]

Fluorescent probes have been used to study cellular processes for almost a century due to their unique sensitivity and selectivity. This sensitivity arises from the fact that one fluorophore molecule can be repeatedly excited and detected. Unless the fluorophore is destroyed in the excited state, which is possible through photobleaching, the same fluorophore can produce thousands of detectable photons. The alternative has been to monitor specific internal substrates to follow cellular target inhibition, but with proteases this is often not possible. In our case we can, in principle, follow LC3B-II build-up, but there is no guarantee that other factors do not interfere with this.

One of the drawbacks of a fluorescent assay is that compounds that are inherently fluorescent or fluorescence quenchers will complicate the results and interfere with the fluorescent readout of the assay.

The main research objective and focus of this chapter was to further develop the irreversible inhibitors into fluorescent compounds that could bind ATG4B in cells to provide visual localization. These fluorescent inhibitors could potentially be used in a type of enzyme-occupancy cell assay. A potent fluorescent inhibitor could bind ATG4B in cells as a so-called ‘label’. Competitive inhibitors could then be introduced to compete

with the fluorescent compounds. Ideally, the probe would be selective for the target protein, but if not (usually the case), techniques such as western blotting or immunoprecipitation could be used to evaluate the target even if other proteins were labelled. A suitable potent cell permeable competitive inhibitor would compete for any 'labelling' of ATG4B within cells.

The results from the initial HTS screen were interesting in the sense that a positive hit was found for a compound very similar to chloromethyl ketone **7a**. The hit, *Z*-*L*-Phe-CMK, compound **40**, is structurally interesting. Depending on how certain bonds are rotated, the Cbz (carboxybenzyl-) protecting group could be binding in the phenylalanine binding pocket, which would suggest that substitution at the α-carbon of the chloromethyl ketone is tolerated. However, it could also mean that *N*-substitution is also well tolerated. It was discovered, in studies by Dr. Kumar, that longer and longer *N*-capping acid chains, up to suberic acid, were optimal for substrate activity. This could imply that hydrophobic interactions at the *N*-terminus could be exploited for increasing potency. From this information, it was decided that the most facile and efficient method of introducing fluorescence would be to append a fluorophore directly onto the molecule of interest and the easiest way to do that would be to substitute the *N*-hemisuccinic acid of compound **7a/b** with an *N*-dansyl fluorescent group. It was also proposed to synthesize truncated versions of the fluorescent inhibitors that more closely resemble **40**. Figure 24 represents the different possible orientations compound **40**, *Z*-*L*-Phe-CMK, can adopt, overlayed on compound **7a**.

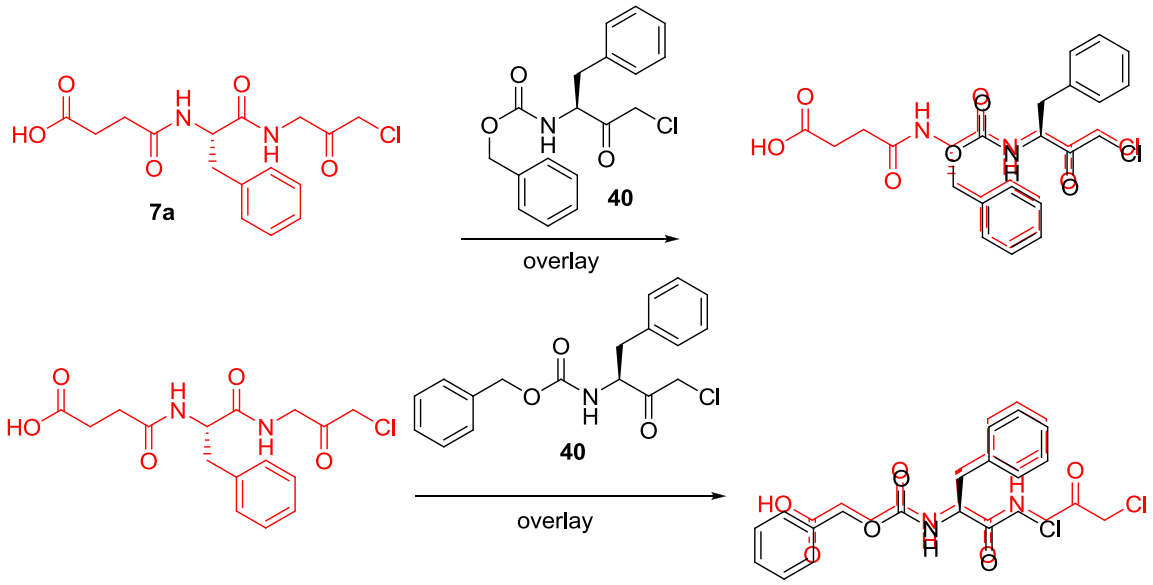
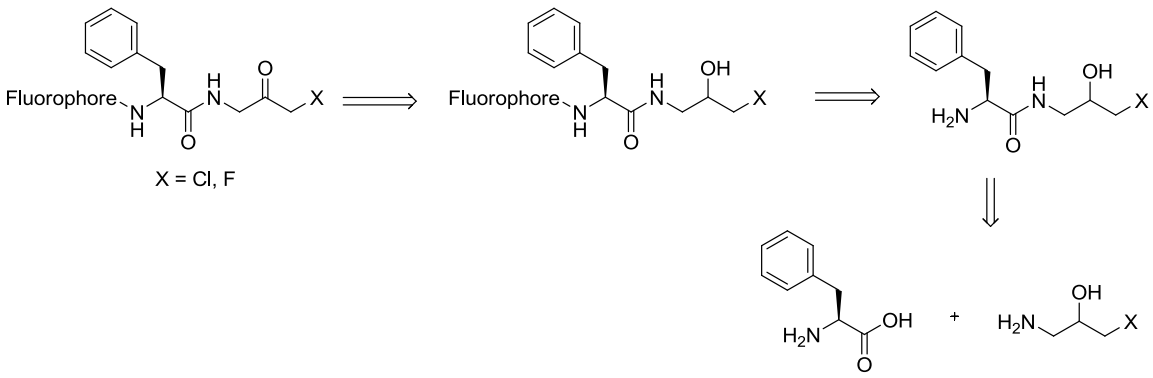


Figure 24: An illustration of potential binding modes for *Z*-*L*-Phe-CMK (**40**) shown as an overlay on top of compound **7a** (red). Top: Shows the *N*-Cbz group of *Z*-*L*-Phe-CMK potentially mimicking the phenylalanine of **7a**. Bottom: Shows the *N*-Cbz group of *Z*-*L*-Phe-CMK mimicking the *N*-hemisuccinic acid of **7a**.

4.2. Retrosynthetic Analysis

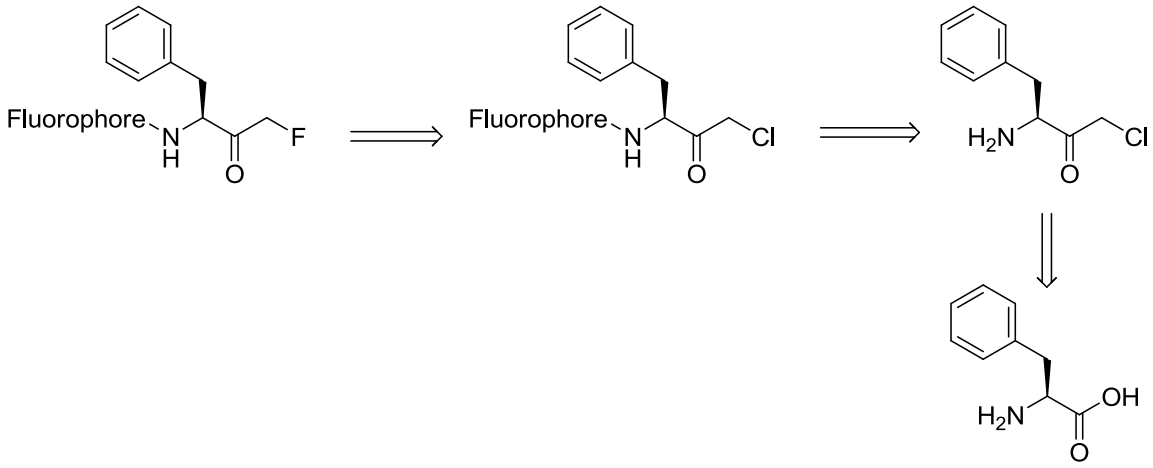
The retrosynthetic analysis for the target compounds **30a/b** can be seen in scheme 5, below.

Scheme 5: Retrosynthetic analysis of irreversible fluorescent inhibitors.



The retrosynthetic analysis is very similar to the first generation of inhibitors with the only difference being the *N*-terminal capping group.

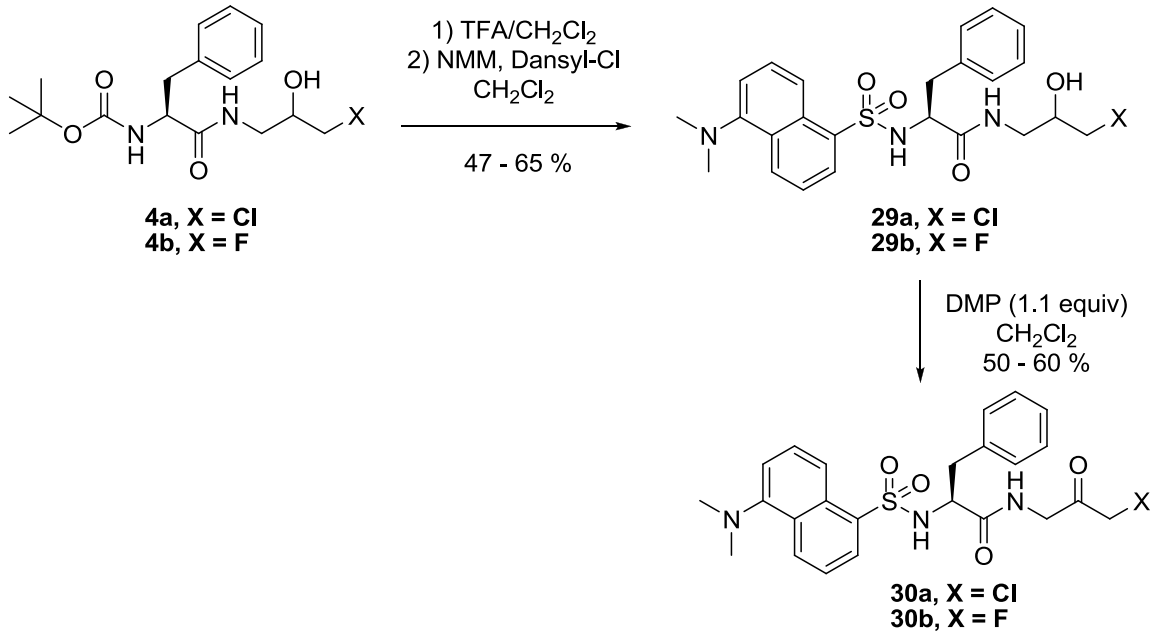
Scheme 6: Retrosynthetic analysis of truncated fluorescent inhibitors



It was envisaged that a fluoromethyl ketone could possibly be made in a more convergent manner by direct displacement of the chloride on the chloromethyl ketone **30a**. The chloromethyl ketone compounds are known and can be easily generated from the corresponding Boc-protected phenylalanine, using thionyl chloride to generate an acid chloride followed by homologation with diazomethane. Of the many potential fluorophores commercially available, dansyl-chloride was a logical first choice since its use in biological systems is well documented and it was on hand at the time of investigation. Fluorescein was another option; however we wanted to keep the inhibitor molecule structurally as small as possible due to the hindered active site present in ATG4B.

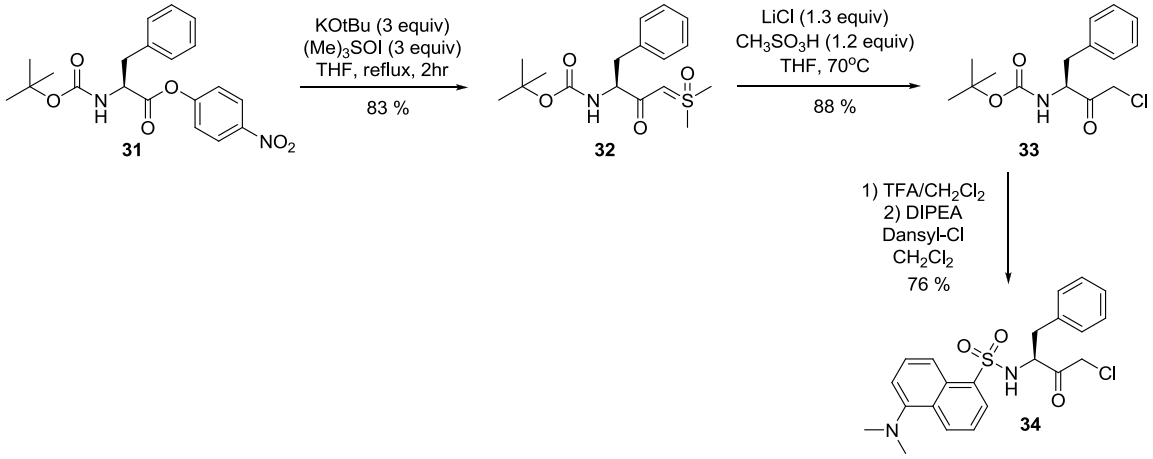
4.3. Synthesis

Scheme 7: Synthesis of fluorescent irreversible inhibitors



From Scheme 7 above, the synthesis for both the chloro and fluoro analogues began from compounds **4a** and **4b**. Removal of the Boc protecting group was achieved using trifluoroacetic acid (TFA) in a 50:50 mixture in dichloromethane. The reaction mixture was then evaporated, resuspended in methanol and evaporated again several times to remove any remaining TFA. The residue was resuspended in dichloromethane and basified using *N*-methylmorpholine (NMM) before being added to a stirring solution of dansyl chloride and NMM dissolved in dichloromethane. The progress of the reaction was monitored *via* TLC by looking at the disappearance of starting dansyl chloride and the appearance of a new fluorescent spot which was isolated by flash column chromatography to yield **29a** and **29b** in good yield. Oxidation of the secondary alcohol up to a ketone was achieved using Dess-Martin periodinane, which again resulted in good yields of compounds **30a** and **30b**.

Scheme 8: Synthesis of truncated fluorescent irreversible inhibitors

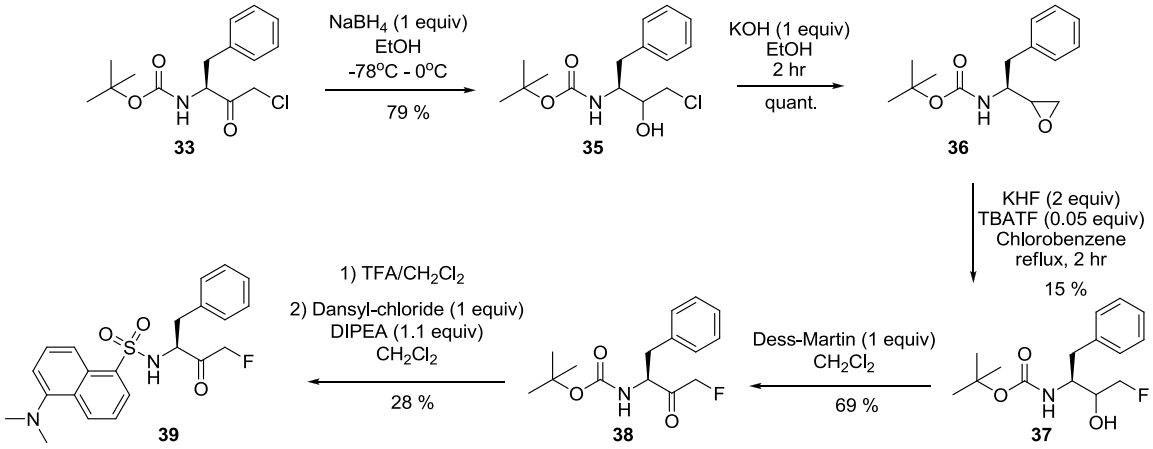


The truncated analogues **34** and **39** were not as straight forward to synthesize. Initial attempts to generate an acid chloride from Boc-protected phenylalanine failed. Acid chlorides generated from thionyl chloride, oxalyl chloride or various chloroformate reagents, such as isobutylchloroformate, were not isolable by column chromatography. Attempts to directly use the crude residue in the subsequent reaction were also met with failure. The homologation of acid chlorides requires the use of diazomethane, a notoriously toxic, shock sensitive and explosive reagent. Due to the explosive nature of such a reagent, it must be prepared on an individual basis using specialized glass ware. A stable derivative of diazomethane, TMS-diazomethane, is commercially available and is not as volatile although it is still rather toxic. Treatment of the crude residue, after attempted acid chloride generation, with TMS-diazomethane ultimately failed to give the desired product. The main by-product appeared to be the methyl ester of the starting phenylalanine. There is much literature precedence for the use of TMS-diazomethane on acid groups to convert them to the corresponding methyl esters.[74]

Success came by adapting methods described in a paper published in 2004 where the authors generated chloromethyl ketones from the sulfoxonium ylids generated from the corresponding para-nitro phenyl ester.[75] It is noted within the publication that in order to avoid racemisation one should start from the para-nitro phenyl ester rather than the methyl-ester. The reaction calls for refluxing trimethylsulfoxonium iodide and potassium *tert*-butoxide in THF for 2 hours to form the ylide. After reflux, the reaction

mixture is cooled to 0 ^{o}C before introducing the starting para-nitrophenyl ester, in this case **31**. The mixture was stirred overnight before quenching the base with the addition of water. The sulfoxonium ylide **32** was isolated by extraction with ethyl acetate, dried over sodium sulphate and evaporated under reduced pressure and used for the next step without further purification. The subsequent step involved the use of lithium chloride and methanesulfonic acid to produce HCl *in situ* and proceeded smoothly to displace DMSO to afford the chloromethyl ketone **33** in good yields. The final two steps of the synthesis involved a simple Boc-deprotection using TFA followed by an S_N2 type displacement reaction to append the fluorescent dansyl moiety onto the final compound's *N*-terminus.

Scheme 9



The fluoromethyl ketone (FMK) analogue **39** could not be obtained by a simple displacement of chloride by a fluoride source. Attempts to use tetrabutylammonium fluoride directly on compound **34** only led to decomposition of starting material. Initial attempts at replacing LiCl with LiF also failed to generate the desired product. However, it is possible to convert the chloromethyl ketone **33** into epoxide **36** by first reducing the ketone to an alcohol, then subsequent treatment of the secondary alcohol with base facilitated the epoxide formation. The epoxide could then be opened by a fluoride source, in this case, tetrabutylammonium dihydrogen trifluoride (TBATF). Subsequent oxidation and the appending of the dansyl fluorescent moiety proceeded in a similar

manner as for the chloromethyl ketone **34**, giving the desired compound **39** in acceptable yield.

4.4. Inhibition and Kinetic Studies

With the target compounds in hand and the appropriate assays having been established, the kinetics of inhibition was investigated. By design, the inhibitors have been synthesized to contain an electrophilic 'warhead'. The nature of such a warhead is to covalently bond to a nucleophilic active site residue of an enzyme irreversibly. However, we know there are several potentially reactive cysteine thiols in ATG4B and these could compete and complicate the analysis. Furthermore, we have the ability to monitor the alkylation by MS and to see if mono or polyalkylation is taking place. If alkylation occurs at the active site thiol then this is manifested by observing a time-dependent effect on enzyme activity over time. The interaction between an irreversible inhibitor and an enzyme is generally a two-step reaction with the first step typically being reversible and diffusion controlled non-covalent binding. The second (slower) step is the irreversible covalent modification of the enzyme. For this reason, inactivation rate constants are represented as k_{inact}/K_I. As can be seen from the figure below, both **7a** (TN74) and **7b** (TN160) exhibit time-dependent inhibition that is characteristic of irreversible inhibitors.

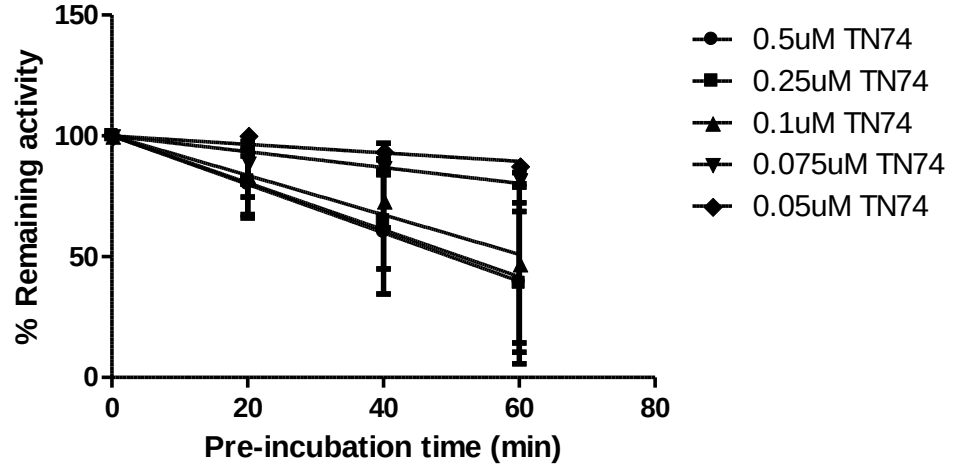


Figure 25a: A plot of the percentage of remaining enzyme activity versus pre-incubation with inhibitor time. Compound **7a** displays time-dependent inhibition characteristic of irreversible inhibitors.

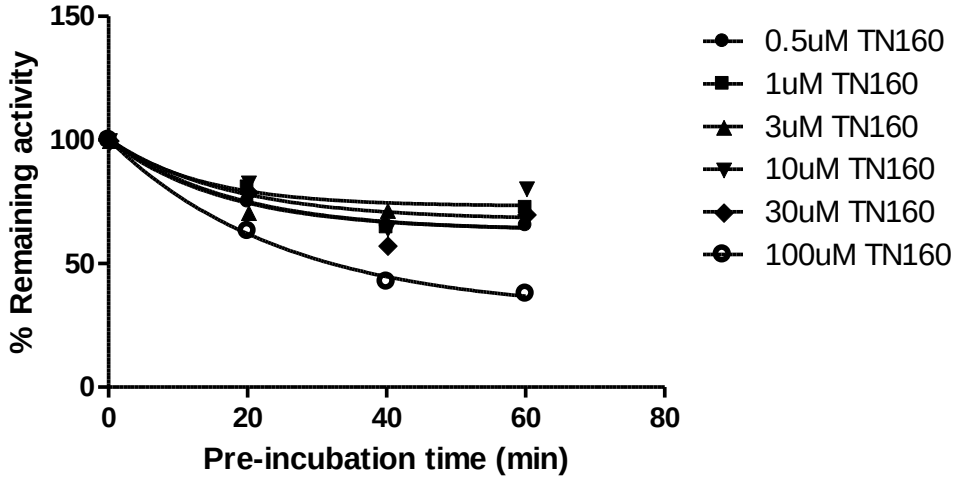


Figure 25b: A plot of the percentage of remaining enzyme activity versus pre-incubation with inhibitor time. Compound **7b** displays time-dependent inhibition characteristic of irreversible inhibitors.

Suitable concentrations and time ranges were selected based on preliminary experiments. The inhibitors were allowed to incubate with ATG4B for various periods of time before adding the FRET-LC3 substrate and monitoring fluorescence changes.

Unfortunately, FMK **7b** did not show the expected linear correlation, even at the lowest concentrations tested. The data from the CMK **7a** can be fit to a first-order rate equation:

$$\% \text{ remaining activity} = 100 \times e^{-(kobs \times t)} \qquad (18)$$

The above equation defines the percentage of remaining enzyme activity as a function of time that is associated with a rate constant (k_{obs}). Thus, at time zero (no pre-incubation with inhibitor) the remaining activity is 100%. Substituting known times and remaining activity, k_{obs} can be calculated for each inhibitor concentration. A plot of k_{obs} vs [I] should yield a Michaelis-Menten like curve, which can be fit to the equation:

$$k_{obs} = k_{inact} \times [I] / (K_I + [I]) \qquad (19)$$

The above equation is analogous to the Michaelis-Menten equation outlined in the first chapter. The rate of the reaction is defined by the rate of inactivation (k_{inact}) which varies as a hyperbolic function of the inhibitor concentration (I) with K_I and k_{inact} as the two parameters that define the curve. A plot of k_{obs} vs [I] for CMK **7a** can be seen in the figure below.

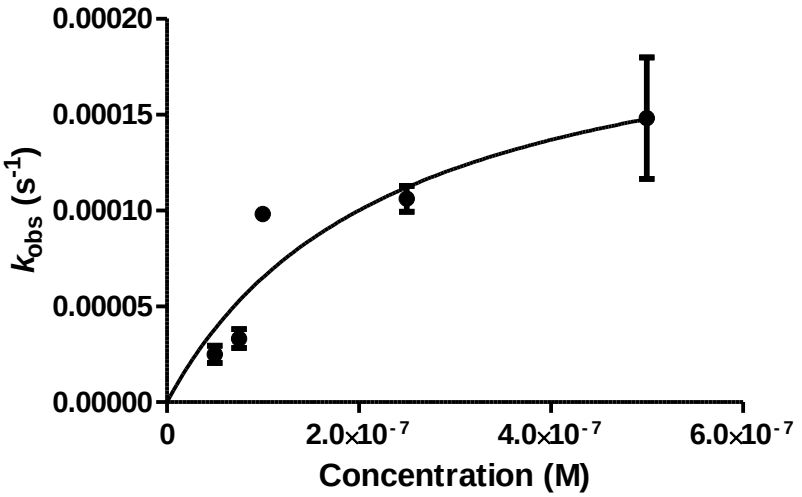


Figure 26: A plot of k_{obs} versus **7a** concentration. This plot was fit using GraphPad 5 software with a non-linear regression analysis to the equation $Y=k_{inact}*X/(K_I+X)$. K_I = 200 nM ± 100 nM, k_{inact} = 2.0 x 10^{-4} s^{-1} ± 0.5 x 10^{-4} s^{-1}, k_{inact}/K_I = 1000 s^{-1} M^{-1} ± 700 s^{-1} M^{-1}.

Performing the analysis this way results in a K_I of 200 nM and a k_{inact} of 2.0 x 10^{-4} s^{-1}. This implies that binding is strong due to a K_I value in sub-micromolar ranges. The rate of inactivation appears rapid, but the half-life of the EI complex is calculated to be approximately 1 hr with observable saturation. Unfortunately, the same plot could not be derived for **7b**. From the time-dependent analysis in Figure 25 above, **7b** did not show a linear correlation with respect to pre-incubation time. It was speculated that this could mean the enzyme was being re-activated at higher inhibitor concentrations, possibly due to the products of the reaction reactivating the enzyme. However, complications arising from the fluorescent read-out of the initial assay are more probable. Alternatively, to avoid such complications, direct measurements of the rate of alkylation of ATG4B can be made using an LC/MS system. By monitoring the mass change of ATG4B over time due to irreversible binding to the inhibitor, one can derive an inactivation rate constant directly; in this case it would elucidate k_{inact} directly. This technique is particularly useful when trying to analyze fluorescent compounds. Preliminary experiments were conducted in order to choose appropriate concentration ranges and incubation times in order to ensure that the inhibitor and enzyme react in a 1:1 ratio. Due to the reactive nature of some of the irreversible inhibitors it is possible to alkylate ATG4B multiple times. This is also due to the fact that ATG4B itself contains 13 cysteine residues, many of which react with CMK **7a**. It can be seen from the figures below that CMK **7a** reacts multiple times with ATG4B even under very short reaction times (Figure 27 middle, expected mass: 46204 Da ATG4B, observed mass: 48114 Da, ATG4B + 1910 Da), while FMK **7b** only reacts once with ATG4B (Figure 27 bottom, expected mass: 46204 Da ATG4B, observed mass: 46523 Da, ATG4B + 319 Da). This ATG4B construct was provided by the SGC in Toronto and differs slightly from the predicted mass on page 25, but the enzymatic activity was evaluated to be the same by parallel reactions monitoring cleavage of proLC3B.

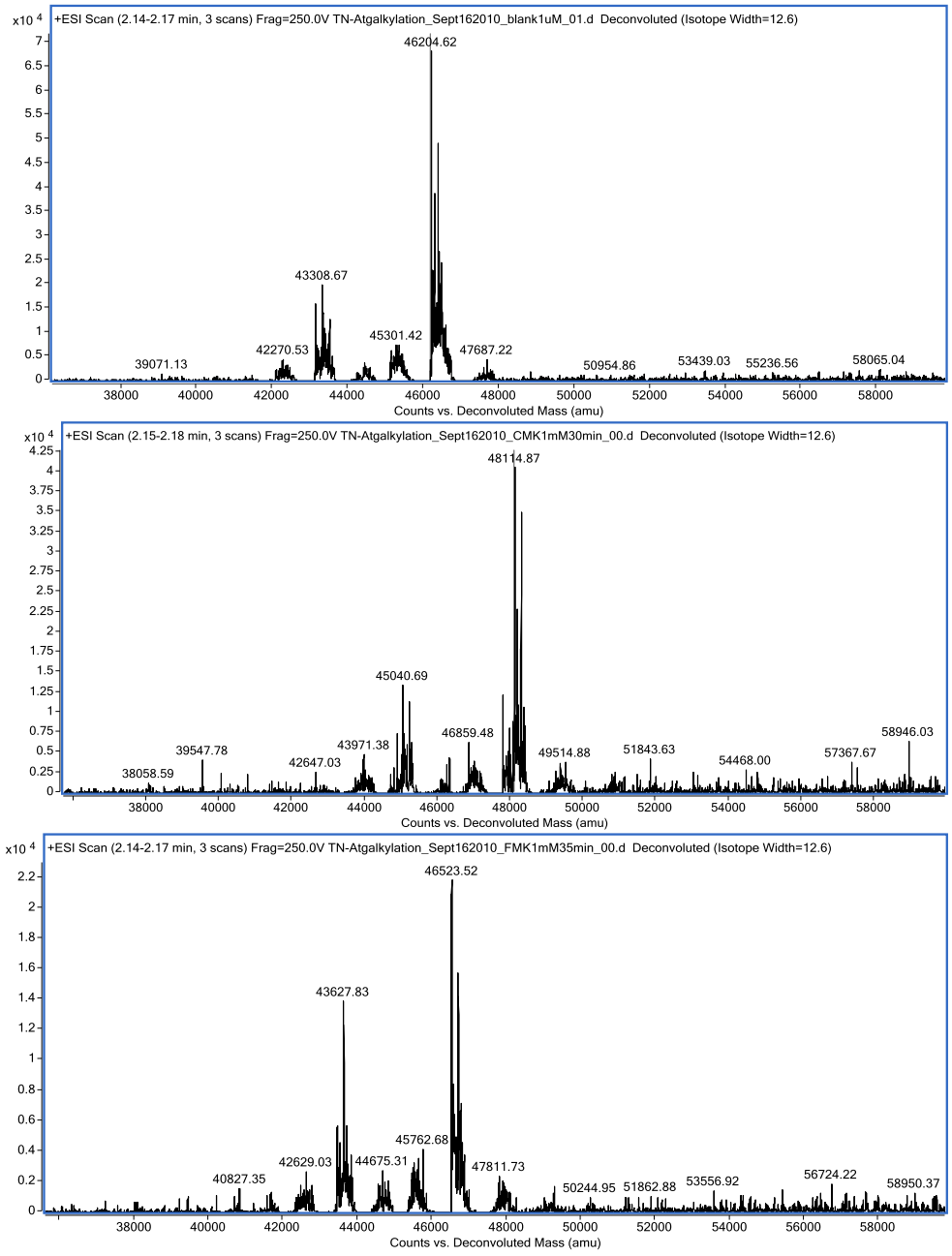


Figure 27: Top) Deconvoluted mass spectrum of ATG4B with parent mass of 46204 Da. Expected mass: 46335 Da, observed mass: 46204 Da for ATG4B. **Middle)** Deconvoluted mass spectrum of ATG4B pre-incubated with CMK **7a** for 30 min at room temperature. A mass of 48114 Da corresponds to a mass addition of 13 molecules of inhibitor to ATG4B. **Bottom)** Deconvoluted mass spectrum of ATG4B pre-incubated with FMK **7b** for 30 min at room temperature. A mass of 46523 Da corresponds to 1 molecule of inhibitor bound to ATG4B.

Presumably the alkylation occurs at the active site of the enzyme first. Indeed this seems to be the case when a sample of mono-alkylated ATG4B was assayed for residual activity against FRET-LC3. Mono-alkylated ATG4B was inactive using FRET-LC3 as substrate (data not shown) and as measured by LC/MS. We observed that the mono-alkylation event was sufficient to completely inhibit ATG4B activity. The fluorescent inhibitors **30a**, **30b**, and **34** were analyzed by LC/MS and the results can be seen in the figures below. The plots were fitted to a line using linear regression analysis performed by GraphPad 5 software.

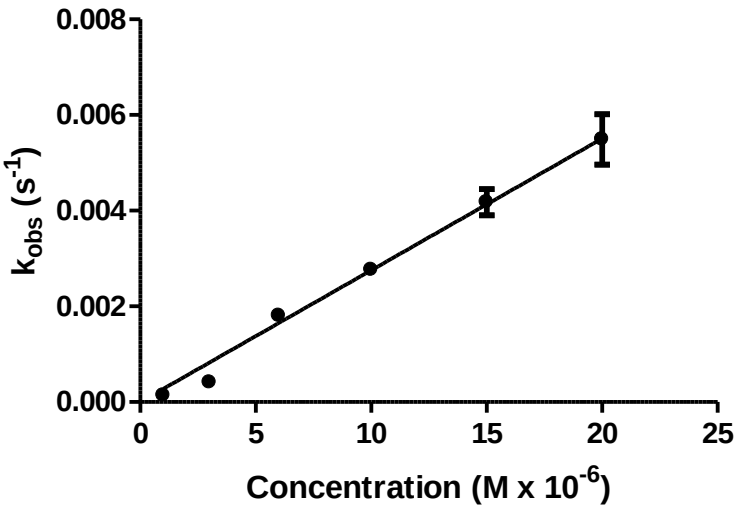


Figure 28: A plot of k_{obs} versus **30a** concentration. The second order rate constant k_{inact}/K_I = 275 s^{-1} M^{-1} is calculated from the slope of the plot using GraphPad 5 software. k_{obs} was derived from monitoring the rate of ATG4B mono-alkylation.

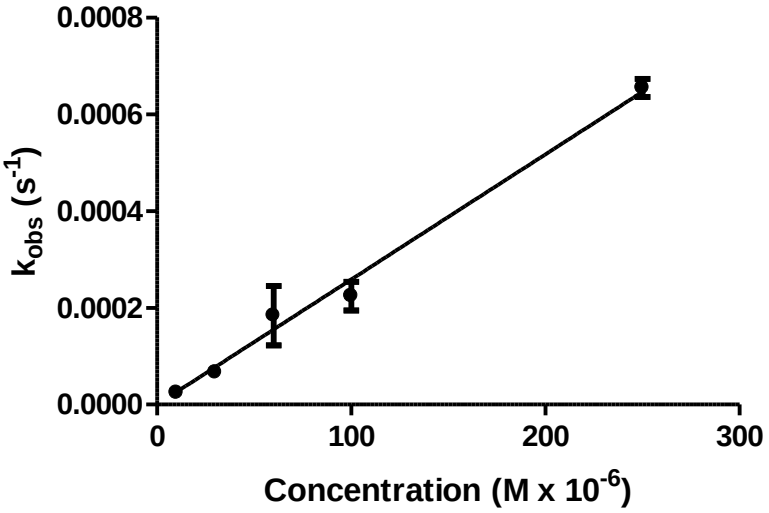


Figure 29: A plot of k_{obs} versus **30b** concentration. The second order rate constant k_{inact}/K_I = 2.58 s^{-1} M^{-1} is calculated from the slope of the plot using GraphPad 5 software. k_{obs} was derived from monitoring the rate of ATG4B mono-alkylation.

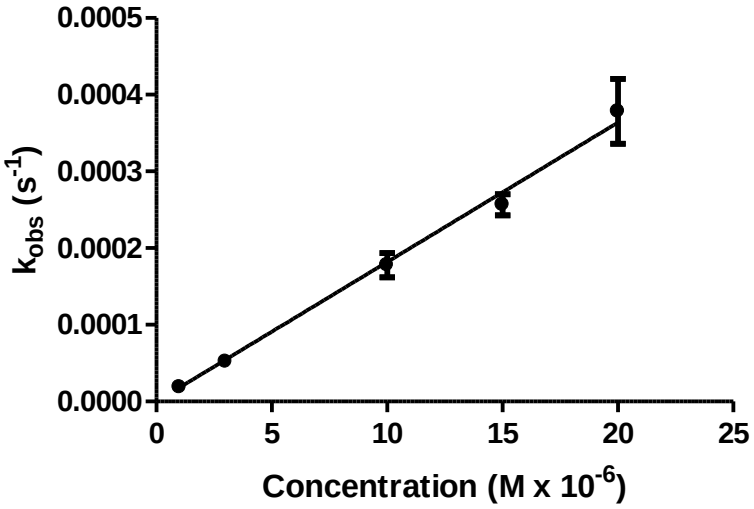


Figure 30: A plot of k_{obs} versus **34** concentration. The second order rate constant k_{inact}/K_I = 18.1 s^{-1} M^{-1} is calculated from the slope of the plot using GraphPad 5 software. k_{obs} was derived from monitoring the rate of ATG4B mono-alkylation.

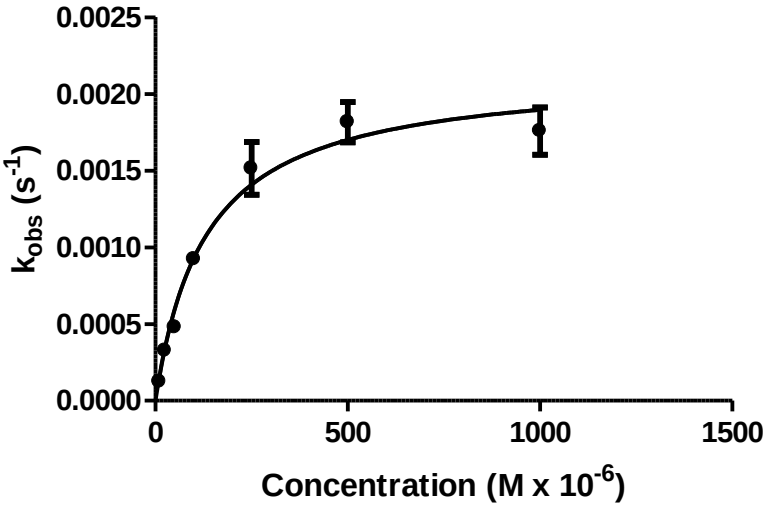


Figure 31: A plot of k_{obs} versus **7b** concentration. The data was fitted to the equation: $Y = (k_{inact} * X) / (K_I + X)$ using non-linear regression analysis performed by GraphPad 5 software. $k_{inact} = 2.14 \times 10^{-3}\ s^{-1}$, $K_I = 130\ \mu M$, $k_{inact}/K_I = 16.4\ s^{-1}\ M^{-1}$.

The rate constants (k_{obs}) for each inhibitor were derived by first measuring the ratio of free-enzyme to alkylated enzyme over time. This data was then fitted to the first order rate equation:

$$A = A_o e^{(-kobs*t)} \qquad (20)$$

Which can be rewritten as:

$$\ln (A/A_o) = -k_{obs} * t \qquad (21)$$

Where A is the amount of free enzyme at time = t, A_o is the amount of inactivated enzyme at time = t, k_{obs} is the observed rate constant and t is time. The rate constant k_{obs} can then be plotted versus inhibitor concentration. A linear plot implies a weak binding constant K_I. It can be seen from Figures 28-31 that only one compound displayed saturation, **7b**. Although **30b** was another compound that only reacts once with ATG4B across all concentrations tested, it still did not show saturation kinetics. With the exception of **7b**, none of the fluorescent compounds displayed saturation. This implies that the K_I is much higher than the inhibitor concentrations used for the experiments. All of the time points measured represented a single alkylation. Using

higher than 20 µM of inhibitor (**30a**, **34**) resulted in multiple alkylations over the time points tested. As a result, only the 2nd order rate constant k_{inact}/K_I could be derived for the fluorescent compounds.

In order to further validate that these irreversible inhibitors are indeed binding to the active-site of ATG4B, protection experiments were conducted. This involves examining the rate of inhibition in the presence of a known substrate. The idea is that there will be competition for the active-site of the enzyme and the presence of a known substrate should reduce, or protect, the enzyme from inhibition at suitable concentrations.

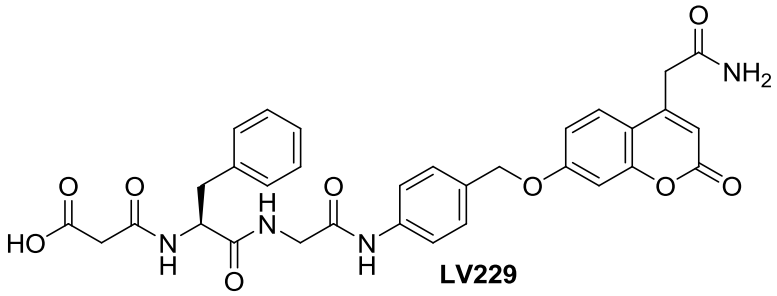


Figure 32: Structure of a synthetic fluorescent peptide substrate (**LV229**) for monitoring ATG4B activity.

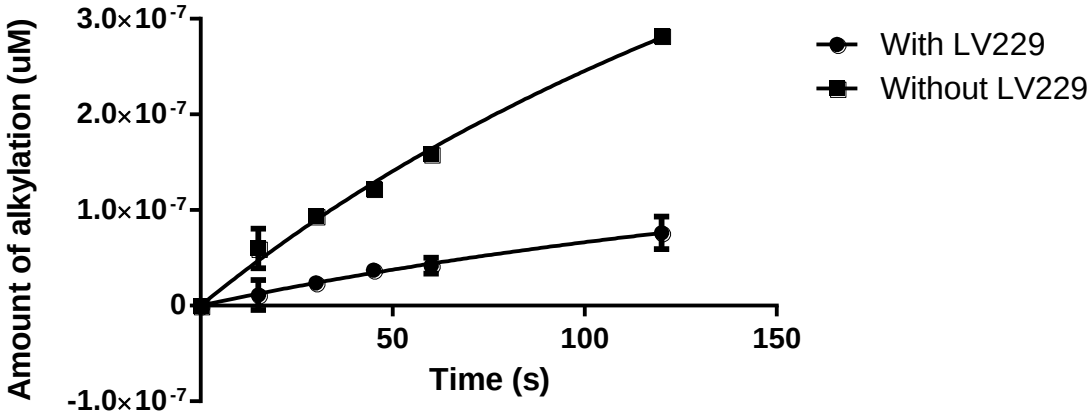


Figure 33: Rate of alkylation of 1 µM ATG4B by **30a** (10 µM) in the presence of fluorescence peptide substrate LV229 (600 µM).

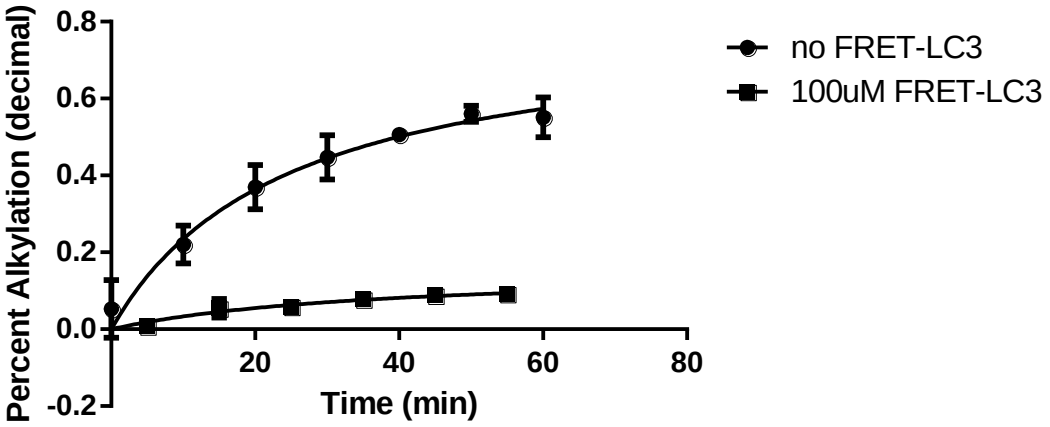


Figure 34: Rate of alkylation of 1 µM ATG4B by **7b** (50 µM) in the presence of FRET-LC3 substrate (100 µM).

It can be seen from the above figures that the rate of alkylation of ATG4B is significantly reduced in the presence of a known substrate. Figure 33 shows that in the presence of a synthetic peptide substrate (**LV229**), which was developed by post-doctoral fellows Dr. Nag Kumar and Dr. Lubomir Vesenkov[76], the ability of **30a** to alkylate ATG4B is reduced. One may even be able to predict the level of protection based on known K_M values of the substrates following the equation below.

$$K_{Iapp} = K_I \left(1 + \frac{[S]}{K_M}\right) \qquad (17)$$

Earlier work by Dr. Vesenkov showed that his peptide substrate, **LV229**, has a K_M of approximately 50 µM.[76] For the protection experiment, 12x the K_M value was used. Indeed, with LV229 present, the rate of alkylation was approximately 10 fold slower. This was further demonstrated using **7b** and the FRET-LC3 substrate (Figure 34). In this case, a concentration of 10x the K_M was used and indeed the rate of alkylation was approximately 10-fold slower in the presence of FRET-LC3. This data taken together with the mono-alkylation evident from the mass shift studies, it is safe to say that these classes of irreversible inhibitors are indeed targeting the active site of ATG4B, binding irreversibly and inhibiting ATG4B activity *in vitro*.

Trypsinization of ATG4B reveals binding location

In order to further validate inhibition and binding at the active site of ATG4B, digestion experiments were performed on enzyme samples treated with inhibitor **30b**. Monoalkylated ATG4B samples were digested using trypsin and the resulting peptide fragments were analyzed by LC/MS. Table 2 shows the predicted peptide fragments resulting from a trypsinization.

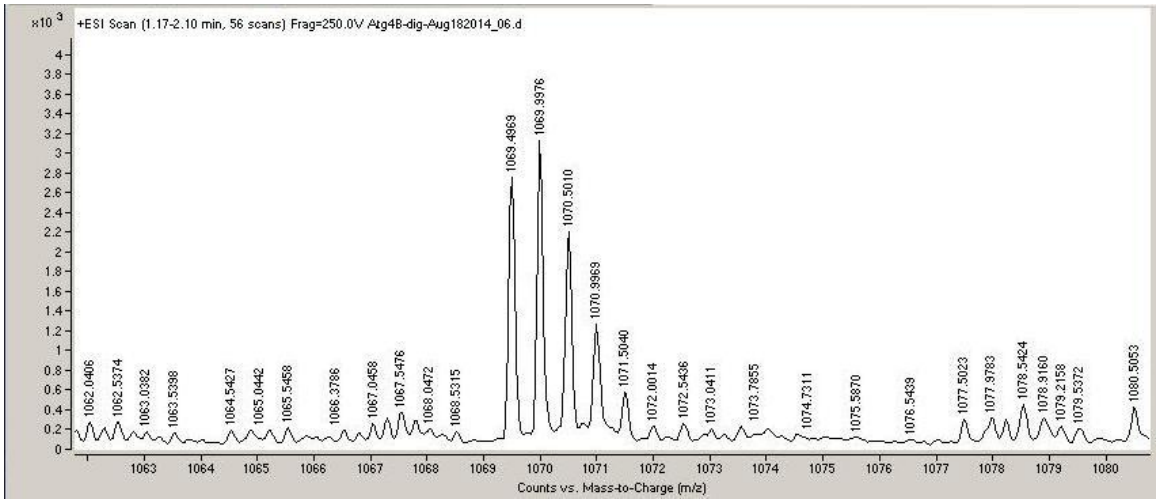


Figure 35: Mass spectrum of a tryptic sample of ATG4B showing a mass of 1069 m/z. This corresponds to a predicted peptide fragment NFPAIGGTGPTSDTGWGCMLR of mass 2138 Da detected with a +2 charge, which contains the putative active site cysteine residue Cys^{74}. Masses are detected as a mass-to-charge ratio.

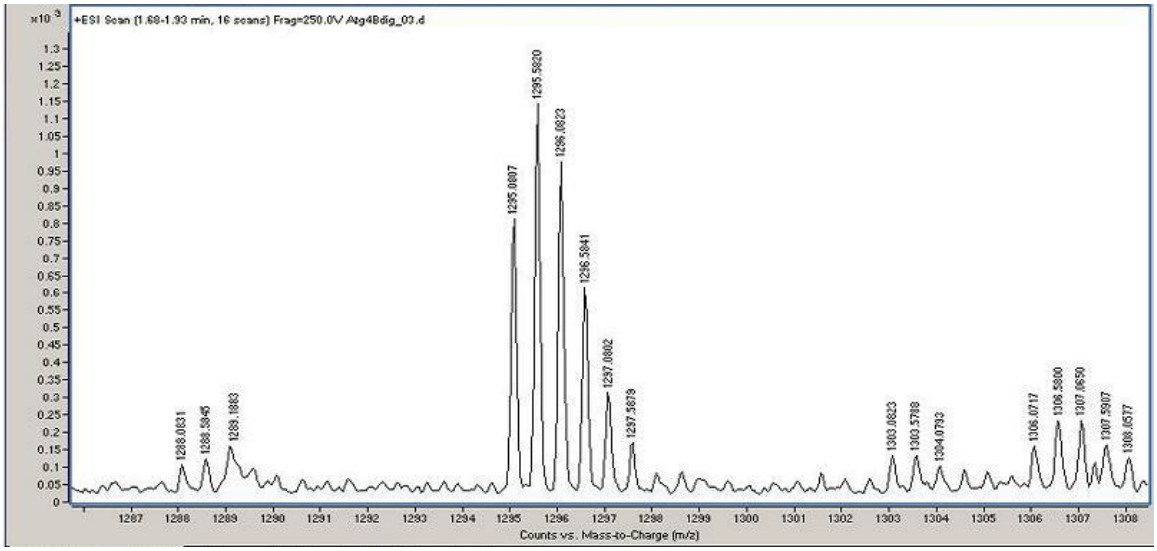


Figure 36: Mass spectrum of a tryptic sample of monoalkylated ATG4B showing 1295 m/z. This corresponds to a peptide fragment of mass 2590 Da detected with a +2 charge.

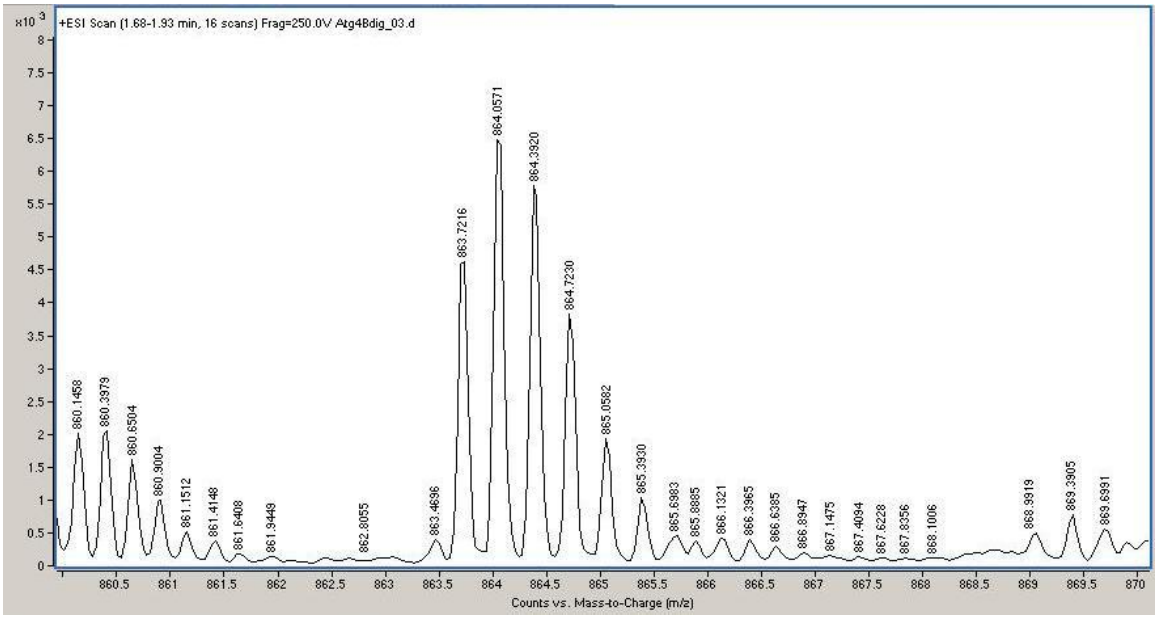


Figure 37: Mass spectrum of a tryptic sample of monoalkylated ATG4B showing 864 m/z. This corresponds to the same peptide fragment of mass 2590 detected with a +3 charge.

Table 2: Peptides resulting from the trypsinization of ATG4B as calculated by ExPASy.[77]

Position of cleavage site	Resulting peptide sequence	Peptide length (aa)	Peptide mass (Da)
17	MGSSHHHHHHSSGLVPR	17	1900.067
31	GSMDAATLTYDTLR	14	1514.670
50	FAEFEDFPETSEPVWILGR	19	2269.495
58	YSIFTEK	7	887.000
68	DEILSDVASR	10	1104.182
74	LWFTYR	6	885.033
96	NFPAIGGTGPTSDTGWG**C**MLR	21	2138.400
109	CGQMIFAQALVCR	13	1439.772
113	HLGR	4	481.555
137	QPDSYFSVLNAFIDR	15	1771.947
156	DSYYSIHQIAQMGVGEGK	18	1983.183
172	SIGQWYGPNTVAQVLK	16	1761.010
199	LAVFDTWSSLAVHIAMDNTVVMEEIR	26	2948.402

Trypsin is an enzyme that cleaves peptide bonds on the C-terminal side of lycine and arginine. In this way, characteristic peptide fragments can be analyzed and mapped according to their expected fragmentation patterns. Trypsin is known to cleave after arginine and lysine residues and thus the predicted fragment containing the active site cysteine should have a mass of 2138.4 Da, which can be seen in table 2, entry 7. With HPLC-MS we observed a peak at retention time of 7.2 min and it can be seen from Figure 35 that 1069 m/z is observed for this peptide fragment. This mass matches with the expected peptide of mass 2138 Da detected with a +2 charge. The peptide fragment 2138 Da is expected to contain the active site cysteine residue (Cys^{74}). Analysis of the monoalkylated tryptic sample of ATG4B reveals a peptide fragment of mass 1295 m/z (Figure 36). This mass corresponds with the expected mass if one molecule of inhibitor is covalently bound to the peptide fragment at the active site cysteine residue (Cys^{74}) [(2138 + 452) ÷ 2] = 1295. This is further validated by the observation of a peak at 864 m/z, which corresponds to the labelled active-site fragment detected with a +3 charge (Figure 37). Other tryptic peptides can be observed and are unchanged after labeling (data not shown) resulting from the digestion of ATG4B which validates that these fragmentation patterns are a result of ATG4B digestion with trypsin and that this class of irreversible inhibitor is indeed active-site directed.

4.5. Experimental

4.5.1. Kinetic experiments

Inhibition experiments

Reducing buffer used throughout is 50 mM Tris, 35 mM NaCl, 5 mM TCEP, pH 8. Inhibitor stock solutions were made in DMSO at 10 mM concentrations and kept at 0 ^{o}C for storage. Inhibitor stock solutions were diluted in DMSO to suitable concentrations to ensure that no more than 2 µL (4%) of DMSO would be needed. ATG4B frozen stocks were diluted with the appropriate reducing buffer and allowed to incubate for 1 h at 0 ^{o}C prior to incubation with inhibitor at 25 ^{o}C. All black 384-well plates were loaded with 23 µL reducing buffer, 2 µL inhibitor, and 5 µL ATG4B. Addition of ATG4B required manual mixing (pipetting up and down 3 times) to ensure a homogenous mixture. ATG4B was

then incubated with the inhibitors for up to 1 hour at 25 ^{o}C prior to addition of 20 µL of fluorescent protein substrate. Experiments were performed in triplicate. Fluorescence changes at 510 nm were monitored. Inhibition rate was compared to an uninhibited control which contained 4% DMSO, but no inhibitor.

Time-Dependent Inhibition experiment

Inhibitor solutions were made in serial dilutions using DMSO starting from a 10 mM stock solution. Final inhibitor concentrations ranged from 0.05 µM – 10 µM in the case of **7a** and 0.5 µM – 100 µM in the case of **7b**. 2 µL of inhibitor were then added into an all black 96-well plate treated with a non-stick surface (Corning). 48 µL of reducing buffer was then manually pipetted into each well containing inhibitor. A stock aliquot of ATG4B (10 µL) in buffer (50 mM Tris, 500 mM NaCl, pH 8.0) was diluted to 240 nM using reducing buffer and was allowed to incubate for at least 1 hour at 0 ^{o}C prior to assay measurements. 20 µL of ATG4B (240 nM) were then dispensed into each well containing inhibitor and allowed to incubate for 20, 40, and 60 minutes. After each pre-incubation, 30 µL of FRET-LC3 substrate (YFP-LC3B-EmGFP; 11.7 µM in 50 mM Tris, 35 mM NaCl, 2 mM TCEP, pH 8) was dispensed into the appropriate wells containing enzyme and inhibitor bringing the final volume to 100 µL. Fluorescence was measured every 2 minutes for 16 minutes exciting at 450 nm and monitoring the emission at 510 nm. All experiments were performed in triplicate and the percent remaining activity was determined by comparison with controls containing DMSO, but no inhibitor or no enzyme respectively.

CMK FMK labelling experiments

ATG4B labelling experiments were performed using a final concentration of 1 µM enzyme and either 10 µM or 1000 µM of inhibitor in a final reaction volume of 50 µL. Final concentrations of DMSO did not exceed 4%. A small volume glass sleeve was inserted into a standard clear HPLC vial and filled with 43 µL of the reducing Tris buffer. 2 µL of inhibitor (0.25 mM or 25 mM) was then added to the vial and mixed manually to ensure homogeneity. 5 µL of ATG4B (10 µM) was then added to the vial and the mixture was allowed to incubate for 1 hour. The rate of alkylation can be monitored by sequential injections into the LC/MS every 5 minutes. Experiments were performed in duplicate and compared to a negative control (no inhibitor). Analysis of the MS trace

was performed using Agilent MassHunter Workstation. Deconvolution range was set to elucidate proteins between 20000 – 60000 Da deconvoluting between 700 m/z and 1800 m/z of the MS spectrum. Mass of native ATG4B was found to be 46204 Da at time zero. After 1 hour the 46204 Da mass peak was completely suppressed and a new mass of 46524 Da was found in the case of the FMK. A new peak of 48114 Da was found in the case of the CMK.

Rate of alkylation

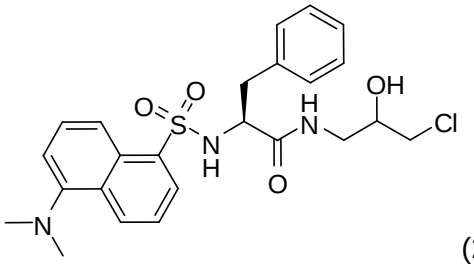
Similar to the method above for the labelling experiments, inhibitor solutions were prepared in serial from a 10 mM DMSO stock solution. 5 µL of each inhibitor solution in DMSO was then incubated with 1 µM ATG4B with a final volume of 100 µL for 15 s – 2 min. After incubation, the reaction mixture (100 µL) was then applied to a desalting column (Fischer cat. #PI-89882, Thermo Scientific No.:89882) to rapidly remove the inhibitor from the mixture. The desalting column was first centrifuged at 1500 g for 1 min and then washed with buffer solution 2 times before loading samples. 100 µL of sample was then loaded onto the column and centrifuged at 1500 g for 1 min. The flow through was then analyzed as described above.

Trypsin digest protocol

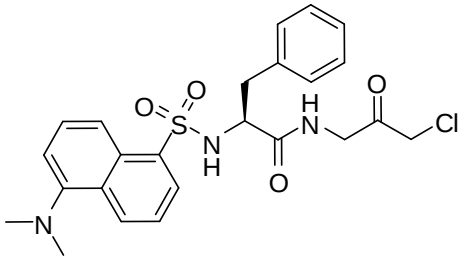
Monoalkylated ATG4B samples were prepared by first flash thawing a frozen aliquot of ATG4B (20 µL, 1 mM) and diluting the sample up to 100 µL using the standard reducing buffer. ATG4B was allowed to incubate with the reducing buffer at 0 ^{o}C for 1 hr. For monoalkylation, 5 µL of inhibitor dissolved in DMSO (8 mM) was then added and the mixture was allowed to incubate overnight at 4 ^{o}C. Complete monoalkylation was confirmed by analyzing a small aliquot of the solution (~1 µL) by LC/MS. The inhibitor was then removed from the mixture by applying the solution to a de-salting column. Confirmation of complete removal of inhibitor was done by analysing another aliquot (1 µL) of the mixture after subjecting it to the de-salting column and observing the disappearance of inhibitor. 5 µL of aqueous TCEP solution (0.5 M) was then added along with 45mg of urea. The resulting mixture was allowed to incubate at 37 ^{o}C for 1 hr. The mixture was then diluted with 850 µL of aqueous Tris solution (40 mM). 45 µL of trypsin (1 µg/µL in 1 mM HCl) was then added to bring the final volume to 1000 µL. The mixture was then incubated at 37 ^{o}C overnight. The tryptic activity was terminated

by the addition of 4 µL of conc. HCl. The sample was analyzed by LC/MS without further purification using a C8 guard column and eluting with 95:5 CH_3CN:water.

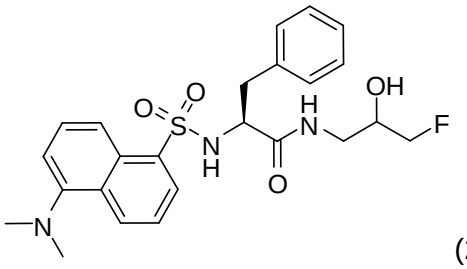
4.5.2. Characterization of compounds



(2-*S*)-*N*-(3-chloro-2-hydroxypropyl)-2-(5-(dimethylamino)naphthalene-1-sulfonamido)-3-phenylpropanamide, **29a**. Initial compound **4a** (0.285 g, 1.110 mmol) was deprotected according to standard conditions involving TFA. The free amine was then resuspended in 2 mL CH_2Cl_2. Basicity was assessed using pH paper and adjusted to pH 8 using approximately 2 equivalents of N-methylmorpholine. In a separate flask, dansyl-chloride (0.200 g, 0.740 mmol) and N-methylmorpholine (0.0815 mL, 0.740 mmol) was dissolved in 5 mL CH_2Cl_2. The free amine of **4a** was added drop-wise to the dansyl-chloride mixture and allowed to stir at r.t. until the dansyl-chloride was consumed. The reaction was maintained at a basic pH by periodic additions of N-methylmorpholine whenever a precipitate was observed or by checking with pH paper. Upon complete consumption of dansyl-chloride the reaction was concentrated under reduced pressure. The crude residue was purified by flash column chromatography to afford an off-white solid in 46% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ_{ppm} = 8.58 (d, *J* = 8.5 Hz, 1H), 8.22 – 8.18 (m, 1H), 8.04 (dd, *J* = 20.9, 8.6 Hz, 1H), 7.55 – 7.46 (m, 2H), 7.20 (dd, *J* = 7.4, 3.9 Hz, 1H), 7.03 – 6.97 (m, 3H), 6.80 – 6.76 (m, 2H), 5.21 (dd, *J* = 16.2, 6.3 Hz, 1H), 3.90 – 3.81 (m, 2H), 3.56 – 3.34 (m, 4H), 3.34 – 3.12 (m, 1H), 2.93 (s, 6H), 2.80 – 2.71 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ_{ppm} = 171.92 (rotomer), 171.84, 134.51, 134.47, 132.65, 131.25, 130.31, 130.28, 129.27, 129.25, 128.87, 128.77, 128.67, 128.66, 127.35, 123.11, 115.42, 70.66 (rotomer), 70.57, 60.40, 58.38, 58.23 (rotomer), 46.23, 46.22 (rotomer), 45.46, 43.23, 43.01 (rotomer), 37.95, 28.22. mp = 152 – 156^{o}C. HRMS (ESI TOF) m/z calculated for $C_{24}H_{28}ClN_3O_4S$: 512.1387 m/z (M+Na), found: 512.1366 m/z (M+Na).

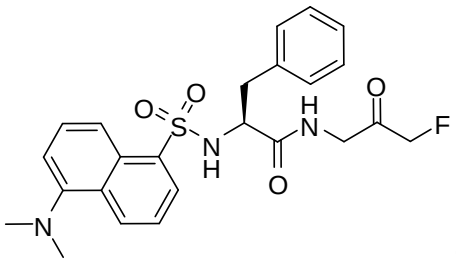


(*S*)-*N*-(3-chloro-2-oxopropyl)-2-(5-(dimethylamino)naphthalene-1-sulfonamido)-3-phenylpropanamide, **30a**. The alcohol **29a** (0.180 g, 0.360 mmol) was dissolved in 3 mL CH_2Cl_2 and to that solution Dess-Martin Periodinane (0.170 g, 0.400 mmol) was added according to the general oxidation procedure. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford **30a** as a yellow oil in 20 % yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 8.58 (d, *J* = 8.2 Hz, 1H), 8.18 (d, *J* = 7.1 Hz, 1H), 8.08 (d, *J* = 8.6 Hz, 1H), 7.59 – 7.42 (m, 2H), 7.21 (d, *J* = 7.5 Hz, 1H), 7.02 (t, *J* = 5.0 Hz, 1H), 7.00 – 6.92 (m, 3H), 6.84 – 6.73 (m, 2H), 5.32 (s, 1H), 4.28 (ddd, *J* = 19.4, 5.6 Hz, 1H), 4.16 (ddd, *J* = 19.4, 5.6 Hz, 1H), 4.14 (d, *J* = 2 Hz, 2H), 3.95 (q, *J* = 6.5 Hz, 1H), 2.92 (m, 7H), 2.80 (dd, *J* = 14.5, 6 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 197.83, 170.88, 134.50, 133.18, 131.03, 130.12, 129.28, 128.82, 128.76, 128.59, 128.53, 127.19, 123.16, 115.44, 58.25, 47.23, 46.26, 45.51, 42.78, 37.92. mp = 75 – 80^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{24}H_{26}ClN_3O_4S$: 510.1230 m/z (M+Na), found: 510.1212 m/z (M+Na) = -56.3^o (c 2.7x10^{-3} g/mL, EtOH).

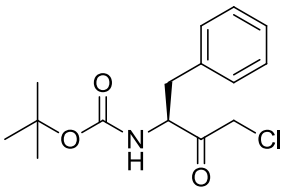


(2-*S*)-2-(5-(dimethylamino)naphthalene-1-sulfonamido)-*N*-(3-fluoro-2-hydroxypropyl)-3-phenylpropanamide, **29b**. The free amine of compound **4b** (0.450 g, 1.850 mmol) was obtained in the same manner as for the free amine resulting from **4a** described above and was introduced to a stirring solution of dansyl-chloride (0.500 g, 1.850 mmol) dissolved in 10 mL CH_2Cl_2 similarly to the method described above for **29a**. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford **29b** as a yellow oil in 46% yield. ^{1}H NMR (500 MHz, DMSO) δ_{ppm} = 8.35 (m, 2H), 8.18 (t, *J* = 8.9 Hz, 1H), 7.93

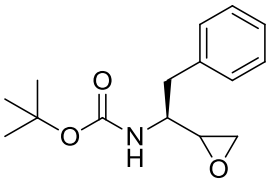
(dt, *J* = 21.4, 5.8 Hz, 1H), 7.85 (ddd, *J* = 18.0, 7.2, 0.8 Hz, 1H), 7.53 – 7.46 (m, 1H), 7.43 (ddd, *J* = 10.4, 8.4, 7.5 Hz, 1H), 7.20 (dd, *J* = 7.5, 1.4 Hz, 1H), 7.03 – 6.91 (m, 5H), 5.13 (t, *J* = 5.2 Hz, 1H), 4.21 – 3.82 (m, 3H), 3.53 – 3.36 (m, 1H), 2.93 – 2.72 (m, 9H), 2.60 (m, 1H). ^{13}C NMR (126 MHz, DMSO) δ_{ppm} = 171.27, 171.03 (rotomer), 151.49, 137.36, 137.24, 136.79, 136.74 (rotomer), 129.62, 129.59 (rotomer), 129.42, 128.39, 128.25, 128.06, 128.02, 127.91, 127.89 (rotomer), 126.56, 126.51, 123.56, 120.08, 115.36, 85.83, 85.78 (rotomer), 84.49, 84.45 (rotomer), 68.36, 68.27 (rotomer), 68.21, 68.12 (rotomer), 58.41, 45.55. mp = 132 – 140^{o}C. HRMS (ESI TOF) m/z calculated for $C_{24}H_{28}FN_3O_4S$: 496.1682 m/z (M+Na), found: 496.1693 m/z (M+Na).



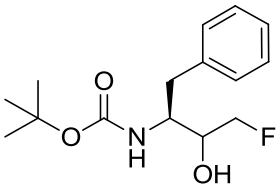
(*S*)-2-(5-(dimethylamino)naphthalene-1-sulfonamido)-*N*-(3-fluoro-2-oxopropyl)-3-phenylpropanamide, **30b**. The secondary alcohol **29b** (0.200 g, 0.420 mmol) was oxidized to the corresponding ketone **30b** in accordance to the general oxidation procedure described above. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford **30b** as a yellow crystalline solid in 50% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 8.58 (d, *J* = 8.3 Hz, 1H), 8.20 (d, *J* = 7.2 Hz, 1H), 8.09 (d, *J* = 8.6 Hz, 1H), 7.59 – 7.43 (m, 2H), 7.21 (d, *J* = 7.5 Hz, 1H), 6.97 (m, 4H), 6.84 – 6.72 (m, 2H), 5.26 (d, *J* = 5.6 Hz, 1H), 4.97 (s, 1H), 4.87 (s, 1H), 4.29 (ddd, *J* = 20.0, 5.6, 1.8 Hz, 1H), 4.14 (ddd, *J* = 20.0, 4.7, 1.7 Hz, 1H), 3.96 (q, *J* = 6.7 Hz, 1H), 3.02 – 2.88 (m, 7H), 2.78 (dd, *J* = 14.1, 5.9 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 201.43, 201.26, 170.79, 134.53, 133.29, 130.98, 130.11, 129.30, 128.86, 128.74, 128.53, 127.19, 123.22, 115.45, 85.19, 83.73, 58.23, 46.63, 45.52, 37.95. mp = 75 – 80^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{24}H_{26}FN_3O_4S$: 471.1628 m/z (M+), found: 472.1719 m/z (M+H) = -34.0^{o} (c 1x10^{-3} g/mL, MeOH).



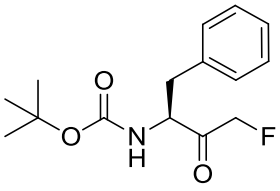
(*S*)-*tert*-butyl 4-chloro-3-oxo-1-phenylbutan-2-ylcarbamate, **33**. This compound was obtained following a literature procedure. All spectroscopic data match the data reported.[75]



tert-butyl (1-*S*)-1-(oxiran-2-yl)-2-phenylethylcarbamate, **36**. This compound was obtained following a literature procedure. All spectroscopic data match the data reported.[75]

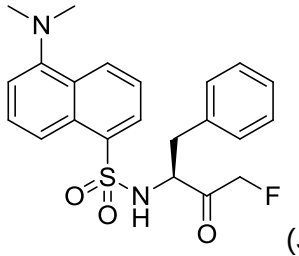


tert-butyl (2-*S*)-4-fluoro-3-hydroxy-1-phenylbutan-2-ylcarbamate, **37**. This compound was obtained following a literature procedure. All spectroscopic data match the data reported.[75]

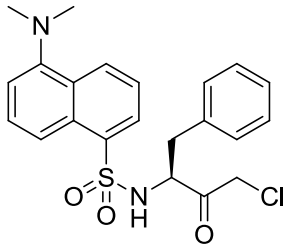


(*S*)-*tert*-butyl 4-fluoro-3-oxo-1-phenylbutan-2-ylcarbamate, **38**. A 25 mL round bottom flask was charged with **37** (0.030 g, 0.110 mmol) dissolved in CH_2Cl_2 (3 mL). Dess-Martin Periodinane (0.050 g, 0.120 mmol) was then added and the reaction mixture was allowed to stir overnight at r.t. Upon consumption of starting material (*via* TLC) the reaction was diluted with CH_2Cl_2 (10 mL) and washed with a saturated solution of aqueous $Na_2S_2O_3$ (4 x 10 mL) followed by brine (3 x 10 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford **38** as a white solid in 70% yield. ^{1}H NMR (400 MHz, $CDCl_3$)

δ_{ppm} = 7.38 – 7.26 (m, 3H), 7.24 – 7.14 (m, 2H), 5.01 (d, *J* = 16.5 Hz, 1H), 4.88 (dd, *J* = 16.5, 5.7 Hz, 1H), 4.76 (m, 1H), 3.15 (dd, *J* = 13.9, 6.4 Hz, 1H), 3.02 (d, *J* = 6.8 Hz, 1H), 1.43 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ_{ppm} = 171.15, 170.91, 135.48, 129.22, 128.85, 127.30, 60.40, 57.15, 37.20, 28.24, 21.06, 14.20.



(*S*)-5-(dimethylamino)-*N*-(4-fluoro-3-oxo-1-phenylbutan-2-yl)naphthalene-1-sulfonamide, **39**. *N*-Boc protected compound **38** (0.120 g, 0.440 mmol) was deprotected using trifluoroacetic acid as described in the general *N*-Boc deprotection procedure. The crude residue was resuspended in CH_2Cl_2 (1 mL) then added to a stirring solution of dansyl-chloride (0.120 g, 0.440 mmol) and diisopropylethylamine (0.150 mL, 0.870 mmol). The reaction was monitored *via* TLC looking at disappearance of starting dansyl-chloride. Upon complete consumption of starting material, the reaction was concentrated under reduced pressure. The crude residue was purified by flash column chromatography (12% EtOAc – 100% EtOAc in hexanes) to afford **39** as a yellow oil in 44% yield. ^{1}H NMR (500 MHz, $CDCl_3$) δ_{ppm} = 8.55 (d, *J* = 8.5 Hz, 1H), 8.18 – 8.10 (m, 2H), 7.56 – 7.47 (m, 2H), 7.20 (d, *J* = 7.5 Hz, 1H), 7.07 – 7.02 (m, 3H), 6.84 (dd, *J* = 7.6, 1.3 Hz, 2H), 5.32 (d, *J* = 8.0 Hz, 1H), 4.79 (d, *J* = 45 Hz, 2H), 4.36 (td, *J* = 7.8, 1.7 Hz, 1H), 2.92 (s, 6H), 2.87 (d, *J* = 5.9 Hz, 1H), 2.82 (dd, *J* = 13.9, 7.2 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ_{ppm} = 204.04, 151.97, 134.09, 133.67, 131.04, 129.86, 129.70, 129.36, 128.91, 128.64, 128.58, 127.33, 122.95, 118.55, 115.37, 59.75, 45.43, 37.73. HRMS (ESI, TOF) m/z calculated for $C_{22}H_{23}FN_2O_3S$: 414.1413 m/z (M+), 437.1311 m/z (M+Na), found: 415.1532 m/z (M+H), found: 437.1353 m/z (M+Na).



(*S*)-*N*-(4-chloro-3-oxo-1-phenylbutan-2-yl)-5-(dimethylamino)naphthalene-1-sulfonamide, **34**. *N*-Boc protected compound **33** (0.200 g, 0.670 mmol) was deprotected using trifluoroacetic acid as described in the general *N*-Boc deprotection procedure above. The crude residue was then resuspended in CH_2Cl_2 (1 mL) and added drop-wise to a stirring solution of dansyl-chloride (0.181 g, 0.671 mmol) and diisopropylethylamine (0.230 mL, 1.340 mmol) dissolved in 10 mL CH_2Cl_2. After 5 hr at r.t. the starting material was consumed and the reaction was diluted with 40 mL CH2Cl2 and washed with 50:50 brine:H_2O (3 x 10 mL), 1M aqueous HCl (2 x 10 mL), then brine (3 x 10 mL). The organic layer was dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude residue was purified by flash column chromatography to afford **34** as a yellow crystalline solid in 76% yield. ^{1}H NMR (500 MHz, DMSO) δ_{ppm} = 8.83 (d, *J* = 8.7 Hz, 1H), 8.37 (d, *J* = 8.5 Hz, 1H), 8.12 (d, *J* = 8.6 Hz, 1H), 7.80 (d, *J* = 7.9 Hz, 1H), 7.55 – 7.49 (m, 1H), 7.45 – 7.38 (m, 1H), 7.23 (d, *J* = 7.5 Hz, 1H), 6.93 – 6.85 (m, 5H), 4.62 (dd, J_{AB} = 20 Hz, J_{ABX} = 5 Hz, 2H), 4.15 (td, *J* = 9.4, 4.8 Hz, 1H), 2.95 (dd, *J* = 13.9, 4.7 Hz, 1H), 2.84 (s, 6H), 2.54 (dd, *J* = 9.1, 4.6 Hz, 1H). ^{13}C NMR (126 MHz, DMSO) δ_{ppm} = 200.83, 136.65, 136.01, 129.95, 129.45, 129.23, 129.17, 128.41, 128.13, 128.06, 126.64, 123.66, 119.75, 115.55, 61.50, 48.00, 45.57, 36.19. mp = 69 – 75^{o}C. HRMS (ESI, TOF) m/z calculated for $C_{22}H_{23}ClN_2O_3S$: 453.1016 m/z (M+Na), found: 453.1012 m/z (M+Na).

Chapter 5. Conclusions and future work

In conclusion, the body of work represented in this thesis highlights the development of useful assays to monitor ATG4B activity. The LC/MS assay offers broad utility by being able to detect covalent modifications occurring to a particular enzyme of interest. It was discovered through the use of LC/MS that compounds such as the CMK **7a** react multiple times with a single enzyme molecule while less reactive FMK **7b** only reacts once. This level of insight may not have been possible by looking at the time-dependent progress curves alone and serves to highlight some of the potential complications when investigating irreversible inhibitors. The FRET-based assay was designed with HTS applications in mind and was successfully adapted for such purposes, which was highlighted by the screening conducted in collaboration with the CDRD. The initial hits obtained from the screening effort did not provide any useful lead compounds, however did provide insight into the types of inhibitors that may be effective. Of the hits obtained, *Z*-*L*-Phe-CMK provided structural insights into the design of the first and second generation of inhibitors.

Small molecule irreversible inhibitors of ATG4B were designed, synthesized and tested for activity towards inhibiting ATG4B. Various kinetic analyses approaches were used and helped to validate our predictions about irreversible inhibition and covalent modifications occurring at various sites in and around ATG4B. A direct measure of k_{inact} is made possible by the use of LC/MS monitoring the rate of alkylation of ATG4B which is directly proportional to the rate of its inactivation, k_{inact}. This method alleviates some of the complicating factors when investigating fluorescent compounds using a fluorescence-based assay. This method of measuring k_{inact} is validated by the fact that these types of irreversible inhibitors are indeed active-site directed, as evidenced by the trypsin digest studies and protection experiments.

The next logical step for this research project would be to test these fluorescent compounds in the context of a whole cell and to monitor inhibition of autophagy with

respect to inhibition of ATG4B. Preliminary experiments conducted in collaboration with Dr. Sharon Gorski revealed that is was possible to label ATG4B within cells, but with varying degrees of success. It is speculated that the CMK's would be too reactive to be useful in a whole cell, while the FMK may not be reactive enough. Many other enzymes and proteases may be present in the context of a whole cell, particularly other ATG homologues. It is possible for the fluorescent compounds (**30a**, **30b**, **34**) to interact, alkylate, or inhibit any of the other cysteine proteases and proteins bearing free cysteine thiols. We also have undertaken preliminary work to determine if these irreversible inhibitors are selective and this exploratory study has already shown that CMK **7a** and FMK **7b** do indeed alkylate ATG4A in much the same way as they do for ATG4B. It is also speculated that other ATG4 homologs may become up-regulated or may simply take over once ATG4B has been inhibited due to redundant pathways. Currently, work is being done to optimize dosing regimes in order to achieve sufficiently high inhibitor concentrations within cells.

References

1. Mariño, G. *et al.* Human autophagins, a family of cysteine proteinases potentially implicated in cell degradation by autophagy. *J. Biol. Chem.* **278,** 3671–8 (2003).

2. Cuervo, A. M. Autophagy: in sickness and in health. *Trends Cell Biol.* **14,** 70–7 (2004).

3. Nguyen, T. G. *et al.* Development of fluorescent substrates and assays for the key autophagy-related cysteine protease enzyme, ATG4B. *Assay Drug Dev. Technol.* **12,** 176–89 (2014).

4. Yang, Z. & Klionsky, D. J. Eaten alive: a history of macroautophagy. *Nat. Cell Biol.* **12,** 814–22 (2010).

5. Komatsu, M. *et al.* Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441,** 880–4 (2006).

6. Hara, T. *et al.* Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441,** 885–9 (2006).

7. White, E. Deconvoluting the context-dependent role for autophagy in cancer. *Nat. Rev. Cancer* **12,** 401–10 (2012).

8. Amaravadi, R. K. *et al.* Principles and current strategies for targeting autophagy for cancer treatment. *Clin. Cancer Res.* **17,** 654–66 (2011).

9. Mancias, J. D. & Kimmelman, A. C. Targeting Autophagy Addiction in Cancer. *Oncotarget* **2,** 1302–1306 (2011).

10. Rosenfeldt, M. T. *et al.* P53 Status Determines the Role of Autophagy in Pancreatic Tumour Development. *Nature* **504,** 296–300 (2013).

11. Rao, S. *et al.* A dual role for autophagy in a murine model of lung cancer. *Nat. Commun.* **5,** 3056 (2014).

12. Qadir, M. a *et al.* Macroautophagy inhibition sensitizes tamoxifen-resistant breast cancer cells and enhances mitochondrial depolarization. *Breast Cancer Res. Treat.* **112,** 389–403 (2008).

13. Dragowska, W. H. *et al.* Induction of autophagy is an early response to gefitinib and a potential therapeutic target in breast cancer. *PLoS One* **8,** e76503 (2013).

14. Chittaranjan, S. *et al.* Autophagy inhibition augments the anticancer effects of epirubicin treatment in anthracycline-sensitive and -resistant triple-negative breast cancer. *Clin. Cancer Res.* **20,** 3159–73 (2014).

15. Rothe, K. *et al.* The core autophagy protein ATG4B is a potential biomarker and therapeutic target in CML stem/progenitor cells. *Blood* **123,** 3622–3634 (2014).

16. Shimizu, S. *et al.* Role of Bcl-2 family proteins in a non-apoptotic programmed cell death dependent on autophagy genes. *Nat Cell Biol* **6,** 1221–1228 (2004).

17. Falgueyret, J.-P. *et al.* Lysosomotropism of Basic Cathepsin K Inhibitors Contributes to Increased Cellular Potencies against Off-Target Cathepsins and Reduced Functional Selectivity. *J. Med. Chem.* **48,** 7535–7543 (2005).

18. Kabeya, Y. *et al.* LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J. Cell Sci.* **117,** 2805–12 (2004).

19. Weidberg, H., Shpilka, T., Shvets, E. & Elazar, Z. LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis. *EMBO J.* **29,** 1792–1802 (2010).

20. Novak, I. *et al.* Nix is a selective autophagy receptor for mitochondrial clearance. *EMBO Rep.* **11,** 45–51 (2010).

21. Klionsky, D. J. *et al.* Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* **8,** 445–544 (2012).

22. Webber, J. L., Young, A. R. J. & Tooze, S. a. Atg9 Trafficking in Mammalian Cells. *Autophagy* **3,** 54–56 (2014).

23. Chan, E. Y. W., Longatti, A., McKnight, N. C. & Tooze, S. a. Kinase-inactivated ULK proteins inhibit autophagy via their conserved C-terminal domains using an Atg13-independent mechanism. *Mol. Cell. Biol.* **29,** 157–71 (2009).

24. Tanida, I. *et al.* Apg7p/Cvt2p: A Novel Protein-activating Enzyme Essential for Autophagy. *Mol. Biol. Cell* **10,** 1367–1379 (1999).

25. Tanida, I., Tanida-Miyake, E., Komatsu, M., Ueno, T. & Kominami, E. Human Apg3p/Aut1p homologue is an authentic E2 enzyme for multiple substrates, GATE-16, GABARAP, and MAP-LC3, and facilitates the conjugation of hApg12p to hApg5p. *J. Biol. Chem.* **277,** 13739–44 (2002).

26. Chung, T., Phillips, A. R. & Vierstra, R. D. ATG8 lipidation and ATG8-mediated autophagy in Arabidopsis require ATG12 expressed from the differentially controlled ATG12A AND ATG12B loci. *Plant J.* **62,** 483–93 (2010).

27. Mizushima, N. Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *J. Cell Sci.* **116,** 1679–1688 (2003).

28. Thompson, A. R., Doelling, J. H., Suttangkakul, A. & Vierstra, R. D. Autophagic Nutrient Recycling in Arabidopsis Directed by the ATG8 and ATG12 Conjugation Pathways 1. *Plant Physiol.* **138,** 2097–2110 (2005).

29. Sugawara, K. *et al.* Structural basis for the specificity and catalysis of human Atg4B responsible for mammalian autophagy. *J. Biol. Chem.* **280,** 40058–65 (2005).

30. Li, M. *et al.* Kinetics comparisons of mammalian Atg4 homologues indicate selective preferences toward diverse Atg8 substrates. *J. Biol. Chem.* **286,** 7327–38 (2011).

31. Shu, C.-W. *et al.* Synthetic substrates for measuring activity of autophagy proteases: autophagins (Atg4). *Autophagy* **6,** 936–47 (2010).

32. Li, M., Chen, X., Ye, Q., Vogt, A. & Yin, X. A high-throughput FRET-based assay for determination of Atg4 activity. *Autophagy* **3,** 401–412 (2012).

33. Shu, C.-W. *et al.* High-throughput fluorescence assay for small-molecule inhibitors of autophagins/Atg4. *J. Biomol. Screen.* **16,** 174–82 (2011).

34. Voet, D. & Voet, J. *Principles in Biochemistry*. (Kaye Pace, 2010).

35. Gray, P. J. & Dugglebyt, R. G. Analysis of kinetic data for irreversible enzyme inhibition. *Biochem. J.* **257,** 419–424 (1989).

36. Kumanomidou, T. *et al.* The crystal structure of human Atg4b, a processing and de-conjugating enzyme for autophagosome-forming modifiers. *J. Mol. Biol.* **355,** 612–8 (2006).

37. Martinez-Lopez, N., Athonvarangkul, D., Mishall, P., Sahu, S. & Singh, R. Autophagy proteins regulate ERK phosphorylation. *Nat. Commun.* **4,** 2799 (2013).

38. Tanida, I., Ueno, T. & Kominami, E. Human light chain 3/MAP1LC3B is cleaved at its carboxyl-terminal Met121 to expose Gly120 for lipidation and targeting to autophagosomal membranes. *J. Biol. Chem.* **279,** 47704–10 (2004).

39. Hemelaar, J., Lelyveld, V. S., Kessler, B. M. & Ploegh, H. L. A single protease, Apg4B, is specific for the autophagy-related ubiquitin-like proteins GATE-16, MAP1-LC3, GABARAP, and Apg8L. *J. Biol. Chem.* **278,** 51841–50 (2003).

40. An, J. *et al.* A novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. *J. Med. Chem.* **52,** 2667–72 (2009).

41. Satoo, K. *et al.* The structure of Atg4B-LC3 complex reveals the mechanism of LC3 processing and delipidation during autophagy. *EMBO J.* **28,** 1341–50 (2009).

42. Falgueyret, J.-P. *et al.* Lysosomotropism of Basic Cathepsin K Inhibitors Contributes to Increased Cellular Potencies against Off-Target Cathepsins and Reduced Functional Selectivity. *J. Med. Chem.* **48,** 7535–7543 (2005).

43. Schoellmann, G. & Shaw, E. No Title. *Biochemistry* **2,** 252 (1963).

44. Shaw, E., Mares-Guia, M. & Cohen, W. No Title. *Biochemistry* **4,** 2219 (1965).

45. Powers, J. C., Asgian, J. L., Ekici, Ö. D. & James, K. E. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem. Rev.* **102,** 4639–4750 (2002).

46. Leung-Toung, R. *et al.* Thiol Proteases: Inhibitors and Potential Therapeutic Targets. *Curr. Med. Chem.* **13,** 547–581 (2006).

47. VX-799. at <http://adisinsight.springer.com/drugs/800016862>

48. Hayden, F. G. *et al.* Phase II , Randomized , Double-Blind , Placebo-Controlled Studies of Ruprintrivir Nasal Spray 2-Percent Suspension for Prevention and Treatment of Experimentally Induced Rhinovirus Colds in Healthy Volunteers. **47,** 3907–3916 (2003).

49. Yang, W. *et al.* MX1013, a dipeptide caspase inhibitor with potent in vivo antiapoptotic activity. *Br. J. Pharmacol.* **140,** 402–12 (2003).

50. Leung-toung, R., Li, W., Tam, T. F. & Karimian, K. Thiol-Dependent Enzymes and Their Inhibitors : A Review. **9,** 979–1002 (2002).

51. Schirmeister, T. & Klockow, a. Cysteine Protease Inhibitors Containing Small Rings. *Mini-Reviews Med. Chem.* **3,** 585–596 (2003).

52. Fukushima, K., Arai, M., Kohno, Y., Suwa, T. & Satoh, T. An epoxysuccinic acid derivative(loxistatin)-induced hepatic injury in rats and hamsters. *Toxicol. Appl. Pharmacol.* **105,** 1–12 (1990).

53. Tachikura, T. The teratogenic effects of E-64 on rat embryogenesis. *Acta Paediatr. Jpn.* **32,** 495–501 (1990).

54. Toyo-oka, T. *et al.* No Title. *Arzneimittelforschung* **34,** 671–675 (1986).

55. Eilon, G. F., Gu, J., Slater, L. M., Hara, K. & Jacobs, J. W. Tumor apoptosis induced by epoxide-containing piperazines, a new class of anti-cancer agents. *Cancer Chemother. Pharmacol.* **45,** 183–191 (2000).

56. Greenspan, P. D. *et al.* N-Arylaminonitriles as bioavailable peptidomimetic inhibitors of cathepsin B. *Bioorg. Med. Chem. Lett.* **13,** 4121–4124 (2003).

57. No Title. at <http://ag.arizona.edu/calpains/nomenclature.html>

58. Lubisch, W. *et al.* Benzoylalanine-Derived Ketoamides Carrying Vinylbenzyl Amino Residues: Discovery of Potent Water-Soluble Calpain Inhibitors with Oral Bioavailability. *J. Med. Chem.* **46,** 2404–2412 (2003).

59. Lubisch, W., Hofmann, H. P., Treiber, H. J. & Möller, A. Synthesis and biological evaluation of novel piperidine carboxamide derived calpain inhibitors. *Bioorg. Med. Chem. Lett.* **10,** 2187–2191 (2000).

60. Lubisch, W. & Möller, A. Discovery of phenyl alanine derived ketoamides carrying benzoyl residues as novel calpain inhibitors. *Bioorg. Med. Chem. Lett.* **12,** 1335–1338 (2002).

61. Gauthier, J. Y. *et al.* The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorg. Med. Chem. Lett.* **18,** 923–928 (2008).

62. Campbell, R. E. Fluorescent-protein-based biosensors: modulation of energy transfer as a design principle. *Anal. Chem.* **81,** 5972–9 (2009).

63. Shaner, N. C., Patterson, G. H. & Davidson, M. W. Advances in fluorescent protein technology. *J. Cell Sci.* **120,** 4247–60 (2007).

64. Bunt, G. & Wouters, F. S. Visualization of molecular activities inside living cells with fluorescent labels. *Int. Rev. Cytol.* **237,** 205–77 (2004).

65. U.S. National Library of Medicine. at <http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=588400>

66. Rauber, P., Angliker, H., Walker, B. & Shaw, E. The synthesis of peptidylfluoromethanes and their properties inhibitors of serine proteinases and cysteine proteinases. *Biochem. J.* **239,** 633–640 (1986).

67. Omura, K., Sharma, A. K. & Swern, D. Dimethyl sulfoxide-trifluoroacetic anhydride. New reagent for oxidation of alcohols to carbonyls. *J. Org. Chem.* **41,** 957–962 (1976).

68. Montalbetti, C. A. G. N. & Falque, V. Amide bond formation and peptide coupling. *Tetrahedron* **61,** 10827–10852 (2005).

69. Srinivasan, R., Uttamchandani, M. & Yao, S. Q. Rapid Assembly and in Situ Screening of Bidentate Inhibitors of Protein Tyrosine Phosphatases. *Org. Lett.* **8,** 713–716 (2006).

70. Ireland, R. E. & Liu, L. An improved procedure for the preparation of the Dess-Martin periodinane. *J. Org. Chem.* **58,** 2899–2899 (1993).

71. Frigerio, M. & Santagostino, M. A User-Friendly Entry to 2-Iodoxybenzoic Acid (IBX). *J. Org. Chem.* **64,** 4537–4538 (1999).

72. Sabitha, G., Babu, R. S., Rajkumar, M. & Yadav, J. S. Cerium(III) chloride promoted highly regioselective ring opening of epoxides and aziridines using

NaN(3) in acetonitrile: a facile synthesis of 1,2-azidoalcohols and 1,2-azidoamines. *Org. Lett.* **4,** 343–345 (2002).

73. Tamami, B. & Mahdavi, H. Synthesis of azidohydrins from epoxides using quaternized amino functionalized cross-linked polyacrylamide as a new polymeric phase-transfer catalyst. *Tetrahedron Lett.* **42,** 8721–8724 (2001).

74. Kühnel, E. *et al.* Mechanism of methyl esterification of carboxylic acids by trimethylsilyldiazomethane. *Angew. Chem. Int. Ed. Engl.* **46,** 7075–8 (2007).

75. Wang, D. *et al.* One-Carbon Chain Extension of Esters to r -Chloroketones : A Safer Route without Diazomethane. *J. Org. Chem.* 1629–1633 (2004).

76. Vezenkov, L., Honson, N.S., Kumar, N.S., Bosc, D., Kovacic, S., Nguyen, T.G., P. & T.A., Young, R. N. Development of fluorescent peptide substrates and assays for the key autophagy-initiating cysteine protease enzyme, ATG4B. *Bioorg. Med. Chem.* (2015).

77. Bioinformatics, S. I. of. PeptideCutter. at <http://web.expasy.org/peptide_cutter/>