STRUCTURE-BASED DRUG DISCOVERY AGAINST HUMAN ENL YEATS

DOMAIN AND SARS-COV-2 MAIN PROTEASE

A Dissertation

by

XINYU MA

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe Liu
Committee Members,	Tadhg Begley
	Kevin Burgess
	Pingwei Li
Head of Department,	Simon North

August 2021

Major Subject: Chemistry

ABSTRACT

Structure-based drug design is a drug discovery strategy where rational design of drug molecules take place based on the structural information of therapeutic targets. With the development of structural biology technologies such as protein crystallography and cryo-electron microscopy, which results in the availability of more and more proteins in a higher and higher resolution, structure-based drug design has become one of the most useful drug discovery strategy in both academia and pharmaceutical industry. This dissertation discusses applying structure-based drug design strategies in inhibitor development targeting ENL (eleven-nienteen leukemia) protein, which is an important protein in the *mix lineage leukemia* (MLL)-rearranged leukemia, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) main protease, a vital viral enzyme for its replication.

Chapter I is a brief introduction to the topics of this dissertation. Starting with a short introduction of the concept of structure-based drug design, it then mainly discusses the molecular mechanism of the pathogenesis and potential therapeutic targets of the following two diseases: MLL-rearranged leukemia and COVID-19.

Chapter II describes the development of a series of selective ENL YEATS domain inhibitors. ENL is a histone acetylation reader essential for disease maintenance in acute leukemias, especially the MLL-rearranged leukemia. The function of ENL is dependent on the recognition of histone acetylation by its YEATS domain, suggesting that inhibition of the ENL YEATS domain is a potential strategy to treat MLL-rearranged leukemia. In

our study, high-throughput screening of a small molecule library was carried out to identify inhibitors for the ENL YEATS domain. Structure–activity relationship studies of the hits and structure-based inhibitor design led to two compounds with IC_{50} values below 100 nM in inhibiting the ENL-acetyl-H3 interaction. Both compounds and their precursor displayed strong selectivity toward the ENL YEATS domain over all other human YEATS domains. One of these compounds also exhibited on-target inhibition of ENL in cultured leukemia cells and a synergistic effect with the BET bromodomain inhibitor JQ1 in killing leukemia cells. Together, we have developed selective inhibitors for the ENL YEATS domain, providing the basis for further medicinal chemistry-based optimization to advance both basic and translational research of ENL.

Chapter III and IV describes the development of SARS-CoV-2 main protease inhibitors and the assessment of their selectivity against host proteases. The COVID-19 pathogen, SARS-CoV-2, requires its main protease (SC2Mpro) to digest two of its translated long polypeptides to form mature viral proteins that are essential for viral replication and pathogenesis. Inhibition of this vital proteolytic process is effective in preventing the virus from replicating in infected cells and therefore provides a potential COVID-19 treatment option. Guided by previous medicinal chemistry studies about SARS-CoV main protease (SC1Mpro), we designed and synthesized a series of peptidyl aldehyde inhibitors that reversibly covalently bind to the active cysteine of SC2Mpro. The most potent compound has an IC_{50} of 8.3 nM. Crystallographic analysis confirmed the covalent linkage between the aldehyde inhibitors and active cysteine and showed structural rearrangement of the apoenzyme to accommodate the inhibitors. Two inhibitors

completely prevented the SARS-CoV-2-induced cytopathogenic effect in Vero E6 cells at 2.5–5 µM and A549/ACE2 cells at 0.16–0.31 µM.

Even though a number of inhibitors have been developed for the SARS-CoV-2 main protease as potential COVID-19 medications, little is known about their selectivity. Using enzymatic assays, we characterized inhibition of TMPRSS2, furin, and cathepsin B/K/L by 11 previously developed Mpro inhibitors. Our data revealed that all these inhibitors are inert toward TMPRSS2 and furin. Diaryl esters also showed low inhibition of cathepsins. However, all aldehyde inhibitors displayed high potency in inhibiting three cathepsins. A cellular analysis indicated high potency of MPI5 and MPI8 in inhibiting lysosomal activity, which is probably attributed to their inhibition of cathepsins. Among all aldehyde inhibitors, MPI8 shows the best selectivity toward cathepsin L. With respect to cathepsin B and K. MPI8 is the most potent compound among all aldehyde inhibitors in inhibiting SARS-CoV-2 in Vero E6 cells. Cathepsin L has been demonstrated to play a critical role in the SARS-CoV-2 cell entry. By selectively inhibiting both SARS-CoV-2 M^{Pro} and the host cathepsin L, MPI8 potentiates dual inhibition effects to synergize its overall antiviral potency and efficacy. Due to its high selectivity toward cathepsin L that reduces potential toxicity toward host cells and high cellular and antiviral potency, we urge serious consideration of MPI8 for preclinical and clinical investigations for treating COVID-19.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Wenshe Liu and Professor Tadhg Begley and Professor Kevin Burgess of the Department of Chemistry and Professor Pingwei Li of the Department of Biochemistry and Biophysics.

In Chapter II, the design and synthesis of inhibitors were carried out in collaboration with Dr. Sukant Das and Dr. Shiqing Xu. Cellular-based assays were conducted by Dr. Xiaobing Shi, Dr. Hong Wen and Dr. Longxia Xu. NMR titration experiments were carried out by Dr. Tatiana G. Kutateladze and Dr. Brianna J. Klein.

In Chapter III, the design and synthesis of inhibitors were completed in collaboration with Dr. Yugendar Reddy Alugubelli, Dr. Yuying Ma, Zhi Geng and Dr. Shiqing Xu. Expression and purification of SARS-CoV-2 main protease was carried out by Dr. Kai Yang. IC_{50} measurements were conducted in collaboration with Dr. Erol C. Vatansever. Crystallography experiments and analysis were carried out by Dr. Kai Yang. Cellular cytopathic effect assay was carried out by Dr. Chien-Te K. Tseng and Aleksandra K. Drelich.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

NOMENCLATURE

3CLpro	3C-like Protease
ACE2	Angiotensin-converting Enzyme 2
BET	Bromodomain and Extraterminal Domain
BRD	Bromodomain
CBP	Cereblon-binding Protein
Cryo-EM	Cryo-electron Microscopy
DCM	Dichloromethane
DMF	N,N'-dimethylformamide
DMSO	Dimethyl Sulfoxide
DOT1L	Disruptor of Telomeric Silencing 1-like
EC_{50}	Half Maximal Effective Concentration
EDTA	Ethylenediaminetetraacetic Acid
EMEM	Eagle's Minimal Essential Medium
ENL	Elevin Nineteen Leukemia
ESI-MS	Electrospray Ionization Mass Spectrometry
EtOAc	Ethyl Acetate
FBS	Fetal Bovine Serum
HCl	Hydrogen Chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
HPLC	High Pressure Liquid Chromatography

IC$_{50}$	Half Maximal Inhibitory Concentration
IPTG	Isopropyl β-1-thiogalactopyranoside
kb	kilo base
kDa	kilodalton
LC-MS	Liquid Chromatography-Mass Spectrometry
MeOH	Methanol
MLL	Mixed Lineage Leukemia
MPI	Main Protease Inhibitor
Mpro	Main Protease
mRNA	Messenger Ribonucleic Acid
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffered Saline
PDB	Protein Data Bank
PHD	Plant Homeodomain
PLpro	Papain-like Protease
PMSF	Phenylmethylsulfonyl Fluoride
PROTAC	Proteolysis Targeting Chimera
PTM	Post-translational Modification
SARS	Severe Acute Respiratory Syndrome
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SC1Mpro	Severe Acute Respiratory Syndrome Coronavirus Main Protease

SC2Mpro	Severe Acute Respiratory Syndrome Coronavirus 2 Main Protease
SUMO	Small Ubiquitin-like Modifier
THF	Tetrahydrofuran
Tris	Tris(hydroxymethyl)aminomethane

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Concepts of Structure-based Drug Design

1.1.1 Overview of Structure-based Drug Design Process

Structure-based drug design (SBDD) is a drug design process which mainly relies on the knowledge of the three-dimensional structure of a drug target, usually a biological macromolecule such as a protein. It was first introduced in 1980s[1]. Over the 40 years, as the completion of human genome project and the outburst of technologies in structural biology, proteomics, computational chemistry and cheminformatics, it has become one of the most important drug design strategies in pharmaceutical industry and academic research institutions.

The overall process of SBDD is an iterative one (**Figure 1**). The first step of SBDD is to identify a druggable target, also called a receptor, in most cases a protein. Then the three-dimensional structure of the drug target needs to be obtained, usually by using experimental methods such as X-ray crystallography or nuclear magnetic resonance (NMR). With the three-dimensional structure of the receptor, a potential ligand binding site should be identified. Once the structural information is ready, a list of hits can be generated from various methods. These hits need to be evaluated in various biochemical assays for its potency against the drug target, as well as the structural information of how they bind to the target. According to the feedback, the hits will need to be further optimized

for stronger binding affinity and better potency, which are called leads, before moving

into the next stage of preclinical studies and clinical studies. It may take several rounds of

such feedback cycles to achieve a satisfying drug candidate.
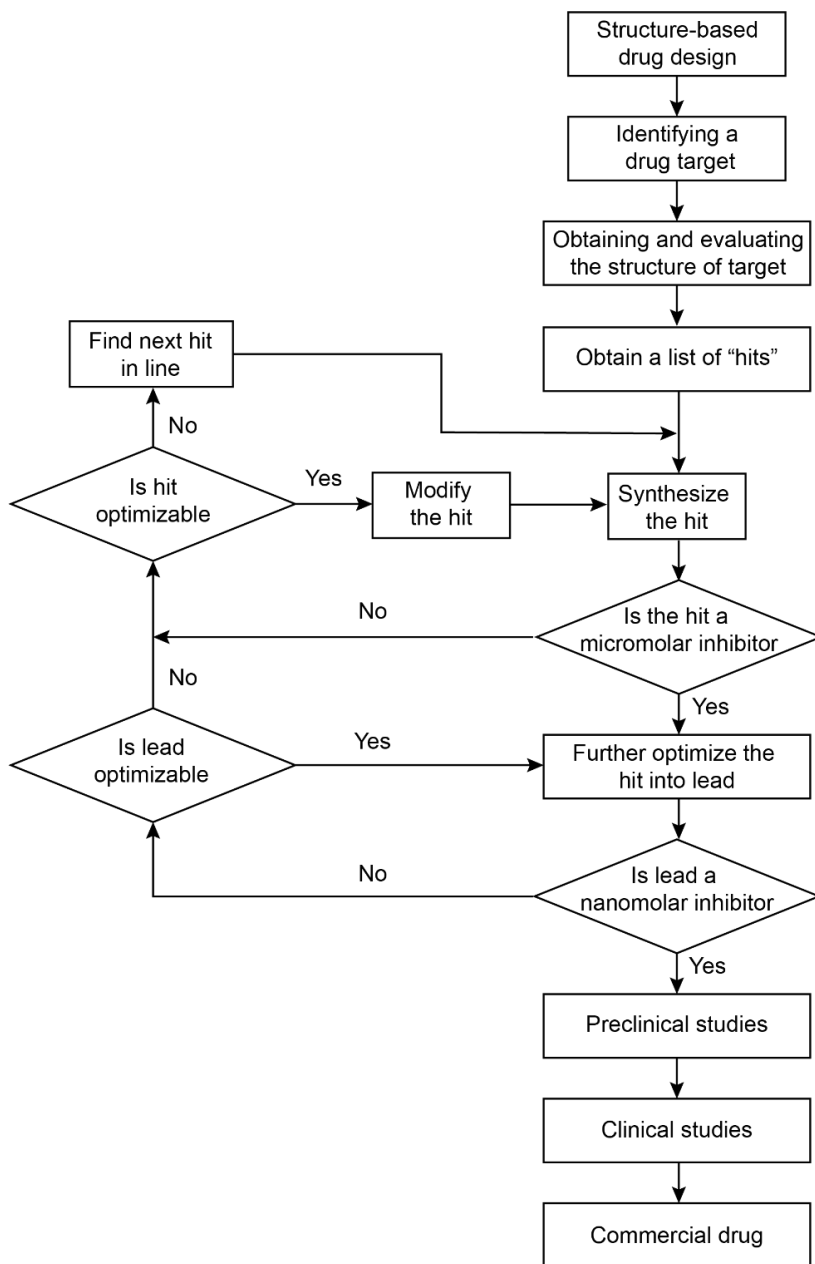
Figure 1. A schematic flowchart for the process of structure-based drug design (modified from reference 1).

1.1.2 Identifying a Drug Target

As the first step of SBDD, the identification of a meaningful and druggable target is apparently fundamental to a successful drug design process. This step is usually built on a biological or biochemical basis. An ideal drug target is usually a protein that is closely related to human phycology or diseases while its functions can be modulated upon the binding of a ligand. The most common proteins that obviously meet these conditions are enzymes, receptors, and ion channels, which consists of most known drug targets. Their functions rely on the intrinsic binding of native ligands (substrates, cofactors, or signal molecules) and therefore can be modulated by analogues relatively easily.

Besides enzymes and receptors, more and more protein-protein interactions are also considered potential drug candidates recently. By using a small molecule ligand to block the protein-protein interaction interface, the downstream signal transduction can be stopped, inducing or suppressing corresponding biological effects. Such examples are particularly common among transcription factors, which are recruited to and usually only function as part of large protein complexes.

Nevertheless, there are still many proteins that do not have a function-related binding pocket, which means the function of the protein cannot be altered even though a high affinity ligand is bound to it. These proteins were traditionally called non-druggable targets. However, with emerging induced protein degradation techniques such as proteolysis-targeting chimera (PROTAC), these proteins can be simply degraded specifically by the induction of drugs. A typical example for this type of targets includes the famous Alzheimer-causing *tau* protein.

1.1.3 Obtaining and Evaluating the Structure of Drug Target

Once a drug target is picked, the next step is to obtain the three-dimensional structure of it. The accuracy of the structures is of great importance to SBDD since all the following design is based on it. The most common method to obtain high-resolution three-dimensional protein structure is still X-ray crystallography, despite it has a history of over half a century. Additionally, newer experimental methods such as nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM), as their resolution getting higher and applicability getting broader, are also being more and more used in the determination of protein structure for SBDD purposes. Besides experimental methods, computational modeling is also a potential method for obtaining three-dimensional structures of drug targets. Although the accuracy of computationally modeled protein structure is still a concern and not widely accepted as a reliable source of protein structure, given the rapid development of AI or machine learning technologies, it is still likely that AI-aided computational methods would be another important source of protein structure in the future.

With an accurate protein structure, the next step for SBDD is to identify the binding site of the drug target. For enzymes and receptors, it usually means the active site of the enzyme or receptor where substrates, cofactors or signal molecules bind to. The most straightforward methods to identify the binding site is to obtain the structure of substrate-bound, cofactor-bound or signal molecule-bound protein through various structural biology methods. Another kind of binding site is allosteric binding site. Allosteric binding sites are not directly related to the binding of substrates, cofactors or signal molecules.

However, when a ligand binds to an allosteric binding site, it triggers conformational change of the protein which makes the active site unable to acceptor their intrinsic ligands. The identification of allosteric binding site is less likely to be a rational process but more an occasional discovery, where in many cases is preceded with the discovery of an allosteric ligand. Other than enzymes and receptors, protein-protein interaction (PPI) is also an important, however, more challenging drug target. Unlike enzymes or receptors whose active sites are usually well-defined, PPI interfaces are usually large, flat and flexible surfaces, which makes them traditionally consider "undruggable" targets. However, many PPIs are characterized with "hot-spots", which means one or a few amino acid residues play the most vital role in the interaction. These "hot-spots" can be starting point for structure-based design[2].

1.1.4 Drug Design Methods

After the structure and binding site of a drug target are identified, the next step is the identification of lead compounds. Various strategies are available in this process, including traditional medicinal chemistry strategy that is guided by chemical modification of a naturally existing ligand, high-throughput screening techniques and more recent computer-aided strategies. In most cases, these strategies are used in combination to achieve the best outcome of drug discovery.

Modifying an Existing Compound

Enzymes and receptors have their intrinsic ligands. These ligands provide medicinal chemists with a good chemical template for modifications and optimizations. A

5

typical approach of this strategy is to convert an intrinsic ligand into a toxic ligand which occupies the active site but does not undergo the normal biological function (enzymatic reaction for an enzyme or signal transduction for a receptor). Before the application of modern drug discovery technologies such as high-throughput screening or computer-aided drug design, this approach is the primary drug discovery approach and is still an important approach in modern drug discovery.

High-throughput Screening

With the development of combinatorial chemistry, automated synthesis and compound evaluation, the preparation and evaluation of vast compound libraries is made possible. This lays down the basis of high-throughput screening. High-throughput screening greatly enlarges the chemical spaces of lead compounds from structural analogues of intrinsic ligands to a wide variety of chemical structures that can hardly be rationally identified with current understanding.

Molecular Docking and Virtual Screening

Molecular docking and virtual screening are now important supplementary methods for the identification of lead compounds in drug discovery. Its most obvious advantage is the convenience since it does not require any wet-lab experimentations. However, it comes with a price of relatively low accuracy and low reliability compared with other strategies. Nevertheless, with the development of computational chemistry and more in-depth understanding of molecular dynamics, a more accurate and dynamic protein-ligand binding model could be developed, which could greatly boost the accuracy of molecular docking and facilitate more efficient drug design.

6

1.2 MLL-rearranged Leukemias and Targeted Therapies

1.2.1 MLL1 Gene and MLL-rearranged Leukemias

Mixed-lineage leukemia 1 (MLL1) gene is located on chromosome 11q23. It is often found to be disrupted in a specific group of acute leukemia, called MLL-rearranged leukemia (MLL-r leukemia), which makes up 80% of acute leukemia cases in infants and 10% of all leukemia cases. The MLL-r leukemia, being one of the most aggressive and deadly leukemia subtypes, usually has much higher resistance to traditional chemotherapies compared with non-MLL-rearranged leukemias, leading to a low 5-year survival rate. Besides the lives it claims, those who survived also suffer from long-term adverse effects due to the intensive treatment they received and high risk of relapse[3].

Normal MLL1 gene expresses a 431-kDa nuclear protein with histone methyltransferase activity. The structure of MLL was first depicted by Tkachuk *et al*. and Gu *et al*[4, 5]. On its N-terminal side, MLL contains a Menin binding domain, which plays an important role in the functions of MLL fusion proteins and the initiation of leukemogenesis[6]. The N-terminus also contains AT-hook motifs for DNA binding, speckled nuclear localization domains 1 and 2 (SNL-1 and SNL-2), and two repression domains (RD1 and RD2). The middle part of MLL contains four plant homeodomain (PHD) fingers and a bromodomain. The C-terminal part of MLL contains a transcriptional activating domain (AD) and a SET domain, which is the catalytical domain for its methyltransferase activity (**Figure 2**).
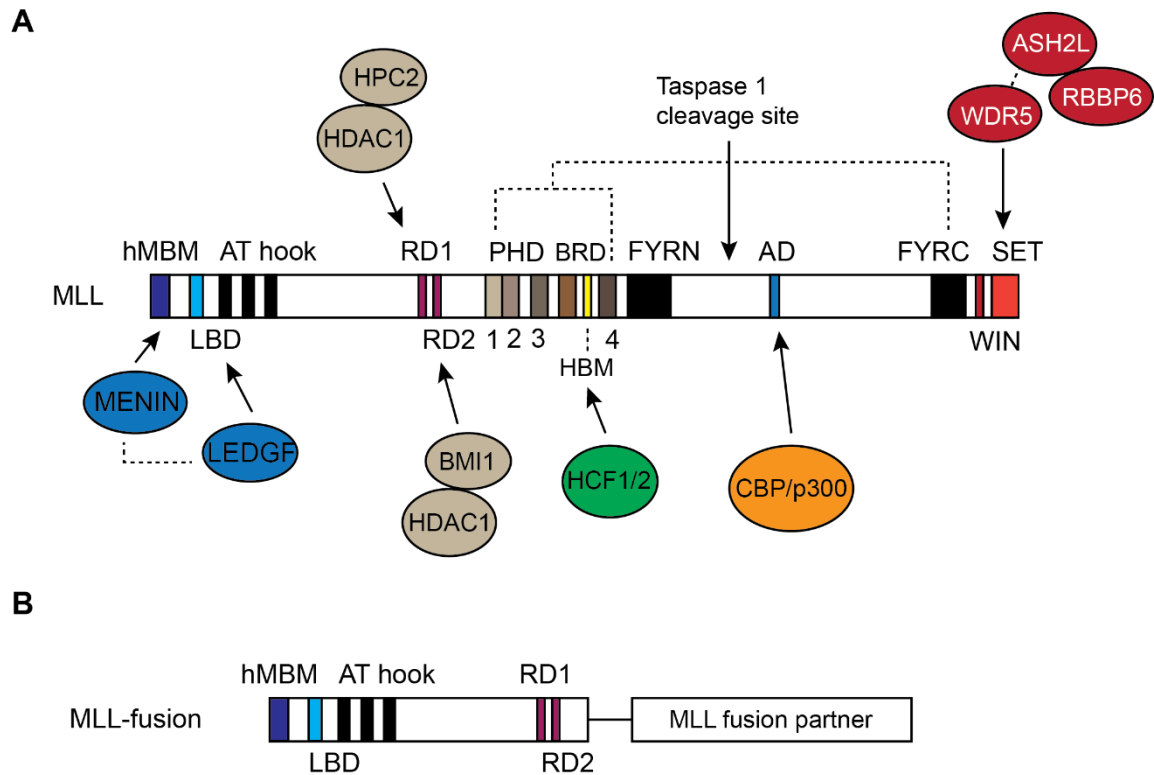
Figure 2. Overview of MLL and MLL-fusion protein structures (modified from reference 8). **(A)** Structure of native MLL protein. The interactions between MLL and other proteins are shown by arrows and dashed lines. **(B)** Structure of MLL-fusion proteins.

After the translation, the MLL polypeptide is proteolytically cleaved by Taspase 1 into two polypeptides[7], named MLL^N and MLL^C. These two polypeptides form a complex through the interaction between PHD domains and FYRN domain on MLL^N and the FYRC domain on MLL^C. The proteolysis and the formation of MLL^N-MLL^C complex is the first step to the maturation and stabilization of MLL. Further, the complex associates with many other cofactors to form a larger biological complex which plays essential role in gene transcription. These cofactors include MENIN, HCF1/2, ASH2L, RBBP6, WDR5 CBP/p300, etc. (**Figure 2A**)[8].

In MLL-r leukemias, due to chromosomal translocation, MLL protein lost its C-terminal part, including all four PHD fingers, bromodomain, activating domain and SET domain[9]. The N-terminal part, including AT-hook DNA binding motifs and repression domains, is found to be fused with over 60 different kinds of proteins, including nuclear proteins, cytoplasmic proteins, and membrane-bound proteins[10]. Among all these fusion partners of MLL, nuclear proteins, such as AF4, AF9, ENL, which directly or indirectly recruit components for the cellular transcriptional machinery, accounts for the vast majority of MLL-r leukemias[11] (**Figure 2B**).

1.2.2 Molecular Basis for MLL-rearranged Leukemogenesis

MLL is required for the expression of several posterior homeobox (*Hox*) genes, some of which are known to facilitate the proliferation of immature hematopoietic progenitors[12, 13]. MLL fusion proteins, generally targeting the same set of genes as wild-type MLL, constitutively activates the transcription of these genes, leading to over-proliferation of immature hematopoietic progenitors and causing leukemia eventually[14]. It is predicted that the MLL portion of MLL fusion proteins plays the most important part in the recognition of target genes, most likely attributing to the N-terminal Menin binding domain[6], while their fusion partners causes the constitutive transcriptional activation[8]. Despite of their huge diversity, the fusion partners of MLL can be generally categorized into four types according to their molecular mechanisms of leukemogenesis.

Recruitment of AEP Complex

One type of the MLL fusion partners is AF4 family proteins (including AF4, AF5q31 and LAF4) and ENL family proteins (ENL and AF9), which accounts for the most cases in MLL-r leukemias[10, 15]. These proteins are all members of the AF4 family/ENL family/P-TEFb complex (AEP complex), which plays an important role in the transcriptional elongation process and is also referred to as super elongation complex (SEC)[16-18]. The fusion between AEP complex members and MLL caused the recruitment of other AEP proteins and the formation of MLL-AEP hybrid complex to MLL-targeted genes and hence the abnormal transcription of those genes, which leads to leukemia[19] (**Figure 3**).
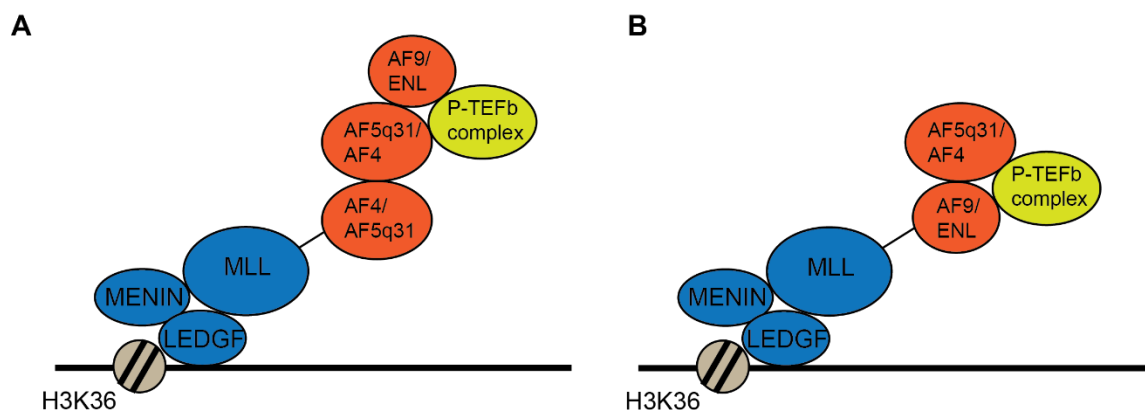
Figure 3. Formation of MLL/AEP hybrid complex (modified from reference 8). (A) MLL/AEP hybrid complex formed through fusion with AF4 family proteins. **(B)** MLL/AEP hybrid complex formed though fusion with ENL family proteins.

Recruitment of DOT1L Complex

AF10 family proteins are another type of commonly seen MLL fusion partners, which makes up approximately 8% of total MLL-r leukemia cases[10]. AF10 and its homologue AF17 are known to form a complex with histone H3K79 methyltransferase DOT1L[20]. There is evidence showing that the methyltransferase activity of DOT1L plays

a critical part in the leukemogenesis of MLL-AF10 fusion proteins[21]. Inhibiting DOT1L activities also showed promising results in suppressing MLL-r leukemia cells[22]. It is noteworthy that ENL family proteins can also form complex with DOT1L and AF10[23, 24], but the leukemogenic transformation of MLL-ENL/AF9 is primarily caused by MLL-AEP hybrid complex[24]. Nevertheless, DOT1L also indirectly facilitate MLL-ENL/AF9-mediated-leukemogenesis through H3K79 methylation in MLL-AF9/ENL targeted loci[25-27] (**Figure 4A**).
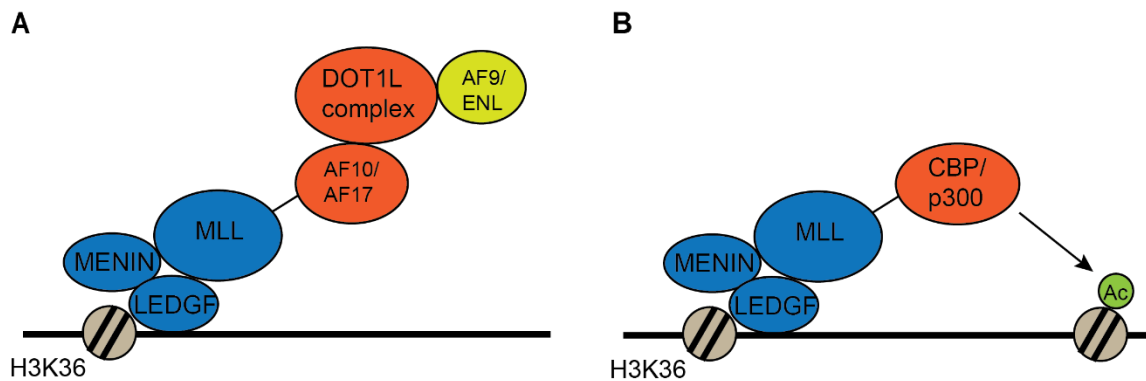
Figure 4. **Formation of MLL/DOT1L hybrid complex (A) and MLL-CBP/p300 fusion protein (B) (modified from reference 8).**

Mimicry of MLL Active Form by CBP/p300 Association

CBP/p300 family proteins are also found to be fused with MLL in rare cases of MLL-associated leukemias. MLL-CBP family fusion proteins transform hematopoietic progenitors through the bromodomain and histone acetyltransferase domain of CBP[28], presumably by enhancing histone acetylation at MLL target chromatin loci which leads to higher chromatin accessibility. Besides, MLL is also found to fuse with CBP binding proteins, such as AFX, in which the binding between AFX and CBP is crucial for the leukemogenesis[29]. Since wild-type MLL also associate with CBP[16, 30, 31], it is likely that

this category of MLL fusion proteins confer leukemic effects through the mimicry of wild-type MLL activity by the involvement of CBP family proteins[8, 32] (**Figure 4B**).

Fusion-partner-mediated MLL Dimerization

Some MLL fusion partners contain dimerization or oligomerization domains. When MLL is fused with these proteins, the dimerization or oligomerization between the fusion partner portion results in the dimerization or oligomerization of the entire MLL fusion protein, which confers leukemogenic transformation[33-36]. Such mechanism might explain the vast diversity of MLL fusion partners, especially those cytoplasmic proteins which have little involvement in the regulation of gene transcription. The most common fusion partner under this category is AF6, consisting of 4% of all MLL-r leukemia[10]. Although the molecular mechanism of how MLL dimerization triggers oncogene expression is still elusive, there is evidence indicating the transforming ability of MLL-AF6 is also related to DOT1L and AEP complexes[15, 37] (**Figure 5**).
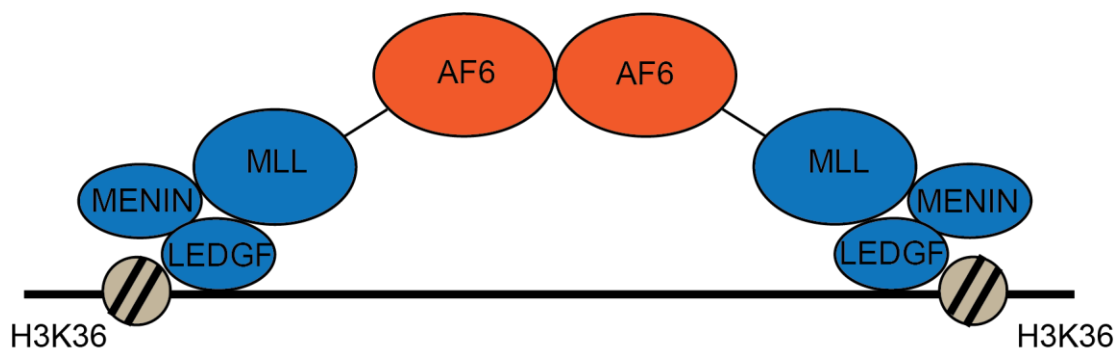
Figure 5. Formation of MLL-AF6 fusion protein and mediated MLL dimerization (modified from reference 8).

1.2.3 Development of Targeted Therapeutics for MLL-rearranged Leukemias

MLL-r leukemias typically are more resistant to traditional chemotherapies than other types of leukemias and are notorious for their poor outcomes of treatment. Thus, targeted therapeutics for the treatment of MLL-r leukemias are in great need.

With improvement in the understanding of pathogenesis of MLL-r leukemias, there have been several targeted therapeutics specifically developed for MLL-r leukemias. These inhibitors include agents that directly disrupt the interacting between Menin and MLL[38, 39], or interfere with the formation or the function of the transcriptional complexes that are recruited to MLL target genes, such as DOT1L inhibitors[22, 40-42], BET inhibitors[43-45], CBP/p300 inhibitors[46, 47] and LSD-1 inhibitors[48-51]. Besides, some agents targeting other non-MLL-specific pathways are also found effective against MLL-r leukemias, including BCL-2 inhibitors[52], FLT3 inhibitors[53] and etc. This introduction will mainly focus on the development of inhibitors that target pathways more specific to MLL rearrangement.

Menin-MLL Inhibitors

Menin binds to MLL through a five amino-acid binding motif on the N-terminal section of MLL, which is usually preserved in MLL fusion proteins. Inhibitors that interrupt the interaction between MLL and Menin can prevent MLL from recognizing and activating downstream oncogenes and thus inhibit the proliferation of MLL-r leukemia cells. The first Menin-MLL interaction inhibitor was reported by Grembecka *et al* in 2012, where a small molecule named ML-2 with piperazinyl pyrimidine fragment showed sub-micromolar binding affinity to Menin[38]. Treating MLL-AF9/ENL leukemia cells with the compound also inhibited cell proliferation and downregulated the expression of *Hoxa*9

13

and Meis1 genes, which are known MLL-targeted oncogenes. These results demonstrated the feasibility of Mein-MLL interaction as a therapeutic target.

Based on the results of ML-2 and the crystal structure of the ML2-MLL complex[54], structural based drug development leads to a much more potent inhibitor named VTP50469 (**Figure 6**), which also showed good efficacy in animal models and PK/PD properties[39]. Currently, a close analogue of this compound, SNDX-5613 (**Figure 6**) is under clinical trial for the treatment of MLL-r leukemia.
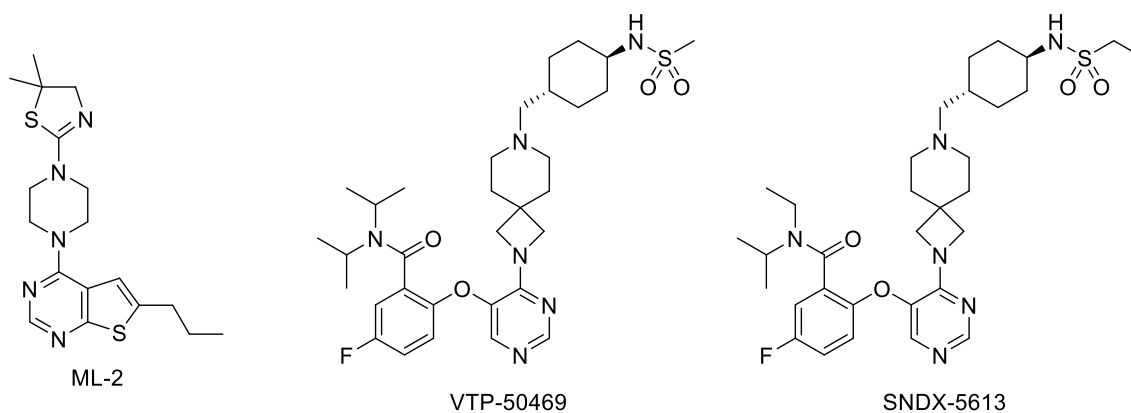
Figure 6. Chemical structures of ML-2, VTP-50469 and SNDX-5613.

DOT1L Inhibitors

DOT1L is a histone H3K79 methyltransferase that uses S-adenosylmethionine (SAM) as its methyl donor. As mentioned above, it can form complexes with several MLL fusion partners and thus plays an important role in MLL-related leukemogenesis[11]. According to the enzyme mechanism and the structure of its active pocket, early inhibitor development focused mostly on chemical analogues of SAM. Among them, EPZ004777 (**Figure 7**) was the first DOT1L inhibitor that showed anti-proliferative activity against MLL-r leukemia cells[22]. Based on that result, further development gave a highly potent

14

and selective inhibitor named pinometostat (EPZ-5676, **Figure 7**), which displayed rapid anti-leukemia efficacy in mouse model[40]. Its phase I clinical trial has shown some moderate efficacy toward a subset of MLL-r leukemia patients[41] and combination therapies with other anti-cancer agents are also under evaluation.

One of the drawbacks of SAM-based DOT1L inhibitors is the low bioavailability due to the adenosyl fragment in a molecule, which renders them unsuitable for oral administration. To overcome the pharmacokinetic limitations of SAM-based inhibitors, Stauffer *et al* applied a high-throughput screening (HTS) strategy and identified a hit with single-digit micromolar binding affinity to DOT1L[42]. Instead of binding to the active pocket of DOT1L, this compound binds to an adjacent pocket to DOT1L active site and is still mutually exclusive with SAM. Further structural optimization of this hit lead to a potent, bioavailable inhibitor (compound 10, **Figure 7**) that displayed promising anti-leukemia efficacy in cells and mouse models.
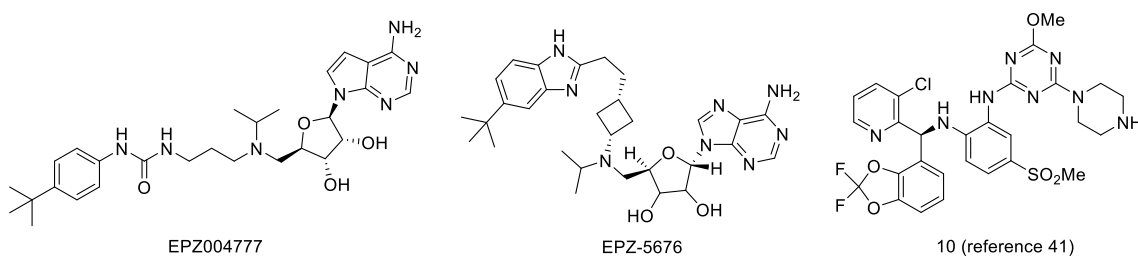
Figure 7. Chemical structures of EPZ004777, EPZ-5676 and compound 10.

BET Inhibitors

Bromodomains (BD) are important epigenetic readers that specifically recognize lysine acetylation on histones, which usually marks the accessibility of chromatin and activation of gene transcription. Bromodomain containing protein 4 (BRD4), a member of

15

bromodomain and extra-terminal (BET) family consisting of four members (BRD2, BRD3, BRD4 and BRDT) is found to play key roles in MLL-r leukemia, presumably through the association with MLL-fusion-related complexes and activation of oncogene transcription[43, 44].

The earliest BET inhibitors were designed based on diazepine core structures, including JQ-1[55] and I-BET[56] (**Figure 8**). In subsequent studies, JQ-1 showed sub-micromolar activity against the proliferation of multiple leukemia cell lines including those with MLL rearrangement[43]. Despite of its high potency, JQ-1 has not been evaluated in any clinical trials due to its short half-life *in vivo*, but it has been widely used as a chemical probe to study the biological functions of BET proteins. Soon after JQ-1, I-BET151 was reported as another small molecule BET protein inhibitor[44]. I-BET151 has a non-diazepine core structure but also binds to bromodomains the same manner as JQ-1. It also showed high efficacy selectively against MLL-fusion leukemia cells.

In 2013, Fish *et al* and Picaud *et al* identified a chemical probe named PFI-1 (**Figure 8**) with a novel dihydroquinazoline-2-one structure as a potent BET inhibitor[45, 57]. Compared with JQ-1 and I-BET151, PFI-1 adopts a slightly different binding pattern, although they all occupy the same N-acetyl lysine recognition pocket. *In vitro* and *in vivo* studies demonstrated its efficacy against several leukemia cell lines with MLL-AF4 and MLL-AF9 fusions.

It is noteworthy that BET inhibitors showed preclinical indications in a wide diversity of cancers and some non-oncology diseases such as diabetes and cardiovascular diseases, which implies BET family proteins play complicated roles in different cells and

tissues[58, 59]. This indicates the selectivity among different BDs in BET family should be an important factor to consider in the development of new inhibitors.
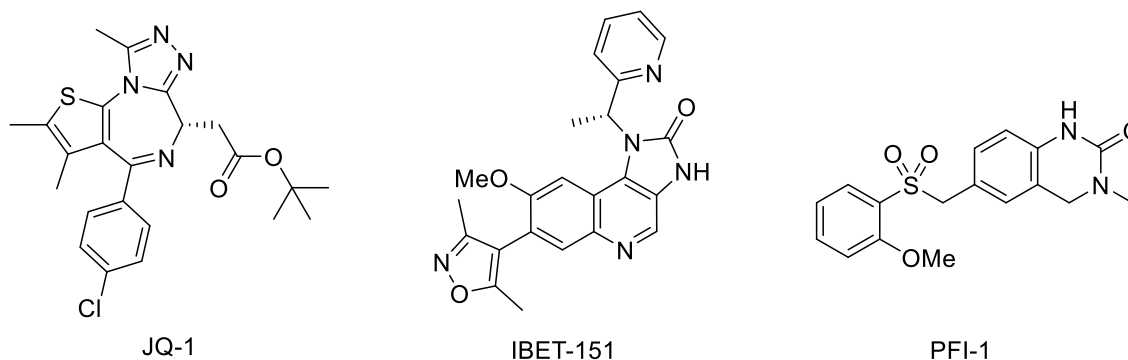
Figure 8. Chemical structures of JQ-1, IBET-151 and PFI-1.

CBP/p300 Inhibitors

CBP/p300 family proteins are two closely related histone acetyltransferases (HAT), containing a bromodomain and a catalytical HAT domain. In leukemias with MLL-CBP/p300 fusion, the bromodomain and the HAT domain play key roles in maintenance of the leukemogenic transformation ability of the fusion protein[28]. This inspires drug development by targeting the bromodomain or HAT domain of CBP/p300.

Based on previously reported BET bromodomain inhibitors such as JQ-1, which also showed weak inhibition of CBP/p300 bromodomain, Popp *et al* discovered I-CBP112 (**Figure 9**) as a selective inhibitor for CBP/p300 bromodomain through screening a commercial library and a comprehensive structure-activity relationship study[60]. I-CBP112 has a 134 nM affinity to CBP bromodomain. With a 2,3,4,5-tetrahydro-1,4-benzoxazepine backbone, it functions similarly to other BET inhibitors by mimicking N-acetyl lysine in the binding pocket. Picaud *et al* performed *in vitro* and *in vivo* assays. Although I-CBP112 did not show immediate cytotoxicity to leukemia cells, it inhibited

17

the clonogenic growth of multiple leukemia cell lines including the one with MLL-CBP fusion and induced the morphological change of cells, implying I-CBP112 might induced differentiation of those leukemia cells[46].
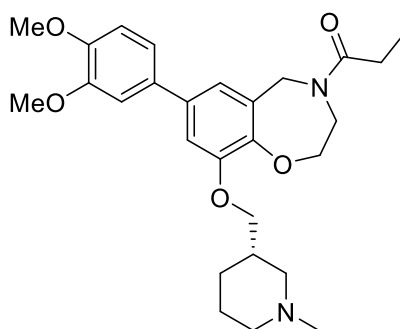
ICBP-112

Figure 9. Chemical structure of ICBP-112.

LSD1 Inhibitors

LSD1 is a flavin-dependent monoamine oxidase that specifically demethylates H3K4 and H3K9. Despite that no evidence shows LSD1 directly associates with MLL or MLL-related complexes, it is speculated that its chromatin-modulating function might make it a potential drug target for MLL-r leukemia as well. In 2012, Harris *et al.* reported that tranylcypromine (TCP, a known monoamine oxidase inhibitor that is used for depression treatment, **Figure 10**) and their more selective analogues impaired clonogenic potential and induced differentiation of MLL-r leukemia cells[48], demonstrating the feasibility of this speculation.

With this starting point, more potent and selective LSD1 inhibitors were developed based on the structure of TCP. In 2016, Feng *et al* reported a series of compounds that showed low nanomolar IC_{50}s *in vitro* against LSD1 specifically (**Figure 10**)[49]. These

compounds all bear the same N-substituted piperazinyl amide moiety on the original primary amine of TCP and different 4-position substitutions on the phenyl ring. Among them, compound 1 with a flexible 4-benzyl substitution showed the highest potency with an IC_{50} of 9.8 nM and similar EC_{50} for MLL-rearranged leukemia cells, while their EC_{50} against non-MLL leukemia cells were over 20 µM[49]. In 2018, Maes *et al.* reported ORY-1001 (**Figure 10**), also derived from TCP, as another potent and selective LSD1 inhibitor that showed strong inhibition both *in vitro* and *in vivo* against MLL-rearranged leukemia cells[50]. The mechanism of action of LSD1 inhibitors in leukemia cells were further confirmed by the gene-specific increase of H3K4me2 level. This compound together with azacitidine as a combined therapy is currently under phase II clinical trial for the treatment of AML.

tranylcypromine   compound 1 (reference 48)   ORY-1001
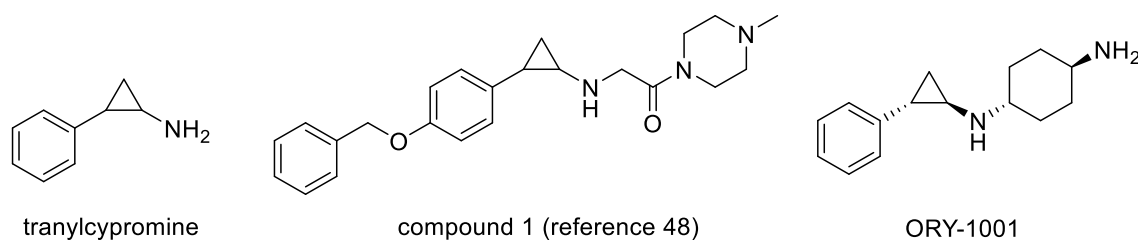
Figure 10. Chemical structures of tranylcypromine, compound 1 and ORY-1001.

1.2.4 Summary

MLL-rearranged leukemia is one of the most aggressive subtypes of leukemia. From the molecular basis, the pathogenic mechanism is due to the formation of abnormal MLL-fusion proteins caused by chromosomal translocation. These MLL-fusion proteins

19

interfere with the transcription of genes that control the proliferation and differentiation of immature hematopoietic progenitors, eventually leading to leukemias.

There are over 60 MLL-fusion partners that have been identified. Most of them are nuclear proteins that are part of cellular transcription machinery. Through identifying different MLL-fusion partners and understanding the biological functions they play in MLL-related leukemogenesis, specific targeted therapies can be developed for the treatment of MLL-r leukemias. However, due to the diverse and complex nature of MLL-related leukemogenesis, many of the biological processes remain to be elucidated, which leaves great challenge and space for further research and drug development.

1.3 SARS-CoV-2 and Its Main Protease as Drug Target

1.3.1 COVID-19 and SARS-CoV-2 Life Cycle

The COVID-19 pandemic emerged in December 2019 has greatly influenced the globe since its outbreak. By June 2021, it has affected over 200 countries and caused over 3 million of deaths in the world. Despite of the successful prevention of wide transmission in some countries due to strict lockdown policies or rapid vaccination for the public, the much of global population is still under immediate threaten of this deadly virus, especially in developing countries in Africa and southeastern Asia due to poor infrastructure, lack of public health resources and limited supply of vaccines. Recent reports of highly contagious mutants of the virus also indicate the severity of the situation. Under such an urgency, it is very unlikely that the global society can stop the pandemic purely relying on

vaccination without the combination with effective COVID-19 treatments. The development of anti-SARS-CoV-2 therapies is of great importance and urgency right now.

SARS-CoV-2 is the pathogen that causes COVID-19. It is an RNA virus that use positive single-stranded RNA as its genome. The genome of SARS-CoV-2 has a length of 29.9 kb, which shares about 82% sequence identity with SARS-CoV and $> 90\%$ sequence identity for essential enzymes and structural proteins[61]. SARS-CoV-2 has four structural proteins: spike (S) protein, envelope (E) protein, membrane (M) protein and nucleocapsid (N) protein. These proteins share high sequence identity with that of SARS-CoV and MERS-CoV.

SARS-CoV-2 relies on its Spike protein to recognize its host cell-surface receptor, ACE2 receptor, through the receptor binding domain (RBD). Upon binding, Spike protein is preactivated by membrane serine proteases TMPRSS2 and furin which facilitates the fusion of virus into host cell membrane and its endocytosis[62, 63]. In endosomes and lysosomes, Spike protein is further digested by proteases, primarily cathepsin L[64], which results in the release of viral RNA into cytosol. Then the viral RNA will hijack the host ribosomes for translation, producing polypeptides that will undergo proteolytic cleavage to form 15 essential non-structural proteins (NSP, **Table 1**). These NSPs will then start to function in different aspects of the replication and release of virus, including the replication of viral RNA, proteolytic maturing of viral proteins, the packaging of and the exocytosis of new virions. Among these NSPs, Nsp3 (papain-like protease), Nsp5 (main protease) and Nsp12 (RNA-dependent RNA polymerase) are proposed to play the most essential roles in the replication process and thus are the most studied and the most

promising therapeutic targets for the development of antiviral agents against SARS-CoV-2[61].

Table 1. List of non-structural proteins in SARS-CoV-2 and their molecular functions[61].

Protein Name	Description and Proposed Function
Nsp1	N-terminal product of the viral replicase
Nsp2	Replicase product essential for proofreading viral replication
Nsp3	Papain-like proteinase, responsible for cleavage of the translated polyprotein into distinct proteins
Nsp4	A membrane-spanning protein believed to anchor the viral replication-transcription complex to modified ER membranes
Nsp5	Main protease, involved in viral polyprotein processing during replication
Nsp6	Putative transmembrane domain; initial induction of autophagosomes from host endoplasmic reticulum
Nsp7	Cofactor of RNA-dependent RNA polymerase
Nsp8	Cofactor of RNA-dependent RNA polymerase
Nsp9	A single-stranded RNA-binding viral protein
Nsp10	Growth-factor-like protein contains two zinc-binding motifs, plays an essential role in viral mRNAs cap methylation
Nsp11	Made of 13 amino acids; unknown function

Nsp12	RNA-dependent RNA polymerase, responsible for replication and transcription of the viral RNA genome
Nsp13	A helicase core domain that binds ATP
Nsp14	Proofreading Exoribonuclease domain with exoribonuclease activity acting in a 3′ to 5′ direction and N7-guanine methyltransferase activity
Nsp15	Mn^{2+}-dependent endoribonuclease
Nsp16	2′-O-ribose methyltransferase that mediates mRNA cap 2′-O-ribose methylation to the 5′-cap structure of viral mRNAs

1.3.2 Potential Therapeutic Targets for SARS-CoV-2

Main Protease (Mpro) and Papain-like Protease (PLpro)

Mpro (also referred to as 3CLpro, Nsp5) and PLpro (Nsp3) of SARS-CoV-2 are two cysteine proteases that cleaves the polypeptide translated from viral RNA into the non-structural and structure proteins which play a vital role in the replication of packaging of new viruses.

After the translation of viral mRNA, the translated polypeptide will undergo an autocatalytic process that produces mature Mpro. Then Mpro will catalyze the cleavage of the polypeptide at 11 different sites to release Nsp4 to Nsp16[65]. Therefore, Mpro is considered as one of the most important NSPs of the virus. Mpro of SARS-CoV-2 has a remarkably high sequence identity of 96% to SARS-CoV Mpro, implying that previously developed inhibitors for SARS-CoV Mpro are very likely to have good potency on SARS-CoV-2 Mpro as well. This provides a good starting point for new inhibitor design.

PLpro cleaves the translated viral polypeptide at 3 different sites that releases Nsp1, Nsp2 and Nsp3. Besides, it also has deubiquitinase activity against host cell proteins, such as the cleavage of ubiquitin like interferon-stimulated gene 15 protein (ISG15) from interferon responsive factor 3 (IRF3), which leads to attenuation of host immune responses[66]. The sequence identity between SARS-CoV PLpro and SARS-CoV-2 PLpro is 83%, which is not as high as that of Mpro, The crystal structures of both proteases have been determined which greatly facilitated the inhibitor design process[67, 68].

RNA-dependent RNA Polymerase (RdRP)

A crucial process in SARS-CoV-2 life cycle is the replication of its genomic RNA. This process is accomplished by a multi-component machinery consisting of several NSPs including Nsp7, Nsp8 and Nsp12. Among them, Nsp12 is a key component, functioning as the RNA-dependent RNA polymerase (RdRP) which catalyzes the synthesis of viral RNA. Therefore, Nsp12 is considered as an important therapeutic target, while Nsp7 and Nsp8 are considered cofactors that help Nsp12 bind more tightly to RNA[69, 70]. But disruption the formation of Nsp7/Nsp8/Nsp12 could also be another target.

Similar to Mpro, Nsp12 (RdRP) of SARS-CoV-2 also shares a remarkable 96% sequence identity to that of SARS-CoV. The complex structure of RdRP with Nsp7 and Nsp8 has also been determined with cryo-electron microscopy at a resolution of 2.9 Å[71].

Spike Protein

Spike protein is a structural glycoprotein located on the surface of the virus. It is the major protein that interacts with the ACE2 receptor on the surface of host cells and mediates the invasion into cells. Spike protein recognizes ACE2 receptor through its RBD

domain. The structure of the complex between SARS-CoV-2 Spike and ACE2 receptor has also been determined[72, 73]. Compared with SARS-CoV Spike protein, SARS-CoV-2 Spike protein has a comparable binding affinity to ACE2 receptor and its RBD domain solely even has a higher binding affinity[63]. Given the nature of the interaction between Spike and ACE receptor as a protein-protein interaction, rational inhibitor design could be challenging. But Spike protein could be a promising target for the development of vaccines and antibody drugs.

Another possible therapeutic target related to Spike protein comes from its pre-activation by host proteases. Upon binding to ACE2 receptor, Spike protein will be activated through proteolytic cleavage by human proteases such as TMPRSS2 and furin[62, 63]. Lysosome cathepsins also facilitate the entry of SARS-CoV-2[64]. Inhibition of these host proteases have already shown inhibitory effects of viral entry[62-64]. Therefore, selective inhibition of some host proteases that play a vital role in facilitating viral entry could be a potential strategy against SARS-CoV-2.

1.3.3 Drug Development against SARS-CoV-2 Mpro

Viral protease is one of the most common therapeutic targets for the development of antiviral agents. Since the last outbreak of SARS-CoV in 2003, SARS-CoV-2 main protease has drawn much attention for drug discovery research as a potential antiviral therapeutic target. A number of potent main protease inhibitors have been developed[74]. Since the high sequence identity between SARS-CoV-2 main protease and SARS-CoV main protease, many SARS-CoV main protease inhibitors have the potential to potently

inhibit SARS-CoV-2 main protease as well and serve as the starting point for the discovery

and optimization of SARS-CoV-2 main protease inhibitors.

Covalent Inhibitors

SARS-CoV and SARS-CoV-2 main proteases are both cysteine proteases. This

feature makes the development of covalent inhibitors that chemically react and inactivate

the catalytical cysteine an attractive option. Many covalent inhibitors based on various

warheads, including aldehydes and their bisulfate adducts[75, 76], nitriles[77], ketoamides[78],

active esters[79], vinyl sulfones[80] and acrylates[81], have been developed. Some of these

inhibitors are shown in **Figure 11**. Many of these inhibitors have nanomolar potency

against SARS-CoV-2 main protease *in vitro*. Based on the mechanism of action, these

covalent inhibitors can be further categorized into reversible covalent inhibitors, including

aldehydes, nitriles, ketoamides and active esters, and irreversible covalent inhibitors

including vinyl sulfones.
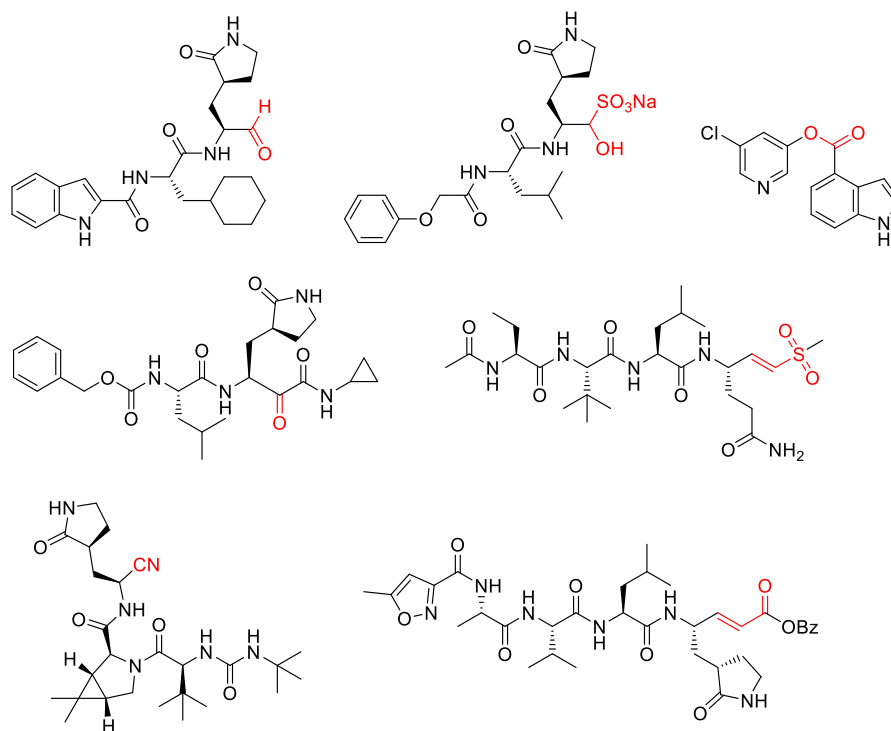
Figure 11. Structures of selected covalent inhibitors for SARS-CoV-2 main protease. The warhead functional groups that react with active cysteine are highlighted.

Non-covalent Inhibitors

Compared to covalent inhibitors, non-covalent inhibitors are being paid less attention to by medicinal chemists, partially due to their less potency (they typically have an IC_{50} between 1 to 10 µM) compared with their covalent counterparts with nanomolar inhibition. Nevertheless, there are still quite a few non-covalent inhibitors identified. Many of these compounds are identified through high-throughput screening[81, 82] or isolated and derived from traditional Chinese medicine that showed some efficacy in China's early-stage clinical treatment to COVID-19 patients[83, 84].

27
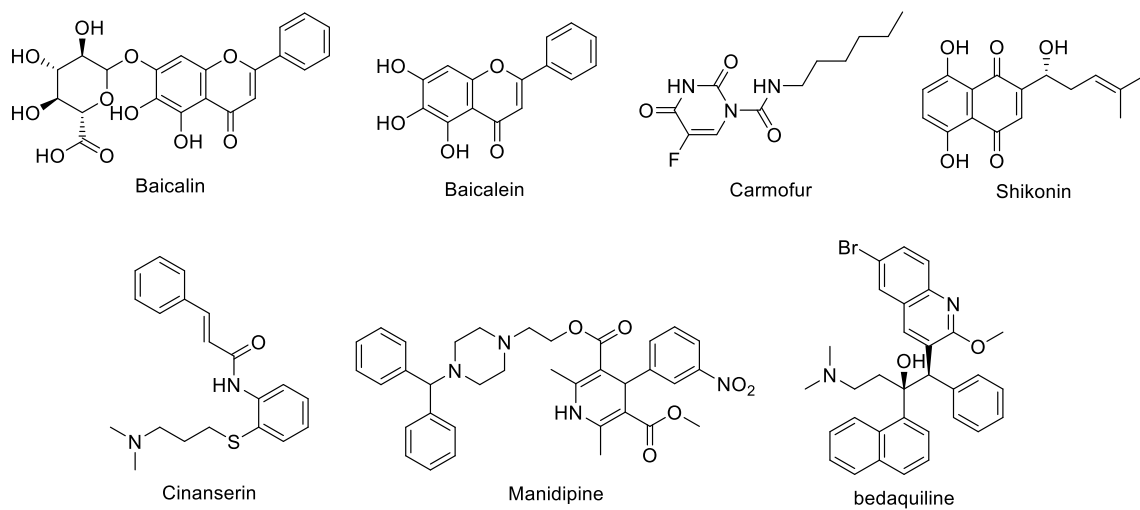
Figure 12. Structures of selected examples of non-covalent inhibitors for SARS-CoV-2 main protease.

CHAPTER II

STRUCTURE-BASED DRUG DEVELOPMENT FOR HUMAN ENL YEATS

DOMAIN*

2.1 Introduction

Post-translational modifications (PTMs) of histones play an important role in the epigenetic regulation of gene expression. These modifications serve as binding sites to recruit reader proteins, which in turn transduce the epigenetic signals into downstream functional outcomes[85, 86]. In addition to small compounds that modulate enzymatic activities of the histone-modifying enzymes, perturbations of reader-histone interactions also provide attractive therapeutic potentials. One such example is the BET bromodomain inhibitors[44, 55]. Bromodomains are known as readers of histone acetylation[87]. Recent studies from our laboratories and others have identified the YEATS domains as a new family of epigenetic readers that bind to not only histone acetylation but also other types of acylations such as crotonylation[88-98].

The YEATS domain, named after its five founding members (Yaf9, ENL, AF9, Taf14 and Sas5), is evolutionarily conserved from yeast to human[99]. The human genome encodes four YEATS domain-containing proteins: ENL, AF9, YEATS2 and GAS41 that all associate with chromatin-associated protein complexes[100, 101]. ENL and AF9 are paralogues that share a similar protein structure including a highly conserved YEATS

domain. Both ENL and AF9 are subunits of the super elongation complex (SEC) and the complex of the histone H3K79 methyltransferase DOT1L, but mutually exclusive[24, 102]. We and others previously showed that ENL, but not AF9, is required for disease maintenance in acute leukemias, in particular the MLL-rearranged leukemia[96, 103]. Depletion of ENL or disrupting the interaction between its YEATS domain and acetylated histones suppresses leukemia progression. In addition, hotspot ENL YEATS domain mutations were found in Wilms' tumor patients[104, 105]. We showed that the reader function of the ENL YEATS domain is indispensable for the gain-of-function mutations in the oncogenesis of Wilms' tumor[106]. Together, all these studies suggest that the YEATS domain of ENL is an attractive therapeutic target.

The acetyllysine binding pocket of the ENL YEATS domain is a long and narrow hydrophobic channel, making it a potentially good target for developing small-molecule inhibitors[96]. Indeed, recent publications of acetyllysine competitive small compounds and peptide-mimic chemical probes demonstrate that the ENL YEATS domain is pharmacologically tractable[107-113]. The peptide-mimic chemical probes showed slightly higher potency to the ENL YEATS domain than other YEATS domains, largely due to interactions outside of the acetyllysine binding pocket[111]. In contrast, the small molecule ENL inhibitors reported so far failed to distinguish ENL from its close paralogue AF9. In addition, none of these small molecule compounds showed significant impact on ENL-dependent leukemia cell growth, suggesting that development of potent, selective ENL YEATS domain inhibitors is in great need. Here we report the discovery of small-molecule compounds that exhibit preferential binding to ENL compared to AF9 and other YEATS

domain proteins. Two compounds, **11** and **24**, displayed IC$_{50}$ values below 100 nM in inhibiting the ENL-acetyl-H3 interaction in vitro. In leukemia cells, compound **7** reduced ENL target gene expression and suppressed leukemia cell growth. In addition, **7** exhibited a synergistic effect with the BET bromodomain inhibitor JQ1 in killing leukemia cells. Our study provided valuable selective ENL chemical probes and potential leads for further medicinal chemistry-based optimization to advance both basic and translational research of ENL.

2.2 Results and Discussion

2.2.1 High-throughput Library Screen for ENL YEATS Domain Inhibitors

In order to identify small molecule inhibitors for the YEATS domain of ENL, we first established an AlphaScreen assay system for high-throughput screening (HTS) of small molecule compounds. In this assay system, two analytes, a 6×His-tagged ENL YEATS domain (His-ENL) and a biotin-H3K9ac peptide (histone H3 residues 1-21 with an acetylation at Lys 9) were immobilized on Perkin Elmer Ni^{2+}-chelating acceptor and streptavidin donor beads, respectively (**Figure 13**). Protein and peptide dose-response assays determined optimal concentrations of His-ENL and biotin-H3K9ac to be 100 nM and 30 nM, respectively (**Figure 13B** and **Figure 13C**). We also determined the optimized Alpha-beads concentration to be 10 µg/mL. This assay system was further evaluated in a high-throughput setting in 384-well plates. Inter-plate variations were measured between two separate plates and on two separate days, yielding robust and highly reproducible

results with high signal/background (S/B) ratio (39.02), low coefficient of variation

(3.5%), and an excellent Z' factor (0.92) (**Figure 13D**). DMSO tolerance of the assay (0.1-

1%) indicated that the Alpha signals were maintained at 95% and 85% in the presence of

0.1% and 0.5% DMSO, respectively. We also set up a counter assay using a biotin-14xHis

peptide to eliminate compounds that interfere with AlphaScreen assay components.

Together, these data demonstrate that the AlphaScreen assay we developed is suitable for

high-throughput screening of ENL inhibitors, with superior sensitivity and reproducibility.
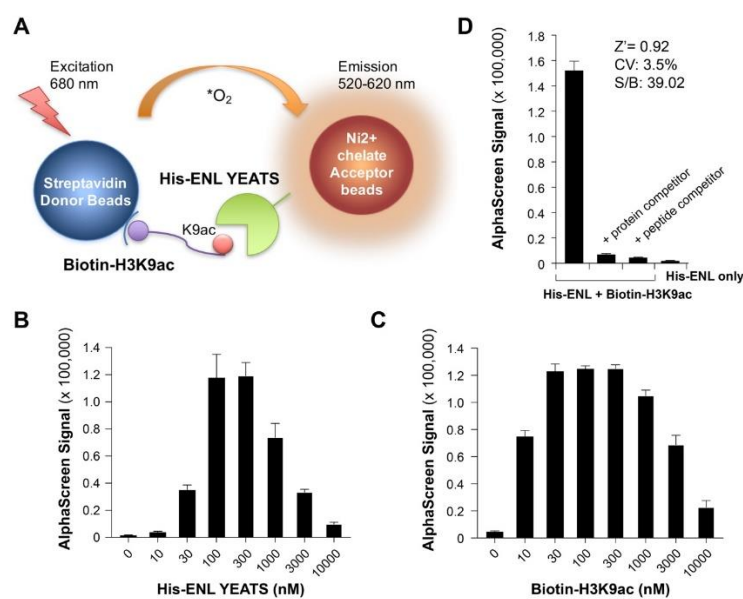
Figure 13. Development of an AlphaScreen assay detecting the interaction between His-ENL YEATS domain and biotinylated H3K9ac peptide. (**A**) A schematic representation of the developed AlphaScreen assay. (**B**) Alpha signals when different concentrations of His-ENL YEATS were titrated into 30 nM of H3K9ac peptide. (**C**) Alpha signals when different concentrations of the H3K9ac peptide were titrated into 100 nM of His-ENL YEATS. (**D**) The developed AlphaScreen assay produces robust and highly reproducible signals in the detection of the interaction between His-ENL YEATS and H3K9ac peptide.

After adapting the AlphaScreen-based HTS system to an automated format for

ENL (100 nM His-ENL, 10 nM biotin-H3K9ac, 0.1% DMSO, and 2.5 µg/mL Alpha

beads), we proceeded to screen a small molecule library of 66,625 compounds with

diverse chemical scaffolds. Non-fragment compounds were screened at a concentration of 10 µM and fragment-based compounds were screened at a concentration of 50 µM. In the primary screen, we obtained 4648 hits with above 50% inhibition. Confirmation and counter assays yielded 524 compounds with above 60% inhibition of the His-ENL–H3K9ac interaction and below 20% inhibition of the counter screen. We then subjected the top 100 compounds to full dose-response curve validation and obtained 37 compounds with IC_{50} values below 5 µM, including 8 compounds with IC_{50} below 1 µM.

2.2.2 Structure-based Inhibitor Design and Structure–activity Relationship Studies

Among the top 8 hits that have an IC_{50} value below 1 µM, five, named as **1-5**, are structurally similar and share a same pharmacophore [1,2,4]triazolo[4,3-a]pyridine-6-amide, suggesting a preferential binding of this pharmacophore to the ENL YEATS domain (**Figure 14A**). All these 5 compounds also contain an aryl substituent at the amide nitrogen side, allowing them to be generally defined as *N,C*-diarylamides. To understand how these compounds interact with ENL, we performed docking analysis using an existing crystal structure of the ENL YEATS domain (the PDB entry: 5j9s). The results showed that all 5 compounds fit nicely to the acetyllysine binding channel of ENL (**Figure 14B**). The compounds are bound to the ENL YEATS domain with a similar orientation as an acetyllysine in a native histone ligand. Similar to acetyllysine side chain amide, the amide in **1-5** is poised to form two hydrogen bonds with S58 and Y78. Although the two aromatic rings can flip to bind either side of the channel, both potentially form pi stacking and van der Waals interactions with residues F28, H56, F59, Y78, and F81 in ENL for preferential

binding (**Figure 14C**). The modelling analysis also indicated that **1**-**5** occupy almost fully

the acetyllysine binding channel of ENL.
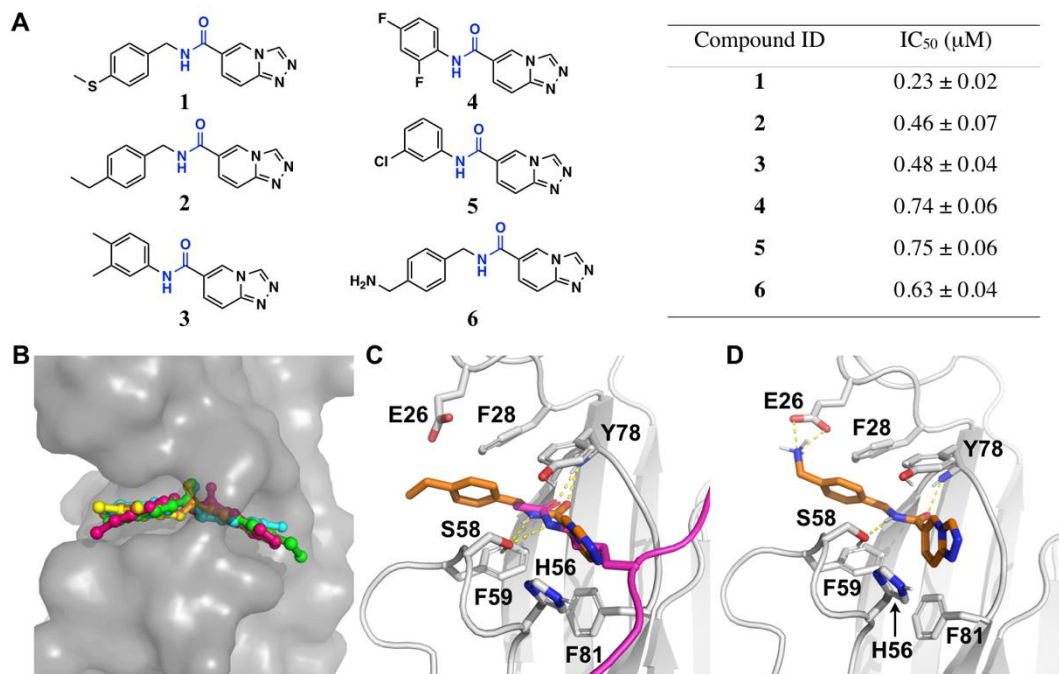
Figure 14. **Structural modeling of the five initial HTS hits, 1-5, and an amine analog 6 with the ENL YEATS domain.** (**A**) Chemical structures of compounds 1-6 (left) and their IC_{50} values in inhibiting the His-ENL–H3K9ac interaction in AlphaScreen assay (right). (**B**) Structural model showing binding of 1-5 to the ENL YEATS domain. Modeling was based on the crystal structure of the ENL YEATS domain (PDB entry: 5j9s). 1-5 are shown in stick representation and the ENL YEATS domain is shown in contoured surface structure. Atoms in ENL are colored in gray, compound 1 in green, 2 in pink, 3 in yellow, 4 in cyan, and 5 in orange. (**C**) The modeled interaction of 2 with ENL. The Ca atoms of 2 are colored in orange and two hydrogen bonds (dashed lines) to S58 and Y78 of ENL are colored in yellow. Acetyllysine in the H3K27ac ligand in the original crystal structure is colored in hotpink and its two hydrogen bonds with E58 and Y78 are shown for comparison. (**D**) The modeled interaction of 6 with ENL. The Ca atoms of 6 are colored in orange. The amine in 6 shows a salt-bridge interaction with E26 in ENL.

Since ENL has relatively flat interfaces on the two sides of the acetyllysine binding

channel, there is a little space for chemical maneuvers of **1**-**5** for improved binding.

However, we noticed that E26, a residue at the edge of the acetyllysine binding channel

can potentially flip its side chain toward the acetyllysine binding channel to interact with

a ligand such as **2** (**Figure 14C**). We deemed that by adding a positively charged amine or amidine to **2** it is possible that a salt-bridge interaction with E26 can be introduced for strong binding to ENL. Therefore, we synthesized compound **6** (**Scheme 1A**) and tested its inhibition of the interaction between His-ENL and biotin-H3K9ac. The determined IC_{50} value for **6** was 0.63 μM, which is very similar to that for **2** (**Figure 14A** and **Figure 15**). Since the introduction of an amine makes the compound more favorable to dissolve in water, some strong interaction with ENL is necessary to compensate the energy loss due to desolvation when **6** binds ENL. This is supported by the salt-bridge interaction observed when we modelled **6** to the acetyllysine binding channel of ENL. In the modelled structure, **6** interacts with ENL similar to **2** except that it engages E26 for a salt-bridge interaction (**Figure 14D**). We have also attempted to co-crystalize ENL with **6** for crystal structure determination, but unfortunately it has not been successful.
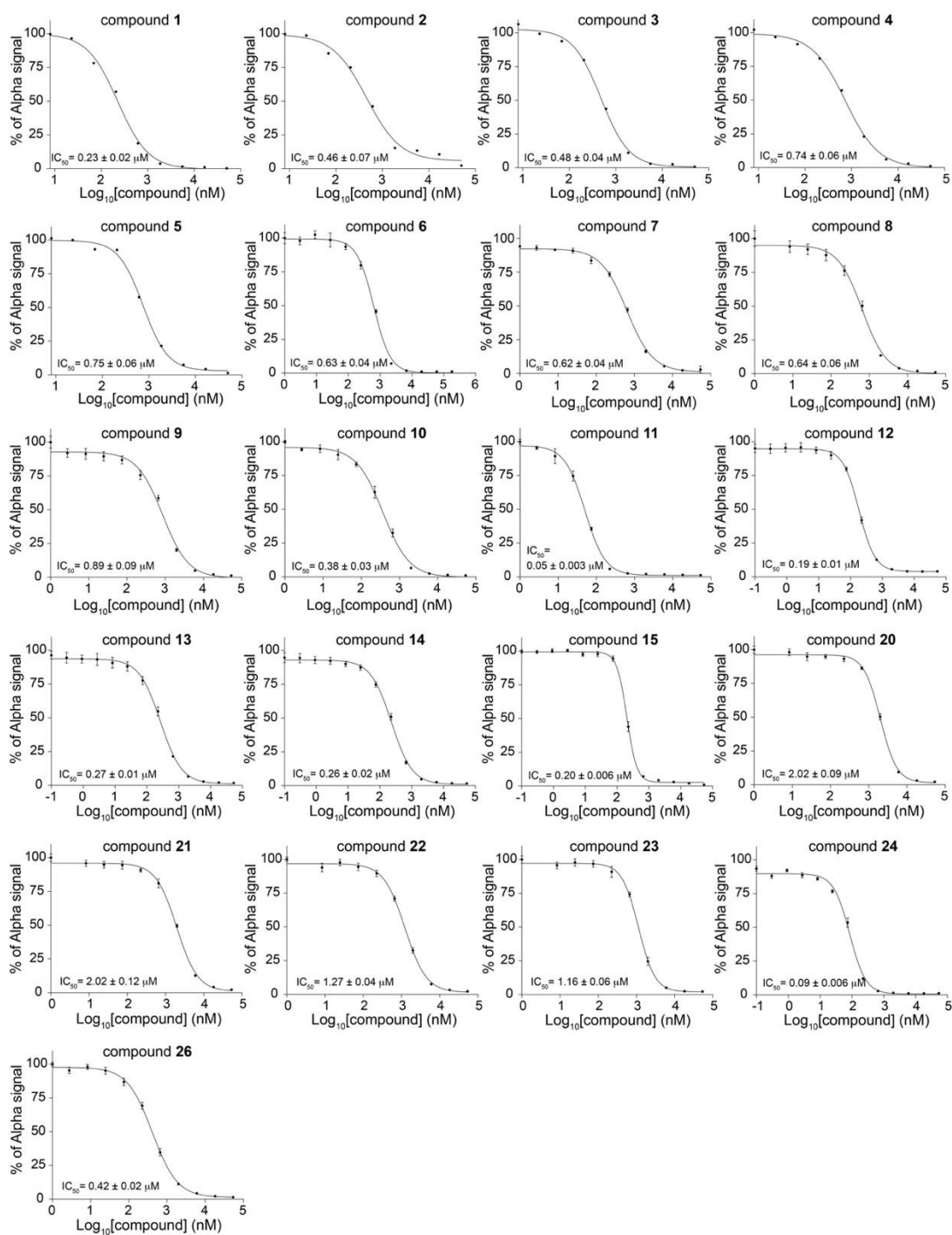
Figure 15. IC$_{50}$ determination of compounds 1-28 by AlphaScreen assay. Compounds were subjected to a series of 3-fold dilutions from 54 mM for dose response curve AlphaScreen assays. IC$_{50}$ values were determined from the plot using nonlinear regression of variable slope (four parameters) and curve fitting performed by the GraphPad Prism software.

Scheme 1. Synthesis of compounds 6-19.
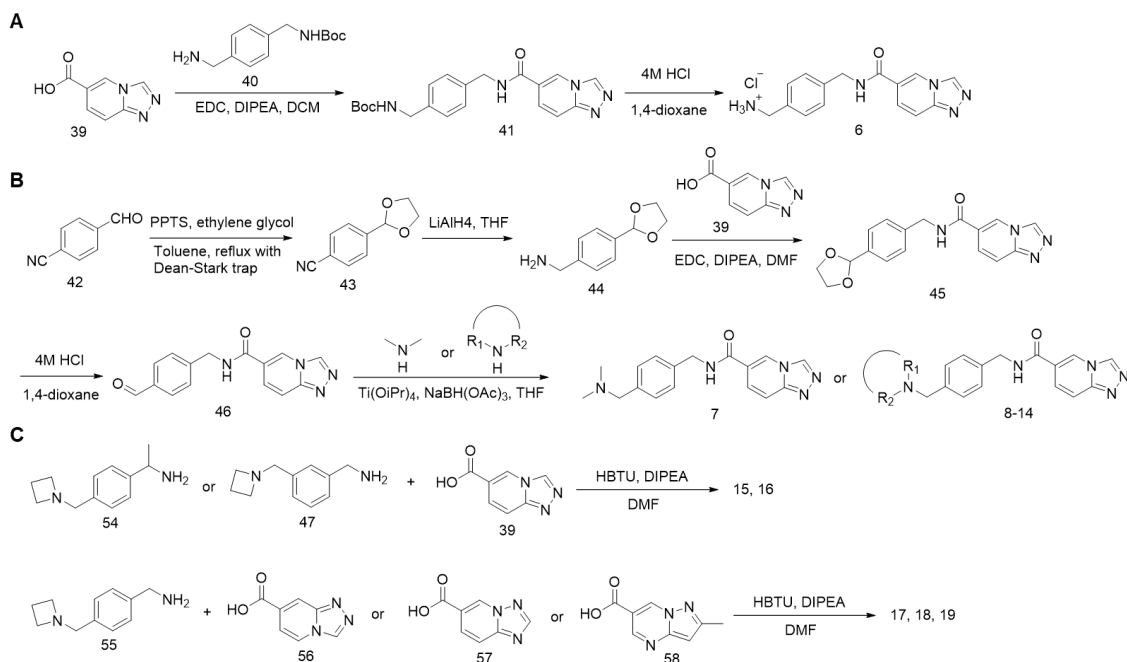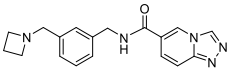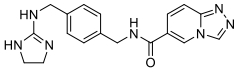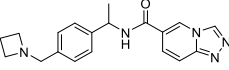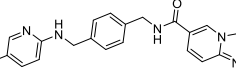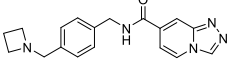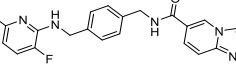
Encouraged by the results from **6**, we expanded the scope of substitution groups on both sides of the amide bond of hit compounds for a comprehensive structure–activity relationship (SAR) study. A major focus was to maintain a positive charged amine, amidine or guanidine as in **6** but tune the ENL binding as well as lower the energy loss due to desolvation by adding different alkyl substituents to the amine, amidine or guanidine. The inhibition potency of all compounds was first tested at 1 µM and 0.1 µM. Promising compounds were then subjected to a more accurate AlphaScreen assay for IC_{50} determination. We first started with replacing the primary amine of **6** with different kinds of tertiary amines through reductive amination of key aldehyde intermediate 46 (**Scheme 1B**), which resulted in **7-11** (**Table 2**). IC_{50} measurement showed that compounds tended to be more potent as the ring size of the substitutional groups on the tertiary amine decreased (**Figure 15**). Among them, **11** that has a four-membered ring azetidine moiety

exhibited the most potency with an IC_{50} value as 51 nM in inhibiting the interaction between His-ENL and biotin-H3K9ac. These results suggest that the azetidine ring assists the binding to ENL. Further modifications were then introduced to **11** to afford **12**-**14** with different alkyl groups on the 2' position of azetidine, which hopefully could increase the electron density on the N atom and enhance the interaction between azetidine and Glu26. However, these compounds displayed lower potency than **11**. Given that Glu26 is located at a loop area with much conformational flexibility, we moved the azetidine moiety from the para to meta position affording **15** but did not result in an increase of potency. We also tried to slightly increase the rigidity of the molecule by adding a methyl group to the benzylic carbon affording **16**. However, it significantly reduced the inhibition potency. We also substituted the triazolopyridine moiety with similar heterocycles to afford **17**-**19**, but none of these compounds outcompeted **11** (**Table 2** and **Figure 15** and **Scheme 1C**).

Table 2. Chemical structures and IC_{50} values of 7-28

ID	Structure	IC_{50} (nM)	ID	Structure	IC_{50} (nM)
7		621 ± 40	18		> 5000
8		636 ± 64	19		> 5000
9		891 ± 91	20		2015 ± 92
10		379 ± 27	21		2016 ± 120
11		51 ± 2.9	22		1270 ± 39
12		185 ± 10	23		1156 ± 57
13		266 ± 13	24		85 ± 5.5
14		264 ± 21	25		> 1000

38

15		202 ± 5.8	**26**		417 ± 22
16		> 1000	**27**		> 5000
17		> 5000	**28**		> 2000

In addition to amine derivatives, we also designed a series of amidine derivatives based on compound **6** to afford **20**-**23**. These compounds were synthesized from corresponding nitrile intermediates followed by acid-catalyzed ethanolysis and then ammonolysis (**Scheme 2A**). However, none of these compounds showed improved potency. Although an amidine or guanidine tend to form a stronger salt bridge with a carboxylate than an amine, it may have a higher desolvation energy than an amine, contributing to weaken binding to ENL. For this reason, we focused the synthesis of additional amidine and guanidine derivates **24**-**28** that have higher hydrophobicity than **20**-**23**. These compounds were synthesized by directly reacting **6** with corresponding N-heterocycle building blocks, except for compound **27**, which was made through 5-fluoro-2-aminopyridine due to the inadequate reactivity of 5-fluoro-2-chloropyridine in the reaction with 6 (**Scheme 2B**). Among them, **24** exhibited an IC_{50} value as 85 nM. **11** and **24** are the two most potent compounds in our compound series. We further evaluated their binding to ENL using the surface plasmon resonance (SPR) analysis. His-ENL was immobilized on dextran-coated Au chips through EDC/NHS coupling, followed by flow-through of a buffer containing different concentrations of **11** and **24**. The responses in sensorgrams were fitted to the Langmuir 1:1 binding kinetics model to obtain both association and dissociation rate constants, from which K_d values were then determined

39

(**Figure 18A**). Compared to the kinetics of typical small molecule-protein interactions, both association and dissociation of **11** and **24** toward ENL are relatively slow (association: 1800 and 1600 $M^{-1} \cdot s^{-1}$; dissociation: 8.3×10^{-5} and 7.0×10^{-5} s^{-1} respectively). Their determined K_d values by SPR were 45 and 46 nM, respectively. As far as we know, **11** and **24** are the two most potent inhibitors for ENL that have so far been developed.

Scheme 2. Synthesis of compounds 20-28
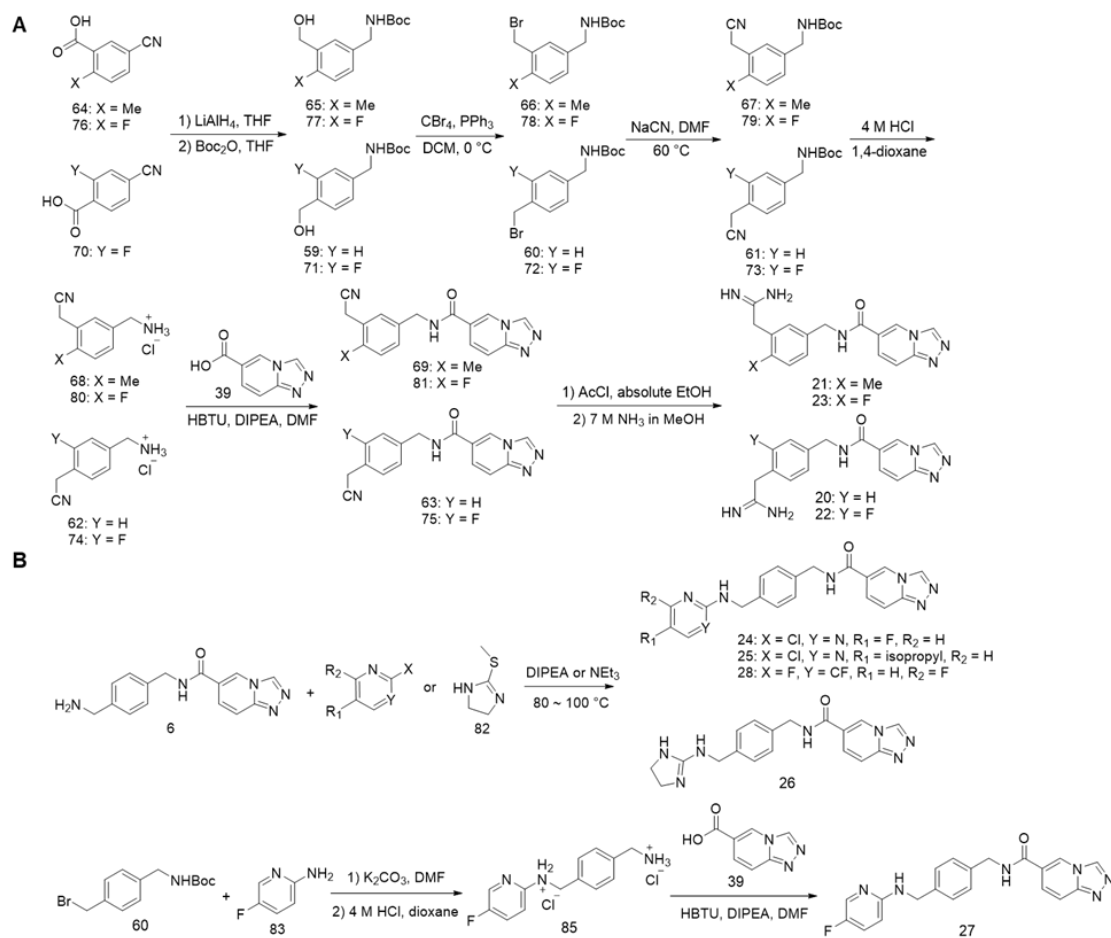
2.2.3 Structural Insights of The Small Molecule Inhibitors Binding to the ENL YEATS Domain

To study the molecular basis of **7**, **11**, and **24** binding to ENL, we attempted to co-crystalize the ENL YEATS domain with these compounds, but it was not successful. We then modelled these compounds to the acetyllysine binding pocket of the ENL YEATS domain by docking analysis. **7** and **11** were docked in their protonated form while **24** was docked in the neutral form given it is less likely to be much protonated under physiological pH. In the modelled structures (**Figure 16**), all three compounds interact with ENL similar as **6** (**Figure 14D**). The amide forms two hydrogen bonds with S58 and Y78. The triazolopyridine ring was involved in pi stacking interactions with H56 in a parallel configuration and with Y78 in a T-shaped configuration. The phenyl group was also involved in pi stacking interactions with F28 and F59, both in a T-shaped configuration. Importantly, the amine of **7**, the azetidine of **11**, and the guanidine of **24** are all within 3 Å to E26, suggesting a common salt bridge or hydrogen bond interaction that stabilizes their interactions with ENL (**Figure 16**).
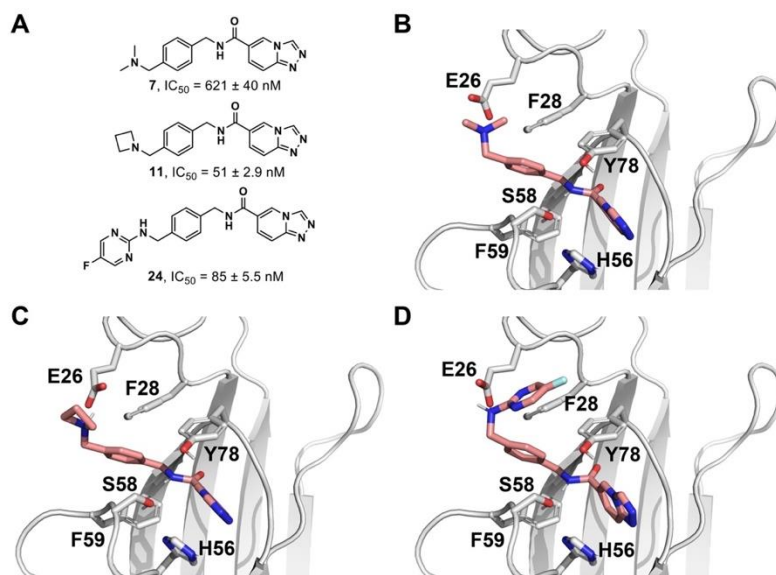
Figure 16. Compounds 7, 11 and 24 and their docking models bound to the ENL YEATS domain. **(A)** Chemical structures of compounds 7, 11 and 24 and their IC_{50} values in inhibiting the His-ENL−H3K9ac interaction in AlphaScreen assay. **(B-D)** The molecular docking models of compounds **7 (B)**, **11 (C)**, and **24 (D)** bound to the ENL YEATS domain. Modeling was based on the crystal structure of the ENL YEATS domain (PDB: 5j9s). Compounds are shown in stick and the ENL YEATS domain is shown as cartoon in gray. Compound-interacting residues of ENL are highlighted and shown in stick.

To experimentally validate that **7** binds to the acyl-lysine binding pocket of ENL, we compared the binding of **7** and a H3K27cr peptide (histone H3 residues **22**-**31** with a crotonylation at K27) to ENL by nuclear magnetic resonance (NMR) spectroscopy (Klein et al., 2014). We expressed ^{15}N-labelled His-ENL and recorded its ^{1}H, ^{15}N heteronuclear single quantum coherence (HSQC) spectra while **7** or the H3K27cr peptide was titrated into the sample (**Figure 17** and **Figure 18B**). As expected, the H3K27cr peptide induced large chemical shift perturbations (CSPs) in the ENL YEATS domain, which were in the intermediate to fast exchange regime on the NMR timescale. Addition of **7** caused CSPs in the intermediate to slow exchange regime, indicating that the ENL YEATS domain binds to **7** tighter than to the H3K27cr peptide. An overall similar pattern of CSPs observed

in both experiments suggest that **7** and the H3K27cr peptide occupy the same binding site

in the ENL YEATS domain.
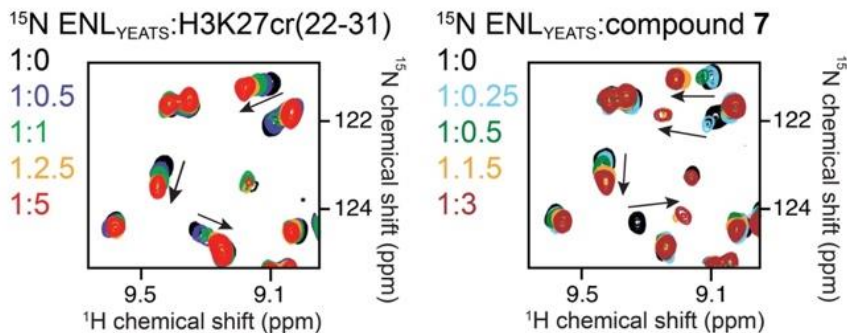
Figure 17. Compound 7 and the H3K27cr peptide occupy the same binding site of the ENL YEATS domain. Superimposed $^1H,^{15}N$ HSQC spectra of His-ENL collected as H3K27cr (H3 residues 22-31, left) or **7** (right) was added stepwise. Spectra are color-coded according to the protein:ligand molar ratios.
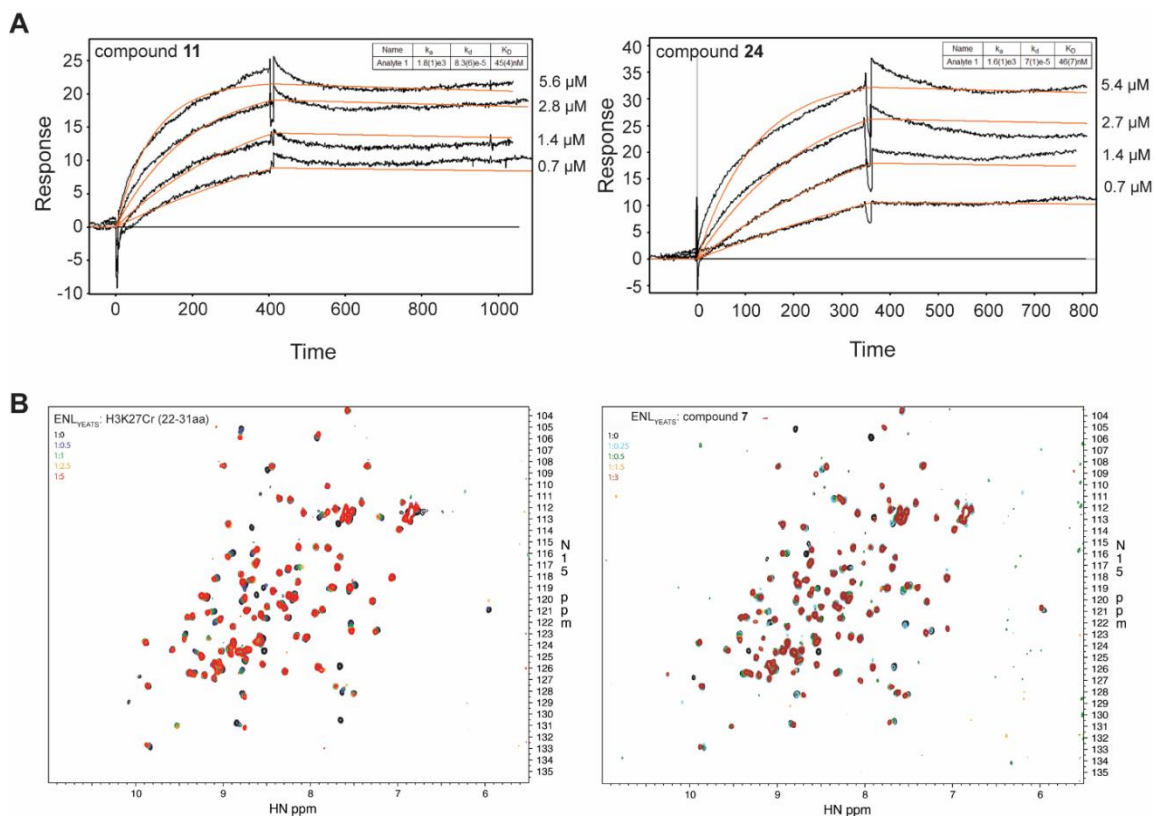
Figure 18. SPR and NMR analysis of compound 11, 24 or 7. (A) Sensorgrams of SPR experiments and the fitted Langmuir 1:1 binding kinetic model with compound **11** (left panel) and

24 (right panel). **(B)** Overlay of 1H,15N HSQC spectra of 15N-labeled ENL YEATS domain collected before and after the H3K27cr (aa 22-31 of H3) peptide (left panel) or compound **7** (right panel) was added stepwise. Spectra are color coded according to the protein-peptide molar ratio as indicated.

2.2.4 High Selectivity of Compound 1, 7, 11, 24 toward the ENL YEATS Domain over Other Human YEATS Domains

To determine whether the small molecule inhibitors are selective toward ENL among the four human YEATS domains, we assessed **7**, **11**, and **24** in their inhibition of ENL, AF9, GAS41, and YEATS2 in peptide pulldown assays. We also included the original hit 1 in the assays for comparison. We used the H3K9ac peptide for AF9 and ENL, the H3K27ac peptide for GAS41, and the H3K27cr peptide for YEATS2, as these peptides are the preferred ligands of the corresponding YEATS domains[89, 93, 96, 98]. We found 1 µM of **11** and **24** and 5 µM of **1** and **7** strongly inhibited the binding of ENL to H3K9ac, whereas at even a 20 µM concentration, none of these compounds showed notable inhibition to AF9, GAS41, or YEATS2 binding to their corresponding acylated histone peptides (**Figure 19A**). We further measured the IC_{50} values of **1**, **7**, **11** and **24** in their inhibition of the binding of four human YEATS domains to their preferred histone peptide ligands using AlphaScreen assays. All four compounds displayed preferential inhibition of ENL over the other three YEATS domains. Compound 1 showed ~4-fold higher potency toward ENL than AF9, whereas no detectable inhibition was measured for the YEATS2 or GAS41 YEATS domain. Compounds **7**, **11**, and **24** exhibited even higher specificity to ENL. Particularly, the IC_{50} value of **11** to ENL was ~20-fold lower than that to AF9 (ENL IC_{50} 51 nM and AF9 IC_{50} 984 nM) (**Figure 19B**). As the previously reported

small molecule ENL inhibitors are not able to differentiate between ENL and AF9, our compounds provide promising scaffolds for further development of ENL-specific inhibitors for the study of ENL biology and for disease intervention.
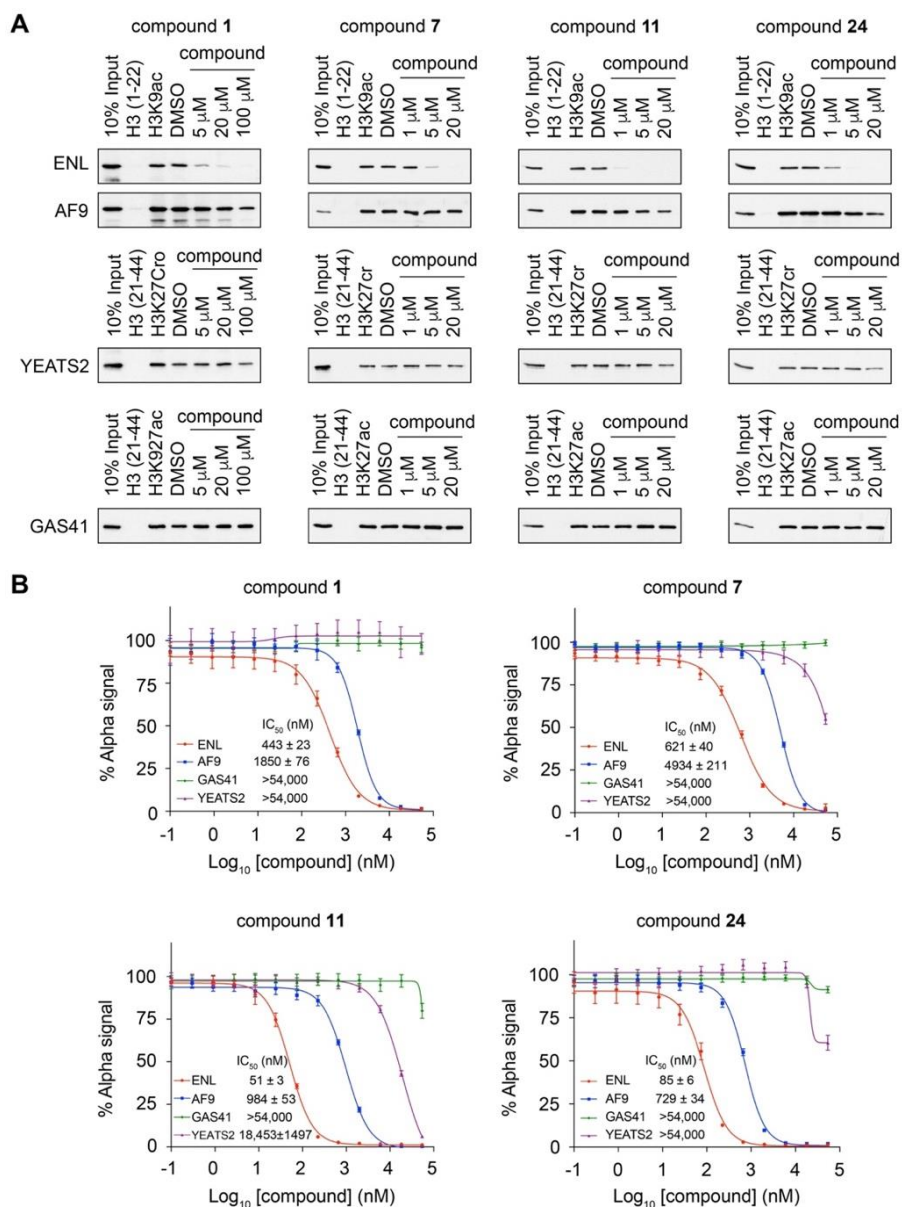
Figure 19. Compounds 7, 11, and 24 are highly specific to ENL over other YEATS domains. **(A)** Peptide pulldowns of ENL, AF9, YEATS2, and GAS41 with the indicated acylated histone peptides with or without **1**, **7**, **11**, and **24**. Unmodified histone peptides were used as negative controls to the acylated peptides and DMSO as a negative control to compound treatment. **(B)**

AlphaScreen measurement of IC_{50} of **1**, **7**, **11**, and **24** in inhibition of YEATS domains binding to the corresponding acylated histone peptides as in **(A)**.

2.2.5 Inhibition of Endogenous ENL Protein by Compound 7 in MLL-rearranged Cell Lines

To explore the small molecule ENL inhibitors we developed in biological applications, we first analyzed their cellular effects in MV4;11 and MOLM13 cells, two *MLL*-rearranged cell lines whose growth is dependent on ENL[96, 103]. We screened 15 compounds with *in vitro* IC_{50} values lower than 2 µM, and we found **7** as the most potent compound in cell growth inhibition (**Figure 20**). The discrepancy between *in vitro* IC_{50} values and cellular efficacy is possibly related to cell permeability or compound stability in cells. Compound **7** exhibited ~ 40% inhibition of MOLM13 cell growth at 5 µM and 80% inhibition at 10 µM concentrations, while about double amounts of the compound were needed to achieve similar levels of inhibition in MV4;11 cells (**Figure 21A**). In contrast, U2OS cells, an ENL-independent cell line, showed little or no response to the treatment with **7**, even at 50 µM.
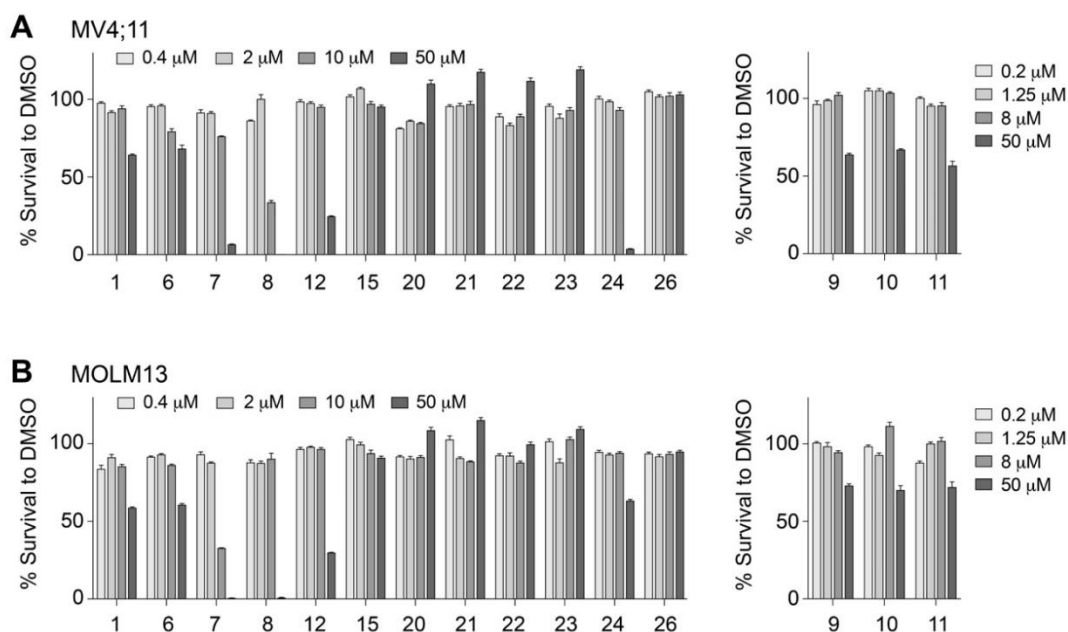
Figure 20. ENL inhibitors inhibit leukemia cell growth. Cell growth inhibition of ENL inhibitors at the indicated concentrations in MV4;11 **(A)** and MOLM13 **(B)** cells. Survived cells were calculated as % relative to DMSO treated cells.

Next, we asked whether the growth inhibition effect was caused by on-target inhibition of the endogenous ENL protein. In this regard, we carried out cellular thermal shift assay (CETSA) to evaluate thermal stability of the ENL protein in MV4;11 and MOLM13 cells treated with **7**. As the AF9 protein is undetectable in these cells with commercial antibodies, we evaluated thermal stability of the GAS41 protein for comparison. Compared to the DMSO-treated cells, we detected higher abundance of soluble ENL proteins in cells treated with **7**, indicating that **7** bound and stabilized ENL proteins (Jafari et al., 2014). In contrast, the thermal stability of GAS41 proteins showed little or no difference between DMSO and **7** treatment (**Figure 21B** and **Figure 21C**). The CETSA results suggest specific engagement of ENL with **7** in living cells.

We also evaluated the expression of two ENL target genes, *HOXA9* and *MYC*, in MOLM13 cells. Compound **7** effectively suppressed *HOXA9* gene expression at as low as 2.5 µM of drug concentration. And at 10 µM, it suppressed ~ 80% of the expression of both *HOXA9* and *MYC* genes (Figure 5D), suggesting potent on-target effect of the ENL inhibitor.
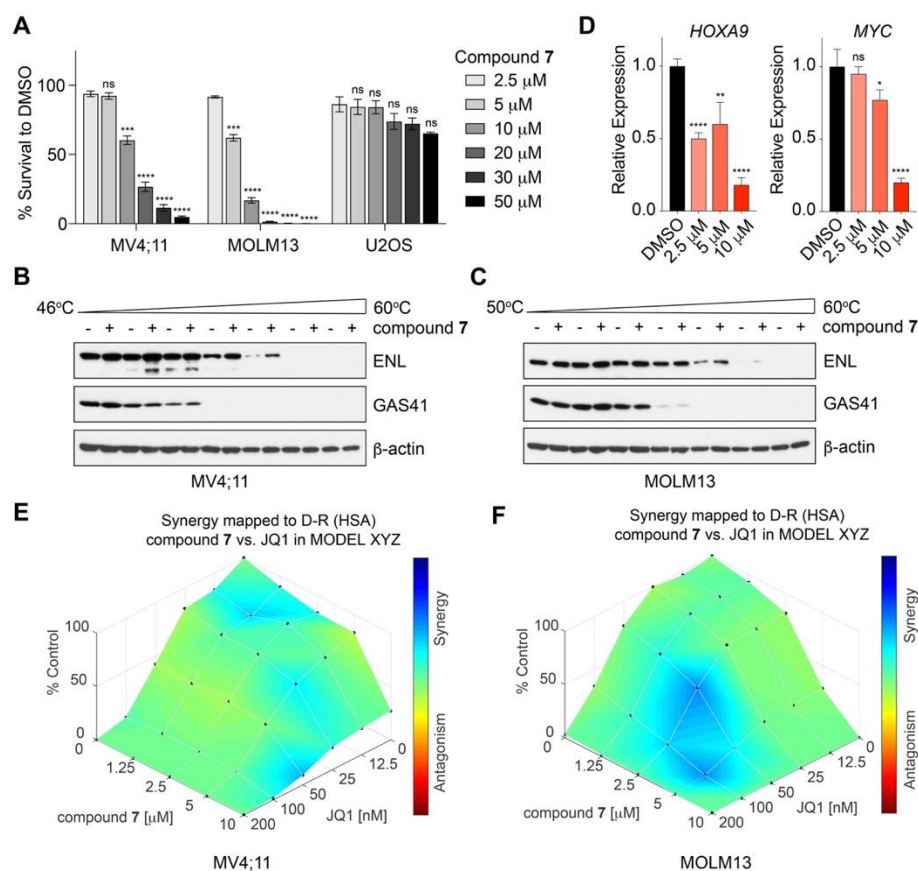
Figure 21. Compound 7 exhibits on-target effect of ENL inhibition in MLL-rearranged leukemia cell lines. (**A**) **7** inhibits leukemia cell growth. Cell growth inhibition of **7** at various concentrations in MV4;11, MOLM13, and U2OS cells. Survived cells were calculated as % relative to DMSO treated cells. (**B-C**) Cellular thermal shift assays in MV4;11 (**B**) and MOLM13 (**C**) cells treated with 20 μM **7** at the indicated temperatures. β-actin was used as a loading control. (**D**) RT-qPCR analysis of HOXA9 and MYC gene expression in MOLM13 cells treated with **7** or the DMSO negative control. (**E-F**) **7** shows a synergistic effect with JQ1. MV4;11 (**E**) and MOLM13 (**F**) cells were treated with indicated doses of **7** and JQ1 or DMSO for 6 days. Survived cells were calculated as % relative to DMSO treated cells. Synergistic interactions were analyzed using the Combenefit software. Data in (**A**) and (**D**) are shown as mean ± SEM of three

independent experiments, two-tailed Student's t test. SEM, standard error of mean. ns not significant, * P < 0.05, ** P < 0.005, *** P < 0.001, **** P < 0.0001.

2.2.6 Compound 7 Exhibits a Synergistic Effect with JQ1

Previously we found that CRISPR/Cas9-mediated ENL knockout sensitized leukemia cells to JQ1, an effective inhibitor of BET bromodomain proteins including BRD4 [55, 96]. An intriguing question was whether **7** has any synergy with JQ1 in killing leukemia cells. To answer this question, we carried out combinatory treatment of MV4;11 and MOLM13 cells with series of concentrations of **7** (0 to 10 µM) and JQ1 (0-200 nM). In both cell lines, we observed synergistic effect between ENL inhibition and JQ1 (**Figure 21E** and **Figure 21F**), demonstrating therapeutic potentials for future exploration in disease treatment [114].

2.2.7 Discussion

Great Pharmacological Value of YEATS Domain

The YEATS domain is a newly identified family of histone acylation readers. The four human YEATS domain-containing proteins, ENL, AF9, YEATS2, and GAS41, are subunits of protein complexes involved in chromatin and transcription regulation[100, 101]. The evolutionarily conserved histone-reading function of the YEATS domains is essential for the functionality of all the YEATS domain proteins in both yeast and human[88-98]. Dysregulation of the YEATS domain-containing proteins has been associated with various human diseases, including cancers. We and others showed that ENL and particularly its YEATS domain is essential for disease maintenance and progression of acute leukemias[96,

[103]. Recently, we also found that the reader function of the ENL YEATS domain is indispensable for the aberrant gene activation and tumorigenesis caused by the gain-of function ENL YEATS domain mutations identified in Wilms' tumor patients[106]. In addition, YEATS2 and GAS41 are frequently amplified in various types of human cancers[115-117]. All these studies suggest that the YEATS domains are promising drug targets and, therefore, targeting the YEATS domains may provide a novel therapeutic approach for a broad spectrum of human cancers.

Selectivity among Different Human YEATS Domains

Developing YEATS domain inhibitors has been a research focus of the epigenetic reader field in recent years. The initial efforts were focused on targeting the YEATS domain of ENL, because of great therapeutic potentials. Both small molecule chemical compounds and peptide-mimic probes have recently been developed as acetyllysine competitive inhibitors of the ENL YEATS domain[107-113]. However, target selectivity has been a big challenge because the YEATS domains share high structural similarity, especially between ENL and its close homologue AF9. The few small molecule ENL inhibitors reported so far have poor specificity that fail to distinguish ENL from AF9. The peptide-mimic chemical probes developed by the Li group showed slightly higher potency to the ENL YEATS domain than other YEATS domains, largely due to interactions outside of the acetyllysine binding pocket[111]. These results suggest that targeting both the acyllysine-binding pocket and additional proximal sites outside of the binding pocket might be a good approach to develop specific inhibitors. Indeed, based on this concept,

the Li group has recently developed a conformationally preorganized cyclopeptide that showed a 38-fold higher binding affinity toward AF9 YEATS over ENL[110].

Despite the success in developing peptide-mimic chemical probes specific to AF9, ENL specific inhibitors were still lacking. Because ENL, but not AF9, is essential for *MLL*-rearranged acute leukemias and ENL mutant Wilms' tumors, it is in urgent need to develop ENL specific inhibitors for further drug development. In our study, through HTS we identified compound **1**, which showed a 4-fold preference towards ENL over AF9 YEATS. After several rounds of structure-based inhibitor design and structure−activity relationship studies, we were able to develop several compounds with much better selectivity. In particular, the IC$_{50}$ value of compound **11** to ENL was ~20-fold lower over AF9, ~360-fold lower over YEATS2, and more than 1,000-fold lower over GAS41, providing a good lead for future drug development.

Possible Explanations for the Observed Selectivity

The selectivity of our compounds to ENL over AF9 is intriguing, given that the AF9 YEATS domain has a 10-fold higher affinity than ENL YEATS in acyllysine binding. The YEATS domains of AF9 and ENL share high degree of structural similarity[93, 96]. It is not clear what interactions contribute to the ENL selectivity. By comparing the modelled structures of compounds **1**, **7**, **11** and **24** docked to the acetyllysine binding pocket of the ENL and AF9 YEATS domains, we observed that the triazolopyridine pharmacophore of these compounds adopts conformations to form stronger pi-pi interaction with H56 residue in ENL than in AF9 YEATS domain. When bound with the ENL YEATS domain, the distances between the triazole rings and the imidazole rings range from 3.4 to 3.8 Å and

their dihedral angles range from 20° to 22°, whereas in the case of AF9 YEATS domain, the distances between them increase to 4.3-4.6 Å and their dihedral angles also increase to a range of 28°-37°, both leading to weaker pi-pi interactions compared to those in the ENL YEATS domain (**Figure 22**). Additionally, the salt bridge interaction in the case of **7** and **11** and hydrogen bond for **24** with E26 in the ENL YEATS domain may also contribute to selectivity. Further structural study of YEATS domains in complex with these compounds will provide insights to guide future development of more potent and selective ENL YEATS domain inhibitors.
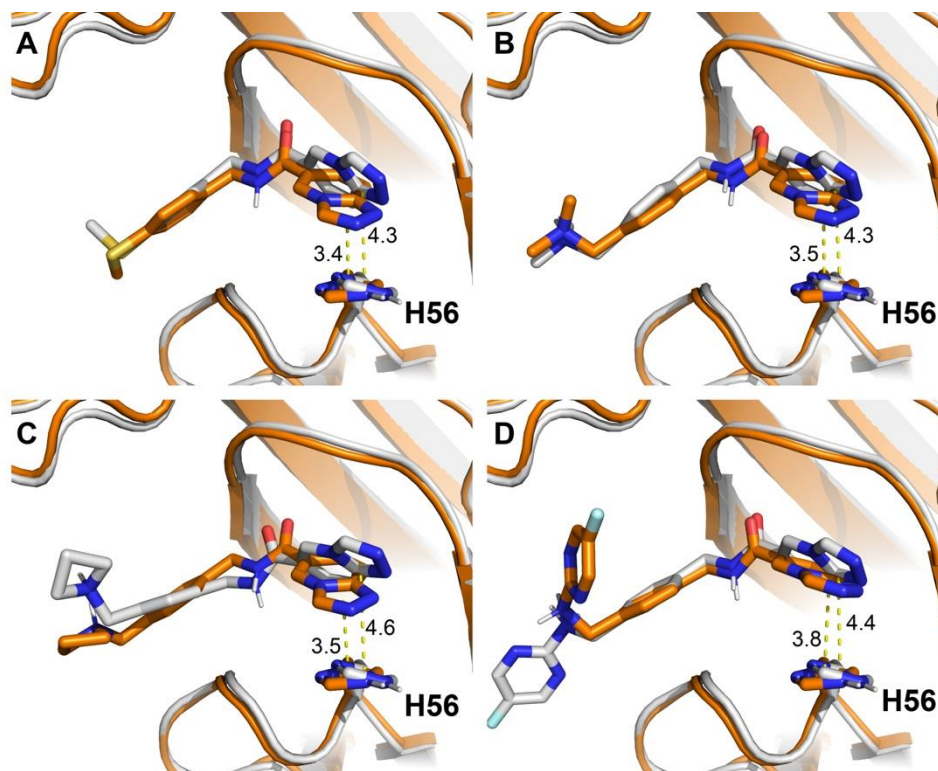
Figure 22. The triazolopyridine pharmacophore of compounds 1, 7, 11 and 24 adopt comformations to form stronger pi-pi interactions with H56 residue in ENL than in AF9 YEATS domain. The molecular docking models comparison of compounds **1 (A)**, **7 (B)**, **11 (C)**, and **24 (D)** bound to the YEATS domain of AF9 (white colored) and ENL (orange colored). Modeling was based on the PDB entries 5j9s (ENL) and 4tmp (AF9).

Potential of Compound 7 as a Medicinal Chemistry Lead or Chemical Probe

In leukemia cells, our synthesized compound **7** exhibited clear on-target cellular effects in reducing ENL target gene expression and suppressing leukemia cell growth. In addition, consistent with previous results of genetic ENL ablation, **7** exhibited a synergistic effect with the BET bromodomain inhibitor JQ1 in killing leukemia cells. The cellular effects of our compounds are superior to all reported ENL inhibitors. Overall, our study provides valuable selective ENL small molecule inhibitors that can serve as potential leads for further medicinal chemistry-based optimization to advance both basic and translational research of ENL. It also provides a molecular platform for the development of more complicated, multifunctional probes for applications such as visualization or targeted degradation in cells.

2.3 Conclusion

In this study, we carried out high-throughput screening of a small molecule library of > 66,000 compounds against the ENL YEATS domain and identified a series of hit molecules that share a [1,2,4]triazolo[4,3-a]pyridine-6-amide pharmacophore and a common *N*,*C*-diarylamide scaffold. By introducing a potential salt bridge interaction with E26 in ENL, we were able to generate compounds with IC_{50} and K_d values less than 100 nM. Importantly, our compounds outcompeted the previously reported ENL inhibitors by showing high selectivity toward ENL over AF9, the close paralogue of ENL. Furthermore, compound **7** exhibited on target effect in inhibiting ENL target gene expression and

leukemia cell growth. Our ENL-specific YEATS domain inhibitors provide the basis for development of potent ENL-specific chemical probes in the future.

2.4 Experimental Details

2.4.1 Protein Expression and Purification

The cDNA encoding sequences of four human YEATS domains: ENL (aa 1-145), AF9 (aa 1-145), full-length GAS41 and YEATS2 (aa 201-332) were cloned in pGEX-6P-1 and pET19b expression vectors, respectively. The His-tagged YEATS proteins were expressed in E. coli Rosetta-2 (DE3) pLysS cells in the presence of 0.2 mM isopropyl-β-D-1-thioglactopyranoside (IPTG) for 18 h at 16 ℃. The His-tagged YEATS proteins were purified using Ni-NTA resins following the manufacture's instruction. The eluted protein was dialyzed in a buffer containing 50 mM HEPES (pH7.4), 100 mM NaCl and 20% glycerol to remove imidazole. Proteins were adjusted to 0.5 mg/mL, aliquoted and stored at -80 ℃. Each batch of purified protein was tested in AlphaScreen assay conditions discussed as following. The GST-tagged proteins used in peptide pulldown assays were expressed in the same way and purified using Glutathion Sepharose resins (GE Healthcare).

2.4.2 AlphaScreem Assau Setup and High-throughput screening

The AlphaScreen assay was carried out in 384-well plates. Manual assay setup was performed in 30 uL reaction in Alpha Reaction Buffer (50 mM HEPES pH7.4, 100 mM

NaCl, 0.1% BSA, and 0.05% CHAPS) with final concentrations of 100 nM His-ENL YEATS, 30 nM Biotin-H3K9ac, and 10 µg/mL of Alpha donor and acceptor beads. During the automation step, we were able to reduce the assay volume to 20 µL per well while maintained the quality and robustness of the assay, with the optimal final concentrations of His-ENL YEATS (100 nM), biotin-H3K9Ac (10 nM), DMSO (0.1%), and Alpha-beads (2.5 µg/mL). Protein, peptide and compounds were mixed and incubated for 1 h at room temperature before adding the Alpha beads. Alpha signals were detected by an EnVision microplate reader equipped with an Alpha laser (PerkinElmer).

High-throughput screen was performed at the Texas Screening Alliance for Cancer Therapeutics (TxSACT) facility. The 66,625 compounds screened were from Maybridge HitFinder Set (14,080), Chembridge Diversity Set (12,900), Chembridge Kinase Set (11,250), Chembridge Fragment Library (4,000), ChemDiv Fragment Collection (14,143), Legacy Collection (2,092), MicroSource Spectrum Collection (2,000), LOPAC Collection (1,275), Selleck Kinase and Bioactive Collection (2,260), NCI Diversity (1595), NCI Mechanistic collection (820), and NCI natural products (210). In the primary HTS, fragments were screened at 50 µM and non-fragment compounds were screened at 10 µM. After the single shot screen and hits triage, 990 hits were picked for confirmation assay, and counter assay with Biotin-14xHis peptide.

2.4.3 IC$_{50}$ determination with AlphaScreen assay

The AlphaScreen assay conditions are essential the same as the one used in high-throughput screen. The protein concentrations of AF9, Gas41 and YEATS2 are 30, 100,

and 100 nM, respectively, and the peptide concentrations are 30 nM. All assays have been validated using protein and peptide competitors. For IC_{50} determination, compounds were subjected to eight 3-fold serial dilutions, for a total of nine concentrations ranging from 50 μM to 8 nM for dose response curve AlphaScreen assays. IC_{50} values were determined from the plot using nonlinear regression of variable slope (four parameters) and curve fitting performed using GraphPad Prism.

2.4.3 Modeling of inhibitors bound with ENL and AF9 YEATS domains

Molecular docking of target compounds was carried out using AutoDock 4 (Morris *et al.*, 2009). The initial conformations of target compounds were first generated and MM2 minimized by PerkinElmer Chem3D software. Structures of the ENL and AF9 YEATS domains were obtained from PDB 5J9S and 4TMP respectively, with H3K27Ac and H3K9Ac deleted from the complexes. Structures of the YEATS domains were then pre-processed in MGLTools 1.5.6 to remove water molecules and add polar hydrogens. The grid box was set to be centred at coordinate (x = 27.352, y = -42.139, z = 3.0) for ENL, and at coordinate (x = 52.734, y = 10.522, z = -11.134) for AF9, with a size of 40 × 40 × 40 npts, which is big enough to contain the binding channel and surrounding amino acid residues. Glu26 residue of ENL YEATS domain was set to be flexible. Target compounds were then docked in the grid box. The conformations with lowest binding energies were converted to PDB files for visualization.

2.4.4 Compound synthesis

All reagents and solvents for synthesis were purchased from commercial sources and used without purification. All glassware was flame-dried prior to use. Thin layer chromatography (TLC) was carried out on aluminium plates coated with 60 F254 silica gel. TLC plates were visualized under UV light (254 nm or 365 nm) or stained with 5% phosphomolybdic acid. Normal phase column chromatography was carried out using a Yamazen Smart Flash AKROS system. Analytical reverse-phase high pressure liquid chromatography (RP-HPLC) was carried out on Shimadzu LC20 HPLC system with an analytical C18 column. Semi-preparative HPLC was carried out on the same system with a semi-preparative C18 column. The mobile phases for were H_2O with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) if not mentioned otherwise. NMR spectra were recorded on a Bruker Avance Neo 400 MHz or Varian INOVA 300 MHz spectrometer in specified deuterated solvents. High resolution electrospray ionization mass spectrometry (HRMS-ESI) was carried out on a Thermo Scientific Q Exactive Focus system. The purities of compounds were confirmed by NMR and analytical HPLC-UV as ≥ 95%.

Scheme 3. Synthesis of compound 6.
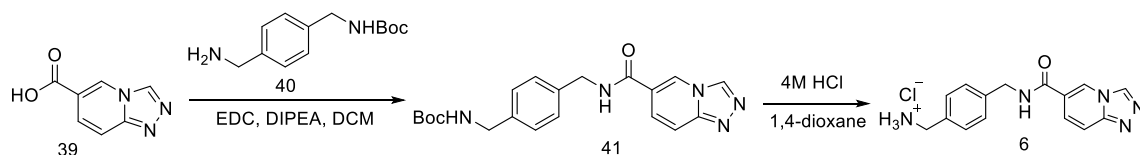
tert-Butyl (4-(([1,2,4]triazolo[4,3-a]pyridine-6-carboxamido)methyl)benzyl) carbamate (41).

To a solution of **39** (1 mmol, 163 mg) and **40** (1 mmol, 236 mg) in dry DMF (5 mL), was added DIPEA (2 mmol, 258 mg), and EDCI (1.2 mmol, 230 mg). The resulting

solution was stirred under room temperature overnight. Then the solution was diluted with EtOAc (50 mL) and washed with saturated NaHCO₃ solution (2 × 50 mL), 1 M HCl (2 × 50 mL) and saturated brine (50 mL). The organic layers were then dried with anhydrous Na₂SO₄ and then concentrated. The residue was purified by column chromatography (silica gel, 10% MeOH/DCM as eluent) to yield **41** as light yellow solid (250 mg, 66%).

(4-((([1,2,4]Triazolo[4,3-a]pyridine-6-carboxamido)methyl)phenyl)methanamine hydrochloride (6).

To a solution of **41** (0.5 mmol, 190 mg) in 5 mL of 1,4-dioxane was added 10 mL of 4 M HCl solution in 1,4-dioxane. The resulting solution was stirred under room temperature for 2 h. Then the reaction mixture was concentrated to dryness *in vacuo* to yield **6** as light yellow solid (150 mg, 95%). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 9.49 (s, 1H), 9.24 (s, 1H), 8.29 (dt, J = 9.7, 1.6 Hz, 1H), 8.10 (d, J = 9.6 Hz, 1H), 7.52 – 7.42 (m, 4H), 4.66 (s, 2H), 4.19 (s, 2H). ^{13}C NMR (101 MHz, D₂O) δ 165.4, 145.4, 138.5, 137.8, 134.1, 131.8, 129.2, 128.0, 127.4, 125.1, 112.1, 43.5, 42.7. ESI-HRMS (m/z): calculated for C₁₅H₁₆N₅O (M+H)$^+$: 282.1349, found: 282.1344.

Scheme 4. Synthesis of compounds 46.
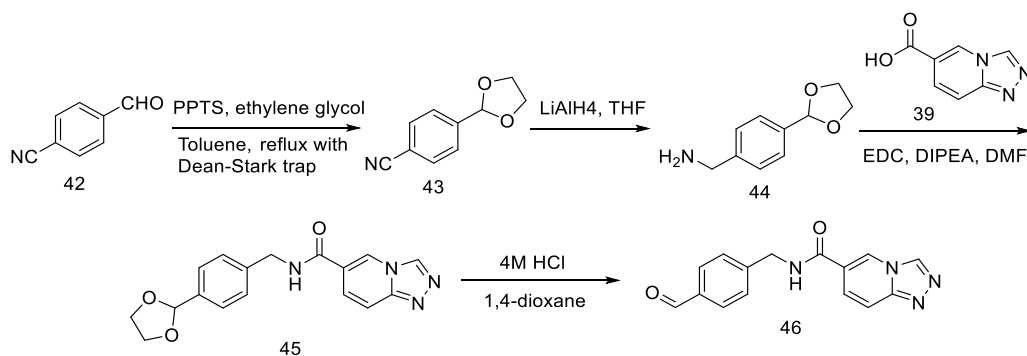
4-(1,3-Dioxolan-2-yl)benzonitrile (43).

To a solution of 42 (38 mmol, 5.0 g) and ethylene glycol (76 mmol, 4.2 mL) in toluene (50 mL) was added pyridinium p-toluenesulfonate (4 mmol, 0.96 g). The resulting solution was heated to reflux with a Dean-Stark trap for 4 h. The resulting solution was then concentrated *in vacuo* and the residue was then purified by column chromatography (silica gel, 10% EtOAc/hexanes as eluent) to yield 43 as white solid (5.25 g, 79%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.67 (d, *J* = 8.3 Hz, 2H), 7.58 (d, *J* = 8.3 Hz, 2H), 5.84 (s, 1H), 4.17 – 3.99 (m, 4H).

(4-(1,3-Dioxolan-2-yl)phenyl)methanamine (44).

LiAlH$_4$ (0.99 g, 26 mmol) was suspended in dry THF (50 mL) and was cooled under 0 °C. A solution of 43 (4.5 g, 25.7 mmol) in dry THF (50 mL) was added dropwise to the LiAlH$_4$ suspension under the same temperature. After the addition, the reaction mixture was warmed up to room temperature and stirred for 4 h. Water (3 mL) was added dropwise followed by 2 M aqueous NaOH solution (3 mL) and then water (3 mL). The precipitate was filtered and washed with THF. The combined filtrate was then dried with anhydrous Na$_2$SO$_4$ and evaporated *in vacuo*. The residue was then purified by column chromatography (silica gel, 100% EtOAc as eluent) to yield 44 as colorless oil to white solid (3.0 g, 67%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.44 (d, *J* = 8.1 Hz, 2H), 7.30 (d, *J* = 8.1 Hz, 2H), 5.78 (s, 1H), 4.15 – 3.95 (m, 4H), 3.85 (s, 2H).

N-(4-(1,3-dioxolan-2-yl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (45).

To a solution of 44 (2 mmol, 358 mg) and 39 (2 mmol, 326 mg) in DMF (10 mL) was added DIPEA (4 mmol, 516 mg) and EDCI (2.4 mmol, 460 mg). The resulting solution was then stirred under room temperature overnight. The reaction mixture then

diluted with EtOAc (50 mL), then washed with saturated NaHCO$_3$ solution (2 × 50 mL), 1 M HCl (2 × 50 mL), and brine (50 mL). The organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to yield 45 as white solid (520 mg, 80%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.36 (d, J = 0.8 Hz, 1H), 9.25 (t, J = 5.9 Hz, 1H), 9.14 (t, J = 1.4 Hz, 1H), 7.90 – 7.76 (m, 2H), 7.44 – 7.33 (m, 4H), 5.70 (s, 1H), 4.52 (d, J = 5.8 Hz, 2H), 4.09 – 3.88 (m, 4H).

N-(4-formylbenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (46).

To a solution of 45 (520 mg, 1.6 mmol) in1,4-dioxane (5 mL) was added 5 mL of 4 M HCl solution in 1,4-dioxane. The resulting solution was stirred for 2 h and concentrated *in vacuo* to yield 46 as yellowish solid (390 mg, 87%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.99 (s, 1H), 9.42 – 9.31 (m, 2H), 9.16 (t, J = 1.4 Hz, 1H), 7.92 – 7.78 (m, 4H), 7.57 (d, J = 7.9 Hz, 2H), 4.61 (d, J = 5.9 Hz, 2H).

Scheme 5. Synthesis of compound 7.
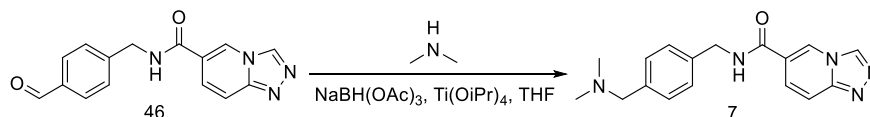
N-(4-((dimethylamino)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (7).

To a solution of **46** (35 mg, 0.12 mmol) in THF (3 mL) was added a solution of dimethylamine in THF (1 M, 0.25 mL) and Ti(O^{i}Pr)$_4$ (0.36 mmol, 102 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.36 mmol, 76 mg) was added and the mixture was refluxed under N$_2$. The reaction was monitored by TLC.

Small additional portions of NaBH(OAc)$_3$ were added to drive the reaction to completion. Upon complete consumption of 46, the reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2 × 20 mL). The combined organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **7** as white solid (11 mg, 30%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.29 (d, $J = 0.8$ Hz, 1H), 9.09 (t, $J = 1.4$ Hz, 1H), 7.91 – 7.76 (m, 2H), 7.57 – 7.44 (m, 4H), 4.65 (s, 2H), 4.29 (s, 2H), 2.83 (s, 6H). ^{13}C NMR (101 MHz, DMSO) δ 164.1, 148.8, 141.0, 138.1, 131.6, 129.5, 128.1, 127.4, 127.0, 121.5, 114.9, 59.5, 42.9, 41.8. ESI-HRMS (m/z): calculated for C$_{17}$H$_{20}$N$_5$O (M+H)$^+$: 310.1662, found: 310.1654.

Scheme 6. Synthesis of compound 8.
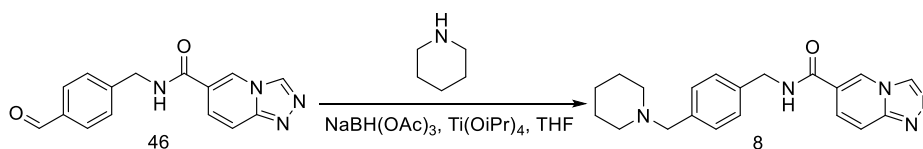
N-(4-(piperidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (8).

To a solution of **46** (35 mg, 0.12 mmol) in THF (3 mL) was added piperidine (20 mg, 0.24 mmol) and Ti(O^{i}Pr)$_4$ (0.36 mmol, 102 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.36 mmol, 76 mg) was added and the mixture was refluxed under N$_2$. The reaction was monitored by TLC. Small portions of NaBH(OAc)$_3$ were added to drive the reaction to completion. Upon complete consumption of **46**, the reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the

precipitate was filtered. The filtrate was then extracted by EtOAc (2×20 mL). The combined organic layers were dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **8** as white solid (12 mg, 28%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.26 (d, $J = 0.8$ Hz, 1H), 9.04 (t, $J = 1.4$ Hz, 1H), 7.87 – 7.76 (m, 2H), 7.38 – 7.29 (m, 4H), 4.60 (s, 2H), 3.50 (s, 2H), 2.37 (t, $J = 5.1$ Hz, 4H), 1.58 (p, $J = 5.6$ Hz, 4H), 1.45 (p, $J = 5.6$ Hz, 2H). ^{13}C NMR (101 MHz, MeOD) δ 164.8, 149.0, 137.5, 137.5, 135.9, 129.9, 127.4, 127.3, 126.4, 122.5, 114.2, 62.9, 53.8, 43.1, 25.0, 23.7. ESI-HRMS (m/z): calculated for $C_{20}H_{24}N_5O$ (M+H)$^+$: 350.1975, found: 350.1970.

Scheme 7. Synthesis of compound 9.
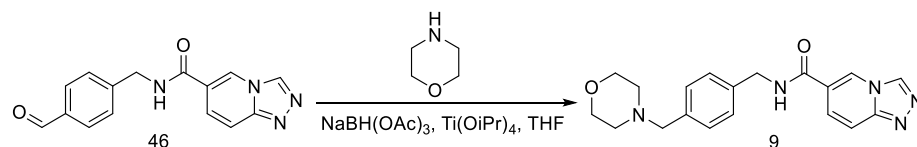
N-(4-(morpholinomethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (9).

To a solution of **46** (35 mg, 0.12 mmol) in THF (3 mL) was added morpholine (21 mg, 0.24 mmol) and Ti(O^{i}Pr)$_4$ (0.36 mmol, 102 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.36 mmol, 76 mg) was added and the mixture was refluxed under N_2. The reaction was monitored by TLC. Small portions of additional NaBH(OAc)$_3$ were added to drive the reaction to completion. Upon complete consumption of **46**, the reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2×20 mL). The combined organic layers were dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20%

methanol/EtOAc as eluent) to give **9** as white solid (15 mg, 43%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.37 (s, 1H), 9.22 (t, J = 5.9 Hz, 1H), 9.14 (t, J = 1.4 Hz, 1H), 7.87 − 7.75 (m, 2H), 7.32 − 7.24 (m, 4H), 4.49 (d, J = 5.8 Hz, 2H), 3.55 (t, J = 4.7 Hz, 4H), 3.42 (s, 2H), 2.37 − 2.27 (m, 4H). ^{13}C NMR (101 MHz, DMSO) δ 163.5, 148.4, 137.8, 137.6, 136.5, 129.0, 127.3, 126.8, 126.4, 121.1, 114.4, 66.2, 62.1, 53.1, 42.5. ESI-HRMS (m/z): calculated for $C_{19}H_{22}N_5O_2$ (M+H)$^{+}$: 352.1768, found: 352.1764.

Scheme 8. Synthesis of compound 10.
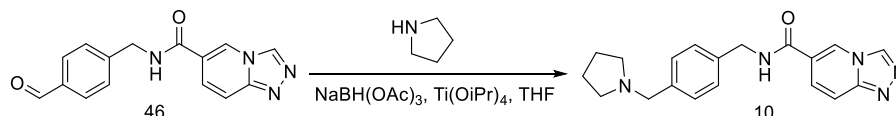
N-(4-(pyrrolidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (10).

To a solution of **46** (35 mg, 0.12 mmol) in THF (3 mL) was added pyrrolidine (17 mg, 0.24 mmol) and Ti(O^{i}Pr)$_4$ (0.36 mmol, 102 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.36 mmol, 76 mg) was added and the mixture was refluxed under N$_2$. The reaction was monitored by TLC. Small portions of NaBH(OAc)$_3$ were added to drive the reaction to completion. Upon complete consumption of **46**, the reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2 × 20 mL). The combined organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **10** as white solid (15 mg, 38%). ^{1}H NMR (300 MHz, Methanol-d_4) δ 9.28 (s, 1H), 9.07 (d, J = 1.6 Hz, 1H), 7.86 (dd, J = 9.7, 1.7 Hz, 1H), 7.79

(d, J = 9.7 Hz, 1H), 7.41 (s, 4H), 4.61 (s, 2H), 3.94 (s, 2H), 3.02 – 2.76 (m, 4H), 1.97 – 1.88 (m, 4H). ^{13}C NMR (75 MHz, CD$_3$OD) δ 164.8, 149.0, 138.8, 137.5, 134.0, 129.7, 127.7, 127.4, 126.4, 122.4, 114.2, 58.8, 53.4, 43.0, 22.6. ESI-HRMS (m/z): calculated for C$_{19}$H$_{22}$N$_5$O (M+H)$^+$: 336.1819 (M+H); found: 336.1810.

Scheme 9. Synthesis of compound 11.

N-(4-(azetidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (11).

To a solution of **46** (35 mg, 0.12 mmol) in THF (3 mL) was added azetidine hydrochloride (22 mg, 0.24 mmol) and Ti(O^{i}Pr)$_4$ (0.36 mmol, 102 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.36 mmol, 76 mg) was added and the mixture was refluxed under N$_2$. The reaction was monitored by TLC. Small portions of additional NaBH(OAc)$_3$ were added to drive the reaction to completion. Upon complete consumption of **46**, the reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2 × 20 mL). The combined organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **11** as white solid (15 mg, 40%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.29 (d, J = 0.8 Hz, 1H), 9.09 (t, J = 1.4 Hz, 1H), 7.91 – 7.76 (m, 2H), 7.52 – 7.42 (m, 4H), 4.63 (s, 2H), 4.30 (s, 2H), 4.06 (t, J = 8.1 Hz, 4H), 2.47 (p, J = 8.1 Hz, 2H). ^{13}C NMR (101 MHz, MeOD) δ 166.3, 150.4, 141.7, 138.9, 131.2, 130.9, 129.5, 128.8,

127.8, 123.7, 115.6, 59.2, 55.2, 44.3, 17.2. ESI-HRMS (m/z): calculated for $C_{18}H_{20}N_5O$ $(M+H)^+$: 322.1662; found: 322.1654.

Scheme 10. Synthesis of compound 12.
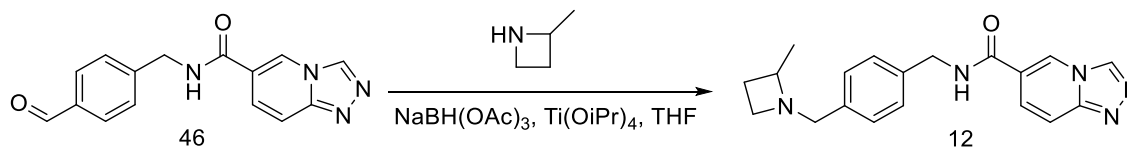
N-(4-((2-methylazetidin-1-yl)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (12).

To a solution of 46 (58 mg, 0.2 mmol) in DMF (1 mL) was added 2-methylazetidine hydrochloride (43 mg, 0.4 mmol) and $Ti(O^iPr)_4$ (0.6 mmol, 170 mg). The mixture was stirred under room temperature for 10 min. Then $NaBH(OAc)_3$ (0.6 mmol, 127 mg) was added and the mixture was heated to 70~75 °C for 24 h under N_2. The reaction mixture was diluted with saturated $NaHCO_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2×30 mL). The combined organic layers were dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give 12 as white solid (58 mg, 88%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.22 (d, J = 0.8 Hz, 1H), 9.02 (t, J = 1.4 Hz, 1H), 7.80 (dd, J = 9.6, 1.6 Hz, 1H), 7.71 (dt, J = 9.6, 1.0 Hz, 1H), 7.31 – 7.20 (m, 4H), 4.54 (s, 2H), 3.58 (d, J = 12.4 Hz, 1H), 3.46 (d, J = 12.4 Hz, 1H), 3.35 – 3.27 (m, 1H), 3.24 – 3.15 (m, 1H), 2.89 (dt, J = 9.8, 7.7 Hz, 1H), 2.04 (dtd, J = 10.2, 7.8, 2.3 Hz, 1H), 1.72 (tt, J = 10.0, 8.5 Hz, 1H), 0.97 (d, J = 6.2 Hz, 3H). ^{13}C NMR (101 MHz, CD$_3$OD) δ 167.0, 151.2, 139.8, 139.7, 138.3, 131.5, 129.7, 129.6, 128.6, 124.7,

116.4, 64.8, 63.7, 53.1, 45.3, 27.4, 22.1. ESI-HRMS (m/z): calculated for $C_{19}H_{22}N_5O$ $(M+H)^+$: 336.1819; found: 336.1814.

Scheme 11. Synthesis of compound 13.
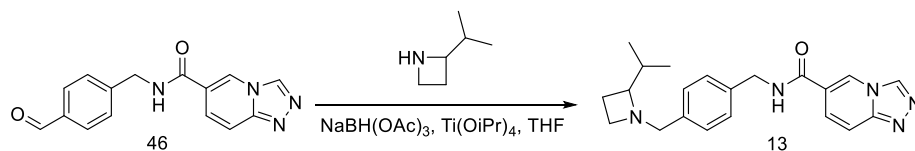
N-(4-((2-isopropylazetidin-1-yl)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (13).

To a solution of **46** (58 mg, 0.2 mmol) in DMF (1 mL) was added 2-isopropylazetidine hydrochloride (54 mg, 0.4 mmol) and Ti(O^{i}Pr)$_4$ (0.6 mmol, 170 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.6 mmol, 127 mg) was added and the mixture was heated to 70~75 °C for 24 h under N$_2$. The reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2×30 mL). The combined organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **13** as white solid (61 mg, 84%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.29 (dt, J = 1.9, 0.9 Hz, 1H), 9.08 (ddt, J = 4.2, 2.8, 1.4 Hz, 1H), 7.91 – 7.77 (m, 2H), 7.46 – 7.33 (m, 4H), 4.62 (s, 2H), 4.18 – 4.03 (m, 1H), 3.90 – 3.65 (m, 1H), 3.43 (s, 1H), 3.39 – 3.34 (m, 1H), 3.22 (d, J = 8.8 Hz, 1H), 2.25 (d, J = 10.6 Hz, 1H), 2.07 – 1.95 (m, 1H), 1.85 (dd, J = 14.8, 8.4 Hz, 1H), 0.95 (dd, J = 6.7, 0.8 Hz, 3H), 0.85 (dd, J = 6.7, 2.8 Hz, 3H). ESI-HRMS (m/z): calculated for $C_{21}H_{25}N_5O$ $(M+H)^+$: 364.2132; found: 364.2128.

Scheme 12. Synthesis of compound 14.
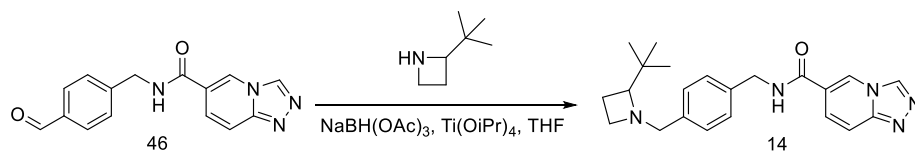
N-(4-((2-(tert-butyl)azetidin-1-yl)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (14).

To a solution of **46** (48 mg, 0.165 mmol) in DMF (1 mL) was added 2-tert-butyl azetidine hydrochloride (50 mg, 0.33 mmol) and Ti(O^{i}Pr)$_4$ (0.495 mmol, 141 mg). The mixture was stirred under room temperature for 10 min. Then NaBH(OAc)$_3$ (0.495 mmol, 105 mg) was added and the mixture was heated to 70~75 °C for 24 h under N$_2$. The reaction mixture was diluted with saturated NaHCO$_3$ solution (20 mL) and the precipitate was filtered. The filtrate was then extracted by EtOAc (2 × 30 mL). The combined organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to give **14** as white solid (50 mg, 80%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 9.30 (d, J = 0.8 Hz, 1H), 9.07 (t, J = 1.4 Hz, 1H), 7.92 – 7.79 (m, 2H), 7.40 – 7.29 (m, 4H), 4.61 (s, 2H), 3.95 (d, J = 12.9 Hz, 1H), 3.47 – 3.37 (m, 1H), 3.19 – 2.98 (m, 2H), 2.76 (d, J = 9.0 Hz, 1H), 2.01 – 1.85 (m, 2H), 0.93 (s, 9H). ESI-HRMS (m/z): calculated for C$_{22}$H$_{27}$N$_5$O (M+H)$^+$: 378.2288; found: 378.2286.

Scheme 13. Synthesis of compound 15.
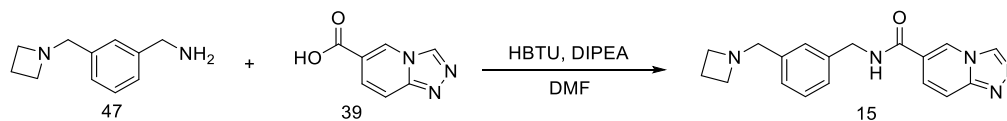
N-(3-(azetidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (15).

To a solution of **47** (22 mg, 0.125 mmol) and **39** (27 mg, 0.14 mmol) in DMF (0.4 mL) was added HBTU (52 mg, 0.14 mmol) and DIPEA (33 mg, 0.25 mmol). The resulting solution was then stirred under room temperature overnight. The reaction mixture then diluted with EtOAc (30 mL), then washed with saturated NaHCO$_3$ solution (2 × 5 mL), and brine (5 mL). The organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to yield **15** as white solid (25 mg, 62%). ^{1}H NMR (400 MHz, Methanol-d4) δ 9.30 (d, J = 0.9 Hz, 1H), 9.09 (t, J = 1.4 Hz, 1H), 7.94 – 7.79 (m, 2H), 7.38 – 7.30 (m, 3H), 7.24 (d, J = 1.0 Hz, 1H), 4.63 (s, 2H), 3.63 (s, 2H), 3.31 (t, J = 7.2 Hz, 4H), 2.21 – 2.07 (m, 2H). ESI-HRMS (m/z): calculated for C$_{18}$H$_{20}$N$_5$O (M+H)$^+$: 322.1662; found: 322.1656.
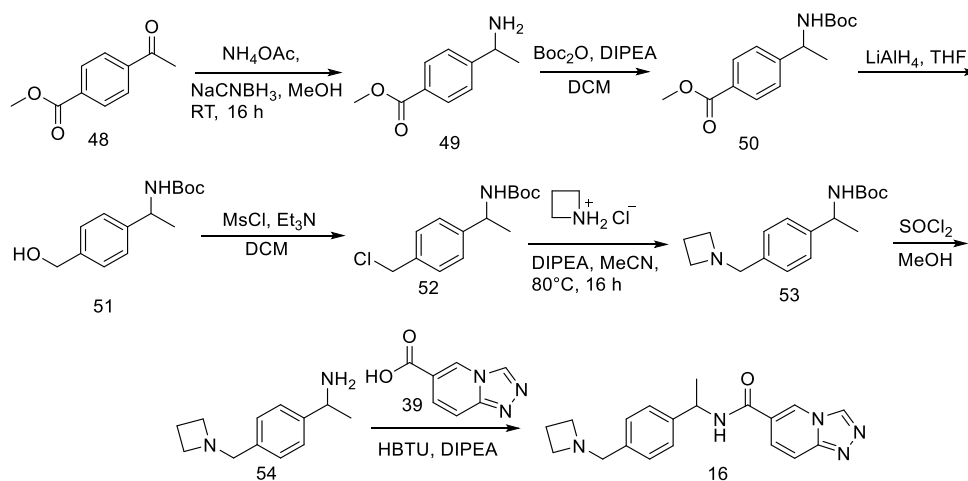
Methyl 4-(1-aminoethyl)benzoate (49).

Methyl 4-acetylbenzoate **46** (500 mg, 2.8 mmol), ammonium acetate (1.29 g, 16.8 mmol) and sodium cyanoborohydrate (263 mg, 4.2 mmol) were dissolved in 10 ml methanol and solution was stirred at room temperature for 16 h. The reaction mixture was

concentrated and acidified with 2 M HCl (5 mL), then extracted with DCM. The aqueous layer was basified with solid NaHCO$_3$ and extracted with DCM (2×30 mL). the combined DCM layers were dried over Na$_2$SO$_4$ and concentrated. The residue was used without further purification.

Methyl 4-(1-((tert-butoxycarbonyl)amino)ethyl) benzoate (50).

Methyl 4-(1-aminoethyl)benzoate **49** (250 mg, 1.39 mmol) was dissolved in DCM (5 mL) and Boc anhydride (348 mg, 1.6 mmol), DIPEA (0.5 ml, 2.7 mmol) and DMAP (17 mg, 0.139 mmol) were added and stirred for overnight. The reaction was washed with water and extracted with ethyl acetate (2×20 mL). the combined organic layers were dried over Na$_2$SO$_4$, concentrated and purified by silica gel column chromatography (20% EtOAc/Hexane) to yield **50** as white solid (300 mg, 77%).

tert-Butyl (1-(4-(hydroxymethyl)phenyl)ethyl) carbamate (51).

Methyl 4-(1-((tert-butoxycarbonyl)amino)ethyl)benzoate **50** (0.3 g, 1.0 mmol) was dissolved in THF (3 mL) and the solution cooled to below -5 °C in an ice/salt bath. LiAlH$_4$ (2 M in THF, 1 mL) was added dropwise over 10 min. Upon completion of addition, the reaction was stirred at 0 °C for 75 min. Water (0.16 mL) was added dropwise followed by 2 M aqueous NaOH solution (0.16 mL) and then water (0.16 mL). The suspension was stirred for 15 min and then diluted with EtOAc (15 mL). The mixture was dried over Na$_2$SO$_4$ and filtered and the resulting filtrate was concentrated *in vacuo* to afford the title compound, which was used without further purification (215 mg, 80%).

tert-Butyl (1-(4-(chloromethyl)phenyl)ethyl) carbamate (52).

To a stirred solution of **51** (200 mg, 0.8 mmol) in DCM (5 mL) was added methanesulfonyl chloride (108 mg, 0.95 mmol) and triethylamine (0.23 mL, 1.6 mmol). The solution was stirred for 16 h at room temperature then washed with water and brine. After separation, the organic phase was dried over Na_2SO_4, filtered and concentrated. The residue was purified by silica gel column chromatography (EtOAc/hexane 0 to 80%) to yield **52** white solid (100 mg, 46% yield).

tert-Butyl (1-(4-(azetidin-1-ylmethyl)phenyl)ethyl) carbamate (53).

To a stirred solution of **52** (100 mg, 0.37 mmol) in acetonitrile (5 mL) was added azetidine hydrochloride (41 mg, 0.44 mmol) and DIPEA (0.2 ml, 1.1 mmol). The solution was stirred at 80 °C for 16 h. The reaction mixture was diluted with water and extracted with DCM. The DCM layers were dried over Na_2SO_4 and concentrated to yield the crude product, which was used without further purification (100 mg).

1-(4-(Azetidin-1-ylmethyl)phenyl)ethan-1-amine (54).

To a stirred solution of **53** (100 mg. 0.3 mmol), 4 M HCl in 1,4-dioxane (0.5 ml, 1.8 mmol) was added and stirred at room temperature for 1 h. The reaction mixture was concentrated to yield 54 as off-white solid. ESI-HRMS (m/z): calculated for $C_{12}H_{19}N_2$ $(M+H)^+$: 191.1543; found: 191.1539.

N-(1-(4-(azetidin-1-ylmethyl)phenyl)ethyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (16).

To a stirred solution of **39** (50 mg, 0.3 mmol) and **54** (58 mg, 0.3 mmol) in DMF (1 ml) was added HBTU (136 mg, 0.36 mmol) and DIPEA (0.1 ml, 0.6 mmol). The solution was stirred at room temperature for 16 h. The reaction mixture was diluted with

water and extracted with DCM. The DCM layers were dried over Na_2SO_4 and concentrated to give the crude product, which was purified by flash chromatography to yield compound **16** as off-white solid (30 mg, 34%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.37 (s, 1H), 9.13 (s, 1H), 8.99 (d, *J* = 7.9 Hz, 1H), 7.80 (q, *J* = 9.7 Hz, 2H), 7.32 (d, *J* = 7.5 Hz, 2H), 7.21 (d, *J* = 7.8 Hz, 2H), 5.14 (dd, *J* = 13.9, 7.1 Hz, 1H), 3.47 (s, 2H), 3.08 (t, *J* = 6.8 Hz, 4H), 1.94 (dd, *J* = 13.4, 6.9 Hz, 2H), 1.47 (d, *J* = 7.0 Hz, 3H). ESI-HRMS (m/z): calculated for $C_{19}H_{22}N_5O$ (M+H)$^+$: 336.1819; found: 336.1808.

Scheme 14. Synthesis of compound 17.
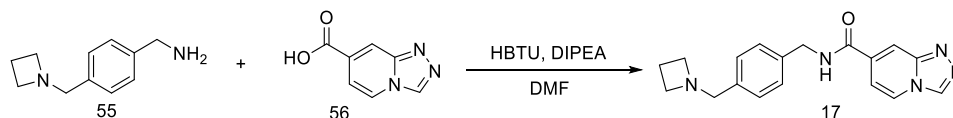
N-(4-(azetidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-7-carboxamide (17).

To a solution of **55** (44 mg, 0.25 mmol) and **56** (45 mg, 0.275 mmol) in DMF (1 mL) was added HBTU (104 mg, 0.275 mmol) and DIPEA (65 mg, 0.5 mmol). The resulting solution was then stirred under room temperature overnight. The reaction mixture then diluted with EtOAc (30 mL), then washed with saturated $NaHCO_3$ solution (2 × 5 mL), and brine (5 mL). The organic layers were dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to yield **17** as white solid (52 mg, 65%). ^{1}H NMR (400 MHz, Methanol-d4) δ 9.27 (s, 1H), 8.58 (d, *J* = 7.2 Hz, 1H), 8.26 (s, 1H), 7.44 (d, *J* = 7.3 Hz, 1H), 7.37 (d, *J* = 7.8 Hz, 2H), 7.29 (d, *J* = 7.1 Hz, 2H), 4.61 (s, 2H), 3.61 (s,

2H), 3.32 – 3.26 (m, 4H), 2.12 (p, J = 7.2 Hz, 2H). ESI-HRMS (m/z): calculated for

$C_{18}H_{20}N_5O$ (M+H)$^+$: 322.1662; found: 322.1653.

Scheme 15. Synthesis of compound 18.
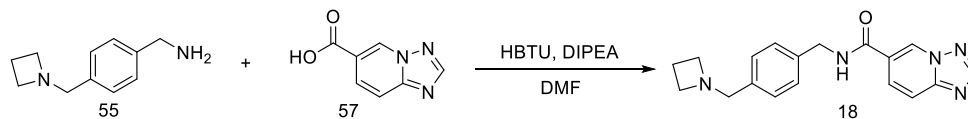
N-(4-(azetidin-1-ylmethyl)benzyl)-[1,2,4]triazolo[1,5-a]pyridine-6-carboxamide

(18).

To a solution of **55** (44 mg, 0.25 mmol) and **57** (45 mg, 0.275 mmol) in DMF (1

mL) was added HBTU (104 mg, 0.275 mmol) and DIPEA (65 mg, 0.5 mmol). The

resulting solution was then stirred under room temperature overnight. The reaction

mixture then diluted with EtOAc (30 mL), then washed with saturated $NaHCO_3$ solution

(2×5 mL), and brine (5 mL). The organic layers were dried with anhydrous Na_2SO_4 and

concentrated *in vacuo*. The residue was then purified by column chromatography (silica

gel, 20% methanol/EtOAc as eluent) to yield **18** as white solid (49 mg, 61%). ^{1}H NMR

(400 MHz, Methanol-d4) δ 9.34 (s, 1H), 8.52 (d, J = 1.4 Hz, 1H), 8.14 (dd, J = 9.3, 1.9

Hz, 1H), 7.86 (d, J = 9.3 Hz, 1H), 7.38 (d, J = 7.7 Hz, 2H), 7.30 (d, J = 7.3 Hz, 2H), 4.62

(s, 2H), 3.65 (s, 2H), 3.36 (d, J = 4.2 Hz, 4H), 2.14 (p, J = 7.2 Hz, 2H). ESI-HRMS (m/z):

calculated for $C_{18}H_{20}N_5O$ (M+H)$^+$: 322.1662; found: 322.1655.

Scheme 16. Synthesis of compound 19.
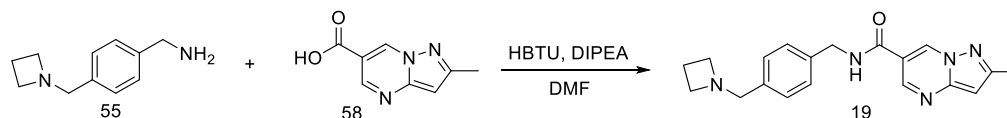
N-(4-(azetidin-1-ylmethyl)benzyl)-2-methylpyrazolo[1,5-a]pyrimidine-6-carboxamide (19).

To a solution of **55** (44 mg, 0.25 mmol) and **58** (49 mg, 0.275 mmol) in DMF (1 mL) was added HBTU (104 mg, 0.275 mmol) and DIPEA (65 mg, 0.5 mmol). The resulting solution was then stirred under room temperature overnight. The reaction mixture then diluted with EtOAc (30 mL), then washed with saturated NaHCO$_3$ solution (2 × 5 mL), and brine (5 mL). The organic layers were dried with anhydrous Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 20% methanol/EtOAc as eluent) to yield **19** as white solid (49 mg, 59%). ^{1}H NMR (400 MHz, Methanol-d4) δ 9.29 (d, J = 2.2 Hz, 1H), 8.88 (d, J = 2.2 Hz, 1H), 7.38 (d, J = 7.9 Hz, 2H), 7.29 (d, J = 8.0 Hz, 2H), 6.59 (s, 1H), 4.61 (s, 2H), 3.62 (s, 2H), 3.29 (d, J = 7.2 Hz, 4H), 2.53 (s, 3H), 2.13 (p, J = 7.2 Hz, 2H). ESI-HRMS (m/z): calculated for C$_{19}$H$_{22}$N$_5$O (M+H)$^+$: 336.1819; found: 336.1812.

Scheme 17. Synthesis of compound 20.
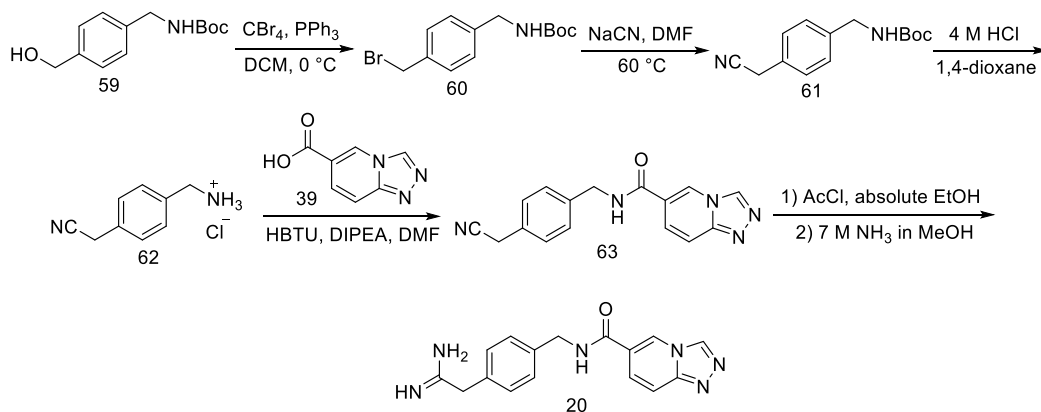
tert-Butyl (4-(bromomethyl)benzyl)carbamate (60).

To a solution of 59 (238 mg, 1 mmol) in DCM (2 mL) was added triphenylphosphine (316 mg, 1.2 mmol). Then carbon tetrabromide (400 mg, 1.2 mmol) was added in portions under ice water bath. The reaction was left under ice water bath for another 3 h. Then the reaction mixture was filtered and concentrated *in vacuo*. The residue was purified by column chromatography (10:1 hexanes/EtOAc) to yield 60 as white solid (252 mg, 83%).

tert-Butyl (4-(cyanomethyl)benzyl)carbamate (61).

To a solution of 60 (150 mg, 0.5 mmol) in 2 mL of DMF was added NaCN (50 mg, 1 mmol). The reaction mixture was then stirred under 60 °C for 4 h. The reaction mixture was cooled and diluted with water and extracted with DCM. Combined DCM layers were dried over Na_2SO_4 and then concentrated to yield 61 as white solid, which was used without further purification (96 mg, 78%).

N-(4-(cyanomethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (63).

To a solution of 61 (96 mg, 0.39 mmol) in 1 mL of 1,4-dioxane was added 4 M HCl solution in dioxane (4 mL). The reaction mixture was stirred at room temperature for 1 h and then concentrated to dryness *in vacuo*. The residue was dissolved in 2 mL of DMF, to which was added 39 (64 mg, 0.39 mmol), DIPEA (155 mg, 1.2 mmol) and HBTU (175 mg, 0.46 mmol). The reaction mixture was stirred at room temperature overnight and then diluted with DCM (20 mL), washed with saturated $NaHCO_3$ solution (2 × 20 mL), 1 M HCl (2 × 20 mL) and brine. The combined DCM layer was dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by flash chromatography (0-10% MeOH/DCM) to yield 63 as pale yellow solid (58 mg, 52%). ^{1}H NMR (300 MHz, DMSO-

d$_6$) δ 9.37 (s, 1H), 9.29 (t, *J* = 5.7 Hz, 1H), 9.15 (s, 1H), 7.84 (d, *J* = 9.7 Hz, 1H), 7.79 (d,

J = 9.7 Hz, 1H), 7.34 (q, *J* = 8.0 Hz, 4H), 4.50 (d, *J* = 5.8 Hz, 2H), 4.01 (s, 2H).

N-(4-(2-amino-2-iminoethyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide

(20).

To 1.5 mL of absolute EtOH was dropwise added 1 mL of acetylchloride under N_2

at 0 °C. The solution was stirred at 0 °C for 5 min before a solution of 63 (20 mg, 0.069

mmol) in absolute EtOH (0.5 mL) was added. The reaction mixture was stirred at room

temperature for 36 h. The reaction mixture was then evaporated to dryness under high

vacuum. To the residue was added a 7 M NH_3 solution in methanol (1 mL). The reaction

mixture was then stirred overnight and concentrated *in vacuo*. The residue was dissolved

in 1 M HCl solution, and washed with EtOAc to remove residual 63, then evaporated to

dryness to yield 20 as its hydrochloride salt (17 mg, 71%). ^{1}H NMR (300 MHz, D_2O) δ

9.44 (s, 1H), 9.19 (s, 1H), 8.20 (d, *J* = 9.4 Hz, 1H), 8.05 (d, *J* = 9.6 Hz, 1H), 7.45 (d, *J* =

8.0 Hz, 2H), 7.38 (d, *J* = 8.0 Hz, 1H), 4.64 (s, 2H), 3.88 (s, 2H). ESI-HRMS (m/z):

calculated for $C_{16}H_{17}N_6O$ (M+H)$^+$: 309.1458; found: 309.1451.

Scheme 18. Synthesis of compound 21.
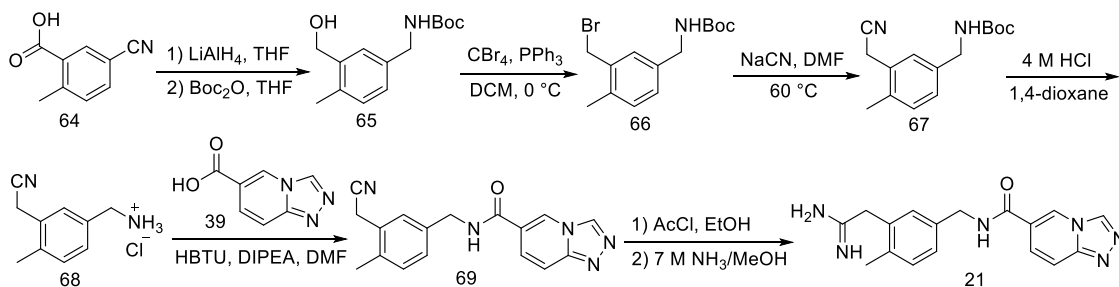
tert-Butyl (3-(hydroxymethyl)-4-methylbenzyl) carbamate (65).

2-Methyl-5-cyanobenzoic acid **64** (5 mmol, 0.81 g) was dissolved in anhydrous THF (15 mL). A solution of LiAlH$_4$ in THF (1.0 M, 20 mL) was added dropwise to the solution under N$_2$ at 0 °C. After completion of addition, the reaction mixture was heated to reflux overnight. The solution was then cooled to room temperature and then to 0 °C. Water (5 mL) was added dropwise, followed by 2 M NaOH solution (5 mL). After stirring for another 10 min, the mixture was filtered over celite. To the filtrate was added Boc$_2$O (5 mmol, 1.09 g) and the solution was stirred at room temperature for 4 h. The solution was then concentrated *in vacuo* to yield crude **65** as yellow oil (0.75 g, 60%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.28 (s, 1H), 7.16-7.09 (m, 2H), 4.84 (brs, 1H), 4.68 (d, *J* = 5.2 Hz, 2H), 4.27 (d, *J* = 5.9 Hz, 2H), 2.32 (s, 3H), 1.66 (s, 1H), 1.45 (s, 9H).

tert-Butyl (3-(bromomethyl)-4-methylbenzyl) carbamate (66).

To a solution of **65** (1.2 mmol, 300 mg) in DCM (10 mL) was added triphenylphosphine (1.44 mmol, 380 mg). Then carbon tetrabromide (1.44 mmol, 480 mg) was added in portions at 0 °C. The reaction was stirred at 0 °C for 3 h. Then the reaction mixture was filtered and concentrated *in vacuo*. The residue was purified by column chromatography (hexanes/EtOAc=10:1) to yield **66** as white solid (284 mg, 76%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.23 (s, 1H), 7.19-7.13 (m, 2H), 4.81 (brs, 1H), 4.59 (s, 2H), 4.27 (d, *J* = 5.9 Hz, 2H), 2.40 (s, 3H), 1.46 (s, 9H).

tert-Butyl (3-(cyanomethyl)-4-methylbenzyl) carbamate (67).

To a solution of **66** (156 mg, 0.5 mmol) in 5 mL of DMF was added NaCN (50 mg, 1 mmol). The reaction mixture was then stirred under 60 °C for 4 h. The reaction mixture was cooled and diluted with water and extracted with DCM. Combined DCM

layers were dried over Na_2SO_4 and then concentrated to yield **67** as white solid, which was used without further purification (78 mg, 60%).

N-(3-(cyanomethyl)-4-methylbenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (69).

To a solution of **67** (78 mg, 0.3 mmol) in 1 mL of 1,4-dioxane was added 4 M HCl solution in dioxane (4 mL). The reaction mixture was stirred at room temperature for 1 h and then concentrated to dryness *in vacuo*. The residue was dissolved in 2 mL of DMF, to which was added **39** (49 mg, 0.3 mmol), DIPEA (116 mg, 0.9 mmol) and HBTU (137 mg, 0.36 mmol). The reaction mixture was stirred at room temperature overnight and then diluted with DCM (20 mL), washed with saturated $NaHCO_3$ solution (2×20 mL), 1 M HCl (2×20 mL) and brine. The combined DCM layer was dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by flash chromatography (0-10% MeOH/DCM) to yield **69** as white solid (46 mg, 50%). ^{1}H NMR (300 MHz, DMSO-d_6) δ 9.37 (t, $J = 0.7$ Hz, 1H), 9.24 (t, $J = 5.9$ Hz, 1H), 9.14 (s, 1H), 7.88-7.76 (m, 2H), 7.32 (s, 1H), 7.25-7.18 (m, 2H), 4.47 (d, $J = 5.9$ Hz, 2H), 3.98 (s, 2H), 2.27 (s, 3H).

N-(3-(2-amino-2-iminoethyl)-4-methylbenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (21).

To 1.5 mL of absolute EtOH was dropwise added 1 mL acetylchloride under N_2 at 0 °C. The solution was stirred at 0 °C for 5 min before a solution of **69** (20 mg, 0.066 mmol) in absolute EtOH (0.5 mL) was added. The reaction mixture was stirred at room temperature for 36 h. The reaction mixture was then evaporated to dryness under high vacuum. To the residue was added a 7 M NH_3 solution in methanol (1 mL). The reaction

mixture was then stirred overnight and concentrated *in vacuo*. The residue was dissolved in 1 M HCl solution, and washed with EtOAc to remove residual **69**, then evaporated to dryness to yield **21** as its hydrochloride salt (14 mg, 60%). ^{1}H NMR (300 MHz, D$_2$O) δ 9.46 (s, 1H), 9.22 (s, 1H), 8.33 (d, *J* = 9.3 Hz, 1H), 8.08 (d, *J* = 9.2 Hz, 1H), 7.31 – 7.13 (m, 3H), 4.50 (s, 2H), 3.83 (s, 2H), 2.14 (s, 3H). ESI-HRMS (m/z): calculated for C$_{17}$H$_{19}$N$_6$O (M+H)$^+$: 323.1615; found: 323.1611.

Scheme 19. Synthesis of compound 22.
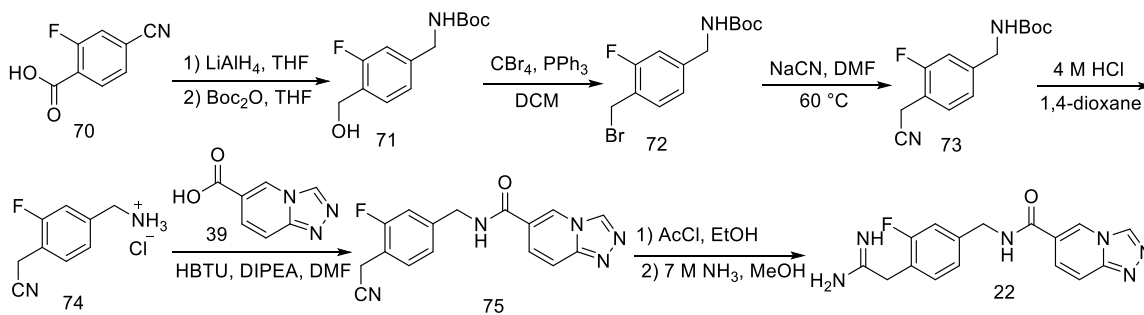
tert-Butyl (3-fluoro-4-(hydroxymethyl)benzyl) carbamate (71).

2-Fluoro-4-cyanobenzoic acid 70 (5 mmol, 0.83 g) was dissolved in anhydrous THF (15 mL). A solution of LiAlH$_4$ in THF (1.0 M, 20 mL) was added dropwise to the above solution under N$_2$ at 0 °C. After completion of addition, the reaction mixture was heated to reflux overnight. The solution was then cooled to room temperature and then to 0 °C. Water (5 mL) was added dropwise, followed by 2 M NaOH solution (5 mL). After stirring for another 10 min, the mixture was filtered over celite. To the filtrate was added Boc$_2$O (5 mmol, 1.09 g) and the solution was stirred at room temperature for 4 h. The solution was then concentrated *in vacuo* to yield crude 71 as yellow oil (0.71 g, 55%).

tert-Butyl (4-(bromomethyl)-3-fluorobenzyl) carbamate (72).

To a solution of 71 (1.2 mmol, 306 mg) in DCM (10 mL) was added triphenylphosphine (1.44 mmol, 380 mg). Then carbon tetrabromide (1.44 mmol, 480 mg) was added in portions at 0 °C. The reaction was stirred at 0 °C for 3 h. Then the reaction mixture was filtered and concentrated *in vacuo*. The residue was purified by column chromatography (hexanes/EtOAc=10:1) to yield 72 as white solid (250 mg, 66%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.34 (t, *J* = 7.8 Hz, 1H), 7.13 – 6.95 (m, 2H), 4.88 (brs, 1H), 4.50 (s, 2H), 4.30 (d, *J* = 6.2 Hz, 2H), 1.46 (s, 9H).

tert-Butyl (4-(cyanomethyl)-3-fluorobenzyl) carbamate (73).

To a solution of 72 (159 mg, 0.5 mmol) in 5 mL of DMF was added NaCN (50 mg, 1 mmol). The reaction mixture was then stirred under 60 °C for 4 h. The reaction mixture was cooled and diluted with water and extracted with DCM. Combined DCM layers were dried over Na_2SO_4 and then concentrated to yield 73 as white solid, which was used without further purification (85 mg, 64%).

N-(4-(cyanomethyl)-3-fluorobenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (75).

To a solution of 73 (79 mg, 0.3 mmol) in 1 mL of 1,4-dioxane was added 4 M HCl solution in dioxane (4 mL). The reaction mixture was stirred at room temperature for 1 h and then concentrated to dryness *in vacuo*. The residue was dissolved in 2 mL of DMF, to which was added 39 (49 mg, 0.3 mmol), DIPEA (116 mg, 0.9 mmol) and HBTU (137 mg, 0.36 mmol). The reaction mixture was stirred at room temperature overnight and then diluted with DCM (20 mL), washed with saturated $NaHCO_3$ solution (2 × 20 mL), 1 M HCl (2 × 20 mL) and brine. The combined DCM layer was dried over Na_2SO_4 and

concentrated *in vacuo*. The residue was purified by flash chromatography (0-10% MeOH/DCM) to yield 75 as white solid (51 mg, 55%). ^{1}H NMR (300 MHz, Methanol-d_4) δ 9.29 (s, 1H), 9.09 (s, 1H), 7.83 (t, J = 8.7 Hz, 2H), 7.46 (t, J = 7.8 Hz, 1H), 7.34 – 7.16 (m, 2H), 4.63 (s, 2H), 3.92 (s, 2H).

N-(4-(2-amino-2-iminoethyl)-3-fluorobenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (22).

To 1.5 mL of absolute EtOH was added dropwise 1 mL of acetylchloride under N_2 at 0 °C. The solution was stirred at 0 °C for 5 min before a solution of 75 (20 mg, 0.065 mmol) in absolute EtOH (0.5 mL) was added. The reaction mixture was stirred at room temperature for 36 h. The reaction mixture was then evaporated to dryness under high vacuum. To the residue was added a 7 M NH_3 solution in methanol (1 mL). The reaction mixture was then stirred overnight and concentrated *in vacuo*. The residue was dissolved in 1 M HCl solution, and washed with EtOAc to remove residual 75, then evaporated to dryness to yield 22 as its hydrochloride salt (10 mg, 47%). ^{1}H NMR (300 MHz, D_2O) δ 9.46 (s, 1H), 9.22 (s, 1H), 8.22 (dd, J = 9.6, 1.5 Hz, 1H), 8.06 (d, J = 9.6 Hz, 1H), 7.41 (t, J = 8.0 Hz, 1H), 7.33 – 7.14 (m, 2H), 4.63 (s, 2H), 3.93 (s, 2H). ESI-HRMS (m/z): calculated for $C_{16}H_{16}FN_6O$ (M+H)$^+$: 327.1364; found: 327.1352.

Scheme 20. Synthesis of compound 23.
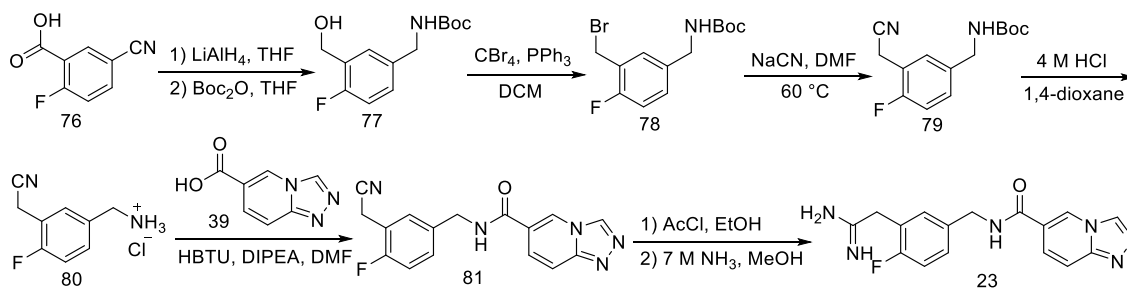
tert-Butyl (4-fluoro-3-(hydroxymethyl)benzyl) carbamate (77).

2-Fluoro-5-cyanobenzoic acid **76** (5 mmol, 0.83 g) was dissolved in anhydrous THF (15 mL). A solution of LiAlH$_4$ in THF (1.0 M, 20 mL) was added dropwise to the solution under N$_2$ at 0 °C. After completion of addition, the reaction mixture was heated to reflux overnight. The solution was then cooled to room temperature and then to 0 °C. Water (5 mL) was added dropwise, followed by 2 M NaOH solution (5 mL). After stirring for another 10 min, the mixture was filtered over celite. To the filtrate was added Boc$_2$O (5 mmol, 1.09 g) and the solution was stirred at room temperature for 4 h. The solution was then concentrated *in vacuo* to yield crude **77** as yellow oil (0.81 g, 59%).

tert-Butyl (3-(bromomethyl)-4-fluorobenzyl) carbamate (78).

To a solution of **77** (1.2 mmol, 306 mg) in DCM (10 mL) was added triphenylphosphine (1.44 mmol, 380 mg). Then carbon tetrabromide (1.44 mmol, 480 mg) was added in portions at 0 °C. The reaction was stirred at 0 °C for 3 h. Then the reaction mixture was filtered and concentrated *in vacuo*. The residue was purified by column chromatography (hexanes/EtOAc=10:1) to yield **78** as white solid (234 mg, 62%). ^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.34 – 7.27 (m, 1H), 7.25-7.18 (m, 1H), 7.05-6.96 (m, 1H), 4.87 (brs, 1H), 4.49 (s, 2H), 4.29 (d, J = 6.2 Hz, 2H), 1.46 (s, 9H).

tert-Butyl (3-(cyanomethyl)-4-fluorobenzyl) carbamate (79).

To a solution of **78** (159 mg, 0.5 mmol) in 5 mL of DMF was added NaCN (50 mg, 1 mmol). The reaction mixture was then stirred under 60 °C for 4 h. The reaction mixture was cooled and diluted with water and extracted with DCM. Combined DCM

layers were dried over Na_2SO_4 and then concentrated to yield **79** as white solid, which was used without further purification (79 mg, 60%).

N-(3-(cyanomethyl)-4-fluorobenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (81).

To a solution of **79** (79 mg, 0.3 mmol) in 1 mL of 1,4-dioxane was added 4 M HCl solution in dioxane (4 mL). The reaction mixture was stirred at room temperature for 1 h and then concentrated to dryness *in vacuo*. The residue was dissolved in 2 mL of DMF, to which was added **39** (49 mg, 0.3 mmol), DIPEA (116 mg, 0.9 mmol) and HBTU (137 mg, 0.36 mmol). The reaction mixture was stirred at room temperature overnight and then diluted with DCM (20 mL), washed with saturated $NaHCO_3$ solution (2×20 mL), 1 M HCl (2×20 mL) and brine. The combined DCM layer was dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by flash chromatography (0-10% MeOH/DCM) to yield **81** as white solid (57 mg, 62%). ^{1}H NMR (300 MHz, DMSO-d_6) δ 9.37 (s, 1H), 9.29 (t, $J = 5.6$ Hz, 1H), 9.14 (s, 1H), 7.85 (d, $J = 9.6$ Hz, 1H), 7.78 (dd, $J = 9.7, 1.5$ Hz, 1H), 7.48 – 7.20 (m, 3H), 4.50 (d, $J = 6.1$ Hz, 2H), 4.05 (s, 2H).

N-(4-(2-amino-2-iminoethyl)-3-fluorobenzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (23).

To 1.5 mL of absolute EtOH was dropwise added 1 mL of acetylchloride under N_2 at 0 °C. The solution was stirred at 0 °C for 5 min before a solution of **81** (20 mg, 0.065 mmol) in absolution EtOH (0.5 mL) was added. The reaction mixture was stirred at room temperature for 36 h. The reaction mixture was then evaporated to dryness under high vacuum. To the residue was added a 7 M NH_3 solution in methanol (1 mL). The reaction

mixture was then stirred overnight and concentrated *in vacuo*. The residue was dissolved in 1 M HCl solution, and washed with EtOAc to remove residual **81**, then evaporated to dryness to yield **23** as its hydrochloride salt (13 mg, 60%). ^{1}H NMR (300 MHz, D$_2$O) δ 9.51 (s, 1H), 9.27 (s, 1H), 8.35 (dd, $J = 9.6, 1.6$ Hz, 1H), 8.13 (d, $J = 9.6$ Hz, 1H), 7.50 – 7.34 (m, 2H), 7.27 – 7.10 (m, 1H), 4.62 (s 2H), 3.91 (s, 2H). ESI-HRMS (m/z): calculated for C$_{16}$H$_{16}$FN$_6$O (M+H)$^+$: 327.1364; found: 327.1359.

Scheme 21. Synthesis of compound 24.
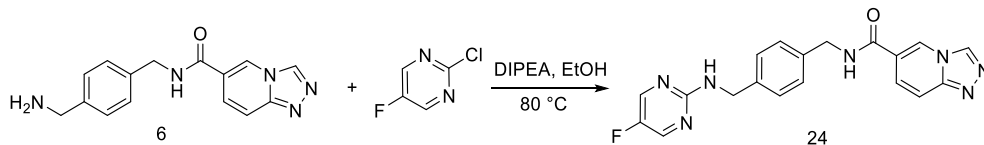
N-(4-((((5-fluoropyrimidin-2-yl)amino)methyl)benzyl) -[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (24).

To a stirred solution of 2-chloro-5-fluoropyrimidine (24 mg, 0.18 mmol) and amine **6** (48 mg. 0.15 mmol) in ethanol (1 mL) was added DIPEA (0.1 mL, 0.57 mmol) and the reaction mixture was heated to 75 °C for 36 h. The reaction was concentrated and purified by flash chromatography (silica gel, 5% methanol/EtOAc as eluent) to yield **24** as light-yellow solid (22 mg, 39%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.36 (s, 1H), 9.21 (t, $J = 5.9$ Hz, 1H), 9.12 (t, $J = 1.4$ Hz, 1H), 8.32 (d, $J = 1.0$ Hz, 2H), 7.86 – 7.73 (m, 3H), 7.26 (s, 4H), 4.46 (d, $J = 5.9$ Hz, 2H), 4.43 (d, $J = 6.4$ Hz, 2H). ^{13}C NMR (101 MHz, DMSO) δ 163.9, 159.9, 153.3, 150.9, 148.8, 146.1, 145.8, 139.4, 138.0, 137.8, 127.8, 127.5, 127.2, 126.9, 121.6, 114.9, 44.7, 43.0. ESI-HRMS (m/z): calculated for C$_{19}$H$_{17}$FN$_7$O (M+H)$^+$: 378.1473; found: 378.1465.

Scheme 22. Synthesis of compound 25.
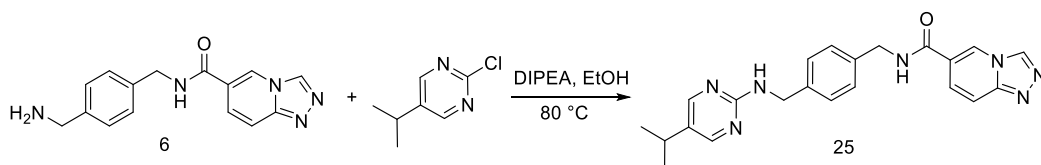
N-(4-(((5-isopropylpyrimidin-2-yl)amino)methyl) benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (25).

To a stirred solution of amine **6** (50 mg, 0.17 mmol) and 2-chloro-5-isopropylpyrimidine (15 mg, 0.17 mmol) in ethanol (2 mL) was added DIPEA (68 mg, 0.53 mmol) and heated to 80 °C for 48 h. The reaction mixture was concentrated and purified by flash chromatography (0-10% methanol/DCM) to yield **25** as off-white solid (10 mg, 18%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.35 (d, J = 0.8 Hz, 1H), 9.21 (t, J = 5.8 Hz, 1H), 9.13 (t, J = 1.4 Hz, 1H), 8.16 (s, 2H), 7.86 – 7.73 (m, 2H), 7.49 (t, J = 6.4 Hz, 1H), 7.26 (s, 4H), 4.49 – 4.40 (m, 4H), 2.71 (hept, J = 6.9 Hz, 1H), 1.15 (d, J = 6.9 Hz, 7H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 163.9, 161.7, 156.5, 148.8, 139.9, 138.0, 137.7, 129.2, 127.7, 127.5, 127.2, 127.0, 121.6, 114.9, 44.3, 43.0, 28.7, 24.0. ESI-HRMS (m/z): calculated for $C_{22}H_{23}N_7ONa$ (M+Na)$^+$: 424.1856, found: 424.1849.

Scheme 23. Synthesis of compound 26.
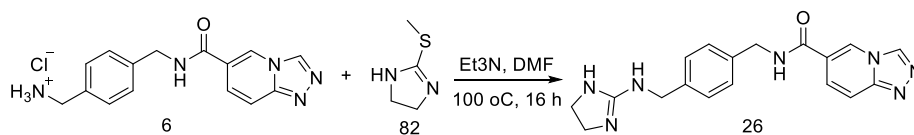
N-(4-((((4,5-dihydro-1H-imidazol-2-yl)amino)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (26).

To a stirred solution of amine **6** (95 mg, 0.3 mmol) and **82** (70 mg, 0.6 mmol) in DMF (2 mL) was added triethylamine (91 mg, 0.9 mmol) and heated to 100 °C for 16 h.

The reaction mixture was concentrated under reduce pressure. Water was added to the crude product and the solid was filtered and concentrated under reduced pressure to provide the crude compound **26** (20 mg, 19%). The crude compound was purified by RP-HPLC (HPLC gradient: 0-70 min: 95% A to 50% A). ^{1}H NMR (300 MHz, DMSO-d_6) δ 9.38 (s, 1H), 9.28 (s, 1H), 9.14 (s, 1H), 8.67 (s, 1H), 7.82 (q, J = 9.7 Hz, 2H), 7.46 – 7.18 (m, 4H), 4.49 (d, J = 5.9 Hz, 2H), 4.35 (d, J = 5.9 Hz, 2H), 3.60 (s, 4H). ESI-HRMS (m/z): calculated for $C_{18}H_{20}N_7O$ (M+H)$^+$: 350.1724; found: 350.1724.

Scheme 24. Synthesis of compound 27.
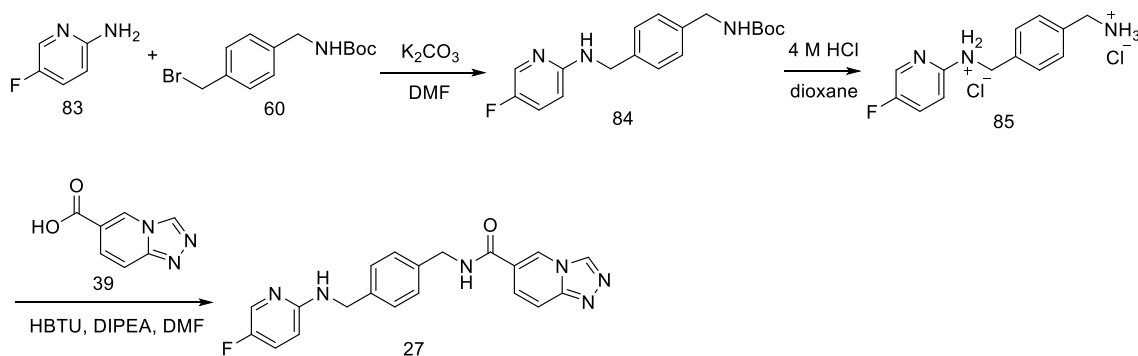
tert-Butyl (4-(((5-fluoropyridin-2-yl)amino)methyl) benzyl)carbamate (84).

To a stirred solution of **83** (1 mmol, 300 mg) in 5 mL DMF was added 60 (1 mmol, 112 mg) and K_2CO_3 (1 mmol, 138 mg). The reaction mixture was stirred at room temperature overnight. The reaction mixture was diluted with water (25 mL) and then extracted with DCM (2 × 20 mL). The organic layer was dried over anhydrous Na_2SO_4 and concentrated. The residue was then purified by flash chromatography to yield **84** as white solid (100 mg, 30%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.96 (d, J = 3.0 Hz, 1H), 7.31 (d, J = 7.9 Hz, 2H), 7.25 (d, J = 8.2 Hz, 2H), 7.19 (ddd, J = 9.0, 7.9, 3.0 Hz, 1H), 6.32 (dd,

J = 9.1, 3.4 Hz, 1H), 4.84 (s, 2H), 4.46 (d, J = 5.6 Hz, 2H), 4.30 (d, J = 6.0 Hz, 2H), 1.46 (s, 9H).

N-(4-(aminomethyl)benzyl)-5-fluoropyridin-2-amine dihydrochloride (85).

To a stirred solution of **84** (0.3 mmol, 100 mg) in 1,4-dioxane (1 mL) was add a 4 M HCl solution in dioxane (4 mL). The solution was stirred at room temperature for 1 h. The reaction mixture was then concentrated *in vacuo* to yield **85** as white solid (85 mg, 95%), which was used without further purification.

N-(4-(((5-fluoropyridin-2-yl)amino)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (27).

To a stirred solution of **85** (54 mg, 0.18mmol) and **39** (32 mg, 0.20 mmol) in DMF (2 mL) was added DIPEA (0.15 mL, 0.8 mmol) and HBTU (80 mg, 0.22 mmol). The solution was then stirred at room temperature overnight. The reaction mixture was then diluted with water (20 mL) and extracted with EtOAc (2 × 20 mL). The organic layer was dried over anhydrous Na_2SO_4 and concentrated. The residue was then purified by flash chromatography to provide **27** as pale yellow solid (38 mg, 56%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.36 (d, J = 0.8 Hz, 1H), 9.21 (t, J = 5.9 Hz, 1H), 9.13 (t, J = 1.4 Hz, 1H), 7.90 (d, J = 3.1 Hz, 1H), 7.87 − 7.73 (m, 2H), 7.33 (td, J = 8.8, 3.1 Hz, 1H), 7.28 (s, 4H), 7.04 (t, J = 6.0 Hz, 1H), 6.51 (dd, J = 9.1, 3.6 Hz, 1H), 4.47 (d, J = 5.9 Hz, 2H), 4.40 (d, J = 6.0 Hz, 2H). ^{13}C NMR (101 MHz, DMSO) δ 163.5, 155.7, 153.7, 151.3, 148.4, 139.2, 137.6, 137.4, 133.5, 133.3, 127.3, 127.2, 126.7, 126.5, 125.1, 124.9, 121.1, 114.4, 108.8, 108.7, 44.4, 42.6. ESI-HRMS (m/z): calculated for $C_{20}H_{17}FN_6O$ (M+H)[+]: 377.1521; found: 377.1520.

Scheme 25. Synthesis of compound 28.
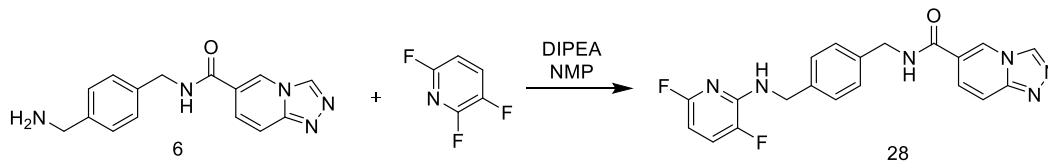
N-(4-(((3,6-difluoropyridin-2-yl)amino)methyl)benzyl)-[1,2,4]triazolo[4,3-a]pyridine-6-carboxamide (28).

To a stirred solution of 2,3,6-trifluoropyridine (32 mg, 0.24 mmol) and amine **6** (63 mg. 0.2 mmol) in N-methyl-2-pyrrolidone (NMP, 0.5 mL) was added DIPEA (0.1 mL, 0.57 mmol), and the reaction mixture was heated to 90 °C overnight. The reaction mixture was cooled and then diluted with EtOAc (25 mL), then washed with brine (2 x 5 mL). The organic layers were dried with anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was then purified by column chromatography (silica gel, 5% methanol/EtOAc as eluent) to yield **28** as light-yellow solid (30 mg, 38%). ^{1}H NMR (400 MHz, Methanol-d4) δ 9.27 (d, J = 1.0 Hz, 1H), 9.04 (q, J = 1.4 Hz, 1H), 7.90 – 7.76 (m, 2H), 7.41 – 7.25 (m, 5H), 6.02 (dtd, J = 8.2, 3.1, 2.6, 1.3 Hz, 1H), 4.59 (d, J = 1.1 Hz, 2H), 4.57 (s, 2H). ESI-HRMS (m/z): calculated for $C_{20}H_{17}F_2N_6O$ (M+H)$^+$: 395.1426; found: 395.1426.

2.4.5 Surface plasmon resonance (SPR) assay

All SPR experiments were performed on a Biosensing BI-4500 instrument with 1 × PBS with 0.1% DMSO as running buffer and a flow rate at 60 µL/min. His-ENL YEATS was immobilized through EDC/NHS coupling on CM dextran coated sensor chips (Biosensing). Sensor chip was activated by flowing 200 µL 0.05 M NHS plus 0.2 M EDC

solution through the surface. Then 200 µL of 6 µM His-ENL YEATS in 10 mM NaOAc/HOAc, 150 mM NaCl, pH 5.5 was injected and flowed through the activated surface. 100 µL of 1 M ethanolamine pH 7.8 solution was injected to block the remaining activated ester on the surface. Compound 11 and 24 was dissolved in running buffer and subjected to a 2-fold serial dilution for a total of 4 concentrations ranging from 5.6 µM to 0.7 µM (for 24 5.4 µM to 0.675 µM). For each cycle, 350 µL of compound solution was injected and flow through the surface followed by a 600 s delay for dissociation. Prior to the first cycle, 350 µL of running buffer was injected for baseline calibration. A control flow channel was set up in parallel without His-ENL YEATS immobilization to subtract non-specific binding signals. The data analysis was performed on the kinetic analysis software provided by Biosensing Instrument Inc. and fitted into the Langmuir 1:1 binding model.

2.4.6 Nuclear magnetic resonance (NMR) experiments for ENL YEATS domain

For NMR experiments, the YEATS domain of ENL (aa 1-148) was expressed as a C-terminal, uncleavable 6xHis fusion protein (plasmid was a generous gift from Oleg Fedorov). The 15N-labeled protein was expressed in E. coli Rosetta-2 (DE3) pLysS cells grown in NH_4Cl (Sigma-Aldrich) minimal media. After induction with IPTG (final concentration 0.5 mM) (Gold biotechnology) for 18 hrs at 16 °C, cells were harvested via centrifugation and lysed by sonication. The uniformly ^{15}N-labeled YEATS domain was incubated with Ni-NTA resins (Thermo Fisher Sci), washed and eluted with imidazole.

The protein was further purified by size exclusion chromatography, concentrated (Millipore) and stored at -80 °C.

NMR experiments were performed at 298 K on a Varian INOVA 600 MHz spectrometer. The ^{1}H,^{15}N HSQC spectra of 0.2 mM uniformly ^{15}N-labeled YEATS domain (25-50 mM Tris-HCl pH 7.5 buffer, supplemented with 150 mM NaCl, 0.2 mM TCEP and 10% D_2O) were collected in the presence of increasing amount of H3K27cr (aa 22-31) peptide (synthesized by Synpeptide) or compound **7**. NMR data were processed and analyzed with NMRPipe and NMRDraw as previously described[118].

2.4.7 Competitive peptide pulldown assay

Compounds at indicated concentrations were mixed with 2 μg of GST-fused YEATS proteins in 300 μL binding buffer (50 mM Tris-HCl pH7.5, 250 mM NaCl, 0.1% NP-40, 1 mM PMSF) and rotated at 4ºC for 1h. Then, 0.5 μg of biotinylated histone peptides with different modifications were added and incubated for 4 h. Streptavidin magnetic beads (Amersham) were added to the mixture, and the mixture was incubated for 1 h with rotation. The beads were then washed three times and analyzed using SDS-PAGE and western blotting.

2.4.8 IC$_{50}$ determination in the inhibition of YEATS domains

To assess the specificity of **7**, **11** and **24**, their IC$_{50}$ values in inhibition of the four human YEATS domain proteins binding to targeted histone peptides were determined in AlphaScreen assays. Compounds were subjected to twelve 3-fold serial dilutions, for a

total of thirteen concentrations ranging from 54 µM to 0.1 nM for dose response curve. IC_{50} values were determined from the plot using nonlinear regression of variable slope (four parameters) and curve fitting performed by the GraphPad Prism software.

2.4.9 Cell growth inhibition assay

Human Leukemia MV4;11 and MOLM13 cells were maintained in RPMI-1640 (Cellgro) supplemented with 10% fetal bovine serum (Sigma). Human U2OS cells were maintained in DMEM (Cellgro) supplemented with 10% fetal bovine serum. Five thousand cells were seeded in 96-well plate in 100 µL medium, treated with DMSO or compounds at indicated concentrations for 3 days. Cell viability was measured using the CellTiter-Glo luminescent cell viability assay kit (Promega) according to the manufacturer's instruction. Survived cells were calculated as % relative to DMSO treated cells.

2.4.10 RNA extraction and RT-qPCR

Total RNA was extracted using the RNeasy plus kit (Qiagen) and reverse transcribed using an iScript cDNA synthesis kit (Bio-Rad). Quantitative real-time PCR (qPCR) analyses were performed as described previously using PowerUp SYBR Green PCR Master Mix and the Bio-Rad CFX96 real-time PCR detection system. Gene expressions were calculated following normalization to GAPDH levels using the comparative Ct (cycle threshold) method.

2.4.11 Combinatorial treatment of compound 7 and JQ1

MOLM13 or MV4;11 cells were treated with DMSO, compound **7**, JQ1 or combination of compounds **7** and JQ1 at indicated concentrations for 6 days. Cell viability was measured using the CellTiter-Glo luminescent cell viability assay kit (Promega). Survived cells were calculated as % relative to DMSO treated cells. Synergistic interactions were analyzed and visualized using the Combenefit software[114].

CHAPTER III

STRUCTURE-BASED DRUG DEVELOPMENT FOR SARS-COV-2 MAIN PROTEASE*

3.1 Introduction

Coronaviruses (CoVs) are a group of related RNA viruses that cause diseases in a wide range of vertebrates including humans and domestic animals[119]. HCoV-229E and HCoV-OC43, known as human pathogens[120, 121]. Before 2003, there were only two CoVs, The SARS pandemic in 2003 led to the revelation of SARS-CoV-1, a pathogen causing a severe respiratory infection[122]. The subsequent surge in CoV research resulted in the discovery of two additional human CoVs, HCoV-NL63 and HCoV-HKU1, that are mildly pathogenic[123, 124]. One addition to this group was MERS-CoV that emerged in 2012 as a pathogen causing a severe respiratory infection[125]. Although SARS-CoV-1 and MERS-CoV are highly lethal pathogens, the public health, social, and economic damages that they have caused are diminutive in comparison to that from SARS-CoV-2, a newly emerged human CoV pathogen that causes COVID-19[126]. Rival only to the 1918 influenza pandemic, the COVID-19 pandemic has led to catastrophic impacts worldwide.

To alleviate catastrophic damages of COVID-19 on public health, society and economy, finding timely treatment options is of paramount importance.

Similar to all other CoVs, SARS-CoV-2 is an enveloped, positive-sensed RNA virus with a genome of nearly 30 kb in size[127]. Its genome encodes 10 open reading frames (ORFs). The largest ORF, ORF1ab encompasses more than two thirds of the whole genome. Its translated products, ORF1a (~500 kDa) and ORF1ab (~800 kDa)[128], are very large polypeptides that undergo proteolytic cleavage to form 15 mature proteins. These are nonstructural proteins (Nsps) that are essential for the virus to modulate human cell hosts for efficient viral protein expression, viral genome replication, virion packaging, and viral genomic RNA processing. The proteolytic cleavage of ORF1a and ORF1ab is an autocatalytic process. Two internal polypeptide regions, Nsp3 and Nsp5, possess cysteine protease activity that cleaves themselves, and all other Nsps, from the two polypeptides. Nsp3 is commonly referred to as papain-like protease (PLpro and Nsp5 as 3C-like protease (3CLPro protease (Mpro)[129]. Although we have yet to understand SARS-CoV-2 biology and COVID-19 pathogenesis, previous studies of SARS-CoV-1 have established that activity of both PLpro and Mpro is essential to viral replication and pathogenesis. Of the two proteases, Mpro processes 12 out of the total 15 Nsps; inhibition of this enzyme is anticipated to have more significant impacts on the viral biology than that of PLpro. Therefore, small molecule medicines that potently inhibit SARS-CoV-2 Mpro are potentially effective treatment options for COVID-19[67, 81, 130]. In this work we report our progress in the development of potent SC2Mpro inhibitors.
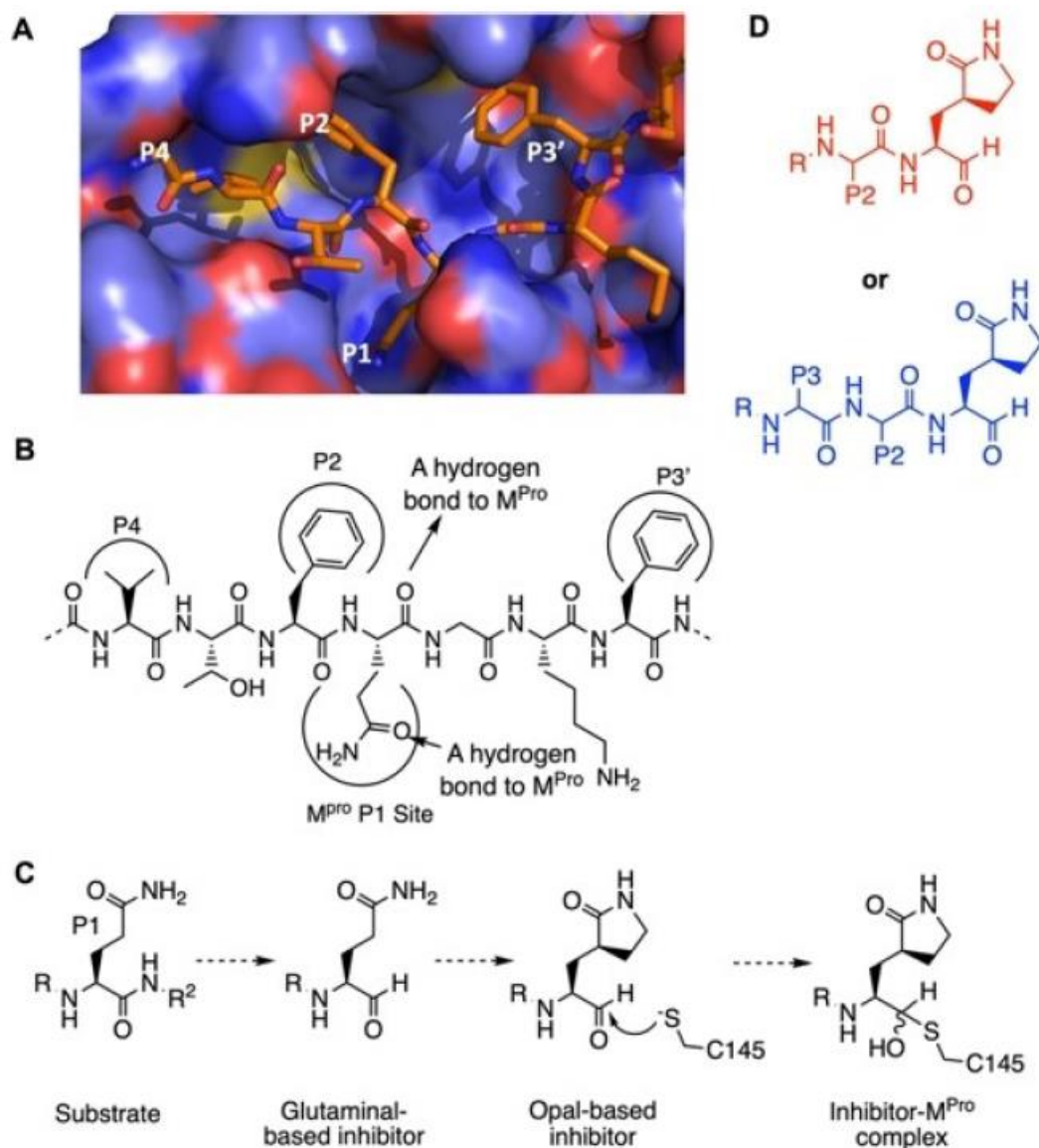
Figure 23. The design of SC2Mpro inhibitors based on medicinal chemistry learned from SC1Mpro studies. (A) The structure of SC1Mpro complexed with a peptide substrate (based on PDB ID: 5B6O). Active-site cavities that bind P1, P2, P4, and P3' residues in the substrate are labeled. **(B)** Schematic diagram that shows interactions between SC1Mpro and a substrate. **(C)** A scheme in which a substrate P1 residue is converted to glutaminal and then β-(S-2-oxopyrrolidin-3-yl)-alaninal (Opal) to form a reversible covalent inhibitor that reacts with the SC1Mpro active-site cysteine C145. **(D)** Scaffold structures of Opal-based inhibitors designed for SC2Mpro.

3.2 Methods and Results

94

3.2.1 The Design of β-(S-2-oxopyrrolidin-3-yl)-alaninal (Opal)-Based, Reversible

Covalent Inhibitors for SC2Mpro.

Although we are at the inaugural stage of learning medicinal chemistry to inhibit SC2Mpro, much has been learned from studies of SARS-CoV-1 Mpro (SC1Mpro) that shares 96% sequence identity with SC2Mpro[74]. SC1Mpro has a large active site that consists of several smaller pockets for the recognition of residues at P1, P2, P4, and P3′ positions in a protein substrate (**Figure 23A**)[131]. P4 is typically a small hydrophobic residue, whereas P2 and P3' are large. For all Nsps that are processed by SC1Mpro and SC2Mpro, Gln is the P1 residue at their cleavage sites. In order to bind the P1 Gln, SC1Mpro forms strong van der Waals interactions with the Gln side chain, and also uses two hydrogen bonds with the Gln side chain amide oxygen and α-carbonyl oxygen atoms (**Figure 23B**). Previous efforts in the development of irreversible covalent inhibitors for SC1Mpro primarily focused on fixing the P1 residue as a more potent β-S-2-oxopyrrolidine-containing Gln analog and changing the scissile backbone amide to an alkene Michael acceptor in order to react with the active-site cysteine C145, as well as varying substituents on two sides to improve potency[132]. The enhanced potency from the use of the β-S-2-oxopyrrolidine-containing Gln analog is most probably due to the reduction of entropy loss during the binding of SC1Mpro to the more rigid lactam compared to the flexible Gln. Although converting the scissile backbone amide to a Michael acceptor in a SC1Mpro ligand turns it into a covalent inhibitor, it eliminates the critical hydrogen bond between the P1 α-carbonyl oxygen and SC1Mpro. Therefore, most

95

Michael acceptor inhibitors developed for SC1Mpro and recently for SC2Mpro tend to have efficacy with low micromolar or sub-micromolar IC_{50} values rather than low nanomolar levels[132]. To maintain this critical hydrogen bond and exploit a covalent interaction with C145 to form a hemiacetal for high affinity, both aldehyde and ketoamide moieties have been used to replace the P1C-side α-amide to develop potent reversible covalent inhibitors for SC1Mpro. For aldehyde-based inhibitors, a typical potent inhibitor contains Opal at the P1 site that consists of a β-S-2-oxopyrrolidine side chain and an α-aldehyde for both taking advantage of strong interactions with the SC1Mpro P1-binding pocket and the formation of a reversible covalent bond with C145 (**Figure 23C**).

Typical examples of this design include GC376, which was originally developed for Mpro from feline infectious peritonitis (FIP) CoV and two inhibitors, 11a and 11b, which were recently developed for SC2Mpro[75, 133]. Given its relative simplicity, we have followed a similar scheme according to structure diagrams shown in **Figure 23D** to design and synthesize reversible covalent inhibitors for SC2Mpro and pursued structural variations at P2, P3, and R positions for improved potency.

3.2.2 Synthesis and IC₅₀ Characterization of SC2Mpro Inhibitors (MPIs)

3.2.2 Synthesis and IC$_{50}$ Characterization of SC2Mpro Inhibitors (MPIs)

GC376 (**Figure 24A**) has confirmed potency against SC1Mpro[134]. We purchased it as a potential SC2Mpro inhibitor. We designed two similar dipeptidyl compounds MPI1-2 (**Figure 24A**) and synthesized them according to a synthetic scheme shown in Scheme S1 in the Supporting Information. Both MPI1 and MPI2 have Phe at the P2 site which was previously shown to contribute to strong bonding to SC1Mpro[132]. MPI2 has

also an *o*-fluoro-*p*-chlorocinnamyl group as an N-terminal cap. This group is more rigid than the CBZ group and therefore possibly introduces a strong interaction with the P4-binding pocket in SC2Mpro[135]. To characterize IC_{50} values of all three molecules for inhibition of SC2Mpro, we recombinantly expressed a His-6-SUMO-SC2Mpro fusion protein in Escherichia coli and purified and digested this protein with SUMO protease to obtain intact SC2Mpro with more than 95% purity. We used a previously described fluorescent peptide assay to measure the IC_{50} values for GC376, MPI1, and MPI2 as 31 ± 4, 100 ± 23, and 103 ± 14 nM, respectively (**Figure 24B**)[130]. Our determined IC_{50} value for GC376 agrees well with that from Ma et al[136]. In the light of the publication of inhibitors 11a and 11b that showed similar IC_{50} values as 53 ± 5 and 40 ± 2 nM, respectively[75], we shifted our focus from the synthesis of bipeptidyl inhibitors to that of tripeptidyl inhibitors. By adding one more residue to the design of inhibitors, additional interactions with SC2Mpro might be achieved to improve potency. In the design of SC1Mpro inhibitors, Leu, Phe, and Cha (cyclohexylalanine) are three residues used frequently at the P2 site and Val and Thr(tBu) (O-tert-butylthreonine) are two residues used frequently at the P3 site[132]. Installation of these residues at two sites and including CBZ as a N-terminal cap led to the design of six compounds MPI3-8 (**Figure 24B**). We added one additional compound MPI9 that has an *o*-fluoro-*p*-chlorocinnamyl cap to this series to compare the effect of the two N-terminal caps on the inhibitor potency for SC2Mpro. We synthesized all seven compounds according to a synthetic scheme presented in Scheme S2 and characterized their IC_{50} values using the fluorescent peptide assay. As shown in **Figure 24B**, all inhibitors have IC_{50} values below 100 nM, except for

97

MPI8 that has an IC_{50} value as 105 ± 22 nM. The most potent compound is MPI3 with an IC_{50} value as 8.5 ± 1.5 nM, followed by MPI4 and MPI5 with IC_{50} values as 15 ± 5 and 33 ± 2 nM, respectively. We also synthesized 11a (named as MPI10 in our series) according to the procedure in Dai et al. and used it as a positive control in our enzyme and viral inhibition analyses[75]. Using our fluorescent peptide assay, we determined the IC_{50} value of 11a as 31 ± 3 nM. As far as we know, MPI3 is the most potent SC2Mpro inhibitor that has been reported so far. From the perspective of enzyme inhibition, Leu and Val are optimal residues at P2 and P3 sites in an inhibitor for improved affinity for SC2Mpro and CBZ also enhances affinity compared to the *o*-fluoro-*p*-chlorocinnamyl as a N-terminal capping group.
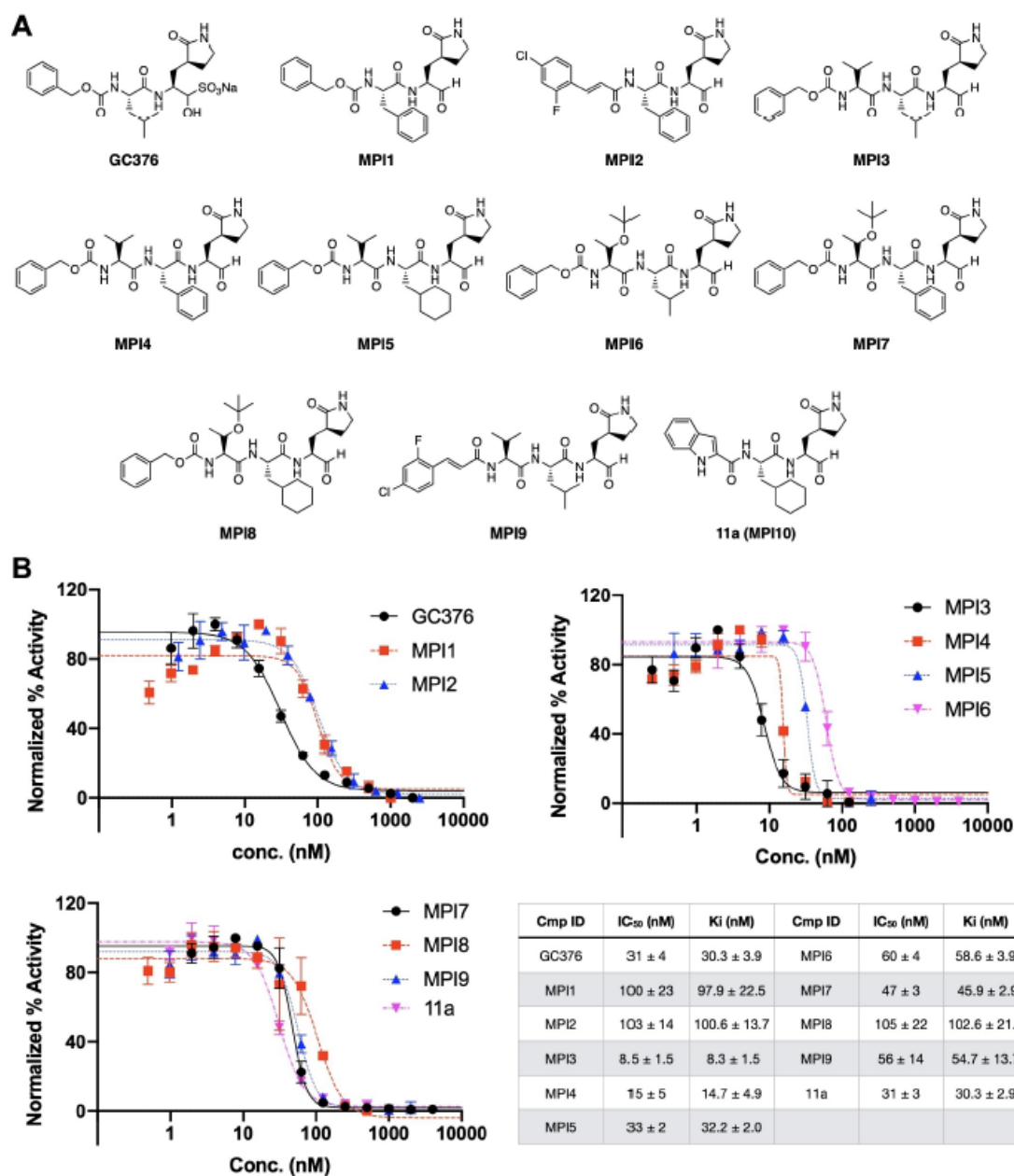
Figure 24. SC2Mpro inhibitors and their IC$_{50}$ values. (A) Structures of GC376 and ten Opal-based inhibitors. **(B)** The inhibition curves of all 11 inhibitors toward SC2Mpro. Triplicate experiments were performed for each compound. The determined IC$_{50}$ values and Ki values are presented as mean $\pm$ standard error (SE) in the associated table.

3.2.3 Structural Characterization of SC2Mpro Interactions with Opal-based Inhibitors

In order to understand how our designed inhibitors interact with SC2Mpro at its active site, we screened crystallization conditions for apo-SC2Mpro, soaked apo-SC2Mpro crystals with different inhibitors, and determined the crystal structures of these inhibitors in complex with SC2Mpro. We used Hampton Research Crystal Screen and Index kits to perform initial screening and identified several conditions that yielded single crystals of apo-SC2Mpro. For all conditions, crystals were in a thinplate shape. The best crystallization condition contained 0.2 M dibasic ammonium phosphate and 17% PEG 3,350. We refined the structure of apo-SC2Mpro against diffraction data to 1.6 Å resolution (PDB ID: 7JPY). In the apoenzyme crystals, SC2Mpro existed as a monomer in the crystallographic asymmetric unit and packed relatively densely. The active site of each monomer stacked upon another monomer (two representative monomers are shown in red and blue respectively in **Figure 25A**). This close contact and dense protein packing made the diffusion of inhibitors to the active site quite slow. We soaked apo SC2Mpro crystals with all nine inhibitors that we synthesized and collected and processed their X-ray diffraction data for structural determination. For crystals that we soaked with the inhibitors for just 2 h, we did not find observable ligand electron density at the enzyme active site.
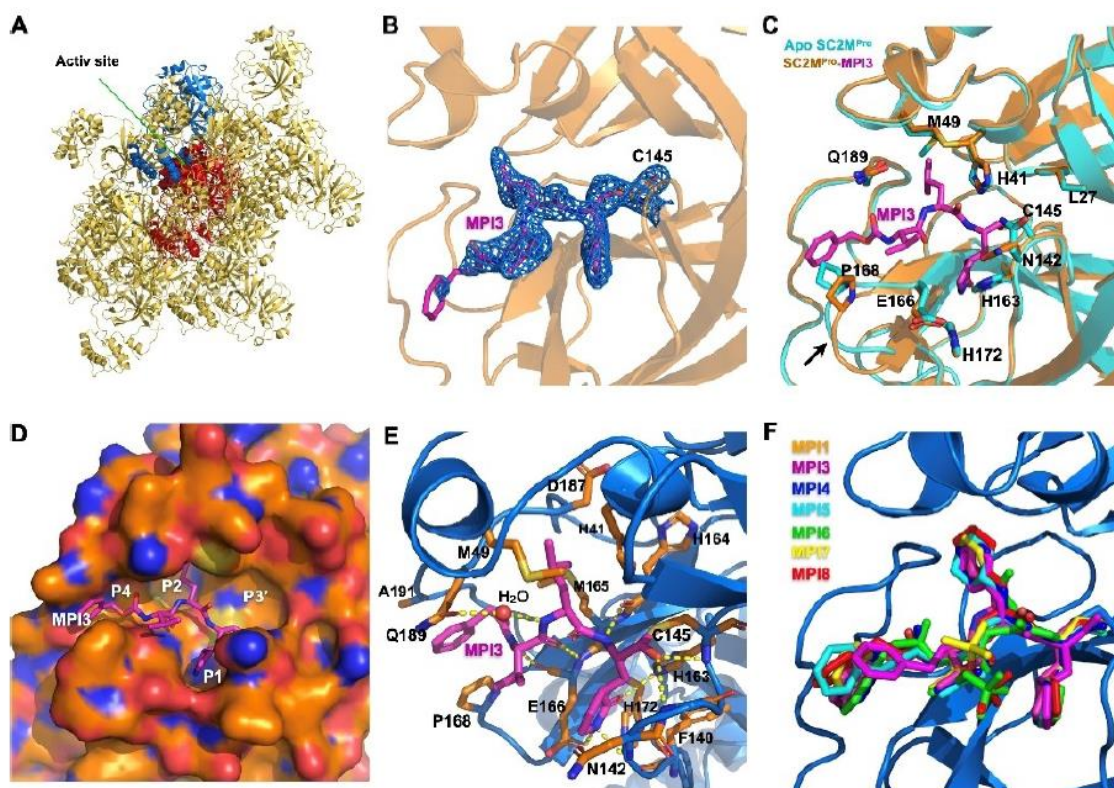
Figure 25. X-ray crystallography analysis of SC2Mpro in its apo-form and complexes with different inhibitors. (A) The packing of apo-SC2Mpro in its crystals. An asymmetric unit monomer in the center is colored in red. Its active site is presented as a concave surface. Another monomer that stacks on the active site of the monomer is colored in blue. **(B)** A contoured 2Fo-Fc map at the 1 σ level around MPI3 and C145 in the active site of SC2Mpro. A covalent bond between MPI3 and C145 is observable. **(C)** Structure overlay between apo-SC2Mpro and the SC2Mpro-MPI3 complex. A black arrow points to a region that undergoes structure rearrangement from the apoenzyme to accommodate MPI3 in the SC2Mpro-MPI3 complex. **(D)** Occupation of the active-site cavity of SC2Mpro by MPI3. The enzyme is shown in its surface presentation mode. **(E)** Extensive hydrogen bonding and van del Waals interactions between SC2Mpro and MPI3. The backbone of SC2Mpro is colored in blue and side chain carbon atoms in orange. Hydrogen bonds between MPI3 and SC2Mpro are depicted as yellow dashed lines. **(F)** Overlay of seven Opal-based inhibitors at the active site of SC2Mpro. Color coded for the names is shown in the figure. All images were made by using the program PyMOL. The PDB IDs for SC2Mpro in its apo form and complexes with inhibitors are 7JPY (apo), 7JPZ (MPI1), 7JQ0 (MPI3), 7JQ1 (MPI4), 7JQ2 (MPI5), 7JQ3 (MPI6), 7JQ4 (MPI7), and 7JQ5 (MPI8).

For seven inhibitors including MPI1 and MPI3-8, we performed two-day soaking

and observed clear electron density in the difference maps in the active site of the enzyme.

For MPI2 and MPI9, we were not able to determine structures of their complexes with

101

SC2Mpro due to cracking of the crystals upon soaking with the inhibitors. For MPI3, the electron density around the P1, P2, and P3 residues were well defined, and the covalent interaction between the C145 side-chain thiolate and the Opal aldehyde to form a hemiacetal was clearly observable (**Figure 25B**; PDB ID: 7JQ0). The electron density around CBZ was very weak indicating flexible CBZ binding around the enzyme P4-binding pocket. **Figure 25C** shows the superposition of apo-SC2Mpro and the SC2Mpro-MPI3 complex structures. The two structures display very little overall variation with RMSD as 0.2 Å. Around the active site in the two structures, large structural rearrangements exist for residues M49 and N142 and the loop region that contains P168. In apoenzyme, the side chain of M49 folds into the P2-binding pocket. It flips toward the solvent to make space available for the binding of the P2 Leu in MPI3. The side chain of N142 rotates by almost 180° between the two structures and adopts a conformation in the SC2Mpro-MPI3 complex that closely caps the P1-binding site for strong van der Waals interactions with the Opal residue in MPI3. In the SC2Mpro-MPI3 complex, the P168-containing loop is pushed away from its original position in the apoenzyme, probably by interaction with the CBZ group, which triggers a position shift for the whole loop. Except for M49, N142, and the P168-containing loop, structural orientations of all other residues at the active site closely resemble each other in the two structures. In the active site, MPI3 occupies the P1, P2, and P4-binding pockets and leaves the large P3'-binding pocket empty (**Figure 25D**). Extensive hydrogen bonding and van der Waals interactions in addition to the covalent interaction with C145 contribute to the strong binding of MPI3 to SC2Mpro (**Figure 25E**). Residues F140, N142, H163, E166, and H172 form a small cage

to accommodate the Opal side chain. Three hydrogen bonds form between the Opal lactam amide and the E166 side chain carboxylate, H163 imidazole, and F140 backbone carbonyl oxygen. The precise fitting of Opal into the P1-binding pocket and the formation of three hydrogen bonds explain the preferential binding of the Opal side chain to this pocket. In the SC2Mpro-MPI3 complex, M49 flips from the P2-binding pocket to leave space for the binding of the P2 Leu in MPI3. Residues H41, M49, M165 and D187, backbones of the M165-containing strand, and the D187-containing loop form a hydrophobic pocket that is in a close range of van der Waals interactions with the P2 Leu in MPI3. We observe Leu as the best residue in this position probably due to this close van der Waals interaction range for the recognition of the P2 Leu side chain. The enzyme has no P3-binding pocket. However, the P3 Val in MPI3 positions its side chain in van der Waals interaction distance to E166 and P168. In the structure, CBZ narrowly fits into the P4-binding pocket and the channel formed between the P168- and Q192-containing loops. The P168 loop rearranges its position from that in apoenzyme to accommodate the CBZ group. The CBZ group also has weak electron density. These observations indicate that CBZ is not an optimal structural moiety for interaction at these sites. Besides interactions involving side chains and the CBZ group in MPI3, its two backbone amides and carbamate form 6 hydrogen bonds with the enzyme. Two of them are formed between the P3 Val in MPI3 and the backbone amino and carbonyl groups of E166 in SC2Mpro. One water molecule mediates a hydrogen bond bridge between the P2 Leu amino group in MPI3 and the Q189 side chain amide in SC2Mpro. For the P1 Opal residue in MPI3, its α-amino group forms a hydrogen bond with the H164 α-carbonyl oxygen in the enzyme. The original aldehyde oxygen in

MPI3 forms two hydrogen bonding interactions, one with the α-amino group of G143 and the other the C145 α-amine in SC2Mpro. The two hydrogen bonds are probably the reason that Opal-based reversible covalent inhibitors are typically stronger than Michael acceptor inhibitors, in which the original scissile amide is replaced with an alkene, for inhibition of Mpro enzymes. In the structures of SC2Mpro complexes with the other 6 inhibitors, we observed similar structure rearrangements at M49, N142, and the P168-containing loop to accommodate inhibitors and a covalent interaction (PDB IDs: 7JPZ, 7JQ1, 7JQ2, 7JQ3, 7JQ4, 7JQ5).

3.2.4 SARS-CoV-2 Inhibition Analysis of GC376, MPI1-8, and 11a

To evaluate our molecules' ability to inhibit SARS-CoV-2, we conducted a live virus-based microneutralization assay in Vero E6 cells. Vero E6 is a kidney epithelial cell line isolated from African Green Monkey. It has been used widely as a model system for human CoV studies[137]. We tested 10 molecules including GC376, MPI1-8, and 11a in a concentration range from 80 nM to 10 μM and recorded cytopathogenic effect (CPE) observed in SARS-CoV-2-infected Vero E6 cells that we cultured in the presence of different concentrations of inhibitors. 11a was included as a positive control. For each condition, we conducted two repeats. Although it was disappointing that MPI3 was not able to completely prevent CPE at all tested concentrations, several inhibitors abolished CPE: GC376, MPI2, MPI6, and 11a at 10 μM, MPI5 at 5 μM, MPI7 at 2.5-5 μM, and MPI8 at 2.5 μM (**Figure 26A**). Three compounds MPI5, MPI7 and MPI8 performed better than GC376 that has been recently explored by Anivive Lifesciences for the treatment of

COVID-19 and 11a that has been considered for COVID-19 clinical studies[130]. As we only recorded complete abolition of CPE, the real EC_{50} values for these compounds are expected to be much lower than lowest observed concentrations for CPE abolishment.
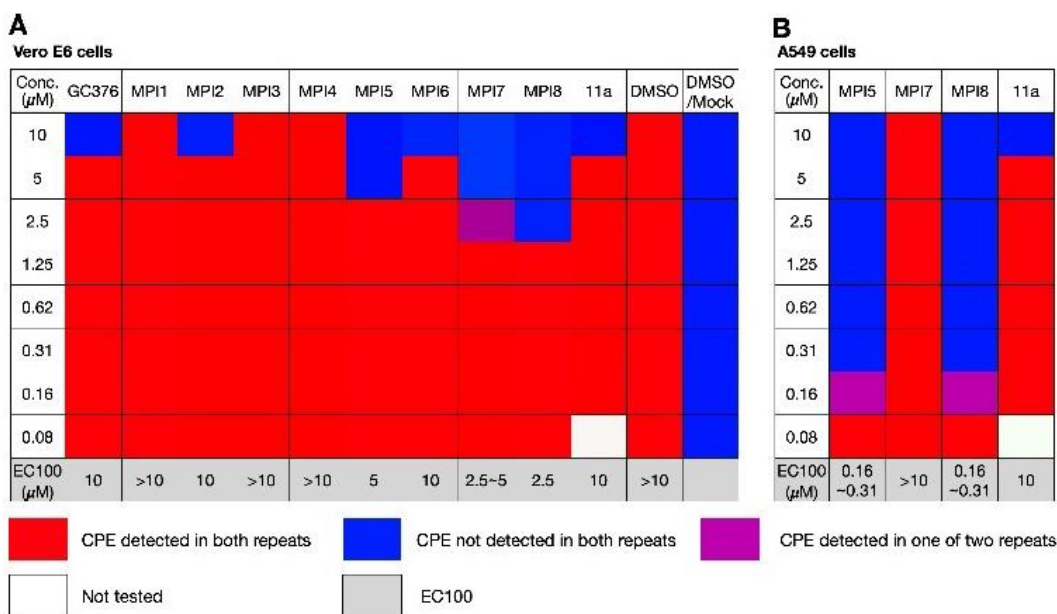
Figure 26. The SARS-CoV-2 viral inhibition results of selected inhibitors in (**A**) Vero E6 and (**B**) A549/ACE2 cells. CPE: cyto-pathogenic effect.

Encouraged by our results in Vero E6 cells, we tested the three most potent compounds MPI5, MPI7, and MPI8 and also 11a in A549/ACE2 cells. The A549/ACE2 cell line was derived from human alveolar epithelial cells. It mimics the SARS-CoV-2 infection of the human respiratory tract system better than Vero E6[138]. We tested a same concentration range for all four compounds. MPI7 was not able to completely abolish CPE at all tested conditions. However, both MPI5 and MPI8 performed much better than in Vero E6 cells with complete abolition of CPE at 160-310 nM and much better than 11a (**Figure 26B**). 11a displayed potency similar to that shown in Vero E6 cells. Given that real EC_{50} values are expected to be lower than the lowest observed concentration for CPE

105

abolishment, MPI5 and MPI8 are, as far as we know, the most potent anti-SARS-CoV-2 small molecules in infected cells that have been reported so far.

3.3 Discussion and Conclusion

Guided by previous medicinal chemistry studies about SC1Mpro, we designed and synthesized a number of Opal-based dipeptidyl and tripeptidyl inhibitors that potently inhibit SC2Mpro, an essential enzyme for SARS-CoV-2, the pathogen of COVID-19. As the most potent inhibitor of SC2Mpro, MPI3 displayed an IC_{50} value of 8.5 nM. As far as we know, this is the lowest reported IC_{50} for known SC2Mpro inhibitors. During the search of optimal conditions for IC_{50} characterizations, we noticed that 10 nM was the lowest SC2Mpro concentration that could provide reliable activity[130]. To characterize the Ki value of an inhibitor, the Km value for a used substrate needs to be determined. The substrate we used for the inhibitor characterization has a sequence as Dabcyl-KTSAVLQSGFRKME-Edans. At a fixed enzyme concentration at 20 nM, the substrate cleavage rate was roughly proportional to the added substrate concentration up to 200 μM. Above 200 μM, the substrate had a significant quenching effect and was also not well soluble. Based on our results, the Km value is about 422.4 μM. Based on this determined Km value, Ki for all the inhibitors are calculated and presented in **Figure 24B**. X-ray crystallography analysis of the SC2Mpro MPI3 complex revealed that MPI3 fits precisely in the P1- and P2-binding pockets at the SC2Mpro active site (PDB ID: 7JQ0). Strong van der Waals interactions at the P1- and P2-binding pockets, nine hydrogen bonds with

106

active-site residues, and the covalent interaction with C145 necessitate high affinity of MPI3 to SC2Mpro. The N-terminal capping group of MPI3 and other inhibitors are not well defined in the crystal structures, indicating an unfitting size for this group or relatively loosely bound pattern in P4-binding pocket. Optimization on size or ligand-protein interacting to introduce stronger interaction between ligand and SC2Mpro at this site would contribute to the generation of more potent inhibitors in the future. Although MPI3 is the most potent inhibitor for the enzyme, its cellular activity in inhibiting SARS-CoV-2 is much lower than several other inhibitors we have generated. A likely reason is its lower cellular stability. MPI3 has Leu and Val at its P2 and P3 sites respectively. Both are naturally occurring amino acids that are expected to be targeted by both extracellular and cellular proteases. Since Leu and Val are optimal residues at two sites, modest changes based on these structures will be necessary for both maintaining high potency in inhibiting SC2Mpro and improving cellular stability for enhanced cellular activity in inhibiting the virus. As such, Val and Leu analogs at these two sites need to be explored. Since both MPI5 and MPI8 show high anti-SARS-CoV-2 activity in both Vero E6 and ACE2+ A549 cells and each has Cha at their P2 site, we suggest maintaining Opal and Cha at P1 and P2 sites and varying the residue at P3 and the N-terminal capping moiety to improve anti-SARS-CoV-2 activity in cells. Based on our structures of SC2Mpro complexes with seven inhibitors, the P1 Opal occupies precisely the P1-binding pocket in SC2Mpro and three hydrogen bonds to the Opal lactam amide are critical in maintaining strong binding to SC2Mpro. Chemical space to manipulate the P1 residue in an inhibitor for improved binding to SC2Mpro is minimal. But one direction that may be explored is to introduce

107

additional heteroatom(s) to Opal for the formation of hydrogen bond(s) with the N142 side chain am-ide. In the SC2Mpro-MPI3 complex, the N142 side chain flips by about 180° from its position in apoenzyme to form a closed P1-binding pocket. However, only van der Waals inter-actions with Opal are involved with N142. Given the close distance between the Opal side chain and the side chain amide of N142, some hydrogen bonds may be designed for improved potency. In all our designed inhibitors, an Opal aldehyde is involved in the formation of a covalent interaction with C145. This design, although necessary for the formation of a hemiacetal covalent complex, effectively excludes the exploration of the P3'-binding site in SC2Mpro for improved potency in a designed inhibitor. **Figure 25D** illustrates that the P3'-binding pocket is completely empty. In our early discussion, we argued that it is critical to maintain the hydrogen bond between the scissile amide oxygen in a substrate and SC2Mpro for high affinity. Changing the scissile amide to an aldehyde in an inhibitor is effective in maintaining this hydrogen bond and allows a covalent interaction with C145. Two hydrogen bonds formed between the hemiacetal alcohol and SC2Mpro contribute to high potency of this group of molecules.

In our study, cell-based anti-SARS-CoV-2 activity of our designed inhibitors do not correlate with their IC_{50} values in inhibiting SC2Mpro. This is expected since cellular stability and other features of these inhibitors are very different. However, information regarding both enzyme inhibition IC_{50} values and anti-SARS-CoV-2 activity is critical for the design of a new generation of inhibitors that perform excellent in both aspects. Given that MPI3 has already reached a single-digit nanomolar IC_{50} value and MPI5 and MPI8 display high potency in inhibiting SARS-CoV-2, merging features of the three molecules

will lead to inhibitors with extreme potency in inhibiting the virus. Our antiviral assays indicated that MPI5 and MPI8 performed much better than GC376 and 11a, two molecules that have been explored for COVID-19 preclinical and clinical tests. These two molecules are ready for preclinical analysis that we are actively exploring. We noticed in our antiviral assays that MPI5 and MPI8 have much higher potency in A549/ACE2 cells than in Vero E6 cells. These two cell lines have different host protease proteomes. It is likely that MPI5 and MPI8 inhibit some host proteases that serve critical functions in the SARS-CoV-2 entry into and replication in host cells and therefore exert different SARS-CoV-2 inhibition.

3.4 Experimental Details

3.4.1 Recombinant SC2Mpro Protein Expression and Purification

The construct pET28a-His-SUMO-SC2Mpro construct was made based on the a pET28a plasmid modified with N-terminal His-SUMO tag. The gene encoding SC2Mpro was amplified from a previous plasmid pBAD-sfGFP-Mpro using the forward primer 5'-CGCGGATCCGGGTTTCGCAAG-3' and the reverse primer 5'-CCGCTCGAGTTACTGAAAAGTTACGCC-3'. The amplified PCR product was digested with BamHI and XhoI and ligated into the vector pET28a-His-SUMO plasmid digested with the same restriction enzymes. The gene sequence of His-SUMO-SC2Mpro was verified by sequencing at Eton Bioscience Inc.

The pET28a-His-SUMO-SC2Mpro construct was transformed into *E. coli* strain BL21(DE3). Cells were cultured at 37° C in 6 L 2xYT medium with kanamycin (50 μg/mL) for 3 h and induced with isopropyl-D-1-thiogalactoside (IPTG) at final concentration of 1 mM when the OD_{600} reached 0.8. After 3 h, cells were harvested by centrifugation at 12,000 rpm, 4 °C for 30 min. Cell pellets were resuspended in 150 mL buffer A (20 mM Tris, 100 mM NaCl, 10 mM imidazole, pH 8.0) and then lysed by sonication on ice. The lysate was clarified by centrifugation at 16,000 rpm, 4 °C for 30 min. The supernatant was loaded onto a nickel-chelating column with High Affinity Ni-Charged Resin (GenScript) and washed with 10 column volumes of buffer A to remove unspecific binding proteins, followed by elution using buffer B (20 mM Tris, 100 mM NaCl, 250 mM imidazole, pH 8.0). The protein eluates were subjected to buffer exchange with buffer C (20 mM Tris, 10 mM NaCl, 1mM dithiothreitol (DTT), pH 8.0) by using HiPrep 26/10 desalting column (GE Healthcare). The His-SUMO-SC2Mpro proteins were digested with SUMO protease overnight at 4 °C. The digested protein was applied to nickel-chelating column again to remove the His-tagged SUMO protease, the His-SUMO tag, and protein with uncleaved His-SUMO tag. The tag-free SC2Mpro protein was loaded onto an anion-exchange column with Q Sepharose, Fast Flow (GE Healthcare) equilibrated with buffer C for further purification. The column was eluted by buffer D (20 mM Tris, 1 M NaCl, 1 mM DTT, pH 8.0) with a linear gradient ranging from 0 to 500 mM NaCl (10 column volumes buffer). Fractions eluted from the anion-exchange column were condensed and loaded to size exclusion column with HiPrep 16/60 Sephacryl S-100 HR (GE Healthcare) pre-equilibrated with buffer E (20 mM Tris, 100 mM NaCl, 1 mM

DTT, 1 mM EDTA, pH 7.8). The eluted SC2Mpro protein in buffer E was concentrated to 20 mg/mL and stored in -80 °C for further use.

3.4.2 The Determination of Km for Sub3

The assays were carried out with 20 nM enzyme and various concentration of Sub3, a fluorogenic substrate we purchased from BaChem (DABCYL-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-Met-Glu-EDANS) at 37 °C. Aliquot was taken out at indicated time points and diluted 10 times to stop the reaction. Fluorescent intensity was recorded immediately. Data treatment were done with Graph Pad Prism 8.0 software. First 14 min were analyzed by linear regression for initial reaction rate analyses. The initial reaction rates were used to determine the K_m value by fitting with Michaelis-Menten non-linear regression.

3.4.3 IC$_{50}$ Analysis

The assays were carried out with 20 nM enzyme (except for MPI3, for which 10 nM enzyme was used) and 10 μM substrate at 37 C with continuous shaking. All the analyses were carried out in triplicate. The substrate (DABCYL-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-Met-Glu-EDANS) was purchased from Bachem and stored as 1 mM solution in 100% DMSO. Enzyme activity was monitored by fluorescence with excitation at 336 nm and emission at 455 nm wavelength. The dilution buffer (used for enzyme and substrate dilution) is 10 mM $Na_xH_yPO_4$, 10mM NaCl, 0.5 mM EDTA, pH 7.6. Final composition of the assay buffer is 10 mM $Na_xH_yPO_4$, 10 mM NaCl, 0.5 mM

EDTA, 2 μM DTT (coming from enzyme stock solution), pH 7.6 with 1.25% DMSO. All the inhibitors were stored as 10 mM in 100% DMSO solutions in -20 °C freezer.

IC$_{50}$ analysis, the inhibitor was diluted to 400-fold times higher than the highest working concentration to make the secondary stock solution (i.e. if the highest working concentration of inhibitor is 2 μM, then the inhibitor was diluted from its 10 mM stock solution to 800 μM in DMSO). 10 μL from this secondary stock solution was added to the 990 uL of dilution buffer. Serial dilutions were carried out in dilution buffer containing 1% DMSO to ensure all the inhibitor serial dilutions are at 1% DMSO. 25 μL of each inhibitor solution were added to 96-well plate with multichannel pipettor. Next, 25 μL of 80 nM enzyme solution (diluted from 10 μM enzyme storage solution in 10 mM Na$_x$H$_y$PO$_4$,10 mM NaCl, 0.5 mM EDTA, pH 7.6, 1 mM DTT with dilution buffer) were added by multichannel pipettor and mixed by pipetting up and down three times. Then, the enzyme-inhibitor solution was incubated at 37 °C for 30 minutes. During incubation period, 20 μM of the substrate solution is prepared by diluting from 1 mM stock solution with dilution buffer. When the incubation period is over, 50 μL of the 20 μM substrate solution added to each well by multichannel pipettor and the assay started. Data recording were stopped after 30 minutes. Data treatment were done with Graph Pad Prism 8.0 software. First 0-300 seconds were analyzed by linear regression for initial slope analyses. Then, the initial slopes were normalized and IC$_{50}$ values were determined by inhibitor vs response - variable slope (four parameters).

3.4.4 Crystallization of SC2Mpro

A freshly prepared SC2Mpro protein solution at a concentration of 10 mg/mL was cleared by centrifugation at 14,000 rpm, 10 min. Next, a basic screen with the commercially available screening kits (Hampton Research IndexTM, Crystal ScreenTM 1 and 2, PEGRxTM 1 and 2, PEG/IonTM 1 and 2) were performed employing the sitting-drop vapor-diffusion method at 18 °C. 1.0 µL of SC2Mpro protein solution and 1.0 µL of reservoir buffer were mixed to equilibrate against 100 µL reservoir solution. Crystals appeared overnight under over 50 conditions. The most promising crystal was found under condition No.44 of PEG/IonTM (0.2 M Ammonium phosphate dibasic, 20% w/v PEG3350, pH 8.0). Subsequent optimization was performed by adjusting the temperature and concentration of protein and precipitant. The best plate-like crystals were obtained at 25 °C from 0.2 M Ammonium phosphate dibasic, 17% w/v PEG3350, pH 8.0, with a SC2Mpro protein concentration of 14 mg/ml. Overnight growing crystals were washed with cryo-protectant containing mother liquor plus gradually increasing glycerol (5%, 10%, 15%, 20%, 25% and 30%). Cryo-protected crystals were fished for data collection.

3.4.5 Crystallization of SC2Mpro in Complex with Inhibitors

Soaking was performed to produce SC2Mpro-inhibitor complex crystals. Overnight growing SC2Mpro crystals were washed with reservoir solution three times in situ. Subsequently, the crystals were washed three times with reservoir solution plus 0.5 mM inhibitor and 2% DMSO (Inhibitors were dissolved to 25 mM in 100% DMSO). The mixture was incubated at 25 °C for 48 h. The cryoprotectant solution contained mother

liquor plus 30% glycerol, 0.5 mM inhibitor and 2% DMSO. Cryo-protected crystals were fished for data collection.

3.4.6 Data collection and Structure Determination

The data of SC2Mpro with MPI6 and MPI8 were collected on a Rigaku R-AXIS IV++ image plate detector. All the other data were collected at the Advanced Light Source (ALS) beamline 5.0.2 using a Pilatus3 6M detector. The diffraction data were indexed, integrated and scaled with iMosflm.1 All crystals are in space group C121. All the structures were determined by molecular replacement using the structure model of the free enzyme of the SARS-CoV-2 (2019-nCoV) main protease [Protein Data Bank (PDB) ID code 6Y2E] as the search model using Phaser in the Phenix package.2-3 *JLigand* and *Sketcher* from the CCP4 suite were employed for the generation of PDB and geometric restraints for the inhibitors. The inhibitors were built into the Fo-Fc density by using *Coot*.4 Refinement of all the structures was performed with Real-space Refinement in Phenix.3 Details of data quality and structure refinement are summarized in Table S1. All structural figures were generated with PyMOL (https://www.pymol.org).

3.4.7 SARS-CoV-2 Inhibition by a Cell-based Assay

A slightly modified cytopathic effect (CPE)-based microneutralization assay was used to evaluate the drug efficacy against SARS-CoV-2 infection. Briefly, confluent African green monkey kidney cells (Vero E6) or human alveolar epithelial A549 cells stably expressing human ACE2 viral receptor, designated A549/hACE2, grown in 96-

wells microtiter plates were pretreated with serially 2-folds diluted individual drugs for two hours before infection with 100 or 500 infectious SARS-CoV-2 (US_WA-1 isolate) particles in 100 μL EMEM supplemented with 2% FBS, respectively. Cells pre-treated with parallelly diluted DMSO with or without virus were included as positive and negative controls, respectively. After cultivation at 37 °C for 3 (Vero E6) or 4 days (A549/hACE2), individual wells were observed under the microcopy for the status of virus-induced formation of CPE. The efficacy of individual drugs was calculated and expressed as the lowest concentration capable of completely preventing virus-induced CPE in 100% (EC_{100}) or 50% (EC_{50}) of the wells. All compounds were dissolved in 100% DMSO as 10 mM stock solutions before subjecting to dilutions with culture media.

3.4.8 The synthesis of inhibitors MPI1-9

Scheme 26. Synthesis of compound 89.
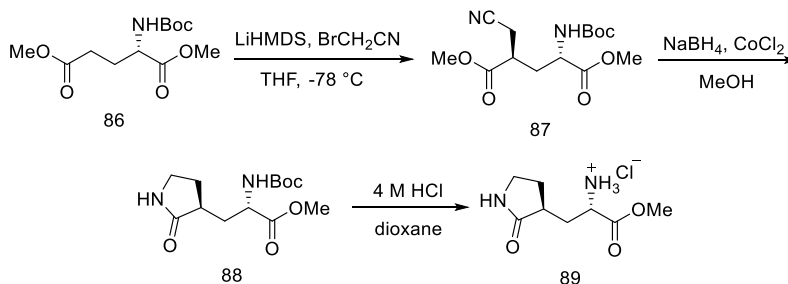
Dimethyl (2S,4R)-2-((tert-butoxycarbonyl)amino)-4-(cyanomethyl)pentanedioate (87)

A solution of *N*-Boc-glutamic acid dimethyl ester (3 g, 11 mmol, 1 equiv.) in anhydrous THF (20 mL) was cooled under -78 °C. Then 24 mL of 1 M LiHMDS solution in THF (24 mmol, 2.18 equiv.) was added to the solution dropwise. After addition, the

solution was stirred under -78 °C for 1 h. Meanwhile, boromoacetonitrile was stirred with activated basic alumina for 2 h and then filtered. Freshly dried and filtered bromoacetonitrle (1.4 g, 11.8 mmol, 1.06 equiv.) was then added dropwise to the dianion solution. The solution was then stirred under -78 °C for 3~5 h, until TLC confirms complete consumption of the starting material. Then the reaction was quenched with precooled methanol (1 mL) in one portion and stirred under the same temperature for 30 min. The methoxide solution was then quenched with pre-cooled AcOH/THF (1 mL in 6 mL THF) in one portion and stirred for another 30 min under the same temperature. Then the cooling bath was removed. The reaction mixture was allowed to warm up to room temperature and poured into 50 mL of saturated brine solution. The layers were separated, and the organic layer was then concentrated to give dark oil. Then to the residue was added 4 g of silica gel, 1 g of activated charcoal and 50 mL of dichloromethane. The slurry was stirred for 1 h, and then filtered and washed with another 50 mL of dichloromethane. The filtrate was then concentrated to give brown oil, which was used without further purification.

Methyl (S)-2-((tert-butoxycarbonyl)amino)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (88)

To a pre-cooled solution of $CoCl_2 \cdot 6H_2O$ (1.54 g, 6.5 mmol) and **13** (11 mmol, crude) in methanol under 0 °C was added $NaBH_4$ (44 mmol, 1.67 g) in portions over 30 min. The reaction was exothermic and produces copious amount of hydrogen and black precipitate. The reaction mixture was stirred under room temperature for 24 h, and then concentrated on vacuo. The residue oil was then poured into 10% citric acid and filtered.

The filtrate was then extracted with ethyl acetate twice. The organic layer was then dried over anhydrous Na_2SO_4 and concentrated. The residue was purified by flash chromatography to afford **88** as light-yellow oil (2.1 g, 66%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 6.58 (s, 1H), 5.58 (d, $J = 8.5$ Hz, 1H), 4.19-4.36 (m, 1H), 3.71 (s, 3H), 3.23-3.39 (m, 2H), 2.36-2.54 (m, 2H), 2.04-2.19 (m, 1H), 1.73-1.90 (m, 1H), 1.41 (s, 9H).

(S)-1-Methoxy-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-amine hydrochloride (89)

To a solution of **88** (2.1 g, 7.3 mmol) in 1,4-dioxane (10 mL) was added dropwise a HCl solution in 1,4-dioxane (4 M, 10 mL). The resulting solution was stirred at room temperature for 1 h. Then residue was then concentrated *on vacuo* to afford **5** as light-yellow hydroscopic crystal (1.5 g, 92%). ^{1}H NMR (400 MHz, d$_6$-DMSO): δ 8.72 (s, 3H), 7.97 (s, 1H), 4.13-4.24 (m, 1H), 3.76 (s, 3H), 3.12-3.24 (m, 2H), 2.54-2.65 (m, 1H), 2.23-2.34 (m, 1H), 2.01-2.10 (m, 1H), 1.83-1.92 (m, 1H), 1.62-1.73 (m, 1H).

Scheme 27. Synthesis of MPI1
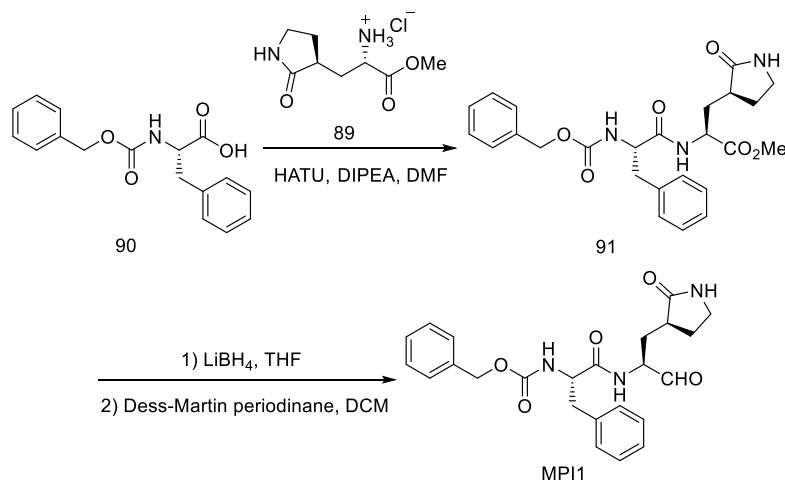
Methyl (*S*)-2-((*S*)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate (91)

To a solution of **90** (2 mmol, 0.44 g) and **89** (2 mmol, 0.44 g) in anhydrous DMF (10 mL) was added DIPEA (4 mmol, 0.52 g) and was cooled to 0 °C. HATU (2.2 mmol, 0.84 g) was added to the solution under 0 °C and then stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (50 mL) and washed with saturated NaHCO3 solution (2×20 mL), 1 M HCl solution (2×20 mL), and saturated brine solution (2×20 mL) sequentially. The organic layer was dried over anhydrous Na_2SO_4 and then concentrated *on vacuo*. The residue was then purified with flash chromatography (50-100% EtOAc in hexanes as the eluent) to afford **91** as white solid (520 mg, 56%). ^{1}H NMR (400 MHz, d_6-DMSO): δ 8.59 (d, *J* = 7.9 Hz, 1H), 7.66 (s, 1H), 7.53 (d, *J* = 8.5 Hz, 1H), 7.03-7.43 (m, 10H), 4.93 (q, *J* = 12.3, 11.8 Hz, 2H), 4.31-4.42 (m, 1H), 4.22-4.31 (m, 1H), 3.63 (s,3H), 2.91-3.18 (m, 3H), 2.68-2.79 (m, 1H), 2.23-2.36 (m, 1H), 2.01-2.17 (m, 2H), 1.53-1.67 (m, 2H).

Benzyl ((*S*)-1-oxo-1-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-3-phenylpropan-2-yl)carbamate (MPI1)

To a solution of **91** (0.1 mmol, 47 mg) in anhydrous dichloromethane (5 mL) was added a solution of LiBH4 in anhydrous THF (2 M, 0.1 mL, 0.2 mmol) at 0 °C. The resulting solution was stirred at the same temperature for 3 h. Then a saturated solution of NH4Cl (5 mL) was added dropwise to quench the reaction. The layers were separated, and the organic layer was washed with saturated brine solution (2×10 mL), dried over anhydrous Na2SO4. and evaporated to dryness. The residue was then dissolved in

anhydrous dichloromethane (5 mL) and cooled to 0 °C. Dess-Martin periodinane (0.2 mmol, 85 mg) was added to the solution. The reaction mixture was then stirred at room temperature overnight. Then the reaction was quenched with a saturated $NaHCO_3$ solution containing 10% $Na_2S_2O_3$. The layers were separated. The organic layer was then washed with saturated brine solution (2×10 mL), dried over anhydrous Na_2SO_4 and evaporated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **MPI1** as white solid (30 mg, 65%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 9.26 (s, 1H), 8.18 (s, 1H), 7.46-7.08 (m, 10H), 5.54 (s, 1H), 5.44 (d, $J =$ 9.2 Hz, 1H), 5.11 (s, 2H), 4.65-4.53 (m, 1H), 4.30-4.16 (m, 1H), 3.37-3.24 (m, 2H), 3.23-3.14 (m, 1H), 3.07 (dd, $J = 13.6, 6.5$ Hz, 1H), 2.40-2.28 (m, 1H), 2.27-2.19 (m, 1H), 1.92-1.73 (m, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 199.98, 180.16, 172.13, 155.96, 136.39, 129.62, 128.73, 128.65, 128.29, 128.15, 127.14, 67.11, 58.05, 56.14, 40.72, 38.96, 38.17, 31.08, 29.53, 28.88; ESI-MS calcd for $C_{24}H_{28}N_3O_5$ (M+H$^+$): 438.2; found 438.3.

Scheme 28. Synthesis of MPI2.

119
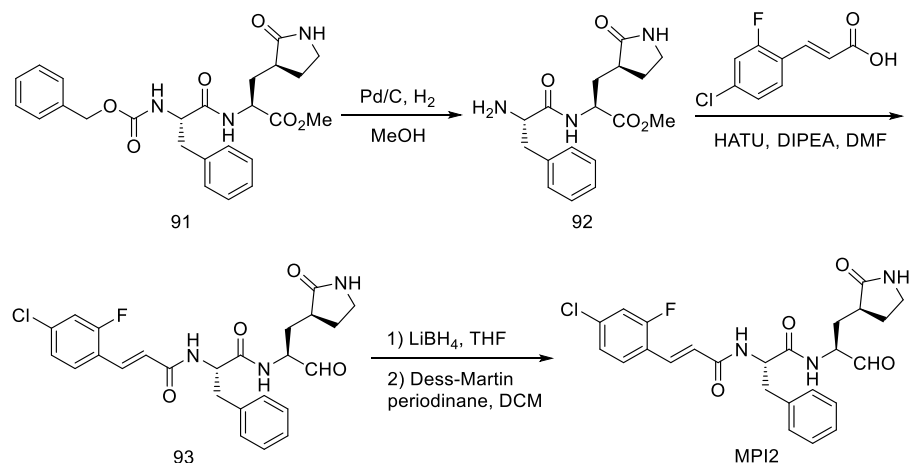
Methyl (*S*)-2-((*S*)-2-((*E*)-3-(3-chloro-5-fluorophenyl)acrylamido)-3-phenylpropanamido)-3-((*S*)-2-oxopyrrolidin-3-yl)propanoate (93)

To a solution of **91** (0.25 mmol, 116 mg) in methanol was added 10 % Pd/C (26 mg). The mixture was then stirred with hydrogen balloon at room temperature for 3 h. The catalyst was then filtered off and the solution was evaporated on vacuo to afford **92** as white solid, which was used without purification. To a solution of **92** in 2 mL dry DMF was added DIPEA (0.5 mmol, 65 mg) and cooled to 0 °C. Then HATU (0.3 mmol, 114 mg) was added to the solution at the same temperature. The solution was stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (20 mL) and washed with saturated $NaHCO_3$ solution (2×10 mL), 1 M HCl solution (2×10 mL), and saturated brine solution (2×10 mL) sequentially. The organic layer was dried over anhydrous Na_2SO_4 and then concentrated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **93** as white solid (80 mg, 62%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.92 (d, *J* = 6.7 Hz, 1H), 7.62 (d, *J* = 15.8 Hz, 1H), 7.40 (t, *J* = 8.3 Hz, 1H), 7.32-7.17 (m, 5H), 7.11 (td, *J* = 8.7, 7.4, 3.0 Hz, 2H), 6.61-6.43 (m, 2H), 5.89 (s, 1H), 5.02 (dd, *J* = 8.2, 5.9 Hz, 1H), 4.51-4.33 (m, 1H), 3.72 (s, 3H), 3.38-3.25 (m, 2H), 3.25-3.11 (m, 2H), 2.44-2.31 (m, 1H), 2.27-2.19 (m, 1H), 2.17-2.05 (m, 1H), 1.95-1.74 (m, 2H).

(*E*)-3-(3-chloro-5-fluorophenyl)-*N*-((*S*)-1-oxo-1-(((*S*)-1-oxo-3-((*S*)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-3-phenylpropan-2-yl)acrylamide (MPI2)

To a solution of **93** (0.1 mmol, 52 mg) in anhydrous dichloromethane (5 mL) was added a solution of $LiBH_4$ in anhydrous THF (2 M, 0.1 mL, 0.2 mmol) at 0 °C. The

resulting solution was stirred at the same temperature for 3 h. Then a saturated solution of NH$_4$Cl (5 mL) was added dropwise to quench the reaction. The layers were separated, and the organic layer was washed with saturated brine solution (2 × 10 mL), dried over anhydrous Na2SO4. and evaporated to dryness. The residue was then dissolved in anhydrous dichloromethane (5 mL) and cooled to 0 ° C. Dess-Martin periodinane (0.2 mmol, 85 mg) was added to the solution. The reaction mixture was then stirred at room temperature overnight. Then the reaction was quenched with a saturated NaHCO$_3$ solutioncontaining 10% Na2S2O3. The layers were separated. The organic layer was then washed with saturated brine solution (2×10 mL), dried over anhydrous Na2SO4 and evaporated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **MPI2** as white solid (27 mg, 55%). ^{1}H NMR (400 MHz, DMSO-d$_6$): δ 9.30 (s, 1H), 8.67 (d, J = 7.6 Hz, 1H), 8.58 (d, J = 8.2 Hz, 1H), 7.68-7.61 (m, 2H), 7.52 (dd, J = 10.8, 2.1 Hz, 1H), 7.41-7.32 (m, 2H), 7.30-7.15 (m, 5H), 6.81 (d, J = 16.0 Hz, 1H), 4.75-4.67 (m, 1H), 4.21-4.12 (m, 1H), 3.18-3.04 (m, 3H), 2.89 (dd, J = 13.8, 9.3 Hz, 1H), 2.25-2.05 (m, 2H), 1.88 (ddt, J = 13.9, 11.3, 5.6 Hz, 1H), 1.67-1.55 (m, 2H); ^{13}C NMR (101 MHz, CDCl$_3$) δ 199.87, 179.97, 171.97, 165.18, 162.26, 159.71, 136.30, 133.45, 130.13, 129.55, 128.59, 127.05, 124.96, 123.37, 121.40, 116.82, 57.99, 54.27, 40.62, 38.83, 38.10, 29.42, 28.75; ESI-MS: calcd for C$_{25}$H$_{26}$ClFN$_3$O$_4$ (M+H+): 486.1; found 486.1.

Scheme 29. Synthesis of MPI3.
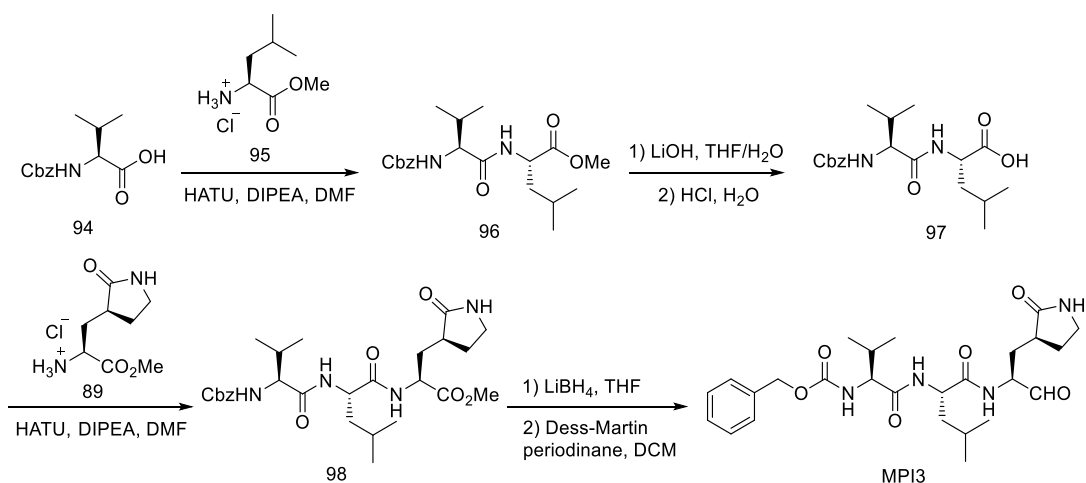
(S)-Methyl 2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-4-methylpentanoate (96)

The amino acid methyl ester hydrochloride **95** (1.0 g, 5.52 mmol) and the Cbz-protected amino acid **94** (1.88 g, 6.08 mmol) were dissolved in dry DMF (20 mL) and the reaction was cooled to 0 °C. HATU (2.52 g, 6.62 mmol) and DIPEA (3.92 mL, 22.08 mmol) were added, and the reaction mixture was allowed warm up to room temperature and stirred for 12 h. The mixture was then poured into water (50 mL) and extracted with ethyl acetate (4×20 mL). The organic layer was washed with aqueous hydrochloric acid 10% v/v (2×20 mL), saturated aqueous $NaHCO_3$ (2×20 mL), brine (2×20 mL) and dried over Na_2SO_4. The organic phase was evaporated to dryness and the crude material purified by silica gel column chromatography (15-50% EtOAc in n-hexane as the eluent) to afford **96** white solid (1.82, 69%). ^{1}H NMR ($CDCl_3$, 400MHz) δ 7.28-7.22 (m, 5H), 6,28 (d, J = 7.72 Hz, 1H), 5.35 (d, J = 8.64 Hz, 1H), 5.03 (s, 2H), 4.56-451 (m, 1H), 3.97 (t, J = 8.12 Hz, 1H), 3.65 (s, 3H), 2.19-1.98 (m, 1H), 1.62-1.43 (m, 3H), 0.92-0.83 (m, 12H); [13]C

NMR (CDCl$_3$, 100MHz) δ 173.2, 171.1, 156., 136.2, 128.5 (2C), 128.2, 128 (2C), 67.0, 60.2, 52.3, 50.7, 31.3, 2.8, 22.7, 21.9, 19.1, 17.8.

(*S*)-2-((*S*)-2-(((Benzyloxy)carbonyl)amino)-3-methylbutanamido)-4-methylpentanoic acid (97)

The dipeptide methyl ester **96** (500 mg, 1.14 mmol) was dissolved in THF/H$_2$O (1:1, 10 mL). LiOH (114 mg, 2.86 mmol) was added at 0 °C. The mixture was stirred at room temperature overnight. Then THF was removed *in vacuo* and the aqueous layer was acidified with 1 M HCl and extracted with dichloromethane (3×10 mL). The organic layer was dried over anhydrous Na$_2$SO$_4$ and concentrated to yield **97** as white solid (315 mg, 65%). ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.26-7.23 (m, 5H), 6.65 (d, *J* = 7.88, 1H), 5.68 (d, *J* = 4.64 Hz, 1H), 5.03 (s, 2H), 4.56-4.45 (m, 1H), 3.97 (t, *J* = 7.92 Hz, 1H), 2.02-1.96 (m, 1H), 1.66-1.46 (m, 3H), 0.85 (dd, *J* = 7.36, 13.3 Hz, 12H); ^{13}C NMR (CDCl$_3$, 100 MHz): δ 176.0, 171.8, 156.7, 136.1, 128.5 (2C), 128.2, 128.0 (2C), 67.2, 60., 50.8, 41.1, 31.1, 24.8, 22.8, 21.8, 19.1, 18.0.

(5*S*,8*S*,11*S*)-Methyl 8-isobutyl-5-isopropyl-3,6,9-trioxo-11-(((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (98)

The methyl (S)-2-amino-3-((S)-2-oxopyrrolidin-3-yl)propanoate hydrochloride **89** (150 mg, 0.657 mmol) and the dipeptide **97** (270 mg, 0.743 mmol) were dissolved in dry DMF (10 mL) and the reaction was cooled to 0 °C. HATU (308 mg, 0.788 mmol) and DIPEA (0.48 mL, 2.63 mmol) were added, and the reaction mixture was allowed to warm up to room temperature and stirred for 12 h. The mixture was then poured into water (20 mL) and extracted with ethyl acetate (4×20 mL). The organic layer was washed with

aqueous hydrochloric acid 10% v/v (2×20 mL), saturated aqueous NaHCO$_3$ (2×20 mL), brine (2×20 mL) and dried over Na2SO4. The organic phase was evaporated to dryness and the crude material purified by silica gel column chromatography afford **98** as white solid (250 mg, 70%). ^{1}H NMR (CDCl$_3$, 400 MHz) δ 7.89 (d, J = 7.16 Hz, 1H), 7.28-7.20 (m, 5H), 7.00 (d, J = 8.24 Hz, 1H), 6.65 (brs, 1H), 5.47 (d, J = 8.8 Hz, 1H), 5.01 (s, 2H), 4.59-4.51 (m, 1H), 4.46-4.38 (m, 1H), 3.94 (t, J = 7.84 Hz, 1H), 3.63 (s, 3H), 3.28-3.18 (m, 2H), 2.38-2.22 (m, 2H), 2.29-1.98 (m, 2H), 1.79-1.70 (m, 1H), 1.68-1.51 (m, 2H), 1.50-1.25 (m, 2H), 0.86-0.81 (m, 12H); ^{13}C NMR (CDCl$_3$, 100 MHz) δ 179.8, 172.5, 172.1, 171.2, 156.5, 136.2, 128.5 (2C), 128.2, 128.0 (2C), 67.1, 60.5, 52.4, 51.7, 51.1, 42.0, 40.5, 38.3, 33.1, 31.1, 28.1, 24.6, 22.8, 22.0, 19.7, 19.1.

Benzyl ((S)-3-methyl-1-((((S)-4-methyl-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)pentan-2-yl)amino)-1-oxobutan-2-yl)carbamate (MPI3)

To a stirred solution of compound **98** (120 mg, 0.254 mmol) in THF (8 mL) was added LiBH$_4$ (2.0 M in THF, 0.636 mL, 1.27 mmol) in several portions at 0 °C under a nitrogen atmosphere. The reaction mixture was stirred at 0 °C for 1 h, then allowed to warm up to room temperature, and stirred for an additional 2 h. The reaction was quenched by the drop wise addition of 1.0 M HCl (aq) (1.2 mL) with cooling in an ice bath. The solution was diluted with ethyl acetate and H$_2$O. The phases were separated, and the aqueous layer was extracted with ethyl acetate (3×15 mL). The organic phases were combined, dried over MgSO$_4$, filtered, and concentrated *in vacuo* to give a yellow oily residue. The residue was dissolved in CH$_2$Cl$_2$ (6 mL) and cooled at 0 °C. Dess-Martin reagent (180 mg, 0.427 mmol, 3 equiv) was added at 0 °C. The resulting mixture was then

stirred at room temperature for 3 h. Then the reaction was quenched with a saturated NaHCO$_3$ solution containing 10 % Na$_2$S$_2$O$_3$. The layers were separated. The organic layer was then washed with saturated brine solution, dried over anhydrous Na$_2$SO$_4$ and concentrated *on vacuo*. The residue was then purified with flash chromatography to afford **MPI3** as white solid (45 mg, 64%). ^{1}H NMR (CDCl$_3$, 400 MHz): δ 9.41 (s, 1H), 8.18 (d, J = 5.12 Hz, 1H), 7.28 - 7.26 (m, 5H), 6.63 (d, J = 7.0 Hz, 1H), 6.15 (brs, 1H), 5.40 (d, J = 6.88 Hz, 1H), 5.02 (s, 2H), 4.52 - 4.50 (m, 1H), 4.27 - 4.25 (m, 1H), 3.94 - 3.92 (m, 2H), 3.27 - 3.23 (m, 2H), 2.35-2.27 (m, 2H), 2.07-2.05 (m, 1H), 1.90 - 1.81 (m, 2H), 1.76-1.71 (m, 1H), 1.57-1.43 (m, 3H), 0.85 (dd, J = 6.76, 14.72 Hz, 12H); ^{13}C NMR (CDCl$_3$, 100 MHz): δ 199.6, 180.0, 173.2, 171.4, 156.6, 136.2, 128.6 (2C), 128.2, 128.1 (2C), 67.1, 60.6, 57.5, 51.2, 41.7, 40.6, 38.0, 31.0, 29.8, 28.4, 24.8, 22.9, 21.9, 19.2, 17.8.

Scheme 30. Synthesis of MPI4.
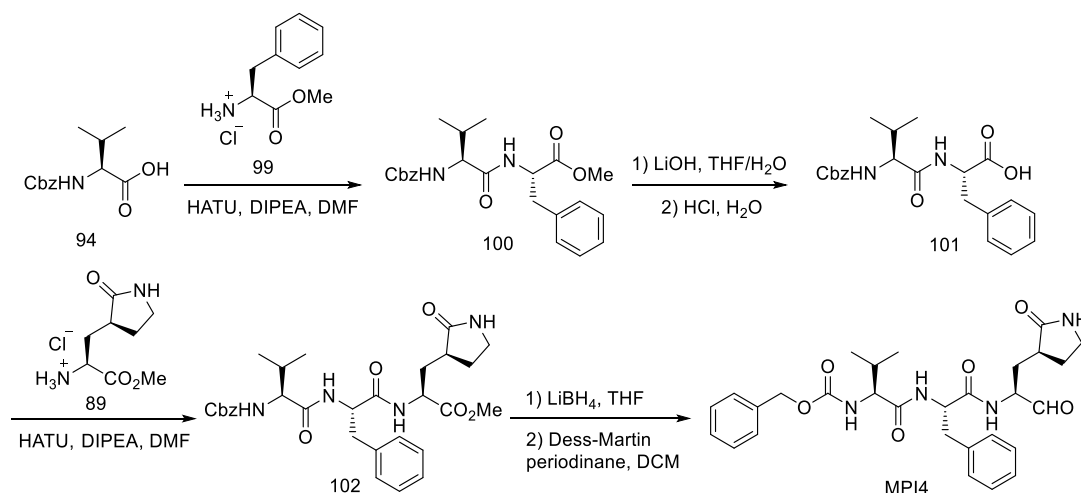
Methyl ((benzyloxy)carbonyl)-L-valyl-L-phenylalaninate (100)

To a solution of **94** (5 mmol, 1.25 g) and **95** (5 mmol, 1.07g) in anhydrous DMF (20 mL) was added DIPEA (10 mmol, 1.29 g) and was cooled to 0 ° C. HATU (5.5 mmol,

2.09 g) was added to the solution under 0 °C and then stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (100 mL) and washed with saturated $NaHCO_3$ solution (2×50 mL), 1 M HCl solution (2×50 mL), and saturated brine solution (2×50 mL) sequentially. The organic layer was dried over anhydrous Na_2SO_4 and then concentrated *on vacuo*. The residue was then purified with flash chromatography (15-50% EtOAc in hexanes as the eluent) to afford **100** as white soild (1.52 g, 74%).

((Benzyloxy)carbonyl)-L-valyl-L-phenylalanine (101)

100 (1 mmol, 470 mg) was dissolved in 5 mL of THF. A solution of LiOH · H_2O (2 mmol, 84 mg) in 5 mL H_2O was added to the solution. The mixture was stirred at room temperature overnight. Then THF was removed *in vacuo* and the aqueous layer was acidified with 1 M HCl and extracted with dichloromethane (3×10 mL). The organic layer was dried over anhydrous Na_2SO_4 and evaporated to give **101** as white solid (312 mg, 76 %). ^{1}H NMR (400 MHz, CD_3OD): δ 7.29 - 7.42 (m, 5H), 7.16 - 7.29 (m, 5H), 5.11 (s, 2H), 4.73 – 4.68 (m, 1H), 3.94 (d, J = 7.3 Hz, 1H), 3.21 (dd, J = 13.9, 5.2 Hz, 1H), 3.01 (dd, J = 13.9, 8.6 Hz, 1H), 2.06 – 1.95 (m, 1H), 0.91 (dd, J = 8.5, 6.7 Hz, 6H).

Methyl (5S,8S,11S)-8-benzyl-5-isopropyl-3,6,9-trioxo-11-(((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (102)

To a solution of **101** (0.4 mmol, 160 mg) and **89** (0.4 mmol, 88 mg) in anhydrous DMF (2 mL) was added DIPEA (0.8 mmol, 103 mg) and was cooled to 0 °C. HATU (0.44 mmol, 167 mg) was added to the solution under 0 °C and then stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (20 mL) and washed

with saturated NaHCO$_3$ solution (2×10 mL), 1 M HCl solution (2×10 mL), and saturated

brine solution (2×10 mL) sequentially. The organic layer was dried over anhydrous

Na$_2$SO$_4$ and then concentrated *in vacuo*. The residue was then purified with flash

chromatography (1-10% methanol in dichloromethane as the eluent) to afford **102** as white

solid (151 mg, 67%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.71 (d, J = 7.1 Hz, 1H), 7.42 – 7.30

(m, 5H), 7.25 – 7.13 (m, 5H), 6.78 (d, J = 8.5 Hz, 1H), 5.83 (s, 1H), 5.26 (d, J = 8.7 Hz,

1H), 5.10 (d, J = 4.2 Hz, 2H), 4.83 (q, J = 6.9 Hz, 1H), 4.51 – 4.41 (m, 1H), 3.96 (dd, J =

8.6, 6.2 Hz, 1H), 3.70 (s, 3H), 3.37 – 3.25 (m, 2H), 3.11 (d, J = 6.3 Hz, 2H), 2.44 – 2.31

(m, 1H), 2.26 – 2.13 (m, 1H), 2.13 – 1.99 (m, 2H), 1.93 – 1.74 (m, 2H), 0.91 (d, J = 6.8

Hz, 3H), 0.82 (d, J = 6.8 Hz, 3H).

Benzyl ((S)-3-methyl-1-oxo-1-(((S)-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-3-phenylpropan-2-yl)amino)butan-2-yl)carbamate (MPI4)

To a solution of **102** (0.1 mmol, 57 mg) in anhydrous dichloromethane (5 mL) was

added a solution of LiBH$_4$ in anhydrous THF (2 M, 0.1 mL, 0.2 mmol) at 0 °C. The

resulting solution was stirred at the same temperature for 3 h. Then a saturated solution of

NH$_4$Cl (5 mL) was added dropwise to quench the reaction. The layers were separated, and

the organic layer was washed with saturated brine solution (2×10 mL), dried over

anhydrous Na$_2$SO$_4$. and evaporated to dryness. The residue was then dissolved in

anhydrous dichloromethane (5 mL) and cooled to 0 °C. Dess-Martin periodinane (0.2

mmol, 85 mg) was added to the solution. The reaction mixture was then stirred at room

temperature overnight. Then the reaction was quenched with a saturated NaHCO$_3$ solution

containing 10 % Na$_2$S$_2$O$_3$. The layers were separated. The organic layer was then washed

with saturated brine solution (2×10 mL), dried over anhydrous Na_2SO_4 and evaporated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **MPI4** as white solid (30 mg, 65 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 9.26 (s, 1H), 8.18 (s, 1H), 7.46 – 7.08 (m, 10H), 5.54 (s, 1H), 5.44 (d, J = 9.2 Hz, 1H), 5.11 (s, 2H), 4.65 – 4.53 (m, 1H), 4.30 – 4.16 (m, 1H), 3.37 – 3.24 (m, 2H), 3.23 – 3.14 (m, 1H), 3.07 (dd, J = 13.6, 6.5 Hz, 1H), 2.40 – 2.28 (m, 1H), 2.27 – 2.19 (m, 1H), 1.92 – 1.73 (m, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 199.98, 180.16, 172.13, 155.96, 136.39, 129.62, 128.73, 128.65, 128.29, 128.15, 127.14, 67.11, 58.05, 56.14, 40.72, 38.96, 38.17, 31.08, 29.53, 28.88; ESI-MS calcd for $C_{24}H_{28}N_3O_5$ (M+H[+]): 438.2; found 438.3.

Scheme 31. Synthesis of MPI5.
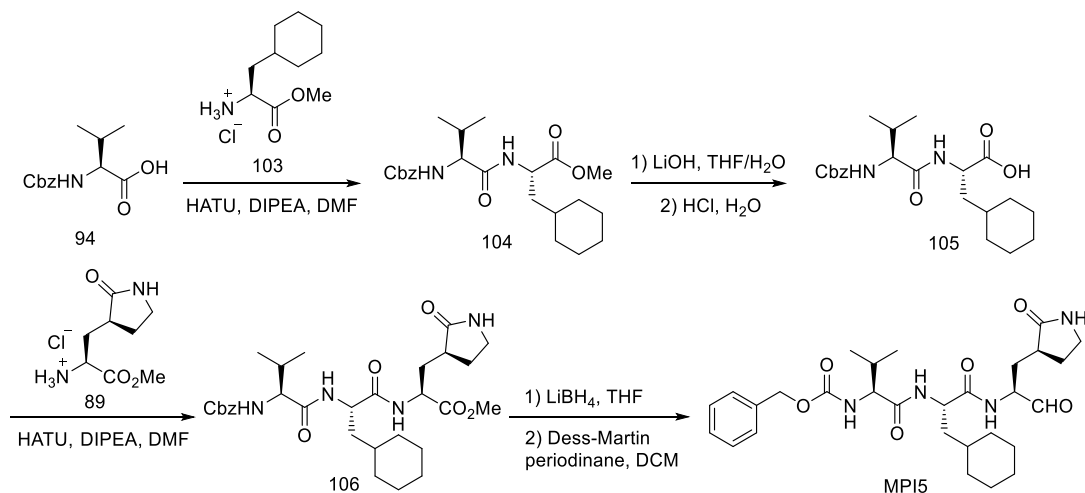
Methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-cyclohexylpropanoate (104).

To a solution of **94** (2g, 7.95 mmol, 1.0 equiv) in anhydrous DMF (15 mL) at 0 °C, and then **103** (1.8 g, 7.95 mmol, 1,0 equiv), HATU (4.5 g, 12.0 mmol, 1.5 equiv),

DIPEA (7.0 mL, 40.0 mmol, 5.0 equiv) was added sequentially. The mixture was stirred at room temperature for 6 h. The mixture was diluted with EtOAc and washed with water, 1M HCl, sat. NaCl, dried over Na_2SO_4, and concentrated. The residue was purified by column chromatography (EtOAc: Hexane = 1:2 v/v) to afford the pure product **104** as a white solid (2.7 g, 81%). ^{1}H NMR (400 MHz, CD_3OD-d_4) δ 7.4 – 7.3 (m, 5H), 5.1 (d, J = 1.6 Hz, 2H), 4.5 (dd, J = 9.6, 5.6 Hz, 1H), 4.0 (d, J = 7.5 Hz, 1H), 3.7 (s, 3H), 2.2 – 2.0 (m, 1H), 1.8 – 1.6 (m, 7H), 1.4 (tdd, J = 11.0, 6.9, 3.6 Hz, 1H), 1.3 – 1.1 (m, 3H), 1.0 (dd, J = 11.7, 6.7 Hz, 7H), 0.9 – 0.8 (m, 1H). ^{13}C NMR (100 MHz, CD_3OD-d_4) δ 174.5, 174.3, 158.5, 138.2, 129.4, 129.4, 129.0, 128.8, 128.8, 67.6, 61.9, 52.5, 51.3, 39.9, 35.2, 34.7, 33.1, 32.0, 27.5, 27.3, 27.1, 19.7, 18.7.

(S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-methylbutanamido)-3-cyclohexylpropanoic acid (105).

To a solution of **104** (400 mg, 1.2 mmol, 1.0 equiv) in 1:1 THF/H_2O (8 mL) was added LiOH • H_2O (200 mg, 4.8 mmol, 4.0 equiv). The reaction was stirred at room temperature for 2 h. After completion, the reaction mixture was neutralized with 1M HCl solution and extracted with EtOAc. The organic layer was washed with sat. NaCl, dried over Na_2SO_4 and concentrated to afford the product **105** (310 mg, yield 80%) as a white solid. The residue was used in the next without further purification.

Methyl (5S,8S,11S)-8-(cyclohexylmethyl)-5-isopropyl-3,6,9-trioxo-11-(((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (106).

To a solution of **105** (300 mg, 0.74 mmol, 1.0 equiv) in anhydrous DMF (5 mL) at 0 °C, and then **89** (165 mg, 0.74 mmol, 1.0 equiv), HATU (400 mg, 1.05 mmol, 1.5 equiv),

DIPEA (610 μL, 3.7 mmol, 5.0 equiv) was added sequentially. The mixture was stirred at room temperature for 6 h. The mixture was diluted with EtOAc and washed with water, 1M HCl, sat. NaCl, dried over Na_2SO_4, and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:20 v/v) to afford the pure product **106** as a white solid (250 mg, 60%).

Benzyl ((S)-1-(((S)-3-cyclohexyl-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)propan-2-yl)amino)-3-methyl-1-oxobutan-2-yl)carbamate (MPI5).

To a solution of **106** (250 mg, 0.44 mmol, 1.0 equiv) in anhydrous THF (10 mL) at 0 °C was added $LiBH_4$ (1.0 M in THF, 1.32 mL, 1.32 mmol, 3.0 equiv). The mixture was stirred at room temperature for 2 h. After the reaction was completed, excess reactants were consumed by slow addition of H_2O. The mixture was diluted with H_2O and extracted with EtOAc, washed with sat. NaCl, dried over Na_2SO_4 and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:15 v/v) to afford a white solid (130 mg, 54%). The solid (130 mg, 0.24 mmol, 1.0 equiv) was dissolved in anhydrous DCM (10 mL). Dess-Martin reagent (200 mg, 0.48 mmol, 2.0 equiv) was added slowly at 0 °C. Then the reaction mixture was stirred at room temperature for 1 h. A solution of $NaHCO_3$ and $Na_2S_2O_3$ was added to quench the reaction. After 10 min, the mixture was washed with water, sat. NaCl, dried over Na_2SO_4 and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:15 v/v) to afford **MPI5** as white solid (60 mg, 47%). ^1H NMR (400 MHz, $CDCl_3$-*d*) δ 9.4 (s, 1H), 8.1 (d, *J* = 6.9 Hz, 1H), 7.3 (s, 5H), 7.1 (d, *J* = 8.5 Hz, 1H), 6.7 (s, 1H), 5.6 (d, *J* = 8.8 Hz, 1H), 5.0 (q, *J* = 12.3

Hz, 2H), 4.6 (td, *J* = 8.9, 5.7 Hz, 1H), 4.3 (p, *J* = 5.0 Hz, 1H), 4.0 (t, *J* = 7.8 Hz, 1H), 3.3

– 3.1 (m, 2H), 2.4 – 2.2 (m, 2H), 2.1 – 1.9 (m, 2H), 1.8 (ddd, *J* = 13.6, 7.3, 4.1 Hz, 1H),

1.7 – 1.4 (m, 8H), 1.3 – 1.2 (m, 1H), 1.1 – 1.0 (m, 3H), 0.8 (dd, *J* = 13.4, 6.9 Hz, 8H). 13C

NMR (100 MHz, CDCl3-*d*) δ 199.6, 180.1, 173.5, 171.5, 156.7, 136.3, 128.7, 128.7,

128.3, 128.1, 128.1, 67.2, 60.7, 57.4, 51.2, 40.7, 40.3, 38.0, 34.3, 33.6, 32.6, 31.1, 30.0,

28.4, 26.5, 26.3, 26.2, 19.3, 19.3. ESI-MS calcd for $C_{29}H_{43}N_4O_6^+$ (M+H$^+$): 543.3; found

543.3.

Scheme 32. Synthesis of MPI6.
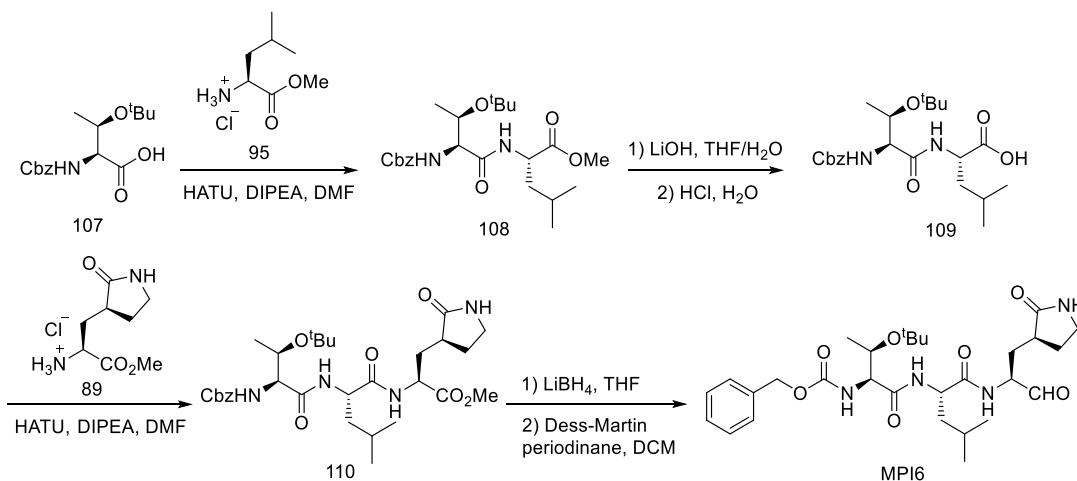
(S)-Methyl 2-((2S,3R)-2-(((benzyloxy)carbonyl)amino)-3-(tert-butoxy)butanamido)-4-methylpentanoate (108)

95 (1.0 g, 5.52 mmol) and the **107** (1.52 g, 6.08 mmol) were dissolved in dry DMF

(20 mL) and the reaction was cooled to 0 °C. HATU (2.52 g, 6.62 mmol) and DIPEA (3.92

mL, 22.08 mmol) were added, and the reaction mixture was allowed warm up to room

temperature and stirred for 12 h. The mixture was then poured into water (50 mL) and

extracted with ethyl acetate (4×20 mL). The organic layer was washed with aqueous

131

hydrochloric acid 10% v/v (2×20 mL), saturated aqueous $NaHCO_3$ (2×20 mL), brine (2×20 mL) and dried over Na2SO4. The organic phase was evaporated to dryness and the crude material purified by silica gel column chromatography (15-50% EtOAc in hexanes as the eluent) to afford **108** as colorless oil (1.71 g, 71%). ^{1}H NMR (CDCl3, 400 MHz): δ 7.58 (d, *J* = 7.64 Hz, 1H), 7.28-7.21 (m, 5H), 5.88 (d, *J* = 4.68 Hz, 1H), 5.04 (ABq, *J* = 12.16 Hz, 2H), 4.47-4.41 (m, 1H), 4.15-4.08 (m, 2H), 3.65 (s, 3H), 1.63-1.46 (m, 3H), 1.23 (s, 9H), 1.03 (d, *J* = 6.28, 3H), 0.86 (dd, *J* = 3.84, 5.92 Hz, 6H); ^{13}C NMR (CDCl3, 100 MHz) δ 172.9, 169.4, 156.1, 136.3, 128.5 (2C), 128.0, 127.9, 75.5, 66.8, 60.4, 58.4, 52.2, 51.1, 41.3, 28.2 (3C), 25.0 22.8, 21.9, 16.4.

(S)-2-((2S,3R)-2-(((Benzyloxy)carbonyl)amino)-3-(tert-butoxy)butanamido)-4-methylpentanoic acid (109)

108 (500 mg, 1.32 mmol) was dissolved in THF/H$_2$O (1:1, 6.0 mL), and LiOH (138 mg, 3.30 mmol) was added at 0 °C. The mixture was stirred at room temperature overnight. Then THF was removed *in vacuo* and the aqueous layer was acidified with 1 M HCl and extracted with dichloromethane (3×10 mL). The organic layer was dried over anhydrous Na$_2$SO$_4$ and concentrated to yield **109** as white solid (350 mg, 70%). ^{1}H NMR (CDCl3, 400 MHz): δ 7.62 (d, *J* = 7.44 Hz, 1H), 7.29-7.22 (m, 5H), 5.94 (d, *J* = 5.28 Hz, 1H), 5.04 (ABq, *J*=12.32 Hz, 2H), 4.45-4.40 (m, 1H), 4.17-4.08 (m, 2H), 1.68-1.50 (m, 3H), 1.21 (s, 9H), 1.02 (d, *J* = 6.32 Hz, 3H), 0.87 (m, 6H); ^{13}C NMR (CDCl3, 100 MHz) δ 177.1, 169.9, 156.2, 136.2, 128.5 (2C), 128.2, 128.0 (2C), 75.6, 66.97, 66.92, 58.4, 51.1, 41.0, 28.2 (3C), 25.0, 22.8, 21.8, 16.5.

(5S,8S,11S)-Methyl 5-((R)-1-(tert-butoxy)ethyl)-8-isobutyl-3,6,9-trioxo-11-(((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (110)

89 (140 mg, 0.606 mmol) and **109** (242 mg, 0.666 mmol) were dissolved in dry DMF (6 mL), and the reaction was cooled to 0 °C. HATU (276 mg, 0.727 mmol) and DIPEA (0.43 mL, 2.42 mmol) were added, and the reaction mixture was allowed warm up to room temperature and stirred for 12 h. The mixture was then poured into water (20 mL) and extracted with ethyl acetate (4×20 mL). The organic layer was washed with aqueous hydrochloric acid 10% v/v (2×20 mL), saturated aqueous $NaHCO_3$ (2×20 mL), brine (2×20 mL) and dried over Na2SO4. The organic phase was evaporated to dryness and the crude material purified by silica gel column chromatography afford **110** as white solid (240 mg, 64%). ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.64 (d, J = 7.08 Hz, 1H), 7.39 (d, J = 7.92 Hz, 1H), 7.31-7.20 (m, 5H), 6.34 (brs, 1H), 5.87 (d, J = 5.0 Hz, 1H), 5.03 (ABq, J = 12.36 Hz, 2H), 4.50-4.43 (m, 1H), 4.50-4.43 (m, 1H), 4.40-4.34 (m, 1H), 4.11-4.09 (m, 2H), 3.67 (s, 3H), 3.29-3.15 (m, 2H), 2.38-2.23 (m, 2H), 2.10-2.02 (m, 1H), 1.81-1.61 (m, 4H), 1.51-1.44 (m, 1H), 1.18 (s, 9H), 0.99 (d, J = 5.76, 3H), 0.86 (dd, J = 5.6, 14.32, 6H). ^{13}C NMR (CDCl$_3$, 100 MHz): δ 179.7, 172.2, 169.5, 136.2, 128.6 (2C), 128.3, 128.1 (2C), 77.1, 75.4, 67.0, 66.7, 58.9, 52.4, 51.9, 51.1, 41.7, 40.5, 38.2, 33.0, 28.2 (3C), 28.2, 24.7, 22.8, 22.2, 17.2.

Benzyl ((2S,3R)-3-(tert-butoxy)-1-(((S)-4-methyl-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)pentan-2-yl)amino)-1-oxobutan-2-yl)carbamate (MPI6)

To a stirring solution of compound **110** (120mg, 0.225 mmol) in THF (5 mL) was added LiBH$_4$ (2.0 M in THF, 0.56 mL, 1.12 mmol) in several portions at 0 °C under a nitrogen atmosphere. The reaction mixture was stirred at 0 °C for 1 h, then allowed to warm up to room temperature, and stirred for an additional 2 h. The reaction was quenched by the drop wise addition of 1.0 M HCl (1.2 mL) with cooling in an ice bath. The solution was diluted with ethyl acetate and H$_2$O. The phases were separated, and the aqueous layer was extracted with ethyl acetate (3×15 mL). The organic phases were combined, dried over Na$_2$SO$_4$, filtered, and concentrated to give a yellow oily residue. Column chromatographic purification of the residue (6% MeOH in CH2Cl2 as the eluent) afforded a white solid (85 mg, 68%). To a solution of the solid (70 mg, 0.138 mmol) in DCM (6 mL) was Dess-Martin reagent (180 mg, 3 equiv). The resulting mixture was stirred at room temperature for 3 h. Then the reaction was quenched with a saturated NaHCO$_3$ solution containing 10 % Na$_2$S$_2$O$_3$. The layers were separated. The organic layer was then washed with saturated brine solution, dried over anhydrous Na$_2$SO$_4$ and concentrated *in vacuum*. The residue was then purified with flash chromatography afford **MPI6** as white solid (41 mg, 59%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 9.44 (s, 1H), 8.04 (d, J = 6.24 Hz, 1H), 7.41 (d, J= 7.64 Hz, 1H), 7.31-7.27 (m, 5H), 6.14 (brs, 1H), 5.85 (d, J= 4.68 Hz, 1H), 5.05 (ABq, J = 12.2 Hz, 2H), 4.45-4.38 (m, 1H), 4.34-4.28 (m, 1H), 4.18-4.05 (m, 2H), 3.30-3.11 (m, 2H), 2.48-2.33 (m, 1H), 2.33-2.21 (m, 1H), 2.11-1.92 (m, 2H), 1.91-1.89 (m, 1H), 1.87-1.58 (m, 1H), 1.57-1.53 (m, 2H), 1.52-1.48 (m, 1H), 1.21 (s, 9H), 1.01 (d, J = 5.92 Hz, 3H), 0.89 (dd, J = 6.12, 11.56 Hz, 6H). ^{13}C NMR (CDCl$_3$, 100 MHz): 199.6, 179.9, 172.8,

169.7, 156.2, 136.1, 128.6 (2C), 128.3, 128.1 (2C), 77.1, 75.4, 67.0, 66.7, 58.9, 57.6, 52.1, 41.6, 40.5, 37.9, 29.7, 28.5, 28.2 (3C), 24.9, 22.9, 22.1, 17.3.

Scheme 33. Synthesis of MPI7.
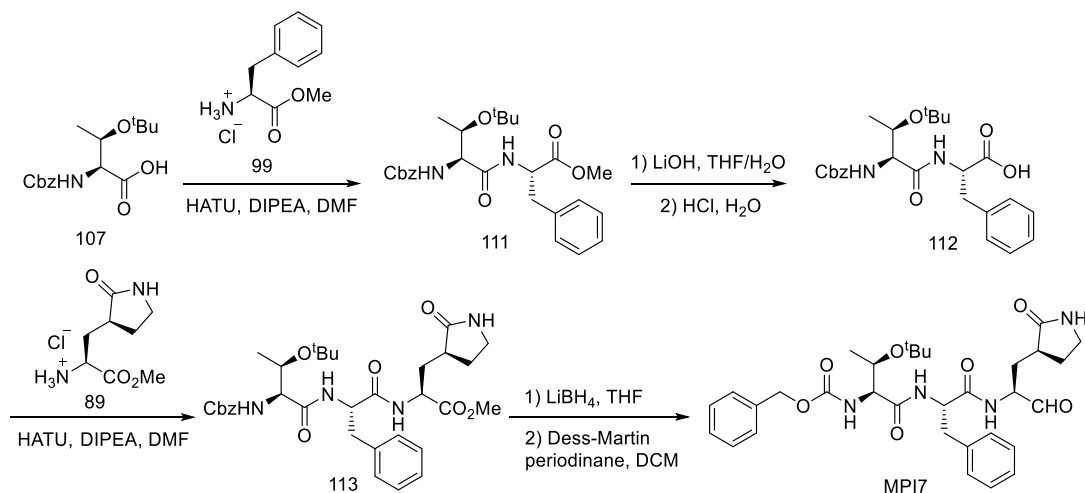
Methyl N-((benzyloxy)carbonyl)-O-(tert-butyl)-L-allothreonyl-L-phenylalaninate (111)

To a solution of **107** (5 mmol, 1.55 g) and **99** (5 mmol, 1.07g) in anhydrous DMF (20 mL) was added DIPEA (10 mmol, 1.29 g) and was cooled to 0 °C. HATU (5.5 mmol, 2.09 g) was added to the solution under 0 °C and then stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (100 mL) and washed with saturated NaHCO$_3$ solution (2×50 mL), 1 M HCl solution (2×50 mL), and saturated brine solution (2×50 mL) sequentially. The organic layer was dried over anhydrous Na$_2$SO$_4$ and then concentrated i*n vacuo*. The residue was then purified with flash chromatography (15-50% EtOAc in hexanes as the eluent) to afford **111** as colorless oil (1.64 g, 70%).

N-((benzyloxy)carbonyl)-O-(tert-butyl)-L-allothreonyl-L-phenylalanine (112)

111 (1 mmol, 470 mg) was dissolved in 5 mL of THF. A solution of LiOH·H$_2$O (2 mmol, 84 mg) in 5 mL H$_2$O was added to the solution. The mixture was stirred at room temperature overnight. Then THF was removed *in vacuo* and the aqueous layer was acidified with 1 M HCl and extracted with dichloromethane (3×10 mL). The organic layer was dried over anhydrous Na$_2$SO$_4$ and concentrated to yield **112** as white solid (330 mg, 72 %). ^{1}H NMR (400 MHz, CD$_3$OD): δ 7.29-7.43 (m, 5H), 7.11-7.29 (m, 5H), 5.11 (s, 2H), 4.71 (t, *J* = 6.2 Hz, 1H), 4.14 (d, *J* = 3.9 Hz, 1H), 3.98-4.07 (m, 1H), 3.17 (dd, *J* = 13.9, 5.6 Hz, 1H), 3.05 (dd, *J* = 13.8, 6.8 Hz, 1H), 0.98-1.16 (m, 12H).

Methyl (5S,8S,11S)-8-benzyl-5-((R)-1-(tert-butoxy)ethyl)-3,6,9-trioxo-11-(((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (113)

To a solution of **112** (0.4 mmol, 182 mg) and **89** (0.4 mmol, 88 mg) in anhydrous DMF (2 mL) was added DIPEA (0.8 mmol, 103 mg) and was cooled to 0 °C. HATU (0.44 mmol, 167 mg) was added to the solution under 0 °C and then stirred at room temperature overnight. The reaction mixture was then diluted with ethyl acetate (20 mL) and washed with saturated NaHCO$_3$ solution (2×10 mL), 1 M HCl solution (2×10 mL), and saturated brine solution (2×10 mL) sequentially. The organic layer was dried over anhydrous Na$_2$SO$_4$ and then concentrated *on vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **113** as white solid (180 mg, 72%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.44 – 7.19 (m, 10H), 5.90 (d, *J* = 5.4 Hz, 1H), 5.52 (s, 1H), 5.12 (q, *J* = 12.3 Hz, 2H), 4.77 (q, *J* = 6.8 Hz, 1H), 4.65 – 4.53 (m, 1H), 4.22 – 4.09 (m, 2H), 3.72 (s, 3H), 3.40 – 3.29 (m, 2H), 3.21 (dd, *J* = 14.0, 6.4

Hz, 1H), 3.09 (dd, J = 13.9, 6.1 Hz, 1H), 2.50 – 2.40 (m, 1H), 2.35 – 2.26 (m, 1H), 2.18 –

2.05 (m, 1H), 1.95 – 1.77 (m, 2H), 1.29 (s, 2H), 1.17 (s, 9H), 1.08 (d, J = 6.2 Hz, 3H).

Benzyl ((2S,3R)-3-(tert-butoxy)-1-oxo-1-(((S)-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)-3-phenylpropan-2-yl)amino)butan-2-yl)carbamate (MPI7)

To a solution of **113** (0.1 mmol, 62 mg) in anhydrous dichloromethane (5 mL) was

added a solution of LiBH$_4$ in anhydrous THF (2 M, 0.1 mL, 0.2 mmol) at 0 °C. The

resulting solution was stirred at the same temperature for 3 h. Then a saturated solution of

NH$_4$Cl (5 mL) was added dropwise to quench the reaction. The layers were separated, and

the organic layer was washed with saturated brine solution (2×10 mL), dried over

anhydrous Na$_2$SO$_4$. and evaporated to dryness. The residue was then dissolved in

anhydrous dichloromethane (5 mL) and cooled to 0 °C. Dess-Martin periodinane (0.2

mmol, 85 mg) was added to the solution. The reaction mixture was then stirred at room

temperature overnight. Then the reaction was quenched with a saturated NaHCO$_3$ solution

containing 10 % Na$_2$S$_2$O$_3$. The layers were separated. The organic layer was then washed

with saturated brine solution (2×10 mL), dried over anhydrous Na2SO4 and evaporated

on vacuo. The residue was then purified with flash chromatography (1-10% methanol in

dichloromethane as the eluent) to afford **MPI7** as white solid (27 mg, 45 %). ^{1}H NMR

(400MHz, CDCl$_3$): δ 9.28 (s, 1H), 7.80 (d, J = 6.5 Hz, 1H), 7.44 – 7.09 (m, 10H), 5.87 (d,

J = 5.6 Hz, 1H), 5.71 (s, 1H), 5.08 (q, J = 11.1 Hz, 2H), 4.76 (q, J = 7.1 Hz, 1H), 4.32 (q,

J = 7.0 Hz, 1H), 4.19 – 4.04 (m, 2H), 3.38 – 3.26 (m, 2H), 3.19 (dd, J = 13.8, 6.3 Hz, 1H),

3.07 (dd, J = 13.8, 6.8 Hz, 1H), 2.41 – 2.22 (m, 2H), 1.88 (t, J = 6.7 Hz, 2H), 1.84 – 1.73

(m, 1H), 1.30 – 1.22 (m, 1H), 1.16 (s, 9H), 1.05 (d, J = 6.1 Hz, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 199.60, 179.68, 171.55, 169.49, 156.32, 136.34, 136.28, 129.49, 128.87, 128.73, 128.40, 128.20, 127.35, 84.26, 75.61, 67.15, 66.80, 59.18, 57.74, 54.53, 40.53, 38.32, 37.82, 28.80, 28.23, 17.63. ESI-MS calcd for C$_{32}$H$_{43}$N$_4$O$_7$ (M+H$_+$): 595.3; found: 595.4.

Scheme 34. Synthesis of MPI8.
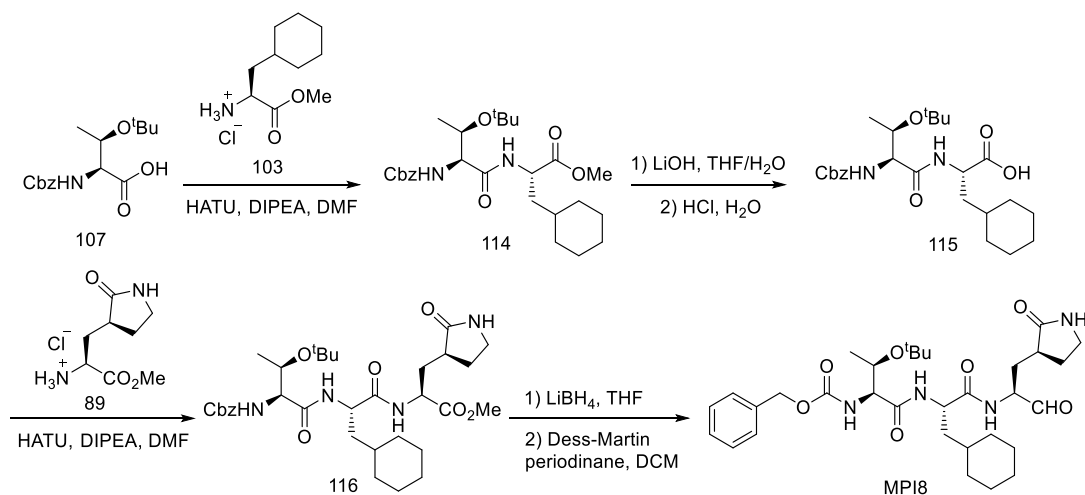
Methyl (S)-2-((2S,3R)-2-(((benzyloxy)carbonyl)amino)-3-(tert-butoxy)butanamido)-3-cyclohexylpropanoate (114).

To a solution of **107** (2 g, 6.46 mmol, 1.0 equiv), in anhydrous DMF (15 mL) at 0 °C, and then **103** (1.4 g, 6.46 mmol, 1.0 equiv), HATU (3.7 g, 9.70 mmol, 1.5 equiv) and DIPEA (5.7 mL, 32.3 mmol, 5.0 equiv) were added sequentially. The mixture was stirred at room temperature for 6 h. The mixture was diluted with EtOAc and washed with water, 1M HCl, sat. NaCl, dried over Na$_2$SO$_4$, and concentrated. The residue was purified by column chromatography (EtOAc: Hexane = 1:4 v/v) to afford the pure product **114** as colorless oil (2.5 g, 83%). ^{1}H NMR (400 MHz, CDCl$_3$-d) δ 7.6 (d, J = 7.8 Hz, 1H), 7.3 – 7.1 (m, 5H), 5.9 (d, J = 5.3 Hz, 1H), 5.1 – 4.9 (m, 2H), 4.4 (td, J = 8.4, 5.1 Hz, 1H), 4.2 –

4.1 (m, 2H), 3.6 (s, 3H), 1.7 – 1.5 (m, 6H), 1.5 (ddd, J = 14.2, 9.0, 5.7 Hz, 1H), 1.2 (s, 9H), 1.2 – 1.1 (m, 3H), 1.0 (d, J = 6.3 Hz, 4H), 0.9 – 0.7 (m, 2H). ^{13}C NMR (100 MHz, CDCl$_3$-d) δ 172.7, 169.2, 155.9, 136.2, 128.3, 128.3, 127.8, 127.7, 127.7, 75.2, 66.7, 66.5, 60.0, 51.8, 50.2, 39.5, 34.0, 33.3, 32.3, 28.0, 28.0, 28.0, 26.1, 26.0, 25.8, 20.7.

(S)-2-((2S,3R)-2-(((benzyloxy)carbonyl)amino)-3-(tert-butoxy)butanamido)-3-cyclohexylpropanoic acid (115).

To a solution of **114** (400 mg, 0.84 mmol, 1.0 equiv) in 1:1 THF/H$_2$O (8 mL) was added LiOH·H$_2$O (140 mg, 3.4 mmol, 4.0 equiv). The reaction was stirred at room temperature for 2 h. After completion, the reaction mixture was neutralized with 1M HCl solution and extracted with EtOAc. The organic layer was washed with sat. NaCl, dried over Na$_2$SO$_4$ and concentrated to afford the product **115** (330 mg, 85%) as a white solid. The residue was used in the next without further purification.

Methyl (5S,8S,11S)-5-((R)-1-(tert-butoxy)ethyl)-8-(cyclohexylmethyl)-3,6,9-trioxo-11-((((S)-2-oxopyrrolidin-3-yl)methyl)-1-phenyl-2-oxa-4,7,10-triazadodecan-12-oate (116).

To a solution of **115** (300 mg, 0.65 mmol, 1.0 equiv) in anhydrous DMF (5 mL) at 0 °C, and then **89** (144 mg, 0.65 mmol, 1.0 equiv), HATU (370 mg, 0.98 mmol, 1.5 equiv), DIPEA (580 μL, 3.25 mmol, 5.0 equiv) was added sequentially. The mixture was stirred at room temperature for 6 h. The mixture was diluted with EtOAc and washed with water, 1M HCl, sat. NaCl, dried over Na$_2$SO$_4$, and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:15 v/v) to afford the pure product **116** as a white solid (265 mg, 65%).

Benzyl ((2S,3R)-3-(tert-butoxy)-1-(((S)-3-cyclohexyl-1-oxo-1-(((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)amino)propan-2-yl)amino)-1-oxobutan-2-yl)carbamate (MPI8).

To a solution of **116** (250 mg, 0.40 mmol, 1.0 equiv) in anhydrous THF (10 mL) at 0 °C was added LiBH$_4$ (1.0 M in THF, 1.2 mL, 1.20 mmol, 3.0 equiv). The mixture was stirred at room temperature for 2 h. After the reaction was completed, excess reactants were consumed by slow addition of H$_2$O. The mixture was diluted with H$_2$O and extracted with EtOAc, washed with sat. NaCl, dried over Na$_2$SO$_4$ and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:15 v/v) to afford a white solid (160 mg, 66%). The solid (160 mg, 0.26 mmol, 1.0 equiv) was dissolved in anhydrous DCM (10 mL). Dess-Martin reagent (225 mg, 0.52 mmol, 2.0 equiv) was added slowly at 0 °C. Then the reaction mixture was stirred at room temperature for 1 h. A solution of NaHCO$_3$ and Na$_2$S$_2$O$_3$ was added to quench the reaction. After 10 min, the mixture was washed with water, sat. NaCl, dried over Na$_2$SO$_4$ and concentrated. The residue was purified by column chromatography (MeOH: DCM = 1:15 v/v) to afford **MPI8** as white solid (82 mg, 53%). ^{1}H NMR (400 MHz, CDCl$_3$-*d*) δ 9.5 (s, 1H), 8.1 (d, *J* = 6.8 Hz, 1H), 7.5 (q, *J* = 8.4, 7.6 Hz, 1H), 7.4 − 7.3 (m, 5H), 6.8 (s, 1H), 5.9 (d, *J* = 5.6 Hz, 1H), 5.1 (q, *J* = 12.1 Hz, 2H), 4.5 (td, *J* = 8.6, 5.6 Hz, 1H), 4.4 (q, *J* = 9.9, 7.2 Hz, 1H), 4.2 − 4.1 (m, 2H), 3.2 (p, *J* = 8.4, 7.1 Hz, 2H), 2.5 − 2.4 (m, 1H), 2.3 − 2.2 (m, 1H), 2.0 (ddt, *J* = 16.2, 11.3, 5.6 Hz, 1H), 1.8 (ddd, *J* = 13.3, 7.9, 4.2 Hz, 1H), 1.8 − 1.5 (m, 8H), 1.4 − 1.1 (m, 12H), 1.0 (d, *J* = 6.2 Hz, 4H), 1.0 − 0.8 (m, 2H). ^{13}C NMR (100 MHz, CDCl$_3$-*d*) δ 199.7, 180.0, 173.0, 169.8, 156.3, 136.2, 128.6, 128.6, 128.3, 128.1, 128.1, 75.4, 67.0, 66.8, 59.0,

57.3, 51.4, 40.6, 40.2, 37.9, 34.2, 34.2, 33.6, 32.7, 29.9, 28.3, 28.3, 28.3, 26.4, 26.2, 26.0, 17.4. ESI-MS calcd for $C_{32}H_{49}N_4O_7^+$ (M+H$^+$): 601.3; found 601.3.

Scheme 35. Synthesis of MPI9.
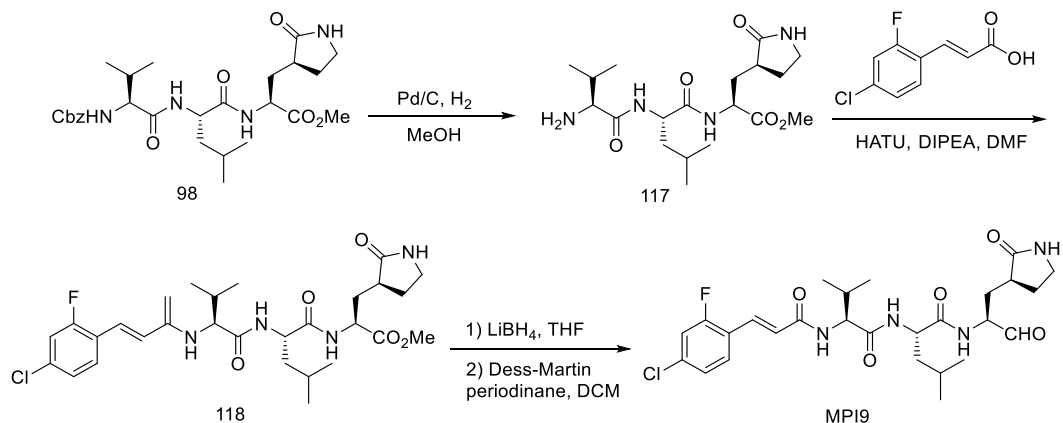
Methyl (S)-2-((S)-2-((S)-2-((E)-3-(4-chloro-2-fluorophenyl)acrylamido)-3-methylbutanamido)-4-methylpentanamido)-3-((S)-2-oxopyrrolidin-3-yl)propanoate (118)

To a solution of **98** (106 mg, 0.2 mmol) in methanol (5 mL) was added 10% Pd/C (21 mg). The mixture was then stirred with hydrogen balloon at room temperature for 3 h. The reaction mixture was then filtered, and the filtrate was concentrated *in vacuo* to afford **117** as colorless oil, which was used without further purification. To a solution of **117** in 2 mL anhydrous DMF was added 2-fluoro-4-chlorocinnamic acid (40 mg, 0.2 mmol) and DIPEA (0.4 mmol, 52 mg). The solution was cooled to 0 °C, follow by the addition of HATU (0.24 mmol, 91 mg). The solution was then allowed to warm up to room temperature and stirred overnight. The reaction mixture was then diluted with ethyl acetate (20 mL) and washed with saturated NaHCO$_3$ solution (2×10 mL), 1 M HCl solution (2×10 mL), and saturated brine solution (2×10 mL) sequentially. The organic layer was dried

141

over anhydrous Na_2SO_4 and then concentrated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in dichloromethane as the eluent) to afford **118** as white solid (75 mg, 65%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 8.46 (d, *J* = 8.2 Hz, 1H), 8.25 (d, *J* = 8.9 Hz, 1H), 8.11 (d, *J* = 7.7 Hz, 1H), 7.66 (dd, *J* = 17.7, 9.4 Hz, 2H), 7.54 (dd, *J* = 10.7, 2.1 Hz, 1H), 7.44 (d, *J* = 16.0 Hz, 1H), 7.38 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.00 (d, J = 16.0 Hz, 1H), 4.42 – 4.24 (m, 3H), 3.62 (s, 3H), 3.16 (t, *J* = 9.2 Hz, 1H), 3.06 (q, *J* = 9.3, 8.9 Hz, 1H), 2.36 – 2.27 (m, 1H), 2.12 – 1.95 (m, 3H), 1.62 (dtd, *J* = 20.8, 11.7, 10.8, 6.9 Hz, 3H), 1.46 (t, *J* = 7.3 Hz, 2H), 0.95 – 0.81 (m, 12H).

(S)-2-((S)-2-((E)-3-(4-chloro-2-fluorophenyl)acrylamido)-3-methylbutanamido)-4-methyl-N-((S)-1-oxo-3-((S)-2-oxopyrrolidin-3-yl)propan-2-yl)pentanamide (MPI9)

To a solution of **118** (0.1 mmol, 58 mg) in anhydrous tetrahydrofuran (5 mL) was dropwise added a solution of LiBH$_4$ in tetrahydrofuran (2 M, 0.1 mL, 0,2 mmol) at 0 °C. The solution was stirred under the same temperature for 3 h. Then a saturated solution of NH$_4$Cl (5 mL) was added dropwise to quench the reaction. The layers were separated, and the organic layer was washed with saturated brine solution (2×10 mL), dried over anhydrous Na_2SO_4. and evaporated to dryness. The residue was then dissolved in anhydrous dichloromethane (5 mL) and cooled to 0 °C. Dess-Martin periodinane (0.2 mmol, 85 mg) was added to the solution. The reaction mixture was then stirred at room temperature overnight. Then the reaction was quenched with a saturated NaHCO$_3$ solution containing 10% $Na_2S_2O_3$. The layers were separated. The organic layer was then washed with saturated brine solution (2×10 mL), dried over anhydrous Na_2SO_4 and evaporated *in vacuo*. The residue was then purified with flash chromatography (1-10% methanol in

dichloromethane as the eluent) to afford **MPI9** as white solid (32 mg, 58%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 9.41 (s, 1H), 8.46 (d, J = 7.9 Hz, 1H), 8.26 (d, J = 8.7 Hz, 1H), 8.17 (d, J = 7.6 Hz, 1H), 7.66 (t, J = 8.3 Hz, 1H), 7.62 (s, 1H), 7.53 (dd, J = 10.6, 2.1 Hz, 1H), 7.43 (d, J = 15.9 Hz, 1H), 7.37 (dd, J = 8.5, 2.0 Hz, 1H), 7.00 (d, J = 15.9 Hz, 1H), 4.42 – 4.18 (m, 3H), 3.16 (t, J = 9.1 Hz, 1H), 3.06 (td, J = 9.4, 7.1 Hz, 1H), 2.36 – 2.21 (m, 1H), 2.13 (dt, J = 14.0, 7.8 Hz, 1H), 2.01 (h, J = 6.6 Hz, 1H), 1.88 (ddd, J = 14.9, 11.4, 4.0 Hz, 1H), 1.70 – 1.56 (m, 3H), 1.55 – 1.39 (m, 2H), 0.97 – 0.73 (m, 12H); ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 201.24, 178.73, 173.08, 171.32, 165.00, 161.93, 159.41, 135.01, 134.90, 130.74, 130.63, 126.15, 125.84, 122.35, 122.23, 117.35, 117.09, 58.07, 56.54, 51.68, 41.08, 37.59, 31.34, 29.82, 27.75, 24.67, 23.30, 22.24, 19.59, 18.53. ESI-MS: calcd for C$_{27}$H$_{37}$ClFN$_4$O$_5$ (M+H$^+$): 551.2; found 551.4.

CHAPTER IV

SELECTIVITY OF MPIS BETWEEN SARS-COV-2 MAIN PROTEASE AND

HUMAN HOST PROTEASES*

4.1 Introduction

The COVID-19 pandemic has now already infected all seven continents, over 200 countries and over 100 million of people so far, among which more than two million people died. Since the identification of the pathogenic virus, SARS-CoV-2, intense efforts have been put into the research of the virus and the development of drugs and vaccines. There have been quite a few vaccines that showed good effectiveness being approved and distributed. Nevertheless, more and more reported virus variants with mutations on the spike protein of the virus could possibly diminish the effectiveness and expose human beings under threaten again in the future. In such a sense, the development of an antiviral drug against SARS-CoV-2 is as important as that of vaccines. Although there are still a lot to elucidate about SARS-CoV-2 biology and COVID-19 pathogenesis, previous studies of SARS-CoV-1 have demonstrated that the main protease (Mpro) is essential to viral replication and pathogenesis. Mpro has been validated as an antiviral drug target for SARS-CoV-2. Unlike Spike that is highly mutable, Mpro is highly conserved as

exemplified by the 96% sequence identity between SARS-CoV-1 Mpro and SARS-CoV-2 Mpro (SC2Mpro). Small-molecule medicines that potently inhibit SC2Mpro could be potentially effective treatment for COVID-19. A good number of inhibitors have been identified to inhibit SC2Mpro and some of them have shown promising *in vitro* and *in vivo* efficacy.

Selectivity between viral proteases and human proteases should always be paid enough attention in drug development. SC2Mpro is a cysteine protease. Its enzymatic activity relies on its Cys145 residue at the active site. Thus, much effort in SC2Mpro inhibitor development focused on covalent inhibitors with aldehyde and α-ketoamide warheads, which can form reversible covalent bond with Cys145 through nucleophilic addition. Nevertheless, these covalent warheads also render the inhibitors potentially able to form covalent bonds with human cysteine proteases or serine proteases through the same chemistry. Currently many SCM2pro covalent inhibitors are developed based on the substrate preference of SCM2pro and adopt a peptidomimetic structure, which also renders the inhibitors more likely to bind to and inhibit human proteases. However, there was also evidence showing certain human proteases function as preprocessor of SARS-CoV-2 spike proteins and facilitate virus entry, including cathepsin L and transmembrane protease, serine 2 (TMPRSS2)[62-64]. After SARS-CoV-2 is internalized into an endosome, cathepsin L potentiates its membrane fusion with the endosome for the release of the virus RNA genome into the host cytosol. Thus, cathepsin L is a potential antiviral drug target for SARS-CoV-2.[62-64] A dual inhibitor design strategy that selectively target both SCM2pro and other important human proteases associated with SARS-CoV-2 life cycle

145

such as cathepsin L would provide improved overall antiviral potency and efficacy. One potential concern is that these inhibitors can also target other similar host proteases leading to undesirable off-target effects by non-specific interactions. Although a number of SCM2pro inhibitors have been developed to combat COVID-19, little is known about their selectivity that is an important aspect for evaluating drugs. Herein, to explore the dual and selective inhibitors, we tested the *in vitro* inhibitory potency of 14 previously reported SCM2pro inhibitors (including 11 peptidomimetic aldehydes and 3 indole chloropyridinyl esters) against SCM2pro, cathepsin B/L/K, furin, and TMPRSS2. We also tested the intracellular inhibition of lysosome activities of 4 aldehyde inhibitors, which previously showed antiviral activity in cells and proposed their possible dual-inhibition mechanism of action against SCM2pro and cathepsin L.
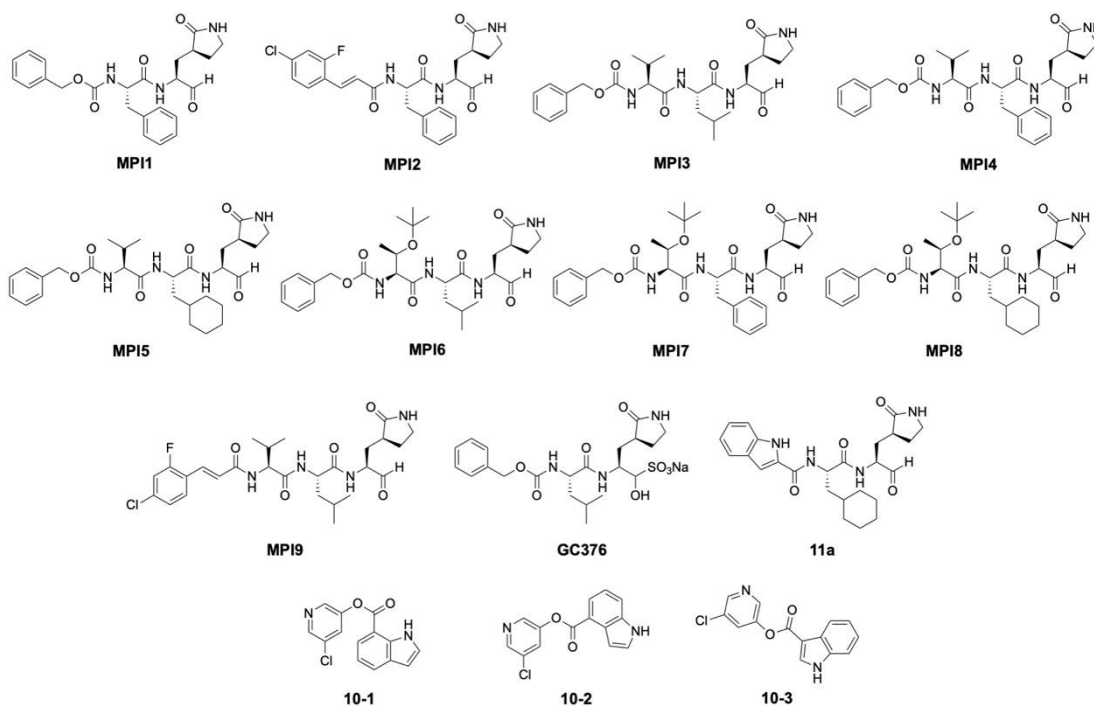
Figure 27. Structures of the tested compounds in this study, MPI1-9, 11a, GC376, 10-1, 10-2, and 10-3.

146

4.2 Methods and Results

4.2.1 Evaluation of Potency against Cathepsins and Other Host Proteases in vitro

In 2020 our group reported MPI1-MPI9 as a series of potent peptidyl aldehyde inhibitors for SC2Mpro[139], with the most potent one having a single-digit nanomolar IC_{50} value. In the following cytopathic effect (CPE) assay, two compounds (MPI5 and MPI8) showed low-micromolar EC_{50} values against SARS-CoV-2 virus in Vero E6 cells and submicromolar EC_{50} values in ACE2-expressing A549 cells, while other MPIs did not inhibit CPE in cells despite they have much lower IC_{50} values compared to MPI5 and MPI8. The higher antiviral efficacy of MPI5 and MPI8 in $ACE2^+$ A549 cells compared to that in Vero E6 cells implies that these inhibitors might also block ACE2-mediated viral entry mechanism in addition to SC2Mpro inhibition. As peptidyl aldehydes, MPIs are very likely to covalently inhibit human cysteine or serine proteases the same way as SC2Mpro. Some human proteases play an important role in facilitating viral entry, including cathepsin L and TMPRSS2. If MPIs can also target these human proteases, it is possible that the observed efficacy of SARS-CoV-2 inhibition in cells is a result of dual-inhibition of both SC2Mpro and protease-mediated SARS-CoV-2 entry.
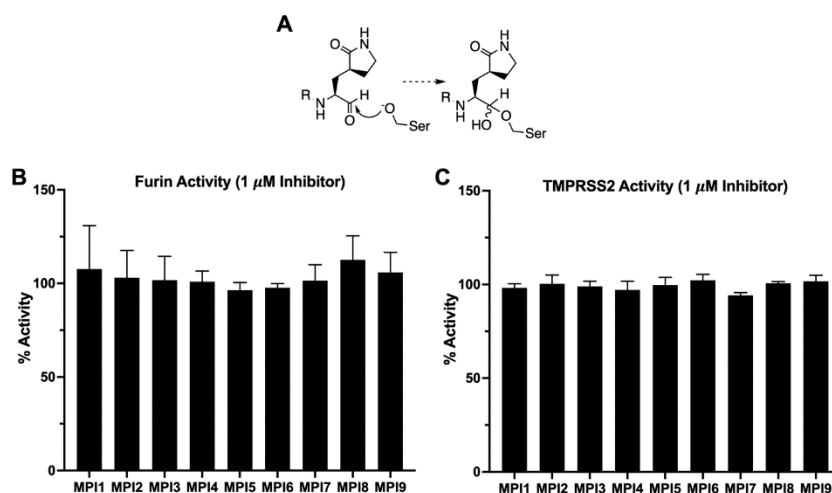
Figure 28. Peptidyl aldehydes inhibiting host serine proteases such as furin and TMPRSS2. **(A)** The mechanism of peptidyl aldehydes reacting with catalytical serine residue of serine protease. **(B)** Normalized furin activity after incubation with 1 µM MPI1-9 for 30 min. **(C)** Normalized TMPRSS2 activity after incubation with 1 µM MPI1-9 for 30 min.

Previously we reported a cellular assay by which we can specifically evaluate the potency of inhibitors against SC2Mpro in cells. We confirmed that all MPIs inhibited SC2Mpro activity in cells to different extents. Among them, MPI8 has the lowest cellular IC_{50} of 0.031 µM. As the next step to investigate the possibility of dual-target mechanism, 4e first determined the *in vitro* IC_{50} of MPI1-9 against cathepsin L, cathepsin B, cathepsin K, furin, and TMPRSS2 by a kinetic assay where the rate of the hydrolysis of a fluorogenic substrate by the proteases was measured. The protocols were adapted and modified from published literatures[19, 140, 141]. The substrates were chosen to be Z-Phe-Arg-AMC for cathepsin B/L/K, Pyr-Arg-Thr-Lys-Arg-AMC for furin and Boc-Gln-Ala-Arg-AMC for TMPRSS2. The substrate concentrations were 20 µM for cathepsin assays and 10 µM for furin and TMPRSS2 assays. The concentration of enzymes used in the assay was determined to be 2 nM for cathepsin L, 1 nM for cathepsin K, 5 nM for cathepsin B, 10

nM for furin and 1 μM for TMPRSS2, so that an optimal signal-to-noise ratio could be reached. The results of enzymatic inhibitory assays showed that all tested MPIs are also potent inhibitors for cathepsin L, cathepsin B and cathepsin K with IC_{50} as low as 80 pM (**Table 3**, **Figure 29**), while furin and TMPRSS2 are not inhibited at all even with 1 μM MPIs (**Figure 28**). Furthermore, there is dramatic difference between the potency of MPIs against different cathepsins: many MPIs showed a higher potency in the inhibition of cathepsin L than cathepsin B and cathepsin K. The selectivity between cathepsin L/K and cathepsin L/B ranges from 2.7 folds of MPI9 to 147 folds of MPI8, and 35 folds of MPI5 to 846 folds of MPI6. A few of them, including MPI4, MPI5, MPI7 and MPI8 showed moderate to strong selectivity to both cathepsin B and K (**Table 3**).
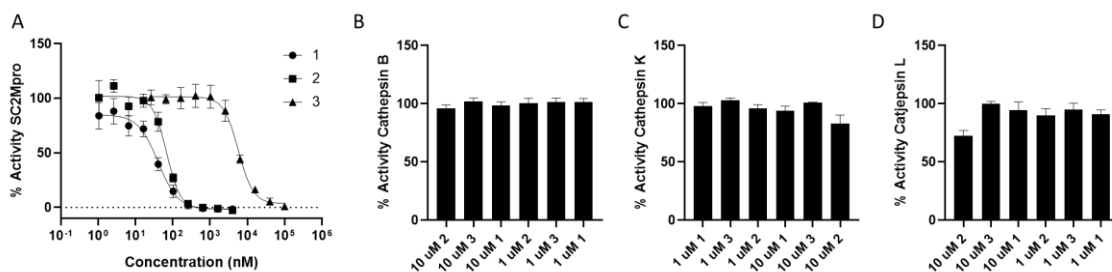
Figure 29. Indole chloropyridinyl esters showed strong inhibition against SC2Mpro but little inhibition against cathepsin L, cathepsin B or cathepsin K. (A) IC_{50} curves for **10-1**, **10-2**, and **10-3** against SC2Mpro. **(B)** Percentage of cathepsin B activity after treatment of **10-1**, **10-2**, and **10-3**. **(C)** Percentage of cathepsin K activity after treatment of **10-1**, **10-2**, and **10-3**. **(D)** Percentage of cathepsin L activity after treatment of **10-1**, **10-2**, and **10-3**.

Besides MPIs developed by our group, we also tested other reported SC2Mpro inhibitors, including **11a**[75], GC376[142] and three indole chloropyridinyl esters (**10-1**, **10-2** and **10-3**)[135]. GC376 and **11a** are two investigational drugs undergoing clinical trials for the treatment of COVID-19 patients in United States. They are also peptidyl aldehydes and share much structural similarity to our MPIs which showed sub-micromolar IC_{50}

against SC2Mpro. As for indole chloropyridinyl esters, we also measured their IC_{50}

against SC2Mpro. Despite of their relatively simple structures, compound **10-1** and **10-2**

exhibited IC_{50}s of 38 nM and 67 nM (**Figure 29**), which are similar to those of many

peptidyl aldehyde inhibitors, but **10-3** showed a much higher IC_{50} of 7.6 µM. We then

measured the IC_{50}s of these five compounds against cathepsin B/L/K, furin and

TMPRSS2. Similarly, all these compounds showed no inhibition against furin or

TMPRSS2 at 1 µM, but they acted dramatically differently with cathepsins. **11a** and

GC376 still turned out to be potent cathepsin B/L/K inhibitors with an IC_{50} of 0.14 nM of

GC376 and 0.23 nM of **11a**. However, their selectivity between cathepsin L and cathepsin

B/K are poorer compared to many MPIs (**Table 3**). But compound **10-1**, **10-2** and **10-3**

displayed no inhibition or minimal inhibition against cathepsins up to the concentration of

10 µM (**Figure 29**).

Table 3. IC_{50} of MPI1-9, 11a, and GC376 against cathepsin L, cathepsin B, and cathepsin K.
Parenthesized numbers indicate selective indices with respect to the activity toward cathepsin L.

Compound	Cathepsin L IC_{50} (nM)	Cathepsin B IC_{50} (nM)	Cathepsin K IC_{50} (nM)
MPI1	0.079 ± 0.029	25 ± 2 (320)	0.88 ± 0.12 (11)
MPI2	0.19 ± 0.03	4.1 ± 0.4 (22)	0.78 ± 0.06 (4.1)
MPI3	0.30 ± 0.07	112 ± 10 (370)	1.18 ± 0.18 (3.9)
MPI4	0.33 ± 0.04	46 ± 3 (139)	49 ± 16 (150)
MPI5	2.3 ± 1.9	80 ± 6 (35)	134 ± 18 (58)

MPI6	0.45 ± 0.08	380 ± 21 (840)	1.07 ± 0.21 (2.4)
MPI7	0.61 ± 0.15	64 ± 7 (105)	51 ± 6 (84)
MPI8	**1.2 ± 1.0**	**230 ± 20 (192)**	**180 ± 50 (150)**
MPI9	0.56 ± 0.06	147 ± 10 (262)	1.5 ± 0.2 (2.7)
11a	0.14 ± 0.04	4.9 ± 0.4 (35)	1.4 ± 0.2 (10)
GC376	0.23 ± 0.05	98 ± 5 (426)	0.35 ± 0.05 (1.5)
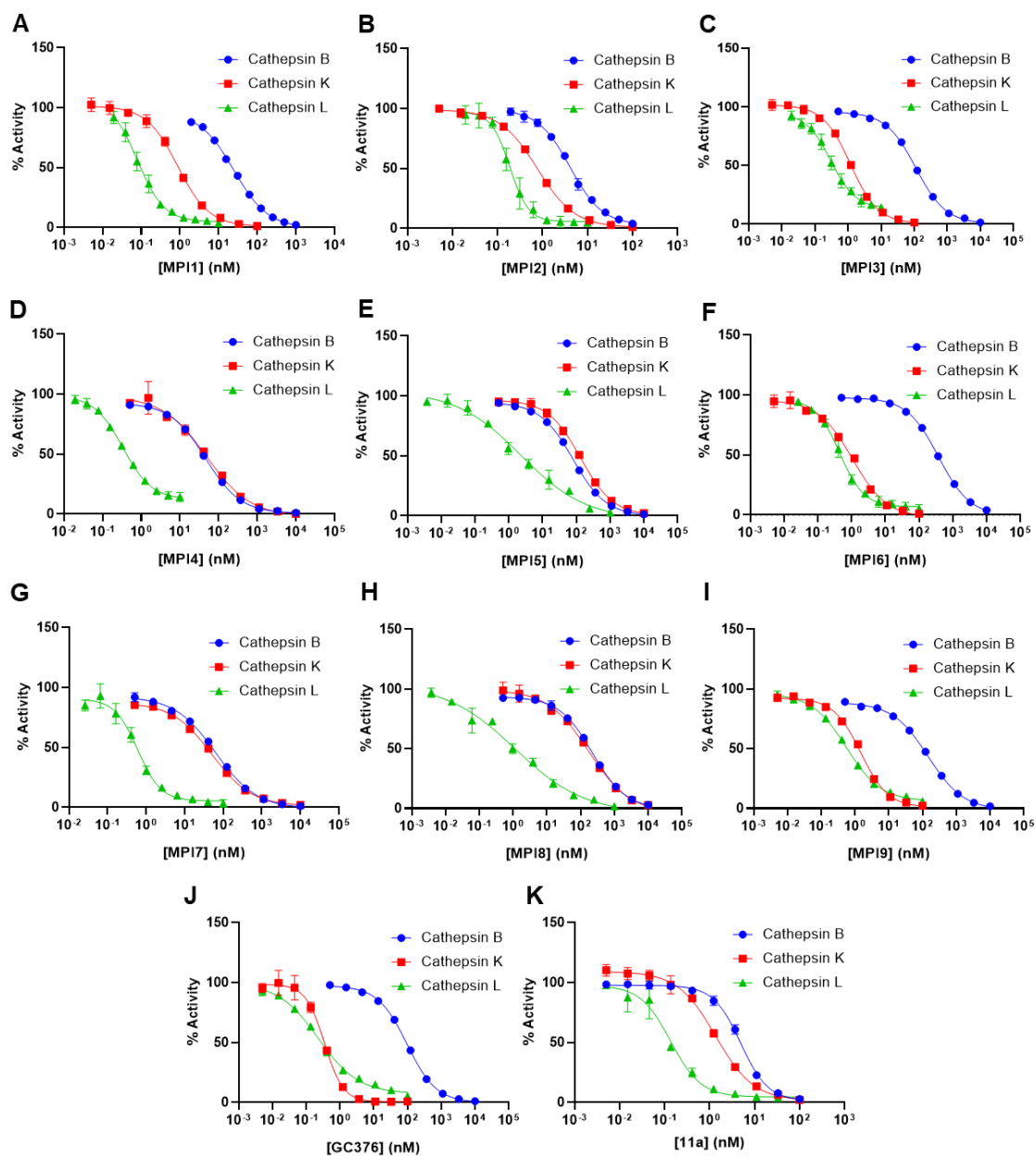
Figure 30. IC$_{50}$ curves of tested peptidyl aldehydes against cathepsin B, cathepsin K, cathepsin L. (**A**) MPI1. (**B**) MPI2. (**C**) MPI3. (**D**) MPI4. (**E**) MPI5. (**F**) MPI6. (**G**) MPI7. (**H**) MPI8. (**I**) MPI9. (**J**) GC376. (**K**) 11a.

4.2.2 Evaluation of Lysosomal Activity Inhibition in Cells

152

Given the evidence reported earlier that cathepsin L plays key roles in the entry of SARS-CoV-2, it is very likely that these peptidyl aldehydes can block SARS-CoV-2 entry through cathepsin L inhibition in cells. However, when the potency of cathepsin L inhibition is compared, MPI5 and MPI8 are still the least potent compounds among these peptidyl aldehydes, leaving it questionable whether MPI5 and MPI8 are genuinely inhibiting cathepsin L activity, or more generally, lysosomal or endosomal activities in cells as they do in vitro. To confirm that, we further used HEK293T cells to perform intracellular lysosomal activity assay on MPI5, MPI8, 11a and GC376, the four inhibitors that showed SARS-CoV-2 inhibition efficacy in the cellular level. The cells were first treated with the inhibitors. Then a fluorogenic substrate was used to assess the lysosomal activity of the treated cells. Lysosomotropic agent bafilomycin A1, which inhibit lysosomal activity, was used as control. The results showed that both MPI5, MPI8, 11a and GC376 inhibit cellular lysosomal activity at a concentration as low as 2.5 µM (**Figure 31**), strongly suggesting that these compounds inhibit SARS-CoV-2 replication in cells via a dual-target mechanism.
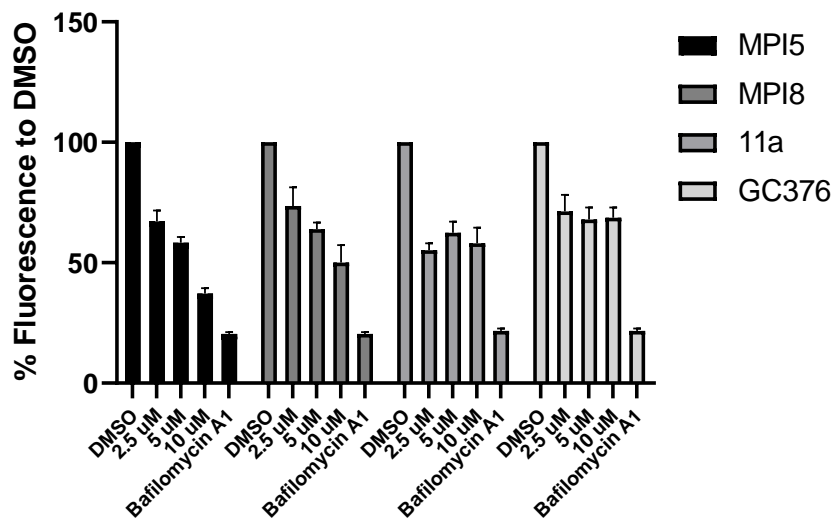
Figure 31. Inhibition of cellular lysosomal activity by MPI5, MPI8, 11a, and GC376. The intensity of cellular fluorescence indicated how much fluorogenic substrate was degraded by lysosomes, representing cellular lysosomal activity. For each compound, three concentrations (10 µM, 5 µM, and 2.5 µM) were used and the fluorescence signals were normalized to that of DMSO group.

4.3 Discussion and Conclusion

4.3.1 Discussion

It is quite intriguing that all peptidyl aldehyde inhibitors we tested showed remarkable potency in the inhibition of cathepsin B/L/K *in vitro*, especially cathepsin L. This observation might not be pure coincidence but have underlying connections with the substrate specificity of cathepsin L and SC2Mpro. According to SC2Mpro crystal structure[67], glutamine residue or γ-lactam residue are considered the best fit for its S1 pocket, which overlaps with the substrate specificity of cathepsin L in which glutamine and glycine are the most favored residues at P1 position[143]. At P2 position, SC2Mpro

favors leucine the most and accommodate many other hydrophobic residues as well[80]. This also overlaps with P2 specificity of cathepsin L of aliphatic residues. The overlapped substrate specificity between these two proteases makes peptidomimetic inhibitors designed based on the substrate specificity of one of them naturally more likely to inhibit the other one as well. This is also implying that a non-peptidomimetic inhibitor should be more selective between SC2Mpro and cathepsin L.We have also demonstrated MPI5, MPI8, **11a** and GC376 can inhibit intracellular lysosomal activity, which unveiled more complicated pharmacology of these peptidyl aldehyde compounds in cells. From one perspective, by inhibiting the activity of cathepsin L and lysosomes in cells, these Mpro inhibitors can also block SARS-CoV-2 entry, providing them with a dual mechanism of action other than mono-specific inhibition of SC2Mpro activity. One obvious advantage of the dual-target mechanism is the less susceptibility to mutation of virus and consequent drug resistance. However, from the other perspective, the inclusion of a human protease as immediate drug targets also means a higher chance of adversary effects and may significantly lower their therapeutic indexes, limiting their potential as an anti-SARS-CoV-2 therapy. As far as we know, cathepsin L is the only enzyme in the cathepsin family that showed clear relevance to SARS-CoV-2 entry. This makes it particularly a concern for those inhibitors that showed relatively poor selectivity between different cathepsins. Among the four compounds that showed anti-lysosomal activity in cells, GC376 has barely any selectivity between cathepsin K and cathepsin L, **11a** and MPI5 have some moderate selectivity and MPI8 showed to be the most selective one despite its relatively

155

low potency, which implies that MPI8 might be a better candidate for the dual-target inhibition of SARS-CoV-2.

On the other hand, chloropyridinyl esters **10-1** and **10-2** exhibited excellent selectivity between SC2Mpro and the human proteases we tested. Nevertheless, despite of their good *in vitro* potency against SC2Mpro, these inhibitors are very likely to suffer from poor *in vivo* stability due to the chemical nature of esters. Further study into their binding mode with SC2Mpro may provide more information to structural modifications leading to more potent and stable selective SC2Mpro inhibitors.

4.3.2 Conclusion

In this study, we tested the in vitro IC_{50} against cathepsin B, cathepsin L, cathepsin K, furin, and TMPRSS2 of 11 peptidyl aldehyde and 3 chloropyridinyl ester SC2Mpro inhibitors. All peptidyl aldehydes were found to be potent cathepsin inhibitors, especially potent against cathepsin L. All the four peptidyl aldehyde compounds that can inhibit SARS-CoV-2 replication in cells (MPI5, MPI8, 11a, and GC376) have also shown inhibition of lysosomal activity in HEK293T cells, implying these inhibitors have a dual-target mechanism of action. Among them, MPI8 has the lowest cellular SC2Mpro IC_{50} and a good selectivity between cathepsin L and cathepsin B/K, making it a better candidate for the dual-target inhibition purpose. In contrast, chloropyridinyl esters 1 and 2 showed excellent selectivity between SC2Mpro and the human proteases tested, which could be a starting point for the development of selective SC2Mpro inhibitors.

4.4 Experimental Details

4.4.1 Inhibition assay for cathepsin L

The assay was performed in the following assay buffer: 100 mM MES-NaOH solution (pH 5.5) containing 2.5 mM EDTA, 2.5 mM DTT and 9% DMSO. The stock solution of the enzyme was diluted to 20 nM with assay buffer containing 2.5 mM DTT. Stock solutions of inhibitors were prepared in DMSO. A 10 mM stock solution of the fluorogenic substrate Z-Phe-Arg-AMC was diluted with assay buffer. The final concentration in enzymatic assay of DMSO was 10%, and those of the substrate and cathepsin L was 20 µM and 2 nM, respectively. Into a well containing 39 µL assay buffer, 1 µL inhibitor solution (or DMSO) and 10 µL diluted solution of cathepsin L were added and mixed thoroughly, and then incubated at 37 ℃ for 30 min. The reaction was initiated by adding 50 µL diluted solution of substrate and the fluorescence intensity at 440 nm under 360 nm excitation was measured. Experiments were performed in triplicate with at least ten different concentration inhibitor and both positive and negative controls. The initial rate was calculated according to the fluorescent intensity in the first five minutes by linear regression, which was then normalized according to the initial rate of positive and negative controls. IC_{50} curve was simulated by GraphPad 8.0 using sigmoidal model (four parameters).

4.4.2 Inhibition assay for cathepsin B

The assay was performed in the following assay buffer: 100 mM MES-NaOH solution (pH 6.0) containing 2.5 mM EDTA, 2.5 mM DTT, 0.001% Tween 20 and 9% DMSO. The stock solution of the enzyme was diluted to 50 nM with assay buffer containing 2.5 mM DTT. Stock solutions of inhibitors were prepared in DMSO. A 10 mM stock solution of the fluorogenic substrate Z-Phe-Arg-AMC was diluted with assay buffer. The final concentration in enzymatic assay of DMSO was 10%, and those of the substrate and cathepsin B was 20 µM and 5 nM, respectively. Into a well containing 39 µL assay buffer, 1 µL inhibitor solution (or DMSO) and 10 µL diluted solution of cathepsin B were added and mixed thoroughly, and then incubated at 37 ℃ for 30 min. The reaction was initiated by adding 50 µL diluted solution of substrate and the fluorescence intensity at 440 nm under 360 nm excitation was measured. Experiments were performed in triplicate with at least ten different concentration inhibitor and both positive and negative controls. The initial rate was calculated according to the fluorescent intensity in the first five minutes by linear regression, which was then normalized according to the initial rate of positive and negative controls. IC_{50} curve was simulated by GraphPad 8.0 using sigmoidal model (four parameters).

4.4.3 Inhibition assay for cathepsin K

The assay was performed in the following assay buffer: 100 mM MES-NaOH solution (pH 5.5) containing 2.5 mM EDTA, 2.5 mM DTT and 9% DMSO. The stock solution of the enzyme was diluted to 10 nM with assay buffer containing 2.5 mM DTT. Stock solutions of inhibitors were prepared in DMSO. A 10 mM stock solution of the

fluorogenic substrate Z-Phe-Arg-AMC was diluted with assay buffer. The final concentration in enzymatic assay of DMSO was 10%, and those of the substrate and cathepsin K was 20 µM and 1 nM, respectively. Into a well containing 39 µL assay buffer, 1 µL inhibitor solution (or DMSO) and 10 µL diluted solution of cathepsin K were added and mixed thoroughly, and then incubated at 37 °C for 30 min. The reaction was initiated by adding 50 µL diluted solution of substrate and the fluorescence intensity at 440 nm under 360 nm excitation was measured. Experiments were performed in triplicate with at least ten different concentration inhibitor and both positive and negative controls. The initial rate was calculated according to the fluorescent intensity in the first five minutes by linear regression, which was then normalized according to the initial rate of positive and negative controls. IC_{50} curve was simulated by GraphPad 8.0 using sigmoidal model (four parameters).

4.4.4 Inhibition assay for furin

The assay was performed in the following assay buffer: 100 mM HEPES buffer (pH 7.0) containing 0.2% Triton X-100, 2 mM $CaCl_2$, 0.02% sodium azide, and 1 mg/mL BSA. The stock solution of the enzyme was diluted to 100 nM with assay buffer. Stock solutions of inhibitors were prepared in DMSO. A 10 mM stock solution of the fluorogenic substrate Pyr-Arg-Thr-Lys-Arg-AMC was diluted with assay buffer. The final concentration in enzymatic assay of the substrate and furin was 10 µM and 10 nM, respectively. Into a well containing 39 µL assay buffer, 1 µL inhibitor solution (or DMSO) and 10 µL diluted solution of furin were added and mixed thoroughly, and then incubated

at 37 ℃ for 30 min. The reaction was initiated by adding 50 µL diluted solution of substrate and the fluorescence intensity at 440 nm under 360 nm excitation was measured. Experiments were performed in triplicate with at least ten different concentration inhibitor and both positive and negative controls. The initial rate was calculated according to the fluorescent intensity in the first five minutes by linear regression, which was then normalized according to the initial rate of positive and negative controls. IC_{50} curve was simulated by GraphPad 8.0 using sigmoidal model (four parameters).

4.4.5 Inhibition assay for TMPRSS2

The assay was performed in the following assay buffer: 50 mM Tris (pH 8.0) containing 150 mM NaCl, and 0.01% Tween20. The stock solution of the enzyme was diluted to 10 µM with assay buffer. Stock solutions of inhibitors were prepared in DMSO. A 10 mM stock solution of the fluorogenic substrate Boc-Gln-Ala-Arg-AMC was diluted with assay buffer. The final concentration in enzymatic assay of the substrate and TMPRSS2 was 10 µM and 1 µM, respectively. Into a well containing 39 µL assay buffer, 1 µL inhibitor solution (or DMSO) and 10 µL diluted solution of TMPRSS2 were added and mixed thoroughly, and then incubated at 37 ℃ for 30 min. The reaction was initiated by adding 50 µL diluted solution of substrate and the fluorescence intensity at 440 nm under 360 nm excitation was measured. Experiments were performed in triplicate with at least ten different concentration inhibitor and both positive and negative controls. The initial rate was calculated according to the fluorescent intensity in the first five minutes by linear regression, which was then normalized according to the initial rate of positive and

negative controls. IC_{50} curve was simulated by GraphPad 8.0 using sigmoidal model (four parameters).

4.4.6 Intracellular lysosomal activity assay

The protocol was adapted from that provided by Abcam with some modifications. HEK293T cells were grown in DMEM media containing 10 % FBS in standard 24-well plate and incubated under 37 °C, 5 % CO_2 overnight. Media was then removed and replaced with fresh DMEM media containing different concentrations of test compounds with 0.1 % DMSO (experimental group), 0.1 % DMSO (negative control group) or 1 × bafilomycin A1 (positive control group). The cells were incubated under 37 °C, 5 % CO_2 for 2 hours. The media was then removed and replaced with fresh DMEM media containing 0.5 % FBS and the same concentration of test compounds, DMSO or bafilomycin. 15 µL of self-quenched fluorogenic substrate was added to each well per 1 mL of media. The cells were incubated under 37 °C, 5 % CO_2 for another 2 hours. The media was then removed, and the cells were harvested, washed with 1 mL ice cold assay buffer (provided in assay kit) twice and resuspended in PBS buffer for flow cytometry analysis under 488 nm excitation. Experiments were performed in triplicate.

CHAPTER V

CONCLUSIONS AND PROSPECTIVES

This dissertation presents the results of applying structure-based drug design strategies in the development of human ENL YEATS domain inhibitors and SARS-CoV-2 main protease inhibitors. Both projects afforded several potent inhibitors against respective targets, once more proving the importance and great potential of utilizing structural information of drug target in early-stage drug discovery.

For human ENL YEATS domain, we have successfully developed a series of potent inhibitors that are able to effectively block the recognition of H3K9 acetylation by ENL. One of them displayed good efficacy in MLL-rearranged leukemia cells, in terms of both the inhibition of cell proliferation and the downregulation of oncogene transcription. However, the biological processes where ENL are involved in MLL-rearranged leukemia are still not totally clear. A lot of research is still needed to further elucidate the functions ENL play in the leukemogenesis process. In such a sense, our developed ENL YEATS domain inhibitors could also serve as a template for the development of different chemical probes and be used in the research of ENL-related cellular processes. These studies will in return provide medicinal chemists with more useful information about ENL, MLL-rearranged leukemias, and related proteins and biological processes, which could be very helpful for the development of more potent and specific drugs for leukemias and other related diseases.

For SARS-CoV-2 main protease, we have also successfully developed a series of potent SARS-CoV-2 main protease inhibitors. Among them, MPI8 showed excellent SARS-CoV-2 main protease inhibition potency both *in vitro* and in human cell lines. It also showed good efficacy in inhibiting cytopathic effects in SARS-CoV-2 infected mammalian cells. Moreover, MPI8 is also a cathepsin L inhibitor, which enables it to block the major cell entry pathway of SARS-CoV-2 while still maintain a good selectivity among other host proteases. Collectively, MPI8 is a very promising drug candidate for the treatment of COVID-19. Given the still severe situation of the global COVID-19 pandemic, especially the difficulty in the manufacturing, transportation, distribution, and storage of COVID-19 vaccines in less developed areas of the world, an easy-to-produce, shelf-stable, and preferably orally bioavailable antiviral agent for SARS-CoV-2 is in great demand. Our developed MPI8, along with many other drug candidates that have been or are being developed, provides a hope to control and contain, if not to defeat the COVID-19 pandemic in the future.

REFERENCES

[1]   Anderson, A.C., The Process of Structure-Based Drug Design. *Chemistry & Biology*, **2003**, 10 (9), 787-797.

[2]   Lu, H., Q. Zhou, J. He, Z. Jiang, C. Peng, R. Tong, and J. Shi, Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduction Targeted Ther.*, **2020**, 5 (1), 213.

[3]   Winters, A.C. and K.M. Bernt, MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Frontiers in Pediatrics*, **2017**, 5 (4).

[4]   Tkachuk, D.C., S. Kohler, and M.L. Cleary, Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell*, **1992**, 71 (4), 691-700.

[5]   Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, and E. Canaani, The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. *Cell*, **1992**, 71 (4), 701-708.

[6]   Yokoyama, A., T.C.P. Somervaille, K.S. Smith, O. Rozenblatt-Rosen, M. Meyerson, and M.L. Cleary, The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. *Cell*, **2005**, 123 (2), 207-218.

[7]   Hsieh, J.J.D., P. Ernst, H. Erdjument-Bromage, P. Tempst, and S.J. Korsmeyer, Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol. Cell. Biol.*, **2003**, 23 (1), 186-194.

[8]   Yokoyama, A., Molecular mechanisms of MLL-associated leukemia. *Int. J. Hematol.*, **2015**, 101 (4), 352-361.

[9]   Zeleznik-Le, N.J., A.M. Harden, and J.D. Rowley, 11q23 translocations split the AT-hook cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proc. Natl. Acad. Sci. U. S. A.*, **1994**, 91 (22), 10610.

[10]   Meyer, C., T. Burmeister, D. Gröger, G. Tsaur, L. Fechina, A. Renneville, R. Sutton, N.C. Venn, M. Emerenciano, M.S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal, V.H.J. van der Velden, E. Delabesse, T.S. Park, S.H. Oh, M.L.M. Silva, T. Lund-Aho, V. Juvonen, A.S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenkova, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. zur Stadt, J. Lentes, G. Göhring, A. Kustanovich, O. Aleinikova, B.W. Schäfer, S. Kubetzko, H.O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J.M. Cayuela, E. Clappier, C.N. Alonso, C.M. Zwaan, M.M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Möricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O.A. Haas, R. Panzer-Grümayer, L. Sedék, T.

Szczepański, A. Caye, L. Suarez, H. Cavé, and R. Marschalek, The MLL recombinome of acute leukemias in 2017. *Leukemia*, **2018**, 32 (2), 273-284.

[11] Deshpande, A.J., J. Bradner, and S.A. Armstrong, Chromatin modifications as therapeutic targets in MLL-rearranged leukemia. *Trends in Immunology*, **2012**, 33 (11), 563-570.

[12] Jude, C.D., L. Climer, D. Xu, E. Artinger, J.K. Fisher, and P. Ernst, Unique and Independent Roles for MLL in Adult Hematopoietic Stem Cells and Progenitors. *Cell Stem Cell*, **2007**, 1 (3), 324-337.

[13] Kroon, E., J. Krosl, U. Thorsteinsdottir, S. Baban, A.M. Buchberg, and G. Sauvageau, Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *The EMBO Journal*, **1998**, 17 (13), 3714-3725.

[14] Krivtsov, A.V., D. Twomey, Z. Feng, M.C. Stubbs, Y. Wang, J. Faber, J.E. Levine, J. Wang, W.C. Hahn, D.G. Gilliland, T.R. Golub, and S.A. Armstrong, Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9. *Nature*, **2006**, 442 (7104), 818-822.

[15] Yokoyama, A., M. Lin, A. Naresh, I. Kitabayashi, and M.L. Cleary, A Higher-Order Complex Containing AF4 and ENL Family Proteins with P-TEFb Facilitates Oncogenic and Physiologic MLL-Dependent Transcription. *Cancer Cell*, **2010**, 17 (2), 198-212.

[16] Okuda, H., A. Kanai, S. Ito, H. Matsui, and A. Yokoyama, AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-dependent transcription. *Nat. Commun.*, **2015**, 6 (1), 8869.

[17] Shilatifard, A., W.S. Lane, K.W. Jackson, R.C. Conaway, and J.W. Conaway, An RNA Polymerase II Elongation Factor Encoded by the Human ELL Gene. *Science*, **1996**, 271 (5257), 1873.

[18] Peterlin, B.M. and D.H. Price, Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, **2006**, 23 (3), 297-305.

[19] Becker, G.L., F. Sielaff, M.E. Than, I. Lindberg, S. Routhier, R. Day, Y. Lu, W. Garten, and T. Steinmetzer, Potent Inhibitors of Furin and Furin-like Proprotein Convertases Containing Decarboxylated P1 Arginine Mimetics. *J. Med. Chem.*, **2010**, 53 (3), 1067-1075.

[20] Mohan, M., H.M. Herz, Y.H. Takahashi, C. Lin, K.C. Lai, Y. Zhang, M.P. Washburn, L. Florens, and A. Shilatifard, Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev.*, **2010**, 24 (6), 574-89.

[21] Okada, Y., Q. Feng, Y. Lin, Q. Jiang, Y. Li, V.M. Coffield, L. Su, G. Xu, and Y. Zhang, hDOT1L Links Histone Methylation to Leukemogenesis. *Cell*, **2005**, 121 (2), 167-178.

[22] Daigle, Scott R., Edward J. Olhava, Carly A. Therkelsen, Christina R. Majer, Christopher J. Sneeringer, J. Song, L.D. Johnston, Margaret P. Scott, Jesse J. Smith, Y. Xiao, L. Jin, Kevin W. Kuntz, R. Chesworth, Mikel P. Moyer, Kathrin M. Bernt, J.-C. Tseng, Andrew L. Kung, Scott A. Armstrong, Robert A. Copeland, Victoria M. Richon, and Roy M. Pollock, Selective Killing of Mixed

Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. *Cancer Cell*, **2011**, 20 (1), 53-65.

[23]   Mueller, D., C. Bach, D. Zeisig, M.-P. Garcia-Cuellar, S. Monroe, A. Sreekumar, R. Zhou, A. Nesvizhskii, A. Chinnaiyan, J.L. Hess, and R.K. Slany, A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, **2007**, 110 (13), 4445-4454.

[24]   Biswas, D., T.A. Milne, V. Basrur, J. Kim, K.S.J. Elenitoba-Johnson, C.D. Allis, and R.G. Roeder, Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proc. Natl. Acad. Sci. U. S. A.*, **2011**, 108 (38), 15751.

[25]   Nguyen, A.T., O. Taranova, J. He, and Y. Zhang, DOT1L, the H3K79 methyltransferase, is required for MLL-AF9–mediated leukemogenesis. *Blood*, **2011**, 117 (25), 6912-6922.

[26]   Bernt, Kathrin M., N. Zhu, Amit U. Sinha, S. Vempati, J. Faber, Andrei V. Krivtsov, Z. Feng, N. Punt, A. Daigle, L. Bullinger, Roy M. Pollock, Victoria M. Richon, Andrew L. Kung, and Scott A. Armstrong, MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer Cell*, **2011**, 20 (1), 66-78.

[27]   Jo, S.Y., E.M. Granowicz, I. Maillard, D. Thomas, and J.L. Hess, Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. *Blood*, **2011**, 117 (18), 4759-4768.

[28]   Lavau, C., C. Du, M. Thirman, and N. Zeleznik-Le, Chromatin-related properties of CBP fused to MLL generate a myelodysplastic-like syndrome that evolves into myeloid leukemia. *The EMBO Journal*, **2000**, 19 (17), 4655-4664.

[29]   So, C.W. and M.L. Cleary, MLL-AFX Requires the Transcriptional Effector Domains of AFX To Transform Myeloid Progenitors and Transdominantly Interfere with Forkhead Protein Function. *Mol. Cell. Biol.*, **2002**, 22 (18), 6542.

[30]   Goto, N.K., T. Zor, M. Martinez-Yamout, H.J. Dyson, and P.E. Wright, Cooperativity in Transcription Factor Binding to the Coactivator CREB-binding Protein (CBP): THE MIXED LINEAGE LEUKEMIA PROTEIN (MLL) ACTIVATION DOMAIN BINDS TO AN ALLOSTERIC SITE ON THE KIX DOMAIN*. *J. Biol. Chem.*, **2002**, 277 (45), 43168-43174.

[31]   Ernst, P., J. Wang, M. Huang, R.H. Goodman, and S.J. Korsmeyer, MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein. *Mol. Cell. Biol.*, **2001**, 21 (7), 2249.

[32]   Miyamoto, R., H. Okuda, A. Kanai, S. Takahashi, T. Kawamura, H. Matsui, T. Kitamura, I. Kitabayashi, T. Inaba, and A. Yokoyama, Activation of CpG-Rich Promoters Mediated by MLL Drives MOZ-Rearranged Leukemia. *Cell Reports*, **2020**, 32 (13), 108200.

[33]   Eguchi, M., M. Eguchi-Ishimae, and M. Greaves, The small oligomerization domain of gephyrin converts MLL to an oncogene. *Blood*, **2004**, 103 (10), 3876-3882.

[34]   Ono, R., H. Nakajima, K. Ozaki, H. Kumagai, T. Kawashima, T. Taki, T. Kitamura, Y. Hayashi, and T. Nosaka, Dimerization of MLL fusion proteins and

FLT3 activation synergize to induce multiple-lineage leukemogenesis. *The Journal of Clinical Investigation*, **2005**, 115 (4), 919-929.

[35] Liedtke, M., P.M. Ayton, T.C.P. Somervaille, K.S. Smith, and M.L. Cleary, Self-association mediated by the Ras association 1 domain of AF6 activates the oncogenic potential of MLL-AF6. *Blood*, **2010**, 116 (1), 63-70.

[36] Martin, M.E., T.A. Milne, S. Bloyer, K. Galoian, W. Shen, D. Gibbs, H.W. Brock, R. Slany, and J.L. Hess, Dimerization of MLL fusion proteins immortalizes hematopoietic cells. *Cancer Cell*, **2003**, 4 (3), 197-207.

[37] Deshpande, A.J., L. Chen, M. Fazio, A.U. Sinha, K.M. Bernt, D. Banka, S. Dias, J. Chang, E.J. Olhava, S.R. Daigle, V.M. Richon, R.M. Pollock, and S.A. Armstrong, Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood*, **2013**, 121 (13), 2533-2541.

[38] Grembecka, J., S. He, A. Shi, T. Purohit, A.G. Muntean, R.J. Sorenson, H.D. Showalter, M.J. Murai, A.M. Belcher, T. Hartley, J.L. Hess, and T. Cierpicki, Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nat. Chem. Biol.*, **2012**, 8 (3), 277-284.

[39] Krivtsov, A.V., K. Evans, J.Y. Gadrey, B.K. Eschle, C. Hatton, H.J. Uckelmann, K.N. Ross, F. Perner, S.N. Olsen, T. Pritchard, L. McDermott, C.D. Jones, D. Jing, A. Braytee, D. Chacon, E. Earley, B.M. McKeever, D. Claremon, A.J. Gifford, H.J. Lee, B.A. Teicher, J.E. Pimanda, D. Beck, J.A. Perry, M.A. Smith, G.M. McGeehan, R.B. Lock, and S.A. Armstrong, A Menin-MLL Inhibitor Induces Specific Chromatin Changes and Eradicates Disease in Models of MLL-Rearranged Leukemia. *Cancer Cell*, **2019**, 36 (6), 660-673.e11.

[40] Daigle, S.R., E.J. Olhava, C.A. Therkelsen, A. Basavapathruni, L. Jin, P.A. Boriack-Sjodin, C.J. Allain, C.R. Klaus, A. Raimondi, M.P. Scott, N.J. Waters, R. Chesworth, M.P. Moyer, R.A. Copeland, V.M. Richon, and R.M. Pollock, Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood*, **2013**, 122 (6), 1017-1025.

[41] Stein, E.M., G. Garcia-Manero, D.A. Rizzieri, R. Tibes, J.G. Berdeja, M.R. Savona, M. Jongen-Lavrenic, J.K. Altman, B. Thomson, S.J. Blakemore, S.R. Daigle, N.J. Waters, A.B. Suttle, A. Clawson, R. Pollock, A. Krivtsov, S.A. Armstrong, J. DiMartino, E. Hedrick, B. Löwenberg, and M.S. Tallman, The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood*, **2018**, 131 (24), 2661-2669.

[42] Stauffer, F., A. Weiss, C. Scheufler, H. Möbitz, C. Ragot, K.S. Beyer, K. Calkins, D. Guthy, M. Kiffe, B. Van Eerdenbrugh, R. Tiedt, and C. Gaul, New Potent DOT1L Inhibitors for in Vivo Evaluation in Mouse. *ACS Med. Chem. Lett.*, **2019**, 10 (12), 1655-1660.

[43] Zuber, J., J. Shi, E. Wang, A.R. Rappaport, H. Herrmann, E.A. Sison, D. Magoon, J. Qi, K. Blatt, M. Wunderlich, M.J. Taylor, C. Johns, A. Chicas, J.C. Mulloy, S.C. Kogan, P. Brown, P. Valent, J.E. Bradner, S.W. Lowe, and C.R. Vakoc, RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*, **2011**, 478 (7370), 524-528.

[44] Dawson, M.A., R.K. Prinjha, A. Dittmann, G. Giotopoulos, M. Bantscheff, W.-I. Chan, S.C. Robson, C.-w. Chung, C. Hopf, M.M. Savitski, C. Huthmacher, E. Gudgin, D. Lugo, S. Beinke, T.D. Chapman, E.J. Roberts, P.E. Soden, K.R. Auger, O. Mirguet, K. Doehner, R. Delwel, A.K. Burnett, P. Jeffrey, G. Drewes, K. Lee, B.J.P. Huntly, and T. Kouzarides, Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, **2011**, 478 (7370), 529-533.

[45] Picaud, S., D. Da Costa, A. Thanasopoulou, P. Filippakopoulos, P.V. Fish, M. Philpott, O. Fedorov, P. Brennan, M.E. Bunnage, D.R. Owen, J.E. Bradner, P. Taniere, B. Sullivan, S. Müller, J. Schwaller, T. Stankovic, and S. Knapp, PFI-1, a Highly Selective Protein Interaction Inhibitor, Targeting BET Bromodomains. *Cancer Research*, **2013**, 73 (11), 3336.

[46] Picaud, S., O. Fedorov, A. Thanasopoulou, K. Leonards, K. Jones, J. Meier, H. Olzscha, O. Monteiro, S. Martin, M. Philpott, A. Tumber, P. Filippakopoulos, C. Yapp, C. Wells, K.H. Che, A. Bannister, S. Robson, U. Kumar, N. Parr, K. Lee, D. Lugo, P. Jeffrey, S. Taylor, M.L. Vecellio, C. Bountra, P.E. Brennan, A. Mahony, S. Velichko, S. Müller, D. Hay, D.L. Daniels, M. Urh, N.B. La Thangue, T. Kouzarides, R. Prinjha, J. Schwaller, and S. Knapp, Generation of a Selective Small Molecule Inhibitor of the CBP/p300 Bromodomain for Leukemia Therapy. *Cancer Research*, **2015**, 75 (23), 5106.

[47] Lasko, L.M., C.G. Jakob, R.P. Edalji, W. Qiu, D. Montgomery, E.L. Digiammarino, T.M. Hansen, R.M. Risi, R. Frey, V. Manaves, B. Shaw, M. Algire, P. Hessler, L.T. Lam, T. Uziel, E. Faivre, D. Ferguson, F.G. Buchanan, R.L. Martin, M. Torrent, G.G. Chiang, K. Karukurichi, J.W. Langston, B.T. Weinert, C. Choudhary, P. de Vries, A.F. Kluge, M.A. Patane, J.H. Van Drie, C. Wang, D. McElligott, E. Kesicki, R. Marmorstein, C. Sun, P.A. Cole, S.H. Rosenberg, M.R. Michaelides, A. Lai, and K.D. Bromberg, Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature*, **2017**, 550 (7674), 128-132.

[48] Harris, William J., X. Huang, James T. Lynch, Gary J. Spencer, James R. Hitchin, Y. Li, F. Ciceri, Julian G. Blaser, Brigit F. Greystoke, Allan M. Jordan, Crispin J. Miller, Donald J. Ogilvie, and Tim C.P. Somervaille, The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer Cell*, **2012**, 21 (4), 473-487.

[49] Feng, Z., Y. Yao, C. Zhou, F. Chen, F. Wu, L. Wei, W. Liu, S. Dong, M. Redell, Q. Mo, and Y. Song, Pharmacological inhibition of LSD1 for the treatment of MLL-rearranged leukemia. *J. Hematol. Oncol.*, **2016**, 9 (1), 24.

[50] Maes, T., C. Mascaró, I. Tirapu, A. Estiarte, F. Ciceri, S. Lunardi, N. Guibourt, A. Perdones, M.M.P. Lufino, T.C.P. Somervaille, D.H. Wiseman, C. Duy, A. Melnick, C. Willekens, A. Ortega, M. Martinell, N. Valls, G. Kurz, M. Fyfe, J.C. Castro-Palomino, and C. Buesa, ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment of Acute Leukemia. *Cancer Cell*, **2018**, 33 (3), 495-511.e12.

[51]   Cusan, M., S.F. Cai, H.P. Mohammad, A. Krivtsov, A. Chramiec, E. Loizou, M.D. Witkin, K.N. Smitheman, D.G. Tenen, M. Ye, B. Will, U. Steidl, R.G. Kruger, R.L. Levine, H.Y. Rienhoff, Jr., R.P. Koche, and S.A. Armstrong, LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBPα-dependent enhancers in AML. *Blood*, **2018**, 131 (15), 1730-1742.

[52]   Benito, Juliana M., L. Godfrey, K. Kojima, L. Hogdal, M. Wunderlich, H. Geng, I. Marzo, Karine G. Harutyunyan, L. Golfman, P. North, J. Kerry, E. Ballabio, Triona N. Chonghaile, O. Gonzalo, Y. Qiu, I. Jeremias, L. Debose, E. O'Brien, H. Ma, P. Zhou, R. Jacamo, E. Park, Kevin R. Coombes, N. Zhang, Deborah A. Thomas, S. O'Brien, Hagop M. Kantarjian, Joel D. Leverson, Steven M. Kornblau, M. Andreeff, M. Müschen, Patrick A. Zweidler-McKay, James C. Mulloy, A. Letai, Thomas A. Milne, and M. Konopleva, MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199. *Cell Reports*, **2015**, 13 (12), 2715-2727.

[53]   Stam, R.W., M.L. den Boer, P. Schneider, P. Nollau, M. Horstmann, H.B. Beverloo, E. van der Voort, M.G. Valsecchi, P. de Lorenzo, S.E. Sallan, S.A. Armstrong, and R. Pieters, Targeting FLT3 in primary MLL-gene–rearranged infant acute lymphoblastic leukemia. *Blood*, **2005**, 106 (7), 2484-2490.

[54]   Shi, A., M.J. Murai, S. He, G. Lund, T. Hartley, T. Purohit, G. Reddy, M. Chruszcz, J. Grembecka, and T. Cierpicki, Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood*, **2012**, 120 (23), 4461-4469.

[55]   Filippakopoulos, P., J. Qi, S. Picaud, Y. Shen, W.B. Smith, O. Fedorov, E.M. Morse, T. Keates, T.T. Hickman, I. Felletar, M. Philpott, S. Munro, M.R. McKeown, Y. Wang, A.L. Christie, N. West, M.J. Cameron, B. Schwartz, T.D. Heightman, N. La Thangue, C.A. French, O. Wiest, A.L. Kung, S. Knapp, and J.E. Bradner, Selective inhibition of BET bromodomains. *Nature*, **2010**, 468 (7327), 1067-1073.

[56]   Nicodeme, E., K.L. Jeffrey, U. Schaefer, S. Beinke, S. Dewell, C.-W. Chung, R. Chandwani, I. Marazzi, P. Wilson, H. Coste, J. White, J. Kirilovsky, C.M. Rice, J.M. Lora, R.K. Prinjha, K. Lee, and A. Tarakhovsky, Suppression of inflammation by a synthetic histone mimic. *Nature*, **2010**, 468 (7327), 1119-1123.

[57]   Fish, P.V., P. Filippakopoulos, G. Bish, P.E. Brennan, M.E. Bunnage, A.S. Cook, O. Federov, B.S. Gerstenberger, H. Jones, S. Knapp, B. Marsden, K. Nocka, D.R. Owen, M. Philpott, S. Picaud, M.J. Primiano, M.J. Ralph, N. Sciammetta, and J.D. Trzupek, Identification of a Chemical Probe for Bromo and Extra C-Terminal Bromodomain Inhibition through Optimization of a Fragment-Derived Hit. *J. Med. Chem.*, **2012**, 55 (22), 9831-9837.

[58]   Alqahtani, A., K. Choucair, M. Ashraf, D.M. Hammouda, A. Alloghbi, T. Khan, N. Senzer, and J. Nemunaitis, Bromodomain and extra-terminal motif inhibitors: a review of preclinical and clinical advances in cancer therapy. *Future Science OA*, **2019**, 5 (3), FSO372.

[59] Cochran, A.G., A.R. Conery, and R.J. Sims, Bromodomains: a new target class for drug development. *Nat. Rev. Drug Discovery*, **2019**, 18 (8), 609-628.

[60] Popp, T.A., C. Tallant, C. Rogers, O. Fedorov, P.E. Brennan, S. Müller, S. Knapp, and F. Bracher, Development of Selective CBP/P300 Benzoxazepine Bromodomain Inhibitors. *J. Med. Chem.*, **2016**, 59 (19), 8889-8912.

[61] Naqvi, A.A.T., K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, S.M. Atif, G. Hariprasad, G.M. Hasan, and M.I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, **2020**, 1866 (10), 165878.

[62] Hoffmann, M., H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M.A. Müller, C. Drosten, and S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, **2020**, 181 (2), 271-280.e8.

[63] Shang, J., Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, and F. Li, Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.*, **2020**, 117 (21), 11727-11734.

[64] Ou, X., Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, and Z. Qian, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.*, **2020**, 11 (1), 1620.

[65] Yang, H., W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, and Z. Rao, Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLOS Biology*, **2005**, 3 (10), e324.

[66] Shin, D., R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, L. Schulz, M. Widera, A.R. Mehdipour, G. Tascher, P.P. Geurink, A. Wilhelm, G.J. van der Heden van Noort, H. Ovaa, S. Müller, K.-P. Knobeloch, K. Rajalingam, B.A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, and I. Dikic, Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, **2020**, 587 (7835), 657-662.

[67] Zhang, L., D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, **2020**, 368 (6489), 409-412.

[68] Osipiuk, J., S.-A. Azizi, S. Dvorkin, M. Endres, R. Jedrzejczak, K.A. Jones, S. Kang, R.S. Kathayat, Y. Kim, V.G. Lisnyak, S.L. Maki, V. Nicolaescu, C.A. Taylor, C. Tesar, Y.-A. Zhang, Z. Zhou, G. Randall, K. Michalska, S.A. Snyder, B.C. Dickinson, and A. Joachimiak, Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.*, **2021**, 12 (1), 743.

[69] Imbert, I., J.-C. Guillemot, J.-M. Bourhis, C. Bussetta, B. Coutard, M.-P. Egloff, F. Ferron, A.E. Gorbalenya, and B. Canard, A second, non-canonical RNA-

dependent RNA polymerase in SARS Coronavirus. *The EMBO Journal*, **2006**, 25 (20), 4933-4942.

[70] Subissi, L., C.C. Posthuma, A. Collet, J.C. Zevenhoven-Dobbe, A.E. Gorbalenya, E. Decroly, E.J. Snijder, B. Canard, and I. Imbert, One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U. S. A.*, **2014**, 111 (37), E3900-E3909.

[71] Gao, Y., L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, and Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, **2020**, 368 (6492), 779-782.

[72] Yan, R., Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, **2020**, 367 (6485), 1444-1448.

[73] Shang, J., G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li, Structural basis of receptor recognition by SARS-CoV-2. *Nature*, **2020**, 581 (7807), 221-224.

[74] Morse, J.S., T. Lalonde, S. Xu, and W.R. Liu, Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem*, **2020**, 21 (5), 730-738.

[75] Dai, W., B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, and H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*, **2020**, 368 (6497), 1331-1335.

[76] Sacco, M.D., C. Ma, P. Lagarias, A. Gao, J.A. Townsend, X. Meng, P. Dube, X. Zhang, Y. Hu, N. Kitamura, B. Hurst, B. Tarbet, M.T. Marty, A. Kolocouris, Y. Xiang, Y. Chen, and J. Wang, Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Science Advances*, **2020**, 6 (50), eabe0751.

[77] Chuck, C.P., C. Chen, Z.H. Ke, D.C.C. Wan, H.F. Chow, and K.B. Wong, Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. *Eur. J. Med. Chem.*, **2013**, 59, 1-6.

[78] Zhang, L.L., D.Z. Lin, Y. Kusov, Y. Nian, Q.J. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E.J. Snijder, H. Liu, and R. Hilgenfeld, alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.*, **2020**, 63 (9), 4562-4578.

[79] Wu, C.Y., K.Y. King, C.J. Kuo, J.M. Fang, Y.T. Wu, M.Y. Ho, C.L. Liao, J.J. Shie, P.H. Liang, and C.H. Wong, Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3CL protease. *Chemistry & Biology*, **2006**, 13 (3), 261-268.

[80] Rut, W., K. Groborz, L. Zhang, X. Sun, M. Zmudzinski, B. Pawlik, X. Wang, D. Jochmans, J. Neyts, W. Młynarski, R. Hilgenfeld, and M. Drag, SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.*, **2021**, 17 (2), 222-228.

[81] Jin, Z., X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, and H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, **2020**, 582 (7811), 289-293.

[82] Ghahremanpour, M.M., J. Tirado-Rives, M. Deshmukh, J.A. Ippolito, C.-H. Zhang, I. Cabeza de Vaca, M.-E. Liosi, K.S. Anderson, and W.L. Jorgensen, Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.*, **2020**, 11 (12), 2526-2533.

[83] Jo, S., S. Kim, D.Y. Kim, M.S. Kim, and D.H. Shin, Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro. *J. Enzyme. Inhib. Med. Chem.*, **2020**, 35 (1), 1539-1544.

[84] Su, H., S. Yao, W. Zhao, M. Li, J. Liu, W. Shang, H. Xie, C. Ke, M. Gao, K. Yu, H. Liu, J. Shen, W. Tang, L. Zhang, J. Zuo, H. Jiang, F. Bai, Y. Wu, Y. Ye, and Y. Xu, Discovery of baicalin and baicalein as novel, natural product inhibitors of SARS-CoV-2 3CL protease in vitro. *bioRxiv*, **2020**.

[85] Jenuwein, T. and C.D. Allis, Translating the histone code. *Science*, **2001**, 293 (5532), 1074-80.

[86] Strahl, B.D. and C.D. Allis, The language of covalent histone modifications. *Nature*, **2000**, 403 (6765), 41-5.

[87] Dhalluin, C., J.E. Carlson, L. Zeng, C. He, A.K. Aggarwal, and M.M. Zhou, Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, **1999**, 399 (6735), 491-6.

[88] Andrews, F.H., S.A. Shinsky, E.K. Shanle, J.B. Bridgers, A. Gest, I.K. Tsun, K. Krajewski, X.B. Shi, B.D. Strahl, and T.G. Kutateladze, The Taf14 YEATS domain is a reader of histone crotonylation. *Nat. Chem. Biol.*, **2016**, 12 (6), 396-U33.

[89] Hsu, C.C., J. Shi, C. Yuan, D. Zhao, S. Jiang, J. Lyu, X. Wang, H. Li, H. Wen, W. Li, and X. Shi, Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer. *Genes Dev.*, **2018**, 32 (1), 58-69.

[90] Hsu, C.C., D. Zhao, J. Shi, D. Peng, H. Guan, Y. Li, Y. Huang, H. Wen, W. Li, H. Li, and X. Shi, Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discovery*, **2018**, 4, 28.

[91] Klein, B.J., S. Ahmad, K.R. Vann, F.H. Andrews, Z.A. Mayo, G. Bourriquen, J.B. Bridgers, J. Zhang, B.D. Strahl, J. Cote, and T.G. Kutateladze, Yaf9 subunit of the NuA4 and SWR1 complexes targets histone H3K27ac through its YEATS domain. *Nucleic Acids Research*, **2018**, 46 (1), 421-430.

[92] Li, Y.Y., B.R. Sabari, T. Panchenko, H. Wen, D. Zhao, H.P. Guan, L.L. Wan, H. Huang, Z.Y. Tang, Y.M. Zhao, R.G. Roeder, X.B. Shi, C.D. Allis, and H.T. Li,

Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Molecular Cell*, **2016**, 62 (2), 181-193.

[93]   Li, Y., H. Wen, Y. Xi, K. Tanaka, H. Wang, D. Peng, Y. Ren, Q. Jin, S.Y.R. Dent, W. Li, H. Li, and X. Shi, AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. *Cell*, **2014**, 159 (3), 558-571.

[94]   Mi, W., H. Guan, J. Lyu, D. Zhao, Y. Xi, S. Jiang, F.H. Andrews, X. Wang, M. Gagea, H. Wen, L. Tora, S.Y.R. Dent, T.G. Kutateladze, W. Li, H. Li, and X. Shi, YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nat. Commun.*, **2017**, 8 (1), 1088.

[95]   Shanle, E.K., F.H. Andrews, H. Meriesh, S.L. McDaniel, R. Dronamraju, J.V. DiFiore, D. Jha, G.G. Wozniak, J.B. Bridgers, J.L. Kerschner, K. Krajewski, G.M. Martin, A.J. Morrison, T.G. Kutateladze, and B.D. Strahl, Association of Taf14 with acetylated histone H3 directs gene transcription and the DNA damage response. *Genes Dev.*, **2015**, 29 (17), 1795-800.

[96]   Wan, L.L., H. Wen, Y.Y. Li, J. Lyu, Y.X. Xi, T. Hoshii, J.K. Joseph, X.L. Wang, Y.H.E. Loh, M.A. Erb, A.L. Souza, J.E. Bradner, L. Shen, W. Li, H.T. Li, C.D. Allis, S.A. Armstrong, and X.B. Shi, ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature*, **2017**, 543 (7644), 265-+.

[97]   Zhang, Q., L. Zeng, C.C. Zhao, Y. Ju, T. Konuma, and M.M. Zhou, Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure*, **2016**, 24 (9), 1606-1612.

[98]   Zhao, D., H. Guan, S. Zhao, W. Mi, H. Wen, Y. Li, Y. Zhao, C.D. Allis, X. Shi, and H. Li, YEATS2 is a selective histone crotonylation reader. *Cell Research*, **2016**, 26 (5), 629-32.

[99]   Le Masson, I., D.Y. Yu, K. Jensen, A. Chevalier, R. Courbeyrette, Y. Boulard, M.M. Smith, and C. Mann, Yaf9, a novel NuA4 histone acetyltransferase subunit, is required for the cellular response to spindle stress in yeast. *Mol. Cell. Biol.*, **2003**, 23 (17), 6086-102.

[100]   Schulze, J.M., A.Y. Wang, and M.S. Kobor, YEATS domain proteins: a diverse family with many links to chromatin modification and transcriptionThis paper is one of a selection of papers published in this Special Issue, entitled CSBMCB's 51st Annual Meeting – Epigenetics and Chromatin Dynamics, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.*, **2009**, 87 (1), 65-75.

[101]   Schulze, J.M., A.Y. Wang, and M.S. Kobor, Reading chromatin Insights from yeast into YEATS domain structure and function. *Epigenetics*, **2010**, 5 (7), 573-577.

[102]   He, N., C.K. Chan, B. Sobhian, S. Chou, Y. Xue, M. Liu, T. Alber, M. Benkirane, and Q. Zhou, Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proc. Natl. Acad. Sci. U. S. A.*, **2011**, 108 (36), E636-45.

[103]   Erb, M.A., T.G. Scott, B.E. Li, H.F. Xie, J. Paulk, H.S. Seo, A. Souza, J.M. Roberts, S. Dastjerdi, D.L. Buckley, N.E. Sanjana, O. Shalem, B. Nabet, R. Zeid, N.K. Offei-Addo, S. Dhe-Paganon, F. Zhang, S.H. Orkin, G.E. Winter, and J.E.

Bradner, Transcription control by the ENL YEATS domain in acute leukaemia. *Nature*, **2017**, 543 (7644), 270-+.

[104] Gadd, S., V. Huff, A.L. Walz, A. Ooms, A.E. Armstrong, D.S. Gerhard, M.A. Smith, J.M.G. Auvil, D. Meerzaman, Q.R. Chen, C.H. Hsu, C. Yan, C. Nguyen, Y. Hu, L.C. Hermida, T. Davidsen, P. Gesuwan, Y. Ma, Z. Zong, A.J. Mungall, R.A. Moore, M.A. Marra, J.S. Dome, C.G. Mulligan, J. Ma, D.A. Wheeler, O.A. Hampton, N. Ross, J.M. Gastier-Foster, S.T. Arold, and E.J. Perlman, A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.*, **2017**, 49 (10), 1487-1494.

[105] Perlman, E.J., S. Gadd, S.T. Arold, A. Radhakrishnan, D.S. Gerhard, L. Jennings, V. Huff, J.M. Guidry Auvil, T.M. Davidsen, J.S. Dome, D. Meerzaman, C.H. Hsu, C. Nguyen, J. Anderson, Y. Ma, A.J. Mungall, R.A. Moore, M.A. Marra, C.G. Mulligan, J. Ma, D.A. Wheeler, O.A. Hampton, J.M. Gastier-Foster, N. Ross, and M.A. Smith, MLLT1 YEATS domain mutations in clinically distinctive Favourable Histology Wilms tumours. *Nat. Commun.*, **2015**, 6, 10013.

[106] Wan, L., S. Chong, F. Xuan, A. Liang, X. Cui, L. Gates, T.S. Carroll, Y. Li, L. Feng, G. Chen, S.P. Wang, M.V. Ortiz, S.K. Daley, X. Wang, H. Xuan, A. Kentsis, T.W. Muir, R.G. Roeder, H. Li, W. Li, R. Tjian, H. Wen, and C.D. Allis, Impaired cell fate through gain-of-function mutations in a chromatin reader. *Nature*, **2020**, 577 (7788), 121-126.

[107] Asiaban, J.N., N. Milosevich, E. Chen, T.R. Bishop, J. Wang, Y. Zhang, C.J. Ackerman, E.N. Hampton, T.S. Young, M.V. Hull, B.F. Cravatt, and M.A. Erb, Cell-Based Ligand Discovery for the ENL YEATS Domain. *ACS Chem. Biol.*, **2020**, 15 (4), 895-903.

[108] Christott, T., J. Bennett, C. Coxon, O. Monteiro, C. Giroud, V. Beke, S.L. Felce, V. Gamble, C. Gileadi, G. Poda, R. Al-Awar, G. Farnie, and O. Fedorov, Discovery of a Selective Inhibitor for the YEATS Domains of ENL/AF9. *SLAS Discov.*, **2019**, 24 (2), 133-141.

[109] Heidenreich, D., M. Moustakim, J. Schmidt, D. Merk, P.E. Brennan, O. Fedorov, A. Chaikuad, and S. Knapp, Structure-Based Approach toward Identification of Inhibitory Fragments for Eleven-Nineteen-Leukemia Protein (ENL). *J. Med. Chem.*, **2018**, 61 (23), 10929-10934.

[110] Jiang, Y., G. Chen, X.M. Li, S. Liu, G. Tian, Y. Li, X. Li, H. Li, and X.D. Li, Selective Targeting of AF9 YEATS Domain by Cyclopeptide Inhibitors with Preorganized Conformation. *J. Am. Chem. Soc.*, **2020**, 142 (51), 21450-21459.

[111] Li, X., X.M. Li, Y. Jiang, Z. Liu, Y. Cui, K.Y. Fung, S.H.E. van der Beelen, G. Tian, L. Wan, X. Shi, C.D. Allis, H. Li, Y. Li, and X.D. Li, Structure-guided development of YEATS domain inhibitors by targeting pi-pi-pi stacking. *Nat. Chem. Biol.*, **2018**, 14 (12), 1140-1149.

[112] Moustakim, M., T. Christott, O.P. Monteiro, J. Bennett, C. Giroud, J. Ward, C.M. Rogers, P. Smith, I. Panagakou, L. Díaz-Sáez, S.L. Felce, V. Gamble, C. Gileadi, N. Halidi, D. Heidenreich, A. Chaikuad, S. Knapp, K.V.M. Huber, G. Farnie, J. Heer, N. Manevski, G. Poda, R. Al-awar, D.J. Dixon, P.E. Brennan, and O.

Fedorov, Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angew. Chem., Int. Ed.*, **2018**, 57 (50), 16302-16307.

[113] Ni, X., D. Heidenreich, T. Christott, J. Bennett, M. Moustakim, P.E. Brennan, O. Fedorov, S. Knapp, and A. Chaikuad, Structural Insights into Interaction Mechanisms of Alternative Piperazine-urea YEATS Domain Binders in MLLT1. *ACS Med. Chem. Lett.*, **2019**, 10 (12), 1661-1666.

[114] Di Veroli, G.Y., C. Fornari, D. Wang, S. Mollard, J.L. Bramhall, F.M. Richards, and D.I. Jodrell, Combenefit: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics*, **2016**, 32 (18), 2866-8.

[115] Fischer, U., P. Meltzer, and E. Meese, Twelve amplified and expressed genes localized in a single domain in glioma. *Hum. Genet.*, **1996**, 98 (5), 625-8.

[116] Pikor, L.A., W.W. Lockwood, K.L. Thu, E.A. Vucic, R. Chari, A.F. Gazdar, S. Lam, and W.L. Lam, YEATS4 is a novel oncogene amplified in non-small cell lung cancer that regulates the p53 pathway. *Cancer Research*, **2013**, 73 (24), 7301-12.

[117] Zimmermann, K., K. Ahrens, S. Matthes, J.M. Buerstedde, W.H. Stratling, and L. Phi-van, Targeted disruption of the GAS41 gene encoding a putative transcription factor indicates that GAS41 is essential for cell viability. *J. Biol. Chem.*, **2002**, 277 (21), 18626-31.

[118] Klein, B.J., L. Piao, Y. Xi, H. Rincon-Arano, S.B. Rothbart, D. Peng, H. Wen, C. Larson, X. Zhang, X. Zheng, M.A. Cortazar, P.V. Pena, A. Mangan, D.L. Bentley, B.D. Strahl, M. Groudine, W. Li, X. Shi, and T.G. Kutateladze, The histone-H3K4-specific demethylase KDM5B binds to its substrate and product through distinct PHD fingers. *Cell Reports*, **2014**, 6 (2), 325-35.

[119] Adachi, S., T. Koma, N. Doi, M. Nomaguchi, and A. Adachi, Commentary: Origin and evolution of pathogenic coronaviruses. *Frontiers in Immunology*, **2020**, 11 (811).

[120] McIntosh, K., J.H. Dees, W.B. Becker, A.Z. Kapikian, and R.M. Chanock, Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. U. S. A.*, **1967**, 57 (4), 933-940.

[121] Hamre, D. and J.J. Procknow, A New Virus Isolated from the Human Respiratory Tract. *Proc. Soc. Exp. Biol. Med.*, **1966**, 121 (1), 190-193.

[122] Drosten, C., S. Günther, W. Preiser, S. van der Werf, H.-R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A.M. Fouchier, A. Berger, A.-M. Burguière, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.-C. Manuguerra, S. Müller, V. Rickerts, M. Stürmer, S. Vieth, H.-D. Klenk, A.D.M.E. Osterhaus, H. Schmitz, and H.W. Doerr, Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.*, **2003**, 348 (20), 1967-1976.

[123] van der Hoek, L., K. Pyrc, M.F. Jebbink, W. Vermeulen-Oost, R.J.M. Berkhout, K.C. Wolthers, P.M.E. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout, Identification of a new human coronavirus. *Nature Medicine*, **2004**, 10 (4), 368-373.

[124] Woo, P.C.Y., S.K.P. Lau, C.-m. Chu, K.-h. Chan, H.-w. Tsoi, Y. Huang, B.H.L. Wong, R.W.S. Poon, J.J. Cai, W.-k. Luk, L.L.M. Poon, S.S.Y. Wong, Y. Guan, J.S.M. Peiris, and K.-y. Yuen, Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J. Virol.*, **2005**, 79 (2), 884-895.

[125] Chan, J.F.W., K.S.M. Li, K.K.W. To, V.C.C. Cheng, H. Chen, and K.-Y. Yuen, Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic? *J. Infect.*, **2012**, 65 (6), 477-489.

[126] Sohrabi, C., Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, **2020**, 76, 71-76.

[127] Khailany, R.A., M. Safdar, and M. Ozaslan, Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, **2020**, 19, 100682.

[128] Phan, T., Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution*, **2020**, 81, 104260.

[129] Chen, Y., C. Yiu, and K. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research*, **2020**, 9 (129).

[130] Vatansever, E.C., K.S. Yang, A.K. Drelich, K.C. Kratch, C.-C. Cho, K.R. Kempaiah, J.C. Hsu, D.M. Mellott, S. Xu, C.-T.K. Tseng, and W.R. Liu, Bepridil is potent against SARS-CoV-2 in vitro. *Proc. Natl. Acad. Sci. U. S. A.*, **2021**, 118 (10), e2012201118.

[131] Muramatsu, T., C. Takemoto, Y.-T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu, and S. Yokoyama, SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U. S. A.*, **2016**, 113 (46), 12997-13002.

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

[133] Pedersen, N.C., Y. Kim, H. Liu, A.C. Galasiti Kankanamalage, C. Eckstrand, W.C. Groutas, M. Bannasch, J.M. Meadows, and K.-O. Chang, Efficacy of a 3C-like protease inhibitor in treating various forms of acquired feline infectious peritonitis. *Journal of Feline Medicine and Surgery*, **2017**, 20 (4), 378-392.

[134] Kim, Y., S. Lovell, K.-C. Tiew, S.R. Mandadapu, K.R. Alliston, K.P. Battaile, W.C. Groutas, and K.-O. Chang, Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. *J. Virol.*, **2012**, 86 (21), 11754-11762.

[135] Ghosh, A.K., G. Gong, V. Grum-Tokars, D.C. Mulhearn, S.C. Baker, M. Coughlin, B.S. Prabhakar, K. Sleeman, M.E. Johnson, and A.D. Mesecar, Design, synthesis and antiviral efficacy of a series of potent chloropyridyl ester-derived SARS-CoV 3CLpro inhibitors. *Bioorg. Med. Chem. Lett.*, **2008**, 18 (20), 5684-5688.

[136] Ma, C., M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, and J. Wang, Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Research*, **2020**, 30 (8), 678-692.

[137] Ng, M.-L., S.-H. Tan, E.-E. See, E.-E. Ooi, and A.-E. Ling, Proliferative growth of SARS coronavirus in Vero E6 cells. *J. Gen. Virol.*, **2003**, 84 (12), 3291-3303.

[138] Yen, Y.-T., F. Liao, C.-H. Hsiao, C.-L. Kao, Y.-C. Chen, and B.A. Wu-Hsieh, Modeling the Early Events of Severe Acute Respiratory Syndrome Coronavirus Infection In Vitro. *J. Virol.*, **2006**, 80 (6), 2684-2693.

[139] Yang, K.S., X.R. Ma, Y. Ma, Y.R. Alugubelli, D.A. Scott, E.C. Vatansever, A.K. Drelich, B. Sankaran, Z.Z. Geng, L.R. Blankenship, H.E. Ward, Y.J. Sheng, J.C. Hsu, K.C. Kratch, B. Zhao, H.S. Hayatshahi, J. Liu, P. Li, C.A. Fierke, C.K. Tseng, S. Xu, and W.R. Liu, A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors. *ChemMedChem*, **2021**, 16 (6), 942-948.

[140] Ren, X.F., H.W. Li, X. Fang, Y. Wu, L. Wang, and S. Zou, Highly selective azadipeptide nitrile inhibitors for cathepsin K: design, synthesis and activity assays. *Org. Biomol. Chem.*, **2013**, 11 (7), 1143-8.

[141] Shrimp, J.H., S.C. Kales, P.E. Sanderson, A. Simeonov, M. Shen, and M.D. Hall, An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19. *ACS Pharmacol. Transl. Sci.*, **2020**, 3 (5), 997-1007.

[142] Fu, L., F. Ye, Y. Feng, F. Yu, Q. Wang, Y. Wu, C. Zhao, H. Sun, B. Huang, P. Niu, H. Song, Y. Shi, X. Li, W. Tan, J. Qi, and G.F. Gao, Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.*, **2020**, 11 (1), 4417.

[143] Biniossek, M.L., D.K. Nägler, C. Becker-Pauly, and O. Schilling, Proteomic Identification of Protease Cleavage Sites Characterizes Prime and Non-prime Specificity of Cysteine Cathepsins B, L, and S. *J. Proteome Res.*, **2011**, 10 (12), 5363-5373.