Profiling of Substrate-Specificity and Rational Design of Peptidomimetic Inhibitors for 3C-Like Proteases of Coronaviruses

CHUCK, Chi Pang

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Biochemistry (Medicine)

The Chinese University of Hong Kong
September 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thesis/Assessment Committee

Professor SHAW Pang Chui (Chair)

Professor WAN Chi Cheong, David (Thesis Supervisor)

Professor WONG Kam Bo (Thesis Co-Supervisor)

Professor FONG Wing Ping (Committee Member)

Professor CHENG Yuen Kit (External Examiner)

Abstract

3C-like protease ($3CL^{pro}$) of severe acute respiratory syndrome-coronavirus (SARS-CoV) is required for autoprocessing of the polyproteins 1a and 1ab, and is a potential target for treating coronaviral infection. To obtain a thorough understanding of its substrate preference, we created a substrate library of 19×8 variants by performing saturation mutagenesis on the autocleavage sequence at P5 to P3' positions. The substrate sequences were inserted between cyan and yellow fluorescent proteins so that the cleavage rates were monitored by *in vitro* fluorescence resonance energy transfer (FRET). The relative cleavage rate for different substrate sequences was correlated with various structural properties. P5 and P3 positions prefer residues with high β-sheet propensity; P4 prefers small hydrophobic residues; P2 prefers hydrophobic residues without β-branch. Gln is the best residue at P1 position, but observable cleavage can be detected with His and Met substitutions. P1' position prefers small residues, while P2' and P3' positions have no strong preference on residue substitutions. Noteworthy, solvent exposed sites such as P5, P3 and P3' positions favour positively charged residues over negatively charged one, suggesting that electrostatic interactions may play a role in catalysis. A super-active substrate, which combined the preferred residues at P5 to P1 positions, was found to have 2.8 fold higher activity than the wild-type sequence.

Inhibition of SARS-CoV $3CL^{pro}$ proteolytic activity suppresses virion replication and virus-induced cytopathic effects. Peptidomimetic inhibitors with nitrile warheads, which inhibit Cys protease activity, have been applied for clinical therapy. To investigate whether the nitrile group can target $3CL^{pro}$, a series of nitrile-based peptidomimetic inhibitors with various protective groups, peptide length and

peptide sequences were synthesized. Inhibitor potency in terms of IC_{50} and K_i values was determined by FRET assay. Most of these nitrile-based inhibitors in micromolar range can significantly reduce $3CL^{pro}$ activity. The most potent inhibitor is the tetrapeptidomimetic inhibitor linked with carbobenzyloxy (cbz) group 'cbz-AVLQ-CN' with IC_{50} and K_i values of 5.9 ± 0.6 μM and 0.62 ± 0.11 μM respectively. Crystal structures of $3CL^{pro}$-inhibitor complexes demonstrated that nitrile warhead covalently bonded to Cys145, while P1 – P4 residues interacted with $3CL^{pro}$ as substrate bound. The cbz group in 'cbz-AVLQ-CN' flipped into a cavity of Glu166 – Pro168, providing an extra binding force to enhance inhibitor potency. In conclusion, the nitrile-based peptidomimetic inhibitor with cbz group is a convincing model for drug development.

Substrate specificities of various $3CL^{pro}$ were further investigated by using the substrate library of SARS-CoV $3CL^{pro}$. Among various viral strains, the proteases of HCoV-NL63, HCoV-OC43 and infectious bronchitis virus (IBV) were selected from group I, IIa and III respectively for specificity profiling. Their proteolytic rates against 19×8 variants were obtained by FRET assay, and correlated with structural properties of substituting residues. Like SARS-CoV $3CL^{pro}$ in group IIb, these $3CL^{pro}$ consistently prefer small hydrophobic P4 residues, positively charged P3 residues, hydrophobic P2 residues without β-branch, P1-Gln and small P1' residues. These proteases also tend to accommodate P5 and P3' residues with positive charge, and P2' residues with small size. In contrast, their preferences on secondary structure are diverse. Correlation was found between IBV $3CL^{pro}$ activity and β-sheet propensity at P5 position, while no strong correlation with secondary structure propensities was observed in HCoV-NL63 and HCoV-OC43. Collectively, all $3CL^{pro}$ share universal preferences on charge, side chain volume and hydrophobicity, but not secondary

structure. Their relative activities against universal and specific super-active substrates were elevated to 1.4 – 4.3, showing synergetic effects by combining preferred residues. These substrates were examined by group I HCoV-229E and group IIa HCoV-HKU1 in parallel. Their activities were highly comparable to those of other group members.

摘要

嚴重急性呼吸系統綜合症的冠狀病毒（SARS-CoV）的類 3C 蛋白酶（$3CL^{pro}$）在多蛋白 1a 和 1ab 的自我切割過程中是必需的，因此它是一種治療受冠狀病毒感染的潛在標靶。要徹底了解它對底物的偏好，我們在自我切割序列的 P5 至 P3'位置上進行飽和突變序列，從而創建了一個包含了 19×8 個變種的底物庫。在青色和黃色螢光蛋白之間插入底物序列，令其切割率可被體外螢光共振能量轉移（FRET）監測。找出不同的底物序列的相對切割率和各種結構特性之間的相關性。P5 和 P3 位置傾向高 β-折疊傾向殘基；P4 位置傾向細小和疏水殘基；P2 位置傾向沒有 β-支鏈的疏水殘基；谷氨醯胺是 P1 位置中最佳的殘基，但被組胺酸和蛋胺酸替換後仍可被切割。P1'位置傾向細小殘基，而 P2'和 P3'位置對於替換沒有強烈的偏好。值得注意的是，暴露於溶劑中的位置，如 P5、P3 和 P3'，傾向正電荷殘基多於負電荷殘基，使人聯想到靜電作用可能在催化作用中擔當一個角色。由 P5 至 P3'位置所傾向的殘基而組合成的超活躍底物，其活性比野生序列高 2.8 倍。

抑制 SARS-CoV $3CL^{pro}$ 蛋白酶的活性可減低冠狀病毒粒子的複製率和引起的細胞病變效應。連結腈彈頭的類肽抑制劑透過抑制半胱氨酸蛋白酶的活性，經已發展成用作臨床治療用途。爲了調查腈能否標靶 $3CL^{pro}$，我們合成了一系列以腈爲基礎的類肽抑制劑，當中包含了各種保護基、肽長度和肽序列。利用 FRET 檢測量度它們的效力，以 IC_{50} 和 K_i 值爲依據。大部份以腈爲基礎的抑制劑在微莫耳濃度下，可明顯減低蛋白酶的活性。最具效力的抑製劑是連上苄氧羰基（cbz）的類四肽抑製劑'cbz-AVLQ-CN'，其 IC_{50} 和 K_i 值分別爲

5.9 ± 0.6 微莫耳和 0.62 ± 0.11 微莫耳。從 $3CL^{pro}$-抑製劑複合物的晶體結構中顯示，腈彈頭與半胱氨酸 145 連結，而 P1 至 P4 殘基像底物一樣與 $3CL^{pro}$ 結合。'cbz-AVLQ-CN'當中的 cbz 翻進了一個由谷胺酸 166 至脯胺酸 168 所組成的空洞，提供額外的給合力，從而提高抑製劑的效力。結括而言，以腈爲基礎的 cbz 類肽抑製劑是一個可靠的模型用作藥物發展。

利用由 SARS-CoV $3CL^{pro}$ 自我切割序列所組成的底物庫，調查各種 CoVs 的 $3CL^{pro}$ 的綜合底物特徵。在各種病毒種類之中，HCoV-NL63、HCoV-OC43 和雞傳染性支氣管炎病毒（IBV）被分別選定爲第 I 組、第 IIa 組和第 III 組的代表，爲特徵剖析之用。透過 FRET 檢測量度它們對於 19×8 個變種的蛋白水解率，以及找出與代替殘基的結構特性之間的關聯。像如第 IIb 組的 SARS-CoV $3CL^{pro}$，這些 $3CL^{pro}$ 一貫傾向細小而疏水的 P4 殘基；帶正電荷的 P3 殘基；沒有 β-支鍵的疏水 P2 殘基；P1-谷氨醯胺及細小的 P1' 殘基。它們亦傾向於接納帶正電荷的 P5 和 P3'殘基，和細小的 P2'殘基。相比之下，它們對二級結構的偏好是不同的。IBV $3CL^{pro}$ 活性與 P5 位置的 β-折疊傾向相關，而 HCoV-NL63 和 HCoV-OC43 $3CL^{pro}$ 與二級結構沒有很強的相關性。整體而言，所有 $3CL^{pro}$ 對於電荷、細小支鍵和疏水性的偏好是普遍的，而對二級結構的偏好是分歧的。通用和專用超活躍底物的相對活性爲 1.4 – 4.3，証明了連結傾向殘基可產生協同作用。這些底物同時被第 I 組的 HCoV-229E 和第 IIa 組的 HCoV-HKU1 檢測，它們的活性與同組病毒的活性高度類近。

Table of Contents

Chapter 3 – FRET Assay Establishment

Chapter 4 – Substrate Specificity of SARS-CoV 3CL^{pro}

Chapter 5 – Peptidomimetic Inhibitors Targeting SARS-CoV 3CL^{pro}

Chapter 6 – Broad Substrate Specificity of Various CoV 3CL^{pro}

Chapter 7 – Conclusive Remarks and Future Perspective

List of Figures

LIST OF TABLES

Abbreviations

$3CL^{pro}$	3C-like protease
BCoV	bovine coronavirus
Boc	*tert*-butyl carbamate
Cbz	carbobenzyloxy
CFP	cyan fluorescent protein
CoV	coronavirus
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
FIPV	feline infectious peritonitis virus
FRET	fluorescence resonance energy transfer
HCoV	human coronavirus
His_6-MBP	poly-Histidine-maltose binding protein
His_6-SUMO	poly-Histidine-small ubiquitin-related modifier
IBV	infectious bronchitis virus
IMAC	immobilized metal ion affinity chromatography
K_m'	apparent K_m
MES	2-(*N*-morpholino)ethanesulfonic acid
MHV	mouse hepatitis virus
Miu	5-methylisourea
OD	optical density
PEDV	porcine epidemic diarrhea virus
PEG	polyethylene glycol
PHEV	porcine hemagglutinating encephalomyelitis virus
SARS	severe acute respiratory syndrome
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TCEP	tris(2-carboxyethyl)phosphine
TGEV	transmissible gastroenteritis virus
WT	wild-type
YFP	yellow fluorescent protein

Acknowledgements

I would like to express my deepest thanks to my supervisors, Professor Kam-Bo Wong and Professor David Chi-Cheong Wan, for providing me valuable advices, encouragement and an atmosphere of research freedom. Also, I express my acknowledgement to my labmates in SC151, especially Mr. Ivan Yu-Hang Fong and Mr. Kwok-Ho Chan, who kindly shared a lot of ideas and experiences on crystallization and data processing. I am grateful to Mr. Chao Chen for the inhibitor synthesis and Mr. Oscar Wan-Gang Gu for helping in the *in vivo* cell-based assay development. Last but not least, I dedicate this dissertation to my family and friends to thank for their support and patience. They made every enterprise to devote their time and effort to my life.

Chapter 1 – Introduction

1.1 SARS-CoV

Severe acute respiratory syndrome (SARS) is a novel pneumonia infected more than 8,000 patients in 2003 (Kuiken *et al.*, 2003). Due to deficiency of knowledge, the patients were treated by non-specific drugs such as interferons and ribavirin, leading to 916 deaths eventually (Cinatl *et al.*, 2005). Its causative agent was a spherical virion with 60 – 140 nm of diameter named SARS-coronavirus (SARS-CoV) (Ksiazek *et al.*, 2003; Tse *et al.*, 2004). Genetic materials and complex projections were found in centre and on surface respectively. Single-stranded RNA viral genome encodes two polyproteins consisted of 15 non-structural proteins (Figure 1.1) (Marra *et al.*, 2003; Rota *et al.*, 2003; Thiel *et al.*, 2003). Most of these non-structural proteins are responsible for virion replication. Activation of these proteins requires autocleavages by 3C-like protease ($3CL^{pro}$) and papain-like protease.

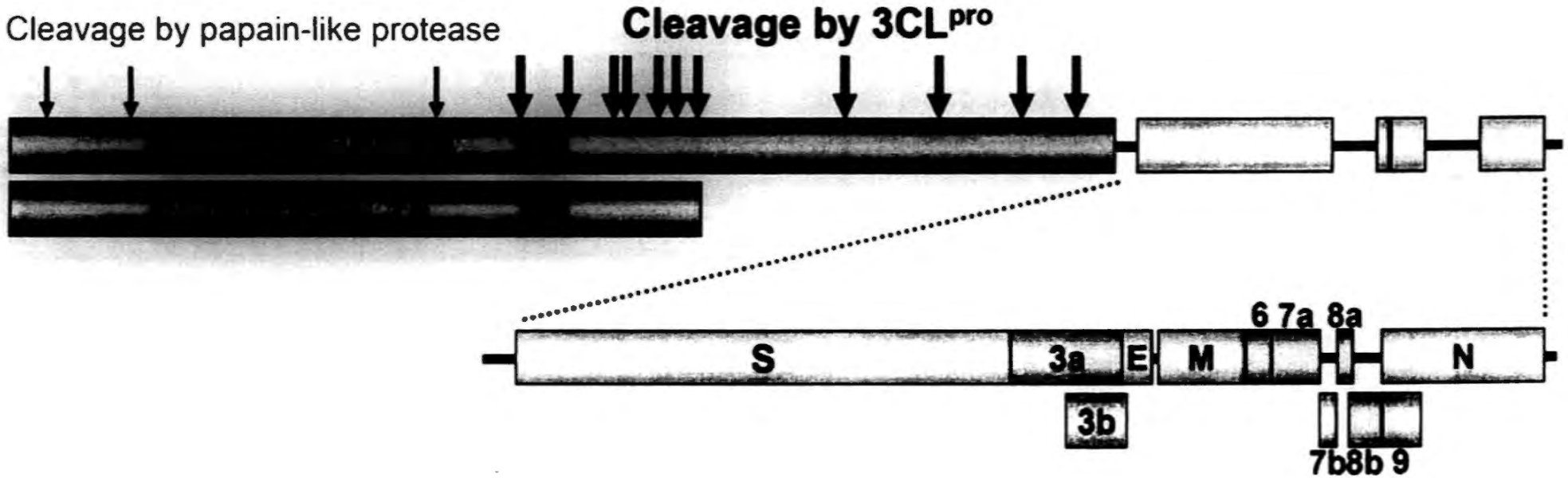


Figure 1.1 – Genomic organization of SARS-CoV. Single-stranded plus sense RNA of 30 kilobase pairs encodes two polyproteins (red), four structural proteins (green) and nine accessory proteins (blue). Cleavages of polyproteins by $3CL^{pro}$ (black arrows) and papain-like protease (grey arrows) activate other non-structural proteins for virion replication. Darken domains represent position of $3CL^{pro}$ in polyproteins.

1.2 SARS-CoV $3CL^{pro}$

1.2.1 Proteolysis by Cys and His Dyad

The $3CL^{pro}$ is a Cys protease that cleaves peptide bond between P1 and P1' positions by Cys145 and His41 dyad (Yang *et al.*, 2003). Substitution of either Cys145 or His41 to Ala resulted in loss of proteolytic activity (Chen *et al.*, 2005b; Huang *et al.*, 2004). During catalysis, Cys145 is at first ionized by donating H atom to His41, followed by bonding with P1 carbonyl group by nucleophilic attack (Figure 1.2) (Solowiej *et al.*, 2008). The P1 – P1' peptide bond is then broken to release C-terminal substrate. Afterwards, intermediate is deacylated by general base catalytic mechanism that His41 accepts H atom from water molecule, and OH group reacts with the P1 carbonyl group to release N-terminal substrate.

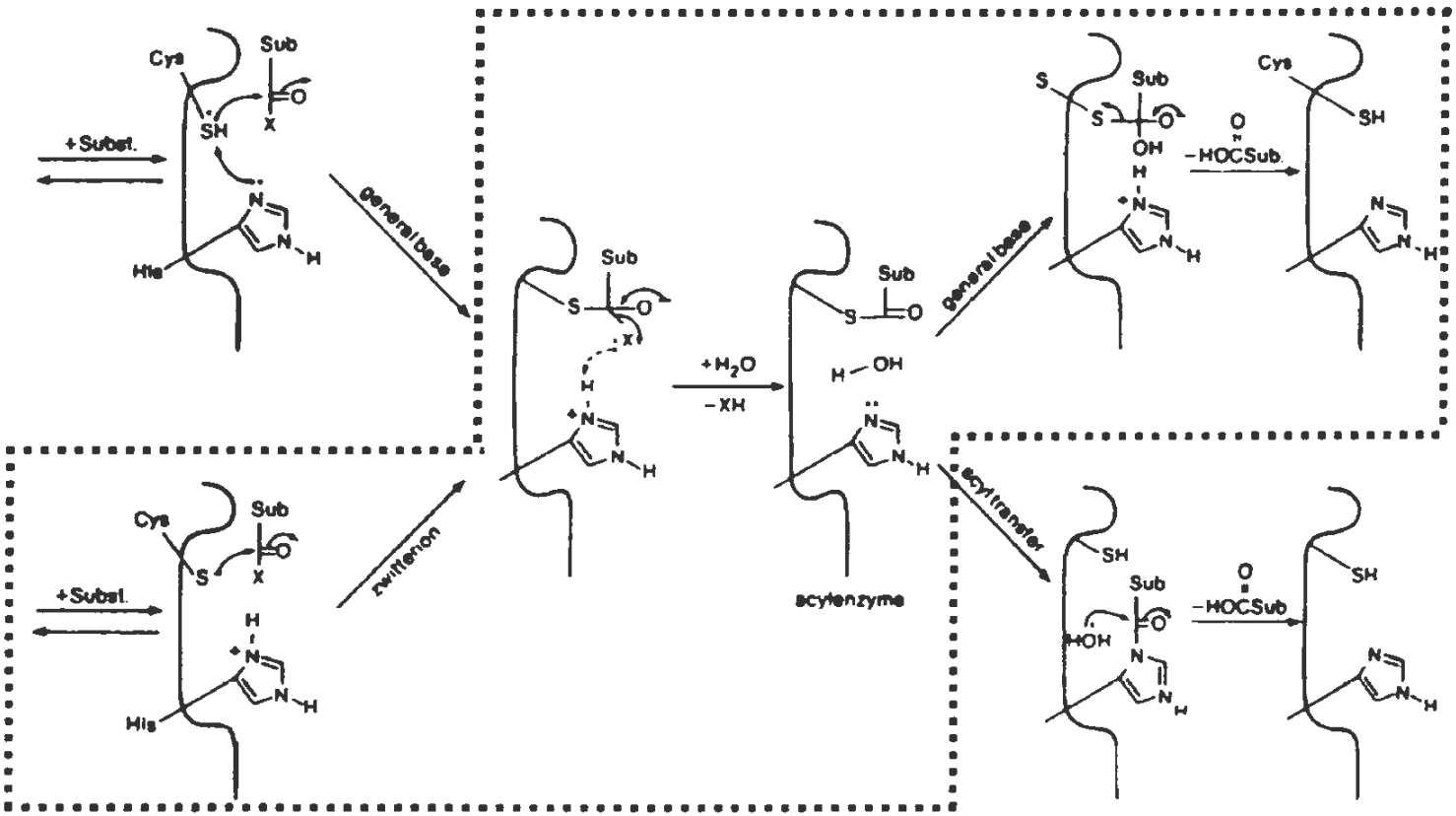


Figure 1.2 – Catalytic mechanism of SARS-CoV 3CL^{pro}. The schematic diagram adapted from street *et al.* shows general base catalytic mechanism and thiolate-imidazolium zwitterionic pair catalytic mechanism of Cys protease (Street *et al.*, 1985). During catalysis, acylation is initiated by the zwitterionic pair, while deacylation is catalyzed by general base mechanism. Highlighted pathway indicates 3CL^{pro} catalytic mechanism.

1.2.2 Homodimer Consisted of Three Domains

Native $3CL^{pro}$ is a homodimer. Each protomer of 34 kDa is divided into three chymotrypsin-like domains (Figure 1.3) (Bacha *et al.*, 2004; Hsu *et al.*, 2005; Yang *et al.*, 2003). Domain I (residue 8–101) and II (residue 102–184) form a substrate-binding cleft, while domain III (residue 201–303) interacts with each others for dimerization. The $3CL^{pro}$ is active only in dimeric form (Lin *et al.*, 2008; Shi *et al.*, 2008).

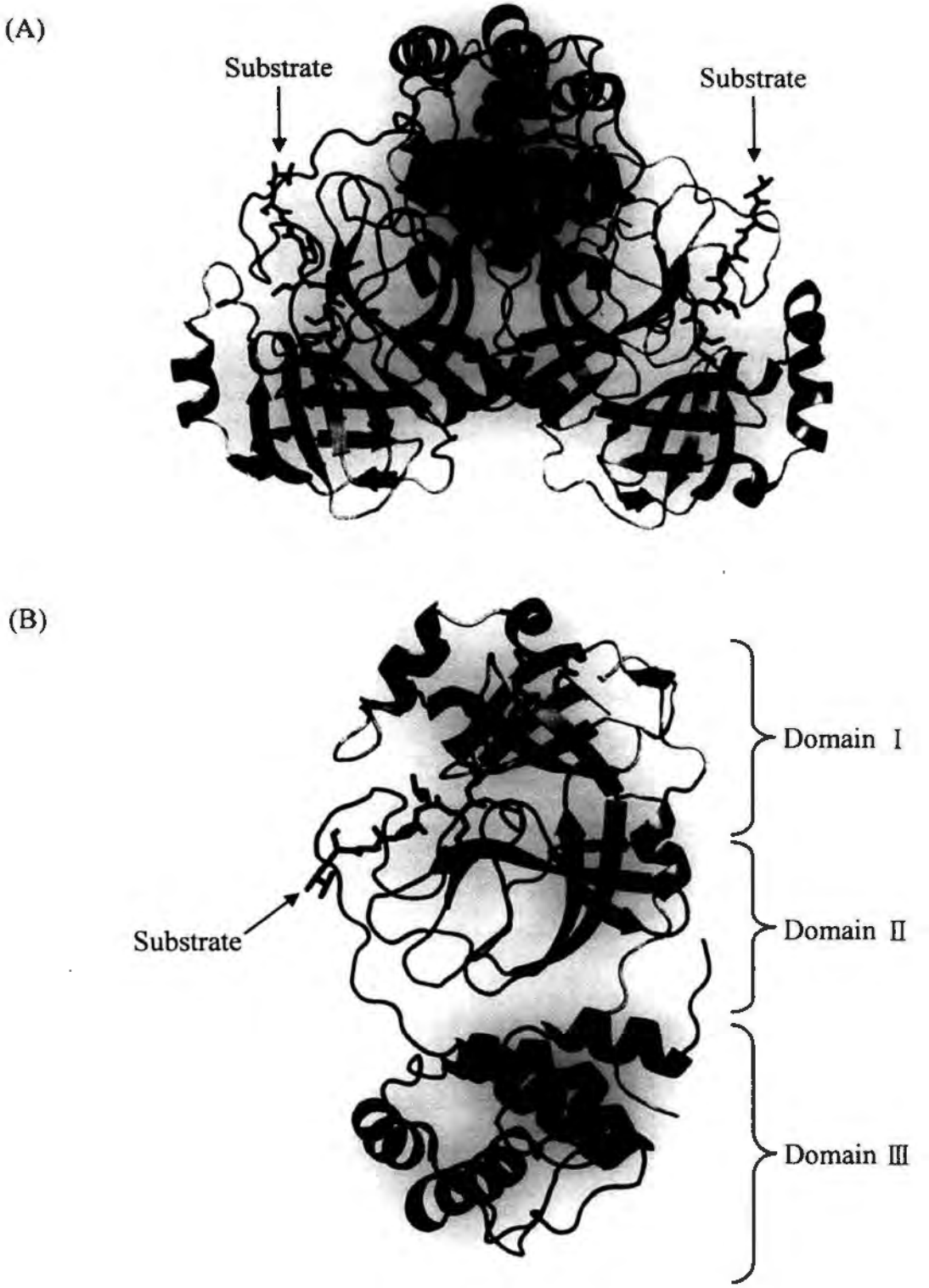


Figure 1.3 – SARS-CoV 3CL^{pro} is homodimer. **(A)** One native 3CL^{pro} molecule comprises two protomers (blue and green), which individually interacts with one substrate (red). **(B)** Protomers are divided into three chymotrypsin-like domains. Domain I (pale blue) and II (blue) form a substrate binding cleft, while domain III (purple) interacts with each other for dimerization (PDB: 2Q6G) (Xue *et al.*, 2008).

1.3 Substrate Specificity of SARS-CoV $3CL^{pro}$

1.3.1 Conservation of Cleavage Site Sequences

Gln is absolutely conserved at P1 position among 11 of $3CL^{pro}$ cleavage sites in the polyproteins (Figure 1.4) (Kiemer *et al.*, 2004; Thiel *et al.*, 2003). P2 position accommodates hydrophobic residues with large side chains such as Leu and Phe, while P1' position tolerates small residues such as Ala and Ser. Autocleavage site at N-terminus of $3CL^{pro}$ (TSAVLQSGFRKM) is the most preferred substrate sequence among these cleavage sites (Fan *et al.*, 2004). $3CL^{pro}$-substrate complex exhibited hydrogen bonding interaction between $3CL^{pro}$ and substrate residues at P4 – P2' positions (Figure 1.5) (Xue *et al.*, 2008). Substituting P5 – P3' residues was found to significantly alter $3CL^{pro}$ activity (Fan *et al.*, 2005); however, comprehensive studies on substrate specificity at these positions are scarce.

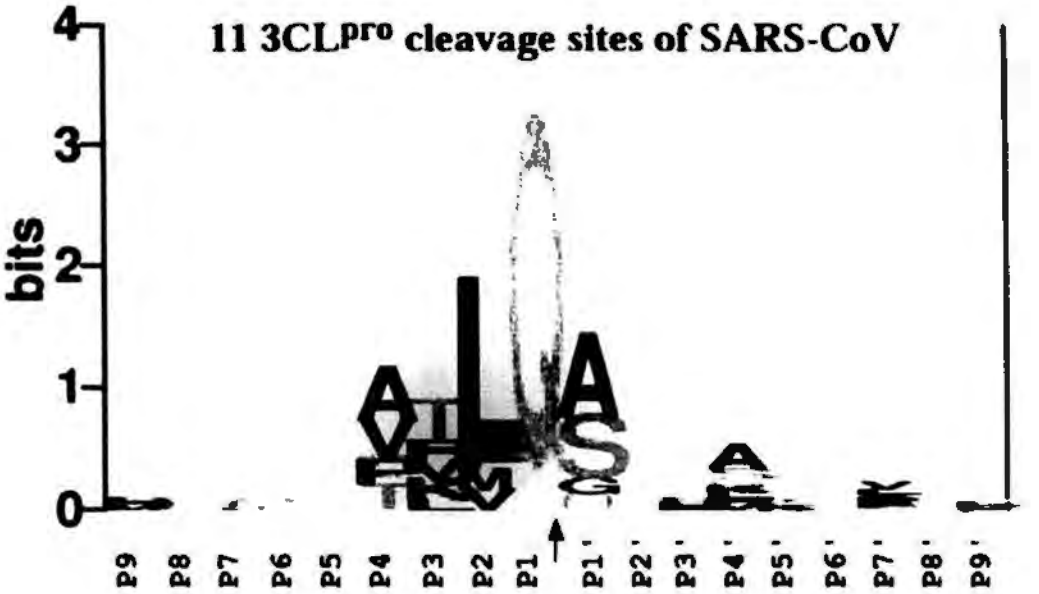


Figure 1.4 – Conservation of 11 cleavage sites of SARS-CoV 3CL^{pro}. Heights of residues represent their frequencies (Thiel *et al.*, 2003). Gln is always found at P1 position. P2 position accommodates Leu, while P1' position tolerates Ala and Ser.

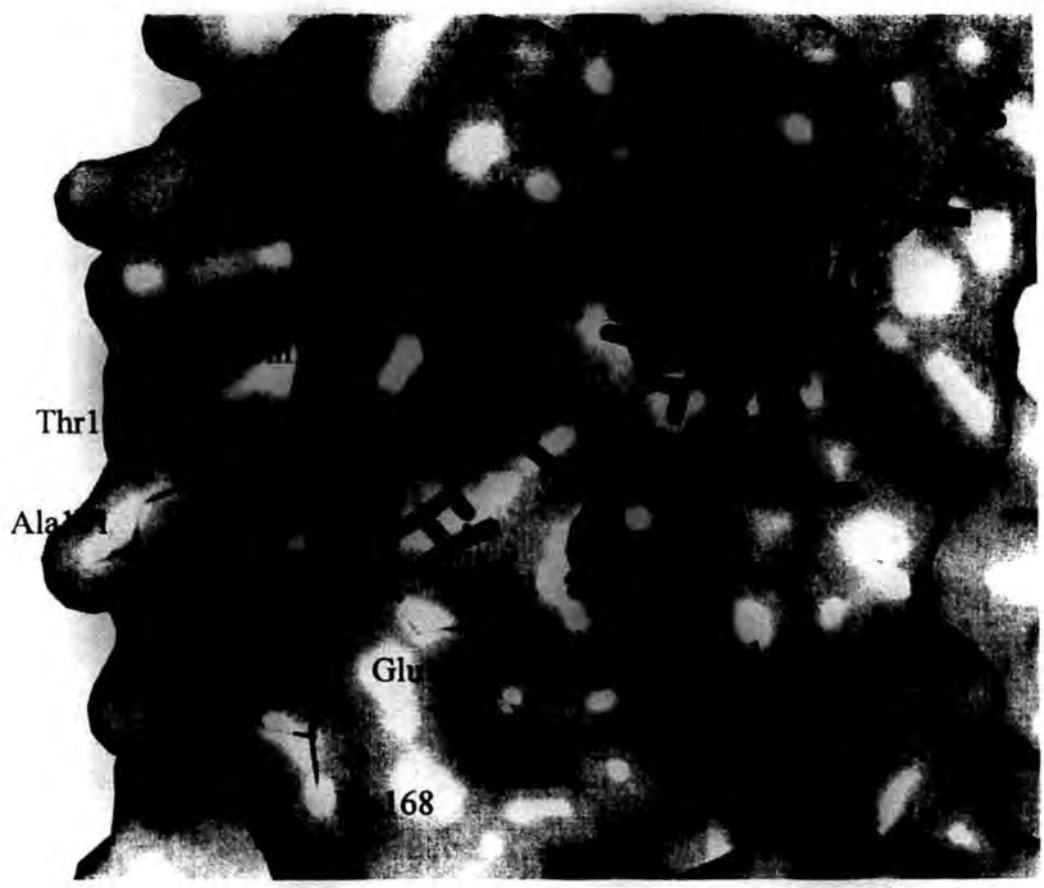


Figure 1.5 – Interaction between SARS-CoV $3CL^{pro}$ and autocleavage sequence. Autocleavage sequence TSAVLQ↓SGFRK (red stick) interacted with substrate binding cleft of $3CL^{pro}$ (grey surface). Side chains of P4, P2, P1 and P1' residues inserted into cavities, while those of P5, P3 and P3' residues were solvent-exposed. P4 to P2' residues hydrogen-bonded with residues (purple lines) of the $3CL^{pro}$ (PDB: 2Q6G) (Xue *et al.*, 2008).

1.3.2 Drawbacks of Previous Studies

Chu *et al.* synthesized peptide substrates with single residue substitution at each of the P4, P3, P2, P1', P2' and P3' positions (Chu *et al.*, 2006). The cleavage of these peptide substrates by the $3CL^{pro}$ was detected by mass spectrometry. They showed that cleavage was detected only when Leu and Phe were present at P2 position. Peptide substrates with acidic residues at P1' position and with Ile/Leu at P2' position were not cleavable. All substitutions at P4 and P3' positions resulted in cleavable substrates. Their detection method can only determine qualitatively whether the peptide substrate is cleavable, but not the relative cleavage rate of different substitutions. Goetz *et al.* profiled the specificity at P4 to P1 positions using a fully degenerate library of tetrapeptides linked with fluorogenic groups at the C-terminus (Goetz *et al.*, 2007). Contradictory to common belief that P1 position only takes Gln, they showed that the $3CL^{pro}$ can cleave the peptide substrates containing His at P1 position equally well. However, in their hands, peptide substrates with Phe at P2 position have no observable cleavage, which is inconsistent with the observation that Phe is naturally occurring at this position of the autocleavage sequence of polyproteins. It is, therefore, unknown whether the tetrapeptide is a good model for substrate specificity for the $3CL^{pro}$.

1.3.3 Comprehensive Specificity Profiling by Using the Protein-Based Substrate Library

To determine a comprehensive-preference, we first selected a 12-residue autocleavage sequence (TSAVLQSGFRKM) located at the $3CL^{pro}$ N-terminus as wild-type (WT) substrate model. This sequence has been broadly applied to $3CL^{pro}$ studies, because it is the most active among 11 known cleavage sites, and these 12 residues covers all essential amino acids for protease-substrate interaction. Next, saturation mutatgenesis was performed at P5 to P3' positions in order to generate a substrate library of 19×8 variants. 20 substrates were included at each position, so the sample size is sufficient to deduce the positional preferences, and thus the overall $3CL^{pro}$ preference. To measure the proteolytic rate on these substrates, each autocleavage sequence was linked to cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP). Cleavage on substrate decreased in fluorescence intensity at 530 nm emitted from YFP. Reaction rate could hence be determined by monitoring fluorescence change. The fluorescent proteins were excluded from the autocleavage sequence; minimizing undesirable influence on the proteolysis during protease-substrate interaction. In addition, change in $3CL^{pro}$ termini may affect the protease activity as well as substrate specificity, hence the $3CL^{pro}$ was produced by factor Xa restriction at SGIEGR↓SGFRKM so that protease with native termini was produced.

1.4 Inhibitor targeting SARS-CoV $3CL^{pro}$

1.4.1 A Potential Drug against SARS

Autoprocessing by $3CL^{pro}$ releases other non-structural proteins for virion replication, suggesting that inhibiting $3CL^{pro}$ proteolysis disrupts viral life cycle, and is a convincing strategy against CoV-induced diseases. The protease inhibition has been verified to reduce production rates of viral RNA as well as proteins, and CoV-induced cytopathic effects (Chen *et al.*, 2005a; Li *et al.*, 2005; Wu *et al.*, 2004; Yang *et al.*, 2005).

1.4.2 Targeting 3CL^{pro} by Peptidomimetic Inhibitors

A number of inhibitors targeted SARS-CoV 3CL^{pro} have been discovered or synthesized, including chemical compounds, natural compounds and peptidomimetic inhibitors. Peptidomimetic inhibitor is a substrate-like peptide linked with a warhead. The peptide binds to substrate-binding cleft of the protease, while the warhead is responsible for inactivating the catalytic site. Peptidomimetic inhibitors can inhibit SARS-CoV 3CL^{pro} activity in both of *in vitro* and *in vivo* conditions with low cytotoxicity against host cells (Lee *et al.*, 2009; Lee *et al.*, 2005; Yang *et al.*, 2005; Yin *et al.*, 2007). Furthermore, peptidomimetic inhibitors have been used as drugs for clinical treatment of autoimmune diseases, neurodegenerative diseases and parasitic diseases (Hauff *et al.*, 2005; Randolph & DeGoey, 2004; Steverding *et al.*, 2006). Hence, peptidomimetic inhibitor is a convincing model to be developed into drugs against SARS.

1.4.3 Nitrile Warhead was Applied for Inhibition

The warhead plays an important role in potency of the overall inhibitor. Nitrile group is an electrophile and a well-known warhead targeted active sites of Cys proteases by formation of reversible covalent bond (Figure 1.6) (Brisson *et al.*, 1986; Moon *et al.*, 1986). This bioactive warhead has been broadly applied for development of inhibitors and drugs against cathepsins-related diseases, such as cancer, autoimmune disorders and osteoporosis (Black; Frizler *et al.*). Thereby, we aim at investigating whether nitrile-based peptidomimetic inhibitor can inhibit SARS-CoV $3CL^{pro}$ activity, and is a good model for further drug development. We produced a series of inhibitors with various protective groups, peptide length and peptide sequences. Their inhibitor potency and binding affinity in terms of IC_{50} and K_i values respectively were determined by fluorescence resonance energy transfer (FRET) assays. Potent inhibitors were soaked into $3CL^{pro}$ crystals for structural determination of protease-inhibitor complexes, which helped in investigating how inhibitors target with the $3CL^{pro}$ and structure-potency relationship.

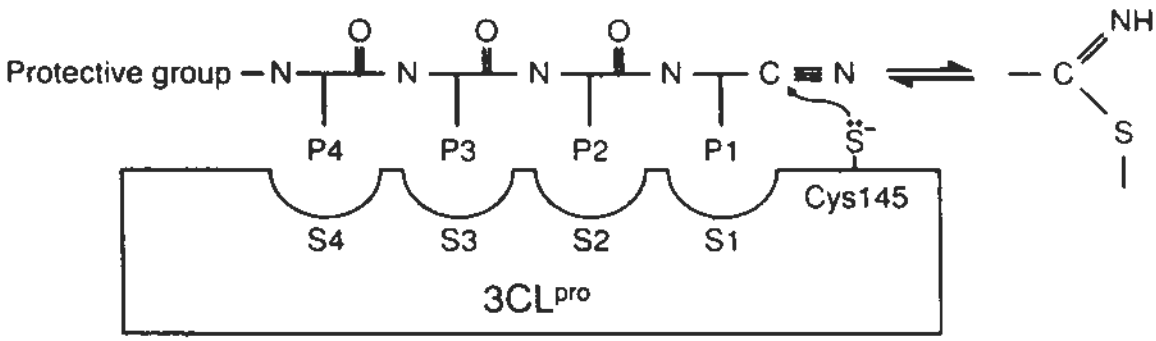


Figure 1.6 – Proposed inhibition mechanism of nitrile-based peptidomimetic inhibitor. Peptide of the inhibitor binds to substrate binding cleft of the $3CL^{pro}$ (grey). Electrophilic nitrile group reacts with nucleophilic Cys145 to form a reversible covalent bond. Bonded Cys145 loses ability to attack substrate, leading to protease inactivation. 'P1 to P4' represent side chains of residues at these positions, while 'S1 to S4' represent $3CL^{pro}$ subsites.

1.5 Substrate Specificity of $3CL^{pro}$ of Various CoVs

1.5.1 Similarities among Various $3CL^{pro}$

A number of CoVs have been discovered from human, farm animals and domesticated pets. Based on $3CL^{pro}$ sequences, they are classified into three major groups (Figure 1.7) (Rota *et al.*, 2003). At least five strains of human-infected CoVs have been found in group I and II. SARS-CoV $3CL^{pro}$ (group IIb) shared 40 – 50 % of protein sequence similarity with other $3CL^{pro}$.

$3CL^{pro}$ structures of HCoV-229E (group I), transmissible gastroenteritis virus (TGEV) (group I), HCoV-OC43 (group IIa), SARS-CoV (group IIb) and IBV (group III) have been solved, and are structurally comparable (Anand *et al.*, 2003; Xue *et al.*, 2008; Zhao *et al.*, 2008). All of them are homodimer consisted of three domains. Domain I and II form substrate binding cleft, while domain III is responsible for dimerization. All $3CL^{pro}$ consist of Cys and His dyad, indicating that they share the common catalytic mechanism. Superimposing their structures showed that main chains of their substrate binding clefts are highly similar (Figure 1.8). Some of the residues within 6 Å of SARS-CoV autocleavage sequence are conserved (Figure 1.9). Moreover, high conservations at P2 to P1' positions have been found among cleavage sites of various $3CL^{pro}$ (Figure 1.10) (Hegyi & Ziebuhr, 2002; Kiemer *et al.*, 2004; Thiel *et al.*, 2003). Gln is absolutely present at P1 position. Leu and small residues are frequently found at P2 and P1' positions respectively.

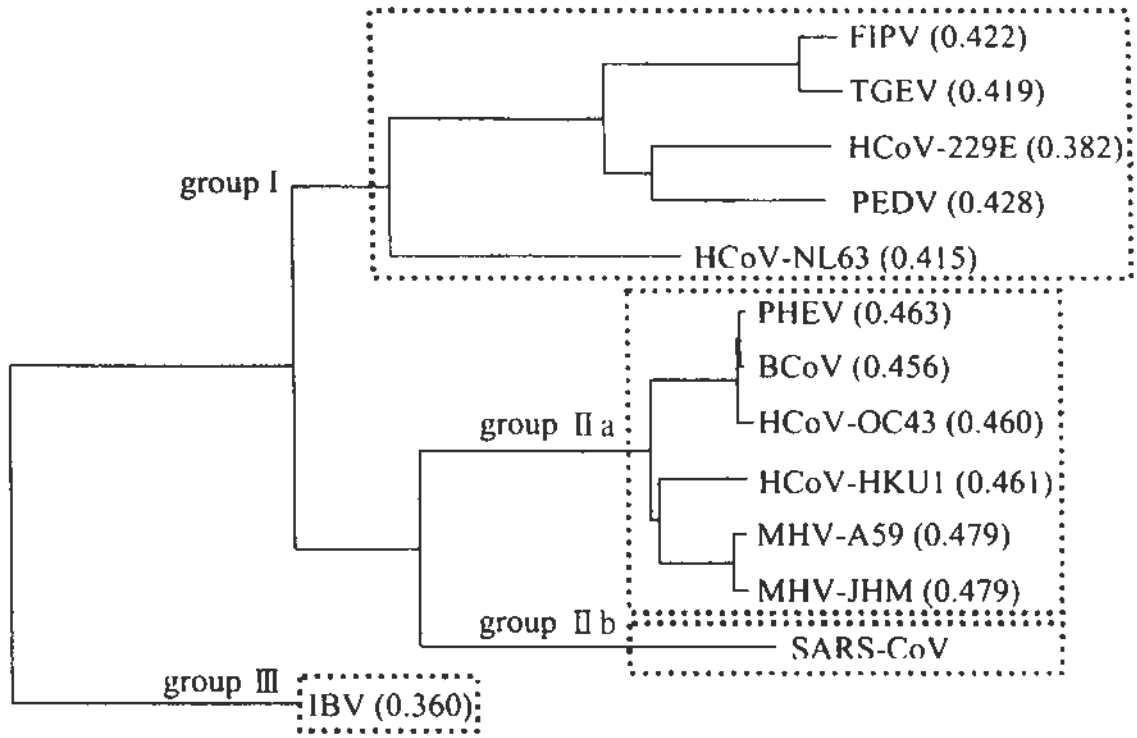


Figure 1.7 – 3CL^{pro} were classified into three major groups. 3CL^{pro} sequences of 13 of CoVs were aligned to obtain a phylogenetic tree. They were classified into three major groups. The CoVs include feline infectious peritonitis virus (FIPV) (GeneID: 5075762), TGEV (GeneID: 1724654), HCoV-229E (GeneID: 918764), porcine epidemic diarrhea virus (PEDV) (GeneID: 935181), HCoV-NL63 (GeneID: 2943501), porcine hemagglutinating encephalomyelitis virus (PHEV) (GeneID: 3882095), bovine coronavirus (BCoV) (GeneID: 921688), HCoV-OC43 (GeneID: 5896646), HCoV-HKU1 (GeneID: 3200429), mouse hepatitis virus (MHV)-A59 (GeneID: 1489749), MHV-JHM (GeneID: 3283258), SARS-CoV and IBV (GeneID: 1489740). The phylogenetic tree was created by computer programs ClusterX 2.0 and PhyloDraw. Numbers in parenthesis indicate sequence identities compared to SARS-CoV 3CL^{pro}.

Figure 1.8 – Structural similarity of substrate binding clefts among $3CL^{pro}$. $3CL^{pro}$ structures of group I members HCoV-229E (PDB: 1P9S, green ribbon) and TGEV (PDB: 1P9U, cyan ribbon), group IIa member HCoV-HKU1 (PDB: 3D23, blue ribbon), group IIb member SARS-CoV (PDB: 2Q6G, red ribbon) and group III member IBV (PDB: 2Q6D, purple ribbon) were superimposed. Autocleavage site (red stick) of SARS-CoV $3CL^{pro}$ indicates location of substrate binding clefts. Conformations of their main chains near the autocleavage site are comparable (Anand *et al.*, 2003; Xue *et al.*, 2008; Zhao *et al.*, 2008).

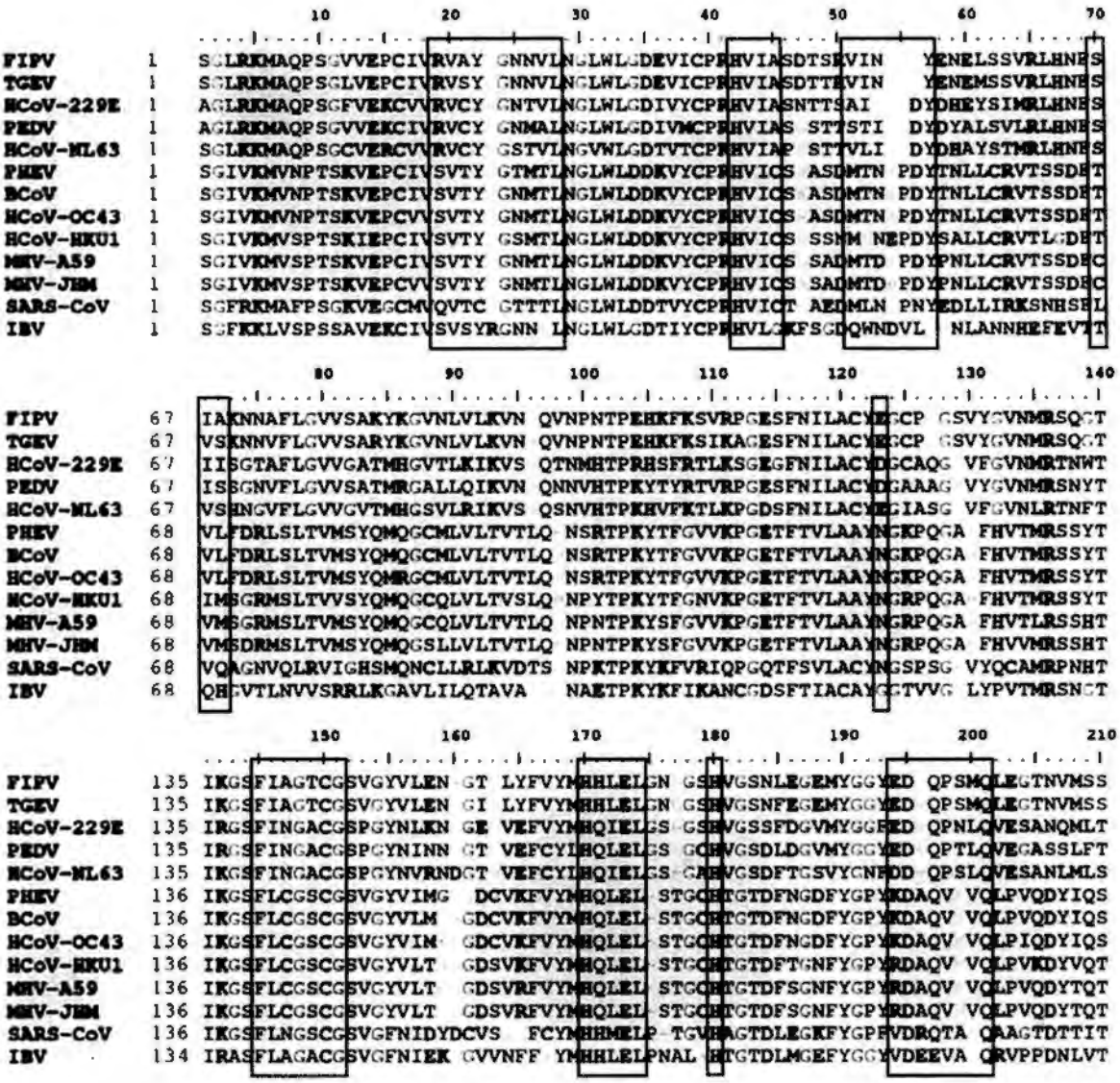


Figure 1.9 – Conservation of substrate binding clefts among 3CL^{pro}. 3CL^{pro} sequences were aligned, and residues within 6 Å of SARS-CoV autocleavage site were highlighted. These sequences include group I members FIPV (GENE ID: 5075762), TGEV (GENE ID: 1724653), HCoV-229E (GENE ID: 918764), PEDV (GENE ID: 935181) and HCoV-NL63 (GENE ID: 2943501), group IIa members PHEV (GENE ID: 3882095), BCoV (GENE ID: 921688), HCoV-OC43 (gb|AAX85666.1|), HCoV-HKU1 (GENE ID: 3200429), MHV-A59 (GENE ID: 1489749), MHV-JHM (GENE ID: 3283258), group IIb member SARS-CoV (GENE ID: 1489680) and group III member IBV (GENE ID: 1489740).

Figure 1.10 – Conservation of cleavage sites among 3CL^{pro}. Frequency of residues in cleavage sites of TGEV (NC_002306), HCoV-229E (NC_002645), PEDV (NC_003436), MHV-A59 (NC_001846), BCoV (AF391542) and IBV (NC_001451) demonstrated that P2 to P1' residues are highly conserved. Gln is always present at P1 position. Hydrophobic and small residues are frequently found at P2 and P1' positions respectively (Thiel *et al.*, 2003).

1.5.2 Profiling of Universal and Specific Specificity among 3CL^{pro}

Previous studies were mainly focused on profiling of substrate specificity of SARS-CoV 3CL^{pro}, while information about that of other CoVs was scarce. Only substrate sequences of various CoVs have been compared (Hegyi & Ziebuhr, 2002; Thiel *et al.*, 2003). We therefore aim at profiling of substrate specificity of these 3CL^{pro}. The high conservation at P2 to P1' positions and structural similarities of substrate binding clefts indicate that other 3CL^{pro} is able to recognize and cleave substrate sequence of SARS-CoV 3CL^{pro}. Comprehensive substrate specificity of other 3CL^{pro} was hence profiled by using the substrate library created by saturation mutatgenesis of autocleavage site at P5 to P3' positions of SARS-CoV 3CL^{pro}. There are a number of 3CL^{pro} with high protein sequence identities divided into three major groups, so HCoV-NL63, HCoV-OC43, SARS-CoV and IBV were selected as representatives of group I, IIa, IIb and III respectively for the investigation. Their relative activities against 19×8 substrate variants were obtained by the FRET assay, and correlated to structural properties of substituting residues. The specificity spectra demonstrated positional preferences of 3CL^{pro} of each group. Besides, universal preferences among various 3CL^{pro} were determined by comparing with their preferences, providing insight in development of 3CL^{pro} inhibitor in broad-spectrum. Universal and specific super-active substrate sequences were produced by combining preferred residues, and were examined by those four CoVs as well as HCoV-229E (group I) and HCoV-HKU1 (group IIa). Results from 3CL^{pro} of two additional HCoVs could verify consistency of 3CL^{pro} preferences in the same group.

1.6 Objectives of the Study

1) To establish FRET assay for measuring $3CL^{pro}$ activity on the protein-based substrate (Chapter 3).

2) To profile comprehensive substrate specificity of SARS-CoV $3CL^{pro}$ at P5 to P3' positions (Chapter 4).

3) To produce nitrile-based peptidomimetic inhibitors, examine their potency on SARS-CoV $3CL^{pro}$ and determine structures of $3CL^{pro}$-inhibitor complexes (Chapter 5).

4) To investigate specificity of $3CL^{pro}$ in other groups (Chapter 6).

Chapter 2 – Materials and Methods

2.1 Materials

2.1.1 Reagents

Table 2.1 – Reagent components.

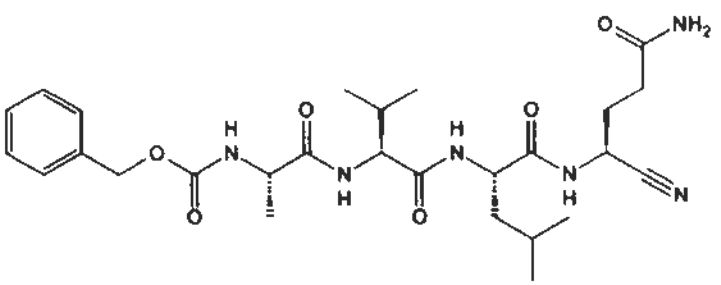
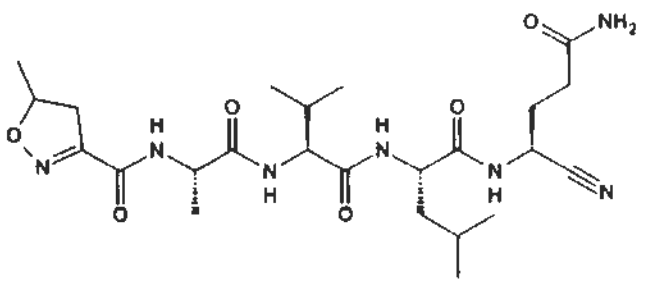
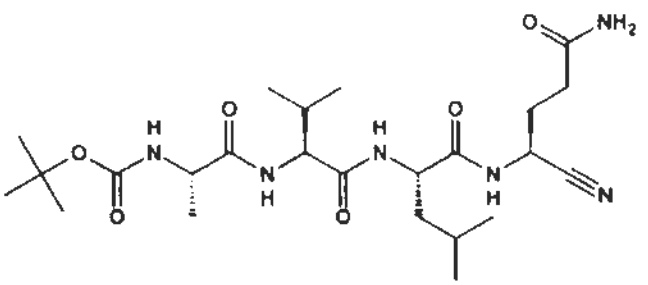
Reagents	Components
1% (w/v) agarose Gel	1 % (w/v) agarose 0.1% (v/v) ethidium bromide TAE buffer
Coomassie blue staining solution	45 % (v/v) EtOH 10 % (v/v) acetic acid 0.24 % (w/v) bromophenol blue
Coomassie blue destaining solution	30 % (v/v) EtOH 10 % (v/v) acetic acid
6X DNA loading dye	0.25 % (w/v) bromophenol blue 0.25 % (w/v) xylene cyanol FF 30 % (v/v) glycerol
Luria-Broth (LB) medium	10 g/L NaCl 10 g/L tryptone 5 g/L yeast extract
LB plate	LB medium 2%(w/v) agar
Mass spectrometry destaining solution	75 mM NH_4HCO_3 40 % EtOH
Matrix solution	10mg/ml cinnamic acid 0.1 % trifluoroacetic acid 50 % (v/v) acetonitrile
RF1 solution	100 mM RbCl 50 mM $MgCl_2$ 30 mM potassium acetate 10 mM $CaCl_2$ 15 % (v/v) glycerol Adjusted to pH 5.8
RF2 solution	10mM 3-(N-morpholino) propanesulfonic acid 10 mM RbCl 75 mM $CaCl_2$ 15 % (v/v) glycerol Adjusted to pH 6.8

(P.T.O)

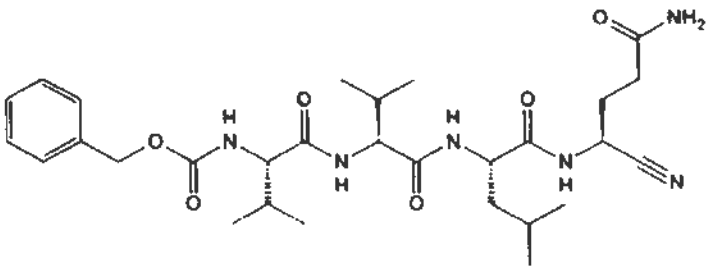
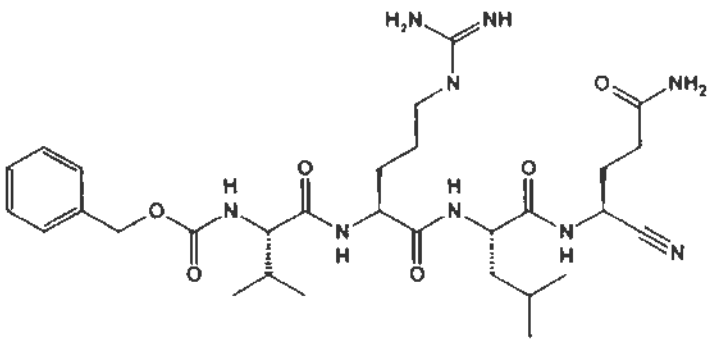
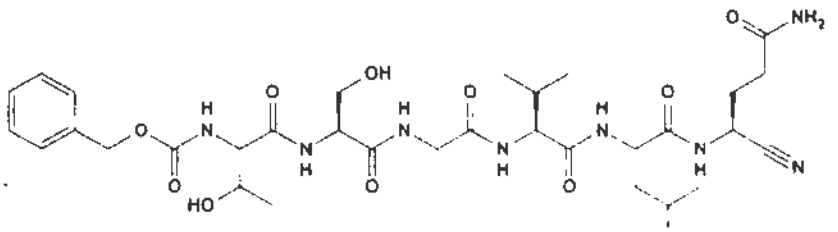
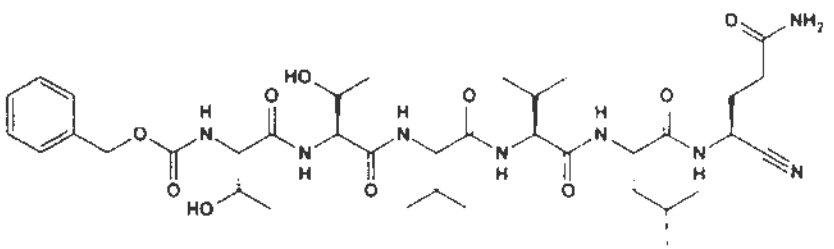
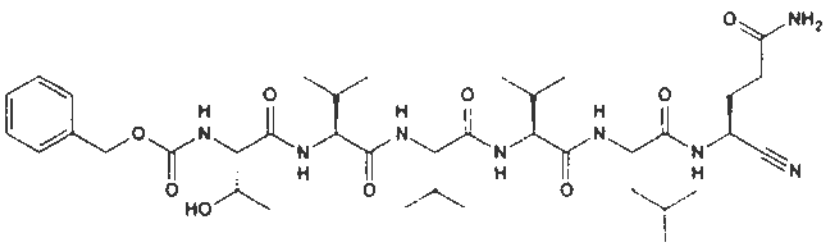
2X SDS loading dye	50 mM Tris-HCl (pH 6.8) 100 mM DTT 2 % (w/v) SDS 0.1 % (w/v) bromophenol blue 5 % (v/v) glycerol
12.5 % or 15 % SDS-polyacrylamide running gel	12.5 % / 15 % (w/v) of acrylamide/bisacrylamide (29:1) 375 mM Tris, pH 8.8 0.1 % (w/v) SDS 0.5 % (v/v) TEMED 1 mM of ammonium perfsulfate
SDS-polyacrylamide stacking gel	5 % (w/v) of acrylamide/bisacrylamide (29:1) 125 mM Tris, pH 6.8 0.1 % (w/v) SDS 0.5 %(v/v) TEMED 1 mM of ammonium perfsulfate
TAE buffer	40 mM Tris-HCl 20 mM acetic acid 1 mM EDTA, pH 8.0
Trypsin solution	80 ng/uL trypsin 10 mM $(NH_4)_2CO_3$

2.1.2 Nitrile-Based Peptidomimetic Inhibitors

Table 2.2 – Chemical structures of nitrile-based peptidomimetic inhibitors. All inhibitors were synthesized by Mr. Chao CHEN, Department of Chemistry, The Chinese University of Hong Kong.

Name	Chemical structure
cbz-AVLQ-CN	
miu-AVLQ-CN	
boc-AVLQ-CN	

(P.T.O.)

cbz-VVLQ-CN	
cbz-VRLQ-CN	
cbz-TSAVLQ-CN	
cbz-TTVVLQ-CN	
cbz-TVVVLQ-CN	

2.2 General Techniques

2.2.1 DNA Manipulation Techniques

2.2.1.1 Competent Cell Preparation

Escherichia coli (*E. coli*) strain DH5α was streaked on LB plate and incubated at 37 ℃ overnight. Single colony was inoculated into 2.5 mL of LB medium and shaken at 37 ℃ overnight. The inoculum was transferred to 250mL of LB medium and shaken at 37 ℃ until mid-log phase. Cells were harvested and chilled at 4 ℃ for 15 minutes in 100mL of RF1 solution. Harvesting and chilling steps were repeated with substitution by 10mL of RF2 solution. Aliquots of competent cells were stored at -80 ℃ for future use. To prepare another strain BL21 (DE3) pLysS, extra 50 mg/L chloramphenicol was added for LB plate and medium for selection. After preparation, transformation of 1 pg standard plasmid pUC19 was performed to verify that transformation efficiency was higher than 10^7 colony-forming unit / μg.

2.2.1.2 Transformation of Plasmids into Competent Cells

DH5α competent cells mixed with plasmid was stayed at 4 ℃ for 10 minutes; heat shocked at 42 ℃ for 2 minutes and stayed at 4 ℃ for another 5 minutes. LB medium was added and shaken at 37 ℃ for 45 minutes for recovery. The cells were spun down and spread to LB plate with 100 mg/L ampicillin, followed by incubating at 37 ℃ overnight. For transformation into BL21 (DE3) pLysS, extra 50 mg/L chloramphenicol was added for LB plate.

2.2.1.3 Plasmid Preparation by Purification Kit

Plasmid preparation procedures referred to the protocol of "DNA-spin™ Plasmid DNA Purification Kit" (Catalog number: INTRON#17093). In brief, plasmid was transformed into DH5α competent cells, followed by inoculating in 5 mL of LB medium and shaking at 37 ℃ overnight. Cells were spun down and resuspended in 250 μL of Resuspension buffer, followed by adding 250 μL of Lysis buffer with mixing gently. After staying for 5 minutes, 350 μL of Neutralization Buffer was added for bacterial protein precipitation. Precipitants were deposited by centrifugation at 13,000 g for 10 minutes. Supernatant was transferred into DNA-spin column, followed by centrifugation at 13,000 g for 30 seconds. The column was washed by 700 μL of Washing buffer B. Centrifugation step was repeated for another 2 minutes to entirely remove solvents. Plasmid was eluted by adding 50 μL of distilled water.

2.2.1.4 DNA Quantification by Measuring OD_{260}

Optical density at 260 nm (OD_{260}) was measured by UV spectrophotometer. Distilled water was used as blank. The DNA concentration was calculated by equation:

$$\text{DNA concentration (ng / μL)} = OD_{260} \times 50$$

OD_{280} was measured in parallel to determine DNA quality. Approximately 1.8 of OD_{260}/OD_{280} ratio revealed that bacterial protein and solvent were absent in the sample.

2.2.1.5 Agarose Gel Electrophoresis and DNA Extraction

DNA sample mixed with 6X DNA loading dye was loaded into well of 1 % (w/v) agarose gel. 100 base pair ladder DNA marker was loaded in parallel as reference. Electricity at voltage of 130 V was supplied for electrophoresis. DNA was visualized under UV light illumination at wavelength of 340 nm.

Target DNA inside agarose gel was sliced and extracted by "DNA gel extraction kit" (Catalog number: H-206). Each sliced gel of 0.2 g was dissolved in 500 μL of Resuspension buffer at 70 ℃, followed by transferring to the column and centrifuging at 13,000 g for 30 seconds. The column was washed by 700 μL of Washing buffer, and centrifuged for additional 2 minutes to completely eliminate residuals. DNA was eluted by 30 μL of distilled water.

2.2.2 Protein Manipulation Techniques

2.2.2.1 SDS-PAGE

Protein sample mixed with 2X SDS loading dye was heated at 100 ℃ for 5 minutes for denaturation. It was then loaded into 12.5 % or 15 % sodium dodecyl sulfate (SDS)-polyacrylamide gel with 0.75 mm thickness. 0.5 μL of low molecular weight protein marker (Catalog number: 17-0446-01) was loaded in parallel as reference. Electricity at current of 60 mA was supplied for SDS-polyacrylamide gel electrophoresis (SDS-PAGE) until dye front reached bottom of the gel. The gel was stained in SDS-PAGE staining solution, followed by destaining in SDS-PAGE detaining solution, till apparent protein bands were visualized.

2.2.2.2 *Protein Quantification by Measuring OD*

Extinction coefficients of $3CL^{pro}$ at 280 nm were determined by submitting protein sequence to "ProtParam" (Table 2.3) (Wilkins *et al.*, 1999). The coefficient of protein-based substrate at 430 nm was experimentally determined by using gel-filtration purified WT substrate. The coefficient of Dabcyl in peptide-based substrate was 15,100 cm^{-1} M^{-1}. OD was measured by Beckman Coulter DU800 spectrophotometer or NanoDrop spectrophotometer. Sample buffer was used as blank. The protein concentration was calculated by equation:

$$\text{Protein Concentration (M)} = \text{OD} / \text{Extinction coefficient } (cm^{-1}\ M^{-1})$$

Table 2.3 – Extinction coefficients of $3CL^{pro}$ and substrates.

Protein	OD at wavelength (nm)	Extinction coefficient (cm^{-1} M^{-1})
SARS-CoV $3CL^{pro}$	280	33,265
HCoV-NL63 $3CL^{pro}$	280	39,880
HCoV-OC43 $3CL^{pro}$	280	42,860
IBV $3CL^{pro}$	280	49,390
HCoV-229E $3CL^{pro}$	280	43,890
HCoV-HKU1 $3CL^{pro}$	280	45,840
Protein-based substrate	280	49,280
Protein-based substrate	430	26,370
DABCYL (Peptide-based substrate)	472	15,100

2.3 Profiling of Substrate Specificity of $3CL^{pro}$

2.3.1 Production of $3CL^{pro}$

2.3.1.1 Plasmid Construction

DNA fragment encoding SARS-CoV $3CL^{pro}$ (GENE ID: 1489680) was amplified by polymerase chain reaction based on SARS-CoV strain CUHK-Su10 sequence (GenBank AY282752) (Tsui *et al.*, 2003), while those encoding $3CL^{pro}$ of HCoV-NL63 (GENE ID: 2943501), HCoV-OC43 (gb|AAX85666.1|), IBV (GENE ID: 1489740), HCoV-229E (GENE ID: 918764) and HCoV-HKU1 (GENE ID: 3200429) were purchased from Mr. Gene Ltd. (Appendix 2.1 – 2.6).

$3CL^{pro}$ coding sequences of SARS-CoV and IBV were treated by restriction enzymes *Sac*I and *Xho*I and cloned into fusion-protein expression vector pET3a, so that $3CL^{pro}$ were tagged with poly-Histidine-maltose binding protein (His_6-MBP) at N-terminus. Factor Xa cleavage sequence was inserted between the $3CL^{pro}$ and tag for subsequent removal of the tag by factor Xa digestion.

$3CL^{pro}$ coding sequences of HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1 were restricted by *Age*I and *Nco*I and ligated into fusion-protein expression vector pRSETA to produce recombinant $3CL^{pro}$ tagged with poly-Histidine-small ubiquitin-related modifier (His_6-SUMO) (GENE ID: 7341) at N-terminus.

2.3.1.2 Expression and Purification

Expression of SARS-CoV $3CL^{pro}$ in BL21(DE3)pLysS was induced by adding 0.4 mM isopropyl β-D-1-thiogalactopyranoside during mid-log phase, while those of other CoVs were induced by 0.1 mM isopropyl β-D-1-thiogalactopyranoside. SARS-CoV $3CL^{pro}$ was expressed at 37 ℃ for 4 hours, while others were expressed at 25 ℃ overnight.

For substrate specificity profiling of SARS-CoV $3CL^{pro}$, the recombinant protein was released by sonication in buffer A (20 mM Tris, 20 mM NaCl, pH 7.8) with 10mM imidazole. Soluble fraction was subjected to immobilized metal ion affinity chromatography (IMAC) column and the recombinant protein was eluted by buffer A with a gradient of 10 – 300 mM imidazole. The His_6-MBP tag was removed by factor Xa digestion in 20 mM Tris, 50 mM NaCl, 2 mM $CaCl_2$, pH 7.4 at room temperature overnight, and by IMAC. The $3CL^{pro}$ was purified by G75 size exclusion column and stored in buffer A. Elution profile showed that the protease was dimeric in solution. OD_{280} was measured for protein quantification.

For substrate specificity profiling of various CoV $3CL^{pro}$, the proteins were purified and stored in buffer B (20 mM Tris, 150 mM NaCl, 1mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.8). His_6-SUMO tag was removed by SUMO protease. Other purification steps were same as procedures mentioned above.

2.3.2 Production of Protein-Based Substrate Library

2.3.2.1 Plasmid Construction

Fusion protein plasmid pET3a encoding the recombinant WT substrate (His_6-CFP-TSAVLQSGFRKM-YFP), which consisted of 12-residue autocleavage sequence inserted between CFP and YFP, was constructed (Appendix 2.7).

The plasmid was mutated by overlapping PCR and quikchange for producing substrate variants. For overlapping PCR, forward primers included DNA fragments encoding mutated autocleavage sequences at 5' overhang, so the resulting products had both DNA sequences of the mutated autocleavage sequences and YFP (Appendix 2.8 – 2.9). The *BamH*I and *Kpn*I treated products were subcloned into the fusion protein vector. For quikchange, primer sequences were designed with aid of "Tm calculator" (http://www.stratagene.com/qpcr/tmcalc.aspx) (Appendix 2.10). Amplified plasmids were restricted by *Dpn*I and transformed into DH5α.

2.3.2.2 Expression and Purification

The substrates were expressed by supply of 0.1 mM isopropyl β-D-1-thiogalactopyranoside and shaking at 22 ℃ overnight. They were purified by IMAC and dialyzed in buffer A or B. OD_{430} was measured for determining substrate concentration.

2.3.3 Determining SARS-CoV $3CL^{pro}$ Activity

2.3.3.1 FRET Assay

For substrate specificity profiling of SARS-CoV, 35 μM of the recombinant substrate was rapidly mixed with 1 – 4 μM of $3CL^{pro}$ in 96-well black Optiplate. Cleavage of the substrate was performed in buffer A was monitored by FRET using EnVision 2101 Multilabel Reader. The reaction mixture was excited by light passing though a 430 nm filter (bandwidth at 8 nm), while the intensity of emitted fluorescence passing though a 530 nm filter (bandwidth at 10 nm) was recorded in every 30 seconds. For Cys variants, 2.5 mM TCEP was added for preventing disulphide bond formation. The assay for each substrate was performed in triplicate. Substrate variants with relative activities of < 0.3 and < 0.1 were further examined by 2 – 8 μM and 4 – 16 μM of $3CL^{pro}$. Super-active substrate variants were examined by 1 – 1.75 μM of $3CL^{pro}$.

Substrate specificity profiling of various CoV $3CL^{pro}$ were determined by using 0.5 μM, 1 μM, 2 μM and 1 μM of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV respectively. Activity on universal and specific super-active substrates was determined by 0.2 – 0.5 μM, 0.1 – 0.25 μM, 0.2 – 0.5 μM, 0.2 – 0.5 μM, 0.4 – 1 μM and 0.4 – 1 μM of $3CL^{pro}$ of HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1, SARS-CoV and IBV respectively. Cleavage was performed in buffer B.

2.3.3.2 *Determining Specific Activity and Relative Activity*

Observed rate constant, k_{obs}, was obtained by fitting the emitted fluorescence at 530 nm to single exponential decay equation:

$$F_{S-P} \times e^{(-k_{obs} \times t)} + F_P$$

where F_{S-P} is fluorescence intensity difference between substrate and product, t is time in terms of minutes and F_P is fluorescence intensity of product. The fitting process was performed by computer program "KaleidaGraph 4.0". Graph of k_{obs} against $3CL^{pro}$ concentration was plotted. Specific activity, $k_{obs}/[3CL^{pro}]$, on the substrate was determined by measuring slope of the graph.

Average of specific activity on substrate variant, A_{VAR}, was normalized against average of specific activity on WT substrate, A_{WT}, to obtain the relative activity:

$$\text{Relative activity} = \frac{A_{VAR}}{A_{WT}}$$

Relative standard deviation (SD) was determined by equation:

$$\text{Relative SD} = \frac{A_{VAR}}{A_{WT}} \times \sqrt{\left(\frac{SD_{VAR}}{A_{VAR}}\right)^2 + \left(\frac{SD_{WT}}{A_{WT}}\right)^2}$$

where SD_{VAR} and SD_{WT} are SD of substrate variant and WT substrate respectively.

2.3.3.3 N-terminal Sequencing

Proteins in SDS-polyacrylamide gel without staining were transferred to polyvinylidene fluoride membrane by supply of electricity at voltage of 20 V for 30 minutes. The membrane was stained by coomassie blue staining solution for 5 minutes, and destained by distilled water until visualizing protein band. Target bands on the membrane were sliced and sequenced by Precise Peptide Sequencing System 492, which was operated by Ms. Doris Ho, Department of Biochemistry, The Chinese University of Hong Kong.

2.3.3.4 Mass spectrometry

Target bands in SDS-polyacrylamide gel were isolated and sliced into small pieces. Samples were destained by shaking in 1 mL of coomassie blue destaining solution until colourless. After washing by nano-pure water, they were incubated two times in 100 μL of 200 mM NH_4HCO_3 for 10 minutes, followed by 100 μL of acetonitrile for 5 minutes. The dried gel slices were immersed in 10 μL of trypsin solution on ice for 30 minutes. 10 mM $(NH_4)_2CO_3$ was added until the slices were just covered, followed by incubation at 30 ℃ overnight for trypsin digestion. 2 μL of supernatant was transferred to Matrix-Assisted Laser Desorption Ionisation sample plate. 0.5 μL of matrix solution was added for crystallization. The samples were subject to ABI 4700 MALDI-ToF-ToF Tandem mass spectrometry system, which was operated by Mr. Denis Ip, Department of Biochemistry, The Chinese University of Hong Kong. Results were analyzed and exported by computer program "GPS explorer"

2.3.4 Correlating with Structural Properties of Substituting Residues

The relative activity was correlated with scales for structural properties of substituting residues, including side chain volume, hydrophobicity, and α-helix and β-sheet propensities (Appendix 2.11) (Chou & Fasman, 1978; Kyte & Doolittle, 1982; Lee *et al.*, 2008). Coefficients and p-values of the correlations were obtained.

2.4 Development of Peptidomimetic Inhibitors

2.4.1 Determining Inhibitor Potency

2.4.1.1 IC_{50} Determination

Inhibitors at various concentrations were pre-mixed with $3CL^{pro}$, followed by determining the specific activities by FRET assay. These specific activities were normalized against that of $3CL^{pro}$ without inhibitor to determine relative activity. Graph of relative activity against inhibitor concentration in log scale was plotted. IC_{50} was obtained by fitting four parameter logistic equation:

$$\text{Relative activity} = \frac{1}{1 + \left(\frac{[\text{Inhibitor}]}{IC_{50}}\right)^{\text{Slope}}}$$

2.4.1.2 K_m Determination

Peptide-based substrate, DABCYL-KTSAVLQSGFRKME-EDANS, with > 95 % of purity was synthesized by GL Biochem (Shanghai) Ltd. Its concentration was determined by measuring OD_{472}. 25 – 50 μM of the peptide-based substrate was rapidly mixed with 1 μM of the $3CL^{pro}$ in 96-well black Optiplate. The reaction condition was 20mM Tris, 20mM NaCl, 2.5 % (v/v) dimethyl sulfoxide (DMSO), pH 7.8. Cleavage was monitored by FRET using EnVision 2101 Multilabel Reader. The reaction mixture was excited by light passing though 340 nm filter (bandwidth at 60 nm), while the intensity of emitted fluorescence passing though 486 nm filter (bandwidth at 10 nm) was recorded in every 30 seconds. The assay for each substrate was performed in triplicate.

k_{obs} was obtained by fitting the emitted fluorescence at 486 nm to single exponential decay equation. Reaction rate was determined by multiplying k_{obs} by

substrate concentration. K_m was then obtained by fitting kinetic from graph of reaction rate against substrate concentration by the Michaelis-Menten equation:

$$\text{Reaction Rate} = \frac{k_{cat} \times [\text{Substrate}]}{[\text{Substrate}] + K_m}$$

where k_{cat} and K_m are catalytic constant and substrate concentration at half-maximal velocity respectively.

2.4.1.3 K, Determination

FRET assay was performed by using the $3CL^{pro}$ pre-mixed with 1 - 5 μM of 'cbz-AVLQ-CN', 7 – 35 μM of 'miu-AVLQ-CN', 10 - 50 μM of 'boc-AVLQ-CN' and 2 – 10 μM of 'cbz-TSAVLQ-CN', to determine apparent K_m (K_m') under each of inhibitor concentrations. Graph of K_m' against inhibitor concentration was plotted to yield a straight line. K_i was determined by rearranged equation:

$$K_m' = K_m \times (1 + \frac{[\text{Inhibitor}]}{K_i})$$

$$K_m' = \frac{K_m}{K_i} \times [\text{Inhibitor}] + K_m$$

2.4.2 Structure Determination of $3CL^{pro}$-Inhibitor Complexes

2.4.2.1 Griding and Soaking

Crystal of native SARS-CoV $3CL^{pro}$ was produced by hanging drop method. 5 µL of purified $3CL^{pro}$ (5 mg/mL) was mixed with mother liquor (50mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 5.0 – 6.5, 2.5 – 10 % (w/v) of polyethylene glycol (PEG) 6000, 10 % (v/v) glycerol, 3 % (v/v) DMSO, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM dithiothreitol (DTT)) in 1:1 ratio, and stayed at 16 ℃ until crystal formation. Single crystal was transferred to 5 µL of mother liquor with inhibitor in excess and stayed overnight at 16 ℃, so that the inhibitor was soaked into the crystal. Before X-ray diffraction, the crystal was immersed in cryoprotectant (mother liquor with extra 20 % (v/v) glycerol) for 5 minutes.

2.4.2.2 X-Ray Diffraction and Data Collection

X-ray source was provided by Rigaku R-AXIS IV x-ray diffractometor and image was captured by R-AXIS IV++ imaging plate detector. Single crystal was mounted in a loop, placed onto gonionmeter and positioned to the centre. Cryojet was adjusted for freezing the crystal by nitrogen gas at 110 K. Preliminary data was obtained by collecting two diffraction images at 0 ℃ and 90 ℃ with maximal resolution of 1.6 Å with exposure for 2 minutes, to predict resolution limit and optimal exposure time. The images were further analyzed by a computer program MOSFLM to estimate space group, mosaicity and strategy for data collection.

As space group of $3CL^{pro}$ was found to be $P2_1$, 180 diffraction images that each image was 1° of oscillation angle, were collected. Resolution limit and

exposure time were adjusted to approximately 1.8 2.5 and 8 - 10 minutes respectively, depending on crystal quality.

2.4.2.3 Data Processing, Molecular Replacement and Structure Refinement

Structures of SARS-3CL^{pro}-inhibitor complexes were built by molecular replacement with computer program "python-based hierarchical environment for integrated xtallography" using SARS-CoV 3CL^{pro} (PDB ID: 1WOF) as the search model (Adams *et al.*; Yang *et al.*, 2005). The restrained refinement was performed by computer programs "coot" and "PHENIX" (Emsley & Cowtan, 2004). The stereochemistry was examined by web tool "MolProbity" (Davis *et al.*, 2007).

Chapter 3 – FRET Assay Establishment

An assay determining $3CL^{pro}$ activity against substrate variants is required for substrate specificity profiling, so a fluorescence pair linking substrate sequence was constructed. Cleavage of the sequence could be determined by monitoring change in FRET efficiency. The activity under various concentrations of the $3CL^{pro}$ and substrate was measured. These concentrations were standardized for profiling.

3.1 Principle of FRET Assay

SARS-CoV $3CL^{pro}$ activity was measured by using the protein-based substrate consisted of 12-residue autocleavage sequence (TSAVLQ↓SGFRKM) inserted between CFP and YFP (Figure 3.1). Cleavage of peptide bond between P1 and P1' positions by $3CL^{pro}$ led to separation of the fluorescence proteins and decrease in FRET efficiency. The proteolytic rate could therefore be measured by monitoring emitted fluorescence intensity.

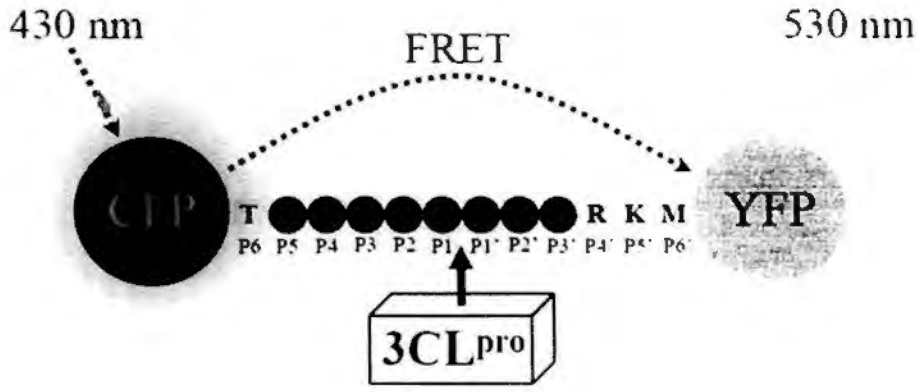


Figure 3.1 – Schematic diagram illustrating measurement of $3CL^{pro}$ activity on the protein-based substrate by FRET. When CFP is excited at 430 nm, YFP emits fluorescence at 530 nm through FRET. Cleavage of the peptide bond at SAVLQ↓SGF by the $3CL^{pro}$ separates CFP and YFP, leading to a decrease in the emitted fluorescence at 530 nm.

3.2 Production of SARS-CoV $3CL^{pro}$ and Protein-Based WT Substrate

A plasmid encoding the recombinant SARS-CoV $3CL^{pro}$ tagged with His_6-MBP tag at N-terminus was constructed (Figure 3.2). The recombinant protease was expressed in BL21 (DE3) pLysS at 37 ℃ for four hours and purified by IMAC. The His_6-MBP tag was removed by factor Xa digestion and by IMAC. The protease was further purified by size exclusion chromatography. The final purity was higher than 90 % (Figure 3.3). Elution profile of size exclusion chromatography showed that the $3CL^{pro}$ was dimeric in solution (Figure 3.4), indicating that the protease was in native conformation.

The protein-based WT substrate, His_6-CFP-TSAVLQSGFRKM-YFP, was expressed in BL21 (DE3) pLysS at 22 ℃ overnight and purified by IMAC. The purity was higher than 80 % (Figure 3.5).

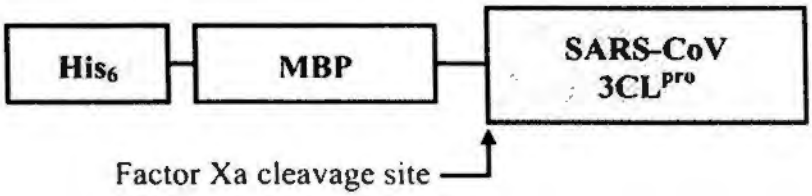


Figure 3.2 – Schematic diagram illustrating construct of recombinant SARS-CoV $3CL^{pro}$. The recombinant protein consists of $3CL^{pro}$ tagged by His_6-MBP at N-terminus. Factor Xa cleavage site (IEGR↓SGFRKM) is located at inserted between the $3CL^{pro}$ and tag, so that the $3CL^{pro}$ with native N-terminus is produced after cleavage.

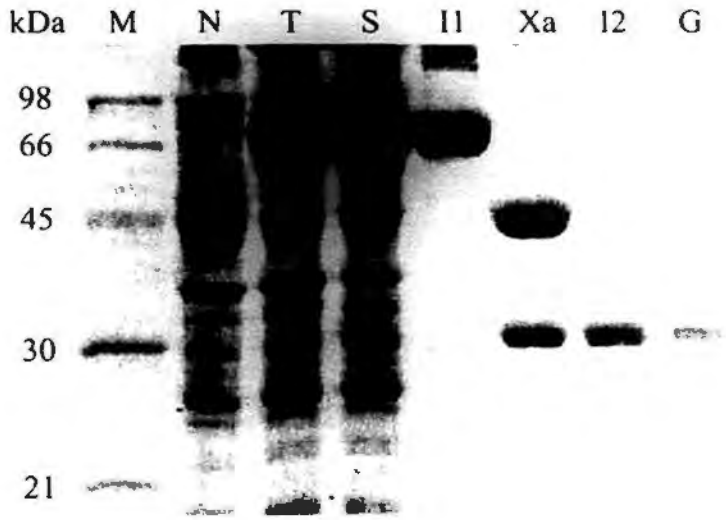


Figure 3.3 – Purification of SARS-CoV 3CL^{pro}. Compared to non-induced control (lane: N), recombinant 3CL^{pro} of 79 kDa was present in total lysate (lane: T) after induction. Soluble fraction (lane: S) was subject to IMAC (lane: I1). After digestion by factor Xa (lane: Xa), His_6-MBP tag was removed by IMAC (lane: I2). The 3CL^{pro} of 34 kDa was finally purified by G75 size exclusion chromatography (lane: G). Purity of the final product was higher than 90 %. Protein marker was loaded in parallel (lane: M).

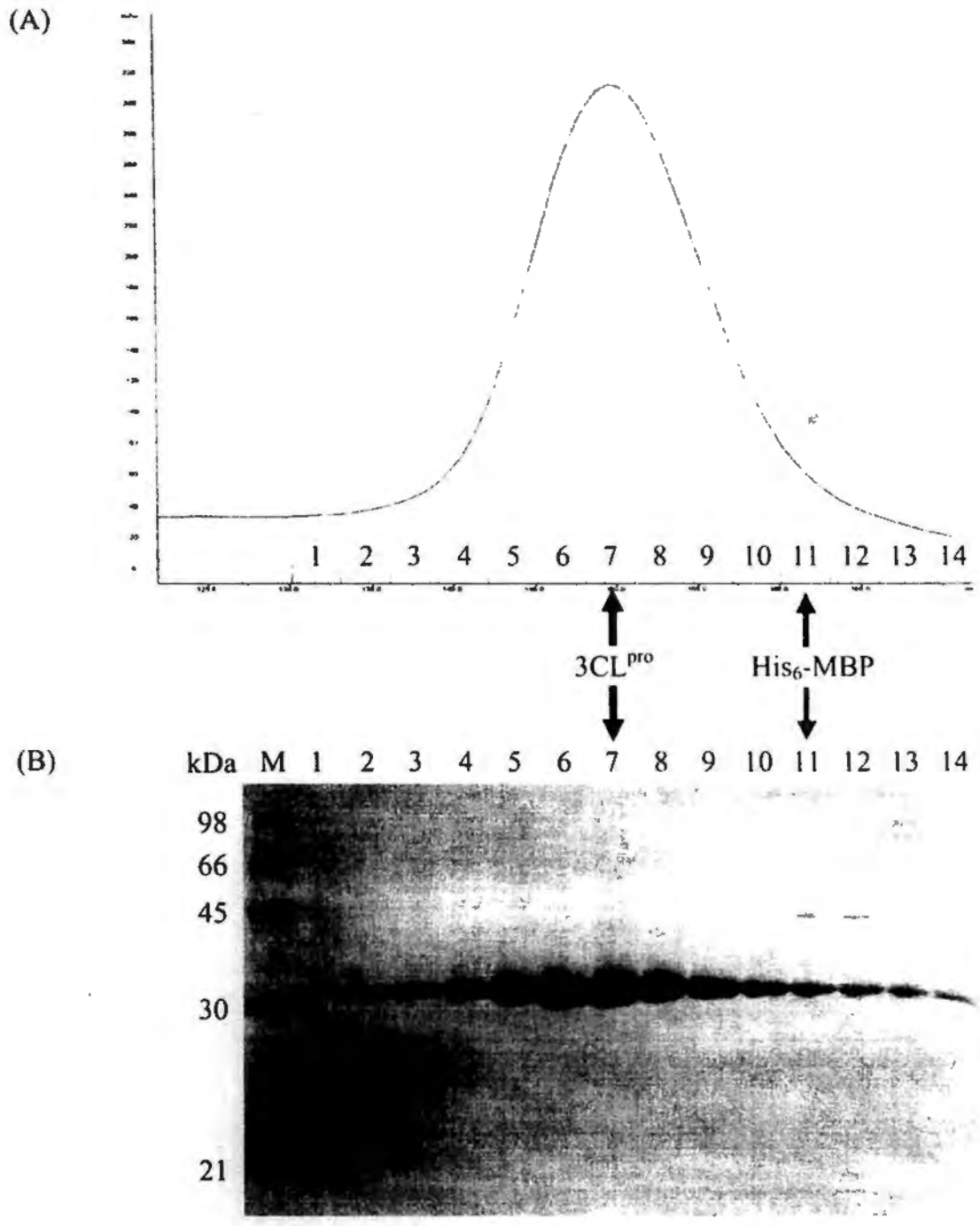


Figure 3.4 – Purified SARS-CoV 3CL^{pro} was dimeric. (A) 3CL^{pro} was subjected to G75 size exclusion chromatography after removal of His_6-MBP tag. Retention volumes of the 3CL^{pro} and tag in trace amount were 150 mL (fraction 7) and 175 mL (fraction 11) respectively, indicating that molecular weight of the 34 kDa protease is higher than that of the 45 kDa tag. **(B)** Protein contents in fraction 1 – 14 were examined.

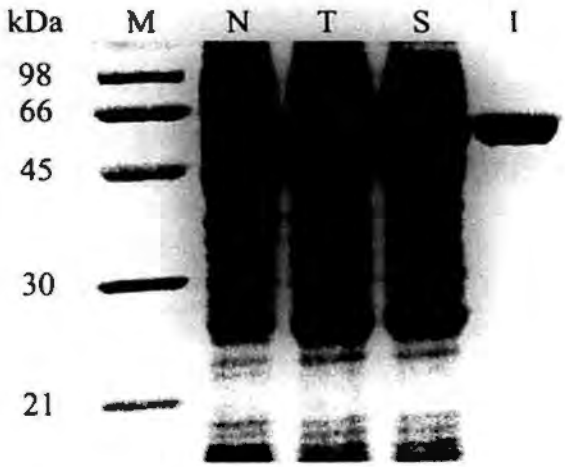


Figure 3.5 – Purification of the protein-based substrate by IMAC. Compared to non-induced control (lane: N), the protein-based substrate of 58 kDa was present in total lysate (lane: T) after induction. Soluble fraction (lane: S) was subject to IMAC (lane: I). Purity of the substrate was higher than 80 %.

3.3 Cleavage of the Substrate by $3CL^{pro}$

3.3.1 Cleavage at Peptide Bond between P1 and P1' Positions

To determine whether the substrate was cleavable by the $3CL^{pro}$, they were mixed and stayed at room temperature for one hour. The substrate of 58 kDa was cleaved into two fragments of 28 and 30 kDa (Figure 3.6). N-terminal sequencing results showed that the five residues at N-terminus of the 28 kDa fragment were SGFRK, verifying that the protease can specifically cleave the substrate at peptide bond between P1 and P1' positions (Figure 3.7).

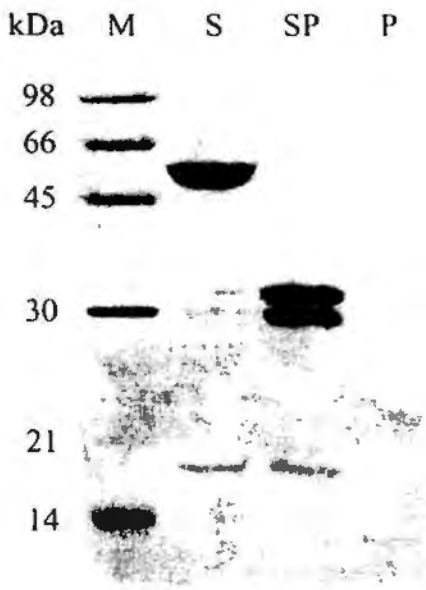


Figure 3.6 – Cleavage of the substrate into two fragments by SARS-CoV 3CL^{pro}. 35 μM of the purified substrate was digested by 4 μM of the purified 3CL^{pro} at room temperature for one hour. The substrate of 58 kDa (lane: S) was separated into two fragments of 28 kDa and 30 kDa (lane: SP) after digestion. A protease sample without substrate (lane: P) was loaded in parallel as control.

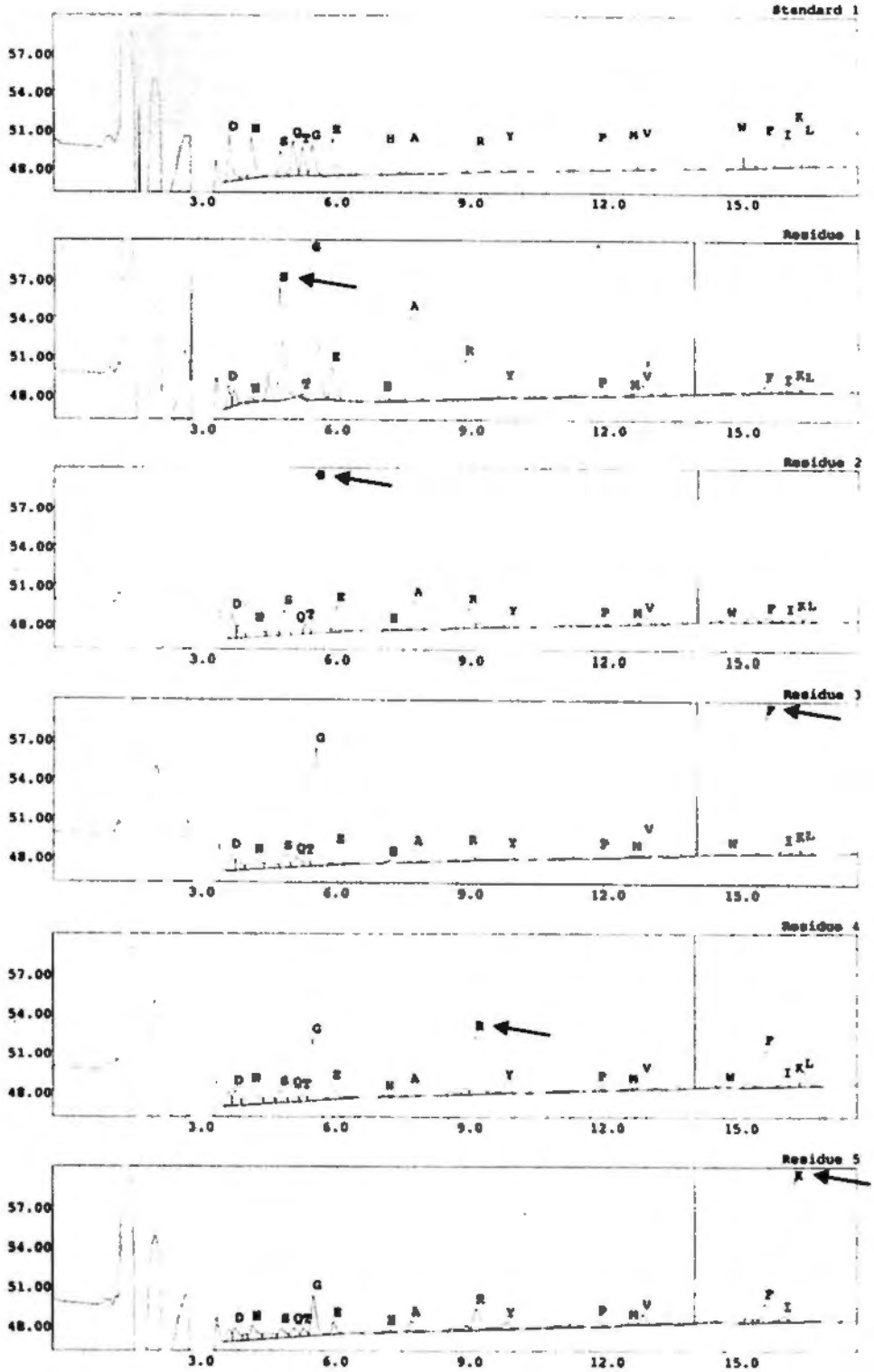


Figure 3.7 – N-terminal sequencing of the cleaved substrate fragment. The 28 kDa fragment was subject to N-terminal sequencing. Five residues at the N-terminus were SGFRK. Arrows indicate the residues that peaks were higher than standard peaks.

3.3.2 Decrease in FRET Efficiency

To examine whether cleavage of the substrate cause decrease in FRET efficiency, the substrate mixed with the $3CL^{pro}$ was excited by light at 430 nm and emitted fluorescence at 530 nm was monitored for every 30 seconds. A time-dependent decrease in the fluorescence intensity was observed (Figure 3.8), indicating that the cleavage rate can be determined by FRET.

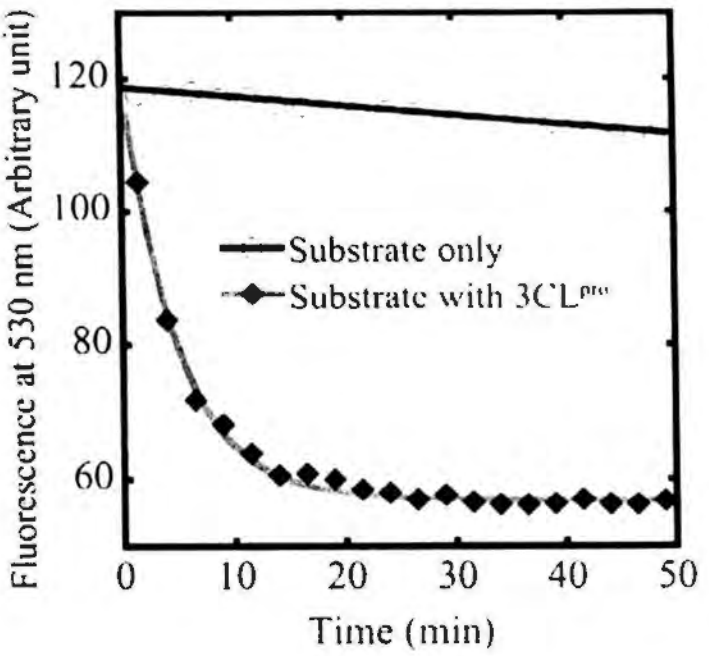


Figure 3.8 – Cleavage caused decrease in FRET efficiency. During cleavage by the $3CL^{pro}$, the substrate was excited by light at 430 nm while the emitted fluorescence intensity at 530 nm was monitored for every 30 seconds. The intensity was reduced in a time-dependency. A substrate sample without protease was assayed in parallel as control.

3.4 *k_{obs} and Specific Activity of $3CL^{pro}$ against the Substrate*

k_{obs} of decrease in fluorescence was determined by fitting kinetics to single exponential decay equation:

$$F_{S-P} \times e^{(-k_{obs} \times t)} + F_P$$

where F_{S-P} is difference in fluorescence intensity between the substrate and product, F_P is fluorescence intensity of the product, and t is time in terms of minutes respectively. k_{obs} values were measured at 1 – 16 μM of $3CL^{pro}$ (Figure 3.9A). Graph of k_{obs} against $3CL^{pro}$ concentration was plotted to yield a straight line. The specific activity, $k_{obs}/[3CL^{pro}]$, of the $3CL^{pro}$ against WT autocleavage sequence was determined by slope of the straight line, which was 71 ± 11 $min^{-1}mM^{-1}$ in overall (Figure 3.9B).

k_{obs} values were directly proportional to $3CL^{pro}$ concentration at 1 – 16 μM, indicating the specific activity was constant by using the $3CL^{pro}$ at those concentrations. To simplify the procedures, 1 – 4 μM of the $3CL^{pro}$ were used for substrate specificity profiling. The concentrations were raised to 2 – 8 μM and 4 – 16 μM when the relative activity was < 0.3 and < 0.1 respectively.

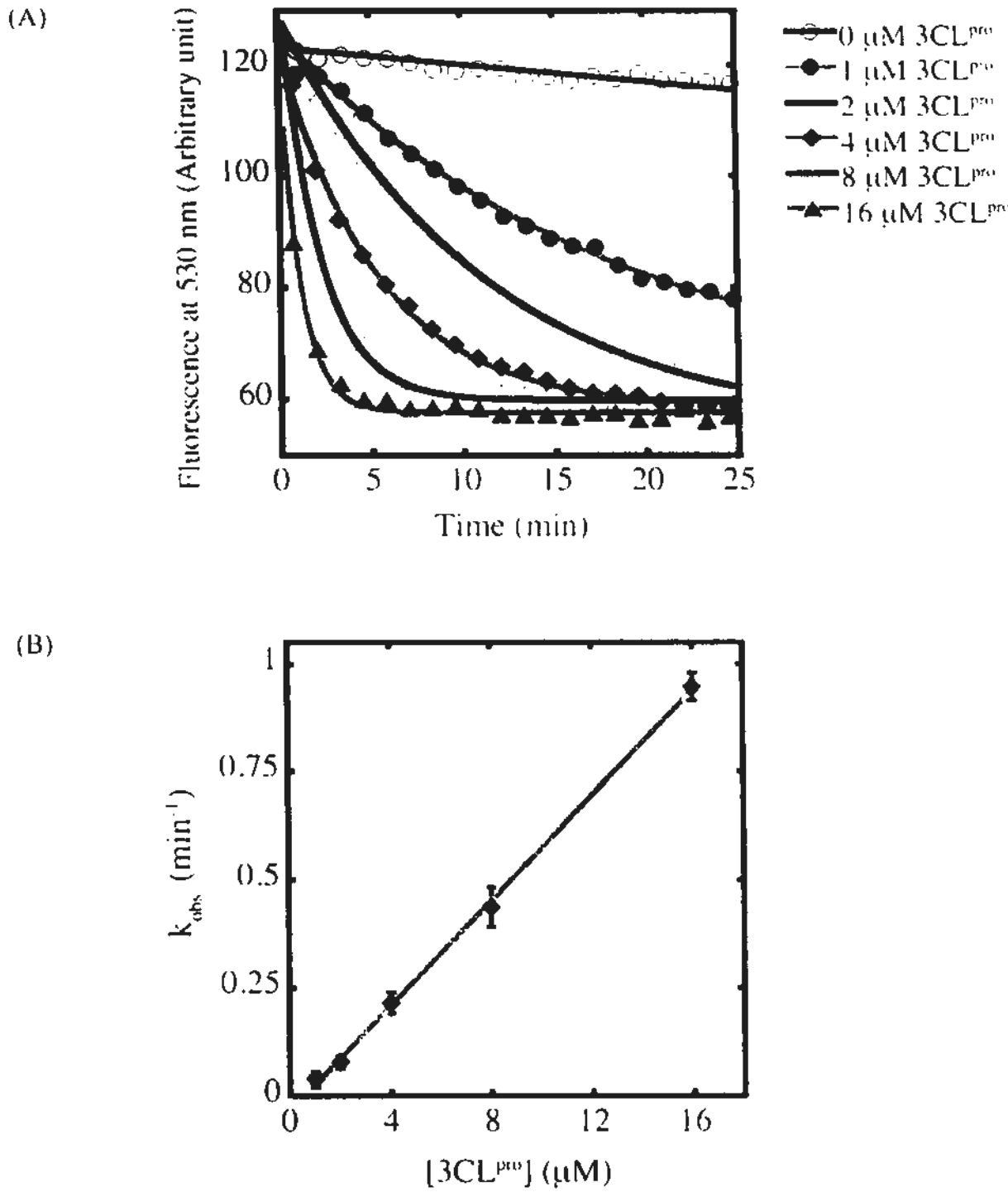


Figure 3.9 – Determining k_{obs} and specific activity. **(A)** The protein-based substrate cleaved by 1 – 16 μM of SARS-CoV $3CL^{pro}$ led to decrease in fluorescence at 530 nm. k_{obs} was obtained by fitting data to single exponential decay equation. **(B)** The plot of k_{obs} against $[3CL^{pro}]$ yielded a straight line. The specific activity, $k_{obs}/[3CL^{pro}]$, was determined by slope of the line.

3.5 Normalizing Substrate Concentration

3.5.1 Cleavage Rate is Substrate Concentration-Dependent

Certain amount of impurities is present in the substrate purified by only IMAC. The impurities caused overestimation of the substrate concentration when the concentration was determined by measuring OD_{280}. Different amount of impurities in each sample led to variation of substrate concentrations. The specific activity is substrate concentration-dependent because it equals to $\frac{k_{cat}}{K_m + [Substrate]}$ (Figure 3.10A). Varying the substrate concentration at 20 – 50 μM caused change in $3CL^{pro}$ activity, demonstrating the dependence (Figure 3.10B). Hence, normalizing concentrations of substrate variants by another more accurate method is required for substrate specificity profiling.

(A)

$$k_{obs} \times [Substrate] = \frac{k_{cat} \times 3CL^{Pro} \times [Substrate]}{K_m + [Substrate]}$$

$$\frac{k_{obs}}{3CL^{Pro}} = \frac{k_{cat}}{K_m + [Substrate]}$$

(B)

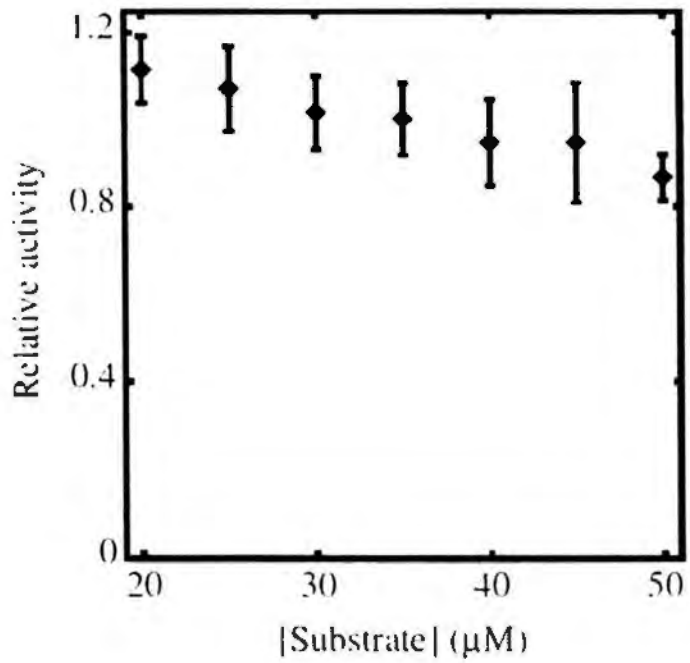
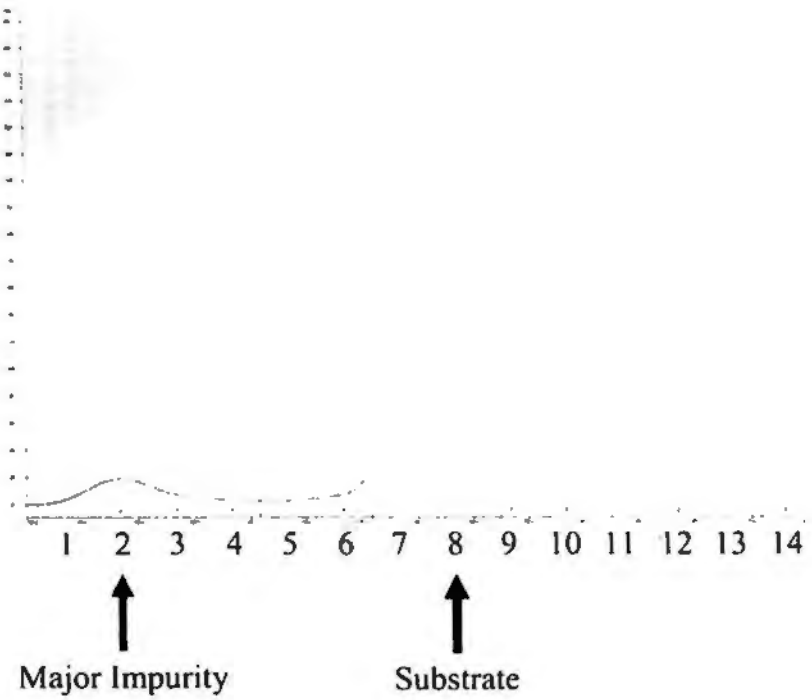


Figure 3.10 – Increase in substrate concentration reduced $3CL^{pro}$ activity. (A) Rearranged Michaelis-Menten equation demonstrates that the specific activity, $k_{obs}/[3CL^{pro}]$, depends on the substrate concentration. **(B)** Increase in substrate concentration from 20 to 50 μM caused that relative activity was decreased from 1.11 to 0.87. The relative activity was determined by the specific activity relative to that using 35 μM of the substrate.

3.5.2 Determining Substrate Concentration by measuring OD_{430}

An alternative method to obtain the substrate concentration was OD_{430} measurement. Light at 430 nm is absorbed by CFP of the protein-based substrate. Extinction coefficient of CFP at OD_{430} is experimentally determined by the substrate with high purity and known concentration. The substrate with high purity was obtained by purification by size exclusion chromatography, and the purity was improved to > 90 % (Figure 3.11). OD_{430} and OD_{280} of the substrate with high purity were measured, followed by multiplying ratio of OD_{430}:OD_{280} by the coefficient at 280 nm, to determine the coefficient at 430 nm, which was 27,500 M^{-1} cm^{-1} (Figure 3.12).

(A)



(B)

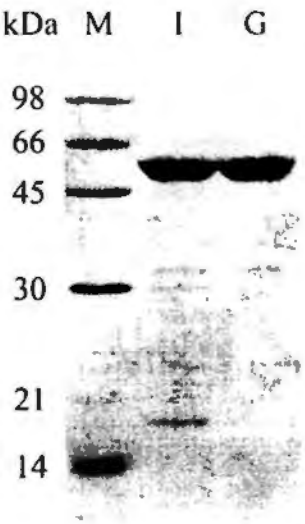


Figure 3.11 – Purifying the substrate by size exclusion chromatography. **(A)** The substrate purified by IMAC was subject to size exclusion chromatography. The substrate and impurity were eluted in 150 mL (fraction 8) and void volume (fraction 2) respectively. **(B)** Compared to input (lane I), purity of the substrate was raised and partial impurities were removed (lane G). The substrate concentrations in lane I and G were similar after measuring OD_{430} and normalization.

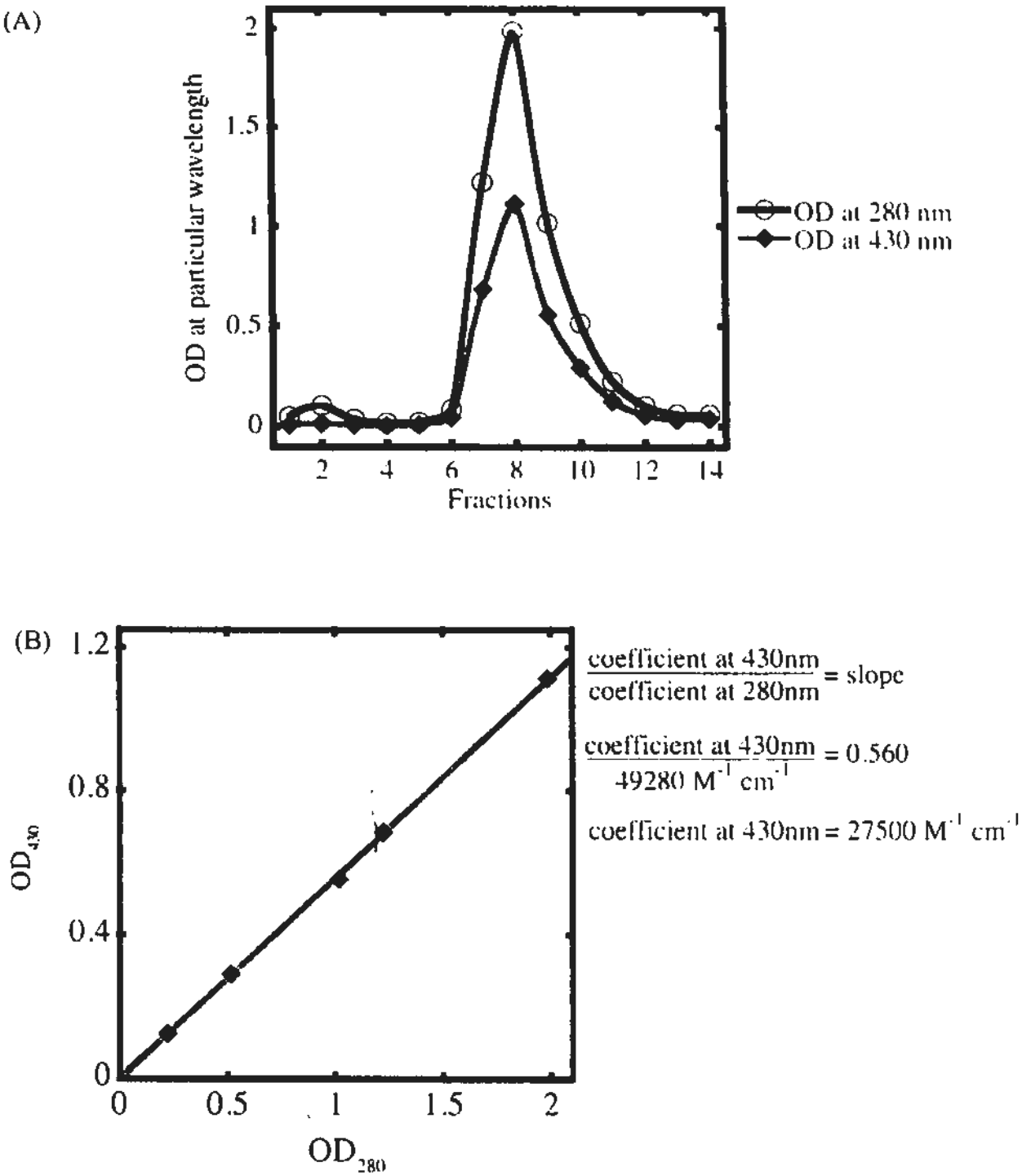


Figure 3.12 – Determining extinction coefficient of the substrate at 430 nm. **(A)** OD_{430} and OD_{280} of the highly pure substrate were measured. **(B)** Ratio of OD_{430}:OD_{280} was determined by measuring fractions 7 – 11. The ratio of 0.560 was multiplied by the coefficient at 280 nm, 49,280 M^{-1} cm^{-1}, to obtain the coefficient at 430 nm, which was 27,500 M^{-1} cm^{-1}.

3.6 Summary

Homodimeric SARS-CoV $3CL^{pro}$ with native termini and the protein-based substrate consisted of substrate sequence (TSAVLQ↓SGFRKM) inserted between CFP/YFP pair were expressed and purified. $3CL^{pro}$ specifically cleaved peptide bond of the substrate between P1 and P1' positions, leading to decrease in FRET efficiency. Specific activity of the protease, $k_{obs}/[3CL^{pro}]$, against 35 μM of the WT substrate was determined by monitoring the emitted fluorescence at 530 nm, and was 71 ± 11 mM^{-1} min^{-1}. $3CL^{pro}$ concentration was directly proportional to the specific activity, while increase in substrate concentration caused a reduction. Thus, the substrate specificity would be profiled by using 1 – 4 μM of $3CL^{pro}$. The substrate with relative activity of < 0.3 and < 0.1 would be further examined by using $3CL^{pro}$ at 2 – 8 μM and 4 – 16 μM. In addition, as protease activity is substrate concentration-dependent, the substrate concentration was obtained by measuring OD_{430} and was normalized to 35 μM.

Chapter 4 – Substrate Specificity of SARS-CoV 3CL^{pro}

Substrate specificity of SARS-CoV 3CL^{pro} can be used for improving computational simulation for protease-substrate/inhibitor interaction, and may help in a rational based design of potent inhibitors. Previous studies on comprehensive specificity are however scarce and are unable to clearly exhibit the specificity at P5 to P3' positions. We therefore established a substrate library of 19×8 variants for substrate specificity profiling. The protease activity against each of variants was determined by FRET assay, and correlated with structural properties of substituting residues, so that substrate specificity at each of the positions was obtained.

4.1 Substrate Specificity Profiling

Saturation mutagenesis at P5 to P3' positions of the autocleavage sequence was performed to create a substrate library of 19×8 variants (Figure 4.1). The relative cleavage rate of 3CL^{pro} against these substrate sequences was measured (Figure 4.2 and table 4.1). For examining Cys substituting variants, 2.5 μM TCEP was added to prevent disulphide bond formation. As TCEP up-regulated the protease activity, the specific activity against Cys substituting variants was normalized by that against the WT substrate assayed under same reaction condition (Figure 4.3).

In general, solvent-exposed sites such as P5, P3, and P3' positions were less selective than the others. The most selective site was P1 position – cleavage was only observable with Gln, His or Met. Substrate sequences with Pro substitutions at P3, P1', P2' positions were non-cleavable. Solvent-exposed sites such as P5, P3 and

P3' positions preferred positively charged substitutions, as the protease activity on the Arg/Lys-substituting variants was consistently higher than that of the Asp/Glu-substituting variants. The difference was the largest at P3 position, where positively charged substitutions resulted in 12-fold higher proteolytic activity. P5 and P3' variants with positive charges were also 3-fold higher in activity. The preference on charged residues indicated that electrostatic interaction, which is long-range in nature, may play a role in the 3CL^{pro} catalysis. One of the possibilities is that the positive charges stabilize transition state of thcatalysis. It is expected that the carboxylate group at P1 residue will be converted to an oxyanion during the formation of the transition state. Presence of positive charges near the active site may electrostatically stabilize the oxyanion and thus promote catalysis. Another possibility is a direct electrostatic interaction between positively charged residues of substrate and negatively charged residues of 3CL^{pro}. There is a Glu166 located at substrate binding cleft that can interact with P3 residue. This may explain why P3 position has the strongest preference for positively charged residues.

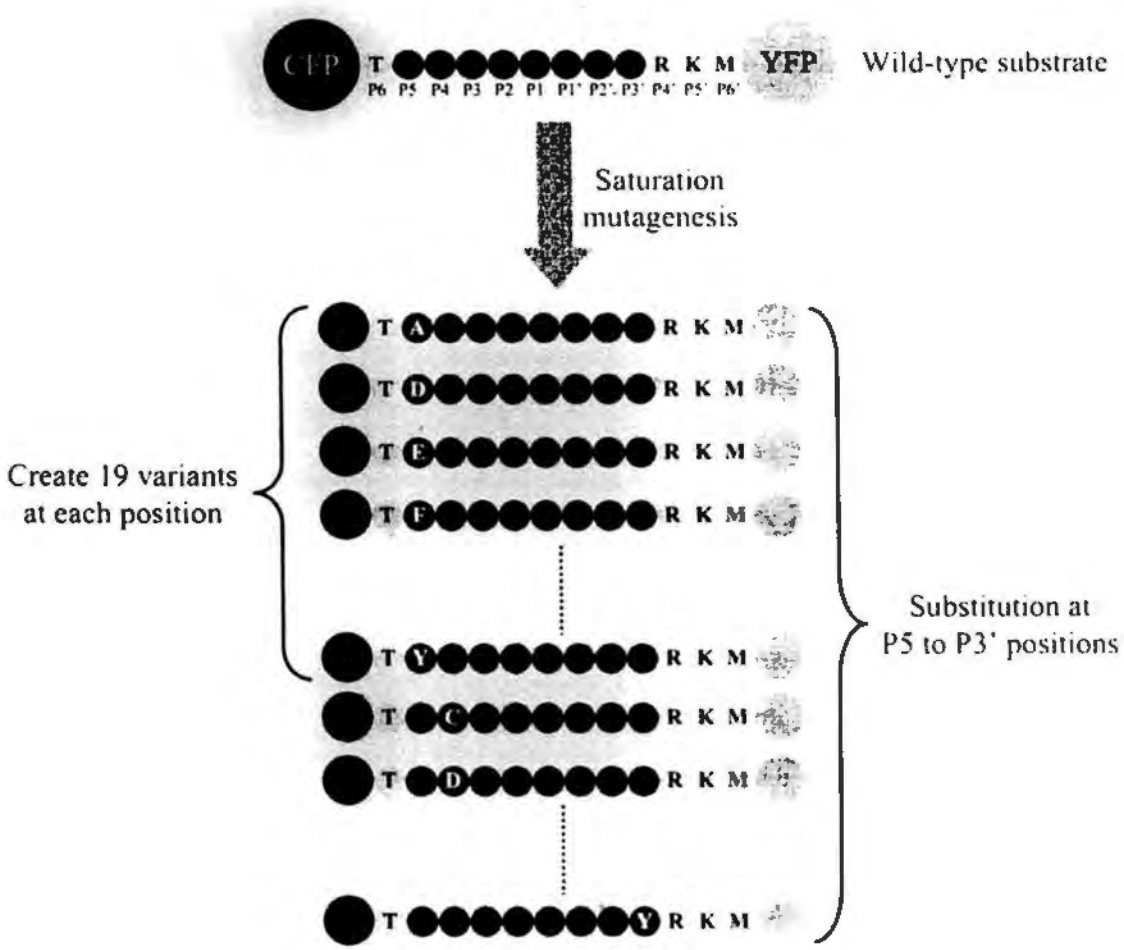


Figure 4.1 – Schematic diagram illustrating create of the substrate library. P5 to P3' residues (SAVLQ↓SGF) of the autocleavage sequence were substituted to other 19 nature-occurring residues (white) by saturation mutagenesis, to create 19×8 substrate variants.

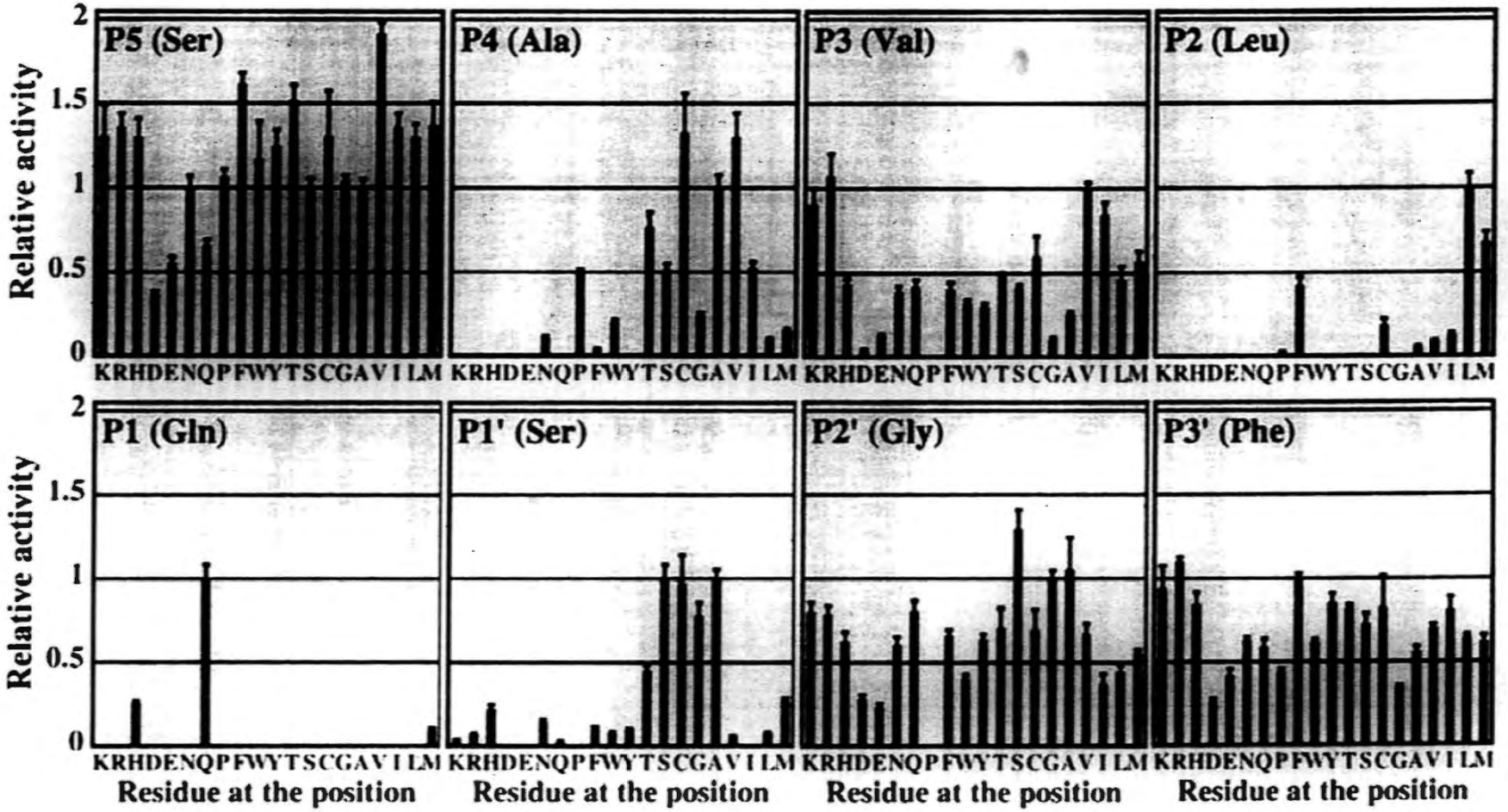


Figure 4.2 – Profiling the substrate specificity at P5 to P3' positions. Specific activity on each of 19×8 variants was determined, and normalized by that on WT substrate to obtain the relative activity.

Table 4.1 – SARS-CoV 3CL^{pro} relative activity on the substrate variants.

'ND' stands for non-detectable cleavage.

Residue	Position at the autocleavage sequence (WT residue)							
	P5(Ser)	P4(Ala)	P3(Val)	P2(Leu)	P1(Gln)	P1'(Ser)	P2'(Gly)	P3'(Phe)
K	1.30±0.19	ND	0.91±0.09	ND	ND	0.03±0.01	0.80±0.05	0.94±0.13
R	1.36±0.08	ND	1.07±0.13	ND	ND	0.07±0.01	0.79±0.05	1.10±0.03
H	1.30±0.11	ND	0.43±0.03	ND	0.26±0.02	0.22±0.02	0.62±0.06	0.84±0.07
D	0.37±0.02	ND	0.04±0.01	ND	ND	ND	0.28±0.02	0.26±0.02
E	0.55±0.04	ND	0.13±0.01	ND	ND	ND	0.23±0.02	0.42±0.04
N	1.01±0.06	0.11±0.01	0.39±0.03	ND	ND	0.15±0.01	0.60±0.05	0.62±0.02
Q	0.66±0.04	ND	0.41±0.04	ND	1.00±0.08	0.03±0.00	0.81±0.06	0.59±0.05
P	1.06±0.05	0.49±0.03	ND	0.03±0.00	ND	ND	ND	0.43±0.02
F	1.62±0.06	0.04±0.00	0.40±0.04	0.42±0.05	ND	0.11±0.01	0.66±0.04	1.00±0.03
W	1.17±0.22	0.20±0.02	0.32±0.02	ND	ND	0.08±0.01	0.40±0.03	0.61±0.02
Y	1.25±0.09	ND	0.30±0.02	ND	ND	0.10±0.01	0.64±0.03	0.85±0.06
T	1.52±0.09	0.77±0.09	0.48±0.02	ND	ND	0.45±0.04	0.71±0.12	0.82±0.02
S	1.00±0.05	0.51±0.04	0.41±0.02	ND	ND	1.00±0.08	1.29±0.12	0.72±0.07
C	1.30±0.27	1.32±0.24	0.60±0.12	0.18±0.04	ND	0.97±0.18	0.69±0.12	0.83±0.19
G	1.03±0.04	0.24±0.02	0.11±0.01	ND	ND	0.78±0.08	1.00±0.05	0.35±0.01
A	1.00±0.05	1.00±0.08	0.26±0.01	0.06±0.01	ND	0.99±0.06	1.05±0.19	0.56±0.03
V	1.92±0.07	1.30±0.15	1.00±0.04	0.09±0.01	ND	0.06±0.01	0.67±0.06	0.70±0.03
I	1.35±0.09	0.53±0.03	0.84±0.07	0.13±0.01	ND	ND	0.37±0.06	0.81±0.08
L	1.30±0.08	0.10±0.01	0.45±0.08	1.00±0.08	ND	0.08±0.01	0.44±0.04	0.64±0.02
M	1.37±0.14	0.15±0.01	0.57±0.06	0.68±0.06	0.10±0.01	0.27±0.02	0.56±0.02	0.61±0.04

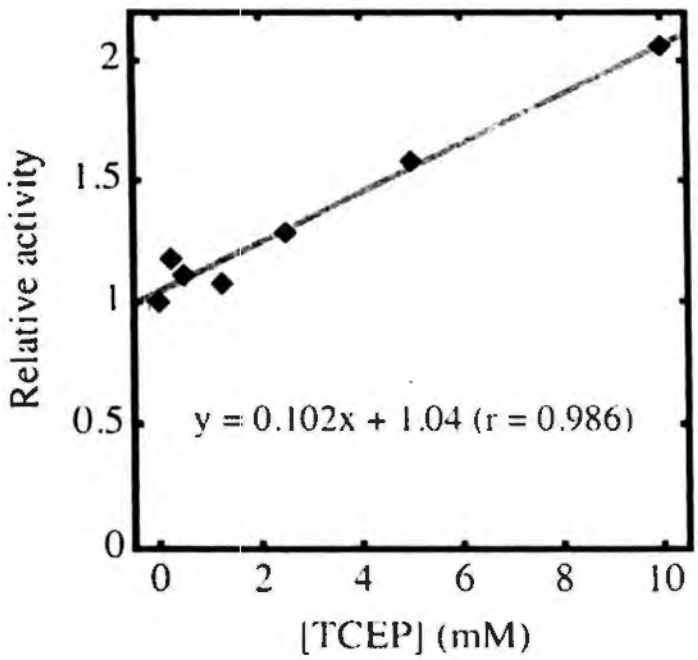


Figure 4.3 – TCEP increased SARS-CoV 3CL^{pro} activity. FRET assay was performed in presence of 0.25 – 10 mM of TCEP. 3CL^{pro} activity relative to control without TCEP was determined. The relative activity was directly proportional to TCEP concentration.

4.2 Correlation between 3CL^{pro} Activity and Structural Properties of Substituting Residues

To find out the preferred properties of substituting residues, 3CL^{pro} activity was correlated with scales for side chain volume (Lee *et al.*, 2008), hydrophobicity (Kyte & Doolittle, 1982), and α-helix and β-sheet propensities (Chou & Fasman, 1978) (Appendix 2.12). The correlation coefficients (r) and p-values were shown in tables 6.2. Significant correlations with p-value < 0.01 were observed in a few cases.

Table 4.2 – Correlation between SARS-CoV 3CL^{pro} activity and structural properties of substituting residues. The relative activity was correlated with scales for side chain volume, hydrophobicity, and α-helix and β-sheet propensities of the substituting residues (Chou & Fasman, 1978; Kyte & Doolittle, 1982; Lee *et al.*, 2008). The correlation coefficients and p-values (in parenthesis) were reported. Significant correlations with p-value < 0.01 were bolded and marked with asterisks.

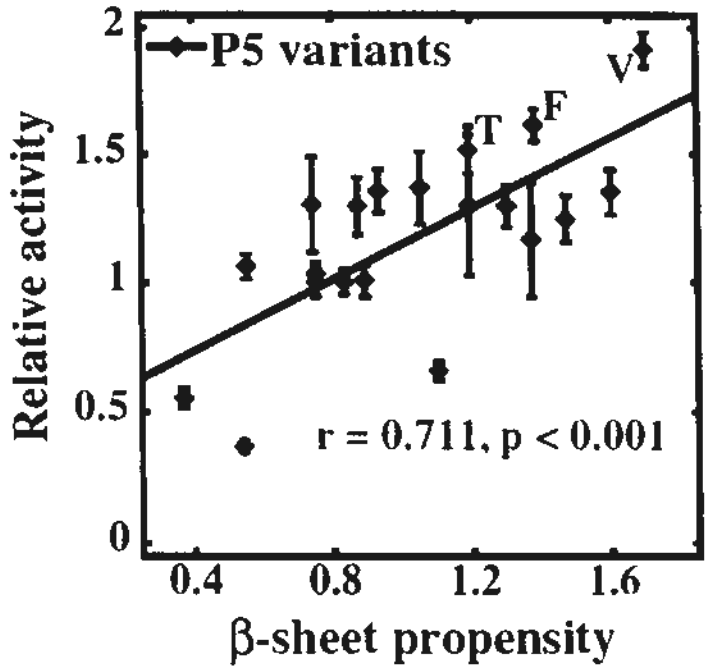
Position	Side chain volume	Hydrophobicity	α-helix propensity	β-sheet propensity
P5	0.331 (0.154)	**0.573 (0.008)***	-0.064 (0.789)	**0.711 (<0.001)***
P4	-0.424 (0.063)	**0.587 (0.006)***	-0.147 (0.536)	0.315 (0.176)
P3	0.338 (0.144)	0.221 (0.349)	0.170 (0.473)	0.510 (0.022)
P2	0.255 (0.277)	**0.590 (0.006)***	0.379 (0.100)	0.304 (0.192)
P1	0.038 (0.873)	-0.269 (0.252)	0.126 (0.595)	0.021 (0.931)
P1'	**-0.660 (0.002)***	0.233 (0.323)	-0.222 (0.347)	-0.143 (0.548)
P2'	-0.363 (0.116)	0.022 (0.926)	-0.097 (0.685)	0.048 (0.841)
P3'	0.496 (0.026)	0.094 (0.695)	-0.017 (0.944)	0.486 (0.030)

4.3 Substrate Specificity at P5 to P3' positions

4.3.1 P5 Position Prefers Residues with High β-sheet Propensity

All substitutions at P5 position were cleavable, and the relative activity ranged from 0.37 to 1.92. Many substitutions resulted in activity significantly higher than that for WT substrate (Figure 4.2). S5V (1.92 ± 0.07) was the most preferred substrate variant, followed by S5F (1.62 ± 0.06) and S5T (1.52 ± 0.09). A strong correlation was observed between the relative activity and β-sheet propensity ($r = 0.711$, $p < 0.001$) (Table 4.2 and figure 4.4A). The relative activity also correlated well with the hydrophobicity of substituting residues ($r = 0.573$, $p = 0.008$) (Figure 4.4B), but the coefficient was lower than that of β-sheet propensity, so P5 preference tended to prefer β-sheet propensity.

(A)



(B)

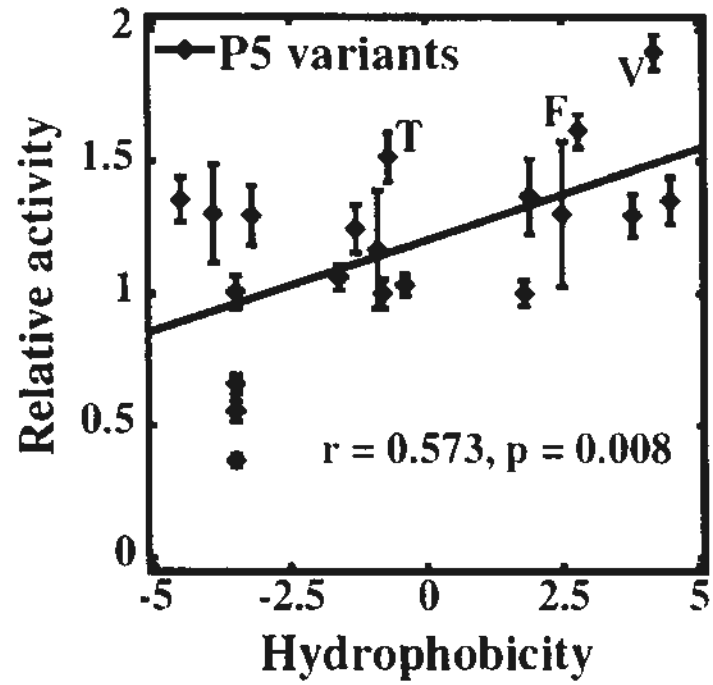


Figure 4.4 – P5 position preferred residues with high β-sheet propensity and hydrophobicity. The relative activity against P5 variants correlated with **(A)** β-Sheet propensity ($r = 0.711$, $p < 0.001$) and **(B)** hydrophobicity ($r = 0.573$, $p = 0.008$).

4.3.2 P4 Position Prefers Small Hydrophobic Residues

The best substitutions were Cys and Val, with relative activity of 1.32 ± 0.24 and 1.30 ± 0.15, respectively (Figure 4.2). The relative activity correlated with hydrophobicity ($r = 0.587$, $p = 0.006$) (Table 4.2), yet the data were not well fitted to the trend line (Figure 4.5A). In most of the active variants, side chain volumes of P4 residues were < 80 Å^3 (Figure 4.5B). The correlation was more evident ($r = 0.942$, $p < 0.001$) when large variants were excluded from the analysis (Figure 4.5C). From the crystal structure of 3CL^{pro}-substrate complex, the side chain of P4 is completely buried inside a small hydrophobic pocket (Xue *et al.*, 2008). Our data suggest that for those residues that are small enough to fit into the binding pocket, the relative activity is directly proportional to the hydrophobicity of the substituting residues. No observable cleavage was detected for charged residues (Arg, Asp, Glu and Lys), probably due to the high desolvation penalty for burial of charges inside the hydrophobic pocket.

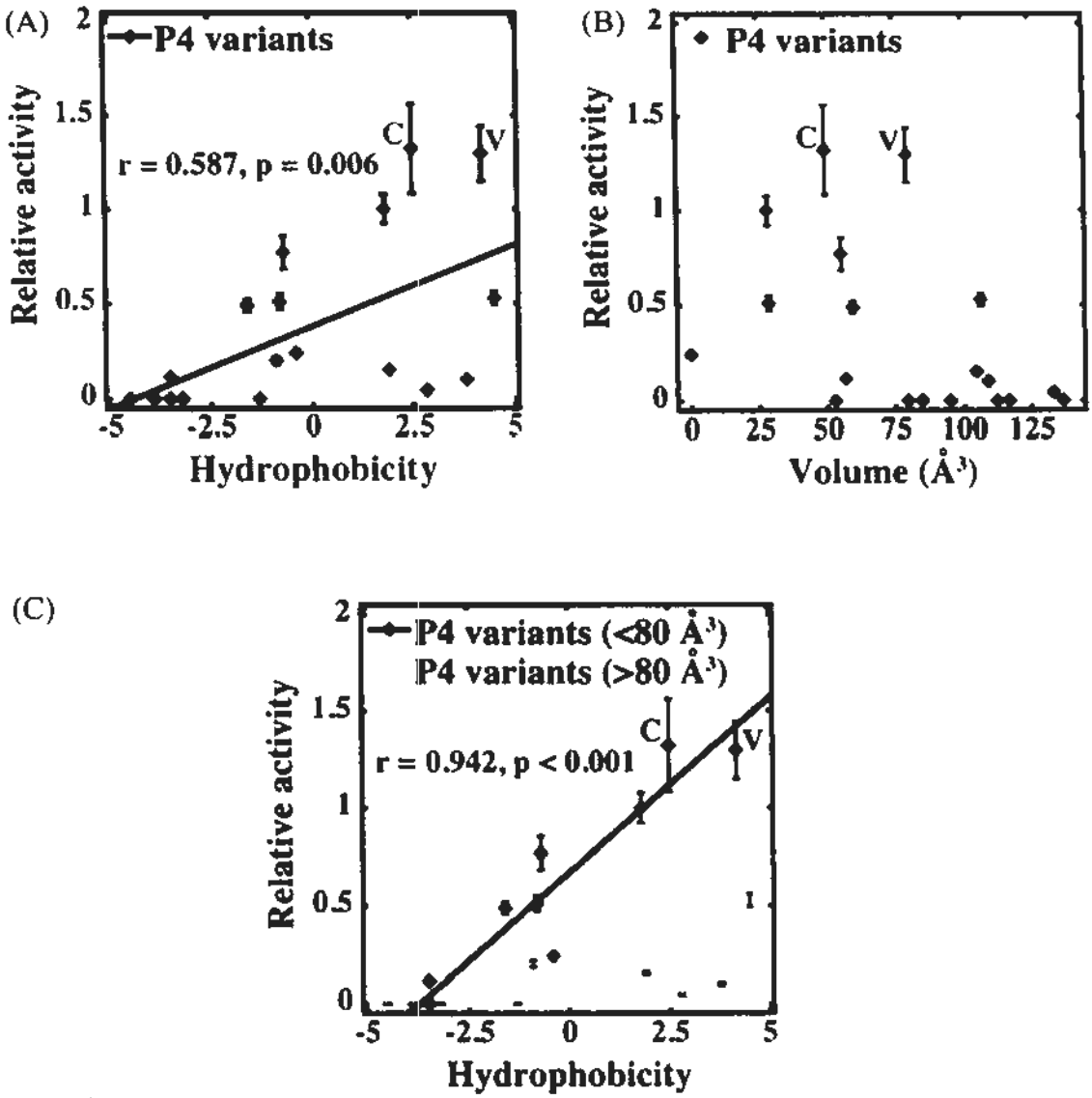


Figure 4.5 – P4 position preferred small hydrophobic residues. **(A)** A significant correlation was observed for hydrophobicity (r = 0.587, p = 0.008), but data were deviated from the trend line. **(B)** P4 residues of active variants were almost < 80 Å^3 in side chain volume **(C)** The correlation was improved (r = 0.942, p < 0.001) when only residues with side chain volumes of < 80 Å^3 (Ala, Asn, Asp, Cys, Glu, Gly, Pro, Ser, Thr and Val) were included.

4.3.3 P3 Position Prefers Residues with High β-sheet Propensity

The relative activity for P3 variants correlated well with β-sheet propensity ($r = 0.510$, $p = 0.022$) (Table 4.1). As discussed above, P3 position favors positively charged residues over negatively charged one (Figure 4.2). After excluding the charged residues, we found that the 3CL^{pro} activity was directly proportional to β-sheet propensity of the substituting residues ($r = 0.729$, $p = 0.001$) (Figure 4.6).

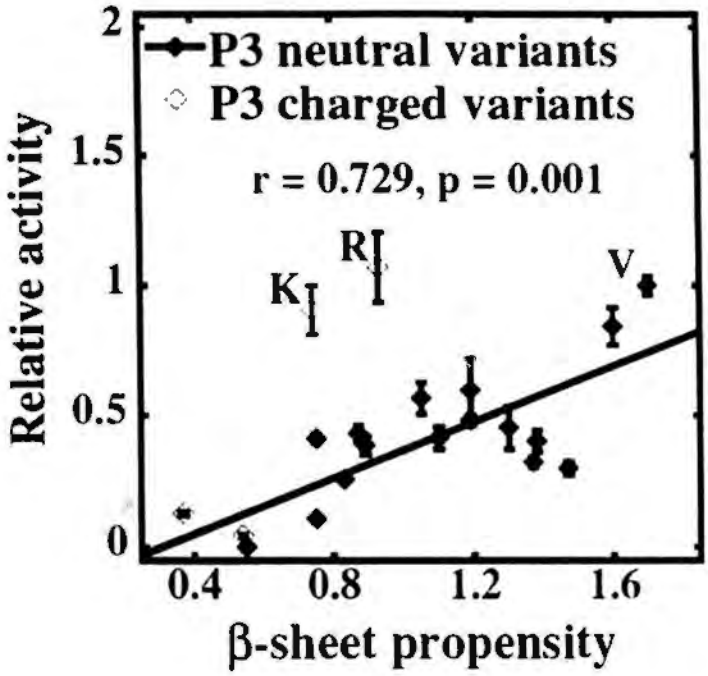


Figure 4.6 – P3 position preferred residues with high β-sheet propensity. A strong correlation (r = 0.729, p = 0.001) was found between the activity against neutral P3 variants (20 substrates except V3D, V3E, V3K and V3R) and β-sheet propensity.

4.3.4 P2 Position Prefers Hydrophobic Residues without β-Branch

Detectable cleavage was only observed for hydrophobic substitutions at P2 position (Figure 4.7). When all 20 residues were included in the correlation analysis, the relative activity was found to correlate with hydrophobicity ($r = 0.590$, $p = 0.006$) (Table 4.2). The most favored residue at P2 position was Leu (1.00 ± 0.08), followed by Met (0.68 ± 0.06) and Phe (0.42 ± 0.05). On the other hand, β-branched residues like Ile (0.13 ± 0.01) and Val (0.09 ± 0.01) were less preferred, although their hydrophobicity is similar to that of Leu. Taken together, our results suggest that P2 position prefers hydrophobic residues without β-branch.

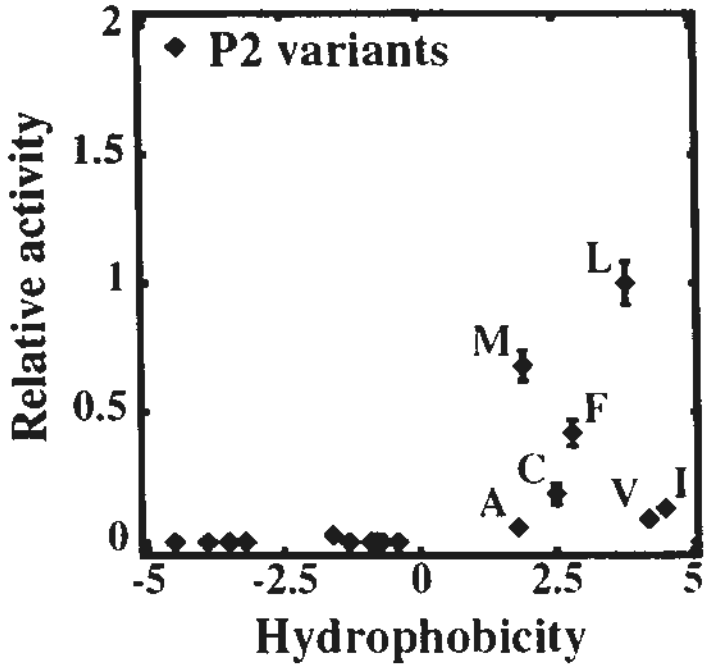


Figure 4.7 – P2 position preferred hydrophobic residues without β-branch. Only variants with hydrophobic residues (Ala, Cys, Ile, Leu, Met, Phe, and Val) at P2 position were cleavable.

4.3.5 P1 Position Tolerates His and Met

The substrate was cleavable when P1 position was a Gln, His or Met (Figure 4.2). Other substitutions were not cleavable. The most favorable residue was Gln, which is an invariant residue at P1 position of the $3CL^{pro}$ substrate sequences. Substitution to His or Met resulted in reduced relative activities of 0.26 ± 0.02 and 0.10 ± 0.01, respectively. Our observation that P1-His was cleavable is consistent with another study by Goetz *et al.* based on tetrapeptide substrates (Goetz *et al.*, 2007). However, in their cases, the activity of P1-His substrate was even higher than that of the WT sequence of P1-Gln. In the crystal structure of $3CL^{pro}$-substrate complex, the $O\varepsilon_1$ and $N\varepsilon_2$ atoms of P1-Gln form hydrogen-bonds to $N\varepsilon_2$ atom of His163 and backbone carbonyl group of Phe140, respectively (Figure 4.8). We modeled how $3CL^{pro}$ recognizes P1-His using SWISS-PDBViewer (Guex & Peitsch, 1997). In the modeled structure, although P1-His can fit into substrate binding pocket without steric hindrance, it is no longer in an optimal position to form hydrogen bonds with His163 and Phe140 (Figure 4). Instead, the $N\varepsilon_2$ atom of P1-His position can form a hydrogen bond with the amide group of Asn142. From this point of view, substitution of His at P1 position should weaken the enzyme-substrate interaction, which justified our observation that the P1-His is a poorer substrate than P1-Gln.

On the other hand, we also modeled Met at P1 position (Figure 4.8). The structure indicated that P1-Met can fit to the pocket, yet cannot interact with other residue by hydrogen bonding. So the activity on P1-Met substrate was lower than that on P1-Gln and P1-His.

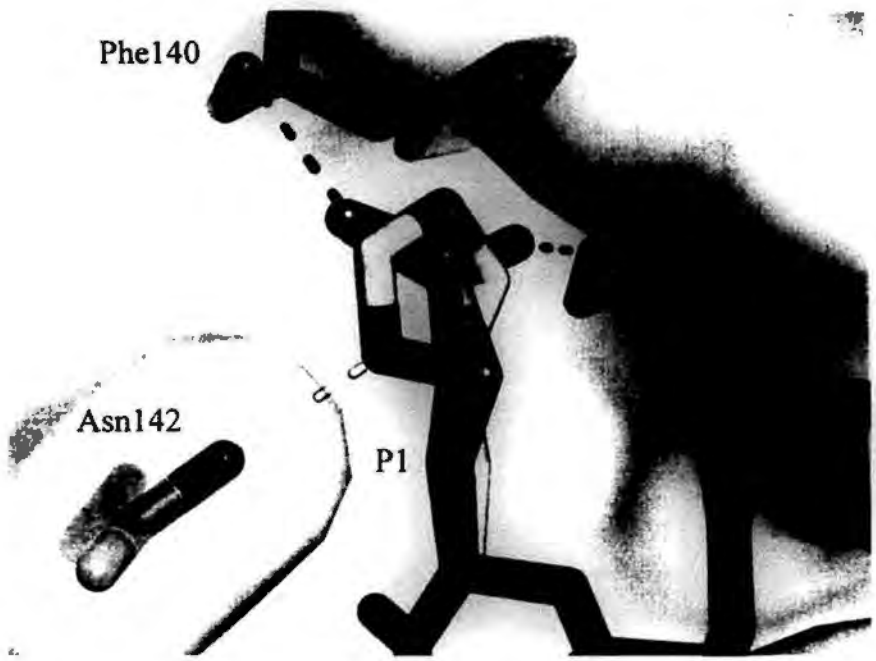


Figure 4.8 – Modeling how 3CL^{pro} recognizes P1-His and P1-Met. In the 3CL^{pro}-substrate complex (PDB: 2Q6G), amide group of P1-Gln of the WT substrate sequence (cyan) forms hydrogen-bonds with the Nε_2 atom of His163 and the backbone carbonyl group of Phe140 (Xue *et al.*, 2008). P1-Gln was substituted to His (yellow) and Met (orange) *in silico* using the program SWISS-PDBViewer (Guex & Peitsch, 1997). The rotamers of P1-His and P1-Met were selected to avoid steric hindrance and to optimize for hydrogen bond formation. The modeled structure was then energy minimized using a GROMOS force-field implemented in SWISS-PDBViewer. It was found that P1-His and P1-Met can fit into the substrate binding pocket and P1-His can form hydrogen bond to the amide group of side chain of Asn142.

4.3.6 P1' Position Prefers Small Residues

The relative activity of P1' variants negatively correlated with the side chain volume ($r = -0.660$, $p = 0.002$), suggesting that P1' position preferred small residues (Table 4.2). Ser (1.00 ± 0.08), Ala (0.99 ± 0.06), Cys (0.97 ± 0.18), and Gly (0.78 ± 0.08) at P1' position were apparently higher in activity (Figure 4.2). Substitutions with residues larger than Cys resulted in dramatic decreases in the relative activity (Figure 4.9). Taken together, P1' position prefers small residues with side chain volumes less than 50 $Å^3$.

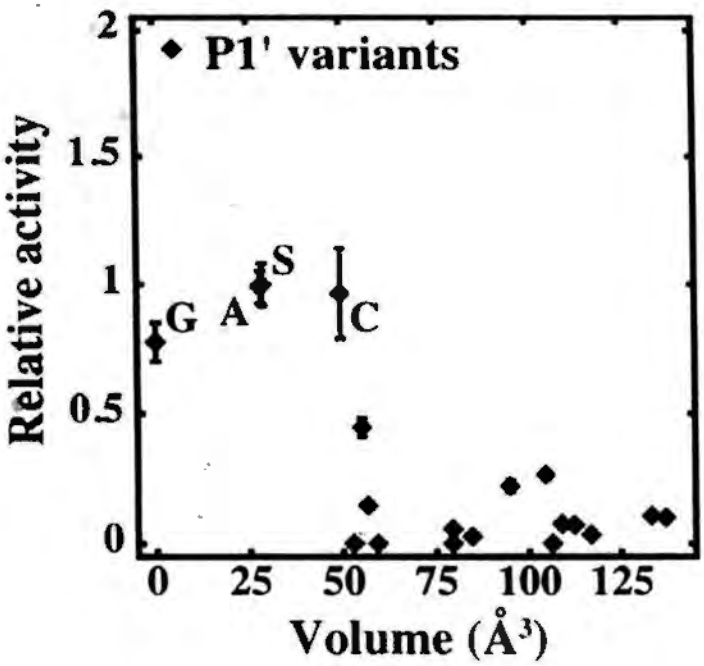


Figure 4.9 – P1' position preferred small residues. The relative activity on P1' variants with side chain volume of < 50 Å^3 (Ala, Cys, Gly and Ser) were higher than that on others.

4.3.7 P2' Position Tends to Prefer Small Residues

No significant correlation was found at P2' position (Table 4.2). However, it was noted that small residues such as Gly (1.00 ± 0.05), Ala (1.05 ± 0.19) and Ser (1.29 ± 0.12) tend to have higher relative activity than the other large residues at P2' position (Figure 4.10).

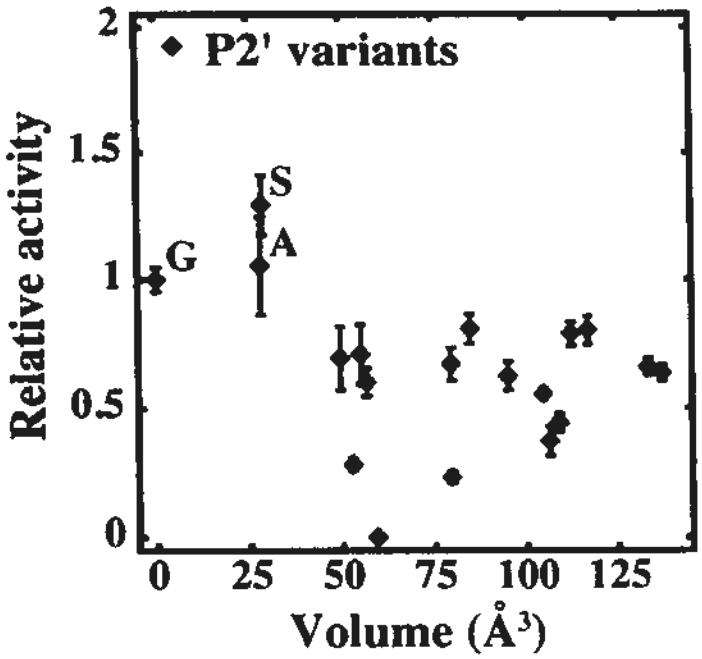


Figure 4.10 – P2' position tended to prefer small residues. The relative activity on variants with Ala, Gly and Ser at P2' position was higher than that on others.

4.3.8 P3' Position has No Strong Correlation with Any Property

No strong correlation was observed at P3' position (Table 4.2). This position only tended to prefer positively charged residues, as relative activities on F3'R (1.10 ± 0.03) and F3'K (0.94 ± 0.13) were 3-fold higher than that on F3'D (0.26 ± 0.02) and F3'E (0.42 ± 0.04) (Figure 4.2).

4.4 Combining Preferred Residues Generated 'Super-Active' Substrate Sequences

Substitutions to Phe, Thr and Val at P5 position and to Val at P4 position resulted in significant increases in 3CL^{pro} activity (Figure 4.2 and table 4.1). Also, P3 position favors positively charged residues. To test if a 'super-active' substrate sequence can be generated by combining the best substitutions at these positions, three variants with double-substitution (FVVLQ↓SGF, TVVLQ↓SGF and VVVLQ↓SGF) and three variants with triple-substitution (FVRLQ↓SGF, TVRLQ↓SGF and VVRLQ↓SGF) were created. The relative activity of 3CL^{pro} against these substrate sequences was determined (Figure 4.11). In general, the relative activity was further increased by introduction of more favorable substitutions. Triple substitution resulted in the best substrate sequence, TVRLQ↓SGF, with a relative activity of 2.84 ± 0.25. Noteworthy, docking simulation by Phakthanakanok *et al.* ranked TVKLQ↓AGF and TVRLQ↓AGF as the sequences with the lowest docking energy for 3CL^{pro}-substrate interaction (Phakthanakanok *et al.*, 2009).

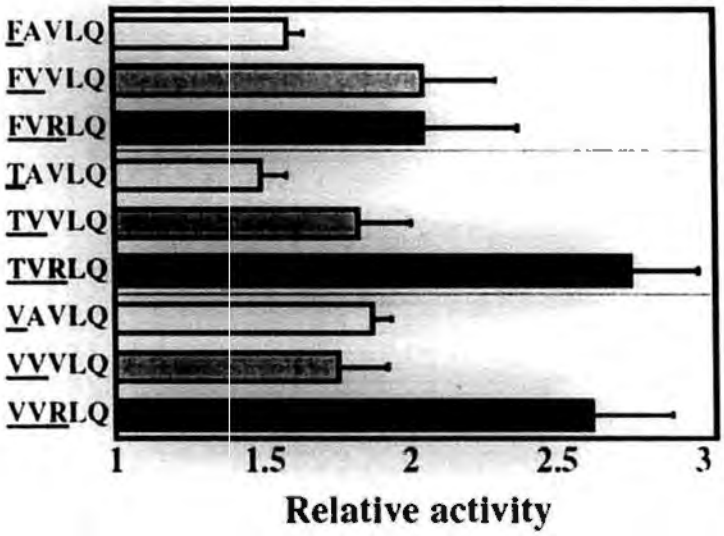


Figure 4.11 – Increase in 3CL^{pro} activity by combining the most preferred residues at P5 to P1 positions. Three 'super-active' substrate variants with double-substitution (grey bar) and three variants with triple-substitution (solid bar) were created, and their relative activities were measured. The relative activities of FVVLQ↓SGF, TVVLQ↓SGF, VVVLQ↓SGF, FVRLQ↓SGF, TVRLQ↓SGF and VVRLQ↓SGF were 2.11 ± 0.26, 1.87 ± 0.19, 1.80 ± 0.17, 2.10 ± 0.34, 2.84 ± 0.25 and 2.71 ± 0.29, respectively.

4.5 Summary

SARS-CoV 3CL^{pro} activity against the substrate library of 19×8 variants was examined by FRET assay. The activity was correlated to structural properties of substituting residues to determine the specificity at P5 to P3' positions (Table 4.3). P5 position prefers residues with high β-sheet propensity, and also tends to accommodate hydrophobic residues and positively charged residues. P4 position prefers small hydrophobic residues, and A4V is the most preferred variant. P3 position clearly prefers residues with positive charge and high β-sheet propensity. P2 position only tolerates hydrophobic residues without β-branch. The 3CL^{pro} activity against variants with P2-Leu is the highest. P1 position absolutely prefers Gln. Variants with P1-His and P1-Met are cleavable in a low rate. P1' position only tolerates small residues. Any P1' variants with side chain of > 50 Å^3 lead to dramatic decrease in activity. P2' and P3' positions tend to accept small and positively charged residues respectively, but these preferences are not obvious as those at other positions. A series of super-active substrates were generated by combining preferred residues at P5 to P1 positions. The combinations generally increase the protease activity. The most active substrate with sequence 'TVRLQ' has 2.8 fold higher activity than the WT sequence.

Table 4.3 – Summary of SARS-CoV 3CL^{pro} substrate specificity at P5 to P3' positions.

Position	WT residue	Preference	The most preferred residue (relative activity ± SD)
P5	Ser	Residues with high β-sheet propensity Positively charged residues	Val (1.92 ± 0.07) Phe (1.62 ± 0.06) Thr (1.52 ± 0.09)
P4	Ala	Small hydrophobic residues	Cys (1.32 ± 0.24) Val (1.30 ± 0.15)
P3	Val	Positively charged residues Residues with high β-sheet propensity	Arg (1.07 ± 0.13) Val (1.00 ± 0.04)
P2	Leu	Hydrophobic residues without β-branch	Leu (1.00 ± 0.08)
P1	Gln	Gln	Gln (1.00 ± 0.08)
P1'	Ser	Small residues with side chain < 50 Å^3	Ser (1.00 ± 0.08) Ala (0.99 ± 0.06) Cys (0.97 ± 0.18)
P2'	Gly	Small residues	Ser (1.29 ± 0.12)
P3'	Phe	Positively charged residues	Arg (1.10 ± 0.03)

Chapter 5 – Peptidomimetic Inhibitors Targeting SARS-CoV 3CL^{pro}

SARS-CoV 3CL^{pro} inhibition suppresses virion replication and viral-induced cytopathic effect, and is a potential strategy against SARS. Nitrile-based peptidomimetic inhibitors have been widely used in therapy of cathepsins-related diseases. To investigate whether nitrile group inhibits SARS-CoV 3CL^{pro}, we produced a series of nitrile-based peptidomimetic inhibitors and examined their potency in terms of IC_{50} and K_i values. Potent inhibitors were soaked into 3CL^{pro} crystals to obtain protease-inhibitor complexes so as to understand structure-activity relationship of inhibitors.

5.1 Nitrile Group is an Effective Warhead against 3CL^{pro}

Nitrile group inactivates Cys protease by forming reversible covalent bond with Cys residue in catalytic site. To examine whether the nitrile warhead is functional in inhibiting 3CL^{pro} activity, we produced peptidomimetic inhibitors that nitrile groups linking autocleavage sequence-like peptides (AVLQ) with various protective groups including carbobenzyloxy (cbz), 5-methylisourea (miu) and *tert*-butyl carbamate (boc), followed by determining inhibitor potencies in terms of IC_{50} (Table 5.1).

The inhibitors 'cbz-AVLQ-CN', 'miu-AVLQ-CN' and 'boc-AVLQ-CN' at different concentrations were pre-mixed with 3 μM of SARS-CoV 3CL^{pro}. The protease activity was obtained by FRET assay using protein-based WT substrate. As extra DMSO was added to the reaction mixture, so influence of DMSO on 3CL^{pro}

activity was investigated. Concentration of DMSO was inversely proportional to the activity, so reaction condition of the FRET assay for IC_{50} determination was standardized to 2.5 % (v/v) DMSO (Figure 5.1). IC_{50} was obtained by fitting kinetic to sigmoidal dose-response equation:

$$\text{Relative activity} = \frac{1}{1 + \left(\frac{[\text{Inhibitor}]}{IC_{50}}\right)^{\text{Slope}}}$$

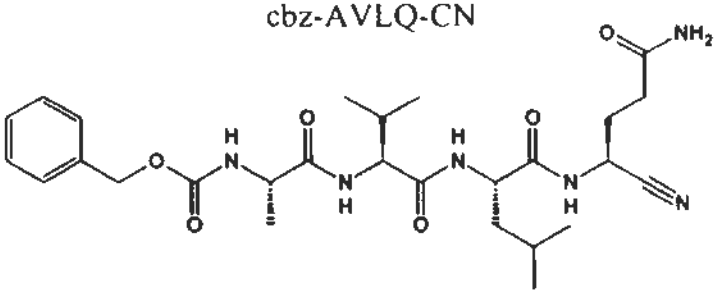
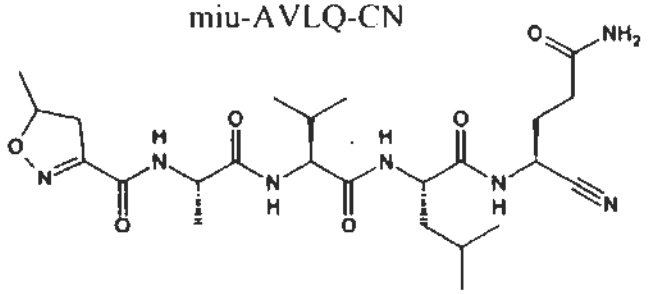
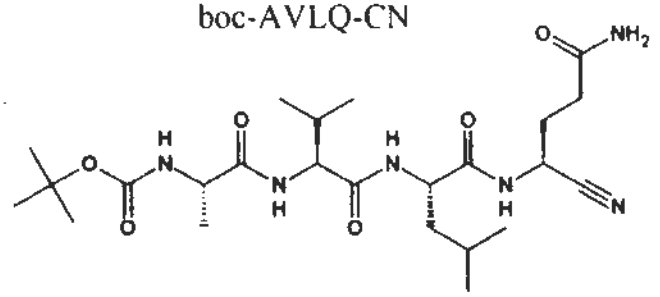
The IC_{50} values of 'cbz-AVLQ-CN', 'miu-AVLQ-CN' and 'boc-AVLQ-CN' were 5.9 ± 0.6 μM, 45 ± 3 μM, and 59 ± 5 μM respectively (Figure 5.2). All of the examined nitrile-based inhibitors could inhibit $3CL^{pro}$ activity in micromolar range, showing that nitrile group is an effective warhead against the protease. In addition, the inhibitor with cbz group was more potent than others, suggesting cbz group can enhance inhibitor potency.

K_i values of these nitrile-based peptidomimetic inhibitors were measured, in order to investigate how tightly these inhibitors bind to $3CL^{pro}$. First, K_m value of 1 μM of SARS-CoV $3CL^{pro}$ against 12-residue autocleavage sequence was obtained by FRET assay using various concentration of peptide-based substrate linked DABCYL/EDANS fluorescent quenching pair (DABCYL–KTSAVLQ↓SGFRKME–EDANS) (Figure 5.3). The peptide-based substrate, having higher solubility than the protein-based one, is more appropriate for K_i determination. The peptide-based substrate tended to precipitate in concentration higher than 150 μM, thus 25 – 150 μM of the peptide-based substrate was used. Cleavage of the peptide by $3CL^{pro}$ separated the fluorophore EDANS from the quencher DABCYL, and resulted in increases in fluorescence at 486 nm (Figure 5.3). Reaction rates (k_{obs}×substrate concentration) were determined by monitoring increase in the fluorescence intensity. A graph of reaction rate against substrate concentration was plotted and kinetic was

fitted by Michaelis-Menten equation to obtain K_m value, which was 194 ± 9 μM (Figure 5.4).

Next, K_m' values of $3CL^{pro}$ pre-incubated with 1 – 5 μM of 'cbz-AVLQ-CN', 7 – 35 μM of 'miu-AVLQ-CN' and 10 – 50 μM of 'boc-AVLQ-CN' were determined (Figure 5.5). Graphs of K_m' verse inhibitor concentrations were plotted to yield straight lines and thus K_i values (Figure 5.6). The K_i values of 'cbz-AVLQ-CN', miu-AVLQ-CN and boc-AVLQ-CN were 0.62 ± 0.11 μM, 7.0 ± 0.7 μM and 11.4 ± 1.1 μM respectively, demonstrating these nitrile-based inhibitors can bind tightly to the $3CL^{pro}$. The K_i value of inhibitor with cbz was consistently lower than that with miu and boc, showing that presence of cbz group leads to a high binding affinity towards the $3CL^{pro}$.

Table 5.1 – Chemical structures, IC_{50} and K_i values of 'cbz-AVLQ-CN', 'miu-AVLQ-CN' and 'boc-AVLQ-CN'.

Chemical structure	IC_{50} (μM)	K_i (μM)
cbz-AVLQ-CN 	5.9 ± 0.6	0.62 ± 0.11
miu-AVLQ-CN 	45 ± 3	7.0 ± 0.7
boc-AVLQ-CN 	59 ± 5	11.4 ± 1.1

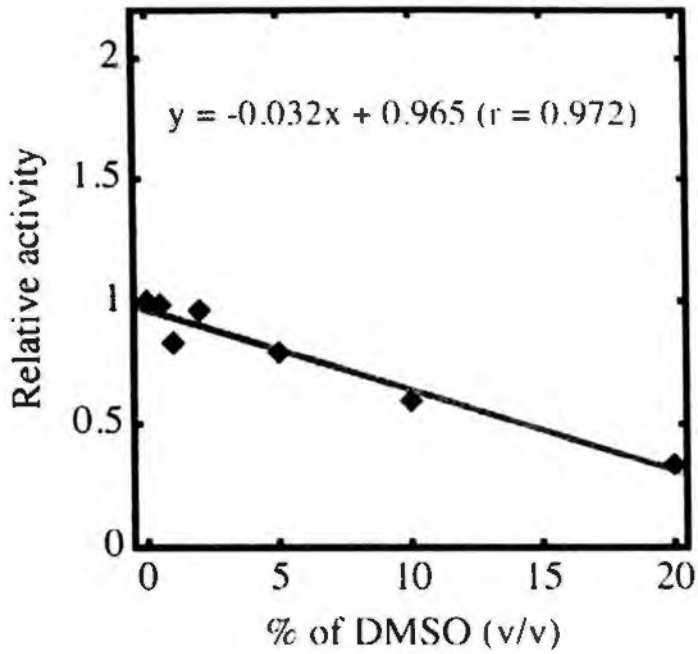


Figure 5.1 – DMSO decreased 3CL^{pro} activity. FRET assay was performed in 1 – 20 % (v/v) of DMSO. 3CL^{pro} activity relative to control without DMSO was determined. The relative activity was reversely proportional to DMSO concentration.

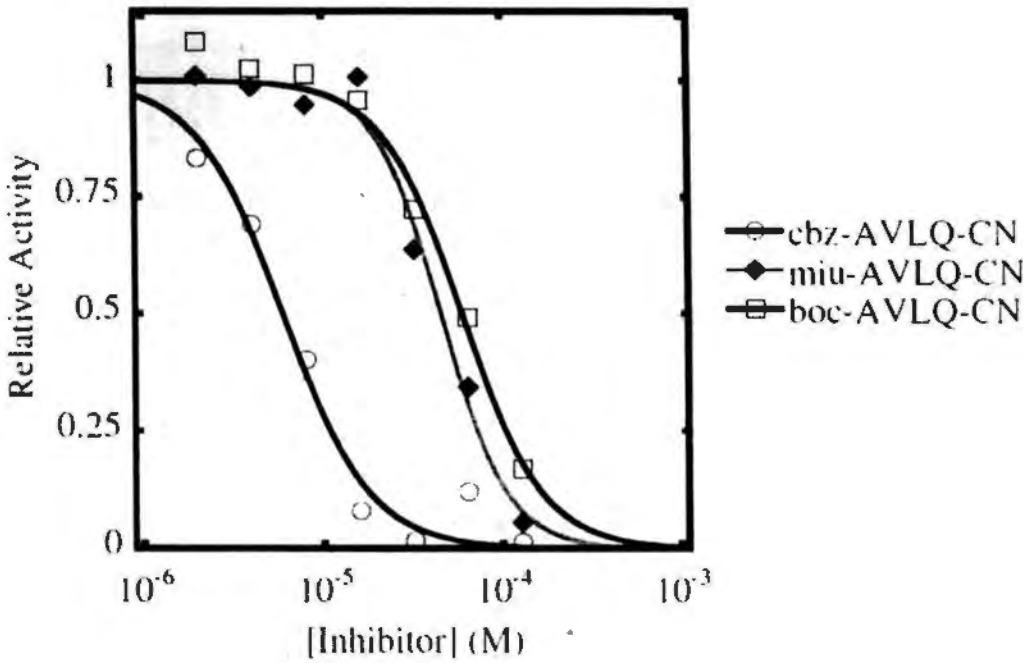


Figure 5.2 – Nitrile-based inhibitors inhibited 3CL^{pro} activity. 2 – 128 μM of inhibitors 'cbz-AVLQ-CN', 'miu-AVLQ-CN' and 'boc-AVLQ-CN' were pre-mixed with 3 μM of SARS-CoV 3CL^{pro}. Relative activity was determined by FRET assay. IC_{50} values were obtained by fitting kinetics by sigmoidal dose-response equation. Their IC_{50} values were 5.9 ± 0.6, 45 ± 3, and 59 ± 5 μM respectively. All inhibitors reduced the protease activity. 'cbz-AVLQ-CN' was the most potent.

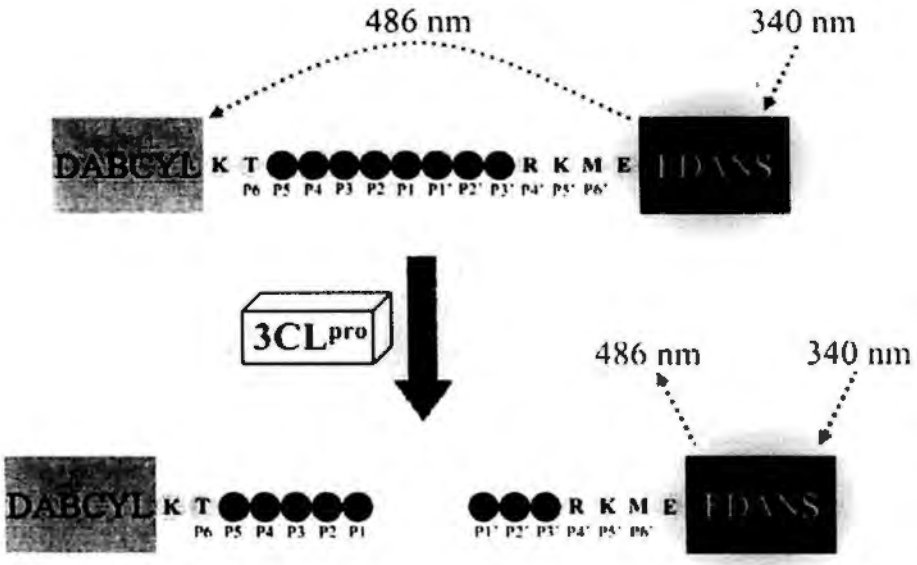
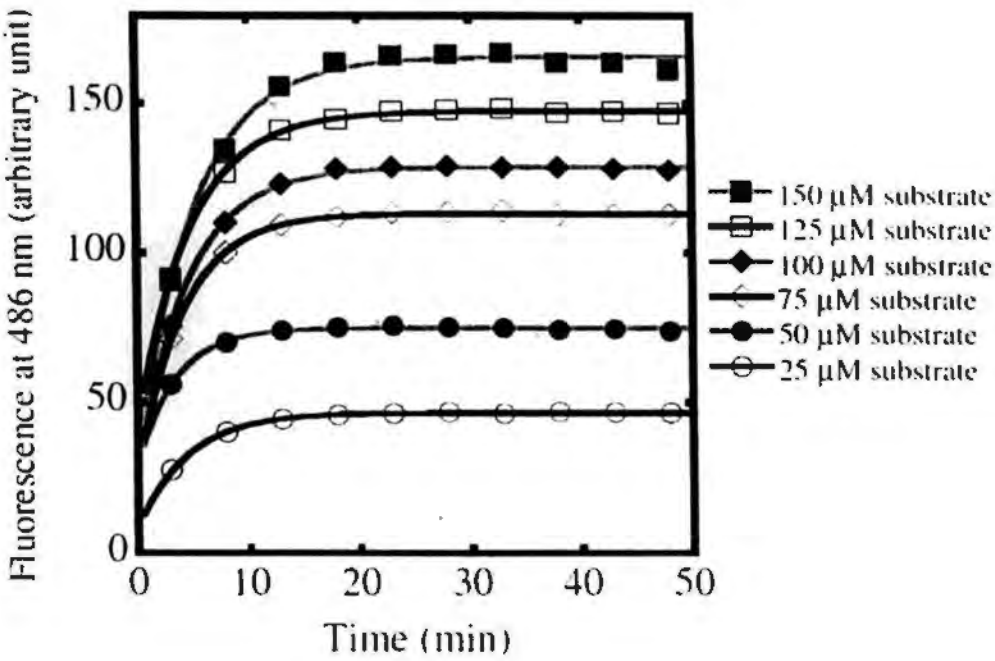


Figure 5.3 – Schematic diagram illustrating measurement of cleavage of peptide-based substrate by FRET. The peptide-based substrate consisted of 12-residue autocleavage sequence (TSAVLQ↓SGFRKM) inserted between fluorophore EDANS and quencher DABCYL. EDANS in intact substrate is excited by light at 340 nm to emit fluorescence at 486nm but quenched by the linked DABCYL. Cleavage of substrate sequence separates EDANS and DABCYL, causing increase in the emitted fluorescence intensity.

(A)



(B)

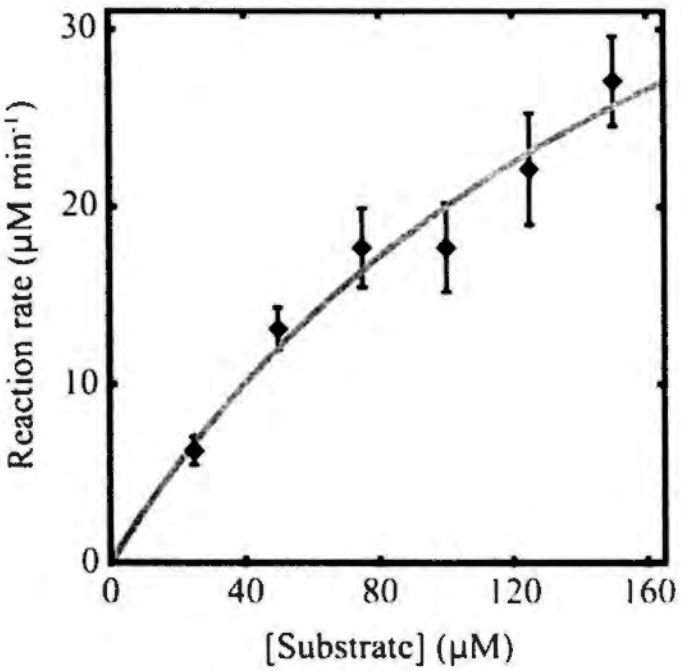


Figure 5.4 – K_m value of 3CL^{pro} against the peptide-based substrate. **(A)** Cleavage of 25 – 150 μM of the peptide-based substrate by 1 μM of the 3CL^{pro} caused increase in the emitted fluorescence at 486 nm. k_{obs} values were determined by fitting single exponential decay equation. **(B)** Plot of reaction rate (k_{obs} × substrate concentration) against substrate concentration yielded a curve. K_m value obtained by Michaelis-Menten equation was 194 ± 9 μM.

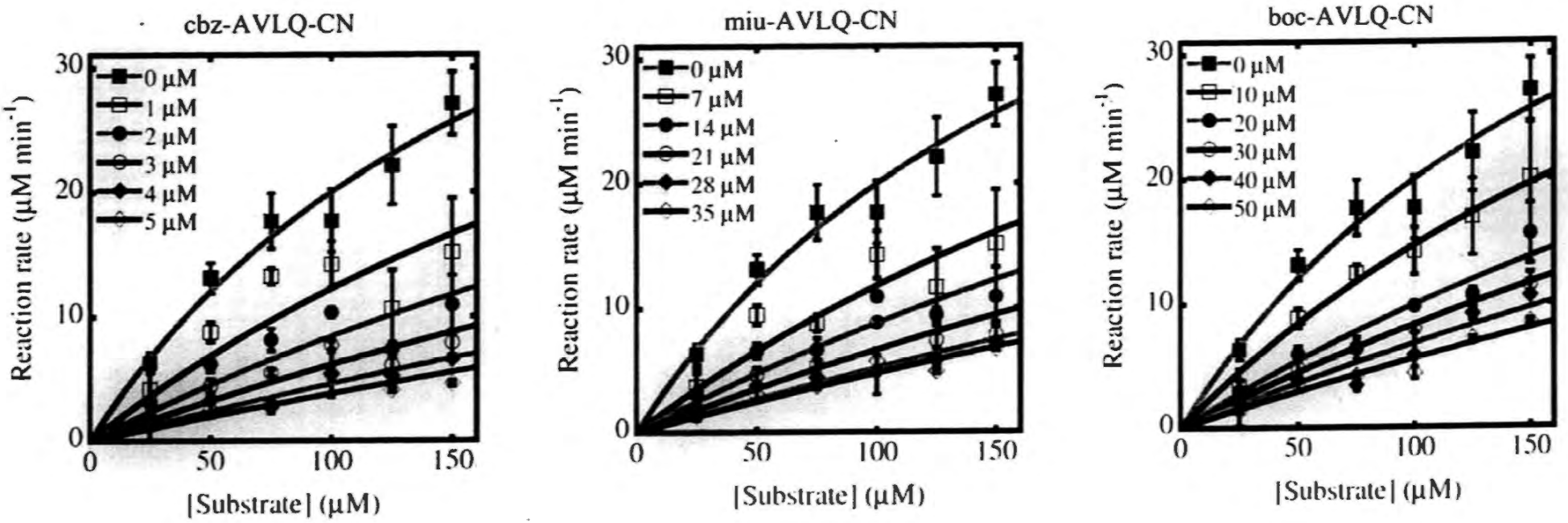


Figure 5.5 – K_m' values of nitrile-based tetrapeptidomimetic inhibitors. Km' values of 1 μM of SARS-CoV 3CL^{pro} pre-mixed with 1 – 5 μM of 'cbz-AVLQ-CN', 7 – 35 μM of 'miu-AVLQ-CN' and 10 – 50 μM of 'boc-AVLQ-CN' were determined. Higher inhibitor concentrations were provided, higher K_m' values were obtained.

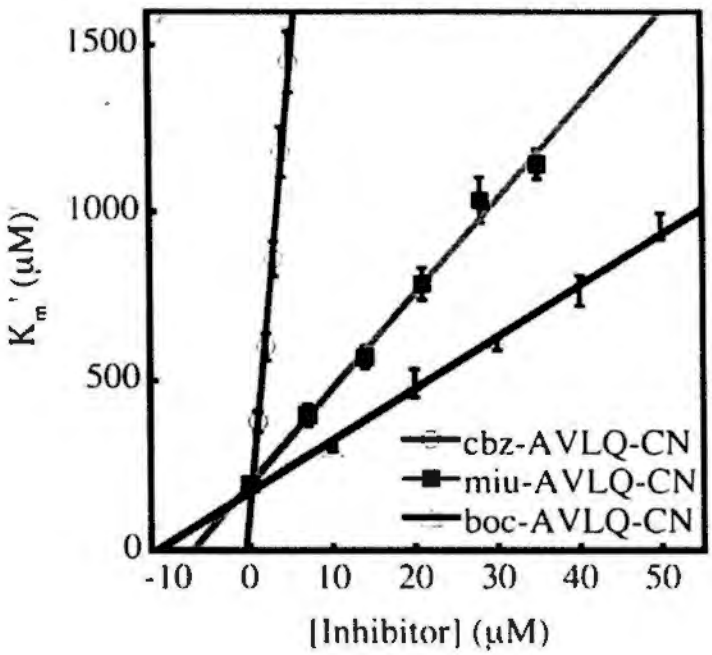
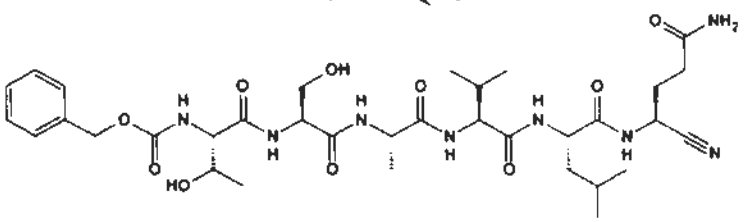


Figure 5.6 – K_i values of nitrile-based tetrapeptidomimetic inhibitors. Plots of K_m' values against inhibitor concentration yielded straight lines. Y-intercept (K_m) was divided by slope of the lines (K_m/K_i) to obtain K_i values of 'cbz-AVLQ-CN', 'miu-AVLQ-CN' and 'boc-AVLQ-CN', which were 0.62 ± 0.11 μM, 7.0 ± 0.7 μM and 11.4 ± 1.1 μM respectively.

5.2 *Lengthening Peptide has No Improvement on Inhibitor Potency*

Inhibitors with longer peptide may be more potent (Akaji *et al.*, 2008; Shao *et al.*, 2008), so inhibitor with hexpeptide 'cbz-TSAVLQ-CN' was synthesized and examined (Table 5.2). Its IC_{50} and K_i values were 7.7 ± 0.9 μM and 5.9 ± 0.1 μM respectively (Figures 5.7 – 5.9). Its potency was similar to that of 'miu-AVLQ-CN', but was lower than that of 'cbz-AVLQ-CN'.

Table 5.2 – Chemical structure, IC_{50} and K_i values of 'cbz-TSAVLQ-CN'.

Chemical structure	IC_{50} (μM)	K_i (μM)
cbz-TSAVLQ-CN 	7.7 ± 0.9	5.4 ± 0.5

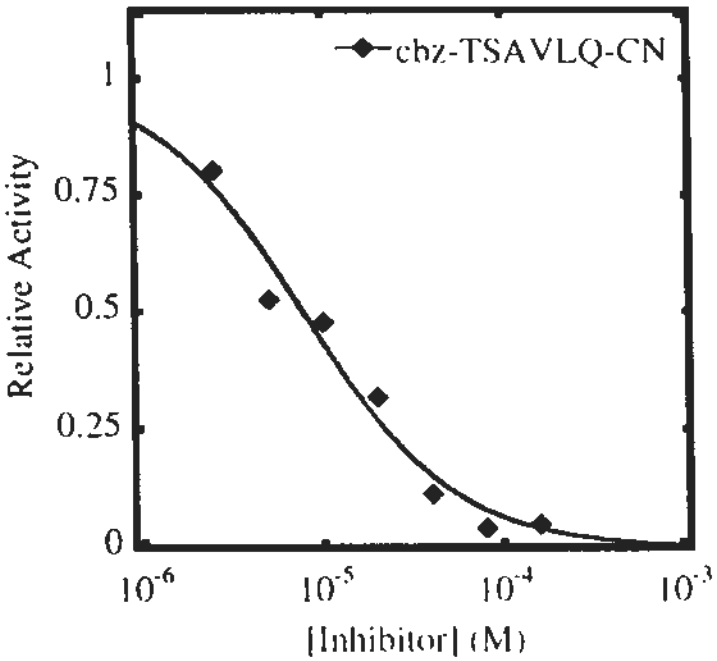


Figure 5.7 – Lengthening inhibitor had no improvement on IC_{50} value. 2.5 – 160 μM of cbz-TSAVLQ-CN was pre-mixed with 3 μM of SARS-CoV 3CL^{pro}, followed by determining relative activity by FRET assay. Curve was fitted by sigmoidal dose-response equation to obtain IC_{50} value, which was 7.7 ± 0.9 μM.

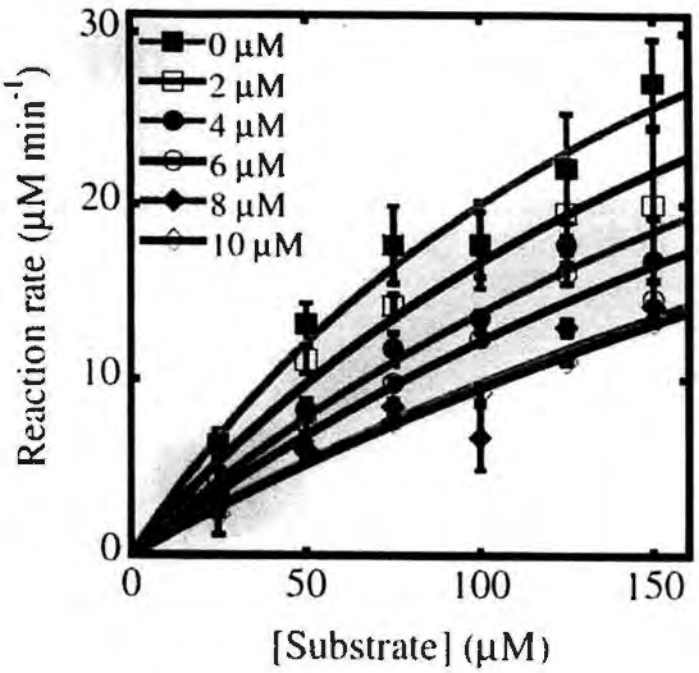


Figure 5.8 – K_m' values of 'cbz-TSAVLQ-CN'. 1 μM of SARS-CoV 3CL^{pro} was pre-mixed with 2 – 10 μM of 'cbz-TSAVLQ-CN', followed by determining its K_m' value.

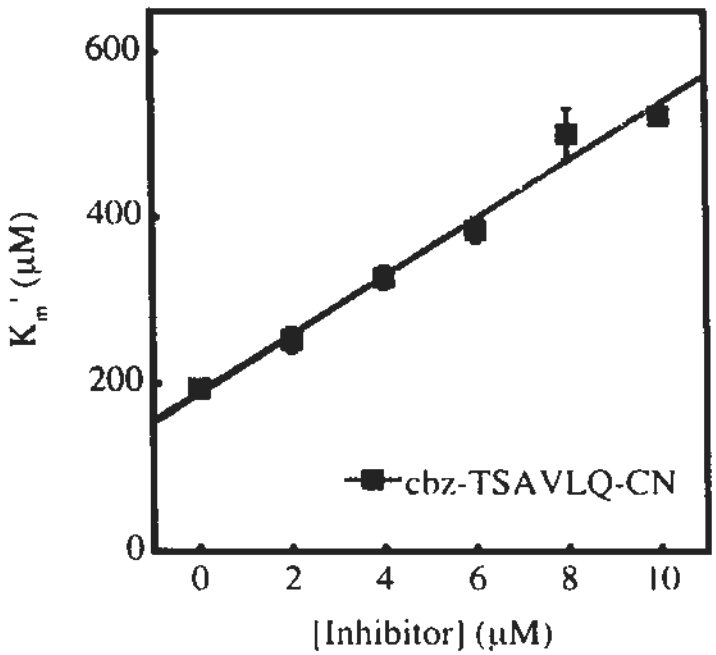
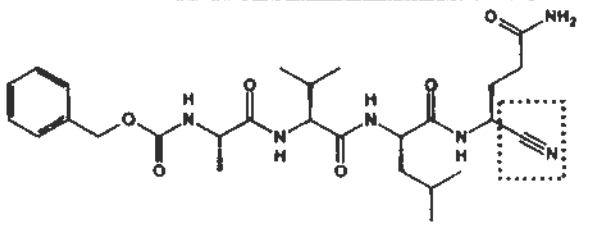
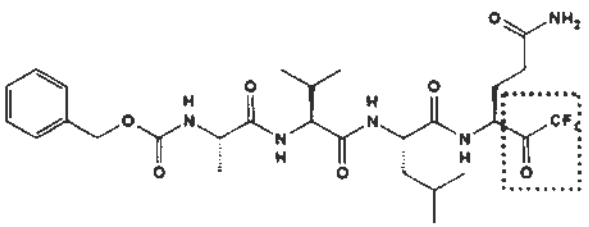
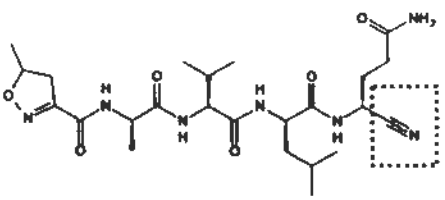
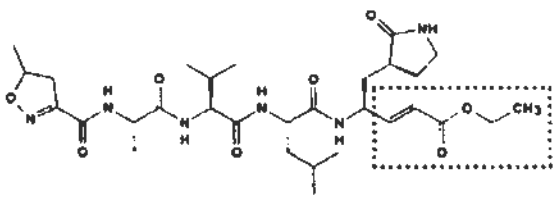
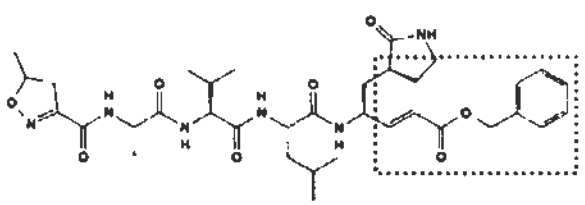


Figure 5.9 – K_i value of 'cbz-TSAVLQ-CN'. Plot of K_m' values against inhibitor concentration yielded straight lines. K_i value was obtained by y-intercept (K_m) divided by slope of the line (K_m/K_i). K_i value of 'cbz-TSAVLQ-CN' was 5.4 ± 0.5 μM.

5.3 Reactivity Comparison of Nitrile Group with Other Warheads

To understand reactivity of nitrile group on the 3CL^{pro}, K_i values of nitrile-based inhibitors were compared with other structurally comparable inhibitors with different warheads (Table 5.3). Regnier *et al.* synthesized tetrapeptidomimetic inhibitor with trifluoromethylketone warhead and cbz group (Regnier *et al.*, 2009). Its K_i value was 135 ± 32 μM, which was obviously higher than K_i value of 'cbz-AVLQ-CN' (0.62 ± 0.11 μM). On the other hand, Yang *et al.* selected α,β-unsaturated ethyl ester and benzyl ester as warheads to produce peptidomimetic inhibitors with K_i values of 10.7 ± 1.0 μM and 9.0 ± 0.8 μM respectively (Yang *et al.*, 2005). Our inhibitor 'miu'-AVLQ-CN' with K_i value of 7.0 ± 0.7 μM was slightly higher in potency. The compared inhibitors consisted of same peptide sequences and protective groups, so difference in potency demonstrated potency of warheads. The nitrile was more potent than others, and is a good model for drug development against SARS-CoV 3CL^{pro}.

Table 5.3 – Comparison of K_i values of nitrile group and other warheads. K_i values of tetrapeptidomimetic inhibitors with diverse warheads (highlighted) were showed (Regnier *et al.*, 2009; Yang *et al.*, 2005). Inhibitors with nitrile warhead were more potent than those with trifluoromethylketone and α,β-unsaturated esters

Chemical structures	K_i (μM)	References
	0.62 ± 0.11	Our work
	135 ± 32	(Regnier *et al.*, 2009)
	7.0 ± 0.7	Our work
	10.7 ± 1.0	(Yang *et al.*, 2005)
	9.0 ± 0.8	(Yang *et al.*, 2005)

5.4 Structure Determination of 3CL^{pro}-Inhibitor Complexes

5.4.1 Structures of 3CL^{pro}-Inhibitor Complexes are Comparable

To understand the inhibitory mechanism of nitrile-based peptidomimetic inhibitors against the 3CL^{pro}, inhibitors 'cbz-AVLQ-CN', 'miu-AVLQ-CN', 'boc-AVLQ-CN' and 'cbz-TSAVLQ-CN' were soaked into crystals of SARS-CoV 3CL^{pro}. Structures of 3CL^{pro}-inhibitor complexes at 1.95 – 2.5 Å resolutions were solved by X-ray diffraction and molecular replacement (Table 5.4). All soaked inhibitors were present in the substrate binding clefts of two protomers, and the nitrile groups were covalently bonded with Cys145 (Figure 5.10). F_o-F_c maps showed that interaction between all of their P1 – P4 residues and 3CL^{pro} was similar to 3CL^{pro}-substrate complex (Xue *et al.*, 2008). The maps also indicated positions of cbz and boc linked tetrapeptidomimetic inhibitors, but not that of miu in 'miu-AVLQ-CN' and 'cbz-TS' in 'cbz-TSAVLQ-CN'.

Superimposition of 3CL^{pro}-inhibitor complexes showed that positions of their nitrile group and Cys145 were almost identical (Figure 5.11). Conformations of P1 – P3 residues among these inhibitors were highly similar, while that of P4 residues were slightly varied. In the meantime, 3CL^{pro} structures bound with different inhibitors were highly comparable. All of these 3CL^{pro} were homodimer consisted of 3 domains, and no major structural change was found. When structure of 3CL^{pro} in apo form was further superimposed and compared, loops of Thr45 – Glu47 in 3CL^{pro}-inhibitor complexes tended to be far from the inhibitors, and loops of Arg188 – Ala191 in the complexes were closer to inhibitors. The structural change suggested that the inhibitors repel the loops of Thr45 – Glu47 and attract the loops of Arg188 – Ala 191.

Table 5.4 – Crystallographic data and refinement statistics. All crystals were produced by using 2.5 mg/ml of SARS-CoV 3CL^{pro} in 100 mM MES, pH 6.0, 8.5 % (w/v) PEG6000, 10 % (v/v) glycerol, 1 mM EDTA and 1 mM DTT at 16 ℃. Numbers in parenthesis indicate data of highest resolution shells. Ramachandran plot analysis was checked by using MolProbity (Davis *et al.*, 2007).

	Native (No inhibitor)	cbz-AVLQ-CN	boc-AVLQ-CN	miu-AVLQ-CN	cbz-TSAVLQ-CN
Diffraction data					
Wavelength (Å)	1.54	1.54	1.54	1.54	1.54
Resolution limit (Å)	27.28 2.20 (2.32 2.20)	27.20 1.95 (2.06 1.95)	34.10 2.20 (2.32 2.20)	24.61 1.95 (2.06 1.95)	29.61 2.50 (2.64 2.50)
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell constants					
a (Å)	52.43	52.27	52.32	52.32	52.10
b (Å)	96.74	96.40	97.10	96.49	96.43
c (Å)	67.72	67.70	67.41	67.64	67.69
α (°)	90.00	90.00	90.00	90.00	90.00
β (°)	102.73	103.26	103.17	103.30	103.48
γ (°)	90.00	90.00	90.00	90.00	90.00
Total no. of reflections	120610 (17160)	171306 (24820)	121516 (17360)	169495 (23699)	79787 (11394)
No. of unique reflections	33503 (4859)	47614 (6946)	33359 (4844)	46922 (6695)	22586 (3258)
Multiplicity	3.6 (3.5)	3.6 (3.6)	3.6 (3.6)	3.6 (3.5)	3.5 (3.5)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	98.6 (96.8)	99.9 (100.0)
R_{merge} (%)	0.119 (0.380)	0.089 (0.223)	0.109 (0.382)	0.081 (0.314)	0.098 (0.305)
I/σ	6.6 (2.9)	13.2 (3.6)	12.3 (2.2)	13.2 (3.0)	8.5 (3.4)
Structure refinement					
R_{work}	0.1872	0.1845	0.1818	0.1877	0.1883
R_{free}	0.2252	0.2188	0.2316	0.2270	0.2347
rmsd from ideal values					
Bond length (Å)	0.006	0.008	0.005	0.005	0.005
Bond angle (°)	0.963	1.232	0.933	1.057	1.018
Ramachandran plot analysis					
In preferred regions (%)	96.85	95.18	95.85	96.34	93.20
In allowed regions (%)	2.82	4.49	3.99	3.33	6.14
In disallowed regions (%)	0.33	0.33	0.17	0.33	0.66

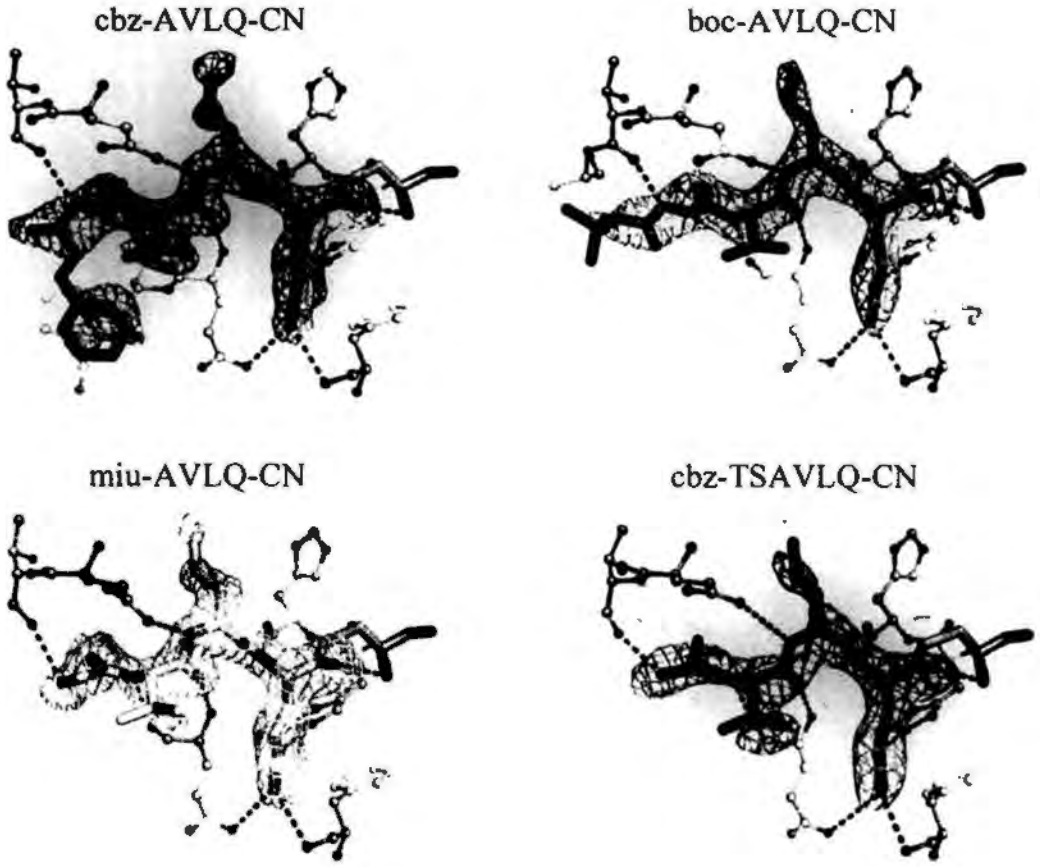


Figure 5.10 – F_o-F_c maps of 3CL^{pro}-inhibitor complexes. $F_o - F_c$ maps (mesh) of the inhibitors were contoured at 2.0 σ level. All of nitrile groups covalently bonded with Cys145, while P4 – P1 residues were consistently found in substrate binding cleft. The map defined positions of cbz group in 'cbz-AVLQ-CN' (red) and boc group in 'boc-AVLQ-CN' (green), but not miu group in 'miu-AVLQ-CN' (yellow) and 'cbz-TS' in 'cbz-TSAVLQ-CN' (blue).

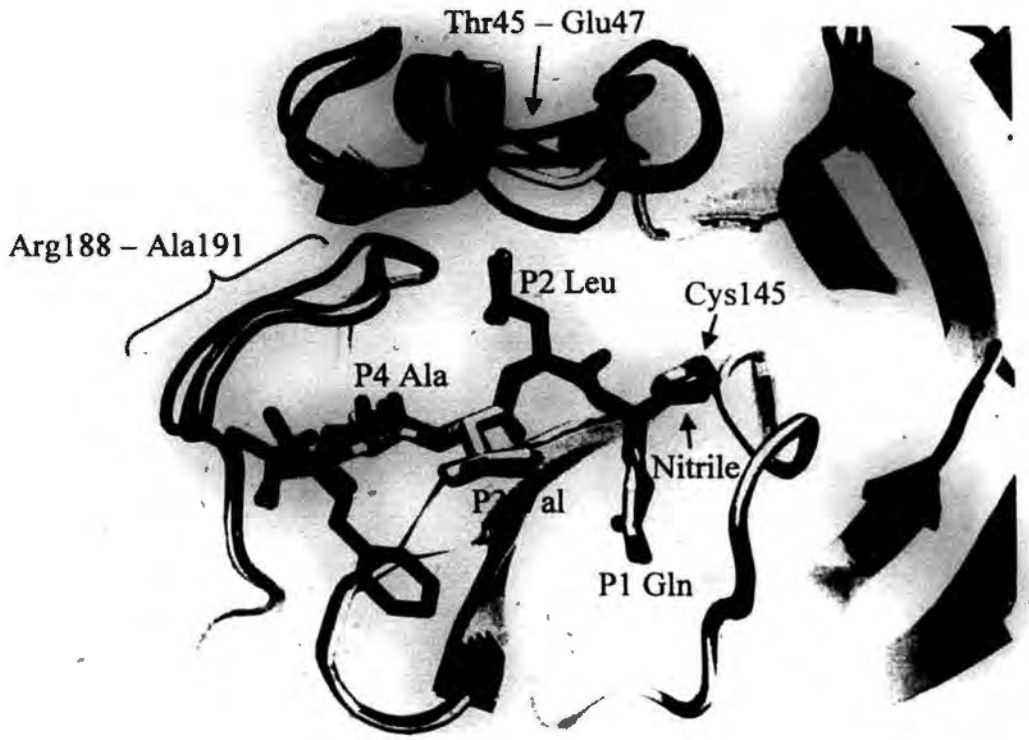
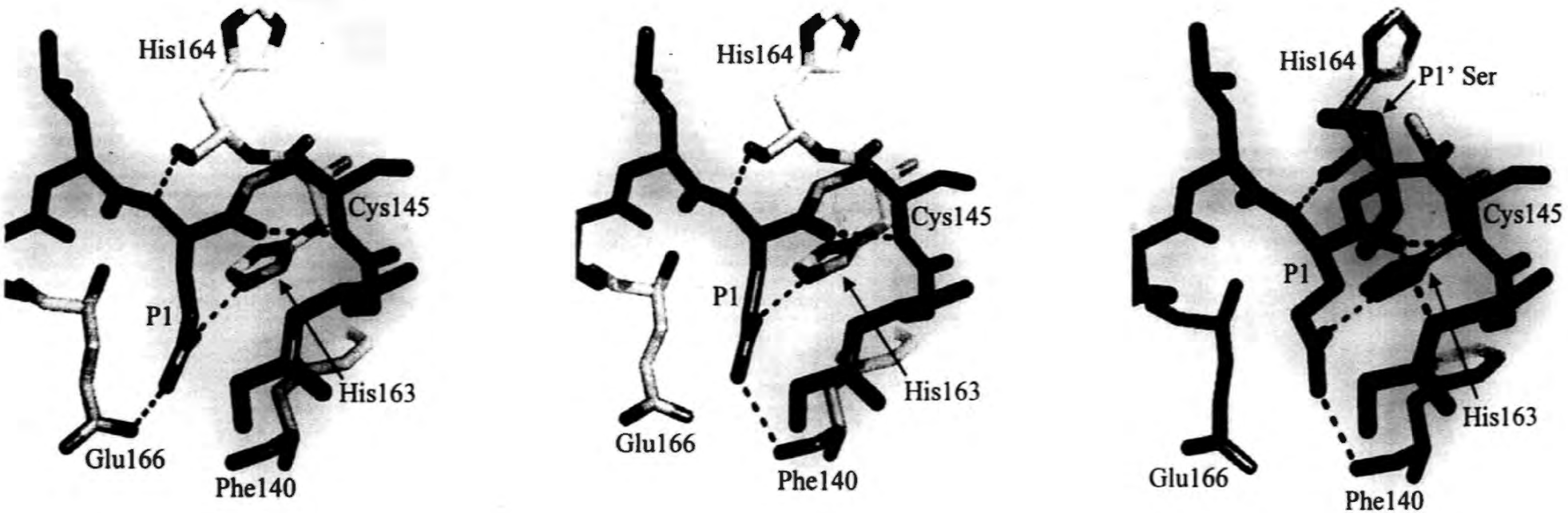


Figure 5.11 – Superimposition of 3CL^{pro}-inhibitor complexes. Structures of 3CL^{pro} with 'cbz-AVLQ-CN', 'boc-AVLQ-CN' (green), 'miu-AVLQ-CN' (yellow), (red), 'cbz-TSAVLQ-CN' (blue) and in apo form (purple) were superimposed. Conformations of all of nitrile group and P1 – P3 residues were highly comparable, while that of P4 residues were slightly varied. 3CL^{pro} bound with inhibitors shared a high similarity. By comparing with 3CL^{pro} in apo form (purple), loops of Thr45 – Glu47 in 3CL^{pro}-inhibitor complexes tended to far from the inhibitors, and loops of Arg188 – Ala191 were closer to inhibitors. No major structural change was found in other regions.

5.4.2 Nitrile Group Covalently Bonded with Cys145

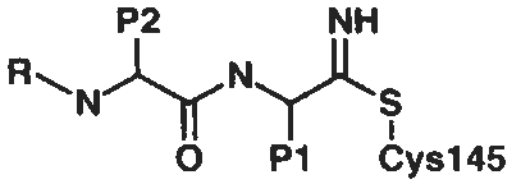
The nitrile group has been proven to inactivate $3CL^{pro}$ activity. To investigate whether the inhibitory mechanism on $3CL^{pro}$ resembles that on other Cys proteases, structure of nitrile-Cys145 was studied. In $3CL^{pro}$-complex, C atom of the nitrile group was covalently bonded with the thiol group of Cys145 and in sp^2 hybridization (Figure 5.12). N atom of the nitrile group was hydrogen-bonded with amide group of Cys145. $O\varepsilon_1$ atom of P1-Gln was hydrogen bonded with $N\varepsilon_2$ atom of His163 in both protomer A and B. However, $N\varepsilon_2$ atoms in protomer A and B bonded to side chain of Glu166 and carbonyl group of Phe140 respectively. P1 backbone amide group was hydrogen bonded to His164. By comparing with structure of $3CL^{pro}$-substrate complex, positions of N atom of nitrile group and O atom of P1-carbonyl group were comparable, but only the carbonyl group could interact with amide groups of both of Gly143 and Cys145. Other interactions between $3CL^{pro}$ and P1-Gln were consistent.

As the nitrile group is structurally comparable to carbonyl group of P1-Gln, the $3CL^{pro}$-inhibitor mimics acyl-intermediate that the carbonyl group was bonded with Cys145 (Figure 5.13). Interaction between the inhibitor and $3CL^{pro}$ is similar to that in acyl-intermediate, suggesting the nitrile-based inhibitor can bind to $3CL^{pro}$ as substrate binds.

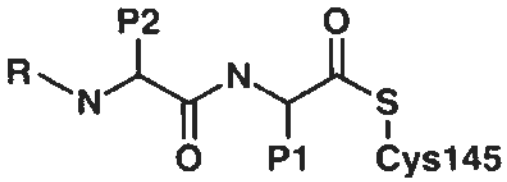


(A) Protomer A (B) Protomer B (C) 3CL^{pro}-substrate complex

Figure 5.12 – Nitrile group and P1-Gln interacted with 3CL^{pro}. (A, B) Nitrile warhead at P1 position covalently bonded with thiol group and hydrogen-bonded with amide group of Cys145. At P1 residue, amide group and $O\varepsilon_1$ consistently interacted with His164 and His163 respectively. $N\varepsilon_2$ in protomer A and B tended to bond with Glu166 and Phe140 respectively. **(C)** 3CL^{pro}-substrate complex (PDB:2Q6G) showed a comparable interaction (Xue *et al.*, 2008).



3CL^{pro}-inhibitor complex



Acyl-intermediate

Figure 5.13 – 3CL^{pro}-inhibitor complexes mimicked acyl-intermediate. Nitrile group of inhibitors covalently bonded to Cys145, mimicking P1-carbonyl group in acyl-intermediate.

5.4.3 Interaction between P1 – P4 Residues of the Inhibitors and Substrate Binding Cleft

Besides the warhead, interaction between peptide of the inhibitor and $3CL^{pro}$ also determines inhibitor potency, thus how the peptide bound to substrate binding cleft was investigated. Side chains of P1, P2 and P4 residues were enclosed by cavities, while P3 side chain was solvent-exposed (Figure 5.14). Amide group of P2-Leu was hydrogen-bonded with Gln189. Amide and carbonyl groups of P3-Val were hydrogen-bonded with main chain of Glu166. Amide group of P4-Ala was hydrogen-bonded with Ala190. Those interactions were same as that between $3CL^{pro}$ and substrate (Xue *et al.*, 2008).

As mentioned in section 5.4.1 (P.106), in presence of the inhibitors, loop of Arg188 – Ala191 moved towards the inhibitors. This movement favored these two hydrogen-bond formation between inhibitor and $3CL^{pro}$. On the other hand, $3CL^{pro}$-inhibitor structures in previous section demonstrated that the inhibitor repelled loop of Thr45 – Glu47. By comparing with $3CL^{pro}$ in apo form, side chain of Met49 was also flipped to another side (Figure 5.15). These changes are required for accommodation of side chain of P2-Leu. Otherwise, Met49 and P2-Leu were crashed.

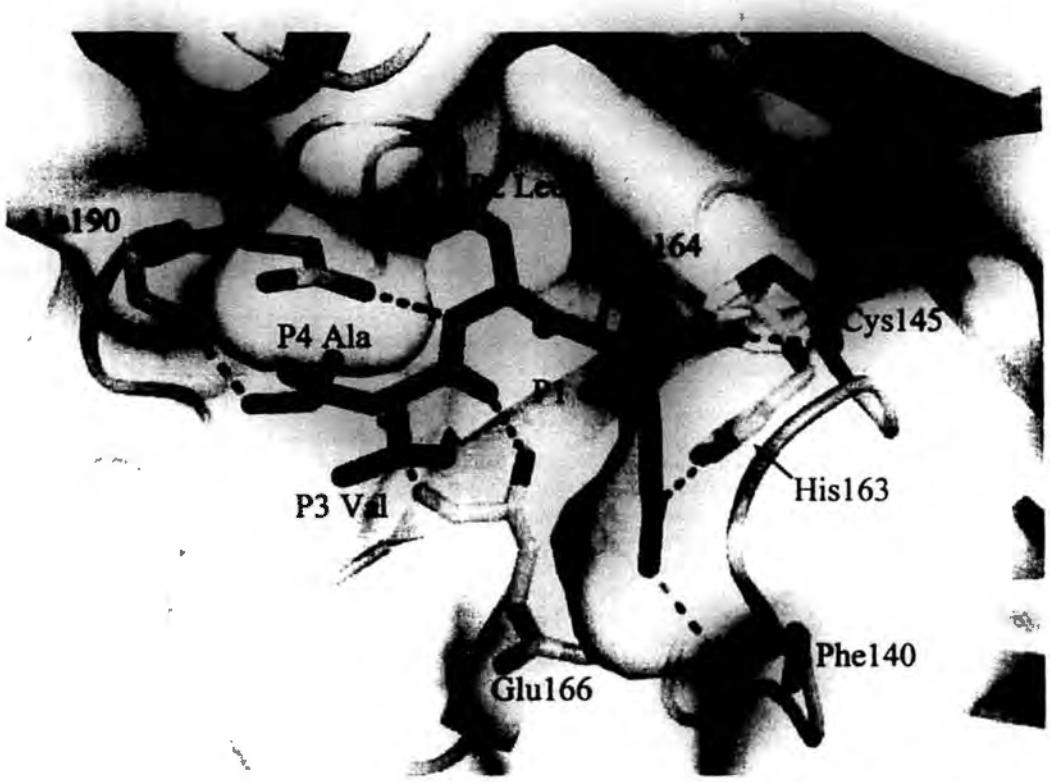


Figure 5.14 – Interaction between peptide of the inhibitor and 3CL^{pro}. Complex structure of 3CL^{pro} and 'cbz-AVLQ-CN' showed that nitrile group was bonded with Cys145. $N\varepsilon_2$ atom, $O\varepsilon_1$ atom and amide group of P1-Gln bonded with Phe140 (protomer B), His163 and His164 respectively. Amide group of P2-Leu bonded with Gln189. Amide and carbonyl groups of P3-Val bonded with Glu166. Amide group of P4-Ala bonded with Ala190.

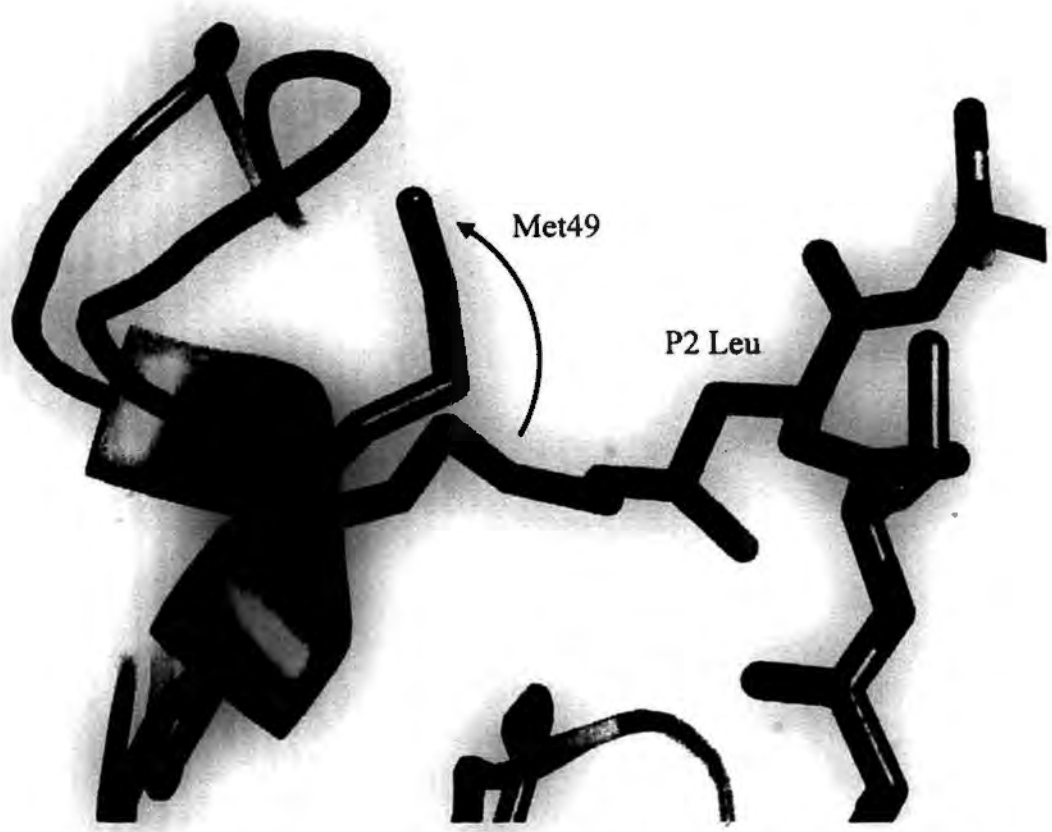


Figure 5.15 – Met49 flipped to prevent steric hindrance of P2-Leu. $3CL^{pro}$ in apo form (purple) and that bound with 'cbz-AVLQ-CN' (red) were compared. Met49 flipped so that side chain of P2-Leu was tolerated.

5.4.4 Cbz Group Flipped into Cavity of Glu166 – Pro168

Tetrapeptidomimetic inhibitor with cbz group is obviously more potent than that with other protective groups. To investigate how cbz group enhances inhibitor potency, structures of protective groups of various $3CL^{pro}$-inhibitor complexes were compared (Figure 5.16). Cbz group linked tetrapeptidomimetic inhibitor was flipped and fitted into a shallow cavity consisted of Glu166 – Pro 168, while boc group was close to Thr190 – Ala191 and pointed towards environment. Miu group and 'cbz-TS' in hexpeptidomimetic inhibitor were unstructured in our $3CL^{pro}$-inhibitor complexes. We predicted locations of miu and 'cbz-TS' by using $3CL^{pro}$-inhibitor complex consisted of 'miu-AVLQ' and $3CL^{pro}$-substrate complexes reported in previous studies (Xue *et al.*, 2008; Yang *et al.*, 2005). Miu and 'cbz-TS' were expected to be near Thr 190 – Ala191 as boc group. The unique interaction of cbz group in tetrapeptidomimetic inhibitor is believed to enhance overall binding affinity towards $3CL^{pro}$, and thus increase inhibitor potency.

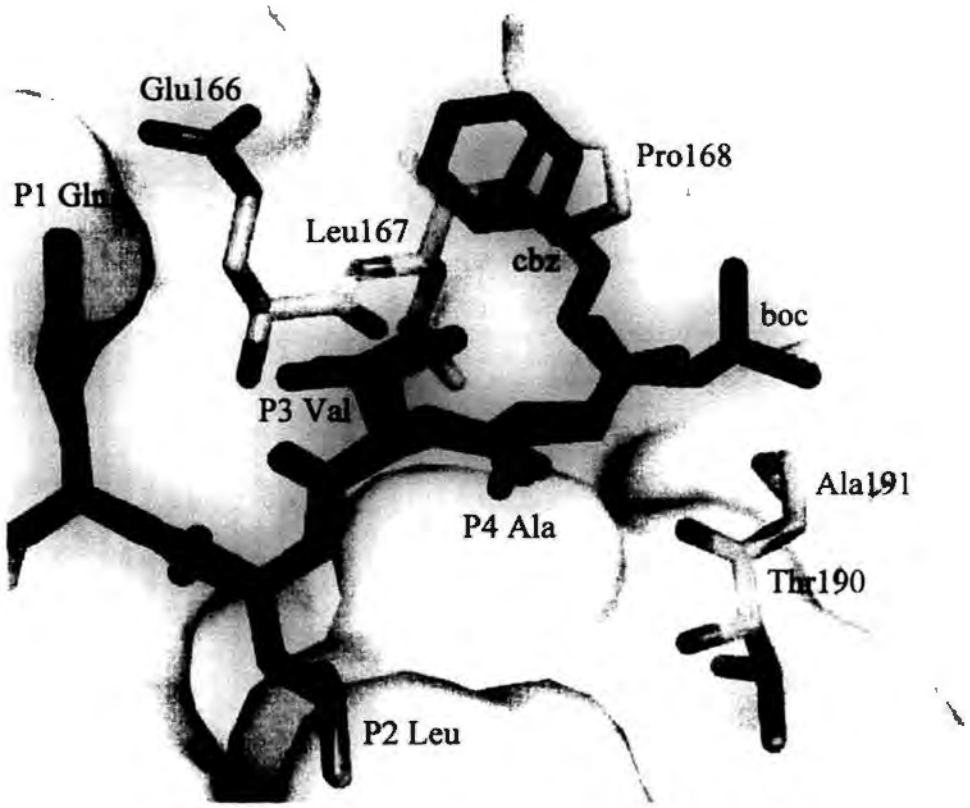


Figure 5.16 – Cbz group specifically flipped into cavity of Glu166 – Pro168. cbz group of 'cbz-AVLQ-CN' (red) flipped into cavity of Glu166 – Pro168 (orange), while boc group of 'boc-AVLQ-CN' (green) was close to Thr190 and Ala191 (dark green), and pointed towards environment.

5.5 Residue Substitution has No Improvement on Inhibitor Potency

5.5.1 Substitution to P4-Val has No Improvement on IC_{50} value

Most of substrate variants with P4-Val caused a higher $3CL^{pro}$ activity. In attempt to enhance potency of 'cbz-AVLQ-CN', Ala at P4 position was substituted to Val to create 'cbz-VVLQ-CN'. Nonetheless, its IC_{50} value was 18 ± 1 μM (Figure 5.17), which was lower in potency than 'cbz-AVLQ-CN' with IC_{50} value of 5.9 ± 0.6 μM. As secondary structure propensity of Ala and Val were different, the substitution of P4 residue might not favor flipping of cbz group into the cavity of Glu166 – Pro168, and reduced the potency.

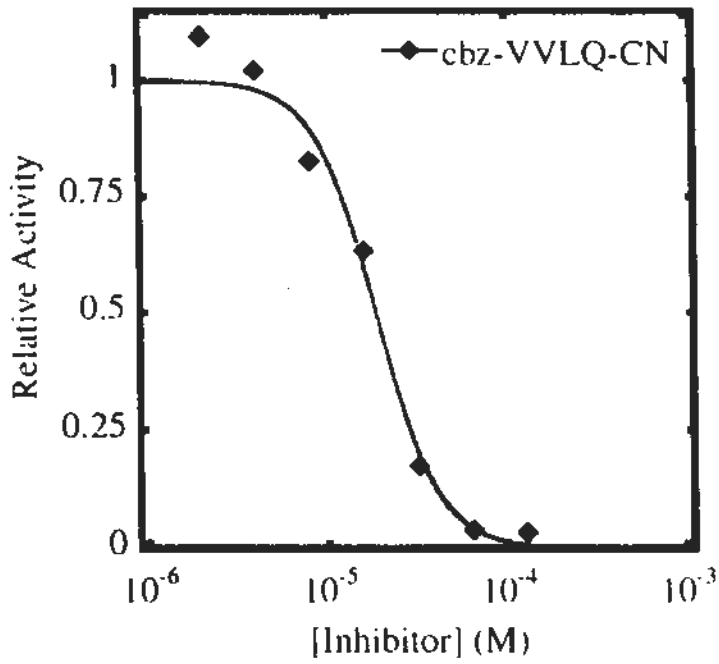


Figure 5.17 – Substitution to P4-Val had no improvement on IC_{50} value. 2 – 128 μM of cbz-VVLQ-CN was pre-mixed with 3 μM of SARS-CoV 3CL^{pro}, followed by determining relative activity by FRET assay. Curve was fitted by sigmoidal dose-response equation to obtain IC_{50} value, which was 18 ± 1 μM.

5.5.2 Substitution to Alkaline P3-Arg Converts Nitrile to Amide

Results of the super-active substrate sequences showed that SARS-CoV $3CL^{pro}$ favored 'VR' at P4 and P3 positions, thus 'cbz-VRLQ-CN' was synthesized and examined. A 648 Da non-target product was however present in the final product (Figure 5.18). We speculated that alkaline condition provided by Arg catalyzed conversion of nitrile group (−C≡N) to amide group (−CO−NH_2). Loss of nitrile group resulted in lowering potency. The IC_{50} value was higher than 100 μM.

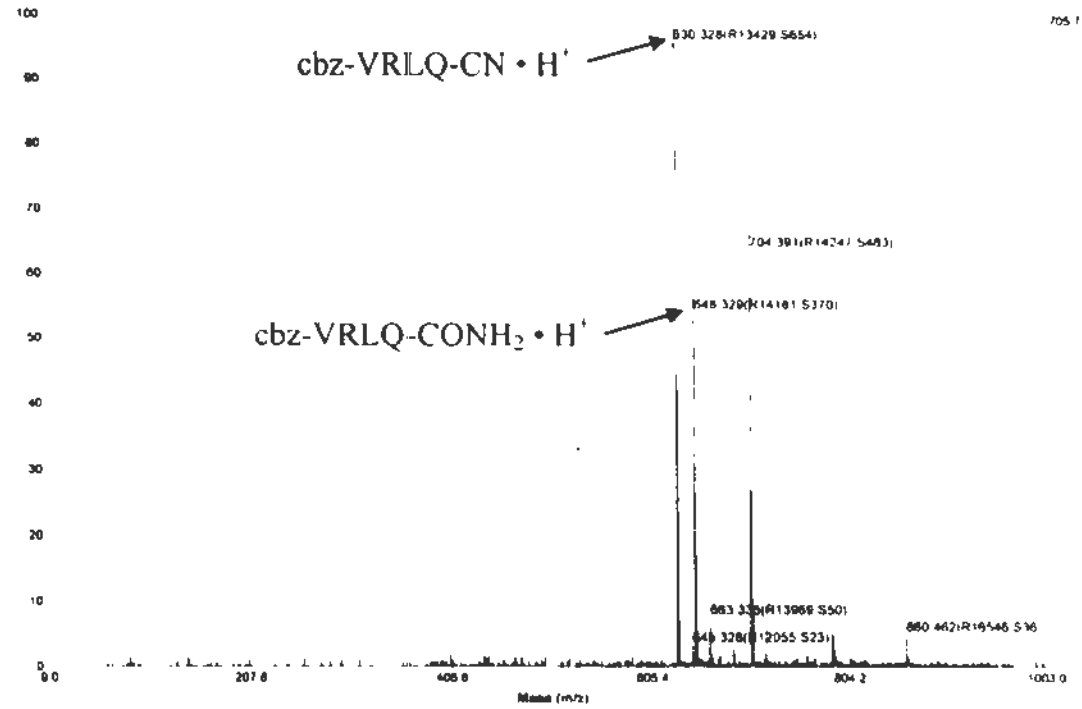


Figure 5.18 – Conversion of nitrile to amide under alkaline condition. In mass spectrometry fingerprint of 'cbz-VRLQ-CN', two of the major products with 630 Da and 648 Da were predicted to be cbz-VRLQ-CN•H$^+$ and cbz-VRLQ-CONH$_2$•H$^+$ respectively.

5.5.3 Substitution of Hexpeptide Causes Precipitation

Based on the super-active substrate sequences TVVLQ (1.87 ± 0.19) and VVVLQ (1.80 ± 0.17) without Arg, hexpeptidomimetic inhibitor was substituted to produce 'cbz-TTVVLQ-CN' and 'cbz-TVVVLQ-CN'. They were nevertheless insoluble in aqueous solution and no inhibitory effect was observed.

5.6 Summary

A series of nitrile-based peptidomimetic inhibitors with various protective groups, peptide sequences and peptide length were produced, followed by determining their IC_{50} and K_i values (Table 5.5). Most of their IC_{50} and K_i values were in micromolar range, indicating nitrile-based inhibitors are effective in $3CL^{pro}$ inhibition. The most potent inhibitor was 'cbz-AVLQ-CN', with IC_{50} and K_i values of 5.9 ± 0.6 μM and 0.62 ± 0.11 μM respectively. Crystal structures of $3CL^{pro}$-inhibitor complexes showed that nitrile group covalent bonded with Cys145, resulting in inhibition. Their P1 – P4 residues bound to substrate binding cleft of $3CL^{pro}$ as protease-substrate complex. Cbz group in 'cbz-AVLQ-CN' flipped into the cavity consisted of Glu166 – Pro168, enhancing binding affinity of the inhibitor towards $3CL^{pro}$.

Table 5.5 – Summary of IC_{50} and K_i values of nitrile-based peptidomimetic inhibitors. 'ND' stands for non-detectable.

Name of inhibitor	IC_{50} (μM)	K_i (μM)
cbz-AVLQ-CN	5.9 ± 0.6	0.62 ± 0.11
miu-AVLQ-CN	45 ± 3	7.0 ± 0.7
boc-AVLQ-CN	59 ± 5	11.4 ± 1.1
cbz-TSAVLQ-CN	7.7 ± 0.9	5.4 ± 0.5
cbz-VVLQ-CN	18 ± 1	/
cbz-VRLQ-CN	> 100	/
cbz-TTVVLQ-CN	ND	/
cbz-TVVVLQ-CN	ND	/

Chapter 6 – Broad Substrate Specificity of Various 3CL^{pro}

A number of research studies have focused on the development of 3CL^{pro} inhibitor in broad spectrum, but information about substrate specificity of various 3CL^{pro} is rare. As cleavage site sequences of all 3CL^{pro} at P2 – P1' positions are highly conserved, other 3CL^{pro} may be able to cleave autocleavage sequence of SARS-CoV 3CL^{pro}. Comprehensive substrate specificity of other 3CL^{pro} can hence be profiled by using the substrate library of SARS-CoV 3CL^{pro} with the same procedures. To determine the general specificity of all 3CL^{pro}, we selected one viral strain from each of the three groups as representative for profiling.

6.1 Broad Substrate Specificity Profiling

6.1.1 Specific Activities of Group Representatives

3CL^{pro} of HCoV-NL63 (group I), HCoV-OC43 (group IIa) and IBV (group III) were expressed and purified. Since they were only soluble in 20 mM Tris, pH 7.8, 150 mM of NaCl and 1 mM TCEP, which was diverse from reaction condition in chapter 4, SARS-CoV 3CL^{pro} (group IIb) was profiled in parallel to study effect of change in reaction condition on the substrate specificity. Proteolytic rates of these four 3CL^{pro} were examined by FRET assay using the protein-based WT substrate consisted of autocleavage sequence of SARS-CoV 3CL^{pro}. All proteases could cleave the autocleavage sequence, showing that the library could be applied for the substrate specificity profiling. 3CL^{pro} specific activities of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV were 443 ± 11, 124 ± 13, 180 ± 5 and

174 ± 19 mM^{-1} min^{-1} respectively (Table 6.1). The activities of group II and III members were comparable, while that of group I member was higher than others.

When the reaction condition was changed from 20 mM of NaCl to 150 mM of NaCl and 1 mM of TCEP, SARS-CoV 3CL^{pro} activity was elevated from 71 ± 11 to 180 ± 5 mM^{-1} min^{-1}. Presence of TCEP has been proven to increase the protease activity in section 5.1. The activity was also raised by increase in NaCl concentration until 200 mM (Figure 6.1).

Table 6.1 – Specific activities of group representatives against WT substrate of SARS-CoV 3CL^{pro}.

CoV	Group	Specific activity, $k_{obs}/[3CL^{pro}]$ (mM^{-1} min^{-1})
HCoV-NL63	I	443 ± 11
HCoV-OC43	IIa	124 ± 13
SARS-CoV	IIb	180 ± 5
IBV	III	174 ± 19

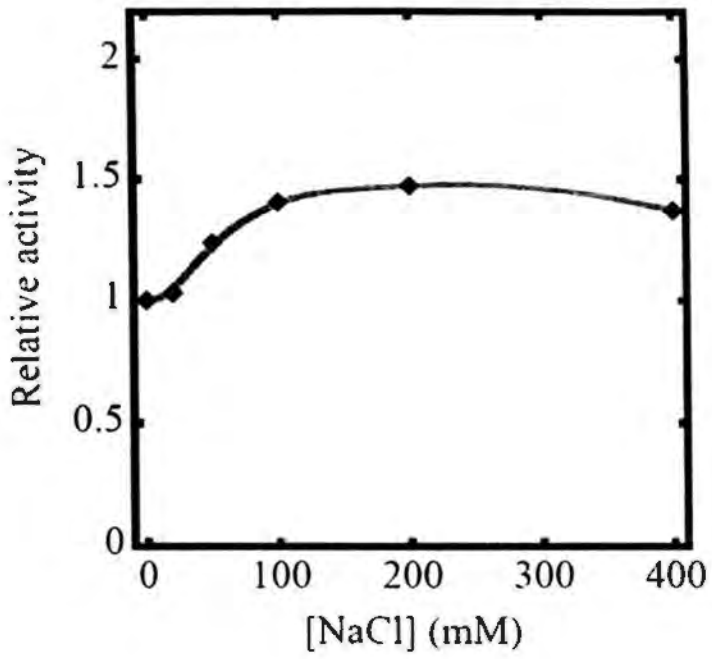


Figure 6.1 – Increase in NaCl concentration until 200 mM elevated SARS-CoV 3CL^{pro} activity. The protease activity was examined under 0 – 400 mM of NaCl. Specific activities under various NaCl concentrations were normalized by that in absence of NaCl to obtain relative activity. The relative activity was elevated when NaCl concentration was increased until 200 mM.

6.1.2 Determining 3CL^{pro} Activities of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV against the Substrate Library

To profile their substrate specificity, 3CL^{pro} activities of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV against each of the 19×8 substrate variants were measured (Figure 6.2 and appendix 6.1 – 6.8). Substrate specificity spectra showed that selectivity at each of the P5 to P3' positions among various 3CL^{pro} were basically comparable. P1 position was the most selective, as it only accommodated Gln, His and Met. P5, P3, and P3' positions are less selective. All variants except V3P were cleavable. Positively charged residues (Arg and Lys) at P3 position was more favored than negatively charged ones (Asp and Glu). P5 and P3' positions also tended to have same preference on charge but less obvious.

Activities of these 3CL^{pro} were obtained by a simplified FRET assay using 3CL^{pro} at one particular concentration instead of a range for workload reduction. To examine whether change in reaction condition influence results of substrate specificity, 3CL^{pro} activity of SARS-CoV reported in figure 6.2 were compared to that stated in figure 4.2. Comparison indicated that their activities against same variants were generally comparable (Figure 6.3), demonstrating that the substrate specificity were independent of concentrations of protease, NaCl and TCEP.

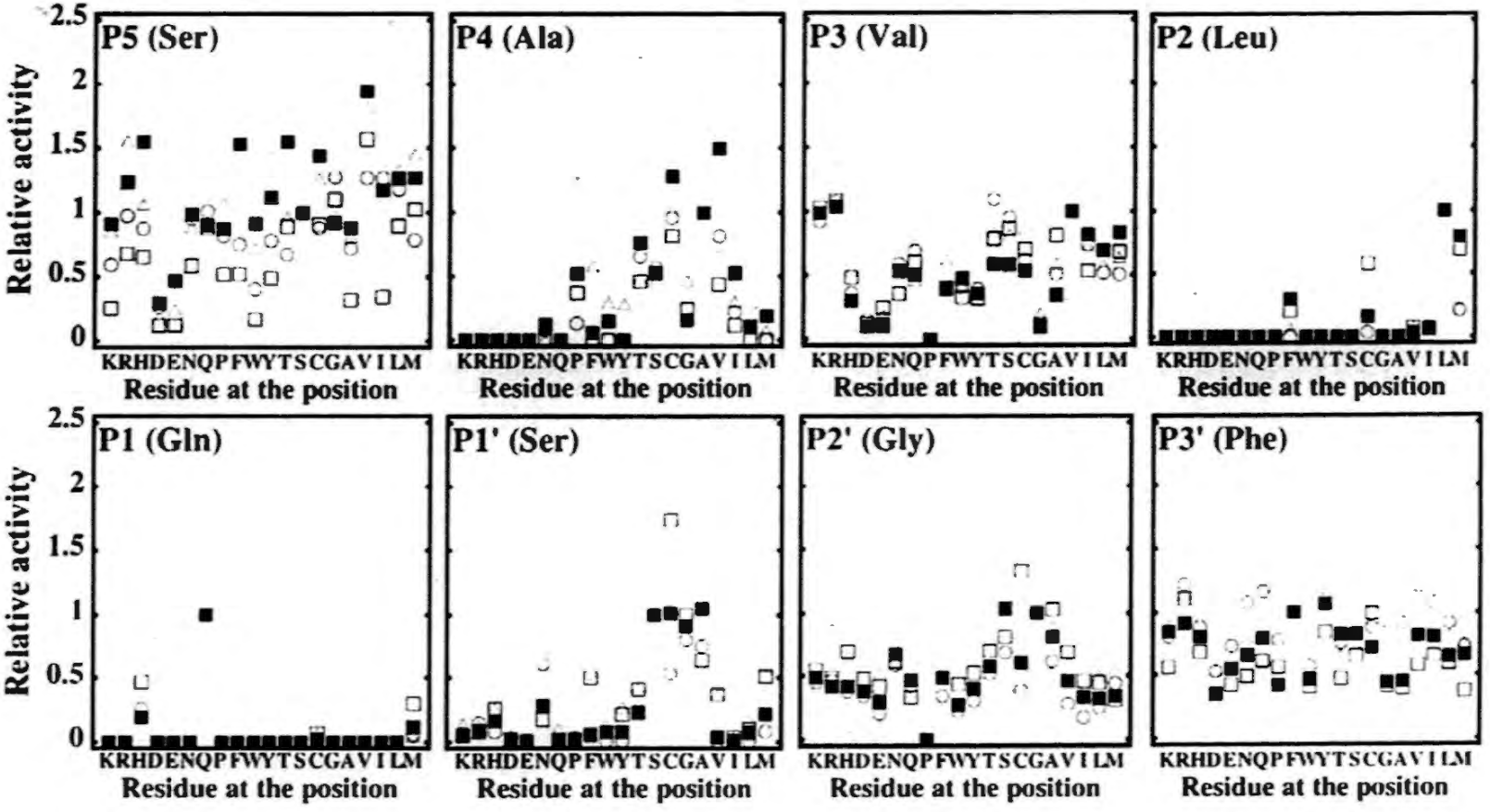


Figure 6.2 – Substrate specificity of 3CL^{pro} of group representatives. Relative activities of 3CL^{pro} of HCoV-NL63 (○), HCoV-OC43 (□), SARS-CoV (■) and IBV (△) against 19×8 variants were determined by simplified FRET assay.

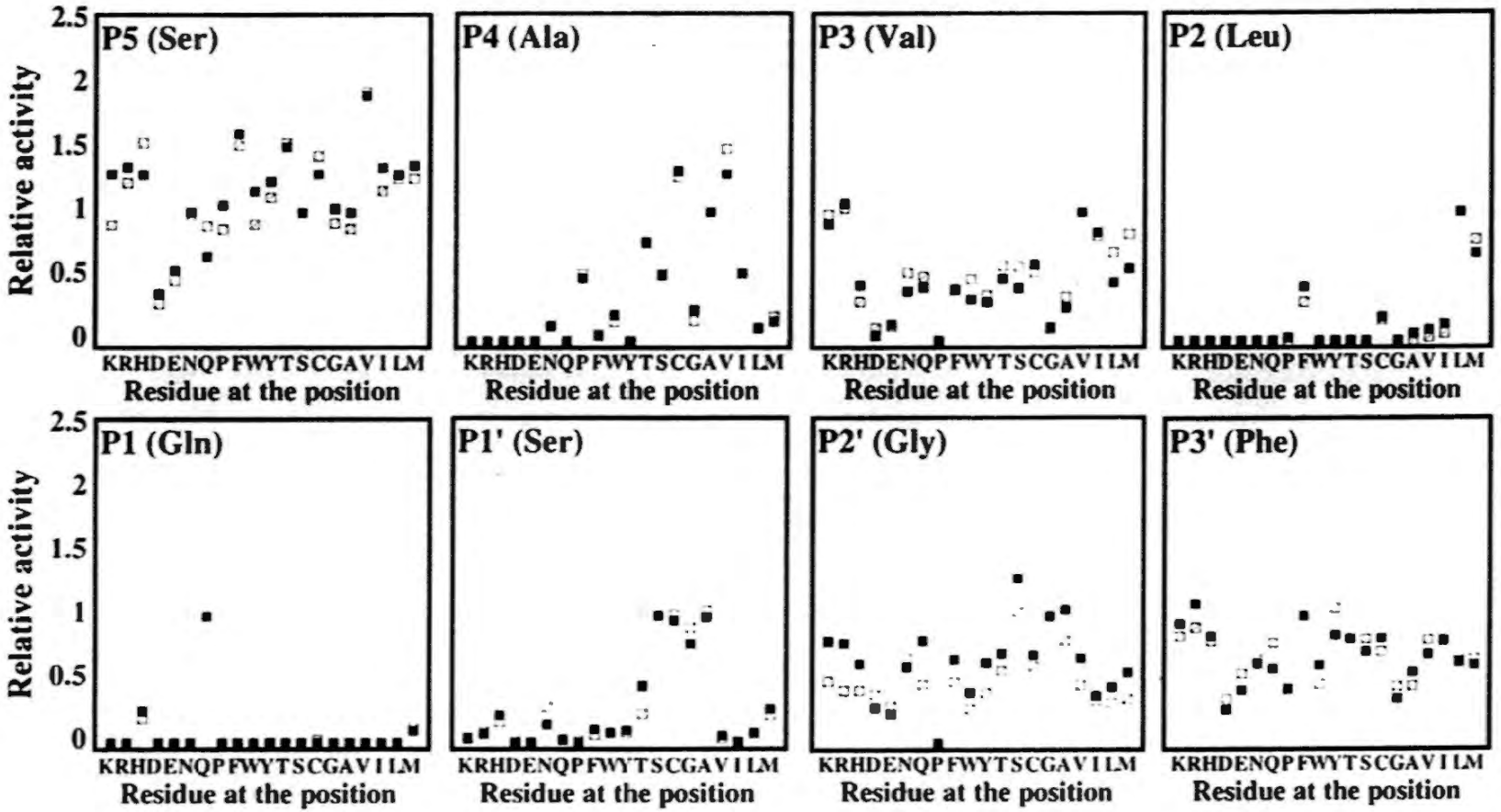


Figure 6.3 – Effects of change in reaction condition on SARS-CoV 3CL^{pro} activity. Relative activities of SARS-CoV 3CL^{pro} against 19×8 variants were determined under 20 mM Tris, pH 7.8 and 20 mM NaCl by using 1 – 4 μM of protease (▨), and 20 mM Tris, pH 7.8, 150 mM NaCl and 1 mM TCEP (■) by using 2 μM of protease.

6.2 Correlation between Activity and Structural Properties of Substituting Residues

3CL^{pro} activities of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV were correlated with structural properties of substituting residues to further find out the substrate specificity (Appendix 2.12) (Chou & Fasman, 1978; Kyte & Doolittle, 1982; Lee *et al.*, 2008). The correlation coefficients (r) and p-values were shown in table 6.2. Significant correlations with p-value < 0.01 were found in some of the cases.

Correlations of SARS-CoV 3CL^{pro} determined under different reaction conditions were compared (Table 6.3). Both of the protease activities showed correlations with β-sheet propensity at P5 position, hydrophobicity at P4 position and side chain volume at P1' position. However, by comparing to correlation reported in Chapter 4, correlation coefficients of hydrophobicity at P5 and P2 positions were slightly decreased from 0.573 (p = 0.008) and 0.590 (p = 0.006) to 0.539 (p = 0.014) and 0.511 (p = 0.021) respectively. Meanwhile, correlation with side chain volume at P2' position (r = -0.628, p = 0.003) was improved after change in reaction condition. Results in chapter 4 showed that SARS-CoV 3CL^{pro} tended to prefer small P2' residues but no significant correlation was found (r = -0.363, p = 0.116). In summary, correlations were slightly varied by change in reaction condition but no contradicting result was observed.

Table 6.2 – Correlation between various 3CL^{pro} activities and structural properties of substituting residues. 3CL^{pro} relative activities of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV were correlated with scales for side chain volume, hydrophobicity, and α-helix and β-sheet propensities of substituting residues (Chou & Fasman, 1978; Kyte & Doolittle, 1982; Lee *et al.*, 2008). The correlation coefficients and p-values (in parenthesis) were reported. Significant correlations with p-value < 0.01 are bolded and marked with asterisks. 'Volume', 'hydrophobic', 'α-helix' and 'β-sheet' stand for side chain volume, hydrophobicity and α-helix and β-sheet propensities respectively.

Position	Property	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
P5	Volume	-0.229 (0.332)	-0.313 (0.180)	0.192 (0.416)	0.313 (0.179)
	Hydrophobic	0.466 (0.039)	0.397 (0.083)	0.539 (0.014)	**0.646 (0.002)***
	α-helix	-0.283 (0.227)	-0.257 (0.275)	-0.107 (0.653)	-0.089 (0.708)
	β-sheet	0.423 (0.063)	0.362 (0.177)	**0.704 (0.001)***	**0.759 (<0.001)***
P4	Volume	-0.529 (0.017)	**-0.618 (0.004)***	-0.397 (0.083)	-0.369 (0.110)
	Hydrophobic	0.492 (0.027)	0.387 (0.092)	**0.594 (0.006)***	0.499 (0.025)
	α-helix	-0.089 (0.709)	-0.154 (0.518)	-0.118 (0.621)	-0.320 (0.170)
	β-sheet	0.198 (0.402)	-0.011 (0.962)	0.335 (0.148)	0.109 (0.647)
P3	Volume	0.079 (0.742)	0.090 (0.706)	0.364 (0.114)	0.219 (0.354)
	Hydrophobic	0.099 (0.677)	0.180 (0.448)	0.257 (0.273)	0.046 (0.847)
	α-helix	0.045 (0.851)	0.295 (0.207)	0.245 (0.298)	0.144 (0.544)
	β-sheet	0.402 (0.079)	0.303 (0.193)	0.519 (0.019)	0.263 (0.262)
P2	Volume	0.175 (0.461)	0.146 (0.539)	0.242 (0.303)	0.165 (0.487)
	Hydrophobic	0.416 (0.068)	**0.561 (0.010)***	0.511 (0.021)	0.490 (0.028)
	α-helix	0.251 (0.286)	0.253 (0.283)	0.378 (0.101)	0.231 (0.328)
	β-sheet	0.209 (0.377)	0.272 (0.246)	0.242 (0.304)	0.266 (0.257)
P1	Volume	0.030 (0.902)	0.059 (0.805)	0.033 (0.891)	0.020 (0.933)
	Hydrophobic	-0.272 (0.246)	-0.240 (0.307)	-0.250 (0.288)	-0.264 (0.262)
	α-helix	0.106 (0.656)	0.171 (0.472)	0.130 (0.585)	0.092 (0.701)
	β-sheet	0.022 (0.926)	0.009 (0.971)	-0.030 (0.901)	0.027 (0.910)
P1'	Volume	**-0.741 (<0.001)***	-0.510 (0.021)	**-0.685 (0.001)***	**-0.628 (0.003)***
	Hydrophobic	0.021 (0.930)	0.375 (0.103)	0.200 (0.398)	0.293 (0.210)
	α-helix	-0.365 (0.114)	-0.315 (0.176)	-0.252 (0.284)	-0.306 (0.189)
	β-sheet	-0.288 (0.218)	0.051 (0.830)	-0.192 (0.418)	-0.042 (0.862)
P2'	Volume	**-0.585 (0.007)***	-0.511 (0.021)	**-0.628 (0.003)***	-0.364 (0.115)
	Hydrophobic	-0.138 (0.562)	0.239 (0.310)	0.039 (0.870)	0.007 (0.977)
	α-helix	-0.225 (0.341)	-0.210 (0.373)	0.242 (0.305)	-0.156 (0.511)
	β-sheet	-0.227 (0.337)	0.068 (0.777)	-0.115 (0.630)	0.134 (0.574)
P3'	Volume	-0.069 (0.772)	0.303 (0.195)	0.431 (0.058)	0.267 (0.255)
	Hydrophobic	0.144 (0.544)	0.079 (0.740)	0.087 (0.715)	-0.001(0.997)
	α-helix	-0.229 (0.332)	-0.260 (0.269)	-0.070 (0.769)	-0.416 (0.068)
	β-sheet	0.371 (0.107)	0.344 (0.137)	0.537 (0.015)	0.199 (0.401)

Table 6.3 – Effects of change in reaction condition on correlation. The relative activity was correlated with scales for side chain volume, hydrophobicity, and α-helix and β-sheet propensities of substituting residues (Chou & Fasman, 1978; Kyte & Doolittle, 1982; Lee *et al.*, 2008). The correlation coefficients and p-values (in parenthesis) were reported. Significant correlations with p-value < 0.01 are bolded and marked with asterisks. 'Volume', 'hydrophobic', 'α-helix' and 'β-sheet' stand for side chain volume, hydrophobicity and α-helix and β-sheet propensities respectively.

Position	Property	SARS-CoV (20mM Tris, pH7.8, 20mM NaCl)	SARS-CoV (20mM Tris, pH7.8 150mM NaCl, 1mM TCEP)
P5	Volume	0.331 (0.154)	0.192 (0.416)
	Hydrophobic	**0.573 (0.008)***	0.539 (0.014)
	α-helix	-0.064 (0.789)	-0.107 (0.653)
	β-sheet	**0.711 (<0.001)***	**0.704 (0.001)***
P4	Volume	-0.424 (0.063)	-0.397 (0.083)
	Hydrophobic	**0.587 (0.006)***	**0.594 (0.006)***
	α-helix	-0.147 (0.536)	-0.118 (0.621)
	β-sheet	0.315 (0.176)	0.335 (0.148)
P3	Volume	0.338 (0.144)	0.364 (0.114)
	Hydrophobic	0.221 (0.349)	0.257 (0.273)
	α-helix	0.170 (0.473)	0.245 (0.298)
	β-sheet	0.510 (0.022)	0.519 (0.019)
P2	Volume	0.255 (0.277)	0.242 (0.303)
	Hydrophobic	**0.590 (0.006)***	0.511 (0.021)
	α-helix	0.379 (0.100)	0.378 (0.101)
	β-sheet	0.304 (0.192)	0.242 (0.304)
P1	Volume	0.038 (0.873)	0.033 (0.891)
	Hydrophobic	-0.269 (0.252)	-0.250 (0.288)
	α-helix	0.126 (0.595)	0.130 (0.585)
	β-sheet	0.021 (0.931)	-0.030 (0.901)
P1'	Volume	**-0.660 (0.002)***	**-0.685 (0.001)***
	Hydrophobic	0.233 (0.323)	0.200 (0.398)
	α-helix	-0.222 (0.347)	-0.252 (0.284)
	β-sheet	-0.143 (0.548)	-0.192 (0.418)
P2'	Volume	-0.363 (0.116)	**-0.628 (0.003)***
	Hydrophobic	0.022 (0.926)	0.039 (0.870)
	α-helix	-0.097 (0.685)	0.242 (0.305)
	β-sheet	0.048 (0.841)	-0.115 (0.630)
P3'	Volume	0.496 (0.026)	0.431 (0.058)
	Hydrophobic	0.094 (0.695)	0.087 (0.715)
	α-helix	-0.017 (0.944)	-0.070 (0.769)
	β-sheet	0.486 (0.030)	0.537 (0.015)

6.3 Substrate Specificity of Various 3CL^{pro}

6.3.1 3CL^{pro} of SARS-CoV and IBV Favor P5 Residues with High β-sheet Propensity

All P5 variants were cleavable by these 3CL^{pro} (Figure 6.2). Most of the substitutions caused increase in the protease activities of SARS-CoV and IBV, but decrease in that of HCoV-NL63 and HCoV-OC43. The universal preferred P5 residue was Val (HCoV-NL63: 1.27 ± 0.02; HCoV-OC43: 1.56 ± 0.10; SARS-CoV: 1.92 ± 0.07 and IBV: 1.81 ± 0.11). In addition, positively charged variants were consistently 2 – 6 fold higher in relative activity than negatively charged ones. Significant correlations were found between β-sheet propensity and 3CL^{pro} activities of SARS-CoV ($r = 0.704$, $p = 0.001$) and IBV ($r = 0.759$, $p < 0.001$), but not HCoV-NL63 ($r = 0.423$, $p = 0.063$) and HCoV-OC43 ($r = 0.362$, $p = 0.177$) (Table 6.2 and figure 6.4). Another correlation was specifically found between hydrophobicity and the activity of IBV 3CL^{pro} ($r = 0.646$, $p = 0.002$), but its coefficient was lower than that of β-sheet propensity (Figure 6.5).

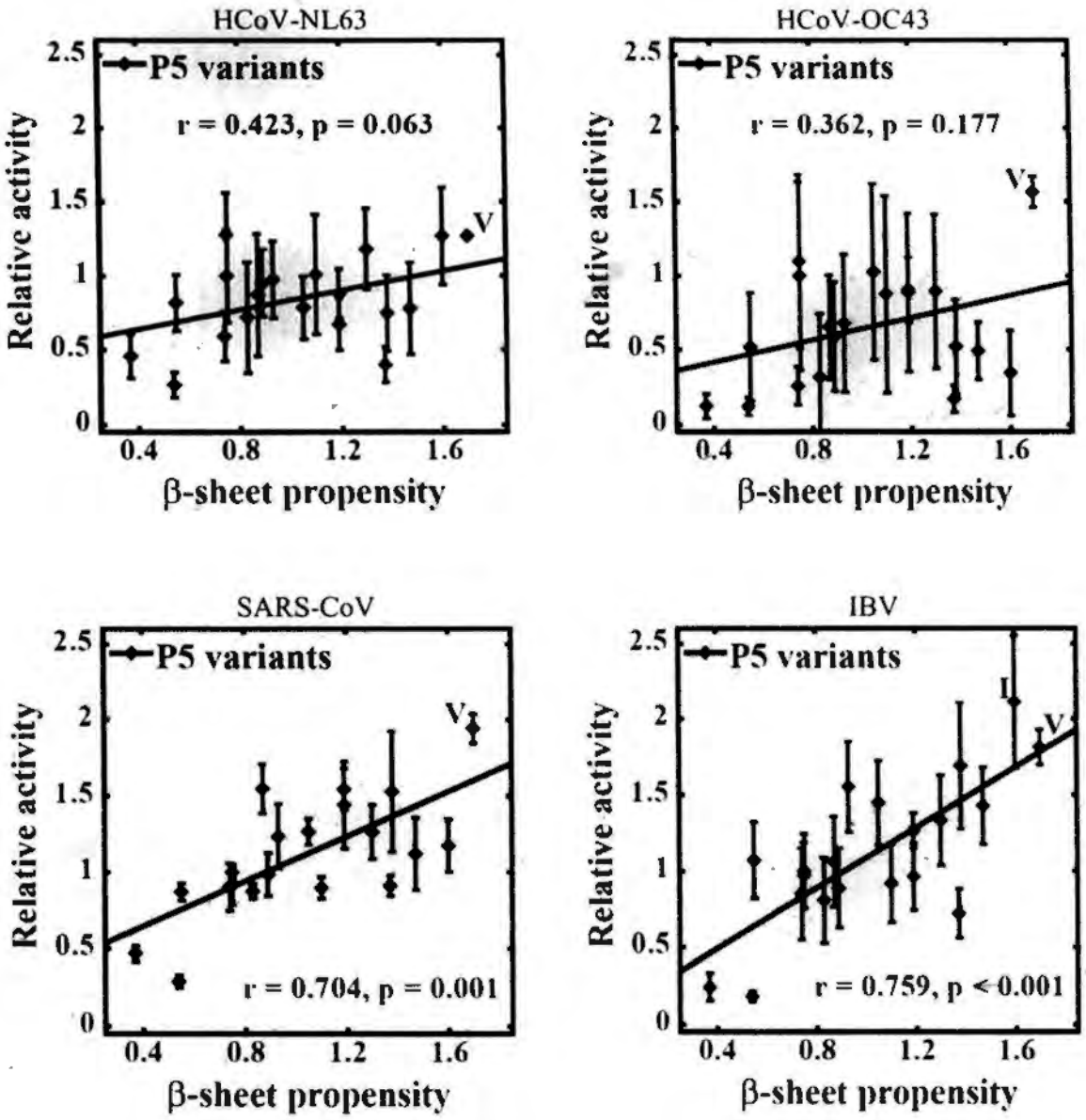


Figure 6.4 – Only 3CL^{pro} of SARS-CoV and IBV preferred P5 residues with high β-Sheet propensity. Correlation was found between β-Sheet propensity at P5 position and the relative activity of 3CL^{pro} of SARS-CoV ($r = 0.704$, $p = 0.001$) and IBV ($r = 0.759$, $p < 0.001$), but not HCoV-NL63 ($r = 0.423$, $p = 0.063$) and HCoV-OC43 ($r = 0.362$, $p = 0.177$).

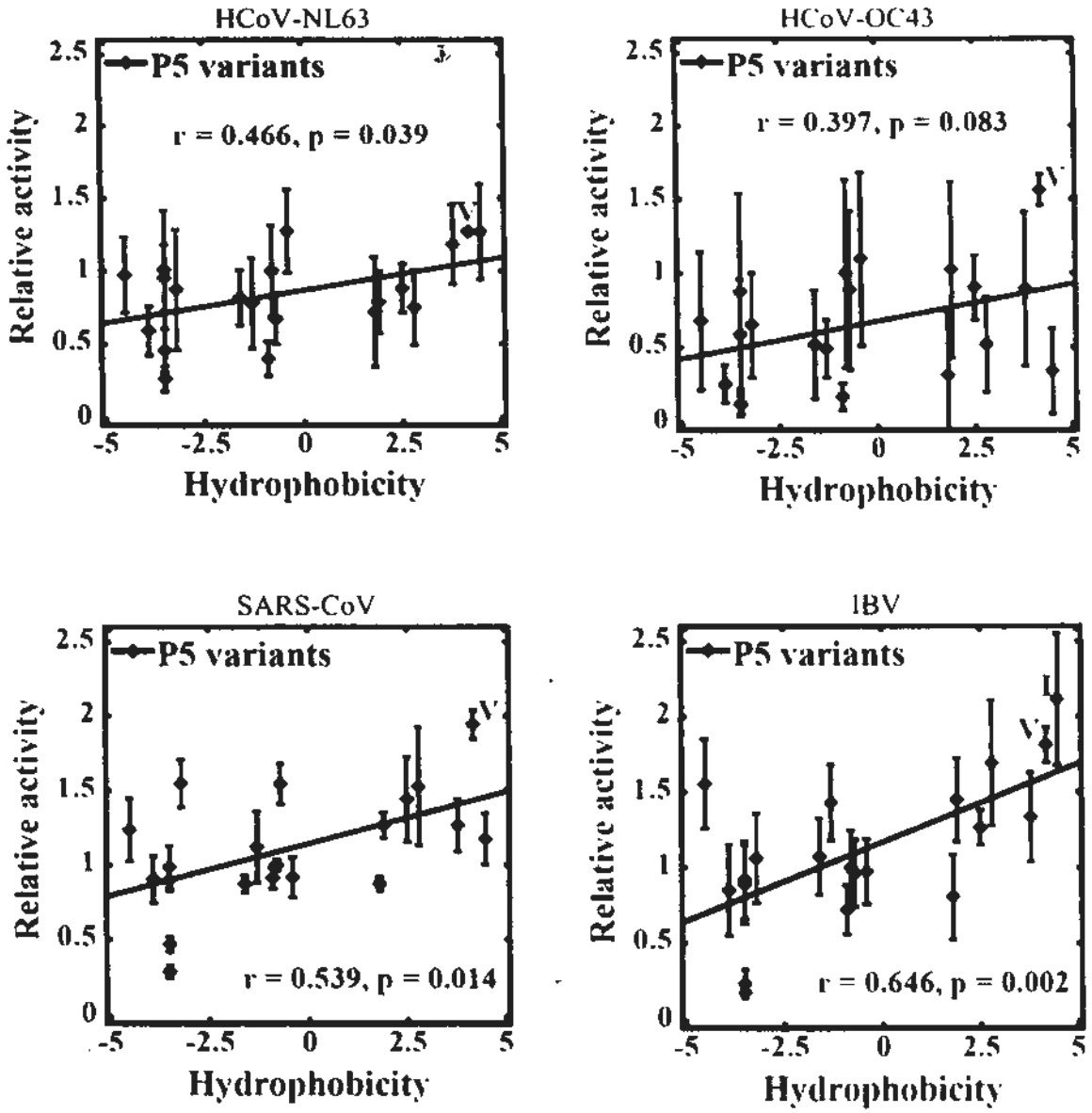


Figure 6.5 – Only IBV 3CL^{pro} preferred P5 residues with high hydrophobicity. Only the relative activity of IBV 3CL^{pro} (r = 0.646, p = 0.002) against P5 variants was significantly correlated to hydrophobicity.

6.3.2 *3CL^{pro} Universally Favor Hydrophobic P4 Residues with Side Chain Volume of < 70 Å^3*

SARS-CoV 3CL^{pro} has been proven to favor hydrophobic P4 residues with side chain volume of < 80 Å^3. Nevertheless, the correlations were lower when other 3CL^{pro} were analyzed (Figure 6.6). We found that only SARS-CoV 3CL^{pro} favored A4V (1.49 ± 0.31). 3CL^{pro} cleavage rates of HCoV-NL63 (0.81 ± 0.05), HCoV-OC43 (0.44 ± 0.04) and IBV (0.67 ± 0.16) were obviously lower (Figure 6.2). Their commonly preferred residues were Ala (HCoV-NL63: 1.00 ± 0.25; HCoV-OC43: 1.00 ± 0.23, SARS-CoV: 1.00 ± 0.20 and IBV: 1.00 ± 0.13) and Cys (HCoV-NL63: 0.95 ± 0.37; HCoV-OC43: 0.81 ± 0.33; SARS-CoV: 1.27 ± 0.27 and IBV: 1.01 ± 0.19), which were smaller than Val, suggesting that S4 pockets of 3CL^{pro} of other CoVs were shallower than that of SARS-CoV. When relative activities of other 3CL^{pro} against P4 variants with side chain volume of < 70 Å^3 were analyzed, correlations with hydrophobicity were increased.

3CL^{pro} activity of IBV against A4P (1.23 ± 0.26) was apparently deviated from the trend line and higher than that of the other 3CL^{pro} (HCoV-NL63: 0.13 ± 0.05; HCoV-OC43: 0.37 ± 0.09 and SARS-CoV: 0.52 ± 0.13), indicating that IBV 3CL^{pro} prefers the exceptional dihedral angle of Pro at P4 position. After neglecting this exceptional data, the activity of IBV 3CL^{pro} was highly correlated to hydrophobicity ($r = 0.994$, $p < 0.001$).

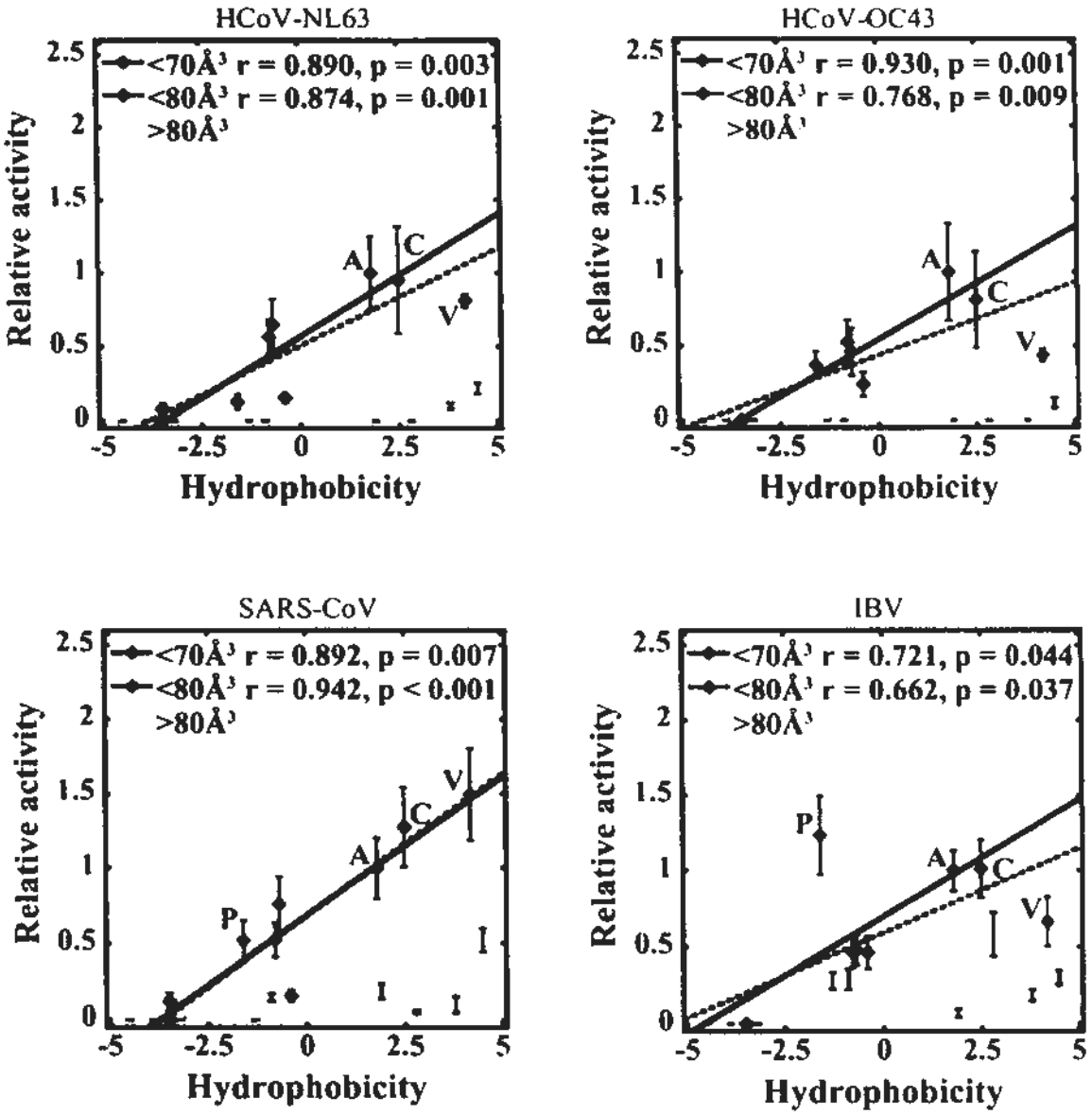


Figure 6.6 – 3CL^{pro} universally preferred hydrophobic P4 residues with side chain volume of < 70 Å^3. Correlations with hydrophobicity were improved when relative activities of 3CL^{pro} of HCoV-NL63, HCoV-OC43 and IBV against P4 variants with side chain volume of < 70 Å^3 instead of < 80 Å^3 were analyzed.

6.3.3 Only SARS-CoV 3CL^{pro} Favors P3 Residues with High β-Sheet Propensity

All examined 3CL^{pro} highly preferred positively charged P3 residues, including Arg (HCoV-NL63: 1.07 ± 0.28; HCoV-OC43: 1.09 ± 0.28; SARS-CoV: 1.03 ± 0.44; IBV: 1.34 ± 0.34) and Lys (HCoV-NL63: 0.92 ± 0.07; HCoV-OC43: 1.03 ± 0.13; SARS-CoV: 0.98 ± 0.20; IBV: 1.69 ± 0.10) (Figure 6.2). Glu166 was absolutely conserved among all 3CL^{pro}, further supporting the hypothesis of ionic interaction between Glu166 and positively charged P3 residue.

SARS-CoV 3CL^{pro} preferred Val (1.00 ± 0.19) because of its high β-sheet propensity (Figure 6.7). A strong correlation was observed between the relative activity of SARS-CoV 3CL^{pro} against neutral P3 variants (20 substrates except V3D, V3E, V3K and V3R) and β-sheet propensity ($r = 0.665$, $p = 0.005$). Yet 3CL^{pro} of other CoVs favored Val as well (HCoV-NL63: 1.00 ± 0.20; HCoV-OC43: 1.00 ± 0.24 and IBV: 1.00 ± 0.10), their correlations with β-sheet propensity was weaker than that of SARS-CoV. Why P3-Val is universally preferred is still under investigation.

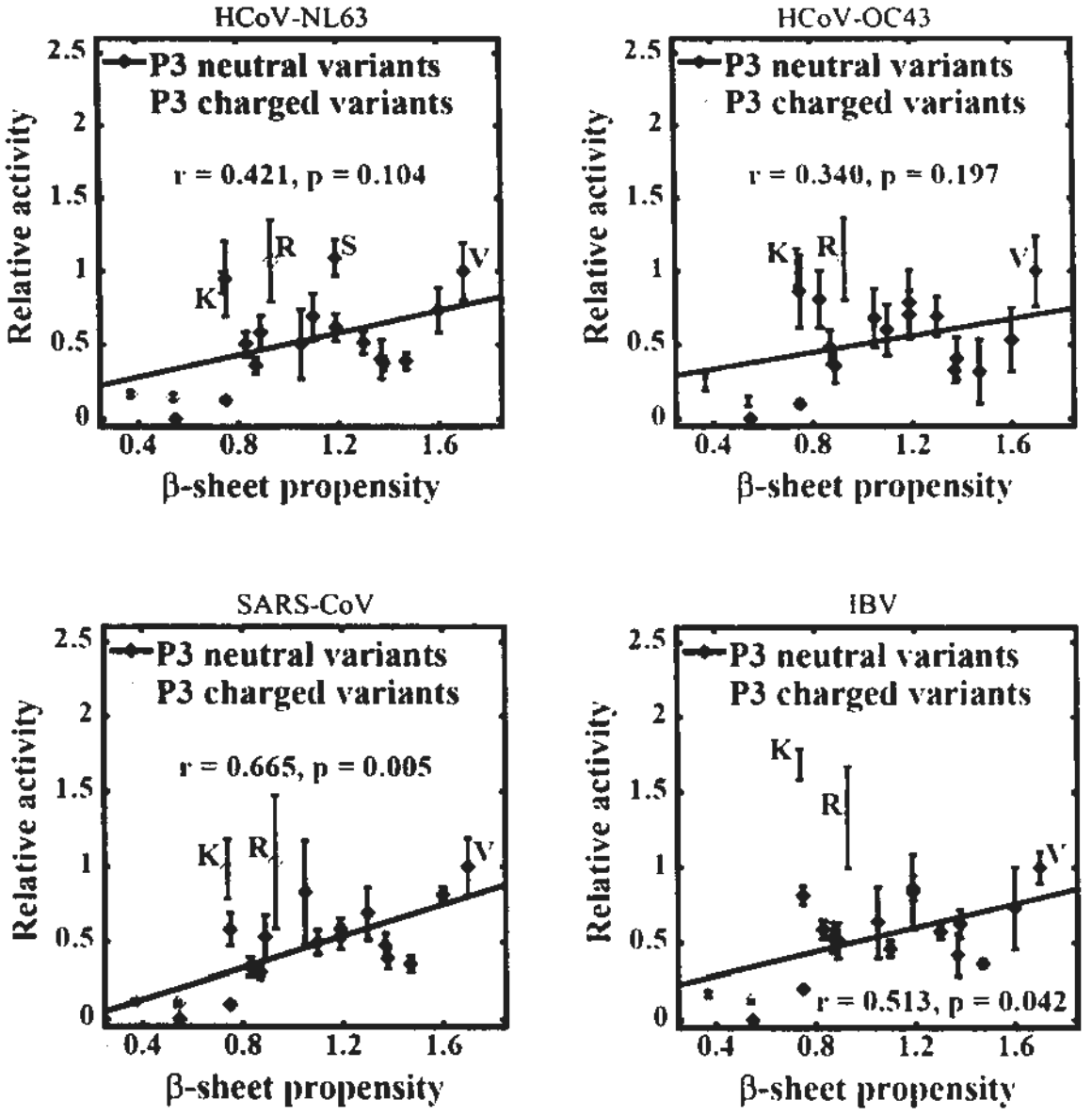


Figure 6.7 – Only SARS-CoV 3CL^{pro} preferred P3 residues with high β-sheet propensity. The relative activity of SARS-CoV 3CL^{pro} (r = 0.665, p = 0.005) against P3 neutral variants (20 substrates except V3D, V3E, V3K, V3R) was significantly correlated with β-sheet propensity.

6.3.4 All 3CL^{pro} Favor Hydrophobic P2 Residues without β-Branch

Substrate specificities of various 3CL^{pro} were alike at P2 position. Only variants with hydrophobic P2 residues could be cleaved (Figures 6.2 and 6.8). Their universal favorite residue was Leu (HCoV-NL63: 1.00 ± 0.28; HCoV-OC43: 1.00 ± 0.40; SARS-CoV: 1.00 ± 0.28 and IBV: 1.00 ± 0.35), followed by Met (HCoV-NL63: 0.21 ± 0.04; HCoV-OC43: 0.69 ± 0.34; SARS-CoV: 0.79 ± 0.34 and IBV: 0.22 ± 0.10). β-branched residues Val (HCoV-NL63: 0.00 ± 0.00; HCoV-OC43: 0.08 ± 0.03; SARS-CoV: 0.03 ± 0.01 and IBV: 0.07 ± 0.02) and Ile (HCoV-NL63: 0.08 ± 0.02; HCoV-OC43: 0.06 ± 0.02; SARS-CoV: 0.06 ± 0.02 and IBV: 0.08 ± 0.03) were structurally similar to Leu but consistently repelled, revealing that all 3CL^{pro} prefer hydrophobic residues without β-branch.

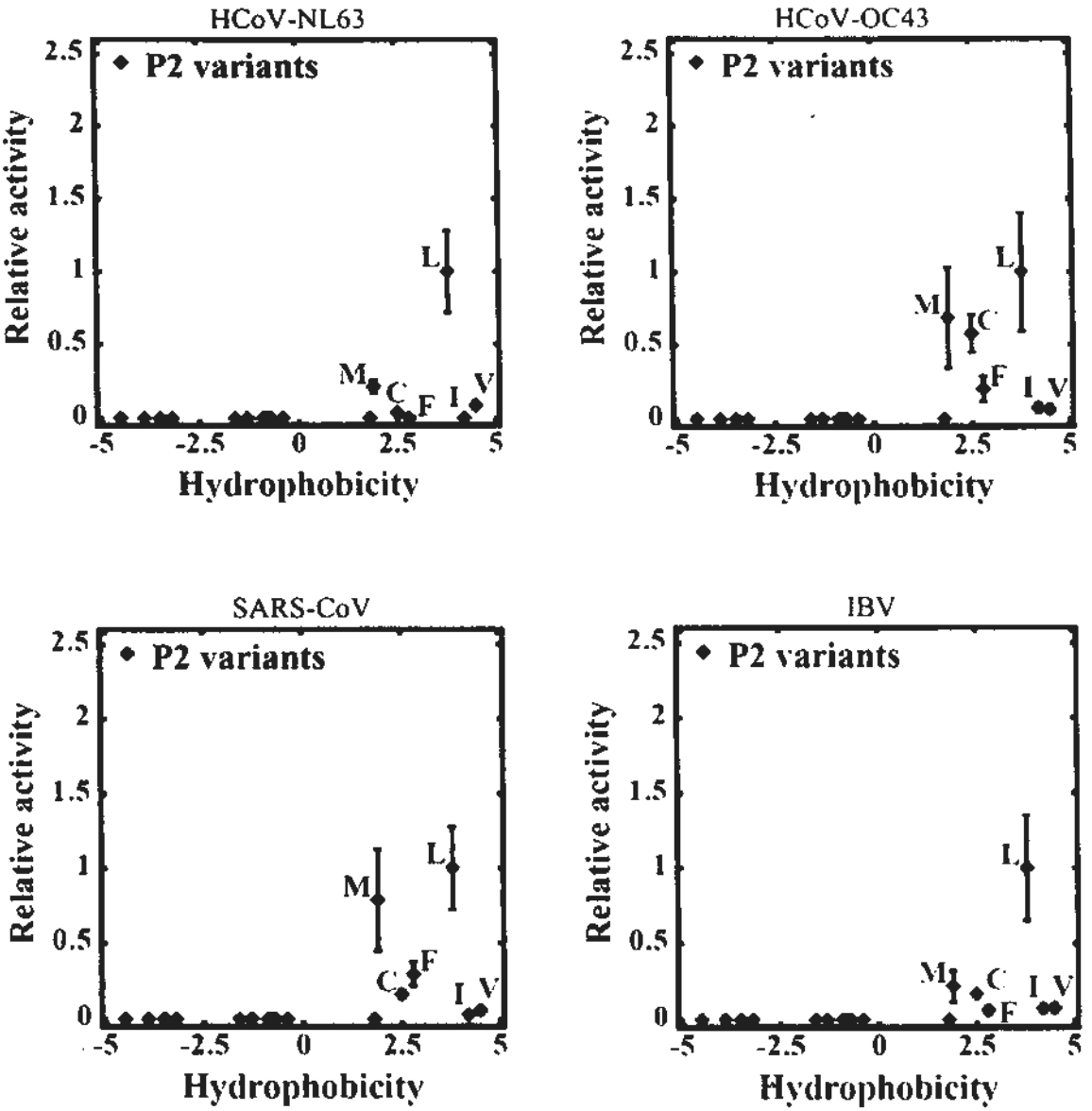


Figure 6.8 – All 3CL^{pro} preferred hydrophobic P2 residues without β-branch. All cleavable variants consisted of hydrophobic P2 residues. Their activities were the highest when Leu, followed by Met, was present at P2 position. β-branched residues such as Val and Ile were consistently repelled.

6.3.5 *All 3CL^{pro} Tolerate His and Met at P1 Position*

All 3CL^{pro} highly favored P1-Gln (Figure 6.2). P1 variants Q1H and Q1M could also be consistently cleaved in low proteolytic rates. Besides, Q1C tended to be cleavable but the rate was nearly undetectable. No cleavage was detected in other substitutions. In Figure 4.8 (P.78), model of P1 substitution indicated that His fits into S1 subsite and hydrogen-bonds with Asn142. Nevertheless, Asn142 was not highly conserved among various 3CL^{pro}. Ala and Cys could be found at this position, implying that hydrogen-bond formation between 3CL^{pro} and P1 residue is unnecessary.

6.3.6 *All 3CL^{pro} Favor Small P1' Residues*

High selectivity at P1' position was observed among all 3CL^{pro} (Figure 6.2). Negative correlations were found between P1' side chain volumes and 3CL^{pro} activities of HCoV-NL63 ($r = -0.741$, $p < 0.001$), SARS-CoV ($r = -0.685$, $p = 0.001$) and IBV ($r = -0.628$, $p = 0.003$) (Table 6.2). Small P1' residues, such as Gly, Ala and Ser, were universally preferred, while substitutions to residues larger than Cys resulted in dramatic decreases in activity (Figure 6.9). Collectively, all 3CL^{pro} prefers small P1' residues with side chain volume of < 50 Å^3.

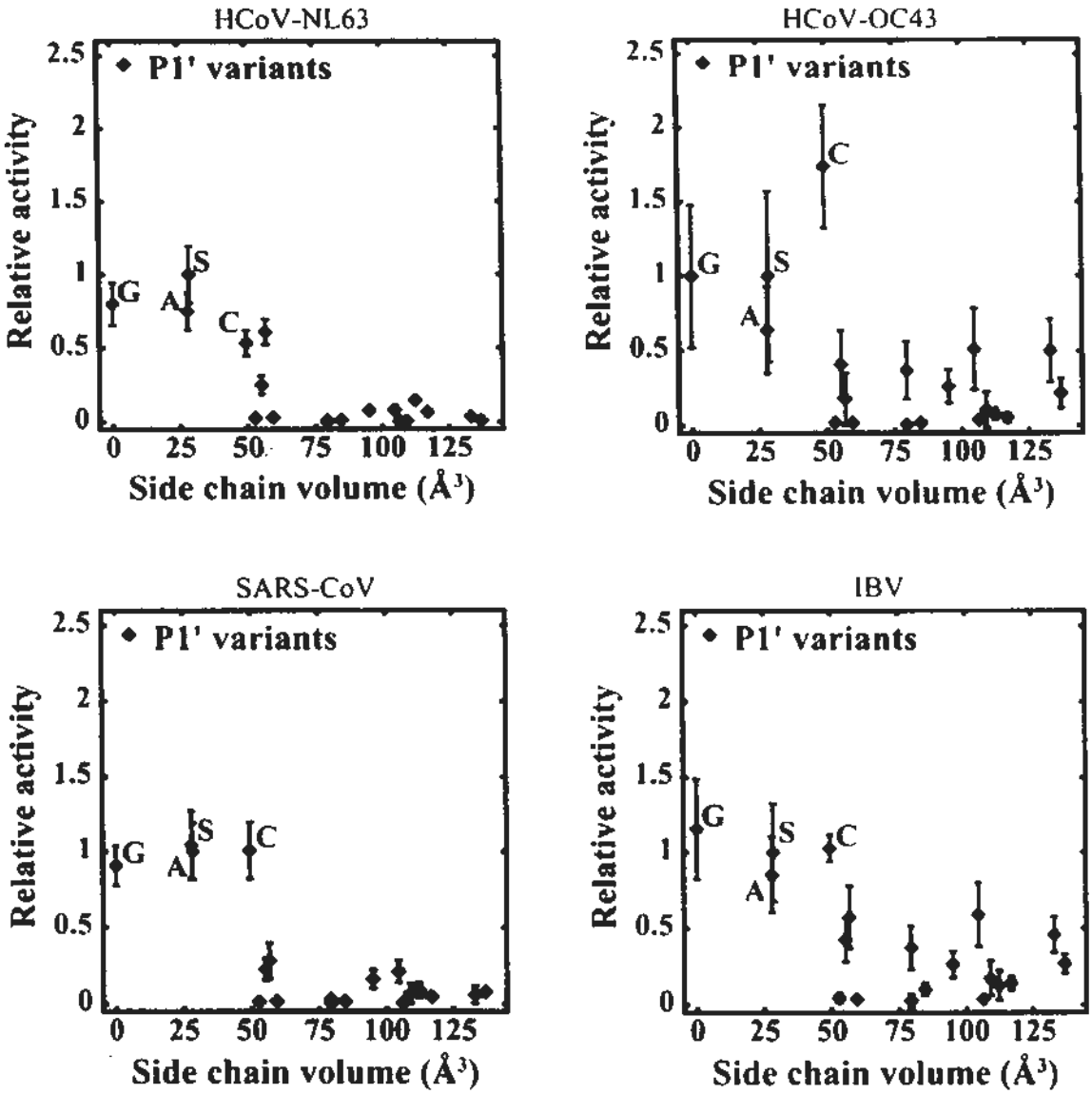


Figure 6.9 – All 3CL^{pro} preferred small P1' residues. The relative activities of all 3CL^{pro} against small P1' variants such as Gly, Ala and Ser were higher than those against large variants. Their activities were dramatically decreased when P1' side chain volume was > 50 Å^3.

6.3.7 All 3CL^{pro} Tend to Favor Small P2' Residues

Most of P2' variants were consistently cleavable by various 3CL^{pro} in low proteolytic rates (Figure 6.2). Only G2'P was found to be non-cleavable. Strong correlations were observed between side chain volume and 3CL^{pro} activities of HCoV-NL63 ($r = -0.585$, $p = 0.007$) and SARS-CoV ($r = -0.628$, $p = 0.003$), suggesting that the activities related to sizes of P2 residues. Plots of the activities against the volume demonstrated that 3CL^{pro} tend to prefer small residues in general, such as Gly, Ala and Ser (Figure 6.10). Large P2' residues were still cleavable in low rates.

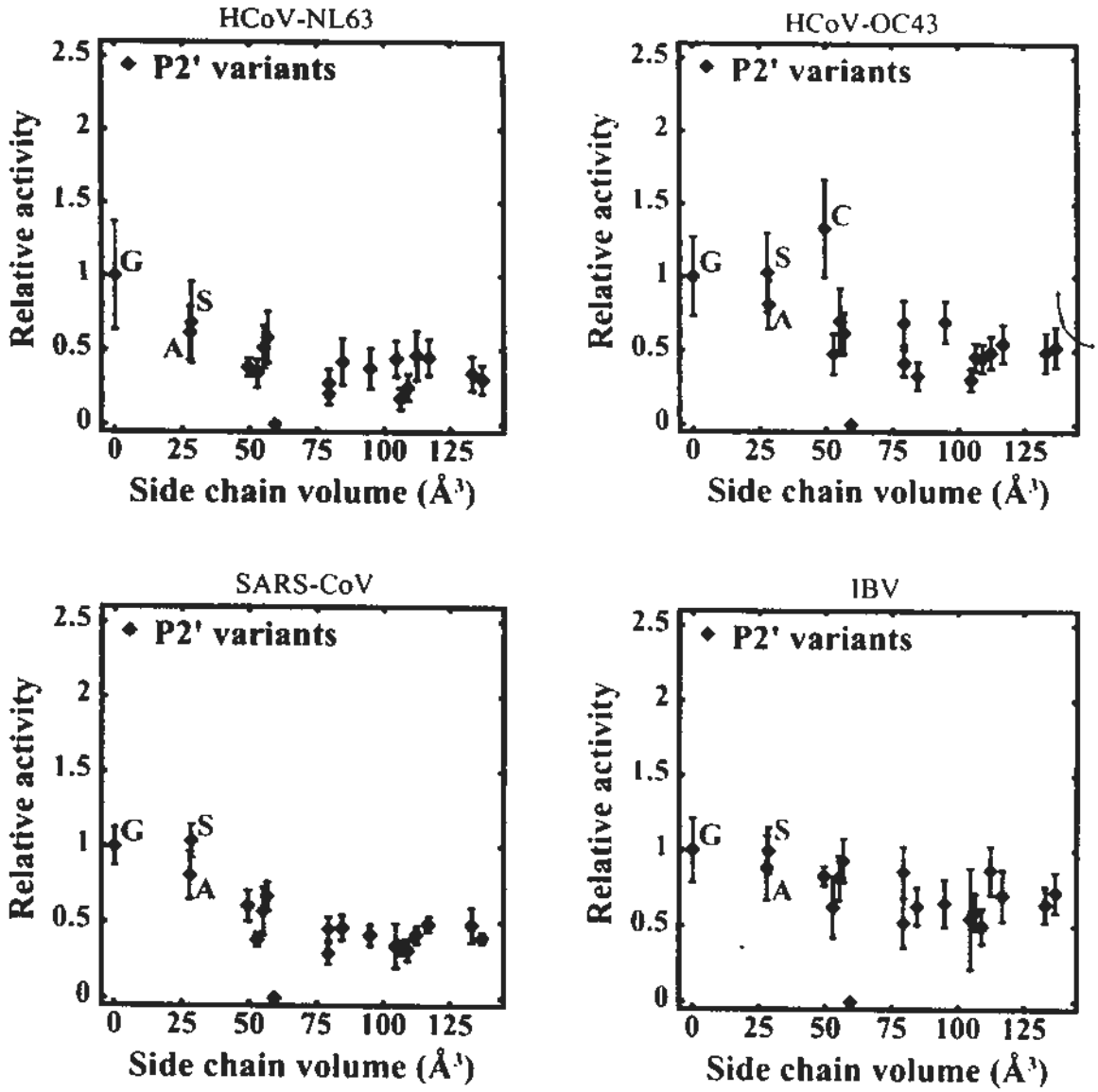


Figure 6.10 – All 3CL^{pro} preferred small P2' residues. The relative activities of all 3CL^{pro} against small P2 variants such as Gly, Ala and Ser were higher than other variants. Increase in P2' side chain volume caused gradient decrease in the activities.

6.3.8 No Strong Correlation is Found at P3' Position

Most of substitutions at P3' position led to reduction of all 3CL^{pro} activities. P3' residues with positively charged residues were relatively preferred, and were approximately 2-fold higher in activity than that with negatively charged ones. No strong preference was found with other structural properties.

6.4 Universal and Specific 'Super-Active' Substrate Sequences

The relative activities against single substitution substrate variants showed that all 3CL^{pro} prefer P5-Val, P4-Ala, P3-Arg/Lys, P2-Leu and P1-Gln while 3CL^{pro} of SARS-CoV and IBV individually favor substitutions to P4-Val and P4-Pro respectively. Those preferred residues were combined to generate four substrates with universal 'super-active' sequences (VARLQ↓SGF and VAKLQ↓SGF) or specific 'super-active' sequences (VPRLQ↓SGF and VPKLQ↓SGF) specifically for IBV. These 'super-active' substrates were examined by 3CL^{pro} of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV to study synergistic effect. The substrates were also tested by another two strains, HCoV-229E (group I) and HCoV-HKU1 (group IIa), to investigate similarity of the specificity among group members.

Relative activities of these 3CL^{pro} against universal 'super-active' substrates were 1.4 – 3.5, which were higher than those against corresponding single substitution substrates S5V, V3R and V3K (Table 6.4), showing that combination of preferred residues could raise the activity. When specific 'super-active' substrate (VPRLQ↓SGF) was examined, the activity of IBV 3CL^{pro} was further elevated to 4.33 ± 0.98. No significant advantage was found by substituting P3-Arg to Lys.

Proteolytic rates of 3CL^{pro} within the same group (HCoV-NL63 and HCoV-229E in group I; HCoV-OC43 and HCoV-HKU1 in group IIa) were similar, whatever WT substrate, single substitution substrates and 'super-active' substrates were examined, revealing that 3CL^{pro} of the same group share common substrate specificities.

Table 6.4 – Relative activity against universal and specific super-active substrate variants. Preferred residues in single substitution variants were combined to produce universal and specific super-active variants, which were examined by 6 of 3CL^{pro}. Substituting residues of the substrates were underlined.

	HCoV-NL63 (group I)	HCoV-229E (group I)	HCoV-OC43 (group IIa)	HCoV-HKU1 (group IIa)	SARS-CoV (group IIb)	IBV (group III)
Specific activity (mM^{-1} min^{-1})	443 ± 11	902 ± 11	124 ± 13	191 ± 28	180 ± 5	174 ± 19
Single variants						
SA<u>R</u>LQ↓	1.27 ± 0.06	1.09 ± 0.06	1.43 ± 0.28	1.66 ± 0.34	0.84 ± 0.06	1.77 ± 0.29
SA<u>K</u>LQ↓	1.24 ± 0.08	1.14 ± 0.21	1.52 ± 0.20	1.25 ± 0.19	0.80 ± 0.15	2.01 ± 0.26
S<u>P</u>VLQ↓	0.08 ± 0.02	0.05 ± 0.01	0.48 ± 0.14	0.72 ± 0.17	0.75 ± 0.08	1.08 ± 0.19
<u>V</u>AVLQ↓	1.20 ± 0.07	1.32 ± 0.10	1.59 ± 0.22	1.58 ± 0.23	1.53 ± 0.11	1.64 ± 0.18
Universal variant						
<u>V</u>A<u>R</u>LQ↓	1.70 ± 0.07	1.55 ± 0.02	2.43 ± 0.35	1.93 ± 0.28	1.70 ± 0.17	3.24 ± 0.37
<u>V</u>A<u>K</u>LQ↓	1.50 ± 0.10	1.65 ± 0.07	2.55 ± 0.39	2.30 ± 0.57	1.42 ± 0.11	3.45 ± 0.44
Specific variant						
<u>VPR</u>LQ↓	0.15 ± 0.04	0.09 ± 0.02	0.91 ± 0.12	1.30 ± 0.24	0.99 ± 0.12	4.33 ± 0.98
<u>VPK</u>LQ↓	0.13 ± 0.02	0.14 ± 0.02	1.08 ± 0.14	1.82 ± 0.40	1.04 ± 0.17	3.95 ± 0.81

6.5 Summary

Comprehensive substrate specificity at P5 to P3' positions of 3CL^{pro} of HCoV-NL63 (group I), HCoV-OC43 (group IIa), SARS-CoV (group IIb) and IBV (group III) were profiled by using the substrate library of 19×8 variants (Table 6.5). These 3CL^{pro} consistently favored small hydrophobic P4 residues, positively charged P3 residues, hydrophobic P2 residues without β-branch, Gln at P1 position and small residues at P1' position. Their activities also tended to be higher when substrate variants were substituted to positively charged residues at P5 and P3' positions, and small residues at P2' position. In contrast, their preferences on secondary structures were diverse. β-sheet propensity was only correlated to relative activity of SARS-CoV 3CL^{pro} against P5 and P3 variants, and that of IBV 3CL^{pro} against P5 variants. Only 3CL^{pro} of SARS-CoV and IBV were favored P4-Val and P4-Pro respectively. The results were independent of reaction conditions such as concentrations of NaCl, TCEP and proteases. Relative activities of those 3CL^{pro} against super-active substrates with combination of preferred residues were further elevated to 1.4 – 4.3. Relative activities of 3CL^{pro} in the same group against different substrate variants were highly comparable.

Table 6.5 – Summary of substrate specificity of various 3CL^{pro} at P5 to P3' positions.

Position	Universal preference	Universal preferred residue
P5	Positively charged residues	Val
P4	Small hydrophobic residues	Ala / Cys
P3	Positively charged residues	Arg / Lys / Val
P2	Hydrophobic residues without β-branch	Leu
P1	Gln	Gln
P1'	Small residues	Gly / Ala / Ser
P2'	Small residues	Gly / Ala / Ser
P3'	Positively charged residues	Arg / Phe / Tyr

Position	Individual preference (Correlated CoV)	Individual preferred residue
P5	Residues with high β-sheet propensity (SARS-CoV and IBV)	Val (SARS-CoV) Pro (IBV)
P4		
P3	Residues with high β-sheet propensity (SARS-CoV)	

Chapter 7 – Conclusive Remarks and Future Perspectives

7.1 Conclusive remarks

The substrate specificity of SARS-CoV $3CL^{pro}$ was profiled using a library of protein substrates. Unlike previous studies that based on short peptide substrates (Fan *et al.*, 2005; Goetz *et al.*, 2007), the use of the protein substrate, which consists of a 12-residue autocleavage sequence, should better mimic the polyprotein substrate of SARS-CoV. The effect of substitution of each of the residues at P5 to P3' positions were investigated. The comprehensive data obtained allowed us to quantitatively correlate the substrate specificity in terms of side chain volume, hydrophobicity and secondary structure propensities. Not only our results are consistent with some of the previous observations, novel insights into the substrate specificity were obtained in this study. First, positively charged residues are consistently preferred over negatively charged ones at solvent-exposed positions such as P5, P3, P3'. Second, the $3CL^{pro}$ activity is directly proportional to hydrophobicity for small residues at P4, and to β-sheet propensities at P5 and P3 positions. Third, residues larger than Cys are not favored at P1' position. Fourth, the most favored residue at P1 position is Gln, but P1-His and P1-Met are also cleavable. Our results suggest the existence of a strong structure-activity relationship between $3CL^{pro}$ and its substrates. The substrate specificity profiled in this study can be used as a benchmark for better computational simulation for $3CL^{pro}$-substrate/inhibitor interaction, and may provide a guideline for a rational based design of potent inhibitors.

Potent peptidomimetic inhibitors require an effective and bioactive warhead for inactivating catalytic site. Nitrile group is a well-known warhead applied for targeting Cys proteases, but no studies proved inhibitory effect against $3CL^{pro}$. We produced a series of nitrile-based peptidomimetic inhibitors and examined their potency in terms of IC_{50} and K_i values against SARS-CoV $3CL^{pro}$. These inhibitors can significantly reduce $3CL^{pro}$ activity in micromolar range, indicating that nitrile group can effectively inhibit $3CL^{pro}$ proteolysis. The most potent inhibitors is 'cbz-AVLQ-CN', with IC_{50} and K_i values of 5.9 ± 0.6 μM and 0.62 ± 0.11 μM respectively. $3CL^{pro}$-inhibitor complexes demonstrated that nitrile group inactivated $3CL^{pro}$ proteolysis by covalently bonding with Cys145. Peptides of the inhibitors could interact with $3CL^{pro}$ as substrate. Cbz group in 'cbz-AVLQ-CN' flipped into a cavity of Glu166 – Pro168, leading to the outstanding potency. Taken together, 'cbz-AVLQ-CN' is a convincing model for $3CL^{pro}$ drug development.

A number of researches focused on development of wide-spectrum inhibitors but understanding on universal substrate specificity of $3CL^{pro}$ was deficient. Our profiles have demonstrated comprehensive specificities of $3CL^{pro}$ in various groups. $3CL^{pro}$ activities of HCoV-NL63 (group I), HCoV-OC43 (group IIa), SARS-CoV (group IIb) and IBV (group III) against the substrate library of 19× 8 variants were examined by FRET assay, and were correlated with structural properties of substituting residues. Preferences of these $3CL^{pro}$ on charge, side chain volume and hydrophobicity were generally universal. All $3CL^{pro}$ clearly preferred small hydrophobic P4 residues, positively charged P3 residues, hydrophobic P2 residues without β-branch, P1-Gln and small P1' residues. They also tended to favor positively charged P5 and P3 residues, and small P2' residues, but the preferences were less obvious. On the contrary, effects of secondary structure propensities on

those $3CL^{pro}$ activities were diverse. Only $3CL^{pro}$ of SARS-CoV were correlated well to β-sheet propensity at P5 and P3 positions. Correlation was found only between IBV $3CL^{pro}$ activity and β-sheet propensity at P5 position. No any significant correlation related to secondary structure propensity was observed when $3CL^{pro}$ of HCoV-NL63 and HCoV-OC43 were examined. Based on the preference, universal substrate sequences (VARLQ↓ / VAKLQ↓) and substrate sequences (VPRLQ↓ / VPKLQ↓) specific for IBV $3CL^{pro}$ were found. Our results provide novel information on similarity and difference of substrate specificities among various $3CL^{pro}$. It is important for development of drugs targeting CoVs in wide-spectrum and novel CoV in the future.

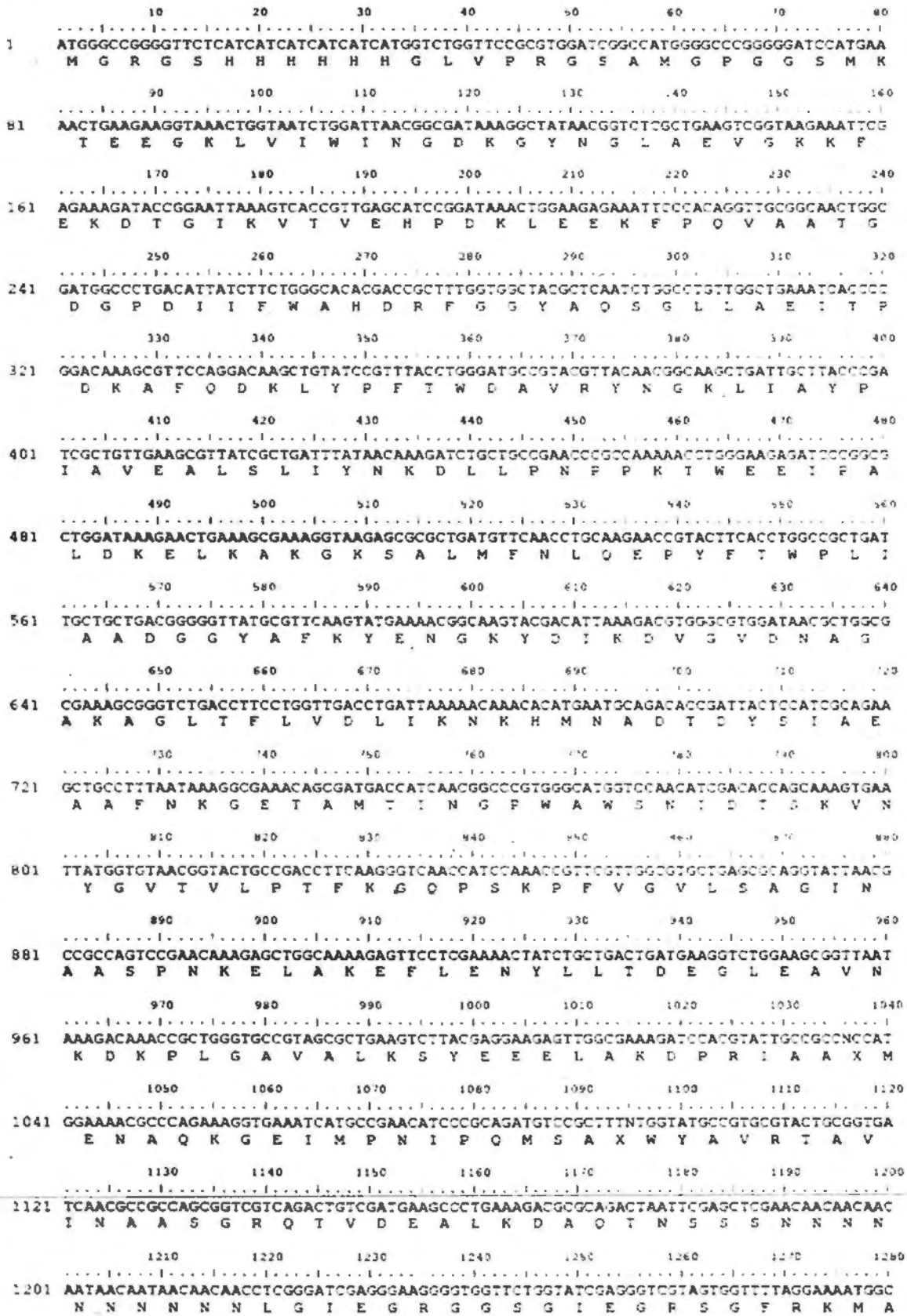
7.2 Future Perspectives

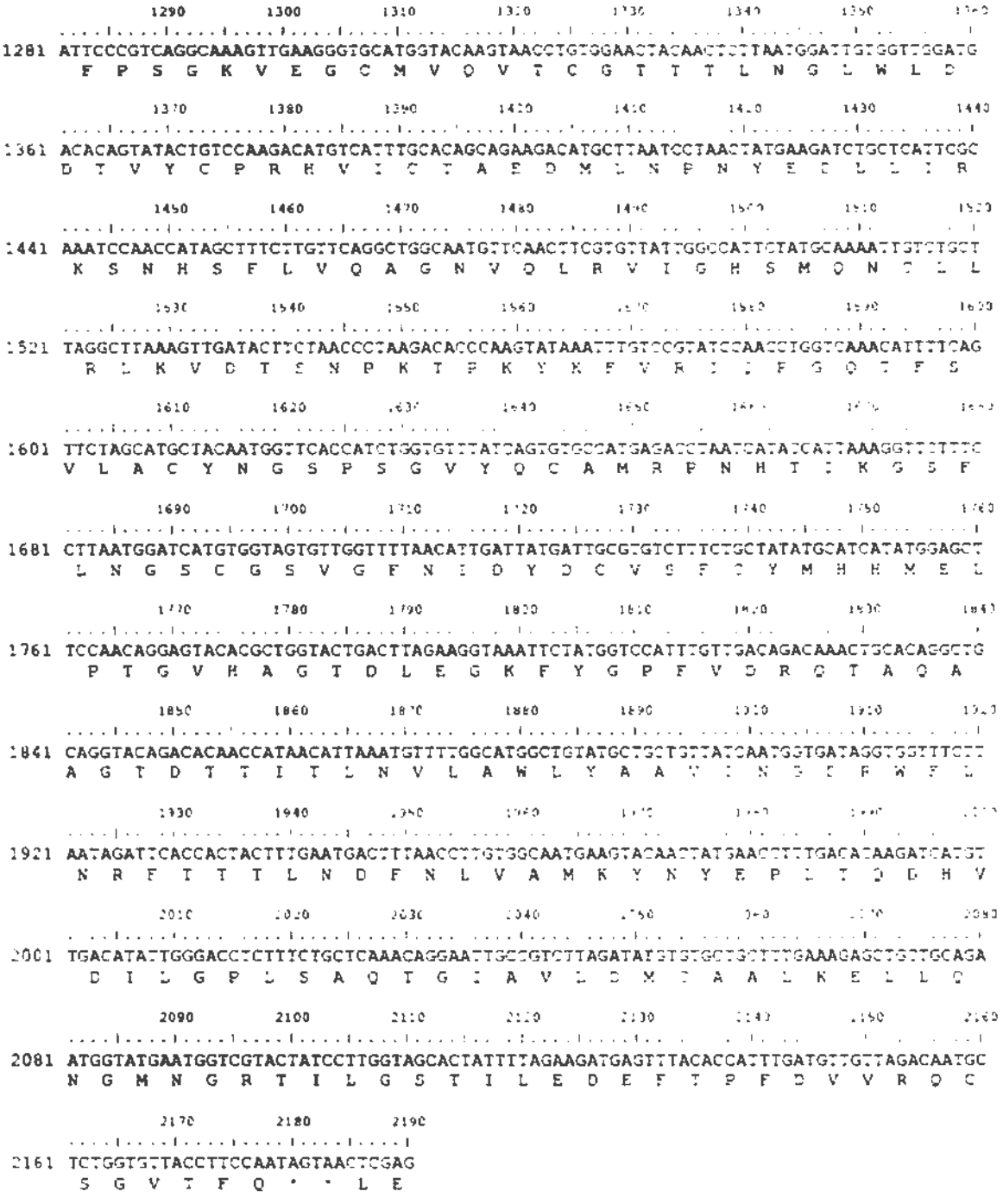
A number of research studies showed that the covalent bond between nitrile and Cys residue is reversible (Brisson *et al.*, 1986; Moon *et al.*, 1986). To determine whether our nitrile-based inhibitors are reversible, we will attempt to remove inhibitors from $3CL^{pro}$-inhibitor complex, followed by examining whether the protease activity can be recovered.

'cbz-AVLQ-CN' has been showed to effectively reduce $3CL^{pro}$ activity of SARS-CoV. Meanwhile, substrate specificities among various $3CL^{pro}$ are generally comparable. Therefore, potency of 'cbz-AVLQ-CN' on other $3CL^{pro}$ will be examined to investigate whether it can target $3CL^{pro}$ in broad spectrum. It will also be examined by a cell-based assay to justify its *in vivo* inhibitory effect.

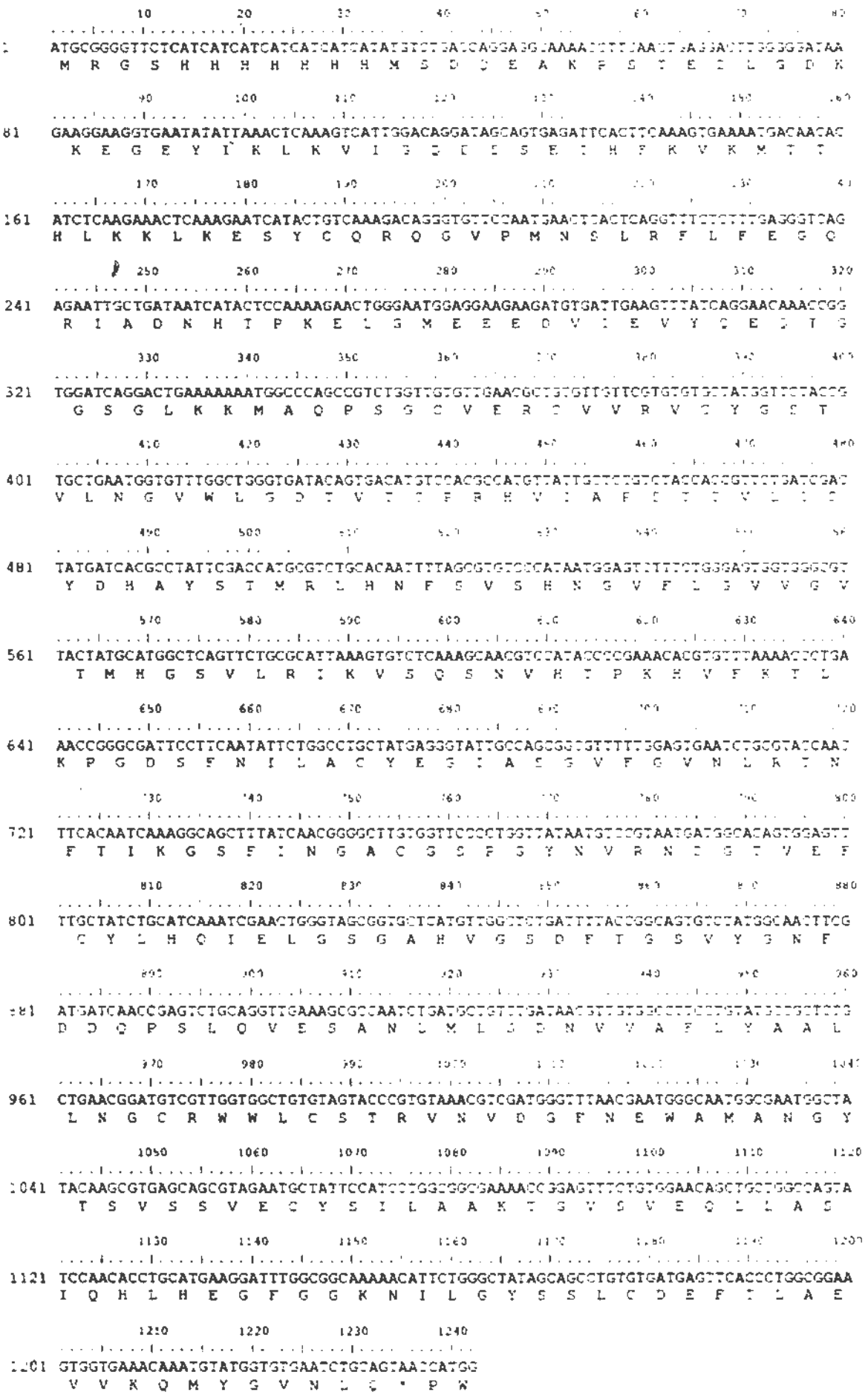
Appendixes

Appendix 2.1 – Sequences of SARS-CoV 3CL^{pro} with His$_6$-MBP tag.

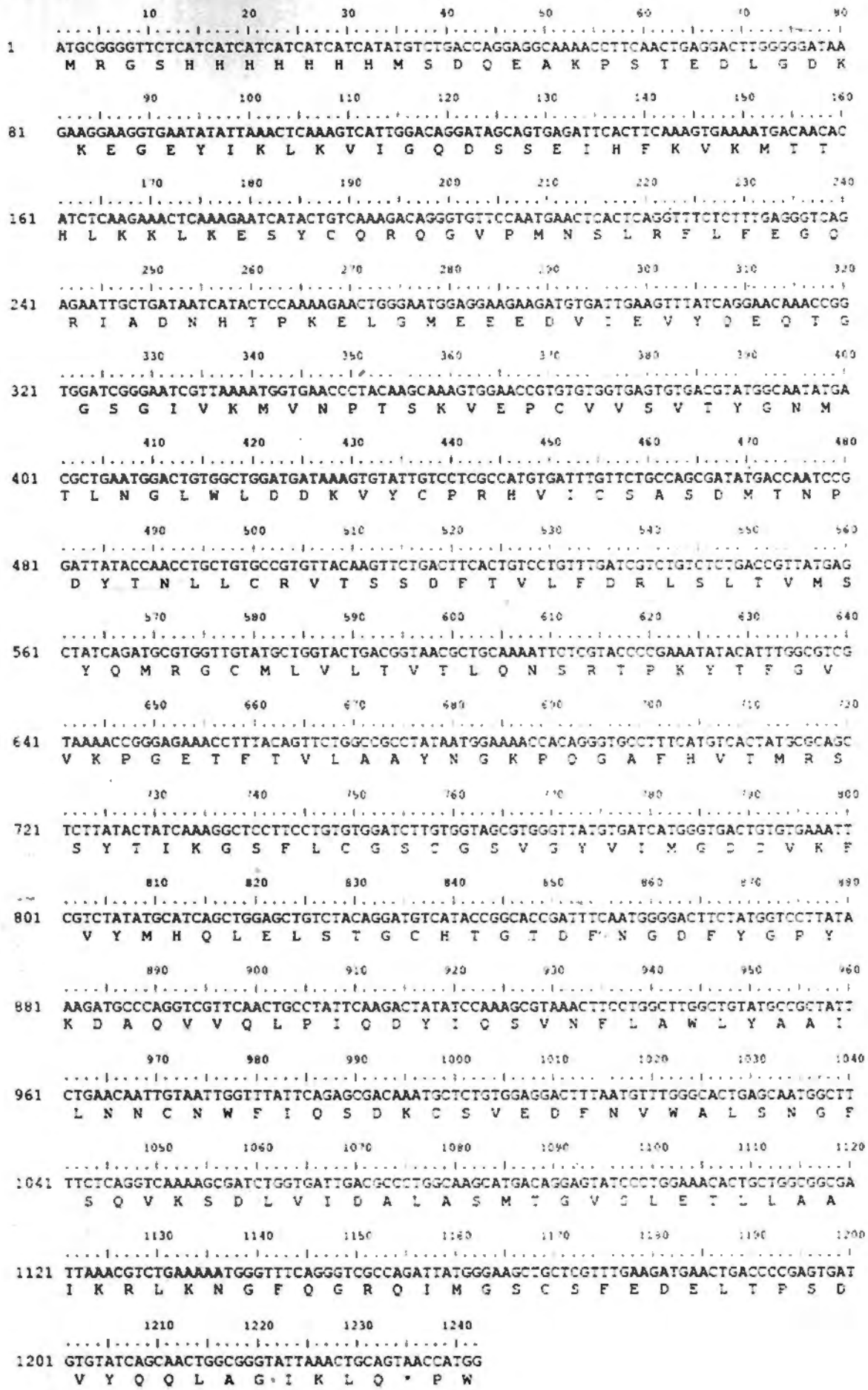




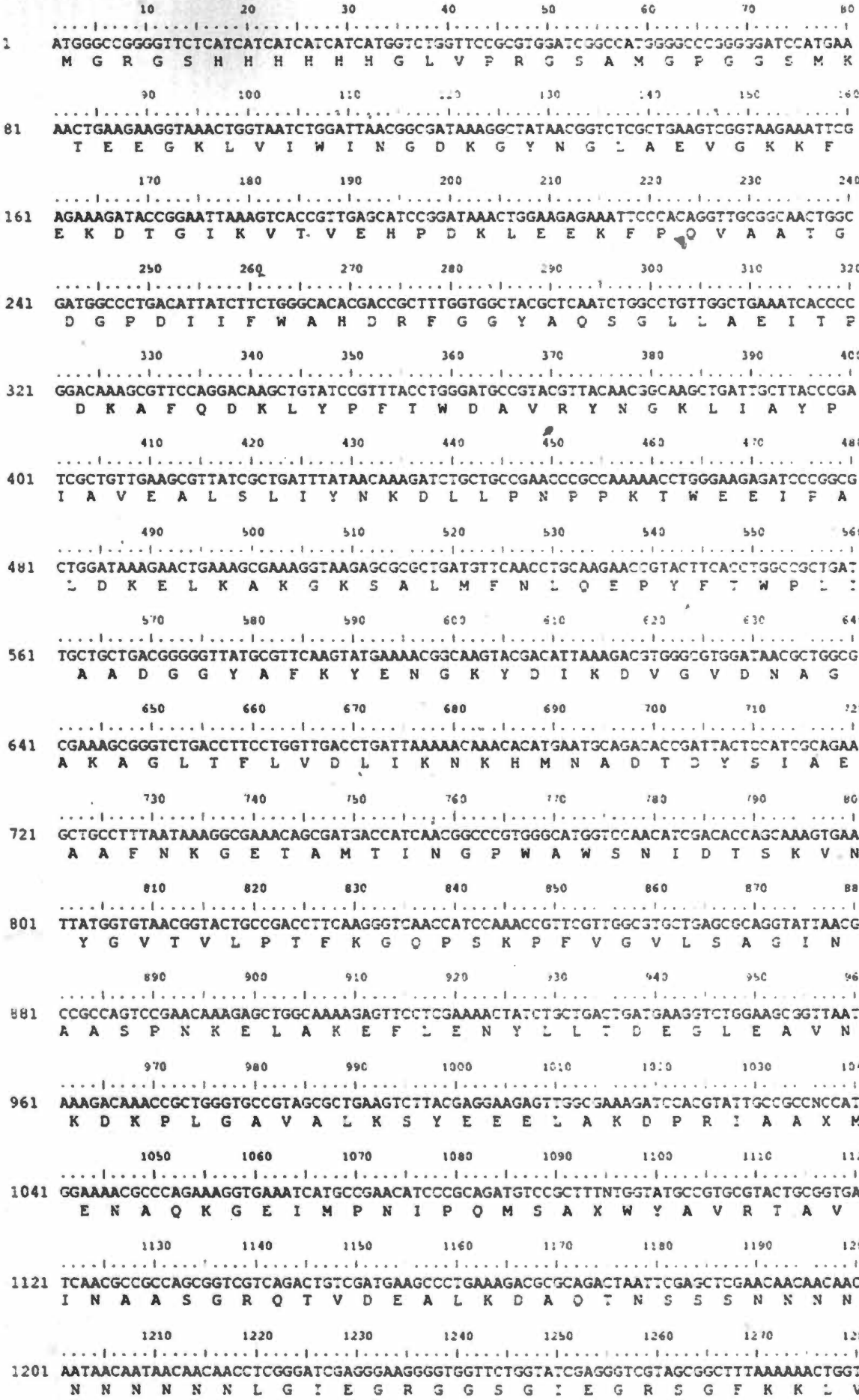
Appendix 2.2 – Sequences of HCoV-NL63 3CL^{pro} with His$_6$-SUMO tag.

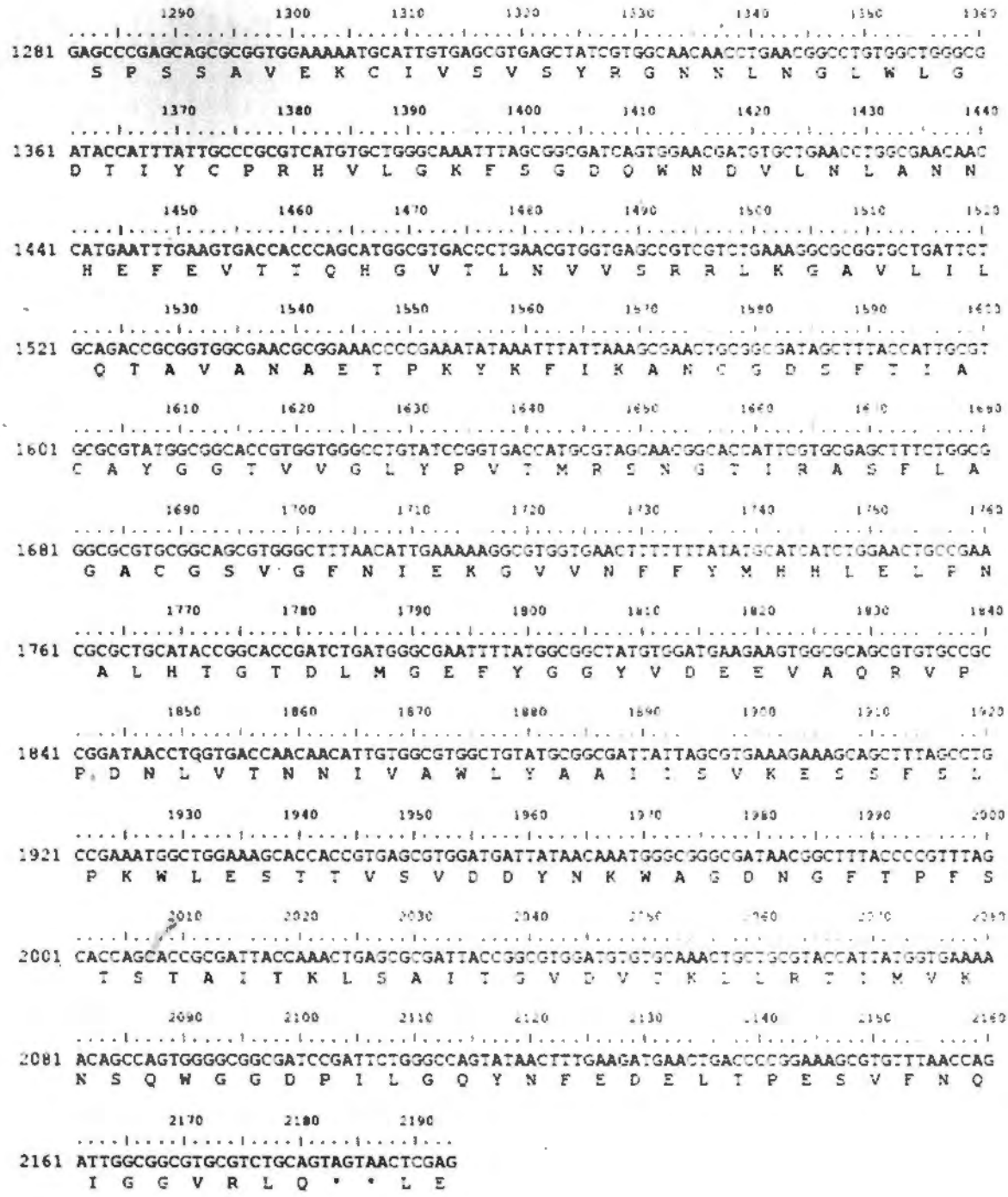


Appendix 2.3 – Sequences of HCoV-OC43 3CL^{pro} with His$_6$-SUMO tag.



Appendix 2.4 – Sequences of IBV $3CL^{pro}$ with His_6-MBP tag.

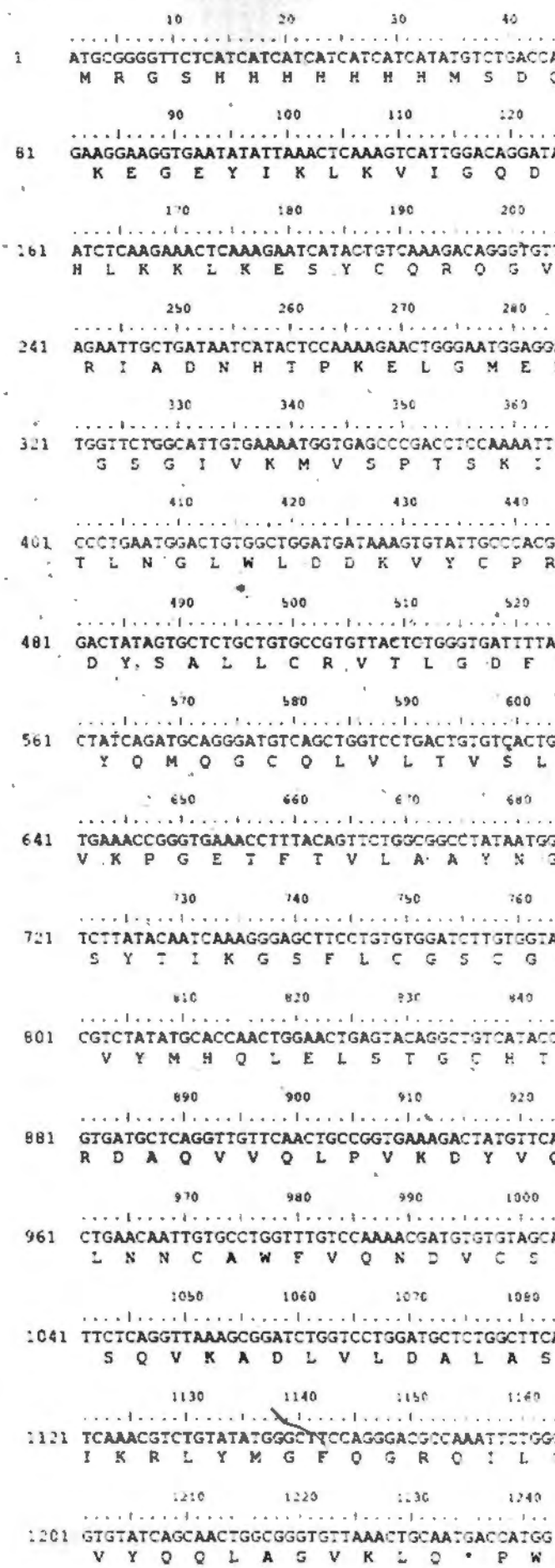



1290 1300 1310 1320 1330 1340 1350 1360
1281 GAGCCCGAGCAGCGCGGTGGAAAAATGCATTGTGAGCGTGAGCTATCGTGGCAACAACCTGAACGGCCTGTGGCTGGGCG
S P S S A V E K C I V S V S Y R G N N L N G L W L G
1370 1380 1390 1400 1410 1420 1430 1440
1361 ATACCATTTATTGCCCGCGTCATGTGCTGGGCAAATTTAGCGGCGATCAGTGGAACGATGTGCTGAACCTGGCGAACAAC
D T I Y C P R H V L G K F S G D Q W N D V L N L A N N
1450 1460 1470 1480 1490 1500 1510 1520
1441 CATGAATTTGAAGTGACCACCCAGCATGGCGTGACCCTGAACGTGGTGAGCCGTCGTCTGAAAGGCGCGGTGCTGATTCT
H E F E V T T Q H G V T L N V V S R R L K G A V L I L
1530 1540 1550 1560 1570 1580 1590 1600
1521 GCAGACCGCGGTGGCGAACGCGGAAACCCCGAAATATAAATTTATTAAAGCGAACTGCGGCGATAGCTTTACCATTGCGT
Q T A V A N A E T P K Y K F I K A N C G D S F T I A
1610 1620 1630 1640 1650 1660 1670 1680
1601 GCGCGTATGGCGGCACCGTGGTGGGCCTGTATCCGGTGACCATGCGTAGCAACGGCAGCATTCGTGCGAGCTTTCTGGCG
C A Y G G T V V G L Y P V T M R S N G T I R A S F L A
1690 1700 1710 1720 1730 1740 1750 1760
1681 GGCGCGTGCGGCAGCGTGGGCTTTAACATTGAAAAAGGCGTGGTGAACTTTTTTATATGCATCATCTGGAACTGCCGAA
G A C G S V G F N I E K G V V N F F Y M H H L E L P N
1770 1780 1790 1800 1810 1820 1830 1840
1761 CGCGCTGCATACCGGCACCGATCTGATGGGCGAATTTTATGGCGGCTATGTGGATGAAGAAGTGGCGCAGCGTGTGCCGC
A L H T G T D L M G E F Y G G Y V D E E V A Q R V P
1850 1860 1870 1880 1890 1900 1910 1920
1841 CGGATAACCTGGTGACCAACAACATTGTGGCGTGGCTGTATGCGGCGATTATTAGCGTGAAAGAAAGCAGCTTTAGCCTG
P D N L V T N N I V A W L Y A A I I S V K E S S F S L
1930 1940 1950 1960 1970 1980 1990 2000
1921 CCGAAATGGCTGGAAAGCACCACCGTGAGCGTGGATGATTATAACAAATGGGCGGGCGATAACGGCTTTACCCCGTTTAG
P K W L E S T T V S V D D Y N K W A G D N G F T P F S
2010 2020 2030 2040 2050 2060 2070 2080
2001 CACCAGCACCGCGATTACCAAACTGAGCGCGATTACCGGCGTGGATGTGTGCAAACTGCTGCGTACCATTATGGTGAAAA
T S T A I T K L S A I T G V D V C K L L R T I M V K
2090 2100 2110 2120 2130 2140 2150 2160
2081 ACAGCCAGTGGGGCGGCGATCCGATTCTGGGCCAGTATAACTTTGAAGATGAACTGACCCCGGAAAGCGTGTTTAACCAG
N S Q W G G D P I L G Q Y N F E D E L T P E S V F N Q
2170 2180 2190
2161 ATTGGCGGCGTGCGTCTGCAGTAGTAACTCGAG
I G G V R L Q * * L E

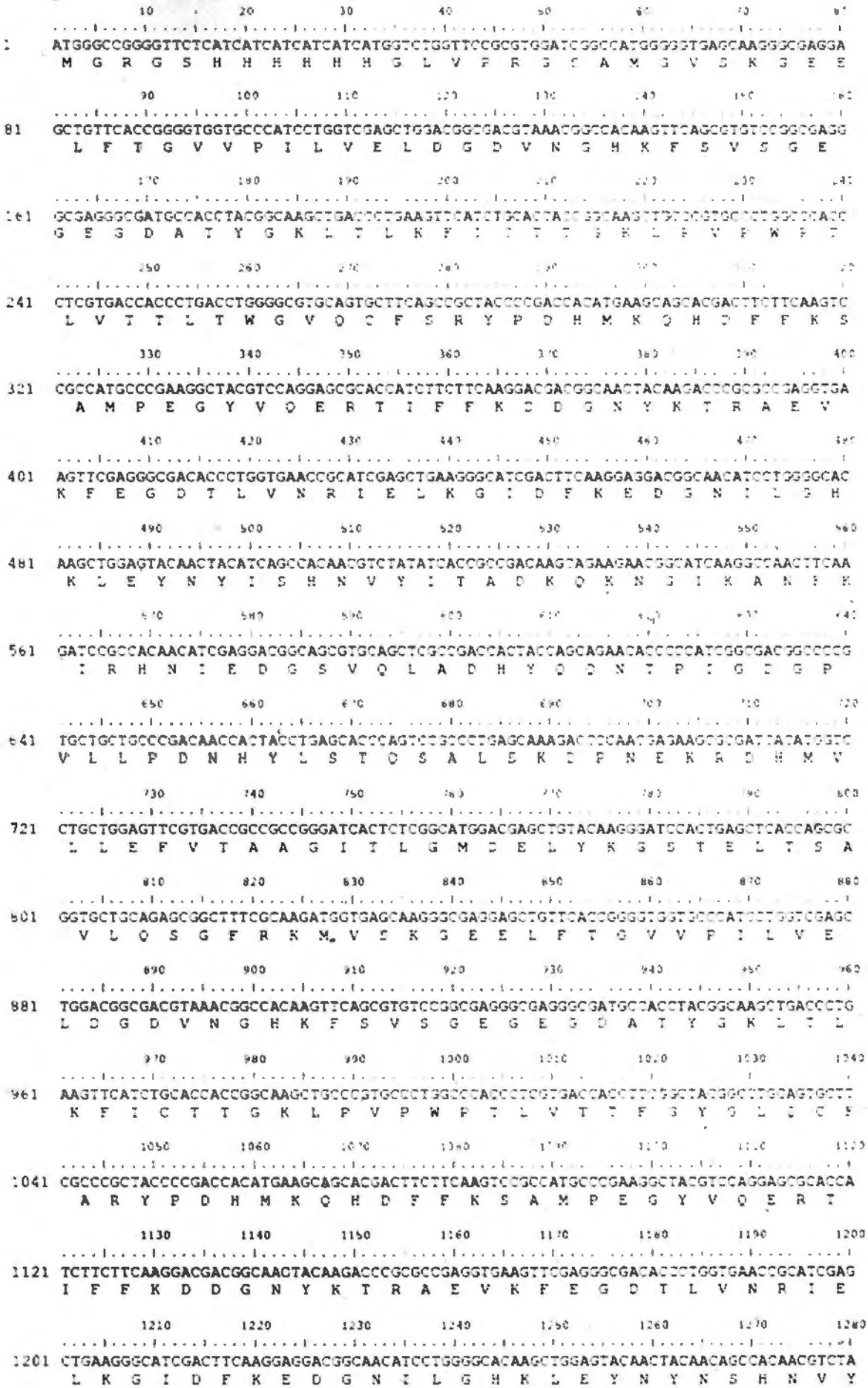
Appendix 2.5 – Sequences of HCoV-229E 3CL^{pro} with His$_6$-SUMO tag.

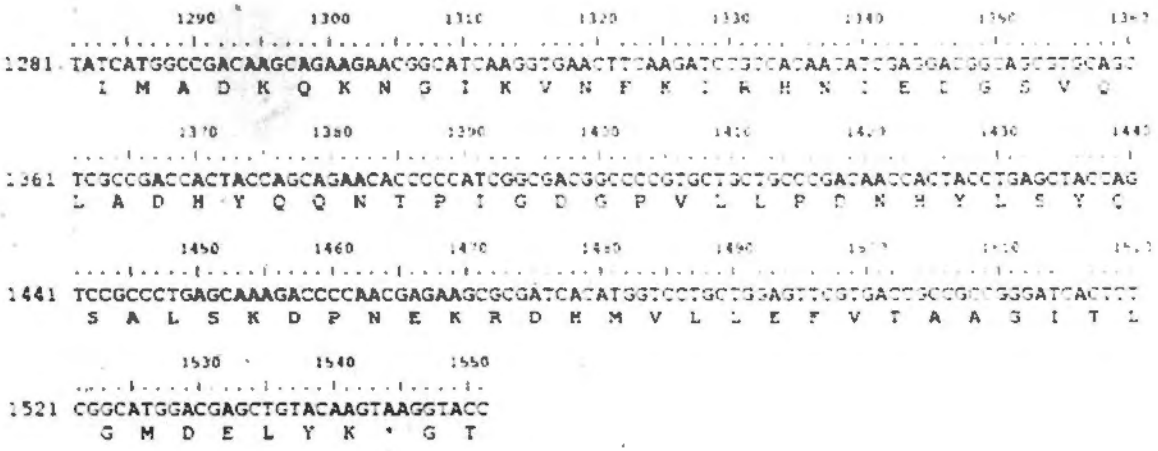
```
          10         20         30         40         50         60         70         80
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1     ATGCGGGGTTCTCATCATCATCATCATCATCATATGTCTGACCAGGAGGCAAAACCTTCAACTGAGGACTTGGGGGATAA
        M  R  G  S  H  H  H  H  H  H  H  M  S  D  Q  E  A  K  P  S  T  E  D  L  G  D  K

          90        100        110        120        130        140        150        160
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
81    GAAGGAAGGTGAATATATTAAACTCAAAGTCATTGGACAGGATAGCAGTGAGATTCACTTCAAAGTGAAAATGACAACAC
        K  E  G  E  Y  I  K  L  K  V  I  G  Q  D  S  S  E  I  H  F  K  V  K  M  T  T

          170        180        190        200        210        220        230        240
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
161   ATCTCAAGAAACTCAAAGAATCATACTGTCAAAGACAGGGTGTTCCAATGAACTCACTCAGGTTTCTCTTTGAGGGTCAG
        H  L  K  K  L  K  E  S  Y  C  Q  R  Q  G  V  P  M  N  S  L  R  F  L  F  E  G  Q

          250        260        270        280        290        300        310        320
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
241   AGAATTGCTGATAATCATACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATTGAAGTTTATCAGGAACAAACCGG
        R  I  A  D  N  H  T  P  K  E  L  G  M  E  E  E  D  V  I  E  V  Y  Q  E  Q  T  G

          330        340        350        360        370        380        390        400
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
321   TGGTGCTGGTCTGCGTAAAATGGCCCAACCTTCAGGCTTTGTCGAGAAATGTGTCGTTCGTGTGTGCTATGGCAATACTG
        G  A  G  L  R  K  M  A  Q  P  S  G  F  V  E  K  C  V  V  R  V  C  Y  G  N  T

          410        420        430        440        450        460        470        480
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
401   TGCTGAATGGACTGTGGCTGGGTGATATCGTCTATTGTCCTCGCCATGTGATTGCTTCCAACACTACCAGTGCTATCGAT
        V  L  N  G  L  W  L  G  D  I  V  Y  C  P  R  H  V  I  A  S  N  T  T  S  A  I  D

          490        500        510        520        530        540        550        560
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
481   TATGACCACGAGTATAGTATCATGCGCCTGCACAACTTTTCTATCATCAGCGGCACCGCTTTTCTGGGTGTCGTAGGAGC
        Y  D  H  E  Y  S  I  M  R  L  H  N  F  S  I  I  S  G  T  A  F  L  G  V  V  G  A

          570        580        590        600        610        620        630        640
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
561   AACAATGCATGGTGTTACACTGAAAATCAAAGTGAGCCAGACTAATATGCATACCCCTCGTCACTCATTCCGTACTCTGA
        T  M  H  G  V  T  L  K  I  K  V  S  Q  T  N  M  H  T  P  R  H  S  F  R  T  L

          650        660        670        680        690        700        710        720
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
641   AATCCGGTGAGGGCTTTAACATCCTGGCCTGCTATGATGGCTGTGCTCAGGGTGTGTTTGGCGTGAATATGCGCACAAAT
        K  S  G  E  G  F  N  I  L  A  C  Y  D  G  C  A  Q  G  V  F  G  V  N  M  R  T  N

          730        740        750        760        770        780        790        800
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
721   TGGACTATCCGTGGCAGCTTTATTAACGGCGCCTGTGGTTCTCCTGGCTATAACCTGAAAAACGGCGAAGTGGAATTCGT
        W  T  I  R  G  S  F  I  N  G  A  C  G  S  P  G  Y  N  L  K  N  G  E  V  E  F  V

          810        820        830        840        850        860        870        880
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
801   GTATATGCACCAGATTGAGCTGGGTTCTGGATCTCATGTGGGTTCGAGCTTCGATGGTGTGATGTATGGTGGCTTCGAAG
        Y  M  H  Q  I  E  L  G  S  G  S  H  V  G  S  S  F  D  G  V  M  Y  G  G  F  E

          890        900        910        920        930        940        950        960
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
881   ATCAGCCGAATCTGCAGGTCGAAAGCGCTAATCAGATGCTGACCGTGAACGTCGTTGCCTTTCTGTATGCCGCCATTCTG
        D  Q  P  N  L  Q  V  E  S  A  N  Q  M  L  T  V  N  V  V  A  F  L  Y  A  A  I  L

          970        980        990        1000       1010       1020       1030       1040
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
961   AACGGCTGTACTTGGTGGCTGAAAGGTGAGAAACTGTTTGTGGAGCATTATAACGAGTGGGCTCAAGCCAATGGTTTCAC
        N  G  C  T  W  W  L  K  G  E  K  L  F  V  E  H  Y  N  E  W  A  Q  A  N  G  F  T

          1050       1060       1070       1080       1090       1100       1110       1120
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1041  TGCCATGAACGGTGAAGATGCCTTTAGCATTCTGGCAGCCAAAACAGGTGTGTGTGTTGAGCGTCTGCTGCACGCAATTC
        A  M  N  G  E  D  A  F  S  I  L  A  A  K  T  G  V  C  V  E  R  L  L  H  A  I

          1130       1140       1150       1160       1170       1180       1190       1200
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1121  AAGTCCTGAATAACGGGTTTGGTGGGAAACAAATTCTGGGCTATTCCTCCCTGAATGATGAGTTCTCCATCAACGAAGTC
        Q  V  L  N  N  G  F  G  G  K  Q  I  L  G  Y  S  S  L  N  D  E  F  S  I  N  E  V

          1210       1220       1230
....|....|....|....|....|....|....
1201  GTTAAACAAATGTTCGGCGTCAACCTGCAATGACCATGG
        V  K  Q  M  F  G  V  N  L  Q  *  P  W
```

Appendix 2.6 – Sequences of HCoV-HKU1 3CL^{pro} with His$_6$-SUMO tag.



Appendix 2.7 – Sequences of protein-based WT substrate.




1290 1300 1310 1320 1330 1340 1350
1281 TATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC
I M A D K Q K N G I K V N F K I R H N I E D G S V Q
1370 1380 1390 1400 1410 1420 1430 1440
1361 TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAG
L A D H Y Q Q N T P I G D G P V L L P D N H Y L S Y Q
1450 1460 1470 1480 1490 1500 1510 1520
1441 TCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCT
S A L S K D P N E K R D H M V L L E F V T A A G I T L
1530 1540 1550
1521 CGGCATGGACGAGCTGTACAAGTAAGGTACC
G M D E L Y K * G T

Appendix 2.8 – Primer sequences for overlapping PCR of substrate library.

Mutated sequences and restriction sites are highlighted and underlined respectively.

Variants	Primer sequences
S5A	GATATCGAGCTCACCGCGGCGGTGCTGCAGAGCGGC
S5C	GATATCGAGCTCACCTGCGCGGTGCTGCAGAGCGGC
S5D	GATATCGAGCTCACCGATGCGGTGCTGCAGAGCGGC
S5E	GATATCGAGCTCACCGAGGCGGTGCTGCAGAGCGGC
S5F	GATATCGAGCTCACCTTTGCGGTGCTGCAGAGCGGC
S5G	GATATCGAGCTCACCGGGGCGGTGCTGCAGAGCGGC
S5H	GATATCGAGCTCACCCATGCGGTGCTGCAGAGCGGC
S5I	GATATCGAGCTCACCATTGCGGTGCTGCAGAGCGGC
S5K	GATATCGAGCTCACCAAGGCGGTGCTGCAGAGCGGC
S5L	GATATCGAGCTCACCCTGGCGGTGCTGCAGAGCGGC
S5M	GATATCGAGCTCACCATGGCGGTGCTGCAGAGCGGC
S5N	GATATCGAGCTCACCAACGCGGTGCTGCAGAGCGGC
S5P	GATATCGAGCTCACCCCGGCGGTGCTGCAGAGCGGC
S5Q	GATATCGAGCTCACCCAGGCGGTGCTGCAGAGCGGC
S5R	GATATCGAGCTCACCGGGGCGGTGCTGCAGAGCGGC
S5T	GATATCGAGCTCACCACGGCGGTGCTGCAGAGCGGC
S5V	GATATCGAGCTCACCGTGGCGGTGCTGCAGAGCGGC
S5W	GATATCGAGCTCACCTGGGCGGTGCTGCAGAGCGGC
S5Y	GATATCGAGCTCACCTATGCGGTGCTGCAGAGCGGC
V3A	GATATCGAGCTCACCAGCGCGGCGCTGCAGAGCGGCTTTCGC
V3C	GATATCGAGCTCACCAGCGCGTGCCTGCAGAGCGGCTTTCGC
V3D	GATATCGAGCTCACCAGCGCGGATCTGCAGAGCGGCTTTCGC
V3E	GATATCGAGCTCACCAGCGCGGAGCTGCAGAGCGGCTTTCGC
V3F	GATATCGAGCTCACCAGCGCGTTTCTGCAGAGCGGCTTTCGC
V3G	GATATCGAGCTCACCAGCGCGGGGCTGCAGAGCGGCTTTCGC
V3H	GATATCGAGCTCACCAGCGCGCACCTGCAGAGCGGCTTTCGC
V3I	GATATCGAGCTCACCAGCGCGATTCTGCAGAGCGGCTTTCGC
V3K	GATATCGAGCTCACCAGCGCGAAGCTGCAGAGCGGCTTTCGC
V3L	GATATCGAGCTCACCAGCGCGCTGCTGCAGAGCGGCTTTCGC
V3M	GATATCGAGCTCACCAGCGCGATGCTGCAGAGCGGCTTTCGC
V3N	GATATCGAGCTCACCAGCGCGAACCTGCAGAGCGGCTTTCGC
V3P	GATATCGAGCTCACCAGCGCGCCGCTGCAGAGCGGCTTTCGC
V3Q	GATATCGAGCTCACCAGCGCGCAGCTGCAGAGCGGCTTTCGC
V3R	GATATCGAGCTCACCAGCGCGCGGCTGCAGAGCGGCTTTCGC
V3S	GATATCGAGCTCACCAGCGCGTCGCTGCAGAGCGGCTTTCGC
V3T	GATATCGAGCTCACCAGCGCGACGCTGCAGAGCGGCTTTCGC
V3W	GATATCGAGCTCACCAGCGCGTGGCTGCAGAGCGGCTTTCGC
V3Y	GATATCGAGCTCACCAGCGCGTATCTGCAGAGCGGCTTTCGC
F3'A	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCGCGCGCAAGATGGTGAGCAAG
F3'C	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCTGCCGCAAGATGGTGAGCAAG
F3'D	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCGATCGCAAGATGGTGAGCAAG
F3'E	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCGAACGCAAGATGGTGAGCAAG
F3'G	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCGGTCGCAAGATGGTGAGCAAG
F3'H	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCCATCGCAAGATGGTGAGCAAG
F3'I	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCATTCGCAAGATGGTGAGCAAG
F3'K	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCAAACGCAAGATGGTGAGCAAG
F3'L	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCCTGCGCAAGATGGTGAGCAAG
F3'M	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCATGCGCAAGATGGTGAGCAAG
F3'N	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCAACCGCAAGATGGTGAGCAAG
F3'P	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCCCGCGCAAGATGGTGAGCAAG
F3'Q	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCCAGCGCAAGATGGTGAGCAAG
F3'R	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCCGTCGCAAGATGGTGAGCAAG
F3'S	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCTCGCGCAAGATGGTGAGCAAG
F3'T	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCACCCGCAAGATGGTGAGCAAG
F3'V	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCGTGCGCAAGATGGTGAGCAAG
F3'W	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCTGGCGCAAGATGGTGAGCAAG
F3'Y	GATATCGAGCTCACCAGCGCGGTGCTGCAGAGCGGCTATCGCAAGATGGTGAGCAAG
Reverse primer	TAGGGCGGTACCTTACTTGTACAGCTC

Appendix 2.9 – Primer sequences for overlapping PCR of super-active substrate variants. Mutated sequences and restriction sites are underlined and highlighted respectively.

Substrate sequence	Primer sequence
TTVVLQ↓SGFRKM	TAGGGCGAGCTCACCACGGTGGTGCTGCAGAGCGGC
TFVVLQ↓SGFRKM	TAGGGCGAGCTCACCTTTGTGGTGCTGCAGAGCGGC
TVVVLQ↓SGFRKM	TAGGGCGAGCTCACCGTGGTGGTGCTGCAGAGCGGC
TTVRLQ↓SGFRKM	TAGGGCGAGCTCACCACGGTGCGGCTGCAGAGCGGC
TFVRLQ↓SGFRKM	TAGGGCGAGCTCACCTTTGTGCGGCTGCAGAGCGGC
TVVRLQ↓SGFRKM	TAGGGCGAGCTCACCGTGGTGCGGCTGCAGAGCGGC

Appendix 2.10 – Forward primer sequences for quikchange mutatgenesis of substrate library. Mutated sequences are underlined. Reverse primers are reversely complementary to the forward primers.

Variants	Primer sequences
S5D	CCACTGAGCTCACCGATGCGGTGCTGCAGAGC
S5E	CCACTGAGCTCACCGAGGCGGTGCTGCAGAGC
S5F	CCACTGAGCTCACCTTTGCGGTGCTGCAGAGC
S5G	CCACTGAGCTCACCGGGGCGGTGCTGCAGAGC
S5H	CCACTGAGCTCACCCATGCGGTGCTGCAGAGC
S5I	CCACTGAGCTCACCATTGCGGTGCTGCAGAGC
S5K	CCACTGAGCTCACCAAGGCGGTGCTGCAGAGC
S5L	CCACTGAGCTCACCCTGGCGGTGCTGCAGAGC
S5M	CCACTGAGCTCACCATGGCGGTGCTGCAGAGC
S5N	CCACTGAGCTCACCAACGCGGTGCTGCAGAGC
S5P	CCACTGAGCTCACCCCGGCGGTGCTGCAGAGC
S5Q	CCACTGAGCTCACCCAGGCGGTGCTGCAGAGC
S5R	CCACTGAGCTCACCCGGGCGGTGCTGCAGAGC
S5T	CCACTGAGCTCACCACGGCGGTGCTGCAGAGC
S5V	CCACTGAGCTCACCGTGGCGGTGCTGCAGAGC
S5W	CCACTGAGCTCACCTGGGCGGTGCTGCAGAGC
S5Y	CCACTGAGCTCACCTATGCGGTGCTGCAGAGC
V3C	GAGCTCACCAGCGCGTGCCTGCAGAGCGGCTTTC
V3D	GAGCTCACCAGCGCGGATCTGCAGAGCGGCTTTC
V3F	GAGCTCACCAGCGCGTTTCTGCAGAGCGGCTTTC
V3G	GAGCTCACCAGCGCGGGGCTGCAGAGCGGCTTTC
V3H	GAGCTCACCAGCGCGCACCTGCAGAGCGGCTTTC
V3I	GAGCTCACCAGCGCGATTCTGCAGAGCGGCTTTC
F3'A	GGTGCTGCAGAGCGGCGCGCGCAAGATGGTGAGC
F3'C	GGTGCTGCAGAGCGGCTGCCGCAAGATGGTGAGC
F3'D	GGTGCTGCAGAGCGGCGATCGCAAGATGGTGAGC
F3'E	GGTGCTGCAGAGCGGCGAACGCAAGATGGTGAGC
F3'G	GGTGCTGCAGAGCGGCGGTCGCAAGATGGTGAGC
F3'H	GGTGCTGCAGAGCGGCCATCGCAAGATGGTGAGC
F3'I	GGTGCTGCAGAGCGGCATTCGCAAGATGGTGAGC
F3'K	GGTGCTGCAGAGCGGCAAACGCAAGATGGTGAGC
F3'L	GGTGCTGCAGAGCGGCCTGCGCAAGATGGTGAGC
F3'M	GGTGCTGCAGAGCGGCATGCGCAAGATGGTGAGC
F3'N	GGTGCTGCAGAGCGGCAACCGCAAGATGGTGAGC
F3'P	GGTGCTGCAGAGCGGCCCGCGCAAGATGGTGAGC
F3'Q	GGTGCTGCAGAGCGGCCAGCGCAAGATGGTGAGC
F3'R	GGTGCTGCAGAGCGGCCGTCGCAAGATGGTGAGC
F3'S	GGTGCTGCAGAGCGGCTCGCGCAAGATGGTGAGC
F3'T	GGTGCTGCAGAGCGGCACCCGCAAGATGGTGAGC
F3'V	GGTGCTGCAGAGCGGCGTGCGCAAGATGGTGAGC
F3'W	GGTGCTGCAGAGCGGCTGGCGCAAGATGGTGAGC
F3'Y	GGTGCTGCAGAGCGGCTATCGCAAGATGGTGAGC

Appendix 2.11 – Scales for quantification of structural properties. Side chain volume was derived from the partial molar volume of amino acids reported in Lee *et al.* 2008. Scales of hydrophobicity and secondary structure propensities were obtained from Kyte & Doolittle, 1982, and Chou & Fasman, 1978 respectively.

Residue	Side chain volume ($Å^3$)	Hydrophobicity	α-helix propensity	β-sheet propensity
A	28.0	1.8	1.42	0.83
C	49.5	2.5	0.70	1.19
D	52.8	-3.5	1.01	0.54
E	79.5	-3.5	1.51	0.37
F	133.2	2.8	1.13	1.38
G	0	-0.4	0.57	0.75
H	95.0	-3.2	1.00	0.87
I	106.3	4.5	1.08	1.60
K	116.8	-3.9	1.16	0.74
L	109.0	3.8	1.21	1.30
M	104.7	1.9	1.45	1.05
N	56.7	-3.5	0.67	0.89
P	59.3	-1.6	0.57	0.55
Q	84.7	-3.5	1.11	1.10
R	112.3	-4.5	0.98	0.93
S	28.5	-0.8	0.77	0.75
T	55.2	-0.7	0.83	1.19
V	79.5	4.2	1.06	1.70
W	170.2	-0.9	1.08	1.37
Y	137.0	-1.3	0.69	1.47

Appendix 6.1 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P5 variants.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	0.59 ± 0.17	0.26 ± 0.13	0.90 ± 0.16	0.85 ± 0.30
R	0.97 ± 0.26	0.68 ± 0.46	1.24 ± 0.21	1.55 ± 0.30
H	0.87 ± 0.41	0.65 ± 0.35	1.55 ± 0.16	1.06 ± 0.30
D	0.27 ± 0.09	0.12 ± 0.06	0.29 ± 0.04	0.18 ± 0.03
E	0.46 ± 0.15	0.12 ± 0.08	0.47 ± 0.05	0.24 ± 0.09
N	0.95 ± 0.23	0.59 ± 0.37	0.99 ± 0.14	0.89 ± 0.26
Q	1.01 ± 0.40	0.88 ± 0.66	0.90 ± 0.07	0.92 ± 0.26
P	0.82 ± 0.19	0.52 ± 0.36	0.87 ± 0.06	1.07 ± 0.25
F	0.75 ± 0.26	0.52 ± 0.32	1.53 ± 0.39	1.69 ± 0.42
W	0.40 ± 0.12	0.17 ± 0.09	0.91 ± 0.07	0.72 ± 0.16
Y	0.78 ± 0.31	0.49 ± 0.19	1.12 ± 0.24	1.43 ± 0.25
T	0.67 ± 0.18	0.89 ± 0.53	1.55 ± 0.14	0.96 ± 0.22
S	1.00 ± 0.32	1.00 ± 0.64	1.00 ± 0.04	1.00 ± 0.25
C	0.88 ± 0.17	0.91 ± 0.22	1.44 ± 0.28	1.26 ± 0.11
G	1.27 ± 0.29	1.10 ± 0.59	0.92 ± 0.14	0.97 ± 0.22
A	0.72 ± 0.37	0.32 ± 0.42	0.87 ± 0.05	0.81 ± 0.28
V	1.27 ± 0.02	1.56 ± 0.10	1.94 ± 0.10	1.81 ± 0.11
I	1.27 ± 0.33	0.34 ± 0.29	1.17 ± 0.17	2.11 ± 0.44
L	1.18 ± 0.28	0.90 ± 0.52	1.27 ± 0.18	1.33 ± 0.30
M	0.79 ± 0.21	1.02 ± 0.60	1.27 ± 0.09	1.45 ± 0.28

Appendix 6.2 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P4 variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	ND	ND	ND	ND
R	ND	ND	ND	ND
H	ND	ND	ND	ND
D	ND	ND	ND	ND
E	ND	ND	ND	ND
N	0.08 ± 0.04	ND	0.12 ± 0.06	ND
Q	ND	ND	ND	ND
P	0.13 ± 0.05	0.37 ± 0.09	0.52 ± 0.13	1.23 ± 0.26
F	ND	ND	0.05 ± 0.01	0.58 ± 0.14
W	ND	ND	0.15 ± 0.02	0.29 ± 0.07
Y	ND	ND	ND	0.28 ± 0.05
T	0.65 ± 0.17	0.46 ± 0.16	0.76 ± 0.18	0.48 ± 0.10
S	0.57 ± 0.12	0.52 ± 0.15	0.52 ± 0.11	0.47 ± 0.10
C	0.95 ± 0.37	0.81 ± 0.33	1.27 ± 0.27	1.01 ± 0.19
G	0.15 ± 0.03	0.24 ± 0.08	0.16 ± 0.04	0.46 ± 0.11
A	1.00 ± 0.25	1.00 ± 0.33	1.00 ± 0.20	1.00 ± 0.13
V	0.81 ± 0.05	0.44 ± 0.04	1.49 ± 0.31	0.67 ± 0.16
I	0.22 ± 0.04	0.11 ± 0.03	0.52 ± 0.08	0.30 ± 0.05
L	0.10 ± 0.02	ND	0.10 ± 0.05	0.19 ± 0.04
M	ND	ND	0.19 ± 0.05	0.07 ± 0.03

Appendix 6.3 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P3 variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	0.92 ± 0.07	1.03 ± 0.13	0.98 ± 0.20	1.69 ± 0.10
R	1.07 ± 0.28	1.09 ± 0.28	1.03 ± 0.44	1.34 ± 0.34
H	0.36 ± 0.05	0.48 ± 0.11	0.30 ± 0.04	0.54 ± 0.10
D	0.15 ± 0.02	0.12 ± 0.03	0.10 ± 0.02	0.12 ± 0.01
E	0.17 ± 0.02	0.25 ± 0.06	0.11 ± 0.02	0.17 ± 0.02
N	0.59 ± 0.11	0.36 ± 0.12	0.53 ± 0.15	0.52 ± 0.12
Q	0.69 ± 0.16	0.60 ± 0.17	0.50 ± 0.08	0.46 ± 0.05
P	ND	ND	ND	ND
F	0.38 ± 0.05	0.41 ± 0.14	0.39 ± 0.07	0.62 ± 0.09
W	0.41 ± 0.13	0.33 ± 0.09	0.48 ± 0.08	0.42 ± 0.14
Y	0.39 ± 0.06	0.32 ± 0.22	0.35 ± 0.05	0.37 ± 0.03
T	1.09 ± 0.12	0.79 ± 0.22	0.59 ± 0.07	0.84 ± 0.25
S	0.95 ± 0.26	0.86 ± 0.25	0.58 ± 0.11	0.82 ± 0.06
C	0.62 ± 0.09	0.70 ± 0.16	0.53 ± 0.08	0.86 ± 0.08
G	0.13 ± 0.02	0.10 ± 0.02	0.09 ± 0.01	0.20 ± 0.01
A	0.51 ± 0.08	0.81 ± 0.19	0.34 ± 0.06	0.59 ± 0.06
V	1.00 ± 0.20	1.00 ± 0.24	1.00 ± 0.19	1.00 ± 0.10
I	0.74 ± 0.15	0.53 ± 0.22	0.82 ± 0.05	0.73 ± 0.27
L	0.52 ± 0.07	0.69 ± 0.13	0.69 ± 0.17	0.57 ± 0.05
M	0.51 ± 0.24	0.68 ± 0.20	0.83 ± 0.34	0.64 ± 0.23

Appendix 6.4 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P2 variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	ND	ND	ND	ND
R	ND	ND	ND	ND
H	ND	ND	ND	ND
D	ND	ND	ND	ND
E	ND	ND	ND	ND
N	ND	ND	ND	ND
Q	ND	ND	ND	ND
P	ND	ND	ND	ND
F	ND	0.20 ± 0.09	0.30 ± 0.08	0.06 ± 0.02
W	ND	ND	ND	ND
Y	ND	ND	ND	ND
T	ND	ND	ND	ND
S	ND	ND	ND	ND
C	0.03 ± 0.01	0.58 ± 0.13	0.16 ± 0.02	0.17 ± 0.02
G	ND	ND	ND	ND
A	ND	ND	ND	ND
V	ND	0.08 ± 0.03	0.03 ± 0.01	0.07 ± 0.02
I	0.08 ± 0.02	0.06 ± 0.02	0.06 ± 0.02	0.08 ± 0.03
L	1.00 ± 0.28	1.00 ± 0.40	1.00 ± 0.28	1.00 ± 0.35
M	0.21 ± 0.04	0.69 ± 0.34	0.79 ± 0.34	0.22 ± 0.10

Appendix 6.5 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P1 variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	ND	ND	ND	ND
R	ND	ND	ND	ND
H	0.26 ± 0.08	0.47 ± 0.08	0.19 ± 0.03	0.25 ± 0.12
D	ND	ND	ND	ND
E	ND	ND	ND	ND
N	ND	ND	ND	ND
Q	1.00 ± 0.06	1.00 ± 0.13	1.00 ± 0.08	1.00 ± 0.05
P	ND	ND	ND	ND
F	ND	ND	ND	ND
W	ND	ND	ND	ND
Y	ND	ND	ND	ND
T	ND	ND	ND	ND
S	ND	ND	ND	ND
C	0.02 ± 0.00	0.07 ± 0.02	0.02 ± 0.01	0.06 ± 0.02
G	ND	ND	ND	ND
A	ND	ND	ND	ND
V	ND	ND	ND	ND
I	ND	ND	ND	ND
L	ND	ND	ND	ND
M	0.06 ± 0.09	0.30 ± 0.07	0.12 ± 0.04	0.04 ± 0.03

Appendix 6.6 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P1' variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	0.07 ± 0.02	0.05 ± 0.03	0.05 ± 0.01	0.14 ± 0.05
R	0.15 ± 0.02	0.08 ± 0.04	0.09 ± 0.05	0.13 ± 0.10
H	0.08 ± 0.02	0.26 ± 0.11	0.16 ± 0.06	0.26 ± 0.09
D	0.03 ± 0.01	0.02 ± 0.02	0.01 ± 0.01	0.05 ± 0.03
E	ND	ND	0.01 ± 0.01	0.03 ± 0.04
N	0.61 ± 0.09	0.18 ± 0.17	0.28 ± 0.12	0.58 ± 0.21
Q	0.01 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.10 ± 0.04
P	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.04 ± 0.01
F	0.04 ± 0.01	0.50 ± 0.21	0.06 ± 0.06	0.46 ± 0.12
W	0.01 ± 0.01	0.08 ± 0.03	0.08 ± 0.02	0.12 ± 0.03
Y	0.01 ± 0.01	0.22 ± 0.10	0.08 ± 0.02	0.27 ± 0.06
T	0.25 ± 0.06	0.41 ± 0.23	0.23 ± 0.07	0.43 ± 0.15
S	1.00 ± 0.19	1.00 ± 0.57	1.00 ± 0.19	1.00 ± 0.32
C	0.54 ± 0.09	1.74 ± 0.41	1.01 ± 0.19	1.03 ± 0.09
G	0.80 ± 0.14	1.00 ± 0.48	0.91 ± 0.13	1.16 ± 0.33
A	0.75 ± 0.13	0.64 ± 0.29	1.05 ± 0.23	0.85 ± 0.25
V	0.01 ± 0.02	0.37 ± 0.19	0.04 ± 0.01	0.37 ± 0.14
I	0.00 ± 0.01	0.04 ± 0.02	0.01 ± 0.01	0.04 ± 0.02
L	0.01 ± 0.01	0.10 ± 0.12	0.07 ± 0.06	0.18 ± 0.11
M	0.08 ± 0.03	0.51 ± 0.27	0.21 ± 0.07	0.59 ± 0.21

Appendix 6.7 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P2' variants. 'ND' stands for non-detectable cleavage.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	0.45 ± 0.12	0.55 ± 0.13	0.49 ± 0.05	0.70 ± 0.17
R	0.46 ± 0.16	0.49 ± 0.11	0.42 ± 0.05	0.87 ± 0.16
H	0.38 ± 0.14	0.70 ± 0.14	0.42 ± 0.07	0.65 ± 0.16
D	0.34 ± 0.09	0.48 ± 0.14	0.38 ± 0.04	0.62 ± 0.20
E	0,20 ± 0.07	0.42 ± 0.09	0.29 ± 0.07	0.52 ± 0.16
N	0.59 ± 0.18	0.62 ± 0.14	0.67 ± 0.09	0.93 ± 0.15
Q	0.42 ± 0.16	0.33 ± 0.10	0.46 ± 0.08	0.63 ± 0.12
P	ND	ND	ND	ND
F	0.34 ± 0.12	0.49 ± 0.13	0.48 ± 0.11	0.64 ± 0.12
W	0.22 ± 0.06	0.43 ± 0.17	0.27 ± 0.04	0.56 ± 0.09
Y	0.30 ± 0.09	0.53 ± 0.14	0.40 ± 0.04	0.72 ± 0.13
T	0.52 ± 0.14	0.70 ± 0.23	0.57 ± 0.16	0.82 ± 0.15
S	0.68 ± 0.27	0.81 ± 0.16	1.04 ± 0.11	1.00 ± 0.15
C	0.38 ± 0.06	1.33 ± 0.33	0.61 ± 0.10	0.83 ± 0.06
G	1.00 ± 0.37	1.00 ± 0.27	1.00 ± 0.12	1.00 ± 0.21
A	0.62 ± 0.18	1.03 ± 0.27	0.81 ± 0.16	0.88 ± 0.21
V	0.28 ± 0.09	0.69 ± 0.15	0.46 ± 0.08	0.86 ± 0.17
I	0.17 ± 0.07	0.46 ± 0.09	0.33 ± 0.05	0.60 ± 0.12
L	0.25 ± 0.09	0.45 ± 0.10	0.32 ± 0.07	0.50 ± 0.12
M	0.44 ± 0.12	0.31 ± 0.07	0.35 ± 0.15	0.55 ± 0.34

Appendix 6.8 – The relative activity of $3CL^{pro}$ of HCoV-NL63, HCoV-OC43, SARS-CoV and IBV on P3' variants.

Residue	HCoV-NL63	HCoV-OC43	SARS-CoV	IBV
K	0.80 ± 0.28	0.56 ± 0.20	0.84 ± 0.30	0.89 ± 0.30
R	1.22 ± 0.11	1.11 ± 0.15	0.91 ± 0.05	1.14 ± 0.09
H	0.88 ± 0.36	0.68 ± 0.34	0.80 ± 0.28	0.80 ± 0.28
D	0.53 ± 0.28	0.34 ± 0.17	0.35 ± 0.08	0.53 ± 0.17
E	0.73 ± 0.23	0.42 ± 0.17	0.55 ± 0.12	0.65 ± 0.16
N	1.08 ± 0.35	0.49 ± 0.16	0.66 ± 0.15	0.79 ± 0.19
Q	1.16 ± 0.35	0.61 ± 0.18	0.79 ± 0.21	0.78 ± 0.22
P	0.78 ± 0.39	0.56 ± 0.30	0.42 ± 0.16	0.99 ± 0.45
F	1.00 ± 0.39	1.00 ± 0.39	1.00 ± 0.31	1.00 ± 0.28
W	0.57 ± 0.20	0.41 ± 0.23	0.46 ± 0.12	0.70 ± 0.18
Y	1.10 ± 0.34	0.84 ± 0.26	1.07 ± 0.26	0.97 ± 0.23
T	0.75 ± 0.25	0.47 ± 0.24	0.82 ± 0.28	0.65 ± 0.19
S	1.22 ± 0.38	0.65 ± 0.20	0.82 ± 0.20	0.91 ± 0.20
C	0.88 ± 0.13	0.99 ± 0.22	0.72 ± 0.12	0.94 ± 0.09
G	0.92 ± 0.34	0.41 ± 0.13	0.44 ± 0.12	0.80 ± 0.17
A	0.91 ± 0.32	0.39 ± 0.14	0.45 ± 0.10	0.59 ± 0.19
V	1.21 ± 0.36	0.58 ± 0.20	0.81 ± 0.21	0.80 ± 0.21
I	1.08 ± 0.35	0.65 ± 0.20	0.81 ± 0.21	0.82 ± 0.17
L	0.91 ± 0.29	0.59 ± 0.17	0.65 ± 0.18	0.81 ± 0.17
M	0.73 ± 0.23	0.37 ± 0.14	0.66 ± 0.24	0.76 ± 0.31

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

Akaji, K., Konno, H., Onozuka, M., Makino, A., Saito, H. & Nosaka, K. (2008). Evaluation of peptide-aldehyde inhibitors using R188I mutant of SARS 3CL protease as a proteolysis-resistant mutant. *Bioorg Med Chem* **16**, 9400-9408.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.

Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry* **43**, 4906-4912.

Black, W. C. Peptidomimetic inhibitors of cathepsin K. *Curr Top Med Chem* **10**, 745-751.

Brisson, J. R., Carey, P. R. & Storer, A. C. (1986). Benzoylamidoacetonitrile Is Bound as a Thioimidate in the Active-Site of Papain. *Journal of Biological Chemistry* **261**, 9087-9089.

Chen, L., Gui, C., Luo, X., Yang, Q., Gunther, S., Scandella, E., Drosten, C., Bai, D., He, X., Ludewig, B., Chen, J., Luo, H., Yang, Y., Zou, J., Thiel, V., Chen, K., Shen, J., Shen, X. & Jiang, H. (2005a). Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J Virol* **79**, 7095-7103.

Chen, S., Chen, L. L., Luo, H. B., Sun, T., Chen, J., Ye, F., Cai, J. H., Shen, J. K., Shen, X. & Jiang, H. L. (2005b). Enzymatic activity characterization of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer technique. *Acta Pharmacol Sin* **26**, 99-106.

Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* **47**, 45-148.

Chu, L. H., Choy, W. Y., Tsai, S. N., Rao, Z. & Ngai, S. M. (2006). Rapid peptide-based screening on the substrate specificity of severe acute respiratory syndrome (SARS) coronavirus 3C-like protease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci* **15**, 699-709.

Cinatl, J., Jr., Michaelis, M., Hoever, G., Preiser, W. & Doerr, H. W. (2005). Development of antiviral therapy for severe acute respiratory syndrome. *Antiviral Res* **66**, 81-97.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

Fan, K., Ma, L., Han, X., Liang, H., Wei, P., Liu, Y. & Lai, L. (2005). The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochem Biophys Res Commun* **329**, 934-940.

Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J. & Lai, L. (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem* **279**, 1637-1642.

Frizler, M., Stirnberg, M., Sisay, M. T. & Gutschow, M. Development of nitrile-based peptidic inhibitors of cysteine cathepsins. *Curr Top Med Chem* **10**, 294-322.

Goetz, D. H., Choe, Y., Hansell, E., Chen, Y. T., McDowell, M., Jonsson, C. B., Roush, W. R., McKerrow, J. & Craik, C. S. (2007). Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. *Biochemistry* **46**, 8744-8752.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

Hauff, K., Zamzow, C., Law, W. J., De Melo, J., Kennedy, K. & Los, M. (2005). Peptide-based approaches to treat asthma, arthritis, other autoimmune diseases and pathologies of the central nervous system. *Arch Immunol Ther Exp (Warsz)* **53**, 308-320.

Hegyi, A. & Ziebuhr, J. (2002). Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol* **83**, 595-599.

Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., Shr, H. L., Chang, G. G., Wang, A. H. & Liang, P. H. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem* **280**, 31257-31266.

Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* **43**, 4568-4574.

Kiemer, L., Lund, O., Brunak, S. & Blom, N. (2004). Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC Bioinformatics* **5**, 72.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J. & Anderson, L. J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953-1966.

Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D., Laman, J. D., de Jong, T., van Doornum, G., Lim, W., Ling, A. E., Chan, P. K., Tam, J. S., Zambon, M. C., Gopal, R., Drosten, C., van der Werf, S., Escriou, N., Manuguerra, J. C., Stohr, K., Peiris, J. S. & Osterhaus, A. D. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* **362**, 263-270.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Lee, C. C., Kuo, C. J., Ko, T. P., Hsu, M. F., Tsui, Y. C., Chang, S. C., Yang, S., Chen, S. J., Chen, H. C., Hsu, M. C., Shih, S. R., Liang, P. H. & Wang, A. H. (2009). Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *J Biol Chem* **284**, 7646-7655.

Lee, S., Tikhomirova, A., Shalvardjian, N. & Chalikian, T. V. (2008). Partial molar volumes and adiabatic compressibilities of unfolded protein states. *Biophys Chem* **134**, 185-199.

Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J Mol Biol* **353**, 1137-1151.

Li, S. Y., Chen, C., Zhang, H. Q., Guo, H. Y., Wang, H., Wang, L., Zhang, X., Hua, S. N., Yu, J., Xiao, P. G., Li, R. S. & Tan, X. (2005). Identification of natural compounds with antiviral activities against SARS-associated coronavirus. *Antiviral Res* **67**, 18-23.

Lin, P. Y., Chou, C. Y., Chang, H. C., Hsu, W. C. & Chang, G. G. (2008). Correlation between dissociation and catalysis of SARS-CoV main protease. *Arch Biochem Biophys* **472**, 34-42.

Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., Pandoh, P. K., Petrescu, A. S., Robertson, A. G., Schein, J. E., Siddiqui, A., Smailus, D. E., Stott, J. M., Yang, G. S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T. F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G. A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R. C., Krajden, M., Petric, M., Skowronski, D. M., Upton, C. & Roper, R. L. (2003). The Genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404.

Moon, J. B., Coleman, R. S. & Hanzlik, R. P. (1986). Reversible Covalent Inhibition of Papain by a Peptide Nitrile - C-13 Nmr Evidence for a Thioimidate Ester Adduct. *Journal of the American Chemical Society* **108**, 1350-1351.

Phakthanakanok, K., Ratanakhanokchai, K., Kyu, K. L., Sompornpisut, P., Watts, A. & Pinitglang, S. (2009). A computational analysis of SARS cysteine proteinase-octapeptide substrate interaction: implication for structure and active site binding mechanism. *BMC Bioinformatics* **10 Suppl 1**, S48.

Randolph, J. T. & DeGoey, D. A. (2004). Peptidomimetic inhibitors of HIV protease. *Curr Top Med Chem* **4**, 1079-1095.

Regnier, T., Sarma, D., Hidaka, K., Bacha, U., Freire, E., Hayashi, Y. & Kiso, Y. (2009). New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CL(pro) inhibitors. *Bioorg Med Chem Lett* **19**, 2722-2727.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D., Drosten, C., Pallansch, M. A., Anderson, L. J. & Bellini, W. J. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394-1399.

Shao, Y. M., Yang, W. B., Kuo, T. H., Tsai, K. C., Lin, C. H., Yang, A. S., Liang, P. H. & Wong, C. H. (2008). Design, synthesis, and evaluation of trifluoromethyl ketones as inhibitors of SARS-CoV 3CL protease. *Bioorg Med Chem* **16**, 4652-4660.

Shi, J., Sivaraman, J. & Song, J. (2008). Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol* **82**, 4620-4629.

Solowiej, J., Thomson, J. A., Ryan, K., Luo, C., He, M., Lou, J. & Murray, B. W. (2008). Steady-state and pre-steady-state kinetic evaluation of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro cysteine protease: development of an ion-pair model for catalysis. *Biochemistry* **47**, 2617-2630.

Steverding, D., Caffrey, C. R. & Sajid, M. (2006). Cysteine proteinase inhibitors as therapy for parasitic diseases: advances in inhibitor design. *Mini Rev Med Chem* **6**, 1025-1032.

Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* **84**, 2305-2315.

Tse, G. M., To, K. F., Chan, P. K., Lo, A. W., Ng, K. C., Wu, A., Lee, N., Wong, H. C., Mak, S. M., Chan, K. F., Hui, D. S., Sung, J. J. & Ng, H. K. (2004). Pulmonary pathological features in coronavirus associated severe acute respiratory syndrome (SARS). *J Clin Pathol* **57**, 260-265.

Tsui, S. K., Chim, S. S. & Lo, Y. M. (2003). Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N Engl J Med* **349**, 187-188.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552.

Wu, C. Y., Jan, J. T., Ma, S. H., Kuo, C. J., Juan, H. F., Cheng, Y. S., Hsu, H. H., Huang, H. C., Wu, D., Brik, A., Liang, F. S., Liu, R. S., Fang, J. M., Chen, S. T., Liang, P. H. & Wong, C. H. (2004). Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc Natl Acad Sci U S A* **101**, 10012-10017.

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X. C., Liao, M., Bartlam, M. & Rao, Z. (2008). Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J Virol* **82**, 2515-2527.

Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M. & Rao, Z. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol* **3**, e324.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* **100**, 13190-13195.

Yin, J., Niu, C., Cherney, M. M., Zhang, J., Huitema, C., Eltis, L. D., Vederas, J. C. & James, M. N. (2007). A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. *J Mol Biol* **371**, 1060-1074.

Zhao, Q., Li, S., Xue, F., Zou, Y., Chen, C., Bartlam, M. & Rao, Z. (2008). Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. *J Virol* **82**, 8647-8655.