

January 2015

Understanding the determinants for substrate recognition, regulation of enzymatic activity and the development of broad-spectrum inhibitors of coronavirus 3-chymotrypsin-like proteases

Sakshi Tomar
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Recommended Citation

Tomar, Sakshi, "Understanding the determinants for substrate recognition, regulation of enzymatic activity and the development of broad-spectrum inhibitors of coronavirus 3-chymotrypsin-like proteases" (2015). *Open Access Dissertations*. 1322.
https://docs.lib.purdue.edu/open_access_dissertations/1322

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Sakshi Tomar

Entitled

UNDERSTANDING THE DETERMINANTS FOR SUBSTRATE RECOGNITION, REGULATION OF ENZYMATIC
ACTIVITY AND THE DEVELOPMENT OF BROAD-SPECTRUM INHIBITORS OF CORONAVIRUS
3-CHYMOTRYPSIN-LIKE PROTEASES

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Carol B. Post

Chair

Cynthia V. Stauffacher

Andrew D. Mesecar

Angeline M. Lyon

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Andrew D. Mesecar

Approved by: Jeffery R. Lucas

Head of the Departmental Graduate Program

11/30/2015

Date

UNDERSTANDING THE DETERMINANTS FOR SUBSTRATE
RECOGNITION, REGULATION OF ENZYMATIC ACTIVITY AND THE
DEVELOPMENT OF BROAD-SPECTRUM INHIBITORS OF CORONAVIRUS
3-CHYMOTRYPSIN-LIKE PROTEASES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Sakshi Tomar

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

This dissertation is dedicated to my mom, Sunita Tomar, for her endless love and support.

ACKNOWLEDGMENTS

There are many people who provided constant support and encouragement during my time in the Mesecar lab. First and foremost, I would like to express my sincere gratitude to my advisor Professor Andrew Mesecar. He gave me guidance and encouragement when experiments failed and patted me on the back when I succeeded. Over the past five years, Professor Mesecar did not only help me grow as a scientist but gave me opportunities to be a mentor, a leader, a public speaker and a collaborator. I will forever be grateful for the time he spent sitting through my practice talks for departmental seminars or job interviews giving me suggestions for improvement. Special thanks also goes to Professor Mesecar's wife, Gail Mesecar. Andy and Gail hosted the best of parties. The warm welcome that Gail always extended at her house made me feel at home while being thousands of miles away from home. I would also like to thank my committee members, Dr. Cynthia Stauffacher, Dr. Carol Post, Dr. Joe Kappock and Dr. Angeline Lyon for their incredible supervision and guidance.

I would like to thank our collaborators Dr. Shaun Stauffer, Dr. Arun Ghosh, Dr. Mark Denison and Dr. Susan Baker for their expertise and ideas that lend me several publications in the top peer-reviewed journals.

I would like to acknowledge current and past Mesecar lab members for being wonderful colleagues and their constant help and support. I consider myself fortunate to have been a part of the lab that has only made me love science more than even before, to say the least. I would like to thank Dr. Aimee Egger who mentored me during my rotation in the Mesecar lab and again when I officially joined the Mesecar lab. I appreciate her for the mentorship she provided, her patience with answering my non-ending list of questions and her encouragement when I doubted myself. She is still the person I look up to as a scientist, a colleague, a friend and a mentor. I would also like to extend my thanks to my lab mates Dr. Yahira Baez Santos, Dr.

Katherine Jensen, Dr. Sergey Savinov, Dr. Soma Mukhopadhyay, Dr. Sarah St. John, Yafang Chen, Julia Luciano, Qing Zhou, Nicole Hjortland, Yu-Chen Yen, Aya Ryuzoji, Xin Wen, Brandon Andon, Jozlyn Clasman, Corey Moore and Dr. Renatta Everett. They all have made my graduate school experience wonderful. I would also like to thank the exceptional undergraduate students, Rebecca Tweedell and Melanie Johnston, whom I had the chance to train and work with.

I would like to thank the friends I made at Purdue who I laughed with, shared minutiae of everyday life with and celebrated several accomplishments with, including Mani, Shiv, Uma, Yafang, Qing, Duy, Satchal, Swapna, Ranjan, Anwesha and Harsh.

Finally, thanks is very small of a word for the unconditional love and support that my family, esp. my mom, and my best friends Vandita Rana and Rimpi Rajput provided me. They endured my frustrations when days were gloomy, laughed with me to cheer me up, made me believe in myself when in self-doubt and celebrated my success. I am truly blessed to have such wonderful friends and family.

Finally, I always carry a fighting spirit inside me and believe in what Paulo Coehlo said in *The Alchemist*-

“When a person really desires something, all the universe conspires to help that person to realize his dream.”

TABLE OF CONTENTS

	Page
LIST OF TABLES	xii
LIST OF FIGURES	xiv
ABBREVIATIONS	xviii
ABSTRACT	xx
CHAPTER 1. INTRODUCTION	1
1.1 Taxonomy and classification of coronaviruses	1
1.2 Human coronaviruses	2
1.3 Animal coronaviruses	4
1.4 Zoonotic transmission of coronaviruses	5
1.5 Coronavirus lifecycle	7
1.5.1 Coronavirus proteases as drug targets	10
1.6 Coronavirus 3CL ^{pro} protease	12
1.6.1 Enzymatic reaction mechanism for coronavirus 3CL ^{pro}	12
1.6.2 3CL ^{pro} is catalytically active as a dimer	14
1.6.3 Model for auto-release of SARS-CoV 3CL ^{pro} from the polypro- tein	17
1.7 Statement of intent	17
CHAPTER 2. EXPERIMENTAL PROCEDURES	20
2.1 Recipes for expression and purification of 3CL ^{pro} enzymes	20
2.1.1 Recipe for 1 liter LB medium	20
2.1.2 Recipe for 1 liter Super-LB medium for auto-induction	20
2.1.3 Buffers for purification of 3CL ^{pro} enzymes	21
2.2 Expression and purification of SARS-CoV 3CL ^{pro}	26
2.2.1 DEAE anion-exchange chromatography	27

	Page	
2.2.2	Hydrophobic-interaction chromatography	28
2.2.3	Mono-Q anion-exchange chromatography	28
2.2.4	Gel-filtration chromatography	29
2.3	Expression and purification of MHV 3CL ^{pro}	31
2.3.1	Hydrophobic-interaction chromatography	32
2.3.2	DEAE anion-exchange chromatography	32
2.3.3	Gel-filtration chromatography	33
2.4	Expression and purification of HKU1-CoV 3CL ^{pro}	33
2.4.1	Hydrophobic-interaction chromatography	35
2.4.2	DEAE anion-exchange chromatography	36
2.4.3	Gel-filtration chromatography	36
2.5	Expression and purification of OC43-CoV 3CL ^{pro}	37
2.5.1	Hydrophobic-interaction chromatography	39
2.5.2	DEAE anion-exchange chromatography	39
2.5.3	Gel-filtration chromatography	40
2.6	Expression and purification of HKU5-CoV 3CL ^{pro}	40
2.6.1	Hydrophobic-interaction chromatography	43
2.6.2	DEAE anion-exchange chromatography	43
2.6.3	Mono-Q anion-exchange chromatography	44
2.7	Expression and purification of MERS-CoV 3CL ^{pro}	46
2.8	Sample preparation for SDS-PAGE analysis	46
2.9	Concentration determination	47
2.9.1	Mini-bradford assay	47
2.9.2	OD ₂₈₀ based determination of protein concentration	48
2.10	Kinetic assays	48
2.10.1	Determination of fluorescence extinction coefficient for the substrate UIVT3	49
2.10.2	Determination of enzymatic efficiency	50

	Page
2.11 Inhibition assays	51
2.11.1 % inhibition determination	52
2.11.2 IC ₅₀ determination	52
2.12 Crystallization of SARS-CoV 3CL ^{PRO} in complex with inhibitors . .	54
2.13 Data collection and structure refinement	54
2.13.1 Procedure	54
2.13.2 Results	56
2.14 Summary	56
CHAPTER 3. SARS-COV 3CL ^{PRO} INHIBITOR DEVELOPMENT	63
3.1 Different classes of SARS-CoV 3CL ^{PRO} inhibitors	63
3.1.1 Natural product inhibitors	63
3.1.2 Non-peptidic inhibitors	65
3.1.3 Metal-conjugated inhibitors	65
3.1.4 Bifunctional boronic acid inhibitors	67
3.1.5 C ₂ -symmetric diol inhibitors	67
3.1.6 Anilide inhibitors	67
3.1.7 Inhibitors with ‘warhead’ functionalities	68
3.2 Development of non-covalent reversible inhibitors of SARS-CoV 3CL ^{PRO}	71
3.3 Results	73
3.3.1 Furanyl amide scaffold - SAR	73
3.3.2 Benzotriazole scaffold - SAR	84
3.4 Identification of inhibitors with broad-spectrum activity against 3CL ^{PRO} from different coronaviruses	92
3.5 X-ray crystal structure determination of SARS-CoV 3CL ^{PRO} -inhibitor complexes	96
3.5.1 X-ray crystal structures in complex with P ₃ -truncated benzo- triazole inhibitors	96
3.5.2 X-ray crystal structures in complex with furanyl-amide inhibitors	97
3.6 Summary	103

	Page
CHAPTER 4. MECHANISTIC BASIS FOR THE TEMPERATURE SENSITIVITY OF 3CL ^{PRO} MUTANTS OF MURINE HEPATITIS VIRUS (MHV)	104
4.1 Abstract	104
4.2 Introduction	105
4.3 Materials and methods	107
4.3.1 Construct design and expression of MHV 3CL ^{PRO}	107
4.3.2 Protein purification	107
4.3.3 Thermal inactivation at permissive and non-permissive temperatures	107
4.3.4 Determining secondary structural changes at permissive and non-permissive temperatures	109
4.3.5 Determination of melting temperature, T_m , using circular dichroism	109
4.3.6 Homology Modeling	110
4.3.7 Molecular Dynamics Simulations	110
4.3.8 CNAalysis	111
4.3.9 Mutual Information Analysis	111
4.4 Results	112
4.4.1 Protein expression and purification	112
4.4.2 The H134Y mutation rescues the <i>-ts</i> V148A mutant from thermal inactivation at non-permissive temperature	112
4.4.3 Loss in secondary structure accompanies thermal inactivation of V148A	113
4.4.4 The difference in the melting temperatures suggests global destabilization of V148A	118
4.4.5 Distant residues V148 and H134 are connected through hydrogen bonding and hydrophobic networks	118
4.4.6 Significance of mutation of surface residue histidine to more hydrophobic tyrosine at position 134	121
4.4.7 Constraint Network Analysis predicts residue 148 as a weak spot	121

	Page	
4.4.8	Distinct residue fluctuations for amino acids 45-70 in the mutants and the wildtype	123
4.4.9	Mutual Information Analysis	124
4.4.10	Solvent Accessible Surface Area	124
4.5	Discussion	128
4.5.1	Position 148 is a structural ‘weak spot’	129
4.5.2	Y134 acts as a a general facilitator for 3CL ^{PRO} stabilization .	130
4.5.3	Implications for the prediction of viral <i>-ts</i> and corresponding suppressor mutations in the design of viral vaccines	132
CHAPTER 5. LIGAND-INDUCED DIMERIZATION REGULATES MERS-COV 3CL ^{PRO}		134
5.1	Abstract	134
5.2	Introduction	135
5.3	Experimental Procedures	137
5.3.1	Construct design and expression of MERS-CoV 3CL ^{PRO} . . .	137
5.3.2	MERS-CoV 3CL ^{PRO} purification	137
5.3.3	Western-blot analysis	141
5.3.4	MALDI mass-spectral analysis	141
5.3.5	Synthesis of compounds <i>1-11</i>	142
5.3.6	Fluorescence-based kinetic assays	142
5.3.7	Determination of enzymatic efficiency	143
5.3.8	Influence of dimerization on the activity of 3CL ^{PRO} enzymes .	143
5.3.9	Inhibition assays	144
5.3.10	Analytical ultracentrifugation (AUC) analysis	146
5.3.11	SEC-MALS analysis	146
5.3.12	MERS-CoV 3CL ^{PRO} activation and inhibition by a non-covalent inhibitor	147
5.3.13	MERS-CoV 3CL ^{PRO} Crystallization, X-ray Data collection and Structure determination	148
5.4	Results	149

	Page	
5.4.1	Production of MERS-CoV 3CL ^{pro} with authentic N- and C-termini	149
5.4.2	MERS-CoV 3CL ^{pro} hydrolyzes a fluorescent peptide substrate with lower efficiency than other 3CL ^{pro} enzymes	151
5.4.3	MERS-CoV 3CL ^{pro} is a weakly associated dimer	153
5.4.4	MERS-CoV 3CL ^{pro} inhibition by designed peptidomimetic compounds	155
5.4.5	Weak association of the MERS-CoV 3CL ^{pro} dimer is supported by AUC and SEC-MALS studies	158
5.4.6	MERS-CoV 3CL ^{pro} undergoes extensive ligand-induced dimerization	161
5.4.7	MERS-CoV 3CL ^{pro} is activated by ligand-induced dimerization	162
5.4.8	X-ray structure of MERS-CoV 3CL ^{pro} -compound <i>6</i> complex	165
5.4.9	MERS-CoV 3CL ^{pro} has a smaller S_2 pocket than SARS-CoV 3CL ^{pro}	168
5.4.10	X-ray structure of MERS-CoV 3CL ^{pro} in complex with a non-covalent inhibitor	169
5.4.11	Interactions at the 3CL ^{pro} dimer interface	170
5.4.12	Analysis of non-conserved residues of MERS-CoV 3CL ^{pro} . .	173
5.5	Discussion	175
5.5.1	A Model for regulation of the enzymatic activity of MERS-CoV 3CL ^{pro} during polyprotein processing	175
5.5.2	Non-conserved amino acids of MERS-CoV 3CL ^{pro} may regulate the dimer formation	178
5.5.3	Development of 3CL ^{pro} inhibitors with broad-spectrum specificity	180
5.5.4	Potential complexity in the development of MERS-CoV 3CL ^{pro} inhibitors as antiviral agents	181
CHAPTER 6. CHARACTERIZATION OF DRUG-RESISTANT Y22C MUTANT OF HKU1 3CL ^{PRO}		183
6.1	Material and Methods	183
6.1.1	<i>in vitro</i> enzymatic assay and IC ₅₀ determination	183

	Page
6.1.2 Enzyme reactivation assay	184
6.1.3 Molecular Dynamics Simulations and Analysis	185
6.2 Results	186
6.2.1 Mechanism for drug-resistance of Y22C mutant of HKU1 3CL ^{pro}	186
6.3 Discussion	189
CHAPTER 7. SUMMARY	192
REFERENCES	194
VITA	207

LIST OF TABLES

Table	Page
1.1 Representative coronaviruses and their cellular receptors	9
2.1 Purification summary of SARS-CoV 3CL ^{pro} from 2 L culture of <i>E.coli</i> BL21-DE3	29
2.2 Purification summary of MHV 3CL ^{pro} from 3 L culture of <i>E.coli</i> BL21-DE3	33
2.3 Purification summary of HKU1 3CL ^{pro} from 1 L culture of <i>E.coli</i> BL21-DE3	37
2.4 Purification summary of OC43 3CL ^{pro} from 1 L culture of <i>E.coli</i> BL21-DE3	42
2.5 Purification summary of HKU5-CoV 3CL ^{pro} from 1 L culture of <i>E.coli</i> BL21-DE3	44
2.6 Molar extinction coefficient values for different 3CL ^{pro} enzymes calculated using Expasy Protparam tool	49
2.7 Enzymatic efficiency (apparent k_{cat}/K_M) for different 3CL ^{pro} enzymes .	51
2.8 Crystallization conditions for SARS-CoV 3CL ^{pro} -Inhibitor complexes .	55
2.9 X-ray data collection and refinement statistics for SARS-CoV 3CL ^{pro} in complex with CQ3 and R30	57
2.10 X-ray data collection and refinement statistics for SARS-CoV 3CL ^{pro} in complex with R2Y and R2X	58
2.11 X-ray data collection and refinement statistics for SARS-CoV 3CL ^{pro} in complex with 886 and CUG	59
2.12 X-ray data collection and refinement statistics for SARS-CoV 3CL ^{pro} in complex with FMW and KWW	60
2.13 X-ray data collection and refinement statistics for SARS-CoV 3CL ^{pro} in complex with XM2	61
4.1 Purification summary of MHV 3CL ^{pro} from 3 L culture of <i>E.coli</i> BL21-DE3	112
4.2 Thermal inactivation of MHV 3CL ^{pro} wildtype and mutants upon incubation at 30 °C and 40 °C	115

Table	Page
4.3 Loss of secondary structure and T_m determination through measurement of CD signal at 222 nm	116
4.4 Unfolding nuclei predicted by CNA webserver	122
4.5 SASA of various residues and groups of residues	127
5.1 Purification summary of MERS-CoV 3CL ^{pro} per liter of <i>E.coli</i> BL21-DE3	151
5.2 Comparison of the apparent turnover number, k_{cat} , and the monomer-dimer dissociation constant, K_d , for different 3CL ^{pro} enzymes	153
5.3 Chemical structures and inhibitory activity of compounds 1 to 11 against MERS-CoV 3CL ^{pro}	157
5.4 X-ray data collection and refinement statistics	166

LIST OF FIGURES

Figure	Page
1.1 Classification and taxonomy of coronaviruses	2
1.2 Phylogeny and host tropism of representative coronaviruses	4
1.3 Coronavirus virion structure and proteins	7
1.4 Coronavirus life cycle	8
1.5 Coronavirus genome and non-structural proteins	11
1.6 Detailed dimeric structure of SARS-CoV 3CL ^{pro}	13
1.7 Enzymatic reaction mechanism for SARS-CoV 3CL ^{pro}	15
1.8 Proposed model for auto-release of SARS-CoV 3CL ^{pro} form the polyprotein	18
2.1 Construct design for 3CL ^{pro} enzymes	27
2.2 Purification of SARS-CoV 3CL ^{pro}	30
2.3 Purification of MHV 3CL ^{pro}	34
2.4 Purification of HKU1 3CL ^{pro}	38
2.5 Purification of OC43 3CL ^{pro}	41
2.6 Purification of HKU5-CoV 3CL ^{pro}	45
2.7 Formation of 3CL ^{pro} high-order aggregates on SDS-PAGE	47
2.8 Representative dose response curve for IC ₅₀ determination	53
3.1 Inhibitory activity of natural product and non-peptidic inhibitors of SARS-CoV 3CL ^{pro}	64
3.2 Inhibitory activity of metal-conjugated, boronic-acid based, C ₂ -symmetric diol and anilide inhibitors of SARS-CoV 3CL ^{pro}	66
3.3 Inhibitory activity of compounds with reactive ‘warhead’ functionalities	68
3.4 Chemical scaffolds identified through high-throughput screening (HTS) for SARS-CoV 3CL ^{pro} inhibitor development	72
3.5 Furanyl amide scaffold - SAR of varying functional groups in S ₁ subsite of SARS-CoV 3CL ^{pro}	74

Figure	Page
3.6 Furanyl amide scaffold - SAR of varying functional groups in S_1' subsite of SARS-CoV 3CL ^{pro}	77
3.7 Furanyl amide scaffold - SAR of varying functional groups in S_3 subsite of SARS-CoV 3CL ^{pro}	79
3.8 Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL ^{pro}	80
3.9 Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL ^{pro} <i>cont.</i>	82
3.10 Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL ^{pro} <i>cont.</i>	83
3.11 Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL ^{pro} <i>cont.</i>	84
3.12 Binding mode of benzotriazole scaffold in the active site of SARS-CoV 3CL ^{pro}	85
3.13 Inactive analogues of benzotriazole scaffold	86
3.14 P_3 -extended benzotriazole scaffold - SAR of varying functional groups in S_1' - S_2 subsite of SARS-CoV 3CL ^{pro}	87
3.15 Inhibitory activity of P_3 -extended (R30) versus P_3 -truncated (XTF) benzotriazole analogues	88
3.16 P_3 -truncated benzotriazole scaffold - SAR of varying functional groups in S_1' - S_2 subsite of SARS-CoV 3CL ^{pro}	90
3.17 P_3 -truncated benzotriazole scaffold - SAR of varying functional groups in S_1' - S_2 subsite of SARS-CoV 3CL ^{pro} <i>cont.</i>	91
3.18 Chemical structures of broad spectrum 3CL ^{pro} inhibitors	94
3.19 Broad spectrum 3CL ^{pro} inhibitors	95
3.20 X-ray crystal structure of SARS-CoV 3CL ^{pro} in complex with P_3 -truncated benzotriazole inhibitors CQ3 and R30	97
3.21 X-ray crystal structure of SARS-CoV 3CL ^{pro} in complex with P_3 -truncated benzotriazole inhibitors R2Y and R2X	98
3.22 Superposition of inhibitors CQ3, R30, R2Y and R2X from the X-ray crystal structures of SARS-CoV 3CL ^{pro} in complex with inhibitors	99
3.23 X-ray crystal structure of SARS-CoV 3CL ^{pro} in complex with furanyl-amide inhibitors 886, CUG, FMW, KWW and XM2.	100

Figure	Page
3.24 Superposition of inhibitors REY, 886, CUG, FMW, KWW and XM2 from the X-ray crystal structures of SARS-CoV 3CL ^{pro} in complex with inhibitors	101
3.25 SARS 3CL ^{pro} -XM2 complex crystal structure	102
4.1 Thermal inactivation of MHV 3CL ^{pro} wildtype and the mutants	114
4.2 Protein unfolding characterized through circular dichroism	117
4.3 Network of residues connecting positions 148 and 134 in MHV 3CL ^{pro}	119
4.4 Structural weak spots in MHV 3CL ^{pro}	122
4.5 RMSF fluctuations in the alpha carbons of MHV 3CL ^{pro}	125
4.6 Distance between the centers of mass of the catalytic dyad	126
5.1 Purification of MERS-CoV 3CL ^{pro}	140
5.2 Production of MERS-CoV 3CL ^{pro} with authentic N- and C-termini	150
5.3 Comparison of enzymatic efficiencies (k_{cat}/K_M) of 3CL ^{pro} enzymes from different CoVs	152
5.4 Dependence of the enzymatic activity of MERS-CoV, HKU4-CoV, HKU5-CoV and SARS-CoV 3CL ^{pro} s on the total enzyme concentration	154
5.5 Progress curves for the MERS-CoV 3CL ^{pro} catalyzed reaction in the presence of compound 6	156
5.6 AUC-SV raw boundary profiles for MERS-CoV 3CL ^{pro}	159
5.7 AUC-SV and SEC-MALS analysis of MERS-CoV 3CL ^{pro}	160
5.8 Activation of MERS-CoV 3CL ^{pro} via ligand-induced dimerization	163
5.9 X-ray crystal structure of MERS-CoV 3CL ^{pro} in complex with inhibitors	167
5.10 Comparison of X-ray crystal structures of 3CL ^{pro} dimers from MERS-CoV, HKU4-CoV and SARS-CoV	171
5.11 Sequence alignment of 3CL ^{pro} enzymes from MERS-CoV, HKU5-CoV, HKU4-CoV and SARS-CoV	172
5.12 Analysis of the non-conserved amino acids of MERS-CoV 3CL ^{pro}	174
5.13 Proposed model for polyprotein processing in MERS-CoV regulated by ligand-induced dimerization of MERS-CoV 3CL ^{pro}	176
6.1 Mechanism for drug-resistance of Y22C mutant of HKU1 3CL ^{pro}	188

Figure	Page
6.2 Reaction mechanism for hydrolysis of the ester-based inhibitor CE-5 by HKU1 3CL ^{pro}	189
6.3 Molecular dynamics simulation of HKU1 3CL ^{pro} wild-type and Y22C mutant	190

ABBREVIATIONS

Å	Angstrom
BSA	bovine serum albumin
C	celsius
DEAE	diethylaminoethyl
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetracarboxylic acid
FPLC	fast pressure liquid chromatography
IC ₅₀	concentration to inhibit 50% enzymatic activity
IPTG	isopropyl β -D-1-thiogalactopyranoside
LRL-CAT	Lilly Research Laboratories Collaborative Access Team
LS-CAT	Life Sciences Collaborative Access Team
mAU	milli absorbance units
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	messenger RNA
nM	nanomolar
nm	nanometer
OD	optical density
PDB	protein data bank
rpm	revolutions per minute
rmsd	root mean square deviation
rmsf	root mean square fluctuation
SDS	sodium dodecyl sulfate

SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TBS	Tis-buffered saline
TBST	Tis-buffered saline with Tween20
UV	ultraviolet
V	volts
WHO	world health organization
μM	micromolar

ABSTRACT

Tomar, Sakshi PhD, Purdue University, December 2015. Understanding the determinants for substrate recognition, regulation of enzymatic activity and the development of broad-spectrum inhibitors of coronavirus 3-chymotrypsin-like proteases. Major Professor: Andrew D. Mesecar.

Coronaviruses include lethal human pathogens like severe acute respiratory syndrome coronavirus (SARS-CoV) and the recently emerged Middle-east respiratory coronavirus (MERS-CoV). Coronavirus also impact global economy by infecting farm animals like pigs (porcine epidemic diarrhea virus, PEDV), cows (bovine coronavirus, BCoV) and poultry (avian infectious bronchitis virus, IBV). Moreover, the global distribution of bats that naturally harbor one or more coronavirus strains heightens the likelihood of emergence of a novel highly pathogenic coronavirus in the near future. To combat infections of existing and emerging coronaviruses, it is important to identify coronavirus drug targets that can be utilized for the development of broad-spectrum anti-coronaviral therapeutics. Viral encoded 3-Chymotrypsin-like protease (3CL^{pro}) is essential for viral polyprotein processing to release non-structural proteins that form the replicase complex machinery for viral genome replication. Due to its indispensable role in coronaviral replication, 3CL^{pro} is an attractive drug target. Moreover, high sequence conservation in the vicinity of active site among 3CL^{pro} proteases from different coronavirus subclasses make them an excellent target for the development of broad-spectrum therapeutics for coronavirus infections. The overarching goal of this project is to investigate enzymatic and structural properties of multiple 3CL^{pro} enzymes encompassing different coronavirus subclasses. Understanding the determinants of structural and functional disparity between different 3CL^{pro} enzymes and the factors regulating these properties will aid in the design of broad-spectrum inhibitors of 3CL^{pro} enzymes.

Here, the successful expression and purification of six different coronavirus 3CL^{pro} enzymes is reported. These 3CL^{pro} enzymes include four 3CL^{pro} enzymes from human-coronaviruses SARS, MERS, OC43 and HKU1; 3CL^{pro} from bat coronavirus HKU5, which is closely related to MERS-CoV, and 3CL^{pro} from murine hepatitis virus (MHV), which serves as a good model system to study virus replication in cell culture. Successful expression of these enzymes was achieved in *E.coli* BL21-DE3 cells using codon-optimized 3CL^{pro} genes that were cloned into pET-11a expression vectors. During construct design, either the TEV-protease cleavage site or the 3CL^{pro} auto-cleavage site was inserted between the N-terminus His₆-tag and the 3CL^{pro} gene, resulting in the expression of 3CL^{pro} enzymes with authentic N-termini. The auto-cleavage of His₆ affinity-tag during protein expression in *E.coli* cells prohibits the use of immobilized metal ion chromatography (IMAC) for protein purification. Therefore, a native step purification protocol involving 2-4 sequential steps including different combinations of hydrophobic-interaction, ion-exchange and gel-filtration chromatographies was utilized to obtain highly pure 3CL^{pro} enzymes for kinetic and structural studies.

For kinetic studies, a FRET-based peptide substrate was used that was custom synthesized by our lab based on the auto-cleavage sequence of multiple 3CL^{pro} enzymes. Due to the inner-filter effect, studies were limited to use this substrate only at low concentrations ($<4 \mu\text{M}$). Due to issues associated with inner-filter effect, combined with high sub-mM K_M values that have been reported, only the apparent values of k_{cat}/K_M could be determined for 3CL^{pro} enzymes. Under the assay conditions utilized, it was determined that MHV 3CL^{pro} and HKU1 3CL^{pro} enzymes were the most efficient enzymes, while 3CL^{pro} from MERS-CoV was the least efficient enzyme. Enzymatic efficiency of MERS-CoV 3CL^{pro} was 10-fold lower than the enzymatic activity of 3CL^{pro} enzymes from MHV and HKU1, and 5-fold lower than that of SARS-CoV 3CL^{pro}. Therefore, the factors that might cause a decrease in the enzymatic activity of MERS-CoV 3CL^{pro} were further explored.

Since dimerization has previously been shown to regulate the enzymatic activity of SARS-CoV 3CL^{pro}, it was hypothesized that the lower enzymatic efficiency of MERS-CoV 3CL^{pro} stems from a weaker dimer formation in MERS-CoV 3CL^{pro}. Interestingly, further kinetic and biophysical investigations revealed that the monomer-dimer dissociation constant K_d for MERS-CoV 3CL^{pro} is significantly higher (high μM) compared to K_d values determined for 3CL^{pro} enzymes from SARS-CoV and closely related HKU4 and HKU5 coronaviruses (low nM). These results suggested that lower enzymatic efficiency of MERS-CoV 3CL^{pro} stems from a weaker dimer formation in MERS-CoV 3CL^{pro}. Weaker dimer formation of MERS-CoV 3CL^{pro} would implicate that inside a virus-infected cell either higher expression levels of 3CL^{pro} will need to be achieved to form an active dimer and/or a mechanism like ligand-induced dimerization must exist to promote dimer formation even at low expression levels. Therefore, whether MERS-CoV 3CL^{pro} is capable of undergoing ligand-induced dimerization was determined. For these experiments, a substrate-mimetic peptide covalent inhibitor of MERS-CoV 3CL^{pro} was utilized. Analytical ultracentrifugation (AUC) studies of MERS-CoV 3CL^{pro} performed in the presence of different stoichiometric ratios of the covalent inhibitor revealed a significantly higher population of dimeric species compared to MERS-CoV 3CL^{pro} analyzed in the absence of any ligand. Similar results were obtained in the presence of a small, non-covalent substrate-mimetic compound. In fact, measuring the activity levels of MERS-CoV 3CL^{pro} in the presence of varying concentrations of the non-covalent, substrate-mimetic compound demonstrated that the enzyme is activated under low compound concentrations, and that the compound displays any inhibitory activity only at higher concentrations. Experiments performed at three different concentrations of MERS-CoV 3CL^{pro} yielded similar results suggesting that activation via ligand-induced dimerization regulates the activity of MERS-CoV 3CL^{pro} during virus replication.

To determine the structural basis for weak dimer formation and ligand-induced dimerization of MERS-CoV 3CL^{pro}, crystallization of MERS-CoV 3CL^{pro} both in the apo form and in complex with the substrate-mimetic compounds was attempted.

Although, the crystallization attempts for the apo form of MERS-CoV 3CL^{pro} were unsuccessful, high-resolution diffraction quality crystals of MERS-CoV 3CL^{pro} in complex with the covalent and non-covalent substrate-mimetic compounds were obtained. MERS-CoV 3CL^{pro}-compound complexes crystallized as dimers. Detailed analysis and comparison of the MERS-CoV 3CL^{pro} dimer structure with the dimer structures of SARS-CoV 3CL^{pro} and HKU4-CoV 3CL^{pro} did not reveal any differences that would clearly suggest the structural basis for weaker dimer formation or ligand-induced dimerization of MERS-CoV 3CL^{pro}. Furthermore, based on sequence alignment of MERS-CoV 3CL^{pro} with closely related HKU4-CoV and HKU5-CoV 3CL^{pro} enzymes, amino acids that participate in dimer formation are conserved in all the three enzymes. Based on these observations, it is speculated that the non-conserved residues of MERS-CoV 3CL^{pro} that are distant from the dimer interface may regulate dimer formation via long-range interactions. Further mutagenesis studies would be required to identify residues that regulate dimer formation specifically in MERS-CoV 3CL^{pro}.

Besides regulating 3CL^{pro} activity via influencing the dimer formation, long-range interactions can regulate the enzymatic activity of 3CL^{pro} enzymes by impacting their structural stability. Our collaborators identified mutants of MHV 3CL^{pro}, V148A, that display a temperature-sensitive (*-ts*) growth phenotype. Surprisingly, the second-site compensatory mutation (V148A/H134Y) that emerged would be structurally distant ($>15 \text{ \AA}$) from the original temperature-sensitive mutation. It was suggested that these distant mutations communicate through long-range interactions in MHV 3CL^{pro}. To determine the exact mechanism for *-ts* and compensatory phenotypes, this study reports the characterization of thermal stability of these mutants. The mutant proteins were purified using a protocol similar to the purification protocol for the wildtype MHV 3CL^{pro}. It was demonstrated that the *-ts* V148A mutant undergoes rapid loss in activity after incubation at 40 °C, while both the single (H134Y) and double (V148A/H134Y) second site compensatory mutants maintain their enzymatic activity for two hours at 40 °C. Determination of melting temperatures (T_m)

by measuring the circular dichroism signal at 222 nm as a function of temperature over the range of 10 °C to 90 °C demonstrated that the *-ts* V148A mutant undergoes faster global structural destabilization compared to the wildtype or H134Y and V148A/H134Y enzymes. Analysis of a structural model of the wildtype MHV 3CL^{pro} enzymes reveals that the amino acid at position 148 would be coupled to amino acid at position 134 through a network of hydrogen-bonding and hydrophobic interactions. It is speculated that the *-ts* and compensatory mutations influence the structural stability of MHV 3CL^{pro} by altering this hydrogen-bonding and hydrophobic network. Based on sequence analysis of other 3CL^{pro} enzymes, it is further suggested that amino acid Y134 acts as a general facilitator of structural stabilization for 3CL^{pro} enzymes.

Long-range interactions do not only influence the dimer formation and structural stability of 3CL^{pro} enzymes as discussed in previous sections, they can have a direct impact on substrate binding and catalysis. This study also reports the characterization of a physiologically relevant drug-resistant mutation (Y22C) that is distant from the substrate binding site in HKU1 3CL^{pro}. Biochemical and kinetic analysis reveals that despite being structurally distant from the active site, the Y22C mutation directly impacts the binding and hydrolysis of a reactive ester inhibitor.

All these results suggest that long-range interactions regulate 3CL^{pro} activity through multiple mechanisms. Long-range interactions can modulate 3CL^{pro} enzymatic activity either by influencing dimer formation, structural stability or directly altering substrate binding and catalysis. Impact of long-range interactions on the structure and function of 3CL^{pro} enzymes is not well understood and studies reported in this dissertation provide few of the initial models for mechanisms utilized by long-range interactions to regulate the function of 3CL^{pro} enzymes.

Finally, one of the goals of this project is to identify broad-spectrum inhibitors of 3CL^{pro} enzymes. This study reports the identification of low-micromolar and sub-micromolar non-covalent inhibitors of SARS-CoV 3CL^{pro}. These compounds were synthesized by our collaborators as part of a focused library of 237 compounds that

were developed based on the initial hits from a high-throughput screen. Several compounds that were able to target multiple 3CL^{pro} enzymes were also identified. Structure activity relationships (SARs) of these compounds are being established in conjunction with the determination of high-resolution crystal structures of SARS-CoV 3CL^{pro} in complex with multiple inhibitors from different chemical scaffolds. Structural analysis reveals that most of the direct interactions formed between inhibitor molecules and SARS-CoV 3CL^{pro} involve either the peptide backbone or the side chains of amino acids that are conserved across all 3CL^{pro} enzymes. Engagement of conserved structural features of SARS-CoV 3CL^{pro} by these inhibitors open avenues for the development of more potent second-generation inhibitors that can inhibit multiple 3CL^{pro} enzymes.

CHAPTER 1. INTRODUCTION

1.1 Taxonomy and classification of coronaviruses

Coronaviruses are enveloped (+) single-stranded (ss) RNA viruses. Genus “coronavirus” of *Coronaviridae* family belongs to the order *Nidovirales* (Figure 1.1). Members in the order *Nidovirales*, including coronaviruses, are characterized by the synthesis of 3’ co-terminal nested set of subgenomic viral mRNAs (sg mRNA) during infection [1]. The name “coronavirus” is derived from the Latin word *corona*, which means crown and refers to the crown-like morphology of the virions as seen under electron microscopy. This morphology is provided by the surface spike protein S of the virus and is important for interaction with host receptors.

Based on phylogenetic analysis, coronaviruses can be further classified as alphacoronavirus, betacoronavirus, gammacoronavirus and deltacoronavirus (Figure 1.2) [2]. Designation alpha, beta and gammacoronavirus replaced the former classification into groups 1, 2 and 3, respectively. Lineage 1a of alphacoronaviruses includes feline coronavirus (FCoV), feline infectious peritonitis virus (FIPV) and pig TGEV, while lineage 1b encompasses human coronaviruses NL63 and 229E, porcine epidemic diarrhea virus (PEDV) and several species of bat coronaviruses. Human coronaviruses HCoV-HKU1, HCoV-OC43 belong to lineage 2a of betacoronaviruses along with murine hepatitis virus (MHV) and bovine coronavirus BCoV, lineage 2b of betacoronaviruses includes SARS-CoV and several strains of SARS-like coronaviruses from bats (BtSARS), MERS-CoV and Bat coronaviruses HKU4 and HKU5 cluster in lineage 2c, and several strains of bat coronavirus HKU9 constitute lineage 2d of betacoronaviruses. Gammacoronaviruses include avian infectious bronchitis virus (IBV) and primarily infect birds. Deltacoronavirus is a newly characterized class and include several bird coronaviruses and strains of porcine coronavirus HKU15.

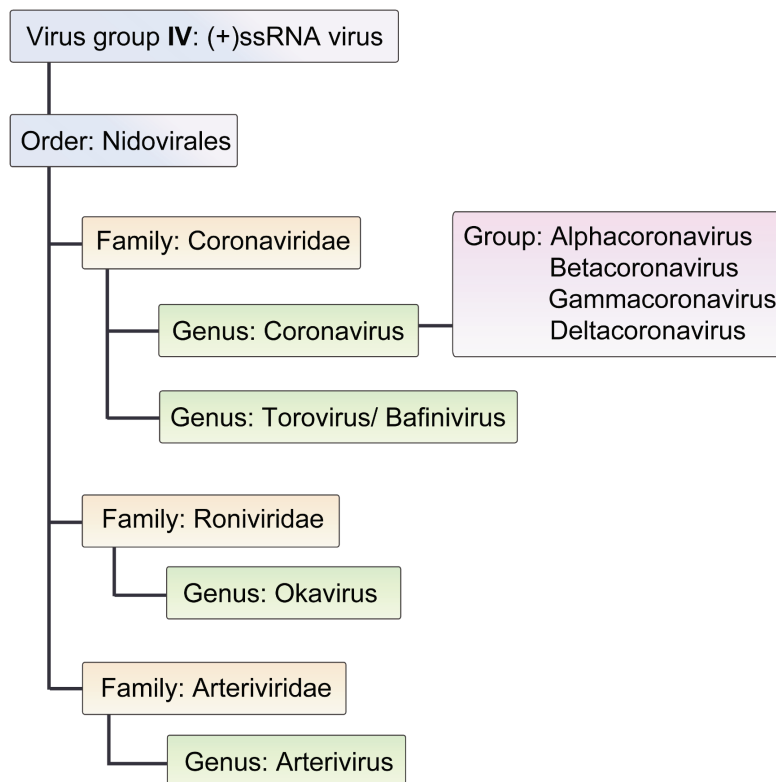


Fig. 1.1.: **Classification and taxonomy of coronaviruses.** According to Baltimore classification, coronaviruses belong to Group IV that encompasses all the viruses containing positive-sense single-stranded RNA genomes. Family *Coronaviridae*, along with *Roniviridae* and *Arteriviridae*, constitute the order *Nidovirales*. Based on genome sequencing, coronaviruses can further be classified as alpha, beta, gamma and deltacoronaviruses.

1.2 Human coronaviruses

Coronaviruses gained prominence in 2003 with the emergence of the global severe acute respiratory syndrome (SARS) epidemic. The early human cases of SARS infection were reported in the Guangdong province in China in November 2002. The infection, however, quickly spread to 30 different countries globally, infecting over 8000 individuals with a case fatality rate of 10% [3–6]. Soon after the initial reports of SARS infection, the etiological agent was identified as a coronavirus, subsequently named SARS-CoV. SARS-CoV quickly evolved in human hosts to allow efficient human-to-human transmission [1, 3–6]. SARS-CoV infection was rapidly

contained through effective global health response via partnerships between WHO and 115 national health services. [5, 6]

Prior to the emergence of SARS-CoV, two human coronaviruses (HCoV) had been identified. HCoV-229E and HCoV-OC43 were identified in the mid-1960s (Figure 1.2) [7–9]. Infection from these viruses cause mild, self-contained cold-like symptoms, and are considered harmless. With renewed interest in the coronavirus research after the SARS outbreak, two new human coronaviruses, HCoV-NL63 and HCoV-HKU1, were discovered in 2004 and 2005, respectively (Figure 1.2) [10–12]. Symptoms associated with NL63 and HKU1 infections include mild flu-like symptoms, however, severe outcomes have been reported in infants, elderly and immuno-compromised individuals [10–12].

News of coronavirus infections emerged again in September 2012, when the first confirmed case of a novel human coronavirus infection was reported in a patient from Qatar [13, 14]. Since the initial cases of this new human coronavirus infection were reported in countries from the Middle East, the virus was subsequently named Middle East respiratory syndrome coronavirus (MERS-CoV) (Figure 1.2) [15]. As of October 22, 2015, a total of 1599 cases of MERS-CoV infection have been reported from 27 different countries globally. These cases include endemic cases as well as travel-associated cases [16]. 574 global deaths have so far been reported by the WHO bringing the case fatality rate from MERS-CoV infections to 35%. Clinical outcomes from MERS-CoV infections can range from being asymptomatic to severe cases of respiratory failure and gastrointestinal symptoms. Severe cases have mostly been reported in older patients with weakened immune system and underlying diseased conditions. Although human-to-human transmission of MERS-CoV is not efficient, cases of virus transmission between patients and healthcare staff upon close contact have been reported [13, 17–20]. Currently, no licensed vaccine or antiviral treatment exists for MERS-CoV. Recent reports, however, on the success of DNA vaccines based on viral S protein in rhesus macaques and isolation of a potent MERS-CoV-neutralizing

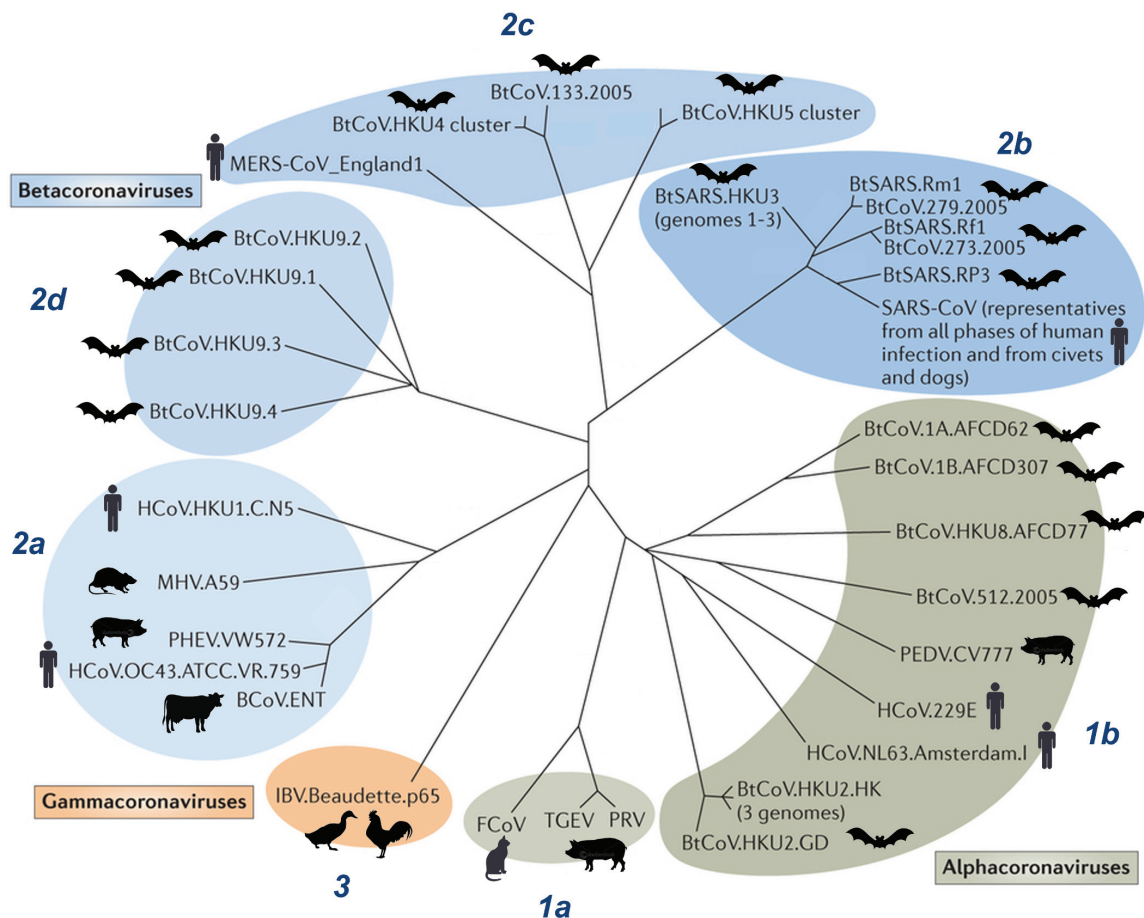


Fig. 1.2.: **Phylogeny and host tropism of representative coronaviruses.** Based on full genome sequencing, coronaviruses can be classified into four distinct groups: Alphacoronaviruses (green), Betacoronaviruses (blue), Gammacoronaviruses (orange), and Deltacoronaviruses (not shown). Taxonomic classification based on former designations into Groups 1, 2 and 3 are also shown. Host tropism for representative coronaviruses in each subgroup is displayed. Figure adapted from Graham et al. [2]

antibody holds promise for the development of treatment regimens against infection from MERS-CoV [21, 22].

1.3 Animal coronaviruses

Besides humans, coronaviruses also infect a variety of animals, which could be life-threatening for animals in certain cases (Figure 1.2). For example, Porcine epidemic

diarrhea virus (PEDV) first appeared in the United States in 2013 and has so far been reported to have killed approximately 7 million swine with a concomitant increase in pork prices [23,24]. Although the virus is not transmissible to humans, it can cause up to a 75-100 % loss of the infected piglets. Transmissible gastroenteritis virus (TGEV) is another coronavirus that infects pigs, and similar to PEDV causes severe diarrhea in piglets [1]. Coronavirus infections have been reported in other farm animals as well. Infectious bronchitis virus (IBV) infects chicken, while bovine coronavirus (BCoV) cause respiratory tract infection and diarrhea in cattle [1,25,26].

Coronaviruses can also infect companion animals like cats and dogs. Infection with feline infectious peritonitis virus (FIPV), a highly virulent form of feline coronavirus (FCoV), can be fatal for cats [1,27]. Canine coronavirus infections are usually mild, but can be fatal if associated with simultaneous infection with canine parvovirus.

Infection from all the aforementioned animal viruses imposes a significant economic or emotional burden to the animal owner and further highlights the significance of understanding coronavirus biology and potential development of antiviral therapeutics to combat existing and emerging coronavirus infections.

1.4 Zoonotic transmission of coronaviruses

SARS-CoV crossed species from Himalayan palm civets and raccoon dogs to infect humans [1,28]. The zoonotic animal-to-human transmission of SARS-CoV is believed to have occurred in the live animal wet markets in China, where the animal handlers became infected after handling of the infected live animals [1,29]. The virus underwent rapid adaptation in both the animal and the human hosts; however, further investigations suggested that the palm civets or the raccoon dogs are not the natural reservoirs of SARS-like coronaviruses. Isolation of a SARS-like virus from the Chinese horseshoe bats, habitation of these bats near the live animal markets and further genetic and serological analysis suggests that the horseshoe bats are natural reservoirs of SARS-like CoV (Figure 1.2) [1,2]. The bat SARS-like CoV spread from

bats to the intermediate animal hosts, like palm civets, before adaptation to infect the human population.

Moreover, several studies published recently strongly support bats as the natural reservoir for the recently emerged MERS-CoV as well [30–33]. β -CoV lineage 2c members (bat coronaviruses HKU4 and HKU5) share close phylogenetic relationship with MERS-CoV (Figure 1.2), suggesting a potential bat origin for MERS-CoV. Additionally, neutralizing antibodies against MERS-CoV have been isolated from dromedary camels, implicating dromedary camels as the intermediary host for MERS-CoV [34, 35]. In a more recent study, Wang et al. demonstrated that the receptor binding domain of the S protein of Bat-CoV HKU4 can recognize the human receptor CD26 (also known as Dipeptidyl peptidase-4 or DPP4) for cell entry [32]. These findings further solidify the notion of a bat origin for MERS-CoV.

SARS-CoV and MERS-CoV are not the only examples of cross-species transmission of coronaviruses. Genomic sequence similarity between bovine coronavirus (BCoV) and human coronavirus OC43 suggests zoonotic transmission of the virus from bovine to human host around 100 years ago [1]. Emergence of transmissible gastroenteritis virus (TGEV) in pigs has been suggested to be the result of a cross-species transmission of canine coronavirus CCoV-II, which in turn has emerged from a recombination event between an unknown coronavirus, canine coronavirus CCoV-I and feline coronavirus FCoV-I [1].

Over 1,100 species of bats have been documented so far. Global prevalence of bats and their habitation proximity to humans and animals enhances their potential to harbor novel coronaviruses capable of bat-to-animal and bat-to-human transmission. Moreover, since bats serve as natural reservoir for multiple coronaviruses (Figure 1.2), there exists a constant threat for the re-emergence of potentially lethal human or animal coronaviruses. Therefore, continuous research on coronavirus replication and pathogenesis is mandated for the development of effective antiviral therapeutics.

1.5 Coronavirus lifecycle

Coronavirus particles are pleomorphic with diameters ranging from 70-120 nm [2]. The positive-sense RNA genome (25-32 kb) is present inside the virus particle in complex with nucleocapsid protein (N) (Figure 1.3). The N protein-RNA genome complex is surrounded by a host-membrane-derived lipid bilayer envelope. The surface of the lipid bilayer envelope is decorated with viral structural proteins: Spike (S), Membrane (M) and Envelope (E) proteins. The spike glycoprotein interacts with host-specific cellular receptor, and mediates the virus attachment and entry [2]. Low sequence conservation of receptor-binding domain of the spike glycoprotein allows variability in the usage of host cellular receptor among different coronaviruses and contributes towards the broad host- and cell-tropism of coronaviruses [2]. A brief list of representative coronaviruses and their cellular receptors is provided in table 1.1 [1]. The ion channel activity of E protein is essential for virus replication, and this integral membrane protein also plays role in virus assembly and budding [1].

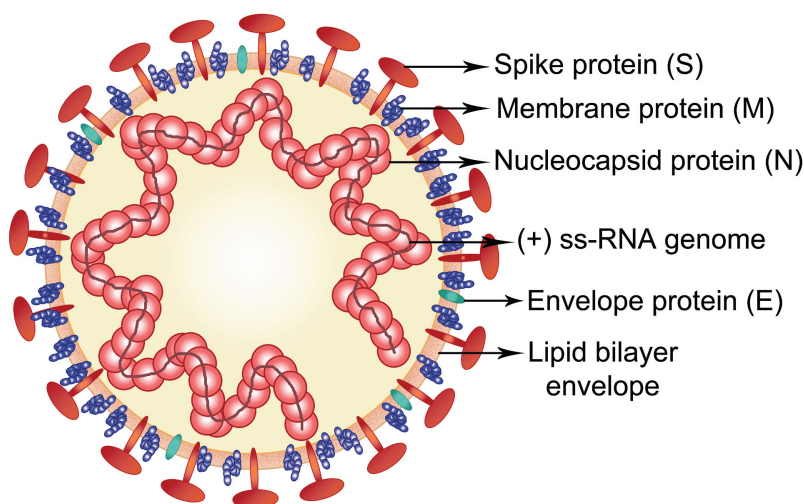


Fig. 1.3.: **Coronavirus virion structure and proteins.** Schematic illustration of a coronavirus virion is shown with the structural proteins (S, M, N and E), the positive-sense single-stranded RNA genome and the lipid bilayer envelope.

The coronavirus lifecycle (Figure 1.4) commences with the attachment of the virus particle on the host cell via interaction between the virus spike protein and the

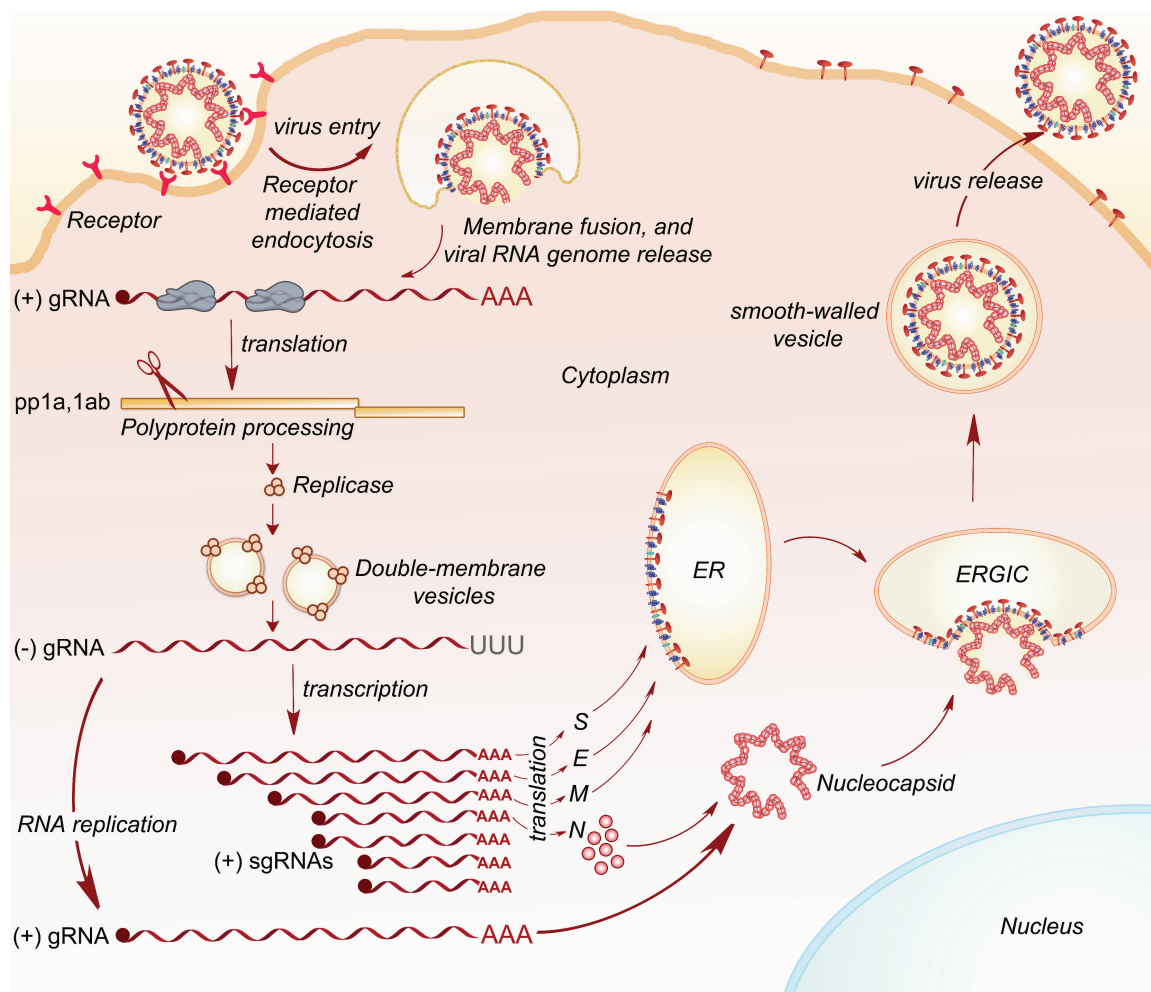


Fig. 1.4.: **Coronavirus life cycle.** Virus entry is mediated through interaction between the viral spike protein and host cell surface receptor. Viral genomic RNA (gRNA) is released into the cytoplasm and translated into two overlapping polyproteins pp1a/1ab using host translational machinery. The polyproteins undergo proteolytic processing using viral encoded cysteine proteases (PL^{Pro} and $3CL^{Pro}$) and release sixteen non structural proteins (nsp1-16). The non-structural proteins assemble to form replicase complex ER-derived double membrane vesicles and facilitate the transcription of (-)gRNA, a nested set of subgenomic mRNAs (sgRNA) and also the replication of (+)gRNA. sgRNAs are translated into structural proteins (S, E, M and N) and accessory proteins. N protein assembles with the newly synthesized (+)gRNA to form the nucleocapsid, while other structural proteins S, E and M traffic through ER and combines with the nucleocapsid at the endoplasmic-reticulum golgi intermediate compartment (ERGIC). The virions are packaged into smooth-walled vesicles and the mature virus particles exit the host cell via exocytosis.

Table 1.1.
Representative coronaviruses and their cellular receptors

Group	Subgroup	Host	Virus	Receptor (Co-receptor)
Alpha	1a	Bat [†] Cat Dog Pig	BtCoV FCoV, FIPV CCoV TGEV	Unknown APN, APN APN APN (Sialic Acid)
	1b	Human Human Pig	HCoV-229E HCoV-NL63 PEDV	APN ACE2 APN
Beta	2a	Cattle Human Human Mouse Pig	BCoV HCoV-HKU1 HCoV-OC43 MHV PHEV	Sialic Acid Unknown Unknown (Sialic Acid) CEACAM1a (Sialic Acid) Unknown
	2b	Bat [†] Human	Bat-SCoV SARS-CoV	ACE2? ACE2 (DC-SIGN, DC-SIGNR, LSECtin)
	2c	Human Bat	MERS-CoV BtCoV-HKU4	DPP4 (also known as CD26) DPP4
Gamma	3a	Chicken	IBV	Unknown (Sialic Acid)
	3b	Beluga whale	SW1	Unknown
	3c	Thrush	ThCoV-HKU12	Unknown
Delta	-	Bird	Bird CoVs (multiple species)	Unknown

[†] Several species of bat coronaviruses have been identified and classified as members of genogroup alphacoronavirus or betacoronavirus. BtCoV, bat coronavirus; FCoV, feline coronavirus; FIPV, feline infectious peritonitis virus; CCoV, canine coronavirus; TGEV, transmissible gastroenteritis virus; HCoV, human coronavirus; MHV, murine hepatitis virus; PHEV, porcine hemagglutinating encephalomyelitis virus; Bat-SCoV, bat SARS-related coronavirus; SARS-CoV, severe acute respiratory syndrome coronavirus; MERS-CoV, Middle-East respiratory syndrome coronavirus; IBV, infectious bronchitis virus; ThCoV, thrush coronavirus; APN, aminopeptidase N; ACE2, angiotensin-converting enzyme 2; CEACAM1a, carcinoembryonic cell adhesion molecule 1a; DC-SIGN, dendritic cell-specific ICAM3-grabbing non-integrin; DC-SIGNR, DC-SIGN-related protein; LSECtin, liver and lymph node sinusoidal C-type lectin; DPP4, dipeptidyl peptidase 4; CD, cluster of differentiation. Table is adapted from Perlman et al. [1]

host cellular receptor. Interaction with the cellular receptor triggers conformational changes in the spike protein that mediate membrane fusion and entry. Next, the

positive sense RNA genome is released into the cytoplasm and two-thirds of the viral genome translated into viral replicase polyproteins 1a and 1ab (Figure 1.5.A). Processing of the polyproteins into 16 non-structural proteins (nsp1-16) is mediated by two viral proteases, papain-like protease (PL^{pro}) or 3 chymotrypsin-like protease (3CL^{pro}). PL^{pro} mediates the release of nsp1, nsp2 and nsp3, while the protease activity of 3CL^{pro} releases nsp4 through nsp16 (Figure 1.5.A). The non-structural proteins assemble at the ER-derived double membrane vesicles to form the replicase complex, which carries out the synthesis of full-length genomic RNA as well as a nested set of sub-genomic mRNA (sgRNA). Figure 1.5.B summarizes the role of different non-structural proteins during virus genome replication and synthesis of sgRNA [1]. Translation of sgRNA produce structural protein S, E, M and N. S, E and M structural proteins mature and traffic through the endoplasmic reticulum, and combines with the N-protein encapsidated viral genome in the ER-golgi intermediate complex (ERGIC). The complete virion assembles in the ERGIC and exits the host cell via exocytosis.

1.5.1 Coronavirus proteases as drug targets

As mentioned earlier, coronavirus polyprotein processing requires the proteolytic activity of two different virus encoded proteases. Proteolytic activity of a single (or two in some cases) papain-like protease, PL^{pro}, is responsible for the release of nsp1 to nsp3 via proteolytic processing at three distinct cleavage sites nsp1|nsp2, nsp2|nsp3 and nsp3|nsp4 [36]. PL^{pro} is encoded as a protease domain within the large multi-functional nsp3 protein. In addition to the protease activity, PL^{pro} also possesses the ability to remove ubiquitin and ISG15 (interferon-stimulated gene 15) molecules from a variety of cellular substrates [36]. Since a variety of cellular proteins undergo ubiquitination and ISGylation during activation of innate immune response, the deubiquitination and deISGylation activities of PL^{pro} play important roles in the viral antagonism of host innate immune response [36].

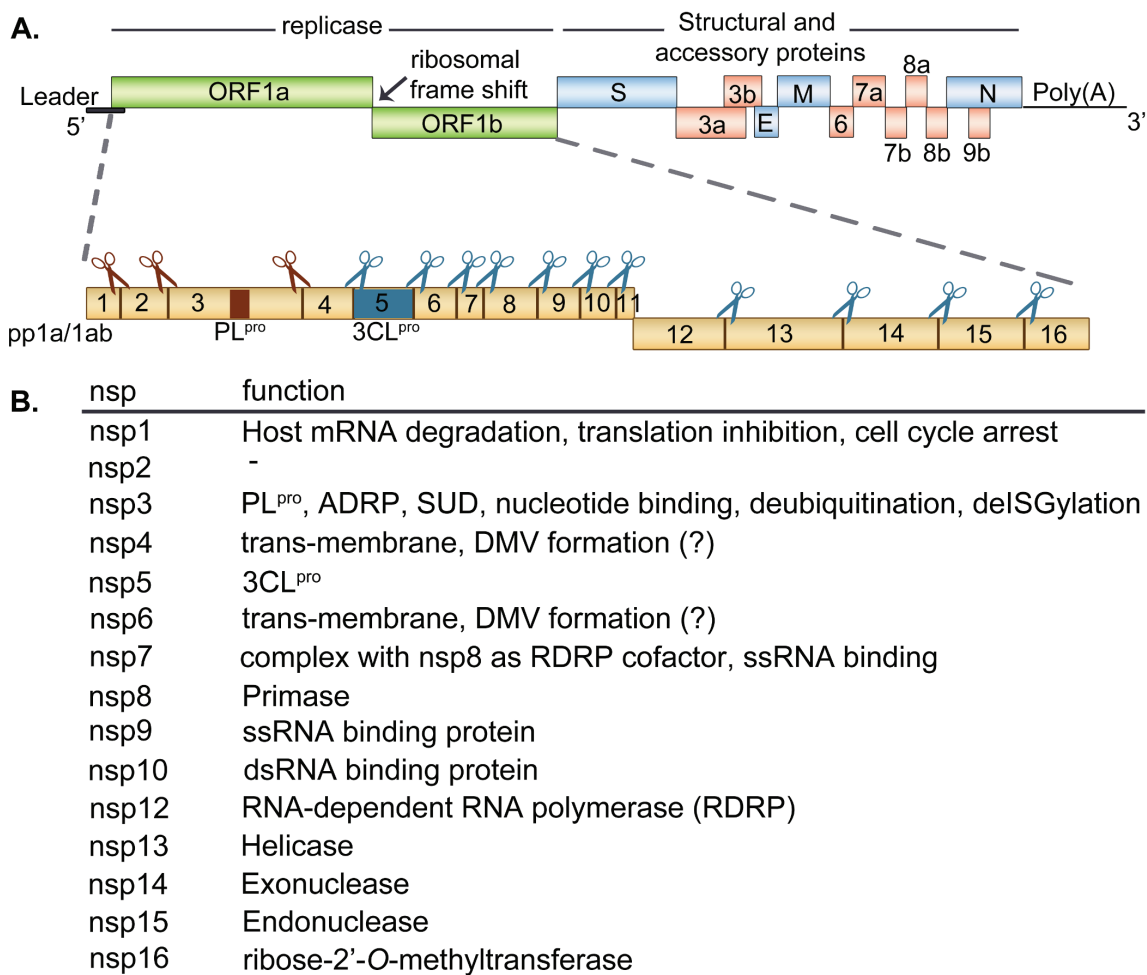


Fig. 1.5.: **Coronavirus genome and non-structural proteins.** **A.** Schematic illustration of SARS-CoV genome arrangement. ORFs (ORF1a and 1b) for the replicase polyproteins along with the ORF's for structural and accessory proteins are depicted. The first two-thirds of the genome is translated into two overlapping polyproteins 1a and 1ab, where translation of 1ab results from the presence of a ribosome frame-shift element at the termination site for 1a. The polyproteins are processed by viral encoded proteases PL^{pro} (red box) and 3CL^{pro} (blue box). PL^{pro} and 3CL^{pro} cleavage sites are shown as red and blue scissors, respectively. nsp1-16 are released from the polyproteins and form the replicase complex. **B.** nsp's are listed along with their associated functions. PL^{pro}, papain-like protease, ADRP, ADP-ribose-1''-monophosphatase; SUD, SARS unique domain; 3CL^{pro}, 3 Chymotrypsin-like protease; RDRP, RNA-dependent RNA polymerase; DMV, double-membrane vesicle.

The second viral encoded protease, 3CL^{pro} or 3 chymotrypsin-like protease, undertakes polyprotein processing at eleven distinct cleavage sites, and is responsible for the release of nsp4 through nsp16 [1]. 3CL^{pro} has also been referred as non-structural protein 5 (nsp5) and Main protease (M^{pro}) in the literature. Both PL^{pro} and 3CL^{pro} have been targeted for inhibitor design for the development of antiviral therapeutics against existing and emerging coronaviruses [36–39].

1.6 Coronavirus 3CL^{pro} protease

The coronavirus 3CL^{pro} enzyme is essential for the polyprotein processing at eleven distinct cleavage sites, albeit with different efficiencies [1, 40, 41]. Each 3CL^{pro} protomer is structurally organized into three domains (Figure 1.6.A). Domains 1 and 2 constitute the catalytic fold with a typical chymotrypsin-like fold. In SARS-CoV 3CL^{pro}, His41 and Cys 145 form a catalytic dyad (Figure 1.6.B) that is present in the cleft between the domains 1 (amino acids 1-101) and 2 (amino acids 102-185). Domain 3 (amino acids 201-306) is unique to 3CL^{pro} enzymes, when compared to chymotrypsin, and contributes significant number of residues that are essential for 3CL^{pro} dimerization [42–48].

1.6.1 Enzymatic reaction mechanism for coronavirus 3CL^{pro}

Unlike a typical serine protease that contains a Cys-His-Asp catalytic triad, the active site of SARS-CoV 3CL^{pro} is comprised of a Cys145-His41 catalytic dyad. A conserved water molecule has been observed in the crystal structures of SARS-CoV 3CL^{pro} [43, 49, 50] and is suggested to assist in catalysis by stabilizing the active conformational orientation of His41 (Figure 1.7, E) [51, 52]. His41 extracts one proton from the side chain thiol of Cys145 via hydrogen bonding interaction between the side chain nitrogen on His41 and the thiol side chain of Cys145 (Figure 1.7, Step a). Next, the thiolate nucleophile attacks the carbonyl carbon of the scissile bond in the substrate (Figure 1.7, formation of ES complex). The *pi*-electrons of the substrate

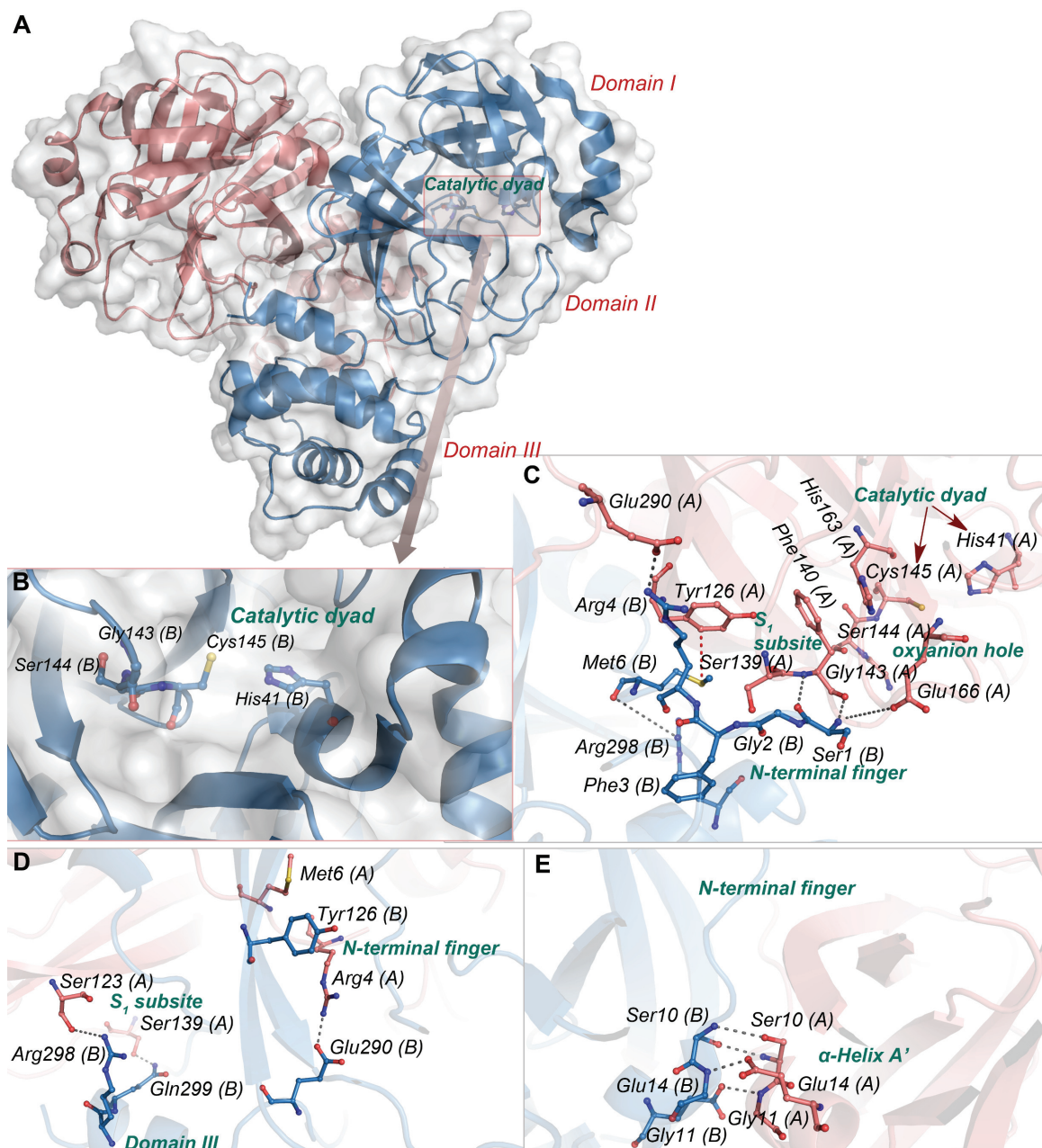


Fig. 1.6.: Detailed dimeric structure of SARS-CoV 3CL^{pro}. **A.** SARS-CoV 3CL^{pro} dimer with monomers A (salmon) and B (blue) is depicted. Domains I, II and III are labeled. The catalytic dyad, composed of His41 and Cys145, is present in the cleft between domains I and II. **B.** Active site residues, His41 and Cys145, are illustrated in ball-and-stick model. Gly143, Ser144 and Cys145 form the oxyanion hole. Primary interactions at the dimer interface involve: **C.** interactions between *S*₁ subsite of monomer A (salmon) and the N-terminal finger (residues 1-7) of monomer B (blue); **D.** interactions of the *S*₁ subsite and the N-terminal finger of monomer A with residues from domain III of monomer B; **E.** interactions between A' α -helices (amino acids 10-15) from each monomer.

carbonyl group are pushed onto the oxygen leading to the formation of a transient tetrahedral intermediate (TI-1) with an oxygen anion (oxyanion) (Figure 1.7, formation of FP complex). The oxyanion is transiently stabilized by hydrogen bonding interactions with the backbone amide of residues in the oxyanion hole (Gly143 and Cys145). Next, the oxyanion electrons are pushed back to reform the carbonyl double bond at the C-terminus of the scissile bond. The scissile bond is cleaved resulting in the formation of thioester enzyme intermediate (Figure 1.7, species F), and the C-terminus of the cleaved peptide substrate dissociates from the active site after abstracting a proton from the side chain nitrogen of His41.

Another water molecule enters the active site via hydrogen bond formation with His41 and attacks the carbonyl carbon of the thioester intermediate. The X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with peptidomimetic inhibitors provide clear evidence for the presence of this water molecule within hydrogen-bonding distance of His41 (PDB ID: 2ALV and 2QIQ; [53, 54]). This results in the formation of second tetrahedral intermediate (TI-2) and abstraction of one proton from water by His41 (Figure 1.7, formation of FQ complex). The oxyanion of TI-2 is transiently stabilized by backbone amides of Gly143 and Cys145. Next, electrons from the oxyanion reform the carbonyl double bond while the the Cys145 sulfur abstracts a proton from His41. Finally, the bond between Cys145 sulfur and carbonyl carbon from the substrate breaks, regenerating Cys145 thiol side chain, and the N-terminus half of the cleaved substrate dissociates from the active site.

1.6.2 3CL^{pro} is catalytically active as a dimer

Using recombinantly expressed and purified protein, several studies have shown that SARS-CoV 3CL^{pro} exists as a tightly associated dimer in solution in a concentration dependent manner [45]. X-ray crystal structures of SARS-CoV 3CL^{pro}, either in apo or ligand-bound forms, also reveal the enzyme as a dimer, suggesting this quaternary structural form as the biological functional unit [43–47]. In fact, mutation of

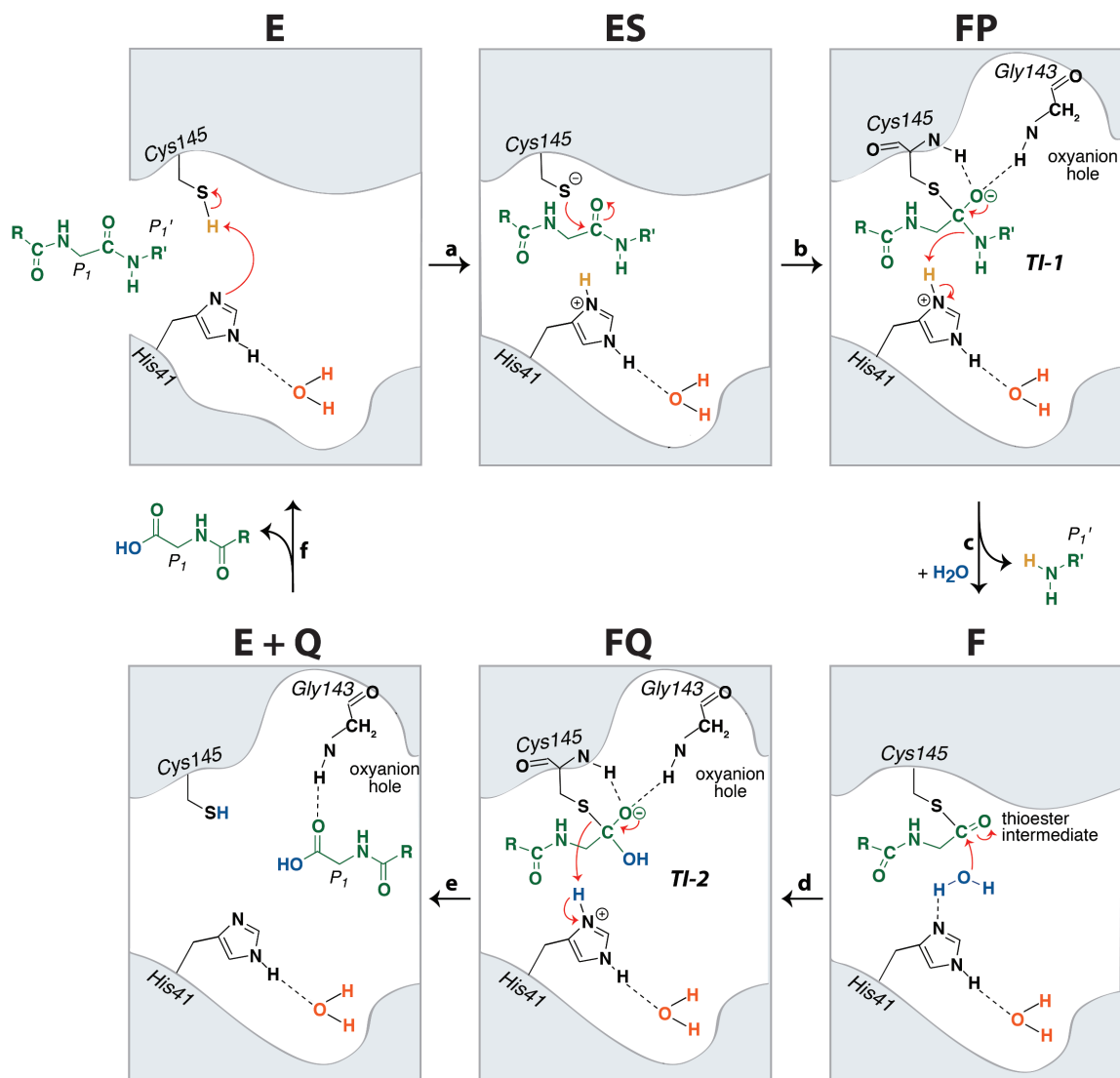


Fig. 1.7.: **Enzymatic reaction mechanism for SARS-CoV 3CL^{pro}.** Active site residues, Cys145 and His41, that form the catalytic dyad are shown. Oxyanion hole residues Gly143 and Cys145 are also shown. Active site water molecule that assists in making His41 a strong base is shown in orange, while catalytic water molecule that attacks the carbonyl oxygen of the scissile bond to form TI-2 is shown in blue. The peptide substrate is depicted in green. Figure is adapted from Baez-Santos et al. [36]

residues that interact across the dimer interface results in a partial or complete loss of dimer formation with a concomitant loss in the enzymatic activity. X-ray crystal structures of the dimer illustrate that the two monomers orient almost perpendicular to each other (Figure 1.6.A). In the active conformation, the S_1 substrate binding

pocket (Figure 1.6.C) is present an open conformation to allow the binding of the corresponding P_I substrate residue. Additionally, the oxyanion hole (Figure 1.6.C) of the active conformation is in the correct orientation to donate two hydrogen bonds from the main chain amides (Gly143 and Cys145) to stabilize the oxyanion of the first tetrahedral intermediate formed during the catalytic cycle. Crystal structures of monomeric SARS-CoV 3CL^{pro} mutants suggest that the loss in enzymatic activity for the monomer originates from the inaccessible collapsed conformation of both the S_I subsite and the oxyanion hole [47,55].

Interactions at the dimer interface of SARS-CoV 3CL^{pro}

Residues that interact to form the 3CL^{pro} dimer interface primarily cluster at the following hot spots: 1) the N-terminal finger (amino acids 1-7); 2) the N-terminal helix (also known as the A' α -helix; amino acids 10-15); 3) amino acids near the S_I subsite; and 4) amino acids in domain III, mainly at the C-terminus. The N-terminal finger residues from one monomer make hydrogen bonding and hydrophobic interactions with several amino acids that form the S_I subsite and the oxyanion hole of the opposing monomer (Figure 1.6.C). Interaction between Arg4 (domain I) and Glu290 (domain III) across the dimer interface is critical (Figure 1.6.D) [55], since mutation of any of these residues drastically ablates the dimer formation and the enzymatic activity. Arg298 is another important residue that is essential for the dimer formation and its mutation results in complete loss of dimerization and enzymatic activity [47]. Mutation studies have further highlighted the significance of residues of the N-terminal helix in the dimer formation (Figure 1.6.E) [56].

Since residues from both the N- and C-termini participate in dimer formation, it is absolutely critical to use a recombinant enzyme without any modification (addition of affinity tags or deletion of residues) on the termini for experimental studies. Furthermore, several groups (including ours) have shown that SARS-CoV 3CL^{pro} with

‘authentic’ N- and C-termini forms a stable dimer that has orders of magnitude higher activity compared to the enzyme with modifications at the termini [45].

1.6.3 Model for auto-release of SARS-CoV 3CL^{pro} from the polyprotein

As discussed in the previous section, SARS-CoV 3CL^{pro} is a functional dimer and modifications at either termini interfere with the dimer formation. Since the dimer is the active form of the enzyme, it is intriguing how 3CL^{pro} liberates itself from the polyprotein in the context of viral replication. Based on several studies performed *in-Vitro* with the recombinantly purified protein, the following model has been proposed to describe the auto-cleavage of SARS-CoV 3CL^{pro} (Figure 1.8) [43, 57,58]. First, two immature monomers present in two different polyproteins approach each other and form an ‘immature’ dimer conformation through interactions between their third domains. This results in a domain swap where the N-terminus region of the polyprotein immediately upstream of 3CL^{pro} interjects itself into the active site of the opposing monomer, leading to the formation of an immature dimer (Figure 1.8, Step 1). Next, the polyprotein is cleaved at the N-terminus auto-cleavage site of SARS-CoV 3CL^{pro} (Figure 1.8, Step 2). In Step 3, the uncleaved C-terminus of the dimer is aligned in the active site of another dimer. The uncleaved C-termini of the dimer is then processed liberating an active mature dimer of SARS-CoV 3CL^{pro} (Step 4). Once an active dimer is formed, it can process other cleavage sites in the polyprotein (Step 5).

1.7 Statement of intent

Emergence of two highly pathogenic human coronaviruses in the 21st century highlights the significance of studying coronavirus biology and identify coronavirus drug targets. Coronavirus 3CL^{pro} protease has been utilized by several groups for the development of anti-coronaviral therapeutics. The overarching goal of this project is to investigate enzymatic and structural properties of multiple 3CL^{pro} enzymes

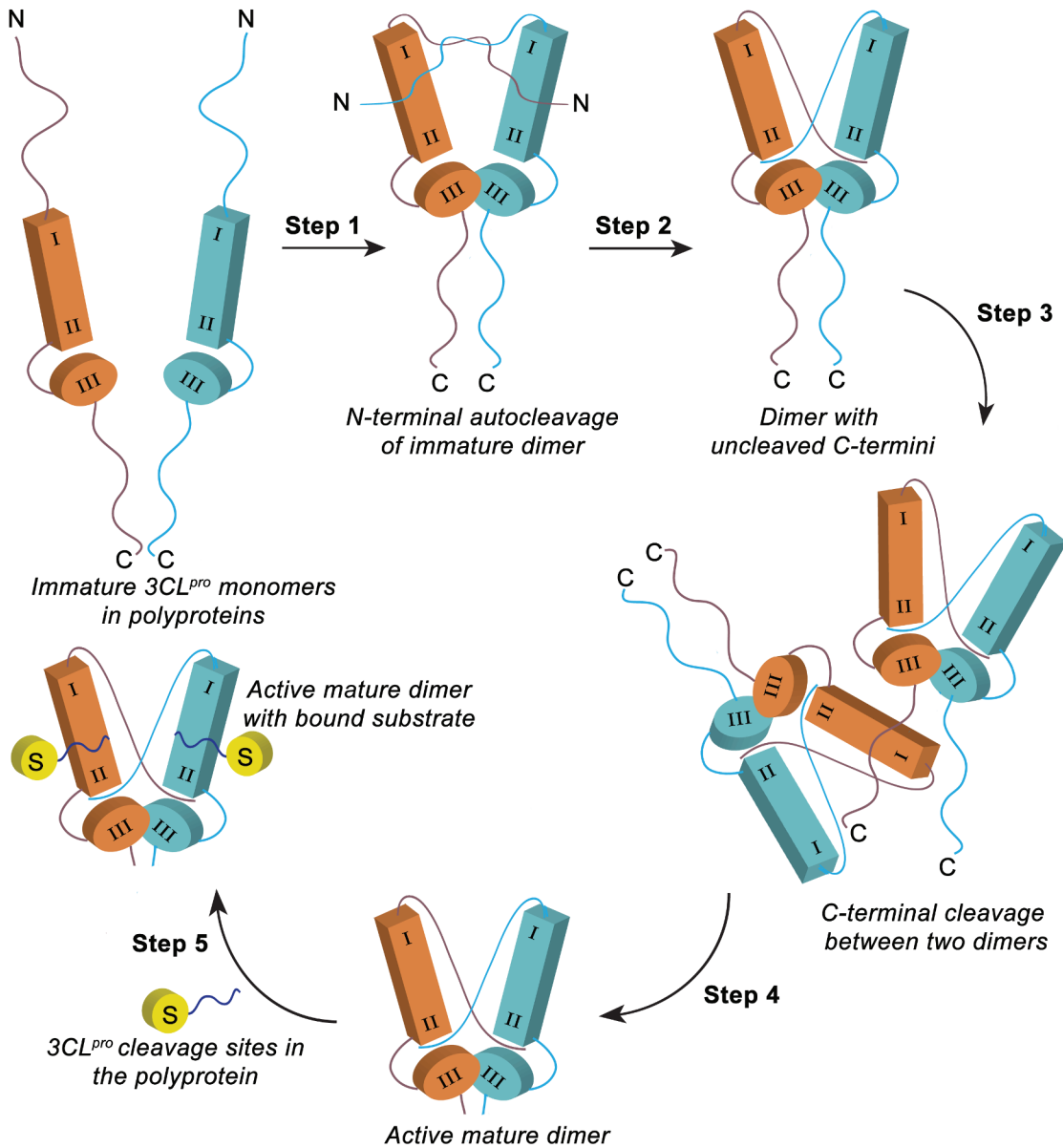


Fig. 1.8.: **Proposed model for auto-release of SARS-CoV 3CL^{pro} from the polyprotein.** (adapted from Chen et al. [57]) SARS-CoV 3CL^{pro} is illustrated as rectangular box (domain I and II) and cylinder (domain III). N- and C-termini are labeled. 3CL^{pro} substrate (cleavage sites in the polyprotein) is shown as yellow cylinder labeled 'S'. Steps for SARS-CoV 3CL^{pro} auto-release from the polyprotein and subsequent 3CL^{pro}-mediated processing of the polyprotein have been described in details in the text.

encompassing different coronavirus subclasses. Understanding the determinants of structural and functional disparity between different 3CL^{pro} enzymes and the fac-

tors regulating these properties will aid in the design of broad-spectrum inhibitors of 3CL^{pro} enzymes.

CHAPTER 2. EXPERIMENTAL PROCEDURES

Parts of the data and text in this chapter have been published in various journal articles [45, 59, 60].

2.1 Recipes for expression and purification of 3CL^{pro} enzymes

2.1.1 Recipe for 1 liter LB medium

10 gm tryptone

5 gm yeast-extract

10 gm Sodium chloride

pH was adjusted to 7.5 before adjusting the final volume to 1 liter, followed by sterilizing the media by autoclaving.

2.1.2 Recipe for 1 liter Super-LB medium for auto-induction

3 gm KH_2PO_4

6 gm Na_2HPO_4

20 gm tryptone

5 gm yeast-extract

5 gm Sodium chloride

pH was adjusted to 7.2 before adjusting the final volume to 1 liter, followed by sterilizing the media by autoclaving.

8% lactose and 10% glucose were prepared in autoclaved water and then filter-sterilized. 60% glycerol was prepared in water and then autoclaved. For auto-induction, 25-50 mL of overnight bacterial culture was inoculated in 1 liter of Super-LB media with 25 mL of 8% lactose, 5 mL of 10% glucose, 10 ml of 60% glycerol and

100 $\mu\text{g}/\text{mL}$ of carbenicillin. Cells were grown in an INFORS multitron incubation shaker with an orbital diameter of 2.5 cm at 25 °C for 22-24 hours at 180 rpm.

2.1.3 Buffers for purification of 3CL^{pro} enzymes

SARS-CoV 3CL^{pro} purification buffers

•**Buffer A**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer B**

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

•**Buffer C**

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Ammonium sulfate

5 mM BME; freshly added right before purification

•**Buffer D**

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

Storage buffer was prepared by adding 10% glycerol to buffer D

MHV 3CL^{pro} purification buffers

•**Buffer A**

50 mM Tris pH-7.5

0.05 mM EDTA

0.2 M Ammonium sulfate

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer B**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer C**

20 mM MES pH-6.0

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer D**

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

•**Buffer E**

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

Storage buffer was prepared by adding 5% glycerol to buffer E

HKU1 3CL^{pro} purification buffers

•**Buffer A**

50 mM Tris pH-7.5

0.05 mM EDTA

0.2 M Ammonium sulfate

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer B**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer C**

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

•**Buffer D**

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

Storage buffer was prepared by adding 10% glycerol to buffer D

OC43 3CL^{pro} purification buffers

•**Buffer A**

50 mM Tris pH-7.5

0.05 mM EDTA

0.2 M Ammonium sulfate

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer B**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•**Buffer C**

20 mM MES pH-6.0

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

•Buffer D

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

•Buffer E

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

Storage buffer was prepared by adding 10% glycerol to buffer E

HKU5-CoV 3CL^{pro} purification buffers**•Buffer A**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

10% glycerol

•Buffer B

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Ammonium sulfate

5 mM β -mercaptoethanol (BME); freshly added right before purification

10% glycerol

•Buffer C

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

10% glycerol

•Buffer D

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

10% glycerol

MERS-CoV 3CL^{pro} purification buffers**•Buffer A**

20 mM Tris pH-7.5

0.05 mM EDTA

5 mM β -mercaptoethanol (BME); freshly added right before purification

10% glycerol

•Buffer B

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Ammonium sulfate

5 mM β -mercaptoethanol (BME); freshly added right before purification

10% glycerol

•Buffer C

50 mM Tris pH-7.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

10% glycerol

•Buffer D

20 mM MES pH-5.5

0.05 mM EDTA

5 mM BME; freshly added right before purification

10% glycerol

•Buffer E

50 mM MES pH-5.5

0.05 mM EDTA

1.0 M Sodium chloride

5 mM BME; freshly added right before purification

10% glycerol

•Buffer F

25 mM HEPES pH-7.5

2.5 mM dithiothreitol (DTT)

10% glycerol

2.2 Expression and purification of SARS-CoV 3CL^{pro}

The codon-optimized gene for SARS-CoV 3CL^{pro} was subcloned into the pET-11a expression vector by Grum-Tokers et al. [45] with nucleotides coding for a TEV protease cleavage site between the N-terminal (His)₆-tag and the first amino acid for SARS-CoV 3CL^{pro} (Figure 2.1). This construct results in the expression of SARS-CoV 3^{pro} without any N-terminal or C-terminal extension.

For protein expression, *E. coli* BL21-DE3 electro-competent cells were transformed via electroporation with pET-11a plasmid containing the gene for SARS-CoV 3CL^{pro} as described above. The transformed cells were grown overnight at 37 °C on LB agar supplemented with 50 µg/mL of carbenicillin. A single colony was picked from the agar plate and was used to inoculate 100 mL of LB media supplemented with 100 µg/mL of carbenicillin. The cells were grown in the incubation shaker at 37 °C for overnight at 200 rpm. 25 mL of overnight culture was used to inoculate one liter of Super-LB media supplemented with 100 µg/mL of carbenicillin and the protein was expressed through auto-induction for 22-24 hours at 25 °C. The cells were harvested by centrifugation at 5000 × g for 20 minutes at 4 °C, and the pellets were stored at

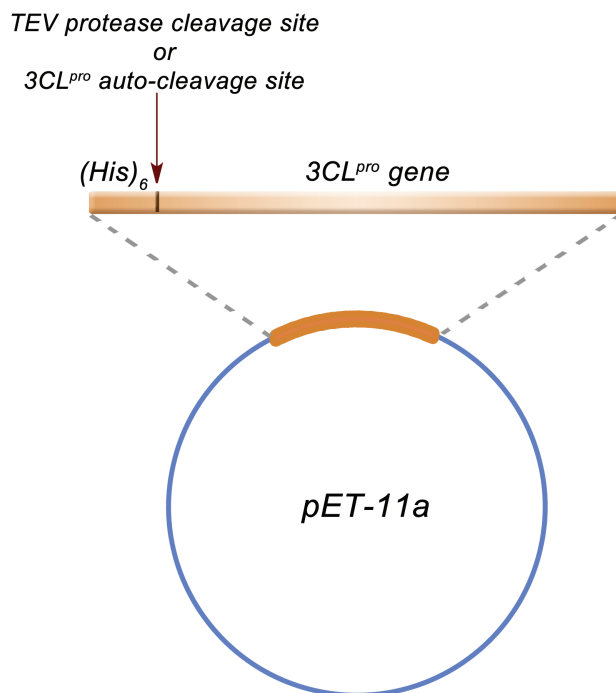


Fig. 2.1.: **Construct design for 3CL^{pro} enzymes.** Codon-optimized gene for 3CL^{pro} protease was cloned into a pET-11a expression vector with an N-terminal (His)₆-tag followed by either the TEV protease cleavage site or the nsp4|nsp5 auto-cleavage site for the corresponding 3CL^{pro} enzyme.

−80 °C until further use. Approximately 22 grams of cell pellet was obtained from 2 L of bacterial culture under the given expression conditions.

The frozen pellets from 2 L of bacterial cell culture were thawed on ice and re-suspended in 5 mL of Buffer A per gram of cell pellet with 500 μg of lysozyme and a small amount of DNase. The cells were then lysed via sonication at 60 % amplitude, with 10 seconds on and 20 seconds off cycles, using a 400 W model Branson sonifier. Cell debris was removed from the cleared lysate by centrifuging at 29,000 × g for 20 minutes.

2.2.1 DEAE anion-exchange chromatography

The clarified cell lysate (120 mL) was loaded at a flow rate of 3 mL/min onto a 60 mL DEAE anion-exchange column (XK 26/20, Amersham Biosciences) equilibrated

with Buffer A. The unbound protein fraction from the load was collected. Once all the sample was loaded, the column was washed with $2 \times$ column volume (120 mL) of Buffer A to remove additional unbound protein fraction. A $2 \times$ column volume linear gradient to 100% Buffer B was finally used to elute the bound protein fraction; 10 mL fractions were collected. Based on SDS-PAGE analysis (Figure 2.2.A) and specific activity measurements, the majority of SARS-CoV 3CL^{pro} was present in the unbound fraction (flow-through).

2.2.2 Hydrophobic-interaction chromatography

Solid ammonium sulfate (70.91 gm) was added to the flow-through (190 mL) obtained from the previous step to a final 60% saturation through gradual mixing on ice using a magnetic stirrer. The sample was then centrifuged at $29,000 \times g$ for 20 minutes at 4 °C. The pellet of precipitated proteins was then resuspended in 65 mL of Buffer C. Sample was loaded at a flow rate of 3 mL/min onto a 30 mL Phenyl Sepharose 6 fast-flow high-sub column (XK 16/20, Amersham Biosciences) equilibrated with Buffer C. The column was then washed with $4 \times$ column volume (120 mL) of Buffer C at a flow rate of 3 mL/min. The protein was eluted using a $10 \times$ column volume (300 mL) linear gradient to 100% Buffer A. The fractions (5 ml) were collected and those containing SARS-CoV 3CL^{pro}, as judged through SDS-PAGE analysis (Figure 2.2.B) and specific activity measurements, were pooled (40 mL) and exchanged into 2 L of Buffer A via overnight dialysis in dialysis tubing (10,000 MWCO SnakeSkin[®], Thermo Scientific).

2.2.3 Mono-Q anion-exchange chromatography

Following dialysis, the protein sample was filtered through a 0.22 μm pore size Millex-GP filter (Millipore) to remove any precipitated protein. The filtered sample was then loaded at a flow-rate of 2 mL/min onto a 8 mL Mono-Q 10/100 column (Amersham Biosciences) equilibrated in Buffer A. Protein was eluted at a flow rate of

3 mL/min using a $10 \times$ column volume (80 mL) and linear gradient to 100% Buffer B (Figure (Figure 2.2.C)). Fractions (3 mL) were collected and those containing SARS-CoV 3CL^{pro} were pooled (9 mL) and concentrated to approximately 4 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore).

2.2.4 Gel-filtration chromatography

As the final purification step, the concentrated protein sample was loaded onto a prep grade Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with Buffer D. Protein was eluted isocratically at a flow rate of 1 mL/min with buffer D (Figure 2.2.D). Fractions (5 ml) containing SARS-CoV 3CL^{pro} were pooled (total volume of 20 mL) and concentrated to approximately 1.5 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore). For final storage of the purified SARS-CoV 3CL^{pro} enzyme, 300 μ L protein aliquots containing glycerol to a final concentration of 10% were placed into 1 mL screw-cap vials, flash-frozen under liquid nitrogen and then stored at -80 °C until further use.

A summary of the percent enzyme yield, total activity units, and the fold-purification after each chromatographic step is summarized in Table 2.1. Approximately 13 mg of highly pure SARS-CoV 3CL^{pro} can be obtained from 2 liters of bacterial cell culture.

Table 2.1.
Purification summary of SARS-CoV 3CL^{pro} from 2 L culture of *E.coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	602	10115	17	1	100
DEAE	478	9231	19	1	91
Phenyl-Sepharose	31	2035	66	4	20
Mono-Q	23	2052	91	5	20
Superdex 75	13	2514	199	12	25

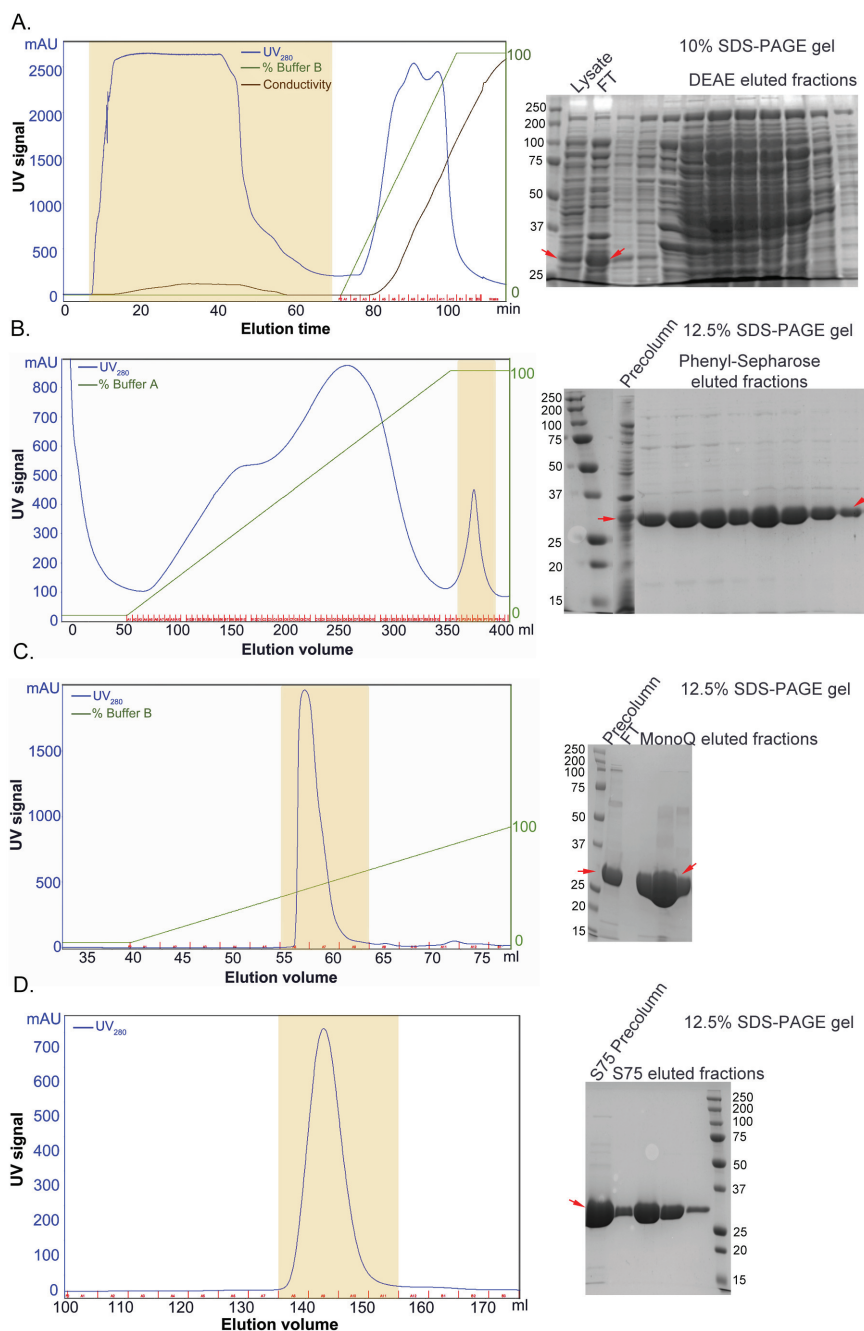


Fig. 2.2.: **Purification of SARS-CoV 3CL^{pro}**. **A.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrows indicate the expected size of SARS-CoV 3CL^{pro}; FT - flow through (unbound fraction). **B.** *Left*-Elution profile from Phenyl-sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **C.** *Left*-Elution profile from Mono-Q column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **D.** *Left*-Elution profile from Superdex 75 column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile.

2.3 Expression and purification of MHV 3CL^{pro}

The gene encoding 3CL^{pro} of MHV (amino acids 3334-3636 in the replicase polyprotein, GenBank: AAU06352.1) was codon optimized for optimal expression in *E. coli* (BioBasic Inc) and was subcloned by Dr. Aimee Egger (Mesecar lab) into pET-11a expression vector with an N-terminal (His)₆-tag and TEV protease cleavage site (Figure 2.1). The plasmids with V148A, H134Y and V148A/H134Y mutations were provided by Dr. Christopher Stobart (Denison lab, Vanderbilt University).

For protein expression, *E. coli* BL21-DE3 electro-competent cells were transformed via electroporation with pET-11a plasmid containing the wildtype and mutant genes. The transformed cells were grown overnight at 37 °C on LB agar supplemented with 50 µg/mL of carbenicillin. A single colony was picked from the agar plate and was used to inoculate 100 mL of LB media supplemented with 100 µg/mL of carbenicillin. The cells were grown in an incubator shaker at 37 °C for overnight at 200 rpm. 20 mL of overnight culture was used to inoculate one liter of LB media supplemented with 100 µg/mL of carbenicillin and the cells were grown at 37 °C, 200 rpm. When the OD₆₀₀ of the culture approached 0.6, protein expression was induced via addition of IPTG to a final concentration of 0.15 mM. Expression of the wildtype and H134Y mutant was induced at 37 °C for 4 hours at 180 rpm, while V148A and V148A/H134Y mutants were expressed at 25 °C for 4 hours at 180 rpm. Cells were harvested by centrifugation at 5000 × g for 20 minutes at 4 °C, and the pellets were stored at -80 °C until further use. Approximately 6 grams of cell pellet was obtained from 3 L of bacterial culture under the given expression conditions.

The frozen pellets from 3 L of bacterial cell culture were thawed on ice and re-suspended in 5 mL of Buffer A per gram of cell pellet with 50 µg of lysozyme and 10 µg of DNase. The cells were then lysed via sonication at 50% amplitude using a 400 W model Branson sonifier and cell debris was removed from the cleared lysate by centrifuging at 29,000 × g for 20 minutes.

2.3.1 Hydrophobic-interaction chromatography

The cleared lysate was loaded at a flow rate of 2 mL/min onto a 30 mL Phenyl Sepharose 6 Fast Flow High Sub column (XK 16/20, Amersham Biosciences) equilibrated with Buffer A. The unbound protein fraction was removed from the load by washing with $3 \times$ column volumes of Buffer A at a flow rate of 3 mL/min. The bound protein fraction was then eluted using a $3 \times$ column volume linear gradient to 100% Buffer B (20 mM Tris pH-7.5, 0.05 mM EDTA and 10 mM BME); 6 mL fractions were collected. The fractions were analyzed through SDS-PAGE (Figure 2.3.A) and specific activity determination. The fractions containing MHV 3CL^{pro} were pooled and pH of the pooled fractions was manually adjusted to 6.0 by gradual addition of solid MES [2-(*N*-morpholino)ethanesulfonic acid] on a magnetic stirrer. For V148A/H134Y, 1 ml of Buffer D was added for every 10 ml of the pooled sample before pH was adjusted. After adjusting the pH, any precipitated protein was removed by filtering through a 0.22 μ m pore size Millex-GP filter (Millipore).

2.3.2 DEAE anion-exchange chromatography

The protein sample from the previous step was loaded at a flow rate of 2 mL/min onto a 60 mL DEAE anion-exchange column (XK 26/20, Amersham Biosciences) equilibrated with Buffer C. The unbound protein fraction was collected in a separate beaker. Once the sample was loaded, the column was washed with $1 \times$ column volume of Buffer C at a flow rate of 3 mL/min to remove additional unbound protein fraction. Finally, $3 \times$ column volume linear gradient to 100% Buffer D was used to elute the bound protein fraction; 5 mL fractions were collected. Based on SDS-PAGE analysis (Figure 2.3.A) and specific activity measurements, the majority of 3CL^{pro} was present in the unbound fraction. The protein was buffer exchanged via overnight dialysis into 2 L of Buffer E using a dialysis tubing (10,000 MWCO SnakeSkin[®] dialysis tubing, Thermo Scientific). The dialyzed sample was concentrated to 1-2 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore) and used for the kinetic

assays. For the CD experiments, gel-filtration chromatography was employed as the final purification step. All the thermal inactivation and CD experiments (Chapter 4) were performed using freshly purified (i.e. not frozen) proteins.

2.3.3 Gel-filtration chromatography

The concentrated protein sample was loaded onto a prep grade Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with Buffer E. A flow rate of 1 mL/min of buffer E was used for isocratic elution of the protein. The fractions containing 3CL^{pro} were pooled and concentrated to approximately 1-2 mg/mL. Total activity units (μ M product/min), specific activity (units/mg) and milligrams of protein obtained (BioRad protein assay) were determined after each chromatographic step to calculate the final protein yield.

Table 2.2.
Purification summary of MHV 3CL^{pro} from 3 L culture of *E.coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	199	19255	97	1	100
Phenyl-Sepharose	26	7429	283	3	39
DEAE	10	4821	461	5	25

2.4 Expression and purification of HKU1-CoV 3CL^{pro}

The gene encoding 3CL^{pro} of HKU1 3CL^{pro} was codon optimized for optimal expression in *E.coli* (BioBasic Inc) and was subcloned by Dr. Aimee Egler (Mesecar lab) into pET-11a expression vector with an N-terminal (His)₆-tag and TEV protease cleavage site (Figure 2.1). The construct was verified by DNA sequencing at the Purdue University Genomics Core Facility.

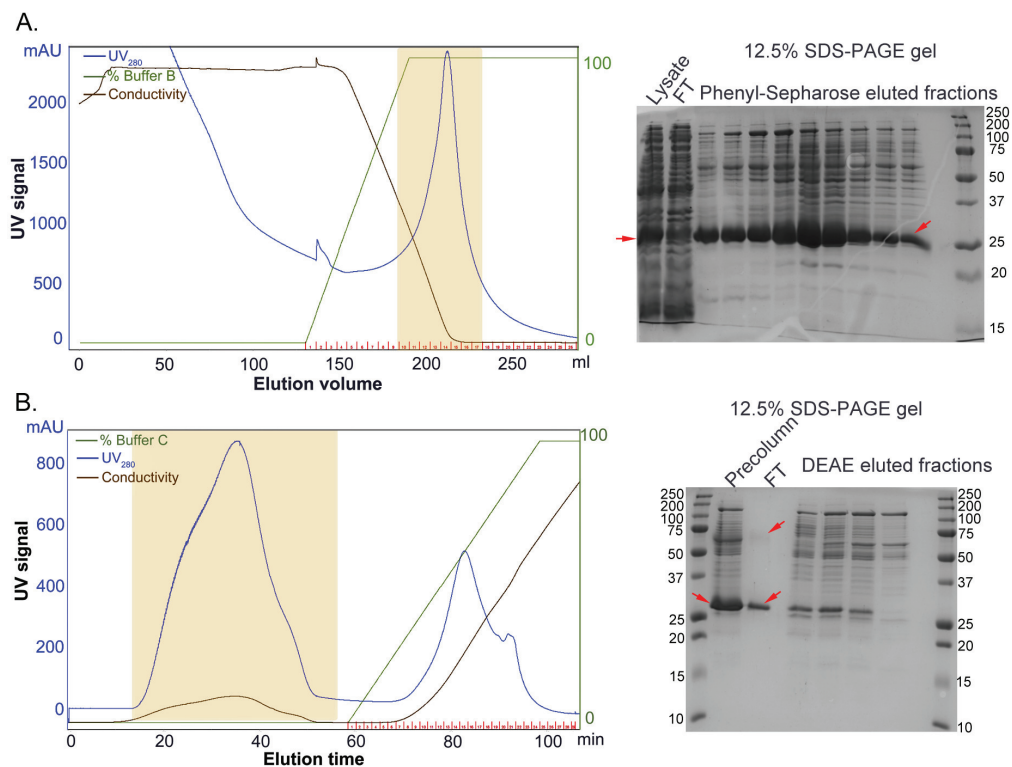


Fig. 2.3.: **Purification of MHV 3CL^{pro}**. **A.** *Left*-Elution profile from Phenyl-sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrows indicate the expected size of MHV 3CL^{pro}; FT - flow through (unbound fraction). **B.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile.

For protein expression, *E. coli* BL21-DE3 electro-competent cells were transformed via electroporation with pET-11a plasmid containing the gene for HKU1 3CL^{Pro}. The transformed cells were grown overnight at 37 °C on LB agar supplemented with 50 µg/mL of carbenicillin. A single colony was picked from the agar plate and was used to inoculate 100 mL of LB media supplemented with 100 µg/mL of carbenicillin. The cells were grown in an incubator shaker at 37 °C for 8 hours at 200 rpm. 50 mL of the bacterial culture was used to inoculate one liter of LB media supplemented with 100 µg/mL of carbenicillin and the cells were grown at 37 °C, 200 rpm. When the OD₆₀₀ of the culture approached 0.6, protein expression was induced via addition of IPTG to a final concentration of 0.5 mM for 12-15 hours at 18 °C, 180 rpm. The cells were harvested by centrifugation at 5000 × g for 20 minutes at 4 °C, and the pellets were stored at −80 °C until further use. Approximately 12 grams of cell pellet was obtained from 3 L of bacterial culture under the given expression conditions.

The frozen pellets from 3 L of bacterial cell culture were thawed on ice and re-suspended in 5 mL (57 mL) of Buffer A per gram of cell pellet with 50 µg of lysozyme and 10 µg of DNase. The cells were then lysed via sonication at 60% amplitude using a 400 W model Branson sonifier and the cell debris was removed from the cleared lysate by centrifuging at 29,000 × g for 20 minutes.

2.4.1 Hydrophobic-interaction chromatography

The cleared lysate (61 mL) was loaded at a flow rate of 2 mL/min onto a 30 mL Phenyl Sepharose 6 Fast Flow High Sub column (XK 16/20, Amersham Biosciences) equilibrated with Buffer A. The unbound protein fraction was removed from the load by washing with 3 × column volumes of Buffer A at a flow rate of 3 mL/min. The bound protein fraction was then eluted using a 2 × column volume (60 mL) linear gradient to 100% Buffer B (20 mM Tris pH-7.5, 0.05 mM EDTA and 10 mM BME); 6 mL fractions were collected. The fractions were analyzed through SDS-PAGE (Figure

2.4.A) and specific activity determination. The fractions containing HKU1 3CL^{pro} were pooled (66 mL).

2.4.2 DEAE anion-exchange chromatography

The pooled fractions from the previous step were loaded at a flow rate of 2 mL/min onto a 60 mL DEAE anion-exchange column (XK 26/20, Amersham Biosciences) equilibrated with Buffer B. The unbound protein fraction was collected. Once the sample was loaded, the column was washed with 1 × column volume of Buffer B at a flow rate of 3 mL/min to remove additional unbound protein fraction. A 8 × column volume (480 mL) linear gradient to 100% Buffer C was finally used to elute the bound protein fraction; 5 mL fractions were collected. Based on SDS-PAGE analysis (Figure 2.4.B) and the specific activity measurements, the majority of HKU1 3CL^{pro} was present in the unbound fraction (flow-through). The DEAE flow-through was concentrated to approximately 10 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore) for next step of purification.

2.4.3 Gel-filtration chromatography

As the final purification step, the concentrated protein sample was loaded onto a prep grade Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with Buffer D. A flow rate of 1 mL/min of buffer D was used for isocratic elution of the protein. The fractions containing HKU1 3CL^{pro} were pooled (Figure 2.4.C) and concentrated to approximately 5 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore). Total activity units (μ M product/min), specific activity (units/mg) and milligrams of protein obtained (BioRad protein assay) were determined after each chromatographic step to calculate the final protein yield. A summary of the percent enzyme yield, total activity units, and the fold-purification after each chromatographic step is summarized in Table 4.3. Approximately 9 mg of highly pure HKU1-CoV 3CL^{pro} can be obtained per liter of bacterial cell culture.

For final storage of the purified HKU1 3CL^{pro} enzyme, 300 μ L protein aliquots containing glycerol to a final concentration of 10% were placed into 1 mL screw-cap vials, flash-frozen under liquid nitrogen and then stored at -80 °C until further use.

Table 2.3.
Purification summary of HKU1 3CL^{pro} from 1 L culture of *E.coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	231	5952	26	1	100
Phenyl-Sepharose	18	1874	105	4	31
DEAE	10	1525	159	6	26
Superdex 75	9	1176	136	5	20

2.5 Expression and purification of OC43-CoV 3CL^{pro}

The gene encoding 3CL^{pro} of OC43 coronavirus 3CL^{pro} was codon optimized for optimal expression in *E.coli* (BioBasic Inc) and was subcloned by Dr. Aimee Egger (Mesecar lab) into pET-11a expression vector with an N-terminal (His)₆-tag and TEV protease cleavage site (Figure 2.1). The construct was verified by DNA sequencing at the Purdue University Genomics Core Facility.

For protein expression, *E. coli* BL21-DE3 electro-competent cells were transformed via electroporation with pET-11a plasmid containing the gene for OC43 3CL^{pro}. The transformed cells were grown overnight at 37 °C on LB agar supplemented with 50 μ g/mL of carbenicillin. A single colony was picked from the agar plate and was used to inoculate 100 mL of LB media supplemented with 100 μ g/mL of carbenicillin. The cells were grown in an incubator shaker at 37 °C overnight at 200 rpm. 50 mL of overnight culture was used to inoculate one liter of Super-LB media supplemented with 100 μ g/mL of carbenicillin and the protein was expressed through auto-induction for 22-24 hours at 25 °C. Cells were harvested by centrifugation.

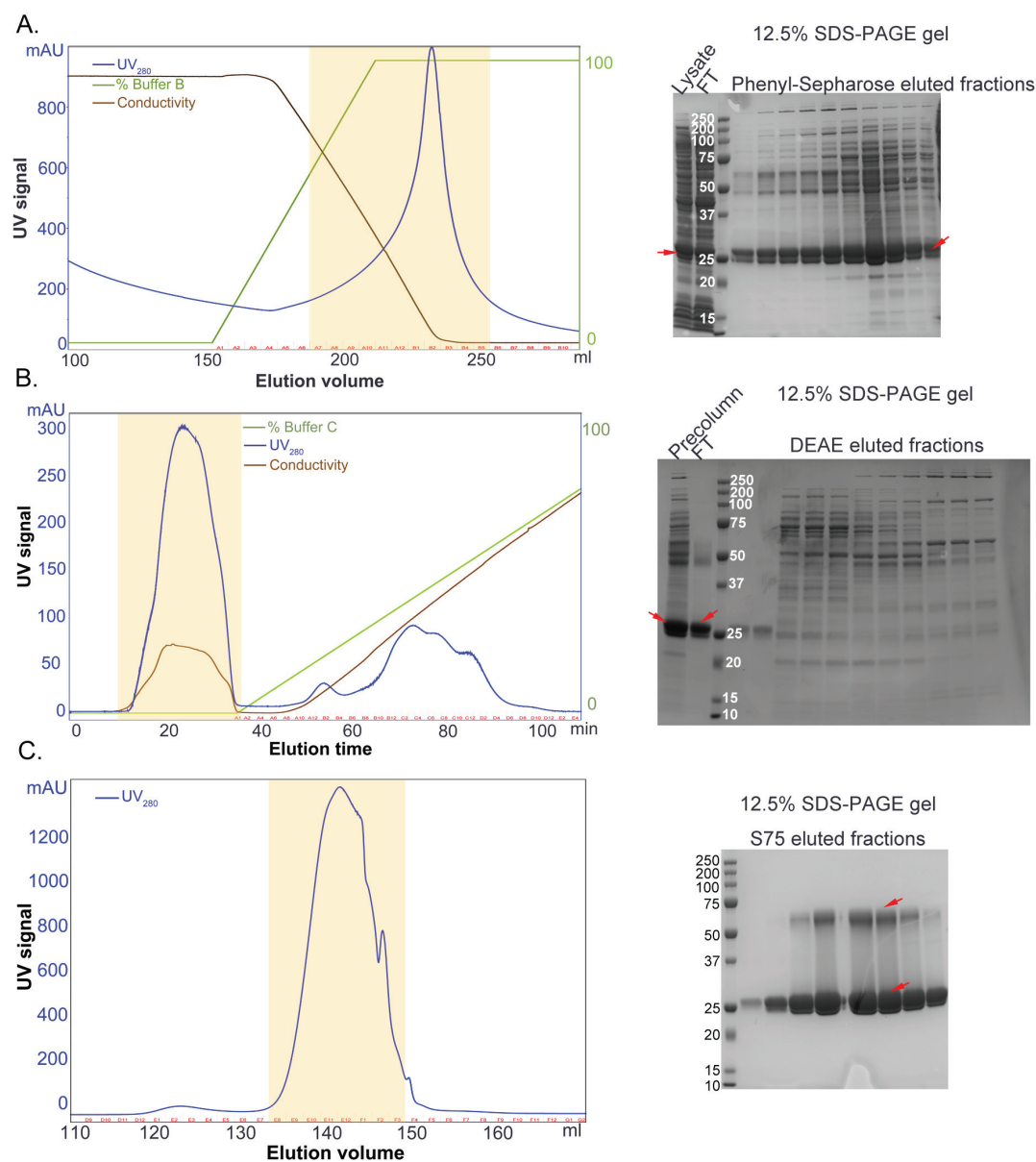


Fig. 2.4.: **Purification of HKU1 3CL^{pro}**. **A.** *Left*-Elution profile from Phenyl-sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrows indicate the expected size of HKU1 3CL^{pro}; FT - flow through (unbound fraction). **B.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **C.** *Left*-Elution profile from Superdex 75 column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile.

gation at $5000 \times g$ for 20 minutes at $4\text{ }^{\circ}\text{C}$, and the pellets were stored at $-80\text{ }^{\circ}\text{C}$ until further use. Approximately 13 grams of cell pellet was obtained from 1 liter of bacterial culture under the given expression conditions.

The frozen pellet from 1 L of bacterial cell culture was thawed on ice and re-suspended in 5 mL of Buffer A per gram of cell pellet with $50\text{ }\mu\text{g}$ of lysozyme and $10\text{ }\mu\text{g}$ of DNase. The cells were then lysed via sonication at 60% amplitude using a 400 W model Branson sonifier and the cell debris was removed from the cleared lysate by centrifuging at $29,000 \times g$ for 20 minutes.

2.5.1 Hydrophobic-interaction chromatography

The clarified lysate was loaded at a flow rate of 2 mL/min onto a 30 mL Phenyl Sepharose 6 Fast Flow High Sub column (XK 16/20, Amersham Biosciences, Piscataway, NJ) equilibrated with Buffer A. Once all the lysate was loaded onto the column, additional buffer A was passed through the column at 3 mL/min until UV_{280} stabilized. The unbound protein fractions under distinct peaks were collected in separate beakers. Once UV_{280} stabilized, gradient to 60% Buffer B was immediately started and 10 ml fractions were collected. A gradient of 60% Buffer B eluted the majority of OC43 3CL from the column. Once UV_{280} went down and stabilized again, the gradient was switched from 60% to 100% of Buffer B in $1 \times$ column volume of Buffer B (30 mL). The fractions were analyzed through SDS-PAGE (Figure 2.5.A) and the specific activity determination. The fractions containing OC43 3CL^{Pro} were pooled and exchanged into 2 L of Buffer C via overnight dialysis in a dialysis tubing (10,000 MWCO SnakeSkin[®], Thermo Scientific).

2.5.2 DEAE anion-exchange chromatography

The dialyzed sample from the previous step was loaded at a flow rate of 2 mL/min onto a 60 mL DEAE anion-exchange column (XK, 26/20, Amersham Biosciences) equilibrated with Buffer C. The unbound protein fraction was collected. Once the

sample was loaded, the column was washed with additional Buffer C at a flow rate of 3 mL/min until UV_{280} stabilized. A $2 \times$ column volume (120 mL) linear gradient to 100% Buffer D was finally used to elute the bound protein fraction; 10 mL fractions were collected. Based on SDS-PAGE analysis (Figure 2.5.B) and the specific activity measurements, the majority of OC43 3CL^{pro} was present in the unbound protein fraction (flow-through). The DEAE flow-through was concentrated to approximately 10 mg/mL Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore) for next step of purification.

2.5.3 Gel-filtration chromatography

As the final purification step, the concentrated protein sample was loaded onto a prep grade Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with Buffer E. A flow rate of 1 mL/min of buffer E was used for isocratic elution of the protein. The fractions containing OC43 3CL^{pro} were pooled (Figure 2.5.C) and concentrated to approximately 5 mg/mL. Total activity units (μ M product/min), specific activity (units/mg) and milligrams of protein obtained (BioRad protein assay) were determined after each chromatographic step to calculate the final protein yield. A summary of the percent enzyme yield, total activity units, and the fold-purification after each chromatographic step is summarized in Table 4.4. Approximately 30 mg of highly pure OC43 3CL^{pro} can be obtained per liter of bacterial cell culture.

For final storage of the purified OC43 3CL^{pro} enzyme, 300 μ L protein aliquots containing glycerol to a final concentration of 10% were placed into 1 mL screw-cap vials, flash-frozen under liquid nitrogen and then stored at -80 °C until further use.

2.6 Expression and purification of HKU5-CoV 3CL^{pro}

The gene encoding 3CL^{pro} of HKU5-CoV 3CL^{pro} was codon optimized for optimal expression in *E.coli* and was subcloned by BioBasic Inc into pET-11a expression

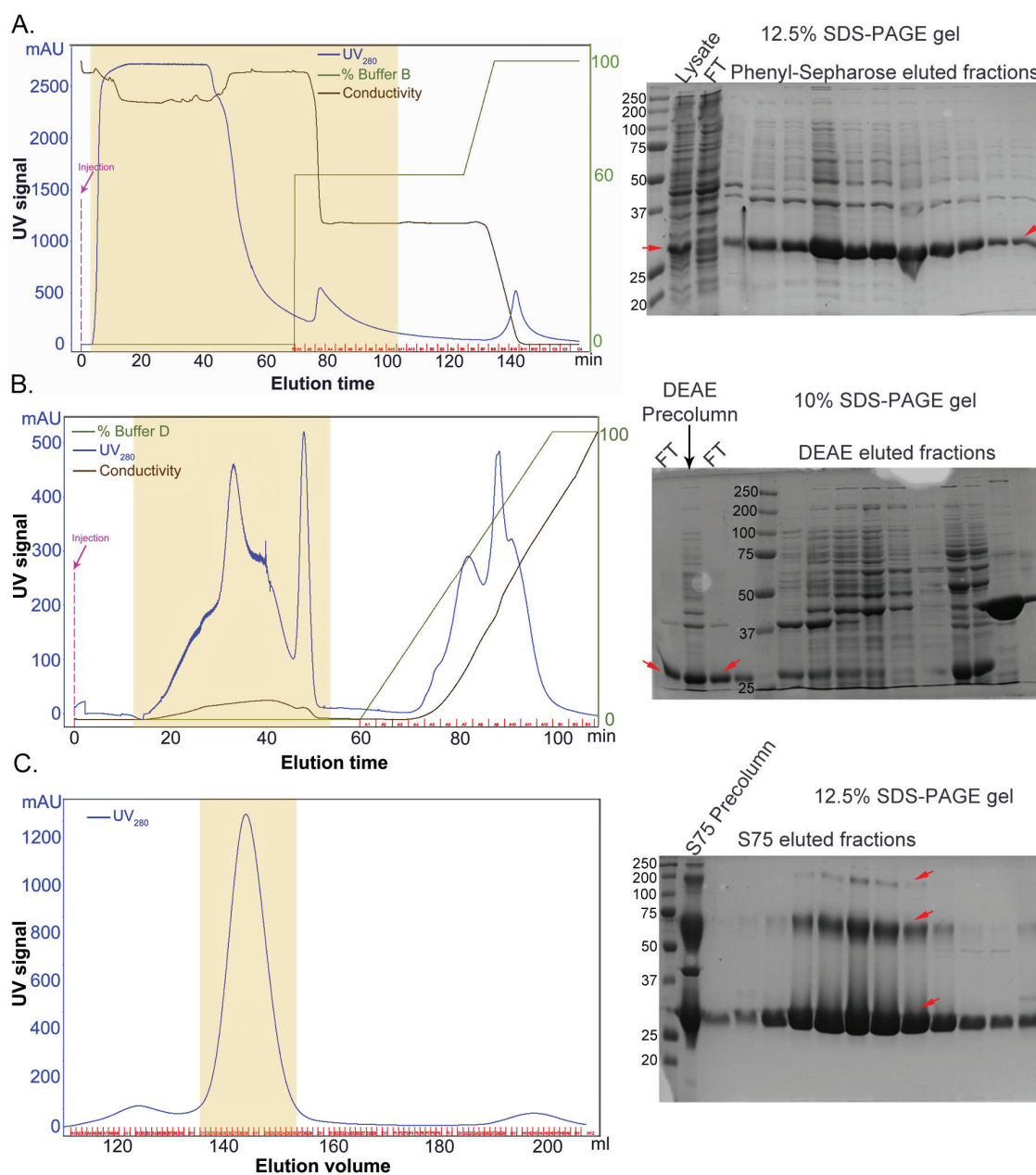


Fig. 2.5.: **Purification of OC43 3CL^{PRO}.** **A.** *Left*-Elution profile from Phenyl-sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrows indicate the expected size of OC43 3CL^{PRO}. **B.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **C.** *Left*-Elution profile from Superdex 75 column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. FT - flow through (unbound fraction).

Table 2.4.
Purification summary of OC43 3CL^{pro} from 1 L culture of *E. coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	1014	7811	8	1	100
Phenyl-Sepharose	79	4579	58	8	59
DEAE	39	6634	172	22	85
Superdex 75	30	1547	51	7	20

vector with an N-terminal (His)₆-tag followed by nsp4|nsp5 auto-cleavage site. This construct results in the expression of HKU5-CoV 3CL^{pro} without any N-terminal or C-terminal extension.

For protein expression, *E. coli* BL21-DE3 electro-competent cells were transformed via electroporation with pET-11a plasmid containing the gene for HKU5-CoV 3CL^{pro} as described above. The transformed cells were grown overnight at 37 °C on LB agar supplemented with 50 µg/mL of carbenicillin. A single colony was picked from the agar plate and was used to inoculate 100 mL of LB media supplemented with 100 µg/mL of carbenicillin. The cells were grown in an incubator shaker at 37 °C overnight at 200 rpm. 25 mL of overnight culture was used to inoculate one liter of Super-LB media supplemented with 100 µg/mL of carbenicillin and the protein was expressed through auto-induction for 22-24 hours at 25 °C. The cells were harvested by centrifugation at 5000 × g for 20 minutes at 4 °C, and the pellets were stored at −80 °C until further use. Approximately 12 grams of cell pellet was obtained from 1 L of bacterial culture under the given expression conditions.

the frozen pellet from 1 L of bacterial cell culture was thawed on ice and re-suspended in 4 mL (50 mL) of Buffer A per gram of cell pellet with 100 µg of lysozyme and a small amount of DNase. The cells were then lysed using a single pass through French press at 1200 psi and the cell debris was removed from the cleared lysate by centrifuging at 29,000 × g for 30 minutes. Solid ammonium sulfate was added to

the cleared lysate (50 mL) to a final concentration of 1 M (3.56 gm) through gradual mixing on ice using a magnetic stirrer.

2.6.1 Hydrophobic-interaction chromatography

The cleared lysate, mixed with ammonium sulfate, was loaded at a flow rate of 3 mL/min onto a 30 mL Phenyl Sepharose 6 fast-flow high-sub column (XK 16/20, Amersham Biosciences, Piscataway, NJ) equilibrated with Buffer B. The column was then washed with $5 \times$ column volume (150 mL) of Buffer B at a flow rate of 3 mL/min to remove the unbound protein fraction. The bound protein fraction was eluted using a $5 \times$ column volume (150 mL) linear gradient to 100% Buffer A. The fractions (5 mL) were collected and those containing HKU5-CoV 3CL^{pro} as judged through SDS-PAGE analysis (Figure 2.6.A) and the specific activity measurements, were pooled (55 mL) and exchanged into 2 L of Buffer A via overnight dialysis in a dialysis tubing (10,000 MWCO SnakeSkin[®], Thermo Scientific).

2.6.2 DEAE anion-exchange chromatography

The dialyzed sample from the previous step was loaded at a flow rate of 3 mL/min onto a 60 mL DEAE anion-exchange column (XK 226/20, Amersham Biosciences) equilibrated with Buffer A. The column was then washed with 2 column volume (120 mL) of Buffer A at a flow rate of 3 mL/min. A linear gradient to 50% Buffer C in $5 \times$ column volume (300 mL) was used to elute the bound protein fraction (Figure 2.6.B). The fractions (5 mL) were collected and those containing HKU5-CoV 3CL^{pro} were pooled (50 mL) and dialyzed for 4 hours in 4 L of Buffer A using a dialysis tubing (10,000 MWCO SnakeSkin[®], Thermo Scientific).

2.6.3 Mono-Q anion-exchange chromatography

Following dialysis, the protein sample was filtered by passing through a 0.22 μm pore size Millex-GP filter (Millipore) to remove any precipitated protein. The filtered sample was then loaded at a flow-rate of 1 mL/min onto a 8 mL Mono-Q 10/100 column (Amersham Biosciences) equilibrated in Buffer A. The protein was eluted at a flow rate of 2 mL/min using a 25 \times column volume (200 mL) and linear gradient to 50% Buffer B (Figure 2.6.C). The fractions (2 mL) were collected and those containing HKU5-CoV 3CL^{pro} were pooled and dialyzed overnight in Buffer D. For final storage of the purified HKU5-CoV 3CL^{pro} enzyme, 300 μL protein aliquots were placed into 1 mL screw-cap vials, flash-frozen under liquid nitrogen and then stored at $-80\text{ }^{\circ}\text{C}$ until further use for kinetic characterization.

A summary of the percent enzyme yield, total activity units, and the fold-purification after each chromatographic step is summarized in Table 4.5. Approximately 28 mg of highly pure HKU5-CoV 3CL^{pro} can be obtained from 1 liter of bacterial cell culture.

Gel-filtration chromatography was employed as the final purification step only to prepare the protein sample for crystallization.

Table 2.5.
Purification summary of HKU5-CoV 3CL^{pro} from 1 L culture of *E.coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	486	9407	19	1	100
Phenyl-Sepharose	84	4857	58	3	52
DEAE	37	4459	122	6	47
Mono-Q	28	3767	136	7	40

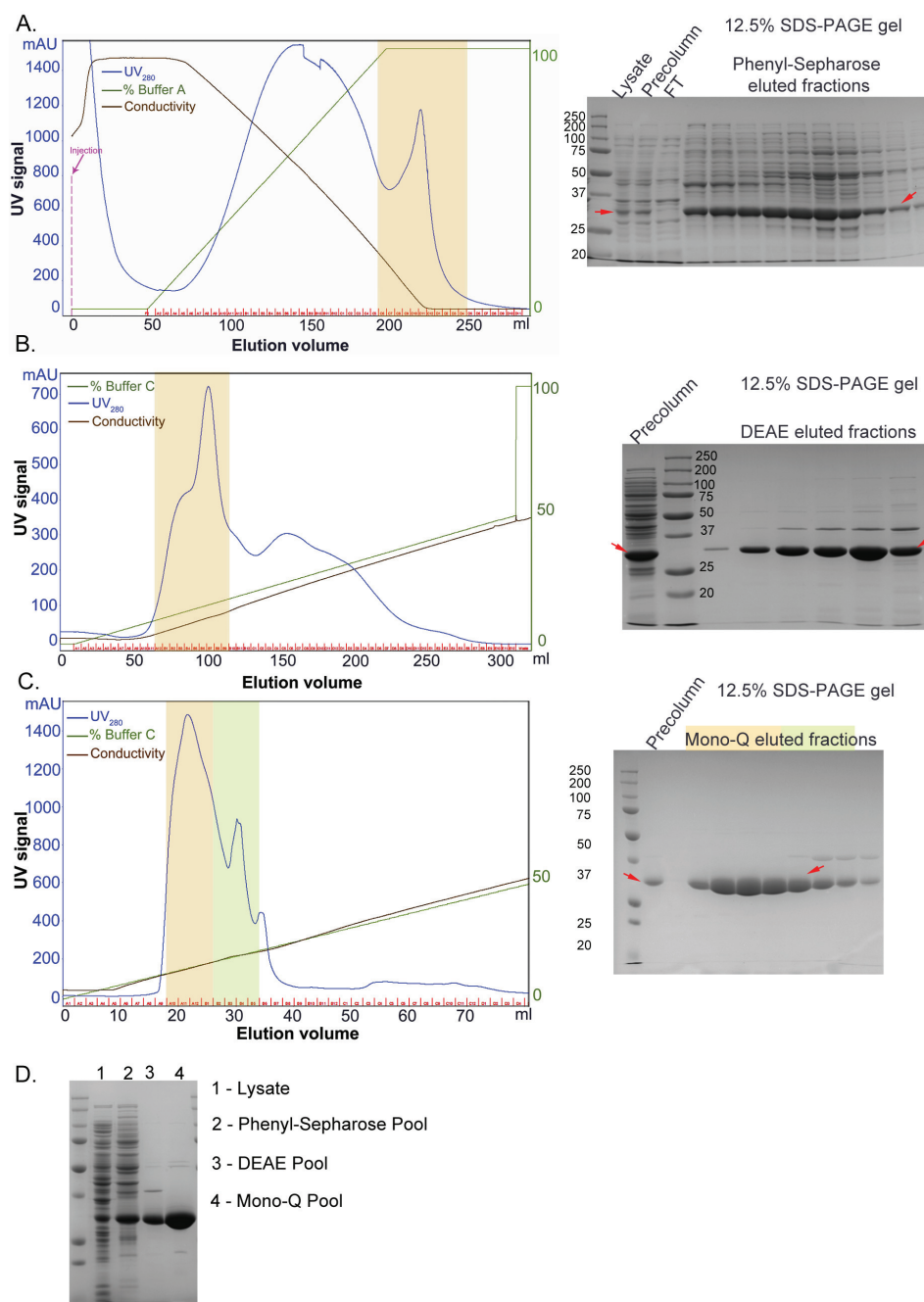


Fig. 2.6.: **Purification of HKU5-CoV 3CL^{Pro}**. **A.** *Left*-Elution profile from Phenyl-Sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrows indicate the expected size of HKU1 3CL^{Pro}; FT - flow through (unbound fraction). **B.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **C.** *Left*-Elution profile from Mono-Q column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **D.** Final gel showing sample purification achieved after each purification step.

2.7 Expression and purification of MERS-CoV 3CL^{pro}

The expression and purification protocol for MERS-CoV 3CL^{pro} has been described in details in Chapter 5.

2.8 Sample preparation for SDS-PAGE analysis

Some of the 3CL^{pro} enzymes we investigated exhibit a peculiar pattern when analyzed on the SDS-PAGE. We observed that boiling of the protein for SDS-PAGE sample preparation results in the formation of high-order protein aggregates displaying a ladder-like pattern (Figure 2.7). Western-blot analysis of MHV 3CL^{pro} using an anti-3CL^{pro} antibody confirmed that the higher molecular weight bands on the SDS-PAGE are in fact higher order aggregates of MHV 3CL^{pro} monomer. Therefore, samples were prepared by mixing the sample volume containing 5-10 μg protein with appropriate volume of SDS-PAGE loading buffer containing 1 mM of fresh DTT, instead of BME as the reducing agent.

Recipe for 5 \times SDS sample loading buffer

0.2 M Tris-HCl, pH-6.8

10% SDS

10 (or 20)% glycerol

0.02% Bromophenolblue

5 mM DTT (added fresh from 1M stock, right before preparing samples for running on SDS-PAGE)

For MHV 3CL^{pro}, samples were then heated at 70 °C for 2 minutes. Heating of the samples was completely avoided for HKU1 3CL^{pro} and OC43 3CL^{pro} as it results in the formation of protein aggregates that accumulate in the stacking gel. For MERS-CoV 3CL^{pro} and HKU5-CoV 3CL^{pro}, samples were heated at 95 °C for 2 minutes; while samples for SARS-CoV 3CL^{pro} were boiled at 95 °C for 5 minutes.

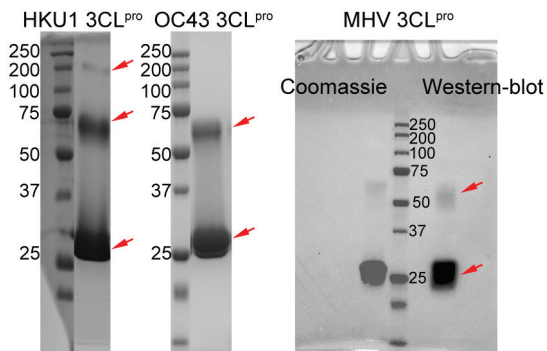


Fig. 2.7.: **Formation of 3CL^{pro} high-order aggregates on SDS-PAGE.** High-order aggregate formation during SDS-PAGE analysis of 3CL^{pro} enzymes from HKU1, OC43 and MHV is shown. Red arrows indicate expected molecular weights for monomer, and higher-order 3CL^{pro} aggregates. Right panel - Western blot analysis of MHV 3CL^{pro} using anti-3CL^{pro} antibody is shown alongside coomassie staining for the sample.

2.9 Concentration determination

Protein concentration was determined using two different methods depending upon the purification of the sample. Protein concentration was followed during protein purification using a Bradford assay that was performed in a 96-well, clear flat-bottom microtiter plate. For purified samples, protein concentration was determined by measuring the absorbance at 280 nm and the theoretical molar extinction coefficient of the protein determined through ExPASy's ProtParam tool.

2.9.1 Mini-bradford assay

- Dye reagent was prepared by diluting concentrate dye (from BIORAD) with DI water in 1:4 ratio.

- BIORAD BSA stock was prepared in DI water with final BSA concentration of 1.36 mg/mL. Next, 0.5 mg/mL working stock of BSA was prepared by diluting 36.8 μ L of BIORAD BSA stock with 63.2 μ L of DI water.

- Five dilutions of BSA standard were prepared by dispensing 0, 2.5, 5.0, 7.5 and 10 μ L of BSA working stock into five different wells of a 96-well clear flat-bottom

microtiter plate. These volumes correspond to 0, 0.125, 0.25, 0.375 and 0.5 mg/mL of BSA, respectively for the standard curve.

- 10 μ L of each sample was dispensed in the microtiter plate for duplicate readings, followed by addition of 200 μ L of dye reagent to the sample wells as well as the BSA standard wells. Samples and the dye were mixed thoroughly for 1-2 minutes using a microplate mixer.

- Absorbance of the samples was measured at 595 nm using BioTek Synergy H1 plate reader. If OD₅₉₅ of any sample was not within the linear range for BSA standards, OD₅₉₅ was measured again after adjusting the sample volume dispensing in the well.

- A linear trendline was created for BSA standards with OD₅₉₅ on the y-axis and concentration (mg/mL) on the x-axis. Finally, calculated slope and intercept of the trendline were used to calculate the concentration of other protein samples.

2.9.2 OD₂₈₀ based determination of protein concentration

The optical density at 280nm (OD₂₈₀) of the reference buffer was measured using a 1 cm pathlength cuvette. Next, the OD₂₈₀ of purified protein sample was measured after dilution in the reference buffer. The final OD₂₈₀ value was calculated after subtracting the reference OD value from the OD value for protein sample. The protein concentration (M) was then determined using the Beer-Lambert equation. The values for molar extinction coefficient were calculated from the primary sequence of the protein using Expasy's Protparam tool. Molar extinction coefficient values for different 3CL^{pro} enzymes are provided in Table 2.6.

2.10 Kinetic assays

The enzymatic activity of 3CL^{pro} enzymes was measured using the following custom synthesized peptide: (HilyteFluorTM-488)-ESATLQSGLRKAK-(QXLTM-520)-NH₂ (AnaSpec, Inc.). The HilyteFluorTM-488 fluorescence group is internally quenched

Table 2.6.
Molar extinction coefficient values for different 3CL^{pro} enzymes
calculated using ExPASy ProtParam tool

3CL ^{pro}	Mol.Wt. (kDa)	Extinction coefficient (M ⁻¹ .cm ⁻¹)
SARS-CoV	33845.7	32890
MHV	33158.9	41370
HKU1-CoV	33199.1	45840
OC43-CoV	33477.4	42860
HKU5-CoV	33459.4	42400
MERS-CoV	33330.2	43890

by QXLTM-520 dye. This substrate works as a generic peptide substrate for 3CL^{pro} enzymes and was designed based on the nsp4|nsp5 cleavage sequence for many coronavirus 3CL^{pro} enzymes. The rate of enzymatic activity was determined at 25 °C by following the increase in fluorescence ($\lambda_{\text{excitation}} = 485 \text{ nm}$, $\lambda_{\text{emission}} = 528 \text{ nm}$, bandwidths = 20 nm) of Hilyte Fluor-488 upon peptide hydrolysis by the enzyme as a function of time. Assays were conducted in black, half-area, 96-well plates (Corning) in assay buffer (50 mM HEPES pH-7.5, 0.1 mg/mL BSA, 0.01% Triton X-100 and 2 mM DTT) using a final reaction volume of 100 μL . The resulting fluorescence was monitored using a BioTek Synergy H1 plate reader. The rate of the reaction in arbitrary fluorescence units per sec (AFU/sec) was determined by measuring the initial slope of the progress curves, which were then converted to units of μM of product produced per min ($\mu\text{M}/\text{min}$) using experimentally determined values of ‘fluorescence extinction coefficient’. All reactions were carried out in triplicate.

2.10.1 Determination of fluorescence extinction coefficient for the substrate UIVT3

To calculate the value of fluorescence extinction coefficient, 100 μL reactions were set up. Substrate concentrations were determined gravimetrically. 20 μL of 5 \times working stocks of varying substrate concentrations were dispensed in duplicate in

black, half-area, 96-well plates. The final substrate concentrations varied over the range from 0 to 2 μM . 80 μL of high concentration of 3CL^{pro} enzyme prepared in assay buffer was dispensed in one set of reaction wells with substrate. 80 μL of assay buffer without enzyme was mixed in second set of reaction wells. The resulting fluorescence was monitored using a BioTek Synergy H1 plate reader. The maximum fluorescence released after complete turnover of the substrate into products was measured for each substrate concentration. the background fluorescence released at each substrate concentration was also measured from the reaction wells without enzyme. ΔAFU was calculated by subtracting background AFU from maximum AFU at each substrate concentration. The fluorescence extinction coefficients ($\text{AFU}\cdot\mu\text{M}^{-1}$) can be determined from the slope of the line that results from a plot of ΔAFU (y-axis) against the substrate concentration (x-axis).

2.10.2 Determination of enzymatic efficiency

The apparent enzymatic efficiency ($k_{\text{cat}}/K_{\text{M}}$) for each of the 3CL^{pro} enzymes was determined by measuring the rate of enzymatic activity as a function of varying substrate concentration in 100 μL reactions. Reactions were initiated by the addition of enzyme to the wells of an assay plate containing varying concentrations of substrate. The final substrate concentrations varied over the range from 0 to 2 μM . The final enzyme concentrations for each 3CL^{pro} studied were the following: SARS-CoV 3CL^{pro} at 100 nM, MHV 3CL^{pro} at 100 nM, HKU1 3CL^{pro} at 100 nM, OC43 3CL^{pro} at 100 nM, HKU5-CoV 3CL^{pro} at 250 nM, HKU4-CoV 3CL^{pro} at 200 nM and MERS-CoV 3CL^{pro} at 1 μM . Since 3CL^{pro} enzymes cannot be saturated with this substrate at a substrate concentration that would still allow accurate fluorescent measurements without the inner filter effect, only the apparent $k_{\text{cat}}/K_{\text{M}}$ values can be determined from the slope of the line that results from a plot of the enzymatic activity (y-axis), normalized for the total enzyme concentration, against the substrate concentration (x-axis). Under the given assay conditions, MHV 3CL^{pro} and HKU1 3CL^{pro} enzymes

were the most efficient enzymes (Table 2.7), while 3CL^{pro} from MERS-CoV was the least efficient enzyme (Table 2.7). Enzymatic efficiency of MERS-CoV 3CL^{pro} was 10-fold lower than the enzymatic activity of 3CL^{pro} enzymes from MHV and HKU1, and 5-fold lower than that of SARS-CoV 3CL^{pro} (Table 2.7). Therefore, the factors that might cause a decrease in the enzymatic activity of MERS-CoV 3CL^{pro} were further explored and have been discussed in Chapter 5.

Table 2.7.
Enzymatic efficiency (apparent $k_{\text{cat}}/K_{\text{M}}$) for different 3CL^{pro} enzymes

3CL ^{pro}	apparent $k_{\text{cat}}/K_{\text{M}}$ ($\times 10^2$) ($\mu\text{M}^{-1}.\text{min}^{-1}$)
SARS-CoV	15.5 \pm 0.9
MHV	34.1 \pm 0.6
HKU1-CoV	31.0 \pm 0.5
OC43-CoV	25.1 \pm 2.6
HKU5-CoV	8.8 \pm 0.1
MERS-CoV	3.1 \pm 0.03

2.11 Inhibition assays

Based on initial hits from a high-throughput screen of $\sim 300,000$ compounds conducted to develop small-molecule reversible inhibitors of SARS-CoV 3CL^{pro} (see Chapter 3 for details), our collaborators synthesized a focused library of 237 compounds. To determine the inhibitory activity of these compounds for 3CL^{pro} enzymes, the % Inhibition of 3CL^{pro} in the presence of 100 μM of compound was first determined. For compounds displaying more than 50% inhibition, the IC₅₀ values were determined from a dose response curve obtained by calculating % Inhibition of 3CL^{pro} over a range of concentrations of the compound.

2.11.1 % inhibition determination

To determine the percent inhibition for compounds, the total concentration of the substrate was fixed at 2.0 μM , and the enzymes were fixed at 250 nM for SARS-CoV 3CL^{pro} and OC43 3CL^{pro}, 80 nM for MHV 3CL^{pro} and HKU1 3CL^{pro}, and 1 μM for HKU5-CoV 3CL^{pro} and MERS-CoV 3CL^{pro}. Next, 100 \times working stocks (10 mM) of the compounds were prepared from 40 mM stocks by diluting in DMSO. DMSO working stocks of the compounds were then diluted hundred-fold (1 μL) to a final concentration of 100 μM in 80 μL of the enzyme solution in the wells of black, 96-well, half-area plates, and incubated for 10 minutes after mixing on a plate-mixer. After 10 minutes, the enzymatic activity was measured as the initial slope of the progress curve, obtained by initiating the reaction with 20 μL of 10 μM substrate. The % Inhibition was calculated using Equation 2.1.

$$\% \text{ Inhibition} = \left[1 - \frac{\text{Rate}_{\text{sample}} - \text{Rate}_{\text{neg}}}{\text{Rate}_{\text{pos}} - \text{Rate}_{\text{neg}}} \right] \times 100 \quad (2.1)$$

In Equation 2.1, $\text{Rate}_{\text{sample}}$ is the initial slope of the progress curve in AFU/sec measured in the presence of the compound, Rate_{pos} is the initial slope measured in the absence of any compound and Rate_{neg} is the baseline substrate hydrolysis calculated in the absence of enzyme. All the reactions were carried out in triplicate and contained a final DMSO concentration of 1%.

2.11.2 IC₅₀ determination

For compounds displaying more than 50% inhibition, a more extensive characterization was conducted. The IC₅₀ value refers to the inhibitor concentration required to achieve half maximum inhibition of the enzyme. The IC₅₀ values for the inhibitors were determined through measuring a dose response with subsequent analysis. To determine the IC₅₀ value, 100 \times working stocks over a range of concentration for the

compounds were prepared from serial dilutions of 40 mM DMSO stocks. Next, DMSO working stocks of the compounds were diluted one hundred-fold, to obtain their respective final concentrations in 80 μL of the enzyme solution. The compounds were allowed to incubate for 10 minutes. After 10 minutes, the enzymatic activity was measured as initial slope of the progress curve, obtained by initiating the reaction with 20 μL of 10 μM substrate. The final concentrations of 3CL^{PRO} enzymes for IC₅₀ determination were kept the same as used for % Inhibition determination. The final substrate concentration in the reaction well was 2.0 μM , and the inhibitor concentration ranged from 120 μM to 0.313 μM . The IC₅₀ values were calculated by fitting the data into Equation 2.2 using non-linear regression program SigmaPlot.

$$\% \text{ Inhibition} = \frac{\text{Max.}\% \text{ Inhibition} \times [\text{Inhibitor}]}{[\text{Inhibitor}] + \text{IC}_{50}} \quad (2.2)$$

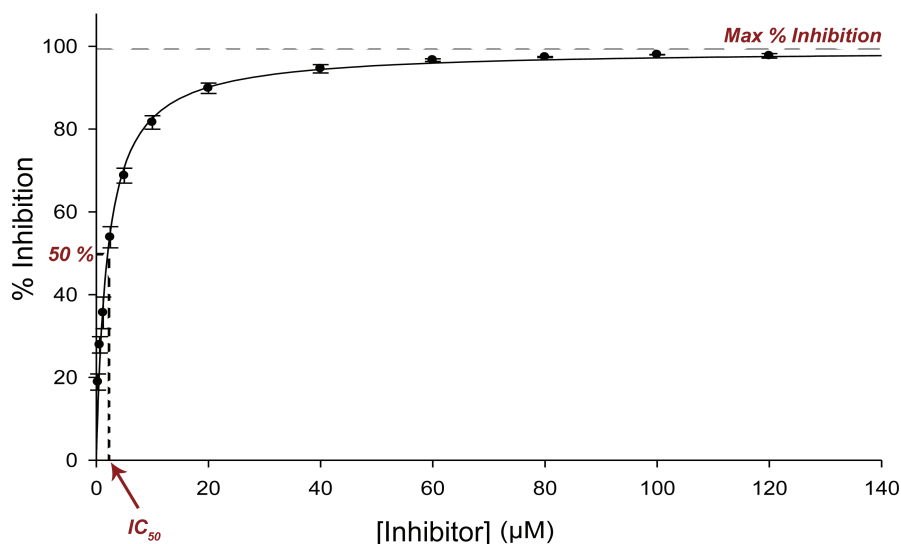


Fig. 2.8.: **Representative dose response curve for IC₅₀ determination.** Maximum % inhibition represents the extrapolated value for % inhibition obtained from fitting the % Inhibition versus [Inhibitor] data into Equation 2.2. IC₅₀ value represents the concentration of inhibitor required to achieve 50% of Maximum inhibition.

A representative dose response curve for IC₅₀ determination is shown in Figure 2.8. The inhibitory activity of inhibitors is discussed in detail in Chapter 3.

2.12 Crystallization of SARS-CoV 3CL^{pro} in complex with inhibitors

The purified SARS-CoV 3CL^{pro} was concentrated to 5.0-2.5 mg/mL in buffer (25 mM HEPES pH-7.5 with 2.5 mM DTT). Inhibitor complexes of SARS-CoV 3CL^{pro} are listed in Table 2.8 and were formed by incubating SARS-CoV 3CL^{pro} with the compounds in a 1:3 stoichiometric ratio at 4 °C for at least one hour. After iterative rounds of optimization of protein concentration, drop ratio and the incubation temperature, the crystals of SARS-CoV 3CL^{pro}inhibitor complexes suitable for X-ray diffraction were grown by the hanging-drop, vapor diffusion method at 4 °C using 2 μ L:1 μ L protein:reservoir buffer ratio. Crystallization conditions for all the SARS-CoV 3CL^{pro}:Inhibitor complexes are summarized in Table 2.8. All the crystallization solutions were kept in the cold room before setting up the crystal trays. For X-ray data collection, crystals were flash-cooled in liquid nitrogen after dragging the crystals through a cryo-solution that contained the crystallization solution supplemented with 15% 2-methyl-2,4-pentanediol (MPD) and 500 μ M of corresponding inhibitor. Crystallization conditions for SARS-CoV 3CL^{pro}-Inhibitor complexes are summarized in Table 2.8.

High-resolution diffraction quality crystals could also been obtained for MERS-CoV 3CL^{pro} in complex with both the non-covalent and covalent inhibitors. Details for the crystallization, X-ray data collection and structure refinement of the MERS-CoV 3CL^{pro}-Inhibitor complexes are provided in Chapter 5.

2.13 Data collection and structure refinement

2.13.1 Procedure

The X-ray diffraction data were collected for the SARS-CoV 3CL^{pro}-Inhibitor complexes at the Life Sciences Collaborative Access Team (LS-CAT) Sector 21 at the Advanced Photon Source, Argonne National Laboratory. The data were processed and scaled using HKL2000 version 706 (46). The method of rigid body refinement

Table 2.8.
Crystallization conditions for SARS-CoV 3CL^{Pro}-Inhibitor complexes

Inhibitor	Reservoir buffer	[Protein]	[Inhibitor]	Drop ratio	Cryo-solution
CQ3	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 8% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M CQ3
R30	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 7% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M R30
R2Y	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 7% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M R2Y
R2X	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 8% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M R2X
886	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 6.5% PEG-10,000, 3 mM DTT	2.5 mg/mL	250 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M 886
CUG	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 8% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M CUG
FMW	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 8% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M FMW
KWW	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 6.5% PEG-10,000, 3 mM DTT	2.5 mg/mL	250 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M KWW
XM2	0.05 M MES pH-6.0, 0.04 M KCl, 1% MPD, 7% PEG-10,000, 3 mM DTT	5 mg/mL	500 μ M	2:1	Reservoir buffer + 15% MPD + 500 μ M XM2

was used to obtain the initial phases using the program Refmac5 in ccp4 suite version 6.2.0. The X-ray structure of SARS-CoV 3CL^{Pro} in complex with REY (PDB ID: 3V3M) was used as a phasing model (32). The inhibitor molecules were manually added to the corresponding pdb files using COOT and then refined in the Phenix suite. Each structure was then refined using iterative cycles of refinement using Phenix Refine coupled to manual model building using COOT (48) based on the $F_o - F_c$ and $2F_o - F_c$ maps. The coordinates and molecular library files for inhibitor molecules were built using the program Sketcher in the ccp4 suite. Water molecules were added to the peaks in residual ($F_o - F_c$) density maps that were greater than

3σ using the “Find Water” function in COOT. MolProbity was used to assess the structural quality of the final model (49). Pymol was used to generate the figures of all the structures (50).

2.13.2 Results

High-resolution crystal structures of SARS-CoV 3CL^{pro} for all the inhibitors listed in Table 2.8 were determined. All the SARS-CoV 3CL^{pro}-inhibitor complexes crystallized in the space group *C2* with one molecule in the asymmetric unit. Summaries of the statistics for the X-ray data collection and refinement are provided in Tables 2.9-2.13. Analysis of inhibitor binding in the X-ray crystal structures is discussed in detail in Chapter 3.

2.14 Summary

Coronavirus 3CL^{pro} is a high profile target for the development anti-coronaviral therapeutics. In our effort to develop broad spectrum inhibitors of 3CL^{pro} enzymes from different coronavirus classes, we aimed to characterize 3CL^{pro} enzymes from SARS-CoV, MHV, HKU1, OC43, HKU5-CoV and MERS-CoV to understand their similar and distinct features. Towards this goal, we report successful expression and purification of these 3CL^{pro} enzymes. For protein expression, the gene sequence was codon-optimized in order to achieve optimal expression of 3CL^{pro} enzymes in *E. coli* BL21-DE3 cells. 3CL^{pro} gene was cloned into pET-11a expression vector with an N-terminal His₆-tag and either a TEV protease cleavage site or 3CL^{pro} auto-cleavage site between the tag and 3CL^{pro} gene sequence. This construct results in the expression of 3CL^{pro} with authentic N-terminus during expression in bacterial cells. For protein expression, a combination of chromatographic steps including hydrophobic-interaction, ion-exchange and gel-filtration were utilized to obtain highly pure 3CL^{pro} enzymes. Using the purification protocols described in this chapter, sufficient quantities of all the 3CL^{pro} enzymes were achieved for kinetic characterization, inhibition

Table 2.9.
X-ray data collection and refinement statistics for SARS-CoV 3CL^{pro} in
complex with CQ3 and R30

	Beamline: LS-CAT Sector 21 ID-F	
	CQ3	R30
Data collection		
Wavelength (Å)	0.97872	0.97872
Resolution range (Å)	100-1.28 (1.30-1.28) ^a	50.00-1.33 (1.35-1.33) ^a
Protein monomers in asymmetric units	1	1
Space group	<i>C2</i>	<i>C2</i>
Unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	107.93, 82.42, 53.49	108.16, 82.01, 53.49
α , β , γ (°)	90, 104.03, 90	90, 104.07, 90
Total number of reflections	720,003	681,777
Number of unique reflections	116,580	103,693
Multiplicity	3.8 (2.6) ^a	3.9 (4.0) ^a
Completeness (%)	97.0 (90.6) ^a	99.1 (99.9) ^a
Mean $I/\sigma I$	12.34 (1.97) ^a	13.46 (3.63) ^a
R_{merge} (%) ^b	11.3 (50.9) ^a	7.9 (35.4) ^a
Refinement		
Resolution range (Å)	64.76-1.28	26.23-1.33
Number of reflections in working set	94,233	96,818
Number of reflections in test set	4,977	4,851
R_{work} (%) ^c	16.6	17.7
R_{free} (%) ^c	18.4	19.9
Number of non-hydrogen atoms	3,118	3,125
Protein / water	2,577/472	2,499/566
RMSD - bond lengths (Å)	0.012	0.010
RMSD - bond angles (°)	1.84	1.24
Ramachandran favored (%)	99	99
Ramachandran outliers (%)	0	0
Molprobit clash score	4.7	2.1
Average B-factor (Å ²)	20	21
Protein	18	18
Ligands	29	22
Solvent	31	32

^a Values in parentheses are for highest-resolution shell.
^b $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$.
^c R_{work} and $R_{\text{free}} = h(|F(h)_o| - |F(h)_c|) / h|F(h)_o|$ for reflections in the working and test sets, respectively.

Table 2.10.
X-ray data collection and refinement statistics for SARS-CoV 3CL^{pro} in
complex with R2Y and R2X

	Beamline: LS-CAT Sector 21 ID-F	
	R2Y	R2X
Data collection		
Wavelength (Å)	0.97872	0.97872
Resolution range (Å)	100-1.21 (1.23-1.21) ^a	50.00-1.33 (1.35-1.33) ^a
Protein monomers in asymmetric units	1	1
Space group	<i>C</i> 2	<i>C</i> 2
Unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	108.05, 82.1, 53.43	108.25, 81.79, 53.43
α , β , γ (°)	90, 104.19, 90	90, 104.06, 90
Total number of reflections	181,3044	758,292
Number of unique reflections	137,251	103,427
Multiplicity	8.2 (7.7) ^a	4.0 (4.0) ^a
Completeness (%)	99.7 (100.0) ^a	99.4 (99.7) ^a
Mean <i>I</i> / σ <i>I</i>	17.92 (6.75) ^a	15.47 (3.39) ^a
<i>R</i> _{merge} (%) ^b	11.3 (31.4) ^a	7.4 (36.3) ^a
Refinement		
Resolution range (Å)	25.87-1.21	64.52-1.33
Number of reflections in working set	135,507	91,361
Number of reflections in test set	6,818	4,833
<i>R</i> _{work} (%) ^c	12.6	16.5
<i>R</i> _{free} (%) ^c	14.8	18.1
Number of non-hydrogen atoms	3,220	3,116
Protein / water	2,535/625	2,547/510
RMSD - bond lengths (Å)	0.013	0.012
RMSD - bond angles (°)	1.51	1.85
Ramachandran favored (%)	99	99
Ramachandran outliers (%)	0	0
Molprobit clash score	5.7	4.7
Average B-factor (Å ²)	24.30	19.50
Protein	20	17
Ligands	32	22
Solvent	40	30

^a Values in parentheses are for highest-resolution shell.
^b $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$.
^c R_{work} and $R_{\text{free}} = h(|F(h)_o| - |F(h)_c|) / h|F(h)_o|$ for reflections in the working and test sets, respectively.

Table 2.11.
X-ray data collection and refinement statistics for SARS-CoV 3CL^{pro} in
complex with 886 and CUG

	Beamline: LS-CAT Sector 21 ID-F	
	886	CUG
Data collection		
Wavelength (Å)	0.97872	0.97872
Resolution range (Å)	50.00-1.33 (1.35-1.33) ^a	100.0-1.38 (1.40-1.38) ^a
Protein monomers in asymmetric units	1	1
Space group	<i>C</i> 2	<i>C</i> 2
Unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	108.10, 81.59, 53.53	107.92, 82.14, 53.56
α , β , γ (°)	90, 104.2, 90	90, 104.22, 90
Total number of reflections	738,039	724,368
Number of unique reflections	103,035	92,788
Multiplicity	4.0 (3.2) ^a	4.0 (3.4) ^a
Completeness (%)	97.1 (94.0) ^a	99.3 (98.2) ^a
Mean <i>I</i> / σ <i>I</i>	14.38 (1.79) ^a	12.94 (2.00) ^a
<i>R</i> _{merge} (%) ^b	8.3 (57.9) ^a	8.6 (51.5) ^a
Refinement		
Resolution range (Å)	32.12-1.33	32.21-1.38
Number of reflections in working set	87,838	81,098
Number of reflections in test set	4,370	4,042
<i>R</i> _{work} (%) ^c	17.8	18.3
<i>R</i> _{free} (%) ^c	20.6	20.8
Number of non-hydrogen atoms	3,120	3,057
Protein / water	2542/510	2551/457
RMSD - bond lengths (Å)	0.008	0.016
RMSD - bond angles (°)	1.23	1.22
Ramachandran favored (%)	99	99
Ramachandran outliers (%)	0	0
Molprobit clash score	1.3	2.1
Average B-factor (Å ²)	20	21
Protein	18	19
Ligands	22	33
Solvent	30	32

^a Values in parentheses are for highest-resolution shell.
^b $R_{\text{merge}} = \frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_i I_i(h)}$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$.
^c R_{work} and $R_{\text{free}} = \frac{h(|F(h)_o| - |F(h)_c|)}{h|F(h)_o|}$ for reflections in the working and test sets, respectively.

Table 2.12.
X-ray data collection and refinement statistics for SARS-CoV 3CL^{pro} in
complex with FMW and KWW

	Beamline: LS-CAT Sector 21 ID-F	
	FMW	KWW
Data collection		
Wavelength (Å)	0.97872	0.97872
Resolution range (Å)	50.00-1.33 (1.35-1.33) ^a	100.00-1.65 (1.68-1.65) ^a
Protein monomers in asymmetric units	1	1
Space group	<i>C</i> 2	<i>C</i> 2
Unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	108.05, 82.09, 53.56	108.19, 82.35, 53.77
α , β , γ (°)	90, 104.16, 90	90, 104.52, 90
Total number of reflections	655,855	958,009
Number of unique reflections	103,903	53,109
Multiplicity	2.1 (2.0) ^a	3.5 (3.6) ^a
Completeness (%)	97.3 (91.1) ^a	98.1 (97.9) ^a
Mean $I/\sigma I$	23.24 (1.46) ^a	15.00 (2.59) ^a
R_{merge} (%) ^b	6.2 (42.4) ^a	8.3 (40.5) ^a
Refinement		
Resolution range (Å)	25.93-1.33	32.37-1.66
Number of reflections in working set	103,487	46,878
Number of reflections in test set	5,126	2,372
R_{work} (%) ^c	18.1	17.7
R_{free} (%) ^c	19.3	20.5
Number of non-hydrogen atoms	3,137	3,239
Protein / water	2,595/492	2,581/610
RMSD - bond lengths (Å)	0.009	0.010
RMSD - bond angles (°)	1.19	1.17
Ramachandran favored (%)	99	98
Ramachandran outliers (%)	0	0
Molprobit clash score	2.2	3.2
Average B-factor (Å ²)	22	18
Protein	20	14
Ligands	36	24
Solvent	32	31
^a Values in parentheses are for highest-resolution shell. ^b $R_{\text{merge}} = \frac{\sum_h \sum_i I_i(h) - \langle I(h) \rangle }{\sum_h \sum_i I_i(h)}$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$. ^c R_{work} and $R_{\text{free}} = \frac{h(F(h)_o - F(h)_c)}{h F(h)_o }$ for reflections in the working and test sets, respectively.		

Table 2.13.
X-ray data collection and refinement statistics for SARS-CoV 3CL^{pro} in
complex with XM2

	Beamline: LS-CAT Sector 21 ID-F XM2
Data collection	
Wavelength (Å)	0.97872
Resolution range (Å)	50.00-1.33 (1.35-1.33) ^a
Protein monomers in asymmetric units	1
Space group	<i>C</i> 2
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	108.11, 81.71, 53.59
α , β , γ (°)	90, 104.15, 90
Total number of reflections	701,632
Number of unique reflections	103,708
Multiplicity	4.0 (3.2) ^a
Completeness (%)	99.6 (96.7) ^a
Mean $I/\sigma I$	17.02 (1.83) ^a
R_{merge} (%) ^b	7.6 (53.2) ^a
Refinement	
Resolution range (Å)	32.13-1.33
Number of reflections in working set	90,059
Number of reflections in test set	4,506
R_{work} (%) ^c	17.4
R_{free} (%) ^c	20.0
Number of non-hydrogen atoms	3,157
Protein / water	2,575/506
RMSD - bond lengths (Å)	0.011
RMSD - bond angles (°)	1.17
Ramachandran favored (%)	99
Ramachandran outliers (%)	0
Molprobit clash score	1.5
Average B-factor (Å ²)	20
Protein	18
Ligands	26
Solvent	30
^a Values in parentheses are for highest-resolution shell.	
^b $R_{\text{merge}} = \Sigma_h \Sigma_i I_i(h) - \langle I(h) \rangle / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$.	
^c R_{work} and $R_{\text{free}} = h(F(h)_o - F(h)_c) / h F(h)_o $ for reflections in the working and test sets, respectively.	

analysis and crystallization attempts. Successful crystallization conditions for SARS-CoV 3CL^{pro} and MERS-CoV 3CL^{pro} in complex with different inhibitors have also been discussed. These crystals diffracted to high resolution and were instrumental in getting molecular insights into inhibitor-3CL^{pro} interactions.

CHAPTER 3. SARS-COV 3CL^{PRO} INHIBITOR DEVELOPMENT

Parts of the data and text in this chapter have been published in various journal articles [38, 39, 60].

Introduction

3.1 Different classes of SARS-CoV 3CL^{PRO} inhibitors

As discussed in Chapter 1, coronavirus encoded proteases PL^{PRO} and 3CL^{PRO} are indispensable for virus replication and thus serve as excellent targets for therapeutic development. Several groups, including ours, have published crystal structures of SARS-CoV 3CL^{PRO} in complex with inhibitors providing critical insights into the binding interactions between the inhibitor molecules and the active site. These inhibitors can be classified based on their chemical scaffolds, and discussed below.

3.1.1 Natural product inhibitors

Chen et al. screened a library of 720 natural products and drugs to identify compounds that can inhibit SARS-CoV 3CL^{PRO} in an HPLC-based assay [61]. Initial hits identified from this library (Figure 3.1, compound **1**) are structurally similar to pure ingredients from natural teas. Next, the authors utilized a fluorometric assay to test the inhibitory activity of known ingredients isolated from tea extracts. They identified several tea polyphenols that could inhibit SARS-CoV 3CL^{PRO} with low μM IC₅₀ values. The most potent tea polyphenol in their assays, compound **2**, inhibited SARS-CoV 3CL^{PRO} with an IC₅₀ of 9.5 μM . It is to be noted that flavonoids and polyphenolic compounds have been reported in the literature to form aggregates that

promiscuously inhibits a variety of enzymes and are considered pan-assay interference compounds (PAINS) [62–64].

Wen et al. tested a small library of 221 compounds to identify phytochemical compounds that can inhibit SARS-CoV dependent cytopathic effects and viral replication in Vero E6 cell-based assays [65]. They identified 22 compounds that have anti-SARS biological activity. Of these compounds, the majority were classified as terpenoids and lignoids isolated from medicinal plants. Further characterization of these compounds in a fluorometric assay using purified SARS-CoV 3CL^{pro} demonstrated that these compounds are competitive inhibitors of SARS-CoV 3CL^{pro}. Compounds **3** (terpenoid) and **4** (lignoid) with the highest inhibitory activity against SARS-CoV 3CL^{pro} are shown in Figure 3.1.

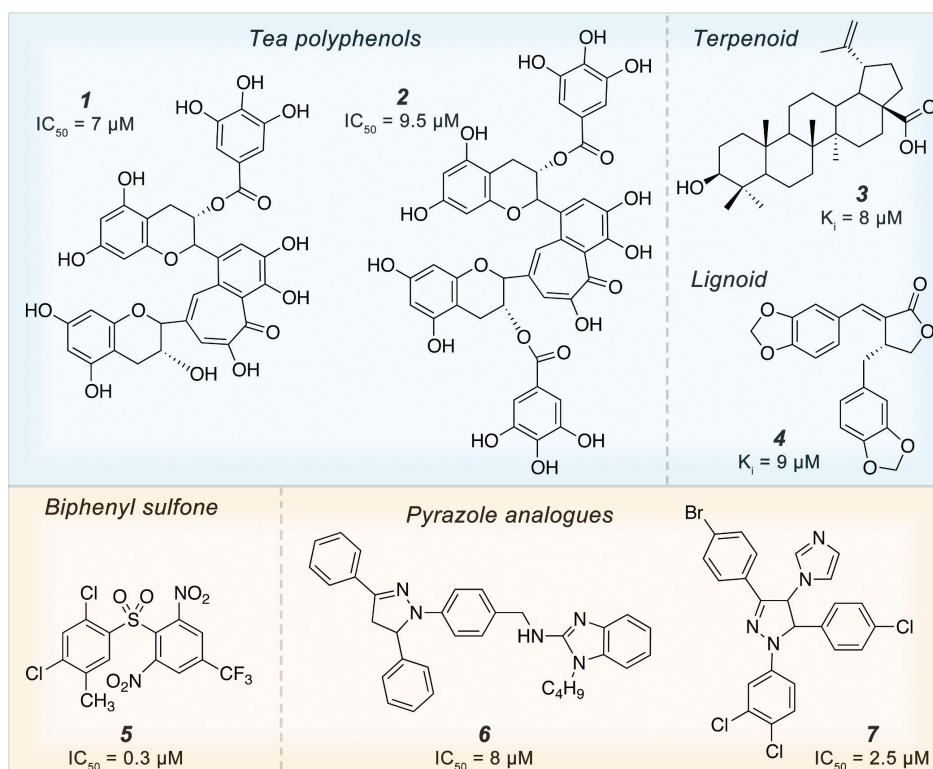


Fig. 3.1.: Inhibitory activity of natural product and non-peptidic inhibitors of SARS-CoV 3CL^{pro}. Compounds **1-4** belong to different classes of natural product inhibitors of 3CL^{pro} enzyme. Biphenyl sulfones and pyrazole analogues are included in the non-peptidic class of inhibitors. Inhibitory activity is represented as either IC₅₀ value or K_i value.

3.1.2 Non-peptidic inhibitors

Biphenyl sulfones

Lu et al. performed a structure-based virtual screening on 58,855 small-molecule compounds from the Maybridge library [50]. Based on the template built from the virtual-screen hits, followed by inhibition assays, 21 compounds were identified that inhibited SARS-CoV 3CL^{pro} with IC₅₀ values < 50 μ M. X-ray crystal structure with the most potent inhibitor, compound **5** in Figure 3.1, revealed a unique binding orientation of the inhibitor molecule in the *S*₃-*S*₅ pockets of the enzyme. Compound **5** interacts with the enzyme through an extensive network of hydrogen-bonding and hydrophobic interactions. The inhibitor molecule also induces conformational changes in the active site that moves His41 of the catalytic dyad away from Cys145, essentially disrupting the catalytic machinery.

Pyrazole analogues

High-throughput screening of a 6800 compound library led to the identification of one compound (compound **6**, Figure 3.1) that not only inhibited 3CL^{pro} from SARS-CoV, but also inhibited 229E-CoV^{pro}, as well as 3C^{pro} protease from picornaviruses RV14 and EV71. Kuo et al. further evaluated the inhibitory activity of several analogues of compound **6** from another library [66]. The most potent of these analogues, compound **7** showed a broad-spectrum specificity for other viral proteases (Figure 3.1).

3.1.3 Metal-conjugated inhibitors

Several metal ions, including Hg²⁺ and Zn²⁺, can react with or strongly interact (coordinate) with the active site Cys of Cys proteases and inhibit their activities. Phenylmercuric acetate (PMA) is used as a bactericidal and antimicrobial preservative in several drug preparations. Hsu et al. screened a library of 960 bioactives and

commercially available drugs to identify inhibitors of SARS-CoV 3CL^{pro} [67]. They identified several Hg and Zn containing compounds that act as potent inhibitors of SARS-CoV 3CL^{pro} with nanomolar K_i values. Further analysis of the X-ray crystal structures of SARS-CoV 3CL^{pro} in the inhibitor bound forms by Lee et al revealed that Hg containing compound PMA (Figure 3.2, compound **8**) binds in the S_3 subsite and coordinates with Cys44, Met49 and Tyr54 (Figure 3.2, compound **8**); however Zn based compounds bind in the active site and coordinate with the His41-Cys145 catalytic dyad (Figure 3.2, compound **9**) [68].

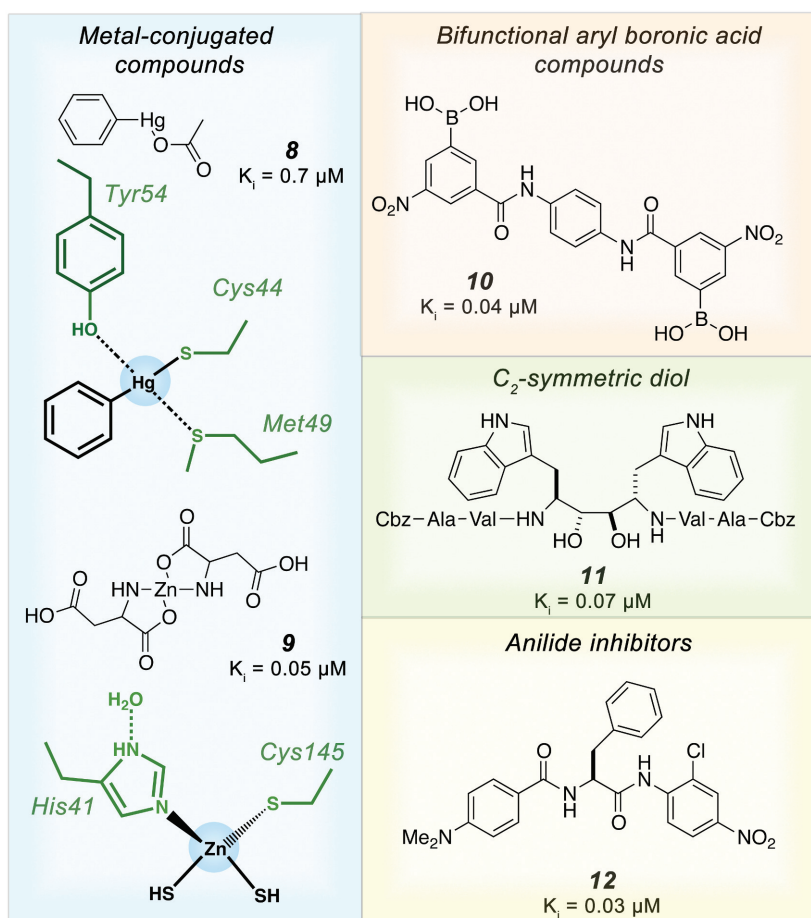


Fig. 3.2.: Inhibitory activity of metal-conjugated, boronic-acid based, C_2 -symmetric diol and anilide inhibitors of SARS-CoV 3CL^{pro}. 3CL^{pro} amino acids that coordinate with the metal in the metal-conjugated inhibitors are shown. Inhibitory activity is represented as either IC_{50} value or K_i value.

3.1.4 Bifunctional boronic acid inhibitors

Bacha et al. predicted that the conserved serine cluster (Ser139, Ser144, Ser147) of SARS-CoV 3CL^{pro} can be targeted for the development of potent inhibitors. Given the reactivity of boronic acid with hydroxyl groups, they tested inhibitory activity of boronic acid based compounds against SARS-CoV 3CL^{pro} [69]. They found one compound that inhibited the enzyme with a K_i value of 4.5 μ M. Further testing of different analogues of this compound led to the identification of compound **10** (Figure 3.2) that inhibits SARS-CoV 3CL^{pro} with nanomolar potency.

3.1.5 C_2 -symmetric diol inhibitors

Shao et al. initially observed that TL-3, a C_2 -symmetric diol non-covalent inhibitor of HIV-protease, also inhibits SARS-CoV 3CL^{pro} [70]. Aided by computational modeling, they developed a series of lead compounds. The most potent compound, compound **11** (Figure 3.2), achieved nanomolar potency and selectivity for the inhibition of SARS-CoV 3CL^{pro}.

3.1.6 Anilide inhibitors

Based on previous reports of enhancement in the inhibitory potency of AG7088 analogues by placing L-phenylalanine at the P_1 position, Shie et al. designed a series of analogues [71]. These analogues were derivatives of chloro *para*-aniline that contained L-phenylalanine at the P_1 position and varying substitutions at other P_x positions. The most potent analogue, compound **12** (Figure 3.2), inhibited SARS-CoV 3CL^{pro} with an IC_{50} value of 60 nM and K_i value of 30 nM. Compound **12** was stable and did not undergo hydrolysis in the presence of enzyme upon several hours of incubation.

3.1.7 Inhibitors with ‘warhead’ functionalities

Traditional approaches for the development of inhibitors for cysteine-active-center proteases tend to favor irreversible inhibitors, which consist of a substrate-mimetic sequence attached to a reactive warhead group such as an ester, aldehyde, ketone or Michael-acceptor. The warhead group attacks the active site cysteine and forms a covalent complex.

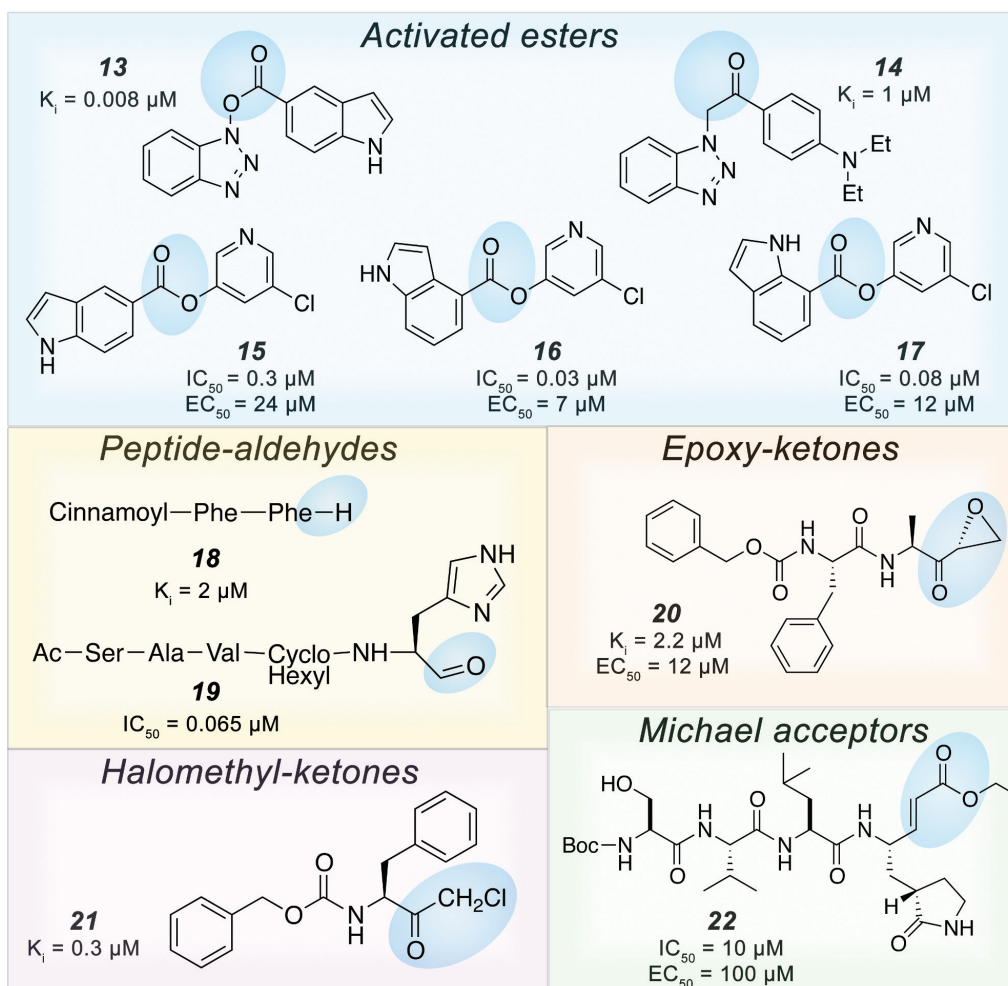


Fig. 3.3.: Inhibitory activity of compounds with reactive ‘warhead’ functionalities. Warhead groups are highlighted in blue circles. Inhibitory activity is represented as either IC_{50} value or K_i value. EC_{50} value represents the inhibitor concentration that blocks virus replication by 50%.

Reactive ester inhibitors

Benzotriazole esters- Wu et al designed a series of reactive benzotriazole based esters with nanomolar potency, compound **13** Figure 3.3 [72]. These compounds displayed a time and concentration dependent decrease in the enzymatic activity indicating modification of the active site cysteine. Further optimization of these compounds to develop more stable non-covalent inhibitors of SARS-CoV 3CL^{pro} yielded compound **14** (Figure 3.3) that inhibited SARS-CoV 3CL^{pro} with a K_i value of 1 μ M. In compound **14**, the ester oxygen was replaced with a carbon.

Chloropyridyl esters- Our lab, in collaboration with the Ghosh lab at Purdue University, developed a series of chloropyridyl esters that not only inhibited SARS-CoV 3CL^{pro} with nanomolar potency, but also blocked virus replication with low micromolar EC₅₀ values (compounds **15-17**, Figure 3.3) [37].

Peptide aldehyde inhibitors

Several groups have reported the development of highly potent aldehyde based inhibitors [73, 74]. The terminal-carboxyl of the P_1 residue is substituted with an aldehyde group in these compounds. These compounds are mechanism-based inhibitors that covalently modify the active site cysteine. Surprisingly though, kinetic characterization and X-ray crystal structure determination of compound **19** (Figure 3.3) in complex with SARS-CoV 3CL^{pro} indicated that compound **19** acts as a competitive inhibitor of the enzyme without modifying the active site cysteine [74].

Epoxy-ketone inhibitors

Based on P_1 - P_4 substrate specificity of 3CL^{pro}, Goetz et al. developed a library of 2000 small-molecules, containing reactive warheads, for high-throughput screening [75]. Optimization of their initial epoxide ketone scaffold resulted in the development of α,β -epoxy ketone compound **20** (Figure 3.3) that inhibited SARS-CoV 3CL^{pro}

with low micromolar potency and also effectively blocked virus replication. Further analysis of inhibitor-bound crystal structure revealed that the epoxide ring opens upon nucleophilic attack by the active site cysteine and the linearized oxirane occupies the S_1 subsite in the enzyme active site.

Halomethyl ketone inhibitors

Halomethyl ketones, such as compound **21** (Figure 3.3), have been identified as potent inhibitors of SARS-CoV 3CL^{pro}. Inhibition kinetics and structural studies reveal that this class of compounds inhibit the enzyme via S-alkylation of active site cysteine. The halomethyl carbon of the inhibitor undergoes nucleophilic attack by Cys145, which results in the formation of a thioether bond [76].

Michael acceptors

Our lab has reported the development of Michael-acceptor based inhibitors of SARS-CoV 3CL^{pro}. These α,β -unsaturated esters, such as compound **22** (Figure 3.3), inactivate the enzyme in a time and concentration dependent manner suggesting inactivation via covalent modification of active site cysteine [54]. Structural studies further reveal the formation of a 1.8 Å bond between the gamma-sulfur of the active site Cys and the electrophilic β -carbon of the Michael acceptor. The P_1 -lactam mimics the P_1 -glutamine and occupies the S_1 subsite in the enzyme. These peptidomimetic inhibitors interact with the enzyme through a series a hydrogen-bonding interactions. Some of the inhibitors in this class also display antiviral activity, albeit at high concentration [54].

Despite the straightforward approach to design potent irreversible inhibitors of cysteine proteases, they tend to display off-target side-effects and toxicity in cellular environments by modifying reactive cysteines of cellular proteins [77, 78]. Moreover, none of the 39 marketed protease inhibitors that act via covalent modification of active site amino acid target cysteine proteases, underlining the challenge of utilizing such

inhibitors as drugs [38, 77, 78]. Therefore, it is important to design potent reversible inhibitors that can specifically target 3CL^{pro} without having off-target side effects.

3.2 Development of non-covalent reversible inhibitors of SARS-CoV 3CL^{pro}

The Mesecar lab, in collaboration with Craig Lindsley's and Shaun Stauffer's lab (Vanderbilt University), synthesized a focused library of 237 compounds to optimize two unique chemical scaffolds (furan-yl-amide and benzotriazole; scaffolds 2 and 4, respectively, in Figure 3.4) to develop small molecule, reversible inhibitors of SARS-CoV 3CL^{pro}. These scaffolds were identified based on initial hits from a high-throughput screen of ~300,000 compounds conducted in 2009 as a part of NIH molecular libraries initiative in collaboration with the Scripps Research Institute Molecular Screening Center (SRIMSC), Figure 3.4 [38, 39]. Scaffolds 2 and 4 were preferred over scaffolds 3 and 4 for hit to lead development for several reasons. Firstly, scaffolds 2 and 4 displayed slightly higher inhibition of SARS-CoV 3CL^{pro} compared to other scaffolds. Additionally, scaffolds 2 and 4 exhibited higher cluster representation among the 101 confirmed HTS hits. Finally, these scaffolds were more amenable to chemical modifications allowing rapid expansion of the chemical library.

Different substitutions were made at R₁, R₂ and R₃ positions to synthesize compounds in this library. Several compounds, with both furan-yl-amide and benzotriazole scaffolds, that inhibit SARS-CoV 3CL^{pro} with low micromolar or sub-micromolar potency were identified partly through work done by previous Mesecar lab members Valerie Grum-Tokars and partly through work described in this dissertation. Determination of X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with multiple inhibitors from both the classes provide insights into enzyme-inhibitor interactions. In this chapter, structure activity relationships (SARs) of compounds in our in-house 237 compound library and identification of specific functional groups that engage in hydrogen-bonding interactions with the enzyme, and are therefore, essential for inhibitory activity of the compounds are discussed.

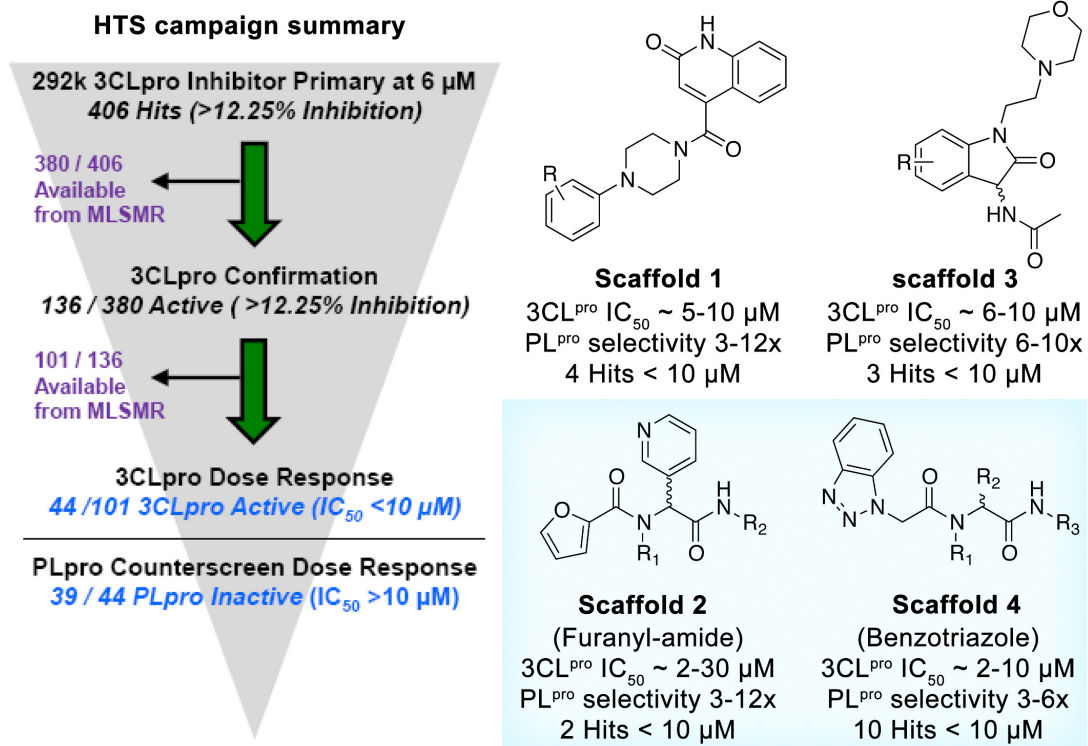


Fig. 3.4.: Chemical scaffolds identified through high-throughput screening (HTS) for SARS-CoV 3CL^{pro} inhibitor development. Left panel represents a summary of the HTS campaign. Of the original 406 hits identified from HTS, 380 compounds that were available from Molecular Libraries Small Molecule Repository (MLSMR) were re-evaluated to confirm their inhibitory activity against SARS-CoV 3CL^{pro}. 136 compounds inhibited SARS-CoV 3CL^{pro} with >12.5 % inhibition at 10 μ M compound concentration. Out of 101 active hits from the confirmation screen, 39 compounds inhibited SARS-CoV 3CL^{pro} with $IC_{50} < 10 \mu$ M and displayed no significant inhibitory activity against PL^{pro}. The initial 101 compounds can be classified into four different scaffolds (right panel) based upon chemical structural clustering. Scaffolds 2 and 4 were utilized for further hit to lead development. Figure has been adapted from [79].

Although these compounds were initially developed to inhibit SARS-CoV 3CL^{pro}, this study reports the evaluation of the inhibitory activity of these compounds against a panel of 3CL^{pro} enzymes from other human CoVs (SARS, OC43, HKu1, HKU5 and MERS) as well as MHV. The latter was chosen since it is a convenient model system for studying CoV replication in cell culture. Several compounds that can inhibit multiple 3CL^{pro} enzymes and therefore, have great potential to be developed as broad-

spectrum inhibitors of 3CL^{pro} enzymes from existing and emerging coronaviruses were identified.

3.3 Results

3.3.1 Furanyl amide scaffold - SAR

Early during the lead development, the Mesecar lab determined the crystal structure of SARS-CoV 3CL^{pro} in complex with a furanyl amide scaffold analogue [38]. The structure revealed the binding orientation of the inhibitor in the enzyme active site. For convenience, structure activity relationships (SAR) discussed below are shown based on SARS-CoV 3CL^{pro} subsites (S_1 '- S_3) that are expected to accommodate varying functional groups at different positions in the furanyl amide scaffold (See top panel of Figure 3.5).

Varying functional groups in S_1 subsite

●Stereo-preference

Based on the inhibitory activities of compounds VNM (racemic mixture), REY (*R*-stereoisomer) and RF9 (*S*-stereoisomer), it is evident that the active site residues impose stereo-selectivity on the functional group that binds in the S_1 subsite of SARS-CoV 3CL^{pro} (Figure 3.5, green box). The *R*-stereoisomer is active (REY), while *S*-stereoisomer (RF9) showed a complete loss of inhibitory activity.

●Position of pyridine nitrogen

Changing the position of the pyridine nitrogen from *meta* to *ortho* or *para* results in a total loss of inhibitory activity (Figure 3.5, blue box). Pyrimidine (QEF), pyrazine (886) and pyridazine (884) with at least one nitrogen at the *meta* position

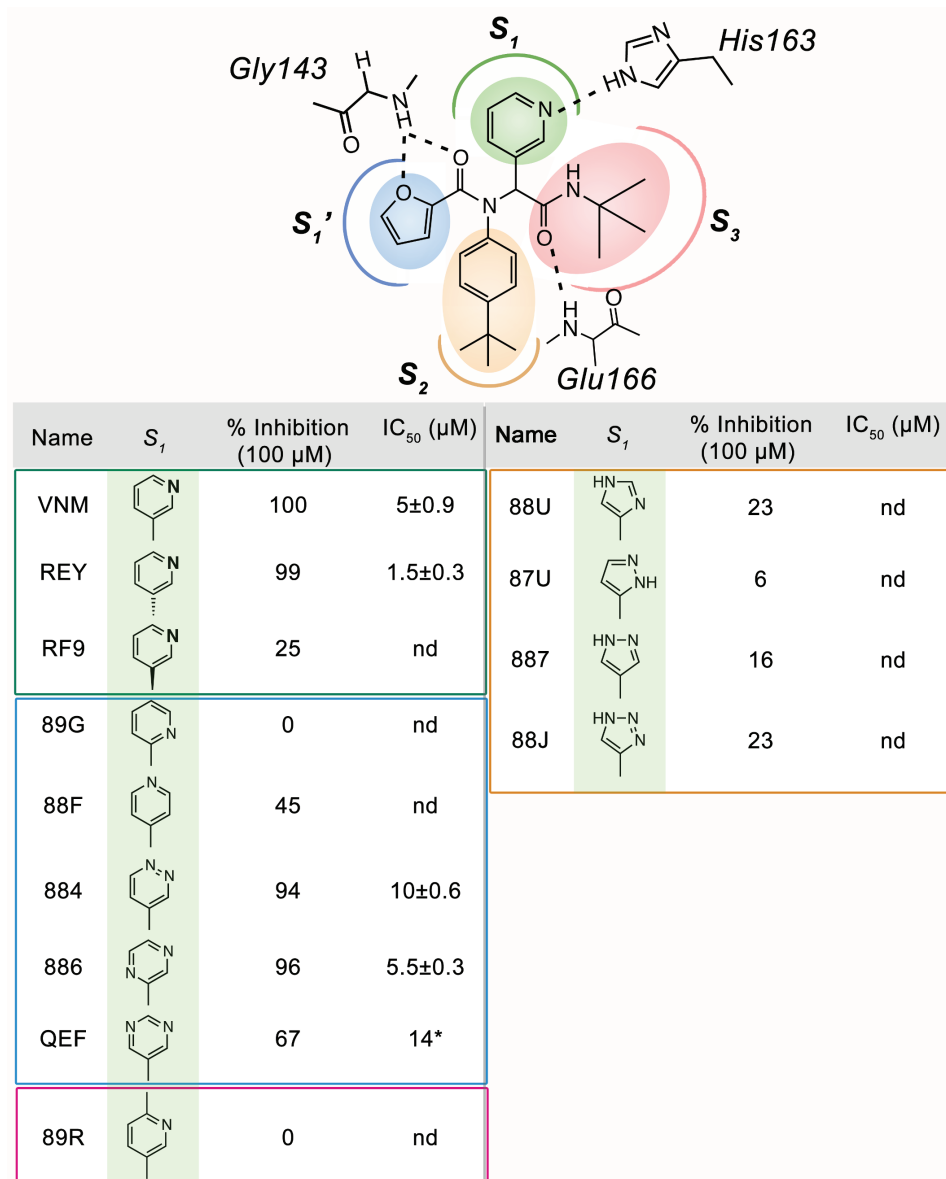


Fig. 3.5.: **Furanyl amide scaffold - SAR of varying functional groups in S_1 subsite of SARS-CoV 3CL^{Pro}.** Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{Pro} is illustrated in the top panel. Subsites S_1 '- S_3 with the amino acids that form hydrogen-bonding interactions (dashed lines) with the inhibitors are also shown. *green box* - stereoselectivity in the S_1 subsite; *blue box* - effect of varying positions of the pyridine nitrogen; *pink box* - effect of addition of methyl group at the *para*-position of pyridine ring; *orange box* - substitution of pyridine ring with five-membered heteroaromatic rings. *Standard error for IC_{50} value is not available.

are also active. The observation is consistent with the formation of hydrogen-bond between the pyridine *meta*-nitrogen and the imidazole ring -NH of conserved His163.

●Para-methyl pyridinyl group

Addition of a methyl group to the pyridine ring (89R) results in a complete loss of activity possibly due to steric clashes, suggesting a compact S_1 sub-site (Figure 3.5, pink box). In fact, based on the X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with the compound REY (PDB ID: 3V3M) [38], distance between the carbon at the position 4 of the pyridine ring and the methylene carbon of Phe140 side chain in the S_1 pocket is only 3.6 Å, which would not only binding of compound 89R due to steric hindrance.

●Six-membered versus five-membered heteroaromatic rings

All five-membered heteroaromatic ring systems (imidazole, pyrazole and triazole) failed to show any inhibitory activity against SARS-CoV 3CL^{pro} suggesting that six-membered aromatic heterocycles are preferred over five-membered heterocycles (Figure 3.5, orange box). Reduction in the inhibitory activity for compounds 87U, 887 and 88J can be explained by a reduction in their hydrogen-bond accepting capability due to a decrease in the pK_a values for the diazole (87U and 887, pK_a = 2.5) and triazole (88J) functional groups compared to the pK_a of pyridine group (VNM, pK_a = 5.2). Reduction in the inhibitory potency for the compound 88U is possibly explained by an increase in the distance between -N of the imidazole functional group in 88U and the -NH of His163, thereby decreasing the hydrogen-bond strength.

Varying functional groups in S_1' subsite

●Substitution on the furanyl ring

Addition of a single methyl group to the furan ring results in a significant loss of inhibitory potency (RWW, RVC), whereas addition of larger $-CF_3$ or benzyl groups completely abolishes inhibition (Figure 3.6, green box). The loss of inhibition is possibly due to steric clashes with Gly143 as suggested by the crystal structure of SARS-CoV 3CL^{pro}-REY complex, suggesting that the S_1' pocket is compact.

●Six-membered substituted or hetero-aromatic rings

Replacement of furan ring by six-membered hetero-aromatic or substituted aromatic ring results in a complete loss of inhibition activity (Figure 3.6, pink box) either due to the bulkier size of these rings or due to the loss of proper orientation of hydrogen-bond acceptor. Furanyl oxygen forms hydrogen-bond with backbone carbonyl oxygen of conserved Gly143.

●Non-aromatic cyclic or acyclic groups

Substitution of the furan ring with non-aromatic cyclic or acyclic groups (Figure 3.6, orange box) results in a partial loss of activity, suggesting a preference for aromatic heterocycles over non-aromatic groups.

●Five-membered heteroaromatic rings

Replacing the furan ring of REY with pyrrole (QEM) or pyrazole ring (RWQ and RW3) results in a significant loss of activity possibly due to the loss of hydrogen-bond with Gly143 (Figure 3.6, blue box). Oxazole (QE7, RWB) is tolerable in the S_1' subsite of SARS-CoV 3CL^{pro}, albeit with a higher IC₅₀ value. Finally, XM2 with an imidazole group is the second best inhibitor of SARS-CoV 3CL^{pro}, after REY, with

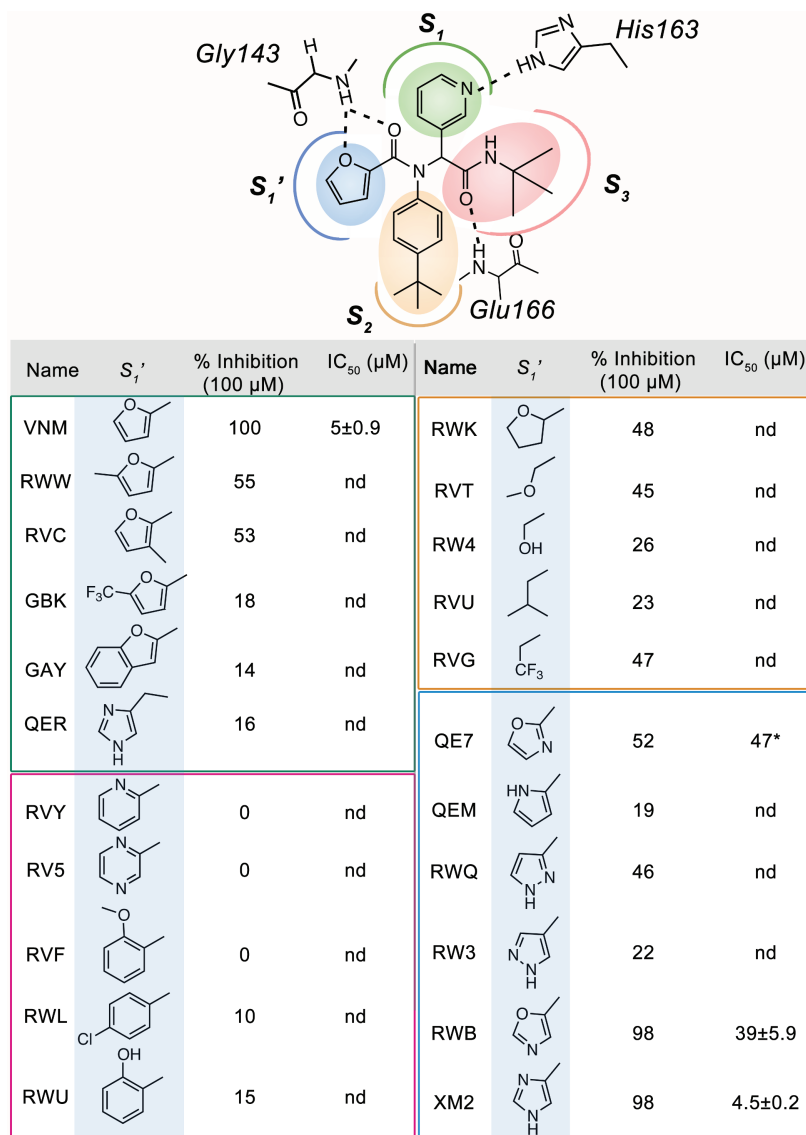


Fig. 3.6.: **Furanyl amide scaffold - SAR of varying functional groups in $S_{1'}$ subsite of SARS-CoV 3CL^{pro}.** Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites $S_{1'}$ - S_3 with the amino acids that form hydrogen-bonding interactions (dashed lines) with the inhibitors are also shown. *green box* - effect of substitutions in the furanyl ring; *pink box* - effect of replacing the furan ring with six-membered heteroaromatic or substituted ring systems; *orange box* - effect of replacing the furan ring with non-aromatic cyclic or acyclic groups; *blue box* - effect of replacing the furan ring with varying five-membered heteroaromatic rings. *Standard error for IC₅₀ value is not available.

an IC_{50} value of $4.5 \mu M$. The imidazole ring of XM2 contains both hydrogen-bonding acceptor and donor groups, and might be involved in the formation of additional hydrogen-bonding interactions with the enzyme. In fact, X-ray crystal structure of SARS-CoV $3CL^{pro}$ in complex with the compound XM2 (Section 3.5.2) reveals that the imidazole -NH of XM2 forms additional hydrogen-bonding interactions with the enzyme through a water-mediated network. It is likely that the oxazole -N of compound RWB engages SARS-CoV $3CL^{pro}$ in a similar manner. When compared to RWB, the reduction in the inhibition activity of compound QE7 can be explained by the loss of hydrogen-bonding interactions as a result of the change in oxazole -N position in QE7.

Varying functional groups in the S_3 subsite

Compounds CTR and L3V (Figure 3.7) displayed a significant loss in the inhibitory activity against SARS-CoV $3CL^{pro}$. The loss in activity can possibly be explained by deletion of the carbonyl oxygen of the inhibitor that forms an important hydrogen-bond with the backbone amide group of Glu166 as shown with other inhibitors in SARS $3CL^{pro}$ (Figure 3.7).

Compound T8J (Figure 3.7) contains an ester group instead of an amide group (VNM). As amides are more polarized than esters, the strength of hydrogen-bond with Glu166 will be weaker in T8J. Additionally, replacement of tert-butyl group with a methyl group (L3V) will also decrease the van der Waals interactions with the enzyme. Additional compounds would need to be synthesized in order to test the amide hydrogen-bond strength and van der Waals interaction hypotheses.

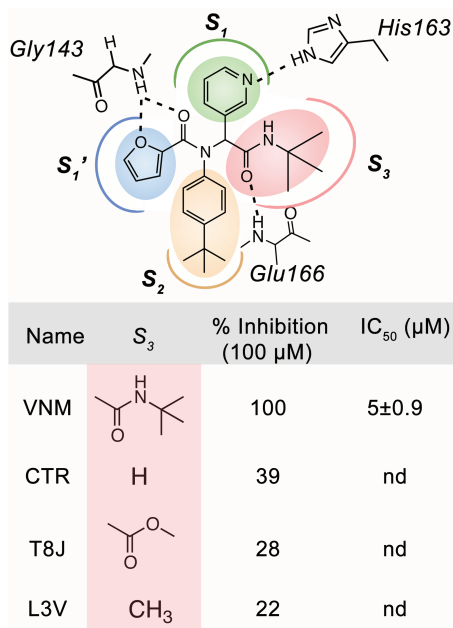


Fig. 3.7.: **Furanyl amide scaffold - SAR of varying functional groups in S_3 subsite of SARS-CoV 3CL^{pro}.** Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites S_1 '- S_3 with the amino acids that form hydrogen-bonding interactions (dashed lines) with the inhibitors are also shown. Formation of a weaker hydrogen-bond (T8J) with Glu166 or complete disruption of the hydrogen-bond (CTR and L3V) possibly influence the inhibitory activity of the given compounds.

Varying functional groups in S_2 subsite

●Non-aromatic ring substitutions

Compounds QRD, QR9 and QRA are completely inactive against SARS-CoV 3CL^{pro} (Figure 3.8, green box), indicating that non-aromatic or significantly bulky ring structures in the S_2 pocket are unfavorable for binding and hence inhibitory activity.

●*para*-substituted halophenyls

VGL with a fluorophenyl group does not have any inhibitory activity against SARS-CoV 3CL^{pro} (Figure 3.8, pink box). Compound GB8 with an iodophenyl sub-

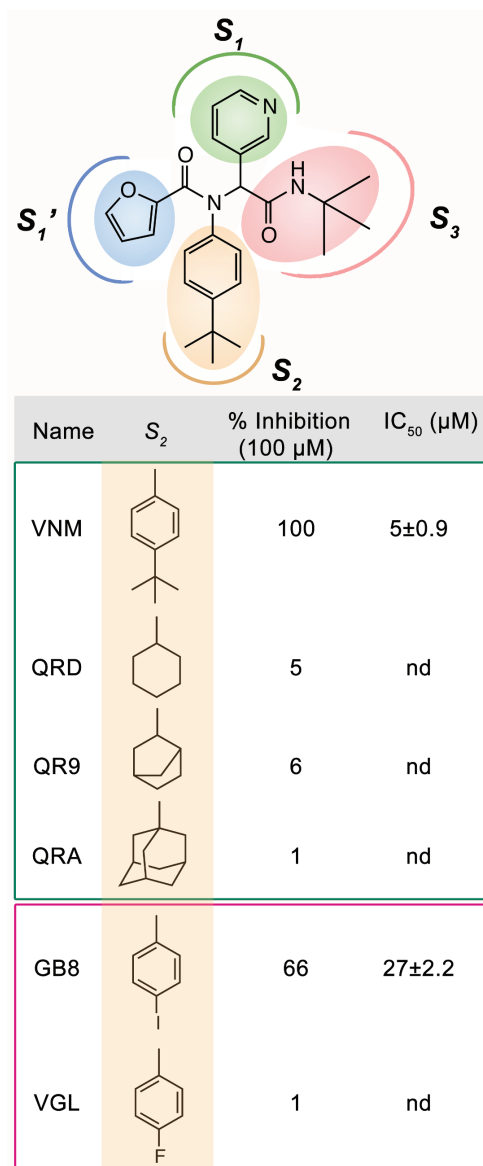


Fig. 3.8.: **Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL^{pro}.** Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites S_1' - S_3 are also shown. *green box* - effect of non-aromatic ring substitutions; *pink box* - effect of *para*-substituted halophenyls.

stitution displayed higher % inhibition compared to VGL when tested at 100 μM inhibitor concentration with an IC_{50} value of 27 μM for SARS-CoV 3CL^{pro} (Figure 3.8, pink box).

●Size of alkyl or aryl groups attached to benzene ring at S_2 position

Attachment of either a *tert*-butyl group (VNM) or a *sec*-butyl (XPC) group to the benzene ring is the most favorable addition for inhibition of SARS-CoV 3CL^{pro} (Figure 3.9, green box). Compared to the *tert*- or *sec*-butyl groups of VNM and XPC, as the size of this group is increased or decreased, the inhibitory activity decreases (VFM and 893 in Figure 3.9, green box). The SARS-CoV 3CL^{pro} S_2 subsite still, however, tolerates a broader size range of chain substitutions attached to the benzene ring. Compound GDF with a phenyl-ethylene group attached to the benzene ring (Figure 3.9, green box) is most likely very bulky to be accommodated in the S_2 subsite and therefore, does not display good inhibitory activity.

●Alkoxy-phenyls

For alkoxy-phenyl substitutions in the S_2 pocket of SARS-CoV 3CL^{pro}, inhibitory activity increases as the size of alkyl chain increases (isopropoxy-phenyl is the most favored) with a preference for compact branched chains (KWW, CUG) over long flexible chains (CUF, GB0) (Figure 3.9, orange box).

●Substituted and heteroaromatic biphenyl rings

The unsubstituted biphenyl compound (GBN) is a potent inhibitor of SARS-CoV 3CL^{pro}. Substitution of cyano or fluoro groups are more preferred at *ortho* positions than *meta* and *para* positions (Figure 3.10). However, FMW, with a pyrimidinyl-phenyl, has better inhibitory activity against SARS-CoV 3CL^{pro} compared to pyridinyl-phenyl (FN9 and 88H), (Figure 3.11). Compound FN8 with fluoro-pyridinyl phenyl displays poor inhibitory activity. All these observations suggest that the pyrimidine nitrogens in FMW might participate in hydrogen-bonding interactions with the enzyme and these interactions are disrupted in FN9, 88H, FN8. Additionally, the number and position of the nitrogen atoms in the phenyl ring differentially affect

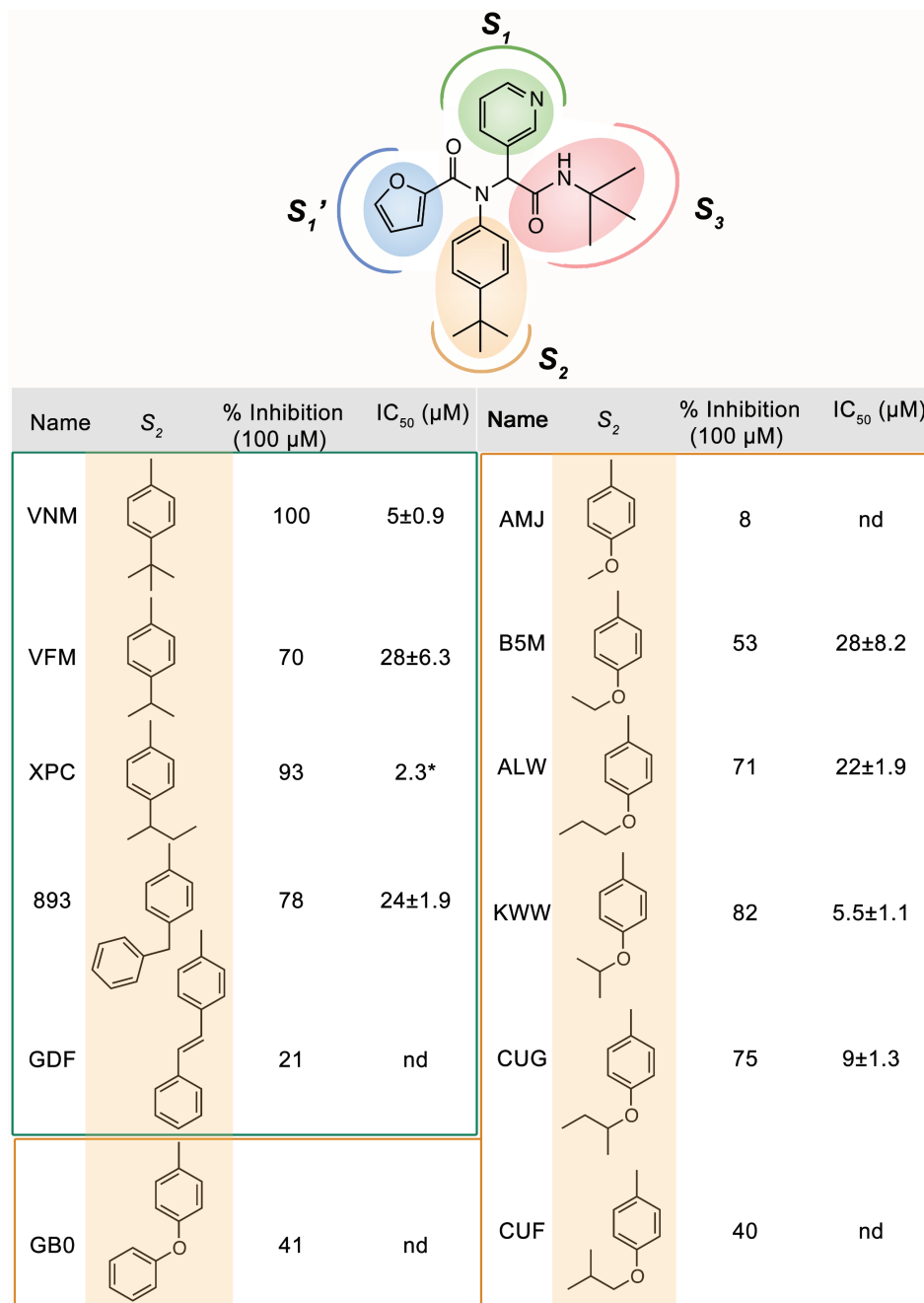


Fig. 3.9.: Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL^{pro} *cont...* Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites S_1 '- S_3 are also shown. *green box* - effect of size of alkyl or aryl groups attached to benzene ring; *orange box* - effect of alkoxy-phenyl groups. *Standard error for IC_{50} value is not available.

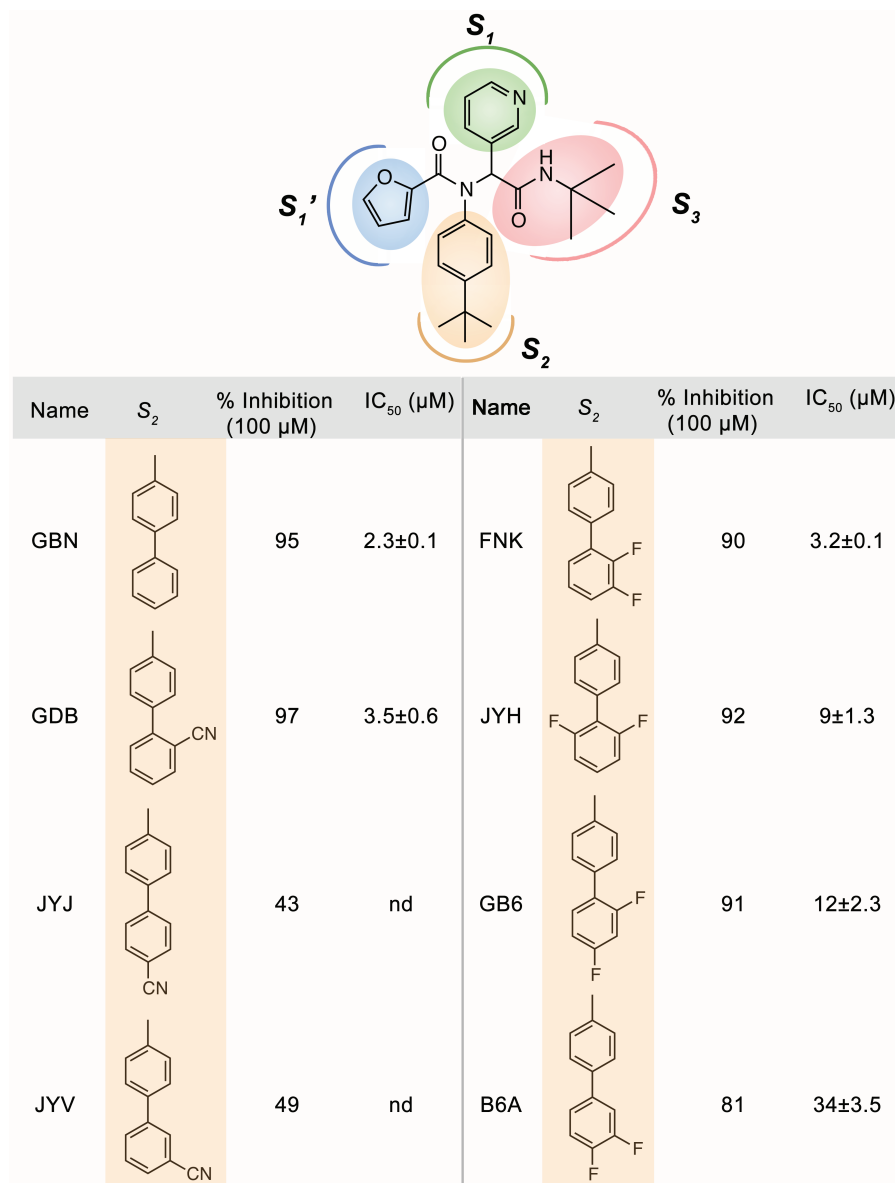


Fig. 3.10.: **Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL^{pro} cont...** Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites S_1 '- S_3 are also shown. Effect of substituted biphenyl groups, in the S_2 pocket, on the inhibitory activity of compounds is represented.

the electron cloud of aromatic ring systems that might influence the interactions of these inhibitors with amino acids in the S_2 subsite of SARS-CoV 3CL^{pro}.

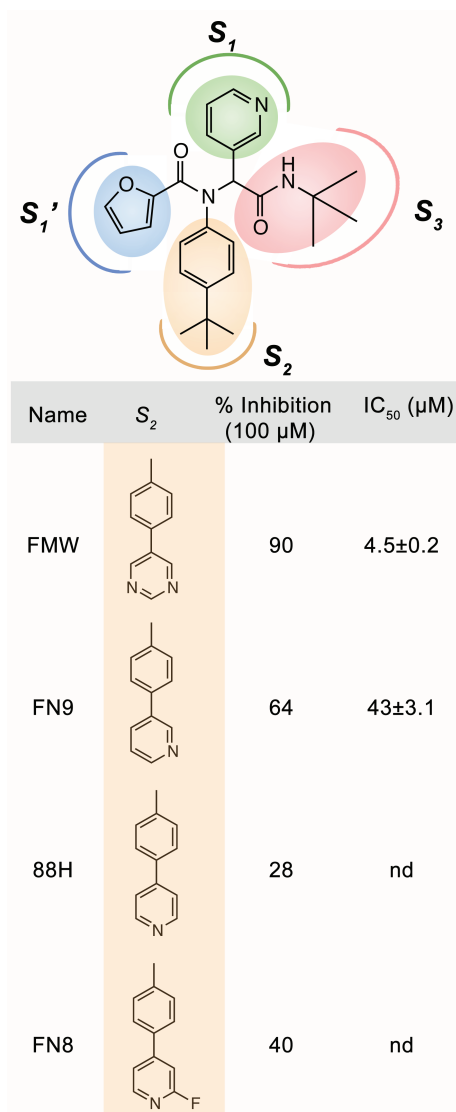


Fig. 3.11.: Furanyl amide scaffold - SAR of varying functional groups in S_2 subsite of SARS-CoV 3CL^{pro} *cont...* Binding mode schematic of reference compound VNM in the active site of SARS-CoV 3CL^{pro} is illustrated in the top panel. Subsites S_1 '- S_3 are also shown. Effect of heteroaromatic biphenyl groups, in the S_2 pocket, on the inhibitory activity of compounds is represented.

3.3.2 Benzotriazole scaffold - SAR

Early during the lead development, our lab determined the X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with a benzotriazole scaffold analogue [39]. The structure revealed an induced-fit binding mechanism and the precise orientation of the

inhibitor in the enzyme active site. For convenience, structure activity relationships (SAR) are shown based on the SARS-CoV 3CL^{pro}s subsites (S_1 '- S_4) that are expected to accommodate varying functional groups at different positions in the benzotriazole scaffold.

Analysis of structure activity relationships (SAR) for the benzotriazole scaffold analogues primarily focused on: 1) replacing the benzotriazole group in the S_1 subsite with other hydrogen-bond accepting functional groups; 2) modifications of N-alkyl phenylamide in the S_1 '- S_2 site; and 3) modifications of S_1 '- S_2 N-alkyl phenylamide in P_3 -truncated benzotriazole analogues, Figure 3.12.

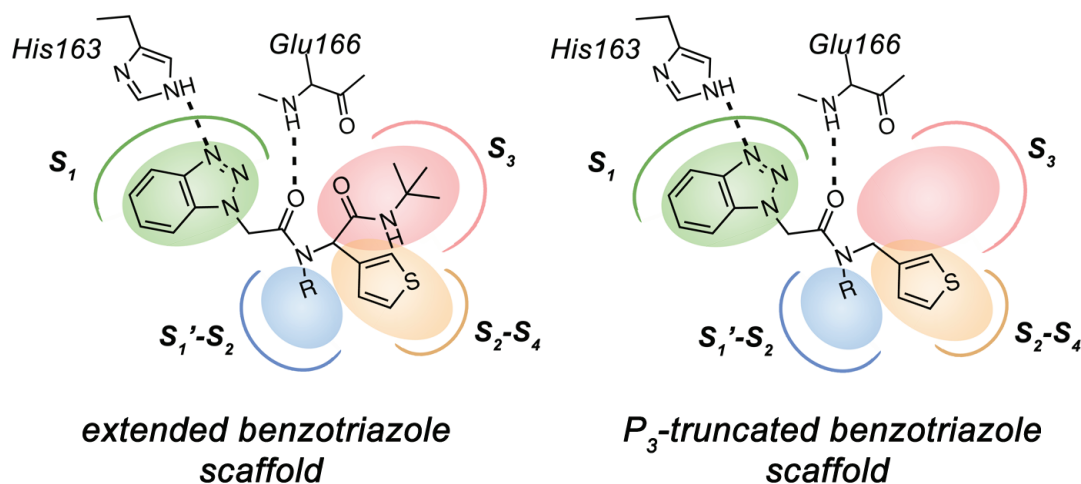


Fig. 3.12.: **Binding mode of benzotriazole scaffold in the active site of SARS-CoV 3CL^{pro}.** *Left-* The binding mode for an extended, benzotriazole-based scaffold is illustrated. *Right-* The binding mode for a P_3 -truncated benzotriazole based scaffold is illustrated. Subsites S_1 '- S_4 with the amino acids that form hydrogen-bonding interactions (dashed lines) with the inhibitors are also shown.

Replacing the benzotriazole group in the S_1 subsite (Inactive analogues)

The X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with a benzotriazole scaffold analogue revealed that the N-3 nitrogen of the benzotriazole group hydrogen-bonds to the His163 -NH in the S_1 subsite. Due to the formation of this important

hydrogen-bond, analogues lacking functional groups with the correct position and electronic configuration for the hydrogen-bond acceptor were inactive, Figure 3.13.

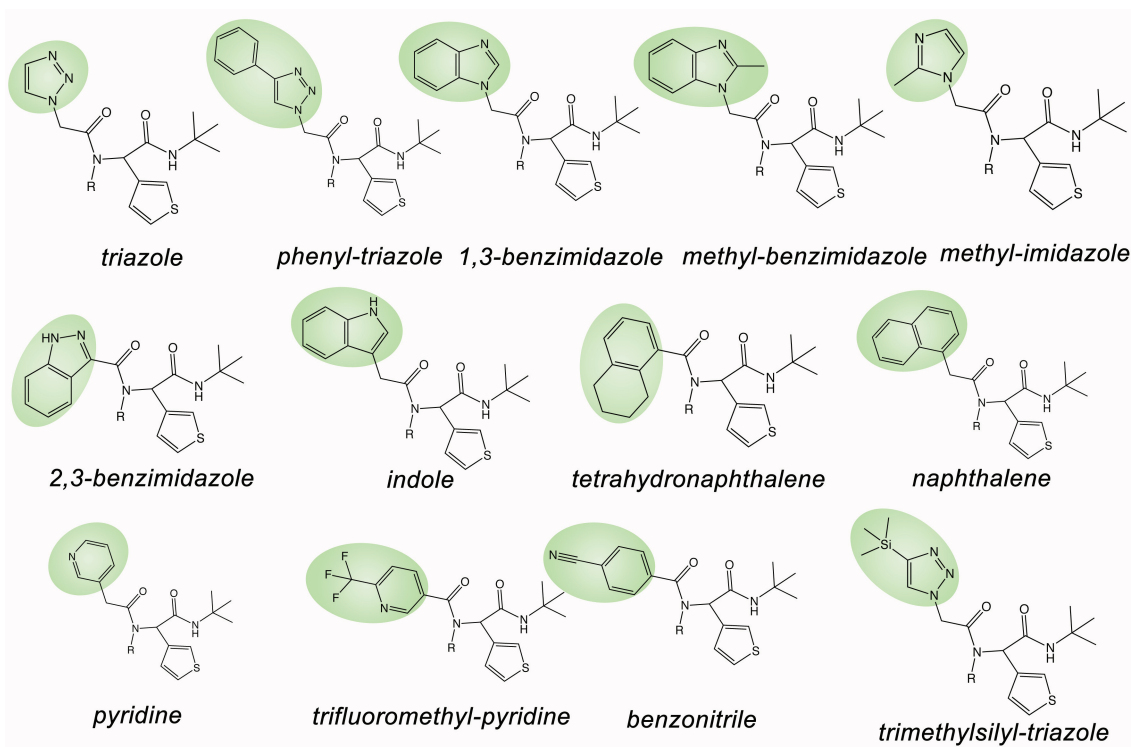


Fig. 3.13.: **Inactive analogues of benzotriazole scaffold.** Chemical structures of benzotriazole scaffold analogues that did not inhibit SARS-CoV 3CL^{pro} are shown. Highlighted in green is the functional group that replaced the benzotriazole group.

Modifications of N-alkyl phenylamide for the S_1' - S_2 subsite in P_3 -extended benzotriazole analogues

●Amide versus sulfonamide

Replacement of the amide group of XPY attached to the benzene ring with sulfonamide (XPK) reduced the inhibitory activity with a three-fold increase in the IC₅₀ value for SARS-CoV 3CL^{pro} (Figure 3.14).

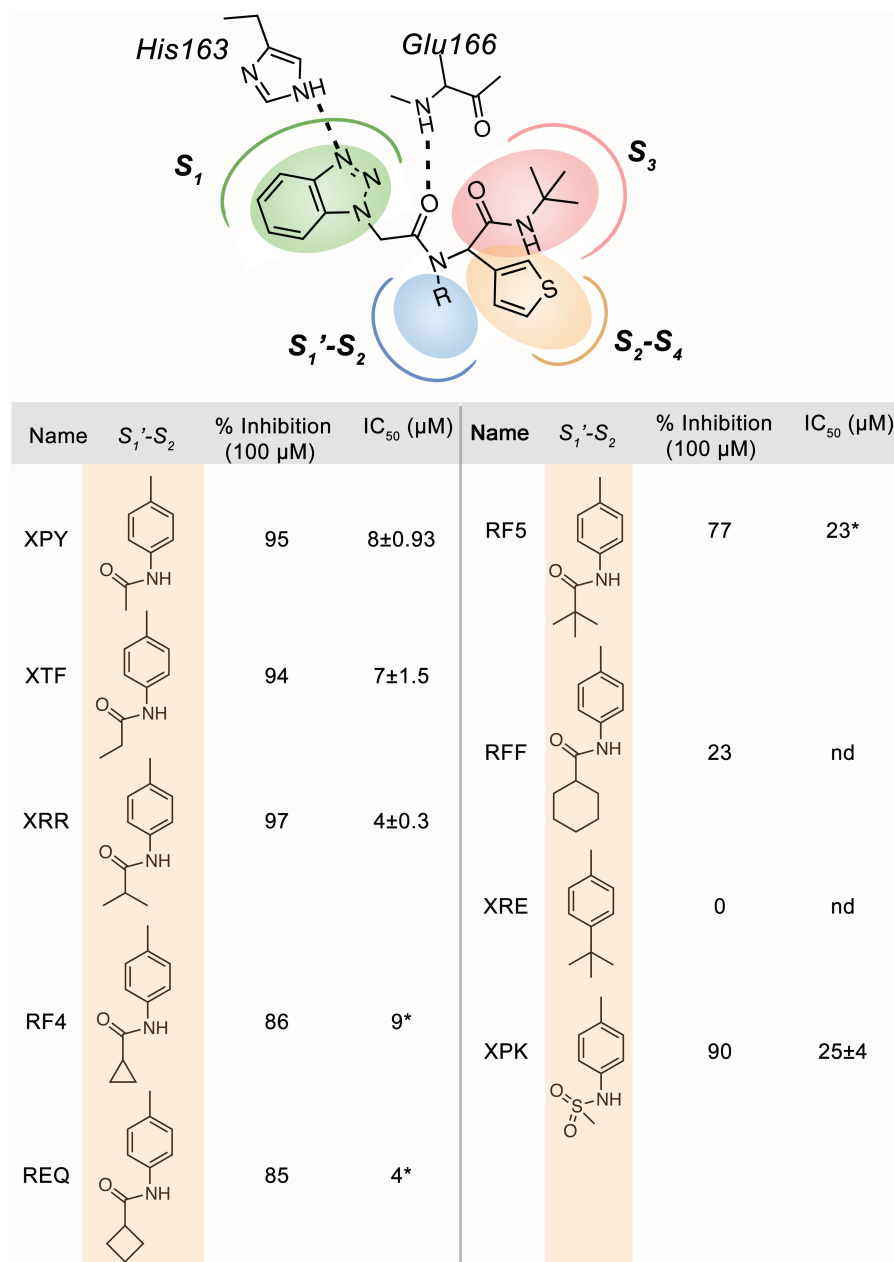


Fig. 3.14.: P_3 -extended benzotriazole scaffold - SAR of varying functional groups in $S_1'-S_2$ subsite of SARS-CoV 3CL^{pro}. Effect of varying size of alkyl chains attached to the amide is shown. *Standard error for IC_{50} value is not available.

●Size of alkyl chain attached to amide group

Among these compounds, XRR (isopropyl amide) and REQ (cyclobutyl amide) display maximum inhibition of SARS-CoV 3CL^{pro} (Figure 3.14). An increase (RF5,

RFF) or decrease (XPY, XTF, RF4) in the size of alkyl chain results in a slight decrease in the inhibitory activity, possibly as a result of steric clashes or decrease in van der Waal interactions, respectively.

***P*₃-extended versus *P*₃-truncated benzotriazole analogues**

The X-ray crystal structure of SARS-CoV 3CL^{PRO} in complex with the *P*₃-extended benzotriazole analogue suggested that the *P*₃ functional group is relatively solvent exposed, and might impose an entropic penalty for solvation [39]. Therefore, a series of second generation *P*₃-truncated benzotriazole analogues were designed. In general, the inhibitory activity of *P*₃-truncated benzotriazole analogues was higher than the corresponding *P*₃-extended benzotriazole analogues. For example, the IC₅₀ value of *P*₃-truncated R30 is more than two-folds lower than the corresponding *P*₃-extended XTF, Figure 3.15.

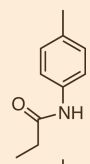
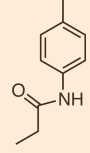
Name	S ₁ '-S ₂	% Inhibition (100 μM)	IC ₅₀ (μM)
R30		97	2.9±0.1
XTF		94	7±1.5

Fig. 3.15.: Inhibitory activity of *P*₃-extended (R30) versus *P*₃-truncated (XTF) benzotriazole analogues.

Modifications of N-alkyl phenylamide for the S_1' - S_2 site in P_3 -truncated benzotriazole analogues

●Size of alkyl chain attached to phenyl amide

For SARS-CoV 3CL^{pro}, inhibitory activity decreases with increase in the size of alkyl chain (Figure 3.16). The ethyl amide (R30) is the most potent compound in this series with an IC₅₀ value of 2.9 μ M, while the bulky cyclohexyl amide (RF1) and phenyl amide (RFM) displayed almost a total loss in inhibitory activity.

●Long alkoxy groups attached to phenyl amide

Longer flexible chains without branching (R2Y, R2X in Figure 3.17) inhibit SARS-CoV 3CL^{pro} with IC₅₀ values <10 μ M. However, branched bulkier alkyl chain (REL) show a loss of inhibitory activity possibly due to steric clashes.

It was observed that compared to the benzamide group (RFM, Figure 3.16), benzalamine group (R2W, Figure 3.17) displayed strikingly higher inhibitory activity. This observation suggests that the increase in flexibility due to loss of amide oxygen allows this group to better accommodate and interact in the S_1' - S_2 pocket of the enzyme. Based on this observation, a series of biaryl compounds were evaluated.

●Biaryls

Compounds with biaryls in the S_1' - S_2 pocket dramatically increased the inhibitory potency of the inhibitors. Phenyl biaryl (T8L, Figure 3.17) is the most potent non-covalent inhibitor of SARS-CoV 3CL^{pro} identified so far. Other inhibitors in this group also display low micromolar and submicromolar potency for inhibition of SARS-CoV 3CL^{pro}.

In summary, several non-covalent inhibitors of SARS-CoV 3CL^{pro} that display low micromolar and submicromolar inhibition potency have been identified. Potent inhibitors were identified in both the furanyl amide and benzotriazole scaffold series.

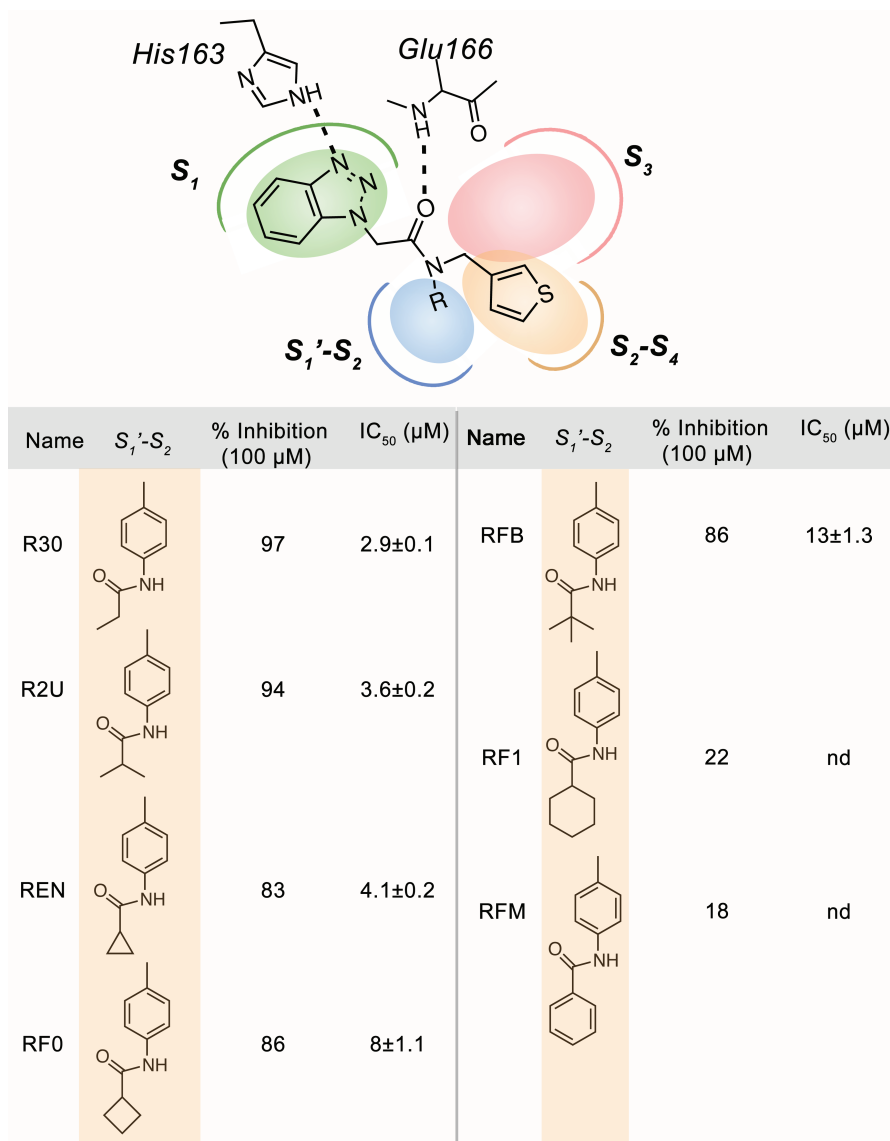


Fig. 3.16.: P_3 -truncated benzotriazole scaffold - SAR of varying functional groups in $S_1'-S_2$ subsite of SARS-CoV 3CL^{pro}. Effect of varying size of alkyl chain attached to phenyl amide is shown.

In general, the P_3 -truncated benzotriazole analogues were more potent than corresponding P_3 -extended benzotriazole analogues. The most potent compound from the furanyl amide series, REY, also exhibit antiviral activity against SARS-CoV in Vero-E6 cells with an EC_{50} value of 13 μM [38]. Experiments to determine the antiviral activity of the potent benzotriazole analogues failed to show antiviral efficacy. Presence of three polar nitrogen atoms in the benzotriazole ring of these analogues may

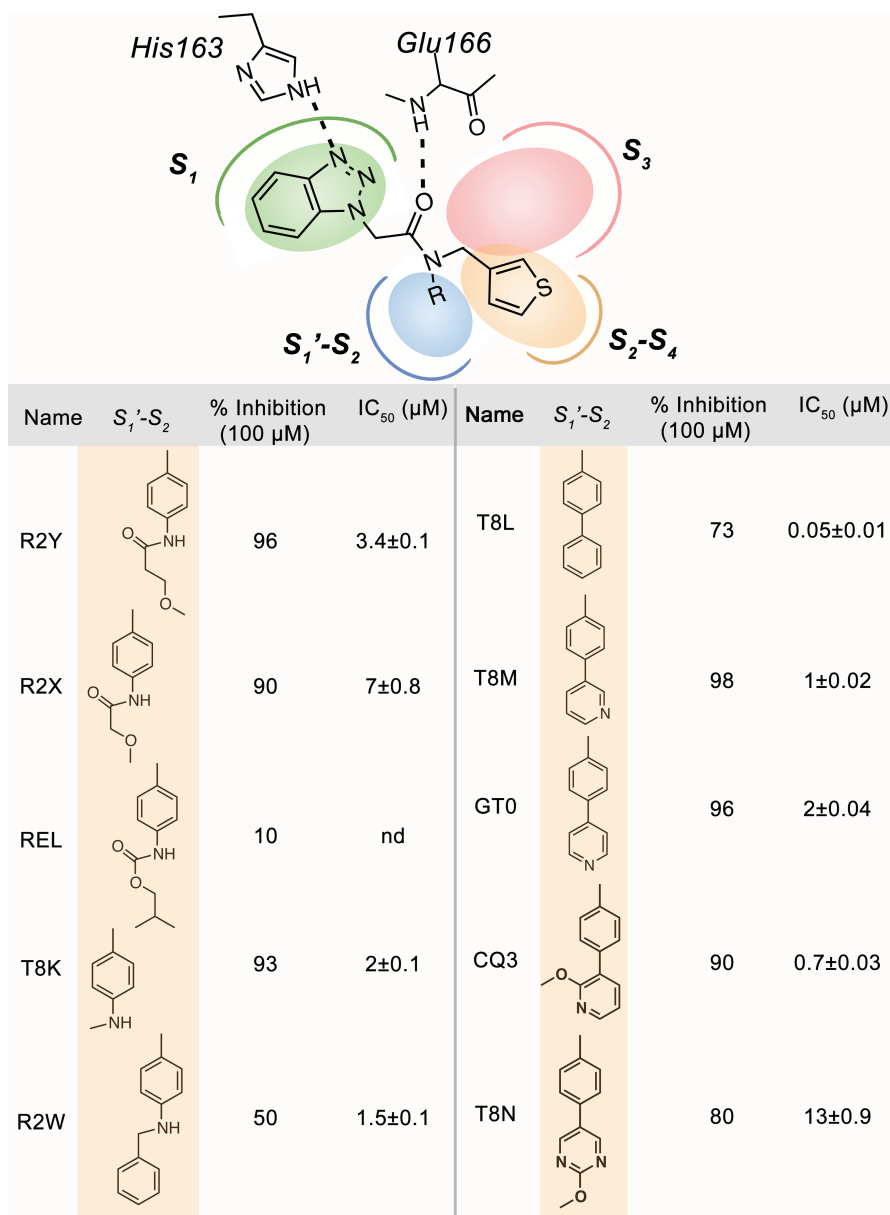


Fig. 3.17.: P_3 -truncated benzotriazole scaffold - SAR of varying functional groups in S_1' - S_2 subsite of SARS-CoV 3CL^{pro} *cont...* Effect of long alkoxy and groups attached to phenyl amide and biaryls is shown.

potentially result in poor cell permeability that would explain their poor antiviral activity.

3.4 Identification of inhibitors with broad-spectrum activity against 3CL^{pro} from different coronaviruses

One of the goals of this dissertation project on coronavirus 3CL^{pro} research is to understand the determinants for molecular recognition of 3CL^{pro} protease across different coronavirus subclasses. Insights into similar and distinct features utilized by different 3CL^{pro} proteases for ligand interaction will aid in the development of potent broad-spectrum inhibitors that can be used as antiviral agents against existing and emerging human CoV infections.

It is hypothesized that by engaging interactions between the inhibitor and the protein backbone and/or conserved residues, potent broad-spectrum inhibitors of human CoV 3CL^{pro} can be developed. This hypothesis is supported by the following observations. First, 3CL^{pro} from different subclasses of human CoV share a minimum of 40% sequence identity. Moreover, the sequence identity is maintained more rigorously in the vicinity of 3CL^{pro} active site. Second, recent success in efforts to develop inhibitors of HIV protease ([78, 80–82]) presents a proof-of-principle foundation for an approach capitalizing on conserved structural features to develop broad-spectrum antiviral agents. One of the main goals of this study was to develop both covalent and non-covalent inhibitors of 3CL^{pro}. Success with rhinovirus protease inhibitor AG-7088 into Phase I clinical trials demonstrated the potential of compounds with covalent mode of action to be advanced as effective antiviral agents [83]. Additionally, insights gained from the successful identification of several reversible non-covalent inhibitors of SARS 3CL^{pro} can be utilized to develop inhibitors targeting 3CL^{pro} from other coronavirus subclasses [38, 39].

In this study, an in-house library of 237 non-covalent compounds was tested against a panel of 3CL^{pro} enzymes from different coronaviruses. 3CL^{pro} enzymes from SARS, OC43, MHV, HKU1, HKU5 and MERS coronaviruses were included. SARS, OC43, KHU1 and MERS are human coronaviruses, MHV infects mice, and HKU5 is a bat coronavirus closely related to MERS-CoV. The materials and methods

utilized for testing the compound library are described in detail in Chapter 2. Briefly, the percent inhibition of each enzyme was determined in the presence of 100 μM of compound. Compounds displaying more than 50% inhibition at this concentration were further evaluated in a dose response assay to determine the IC_{50} values. Compounds exhibiting inhibitory activity against multiple 3CL^{pro} enzymes are listed in Figures 3.18 and 3.19.

Several compounds belonging to both the furanyl-amide and benzotriazole scaffolds were identified that displayed inhibitory against multiple 3CL^{pro} enzymes. Since the compound library was originally synthesized for SARS-CoV 3CL^{pro} enzymes, it was not surprising that most of the compounds listed in Figure 3.19 displayed better inhibitory activity against SARS-CoV 3CL^{pro} compared to other 3CL^{pro} enzymes we tested. Among the furanyl-amide analogues, it was observed that 3CL^{pro} enzymes from OC43, MHV and HKU1 can only accommodate small alkyl chains attached to the phenyl ring at the P_2 position, suggesting that the P_2 pocket of these enzyme is smaller than SARS-CoV 3CL^{pro} . For example, compound FMW with a bulky substitution at the P_2 position inhibits SARS-CoV 3CL^{pro} with an IC_{50} value of 4.5 μM (Figure 3.11) but displays <20 % inhibition of other 3CL^{pro} enzymes tested here. Compound XM2, followed by compounds QEF and RWB, are the most potent inhibitors in this class with broad spectrum activity against multiple 3CL^{pro} enzymes (Figure 3.19). Both of these compounds contain an imidazole ring in place of a furan ring at the position that occupies the S_1 subsite of 3CL^{pro} . The imidazole ring would not only maintain the hydrogen-bond with Gly143, but it also contains a hydrogen-bond donor that might engage in additional interactions with the enzymes. It is predicted that the cross-reactivity of XM2 and QEF emanates from the formation of additional hydrogen-bonding interactions with the enzymes.

Among the benzotriazole analogues, it was observed that P_3 -truncated benzotriazole analogues are better cross-reactive inhibitors than the corresponding P_3 -extended benzotriazole analogues. Most of the cross-reactive inhibitors in this class displayed lowest inhibitory activity against MHV 3CL^{pro} . Compounds RF4, R2Y, R2X, T8M

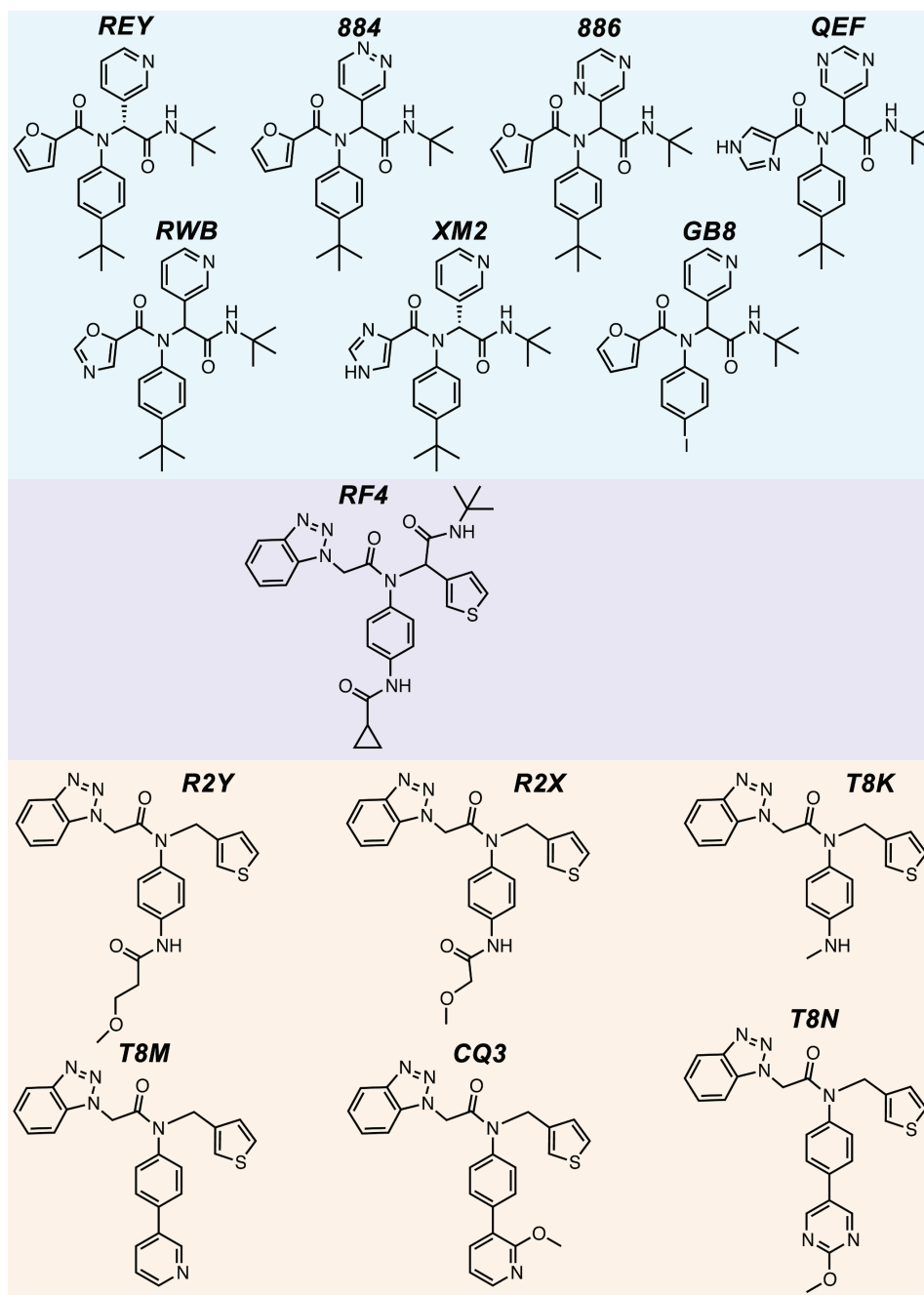


Fig. 3.18.: **Chemical structures of broad spectrum 3CL^{pro} inhibitors.** Inhibitors from the furanyl amide, *P*₃-extended benzotriazole and *P*₃-truncated benzotriazole scaffolds are highlighted in cyan, purple and orange, respectively.

and T8N inhibited all the 3CL^{pro} enzymes tested, albeit with different potencies. Although CQ3 did not display good inhibitory activity against MHV 3CL^{pro}, it was

compound	3CL ^{pro} % Inhibition (IC ₅₀ μM)				
	SARS	OC43	MHV	HKU1	HKU5
REY	99 (1.5±0.3)	69 (30±4.7)	38 (nd)	57 (62±7.5)	73 (43±5.8)
884	94 (10±0.6)	56 (33±11)	67 (6±2.3)	66 (43±6.6)	28 (nd)
886	96 (5.5±0.3)	29 (nd)	77 (3.8±0.9)	77 (27±1.9)	17 (nd)
QEF	67 (14*)	81 (16±2.7)	72 (37±3.7)	86 (15±0.9)	87 (38±4.7)
RWB	98 (39±5.9)	98 (8.3±1.5)	100 (7.4±0.4)	99 (10±0.7)	89 (23±2.4)
XM2	98 (4.5±0.2)	93 (4.1±0.5)	92 (1.4±0.2)	92 (5±0.2)	94 (20±4.5)
GB8	66 (27)	81 (24±2.9)	94 (11±1.6)	95 (10±2.1)	52 (53±7)
RF4	86 (9*)	84 (8.8±0.5)	81 (31±2.2)	96 (111±36)	87 (9.4±0.6)
R2Y	96 (3.4±0.1)	92 (5±0.2)	64 (41±6.5)	66 (26±1.7)	96 (4.1±0.3)
R2X	90 (7±0.8)	91 (7±0.7)	51 (60±8.1)	68 (27±3.7)	94 (6.2±0.3)
T8K	93 (2.1±0.1)	84 (10±1.0)	36 (nd)	55 (29±3.3)	94 (3.1±0.2)
T8M	98 (1±0.02)	95 (2.4±0.2)	72 (26±2.9)	84 (6.2±0.4)	98 (3.4±0.2)
CQ3	90 (0.7±0.03)	77 (1±0.1)	43 (nd)	55 (1.6±0.2)	91 (0.9±0.05)
T8N	80 (13±0.9)	90 (3.7±0.5)	67 (25±1.8)	73 (12±0.7)	87 (8.3±3.5)

Fig. 3.19.: **Broad spectrum 3CL^{pro} inhibitors.** A list of inhibitors with broad spectrum inhibitory activity against multiple 3CL^{pro} enzymes is provided. Inhibitors from furanyl amide, *P*₃-extended benzotriazole and *P*₃-truncated benzotriazole scaffolds are highlighted in cyan, purple and orange, respectively. % inhibition was determined at final compound concentration of 100 μM. IC₅₀ values were determined from a dose-response curve; IC₅₀ values are given in parenthesis. nd - not determined. *Standard error for IC₅₀ value is not available.

able to achieve sub-micromolar inhibitory activity against 3CL^{pro} from SARS-CoV and HKU5-CoV.

3.5 X-ray crystal structure determination of SARS-CoV 3CL^{pro}-inhibitor complexes

Crystal structure analysis is a great tool to validate SARs and gain further insights into interactions important for specificity versus cross-reactivity of inhibitors. In addition to the reported crystal structure of the most potent compound ML188 (or REY), nine high-resolution crystal structures of inhibitors (five from the furanyl amide series and four from the benzotriazole series) bound to SARS-CoV 3CL^{pro}, including complex with the cross-reactive XM2 agent, were determined.

3.5.1 X-ray crystal structures in complex with *P*₃-truncated benzotriazole inhibitors

Figures 3.20-3.21 show the crystal structures of SARS-CoV 3CL^{pro} bound to four compounds (CQ3, R30, R2Y and R2X) from the benzotriazole series. All these compounds have similar binding orientations in the active site and form equivalent hydrogen bonds with His163 and Glu166 (Figure 3.22). Compared to the compounds in the furanyl-amide series (discussed in the next section), which form at least four hydrogen bonds with SARS-CoV 3CL^{pro}, the benzotriazole analogues form only two hydrogen bonds with the enzyme. Interestingly, these inhibitors display an induced-fit binding mode, which is slightly different from the traditional S_1' - S_4 subsites that accommodate the peptidomimetic compounds. The thiophene ring occupies the S_2 - S_4 subsites, while the other functional group occupies the S_1' - S_2 subsites. Rearrangement of Gln189 and Met49 is largely responsible for the formation of the induced-fit binding pockets. In addition to the conserved hydrogen-bond formation with His163 and Glu166, the alkoxy oxygen of compounds R2Y and R2X might also engage Thr25 with hydrogen-bonding interactions in the S_1' - S_2 pocket.

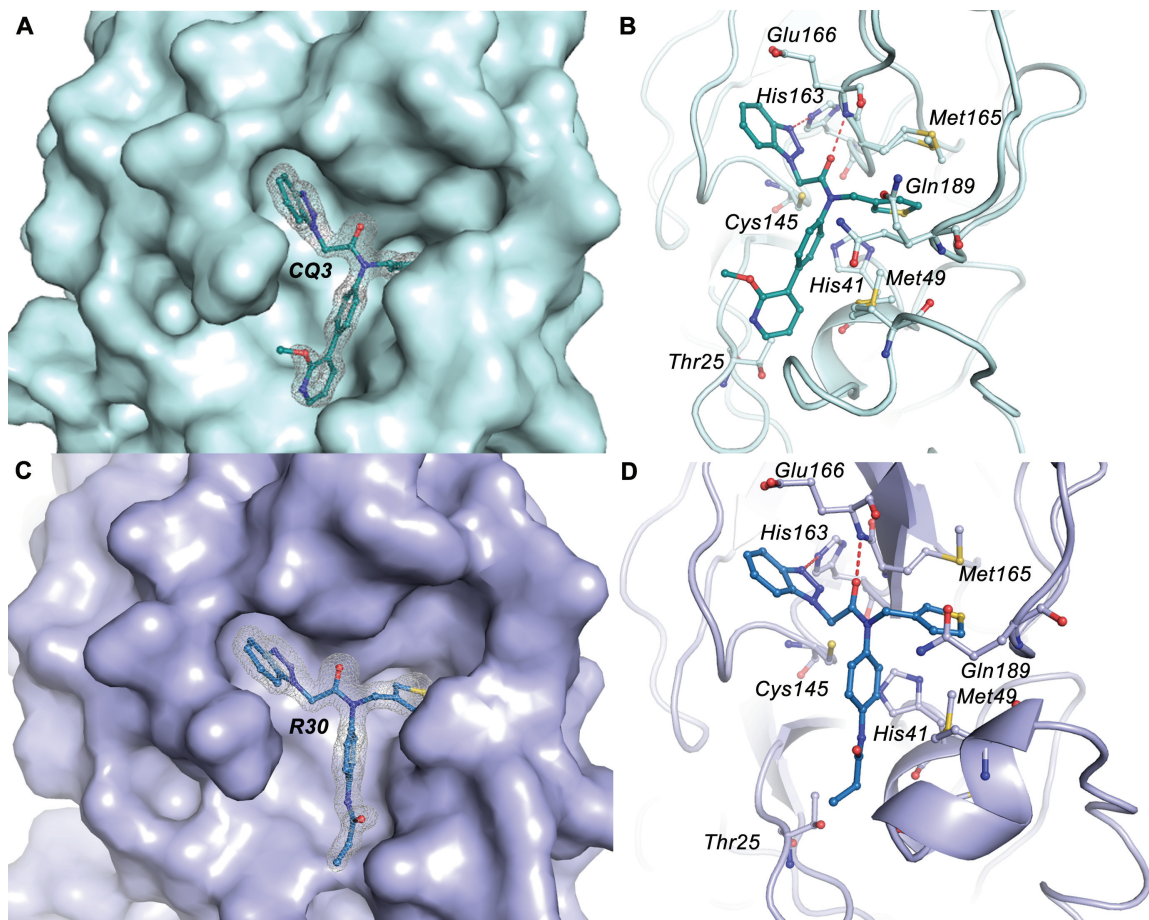


Fig. 3.20.: X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with *P*₃-truncated benzotriazole inhibitors CQ3 and R30. **A.** and **C.** Solvent-accessible surface of SARS-CoV 3CL^{pro} in complex with inhibitors CQ3 and R30, respectively. Inhibitor molecules are displayed in ball and stick models. Electron density associated with the inhibitors is shown as a $F_o - F_c$ electron density difference map contoured to 3σ (mesh). **B.** and **D.** SARS-CoV 3CL^{pro}-CQ3 and SARS-CoV 3CL^{pro}-R30 complexes, respectively, are illustrated with the SARS-CoV 3CL^{pro} backbone represented as a ribbon model and relevant amino acids that interact with the inhibitors represented as ball and sticks. Hydrogen bonds are depicted as red dashed lines.

3.5.2 X-ray crystal structures in complex with furanyl-amide inhibitors

Figure 3.23 show the crystal structures of SARS-CoV 3CL^{pro} in complex with compounds 886, CUG, FMW, KWW and XM2. All these compounds have similar binding orientations in the active site and form equivalent hydrogen bonds with His163

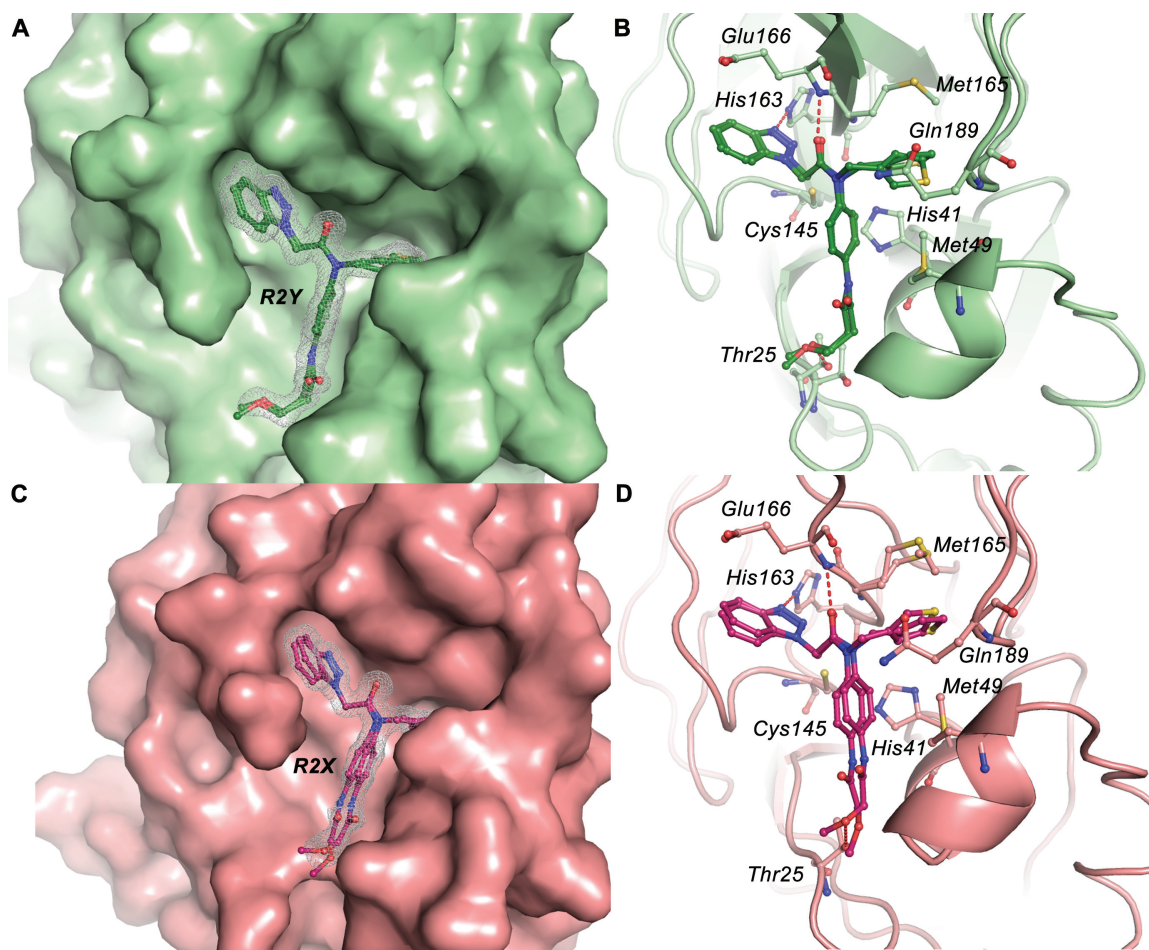


Fig. 3.21.: X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with *P*₃-truncated benzotriazole inhibitors R2Y and R2X. **A.** and **C.** Solvent-accessible surface of SARS-CoV 3CL^{pro} in complex with inhibitors R2Y and R2X, respectively. Inhibitor molecules are displayed in ball and stick models. Electron density associated with the inhibitors is shown as a $F_o - F_c$ electron density difference map contoured to 3σ (mesh). **B.** and **D.** SARS-CoV 3CL^{pro}-R2Y and SARS-CoV 3CL^{pro}-R2X complexes, respectively, are illustrated with the SARS-CoV 3CL^{pro} backbone represented as a ribbon model and relevant amino acids that interact with the inhibitors represented as ball and sticks. Hydrogen bonds are depicted as red dashed lines.

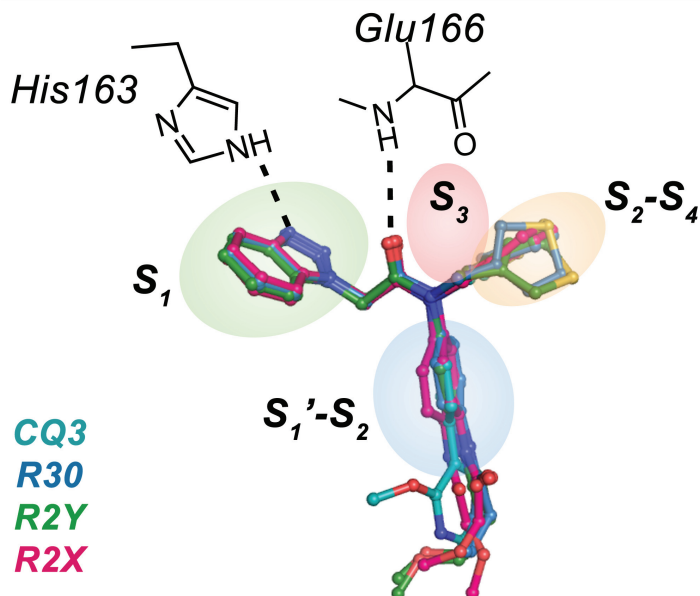


Fig. 3.22.: Superposition of inhibitors CQ3, R30, R2Y and R2X from the X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with inhibitors. X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with P_3 -truncated benzotriazole inhibitors CQ3, R30, R2Y and R2X reveal similar binding orientations in the active site of the enzyme. These inhibitor molecules display an induced-fit binding mode with different functional groups binding in the S_1 , S_1' - S_2 , S_2 - S_4 and S_1 subsites of the enzyme. Conserved hydrogen-bonding interactions with His163 and Glu166 are depicted as dashed lines.

and Glu166, Figure 3.24. Compared to the benzotriazole analogues, which form two hydrogen bonds with SARS-CoV 3CL^{pro}, furanyl-amide analogues form at least four direct hydrogen bonds with the enzyme. These inhibitors occupy the traditional S_1' - S_3 subsites that accommodate the peptidomimetics. Furan oxygen (or imidazole nitrogen) and amide oxygen form bifurcated hydrogen-bonds with the backbone amide of Gly143 in the S_1' subsite. Pyridine nitrogen engages His163 side chain with a hydrogen-bond in the S_1 pocket. Second amide oxygen of the inhibitor backbone interacts with Glu166 through a hydrogen-bond formation.

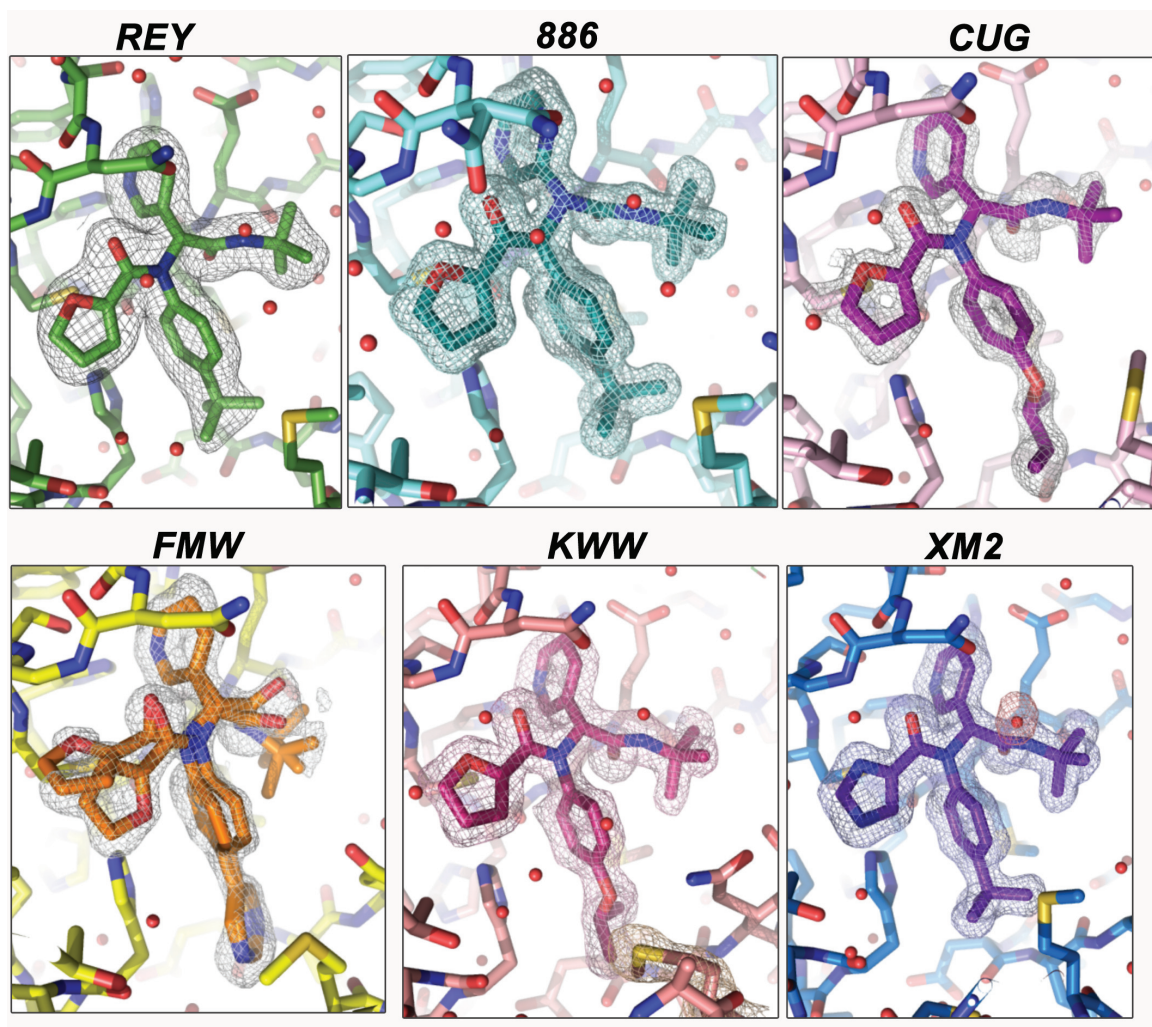


Fig. 3.23.: X-ray crystal structure of SARS-CoV 3CL^{pro} in complex with furanyl-amide inhibitors 886, CUG, FMW, KWW and XM2. Inhibitor molecules and the neighboring amino acid side chains are displayed in ball and stick models. Water molecules are represented as red spheres. Electron density associated with the inhibitors is shown as a $F_o - F_c$ electron density difference map contoured to 3σ (shown as mesh). Previously solved crystal structure of SARS-CoV 3CL^{pro} in complex with compound REY is shown for reference.

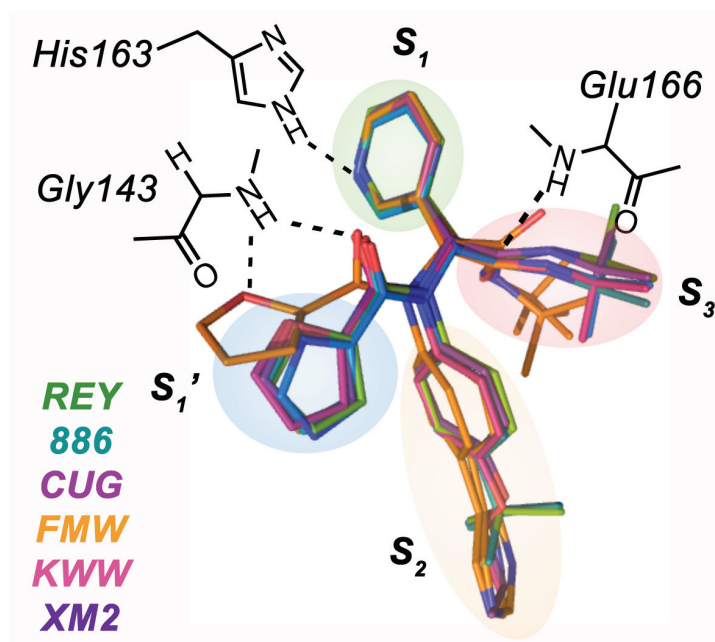


Fig. 3.24.: **Superposition of inhibitors REY, 886, CUG, FMW, KWW and XM2 from the X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with inhibitors.** X-ray crystal structures of SARS-CoV 3CL^{pro} in complex with furanyl-amide inhibitors REY, 886, CUG, FMW, KWW and XM2 reveal similar binding orientations in the active site of the enzyme. Different functional groups of these inhibitor molecules occupy S_1' , S_1 , S_2 and S_3 subsites of the enzyme in a manner similar to the binding of substrate-mimetic compounds. Conserved, direct hydrogen-bonding interactions with Gly143, His163 and Glu166 are depicted as dashed lines.

SARS-CoV 3CL^{pro}-XM2 crystal structure

Cross-reactive inhibitor XM2 engages the enzyme via hydrogen-bonding interactions with conserved structural elements (either backbone amides or side chains of conserved residues) of SARS-CoV 3CL^{pro}. These interactions are similar to the interactions formed by other inhibitors in this class. Interestingly though, as predicted through SAR analysis, the imidazole ring of XM2 forms a network of four additional hydrogen-bonds through bridging water molecules. These hydrogen-bonds engage backbone of SARS-CoV 3CL^{pro} via Val42 and Cys44, and conserved His41 of the catalytic dyad. Additional structural analysis reveals that MPD (2-Methyl-2,4-pentanediol) molecule, added as a cryo-protectant during crystallization, binds

at the bottom of S_2 pocket in the active site and can be displaced by corresponding non-conserved Met25 residue in other 3CL^{pro} enzymes (Thr25 in SARS-CoV 3CL^{pro}).

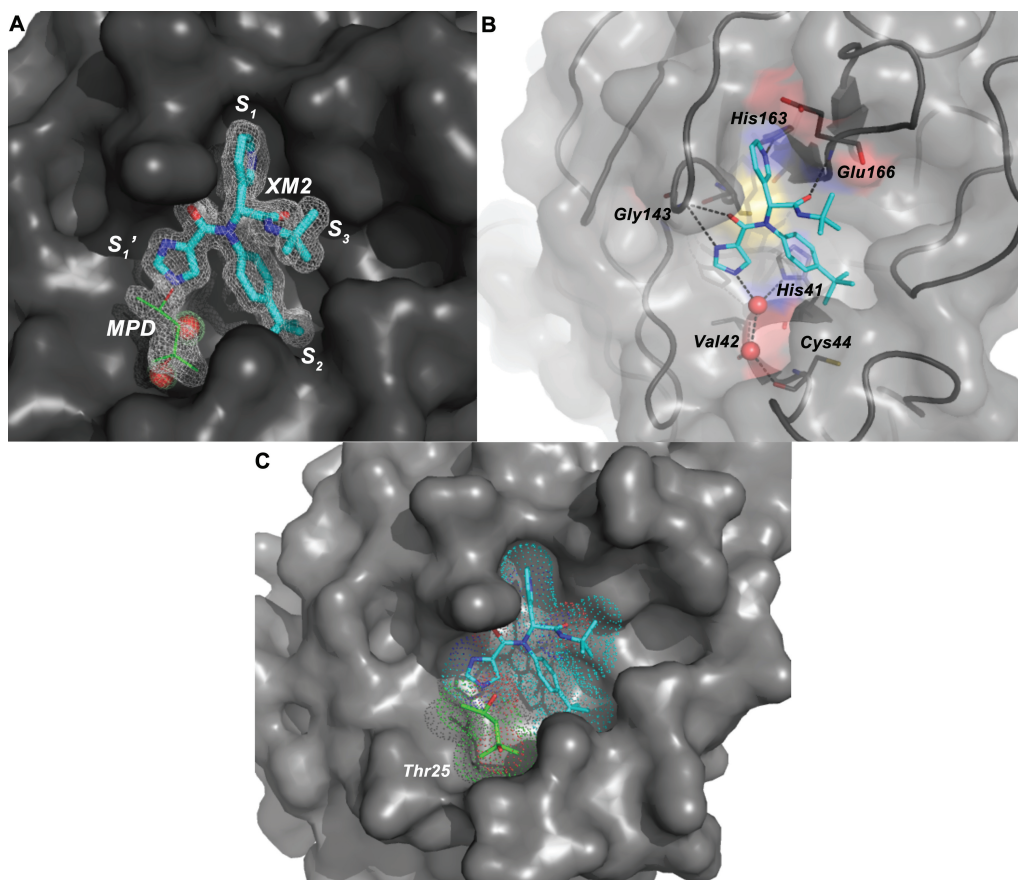


Fig. 3.25.: **SARS 3CL^{pro}-XM2 complex crystal structure.** **A.** Solvent-accessible surface of SARS-CoV 3CL^{pro} in complex with inhibitor XM2, respectively. Inhibitor molecule and solvent molecule MPD are displayed as stick models. Electron density associated with the ligands is shown as a $F_o - F_c$ electron density difference map contoured to 3σ (mesh). Water molecules are shown as red spheres. **B.** Interactions between XM2 and SARS 3CL^{pro} active site residues are illustrated. Hydrogen bonds are shown as dashed lines. Water molecules are shown as red spheres. **C.** MPD molecule (green sticks) was modeled near the active site in the extra electron density observed in $F_o - F_c$ map. MPD molecule (added as cryo-protectant) found near the active site in SARS 3CL^{pro} can be displaced by non-conserved Met25 residue in other 3CL^{pro} enzymes, while this residue is Thr25 in SARS 3CL^{pro}).

3.6 Summary

In an effort to develop potent non-covalent inhibitors of SARS-CoV 3CL^{pro}, we designed a focused in-house library of 237 compounds. The compound library was designed based on the initial hits from a high-throughput screen of approximately 30,000 compounds. We identified several potent (low micromolar furanyl-amide analogues, sub-micromolar benzotriazole analogue) inhibitors of SARS-CoV 3CL^{pro}. SAR of these compounds was developed and suggested significance of key functional groups that occupy different subsites in SARS-CoV 3CL^{pro}. Based on SAR, *S*₂ subsite is the most variable site and can tolerate a variety of substitutions in SARS-CoV 3CL^{pro}. Several SARS-CoV 3CL^{pro} inhibitors also displayed cross-reactivity against 3CL^{pro} enzymes from coronaviruses OC43, MHV, HKU1 and HKU5. The X-ray crystal structure analysis of SARS-CoV 3CL^{pro} in complex with the most potent broad-spectrum inhibitor XM2 suggests that formation of additional water-mediated hydrogen-bonds by XM2 might be playing a crucial role in cross-reactivity. These results suggest that designing of potent non-covalent inhibitors of 3CL^{pro} enzymes that also display cross-reactive activity can be achieved.

CHAPTER 4. MECHANISTIC BASIS FOR THE TEMPERATURE
SENSITIVITY OF 3CL^{PRO} MUTANTS OF MURINE HEPATITIS VIRUS (MHV)

Parts of the data and text in this chapter have been included in a manuscript entitled “Mechanistic basis for the temperature sensitivity of 3CL^{pro} mutants of murine hepatitis virus” and is currently under preparation.

4.1 Abstract

All coronaviruses (CoVs) require the protease activity of viral-encoded nsp5 (3CL^{PRO}) to process the viral replicase polyproteins 1a and 1ab. Domains 1 and 2 of 3CL^{PRO} form the catalytic fold, while unique domain 3 is critical for 3CL^{PRO} dimerization. For the CoV murine hepatitis virus (MHV), our collaborator, Prof. Mark Denison at Vanderbilt University, recently identified multiple temperature sensitive (*-ts*) mutations in domains 2 and 3 that are each complemented by a common, structurally distant revertant H134Y in domain 2, implicating long-distance, intramolecular communication in regulation of CoV 3CL^{PRO} activity. However, the structural and biochemical bases for the *-ts* and revertant phenotypes were not determined. In this study, biochemical, biophysical and structural analyses of these physiological relevant mutations of MHV 3CL^{PRO} were performed. It is demonstrated that V148A mutation results in the global destabilization of the protein structure, which denatures at elevated temperatures. Interestingly, the H134Y mutation increases the melting temperature (T_m) by 7 °C compared to the V148A mutant. Moreover, the V148A/H134Y double mutant is more stable than the V148A mutant, albeit less stable than the wildtype or the single mutant H134Y. Structural analysis reveals that V148A occupies a structural ‘weak spot’, while the H134Y mutation provides a rather general mechanism for enhancing the structural stability of MHV 3CL^{PRO}. Emergence of H134Y to suppress

the *-ts* phenotype of multiple independent mutants of MHV 3CL^{pro} as well as the natural existence of a tyrosine residue at position 134 in several other CoV 3CL^{pro} enzymes highlight the potential evolutionary significance of this residue as a facilitator of structural stability for multiple 3CL^{pro} enzymes.

4.2 Introduction

CoVs are positive-strand RNA viruses that include important human pathogens like severe acute respiratory syndrome coronavirus (SARS-CoV) and the recently emerged Middle East respiratory syndrome coronavirus (MERS-CoV). Besides infecting humans, CoVs are also pathogenic to a variety of other vertebrates including but not limited to livestock, mouse, cats, dogs etc. Upon entry inside the host cells, the viral positive-strand RNA translates into replicase polyproteins 1a and 1ab. CoV 3CL^{pro} protease is a non-structural protein that processes the viral polyproteins into mature non-structural proteins required for the assembly of replicase complex [1,40,41,84]. Due to its indispensable role in viral replication, 3CL^{pro} serves as an attractive drug target for the development anti-coronaviral therapeutics [37–39,53]. Structures of several 3CL^{pro} from different alpha, beta and gamma CoVs have been determined and reveal an overall similar structural architecture [60,85,86]. 3CL^{pro} is comprised of three distinct domains; domains 1 and 2 form the typical chymotrypsin-like fold that contains the His41-Cys145 catalytic dyad and the substrate-binding site [42]. CoV 3CL^{pro} domain 3 is unique among chymotrypsin-like enzymes and has been shown to be important for 3CL^{pro} dimerization [47,48]. The role of amino acids present at the dimer interface or substrate-binding pocket that regulates 3CL^{pro} structure and activity has been illustrated in great depth through mutagenesis, biochemical and structural studies [43–47,60]. However, we are only beginning to understand the mechanisms by which a distant residue can modulate dimerization or catalytic activity of 3CL^{pro} through long-range interactions [60,87–89].

Previous studies have suggested the role of long-range interactions in regulating the activity of 3CL^{pro} of murine hepatitis virus (MHV). Sparks et al. showed that a temperature sensitive (*-ts*) V148A mutation in MHV 3CL^{pro} impairs viral growth and polyprotein processing at 40 °C [90]. Emergence of a second site H134Y mutation, physically distant from V148A mutation site, suppresses the *-ts* phenotype [90,91]. Moreover, MHV with two other independent *-ts* mutations, S133A and F219L, also select for H134Y as the second site compensatory mutation [90,91]. Since these mutations showed differential impairment in the processing of nsp8 and nsp5 in the virus-infected cells, it was suggested that the altered polyprotein processing at specific cleavage sites results in the *-ts* phenotype of these mutant viruses [91]. It was also suggested that the suppression of *-ts* phenotype by the distant H134Y mutation is mediated through long-range interactions, and the long-range interactions play a significant role in regulating 3CL^{pro} enzymatic activity during polyprotein processing and virus replication [90,91]. However, the structural basis of temperature sensitivity of mutant enzymes and the suppression of *-ts* phenotype by H134Y mutation was not elucidated.

Mutation of these residues may influence MHV 3CL^{pro} in a variety of ways in the virus-infected cells. These mutations can impact enzymatic activity by altering substrate binding site or catalysis, influence the kinetic and thermodynamic stability of protein structure, impair polyprotein processing at the rate-limiting cleavage step or alter interaction of 3CL^{pro} with a viral/host binding partner. To understand the mechanistic basis for temperature sensitivity of the physiologically relevant *-ts* V148A mutation of MHV 3CL^{pro} and its rescue by the distant H134Y mutation, it was hypothesized that the V148A mutation leads to significant destabilization of the protein structure at elevated temperatures. A series of kinetic, biophysical and molecular dynamics studies were conducted on the *in vitro* purified wildtype, V148A, H134Y and V148A/H134 enzymes. Utilizing a combination of biochemical investigations, molecular dynamics (MD) simulations and constraint network analysis, a potential mechanism for the observed phenotypes in the MHV 3CL^{pro} variants is proposed.

Moreover, to the best of our knowledge, this study is the first to report biochemical characterization of physiologically relevant mutations of coronavirus 3CL^{pro}. Finally, structural analysis of the H134Y mutation suggests a general mechanism for improved stability of 3CL^{pro} that may have implications when utilizing structurally unstable *-ts* mutant viruses for vaccine design.

4.3 Materials and methods

4.3.1 Construct design and expression of MHV 3CL^{pro}

The construct design and protocol for the expression of MHV 3CL^{pro} wildtype and the mutants have been described in Chapter 2, Section 2.3.

4.3.2 Protein purification

The purification protocol for MHV 3CL^{pro} wildtype has been described in detail in Chapter 2. For the CD experiments (section 4.3.4), gel-filtration chromatography was employed as the final purification step. All experiments were performed using freshly purified (i.e. not frozen) proteins.

Total activity units (μM product/min), specific activity (units/mg) and milligrams of protein obtained (BioRad protein assay) were determined after each chromatographic step to calculate the final protein yield.

4.3.3 Thermal inactivation at permissive and non-permissive temperatures

Aliquots of freshly purified proteins with a concentration of 10 μM were transferred to microcentrifuge tubes and kept at ambient temperature for 10-15 minutes before incubating at higher temperatures. The samples were then incubated at ambient temperature and either at 30 °C or 40 °C in a circulating water bath for a period of two hours. Samples incubated at room temperature served as controls. Aliquots

were taken out at different time points within a two hour period and one hundred-fold dilutions were prepared in an assay buffer containing 50 mM HEPES pH-7.5, 0.1 mg/mL BSA, 0.01% Triton X-100 and 2 mM DTT at ambient temperature. Then, 80 μ L of the enzyme mixed in assay buffer was dispensed into the wells of a black, half-area, 96-well plate (Corning). Data were collected in triplicate. The cleavage of a peptide substrate with an internal FRET donor-quencher pair, (HilyteFluorTM-488)-ESATLQSGLRKAK-(QXLTM-520)-NH₂ (AnaSpec, Inc.), was used to determine the protease activity of the enzymes. The reactions were initiated by adding 20 μ L of 10 μ M substrate to 80 μ L of the enzyme solution. The total reaction volume was 100 μ L with final enzyme and substrate concentrations of 80 nM and 2 μ M, respectively. The increase in fluorescence over time was monitored in a BioTek Synergy H1 plate reader using a filter cube with 20 nm bandwidths for excitation and emission wavelengths of 485 nm and 528 nm, respectively. Enzymatic activity was determined by measuring the initial slope of the progress curve. Percent (or fractional) residual activity, calculated with respect to the activity measured at 0 min incubation, was plotted as a function of incubation time at a given temperature.

For samples displaying a reduction in the enzymatic activity upon incubation at 30 °C or 40 °C, fractional residual activity versus time of incubation (min) data were fit into Equation 4.1 using non-linear regression program GraphPad Prism version 6.0 and the value of decay rate constant was determined.

$$Y = \text{Plateau} + (Y_0 - \text{Plateau}) \times \exp^{-k(X-X_0)} \quad \text{If } X > X_0, Y_0 \quad (4.1)$$

In Equation 4.1, Y is the fractional residual activity, X is the time of incubation in minutes, X_0 is the time at which the decay begins, Y_0 is the average Y value upto time X_0 , plateau is the Y value at infinite times, k is the decay rate constant.

4.3.4 Determining secondary structural changes at permissive and non-permissive temperatures

To monitor changes in the protein secondary structure, far-UV CD signal was measured for the wildtype and mutant enzymes upon incubation at permissive (30 °C) and non-permissive (40 °C) temperatures. CD signal was monitored between 205–250 nm using Chirascan CD spectrophotometer (Applied Photophysics). CD signal for 0.1 M potassium phosphate buffer, pH-7.5 was measured as a reference. Enzymes were diluted to a final concentration of 1.0 μ M in 0.1 M potassium phosphate buffer, pH-7.5 and transferred to a 1 cm path length CD cuvette with Teflon stopper to prevent evaporation. Samples were constantly stirred while the CD signal was measured at regular intervals for a period of two hours at given temperatures.

4.3.5 Determination of melting temperature, T_m , using circular dichroism

To compare the stability of the wild type and mutant enzymes, CD signal at 222 nm was monitored using Chirascan CD spectrophotometer (Applied Photophysics) to determine the melting temperature, T_m . Samples were prepared as described in the previous section. Samples were constantly stirred as the temperature was raised continuously from 10 °C to 90 °C at a constant rate of 0.5 °C /min. Prior to the thermal scans for protein samples, 222 nm CD signal for 0.1 M potassium phosphate buffer, pH-7.5 was measured at 10 °C as reference. CD signal, obtained after subtracting the reference signal, was plotted as a function of temperature, and the melting temperature (T_m) was calculated as the inflection point of the melting curves using non-linear regression program SigmaPlot version 10. Final T_m value for each enzyme was calculated as the average of T_m values determined from three independent experiments.

4.3.6 Homology Modeling

A homology model of MHV 3CL^{pro} was built using the Prime module implemented in Schrödinger Suite [92]. 3CL^{pro} from human coronavirus HKU1 (PDB ID 3D23) was selected as the template for model building [86]. The sequence of HKU1 3CL^{pro} is 84% identical to that of MHV 3CL^{pro} and shares 92% overall sequence similarity. The model was built using an energy-based refinement and the model was then minimized using the OPLS2005 force-field. From this minimized structure, three mutant structures were created; V148A, H134Y, and the revertant double mutant V148A/H134Y. These mutations were made in the Maestro interface [92]. These studies were performed in collaboration with Dr. Laura Kingsley (Mesecar Lab).

4.3.7 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed using Gromacs 5.0.2 [93,94] using the Amber99SB-ILND forcefield [95]. Each system was prepared by solvating in TIP3P waters and adding the appropriate counter ions to neutralize the system, in this case four sodium atoms. The system was minimized using 1000 steps steepest descent and particle mesh Ewald (PME) summation with a grid size of 0.12 nm. A 2 fs integration time step was used. Following minimization, the surrounding water network was relaxed using 500 ps simulation in which all non-water atoms were restrained. Next, a 10 ns equilibration was performed wherein all atoms were free to move. Finally, a 50 ns production run was completed. This procedure was carried out for each simulation at both 30 °C and 40 °C, resulting in a total of eight simulations.

Upon completion of the simulations, the root mean square fluctuation (RMSF) was calculated at the alpha-carbon for each residue using the `g_rmsf` utility of gromacs [96]. In addition, the distance between the centers of mass of the catalytic dyad residues (H41 and C145) was calculated using the `gmx distance` utility [96]. The `gmx sasa` and `gmx select` tools were used to analyze the solvent accessible surface area (all defaults were used) and to count nearby waters (within 5.0 Å) of the group

of interest, respectively [96] These studies were performed in collaboration with Dr. Laura Kingsley (Mesecar Lab).

4.3.8 CNAalysis

Snapshots from the MD simulations were extracted every 100 ps resulting in a total of 501 snapshots for each simulation. These snapshots were used as an input into the CNA webserver and all default settings were used. The output of this analysis provides information on the predicted temperature transition points during the simulated unfolding of the protein and also information about the relative rigidity of various residues [97]. However, for our analysis we were specifically interested in unfolding nuclei, also known as ‘weak spots’ in the protein.

The top-10 predicted weak spots were calculated for each simulation. The number of times this residue appeared in all simulations (frequency) as well as the average score it received in those simulations (score) were calculated. From this list, residues that were predicted in half or more of the ensembles were collected These studies were performed in collaboration with Dr. Laura Kingsley (Mesecar Lab).

4.3.9 Mutual Information Analysis

The pairwise correlation between the residues was calculated using the MutInf program [98]. All standard settings were used and correlation was calculated based on the phi, psi, and chi angles of each residue which were calculated using the g_chi module of gromacs [96]. The 50 ns simulations were broken down into five “runs” each containing 1000 snapshots which were taken every 10 ps. Notably, this number of frames at this interval is on the low end of the author’s recommendations, but is not believed to affect the results. The data was then converted into a heatmap using the R-code developed by Sebastian Raschka [99]. The diagonals were set to equal zero rather than 1.0 for clarity These studies were performed in collaboration with Dr. Laura Kingsley (Mesecar Lab).

4.4 Results

4.4.1 Protein expression and purification

The canonical cleavage site of 3CL^{pro} closely resembles the TEV cleavage site ENLYFQ↓S. As a result, 3CL^{pro} cleaves between the Q and S amino acids from the construct, resulting in the removal of N-terminus (His)₆-tag and the production of 3CL^{pro} with an authentic N-terminus during expression in *E.coli*. Cleavage of the His-tag was confirmed from inability of the protein to bind nickel beads.

MHV 3CL^{pro} wildtype and mutants were purified using two (or three for CD experiments) sequential chromatographic steps. The final protein yield was determined by measuring the total units of activity, specific activity and the total amount of protein obtained after each chromatographic step during the entire purification process. Approximately 10 mg of wildtype MHV 3CL^{pro} can be obtained from 3 L of bacterial culture. Similar yields were obtained for the mutant proteins. Table 5.1 represents the purification summary for the wildtype MHV 3CL^{pro}.

Table 4.1.
Purification summary of MHV 3CL^{pro} from 3 L culture of *E.coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	199	19,255	97	1	100
Phenyl-Sepharose	26	7,429	283	3	39
DEAE	10	4,821	461	5	25

4.4.2 The H134Y mutation rescues the *-ts* V148A mutant from thermal inactivation at non-permissive temperature

A multitude of factors may contribute to the temperature sensitivity of V148A mutant inside the virus-infected cells. For example, the loss of the enzymatic activity

due to local structural changes in the active site, global unfolding of the protein structure, impaired polyprotein processing at the rate-limiting cleavage step or altered interaction with a viral/host binding partner may all result in *-ts* phenotype. To investigate whether the temperature sensitivity of the V148A mutant stems from the loss of activity at non-permissive temperature, we compared the ability of wildtype and mutant enzymes to retain protease activity upon incubation at permissive and non-permissive temperatures over a period of two hours.

All the enzymes maintained their full activity for two hours upon incubation at 30 °C (Figure 4.1.A, Table 4.2). Interestingly, the *-ts* V148A mutant lost > 90% of its activity during two hours incubation at the non-permissive temperature of 40 °C, while the wildtype enzyme maintained nearly the full activity (Figure 4.1.B, Table 4.2). Moreover, the V148A/H134Y enzyme also retained its full activity at 40 °C. The H134Y mutant behaved similar to the wildtype enzyme. The loss of enzymatic activity of V148A at 40 °C suggests that thermal inactivation at non-permissive temperature largely contributes towards temperature sensitivity of this mutant in virus-infected cells. Moreover, addition of the H134Y compensatory mutation renders more stability resulting in slower thermal inactivation of V148A/H134Y revertant. These results establish a clear correlation between the *in vitro* thermal inactivation behavior of the purified enzymes and the temperature-sensitive and reversion phenotypes observed in virus-infected cells.

4.4.3 Loss in secondary structure accompanies thermal inactivation of V148A

The loss in enzymatic activity can be achieved via local structural changes in the active site and does not mandate a total loss of protein structure. To gain insights into the biophysical mechanism for thermal inactivation behavior of the wildtype and mutant enzymes, the change in protein secondary structure at 30 °C and 40 °C was examined. CD signals were measured between 205–250 nm at regular time intervals

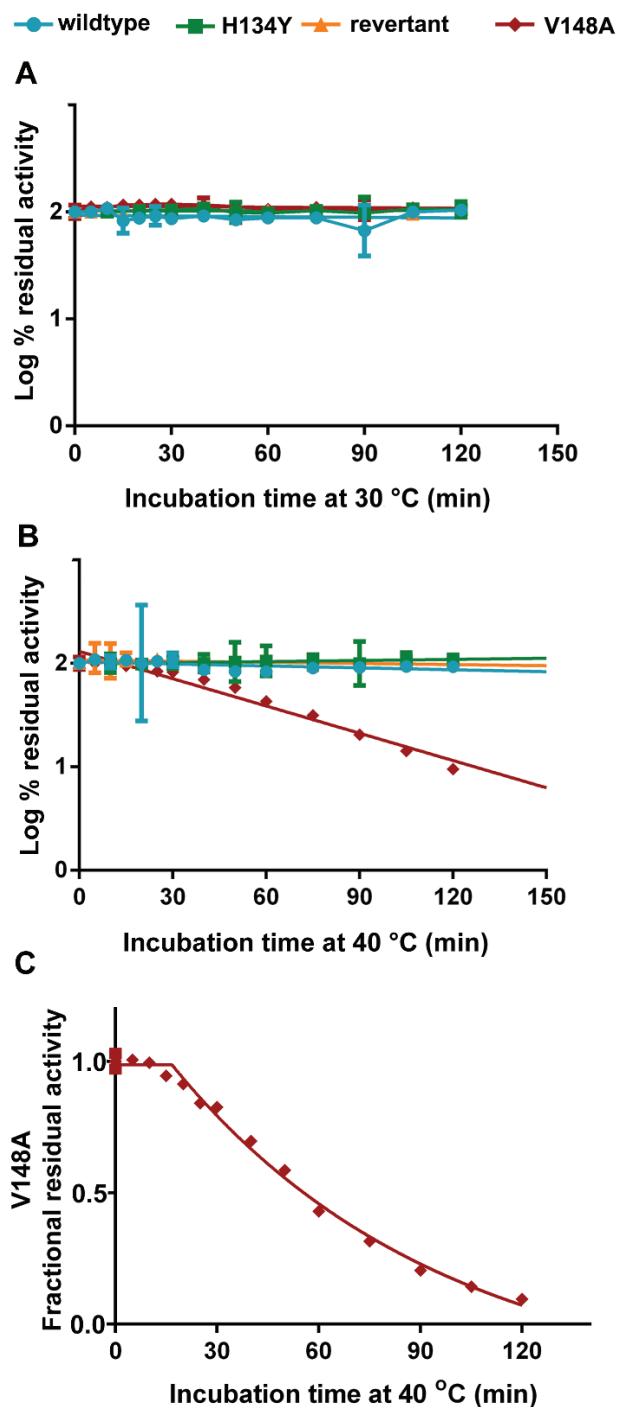


Fig. 4.1.: Thermal inactivation of MHV 3CL^{PRO} wildtype and the mutants. Log % residual activity versus time of incubation plots at **A.** 30 °C, and **B.** 40 °C. **C.** Plot of fractional residual activity versus incubation time at 40 °C was utilized to calculate the rate constant for decay of V148A by fitting the data into Equation 4.1.

Table 4.2.
Thermal inactivation of MHV 3CL^{pro} wildtype and mutants upon
incubation at 30 °C and 40 °C for 120 minutes

Enzyme	*Viral phenotype _{40°C}	% Residual activity _{30°C}	% Residual activity _{40°C}	$k_{40°C}$ ($\times 10^3$) (min ⁻¹)	Half-life _{40°C} (min)
Wildtype	Wildtype	112 ± 4	114 ± 1	ND	≫ 120
H134Y	Wildtype	106 ± 4	111 ± 4	ND	≫ 120
V148A/H134Y	revertant	100 ± 2	99 ± 3	ND	≫ 120
V148A	- <i>ts</i>	107 ± 1	10 ± 1	12.6 ± 1.4	55 ± 6

over a period of two hours at a given temperature. The CD signal at 222 nm was then utilized to calculate the percent change in secondary structure (Table 4.3).

Over a period of two hours, wildtype and H134Y enzymes maintained > 90% of their secondary structure at 30 °C and 40 °C. Both V148A and V148A/H134Y mutant enzymes were stable at 30 °C for two hours (Figures 4.2.A and 4.2.C, respectively). However, we observed a rapid loss of the secondary structure for V148A at 40 °C (Figure 4.2.B). As determined from the CD signal at 222 nm, V148A lost more than 75% of the secondary structure over a period of two hours of incubation at 40 °C. In contrast, the V148A/H134Y revertant maintained approximately 80% of its secondary structure at 40 °C (Figure 4.2.D). These results clearly demonstrate that the structural stability of V148A mutant is lower than the V148A/H134Y mutant at the non-permissive temperature of 40 °C. These biophysical data corroborate directly with the thermal inactivation data suggesting that structural unfolding accompanies thermal inactivation of V148A, while addition of H134Y mutation rescues the activity by providing additional structural stability to the V148A/H134Y revertant mutant.

Table 4.3.
Loss of secondary structure and T_m determination through measurement of CD signal at 222 nm

Enzyme	*Viral phenotype _{40°C}	% Residual CD signal _{40°C}	T_m °C
Wildtype	Wildtype	> 90	52.07 ± 0.17
H134Y	Wildtype	> 90	54.77 ± 0.52
V148A/H134Y	revertant	~ 80	49.69 ± 0.46
V148A	- <i>ts</i>	< 25	47.33 ± 0.24

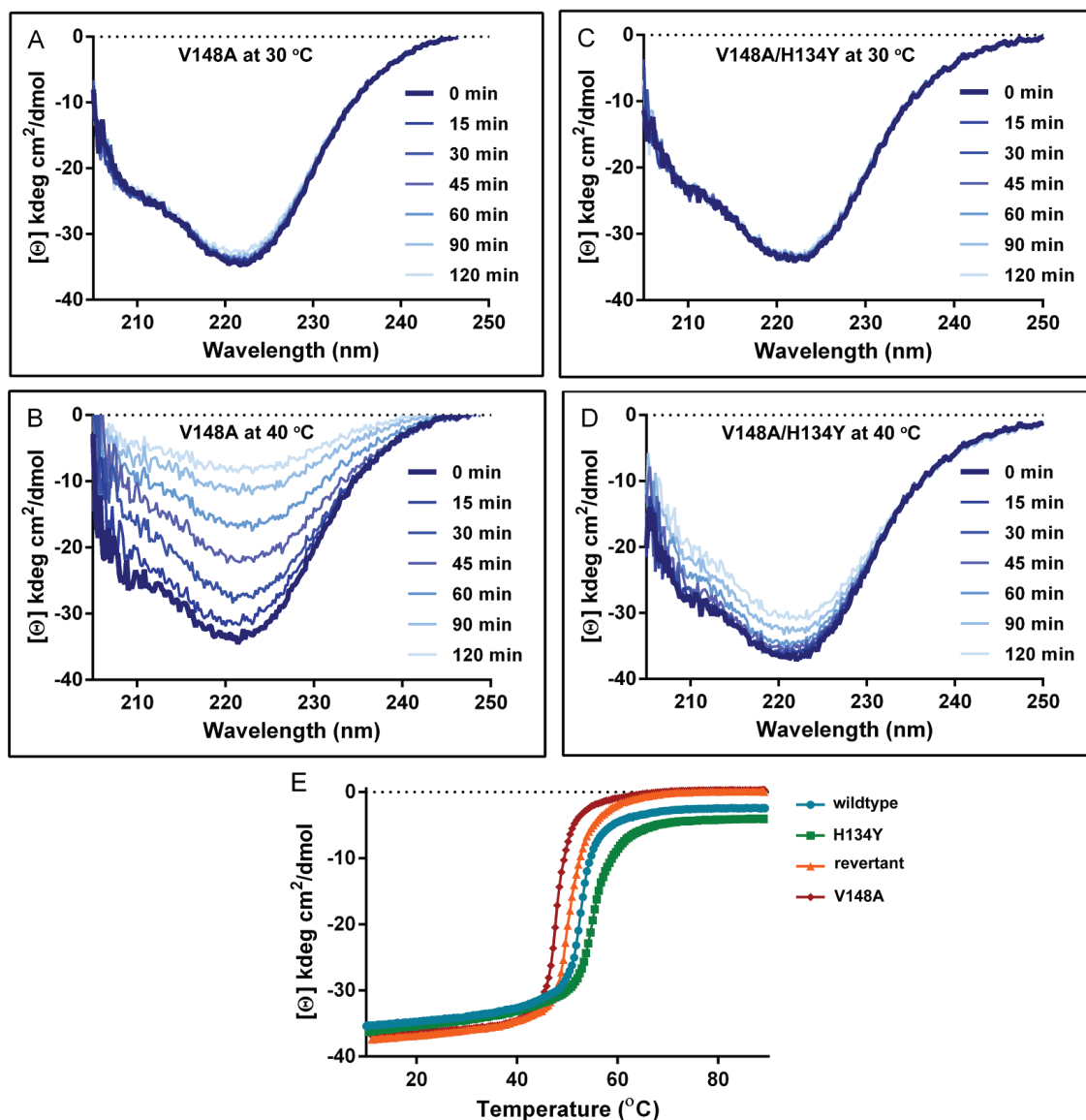


Fig. 4.2.: **Protein unfolding characterized through circular dichroism.** **A. - D.** Changes in protein secondary structure as determined through the CD signal (205-250 nm) during two hours incubation of V148A and revertant V148A/H134Y at 30 °C and 40 °C is shown. Both enzymes maintain their secondary structures at 30 °C for two hours. More than 75% of the secondary structure (signal at 222 nm) is lost for the V148A mutant over a period of two hours incubation at 40 ; V148A/H134Y revertant maintains approximately 80% of its secondary structure at this temperature. **E.** Melting curves, represented as CD signal at 222 nm as the temperature was raised from 10 to 90 , are shown for the wildtype (blue), H134Y (green), revertant V148A/H134Y (orange) and V148A (red). Melting temperatures, T_m in °C, for the wildtype, H134Y, revertant and V148A are 52.07 ± 0.17 , 54.77 ± 0.52 , 49.69 ± 0.46 and 47.33 ± 0.24 , respectively.

4.4.4 The difference in the melting temperatures suggests global destabilization of V148A

Melting temperatures (T_m) were determined as the inflection point of the melting curves for the wildtype and mutant enzymes. The CD melting curve for each enzyme is shown in Figure 4.2.E and the T_m values are provided in Table 4.3. The T_m value for the V184A *-ts* mutant is 5 °C lower than the wildtype enzyme whereas the V148A/H134Y mutant is only 2.5 °C lower, supporting our previous observations that the V148A mutant is thermally less stable at the non-permissive temperature of 40 °C. Interestingly, the single H134Y mutant that leads to compensation of the *-ts* phenotype of V184A is about 2.5 °C more stable than the wildtype enzyme. This demonstrates that the loss in activity observed for *-ts* V148A at non-permissive temperature is due to global destabilization of protein structure that causes a time-dependent loss of enzymatic activity at elevated temperatures. Moreover, the V148/H134Y mutant is better able to maintain its structure and activity at the non-permissive temperature due to the additional stability provided by H134Y mutation. Taken together, the biochemical and biophysical data corroborate directly with the viral data published previously [90,91].

4.4.5 Distant residues V148 and H134 are connected through hydrogen bonding and hydrophobic networks

To determine the structural mechanism by which mutation of residues at positions 134 and 148 are influencing protein stability, the structural model of MHV 3CL^{pro} wildtype was analyzed. A structural model of MHV 3CL^{pro} was built using the X-ray crystal structure of HKU1 3CL^{pro} as a template in homology modeling. In the resulting structural model, residues V148 and H134 are present at 15.6 Å (C-alpha distance) apart from each other. It is intriguing how a residue at position 134, which is distant from position 148, compensates for the structural instability emerging from a mutation at position 148.

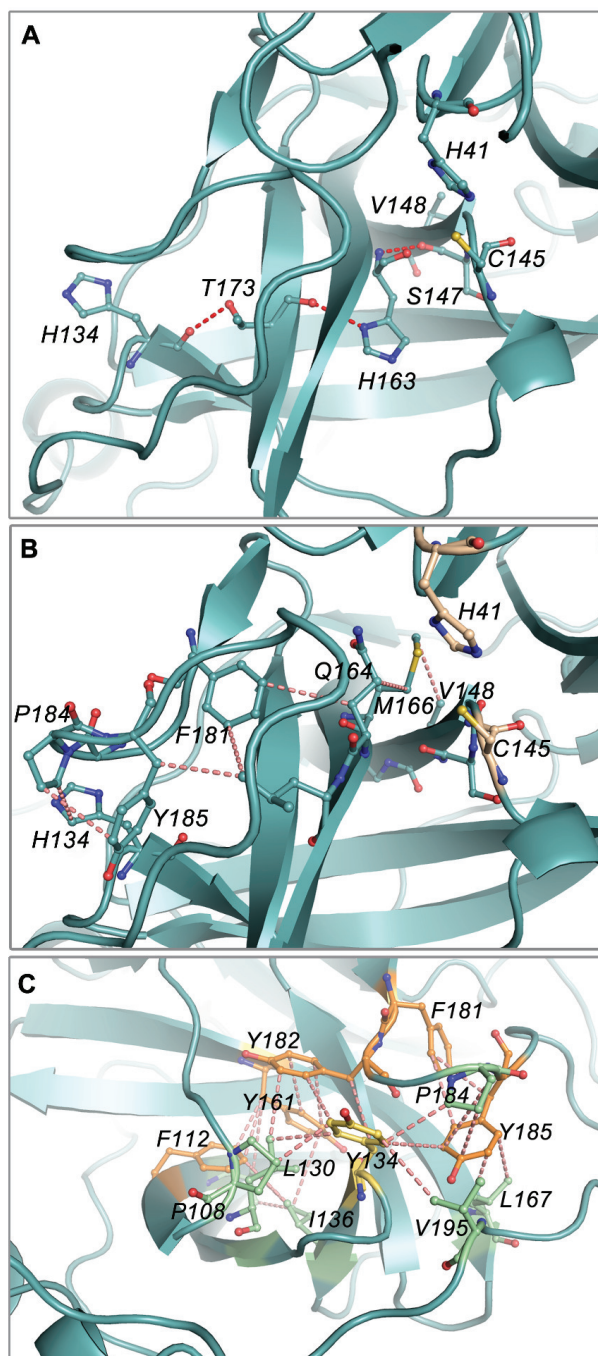


Fig. 4.3.: **Network of residues connecting positions 148 and 134 in MHV 3CL^{pro}.** **A.** Network of residues (T173, H163 and S147) connecting amino acids at positions 134 and 148 through hydrogen bonds (red dashed lines) are shown. **B.** Network of hydrophobic residues connecting amino acids at positions 134 and 148. Hydrophobic interactions are shown as pink dashed lines. **C.** Y134 (yellow sticks) mutation is expected to extend the underlying hydrophobic network through a cluster of aromatic residues (orange sticks) and other hydrophobic residues (pale green sticks).

Long-range interactions have been shown to manifest through a network of non-covalent interactions including hydrogen bonds, salt-bridges and hydrophobic interactions [100,101]. To probe the network of non-covalent interactions connecting residues H134 and V148, the shortest path between these two residues comprising of either hydrophobic contacts, or hydrogen bonds was examined. An inspection of the structural model revealed that only three hydrogen-bonding interactions are required to connect H134 to V148 (Figure 4.3.A). The backbone carbonyl oxygen of H134 forms a hydrogen bond with the side-chain hydroxyl group of T173. T173 interacts with H163 through a hydrogen bond between its backbone carbonyl oxygen and the imidazole NH of H163. H163 further engages amino acid S147, adjacent to V148, through a hydrogen bonding interaction between backbone NH of H163 and backbone carbonyl oxygen of S147. In addition to the hydrogen-bonding network between H134 and V148, the network of hydrophobic residues (within 5 Å of each other) that bridges these two residues was also examined.

Based on the structural analysis, five residues that are involved in connecting H134 and V148 through hydrophobic interactions were identified (Figure 4.3.B). These residues are P184, Y185, F181, Q164 and M166. H134 is present within hydrophobic contact distance from P184, which also forms hydrophobic contacts with the aromatic ring of Y185. The β -carbon of Y185 interacts with the aromatic ring of F181 through hydrophobic contacts, which in turn forms hydrophobic contacts with the aliphatic β -carbon of Q164. The γ -carbon of Q164 lies within hydrophobic contact distance of aliphatic γ -carbon of M162. Residue M162 is the closest amino acid to V148 in this hydrophobic network, where its γ -carbon forms hydrophobic contacts with the side chain of V148.

It is likely that the mutation of V148 to alanine disrupts the aforementioned hydrogen bonding or hydrophobic network, while mutation of H134 to tyrosine either preserves this network or form additional hydrogen bonding or hydrophobic networks to compensate for the loss of these interactions in the V148A mutant.

4.4.6 Significance of mutation of surface residue histidine to more hydrophobic tyrosine at position 134

In contrast to the popular belief that introduction of hydrophobic residues at the protein surface is thermodynamically unfavorable, several studies have demonstrated the significance of a surface hydrophobic residue in enhancing the protein structural stability [102–104]. These surface hydrophobic residues, which are usually present in cavities, enhance protein structural stability either by extending the hydrophobic network, especially aromatic-aromatic interactions, of underlying residues or by creating a hydrophobic barrier that protects the underlying buried residues and secondary structural elements from solvent molecules. The structural model of the H134Y mutant was therefore analyzed to examine the formation of extended hydrophobic network. Interestingly, it was observed that Y134 extends the underlying cluster of aromatic residues from the buried core of domain 2 to the surface (Figure 4.3.C). Moreover, the residue at position 134 (histidine or tyrosine) lies on one side of a small cavity formed largely by hydrophobic amino acids. We speculate that mutation at position 134 from histidine to more hydrophobic tyrosine assists in creating a protective hydrophobic shield from waters to form unfavorable interactions with underlying hydrophobic residues in the cavity.

4.4.7 Constraint Network Analysis predicts residue 148 as a weak spot

Computational tools were utilized to gain more insights into the molecular mechanism by which the V148A and H134Y mutations influence the activity and stability of MHV 3CL^{pro}. A constraint network analysis was performed on each of the trajectories using the CNA webserver developed by Krüger et al. [97]. This webserver was used to locate and analyze residues that may be prone to unfolding, also called ‘weak spots’ in each structural ensemble.

The top predicted weak spots from all simulations are shown in Table 4.4 and graphically in Figure 4.4.A (the wildtype 30 °C ensemble was selected as a represen-

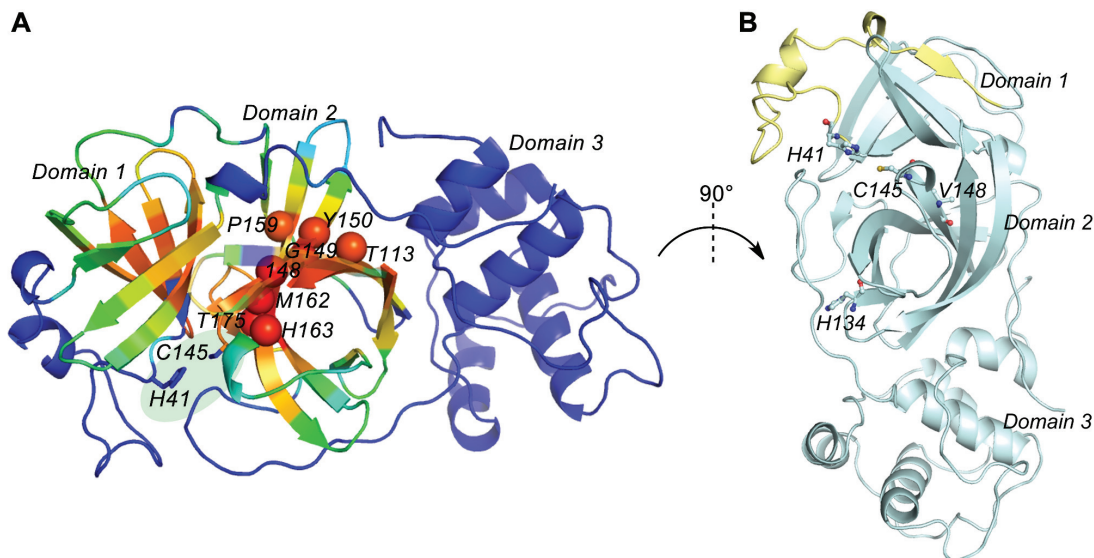


Fig. 4.4.: **Structural weak spots in MHV 3CL^{Pro}.** **A.** Location of the top predicted unfolding nuclei for all simulations are shown in spheres. The active site residues (H41 and C145) have been shown in sticks and the active site is highlighted in green. The structure is colored by the unfolding score (colored high to low score in red to blue) predicted in the wildtype 30 °C ensemble by the CNA webserver. In all structures, domain 3 was found to be very stable while domain 2 was found to contain most, if not all of the identified weak spots. **B.** Amino acids (40-70) demonstrating different behaviors between the 30 and 40 °C simulations in V148A mutant are shown in yellow; at 40 °C these residues are more mobile than at 30 °C based on RMSF and solvent accessible surface area (SASA) calculations.

Table 4.4.
Unfolding nuclei predicted by CNA webserver

Position	Residue	Frequency	Score	Average rank
148	V/A	7	63.63	1
149	G	7	59.19	4
150	Y	7	57	7
162	M	6	58.61	5
113	T	4	56.69	8
159	F	4	55.19	8
163	H	4	59.73	5
175	T	4	58.48	3

The residue number and name are given along with the number of times the residue was predicted in the top-10 positions across all simulations (maximum of 8 was possible). From those residues found in top-10 for each simulation the average score and average rank was calculated.

tative, but all were relatively similar). In seven of the eight total ensembles residue 148 was predicted to be in the first or second most likely point where unfolding may occur. Nearby residues, including 149 and 150, were also predicted as weak spots in the protein and together, these three residues form a large portion of a β -sheet immediately following the loop containing the active site cysteine (C145) (Figure 4.4.A). Interestingly, all frequently predicted weak spots occurred in the β -sheets of domain 2 (Figure 4.4.A).

Notably, residue 148 was not predicted as a weak spot in the 40 °C simulation of the wildtype protein. This was the only simulation in which this residue was not in the top-1 or top-2 position, in fact it was predicted in the 16th position with a score of 48.5. However, compared to the other simulations, the 40 °C wildtype simulation had a much tighter data distribution. For instance, the top-1 ranked residue (Thr175) in the wildtype scored a 53.69, just 4 units larger than the tenth ranked residue (Leu87). In comparison to the other simulations, the scores and the distribution for the 40 °C wildtype were lower and narrower on average. This may suggest that the tested mutations exasperate the weak spot at residue 148, but that there may be other vulnerable residues in the wildtype structure.

4.4.8 Distinct residue fluctuations for amino acids 45-70 in the mutants and the wildtype

In addition to assessing the weak spots in the protein, the root mean square fluctuation (RMSF) of each residue was also calculated over the length of the 50 ns trajectory (Figure 4.5). In general for each simulation, the 30 °C and 40 °C simulations are relatively similar in RMSF, with the exception of residues 45-70 which form one side of the active site cleft. In the V148A mutant (Figure 4.5.C), there is a notable difference between the 30 °C and 40 °C simulations in this region; at 40 °C these residues appear to be more mobile than at 30 °C. This is in contrast to the wildtype (Figure 4.5.A) and H134Y mutant (Figure 4.5.D) wherein the fluctuations in this

section of the protein are approximately equal or even slightly less flexible in the 40 °C simulations versus the 30 °C simulations. The RMSF profile of these residues in the revertant (Figure 5B) falls in between the wildtype and the V148A mutant, where from residues 45-60 it resembles more closely the V148A mutant, but from 60-70 more closely parallels the wildtype (Figure 4.5.E).

While the RMSF reveals differences between the wildtype and the V148A mutant on one side of the active site cleft, the active site residues themselves do not appear to be highly influenced. Figure 4.6 shows a running average of the distance between the centers of mass of the active site cysteine and histidine at 30 °C (Figure 4.6.A) and 40 °C (Figure 4.6.B). There is not a noticeable difference between any of the simulations or across the two temperatures. In all cases, the average distance between the residues is approximately 1.28 nm.

4.4.9 Mutual Information Analysis

This finding suggests that the structural implications of the mutation do not directly influence the active site, but likely influence surrounding residues in an allosteric manner. To further investigate this possibility, mutual information matrices were calculated using the MutInf program [98]. While there were several minor differences observed between the V148A mutant and the wildtype, many of these changes were not reflected in the revertant. In other words, it would be expected that similar patterns would be observed in both the revertant and the wildtype at 40 °C and that these patterns would be altered in the V148A mutant. However, such consistent patterns were minor, if apparent at all.

4.4.10 Solvent Accessible Surface Area

Based on the changes observed in the RMSF in the cleft region, this region along with the mutation sites and active sites, were further assessed by calculating the solvent accessible surface area (SASA).

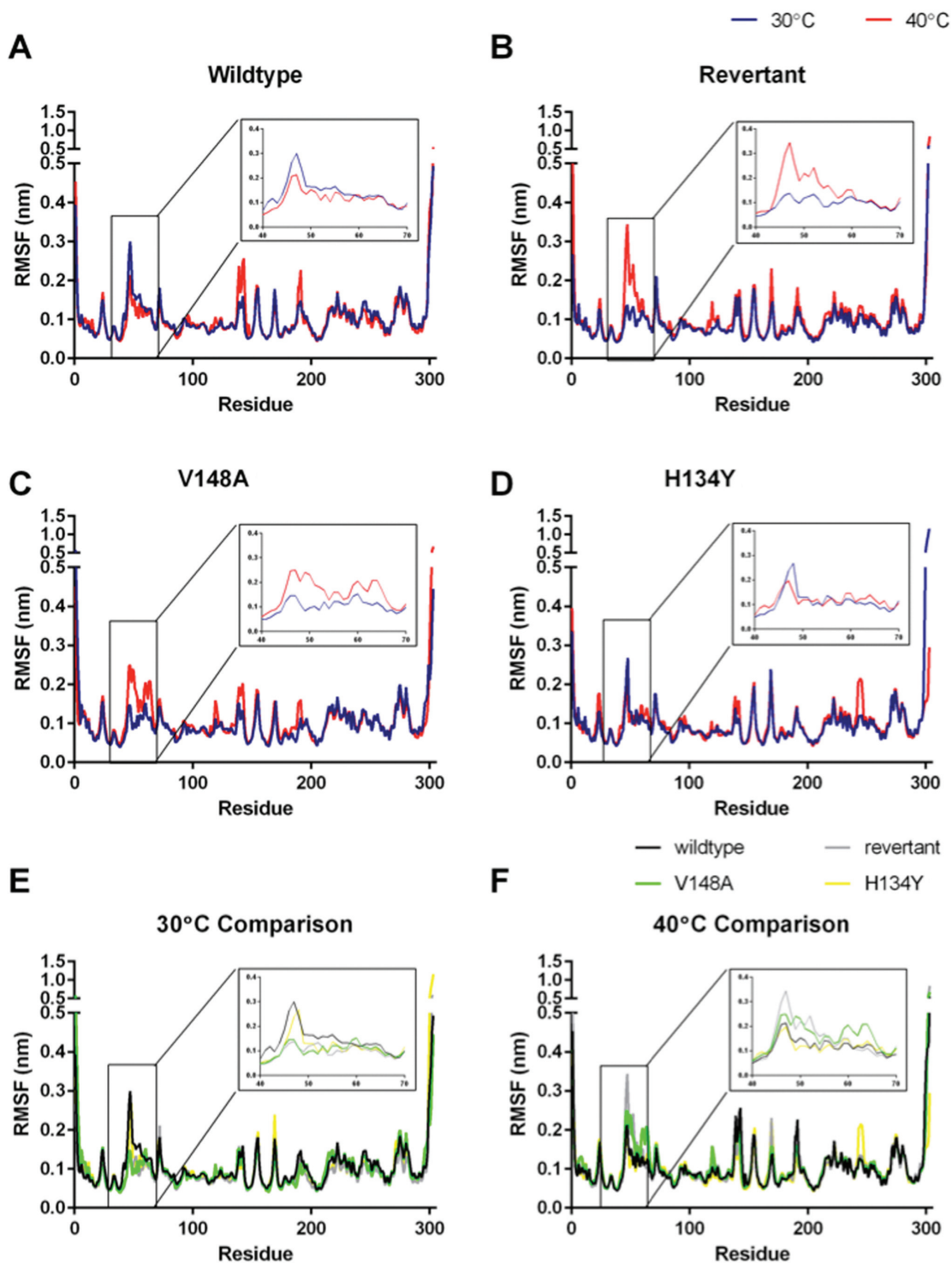


Fig. 4.5.: RMSF fluctuations in the alpha carbons of MHV 3CL^{Pro}. **A.** wildtype, **B.** V148A/H134Y revertant, **C.** V148A, **D.** H134Y. An overlay of the 30 °C and 40 °C simulations of the wildtype, H134Y, V148A mutant and V148A/H134Y revertant are shown in panels **E.** and **F.**, respectively.

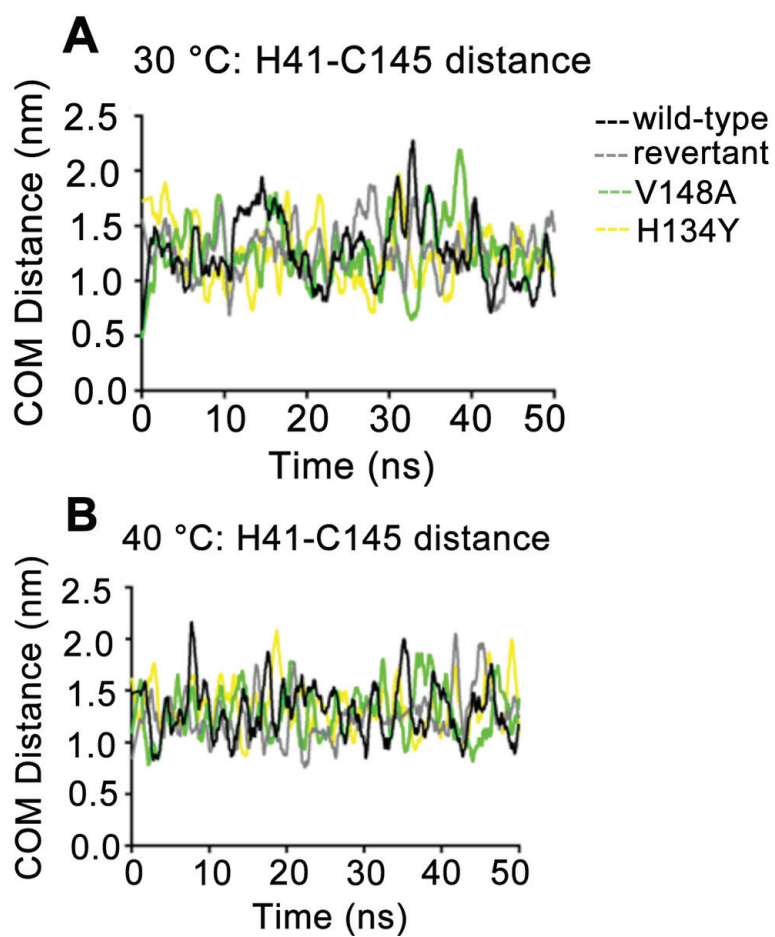


Fig. 4.6.: Distance between the centers of mass of the catalytic dyad (H41 and C145). **A.** at 30 °C, and **B.** 40 °C. For clarity, the distances have been plotted as a running average with a step size of 1 ns (100 points).

Not surprisingly the SASA of the entire protein was nearly identical across all simulations (Table 4.5). This suggests that the mutations and/or temperature changes do not drastically alter the stability or overall fold of the protein. In addition, the SASA of both mutation sites was relatively constant across all simulations. Position 148 is less exposed than position 134, but overall the SASA of these residues is relatively constant, suggesting that the source of the observed activity change is distant from the mutation sites.

Table 4.5.
SASA of various residues and groups of residues

	MHV 3CL ^{pro}	All residues	Position 148	Position 134	Catalytic dyad	Active site cleft ^a	Waters around catalytic dyad ^b
30 °C simulations	Wildtype	145.8±2.4	0.00±0.01	0.57±0.09	0.4±0.19	17.0±0.77	16±4
	Revertant	147.6±2.2	0.01±0.02	0.79±0.1	0.22±0.10	16.0±0.63	12±2
	V148A	147.2±2.1	0.01±0.02	0.58±0.1	0.25±0.1	16.2±0.70	11±2
	H134Y	147.6±2.4	0.00±0.01	0.8±0.11	0.3±0.12	16.9±0.58	13±2
40 °C simulations	Wildtype	145.3±2.4	0.0±0.1	0.66±0.11	0.5±0.20	16.2±0.67	17±4
	Revertant	147.3±2.7	0.02±0.03	0.8±0.12	0.4±0.19	17.0±0.12	14±4
	V148A	146.2±2.7	0.0±0.1	0.63±0.10	0.5±0.22	17.681±0.91	16±4
	H134Y	145.9±2.5	0.0±0.1	0.8±0.11	0.3±0.13	16.6±0.684	10±3

3

^a refers to amino acids 40-70

^b waters within 5 Å of catalytic dyad

While SASA of the mutation sites, catalytic dyad, and the overall protein were consistent across the various forms of the protein, the active site cleft showed subtle variations in the 40 °C V148A mutant. Compared to the V148A mutant simulation at 30 °C, the 40 °C simulation showed a slight increase in SASA for part of the active site cleft (residues 40-70, highlighted in yellow in Figure 4.4B). Compared to the wildtype, revertant and H134Y mutant simulations, the SASA of active site cleft in

the 40 °C V148A mutant is not only larger, but also has a larger standard deviation, suggesting that this region is more variable in the V148A mutant at 40 °C than in the other simulations (Table 4.5). This finding was in agreement with the RMSF calculations reported above.

Although the SASA of the catalytic dyad remained relatively unchanged, we evaluated whether the surrounding topology was affected by the mutation. To do this, the number of waters present within 5.0 Å of either active site residue over the length of the trajectory was calculated (Table 4.5). If the active site cleft is destabilized, it is possible that this could alter the number of water molecules able to penetrate into the active site cleft and therefore alter activity without altering the SASA of the catalytic dyad itself. In general, the number of waters near the active site increased with increasing temperature. However, the most drastic change between the 30 °C and 40 °C simulations was observed for the V148A mutant, where the number of waters increased from an average of 11 to an average of 16. While the average number of waters does not differ greatly in comparison to the other simulations, the large change observed between the two temperatures may indicate a greater destabilization of the V148A mutant protein at this temperature as compared to the other variants tested.

4.5 Discussion

The coronavirus 3CL^{pro} is an essential enzyme that processes the replicase polyproteins 1a and 1ab at eleven distinct cleavage sites. Structures of 3CL^{pro} from several coronaviruses have been determined. Several groups have reported the *in vitro* characterization of structural and functional determinants of 3CL^{pro} from SARS-CoV using purified protein [43, 45–47, 89]. Amino acids that directly influence substrate binding, catalysis and dimer formation due to their physical proximity to the substrate binding pocket or the dimer interface have been identified through kinetic, biophysical and structural studies [43, 45–47, 49, 60]. Interestingly, a few groups have also identified amino acids that influence dimerization or catalysis through long-range

interactions [60,87–89]. Surprisingly though, the ramifications of mutating the amino acids that influence catalysis or dimerization either through direct or long-range interactions have rarely been investigated in the context of coronavirus replication. Our collaborators previously identified mutations (V148A, S133A and F219L) in MHV 3CL^{pro} that impart temperature sensitive phenotype to the recombinant viruses carrying these mutations [90,91]. They also reported the emergence of a common second site compensatory mutation (H134Y) that partially or completely suppresses the temperature sensitive phenotype and was implicated to communicate with the temperature sensitive mutations through long-range interactions. This study highlights the first, to our knowledge, *in vitro* characterization of physiologically relevant mutations of 3CL^{pro}. Our biochemical, biophysical and molecular-dynamics based structural analysis provide the mechanistic basis for temperature sensitive and reversion phenotypes of V148A and H134Y mutations, respectively in 3CL^{pro} from MHV.

4.5.1 Position 148 is a structural ‘weak spot’

The *-ts* phenotype of V148A mutation may ensue via perturbation of a variety of pathways in the virus-infected cells. The *-ts* V148A mutation can directly impact protease activity of 3CL^{pro}, influence the kinetic and thermodynamic stability of protein structure, impair polyprotein processing at the rate-limiting cleavage step or alter interaction of 3CL^{pro} with a viral/host binding partner. Other groups have reported alteration of 3CL^{pro}-mediated polyprotein processing upon introduction of mutations in nsp3 and nsp10, suggesting interaction of 3CL^{pro} with other viral proteins in the replicase complex. Differences in the cleavage site specificity, rather than the protease activity, have also been reported for certain mutations of nsp4 protease from equine arteritis virus. Based on the processing of nsp5 (3CL^{pro}) and nsp8 in the *-ts* V148A virus, Stobart et al. suggested that the temperature sensitive phenotype may have resulted from altered processing at different polyprotein cleavage sites. It was hypothesized that the *-ts* phenotype of V148A mutant arises from altered structural stability

of the mutant protein at the non-permissive temperature of 40 °C. The biochemical examination of V148A mutant enzyme reveals that V148A mutation results in global destabilization of the protein structure which then rapidly falls apart at elevated temperature. Constraint network analysis (CNA) based structural investigation reveals that the residue 148 is a structural weak spot and the major unfolding nuclei are clustered in the β -sheets of domain 2 near the V148A substitution. This area forms the core of the second domain and is located just past the active site cysteine (C145). It was interesting that majority of the weak spots are identified in domain 2; inherent flexibility is usually expected to exist in a catalytic domain for optimal enzymatic activity in general and for interaction with a variety of substrates, eleven polyprotein cleavage sites, specifically for 3CL^{pro}.

4.5.2 Y134 acts as a general facilitator for 3CL^{pro} stabilization

Viral evolution necessitates viral proteins to accrue mutations that provide selective advantage for adaptation in a new host or under stressed environmental conditions like elevated temperatures or introduction of drugs. For example, development of resistance against a drug often involves introduction of a mutation in the vicinity of the binding site. Likewise, emergence of compensatory mutations in physical proximity to the cognate temperature sensitive mutation is common occurrence. Emergence of distant second-site H134Y compensatory mutation that suppresses the temperature sensitive phenotype of V148A mutant virus was interesting. Our biochemical analysis demonstrates that H134Y mutation indeed provides extra stability to V148A MHV 3CL^{pro} mutant. To identify the structural basis for improved stability of the H134Y mutant, the structural model of this mutant was analyzed. Numerous studies have described the stabilizing role of a surface hydrophobic residue [102–104]. Stabilizing effect of a surface hydrophobic residue is usually derived from the expansion of underlying hydrophobic network and the formation of a shield that excludes water molecules from destabilizing the surface layer secondary structures. Analysis of the

structural models of wildtype and the V148A/H134Y revertant suggests that mutation to an aromatic tyrosine residue at position 134 stabilizes the protein structure in a similar manner. Moreover, specific mechanism of reversion phenotype is not recognized through MD simulation analysis of V148A/H134Y mutant. These data suggest that residue Y134 acts as a general facilitator for structural stability of MHV 3CL^{pro}.

The role of Y134 as a gatekeeper residue for structural stability is further exemplified from the emergence of H134Y compensatory mutation to suppress the temperature sensitive phenotypes of two other independent *-ts* mutations, S133A and F219L, in MHV 3CL^{pro} as previously reported. Furthermore, 3CL^{pro} enzymes from other closely related β -coronaviruses in genogroup 2a, HKU1 and OC43, contain a tyrosine at position 134. Interestingly, introduction of V148A mutation in chimeric 3CL^{pro} from HKU1 and OC43 engineered in MHV background does not result in *-ts* phenotype [105]. Moreover, V148A/Y134H chimeric OC43 3CL^{pro} exhibits *-ts* phenotype [105]. All these results suggest that site 134 acts as an evolutionary sentry for structural stability of MHV 3CL^{pro} that can be readily mutated to tyrosine to provide extra stability when required.

While one possible interpretation of the data is provided here, taken together, the findings indicate that the mechanisms behind the temperature sensitive nature of this enzyme are subtle and do not involve a readily observable or drastic conformational change which is not surprising. In fact, it has been suggested that in some cases changes in the entropic/enthalpic landscape rather than backbone conformational changes may be responsible for the allostery observed in some systems [106]. Dissecting and analyzing the specific structural implications of temperature is an evolving field and one we are only beginning to understand [107].

4.5.3 Implications for the prediction of viral *-ts* and corresponding suppressor mutations in the design of viral vaccines

Temperature sensitive mutations not only serve as powerful tools for studying the function of specific genes, they have long been utilized for the development of attenuated viral vaccines. Understanding the structural implications of temperature on these mutations is, therefore, of considerable significance. Several groups have reported the genetic instability of temperature sensitive mutations in an attenuated virus [108]. Reversion of the temperature-sensitive mutation to the wildtype as well as the emergence of second-site compensatory mutations are common occurrences and may have detrimental effects on the efficacy of attenuated virus as vaccine [108]. However, studies investigating the effect of these specific mutations on the structural and biochemical properties of the viral proteins harboring them, esp. the compensatory mutations, remain largely unexplored.

The Mesecar group, in collaboration with Prof. Susan Baker's group (Loyola University), recently reported the generation of an attenuated murine hepatitis virus via incorporation of a temperature-sensitive mutation in the Ubl2 domain of MHV papain-like protease PLP2 [109]. The attenuated mutant virus was capable of eliciting immune response in mice. Interestingly, this mutation was able to influence the enzymatic activities and stability of purified MHV PLP2 despite its presence in an entirely distinct Ubl2 domain than the active site, suggestive of long-range interactions between the Ubl2 domain and the active site.

In the current study, the mechanistic basis of temperature-sensitivity of a previously identified mutation in MHV 3CL^{pro} was elucidated. We also provided the biochemical and structural basis for reversion of the temperature-sensitive phenotype by a distant second-site compensatory mutation. Existence of long-range interactions is not only limited to MHV 3CL^{pro} but has also been reported by several groups, including ours, for 3CL^{pro} from SARS-CoV [87–89,91] and MERS-CoV [60]. Therefore, it is important to understand the protein structural dynamics that allows interaction

between distant residues. While considering a specific *-ts* mutation for the development of attenuated viral vaccines, emergence of distant second site compensatory mutations specifically highlight the need for powerful computational tools to predict the location for second-site compensatory mutations in advance. From an evolutionary perspective, emergence of compensatory mutation at a completely conserved site is highly unlikely. The virus is more likely to mutate at a less conserved site to a residue that is present in other closely related viruses. The fact that position 134 is naturally occupied by a tyrosine in 3CL^{PRO} from other coronaviruses like HKU1, OC43, BtCoV HKU8 and PEDV suggests that MHV strategically selected for a compensatory residue that has already been sampled for a functional protease in other coronaviruses. Although current methods are limited to either predicting *-ts* mutations or thermo-stable mutations, with recent advancements in computational performance we expect to see the development of computational tools for predicting and simulating the emergence of second site compensatory mutations in the near future.

CHAPTER 5. LIGAND-INDUCED DIMERIZATION REGULATES MERS-COV 3CL^{PRO}

Parts of the data and text in this chapter have been published in the journal article entitled “Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro): implications for nsp5 regulation and the development of antivirals.” [60]

5.1 Abstract

All coronaviruses, including the recently emerged Middle East respiratory syndrome coronavirus (MERS-CoV) from the β -CoV subgroup, require the proteolytic activity of nsp5 protease (aka 3C-like protease, 3CL^{PRO}) during virus replication, making it a high value target for the development of anti-coronavirus therapeutics. Kinetic studies indicate that in contrast to 3CL^{PRO} from other β -CoV 2c members including HKU4 and HKU5, MERS-CoV 3CL^{PRO} is less efficient at processing a peptide substrate due to MERS-CoV 3CL^{PRO} being a weakly associated dimer. Conversely, HKU4, HKU5 and SARS-CoV 3CL^{PRO} enzymes are tightly associated dimers. AUC studies support that MERS-CoV 3CL^{PRO} is a weakly associated dimer ($K_d \sim 52 \mu\text{M}$) with a slow off-rate. Peptidomimetic inhibitors of MERS-CoV 3CL^{PRO} were synthesized and utilized in AUC experiments and demonstrate that MERS-CoV 3CL^{PRO} undergoes significant ligand-induced dimerization. Kinetic studies also revealed that designed reversible inhibitors act as activators at low compound concentration as a result of induced dimerization. Primary sequence comparisons and X-ray structural analyses of two MERS-CoV 3CL^{PRO}-inhibitor complexes, determined to 1.6 Å, reveal remarkable structural similarity of the dimer interface with 3CL^{PRO} from HKU4-CoV and HKU5-CoV. Despite this structural similarity, substantial differences in the dimerization

ability suggest that long-range interactions by the non-conserved amino acids distant from the dimer interface may control MERS-CoV 3CL^{pro} dimerization. Activation of MERS-CoV 3CL^{pro} through ligand-induced dimerization appears to be unique within the genogroup 2c and may potentially increase the complexity in the development of MERS-CoV 3CL^{pro} inhibitors as antiviral agents.

5.2 Introduction

Coronaviruses (CoVs) are enveloped, positive strand RNA viruses that infect a variety of vertebrates. Coronaviruses (CoVs) are enveloped, positive-strand RNA viruses that infect a variety of vertebrates, including bats, livestock, pets, poultry and humans [1, 25, 26]. Although human CoVs cause respiratory illnesses of mild to moderate severity [8–11, 110, 111], two recently emerged CoVs, SARS-CoV (severe acute respiratory syndrome coronavirus) and MERS-CoV (Middle East respiratory syndrome coronavirus), have demonstrated their potential to become a serious threat to public health. The MERS-CoV emerged late in 2012 and unlike its predecessor SARS-CoV, MERS-CoV continues to exhibit up to a 35 % case-fatality rate [15, 112, 113].

Based on the sequence analysis of seven genes of the replicase domain, MERS-CoV has been classified as a β -CoV genogroup 2c member, along with closely related bat coronaviruses HKU5 (*Pipistrellus* bat) and HKU4 (*Tylonycteris* bat) [114, 115]. Increasing evidence suggests that bats may serve as zoonotic reservoirs for MERS-CoV [30, 31]. Evidence presented by recent studies also supports the local zoonotic transmission of MERS-CoV from dromedary camels to humans [34, 35]. Alarmingly, human-to-human transmission during close contact, especially in elderly or patients with underlying health conditions, has also been reported for MERS-CoV [13, 17–19]. In the wake of the recent upsurge in the laboratory-confirmed cases of MERS-CoV, including two recently identified cases in the United States [116], there is an urgent

need to study and characterize the properties of important drug targets of MERS-CoV for the development of effective therapeutics.

Coronaviruses express a >800 kDa replicase polyprotein, which is processed by viral 3CL^{pro} protease (or nsp5) at eleven distinct cleavage sites to yield intermediate and mature non-structural proteins (nsp) responsible for many aspects of virus replication [1, 40, 41, 84]. Due to its indispensable role in virus lifecycle, 3CL^{pro} is an important target for therapeutic intervention against coronavirus infections [37–39, 49, 53, 85, 117].

A number of kinetic, biophysical and X-ray structural studies have demonstrated that SARS-CoV 3CL^{pro} is only active *in vitro* as a tightly associated dimer with a dimer dissociation constant (K_d) in the low nM range [43–47]. The addition or deletion of amino acids, e.g. (His)₆-affinity tags, at either the N- or C-terminus drastically reduces the enzymatic rate and decreases the ability of SARS-CoV 3CL^{pro} to dimerize [45]. Although the cellular evidence for the auto-cleavage mechanism (*cis* vs *trans*) of 3CL^{pro} is lacking, models for how 3CL^{pro} cleaves itself from the polyprotein to form the mature dimer have been proposed based on *in Vitro* studies using purified 3CL^{pro} [43, 57, 58]. A current model posits that two inactive 3CL^{pro} molecules within two separate polyproteins recognize each other and form an immature dimer capable of cleaving the nsp4|nsp5 and nsp5|nsp6 sites *in trans*, followed by formation of an active and mature dimer that can then rapidly process other cleavage sites and multiple polyproteins. It has also been proposed that substrate-induced dimerization regulates the enzymatic activity of SARS-CoV 3CL^{pro} during virus replication; however, no experimental evidence of this has ever been demonstrated in infected cells [58]. Although our knowledge of SARS-CoV 3CL^{pro} is extensive, the dimerization properties of 3CL^{pro} from MERS-CoV and other coronaviruses, as well as the factors regulating their enzymatic activity, remain largely unknown.

To understand the properties of MERS-CoV 3CL^{pro}, a series of kinetic, biophysical and X-ray structural studies were conducted. Herein, the detailed kinetic and biophysical analysis of MERS-CoV 3CL^{pro} activity and dimerization is reported. These kinetic and biophysical studies provide evidence for a weakly associated MERS-CoV

3CL^{pro} dimer. In addition, utilizing the previous knowledge on the design of potent SARS-CoV 3CL^{pro} peptidic inhibitors, a series of inhibitors of MERS-CoV 3CL^{pro} that exhibit low micromolar potency were designed. It is demonstrated that MERS-CoV 3CL^{pro} requires the binding of a ligand for dimer formation, indicating that ligand-induced dimerization is likely a key mechanism in the regulation of MERS-CoV 3CL^{pro} activity during virus infection.

5.3 Experimental Procedures

5.3.1 Construct design and expression of MERS-CoV 3CL^{pro}

The gene encoding 3CL^{pro} protease of MERS-CoV (amino acid residues 3248—3553 in the replicase polyprotein, GenBank: AHC74086.1) was codon optimized for optimal expression in *E. coli* (BioBasic Inc). The gene was subcloned into pET-11a expression vector with an N-terminal (His)₆-tag followed by nsp4|nsp5 auto-cleavage site using the forward primer 5'-ATATACATATGCACCACCACCACCACAGCGGTGTTCTGCAGTCTGGTC-3' and the reverse primer 5'-GACGGATCCTTACTGCATCAC AACACCCATGATCTGC-3'. Construct was verified by DNA sequencing at the Purdue University Genomics Core Facility. This construct results in the expression of MERS-CoV 3CL^{pro} without any N-terminal or C-terminal extension. MERS-CoV 3CL^{pro} was expressed through auto-induction in *E. coli* BL21-DE3 cells in the presence of 100 µg/mL of carbenicillin as described previously [118]. Cells were harvested by centrifugation at 5000 × g for 20 minutes at 4 °C, and the pellets were stored at −80 °C until further use.

5.3.2 MERS-CoV 3CL^{pro} purification

Frozen pellets from 4 L of bacterial cell culture were thawed on ice and re-suspended in 250 mL of Buffer A [20 mM Tris pH-7.5, 0.05 mM EDTA, 10% glycerol and 5 mM β-mercaptoethanol (BME)], containing 500 µg of lysozyme and a small

amount of DNase. Cells were then lysed using a single pass through French press at 1200 psi and cell debris was removed from the cleared lysate by centrifuging at $29,000 \times g$ for 30 minutes. Solid ammonium sulfate was added to the cleared lysate to a final concentration of 1 M through gradual mixing on ice.

Hydrophobic-interaction chromatography

The cleared lysate, mixed with ammonium sulfate, was loaded at a flow rate of 3 mL/min onto a 60 mL Phenyl Sepharose 6 fast-flow high-sub column (XK 26/20, Amersham Biosciences, Piscataway, NJ) equilibrated with Buffer B (50 mM Tris pH-7.5, 1 M ammonium sulfate, 0.05 mM EDTA, 10% glycerol and 5 mM BME). The column was then washed with $5 \times$ column volume (300 mL) of Buffer B at a flow rate of 4 mL/min. Protein was eluted using a $5 \times$ column volume (300 mL) linear gradient to 100% Buffer A. Fractions (12 ml) were collected and those containing MERS-CoV 3CL^{pro}, as judged through SDS-PAGE analysis (Figure 5.1.A) and specific activity measurements, were pooled (120 mL) and exchanged into 2 L of Buffer A via overnight dialysis in a dialysis tubing (10,000 MWCO SnakeSkin[®], Thermo Scientific).

DEAE anion-exchange chromatography

The dialyzed sample from the previous step was loaded at a flow rate of 3 mL/min onto a 120 mL DEAE anion-exchange column (XK 26/40, Amersham Biosciences) equilibrated with Buffer A. The column was then washed with 2 column volume (240 mL) of Buffer A at a flow rate of 4 mL/min. A linear gradient (total volume 480 ml) to 40% Buffer C (50 mM Tris pH-7.5, 1 M NaCl, 0.05 mM EDTA, 10% glycerol and 5 mM BME) was used to elute the protein (Figure 5.1.B). Fractions (6 ml) were collected and those containing MERS-CoV 3CL^{pro} were pooled (66 mL) and dialyzed for 4 hours in 4 L of Buffer D (20 mM MES pH-5.5, 0.05 mM EDTA, 10% glycerol and 5 mM BME).

Mono S cation-exchange chromatography

Following dialysis, the pH of the sample was manually adjusted to 5.5 using 1 M solution of MES pH-5.5, and any precipitated protein was removed by filtering through a 0.22 μm pore size Millex-GP filter (Millipore). The filtered sample was then loaded at a flow-rate of 2 mL/min onto a 8 mL Mono S 10/100 column (Amersham Biosciences) equilibrated in Buffer D. The column was then washed with $5 \times$ column volume (40 mL) of Buffer D at a flow rate of 2 mL/min. Protein was eluted using a $25 \times$ column volume (200 mL) and linear gradient to 50% Buffer E (50 mM MES pH-5.5, 1 M NaCl, 0.05 mM EDTA, 10% glycerol and 5 mM BME) (Figure 5.1.C). Fractions (2 mL) were collected and those containing MERS-CoV 3CL^{pro} were pooled (22 mL) and concentrated to approximately 5 mg/mL using Amicon[®] Ultra 15 mL Centrifugal Filters (Millipore).

Gel-filtration chromatography

As the final purification step, the concentrated protein sample was loaded onto the prep grade Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with Buffer F [25 mM HEPES pH-7.5, 10% glycerol, 2.5 mM dithiothreitol (DTT)]. Protein was eluted isocratically at a flow rate of 1 mL/min with buffer F (Figure 5.1.D). Fractions (2 ml) containing MERS-CoV 3CL^{pro} were pooled (total volume of 34 mL) and concentrated to approximately 5 mg/mL. For final storage of the purified MERS-CoV 3CL^{pro} enzyme, 300 μL protein aliquots were placed into 1 mL screw-cap vials, flash-frozen under liquid nitrogen and then stored at $-80\text{ }^{\circ}\text{C}$ until further use.

Purification of SARS-CoV 3CL^{pro} and HKU5-CoV 3CL^{pro} has been described in Chapter 2. HKU4-CoV 3CL^{pro} was purified using a protocol similar to the purification of HKU5-CoV 3CL^{pro} [119].

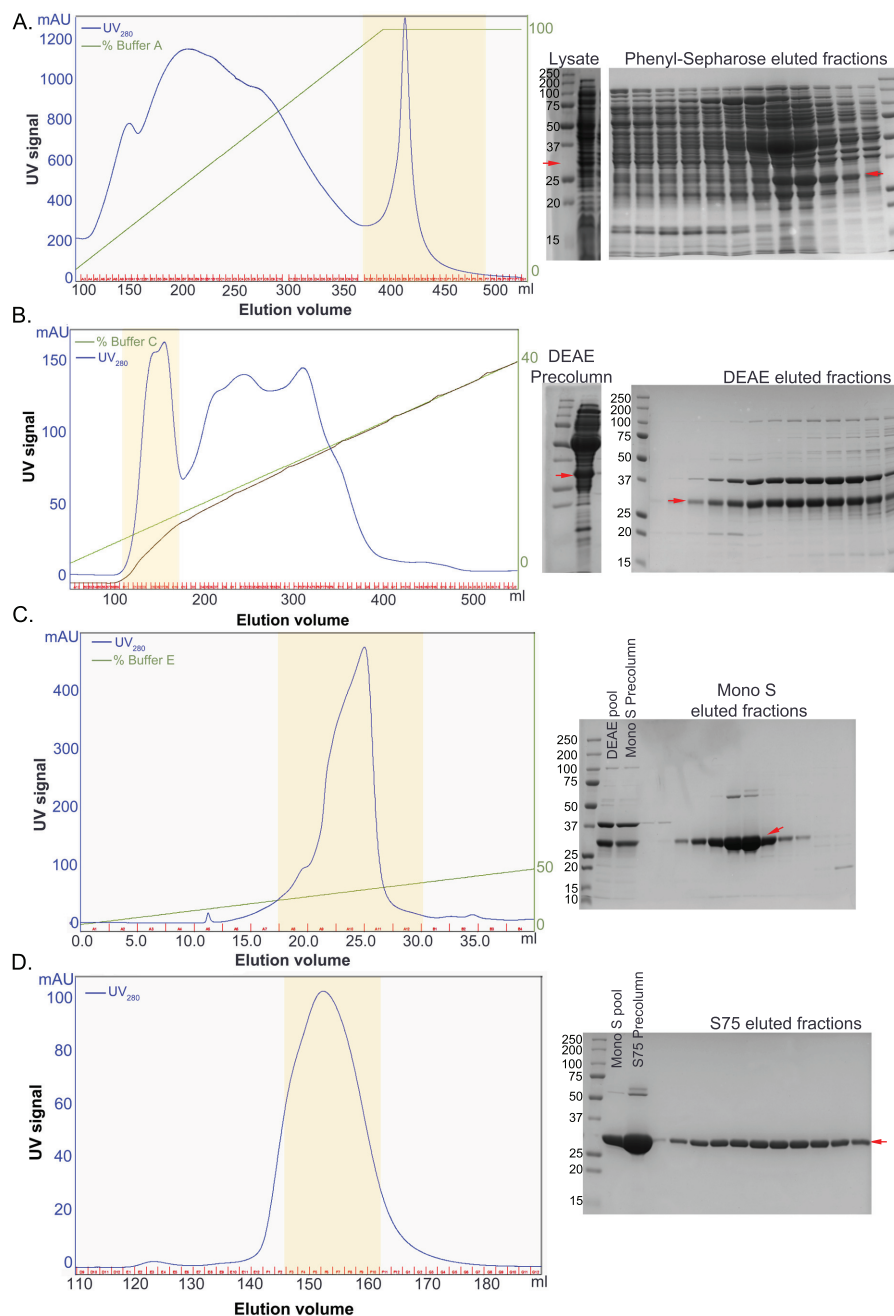


Fig. 5.1.: **Purification of MERS-CoV 3CL^{pro}**. **A.** *Left*-Elution profile from Phenyl-sepharose column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. MW marker sizes are indicated. Red arrow indicates the expected size of MERS-CoV 3CL^{pro}. **B.** *Left*-Elution profile from DEAE column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **C.** *Left*-Elution profile from Mono S column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile. **D.** *Left*-Elution profile from Superdex 75 column. *Right*-SDS-PAGE analysis of the eluted fractions highlighted in yellow in the elution profile.

5.3.3 Western-blot analysis

For western-blot analysis, the samples were prepared by mixing 2 μg each of the purified MERS-CoV 3CL^{pro}, untagged SARS-CoV 3CL^{pro} (negative control) and a His-tagged protein (positive control) with SDS-PAGE loading buffer (containing fresh DTT to a final concentration of 1 mM as the reducing agent instead of BME) and heated at 90 °C for 2 minutes. The samples were loaded in duplicate along with the prestained SDS-PAGE standards (Bio-Rad) in the wells of a 12.5% SDS-PAGE gel, both for Coomassie staining and western-blot analysis. The samples were run for 40 minutes at a constant voltage of 180 V. The gel was then cut into two; the first half was analyzed through regular Coomassie staining and the second half was used for western blot analysis. For western blot, the proteins were transferred from the gel onto a PVDF membrane at 4 °C in 1X transfer buffer (25 mM Tris-HCl, 192 mM glycine, 10% methanol and 0.01% sodium dodecyl sulfate). The membrane was then blocked for 2 hours at room temperature in 5% non-fat dry milk solution prepared in 1X TBST buffer (50 mM Tris base, 150 mM NaCl and 0.05% Tween 20). The membrane was then rinsed thrice in TBST buffer for 5 minutes each. After washing, the membrane was incubated with HRP conjugated anti-His antibody from GenScript (1:1000 dilution prepared in TBST buffer) for 1 hour at room temperature. The membrane was then rinsed again in TBST buffer for three times. Next, the membrane was incubated for 5 minutes with SuperSignal[®] West Dura Extended Duration substrate (Thermo Scientific) and finally imaged using Chemi with Markers setting of FluorChem E imaging system from ProteinSimple.

5.3.4 MALDI mass-spectral analysis

For molecular weight determination through MALDI-TOF analysis, purified MERS-CoV 3CL^{pro} was diluted to a concentration of 1 mg/mL in 25 mM HEPES pH-7.5 buffer. Matrix was prepared by dissolving Sinapinic acid in a 70% acetonitrile solution in deionized water with 0.1% TFA. The protein sample was mixed with the matrix at

a ratio of 1:2. 1 μL of the protein-matrix mix was spotted on the MALDI sample plate and allowed to dry completely before the plate was loaded into the mass-spectrometer. MALDI-TOF analysis was performed in linear mode on an Applied Biosystems 4700 Proteomics analyzer (Framingham, MA) available in the Purdue Proteomics Facility.

5.3.5 Synthesis of compounds **1-11**

The peptidomimetic compounds with Michael acceptor groups (Table 5.3) were synthesized via very similar methods to those published previously [53,54]. Synthesis of non-covalent peptidomimetic compounds **10** and **11** has been described previously [39].

5.3.6 Fluorescence-based kinetic assays

The enzymatic activity of 3CL^{pro} was measured using the following custom synthesized peptide: (HilyteFluorTM-488)-ESATLQSGLRKAK-(QXLTM-520)-NH₂ (AnaSpec, Inc.). The HilyteFluorTM-488 fluorescence group is internally quenched by QXLTM-520 dye. This substrate works as a generic peptide substrate for 3CL^{pro} enzymes and was designed based on the nsp4|nsp5 cleavage sequence for many coronavirus 3CL^{pro} enzymes. The rate of enzymatic activity was determined at 25 °C by following the increase in fluorescence ($\lambda_{\text{excitation}} = 485 \text{ nm}$, $\lambda_{\text{emission}} = 528 \text{ nm}$, bandwidths = 20 nm) of Hilyte Fluor-488 upon peptide hydrolysis by the enzyme as a function of time. Assays were conducted in black, half-area, 96-well plates (Corning) in assay buffer (50 mM HEPES pH-7.5, 0.1 mg/mL BSA, 0.01% Triton X-100 and 2 mM DTT) using a final reaction volume of 100 μL . The resulting fluorescence was monitored using a BioTek Synergy H1 plate reader. The rate of the reaction in arbitrary fluorescence units per sec (AFU/sec) was determined by measuring the initial slope of the progress curves, which were then converted to units of μM of product produced per min ($\mu\text{M}/\text{min}$) using experimentally determined values of ‘fluorescence extinction coefficient’ as described in Chapter 2. All reactions were carried out in triplicate.

5.3.7 Determination of enzymatic efficiency

The apparent enzymatic efficiency for each of the 3CL^{pro} enzymes was determined by measuring the rate of enzymatic activity as a function of varying substrate concentration in 100 μ L reactions. Reactions were initiated by the addition of enzyme to the wells of an assay plate containing varying concentrations of substrate. The final substrate concentrations varied over the range from 0 to 2 μ M. The final enzyme concentrations for each 3CL^{pro} studied were the following: MERS-CoV 3CL^{pro} at 1 μ M, SARS-CoV 3CL^{pro} at 100 nM, HKU5-CoV 3CL^{pro} at 250 nM and HKU4-CoV 3CL^{pro} at 200 nM. Since 3CL^{pro} enzymes cannot be saturated with this substrate at a substrate concentration that would still allow accurate fluorescent measurements without the inner filter effect, only the apparent k_{cat}/K_M values can be determined from the slope of the line that results from a plot of the enzymatic activity (y-axis), normalized for the total enzyme concentration, against the substrate concentration (x-axis).

5.3.8 Influence of dimerization on the activity of 3CL^{pro} enzymes

The dependence of the enzymatic activity on the total enzyme concentration was determined using the FRET-based assay described above. The final enzyme concentrations were varied over a concentration range from 2 μ M to 100 nM for MERS-CoV 3CL^{pro}, 500 nM to 10 nM for SARS-CoV 3CL^{pro}, 250 nM to 0.6 nM for HKU5-CoV 3CL^{pro} and 200 nM to 10 nM for HKU4-CoV 3CL^{pro}. Reactions were initiated by the addition of substrate, at a final concentration of 2 μ M, to the assay plates containing varying enzyme concentrations in the assay buffer. Initial rates were determined from the initial slopes of the progress curves at each enzyme concentration.

The rates of the 3CL^{pro} catalyzed reactions measured over a range of enzyme concentrations can be fit to either Equation 5.1 or Equation 5.2 to determine the values of the dissociation constant for the monomer-dimer equilibrium as well as the

turnover numbers. Non-linear regression and the program TableCurve 2D version 4.0 were used to fit the data to either Equation 5.1 or Equation 5.2 below [120].

$$V_{\max} = k_{\text{cat},\text{M}} \frac{-K_{\text{d}} + \sqrt{K_{\text{d}}^2 + 8K_{\text{d}}C_{\text{T}}}}{4} + k_{\text{cat},\text{D}} \frac{K_{\text{d}} + 4C_{\text{T}} - \sqrt{K_{\text{d}}^2 + 8K_{\text{d}}C_{\text{T}}}}{8} \quad (5.1)$$

In Equation 5.1, V_{\max} is the rate of the enzymatic activity calculated at each enzyme concentration (C_{T}), K_{d} is the monomer-dimer equilibrium dissociation constant, and $k_{\text{cat},\text{M}}$ and $k_{\text{cat},\text{D}}$ are the turnover numbers for the monomer and the dimer, respectively.

$$V_{\max} = k_{\text{cat}}[D] = k_{\text{cat}} \frac{K_{\text{d}} + 4C_{\text{T}} - \sqrt{K_{\text{d}}^2 + 8K_{\text{d}}C_{\text{T}}}}{8} \quad (5.2)$$

For Equation 5.2, V_{\max} , C_{T} and K_{d} have been described previously, and k_{cat} is the turnover number for the dimer only.

5.3.9 Inhibition assays

To determine the percent inhibition for compounds **1** to **9**, the total concentration of the substrate was fixed at 1.0 μM , and the enzymes was fixed at 250 nM for SARS-CoV 3CL^{pro}, HKU5-CoV 3CL^{pro}, HKU4-CoV 3CL^{pro} and at 500 nM for MERS-CoV 3CL^{pro}. DMSO stocks (100 \times) of the compounds were diluted hundred-fold to a final concentration of 50 μM in 80 μL of the enzyme solution and incubated for 20 minutes. After 20 minutes, the enzymatic activity was measured as initial slope of the progress curve, obtained by initiating the reaction with 20 μL of 5 μM substrate. % Inhibition was calculated using Equation 5.3.

$$\% \text{ Inhibition} = \left[1 - \frac{\text{Rate}_{\text{sample}} - \text{Rate}_{\text{neg}}}{\text{Rate}_{\text{pos}} - \text{Rate}_{\text{neg}}} \right] \times 100 \quad (5.3)$$

In Equation 5.3, $\text{Rate}_{\text{sample}}$ is the initial slope of the progress curve in AFU/sec measured in the presence of the compound, Rate_{pos} is the initial slope measured in the absence of any compound and Rate_{neg} is the baseline substrate hydrolysis calculated in the absence of enzyme. All the reactions were carried out in triplicate and contained a final DMSO concentration of 1%. For compounds displaying more than 50% inhibition, a more extensive characterization of the inactivation kinetics was performed through progress curve analysis. To the reaction well, 20 μL of 5 μM substrate was added to a final concentration of 1 μM and the total inhibitor concentration $[I]_{\text{total}}$ was varied from 0 to 50 μM . The reaction was initiated with the addition of 80 μL of MERS-CoV 3CL^{pro} to a final concentration of 500 nM. Fluorescence intensity was then measured over time as AFU_t for a period of 70 minutes. Equation 5.4 describes the resulting time-course of reaction.

$$[P]_t = \frac{v_i}{k_{\text{obs}}} [1 - \exp(-k_{\text{obs}} \times t) + [P]_i] \quad (5.4)$$

In Equation 5.4, v_i is the initial velocity of the reaction, k_{obs} is the observed first-order rate constant for the reaction in the absence and presence of inhibitor, t is the time in minutes, $[P]_t$ is the concentration of product produced at time t and $[P]_i$ is the initial product concentration which is zero. Product concentrations were calculated from the values of AFU_t , using experimentally determined ‘fluorescence extinction coefficient’. The resulting values of $[P]_t$ were then plotted against time t and the data were fit to Equation 5.4 with $[P]_i = 0$ using the non-linear regression program TableCurve 2D to derive the fitted parameters v_i and k_{obs} and their associated errors Δv_i and Δk_{obs} . Values for each k_{obs} were then plotted against $[I]_{\text{total}}$ and the data were fit to Equation 5.5.

$$k_{\text{obs}} = \frac{k_{\text{inact}} [I]_{\text{total}}}{K_I + [I]_{\text{total}}} \quad (5.5)$$

In Equation 5.5, k_{inact} defines the maximum rate of inactivation at infinite inhibitor concentration, and K_I defines the concentration of inhibitor that yields a rate of inactivation equal to $1/2k_{\text{inact}}$. The half-life of inactivation at infinite inhibitor concentration, which is a measure of inactivation efficiency, is defined as $t_{1/2}^{\infty} = 0.693/k_{\text{inact}}$.

5.3.10 Analytical ultracentrifugation (AUC) analysis

To determine the oligomeric state of MERS-CoV 3CL^{pro}, sedimentation velocity experiments were performed at 20 °C on the Beckman-Coulter XLA ultracentrifuge using varying concentrations of MERS-CoV 3CL^{pro} (4 to 23 μM) in 25 mM HEPES pH-7.5, 50 mM NaCl and 1 mM TCEP at 50,000 rpm. In order to characterize the effect of the ligand on the monomer-dimer equilibrium of MERS-CoV 3CL^{pro}, sedimentation velocity experiments were conducted on the Beckman-Coulter XLI instrument using different stoichiometric ratios of MERS-CoV 3CL^{pro} with compounds **6** and **10**. Samples were prepared by mixing 25 μM of MERS-CoV 3CL^{pro} with 25, 50 and 100 μM of compound **6** or **10** and incubating the mixture overnight at 4 °C before performing the experiments. Absorbance optics (280 nm) and interference optics were utilized for protein detection. Solvent density, viscosity and partial specific volumes were calculated using SEDNTERP. SEDPHAT was used to fit the data to monomer-dimer self-association model to estimate the sedimentation coefficients (S), apparent molecular weights, K_d and k_{off} from size distribution analysis. To obtain exact molecular weights, sedimentation equilibrium experiments were performed at concentrations of 3 and 17 μM MERS-CoV 3CL^{pro}. The experiments were done at 20 °C utilizing a 2-channel centerpiece and run at multiple speeds (8100, 13,800 and 24,000 rpm) in a AN-60 Ti rotor.

5.3.11 SEC-MALS analysis

Size-exclusion chromatography (SEC) coupled with on-line multi-angle light scattering (MALS) analysis was utilized to determine the molecular weight and homo-

geneity of the protein sample. SEC-MALS analysis was performed on 100 μL aliquots of purified MERS-CoV 3CL^{pro} at three different concentrations, 90 μM (or 3 mg/ml), 45 μM and 22.5 μM , in buffer containing 25 mM HEPES pH-7.5, 50 mM NaCl and 1 mM TCEP. Superdex 75 analytical gel filtration column (GE healthcare) was used at a flow rate of 0.5 ml/min for size exclusion and was coupled to a Dawn Heleos MALS instrument (Wyatt technology) and an Optilab rEX instrument (Wyatt technology). Weighted-average molecular weight and polydispersity index was determined by measuring the intensity of Rayleigh scattering as a function of the differential refractive index and the angular dependence of the scattered light within the horizontal plane.

5.3.12 MERS-CoV 3CL^{pro} activation and inhibition by a non-covalent inhibitor

The rates of the MERS-CoV 3CL^{pro} catalyzed reactions were determined at final enzyme concentrations of 0.5, 1.0 and 2.0 μM and in the absence and presence of varying concentrations (0.1 μM to 60 μM) of compound **10**. The substrate concentration was fixed at 2.0 μM . DMSO stocks (100 \times) of compound **10** were diluted hundred-fold in 80 μL of enzyme solution and incubated for 10 minutes. At the same time, a zero-inhibitor control reaction was set up by mixing DMSO to a final concentration of 1% into 80 μL of enzyme solution. After 10 minutes, the rate of the enzymatic activity was measured as the initial slope of the progress curve, obtained by initiating the reaction with 20 μL of 10 μM substrate. Equation 5.6 was utilized to calculate the percent activity.

$$\% \text{ Activity} = \left[\frac{\text{Rate}_{\text{sample}} - \text{Rate}_{\text{neg}}}{\text{Rate}_{\text{pos}} - \text{Rate}_{\text{neg}}} \right] \times 100 \quad (5.6)$$

$\text{Rate}_{\text{sample}}$, Rate_{pos} and Rate_{neg} are described above for Equation 5.3.

5.3.13 MERS-CoV 3CL^{pro} Crystallization, X-ray Data collection and Structure determination

Purified MERS-CoV 3CL^{pro} was concentrated to 1.6 mg/mL in 25 mM HEPES pH-7.5, and 2.5 mM DTT. Inhibitor complexes of MERS-CoV 3CL^{pro} with compounds **6** and **11** were formed by incubating MERS-CoV 3CL^{pro} with the compounds in a 1:3 stoichiometric ratio at 4 °C overnight. After iterative rounds of optimization of the crystallization conditions based on the initial hits obtained from high-throughput screening of Qiagen Nextel Screens, crystals of MERS-CoV 3CL^{pro} inhibitor complexes suitable for X-ray diffraction were grown by the hanging-drop, vapor diffusion method at 20 °C in 0.2 M sodium acetate, 0.1 M Bis-Tris pH-7.0 and 20% PEG-3350 for the MERS-CoV 3CL^{pro}-**6** complex, and 0.2 M ammonium acetate, 0.1 M Bis-Tris pH-5.5, 12% PEG-3350 for the MERS-CoV 3CL^{pro}-**11** complex. For X-ray data collection, crystals were flash-cooled in liquid nitrogen after dragging the crystals through a cryo-solution that contained the crystallization solution supplemented with 15% 2-methyl-2,4-pentanediol.

X-ray diffraction data were collected for MERS-CoV 3CL^{pro}-**6** and MERS-CoV 3CL^{pro}-**11** complexes at the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) Sector 31 and the Life Sciences Collaborative Access Team (LS-CAT) Sector 21 at the Advanced Photon Source, Argonne National Laboratory, respectively. Data were processed and scaled using Mosflm version 7.0.5 [121] and HKL2000 version 706 [122]. The method of molecular replacement was used to obtain initial phases using the program PHASER-MR in Phenix suite version 1.8.4 [123].

For MERS-CoV 3CL^{pro}-**6** complex, the X-ray structure of SARS-CoV 3CL^{pro} (PDB ID: 3V3M) was used as a phasing model [38]. The final MERS-CoV 3CL^{pro}-**6** complex structure was then used to calculate the initial phases for the MERS-CoV 3CL^{pro}-**11** complex model. Automated model building using Autobuild in Phenix was initially used to build a preliminary model of the MERS-CoV 3CL^{pro}-**6** inhibitor complex. Each structure was then refined using iterative cycles of refinement using

Phenix Refine coupled to manual model building using COOT [124] based on $F_o - F_c$ and $2F_o - F_c$ maps. Coordinates and molecular library files for inhibitor molecules were built using the program eLBOW in the Phenix suite. Water molecules were added to peaks in residual ($F_o - F_c$) density maps that were greater than 3σ using the “Find Water” function in COOT. MolProbity was used to assess structural quality of the final model [125]. The measured structure factor amplitudes and the atomic coordinates for the final structures were deposited in the Protein Data Bank with accession codes 4RSP (MERS-CoV 3CL^{pro}-**6** complex) and 4YLU (MERS-CoV 3CL^{pro}-**11** complex), respectively. Structural superposition was performed using the method of Least-Squares fitting (LSQ) of C-alpha atoms in COOT. Pymol was used to generate figures of all the structures [126].

5.4 Results

5.4.1 Production of MERS-CoV 3CL^{pro} with authentic N- and C-termini

Insertion of the nsp4|nsp5 cleavage site between N-terminal (His)₆-tag and the coding region for MERS-CoV 3CL^{pro} (Figures 5.2.A) results in auto-processing of the His-tag, and over-expression of MERS-CoV 3CL^{pro} without any N-terminal extension in *E. coli* BL21-DE3 cells. MERS-CoV 3CL^{pro} was purified to high purity (Figure 5.2.B) and an overall yield of 10% using four sequential chromatographic steps. A summary of the percent enzyme yield, total activity units, and the fold-purification after each chromatographic step is summarized in Table 5.1. Approximately 12 mg of highly pure MERS-CoV 3CL^{pro} can be obtained per liter of bacterial cell culture.

Western blot analysis of purified MERS-CoV 3CL^{pro} using an anti-(His)₆ antibody confirmed the absence of the N-terminus (His)₆-tag associated with the expression plasmid (Figure 5.2.C). To further verify the production of the enzyme with correct N- and C-termini, the molecular weight of purified MERS-CoV 3CL^{pro} was determined by MALDI-TOF analysis (Figure 5.2.D). MALDI calculated molecular weight of purified MERS-CoV 3CL^{pro} is 33.4 kDa, which is close to the theoretical molecular

weight of 33.3 kDa for the authentic/mature MERS-CoV 3CL^{pro} monomer. These results demonstrate that the N-terminal (His)₆-tag is auto-catalytically removed by MERS-CoV 3CL^{pro} during its expression in *E. coli*, indicating MERS-CoV 3CL^{pro} is enzymatically active when expressed in *E. coli*.

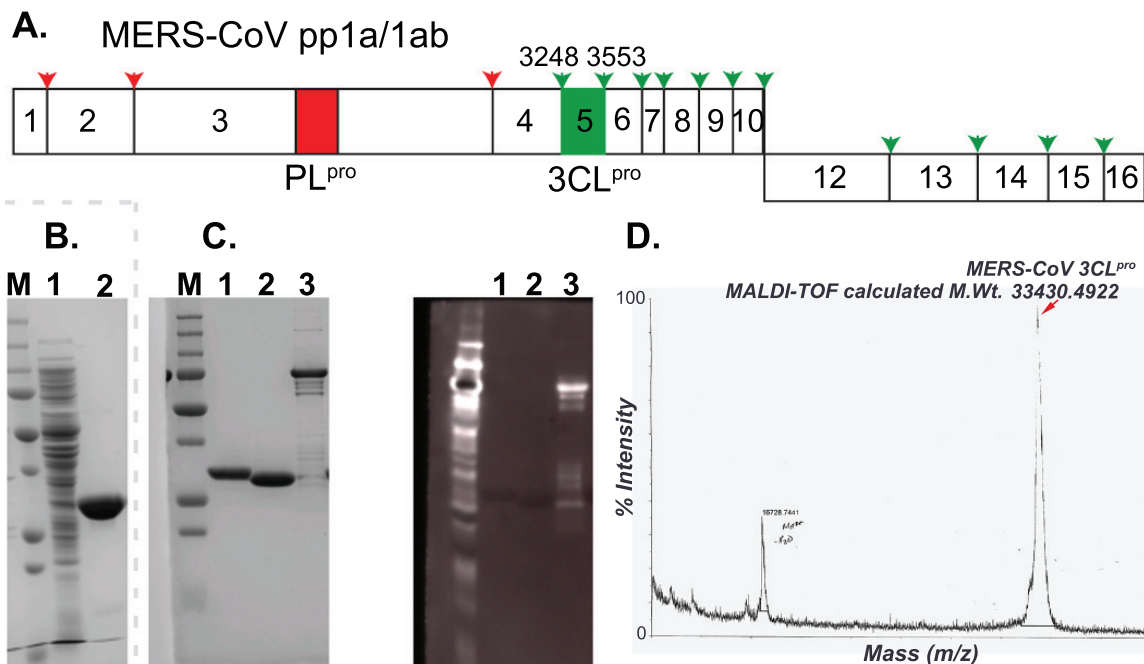


Fig. 5.2.: Production of MERS-CoV 3CL^{pro} with authentic N- and C-termini. A. Schematic diagram of MERS-CoV polyproteins pp1a/1ab. PL^{pro} domain within the nsp3 is shown in red while the 3CL^{pro} (nsp5) is shown in green. Red and green arrows indicate the cleavage sites processed by PL^{pro} and 3CL^{pro} respectively. **B.** Purification of MERS-CoV 3CL^{pro} as analyzed on SDS-PAGE gel. Lane M represents the molecular weight markers, lane 1 is the crude lysate obtained after lysing the bacterial cells expressing MERS-CoV 3CL^{pro} and lane 2 is the purified MERS-CoV 3CL^{pro}. **C.** Verification of the production of MERS-CoV 3CL^{pro} with authentic N- and C-termini. On the left is the gel analysed through coomassie staining; on the right are the same samples analysed through western blot using an anti-His antibody. Lane M represents the molecular weight markers, Lane 1 is the purified MERS-CoV 3CL^{pro}, lane 2 is the purified SARS-CoV 3CL^{pro} (negative control) and lane 3 is a protein with (His)₆ N-terminal tag (positive control). **D.** Molecular weight of purified MERS-CoV 3CL^{pro} was determined to be 33430 Da using MALDI-TOF. MALDI-TOF calculated molecular weight is close to the theoretical molecular weight of 33330 Da for MERS-CoV 3CL^{pro} monomer with authentic N-terminus.

Table 5.1.
Purification summary of MERS-CoV 3CL^{pro} per liter of *E. coli*
BL21-DE3

Sample	Protein (mg)	Total activity Units	Specific activity (Units/mg)	Fold purification	% Yield
Lysate	1102	1168	1	1	100
Phenyl-Sepharose	219	185	1	1	16
DEAE	22	189	8	8	16
Mono-S	15	142	9	9	12
Superdex 75	12	114	10	10	10

5.4.2 MERS-CoV 3CL^{pro} hydrolyzes a fluorescent peptide substrate with lower efficiency than other 3CL^{pro} enzymes

A FRET-based peptide substrate was used to measure the enzymatic activity of MERS-CoV 3CL^{pro} as a function of substrate concentration over a substrate concentration range from 0 to 2.0 μM (Figure 5.3.A). It was observed that MERS-CoV 3CL^{pro} cannot be saturated by the substrate over this concentration range, which is typical for other coronavirus 3CL^{pro} enzymes since the K_M values for peptide substrates approach 1 mM [127–130]. Therefore, the slope of the kinetic response of MERS-CoV 3CL^{pro} to increasing substrate concentration was determined to derive an apparent (k_{cat}/K_M) value, which is a measure of enzymatic efficiency. Next, the determination and comparison of the apparent (k_{cat}/K_M) values for 3CL^{pro} enzymes from SARS-CoV, HKU5-CoV and HKU4-CoV under similar experimental conditions were performed (Figure 5.3.B). MERS-CoV 3CL^{pro} is able to hydrolyze the peptide substrate, however, the enzymatic efficiency of MERS-CoV 3CL^{pro} ($k_{\text{cat}}/K_M = 3.1 \pm 0.03 \times 10^{-2} \mu\text{M}^{-1} \text{min}^{-1}$) is noticeably lower than other 3CL^{pro} enzymes tested. Specifically, MERS-CoV 3CL^{pro} was five-fold less efficient at processing the peptide substrate when compared to SARS-CoV 3CL^{pro}. Even among the β -CoVs from the same 2c genogroup (MERS, HKU5 and HKU4), MERS-CoV 3CL^{pro} was the least efficient enzyme.

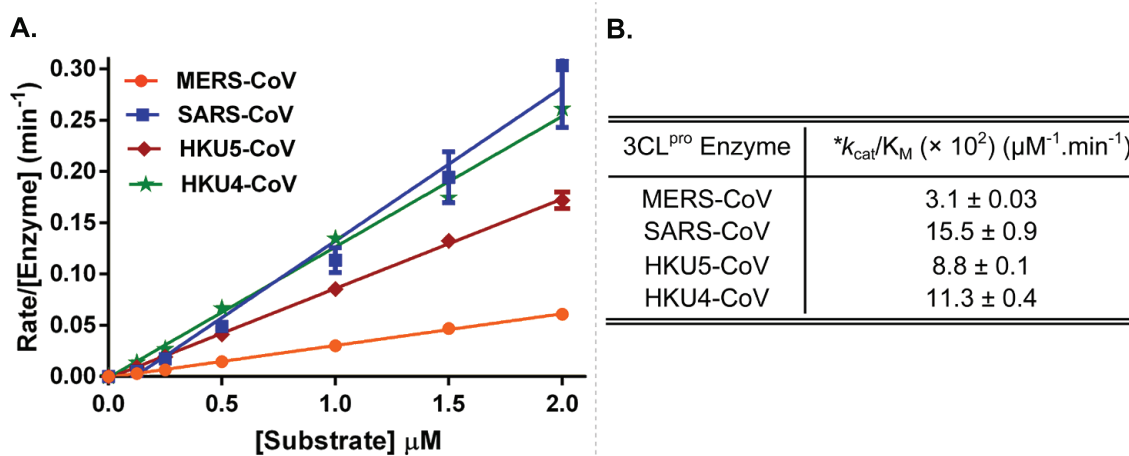


Fig. 5.3.: Comparison of enzymatic efficiencies (k_{cat}/K_M) of 3CL^{pro} enzymes from different CoVs. **A.** Rates for the enzymatic activity, normalized to the total enzyme concentration, are plotted as a function of varying substrate concentrations. Total concentration of each enzyme in the final reaction is following: MERS-CoV 3CL^{pro} at 1 μM , SARS-CoV 3CL^{pro} at 100 nM, HKU5-CoV 3CL^{pro} at 250 nM and HKU4-CoV 3CL^{pro} at 200 nM. Slope of the line represents the apparent value of k_{cat}/K_M . Error bars represent the standard deviation for triplicate data. **B.** *Apparent value of k_{cat}/K_M for the non-saturable substrate, calculated as the slope of the linear plot from panel A.

5.4.3 MERS-CoV 3CL^{pro} is a weakly associated dimer

Since a dimer has consistently been shown to be the catalytically active form of all 3CL^{pro} enzymes studied to date, the hypothesis that the lower enzymatic efficiency of MERS-CoV 3CL^{pro} is a result of the reduction in its ability to dimerize was tested. Therefore, the dependence of the enzymatic activity of MERS-CoV 3CL^{pro} on the total enzyme concentration was determined and compared to other 3CL^{pro} enzymes from HKU4, HKU5 and SARS coronaviruses (Figure 5.4).

Table 5.2.

Comparison of the apparent turnover number, k_{cat} , and the monomer-dimer dissociation constant, K_{d} , for different 3CL^{pro} enzymes

3CL ^{pro}	Non-linear fitting of kinetic data ^a	
	k_{cat} (min ⁻¹) ^b	K_{d} (μM)
MERS-CoV	0.2 ± 0.02	7.8 ± 1.3
SARS-CoV	0.47 ± 0.03	0.06 ± 0.01
HKU5-CoV	0.53 ± 0.02	0.06 ± 0.01
HKU4-CoV	0.84 ± 0.07	0.1 ± 0.03

^a Values determined through non-linear fitting of the kinetic data to Equation 5.2.

^b k_{cat} represents the apparent turnover number.

It is immediately apparent from the data plotted in Figure 5.4 that the response of MERS-CoV 3CL^{pro} enzymatic activity to an increasing enzyme concentration is non-linear. The strong curvature suggests that a dimer is either the most active form or the only active form of MERS-CoV 3CL^{pro}. To determine the mechanism of dimerization, the data in Figure 5.4 were first fit to Equation 5.1 (see Experimental Procedures), which describes a model where both the monomer and the dimer are active. A fit of the data to Equation 5.1 yielded a negative turnover value for the monomer ($k_{\text{cat},\text{M}}$), suggesting the monomer is inactive and that the dimer is the only active form of the enzyme. Therefore, the data were fit to Equation 5.2 (see Experimental Procedures), which considers only the dimer as the active form of the enzyme. The kinetic data for all four 3CL^{pro} enzymes, MERS-CoV, HKU4-CoV, HKU5-CoV and SARS-CoV,

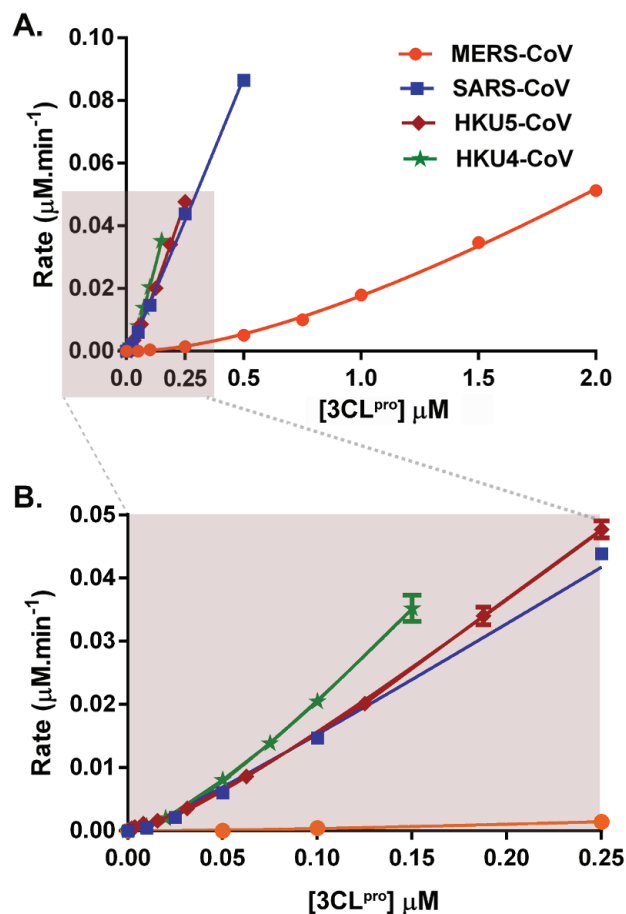


Fig. 5.4.: Dependence of the enzymatic activity of MERS-CoV, HKU4-CoV, HKU5-CoV and SARS-CoV 3CL^{pro}s on the total enzyme concentration. **A.** Kinetic response of each CoV 3CL^{pro} to increasing enzyme concentration is plotted along with the resulting fit of the data to Equation 5.2. Resulting values for the apparent turnover number, k_{cat} , and the monomer-dimer equilibrium constant, K_{d} , are shown in Table 5.2. Final enzyme concentrations varied over the concentration ranges of 2 μM to 100 nM for MERS-CoV 3CL^{pro}, 500 nM to 10 nM for SARS-CoV 3CL^{pro}, 250 nM to 0.6 nM for HKU5-CoV 3CL^{pro} and 200 nM to 10 nM for HKU4-CoV 3CL^{pro}. Final substrate concentration was fixed at 2 μM . Experiments were done in triplicate. Error bars represent the standard deviation for triplicate data. Shaded box represents the data that are plotted in panel B. **B.** Enlarged view of the fitted data at low total enzyme concentrations, marked in shaded box in panel A, illustrating the non-linear dependence of enzymatic activity on the total concentrations of 3CL^{pro} from SARS-CoV, HKU5-CoV and HKU4-CoV.

fit well to this model and the resulting values for the monomer-dimer equilibrium dissociation constant, K_d , and apparent turnover number, k_{cat} , for each enzyme are provided in Table 5.2.

The lower k_{cat} value for MERS-CoV 3CL^{pro}, when compared to other coronavirus 3CL^{pro} enzymes, indicates a moderate reduction (2- to 4-fold) in its ability to turnover the substrate, which is consistent with the observed lower apparent (k_{cat}/K_M) value. In contrast, there is a substantial reduction in the ability of MERS-CoV 3CL^{pro} to dimerize compared to the other 3CL^{pro} enzymes. Based on the K_d values, the capacity of MERS-CoV 3CL^{pro} to dimerize is approximately 78- to 130-fold weaker than the other enzymes (Table 5.2). These results indicate that the MERS-CoV 3CL^{pro} dimer is much more weakly associated than the other coronavirus 3CL^{pro} enzymes studied and raises questions as to the structural and mechanistic differences among the 3CL^{pro} enzymes that ultimately regulate protease activity during coronavirus replication.

5.4.4 MERS-CoV 3CL^{pro} inhibition by designed peptidomimetic compounds

In an effort to develop potent inhibitors of MERS-CoV 3CL^{pro}, our collaborator (Prof. Arun Ghosh, Purdue University) designed and synthesized nine peptidomimetic compounds containing a Michael acceptor group, i.e. an α,β -unsaturated carbonyl, capable of irreversibly reacting with the active site cysteine of MERS-CoV 3CL^{pro} (Table 5.3). These compounds were designed and synthesized based on our understanding and knowledge of the interactions of similar inhibitor molecules with SARS-CoV 3CL^{pro} [37,53]. At a concentration of 50 μ M, compounds **6** to **9** displayed more than 50% inhibition of MERS-CoV 3CL^{pro} and were further evaluated for their ability to inactivate the enzyme in a time- and concentration-dependent manner (Figure 5.5). Data from the kinetic progress curve for compound **6** (Figure 5.5), as well as for compounds **7** to **9**, were fit to the appropriate equations (see Experimental

Procedures) to obtain the kinetic parameters, k_{inact} , $t_{1/2}^{\infty}$ and K_I , and the resulting values are provided in Table 5.3.

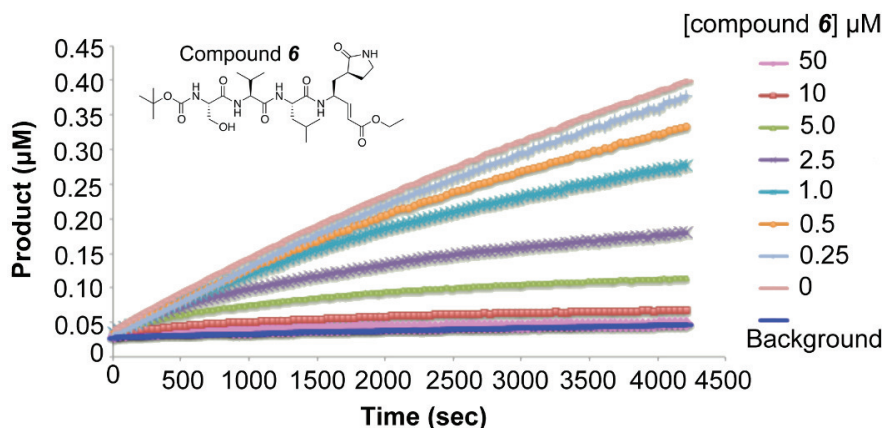


Fig. 5.5.: **Progress curves for the MERS-CoV 3CL^{pro} catalyzed reaction in the presence of compound 6.** Time-dependent hydrolysis of 1 μM of substrate catalyzed by 500 nM of MERS-CoV 3CL^{pro} was measured over a time period of 70 minutes and at fixed variable concentrations of compound 6 ranging from 0 to 50 μM . Values for the inactivation kinetic parameters k_{inact} , $t_{1/2}^{\infty}$ and K_I were calculated by fitting the progress curve data to Equation 5.4 and Equation 5.5. Chemical structure of compound **6** is shown in inset.

Out of the nine compounds tested, four compounds, **6** to **9**, were identified as micromolar inhibitors of MERS-CoV 3CL^{pro} with K_I values less than 10 μM , (Table 5.3). Analysis of structure-activity relationships (SAR) of these compounds suggests that the S_2 subsite pocket of MERS-CoV 3CL^{pro} is small and can only accommodate a smaller P_2 -isobutyl substituent (compounds **6** to **9**), but not bigger substituents such as P_2 -benzyl or P_2 -isobutylenyl (compounds **1** to **5**). It was also observed that replacing the P_4 -ethoxy (Compound **6**) with P_4 -isopropoxy (compounds **7** and **8**) has no effect on the inhibitory activity of the compounds. Finally, these compounds provide an excellent chemical scaffold to study the molecular details of interactions of substrate-like compounds with the enzyme and to develop more potent inhibitors of MERS-CoV 3CL^{pro} for therapeutic intervention.

To evaluate broad-spectrum specificity of these compounds, the percent inhibition of SARS-CoV 3CL^{pro}, HKU5-CoV 3CL^{pro} and HKU4-CoV 3CL^{pro} after 20 minutes incubation in the presence of 50 μM of compounds **6** to **9** was also determined.

Table 5.3.
Chemical structures and inhibitory activity of compounds **1** to **11**
against MERS-CoV 3CL^{pro}

Peptidomimetic compounds with Michael-acceptor groups					Non-covalent peptidomimetics	
Cmpd [†]	% Inhi [*]	k_{inact} ^b	$t_{1/2}^{\infty}$ ^c	K_I ^d	Cmpd	IC ₅₀ ^e
1 ^a	46	nd	nd	nd	10	>100
2 ^a	11	nd	nd	nd	11	>100
3 ^a	21	nd	nd	nd		
4 ^a	0	nd	nd	nd		
5 ^a	46	nd	nd	nd		
6	99	0.81 ± 0.08	0.86 ± 0.08	3.6 ± 0.8		
7	100	0.84 ± 0.05	0.83 ± 0.05	4.7 ± 0.6		
8	100	1.12 ± 0.20	0.62 ± 0.11	9.0 ± 2.3		
9	100	1.13 ± 0.20	0.61 ± 0.11	9.9 ± 2.6		

[†] The Michael-acceptor group for Compound **1** is shaded to highlight this group for all the compounds; ^{*} % Inhibition measured as % loss in enzymatic activity after 20 minutes incubation of 500 nM of MERS-CoV 3CL^{pro} with 50 μM of compound; ^a k_{inact} , $t_{1/2}^{\infty}$ and K_I were not determined (nd) for compounds **1** to **5** that showed <50% inhibition of MERS-CoV 3CL^{pro}; ^b k_{inact} ($\times 10^3$)(sec⁻¹); ^c $t_{1/2}^{\infty}$ ($\times 10^{-3}$)(sec); ^d K_I (μM); ^e IC₅₀ (μM) calculated from a dose response curve determined after 10 minutes incubation of 1 μM MERS-CoV 3CL^{pro} with varying concentrations of compounds.

Except for compound **9** which inhibited SARS-CoV 3CL^{pro} by 76%, 100% inhibition of all other enzymes in the presence of compounds **6** to **9** was observed. Furthermore, progress curve analysis of HKU5-CoV 3CL^{pro} and HKU4-CoV 3CL^{pro} in the presence of varying concentrations of compounds **6** to **9** was performed. The K_I values of compounds **6** to **9** for HKU5-CoV 3CL^{pro} are 0.49 ± 0.16 , 0.60 ± 0.21 , 1.30 ± 0.53 , and $0.47 \pm 0.06 \mu\text{M}$, respectively. The K_I values of compounds **6** to **9** for HKU4-CoV 3CL^{pro} are 0.39 ± 0.14 , 0.50 ± 0.17 , 0.85 ± 0.33 , and $0.64 \pm 0.25 \mu\text{M}$, respectively. These data suggest that peptidomimetic compounds **6** to **9** have the potential to be developed as coronavirus 3CL^{pro} inhibitors with broad-spectrum specificity.

5.4.5 Weak association of the MERS-CoV 3CL^{pro} dimer is supported by AUC and SEC-MALS studies

To further explore the mechanism of MERS-CoV 3CL^{pro} dimerization, analytical ultracentrifugation sedimentation velocity (AUC-SV) studies were performed at varying concentrations of MERS-CoV 3CL^{pro} (Figures 5.6.A, 5.7.A and 5.7.B). Unlike enzyme kinetics, AUC allows determination of the monomer-dimer equilibrium constant (K_d) in the absence of substrate. MERS-CoV 3CL^{pro} displayed a continuous size distribution at different protein concentrations. Two distinct peaks corresponding to monomer (2.9 S) and dimer (3.9 S) species are observed, with the dimer peak becoming more pronounced at higher enzyme concentrations (Figure 5.7.A). The AUC data were fitted to a monomer-dimer equilibrium model to determine the values for K_d and k_{off} , where K_d is the equilibrium dissociation constant for a monomer from the dimer and k_{off} is the rate constant for dissociation of the monomer from the dimer. The resulting best-fit value for K_d is $52 \pm 5 \mu\text{M}$ and that for k_{off} is 10^{-4} sec^{-1} . The K_d value of $52 \mu\text{M}$ for MERS 3CL^{pro} is dramatically different from SARS-CoV 3CL^{pro}, which has reported K_d values ranging from low nM up to $10 \mu\text{M}$ depending on the enzyme construct used and the experimental conditions and methods utilized to determine the dissociation constant [45]. The dimer affinity of MERS-CoV 3CL^{pro}

is substantially weaker than that for SARS-CoV 3CL^{pro}, when comparing the same enzyme construct, i.e. the enzyme without any N- or C-terminal modifications. The AUC-SV calculated K_d value for MERS-CoV 3CL^{pro} is $\sim 150,000$ -times higher than the value of 0.35 nM determined for SARS-CoV 3CL^{pro} [43].

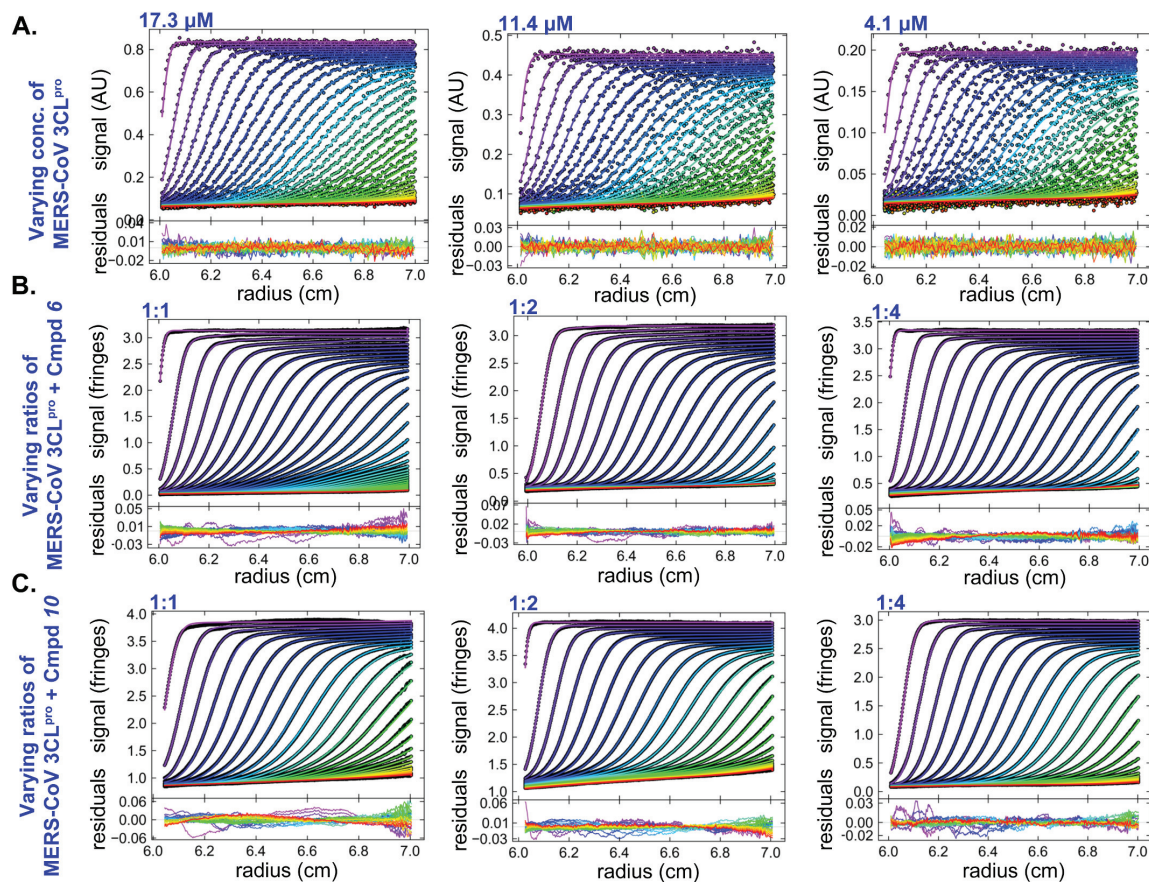


Fig. 5.6.: AUC-SV raw boundary profiles for MERS-CoV 3CL^{pro}. **A.** Raw boundary profiles from absorbance signal at 280 nm for varying concentrations of MERS-CoV 3CL^{pro}. **B.** Raw boundary profiles from interference optics signal for MERS-CoV 3CL^{pro} in complex with varying stoichiometric ratios of compound **6**. **C.** Raw boundary profiles from interference optics signal for MERS-CoV 3CL^{pro} in complex with varying stoichiometric ratios of compound **10**. Points represent the experimental data and solid lines represent the best fit to Lamm equation using the SEDFIT program.

The AUC results (Figure 5.7.A) show that the monomer peak at ~ 2.9 S does not gradually shift peak position towards the dimer peak at ~ 3.9 S with increasing concentrations of MERS-CoV 3CL^{pro}; rather, the two peaks change in area, which

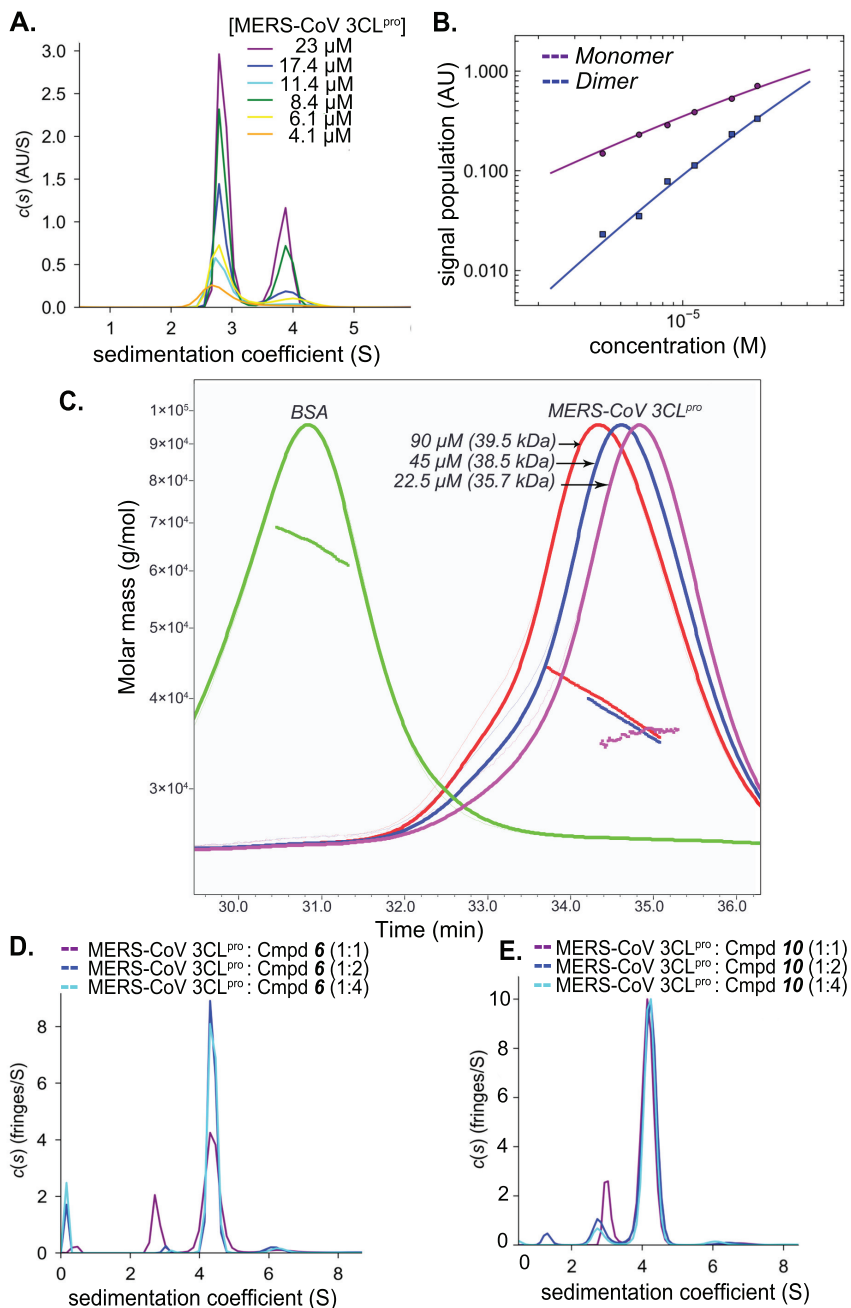


Fig. 5.7.: AUC-SV and SEC-MALS analysis of MERS-CoV 3CL^{pro}. **A.** Sedimentation coefficient distribution for varying concentrations of MERS-CoV 3CL^{pro} (4.1 to 23 μM) with sedimentation coefficient values of 2.9 S and 3.9 S for the monomer and the dimer, respectively. Best-fit value for AUC-SV calculated K_d is $52 \pm 5 \mu\text{M}$. **B.** Population isotherm showing contribution of monomer (purple) and dimer (blue), respectively. Solid lines represent the best fit to a monomer-dimer equilibrium model. **C.** Molecular weight determination through SEC-MALS analysis of varying MERS-CoV 3CL^{pro} concentrations indicate monomer as the major species in solution. **D.** Sedimentation coefficient distribution of MERS-CoV 3CL^{pro} (25 μM) in the presence of different stoichiometric ratios of compound **6** (25, 50 and 100 μM) and **E.** compound **10** (25, 50 and 100 μM).

is indicative of very slow monomer-dimer exchange rate ($k_{\text{off}} \sim 10^{-4} \text{ sec}^{-1}$) and the formation of hydrodynamically stable monomer and dimer species (Figure 5.7.B) [131]. This k_{off} value is 1,000-times slower than the k_{off} value (10^{-1} sec^{-1}) reported for SARS-CoV 3CL^{pro} indicating that the SARS-CoV enzyme has a significantly more rapid monomer-dimer exchange rate [132]. These observations support a model whereby the MERS-CoV 3CL^{pro} dimer is weakly associated, suggesting the enzyme exists mainly as a monomer in solution.

SEC-MALS analysis also confirmed MERS-CoV 3CL^{pro} monomer as the major species in solution (Figure 5.7.C). Three tested MERS-CoV 3CL^{pro} concentrations eluted at slightly different retention times. SEC-MALS calculated molecular weights for 90 μM , 45 μM and 22.5 μM concentration of MERS-CoV 3CL^{pro} are 39.5 kDa, 38.5 kDa and 35.7 kDa, respectively. Calculated average molecular weight is 37.9 ± 2 kDa, which is close to the theoretical molecular weight of 33.3 kDa for the monomer and is clearly indicative of monomer as the major species in the solution.

5.4.6 MERS-CoV 3CL^{pro} undergoes extensive ligand-induced dimerization

The weak association of MERS-CoV 3CL^{pro} monomers engenders questions such as, “Are higher levels of expression of 3CL^{pro} in MERS-CoV infected cells necessary to allow formation of active dimer?” “Are other mechanisms such as substrate or ligand-induced dimerization involved in activating 3CL^{pro}?” To explore the latter question of ligand-induced dimerization of MERS-CoV 3CL^{pro}, AUC experiments were performed in the presence of compound **6**, which acts as a substrate mimetic and mechanism-based inhibitor, aka suicide substrate. Peptidomimetic compounds such as compound **6**, which contain a Michael acceptor group, interact and react with the active site cysteine of cysteine proteases to covalently modify them. The compound **6** was utilized to form a covalent MERS-CoV 3CL^{pro}-inhibitor **6** complex that is stable over long periods of time making it amenable to analysis by AUC-SV

experiments. In contrast, incubation of a normal peptide substrate with the enzyme would lead to immediate hydrolysis of the substrate and dissociation of the products from the enzyme, confounding AUC experiments and subsequent data analysis.

MERS-CoV 3CL^{pro} was incubated with varying concentrations of compound **6** in stoichiometric ratios of (1:1, 1:2 and 1:4). The modified enzyme was then subjected to AUC studies to determine the influence of compound **6** on the monomer-dimer equilibrium (Figures 5.6.B, 5.7.D). A significant shift in the area under 2.9 S peak (monomer) to 4.1 S peak (dimer) is detected upon addition of increasing concentrations of compound **6** (Figure 5.7.D). Similar results were obtained when AUC studies were performed utilizing a complex of MERS-CoV 3CL^{pro} with a non-covalent peptidomimetic inhibitor (Compound **10**, Figures 5.6.C and 5.7.E). The transition of MERS-CoV 3CL^{pro} from monomer to dimer in the presence of compounds **6** and **10** suggests that the enzyme undergoes extensive dimerization upon substrate binding.

5.4.7 MERS-CoV 3CL^{pro} is activated by ligand-induced dimerization

The observed ligand-induced dimerization of MERS-CoV 3CL^{pro}, as demonstrated through AUC studies, prompted us to investigate whether or not the enzymatic activity of MERS-CoV 3CL^{pro} could be increased at low concentrations of a compound via ligand-induced dimerization. To do so, a non-covalent peptidomimetic compound (compound **10**, Figure 5.8.A) that we previously identified as an inhibitor of SARS-CoV 3CL^{pro} was used. Due to the time-dependent, irreversible nature of the reaction between compound **6** and MERS-CoV 3CL^{pro}, use of compound **6** was not ideal for these kinetic studies as it would further complicate kinetic data analysis.

The kinetic response of MERS-CoV 3CL^{pro} to increasing concentrations of compound **10** was first measured at a single enzyme concentration of 1.0 μM (Figure 5.8.A). Interestingly, an increase in the activity of MERS-CoV 3CL^{pro}, as high as 195%, was observed in the presence of low inhibitor concentrations (0.1 to 20 μM). Inhibition of enzymatic activity was observed only at higher inhibitor concentrations

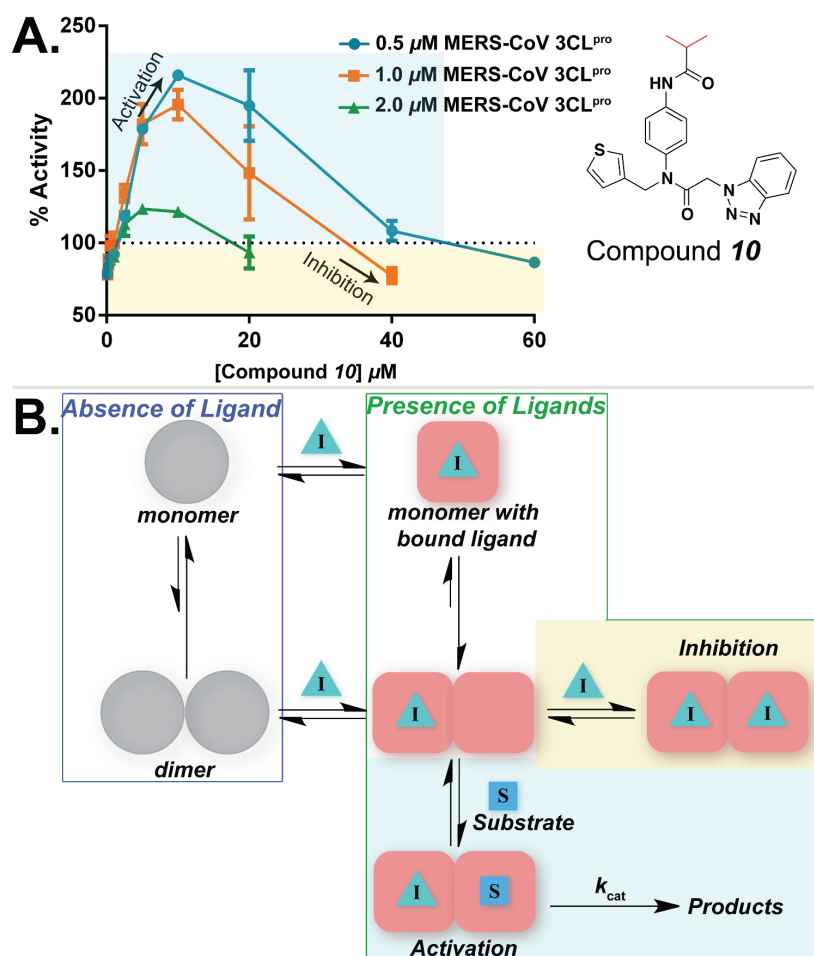


Fig. 5.8.: Activation of MERS-CoV 3CL^{pro} via ligand-induced dimerization. A. Enzymatic activity of 0.5, 1.0 and 2.0 μM MERS-CoV 3CL^{pro} was measured in the absence and presence of varying concentrations of compound **10**. Substrate concentration was fixed at 2.0 μM . % Activity, normalized to zero-inhibitor enzymatic activity, was plotted as a function of increasing inhibitor concentration. Error bars represent the standard deviation for triplicate data. Increase in enzymatic activity (highlighted in cyan shaded box) is observed in the presence of low concentrations of compound **10**. Inhibition of enzymatic activity is observed at higher inhibitor concentrations (highlighted in yellow shaded box).

B. Kinetic model describing the equilibrium between different species of MERS-CoV 3CL^{pro} that are formed in the absence (blue box) and presence (green box) of a ligand is shown. Based on the AUC calculated K_d value of $\sim 52 \mu\text{M}$, MERS-CoV 3CL^{pro} primarily exists as a monomer in solution in the absence of a ligand. Upon ligand binding (inhibitor “I” in our case) to the monomer, the monomer-dimer equilibrium shifts towards dimer formation. Next, under lower inhibitor concentrations (cyan shaded box), substrate binds in the second active site and catalysis takes place. However, under higher inhibitor concentrations (yellow shaded box), inhibitor directly competes with the substrate for the second active site and inhibition of the enzymatic activity is observed.

(40 μM or greater). These results suggest that at low concentrations, compound **10** binds to a monomer and induces the formation of a dimer. The resulting dimer then has one free active site that is capable of processing the substrate. At higher concentrations of inhibitor, the substrate and inhibitor directly compete for the free active site.

The model of activation and inhibition suggested by the data at 1 μM of enzyme would predict that at higher enzyme concentrations, less activation by a compound would be observed at lower inhibitor concentrations and inhibition of activity would be detected at lower inhibitor concentrations since the equilibrium would be pushed towards dimer formation. In contrast, lower enzyme concentrations would result in higher activation by compounds and inhibition by the compound would occur at significantly higher compound concentrations. Therefore, the activity of MERS-CoV 3CL^{pro} was further measured at two additional enzyme concentrations (0.5 and 2.0 μM) in the presence of varying concentrations of compound **10**. Remarkably, it was observed that the activation effect was most pronounced at the lowest MERS-CoV 3CL^{pro} concentration tested (0.5 μM), and the effect decreased as the enzyme concentration was increased (1.0 and 2.0 μM) (Figure 5.8.A). Moreover, inhibition by compound **10** occurred at lower compound concentrations when higher concentrations of enzymes were used. These observations further support a model whereby enzyme activation can occur through ligand-induced dimerization.

The activation and inhibition of MERS-CoV 3CL^{pro} by compound **10** can be explained by a simple kinetic model depicted in Figure 5.8.B. The MERS-CoV 3CL^{pro} monomer exists in equilibrium with the dimer and their relative concentrations depend on the total enzyme concentration. In the absence of substrate or compound, the K_d value is 52 μM and the equilibrium is represented by the grey spheres (blue box) in Figure 5.8.B. The monomer is unable to hydrolyze the substrate and is, therefore, inactive. Binding of inhibitor (blue triangle) to the monomer results in monomer-to-dimer switch leading to the formation of a dimer that contains inhibitor bound in one of the active sites. Once the dimer is formed, the substrate binds in the second

active site and catalysis takes place. Under high inhibitor concentrations, however, the inhibitor molecule directly competes with substrate for the free dimer active site and inhibition of the enzymatic activity is observed as a result.

The induced dimerization and activation is also expected to be observed in the presence of the substrate. Indeed, the monomer-dimer kinetic studies performed in Figure 5.4 were performed at a fixed concentration of substrate at 2 μM . In this experiment, the K_d value for the MERS-CoV 3CL^{pro} dimer was determined to be 7.8 μM , which is lower than the K_d value determined in the absence of substrate using AUC, thereby supporting substrate-induced dimerization. Given the high K_M of 3CL^{pro} for the peptide substrate [127–130], even higher substrate concentrations would be required to observe substrate activation in a plot of catalytic activity versus substrate concentration. However, the use of the FRET-based substrate is limited to only low substrate concentrations due to a significant inner filter effect at higher concentrations of substrate. Therefore, a compound that both mimics substrate and has higher binding affinity can act as a useful surrogate for the substrate, allowing the observation of ligand-induced dimerization and activation even at low substrate concentrations.

5.4.8 X-ray structure of MERS-CoV 3CL^{pro}-compound **6** complex

To gain atomic level detail and molecular insight into the mechanism for substrate-induced dimerization of MERS-CoV 3CL^{pro}, crystallization and X-ray structure determination of the unliganded MERS-CoV 3CL^{pro} monomer and the MERS-CoV 3CL^{pro} covalently modified with compound **6** were attempted. Unfortunately, all the attempts to crystallize the unliganded MERS-CoV 3CL^{pro} monomer failed after multiple attempts, but we were able to crystallize and determine the X-ray structure of MERS-CoV 3CL^{pro} in complex with compound **6** to a resolution of 1.6 Å. The statistics for X-ray data collection, processing and refinement are summarized in Table 5.4. The MERS-CoV 3CL^{pro}-**6** complex crystallized as a biologically relevant,

symmetrical dimer in space group $C2$ with one monomer in the asymmetric unit. Electron density for the entire protein was clearly visible and strong electron density ($F_o - F_c > 4\sigma$) was present for compound **6** within the active site (Figure 5.9.A).

Table 5.4.
X-ray data collection and refinement statistics

Beamline:	MERS-CoV 3CL ^{PRO} - 6 LRL-CAT Sector 31 ID-D	MERS-CoV 3CL ^{PRO} - 11 LS-CAT Sector 21 ID-G
Data collection		
Wavelength (Å)	0.9793	0.9786
Resolution range (Å)	19.35-1.62 (1.68-1.62) ^a	50.00-2.10 (2.14-2.10) ^a
Protein monomers in asymmetric units	1	4
Space group	$C2$	$P2_1$
Unit cell dimensions		
a, b, c (Å)	106.49, 57.31, 48.88	63.44, 114.93, 92.34
α, β, γ (°)	90, 112.78, 90	90, 90.89, 90
Total number of reflections	63855	816216
Number of unique reflections	32851	76865
Multiplicity	1.9 (1.9) ^a	2.2 (2.2) ^a
Completeness (%)	95.0 (93.8) ^a	96.8 (93.8) ^a
Mean $I/\sigma I$	5.2 (1.3) ^a	11.17 (1.83) ^a
R_{merge} (%) ^b	8.3 (67.2) ^a	8.8 (58.6) ^a
Refinement		
Resolution range (Å)	19.35-1.62	42.59-2.10
Number of reflections in working set	30824	76623
Number of reflections in test set	2026	2019
R_{work} (%) ^c	17.8	15.91
R_{free} (%) ^c	21.7	21.51
Number of non-hydrogen atoms		
Protein / water	2380 / 208	9383 / 995
RMSD - bond lengths (Å)	0.007	0.013
RMSD - bond angles (°)	1.09	1.35
Ramachandran favored (%)	99	98
Ramachandran outliers (%)	0	0
Molprobrity clash score	3.3	1.94
Average B-factor (Å ²)	20.4	33.1
Protein	19.8	32.5
Ligands	16.6	41.1
Solvent	27.7	37.9
^a Values in parentheses are for highest-resolution shell.		
^b $R_{\text{merge}} = \sum_h \sum_i I_i(h) - \langle I(h) \rangle / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the i^{th} measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I(h)$.		
^c R_{work} and $R_{\text{free}} = h(F(h)_o - F(h)_c) / h F(h)_o $ for reflections in the working and test sets, respectively.		

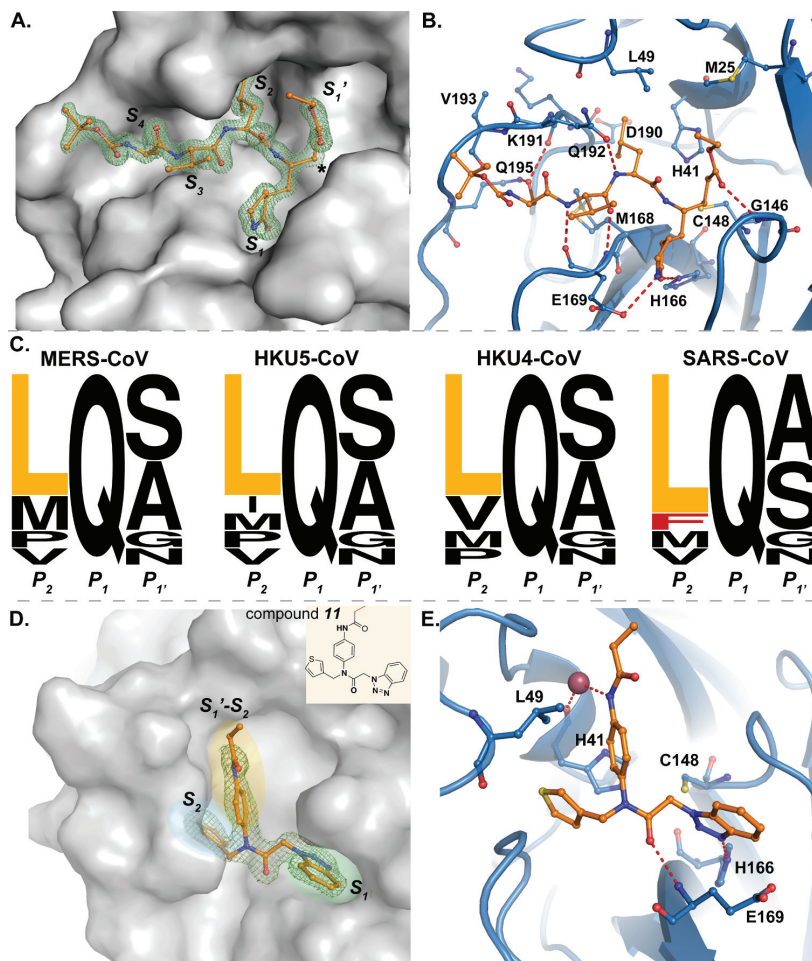


Fig. 5.9.: X-ray crystal structure of MERS-CoV 3CL^{pro} in complex with inhibitors. **A.** Solvent-accessible surface (grey shaded surface) of MERS-CoV 3CL^{pro}-compound **6** complex. Compound **6** is displayed in ball and stick model. Electron density associated with compound **6** is shown as a $F_o - F_c$ electron density difference map contoured to 3σ (green mesh). Substrate binding pockets S_4 - S_1' are labeled, where * indicates the electrophilic carbon of compound **6** that forms a C-S covalent bond with the active site cysteine C148. **B.** MERS-CoV 3CL^{pro}-compound **6** complex with the MERS-CoV 3CL^{pro} backbone represented as a ribbon model and relevant amino acids that interact with compound **6** represented as ball and sticks. Hydrogen bonds are depicted as red dashed lines. **C.** Sequence logos showing amino acid conservation for the eleven polyprotein cleavage sites of different 3CL^{pro} enzymes (MERS-CoV, HKU5-CoV, HKU4-CoV and SARS-CoV), generated using the WebLogo server [133]. Residues P_2 - P_1' are shown. Height of each letter corresponds to the amino acid conservation at that position. **D.** Solvent-accessible surface (grey shaded surface) of MERS-CoV 3CL^{pro}-compound **11** complex. Compound **11** is displayed in ball and stick model. Electron density associated with compound **11** is shown as a $2F_o - F_c$ electron density difference map contoured to 1.5σ (green mesh). Functional groups of compound **11** with their corresponding binding pockets are highlighted in yellow, green and blue ellipses. Chemical structure of compound **11** is shown in inset. **E.** Interactions between MERS-CoV 3CL^{pro} and compound **11** are illustrated.

5.4.9 MERS-CoV 3CL^{pro} has a smaller S_2 pocket than SARS-CoV 3CL^{pro}

The active site of MERS-CoV 3CL^{pro} bound with compound **6** is shown in Figures 5.9.A and 5.9.B. Compound **6** is covalently bound to the active site cysteine (C148) via a 1.8 Å bond between the gamma-sulfur and the electrophilic β -carbon of the Michael acceptor. The P_1 '-ethyl ester carbonyl, which mimics the carbonyl of the scissile bond in a substrate, forms a hydrogen bond with the backbone NH of G146 that forms part of the oxyanion hole (Figure 5.9.B). Within the S_1 subsite, the P_1 -lactam carbonyl, which is a surrogate for the amide of P_1 -glutamine of substrates, participates in a hydrogen bonding interaction with the imidazole ring of H166, and the P_1 -lactam NH forms a hydrogen bond with the carboxylate oxygen of E169. The P_2 -backbone amide NH forms a hydrogen bond with the side chain carbonyl of Q192 (Figure 5.9.B). The P_2 -leucine side chain atoms of the inhibitor make hydrophobic contacts with the side chains of M168 and L49 that line the S_2 subsite pocket. Moreover, compared to the equivalent residue T25 in SARS-CoV 3CL^{pro}, M25 in the S_2 pocket of MERS-CoV 3CL^{pro} is expected to reduce the size of the hydrophobic pocket, which is supported by our observed SAR described above.

The smaller size of the S_2 pocket in MERS-CoV 3CL^{pro} is also consistent with the preference for a smaller leucine residue at the P_2 position of cleavage sites instead of a bulkier phenylalanine or methionine residue. Indeed, analysis of the preference for leucine or phenylalanine at the P_2 position for the eleven 3CL^{pro} cleavage sites within the polyprotein of MERS-CoV shows that none of the eleven cleavage sites contain a phenylalanine residue at this position (Figure 5.9.C). Leucine is the predominantly favored residue at this position followed by methionine. Analysis of the cleavage sites from SARS-CoV, HKU4-CoV and HKU5-CoV shows that none of the eleven cleavage sites from group 2c members (MERS-CoV, HKU4-CoV and HKU5-CoV) contain a phenylalanine residue at the P_2 position; however, the SARS-CoV nsp5|nsp6 cleavage site contains a phenylalanine residue at this position.

Other interactions are also observed to play a significant role in stabilizing the MERS-CoV 3CL^{pro}-compound **6** complex. The P_3 -carbonyl and P_3 -NH participate in hydrogen bonding interactions with the backbone NH and carbonyl of E169. The P_4 -serine side chain is within hydrogen bonding distance of the side chain carboxamide of Q195 and the backbone carbonyl of K191.

5.4.10 X-ray structure of MERS-CoV 3CL^{pro} in complex with a non-covalent inhibitor

High-resolution diffraction quality crystals of MERS-CoV 3CL^{pro} were obtained in complex with compound **11**, which has an almost identical chemical structure as that of compound **10** (Figure 5.9.D). Our group previously showed that compounds similar to **10** and **11** act as potent non-covalent inhibitors of 3CL^{pro} from SARS-CoV [39]. The X-ray structure of compound **11** bound to MERS-CoV 3CL^{pro} was determined to a resolution of 2.1 Å and the X-ray data collection, processing and refinement statistics are summarized in Table 5.4. The MERS-CoV 3CL^{pro} **11** complex crystallized in space group $P21$ with two biologically relevant dimers in the asymmetric unit. The overall RMSD between the C-alpha atoms of the four chains was less than 1 Å, with the highest C-alpha RMSD of 0.719 Å between chains C and D. Strong electron density ($F_o - F_c > 4\sigma$) was present for compound **11** within all the 4 active sites of the two dimers (Figure 5.9.D).

The binding orientation for compound **11** in the active site of MERS-CoV 3CL^{pro} is similar to the binding orientation of related compounds in the active site of SARS-CoV 3CL^{pro} (PDB ID: 4MDS). The benzotriazole group binds in the S_1 subsite, phenyl propionamidyl occupies the S_1' - S_2 subsite and the thiophene group binds in the S_2 subsite. Compound **11** also forms two direct and one water-mediated hydrogen bond interactions with amino acids in the MERS-CoV 3CL^{pro} active site (Figure 5.9.E). The N-(3) of the benzotriazole ring forms a hydrogen bond with the side chain ϵ -nitrogen of conserved H166, and the central acetamide oxygen forms a hydrogen

bond with the backbone NH of conserved E169. The NH of phenyl propionamidyl group interacts with backbone carbonyl oxygen of the catalytic H41 residue through a water-mediated hydrogen bond and the imidazole ring of H41 engages with the phenyl ring of phenyl propionamidyl group through T-shaped π -stacking. The phenyl ring also form hydrophobic contacts with L49.

5.4.11 Interactions at the 3CL^{pro} dimer interface

Analysis of the MERS-CoV 3CL^{pro}-**6** and MERS-CoV 3CL^{pro}-**11** crystal structures reveal key differences between the dimer interface of MERS-CoV and SARS-CoV 3CL^{pro} (PDB ID: 2ALV) [53] (Figure 5.10). Two arginine residues, R4 and R298 (Figures 5.10.A-5.10.C), form some of the key interactions at the dimer interface of SARS-CoV 3CL^{pro}, and mutation of either of these amino acids results in a drastic loss of dimerization in SARS-CoV 3CL^{pro} [46, 47]. Interestingly, these two arginine residues (R4 and R298) are substituted in MERS-CoV 3CL^{pro} by two hydrophobic residues (V4 and M298) that are unable to participate in the formation of hydrogen bonds or salt-bridges. Therefore, it was initially thought that the loss of these key interactions might simply explain the >100,000-fold weaker dimerization observed for MERS-CoV 3CL^{pro} compared to SARS-CoV 3CL^{pro}. Surprisingly though, structural analysis of the dimer interface from the available X-ray structure of HKU4-CoV 3CL^{pro} (PDB ID 2YNB, unpublished; Figures 5.10.B and 5.10.C), and primary sequence alignment of 3CL^{pro} from MERS-CoV, HKU5-CoV, HKU4-CoV and SARS-CoV (Figure 5.11) revealed that V4 and M298 are conserved between all the β -CoV 2c members studied here. Substantial differences between the ability of MERS-CoV 3CL^{pro} and HKU4/HKU5-CoV 3CL^{pro} to dimerize, despite their high sequence identity, led us to the hypothesis that non-conserved residues between 3CL^{pro} enzymes from MERS-CoV and other β -CoV 2c members that are remote from the dimer interface may play a significant role in dimer formation.

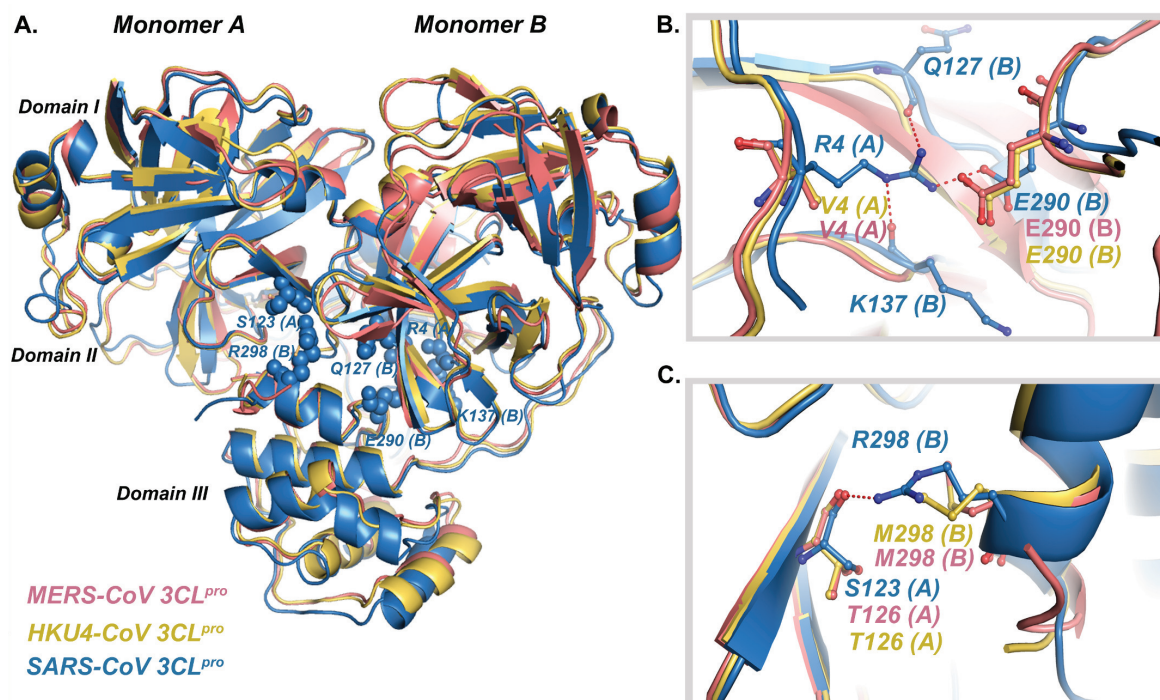


Fig. 5.10.: Comparison of X-ray crystal structures of 3CL^{pro} dimers from MERS-CoV, HKU4-CoV and SARS-CoV. **A.** Superposition of dimers of MERS-CoV 3CL^{pro} (pink color), HKU4-CoV 3CL^{pro} (yellow color, PDB ID: 2YNB) and SARS-CoV 3CL^{pro} (blue color, PDB ID: 2ALV). For SARS-CoV 3CL^{pro}, residues R4 and S123 from monomer A, and residues Q127, K137, E290 and M298 from monomer B are represented as spheres. **B.** For SARS-CoV 3CL^{pro}, interactions between the side chain of R4 from monomer A and Q127, E290 and K137 residues from monomer B are shown. The corresponding residues in MERS-CoV 3CL^{pro} and HKU4-CoV 3CL^{pro} are V4 in monomer A and E290 in monomer B, which do not interact at the dimer interface. **C.** For SARS-CoV 3CL^{pro}, S123 from monomer A engages in hydrogen bonding with R298 from monomer B across the dimer interface. The corresponding residue in monomer B of MERS-CoV 3CL^{pro} and HKU4-CoV 3CL^{pro} is M298, which does not participate in any interaction with T126 from monomer A across the dimer interface.

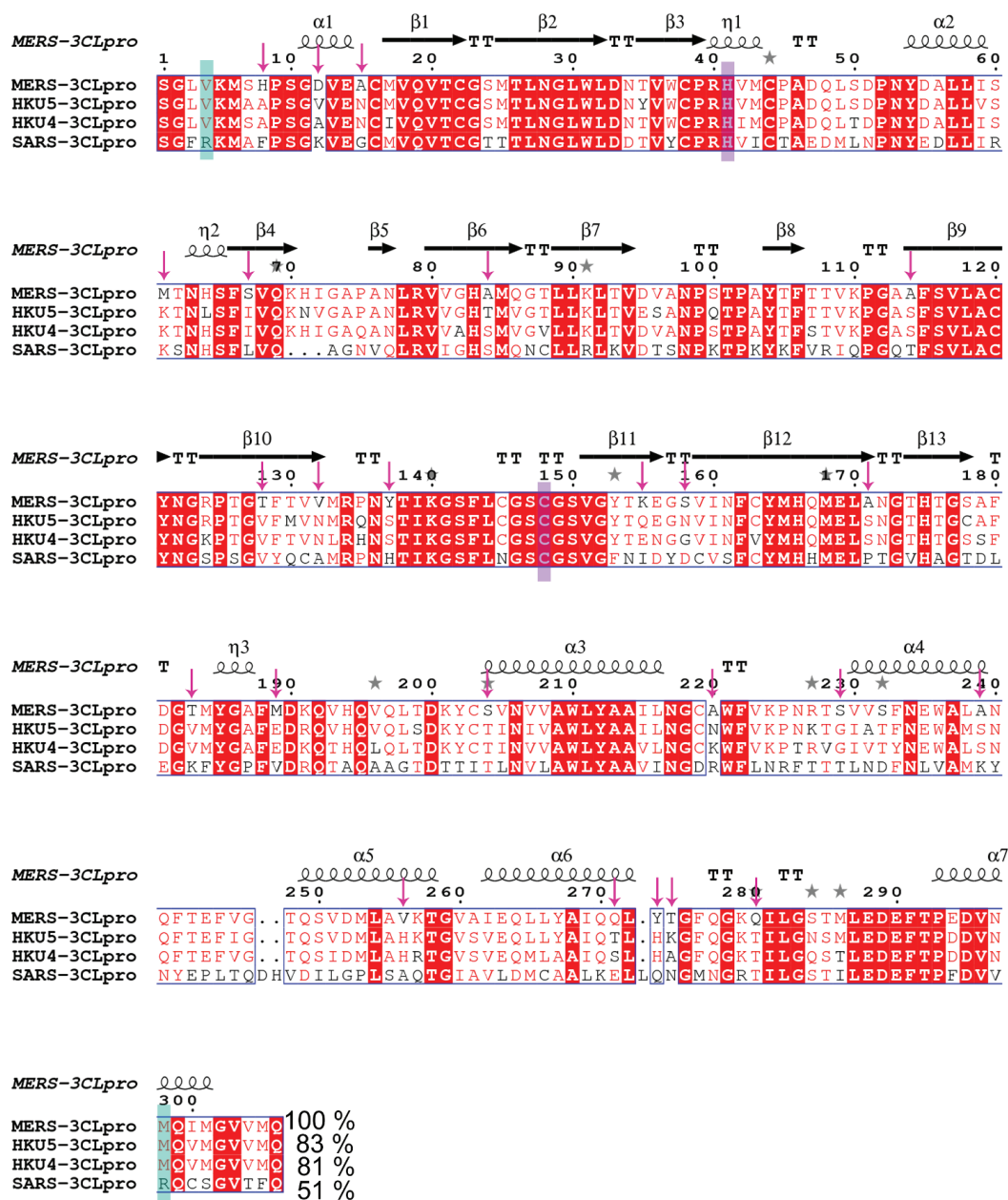


Fig. 5.11.: Sequence alignment of 3CL^{pro} enzymes from MERS-CoV, HKU5-CoV, HKU4-CoV and SARS-CoV. Programs MultAlin [134] and ESPript [135] were used for the sequence alignment and visualization. Secondary structural elements of MERS-CoV 3CL^{pro} are represented as spirals for alpha-helix, arrows for beta-strands, η for 3_{10} helix and T for beta-turns. Residues V4 and M298 in MERS-CoV, HKU5-CoV, HKU4-CoV 3CL^{pro}, and R4 and R298 in SARS-CoV are shown in green box; catalytic residues H41 and C148 are highlighted in purple box. The non-conserved residues of MERS-CoV 3CL^{pro} are marked with pink arrows. % Identity with MERS-CoV 3CL^{pro} is shown.

5.4.12 Analysis of non-conserved residues of MERS-CoV 3CL^{pro}

Analysis of our current crystal structures does not reveal a clear mechanism for the monomer to dimer switch of MERS-CoV 3CL^{pro} upon ligand binding. Therefore, attempts to identify the non-conserved residues in MERS-CoV 3CL^{pro} that might affect enzymatic activity due to their proximity to key residues involved in substrate binding and/or dimer formation were made next.

Based on a sequence alignment, MERS-CoV 3CL^{pro} contains ~ 24 non-conserved amino acids (pink arrows in Figure 5.11). Upon analyzing the position of these amino acids in the crystal structure, it was observed that a remarkable number of these amino acids are present in the loop regions. Figure 5.12.A illustrates the non-conserved residues present in the loop regions as grey (monomer A) and pink (monomer B). Interestingly, it was also observed that there are hotspots in the protein structure where most of these amino acids are clustered. These hotspots include the N-terminal region, the active site region, the inter-domain loop (loop between the catalytic fold and domain III) and the domain III. In MERS-CoV 3CL^{pro}, non-conserved amino acid H8, which forms van der Waals contacts with K155 of the same monomer and T128 of the other monomer, is present at the end of the N-terminal finger (Figures 5.12.B and 5.12.C), while amino acids D12 and A15 are part of the N-terminal helix (Figure 5.12.B). Additionally, amino acids T128, K155 and S158 are present within 6 Å of the N-terminal region (Figure 5.12.B). Substitution to these amino acids in MERS-CoV 3CL^{pro} might have changed the protein dynamics in a way that only ligand binding populates the monomer conformation, which is more amenable to dimer formation.

It was also observed that some of the non-conserved residues in MERS-CoV 3CL^{pro} are located in proximity to the substrate-binding site and might contribute towards ligand-induced dynamic changes favorable for dimer formation. For example, non-conserved amino acid M61 forms hydrophobic interactions with M43, which in turn is in close proximity to the catalytic residue H41 (Figure 5.12.D). Residue A171 is present on a loop and this loop, along with conserved residues H166 and H175, form

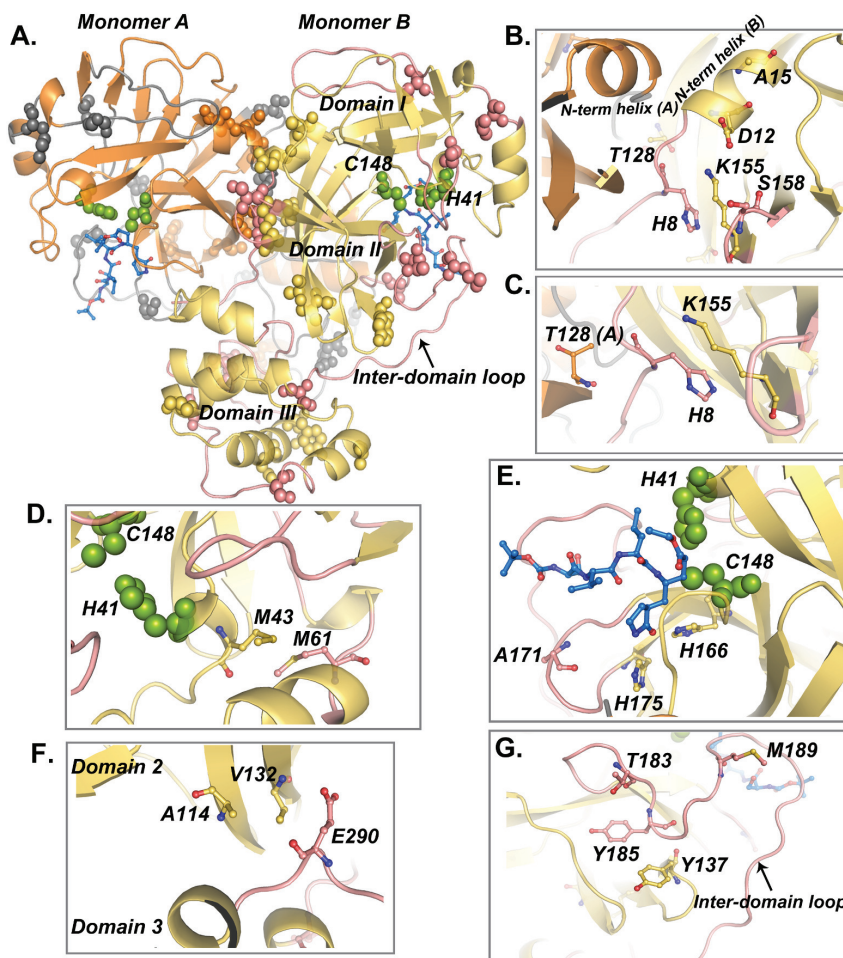


Fig. 5.12.: **Analysis of the non-conserved amino acids of MERS-CoV 3CL^{pro}.** **A.** Cartoon illustration of MERS-CoV 3CL^{pro} dimer with monomers A and B colored in orange and yellow, respectively. Non-conserved residues that are present in the loop regions are shown as spheres in grey and pink for monomers A and B, respectively. Other non-conserved residues are represented as spheres with the corresponding chain color. Domains I, II, III, and the inter-domain loop are labeled. Catalytic residues H41 and C148 are shown as green spheres. Inhibitor molecule is shown in both active sites in blue sticks. **B-G.** Residues of Monomer B are shown (yellow and pink), unless otherwise labeled. **B.** Clustering of some of the non-conserved amino acids, H8, D12, A15, T128, K155 and S158, near the N-terminal region is shown. N-terminal helices for both monomers are labeled. **C.** H8 from the N-terminal region forms van der Waals contacts with K155 of the same monomer and T128 of the other monomer in the dimer. **D.** Non-conserved residue M61 forms hydrophobic contacts with M43 residue, which is in close proximity to catalytic residue H41. **E.** Loop containing the non-conserved residue A171 forms the S_1 pocket along with residues H166 and H175. **F.** V132 forms hydrophobic contacts with a residue within the same domain (A114), as well as E290 from domain III. **G.** Non-conserved residue Y137 makes hydrophobic contacts with Y185; Y185 along with two other non-conserved residues T183 and M189 are present on the inter-domain loop.

the S_1 subsite for binding the P_1 amino acid of the substrate (Figure 5.12.E). In addition to its influence on substrate binding, A171 may also contribute towards dimer formation upon substrate binding due to its close proximity with E169. This glutamate residue in SARS-CoV 3CL^{pro} (E166) has been established as a key residue linking substrate-binding site to dimer interface [132]. V132 forms hydrophobic interaction with other non-conserved residue A114 within domain II (Figure 5.12.F). Additionally, V132 is present within van der Waals contact distance of E290 from extra-helical domain III (Figure 5.12.F). It is noteworthy that E290 forms a salt-bridge with R4 across the dimer interface in SARS-CoV 3CL^{pro}. However, this interaction is not formed in MERS-CoV 3CL^{pro} due to the substitution of R4 with V4. Y137 forms hydrophobic contacts with the conserved residue Y185 (Figure 5.12.G).

Besides amino acid V132 that connects domains II and III, residue Y185, along with two other non-conserved residues, T183 and M189, is present on the inter-domain loop that connects the catalytic fold (domains I and II) with the extra helical domain III (Figure 5.12.G). Flexibility within these residues might affect the orientation of domain III required for dimer formation.

5.5 Discussion

5.5.1 A Model for regulation of the enzymatic activity of MERS-CoV 3CL^{pro} during polyprotein processing

Enzymatic activity of coronavirus 3CL^{pro} is required for the processing of viral polyproteins at eleven distinct cleavage sites, allowing the release of non-structural proteins that subsequently form a replication complex for virus genome replication. Due to its indispensable role in the virus life cycle, regulation of the enzymatic activity of 3CL^{pro} is instrumental for efficient replication of coronaviruses. Based on the experimental results discussed in this dissertation, a model to explain the mechanism for regulating the enzymatic activity of MERS-CoV 3CL^{pro} in the context of polyprotein processing during virus infection is proposed (Figure 5.13).

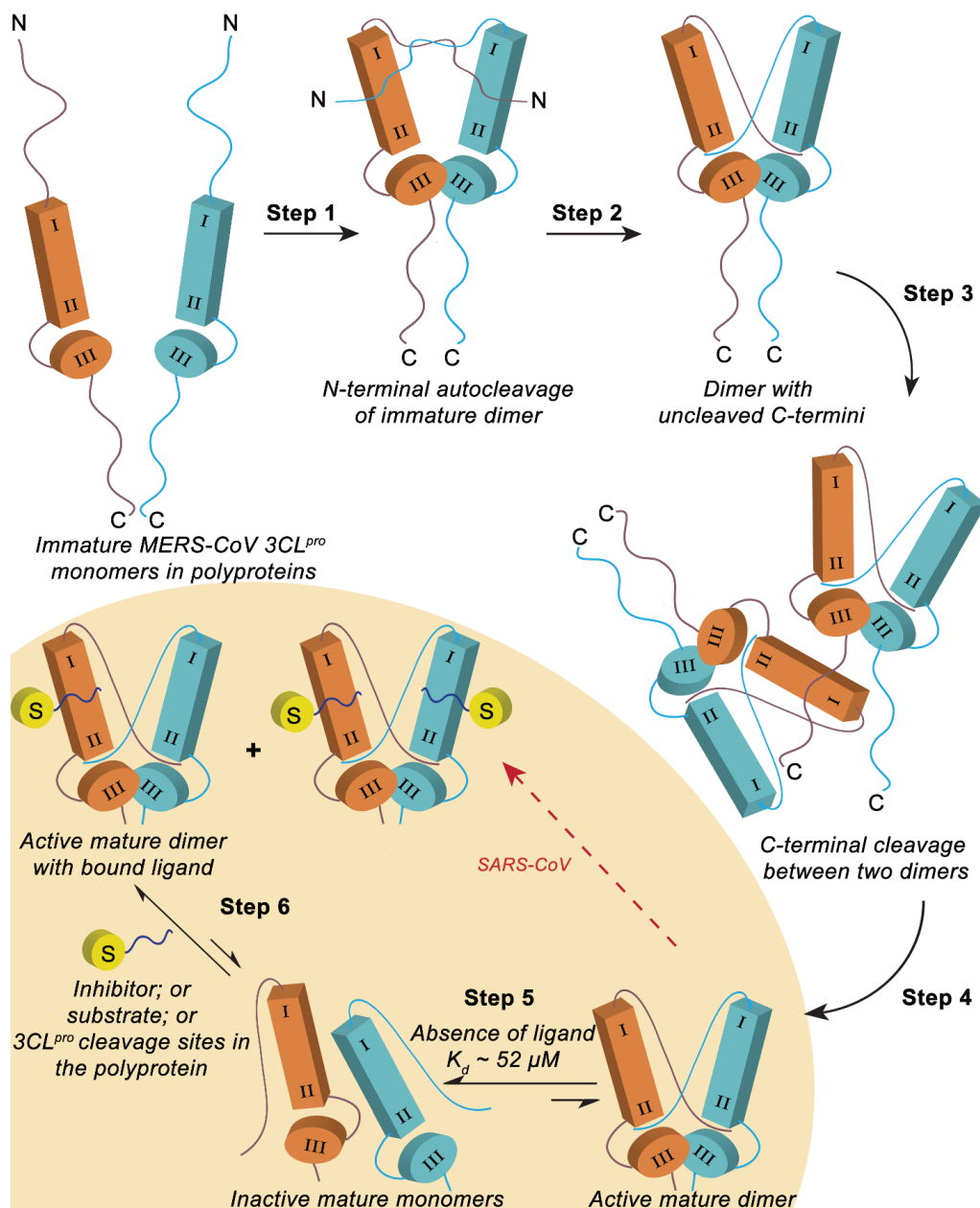


Fig. 5.13.: **Proposed model for polyprotein processing in MERS-CoV regulated by ligand-induced dimerization of MERS-CoV 3CL^{pro}.** MERS-CoV 3CL^{pro} domains I and II are together represented as the rectangular box, and domain III is represented as a cylinder. The N- and C-termini are labeled, and the yellow cylinder labeled 'S' represents a ligand that can be a peptide inhibitor, peptide substrate, or 3CL^{pro} cleavage sites in the polyprotein. Various steps required for the auto-release of 3CL^{pro} from the polyprotein and subsequent processing of the polyprotein cleavage sites are described in the main manuscript. Suggested by our AUC and kinetic studies, shaded region (Steps 5 and 6) highlights the additional steps MERS-CoV 3CL^{pro} would undertake during polyprotein processing and have been described in the kinetic model depicted in Figure 5.8.B.

A number of *in Vitro* studies performed on SARS-CoV 3CL^{pro} have established the mechanism for 3CL^{pro} auto-release from the polyprotein [43, 57, 58]. Based upon these studies and our data on MERS-CoV 3^{pro}, the polyprotein processing model for MERS-CoV 3^{pro} is proposed in Figure 5.13. The steps proposed for auto-release of MERS-CoV 3^{pro} from the polyprotein (Steps 1 to 4, Figure 5.13) have been adapted from Chen et al. [57], where it is suggested that N-terminal auto-processing does not require the formation of a mature 3CL^{pro} dimer for SARS-CoV. Based on the differences between the properties of SARS-CoV 3CL^{pro} and MERS-CoV 3CL^{pro}, as highlighted in our studies, two additional steps (Steps 5 and 6, Figure 5.13) that MERS-CoV 3CL^{pro} may need to utilize for efficient polyprotein processing were added. In Step 1, two immature MERS-CoV 3CL^{pro} monomers in the polyprotein approach each other and form an immature dimer via interactions between domain III, which allows each of the monomers to insert their N-termini into the active site of the other monomer. In Step 2, the N-termini are cleaved and the dimer with uncleaved C-termini adopts a conformation similar to the mature dimer. Our observation of auto-cleavage of the N-terminal (His)₆-tag from MERS-CoV 3CL^{pro} during expression in bacterial cells supports Steps 1 and 2, where formation of an immature dimer capable of auto-processing the N-terminus occurs. In Step 3, two dimers with uncleaved C-termini approach each other, followed by insertion of the C-terminus from one dimer into one of the active sites of the other dimer. In Step 4, the C-termini are cleaved and mature dimer is released from the polyprotein.

For SARS-CoV, the 3CL^{pro} dimer formed in Step 4 continues to process cleavage sites in the polyprotein, effectively skipping Steps 5 and 6 (red arrow in Figure 5.13) since the dimer is tightly associated. However, the high K_d value of MERS-CoV 3CL^{pro} dimer suggests that the active and mature dimer may dissociate into inactive, mature monomers in the absence of any ligand (Step 5). In order for polyprotein processing to proceed, another step (Step 6) must occur. In Step 6, a substrate ‘S’, e.g. one of the eleven polyprotein cleavage sites, would induce dimer formation and, hence, activate catalysis and cleavage at the substrate recognition sites. Our AUC

results and the kinetic activation studies performed in the absence and presence of inhibitors support Steps 5 and 6 where the inactive but mature monomers require binding of a ligand to undergo ligand-induced dimerization and formation of an active, mature dimer that can then process the polyprotein cleavage sites.

5.5.2 Non-conserved amino acids of MERS-CoV 3CL^{pro} may regulate the dimer formation

Long-range interactions have been reported to modulate dimerization and activity of 3CL^{pro} enzymes. Barrila et al. demonstrated that mutation of a conserved amino acid S147, which is distant from the dimer interface, results in a total loss of dimerization and enzymatic activity of SARS-CoV 3CL^{pro} [87]. Although S147 does not form direct interactions at the dimer interface, disruption of the dimer upon mutation stems from the fact that S147 makes several interactions with other residues involved in forming a hydrogen-bonding network within SARS-CoV 3CL^{pro}. Site-directed mutagenesis studies on domain III of SARS-CoV 3CL^{pro}, where N214A and S284-T285-I286/A mutants were characterized, revealed that in spite of being present on an entirely different domain, these residues affect catalysis through a network of residues undergoing correlated motions across the entire protease [88, 89]. Utilizing 3CL^{pro} temperature sensitive mutants of MHV, Stobart et al. have also demonstrated that second-site mutation physically distant from the temperature sensitive mutation suppresses the temperature sensitive phenotype through long-range interactions, thereby regulating 3CL^{pro} enzymatic activity during polyprotein processing and virus replication [91].

Our studies also suggest that long-range interactions among the non-conserved residues can significantly alter the properties of MERS-CoV 3CL^{pro}. A detailed analysis of non-conserved residues of MERS-CoV 3CL^{pro} among β -CoV 2c members identified hotspots, including the N-terminal finger and helix, the active site region, the inter-domain loop and the domain III, where these residues are clustered. Several

studies done on SARS-CoV 3CL^{pro} have demonstrated that amino acids from the N-terminal finger, the N-terminal helix and domain III significantly contribute towards dimer formation.

In addition to the direct interactions at the dimer interface, correct orientation between the catalytic fold and domain III is also crucial for dimer formation. Wu et al. showed that the most dramatic difference between the crystal structures of monomer and ligand-bound dimer of R298A mutant of SARS-CoV 3CL^{pro} was a 33° rotation of domain III [47, 48]. This rotation results in a steric clash between domain III from two monomers and would essentially block dimer formation. However, upon addition of a ligand, domain III of the R298A mutant adopts the correct orientation and results in the formation of a dimer structure. Similar to the SARS-CoV 3CL^{pro} R298A mutant, ligand binding into the active site of MERS-CoV 3CL^{pro} monomer possibly stabilizes the inter-domain loop conformation that maintains domain III in the correct orientation for dimer formation. Most of the non-conserved residues within domain III are present on the surface, and also are distant from the dimer interface. These residues may be involved in providing the flexibility required for conformational changes during the monomer to dimer switch.

Several amino acids have been identified in MERS-CoV 3CL^{pro} that may contribute to the dimer formation upon ligand binding. However, single amino acid mutagenesis alone is unlikely to reveal significant differences in the dimerization properties. As demonstrated by Myers et al. [101] for ornithine decarboxylase, response of single amino acid to ligand binding may be limited to only local conformational changes and may not have significant contribution towards dimer stability. However, local conformational changes in a network of residues may propagate larger effects that stabilize dimer formation upon ligand binding. Analysis of the non-conserved residues of MERS-CoV 3CL^{pro} discussed here sets forth a framework to perform systematic single or multiple mutagenesis studies to gain insights into the mechanism for ligand-induced dimerization of the enzyme.

5.5.3 Development of 3CL^{pro} inhibitors with broad-spectrum specificity

Insights into the mechanistic and structural similarities as well as differences between 3CL^{pro} enzymes from different coronavirus subgroups are instrumental for the development of 3CL^{pro} inhibitors with broad-spectrum specificity. To evaluate the broad-spectrum specificity of our peptidomimetic compounds, their inhibitory activity against 3CL^{pro} from MERS-CoV, SARS-CoV, HKU5-CoV and HKU4-CoV was determined. The inhibitory data and K_i values clearly show that compounds **6** to **9** inhibit all the 3CL^{pro} enzymes tested here. The X-ray structure of MERS-CoV 3CL^{pro} in complex with compound **6** revealed that out of eight direct hydrogen bonds formed between compound **6** and MERS-CoV 3CL^{pro}, four of these hydrogen bonds involve interactions with conserved structural elements of the peptide backbone of the enzyme. Furthermore, the amino acids that form hydrogen bonds with compound **6** through side chain interactions are conserved in all the coronavirus 3CL^{pro} evaluated here, as well as 3CL^{pro} from other β -coronaviruses like MHV, OC43 and HKU1. These results suggest that canonical structural features exist among the 3CL^{pro} enzymes that can be exploited for structure-based design of broad-spectrum inhibitors.

For the non-covalent inhibitor compound **11**, the X-ray structure reveals two direct hydrogen-bonding interactions between the compound and MERS-CoV 3CL^{pro}. One of the hydrogen bonds forms with the side chain ϵ -nitrogen of conserved H166, and the second involves the backbone NH of conserved E169. It is speculated that these interactions remain conserved in other 3CL^{pro} enzymes as well, since H166 and E169 amino acids are conserved in all 3CL^{pro} enzymes. In fact, the crystal structure of SARS-CoV 3CL^{pro} in complex with an inhibitor similar to compound **11** (PDB ID: 4MDS) reveals that the interactions of the inhibitor with the amino acids H166 and E169 are conserved.

The identification of 3CL^{pro}-inhibitor interactions utilizing conserved elements of the protein structure, including the peptide backbone and conserved side chains of

active site residues, suggests that the development of broad-spectrum inhibitors of coronavirus 3CL^{pro} is feasible.

Our studies here demonstrate the unique properties of MERS-CoV 3CL^{pro} among β -CoV 2c members, evident from the requirement for a ligand to induce dimerization. Although the peptidomimetic compounds containing a Michael acceptor group (for example, compounds **6** to **9**) induce dimer formation of MERS-CoV 3CL^{pro}, the irreversible nature of their reaction with the active site cysteine ensures complete inhibition of the enzyme at stoichiometric ratios in a time dependent manner. On the contrary, non-covalent peptidomimetic compounds (for example, compounds **10** and **11**) inhibit the enzymatic activity of MERS-CoV 3CL^{pro} only at high compound concentrations. Based on these observations, compounds that irreversibly modify the 3CL^{pro} active site may serve as better candidates for the development of inhibitors for MERS-CoV 3CL^{pro}.

5.5.4 Potential complexity in the development of MERS-CoV 3CL^{pro} inhibitors as antiviral agents

Induced dimerization of MERS-CoV 3CL^{pro}, as seen in the presence of peptidomimetic inhibitors, has significant implications in the development of antiviral agents targeting MERS-CoV 3CL^{pro}. As a consequence of enzyme activation, the development of an effective antiviral agent may necessitate the development of a compound that can inhibit the MERS-CoV 3CL^{pro} monomer and stabilize it without inducing dimerization and/or inhibit the active sites of the dimer at low doses, ensuring inactivation of both the active sites within the dimer. On the contrary, it is also possible that the presence of an inhibitor could enhance the activity of MERS-CoV 3CL^{pro} to an extent that results in a complete loss of the temporal and spatial regulation of the enzymatic activity, thereby disrupting viral genome replication. Ramifications of ligand-induced dimerization and activation of MERS-CoV

3CL^{pro}, as seen in the presence of lower concentrations of inhibitor, will need to be further explored in virus infected cells.

CHAPTER 6. CHARACTERIZATION OF DRUG-RESISTANT Y22C MUTANT OF HKU1 3CL^{PRO}

Parts of the data and text in this chapter have been included in a manuscript entitled “Chimeric murine hepatitis virus identifies nsp5 protease (3CL^{PRO}) inhibitors active against subgroup 2a beta coronaviruses HKU1 and OC43” and is currently under preparation in collaboration with the Denison group at Vanderbilt University.

Compound CE-5 (Figure 6.2) is an ester-based small molecule inhibitor of multiple 3CL^{PRO} enzymes [37, 59, 136]. CE-5 also displays antiviral activity and has been shown to efficiently inhibit replication of SARS-CoV, MHV, HKU5-CoV and MERS-CoV [37, 59, 136]. Our collaborators identified a mutation, Y22C, in HKU1 3CL^{PRO} that renders the virus harboring this mutation resistance to inhibitor CE-5 (unpublished data). It is interesting that emergence of mutation Y22C, which is structurally distant from the active site of HKU1 3CL^{PRO}, confers drug resistance to the enzyme. In order to explore the mechanism by which Y22C mutation confers drug resistance towards CE-5, enzyme inhibition and reactivation assays were performed using purified enzymes. Based on the molecular dynamics simulation that were performed on the wild-type and structural model of Y22C mutant enzymes, a potential mechanism for drug resistance for this mutant is proposed.

6.1 Material and Methods

6.1.1 *in vitro* enzymatic assay and IC₅₀ determination

Wild-type and Y22C drug-resistant mutant of HKU1 3CL^{PRO} were over-expressed in *E. coli* BL21-DE3 cells and purified using the purification protocol described in Chapter 2. Enzymatic efficiency (apparent $k_{\text{cat}}/K_{\text{M}}$) was determined at room-

temperature and 37 °C by measuring the rate of hydrolysis of the FRET-based peptide substrate as a function of substrate concentration varied in the range from 0 to 2 μM . Enzyme solution was prepared in assay buffer to a final concentration of 100 nM, and 5 \times stocks of varying substrate concentrations were incubated separately at either room temperature or 37 °C in the wells of a 96-well microtiter plate. After 15 minutes, the reaction was started by adding 20 μL of substrate to 80 μL of enzyme solution. Increase in fluorescence over time was then monitored in a BioTek Synergy H1 plate reader using an excitation wavelength of 485/20 nm and emission wavelength of 528/20 nm. Enzymatic activity was converted from units of RFU/s to $\mu\text{M}/\text{min}$ using experimentally determined value of fluorescence extinction coefficient. Apparent $k_{\text{cat}}/K_{\text{M}}$ values were determined by calculating slope of the linear plot between rate of enzymatic activity, normalized for total enzyme concentration, and substrate concentration i.e. $\text{Rate}/[\text{Enzyme}]$ versus $[\text{Substrate}]$.

For IC_{50} determination, 100 \times DMSO stocks of varying concentrations of compound CE-5 (2.5 μM to 5 nM) were diluted hundred-fold in 80 μL of the enzyme solution and incubated for 20 minutes at 37 °C. After 20 minutes, enzymatic activity was measured as initial slope of the progress curve, obtained by initiating the reaction with 20 μL of 10 μM substrate. Final enzyme and substrate concentrations in 100 μL reactions were 100 nM and 2 μM , respectively. Percent inhibition was plotted as a function of inhibitor concentration, and IC_{50} values were calculated by fitting the data to a dose-response curve using non-linear regression program GraphPad Prism.

6.1.2 Enzyme reactivation assay

To monitor complete hydrolysis of CE-5 from the active site and consequential reactivation of the enzyme, enzymatic activity of both the wild-type and Y22C mutant was measured after incubating in the presence of CE-5 over a period of 7 hours at 37 °C. 10 \times enzyme stocks were incubated with either 2 M equivalents of CE-5 or DMSO (uninhibited enzyme control) at 37 °C. Final DMSO concentration was 1%. 10 μL

aliquots were taken at specific time points post incubation and mixed with 70 μL of assay buffer. Reaction was initiated by the addition of 20 μL of 10 μM substrate. Final concentrations of enzyme, CE-5 and substrate in the reaction were 100 nM, 200 nM and 2 μM , respectively. Enzymatic activity was measured as initial slope of the progress curve for increase in fluorescence over time. Percent residual activity was calculated as the percent activity of uninhibited enzyme at the same time point and plotted as a function of time.

6.1.3 Molecular Dynamics Simulations and Analysis

To assess the influence of the Y22C mutation on HKU1 3CL^{pro}, molecular dynamics (MD) simulations were performed using the Desmond Molecular Dynamics tool [137] implemented in the Schrödinger software suite. All the molecular dynamics simulations and subsequent analyses were performed by Dr. Laura Kingsley (Mescar Lab). The crystal structure, PDB ID: 3D23 [86], of HKU1 3CL^{pro} was used for all simulations with the co-crystallized ligand removed. For the mutant form of the protein, the mutation was made using the Maestro [138] interface prior to protein preparation.

In each case, the protein was prepared using the Protein Preparation Wizard module in Schrödinger. In general, this process involves the assignment of bond orders, addition of hydrogens, rebuilding of missing atoms or residues, and capping terminal residues.

The systems were prepared for the Desmond MD run using the System Setup panel in Maestro [137]. The systems were solvated in a cubic box of TIP3P water molecules using a distance of 15 Å between the solute and edge of the solvent box. Ions were added to neutralize the overall charge of the system. Each system was minimized using the steepest descent method until a gradient threshold of 25kcal/mol/Å was reached. Prior to the production run, the simulation was relaxed using the default protocol in Desmond; 12 ps NVT ensemble simulation at 10K with the non-hydrogen

atoms restrained, followed by 12 ps NPT ensemble simulation at 10K and 1 atm with non-hydrogen atoms restrained, next the temperature was raised to 300K for a 12 ps NPT ensemble simulation with the non-hydrogen atoms restrained, and finally a 24 ps NPT ensemble with no restraints. An additional equilibration step of 1ns was also completed prior to the production run. A 10 ns production run was then completed at a temperature of 300K using a timestep of 2 fs. Long-range interactions were calculated using the Smooth Particle Mesh Ewald method and a cutoff of 9 Å.

The simulations were visualized using VMD, and the root mean squared deviation (RMSD) of all backbone atoms and the backbone atoms at the site of the mutation and immediately following (residues 22-30) were calculated using the Simulation Event Analysis panel in Schrödinger [137]. Distance measurements between the oxygen and nitrogen atoms in the backbones of residues 22 and 25, respectively were also completed using this module.

6.2 Results

6.2.1 Mechanism for drug-resistance of Y22C mutant of HKU1 3CL^{Pro}

To investigate the mechanism for drug-resistance of Y22C mutant, the wild-type HKU1 3CL^{Pro} and the HKU1-Y22C mutant enzyme were expressed and purified using the purification protocol described in Chapter 2. At room temperature, the wild-type and Y22C mutant enzymes were equally efficient at hydrolyzing the peptide substrate with apparent $k_{\text{cat}}/K_{\text{M}}$ values of $7.5 \pm 0.03 \times 10^{-2} \mu\text{M}^{-1} \text{min}^{-1}$ and $7.4 \pm 0.2 \times 10^{-2} \mu\text{M}^{-1} \text{min}^{-1}$, respectively. However, the enzymatic efficiency of Y22C (apparent $k_{\text{cat}}/K_{\text{M}} = 2.1 \pm 0.2 \times 10^{-2} \mu\text{M}^{-1} \text{min}^{-1}$) was noticeably lower than for the wild-type (apparent $k_{\text{cat}}/K_{\text{M}} = 10.3 \pm 0.2 \times 10^{-2} \mu\text{M}^{-1} \text{min}^{-1}$) at the physiological temperature of 37 °C. Significant difference in the enzymatic efficiencies of wild-type and Y22C mutant at 37 clearly indicates that the active sites of these two enzymes are distinct. Even subtle differences in the active site can manifest into altered interactions with CE-5 inhibitor, thereby, conferring drug resistance to Y22C mutant.

To test the mechanism of drug-resistance, the IC_{50} values of CE-5 for both the wild-type and Y22C mutant of HKU1 3CL^{pro} were determined at 37 °C (Figure 6.1). IC_{50} values of CE-5 for wild-type and Y22C mutant were 115 ± 5 nM and 153 ± 12 nM, respectively (Figure 6.1), a statistically significant ($P < 0.01$), increase of 33 % in the IC_{50} value for Y22C mutant over the wild-type enzyme. While this might contribute towards drug-resistance, this subtle difference could not completely explain the resistance phenotype. Since the IC_{50} values were determined following 20 minutes incubation of the enzymes with CE-5 inhibitor, difference in the IC_{50} values only report differences leading up to the formation of thioester intermediate in the ester hydrolysis reaction pathway (Figure 6.2).

Drug-resistance could also ensue from a faster release of the second hydrolysis product after formation of the thioester intermediate Figure 6.2, thereby regenerating the active enzyme. To probe this possibility, the residual activity of enzymes incubated in the presence of inhibitor CE-5 was measured in a time-dependent manner over a period of seven hours at 37 °C (Figure 6.1). The wild-type enzyme required more than 5 hours to restore 50 % activity. In contrast, the Y22C mutant 3CL^{pro} recovered 50 % of its activity in less than 3 hours. After seven hours, Y22C completely restores its activity, while only 60 % of the wild-type enzyme was reactivated. These studies clearly suggest that in addition to increase in IC_{50} value, faster restoration of the enzymatic activity via release of the second hydrolysis product is the likely mechanism for drug-resistance of Y22C mutant.

To test for correlates of molecular activity, molecular dynamics simulations were performed on the wild-type and Y22C mutant of HKU1 3CL^{pro}. The Y22C mutation occurs at the end of a β 1-strand and is immediately followed by a loop leading to a second β strand, which lies directly adjacent to the active site. In the wild-type simulations the tip of this loop is held together by a hydrogen bond between the nitrogen in the backbone of Y22 and the oxygen in the backbone of M25. Due to this hydrogen bond the distance between the backbones of these two residues at the tip of the loop is well maintained (Figure 6.3- inset). However, in the Y22C mutant

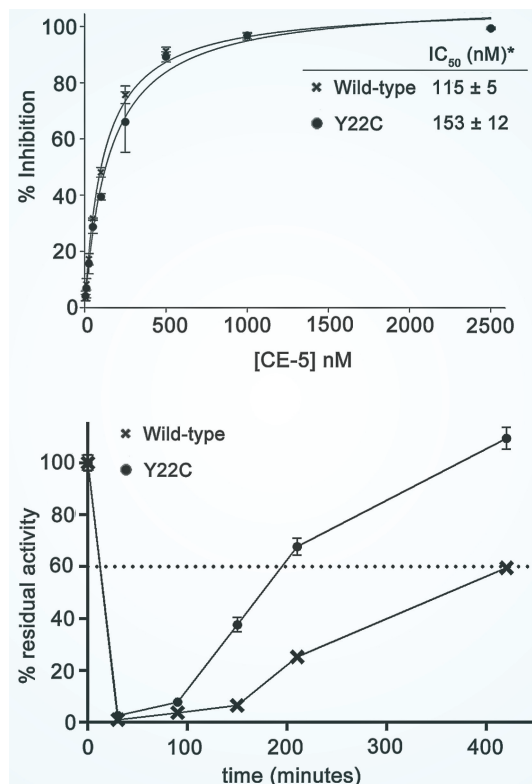


Fig. 6.1.: **Mechanism for drug-resistance of Y22C mutant of HKU1 3CL^{pro}.** *Top*- IC₅₀ values were determined from dose-dependent % inhibition of wild-type and Y22C mutant after 20 minutes of incubation with CE-5 at 37 °C. * indicates the difference in IC₅₀ values for Y22C and wild-type enzymes being statistical significant with a P-value <0.01. *Bottom*- Time dependent restoration of the enzymatic activity in the presence of CE-5, calculated as % residual activity of uninhibited enzyme control, is plotted for Y22C and wild-type enzymes. Restoration of enzymatic activity results from complete hydrolysis and release of CE-5 covalent intermediates from the active site.

enzyme, this hydrogen bond was found to be less stable and was found to break and eventually reform during the simulation (Figure 6.3).

The presence or absence of this hydrogen bond directly influences the flexibility in this region. When broken, the flexibility in this region increases (Figure 6.3-RMSD inset) and M25 was found to adopt an alternative conformation that was not observed in the wild-type simulations (Figure 6.3). This alternative conformation opens a pocket near the active site that would otherwise be occupied by the side chain of M25 (Figure 6.1, right panel- yellow dashed line).

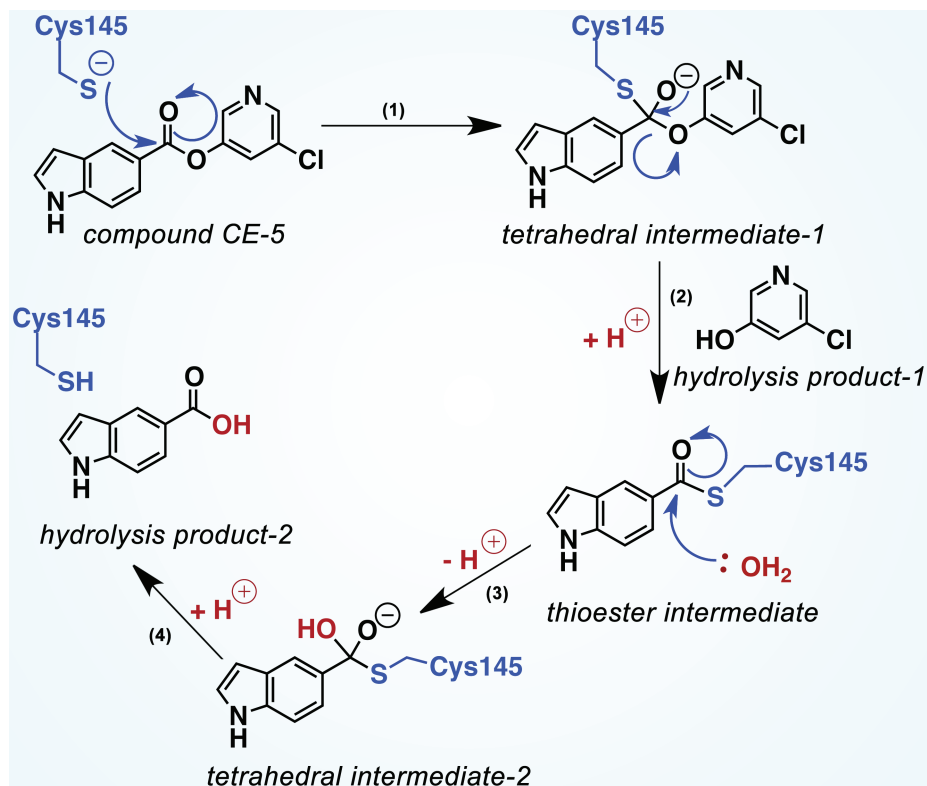


Fig. 6.2.: **Reaction mechanism for hydrolysis of the ester-based inhibitor CE-5 by HKU1 3CL^{pro}.** Catalytic cysteine (Cys145), compound CE-5 and reaction intermediates are illustrated. Reaction steps 1-4 are labeled as well.

6.3 Discussion

Our collaborators identified a mutation within HKU1 3CL^{pro} protease that conferred resistance to the CE-5 inhibitor in the context of virus replication. Ester based inhibitor CE-5 forms different covalent intermediates after reacting with the active site cysteine. Different mutations in 3CL^{pro} can confer drug-resistance through different mechanisms based upon the altered rate of formation of any of the covalent intermediates. Our proposed mechanism for drug-resistance of Y22C mutant of HKU1 3CL^{pro} is distinct from the drug-resistance mechanism of T26I/D65G mutant of MHV 3CL^{pro} reported by Deng et al [136]. For T26I/D65G MHV 3CL^{pro} mutant, an eight fold increase in IC₅₀ value over the wild-type enzyme was reported, suggesting that the likely mechanism for drug-resistance is lowered affinity of the active site for CE-5 and/or slower formation of the thioester intermediate. Moreover, a comparison of

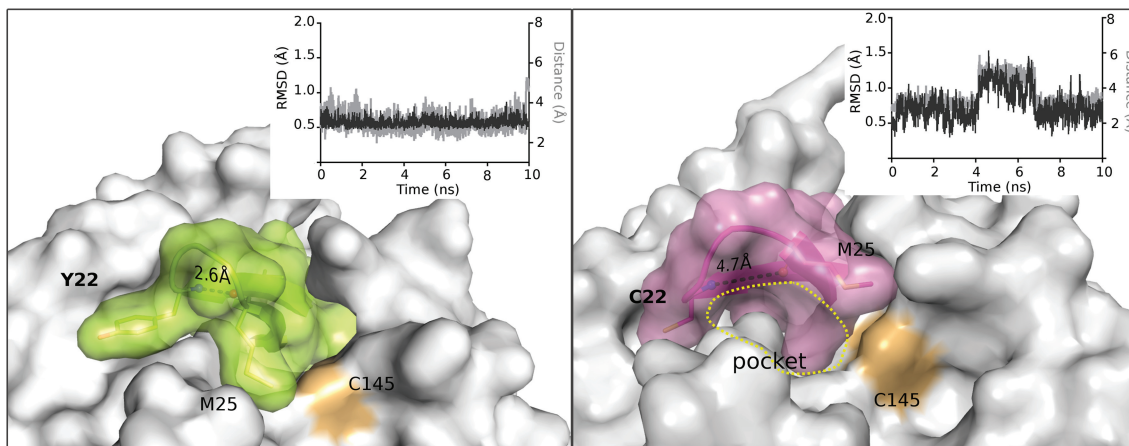


Fig. 6.3.: **Molecular dynamics simulation of HKU1 3CL^{Pro} wild-type and Y22C mutant.** *Left-* Active site of HKU1 3CL^{Pro} wild-type is shown with amino acid Y22, catalytic cysteine C145 and another amino acid M25. *Right-* Active site of HKU1 3CL^{Pro} Y22C mutant is shown with amino acid C22, catalytic cysteine C145 and another amino acid M25. *Inset-* RMSD and distance measurements between the backbone nitrogen and oxygen atoms of residues 22 and 25, respectively are shown.

restoration of the enzymatic activity after longer incubation period with the inhibitor showed contrasting results between the T26I/D65G and Y22C mutant. Release of the second hydrolysis product, measured as restoration of the enzymatic activity, was slower for T26I/D65G mutant compared to the wild-type MHV 3CL^{Pro} as opposed to a faster release that is observed for Y22C mutant of HKU1 3CL^{Pro} in our studies. To gain insight into the structural implications of the Y22C mutant, MD simulations of the wild-type and mutant forms of nsp5 were performed. Our simulations suggest that the Y22C mutation may cause intermittent disruptions of the hydrogen bond between the mutation site and M25 resulting in a local destabilization and “opening” at the tip of the 3CL^{Pro} β 1-sheet. When this hydrogen bond is broken, M25 is free to adopt an alternative conformation (Figure 6.3-right panel) which opens a pocket adjacent to the active site and exposes the active site to solvent. It is proposed that the intermittent opening of this pocket allows water to penetrate the active site more frequently than in the wild-type enzyme and increases the rate of inhibitor hydrolysis. This is in agreement with our kinetic findings that the rate of inhibitor binding is

approximately equivalent between the mutant and wild-type forms, but that the rate of hydrolysis is increased in the mutant form.

CHAPTER 7. SUMMARY

Coronavirus 3-Chymotrypsin-like protease (3CL^{pro}) is an attractive drug target for the development of antiviral therapeutics. High sequence conservation in the vicinity of active site among 3CL^{pro} proteases from different coronavirus subclasses make them an excellent target for the development of broad-spectrum antivirals. The overall goal of this project was to investigate enzymatic and structural properties of multiple 3CL^{pro} enzymes encompassing different coronavirus subclasses. Understanding the determinants of structural and functional disparity between different 3CL^{pro} enzymes and the factors regulating these properties will aid in the design of broad-spectrum inhibitors of 3CL^{pro} enzymes.

Here, we report the successful expression and purification of six different coronavirus 3CL^{pro} enzymes (SARS, MERS, OC43 HKU1, HKU5 and MHV). Since, the N- and C-termini of 3CL^{pro} play an important role in dimer formation, all the 3CL^{pro} enzymes discussed in this project were purified in their authentic form, without any N- or C-termini modification, using 2-4 sequential chromatographic steps.

Using a custom synthesized FRET-based peptide substrate, we determined the enzymatic efficiencies (apparent values of k_{cat}/K_M) for 3CL^{pro} enzymes. In our assay conditions, MHV 3CL^{pro} and HKU1 3CL^{pro} enzymes are the most efficient enzymes, while 3CL^{pro} from MERS-CoV is the least efficient enzyme.

Further investigations with MERS-CoV 3CL^{pro} reveal that this enzyme is a weakly associated dimer and primarily exists in a monomeric form in solution. These results suggest that lower enzymatic efficiency of MERS-CoV 3CL^{pro} stems from a weaker dimer formation. Biophysical and kinetic analysis performed in the presence of substrate-mimetic compounds further revealed that MERS-CoV 3CL^{pro} undergoes significant ligand-induced dimerization. We propose that ligand-induced dimerization can serve as a regulatory mechanism for 3CL^{pro} activation during polyprotein process-

ing in the context of virus replication. Since, structural and sequence analysis did not reveal any differences that would clearly suggest the structural basis for weaker dimer formation or ligand-induced dimerization of MERS-CoV 3CL^{pro}, we further propose long-range interactions between distant non-conserved residues as the regulator of dimerization and activity of MERS-CoV 3CL^{pro}.

Characterization of physiologically relevant temperature-sensitive (V148A) and compensatory (H134Y) mutants of MHV 3CL^{pro} further suggests that long-range interactions can also influence structural stability of 3CL^{pro} enzymes.

We also explored the mechanism for drug-resistant Y22C mutant of HKU1 3CL^{pro}. Y22 residue is distant from both the active site and dimer interface of HKU1 3CL^{pro}. Our studies suggest that distant mutation Y22C can directly influence substrate binding and catalysis, most-likely through long-range interactions.

Finally, despite the varying influences of long-range interactions on dimerization, stability and catalysis of 3CL^{pro} enzymes, we were able to identify several inhibitor molecules that can inhibit multiple 3CL^{pro} enzymes. Inhibitors identified in this study may serve as excellent scaffolds for the development of next-generation compound library with better potency and broad-specificity.

REFERENCES

REFERENCES

- [1] S. Perlman and J. Netland, "Coronaviruses post-sars: update on replication and pathogenesis," *Nature reviews. Microbiology*, vol. 7, no. 6, pp. 439–50, 2009.
- [2] R. L. Graham, E. F. Donaldson, and R. S. Baric, "A decade after sars: strategies for controlling emerging coronaviruses," *Nature Reviews Microbiology*, vol. 11, no. 12, pp. 836–848, 2013.
- [3] J. S. Peiris, "Severe acute respiratory syndrome (sars)," *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, vol. 28, no. 3, pp. 245–7, 2003.
- [4] J. S. Peiris, Y. Guan, and K. Y. Yuen, "Severe acute respiratory syndrome," *Nat Med*, vol. 10, no. 12 Suppl, pp. S88–97, 2004.
- [5] J. D. Cherry, "The chronology of the 2002-2003 sars mini pandemic," *Paediatr Respir Rev*, vol. 5, no. 4, pp. 262–9, 2004.
- [6] N. S. Zhong, B. J. Zheng, Y. M. Li, Poon, Z. H. Xie, K. H. Chan, P. H. Li, S. Y. Tan, Q. Chang, J. P. Xie, X. Q. Liu, J. Xu, D. X. Li, K. Y. Yuen, Peiris, and Y. Guan, "Epidemiology and cause of severe acute respiratory syndrome (sars) in guangdong, people's republic of china, in february, 2003," *Lancet*, vol. 362, no. 9393, pp. 1353–8, 2003.
- [7] S. E. Reed, "The behaviour of recent isolates of human respiratory coronavirus in vitro and in volunteers: evidence of heterogeneity among 229e-related strains," *J Med Virol*, vol. 13, no. 2, pp. 179–92, 1984.
- [8] K. McIntosh, W. B. Becker, and R. M. Chanock, "Growth in suckling-mouse brain of "ibv-like" viruses from patients with upper respiratory tract disease," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 58, no. 6, pp. 2268–73, 1967.
- [9] J. J. Cavallaro and A. S. Monto, "Community-wide outbreak of infection with a 229e-like coronavirus in tecumseh, michigan," *The Journal of infectious diseases*, vol. 122, no. 4, pp. 272–9, 1970.
- [10] L. J. Cui, C. Zhang, T. Zhang, R. J. Lu, Z. D. Xie, L. L. Zhang, C. Y. Liu, W. M. Zhou, L. Ruan, X. J. Ma, and W. J. Tan, "Human coronaviruses hcov-nl63 and hcov-hku1 in hospitalized children with acute respiratory infections in beijing, china," *Advances in virology*, vol. 2011, p. 129134, 2011.
- [11] P. C. Woo, S. K. Lau, C. M. Chu, K. H. Chan, H. W. Tsoi, Y. Huang, B. H. Wong, R. W. Poon, J. J. Cai, W. K. Luk, L. L. Poon, S. S. Wong, Y. Guan, J. S. Peiris, and K. Y. Yuen, "Characterization and complete genome sequence of a novel coronavirus, coronavirus hku1, from patients with pneumonia," *Journal of virology*, vol. 79, no. 2, pp. 884–95, 2005.

- [12] K. Pyrc, B. Berkhout, and L. van der Hoek, “The novel human coronaviruses nl63 and hku1,” *J Virol*, vol. 81, no. 7, pp. 3051–7, 2007.
- [13] M. Cotten, S. J. Watson, P. Kellam, A. A. Al-Rabeeah, H. Q. Makhdoom, A. Assiri, J. A. Al-Tawfiq, R. F. Alhakeem, H. Madani, F. A. AlRabiah, S. Al Hajjar, W. N. Al-nassir, A. Albarrak, H. Flemban, H. H. Balkhy, S. Al-subaie, A. L. Palser, A. Gall, R. Bashford-Rogers, A. Rambaut, A. I. Zumla, and Z. A. Memish, “Transmission and evolution of the middle east respiratory syndrome coronavirus in saudi arabia: a descriptive genomic study,” *Lancet*, vol. 382, no. 9909, pp. 1993–2002, 2013.
- [14] C. M. Coleman and M. B. Frieman, “Emergence of the middle east respiratory syndrome coronavirus,” *PLoS Pathog*, vol. 9, no. 9, p. e1003595, 2013.
- [15] WHO, “Middle east respiratory syndrome coronavirus (mers-cov) saudi arabia,” *WHO - Global Alert and Response*, 2 October 2014.
- [16] —, “Middle east respiratory syndrome coronavirus (mers-cov),” *WHO - Global Alert and Response*, 22 October 2015.
- [17] A. Assiri, A. McGeer, T. M. Perl, C. S. Price, A. A. Al Rabeeah, D. A. Cummings, Z. N. Alabdullatif, M. Assad, A. Almulhim, H. Makhdoom, H. Madani, R. Alhakeem, J. A. Al-Tawfiq, M. Cotten, S. J. Watson, P. Kellam, A. I. Zumla, and Z. A. Memish, “Hospital outbreak of middle east respiratory syndrome coronavirus,” *The New England journal of medicine*, vol. 369, no. 5, pp. 407–16, 2013.
- [18] R. Breban, J. Riou, and A. Fontanet, “Interhuman transmissibility of middle east respiratory syndrome coronavirus: estimation of pandemic risk,” *Lancet*, vol. 382, no. 9893, pp. 694–9, 2013.
- [19] S. Cauchemez, M. D. Van Kerkhove, S. Riley, C. A. Donnelly, C. Fraser, and N. M. Ferguson, “Transmission scenarios for middle east respiratory syndrome coronavirus (mers-cov) and how to tell them apart,” *Euro surveillance : bulletin europeen sur les maladies transmissibles = European communicable disease bulletin*, vol. 18, no. 24, 2013.
- [20] C. M. Coleman and M. B. Frieman, “Coronaviruses: important emerging human pathogens,” *J Virol*, vol. 88, no. 10, pp. 5209–12, 2014.
- [21] D. Corti, J. Zhao, M. Pedotti, L. Simonelli, S. Agnihothram, C. Fett, B. Fernandez-Rodriguez, M. Foglierini, G. Agatic, F. Vanzetta, R. Gopal, C. J. Langrish, N. A. Barrett, F. Sallusto, R. S. Baric, L. Varani, M. Zambon, S. Perlman, and A. Lanzavecchia, “Prophylactic and postexposure efficacy of a potent human monoclonal antibody against mers coronavirus,” *Proc Natl Acad Sci U S A*, vol. 112, no. 33, pp. 10473–8, 2015.
- [22] K. Muthumani, D. Falzarano, E. L. Reuschel, C. Tingey, S. Flingai, D. O. Villarreal, M. Wise, A. Patel, A. Izmirly, A. Aljuaid, A. M. Seliga, G. Soule, M. Morrow, K. A. Kraynyak, A. S. Khan, D. P. Scott, F. Feldmann, R. Lacasse, K. Meade-White, A. Okumura, K. E. Ugen, N. Y. Sardesai, J. J. Kim, G. Kobinger, H. Feldmann, and D. B. Weiner, “A synthetic consensus anti-spike protein dna vaccine induces protective immunity against middle east respiratory syndrome coronavirus in nonhuman primates,” *Sci Transl Med*, vol. 7, no. 301, p. 301ra132, 2015.

- [23] K. Jung and L. J. Saif, "Porcine epidemic diarrhea virus infection: Etiology, epidemiology, pathogenesis and immunoprophylaxis," *Vet J*, vol. 204, no. 2, pp. 134–43, 2015.
- [24] D. Song, H. Moon, and B. Kang, "Porcine epidemic diarrhea: a review of current epidemiology and available vaccines," *Clin Exp Vaccine Res*, vol. 4, no. 2, pp. 166–76, 2015.
- [25] S. Alenius, R. Niskanen, N. Juntti, and B. Larsson, "Bovine coronavirus as the causative agent of winter dysentery: serological evidence," *Acta veterinaria Scandinavica*, vol. 32, no. 2, pp. 163–70, 1991.
- [26] D. M. Barber, P. F. Nettleton, and J. A. Herring, "Disease in a dairy herd associated with the introduction and spread of bovine virus diarrhoea virus," *The Veterinary record*, vol. 117, no. 18, pp. 459–64, 1985.
- [27] P. J. Rottier, K. Nakamura, P. Schellen, H. Volders, and B. J. Haijema, "Acquisition of macrophage tropism during the pathogenesis of feline infectious peritonitis is determined by mutations in the feline coronavirus spike protein," *J Virol*, vol. 79, no. 22, pp. 14 122–30, 2005.
- [28] Y. Guan, B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. Peiris, and L. L. Poon, "Isolation and characterization of viruses related to the sars coronavirus from animals in southern china," *Science*, vol. 302, no. 5643, pp. 276–8, 2003.
- [29] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L. M. Ho, T. H. Lam, T. Q. Thach, P. Chau, K. P. Chan, S. V. Lo, P. Y. Leung, T. Tsang, W. Ho, K. H. Lee, E. M. Lau, N. M. Ferguson, and R. M. Anderson, "Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions," *Science*, vol. 300, no. 5627, pp. 1961–6, 2003.
- [30] A. Annan, H. J. Baldwin, V. M. Corman, S. M. Klose, M. Owusu, E. E. Nkrumah, E. K. Badu, P. Anti, O. Agbenyega, B. Meyer, S. Oppong, Y. A. Sarkodie, E. K. Kalko, P. H. Lina, E. V. Godlevska, C. Reusken, A. Seebens, F. Gloza-Rausch, P. Vallo, M. Tschapka, C. Drosten, and J. F. Drexler, "Human betacoronavirus 2c emc/2012-related viruses in bats, ghana and europe," *Emerging infectious diseases*, vol. 19, no. 3, pp. 456–9, 2013.
- [31] Z. A. Memish, N. Mishra, K. J. Olival, S. F. Fagbo, V. Kapoor, J. H. Epstein, R. Alhakeem, A. Durosinioun, M. Al Asmari, A. Islam, A. Kapoor, T. Briese, P. Daszak, A. A. Al Rabeeah, and W. I. Lipkin, "Middle east respiratory syndrome coronavirus in bats, saudi arabia," *Emerging infectious diseases*, vol. 19, no. 11, pp. 1819–23, 2013.
- [32] Q. Wang, J. Qi, Y. Yuan, Y. Xuan, P. Han, Y. Wan, W. Ji, Y. Li, Y. Wu, J. Wang, A. Iwamoto, P. C. Woo, K. Y. Yuen, J. Yan, G. Lu, and G. F. Gao, "Bat origins of mers-cov supported by bat coronavirus hku4 usage of human receptor cd26," *Cell Host Microbe*, vol. 16, no. 3, pp. 328–37, 2014.
- [33] G. Lu, Q. Wang, and G. F. Gao, "Bat-to-human: spike features determining 'host jump' of coronaviruses sars-cov, mers-cov, and beyond," *Trends Microbiol*, vol. 23, no. 8, pp. 468–78, 2015.

- [34] C. B. Reusken, B. L. Haagmans, M. A. Muller, C. Gutierrez, G. J. Godeke, B. Meyer, D. Muth, V. S. Raj, L. Smits-De Vries, V. M. Corman, J. F. Drexler, S. L. Smits, Y. E. El Tahir, R. De Sousa, J. van Beek, N. Nowotny, K. van Maanen, E. Hidalgo-Hermoso, B. J. Bosch, P. Rottier, A. Osterhaus, C. Gortazar-Schmidt, C. Drosten, and M. P. Koopmans, "Middle east respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study," *The Lancet. Infectious diseases*, vol. 13, no. 10, pp. 859–66, 2013.
- [35] R. A. Perera, P. Wang, M. R. Goma, R. El-Shesheny, A. Kandeil, O. Bagato, L. Y. Siu, M. M. Shehata, A. S. Kayed, Y. Moatasim, M. Li, L. L. Poon, Y. Guan, R. J. Webby, M. A. Ali, J. S. Peiris, and G. Kayali, "Seroepidemiology for mers coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in egypt, june 2013," *Euro surveillance : bulletin europeen sur les maladies transmissibles = European communicable disease bulletin*, vol. 18, no. 36, p. pii=20574, 2013.
- [36] Y. M. Baez-Santos, S. E. St John, and A. D. Mesecar, "The sars-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds," *Antiviral Res*, vol. 115, pp. 21–38, 2015.
- [37] A. K. Ghosh, G. Gong, V. Grum-Tokars, D. C. Mulhearn, S. C. Baker, M. Coughlin, B. S. Prabhakar, K. Sleeman, M. E. Johnson, and A. D. Mesecar, "Design, synthesis and antiviral efficacy of a series of potent chloropyridyl ester-derived sars-cov 3clpro inhibitors," *Bioorganic medicinal chemistry letters*, vol. 18, no. 20, pp. 5684–8, 2008.
- [38] J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S. A. Saldanha, P. Chase, A. Eggler, E. S. Dawson, Y. M. Baez-Santos, S. Tomar, A. M. Mielech, S. C. Baker, C. W. Lindsley, P. Hodder, A. Mesecar, and S. R. Stauffer, "Discovery, synthesis, and structure-based optimization of a series of n-(tert-butyl)-2-(n-arylamido)-2-(pyridin-3-yl) acetamides (ml188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (sars-cov) 3cl protease," *Journal of medicinal chemistry*, vol. 56, no. 2, pp. 534–46, 2013.
- [39] M. Turlington, A. Chun, S. Tomar, A. Eggler, V. Grum-Tokars, J. Jacobs, J. S. Daniels, E. Dawson, A. Saldanha, P. Chase, Y. M. Baez-Santos, C. W. Lindsley, P. Hodder, A. D. Mesecar, and S. R. Stauffer, "Discovery of n-(benzo[1,2,3]triazol-1-yl)-n-(benzyl)acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (sars-cov) 3clpro inhibitors: identification of ml300 and noncovalent nanomolar inhibitors with an induced-fit binding," *Bioorganic medicinal chemistry letters*, vol. 23, no. 22, pp. 6172–7, 2013.
- [40] Y. Lu, X. Lu, and M. R. Denison, "Identification and characterization of a serine-like proteinase of the murine coronavirus mhv-a59," *Journal of virology*, vol. 69, no. 6, pp. 3554–9, 1995.
- [41] J. Ziebuhr, E. J. Snijder, and A. E. Gorbalenya, "Virus-encoded proteinases and proteolytic processing in the nidovirales," *The Journal of general virology*, vol. 81, no. Pt 4, pp. 853–79, 2000.

- [42] K. Anand, G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld, "Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain," *The EMBO journal*, vol. 21, no. 13, pp. 3213–24, 2002.
- [43] M. F. Hsu, C. J. Kuo, K. T. Chang, H. C. Chang, C. C. Chou, T. P. Ko, H. L. Shr, G. G. Chang, A. H. Wang, and P. H. Liang, "Mechanism of the maturation process of sars-cov 3cl protease," *The Journal of biological chemistry*, vol. 280, no. 35, pp. 31 257–66, 2005.
- [44] L. Lai, X. Han, H. Chen, P. Wei, C. Huang, S. Liu, K. Fan, L. Zhou, Z. Liu, J. Pei, and Y. Liu, "Quaternary structure, substrate selectivity and inhibitor design for sars 3c-like proteinase," *Current pharmaceutical design*, vol. 12, no. 35, pp. 4555–64, 2006.
- [45] V. Grum-Tokars, K. Ratia, A. Begaye, S. C. Baker, and A. D. Mesecar, "Evaluating the 3c-like protease activity of sars-coronavirus: recommendations for standardized assays for drug discovery," *Virus research*, vol. 133, no. 1, pp. 63–73, 2008.
- [46] S. Chen, J. Zhang, T. Hu, K. Chen, H. Jiang, and X. Shen, "Residues on the dimer interface of sars coronavirus 3c-like protease: dimer stability characterization and enzyme catalytic activity analysis," *Journal of biochemistry*, vol. 143, no. 4, pp. 525–36, 2008.
- [47] J. Shi, J. Sivaraman, and J. Song, "Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3c-like protease," *Journal of virology*, vol. 82, no. 9, pp. 4620–9, 2008.
- [48] C. G. Wu, S. C. Cheng, S. C. Chen, J. Y. Li, Y. H. Fang, Y. H. Chen, and C. Y. Chou, "Mechanism for controlling the monomer-dimer conversion of sars coronavirus main protease," *Acta crystallographica. Section D, Biological crystallography*, vol. 69, no. Pt 5, pp. 747–55, 2013.
- [49] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao, "The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 23, pp. 13 190–5, 2003.
- [50] I. L. Lu, N. Mahindroo, P. H. Liang, Y. H. Peng, C. J. Kuo, K. C. Tsai, H. P. Hsieh, Y. S. Chao, and S. Y. Wu, "Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease," *Journal of medicinal chemistry*, vol. 49, no. 17, pp. 5154–61, 2006.
- [51] J. Tan, K. H. Verschueren, K. Anand, J. Shen, M. Yang, Y. Xu, Z. Rao, J. Bigalke, B. Heisen, J. R. Mesters, K. Chen, X. Shen, H. Jiang, and R. Hilgenfeld, "ph-dependent conformational flexibility of the sars-cov main proteinase (m(pro)) dimer: molecular dynamics simulations and multiple x-ray structure analyses," *J Mol Biol*, vol. 354, no. 1, pp. 25–40, 2005.
- [52] M. Bartlam, H. Yang, and Z. Rao, "Structural insights into sars coronavirus proteins," *Curr Opin Struct Biol*, vol. 15, no. 6, pp. 664–72, 2005.

- [53] A. K. Ghosh, K. Xi, K. Ratia, B. D. Santarsiero, W. Fu, B. H. Harcourt, P. A. Rota, S. C. Baker, M. E. Johnson, and A. D. Mesecar, "Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors," *Journal of medicinal chemistry*, vol. 48, no. 22, pp. 6767–71, 2005.
- [54] A. K. Ghosh, K. Xi, V. Grum-Tokars, X. Xu, K. Ratia, W. Fu, K. V. Houser, S. C. Baker, M. E. Johnson, and A. D. Mesecar, "Structure-based design, synthesis, and biological evaluation of peptidomimetic sars-cov 3clpro inhibitors," *Bioorganic medicinal chemistry letters*, vol. 17, no. 21, pp. 5876–80, 2007.
- [55] C. Y. Chou, H. C. Chang, W. C. Hsu, T. Z. Lin, C. H. Lin, and G. G. Chang, "Quaternary structure of the severe acute respiratory syndrome (sars) coronavirus main protease," *Biochemistry*, vol. 43, no. 47, pp. 14958–70, 2004.
- [56] S. Chen, T. Hu, J. Zhang, J. Chen, K. Chen, J. Ding, H. Jiang, and X. Shen, "Mutation of gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3c-like protease: crystal structure with molecular dynamics simulations," *J Biol Chem*, vol. 283, no. 1, pp. 554–64, 2008.
- [57] S. Chen, F. Jonas, C. Shen, and R. Hilgenfeld, "Liberation of sars-cov main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode," *Protein cell*, vol. 1, no. 1, pp. 59–74, 2010.
- [58] C. Li, Y. Qi, X. Teng, Z. Yang, P. Wei, C. Zhang, L. Tan, L. Zhou, Y. Liu, and L. Lai, "Maturation mechanism of severe acute respiratory syndrome (sars) coronavirus 3c-like proteinase," *The Journal of biological chemistry*, vol. 285, no. 36, pp. 28134–40, 2010.
- [59] S. Agnihothram, J. Yount, B. L., E. F. Donaldson, J. Huynh, V. D. Menachery, L. E. Gralinski, R. L. Graham, M. M. Becker, S. Tomar, T. D. Scobey, H. L. Osswald, A. Whitmore, R. Gopal, A. K. Ghosh, A. Mesecar, M. Zamboni, M. Heise, M. R. Denison, and R. S. Baric, "A mouse model for betacoronavirus subgroup 2c using a bat coronavirus strain hku5 variant," *MBio*, vol. 5, no. 2, pp. e00047–14, 2014.
- [60] S. Tomar, M. L. Johnston, S. E. St John, H. L. Osswald, P. R. Nyalapatla, L. N. Paul, A. K. Ghosh, M. R. Denison, and A. D. Mesecar, "Ligand-induced dimerization of middle east respiratory syndrome (mers) coronavirus nsp5 protease (3clpro): Implications for nsp5 regulation and the development of antivirals," *The Journal of biological chemistry*, vol. 290, no. 32, pp. 19403–22, 2015.
- [61] C. N. Chen, C. P. Lin, K. K. Huang, W. C. Chen, H. P. Hsieh, P. H. Liang, and J. T. Hsu, "Inhibition of sars-cov 3c-like protease activity by theaflavin-3,3'-digallate (tf3)," *Evid Based Complement Alternat Med*, vol. 2, no. 2, pp. 209–215, 2005.
- [62] B. K. Shoichet, "Screening in a spirit haunted world," *Drug Discov Today*, vol. 11, no. 13-14, pp. 607–15, 2006.
- [63] S. L. McGovern and B. K. Shoichet, "Kinase inhibitors: not just for kinases anymore," *J Med Chem*, vol. 46, no. 8, pp. 1478–83, 2003.
- [64] G. M. Rishton, "Nonleadlikeness and leadlikeness in biochemical screening," *Drug Discov Today*, vol. 8, no. 2, pp. 86–96, 2003.

- [65] C. C. Wen, Y. H. Kuo, J. T. Jan, P. H. Liang, S. Y. Wang, H. G. Liu, C. K. Lee, S. T. Chang, C. J. Kuo, S. S. Lee, C. C. Hou, P. W. Hsiao, S. C. Chien, L. F. Shyur, and N. S. Yang, "Specific plant terpenoids and lignoids possess potent antiviral activities against severe acute respiratory syndrome coronavirus," *Journal of medicinal chemistry*, vol. 50, no. 17, pp. 4087–95, 2007.
- [66] C. J. Kuo, H. G. Liu, Y. K. Lo, C. M. Seong, K. I. Lee, Y. S. Jung, and P. H. Liang, "Individual and common inhibitors of coronavirus and picornavirus main proteases," *FEBS letters*, vol. 583, no. 3, pp. 549–55, 2009.
- [67] J. T. Hsu, C. J. Kuo, H. P. Hsieh, Y. C. Wang, K. K. Huang, C. P. Lin, P. F. Huang, X. Chen, and P. H. Liang, "Evaluation of metal-conjugated compounds as inhibitors of 3cl protease of sars-cov," *FEBS letters*, vol. 574, no. 1-3, pp. 116–20, 2004.
- [68] C. C. Lee, C. J. Kuo, M. F. Hsu, P. H. Liang, J. M. Fang, J. J. Shie, and A. H. Wang, "Structural basis of mercury- and zinc-conjugated complexes as sars-cov 3c-like protease inhibitors," *FEBS letters*, vol. 581, no. 28, pp. 5454–8, 2007.
- [69] U. Bacha, J. Barrila, A. Velazquez-Campoy, S. A. Leavitt, and E. Freire, "Identification of novel inhibitors of the sars coronavirus main protease 3clpro," *Biochemistry*, vol. 43, no. 17, pp. 4906–12, 2004.
- [70] Y. M. Shao, W. B. Yang, H. P. Peng, M. F. Hsu, K. C. Tsai, T. H. Kuo, A. H. Wang, P. H. Liang, C. H. Lin, A. S. Yang, and C. H. Wong, "Structure-based design and synthesis of highly potent sars-cov 3cl protease inhibitors," *Chembiochem*, vol. 8, no. 14, pp. 1654–7, 2007.
- [71] J. J. Shie, J. M. Fang, C. J. Kuo, T. H. Kuo, P. H. Liang, H. J. Huang, W. B. Yang, C. H. Lin, J. L. Chen, Y. T. Wu, and C. H. Wong, "Discovery of potent anilide inhibitors against the severe acute respiratory syndrome 3cl protease," *Journal of medicinal chemistry*, vol. 48, no. 13, pp. 4469–73, 2005.
- [72] C. Y. Wu, K. Y. King, C. J. Kuo, J. M. Fang, Y. T. Wu, M. Y. Ho, C. L. Liao, J. J. Shie, P. H. Liang, and C. H. Wong, "Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3cl protease," *Chemistry biology*, vol. 13, no. 3, pp. 261–8, 2006.
- [73] L. Zhu, S. George, M. F. Schmidt, S. I. Al-Gharabli, J. Rademann, and R. Hilgenfeld, "Peptide aldehyde inhibitors challenge the substrate specificity of the sars-coronavirus main protease," *Antiviral research*, vol. 92, no. 2, pp. 204–12, 2011.
- [74] K. Akaji, H. Konno, H. Mitsui, K. Teruya, Y. Shimamoto, Y. Hattori, T. Ozaki, M. Kusunoki, and A. Sanjoh, "Structure-based design, synthesis, and evaluation of peptide-mimetic sars 3cl protease inhibitors," *Journal of medicinal chemistry*, vol. 54, no. 23, pp. 7962–73, 2011.
- [75] D. H. Goetz, Y. Choe, E. Hansell, Y. T. Chen, M. McDowell, C. B. Jonsson, W. R. Roush, J. McKerrow, and C. S. Craik, "Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the sars coronavirus," *Biochemistry*, vol. 46, no. 30, pp. 8744–52, 2007.

- [76] U. Bacha, J. Barrila, S. B. Gabelli, Y. Kiso, L. Mario Amzel, and E. Freire, "Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease 3cl(pro)," *Chemical biology drug design*, vol. 72, no. 1, pp. 34–49, 2008.
- [77] L. Guterman, "Covalent drugs form long-lived ties (vol 89, pg 19, 2011)," *Chemical Engineering News*, vol. 89, no. 38, pp. 2–2, 2011.
- [78] B. Turk, "Targeting proteases: successes, failures and future prospects," *Nature reviews. Drug discovery*, vol. 5, no. 9, pp. 785–99, 2006.
- [79] J. Jacobs, S. Zhou, E. Dawson, J. S. Daniels, P. Hodder, V. Tokars, A. Mesecar, C. W. Lindsley, and S. R. Stauffer, *Discovery of non-covalent inhibitors of the SARS main proteinase 3CLpro*, Bethesda (MD), 2010, jacobs, Jon Zhou, Sandra Dawson, Eric Daniels, J. Scott Hodder, Peter Tokars, Valerie Mesecar, Andrew Lindsley, Craig W. Stauffer, Shaun R. Book Chapter. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23658941>
- [80] B. A. Larder, K. Hertogs, S. Bloor, C. H. van den Eynde, W. DeCian, Y. Wang, W. W. Freimuth, and G. Tarpley, "Tipranavir inhibits broadly protease inhibitor-resistant hiv-1 clinical samples," *AIDS*, vol. 14, no. 13, pp. 1943–8, 2000.
- [81] J. C. Ruela Correa, D. M. D'Arcy, C. H. dos Reis Serra, and H. R. Nunes Salgado, "Darunavir: a critical review of its properties, use and drug interactions," *Pharmacology*, vol. 90, no. 1-2, pp. 102–9, 2012.
- [82] B. Vergani and S. Rusconi, "Tipranavir in the protease inhibitors arena," *Drugs in RD*, vol. 11, no. 4, pp. 291–3, 2011.
- [83] D. A. Matthews, P. S. Dragovich, S. E. Webber, S. A. Fuhrman, A. K. Patick, L. S. Zalman, T. F. Hendrickson, R. A. Love, T. J. Prins, J. T. Marakovits, R. Zhou, J. Tikhe, C. E. Ford, J. W. Meador, R. A. Ferre, E. L. Brown, S. L. Binford, M. A. Brothers, D. M. DeLisle, and S. T. Worland, "Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3c protease with potent antiviral activity against multiple rhinovirus serotypes," *Proc Natl Acad Sci U S A*, vol. 96, no. 20, pp. 11 000–7, 1999.
- [84] M. M. Lai and D. Cavanagh, "The molecular biology of coronaviruses," *Advances in virus research*, vol. 48, pp. 1–100, 1997.
- [85] K. Anand, J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld, "Coronavirus main proteinase (3clpro) structure: basis for design of anti-sars drugs," *Science*, vol. 300, no. 5626, pp. 1763–7, 2003.
- [86] Q. Zhao, S. Li, F. Xue, Y. Zou, C. Chen, M. Bartlam, and Z. Rao, "Structure of the main protease from a global infectious human coronavirus, hcov-hku1," *Journal of virology*, vol. 82, no. 17, pp. 8647–55, 2008.
- [87] J. Barrila, U. Bacha, and E. Freire, "Long-range cooperative interactions modulate dimerization in sars 3clpro," *Biochemistry*, vol. 45, no. 50, pp. 14 908–16, 2006.

- [88] L. Lim, J. Shi, Y. Mu, and J. Song, “Dynamically-driven enhancement of the catalytic machinery of the sars 3c-like protease by the s284-t285-i286/a mutations on the extra domain,” *PLoS One*, vol. 9, no. 7, p. e101941, 2014.
- [89] J. Shi, N. Han, L. Lim, S. Lua, J. Sivaraman, L. Wang, Y. Mu, and J. Song, “Dynamically-driven inactivation of the catalytic machinery of the sars 3c-like protease by the n214a mutation on the extra domain,” *PLoS Comput Biol*, vol. 7, no. 2, p. e1001084, 2011.
- [90] J. S. Sparks, E. F. Donaldson, X. Lu, R. S. Baric, and M. R. Denison, “A novel mutation in murine hepatitis virus nsp5, the viral 3c-like proteinase, causes temperature-sensitive defects in viral growth and protein processing,” *Journal of virology*, vol. 82, no. 12, pp. 5999–6008, 2008.
- [91] C. C. Stobart, A. S. Lee, X. Lu, and M. R. Denison, “Temperature-sensitive mutants and revertants in the coronavirus nonstructural protein 5 protease (3clpro) define residues involved in long-distance communication and regulation of protease activity,” *Journal of virology*, vol. 86, no. 9, pp. 4801–10, 2012.
- [92] SchrodingerRelease, “Prime, version 3.9, schrodinger, llc, new york, ny, 2015,” 2015-1.
- [93] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, “Gromacs: fast, flexible, and free,” *J Comput Chem*, vol. 26, no. 16, pp. 1701–18, 2005.
- [94] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, “Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, vol. 29, no. 7, pp. 845–54, 2013.
- [95] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, “Improved side-chain torsion potentials for the amber ff99sb protein force field,” *Proteins*, vol. 78, no. 8, pp. 1950–8, 2010.
- [96] M. Abraham, D. Van Der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, “Gromacs user manual version 5.0.2 www.gromacs.org,” 2014.
- [97] D. M. Kruger, P. C. Rathi, C. Pflieger, and H. Gohlke, “Cna web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability, and function,” *Nucleic Acids Res*, vol. 41, no. Web Server issue, pp. W340–8, 2013.
- [98] C. L. McClendon, G. Friedland, D. L. Mobley, H. Amirkhani, and M. P. Jacobson, “Quantifying correlations between allosteric sites in thermodynamic ensembles,” *Journal of Chemical Theory and Computation*, vol. 5, no. 9, pp. 2486–2502, 2009.
- [99] S. Raschka, “<http://sebastianraschka.com/contact.html>.”
- [100] E. Papaleo, G. Renzetti, and M. Tiberti, “Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: a molecular dynamics investigation,” *PLoS One*, vol. 7, no. 4, p. e35686, 2012.

- [101] D. P. Myers, L. K. Jackson, V. G. Ipe, G. E. Murphy, and M. A. Phillips, "Long-range interactions in the dimer interface of ornithine decarboxylase are important for enzyme function," *Biochemistry*, vol. 40, no. 44, pp. 13 230–6, 2001.
- [102] L. C. Tisi and P. A. Evans, "Conserved structural features on protein surfaces: small exterior hydrophobic clusters," *Journal of molecular biology*, vol. 249, no. 2, pp. 251–8, 1995.
- [103] M. Machius, N. Declerck, R. Huber, and G. Wiegand, "Kinetic stabilization of bacillus licheniformis alpha-amylase through introduction of hydrophobic residues at the surface," *Journal of Biological Chemistry*, vol. 278, no. 13, pp. 11 546–11 553, 2003.
- [104] V. G. H. Eijsink, A. Bjork, S. Gaseidnes, R. Sirevag, B. Synstad, B. van den Burg, and G. Vriend, "Rational engineering of enzyme stability," *Journal of Biotechnology*, vol. 113, no. 1-3, pp. 105–120, 2004.
- [105] C. C. Stobart, N. R. Sexton, H. Munjal, X. Lu, K. L. Molland, S. Tomar, A. D. Mesecar, and M. R. Denison, "Chimeric exchange of coronavirus nsp5 proteases (3clpro) identifies common and divergent regulatory determinants of protease activity," *Journal of virology*, vol. 87, no. 23, pp. 12 611–8, 2013.
- [106] A. del Sol, C. J. Tsai, B. Ma, and R. Nussinov, "The origin of allosteric functional modulation: multiple pre-existing pathways," *Structure*, vol. 17, no. 8, pp. 1042–50, 2009.
- [107] R. M. Daniel and M. J. Danson, "Temperature and the catalytic activity of enzymes: a fresh understanding," *FEBS letters*, vol. 587, no. 17, pp. 2738–43, 2013.
- [108] K. A. Hanley, "The double-edged sword: How evolution can make or break a live-attenuated virus vaccine," *Evolution (N Y)*, vol. 4, no. 4, pp. 635–643, 2011.
- [109] A. M. Mielech, X. F. Deng, Y. F. Chen, E. Kindler, D. L. Wheeler, A. D. Mesecar, V. Thiel, S. Perlman, and S. C. Baker, "Murine coronavirus ubiquitin-like domain is important for papain-like protease stability and viral pathogenesis," *Journal of virology*, vol. 89, no. 9, pp. 4907–4917, 2015.
- [110] D. Hamre and J. J. Procknow, "A new virus isolated from the human respiratory tract," *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine*, vol. 121, no. 1, pp. 190–3, 1966.
- [111] L. van der Hoek, K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout, "Identification of a new human coronavirus," *Nature medicine*, vol. 10, no. 4, pp. 368–73, 2004.
- [112] A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, and R. A. Fouchier, "Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia," *The New England journal of medicine*, vol. 367, no. 19, pp. 1814–20, 2012.

- [113] D. Butler, "Clusters of coronavirus cases put scientists on alert," *Nature*, vol. 492, no. 7428, pp. 166–7, 2012.
- [114] S. van Boheemen, M. de Graaf, C. Lauber, T. M. Bestebroer, V. S. Raj, A. M. Zaki, A. D. Osterhaus, B. L. Haagmans, A. E. Gorbalenya, E. J. Snijder, and R. A. Fouchier, "Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans," *MBio*, vol. 3, no. 6, 2012.
- [115] J. F. Chan, K. S. Li, K. K. To, V. C. Cheng, H. Chen, and K. Y. Yuen, "Is the discovery of the novel human betacoronavirus 2c emc/2012 (hcov-emc) the beginning of another sars-like pandemic?" *J Infect*, vol. 65, no. 6, pp. 477–89, 2012.
- [116] CDC, "Mers in the u.s." *CDC, Atlanta, GA.*, June 20 2014.
- [117] H. Yang, M. Bartlam, and Z. Rao, "Drug design targeting the main protease, the achilles' heel of coronaviruses," *Current pharmaceutical design*, vol. 12, no. 35, pp. 4573–90, 2006.
- [118] Y. M. Baez-Santos, S. J. Barraza, M. W. Wilson, M. P. Agius, A. M. Mielech, N. M. Davis, S. C. Baker, S. D. Larsen, and A. D. Mesecar, "X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases," *Journal of medicinal chemistry*, vol. 57, no. 6, pp. 2393–412, 2014.
- [119] S. E. St John, S. Tomar, S. R. Stauffer, and A. D. Mesecar, "Targeting zoonotic viruses: Structure-based inhibition of the 3c-like protease from bat coronavirus hku4-the likely reservoir host to the human coronavirus that causes middle east respiratory syndrome (mers)," *Bioorg Med Chem*, vol. 23, no. 17, pp. 6036–48, 2015.
- [120] Y. Shen, C. Y. Chou, G. G. Chang, and L. Tong, "Is dimerization required for the catalytic activity of bacterial biotin carboxylase?" *Mol Cell*, vol. 22, no. 6, pp. 807–18, 2006.
- [121] A. G. W. Leslie and H. R. Powell, "Processing diffraction data with mos-flm," *Evolving Methods for Macromolecular Crystallography*, vol. 245, pp. 41–51, 2007.
- [122] Z. Otwinowski and W. Minor, "Processing of x-ray diffraction data collected in oscillation mode," *Macromolecular Crystallography, Pt A*, vol. 276, pp. 307–326, 1997.
- [123] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, "Phenix: a comprehensive python-based system for macromolecular structure solution," *Acta Crystallographica Section D-Biological Crystallography*, vol. 66, pp. 213–221, 2010.
- [124] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of coot," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. Pt 4, pp. 486–501, 2010.

- [125] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "Molprobity: all-atom structure validation for macromolecular crystallography," *Acta crystallographica. Section D, Biological crystallography*, vol. 66, no. Pt 1, pp. 12–21, 2010.
- [126] "The pymol molecular graphics system, version 1.7.2.1, schrodinger, llc."
- [127] K. Fan, P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, and L. Lai, "Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3c-like proteinase," *The Journal of biological chemistry*, vol. 279, no. 3, pp. 1637–42, 2004.
- [128] J. Shi, Z. Wei, and J. Song, "Dissection study on the severe acute respiratory syndrome 3c-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors," *The Journal of biological chemistry*, vol. 279, no. 23, pp. 24 765–73, 2004.
- [129] J. E. Blanchard, N. H. Elowe, C. Huitema, P. D. Fortin, J. D. Cechetto, L. D. Eltis, and E. D. Brown, "High-throughput screening identifies inhibitors of the sars coronavirus main proteinase," *Chemistry biology*, vol. 11, no. 10, pp. 1445–53, 2004.
- [130] U. Kaeppeler, N. Stiefl, M. Schiller, R. Vicik, A. Breuning, W. Schmitz, D. Rupprecht, C. Schmuck, K. Baumann, J. Ziebuhr, and T. Schirmeister, "A new lead for nonpeptidic active-site-directed inhibitors of the severe acute respiratory syndrome coronavirus main protease discovered by a combination of screening and docking methods," *Journal of medicinal chemistry*, vol. 48, no. 22, pp. 6832–42, 2005.
- [131] P. H. Brown, A. Balbo, and P. Schuck, "Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation," *Current protocols in immunology / edited by John E. Coligan ... [et al.]*, vol. Chapter 18, p. Unit 18 15, 2008.
- [132] S. C. Cheng, G. G. Chang, and C. Y. Chou, "Mutation of glu-166 blocks the substrate-induced dimerization of sars coronavirus main protease," *Biophysical journal*, vol. 98, no. 7, pp. 1327–36, 2010.
- [133] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, "Weblogo: a sequence logo generator," *Genome Res*, vol. 14, no. 6, pp. 1188–90, 2004.
- [134] F. Corpet, "Multiple sequence alignment with hierarchical clustering," *Nucleic Acids Res*, vol. 16, no. 22, pp. 10 881–90, 1988.
- [135] X. Robert and P. Gouet, "Deciphering key features in protein structures with the new endscript server," *Nucleic Acids Res*, vol. 42, no. Web Server issue, pp. W320–4, 2014.
- [136] X. Deng, S. E. StJohn, H. L. Osswald, A. O'Brien, B. S. Banach, K. Sleeman, A. K. Ghosh, A. D. Mesecar, and S. C. Baker, "Coronaviruses resistant to a 3c-like protease inhibitor are attenuated for replication and pathogenesis, revealing a low genetic barrier but high fitness cost of resistance," *J Virol*, vol. 88, no. 20, pp. 11 886–98, 2014.

- [137] SchrodingerRelease, “Desmond molecular dynamics system, version 4.0, d.e.shaw schrddinger, new york, ny, 2014,” 2014-4.
- [138] —, “Maestro, version 10.0, schrddinger, llc, new york, ny, 2014,” 2014-4.

VITA

VITA

EDUCATION

- Ph.D., 2015, Department of Biological Sciences, Purdue University, West Lafayette, Indiana
- M.Sc., 2009, Department of Biotechnology, Indian Institute of Technology Roorkee, India
- B.Sc., 2007, Department of Biotechnology, ITS Paramedical college, Ghaziabad, India

RESEARCH EXPERIENCE

Graduate Student, Purdue University, August 2010 - December 2015

Advisor: Professor Andrew D. Mesecar

- Independently optimized the expression and purification protocols for several non-tagged 3C-like proteases utilizing multi-step chromatographic methods
- Characterized inhibition kinetics of over 250 novel compounds using steady-state kinetics
- Identified novel non-covalent inhibitors of a target human coronavirus enzyme, 3C-like protease, through high-throughput screening (HTS) and performed structure activity relationship (SAR) analysis for lead optimization
- Screened crystallization conditions in high-throughput format using robotics
- Synchrotron experience for automated X-ray crystallographic data collection (Argonne National Laboratories)
- Used X-ray structure elucidation and structure-based drug design in collaboration with medicinal and computational chemists for optimization of potent inhibitors of 3C-like protease from SARS coronavirus

- Utilized biophysical techniques like CD, AUC, SPR and SEC-MALS for biophysical characterization of 3C-like proteases and ligand binding

Junior Research Fellow, IIT Roorkee, India August 2009 - July 2010

Advisor: Dr. Pravindra Kumar

- Structural elucidation of BphB, protein involved in the aerobic biodegradation of biphenyl and polychlorinated biphenyls
- Structural elucidation of DNA-binding domain of transcription factor KdpE that is involved in K⁺ ion sensing in bacteria

Masters, IIT Roorkee, India August 2007 - May 2009

Advisor: Dr. Pravindra Kumar

- Expression, purification and crystallization of Trypsin inhibitor from natural-product extract

PUBLICATIONS

- **Tomar, S.**, Johnston, M. L., St John, S. E., Osswald, H. L., Nyalapatla, P. R., Paul, L. N., Ghosh, A. K., Denison, M. R., and Mesecar, A. D. (2015) Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro): IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS. *J Biol Chem* 290, 19403-19422
- St John, S. E., **Tomar, S.**, Stauffer, S. R., and Mesecar, A. D. (2015) Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4-The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS). *Bioorganic medicinal chemistry* 23, 6036-6048

- Turlington, M., Chun, A., **Tomar, S.**, Eggler, A., Grum-Tokars, V., Jacobs, J., Daniels, J. S., Dawson, E., Saldanha, A., Chase, P., Baez-Santos, Y. M., Lindsley, C. W., Hodder, P., Mesecar, A. D., and Stauffer, S. R. (2013) Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg Med Chem Lett* 23, 6172-6177
- Stobart, C. C., Sexton, N. R., Munjal, H., Lu, X., Molland, K. L., **Tomar, S.**, Mesecar, A. D., and Denison, M. R. (2013) Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *J Virol* 87, 12611-12618
- Agnihothram, S., Yount, B. L., Jr., Donaldson, E. F., Huynh, J., Menachery, V. D., Gralinski, L. E., Graham, R. L., Becker, M. M., **Tomar, S.**, Scobey, T. D., Osswald, H. L., Whitmore, A., Gopal, R., Ghosh, A. K., Mesecar, A., Zambon, M., Heise, M., Denison, M. R., and Baric, R. S. (2014) A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant. *MBio* 5, e00047-00014
- Jacobs, J., Grum-Tokars, V., Zhou, Y., Turlington, M., Saldanha, S. A., Chase, P., Eggler, A., Dawson, E. S., Baez-Santos, Y. M., **Tomar, S.**, Mielech, A. M., Baker, S. C., Lindsley, C. W., Hodder, P., Mesecar, A., and Stauffer, S. R. (2013) Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease. *J Med Chem* 56, 534-546

POSTER PRESENTATIONS and TALKS

- **Sakshi Tomar**, Melanie Johnston, Mark Denison, Andrew Mesecar (2015). Ligand-induced dimerization of MERS coronavirus 3CL protease (3CLpro): implications for 3CLpro regulation and the development of antivirals. Midwest Enzyme Chemistry Conference, Illinois Institute of Technology, Chicago, IN. September 20th. (Poster)
- **Sakshi Tomar**, Aimee Egger, Valerie Grum-Tokars, Craig W. Lindsley, Shaun R. Stauffer, Anna Mielech, Susan Baker, Andrew Mesecar (2014). Discovery of potent, non-covalent inhibitors of SARS-CoV 3CLpro. Sigma Xi Poster competition, Purdue University, West Lafayette, IN. April 2nd. (Poster)
- **Sakshi Tomar** (2014). Ligand-induced dimerization of MERS coronavirus 3CL protease (3CLpro): implications for 3CLpro regulation and the development of antivirals. Infectious Disease Seminar series, Department of Biological Sciences, Purdue University, West Lafayette, IN. November 6th. (Talk)
- **Sakshi Tomar** (2014). Ligand-induced dimerization of MERS coronavirus 3CL protease (3CLpro): implications for 3CLpro regulation and the development of antivirals. Graduate Student Research In Progress Seminar series, Department of Biological Sciences, Purdue University, West Lafayette, IN. November 6th. (Talk)
- **Sakshi Tomar** (2014). Characterization of 3CL protease (3CLpro) from recently emerged MERS coronavirus. Infectious Disease Seminar series, Department of Biological Sciences, Purdue University, West Lafayette, IN. March 6th. (Talk)
- **Sakshi Tomar**, Aimee Egger, Valerie Grum-Tokars, Craig W. Lindsley, Shaun R. Stauffer, Anna Mielech, Susan Baker, Andrew Mesecar (2013). Discovery of potent, non-covalent inhibitors of SARS-CoV 3CLpro. Midwest Enzyme Chemistry Conference, Loyola University, Chicago, IN. October 12th. (Poster)

- **Sakshi Tomar** (2012). Anticoronaviral drug discovery: An effort towards the development of broad spectrum Coronavirus 3CLpro Inhibitors. Infectious Disease Seminar series, Department of Biological Sciences, Purdue University, West Lafayette, IN. November 8th. (Talk)

FELLOWSHIPS AND AWARDS

- Graduate Research Fellowship from the Purdue Research Foundation, Purdue University (June 2014-May 2015)
- Ross Fellowship, Purdue University (2010-2011)
- Junior Research Fellowship, awarded by CSIR, Ministry of Human Resource and Development, Government of India (2009-2010)
- Fellowship awarded by Department of Biotechnology (DBT), Govt. of India (2007-2009)
- Dr. G. Pande Medal awarded for the female student obtaining highest C.G.P.A. in M.Sc. final year examination for the session 2008-09, IIT Roorkee, India
- Smt. Sushilavati Khosla Medal awarded for the female student obtaining highest C.G.P.A. in M.Sc./M.Tech. (Science) courses for the session 2008-09, IIT Roorkee, India