

2015-08-24

# Structural Studies of Human Norovirus Protease Complexes with RNA and Peptides

Sikandar, Asfandyar

---

Sikandar, A. (2015). Structural Studies of Human Norovirus Protease Complexes with RNA and Peptides (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. doi:10.11575/PRISM/27903

<http://hdl.handle.net/11023/2403>

*Downloaded from PRISM Repository, University of Calgary*

UNIVERSITY OF CALGARY

Structural Studies of Human Norovirus Protease Complexes with RNA and Peptides

by

Asfandyar Sikandar

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTERS OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

August, 2015

© Asfandyar Sikandar 2015

## Abstract

Noroviruses are single-stranded RNA viruses. They encode a protease that cleaves a viral polyprotein at specific sites to produce mature viral proteins. In addition, the protease also binds to viral RNA, and thus is thought to regulate viral replication. However, to date no structural information is available for protease-substrate complexes that might explain the interactions made by peptide residues P'-side of cleavage junctions or RNA. Here I report the work carried out to characterize these interactions in human norovirus protease using X-ray crystallography. The protease was successfully expressed, purified and the crystallization conditions were optimized to grow crystals for structure determination. Unfortunately, RNA and peptide electron density were not observed in co-crystal structures. The packing of protease molecules in one of the crystal forms shows the interaction of protease C-terminal residues with the peptide-binding groove of a neighboring molecule in the crystal, thereby providing the view of a protease-product complex.

## Acknowledgments

It would not have been possible to work on this project without the support and help of people around me.

First, I would like to thank my family for their unconditional support. Then there are my friends Saad Rajput, Saad Amjad, Mujeeb Mufti and Alexander Carleton, who made graduate life fun and were there for me whenever I needed them.

This thesis would not have been possible without the help of our collaborators: Jared May, Prasanth Viswanathan and Brent Korba from the University of Georgetown, USA. They expressed both the wild type and mutant protease for this project. I would like to thank all the current members of Dr. Ng's lab- especially Dr. Md Munan Shaik for his help in refinement and analysis of X-ray diffraction data.

I would also like to extend my gratitude to my committee members, Dr. Zimmerly and Dr. Tieleman, for their suggestions and sound advice.

Last, but certainly not the least, my supervisor Dr. Kenneth Ng for giving me the opportunity to work in his lab. I cannot thank him enough for his patience and trust in my abilities.

## Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Acknowledgments</b> .....	<b>iii</b>
<b>Table of Contents</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>vi</b>
<b>List of Figures</b> .....	<b>vii</b>
<b>List of Abbreviations</b> .....	<b>ix</b>
<b>Physical Units Abbreviation</b> .....	<b>xi</b>
<b>Amino Acid Abbreviations</b> .....	<b>xii</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>a. Viral taxonomy</b> .....	<b>1</b>
<b>b. Epidemiology and pathophysiology</b> .....	<b>2</b>
<b>c. Virus genome organization</b> .....	<b>3</b>
<b>d. NV life cycle</b> .....	<b>4</b>
<b>e. NV Pro</b> .....	<b>7</b>
I. Structural features of NV Pro.....	7
II. Catalytic site and mechanism of proteolysis .....	9
III. Peptide substrate recognition and proteolytic processing order .....	11
<b>f. NV Pro-RNA interaction</b> .....	<b>16</b>
<b>Chapter 2: Objectives of the Study</b> .....	<b>18</b>
<b>a. Interaction of HuNV Pro with RNA oligonucleotides</b> .....	<b>18</b>
<b>b. Interaction of HuNV Pro with P' residues</b> .....	<b>20</b>
<b>Chapter 3: Materials and Methods</b> .....	<b>21</b>
<b>a. Protein expression</b> .....	<b>21</b>
<b>b. Ion exchange chromatography</b> .....	<b>21</b>
<b>c. Gel-filtration chromatography</b> .....	<b>21</b>
<b>d. Dialysis and concentration</b> .....	<b>22</b>
<b>e. RNA and peptide preparation</b> .....	<b>22</b>
<b>f. Crystallization methodology</b> .....	<b>22</b>
I. Sparse matrix screening .....	23
II. Crystal optimization.....	24
III. Soaking .....	25
IV. Crystal harvesting .....	25
<b>g. X-ray data collection and structure determination</b> .....	<b>26</b>
<b>Chapter 4: Results and Discussion</b> .....	<b>27</b>
<b>a. HuNV Pro-RNA interaction</b> .....	<b>27</b>
I. Expression and purification .....	27
II. Screening and hit optimization .....	29
III. Optimization of purification protocol.....	32
IV. Optimization of Thiocyanate condition .....	34

V. Soaking .....	36
VI. X-ray data collection and structure determination.....	37
VII. Structural analysis and discussion of AS59 .....	40
<b>b. HuNV Pro-peptide interaction .....</b>	<b>44</b>
I. Expression and purification .....	44
II. Screening and optimization .....	45
III. X-ray data collection and structure determination .....	47
IV. Structural analysis and discussion of AS106.....	51
<b>Chapter 4: Conclusion and Future Work.....</b>	<b>56</b>
<b>Bibliography .....</b>	<b>59</b>
<b>Appendix.....</b>	<b>65</b>

## List of Tables

TABLE 1: EFFECT OF RNA ON HUNV PRO ACTIVITY..	17
TABLE 2: RNA OLIGONUCLEOTIDE INFORMATION AND SEQUENCE..	22
TABLE 3: SYNTHETIC PEPTIDE INFORMATION AND SEQUENCE..	22
TABLE 4: CRYSTALLIZATION HITS IDENTIFIED FROM SCREENS PERFORMED IN THE PRESENCE OF HUNV PRO AND RNA OLIGONUCLEOTIDES.....	29
TABLE 5: SUMMARY OF CHANGES MADE TO THE ORIGINAL CRYSTALLIZATION CONDITION IDENTIFIED TO SLOW DOWN THE RATE OF CRYSTALLIZATION. ....	32
TABLE 6: SUMMARY OF SALT CONDITIONS THAT YIELDED CRYSTALS.....	35
TABLE 7: SUMMARY OF CONDITIONS USED FOR SOAKING EXPERIMENTS.....	36
TABLE 8: X-RAY DIFFRACTION DATA SETS COLLECTED FOR CO-CRYSTALS GROWN AND/OR SOAKED IN DIFFERENT CONDITIONS.....	38
TABLE 9: CRYSTALLOGRAPHIC DATA COLLECTION AND MODEL REFINEMENT STATISTS FOR AS59.....	39
TABLE 10: CRYSTALLIZATION HITS IDENTIFIED FROM SCREENS PERFORMED IN THE PRESENCE OF MUTANT HUNV PRO AND DIFFERENT PEPTIDES. ....	46
TABLE 11: SUMMARY OF THE CHANGES MADE TO THE CRYSTALLIZATION HITS TO GROW DIFFRACTION QUALITY CRYSTALS. ....	47
TABLE 12: X-RAY DATA SETS COLLECTED FOR HUNV PRO-PEPTIDE CRYSTALS GROWN AND/OR SOAKED IN DIFFERENT CONDITIONS. ....	49
TABLE 13: CRYSTALLOGRAPHIC DATA COLLECTION AND MODEL REFINEMENT STATISTS FOR AS106.....	50

## List of Figures

FIGURE 1: CLASSIFICATION OF NOROVIRUSES INTO 5 GENOGROUPS (GI-V) AND 35 GENOTYPES BASED ON SEQUENCE DIVERSITY IN THE COMPLETE CAPSID PROTEIN .....	2
FIGURE 2: STRUCTURE OF NOROVIRUS CAPSID AND NOROVIRUS GENOME ORGANIZATION .....	4
FIGURE 3: LIFE CYCLE OF NOROVIRUS .....	6
FIGURE 4: STRUCTURE OF NV PRO .....	8
FIGURE 5: STRUCTURAL COMPARISON OF THE NV PRO (PDB-ID: 1QWS) WITH PV (PDB-ID: 1L1N) AND HRV PROTEASES (PDB-ID: 1CQQ) .....	9
FIGURE 6: NV PRO CATALYTIC MECHANISM .....	10
FIGURE 7: THE PROCESSING OF ORF1 POLYPROTEIN .....	14
FIGURE 8: RIBBON REPRESENTATION OF HUNV PRO-PRODUCT COMPLEX STRUCTURE .....	14
FIGURE 9: SURFACE REPRESENTATION OF NV PRO (PDB ID: 4IN2) WITH COLOR-CODED ACTIVE SITE AND SUBSTRATE BINDING POCKETS .....	15
FIGURE 10: SURFACE REPRESENTATION OF PEPTIDE-INDUCED CONFORMATIONAL CHANGES IN S2 AND S4 POCKETS OF HUNV PRO.. .....	15
FIGURE 11: RNA BINDING BY HUNV PRO AND ITS AFFECT ON ENZYME KINETICS .....	17
FIGURE 12: PUTATIVE RNA BINDING SITES ON NV PRO BOUND TO PEPTIDE SUBSTRATE .....	19
FIGURE 13: ION EXCHANGE CHROMATOGRAM AND SDS-PAGE PROFILE OF HUNV PRO . .....	28
FIGURE 14: OPTIMIZATION OF THIOCYANATE CONDITION .....	31
FIGURE 15: X-RAY DIFFRACTION PATTERN OF CRYSTAL WITH AND WITHOUT DEFECTS .....	31
FIGURE 16: HUNV PRO GEL-FILTRATION ELUTION AND SDS-PAGE PROFILE.....	33
FIGURE 17:THE CHANGE IN SIZE OF SEED CRYSTAL TRANSFERRED TO KCL CONDITION. ....	35
FIGURE 18: THE DIMERIC ARRANGEMENT OBSERVED IN AS59 . ....	41
FIGURE 19: COMPARISON OF DIMER STRUCTURES AND INTERFACE OF NV PRO DIMERS WITH AS59. ....	42
FIGURE 20: OVERLAY OF AS59 MONOMER A (RED) AND B (GREEN) SHOWN AS CARTOONS.....	42
FIGURE 21: STRUCTURAL COMPARISON OF AS59 WITH OTHER HUNV PROTEASES.....	43
FIGURE 22:STRUCTURAL COMPARISON OF HUNV PRO CRYSTAL GROWN IN THE PRESENCE (GREEN) AND ABSENCE OF RNA OLIGONUCLEOTIDES (RED). .....	43
FIGURE 23: MUTANT NV PRO GEL-FILTRATION ELUTION AND SDS-PAGE PROFILE (TOP). ....	44
FIGURE 24: HUNV PRO-PEPTIDE CRYSTAL BEFORE (LEFT) AND AFTER OPTIMIZATION (RIGHT).....	46
FIGURE 25: CRYSTAL PACKING OF AS106 IN THE P <sub>2</sub> <sub>1</sub> SPACE GROUP, WITH FOUR MOLECULES IN THE ASYMMETRIC UNIT .....	52
FIGURE 26:OVERLAY OF AS106 MONOMERS REPRESENTED AS RIBBON DIAGRAM.....	53

FIGURE 27: STRUCTURAL COMPARISON OF AS106 MONOMER WITH OTHER NV PROTEASES REPRESENTED AS RIBBON DIAGRAM. ....	53
FIGURE 28: STRUCTURAL COMPARISON OF C-TERMINAL TAIL INTERACTIONS OBSERVED IN HUNV PRO-PEPTIDE COMPLEXES .....	54
FIGURE 29: INTERACTION OF MONOMER C AND D .....	55
FIGURE 30: SCHEMATIC REPRESENTATION OF SCHECHTER AND BERGER NOMENCLATURE .....	65
FIGURE 31: THE PREDICTED BINDING ROUTES FOR P' RESIDUES IN NOROVIRUS PROTEASE .....	66
FIGURE 32: STRUCTURAL COMPARISON OF HUNV PRO DIMERS OBTAINED IN THE PRESENCE OF RNA (RED) AND RNA/PEPTIDE (GREEN).....	68

## List of Abbreviations

CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]-1- propanesulfonate
CDC	Center for disease control and prevention
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FCV	Feline calicivirus
HCA	Hepatitis A virus
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HRV	Human rhinovirus
HuNV	Human norovirus
HBGAs	Histo-blood group antigens
MOPS	3-morpholinopropane-1-sulfonic acid
MNV	Murine norovirus
mRNA	Messenger ribonucleic acid
NV	Norovirus
ORF	Open reading frame
PV	Poliovirus
PEG	Polyethylene glycol
PEG MME	Polyethylene glycol monomethyl ether
PDB	Protein data bank
PDB ID	Protein data bank identification code
Pro	Protease
Pol	Polymerase

ProPol	Protease-Polymerase
RNA	Ribonucleic acid
RdRp	RNA dependent RNA polymerase
RMSD	Root mean square deviation
SaV	Sapovirus
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
VPg	Viral Protein genome-linked
VP1/2	Viral capsid protein

## Physical Units Abbreviation

°	Degree
°C	Degree centigrade
Å	Angstrom
g	Gram
K	Kelvin
kDa	kilo Dalton
M	Molar
mg	Milligram
min	Minute
ml	Milliliter
mM	Millimolar
μL	Microliter

## Amino Acid Abbreviations

Ala /A	Alanine
Arg /R	Arginine
Asn /N	Asparagine
Asp /D	Aspartate
Cys /C	Cysteine
Glu /E	Glutamine
Gln /Q	Glycine
His /H	Histidine
Ile /I	Isoleucine
Leu /L	Leucine
Lys /K	Lysine
Met /M	Methionine
Phe /F	Phenylalanine
Pro /P	Proline
Ser /S	Serine
Thr /T	Threonine
Trp /W	Tryptophan
Tyr /Y	Tyrosine
Val /V	Valine

## Chapter 1: Introduction

### a. Viral taxonomy

Noroviruses, previously referred to as Norwalk-like viruses [1], are a group of small, non-enveloped, positive-sense RNA viruses that belong to the *Caliciviridae* family [2]. The *Caliciviridae* family is divided into five genera: *Vesivirus*, *Lagovirus*, *Nebovirus*, *Sapovirus* and *Norovirus*, of which only *Norovirus* and *Sapovirus* are able to infect humans causing gastroenteritis (more commonly known as the “stomach flu”) [3]. The genus *Norovirus* (NV) based on the gene sequence of the capsid protein is subdivided into at least five genogroups (GI-V), consisting of more than 30 strains (Fig. 1) [2]. Strains from genogroups I, II and IV primarily infect humans and are collectively called human noroviruses (HuNVs). NV strains that infect cattle and mice are found in GIII and GV (Fig. 1). The presence of GII and GIV strains in pigs and dogs has raised questions about zoonotic potential (cross-species transmission) and whether animals represent a reservoir from which more virulent strains may emerge. To date no zoonotic infection has been reported, but additional studies are needed to allay these concerns [3, 4, 5]. The inability of clinically important HuNVs to be cultured *in vitro* has hampered the understanding of the viral life cycle. Currently, murine norovirus (MNV), belonging to GV, is the only NV that replicates in cell cultures and in immunocompromised mice [6, 3, 7]. As a result MNV has become a model of choice to study HuNVs. However, MNV infection in immunocompromised mice does not result in overt signs of gastroenteritis. This contrasts with HuNVs, where healthy individuals are clearly susceptible to symptomatic infections [8, 9]. Therefore, there are limitations to the usefulness of this model system to study HuNVs, which have hampered the understanding of viral attachment to, entry into, and replication within the cells of the human gastrointestinal tract, and consequently the development of much needed anti-HuNV therapeutics and vaccines [10].

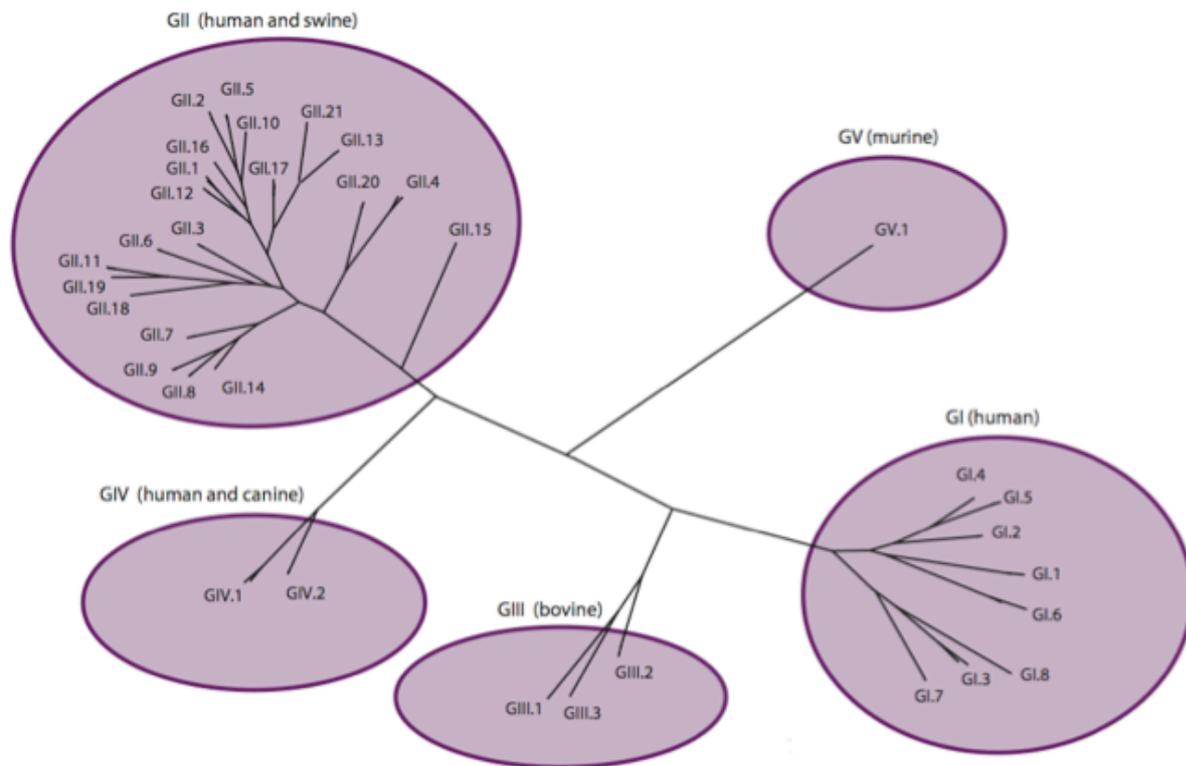


Figure 1: Classification of noroviruses into 5 genogroups (GI-V) and 35 genotypes based on sequence diversity in the complete capsid protein (VP1). Human strains are cluster within GI, II and III [11].

## b. Epidemiology and pathophysiology

HuNVs infect people of all ages and are a leading cause of acute non-bacterial gastroenteritis worldwide – responsible for an estimated 20 million cases, 56,000-71,000 hospitalizations and 500-700 deaths per year in the United States alone [12]. The situation in developing countries is much worse, where it is estimated to be responsible for 1.1 million hospitalization and approximately 200,000 deaths annually [10, 12]. HuNV outbreaks are reported throughout the year, but in temperate climates cold and dry weather has been associated with a short-term increase in NV cases. Summer peaks have also been reported [13, 14] and due to the complex ways in which the environmental factors interact with host factors, it is difficult to fully understand the reasons behind the seasonal variation [15]. However, it has been postulated that the prevalence of HuNV during winter is due to better viral survival, host crowding and reduction in vitamin D levels in humans, which impairs immune responses [16].

Humans are believed to be the only host of human norovirus infection, as to date no cross-species transmission has been reported. Transmission occurs by fecal-oral route with person-to-person and food- or waterborne spread being the most common. Exposure to minute amounts of virus (less than 10 virions) carries a high risk of infection [17]. In addition, NVs have a very short incubation time (1-3 days) and viral shedding by patients can continue up to 1-2 months after the illness. To make matters worse, the virus has been found to be resistant to most disinfectants [18]. These characteristics of the virus along with the unavailability of NV vaccine and anti-NV drugs make it impossible to control or prevent NV infections and outbreaks, leading the Centers for Diseases Control and Prevention (CDC) to classify NVs as a Category B threat [10].

In otherwise healthy individuals, NV infection is rarely fatal and is characterized by acute onset of nausea, vomiting, abdominal cramps, fever and muscle pain that is followed by non-bloody diarrhea that lasts no more than 4 days [18, 19]. However, in immunocompromised patients diarrhea lasting up to 6 months has been reported with a mortality rate as high as 25% [10]. Recently, HuNVs have also been linked to more serious conditions such as necrotizing enterocolitis (inflammation and death of intestinal tissue) [20] and seizures in infants [21].

The norovirus infection is usually self-limiting with no need for hospitalization. The treatment for NV gastroenteritis, like other diarrheal illnesses, requires oral rehydration therapy with fluids and electrolytes. In severe cases, parenteral nutrition (also known as IV nutrition therapy), along with anti-motility and anti-secretory agents are prescribed to prevent dehydration [19, 10]. However, to date there are still no vaccines and/or anti-NV drugs available in the market that might reduce the severity of the infection or prevent it altogether.

### c. Virus genome organization

NVs have a non-enveloped icosahedral capsid that encapsulates the viral genome, which is positive-sense RNA approximately 7.7 kb in length [22, 2]. The genome is covalently attached to viral genome-linked protein (VPg) at the 5' end and polyadenylated at the 3' end. It is organized into three open reading frames (ORF1-3; Fig. 2), with the exception of MNV, which has a fourth

alternative open reading frame [23]. ORF1 encodes a 200 kDa precursor polyprotein, which is co- and post-translationally cleaved by the viral protease into six non-structural proteins in their mature or intermediate forms that are essential for viral replication [3]. The mature non-structural proteins include p48, NTPase, p22, VPg, protease (Pro) and RNA-dependent RNA polymerase (RdRp/Pol) (Fig. 2), whereas the stable intermediate forms include p22VPg and ProPol [24]. The functional importance of these intermediates is not fully understood. On the other hand, structural proteins, which include major capsid (VP1) and minor capsid (VP2) proteins, are encoded by ORF2 and ORF3 [22, 3]. ORF4, which is unique to MNV in Caliciviridae family, overlaps with ORF2 and encodes a protein called virulence factor 1 (VPF1) [22, 3].

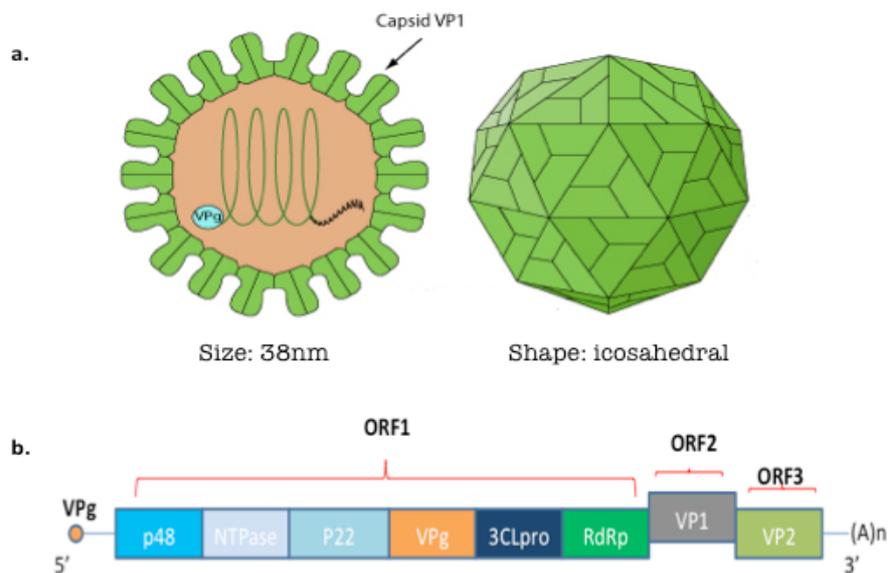


Figure 2: a) Structure of norovirus capsid. The viral genome is encapsulated in a non-enveloped capsid composed of 180 VP1 proteins. The virion is roughly 38nm in diameter (Image adapted from ViralZone, [http://viralzone.expasy.org/all\\_by\\_species/194.html](http://viralzone.expasy.org/all_by_species/194.html)). b) Norovirus genome organization. The viral genome is organized into three open reading frames (ORF 1-3), with the exception of MNV, which has a fourth open reading frame (not shown). The genome is covalently attached to VPg at its 5' end, whereas the 3' end is polyadenylated.

#### d. NV life cycle

The life cycle of NV can be divided into 6 major steps: entry, uncoating, translation, RNA replication, virion assembly, and release. The attachment of NV to cells is mediated by carbohydrate structures present on the cell surface, which in the case of HuNVs include interaction of viral capsid protein (VP1) with histo-blood group antigens (HGBAs) in the gastrointestinal tract [25, 2]. However, this interaction is not sufficient to mediate entry and

binding to an unidentified protein receptor is thought to be required (Fig. 3; steps 1-3) [3]. Following entry, the virus is uncoated (Fig. 3; step 4) and the viral positive-sense RNA is released into the cytoplasm through as-yet-undefined pathways, where it acts as a messenger RNA (mRNA) template for the viral protein synthesis. The 5' end covalently attached VPg interacts with the host cell translation initiation factors to initiate viral genome translation via recruitment of 43S ribosomal pre-imitation complex (Fig. 3; step 5) [26, 27]. The translation of the viral genome, followed by co- and post-translational processing by the viral protease, results in the release of viral non-structural proteins including RdRp, which carries out RNA replication (Fig. 3; step 8) [3].

The first step of RNA replication involves the synthesis of negative strand RNA (Fig. 3; blue colored RNA), which is then used as template for the transcription of genomic and sub-genomic RNA by the viral polymerase (Fig. 3; steps 8-9). The viral genomic and sub-genomic RNAs can then either be used for more rounds of translation (Fig. 3; step 5) or are directed towards virion assembly and exit pathways (Fig. 3; Steps 9-10) [3].

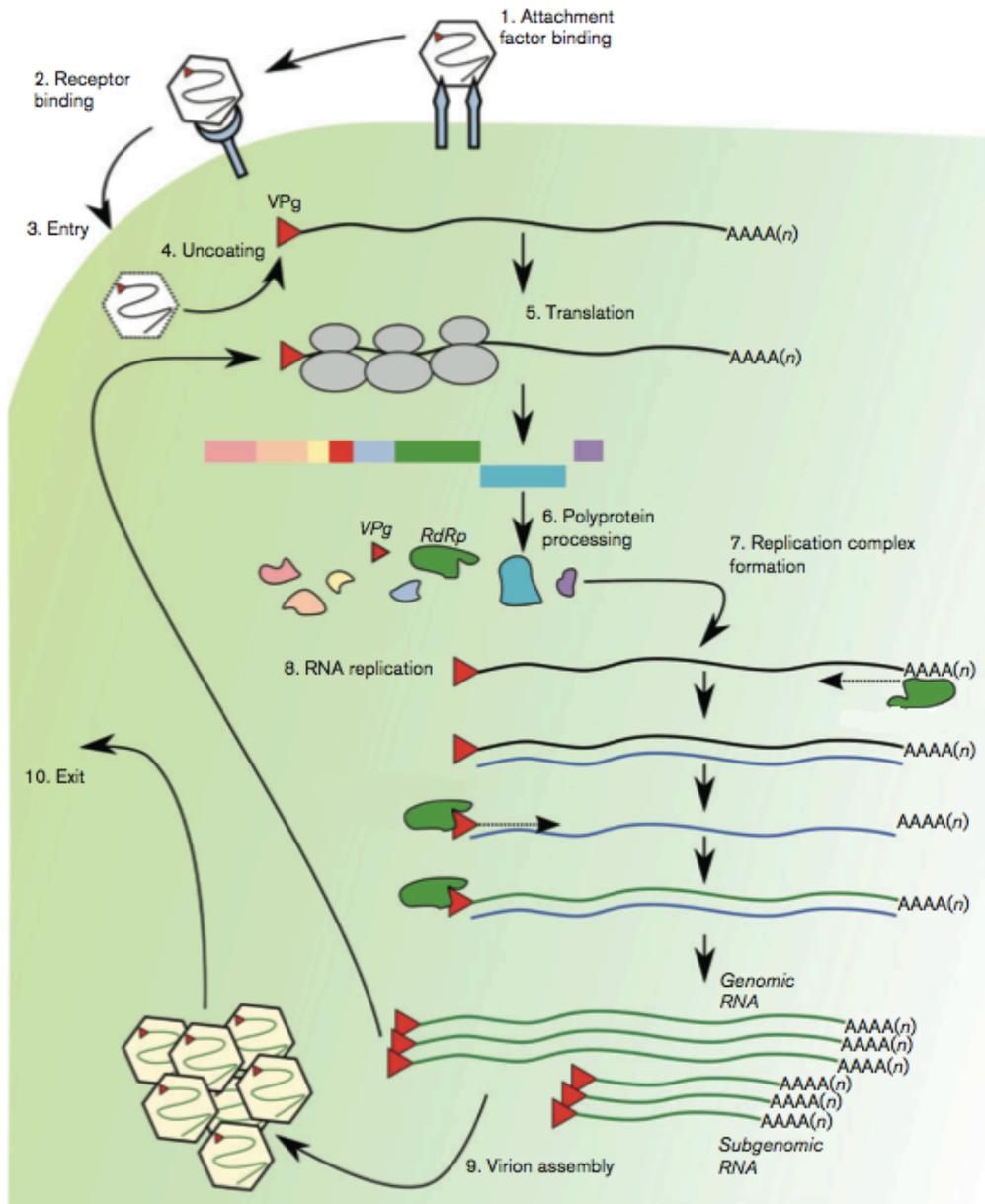


Figure 3: Life cycle of norovirus [3].

## e. NV Pro

NVs encode a single Pro, which is roughly 20 kDa in size. NV Pro plays a pivotal role in the viral life cycle through proteolytic cleavage of ORF1 encoded non-structural polyprotein into their mature or intermediate forms [3]. The intermediate forms of NV Pro include NV Pro precursors protein, ProPol, which possesses both protease and polymerase activity [24]. In related human sapovirus (SaV) and feline calicivirus (FCV), it has been shown that the virus does not produce either mature Pro or Pol, but instead relies on ProPol precursor for both polyprotein processing and RNA replication [28]. Therefore, it is possible that the NV ProPol precursor, like FCV and SaV, is important for viral replication strategies. This is supported by previous studies, which have shown that NV ProPol precursor is an active part of the viral replicase, but the role of NV Pro in this complex is still not clear [29]. Recently, direct ribonucleic acid (RNA) binding was shown to inhibit NV Pro activity in a non-competitive manner [28]. These results provide support for the participation of NV Pro or its precursor ProPol in regulation of viral replication, something that has been demonstrated but not yet fully understood in a wide range of RNA viruses. Although viral protease polyprotein processing activity and the interaction with the viral RNA has been studied in some detail for various positive-sense RNA viruses, including NVs, the underlying mechanism(s) regulating both these activities is not clear. Moreover, it is also not well-defined if either these functions are mutually exclusive or Pro interaction with peptide substrate regulates RNA binding.

### I. Structural features of NV Pro

The first crystal structure of NV Pro was solved in 2005 [30]. Since then numerous other structures of Pro have been solved (Appendix 1), all confirming that it adopts a chymotrypsin-like fold comprised of two domains: N-terminal anti-parallel  $\beta$ -sheet and C-terminal  $\beta$ -barrel domain, separated by a groove where the active site is located. The N-terminal domain is composed of two  $\alpha$ -helices (not labeled) and five well defined  $\beta$ -strands (aI, bI, cI, fI, and gI)-which form a twisted anti-parallel  $\beta$ -sheet (Fig. 4). The N-terminal is connected, via a large loop (amino acids 61-79), to a much larger C-terminal domain that is defined by a six-stranded anti-parallel  $\beta$ -barrel, formed by aII, bII, cII, dII, eII and fII [31, 30, 32].

NV Pro although generally similar to other viral cysteine protease e.g. poliovirus (PV), hepatitis A virus (HAV) and human rhinovirus (HRV), has one distinct structural difference *i.e.* the presence of anti-parallel  $\beta$ -sheet instead of a  $\beta$ -barrel domain in N-terminal (Fig. 5) [30, 31].

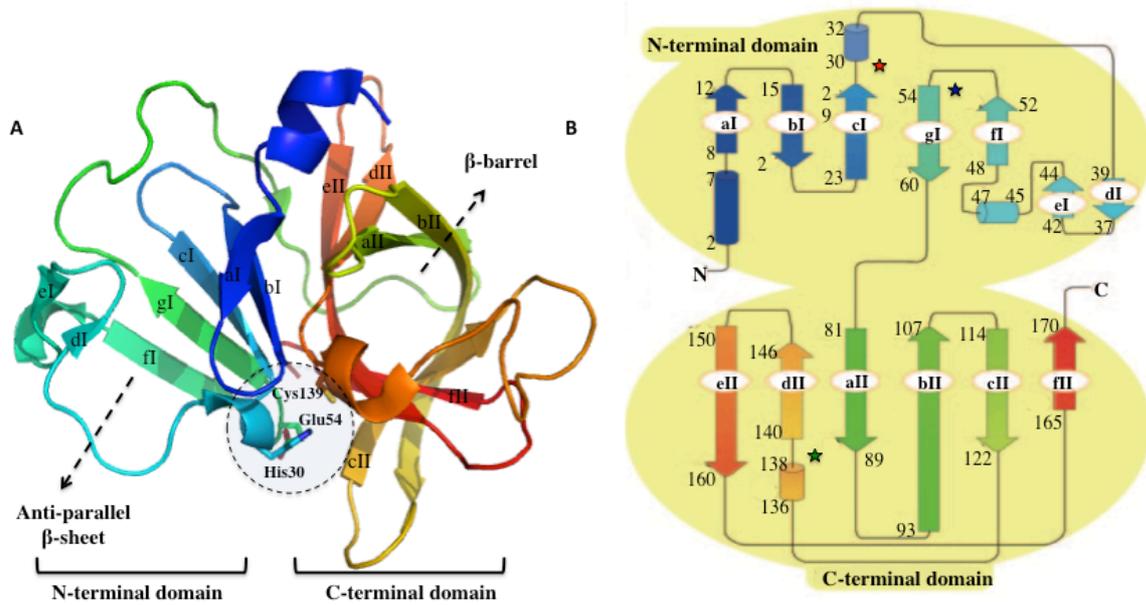


Figure 4: Structure of NV Pro. (A) Cartoon representation of NV Pro (PDB ID: 1QWS) with active site catalytic residues (His30, Glu54 and Cys139) shown as sticks inside a dotted circle. (B) NV Pro secondary structure topology. The location of active site catalytic residues is indicated by different colored stars- His30 (Red star), Glu54 (Blue star) and Cys139 (Green star) [30].

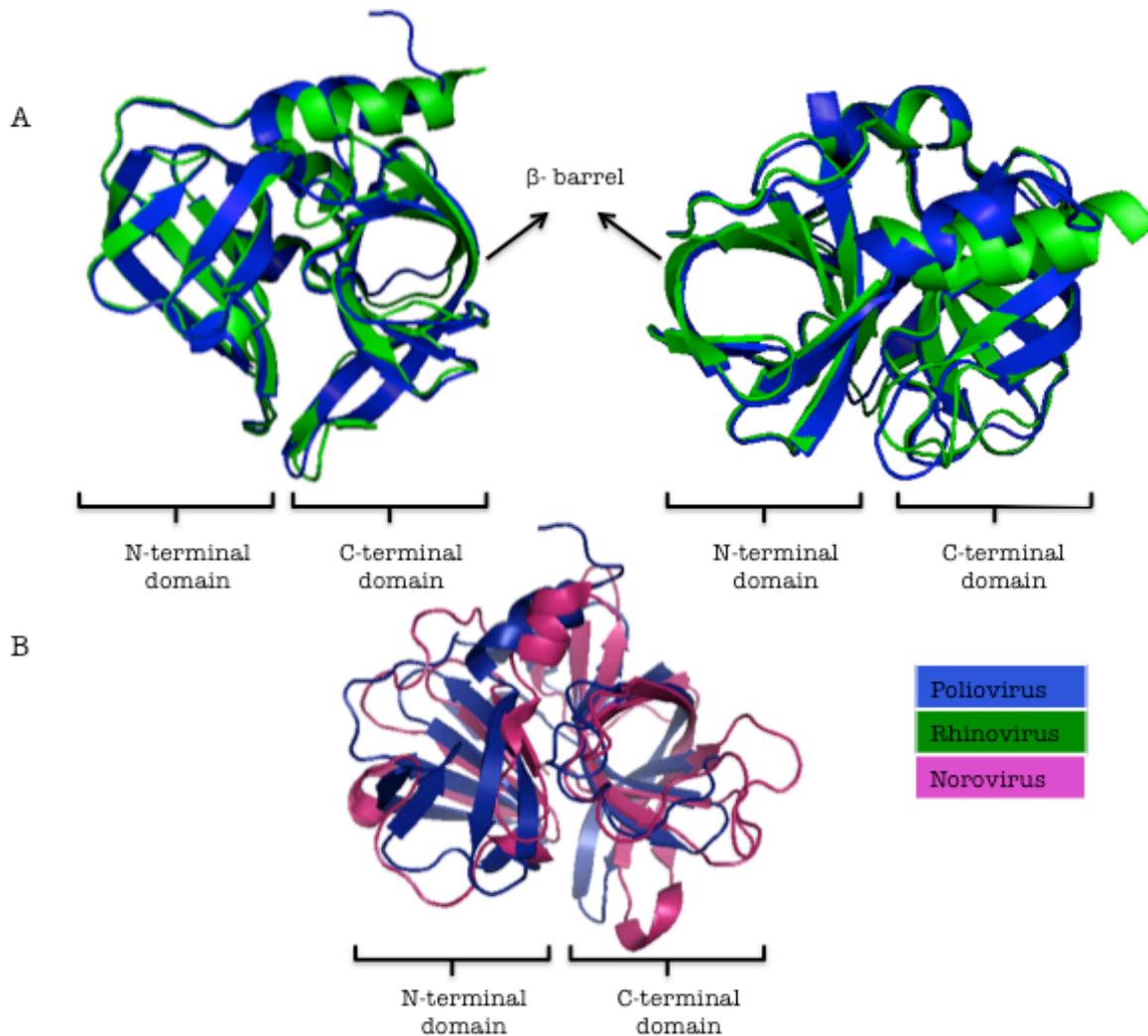


Figure 5: Structural comparison of the NV Pro (PDB-ID: 1QWS) with PV (PDB-ID: 1L1N) and HRV proteases (PDB-ID: 1CQQ). (A) Overlay of PV and HRV proteases.  $\beta$ -barrel domain that is present in both C- and N-terminal domain is labeled (B) Overlay of PV and NV proteases showing their similar overall structure. The cartoon representations of proteases were prepared using PyMOL (Schrödinger, LLC).

## II. Catalytic site and mechanism of proteolysis

The active site of NVs Pro consists of a conserved catalytic triad made up with cysteine (Cys139) as the nucleophile, histidine (His30) as the general base catalyst and the glutamic acid (Glu54) as the anion to orient the imidazole ring of His30. These residues are located deep within the cleft between N- and C-terminal domains (Fig. 4). His30 and Glu54 are part of the N-terminal domain and are located on two separate loops that connect cI-dI and fI-gI strands (Fig. 4). Cys139 on the other hand is part of C-terminal domain and is present on a loop that connects strands cII to dII (Fig. 4) [32, 30]. The peptide substrate binds to the region between the two

domains, mainly through interactions with the  $\beta$ -barrel eII strand, which is part of C-terminal domain (Fig. 4). On binding it adapts a  $\beta$ -strand conformation, which is stabilized primarily by the hydrogen bond interactions with Ala160, Ala158, His157, Gln110 and Arg108. These interactions allow for the correct positioning of the substrate in the active site for proteolysis (breakdown of protein into smaller polypeptides or amino acids) [31].

The catalytic mechanism of NVs Pro is similar to serine proteases, and involves the activation of peptide substrate amide bond via the interaction of carbonyl with cysteine sulfhydryl group (R-SH) [32]. The nucleophilic attack of Cys139 is facilitated by imidazole group of His30, which results in the formation of a tetrahedral intermediate (Fig. 6: Step 1). The oxyanion hole formed by conserved residues (Cys139 and Gly137), not only helps bind the peptide tightly but also stabilizes the negative charge of the tetrahedral intermediate (Fig. 6; Step 2). This is followed by protonation of the tetrahedral intermediate by His30-H<sup>+</sup>, which results in the release of the C-terminal part of the substrate as free peptide and the formation of acylenzyme (Fig 6; Step 2). The next step (Fig. 6; Step 3: hydrolysis) involves an attack by water on the ester bond of the acylenzyme, yielding a second tetrahedral intermediate that collapses to release the second peptide product as carboxylic acid and regenerates serine sulfhydryl group (Fig. 6; Step 4) [33]. The Pro is then ready for the next round of proteolysis.

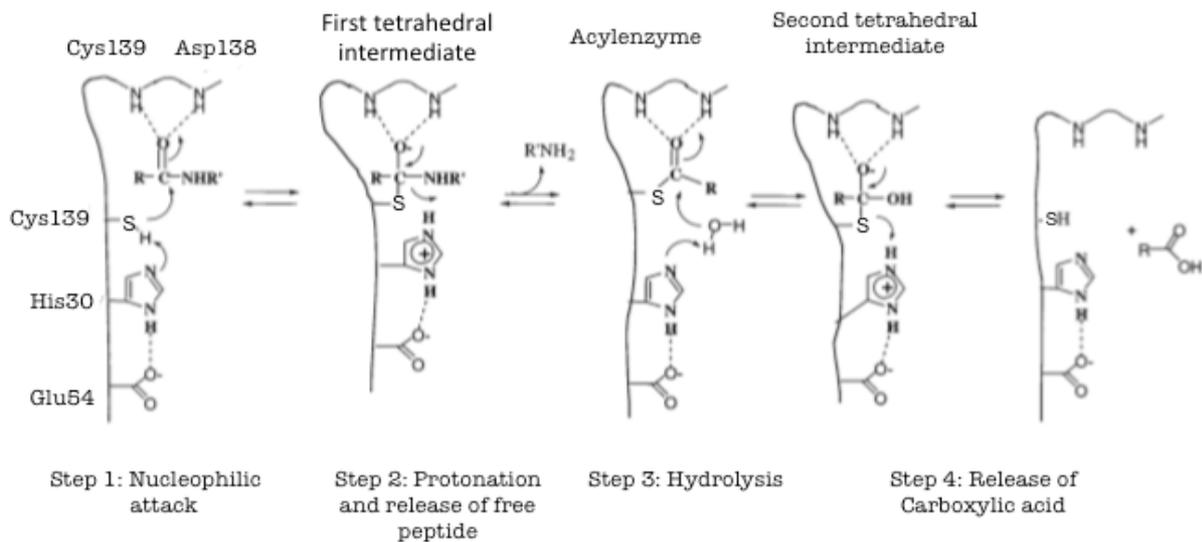


Figure 6: NV Pro catalytic mechanism (adapted from Hedstrom *et al.* [33])

### III. Peptide substrate recognition and proteolytic processing order

The processing of ORF1 by NV Pro occurs at five sites, with either glutamate-glycine (Q-G), glutamic acid-glycine (E-G) or glutamic acid-alanine (E-A) as the cleavage junction. These cleavage sites, based on the processing order *in vivo* and *in vitro* studies, can be grouped into “early and “late” cleavage sites- with Q-G sites cleaved first, followed by E-A and Q-G sites (Fig. 7A) [34, 35, 36]. Mutational studies have shown that substitution of early cleavage site with the late one (Q-G > E-G) is not enough to alter the proteolytic processing order, suggesting that cleavage sites are processed independently of one another and are regulated by the residues flanking the cleavage sites (Fig. 7B) [27, 37]. Recently, kinetic studies have confirmed that the flanking residues, especially P4-P2' (Appendix 2: Schechter and Berger nomenclature), determines the processing order primarily via regulation of NV Pro rate of catalytic efficiency ( $K_{cat}/K_m$ ) [34]. It is primarily the rate of reaction ( $K_{cat}$ ) and not the binding affinity ( $K_m$ ) that determines the preference of Q-G sites over E-G and E-A (Fig. 7C). Moreover, it has been shown that by replacing the P4-P2' residues of early cleavage site with late cleavage site, the processing order can also be altered. All these results suggest that P4-P2' forms a core sequences that contain all the information necessary for the regulation of ORF1 polyprotein processing [35].

The structures of NV Pro-product complexes have been solved by adventitious crystallization. Crystal packing in these complexes resulted in the interaction of C-terminal tail of one molecule, representing the P4-P1 residues, into the active site of another (Fig. 8) [38, 36]. Comparative studies of these Pro-product complexes with unbound (native) NV Pro structures have provided insight into how different amino acid residues are accommodated by the Pro and how substrate-induced conformational changes might regulate its activity. The majority of the interactions with the substrate involve hydrophobic S1 and S2 pockets (Appendix 2: Schechter and Berger nomenclature) located in the N-terminal domain. S1 pocket is formed by His157, Ala160 and Thr134 [32], whereas S2 pocket is defined by the hydrophobic residues Ile109, Glu110, Arg111, and Val114, which are part of a loop that connect bII and cII  $\beta$ -strands (Fig. 9 and 10). S2 pocket is not only larger than S1 pocket but also undergoes considerable conformation change upon

substrate binding (Fig. 9 A, B and C), allowing it to accommodate variation in the P2 position of the substrate (Fig. 7) [36]. Depending on the absence or the presence of particular residues, S2 pocket depicts closed, semi-open or open conformations (Fig. 9). Moreover, the hydrophobic nature of S2 also explains its selectivity for hydrophobic amino acids at P2 position. On the other hand, the S1 pocket does not show significant difference upon substrate binding. This makes sense as S1 pocket binds only to two similar residues *i.e.* either Gln (Q) or Glu (E) residue (Fig. 7A).

S3 pocket is not well defined and the lack of conserved interactions observed in crystal structures with P3 residues explains the diversity of peptide residues at this position in NVs (Fig. 7 B) [36, 38]. In contrast, kinetic studies have implicated P3 residue to play a role in regulation of polyprotein processing by modulating HuNV Pro efficiency [39]. This residue had been overlooked because most of the earlier studies utilized a NV strain (Norwalk) as a standard peptide, which contains His at P3 site. The peptide containing Glu at P3 instead of His was observed to increase the NV Pro enzyme efficiency 4-fold [35]. Based on the modeling studies, it has been proposed that P3-Glu carboxylate group (negatively charged residue) forms favorable electrostatic and hydrogen-bonding interactions with residue-162 (Lys or Arg in GI/GII HuNV) and P1-Glutamine (Gln), thereby increasing the HuNV Pro efficiency by promoting “enzyme:transition-state” complex formation. In contrast, these interactions were not observed in P3-His (positively charged residue) models, supporting a potential role of P3 residues in polyprotein processing [35]. The preference for positively charged residues over negatively charged ones at P3 position, with a 5-14 fold increase in activity, has been previously reported for *Coronavirus* protease [40]. Therefore, it is possible that viral proteases have evolved to utilize charged residues at P3 position as a common strategy to regulate polyprotein processing. However, given its minimal affect in NVs *i.e.* only 4-fold increase in enzyme efficiency, it is likely that P3 residue does not play a major role in HuNV Pro substrate specificity and polyprotein processing.

S4 pocket (Fig. 9) is formed by hydrophobic residues (Met107, Arg108, Ile109, Thr166 and Val 168), which explain its preference for hydrophobic residues at P4 (Phe, Ile, Leu and Ala). S4 also exhibits significant conformational change upon peptide binding (Fig. 10 C, D and E). The

substrate-induced conformational changes are coordinated between S2 and S4, such that the constriction of S2 results in the widening of S4 pocket and vice versa (Fig. 10). The coordinated changes observed are due to bII-cII loop that is shared by both S2 and S4 pocket (Fig. 9) [36]. Beyond S4 there appears to be little specific interactions between the peptide and NV Pro [36, 38]. This observation supports the results of kinetic studies, mentioned earlier, that residues beyond P4 are not important for polyprotein processing.

However, as far as the P' residues are concerned, there are no structural data available which might help explain how they interact with NV Pro. Surprisingly, in kinetic studies the substitution of P2' glycine of the ProPol late cleavage site for early cleavage sites P2' proline, resulted in a 3-fold increase in HuNV efficiency [39]. Modeling studies indicate that the peptide containing glycine at P2' position can be stabilized by additional interactions between glycine and other P' residues [39]. In addition, modeling studies have also indicated that the NV Pro can accommodate P' residues up to P3' [30] or P4' [35]. However, based on the visual inspection of the available NV Pro crystal structures in the presence and absence of peptide substrate or peptide-based inhibitors, there appears to be no major binding groove beyond S1' pocket [41], and P' residues beyond P2' can take two different routes (Appendix 3: Peptide Route 1 and 2) [31]. Based on these results it can be concluded that there is a potential for NV Pro-P' interactions to occur, which may lack specificity beyond P1'. Therefore, further structural studies are needed to help understand how P' residues interact with the enzyme and bring about any substrate-induced conformational changes that might explain their role in binding specificity.

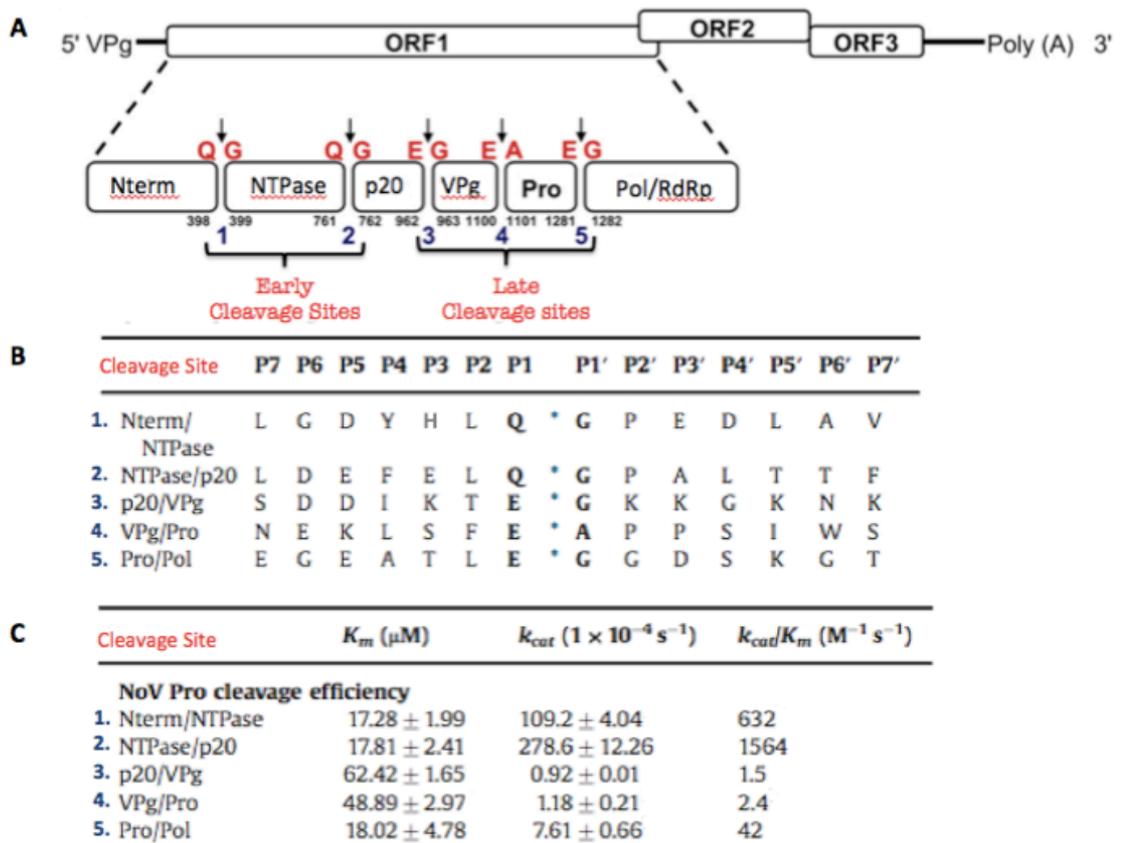


Figure 7: (A) The processing of ORF1 polyprotein occurs at five sites, which can be grouped into “early” and “late” cleavage sites (1-5) [36]. (B) The most common HuNV amino acid sequence (P7-P7’) along each cleavage site (P1-P1’) [34]. (C) Kinetics of HuNV Pro with ORF1 substrates [34].

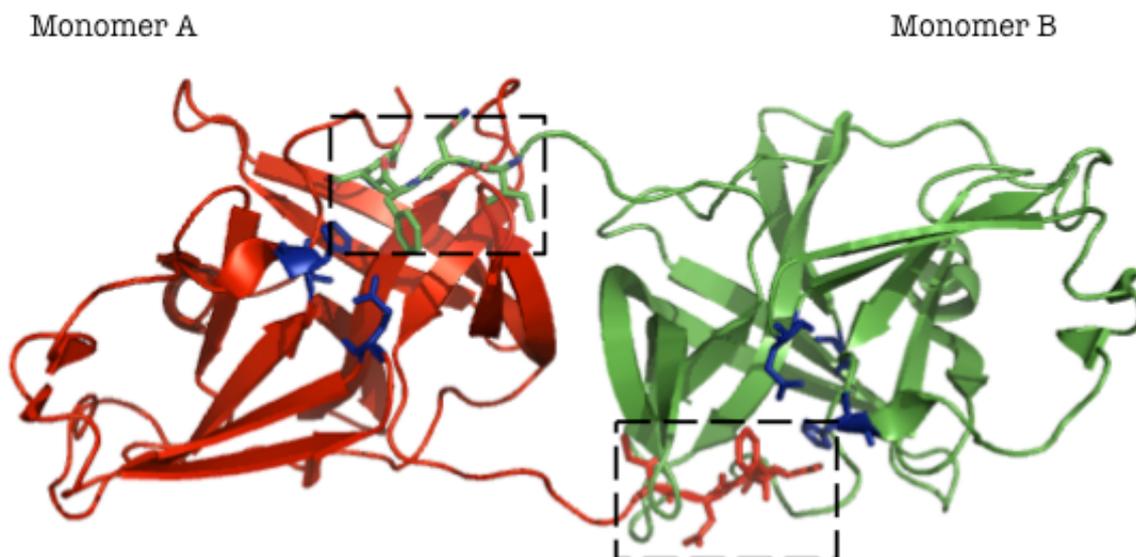


Figure 8: Ribbon representation of HuNV Pro-product complex structure (PDB ID: 4IN1). The two monomers (A and B) interact such that the C-terminal residues *i.e.* 178-181 of one monomer, represented by sticks in dashed boxes, corresponding to positions P4 to P1 of the native substrate are inserted into the active sites (blue sticks) of the second monomer.

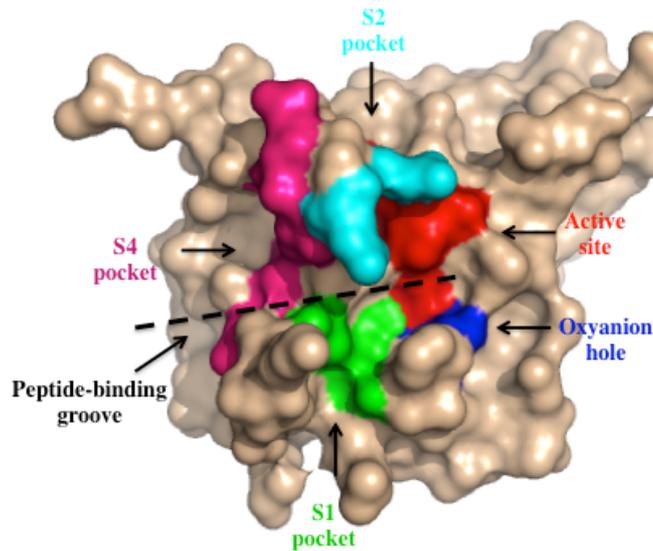


Figure 9: Surface representation of NV Pro (PDB ID: 4IN2) with color-coded active site (red) and substrate binding pockets. The dashed line represents the peptide-binding groove where the peptide substrate binds.

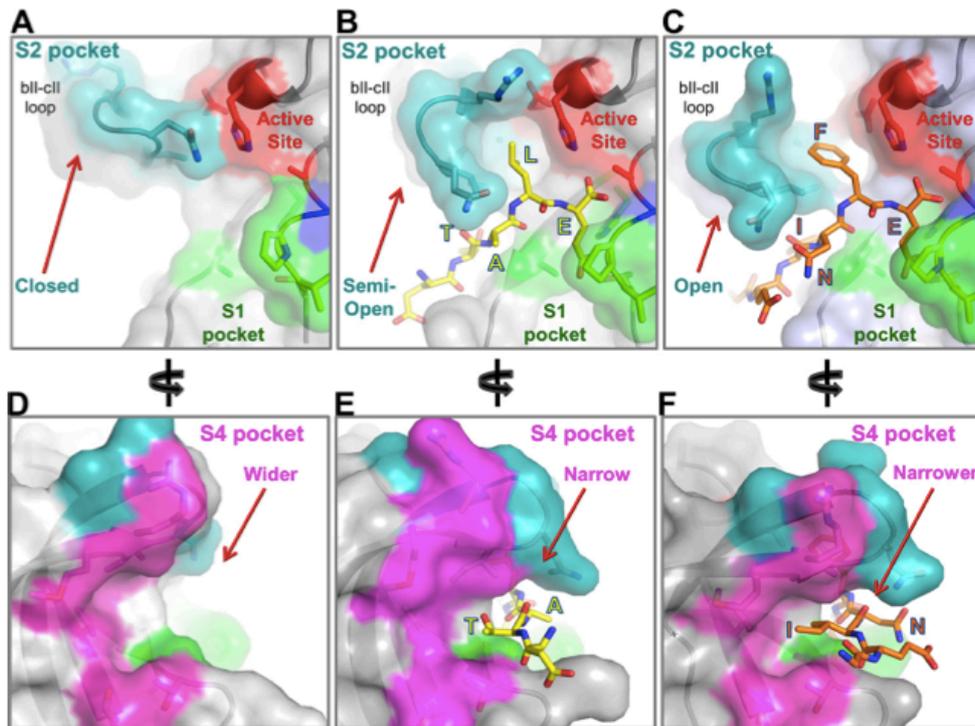


Figure 10: Surface representation of peptide-induced conformational changes in S2 and S4 pockets of HuNV Pro. Pockets exhibit three different conformational states: closed/narrower, semi-open/narrow and open/wider, depending on the absence or presence of particular residue. The peptide sequences (TALE and INFE) are shown along the length of the substrate and represent P4-P1 positions of the natural substrate. The active site is depicted in red, the S1 pocket in green, the S2 pocket in teal, and the S4 pocket in magenta. (A/D) The S2 and S4 pockets in the absence of substrate. (B/E) The S2 and S4 showing semi-open and narrow conformations in the presence of TALE substrate. (C/F) The conformational changes in S2 and S4 pockets to accommodate bulkier substrate residues (F) at P2 position in INFE substrate. S2 adopts an open conformation to accommodate bulkier F (Phe) residue, which is facilitated by the movement of bII-cII loop. This loop is shared by both S2 and S4 such that widening of S2 results in constriction of S4 pocket and vice versa [36].

## f. NV Pro-RNA interaction

Previous studies have indicated that the HuNV ProPol precursor is an active part of the viral replicase. However, no precise role for Pro in this complex was observed [29, 42]. Recently, direct RNA binding by HuNV Pro was shown to inhibit Pro activity in a non-competitive manner (Fig. 11) [28]. These results provide support for the participation of NV Pro in regulating of viral replication. In related PV and HRV proteases, the RNA-binding motif has been mapped to the highly conserved amino acid motif KFRDI, which is absent in NV Pro [43, 44, 45]. Therefore, it is probable that the binding of RNA involves a novel, as-yet-unidentified mechanism and RNA/protein interactions. Moreover, based on the initial enzyme inhibition studies there appears to be little evidence for RNA sequence specificity for the NV Pro, although length was observed to have an effect on inhibition, with longer RNA oligonucleotides being responsible for more potent inhibition (Table 1) [28].

RNA binding by other related cysteine proteases, such as the PV and HRV proteases, is well established [46]. Studies have indicated that the binding of PV and HRV Pro and/or ProPol precursor to the RNA secondary structures, termed cis-acting replication elements (CREs)-located in the untranslated regions (UTRs) and at an internal position within the viral RNA- is important for genome replication [47]. It has been proposed that the binding of CREs to Pro leads to unwinding of these secondary structures, which then serve as the template for uridylylation of primer protein VPg by the viral polymerase to initiate viral replication [48]. Therefore, NV Pro might be involved in the regulation of viral replication, perhaps through a similar mechanism observed in related PV and HRV proteases [48, 46, 28, 49]. However, further studies are warranted to help understand the role of CREs, RNA binding mechanism, interplay between NV Pro peptide and RNA binding sites, and different roles of Pro and ProPol in HuNV life cycle.

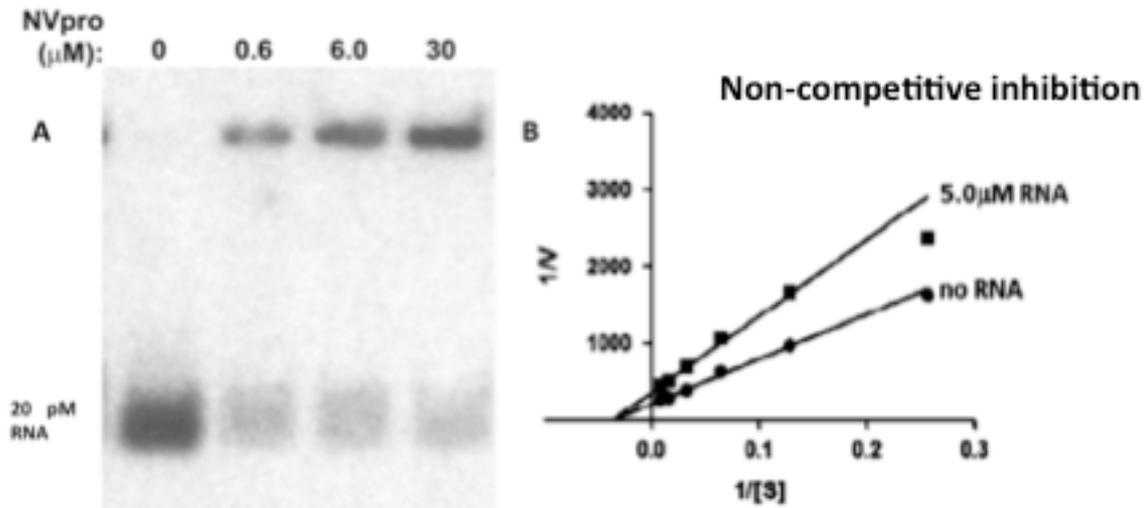


Figure 11: RNA binding by HuNV Pro and its effect on enzyme kinetics. A) RNA binding by HuNV Pro was confirmed by mobility shift electrophoresis. B) Lineweaver-Burk analysis of the effect of RNA on HuNV Pro activity [28].

Table 1: Effect of RNA on HuNV Pro activity.  $IC_{50}$  (2-fold reduction in activity) and  $IC_{90}$  (10-fold reduction in activity) were calculated relative to an inhibited proteolysis reaction. pET32 represent a non-NV RNA [28].

Oligonucleotides (14-mer)	$IC_{50}$ ( $\mu$ M)	$IC_{90}$ ( $\mu$ M)
5'-End, sense	$5.5 \pm 0.1$	$23.3 \pm 1.1$
5'-End antisense	$4.8 \pm 0.5$	$18.6 \pm 1.2$
3'-End, sense	$5.4 \pm 0.1$	$21.5 \pm 0.2$
3'-End antisense	$3.5 \pm 0.2$	$10.7 \pm 0.9$
RNAs		
NV Pro (642nt)	$0.013 \pm 0.001$	$0.46 \pm 0.003$
pET32 (589nt)	$0.017 \pm 0.002$	$0.62 \pm 0.002$

## Chapter 2: Objectives of the Study

The goals of this study were to determine the three-dimensional structure of HuNV Pro in complex with peptides and RNA substrates by X-ray crystallography. Structural data of HuNV Pro-substrate complexes will uncover details of the interactions made by the NV Pro with RNA oligonucleotides and the peptide residues, prime side of the peptide cleavage junctions. This will help identify the mechanism of RNA binding in viral proteases and explain how variation in peptide cleavage sequences regulate HuNV Pro-Peptide binding and activity.

### a. Interaction of HuNV Pro with RNA oligonucleotides

Even though HuNV Pro has been observed to bind viral RNA, the amino acids involved in RNA binding are yet to be determined, which based on the non-competitive nature of inhibition supports the presence of a distinct RNA binding motif on the NV Pro. Since no structures are available for any viral protease in complex with RNA, we have undertaken X-ray crystallographic studies of HuNV Pro in complex with RNA oligonucleotides, which were identified in already published NV Pro-RNA binding studies [28]. These Pro-RNA complexes would allow us to identify the mechanism of RNA binding and to elucidate how RNA can inhibit proteolytic activity in a non-competitive manner. Since the NV Pro residues involved in RNA binding are unknown [28], we propose two putative RNA binding sites on NV Pro (Fig. 12), which due to their distinct positions from peptide-binding site are consistent with the observed non-competitive inhibition. RNA binding proteins are known to bind RNA mostly via positively charged residues that compensates the negative charge of the RNA [50]. Keeping this in mind the two proposed RNA-binding sites on NV Pro-based on the visual inspecting of the surface potential-represent the two largest positive potential patches on the enzyme (Fig. 12). In a distantly related viral cysteine protease, HAV 3C, it has been reported that dimerization results in the formation of an extended RNA-binding site, which improves RNA binding affinity [43]. Dimerization of NV Pro has also been reported in crystallographic studies [51, 31, 52, 36], but the dimer interfaces observed in NV Pro crystallographic studies are quite different from each other [51]. Considering it is unlikely that biologically relevant dimer would have different contacts as observed for NVs Pro, the role of NV Pro dimerization in RNA binding warrants

further studies. The structures of HuNV Pro in complex with different RNA oligonucleotides would also shed light on the possible role of NV Pro dimerization in RNA binding.

To help solve the structure of HuNV Pro in complex with RNA oligonucleotides, the project was divided into the following specific objectives:

- I. To develop efficient expression and purification protocols for wild type HuNV Pro in order to meet the high sample quantity and purity requirements of crystallization.
- II. To identify crystallization conditions that promotes the formation of HuNV-RNA complexes.
- III. To collect and process X-ray data to understand the nature of HuNV Pro-RNA interaction.

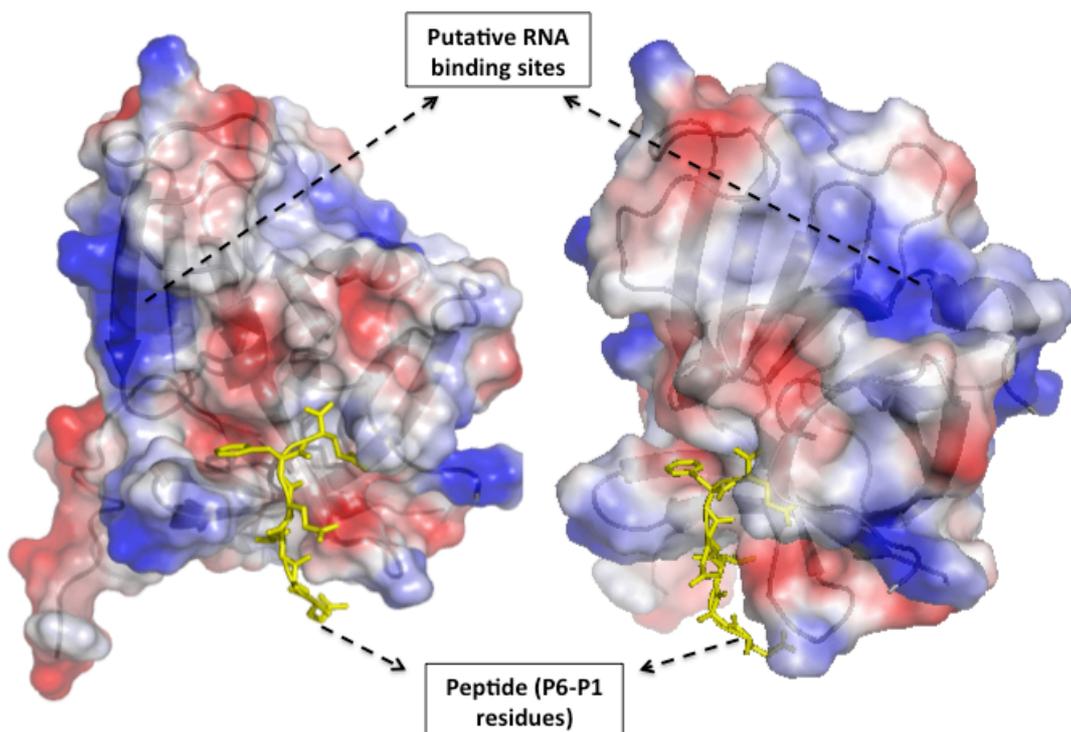


Figure 12: Putative RNA binding sites on NV Pro (PDB ID: 4ASH) bound to peptide substrate (yellow sticks). The electrostatic potential distribution on the protein surface was calculated using PyMOL. The positive potential is colored blue and negative potential red.

## b. Interaction of HuNV Pro with P' residues

As mentioned earlier, there is no structural information available on NV Pro-peptide complexes which might shed light on the interaction made by the Pro with P' residues. It is also not clear how many P' residues can be accommodated by the enzyme. Keeping this in mind, synthetic peptides with a different number of P' residues were used in this study to prevent overhanging residues that might interfere with the crystallization of HuNV Pro-peptide complex. Moreover, to promote substrate binding the sequence of these peptides were based on the early cleavage site (N-term/NTP-ase), as it has the strongest binding affinity to HuNV Pro (Fig. 7 A and C). These peptide sequences also contain a glutamic acid (E) at P3 position. Therefore, this project has the potential to help understand how P' and P3 residues interact with HuNV Pro, thereby providing insight into polyprotein processing. Ultimately, this work would aid in the development of much needed anti-viral treatments for HuNV infection.

To achieve this goal, several specific objectives were developed in an attempt to solve the structure of HuNV Pro-peptide complexes.

- I. To develop an efficient expression system for inactive mutant HuNV Pro (C139A). This is needed to prevent autolysis of synthetic peptides used in co-crystallization studies.
- II. To develop efficient purification protocol in order to meet the high sample quantity and purity requirements required for crystallographic studies.
- III. To identify crystallization conditions that promotes the formation of HuNV Pro-peptide complexes.
- IV. To collect and process X-ray data to determine the structures of mutant HuNV Pro in complex with peptide substrates.

## Chapter 3: Materials and Methods

### a. Protein expression

Our collaborators at the University of Georgetown, US, expressed both the wild type and mutant (C139A) HuNV Pro (strain Hu/G1.1/8FIIa/1968/USA, accession number JX023285) with His<sub>6</sub>-tag at N-terminal following previously described methods [28].

### b. Ion exchange chromatography

The HuNV Pro sample received from the collaborators was purified with ion exchange chromatography. Protein was loaded onto a 5ml SP Sepharose High Performance column (HiTrap SP Hp, GE Healthcare Life Sciences) already equilibrated with 5 column volumes of binding buffer, Buffer A (20 mM MOPS, pH 7.0, 0.5 mM EDTA, pH 8.0, 2 mM DTT and 10% glycerol). The bound protein was eluted in 1 ml fractions with a 30 mL gradient from 0 to 1 M NaCl using elution buffer, Buffer B (20 mM Tris-Cl, pH 7.0, 1 M NaCl, 0.5 mM EDTA, pH 8.0, 2 mM DTT and 10% glycerol). The concentration and the purity of the protein in the fractions were measured with NanoDrop (Thermo Scientific) and SDS-PAGE gel.

### c. Gel-filtration chromatography

Ion exchange fractions of highest purity and concentration were pooled together and concentrated to 4.5 ml by centrifugation (6000-8000 RPM). The concentrated protease sample was then applied to a gel filtration column (HiLoad 16/60 Superdex 200 pg, GE Healthcare), which had been equilibrated with buffer, Buffer C (20 mM MOPS, pH 7.0, 100 mM NaCl, 0.5 mM EDTA, pH 8.0, 5% glycerol, 2 mM DTT). The 1 ml fractions were eluted at a flow rate of 0.2 ml/min and collected in a 48 spot collector tray by auto-fraction collector. The presence of HuNV Pro oligomers were confirmed by running the sample against Gel filtration Markers kit for Protein Molecular Weight 12-200 kDa (Sigma Aldrich).

#### d. Dialysis and concentration

Fractions of highest concentration and purity were pooled together and dialyzed against low salt storage buffer (20 mM Tris-Cl, pH 7.0, 20 mM NaCl, 0.5 mM EDTA, pH 8.0, 2 mM DTT and 10% glycerol) overnight. The dialyzed protein was then concentrated by ultrafiltration until a concentration of approximately 0.12-0.20 mM ( $A_{280} = 3.0-5.0$  mg/ml) was achieved. The purified wild type and mutant HuNV Pro samples were flash frozen as 50  $\mu$ L aliquots in thin-walled PCR tubes and stored at  $-80^{\circ}\text{C}$ .

#### e. RNA and peptide preparation

RNA oligonucleotides of different sequences (Table 2) were ordered from the University of Calgary CORE DNA services, whereas, the synthetic peptides (Table 3) were purchased from CanPeptide Inc. RNA and peptides were neutralized to pH 7.0-7.5, and dissolved in UltraPure<sup>TM</sup> DNase/RNase-Free Distilled Water (Life Technologies) at different concentrations ranging from 10-30 mM.

Table 2: RNA oligonucleotide information and sequence. These sequences are based on the published HuNV Pro-RNA binding studies [28].

RNA Oligonucleotide	Sequence
1. NVpla (10-mer; 5'-end of NV viral genome)	5'-GUGAAUGAUG-3'
2. NVp2 (14-mer; 3'-end of NV viral genome)	5'-UUUAAUUUGAUGUU-3'

Table 3: Synthetic peptide information and sequence. These sequences represent the residues surrounding the early NS2-3 cleavage site [34, 39].

Synthetic Peptide	Sequence
1. Long Peptide (9 amino acids; P5-P4')	Ac-DYELQGPED-NH <sub>2</sub>
2. Short Peptide (5 amino acids; P4-P1')	Ac-YELQG-NH <sub>2</sub>

#### f. Crystallization methodology

The vapor diffusion method was used for the crystallization studies. This is the most widely used method due to the relative ease with which it can be set up compared to other methods. In this method, the protein is mixed with the reservoir solution is equilibrated in a closed system against

the reservoir solution containing different crystallization precipitants e.g. salts, organic solvents and polymers etc., at a much higher concentration. Initially, the protein solution contains precipitants at lower concentration than required for crystallization but, due to the net movement of water vapors from the protein to the reservoir solution, the concentration of precipitants in protein solution increases gradually to a level optimal for crystallization. Whether the protein forms crystals or amorphous solid, depends on many properties of the protein and reservoir solution, including protein purity, protein concentration, temperature, pH, precipitant and its concentration, and ionic strength [53]. Due to its multi-parametric nature, it is difficult to predict crystallization conditions *a priori*. As a result, several thousand crystallization conditions have to be tested, which, with the advent of crystallization robots and commercially available crystallization screens, has become easier. First step towards growing protein crystals is automated screening, using commercially available crystallization screens, to cover a large part of chemical space to determine initial crystallization conditions (called “hits”) under which protein has the propensity to crystallize. These hits are then fine-tuned manually during crystal optimization step to grow diffraction quality crystals for X-ray data collection.

#### I. Sparse matrix screening

In an attempt to co-crystallize the wild type HuNV Pro-RNA and mutant HuNV Pro-peptide complexes, the protein (0.12-0.20 mM) was mixed with varying concentration of synthetic peptides (1.2-5.4 mM) and RNA oligonucleotide (0.60 mM). These protein-substrate mixes were equilibrated for 2 hours to ensure binding of substrates to HuNV Pro, before setting up the crystallization screens. Since there is no way to predict the set of conditions that will give rise to a single, well diffracting crystal, automated setup was used to screen hundreds of crystallization conditions using commercially available sparse matrix crystallization screens (Appendix 4: Sparse Matrix Screening). These sparse matrix screens involve an intentional bias towards combination of crystallization conditions that have worked previously and cover a large chemical space by using many different crystallization reagents [54]. Each screen used contained 96 different conditions, varying in pH, buffer, precipitant and salt concentrations. Using the Matrix Hydra II eDrop pipetting robot (Thermo Scientific) crystallization trials were set up at room

temperature (25 °C) and monitored periodically under a microscope for over a course of ~12 weeks to identify crystallization hits.

The robotic setup involved placing 40  $\mu$ l of crystallization solution into the base (reservoir) of a 96 well sitting drop vapor diffusion plate (Hampton Research). This was followed by mixing 0.4  $\mu$ l of crystallization solution with 0.4  $\mu$ l of HuNV Pro-substrate (RNA or peptide) solution into the top sample container. These plates were then sealed by hand to prevent drying out of the reservoir and sample container using Crystal Clear Sealing Tape (Hampton Research).

## II. Crystal optimization

The hits identified from sparse matrix screening were repeated manually and optimized by systematically varying each component of the crystallization conditions e.g. protein concentration, substrate concentration, pH, salt and its concentration, and precipitant concentration etc. Crystals were grown using hanging drop vapor diffusion method in 24 well VDXm plates (Hampton Research). HuNV Pro (0.12-0.20 mM) was mixed with either RNA (0.60-2.0 mM) or peptide (1.2-6.0 mM) and equilibrated on ice for 2 hours. The protein-substrate solution (1-3  $\mu$ l) was then mixed with crystallization solution (1-3  $\mu$ l) in different ratios e.g. 1:1, 1:2, 2:1 and 3:1, and equilibrated against 0.5 mL crystallization solution (*i.e.* crystallization condition identified from the screens). These plates were then either placed at room temperature or 4° C. Optimization was also carried out for crystals grown in the absence of substrates. These native crystals were used for soaking experiments (explained later; page 25).

As an additional optimization strategy, seeding technique [55] was also used to improve the quality of co- and native crystals. Small seed crystals were transferred by pipette tip or 18 mm mounted Cryoloop<sup>TM</sup> (Hampton Research) of different diameters (25-150  $\mu$ m) to a pre-equilibrated drop containing modified crystallization condition.

### III. Soaking

In addition to co-crystallization of HuNV Pro with either RNA or peptide, the soaking technique was also used [56]. Since this technique relies on pre-formed crystal, it allows one to easily change the crystal conditions, e.g. pH, salt, ligand concentration etc., in order to promote ligand binding. After native and/or co-crystals had grown to the maximum size, they were transferred using a mounted cryloop (Hampton Research) to a drop containing the soaking solution (1-2  $\mu$ l) containing varying concentrations of RNA (0.6–7.5 mM) and/or peptide (2 mM-20 mM) for 30 min to couple of days. The drop was setup using the 24 well VDXm plates (Hampton Research).

Moreover, in order to avoid damage to the crystal either due to the sudden change in crystal environment or physical manipulation of crystal, a modified soaking methodology was also tried to gently alter the condition. The drop (1-2  $\mu$ l) containing the crystal was diluted with soaking solution (1-2  $\mu$ l) and left to equilibrate for 30 min. The next step involved the removal of 1-2  $\mu$ l of solution, followed by addition of soaking solution (1-2  $\mu$ l) and equilibration. This protocol was followed until the drop solution was completely replaced with the soaking solution. Depending on the stability of the crystal in the soaking solution, varying soaking times were tried (30 min – 2 days).

### IV. Crystal harvesting

Due to limited time and resources, only a few crystals without any visible defects were selected for each synchrotron trip. Crystal harvesting involved the transfer of a protein crystal from its growth solution into a suitable 18 mm mounting Cryoloop<sup>TM</sup> (Hampton Research). Depending on the size of the crystal, cryoloop with a suitable diameter (50-200  $\mu$ m) was selected beforehand. Then the coverslip (18 x 18 mm) containing the crystal was gently flipped and placed under the microscope. While looking under the microscope the selected crystal was quickly picked and flash-frozen in liquid nitrogen and transferred into a CrystalCap<sup>TM</sup> vial (Hampton Research), submerged in liquid nitrogen. Finally, the crystals are transferred into a 96-sample ports storage cassette (MiTeGen) and shipped in storage dewar to synchrotron for X-ray data collection.

During harvesting crystal exposure to air is minimized. This prevents and/or reduces damage to the crystal due to dehydration. The average solvent content in protein crystal is around 43% [57] and plays an important role in its structural stability [58]. Therefore, excess dehydration can lead to crystal damage by structural transformation. Moreover, all the crystals were grown in the presence of cryoprotectant e.g. glycerol and polyethylene glycol, and flash frozen quickly to prevent the formation of ice crystals. These ice crystals not only damage the crystal due to sudden expansion of water upon freezing but, as ice also diffracts X-ray very strongly, it can also interfere with the accuracy of measuring protein diffraction [59].

### g. X-ray data collection and structure determination

X-ray diffraction data was collected either at the Stanford Synchrotron Radiation Lightsource (SSRL) or at the Canadian Light Source (CLS). The cryoloop containing the crystal was mounted onto the goniometer, where it is held in a stream of nitrogen gas (100 K). Each crystal was first screened by measuring two X-ray diffraction patterns at two orientations separated by 90° with X-ray beam normal to the plane of the crystal. If the crystal was found to be relatively defect-free then complete X-ray data set was collected, which involved collecting a total of 400-500 frames with 0.5° oscillations and 2s of exposure per frame. Data were processed and scaled using XDS [60]. Molecular replacement (MR) analysis was performed using Phaser [90] with the structure of HuNV Pro (PDB ID: 3UR6; 99% sequence similarity) as the search model. The model building and refinement was carried out using COOT [61] and Refmac [62]. Residues were added and adjusted to fit the electron density map (called model building), followed by refinement to obtain a model which best explains the experimental X-ray data. Progress was monitored by R-factor and R-free values, both of which improve progressively with the improvement in model quality [63]. All the representations of HuNV Pro were generated with PyMol (The PyMol Graphic System, Version 1.7.4 Schrödinger, LLC).

## Chapter 4: Results and Discussion

### a. HuNV Pro-RNA interaction

#### i. Expression and purification

Protein expression and purification are often rate limiting steps in crystallography, as it requires milligram amounts of highly purified protein. Our collaborators successfully expressed wild type HuNV Pro as His<sub>6</sub>-Pro fusion protein (approx. 20 kDa), using previously established protocols [35, 28]. The protein was purified using ion-exchange chromatography. The pattern of elution is presented below (Fig. 12), where the peak 1 and peak 2 correspond to elution of the protein at around 20-30% buffer B (200-300 mM NaCl). Since the SDS-PAGE profile for both peak 1 and 2 is identical, it is possible that peak 1 represents the HuNV Pro dimeric form. The charged residues probably buried in the dimer interface, weakens dimer interactions, compared to the monomer, with the ion-exchange column, thereby resulting in the elution of protein dimer followed by the monomer. Peak fractions were analyzed for purity by SDS-PAGE, and the fractions of the highest purity (Fig. 13) were pooled together and dialyzed against low salt storage buffer. The protein solution was then concentrated by ultrafiltration to 0.12-0.20 mM and SDS-PAGE analysis was performed again to determine sample purity (Fig. 13b). Impurities were still present in the sample after ion exchange, ultrafiltration and dialysis; however, based on the limited amount of protein available and the work of previous project students the quality was deemed to be good enough for crystallization.

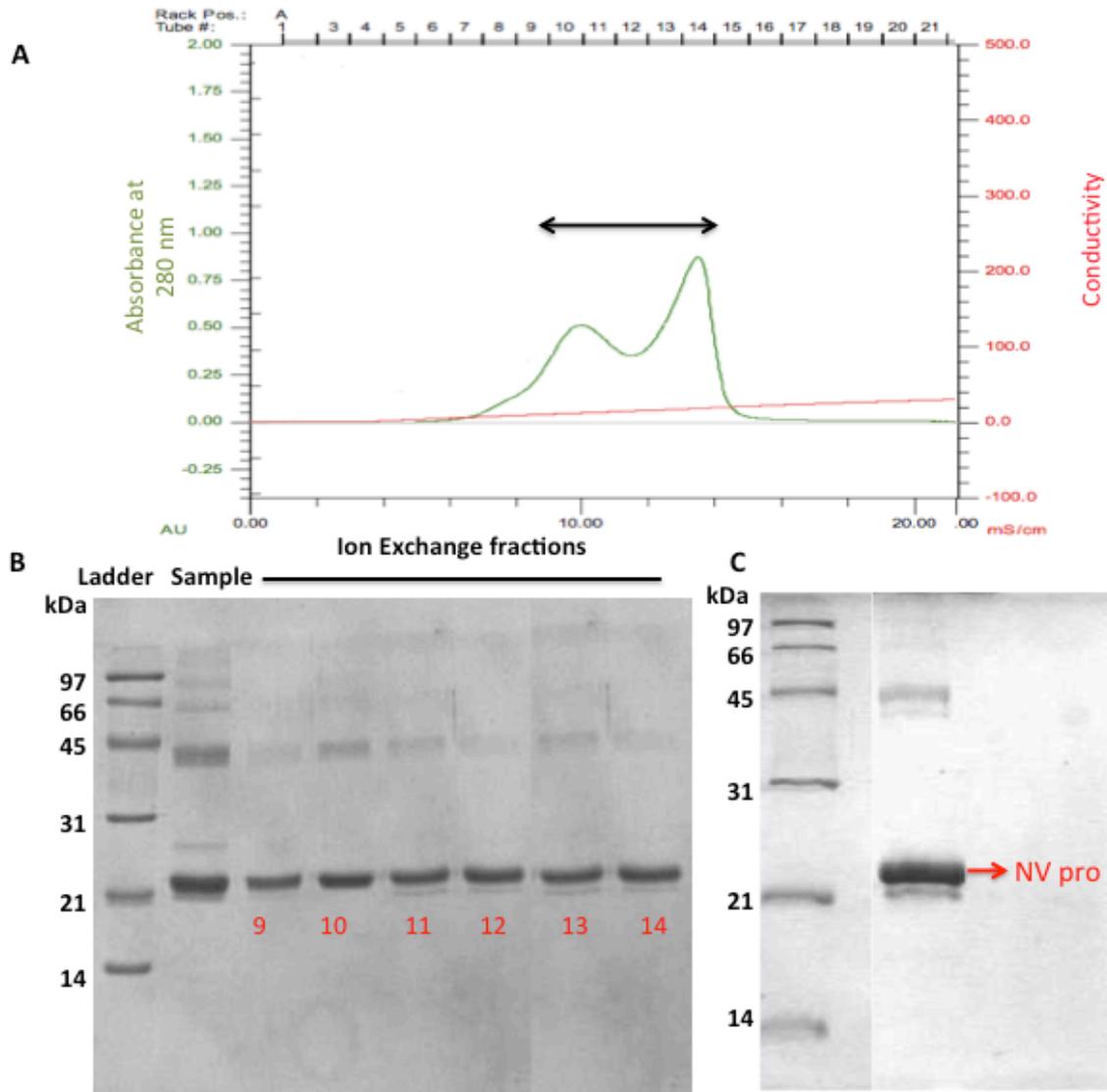


Figure 13: A) Ion exchange chromatogram of HuNV Pro (approx. 20 kDa). The elution profile of protein is shown in green and the arrow indicates fractions that were analyzed by SDS-PAGE. B) Fractions tested for purity by SDS-PAGE. The fractions (9-14) were pooled together, dialyzed and concentrated for crystallization experiments. HuNV Pro received from the collaborators was also loaded (sample next to MW ladder) to observe the improvement in sample purity. C) SDS-PAGE analysis of concentrated (3.0 mg/ml) HuNV Pro sample after purification.

## II. Screening and hit optimization

Since the conditions for crystallization of HuNV Pro in complex with RNA were unknown, a number of commercially available crystallization screens were tested. These screens were carried out in the presence of NVP1a or NVP2 RNA oligonucleotide. NVP1a screens were prepared and analyzed by the project students before I joined the lab, whereas I was responsible for NVP2 screening. Initial screening carried out using Index screen (Hampton Research) was unsuccessful as most drops were clear even after several weeks. The lack of crystal growth or precipitation suggested that either this particular sequence and length of RNA was not suitable for crystallization or the protein sample was under-saturated [64]. Therefore, for the next two screens, the concentration of protein was increased to avoid under-saturation. This increase in protein concentration resulted in the identification of a hit, which is mentioned in Table 4. The thiocyanate condition represented a weak hit characterized by very small, clumpy and extremely layered crystals, which are unsuitable for X-ray data measurement. As a result, the components of the thiocyanate condition were systematically varied and optimized to grow larger, unlayered individual crystals for high-resolution analysis by X-ray diffraction. Ideally further screening should have been conducted to identify additional hits. However, due to the limited amount of protein and RNA oligonucleotides available, the strategy was to focus on the optimization of the single hit identified.

Table 4: Crystallization hits identified from screens performed in the presence of HuNV Pro and RNA oligonucleotides. Previous project students were responsible for the identified of crystallization hit for NVP1a.

Screen (Well No.)	[Protein] mM	RNA	Hit condition
1. Index (H11)	0.12	NVP1a	0.1 M KSCN, 30% PEG MME 2000
2. Index	0.12	NVP2	No hit
3. Index (H11)	0.20	NVP2	0.1 M KSCN, 30% PEG MME 2000
4. Crystal	0.20	NVP2	No hit

Thiocyanate condition was successfully optimized and the crystals of HuNV Pro in the presence of either NVP1a or NVP2, were obtained overnight in 0.10-0.30 M Potassium thiocyanate, 0.10 M Tris-Cl, pH 7.5, and 32-36% PEG MME 2000 (Fig. 12). However, the X-ray diffraction data

collected for these crystals had irregular (smeary) reflections, which is an indication of crystal defects (Fig. 14). A high level of smearing makes it impossible to distinguish between individual diffraction reflections and since each reflection recorded contains information about the position of the atoms in the structure, it was impossible to generate electron-density maps from diffraction patterns showing high degree of smearing.

Defects in crystals are mainly due to the presence of impurities in protein sample and/or fast rate of crystallization [65, 66]. Rapid growth of crystal is frequently associated with the occurrence of packing disorder, lattice distortion, and the incorporation of impurities in the crystal [67]. Impurities can also create and/or enhance crystal disorder in several ways [66, 68, 65]. Since the amount of disorder and defects incorporated into a crystal is at least partially dependent on the rate of crystallization [65], one can typically reduce the amount of defects in a crystal by slowing down the rate of crystallization. The crystals grown in the presence of the original thiocyanate conditions had a very fast rate of crystallization (crystals appeared overnight), therefore, efforts were made to adjust the crystallization conditions in a manner that would slow down the rate of crystallization and reduce the amount of defects incorporated into growing crystals.

Several different approaches were tried to slow down the rate of crystallization including lowering of precipitant concentration, use of paraffin and silicon oil [69], seeding [70], lowering of precipitant concentration after nucleation has occurred [55], temperature variation [71], increase in drop size [72], variation in pH, and the use of additives to increase the solubility of protein [73]. Out of all the methods tried addition of additive (glycerol: 5-10%), use of oils (70  $\mu$ l silicon and 30  $\mu$ l paraffin oil placed on top of the reservoir solution) and lowering of precipitant concentration (28-32% PEG MME 2000) had the most profound affect (Table 5). The slowest rate of crystallization was observed for HuNV Pro-RNA crystals grown in the presence of 0.10 KSCN, 28% PEG MME 2000, 8% (w/v) glycerol, 0.10 M Tris-Cl, pH 7.5-8.5. These crystals appeared in the protein drop after 3-4 days, rather than overnight, and then continued to grow slowly for another 4 days. However, this decrease in the rate of crystallization did not improve crystal quality as X-ray diffraction data collected for all these crystals (Table 5) still had irregular (smeary) reflections (Fig. 15). Consequently, focus was shifted towards the other possible source of crystal disorder *i.e.* HuNV Pro sample purity.

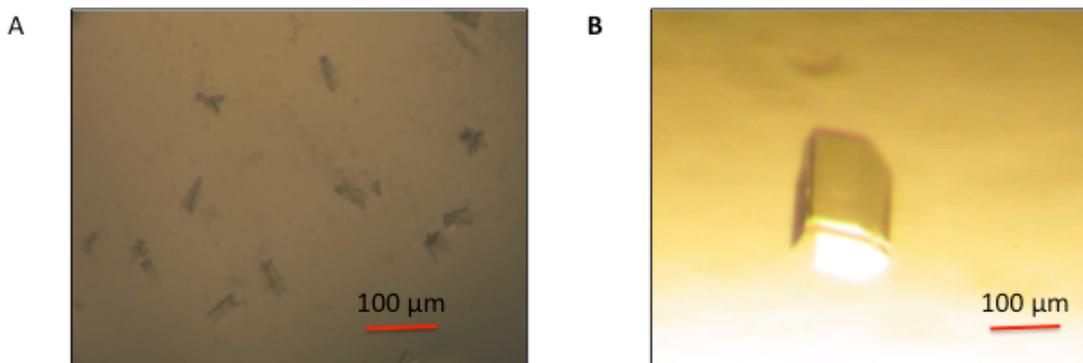
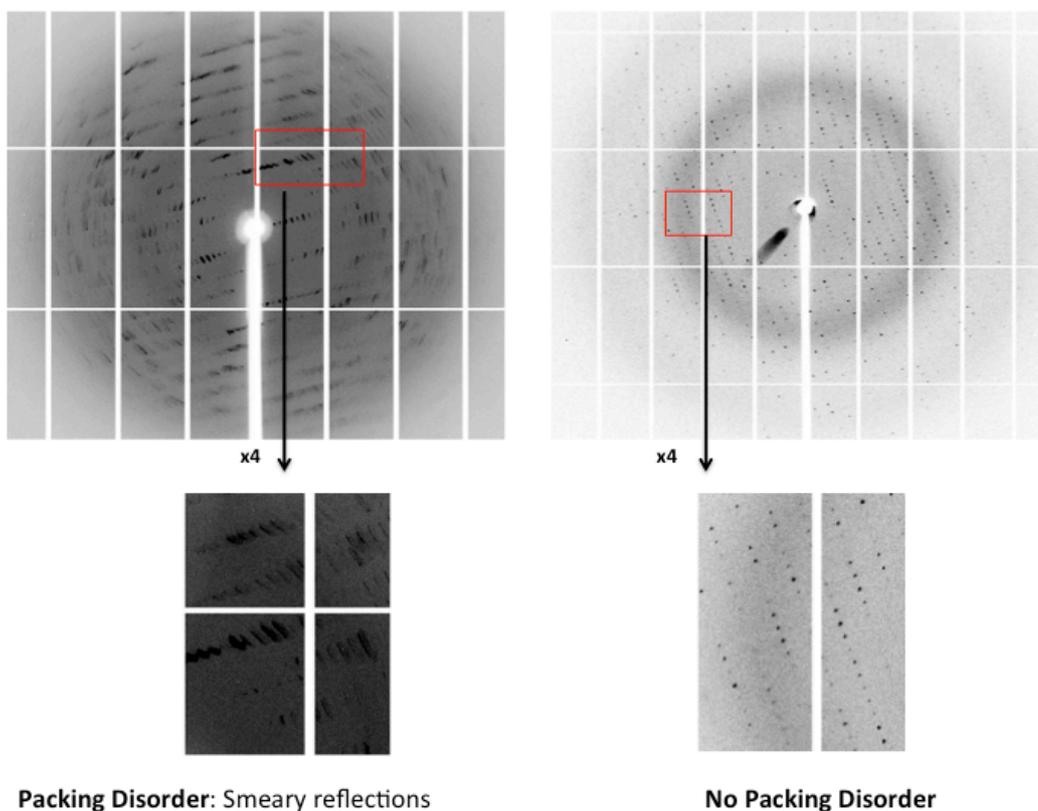


Figure 14: Optimization of thiocyanate condition. A) Small, clumpy and layered crystals obtained in Index screen. B) Several rounds of optimization by systematically varying crystallization parameters led to the growth of larger and individual crystals.



**Packing Disorder: Smearly reflections**                      **No Packing Disorder**  
 Figure 15: X-ray diffraction pattern of crystal with (left) and without defects (right). A small section of X-ray image is magnified by x4 to clearly represent the irregular (smeared) X-ray reflections associated with crystal defects.

Table 5: Summary of changes made to the original crystallization condition identified to slow down the rate of crystallization. All the crystallization experiments were conducted at room temperature, if not stated otherwise.

<b>Original Condition</b>	<b>Rate of Crystallization</b>
1. 0.10-0.30 M KSCN, 0.10 Tris-Cl, pH 7.5, 32-36% PEG MME 2000	Overnight
<b>Optimized Condition</b>	<b>Rate of Crystallization</b>
1. 0.10 M K/NaSCN, 0.10 Tris-Cl, pH 7.5-9.0, 28-34% PEG MME 2000, 4° C	Overnight
2. 0.05-0.10 M K/NaSCN, 0.10 Tris-Cl, pH 8.0-9.0 28-34% PEG MME 2000	1-2 days
3. 0.05-0.10 M K/NaSCN, 0.10 Tris-Cl, pH 7.5-8.5, 28-34% PEG MME 2000, Silicon oil: Paraffin oil (70/30 $\mu$ l)	2-3 days
4. 0.05-0.10 M K/NaSCN, 0.10 Tris-Cl, pH 7.5-8.5, 26% PEG MME 2000, 8% (w/v) glycerol	3-4 days

### III. Optimization of purification protocol

In an attempt to improve the purity of HuNV Pro sample, gel-filtration was added to the purification protocol. Ion-exchange fractions of highest purity were pooled together and loaded onto a pre-equilibrated gel-filtration column. The pattern of elution and the subsequent improvement in protein purity is shown in Fig. 16. Gel-filtration studies showed that HuNV Pro is present in different oligomeric states (predominantly monomer) in solution, which were collected separately. In order to identify the impurity present at approx. 45 kDa, the stained band resolved on the gel was excised with a new razor blade to avoid contamination and sent to the Southern Alberta Mass Spectrometry (SAMS) center, University of Calgary, for analysis. Surprisingly, the results (Appendix 6) revealed that the peptide is HuNV Pro with a MW of 40 kDa *i.e.* protease dimer. Based on this result, it is possible that the bands present at 66 and 97 kDa (Fig. 16) represent higher oligomeric states of HuNV Pro rather than non-protease impurities, as the difference in their MW is approx. 20 kDa, which is equivalent to MW of HuNV Pro monomer. The presence of NV Pro monomer and dimer in solution has been previously reported and it is thought that the concentration and/or the condition of the prepared Pro may influence the monomer-dimer equilibrium of NV Pro [74, 32, 41, 31].

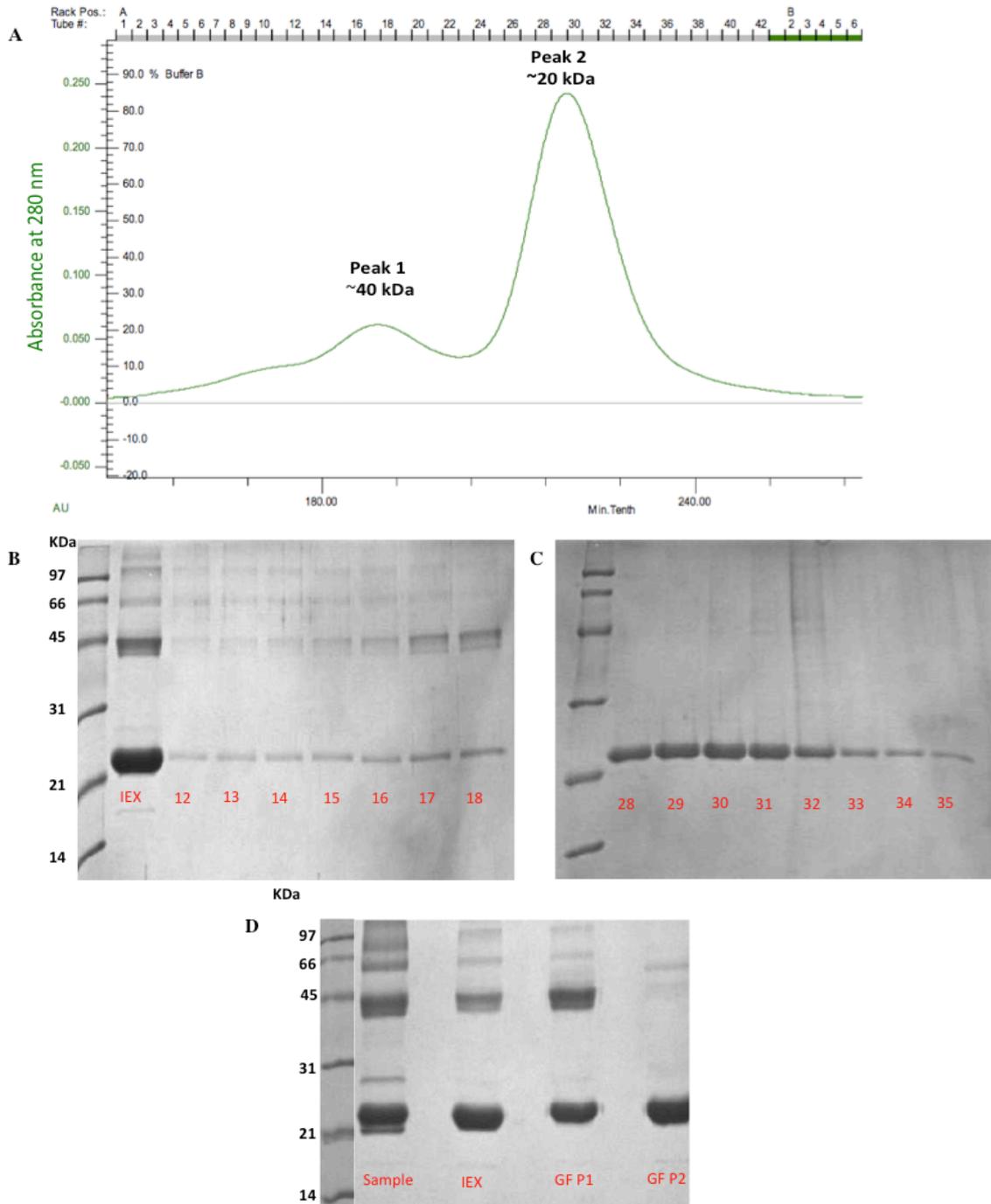


Figure 16: A) HuNV Pro gel-filtration elution profile. Peak 1 and 2 corresponds to HuNV Pro dimeric and monomeric forms. B and C) Fractions were tested for purity by SDS-PAGE and the fractions of highest purity *i.e.* fractions 13-17 (peak 1; B) and 28-33 (peak 2; C) were pooled separately and concentrated for crystallization experiments. D) SDS-PAGE analysis of concentrated protein samples after each purification step. (Sample: protein sample from our collaborators; IEX: sample after ion-exchange chromatography, and GF P1 and GF P2: concentrated peak 1 and peak 2 samples after gel-filtration).

#### IV. Optimization of Thiocyanate condition

HuNV Pro-RNA crystals were grown in the original thiocyanate condition (Table 5) to test if the improvement in purity had an effect on crystal quality. X-ray diffraction data collected for most of these crystals did not suffer from irregular reflections, confirming that the improvement in purity and homogeneity of the protein sample was responsible for the reduction in crystal defects. Protein-ligand complexes are usually sensitive to high salt concentration, since increasing ionic strength can interfere with the interactions needed to stabilize the complex [75]. Keeping this in mind, crystallization experiments were also conducted in the presence of low thiocyanate concentration. In addition, crystallization experiments were also carried out in the presence of non-thiocyanate salts (Table 6). Efforts to completely replace the potassium thiocyanate with other salts failed as the replacement of thiocyanate inhibited spontaneous nucleation and crystallization. Based on these results it was hypothesized that thiocyanate in crystallization conditions is necessary for spontaneous nucleation, but not for crystal growth.

Spontaneous nucleation is kinetically demanding and conditions optimal for nucleation are not ideal to support crystal growth. However, nucleation and growth can be uncoupled by using a technique called seeding [55], which was used as an additional optimization strategy. Seeding involves transfer of crystals or seeds that act as sites for nucleation, to the drops that cannot support spontaneous nucleation. Seed crystals (20-50  $\mu\text{m}$ ) grown in conditions containing thiocyanate and/or thiocyanate in combination with other salts, were transferred to a pre-equilibrated drop containing different precipitant (PEGs), additives, salt and pH etc. Unfortunately, seeding experiments did not yield any new crystals, but seeds in a few conditions did continue to grow. This observation supported our hypothesis that thiocyanate is necessary for spontaneous nucleation, but not for crystal growth. The best quality crystals were obtained when seeds were transferred to a drop containing 0.050-0.10 KCl, 22-26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5. These crystals grew to full size within 10 days after seeding (Fig. 17).

Table 6: Summary of salt combinations that yielded crystals.

Initial crystallization Condition		
Precipitant	Salt	Buffer pH
PEG MME 2000 (22-28%)	Potassium or Sodium thiocyanate (0.050-0.10 M)	Tris-Cl (pH 7.5-9.0)
Salt Optimization		
Precipitant	Salt Combination	Buffer pH
PEG MME 2000 (22-28%) PEG 3000 (22-28%)	Crystals were grown in the presence of Sodium/Potassium thiocyanate (0.010-0.10 M) and: <ol style="list-style-type: none"> <li>1. Sodium Chloride/Bromide/Flouride/Iodide (0.020-0.10 M)</li> <li>2. Potassium Chloride/Bromide/Iodide (0.050-0.10 M)</li> <li>3. Cesium Chloride (0.050 M)</li> <li>4. Calcium Chloride (0.025-0.050 M)</li> <li>5. Ammonium Sulfate/ Acetate (0.020-0.050 M)</li> <li>6. Potassium Acetate (0.020-0.050 M)</li> </ol>	Tris-Cl (pH 7.0-8.0)

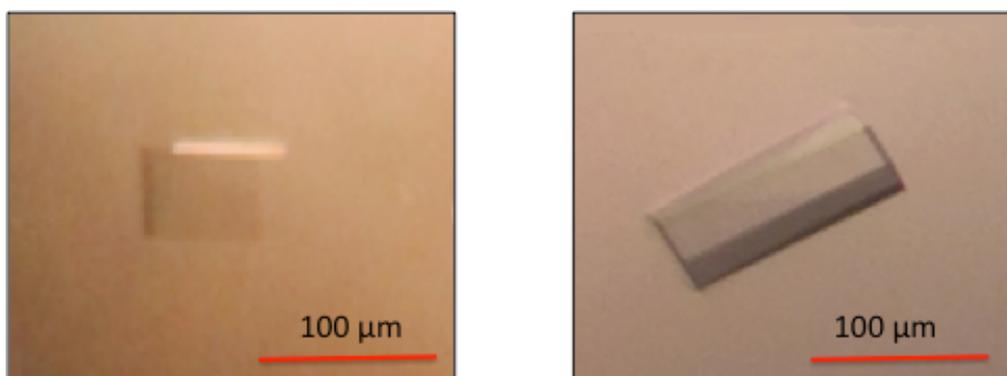


Figure 17: The change in size of seed crystal transferred to potassium chloride (KCl) condition. The crystal on the left shows the crystal seed after 2 days, whereas, the one on the right represents the full size achieved after a few more days.

## V. Soaking

HuNV Pro-RNA binding is weak ( $k_i$ :  $\mu\text{M}$  range) [28], therefore co-crystallization experiments were also performed in the presence of very high RNA concentration in order to saturate the RNA binding site on the protein. However, these experiments were not successful and RNA concentration above 2.5 mM was found to inhibit nucleation and growth. As a result crystals grown in the presence of 0.60-2.5 mM RNA were soaked to further increase RNA concentration to as high as 7.5 mM (40-fold molar excess over HuNV Pro). The absence of RNA in crystals grown and/or soaked in the presence of thiocyanate led to the hypothesis that thiocyanate salt is interfering with RNA binding. To test this hypothesis soaking technique was also used to gently exchange the thiocyanate with other salts and even the condition used in HuNV Pro-RNA binding studies was also tested [28] (Table 7). HuNV Pro proteolytic activity in HEPES has been reported to be at least 2-fold higher than in Tris-Cl [28]. Therefore, HEPES buffer was also used in soaking experiments, instead of Tris-Cl, to check if it promotes RNA binding to HuNV Pro.

Table 7: Summary of conditions used for soaking experiments.

<b>Initial Crystallization Condition</b>				
<b>Precipitant</b>	<b>Salt</b>	<b>Buffer</b>	<b>[RNA] mM</b>	
PEG MME 2000 (22-28%) PEG 3000 (22-28%)	K/NaSCN (0.020-0.20 M)	Tris-Cl (pH 7.5-8.9)	0.60-2.0	
<b>Soaking Condition</b>				
<b>Precipitant</b>	<b>Salt</b>	<b>Buffer</b>	<b>[RNA] mM</b>	<b>Time</b>
1. PEG MME 2000 (22-30%) PEG 3000 (22-30%)	KCl (0.020-0.10 M)	Tris-Cl (pH 7.0-8.0)	2.5-7.0	15 min to 48 hr.
2. PEG MME 2000 (22-30%) PEG 3000 (22-30%)	NaCl (0.020-0.10 M)	Tris-Cl (pH 7.0-8.0)	2.5-7.0	15 min to 48 hr.
3. PEG MME 2000 (22-30%)	NaCl/KCl (0.020- 0.10 M)	HEPES (pH 7.5-8.9)	2.5-7.0	15 min to 24 hr.
4. Condition used in HuNV Pro-RNA binding studies (10 mM HEPES pH 7.6, 0.10% CHAPS, 30% glycerol). The concentration of RNA used was 0.20-7.0 mM with soaking time of 15 min to 4 hr.				

## VI. X-ray data collection and structure determination

Due to limited time given for X-ray data collection at synchrotrons, only a fraction of crystals grown were sent for data collection. Out of 70 crystals sent, only 60 were screened and complete X-ray diffraction data set could only be collected for 25 crystals, as most of these crystals were grown before the purification was optimized and therefore suffered from severe crystal defects. The co-crystallization and soaking conditions along with the maximum resolution for few of the crystals for which complete X-ray data set was collected are mentioned in Table 8. The data was processed as explained earlier (page 28). All these crystals belonged to the space group  $P2_1$  with unit cell dimensions of  $a = 36.71\text{\AA}$ ,  $b = 34.70\text{\AA}$ , and  $c = 110.71\text{\AA}$ . Calculations revealed that the highest probability is for two molecules of HuNV Pro in the asymmetric unit with Matthews coefficient of  $2.21\text{ \AA}^3/\text{Dalton}$  and a solvent content of 22.6% [76]. Generally, crystal with low solvent content diffract better, as the crystal packing tends to be tighter [77]. This explains why the maximum resolution limit for most of the crystal is less than  $2.0\text{ \AA}$  (Table 8). Due to high quality of X-ray data collected for most of the crystals, electron-density maps showed almost the entire polypeptide chain for each of the two molecules in the asymmetric unit including the highly flexible solvent exposed loop (123-134 aa), which is missing in most of the NV Pro structures deposited in protein data bank (PDB). However, electron density for His-tag and C-terminal tail (residues 175-181) was very weak, presumably due to disorder, and so was not included in the HuNV Pro models built. All the 3D models were refined to an R-factor and R-free values of  $\leq 24\%$  and  $\leq 28\%$ . After refinement, visual inspection of all of the models was carried to find electron-density not accounted for by the protein into which RNA oligonucleotide could be modeled. Disappointingly, however, no electron-density was observed for either NVP1a or NVP2 RNA oligonucleotides.

The most important details of data collection parameters and refinement statistics for one of the models (AS59; condition no. 7 in Table 8) are summarized in Table 9. The structural analysis was carried out for all the models built. However, all the models built were found to be structurally identical as indicated by the small root mean square deviation (RMSD) in  $C\alpha$  atom positions of only  $0.144\text{ \AA}$ . Therefore, only one of the structures, AS59, is discussed in detail below.

Table 8: X-ray diffraction data sets collected for co-crystals grown and/or soaked in different conditions. All the crystals were co-crystallized in the presence of RNA, if not stated otherwise.

<b>Crystallization Condition</b>	<b>Soaking Condition and Time</b>	<b>RNA</b>	<b>Resolution Å</b>
1. 0.15 M NaSCN, 36% PEG MME 2000, 0.10 M Tris-Cl, pH 8.5	-	NVP1a (0.6 mM)	2.1
2. 0.10 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5	-	NVP1a (1.5 mM)	2.3
3. 0.10 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 8.0	0.10 M KCl, 27% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5; 30 min	NVP1a (2.5 mM)	2.5
4. 0.10 M NaSCN, 24% PEG MME 2000, 0.10 M Tris-Cl, pH 8.0 (Native)	0.025 M NaCl, 0.025 M NaSCN, 0.003 M MgCl <sub>2</sub> 0.01M Tris-Cl, pH 7.5; 24 hr	NVP1a (7.5 mM)	2.0
5. 0.070 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 8.0	-	NVP2 (2.5 mM)	2.3
6. 0.040 M NaSCN, 0.050 M KCl, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5	-	NVP2 (2.5 mM)	1.9
7. 0.050 M NaSCN, 0.050 M KCl 26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5 (AS59)	-	NVP2 (2.5 mM)	1.6
8. 0.10 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5	0.05 M KCl, 27% PEG MME 2000, 0.01M Tris-Cl, pH 7.5; 30 min	NVP2 (7.5 mM)	1.9
9. 0.15 M NaSCN, 36% PEG MME 2000, 0.10 M Tris-Cl, pH 8.0	0.025 M KCl, 37% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5; 3 hr	NVP2 (7.5 mM)	1.7
10. 0.20 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5 (Native)	0.05 M KCl, 0.10 M, 27% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5; 24 hr	NVP2 (7.5 mM)	1.8

Table 9: Crystallographic data collection and model refinement statistics for AS59. This particular data set was collected at Stanford Synchrotron Radiation Lightsource (SSRL). Values in the brackets represent the R-values for unrefined structure.

<b>Data-integration Statistics</b>	
Space-group	P2 <sub>1</sub>
a, b, c (Å)	37.45 36.18 113.79
$\alpha, \beta, \gamma$ (°)	90.00 98.07 90.00
Unit cell volume (Å <sup>3</sup> )	152511382.9
Solvent content (%)	22.63
<b>Data Collection</b>	
Temperature (K)	100
Detector	PILATUS 6M
No. Of Images	500
Oscillation range (°)	0.50
Wavelength (Å)	0.9795
Resolution (Å)	34.45-1.60
Total observation	206529
Unique reflections	46209
Multiplicity	4.47
I/ $\sigma$ <sub>1</sub>	10.27
Completeness	95.8 %
<b>Refinement</b>	
Resolution (Å)	1.60
R <sub>work</sub>	0.2036 (0.3102)
R <sub>free</sub>	0.2339 (0.3262)
RMS deviation – bonds (Å)	0.0069
RMS deviation- Angles (°)	1.098
Ramachandran plot	
Favored	97.63%
Allowed	2.37%

## VII. Structural analysis and discussion of AS59

As mentioned earlier, there are two molecules in the asymmetric unit (Dimer; Fig. 18). AS59 dimer interface is maintained primarily by hydrophobic interactions, resulting in a total buried surface area of  $909 \text{ \AA}^2$ , which is suggestive of a very weak non-physiological dimer (*i.e.* result of crystal packing) [78, 79]. The dimerization of NV Pro has been previously reported in crystal structures [31, 51, 32]. AS59 dimer interface contacts were compared with other NV Pro dimers using POLYVIEW-3D [80]. Despite the fact that the amino acid sequence is highly conserved in norovirus proteases, the dimer interface contacts were found to be significantly different (Fig. 19). It is unlikely that the biologically relevant dimer would have different contacts as observed in NVs Pro dimers. Therefore, further mutational and functional studies are needed to understand NV Pro dimerization and its biological significance in NV life cycle, if any.

The molecules present in the dimer are very closely related structurally as indicated by the small root mean square deviation (RMSD) in C $\alpha$  atom positions of only  $0.4 \text{ \AA}$  (Fig. 20; monomer A and B). Therefore, only one of the two molecules was used for structural comparison of AS59 with other NV proteases. The overall structure of AS59 monomer is nearly identical to previously solved NV Pro structures, which comes as no surprise given the high degree of protease sequence similarity in NVs ( $\geq 60\%$ ; Fig 21) [38]. As expected, it consists of an incomplete  $\beta$ -barrel N-domain and an anti-parallel  $\beta$ -sheet C-domain. The active site residues are located in the cleft formed at the interface of these two domains (Fig. 20). One noticeable difference observed in AS59 structure was the conformation of solvent exposed loop (121-134 aa; Fig. 21). However, the conformational flexibility of this loop is well documented and is possibly an artifact of crystal packing [38, 31, 51, 32].

Low occupancy of ligands result in less well-defined electron density map for ligands [81]. Therefore, it could be argued that RNA is not visible in our structures due to its low occupancy. However, very high structural similarity of high-resolution structures of HuNV Pro-RNA co-crystals (grown in the presence of as high as 60-fold molar excess of RNA to protein) to HuNV Pro native structures support the lack of RNA binding in our crystallization studies. The only variable regions in these structures are solvent exposed loop (121-134 aa) and His-tag (Fig. 22). The conformational variation of these residues is well documented and cannot be attributed to

RNA binding [31, 38, 32]. Moreover, the loop also lacks positively charged residues required to bind RNA. Therefore, based on the absence of RNA electron density in our models and structural similarity between native and co-crystal structures, it can be concluded that RNA oligonucleotides did not bind to HuNV Pro in our crystallization studies.

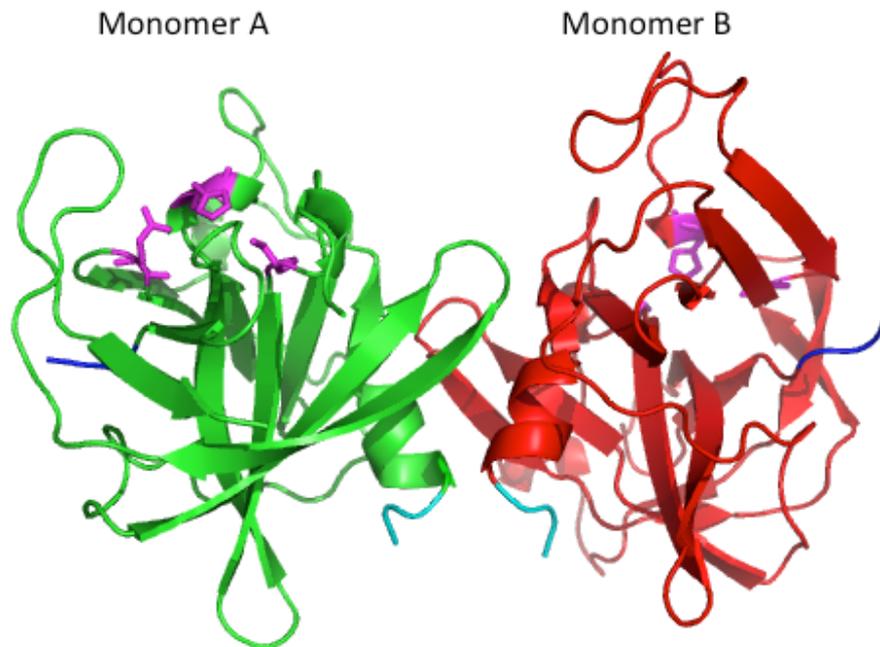


Figure 18: The cartoon representation of dimeric arrangement observed in AS59. The active site (H30, C139 and E54) is shown as magenta sticks, whereas the C and N-terminal residues are colored cyan and blue.

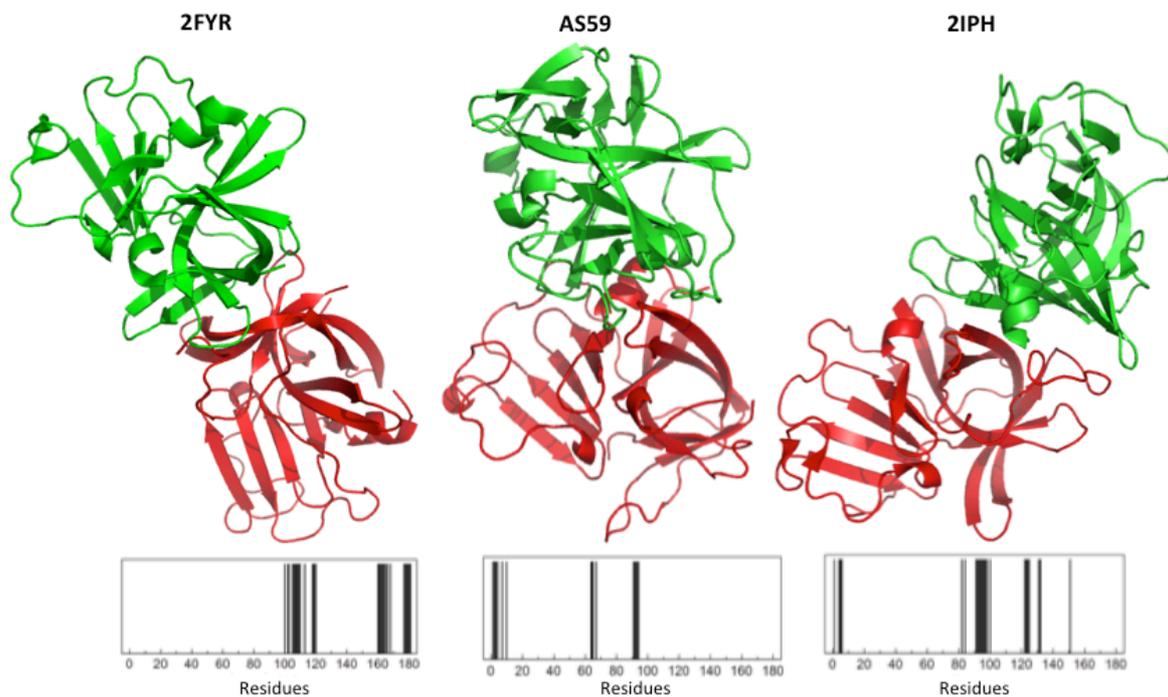


Figure 19: Comparison of dimer structures and interface of NV Pro dimers (PDB ID: 2FYR and 2IPH) with AS59. Plotted below each dimer are residues forming dimer interfaces identified using POLYVIEW-3D software [80].

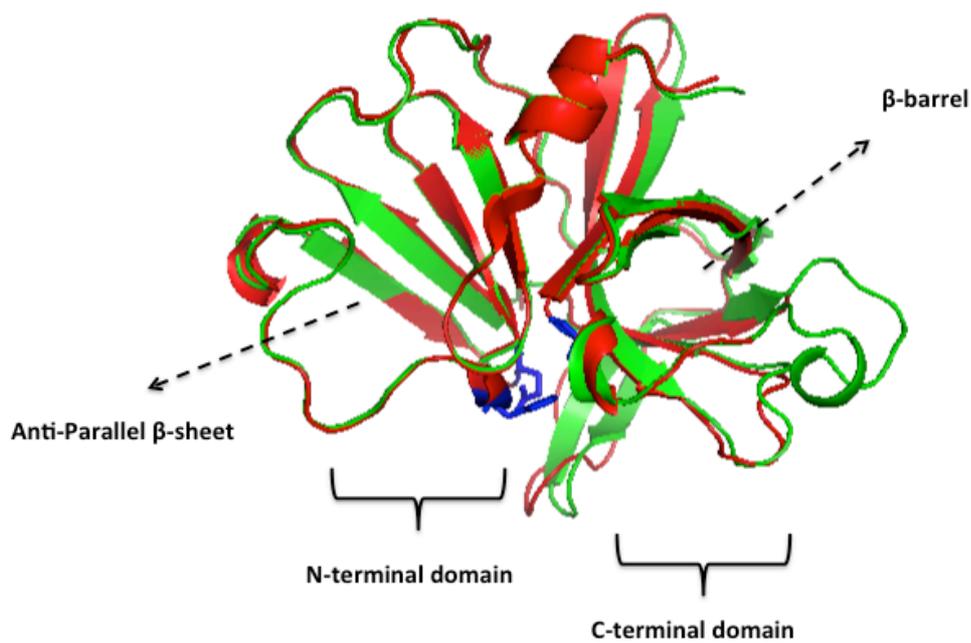


Figure 20: Overlay of AS59 monomer A (Red) and B (green). The side-chains of the active site residues (His 30, Gly54 and Cys139) are shown as sticks (blue).

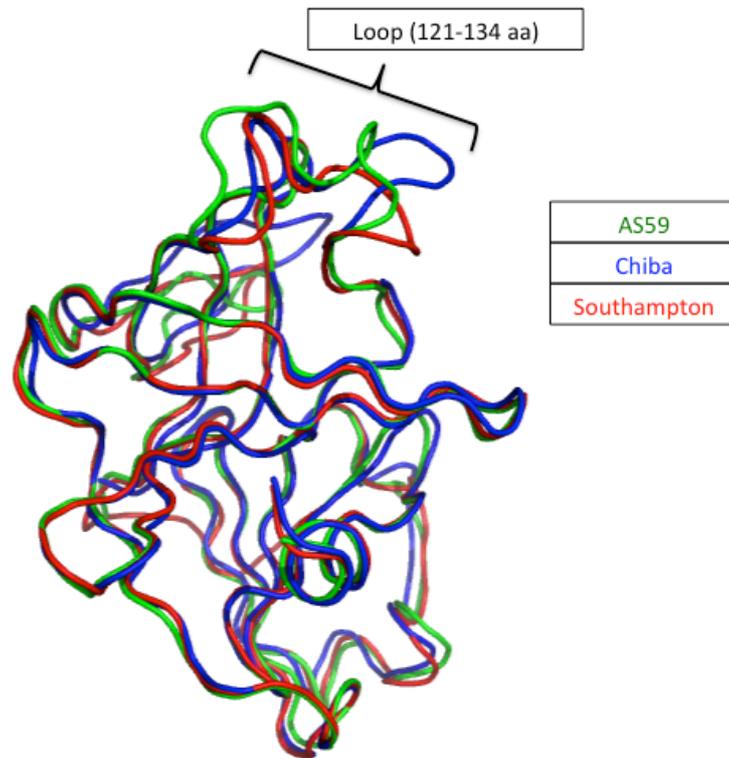


Figure 21: Structural comparison of the backbone of AS59 with other HuNV proteases (PDB ID: 1WQS (Chiba) and 2IPH (Southampton)). The overall structure is very similar, apart from the flexible loop composed of residues 121 to 134. The average root mean square deviations (RMSD) of the backbone atoms of Chiba and Southampton from AS59 are 0.7 and 0.4.

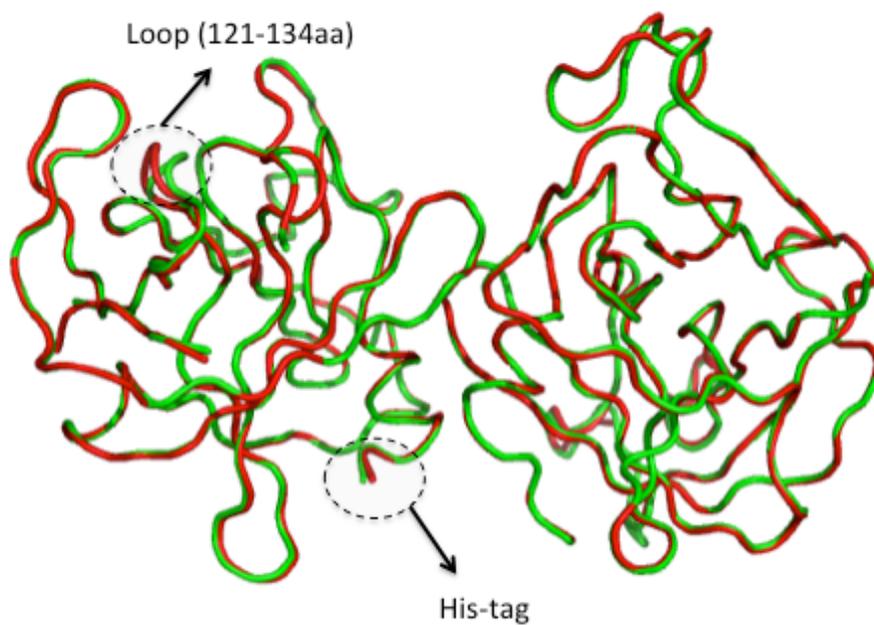


Figure 22: Structural comparison of the backbone of HuNV Pro crystal grown in the presence (Green) and absence of RNA oligonucleotides (Red). The calculated average root mean square deviation of the backbone atoms of these two structures is 0.14 Å. The most variable regions are labeled (Loop and His-tag).

## b. HuNV Pro-peptide interaction

### i. Expression and purification

Our collaborators successfully expressed mutant HuNV Pro (C139A) as His<sub>6</sub>-Pro fusion protein (approx. 20 kDa). Mutant HuNV Pro was purified using a purification protocol established for the wild type protease, which involved ion-exchange chromatography followed by gel-filtration. The gel-filtration elution profile is presented below (Fig. 23). The elution profiles for wild type and mutant HuNV Pro are identical and therefore it was assumed that the peak 1 and peak 2 represent monomeric and dimeric forms of protein. Peak fractions were analyzed for purity by SDS-PAGE (Fig. 23), and the fractions of the highest purity were pooled together, dialyzed against low salt storage buffer and then concentrated by ultrafiltration to 0.20 mM.

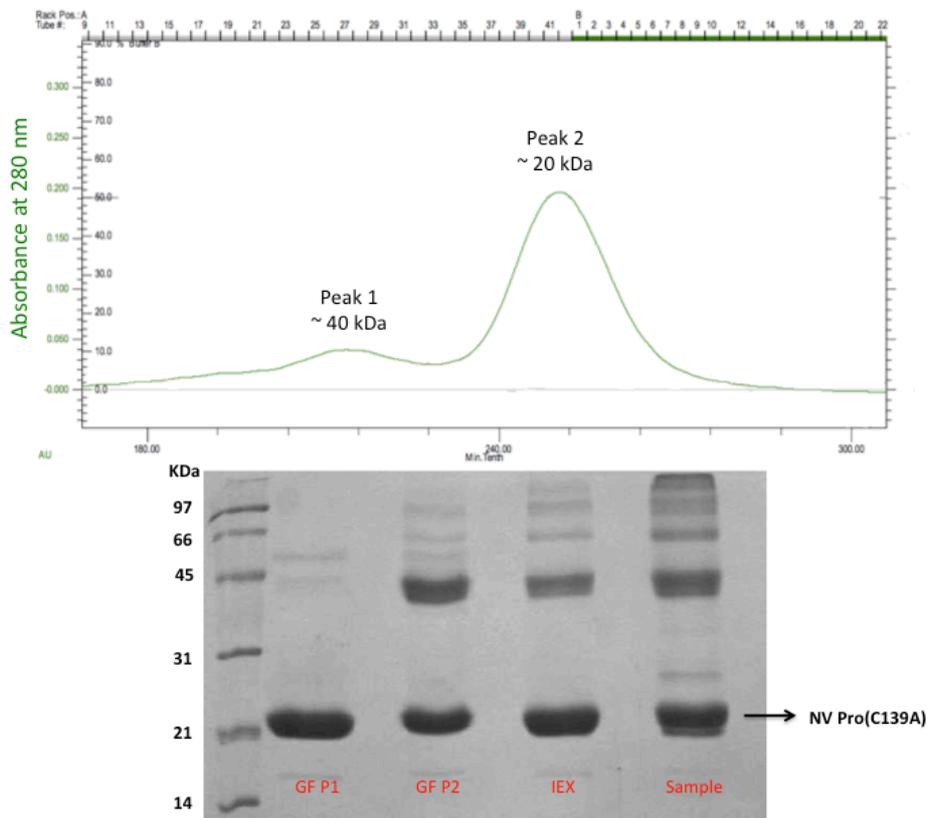


Figure 23: Mutant HuNV Pro gel-filtration elution profile (*Top*). SDS-PAGE analysis of concentrated protein samples after each purification step (*Bottom*). (Sample: protein sample received from our collaborators; IEX: sample after ion-exchange chromatography, and GF P1 and P2: concentrated peak 1 and peak 2 samples after gel-filtration).

## II. Screening and optimization

In order to identify ideal crystallization conditions of mutant HuNV Pro in complex with peptides, a number of commercially available crystallization screens were tested (Appendix 4: Sparse matrix screening). These screens were carried out in the presence of either long or short synthetic peptides, and crystallization hits were obtained for both peptides in a number of conditions, which are mentioned in Table 10. Almost all the crystallization conditions identified were successfully optimized manually to obtain large clean crystals for X-ray data collection (Fig. 24 and Table 11). It was also observed that co-crystallization conditions were not sensitive to the presence of a particular peptide as crystallization conditions yielded crystals with both short and long peptides. This is not surprisingly considering that the crystallization hits identified for both short and long peptides are similar. Even though co-crystallization studies were successful in the presence of 30-fold molar excess of peptide to protein, soaking was used to further increase the peptide concentration up to 20 mM (100-fold molar excess to mutant HuNV Pro).

Peptide-enhanced Pro binding to RNA has been reported in distantly related HAV 3C [43]. Therefore, co-crystallization experiments were also carried out in the presence of mutant HuNV Pro, peptide and RNA oligonucleotides using the conditions mentioned in Table 10. Currently, these conditions are being optimized to grow large clean crystals.

All the structures of mutant HuNV Pro-peptide crystallized with C-terminal of one Pro molecule being inserted into the peptide-binding groove of another molecule, thereby preventing binding of synthetic peptides. To circumvent this problem, synthetic peptides, both short and long, were also co-crystallized with wild HuNV Pro, which always crystallized as a dimer with unoccupied active sites (Fig. 18). These experiments were conducted in the presence of very high concentration of RNA oligonucleotides to inhibit the proteolysis of peptide by wild type HuNV Pro (Table 12).

Table 10: Crystallization hits identified from screens performed in the presence of mutant HuNV Pro and different peptides.

Screen (Well No.)	[Protein] mM	Peptide	Hit condition
1. Index (C6)	0.20	Long	0.1 M NaCl, 0.1 M Bis-Tris pH 6.5, 1.5 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
2. Index (H11)	0.20	Long	0.1 M KSCN, 30% PEG MME 2000
3. JCSG+ (G4)	0.20	Long	0.2 TMAO, 0.1 Tris pH 8.5, 20% (w/v) PEG MME 2000
4. PEGRx (E11)	0.20	Short	1.8 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.1 M Bis-Tris pH 6.5, 2% (w/v) PEG MME 550
5. JCSG+ (G9)	0.20	Short	0.1 M KSCN, 30% PEG MME 2000

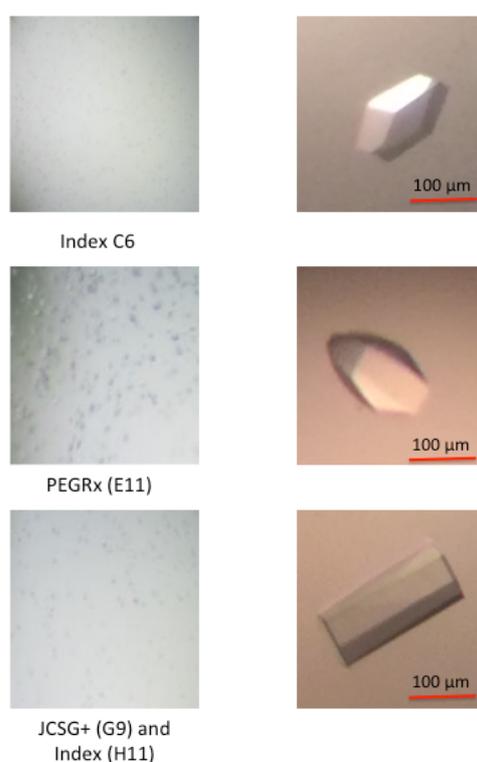


Figure 24: HuNV Pro-peptide crystals before (left) and after optimization (right). JCSG+ (G4) crystallization condition could not be reproduced manually.

Table 11: Summary of the changes made to the crystallization hits to grow diffraction quality crystals.

Hit Condition	Optimized Condition
Index (C6)	0.10 M NaCl, 0.10 M Bis-Tris pH 6.5-7.0, 1.30-1.60 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 5-10% (w/v) glycerol
PEGRx (E11)	1.60-1.65 (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.10 M Bis-Tris pH 6.5-7.5, 0.50-3.0% PEG 550, 6-10% (w/v) glycerol
JCSG+ (G9) and Index (H11)	0.050-0.10 M KSCN, 0.10 M Tris-Cl pH 8.5-8.9, 22-28% (w/v) PEG MME 2000 or 20-26% PEG 6000

### III. X-ray data collection and structure determination

Due to limited time given for X-ray data collection at synchrotrons, only a fraction of crystals grown were sent for data collection. Out of 65 crystals sent, only 40 were screened and complete X-ray diffraction data could only be collected for 20 crystals, as all the crystals grown in the presence of ammonium sulfate diffracted to roughly 5.0 Å resolution. At this resolution it is impossible to distinguish peptide from the HuNV Pro [82]. Therefore, complete X-ray diffraction data sets were not collected for crystals grown in the presence of ammonium sulfate. The co-crystallization and soaking conditions along with the maximum resolution for few of the crystals for which complete X-ray data sets were collected are given in Table 12. The data was processed and refined as explained earlier (page 26). HuNV Pro-peptide crystal structures were refined in the space group P2<sub>1</sub> with 4 molecules per asymmetric unit. Unfortunately, electron density was not observed for either short or long peptide in all of the models built. Instead, crystal packing in all these structures resulted in the insertion of the flexible C-terminal tail (residues 177-181) of one Pro molecule into the peptide-binding groove of another molecule in the crystal.

The C-terminal tail residues were built completely in only one of the models, AS106 (Table 12: condition no. 2), and this is discussed in detail below. However, for two monomers in the asymmetric unit the electron density for the His<sub>6</sub>-tag along with the solvent-exposed loop formed by residues 121-134 and C-terminal tail (residues 175-181) was weak, presumably due to disorder, and so was not included in the model. The flexibility of these residues is consistent with the previously solved structures of NV proteases [38, 31, 30]. The most important details of data collection parameters and refinement statistics for AS106 are summarized in Table 13.

As expected, wild type HuNV Pro crystallized in space group  $P2_1$  with 2 molecules in the asymmetric unit. Disappointingly, no electron density was observed for either RNA or peptide. This supports the results of HuNV Pro-RNA crystallization studies, as RNA binding would have inhibited the proteolytic activity of the wild type Pro, thereby allowing the binding of synthetic peptides into the active site. Therefore, the lack of observed electron density for peptides is presumably due to proteolysis by the wild type HuNV Pro. Since these structures are almost identical to wild type HuNV Pro dimers mentioned earlier they are not discussed below (Appendix: 5).

Table 12: X-ray data sets collected for HuNV Pro-peptide crystals grown and/or soaked in different conditions. All the crystals were co-crystallized in the presence of either short or long peptide, if not stated otherwise.

Crystallization Condition	Soaking Condition and Time	Peptide (mM)	Resolution Å
1. 0.10 M NaSCN, 23% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9	-	Long (2 mM)	2.0
2. 0.05 M NaSCN, 24% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9 (AS106)	-	Long (6 mM)	1.9
3. 0.10 M NaSCN, 24% PEG MME 2000, 0.10 M Tris-Cl, pH 8.5	0.10 M NaSCN, 25% PEG MME 2000, 0.10 M Tris-Cl, pH 8.5; 2 hr	Long (20 mM)	2.8
4. 0.10 M NaSCN, 25% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9	0.05 M NaSCN, 26% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9; 30 min	Long (20 mM)	2.5
5. 0.10 M NaSCN, 28% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9		Short (20 mM)	2.4
6. 0.05 M NaSCN, 30% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9	0.05 M NaSCN, 31% PEG MME 2000, 0.10 M Tris-Cl, pH 7.5; overnight	Short (6 mM)	2.3
7. 0.050 M NaSCN, 30% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9	0.050 M NaSCN, 31% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9	Short (20 mM)	1.7
8. 0.10 M NaSCN, 25 % PEG MME 2000, 1% (w/v) glycerol; 0.10 M Tris-Cl, pH 7.5	0.10 M NaSCN, 26% PEG MME 2000, 1% (w/v) glycerol; 0.10 M Tris-Cl, pH 7.5. Crystal was soaked for 2 hr with RNA followed by peptide for 15 min.	NVP2 (5 mM) Long (20 mM)	1.9
9. 0.10 M NaSCN, 30% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9, 1% (w/v) glycerol. Wild type HuNV Pro was co-crystallized in the presence of NVP2 and small peptide.	0.10 M NaSCN, 30% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9, 1% (w/v) glycerol. Crystal was soaked with RNA followed by peptide for 30 min.	NVP2 (5 mM) Short (20 mM)	1.7
10. 0.10 M NaSCN, 30% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9. Wild type HuNV Pro was co-crystallized in the presence of NVP2 and long peptide.	0.10 M NaSCN, 32% PEG MME 2000, 0.10 M Tris-Cl, pH 8.9. Crystal was soaked with RNA followed by peptide for 2 hr.	NVP2 (5 mM) Long (20 mM)	1.8

Table 13: Crystallographic data collection and model refinement statistics for AS106. This particular data set was collected at Stanford Synchrotron Radiation Lightsource (SSRL). Values in the brackets represent the R-values for unrefined structure.

<b>Data-integration Statistics</b>	
Space-group	P2 <sub>1</sub>
a, b, c (Å)	73.97 35.83 111.54
$\alpha, \beta, \gamma$ (°)	90.00 101.33 90.00
Unit cell volume (Å <sup>3</sup> )	289784.8
Solvent content (%)	38.89
<b>Data Collection</b>	
Temperature (K)	100
Detector	PILATUS 6M
No. Of Images	500
Oscillation range (°)	0.50
Wavelength (Å)	0.9795
Resolution (Å)	55.63-1.90
Total observation	235132
Unique reflections	50477
Multiplicity	4.65
I/ $\sigma$ <sub>1</sub>	7.47
Completeness	97.49 %
<b>Refinement</b>	
Resolution (Å)	1.90
R <sub>work</sub>	0.1845 (0.32941)
R <sub>free</sub>	0.2324 (0.35373)
RMS deviation – bonds (Å)	0.0070
RMS deviation- Angles (°)	1.054
Ramachandran plot	
Favored	96.89%
Allowed	3.11%

#### IV. Structural analysis and discussion of AS106

There are four HuNV Pro monomers (designated A, B, C and D) per asymmetric unit (Fig. 25). The overall structure of each monomer is nearly identical to each other (Fig. 26) and previously solved structures of NV Pro (Fig. 27) with C-terminal tail and the solvent exposed loop (121-134aa) being the most variable regions. Interestingly, crystal packing allowed the flexible C-terminal tail *i.e.* residues 177-181, corresponding to positions P5 to P1 of the native substrate, to associate with the peptide-binding groove of the neighboring Pro molecule in the crystal, thereby generating a *trans* protease-product complex, which represents the interaction of the HuNV Pro at the ProPol junction (Fig. 25). In fact, similar adventitious crystallization had previously provided structural insight into peptide recognition in HuNV proteases (PDB ID: 4IN1 and 4IN2) [37]. Superimposition of 4IN2 with AS106 and the analysis of peptide interaction with HuNV proteases using LigPlot+ [83] revealed that the conformation and the interaction of P5-P1 residues observed in the peptide-binding grooves are identical (Fig. 28). The conformation of the peptide residues beyond P5 differs in these structures due to different crystal packing in AS106 and 4IN2.

The structure explains why we did not observe electron density for synthetic peptides used in the crystallization experiments. The lack of binding in HuNV Pro monomer A and B is evidently due to lack of empty peptide-binding groove (Fig. 25), whereas for molecules C and D it is not clear. Even though the C-terminus in neighboring molecules of monomer C and D could not be modeled beyond residue 173, it is possible that inherently flexible C-terminal tails of these molecules, due to their close proximity to the peptide-binding grooves of monomer C and D, is interfering with peptide binding (Fig. 29). The flexibility of C-terminus in NV proteases is well documented [38, 31] and it may be necessary to allow the Pro to access the cleavage sites during polyprotein processing [36]. Therefore, it can be concluded that we failed to solve HuNV Pro-peptide complexes due to C-terminal tail residues either occupying the peptide-binding groove or interfering with synthetic peptide binding.

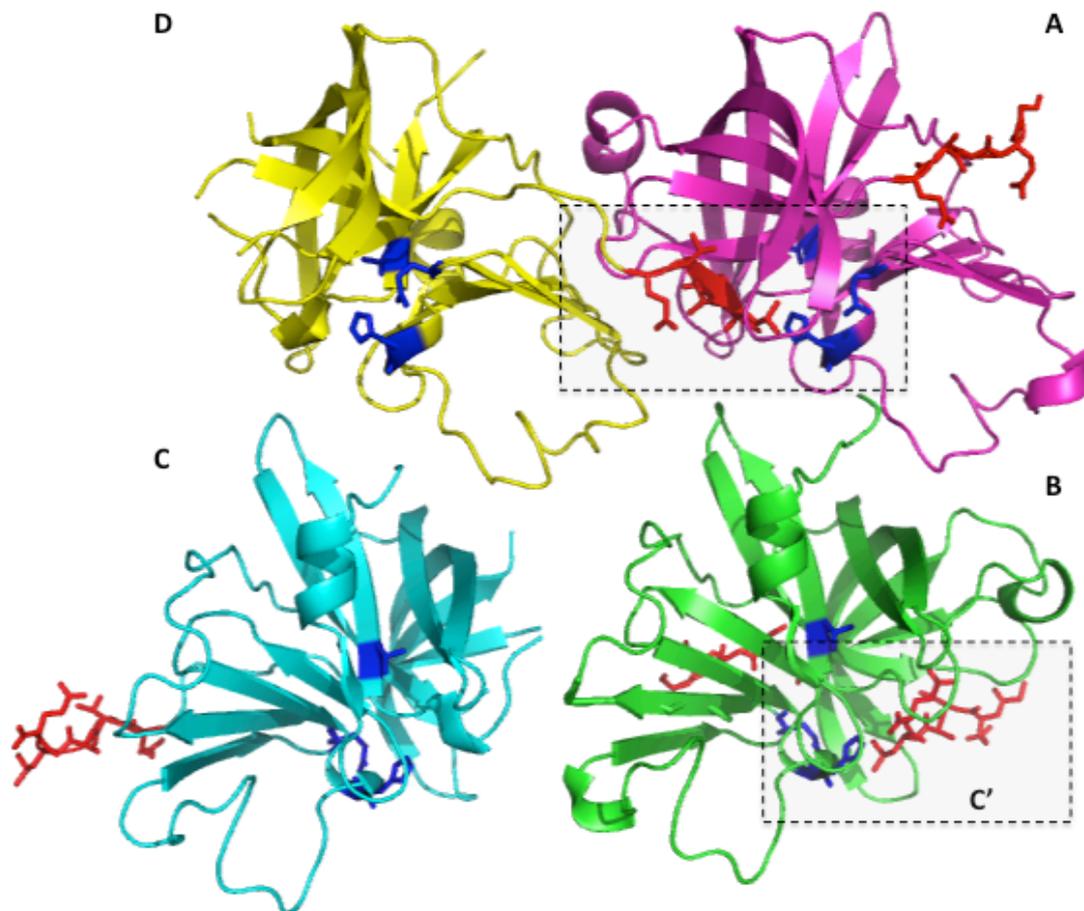


Figure 25: Crystal packing of AS106 in the  $P2_1$  space group, with four monomers in the asymmetric unit (A, B, C and D). The Pro is represented as cartoon diagram, while active site and C-terminal residues are shown as blue and red sticks. The C-terminal tail of molecule D (red stick) is inserted into the peptide-binding groove of monomer A, whereas the neighboring molecule called C' (only C-terminal tail shown as sticks) interacts with the peptide-binding groove of monomer B. The HuNV Pro-tail interactions are shown in dashed boxes.

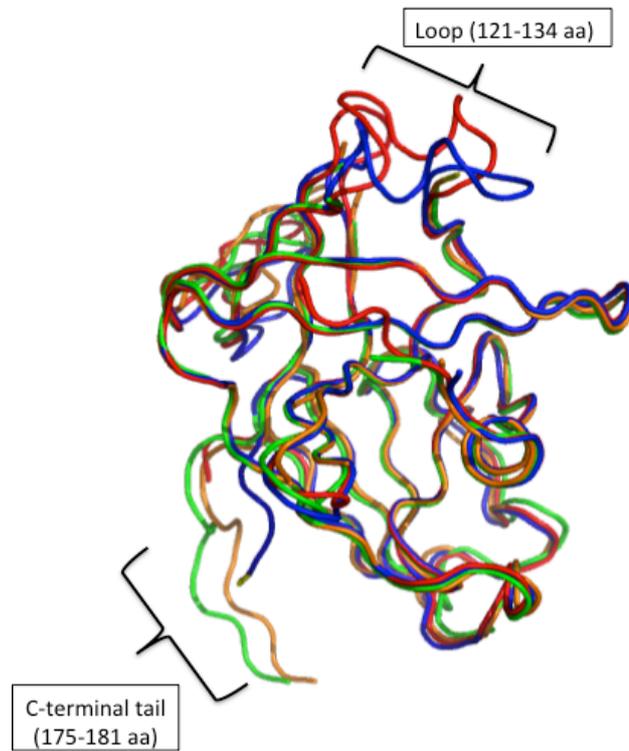


Figure 26: Superimposition of the backbone of AS106 monomers. Average root mean square deviation (RMSD) of the backbone atoms of all the monomer is less than 0.4 Å. The regions with the most variation most are labeled (C-terminal tail and solvent exposed loop).

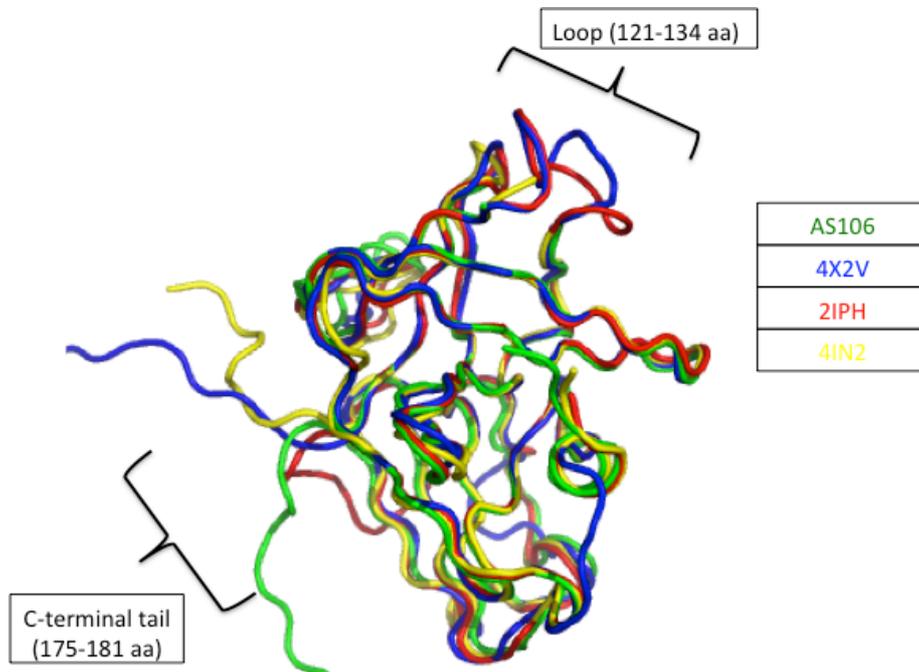


Figure 27: Structural comparison of the backbone of AS106 monomer with other NV proteases (PDB ID: 4X2V, 2IPD and 4IN2). The inherently flexible C-terminal tail and the solvent exposed loop (121-134) are labeled. The RMSD for all the structures is less than 0.6 Å.

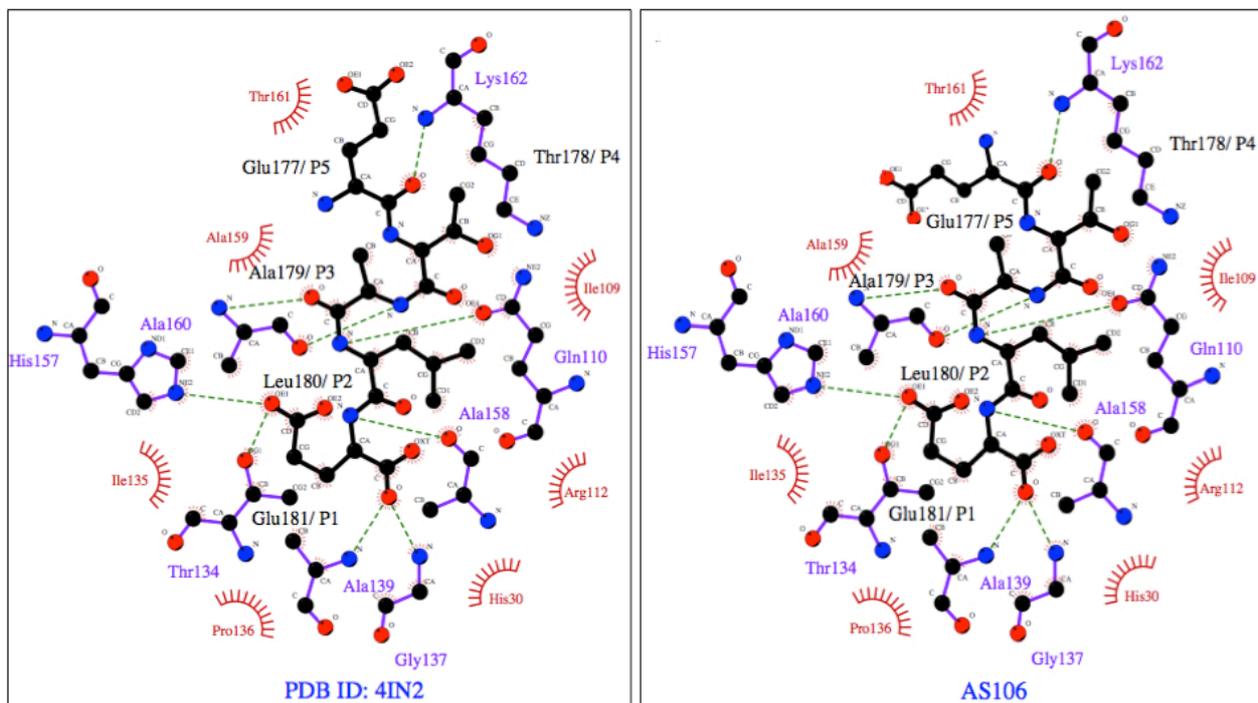
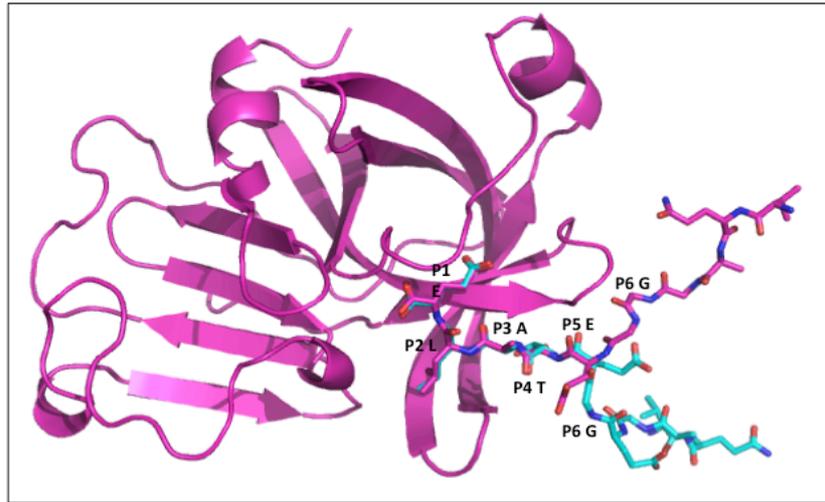


Figure 28: Structural comparison of C-terminal tail interactions observed in HuNV Pro-peptide complexes. (*Top*) Superimposition of C-terminal tail observed in 4IN2 (cyan) and AS106 (magenta) complex structures. C-terminal tails, which represent native peptide substrates, are shown as sticks. Only peptide residues up to P6 are labeled. (*Bottom*) Schematic diagram of HuNV Pro-peptide (P1-P5; black sticks) interactions generated using Ligplot+. The hydrogen bonds are shown as green dotted lines, while the spoked arcs represent residues involved in hydrophobic contacts.

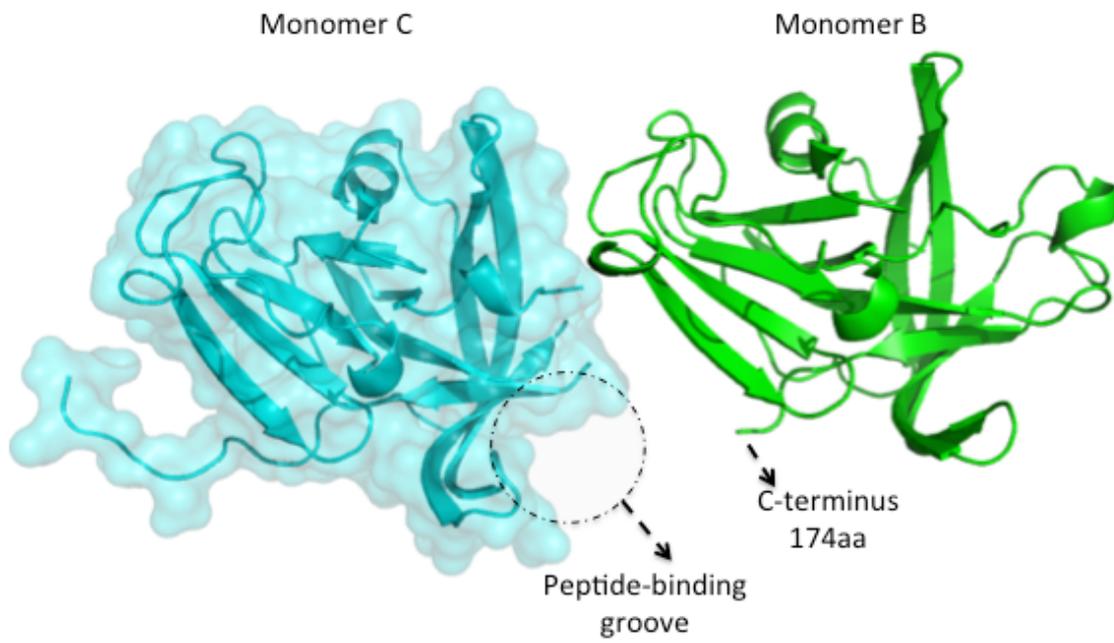


Figure 29: Interaction of monomer C and B. The figure depicts the close proximity of monomer B C-terminus to the peptide-binding groove in monomer C. The interaction of monomer D with its neighboring molecule's C-terminus is identical to CB interaction and therefore not shown.

## Chapter 4: Conclusion and Future Work

The main goal of this research was to characterize the HuNV Pro in complex with different peptides and RNA oligonucleotides. Towards this end, several objectives were achieved. First, protocols were established to express and purify the soluble form of the enzyme. Secondly, crystallization conditions were optimized by systematic variation of each parameter of the initial crystallization conditions to grow high quality crystals in easily reproducible crystallization conditions. Although, the final objective of determining the structures of HuNV Pro in complex with either RNA oligonucleotide or synthetic peptides could not be achieved, the results obtained have highlighted the possible reasons why we failed to crystallize these complexes.

Due to the limited amount of protein and RNA oligonucleotides available, extensive crystallization screening could not be carried out for the HuNV Pro-RNA project. Therefore, it is still possible that further screening might result in an identification of crystallization condition that can stabilize the complex structure, thereby enabling us to solve the first complex structure of viral Pro with RNA. However, very high structural similarity of HuNV Pro-RNA co-crystals grown in the presence of very high RNA concentration and native protein structures suggest that either RNA interaction is transient in nature or it did not interact with HuNV Pro. In either case we need to explore the components of the replication complex and/or RNA sequence(s) that might help stabilize the HuNV Pro-RNA complex. In related Picornaviruses, the interaction of Pro to the viral RNA is facilitated and modulated by Pol [48, 47, 46, 44]. Therefore, the role of NV ProPol in Pro-RNA binding needs to be explored. Previous studies have also indicated that the ProPol, which is a stable precursor, serves as a better protease than mature Pro, suggesting that ProPol might be the dominant form of the viral Pro in infected cells [34]. Therefore, it is also possible that in NVs, like PV, ProPol has superior binding affinity to viral RNA compared to Pro and consequently would help stabilize the protein-RNA complex [84].

Generally, crystallization studies of protein-RNA complexes are sensitive to RNA sequences and length. The binding studies of HuNV Pro to RNA provided little evidence for RNA sequence specificity, but considering that in these studies only a few sequences were tested [28], the role of RNA sequence and length in NV Pro-RNA binding needs to be explored further. Highly

conserved RNA structures, similar to picornaviruses CREs, have also been identified in NVs genome, and are important for replication [85, 86, 87]. It is probable that these RNA secondary structures or short nucleotide sequences within these structures will bind to NV Pro with higher affinity.

The role of RNA sequence and length in HuNV Pro or ProPol-RNA binding can be explored by differential scanning fluorometry (DSF) [88]. DSF monitors the thermal unfolding of protein and can be used to measure the thermal stability of HuNV Pro or its precursor, ProPol, in the presence of different RNA oligonucleotides. The stabilization effect is believed to be proportional to the affinity of substrate for a given protein [89, 88]. Therefore, DSF can be used to calculate the binding affinities of different RNA oligonucleotides with HuNV Pro and its precursor, ProPol. This would help identify the binding preference of viral RNA for either HuNV Pro or ProPol. Furthermore, the right length and sequence of RNA oligonucleotides will ensure better binding to the Pro and promote the formation of HuNV Pro or ProPol-RNA complex crystals by stabilizing the complex [56]. Therefore, crystallization studies using the RNA oligonucleotides with the strongest calculated binding affinities would significantly increase the likelihood of solving the NV Pro-RNA complex structure.

As far as the HuNV Pro-peptide project is concerned, the structural analysis has provided insight into why we failed to crystallize the complex structure. As a result of crystal packing the two peptide-binding grooves out of the four present in the asymmetric unit are already occupied by the C-terminal tail residues, thereby preventing the binding of synthetic peptides used in the study. The lack of binding in the other two unoccupied active sites is possibly due to the interference by the C-terminus residues of the neighboring Pro molecules in the crystal. One way to overcome this issue is to pursue screening in order to find a crystallization condition that might lead to the formation of crystals with different packing in which the C-terminal tail neither interferes with peptide binding nor occupies the peptide-binding groove. It also occurred to us that by modifying the sequence of the C-terminus of inactive NV Pro (C139A) to incorporate P' residues, it may be possible to exploit the same crystal form to obtain structures of HuNV Pro-peptide complexes that would reveal the details of Pro-peptide interactions on the P' side of the cleavage. A similar strategy has been successfully used to gain insight into how the interactions

of P4-P1 residues of early and late cleavage sites differ in HuNV Pro [36]. However, it is well known that even very modest changes to protein constructs may result in different crystal packing. Therefore, it is possible that the C-terminally extended constructs may not work. In addition, NMR studies have also suggested that the NV Pro C-terminus has the propensity to fold into the active site of the molecule it belongs [52]. Considering all these issues, a time and cost-effective approach would be to use C-terminus truncate construct of HuNV Pro for Pro-peptide crystallization studies. Disappointingly, however, aforementioned strategies could not be pursued due to lack of resources.

## Bibliography

1. Zheng, D. P., T. Ando, R. L. Fankhauser, R. S. Beard, R. I. Glass and S. S. Monroe, Norovirus classification and proposed strain nomenclature, *Virology* 346 (2006), no. 2, pp. 312-323.
2. Donaldson, E. F, Lindesmith, L. C., Lobue, A. D., and Baric, R. S., Viral shape-shifting: Norovirus evasion of the human immune system, *Nat Rev Microbiol* 8 (2010), no. 3, pp. 231-241.
3. Thorne, L. G., and Goodfellow, I. G., Norovirus gene expression and replication, *J Gen Virol* 95, 2014, pp. 278-291.
4. Bank-Wolf, B. R., Konig, M., and Thiel, H. J., "Zoonotic aspects of infections with noroviruses and sapoviruses," *Vet microbiol*, vol. 140, 2009 Elsevier B.V, Netherlands, 2010, pp. 204-212.
5. Mesquita, J. R., and Nascimento, M. S., "Serosurvey of veterinary conference participants for evidence of zoonotic exposure to canine norovirus - study protocol," *Virol j*, vol. 9, 2012, p. 250.
6. Wobus, C. E., Thackray, L. B., and Virgin, H. W. T., "Murine norovirus: A model system to study norovirus biology and pathogenesis," *J virol*, vol. 80, 2006, pp. 5104-5112.
7. Vashist, S., Bailey, D., Putics, A. and Goodfellow, I. G., Model systems for the study of human norovirus biology, *Future Virol* 4, 2009, no. 4, pp. 353-367.
8. Karst, S. M., Zhu, S., and Goodfellow, I. G., The molecular pathology of noroviruses, *J Pathol* 235, 2015, no. 2, pp. 06-216.
9. Straub, T. M., Honer zu Bentrup, K., Orosz-Coghlan, P., Dohnalkova, A., Mayer, B. K. Bartholomew, R. A., Valdez, C. O., Bruckner-Lea, C. P. Gerba, Abbaszadegan M., and Nickerson, C. A., In vitro cell culture infectivity assay for human noroviruses, *Emerg Infect Dis* 13, 2007, no. 3, pp. 396-403.
10. Kaufman, S. S., Green, K. Y., and Korba, B. E., Treatment of norovirus infections: Moving antivirals from the bench to the bedside, *Antiviral Res* 105c, 2014, pp. 80-91.
11. Updated norovirus outbreak management and disease prevention guidelines, *MMWR Recomm Rep* 60, 2011, pp. 1-18.
12. Hall, A. J., Lopman, B. A., Payne, D. C., Patel, M. M., Gastanaduy, P. A., J. Vinje and Parashar, U. D., Norovirus disease in the united states, *Emerg Infect Dis* 19, 2013, no. 8, pp. 1198-1205.
13. Lopman, B. A. , Reacher, M., Gallimore, C., Adak, G. K., Gray, J. J., and Brown, D. W., A summertime peak of "winter vomiting disease": Surveillance of noroviruses in england and wales, 1995 to 2002, *BMC Public Health* 3, 2003, 13.
14. Siebenga, J. J., Vennema, H., Zheng, D. P., Vinje, B. E. Lee, X. L. Pang, E. C. Ho, Lim, A., Choudekar, W., Broor, S., Halperin, T., Rasool, N. B., Hewitt, J., Greening, G. E., M. Jin, Z. J. Duan, Y. Lucero, M. O'Ryan, M. Hoehne, E. Schreier, R. M. Ratcliff, P. A. White, N. Iritani, G. Reuter and M. Koopmans, Norovirus illness is a global problem: Emergence and spread of norovirus gii.4 variants, 2001-2007, *J Infect Dis*, 2009, no. 5, pp. 802-812.
15. Lopman, B., Gastanaduy, P., Park, G. W., Hall, A. J., Parashar, U. D., and Vinje, J., "Environmental transmission of norovirus gastroenteritis," *Curr opin virol*, vol. 2, Elsevier, 2012, pp. 96-102.

16. Rohayem, J., "Norovirus seasonality and the potential impact of climate change," *Clin microbiol infect*, vol. 15, 2009, pp. 524-527.
17. Teunis, P. F. , Moe, C. L., Liu, P., S. E. Miller, L. Lindesmith, R. S. Baric, J. Le Pendu and R. L. Calderon, Norwalk virus: How infectious is it?, *J Med Virol* 80, 2008, no. 8, pp. 1468-1476.
18. Patel, M. M., Hall, A. J., J. Vinje and Parashar, U. D., "Noroviruses: A comprehensive review," *J clin virol*, vol. 44, 2009, pp. 1-8.
19. Glass, R. I., Parashar, U. D., and Estes, M. K., Norovirus gastroenteritis, *N Engl J Med* 361, 2009, no. 18.
20. Turcios-Ruiz, R. M., Axelrod, P., St John, K., Bullitt, E., Donahue, J., Robinson and Friss, H. E., "Outbreak of necrotizing enterocolitis caused by norovirus in a neonatal intensive care unit," *J pediatr*, vol. 153, 2008, pp. 339-344.
21. Chen, S. Y., Tsai, C. N., Lai, M. W., Chen, C. Y., Lin, K. L., T. Y. Lin and Chiu, C. H., Norovirus infection as a cause of diarrhea-associated benign infantile seizures, *Clin Infect Dis* 48, 2009, no. 7, 849-855.
22. Hardy, M. E., Norovirus protein structure and function, *FEMS Microbiol Lett* 253, 2005, no. 1, pp. 1-8.
23. McFadden, N., Bailey, D., Carrara, G., Benson, A., Chaudhry, Y., Shortland, A., Heeney, J., Yarovinsky, F., Simmonds, P., Macdonald, A., and Goodfellow, I. G., "Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4," *Plos pathog*, vol. 7, , 2011, p. e1002413.
24. Belliot, G., Sosnovtsev, S. V., Mitra, T., Hammer, C., Garfield, M., and Green, K. Y., In vitro proteolytic processing of the md145 norovirus orf1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells, *J Virol* 77, 2003, no. 20, pp.10957-10974.
25. Shirato, H., Norovirus recognition sites on histo-blood group antigens, *Front Microbiol* 3 2012.
26. Chung, L., Bailey, D., Leen, E. N., Emmott, E. P., Chaudhry, Y., Roberts, L. O., Curry, S., Locker, N. and Goodfellow, I. G., "Norovirus translation requires an interaction between the c terminus of the genome-linked viral protein vpg and eukaryotic translation initiation factor 4g," *J biol chem*, vol. 289, 2014, pp. 21738-21750.
27. Daughenbaugh, K. F., Fraser, C. S., Hershey, J. W., and Hardy, M. E., The genome-linked protein vpg of the norwalk virus binds eif3, suggesting its role in translation initiation complex recruitment, *EMBO J* 22, 2003, no. 11, pp. 2852-2859.
28. Viswanathan, P., May, J., Uhm, S., Yon, C. and Korba, B., Rna binding by human norovirus 3c-like proteases inhibits protease activity, *Virol* 438, 2013, no. 1, pp. 20-27.
29. Belliot, G., Sosnovtsev, S. V., Chang, K. O., McPhie, P., and Green, K. Y., Nucleotidylylation of the vpg protein of a human norovirus by its proteinase-polymerase precursor protein, *Virol* 374, 2008, no. 1, 33-49.
30. Nakamura, K. Y. Someya, T. Kumasaka, G. Ueno, M. Yamamoto, T. Sato, N. Takeda, T. Miyamura and Tanaka, N., A norovirus protease structure provides insights into active and substrate binding site integrity, *J Virol* 79, 2005, no. 21, pp.13685-13693.
31. Hussey, R. J., Coates, L., Gill, R. S., Erskine, P. T., Coker, S. F., Mitchell, E., Cooper, J. B., Wood, S., Broadbridge, R., Clarke, I. N., Lambden, P. R., and Shoolingin-Jordan, P. M., A structural study of norovirus 3c protease specificity: Binding of a designed active site-directed peptide inhibitor, *Biochemistry* 50, 2011, no. 2, pp. 240-249.

32. Zeitler, C. E., Estes, M. K., and Venkataram, B. V., Prasad, "X-ray crystallographic structure of the norwalk virus protease at 1.5-Å resolution," *J virol*, vol. 80, 2006, pp. 5050-5058.
33. Hedstrom, L., "Serine protease mechanism and specificity," *Chem rev*, vol. 102, 2002, pp. 4501-4524.
34. May, J., Korba, B., Medvedev, A. and Viswanathan, P., Enzyme kinetics of the human norovirus protease control virus polyprotein processing order, *Virology* 444, 2013, no. 1-2, pp. 218-224.
35. May, J. , Ng, K. K.-S., Medvedev, A., and Korba, B., "The p4-p2' amino acids surrounding human norovirus polyprotein cleavage sites define the core sequence regulating self-processing order.," *J Virol*, 2014, pp. 01357-01314.
36. Muhaxhiri, Z., Deng, L., Shanker, S., Sankaran, B., Estes, M. K., Palzkill, T., Song, Y., and Prasad, B. V., Structural basis of substrate specificity and protease inhibition in norwalk virus, *J Virol* 87, 2013, no. 8, 4281-4292.
37. Hardy, M. E., Crone, T. J., Brower, J. E., and Ettayebi, K., Substrate specificity of the norwalk virus 3c-like proteinase, *Virus Res* 89, 2002, no. 1, pp. 29-39.
38. Leen, E. N., Baeza, G., and Curry, S., "Structure of a murine norovirus ns6 protease-product complex revealed by adventitious crystallisation," *Plos one*, vol. 7, 2012, p. e38723.
39. May, J., Viswanathan, P., Ng, K. K., Medvedev, A. and Korba, B., The p4-p2' amino acids surrounding human norovirus polyprotein cleavage sites define the core sequence regulating self-processing order, *J Virol* 88, 2014, no. 18, pp.10738-10747.
40. Chuck, C. P., Chow, H. F., Wan, D. C., and Wong, K. B., Profiling of substrate specificities of 3c-like proteases from group 1, 2a, 2b, and 3 coronaviruses, *PLoS One* 6 , 2011, no. 11, pp. e27228.
41. Herod, M. R., Prince, C. A., Skilton, R. J., Ward, V. K., Cooper, J. B., and Clarke, I. N., "Structure-based design and functional studies of novel noroviral 3c protease chimaeras offer insights into substrate specificity," *Biochem j*, vol. 464, 2014, pp. 461-472.
42. Belliot, G., Sosnovtsev, S. V., Chang, K. O., Babu, V., Uche, U., Arnold, J. J., Cameron and Green, K. Y., Norovirus proteinase-polymerase and polymerase are both active forms of rna-dependent rna polymerase, *J Virol* 79, 2005, no. 4, pp. 2393-2403.
43. Peters, H., Kusov, Y. Y., Meyer, S., Benie, A. J., Bauml, E., Wolff, M., Rademacher, C. , Peters, T., and Gauss-Muller, V., "Hepatitis a virus proteinase 3c binding to viral rna: Correlation with substrate binding and enzyme dimerization," *Biochem j*, vol. 385, 2005, pp. 363-370.
44. Parsley, T. B., Cornell, C. T., and Semler, B. L., Modulation of the rna binding and protein processing activities of poliovirus polypeptide 3cd by the viral rna polymerase domain, *J Biol Chem* 274 , 1999, no. 18, pp. 12867-12876.
45. Shih, S. R., Chiang, C., Chen, T. C., Wu, C. N., Hsu, J. T., Lee, J. C., Hwang, M. J., M. L. Li, Chen, G. W., and Ho, M. S., Mutations at kfrdi and vlgk domains of enterovirus 71 3c protease affect its rna binding and proteolytic activities, *J Biomed Sci* 11, 2004, no. 2, pp. 239-248.
46. Shen, M. , Wang, Q., Yang, Y., Pathak, H. B., Arnold, J. J., Castro, C., Lemon, S. M., and Cameron, C. E., Human rhinovirus type 14 gain-of-function mutants for orii utilization define residues of 3c(d) and 3dpol that contribute to assembly and stability of the picornavirus vpg uridylylation complex, *J Virol* 81, 2007, no. 22, pp.12485-12495.

47. Steil, B. P., and Barton, D. J., Cis-active rna elements (cres) and picornavirus rna replication, *Virus Res* 139, 2009, no. 2, pp. 240-252.
48. Amero, C. D., Arnold, J. J., Moustafa, I. M., Cameron, C. E., and Foster, M. P., Identification of the orii-binding site of poliovirus 3c protein by nuclear magnetic resonance spectroscopy, *J Virol* 82, 2008, no. 9, pp. 4363-4370.
49. Bertolotti-Ciarlet, A., Crawford, S. E., Hutson, A. M., and Estes, M. K., The 3' end of norwalk virus mrna contains determinants that regulate the expression and stability of the viral capsid protein vp1: A novel function for the vp2 protein, *J Virol* 77, 2003, no. 21, pp. 11603-11615.
50. Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., and Thornton, J. M., Protein-rna interactions: A structural analysis, *Nucleic Acids Res* 29, 2001, no. 4, pp. 943-954.
51. Takahashi, D., Hiromasa, Y., Kim, Y., Anbanandam, A., Yao, X., Chang, K. O., and Prakash, O., Structural and dynamics characterization of norovirus protease, *Protein Sci* 22, 2013, no. 3, pp. 347-357.
52. Fernandes, H., Leen, E. N., Cromwell, H., Pfeil, M. P., and Curry, S., Structure determination of murine norovirus ns6 proteases with c-terminal extensions designed to probe protease-substrate interactions, *PeerJ* 3, 2015, e798.
53. Rhodes, G., *Crystallography made crystal clear*, Academic Press 2006.
54. Jancarik, J., and Kim, S.-H., Sparse matrix sampling: A screening method for crystallization of proteins, *Journal of Applied Crystallography* 24, 1991, pp. 409-411.
55. Bergfors, T., Seeds to crystals, *J Struct Biol* 142, 2003, no. 1, pp. 66-76.
56. Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R., and Shewchuk, L. M., Crystallization of protein-ligand complexes, *Acta Crystallogr D Biol Crystallogr* 63, 2007, pp. 72-79.
57. Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P. and Minor, W. Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter?, *Protein Sci* 17, 2008, no. 4, pp. 623-632.
58. Nagendra, H. G, N. Sukumar, M. Vijayan, Role of water in plasticity, stability and action of protein: the crystal structures of lysozyme at very low levels of hydration, *Proteins* (1998), vol. 32, pp. 229-40
59. Chapman, M. S., and Somasundaram, T., "De-icing: Recovery of diffraction intensities in the presence of ice rings," *Acta crystallogr d biol crystallogr*, vol. 66, 2010, pp. 741-744.
60. Kabsch, W., Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr D Biol Crystallogr*, vol. 66, 2010, pp. 133-144.
61. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K., "Features and development of coot," *Acta crystallogr d biol crystallogr*, vol. 66, 2010, pp. 486-501.
62. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A., "Refrmac5 for the refinement of macromolecular crystal structures," *Acta crystallogr d biol crystallogr*, vol. 67, 2011, pp. 355-367.
63. Brunger, A. T., Free r value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature*, vol. 355, 1992, pp. 472-475.
64. Luft, J. R., Wolfley, J. R., and Snell, E. H., What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments, *Cryst Growth Des* 11 2011, no. 3, pp. 651-663.

65. Yoshizaki, I., Fukuyama, S., Koizumi, H., Tachibana, M., Kojima, k., Matsuura, Y., Tanaka, M., and Komatsu, H., Impurity-induced defect and its effect on protein crystal perfection., *Journal of Crystal Growth* 290, 2006, pp. 185-191.
66. Caylor, C. L., Dobrianov, I., Lemay, S. G., Kimmer, C., Kriminski, S., Finkelstein, K. D., Zipfel, W., Webb, W. W., Thomas, B. R., Chernov, A. A., and Thorne, R. E., Macromolecular impurities and disorder in protein crystals, *Proteins* 36, 1999, pp. 270-281.
67. Collingsworth, P. D., Bray, T. L., and Christopher, G. L., Crystal growth via computer controlled vapor diffusion, *Crystal growth via computer controlled vopor diffusion*. 219 2000, pp. 283-289.
68. Yoshizaki, I., Fukuyama, S., Koizumi, H., Tachibana, M., Kojima, k., Matsuura, Y., Tanaka, M., and Komatsu, H., Impurity-induced defect and its effect on protein crystal perfection., *Journal of Crystal Growth* 290, 2006, pp. 185-191
69. Chayen, N., A novel technique to control the rate of vapour diffusion, giving larger protein crystals, *Journal of Applied Crystallography* 30, 1997, pp. 198-202.
70. Stura, E. A., and Wilson, I. A., Seeding techniques. In *crystallization of nucleic acids and proteins.*, OUP, New York, 1992.
71. Rosenberger, F., Howard, S. B., Sowers, J. W., and Nyce, T. A., Temperature dependence of protein solubility — determination and application to crystallization in x-ray capillaries, *Journal of Crystal Growth* 129, 1993, pp. 1-12.
72. Forsythe, E. L., Maxwell, D. L., and Pusey, M., Vapor diffusion, nucleation rates and the reservoir to crystallization volume ratio, *Acta Crystallogr D Biol Crystallogr* 58, 2002, pp. 1601-1605.
73. Gilliland, G. L., Tung, M., and Ladner, J. E., The biological macromolecule crystallization database: Crystallization procedures and strategies, *Acta Crystallogr D Biol Crystallogr* 58, 2002, pp. 916-920.
74. Chang, K. O., Takahashi, D., Prakash, O., and Kim, Y., Characterization and inhibition of norovirus proteases of genogroups i and ii using a fluouescence resonance energy transfer assay, *Virology* 423, 2012, pp. 125-133.
75. Batey, R. T., Sagar, M. B., and Doudna, J. A., Structural and energetic analysis of rna recognition by a universally conserved protein from the signal recognition particle, *J Mol Biol* 307, 2001, pp. 229-246.
76. Matthews, B. W., Solvent content of protein crystals, *J Mol Biol* 33, 1968, pp. 491-497.
77. Kantardjieff, K. A., and Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, dna, and protein-nucleic acid complex crystals, *Protein Sci* 12, 2003, pp. 1865-1871.
78. Janin, J., and Rodier, F., Protein-protein interaction at crystal contacts, *Proteins* 23 ,1995, pp. 580-587.
79. Janin, J., Specific versus non-specific contacts in protein crystals, *Nat Struct Biol* 4, 1997, pp. 973-974.
80. Porollo, A., and Meller, J., "Versatile annotation and publication quality visualization of protein complexes using polyview-3d," *Bmc bioinformatics*, vol. 8, 2007, p. 316.
81. Liebeschuetz, J., Hennemann, J., Olsson, T. , and Groom, C. R., "The good, the bad and the twisted: A survey of ligand geometry in protein crystal structures," *J comput aided mol des*, vol. 26, 2012, pp. 169-183.
82. Blow, D. *Outline of crystallography for biologist*, OUP, 2002.

83. Laskowski, R. A., and Swindells, M. B., Ligplot+: Multiple ligand-protein interaction diagrams for drug discovery, *J Chem Inf Model* 51, 2011, no. 10, pp. 2778-2786.
84. Pathak, H. B., Arnold, J. J., Wiegand, P. N., Hargittai, M. R., and Cameron, C. E., Picornavirus genome replication: Assembly and organization of the vpg uridylylation ribonucleoprotein (initiation) complex, *J Biol Chem* 282, 2007, pp. 16202-16213.
85. Simmonds, P., Karakasiliotis, I., Bailey, D., Chaudhry, Y., Evans, D. J., and Goodfellow, I. G., Bioinformatic and functional analysis of rna secondary structure elements among different genera of human and animal caliciviruses, *Nucleic Acids Res* 36, 2008, no. 8, pp. 2530-2546.
86. Bailey, D., Karakasiliotis, I., Vashist, S. , Chung, L. M., Rees, J., McFadden, N., Benson, A., Yarovinsky, F., Simmonds, P., and Goodfellow, I. G., Functional analysis of rna structures present at the 3' extremity of the murine norovirus genome: The variable polypyrimidine tract plays a role in viral virulence, *J Virol* 84, 2010, pp. 2859-2870.
87. Lin, X., Thorne, L., Jin, Z., Hammad, L. A., Li, S., Deval, J., Goodfellow, I. G., and Kao, C. C., Subgenomic promoter recognition by the norovirus rna-dependent rna polymerases, *Nucleic Acids Res*, vol. 43 , 2015, pp. 446-460.
88. Niesen, F. H., Berglund, H., and Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc*, vol. 9, 2007, pp. 2212-2221.
89. Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J., and Edwards, A. M., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination, *Proc Natl Acad Sci*, vol. 9, 2006, , pp. 15835-15840.
90. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J., Phaser crystallographic software, *J Appl Crystallogr*, vol. 40, 2007, pp. 658-674.
91. Schechter , I., and Berger, A., "On the size of the active site in proteases. I. Papain," *Biochem biophys res commun*, vol. 27, 1967, pp. 157-162.
92. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M., Exponentially modified protein abundance index (empai) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, *Mol Cell Proteomics*, vol. 9, 2005, pp. 1265-1272.

## Appendix

### 1. NV Pro structures deposited in PDB to date.

PDB ID	Genogroup	Molecules per asymmetric units
1WQS	GI (Human/Chiba)	4
4ASH	GV (Murine)	2
4X2V	GV (Murine)	4
2LNC	GI (Human/Norwalk)	1
2IPH	GI (Human/Southampton)	2
4X2X	GV (Murine)	1
4X2W	GV (Murine)	2
4X2Y	GV (Murine)	2
4XBB	GI (Human/Norwalk)	1
4XBC	GI (Human/Norwalk)	1
4XBD	GI (Human/Norwalk)	2
3UR6	GI (Human/Norwalk)	2
3UR9	GI (Human/Norwalk)	2
2FYQ	GI (Human/Norwalk)	1
2FYR	GI (Human/Norwalk)	1
4IN1	GI (Human/Norwalk)	1
4IN2	GI (Human/Norwalk)	2

2. Schechter and Berger nomenclature. In this system the amino acids residues (P or P') of a polyprotein bind Pro at sites called sub-sites (S or S'). The amino acids residues and their complementary sub-sites present on the N-terminal side of the scissile bond are labeled P and S, while those C-terminal to the scissile bond are labeled P' and S' [90].

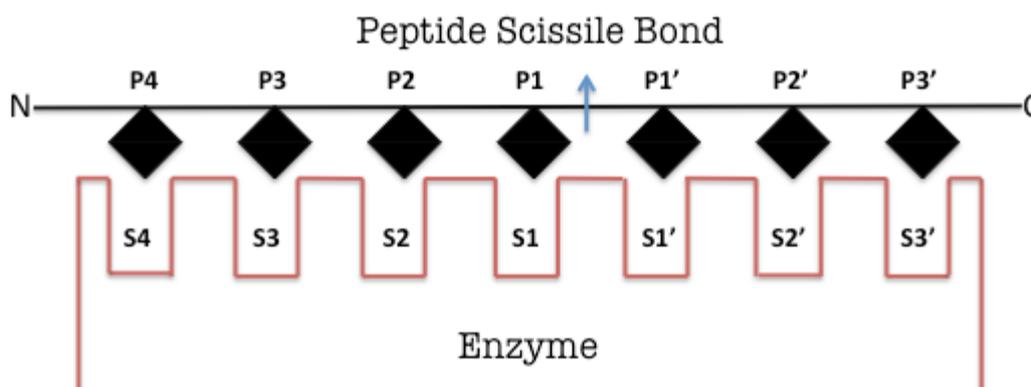


Figure 30: Schematic representation of Schechter and Berger nomenclature.

3. Predicted binding of P' residues in NV Pro-peptide structures. Based on visual inspecting the P1' residue can occupy a cavity that lies close to the active site residues (colored yellow). Beyond P1' residues can follow two different routes (1 and 2). Route 1 is an extension of P1' cavity, and represent a more probable route. However, for substrate to follow route 2, it has to kink considerably at P1'-P2' position. This is not impossible considering that the most common residues present at P2' is Proline, which based on its unique structure is known to create bends in protein structures (based on *Hussey et al.* [31]).

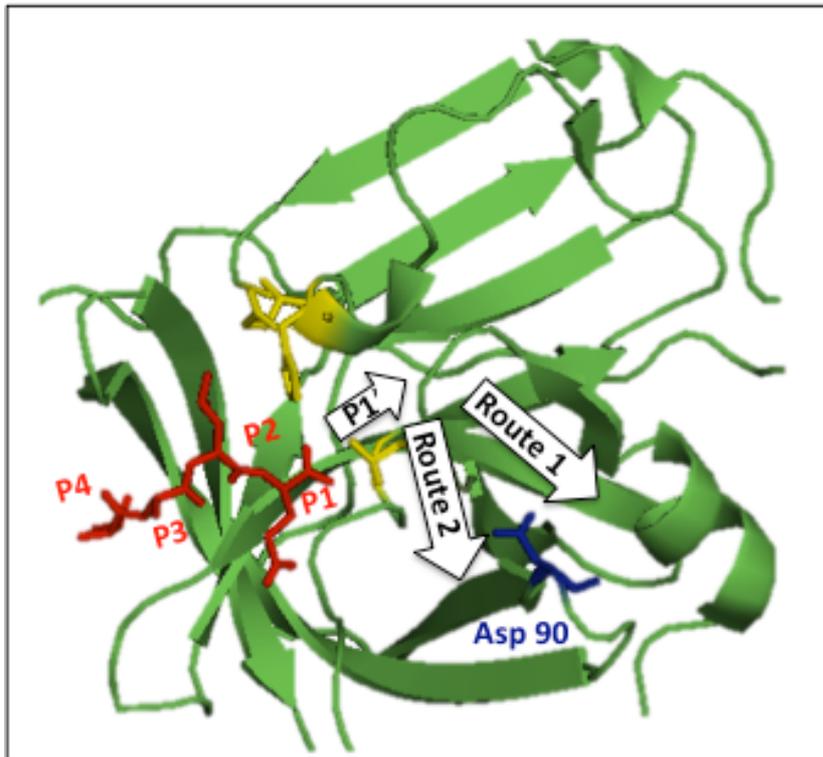


Figure 31: The predicted binding routes for P' residues in Norovirus protease (PDB ID: 4IN2).

4. Sparse Matrix Screening. Commercially available crystallization screens used to identify hits for the wild type NV Pro-RNA and mutant NV Pro-Peptide complexes. All screens were done in the presence of excess RNA or peptide.

<b>Code</b>	<b>Developer</b>	<b>Screen</b>	<b>Temperature °C</b>	<b>[Wild type NV Pro] mM</b>	<b>RNA oligonucleotide (mM)</b>
C2AS005	Hampton Research	Crystal	Room Temperature (25 °C)	0.12	NVP2 (0.6 mM)
C2AS006	Hampton Research	Index	Room Temperature (25 °C)	0.20	NVP2 (0.6 mM)
C2AS007	Hampton Research	Index	Room Temperature (25 °C)	0.20	NVP2 (0.6 mM)
	<b>Developer</b>	<b>Screen</b>	<b>Temperature</b>	<b>[Mutant NV Pro] mM</b>	<b>Peptide</b>
C2AS008	Qiagen	Protein complex	Room Temperature (25 °C)	0.20	Long peptide (1.2 mM)
C2AS009	Qiagen	JCSG+	Room Temperature (25 °C)	0.20	Long peptide (1.2 mM)
C2AS010	Hampton Research	Index	Room Temperature (25°C)	0.20	Long peptide (2.0 mM)
C2AS011	Hampton Research	PEGRx	Room Temperature (25 °C)	0.20	Long peptide (2.0 mM)
C2AS012	Qiagen	Protein Complex	Room Temperature (25 °C)	0.20	Long peptide (2.0 mM)
C2AS013	Hampton Research	Index	Room Temperature (25 °C)	0.20	Short peptide (5.40 mM)
C2AS014	Hampton Research	Crystal	Room Temperature (25 °C)	0.20	Short peptide (5.40 mM)
C2AS015	Qiagen	JCGS+	Room Temperature (25 °C)	0.20	Short peptide (5.40 mM)

5. Structural Comparison of HuNV Pro dimers.

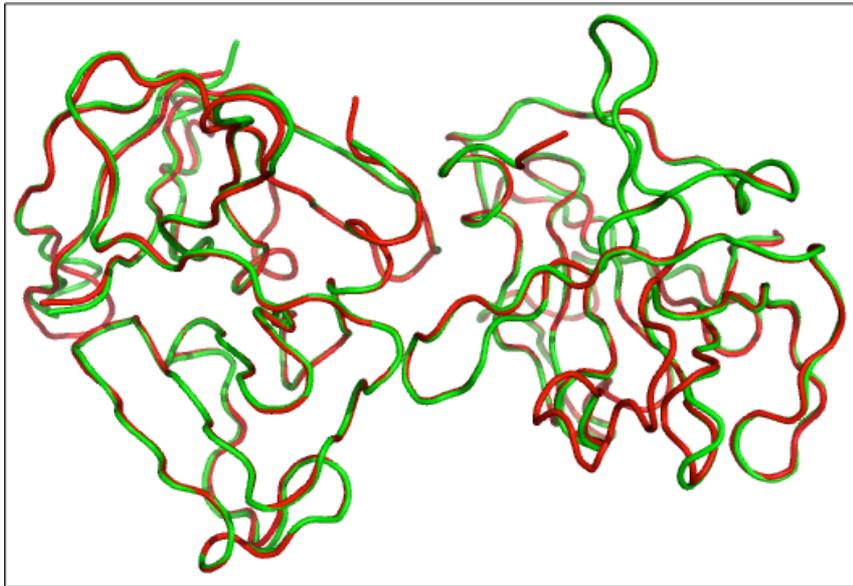


Figure 32: Structural comparison of HuNV Pro dimers obtained in the presence of RNA (Red) and RNA/Peptide (Green). The structures are identical considering that the RMSD difference is only 0.2 Å.

6. Mass Spectroscopy data was interpreted using Mascot software (Matrix Science). The protein score is used to rank the proposed proteins according to the total amount of evidence supporting the identification of each protein. A higher score indicates more evidence. The exponentially modified protein abundance index (emPAI) offers approximate, label free, quantitation for the proteins in a mixture based on the number of peptides observed per protein [92].

Description	Protein Score	Non-duplicate Match	emPAI
Norwalk chymotrypsin-like cysteine protease	776	11	3.11
Enterobacteria phage HK620 cII	31	3	0.29
Pseudomonas aeruginosa K1aB	19	1	0.07

**ELSEVIER LICENSE  
TERMS AND CONDITIONS**

Aug 09, 2015

---

---

This is a License Agreement between Asfandyar Sikandar ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Asfandyar Sikandar
Customer address	225b, Olympus Hall Calgary, AB t2n 4v7
License number	3685080041684
License date	Aug 09, 2015
Licensed content publisher	Elsevier
Licensed content publication	Virology
Licensed content title	Enzyme kinetics of the human norovirus protease control virus polyprotein processing order
Licensed content author	Jared May, Brent Korba, Alexei Medvedev, Prasanth Viswanathan
Licensed content date	September 2013
Licensed content volume number	444
Licensed content issue number	1-2
Number of pages	7
Start Page	218
End Page	224
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic

Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Table 1 and 2
Title of your thesis/dissertation	Structural Studies of Human Norovirus Protease with RNA and Peptides
Expected completion date	Aug 2015
Estimated size (number of pages)	70
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

### INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

### GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier

## ELSEVIER LICENSE TERMS AND CONDITIONS

Aug 09, 2015

This is a License Agreement between Asfandyar Sikandar ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Asfandyar Sikandar
Customer address	225b, Olympus Hall Calgary, AB t2n 4v7
License number	3685060522892
License date	Aug 09, 2015
Licensed content publisher	Elsevier
Licensed content publication	Virology
Licensed content title	RNA binding by human Norovirus 3C-like proteases inhibits protease activity
Licensed content author	Prasanth Viswanathan, Jared May, Sunghae Uhm, Changsuek Yon, Brent Korba
Licensed content date	30 March 2013
Licensed content volume number	438
Licensed content issue number	1
Number of pages	8
Start Page	20
End Page	27
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of	2

## figures/tables/illustrations

Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Fig 5 and 6
Title of your thesis/dissertation	Structural Studies of Human Norovirus Protease with RNA and Peptides
Expected completion date	Aug 2015
Estimated size (number of pages)	70
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

[Terms and Conditions](#)

### INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

### GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."



AMERICAN SOCIETY FOR MICROBIOLOGY

**Title:** Structural Basis of Substrate Specificity and Protease Inhibition in Norwalk Virus

**Author:** Zana Muhaxhiri, Lisheng Deng, Sreejesh Shanker et al.

**Publication:** Journal of Virology

**Publisher:** American Society for Microbiology

**Date:** Apr 15, 2013

Copyright © 2013, American Society for Microbiology

Logged in as:  
Asfandyar Sikandar  
Account #:  
3000925953

[LOGOUT](#)

### Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2015 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)



AMERICAN SOCIETY FOR MICROBIOLOGY

**Title:** A Norovirus Protease Structure Provides Insights into Active and Substrate Binding Site Integrity

**Author:** Kentaro Nakamura, Yuichi Someya, Takashi Kumasaka et al.

**Publication:** Journal of Virology

**Publisher:** American Society for Microbiology

**Date:** Nov 1, 2005

Copyright © 2005, American Society for Microbiology

Logged in as:  
Asfandyar Sikandar  
Account #:  
3000925953

[LOGOUT](#)

### Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2015 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)



**Note:** Copyright.com supplies permissions but not the copyrighted content itself.

**1**  
PAYMENT

**2**  
REVIEW

**3**  
CONFIRMATION

### Step 3: Order Confirmation

**Thank you for your order!** A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us at +1.855.239.3415 Toll Free, M-F between 3:00 AM and 6:00 PM (Eastern), or write to us at [info@copyright.com](mailto:info@copyright.com). This is not an invoice.

**Confirmation Number: 11418861**  
**Order Date: 08/09/2015**

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

#### Payment Information

Asfandyar Sikandar  
asikanda@ucalgary.ca  
+1 (403) 827-1196  
Payment Method: n/a

#### Order Details

##### The Journal of general virology

**Order detail ID:** 67931867  
**Order License Id:** 3685071066542

**ISSN:** 0022-1317

**Publication Type:** Journal

**Volume:**

**Issue:**

**Start page:**

**Publisher:** SOCIETY FOR GENERAL  
MICROBIOLOGY  
**Author/Editor:** SOCIETY FOR GENERAL  
MICROBIOLOGY ; FEDERATION OF  
EUROPEAN MICROBIOLOGICAL  
SOCIETIES

**Permission Status:** **Granted**

**Permission type:** Republish or display content  
**Type of use:** Republish in a thesis/dissertation

**Requestor type:** Academic institution

**Format:** Print, Electronic

**Portion:** image/photo

**Number of images/photos requested:** 1

**Title or numeric reference of the portion(s):** Fig.2

**Title of the article or**

<b>chapter the portion is from</b>	Norovirus gene expression and replication
<b>Editor of portion(s)</b>	N/A
<b>Author of portion(s)</b>	Lucy G. Throne and Ian G. Goodfellow
<b>Volume of serial or monograph</b>	N/A
<b>Issue, if republishing an article from a serial</b>	95
<b>Page range of portion</b>	278-291
<b>Publication date of portion</b>	2014
<b>Rights for</b>	Main product
<b>Duration of use</b>	Current edition and up to 5 years
<b>Creation of copies for the disabled</b>	yes
<b>With minor editing privileges</b>	no
<b>For distribution to</b>	Canada
<b>In the following language(s)</b>	Original language of publication
<b>With incidental promotional use</b>	no
<b>Lifetime unit quantity of new product</b>	Up to 499
<b>Made available in the following markets</b>	educational
<b>The requesting person/organization</b>	Asfandiyar Sikandar (Graduate Student)
<b>Order reference number</b>	

<b>Author/Editor</b>	Asfandyar Sikandar
<b>The standard identifier</b>	1
<b>Title</b>	Structural Studies of Human Norovirus Protease Complexes with RNA and Peptides
<b>Publisher</b>	University of Calgary
<b>Expected publication date</b>	Aug 2015
<b>Estimated size (pages)</b>	70

**Note:** This item will be invoiced or charged separately through CCC's **RightsLink** service. [More info](#) **\$ 0.00**

---

<b>Total order items: 1</b>	<b>This is not an invoice.</b>	<b>Order Total: 0.00 USD</b>
-----------------------------	--------------------------------	------------------------------

**Title:** Serine Protease Mechanism and Specificity  
**Author:** Lizbeth Hedstrom  
**Publication:** Chemical Reviews  
**Publisher:** American Chemical Society  
**Date:** Dec 1, 2002

Logged in as:  
Asfandyar Sikandar  
Account #:  
3000925953

[LOGOUT](#)

Copyright © 2002, American Chemical Society

### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)