Investigating the Main Protease (MPro) of SARS-CoV-2 as a Potential Drug Target

Valerie Giovina Pascetta
University of New Hampshire

Follow this and additional works at: https://scholars.unh.edu/honors
Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation
Pascetta, Valerie Giovina, "Investigating the Main Protease (MPro) of SARS-CoV-2 as a Potential Drug Target" (2022). Honors Theses and Capstones. 621.
https://scholars.unh.edu/honors/621

This Senior Honors Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Honors Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact Scholarly.Communication@unh.edu.
Investigating the Main Protease (MPro) of SARS-CoV-2 as a Potential Drug Target

Valerie Pascetta

PI: Dr. Krisztina Varga

University of New Hampshire, Durham, NH

Biochemistry, Molecular, and Cellular Biology Honors Thesis

Topic: Spectroscopic Analysis of Viral Enzymatic Activity and Inhibition
# Table of Contents

Abstract ........................................................................................................ Page 3

Introduction .................................................................................................. Page 4-8

Methods ....................................................................................................... Page 8-11

Results/Discussion ....................................................................................... Page 11-17

Conclusions .................................................................................................. Page 17

Acknowledgements ....................................................................................... Page 18

References .................................................................................................... Page 19-21
Abstract

The rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 19 (COVID-19) pandemic has claimed the lives of roughly 6.2 million people worldwide as of May 2022. The virus’s main protease ($M^{\text{pro}}$) has been identified as an attractive drug target due to the critical role it plays in the viral life cycle. The roughly 34 kDa $M^{\text{pro}}$ cleaves functional viral polypeptides out of two long polyproteins at conserved cut sites, allowing them to fulfill their role in processes like transcription and replication. Here, we have studied the enzymatic activity and inhibition of $M^{\text{pro}}$ and tested known (carmofur, ebselen, and tideglusib) and novel (CCG-50014, and CCG-203769) inhibitors using fluorescence spectroscopy. We report both the catalytic efficiency of $M^{\text{pro}}$ ($K_{\text{cat}} / K_M = 27,900 \text{ M}^{-1} \text{s}^{-1}$) and the IC$_{50}$ values measured for tideglusib and ebselen ($1.39 \pm 0.2 \mu\text{M}$ and $0.40 \pm 0.05 \mu\text{M}$ respectively) to be comparable to values published by Jin et al. [13]. We also report that the novel CCG-50014 has comparable IC$_{50}$ value ($1.39 \pm 0.22 \mu\text{M}$) to known inhibitors of this enzyme.
This work has been accepted for publication:

Jacob Andrzejczyk, Katarina Jovic, Logan M. Brown, Valerie G. Pascetta, Krisztina Varga, Harish Vashisth. Molecular Interactions and Inhibition of the SARS-CoV-2 Main Protease by a Thiadiazolidinone Derivative. Accepted for publication. Proteins (2022)

My contributions to this work: My contributions included working with Katarina Jovic under the supervision of Dr. Krisztina Varga to express and purify recombinant M<sup>Pro</sup> and perform the inhibition assays with the five tested compounds. I also performed some preliminary activity assays with Katarina Jovic before Logan Brown collected the bulk of the data used to calculate catalytic efficiency. The computational studies were performed by Jacob Andrzejczyk under the supervision of Dr. Harish Vashisth.

Abbreviations: SARS, severe acute respiratory syndrome; MERS, Middle East respiratory syndrome; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; COVID-19, coronavirus disease 2019; ACE2, angiotensin-converting enzyme 2; nonstructural protein, nsp; ORF, open reading frame; main protease, M<sup>Pro</sup>; thiadiazolidine, TDZD; regulators of G-protein signaling, RGS; inhibitory concentration, IC<sub>50</sub>

Introduction

Coronaviruses are a large and diverse family of virus named for their unique crown-like spike proteins that cover the virion surface. These viruses are infectious to humans and animals and cause a variety of prevalent diseases with mild to severe symptoms. Coronavirus infections
typically manifest as moderate upper-respiratory tract illnesses [1]. However, the past two
decades have seen coronavirus outbreaks with increasingly serious and fatal symptoms including
severe acute respiratory syndrome (SARS) in 2002, Middle East respiratory syndrome (MERS)
in 2012, and most notably, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the
causative agent of coronavirus disease 2019 (COVID-19) [1, 2].

SARS-CoV-2 is a novel, β-coronavirus hypothesized to have zoonotic origins in bats [3]
as well as pangolins [4]. Symptoms range from mild; fever, cough, shortness of breath, to severe;
respiratory failure, pneumonia, and death. Outbreaks of the highly infectious COVID-19 virus
were first reported in December of 2019 in Wuhan, China. The virus spread rapidly and, in
March od 2020, a global pandemic was declared by the World Health Organization [3, 5]. The
ongoing pandemic has claimed the lives of almost 6.2 million people worldwide as of May 2022
[6] and has sparked an international research effort to understand the molecular mechanisms of
SARS-CoV-2 infection and develop effective treatments against its disease.

The virus is roughly 65-125 nm in diameter, and has a single stranded, positive sense,
RNA genome. The genome encodes four main structural proteins (S,E,M and N) essential to host
cell recognition and entry. The spike (S) protein is a transmembrane glycoprotein expressed on
the viral surface which interacts with host lower respiratory cell angiotensin-converting enzyme
2 (ACE2) receptors. The envelope (E) protein plays a role in viral maturation. The membrane
(M) protein determines viral envelope shape and assembly. The nucleocapsid (N) protein is an
RNA binding protein that is involved in viral transcription, translation, and replication. The
SARS-CoV-2 genome also encodes sixteen non-structural proteins (nsp 1-16) which have a
variety of functions including suppression of the host immune response, cleavage of
polyproteins, promotion of cytokine expression, and proofreading of the viral genome [3].
Once the S protein attaches to host cell ACE2 receptors, the membranes fuse and viral mRNA ready for translation enters the host cytoplasm. The genome contains fourteen open reading frames (ORFs) that encode both the structural and non-structural proteins. Among the first to be translated are the non-structural protein coding ORF1a and ORF1b which produce two long, overlapping polyproteins pp1a and pp1b. pp1a and pp1b are made up of the sixteen functional nsp polypeptides, including multiple proteins that aid in forming the polymerase and helicase complexes essential to viral transcription and replication. Enzymatic digestion of pp1a and pp1b is required to release the functional polypeptides from these larger polyproteins and allow them to participate in the viral life cycle. The nsp5 enzyme, also known as the main protease (MPro), is largely responsible for this digestion via its at least eleven conserved cut sites along the length of pp1a and pp1b, including its own autolytic cleavage [3].

MPro is a roughly 34 kDa, homodimeric enzyme comprised of 306 amino acids per monomer. Each monomer contains three subdomains, domains I and II (residues 8-101 and 102-184, respectively) are characterized by mainly β-barrel motifs, while domain III (residues 201-306) is primarily comprised of α-helices [7] (Fig. 1A). The MPro active site is situated at the interface of domains I and II and is comprised of key active site residues; M49, G143, H163, H164, E166, P168, Q189 and Y54, G143, H163 which form an oxyanion loop [8] (Fig. 1B). The MPro active site cleaves peptide bonds using a catalytic dyad (Fig. 1B) in which the -SH group of a cysteine residue (C145) donates a proton to the imidazole group of a histidine residue (H41) [9]. This allows for cleavage at the following recognition site sequences: L-Q ↓ (S, A, G) [10, 11].

Because of its pivotal role in the viral life cycle, high sequence conservation with other coronaviruses, and the lack of human homologues, MPro was identified as an attractive drug
target for the treatment of COVID-19. Although vaccine development has tremendously reduced the public health risks of COVID-19, the need for antiviral drugs to treat immunocompromised and un-vaccinated patients is still pressing, especially in the face of rapidly emerging variants. In December of 2021, the FDA announced its emergency authorization of Pfizer’s Paxlovid, an MPro inhibitor and the first antiviral pill authorized to treat COVID-19 [12]. While the development of Paxlovid again marks a major achievement in combating the global COVID-19 pandemic, more research into antiviral development is still necessary to increase drug efficacy and reduce potential adverse side effects.

In this work we have investigated five potential small molecule MPro inhibitors, three of which are known MPro inhibitors (carmofur, tideglusib, and ebselen), and two that are novel (CCG-50014, and CCG-203769). The established inhibitors were pulled from a library of known compounds with applications that were initially unrelated to SARS-CoV-2. They were then repurposed and published in previous studies as effective MPro inhibitors [13, 14]. This study tested the three known inhibitors with the intent of replicating the results previously reported by Jin et al. [13] to optimize inhibition assay conditions and directly compare results with structurally similar inhibitors. The two novel inhibitors were selected based on structural similarity to tideglusib, a thiaiazolidine (TDZD) analog [15]. TDZD derivatives are known to covalently modify the cysteine residues of the regulator of G-protein signaling (RGS) proteins [16, 17, 18]. Because cysteine is an essential part of the MPro active site, it was hypothesized that novel TDZD derived inhibitors would demonstrate comparable levels of MPro inhibition to tideglusib. To investigate this hypothesis, recombinant MPro was expressed, and purified to test the inhibitory effects of the five ligands using fluorescence spectroscopy. These studies were used to complement molecular dynamics simulations for MPro inhibition [19].
Figure 1: Structural details of M\textsuperscript{pro} monomer and active site. (A) M\textsuperscript{pro} monomer with domain I (orange), domain II (purple), and domain III (blue). Key catalytic residues H41 (cyan) and C145 (yellow) also pictured. (B) A zoomed view of the active site with key residues colored, labeled and shown in stick representations: M49 (red), G143 (pink), H163/H164 (gray), E166 (tan), P168 (blue), Q189 (purple), and the backbone atoms of residues that form the oxyanion loop (green). Figure was adapted from Padhi & Tripathi, 2021 [8].

Methods

M\textsuperscript{Pro} Expression and Purification

The expression vector, M\textsuperscript{Pro}-3C pET21b(+), containing the full-length SARS-CoV-2 gene encoding M\textsuperscript{Pro} (NC_045512) was purchased from GenScript, with the gene sequence optimized for expression in Escherichia coli. The final protein product contains four additional amino acids
(GPGG) before the C-terminal 6xHis-tag. Protein expression and purification procedures presented here were adopted and modified from Jin et. al [13]. The vector was transformed into *Escherichia coli* Rosetta-gami B(DE3) cells (EMD Millipore), which were grown for 5-6 hours in Luria-Bertani (LB) broth with 100 µg/mL ampicillin at 37 °C. Once the cells reached OD$_{600}$ 0.7, the over expression of M$^\text{Pro}$ was induced by adding 0.5 mM IPTG to the medium. After 11-12 hours of incubation at 16 °C, the liquid cell culture was centrifuged at 9,559 × $g$ for 25 minutes (4 °C). The obtained *Escherichia coli* pellets were then stored -80 °C.

Protein purification started by thawing out a pellet at 4 °C for 30 minutes. Lysis buffer containing 20 mM Tris-HCl (pH 7.3), 150 mM NaCl, Halt$^\text{TM}$ Protease Inhibitor cocktail (Thermo Scientific) and Benzonase (EMD Millipore) were then added to the pellet. After 15-20 minutes of incubation with Lysis buffer, the cells were fully re-suspended and subjected to four rounds of mechanical cell lysis at 1,500 psi using the G-M French Press (Glen Mills). The cell lysate was then centrifuged at 20,217 × $g$ for 30 minutes (4 °C). The obtained supernatant was filtered through 0.45- and 0.22-µm filters before purification. Since M$^\text{Pro}$ was expressed with a C-terminal 6xHis-tag, the protein was purified using fast protein liquid chromatography system (FPLC; GE Healthcare AKTA purifier 900) equipped with a HisTrap FF 1 mL column. The protein was eluted from the column using an imidazole concentration gradient going up to 500 mM. The selected protein fractions were pooled, and imidazole was removed using Zeba$^\text{TM}$ spin desalting columns (following manufacturer’s recommendation; Thermo Scientific). The sample was brought up to 5 mL prior to subjecting it to another round of Ni-affinity chromatography using a HisTrap HP 1 mL column. Upon collecting and pooling the selected fractions, the protein sample was exchanged into buffer containing 50 mM Tris-HCl and 1 mM EDTA (pH = 7.3) using Zeba$^\text{TM}$ spin desalting columns. The protein purity was assessed using SDS-PAGE. The
protein concentration was measured using JASCO V-650 UV-vis spectrophotometer. Upon determining the MPro purity and concentration, the protein identity was confirmed using LC-MS/MS mass spectrometry.

**Activity and Inhibition Assays**

Since it was previously shown that C-terminal 6His-tag does not alter MPro activity, the protease activity of uncleaved MPro containing the C-terminal 6His-tag was measured [20]. Both the activity and inhibition assays were performed following a modified protocol originally established by Jin et al. [13, 19] Fluorescence resonance energy transfer (FRET) peptide Mca-AVLQSGFRK(Dnp)K (GLBiochem) was used as the MPro substrate (excitation wavelength 325 nm, emission wavelength 392 nm). JASCO FP-8300 spectrofluorometer was used to measure the MPro activity. 1 µM MPro (final concentration) was mixed with varying concentrations of the peptide substrate (1.87 - 40 µM). Initial rates were obtained by fitting the linear portion of the curves to a straight line (typically around 35 - 50 seconds) (Kinetic Analysis, JASCO). K_M and V_max were calculated from the Lineweaver-Burk plot (1/V vs. 1/[S]) using Kinetic Analysis, JASCO. These values were used to determine the K_cat (V_max/[MPro]) as well as the catalytic efficiency (K_cat/K_M).

The degree of inhibition was measured for the three known MPro inhibitors [13, 14] carmofur (Fig. 3), tideglusib (Fig. 4) and ebselen (Fig. 5); and two novel ones, CCG-2203769 (Fig. 6) and CCG-50014 (Fig. 7). The inhibition assays were performed on a microplate reader SpectraMax Me (Molecular Devices). Similarly to the activity assays, fluorescence released the Mca-AVLQSGFRK(Dnp)K peptide cleavage by MPro was measured. Each reaction contained 0.2 µM MPro, 20 µM FRET peptide, varied inhibitor concentration (0.025 - 100 µM), and 5% DMSO in Assay buffer (50 mM Tris-HCl 1 mM EDTA pH 7.3). Data points were recorded every 4
seconds for 5 minutes, at 30 °C. SoftMax Pro 7 (Molecular Devices) was used to calculate the initial rates in units of relative fluorescence per second (RFU/s). The GraphPad Prism software was used to calculate the IC₅₀ values and plot the inhibition curves.

**Results/Discussion**

**Activity Assay**

The activity assay was performed before testing inhibition to verify that the catalytic efficiency of Mₚro was comparable to the previously reported value by Jin et al. (Kₐₚ / Kₘ = 28,500 M⁻¹ s⁻¹) [13]. The Michaelis-Menten plot of [substrate] vs. RFU/min (Fig. 2) showed a Michaelis constant (Kₘ) of 7.88 µM ± 2.62 µM and a maximum rate (Vₘₚ) of 133.2 RFU/min. These values were then used to calculate the turnover number (Kₐₚ = 2.10 s⁻¹ ± 0.61 s⁻¹) and the catalytic efficiency (Kₐₚ / Kₘ = 27,900 M⁻¹ s⁻¹ ± 6,250 M⁻¹ s⁻¹) which was comparable to the value published by Jin et al. [13]. This confirmed that the IC₅₀ values measured in the later inhibition assays could be objectively compared to the previously reported values by Jin et al. [13].
Figure 2: Michaelis-Menten plot showing hyperbolic $M_{Pro}$ saturation. Plot shows [substrate] (uM) against reaction rate (RFU/min). Also shown is the $K_M$ value with the error and $V_{max}$ value.
Inhibition Assays

The inhibition data was analyzed using plots of normalized inhibition versus log [I]. Each inhibition plot exhibited a classic sigmoidal shape used to calculate the IC$_{50}$ values. The IC$_{50}$ values of the three known inhibitors, carmofur (Fig. 3), tideglusib (Fig. 4), and ebselen (Fig. 5) were calculated first for comparison with the values of Jin et al. [13]. The IC$_{50}$ values of tideglusib (1.39 ± 0.2 µM) and ebselen (0.40 ± 0.05 µM) were comparable to reported values [13], while the IC$_{50}$ value for carmofur (4.45 ± 0.52 µM) was approximately 2.5-fold higher, indicating that it may be a slightly weaker inhibitor of M$^{pro}$ than previously reported. As our measured IC$_{50}$ values were comparable to those previously reported [13], it was assumed that the inhibition assay could be used to reliably assess the potential of the two novel inhibitors and the results could be objectively compared to the IC$_{50}$ values published by Jin et al.

Figure 3: Normalized inhibition of M$^{pro}$ by carmofur (left) and carmofur chemical structure (right). Plot shows the normalized inhibition percentage of M$^{pro}$ (y-axis) against the log of the carmofur concentration (nM, x-axis). Also shown is the IC$_{50}$ value with the error, and the proportion of variance value.
Figure 4: Normalized inhibition of M<sub>Pro</sub> by tideglusib (left) and tideglusib chemical structure (right). Plot shows the normalized inhibition percentage of M<sub>Pro</sub> (y-axis) against the log of the tideglusib concentration (nM, x-axis). Also shown is the IC<sub>50</sub> value with the error, and the proportion of variance value.

Figure 5: Normalized inhibition of M<sub>Pro</sub> by ebselen (left) and ebselen chemical structure (right). Plot shows the normalized inhibition percentage of M<sub>Pro</sub> (y-axis) against the log of the ebselen concentration (nM, x-axis). Also shown is the IC<sub>50</sub> value with the error, and the proportion of variance value.
The IC$_{50}$ values of the two novel inhibitors; CCG-50014 and CCG-203769 were then calculated using the same normalized inhibition vs log [I] plots. The IC$_{50}$ value of CCG-50014 was 1.39 ± 0.22 µM (Fig. 6); thus CCG-50014 exhibited inhibition comparable to that tideglusib, as hypothesized. These results are promising for CCG-50014’s potential application as an M$^{\text{Pro}}$ inhibitor. The IC$_{50}$ value calculated for CCG-203769 was 10.95 ± 1.74 µM (Fig. 7). Because this value was higher than CCG-50014 and all the known inhibitors, the compound showed less potential as an M$^{\text{Pro}}$ inhibitor.

![Inhibition vs log [I] plot](image)

**Figure 6:** Normalized inhibition of M$^{\text{Pro}}$ by CCG-50014 (left) and CCG-50014 chemical structure (right). Plot shows the normalized inhibition percentage of M$^{\text{Pro}}$ (y-axis) against the log of the CCG-50014 concentration (nM, x-axis). Also shown is the IC$_{50}$ value with the error, and the proportion of variance value.
Figure 7: Normalized inhibition of $\text{M}^{\text{Pro}}$ by CCG-203769 (left) and CCG-203769 chemical structure (right). Plot shows the normalized inhibition percentage of $\text{M}^{\text{Pro}}$ (y-axis) against the log of the CCG-203769 concentration (nM, x-axis). Also shown is the IC$_{50}$ value with the error, and the proportion of variance value.

The difference in inhibition between CCG-50014 and CCG-203769 was attributed to the different moieties attached to each structure’s thiazole ring. Where CCG-203769 has aliphatic moieties attached to its central thiazole ring (an ethyl and a butyl group Fig. 7), CCG-50015 has aromatic moieties attached (a methylbenzene and a fluorophenyl group Fig. 6). The molecular dynamics studies (by the Vashisth group) complementary to this work showed these aromatic moieties to be a stabilizing feature of CCG-50014’s interaction with the $\text{M}^{\text{Pro}}$ active site. The active site residues located closest to the fluorophenyl side chain were shown to be stabilized in the presence of covalently-bound CCG-50014 through several stacking interactions and hydrogen bonds. While the active site residues that showed increased fluctuations in the presence
of CCG-5001 were those residues that did not have direct interactions with its aromatic side-chains [19].

Conclusions

In this work we have investigated inhibitory the potential of three known M\textsuperscript{Pro} inhibitors: carmofur, tideglusib, and ebselen, and two novel M\textsuperscript{Pro} inhibitors: CCG-50014 and CCG-203769 using fluorescence spectroscopy assays. Both the catalytic efficiency of M\textsuperscript{Pro} (K\text{cat} / K_M = 27,900 M^{-1} s^{-1} \pm 6,250 M^{-1} s^{-1}) and the IC\textsubscript{50} values measured for tideglusib and ebselen (1.39 ± 0.2 µM and 0.40 ± 0.05 µM respectively) are comparable to the values reported by Jin et al. [13]. The IC\textsubscript{50} value measured for carmofur (4.45 ± 0.52 µM) indicated it is a less potent M\textsuperscript{Pro} inhibitor than previously reported [13]. The IC\textsubscript{50} value measured for CCG-50014 (1.39 ± 0.22 µM) was comparable to those of tideglusib, while the value for CCG-203769 (10.95 ± 1.74 µM) signified it is a less potent inhibitor of M\textsuperscript{Pro}.

Molecular dynamics simulations revealed this difference to be because of the moieties attached to each structure’s central thiazole ring. While CCG-50014 has aromatic moieties attached, CCG-203769 has aliphatic moieties. The simulations showed these aromatic moieties to have several stabilizing interactions with the key residues of the active site.
Acknowledgements

This work was supported by the National Institutes of Health (NIH) under grant R35GM138217 (H. V.) and by the Collaborative Research Excellence (CoRE) program at the University of New Hampshire (UNH). Additional support was also provided by the UNH Center of Integrated Biomedical and Bioengineering Research (CIBBR) core facilities through a grant from NIH (P20GM113131). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Support for the computational studies was provided through the following resources: Premise, a central shared high-performance computing cluster at UNH supported by the Research Computing Center and Biomade, a heterogeneous CPU/GPU cluster supported by the NSF EPSCoR award (OIA-1757371). This work also acknowledges and thanks Dr. Richard R. Neubig, Professor of Pharmacology and Toxicology at Michigan State University, for providing the inhibitor CCG-50014 and CCG-203769, and Dr. Douglas E. Frantz, Professor of Chemistry at University of Texas at San Antonio, for providing the three inhibitors tideglusib, ebselen, and carmofur.

I would also like to thank the members of the Varga group at the University of New Hampshire for their support throughout this research, and special thanks to Katarina Jovic for her excellent mentorship, leadership, and hard work throughout this project.
References


