P188G MUTATION ENHANCES PYRROLYSYL-TRNA SYNTHETASE ACTIVITY AND STABILIZES FULL-LENGTH PROTEIN

AND

DISCOVERING DRUGS AGAINST SARS-COV-2 M^{PRO} AND PL^{PRO}

A Dissertation

by

CHIA-CHUAN CHO

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 Ray Liu
Committee Members,	David Barondeau
	Pingwei Li
	Frank Raushel
Head of Department,	Simon W. North

December 2022

Major Subject: Chemistry

ABSTRACT

Aggregation of pyrrolysyl-tRNA synthetase (PylRS) has proven problematic for expression of non-canonical amino acids (ncAAs) in proteins. *C. M. alvus* PylRS (CMaPylRS) is a highly soluble PylRS with a shorter N-terminal domain. While testing the orthogonality between CMaPylRS and *M. mazei* PylRS (MmPylRS), we found that $tRNA^{Pyl}$ (PylT) from *C. M. alvus* stabilized MmPylRS and enhanced its activity 3-fold higher than its cognate tRNA by suppressing an amber codon within green fluorescent protein. We further found that MmPylRS was cleaved to two fragments, aa111-454 and aa190-454 when paired to CMaPylT or MmPylT. When the fragments were expressed separately, the MmPylRS activity was only 40%. We showed that a P188G mutation can stabilize MmPylRS and enhance its activity.

Since 2020, millions have died from SARS-CoV-2 infection, and even more have suffered long-covid symptoms. As such, a cure is urgently in need. Main protease (M^{pro}) and papain-like protease (PL^{pro}) are essential proteases in SARS-CoV-2 virology. We screened deubiquitinase inhibitor and cysteine protease inhibitor libraries against PL^{pro} and identified 4 inhibitors: GRL-0617, TCID, SJB2-043 and PR-619 with IC_{50} under 10 μM using a peptide substrate, Z-LRGG-AMC. When the substrate was switched from Z-LRGG-AMC to Ubiquitin-AMC, the IC_{50} of SJB2-043 improved from 0.56 μM to 0.091 μM. The dramatic improvement suggests a binding mode for this inhibitor to PL^{pro} distinct from other inhibitors. *In vitro* inhibition assays do not always match with *in vivo* assays

due to mechanisms of cellular drug delivery. Because antiviral assays cannot be performed with SARS-CoV-2 in low-BSL laboratories, we developed a cellular assay to evaluate M^{pro} inhibitors to facilitate drug discovery. In this assay, a fusion protein of M^{pro} and eGFP in HEK293T cell was expressed and quantified via fluorescent flow cytometry. The cytotoxicity caused by M^{pro} results in low cell fluorescence, while a potent inhibitor rescues the cell (high fluorescence). Therefore, cellular potency of inhibitors can be quantified according to the inhibitor-dependent fluorescence enhancement. Using this assay, 30 known M^{pro} inhibitors were analyzed. MPI5-8 resulted in excellent antiviral effects and *in vivo* M^{pro} inhibition. This cellular assay reduces the BSL requirement for evaluating M^{pro} inhibitors and significantly facilitates the drug discovery process.

DEDICATION

To my father and mother. Without their love and support, I would have never been able to accomplish this.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Wenshe Ray Liu, and my committee members, Dr. David Barondeau, Dr. Frank Raushel and Dr. Pingwei Li, for their guidance and support throughout the course of this research.

I would also like to thank my colleagues Dr. Ge Yu and Dr. Erol C. Vatansever for helping me out with all the questions I have encountered in the research. Our previous member Dr. Wenyue Cao as well as Dr. Baoyu Zhao from Professor Pingwei Li's lab for teaching me mammalian cell and insect cell experiment skills. I would love to thank my friend Ming-Uei Hung and Che-Hsuan Chang who hosted me when I first came to United States and helped me adapting the life here. Thanks to Lauren Blankenship who helped me a lot throughout the time in lab. Thanks to Han-Jung Lee from department of entomology who motivated me on writing. I would love to thank all the people who I met along my way and helped me coping with my life.

Special thanks to Amy Liu who is always busy dealing with all the laboratory miscellaneous errands but still listened to my complaints and encouraged me to carry on.

Lastly, I would like to thank my mother and father and all my family for their support and encouragement. It would not have been accomplished without them.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Wenshe Ray Liu, Professor David Barondeau and Professor Frank Raushel of the Department of Chemistry, and Professor Pingwei Li of the Department of Biochemistry and Biophysics.

The plasmids used in Chapter II were constructed by Chia-Chuan Cho. All the experiment was conducted by Chia-Chuan Cho. AcdK was synthesized and provided by Dr. Xinyu Ma. Mass spectrums of PylRS fragments were collected by Dr. Yuchen Qiao. H3K36ac quantification was provided by Laruren Blankenship. The results were published in 2022.

In Chapter III, plasmid construction, PL^{pro} purification, Ub-AMC synthesis and assay development were carried out by Chia-Chuan Cho. The inhibition assay was carried out in part by Dr. Tyler Lalonde. The protocol for synthesizing Ub-AMC was provided by Dr. Yuchen Qiao. The docking and simulation results were provided by Dr. Shuhua G. Li and was published in 2022.

The assay established in Chapter IV was conducted by Chia-Chuan Cho together with Dr. Wenyue Cao, and the quantification of *in vivo* M^{pro} inhibition with all the inhibitors was carried out by Chia-Chuan Cho together with Zhi Zachary Geng. All the plasmid constructions were performed by Chia-Chuan Cho. The data processing and the MATLAB script were developed and standardized by Chia-Chuan Cho. The apoptosis

analysis and CP-100356 experiment were carried out by Zhi Zachary Geng. The plague reduction neutralization assay was conducted by Dr. Henry Ji and Dr. Robert Allen in Sorrento Therapeutics. The *in vivo* M^{pro} quantification with western blot from viral infection was carried out by Sankar P. Chaki in Professor Benjamin Neuman's laboratory. The *in vivo* M^{pro} quantification from M^{pro} expressing plasmid transfection was done by Chia-Chuan Cho. The MPI1-9, 10-1, 10-2 and 10-3 were synthesized and provided by Dr. Yugendar R. Alugbelli, Dr. Yuying Ma and Dr. Xinyu Ma. *In vitro* characterization of inhibitors and M^{pro} purification was carried out by Dr. Erol C. Vatansever and Dr. Kai S. Yang. The data was published in 2022.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University.

This work was also made possible in part by National Institute of Health under Grant Numbers R01GM121584 and R01GM127575, National Institute of Allergy and Infectious Disease under Grant Number R21AI64088, Welch Foundation under Grant Number A-1715, and the Texas A&M X-Grant to W.R.L. and S.X.

NOMENCLATURE

ncAA	non-canonical amino acid
E. coli	*Escherichia coli*
TyrRS	Tyrosyl-tRNA synthetase
OMeY	*ortho*-methyl-L-tyrosine
tRNA	transfer ribonucleic acid
PylRS	Pyrrolysyl-tRNA synthetase
PylT	Pyrrolysyl-tRNA
MbPylRS	*Methanosarcina bakeri* PylRS
MmPylRS	*Methanosarcina mazei* PylRS
CMaPylRS	*C. Methanomethylophilus alvus* PylRS
MmPylT	*Methanosarcina mazei* Pyrrolysyl-tRNA
CmaPylT	*C. Methanomethylophilus alvus* Pyrrolysyl-tRNA
mRNA	messenger ribonucleic acid
Pyl	pyrrolysine, noted O as amino acid abbreviation
BocK	N^{ε}-Boc-L-lysine
NTD	N-terminual domain
CTD	C-terminual domain
PTM	Post-translational modification
AcK	N^{ε}-acetyl-L-lysine
PrK	N^{ε}-propionyl-L-lysine
BuK	N^{ε}-butyryl-L-lysine

CrK	N^{ε}-crotonyl-L-lysine
AcdK	N^{ε}-(4-azidobenzoxycarbonyl)-δ, ε-dehydrolysine
AznL	Azidonorlucine
TCEP	Tris(2-carboxyethyl)phosphine
PCR	polymerase chain reaction
epPCR	error-prone polymerase chain reaction
PACE	Phage assisted continuous evolution
PANCE	Phage assisted non-continuous evolution
chPylRS	chimeric PylRS
SARS	Severe acute respiratory syndrome
SARS-CoV	Severe acute respiratory syndrome coronavirus
MERS	Middle east respiratory syndrome
COVID-19	Coronavirus disease 2019
ORF	Open reading frame
ACE2	Angiotensin converting enzyme 2
Cts	Cathepsin
TMPRSS2	Transmembrane serine protease 2
ssRNA	single-stranded RNA
IC_{50}	half maximal inhibition concentration
EC_{50}	half maximal effective concentration
PDB	protein data bank
PL^{pro}	Papain-like protease

M^{pro}	Main protease
RdRP	RNA-dependent RNA polymerase
nsp	non-structural protein
Ub	Ubiquitin
Ubl	Ub-like
Ub-AMC	Ub-7-amino-4-methylcoumarin
ISG15	Interferon-stimulated gene 15
IFNα	Interferon α
IFNAR 1 and 2	Interferon alpha/beta receptor subunit 1 and 2
sfGFP	super folder green fluorescence protein
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
Tris	2-(amino-2-(hydroxymethyl)propane-1,3-diol
PMSF	phenylmethylsulfonyl fluoride
FPLC	Fast protein liquid chromatography
NaCl	sodium chloride
OD_{600}	optical density 600 nm
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ESI-MS	electrospray ionization mass spectrum
aa	amino acid
pp1a/1ab	polypeptide product 1a/1ab
DTT	dithiothreitol

ACPL	Activated cysteine-directed protein ligation
DMSO	dimethyl sulfoxide
Gly-AMC	Glycine 7-amino-4-methylcoumarin
NTCB	2-nitro-5-thiocyanatobenzoic acid
NaOH	Sodium hydroxide
FBS	fetal bovine serum
HEK293T	human embryonic kidney 293 cells
PBS	Phosphate buffered saline
SSC-A/H	Size scatters-Area/Height
FSC-A/H	Forward scatters-Area/Height
FITC-A	Fluorescein isothiocyanate-Area
CPE	cytopathogenic effect
FRET	Förster resonance energy transfer
CFP	Cyan fluorescent protein
YFP	Yellow fluorescent protein

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Genetic Code Expansion

The biological system utilizes 64 genetic codes to afford 21 canonical amino acids in protein biosynthesis. Multiple codes are only translated to a single amino acid.[1] This redundancy highly restricts the possible diversity. In addition, to ensure protein stability, most canonical amino acid side chains are non-reactive functional groups.[2] The lack of reactive groups increases the difficulty on protein modification. The amino acids cysteine has been widely used to conduct protein labeling due to its ability to form disulfide bonds and serve as a strong nucleophile. However, the specificity is not guaranteed when there are multiple cysteines.[3] To overcome this difficulty, several protein ligation methods have been developed to facilitate protein modification. One of them is utilizing the naturally occurring intein ligation. The special intein sequencing allows two proteins to fuse and undergo protein splicing, which removes the intein and results in a traceless fusion protein.[4] With this feature, an expressed protein with intein can ligate with a chemically synthesized peptide which contains non-canonical amino acids (ncAAs) with active functional groups such as alkenes, alkynes, azides or aldehydes. A site-specific modification can be achieved with these functional groups. However, the length of synthesized peptide which contains ncAAs is limited by solid phase peptide synthesis.[5]

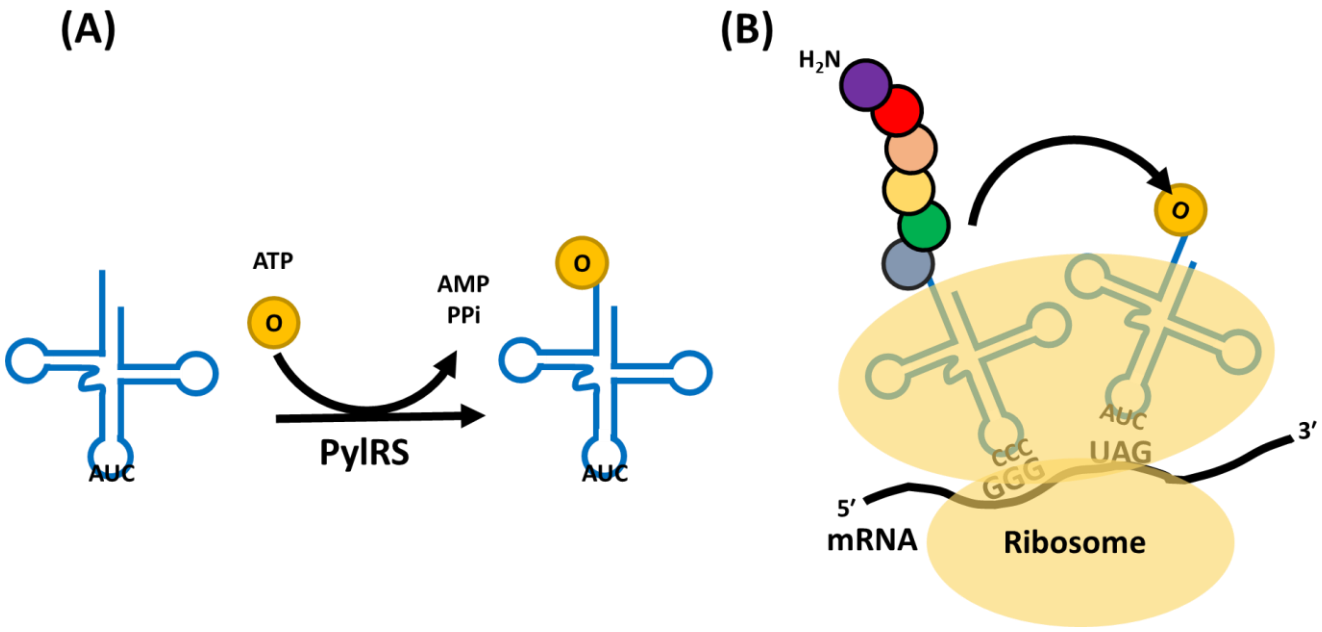

Figure 1 (A) Pyrrolysyl-tRNA synthetase synthesize the amber suppressing pyrrolysyl-tRNA. (B) Pyrrolysyl-tRNA is used in translation to incorporate pyrrolysine in an expressed protein. The arrow indicates the movement of polypeptide chain.

Therefore, genetic code expansion utilizing the redundant genetic code for an extra ncAAs was proposed to facilitate protein chemistry. [6]

The genetic code expansion system is widely used in the chemical biology field as well as in the epigenetic research field. The concept was first proposed by Dr. Peter Schultz an others.[6] His team demonstrated the feasibility of using the amber stop codon as a sense codon to incorporate an ncAA. They first chemically synthesized an amber suppressing tRNA which was then amino acylated with different ncAAs. This tRNA was used in *in vitro* protein expression, demonstrating the concept of expanding the genetic code with amber codon. They further developed a system that site-specifically

incorporates *ortho*-methyl-L-tyrosine (oMeY) in *Escherichia. coli* by an artificially evolved *ortho*-methyl-tyrosyl-tRNA synthetase (oMeYRS) derived from *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase and its tRNA[7]. The fascinating pioneer work of Dr. Peter Schultz inspired scientists to develop versatile tools for chemical biology research.[8] In 2002, a naturally existing amber codon suppressing system was identified in a methanogenesis archaea, *Methanosarcina bakeri*. This system utilizes a naturally evolved pyrrolysyl-tRNA synthetase (PylRS) and its tRNA for reprogramming the amber codon to code 22nd amino acid, pyrrolysine (Pyl, noted O as one letter code). (Fig. 1A and 1B)[9, 10]. This naturally existing amber suppression system has an even greater plasticity than the artificially developed one[11]. PylRS consists of two functional domains: the N-terminal as a tRNA binding domain and the C-terminal as a catalytic domain. The tRNA binding domain recognizes the tertiary structure of tRNA without interacting with the anticodon region.[12] With this feature, the anticodon region was mutated to recognize the other two stop codons for utilizing ochre and opal codons as expanded genetic codons as well.[13] The C-terminal domain catalyzes the amino acylation of Pyl to form pyrrolysyl-tRNA. Structural analysis suggests that the specificity of PylRS for Pyl resulted solely from the van der Waal interactions between the active site pocket and Pyl[14]. This characteristic provides PylRS the ability to recognize several ncAAs with hydrophobic side chains.[11] In addition, it also allows us to engineer the amino acid binding pocket to fit hundreds of different ncAAs for site-specific incorporations in expressed proteins. Furthermore, some of the ncAAs contain active

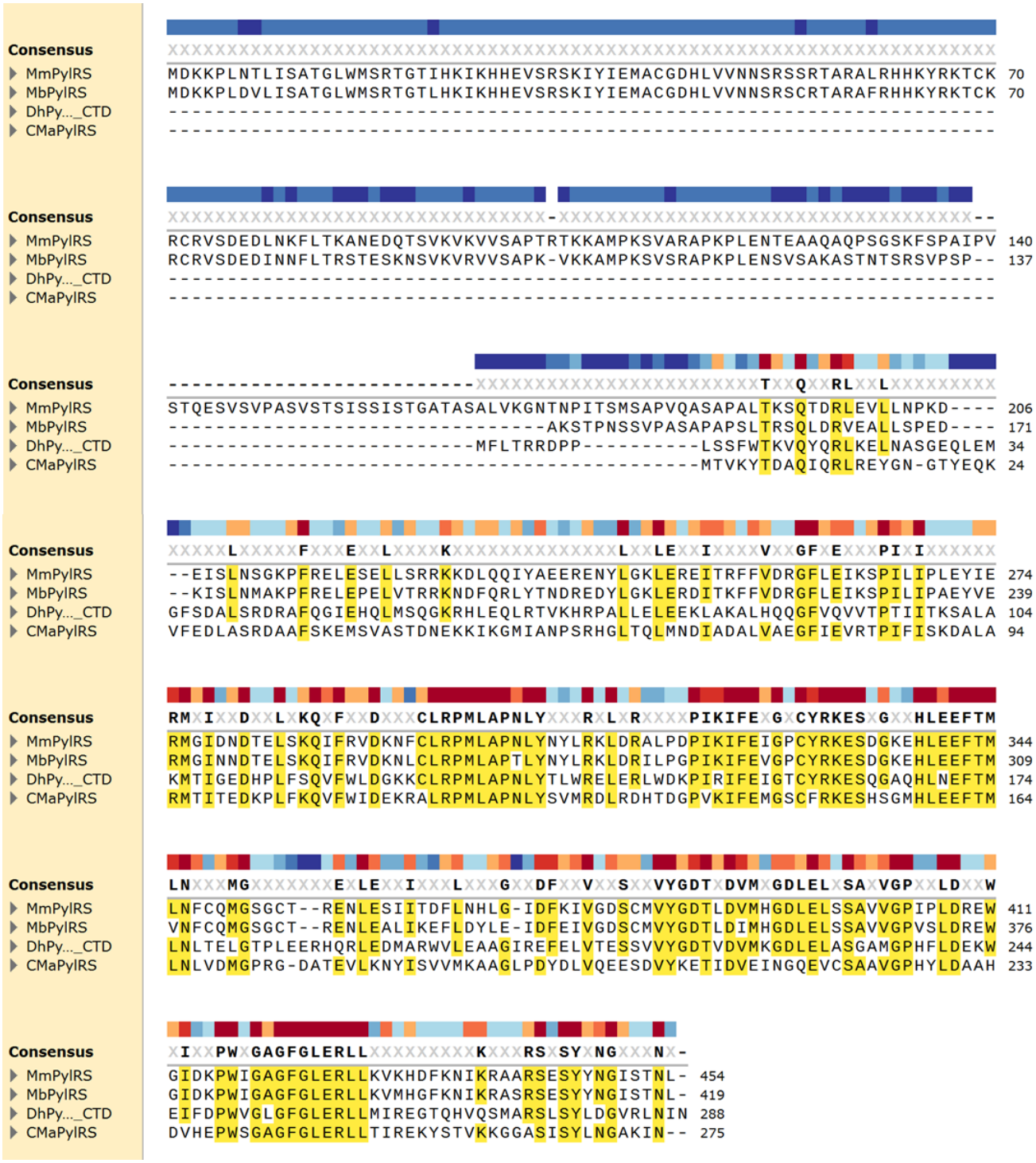

Figure 2 Sequence alignment of PylRS found from *Methanosarcina mazei*, *Methanosarcina bakeri*, *Desulfitobacterium hafniense* and *C. Methanomethylophilus alvus*.

functional groups for advanced chemical modification, such as using propargyl-lysine for click reactions. Although PylRS is significantly versatile, the enzyme tends to aggregate when overexpressed and can lose its activity due to the hydrophobicity of the N-terminal domain (NTD)[15]. To overcome this problem, researchers searched for a more soluble PylRS. Enzymes from *Methanosarcina mazei* (MmPylRS) as well as a split PylRS from *Desulfitobacterium hafniense* (DhPylRS) were found (Fig. 2)[16, 17]. Yet, the solubility of MmPylRS was only slightly improved over that of MbPylRS, and DhPylRS possessed significantly lower catalytic activity compared to its homologs.[18] Several clones of PylRS have been screened for enhanced stability including a R3-11 MmPylRS from our lab and an IPYE MbPylRS shown in Bryson *et al*[19, 20]. The most soluble PylRS was discovered from *C. Methanomethylophilus alvus* (CMaPylRS) with a much shorter NTD, but the CTD had 80% similarity to the CTD of MmPylRS (Fig. 2)[16, 21]. The solubility of CMaPylRS is highly improved from all the previous clones. Taking advantage of its structural similarity against MmPylRS, some of the previously discovered mutants can be directly transferred to CMaPylRS to yield a greater catalytic activity. However, the transfer of mutations doesn't always ensure an active enzyme[22]. It is likely because of a smaller substrate binding pocket comparing to MmPylRS (Fig. 3).[23] More detailed knowledge regarding the molecular structure of CMaPylRS is crucial to further explain experimental results. In summary, the novel characteristics of CMaPylRS result in it being a potential resource to further expand chemical biology research. In addition, the different origin of CMaPylRS from other PylRSs presents the possibility of using multiple non-sense codons to incorporate more ncAAs into one expressed protein[24].

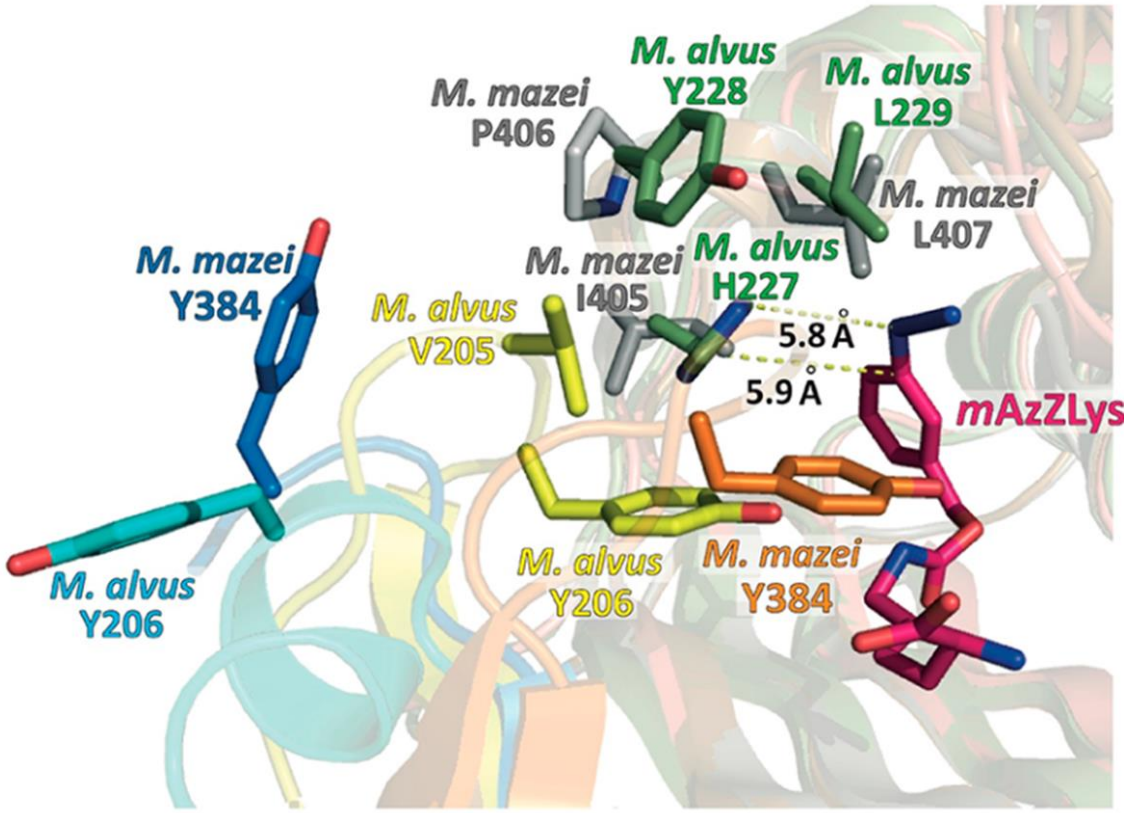

Figure 3 Active site conformation of MmPylRS and CMaPylRS

PylRS and Its Directed Evolution

Structural analysis of PylRS and its cognate tRNA has shown that the specificity of PylRS to Pyl is solely from a simple lock-and-key enzymatic model.[25] This feature allows us to engineer the cavity structure of PylRS active site for novel amino acids. The most efficient way to identify an ncAA-specific PylRS mutatnt is to perform a directed evolution against the wild type PylRS. There are several ways to carry out the evolutions.

One of the ways is to build a PylRS library through saturated mutation.[26] According to the co-crystal of PylRS/Pyl structure, the Pyl fits into the active site with a specific orientation. With this information, we can replace Pyl with other ncAAs in the model to simulate binding modes and identify possible interacting residues (Fig. 4A). For most of the lysine derivatives, N^{ε} on the side chain was acylated with different functional groups.[11] To identify a PylRS mutant for a specific lysine acylation, a PylRS library can be generated by randomizing interacting sites with saturated mutations. With this method many PylRS variants have been screened out for ncAAs such as N^{ε}-acetyl-L-lysine (AcK), N^{ε}-butyryl-L-lysine (BuK), N^{ε}-(4-azidobenzoxycarbonyl)-δ, ε-dehydrolysine (AcdK), Azidonorlucine (AznL) (Fig. 4B).[27-30] AcK is a post-translational modification (PTM)

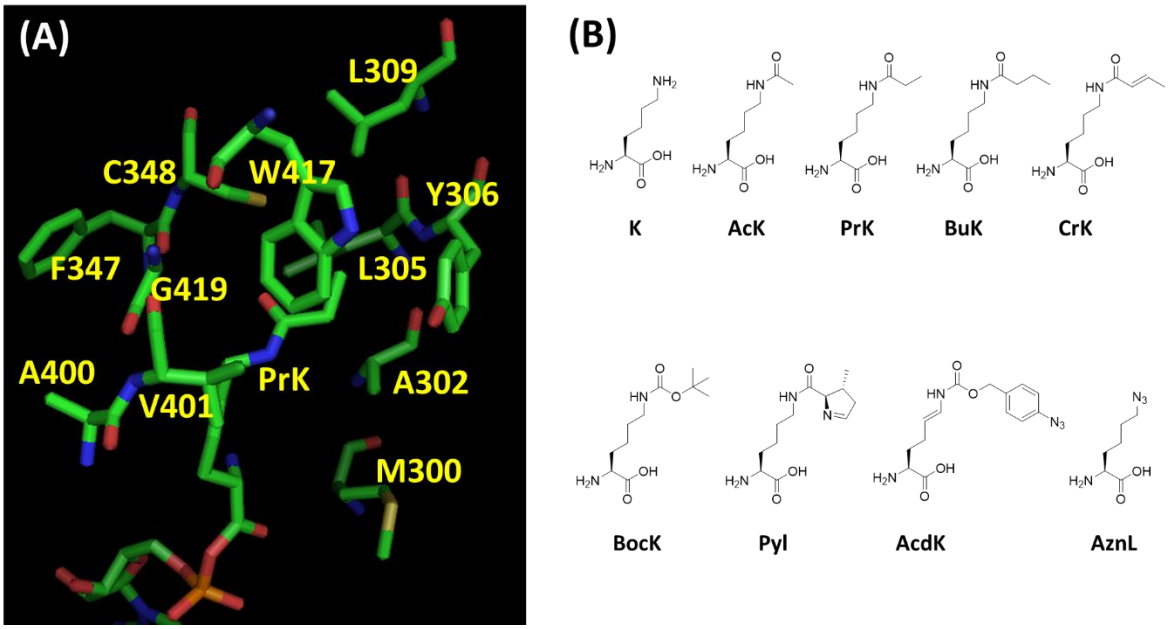

Figure 4 (A) Based on the crystal structure of MmPylRS, 4CH5, the amino acids within 5 Å proximity to N^{ε} are M300-L309, N346-C348, Y384, A400, V401, W417 and G419. (B) Chemical structure of lysine derivatives.

marker that is widely used in nucleosome modification.[31] It was proposed to silence the gene with a tighter binding toward acetylated nucleosomes, yet different sites of lysine acetylation have a more profound function rather than just gene silencing. Genetically incorporating AcK at a specific site of nucleosome allows us to address a detailed view of its functions. For this purpose, a PylRS recognizing AcK was screened and identified as AcKRS1 with L301M, Y306L, L309A and C348F.[32] It has been widely used in protein PTM research. BuK is another PTM marker.[33] The larger acylated group compared to acetylation on the nucleosome likely destabilizes nucleosome structure for gene activation. BuKRS was screened by randomizing Y384 and identified as Y384W. This mutation combining with 4 mutations in AcKRS1 allows us to genetically incorporate N^{ε}-propionyl-L-lysine (PrK). Due to the structure similarity of BuK and N^{ε}-crotonyl-L-lysine (CrK), BuKRS also recognize CrK as a substrate. The alkene group on CrK allows it to undergo nucleophilic addition. A crotonylation on nucleosome leads to more complicated gene modulation. This modification was found to rescue acute kidney failure.[34] It was also found to cause depression when accumulated in mice brain.[35] AcK, PrK, BuK and CrK are lysine acylation with different carbon chains that have been identified as nucleosome modifications.[36] Even though, only the length of carbon chain differs, these changes lead to a profound change in biological function. For some of the PTM such as N^{ε},N^{ε}-dimethyl-L-lysine (Kme2) that can't be genetically incorporated due to the structural similarity to lysine, they need to be synthesized by chemical reactions. AcdK is one of the precursors that can undergo mild reaction to generate allysine without

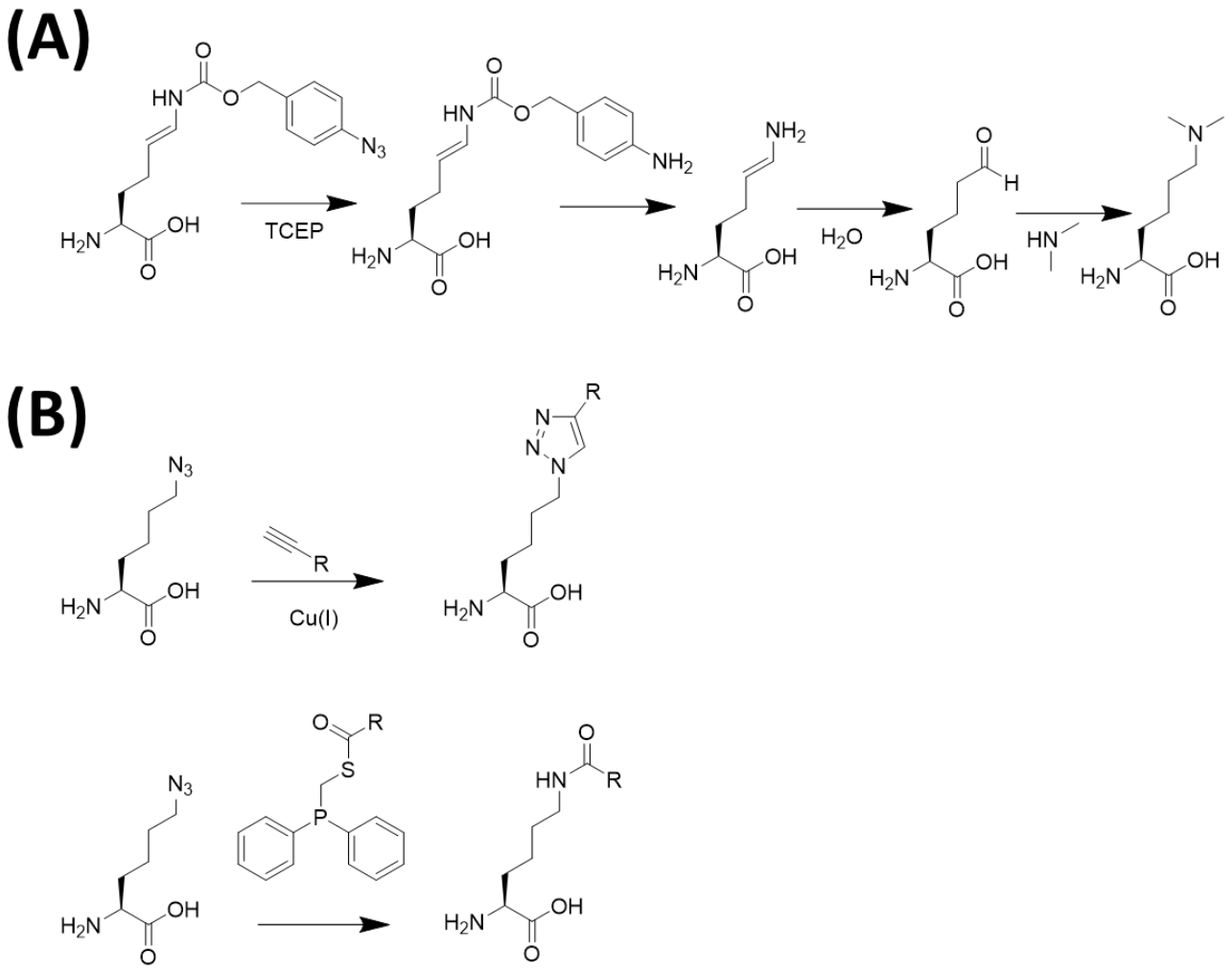

Figure 5 (A) AcdK was used to synthesize lysine demethylation through allysine. (B) AznL was used as for click reaction as well as synthesizing lysine acylation through Staudinger reaction.

degrading expressed protein. The aldehyde group on allysine can then react with dimethylamine to form enamine and furthermore reduce to form Kme2. With this precursor, any dialkylated lysine can be easily synthesized (Fig 5A).[30] To genetically incorporate AcdK, an AcdKRS was selected and then identified as L309T, C348G and

Y384F. AznL is an ncAA with lysine switching the N^{ε} amine to an azide group. The azide group can undergo click reactions to form more lysine modifications on an expressed protein, and can also be used to synthesize any acetylated lysine through the Staudinger reaction (Fig 5B).[29] With the same selection method, AznLRS was identified as Y306L, C348I and Y384F. PylRS has been specifically evolved to recognize much more ncAAs. There are more PylRS variants that incorporate lysine with reactive functional groups.[28, 37] It has been widely used in chemical biology research as a powerful tool for protein labeling and chemical modification.

The solubility of PylRS has been a long-lasting issue for reaching high yields of ncAA incorporation. To overcome this problem, it needs more complicated mutagenesis than just mutating the active site residues. There are several ways to carry out directed evolution for identifying a more stable PylRS. One of the ways is to randomly generate mutation on PylRS sequence by error-prone polymerase chain reaction (epPCR) (Fig. 6A). With this method, a clone with three mutations, R19H, H29R and T122S, in the tRNA binding domain was identified. This mutant was noted as R3-11 MmPylRS, and it was shown to recognize N^{ε}-tert-butyl-L-lysine three folds than the wild type one.[19] Furthermore, the mutations were also shown to be transferrable to existing MmPylRS variants to enhance the activity. Another method is to go through phage-assisted continuous evolution (PACE) or phage-assisted non-continuous evolution (PANCE) (Fig. 6B).[38] This method takes the advantage of the fast-propagating feature of phage and a

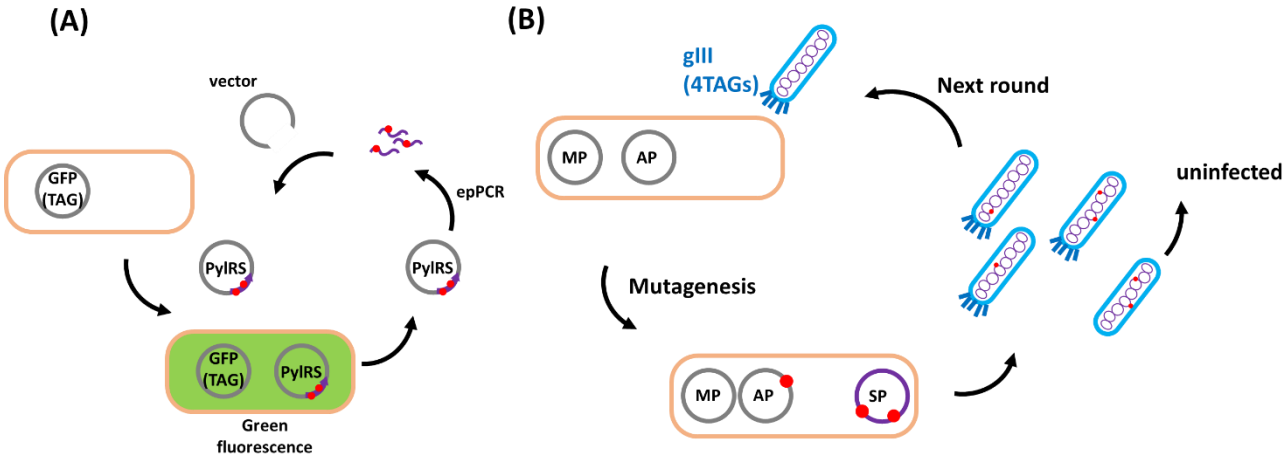

Figure 6 (A) EpPCR was used to generate a library of randomized MmPylRS. SfGFP was encoded with an in-frame amber codon. The improved PylRS increases the green fluorescence from the cell. (B) Phage assisted continuous evolution (PACE) was used to quickly evolve PylRS. MP is the mutagenesis plasmid which contains proteins that generate mutations. AP is an accessory plasmid that encodes gIII with 4 in-frame amber codons as well as PylT. The truncated gIII diminishes the infectivity of a phage. SP is a selection plasmid that encodes PylRS and genes for phage packaging.

recombination incompetent *E. coli*, S2060, which has lower fidelity on repairing mutations. Mutations were generated randomly through a mutagenesis plasmid and passed on to different phage generations. With this method, it can complete multiple rounds of selection in an hour. The time-consuming directed evolution can be greatly shortened to within a week with these advantages. PACE has been used to evolve and increase the activity of a chimeric PylRS (chPylRS) which is a fusion protein of MbPylRS1-149 and MmPylRS185-454.[20] They identified a chPylRS-IPYE with four mutations: V31I, T56P, H62Y and A100E. They also showed these mutations are transferrable to other chPylRS variants. These mutations rescued the activity of chimeric AcK3-RS around 90-fold. Due

to the fast evolution feature, these methods have been used to identify mutants that have better tRNA binding as well as a more active PylRS from different species.[39, 40]

Coronavirus and Drug discovery

The coronavirus was first isolated and identified in 1965 from a nasal swap collected from a patient, B814 strain.[41] This novel virus was then inoculated to volunteers and led to a cold-like syndrome. The novel virus' morphology was visualized with electron microscopy and then categorized as coronavirus in 1968 (Fig. 7A).[42] Since the discovery of the first coronavirus, there have been several disease outbreaks caused by human infectious coronaviruses. These outbreaks include but are not limited to the well-known Severe Acute Respiratory Syndrome (SARS) in Hong Kong in 2003

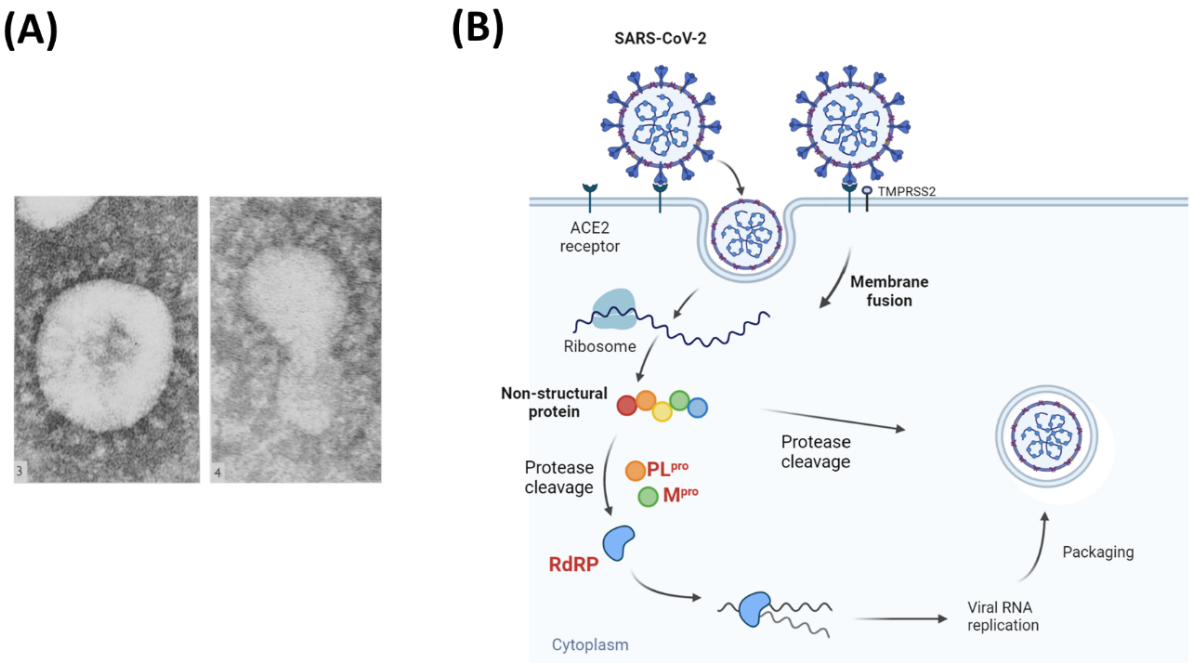

Figure 7 (A) Electron microscopy image of coronavirus B814 from 1965. (B) The brief mechanism of SARS-CoV-2 infection and replication.

caused by SARS-CoV, Middle East Respiratory Syndrome (MERS) in 2012 in Saudi Arabia by MERS-CoV, and the recent pandemic, Coronavirus Disease 2019 (COVID-19), in 2019 caused by SARS-CoV-2.[43] The novelty, highly infectious features, and other uncontrollable factors make it hard to prevent spreading of the virus. Although vaccines from several companies were released in 2021, the high mutation rate of the positive-strand RNA virus, SARS-CoV-2 as well as highly breakthrough infection rate are still a potential risk to people all over the world.[44] Therefore, potent antiviral drugs are still essential to cope with this threat.

Right after the epidemic outbreak in the Wuhan seafood market in China in 2019, the whole genome of the virus Alpha strain was identified and released promptly for viral research. Compared to SARS-CoV-1, they shared 79% genome similarity and sharing 94.4% similarity in the amino acid sequences in 7 conserved domains in open reading frame 1ab (ORF1ab).[45] SARS-CoV-2 is a positive-stranded RNA virus (Fig. 7B) that enters human cells through the interaction between the spike protein on the surface of virus and the Angiotensin Converting Enzyme 2 (ACE2) receptor on the cell surface.[46] Specifically, there are two ways of entry: 1) Directly from the endocytosis after the binding of ACE2 and spike, the RNA is then released into the cytosol by Cathepsin L proteolytic activity or 2) Furin and TMPRSS2 mediated proteolytic releasing.[47] The TMPRSS2 pathway induced cell fusion likely causes the commonly reported symptoms of lost taste and smell. The unique mutations found in the Omicron variant's spike protein (N764K, D796Y, N856K, Q954H, N969K and L981F) reduces TMPRSS2 cleavage of spike

protein. [48] Compared to the Alpha and the Delta strains which contain mutations that enhance cleavage from Furin, the Omicron strain causes more flu-like symptoms.[49] The released viral positive ssRNA contains two important ORFs, ORF1a and ORF1ab. These two ORFs are translated to long polypeptide chains (pp1a and pp1ab) which are cleaved and identified as non-structural proteins 1-16 (nsp1-16).[50] These polypeptide chains contain almost all the essential proteins for viral replication including papain-like protease (PL^{pro}) in nsp3, main protease (M^{pro}) in nsp5, and RNA-dependent RNA polymerase (RdRp) in nsp12.[51] Among all, PL^{pro} cleaves the nascent peptide chain to release nsp1, 2, 3 and 4; while M^{pro} cleaves and releases most of the protein for viral assembly. RdRp transcribes the positive RNA strand to a negative RNA strand and then synthesizes excess positive RNA strands for viral replication. Due to the importance of the above-mentioned enzymes in the virology of SARS-CoV-2, they are the best potential drug targets for fighting the viral infection. Currently, scientists focused on developing 1) Drugs that disrupt the interaction between spike and ACE2; 2) Drugs that inhibits PL^{pro} and M^{pro} proteolytic activity, and 3) Drugs that inhibit RdRp activity.[52] Among them, PL^{pro} and M^{pro} are identified as cysteine proteases. Cysteine proteases are a huge class of enzymes that can leads to several diseases if not well regulated. They are also essential to most pathogenic viruses.[53] Therefore, there are numerous drugs that have been developed to inhibit cysteine proteases, which makes repurposing existing drugs the best and the fastest way to identify a potent drug for curing COVID.

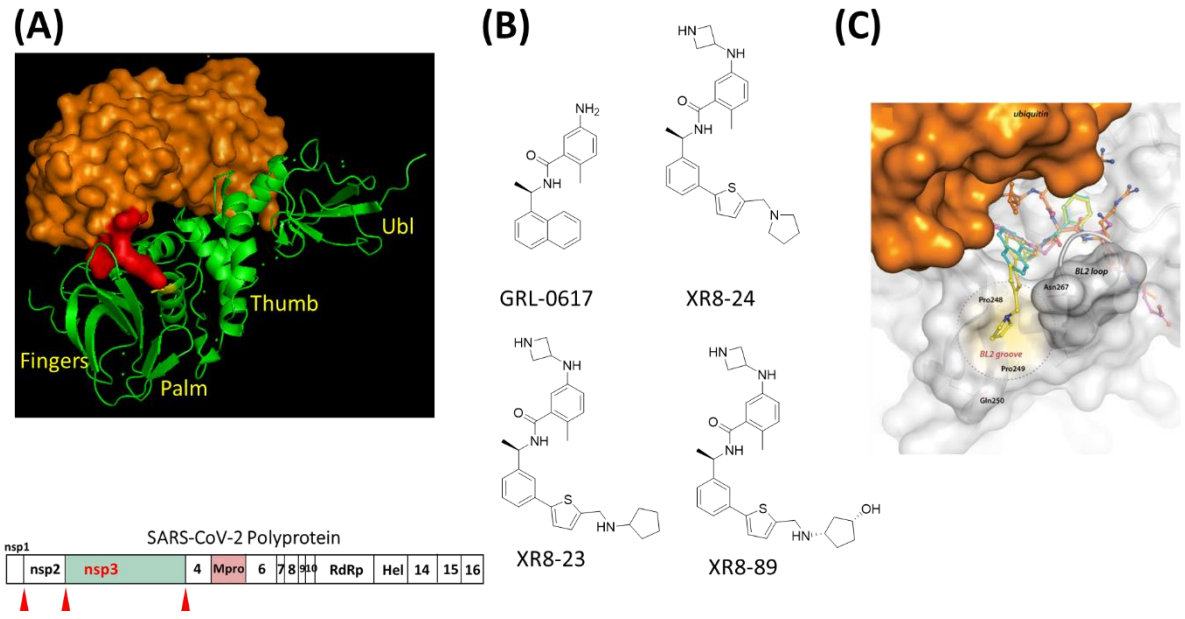

Figure 8 (A) PL^{pro}-mISG15 co-structure: modifying from PDBS: 6YVA. The red residue is peptide, RLRGG. The yellow residue on the loop by RLRGG indicates C111. The red daggers indicate the PL^{pro} cutting sites. (B) GRL-0617 and its derivatives. (C) Co-crystal structure of PL^{pro} and XR8-89.

PL^{pro}: PL^{pro} is a cysteine protease that consists of a unique right-handed finger, palm, and thumb C-terminal domain and a Ub-like (Ubl) N-terminus domain. The catalytic triad, C111, H272 and D286, is located in the palm region. The Ubl domain provides PL^{pro} the ability to interact with Ubl proteins such as Ub and ISGs. The SARS-CoV-2 PL^{pro} has two distinct functions. Firstly, this protease recognizes the peptide chain LXGG/A in between nsp1 and 2, nsp2 and 3, nsp3 and 4 to release functional protein from polypeptide chain (Fig. 8A).[54]Secondly, it provides deubiquitinase activity that interferes with the innate immune response. Compared to the PL^{pro} protease activity which only cleaves 3 sites on the viral polypeptide chain, the deubiquitinase activity is more important for viral

replication. One of the innate antiviral responses is activated by interferon α (IFNα).[55] IFNα binds to its receptor IFNAR1 and IFNAR2 (interferon alpha/beta receptor) complex to activate the interferon-stimulated gene 15 (ISG15) antiviral pathway.[56] ISG15 is a small protein marker for degradation of viral protein. It can also neutralize and aggregate the virus particle by ISGylation. However, the deubiquitinase activity of SARS-CoV-2 PL^{pro} allows it to specifically recognize ISG15 and removes it from viral proteins.[57] Unlike its homologs SARS-CoV and MERS-CoV PL^{pro}, SARS-CoV-2 PL^{pro} specifically removes ISG15s around 350-fold more than deubiquitinylation. This unique function of PL^{pro} allows the virus to evade the immune response.[58] Inhibiting PL^{pro} can not only reduce viral replication but also restores the regular cellular antiviral activity.[59] Several drug screens have been done to identify a specific inhibitor for PL^{pro}.[60] However, due to the shallow and broad cavity of its active site, few good inhibitors were identified except the existing SARS-CoV PL^{pro} inhibitor, GRL-0617, with an IC_{50} of 1 μM.[61] According to the crystal structure of GRL-0617 and SARS-CoV-2 PL^{pro}, functional groups on GRL-0617 have been changed to fit better into the PL^{pro} active site (Fig. 8B).[62] Most potent of all, XR8-89 inhibits PL^{pro} with an IC_{50} around 100 nM. The crystal structure of PL^{pro} and XR8-89 showed that with a longer group than the original naphthalene in GRL-0617, the interaction of inhibitors and PL^{pro} BL2 loop is enhanced (Fig. 8C). This distinct biochemical information could be useful for designing more potent inhibitors. A mutant, SARS-CoV-2 NIB-1 circulated in Bangladesh, with a mutation of V843F (V98F on PL^{pro}) was found within PL^{pro}.[63] This mutation indicates the possibility that a potent inhibitor can lose functionality with unpredictable mutations within the active site. Designing more

inhibitors based on the co-crystal structure of XR8-89 and PL^{pro} can be advantageous as it enables researchers to expand the PL^{pro} inhibitor library and to be prepared for different mutations. There were two irreversible activity-based inhibitors designed based on the peptide substrate of SARS-CoV-2 PL^{pro}, VIR250 and VIR251, reported with the first inhibitor bound crystal structure.[54] However, an irreversible inhibitor is not usually a good drug due to the unknown off-target effects.[64] The unique activity of PL^{pro} provides SARS-CoV-2 the ability to escape from the immune system. A potent drug that specifically targets and inhibits PL^{pro} will be very useful as a treatment for COVID-19. We aimed to identify and repurpose potent PL^{pro} inhibitors from the existing drug libraries.

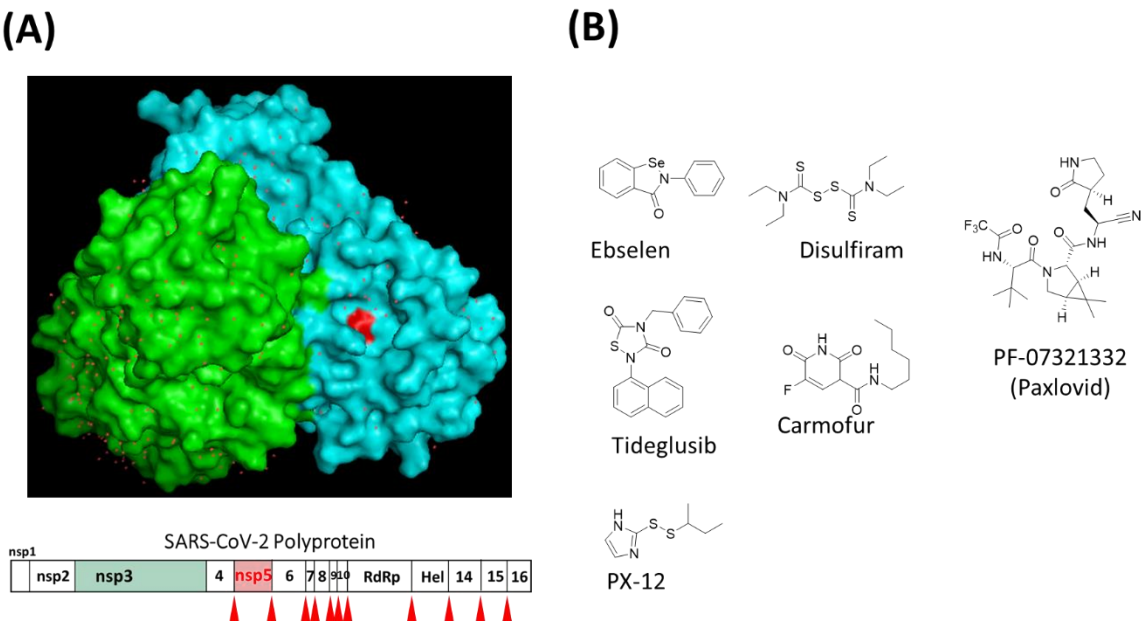

Figure 9 (A) M^{pro} dimer structure. The red residue indicates the active site C145. The red daggers point out the cutting site of M^{pro}. (B) M^{pro} inhibitors identified from high throughput screening and a structure of Pfizer developed M^{pro} inhibitor, PF-07321332.

M^{pro}: M^{pro} (or $3CL^{pro}$: 3C-like protease) is the protease responsible for the maturation of most of the core proteins in SARS-CoV-2. It is encoded in nsp5.[65] It recognizes hydrophobic peptide residues such as KTSAVLQ/S and cuts after the glutamine. Due to the special sequence of coronavirus around the M^{pro} gene, M^{pro} is able to autocleave itself out of the original polypeptide chain and release to form a highly active protease dimer (Fig. 9A). The autocleavage activity allows M^{pro} to remove all the bulky amino acids in front of the first serine and facilitate dimerization. The strong protease activity and the broad substrate preference of M^{pro} leads to extreme cytotoxicity and triggers apoptosis of the infected cells. Inhibiting M^{pro} greatly reduces the viral replication as well as cell illness.[66] Because of the importance of the M^{pro} to the virus, most of the drug discovery research has been designed against it. For example, PAXLOVID, the only FDA approved drug by now, is an M^{pro} inhibitor with an IC_{50} of 3.11 nM.[67]

In the early stage of COVID, a Michael acceptor inhibitor, N3, was used to address the detail enzymology of M^{pro}. The co-crystal structure of N3 showed the detail interaction between N3 and the substrate binding pocket of M^{pro}. SARS-CoV-2 M^{pro} has a catalytic dyad of H41 and C145 which is similar to previously identified coronavirus M^{pro}s. In the crystal structure, M^{pro} consists of two dimers, protomer A and protomer B. The subsite S1 was form with C145 of protomer B and F140, N142, E166, H163 and H172 from protomer A. The interaction within M^{pro} active site between protomer A and B explains why dimerization is essential to obtain its protease activity. The highly charged residues, H163, from protomer A makes glutamine a required residue for P1 site. The S2 site, which

consists of more hydrophobic residues H41, M49, Y54 and Y165, interacts with hydrophobic side chains on P2. P3 was exposed toward solvent, which suggested no specific preference. P4 is surrounded by hydrophobic residues, while P5 and P1' sites interacts with binding pocked through Van Der Waal interaction.[68] Thie details of the interaction of N3 and M^{pro} provides information for designing potential inhibitors.[69, 70] There were several potent inhibitors identified in the early stage of COVID pandemic such as disulfiram, carmofur, ebselen, shikonin, tideglusib and PX-12 (Fig 9B). These inhibitors exhibit *in vitro* IC_{50} values ranging from 0.67 to 21.4 µM. However, ebselen is the only one that exhibited potent antiviral activity with an EC_{50} of 4.67 µM, when it was tested *in vivo* with SARS-CoV-2 infecting vero E6 cells. [68] The inconsistencies between *in vitro* and *in vivo* assay could lead to a huge waste of research availabilities. Even though the *in vivo* antiviral assay resembled the physiological results better, BSL2 laboratories are not allowed to perform it due to safety concerns. To simplify the *in vivo* assay for identifying a potent M^{pro} inhibitor, we developed an *in vivo* assay using fluorescence enhancement to quantify the EC_{50} of each inhibitor.

CHAPTER II *

THE PYRROLYSYL-TRNA SYNTHETASE ACTIVITY CAN BE IMPROVED BY A P188 MUTATION THAT STABILIZES THE FULL-LENGTH ENZYME

Reprinted from Chia-Chuan Cho, Lauren R. Blankenship, Xinyu Ma, Shiqing Xu, and Wenshe Liu *Journal of Molecular Biology* **2022**, 434 (8), 167453 DOI: 10.1016/j.jmb.2022.167453.

Introduction

Pyrrolysine (Pyl) is the 22nd proteinogenic amino acid that is genetically encoded by the amber UAG codon. Its genetic encoding requires pyrrolysyl-tRNA synthetase (PylRS) and its cognate tRNA tRNA^{Pyl} (PylT).[9, 71] The Pyl incorporation system was discovered in 2002 and has been developed as the arguably most used system for undergoing amber suppression-based noncanonical amino acid (ncAA) mutagenesis in cells[27, 72]. So far more than 200 different ncAAs have been genetically encoded using wild-type and engineered PylRS-PylT pairs[11]. These ncAAs have a large variety of chemical functionalities and some of them are post-translationally modified amino acids[12, 27-30, 37, 73]. Applications of these genetically encoded ncAAs include the synthesis of proteins with post-translational modifications for their functional investigations and the engineering of proteins for various purposes such as the synthesis of antibody-drug conjugates.

The widely used PylRS strains for genetic code expansion are from two archaebacteria, *Methanosarcina mazei* and *Methanosarcina bakeri*. They share 83%

sequence identity[12]. However, both PylRSs are unstable due to an unstable N-terminal domain (NTD) that separates from the C-terminal catalytical domain (CTD). NTD is known to recognize PylT for improved binding[15, 39]. However, its easy aggregation leads to an insoluble full-length PylRS or is cleaved from CTD that itself is not catalytically active[15]. To overcome this instability issue, several evolution strategies have been used to optimize PylRS. For example, a more stable and active tetra-mutation IPYE clone was identified for *M. bakeri* PylRS (MbPylRS) and a tri-mutation R3-11 clone was introduced for *M. maze*i PylRS (MmPylRS)[19, 20]. An extreme solution to this problem is the total elimination of NTD, a phenomenon that occurs naturally in *Ca. Methanomethylophilus alvus*. *Ca. M. alvus* PylRS (CMaPylRS) has only a CTD domain that shares 80% similarity with the MmPylRS CTD.[13, 21, 25, 74]

In comparison to MmPylRS and MbPylRS, CMaPylRS has much improved solubility. When expressed in Escherichia coli, full-length MmPylRS was only observed in the cell lysate pellet whereas CMaPylRS was primarily in the cell lysate supernatant (Figure 26). Due to this feature, there has been a trend to switch from using MmPylRS and MbPylRS to using CMaPylRS to conduct genetic code expansion research. However, the majority of ncAAs are genetically encoded using engineered MmPylRS and MbPylRS variants. Mutations in these engineered MmPylRS and MbPylRS variants cannot be simply transferred to CMaPylRS to afford variants for the genetic incorporation of corresponding ncAAs. We have introduced mutations that were found in MmAcKRS1 and MmBuKRS into CMaPylRS to generate corresponding CmaAcKRS and CmaBuKRS.

However, afforded CmaAcKRS and CmaBuKRS clones were not able to mediate the incorporation of N^{ε}-acetyl-L-lysine (AcK) and N^{ε}-crotonyl-L-lysine (CrK), respectively (Figure 27). Therefore, finding a way to stabilize MmPylRS and MbPylRS is still in need to improve current existing systems for the genetic incorporation of ncAAs. Chin et al. has recently shown that CMaPylRS can be combined with MmPylRS for the genetic incorporation of two different ncAAs at two separate codons when they are used together with an evolved *Ca. M. alvus* PylT (CmaPylT) that doesn't cross-interact with MmPylRS and *M. mazei* PylT (MmPylT) that doesn't cross-interact with CMaPylRS.[24] This type of combined uses of PylRS clones will also demand existing MmPylRS and MbPylRS clones to be improved. In this work, we wish to report our finding that a P188G mutation in MmPylRS can significantly stabilize MmPylRS for improved ncAA incorporation.

Material and Methods

Construction of plasmids encoding wild-type MmPylRS or a variant together with CmaPylT

The pEVOL-MmPylRS-CmaPylT plasmid was constructed by replacing CMaPylRS in the pEVOL-CMaPylRS-CmaPylT plasmid with megaWhop PCR. The MmPylRS fragment was amplified from the pEVOL-MmPylRS-mmPylT plasmid using two primers MmPylRS-for and MmPylRS-rev (Table 4). The amplified MmPylRS fragment was then used as a primer to amplify pEVOL-CMaPylRS-CmaPylT. Plasmids pEVOL-MmAcKRS1-CmaPylT, pEVOL-MmAcdKRS-CmaPylT, and pEVOL-MmAznLRS-CmaPylT were cloned in the same way.

Construction of plasmids encoding MmPylRS CTD, split MmPylRS CTD and NTD, and split MmPylRS CTD and sfGFP-Y66F-NTD

The split construct was made by PCR amplification of the whole construct pEVOL-MmPylRS-CmaPylT using two primers split-for and split-rev primer pairs that added a stop codon, a ribosome binding site, and a start codon between MmPylRS aa1–110 and aa111–454. The sfGFPY66F-NTD split construct was then made using megawhop PCR from the construct coding split CTD and NTD. The sfGFP fragment was amplified from the plasmid pBAD-sfGFP using a primer pair sfGFP-for and sfGFP-rev. The amplified fragment was then used as a primer to amplify the split construct. A site-directed mutagenesis was performed to the sequence-confirmed sfGFP-NTD split construct using two primers Y66F-for and Y66F-rev to generate a non-fluorescent sfGFP-Y66F-NTD. The plasmid encoding only the CTD domain was generated by PCR of pEVOL-MmPylRS-CmaPylT using a primer pair CTD-for and CTD-rev to remove the whole NTD region.

Construction of plasmids containing K110A, P188G, P118Q and double mutations in MmPylRS

Plasmid pEVOL-MmPylRS-K110A was made by PCR using a primer pair CTD-for and K110A-rev from a template pEVOL-MmPylRS-MmPylT. The pEVOL-MmPylRS-P188G-MmPylT and pEVOL-MmPylRS-P188Q-MmPylT was made from pEVOL-MmPylRS-MmPylT using primer pairs p188G-for/p188G-rev and p188Q-

for/P188Q-rev, respectively. Double mutations K110A/P188Q or P188G were constructed in a same way. The P188G mutated plasmids of MmAcKRS1, MmAcdKRS, and MmAznLRS were made from their corresponding constructs pEVOL-MmAcKRS1-MmPylT, pEVOL-MmAcdKRS-MmPylT and pEVOL-MmAznLRS-MmPylT by site-directed mutagenesis using two primers P188G- for and P188G-rev.

Purification of MmPylRS

Transformed cells were grown in 10 mL 2YT media (Ampicillin and Chloramphenicol) at 37 ^{o}C overnight. The overnight culture was inoculated into 1L of 2YT and grown at 37 ^{o}C until OD_{600} reached 0.6. The expression of MmPylRS was induced by the addition of 0.2% arabinose for 6 h at 37 ^{o}C. Cells were collected by centrifuging at 6 K rpm for 15 min. The cell pellet was resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 10% glycerol and 1 mM PMSF at pH 7.5). Suspended cells were lysed by sonication with 1 sec on and 4 sec off with a total of 5 min. The lysate was removed by centrifugation at 16 K rpm for 30 min at 4 ^{o}C. Ammonium sulfate precipitation was performed for the supernatant. Precipitation from 0% to 40% was collected by centrifuging at 16 K rpm for 30 min. The pellet was dissolved to buffer A (50 mM HEPES, 100 mM NaCl and 10% glycerol at pH 7.5) and dialyzed against buffer A overnight. Solubilized pellet was then filtrated through 0.45 mm filter and loaded onto a source 15Q column for purification by FPLC. The protein was eluted with a gradient from buffer A to buffer B (50 mM HEPES, 1 M NaCl and 10% glycerol at pH 7.5). The desired

peak was collected, concentrated, and loaded onto a superdex 200 Increase 10/300 GL column for further purification.

Quantification of amber codon suppression

Transformed cells were grown in 6 mL 2YT media at 37 ^{o}C until they reached OD_{600} as 0.6. Cells were then were induced with 0.2% arabinose and split to two 3 mL cultures induced with and without 1 mM NcAA. After 6 h of expression at 37 ^{o}C, cells were collected. 25 mL of the culture was diluted to 100 mL on a 96 well plate. OD_{600} was measured in a BioTek synergy neo2 plate reader. The remaining cells were spun down with 4 K rpm for 15 min. The pellet was collected and resuspend in 0.5 mL lysis buffer (50 mM Tris, 50 mM NaCl, 1 mM PMSF and 1 mg/mL lysozyme at pH 8). Cells were lysed by freezing and thawing between liquid nitrogen and 42 ^{o}C water bath with 20 sec vortexing after thawing. The lysate was clarified by centrifugation at 14 K rpm for 30 min. 100 uL of the supernatant was transferred to a Grenier half area black 96-well plate. The fluorescence of sfGFP was quantified using a BioTek synergy neo2 plate reader with 485 nm excitation and 510 nm emission. The supernatant was collected and incubated with 250 uL Ni-NTA resins for 10 min. The Ni resins were washed with 1 mL of wash buffer (50 mM Tris, 50 mM NaCl and 30 mM imidazole at pH8). The bound sfGFP was then eluted with elution buffer (50 mM Tris, 50 mM NaCl and 300 mM imidazole at pH8) after rocking in room temperature for 10 min. The eluent was then analyzed with SDS-PAGE.

Results and Discussion

Results and Discussion

CmaPylT enhances MmPylRS-mediated amber suppression

We started our research by testing the orthogonality of MmPylRS and CMaPylRS toward each other's cognate PylTs. Our original purpose was to examine whether these independent pairs can be combined to genetically encode two different ncAAs. To test orthogonality of MmPylRS toward CmaPylT, we constructed a plasmid pEVOL-MmPylRS-CmaPylT that encodes MmPylRS and CmaPylT (Figure 10A). This plasmid was used along with a reporter plasmid pBAD-sfGFP134TAG that encodes superfolder green fluorescent protein (sfGFP) with an amber mutation at the 134th amino acid coding position (sfGFP134TAG) to transform *E. coli* Top10 cells. Transformed cells were growth in the presence of 1 mM N$^{\varepsilon}$-Boc-lysine (BocK). Without amber suppression, the sfGFP134TAG gene was expressed as a truncated protein with a premature termination at the amber codon site. Otherwise, full-length sfGFP that was quantified by measuring fluorescence was expressed to indicate amber suppression levels. After 6 h expression, cells were collected and lysed to quantify expressed full-length sfGFP. A similar experiment in which a plasmid pEVOL-MmPylRS-MmPylT that encodes MmPylRS and MmPylT was used as a positive control. MmPylRS showed an amber suppression effect in the presence of CmaPylT slightly more than twice better than that in the presence of MmPylT (Figure 10B). To explore whether this enhanced expression can be transferred to other MmPylRS variants, we paired CmaPylT with engineered MmPylRS variants including MmAcKRS1, MmAcdKRS, MmAznLRS, and MmBuKRS (Mm was added in

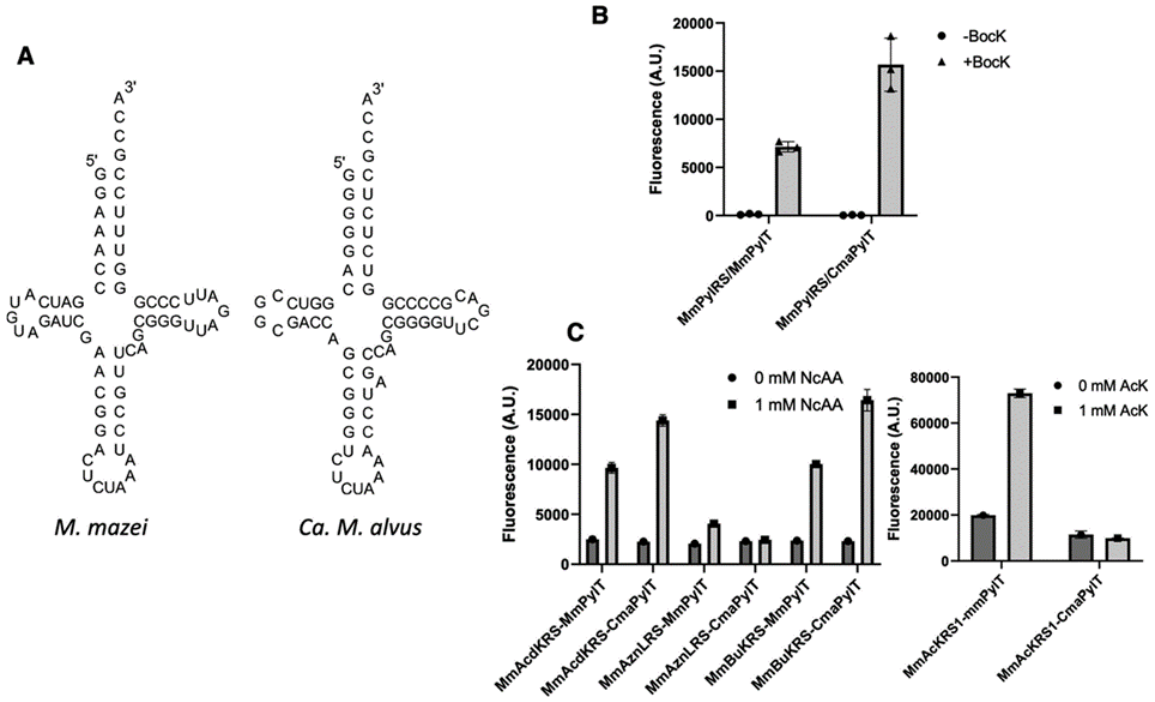

Figure 10 (A) Cloverleaf structures of MmPylT and CmaPylT. (B) The amber suppression efficiency of MmPylRS in the presence of CmaPylT is significantly better than that in the presence of MmPylT (p-value < 0.05). (C) CmaPylT doesn't lead to improved activities in all engineered MmPylRS variants. 1 mM ncAA was used to induce full-length sfGFP expression (p-value < 0.05 except the one from AznLRS-CmaPylT in the presence of AznL was 0.3). Reprinted with permission from reference 22.

the names for these variants to indicate their origin) that were previously evolved for the genetic incorporation of AcK, N^{ε}-(4-azidobenoxycarbonyl)-δ, ε -dehydrolysine (AcdK), azidonorleucine (AznL),and N^{ε}-crotonyl-lysine (CrK) respectively.[28-30] Plasmids similar to pEVOL-MmPylRS-CmaPylT that encode different MmPylRS variants and Cma-PylT were constructed. They were used along with pBAD-sfGFP134TAG to transform Top10 cells for the expression of full-length sfGFP in the presence of 1 mM corresponding ncAA.

Their corresponding MmPylRS variant-MmPylT pairs were used as positive controls. Among four ncAAs, only AcdK and CrK showed enhanced

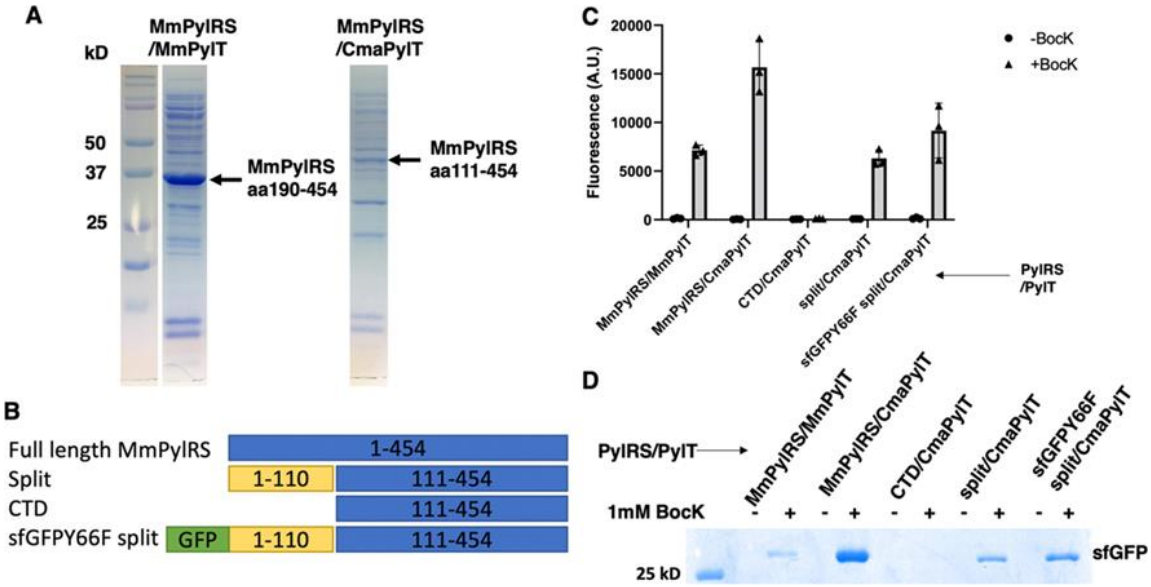

Figure 11 (A) MmPylRS showed different truncated fragments in the presence of MmPylT and CmaPylT. Arrows indicate truncated fragments aa190-454 and aa111-454. (B) Different MmPylRS clones used in combination with CmaPylT for testing amber suppression efficiency. Split means that CTD and NTD were expressed separately (p-value < 0.05). (C) Amber suppression efficiency of different MmPylRS clones in the presence of CmaPylT. (D) Purified full-length sfGFP that was expressed when different MmPylRS clones were combined with CmaPylT. 1 mM BocK was used to induce full-length sfGFP expression. Reprinted with permission from reference 22.

incorporation when their specific MmPylRS variants were paired with CmaPylT. AcK and AznL exhibited reduced incorporation (Figure 10C). Therefore, it is evident that the binding of CmaPylT and its aminoacylation by MmPylRS are two intricately related processes. The binding of a ncAA and its catalysis to form an aminoacyl-CmaPylT might require unique structural dynamics in the binding between MmPylRS and CmaPylT that

varies between different ncAAs. Therefore, not all engineered MmPylRS variants work well with CmaPylT.

CmaPylT prevents MmPylRS from its cleavage after A189.

To investigate the role of CmaPylT in the MmPylRS-mediated amber suppression enhancement, we lysed cells that expressed MmPylRS in the presence of either CmaPylT or MmPylT and analyzed soluble proteins by SDS-PAGE (Figure 11A). In the presence of MmPylT, a truncated soluble MmPylRS fragment whose molecular weight was around 30 kDa was clearly visible. However, this soluble MmPylRS fragment disappeared when CmaPylT was co-expressed. In the presence of CmaPylT, a soluble protein band with a molecular weight around 38 kDa appeared in the SDS-PAGE gel. We purified both soluble proteins and characterized their molecular weights using electrospray ionization mass spectrometry (ESI-MS). The deconvoluted molecular weights of the two proteins are 30,673.3 and 38,335.0 Da, respectively (Figure 28). MmPylRS is known to undergo hydrolysis in *E. coli* and only its CTD is soluble. These two determined molecular weights match well with MmPylRS fragments aa190-454 (theoretical molecular weight: 30672.2 Da) and aa111-454 (theoretical molecular weight: 38,336.7 Da). The MmPylRS NTD domain is aa1-189. This domain is known to bind PylT but how this domain folds is only partially studied. Although the crystal structure of aa1-100 has been determined,[39] the structure of aa101-189 is unknown. Our results indicate that aa111-189 is highly likely an independently folded domain and this domain interacts directly with CmaPylT. This interaction stabilizes the MmPylRS CTD fusion with aa111-189 and consequently

contributes to a strong activity of MmPylRS in the presence of CmaPylT. However, how exactly CmaPylT prevents the cleavage of MmPylRS after A189 needs to be further investigated.

Full-length MmPylRS is essential for an enhanced activity

To examine effects of different domains on the MmPylRS activity, we constructed plasmids encoding MmPylRS fragments and CmaPylT. Three designs were conducted. One contained just the MmPylRS CTD, the second had two fragments aa1-110 and aa111-454 expressed separately, and the third encoded independently expressed aa111-454 and aa1-110 that was fused with sfGFP with a 66F mutation (Figure 11B). A previous publication has shown that two split fragments from a PylRS chimera can be co-expressed to improve genetic incorporation of a ncAA.[20] We hoped to recapitulate this result. Two independently expressed CTD and NTD are also observed in *Desulfitobacterium hafinience* and other organisms.[12] In these organisms, CTD, NTD and PylT form a tertiary complex and NTD is required for the *in vivo* activity of the whole complex. sfGFP is known to improve solubility and folding of its fused proteins.[75, 76] The purpose of its fusion to aa1-110 was to improve the solubility of aa1-110 in *E. coli* so that a more soluble complex between aa111-454 and aa1-110 could be generated to improve genetic incorporation of a ncAA. To prevent fluorescence of the fused sfGFP from interfering with the characterization of fluorescence-based amber suppression levels, we mutated its active site Y66 to F, which eliminates its fluorescence[76]. aa1-110 instead of aa1-189 was used since CmaPylT stabilizes the MmPylRS CTD fused with aa110-aa189. The plasmid that encoded MmPylRS and MmPylT was used as a control. These plasmids along with pBAD-

sfGFP134TAG were used to transform *E. coli* Top10 cells. Transformed cells were grown in the presence of 1 mM BocK for the characterization of amber suppression levels. As shown in Figure 11C, removing the NTD domain from MmPylRS totally killed its amber suppression activity even in the presence of CmaPylT. No expressed full-length sfGFP was detected (Figure 29). Cells expressing split NTD and CTD domains resumed partially amber suppression activity. Although this activity was close to the activity from the MmPylRS-MmPylT pair, it was only about 40% of that from the MmPylRS-CmaPylT pair. The fusion of sfGFP-Y66F to NTD improved the overall amber suppression activity. However, this improved activity was only about 60% of that from the MmPylRS-CmaPylT pair. Collectively, our results approve that the expression of full-length MmPylRS is critical for a high amber suppression activity even when it is coupled with CmaPylT. Since we did not observe an obvious band for full-length MmPylRS in the presence of either MmPylT or CmaPylT in our SDS-PAGE analysis of the cell lysate supernatants as shown in Figure 11A, we suspected that full-length MmPylRS in the presence of CmaPylT might aggregate during the cell lysis process. To confirm this, we directly dissolved cells expressing MmPylRS and CmaPylT in the SDS-PAGE sample loading buffer, boiled the sample and analyzed proteins in whole cells by SDS-PAGE. As shown in Figure 30, the SDS-PAGE gel showed a heavy band corresponding to full-length MmPylRS. However, this band was not observable in cells expressing MmPylRS and MmPylT. Therefore, it is clear that CmaPylT stabilizes full-length MmPylRS. This stabilization likely contributes to the overall high activity of MmPylRS in its mediated BocK incorporation at amber codon in the presence of CmaPylT. How exactly CmaPylT

stabilizes MmPylRS will demand more structural analysis of the interactions between them.

K110A, P188Q and P188G enhance the activity of MmPylRS

Since the activity enhancement of MmPylRS by CmaPylT cannot be simply transferred to engineered MmPylRS variants, we sought other ways to improve activity of engineered MmPylRS variants that can be used together with MmPylT. Our investigation of MmPylRS stabilization by CmaPylT showed that MmPylRS is cleaved after K110 and A189 in *E. coli*. These two cleavages are presumably done by *E. coli* proteases. K110 is a lysine, which is a typical protease recognition site.[77] A189 is not a unique protease recognition site.[78] However, its N-terminal proximal residue is P188. Proline and glycine are known as protein secondary structure breakers.[79] P188 Mutations likely generate large structure changes that disrupt protease recognition at A189. Therefore, we decided to conduct mutagenesis at K110 and P188. To neutralize the charge of K110, we introduced a K110A mutation and for P188 we introduced two mutations P188G and P188Q. The P188G mutation was introduced with the purpose to maintain a possible turn structure and P188Q was introduced to improve the solubility of the P188-containing peptide. Clones with three separate single mutations and that with double mutations including K110A/P188G and K110A/P188Q were constructed. These mutated clones were coupled with MmPylT to test their amber suppression activity. For comparison, we only constructed the K110A mutant clone of MmPylRS for its coupled use with CmaPylT. Only the K110A mutation was introduced because CmaPylT prevents the MmPylRS

cleavage after A189. Plasmids coding these mutants along with either MmPylT or CmaPylT were built. There were coupled with pBAD-sfGFP134TAG to transform *E. coli* Top10 cells. Transformed cells were grown in the presence of 1 mM BocK for the characterization of overall full-length sfGFP expression levels. The K110A mutation improved the activity of MmPylRS in the presence of CmaPylT for about 10% but this improvement was not significant (Figure 31). An SDS-PAGE analysis of cells expressing MmPylRS K110A and CmaPylT showed an enhanced level of full-length MmPylRS in the cell lysate pellet compared to cells expressing wild-type MmPylRS and CmaPylT (Figure 32). Cells expressing MmPylRS K110A and CmaPylT had also close to non-visible truncated MmPylRS fragement aa111-454 in the cell lysate supernatant. On the contrary, cells expressing wild-type MmPylRS and CmaPylT had a high level of truncated MmPylRS fragment aa111-454 (Figure 32). In the presence of MmPylT, all three mutations significantly improved the activity of MmPylRS (Figure 12A). Although clones with double mutations showed higher activities than wild-type MmPylRS, their activities were lower than that from single mutation clones. It is obvious that mutations at the two cleavage sites influence the MmPylRS activity. Although our results showed that tested mutations at two sites cannot be synergistically used likely due to their combined disruption to the overall protein folding, it is highly possible that other combinations of mutations at two sites will lead to better activities in afforded enzyme variants than single mutation clones.

The enzyme activity enhancement effect of P188G can be transferred to engineered MmPylRS variants

To explore whether we can transfer the activity improvement effect of the P188 mutation to engineered MmPylRS variants, we introduced P188G into MmAcKRS1, MmAcdKRS and MmAznLRS, the three MmPylRS clones that we have frequently used to express proteins with posttranslational lysine modifications. Plasmids that encode these clones with the P188G mutation and MmPylT were constructed. They were used together with pBAD-sfGFP134TAG to transform *E. coli* Top10 cells. Transformed cells were grown in the presence of 1 mM corresponding ncAAs. As shown in Figure 12B, the P188G mutation led to 3-fold improvement of AcK incorporation by MmAcKRS1. The genetic

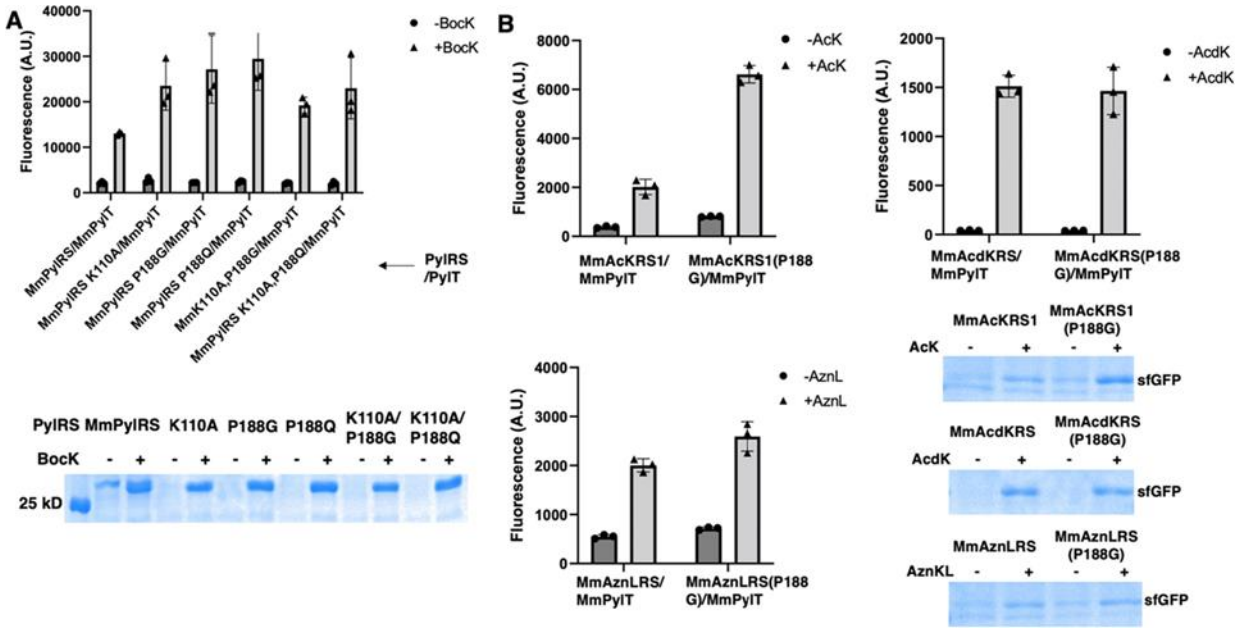

Figure 12 (A) Amber suppression efficiency for different MmPylRS mutants in the presence of MmPylT (p-value < 0.05). (B) Transfer of the P188G mutation to MmAcKRS1, MmAcdKRS and MmAznLRS to improve their amber suppression efficiency in the presence of MmPylT. 1 mM ncAA was used for expression (p-value < 0.05). Reprinted with permission from reference 22.

incorporation of AcK by amber codon has been broadly used for the characterization of posttranslational lysine acetylation functions. However, the approach has suffered from a low incorporation yield. To improve AcK incorporation, we have typically used 5 or 10 mM AcK for expression.[80] The dramatic improvement of AcK incorporation by a single P188G mutation will likely promote broader adoption of the technique by removing a significant hurdle. We have also tested the synthesis of histone H3 K36ac using the MmAcKRS1 P188G clone in combination with MmPylT. In comparison to almost non-detectable H3 K36ac that was expressed using the original MmAcKRS1, MmAcKRS1 P188G led to a high expression level of H3 K36ac in the presence of 5 mM AcK. The 188G mutation did not change the MmAcdKRS activity very much. It led to an arguably same activity as the original MmAcdKRS. Since MmAcdKRS is an efficient enzyme, this is not a surprise. The P188G mutation led to a better performing MmAznLRS. However, the activity improvement in MmAznLRS is not as significant as that was shown in MmAcKRS1. The aminoacylation catalysis requires coordination between aminoacyl-tRNA synthetase and several substrates including ATP, mino acid and tRNA. How exactly a mutation will influence interactions between the enzyme and three substrates and the catalytic process itself is difficult to perceive. However, based on all our collected data, we can conclude that the P188G mutation will likely improve activity of an MmPylRS variant although exact effects are hard to predict.

Conclusion

CmaPylT stabilizes MmPylRS in *E. coli* by preventing its cleavage after A189 although MmPylRS is still prone to the hydrolysis after 110. Mutations at K110 and P188 in general improve the activity of MmPylRS in the presence of MmPylT. The P188G mutation can also be potentially transferred to engineered MmPylRS variants to improve their activities for the coupled use with MmPylT.

CHAPTER III *

DRUG REPURPOSING FOR THE SARS-COV-2 PAPAIN-LIKE PROTEASE

This Chapter is reprinted with the permission of Chia-Chuan Cho, Shuhua G. Li, Tyler J. Lalonde, Kai S. Yang, Ge Yu, Yuchen Qiao, Shiqing Xu, and Wenshe Ray Liu *ChemMedChem* **2022** 17 (1), e2020100455 DOI 10.1002/cmdc.202100455. [81]

Introduction

The current prevailing COVID-19 pandemic has caused global health and economic consequences that are often compared to that of the 1918 influenza pandemic.[82] As of June 1st, 2021, the total number of confirmed global COVID-19 cases has exceeded 170 million, of which more than 3.5 million have succumbed to death (WHO data). Encouragingly, three COVID-19 vaccines developed by Pfizer/BioNtech, Moderna, and Johnson & Johnson have been approved by FDA for human immunization in the United States. Although vaccines are promising in containing the pandemic, their availability does not diminish the urgent need for other effective antiviral drugs. Existing COVID-19 vaccines are targeting the membrane Spike protein of SARS-CoV-2, the pathogen of COVID-19. Spike is highly mutable.[52] New viral strains with critical mutations in Spike such as UK and South Africa SARS-CoV-2 strains have emerged. Effectiveness of vaccines against these new strains will be a to effective antivirals for COVID-19, repurposing of existing drugs has been broadly conducted. This effort has led to the

identifications of a number of potential antivirals for SARS-CoV-2.[83] So far, remdesivir is the only antiviral that has been approved for the treatment of COVID-19. Remdesivir is a nucleoside analog that inhibits RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2. It shows high potency in inhibiting SARS-CoV-2 in vitro but appears to be only modestly effective in treating COVID-19 patients.[84] In view of the ongoing pandemic in its colossal scale and apocalyptic impact, there is a dire need of more effective antivirals with novel mechanisms of action to save lives of COVID-19 patients.

As the pathogen of COVID-19, SARS-CoV-2 has a positive-sensed genomic RNA.[85] It encodes 10 open reading frames (ORFs). As the largest ORF, ORF1ab comprises more than two thirds of the whole genome. In an infected host cell, ORF1ab is translated to two large polypeptide products, pp1a (~ 500 kDa) and pp1ab (~ 800 kDa),[86] by the host protein translation system. The pp1ab polypeptide is formed due to a frameshift during protein translation. Both pp1a and pp1ab need to undergo proteolytic cleavage to form 15 mature proteins. These mature proteins are nonstructural proteins (nsps) essential for the virus in its reproduction and pathogenesis. The proteolytic cleavage of pp1a and pp1ab is an autocatalytic process. Two internal polypeptide regions, nsp3 and nsp5, possess cysteine protease activities that cleave themselves and all other nsps in pp1a and pp1ab. Nsp3 is commonly referred to as papain-like protease (PL^{Pro}) and nsp5 as main protease (M^{Pro}).[87] Both PL^{Pro} and M^{Pro} are essential to SARS-CoV-2. Of the two enzymes, PL^{Pro} recognizes the tetrapeptidyl LXGG motif juxtaposed between viral proteins nsp1 and nsp2, nsp2 and nsp3, and nsp3 and nsp4. Its cleavage after the LXGG motif leads to

the release of nsp1, nsp2, and nsp3 which are essential for host modulation and viral replication.[88] In addition, recent studies have shown that PL^{Pro} can proteolytically remove K48-crosslinked ubiquitin concern.[52] Vaccines are also preventative making them not an option for the treatment of COVID-19 patients. For quick access (Ub) and interferon-stimulated gene 15 product (ISG15) that play important roles in the regulation of host innate immune responses to viral infection.[57, 89] PL^{Pro} is also conserved across various coronaviruses.[90] For these reasons, PL^{Pro} is considered an attractive COVID-19 drug target for the development of antivirals that may be potentially used as broad-spectrum inhibitors for other coronaviruses as well.[91, 92]

In view of the urgent need of antivirals for effective COVID-19 treatments, repurposing approved drugs and late-stage clinical drug candidates against SARS-CoV-2 is an efficient strategy. Significant progress has been made in drug repurposing for M^{Pro}. The first orally administered SARS-CoV-2 M^{Pro} inhibitor PF-07321332 that was developed by Pfizer is currently undergoing evaluation in a Phase 1b multi-dose study in hospitalized COVID-19 participants. In contrast, inhibitors for PL^{Pro} remain relatively less explored because PL^{Pro} is comparably a more challenging drug target. It has a flatter active site pocket than M^{Pro}. Since PL^{pro} is both a cysteine protease and a deubiquitinase, we reasoned that screening of cysteine protease inhibitors and deubiquitinase inhibitors would provide fast identification of PL^{Pro} inhibitors as initial hits for further optimization. In this study, we report our progress in identifying PL^{Pro} inhibitors by screening 33 deubiquitinase and 37 cysteine protease inhibitors. Among these molecules, we identified seven

candidates that can potently inhibit PL^{Pro} with an IC_{50} value below 40 μM. Four molecules have an IC_{50} value below 10 μM. One molecule SBJ2-043 can only partially inhibit PL^{Pro} but has a remarkable low IC_{50} value of 0.56±0.16 μM. More in-depth investigation of these inhibitors in their mechanisms of action will likely reveal unique inhibition mechanisms for PL^{Pro} for advanced development of potent PL^{pro} inhibitors.

Material and Methods

Chemicals

The deubiquitinase inhibitor library and cysteine protease library were purchased from MedChemExpress. Two other deubiquitinase inhibitors, C527 and VLX1579, were purchased from Cayman Chemicals. The PL^{Pro} substrate, Z-LRGG-AMC, was purchased from Cayman Chemicals.

In-silico molecular docking study

Protein preparation: The crystal structure of the SARS-CoV-2 PL^{Pro} in complex with inhibitor GRL0617 (PDB ID: 7CMD) was obtained from the RCSB Protein Data Bank (https://www.rcsb.org/). Only chain A of the homo-tetrameric structure was used in our docking study. The cognate ligand was removed from the structure that was applied in the following redocking test. In AutoDockTools-1.5.7, water molecules were deleted, and polar hydrogens were added to the structure. Finally, the prepared protein structure was converted into a PDBQT file for further docking studies.

Ligand preparation: Compounds in SDF format were converted into PDB files using OpenBabel-3.1.1. These PDB files were further converted into PDBQT files via AutoDockTools-1.5.7 with default options.

Docking parameters and method: In addition to receptor and ligand preparations, AutoDockTools-1.5.7 was also used for grid parameter setting. The cognate ligand of crystal structure 7CMD suggested the inhibitor binding site. A grid box with dimensions 30 × 30 × 30 centered at the coordinates X = 30.0, Y = 12.0, and Z = 30.0 was used to represent the search space. Then we applied AutoDock Vina docking protocol with options of 20 CPUs to use and maximum 100 binding modes to generate. Only modes with the highest binding affinities were shown in this paper.

Docking validation: We redocked the cognate ligand GRL0617 into its cognate receptor in order to verify the correctness of our docking method and parameters. The top row of Figure 7A showed the docked conformation in magenta and the crystal conformation in grey. The superposition of two conformations demonstrated the feasibility of our molecular docking method.

Expression and purification of PL^{Pro}

PL^{Pro} gene fragment was purchased from Integrated DNA Technologies (IDT), amplified with the primer pair that has XbaI and XhoI cleavage sites. The amplified gene was inserted into pBAD-sfGFP vector. In the final construct (pBAD-sfGFP-PL^{Pro}), a

hexahistidine tag is located at the N-terminus of sfGFP. A TEV cutting site is placed after the C-terminus of sfGFP which is followed by PL^{Pro}. Chemically competent Top10 *E. coli* cells were transformed with the pBAD-sfGFP- PL^{Pro} plasmid. Then a single colony was picked, inoculated into 10 mL of 2YT with ampicillin and grew overnight at 37 °C. The overnight culture was inoculated to 1 L of 2YT with 100-fold dilution and grew until OD_{600} (optical density at 600 nm) reached 0.6. Then, the medium was cooled down to 20 °C and induced with 0.2% L-arabinose for 24 h at 20 °C. The cells were collected by centrifuging at 6000 rpm for 30 mins and resuspended in lysis buffer (20 mM Tris, 250 mM NaCl, 5% glycerol, 0.2% TritonX-100 and 1 mg/mL lysozyme at pH 7.8). Collected cells were lysed by sonication at 60% amplitude with 1 sec on 4 sec off for a total of 5 min. The cell debris was removed by centrifuging at 16000 rpm for 30 min and the supernatant was loaded onto 3 mL of Ni-NTA resins (Genscript). A gravity flow Ni-NTA chromatography was performed. The Ni-NTA resin was washed with 10 times resin bed volume of wash buffer (20 mM Tris, 250 mM NaCl and 10 mM imidazole at pH 7.8) and eluted with 20 mL elution buffer (20 mM Tris, 30 mM NaCl and 300 mM imidazole at pH 7.8). The buffer was changed to 20 mM Tris and 30 mM NaCl with 20% glycerol by using HiPrep 26/10 desalting column. Finally, the protein was aliquoted, flash frozen with liquid nitrogen and stored in –80 °C freezer.

Expression and purification of Ub-G76 C-6H

The Ub-G76 C-6H was expressed according to the protocol described before.[93] Briefly, an overnight starting culture was inoculated in 2YT medium using a 1:100 dilution ratio. The cells were grown at 37 °C with 250 rpm. The protein was induced with 1 mM of IPTG overnight at 18 °C when OD_{600} reached 0.7–0.9. The next morning, cells were harvested by centrifugation at 6000 rpm for 30 min and lysed by sonication in lysis buffer (50 mM NaH_2PO_4, pH 7.8, 500 mM NaCl, 5 mM imidazole, 1 mM TCEP). The whole cell lysate was then clarified by centrifugation (16000 rpm, 30 min, 4 °C). The supernatant was collected and 6 M HCl was added dropwise until it reached pH 2. The acidic supernatant was stirred for 5–10 min to allow impurities to precipitate. The impurities were removed by centrifugation (10000 rpm, 30 min, 4 °C) and the supernatant was neutralized to pH 7.8 using 6 M NaOH solution. Subsequently, the neutralized supernatant was loaded onto Ni-NTA resins to perform a gravity flow chromatography. The resin was washed with wash buffer (50 mM NaH_2PO_4, 500 mM NaCl, 30 mM imidazole, 1 mM TCEP at pH 7.8). Then the target protein was eluted with elution buffer (washing buffer substituting 20 mM imidazole with 300 mM imidazole). The elution was desalted to 50 mM ammonium bicarbonate (ABC) buffer using a HiPrep Desalting column and lyophilized into 100 nmol per aliquot for further use.

Synthesis of Ub-AMC

Ub-AMC was synthesized using the Activated Cysteine-directed Protein Ligation (ACPL) technique described before.[93] 100 nmol of lyophilized Ub-G76 C-6H was dissolved in 194.5 μL of 50 mM borate buffer (pH 9) supplied with 0.5 μL of 500 mM

TCEP stock solution. It was then mixed with 300 μL of 40 mM Gly-AMC solution in DMSO to achieve a reaction mixture containing 24 mM Gly-AMC, 20 mM borate and 60% DMSO. The ACPL reaction was initiated by adding 5 μL of 500 mM NTCB stock solution in DMSO and the pH value was carefully adjusted to 9 using 6 M NaOH solution. The reaction was performed at 37 °C for 16 h and terminated by desalting to 50 mM ABC buffer using a HiTrap Desalting column. The desalted sample was further purified using Ni-NTA purification by collecting the flowthrough. The purified Ub-AMC was analyzed by Orbitrap ESI-MS, then flash frozen and stored in 80 °C freezer.

Screening assay

To have a preliminary test of inhibition ability of each drug, a screening assay was performed. 50 μL reaction solutions with 40 nM of PL^{Pro} and 400 μM of each drug (deubiquitinase inhibitors and cysteine protease inhibitors) were preincubated in PL^{Pro} assay buffer (20 mM Tris, 300 mM NaCl, pH 7.5) with 4 mM DTT in 37 °C for 30 min. It was then mixed with 50 μL solutions with 100 μM of Z-LRGG-AMC. The final assay solution had 200 μM of each compound, 20 nM PL^{Pro}, 2 mM DTT and 50 μM Z-LRGG-AMC. The fluorescence of AMC that was generated as a result of PL^{Pro} enzymatic activity was recorded in a plate reader using 380 nm as the excitation wavelength and 440 nm as the emission wavelength. The initial slope of the fluorescence-vs-time graph for each drug was analyzed by calculating the slope of the 0–10 min curve. The calculation was done by using linear regression analysis in GraphPad 9.0.

Inhibition analysis

The inhibition assay was performed in the same way as the screening assay except a series of inhibition concentrations were used. All the reactions were performed in triplicates. Reaction rates that were normalized against the rate in the absence of an inhibitor were used to determine IC_{50} values using GraphPad 9.0.

Results and Discussion

PL^{Pro} is a cysteine protease with a classical Cys-His-Asp catalytic triad (Cys111, His272 and Asp286). It is responsible for processing three cleavage sites in the viral polypeptides pp1a and pp1ab to release nsp1, nsp2, and nsp3. Besides this function, PL^{Pro} has also deubiquitinase and deISGylase activities to help SARS-CoV-2 evade host immune responses.[57, 89, 94] Due to its both general protease and deubiquitinase nature, we screened existing deubiquitinase and cysteine protease inhibitors on their inhibition of PL^{Pro}. To express PL^{Pro} for experimental characterization of existing deubiquitinase and cysteine protease inhibitors, we constructed a PL^{Pro} vector for expression in *E. coli* (Figure 13A). In this construct, PL^{Pro} was fused to the C terminus of superfolder green fluorescent protein (sfGFP) that is known to stabilize a fused partner.[95] A 6×His tag was added to the N terminus of sfGFP for affinity purification using Ni-NTA resins. A TEV cleavage site was inserted between sfGFP and PL^{Pro} for the proteolytic removal of PL^{Pro} from sfGFP by TEV protease. During the purification and treatment of the sfGFP-PL^{Pro} fusion, we noticed that the cleaved PL^{Pro} quickly aggregates. Therefore, for long term storage of PL^{Pro}, we have chosen to save and use sfGFP-PL^{Pro} instead of PL^{Pro} for all our assays. In

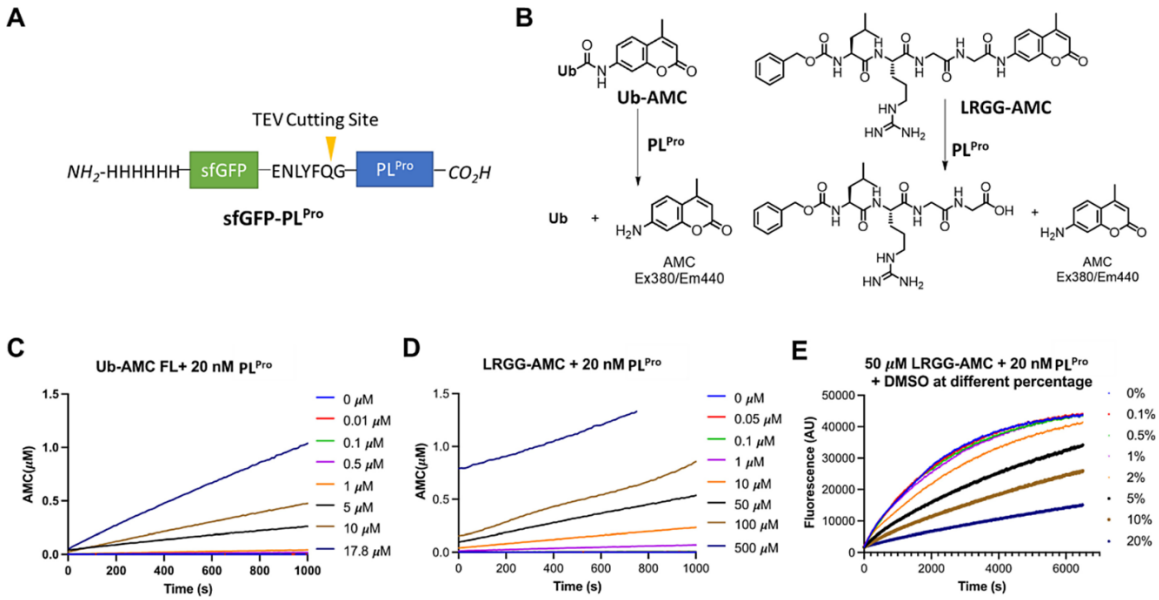

Figure 13 (A) The sfGFP-PL^{Pro} fusion protein whose expression in *E. coli* has been tested. (B) Two PL^{Pro} substrates and their catalytic release of AMC. (C) The catalytic release of AMC at various concentrations of Ub-AMC by 20 nM PL^{Pro}. (D) The catalytic release of AMC at various concentrations of Z-LRGG-AMC by 20 nM PL^{Pro}. (E) The catalytic release of AMC at 50 μM LRGG-AMC by 20 nM PL^{Pro} in the presence of various concentrations of DMSO. Reprinted with permission from reference 81.

order to test catalytic activities of PL^{Pro}, we synthesized a fluorogenic protein substrate Ub-AMC (AMC: 7-amino-4-methylcoumarin) using our recently developed activated cysteine-directed protein ligation technique and purchased a fluorogenic peptide substrate Z-LRGG-AMC (Figure 13B).[93] The hydrolysis of Ub-AMC and Z-LRGG-AMC releases AMC whose strong blue fluorescence can be detected by a fluorometer or a fluorescent plate reader. To obtain an optimal assay condition for inhibitor characterizations, both Ub-AMC and Z-LRGG-AMC are titrated at 20 nM PL^{Pro} (Figure 13C and 13D). Based on the product formation curves, we deemed that 5 μM Ub-AMC and 50 μM Z-LRGG-AMC provide easily detectable linear production formation within

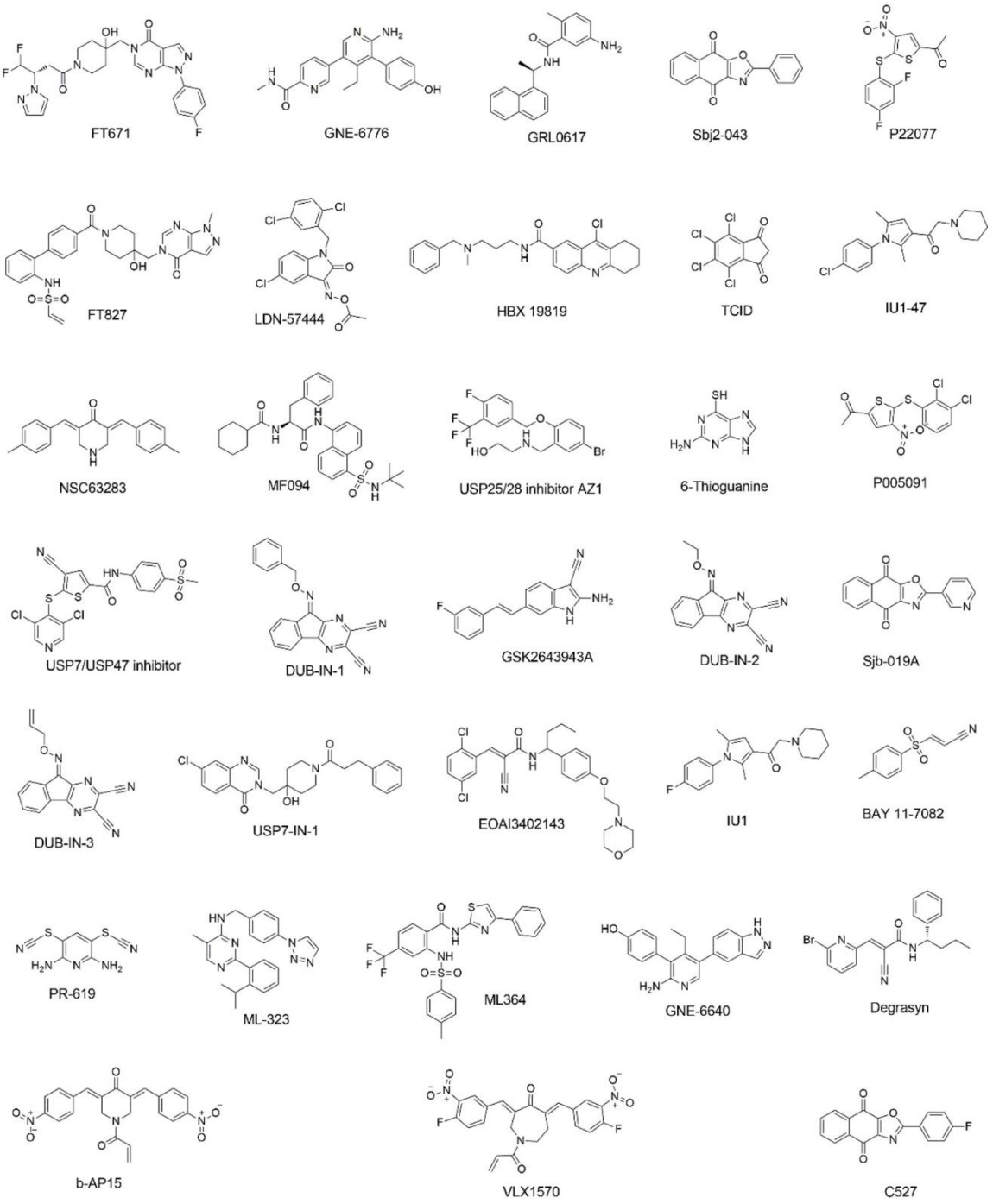

Figure 14 Structures of 33 selected deubiquitinase inhibitors. Reprinted with permission from reference 81.

600 s and therefore used these two conditions in our inhibitor characterization assays. We have also characterized the influence of DMSO to the PL^{Pro} catalysis since most commercial small molecules are provided as 10 mM DMSO stocks. The stability of PL^{Pro} in the presence of DMSO restricts the highest drug concentration we can test. We titrated DMSO from 0.1% to 20% (Figure 13E). It showed that PL^{Pro} had slightly reduced activity at 2% DMSO and this inhibition trend by DMSO increases with the DMSO concentration. Considering that most of commercial small molecules are provided as 10 mM DMSO stocks, their concentrations will be 200 μM when they are diluted by 50-fold leading to 2% DMSO in the final assay conditions. At 2% DMSO, PL^{Pro} remains almost 100% active and is good for quantifying drug inhibition effects. We initiated our search of PL^{Pro} inhibitors by testing 31 deubiquitinase inhibitors (Figure 14) and 37 cysteine protease inhibitors (Figure 15) that we purchased from MedChemExpress. These molecules were provided as 10 mM stocks in DMSO without further purification. Two additional deubiquitinase inhibitors C527 and VLX1570 were acquired from Cayman Scientific (Figure 14). In an initial screening assay, all acquired small molecules were analyzed at a 200 μM final concentration. In this assay, an inhibitor (400 μM) was incubated with 40 nM PL^{Pro}, 4 mM DTT, and 4% DMSO in a reaction buffer containing 20 mM Tris and 300 mM NaCl at pH 7.5 for 30 min. 50 μL of this incubation solution was then mixed with an equal volume of a substrate solution containing 100 μM Z-LRGG-AMC in the reaction buffer to initiate the catalytic release of AMC. The AMC fluorescence (Ex380/Em440) was recorded in a plate reader and the linear slope within the first 10 min was calculated

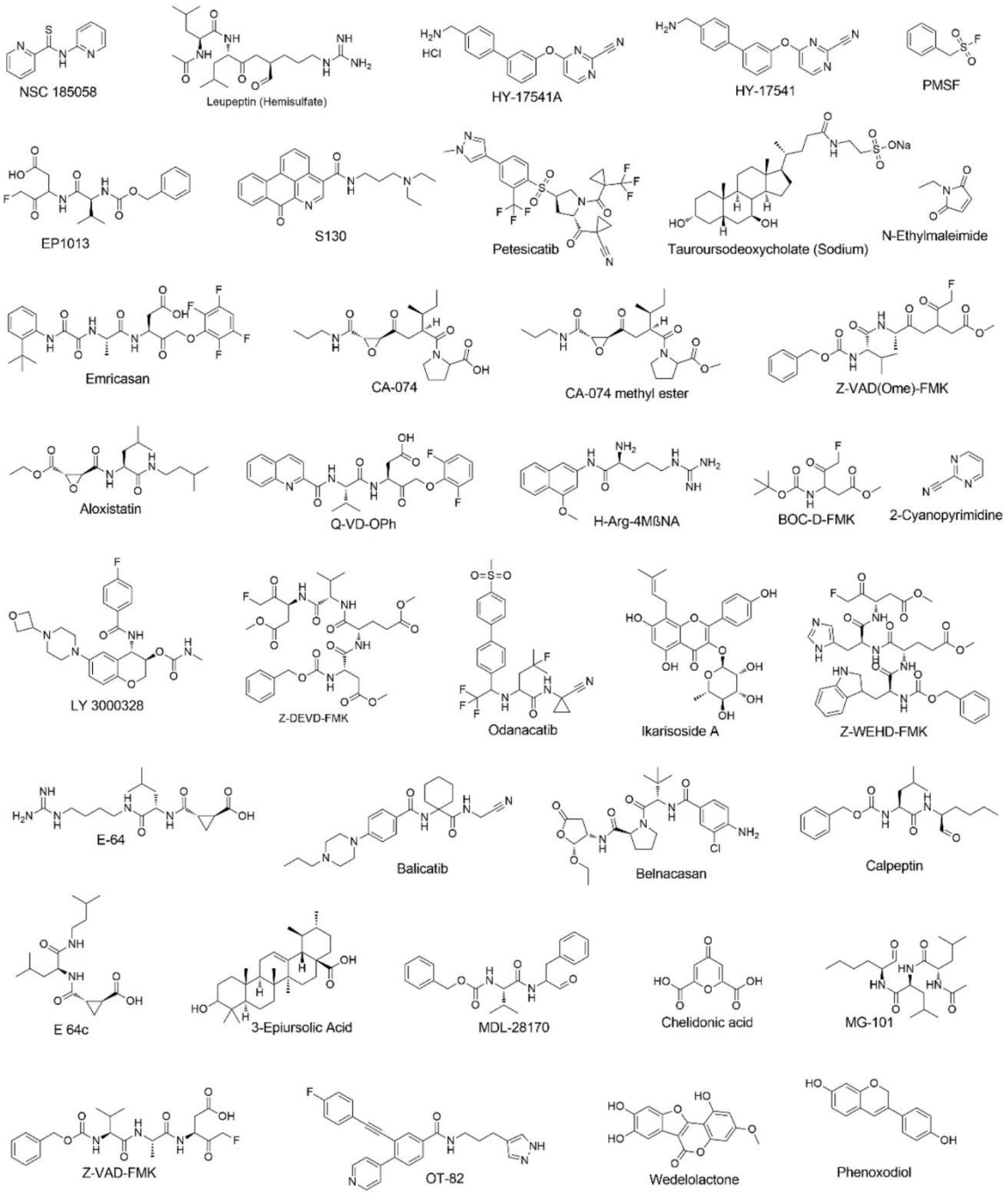

Figure 15 Structures of 37 selected cysteine protease inhibitors. Reprinted with permission from reference 81.

as the initial product formation rate. Among the deubiquitinase inhibitors at 200 μM, GRL0617 and TCID inhibited PL^{Pro} activity almost completely, PR-619 and ML364 had significant quenching effects leading to negative signals detected for the two compounds, and SJB2-043, P22077, LDN-57444, HBX19818, NSC632839, P005091, GSK2643943 A and SJB3-019 A inhibited PL^{Pro} partially but at significant levels (Figure 16A). In contrast, among 37 cysteine protease inhibitors, only S130 displayed strong inhibition of PL^{Pro} at the 200 μM level (Figure 16B and Table 2). The weak PL^{Pro} inhibition by most cysteine protease inhibitors is likely due to a much shallower active site in PL^{Pro} compared to other cysteine proteases.

For 16 small molecules that exhibited significant inhibition of PL^{Pro} at the 200 μM level (PR-619 and ML364 were selected as well), we further characterized their IC_{50} values. For IC_{50} characterization, AMC product formation rates at 20 nM PL^{Pro}, 50 μM Z-LRGG-AMC and varied concentrations of a small molecule were measured in triplicates. Data of the AMC product formation rate vs the small molecule concentration were fitted nonlinearly to a sigmodal inhibition equation to determine IC_{50} values for all compounds in GraphPad Prism 9. All data are presented in Figure 17. For PR-619 and ML364, we adjusted the initial rate by subtracting autofluorescence and adding photobleaching from a provided inhibitor. After this adjustment, we noticed that ML364 did not show strong concentration-dependent inhibition of PL^{Pro} but PR-619 strongly inhibited PL^{Pro} at concentrations above 1 μM. Among all tested compounds, seven has a determined IC_{50}

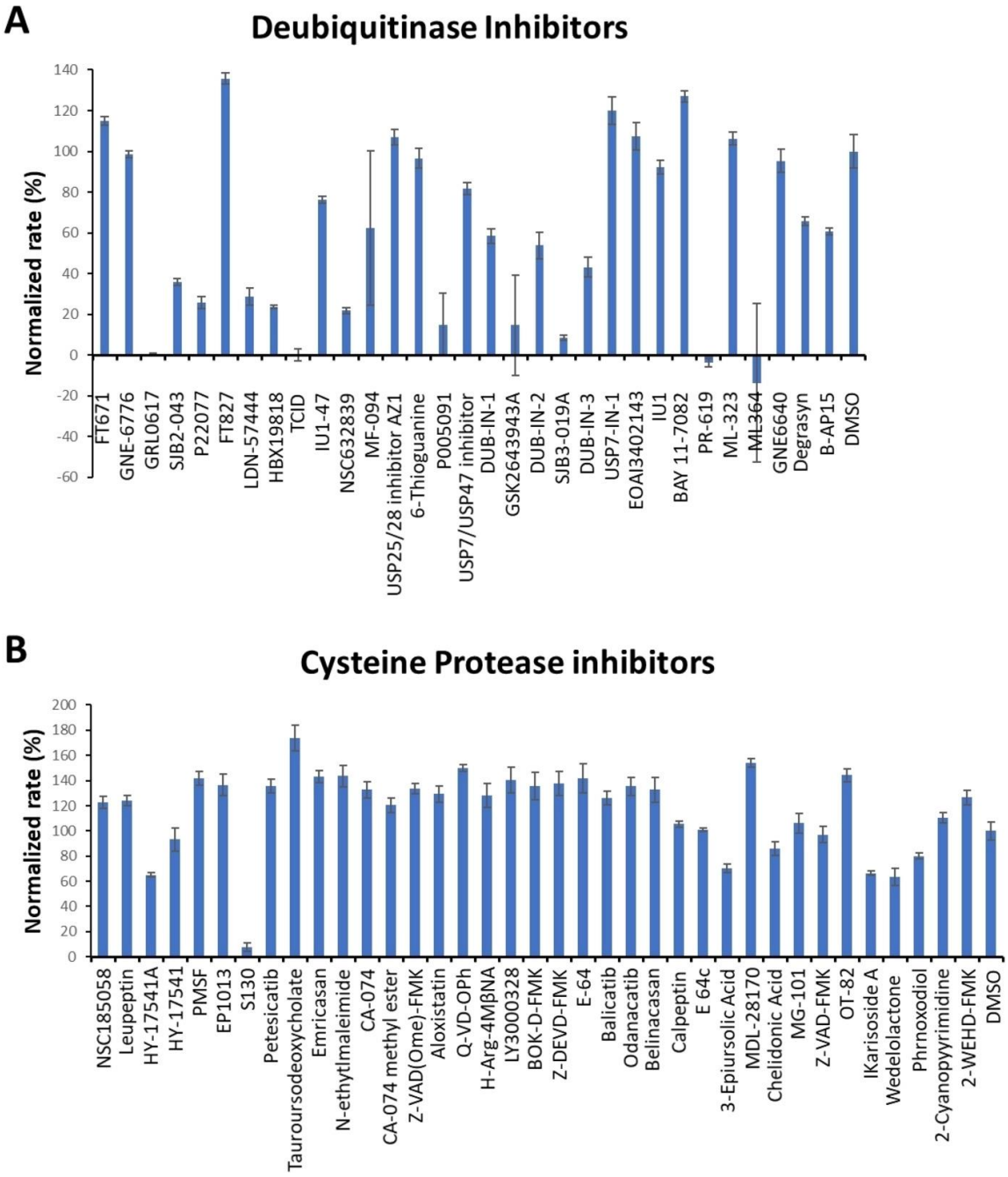

Figure 16 (A) Initial screening of PL^{Pro} inhibition by 33 deubiquitinase inhibitors. (B) Initial screening of PL^{Pro} inhibition by 37 cysteine protease inhibitors. Reaction conditions: 20 nM PL^{Pro}, 50 µM Z-LRGG-AMC, 200 µM inhibitor, 20 mM Tris-HCl, 300 mM NaCl, pH 7.5. Reprinted with permission from reference 81.

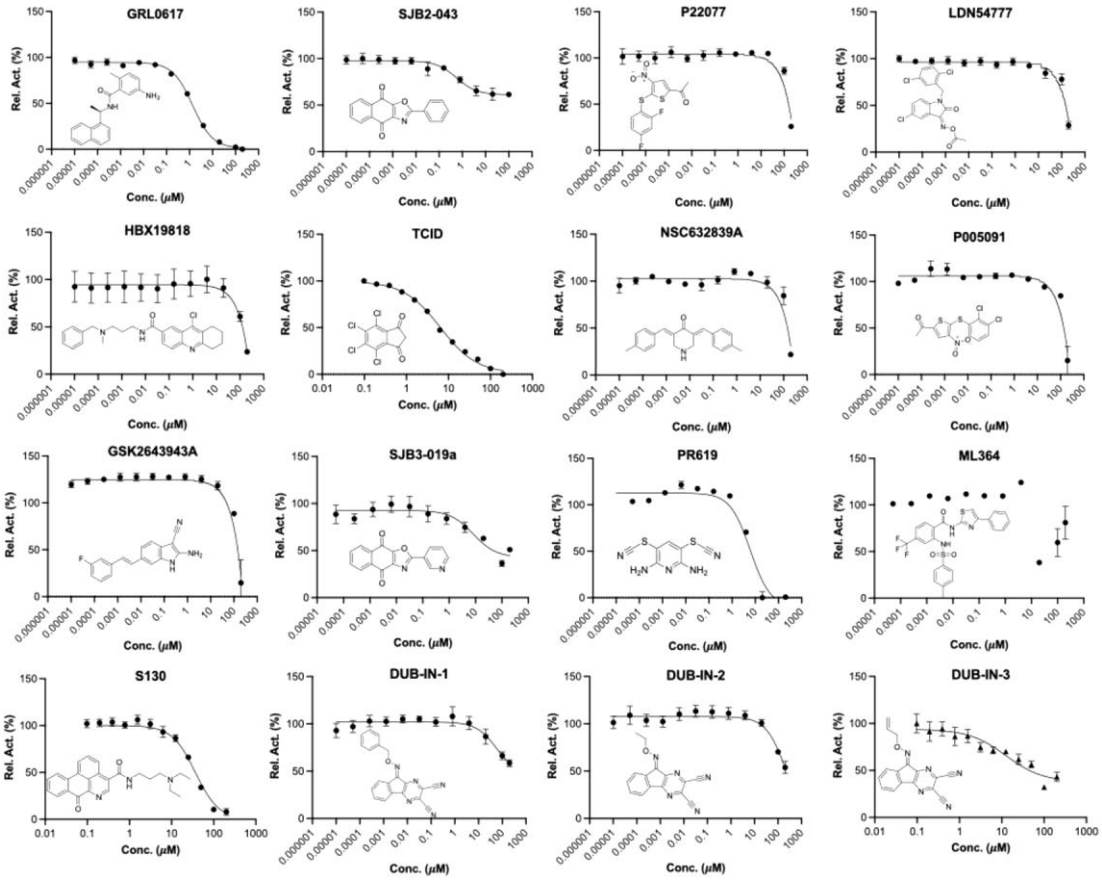

Figure 17 IC_{50} characterization for 16 small molecules on their inhibition of PL^{Pro} using 50 μM LRGG-AMC as a substrate. Experiments at different conditions were performed in triplicates. Reprinted with permission from reference 81.

value below 40 μM. These include SJB2-043 (0.56±0.16 μM), GRL0617 (1.37±0.10 μM), PR-619 (6.1±1.2 μM), TCID (6.42±0.32 μM), SJB3-019 A (8.15±3.59 μM), DUB-IN-3 (12.5±3.9 μM), and S130 (35.0±1.4 μM); see Table 1. GRL0617, TCID, PR-619 and S130 fully inhibited PL^{Pro} at 200 μM. SJB2-043, SJB3-019a and DUB-IN-3 achieved only partial inhibition of PL^{Pro} at 200 μM. GRL0617 is a naphthalene-based compound that was previously shown as a non-covalent inhibitor of PL^{Pro} from SARS-CoV with no significant inhibition of host proteases.[61, 96] PL^{Pro} proteins from SARS-CoV and SARS-

CoV-2 share 83% sequence identity and have a very high level of sequence and structural similarity in their substrate binding pockets.[57, 61, 92, 97] Recent studies have showed that GRL0617 can effectively inhibit PL^{Pro} of SARS-CoV-2.[89, 97, 98] In line with these recent studies, our results show that GRL0617 is effective at inhibiting PL^{Pro} from SARS-CoV-2 (IC_{50}: 1.4 μM). Given its strong potency and relatively small size, structure-activity relationship studies of GRL0617 will likely lead to more potent PL^{Pro} inhibitors with high antiviral activities. TCID is an inhibitor for ubiquitin C-terminal hydrolase L3 with a reported IC_{50} of 0.6 μM.[99] TCID has two ketones that are conjugated to a highly electron withdrawing tetrachlorobenzene. Both ketones are highly prone to hydration and reaction with nucleophiliccysteine residue in PL^{Pro}. It is likely that it interacts with PL^{Pro} covalently. A structural investigation of its complex with PL^{Pro} will serve as a starting point for the development of more potent inhibitors. PR619 is a reversible and cell-permeable inhibitor that broadly inhibits deubiquitinases.[100] The propensity of exchanging the cyano group with a protein thiolate makes PR619 likely to target cysteine proteases broadly. How exactly PR619 inactivates PL^{Pro} needs to be investigated further. S130 is an inhibitor that was originally discovered as an ATG4B inhibitor.[101] ATG4B functions similar to a deubiquitinase and hydrolyzes the preprotein of ATG8. Since mutating the catalytic cysteine in ATG4B does not significantly affect the binding of S130 to ATG4B, S130 likely interacts with ATG4B noncovalently. A similar mechanism to inhibit PL^{Pro} is also expected. S130 has a large aromatic moiety in which four rings are conjugated. It will be interesting to see how this extended large aromatic moiety interacts with PL^{Pro} to achieve strong binding. SJB2-043, SJB3-019a and DUB-IN-3 are three molecules that

inhibit PL^{Pro} only partially at 200 μM. SJB2-043 and SJB3-019a are two structurally similar compounds that inhibit USP1.[102] Both have two quinone oxygens that potentially interact with nucleophilic residues in PL^{Pro}. Since both molecules did not inhibit PL^{Pro} completely but displayed clearly measurable IC_{50} values, they likely bind to an allosteric site of PL^{Pro}. Further investigations of these two molecules in their mechanisms of inhibiting PL^{Pro} will likely reveal novel targeting sites for the development of PL^{Pro} inhibitors. SJB2-043 has the lowest IC_{50} value among all tested small molecules, indicating very strong binding to PL^{Pro}. A potential application of this strong binding is to use it to develop a proteasome targeting chimera for PL^{Pro}. This is a direction that we are actively exploring. DUB-IN-3 is a USP8 inhibitor that has two nitrile groups conjugated to a diazine.[103] The electron deficient nature of the diazine makes the two nitriles reactive toward nucleophilic residues in PL^{Pro}. It is possible that one of these two nitriles will hit on the catalytic cysteine in PL^{pro} to form a reversible covalent complex. The low Hill coefficient of the DUB-IN-3 inhibition curve also indicates a complicated inhibition mechanism.

For six deubiquitinase inhibitors that showed high potency in inhibiting PL^{Pro}-catalyzed AMC release from Z-LRGG-AMC, we have also characterized their inhibition of PL^{Pro} in its hydrolysis of Ub-AMC. Ub-AMC is much larger than Z-LRGG-AMC and

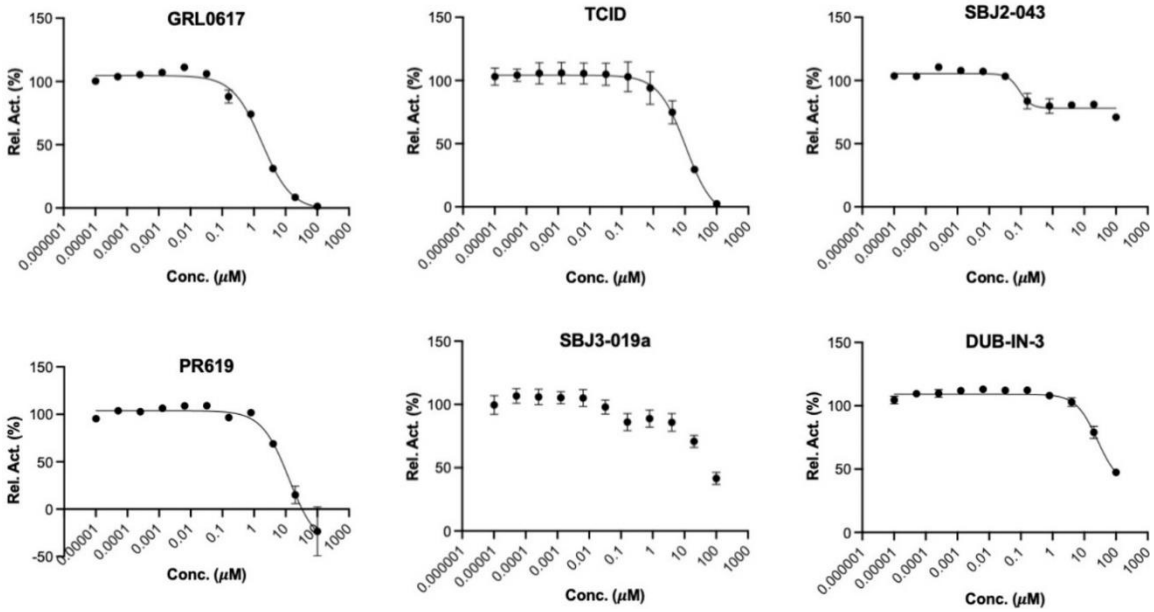

Figure 18 IC_{50} assays for 6 deubiquitinase inhibitors on their inhibition of 20 nM PL^{Pro} using 5 μM Ub-AMC as substrate. Experiments at different conditions were performed in triplicates. Reprinted with permission from reference 81.

involves a much bigger interface than Z-LRGG-AMC to interact with PL^{Pro}. For inhibitors that do not directly target the active site, they may display different inhibition characteristics when Ub-AMC instead of Z-LRGG-AMC is used as a substrate. We tested all six inhibitors in the presence of 20 nM PL^{Pro} and 5 μM Ub-AMC (Figure 18). When 5 μM Ub-AMC was used as a substrate, IC_{50} values of GRL0617, TCID and PR619 were determined as 1.80±0.21, 10.5±2.7 and 12.9±2.4 μM, respectively. These IC_{50} values are higher than but comparable to IC_{50} values determined when 50 μM Z-LRGG-AMC was used as a substrate, indicating that all three inhibitors are likely involved in a similar mechanism in inhibiting PL^{Pro}. GRL0617 is known to bind the PL^{Pro} active site. It is highly likely that TCID and PR619 bind to the PL^{Pro} active site as well. When Ub-AMC was used as a substrate, SJB2-043 had a determined IC_{50} value of 0.091±0.024 μM. This IC_{50} value

is much lower than the determined IC_{50} value when Z-LRGG-AMC was used as a substrate. Similar to the observed pattern when Z-LRGG-AMC was used as a substrate, SJB2-043 did not inhibit PL^{pro} completely. It is obvious that SJB2-043 behaves very differently from GRL0617, TCID and PR619. It binds likely to an allosteric site of PL^{Pro}. This allosteric binding influences apparently the catalytic hydrolysis of Ub-AMC more than Z-LRGG-AMC. This observation is possibly attributed to the much larger binding interface of Ub-AMC that responds more sensitively to the PL^{Pro} structural perturbation. When Ub-AMC was used as a substrate, SBJ3-019a displayed an inhibition curve that is more complicated than that from SBJ2-043. It showed two inhibition stages with the first leading to partial inhibition and the second continuously inhibiting PL^{Pro} without reaching its inhibition plateau at the highest concentration we tested. Since it could not be fitted to a simple inhibition mechanism, we did not calculate its IC_{50} value. When Ub-AMC was used as a substrate, DUB-IN-3 displayed a simple inhibition curve that did not reach its plateau at the highest inhibitor concentration. Its estimated IC_{50} value was above 10 μM and similar to the IC_{50} value determined when Z-LRGG-AMC was used as a substrate. This similarity indicates that DUB-IN-3 likely binds to the PL^{Pro} active site to achieve its inhibition of PL^{Pro}. To understand how some inhibitors interact with PL^{Pro}, we conducted molecular docking using AutoDock Vina. As a starting point for docking, we used the crystal structure of PL^{Pro} bound with GRL0617 (PDB ID: 7CMD).[16a,17] GRL0617 and water molecules were removed from the active site to prepare PL^{Pro} for the docking analysis. To simplify our analysis, we limited docking around the active site. A binding pocket was defined based on the known residues of the S3/S4 binding pocket site of PL^{Pro}.

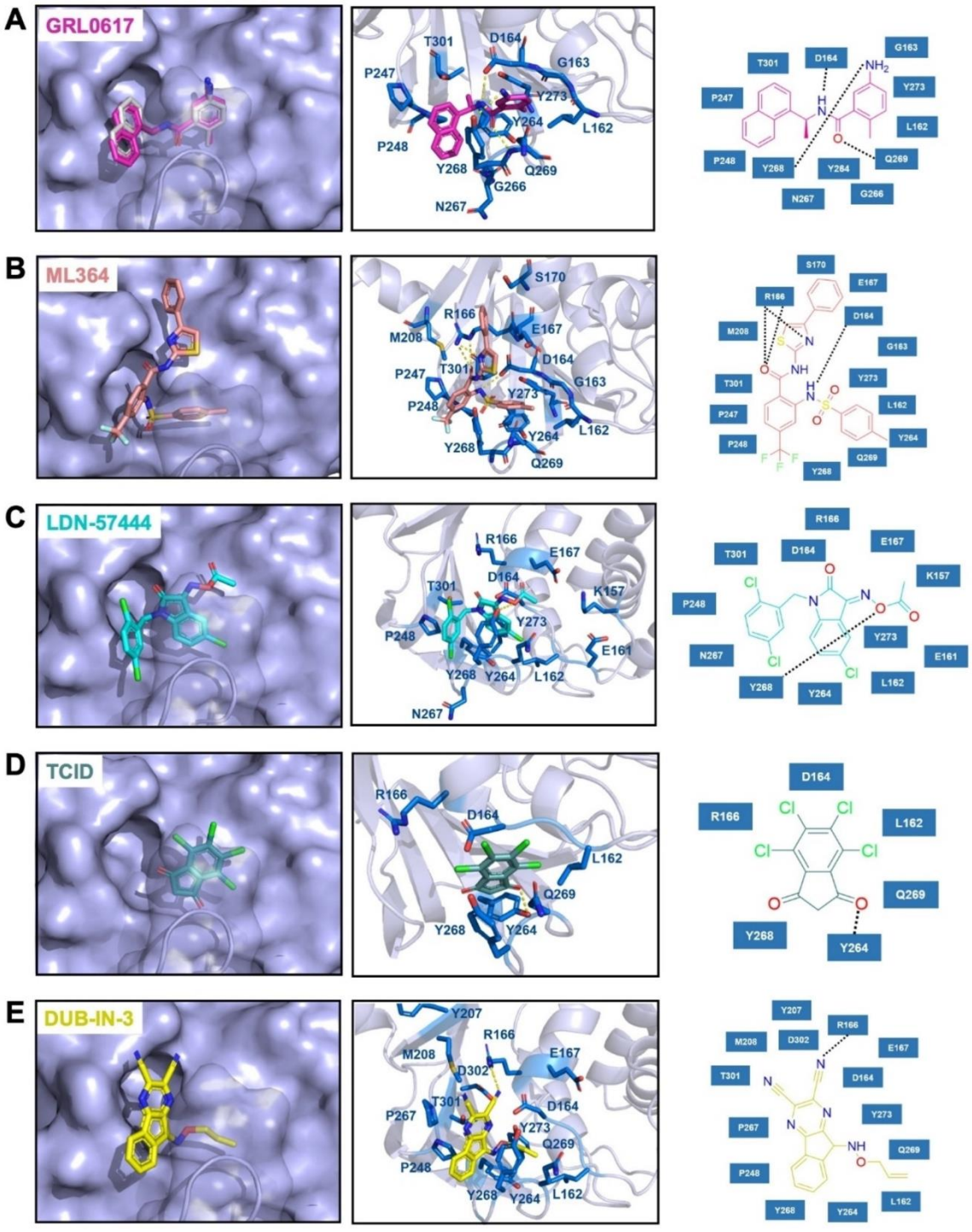

Figure 19 The top binding modes of the selected compounds along with their corresponding interactions within the active site of PL^{Pro}. Reprinted with permission from reference 81.

3D Conformers of selected compounds were generated using OpenBabel. We validated the docking protocol by conducting re-docking of GRL0617. Our docking results suggested a very similar binding mode of GRL0617 with PL^{Pro} as shown in the co-crystal structure (Figure 7A). GRL0617 formed two hydrogen bonds between its amides and PL^{Pro} residues Asp164 and Tyr268 and one hydrogen bond between its oxygen atom and Gln269. In addition, the naphthalene group of GRL0617 forms hydrophobic interactions with aromatic residues Tyr264 and Tyr268. Other inhibitors ML364, LDN-57444, TCID and DUB-IN-3 were also docked against PL^{Pro}. The protein-ligand interactions and their detailed binding modes were illustrated in Figure 19. In the docked model, ML364 forms a hydrogen bond with Asp164 and three hydrogen bonds with Arg166 (Figure 19B). LDN-57444, TCID, and DUB-IN-3 make hydrogen bonds with Tyr268, Tyr264, and Arg166, respectively (Figure 19C E). There is no strong correlation between calculated binding energies and determined IC_{50} values. This discrepancy may be explained by limited factors involved in the calculation that would lead to missing some of the potentially important interactions. Other contributing factors include potential covalent interactions with the enzyme. Further crystallographic investigation will provide critical insights on how these molecules inhibit PL^{pro}.

Conclusion

Since 2003, there have been three coronavirus disease outbreaks. Many researchers have predicted that additional coronavirus diseases will emerge with higher frequency. For both combating the current pandemic and preparing to contain future coronavirus

disease outbreaks, it is imperative to develop small molecule antivirals that may be applied generally to inhibit coronaviruses. Due to the highly conserved sequences among coronaviruses, PL^{Pro} is an attractive drug target for developing broad-spectrum antivirals. In this study, we experimentally characterized 33 deubiquitinase inhibitors and 37 cysteine protease inhibitors on their inhibition of PL^{Pro} from SARS-CoV-2. From this study, we identified seven molecules that potently inhibit PL^{Pro} with an IC_{50} value below 40 μM when Z-LRGG-AMC was used a substrate. Five inhibitors GRL0617, SJB2-043, TCID, SJB3-019 A, and PR-619 have an IC_{50} value below 10 μM. Interestingly, SBJ2-043 only partially inhibits PL^{Pro} but has an outstanding IC_{50} value of 0.56 μM. When Ub-AMC was used as a substrate, an even lower IC_{50} value of 0.091 μM was determined. SJB2-043 binds likely to an allosteric site of PL^{Pro} to exert its inhibition effect. As a pilot study, the current work indicates that drug repurposing for COVID-19 by targeting PL^{Pro} holds promises but in-depth investigation of these inhibitors in their mechanisms of action is necessary for the development of more potent PL^{Pro} inhibitors based on them.

Table 1 IC_{50} values of identified deubiquitinase inhibitors against PL^{Pro}. Reprinted with permission from reference 81.

Name	**IC_{50} (μM)[a]**	**IC_{50} (μM)[b]**
FT671	n.d.	n.d.
GNE-6776	n.d.	n.d.
GRL0617	1.37±0.10	1.80±0.21
SJB2-043	0.56±0.16	0.091±0.024
P22077	>100	n.d.
FT827	n.d.	n.d.
LDN-57444	>100	n.d.
HBX19818	>100	n.d.
TCID	6.42±0.32	10.5±2.7
IU1-47	n.d.	n.d.
NSC63283	>100	n.d.
MF-094	n.d.	n.d.
USP25/28 inhibitor AZ1	n.d.	n.d.
6-Thioguanine	n.d.	n.d.
P005091	>100	n.d.
USP7/USP47 inhibitor	n.d.	n.d.
DUB-IN-1	>100	n.d.
GSK2643943A	>100	n.d.
DUB-IN-2	>100	n.d.

Table1 continued

Name	**IC_{50} (μM)[a]**	**IC_{50} (μM)[b]**
SJB3-019A	8.15±3.59	n.d.
DUB-IN-3	12.5±3.9	>10
USP7-IN-1	n.d.	n.d.
EOAI3402143	n.d.	n.d.
IU1	n.d.	n.d.
BAY 11-7082	n.d.	n.d.
PR-619	6.1±1.2	12.9±2.4
ML-323	n.d.	n.d.
ML364	n.d.	n.d.
GNE-6640	n.d.	n.d.
Degrasyn	n.d.	n.d.
b-AP15	n.d.	n.d.
VLX1570	>100	n.d.
C527	n.d.	n.d.
[a] Determined using Z-LRGG-AMC as a PL^{Pro} substrate (n.d.: not determined). [b] Determined using Ub-AMC as a PL^{Pro} substrate.		

Table 2 IC_{50} values of identified cysteine protease inhibitors against PL^{Pro}. Reprinted with permission from reference 81.

Name	**IC_{50} (μM)[a]**
NSC18058	n.d.
Leupeptin	n.d.
HY-17541 A	n.d.
HY-17541	n.d.
PMSF	n.d.
EP1013	n.d.
S130	35.0±1.4
Petesicatib	n.d.
Tauroursodeoxycholate	n.d.
Emricasan	n.d.
N-Ethylmaleimide	n.d.
CA-074	n.d.
CA-074 methyl ester	n.d.
Z-VAD(OMe)-FMK	n.d.
Aloxistatin	n.d.
Q-VD-OPh	n.d.
H-Arg-4MβNA	n.d.
LY 3000328	n.d.
BOC-D-FMK	n.d.

Table 2 continued

Name	**IC_{50} (μM)[a]**
Z-DEVD-FMK	n.d.
E-64	n.d.
Balicatib	n.d.
Odanacatib	n.d.
Belnacasan	n.d.
Calpetin	n.d.
E 64c	n.d.
3-Epiursolic Acid	n.d.
MDL-28170	n.d.
Chelidonic acid	n.d.
MG-101	n.d.
Z-VAD-FMK	n.d.
OT-82	n.d.
IKarisoside A	n.d.
Wedelolactone	n.d.
Phenoxodiol	n.d.
2-Cyanopyrimidine	n.d.
Z-WEHD-FMKL	n.d.
[a] n.d.: not determined.	

CHAPTER IV *

CELLULAR ACTIVITIES OF SARS-COV-2 MAIN PROTEASE INHIBITORS REVEAL THEIR UNIQUE CHARACTERISTICS

This chapter is reprinted with permission of Wenyue Cao, Chia-Chuan Dean Cho, Zhi Zachary Geng, Namir Shaabani, Xinyu R. Ma, Erol C. Vatansever, Yugendar R. Alugubelli, Yuying Ma, Sankar P. Chaki, William H. Ellenburg, Kai S. Yang, Yuchen Qiao, Robert Allen, Benjamin W. Neuman, Henry Ji, Shiqing Xu, and Wenshe Ray Liu *American Chemistry Science Central Science* **2022**, 8 (2), 192−204. DOI 10.1021/acscentsci.1c00910. Copyright 2022 The Authors.[104]
(https://pubs.acs.org/doi/full/10.1021/acscentsci.1c00910)

Introduction

COVID-19 has paralyzed much of the world. As of December 9, 2021, the total confirmed infections have reached above 267 million, and the total death toll has exceeded 5.2 million worldwide.[105] With vaccines available for COVID-19, many countries have been conducting immunization campaigns hoping that herd immunity will be achieved when the majority of the population is vaccinated.[106] Current COVID-19 vaccines are targeting the Spike protein of SARS-CoV-2, the pathogen of COVID-19.[52] Spike is a

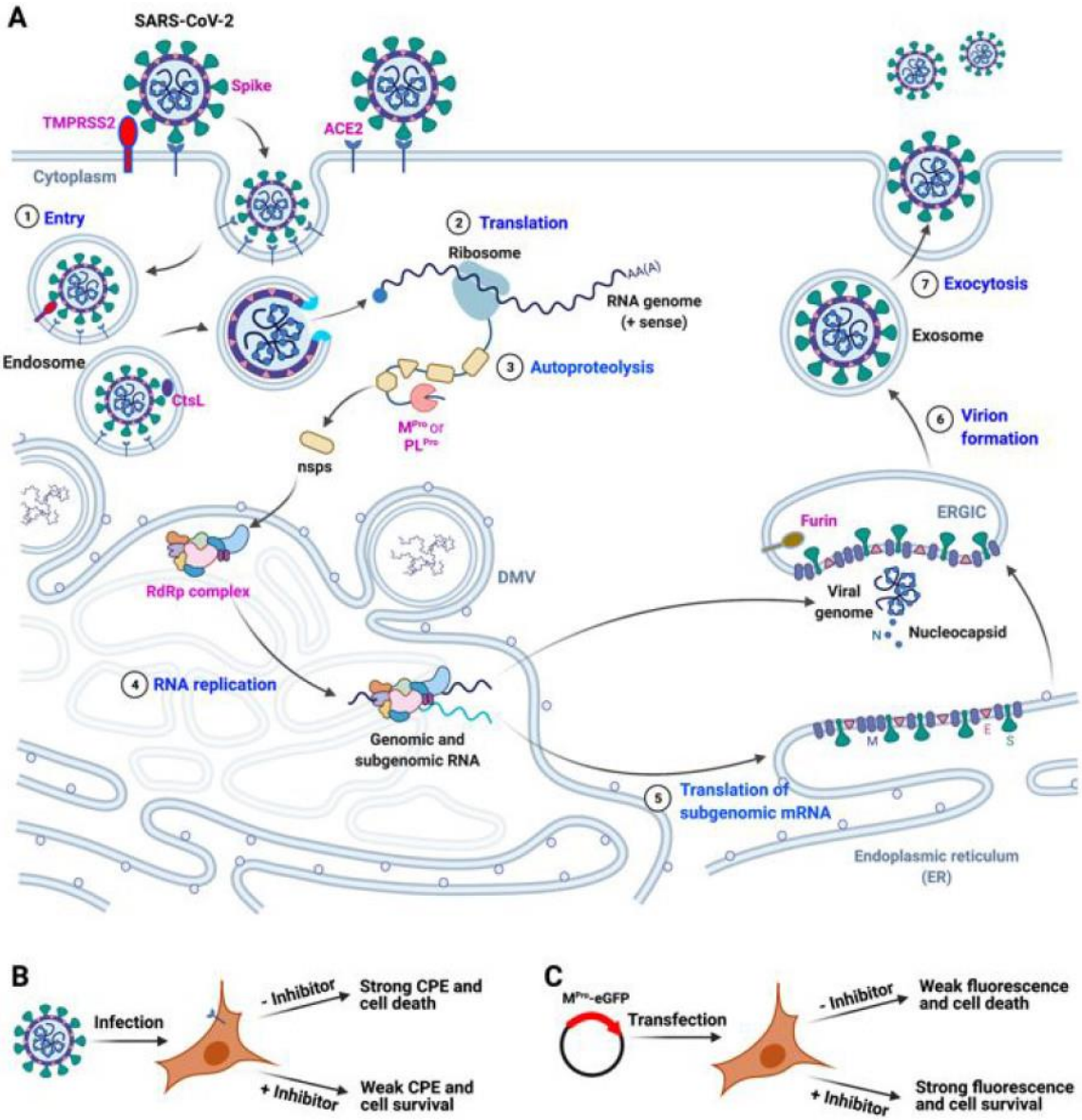

Figure 20 Life cycle of SARS-CoV-2 and two assays for M^{Pro}-targeting antivirals. (A) Cartoon diagram illustrating the life cycle of SARS-CoV-2. Seven sequential steps are labeled in blue. Proteins that are labeled in pink are targets for the development of antivirals. TMPRSS2, CtsL, and furin are three host proteases that prime Spike for viral entry and new virion packaging. ACE2, angiotensin-converting enzyme 2; RdRp, RNA-dependent RNA polymerase. (B) Antiviral assay based on the inhibition of virus-induced CPE and cell death. (C) Antiviral assay based on the inhibition of M^{Pro}-induced apoptosis in host cells and the fluorescence of the expressed M^{Pro}-eGFP fusion protein. Reprinted with permission from reference 104.

weakly conserved protein in a highly mutable RNA virus. Although SARS-CoV-2 shares overall 82% genome sequence identity with SARS-CoV, Spike has only 76% protein sequence identity shared between two origins.[107] The highly mutable nature of Spike has also been corroborated by the continuous identification of new SARS-CoV-2 variants with Spike mutations.[108] The most notable are Alpha, Beta, Delta, and Omicron variants. Accumulated evidence has shown an attenuated activity of approved vaccines against some new SARS-CoV-2 variants.[109] Booster vaccines might be developed for new virus variants. However, the situation will likely turn into an incessant race between viral mutation and vaccine development. The focus on vaccine development that is preventative toward COVID-19 has largely obscured the development of targeted therapeutics that are needed for treating patients with severe symptoms. By targeting a conserved gene in SARS-CoV-2, a small-molecule antiviral will likely be more successful than a vaccine in both prevention and treatment since it is generally easier to manufacture, store, deliver, and administer a small-molecule antiviral than a vaccine.

One demonstrated drug target in SARS-CoV-2 is its main protease (M^{Pro}).[68, 110] Unlike Spike that is highly mutable, M^{Pro} is highly conserved. Its 96% protein sequence identity shared between SARS-CoV and SARS-CoV-2 is much higher than the overall 82% genome sequence identity shared between the two viruses.[52] Much work has also been done in the development of M^{Pro} inhibitors.[111, 112] A strategy that most researchers have been following in the development of M^{Pro} inhibitors is to synthesize an inhibitor, test its enzymatic inhibition, and then carry out its structural and antiviral analysis to obtain

information for next-round optimization. For most medicinal chemists, the bottleneck in this drug discovery process is the antiviral assay that requires the use of a BSL3 facility and is often not accessible. The antiviral assay itself may also lead to misleading results about the real mechanism of an inhibitor. The life cycle of SARS-CoV-2 (Figure 20A) requires a number of proteases that are from either the host or the virus itself. It has been shown that transmembrane protease serine 2 (TMPRSS2) can prime Spike for interactions with the human cell host receptor ACE2 during the virus entry process.[113] Cathepsin L (CtsL) also potentiates the membrane fusion between SARS-CoV-2 with the endosome in infected cells.[114] It has also been suggested that other cathepsins such as cathepsin B (CtsB) serve a role in the SARS-CoV-2 entry.[115] After the SARS-CoV-2 genomic RNA is released into the host cytosol, it is translated by the host ribosome to form two large polypeptides, pp1a and pp1ab. The processing of pp1a and pp1ab to 16 nonstructural proteins (nsps) requires proteolytic functions of two internal protease fragments, nsp3 and nsp5 that are also called papain-like protease (PL^{Pro}) and main protease (M^{Pro}), respectively. Some nsps package into an RNA replicase complex that replicates both genomic and subgenomic RNAs. The translation of subgenomic RNAs leads to essential structural proteins for packaging new virions. Furin is a host protease that can hydrolyze Spike to prime it for new virion packaging and release.[116] Based on our current understanding of SARS-CoV-2 pathogenesis and replication, there are at least three host and two viral proteases serving critical roles in the SARS-CoV-2 life cycle. Inhibition of any of these enzymes will potentially cause a strong antiviral effect. The catalytic similarity between these enzymes makes it possible that a developed inhibitor is

unselective toward these enzymes. M^{Pro}, PL^{Pro}, CtsB, and CtsL are cysteine proteases with a similar catalytic mechanism. TMPRSS2 and furin are serine proteases. Although serine proteases are mechanistically different from cysteine proteases, many currently developed M^{Pro} inhibitors have covalent warheads such as aldehyde and ketone that are prone to form covalent adducts with TMPRSS2 and furin as well.[69, 117] All of these proteases are also localized in different parts of the host cell. Their inhibition requires different characteristics in inhibitors such as cellular permeability and pH sensitivity. A simple antiviral assay for an inhibitor will likely lead to a positive result that is not from the inhibition of M^{Pro} and therefore causes a misunderstanding that can be detrimental to further rounds of lead optimization. Therefore, an assay system that directly reflects M^{Pro} inhibition in the host cell is critical for both the assessment and optimization of M^{Pro} inhibitors. In the current work, we describe such a system and its application in the evaluation of a number of developed and repurposed M^{Pro} inhibitors.

Material and Methods

Chemicals, Reagents, and Cell Lines from Commercial Providers.

We purchased HEK293T/17 cells from ATCC; DMEM with high glucose with GlutaMAX supplement, fetal bovine serum, 0.25% trypsin-EDTA, phenol red, puromycin, Lipofectamine 3000, and dimethyl sulfoxide from Thermo Fisher Scientific; linear polyethylenimine MW 25000 from Polysciences; RealTime-Glo annexin V apoptosis and a necrosis assay kit from Promega; an EndoFree plasmid DNA midi kit from

Omega Biotek; antimycin a from Sigma-Aldrich; GC376 from Selleck Chem; boceprevir, calpeptin, MG-132, telaprevir, and carmofur from MedChemExpress; ebselen from TCI; calpain inhibitors II and XII from Santa Cruz Biotechnology; MG-115 From Abcam; tideglusib, disulfiram, and PX-12 from Cayman Chemical; chloroquine diphosphate from Alfa Aesar; hydroxychloroquine sulfate from Acros Organics; and a fluorogenic M^{Pro} substrate DABCYL-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-Met-Glu-EDANS termed as Sub3 from Bachem. K777 was a gift from Prof. Thomas Meek at Texas A&M University. The syntheses of MPI1−9 and 11a were shown in a previous publication.[69]

Plasmid Construction.

We amplified M^{Pro} with an N-terminal KTSAVLQ sequence using primers FRET-M^{Pro}-for and FRET-M^{Pro}-rev (Table 5) and cloned it into the pECFP-18aa-EYFP plasmid (Addgene, 109330) between XhoI and HindIII restriction sites to afford pECFP-M^{Pro}-EYFP. To construct pLVX-M^{Pro}-eGFP-1, we amplified M^{Pro} with an N-terminal methionine using primers XbaI-M^{Pro}-f and M^{Pro}-HindIII-r (Table 5) and eGFP using primers HindIII-eGFP-f and eGFP-NotI-r. We digested the M^{Pro} fragment using XbaI and HindIII-HF restriction enzymes and the eGFP fragment using HindIII-HF and NotI restriction enzymes. We ligated the two digested fragments together with the pLVX-EF1a-IRES-Puro vector (Takara Bio 631988) that was digested at XbaI and NotI restriction sites. To facilitate the ligation of three fragments, we used a ratio of M^{Pro}, eGFP, and pLVX-EF1a-IRES-Puro digested products of 3:3:1. We constructed pLVX-M^{Pro}-eGFP-2 in the

same way as pLVX-M^{Pro}-eGFP-1 except that we amplified the M^{Pro} fragment using primers XbaI-Cut-M^{pro}-f and M^{Pro}-HindIII-r (Table 5). XbaI-Cut-M^{Pro}-f encodes an MKTSAVLQ sequence for its integration to the M^{Pro} N-terminus. To construct pLVX-M^{Pro}(C145S)-eGFP, two primers M^{Pro}-C145S-f and M^{Pro}-C145S-r (Table 5) were used to carry out a site-directed mutagenesis of pLVX-M^{Pro}-eGFP. All plasmids were sequence confirmed by Sanger sequencing.

Transfection and MPI8 Inhibition Tests Using pECFP-M^{Pro}-EYFP.

We grew 293T cells to 60% confluency and then transfected them with pECFP-M^{Pro}-EYFP using Lipofectamine 3000. We added 10 μM MPI8 at the same time of transfection. After 72 h of incubation, cells were collected and analyzed by a flow cytometer as well as fluorescence microscopy. In order to obtain high-definition images, poly-D-lysine-coated glass bottom plates from Mattek were used for microimaging.

Transfection and Inhibition Tests Using pLVX-M^{Pro}-eGFP-1 and pLVX-M^{Pro}-eGFP-2.

We grew 293T cells to 60% confluency and transfected them with pLVX-M^{Pro}-eGFP-1 or pLVX-M^{Pro}-eGFP-2 using Lipofectamine 3000. We added different concentrations of MPI8 from the nM to μM level at the same time of transfection. After 72 h of incubation, we analyzed the transfected 293T cells using flow cytometry to determine fluorescent cell numbers and the eGFP fluorescence intensity.

Establishment of 293T Cells Stably Expressing M^{Pro}-eGFP.

To establish a 293T cell line that stably expresses M^{Pro}-eGFP, we packaged lentivirus particles using the pLVX-M^{Pro}-eGFP-2 plasmid. Briefly, we transfected 293T cells at 90% confluency with three plasmids including pLVX-M^{Pro}-eGFP-2, pMD2.G, and psPAX2 using 30 mg/mL polyethylenimine. We collected supernatants at 48 and 72 h after transfection separately. We concentrated and collected lentiviral particles from collected supernatant using ultracentrifugation. We then transduced fresh 293T cells using the collected lentivirus particles. After 48 h of transduction, we added puromycin to the culture media to a final concentration of 2 μg/mL. We gradually raised the puromycin concentration 10 μg/mL in 2 weeks. The final stable cells were maintained in media containing 10 μg/mL puromycin.

Apoptosis Analysis.

We performed the apoptosis analysis of the M^{Pro} stable cells and cells transiently transfected with the pLVX-M^{Pro}-eGFP-2 plasmid using the RealTime-Glo annexin V apoptosis and necrosis assay kits from Promega. The cells were maintained in a high glucose DMEM medium supplemented with 10% fetal bovine serum (FBS), plated with a cell density of 5×10^5 cells/mL. We set up five groups of experiments including (1) HEK293T/17, (2) HEK293T/17+MPI8 (1 μM), (3) HEK293T/17 cells stably expressing M^{Pro}-eGFP, (4) HEK 293T/17 cells stably expressing M^{Pro}-eGFP+MPI8 (1 μM), and (5) HEK 293T/17 or HEK 293T/17 cells stably expressing M^{Pro}-eGFP+antimycin A (1 μM). Each experiment was repeated 5 times. The assay was performed according to the

instructor's protocol. Chemiluminescence was recorded at 12, 24, 36, 48, 60, and 72 h after plating the cells. The luminescence readings were normalized using HEK 293T/17 as a negative control.

Cellular M^{Pro} Inhibition Analysis for 29 Selected Compounds.

We grew HEK 293T/17 cells in high-glucose DMEM with GlutaMAX supplement and 10% fetal bovine serum in 10 cm culture plates under 37 °C and 5% CO_2 to ∼80−90% and then transfected cells with the pLVX-M^{Pro}-eGFP-2 plasmid. For each transfection, we used 30 mg/mL polyethylenimine and a total of 8 μg of the plasmid in 500 μL of the opti-MEM medium. We incubated cells with transfecting reagents overnight. On the second day, we removed the medium, washed cells with a PBS buffer, digested them with 0.05% trypsin-EDTA, resuspended the cells in the original growth media, adjusted the cell density to 5×10^5 cells/mL, provided 500 μL of suspended cells in the growth media to each well of a 48-well plate, and then added 100 μL of a drug solution to the growth media. These cells were then incubated under 37 ° C and 5% CO_2 for 72 h before their flow cytometry analysis.

Data Collection, Processing, and Analysis.

The cell was incubated with various concentrations of drugs at 37 °C for 3 days. After 3 days of incubation, we removed the media and then washed cells with 500 μL of PBS to remove dead cells. Cells were then trypsinized and spun down at 800 rpm for 5 min. We removed the supernatant and suspended the cell pellets in 200 μL of PBS. The

fluorescence of each cell sample was collected by a Cytoflex Beckman flow cytometer based on the size scatters (SSC-A and SSC-H) and forward scatters (FSC-A). We gated cells based on SSC-A and FSC-A and then with SSC-A and SSC-H. The eGFP fluorescence was excited by a blue laser (488 nm), and cells were collected at FITC-A (525 nm). After collecting the data, we analyzed and transferred the data to csv files containing information on each cell sample. We then analyzed these files using a self-written MATLAB program for massive data processing. We sorted the FITC-A column from smallest to largest. A 10^6 cutoff was set to separate the column into two groups, larger as positive and smaller as negative. We integrated the positive group and divided the total integrated fluorescence intensity by the total positive cell counts as Flu. Int. shown in all graphs. The standard deviation of positive fluorescence was also calculated. It was then plotted and fitted nonlinearly with an agonist curve (three parameters) against drug concentrations in the program Prism 9 from GraphPad for IC_{50} determination.

Kinetic Recharacterization of Chloroquine and Hydroxychloroquine.

We prepared 10 mM stock solutions of hydroxychloroquine and chloroquine in a PBS buffer and carried out IC_{50} assays for both hydroxychloroquine and chloroquine by measuring activities of 50 nM M^{Pro} against a concentration range of 0−16 μM hydroxychloroquine and chloroquine. Serial dilutions of hydroxychloroquine and chloroquine were carried out in the assay buffer by keeping the PBS concentration the same. First, 100 nM M^{Pro} in the assay buffer (10 mM phosphate, 10 mM NaCl, 0.5 mM EDTA, pH 7.6) was treated with 2 times the working concentration of hydroxychloroquine

and chloroquine at 37 ° C for 30 min. Then, a 20 μM concentration of the fluorogenic M^{Pro} substrate Sub3 (prepared from a 1 mM stock solution of the dye in DMSO) in the assay buffer was added to the reaction mixture to a final concentration of 10 μM. Immediately after the addition of the substrate, we started to monitor the reaction in a BioTek Neo2 plate reader with an excitation wavelength at 336 nm and emission detection at 490 nm. Initial product formation slopes at the first 5 min were calculated by simple linear regression, and data were plotted in GraphPad Prism 9.

Synthesis of 5-Chloropyridin-3-yl 1 H -Indole-7-carboxylate (10-1).

To a solution of 5-chloropyridin-3-ol (1 mmol, 130 mg) and 1H-indole-7-carboxylic acid in anhydrous dichloromethane (DCM), we added DMAP (0.1 mmol, 12 mg) and EDC (1.2 mmol, 230 mg). The resulting solution was stirred at room temperature overnight. Then, the reaction mixture was evaporated *in vacuo*, and the residue was purified with flash chromatography to afford 10-1 as a white solid (210 mg, 77%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 11.34 (s, 1H), 8.65 (dd, J = 10.8, 2.2 Hz, 2H), 8.19 (t, J =2.2 Hz, 1H), 8.03−7.92 (m, 2H), 7.47 (t, J = 2.9 Hz, 1H), 7.23 (t, J = 7.7 Hz, 1H), 6.65 (dd, J=3.1, 1.9 Hz, 1H). ^{13}C NMR (101 MHz, DMSO-d_6): δ 164.7, 147.9, 146.1, 143.0, 134.9, 131.2, 131.0, 130.2, 128.0, 127.9, 125.3, 119.2, 111.3, 102.5. ESI-HRMS: calcd for $C_{14}H_{10}ClN_2O_2^+$, 273.0425; found, 273.0420.

Synthesis of 5-Chloropyridin-3-yl 1 H -Indole-4-carboxylate (10-2).

To a solution of 5-chloropyridin-3-ol (1mmol, 130 mg) and 1H-indole-4-carboxylic acid in anhydrous DCM, we added DMAP (0.1 mmol, 12 mg) and EDC (1.2 mmol, 230 mg). The resulting solution was stirred at room temperature overnight. Then, the reaction mixture was evaporated in vacuo, and the residue was puri fi ed with flash chromatography to afford 10-2 as a white solid (220 mg, 80%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.72 (s, 1H), 8.53 (dd, J=7.2, 2.2 Hz, 2H), 8.10 (dd, J=7.5, 0.9 Hz, 1H), 7.75 (t, J=2.2 Hz, 1 H), 7.71 (dt, J = 8.1, 1.0 Hz, 1H), 7.43 (t, J = 2.9 Hz, 1H), 7.32 (t, J=7.8 Hz, 1H), 7.23 (ddd, J=3.2, 2.1, 0.9 Hz, 1H).^{13}C NMR (101 MHz, $CDCl_3$): δ 165.0, 147.8, 145.7, 141.7, 136.7, 131.9, 130.0, 128.0, 127.2, 124.6, 121.3, 119.3, 117.4, 103.8. ESI-HRMS: calcd for $C_{14}H_{10}ClN_2O_2^+$, 273.0425; found, 273.0420.

Synthesis of 5-Chloropyridin-3-yl 1H-Indole-3-carboxylate (10-3).

To a solution of 5-chloropyridin-3-ol (1 mmol, 130 mg) and 1H-indole-3-carboxylic acid in anhydrous DCM, we added DMAP (0.1 mmol, 12 mg) and EDC (1.2 mmol, 230 mg). The resulting solution was stirred at room temperature overnight. Then, the reaction mixture was evaporated in vacuo, and the residue was puri fi ed with flash chromatography to afford 10-3 as a white solid (190 mg, 69%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 12.27 (s, 1H), 8.58 (dd, J=2.3, 1.0 Hz, 2 H), 8.40 (s, 1H), 8.08 (t, J = 2.2 Hz, 1H), 8.06−8.00 (m, 1H), 7.60−7.51 (m, 1H), 7.31−7.22 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6): δ 162.3, 148.0, 145.6, 142.8, 137.0, 135.1, 131.2, 130.8, 126.2, 123.4, 122.4, 120.8, 113.2, 104.8. ESI-HRMS: calcd for $C_{14}H_{10}ClN_2O_2^+$, 273.0425; found, 273.0420.

Kinetic Characterization of 10-1, 10-2, and 10-3 in Inhibiting M^{Pro}.

We performed M^{Pro} inhibition assays of 10-1, 10-2, and 10-3 according to the same procedure used for the kinetic characterization of hydroxychloroquine and chloroquine.

Characterization of the Cellular Potency of MPI1−9, GC376, 11a, 10-1, 10-2, and 10-3 in the presence of CP-100356.

All cellular M^{Pro} inhibition assays for these 14 compounds were repeated with the addition of CP-100356 in DMSO to a final concentration of 0.5 μM. The overall assay process and analysis were identical to the assays without CP-100356. Plaque Reduction Neutralization Tests of SARS-CoV-2 by MPI5−8. We seeded 18×10^3 Vero cells per well in flat-bottom 96-well plates in a total volume of 200 μL of a culturing medium (DMEM+10% FBS+glutamine) and incubated cells overnight at 37 °C and under 5% CO_2. The next day, we titrated compounds in separate round-bottom 96-well plates using the culturing medium. We then discarded the original medium used for cell culturing and replaced it with 50 μL of compound-containing media from round-bottom plates. We incubated cells for 2 h at 36 °C and under 5% CO_2. After incubation, we added 1000 PFU/50 μL of SARS-COV-2 (USA-WA1/2020) to each well and incubated the plate at 36 °C and under 5% CO_2 for 1 h. After incubation, we added 100 μL of overlay (1:1 of 2% methylcellulose and the culture medium) to each well. We incubated plates for 3 days at 36 °C and under 5% CO_2. Staining was performed by discarding the supernatant, fixing the plates with 4% paraformaldehyde in the PBS buffer for 30 min, and staining with crystal violet. Plaques were then counted.

Results and Discussion

The Rationale and the Establishment of a Cellular M^{Pro} Inhibition Assay for MPI8.

A typical antiviral assay for SARS-CoV-2 is its induced strong cytopathogenic effect (CPE) in host cells that can be quantified by counting viral plaques (Figure 20B). An M^{Pro} inhibitor with high cellular potency will suppress this virus-induced CPE and therefore lead to host cell survival. A good cellular M^{Pro} inhibition assay will need to provide results similar to this CPE suppression process. Our original design for a cellular M^{Pro} inhibition assay was to express M^{Pro} in host cells that is fused with an N-terminal cyan fluorescent protein (CFP) and a C-terminal yellow fluorescent protein (YFP) and characterize the inhibition of autocleavage of this fusion protein in the presence of an inhibitor. M^{Pro} natively cuts off its fused protein at the C-terminus. We put an M^{Pro} digestion site between CFP and M^{Pro} as well. CFP and YFP form a Förster resonance energy transfer (FRET) pair.[118] Without an inhibitor, both CFP and YFP will be cleaved from the fusion protein in host cells, which leads to no FRET signal. In the presence of a potent inhibitor, the fusion protein will be intact and emit strong FRET signals. However, transfection of 293T cells with pECFP-M^{Pro}-EYFP (Figure 33), a plasmid containing a gene coding the CFP-M^{Pro}-YFP fusion protein led to the death of most transfected cells (Figure 34). Repeating this transfection process all led to the exact same result. It is evident that M^{Pro} can exert acute toxicity to its human cell host. The same observation has been made by others and used to develop assays as well.[66] MPI8 is an M^{Pro} inhibitor that our lab developed previously.[69] Antiviral analysis indicated that MPI8 has potency to totally suppress SARS-

CoV-2-induced CPE in ACE2+ A549 cells at 0.2 μM. Given its approved antiviral potency, we used MPI8 as a positive control molecule for the analysis of cellular M^{Pro} inhibition. To alleviate the toxicity that was induced by the expression of CFP-M^{Pro}-YFP, we cultured 293T cells that were transfected with pECFP-M^{Pro}-EYFP in media containing 10 μM MPI8. The presence of MPI8 reduced death of transfected cells sharply. Interestingly, the overall expressed fusion protein was also significantly improved, showing much enhanced, directly detected yellow fluorescence from YFP (Figure 34). This positive correlation between the expression of CFP-M^{Pro}-YFP and the survival of transfected cells is likely due to the shutting-down of translation by active M^{Pro}. Since the measurement of a cell survival-correlated fluorescence increase in the presence of an inhibitor is much simpler than the FRET characterization, we decided to adopt this new way to analyze the cellular potency of M^{Pro} inhibitors. To explore whether M^{Pro} expression is correlated to the death of SARS-CoV-2-infected cells, we used SARS-CoV-2 WA1/2020 to infect both Vero E6 and ACE2+A549 cells; a strong CPE was observed 12 h after infection, which correlated with strong M^{Pro} expression detected by Western blot (Figure 35). Since a FRET system is not necessary, we modified our plasmid to express an M^{Pro}-eGFP fusion protein (Figure 20C) in host cells that can be easily analyzed using fluorescent flowcytometry. The expression of M^{Pro}-eGFP in host cells will trigger cell death that leads to weak fluorescence. This process can be reversed by adding a potent inhibitor with cellular activity. In order to use eGFP fluorescence to accurately represent expressed M^{Pro}, we introduced a Q306G mutation in M^{Pro} to abolish its cleavage of the

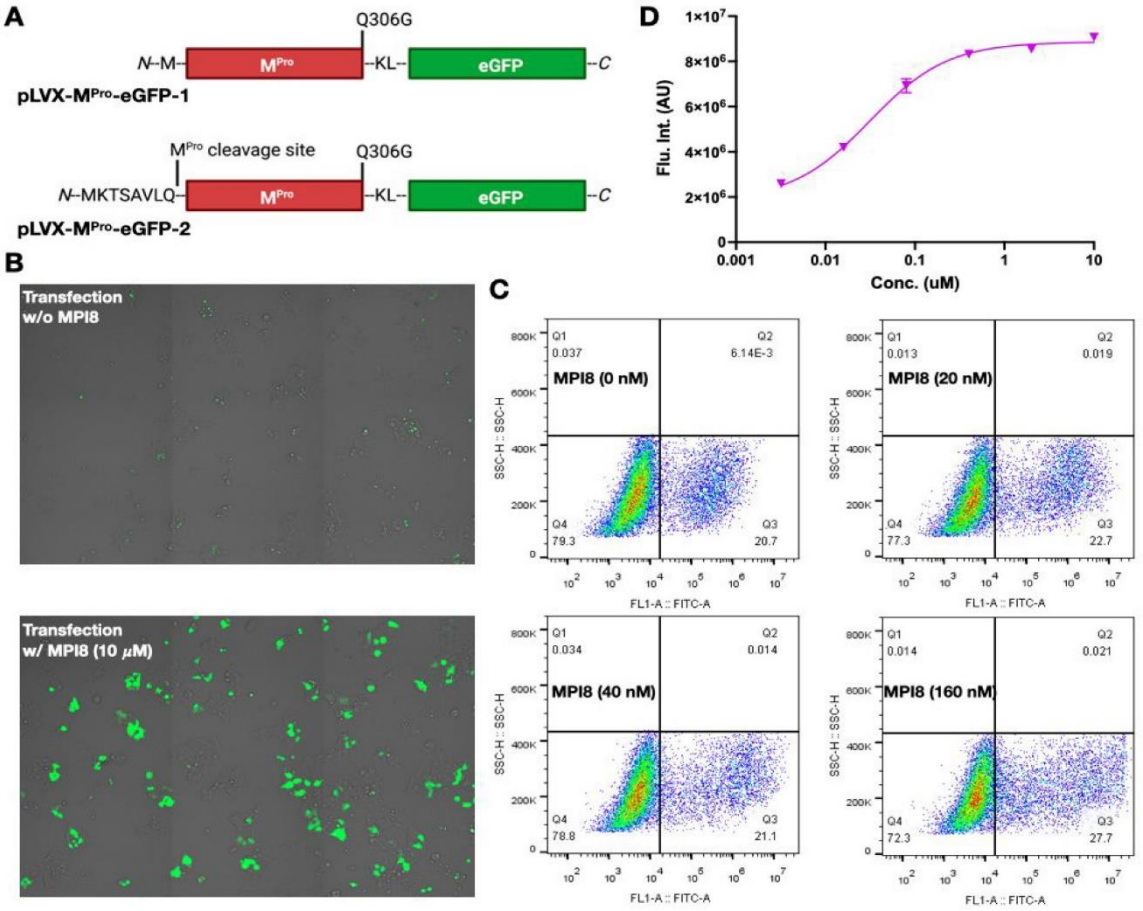

Figure 21 Validation of transiently expressed M^{Pro} and its cellular toxicity for the analysis of cellular potency of MPro inhibitors. (A) Design of two M^{Pro}-eGFP fusions. (B) 293T cells transiently transfected with pLVX-M^{Pro}-eGFP-2 and grown in the absence or presence of 10 μM MPI8. (C) 293T cells that were transiently transfected with pLVX-M^{Pro}-eGFP-2 expressed M^{Pro}-eGFP correlated with the concentration of MPI8 in the growth media. (D) Cellular IC_{50} determination of MPI8. 293T cells were transfected with pLVX-M^{Pro}-eGFP-2 and grown in the presence of different concentrations of MPI8 for 72 h before their sorting using flow cytometry. The average fluorescence intensity for cells with FL1-A signal higher than 2×10^6 was determined and used to plot against the MPI8 concentration. Data were fitted to the three-parameter dose-dependent inhibition mechanism to determine the cellular IC_{50} value. Reprinted with permission from reference 104.

C-terminal eGFP. M^{Pro} requires a free N-terminal serine for strong activity. To achieve this, we built two constructs as shown in Figure 21A and Figure 36. The first construct pLVX-M^{Pro}-eGFP-1 encodes M^{Pro}-eGFP with an N-terminal methionine that relies on host

methionine aminopeptidases for its cleavage. The second construct pLVX-M^{Pro}-eGFP-2 encodes M^{Pro}-eGFP containing a short N-terminal peptide that has an M^{Pro} cleavage site at the end for its autocatalytic release. The transfection of 293T cells with two constructs showed that pLVX-M^{Pro}-eGFP-2 led to more potent toxicity to cells, and this toxicity was effectively suppressed when we provided 10 μM MPI8 in the growth media (Figure 21B). Therefore, we selected pLVX-M^{Pro}-eGFP-2 for all of our following studies. We have noticed that 72 h provided optimal fluorescence detection indicating a slow turnover of M^{Pro}-eGFP. To demonstrate that cellular fluorescence is positively correlated to the concentration of the provided MPI8, we transfected 293T cells with pLVX-M^{Pro}-eGFP-2, grew transfected cells in the presence of four MPI8 concentrations (0, 20, 40, and 160 nM) for 72 h, and then sorted cells using fluorescent flow cytometry (Figure 21C). Both the number and intensity of fluorescent cells (FL1-A signal, $>1\times10^6$) were positively dependent on the provided MPI8 concentration, indicating the feasibility of using the system to characterize the cellular potency of an M^{Pro} inhibitor. To demonstrate this feasibility, we transiently transfected 293T cells with pLVX-M^{Pro}-eGFP-2 and grew transfected cells in the presence of a cascade of MPI8 concentrations between 0.001 and 10 μM. After 72 h, we sorted cells according to their eGFP fluorescence intensity. Cells with an FL1-A signal above 1×10^6 were analyzed. We built a MATLAB script to calculate the average eGFP fluorescence intensity of all analyzed cells and plotted the average eGFP fluorescence intensity against the MPI8 concentration as shown in Figure 21D. The data showed obvious MPI8-induced saturation of M^{Pro}-eGFP expression and fit nicely to a three-parameter dose-dependent inhibition mechanism in GraphPad Prism9 for IC_{50}

determination. The determined cellular IC_{50} value of MPI8 is 31 nM. To confirm that cell survival was from the direct inhibition of M^{Pro} protease activity by MPI8, we constructed pLVX-M^{Pro}(C145S)-eGFP whose encoded M^{Pro} has its active site cysteine mutated to serine. Transfecting 293T cells with this construct led to strong M^{Pro}(C145S)-eGFP expression that was detected by both a fluorescence measurement and Western blot and low cell death regardless of whether 1 μM MPI8 was present (Figure 37 and 38). Similarly, an M^{Pro}-targeting siRNA significantly reduced cellular apoptosis in 293T cells that transiently expressed M^{Pro}-eGFP (Figure 39). These results confirmed that host cell death was due to the protease activity of M^{Pro}. To confirm that MPI8 does not inhibit caspases that serve functions in apoptosis, we used antimycin A to induce apoptosis in 293T cells and cultured treated cells with or without 1 μM MPI8. In both conditions, the detected apoptosis levels were not significantly different (Figure 40). Collectively, these results

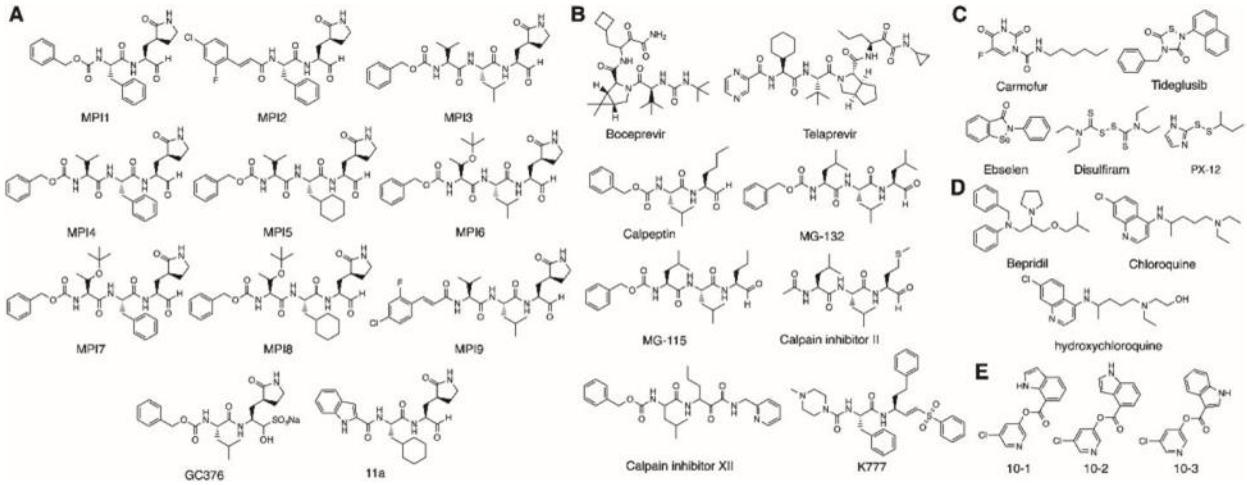

Figure 22 Structures of inhibitors that were evaluated in their cellular inhibition of M^{Pro}. (A) Reversible covalent inhibitors designed for M^{Pro}. (B) Investigational covalent inhibitors that were developed for other targets. (C) Inhibitors that were identified via high-throughput screening. (D) FDA-approved medications that have been explored as M^{Pro} inhibitors. (E) Diaryl esters that have high potency to inhibit M^{Pro}. Reprinted with permission from reference 104.

confirm that MPI8 directly inhibits the protease activity of M^{Pro} in 293T cells to cause overall cell survival and overexpression of M^{Pro}-eGFP.

Since MPI8 is highly effective in inhibiting M^{Pro} in cells, we used it in combination with pLVX-M^{Pro}-eGFP-2 to make stable 293T cells that continuously expressed M^{Pro}-eGFP. Using this stable cell line, we characterized M^{Pro}-induced apoptosis that was detected by antiannexin. After we withdrew MPI8 from the growth media, a strong apoptotic effect started to show after 24 h and continued to increase (Figure 41). Since MPI8 is a reversible covalent inhibitor, the relatively long incubation time for the

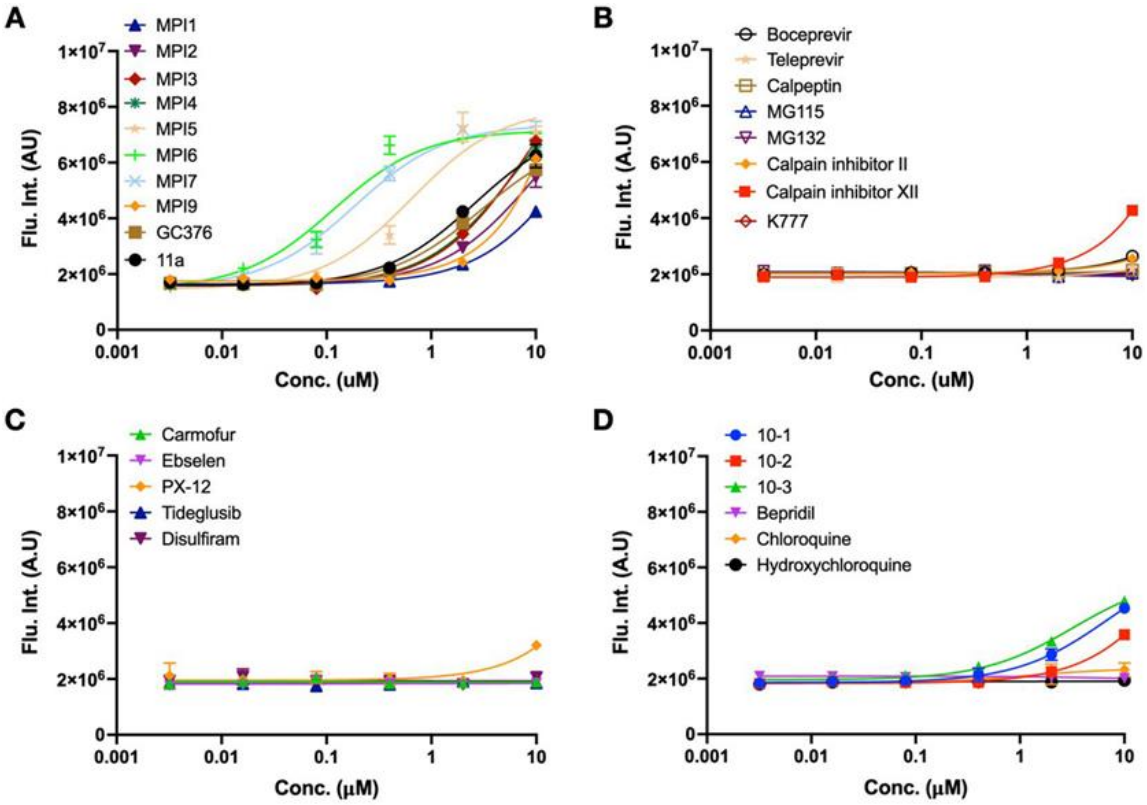

Figure 23 Cellular potency of literature-reported M^{Pro} inhibitors. K777 is included as a potential M^{Pro} inhibitor. Reprinted with permission from reference 104

observation of apoptosis is likely due to its slow release from the M^{Pro} active site. Due to concerns about residual MPI8 and its potential slow release from M^{Pro} in stable cells, we chose to perform a cellular potency characterization of all M^{Pro} inhibitors by performing a transient transfection of 293T cells and then growth in the presence of different inhibitor concentrations.

MPI1−7, MPI9, GC376, and 11a.

MPI8 was one of 9 β -(S-2-oxopyrrolidin-3-yl)alaninal (Opal)-based, reversible covalent M^{Pro} inhibitors (MPI1−9) that we previously developed (Figure 22A).[69] GC376 is a prodrug that dissociates quickly in water to release its Opal component.[119] 11a is another Opal-based, reversible covalent M^{Pro} inhibitor that was developed in 2020.[111] All 11 compounds showed high potency in inhibiting M^{Pro} in an enzymatic assay. Besides MPI8, we tested the cellular potency of all other 10 Opal inhibitors in their cellular inhibition of M^{Pro} by following the exact same procedure that we did for MPI8. As shown in Figure 23A, all tested Opal inhibitors promoted cell survival and the expression of M^{Pro}-eGFP significantly at 10 μM. However, data collected at different concentrations showed that only three inhibitors, MPI5, MPI6, and MPI7, induced the saturation of M^{Pro}-eGFP expression at or below 10 μM. Determined IC_{50} values for MPI5, MPI6, and MPI7 are 0.66, 0.12, and 0.19 μM, respectively (Table 3). Based on the collected data, MPI2−4, MPI9, GC376, and 11a have IC_{50} values higher than 2 μM, and MPI1 has an IC_{50} value higher than 10 μM.

Boceprevir, Telaprevir, Calpeptin, MG-132, MG-115, Calpain Inhibitor II, Calpain Inhibitor XII, and K777.

Drug repurposing research has led to the identification of a number of both FDA-approved and investigational drugs as M^{Pro} inhibitors. These include boceprevir, telaprevir, and calpain inhibitor XII that have an α-ketoamide moiety; and calpeptin, MG-132, and calpain inhibitor II that have an aldehyde for covalent interactions with the M^{Pro} active site cysteine.[117, 120, 121] Some of these compounds display potency in inhibiting SARS-CoV-2 replication in host cells as well. We proceeded to characterize the cellular potency of these inhibitors using our developed cellular assay. K777 is a known CtsL inhibitor with high potency in inhibiting SARS-CoV-2 replication in a human cell host.[122] It has a vinylsulfonate moiety. Due to its propensity to form a permanent covalent adduct with the M^{Pro} active site cysteine, we tested its cellular potency in inhibiting M^{Pro} as well. As shown in Figure 23B, calpeptin, MG115, MG132, telaprevir, and K777 displayed a close to undetectable cellular inhibition of M^{Pro} up to 10 μM; boceprevir and calpain inhibitor II displayed a close to undetectable cellular inhibition of M^{Pro} up to 2 μM and a very weak cellular inhibition of M^{Pro} at 10 μM. Calpeptin XII exhibited the highest cellular inhibition of M^{Pro} among this group of inhibitors, but it has low potency with an estimated IC_{50} value higher than 10 μM.

Carmofur, Tideglusib, Ebselen, Disulfiram, and PX-12.

Drug repurposing research has also shown that carmofur, tideglusib, ebselen, disulfiram, and PX-12 can potently inhibit M^{Pro}. Carmofur is an antineoplastic agent that

generates a permanent thiocarbamate covalent adduct with the M^{Pro} active site cysteine.[123] All of the other four compounds are redox active for covalent conjugation with the M^{pro} active site cysteine. We applied our cellular assay to these drugs as well. As shown in Figure 23C, except for PX-12 that weakly inhibited M^{Pro} in cells at 10 μM, the other four compounds showed an undetectable cellular inhibition of M^{Pro} at all tested inhibitor concentrations. Since all repurposed drugs in this section and the previous section displayed low cellular potency, we recharacterized most of them in an enzyme inhibition assay to confirm their in vitro M^{Pro} inhibition potency. Our determined IC_{50} values shown in Table 3 and Figure 42 are not significantly different from literature-reported results.

Bepridil, Chloroquine, and Hydroxychloroquine.

Using computational docking analysis in combination with experimental examination to guide drug repurposing for COVID-19, we previously showed that bepridil, an antianginal drug, inhibited M^{Pro} and had high potency in inhibiting SARS-CoV-2 replication in host cells.[124] To provide a full picture to understand the mechanism of bepridil in inhibiting SARS-CoV-2, we applied our cellular assay to bepridil. As shown in Figure 23D, bepridil displayed a very weak inhibition of M^{Pro} in cells up to 10 μM. A previous publication reported that chloroquine and hydroxychloroquine are potent inhibitors of M^{Pro}.[125] We applied our cellular assay to these two drugs. At all tested concentrations, both drugs displayed a close to undetectable promotion of M^{Pro}-eGFP expression, which indicates very low M^{Pro} inhibition from both drugs in cells. Using both a commercial and homemade substrate, we recharacterized M^{Pro} enzymatic inhibition by

chloroquine and hydroxychloroquine. Our data (Figure 43) showed that M^{Pro} retained 84% activity at 16 μM chloroquine, and hydroxy-chloroquine did not inhibit M^{Pro} up to 16 μM.

Diaryl Esters 10-1, 10-2, and 10-3.

Benzotriazole esters that were contaminants in a peptide library were accidentally discovered as potent inhibitors of SARS-CoV M^{Pro}.[126] Based on their inhibition mechanism, a number of diaryl esters were developed later as potent SARS-CoV M^{Pro} inhibitors.[127] To show whether similar compounds will also inhibit M^{Pro} of SARS-CoV-2, we synthesized diaryl esters 10-1, 10-2, and 10-3 and characterized their enzymatic inhibition IC_{50} values as 0.067, 0.038, and 7.6 μM, respectively (Figure 44). Using our cellular assay, we characterized all three compounds as well. As shown in Figure 23D, all three compounds display observable potency in inhibiting M^{Pro} to promote M^{Pro}-eGFP expression at 2 and 10 μM. Their cellular M^{Pro} inhibition IC_{50} values are estimated above 10 μM.

Effect of CP-100356 on the Cellular Potency of Peptide-Based M^{Pro} Inhibitors.

CP-100356 is a high-affinity inhibitor of multidrug resistance protein (Mdr-1/gp), a protypical ABC transport that exports toxic substances from the inside of cells. A previous report showed that CP-100356 enhanced the antiviral potency of M^{Pro} inhibitors significantly.[128] To investigate whether CP-100356 improves the cellular M^{Pro}

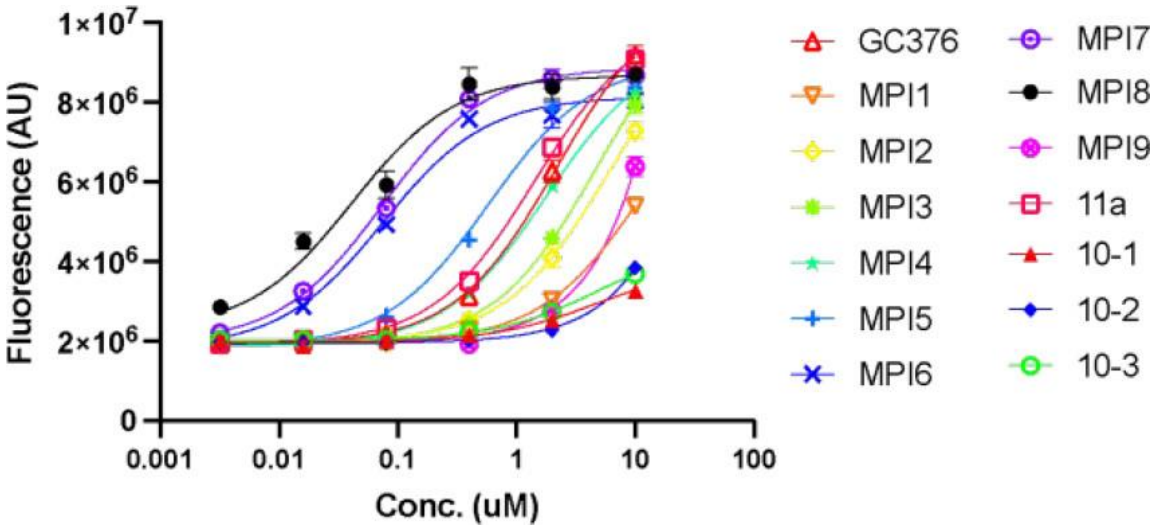

Figure 24 Cellular potency of selected compounds in their inhibition of M^{Pro} in the presence of 0.5 μM CP-100356. Reprinted with permission from reference 104.

inhibition potency of Opal inhibitors, we recharacterized MPI1−9, GC376, and 11a using our cellular assay in the presence of 0.5 μM CP-100356 (Figure 24). Except for MPI8 that showed an inhibition curve in the presence of CP-100356 very similar to that in the absence of CP-100356 and had a determined IC_{50} value of 39 nM, all other Opal inhibitors displayed a better cellular M^{Pro} inhibition curve. MPI5 and MPI6 have IC_{50} values (580 and 75 nM, respectively) in the presence of CP-100356 that are slightly lower than those in the absence of CP-100356. The highest cellular potency improvement that we observed among all compounds was for MPI7. It displayed an IC_{50} value (75 nM) in the presence of CP-100356 that is 60% lower than that in the presence of CP-100356. The cellular potency improvement for MPI4, GC376, and 11a in the presence of CP-100356 also led to characterizable IC_{50} values of 1.8, 2.2, and 1.4 μM, respectively. We did a similar test for 10-1, 10-2, and 10-3. Providing CP-100356 did not significantly change the cellular M^{Pro} inhibition for all three compounds at all tested concentrations.

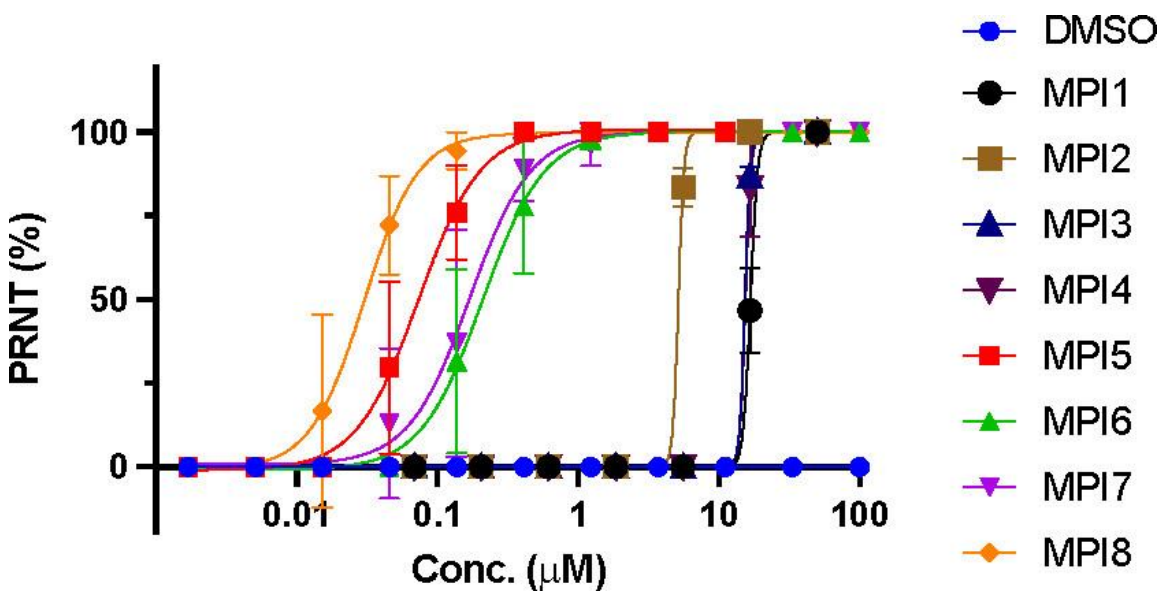

Figure 25 Plaque reduction neutralization tests (PRNTs) of MPI5–8 on their inhibition of SARS-CoV-2 in Vero E6 cells. DMSO was used as a negative control. Determination of Antiviral EC_{50} Values for MPI1−8. Reprinted with permission from reference 104.

Our previous antiviral assay for Opal inhibitors was based on the on−off observation of a CPE in Vero E6 and ACE2+A549 cells. To quantify antiviral EC_{50} values of MPI1−8, we conducted plaque reduction neutralization tests of SARS-CoV-2 in Vero E6 cells in the presence of MPI1−8. we infected Vero E6 cells with SARS-CoV-2, grew infected cells in the presence of different inhibitor concentrations for 3 days, and then quantified SARS-CoV-2 plaque reduction. Based on SARS-CoV-2 plaque reduction in the presence of MPI1−8, we determined antiviral EC_{50} values for MPI1−8 as shown in Figure 25 and Table 3. MPI1−4 displayed a low antiviral potency with estimated EC_{50} values above 5 μM, and MPI5–8 have EC_{50} values determined to be 0.073, 0.21, 0.17, and 0.030 μM, respectively.

Discussion

MPI1−9 were previously developed as potent M^{Pro} inhibitors. All showed enzymatic IC_{50} values around or below 100 nM (Table 3). Among them, MPI3 has the highest potency with an IC_{50} value of 8.5 nM. However, a CPE-based antiviral assay in Vero E6 cells showed that MPI3 weakly inhibited SARS-CoV-2.[69] On the contrary, MPI8 that has an enzymatic IC_{50} value of 105 nM displayed the highest potency in inhibiting SARS-CoV-2. A separate antiviral assay in ACE2+A549 cells showed that MPI8 inhibited the SARS-CoV-2-induced CPE completely at 200 nM MPI8. Overall, the antiviral potency of MPI1−9 based on the on−off observation of the CPE correlates with their cellular M^{Pro} inhibition potency that we have detected using the new cellular assay. In order to confirm that the determined cellular potency results correlate closely with antiviral effects, we quantified antiviral EC_{50} values for MPI1−8 in Vero E6 cells. Overall, the general trends of determined potency for MPI1−8 from two assays correlate well with each other, indicating that the developed cellular assay is valid in assessing the antiviral potency of M^{Pro} inhibitors if these inhibitors act on M^{Pro} alone. CP-100356 improved the cellular potency for most Opal-based inhibitors, although this improvement is as dramatic as reported in Hoffman et al.[128] Therefore, the main reason for the low cellular and antiviral potency of MPI3 and other Opal-based inhibitors might not be their active exportation from cells. Possible reasons that may contribute to the low antiviral and cellular potency for these compounds include their potential low cell permeability and proneness to both

extracellular and intracellular proteolysis. Although MPI8 is not the most potent Opal-based inhibitor according to its in vitro enzymatic inhibition potency, it has the best antiviral and cellular potency. Its determined cellular IC_{50} is 31 nM, which is less than a third of its in vitro enzymatic IC_{50} value. A likely reason is the possible accumulation of MPI8 in cells, which needs to be investigated. Other Opal-based inhibitors with high cellular potency are MPI5, MPI6, and MPI7. All display cellular IC_{50} values below 1 μM. Among all 30 inhibitors that we have tested, MPI5−8 show the highest potency, which warrants their further investigation for possible use in treating COVID-19. As far as we know, MPI8 is the compound with the highest cellular M^{Pro} inhibition potency and the highest SARS-CoV-2 antiviral potency in Vero E6 cells. We urge its preclinical investigation for treating COVID-19.

GC376 is an investigational drug for treating feline infectious peritonitis, a lethal coronavirus disease in cats.[119] Anivive Lifesciences Inc. did clinical investigations to repurpose GC376 for the treatment of COVID-19 patients. Although GC376 has high in vitro enzymatic M^{Pro} Inhibition potency with an IC_{50} value of 30 nM, it shows relatively weak cellular potency (IC_{50} > 2 μM). This weak cellular potency correlates with its antiviral potency that was determined with an EC_{50} value of 3.37 or 0.7 μM from two separate studies.[117, 120] In comparison to MPI8, GC376 is almost 2 orders of magnitude less potent in cellular and antiviral potency. A low cellular permeability and stability likely contribute to this low cellular and antiviral potency. 11a is an M^{Pro} inhibitor that has an antiviral EC_{50} value of 0.53 μM which is not significantly different from those of MPI6

and MPI7.[111] However, its cellular potency is much weaker compared to MPI6 and MPI7. Its estimated cellular IC_{50} value is higher than 2 μM. It is likely that 11a may interfere with other critical process(es) in the SARS-CoV-2 life cycle to exert a potent antiviral effect, which needs to be explored.

Boceprevir and telaprevir are two drugs approved for treating hepatitis C virus infection. Both have shown potency to inhibit M^{Pro} enzymatically, and boceprevir has also been characterized in an antiviral assay to show an EC_{50} value of 1.31 μM.[117] However, both drugs display very weak potency in their cellular M^{Pro} inibition tests. Since we detected very weak cellular potency for boceprevir at 10 μM, boceprevir must hit on other key step(s) in the SARS-CoV-2 pathogenesis and replication pathway to convene its high antiviral effect. An investigation in this possibility will likely lead to the discovery of novel target(s) for COVID-19 drug development. Other aldehyde- and ketone-based inhibitors we have tested include calpeptin, MG-132, MG-115, calpain inhibitor II, and calpain inhibitor XII. Except for calpain inhibitor XII that showed a weak cellular inhibition of M^{Pro} with an estimated IC_{50} value higher than 10 μM, all others exhibited a close to undetectable M^{Pro} inhibition in cells up to 10 μM. Both calpain inhibitor II and XII have demonstrated antiviral potency toward SARS-CoV-2 with an EC_{50} value of 2.07 and 0.49 μM, respectively. Based on our cellular potency analysis of the two compounds, it is clear that their antiviral potency is not primarily from the inhibition of M^{Pro}. Wang et al. have explored compounds with dual functions to inhibit both M^{Pro} and host calpains/cathepsins as antivirals for SARS-CoV-2.[129] These compounds include calpain inhibitor II and XII.

As such, they likely inhibit host proteases to cause potent antiviral effects. K777 weakly inhibited M^{Pro} in a kinetic assay but potently inhibited SARS-CoV-2 in an antiviral assay.[122] It showed undetectable potency in our cellular assay, which confirms that it must target other key process(es) in the SARS-CoV-2 life cycle.

Carmofur, tideglusib, ebselen, disulfiram, and PX-12 were discovered as M^{Pro} inhibitors from high-throughput screening. Although carmofur has an enzymatic IC_{50} value of 1.35 μM and generates a permanent covalent adduct with the M^{Pro} active site cysteine by forming a thiocarbamate, it showed undetectable cellular potency up to 10 μM in our cellular assay. This observation correlates well with its low antiviral potency.[123] The high chemical reactivity of carmofur likely contributes to its low cellular and antiviral potency. Tideglusib, ebselen, disulfiram, and PX-12 are redox-active compounds that can form covalent adducts with the M^{Pro} active site cysteine. Except for PX-12 that showed weak cellular potency at 10 μM, the other three drugs exhibited undetectable cellular potency up to 10 μM. Among the four compounds, only ebselen has been examined in an antiviral assay.[68] It has a determined EC_{50} value of 4.67 μM. Since ebselen showed undetectable cellular M^{Pro} inhibition up to 10 μM, its antiviral potency must be from its interference with other key process(es) in the SARS-CoV-2 life cycle. The revelation of the SARS-CoV-2 inhibition mechanism by ebselen will likely lead to the discovery of novel drug target(s) for COVID-19.

Bepridil is an M^{Pro} inhibitor with an enzymatic IC_{50} value of 72 μM but a much lower antiviral EC_{50} value of 0.46 μM in ACE2+A549 cells. Bepridil is known to inhibit other human viral pathogens as well.[130] We detected a close to undetectable cellular M^{Pro} inhibition potency for bepridil up to 10 μM. This correlates with its relatively high enzymatic IC_{50} value. Therefore, bepridil must use a mechanism different from the inhibition of M^{Pro} to convene its high antiviral potency. This needs to be investigated. Chloroquine and hydroxychloroquine are two repurposed drugs for COVID-19 with demonstrated antiviral EC_{50} values of 5.47 and 0.72 μM, respectively.[131] Although TMPRSS2 was shown as a possible target of chloroquine and hydroxychloroquine,[132] a previous report showed that chloroquine and hydroxychloroquine potently inhibited M^{Pro} in an enzyme inhibition assay.[125] We tested both drugs using the new cellular assay but revealed close to undetectable cellular M^{Pro} inhibition up to 10 μM for both drugs. We recharacterized the enzymatic inhibition of M^{Pro} by both drugs as well. However, we were not able to detect any M^{Pro} inhibition by hydroxychloroquine up to 16 μM, and chloroquine exhibited very weak inhibition of M^{Pro} at 16 μM. Based on our cellular data, enzymatic inhibition data, and data from a separate study,[133] we are confident that both chloroquine and hydroxychloroquine do not potently inhibit M^{Pro}. Their antiviral activities are from different mechanism(s).

10-1, 10-2, and 10-3 are three diaryl esters in which 10-1 and 10-2 displayed high potency in inhibiting M^{Pro} enzymatically. All three compounds displayed a significant cellular M^{Pro} inhibition potency at 10 μM, but their potency is much lower than those of

MPI5−8. Although 10-3 has a much weaker enzymatic inhibition potency than 10-1 and 10-2, its cellular potency is slightly better than those from 10-1 and 10-2. A likely explanation is that 10-3 is more stable than 10-1 and 10-2, which leads to a longer cellular time to convene its cellular M^{Pro} inhibition potency. Therefore, we recommend balancing cellular stability and enzymatic inhibition potency for the future development of diaryl esters as M^{Pro} inhibitors to achieve optimal antiviral effects.

As a protypical ABC transporter inhibitor, CP-100356 can potentially improve the intracellular accumulation of exogenous toxic molecules in cells. Providing CP-100356 improved the cellular activity for all Opal-based inhibitors except MPI8, albeit the improvement is not as great as what was reported for PF-00835231.[128] This is likely due to a low expression of Mdr-1/gp in 293T cells. Since CP-100356 is not an approved drug, its use in combination with an M^{Pro} inhibitor for COVID-19 treatment will face significant hurdles in clearing out toxicity and other clinical concerns. MPI8 showed a similar cellular potency in the presence and absence of CP-100356, suggesting MPI8's high propensity to accumulate inside cells. This explains our observation that the determined cellular M^{Pro} inhibition IC_{50} value for MPI8 was 3-fold less than its determined enzymatic inhibition IC_{50} value. Data related to the use of CP-100356 support that MPI8 is optimal for cellular M^{Pro} inhibition. As the compound with the highest cellular and antiviral potency among all of the literature and new compounds that we have tested in the current study, we urge MPI8 for further investigations in treating COVID-19.

Conclusion

We have developed a cellular assay to determine the cellular potency of SARS-CoV-2 M^{Pro} inhibitors. Unlike an antiviral assay in which the interference of any key step in the SARS-CoV-2 life cycle may lead to a strong antiviral effect, this new cellular assay reveals only cellular M^{Pro} inhibition potency of a compound. It provides precise information that reflects real M^{Pro} inhibition in cells. Using this assay, we characterized 30 M^{Pro} inhibitors. Our data indicated that 11a, boceprevir, ebselen, calpain inhibitor II, calpain inhibitor XII, K777, and bepridil likely interfere with key processes other than the M^{Pro} catalysis in the SARS-CoV-2 pathogenesis and replication pathways to convene their strong antiviral effects. Our results also revealed that MPI8 has the highest cellular potency among all compounds that were tested. It has a cellular M^{Pro} inhibition IC_{50} value of 31 nM. MPI8 has been recently shown with dual inhibition effects against human cathepsin L but not other tested human proteases.[134] As the compound with the highest antiviral potency with an EC_{50} value of 30 nM, we cautiously advocate preclinical tests of MPI8 as a COVID-19 treatment.

Table 3 Determined Enzymatic and Cellular IC_{50} Values in Inhibiting SARS-CoV-2 M^{Pro} for Different Inhibitors. Reprinted with permission from reference 104.

Compound ID	Enzymatic IC_{50} (μM)	Cellular IC_{50} (μM)	Cellular IC_{50} (μM) with CP-100356	Antiviral EC_{50} (μM)	Compound ID	Enzymatic IC_{50} (μM)	Cellular IC_{50} (μM)	Cellular IC_{50} (μM) with CP-100356	Antiviral EC_{50} (μM)
MPI1[69]	0.1000±0023	>10	>2	>5	MG-132	3.9±1.0;[117] 3.0±0.2[d]	n.d.[c,e]		
MPI2[69]	0.103±0.014	>2	>2	>5	Calpain inhibitor II[117]	0.97±0.27	>10		2.07±0.76[a]
MPI3[69]	0.0085±0.0015	>2	>2	>5	Calpain inhibitor XII	0.45±0.06; [117] 0.82±0.08[d]	>10		0.49±0.18[a]
MPI4[69]	0.015±0.005	>2	1.8±0.01	>5	K777[122]	>100	n.d.		0.62[a]
MPI5[69]	0.033±0.002	0.66±0.15	0.58±0.06	0.073±0.007	Carmofur	1.35±0.04;[68, 123] 0.20±0.01[d]	n.d.		>100[b]
MPI6[69]	0.060±0.004	0.12±0.03	0.075±0.008	0.21±0.02	Tideglusib[68]	1.55±0.30; 2.8±0.1[d]	n.d.		
MPI7[69]	0.047±0.003	0.19±0.03	0.075±0.006	0.17±0.02	Ebselen[68]	0.67±0.09; 0.98±0.01[d]	n.d.		4.76±0.80[a]
MPI8[69]	0.105±0.022	0.031±0.002	0.039±0.007	0.030±0.003	Disulfiram[68]	9.35±0.18; 2.2±0.2[d]	n.d.		
MPI9[69]	0.056±0.014	>2	>2		PX-12[68]	21.4±7.1	>10[c]		
GC376	0.030±0.0086[69]	>2	2.2±0.2	3.37±1.68[a]; [117] 0.70[120]	Bepridil[124]	72±3	n.d.		0.46[a]
11a	0.053±0.005;[107] 0.031±0.003[69]	>2	1.4±0.1	0.53±0.01[a]	Chloroquine[125, 131]	3.9±0.2, >10[d]	n.d.		5.47[a]
Boceprevir	4.2±0.6;[117] 8.0±1.5;[120] 7.2±2.3[d]	>>10		1.31±0.58[a]; [117] 15.57[120]	Hydroxychloroquine[125, 131]	2.9±0.3, >10[d]	n.d.		0.72[a]
Telaprevir[121]	15.3	>>10			10-1	0.040±0.004	>10	>10	
Calpeptin[117]	10.7±2.8	n.d.			10-2	0.068±0.005	>10	>10	
MG-115	3.1±1.0; [117] 2.7±0.1[d]	n.d.[c]			10-3	5.72±0.43	>10	>10	

[a]Primary CPE assay. [b]Genomic RNA quantification. [c]Toxic at 10 μM. [d]Determined separately by us. [e]n.d.: not detected

CHAPTER V

CONCLUSIONS

The instability and low activity of PylRS has been a long-lasting problem for expressing proteins incorporated with ncAAs. We discovered the amber suppression tRNA from *C. M. alvus* protects as well as stabilizes PylRS from *M. mazei*. The enhancement rescues the low yield of sfGFP expression by 3-fold or even more. MmPylRS with different digestion fragments pairing to CMaPylT or MmPylT provides us the insight for stabilizing full length MmPylRS by mutating K110 and P188. Although the benefit from each site is not additive, P188G has stabilized and enhanced the amber suppression efficiency well. A detailed crystal structure of MmPylRS and its tRNA could provide extra information for better engineering. However, the only structure available is only from aa1-101 and aa184-454. Aligning with CMaPylRS, interestingly, it matched 80% within aa186-454. A co-crystal structure of CMaPylRS and its tRNA could be very beneficial for engineering a stable MmPylRS.

COVID-19 has built into our daily life since the outbreak in 2020. Although the vaccination has been developed and is widely distributed, there's still increasing cases. Not to mention new mutants are still developing. Fortunately, most of the mutations occurs on the spike instead of the essential enzymes., therefore, antiviral drugs targeting the essential proteases such as M^{pro} and PL^{pro} in the viral replication pathway are active toward all different variants of SARS-CoV-2. In this dissertation, we expressed and purified

SARS-CoV-2 PL^{pro}. Furthermore, we identified several potent inhibitors from a deubiquitinase inhibitor library and a cysteine protease inhibitor library. However, among all the inhibitors, a deubiquitinase inhibitor specifically developed for SARS-CoV-1 PL^{pro}, GRL-0617, shows the best inhibition and IC_{50}. In addition, we found that some inhibitors, such as SJB3-019A, displayed different inhibition patterns with the peptide substrate, Z-LRGG-AMC, and a protein substrate, Ub-AMC. The IC_{50} improved with Ub-AMC. It likely inhibits PL^{pro} with a different mechanism. We also developed a method to evaluate the M^{pro} inhibitor's activity *in vivo*. Due to the special structure of cell membrane, delivering drugs into a cell plays an important role to inhibit targeted protein *in vivo*. It makes some of the *in vitro* assay results poorly matched with *in vivo* assay results. To better evaluate the inhibitor activity, viral experiments have to be carried out in BSL3 or higher laboratory. To reduce the laboratory requirement, we developed an *in vivo* assay that can quantify EC_{50} of inhibitors that can be done in any BSL-2 laboratory. We expressed M^{pro}-Q306G-eGFP in HEK293/17 cells. Low fluorescence was observed due to the M^{pro} cytotoxicity. On the contrary, the potent inhibitors rescue the cells and results in high fluorescence. The inhibitor concentration-dependency of the fluorescence enhancement allows us to easily evaluate the potency of each drug *in vivo*. The simplicity and the availability of this method allows for quick identification of potent inhibitors *in vivo* in most biochemical laboratories. It is greatly beneficial for the current pandemic due to the high mutation rate of SARS-CoV-2. To date, there is only one FDA approved drug that was designed specifically for SARS-CoV-2 M^{pro}. With the increasing infected cases and possible new variants, the benefits from various drugs options could be additive and

economical. In addition, it can be easily adapted to evaluate M^{pro} from different species of coronavirus by switching the expressed protein. It can greatly fasten the drug discovery process for preparing for an unknown coronavirus.

REFERENCES

(1) Koonin, E. V.; Novozhilov, A. S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **2009**, *61* (2), 99-111. DOI: 10.1002/iub.146 From NLM Medline.

(2) Doig, A. J. Frozen, but no accident - why the 20 standard amino acids were selected. *FEBS J* **2017**, *284* (9), 1296-1305. DOI: 10.1111/febs.13982 From NLM Medline.

(3) Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem Asian J* **2009**, *4* (5), 630-640. DOI: 10.1002/asia.200800427 From NLM.

(4) Wang, H.; Wang, L.; Zhong, B.; Dai, Z. Protein Splicing of Inteins: A Powerful Tool in Synthetic Biology. *Front Bioeng Biotechnol* **2022**, *10*, 810180. DOI: 10.3389/fbioe.2022.810180 From NLM PubMed-not-MEDLINE.

(5) Jensen, K. J. Solid-phase peptide synthesis: an introduction. *Methods Mol Biol* **2013**, *1047*, 1-21. DOI: 10.1007/978-1-62703-544-6_1 From NLM.

(6) Anthony-Cahill, S. J.; Griffith, M. C.; Noren, C. J.; Suich, D. J.; Schultz, P. G. Site-specific mutagenesis with unnatural amino acids. *Trends Biochem Sci* **1989**, *14* (10), 400-403. DOI: 10.1016/0968-0004(89)90287-9 From NLM Medline.

(7) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498-500. DOI: 10.1126/science.1060077 From NLM Medline.

(8) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **2010**, *395* (2), 361-374. DOI: 10.1016/j.jmb.2009.10.030 From NLM Medline.

(9) Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **2002**, *296* (5572), 1459-1462. DOI: 10.1126/science.1069588 From NLM Medline. Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **2002**, *296* (5572), 1462-1466. DOI: 10.1126/science.1069556 From NLM Medline.

(10) Paul, L.; Ferguson, D. J., Jr.; Krzycki, J. A. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of Methanosarcina barkeri contain in-frame and read-through amber codons. *J Bacteriol* **2000**, *182* (9), 2520-2529. DOI: 10.1128/JB.182.9.2520-2529.2000 From NLM Medline. James, C. M.; Ferguson, T. K.; Leykam, J. F.; Krzycki, J. A. The amber codon in the gene encoding the monomethylamine methyltransferase isolated from Methanosarcina barkeri is translated as a sense codon. *J Biol Chem* **2001**, *276* (36), 34252-34258. DOI: 10.1074/jbc.M102929200 From NLM Medline. Tharp, J. M.; Ehnbom, A.; Liu, W. R. tRNA(Pyl): Structure, function, and applications. *RNA Biol* **2018**, *15* (4-5), 441-452. DOI: 10.1080/15476286.2017.1356561 From NLM Medline.

(11) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* **2014**, *1844* (6), 1059-1070. DOI: 10.1016/j.bbapap.2014.03.002 From NLM Medline.

(12) Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Soll, D.; Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **2009**, *457* (7233), 1163-1167. DOI: 10.1038/nature07611 From NLM Medline.

(13) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Nureki, O.; Yokoyama, S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta*

Crystallogr Sect F Struct Biol Cryst Commun **2006**, *62* (Pt 10), 1031-1033. DOI: 10.1107/S1744309106036700 From NLM Medline.

(14) Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L. Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem Biol* **2011**, *6* (7), 733-743. DOI: 10.1021/cb200057a From NLM Medline.

(15) Jiang, R.; Krzycki, J. A. PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. *J Biol Chem* **2012**, *287* (39), 32738-32746. DOI: 10.1074/jbc.M112.396754 From NLM Medline.

(16) Yamaguchi, A.; Iraha, F.; Ohtake, K.; Sakamoto, K. Pyrrolysyl-tRNA Synthetase with a Unique Architecture Enhances the Availability of Lysine Derivatives in Synthetic Genetic Codes. *Molecules* **2018**, *23* (10). DOI: 10.3390/molecules23102460 From NLM Medline.

(17) Bhattacharyya, M.; Vishveshwara, S. Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry* **2011**, *50* (28), 6225-6236. DOI: 10.1021/bi200306u From NLM Medline.

(18) Herring, S.; Ambrogelly, A.; Gundllapalli, S.; O'Donoghue, P.; Polycarpo, C. R.; Soll, D. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett* **2007**, *581* (17), 3197-3203. DOI: 10.1016/j.febslet.2007.06.004 From NLM Medline.

(19) Sharma, V.; Zeng, Y.; Wang, W. W.; Qiao, Y.; Kurra, Y.; Liu, W. R. Evolving the N-Terminal Domain of Pyrrolysyl-tRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids. *Chembiochem* **2018**, *19* (1), 26-30. DOI: 10.1002/cbic.201700268 From NLM Medline.

(20) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Soll, D.; Liu, D. R. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol* **2017**, *13* (12), 1253-1260. DOI: 10.1038/nchembio.2474 From NLM Medline.

(21) Meineke, B.; Heimgartner, J.; Lafranchi, L.; Elsasser, S. J. Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem Biol* **2018**, *13* (11), 3087-3096. DOI: 10.1021/acschembio.8b00571 From NLM Medline.

(22) Cho, C. C.; Blankenship, L. R.; Ma, X.; Xu, S.; Liu, W. The Pyrrolysyl-tRNA Synthetase Activity can be Improved by a P188 Mutation that Stabilizes the Full-Length Enzyme. *J Mol Biol* **2022**, *434* (8), 167453. DOI: 10.1016/j.jmb.2022.167453 From NLM Medline.

(23) Seki, E.; Yanagisawa, T.; Kuratani, M.; Sakamoto, K.; Yokoyama, S. Fully Productive Cell-Free Genetic Code Expansion by Structure-Based Engineering of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase. *ACS Synth Biol* **2020**, *9* (4), 718-732. DOI: 10.1021/acssynbio.9b00288 From NLM Medline.

(24) Willis, J. C. W.; Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem* **2018**, *10* (8), 831-837. DOI: 10.1038/s41557-018-0052-5 From NLM Medline.

(25) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci U S A* **2007**, *104* (27), 11268-11273. DOI: 10.1073/pnas.0704769104 From NLM Medline.

(26) Chica, R. A.; Doucet, N.; Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol* **2005**, *16* (4), 378-384. DOI: 10.1016/j.copbio.2005.06.004 From NLM Medline.

(27) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **2008**, *4* (4), 232-234. DOI: 10.1038/nchembio.73 From NLM Medline.

(28) Lee, Y. J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R. A genetically encoded acrylamide functionality. *ACS Chem Biol* **2013**, *8* (8), 1664-1670. DOI: 10.1021/cb400267m From NLM Medline.

(29) Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y. J.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R. A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew Chem Int Ed Engl* **2017**, *56* (6), 1643-1647. DOI: 10.1002/anie.201611415 From NLM Medline.

(30) Wang, Z. A.; Zeng, Y.; Kurra, Y.; Wang, X.; Tharp, J. M.; Vatansever, E. C.; Hsu, W. W.; Dai, S.; Fang, X.; Liu, W. R. A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew Chem Int Ed Engl* **2017**, *56* (1), 212-216. DOI: 10.1002/anie.201609452 From NLM Medline.

(31) Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **1964**, *51*, 786-794. DOI: 10.1073/pnas.51.5.786 From NLM Medline.

(32) Neumann, H.; Hancock, S. M.; Buning, R.; Routh, A.; Chapman, L.; Somers, J.; Owen-Hughes, T.; van Noort, J.; Rhodes, D.; Chin, J. W. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* **2009**, *36* (1), 153-163. DOI: 10.1016/j.molcel.2009.07.027 From NLM Medline. Umehara, T.; Kim, J.; Lee, S.; Guo, L. T.; Soll, D.; Park, H. S. N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* **2012**, *586* (6), 729-733. DOI: 10.1016/j.febslet.2012.01.029 From NLM Medline.

(33) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **2011**, *146* (6), 1016-1028. DOI: 10.1016/j.cell.2011.08.008 From NLM Medline.

(34) Ruiz-Andres, O.; Sanchez-Nino, M. D.; Cannata-Ortiz, P.; Ruiz-Ortega, M.; Egido, J.; Ortiz, A.; Sanz, A. B. Histone lysine crotonylation during acute kidney injury in mice. *Dis Model Mech* **2016**, *9* (6), 633-645. DOI: 10.1242/dmm.024455 From NLM Medline.

(35) Chan, J. C.; Maze, I. Histone Crotonylation Makes Its Mark in Depression Research. *Biol Psychiatry* **2019**, *85* (8), 616-618. DOI: 10.1016/j.biopsych.2019.01.025 From NLM Medline.

(36) Gattner, M. J.; Vrabel, M.; Carell, T. Synthesis of epsilon-N-propionyl-, epsilon-N-butyryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem Commun (Camb)* **2013**, *49* (4), 379-381. DOI: 10.1039/c2cc37836a From NLM Medline.

(37) Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. A genetically encoded photocaged Nepsilon-methyl-L-lysine. *Mol Biosyst* **2010**, *6* (9), 1557-1560. DOI: 10.1039/c002155e From NLM Medline.

(38) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472* (7344), 499-503. DOI: 10.1038/nature09929 From NLM Medline.

(39) Suzuki, T.; Miller, C.; Guo, L. T.; Ho, J. M. L.; Bryson, D. I.; Wang, Y. S.; Liu, D. R.; Soll, D. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat Chem Biol* **2017**, *13* (12), 1261-1266. DOI: 10.1038/nchembio.2497 From NLM Medline.

(40) Fischer, J. T.; Soll, D.; Tharp, J. M. Directed Evolution of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase Generates a Hyperactive and Highly Selective Variant. *Front Mol Biosci* **2022**, *9*, 850613. DOI: 10.3389/fmolb.2022.850613 From NLM PubMed-not-MEDLINE.

(41) Tyrrell, D. A.; Bynoe, M. L. Cultivation of a Novel Type of Common-Cold Virus in Organ Cultures. *Br Med J* **1965**, *1* (5448), 1467-1470. DOI: 10.1136/bmj.1.5448.1467 From NLM Medline.

(42) Almeida, J. D.; Tyrrell, D. A. The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *J Gen Virol* **1967**, *1* (2), 175-178. DOI: 10.1099/0022-1317-1-2-175 From NLM Medline. Virology: Coronaviruses. *Nature* **1968**, *220* (5168), 650-650. DOI: 10.1038/220650b0.

(43) Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A.; et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **2003**, *348* (20), 1967-1976. DOI: 10.1056/NEJMoa030747 From NLM. Zaki, A. M.; van Boheemen, S.; Bestebroer, T. M.; Osterhaus, A. D.; Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **2012**, *367* (19), 1814-1820. DOI: 10.1056/NEJMoa1211721 From NLM Medline. Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395* (10223), 470-473. DOI: 10.1016/S0140-6736(20)30185-9 From NLM Medline.

(44) Tregoning, J. S.; Flight, K. E.; Higham, S. L.; Wang, Z.; Pierce, B. F. Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape. *Nat Rev Immunol* **2021**, *21* (10), 626-636. DOI: 10.1038/s41577-021-00592-1 From NLM Medline.

(45) Low, Z. Y.; Yip, A. J. W.; Sharma, A.; Lal, S. K. SARS coronavirus outbreaks past and present-a comparative analysis of SARS-CoV-2 and its predecessors. *Virus Genes* **2021**, *57* (4), 307-317. DOI: 10.1007/s11262-021-01846-9 From NLM Medline. Rahimi, A.; Mirzazadeh, A.; Tavakolpour, S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* **2021**, *113* (1 Pt 2), 1221-1232. DOI: 10.1016/j.ygeno.2020.09.059 From NLM Medline.

(46) Cevik, M.; Kuppalli, K.; Kindrachuk, J.; Peiris, M. Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ* **2020**, *371*, m3862. DOI: 10.1136/bmj.m3862 From NLM Medline.

(47) Bestle, D.; Heindl, M. R.; Limburg, H.; Van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Hardes, K.; Eickmann, M.; Dolnik, O.; et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance* **2020**, *3* (9). DOI: 10.26508/lsa.202000786 From NLM.

(48) Thakur, V.; Ratho, R. K. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. *J Med Virol* **2022**, *94* (5), 1821-1824. DOI: 10.1002/jmv.27541 From NLM Medline. Rodriguez-Sevilla, J. J.; Guerri-Fernadez, R.; Bertran Recasens, B. Is There Less Alteration of Smell Sensation in Patients With Omicron SARS-CoV-2 Variant Infection? *Front Med (Lausanne)* **2022**, *9*, 852998. DOI: 10.3389/fmed.2022.852998 From NLM PubMed-not-MEDLINE.

(49) Omicron: a shift in the biology of SARS-CoV-2. *Nat Microbiol* **2022**, *7* (8), 1114-1115. DOI: 10.1038/s41564-022-01149-1 From NLM Medline.

(50) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583* (7816), 459-468.

(51) Baez-Santos, Y. M.; Mielech, A. M.; Deng, X.; Baker, S.; Mesecar, A. D. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J Virol* **2014**, *88* (21), 12511-12527. DOI: 10.1128/JVI.01294-14 From NLM Medline.

(52) Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemRxiv* **2020**. DOI: 10.26434/chemrxiv.11728983.v1 10.26434/chemrxiv.11728983 From NLM PubMed-not-MEDLINE.

(53) Gorbalenya, A. E.; Donchenko, A. P.; Blinov, V. M.; Koonin, E. V. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Lett* **1989**, *243* (2), 103-114. DOI: 10.1016/0014-5793(89)80109-7 From NLM Medline. Liu, C. L.; Guo, J.; Zhang, X.; Sukhova, G. K.; Libby, P.; Shi, G. P. Cysteine protease cathepsins in cardiovascular disease: from basic research to clinical trials. *Nat Rev Cardiol* **2018**, *15* (6), 351-370. DOI: 10.1038/s41569-018-0002-3 From NLM Medline. Verma, S.; Dixit, R.; Pandey, K. C. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front Pharmacol* **2016**, *7*, 107. DOI: 10.3389/fphar.2016.00107 From NLM PubMed-not-MEDLINE. Siklos, M.; BenAissa, M.; Thatcher, G. R. Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. *Acta Pharm Sin B* **2015**, *5* (6), 506-519. DOI: 10.1016/j.apsb.2015.08.001 From NLM.

(54) Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.; Snipas, S. J.; El Oualid, F.; Huang, T. T.; Bekes, M.; Drag, M.; et al. Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design. *bioRxiv* **2020**. DOI: 10.1101/2020.04.29.068890 From NLM PubMed-not-MEDLINE.

(55) Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nature Reviews Immunology* **2005**, *5* (5), 375-386. DOI: 10.1038/nri1604.

(56) Perng, Y. C.; Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol* **2018**, *16* (7), 423-439. DOI: 10.1038/s41579-018-0020-5 From NLM Medline. Shemesh, M.; Lochte, S.; Piehler, J.; Schreiber, G. IFNAR1 and IFNAR2 play distinct roles in initiating type I interferon-induced JAK-STAT signaling and activating STATs. *Sci Signal* **2021**, *14* (710), eabe4627. DOI: 10.1126/scisignal.abe4627 From NLM Medline.

(57) Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.; Bernardini, J. P.; Lu, B. G.; Kuchel, N. W.; Grohmann, C.; Shibata, Y.; et al. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J* **2020**, *39* (18), e106275. DOI: 10.15252/embj.2020106275 From NLM Medline.

(58) Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A. R.; Tascher, G.; et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020**, *587* (7835), 657-662. DOI: 10.1038/s41586-020-2601-5 From NLM Medline. Munnur, D.; Teo, Q.; Eggermont, D.; Lee, H. H. Y.; Thery, F.; Ho, J.; van Leur, S. W.; Ng, W. W. S.; Siu, L. Y. L.; Beling, A.; et al. Altered ISGylation drives aberrant macrophage-dependent immune responses during SARS-CoV-2 infection. *Nat Immunol* **2021**, *22* (11), 1416-1427. DOI: 10.1038/s41590-021-01035-8 From NLM Medline.

(59) Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.; Crich, D.; Hogan, R. J.; Tripp, R. A.; Pegan, S. D. Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect Dis* **2020**, *6* (8), 2099-2109. DOI: 10.1021/acsinfecdis.0c00168 From NLM Medline.

(60) Zhao, Y.; Du, X.; Duan, Y.; Pan, X.; Sun, Y.; You, T.; Han, L.; Jin, Z.; Shang, W.; Yu, J.; et al. High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. *Protein Cell* **2021**, *12* (11), 877-888. DOI: 10.1007/s13238-021-00836-9 From NLM Medline.

(61) Ratia, K.; Pegan, S.; Takayama, J.; Sleeman, K.; Coughlin, M.; Baliji, S.; Chaudhuri, R.; Fu, W.; Prabhakar, B. S.; Johnson, M. E.; et al. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc Natl Acad Sci U S A* **2008**, *105* (42), 16119-16124. DOI: 10.1073/pnas.0805240105 From NLM Medline.

(62) Shen, Z.; Ratia, K.; Cooper, L.; Kong, D.; Lee, H.; Kwon, Y.; Li, Y.; Alqarni, S.; Huang, F.; Dubrovskyi, O.; et al. Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity. *J Med Chem* **2022**, *65* (4), 2940-2955. DOI: 10.1021/acs.jmedchem.1c01307 From NLM Medline.

(63) Hossain, M. U.; Bhattacharjee, A.; Emon, M. T. H.; Chowdhury, Z. M.; Ahammad, I.; Mosaib, M. G.; Moniruzzaman, M.; Rahman, M. H.; Islam, M. N.; Ahmed, I.; et al. Novel mutations in NSP-1 and PLPro of SARS-CoV-2 NIB-1 genome mount for effective therapeutics. *J Genet Eng Biotechnol* **2021**, *19* (1), 52. DOI: 10.1186/s43141-021-00152-z From NLM PubMed-not-MEDLINE.

(64) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2010**, *2* (6), 949-964. DOI: 10.4155/fmc.10.21 From NLM Medline.

(65) Thiel, V.; Ivanov, K. A.; Putics, Á.; Hertzig, T.; Schelle, B.; Bayer, S.; Weißbrich, B.; Snijder, E. J.; Rabenau, H.; Doerr, H. W.; et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* **2003**, *84* (Pt 9), 2305-2315. DOI: 10.1099/vir.0.19424-0 From NLM.

(66) Resnick, S. J.; Iketani, S.; Hong, S. J.; Zask, A.; Liu, H.; Kim, S.; Melore, S.; Lin, F.-Y.; Nair, M. S.; Huang, Y. Inhibitors of coronavirus 3CL proteases protect cells from protease-mediated cytotoxicity. *Journal of Virology* **2021**, *95* (14), e02374-02320.

(67) Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; et al. An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374* (6575), 1586-1593. DOI: 10.1126/science.abl4784 From NLM Medline.

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

(69) Yang, K. S.; Ma, X. R.; Ma, Y.; Alugubelli, Y. R.; Scott, D. A.; Vatansever, E. C.; Drelich, A. K.; Sankaran, B.; Geng, Z. Z.; Blankenship, L. R. A quick route to multiple highly potent SARS-CoV-2 main protease inhibitors. *ChemMedChem* **2021**, *16* (6), 942-948.

(70) Katre, S. G.; Asnani, A. J.; Pratyush, K.; Sakharkar, N. G.; Bhope, A. G.; Sawarkar, K. T.; Nimbekar, V. S. Review on development of potential inhibitors of SARS-CoV-2 main protease (M(Pro)). *Futur J Pharm Sci* **2022**, *8* (1), 36. DOI: 10.1186/s43094-022-00423-7 From NLM PubMed-not-MEDLINE.

(71) Polycarpo, C.; Ambrogelly, A.; Ruan, B.; Tumbula-Hansen, D.; Ataide, S. F.; Ishitani, R.; Yokoyama, S.; Nureki, O.; Ibba, M.; Soll, D. Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II lysyl-tRNA synthetases. *Mol Cell* **2003**, *12* (2), 287-294. DOI: 10.1016/s1097-2765(03)00280-6 From NLM Medline.

(72) Namy, O.; Zhou, Y.; Gundllapalli, S.; Polycarpo, C. R.; Denise, A.; Rousset, J. P.; Soll, D.; Ambrogelly, A. Adding pyrrolysine to the Escherichia coli genetic code. *FEBS Lett* **2007**, *581* (27), 5282-5288. DOI: 10.1016/j.febslet.2007.10.022 From NLM Medline.

Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem Biophys Res Commun* **2008**, *371* (4), 818-822. DOI: 10.1016/j.bbrc.2008.04.164 From NLM Medline. Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. A pyrrolysine analogue for protein click chemistry. *Angew Chem Int Ed Engl* **2009**, *48* (9), 1633-1635. DOI: 10.1002/anie.200805420 From NLM Medline. Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. A facile system for encoding unnatural amino acids in mammalian cells. *Angew Chem Int Ed Engl* **2009**, *48* (22), 4052-4055. DOI: 10.1002/anie.200900683 From NLM Medline.

(73) Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *J Am Chem Soc* **2009**, *131* (40), 14194-14195. DOI: 10.1021/ja906603s From NLM Medline. Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc* **2009**, *131* (25), 8720-8721. DOI: 10.1021/ja900553w From NLM Medline. Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W. Traceless and site-specific ubiquitination of recombinant proteins. *J Am Chem Soc* **2011**, *133* (28), 10708-10711. DOI: 10.1021/ja202799r From NLM Medline. Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew Chem Int Ed Engl* **2010**, *49* (18), 3211-3214. DOI: 10.1002/anie.201000465 From NLM Medline. Huang, Y.; Russell, W. K.; Wan, W.; Pai, P. J.; Russell, D. H.; Liu, W. A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. *Mol Biosyst* **2010**, *6* (4), 683-686. DOI: 10.1039/b920120c From NLM Medline. Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R. Genetically encoded cleavable protein photo-cross-linker. *J Am Chem Soc* **2014**, *136* (34), 11860-11863. DOI: 10.1021/ja504371w From NLM Medline. Zhang, M.; Lin, S.; Song, X.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.; Chang, Z.; Chen, P. R. A

genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat Chem Biol* **2011**, *7* (10), 671-677. DOI: 10.1038/nchembio.644 From NLM Medline. Li, J.; Jia, S.; Chen, P. R. Diels-Alder reaction-triggered bioorthogonal protein decaging in living cells. *Nat Chem Biol* **2014**, *10* (12), 1003-1005. DOI: 10.1038/nchembio.1656 From NLM Medline. Xiang, Z.; Ren, H.; Hu, Y. S.; Coin, I.; Wei, J.; Cang, H.; Wang, L. Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat Methods* **2013**, *10* (9), 885-888. DOI: 10.1038/nmeth.2595 From NLM Medline. Hoppmann, C.; Lacey, V. K.; Louie, G. V.; Wei, J.; Noel, J. P.; Wang, L. Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. *Angew Chem Int Ed Engl* **2014**, *53* (15), 3932-3936. DOI: 10.1002/anie.201400001 From NLM Medline. Xiang, Z.; Lacey, V. K.; Ren, H.; Xu, J.; Burban, D. J.; Jennings, P. A.; Wang, L. Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* **2014**, *53* (8), 2190-2193. DOI: 10.1002/anie.201308794 From NLM Medline. Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J Am Chem Soc* **2012**, *134* (29), 11912-11915. DOI: 10.1021/ja3046958 From NLM Medline. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* **2012**, *4* (4), 298-304. DOI: 10.1038/nchem.1250 From NLM Medline. Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J Am Chem Soc* **2012**, *134* (25), 10317-10320. DOI: 10.1021/ja302832g From NLM Medline. Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew Chem Int Ed Engl* **2011**, *50* (17), 3878-3881. DOI: 10.1002/anie.201008178 From NLM Medline. Plass, T.; Milles, S.; Koehler, C.; Szymanski, J.; Mueller, R.; Wiessler, M.; Schultz, C.; Lemke, E. A. Amino acids for Diels-Alder reactions in living cells. *Angew*

Chem Int Ed Engl **2012**, *51* (17), 4166-4170. DOI: 10.1002/anie.201108231 From NLM Medline.

(74) Borrel, G.; Harris, H. M.; Tottey, W.; Mihajlovski, A.; Parisot, N.; Peyretaillade, E.; Peyret, P.; Gribaldo, S.; O'Toole, P. W.; Brugere, J. F. Genome sequence of "Candidatus Methanomethylophilus alvus" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J Bacteriol* **2012**, *194* (24), 6944-6945. DOI: 10.1128/JB.01867-12 From NLM Medline.

(75) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **2006**, *24* (1), 79-88. DOI: 10.1038/nbt1172 From NLM Medline. Tsien, R. Y. The green fluorescent protein. *Annu Rev Biochem* **1998**, *67*, 509-544. DOI: 10.1146/annurev.biochem.67.1.509 From NLM Medline.

(76) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. The case of the missing ring: radical cleavage of a carbon-carbon bond and implications for GFP chromophore biosynthesis. *J Am Chem Soc* **2007**, *129* (11), 3118-3126. DOI: 10.1021/ja063983u From NLM Medline.

(77) Jiang, X.; Zhang, M.; Ding, Y.; Yao, J.; Chen, H.; Zhu, D.; Muramatu, M. Escherichia coli prlC gene encodes a trypsin-like proteinase regulating the cell cycle. *J Biochem* **1998**, *124* (5), 980-985. DOI: 10.1093/oxfordjournals.jbchem.a022216 From NLM Medline.

(78) Gottesman, S. Proteases and their targets in Escherichia coli. *Annu Rev Genet* **1996**, *30*, 465-506. DOI: 10.1146/annurev.genet.30.1.465 From NLM Medline.

(79) Imai, K.; Mitaku, S. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics (Nagoya-shi)* **2005**, *1*, 55-65. DOI: 10.2142/biophysics.1.55 From NLM PubMed-not-MEDLINE.

(80) Wang, W. W.; Angulo-Ibanez, M.; Lyu, J.; Kurra, Y.; Tong, Z.; Wu, B.; Zhang, L.; Sharma, V.; Zhou, J.; Lin, H.; et al. A Click Chemistry Approach Reveals the Chromatin-Dependent Histone H3K36 Deacylase Nature of SIRT7. *J Am Chem Soc* **2019**, *141* (6), 2462-2473. DOI: 10.1021/jacs.8b12083 From NLM Medline. Wang, W. W.; Zeng, Y.; Wu, B.; Deiters, A.; Liu, W. R. A Chemical Biology Approach to Reveal Sirt6-targeted Histone H3 Sites in Nucleosomes. *ACS Chem Biol* **2016**, *11* (7), 1973-1981. DOI: 10.1021/acschembio.6b00243 From NLM Medline.

(81) Cho, C. C.; Li, S. G.; Lalonde, T. J.; Yang, K. S.; Yu, G.; Qiao, Y.; Xu, S.; Ray Liu, W. Drug Repurposing for the SARS-CoV-2 Papain-Like Protease. *ChemMedChem* **2022**, *17* (1), e202100455. DOI: 10.1002/cmdc.202100455 From NLM Medline.

(82) Gates, B. Responding to Covid-19 - A Once-in-a-Century Pandemic? *N Engl J Med* **2020**, *382* (18), 1677-1679. DOI: 10.1056/NEJMp2003762 From NLM Medline. Morens, D. M.; Daszak, P.; Taubenberger, J. K. Escaping Pandora's Box - Another Novel Coronavirus. *N Engl J Med* **2020**, *382* (14), 1293-1295. DOI: 10.1056/NEJMp2002106 From NLM Medline.

(83) Vatansever, E. C.; Yang, K.; Kratch, K. C.; Drelich, A.; Cho, C. C.; Mellott, D. M.; Xu, S.; Tseng, C. K.; Liu, W. R. Bepridil is potent against SARS-CoV-2 In Vitro. *bioRxiv* **2020**. DOI: 10.1101/2020.05.23.112235 From NLM PubMed-not-MEDLINE.

(84) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* **2020**, *395* (10236), 1569-1578. DOI: 10.1016/S0140-6736(20)31022-9 From NLM Medline.

(85) Lam, T. T.; Jia, N.; Zhang, Y. W.; Shum, M. H.; Jiang, J. F.; Zhu, H. C.; Tong, Y. G.; Shi, Y. X.; Ni, X. B.; Liao, Y. S.; et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **2020**, *583* (7815), 282-285. DOI: 10.1038/s41586-020-2169-0 From NLM Medline. Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, -.

R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B.; Lauber, C.; Leontovich, A.; Neuman, B. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol* **2020**, *5* (4), 536-544. Zhou, P.; Yang, X.; Wang, X.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.; Zhu, Y.; Li, B.; Huang, C. Regulation of inflammatory cytokines and inhibition of T and B lymphocytes X. *Wang, XS Zheng, K. Zhao, QJ Chen, F. Deng, LL Liu, B. Yan, FX Zhan, YY Wang, GF Xiao, ZL Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature* **2020**, *579*, 270-273.

(86) Phan, T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol* **2020**, *81*, 104260. DOI: 10.1016/j.meegid.2020.104260 From NLM Medline.

(87) Chen, Y. W.; Yiu, C. B.; Wong, K. Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL (pro)) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Res* **2020**, *9*, 129. DOI: 10.12688/f1000research.22457.2 From NLM Medline.

(88) Naqvi, A. A. T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I. K.; Singh, A.; Atif, S. M.; Hariprasad, G.; Hasan, G. M.; Hassan, M. I. 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.

(89) Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A.; Tascher, G. GJ van de r Heden van Noort, H. *Ovaa, S. Müller, KP Knobeloch, K. Rajalingam, B. A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, I. Dikic, Papain-like proteas e regulates SARS-CoV-2 viral spread and innate immunity, Nature* **2020**.

(90) Baez-Santos, Y. M.; St John, S. E.; Mesecar, A. D. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res* **2015**, *115*, 21-38. DOI: 10.1016/j.antiviral.2014.12.015 From NLM Medline.

(91) McClain, C. B.; Vabret, N. SARS-CoV-2: the many pros of targeting PLpro. *Signal Transduct Target Ther* **2020**, *5* (1), 223. DOI: 10.1038/s41392-020-00335-z From NLM Medline. Totura, A. L.; Bavari, S. Broad-spectrum coronavirus antiviral drug discovery. *Expert Opin Drug Discov* **2019**, *14* (4), 397-412. DOI: 10.1080/17460441.2019.1581171 From NLM Medline. Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent Sci* **2020**, *6* (3), 315-331. DOI: 10.1021/acscentsci.0c00272 From NLM PubMed-not-MEDLINE. Petushkova, A. I.; Zamyatnin Jr, A. A. Papain-like proteases as coronaviral drug targets: Current inhibitors, opportunities, and limitations. *Pharmaceuticals* **2020**, *13* (10), 277.

(92) Rut, W.; Lv, Z.; Zmudzinski, M.; Patchett, S.; Nayak, D.; Snipas, S. J.; El Oualid, F.; Huang, T. T.; Bekes, M.; Drag, M.; et al. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. *Sci Adv* **2020**, *6* (42). DOI: 10.1126/sciadv.abd4596 From NLM Medline.

(93) Qiao, Y.; Yu, G.; Kratch, K. C.; Wang, X. A.; Wang, W. W.; Leeuwon, S. Z.; Xu, S.; Morse, J. S.; Liu, W. R. Expressed Protein Ligation without Intein. *J Am Chem Soc* **2020**, *142* (15), 7047-7054. DOI: 10.1021/jacs.0c00252 From NLM Medline.

(94) Lei, J.; Kusov, Y.; Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res* **2018**, *149*, 58-74. DOI: 10.1016/j.antiviral.2017.11.001 From NLM Medline. Lindner, H. A.; Fotouhi-Ardakani, N.; Lytvyn, V.; Lachance, P.; Sulea, T.; Menard, R. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J Virol* **2005**,

79 (24), 15199-15208. DOI: 10.1128/JVI.79.24.15199-15208.2005 From NLM Medline.
Ratia, K.; Saikatendu, K. S.; Santarsiero, B. D.; Barretto, N.; Baker, S. C.; Stevens, R. C.; Mesecar, A. D. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc Natl Acad Sci U S A* **2006**, *103* (15), 5717-5722. DOI: 10.1073/pnas.0510851103 From NLM Medline.

(95) Liu, M.; Wang, B.; Wang, F.; Yang, Z.; Gao, D.; Zhang, C.; Ma, L.; Yu, X. Soluble expression of single-chain variable fragment (scFv) in Escherichia coli using superfolder green fluorescent protein as fusion partner. *Appl Microbiol Biotechnol* **2019**, *103* (15), 6071-6079. DOI: 10.1007/s00253-019-09925-6 From NLM Medline.

(96) Baez-Santos, Y. M.; Barraza, S. J.; Wilson, M. W.; Agius, M. P.; Mielech, A. M.; Davis, N. M.; Baker, S. C.; Larsen, S. D.; Mesecar, A. D. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J Med Chem* **2014**, *57* (6), 2393-2412. DOI: 10.1021/jm401712t From NLM Medline.

(97) Fu, Z.; Huang, B.; Tang, J.; Liu, S.; Liu, M.; Ye, Y.; Liu, Z.; Xiong, Y.; Zhu, W.; Cao, D.; et al. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat Commun* **2021**, *12* (1), 488. DOI: 10.1038/s41467-020-20718-8 From NLM Medline.

(98) Gao, X.; Qin, B.; Chen, P.; Zhu, K.; Hou, P.; Wojdyla, J. A.; Wang, M.; Cui, S. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm Sin B* **2021**, *11* (1), 237-245. DOI: 10.1016/j.apsb.2020.08.014 From NLM PubMed-not-MEDLINE.

(99) Liu, Y.; Lashuel, H. A.; Choi, S.; Xing, X.; Case, A.; Ni, J.; Yeh, L. A.; Cuny, G. D.; Stein, R. L.; Lansbury, P. T., Jr. Discovery of inhibitors that elucidate the role of UCH-L1 activity in the H1299 lung cancer cell line. *Chem Biol* **2003**, *10* (9), 837-846. DOI: 10.1016/j.chembiol.2003.08.010 From NLM.

(100) Altun, M.; Kramer, H. B.; Willems, L. I.; McDermott, J. L.; Leach, C. A.; Goldenberg, S. J.; Kumar, K. G.; Konietzny, R.; Fischer, R.; Kogan, E.; et al. Activity-based chemical proteomics accelerates inhibitor development for deubiquitylating enzymes. *Chem Biol* **2011**, *18* (11), 1401-1412. DOI: 10.1016/j.chembiol.2011.08.018 From NLM Medline.

(101) Fu, Y.; Hong, L.; Xu, J.; Zhong, G.; Gu, Q.; Gu, Q.; Guan, Y.; Zheng, X.; Dai, Q.; Luo, X.; et al. Discovery of a small molecule targeting autophagy via ATG4B inhibition and cell death of colorectal cancer cells in vitro and in vivo. *Autophagy* **2019**, *15* (2), 295-311. DOI: 10.1080/15548627.2018.1517073 From NLM Medline.

(102) Mistry, H.; Hsieh, G.; Buhrlage, S. J.; Huang, M.; Park, E.; Cuny, G. D.; Galinsky, I.; Stone, R. M.; Gray, N. S.; D'Andrea, A. D.; et al. Small-molecule inhibitors of USP1 target ID1 degradation in leukemic cells. *Mol Cancer Ther* **2013**, *12* (12), 2651-2662. DOI: 10.1158/1535-7163.MCT-13-0103-T From NLM Medline.

(103) Colombo, M.; Vallese, S.; Peretto, I.; Jacq, X.; Rain, J. C.; Colland, F.; Guedat, P. Synthesis and biological evaluation of 9-oxo-9H-indeno[1,2-b]pyrazine-2,3-dicarbonitrile analogues as potential inhibitors of deubiquitinating enzymes. *ChemMedChem* **2010**, *5* (4), 552-558. DOI: 10.1002/cmdc.200900409 From NLM Medline.

(104) Cao, W.; Cho, C. D.; Geng, Z. Z.; Shaabani, N.; Ma, X. R.; Vatansever, E. C.; Alugubelli, Y. R.; Ma, Y.; Chaki, S. P.; Ellenburg, W. H.; et al. Evaluation of SARS-CoV-2 Main Protease Inhibitors Using a Novel Cell-Based Assay. *ACS Cent Sci* **2022**, *8* (2), 192-204. DOI: 10.1021/acscentsci.1c00910 From NLM PubMed-not-MEDLINE.

(105) WHO, S. A. WHO coronavirus disease (COVID-19) dashboard. *World Health Organization* **2020**.

(106) Forni, G.; Mantovani, A. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death & Differentiation* **2021**, *28* (2), 626-639.

(107) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581* (7807), 215-220.

(108) Guruprasad, L. Human SARS CoV-2 spike protein mutations. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89* (5), 569-576.

(109) Jangra, S.; Ye, C.; Rathnasinghe, R.; Stadlbauer, D.; Alshammary, H.; Amoako, A. A.; Awawda, M. H.; Beach, K. F.; Bermúdez-González, M. C.; Chernet, R. L. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *The Lancet Microbe* **2021**, *2* (7), e283-e284.

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

(111) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368* (6497), 1331-1335.

(112) Günther, S.; Reinke, P. Y.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H.; Ehrt, C.; Ewert, W.; Oberthuer, D. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372* (6542), 642-646.

(113) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A. 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. e278.

(114) Zhao, M.-M.; Yang, W.-L.; Yang, F.-Y.; Zhang, L.; Huang, W.-J.; Hou, W.; Fan, C.-F.; Jin, R.-H.; Feng, Y.-M.; Wang, Y.-C. Cathepsin L plays a key role in SARS-CoV-

2 infection in humans and humanized mice and is a promising target for new drug development. *Signal transduction and targeted therapy* **2021**, *6* (1), 1-12.

(115) Pišlar, A.; Mitrović, A.; Sabotič, J.; Pečar Fonović, U.; Perišić Nanut, M.; Jakoš, T.; Senjor, E.; Kos, J. The role of cysteine peptidases in coronavirus cell entry and replication: The therapeutic potential of cathepsin inhibitors. *PLoS pathogens* **2020**, *16* (11), e1009013.

(116) Cheng, Y.-W.; Chao, T.-L.; Li, C.-L.; Chiu, M.-F.; Kao, H.-C.; Wang, S.-H.; Pang, Y.-H.; Lin, C.-H.; Tsai, Y.-M.; Lee, W.-H. Furin inhibitors block SARS-CoV-2 spike protein cleavage to suppress virus production and cytopathic effects. *Cell reports* **2020**, *33* (2), 108254.

(117) Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y. 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.

(118) Song, Y.; Yang, M.; Wegner, S. V.; Zhao, J.; Zhu, R.; Wu, Y.; He, C.; Chen, P. R. A genetically encoded FRET sensor for intracellular heme. *ACS chemical biology* **2015**, *10* (7), 1610-1615.

(119) Kim, Y.; Lovell, S.; Tiew, K.-C.; Mandadapu, S. R.; Alliston, K. R.; Battaile, K. P.; Groutas, W. C.; Chang, K.-O. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *Journal of virology* **2012**, *86* (21), 11754-11762.

(120) Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.; Wu, Y.; Zhao, C.; Sun, H.; Huang, B.; Niu, P. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nature communications* **2020**, *11* (1), 1-8.

(121) Baker, J. D.; Uhrich, R. L.; Kraemer, G. C.; Love, J. E.; Kraemer, B. C. A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. *PloS one* **2021**, *16* (2), e0245962.

(122) Mellott, D. M.; Tseng, C.-T.; Drelich, A.; Fajtová, P.; Chenna, B. C.; Kostomiris, D. H.; Hsu, J.; Zhu, J.; Taylor, Z. W.; Kocurek, K. I. A clinical-stage cysteine protease inhibitor blocks SARS-CoV-2 infection of human and monkey cells. *ACS chemical biology* **2021**, *16* (4), 642-650.

(123) Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nature structural & molecular biology* **2020**, *27* (6), 529-532.

(124) Vatansever, E. C.; Yang, K. S.; Drelich, A. K.; Kratch, K. C.; Cho, C.-C.; Kempaiah, K. R.; Hsu, J. C.; Mellott, D. M.; Xu, S.; Tseng, C.-T. K. Bepridil is potent against SARS-CoV-2 in vitro. *Proceedings of the National Academy of Sciences* **2021**, *118* (10), e2012201118.

(125) Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H. Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proceedings of the National Academy of Sciences* **2020**, *117* (44), 27381-27387.

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

(127) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *Journal of medicinal chemistry* **2016**, *59* (14), 6595-6628.

(128) Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *Journal of medicinal chemistry* **2020**, *63* (21), 12725-12747.

(129) Sacco, M. D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J. A.; Meng, X.; Dube, P.; Zhang, X.; Hu, Y.; Kitamura, N. 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.

(130) DeWald, L. E.; Dyall, J.; Sword, J. M.; Torzewski, L.; Zhou, H.; Postnikova, E.; Kollins, E.; Alexander, I.; Gross, R.; Cong, Y. The calcium channel blocker bepridil demonstrates efficacy in the murine model of marburg virus disease. *The Journal of infectious diseases* **2018**, *218* (suppl_5), S588-S591.

(131) Yao, X.; Ye, F.; Zhang, M.; Cui, C.; Huang, B.; Niu, P.; Liu, X.; Zhao, L.; Dong, E.; Song, C. In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clinical infectious diseases* **2020**, *71* (15), 732-739.

(132) Ou, T.; Mou, H.; Zhang, L.; Ojha, A.; Choe, H.; Farzan, M. Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. *PLoS pathogens* **2021**, *17* (1), e1009212.

(133) Ma, C.; Wang, J. Dipyridamole, chloroquine, montelukast sodium, candesartan, oxytetracycline, and atazanavir are not SARS-CoV-2 main protease inhibitors. *Proceedings of the National Academy of Sciences* **2021**, *118* (8), e2024420118.

(134) Ma, X. R.; Alugubelli, Y. R.; Ma, Y.; Vatansever, E. C.; Scott, D. A.; Qiao, Y.; Yu, G.; Xu, S.; Liu, W. R. MPI8 Is Potent against SARS-CoV-2 by Inhibiting Dually and Selectively the SARS-CoV-2 Main Protease and the Host Cathepsin L. *ChemMedChem* **2022**, *17* (1), e202100456.

CHAPTER VI APPENDIX A

SUPPLEMENTARY INFORMATION OF THE PYRROLYSYL-TRNA SYNTHETASE ACTIVITY CAN BE IMPROVED BY A P188 MUTATION THAT STABILIZES THE FULL-LENGTH ENZYME

Supplementary figure and tables

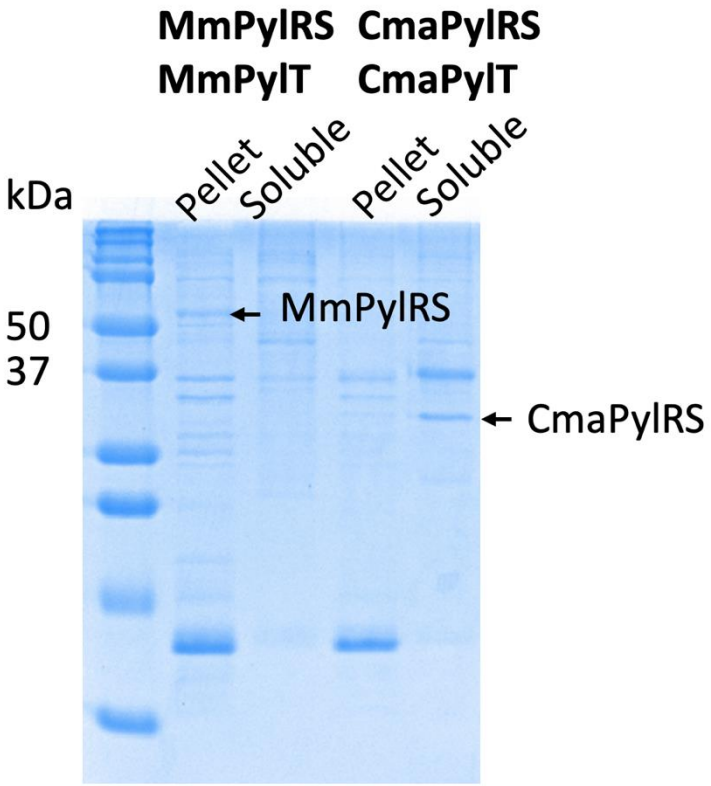

Figure 26 CMaPylRS showed much better solubility than full-length MmPylRS. Reprinted with permission from reference 22.

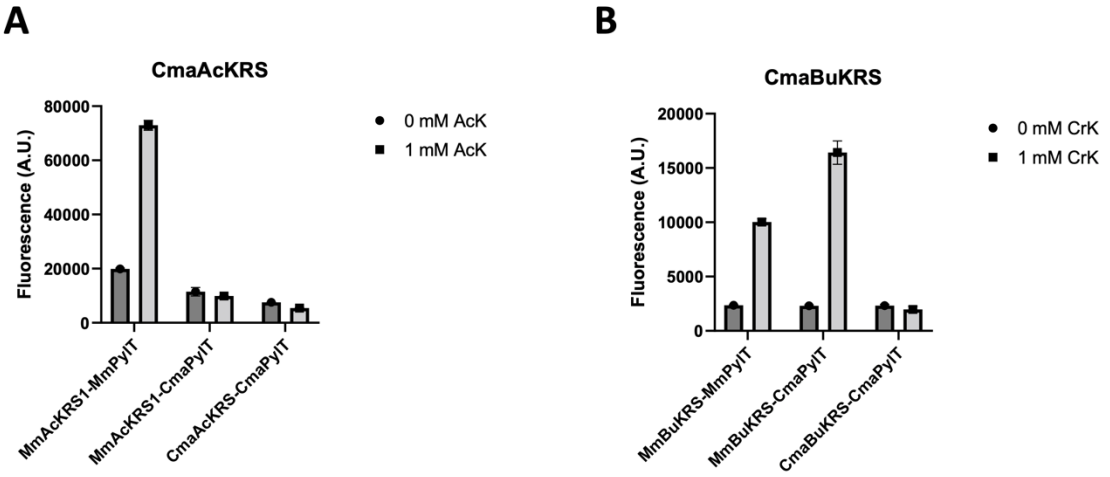

Figure 27 Mutations in ncAA-specific MmPylRS variants cannot be transferred directly to CMaPylRS for genetic incorporation of all targeted ncAAs. (A) CmaAcKRS cannot mediate genetic incorporation of AcK into sfGFP134TAG. (B) CmaBuKRS doesn't work on CrK in the presence of CmaPylT. Reprinted with permission from reference 22.

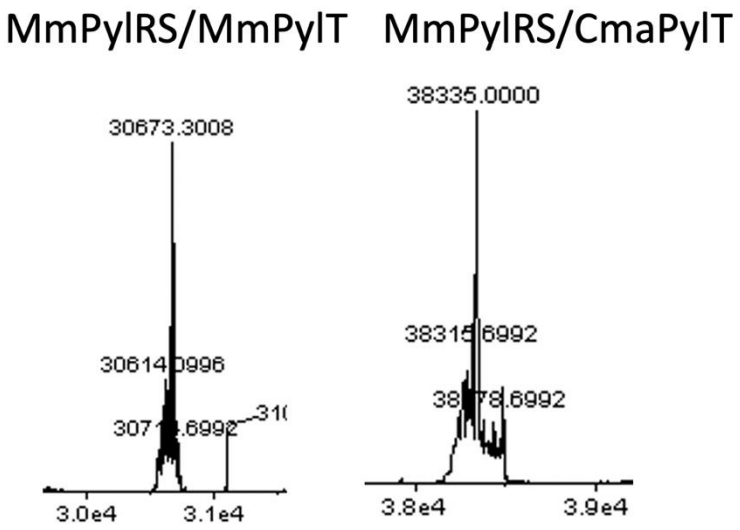

Figure 28 Molecular weights of two different cleavage products of MmPylRS in the presence of MmPylT and CmaPylT. Reprinted with permission from reference 22.

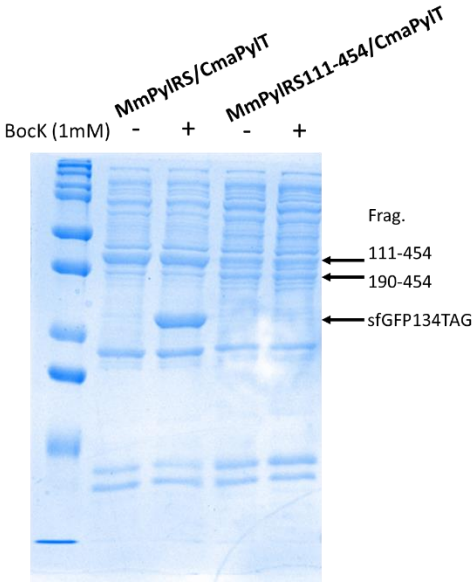

Figure 29 Removing the NTD from MmPylRS totally kill the protein activity. Reprinted with permission from reference 22.

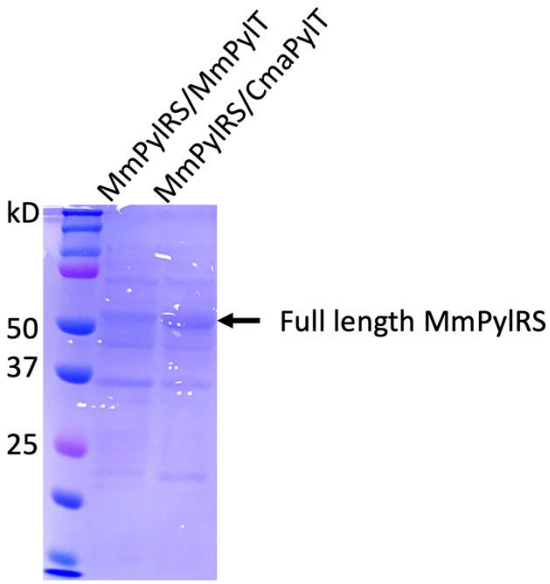

Figure 30 CmaPylT leads to stable full-length MmPylRS in E. coli cells. The band corresponding to full-length MmPylRS did not show in cells expressing MmPylRS and MmPylT. Reprinted with permission from reference 22.

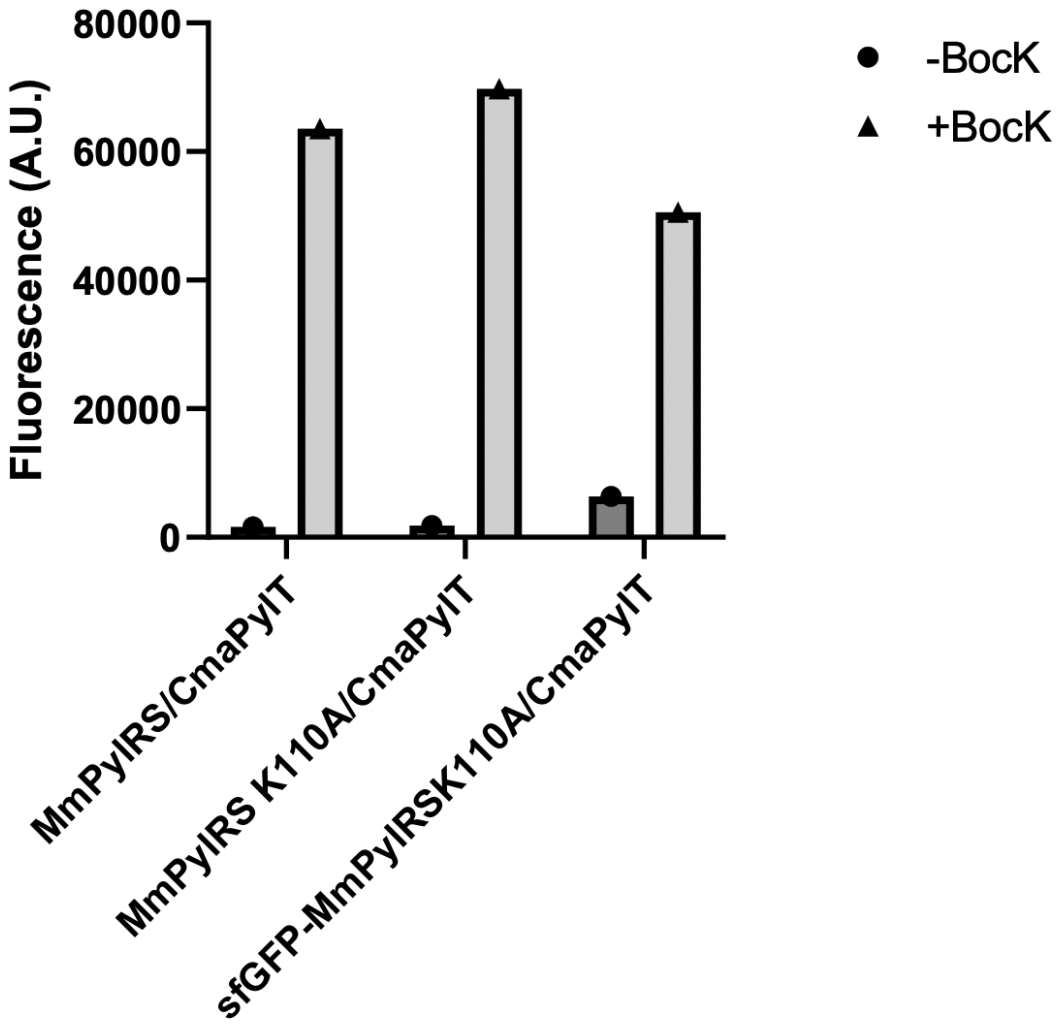

Figure 31 K110A only slightly improved amber suppression activity of MmPylRS in the presence of CmaPylT. Fusing sfGFP to MmPylRS did not improve its amber suppression efficiency. Reprinted with permission from reference 22.

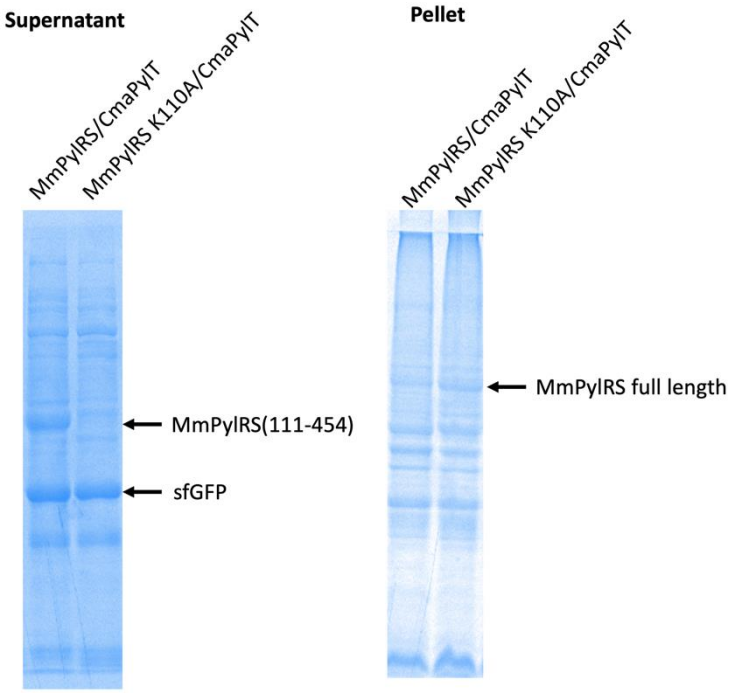

Figure 32 K110A protected MmPylRS from cleavage at aa110. More full-length mmPylRS-K110A was observed in the cell lysate pellet. Much less truncated MmPylRS(111-454) was observed in the cell lysate supernatant. Reprinted with permission from reference 22.

Table 4 Primer list. Reprinted with permission from reference 22.

primer	sequence
mmPylRS-for	ggctaacaggaggaattactagt
mmPylRS-rev	accgtttaaactcagtcgacttacaggttagtgctaatg
Split-for	gaggaattactagtATGtccgttgcacgcgctcc
Split-rev	CTAAGCTCActtcggcatcgctttcttg
sfGFP-for	atacccgtttttttgggctaacaggaggaattactagtatggttagcaaaggtgaagaac
sfGFP-rev	tgcagaaatcagagtattcagcggtttcttatccatTCTAGATTGGAAGTACAGGTTTTC
Y66F-for	ccctgaccTTTggcgttcagtgctttagccgcta
Y66F-rev	aacgccAAAggtcagggtggtcaccagg
CTD-for	tccgttgcacgcgctcc
CTD-rev	catactagtaattcctcctgttagcccaaaaaaa
K110A-rev	GGCcggcatcgctttcttggtg (pair with CTD-for)
P188Q-for	caggcttctgcaCAGgcactgactaaaagccagactgaccgtctgga
P188Q-rev	ttttagtcagtgcCTGtgcagaagcctgaaccggagcagacatgctgg
P188G-for	caggcttctgcaGGCgcactgactaaaagccagactgaccgtctgga
P188G-rev	ttttagtcagtgcgcctgcagaagcctgaaccggagcagacatgctgg

APPENDIX B

SUPPLEMENTARY INFORMATION OF CELLULAR ACTIVITIES OF SARS-COV-2 MAIN PROTEASE INHIBITORS REVEAL THEIR UNIQUE CHARACTERISTICS

Supplementary Methods

Cytopathic effects assay

Vero-E6 with high endogenous ACE2 expression (BEI Resources, NR-53726) or transgenic A549 cells expressing human ACE2 with a C-terminal FLAG tag (BEI Resources, NR-53522) were inoculated with SARS-CoV-2 WA1/2020 (BEI Resources, NR-52281) at a multiplicity of 3 plaque-forming units per cell. After 1 h, the inoculum was removed and cells were washed 3x with phosphate-buffered saline (Thermo Fisher Scientific). 1 h before collection, medium was removed and replaced. At the specified timepoints, cells were either fixed in formalin and stained with crystal violet for microscopy, or lysed with RIPA buffer (10 mM Tris-HCl, 150 mM NaCl, 1% Triton X-100, 0.5% deoxychonate, 0.5% sodium dodecyl sulfate, 1x Roche complete protease inhibitor, 1 mM phenylmethylsulfonyl fluoride, 10 mM dithiothreitol) for western blotting.

SiRNA knocking down of Mpro

The siRNA was designed and synthesized using the custom Designed Silencer® Select siRNA service from ThermoFisher. The sequence was provided in the Table S2. HEK

293T/17 cells were maintained in DMEM supplemented with 10% FBS. Upon a confluency of 80~90%, the pLVX-Puro-MPro-eGFP plasmid was transfected into the cells with PEI. After overnight incubation, the cells were dispersed and plated in white 96-well plates, 100 μL in each well with cell density of 5×105 cells/mL. The next day, medium was replaced with fluorobrite DMEM medium with 10% FBS. One group was set as blank, another one was supplemented with 1 μM MPI8. Another two groups were transfected with 10nM and 30nM siRNA with lipofectamine 3000 twice, with a time gap of 24 h. The apoptosis reading was done with the RealTime-Glo™ annexin V apoptosis assay kit from Promega. Each group was characterized with 3 repeats.

Cell toxicity assay for MPro(C145S)

HEK 293T/17 cells were maintained in DMEM supplemented with 10% FBS. Upon a confluency of 80~90%, pLVX-MPro-eGFP-2 plasmid and PLVX-MPro(C145S)-eGFP plasmids were transfected separately into the cells with PEI, with an amount of 1.6 μg for each well in a 6-well plate. After incubation overnight, the cells were dispersed and plated in a 24-well plate with high glucose fluorobrite DMEM supplemented with 10% FBS as one blank group and the other group supplemented with 1 μM MPI8. fluorescent pictures of the cells were taken at 24, 48 and 72 h, respectively.

Quntification of antimycin A induced cell apoptosis

HEK293T/17 cells were plated in a white 96-well plate with high glucose fluorobrite DMEM supplemented with 10% FBS at a cell density of 5 x 105 cells/mL and 100 μL

each well. One group was supplemented with 1 μM antimycin A, another group was supplemented with 1 μM MPI8 and S3 1 μM antimycin A. The cell apoptosis was tested using the RealTime-Glo™ annexin V apoptosis assay kit from Promega after 72h. Each group was characterized with 3 repeats.

New material introduced

High glucose FluoroBrite™ DMEM, Gibco.

Supplementary figures

Figure 33 The plasmid map of pECFP-M^{Pro}-EYFP. Reprinted with permission from reference 104.

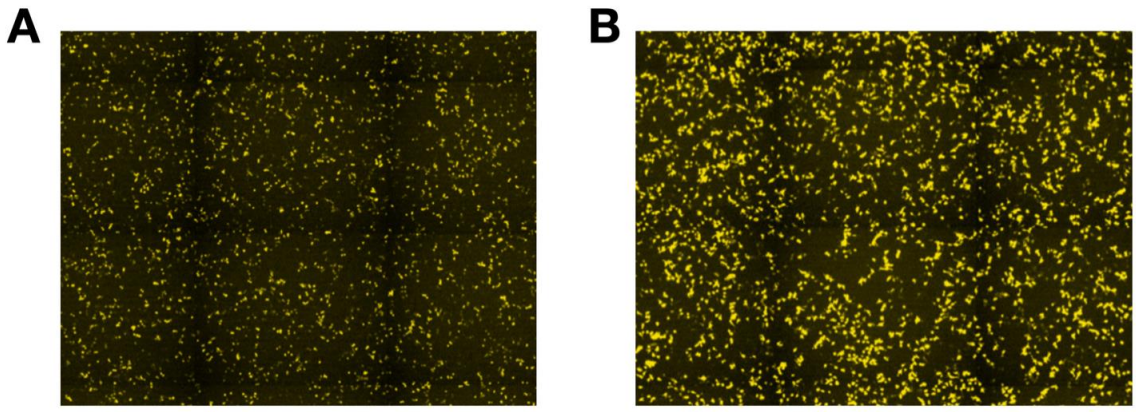

Figure 34 Yellow fluorescence from expressed CFP-M^{Pro}-YFP in 293T cells transfected with pECFP-M^{Pro}-EYFP and grown in the absence (A) or presence (B) of 10 μM MPI8. Reprinted with permission from reference 104.

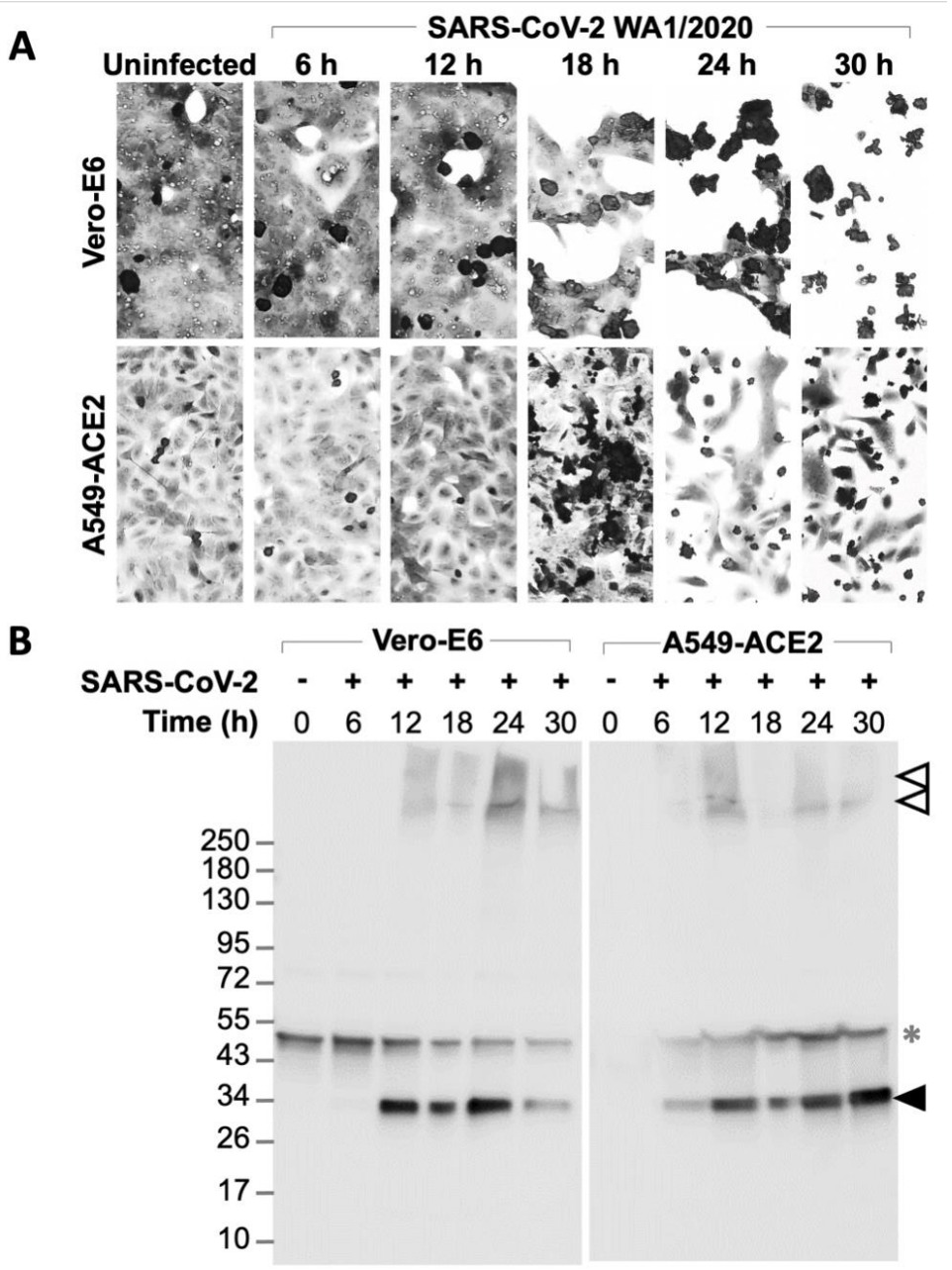

Figure 35 Association of M^{pro} expression with SARS-CoV-2 cytopathic effects. lines were inoculated with SARS-CoV-2 at different times. For each timepoint, one replicate was fixed and stained with crystal violet (A) and a second replicate was lysed for western blot with anti-M^{pro} antibody detection (B). Virus-induced cytopathic effects included extensive cell rounding (small, condensed staining) and detachment from the monolayer. Positions of viral polyprotein precursors (white triangles), ~33.7 kDa fully processed M^{pro} (black triangle) and a nonspecific staining product (asterisk) are indicated. Reprinted with permission from reference 104.

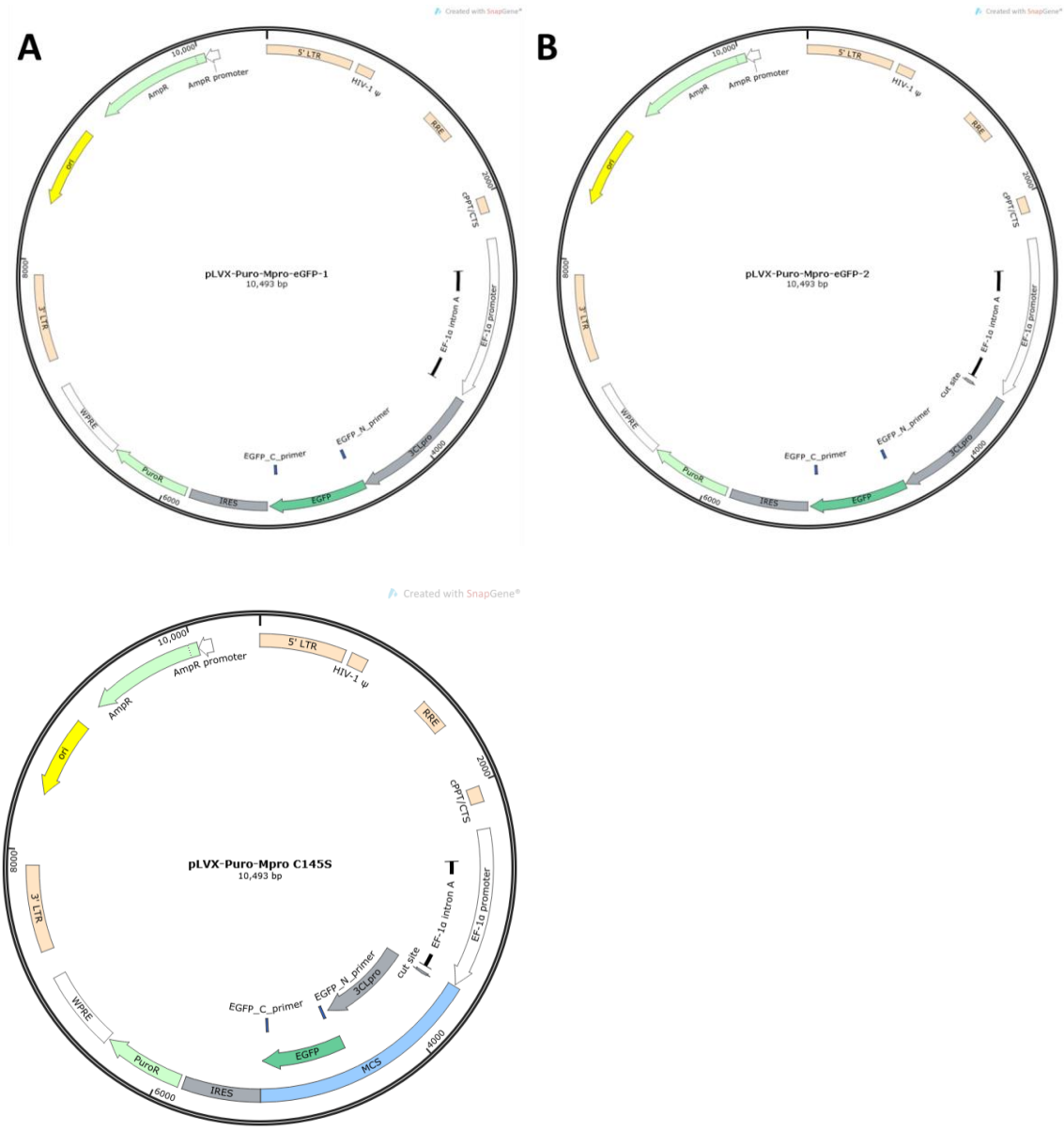

Figure 36 Plasmid maps of pLVX-M^{Pro}-eGFP-1 (A), pLVX-M^{Pro}-eGFP-2 (B) and pLVX-M^{Pro}C145S-eGFP. Reprinted with permission from reference 104.

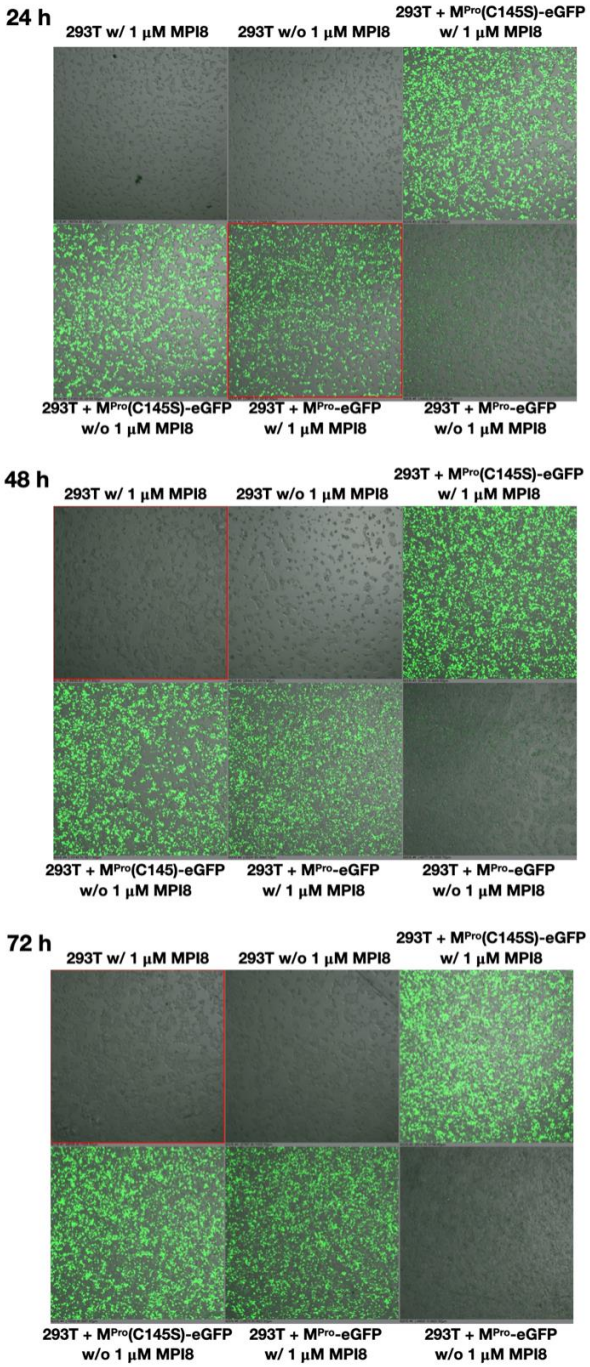

Figure 37 Cellular toxicity was from the protease function of M^{Pro}. 293T cells were transiently transferred with plasmids coding either active M^{Pro}-eGFP or inactive M^{Pro}(C145S)-eGFP and then grown with or without 1 μM MPI8 for 24, 48, and 72 h before they were fluorescently imaged. Reprinted with permission from reference 104.

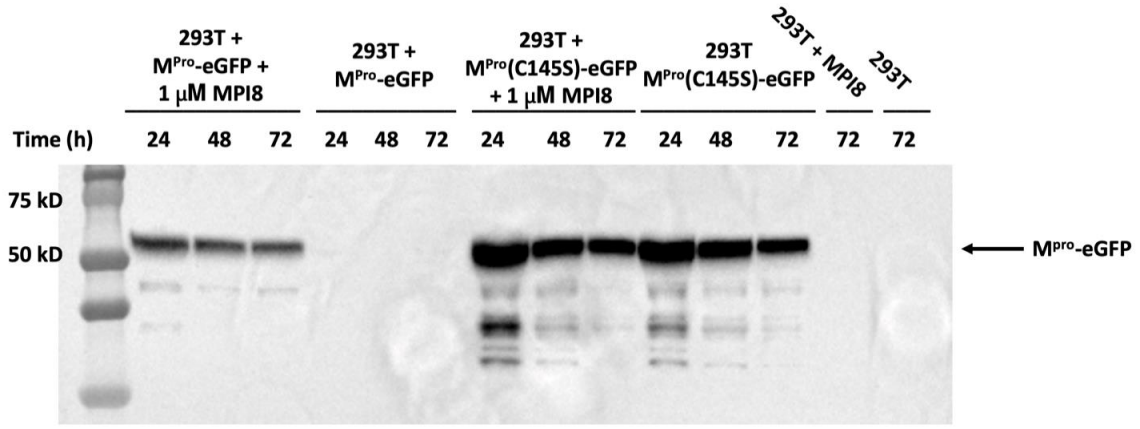

Figure 38 The cellular toxicity of M^{Pro} is from its protease activity. Cells were transfected with plasmids coding active M^{Pro}-eGFP or inactive M^{Pro}(C145S)-eGFP. Without 1 μM MPI8, the expression of M^{Pro}-eGFP led to cell death and no detectable M^{Pro}-eGFP. The addition of 1 μM MPI8 led to cell survival and detectable M^{Pro}-eGFP. However, in either presence or absence of 1 μM MPI8, cells expressing inactive M^{Pro}(C145S)-eGFP showed highly expressed M^{Pro}(C145S)-eGFP. The displayed gel was Western blotting by anti-M^{Pro}. Reprinted with permission from reference 104.

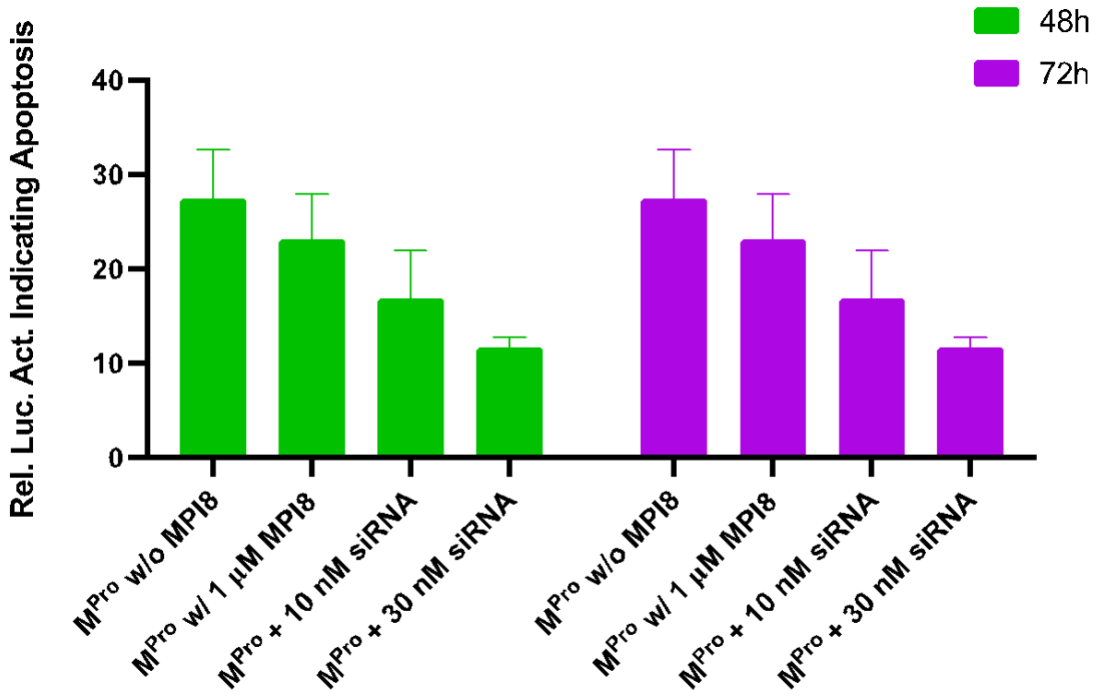

Figure 39 Cellular toxicity from M^{Pro} was inhibited by M^{Pro}-targeting siRNA. 293T cells were transiently transfected with PLVX-M^{Pro}-eGFP-2 and then incubated with or without MPI8 or M^{Pro}-targeting siRNA. SiRNA was transfected with lipofectamine 3000 (ThermoFisher L3000001), according to the protocol (ThermoFisher Document Part No. 100022234), at 24 and 48h after cells were plated respectively. After 48 and 72 h, cellular apoptosis indicating cell death was analyzed using the Promega RealTime-GloTM apoptosis assay kit. Reprinted with permission from reference 104.

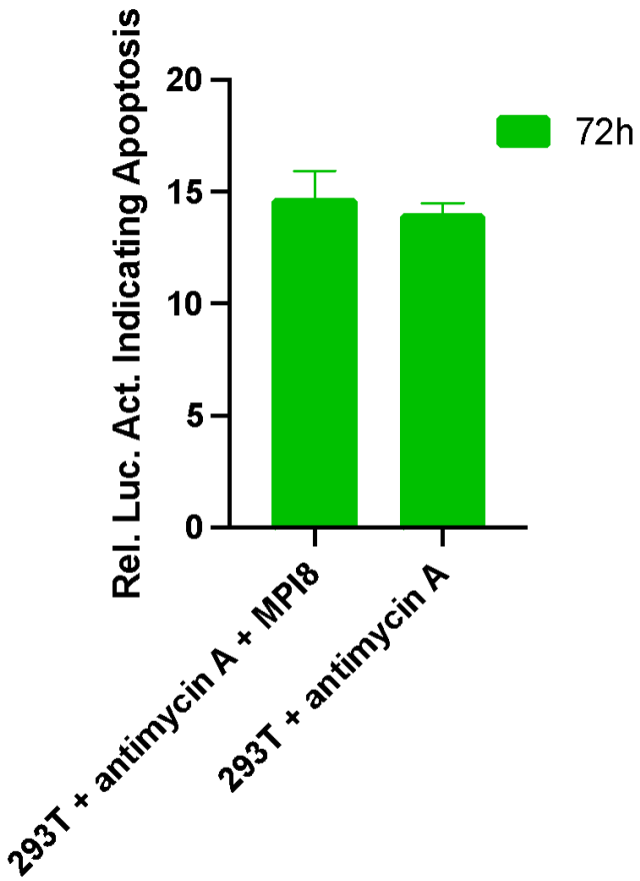

Figure 40 293T cell apoptosis induced by 1 µM antimycin A is not influenced by the addition of 1 µM MPI8. Reprinted with permission from reference 104.

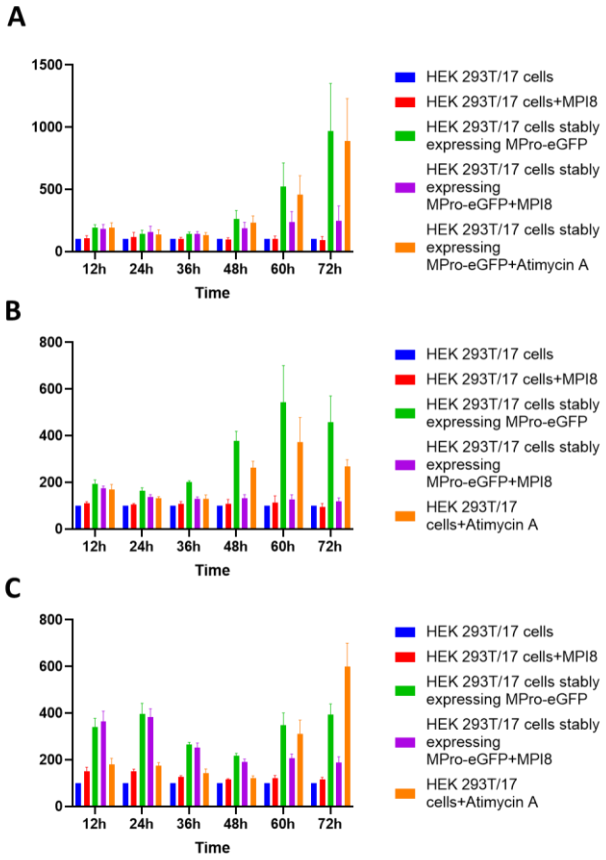

Figure 41 293T/17 cells that stably expressed M^{Pro}-eGFP and were established in the presence of MPI8 exhibited strong apoptosis when MPI8 was withdrawn from the growth media. A, B, and C are three repeats. The cell assay was performed with RealTime-Glo™ Annexin V Apoptosis and Necrosis Assay kit from Promega. HEK 293T/17 and constructed HEK 293T/17 cells stably expressing M^{Pro}-eGFP were used for this cell assay. The cells were maintained in high glucose DMEM medium supplemented with 10% FBS, plated with a cell density of 5×10^5 cells/mL. Five groups of experiments were set:
HEK 293T/17;
HEK 293T/17 + MPI8 (1 μM)
HEK 293T/17 cells stably expressing M^{Pro}-eGFP
HEK 293T/17 cells stably expressing M^{Pro}-eGFP + MPI8 (1 μM)
HEK 293T/17(b&c) or HEK 293T/17 cells stably expressing M^{Pro}-eGFP(a) + Antimycin A (1 μM)
Each experiment has 5 repeats.
The cell assay was performed as instructed by the protocol, luminescence was recorded at 12h, 24h, 36h, 48h, 60h, 72h after plating the cells. The luminescence readings were normalized using HEK 293T/17 as a negative control, which was set to a unit of 100. Reprinted with permission from reference 104.

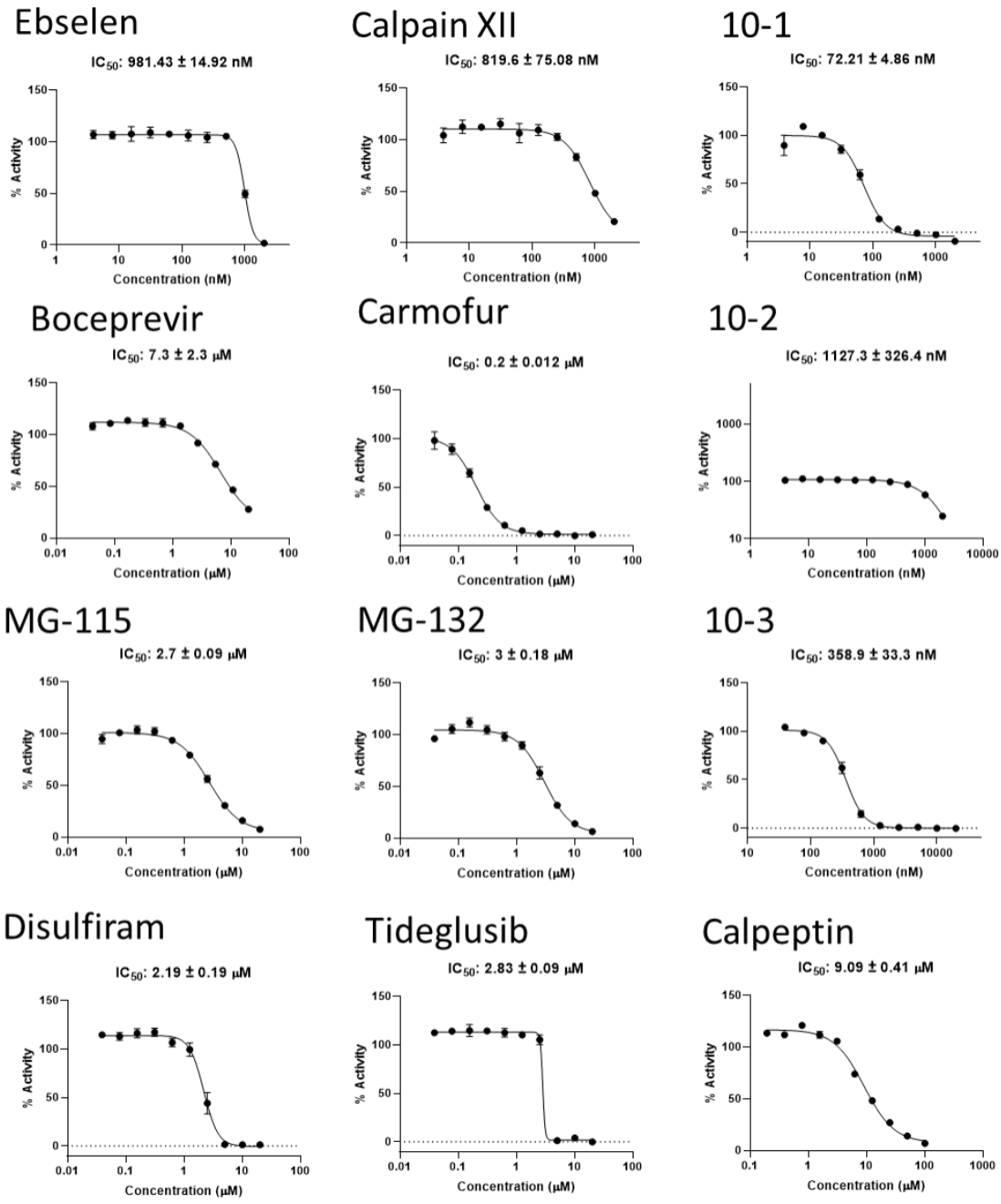

Figure 42 The recharacterization of M^{Pro} inhibition. Reprinted with permission from reference 104.

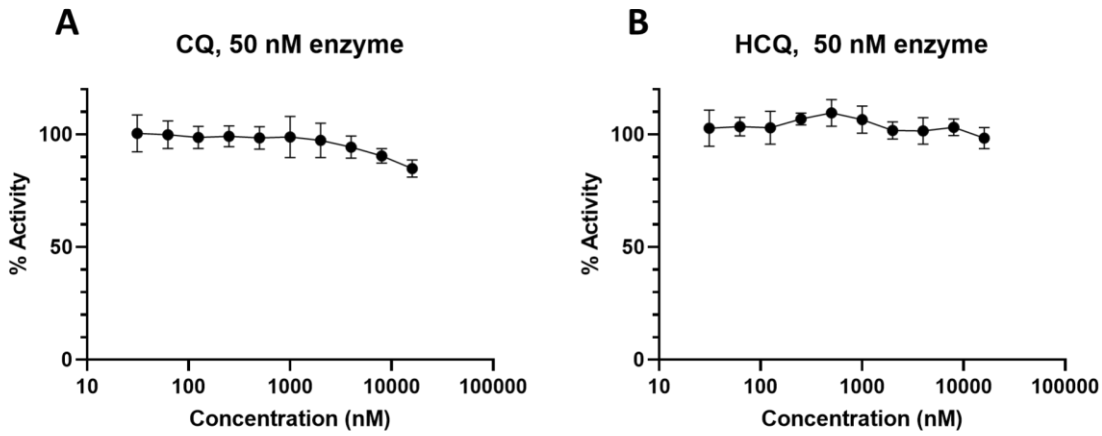

Figure 43 The recharacterization of M^{Pro} inhibition by (A) chloroquine and (B) hydroxychloroquine. Reprinted with permission from reference 104.

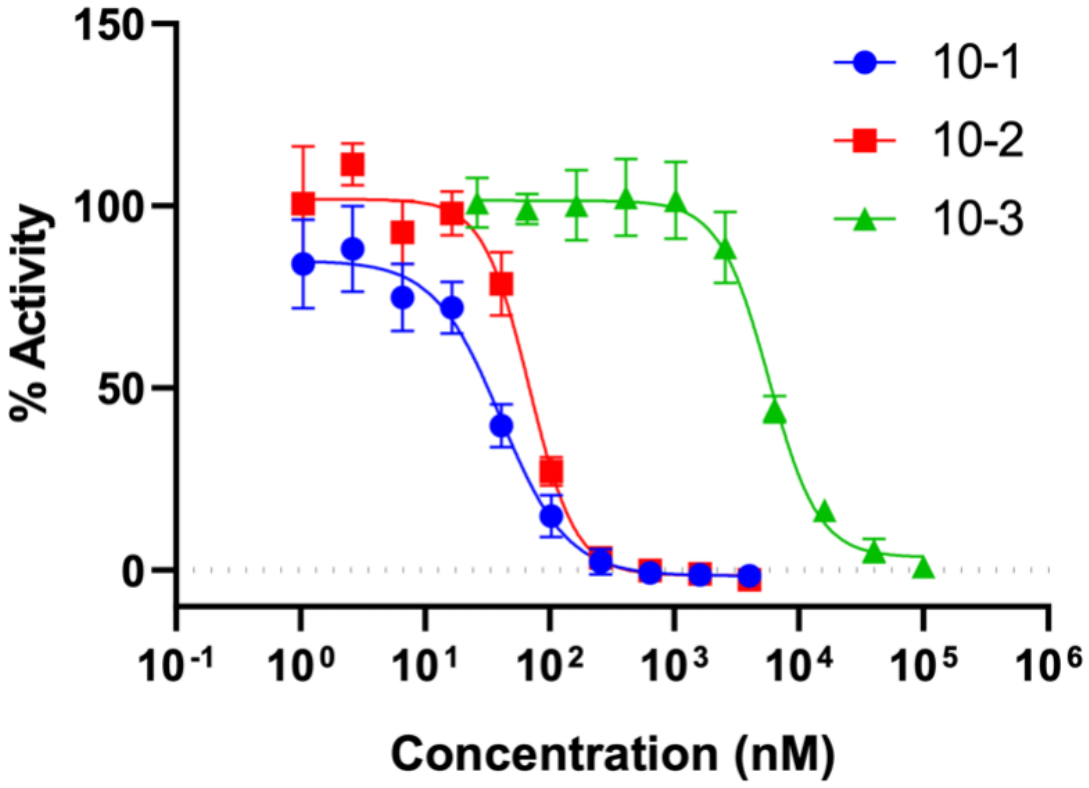

Figure 44 The kinetic characterization of 10-1, 10-2, and 10-3 in their inhibition of M^{Pro}. Reprinted with permission from reference 104.

Primer	**Primer**
FRET-Mpro-for	AGATCTCGAGTCAAAACAAGCGCGGTGC
FRET-Mpro-rev	TTCGAAGCTTGCTGAAAAGTTACGCCGGAAC
XbaI-Mpro-f	TAGTTCTAGAATGTCAGGGTTTCGCAAG
Mpro-HindIII-r	CCATAAGCTTGCCAAAAGTTACGCCGGAACAC
HinIII-eGFP-f	TGGCAAGCTTATGGTGAGCAAGGGC
eGFP-NotI-r	ATCCGCGGCCGCTTACTTGTACAGCTCGTCCATG
XbaI-Cut-Mpro-f	TAGTTCTAGAATGAAAACAAGCGCGGTGCTCCAGTCAG GGTTTCGCAAGATG
Mpro C145S-f	GAACTTCACAATCAAGGGATCGTTCCTGAATGGGAGTAG CGGTTCGGTTGGATTCAATAT
Mpro C145S-r	AAGAGACGCAGTCGTAGTCGATATTGAATCCAACCGAA CCGCTACTCCCATTCAGGAACG

Table 5 The primers and their sequences used in the construction of plasmids. Reprinted with permission from reference 104.

	Sequence
siRNA	UUUCCUUCAAGAUCGGUCCCG

Table 6 siRNA sequence used to knock down Mpro expression. Reprinted with permission from reference 104.