

**Computational Analysis of
Synergistic Inhibitors of the Papain-Like Protease**

BY

SHUYI CAO
B.S., Sichuan University, 2010

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Medicinal Chemistry
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Michael E. Johnson, Chair and Advisor
Pavel A. Petukhov
Brian T. Murphy

This thesis is dedicated to my parents, Xingyuan Cao and Li Wang, without whom it would never have been accomplished.

ACKNOWLEDGEMENT

I would like to thank my thesis committee -- Michael E. Johnson, Pavel A. Petukhov and Brian T. Murphy -- for their unwavering support and assistance. They provided guidance in all areas that helped me accomplish my research goals and enjoy myself in the process. I would also like to acknowledge Dr. Kirk E. Hevener, who offered kind help to the development of the proposal and made important contributions to the research.

The experimental section was done by Dr. Hyun Lee and Lena Truong and I would like to thank for them as well. I am also grateful about the help I got from other lab members from Dr. Michael E. Johnson.

SC

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
1. INTRODUCTION.....	1
1.1 Background: Papain-like cysteine protease of SARS Coronavirus.....	1
1.1.1 The spread and current treatment of SARS	1
1.1.2 The causative agent SARS-CoV	1
1.1.3 SARS Coronavirus replication cycle.....	2
1.1.4 The papain-like protease (PLpro)	4
1.1.5 MERS-CoV.....	7
1.2 Purpose and Significance of this Study	9
1.3 Introduction to Previous Work.....	9
1.3.1 Lead inhibitors.....	9
1.3.2 Fragments with synergistic effects.....	12
1.3.3 Inhibition mechanism of fragments with synergistic activity	12
2. HOT SPOT ANALYSIS OF LIGAND-BOUND PLPRO CRYSTAL STRUCTURE	16
2.1 Introduction and Purpose of the Study.....	16
2.2 Materials and Methods	17
2.2.1 Molecular Simulation of the Ligand-bound PLpro Crystal Structure	17
2.2.2 FTMAP	18
2.3 Results and Discussion.....	18
3. FREE BINDING ENERGY CALCULATION TO DETERMINE POSSIBLE BINDING SITES OF THE FRAGMENTS THAT GREATLY ENHANCED ENZYMATIC INHIBITION OF LEAD INHIBITORS	22
3.1 Introduction and Purpose of the Study.....	22
3.2 Materials and Methods	23
3.2.1 Starting conformations generated by Induced Fit Docking (IFD).....	23
3.2.2 Molecular Dynamic Simulation	24
3.2.3 MM/PBSA	24
3.3 Results and Discussion.....	25
4. DESIGN NEW COMPOUNDS BASED ON THE PREVIOUS STUDY BY FRAGMENT LINKING OR FRAGMENT MERGING METHOD	32
4.1 Introduction and Purpose of the Study.....	32
4.2 Materials and Methods	32
4.2.1 Virtual compound generation by fragment merging	32
4.2.2 Docking.....	35

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
4.3 Results and Discussion.....	35
5. CONCLUSION.....	40

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
I . IC ₅₀ VALUE COMPARISON OF FOUR LEAD INHIBITORS IN COMBINATION WITH FRAGMENT COMPOUNDS FROM ZENOBIA LIBRARY. (EXPERIMENTAL DATA FROM DR. H. LEE).....	14
II . IC ₅₀ VALUE COMPARISONS OF FOUR LEAD INHIBITORS IN COMBINATION WITH FRAGMENT COMPOUNDS F2124-0890 FROM LIFE CHEMICALS LIBRARY. (EXPERIMENTAL DATA FROM DR. H. LEE) ..	15
III. MECHANISM OF EXISTING PLRO INHIBITORS AND SEVERAL FRAGMENTS IDENTIFIED IN ZENOBIA LIBRARY. (EXPERIMENTAL DATA FROM DR. H. LEE)	15
IV. CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0617S WHEN FRAGMENT ZT0834 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.	28
V. CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0667S WHEN FRAGMENT ZT0834 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.	28
VI. CALCULATED BINDING FREE ENERGIES OF INHIBITOR GRL-0617S WHEN FRAGMENT F2124-0890 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.	30
VII. CALCULATED BINDING FREE ENERGIES OF INHIBITOR GRL-0667S WHEN FRAGMENT F2124-0890 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.	30
VIII. THE STRUCTURES OF PROPOSED COMPOUNDS OF MERGING GRL-0617S WITH ZT0834.....	33
IX. THE STRUCTURES OF PROPOSED COMPOUNDS OF MERGING GRL-0667S WITH ZT0834.....	34
X. COMPOUNDS PROPOSED FOR SYNTHESIS AND TESTING BY FRAGMENT-MERGING AND DOCKING CALCULATIONS.....	36

LIST OF FIGURES

<u>FIGURE</u>	<u>PAGE</u>
1. A brief description of the four genera in <i>Nidavirales</i> order	2
2. SARS coronavirus replication cycle.....	3
3. Schematic of the SARS-CoV polyprotein with cleavage sites of the two proteases, 3CLpro (pink) and PLpro (green).	4
4. Cleavage sites of SARS-CoV PLpro. Conserved amino acids are shown in green	5
5. The crystal structure of apo PLpro (PDB id: 2fe8).	5
6. Four lead inhibitors of PLpro in previous studies.....	6
7. The sequence alignment of SARS-CoV PLpro and MERS-CoV PLpro.	8
8. The overlay structures of SARS-CoV PLpro (blue) and MERS-CoV PLpro (tan).	8
9. X-ray structure of inhibitor GRL-0617S (PDB id: 3e9s) and GRL-0667S (PDB id: 3mj5) bounded PLpro.	11
10. The molecular dynamic simulation results of the ligand-bound PLpro crystal structures.	19
11. Fragment binding site analysis results by FTMAP. Two potential binding sites identified by FTMAP using the co-crystal structure of PLpro with inhibitor GRL-0667S are shown	21
12. Potential binding site of the fragment compounds	26
13. Most probable binding conformation of fragment ZT0834 (orange) in combination with lead inhibitors GRL-0617S (magenta)(a) and GRL-0667S (magenta)(b)	29
14. The most probable binding conformation of fragment F2124-0890 (orange) in combination with lead inhibitors GRL-0617S (magenta) (a) and GRL-0667S (magenta) (b)	31
15. Proposed compound for synthesis and testing by fragment-merging strategy based on the proposed conformations of GRL-0617S and ZT0834 from MM/PBSA calculations.....	37

LIST OF ABBREVIATIONS

SARS	Severe Acute Respiratory Syndrome
HTS	High Throughput Screening
SAR	Structure Activity Relationship
SARS-CoV	Severe Acute Respiratory Syndrome Conronavirus
PLpro	Papain-Like Protease
MERS	Middle East Respiratory Syndrome
nsps	non-structural proteins
3CLpro	3 Chymotrypsin-Like protease
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
MM/PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
IFD	Induced Fit Docking
SPR	Surface plasmon resonance
NMR	Nuclear magnetic resonance

SUMMARY

Severe Acute Respiratory Syndrome (SARS) is a contagious viral disease that has a high mortality rate. The Papain-Like protease (PLpro) is an essential enzyme involved in viral replication, which makes it an attractive drug target. In previous studies, several inhibitors, comprising two different scaffolds, were discovered by High Throughput Screening (HTS) followed by Structure Activity Relationship (SAR) based lead optimization. In this study, in order to further optimize the compounds, a fragment screening was conducted in the presence of the previously discovered inhibitors and several compounds with synergistic activity in combination with the lead inhibitors were identified. Enzymatic profiling and computational techniques were applied to search the potential binding site for these fragments, which provided the foundation for proposing the next generation of compounds with novel structures for further synthesis and testing.

Chapter 1

1. INTRODUCTION

1.1 Background: Papain-like cysteine protease of SARS Coronavirus

1.1.1 The spread and current treatment of SARS

An atypical pneumonia-like disease was discovered in Southern China, Guangdong Province in November 2002 (1, 2). The disease quickly spread to many countries in the following several months, including Vietnam, Singapore, Canada, and United States (3). It was named Severe Acute Respiratory Syndrome (SARS) and was identified to be caused by SARS Coronavirus, which is a novel enveloped, single-stranded–RNA virus (4). There were more than 8,400 cases and 789 deaths documented by the World Health Organization by 11 June 2003, with an estimated mortality rate of 10% (5). The disease was quickly controlled after its outbreak, but there is no effective treatment reported so far. Moreover, even though the re-emergence of SARS did not cause any fatality or secondary transmission, it is still a threat to the public health due to potential re-emergence of a closely related virus.

The current treatment of SARS mainly used antipyretics plus the supplemental oxygen. In some serious cases, the immunoglobulin from people who have already recovered from SARS has been given as a treatment (4). However, there is no strong evidence that these treatments work well (6).

1.1.2 The causative agent SARS-CoV

Coronaviruses are large RNA viruses and the regulatory mechanisms that are involved to express their genomes are quite complex (5). Its genome is about 30kb in size, and has three distinct functions. The replicase proteins were encoded by the N-terminal two-thirds of the genome, and the C-terminal one-third of the genome encodes the structural proteins, including the spike, envelope (E), membrane (M), and nucleocapsid (N) proteins and accessory Open Reading Frames (ORFs) (7). In the order *Nidovirales*,

there are four genera, including *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* (**Figure 1**). The viruses in *Alphacoronavirus* genus are associated with human transmitted coronavirus and the viruses in *Gammacoronavirus* and *Deltacoronavirus* are mostly associated with animal transmitted coronavirus. The SARS coronavirus caused a unique respiratory disease which showed both upper and lower respiratory tract infections and also gastroenteritis (8). It belongs to the B lineage of *Betacoronavirus*.

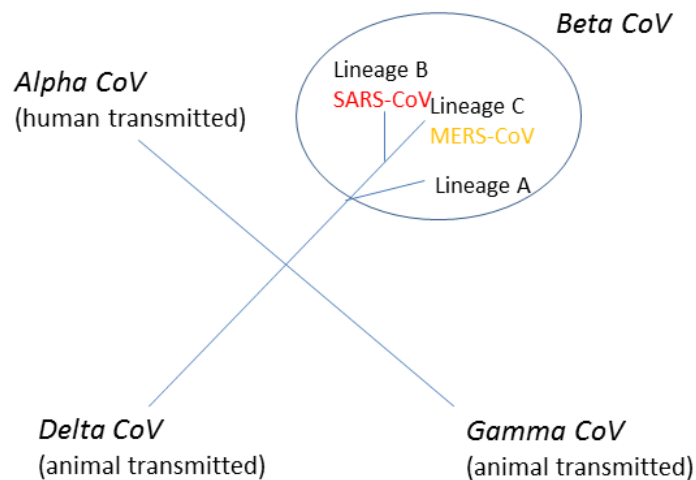


Figure 1. A brief description of the four genera in *Nidavirales* order.

1.1.3 SARS Coronavirus replication cycle

The replication cycle begins when the spike of the SARS coronavirus attaches to the receptor of host cell (**Figure 2**). The cellular target of SARS-CoV is angiotensin-converting enzyme 2 (ACE2), which exists mostly on the epithelial cells (9). ACE-2 is highly expressed in the enterocytes of the small intestine and lung, which explains

important aspects of SARS-CoV tropism (9). Then, the virus is drawn into the cell as the plasma protein infuses it. After being infected by SARS-CoV, the SARS-CoV genomic RNA is released into the cytoplasm of the cell and translated into two long and overlapping polyproteins, pp1a and pp1ab(10). The 3 Chymotrypsin-Like protease (3CLpro) and the papain-like protease (PLpro) are mostly responsible for processing of pp1a/1ab (11) (**Figure 3**), which releases 16 non-structural proteins (nsps). Those non-structural proteins will assemble into complexes and initiate RNA synthesis (11). PLpro generates the multifunctional membrane-associated replicase complex and mediates the viral replication of the SARS-CoV. The structural proteins are also created, including the Spike Glycoprotein S, the Small Envelope Protein E, the Membrane Glycoprotein M and the Nucleocapsid Protein N, which will help the budding of virions and release new virus (12).

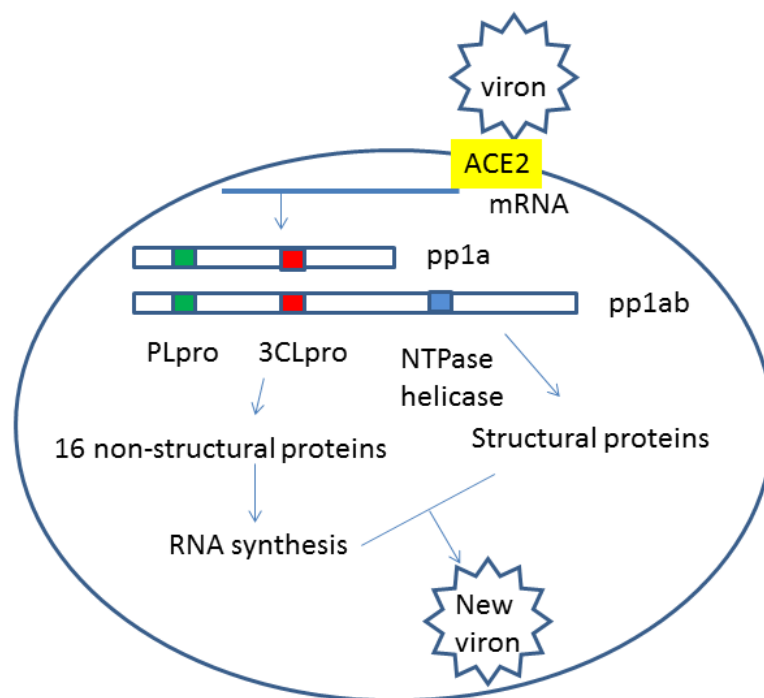


Figure 2. SARS coronavirus replication cycle.

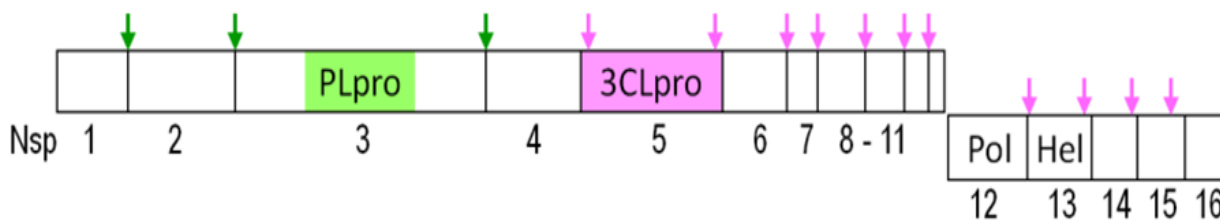


Figure 3. Schematic of the SARS-CoV polyprotein with cleavage sites of the two proteases, 3CLpro (pink) and PLpro (green).

1.1.4 The papain-like protease (PLpro)

PLpro is a cysteine protease which cleaves the polypeptide pp1a at three sites in N-proximal regions and is essential for viral replication. In SARS-CoV, the cleavage of polyprotein pp1a/1ab at three sites by PLpro will catalyze its own release as well as that of other non-structural proteins (nsps), nsp1, nsp2 and nsp3 (13) (**Figure 4**). It locates in a domain within nsp3 of pp1a/pp1ab.

The key features of the PLpro crystal structure were characterized as an intact zinc-binding motif, a Cys-His-Asp catalytic triad and an N-terminal ubiquitin-like domain (11) (**Figure 5**). SARS-CoV PLpro has deubiquitinating activity which plays an important role in many key cellular processes. It is highly likely that the deubiquitination by PLpro leads to a favorable environment for the viruses' replication by mimicking the function of cellular protein. In most cases, the ubiquitin acts as a target for degradation by the proteasome. Therefore, the deubiquitinating properties of PLpro may protect polyproteins of SARS-CoV from degradation, which increases the stability of the host cell and viral proteins (14). The critical function of PLpro in viral replication suggests that PLpro is an attractive target for antiviral drug development.

In prior studies of the Johnson/Ghosh groups, several non-covalent inhibitors of PLpro have been identified through HTS, structure-based drug design and further synthetic optimization [13-15], which can be classified into two scaffolds with different binding modes (**Figure 6**). The two scaffolds both have

SARS-CoV PLpro

		P6	P5	P4	P3	P2	P1	P1+	P2+		
nps 1	R	E	L	N	G	G	A	V	nps 2
nps 2	F	R	L	K	G	G	A	P	nps 3
nps 3	I	S	L	K	G	G	K	I	nps 4

↓

Figure 4. Cleavage sites of SARS-CoV PLpro. Conserved amino acids are shown in green.

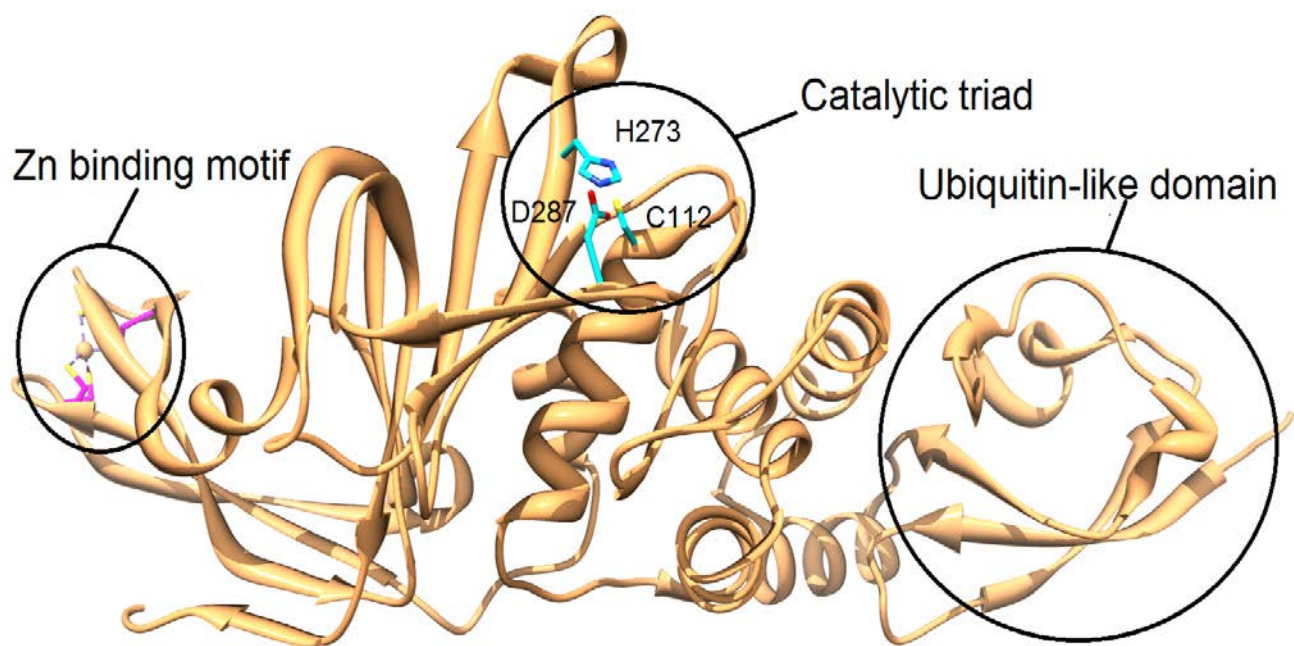


Figure 5. The crystal structure of apo PLpro (PDB id: 2fe8). The key features of PLpro crystal structure were characterized as an intact zinc-binding motif, a Cys-His-Asp catalytic triad and an N-terminal ubiquitin-like domain. The catalytic triad is composed of His273, Asp287 and Cys112. The metal ion zinc is coordinated by four Cysteine residues. The N-terminal domain of PLpro structure is a ubiquitin-like structure.

the naphthyl group, a chiral center and benzyl amide group. From the apo protein crystal structure along with the two protein-ligand complex X-ray structures (one with each class of inhibitor, PDB id: 3e9s and PDB id: 3mj5), there are some unoccupied regions in the crystal structure besides the original inhibitors' binding sites.

The inhibition mechanism is blocking the entrance of the active site, thus preventing the substrate from entering into the catalytic triad.

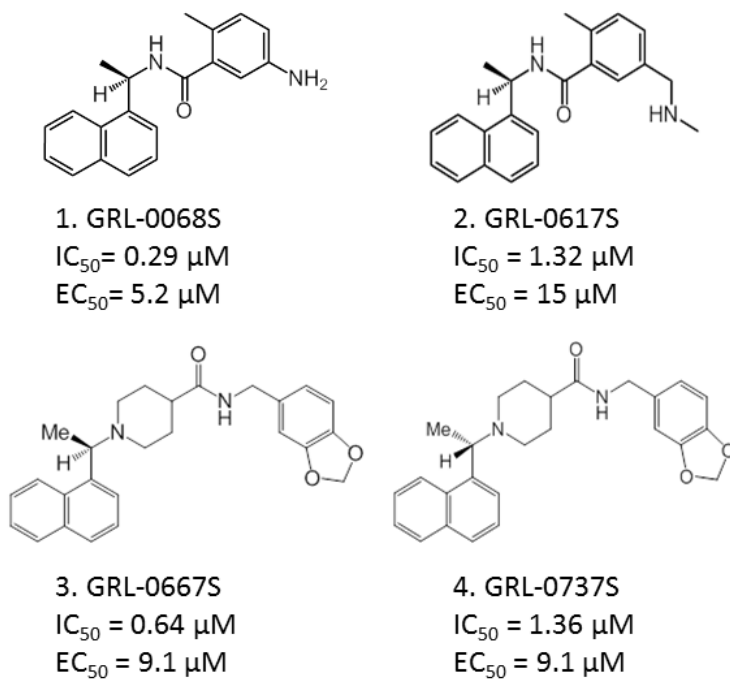


Figure 6. Four lead inhibitors of PLpro in previous studies. The binding modes of GRL-0617S and GRL-0667S are slightly different, which are shown in X-ray crystal structures (PDB id: 3e9s and PDB id: 3mj5).

1.1.5 MERS-CoV

In 2012, a novel human coronavirus was identified from the sputum of a man from Saudi Arabia(15). One particular symptom is acute pneumonia and subsequent multiple organ failure. It is suspected that this virus was transmitted from a bat host reservoir (15). This newly discovered virus was named Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and belongs to the same *Betacoronavirus* family as SARS-CoV. It is also the first human coronavirus in lineage C of the genus *Betacoronavirus*, along with its close relative bat coronaviruses HKU4 and HKU5 (15, 16) (**Figure 1**). It is the sixth coronavirus known to infect humans, following HCoV-229E, -OC43, -NL63, -HKU1 and SARS-CoV.

The genome of MERS-CoV has been sequenced, and also has one PLpro and 3CLpro, similar to SARS (16). From the sequence alignment, the sequence identity and similarity between SARS-PLpro and MERS-CoV PLpro are 32% and 51%, respectively (**Figure 7**). The catalytic triad residues cysteine, histidine and aspartate were conserved between the two sequences. A homology model was made using SARS-CoV PLpro as a template (**Figure 8**), whose RMSD was lower than 0.1 Å with the respect to the template, which means the homology model is quite similar with the template. From the homology model, we can see that the key features of PLpro are conserved, including the catalytic triad, zinc-binding motif and N- terminal ubiquitin-like domain.

Although MERS-CoV shares similar features with SARS-CoV, it also has some unique characteristics. Renal failure is documented as a unique aspect of the MERS-CoV infection, which may provide basis for further insights of this newly discovered disease. The molecular receptor of MERS-CoV is dipeptidyl peptidase 4 (DPP4; also known as CD26), which is conserved among different species such as bats and humans. That may explain the transmission from bats to human beings.



Figure 7. The sequence alignment of SARS-CoV PLpro and MERS-CoV PLpro. The conserved catalytic triad residues were marked with red box.

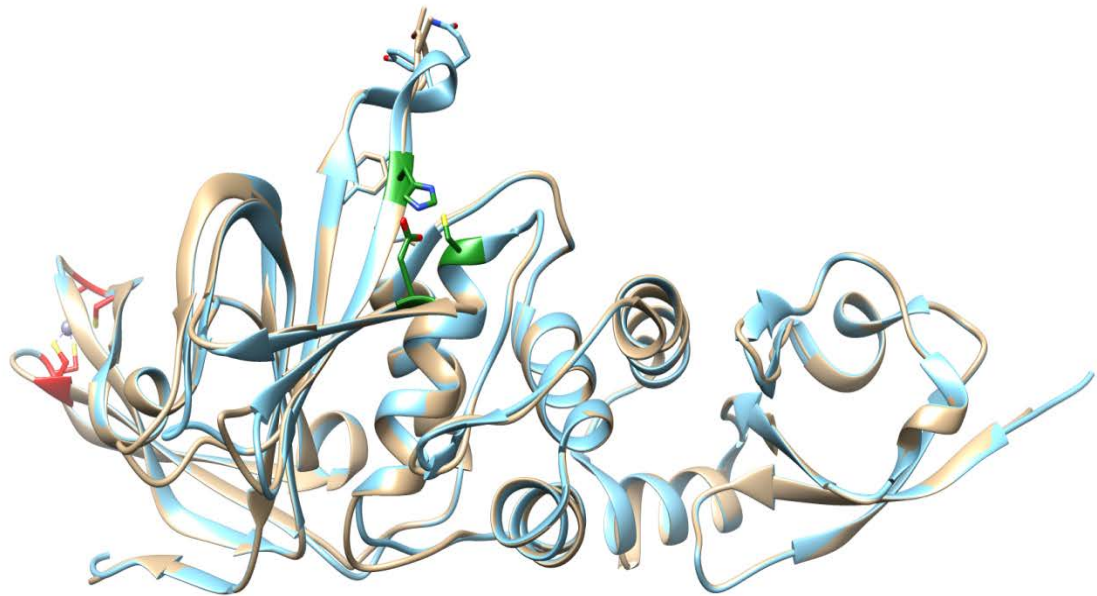


Figure 8. The overlay structures of SARS-CoV PLpro (blue) and MERS-CoV PLpro (tan). The catalytic triad residues are colored in green. The zinc binding motif is colored in red.

1.2 Purpose and Significance of this Study

PLpro plays an essential role in the replication of the SARS Coronavirus. By inhibiting the protease, the cleavage of polyproteins for final assembly of new virions can be stopped. Several inhibitors were found in previous studies, which can mainly be classified into two scaffolds (**Figure 6**). As these four inhibitors show moderate antiviral activity against SARS-CoV, their potencies must be further improved in order to be developed into clinical candidates. The purpose of this study is to search for compounds with better activity and novel scaffolds based on the current lead inhibitors. In this study, an in-house fragment screening was conducted and fragments that showed synergistic effect with the lead inhibitors were identified. These results led us to think about growing the compounds into the nearby unoccupied regions to increase their activity. In order to investigate the most probable binding site for those fragments, we conducted solvent mapping on the PLpro protein surface for so-called “hot spots” to identify potential binding sites. To further investigate the binding sites and binding conformations of the fragment in the presence of lead inhibitors, we started by docking to identify possible binding sites of the fragments, followed by binding free energy calculation to suggest the most probable binding site. Identifying the probable binding sites and binding conformations of the fragments provide a guide for us to link the lead inhibitors with fragments through various linkers to produce more potent inhibitors. Based on the docking results of a virtual collection of the new compounds, we are able to propose the potential new generation of PLpro inhibitors for future synthesis and testing.

1.3 Introduction to Previous Work

1.3.1 Lead inhibitors

In previous work, several non-covalent PLpro inhibitors of different binding modes were identified (17-19). Those inhibitors can be classified into two categories with slightly different binding modes. There are two protein-ligand complex co-crystal structures available in the RCSB Protein Data Bank showing two different binding modes (PDB id:3e9s and PDB id:3mj5) (**Figure 9**). The positions of these two classes of inhibitors overlay with the modeled substrate recognition S3 and S4 subsites of PLpro (17). In the crystal structure of GRL-0617S, when the inhibitors bind to the binding site, the side chain of L163 moves to cradle the *ortho*-methyl of the

benzene ring and simultaneously block access to the catalytic triad (17), however, in the crystal structure of GRL-0667S, the benzodioxolane ring extends to the regions that GRL-0617S did not reach (18). The inhibitors induced drastic conformational changes of the binding site BL2 loop (Gly267-Gly272) (**Figure 9**), compared to the previous apo structure. The naphthyl groups in both structures were partly solvent exposed but form hydrophobic interactions with residue Tyr269, as shown in **Figure 9**. The amide group of GRL-0617S formed two hydrogen bonds with Asp165 and Gln270 in the BL2 loop. In the structure of GRL-0667S, the amide group forms a hydrogen bond with Tyr269 in BL2 loop. Since the lead inhibitors do not bind to the catalytic site, there is a reasonable possibility that some fragment compounds could successfully bind to this location and enhance inhibitory activity of the lead inhibitors.

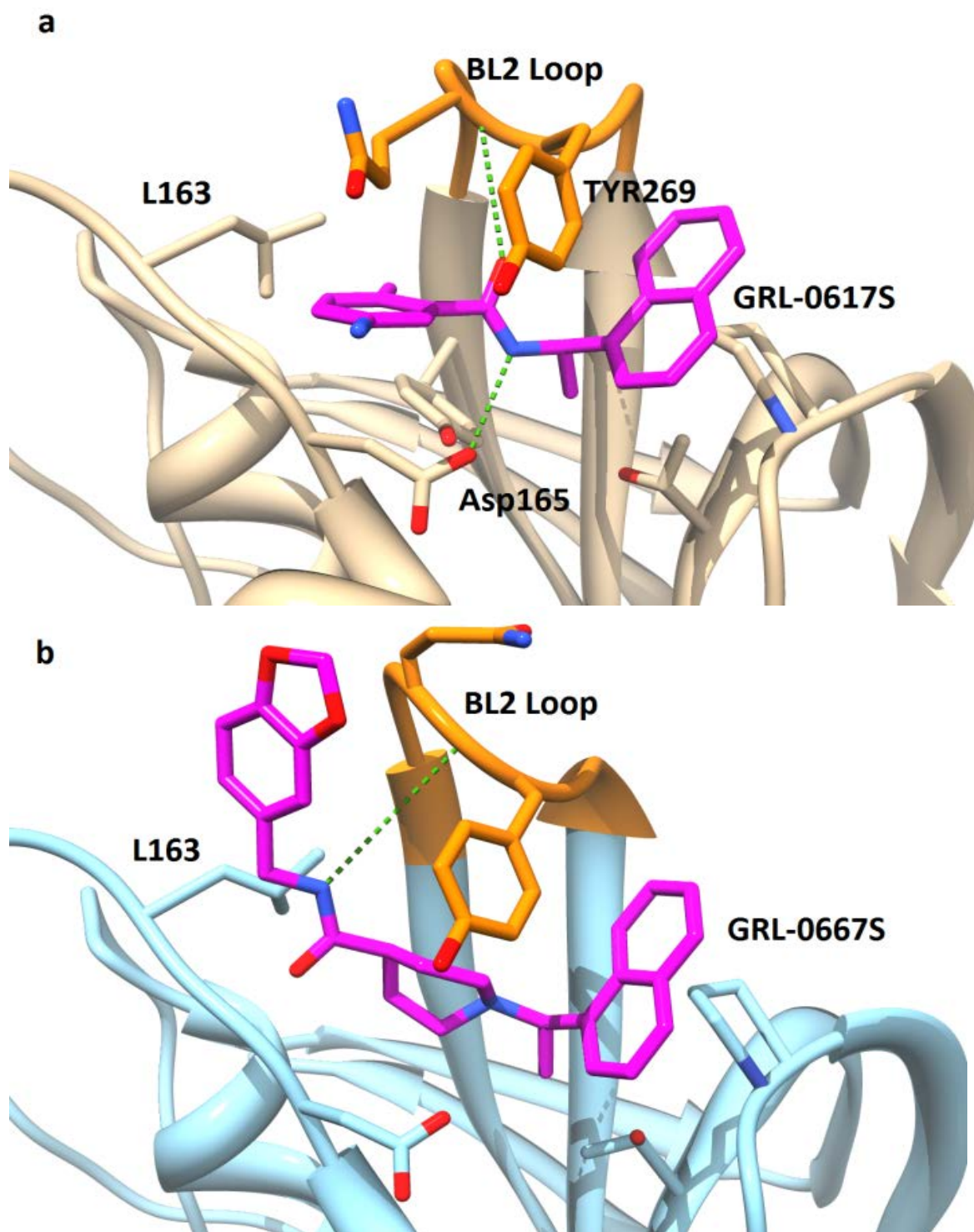


Figure 9. X-ray structure of inhibitor GRL-0617S (PDB id: 3e9s) and GRL-0667S (PDB id: 3mj5) bounded PLpro. (a) The amide group of GRL-0617S formed two hydrogen bonds with Asp165 and Gln270 in Binding Loop 2 (BL2). (b) The amide group of GRL-0667S formed hydrogen bond with Tyr269 in BL2 loop.

1.3.2 Fragments with synergistic effects

In order to explore the unoccupied regions in the PLpro ligand-bound crystal structures, screening against the fragment library was performed. The Zenobia library, consisting of 352 chemically and structurally diverse fragment-like compounds, was experimentally screened in the presence of each of the four previously discovered lead inhibitors. This new combination can increase the IC_{50} up to almost 12 fold over the original active compounds (**TABLE I**). In the Zenobia library, some compounds contain nitro group and also an aldehyde group, which make the activity suspicious. Moreover, in the process of HTS against PLpro, we discovered a compound F2124-0890, which possessed a decent IC_{50} of 10.9 μ M. Since the compound is a fragment-like structure, it was also tested with the presence of lead inhibitors. It turns out that it possessed similar synergistic activity with the lead inhibitors as the Zenobia Library fragments (**TABLE II**). Those new findings show the possibility to optimize potency of the lead inhibitors and to identify new scaffolds.

1.3.3 Inhibition mechanism of fragments with synergistic activity

The reversible activity of the fragments was determined by the dilution method. The reversibility of the six fragments ZT0273, ZT0426, ZT0537, ZT0626 and ZT0834 are over 50%. Even though ZT0408 showed the highest synergistic activity, its reversibility is only 25%. The irreversible fragments may bind to the catalytic cysteine of PLpro. Furthermore, from binding analysis by Surface Plasmon Resonance (SPR), we ruled out the possibility of non-specific binding of the fragments. ZT0273, ZT0537 and ZT0834 exhibited good binding curves. Fragments ZT0408, ZT0414, and ZT0626 did not show a typical dose-response curve as their concentrations were increased; instead they showed a random binding pattern.

In further study of the inhibition mechanism, several compounds, which have the best activity when combined with lead inhibitors (**TABLE III**), were tested for their inhibition mechanism by kinetic studies. Each compound was performed with the enzyme-inhibitor complexes and varying substrate concentrations. Fragment ZT0273, ZT0408, ZT0426, ZT0470, ZT0537, ZT0626 and ZT0834 we tested for inhibition mechanism with respect to the substrate. Alpha (α) is the interaction parameter applied to mixed-type inhibition, which reflects a change of the

affinity between the enzyme and substrate upon binding of inhibitor. All four lead inhibitor and a fragment ZT0426 showed α value bigger than 1, which indicates competitive inhibitors. The α values of fragment compound ZT0626 was found to be 1, which means that the binding of all inhibitors hinders substrate binding to the enzyme. These inhibitors Inhibition constant (K_i) being 265 μM , which indicates that ZT0626 fragment behaves as a noncompetitive inhibitor. The rest of the six fragment compounds showed α value smaller than 1, which act similar to uncompetitive inhibitors. All newly identified fragment compounds showed mixed-type inhibition, which indicates that none of them binds to the catalytic triad of PLpro enzyme. It has also been demonstrated that the synergistic fragment-like compounds do not compete with lead inhibitors at the same binding site, but bind to a separate site of the enzyme in the mutual exclusivity studies.

TABLE I.
IC₅₀ VALUE COMPARISON OF FOUR LEAD INHIBITORS IN COMBINATION WITH FRAGMENT COMPOUNDS
FROMZENOBIA library. (Experimental data from Dr. H. Lee)

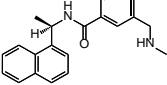
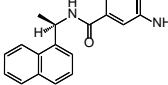
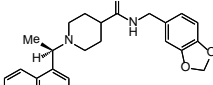
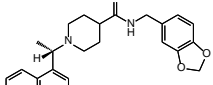
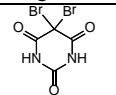
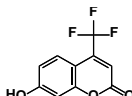
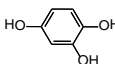


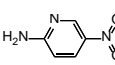

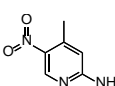
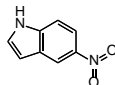
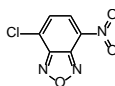
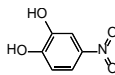
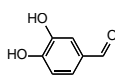
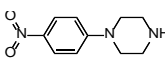
		IC ₅₀ (μM) of Lead Inhibitors			
		1. GRL-0068S	2. GRL-0617S	3. GRL-0667S	4. GRL-0737S
					
		0.29 ± 0.07	1.32 ± 0.08	0.64 ± 0.06	1.36 ± 0.08
IC ₅₀ (μM) of fragments		1 + fragment	2 + fragment	3 + fragment	4 + fragment
ZT0009 188 ± 36		0.23 ± 0.04	2.94 ± 0.09	0.20 ± 0.04	0.79 ± 0.19
ZT0273 108 ± 2		0.056 ± 0.012	0.66 ± 0.01	0.18 ± 0.01	0.49 ± 0.06
ZT0408 392 ± 28		0.049 ± 0.021	2.68 ± 0.52	0.055 ± 0.027	0.16 ± 0.03
ZT0414 250 ± 12		0.093 ± 0.027	1.09 ± 0.05	0.18 ± 0.01	0.75 ± 0.02
ZT0422 369 ± 101		0.12 ± 0.04	1.00 ± 0.04	0.21 ± 0.03	0.84 ± 0.01
ZT0426 212 ± 27		0.032 ± 0.013	0.42 ± 0.09	0.13 ± 0.03	0.25 ± 0.04
ZT0427 357 ± 27		0.15 ± 0.03	0.76 ± 0.14	0.23 ± 0.05	0.95 ± 0.08
ZT0428 269 ± 28		0.10 ± 0.03	0.79 ± 0.05	0.20 ± 0.02	0.45 ± 0.15
ZT0445 312 ± 31		0.14 ± 0.02	1.43 ± 0.50	0.30 ± 0.07	0.95 ± 0.03
ZT0470 388 ± 10		0.050 ± 0.011	1.25 ± 0.01	0.17 ± 0.04	0.44 ± 0.08
ZT0537 427 ± 20		0.051 ± 0.005	0.53 ± 0.06	0.088 ± 0.018	0.21 ± 0.05
ZT0626 223 ± 38		0.048 ± 0.005	0.34 ± 0.13	0.12 ± 0.01	0.26 ± 0.03
ZT0834 279 ± 38		0.036 ± 0.013	0.43 ± 0.08	0.074 ± 0.019	0.37 ± 0.11

TABLE II.

IC₅₀ VALUE COMPARISONS OF FOUR LEAD INHIBITORS IN COMBINATION WITH FRAGMENT COMPOUNDS F2124-0890 FROM LIFE CHEMICALS LIBRARY. (EXPERIMENTAL DATA FROM DR. H. LEE)

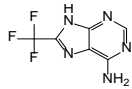
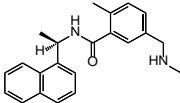
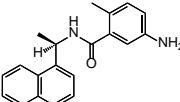
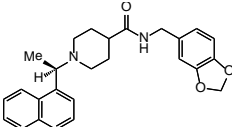
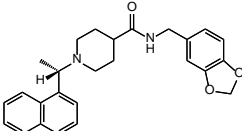
		F1. F2124-0890	
			
		IC ₅₀ = 10.9 ± 0.9 μM	
Lead Inhibitor (I)	Structure	IC ₅₀ (μM) of lead Inhibitor (I)	
		alone	I + F1
1. GRL-0068S		0.23 ± 0.07	0.034 ± 0.008
2. GRL-0617S		1.18 ± 0.05	0.27 ± 0.05
3. GRL-0667S		0.92 ± 0.07	0.22 ± 0.07
4. GRL-0737S		2.2 ± 0.08	0.74 ± 0.06

TABLE III.

MECHANISM OF EXISTING PLRO INHIBITORS AND SEVERAL FRAGMENTS IDENTIFIED IN ZENOBIA LIBRARY. (EXPERIMENTAL DATA FROM DR. H. LEE)

Compound ID	Type of inhibition	K _i (μM)	K _m (μM)	α
No compound			97.4 ± 9.2	
GRL-0068S	Mixed	0.42	177	2.6
GRL-0617S	Mixed	2.00	195	2.0
GRL-0667S	Mixed	0.42	177	2.7
GRL-0737S	Mixed	0.80	190	2.5
ZT0273	Noncompetitive	207	153	1.0
ZT0537	Mixed	234	158	1.5
ZT0834	Mixed	176	192	1.5

Chapter 2

2. HOT SPOT ANALYSIS OF LIGAND-BOUND PLPRO CRYSTAL STRUCTURE

2.1 Introduction and Purpose of the Study

Hot spots are locations on the protein surface that contribute significantly to the ligand binding free energy. Those spots are the main targets for many biological applications including rational drug design (20). Protein sites with high-binding affinity can be identified by X-ray crystallography or nuclear magnetic resonance (NMR) and also computational methods(20) by screening against libraries of small organic molecules that tend to cluster at those regions. The docking result of the organic molecules could represent the behavior of hydrophobic, polar and aromatic moieties. Those favorable positions can be ranked on the basis of the average energy.

From the result of mechanism of inhibition and mutual exclusivity studies, the newly identified fragment-like compounds bind to sites separate from these of the lead inhibitors. Unfortunately this information alone does not provide us with the exact location in the PLpro enzyme of fragments' binding site. For this reason, we investigated all possible binding site candidates. In the crystal structures of GRL-0617S and GRL-0667S, there are some unoccupied spaces besides the binding site of the lead inhibitors, including the catalytic region. In order to assist in the further optimization of the active compounds, we used computational methods to identify the potential binding pockets in the PLpro structures with the presence of the lead inhibitors. The conformations of the ligand-receptor bound co-crystal structures were sampled by molecular dynamics simulation, followed by hot spot analysis by FTMAP server. FTMAP is a server that models the experiment screening approach to sample the surface of a target protein using small organic molecules as probes. It goes through the following steps: rigid body docking of probe molecules, minimization and rescoring of the docking poses, clustering and ranking of the docking poses, searching the consensus sites and finally finding the

probable binding sites (20). The 16 probes used in FTMAP are ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide and N,N-dimethylformamide. The FTMAP server was used to identify potential druggable PLpro protein subpockets by performing a fragment-based search. By comparing different ligand-bound crystal structures, we can find the conserved hot spots that served as the possible pockets for the newly discovered synergistic inhibitors.

2.2 Materials and Methods

2.2.1 Molecular Simulation of the Ligand-bound PLpro Crystal Structure

All molecular dynamics simulation were performed using the AMBER suite of programs (AMBER 11)(21). Two model system were studied, one was the GRL-0617S bound PLpro crystal structure (PDB id :3e9s), one was the GRL-0667S bound PLpro crystal structure (PDB id:3mj5), which are representative crystal structures of each compound scaffold. All crystal structures were obtained from the RCSB Protein Data Bank, and multimer protein complexes were reduced to the functional, monomer form. The bound compound and zinc ions were kept in the systems while other ligands and crystal waters were removed. The parameters of ligands were generated by Antechamber in AmberTools v1.5 using the General Amber Force Field (GAFF) with am1bcc charge (22, 23). Leap was used to neutralize the systems by adding sodium ions and solvating the system using an eight angstrom rectangular TIP3P water box. The systems are prepared in ff99SB Amber Force Field (24). Minimization using SANDER was carried out in two steps, only the water molecules were minimized first for 5,000 steps, while keeping the protein fixed followed by the second stage of minimization of the entire system for another 10,000 steps. Periodic boundary conditions were applied with 8 Å non-bonded cutoff. The systems were heated from 0K to 300K using an NVT ensemble for 50ps with positional restraint on the protein. Then the systems were equilibrated for 100ps using the NPT ensemble without positional restraint. The production phase of the simulation was performed in the NPT ensemble without positional restraint for 50 ns. Further

analysis used the ptraj module from AmberTools 1.5 to compute the average structures every 1 ns and write out the PDB file for the averaged coordinates. A total of 50 snapshots were extracted using this method.

2.2.2 FTMAP

The PDB files obtained above were uploaded to the FTMAP server (<http://ftmap.bu.edu/>). The FTMAP server removes the waters, ions and bound ligands and retaining only the protein structure. Sixteen different probes were used: ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, N,N-dimethylformamide. The output results generated by FTMAP were in the form of PDB files with consensus clusters of the different probes mapped to the surface of the PLpro structure. Clusters sites identified by FTMAP were required to be conserved between all molecular probes and snapshots for selection and further study.

2.3 Results and Discussion

By post analyzing the trajectory files from the 50ns simulation of the lead inhibitor bound co-crystal structures, the RMSD changes of the backbone residues of each structure and the RMSF changes of each residue in each crystal structure over the time scale were obtained (**Figure 10**). In the last 10ns simulation of the GRL-0667S bound structure, there was a dramatic change in the backbone RMSD, due to the sudden increasing mobility of the ubiquitin-like domain. By looking at the RMSF of each residue in total 50ns, we can see that the fluctuations over the simulations were not evenly distributed over all residues. The residues which were mostly responsible for the fluctuations were similar between the two crystal structures, including the zinc-binding motif and catalytic triad part. The only unexpected difference was that the ubiquitin-like domain of the GRL-0667S structure has a significantly larger fluctuation than the GRL-0617S one. The extended benzodioxolane ring part of GRL-0667S may cause some differences in the motion in that region, which were not visible in the crystal structure. There are

also some structural changes around the residue Asp287, which is one the residue from the catalytic triad.

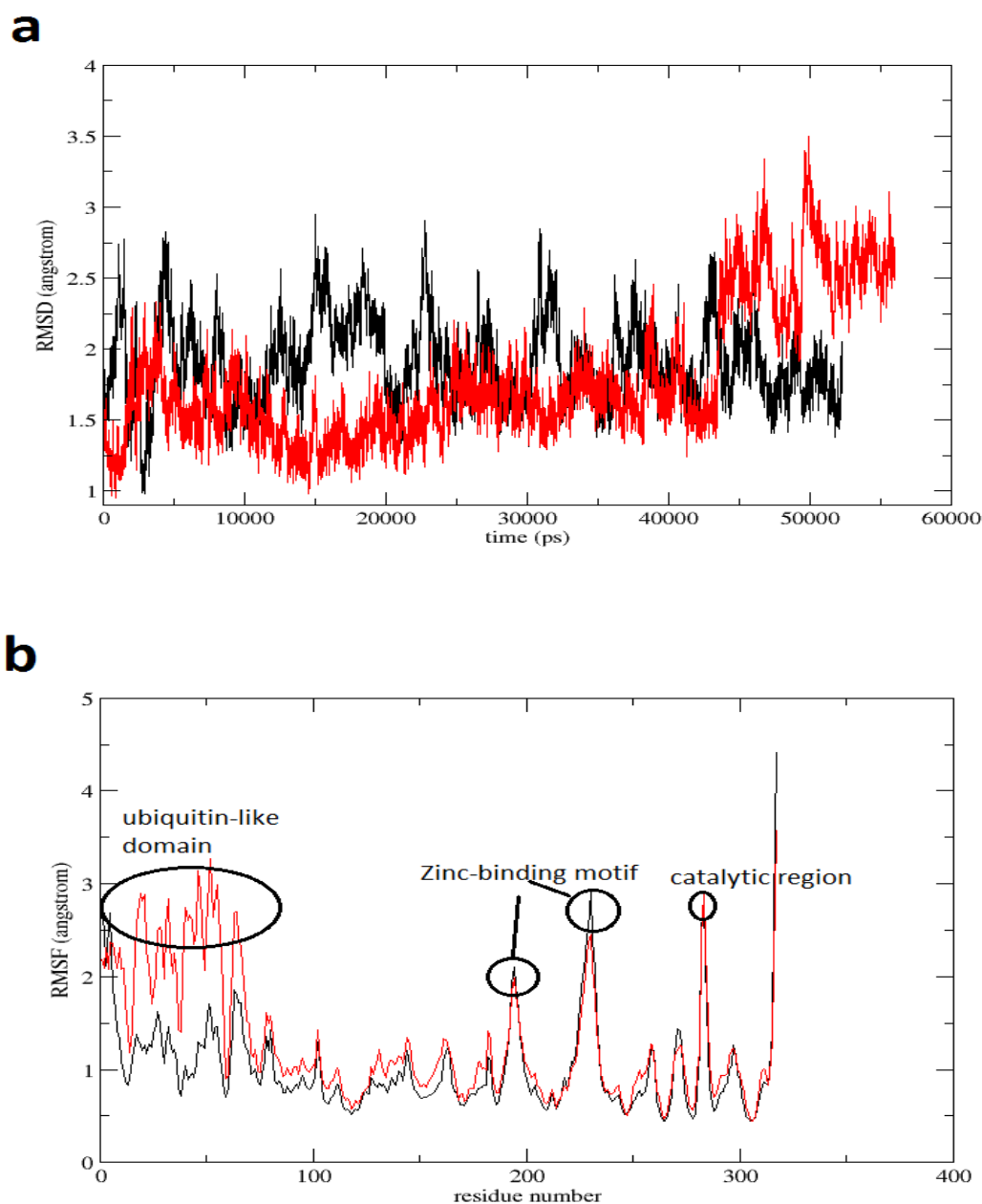


Figure 10. The molecular dynamic simulation results of the ligand-bound PLpro crystal structures. (a) RMSD changes of the backbone residues of each structure over 50ns. (b) RMSF of each residue of each crystal structure in total 50ns. The results for the crystal structure of GLR-0617S (PDB id:3e9s) are colored in black while those from the crystal structure of GRL-0667S (PDB id:3mj5) are colored in red.

In order to identify unoccupied binding pockets for compound optimization, we performed a series of computational solvent mapping experiments, or 'hot spot' analyses, using the FTMAP server. Representative snapshot structures were extracted from the simulations as discussed above (see Materials and Methods) and submitted to the FTMAP server, which performed a fragment-based binding site analysis. Two strong candidates for fragment binding sites were identified by analyzing the original structures and consensus clusters of used probes molecules. Two strong candidates for fragment binding sites were identified by analyzing the original structures and consensus clusters of the probes molecules used. The two candidate sites identified included an extension of the original binding site of the lead inhibitors and one cavity in the palm region, designated candidate sites 2 and 3 (**Figure 11**). Candidate site 2 was larger than the lead inhibitor binding site, which is also the substrate recognition site, and included an additional cavity unoccupied by the lead inhibitors. This binding site has been previously discussed as a possible substrate recognition site. This extended cavity is composed of five residues, Arg167, Glu168, Met209, Asp303, and Thr302. The second position identified by FTMAP, site 3, was located more than 10 Å away from both the lead inhibitor binding site and the catalytic site of the PLpro. It is encompassed by the zinc-binding motif and the palm region. Lastly, the small volume containing the catalytic triad (site 1) was not identified by FTMAP, likely due to the smaller size of this pocket.

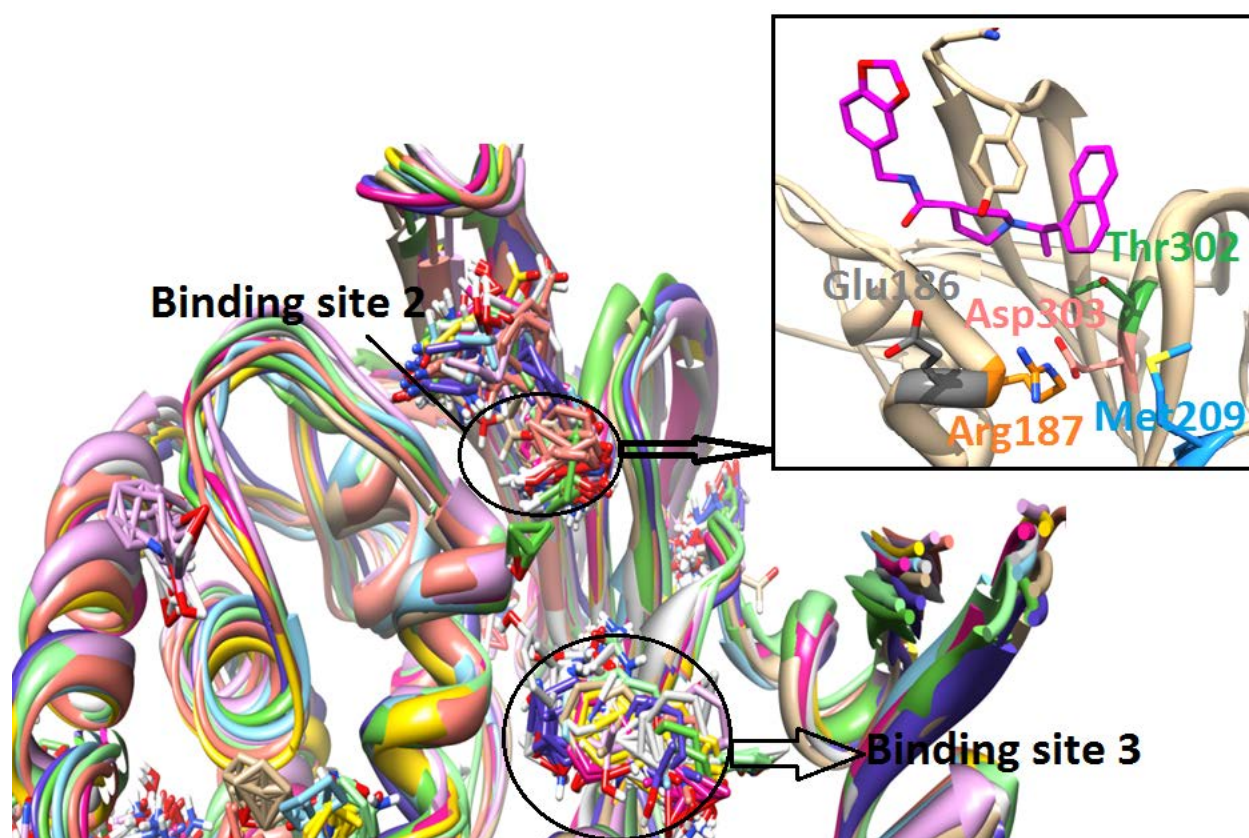


Figure 11. Fragment binding site analysis results by FTMAP. Two potential binding sites identified by FTMAP using the co-crystal structure of PLpro with inhibitor GRL-0667S are shown. The first candidate site (upper circle) is an extension of the lead inhibitor's binding site, formed by Glu186 (Grey), Arg187 (Orange), Met209 (Blue), Thr302 (Green), and Asp303 (Pink). The second candidate site (lower circle) is encompassed by the zinc-binding motif and the palm region.

Chapter 3

3. FREE BINDING ENERGY CALCULATION TO DETERMINE POSSIBLE BINDING SITES OF THE FRAGMENTS THAT GREATLY ENHANCED ENZYMATIC INHIBITION OF LEAD INHIBITORS

3.1 Introduction and Purpose of the Study

In Section 2.1, we found a number of fragments that will greatly enhance enzymatic inhibition of the lead inhibitors. This observation encouraged us to consider the possibility of linking or merging the fragments and lead inhibitors, which may lead to more active compounds. From the experimental results, we exclude the possibility of the catalytic triad being the binding site of the fragments. However, the actual binding site of those fragments still remains a question. In our previous computational studies, we found several possible binding sites in the crystal structures. Simple docking is not sufficient to suggest which binding site is more probable, so the calculation of the free binding energy for each docking complex was used. For free energy binding calculation, Free-energy perturbation (FEP) and Thermodynamic Integration (TI) are more accurate methods, which perform sufficient sampling. However, they are very computational expensive. MMPBSA (Molecular Mechanics/Poisson-Boltzmann Surface Area) is a method for computing binding affinity based on molecular mechanics (force field) and electrostatic calculation (Poisson Boltzmann (PB) or Generalized Born (GB)), which has been successfully applied to several studies to estimate binding affinity (25, 26). It requires much less computational time than FEP and TI. Even though it is just a rougher estimate, compared with FEP and TI, it is still adequate to evaluate plausibility possible binding sites. The binding free energies were calculated by the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

where G_{complex} is the calculated Gibb's free energy of the inhibitor-fragment-enzyme complex, G_{receptor} is the Gibb's free energy for the fragment-enzyme complex, and G_{ligand} is the Gibb's free energy calculated for the inhibitor compounds. The free energy values for each of these terms were estimated as the sum of the four terms shown here:

$$G = E_{\text{MM}} + G_{\text{psolv}} + G_{\text{npsolv}} - TS_{\text{nmode}} \quad (2)$$

Where E_{MM} is the molecular mechanics energy of the molecule expressed as the sum of the internal energy of the molecule plus the electrostatics and van der Waals interactions, G_{psolv} is the polar contribution to the solvation energy of the molecule, G_{npsolv} is the nonpolar solvation energy, T is the absolute temperature, and S is the vibrational, rotational, and translational entropy of the molecule. We picked fragment ZT0834 and F2124-0890 as examples for MMPBSA calculation to suggest the binding site for fragments.

3.2 Materials and Methods

3.2.1 Starting conformations generated by Induced Fit Docking (IFD)

Fragment docking was performed by Glide-XP (extra precision) combined with Induced Fit Docking (IFD)(27, 28) protocol of Schrödinger using default settings in order to explore the possible binding sites of the fragments. The structures of PLpro bound with GRL-0617S (PDB id: 3E9S) and GRL-0667S (PDB id: 3MJ5) were obtained from the RCSB Protein Data Bank (PDB). Prior to the molecular docking, structures of protein targets were prepared using the Protein Preparation Wizard in the Schrödinger software package (Suite 2012). All water molecules in the crystal structures were removed, and multimeric complexes were simplified to monomer from the original PDB structure. Residues which are within the 5 Å of the bound inhibitors were refined by Prime. Two distinct docking runs were performed using this strategy. The first docking run was blind docking (coarse docking), with the docking area defined to be within 10 Å of the lead inhibitors. The subsequent docking runs were defined by the location of the candidate sites identified by the structural and computational analyses, described above. Binding site 1

(the catalytic site) was defined by selecting residues His273, Cys112 and Asp287; binding site 2 was defined by selecting residues Arg167, Glu168, Met209, Asp303, Thr302; and binding site 4 was defined by selecting residues Asn157 Lys158, Asp165 and Glu168. In each case, the docking area was defined to be a 10Å box centered on the selected residues. All eight fragments from Zenobia Library and the one from Life Chemicals were prepared in the tautomer and protonation state at PH 7.4 by Schrödinger's LigPrep (29). All nine active fragments were docked in the coarse docking studies; only fragment ZT0834 and Life Chemicals F2124-0890 were docked in the subsequent, candidate site, docking studies. All the docking poses were ranked by Glide SP score(28). In the second docking run, three conformations of each binding site (binding sites 1, 2 and 4) produced by IFD were selected to be the starting conformations for MM/PBSA calculation. Three different conformations in each binding site were selected including the top score conformation and other two conformations which were most different from the top score one.

3.2.2 Molecular Dynamic Simulation

All molecular dynamics simulations and subsequent MM/PBSA calculations used to predict binding free energies were performed using the AMBER suite of programs (AMBER 11)(21). The preparation of the MD simulations, including parameterization of the fragment (F8) and inhibitors, and preparation and equilibration of the system, were performed exactly as described above in Section 2.2.1 for binding site analyses. The production phase of the simulation was performed in NPT ensemble without positional restraint for 20ns.

3.2.3 MM/PBSA

The python script, MMPBSA.py, included in AmberTools v1.5, was used to perform the MM/PBSA calculations. The first 100 frames from the last 2 nanoseconds (equilibrated) of the simulation trajectory file were used for the MM/PBSA calculations. The ionic strength in molarity was set to 0.1 and the internal dielectric constant was set as 1.0 (default). All other options were set to default settings. The

entropy of the molecule was estimated by normal mode analysis, using 5 regularly spaced snapshots from the first 100 frames in the last 2 nanoseconds of the MD simulations. Ionic strength used in normal mode calculations (in molarity) was set as 0.1.

3.3 Results and Discussion

In the previous section, we found several putative binding sites for fragment binding (**Figure 12**).

Binding site 1 is the catalytic triad, formed by His273 Asp287 and Cys112; binding site 2 is an extended pocket of the lead inhibitors' binding site identified by FTMAP; binding site 3 is wrapped up by the palm region and zinc-binding motif, which is more than 10Å farther away from the lead inhibitors' binding site. In the coarse docking process, an additional shallow and solvent-exposed binding site which is located within 10 Å radius of the lead inhibitors' binding site was identified as binding site 4.

The previous experimental results showed that the mechanism of those fragments behaved the uncompetitive manner with the respect of the substrate. Since the fragments showed synergistic effect with the lead inhibitors, they bind in combination with the inhibitors in a distinct and separate binding site of the enzyme.

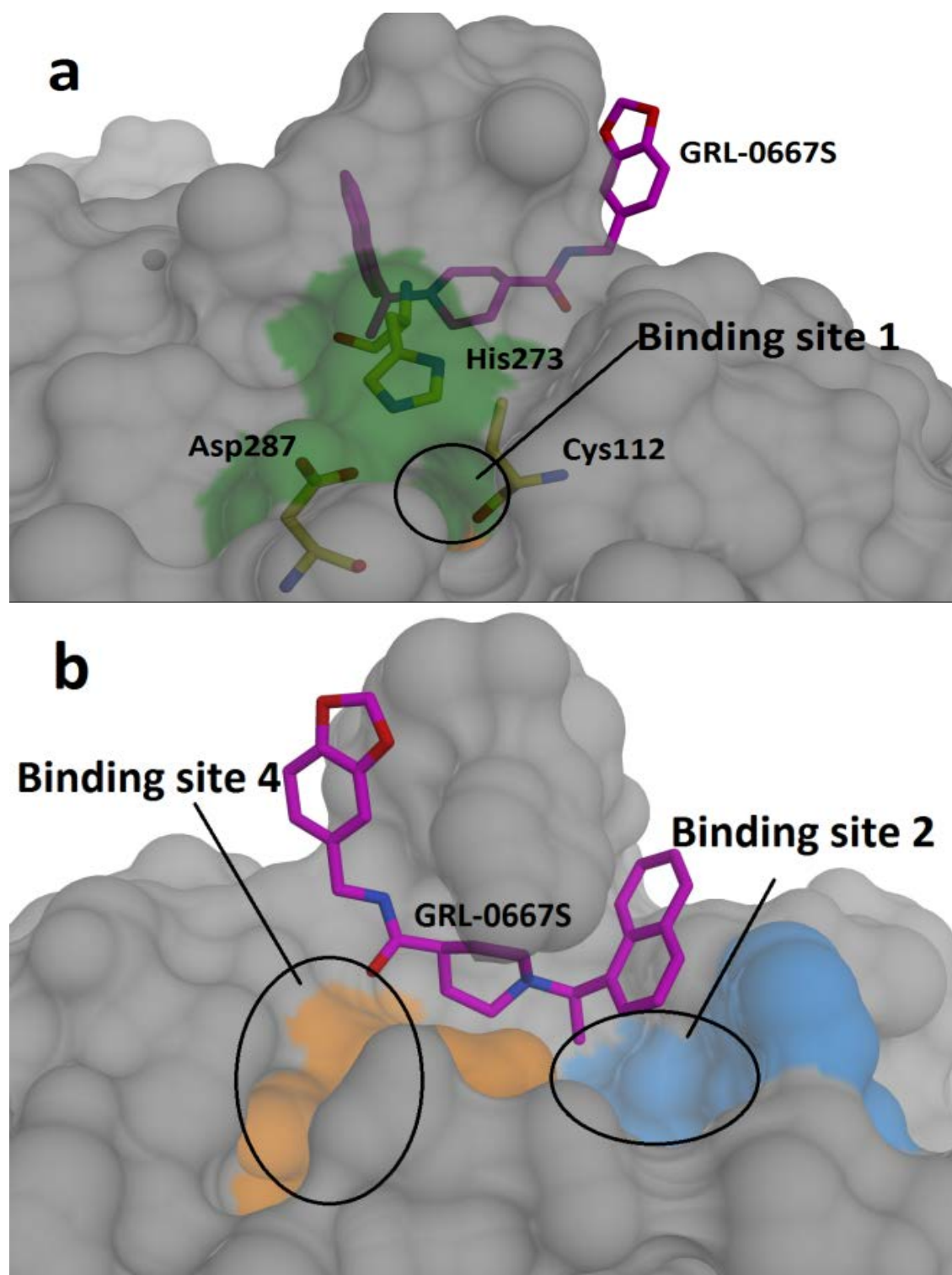


Figure 12. Potential binding site of the fragment compounds. (a) Unoccupied catalytic pocket surrounded by three residues (C112, H273, and D287). This small pocket is considered as potential fragment binding site 1. (b) Two small unoccupied pockets in the near PLpro lead inhibitor binding site. Binding site 2 was identified by FTMAP hot spot analysis, and the third putative binding sites (Binding site 4) were identified by docking studies in the near the currently lead inhibitor binding site. The image from (a) rotated by 90 degrees anticlockwise gives a better view of the binding site 2.

In order to further analyze the putative binding site and to estimate the most probable binding conformations of the fragments within each site, we performed a series of computational binding free energy calculations. Fragment ZT0834 from Zenobia Library and Fragment F2124-0890 from Life Chemicals were selected in the MM/PBSA calculation studies due to their strong synergistic effects on both inhibitors GRL-0617S and GRL-0667S. These studies were performed using site 1, the catalytic site, site 2, the extended inhibitor binding site, and site 4. Binding site 3 was excluded because it is very shallow, highly solvent exposed and located far away from the lead inhibitors' binding site (>10 Å).

During the 20 ns simulation, it was firstly noticed that both Fragment ZT0834 and Fragment F2124-0890 repeatedly escaped from binding site 4 regardless of their starting poses, which indicates that this shallow and solvent-exposed cavity is highly unlikely to be the actual binding site of these fragments. Because of the reason mentioned above, binding site 4 was removed from further MM/PBSA calculation and analysis.

The calculated binding free energy of the lead inhibitors GRL-0617S and GRL-0667S with the fragment ZT0834 placed in each successive candidate site is shown in **TABLE IV** and **TABLE V**. When the fragment was docked into binding site 1, the calculated binding free energy is lower than those calculated for the inhibitor alone as well as the experimentally derived results, which showed clearly that binding site 1, the catalytic triad was not a favorable site for fragment ZT0834 binding. Also, in the experimental section, the inhibition mechanism of ZT0834 was not competitive with the substrate, which was consistent with our calculation results. When the fragment posited at binding site 2, favorable results of calculated binding free energies were obtained. The results were consistent with the synergistic combination of the fragment and the lead inhibitors, indicating that binding site 2 is a more favorable position for fragment ZT0834 binding. The best conformations of ZT0834 placed in combination with the lead inhibitors GRL-0617S and GRL-0667S are shown in **Figure 13**. In binding site 2, there is an anionic pocket formed by Asp303, Thr302, Tyr274, and Asp165, which interacts with the positive

charged piperazine amine group of ZT0834. The lead inhibitors may be further stabilized by an additional edge-to-face Pi stacking interaction between the naphthalene ring system and the aromatic ring of fragment ZT0834 in addition to their established binding features.

TABLE IV.

CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0617S WHEN FRAGMENT ZT0834 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.

Conformation #	Binding site 1			Binding site 2			Inhibitor GRL-0617s alone	
	1	2	3	1	2	3	calculated	experimental
t _{ds}	-18.99	-20.13	-21.07	-19.82	-20.64	-19.96	-21.87	
δh	-20.17	-24.78	-25.83	-29.56	-30.18	-32.74	-27.33	
ΔG(Kcal/mol)	-1.18	-4.65	-4.76	-9.74	-9.54	-12.78	-5.46	-8.7

TABLE V.

CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0667S WHEN FRAGMENT ZT0834 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.

Conformation #	Binding site 1			Binding site 2			Inhibitor GRL-0667S alone	
	1	2	3	1	2	3	calculated	experimental
TΔS	-19.89	-20	-19.94	-20.39	-20.04	-20.11	-20.32	
ΔH	-20.65	-21.75	-22.47	-36.49	-26.73	-24.17	-25.89	
ΔG(Kcal/mol)	-0.76	-1.75	-2.53	-16.1	-6.69	-4.06	-5.57	-7.82

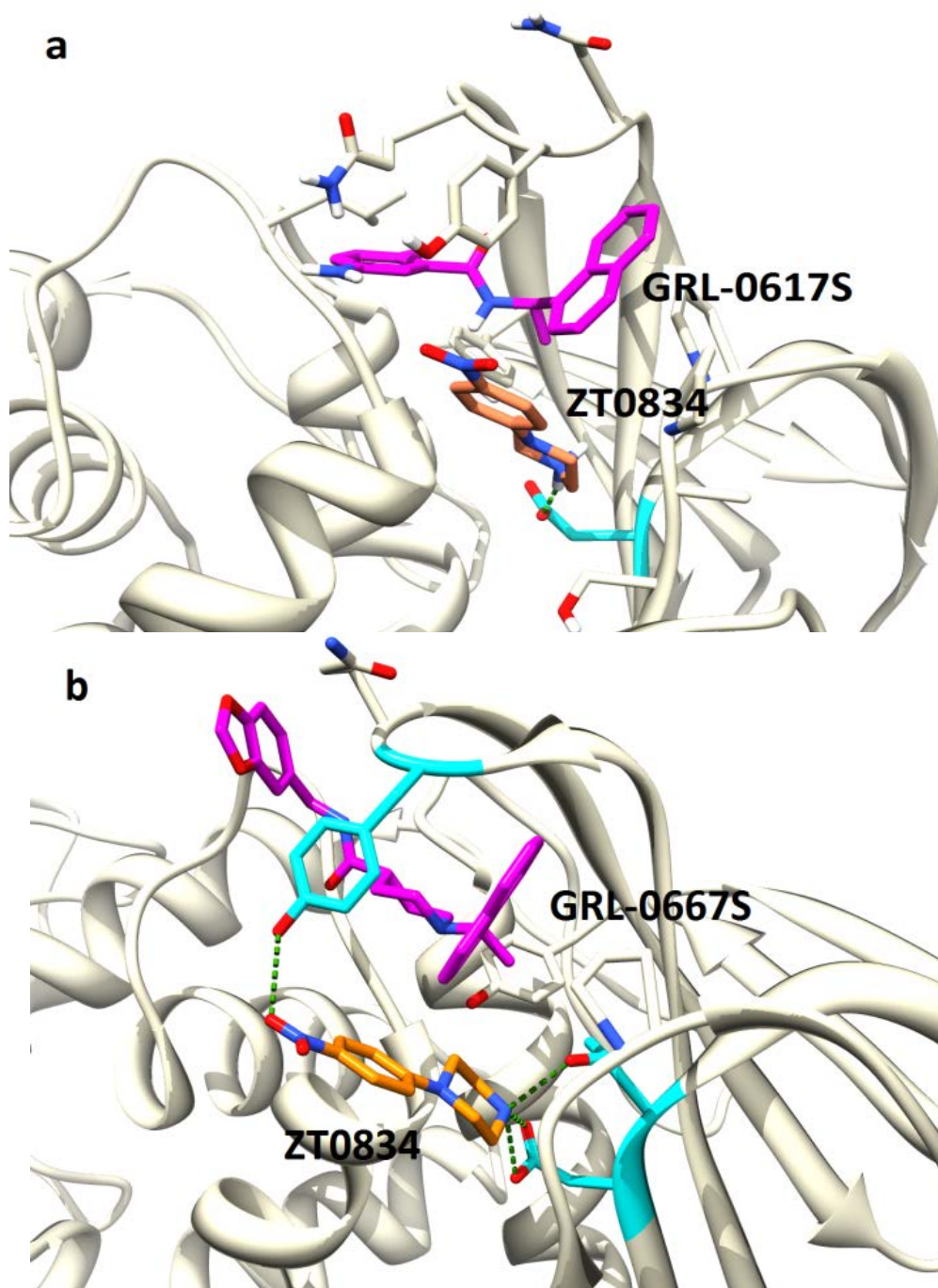


Figure 13. Most probable binding conformation of fragment ZT0834 (orange) in combination with lead inhibitors GRL-0617S (magenta)(a) and GRL-0667S (magenta)(b). Interacting residues of PLpro enzyme to fragment ZT0834 are shown in cyan.

The results of the calculated binding free energies of the lead inhibitors GRL-0617S and GRL-0667S with the fragment F2124-0890 were consistent with the results of fragment ZT0834 (TABLE VI and TABLE VII). The most favorable binding site among the putative binding sites was still binding site 2 and the best conformations obtained from MM/PBSA calculation are shown in **Figure 14**. Most of the interactions were formed by the amine group of F2124-0890 and the charged residues Asp165 and Arg167.

TABLE VI.

CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0617S WHEN FRAGMENT F2124-0890 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.

Conformation #	Binding site 1			Binding site 2			Inhibitor GRL-0617S alone	
	1	2	3	1	2	3	calculated	experimental
TΔS	-16.89	-20.29	-20.23	-21.62	-20.82	-20.76	-21.87	
ΔH	-4.7	-4.7	-15.61	-29.83	-23.43	-36.11	-27.33	
ΔG(Kcal/mol)	12.19	15.79	4.62	-8.21	-2.61	-15.35	-5.46	-8.7

TABLE VII.

CALCULATED BINDING FREE ENERGIES OF LEAD INHIBITOR GRL-0667S WHEN FRAGMENT F2124-0890 WAS PLACED AT DIFFERENT CANDIDATE SITES. THREE DIFFERENT DOCKED CONFORMATIONS AT EACH BINDING SITE WERE SELECTED FOR CALCULATIONS.

Conformation #	Binding site 1			Binding site 2			Inhibitor GRL-0667S alone	
	1	2	3	1	2	3	calculated	experimental
TΔS	-20.78	-19.87	-19.33	-20.82	-19.64	-18.96	-20.32	
ΔH	-26.16	-23.66	-22.57	-30.04	-35.34	-33.75	-25.89	
ΔG(Kcal/mol)	-5.38	-3.79	-3.24	-9.22	-15.7	-14.79	-5.57	-7.82

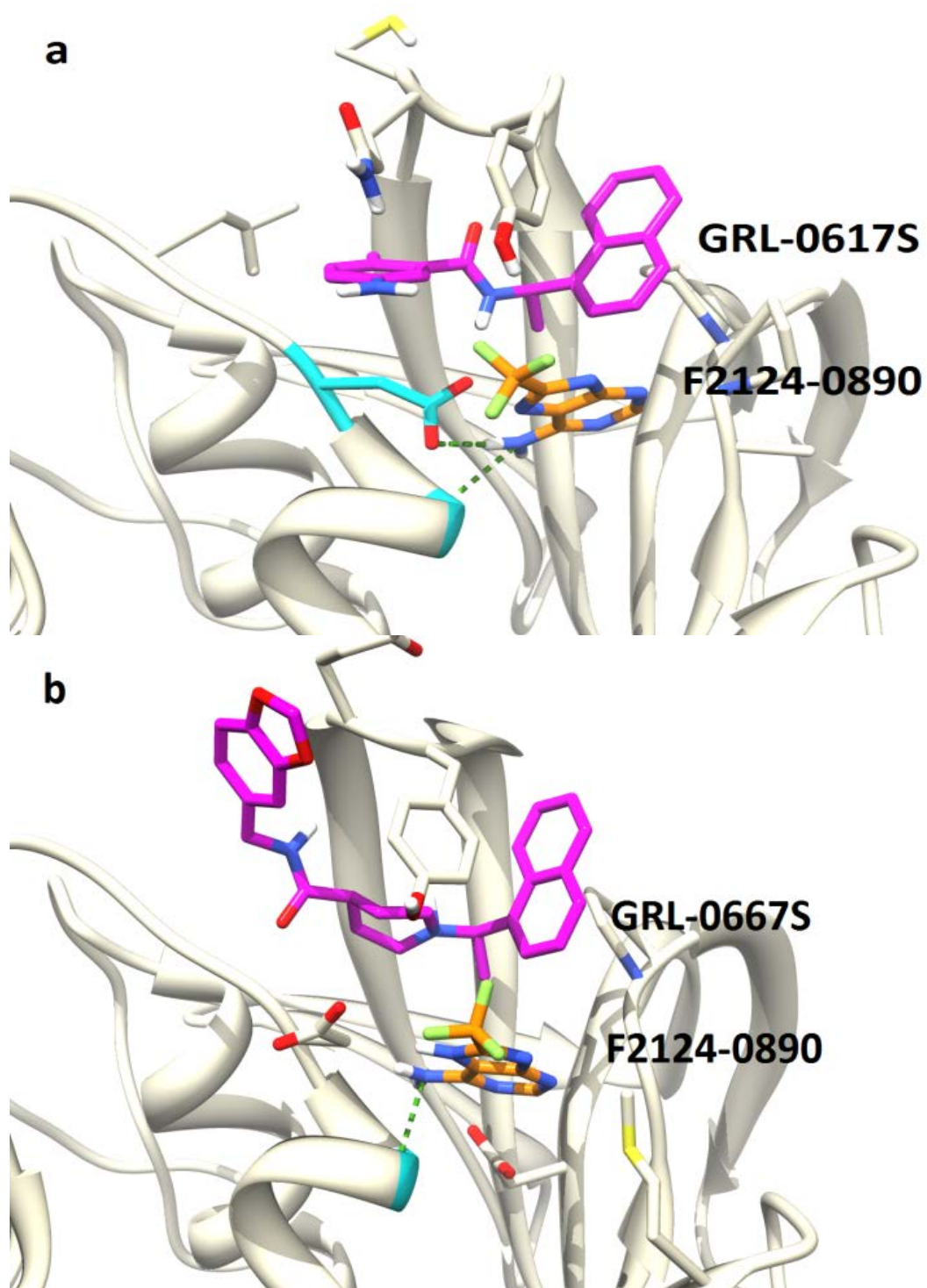


Figure 14. The most probable binding conformation of fragment F2124-0890 (orange) in combination with lead inhibitors GRL-0617S (magenta) (a) and GRL-0667S (magenta) (b). Interacting residues of PLpro enzyme to fragment F2124-0890 are shown in cyan

Chapter 4

4. DESIGN NEW COMPOUNDS BASED ON THE PREVIOUS STUDY BY FRAGMENT LINKING OR FRAGMENT MERGING METHOD

4.1 Introduction and Purpose of the Study

Fragment based drug design has been known as a well-established method for lead optimization in drug discovery and design for the past fifteen years (30). It has been proved that linking two fragment compounds that occupy unique sub-sites of the target can result in a superadditivity of binding affinities, in which the binding free energy of the new compound is much greater than the simple sum of the binding energies of individual fragments (31). From the binding free energy calculations above, the binding site 3 is the most probable position, which is an extended pocket of the lead inhibitor's binding site. It encouraged us to think about the possibility of linking the lead inhibitors with the fragments identified through experimental screening above, which may lead to new compounds of novel scaffold with better activity. The linking and merging strategy was mostly based on recapitulating the binding position of the lead inhibitors as well as key binding features of the identified fragments. Computational drug design was quite useful in modeling the changes of binding affinity that resulted from various linking positions and linker groups.

4.2 Materials and Methods

4.2.1 Virtual compound generation by fragment merging

The fragment merging strategy was based on the conformations from the MM/PBSA calculation. The new compounds were designed by connecting the lead inhibitors with ZT0834. The problematic aromatic nitro part was deducted and the piperazine ring was retained. It was linked with lead inhibitors by various linkers through different points. In order to retain the conformation of the lead

inhibitors, we tried several possible linking strategies, including connecting from the naphthalene ring, chiral methyl group on the chain and also the P-toluidine or benzodioxole by alkyl groups, ester groups and amide groups (**TABLE VIII** and **TABLE IX**).

TABLE VIII.

THE STRUCTURES OF PROPOSED COMPOUNDS OF MERGING GRL-0617S WITH ZT0834.

TABLE IX.

THE STRUCTURES OF PROPOSED COMPOUNDS OF MERGING GRL-0667S WITH ZT0834.

4.2.2 Docking

The binding conformations of the new compound collection were predicted by Glide in Schrödinger software packages. The structures of PLpro bound with GRL-0617S (PDB id: 3E9S) and GRL-0667S (PDB id: 3MJ5) were obtained from RCSB Protein Data Bank (PDB). Prior to the molecular docking, structures of protein targets were prepared using the Schrödinger software package. All water molecules were removed, and multimeric complexes were simplified to the monomer from the original PDB structure. The docking site was defined by the original binding site of the lead inhibitors. The new compound collection was prepared in the tautomer and protonation state at PH 7.4 by Ligprep (29) in Schrödinger. The docking site was defined as a 10 Å box centered on the original binding site of the lead inhibitors. Docking was conducted at SP (standard precision) mode. All the docking poses were ranked by Glide SP (standard precision) score. Proposed compounds were selected by score and visual comparison to the crystallographic position of the lead inhibitor and the predicted binding conformation of the fragment.

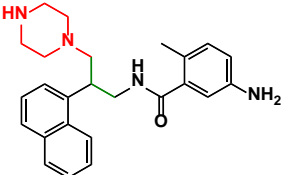
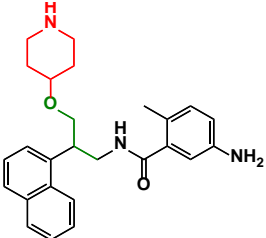
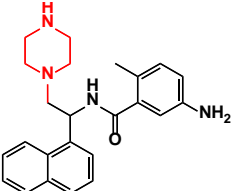
4.3 Results and Discussion

The above section provides the clue of binding site for the Zenobia Library and Life Chemicals fragments, which lays a foundation for further optimizing and prioritizing the lead inhibitors. We chose the fragment ZT0834 from Zenobia library as the starting point for fragment merging and linking strategy. Since the most probable binding site for the fragments, the binding site 2, is an anionic pocket, which is composed of several negative charged residues, we believe that the positive charged piperazine ring is essential for retaining its synergistic activity. Our fragment-merging strategy involved the elimination of a potentially problematic nitro-aromatic ring of the fragment ZT0834 and the direct incorporation of the piperazine ring into the lead inhibitor scaffold using various linking groups and different points. The new compounds collection was tested by docking to the extended binding site defined by the lead inhibitor

and binding site 2. Proposed compounds were selected by docking score and their ability to successfully recapitulate the key binding features of the inhibitor and synergistic fragment, including the lipophilic interactions of the naphthalene ring system and the positioning of the positively charged amine of the piperazine into the key, anionic pocket discussed in the above section. Compounds with best docking scores are shown in **TABLE X**.

TABLE X.

COMPOUNDS PROPOSED FOR SYNTHESIS AND TESTING BY FRAGMENT-MERGING AND DOCKING CALCULATIONS. CHEMICAL STRUCTURES IN BLACK COLOR ARE FROM THE LEAD INHIBITORS, WITH THE RED COLORED MOIETY CAME FROM THE FRAGMENT ZT0834. LINKERS BETWEEN THE LEAD INHIBITOR AND A FRAGMENT ARE SHOWN IN GREEN.

GRL-0617S + ZT0834		
Compound	Structure	Docking Score
M1		-7.837
M2		-6.421
M3		-5.735

All the compounds that could retain the conformation of the lead inhibitors and the piperazine ring from the fragment ZT0834 were linked through the chiral methine group. Other proposed compounds could not recapitulate the binding conformations probably due to the angle between the linking group and the lead inhibitors while forced the piperazine ring out of in the anionic pocket. The docking conformation of the proposed compound M1 is shown in **Figure 15** with comparison of the original binding conformation proposed by MM/PBSA calculations.

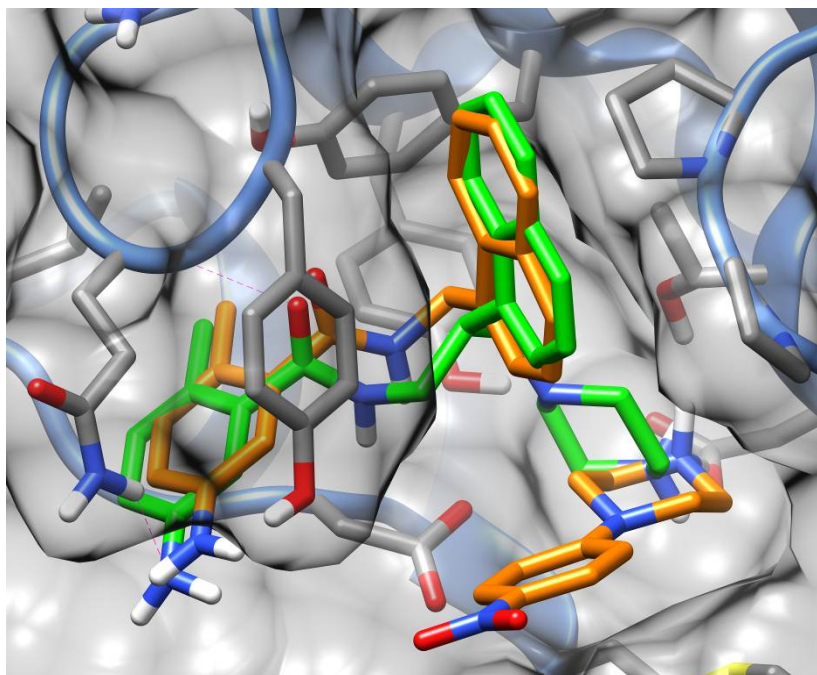


Figure 15. Proposed compound for synthesis and testing by fragment-merging strategy based on the proposed conformations of GRL-0617S and ZT0834 from MM/PBSA calculations.

In summary, in order to optimize current inhibitors of PLpro, a series of fragments from Zenobia Library and Life Chemicals were screened in the presence of the previously discovered inhibitors. Since the lead inhibitors did not bind at the active site of PLpro, fragments which could take over the unoccupied catalytic triad (binding site 1) were expected. In the screening process, several fragments showed synergistic activity with the lead inhibitors were identified. The enzymatic analyses experiments (including mechanism of inhibition and mutual exclusivity studies) alone could not give structural information about the specific binding site and binding conformation of these fragments. To further obtain structural information about the fragments, computational studies, including molecular dynamic simulation, solvent mapping (hot spot analysis), docking and binding free energy calculations were employed. Snapshots of the co-crystal structure were extracted from 50ns conventional simulation trajectories and sent to the FTMAP server. Two strong candidates for fragment binding sites were identified by analyzing the original structures and consensus clusters of used probes molecules. The two candidate sites identified included an extension of the original binding site of the lead inhibitors (binding site 2) and one cavity in the palm region (binding site 3). Another binding site (binding site 4) was discovered through coarse docking (**Figure 8**). There were four candidate binding sites, but neither the solvent mapping nor docking could give the information about the exact binding location for those fragments. Based on the hypothesis that there will be a favorable decrease of the lead inhibitors' binding free energy in combination with fragments, MM/PBSA calculations were used to compare the binding free energy of the lead inhibitors with the fragments placed in each binding site. Binding site 3 was excluded from further analysis since it was more than 10Å away from the original binding site which is unlikely to affect the improved activity observed. From the MM/PBSA free energy calculation result (**TABLE IV-VII**), binding site 2 is the most probable one for fragments binding. Furthermore, a series of compounds which were proposed by merging the fragment compound and the lead inhibitors were

tested by docking into the binding site 2. Those compounds will be synthesized and subjected to enzymatic testing.

Chapter 5

5. CONCLUSION

Fragment-based drug design is widely used in drug development in recent years by seeking candidate fragments for further optimization to design potent and efficient larger compounds. This technique could also be used in optimizing the existing inhibitors of the target by probing the fragments in the unoccupied region near the original binding site. The use of Computer-Aided Drug Design (CADD) accelerates this process by predicting their affinity and binding positions to the target virtually, which assists in prioritizing the compound candidates for future synthesis and testing.

In summary, in order to optimize current inhibitors of PLpro, a series of fragments from the Zenobia Library and Life Chemicals were screened in the presence of the previously discovered inhibitors. Since the lead inhibitors did not bind at the active site of PLpro, fragments which could take over the unoccupied catalytic triad (binding site 1) were expected. In the screening process, several fragments showing synergistic activity with the lead inhibitors were identified. The enzymatic analyses experiments (including mechanism of inhibition and mutual exclusivity studies) alone could not give structural information about the specific binding site and binding conformation of these fragments. Obtaining the structural information from crystallization could also be very challenging and time-consuming, due to the weak binding affinity of the fragments in many cases. The analysis of the synergistic inhibitors binding to the PLpro of Human SARS coronavirus presented here heavily relied on the enzymatic analysis and computational techniques.

To further obtain structural information about the fragments, computational studies, including molecular dynamic simulation, solvent mapping (hot spot analysis), docking and binding free energy calculations were employed. Snapshots of the co-crystal structure were extracted from 50ns

conventional simulation trajectories and sent to the FTMAP server. Two strong candidates for fragment binding sites were identified by analyzing the original structures and consensus clusters of used probes molecules. The two candidate sites identified included an extension of the original binding site of the lead inhibitors (binding site 2) and one cavity in the palm region (binding site 3). Another binding site (binding site 4) was discovered through coarse docking (**Figure 12**). There were four candidate binding sites identified. However, neither the solvent mapping nor docking could give information about the exact binding location for those fragments. Based on the hypothesis that there will be a favorable decrease of the lead inhibitors' binding free energy in combination with fragments, MM/PBSA calculations were used to compare the binding free energy of the lead inhibitors with the fragments placed in each binding site. Binding site 3 was excluded from further analysis since it was more than 10 Å away from the original binding site which is unlikely to affect the improved activity observed. From the MM/PBSA free energy calculation result (**TABLE IV-VII**), binding site 2 is most probable one for fragments binding. Furthermore, a series of compounds which were proposed by merging the fragment compound and the lead inhibitors were tested by docking into the binding site 2. Using the structural information revealed by these studies, we were able to apply a fragment-merging strategy to suggest several next generation compounds with high predicted binding affinity. Those compounds will be synthesized and subjected to the enzymatic testing.

The novel approach described here may be beneficial in other cases, where direct crystallization studies prove to be difficult. It also provides structural information about optimization of lead inhibitors, which could lead to novel scaffolds of PLpro inhibitors.

CITED LITERATURE

1. Zhao Z, Zhang F, Xu M, Huang K, Zhong W, Cai W, Yin Z, Huang S, Deng Z, Wei M, Xiong J, Hawkey PM. Description and clinical treatment of an early outbreak of severe acute respiratory syndrome (SARS) in Guangzhou, PR China. *J Med Microbiol*.52:715-20. 2003.
2. Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, Li PH, Tan SY, Chang Q, Xie JP, Liu XQ, Xu J, Li DX, Yuen KY, Peiris, Guan Y. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet*.362:1353-8. 2003.
3. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, Lui SF, Szeto CC, Chung S, Sung JJ. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med*.348:1986-94. 2003.
4. Peiris JS, Yuen KY, Osterhaus AD, Stohr K. The severe acute respiratory syndrome. *N Engl J Med*.349:2431-41. 2003.
5. Thiel V, Ivanov KA, Putics A, Hertzog T, Schelle B, Bayer S, Weissbrich B, Snijder EJ, Rabenau H, Doerr HW, Gorbalenya AE, Ziebuhr J. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol*.84:2305-15. 2003.
6. Stockman LJ, Bellamy R, Garner P. SARS: systematic review of treatment effects. *PLoS Med*.3:e343. 2006.
7. Frieman M, Baric R. Mechanisms of severe acute respiratory syndrome pathogenesis and innate immunomodulation. *Microbiol Mol Biol Rev*.72:672-85, Table of Contents. 2008.
8. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus. *Nat Med*.10:368-73. 2004.
9. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pohlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci U S A*.102:7988-93. 2005.
10. Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus. *Current Opinion in Microbiology*.7:412-9. 2004.
11. Ratia K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC, Mesecar AD. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc Natl Acad Sci U S A*.103:5717-22. 2006.
12. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Kraiden M,

Petric M, Skowronski DM, Upton C, Roper RL. The Genome sequence of the SARS-associated coronavirus. *Science*.300:1399-404. 2003.

13. Fouchier RA, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, Osterhaus AD. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A*.101:6212-6. 2004.

14. Lindner HA, 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*.79:15199-208. 2005.

15. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*.367:1814-20. 2012.

16. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ, Fouchier RA. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*.3. 2012.

17. Ratia K, Pegan S, Takayama J, Sleeman K, Coughlin M, Baliji S, Chaudhuri R, Fu W, Prabhakar BS, Johnson ME, Baker SC, Ghosh AK, Mesecar AD. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc Natl Acad Sci U S A*.105:16119-24. 2008.

18. Ghosh AK, Takayama J, Rao KV, Ratia K, Chaudhuri R, Mulhearn DC, Lee H, Nichols DB, Baliji S, Baker SC, Johnson ME, Mesecar AD. Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein-ligand X-ray structure and biological evaluation. *J Med Chem*.53:4968-79. 2010.

19. Ghosh AK, Takayama J, Aubin Y, Ratia K, Chaudhuri R, Baez Y, Sleeman K, Coughlin M, Nichols DB, Mulhearn DC, Prabhakar BS, Baker SC, Johnson ME, Mesecar AD. Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. *J Med Chem*.52:5228-40. 2009.

20. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S. FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res*.40:W271-5. 2012.

21. D.A. Case TAD, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H.Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. AMBER 11. 11 ed. San Francisco: University of California; 2010.

22. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*.25:1157-74. 2004.

23. Widge AS, Matsuoka Y, Kurnikova M. Development and initial testing of an empirical forcefield for simulation of poly(alkylthiophenes). *J Mol Graph Model*.27:34-44. 2008.

24. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*.65:712-25. 2006.

25. Zhou Z, Wang Y, Bryant SH. Computational analysis of the cathepsin B inhibitors activities through LR-MMPBSA binding affinity calculation based on docked complex. *J Comput Chem.*30:2165-75. 2009.
26. Zhou Z, Madura JD. Relative free energy of binding and binding mode calculations of HIV-1 RT inhibitors based on dock-MM-PB/GS. *Proteins.*57:493-503. 2004.
27. Prime. 3.0 ed. New York, NY: Schrödinger, LLC; 2011.
28. Glide. 5.7 ed. New York, NY: Schrödinger, LLC; 2011.
29. Ligprep. Schrödinger, LLC. 2.5 ed. New York, NY 2011.
30. Kumar A, Voet A, Zhang KY. Fragment based drug design: from experimental to computational approaches. *Curr Med Chem.*19:5128-47. 2012.
31. Osamu Ichihara JB, Richard J. Law, Mark Whittaker. Compound Design by Fragment-Linking. *Molecular Informatics.*30:298-306. 2011.

VITA

NAME: Shuyi Cao

EDUCATION: B.S., Pharmacy, Sichuan University, Chengdu, China, 2010

M.S., Medicinal Chemistry, University of Illinois at Chicago, Chicago, Illinois, 2013

PROFESSIONAL MEMBERSHIP: American Association of Pharmaceutical Scientists

POSTER PRESENTATION: 1. Shuyi Cao, Hyun Lee, Lena Truong, Kavankumar Patel, and Michael E. Johnson. Computational studies on the cooperative binding of fragment compounds in combination with known Papain-Like protease (PLpro) inhibitors. 2012 American Association of Pharmaceutical Scientists (AAPS) Conference, Chicago, IL

2. Shuyi Cao, Mahmud Tareq Hassan Khan, Michael E. Johnson. QM/MM Calculations to Study the Putative Mechanism of SAICAR Synthetase. 2012 Research Day, College of Pharmacy, University of Illinois at Chicago, IL

PUBLICATIONS: 1. Zhu T, Cao S, Su PC, Patel R, Shah D, Chokshi HB, et al. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. J Med Chem. 2013.

2. Lee H, Cao S, Hevener KE, Truong L, Gatuz JL, Patel K, et al. Synergistic Inhibitor Binding to the Papain-Like Protease of Human SARS Coronavirus: Mechanistic and Inhibitor Design Implications. ChemMedChem. 2013;8(8):1361-72.